

근린가중치행렬이 공간적 자기상관 추정에 미치는 영향

- 서울시를 사례로 -

박 기 호*

A Study on the Effects of Spatial Proximity Weight Matrices on the Spatial Autocorrelation Measures

- The Case of Seoul Administrative Units -

Key-Ho Park*

요약 : 공간분석에 있어서 “근린”의 개념은 다양하게 정의될 수 있으며, 이의 측정과 정량화 방법 또한 임의적이다. 본 연구는 근린가중치를 설정하는 방식이 공간적 자기상관 추정에 어떠한 영향을 주는지 경험적 분석을 통해 밝히는 것을 목적으로 한다. 서울시의 행정구역인 구와 동 단위로 집계된 변수 중 공간적 자기상관 정도를 대표하는 자료를 선정한 후, Moran의 I 지수 측정치가 다양한 공간근린가중치의 설정 방식에 따라 변화하는 정도를 기록하고 이를 해석하였다. 민감도 분석의 결과, 공간근린가중치의 종류와 매개변수 설정에 따라 공간적 자기상관 지수는 매우 예민하고 큰 편차를 가지고 변하는 것으로 확인되었다. 또한 이러한 변화에는 거리조락 등 관련 매개변수에 따라 일련의 패턴이 나타났으며, 이는 공간 자기상관이나 공간회귀 모형의 추정량 산출과 그 유의성 해석에 있어서 공간근린가중치 설정에 대한 탐색적 분석이 선행되어야 함을 시사한다.

주제어 : 공간통계, 공간적 자기상관, 근린가중치행렬, 민감도 분석

ABSTRACT : The concept of spatial "neighborhood" among a set of polygonal areal units may be defined in a variety of ways. The specification and the measurement of spatial proximity weight, however, has not received much attention in the literature. This paper reports an empirical study on the effects of proximity weight specification on the spatial autocorrelation measurement indices. For some representative variables collected over the administrative units of Seoul, the autocorrelation index scores are profiled along the variety of proximity weighting schemes. This study shows that the measurements are quite sensitive to the schemes and their parameters, and that some noticeable patterns seem to exist. Since the autocorrelation measurements are inherently dependent on the weighting schemes, a keen attention to spatial proximity weight matrices may be warranted when studies involving spatial statistics and spatial autoregressive model parameter estimations are conducted.

Key Words : spatial statistics, spatial autocorrelation, spatial proximity weight matrix, sensitivity analysis

*서울대학교 지리학과 부교수(Associate Professor, Department of Geography, Seoul National University)

I. 연구 배경과 목적

1. 연구 배경

지리학을 포함한 공간과학의 경험적 법칙으로 흔히 언급되는 것이 “모든 것은 다른 모든 것과 연관되어 있지만, 가까운 것은 먼 것보다 더욱 관련이 있다”라고 하는 Tobler의 법칙이다(Tobler, 1970). 공간적 자기상관(spatial autocorrelation)은 지표공간에서 관측되는 제 현상의 공간적 연관성을 포착하는 개념으로 일반 통계의 시계열 분석과 그 맥을 같이한다. 공간자기상관은 Moran(1948)에 의해 측정지수가 구체적으로 제안된 이후 지리학 뿐 아니라 경제학 등의 계량적 연구에서 자료의 공간적 성질에 초점을 두는 매우 중요한 분석 기법으로서 그 동안 널리 활용되어 왔다.

최근 국지적 자기상관(Anselin, 1995)과 같은 확장된 개념이 도입되고, GIS와 공간통계 전문 패키지 등 분석도구의 발전에 힘입어 공간적 자기상관은 학계에서 새로운 조명을 받

고 있다. 공간적 자기상관의 개요와 최근까지 축적된 연구 동향 및 성과는 Goodchild(1986), Getis and Ord(1992), Griffith(2003) 등에 잘 정리되어 있으며, <표 1>은 최근 국내 문헌 중 자기상관기법이 활용된 일부 사례이다.

공간적 자기상관 지수는 연구지역 내 단위 구역들(areal units)의 속성값이 그 근린(neighborhood)의 값들과 유사한 정도를 일련의 표준적인 수식에 의해 정량적으로 측정하는 것이다. 이를 위해서는 먼저 단위 구역들 사이의 공간적 근린관계를 정의하고 이를 정량화하는 작업이 필요한데, 이것을 근린가중치행렬(spatial proximity weight matrix)이라 하고 W 로 표기한다. W 는 단위구역 수를 차원으로 하는 정방행 행렬로서 W_{ij} 는 단위 구역 i 와 j 간의 근린 정도를 나타내는 수치이다.

공간적 자기상관의 추정에 있어서 짚어 보아야 할 문제는 공간적 “근린”의 개념 자체가 다양하게 정의될 수 있다는 것이다. 공간적 자기상관 지수와 수식은 표준화 되었지만, 그 수식에 포함된 W 에는 이렇다 할 표준이 없다.

<표 1> 공간적 자기상관을 활용한 국내 연구사례

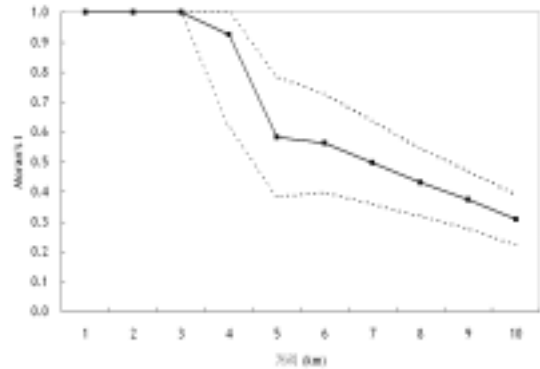
연구자	연 도	연구 제 목	주요 연구 내용
이희연 이용균	2004	Analysis on the Spatial Dimension of the Commercial Domains: the Case of Seoul, Korea	공간 자기상관 지수를 이용한 한국의 상업적 도메인 분포 분석
김광구	2003	공간자기상관(spatial autocorrelation)의 탐색과 공간회귀분석(spatial regression)의 활용	공간 자기상관 지수 및 공간구조 방법 제시
이지영 황철수	2002	공간통계분석을 이용한 지가의 입지값 측정에 관한 연구	공간적 자기상관 관계를 이용한 공간적 특성가격 모형 제시
서경천 이성호	2001	지가의 공간적 변동에 따른 입지대역의 분석에 관한 연구 - 공간적 자기상관을 고려한 방법론을 중심으로	공간적 자기상관 기법을 이용한 헤도닉 모형 개선 방법 제시
김종원	2000	주택시장에서의 공간자기상관의 검증 및 회귀계수의 추정	공간적 데이터의 불편 추정통계량을 구하기 위한 바람직한 모형설정방법 제시

2. 문제 제기

W 를 정의하는 다양한 방법들이 공간통계의 여러 문헌에서 소개되고 있지만(Bailey and Gartrel, 1995; Lee and Wong, 2001), 특정 W 의 선택이 분석 결과에 어떤 영향을 미치는지에 대해서는 논의된 바가 없고, <표 1>의 사례들처럼 대부분의 연구에서는 통계패키지 등에서 제공하는 편의에 따라 단위 구역 사이의 경계선 공유 여부만을 고려한 W 가 여과 없이 사용되고 있는 실정이다.

W 의 설정에 따라 공간적 자기상관 측정치가 상당한 차이를 보이는 것은 <그림 1>의 예에서 쉽게 확인할 수 있다. 이 분석은 서울시 행정구의 노인인구비율에 대하여 단위구역 중심점 간의 거리를 근린의 임계치로 설정하여 이 근린설정의 변화에 따른 자기상관의 변화를 추적한 것으로, 실선은 Moran의 I 를, 점선은 ± 1 표준오차 범위를 표시한다. <그림 1>은 표준화된 공간적 자기상관 지수를 적용할 지라도 연구자가 선택한 W 에 따라 자기상관의 규모가 상대적으로 과대 혹은 과소 추정될 가능성이 충분히 있다는 것을 시사한다.

W 의 생성과 매개변수의 설정이 공간적 자기상관 추정에 어떠한 영향을 미치는지, 아울러 이 영향이 지리적 현상의 특성에 따라 차이가 있는지에 관한 체계적인 기존 연구는 찾아보기 힘들다. 특히 서울시 행정구역이라는 고유한 분할을 기준으로 사회·경제·환경 현상을 분석할 때 W 에 관련한 지침을 제공하는 연구는 시도된 바 없었다.



<그림 1> 근린거리에 따른 자기상관 지수의 변화
(서울시 노인인구비율 : 2000년 구 단위)

3. 연구 목적

본 연구는 근린가중치행렬 생성방법이 공간적 자기상관 추정에 미치는 영향을 경험적 분석을 통해 파악하는 것이 목적이며, 이러한 영향의 민감도는 아래와 같은 질문으로 구체화하여 분석한다.

첫째, 근린구조, 즉 W 의 설정이 공간적 자기상관의 추정에 어떤 영향을 주는가? 추정값의 통계적 유의성과 표준오차에도 영향이 있는가? 이러한 변화에 일정한 패턴이 있는가?

둘째, 표본 크기, 즉 구와 동과 같은 공간단위의 규모에 따라 공간적 자기상관 지수에 대한 W 의 영향과 그 민감도에 차이가 있는가?

셋째, 공간현상에 내재한 자기상관 정도에 따라 W 의 영향이 달라지는가? 공간적 자기상관성이 비교적 큰 변수와 작은 변수에 있어서 W 의 영향력과 민감도가 각각 달라지는가?

II. 연구 방법

1. 분석의 틀과 범위

근린구조의 영향을 관찰하는 민감도 분석이 체계적으로 진행되기 위해서는 <표 2>에 개략적으로 나타난 세 가지 요소가 잘 설계되고 조절되어야 한다. 첫째, 분석 자료의 선정이 적절히 이루어져야 한다. 구체적으로 분석의 대상 지역과 공간적 단위, 그리고 그 단위에서 집계된 변수의 특성에 따라 독립적으로 근린구조의 영향을 관찰할 수 있어야 한다. 이 때 실험 변수들의 정성적, 정량적 대표성에 큰 문제가 없어야 한다. 둘째, 근린구조에 따라 계산되는 공간적 자기상관의 통계적 추정량을 선정해야 한다. 이는 공간적 자기상관은 다양한 지수로 추정될 수 있는 바, 각 대안적 지수들이 근린구조에 의해 영향을 받는 정도가 다를 수 있기 때문이다. 셋째, 본 연구의 핵심이라고 할 수 있는 근린구조의 정의와 그것이 가중치로 정량화되는 유형을 체계적으로 분류하고 이를 실험해야 한다. 아울러 근린가중치 유형별 매개변수를 도출하고 이들을 적절한 범위 내에서 조절하며 지수의 반응을 수집해야 한다.

공간적 자기상관 지수는 공간적 의존성 존재 여부를 말해주는 전역적 지수(global measure)와 어떤 속성 값에 대한 공간적 군집지역을 검출하는 국지적 지수(local measure)로 구분된다. 공간적 자기상관 추정에 미치는 W 의 영향을 분석하는 것은 전역적 지수만으로도 가능하기 때문에 본 연구에서는 전역적 지수로 실험의 범위를 제한하였다.

<표 2> 분석의 틀

사례 지역과 분석 변수 선정
① 단계구분도를 이용한 탐색 ② 정성적/정량적 대표 변수 선정
↓
근린가중치행렬 생성
① 근린구조의 유형화 ② 유형별 근린가중치행렬 생성
↓
공간적 자기상관 지수 산출
① 전역적 공간적 자기상관 지수 선정 ② 근린가중치행렬 유형별 지수 산출
↓
공간적 자기상관 지수 비교·분석
① W 유형별·매개변수별 비교·분석 ② 공간적 규모별 비교·분석 ③ 변수별 비교·분석

공간자료의 분석에서 까다로운 문제 중 하나는 유한한 연구대상지역의 최외곽 경계와 관련된 경계효과(boundary effect)이다. 본 연구의 근린가중치 산출 방법에서는 이러한 경계효과를 보정하지 않았다.

공간 근린을 정의하는 데 있어서 단위 지역 간의 도로망 연결성이나 인구규모 등 지역 간의 상호작용 정도를 근린 가중치에 반영하는 방법이 있다. 이것은 오존농도처럼 자연환경에 관련한 자료의 경우는 예외이나 실제 물류의 이동이 이루어지는 사회, 경제적인 자료를 다룰 때는 의미가 있다. 하지만 본 연구에서는 근린가중치의 영향 여부를 분석하는 것이 주목적이기에 단위 지역들의 거리와 기하학적 배열에 한정하여 근린가중치를 유형화하였다. 근린 설정의 영향이 중대하다고 확인된 후에야 최적의 근린가중치에 대한 논의를 시작할 수 있기 때문이다.

2. 연구 지역

공간근린의 구조와 그 영향에 대한 민감도 분석결과는 사례로 삼은 지역의 공간 단위와 그 배열에 따라 조금씩 다를 수밖에 없다. 본 연구에서는 현실적으로 많이 연구되고 분석되고 있는 서울을 사례로 하며, 특히 다양한 통계 자료가 집계되는 행정구역단위를 대상으로 한다. 2004년 현재 서울시는 25개 행정자치구와 522개 동으로 구성되어 있다. 세부 분석은 동 단위 자료를 중심으로 수행하였으나 공간 단위의 규모에 따른 W 효과의 차이를 보기 위해 구 단위의 분석도 포함하였다.

3. 분석 자료

서울시 행정구역에 대하여 제공되는 통계자료는 그 종류와 특성이 매우 다양하다. 본 연구에서는 구, 동의 공간단위로 집계되며, 서울시에 관한 사회·경제·환경적 연구에서 보편적으로 활용되는 자료를 분석대상으로 삼았다.

다양한 서울시 구·동별 집계자료의 단계구분도를 통해 각 변수들의 자기상관 정도를 정

성적으로 탐색하여 1차 선별을 하였다. 선정과정은 원자료의 구분에 따른 시각적 단순화와 왜곡을 최소화하기 위해 무분류 단계구분도(Tobler, 1973)를 통해 이루어졌으며 본 논문에서는 해당 변수의 최소치와 최대치를 64개의 등간격으로 구분하여 나타내었다.

이들에 대하여 공유경계선 근린 정의에 따른 Moran의 I 를 산출하였고, 특히 여러 분야의 활용도를 고려하여 노인인구비율, 오존농도, 평균 지가, 그리고 생활보호대상자비율을 분석 대상으로 최종 선정하였다. 이들에 대한 원자료 출처와 가공방식은 <표 3>에 정리되었으며, 단계구분도는 <그림 2>에서 <그림 5>에 나타난 바와 같다.

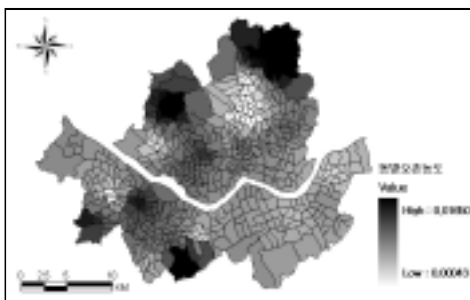
각 변수에 대한 Moran의 I 는 노인인구비율이 0.69, 평균지가의 경우 0.52, 그리고 오존농도는 0.87로 계산되었다. 한편, 생활보호대상자비율의 경우 0.14로서 자기상관성이 미미한 변수이다. 이들 변수에 대해 측정된 자기상관 수치는 p -value가 모두 0으로 통계적으로 유의함을 확인하였다. 이들 수치는 특정 근린가중치행렬로 계산된 것이지만 자기상관지수의 다양한 정도를 각각 대표할 수 있다고 판단되었다.

<표 3> 분석 자료

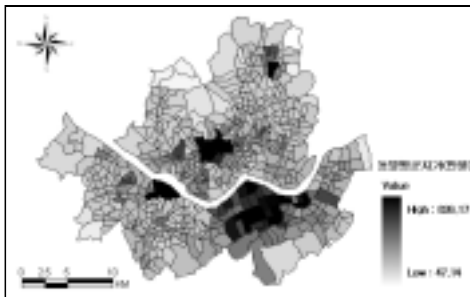
구 분	변 수	연 도	출 처	산 출 방 법
인 구	65세이상 노인인구 비율	2000	지도로 본 서울 2000 (서울시정개발연구원, 2000)	(65세이상인구수/동별인구)×100
경 제	지 가	2000	2000년 공시지가 자료	동별 평균 지가 (지목이 '대'인 필지 대상)
환 경	오존 농도(ppm)	1999.12	1999년 12월 대기환경월보 (환경부, 2000)	동별 평균 오존농도 (IDW 보고결과를 이용한 평균)
사회복지	생활보호대상자비율	1999	지도로 본 서울 2000 (서울시정개발연구원, 2000)	



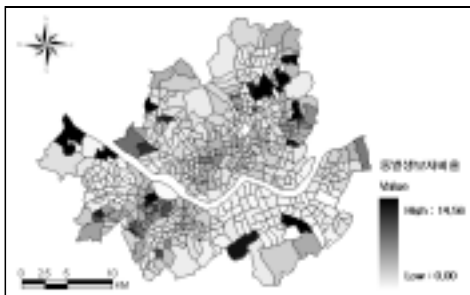
〈그림 2〉 서울시 노인인구비율의 분포



〈그림 3〉 서울시 오존농도의 분포



〈그림 4〉 서울시 평균지가의 분포



〈그림 5〉 서울시 생활보호대상자비율의 분포

4. 공간적 자기상관의 측정

1) 공간적 자기상관의 통계적 추정량

공간적 자기상관의 전역적 지수로 Moran의 I 지수(Moran, 1948), Geary의 c 지수(Geary, 1954), Getis와 Ord가 제안한 G(d) 통계량(Getis and Ord, 1992) 등이 쓰이며, 본 연구에서 이들을 모두 민감도 측정 대상으로 삼아 실험을 수행하였다. 하지만 근린가중치행렬이 이들 지수의 측정치에 미치는 영향에 있어서 큰 차이가 없었기에 본 논문에서는 가장 대표적인 Moran의 I 지수와 관련된 결과만을 밝힌다.

Moran의 I 지수는 공분산 개념을 사용하여 속성값 사이의 유사성을 측정하며, 일반적으로 아래 식을 통해 계산된다.

$$I = \frac{n \sum \sum w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{W \sum (x_i - \bar{x})^2}$$

위 식에서 n 은 구역의 수, x_i 와 x_j 는 각각 i 지역과 j 지역에서 나타나는 x 속성값, \bar{x} 는 전체 지역에 대한 평균이다. w_{ij} 는 i 지역과 j 지역 사이의 근린가중치이며, W 는 근린가중치행렬의 전체 합을 의미한다. Moran의 I 지수 값은 -1 ~ +1 범위의 값을 띠며, 그 해석 방법은 피어슨 상관계수와 유사하다.

2) 공간적 자기상관의 통계적 유의성

특정 지수의 산출이 통계적 추정량이므로 이에 대한 통계적 유의성 여부를 살펴 보는 것이 필요하다. Moran의 I 측정값의 유의성은

그 측정값의 크기와 밀접한 비례 관계가 있다. 일반적으로 유의성이 문제가 되는 것은 자기상관이 0 부근에서 추정되었거나 표본크기가 작은 경우이다. 귀무가설(H_0 : Moran의 $I = 0$)의 기각이 어렵거나 추정오차가 지나치게 큰 경우 자기상관 규모의 판단과 또한 이를 바탕으로 한 공간회귀모형의 적합성이나 회귀계수의 유의성에 많은 제약이 따르게 된다.

Moran의 I 에 대한 유의성은 표준정규점수와 p -value를 산출하거나 표준오차를 통하여 관찰할 수 있다. 이러한 유의성 검정 통계량은 귀무가설 하에서 정규성(normality)이나 확률화(randomization) 가정에 따른 확률분포를 통해 계산된다. 이러한 자기상관의 검정에 대한 구체적인 절차는 Goodchild(1986) 등의 문헌에 잘 나타나 있다.

본 연구에서 실험하는 모든 변수에 있어서 각 유의성 검정 통계량의 구성요소 중 표본 특성 부분이 서울시 행정구역으로 고정되어 있다. 따라서 본 논문에서는 자기상관 측정값의 유의성 논의를 W 와 관련한 검정력 함수의 변화를 통해 전개하며 특히 자기상관의 구간추정량이 되는 표준오차에 있어서 W 의 대안적 설정에 따른 민감도를 분석한다.

5. 근린가중치행렬

대상 지역의 공간적 분할이 단위 구역 폴리곤으로 구성된 경우 W 는 이 폴리곤 사이의 공간적 인접성에 기초하여 작성된다(Bailey and Gatrell, 1995). 이러한 공간적 인접성을 나타내는 가장 기본적인 형태는 이진 연결성(binary

connectivity)으로서, 두 폴리곤이 인접하면 1, 인접하지 않으면 0의 가중치가 설정된다(Lee and Wong, 2001).

기존 연구의 공간적 자기상관 분석에서는 예외 없이 이진 연결성 행렬을 사용해 왔으나, 본 연구에서는 이와 함께 단위 구역의 기하학적 특성에 의한 공유 경계선(shared boundary), 폴리곤 중심점(polygon centroid), K-최근린(K-nearest), 그리고 폴리곤의 중첩(overlaid polygon) 등 다양한 근린가중치행렬 W 를 산출하고 이를 비교한다. 모든 근린가중치행렬은 지수산출을 위한 전처리 과정으로서의 열의 합계를 1로 표준화 했다($\sum_i w_{ij} = 1$).

1) 공유 경계선 비율에 따른 근린가중치



〈그림 6〉 공유 경계선과 비율

〈그림 6〉에 보이는 것처럼 A지역과 B지역이 경계선을 공유할 때, A지역에 대한 B지역의 근린가중치는 폴리곤 A의 총 경계선 길이에 대한 공유 경계선길이의 비율로 정의되고, B지역에 대한 A지역의 근린가중치는 폴리곤 B의 총 경계선 길이에 대한 공유 경계선 길이의 비율로 정의된다(Bailey and Gatrell, 1995).

$$w_{ij} = \frac{l_{ij}}{l_i}$$

l_i : i 의 경계선 길이

l_{ij} : i 와 j 의 공유 경계선 길이

2) 폴리곤 중심점간의 거리에 따른 근린가중치

어떤 지역의 폴리곤 중심점을 P라고 할 때, P로부터 다른 지역들의 폴리곤 중심점 사이의 거리를 기준으로 가중치를 결정할 수 있다. 이때 P로부터의 반경을 매개변수화(parameterized)하여 일정 반경 이내의 폴리곤만을 근린 대상으로 하여 연산하기도 한다(<그림 7>). 또한 거리 d 와 근린지역의 면적 A 등을 가중치 산출에 반영할 수 있다. 거리조락은 d , d^2 , d^3 등의 형태를 취한다(Upton, 1990).

$$w_{ij} = A_j \left(\frac{\frac{1}{(d_{ij})^p}}{\sum_i \frac{1}{(d_{ij})^p}} \right)$$

A_j : j 의 면적

d_{ij} : i 와 j 사이의 중심점 거리

p : 제곱수



<그림 7> 폴리곤 중심점과 거리

3) K-최근린에 따른 근린가중치

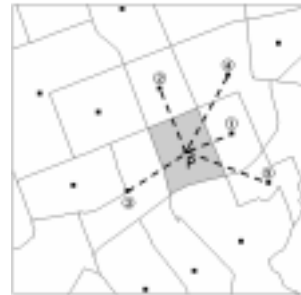
K-최근린 방법은 어떤 지역의 폴리곤으로부터

터 최근린에 위치한 K개의 폴리곤을 근린으로 정의하여 W 를 작성하며, 폴리곤 중심점 사이의 거리를 최근린 측정의 기준으로 이용한다(<그림 8>). 어떤 지역의 폴리곤 중심점을 P라고 할 때, 최근린 매개변수 K는 P로부터의 가장 가까운 거리에 있는 폴리곤(중심점)의 개수로 설정된다(Bailey and Gatrell, 1995).

$$w_{ij} = \left(\frac{\frac{1}{(d_{ij})^p}}{\sum_i \frac{1}{(d_{ij})^p}} \right)$$

d_{ij} : i 와 j 사이의 중심점 거리

p : 제곱수

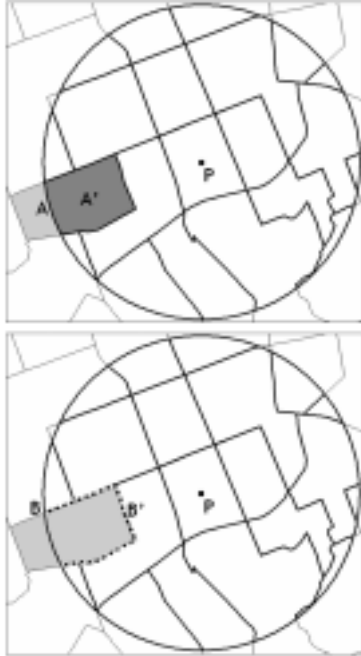


<그림 8> K-최근린에 의한 근린 설정

4) 폴리곤의 중첩에 따른 근린가중치

공유 경계선, 폴리곤 중심점, K-최근린 등 기존 문헌에서 이론적으로 제시되어온 방법과 더불어 본 연구에서는 폴리곤의 중첩면적을 근린가중치로 설정하는 방법을 새롭게 제안한다. 이는 <그림 9>에서처럼 어떤 지역의 폴리곤 중심점 P로부터의 반경을 매개변수로 삼아 일정 반경의 원을 그렸을 때, 이 원과 중첩되는 폴리곤의 면적(A')이나 경계선 길이(B')의 비율을 가중치로 적용하는 것이다. 이 방법은 앞서 제시된 세 가지 방법을 통합하여 개선한

방식으로서, GIS의 기하연산에 의해 산출된다.



〈그림 9〉 폴리곤 중첩과 근린가중치

일정 반경 이내의 폴리곤을 근린 대상으로 하며, 이 폴리곤의 참여 정도에 따라 근린정도

를 단계화함으로써 1차 근린에 대한 공유 경계선 여부 또는 공유 경계선의 비율을 이용하는 방법을 개선하여 정밀도를 높인 것이다.

$$w_{ij} = \frac{A'_j}{A_j}$$

$$w_{ij} = \frac{B'_j}{B_j}$$

A_j : j 의 면적, A'_j : 반경내 j 의 면적

B_j : j 의 경계선길이, B'_j : 반경내 j 의 경계선길이

III. 분석 및 결과

1. 근린가중치행렬의 구성

이진 연결성 행렬은 거리조락의 공간적 효과를 고려하지는 않지만 일반 연구에서 가장 보편적으로 사용되고 있으므로, 본 연구에서는 이를 다른 행렬에 대한 비교 기준으로 설정한다. 이와 함께 공유경계선 비율, 폴리곤 중심점 거리, K-최근린거리 및 폴리곤 중첩을 이용한

〈표 4〉 분석에서 사용된 W 의 유형과 매개변수

분 류	행 렬	매개변수	핵 심 사 항
wmC	$wmC::d$	$d=1..10(km)$	매개변수 반경내의 폴리곤 중심점간 거리
	$wmC::d2$		매개변수 반경내의 폴리곤 중심점간 거리의 제곱
	$wmC::d3$		매개변수 반경내의 폴리곤 중심점간 거리의 세제곱
$wmCA$	$wmCA::d$	$d=1..10(km)$	매개변수 반경내의 폴리곤 중심점간 거리와 해당 폴리곤의 면적
	$wmCA::d2$		매개변수 반경내의 폴리곤 중심점간 거리의 제곱과 해당 폴리곤의 면적
	$wmCA::d3$		매개변수 반경내의 폴리곤 중심점간 거리의 세제곱과 해당 폴리곤의 면적
wmK	$wmK::d$	$k=1..200(개)^{1)}$	매개변수 k 개의 근린과의 폴리곤 중심점간 거리
	$wmK::d2$		매개변수 k 개의 근린과의 폴리곤 중심점간 거리의 제곱
	$wmK::d3$		매개변수 k 개의 근린과의 폴리곤 중심점간 거리의 세제곱
wmO	$wmO::a$	$d=1..10(km)$	매개변수 반경내에 중첩되는 폴리곤 면적
	$wmO::b$		매개변수 반경내에 중첩되는 폴리곤 경계선길이
$wmSB$			공유경계선 길이의 비율

1) 각 동의 중심점으로부터 10km 반경 내에 포함되는 근린의 수는 약 200개임

근린가중치행렬을 구성하여 분석을 실시한다.

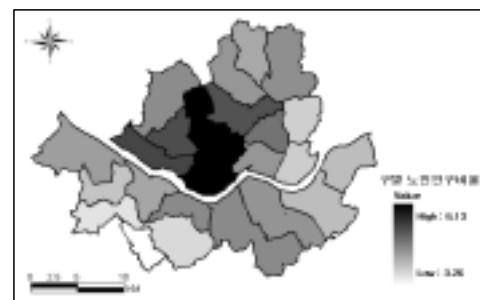
이러한 W 들은 현 통계패키지에서 직접 제공되지 않으므로, 본 연구에서는 ArcObjects와 Visual Basic 6을 플랫폼으로 ArcGIS 확장모듈(Razavi, 2002)을 개발하였다. 폴리곤 중심점 거리, 폴리곤 중심점 거리와 폴리곤 면적, K-최근린, 폴리곤 중첩을 이용하는 방법을 편의상 각각 wmC , $wmCA$, wmK , wmO , $wmSB$ 로 구분하며, 그 세부사항은 <표 4>와 같다.

2. 서울시 구 단위 자료의 분석

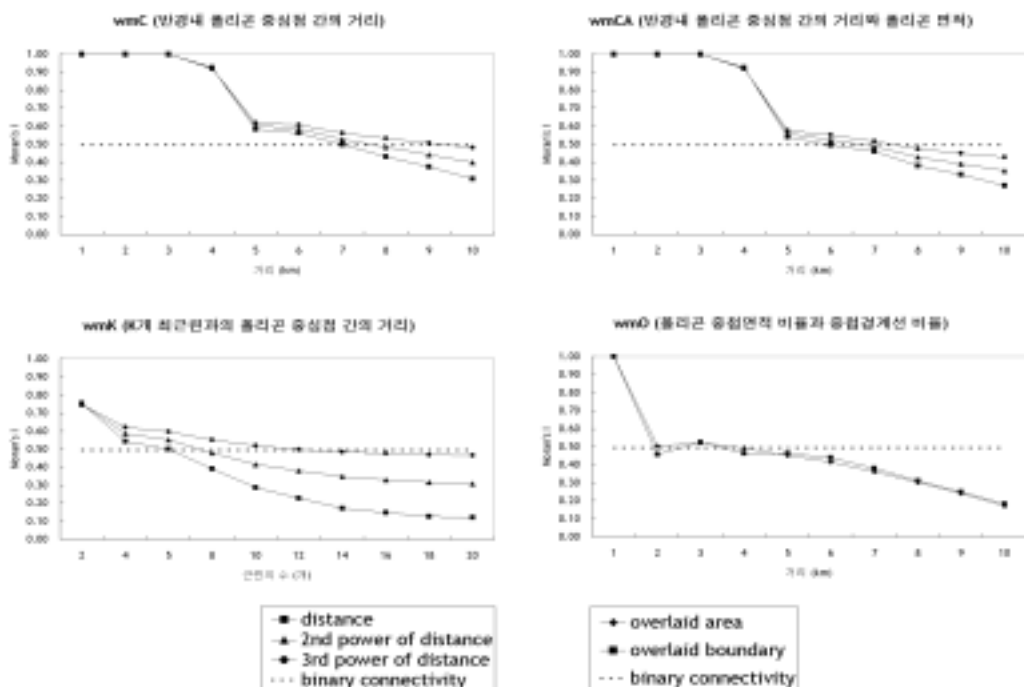
서울시 구 단위에서 앞에서 선정된 4가지 변수를 대상으로, wmC , $wmCA$, wmK , wmO 근린가중치행렬을 적용하여 매개변수별로 공간적 자기상관 지수를 산출하였다. 이 중 노인

인구비율의 분포(<그림 11>)에 대한 결과는 <그림 10>에 정리되어 있다. 서울시 구를 대상으로 하는 공간적 자기상관 평가는 구의 평균 크기를 고려할 때 폴리곤 중심점 간의 거리를 기준으로 약 4km 이상부터 유효함을 알 수 있다.

이진 연결성행렬 $wmBin$ 을 이용했을 경우, 이 변수의 Moran의 I 는 0.5 정도로 측정된다.



<그림 11> 서울시 노인인구비율의 분포



<그림 10> 서울시 노인인구비율의 자기상관 민감도 (단위지역 : 구)

그러나 실험 결과, 대안적인 W 와 매개변수의 조합에 따라, Moran의 I 지수가 0.2 ~ 0.9 정도까지 큰 편차를 보이며 변할 수 있음이 확인되었다. 구 단위의 자료는 통계적 표본의 수가 작아 유의성이 다소 떨어진다. 하지만 구 단위 자료를 이용한 공간회귀 연구도 빈번히 수행되고 있는 만큼 근린가중치 설정에 따른 민감도에 더욱 주의할 필요가 있다.

3. 서울시 동 단위 자료의 분석

동 단위 분석에서도 구 단위 분석과 마찬가지로 4가지 변수를 대상으로, wmC , $wmCA$, wmK , wmO 각 근린가중치행렬을 매개변수 별로 생성하여 공간적 자기상관 지수를 산출

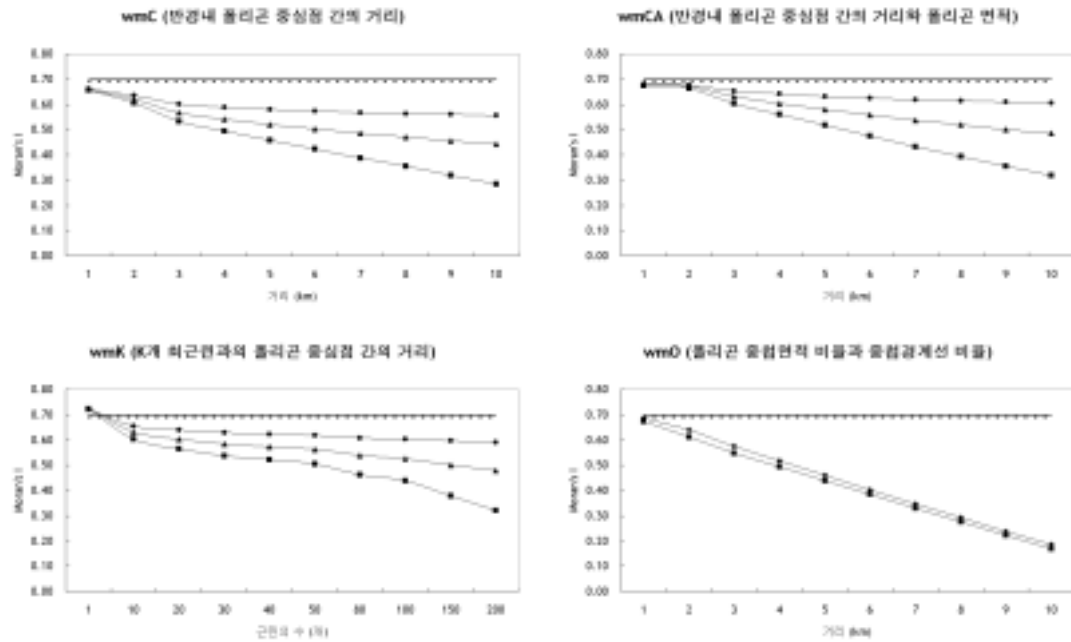
하였다. 각 W 에서의 매개변수 증가에 따른 지수 변화는 〈그림 12〉 ~ 〈그림 15〉와 같다.

노인인구비율, 오존농도, 지가와 같이 공간적 자기상관의 정도가 강한 3가지 변수에서는, 매개변수가 증가할수록(즉, 근린 기준이 완화될수록) 자기상관 지수가 전체적으로 감소하는 경향을 보인다. wmC , $wmCA$, wmK 의 경우 매개변수인 거리 d 의 승수가 높아질수록 자기상관지수의 감소폭이 완만해지며 이는 거리조락의 효과가 강해짐에 따라 근린의 기준도 강화되었기 때문이다.

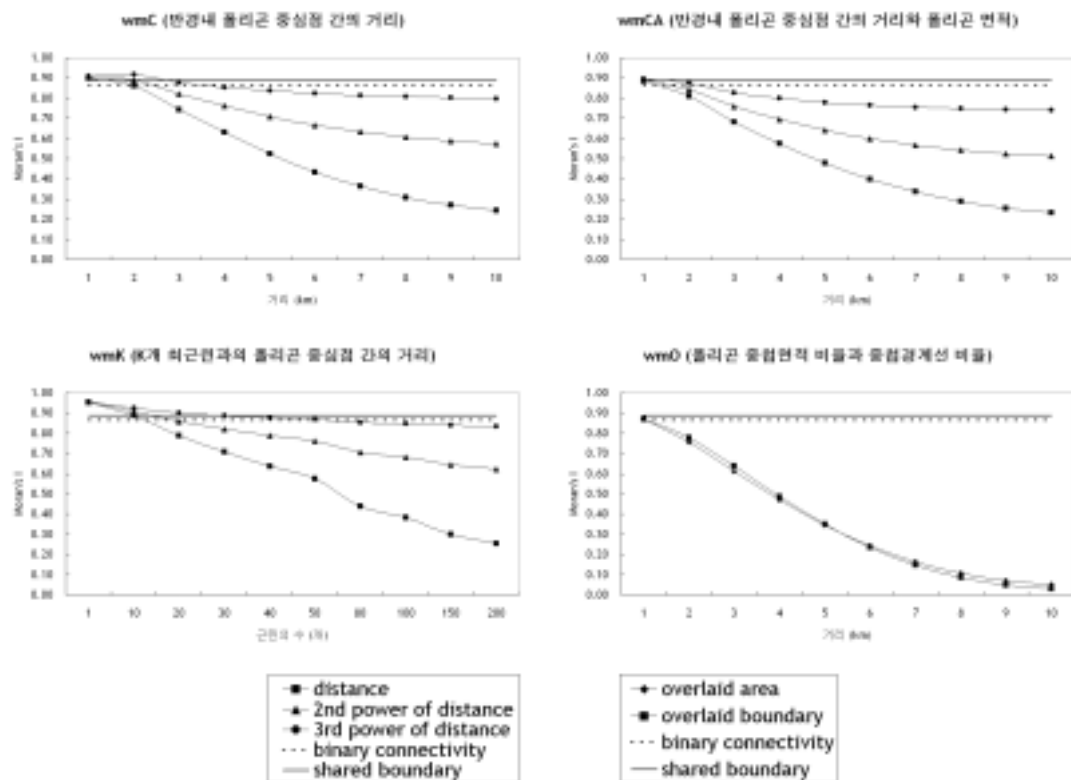
$wmBin$ 을 기준으로 통계적으로 유의한 자기상관이 관찰되지 않는 경우 근린가중치행렬을 바꾸더라도 자기상관의 규모나 그 유의성은 증가하지 않았다.

〈표 5〉 동별 4개변수에 대한 Moran I

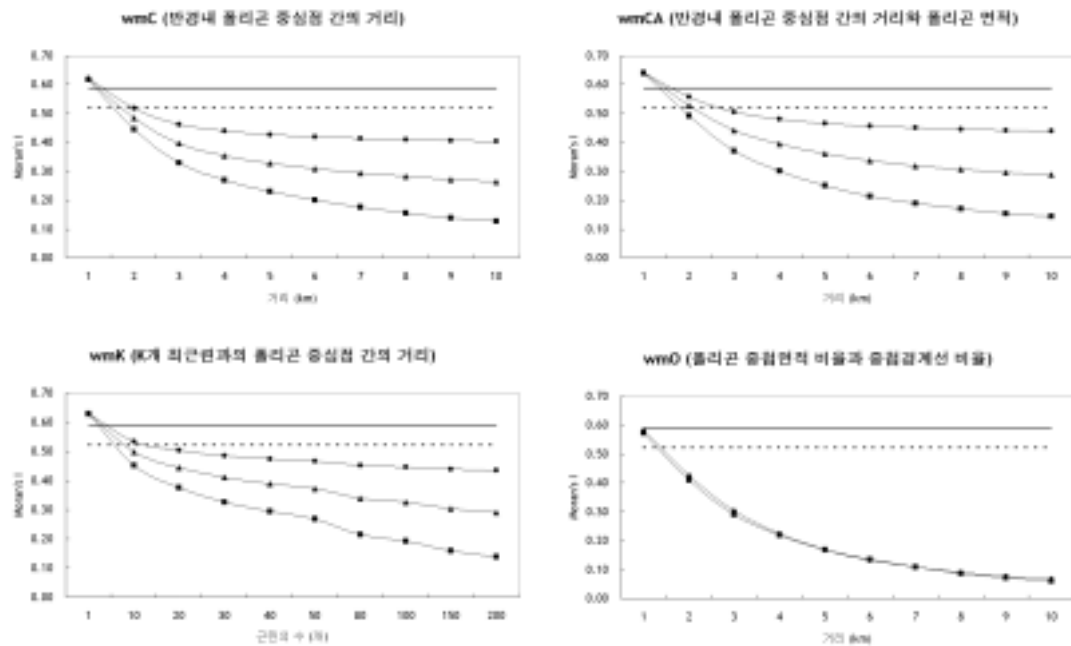
	W	최소값	최대값	평균	표준편차		W	최소값	최대값	평균	표준편차
노 인 인 구 비	wmC::d	0.2844	0.6560	0.4515	0.1219	지 가	wmC::d	0.1267	0.6168	0.2687	0.1567
	wmC::d2	0.4422	0.6585	0.5263	0.0714		wmC::d2	0.2639	0.6194	0.3605	0.1133
	wmC::d3	0.5564	0.6612	0.5884	0.0349		wmC::d3	0.4045	0.6220	0.4526	0.0689
	wmCA::d	0.3190	0.6744	0.4994	0.1254		wmCA::d	0.1434	0.6410	0.2926	0.1644
	wmCA::d2	0.4856	0.6760	0.5762	0.0680		wmCA::d2	0.2885	0.6417	0.3914	0.1155
	wmCA::d3	0.6055	0.6775	0.6351	0.0265		wmCA::d3	0.4402	0.6425	0.4889	0.0651
	wmK::d	0.3213	0.7236	0.5054	0.1147		wmK::d	0.1381	0.6296	0.3049	0.1505
	wmK::d2	0.4801	0.7236	0.5727	0.0701		wmK::d2	0.2911	0.6296	0.4004	0.1031
	wmK::d3	0.5914	0.7236	0.6287	0.0387		wmK::d3	0.4358	0.6296	0.4870	0.0587
	wmO::a	0.1871	0.6856	0.4343	0.1701		wmO::a	0.0672	0.5816	0.2179	0.1710
	wmO::b	0.1685	0.6736	0.4147	0.1684		wmO::b	0.0598	0.5729	0.2120	0.1678
	wmBin		0.6911				wmBin		0.5226		
	wmSB		0.6963				wmSB		0.5881		
오 존 농 도	wmC::d	0.2423	0.8997	0.5274	0.2447	생 활 보 호 대 상 자 비 율	wmC::d	0.0509	0.2931	0.1141	0.0810
	wmC::d2	0.5737	0.9045	0.7153	0.1233		wmC::d2	0.1289	0.2955	0.1740	0.0543
	wmC::d3	0.7981	0.9158	0.8435	0.0439		wmC::d3	0.2108	0.2979	0.2308	0.0280
	wmCA::d	0.2331	0.8845	0.4940	0.2345		wmCA::d	0.0432	0.3073	0.1096	0.0853
	wmCA::d2	0.5129	0.8898	0.6576	0.1348		wmCA::d2	0.1197	0.3084	0.1667	0.0607
	wmCA::d3	0.7408	0.8949	0.7928	0.0556		wmCA::d3	0.2044	0.3097	0.2268	0.0336
	wmK::d	0.2540	0.9538	0.5928	0.2451		wmK::d	0.0422	0.2175	0.1069	0.0581
	wmK::d2	0.6239	0.9538	0.7754	0.1120		wmK::d2	0.1090	0.2175	0.1485	0.0356
	wmK::d3	0.8354	0.9538	0.8792	0.0388		wmK::d3	0.1703	0.2175	0.1850	0.0155
	wmO::a	0.0503	0.8692	0.3692	0.2959		wmO::a	0.0064	0.2575	0.0685	0.0827
	wmO::b	0.0282	0.8721	0.3675	0.3116		wmO::b	0.0066	0.2362	0.0686	0.0790
	wmBin		0.8669				wmBin		0.1977		
	wmSB		0.8878				wmSB		0.2386		



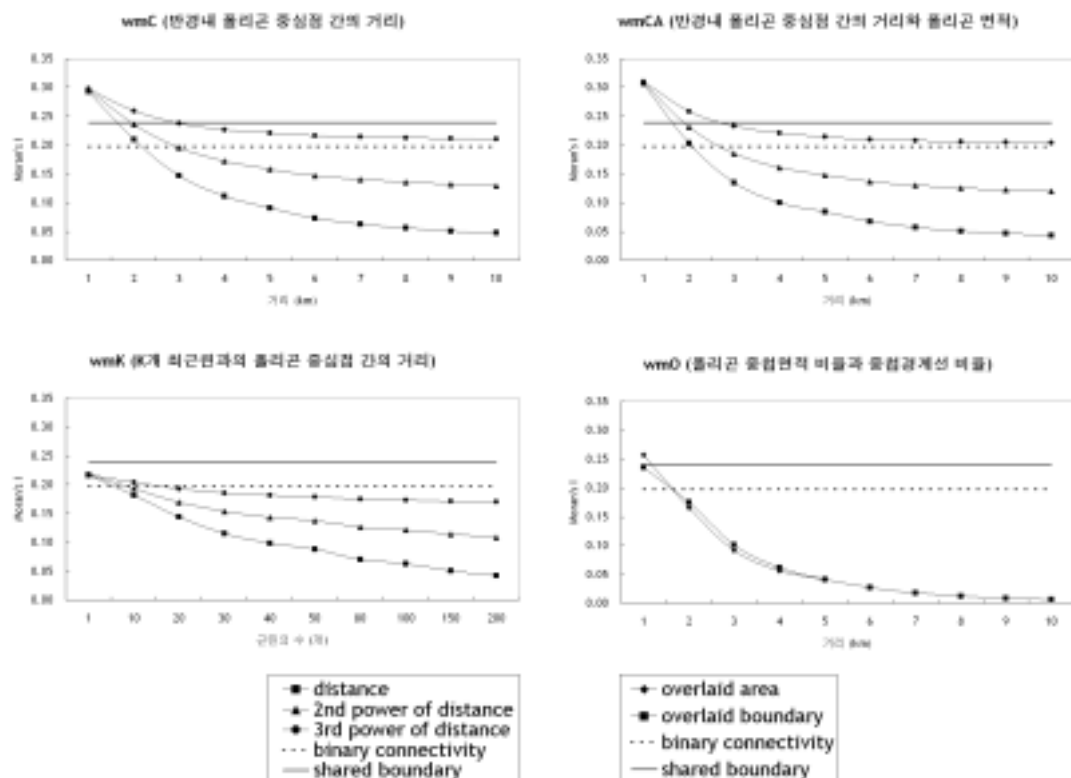
〈그림 12〉 서울시 노인인구비율의 자기상관 민감도 (단위지역 : 동)



〈그림 13〉 서울시 오존농도분포의 자기상관 민감도 (단위지역 : 동)



〈그림 14〉 서울시 평균지가분포의 자기상관 민감도 (단위지역 : 동)



〈그림 15〉 서울시 생활보호대상자 비율의 자기상관 민감도 (단위지역 : 동)

*wmBin*을 이용한 경우 노인인구비율의 자기상관은 구 단위의 경우 0.5에 비해 동 단위는 0.69로 공간규모의 효과가 여실히 드러난다.

생활보호대상자비율처럼 공간적 자기상관의 경향이 미미한 변수의 경우에도 근린 기준이 완화될수록 자기상관 정도가 감소하는 경향을 보인다. 하지만, 감소의 폭은 자기상관성이 있는 변수들에 비해 적은 것으로 나타났다.

위에서 산출한 각 추정값의 통계적 유의성과 표준오차의 상대적 증감 패턴을 관찰했으며, 결과에 대한 해석에 앞서 몇 가지 원론적인 사항은 다음과 같다. 통계적 유의성 검정에 있어서 유의수준이 정해지고 나면 검정력 함수의 크기를 결정하는 독립적인 요소는 추정값의 강도와 표본크기 두 가지이다.

첫째, 다른 조건이 동일할 때, Moran의 *I* 추정값이 0에서 벗어날수록 유의성은 증가한다. 구 라는 같은 표본 조건하에서 추정값의 상대적 강도는 유의성의 상대적 크기로 반영된다. 이것은 표본크기 *n*과 근린행렬 *W*가 동일한 경우 모든 변수에 대한 추정의 표준오차는 동일하기 때문이다(Goodchild, 1986).

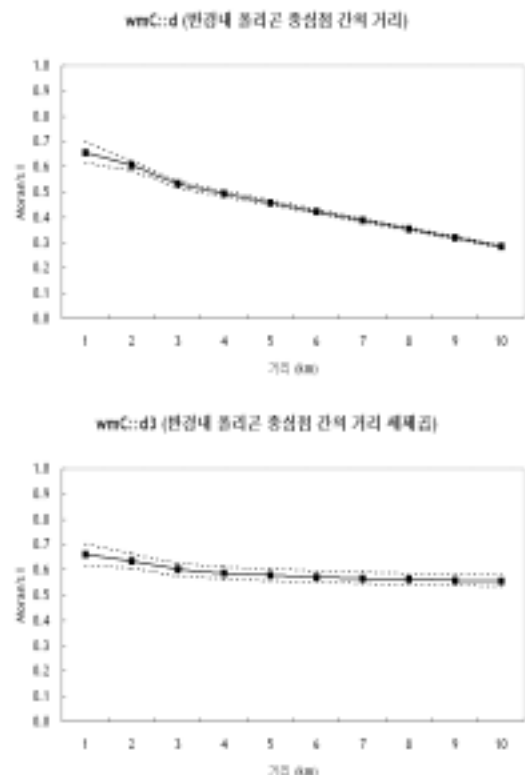
둘째, 동일 조건에서 표본의 크기가 커질수록 유의성 검정을 위한 표준오차가 작아지고 따라서 검정력 함수의 크기도 증가한다. 본 연구의 실험에서 표본의 크기는 두 가지 방식으로 제어된다. (i) 공간단위의 차이에 의한 표본크기의 변화로서, 구 단위와 동 단위의 경우 표본크기에 큰 차이가 있다. (ii) 동일한 공간단위라 하더라도 표본오차를 산출하는 표본크기는 *W*에 따라 조금씩 차이가 있다.

앞서 나왔던 <그림 1>은 노인인구비율을 구 단위로 분석했을 때 표준오차의 규모와 증감을 보여준다. 점선 부분은 거리를 매개변수로

신뢰구간대(confidence belt)가 변하는 양상을 나타낸다. 이와 비교해서 <그림 16>은 동일 변수에 대하여 동 단위 Moran의 *I* 추정치의 표준오차가 매우 작아진 것을 보인다.

한편 *W*의 매개변수인 거리가 증가함에 따라 근린의 수가 증가하게 되고, 같은 거리라 해도 거리조각 승수가 낮을수록 근린의 수는 증가한다. 근린 증가에 따라 표준오차가 일관되게 감소(monotonic decrease)하는 것을 <그림 1>과 <그림 16>에서 확인할 수 있다.

이러한 감소의 정도는 구 단위 자료와 같이 표본크기가 작을 때 더욱 민감하게 두드러진다. 한편 동 단위 자료의 경우 표준오차 자체가 0.01 수준으로 매우 작기 때문에 *W*에 의한 변화도 극미한 차이만 보인다.



<그림 16> Moran의 *I* 표준오차 범위
(서울시 노인인구비율 : 2000년 동 단위)

IV. 요약 및 결론

1. 연구의 요약

공간적 근린의 개념이 매우 다양하게 정의되고 수치화될 수 있다는 것은 잘 알려져 있지만, 이러한 근린의 개념을 활용하는 여러 공간 분석기법들이 근린의 정의에 따라 어떻게 또 얼마나 영향을 받는지에 대해 알려진 바는 거의 없었다.

본 연구는 서울시에 대하여 그 근린구조를 결정하는 근린가중치행렬 W 를 설정하는 방식이 공간자기상관의 추정에 미치는 영향과 그 민감도를 분석하기 위해 다음과 같은 질문으로부터 출발하였다. (i) W 의 생성방법에 따라 공간적 자기상관 지수와 그 유의성은 어떻게 달라지는가? (ii) 구, 동과 같은 공간적 규모에 따라, W 에 대한 공간적 자기상관 지수의 민감도가 달라지는가? (iii) 변수에 내재한 공간적 자기상관 정도에 따라 W 의 영향이 다른 양상을 보이는가?

본 연구를 통해 얻어진 결과는 다음과 같이 요약할 수 있다. 첫째, W 의 선택과 매개변수의 설정은 공간적 자기상관 지수에 무시할 수 없는 큰 영향을 미치는 것으로 보인다. 그 영향력의 크기에는 다소 차이가 있을 수 있지만, 이러한 영향은 구 단위와 동 단위 자료, 자기상관성이 비교적 강하거나 미미하다고 선별된 변수들 모두에서 나타난다.

거리조락은 얼마나 가까운 거리에 있는 구역을 근린으로 정의하느냐는 문제를 실험하는 제어수단이다. 거리조락과 관련한 패턴은 거리

매개변수가 짧아질수록, 또 거리조락 승수가 올라갈수록 자기상관이 상대적으로 과대 추정되는 것이며, 본 연구에서 그 편차의 규모와 민감도를 구체적으로 밝혔다. 본 연구에서 처음 제안된 면적과 경계선 비율 영향요소를 함께 고려하는 wmO 의 경우 거리에 따른 추정치간 편차가 가장 작게 나타난다는 것은 후속 연구가 필요한 주목할 성과이다.

둘째, 공간적 자기상관 지수의 민감도는 주로 W 의 종류에 의해 좌우되지만, 공간적 규모에 따라 상이하게 나타난다. 예를 들어, 노인인구비율의 경우 $wmBin$ 을 사용했을 때 구 단위 자기상관성이 0.5이지만 동 단위 경우 0.69로 측정되었다. 서울시 구 단위 자료의 경우, 단위 구역의 총 수가 상대적으로 적고 경계효과와의 개입이 불가피하므로 자기상관 지수 변화의 패턴이 동 단위 자료에 비해 불규칙적이며, W 의 설정에 특별한 주의를 요한다.

셋째, 자기상관 정도가 상대적으로 높은 변수들이 자기상관성이 미미한 변수에 비해 W 의 종류와 매개변수 설정에 더 민감하게 반응한다. 모든 경우 근린 기준의 완화에 따라 자기상관의 감소가 확인되지만, 자기상관 규모가 비교적 큰 변수들의 경우 큰 값에서 출발하여 극치에 수렴하므로 변화의 폭이 더 커진다.

넷째, 서울시 구 단위 자료의 분석은 표본의 크기가 작기 때문에 W 와 관계없이 자기상관 추정치의 표준오차가 커지게 되어 그 유의성 검정력이 떨어진다. 하지만, W 의 설정에 따라 근린이 많아질수록 그 표준오차는 일정하게 감소한다. 반면, 동 단위로 분석 시 표본이 충분히 크기 때문에 모든 변수의 자기상관 추정

치에 대한 검정력이 높고 이의 표준오차는 거의 0에 가까울 정도로 작다. 따라서 동일 변수에 대해 W 에 따른 검정력과 표준오차의 차이도 무시할 만큼 작은 범위에서 나타난다.

공간적 자기상관 측정에서 지금까지 가장 자주 혹은 유일하게 사용되는 근린구조는 경계선 공유여부에 따른 이진 가중치행렬 $wmBin$ 일 것이다. 따라서 $wmBin$ 에 기반한 추정치를 기준으로 삼고, 이들을 대안적인 근린설정 방식들과 비교 실험하는 것은 본 연구의 주요 관심사 중의 하나였다. $wmBin$ 은 서울시 동 단위 자료를 분석할 경우 근린의 영향권이 매우 짧은 거리로 한정되므로 자기상관을 상대적으로 과대 추정하는 경향이 본 연구를 통해 실증되었다. 이것은 $wmBin$ 의 활용빈도를 고려할 때 의미심장한 결과이며, 센서스 블록처럼 보다 미세한 단위의 자료를 다룰 때 과대추정 특성은 더욱 두드러질 것이다.

한편 공유경계선의 비율을 고려한 가중치행렬 $wmSB$ 의 경우 $wmBin$ 보다 자기상관이 일관되게 높게 추정되는 것도 매우 흥미로운 결과이다. 자기상관 추정치가 0.5 근방으로 나타난 변수에 있어서 이러한 차이는 더욱 두드러지는데, 이는 사회 통계자료의 대부분이 경험적으로 측정되는 자기상관 규모이기에 향후 $wmSB$ 의 역할이 부각되는 실험 결과이다.

서울의 동은 그 크기나 형태가 비교적 정형적이다. 하지만 단위구역의 분할형태가 상당히 불규칙한 지역의 경우 구역 간 공유경계선 비율의 편차는 커지게 된다. 이 때 동일한 근린구조를 사용하지만 $wmBin$ 에 비해 $wmSB$ 가 실제 근린의 영향력과 자기상관성을 보다 정

교하게 추정해낸다. 또한, 두 행렬 모두 동일한 근린의 수를 기준으로 하므로 표준오차는 동일하지만 $wmSB$ 의 경우 측정값이 크므로 유의성 또한 상대적으로 높게 나오게 된다.

지금까지의 분석을 종합하면 특정 현상의 공간분포에 대하여 자기상관성이 있다 혹은 없다고 진단하거나 자기상관이 어떤 크기로 나타났거나 하는 진술은 반드시 어떤 근린구조를 설정했는가에 따라 조건부로 해석되어야 할 만큼 W 의 영향이 크게 작용한다.

2. 연구의 함의

공간자료의 통계적 분석에 있어서 가변공간단위문제(Modifiable Areal Unit Problem: MAUP)는 Openshaw and Taylor(1979)의 실험에서 더욱 구체화 되었다. 공간단위개체들의 규모와 배열에 따라 피어슨 상관계수가 -1에서 +1까지 변할 수 있음을 보고한 이 실험은 공간통계 추정치에 어떤 절대적이며 유일한 타당성 부여가 불가능하다는 것을 시사한다.

Moran의 I 측정값이 0.2에서 0.9까지도 변할 수 있다는 본 실험의 함의도 이와 크게 다르지 않다. 이러한 한계를 고려할 때, 어떤 연구의 맥락과 상관없이 쓸 수 있는 최적의 W 를 제시하는 것은 애당초 불가능한 일이다. 가변공간단위문제와 마찬가지로 어떤 W 가 “올바른” 것이냐 혹은 잘못된 것이냐 하는 논의에 정답은 없으며, 차라리 공간현상의 이해와 해석을 위해 어떤 W 가 더 흥미롭고 또 “쓸모 있는지”를 판단하고 찾아내는 것이 공간통계기법을 사용하는 연구자들의 몫이고 그들의 능력이다.

이러한 예로, 공간자기상관이 확인된 후에 공간회귀모형을 세우는 시도에 있어서 가능한 두 가지 접근을 들 수 있다. 즉, 공간회귀모형에도 자기상관을 설명변수로 동원하는 모형이 있을 수 있고 또한 자기상관을 오차로 간주하고 이를 제거하는 모형도 가능하다(Anselin, 1988). 공간적 자기상관이 모형의 설명변수로 설정되는 경우라면 자기상관을 보다 적극적으로 포착해내는 W 를 찾는 것이 또 다른 외생 설명변수를 추가하는 것보다 유용한 접근이 될 수 있다.

GIS의 보급으로 보다 섬세하고 다양한 근린구조의 계산이 가능해졌다. 대상 지역과 연구의 맥락에 따라 가장 적합한 W 를 찾기 위한 탐색적 분석이 공간통계의 필수 절차로 정형화될 필요가 있으며, 본 연구의 결과는 이러한 탐색에 작은 지침이 될 것으로 기대한다.

참고문헌

- 김광구, 2003, “공간자기상관 (spatial autocorrelation)의 탐색과 공간회귀분석 (spatial regression)”, 『한국행정학회 2003년도 하계학술대회 발표논문집』, pp. 983~1002.
- 김종원, 2000, “주택시장에서의 공간자기상관의 검증 및 회귀계수의 추정”, 『경제학연구』, 48(2): 155~174.
- 서경천 · 이성호, 2001, “지가의 공간적 변동에 따른 입지 지대의 분석에 관한 연구 - 공간적 자기상관을 고려한 방법론을 중심으로”, 『대한국토도시계획학회지』, 36(1): 55~72.
- 서울시정개발연구원, 2000, 『지도로 본 서울 2000』, 서울시정개발연구원.
- 이지영 · 황철수, 2002, “공간통계분석을 이용한 지가의 입지값 추정에 관한 연구”, 『한국GIS학회지』, 10(2): 233~247.
- 환경부, 2000, 『1999년 12월 대기환경월보』, 환경부.
- Anselin, L., 1988, *Spatial Econometrics: Methods and*

- Models*, Dordrecht: Martinus Nijhoff.
- Anselin, L., 1995, “Local indicators of spatial association - LISA”, *Geographical Analysis* 7: 93~115.
- Bailey, T. C. and Gatrell, A. C., 1995, *Interactive Spatial Data Analysis*, Longman Scientific and Technical.
- Geary, R. C., 1954, “The contiguity ratio and statistical mapping”, *Incorporated Statistician*, 5: 115~141.
- Getis, A. and Ord, K., 1992, “The analysis of spatial association by use of distance statistics”, *Geographical Analysis*, 24(3): 186~206.
- Goodchild, M. F., 1986, *Spatial Autocorrelation*, CATMOG 47.
- Griffith, D., 2003, *Spatial Autocorrelation and Spatial Filtering*, Springer.
- Lee, H. Y. and Lee, Y. G., 2004, “Analysis on the Spatial Dimension of the Commercial Domains: the Case of Seoul, Korea”, *Journal of the Korean Geographical Society*, 39(2): 195~211.
- Lee, J. and Wong, D., 2001, *Statistical Analysis with ArcView GIS*, John Wiley & Sons.
- Moran, P., 1948, “The Interpretation of Statistical Maps”, *Journal of Royal Statistical Society*, 10: 243~251.
- Openshaw, S. and Taylor, P., 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*, London, Pion: pp. 127~144.
- Razavi, A. H., 2002, *ArcGIS Developer's Guide for VBA*, OnWord Press.
- Tobler, W., 1970, “A Computer Movie Simulating Urban Growth in the Detroit Region”, *Economic Geography*, 46: 234~240.
- Tobler, W., 1973, “Choropleth Maps without Class Intervals”, *Geographical Analysis*, 5(3): 262~265.
- Upton, G., 1990, “Information from Regional Data”, in Griffith, D. A. (Eds.), *Spatial Statistics, Past Present and Future*, Institute of Mathematical Geography: Monograph 12. Michigan Document Series: Ann Arbor. pp. 315~359.

원 고 접 수 일 : 2004년 8월 23일
최종원고채택일 : 2004년 9월 9일