대한지리학회지 제49권 제5호 2014(761~778)

상업용 토지 가격의 베이지안 추정: 주관적 사전지식과 크리깅 기법의 활용을 중심으로

이창로* · 엄영섭** · 박기호***

A Bayesian Estimation of Price for Commercial Property: Using subjective priors and a kriging technique

Chang Ro Lee* · Young Seob Eum** · Key Ho Park***

요약 : 본 논문은 거래빈도가 낮아 지금껏 적극적으로 시도되지 못한 상업용 토지의 가격을 정확히 추정하고자 하였다. 서울시 상업용 토지 실거래가 자료를 대상으로 선형 결합 형태의 평균 구조(전역적 경향), 지수 형태의 공분산함수 그리고 순수 오차항을 구성요소로 하는 모형을 구축 및 적용하였다. 상권별로 가격수준이 차별적으로 형성되는 상업용 토지 가격의 특성을 감안하여 대표적 공간보간기법인 크리깅 방법을 적용함으로써 지가의 공간적 상관성을 명시적으로 고려하였다. 더 나아가 희소한 자료의 한계를 극복하기 위해 전문가 지식을 사전 확률분포의 형태로 모형에 반영할 수 있는 베이지안 크리깅 방법을 활용하였다. 적용한 모형의 성능은 적합과정에 사용되지 않은 검증 자료를 대상으로 검토하였으며, 전문가 지식의 반영과 공간적 상관성의 명시적 고려를 통해 가격 추정의 정확성이 높아진 사실을 확인하였다. 본 논문은 베이지안 크리깅 기법을 토지 가격 추정에 적용하되, 전문가의 주관적 지식을 명시적으로 모형에 반영하였다는 점 등에서 기존 연구와 차별성을 갖는다. 본 논문의 결과는 거래 자료가 희소한 상황에서도 신뢰성 있게 부동산 가격을 추정해야하는 경우에 유용하게 활용될 수 있을 것으로 기대된다.

주요어 : 주관적 사전지식, 공간적 자기상관성, 베이지안 크리깅, 상업용 토지, 실거래가

Abstract : There has been relatively little study to model price for commercial property because of its low transaction volume in the market. Despite of this thin market character, this paper tried to estimate prices for commercial lots as accurate as possible. We constructed a model whose components consist of mean structure(global trend), exponential covariance function and a pure error term, and applied it to actual sales price data of Seoul. We explicitly took account of spatial autocorrelation of land price by utilizing a kriging technique, a representative method of spatial interpolation, because the land price of commercial lots has feature of differential price forming pattern depending on submarkets they belong to. In addition, we chose to apply a bayesian kriging to overcome data scarcity by incorporating experts' knowledge into prior probability distribution. The chosen model's excellent performance was verified by the result from validation data. We confirmed that the excellence of the model is attributed to incorporating both

이 논문은 교육부와 한국연구재단의 BK21플러스 사업(4-Zero지향 국토공간창조 사업단, 서울대학교 지리학과)의 지원을 받아 수행된 연구결과임

* 서울대학교 지리학과 박사과정(Ph.D Candidate, Department of Geography, Seoul National University), spatialstat@naver.com

** 서울대학교 지리학과 석사과정(Student in the Master's course, Department of Geography, Seoul National University), eys257@gmail.com

*** 서울대학교 지리학과 교수 및 국토문제연구소 겸무 연구원(Professor, Department of Geography, Seoul National University, and Researcher, Institute for Korean Regional Studies), khp@snu.ac.kr

experts' knowledge and spatial autocorrelation in the model construction. This paper is differentiated from previous studies in the sense that it applied the bayesian kriging technique to estimate price for commercial lots and explicitly combined experts' knowledge with data. It is expected that the result of this paper would provide a useful guide for the circumstances under which property price has to be estimated reliably based on sparse transaction data.

Key Words : subjective priors, spatial autocorrelation, bayesian kriging, commercial lot, actual sales price

1. 서론

공정하고 투명한 부동산 거래질서 확립 및 공평과세를 목적으로 정부는 2006년부터 부동산 거래 신고제도(실거래가 신고제도)를 시행하여 오고 있다. 올해로 제도 도입 9년째를 맞이하는 동 제도는 양도소득세 등 국세의 부과기준, 아파트 실거래가 지수 및 지가변동률 산정기준 등 다양한 분야에서 중요한 기초자료로 활용되고 있다. 최근 정부는 축적된 실거래가 자료를 활용할 또 하나의 방안으로 실거래가 기반 공시제도 도입 계획을 발표한 바 있다[1].

이러한 실거래가 신고제도 도입 이후 부동산 중개업체 등에서 발표하는 호가수준이 아니라 실제 거래된 가격을 기초로 부동산 가격을 추정한 연구사례들도 증가하는 추세이다(이창무 외, 2009; 김성우 · 정건섭, 2010; 김종수, 2012 등). 이러한 선행연구 대부분은 비교적 사례가 풍부한 주택, 그 중에서 특히 공동주택(아파트, 다세대주택 등)에 집중되어 있는 편이다. 표 1에서 보듯 2013년 서울시에 신고된 거래건수 중 약 80%가 주택에 해당할 정도로 주거용 부동산은 거래가 빈번하다.

반면 상가, 오피스 빌딩 등 상업용 부동산은 거래가 상대적으로 희소한 편이다. 표 1에서 서울시 토지만의 거래는 2.9%에 불과하며, 이러한 토지 거래에는 상업용, 공업용, 농업용(농경지, 임야) 등이 모두 포함되는데 특히 상업용 토지는 규모가 크고 환가성이 낮아 거래빈도가 더욱 낮은 편이다. 서울시 강남 테헤란로에 위치한 업무용 빌딩의 매각은 언론매체에 보도될 정도로 '특별한' 사건인 것이다.

상업용 부동산은 자료의 수가 절대적으로 부족할 뿐 아니라 거래가격 자체에 상당한 잡음(noise)이 포함되어 있다. 즉 급매에 의한 거래, 매수자 및 매도자의 정보 비대칭, 부동산 중개인의 개입 정도, 탈세를 위한 저가신고의 유인 등 수많은 요인으로 인해 인접하여 위치한 유사 부동산이 동일 시점에 거래된다 하더라도 신고된 가격은 상이할 수 있다. 또한 상업용 부동산에만 존재하는 특유의 거래잡음, 예를 들어 권리금 · 영업 노하우 · 기술력 등 무형자산의 가치, 집기 · 비품 · 인테리어시설 등 동산항목의 가치, 유리한 프랜차이즈 계약 등에 따른 프리미엄 등 부동산 가치라 보기 어려운 다양한 가치가 거래가격에 포함되어 있다. 따라서 신고가격에 비부동산가치가 상당한 비중으로 포함되어 있다면 이러한 거래 역시 분석에서 제외하여야 하며, 이 경우 자료의 부족 현상은 더욱 심각해진다.

낮은 거래빈도와 함께 상업용 부동산의 또 다른 특징은 가격형성의 공간적 자기상관성(spatial autocor-

표 1. 2013년 서울시 실거래가 신고 현황*

구분	전체	토지	건물	공동주택	단독주택	기타
건수	140,866	4,049	1,803	103,874	10,142	20,998
비율(%)	100.0	2.9	1.3	73.7	7.2	14.9

* 국토교통부 감정평가정보체계「거래동향」에서 발췌 · 정리(2014.6 기준)

relation)이다. 상업용 부동산을 포함한 모든 유형의 부동산이 공간자료의 대표적 특징인 자기상관성을 가지고 있지만, 상권별로 뚜렷하게 가격수준이 형성되는 상업용 부동산의 경우 그러한 현상이 더욱 두드러진다.

공간상에서 발생하는 자기상관성은 다양한 방법으로 처리되고 있으며 헤도닉 모형을 적용한 기존의 선행연구를 보면, 속성(attributes)정보에 해당하는 설명변수 또는 공변량(covariates)을 정밀하게 구성하고, 이러한 공변량으로도 설명되지 않는 자료의 변동성은 오차항을 구성하는 단계에서 해결하려는 접근이 많다. 즉, 속성정보를 우선적으로 고려하되, 잔차에 남겨진 공간적 자기상관성은 공간가중행렬의 구성 등을 통해 해결하고자 한다(서경천 · 이성호, 2001; 서교, 2005; 김성우 · 정건섭, 2010 등).

이와 반대되는 접근으로 먼저 거래사례의 위치(location)정보를 명시적으로 고려하되, 활용할 수 있는 속성정보가 있는 경우 부차적으로 모형에 포함시키는 방법이 있다. 예를 들어 기온, 강수량 등을 추정할 때 흔히 사용되는 공간보간(spatial interpolation)기법이 이러한 접근에 해당된다. 이 방법은 '3L 접근법'이라는 표현에서 잘 나타나듯(Kiel & Zabel, 2008) 부동산 가격추정에 있어 가장 중요한 요인이 위치라는 견해에 보다 충실한 방법이다. 또한 이러한 접근은 부동산 가격을 설명하는 공변량을 모두 측정하여 모형에 포함시킬 수는 없으며, 따라서 모형에서의 누락변수(omitted variable) 한계를 인정한 방법으로 보다 현실적이다(Wheeler *et al.*, 2013).

본 연구에서는 거래가 드문 서울시의 상업용 실거래가 자료를 대상으로 헤도닉 모형을 구성하여 희소한 자료에 기초한 모형 구축의 가능성을 검토하였다. 자료의 희소성은 전문가의 지식을 보충적으로 활용하여 극복하고자 하였고, 부동산 가격이 가지는 공간적 자기상관성은 공간보간기법의 일종인 크리깅(kriging)기법을 활용하여 모형에 반영하고자 하였다. 공간보간기법에서도 활용가능한 공변량이 있는 경우 이를 배제하지 않고 모형에 포함시킬 수 있는 유연성이 있음은 앞서 언급한 바 있다.

2. 이론적 검토 및 선행연구 고찰

1) 베이지안 추론(Bayesian inference)

본 연구는 희소한 자료의 한계를 극복하고자 베이지안 접근법(Bayesian approach)을 활용하였다. 기존의 고전적 접근법(classical or frequentist approach)은 모형에 투입된 '데이터만큼' 좋은 결과가 나오게 되어 있으며 투입 데이터가 미흡하면 모형을 통해 산출된 결과 역시 미흡할 수밖에 없다.

데이터의 질이 미흡하거나 양이 충분하지 않은 경우 이러한 부족 부분을 연구자가 보유한 경험이나 지식, 또는 과거에 수행된 선행연구 결과로 보완하려는 방법이 바로 베이지안 접근법이다.

고전적 접근법과 베이지안 접근법의 오랜 논쟁의 핵심은 '사전 확률분포(prior probability distribution)의 주관성'에 있는데, 이러한 주관성에 대해 수많은 비판이 제기되었으나(Dennis, 1996), 베이지안 접근법은 이러한 주관적 지식을 아예 존재하지 않는 것으로 간주하거나 '정성적 판단'이라 하여 배척하기보다는 그러한 주관적 지식이 어떻게 모형에 반영되었는지 적어도 투명하게 설명할 수 있다는 점에서(McCarthy, 2007, p.225) 보다 장점이 많은 접근법이라 할 수 있다.

베이지안 접근은 베이즈 법칙(Bayes' rule)에 기반한 것으로 $\boldsymbol{Y}$를 관찰값 벡터(vector of observed values), θ를 추정할 모수 벡터(vector of parameters)라 하면 다음과 같이 베이즈 법칙을 표현할 수 있다.

$$p(\theta|\boldsymbol{Y})=\frac{p(Y|\theta)p(\theta)}{p(Y)} \quad (1)$$

베이지안 통계와 기존 통계의 가장 큰 차이는 θ의 추정방식에 있다. 기존의 통계적 접근에서 θ는 확률변수(random variable)가 아니며 고정된 값을 갖는다(다만 그 값을 사전에 알지 못할 뿐이다). 따라서 기존의 통계적 접근에서 θ에 대해 확률분포를 논하는 것은 의미가 없다. 반면 베이지안 통계에서 θ는 확률변수이며 그 분포는 θ에 대한 불확실성을 표현한다. 식

(1)에서 $p(\theta)$는 사전 확률분포라 하며 데이터를 관찰하기 전 θ에 대해 연구자가 가지는 사전 지식(또는 신념)이라 할 수 있다. $p(\boldsymbol{Y}|\theta)$는 우도(likelihood)라 하며 최우추정법(Maximum Likelihood Method)에서의 우도와 동일한 개념이다. $p(\boldsymbol{Y})$는 주변확률(marginal probability)로서 원칙적으로 다음과 같은 적분을 통하여 계산할 수 있다.

$$p(Y)=\int p(Y|\theta)p(\theta)d\theta \quad (2)$$

마지막으로 $p(\theta|Y)$는 사후 확률분포(posterior probability distribution)로서 데이터를 관찰한 후 θ에 대해 업데이트된 연구자의 현재 지식(또는 신념)을 의미한다. 이러한 사후 확률분포의 대표값(평균)은 θ에 대한 최적의 추정치로 해석되며, θ의 분산은 그러한 추정치에 대한 불확실성을 나타낸다(Plant, 2012, p.450).

2) 전문가 지식의 추출

베이지안 모형에 투입되는 사전 확률분포는 크게 모호 사전분포(vague prior distribution)와 주관적 사전분포(subjective prior distribution)로 나눌 수 있다[2]. 모호 사전분포는 추정하고자 하는 모수에 정보를 거의 투입하지 않는 반면, 주관적 사전분포는 이전에 획득한 정보를 모수 추정에 반영하고자 한다. 따라서 전자를 통한 분석은 고전적 접근법으로 추정한 결과와 유사하며 '주관성' 논쟁으로부터 비교적 자유롭다. 반면 후자를 통한 분석은 주관성 비판의 대상이 되기도 하지만, 축적된 사전지식을 활용할 수 있는 장점이 있다. 본 연구에서는 주관적 사전분포(지식)의 활용에 초점을 맞춘다.

사전지식은 실험설계(experimental design), 선행연구 검토와 같은 메타분석(meta-analysis), 그리고 관련 전문가로부터의 추출 등 여러 가지 방법을 통해 얻을 수 있다. Choy *et al.*(2009)은 전문가로부터 사전지식을 추출하는 절차에 대하여 비교적 자세하게 정리하였는데 그가 제시한 절차는 표 2와 같다.

표 2의 6개 절차 중 가장 중요하며 실무적으로 수행하기 까다로운 단계는 4번째로서, 사전지식을 측정하는 방법은 직접 측정과 간접 측정으로 나눌 수 있다(Choy *et al.*, 2009). 직접 측정은 전문가에게 모형의 모수값에 대해 직접 질문하는 형태로, 해당 전문가는 질문 내용에 대한 이해 뿐 아니라 특정 모형의 형태에 대해서도 통계적 지식을 가지고 있어야 한다. 반면, 간접 측정은 전문가가 관찰한 현상들(observations)을 질문하는 것에 그친다. 예를 들어 회귀모형에서 공변량 값을 제시하고, 그에 따른 종속변수 값을 예측하게 하는 것이다. 직접 측정은 연구자가 설정한 모형이나 가설에 대해 해당 전문가가 통계적 지식을 함께 가지고 있어야 하므로 실무적으로 수행하기 어려운 반면, 간접 측정은 그러한 제한이 없어 전문가로부터 보다 '자연스럽게' 지식을 추출할 수 있다(Lele & Das, 2000). 표 3은 전문가 지식 추출과정을 명시적으로 설명한 최근의 연구사례를 직접 측정과 간접 측정의 범주로 나누어 제시한 것이다.

표 2. 전문가로부터의 사전지식 추출 절차*

단계	내용
①	사전지식 활용의 목적 및 동기의 설정
②	전문가로부터 추출 가능한 사전지식의 종류 결정
③	우도함수, 사전분포 등 모형의 형태 설정
④	사전지식의 측정방법 결정
⑤	측정된 지식의 검증절차 구성
⑥	전문가 선정 및 사전지식 추출

* Choy *et al.*, 2009, "Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical model"에서 발췌 및 정리

표 3을 보면 직접측정 방법을 적용한 경우 전문가 대답의 편의(bias)를 최소화하기 위해 자신이 대답한 결과에 대해 반복적으로 조정할 수 있는 절차를 두고 있음을 알 수 있다. 간접측정의 경우 대부분 일정한 조건(공변량)을 제시하고, 전문가가 생각하는 종속변수 값을 수집하는 것을 알 수 있는데, 이 경우에도 사후적으로 전문가 간 대답의 일치성 정도를 카파 통계량(Kappa statistic) 등으로 검토하는 것이 일반적이다(Johnson *et al.*, 2010).

본 연구에서는 실무적으로 수월한 간접 측정을 통

표 3. 전문가 지식의 직접측정 및 간접측정 사례

구분	저자	내용
직접 측정	Truong *et al.* (2013)	geostatistics 모형의 핵심요소인 베리오그램(variogram) 모수 값을 직접 질문하되, 자신이 대답한 베리오그램 형태에 대해 시각적으로 확인 및 수정할 수 있는 기능 제공(Web 기반으로 실행)
	Jones & Johnson (2014)	모형의 모수 값(예를 들어 Poisson Rate λ)을 직접 질문하되, 자신이 대답한 모수 값의 분포형태를 시각적으로 확인할 수 있는 그래픽 도구를 함께 제공
간접 측정	Gill & Walker (2005)	법조인, 보수 정치인 및 진보 정치인을 대상으로 니카라과(Nicaragua) 사법체계의 공정성 여부를 이진변수 형태로(예, 아니오) 질의
	Martin *et al.* (2005)	조류 전문가를 대상으로 방목강도(grazing density)에 따른 조류의 서식 가능성을 3가지 답변 형태(증가, 감소, 불변)로 질의
	Johnson *et al.* (2010)	약물 투여 여부에 따른 환자의 향후 3년 간 생존 가능성을 확률 형태로 질의

해 전문가의 지식을 수집하였다.

3) 공간적 자기상관성의 고려

공간자료는 크게 점 자료(point data)와 면 자료(areal data)로 나눌 수 있으며, 본 연구에서는 개별 필지의 가격을 대상으로 하므로 점 자료의 추정 및 예측 방법에 초점을 맞춘다.

점(point)으로 표현되는 공간자료를 모델링하는 가장 일반적인 형태는 다음과 같다(Gelfand, 2012).

$$y(s)=\mu(s)+w(s)+\varepsilon(s) \quad (3)$$

위 식에서 $y(s)$는 지리적 위치 s에서의 종속변수를 의미하며, $\mu(s)$는 이러한 종속변수의 평균 구조(mean structure) 또는 전역적 경향(global trend)을 의미한다. 평균 구조는 공변량과 모수와의 선형관계, 즉 $\mu(s)=\boldsymbol{x}^T(s)\beta$로 표현하는 것이 일반적이다(Banerjee *et al.*, 2004). 공변량은 속성정보(부동산의 경우 면적, 도로조건 등)로 이루어지는 것이 통상이다.

종속변수에 영향을 미치는 체계적 요인들(systematic components)을 공변량을 통해 통제하였다면 나머지 변동성은 오차항으로 반영되며, 비공간자료의 경우 독립적 오차항(백색잡음, white noise) $\varepsilon(s)$로 표현한다.

그러나 공간자료의 가장 큰 특징은 공간적 자기상관성으로, 가까운 곳에 위치한 관찰치들은 서로 유사한 값을 갖기 마련이다(Tobler, 1970). 따라서 식 (3)에서 오차항은 두 부분으로 구분할 수 있는데, $w(s)$는 관찰치들 간의 공간적 자기상관성을 나타내는 임의효과항(spatial random effect term)이며, $\varepsilon(s)$는 비공간적인(non-spatial) 순수 오차항(pure error term)을 나타낸다.

점(point) 자료를 다루는 공간 모델링 분야(문헌에서는 'geostatistics'라 한다)에서 이러한 자료의 공간적 상관성은 크리깅(kriging) 기법을 이용하여 모형에 반영할 수 있다. 크리깅은 공간보간을 위한 대표적인 기법으로, 관찰되지 않은 지점의 예측값을 주변 관찰지점 값의 가중선형조합으로 산출하는 방법이다(Isaaks & Srivastava, 1989). 크리깅은 환경과학 분야에서 가장 일반적으로 사용되는 방법임에도(Webster & Oliver, 2007) 불구하고, 부동산 가격추정 등 사회과학 분야에서는 잘 시도되지 않고 있다.

그러나 해외의 경우 부동산 가격추정을 위한 연구에서 자료의 공간적 상관성을 크리깅 기법으로 모형에 반영하고자 하는 노력을 찾아 볼 수 있다(Militino *et al.*, 2004; Chica-Olmo, 2007; Montero & Larraz, 2011; Kuntz & Helbich, 2014). 부동산 가격 추정을 위한 헤도닉 모형(hedonic model)에서 크리깅 기법을 적용한 상기 연구들은 모두 일종의 보편 크리깅

(universal kriging)을 적용한 예에 해당한다. 보편 크리깅은 종속변수인 부동산 가격에 대해 직접 공간보간을 하는 것이 아니라, 부동산 가격을 예측할 수 있는 체계적 요인들(평균 구조 또는 전역적 경향)을 통제한 후에 잔차, 즉 설명되지 않는 변이에 대해 크리깅 기법을 적용하는 방법이다[3]. 반면 자료에 전역적 경향이 없는 경우 단순 크리깅(simple kriging)이나 정규 크리깅(ordinary kriging)을 적용할 수 있다.

크리깅의 기본적 가정이 자료가 공간적으로 불변성(stationarity)을 유지해야 한다는 것, 즉 자료의 전역적 경향이 없어야 하는 것이므로 상기 연구들의 보편 크리깅 적용은 이론에 합당한 방법이라 할 수 있다. 본 연구에서도 이러한 접근을 따라 자료의 전역적 경향을 제거한 후에 남아 있는 잔차에 대해 크리깅 기법을 적용, 부동산 가격의 공간적 상관성을 모형에 반영하였다.

이러한 보편 크리깅은 일반화최소제곱법(Generalized Least Squares, GLS), 최우추정법(MLE)과 같은 고전적 접근법으로도 추정이 가능하지만, 베이지안 접근법(Bayesian kriging)으로도 역시 추정이 가능하다. 베이지안 크리깅은 비교적 새로운 접근법으로, 역학(epidemiology), 보건지리(medical geography) 분야에서 질병의 발생 패턴을 분석(Lai *et al.*, 2013; Slater & Michael, 2013; Scholte *et al.*, 2014)하는 등 자연 · 환경 분야에서 최근 들어 활발하게 적용되고 있다.

최우추정법(MLE) 및 베이지안 접근법을 통한 크리깅 추정치의 정확성을 비교한 연구가 있기는 하지만(Ghosh & Carriazo-Osorio, 2007 등), 본 연구는 앞서 설명한 전문가의 지식을 모형에 명시적으로 반영하는 것이 주된 목적이므로, 베이지안 크리깅을 적용하였다.

즉, 전문가의 지식은 자료의 전역적 경향을 설명하는데 사용하고, 자료의 나머지 변동성(자기상관성)은 크리깅을 통해 모형에 반영하였다.

3. 모형의 구성 및 적합

1) 베이지안 공간모형의 구성

식(3) $y(s)=\mu(s)+w(s)+\varepsilon(s)$에서 자료의 평균 구조 또는 전역적 경향 $\mu(s)$는 실거래가 자료에서 확보할 수 있는 속성정보(용도지역, 도로조건 등)를 중심으로 구성하였다. 자기상관성 즉, 공간효과를 나타내는 $w(s)$는 '안정적' 공분산 함수(stationary covariance function)에 기초한 임의효과항이며, $\varepsilon(s)$는 $N(0,\tau^2)$을 따르는 독립적인 오차항이다.

식(3) $y(s)$의 주변 공분산 행렬(marginal covariance matrix)은 다음의 형태로 구성할 수 있다(Banerjee *et al.*, 2004).

$$\Sigma=\sigma^2R(\phi)+\tau^2I \tag{4}$$

위 식에서 σ^2는 공간효과와 관련된 분산(spatial variance 또는 partial sill), R은 공간적 상관성을 표현하는 일종의 상관행렬(correlation matrix), τ^2는 $\varepsilon(s)$의 분산(non-spatial variance 또는 nugget)을 각각 의미한다[4].

상관행렬 R은 다음과 같이 표현할 수 있다.

$$R_{ij}=\rho(s_i-s_j;\phi) \tag{5}$$

즉, 관찰치들 간의 거리(s_i-s_j), 상관함수 ρ, 그리고 일종의 거리조락 계수(decay parameter) ϕ로 구성할 수 있다.

따라서 추정해야할 모수들을 θ라 한다면 $\theta=(\beta,\sigma^2,\tau^2,\phi)^T$로 나타낼 수 있고, 사전 확률분포 $p(\theta)$를 적절하게 구성할 경우, 사후 확률분포는 베이즈 법칙에 의해 식(1)과 같이 산출할 수 있다. 이때 우도(likelihood) $p(\boldsymbol{Y}|\theta)$는 다음과 같은 형태로 표현할 수 있다.

$$\boldsymbol{Y}|\theta\sim N(\boldsymbol{X}\beta,\sigma^2R(\phi)+\tau^2I) \tag{6}$$

사전 확률분포는 다음과 같은 독립적인 사전 확률

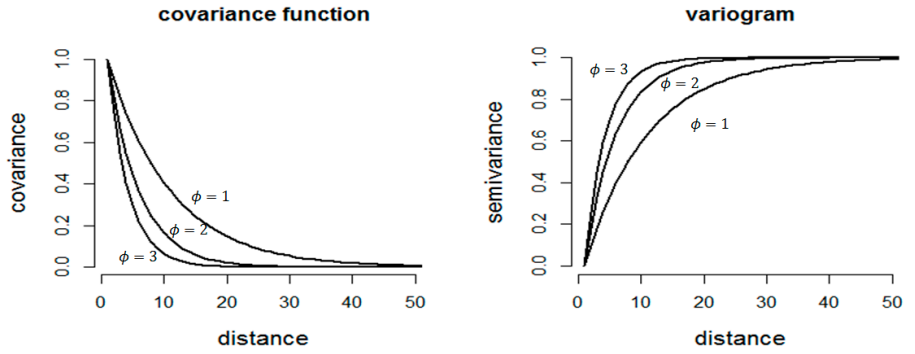

그림 1. 지수 공분산함수 및 베리오그램의 형태

분포 구성이 가능하다(Banerjee *et al.*, 2004)[5].

$$p(\theta)=p(\beta)p(\sigma^2)p(\tau^2)p(\phi) \quad (7)$$

이러한 형태의 모형 구성은 정규 우도(normal likelihood)뿐 아니라 이진변수(binary variable)와 같은 비정규 우도(non-normal likelihood)에도 쉽게 확대하여 적용할 수 있는 유연성을 가지고 있다[6].

마지막으로 상기 모형의 중요한 구성요소인 $w(s)$의 세부적인 구성 절차는 다음과 같다. 부동산 가격과 같은 공간자료가 통계적 모형에서 일반적으로 전제하는 독립성(independence) 가정에 잘 들어맞지 않는다는 것은 널리 인지된 사실이다. 따라서 인근에 위치한 관찰치들은 서로 유사한 값을 갖기 마련이며, 이러한 양의 상관성(positive correlation)은 거리가 멀어짐에 따라 약화된다. 공간자료의 이러한 특징은 공분산함수를 적정하게 구성함으로써 모형에 반영할 수 있다. 본 연구에서는 자료의 공간적 상관성이 자료들 간의 거리에 의해서만 결정된다는 가정(등방성, isotropy) 하에 지수 공분산함수(exponential covariance function)를 활용하여 $w(s)$를 구성하였다[7].

지수 공분산함수는 식(8)의 형태로 표현되며 크리깅 기법에서는 식(9)의 형태(베리오그램, variogram)로 변형하여 사용하는 것이 일반적이다[8].

$$C(t)=\sigma^2\exp(-\phi t),\ t>0 \quad (8)$$

$$\gamma(t)=\tau^2+\sigma^2(1-\exp(-\phi t)),\ t>0 \quad (9)$$

위 식에서 t는 자료 간의 거리를 나타내며, σ^2, τ^2, ϕ는 식(4) 및 식(5)에서의 의미와 동일하다[9]. 공간적 상관성이 점차 약화되어 무시할 수 있는 수준에 이르는 거리(range)를 t_0라 할 경우, $t_0 \approx 3/\phi$로 해석할 수 있다[10].

그림 1은 ϕ값의 변화에 따른(ϕ=1,2,3) 지수 공분산함수 및 이에 대응되는 베리오그램을 보여준다. 공분산함수에서는 거리가 멀어짐에 따라 자료 간 상관도가 떨어지는 것으로, 베리오그램에서는 거리가 멀어짐에 따라 자료 간 상이성(dissimilarity)이 증가하는 것으로 해석할 수 있다. 본 연구에서는 이러한 지수 공분산함수(베리오그램)를 적용하여 공간적 상관성을 모형에 반영하였다.

식(3)~(9)를 통해 모형의 형태를 결정하였으므로 다음 절에서는 모형의 평균 구조에 포함시킬 전문가 지식의 추출과정에 대해 논의한다.

2) 전문가 지식의 추출

본 연구에서는 부동산 가격평가 전문가(감정평가사)가 보유한 지식을 추출하여 모형에 반영하였다. 이때 앞서 구성한 모형의 형태나 모수 값을 직접 질의하기보다는(직접 측정), 전문가가 생각하는 지역별 부동산 가격수준을 관찰하는(간접 측정) 절차를 통해 지식을 추출하였다.

다만 본 연구에서는 별도의 설문지 등을 구성하여 전문가에게 발송하는 절차를 거치지 않고 전문가의

생각이 담겨 있다고 판단되는 기존의 공개된 자료, 즉 서울시 소재 상업용 표준지 공시지가 자료를 활용하였다. 표준지 공시지가는 정부에 의해 매년 안정적으로 관리되는 공식 자료로서 본 연구에서 사용한 상업용 실거래가 자료(123개)보다 그 양이 월등하게 많아 감정평가사가 생각하는 지역별 부동산 가격수준을 추출하는데 적합하다.

표 4는 분석에 사용된 상업용 실거래가 자료와 표준지 자료를 비교한 것인데, 표준지 자료에 기초하여 전문가가 생각하는 부동산 가격수준을 추출하고, 이러한 결과를 주관적 사전 확률분포의 형태로 실거래가에 기초한 본 모형에 반영하였다.

표 4. 상업용 실거래가 및 표준지 자료

구분	실거래가	표준지
개수	123개	13,968개
거래(평가)시점	2013.1.2~12.8	2013.1.1
가격범위(만원/㎡)	147~9,337	50~6,500

3) 모형의 적합: MCMC 시뮬레이션

베이지안 접근법을 실무에 비교적 용이하게 적용할 수 있게 된 것은 MCMC(Markov Chain Monte Carlo) 방법의 개발에 힘입은 바 크다. MCMC는 모수 θ를 사전 확률분포 $p(\theta)$로부터 추출하고, 이를 반복적으로 수정하여 종국에는 의도한 사후 확률분포 $p(\theta|\boldsymbol{Y})$로 근사화시키는 일반적 방법이다(Gelman *et al*. 2004). **MCMC**는 사전 확률분포 및 우도함수가 복잡한 형태를 갖게 되어 분석적 해(**analytic solution**)의 도출이 불가능한 경우에도 모형의 계수를 추정할 수 있는 매우 유연한 분석도구이다.

MCMC는 일종의 시뮬레이션 도구로 일련의 표본을 추출하여 체인(**chain**)을 구성하되, 시뮬레이션 회수가 매우 커질 경우 이러한 체인은 안정적 분포(**stationary distribution**)로 수렴하게 된다. 일단 수렴상태에 이르렀다면 이후 추출되는 표본은 그러한 안정적 분포, 즉 연구자가 의도한 사후 확률분포로부터 도출되었다고 해석한다.

MCMC 시뮬레이션 기법 중 가장 폭넓게 활용되는 방법은 **Gibbs** 샘플링(**Gibbs sampling**)이며 이 방법은 다차원 구조(**multi-dimensional structure**)를 갖는 복잡한 데이터를 한 번에 해결하여 해(**solution**)를 찾는 대신, 이러한 데이터를 저차원 구조(**low-dimensional structure**)로 분해하여 각 부분에 대해 하나씩 순차적으로 해를 찾는 방법이다(**Gelman *et al*., 2004; Geman & Geman, 1984**). 본 연구도 **Gibbs** 샘플링에 기반하여 분석을 수행하였다.

4. 사례 분석 및 결과의 해석

1) OLS 모형: 주요 공변량의 선별

본 연구에서는 2013년 1년 동안 거래된 서울시 상업용 부동산 실거래 자료 123개를 분석대상으로 하였으며, 이 중 각 자치구마다 한 개씩의 사례를 검증자료(validation dataset)로 유보하여(23개) 실제 모형 적합에 사용된 사례는 100개에 해당한다[11]. 그림 2는 이러한 자료의 공간적 분포를 보여준다.

평균 구조를 설명하는데 도움이 될 것으로 추정되는 공변량(속성정보)은 실거래가 자료에 함께 포함된 공시지가 공개항목(용도지역, 용도지구, 경사도, 필지 형상, 도로조건, 유해시설(철로)과의 거리 등 6개 항목)을 중심으로 살펴보았다. OLS 모형의 반복적 적합을 통해 선별된 공변량은 용도지역과 도로조건이며 나머지 항목은 유의성이 없어 모형에서 제외하였다.

본 연구의 지역 범위가 서울시 전체를 대상으로 하고 있고, 상업용 부동산의 가격수준이 자치구별로 편차가 큰 점을 감안하여 용도지역 및 도로조건 외에 지역 더미[자치구별 더미(dummy)변수]를 모형에 추가하였다. 지역 더미는 대표적인 맥락변수(contextual variable)로서 인근지역의 사회·경제적 수준을 나타내는 중요한 변수로 최근 재조명받고 있다(Lawson, 2009, p.164). 예를 들어 교육의 질을 나타내는 학군, 인근지역의 질을 나타내는 사회계층·특정 민족의 구성비 등은 해외의 여러 헤도닉 모형에서 매우 유의하

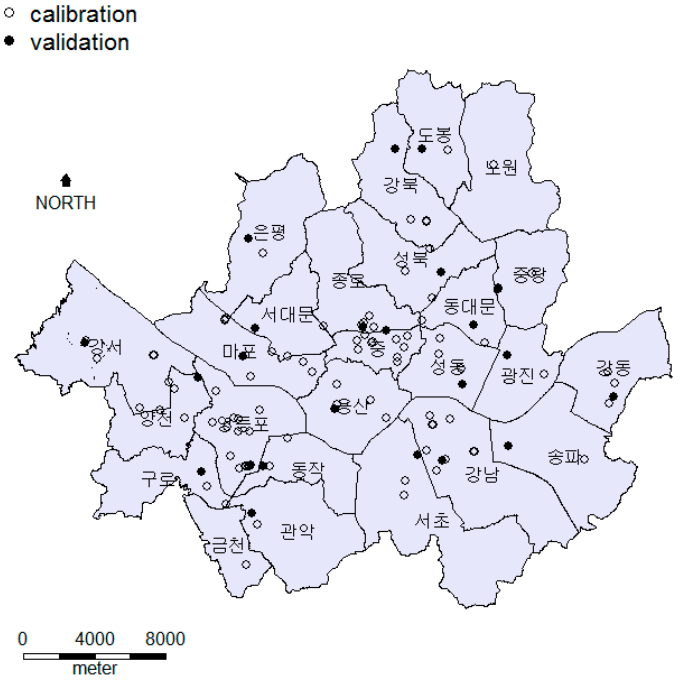

그림 2. 자료의 공간적 분포(적합자료 100개, 검증자료 23개)

게 나타나는 것을 볼 수 있는데, 이 경우 지역 더미를 포함시키면 유의성이 없어지거나 계수의 크기 자체가 작아지는(**0**에 가까워지는) 현상을 흔히 볼 수 있다(**Baranzini *et al.*, 2008, p.227**). 이는 지역 더미 자체에 인근지역의 질을 나타내는 다양한 요소가 포함되어 있음을 의미하며, 이러한 지역 더미를 효율적으로 활용할 경우 지역의 특징을 나타내는 여러 가지 요인들(고용 중심점까지의 거리, 강·호수와의 접근성, 대기오염 수준, 행정구역에 따른 세제 차이, 범죄 발생 빈도 등)을 모형에 효과적으로 반영할 수 있다(**Baranzini *et al.*, 2008, p.139; Clapp *et al.*, 2008**).

헤도닉 모형을 적용한 연구에서 누락변수 문제는 비일비재하며(**Brunauer *et al.*, 2013**) 본 연구는 모든 공변량의 빠짐없는 포착에 주안점을 두지는 않았다. 대신 구득가능한 공변량(용도지역, 도로조건, 지역더미)은 분석에 포함시키되, 이러한 공변량으로 설명되지 않는 부분은 전문가 지식과 자료의 공간적 상관성을 보다 정교하게 고려하여 해결하는데 초점을 두었다.

표 5는 상기 3개의 공변량으로 구성한 OLS 모형 결과를 보여 주며, 용도지역의 경우 기준범주인 주거지역 대비 상업지역이 (+) 부호를, 도로조건의 경우에도 기준범주인 세로 대비 광로나 중로가 (+) 부호를 보이는 등 모형의 전반적인 적합 결과는 일반적인 기대와 일치한다.그러나 표 5의 지역더미를 보면 기준범주인 영등포구[12] 대비 4개 구(강남구+, 강서구−, 중구+, 중랑구−)만 부호의 방향 및 통계적 유의성 측면에서 의미 있게 산출되었고, 나머지 구는 그렇지 않은 것으로 보인다. 이는 자료의 수가 적어 25개에 이르는 자치구별 가격수준의 차별성을 데이터가 구분해낼 수 없음을 의미한다.

2) 전문가 지식을 반영한 모형

인근지역의 특징을 나타내는 맥락변수, 즉 25개 자치구의 지역 더미계수를 보다 정교하게 구성하기 위해 표 4의 표준지 자료(서울시 상업용 표준지 13,968필지)를 기초로 OLS 모형을 적용한 결과는 표 6과 같

다. 표 6의 지역 더미 계수는 전문가가 생각하는 자치구별 상업용 토지의 가격수준으로 해석할 수 있으며, 통계적 측면에서도 대부분 유의하다.

실거래가 자료를 기초로 산출된 자치구별 지가수준(표 5)과 전문가가 생각하는 지가수준(표 6)을 비교하면 그림 3과 같다. 비교의 편의를 위해 지역 더미 계수를 지수(**exponent**)화하였으며, 따라서 영등포구 상업용 토지의 가격수준을 **1.00**으로 보았을 때 강남구의 가격수준은 실거래가 자료에 기초한 경우 **2.85**배[표 **5**의 강남구 **1.05**를 지수화, **exp(1.05)**] 높은 것으로 나타났다. 반면 전문가는 이보다 낮은 약 **2.45**배[표 **6**의 강남구 **0.90**을 지수화, **exp(0.90)**] 정도가 적정한 가격 격차로 판단하고 있다.

또한 강동구나 광진구의 경우 전문가는 영등포구보다 가격수준이 높은 것으로 생각하고 있으나, 실거래가 자료에 기초한 경우 영등포구보다 낮은 것으로 산출되었다. 상기와 같이 상반되는 결과가 발생할 경우, 자료의 수가 보다 많은 표준지 자료에 근거하여 추론하는 것이 신뢰성이 더욱 높다고 할 수 있다.

따라서 본 연구에서는 표 6의 결과를 반영하여 지역 더미에 대한 사전 확률분포를 정하였다. 예를 들어 지역 더미 중 강남구의 경우 식(10)과 같이 사전 확률분포를 지정하였고, 다른 지역 더미의 경우에도 이와 동일한 방식으로 사전 확률분포를 처리하였다.

$$\beta_{\text{강남구}} \sim N(0.90, 0.02^2) \tag{10}$$

반면 지역 더미와 같은 맥락 변수로 보기 어려운 변수에 대해서는 식(11)과 같은 모호 사전분포를 적용하였다[13].

$$\beta \sim N(0.00, 100^2) \tag{11}$$

표 7은 상기와 같이 지정된 사전분포를 기초로 MCMC 방법을 적용한 시뮬레이션 결과를 보여준다. 사후 확률분포의 수렴 여부는 Rhat(Gelman-Rubin

표 5. OLS 모형 적합 결과(실거래가 기준)*

공변량		회귀계수	표준오차	t-value	공변량	회귀계수	표준오차	t-value
상수항		**15.18**	**0.20**	**76.1**	노원구	−0.26	0.54	−0.5
용도지역	개발제한	−0.94	0.59	−1.6	도봉구	−0.62	0.55	−1.1
	공업지역	0.09	0.27	0.3	동대문구	−0.28	0.41	−0.7
	상업지역	**0.85**	**0.15**	**5.9**	동작구	−0.28	0.34	−0.8
	준주거지역	0.10	0.22	0.5	마포구	−0.20	0.26	−0.8
도로조건	**광로**	**0.49**	**0.16**	**3.1**	서대문구	0.48	0.39	1.2
	중로	**0.71**	**0.18**	**3.9**	서초구	0.62	0.41	1.5
	소로	0.25	0.19	1.3	성동구	−0.03	0.31	−0.1
	맹지	−0.05	0.56	−0.1	성북구	0.18	0.34	0.5
강남구		**1.05**	**0.23**	**4.6**	송파구	0.24	0.55	0.4
강동구		−0.20	0.34	−0.6	양천구	0.07	0.31	0.2
강북구		0.09	0.34	0.3	용산구	0.44	0.27	1.6
강서구		**−0.59**	**0.29**	**−2.0**	은평구	−0.10	0.55	−0.2
관악구		−0.01	0.55	0.0	종로구	0.19	0.31	0.6
광진구		−0.53	0.55	−1.0	**중구**	**0.55**	**0.22**	**2.5**
구로구		0.01	0.40	0.0	**중랑구**	**−0.82**	**0.41**	**−2.0**
금천구		−0.09	0.55	−0.2	Adj. R^2 = 0.64			

* (기준범주) 용도지역:주거지역, 도로조건:세로(맹지<세로<소로<중로<광로 순으로 도로 폭이 넓어짐), 지역더미: 영등포구

표 6. OLS 모형 적합 결과(표준지 기준)*

공변량		회귀계수	표준오차	t-value	공변량	회귀계수	표준오차	t-value
상수항		14.72	0.02	970.8	노원구	-0.15	0.02	-6.3
용도지역	개발제한	-0.74	0.06	-11.9	도봉구	-0.25	0.02	-10.8
	공업지역	0.08	0.02	4.4	동대문구	0.02	0.02	1.1
	상업지역	0.57	0.01	70.4	동작구	0.23	0.02	11.3
	준주거지역	0.26	0.01	24.4	마포구	0.33	0.02	17.1
도로조건	광로	0.53	0.01	63.9	서대문구	0.26	0.02	12.4
	중로	0.32	0.01	37.0	서초구	0.71	0.02	36.0
	소로	0.17	0.01	20.8	성동구	0.17	0.02	8.3
	맹지	0.25	0.13	1.9	성북구	0.06	0.02	3.2
강남구		0.90	0.02	48.6	송파구	0.53	0.02	27.0
강동구		0.30	0.02	14.0	양천구	-0.05	0.02	-2.1
강북구		-0.15	0.02	-6.9	용산구	0.55	0.02	26.5
강서구		-0.06	0.02	-2.9	은평구	-0.02	0.02	-1.2
관악구		0.16	0.02	7.8	종로구	0.50	0.02	28.3
광진구		0.17	0.02	8.2	중구	0.52	0.02	30.1
구로구		0.01	0.02	0.6	중랑구	-0.14	0.02	-6.3
금천구		-0.15	0.02	-6.2	Adj. R^2 = 0.66			

* 비교의 편의를 위해 표 5에 존재하지 않는 항목(용도지역 중 녹지지역 등)은 표기 생략

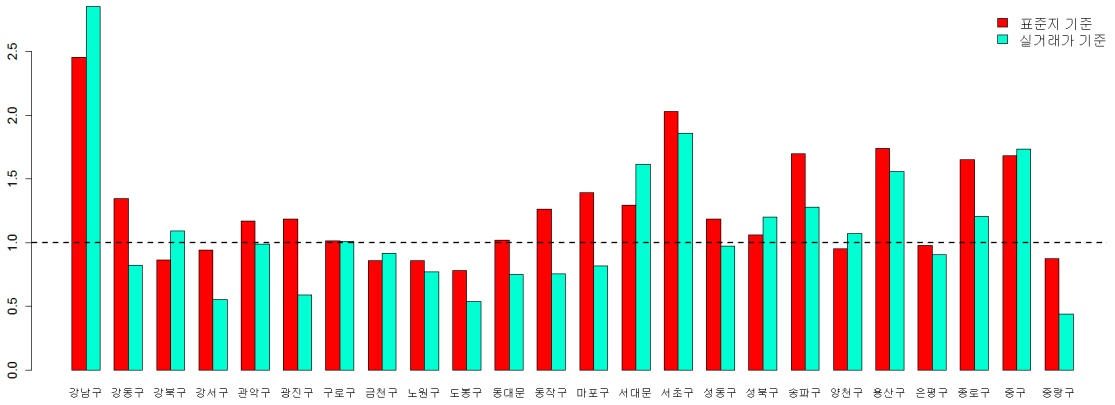

그림 3. 표준지 및 실거래가에 기초한 지역 더미 계수 비교(영등포구=1.00)

convergence statistic)과 유효 시뮬레이션 횟수(effective number of simulation draws)를 통해 확인하였다. Rhat은 체인 간 변동과 체인 내에서의 변동을 비교한 수치로 시뮬레이션 초기에는 그 수치가 매우 크지만 일단 수렴 상태에 도달하면 1.0에 가까워지며, 1.1 이하의 Rhat값을 보이면 의도한 사후 확률분포에 수렴하였다고 판단할 수 있다(Gelman & Hill, 2007, p.358). 유효 시뮬레이션 횟수 역시 최소 100개 이상이 되면 추론상의 문제는 없는 것으로 본다(Plant, 2012, p.469). 표 7을 보면 이러한 수렴 조건을 모두 충족한 것으로 보인다.

표 7에서 용도지역의 경우 상업지역>준주거지

표 7. 전문가 지식을 반영한 모형 적합 결과

공변량		대표값(평균)*	95% CI	Rhat	n.eff	공변량	대표값(평균)*	95% CI	Rhat	n.eff
상수항		15.06	14.83~15.29	1.00	2000	노원구	−0.15	−0.20~−0.11	1.00	2000
용도지역	개발제한	−0.55	−1.54~0.48	1.00	1500	도봉구	−0.25	−0.29~−0.21	1.00	2000
	공업지역	0.09	−0.35~0.49	1.00	1400	동대문구	0.02	−0.02~0.06	1.00	2000
	상업지역	0.80	0.57~1.02	1.00	2000	동작구	0.23	0.19~0.27	1.00	2000
	준주거지역	0.17	−0.18~0.51	1.00	1800	마포구	0.33	0.29~0.37	1.00	880
도로조건	광로	0.60	0.33~0.86	1.00	2000	서대문구	0.26	0.22~0.30	1.00	2000
	중로	0.61	0.30~0.94	1.00	1300	서초구	0.71	0.67~0.75	1.00	1000
	소로	0.23	−0.08~0.52	1.00	1800	성동구	0.17	0.13~0.21	1.00	2000
	맹지	−0.42	−1.48~0.61	1.00	1800	성북구	0.06	0.02~0.10	1.00	2000
강남구		0.90	0.87~0.94	1.00	900	송파구	0.53	0.49~0.57	1.00	1200
강동구		0.30	0.25~0.34	1.00	2000	양천구	−0.05	−0.09~−0.01	1.00	2000
강북구		−0.15	−0.19~−0.11	1.00	2000	용산구	0.55	0.51~0.59	1.00	2000
강서구		−0.06	−0.10~−0.02	1.00	1900	은평구	−0.02	−0.06~0.02	1.00	1100
관악구		0.16	0.12~0.20	1.00	1400	종로구	0.50	0.46~0.53	1.00	1500
광진구		0.17	0.13~0.21	1.00	740	중구	0.52	0.49~0.56	1.00	1400
구로구		0.01	−0.03~0.05	1.00	2000	중랑구	−0.13	−0.17~−0.09	1.00	1100
금천구		−0.15	−0.20~−0.10	1.00	2000	3개 체인, 체인당 4,000번 반복(첫 2,000번은 분석에서 제외)				

* 대표값(평균): 시뮬레이션을 통해 산출된 회귀계수 값들의 대표값으로 평균을 기재, 95% CI(Confidence Interval): 95% 신뢰구간, n.eff: 유효 시뮬레이션 횟수

역>공업지역>개발제한구역의 위계순서는 표 5의 OLS 모형 결과와 동일하다. 다만 도로조건의 경우 표 5의 OLS 결과에서는 중로의 계수(0.71)가 광로(0.49)보다 높았으나 표 7에서는 차이가 미미할 정도로 격차가 줄어들어(중로: 0.61, 광로: 0.60) 일반적인 직관에 보다 부합하는 결과가 산출되었다.

지역 더미의 경우 대다수 자치구에서 지정한 사전 확률분포대로 계수가 산출되었는데, 이는 자치구별 가격 수준을 구분할 만큼 충분한 정보가 실거래가 자료에 없음을 의미한다. 즉, 자치구별 실거래가 자료의 수가 매우 적어(대부분 1개 내지 6개), 지역 더미의 사후 확률분포는 데이터보다는 사전 확률분포에 큰 영향을 받아 추정되었음을 알 수 있다.

3) 전문가 지식과 공간적 상관성을 반영한 모형

전문가 지식을 반영한 위 모형에서 마지막으로 식(4)의 주변 공분산행렬, 즉 τ^2 외에 공간적 상관성을 지수 공분산함수 형태로 구성하여 최종 모형을 적합한 결과는 표 8과 같다[14].

대부분 표 7과 유사한 결과가 산출되었으며, 도로조건의 경우 광로(0.52)>중로(0.36)>소로(0.19)>맹지(−0.55)의 순서로 계수 크기가 산출되어 일반적인 직관과 일치하고 있다. 이는 공간적 상관성을 반영하지 못해 발생한 이전 모형의 오류가 치유되었음을 의미한다. 특히 식(4)의 공간 효과와 관련된 분산 σ^2이 0.163(평균값)인 반면, 순수 오차항의 분산 τ^2은 0.024(평균값)에 불과하여 공간 효과 부분이 오차항 분산의 대부분[약 87%=0.163/(0.024+0.163)]을 설명하고 있는 바, 공간적 임의효과를 모형의 구성요소로 포함시킨 것은 적절한 조치였음을 확인할 수 있다.

공간적 상관성이 점차 약화되어 무시할 수 있는 수준에 이르는 거리, 즉 레인지(range)로 해석할 수 있는 ϕ값을 보면 0.001~0.008의 범위를 보이고 있고, 평균값은 0.003이다. 지리적 거리(m)로 환산할 경우($t_0 \approx 3/\phi$), 약 375m에서 3,000m, 평균적으로 약

표 8. 전문가 지식 및 공간적 상관성을 반영한 모형 적합 결과

공변량		대표값(평균)*	95% CI	Rhat	n.eff	공변량	대표값(평균)*	95% CI	Rhat	n.eff
상수항		15.14	14.94~15.34	1.00	560	노원구	−0.15	−0.20~−0.10	1.00	2000
용도지역	개발제한	−0.59	−1.44~0.23	1.00	2000	도봉구	−0.25	−0.30~−0.21	1.00	2000
	공업지역	0.09	−0.29~0.45	1.00	660	동대문구	0.02	−0.02~0.06	1.00	1700
	상업지역	0.68	0.48~0.91	1.03	70	동작구	0.23	0.19~0.27	1.00	2000
	준주거지역	0.09	−0.19~0.36	1.01	340	마포구	0.33	0.29~0.37	1.00	2000
도로조건	광로	0.52	0.32~0.74	1.00	620	서대문구	0.26	0.22~0.30	1.00	730
	중로	0.36	0.13~0.59	1.00	1600	서초구	0.71	0.67~0.75	1.00	2000
	소로	0.19	−0.05~0.44	1.00	1100	성동구	0.17	0.13~0.21	1.00	2000
	맹지	−0.55	−1.25~0.15	1.00	930	성북구	0.06	0.02~0.10	1.00	2000
강남구		0.90	0.87~0.94	1.00	2000	송파구	0.53	0.49~0.57	1.00	1100
강동구		0.30	0.26~0.34	1.00	2000	양천구	−0.05	−0.09~−0.01	1.00	2000
강북구		−0.15	−0.19~−0.11	1.00	2000	용산구	0.55	0.51~0.59	1.00	2000
강서구		−0.06	−0.10~−0.02	1.00	2000	은평구	−0.02	−0.06~0.02	1.00	2000
관악구		0.16	0.12~0.20	1.01	340	종로구	0.50	0.47~0.53	1.00	2000
광진구		0.17	0.13~0.21	1.00	2000	중구	0.52	0.49~0.55	1.00	2000
구로구		0.01	−0.03~0.05	1.00	2000	중랑구	−0.13	−0.18~−0.09	1.00	2000
금천구		−0.15	−0.20~−0.10	1.00	2000					
τ^2		0.024	0.011~0.045	1.00	570	3개 체인, 체인당 4,000번 반복 (첫 2,000번은 분석에서 제외)				
σ^2		0.163	0.116~0.234	1.00	580					
ϕ		0.003	0.001~0.008	1.00	2000					

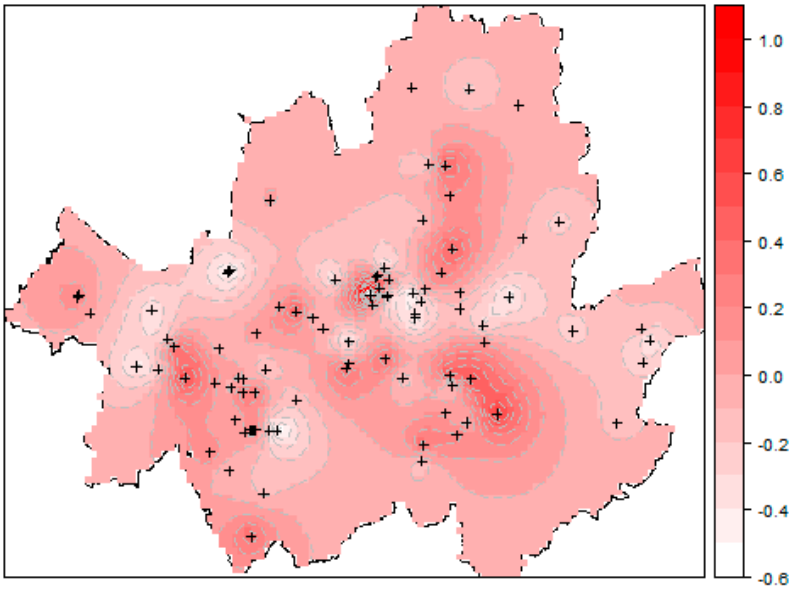

그림 4. $w(s)$ 연속표면

1,000m 내외까지 상업용 토지 가격의 공간적 상관성이 존재하는 것으로 풀이할 수 있다.

식(3)에서 공간적 임의효과를 나타내는 항 $w(s)$는 순수 오차항 $\varepsilon(s)$의 효과를 제거한 후 가격의 공간적 상관성을 나타낸다. 즉 무작위 형태의 잡음(noise)이 아닌 일종의 신호(signal)를 나타내는 항으로서 최종 모형에서 산출된 $w(s)$를 연속표면으로 구축한 결과는 그림 4와 같다.

그림 4의 패턴은 서울시 상업용 토지 가격의 공간적 상관성 정도를 시각화한 것으로 해석할 수 있다. 즉 강남구, 중구 및 종로구는 $w(s)$=0.0(서울시 평균)을 기준으로 이보다 가격 수준이 높게 형성된 지역을, 반대로 서울 서쪽의 강서구와 마포구, 동쪽의 성동구, 남서쪽의 동작구, 관악구 및 금천구는 가격 수준이 상대적으로 낮게 형성된 지역을 의미한다.

이러한 $w(s)$의 공간 패턴은 상업용 토지 가격이 상호 간에 영향을 미치는 지역적 범위, 즉 유사가격권 내지 상권을 의미하는 것으로 풀이할 수 있다. 따라서 $w(s)$의 패턴을 살펴 헤도닉 모형의 적용 범위 등을 정하는데 유용하게 활용할 수 있다.

마지막으로 표 9는 본 논문에서 활용한 3가지 모형(OLS 모형, 전문가 지식 반영 모형, 전문가 지식 및 공간적 상관성 반영 모형)을 검증자료(n=23)에 적용하여 산출된 가격 예측력 결과이다. 모형의 성능을 판단하는 기준으로는 통계학에서 일반적으로 사용하는 잔차 제곱합($\Sigma(Y_i-\hat{Y}_i)^2=\Sigma e_i^2$)과 부동산 대량평가 모형에서 널리 활용되는 지표 COD(Coefficient Of Dispersion)를 사용하였다. COD는 아래와 같은 산식을 통하여 계산된다[15].

$$COD=\frac{\left[\frac{\Sigma|\text{개별비율}-\text{비율들의 중위수}|}{\text{비율들의 개수}}\right]}{\text{비율들의 중위수}}\times 100 \quad (12)$$

식(12)에서 비율은 실제 가격(거래가격) 대비 모형을 통해 산출된 추정가격(estimated price)의 비율을 말한다. COD가 작을수록 실제 가격과 추정가격의 격차가 작은 것으로 해석할 수 있으며, 통상 20.0~25.0 정도를 상한선으로 하여 이보다 작은 값을 갖는 경우 과세표준 산정 등 정부 행정업무에서 받아들일 수 있는 수준으로 간주한다(IAAO, 2010).

표 9. 모형 간 가격 예측력 비교[16]

구분	전문가 지식 반영	공간적 상관성 반영	$\sum_{i=1}^{23} e_i^2$	COD
Model 1	NO	NO	4.35	40.2
Model 2	YES	NO	2.62	27.5
Model 3	YES	YES	1.42	19.4

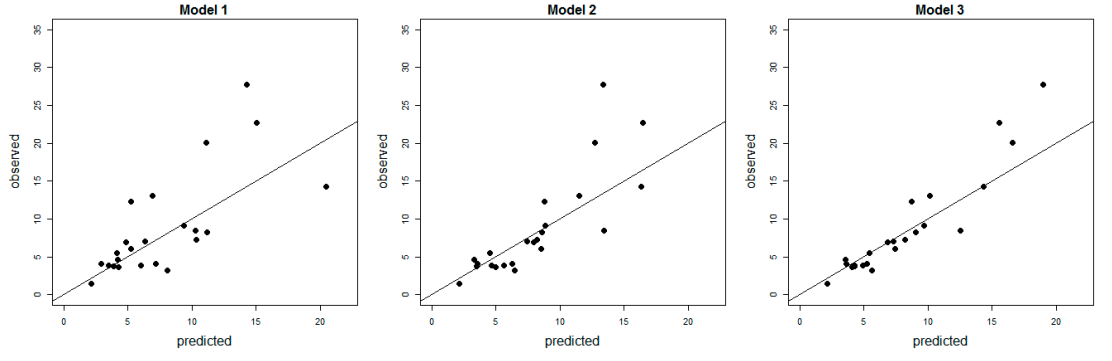

그림 5. 실제 가격과 추정가격의 비교

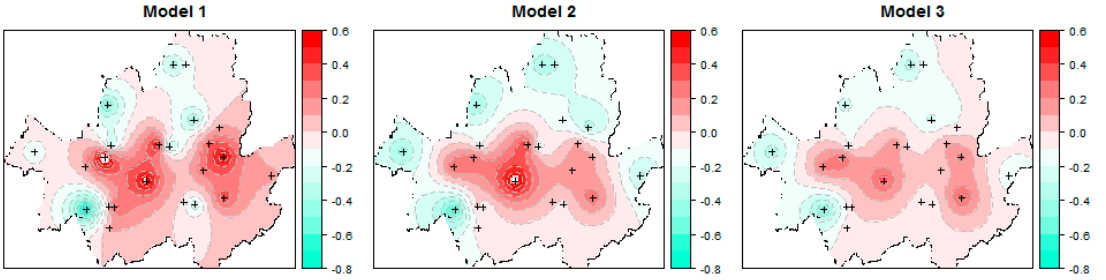

그림 6. 검증자료(n=23) 기준 잔차의 분포 패턴

표 9를 보면 모형 1부터 시작해서 모형 3으로 갈수록 모형 성능이 개선되었음을 알 수 있다. 즉, 모형 1에서 모형 2의 개선은 인근지역의 질을 나타내는 지역 더미에 전문가 지식을 반영한 것에 기인하며, 모형 2에서 모형 3으로의 개선은 잔차에 포함된 지가의 공간적 상관성을 추가적으로 반영한 것에 기인한다. 모형을 정교하게 구성함으로써 가격 예측력이 향상되었다는 사실은 실제 가격(observed)과 추정가격(predicted)의 일치성 정도를 표현한 그림 5 및 검증자료의 잔차를 표현한 그림 6을 통해서도 확인할 수 있다. 특히 그림 6의 경우 OLS 모형에서 뚜렷하게 나타났던 잔차의 공간적 상관성이 최종 모형에서는 어느 정도 완화되었음을 알 수 있다.

5. 결론

본 논문은 거래빈도가 낮아 지금껏 활발하게 시도되지 못한 상업용 토지의 가격을 정확히 추정하는데 초점을 맞추었다.

상업용 토지 가격의 추정을 위해 선형 결합 형태의 평균 구조, 지수 형태의 공분산함수 그리고 순수 오차항을 구성요소로 하는 모형을 구축하여 실거래가 자료에 적용하였다. 상업용 토지 가격의 특성상 상권별로 가격수준이 차별적으로 형성된다는 점을 감안하여 자료의 공간적 상관성을 명시적으로 반영할 수 있는 대표적 공간보간기법인 크리깅 방법을 활용하였다. 더 나아가 희소한 자료의 한계를 극복하기 위해 전문가 지식을 사전 확률분포의 형태로 모형에 반영할 수 있는 베이지안 크리깅 방법을 활용하였다.

베이지안 크리깅은 최근 들어 강수량, 농산물 수확량 추정 등 환경과학 분야에서 적용된 바가 있으나(Diggle & Ribeiro, 2002; Jiang *et al.*, 2009; Heo & Park, 2009), 사회과학 분야에서 시도된 바가 드물고, 또한 이러한 접근법을 적용하였더라도 베이지안 기법의 장점인 사전지식 또는 맥락지식(contextual knowledge)을 활용하지 못한 점이 아쉬움으로 지적되고 있다(Diggle & Ribeiro, 2002).

따라서 본 논문은 베이지안 크리깅 기법을 부동산 가격 추정에 적용하되, 전문가의 주관적 지식을 명시적으로 모형에 반영하였다는 점 등에서 기존 연구와 차별성을 갖는다.

검증 자료의 적합을 통해 제시하였듯 전문가 지식의 반영과 공간적 상관성의 명시적 고려는 가격 추정의 정확성을 높였다. 본 논문의 결과는 거래 자료가 희소한 상황에서도 신뢰성 있게 부동산 가격을 추정해야하는 경우(스키장 등 거래가 거의 없는 특수 부동산에 대해 공시지가를 산정하거나 금융기관이 대출금액을 결정해야하는 경우 등)에 유용하게 활용될 수 있을 것으로 기대된다.

본 논문은 전문가가 생각하는 서울시 자치구별 가격수준 차이를 지역 더미에 반영하였다. 그러나 지역 더미의 공간 스케일을 바꾸어 시도로 상향 또는 동(洞)으로 하향 조정할 경우 본 논문에서와 같은 결과가 나올 것으로 확신할 수는 없다. 이러한 예상은 표준지 데이터의 생성과정과 밀접한 관련이 있다. 현행 표준지 공시지가는 구별로 적정한 가격균형을 유지

하기 위한 감정평가사 간 협의절차 등이 마련되어 있으나, 시도 수준(예를 들어 서울시와 경기도)에서의 가격균형은 그만큼 엄격하게 관리되지 않고 있다. 또한 동의 경우 일부 동에는 표준지 자체가 분포하지 않아 결측 지역이 나타날 가능성이 높다. 본 논문에서는 이러한 점을 감안하여 지역 더미의 스케일을 구로 설정하고 분석을 진행하였다.

마지막으로 본 논문에서 전문가의 주관적 지식은 지역 더미의 구성에 국한하여 활용하였는데, 다른 가격요인(용도지역, 도로조건 등)의 경우에도 전문가의 지식이나 경험치가 있다면 활용이 가능할 것이다.

본 논문을 통해 전문가 지식을 정성적이라 하여 배척하기보다는 적극적으로 활용할 수 있다는 점, 그리고 부동산 가격 형성의 가장 중요한 요소는 바로 '지리적 위치'라는 사실이 의미 있게 부각되기를 기대한다.

주

1) 국토교통부 보도자료(2014.5.6)

2) non-informative prior 및 informative prior라 하기도 한다.

3) 문헌에 따라 보편 크리깅(universal kriging)을 kriging with external drift, regression-kriging 등으로 지칭하기도 한다. 학자에 따라 세부적 정의를 다르게 내리기도 하고 추정방법에도 약간의 차이가 있지만 본 연구에서는 보편 크리깅으로 용어를 통일하였다. 보다 자세한 내용은 Hengl(2009) 참조.

4) τ^2은 다양한 각도에서 해석할 수 있다. 측정 오류(measurement error), 자료의 최소 이격거리보다 더 작은 스케일(scale)에서 발생하는 변동성(microscale variability), 부동산 거래의 경우 매도자 및 매수자의 협상력 등 모형에 포함시키기 어려운 잡음(noise) 등으로 해석할 수 있다.

5) 추정해야할 모수들을 θ라 한다면 $\theta=(\beta,\sigma^2,\tau^2,\phi)^T$로 나타낼 수 있다. 이 경우 모수들 간에 어떠한 상관관계가 있다는 강력한 증거가 없는 한, 각 모수들 간에는 아무런 관계가 없다는 독립적 사전 확률분포를 가정하는 것이 통상이다(Banerjee *et al.*, 2004, p.131). 사전 확률에 대해 독립성을 가정함으로써, 모수 벡터 θ에 대한 사전 확률은 θ를 구성하는 $\beta,\sigma^2,\tau^2,\phi$의 주변 사전 확률분포(marginal prior)를 단순하게 곱하는 형식으로 표현할 수 있게 된다.

6) 공간 일반선형모형(spatial generalized linear model)이라고 하며 링크 함수(link function) η를 통해 다음과 같이 비정규 우도로 쉽게 확대할 수 있다(Diggle *et al.*, 1998).
$g(E(Y(s)))=\eta(s)=\boldsymbol{x}^T(s)\beta+w(s)$

7) 일반적으로 활용되는 공분산함수의 형태는 구형(spherical), 지수(exponential), 가우시안(gaussian) 등이 있으며 그 형태는 유사한 편이다. 동일한 자료에 대해 여러 공분산함수가 모두 잘 들어맞는 경우가 많으므로 함수의 형태 결정은 그리 중요하지 않으며(Ghosh & Carriazo-Osorio, 2007) 본 연구에서는 함수 형태가 보다 단순한 지수 공분산함수를 적용하였다.

8) $\gamma(t)=C(0)-C(t)$의 관계가 있으므로(Banerjee *et al.*, 2004) 어떠한 것을 사용하여도 무방하며, 학문 분야에 따른 관행인 것으로 보인다.

9) 각 계수의 정의와 해석 등에 대해서는 Isaaks & Srivastava(1989) 참조.

10) 공간적 상관성(spatial correlation)이 0.05 이하로 떨어지는 거리를 range로 해석할 경우, $\exp(-\phi t_0)=0.05$가 되고, 이를 정리하면 $t_0\approx 3/\phi$가 된다.

11) 검증자료는 각 자치구 내에서 임의추출(random sampling)하였으며, 서울시 자치구는 25개이나 2개 구(노원구, 금천구)에는 실거래가 자료가 1개씩밖에 존재하지 않아 검증자료를 추출하지 않았다.

12) 자료의 수가 가장 많아(14개) 기준범주로 정하였다.

13) 회귀계수의 기댓값이 0.00이며, 표준편차가 100이라는 의미이므로 β에 대해 어떠한 사전정보도 투입하지 않은 셈이 된다(모호 사전분포). 모호 사전분포의 경우 통상 기댓값은 0.00, 표준편차는 10에서 1000까지 다양하게 부여된다(Gelman *et al.*, 2004; Gelman & Hill, 2007; Gelfand 2012).

14) σ^2 및 ϕ는 Banerjee *et al.*(2004) 및 Gelfand(2012)의 사례를 따라 충분히 무정보적인 분포를 부여하되, MCMC 수렴이 용이하도록 감마분포 $G(0.1,0.1)$을 사전 확률분포로 지정하였다. $G(0.1,0.1)$에서 첫 번째 인자(α)는 shape parameter, 두 번째 인자(β)는 rate parameter로서, 이 경우 평균 1.0($=\alpha/\beta$), 분산 10.0($=\alpha/\beta^2$)이 되어 모수 추정치 범위에 제한을 가할 정도는 아니다.

15) 공시지가 등 국내 가격공시제도에서 가격 적정성은 COD를 기준으로 검토하고 있으며 이는 해외의 경우도 마찬가지이다.

16) 심사자의 의견을 따라 전문가 지식은 반영하지 않되, 공간적 상관성만을 반영한 모형(비베이지언, Regression-kriging 모형 적용)의 잔차제곱합은 2.36, COD는 27.1로 나타나 모형 2(전문가 지식 반영, 공간적 상관성 미반영)와 비슷한 수준의 가격 예측력을 보이고 있다.

참고문헌

국토교통부, 2014, 감정평가정보체계, https://www.kais.kr.

김성우 · 정건섭, 2010, "부산 아파트 실거래가를 이용한 전통적 헤도닉모형과 공간계량모형간의 적합도에 관한 비교연구," 부동산학연구, 16(3), 41-55.

김종수, 2012, "실거래가격을 활용한 개별주택가격의 적정성 분석," 부동산연구, 22(2), 29-56.

서경천 · 이성호, 2001, "공간적 자기회귀모델과 토지시장 분할에 의한 효율적 지가추정에 관한 연구," 국토계획, 36(4), 1-18.

서교, 2005, "헤도닉분석기법과 공간계량경제모형을 이용한 농촌지역 지가의 영향인자 분석," 농촌계획, 11(3), 11-17.

이창무 · 김종현 · 김형태, 2009, "시세 대비 실거래가를 활용한 아파트 호별 세부 특정가격 추정," 국토계획, 44(4), 67-77.

Banerjee, S., Carlin, B.P. and Gelfand, A.E., 2004, *Hierarchical Modeling and Analysis for Spatial Data*, Chapman & Hall/CRC, Boca Raton.

Baranzini, A., Ramirez, J., Schaerer, C. and Thalmann, P., 2008, *Hedonic Methods in Housing Markets: Pricing Environmental Amenities and Segregation*, Springer, New York.

Bruanuer W.A., Lang, S. and Feilmayr, W., 2013, Hybrid multilevel STAR models for hedonic house prices, *Jahrbuch für Regionalwissenschaft*, 33(2), 151-172.

Chica-Olmo, J., 2007, Prediction of housing location price by a multivariate spatial method: cokriging, *Journal of Real Estate Research*, 29, 91-114.

Choy, S.L., O'Leary, R. and Mengersen, K, 2009, Elicitation by design in ecology: using expert opinion to inform priors for Bayesian statistical model, *Ecology*, 90(1), 265-277.

Clapp, J.M., Nanda, A. and Ross, S.L., 2008, Which school attributes matter? The influence of school district performance and demographic composition on property values, *Journal of Urban Economics*, 63(2), 451-466.

Dennis, B., 1996, Discussion: should ecologists become Bayesian?, *Biological Applications*, 6, 1095-1103.

Diggle, P.J., Moyeed, R.A. and Tawn, J.A., 1998, Model-based geostatistics (with discussion), *Applied Statistics*, 47, 299-350.

Diggle, P.J. and Ribeiro, P.J., 2002, Bayesian Inference in Gaussian Model-based Geostatistics, *Geographical & Environmental Modeling*, 6(2), 129-146.

Gelfand, A.E., 2012, Hierarchical modeling for spatial data problems, *Spatial Statistics*, 1, 30-39.

Gelman, A., Carlin, J.B., Stern, H.S. and Rubin, D.B., 2004, *Bayesian Data Analysis*, Chapman & Hall/CRC, Boca Raton.

Gelman, A. and Hill, j., 2007, *Data Analysis Using Regression and Multilevel/Hierarchical Models*, Cambridge University Press, Cambridge.

Geman, S. and Geman, S., 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern. Anal. Mach. Intell.*, 6, 721-741.

Ghosh, G.S. and Carriazo-Osorio, F., 2007, Bayesian and Frequentist Approaches to Hedonic Modeling in a Geo-Statistical Framework, *American Agricultural Economics Association Annual Meeting*, Portland, OR.

Gill, J. and Walker, L.D., 2005, Elicited Priors for Bayesian Model Specifications in Political Science Research, *The Journal of Politics*, 67(3), 841-874.

Hengl, T., 2009, A Practical Guide to Geostatistical Mapping, University of Amsterdam, Amsterdam.

Heo, T.Y. and Park, M.S., 2009, Bayesian Spatial Modeling of Precipitation Data, *The Korean Journal of Applied Statistics*, 22(2), 425-433.

International Association of Assessing Officers, 2010, *Standard On Ratio Study*, IAAO, Kansas City.

Isaaks, E.H. and Srivastava, R.M., 1989, *An Introduction to Applied Geostatistics*, Oxford University Press, New York.

Jiang, P., He, Z., Kitchen, N.R. and Sudduth, K.A., 2009, Bayesian analysis of within-field variability of corn yield using a spatial hierarchical model, *Precision Agric*, 10, 111-127.

Johnson, S.R., Tomlinson, G.A., Hawker, G.A. and Granton, J.T., 2010, A valid and reliable belief

elicitation method for Bayesian priors, *Journal of Clinical Epidemiology*, 63(4), 370-383.

Jones, G. and Johnson, W., 2014, Prior Elicitation: Interactive Spreadsheet Graphics With Sliders Can Be Fun, and Informative, *The American Statistician*, 68(1), 42-51.

Kiel, K.A. and Zabel, J.E., 2008, Location, location, location: The 3L Approach to house price determination, *Journal of Housing Economics*, 17(2), 175-190.

Kuntz, M. and Helbich, M., 2014, Geostatistical mapping of real estate prices: an empirical comparision of kriging and cokriging, *International Journal of Geographical Information Science*, DOI: 10.1080/13658816.2014.906041.

Lai, Y.S., Zhou, X.N., Utzinger, J. and Vounatsou, P., 2013, Bayesian geostatistical modeling of soil-transmitted helminth survey data in the People's Republic of China, *Parasites & Vectors*, 6:359.

Lawson, A.B., 2009, *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*, CRC press, Boca Raton.

Lele, S.R. and Das, A., 2000, Elicited Data and Incorporation of Expert Opinion for Statistical Inference in Spatial Studies, *Mathematical Geology*, 32(4), 465-487.

Martin, T.G., Kuhnert, P.M., Mengersen, K. and Possingham, H.P., 2005, The power of expert opinion in ecological models using bayesian methods: impact of grazing on birds, *Ecological Applications*, 15(1), 266-280.

McCarthy, M.A., 2007, *Bayesian Methods for Ecology*, Cambridge University Press, Cambridge.

Militino, A.F., Ugarte, M.D. and Garcia-Reinaldos, L., 2004, Alternative models for describing spatial dependence among dwelling selling prices, *Journal of Real Estate Finance and Economics*, 29(2), 193-209.

Montero, J. and Larraz, B., 2011, Interpolation methods for geographical data: housing and commercial establishment markets, *Journal of Real Estate Research*, 33, 233-244.

Plant, R.E., 2012, *Spatial data analysis in ecology and agriculture using R*, CRC press, Boca Raton.

Scholte, R., Gosoniu, L., Malone, J.B. and Chammartin, F., 2014, Predictive risk mapping of schistosomiasis in Brazil using Bayesian geostatistical models, *Acta Tropica*, 132, 57-63.

Slater, H. and Michael, E, 2013, Mapping, Bayesian Geostatistical Analysis and Spatial Prediction of Lymphatic Filariasis Prevalence in Africa, *PLoS ONE*, 8(8), e71574.doi:10.1371/journal.pone.0071574.

Tobler, W., 1970, A computer movie simulating urban growth in the Detroit region, *Economic Geography*, 46(2), 234-240.

Truong, P.N., Heuvelink, Gerard B.M. and Gosling, J. P., 2013, Web-based tool for expert elicitation of the variogram, *Computers & Geosciences*, 51, 390-399.

Webster, R. and Oliver, M.A., 2007, *Geostatistics for Environmental Scientists, Statistics in Practice*, John Wiley & Sons, Chichester.

Wheeler, D. C., Paez, A., Spinney, J. and Waller, L. A., 2013, A Bayesian approach to hedonic price analysis, *Papers in Regional Science*, DOI:10.1111/pirs.12003.

교신: 박기호, 151-742, 서울시 관악구 관악로 599, 서울대학교 지리학과(이메일: khp@snu.ac.kr, 전화: 02-880-6453, 팩스: 02-876-9498)

Correspondence: Key Ho Park, Department of Geography, Seoul National University, 599, Gwanangno, Gwanak-gu, Seoul, 151-742, Korea (e-mail: khp@snu.ac.kr, phone: +82-2-880-6453, fax: +82-2-876-9498)

최초투고일 2014. 9. 29
수정일 2014. 10. 21
최종접수일 2014. 10. 24