

연구논문

지역 단위 조사연구와 공간정보의 활용 :  
지리정보시스템과 지리적 가중 회귀분석을 중심으로\*

GIS and Geographically Weighted Regression in the Survey Research of Small Areas

조동기\*\*

Dong - Gi Jo

본 연구는 조사연구의 과정에서 활용 가능한 공간분석의 유용성을 지리정보시스템(GIS)과 공간적 이질성을 고려하는 지리적 가중 회귀분석(GWR)을 통해 탐색한다. 많은 사회현상은 공간적 차원을 포함하고 있으며, GIS, GPS 단말장치, 온라인 위치기반 서비스의 발달로 위치정보의 수집과 활용이 용이해짐에 따라 조사연구의 과정에서 공간정보를 활용하는 분석이 이전보다 훨씬 더 용이해지고 있다. 관찰의 독립성과 오차의 동분산성을 가정하는 전통적 회귀분석은 공간적 의존성을 분석하지 못한다. GWR 분석은 속성정보뿐만 아니라 공간정보를 활용하는 공간분석 기법으로서, 공간적으로 근접한 사례들은 유사성을 가진다는 가정에 따라 지리적 가중 함수를 활용한다. A 기초자치단체 주민들을 대상으로 한 조사연구 자료를 공간정보와 결합시킨 후 간단한 행정만족도 모형을 추정해 본 결과, 지리적 가중 회귀분석은 전통적 회귀분석에 비해 공간적 자기상관의 문제를 극복하고 모형의 부합도를 증가시키는 것으로 나타났다. GWR 결과를 GIS와 결합시켜 독립변수 효과의 공간적 변이를 시각화시켜 봄으로써, 변수들의 효과와 관계를 더 자세하고 풍부하게 이해할 수 있다. 나아가서 이 기법은 특정 변수의 효과가 예외적으로 낮거나 높은 지역을 더 쉽게 밝혀냄으로써 정책방안을 모색하는 데에도 유용하게 활용될 수 있다.

**주제어:** 공간분석, 지리적 가중 회귀분석(GWR), 지리정보시스템(GIS), 자기상관, 모형부합도

This study investigates the utilities of spatial analysis in the context of survey research using Geographical Information System(GIS) and Geographically Weighted Regression (GWR) which take account of spatial heterogeneity. Many social phenomena involve spatial dimension, and with the development of GIS, GPS receiver, and online location-based services, spatial information can be collected and utilized more easily,

\* 이 논문은 동국대학교 학술연구비 지원에 의하여 연구되었음.

\*\* 동국대학교-서울 사회학과 조교수 조동기.

E-mail: infowell@dongguk.edu

and thus application of spatial analysis in the survey research is getting easier. The traditional OLS regression models which assume independence of observations and homoscedasticity of errors cannot handle spatial dependence problem. GWR is a spatial analysis technique which utilizes spatial information as well as attribute information, and estimated using geographically weighted function under the assumption that spatially close cases are more related than distant cases. Residential survey data from a Primary Autonomous District are used to estimate a model of public service satisfaction. The findings show that GWR handles the problem of spatial auto-correlation and increases goodness-of-fit of model. Visualization of spatial variance of effects of the independent variables using GIS allows us to investigate effects and relationships of those variables more closely and extensively. Furthermore, GIS and GWR analyses provide us a more effective way of identifying locations where the effect of variable is exceptionally low or high, and thus finding policy implications for social development.

**Key words** : spatial analysis, GWR, GIS, auto-correlation, goodness-of-fit

## I . 서론

취업률, 출산율, 범죄율, 학력, 소득수준, 삶의 질 등과 같은 많은 사회현상이나 과정은 공간적 위치에 따른 차이를 나타낸다. 최근 들어 지리정보시스템(GIS), GPS 수신기, 온라인 위치기반 서비스(LBS) 등과 같이 공간정보를 수집하고 처리하는 기술이 발달하고 공간 정보를 활용하는 분석기법이 개발됨으로써 조사연구(survey)의 과정에서도 공간적 의존성을 손쉽게 분석할 수 있는 여건이 만들어지고 있다. 공간정보를 수집하고 처리하는 GIS의 발전과 정밀한 수치지도(digital map)의 제작과 보급이 무엇보다 중요한 발전이지만, 휴대 전화를 통한 위치기반 서비스의 확산과 휴대성과 정확도가 높은 GPS 단말기의 보급도 중요한 변화이다. 최근에는 위성 지도나 네비게이션과 같이 인터넷이나 휴대용 단말기를 통한 위치기반 서비스의 활용도 일상화되고 있다.

사회현상의 공간적 변이를 실증적으로 분석하기 위해서는 연구대상이 되는 지역을 포괄하면서도 공간적 정보를 가진 자료가 필요한데, 대규모 센서스가 아닌 일반적인 조사연구의 자료는 그렇지 않은 경우가 많다. 전국적 규모의 일반적 조사연구에서는 자료의 지역적 포괄성이 확보되기 어렵다. 전국을 대상으로 하는 조사연구의 경우 기초자치단체(시·군·

구) 단위에서는 대부분 집락표집(cluster sampling)이 이루어져 다수의 지역이 표본에서 누락되기 때문이다. 그러나 기초자치단체와 같이 소규모 지역을 대상으로 한 조사연구에서는 연구지역을 포괄하는 표집이 가능하기 때문에, 공간정보를 활용하는 자료의 수집과 분석이 충분히 이루어질 수 있다.

조사연구 자료의 분석에서는 회귀분석이 흔히 사용되는데, 일반최소자승법(OLS)으로 추정되는 전통적 회귀분석은 관찰들 간의 독립성과 오차의 동분산성(homoscedasticity)을 가정하기 때문에 이러한 공간적 변이의 측면을 분석하지 못한다. 전통적 회귀분석에서는 표집 위치에 상관없이 관찰들은 서로 독립적이어야 하고, 종속변수의 관찰값과 추정값의 차이인 오차(errors)가 상호 독립적이며 분산이 일정한 것으로 가정한다(김두섭·강남준 2008; Berry & Feldman 1985; Fox 1991). 그러나 공간적으로 근접한 위치에서 표집된 사례일수록 유사한 값을 가지는 경향이 있기 때문에 현실적으로 이러한 가정이 충족되기는 어렵다. 따라서 공간적 이질성 또는 의존성이라고 하는 이러한 특성들이 존재하는 사회현상을 OLS 회귀모형으로 분석하게 되면 모수 추정치의 효율성이 떨어지게 된다.

공간적 변이를 고려하는 다양한 공간통계 분석기법이 발전되어 왔는데, 그 중에서도 지리적 가중 회귀분석(Geographically Weighted Regression: GWR)이 그 대표적인 예라 할 수 있다. 최근에는 GIS와 공간적 분석이 통합되어 공간적 변이를 시각적으로 분석할 수 있게 되었다(DeMers 2004; LeGates 2005; Lo & Yeung 2002). 공간분석은 물리적 공간에 대한 정밀하고도 체계적인 검토를 요구하는데, 공간적 자료를 관리, 분석, 표시하는 효율적인 정보시스템인 GIS는 사회과학자들이 공간분석을 수행할 수 있게 해주는 강력한 도구가 될 수 있다(구자문 1999; 이진 외 2006; 조춘만 2006; 통계청 2006; LeGates 2005). 이러한 환경의 변화에도 불구하고 지리학 등 일부 분야를 제외하면 사회과학적 연구에서 공간정보의 활용은 제한적으로만 이루어져 왔다.

본 논문은 기초자치단체의 주민들을 대상으로 한 조사연구 자료를 이용하여 공간적 변이를 분석하여 봄으로써, 조사연구의 과정에서 공간정보를 활용하고 공간적 의존성을 분석하는 방법의 유용성을 평가해 보고자 한다. 아래에서는 GIS 활용과 GWR 기법에 대해 설명한 후 행정만족도에 대한 간략한 회귀모형을 구성하고 지리적 가중 회귀분석을 실시함으로써, 일반적 회귀분석이 해결하지 못하는 공간적 변이에 대한 공간 분석기법의 장점과 과제를 검토하고자 한다.

## II. GIS의 활용과 지리적 가중 회귀(GWR) 분석

### 1. 조사연구와 GIS의 활용

사회과학적 연구에서 다루어지는 대부분의 자료는 공간적 위치와 관련되어 있으며, 차이나 동질성, 상관관계나 상호작용이 나타나는 지역을 확인하는 작업은 사회과학적 분석에서 중요한 과정의 일부이다. 지리정보시스템(GIS)은 개별 수준의 자료를 맥락정보와 연결시켜 공간적 비교를 가능하게 해주기 때문에 사회과학적 연구에서 강력한 도구로 활용할 수 있다. 이에 따라 GIS의 역사는 그리 길지 않지만 사회과학 분야에의 활용 가치에 대한 인식이 빠르게 확산되고 있다.

GIS는 지리학 분야에서 가장 많이 활용되고 있지만, 사회현상의 공간적 측면에 대한 이해는 여타 사회과학 분야에서도 요구되기 때문에 GIS를 이용한 공간분석은 다양한 분야에서 활용할 수 있다. 예컨대 사회학자들은 성별구성이나 직업구조 등과 같은 쟁점을 탐구하기 위해 센서스 자료를 사용하는데, 이러한 쟁점들은 공간적 차원을 가지고 있기 때문에 GIS를 활용함으로써 센서스 자료에 대한 훨씬 더 풍부한 사회학적 분석을 할 수 있다 (Bertaud & Stephen 2003; Brewer & Suchan 2001; Peters & MacDonald 2004).

도시 내부의 재개발과 대규모 주택단지의 조성, 지하철역의 신설이나 새로운 도로망의 개통 등은 주어진 지역의 공간적 특성을 변화시킴으로써 많은 사회경제적 영향을 미치게 된다. 이러한 변화는 비교적 단기간에 걸쳐 일어나는 경우가 많기 때문에 행정구역을 근간으로 하는 전통적인 표집의 과정에서는 공간적 특성의 변화가 반영되기 어렵다. 그런데 조사설계의 과정에서 GIS를 통해 행정구역 수치지도와 최신의 위성지도나 각종 주제도를 중첩시켜 봄으로써 연구대상 지역의 공간적 특성을 더 잘 이해할 수 있게 된다.

조사연구의 과정에서 GIS를 활용하기 위해서는 연구에서 사용되는 분석의 단위에 대한 위치정보가 필요하다. 위치정보는 주소의 형태로 표집틀(sampling frame) 속에 포함될 수도 있고, 자료를 수집하는 과정에서 사후적으로 파악될 수도 있다. 표집틀 속에 위치정보가 포함되어 있는 경우에는 모집단의 경계를 확인해 볼 수 있고, 실사의 과정에서 실사 위치의 확인과 접근경로 탐색이 가능하기 때문에 다양한 측면의 오차를 줄일 수가 있다.

GIS 활용의 보다 큰 이점은 공간분석 능력이다(Blogh 2003). GIS를 이용한 공간분석은 공간정보와 속성정보를 결합시킴으로써 공간적 변이와 유형을 쉽게 포착할 수 있게 해준다. 공간분석을 통해 특정한 속성을 공유하는 주변 지역이나 특정 현상이 빈발하는 관심

지역을 쉽게 밝혀낼 수도 있다. 또한 주요 GIS에서는 공간적 맥락정보의 활용이 용이해서 시설이나 도로망에 대한 근접성이나 접근성 등과 같이 전통적 조사연구에서는 사용하기 어려운 변수들도 분석에 활용할 수가 있다.

최근까지만 해도 지리학 외부의 GIS 활용은 주로 지도 그래프를 작성하는 용도로 이용했을 정도로, 지도를 통한 조사결과의 보고와 효율적 소통의 측면도 GIS의 중요한 이점 중의 하나이다. 공간정보를 포함하고 있는 자료의 경우, 잘 만들어진 지도 그래프는 다른 형태의 도표나 그래프가 제시할 수 없는 수많은 정보를 효과적으로 전달할 수 있다.

요컨대 조사연구의 실행과정에서 GIS는 연구의 설계, 자료의 수집과 분석, 결과의 보고 등의 과정에서 특히 중요한 기여를 할 수 있다. GIS를 활용함으로써 모집단의 특성을 더 잘 이해할 수 있게 되고 보다 대표성 있는 표본과 더 정확한 추정치를 얻을 수가 있다. 나아가서 공간적 유형을 분석할 수 있고 지도 그래프를 이용함으로써 조사결과의 보고도 효율적으로 이루어질 수 있다.

## 2. 지리적 가중 회귀분석의 논리

지리적 가중 회귀분석(GWR)은 공간적 분석기법의 하나로 속성정보와 더불어 지리적 정보를 활용한다. 일반최소자승법(OLS)을 이용하는 전통적 회귀분석은 공간적으로 안정된(stationary), 그리하여 위치에 따른 차이가 없는 것으로 가정한다. 즉 독립변수의 효과가 연구대상 지역의 모든 곳에서 동일하게 나타나는 것으로 가정한다. 그러나 행정만족도와 같은 많은 사회현상은 접근성과 밀접한 관련이 있고, 특히 본 연구의 대상이 되는 지역과 같이 지역의 특성이 복합적인 경우에는 공간적 변이가 더 많이 나타날 수밖에 없다.

지리적 가중 회귀분석은 바로 이러한 공간적 변이를 분석하는 회귀분석의 한 방법으로(Brundson & Fotheringham 2002; Fotheringham et al. 2002), ‘모든 것은 관련되어 있지만 공간적으로 가까운 것과 더 밀접하게 관련되어 있다’(Tobler 1970)는 소위 ‘지리학 제 1법칙’에 기초한 것이다. 지리적 가중 회귀분석의 핵심은 회귀계수를 상수가 아니라 위치의 함수로 간주한다는 것이다. 즉 회귀계수가 모든 지역에서 동일한 것이 아니라 공간적 위치에 따라 달라지도록 모형화한 것인데, 알려진 어떤 위치  $u$ 에 대해 종속변수  $y$ 와  $m$ 개의 독립변수  $x$ 가 있고, 회귀계수와 오차항이 각각  $\beta$ ,  $\epsilon$ 인 경우 회귀식은 (1)과 같이 표현된다.

$$y_i(u) = \beta_{0i}(u) + \beta_{1i}(u)x_{1i} + \beta_{2i}(u)x_{2i} + \dots + \beta_{mi}(u)x_{mi} + \epsilon_i(u) \quad (1)$$

식(1)이 일반적인 회귀모형과 다른 점은 각 항에 ( $u$ )가 붙는 것인데, 이는 모수의 추정 이 주어진 위치  $u$ 에 대해 이루어지고 그 위치에 대해서만 의미를 가진다는 것을 나타낸다. 회귀계수에 대한 가중치 이루어진다는 점에서 지리적 가중 회귀는 가중최소자승법 (Weighted Least Squares: WLS)의 일종으로 볼 수 있으며, 다만 가중치가 연구대상 지역 내의 위치에 따라 달라진다는 차이가 있다. 이에 따라 회귀계수에 대해 위치에 따른 가중치가 부여되고 추정은 식(2)와 같이 가중최소자승법에 따라 이루어진다.

$$\hat{\beta}(u) = [X^T W(u) X]^{-1} X^T W(u) Y \quad (2)$$

이 식에서 지리적 가중치 행렬  $W(u)$ 의 각 요소는 가중 함수인 커널(kernel)에 따라 계산된다. 커널은 다양한 형태를 취할 수 있지만 일반적으로 가우스 형태인 식(3)이 사용된다.

$$w_i(u) = e^{-0.5(d_i(u)/h)^2} \quad (3)$$

여기서  $w_i(u)$ 는 위치  $u$ 에 대한 사례  $i$ 의 지리적 가중치이고,  $d_i(u)$ 는 사례  $i$ 와 위치  $u$  간의 거리이고,  $h$ 는 대역폭(bandwidth)이라고 하는 일정한 값이다. 대역폭이 커지면 가중치가 1에 근사하고 따라서 지리적 가중 회귀모형은 공간적 변이를 고려하지 않는 일반적 회귀모형에 근사하게 된다.

일반적으로 자료가 수집된 위치인 표본점(sample points)과 모수의 추정이 이루어지는 위치인 회귀점(regression points)이 일치하는데, 이에 따라 추정치와 잔차의 계산이 가능하고 모형의 부합도에 대한 평가도 가능하게 된다. 모형의 부합도는 대역폭의 영향을 많이 받는데, 연구대상 지역에서 표본점들이 규칙적으로 분포하는 경우에는 고정된 대역폭을 사용해도 되지만 불규칙하게 흩어져 있는 경우에 고정된 대역폭으로 추정을 하게 되면 모형의 부합도가 떨어지게 된다. 또한 고정된 대역폭은 표본점이 조밀하게 분포하는 곳에서는 미묘한 공간적 차이를 포착하지 못할 수 있고, 산발적으로 분포하는 곳에서는 추정치의 분산을 증가시킬 위험이 있다.

표본점들이 불규칙하게 분포하는 경우에는 대역폭을 가변적으로 만들어서, 조밀한 지역에서는 커널이 작아지고 산발적인 지역에서는 커널이 커지게 함으로써 모형의 부합도를 증가시킬 수 있다. 일반적으로 모든 회귀점에 대해 동일한 수의 표본점들이 포함되도록 대역폭을 설정하는 식(4)의 방식이 사용된다.

$$w_i(u) = (1 - (d_i(u)/h)^2)^2 \text{ (단, } d_i(u) > h \text{ 이면 } w_i(u) = 0) \quad (4)$$

이 방식의 특징은 어떤 사례의 거리가 대역폭보다 큰 경우 주어진 지역의 모수 추정에서 제외되도록 한다는 것이다.

지리적 가중 회귀분석에서 흔히 사용되는 부합도 측정치는 수정 AIC(corrected Akaike Information Criterion: AICc)이다(Hurvich et al. 1998). 사례수가  $n$ 이고 종속변수와 그 추정치가 각각  $y$ ,  $\hat{y}$ 이고, 잔차의 표준오차 추정치가  $\hat{\sigma}$ 일 때 AICc의 형태는 식(5)와 같다.

$$AIC_c = 2n \log_e(\hat{\sigma}) + n \log_e(2\pi) + n \left( \frac{n + \text{tr}(S)}{n - 2 - \text{tr}(S)} \right) \text{ (단, } S = \frac{\hat{y}}{y}) \quad (5)$$

AICc는 동일한 종속변수에 대해 상이한 독립변수로 구성된 모형을 비교하는 데 유용하며, 아래에서 설명되는 전역모형과 지역모형을 비교하는 데에도 사용된다. 일반적으로 비교되는 두 모형에서 AICc 값의 차이가 4보다 작은 경우 두 모형은 사실상 차이가 없는 것으로 본다(Charton & Fotheringham 2009).

지리적 가중 회귀분석의 결과는 커널의 형태보다는 대역폭에 따른 영향을 더 크게 받는 것으로 알려져 있다(Fotheringham et al. 2002). 최근지점(nearest neighbors)의 수로 표현되는 적정 대역폭은 관찰값과 추정값 간의 차이인 교차검증(cross validation: CV) 값을 최소화하는 방법이나, 관찰값과 추정값의 차이뿐만 아니라 모형의 복잡성을 고려하여 수정 부합지수 AICc 값을 최소화하는 방법에 따라 계산될 수 있다.

일반적 회귀분석과 비교해 볼 때 GWR 분석 결과는 잔차가 훨씬 작고 공간적 의존성도 크게 낮은 반면에 자료와의 부합도는 높으며, 위치정보와 결합되어 있기 때문에 GIS에 입력되어 시각적 분석을 가능하게 해준다는 장점이 있다. 최근 들어 GIS와 공간분석 간의 통합이 촉진된 배경에는 공간적 통계기법의 하나인 GWR의 발전이 자리잡고 있다.

### III. 자료와 모형

#### 1. 자료

본 연구에 사용된 자료는 A 지역(기초자치단체)의 행정동 통폐합을 위한 주민 의견 조사 결과의 일부이다. 총 15개의 행정동으로 구성된 A 지역은 대도시의 중심에 위치하며,

업무지구, 주거지구, 복합지구의 구분이 비교적 뚜렷한 곳이다.<sup>1)</sup> 전체 조사의 표본은 총 1,250 사례로, 주민(통·반장, 주민대표 포함), 행정서비스 이용자(비거주자 포함), 주민센터에서 운영하는 생활문화 프로그램 이용자, 공무원 등을 대상으로 다단계층화집락 표집의 방법으로 구성되었다.

본 분석에 사용된 표본은 A지역 거주민 응답자 중 ‘면접원 사후 기록’란에 주소가 구체적으로 기록되어 위치 확인이 가능한 242 사례이다. 위치정보를 이용한 공간분석을 실시하기 위해서는 속성정보와 공간정보를 연결시키는 지오코딩(geocoding) 과정이 요구된다. 이 분석에서 각 사례의 위치정보 입력은 번지가 기록된 경우에는 지리정보시스템의 주소검색을 이용하였고, 검색 결과 일치하는 목록이 없는 경우에는 동별 상세 지도와 대조하여 위치를 확인하였으며, 번지가 없고 통과 반의 정보만 있는 경우에는 해당 반의 가운데 지점 좌표를 사용하였다.

위치정보를 보유한 표본의 지역별 구성비를 전체 표본의 구성비와 비교해 보면, 복합지구에서는 큰 차이가 없지만 업무지구에서는 다소 높아졌고 주거지구에서는 다소 낮아졌다(〈표 1〉 참조). 업무지구에서 구성비의 차이가 상대적으로 크기는 하지만 전반적으로 심각한 표본의 왜곡은 없는 것으로 보인다.

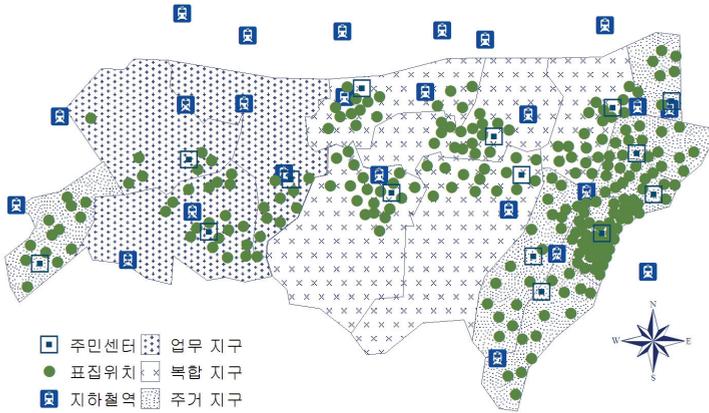
표본의 분포를 지도상으로 표시해 보면 업무지구, 학교나 병원, 기타 대규모 시설이나 녹지가 많은 지역의 행정동에서는 상주 거주민의 수도 적어서, 소수의 사례가 일부 제한된 공간에 분포하는 모습을 나타낸다. 반면에 주거지구의 비중이 높은 행정동에서는 다수의 사례가 비교적 골고루 분포하는 모습을 보인다(〈그림 1〉 참조).

〈표 1〉 지역별 표본 구성비

(단위: 명, %)

지역 구분	본래 조사의 전체 표본	위치정보 보유 표본
업무지구	233(18.6)	35(14.5)
주거지구	592(47.4)	123(50.8)
복합지구	425(34.0)	84(34.7)
합 계	1250(100.0)	242(100.0)

1) 행정동의 업무, 주거, 복합지구로의 분류는 A 자치단체에서 사용하고 있는 통상적인 기준을 따른 것이다.



〈그림 1〉 A 기초자치단체의 지역 구분과 표집 위치

## 2. 모형 설정

공간적 이질성을 고려하는 분석 과정에서는 흔히 전역모형과 지역모형을 구분한다. 전역모형(global model)은 공간적 위치에 따른 차이가 없다는 가정에 따라 설정된 모형이고, 지역모형(local model)은 전역모형을 공간적으로 분해함으로써 공간적 위치에 따른 차이를 고려하는 모형이다. 어떤 변수의 효과가 모든 곳에서 일정하다고 전제하는 전통적 OLS 회귀분석은 전역모형에 해당되는 반면에, 지리적 가중치, 커널, 그리고 대역폭이 결합된 지리적 가중 회귀분석은 지역모형에 해당된다.

본 연구에서 전역모형은 행정만족도를 종속변수로 하고 주민센터 생활문화 프로그램 만족도, 응답자 가구와 주민센터 근접성, 응답자 연령,<sup>2)</sup> 온라인 주민 서비스 만족도, 지하철역 근접성을 독립변수로 하여 구성하였다. 만족도 변수는 리커트(5점) 척도로 측정하였다. 주민센터 및 지하철역 근접성(proximity)은 좌표상의 직선거리로 측정하되 위치 좌표가 근사적으로 입력된 사례가 많다는 점을 감안하여 실제 거리를 5단계로 구분하여, 주민센터 근접성은 200m 단위, 그보다 좀더 조밀하게 분포하는 지하철역 근접성은 150m 단위로 부호화하였다. 연령은 연단위로 입력하였다. 일반적으로 공공서비스의 질에 대한 주민만족도는 접근용이성, 편리성, 신속정확성, 쾌적성, 대응환류성, 업무공정성 등의 측면에서 평

2) 응답자의 관내 거주연수가 보다 적합한 독립변수가 될 것이나, 거주연수 변수의 조작화 문제로 관내 주민의 다수가 10년 이상 장기 거주자라는 지역의 특성을 포착하지 못하였다.

가되지만(김영오 2008), 위의 행정만족도 모형은 예시를 위한 간단한 모형이다. 분석 단계에서는 행정만족도에 영향을 미치는 모든 독립변수를 포함하는 것이 아니라는 점과, 일부 사례의 경우 위치정보가 근사적으로 입력되었다는 점이 고려되어야 할 것이다.

본 연구에서 지역모형은, 종속변수와 독립변수는 전역모형과 동일하지만, 공간적 이질성을 고려하여 추정 방법을 달리한 것이다. GWR 분석에 사용할 수 있는 패키지가 몇 가지 있지만, 본 연구에서는 자료의 시각화에 유용한 ArcGIS의 공간적 통계분석 모듈의 하나인 지리적 가중 회귀모형 도구를 이용하였다.<sup>3)</sup> GWR 분석을 위해서 분석 환경을 설정해야 되는데, 먼저 지리적 가중을 위한 커널(kernel) 함수의 유형은 연구지역 내에서 관찰 사례의 위치가 불규칙하게 분포하므로 식(4)에 따라 가중치를 계산하는 ‘ADAPTIVE’ 방식을 선택하였다. 다음으로 대역폭(bandwidth)이 계산되는 기준에 대한 설정은 교차검증(CV) 값 또는 AICc 값에 따라 자동적으로 계산되게 하거나 직접 입력하는 방법이 있는데, 사전적으로 대역폭을 지정할 만한 충분한 근거가 없는 경우에는 자동적으로 계산되도록 하는 것이 관행이다(Carlton & Fotheringham 2008). 자동으로 설정하는 경우 관찰값과 추정값 간의 차이만 고려하는 CV 방식보다는 그 차이와 더불어 모형의 복잡성을 고려하는 AICc 방식이 선호된다. 본 분석에서도 수정 부합지수 AICc 값을 최소화시키는 방법으로 적정 대역폭을 계산하도록 설정하였다.

전역모형이 연구대상 지역 내의 유사성을 전제로 규칙성의 발견을 추구하는 반면에, 지역모형은 지역 내의 이질성을 전제로 예외적 현상을 찾아내려고 한다. 특히 지역모형의 분석 결과는 GIS와 결합됨으로써 추가적인 분석을 가능하게 해준다. 아래에서는 모형 부합도의 측면에서 지역모형이 전역모형에 비해 우수하다고 할 수 있는지와 공간적 위치에 따라 효과의 변이가 어떻게 나타나는가를 분석하고자 한다.

GIS를 이용한 시각적 분석에서 특히 유의해야 할 부분은 주제 정보를 분류하는 기준을 선택하는 것이다(DeMers 2004; LeGates 2005). 범주 구분에 따른 정보의 왜곡은 다른 시각적 자료에서도 나타날 수 있는 문제이지만(Tufte 2001), 지도의 경우 주제 정보가 공간적 정보와 결합됨으로써 정보의 왜곡이 훨씬 더 심하게 나타날 수 있다. GIS에서는 흔히 동일 급간 분류, 분위수 분류, 자연적 분류, 인위적 분류 방식 등이 사용되는데, 본 연구에는 자연적 분류 방식을 따랐다. 자연적 분류는 자료의 분포 속에서 의미가 통하는 분류 지점을 찾아내는 알고리즘에 따라 자료를 분류하는 방식으로, 일반적인 상황에서는 가장 현실적인 분류방식이라 할 수 있다(LeGates 2005: 89).

3) ArcGIS에서 GWR 모듈은 9.3버전부터 가능하다. 본 연구를 위해 ArcGIS 사용권을 발급해 준 (주)선도소프트에 감사드린다.

## IV. 분석 결과

### 1. 전역모형 추정 결과

회귀모형의 추정에 앞서 <표 2>에 제시된 상관계수(Pearson's  $r$ )를 통해 변수들 간의 상관관계를 분석해 보면, 독립변수들 간에는 상관관계가 거의 없거나 정도가 낮아서 다중공선성(multicollinearity)의 문제는 없는 것으로 판단된다. 생활문화 프로그램에 대한 만족도, 연령, 그리고 주민센터 근접성 변수가 행정만족도 변수와 유의미한 관계를 가지고 있다. 프로그램 만족도 및 주민센터 근접성은 행정만족도와 정적인 관계를 가지는 반면에, 연령과 행정만족도는 부적 관계를 보여 나이가 많은 응답자일수록 행정만족도가 낮다는 점이 흥미롭다. 연령은 생활문화 프로그램 만족도와도 부적 관계를 가지고 있다.

<표 2> 변수들 간의 상관계수

	행정만족도	온라인 만족도	프로그램 만족도	연령	주민센터 근접성
온라인 만족도	0.109				
프로그램 만족도	0.469**	0.119			
연령	-0.194**	0.002	-0.131*		
주민센터 근접성	0.471**	0.032	0.083	-0.092	
지하철역 근접성	0.073	-0.015	0.059	0.002	0.100

주) \*  $p < .05$ , \*\*  $p < 0.01$ 에서 유의미

<표 3> 행정만족도 모형의 회귀분석 결과(OLS)

독립 변수	회귀 계수	t	VIF <sup>1)</sup>
연령	-0.007	-2.023*	1.025
생활문화 프로그램 만족도	0.375	8.237**	1.037
온라인 서비스 만족도	0.026	0.932	1.015
주민센터 근접성	0.410	8.544**	1.014
지하철역 근접성	0.006	0.883	1.013
모형 부합도	$R_{adj.}^2 = 0.408$ , $AICc^{2)} = 445.874$		
공간적 자기상관	Moran 지수(I) = 0.078**		

주) 1) Variance Inflation Factor    2) Akaike's Information Criterion

\*  $p < 0.05$ , \*\*  $p < 0.01$ 에서 유의미

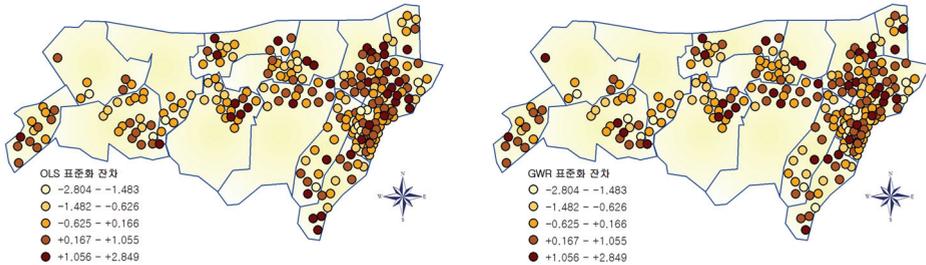
전역모형에 대해 일반최소자승법으로 추정한 회귀분석의 결과가 <표 3>에 제시되어 있다. VIF(Variance Inflation Factor) 지수는 위의 상관계수 분석과 마찬가지로 독립변수의 중복성이 없다는 것을 보여준다. 독립변수의 효과를 보면 연령, 생활문화 프로그램 만족도, 주민센터 근접성 변수가 유의미한 효과를 미치는 반면에, 온라인 서비스 만족도와 지하철역 근접성 변수의 효과는 없다. 연령이 적을수록, 주민센터에서 운영하는 생활문화 프로그램에 대한 만족도가 높을수록, 그리고 주민센터와의 거리가 가까울수록 행정서비스에 대한 만족도가 높아진다.

응답자의 60% 이상이 주민센터로부터 800m 이내의 거리에 거주하고, 지하철역이 주민센터로부터 멀리 떨어진 경우가 많기 때문에 행정서비스에 대한 지하철역 근접성 변수의 효과가 없는 것으로 보인다. 도시의 복잡한 도로망을 감안할 때 직선거리로 측정되는 근접성보다는 도로망이나 소요시간을 고려한 비용거리(cost distance)로 측정되는 접근성(accessibility)을 주목할 필요가 있다. 따라서 주민센터에 대한 접근성을 제대로 분석하려면 지하철역 근접성이 아니라 지하철이나 도로망을 통한 접근성을 고려해야 할 것이다.

전역모형의 부합도를 보면 수정 결정계수( $R_{adj.}^2$ )가 0.408로서 종속변수 변량의 약 41%가 독립변수에 의해 설명되고 있다. 공간적 자기상관(spatial auto-correlation)을 측정하는 잔차들 간의 모란지수(Moran's I)는 0.078로서( $p < .01$ ), 전역모형에서 발견되는 공간적 변이의 유형은 우연적인 요소에 의한 것이 아님을 보여 준다.

## 2. 공간적 자기상관과 모형 부합도 분석

공간적 자기상관을 분석하기 위해서는 먼저 전역모형과 지역모형에서 나온 잔차의 공간적 변이를 살펴볼 필요가 있다. 이 과정에서는 특히 유난히 높거나 낮은 잔차를 보이는 지역이 존재하는지와 잔차가 공간적 자기상관을 보이는지를 검토해야 한다. <그림 2>에 전역모형과 지역모형의 표준화 잔차의 공간적 변이가 제시되어 있는데, 일반최소자승법에 의한 전역모형의 잔차의 분포를 보면(<그림 2>의 (가) 참조), 대체로 주거지구에서 표준화 잔차가 큰 반면에 업무지구에서는 작은 편이다. 이러한 결과는 전역모형이 주거지구에서 행정만족도를 과소예측하는 반면에 업무지구에서는 과대예측하는 경향이 있다는 것을 보여준다. 지역모형의 표준화 잔차의 분포도 비슷한 양태를 보이지만, 전역모형에 비해 지역별 대비 정도가 약간 줄어들고 있다(<그림 2>의 (나) 참조).



(가) 전역모형(OLS)

(나) 지역모형(GWR)

〈그림 2〉 전역모형과 지역모형의 잔차(표준화)의 공간적 변이

공간적 자기상관은 지역별 잔차의 분포에서도 드러나지만, 모란지수(I)를 통해 그 정도를 구체적으로 측정해 볼 수 있다(Fotheringham et al. 2002). 앞서 보았듯이 전역모형에서는 공간적 자기상관이 존재하는 것으로 나타났지만, 지역모형에서는 모란지수가 0.023로서( $p > .05$ ) 자기상관을 의심할 만한 근거가 없다(〈표 4〉 참조). 전역모형의 잔차에서 나타났던 공간적 의존성이 지역모형에서는 지리적 가중을 통해 제거된 것이다.

공간적 이질성을 고려하는 지역모형이 전역모형의 부합도를 개선시키는 정도는 잔차의 분포와 부합도 지수를 통해 평가될 수 있다. 먼저 잔차의 경우 지역별로 잔차의 분산이 고를수록 지역모형의 부합도가 커진다는 점을 고려할 때, 전역모형에 비해 지역모형의 부합도가 다소 증가하는 것으로 보인다. 그러나 〈그림 2〉에서는 이러한 차이가 두드러지지 않는다.

잔차의 분포와 아울러 부합도의 개선 정도를 평가하는 통계치를 살펴볼 필요가 있다(Burnham & Anderson 2002). 전역모형에 대한 회귀분석 결과인 〈표 3〉과 지역모형에 대한 지리적 가중 회귀분석의 결과인 〈표 4〉를 비교해 보면, 일반적인 부합도 측정치인 수정 결정계수( $R_{adj}^2$ )가 전역모형에서 0.408이고 지역모형에서는 0.470으로 나타나 부합도가 약간 개선되고 있다. 두 모형간의 부합도 평가에서 더 선호되는 통계치인 수정 AIC(corrected Akaike Information Criterion: AICc) 지수도 살펴볼 필요가 있다. AICc는 상대적 측도로써 한 값의 크기 자체보다는 두 모형 간에서 나타나는 값의 차이가 중요하다(Carlton & Fotheringham 2008, 2009). 일반적인 관행에 따라 두 모형 간 AICc 값의 차이가 4보다 클 때 모형의 개선이 있는 것으로 간주한다. 전역모형의 AICc 지수는 445.874이고 지역모형에서 그 값은 431.564로서 지역모형의 부합도가 유의미하게 개선된 것으로 보인다.

〈표 4〉 지리적 가중 회귀 분석(GWR) 결과 요약

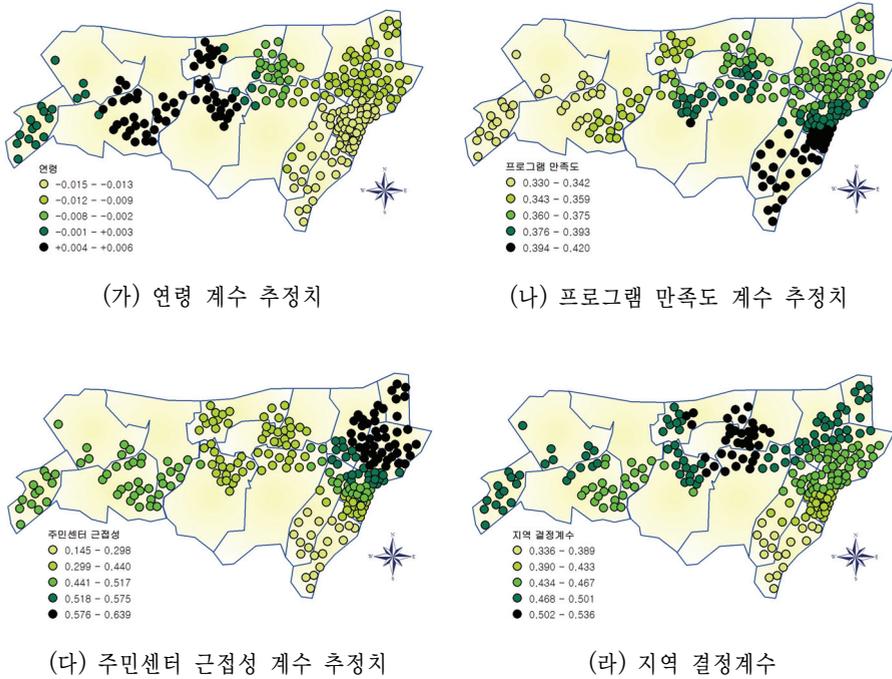
독립 변수	최소값	중앙값	최대값
연령	-0.015	-0.011	0.006
생활문화 프로그램 만족도	0.329	0.371	0.420
온라인 서비스 만족도	0.016	0.027	0.060
주민센터 근접성	0.145	0.473	0.639
지하철역 근접성	-0.107	0.159	-0.029
지역 결정계수	0.336	0.466	0.536
대역폭(최근지점)	168		
모형 부합도	$R_{adj}^2 = 0.470$ , $AICc^* = 431.564$		
공간적 자기상관	Moran 지수(I) = 0.023 (p > .05)		

주) \* Akaike's Information Criterion

### 3. 회귀계수와 결정계수의 공간적 변이

공간적 자기상관과 모형 부합도 분석의 결과에 따르면 행정만족도 모형은 공간적 의존성을 가지고 있기 때문에 지리적 가중 회귀분석을 통해 제대로 분석될 수 있다. 회귀계수와 결정계수의 공간적 변이를 자세히 보기 위해서는 이들 계수를 지도상에 표시하여 시각적 분석을 해 볼 필요가 있다.

지역모형의 계수 추정치를 지도상에 표시해 보면, 전역모형에서 계수 추정치가  $-0.007$ 이었던( $p < .05$ ) 연령의 경우 지역모형에서는  $-0.015$ 에서  $0.006$ 의 범위를 보이는데, 업무지구에서는 양의 효과를 가지는 반면에 주거지구나 복합지구에서는 음의 효과를 가지는 경향이 있다. 효과의 크기(절대값)는 업무지구보다 주거지구에서 더 큰 편이다(〈그림 3〉의 (가) 참조). 프로그램 만족도의 경우 전역모형에서 그 계수가  $0.375$ 였는데( $p < .01$ ) 지역모형에서는 모두 양의 효과를 가지며  $0.329$ 에서  $0.420$ 의 범위를 보인다. 위치에 따른 분포를 보면 연령 계수와 유사한 분포를 보여서 주거지구에서 크고 업무지구에서 작은 편이다(〈그림 3〉의 (나) 참조). 전역모형에서 계수가  $0.410$ 였던( $p < .01$ ) 주민센터 근접성은 지역모형에서는  $0.145$ 에서  $0.639$ 의 범위로 모두 양의 효과를 가진다. 효과의 크기를 보면 주거지구 중에서도 북부에서는 큰 반면에 남부에서는 업무지구나 복합지구에 비해 더 작은 편이다(〈그림 3〉의 (다) 참조).



〈그림 3〉 지역모형의 회귀계수와 결정계수의 공간적 변이

한편 지역모형의 결정 계수는 0.336에서 0.536의 범위를 보이는데, 대체로 보면 복합지구에서 높고 주거지구에서 낮은 편이다(〈그림 3〉의 (라) 참조).

지역 계수를 지도상으로 살펴본 결과, 변수들의 효과가 위치에 따라 차이를 보인다는 점이 드러났다. 연령과 생활문화 프로그램 만족도의 경우는 주거지구에서 효과가 큰 반면에 업무지구에서는 낮은 편이다. 프로그램 만족도와 주민센터 근접성 변수의 경우 지역모형에서 모든 계수의 방향이 전역모형의 계수 방향과 일치한다. 그러나 연령 변수의 경우 전역모형에서는 음의 방향이었지만 지역모형에서 일부의 계수는 양의 방향을 나타낸다. 이러한 결과는 종속변수에 대한 이들 독립변수 효과의 공간적 변이가 복잡하게 나타날 수 있다는 것을 보여준다.

공간적 이질성이 존재하는 경우 종속변수와 독립변수 간의 관계가 모든 위치에서 동일한 것이 아니어서, 독립변수가 동일한 정도로 변화더라도 그것에 영향을 받는 종속변수의 크기는 위치에 따라 달라진다. 지리적 가중 회귀분석은 이러한 공간적 이질성을 분석하는 유용한 기법이라 할 수 있다.

## V. 요약 및 결론

행정만족도에 대한 GWR 분석 결과는 A 지역 내의 위치에 따라 모형에 포함된 독립변수의 효과가 차이가 있다는 것을 보여준다. 이와 같이 연구대상 지역 내에서 상당한 공간적 변이가 존재하는 경우, GWR 기법은 지리적 가중을 통해 이러한 공간적 변이를 효과적으로 분석할 수 있게 해준다. 행정만족도 모형을 통해 본 바와 같이 GWR은 잔차의 변량을 줄이고, 전역모형에서는 발견되기 어려운 공간적 유형을 보여줌으로써 전통적 회귀분석을 개선 시키게 된다. 행정만족도 모형에 포함된 독립변수의 효과는 공간적 변이를 보이고 있고 그로 인해 종속변수와 독립변수들 간의 관계를 표현하는 모형들의 부합도가 공간적 위치에 따라 다르게 나타난다. GWR 분석은 일반최소자승법을 이용하는 전통적 회귀분석에 비해 잔차도 적고 공간적 의존성도 낮아서 모형의 복잡성을 고려하더라도 전체적인 모형의 부합도를 증가시켜 준다.

많은 사회현상은 공간적 차원을 포함하고 있는데 공간이 사회학적 연구에서 중요한 범주로 고려되기 시작한 것은 1980년대 말의 ‘공간적 전환’ 이후이다. 일부 사회현상은 공간적 요소를 고려해야만 제대로 분석할 수 있다는 인식이 확산되면서, 일상생활의 미시적 과정이나 사회나 국가 단위의 거시적 현상을 설명하는 맥락에서 위치, 장소, 그리고 지역이 중요한 변수로 주목받기 시작한 것이다. 초기의 공간에 대한 관심은 주로 마르크스주의 접근에서 비롯되어 공간의 구조적 결정성을 강조하는 경향이 강했다(Castells 1972; Harvey 1989; Lefebvre 1992). 보다 최근의 연구들은 사람들의 인식, 회상, 사고의 과정에서 공간이 구성되고 구조적 속성으로 드러나게 되는 측면을 주목한다(Löw 2008). 이에 따르면 사회적 행위자로서의 개인이 공간을 구성하는 측면과 개인의 행위가 사회경제적 구조뿐만 아니라 공간적 구조에 의해서 제약을 받는 측면이 동시에 존재한다. 즉 공간은 행위의 산물인 동시에 행위를 구조화시키는 작용을 한다는 것이다.

조사연구의 과정에서도 이러한 공간적 차원을 분석하기 위한 연구자들의 노력이 요구된다. 특히 조사연구의 표집단계에서부터 표집대상의 위치정보를 고려한 표집설계가 이루어질 필요가 있다. 최근 들어 정확도가 높은 GPS 단말기와 위치정보 확인 기능이 내장된 휴대전화 단말기의 보급이 늘어나고 있기 때문에 큰 비용 부담이 없이도 조사연구의 과정에서 공간정보를 활용하는 것이 가능해지고 있다. 서두에서 언급되었듯이 조사연구의 과정에서 GIS의 활용은 많은 이점을 가지고 있기 때문에 이를 활용하는 연구가 점차 일반화될 것이다. 이에 따라 속성정보와 공간정보 간의 결합이 더 늘어날 것이고, 그 결과 GWR과

같은 공간분석 기법의 활용도 늘어날 것으로 예상된다.

요컨대 GWR은 자료 속에 들어 있는 공간적 변이를 양화시키고 시각화시켜 주며, 모형의 부합도를 증가시킴으로써 공간적 변이가 중요한 의미를 가지는 지역적 현상을 더 잘 이해할 수 있게 해준다. 속성정보뿐만 아니라 공간정보를 활용하는 공간분석 기법인 GWR 분석은 공간적으로 근접한 사례들은 유사성을 가진다는 가정에 따라 지리적 가중 함수를 활용한다. GWR 분석의 결과는 GIS와 결합되어 더 풍부한 분석을 가능하게 해준다. 종속 변수에 영향을 주는 독립변수 효과의 공간적 변이를 시각화시킴으로써 변수들의 효과와 관계를 더 자세하고 분명하게 분석할 수 있다. 나아가서 특정 변수의 효과가 예외적으로 낮거나 높은 지역을 더 쉽게 찾아내게 함으로써, 그 원인을 추적하여 문제를 개선시키는 방안을 모색할 수 있게 해준다.

그러나 GWR 분석기법이 모든 종류의 공간적 분석에 활용될 수 있는 만병통치약은 아니다. 지역별 계수의 유의도에 대한 검증과 평가가 쉽지 않는 등 아직 개선의 여지도 많다 (Griffith 2008). 나아가서 표집이나 실사의 과정에서 표본의 위치에 대한 자세한 정보가 기록되는 경우 조사연구의 익명성과 비밀보장의 원칙과 위배될 수 있기 때문에, 위치정보 활용과 관련된 연구자의 윤리의식이 더 많이 요구된다.

## 참고 문헌

- 구자문. 1999. “도시분석을 위한 인구주택센서스와 GIS의 연계활용방안 연구—수치지도의 보완과 센서스트랙의 결정.” 《한국지리정보학회지》 제2권 제2호: 27-45.
- 김두섭·강남준. 2008. 《회귀분석: 기초와 응용》(개정판). 나남출판사.
- 김영오. 2008. “주민만족도 평가에 관한 실증적 연구.” 《한국행정논집》 제20권 제2호: 601-633.
- 이건·이명진·서우석·변미리. 2006. “2005년 인구주택총조사의 조사구 설정 방법.” 《조사연구》 7(1): 109-129.
- 조춘만. 2006. “GIS를 이용한 센서스기반 근린주구 및 주민인지 근린주구 간 영역계 비교연구.” 《국토계획》 41(3): 151-162.
- 통계청. 2006. “조사구 웹 기반 관리 시스템 구축”. [http://www.nso.go.kr/board\\_info/BoardAction.do?method=view&board\\_id=137&seq=4&num=4&parent\\_num=0&page=1&sdate=&edate=&search\\_mode=&keyword=&position.](http://www.nso.go.kr/board_info/BoardAction.do?method=view&board_id=137&seq=4&num=4&parent_num=0&page=1&sdate=&edate=&search_mode=&keyword=&position.)
- Berry, W. D. and Feldman, S. 1985. “Multiple Regression in Practice.” *Sage University Paper Series on Quantitative Applications in the Social Sciences* 07-050. Sage Publications.
- Bertaud, A. and Stephen M. 2003. “The Spatial Distribution of Population in 48 World Cities:

- Implications for Economies in Transition.” *The Centre for Urban Land Economic Research*. University of Wisconsin.  
[http://alain-bertraud.com/AB\\_Files/Spatia\\_Distribution\\_of\\_Pop\\_50\\_Cities.pdf](http://alain-bertraud.com/AB_Files/Spatia_Distribution_of_Pop_50_Cities.pdf).
- Blough, D. R. 2003. “Integrating GIS into the Survey Research Process.” *New Directions for Institutional Research* 120: 37–52.
- Brewer, C. and Suchan, T. A. 2001. *Mapping Census 2000: The Geography of U.S. Diversity*. Redlands, CA: ESRI Press.
- Brundson, C., and Fotheringham, A. S. 2002. “Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity.” *Geographical Analysis* 28(4): 281–298.
- Burnham, K. P. and Anderson, D. R. 2002. *Model Selection and Multimodel Inference: A Practical Information–Theoretic Approach* (2nd ed.). New York: Springer.
- Carlton, M. and Fotheringham, A. S. 2008. “Geographically Weighted Regression: A Tutorial on Using GWR in ArcGIS9.3.” [http://ncg.nuim.ie/ncg/GWR/GWR\\_Tutorial.pdf](http://ncg.nuim.ie/ncg/GWR/GWR_Tutorial.pdf).
- Carlton, M. and Fotheringham, A. S. 2009. “Geographically Weighted Regression: White Paper.” [http://ncg.nuim.ie/ncg/GWR/GWR\\_WhitePaper.pdf](http://ncg.nuim.ie/ncg/GWR/GWR_WhitePaper.pdf).
- Castells, M. 1972. *The Urban Question. A Marxist Approach* (Original French Edition, Translated by Alan Sheridan, 1977). MIT Press.
- DeMers, M. N. 2004. *Fundamentals of Geographic Information Systems* (3rd ed.). John Wiley and Sons.
- Fotheringham, A. S., Brundson, C. and Charlton M. E. 2002. *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*. West Sussex, John Wiley and Sons Ltd.
- Fox, J. 1991. Regression Diagnostics, *Sage University Paper Series on Quantitative Applications in the Social Sciences* 07–079. Sage Publications.
- Griffith, D. A. 2008. “Spatial–Filtering–Based Contribution to a Critique of Geographically Weighted Regression(GWR).” *Environment and Planning* 40: 2751–2769.
- Harvey, D. 1989. *The Condition of Postmodernity: An Enquiry into the Origins of Cultural Change*. Wiley–Blackwell.
- Hurvich, C. M., Simonoff, J. S., and Tasi, C. L. 1998. “Smoothing Parameter Selection in Nonparametric Regression Using an Improved Akaike Information Criterion.” *Journal of Royal Statistical Society* (Series B) 60: 271–293.
- Lefèbvre, H. 1992. *The Production of Space*. Wiley–Blacwell.
- LeGates, R. 2005. *Think Globally, Act Regionally: GIS and Data Visualization for Social Science and Policy Research*. ESRI Press.
- Lo, C. P. and Yeung, A. K. W. 2002. *Concepts and Techniques in Geographic Information Systems*. Prentice Hall.

- Löw, M. 2008. "The Constitution of Space: The Structuration of Spaces through the Simultaneity of Effects and Perception." *European Journal of Social Theory* 1(11): 25-49.
- Peters, A. and MacDonald, H. 2004. *Unlocking the Census with GIS*. Redlands, CA: ESRI Press.
- Tobler, W. R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46(2): 234-42.
- Tufte, E. R. 2001. *The Visual Display of Quantitative Information* (2nd ed.). Cheshire, CN: Graphics Press.

[접수 2009/10/12, 수정 2009/11/6, 게재확정 2009/11/9]

