

# 공간데이터의 시공간적 평가방법에 대한 고찰\*

김영호\*\*

## Space-time Analyses of Spatial Data: A Review\*

Youngho Kim\*\*

**요약 :** 공간데이터의 시공간 군집현상에 대한 분석은 근래에 들어 많은 분야에서 관심을 보이고 있지만 국내에서는 아직까지 연구사례가 제한된 상황이다. 공간현상의 시간의 흐름에 따른 군집을 찾아내는 것은 지리학의 주요한 연구분야 중 하나이며, 여러 인접 학분 분야에 많은 영향을 미칠 수 있는 연구분야이다. 이 연구는 공간 현상의 대표적 시공간 분석방법을 요약, 소개한다. 이 과정에서 공간현상의 기본 가정과 공간데이터의 특징, 그리고 공간현상의 형태 및 개념에 대해 정리한다. 이 연구는 국내 공간분석의 지평을 확대에 기여할 것으로 기대된다.

**주요어 :** 공간군집, 시공간 군집, 공간데이터, 핫스팟

**Abstract :** Space-time analysis of geographical data has recently been focused in many academic fields. However, research literature in Korea has been relatively limited in the fields. Finding spatial cluster (clustering) over time has been important subject in geography, and has a potential to make an impact on other fields. This study introduces and summarizes major space-time cluster (clustering) measures in spatial analysis. In the process this study explains characteristics in spatial data, basic assumption in spatial point pattern analysis, and major concepts and definitions of cluster (clustering). This study aim to contribute to extends Korean literatures on spatial data analysis.

**Key Words :** Spatial cluster/clustering, Space-time cluster/clustering, Spatial data, Hotspot

### I. 서론

현대에 들어서 지리학 뿐만이 아니라 많은 인접학문 분야에서 공간현상을 계량적으로 분석하는 연구가 진행되고 있다(Goodchild and Haining, 2004). 공간을 대상으로 하는 연구는 연구 주제에 따라 다양한 방법으로 진행되었지만, 크게 다음의 네 가지로 분류될 수 있다(Fotheringham and Rogerson, 2009; O'Sullivan and Unwin, 2010): 1) 대규모의 데이터를 정리하여 의미 있는 정보를 추출하거나 데이터를 시각화하는 분야, 2) 탐

색적 분석법으로 데이터의 속성을 파악하여 가설을 만들거나 특이값의 존재 여부를 확인하는 분야, 3) 공간 객체의 공간적 분포를 평가하여 이들이 무작위로 존재하는지, 아니면 군집을 형성하는지의 공간적 패턴을 분석하는 분야, 4) 공간 현상의 변화과정을 계량적으로 모델링하여 공간객체의 속성값을 예측하고 공간이용과 관련한 최적의 의사결정을 도출하는 분야.

이렇게 다양하게 진행되는 공간분석 방법에서 이 연구는 공간 객체의 분포를 평가하여 군집을 판단하는 핫스팟(hot spot) 모델링을 대상으로 한다. 일반적으로 핫

\*이 연구는 2012년 연구된 국토연구원의 보고서를 일부 수정 보완한 것임.

\*\*고려대학교 사범대학 지리교육과 조교수 (Assistant Professor, Department of Geography Education, College of Education, Korea University, younghokim@korea.ac.kr)

스팟은 특정 현상의 공간적인 밀집지역을 말하며, 다른 표현으로는 공간현상의 군집(cluster)이라고도 한다. 공간적인 군집에 관한 연구는 많은 문헌에서 잘 정리되어 있다(Bailey and Gatrell, 1995; Cressie, 1991; Elliott *et al.*, 1992; Elliott *et al.*, 1999; Waller and Gotway, 2004). 이 논문에서는 이러한 공간 군집 현상에 대한 전반적인 리뷰 보다는 시공간 군집 현상에 대한 모델을 정리하여 요약, 소개하고자 한다.

공간적 군집 현상을 찾아내어 설명하는 것은 공간을 연구의 대상으로 하는 지리학에서는 매우 중요하고 의미 있는 연구 주제이다. 범죄의 경우에는 범죄 군집 지역을 확인하는 것이 지역 경찰인력의 배치에 도움이 되며(Ratcliffe, 2004), 경찰의 순찰 경로를 결정하는데 참조되기도 한다(Maguire, 2000). 또한 의료, 질병과 관련한 연구에서는 유명한 사례로 John Snow가 사망자의 공간적 군집을 찾아 질병의 원인인 오염된 식수원을 찾는 것에서(Johnson, 2006), 질병과 관련한 지역적 불균형을 찾아내고 의료서비스의 지역적 형성성에 대한 평가까지(Gatrell and Loytonen, 1998), 다양한 분야에서 공간적 군집현상이 이용된다.

근래에 들어서 많은 연구 결과들이 축적되고 있음에도 불구하고 현실적으로 국내에는 공간적 군집은 물론이고 시공간 군집 현상을 소개하는 문헌이 극히 부족한 현실이다. 이 연구는 국내의 시공간 군집현상의 연구가 제한적인 이유를 관련 문헌의 제약에 있다고 판단한다. 따라서 이 연구에서는 현재 많이 이용되는 시공간적 군집현상을 요약 비교 설명한다. 시공간적 군집 현상을 설명하는 과정에서 공간현상의 기본 가정과 공간 데이터의 특징을 간략히 소개하며, 군집의 기본적인 정의, 그리고 시공간적 군집현상을 설명한다.

## II. 공간현상의 기본 가정과 공간데이터의 특징

범죄와 같이 공간상에서 포인트로 표현되는 공간현상은 공간적 분포의 기본 가정을 확인할 필요가 있다. 연구지역에서 포인트로 표현되는 공간 현상의 기본 가정은 CSR(complete spatial randomness) 또는 IRP(independent random process)로 정의된다. 이 기본 가정은 두 가지 조건을 모두 만족하는 형태로 제시될 수 있는데, 첫째,

각각의 사건들은 동일한 확률로 발생해야 하며, 둘째, 개별 사건의 발생 장소는 다른 사건의 영향을 받지 않는다는 것이다(Bailey and Gatrell, 1995). 통계적으로 표현될 때 CSR(또는 IRP)는 균일한 포와송 분포(homogeneous spatial randomness)를 통해 구현된다. 서울시 범죄 발생을 예로 든다면 CSR에서는 서울시의 모든 행정구에서 동일한 범죄율을 갖게 되며, 범죄발생패턴이 특정 지역에서 비슷하게 나타나거나 너무 상이하게 나타나는 패턴이 발견되지 않는다는 것을 말한다. 즉 CSR은 숲에서 나무가 자라는 것과 같이 범죄 발생이 독립적으로 그리고 무작위적으로 일어나는 것을 말한다.

실제적으로 공간 현상들은 대부분 앞에서 제시된 두 가지 조건을 만족시키지 않으며 그러한 특징이 두 가지로 제시된다(Bailey and Gatrell, 1995). 첫째는 일차 영향효과(first order effects)로 공간현상의 평균값이 변화하는 것을 말하며, 전체적, 대규모의 경향으로 설명될 수 있다. 둘째는 이차 영향효과(second order effects)로 공간적 상관구조를 말한다. 이는 특정 지역의 값이 인접한 주변지역의 영향을 받는 것으로 - 국지적, 그리고 소규모의 효과로 설명된다. 이러한 공간현상의 특징을 범죄의 예로 들어 설명한다면, 빈민촌, 치안부재지역과 같은 대규모로 범죄를 양산하는 효과로 인해 특정 지역에 범죄율이 다른 지역보다 확연하게 높게 나타나는 것은 공간현상의 일차 영향효과로 설명이 가능하다. 그리고 이런 우범지역에서 형성된 범죄자들이 다른 지역에서도 범죄활동을 해서 범죄자들의 거주지역 주변으로 범죄율이 높게 변화하는 것은 이차 영향효과로 설명이 가능하다.

우리나라의 경우, 강원랜드와 같은 내국인을 대상으로 하는 대규모 카지노는 그와 관련한 많은 부작용 가운데 대규모의 범죄를 양산한다. 실제로 강원랜드 인근의 5대 범죄 발생 증가율이 서울 서초의 6배에 이르는 것으로 국정감사 결과 나타났다(중앙일보 2011.9.30). 이는 공간적으로 평가할 때 공간현상이 대규모의 전역적 변화를 나타내는 일차 영향효과로 설명할 수 있다. 강원랜드에서 범죄를 일으키는 범죄자들이 영월이나 태백으로 이동하여 인근지역의 범죄율을 높이는 것은 이차 영향효과로 해석할 수 있다. 이론적으로 공간현상의 일차, 이차 영향효과는 명확하게 설명이 가능하지만, 실제적인 공간 데이터에서 이러한 일차, 이차 영향효과는 혼재되어 나타나며, 이 두 효과를 분리하기는 쉽지 않다(Haining, 1990). 일차가 공간상의 구조적인 큰 요인에

의해서 설명되는 반면, 이차는 인구와 같은 개별적인 작은 요인에 의해 형성되기 때문에 일차와 이차 영향효과는 혼재되기 마련이다. 또한 이차 영향효과는 이동성 있기 때문에 주변지역으로의 파급효과가 있으며 명확한 공간적인 범위를 측정하는 것이 쉽지 않다.

### III. 군집현상의 형태 및 개념: 클러스터(cluster)와 클러스터링(clustering)

공간현상의 핫스팟을 나타내는 클러스터와 클러스터링은 엄격하게 다른 현상을 나타내지만, 비슷한 표현으로 혼용되는 경우가 많다. 클러스터와 클러스터링의 엄격한 구분은 Besag and Newell(1991)에 의해 처음으로 제시되었고 Waller와 Gotway의 저서(2004)에 잘 정리되어 있는데 간단하게 요약하면 다음과 같다. 클러스터는 국지적 공간 패턴을 나타내며, 연구지역 내의 특정 케이스들이 일반적인 공간적 분포의 기본 가정과 다른 경우를 일컫는다. 공간적 분포의 기본 가정은 앞서 설명한 CSR로 표현할 수 있으며, 결국 클러스터를 찾는 것은 연구지역에서 통계적으로 가장 유의미한 집단으로 구성된 국지적 핫스팟을 찾아내는 것이라 설명할 수 있다. 따라서 하나의 클러스터는 데이터 내에서 다른 데이터들과 차이가 나는 특징적인 케이스들을 말한다.

클러스터링은 전역적 공간 패턴을 나타내며, 공간현상이 전반적으로 밀집되는 경향을 나타낸다(Waller and Gotway, 2004). 따라서 클러스터링은 클러스터와 같이 특정한 케이스들을 찾아서 제시하는 것이 아니라 데이터를 구성하는 모든 케이스의 전반적인 패턴으로 나타난다. 클러스터링은 공간현상의 전역적 패턴을, 클러스

터는 특정한 케이스들로 구성된 국지적 패턴을 나타내기 때문에, 통계적인 방법으로 분석할 때의 특이 명확히 구별된다. 클러스터링은 통계적 유의성을 검증하기 위해, 전체 연구지역을 대상으로 단 한번의 평가를 진행한다. 하지만 클러스터는 어떠한 케이스들로 통계적으로 유의미한 클러스터가 형성되는지 확인하기 위해 여러 번의 평가를 진행한다. 따라서 범죄 데이터의 경우 일반적으로 클러스터를 찾아내는 방법들이 많이 적용된다. 클러스터링을 확인하는 방법들은 전체적인 공간적 분포의 특징을 찾기 위해 데이터 요약의 측면에서 이용되는 경향이 강하다.

클러스터와 클러스터링 테스트를 더 세분화 하면, 일반테스트(general test)와 집중테스트(focused test)로 나눌 수 있다. Besag과 Newell(1991)에 의하면 일반테스트는(클러스터와 클러스터링 테스트를 포함하여) 연구지역 어느 곳이든 위치의 제한 없이 핫스팟을 찾는 것을 말한다. 반면 집중테스트는 미리 선정된 지역(들)을 기준으로 이 지역에서 핫스팟이 형성되는지를 확인한다. 일반테스트와 집중테스트의 통계적인 차이는 귀무가설(null hypothesis)의 설정에서 나타난다(Waller and Gotway, 2004). 일반테스트는 귀무가설을 연구지역 모든 곳에서 클러스터와 클러스터링이 없는 것을 정하지만, 집중테스트는 특정 지역을 중심으로 한 곳에서 클러스터가 존재하지 않는다는 것을 귀무가설로 설정한다. 예를 들면 '서울시 전역에서 절도가 밀집한 우범지역은 존재하지 않는다'가 일반테스트의 귀무가설에 해당 된다면, '서울시에서 전당포 인근지역에는 절도가 밀집하지 않는다'가 집중테스트의 귀무가설이라 할 수 있다. 따라서 공간상 존재하는 핫스팟 테스트는 공간적인 규모에 따라 클러스터, 클러스터링으로 구분할 수 있으며, 미리 선정된 핫스팟의 요인이 있느냐 없느냐에 따라 일반

표 1. 공간핫스팟 테스트 모델의 구분

	클러스터 테스트(cluster test)	클러스터링 테스트(clustering test)
일반테스트 (general test)	<ul style="list-style-type: none"> <li>• Geographical Analysis Machine (Openshaw and Charlton, 1987)</li> <li>• cluster evaluation permutation statistic (Tumbull et al., 1990)</li> <li>• Spatial scan statistic (Kulldorff and Nagarwalla, 1995)</li> <li>• Local Indicators of spatial autocorrelation (Anselin, 1995)</li> <li>• Local Statistics Model (Getis and Aldstadt, 2004)</li> </ul>	<ul style="list-style-type: none"> <li>• Moran's I (Moran, 1948)</li> <li>• Geary's c (Geary, 1954)</li> <li>• K-function (Ripley, 1976)</li> <li>• Mantel test (Mantel, 1967)</li> </ul>
집중테스트 (focused test)		<ul style="list-style-type: none"> <li>• Score test (Lawson, 1993; Waller et al., 1992)</li> <li>• Stone test (Stone, 1988)</li> </ul>

테스트와 집중테스트로 구분할 수 있다. 표 1 은 핫스팟 테스트 모델을 특성에 따라 구분한 것이다.

#### IV. 핫스팟의 시·공간적 평가방법

공간적 핫스팟은 특정현상의 공간적 밀도가 높은 것, 따라서 개별 케이스 사이의 거리가 상대적으로 가까운 것을 말한다고 할 수 있다(Bailey and Gatrell, 1995). 이러한 핫스팟의 개념을 시·공간으로 확장하면 다음과 같이 설명될 수 있다. 시간도 공간과 같은 거리(시간거리)로 측정하여 멀고 가까움을 비교한다면, 시·공간의 상호작용에 따라 케이스들 사이의 시간거리와 공간거리 모두가 밀집되어 있는 경우는 시·공간적 핫스팟이라고 할 수 있다.

그림 1은 시·공간 클러스터를 개념적으로 표현한 것이다. 공간이 X,Y좌표로 표현되며 공간상의 거리는 수평으로 표현된다. 시간은 Z좌표로 표현되며, 시간상의 거리는 수직으로 표현된다. 따라서 시·공간 클러스터는 케이스들 사이에서 수평과 수직의 거리가 모두 가까운 값으로 구성된다. 그 결과 시·공간 클러스터는 분석 결과로서 특정 공간에서 특정 시간대에 클러스터가 확인된 것을 보여준다.

핫스팟의 공간적 평가방법과 마찬가지로, 시·공간적 평가방법도 독립성, 무작위성에 기초하여 평가한다. 시·공간적 평가방법 또한 클러스터와 클러스터링 테스트, 그리고 일반테스트와 집중테스트로 공간적 평가방법과 동일하게 구분할 수 있다. 하지만 핫스팟의 시·공간적 평가방법은 데이터의 시간적 범위와 연속성의 여부에 따라 소급적 분석방법(retrospective analysis)와 전망적

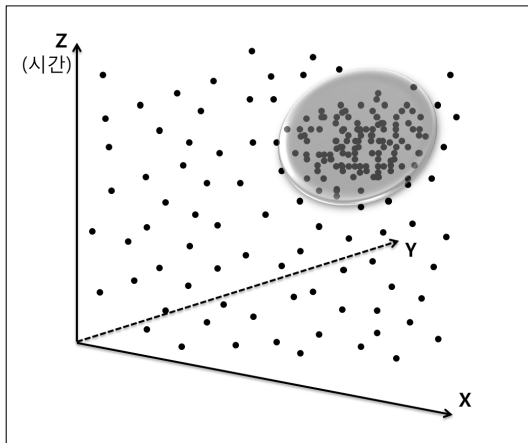


그림 1. 시공간 클러스터의 3D표현

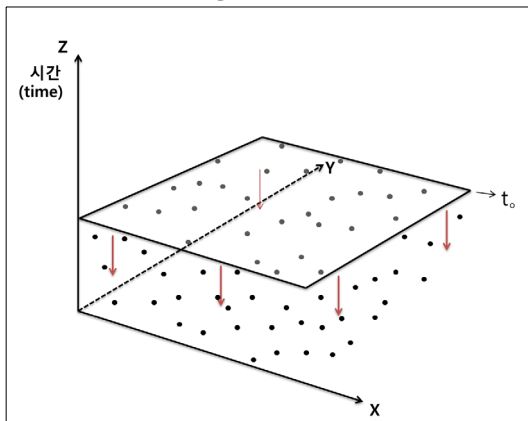


그림 2. 핫스팟의 시공간적 평가방법 분류: 소급적 분석방법

분석방법(prospective analysis) 두 가지로 구분이 가능하다(Sonesson and Bock, 2003). 그림 2와 3은 핫스팟의 시·공간적 평가방법을 분류한 것이다.

그림 2에서 제시한 소급적 분석방법(Retrospective analysis)은 과거의 고정된 데이터를 대상으로 한다. 따라서 사건의 진행이 완료된, 즉 시·공간적으로 제한된 범위에 있는 데이터를 연구대상으로 한다. 분석은 단 한번의 테스트를 통해서 시·공간적 핫스팟을 찾아내는 것으로 진행된다. 분석 결과는 과거의 특정 지점과 시간에서 핫스팟을 발견하는 것으로 나타난다. 일반적으로 소급적 연구방법으로 분석한 핫스팟은 다른 시·공간적 범위에 위치한 핫스팟과 비교하는 목적으로 사용된다.

그림 3에서 보여주는 전망적 분석방법(Prospective analysis)은 과거의 시·공간 데이터에 최신의 새로운 시·공간 데이터가 연속적으로 업데이트되어 실시간에 가까운 상태로 분석이 진행되는 시·공간적 핫스팟 평가방법을 말한다(Rogerson, 1997). 그림 3에서는 데이터가 업데이트됨에 따라  $t_0, t_1, t_2$ 의 개별 시간에서 연속적으로

분석이 진행되었다.

실시간에 가까운 데이터의 업데이트와 그에 따른 실시간에 가깝게 반복되는 지속적인 분석을 특징으로 하기 때문에 전망적 연구방법은 범죄나 질병과 같은 지속적인 모니터링이 필요한 부분에 많이 이용된다. 전망적 연구방법의 가장 큰 장점으로서는 연속적인 관찰이 가능하기 때문에 어떤 공간적으로 위험한 핫스팟이 형성되었을 때, 즉각적으로 확인하여 신속한 대응을 가능하게 한다는 장점이 있다(Elliott and Wartenberg, 2004). 따라서 전망적 연구방법은 연속적으로 데이터를 수집하고, 분석하고, 결과 해석하며, 결과에 따른 예방과 통제와 같은 지속적인 대응을 하는 일련의 과정을 의미한다(Kim and O'Kelly, 2008). 따라서 지속적인 감시나 관찰을 의미하는 서베이런스(surveillance)의 개념으로 많이 적용되며, 연속성의 개념은 시·공간적 연구대상의 속성에 따라 일, 주, 월 단위 등으로 바뀐다. 표 2는 앞으로 언급될 시공간적 핫스팟 분석방법을 분류하였다.

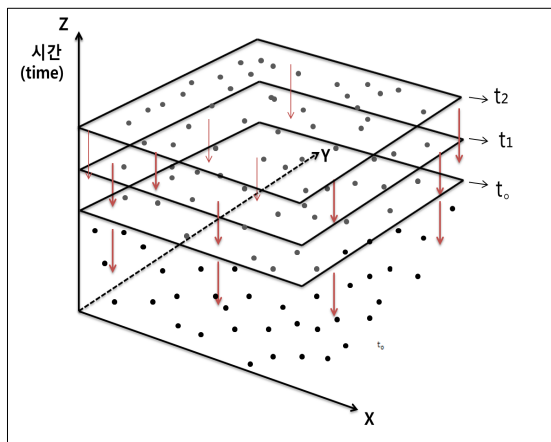


그림 3. 핫스팟의 시공간적 평가방법 분류 : 전망적 분석방법

표 2. 시공간적 핫스팟 분석방법과 분류

연구방법	클러스터 테스트 (cluster test)	클러스터링 테스트 (clustering test)
소급적 연구방법 (Retrospective analysis)	-	<ul style="list-style-type: none"> <li>• Nearest neighbor test (Cuzick and Edwards, 1990)</li> <li>• Knox test (Knox, 1964)</li> </ul>
전망적 연구방법 (prospective analysis)	<ul style="list-style-type: none"> <li>• Space-time scan (Kulldorff, 2001)</li> <li>• Permutation scan (Kulldorff et al., 2005)</li> <li>• Bootstrap-based surveillance model (Kim and O'Kelly, 2008)</li> </ul>	<ul style="list-style-type: none"> <li>• Variants of CUSUM test (Kleinman et al., 2006; Rogerson, 2005)</li> <li>• Jarpe test (Jarpe, 1999)</li> </ul>

## 1. 소급적 연구방법

### 1) Knox 테스트

핫스팟의 시·공간적 평가방법 중, Knox 테스트(Knox, 1964)는 대표적인 소급적 연구방법 모델이라 할 수 있다. Knox테스트는 개별 시·공간 데이터의 상호작용의 거리와 빈도를 조사하여 상호작용의 시·공간적 거리가 한계수준 보다 가까울 경우 시·공간 클러스터로 확인하는 테스트 방법이다. 자연히 Knox테스트는 미리 설정된 시·공간적 거리의 한계수준이 필요하다. 그림 4는 Knox 테스트의 개념도를 나타낸다.

그림 4의 산포도에서 표현되는 점들은 시·공간에 위치한 한 쌍의 케이스들을 나타내며, 산포도는 이들의 시간거리(Y축)와 공간거리(X축)를 표현한 것이다. 그림에 있는 굵은 선은 시간거리와 공간거리의 가깝고 먼 것을 구분하는 한계거리를 나타내는 한계선에 해당한다. 그림에서 수평선은 시간거리의 멀고 가까움을 구분하며, 수직선은 공간거리의 멀고 가까움을 나눈다. 따라서, 그림의 A 지역에 해당하는 케이스들은 공간적인 클러스터에, C지역에 해당하는 케이스들은 시간적인 클러스터에, D지역에 포함된 케이스들은 시·공간 클러스터에 해당한다.

Knox 테스트는 다음과 같이 정의된다.

$$X = \sum_{i=i}^N \sum_{j=1}^{i-1} s_{ij} t_{ij}$$

이 공식에서  $S_{ij}$ 는 공간적 근접성을 나타내는 변수로,

이벤트  $i$ 와  $j$ 가 한계거리 이내에 있으면 1로, 그렇지 않을 경우에는 0으로 표현된다.  $t_{ij}$ 는 시간적 근접성을 나타내며, 공간과 동일하게  $i$ 와  $j$ 가 한계거리 이내에 있으면 1로, 그렇지 않을 경우에는 0으로 표현된다.

앞서 공간현상의 기본 가정에서 언급되었듯이 Knox 테스트의 귀무가설은  $X$ 가 공간적으로 독립되어 있는 포와송 분포를 따르며 기댓값은  $ST/n(n-1)$ 로 정의된다. 여기서  $ST$ 는 시·공간의 거리가 한계치 보다 작은 값들을 케이스들을 나타낸다. 따라서  $X$ 가 귀무가설의 포와송 분포에서 현격하게 큰 값을 가질 때에는 귀무가설이 기각되며, 케이스들이 시공간적으로 클러스터링을 형성한 것으로 해석된다.

Knox테스트는 시·공간 핫스팟의 탐색과 관련하여 많은 장점을 가지고 있으며 많은 연구에 응용되었다. Kuldorff와 Hjalmar에(1999) 따르면 Knox테스트는 배경인구에 관한 추가 정보가 없이도 충분한 분석이 가능하며 모델의 단순성을 장점으로 언급했다. Knox테스트는 이후 많은 학자들에 의해 확장되고, 응용 적용되었다. Baker는(1996) 시·공간의 한계거리를 설정하는 방법을 소개했으며, Diggle 등은(1995) 시·공간 클러스터링과 관련하여 K-function이 어떻게 시·공간 의존성을 나타낼 수 있는지 제시하였다. Mantel은(1967) Knox 테스트가 배경인구에 대한 고려를 하고 있지 않기 때문에, 시간이 경과하면서 인구가 이동하면 Knox 테스트의 결과는 편향될 수 밖에 없음을 지적했다. Kuldorff와 Hjalmar는 이러한 문제점을 해결하기 위해, 수정된 버전의 Knox 테스트를 제시했다. Rogerson(2001)은 Knox 테스트를 국지적 모델로 변형하여 감시 모델에 응용할 수 있게 하

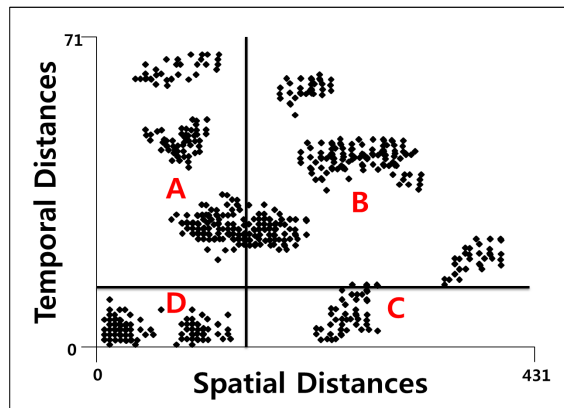


그림 4. Knox 테스트의 개념도

였다. 마지막으로 O'Kelly(1994)는 Knox 테스트에서 아이디어를 차용하여, 최소의 시·공간 거리를 산출하는 방법으로 클러스터의 중심점을 찾는 과정에서 최적의 분할법을 제시하였다.

## 2) Mantel 테스트

Mantel테스트(1967)는 기존의 Knox테스트와 비교할 때 시·공간의 한계치를 연구자가 인위적으로 조작할 필요가 없다는 장점을 가진다. 이 테스트는 연구대상의 모든 조합에서 산출한 시간거리와 공간거리를 곱하는 과정에서 이 두 가지 거리의 상관관계를 찾아낸다. Mantel 수식은 일반 수식과 표준화된 수식의 두 가지로 표현이 가능하다. 표준화 수식은 관찰값이 많을 경우에 평균은 0, 표준편차 값이 1인 표준정규분포를 따르게 한다는 장점이 있다.

(일반수식)

$$Z = \sum_{i=1}^N \sum_{j=1}^N s_{ij} t_{ij}$$

(표준화 수식)

$$r = \frac{1}{N^2 - N - 1} \sum_{i=1}^N \sum_{j=1}^N \frac{(s_{ij} - \bar{s})}{s_s} \frac{(t_{ij} - \bar{t})}{s_t}$$

이 식에서  $\bar{s}$ 와  $\bar{t}$ 는 각각 시간과 공간 거리의 평균값을 나타낸다. 그리고  $s_s$ 와  $s_t$ 는 시간과 공간 거리의 표준편차를 나타낸다. 따라서 표준화 수식은 시간거리와 공간거리 평균0, 표준편차가 1로 변환시킨 값을 사용하며, 결과적으로 결과값이 표준정규분포를 따른다. Mantel 테스트의 귀무가설은 시·공간 거리의 독립성을 가정하며, 이러한 귀무가설은 표준정규분포에 기초한 통계적 유의성을 통해 확인할 수 있다. 또는 몬테카를로 시뮬레이션을 통해서 더 엄격한 통계적 유의성 검증이 가능하다.

Mantel 테스트는 상관관계의 값으로 주어지기 때문에 테스트 값이 통계적으로 유의할 경우에는 시간거리와 공간거리의 유사성이 높은 것으로 해석할 수 있다. 이는 시·공간 거리의 독립성이라는 기본적인 귀무가설을 확인하지만, 실질적인 해석의 과정에서는 주의를 필요로 한다. 시간거리와 공간거리가 모두 크거나, 모두 작은 두 가지 경우로 해석이 가능하기 때문이다. 일반적으로 시·공간적 클러스터링은 시·공간적 거리가 가까운 경우

에 해당하기 때문에 분석 과정에서 두 가지 경우 중 어느 쪽에 해당하는지 추가적인 확인절차가 필요하다.

결과값의 유의성 검증과정에서 연구자가 시공간 거리의 한계치를 인위적으로 조작할 필요가 없다는 것은 Mantel 테스트의 가장 큰 장점이다. Mantel 테스트는 Knox테스트와 같이 배경인구를 고려하지 않기 때문에 오랜 시간의 경과로 배경인구가 변화했을 때에는 왜곡된 결과를 도출할 수 있다.

Mantel 테스트는 시·공간 상호작용에 관한 연구에서 다양하게 이용되었다. 특히 범죄의 시·공간 발생패턴과 관련해서는 Johnson과 Bowers(2004)의 연구가 있다. 강도사건의 시·공간적 분포 패턴을 조사한 이 연구는 범죄의 시·공간적 상호작용의 연속성에 초점을 맞췄다.

## 3) k-Nearest Neighbor(k-NN) 테스트

k-NN 테스트(Jacquez, 1996)는 공간적으로 가까운 거리의 케이스가 시간적으로도 가깝게 위치했는지 확인해 준다. 개별데이터와 시·공간 거리를 이용한다는 점에서 k-NN테스트는 Knox테스트나 Mantel테스트와 유사하지만, 상호작용이 일어나는 스케일을 조정할 수 있다는 점에서 차이점이 부각된다. k-NN테스트는 시·공간 스케일을 정의하는 방식에서 Knox, Mantel 테스트와 다르다. Knox와 Mantel테스트가 절대적인 시간거리, 공간거리로 시·공간 스케일을 정의하는데 반해, k-NN테스트는 절대적인 거리보다는 연구자가 미리 선정한 k근린 케이스의 수 만큼 시·공간적으로 가까운 케이스들을 대상으로 한다. 예를 들어 미리 k근린 케이스의 기준을 3으로 지정했을 경우, 각각의 시간거리와 공간거리가 가장 가까운 3개씩의 케이스가 분석대상에 해당된다. 따라서 시·공간적으로 사건의 밀도가 높은 경우에는 k-NN테스트의 이용되는 케이스들의 절대 시·공간 거리가 가까워지며, 밀도가 낮은 경우에는 반대이다.

$$J_k = \sum_{i=1}^N \sum_{j=1}^N s_{ijk} t_{ijk}$$

시·공간 클러스터링이 존재할 때, 공간적으로 최근린에 해당하는 케이스들은 시간적으로도 k근린에 해당하기 때문에  $J_k$ 는 큰 값을 가지게 된다. 이 공식에서 k는 미리 지정된 기준에 해당하며,  $s_{ijk}$ 는 공간적인 k근린 지표( $s_{ijk} = 1$  공간적 k근린에 해당하는 경우,  $s_{ijk} = 0$

그렇지 않은 경우), 그리고  $t_{ijk}$ 는 시간적인  $k$ 근린 지표 ( $t_{ijk}=1$  시간적  $k$ 근린에 해당하는 경우,  $t_{ijk}=0$  그렇지 않은 경우)에 해당한다. K-NN테스트의 통계적인 유의도 검사는 몬테카를로 시뮬레이션을 통해서 진행된다. K-NN테스트가  $k$ 근린 지표에 기초한 클러스터링 테스트 방법임을 볼 때, 이는 Mantel 테스트의 멀고 가까운 거리의 정도를 비율로서 나타내는 변형이라고도 볼 수 있다.

통계적으로 표현되는 구조는 다르지만, Knox 테스트, Mantel 테스트, 그리고 k-NN테스트는 케이스들의 시·공간적 거리에 기초한 상호작용을 평가한다는 측면에서 공통점이 있으며 시간적으로 고정된 데이터를 사용한다는 특징이 있다. 그 결과 이 테스트들은 과거에 형성된 시·공간적 클러스터링은 찾아낼 수 있지만, 지속적으로 업데이트되는 데이터를 이용하여 현재 진행중인 클러스터링을 찾는 데에는 제약이 있다. 일례로 Knox테스트는 미리 지정된 시·공간적 한계치라는 기준을 가지고 있으며, 이러한 기준을 규칙적으로 업데이트 되는 새로운 데이터에 적용하기에는 문제가 있다. Mantel, k-NN 테스트의 경우에도 시·공간 적으로 고정된 데이터를 사용하기 때문에 연속적으로 업데이트되어 추가되는 정보를 분석할 수 없다.

## 2. 전망적 연구방법

일반적으로 전망적 연구방법은 기존의 데이터에 연속적으로 업데이트되는 자료들을 분석하여 클러스터/클러스터링을 찾아낸다. 비교적 근래 까지는 전망적 연구방법이 주로 시간에 국한된 감시기능으로 진행되어 왔다 (Strat, 2005). 가령 시간에 국한된 연구방법의 예로 Shewhart차트(Shewhart, 1931)는 공장의 생산라인에서 단위시간당 발생하는 불량품의 개수를 모니터하는 과정에서 기준치보다 많은 불량품이 발생하는 경우에 알람을 보내는 방법이다. 이렇게 시간에 국한된 전망적 연구방법은 연속적인 분석이 가능하다는 장점은 가지고 있었지만 공간적인 위치 정보를 담아낼 수 없었기 시·공간을 대상으로 하는 소급적 연구방법과 비교하여 주목 받지 못했다(Farrington and Beale, 1998). 이 부분에서는 시간에 국한된 전망적 연구방법으로 Shewhart차트들과 국지적 Knox 테스트, 최근린 테스트로 확장된 CUSUM 모델, 그리고 Tango 통계로 확장된 CUSUM모델을 소개한다.

### 1) Shewhart차트

Shewhart차트(1931)는 가장 단순한 형태의 관리도로 관찰값이 미리 지정된 한계치보다 커질 때 경보를 올린다. 이것은 현재 진행중인 과정 값이 과거의 수준 보다 높아져서 주의를 필요로 한다는 것으로 해석할 수 있다. 일반적으로 한계치는 연구 대상의 평균값에 표준편차의 일정배수를 더해서 설정한다.

Shewhart차트가 단순히 과거의 최종 관찰값을 사용하는데 반하여, 이를 확장시킨 관리도로 Moving average 차트(Stern and Lightfoot, 1999)가 있다. Moving average 차트는 최종 하나의 관찰값이 아닌 일정한 기간에 동안의 여러 관찰값을 사용한다.

$$y_t = \frac{1}{m} \sum_{i=0}^{m-1} x_{t-i}$$

여기서  $m$ 은 평균을 구하는데 사용된 관찰값의 개수를 말한다. 일반적으로  $m$ 의 크기가 클수록 차트가 안정되어 평균에서 벗어난 작은 변화에도 민감하게 반응하기 때문에 전체적으로 알람의 빈도가 증가하게 된다. 이처럼  $m$ 의 크기가 알람의 빈도에 영향을 미치기 때문에 적절한  $m$ 값을 선택하는 것이 중요하다.

Moving average차트가 복수의 관찰값에 동일한 가중치를 부여하는데 반하여, exponentially weighted moving average(EWMA) 차트(Hunter, 1986)는 값의 관찰시점에 따라 가중치를 다르게 부여한다. 최근의 관찰값에는 상대적으로 많은, 그리고 오래된 관찰값에는 적은 가중치를 부여하여 기대값을 산출한다.

$$y_t = \lambda x_t + (1 - \lambda)y_{t-1} \\ = \lambda \sum_{i=1}^n x_{t-i} (1 - \lambda)^j$$

여기서  $\lambda$ 는 가중치를 정하는 가중치 산출변수이며,  $t=0$ 인 경우에는 과거에 해당하는 관찰값이 없기 때문에  $y_0=0$ 으로 정의한다. 그리고 관찰값이 쌓여갈수록 가중치가  $(1 - \lambda)$ 의 승수로 누적되기 때문에, 위의 수식을  $y_t = \lambda x_t + (1 - \lambda) [\lambda x_{t-1} + (1 - \lambda) \{ \lambda x_{t-2} + (1 - \lambda)(\dots) \}]$ 로 풀어서 정리할 수 있다. 가중치 산출변수  $\lambda$ 가 커질수록, 과거의 관찰값은 적은 영향을 미치며, 최근의 관찰값이 더 많은 영향을 미치게 된다. 일반적으로  $\lambda$ 의 값은 0.1~0.5사이의 값이 이용된다.



## 2) Cumulative Sum 모델

Page에 의해 소개된 Cumulative Sum 모델(1954)은 일련의 과정에 양적인 변화를 측정하는 방법으로 평균으로부터 급격한 변화가 발생할 때 이를 발견하는 모델이다. Cumulative Sum 모델은 일반적으로 귀무가설로 독립정규분포를 가정한다;  $x_t \sim N(\mu_t, \delta_t^2)$ . 이 모델은 새로운 데이터가 결합됨에 따라 반복적으로 정의되며 다음과 같다.

$$y_t = \max\left(0, y_{t-1} + \left(\frac{x_t - \mu_t}{\delta_t} - k\right)\right)$$

$$y_0 = 0$$

여기서  $k$ 는 일반적으로 0.5로 설정되는 패러미터 값이다. Cumulative Sum 모델은 평균값에서 표준편차의  $k$  배 이상으로 차이가 나는 값을 축적해 나간다. 그 결과,  $y_t$  값이 한계치인  $h$  값 보다 커질 때, 과정 중에 유의한 수준의 변화가 발생한 것으로 해석된다.  $h$  값의 설정은 유의한 수준의 변화가 발생하는 평균 기간, 즉 Average Run Length(ARL)에 영향을 미치게 된다.  $h$  값이 클 때에는 ARL의 길게 되며, 이 경우 유의한 수준의 변화가 아닌 데 유의한 수준의 변화로 인식하는 오차의 빈도수가 적게 발생하며,  $h$  값 이 작을 경우에는 그 반대이다. 일반적으로 Cumulative Sum 모델은 정규분포에 기초하여 많이 이용되지만, 수학적인 변형을 통해서 Poisson 또는 Exponential 분포에 적용되기도 한다.

## 3) 최근린 통계법

최근린 통계법(nearest neighbor statistic)(Clark and Evans, 1954)은 최근린 이웃들의 거리의 평균값과 공간적 무작위 패턴에서의 거리의 기대값을 비교하여 클러스터를 확인한다. 이 과정에서 무작위 패턴에서의 거리 기댓값은  $r_e = \frac{1}{2\sqrt{\rho}}$  로 계산된다. 여기서  $\rho$ 는 관찰값의 공간적 밀도를 나타내는 지표로 연구지역의 넓이 ( $a$ )를 관찰값의 개수( $n$ )으로 나눈 값에 해당한다  $\rho = \frac{n}{a}$ . 결과적으로 도출되는 최근린통계값은 관찰 거리값과 기대 거리값의 비율로 다음과 같이 나타난다.

$$R = \frac{r_{obs}}{r_e}$$

결과적으로 공간상의 패턴에서 클러스터링이 존재하는 경우, 관찰된 평균 거리값은 기대 거리값 보다 작아지기 때문에 최근린통계값,  $R$ , 은 1보다 작은 값을 가지게 된다.  $R$ 의 최소값은 0이며, 이 경우는 모든 값들이 한곳에 위치하는 경우에 발생한다. 무작위 패턴에서 발생하는 기대값의 표준편차는  $\delta_{r_e} = \frac{0.26}{\sqrt{n\rho}}$  로 계산되며, 평균과 표준편차를 이용하면 최근린 통계치를 표준정규 분포로 변환할 수 있다. 그리고 이렇게 변환된 값은 Cumulative Sum 모델로 적용이 가능하다.

## 4) 국지적 Knox 테스트

Knox 테스트가 시공간의 클러스터링을 찾아내지만, 국지적 Knox 테스트(Rogerson, 2001)는 시공간의 특정한 클러스터를 형성하는 특정 관찰값을 찾아낸다. 특히 Knox 테스트가 현재로부터 과거의 사건을 일회 분석하는 소급적 연구방법의 특징을 가지고 있는 반면, 국지적 Knox 테스트는 cumulative sum으로 확장되어 연속적인 분석이 가능하기 때문에 전망적 연구방법에 해당한다. 결과적으로 국지적 Knox 테스트는 cumulative sum의 구조에서는 연속적인 감시가 가능하며, 이는 Knox 테스트값의 변화를 측정할 수 있도록 해준다. 통계적으로 유의한 국지적 Knox 값은 특정 관찰값이 다른 관찰값에 시·공간적으로 기대치보다 더 많이 연결되어 있음을 보여준다. 따라서 국지적 Knox 테스트는 어떠한 관찰값이 전역적 Knox 통계치에 시·공간적으로 영향을 미치는지 확인해 준다.

$n_s(i)$ 를 공간상에서 관찰값  $i$ 와 가까운 관찰값의 수라고 정의하고,  $n_t(i)$ 를 시간상에서  $i$ 와 가까운 관찰값, 그리고  $n_{st}(i)$ 를 시공간 모두에서 가까운 관찰값의 수라 정의한다. 여기서  $N_{st}(i)$ 의 분포에 대한 귀무가설이 모든 관찰값들 사이의 시·공간 거리는 동일하다고 가정할 때,  $N_{st}(i)$ 에 해당하는 통계적 분포는 다음과 같이  $n-1, n_s(i), n_t^j(i)$ 의 변수로 구성된 초기하 분포(hypergeometric distribution)의 합으로 표현된다.

$$p\{N_{st}(i) = n_{st}(i)\} = \frac{1}{n} \frac{\sum_{j=1}^n \binom{n_t^j(i)}{n_{st}} \binom{n-1-n_t^j(i)}{n_s(i)-n_{st}}}{\binom{n-1}{n_s(i)}}$$

여기서  $n_t^j(i)$ 는 시간적으로  $i$  관찰값에 가까운 관찰값

의 수이며, 기대값과 분산은 다음과 같은 정의 된다.

$$E[N_{st}(i)] = \frac{2n_t n_s(i)}{n(n-1)}$$

$$V[N_{st}(i)] = \frac{2(n-1)n_t - \sum_{j=1}^n n_t^j(i)^2 n_s(i) \{n-1-n_s(i)\}}{n(n-1)^2(n-2)}$$

이렇게 정의된 기대값과 분산을 이용하여 시·공간적으로 가까운 관찰값을 나타내는  $n_{st}(i)$ 는 정규분포로 변환이 가능해진다.

$$z_i = \frac{n_{st}(i) - E\{N_{st}(i)\} - 0.5}{\sqrt{V\{N_{st}(i)\}}}$$

지금까지 소개된 전망적 연구방법은 기존의 데이터에 새롭게 업데이트되는 데이터를 결합하여 연속적으로 변화하는 기대값을 반영하여 연속적인 감시와 분석이 가능하다는 특징을 가지고 있다. 따라서 전망적 연구방법은 시간의 흐름에 따라 변화하는 범주의 경향이나 특징을 잡아내는 목적에 소급적 연구방법에 비교하여 적합하다. 특히 Shewart 차트, moving average 차트, 그리고 Exponentially weighted moving average 차트는 시간의 흐름에 따른 변화량을 측정하는데 초점을 맞춘다. 기존에 관찰된 값을 토대로 기대값을 산출하고 새롭게 관찰되는 값을 기대값과 비교하는 측면에서 이 세가지 차트는 공통된 특징을 보여준다. 하지만 이 접근법들은 시간의 흐름에 따른 변화량과 그 패턴을 찾아내는 목적에는 적합하지만, 공간적으로 정확한 위치와, 그에 따른 변화의 공간적 패턴을 보여줄 수 없다는 단점을 가지고 있다. 최근된 통계법과 국지적 Knox 테스트는 Cumulative sum의 형식으로 관찰값들의 변화의 연속성을 측정한다는 면에서 전망적 연구방법의 특징을 가지면서도, 정확한 위치정보를 제공하기 때문에 공간적 변화 패턴을 제시한다는 차이점이 있다.

## V. 요약 및 결론

공간현상의 패턴을 연구하는 군집현상은 넓은 응용가능성으로 앞으로도 지리학을 포함한 많은 인접분야에서 다양하게 연구될 것으로 기대된다. 특히 이 연구의 주요

대상인 시공간 군집 현상은 미래의 공간분석에 더 많이 이용될 것으로 기대된다. 지리학의 영역은 앞으로는 단순한 공간적 측면에서 확대되어 시간의 측면을 포함할 것으로 기대되며, 이 과정에서 시공간 분석에 대한 연구의 필요성은 매우 높다(Florax and Vlist, 2003). 본 논문의 목적은 다양한 시공간 연구에 대한 개괄적인 소개가 아니라, 시공간 연구에서도 응용분야가 많은 시공간 군집분석에 대한 최근의 모델을 심층적으로 정리하는 것이다. 이 과정에서 이 논문은 공간현상의 기본 가정과 공간데이터의 특징을 정리했으며, 군집의 형태를 연구 지역 내의 군집 규모에 따라 정의하였다. 또한 시공간 군집현상을 분석 모델의 시간적 방향성에 따라 소급적 분석방법과 전망적 분석방법으로 분류하여 소개하였다.

인문, 자연을 포괄하는 공간현상은 고정된 것이 아니라 연속적으로 변화하는 대상이며, 이러한 공간현상을 현실적, 구체적으로 분석하는 과정에서는 시간과 공간을 동시에 고려하는 것이 필수적이라 하겠다. 시공간적 분석 방법의 연구 스펙트럼이 매우 넓기 때문에 아직 학문적 발전의 초기단계 임에도 본 논문에서 다룰 수 있는 내용은 매우 제한적이다. 따라서 본 논문에서 자세히 다루지 못한 다른 방법론적 고찰은 이후 많은 학자들의 차후 과제로 남겨둔다.

## 참고문헌

- Anselin, L., 1995, Local Indicators of Spatial Association -LISA, *Geographical Analysis*, 27(2), 93-115.
- Bailey, T.C. and Gatrell, A.C., 1995, *Interactive Spatial Data Analysis*, Essex, England: Longman.
- Baker, R.D., 1996, Testing for space-time clusters of unknown size, *Journal of applied statistics*, 23, 543-554.
- Besag, J. and Newell, J., 1991, The detection of clusters in rare diseases, *Journal of the Royal Statistical Society: Series A*, 154, 256-274.
- Clark, P.J. and Evans, F.C., 1954, Distances to nearest neighbor as a measure of spatial relationships in populations., *Ecology*, 35, 445-453.
- Cressie, N., 1991, *Statistics for Spatial Data*, New York: Wiley.

- Diggle, P.J., Chetwynd, A.G., Haggkvist, R. and Morris, S.E., 1995, Second-order analysis of space-time clustering, *Statistical Methods in Medical Research*, 4, 124-136.
- Elliott, P., Cuzick, J., English, D. and Stern, R., 1992, *Geographical & Environmental Epidemiology: Methods for small-area studies*, Oxford: Oxford University Press.
- Elliott, P., Wakefield, J.C., Best, N. and Briggs, D.J., 1999, *Spatial Epidemiology: Methods and Applications*, Oxford: Oxford University Press.
- Elliott, P. and Wartenberg, D., 2004, Spatial epidemiology: Current approaches and future challenges, *Environmental Health Perspectives*, 112, 998-1006.
- Florax, R. and Vlist, A.V.D., 2003, Spatial econometric data analysis: moving beyond traditional models, *International Regional Science Review*, 26(3), 223-243.
- Fotheringham, S. and Rogerson, P., 2009, *The SAGE Hand book of Spatial Analysis*, London: Sage.
- Gatrell, A.C. and Loytonen, M., 1998, *GIS and Health Research: An Introduction*, in: Gatrell, A.C., Loytonen, M. (eds.), *GIS and Health*, Philadelphia: Taylor & Francis.
- Geary, R.C., 1954, The contiguity ratio and statistical mapping, *The incorporated statistician*, 5, 115-145.
- Getis, A. and Aldstadt, J., 2004, Constructing the spatial weights matrix using a local statistic, *Geographical Analysis*, 36, 90-104.
- Goodchild, M.F. and Haining, R.P., 2004, GIS and spatial data analysis: Converging perspectives, *Papers in Regional Science*, 83(1), 363-385.
- Haining, R., 1990, *Spatial data analysis in the social and environmental sciences*, Cambridge: Cambridge University Press.
- Hunter, J.S., 1986, The exponentially weighted moving average, *Journal of Quality Technology*, 18, 203-210.
- Jacquez, G.M., 1996, A k-nearest neighbor test for space-time interaction, *Statistics in Medicine*, 15, 1935-1949.
- Johnson, S., 2006, *The Ghost Map: the story of London's most terrifying epidemic-and how it changed science, cities, and the modern world*, New York: Riverhead Books.
- Johnson, S.D. and Bowers, K.J., 2004, The stability of space-time clusters of burglary, *British Journal of Criminology*, 44(1), 55-65.
- Kim, Y. and O'Kelly, M., 2008, A bootstrap based space-time surveillance model with an application to crime occurrences, *Journal of Geographical Systems*, 10(2), 141-165.
- Knox, G., 1964, The detection of space-time interactions, *Applied Statistics*, 13, 25-29.
- Kulldorff, M. and Hjalmars, U., 1999, The Knox method and other tests for space-time interaction, *Biometrics*, 55, 544-552.
- Kulldorff, M. and Nagarwalla, N., 1995, Spatial disease clusters: detection and inference, *Statistics in Medicine*, 14, 799-810.
- Lawson, A.B., 1993, On the analysis of mortality events associated with a pre-specified fixed point, *Journal of the Royal Statistical Society: Series A*, 156, 363-377.
- Maguire, M., 2000, Policing by risks and targets: Some dimensions and implications of intelligence-led crime control *Policing and Society*, 9, 315-336.
- Mantel, N., 1967, The detection of disease clustering and a generalized regression approach, *Cancer Research*, 27, 209-220.
- Moran, P.A.P., 1948, The interpretation of statistical maps., *Journal of the Royal Statistical Society: Series B*, 10, 245-251.
- O'Kelly, M.E., 1994, Spatial analysis and GIS, in: Fotheringham, A.S., Rogerson, P. (eds.), *Spatial Analysis and GIS*: Taylor and Francis, 65-79.
- O'Sullivan, D. and Unwin, D.J., 2010, *Geographic Information Analysis (2nd ed.)* New Jersey: Wiley.
- Openshaw, S. and Charlton, M., 1987, A mark 1 Geographical Analysis Machine for the automated analysis of point data-sets, *International Journal of Geographical Information Systems*, 1, 335-358.
- Page, E.S., 1954, Continuous inspection schemes,

- Biometrika*, 41, 100-115.
- Ratcliffe, J.H., 2004, The hotspot matrix: a framework for the spatio-temporal targeting of crime reduction, *Police Practice and Research*, 5(1), 05-23.
- Ripley, B.D., 1976, The second-order analysis of stationary point patterns, *Journal of applied probability*, 13, 255-266.
- Rogerson, P., 1997, Surveillance systems for monitoring the development of spatial patterns, *Statistics in Medicine*, 16, 2081-2093.
- Rogerson, P.A., 2001, Monitoring point patterns for the development of space-time clusters, *Journal of Royal Statistical Society A*, 164, 87-96.
- Shewhart, W.A., 1931, *Economic control of quality of manufactured product*, New York: Springer-Verlag.
- Sonesson, C. and Bock, D., 2003, A review and discussion of prospective statistical surveillance in public health, *Journal of Royal Statistical Society A*, 166, 5-21.
- Stern, L. and Lightfoot, D., 1999, Automated outbreak detection: a quantitative retrospective analysis, *Epidemiology and Infection*, 122, 103-110.
- Stone, R., 1988, Investigation of excess environmental risks around putative sources: statistical problems and a proposed test, *Statistics in Medicine*, 7, 649-660.
- Strat, Y.L., 2005, Overview of temporal surveillance, in: Lawson, A.B., Kleinman, K. (eds.), *Spatial and Syndromic Surveillance*: Wiley.
- Turnbull, B.W., Iwano, E.J., Burnett, W.S., Howe, H.L. and Clark, L.C., 1990, Monitoring for clusters of disease: application to leukemia incidence in upstate New York., *American Journal of Epidemiology*, 312 (Suppl. 1), S136-143.
- Waller, L.A. and Gotway, C.A., 2004, *Applied spatial statistics for public health data*, New York: Wiley.
- Waller, L.A., Turnbull, B.W., Clark, L.C. and P., N., 1992, Chronic disease surveillance and testing of clustering of disease and exposure: Application to leukemia incidence and TCE-contaminated dumpsites in upstate New York, *Environmetrics*, 3, 281-300.
- 교신 : 김영호, 136-107, 서울시 성북구 안암동 5가 고려대학교 사범대학 지리교육과 (이메일: younghokim@korea.ac.kr)
- Correspondence : Youngho Kim, 136-107, 5 Anam-dong Seongbuk-gu, Seoul, Korea, Department of Geography Education, College of Education, Korea University (Email: younghokim@korea.ac.kr)

투 고 일: 2012년 10월 10일

심사완료일: 2012년 10월 27일

투고확정일: 2012년 10월 30일