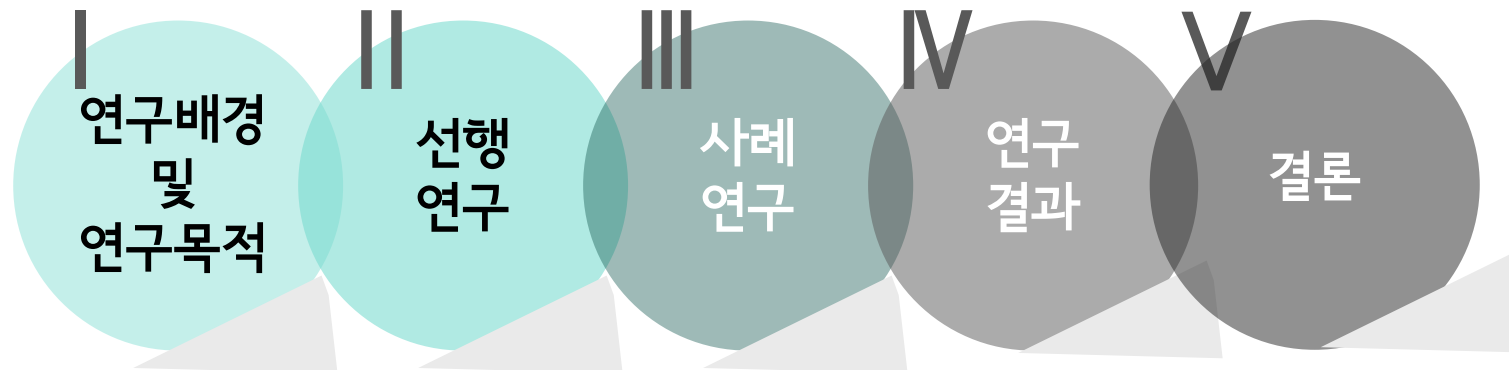


GWL을 적용한 공간 헤도닉 모델링

-서울시 아파트 실거래가격을 대상으로-

진찬우(서울대학교 지리학과 석사과정)
이건학(서울대학교 지리학과 조교수)

목차



I. 연구 배경 및 연구 목적

- 부동산 가격 예측의 중요성
 - 부동산 가격은 사회경제활동의 상호작용을 반영하는 지표
 - 과세와 지역 계획 등의 다양한 분야에서 기초자료로 활용됨
 - 사회적으로 중요한 지표이지만 미시적 공간 단위의 정보 접근이 어려움
 - 현실과 유사한 추정과 예측이 중요
 - 예측의 어려움 : 지역적 이질성, 가격결정요소의 복잡성
- 기존 연구의 한계
 - 공간적 효과에 대한 이해 부족 : 공간적 종속성, 이질성 간과
 - 변수의 제한 : 다중공선성, 적절한 변수 선택의 어려움

I. 연구 배경 및 연구 목적

- 새로운 방법론의 등장
 - 공간 헤도닉 모형
 - 벌점화 회귀 방법
 - 공간 회귀 모형과 벌점화 회귀 방법의 결합
 - 다중공선성 처리에 집중
 - 다차원 데이터 처리 능력 간과
- 연구 목적 : 공간 헤도닉 모델에서 GWL 적용가능성 탐색
 - 국지적으로 다양한 부동산 가격결정요인 탐색
 - GWL 헤도닉 모델 구축
 - 기존 방법론(OLS, GWR)과의 비교를 통한 정확성, 유용성 평가

II. 선행연구

1) 전통적인 헤도닉 모형 : 최소제곱법(Ordinary Linear Square; OLS)

- $Y = XB + \varepsilon \rightarrow \hat{B} = (X^T X)^{-1} X^T Y$
- 공간적 효과(이질성과 종속성)를 고려하지 않음
- 한계 : 변수의 수가 많거나, 변수들간의 다중공선성이 발생하는 경우 연산 불가

2) 공간 헤도닉 모형 : 지리가중회귀(Geographically Weighted Regression; GWR)

- 공간적 이질성 포착에 유리

$$y_i = \beta_{i0} + \sum_k \beta_{ik} x_{ik} + \varepsilon_i$$

- 공간적 종속성 반영

$$\hat{B}_i^{gwr} = (X^T W_i X)^{-1} X^T W_i Y$$

- 한계

- 근린 폭 이내에 반드시 일정 수준의 데이터를 포함 해야 함
- 다중공선성 문제 상존

II. 선행연구

3) 벌점화 헤도닉 모형

- OLS 추정에서의 행렬 연산 문제 해결
- 편의(Bias)와 분산(Variance)간의 상쇄관계(trade-off)를 활용
- Ridge Regression

$$\hat{\beta}^{ridge} = \arg \min_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2$$

- LASSO(Least Absolute Shrinkage and Selection Operator)

$$\hat{\beta}^{lasso} = \arg \min_{\beta} \sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

4) 벌점화 공간 헤도닉 모형 : 지리가중라소(Geographically Weighted LASSO; GWL)

- GWR과 LASSO의 한계 극복

$$\widehat{B}_i^{gwl} = (X^T W_i X + \lambda I)^{-1} X^T W_i Y$$

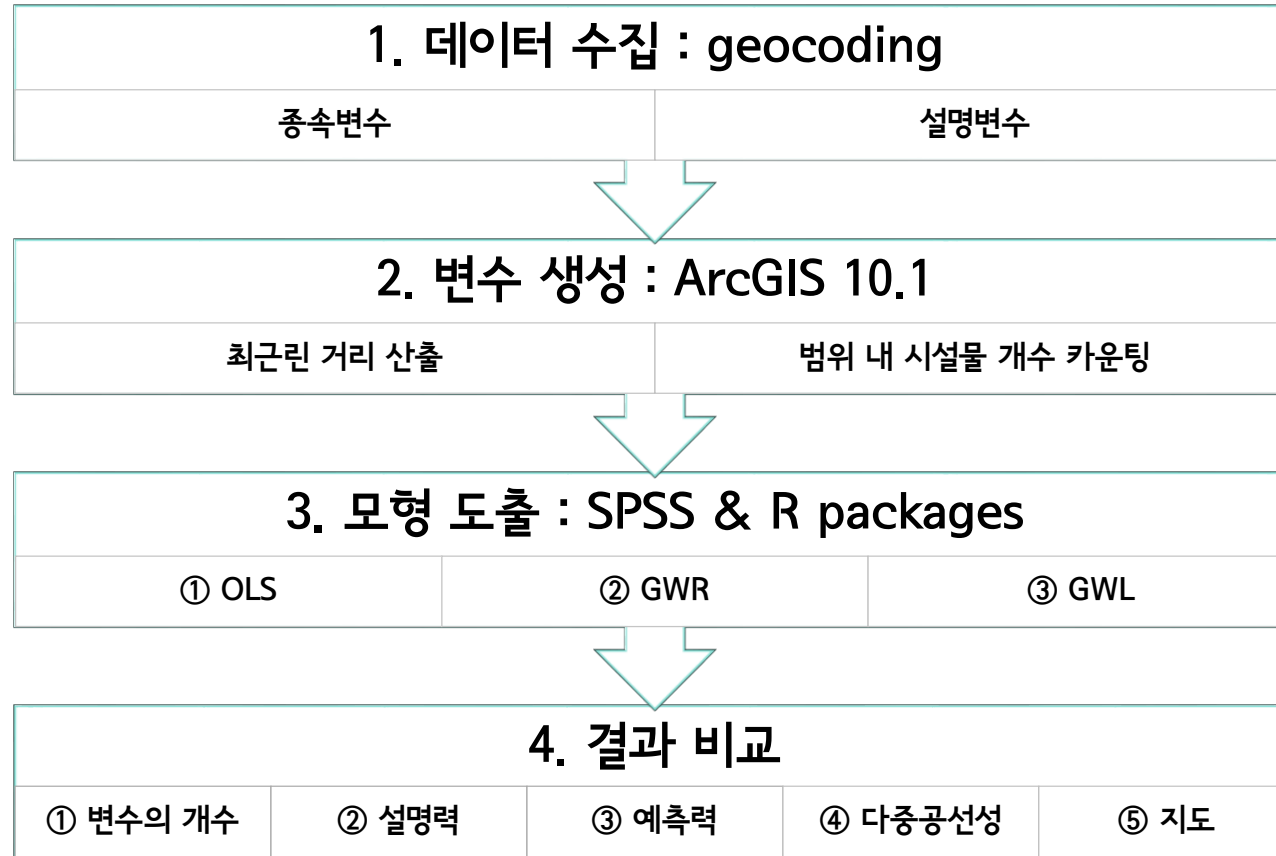
- GWR보다 공간적 이질성을 보다 효과적으로 반영

III. 사례 연구

1) 연구 방법

- OLS, GWR, GWL 모델 구축
 - SPSS, R
 - OLS : SPSS의 단계적 변수 선택법을 이용
 - GWR, GWL : R *gwrr* package
- 각 모형 별 결과 확인
 - 설명력 : $Adj. R^2$ (OLS), 평균 $Adj. R^2$ (GWR, GWL)
 - 예측력 : 평균절대오차(Mean Absolute Error; MAE), 평균제곱근오차(Root Mean Square Error; RMSE)
 - 다중공선성 : $VIF > 10$ (OLS), 전역적 계수 상관 $> |0.7|$ (GWR, GWL)
 - 추정 결과 지도화 : IDW를 이용한 시각화
- 모형 별 결과 비교
 - 모델 간 예측력 차이 검정 : ANOVA, T-test
 - 지도를 통한 오차의 공간적 분포에 대한 시각적 검정

연구흐름도



III. 사례연구

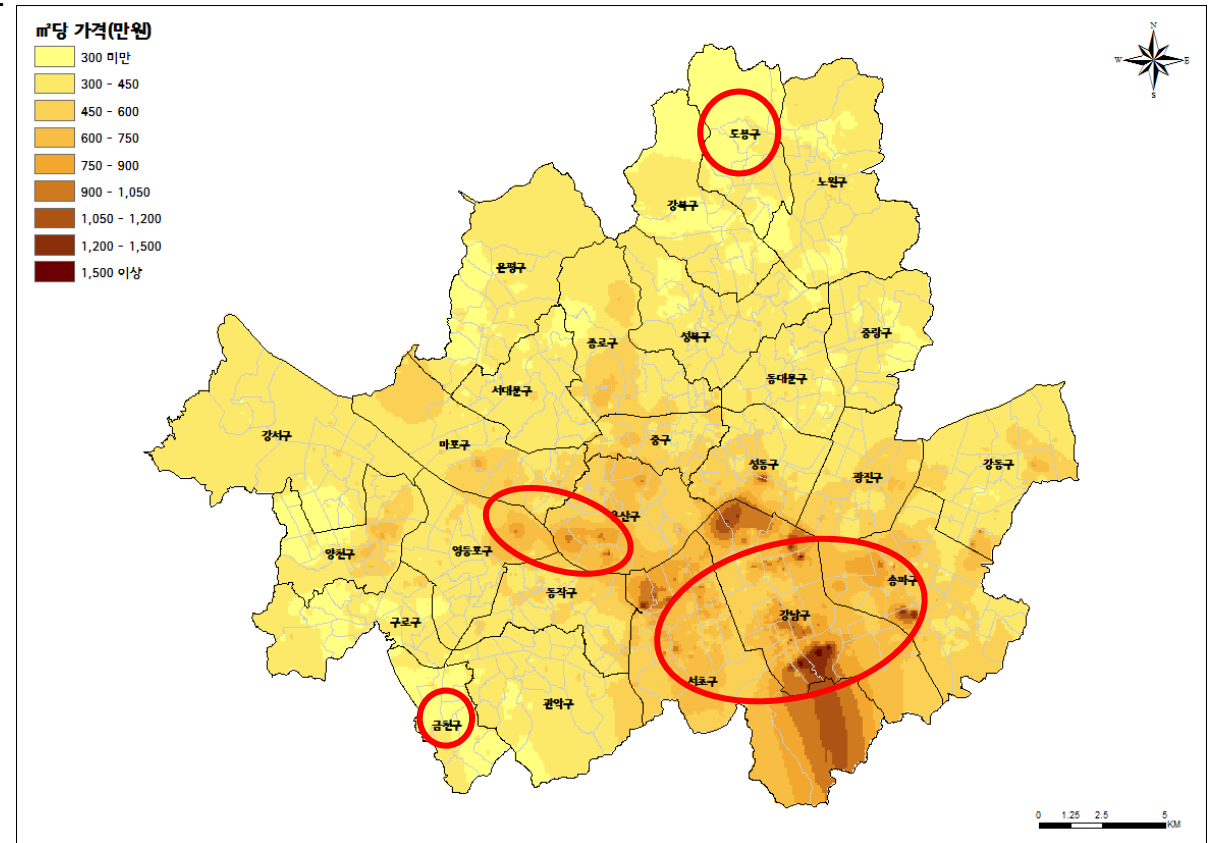
2) 사례지역 및 데이터

■ 종속 변수 : 2013년 서울시 아파트 단지 단위 평균 실거래가

- 포털 제공 실거래가의 아파트 단지별 평균가
- 주택실거래가자료와 대조
- 일반적 사실과 부합하는 분포

■ 독립 변수 : 3개 범주의 39개 변수 생성

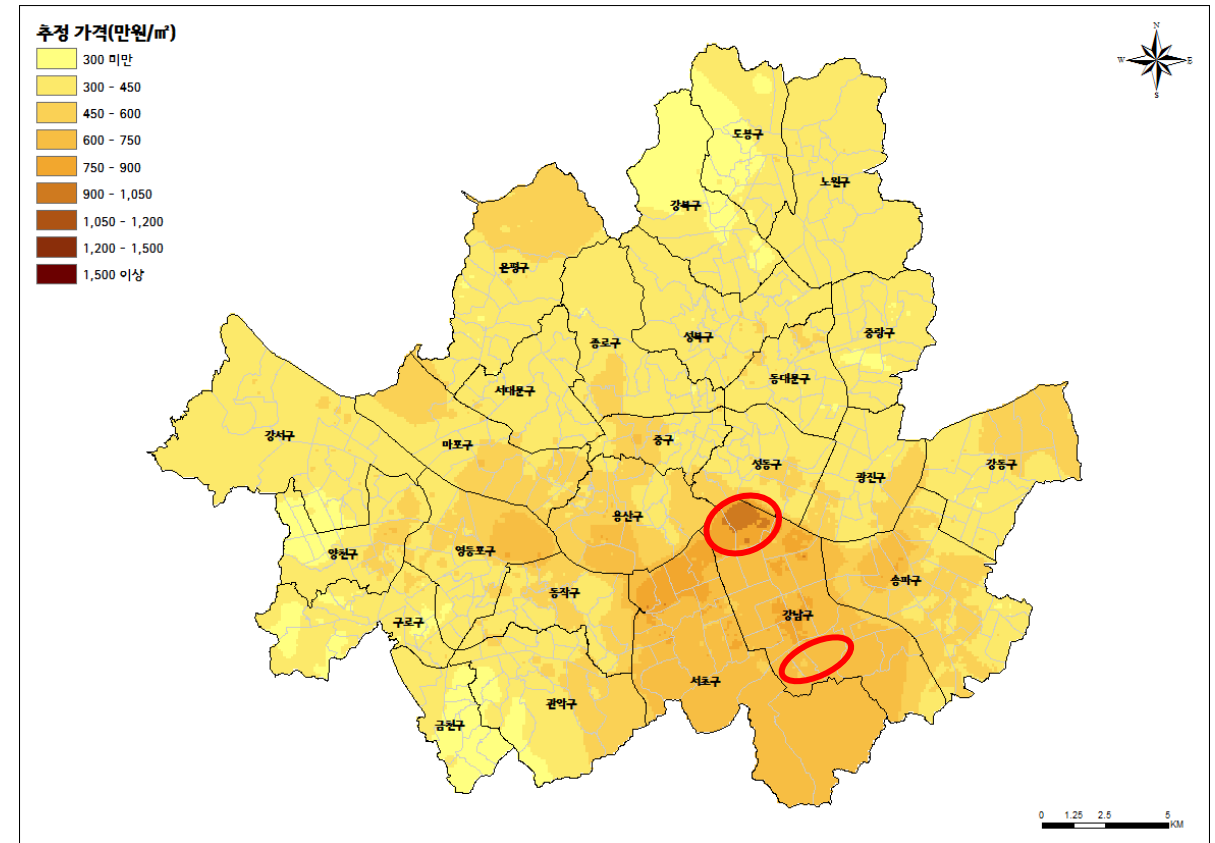
- 구조변수 : 아파트의 건축적 특징 8개
=> 연식, 층수, 면적, 용적률, 건폐율 등
- 입지변수 : 아파트의 입지적 특징 19개
=> 교육, 교통, 상업, 환경, 보건 시설물과의 거리
- 근린변수 : 아파트가 속한 지역의 특징 12개
=> 인구학적 특성, 소득 매개변수, 지역 발전수준 등



IV. 연구결과

1) OLS

- 27개의 유의한(신뢰수준 95%) 변수 도출
- 설명력 : $Adj. R^2 = 0.6646$
- 예측력 : $MAE = 71.146$, $RMSE = 105.211$
- 다중공선성 : VIF 검정 결과 발견되지 않음
- 공간적 분포
 - 전역적 패턴 포착
 - 실제 최고가와와의 큰 오차
 - 지역 간 편차가 두드러지지 않음

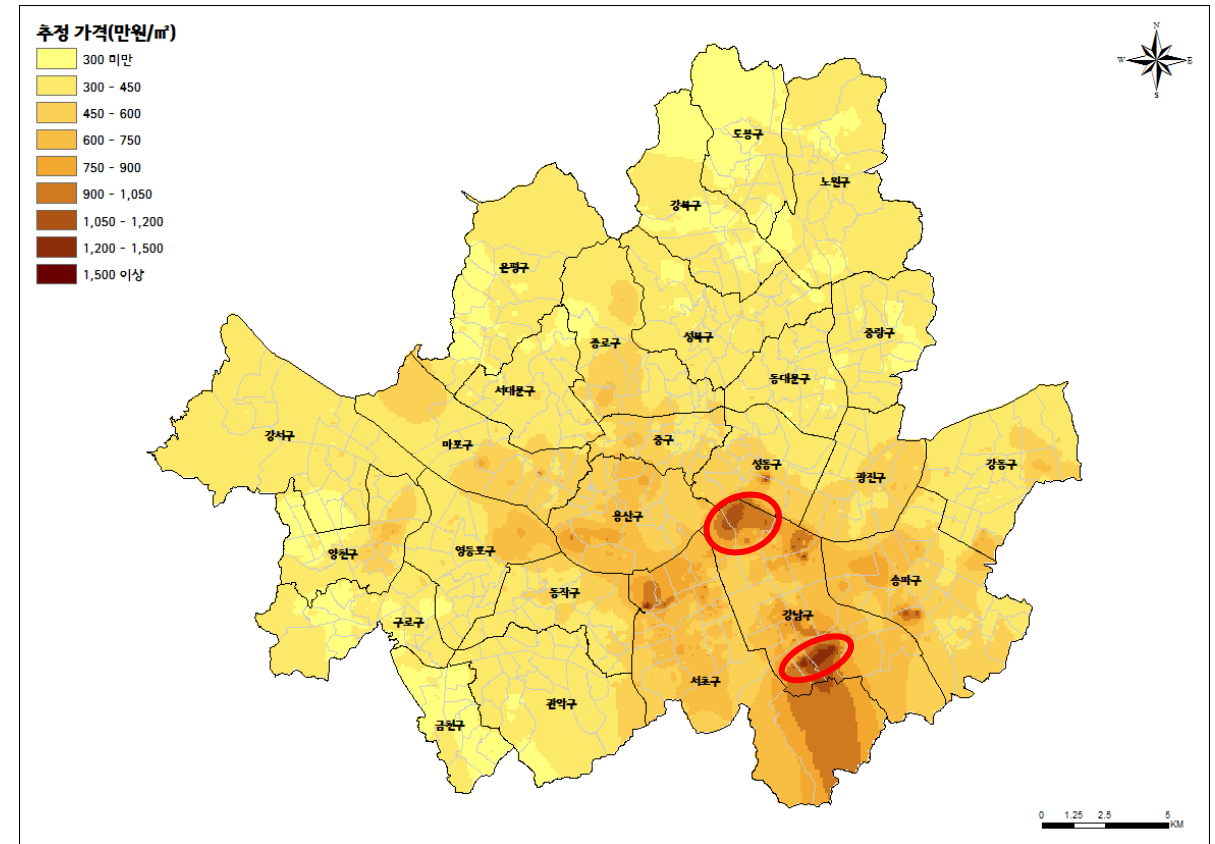


아파트 단지 평균 추정 가격의 분포(OLS)

IV. 연구결과

2) GWR

- 모든 아파트 단지에서 39개의 변수를 포함하는 모델 도출
- 설명력 : 평균 $Adj. R^2 = 0.921$
- 예측력 : $MAE = 31.731$, $RMSE = 50.459$
- 다중공선성 : 있음
 - 도로인접-간선도로거리(0.735^{**})
 - 용적률-평균층수(-0.713^{**})
 - 인구밀도-고용중심지(0.703^{**})
- 공간적 분포
 - 전역적 수준의 지역적 격차가 두드러짐
 - 최고가 추정이 정밀해짐
 - 국지적 가격차가 선명히 확인됨

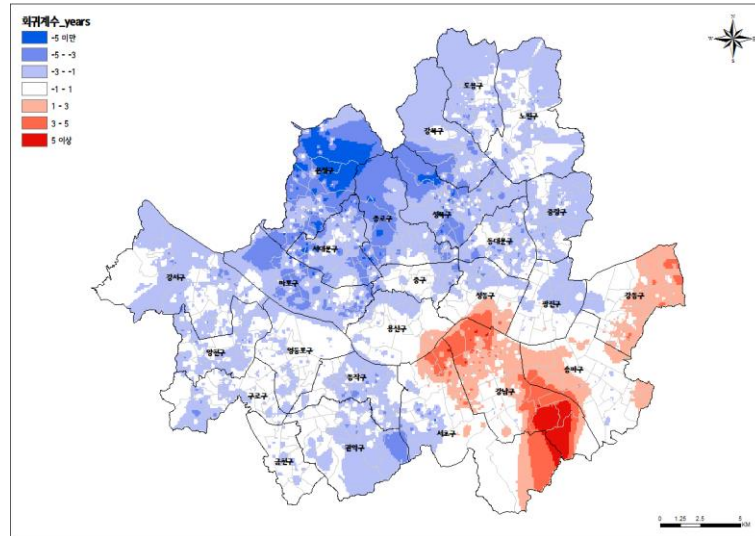


아파트 단지 평균 추정 가격의 분포(GWR)

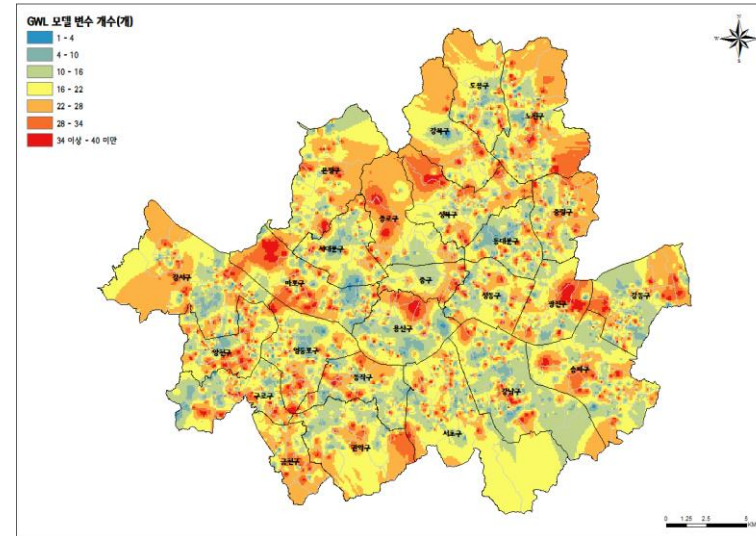
IV. 연구결과

3) GWL

- 모든 단지에서 각기 다른 변수를 포함하는 모델 도출
 - 지역에 따라 포함되는 변수의 개수와 종류가 다름(평균 19개)
 - 예시 1: 비슷한 오차 수준에서 포함되는 변수의 개수가 다름
강서구 C단지 : 20개 vs. 동대문구 J단지 : 11개
 - 예시 2: 비슷한 오차 수준에서 변수의 개수가 같아도 종류가 다름
노원구 D단지 : 층수, 면적, 고용중심지, 노령인구, 학력
동작구 S단지 : 면적, 노령인구, 학력, 자가소유비, 단독주택



회귀계수의 공간적 분포 예시(건축연한)



변수 개수의 공간적 분포

3) GWL

-
- 추정 가격(만원/m²)**
- | |
|---------------|
| 300 미만 |
| 300 - 450 |
| 450 - 600 |
| 600 - 750 |
| 750 - 900 |
| 900 - 1,050 |
| 1,050 - 1,200 |
| 1,200 - 1,500 |
| 1,500 이상 |
- Map labels include: 도봉구, 노원구, 강북구, 은평구, 종로구, 성북구, 중랑구, 강서구, 강동구, 양천구, 영등포구, 동작구, 용산구, 서초구, 강남구, 송파구, 금천구, 구로구.
- Scale: 0 1.25 2.5 5 KM

아파트 단지 평균 추정 가격의 분포(GWL)

IV. 연구결과

4) 모형 비교

■ 모형 별 결과

- GWL이 가장 간단한 모형
- GWR이 가장 설명력 높은 모형
- 예측력은 GWR과 GWL의 결과가 상충됨
- GWR은 다중공선성 문제가 발견됨

	변수의 개수	$Adj. R^2$	MAE	RMSE	다중공선성
OLS	27	0.6646	71.146	105.211	없음
GWR	39	(평균)0.921	31.729	50.459	있음
GWL	(평균)19	(평균)0.913	25.271	52.982	없음

IV. 연구결과

4) 모형 비교

■ 예측력 검증

- ANOVA : 세 모델에서 도출된 MAE와 RMSE 값 모두는 같지 않다고 볼 수 있음
- T-test : GWR과 GWL의 RMSE 차이는 다르다고 보기 어려움

➡ GWL의 예측력이 MAE에 근거하여 우수하다고 할 수 있음

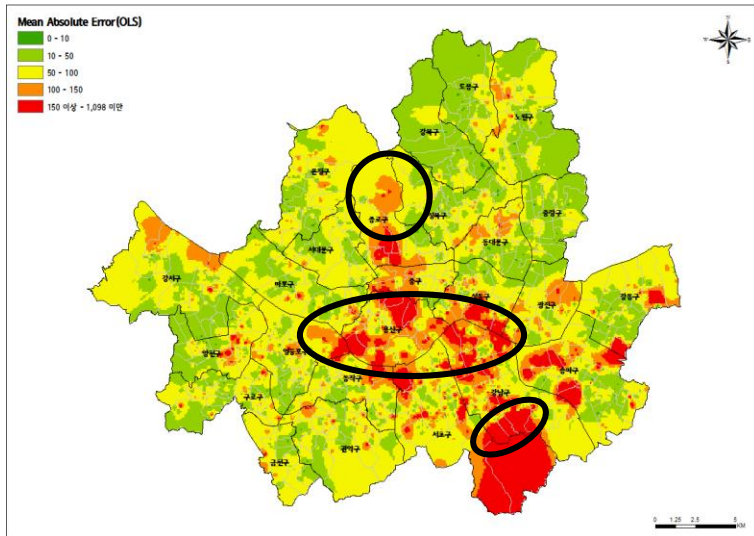
예측력 지표	분석 방법	비교 모델	F(또는 T)값	P값
MAE	ANOVA	OLS-GWR-GWL	611.038	0.000***
	T-Test	OLS-GWR	25.703	0.000***
		OLS-GWL	28.74	0.000***
		GWR-GWL	6.007	0.000***
RMSE	ANOVA	OLS-GWR-GWL	70.1137	0.000***
	T-Test	OLS-GWR	9.036	0.000***
		OLS-GWL	8.649	0.000***
		GWR-GWL	-0.57	0.56

IV. 연구결과

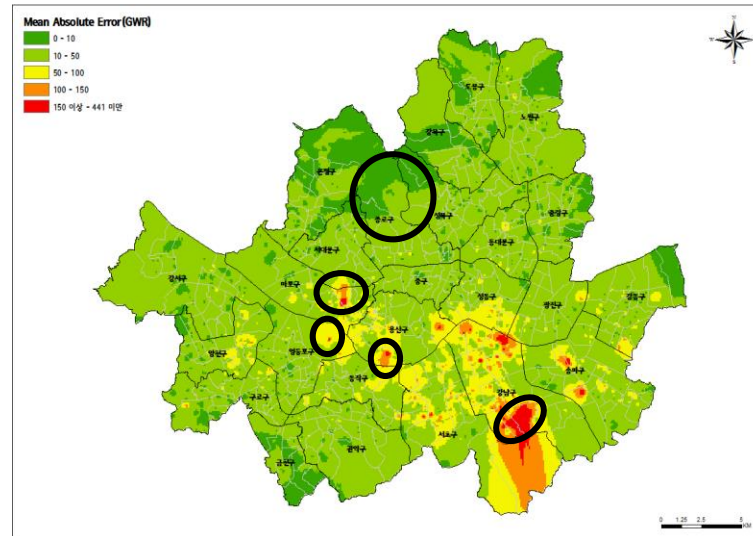
4) 모형 비교

■ 예측 오차(MAE)의 공간적 분포

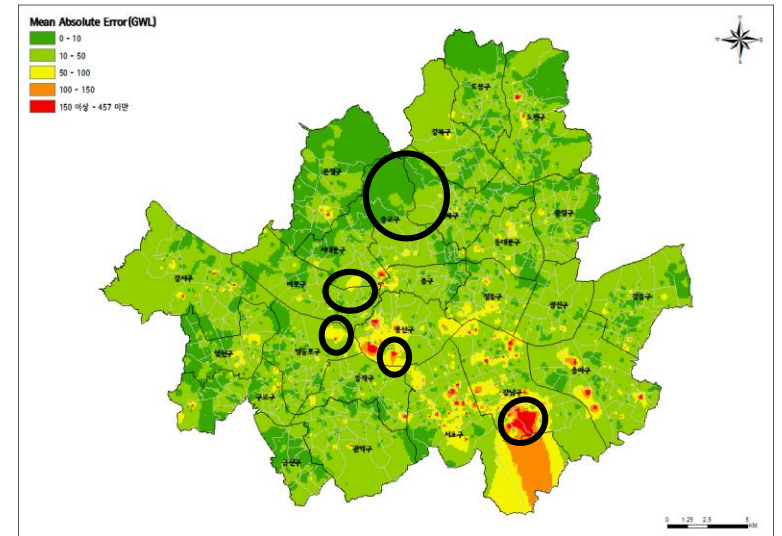
- 모델 별 차이가 유의하고, 해석에 용이한 지표인 MAE를 사용
- 비교를 위한 직관적 수준의 급간 구분



a) OLS



b) GWR



c) GWL

V. 결론

■ 연구 요약

- 2013년 서울시 아파트 실거래가를 기초로 새로운 방법론(GWL)을 도입한 공간 헤도닉 모델을 구축
- 기존의 방법론(OLS, GWR)에 의한 모델과 비교하여 GWL의 정확성과 유용성을 평가
- GWL은 공간적 효과를 고려함과 동시에 고차원 데이터를 효율적으로 분석하는 방법으로 예측력을 높임

■ 연구의 한계 및 후속연구 과제

- 예측력의 과대평가 : 모델 간 상대적인 차이를 분석하는 연구이므로 수용 가능
- 예측력의 일관성 부족 : 후속 연구 필요

■ 연구의 의의

- 부동산 가격 결정요인 집합이 지역별로 다름을 확인
- 부동산 가격 모니터링 : 다수의 변수로 설명되지 않는 지역의 비합리적 가격 요인을 탐색
- 공간빅데이터의 활용 : 다변량 데이터의 효과적인 필터링 가능

참고문헌

- 강창덕, 2010, “GWR 접근법을 활용한 부동산 감정평가 모델 연구,” 부동산연구, 20(2), 107-132.
- 이건학, 김감영, 2013, “개별공시지가와 주택실거래가의 공간적 불일치에 관한 연구,” 대한지리학회지, 48(6), 879-896.
- 이창로, 박기호, 2013, “지가형성요인의 다수준 종단 분석,” 대한지리학회지, 48(2), 272-287.
- Hoerl, A. E., and Kennard, R. W., 1970, Ridge regression, Biased estimation for nonorthogonal problems. *Technometrics*, 12(1), 55-67.
- Löchl, M. and Axhausen, K. W., 2010, Modelling hedonic residential rents for land use and transport simulation while considering spatial effects, *Journal of Transport and Land Use*, 3(2), 39-63.
- Tibshirani, R., 1996, Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society. Series B (Methodological)*, 267-288.
- Wheeler, D. and Tiefelsdorf, M., 2005, Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *Journal of Geographical Systems*, 7(2), 161-187.
- Wheeler, D. C., 2007, Diagnostic tools and a remedial method for collinearity in geographically weighted regression, *Environment and Planning A*, 39(10), 2464-2481.
- Wheeler, D. C., 2009, Simultaneous coefficient penalization and model selection in geographically weighted regression, the geographically weighted lasso. *Environment and planning. A*, 41(3), 722-742.

Thank you

Q & A
