Uncovering drivers of community-level house price dynamics through multiscale geographically weighted regression: a case study of Wuhan, China

Binbin Lu1,3, Yong Ge2\*, Yilin Shi1, Jianghua Zheng3, Paul Harris4

1. School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China

2. State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China

3. College of resource and Environment Sciences, Xinjiang University, Urumqi 830000，China

4. Sustainable Agriculture Sciences North Wyke, Rothamsted Research, Okehampton, UK.

Corresponding author:

Yong Ge

State Key Laboratory of Resources and Environmental Information System, Institute of Geographical Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing, China;

e-mail: [gey@lreis.ac.cn](mailto:gey@lreis.ac.cn)

*Funding*

This work was jointly supported by the National Natural Science Foundation of China (NSFC: 41725006; 42071368; 41871287).

**Uncovering drivers of community-level** **house price dynamics through multiscale geographically weighted regression: a case study of Wuhan, China**

**Abstract**: For buyers, investors and urban policy, understanding drivers of community-level house prices across space and across time, are important for urban management and economic planning. In this study, we interrogated two housing market datasets, one from 2015, the other from 2019, for Wuhan, China, in order to uncover diversities and similarities in the spatial relationships between house price and contextual data; and in the context of increasingly volatile markets. A non-stationary approach was adopted with basic geographically weighted regression (GWR) and multiscale GWR (MGWR), where only the latter enables relationships to vary at their own spatial scale. In terms of model fit, both MGWR (adj. *R*2: 0.94 and 0.97, for 2015 and 2019, respectively) and GWR (adj. *R*2: 0.87 and 0.81) represented an improvement over the usual linear regression (adj. *R*2: 0.11 and 0.09) and the spatial lag mode (adj. *R*2: 0.21 and 0.27). Similarly marked improvements for GWR and for MGWR were found using corrected Akaike Information Criterion (AICc) based fit diagnostics. However, of more importance and via MGWR, the spatially varying drivers of house price were found to operate at a range of spatial scales, that in turn changed in strength and significance between the two study years. Such insights allow for spatially- and temporally-aware decision- and policy-making for housing price control and urban planning, given China’s housing markets can be increasing prone to strong growth coupled with severe depressions.

**Key Words**: Spatial heterogeneity, temporal dynamics, multi-scale, real estate market, urban planning

# 1 Introduction

To provide greater insight into housing market trends, statistical models and machine learning techniques have long been used to help facilitate this (e.g. Kirby 1976, Meen and Meen 2003, Lau and Wei 2018, Hagenauer *et al.* 2019, Credit 2022), some of which focus on market dynamics (e.g. Maurice 1999, Esmaeili *et al.* 2010, Gauvin *et al.* 2011) and its prediction (e.g. Limsombunchai 2004, Deng *et al.* 2017, Yazdani 2021), while others focus on the influential drivers of house price change (e.g. Adair *et al.* 1998, Madsen 2011, Lu *et al.* 2014a, Simlai 2014). Globally, and over the past two decades, housing markets have been particularly volatile, especially those in developed countries or regions where housing markets have displayed irregular patterns of strong growth and severe slumps (Nanthakumaran *et al.* 2000, Bajari *et al.* 2010). For China, housing markets have experienced high volatility caused by rapid urbanization and intensive government interventions (Jiang and Wang 2021). In this respect, elevated house prices in China have become *the* primary indicator of economic development and migration attractiveness for a given city or region (Zhang *et al.* 2015, Wang *et al.* 2017). Consequently, exploring city house price disparities and dynamics can enhance our understanding of inter-city features of development, migration and planning.

To contextualize house price variation and inform the statistical models, factors commonly include consumer behaviour, and various social, economic, locational and demographic characteristics (e.g. Nanthakumaran *et al.* 2000, Orford 2000, Meen and Meen 2003, Leishman and Bramley 2005, Bajari *et al.* 2010, Esmaeili *et al.* 2010, Nagaraja *et al.* 2011). More specifically, such factors have included income, expenditure, construction cost and land price (e.g. Jim and Chen 2006, Madsen 2011, Gasparėnienė *et al.* 2016). In this study, macro-economic, demographic, environment and urban vitality factors are considered.

Geocoded data provides opportunities to geographically explore markets and sub-markets (Fik *et al.* 2003). Commonly, individual property sales, each with unique geo-locations have been studied (e.g. Lu *et al.* 2014a, Yao and Fotheringham 2016). However, a city’s housing market typically represent prices for apartments or flats, where many have the same geo-location, rather than a unique location of a stand-alone house. This situation is particularly true for China (Hu *et al.* 2022). Thus for these markets, analyses are conducted at some aggregated scale; for example: community-level (e.g. Qian *et al.* 2021, You *et al.* 2021), prefecture-level (e.g. Ou *et al.* 2021) and city-level (e.g. Wei and Cao 2017, Garriga *et al.* 2021). As community-level is the fundamental administrative partition of a Chinese city, it is the representative unit in urban planning and construction (Bian *et al.* 2021, Hui *et al.* 2021), and is the aggregation level of choice for modelling house prices (Kang *et al.* 2021).

For this study, attention is placed on capturing spatial heterogeneity in data relationships, rather than spatial autocorrelation effects (Holly *et al.* 2006). Regression models for spatial heterogeneity estimate spatially varying coefficients (SVC) that can be mapped, whereas regressions with homogeneous assumptions typically estimate spatially invariant coefficients (i.e. they are fixed and can’t be mapped) (Harris 2019). There is a rich history of applying SVC models to house price data, such as those using the expansion method (e.g. Can 1992, Kestens *et al.* 2006, Bitter *et al.* 2007), multi-level methods (e.g. Gelfand *et al.* 2007) and moving window kriging (Case *et al.* 2004, Páez *et al.* 2008). The most popular in this respect has been the use of geographically weighted regression (GWR) (e.g. Gao and Asami 2005, Yu 2007, Páez *et al.* 2008, Chen and Truong 2012)

Basic GWR (Brunsdon *et al.* 1996) naively assumes each house price to contextual data relationship locally varies at the same spatial scale; whereas mixed GWR (Brunsdon *et al.* 1999) allows some relationships to be globally fixed while the rest locally vary at the same spatial scale; whereas and finally, multiscale GWR (Fotheringham *et al.* 2017) allows all relationships to vary at their own particular scale. Thus, basic and mixed GWR can be considered as special cases of multiscale GWR (MGWR) (Comber *et al.* 2022); and where MGWR itself has further refinements, such as with different distance metrics (Lu *et al.* 2017) and with different kernel weighting functions (Lu *et al.* 2019) for each relationship.

For this study, basic GWR and basic MGWR models are used, together with standard (and non-spatial) linear regressions for context – each applied at two different time intervals (2015 and 2019), for Wuhan, China. The use of MGWR provides a worthy update to house price studies using only basic (e.g. Paez et al., 2008) and mixed (2016) GWR forms. The structure of this study is organized as follows: in section 2, the community-level house price data and the methodologies of GWR and MGWR are given; in section 3, the results of the model estimations are presented in terms of model fit and significance of coefficient estimates; in section 4, the study results are discussed; and section 5, concludes the study.

# 2 Methods

## 2.1 Study area and data

Wuhan is the capital of Hubei province and the mega-center city in the central region of China. Its total area is 8494.41 square kilometers. As of 2019, Wuhan had a population of around 11,212,000 and gross domestic product (GDP) of around ¥ 1622.3 billion (source: official annual statistical yearbook). In the process of rapid urbanization and in-flow migration, Wuhan’s population increased by 5.7% to that found in 2015. Through in-flow migrations, the average house price of ¥ 9, 500 per square meter (*Yuan/m2*) in 2015 had risen to ¥ 17, 500 *Yuan/m2* in 2019, where this study seeks to investigate the spatial dynamics of this increase. For this study, datasets were collated for central built-up area (CBA) only of Wuhan. The CBA houses around 65% of Wuhan’s total population and has had strong commercial real estate development, also. Clear geographical change in average community-level house price (*Yuan/m2*) in the CBA from 2015 and 2019 is depicted in Figures 1a and 1b, where price increases were not uniform across communities.



(a) 2015 (b) 2019

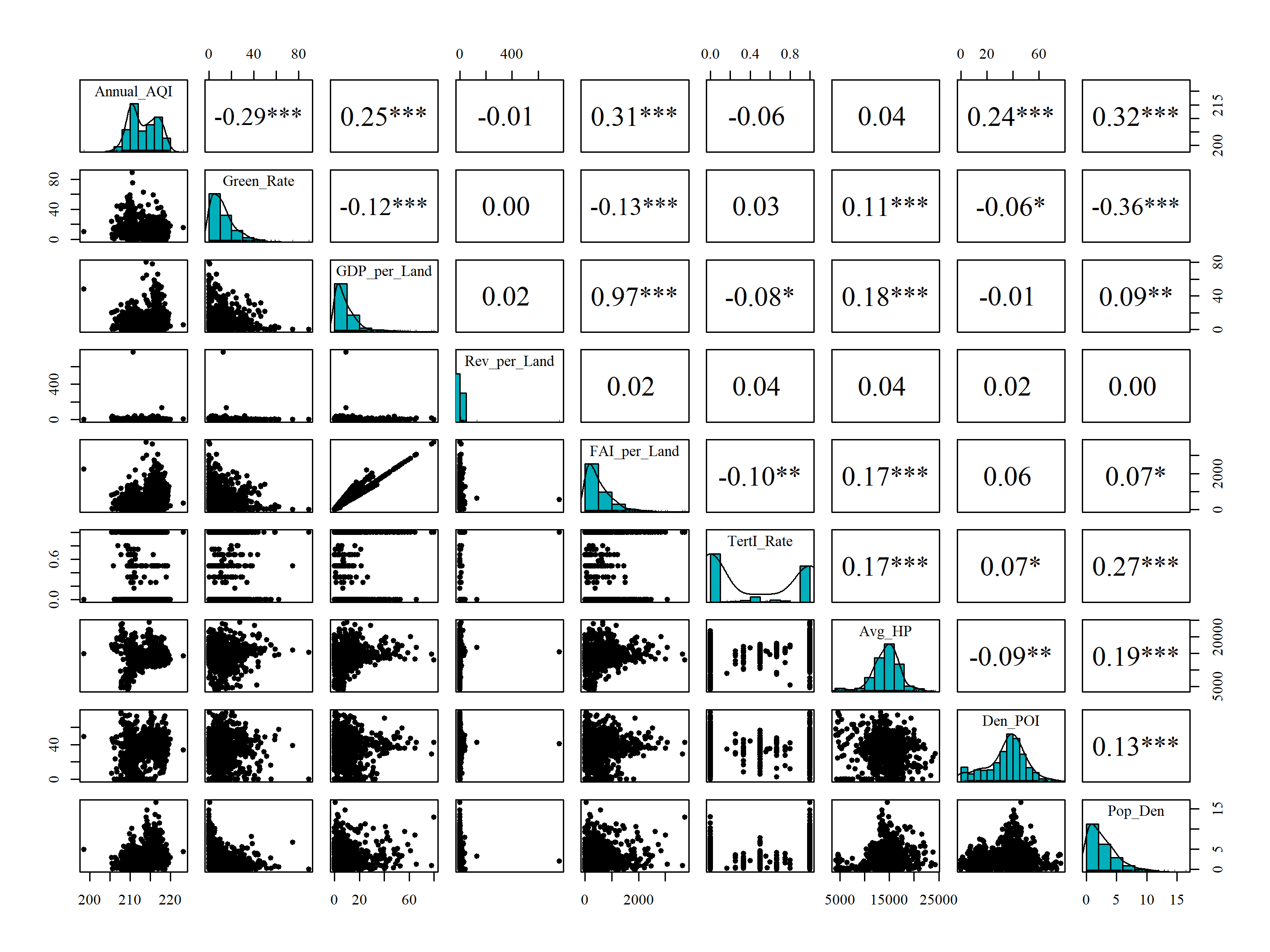
**Figure 1**. Geographical change in average community-level house price per square meter (*Yuan/m2*) in the central built-up area (CBA) of Wuhan from 2015 and 2019.

Table 1 Study variable and their descriptions.

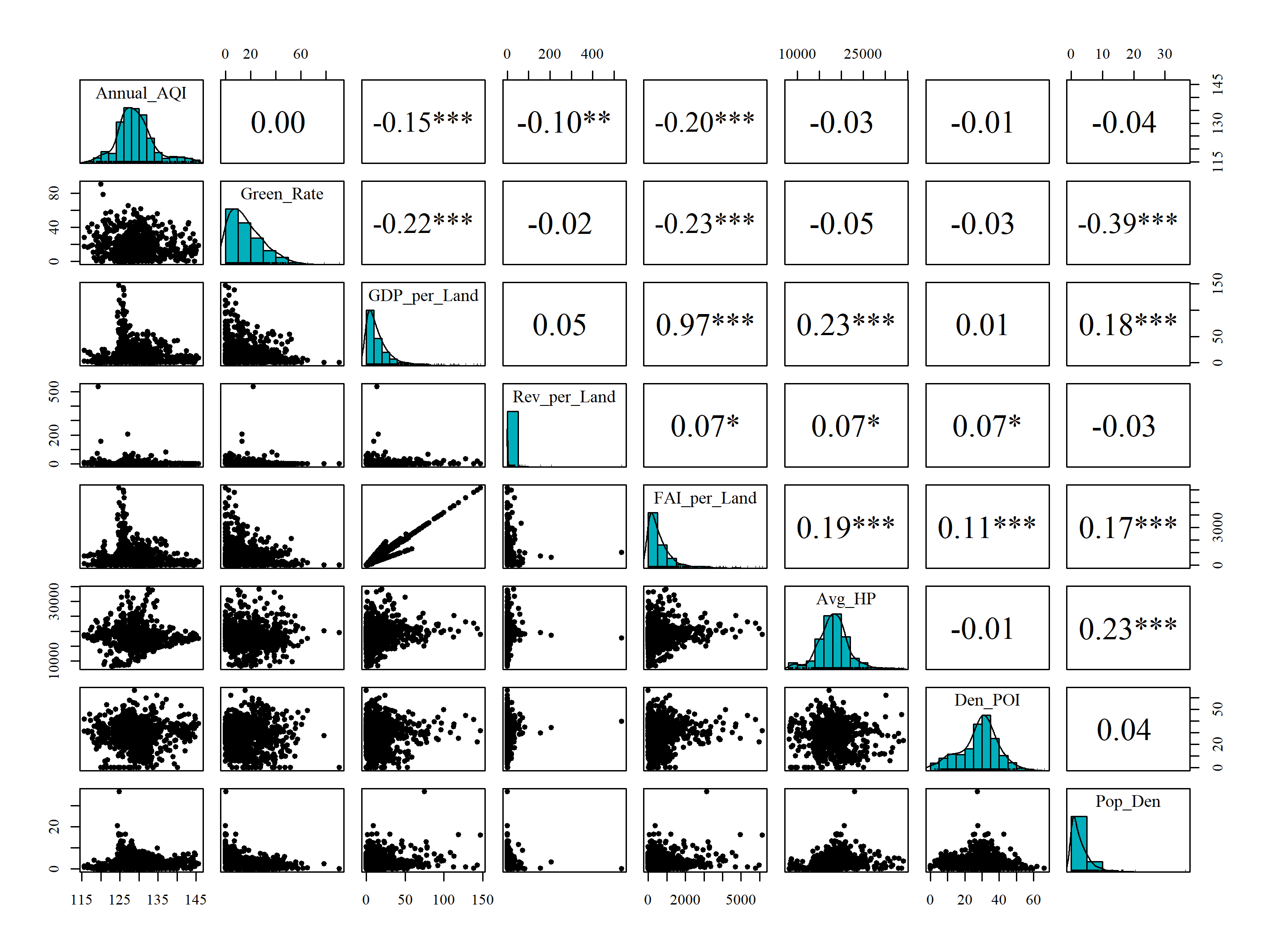
|  |  |
| --- | --- |
| ***Variable*** | **Description** |
| *Avg\_HP* | Average house price at the community-level (*Yuan/m2*) |
| *Annual\_AQI* | Annual air quality index (AQI) |
| *Pop\_Den* | Population density (*\*105 / km2*) |
| *Green\_Rate* | Percentage of green space (%) |
| *GDP\_per\_Land* | GDP per land area (*\*104 Yuan/ km2*) |
| *Rev\_per\_Land* | Revenue per land area (*\*104 Yuan/ km2*) |
| *FAI\_per\_Land* | Fixed assets investment per land area (*\*104 Yuan/ km2*) |
| *TertI\_Rate* | Percentage of tertiary industry (%) |
| *Den\_POI* | Density of places of interest (POI) (*\*10-2/ km2*) |

Expected drivers of community-level house price where collated from multiple data sources, including those for resident population density, economics (land value), environment (air quality index, AQI; green space) and place of interest (POI). Datasets were processed to provide eight variables considered useful in explaining variation in average community-level house price (*Avg\_HP*) for 2015 and for 2019. These were: *Annual\_AQI*, *Pop\_Den, Green\_Rate, GDP\_per\_Land, Rev\_per\_Land, FAI\_per\_Land, TertI\_Rate, Den\_POI* whose descriptions are presented in Table 1. Number of community-level units (sample size) for each study year is 974.

A preliminary non-spatial exploration of this data via histograms, bivariate scatter plots and correlations are shown in Figures 2 and 3, for 2015 and for 2019, respectively. It can be observed that air quality improved over the 5-year period, together with an increase in green space. Some variables have changed little; for example, fixed assets investments (FAI) which tend to depend on prevailing government investment policy. There does not appear to be any strong drivers of house price variation (as the strongest correlation with house price (*Avg\_HP*) is only 0.29 for % tertiary industry(*TertI\_Rate*) in 2019), but locally, relationships may strengthen (i.e., via this study’s GWR and MGWR fits).



**Figure 2**. Scatterplots, correlations and histograms for 2015 data. Significant correlations are indicated at the 5% (marked by \*), 1% (marked by \*\*) and 0.1% (marked by \*\*\*) level.



**Figure 3**. Scatterplots, correlations and histograms for 2019 data. Significant correlations are indicated at the 5% (marked by \*), 1% (marked by \*\*) and 0.01% (marked by \*\*\*) level.

Informed by this exploratory analysis, three of the eight explanatory variables were removed to facilitate robust regression estimations. Firstly, *FAI\_per\_Land* was removed due to its strong positive correlation (collinearity) to *GDP\_per\_Land*. Secondly, *Annual\_AQI* was removed given its very weak correlation to house price. This very weak correlation can be attributable to air quality data being collected at only ten monitoring sites in the CBA. Thirdly, *TertI\_Rate* was removed, as its relatively promising correlation to house price was considered spurious due to its bi-modal behaviour (most values were 0 or 1). Subsequently, the study regressions will be of this general form:

|  |  |
| --- | --- |
|  | (1) |

where is the intercept and where (*i=*1, …, 5) are the five coefficients corresponding to each explanatory variable.

## 2.2 Global regression and GWR models

In the first place, this regression model expressed in Eq. (1) could be regarded as a standard linear regression (LR) and typically estimated by ordinary least squares. As a typical econometric case, the spatial lag model (SLM) is also adopted to explore the neighborhood influence of the dependent variable (Elhorst 2010), and it can be expressed as,

(2)

where is a column vector of the response variable, is the spatial autoregressive coefficient, is the spatial weight matrix, is the intercept, is the dimension vector of 1s, is an matrix of the explanatory variables, is a column vector of coefficients and is random error terms.

Both LR and SLM are global and stationary forms where coefficients are assumed to be stationary and fixed across space. To capture spatially heterogeneous features that reflect spatial variation and diversities in data relationships, GWR can be used which involves a series of location-specific regression estimations (Brunsdon *et al.* 1996, Lu *et al.* 2018). The GWR model can be expressed as:

(3)

where for each community-level unit *i*, *yi* is the average house price, is a location-specific intercept, *βij* is the *j*th location-specific regression coefficient, *xij* is the observed value of the *j*th explanatory variable, *m* is the number of explanatory variables (i.e., *m=*5), *εi*is the random error term by following a normal distribution, and *n* is the number of community units within Wuhan’s CBA. For each regression calibration location *i*, the GWR model is calibrated via a weighted least square approach, expressed as:

|  |  |
| --- | --- |
|  | (4) |

where for each calibration location *i,* is a row vector of estimated coefficients, is an matrix of the explanatory variables with a column of ***1***s for the intercept, is a column vector of the response variable, and is a diagonal matrix of the spatial weightings for each observation. To calculate **,** a kernel weighting function needs to be specified together with an optimized kernel bandwidth. For this study, a bi-square kernel weighting function was chosen together with an adaptive kernel bandwidth (i.e., the number of nearest neighbours used in each local calibration) that was optimally found via a corrected Akaike Information Citerion (AICc) procedure; as detailed in (see Gollini *et al.* 2015).

For this basic form of GWR, only a single bandwidth is found, which directly results in each set of localized coefficients having the same level of spatial smoothing. To allow for different levels of spatial smoothing for each set of localized coefficients, a multiscale GWR (MGWR) can be calibrated that finds multiple, parameter-specific bandwidths (Yang 2014, Fotheringham *et al.* 2017, Leong and Yue 2017, Lu *et al.* 2017, Lu *et al.* 2018); and thus allows an investigation of relationship heterogeneity at relationship-specific spatial scales. The MGWR model can be expressed as:

(5)

where , , …, are the bandwidths for calculating the corresponding set of local coefficients, including the intercept. MGWR is typically calibrated with a back-fitting algorithm, following those used in generalized additive models (Hastie and Tibshirani 1986). In this study, SLM functions of the ***spatialreg R*** package (Bivand and Piras 2015) and MGWR functions of the ***GWmodel* *R*** package (Lu *et al.* 2014b, Gollini *et al.* 2015) were used, which means the back-fitting algorithm detailed in Lu et al. (2017) was applied. As with basic GWR, MGWR was calibrated with bi-square kernel weighting functions with adaptive kernel bandwidths each found optimally via an AICc-based approach.

# 3 Results

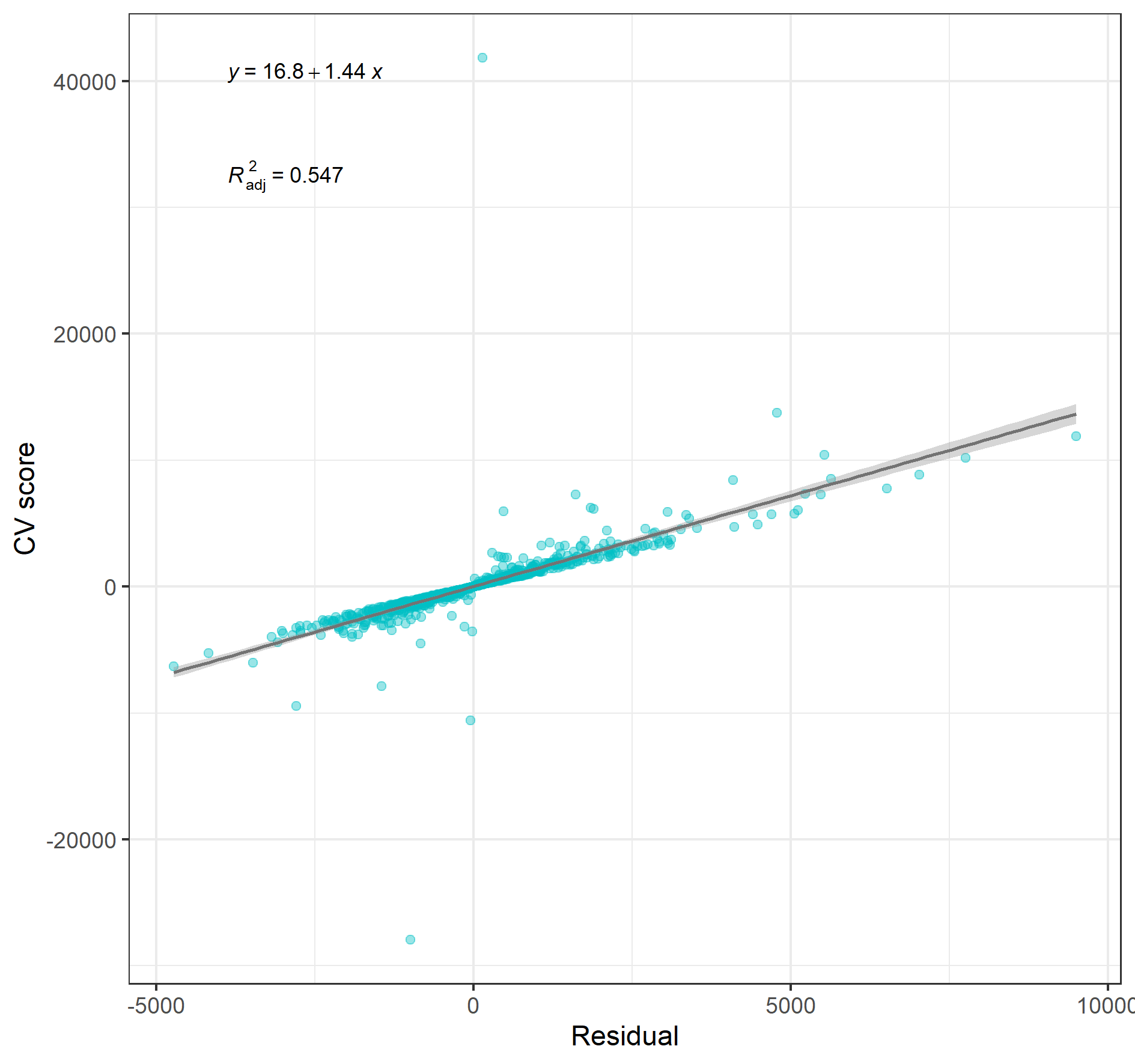
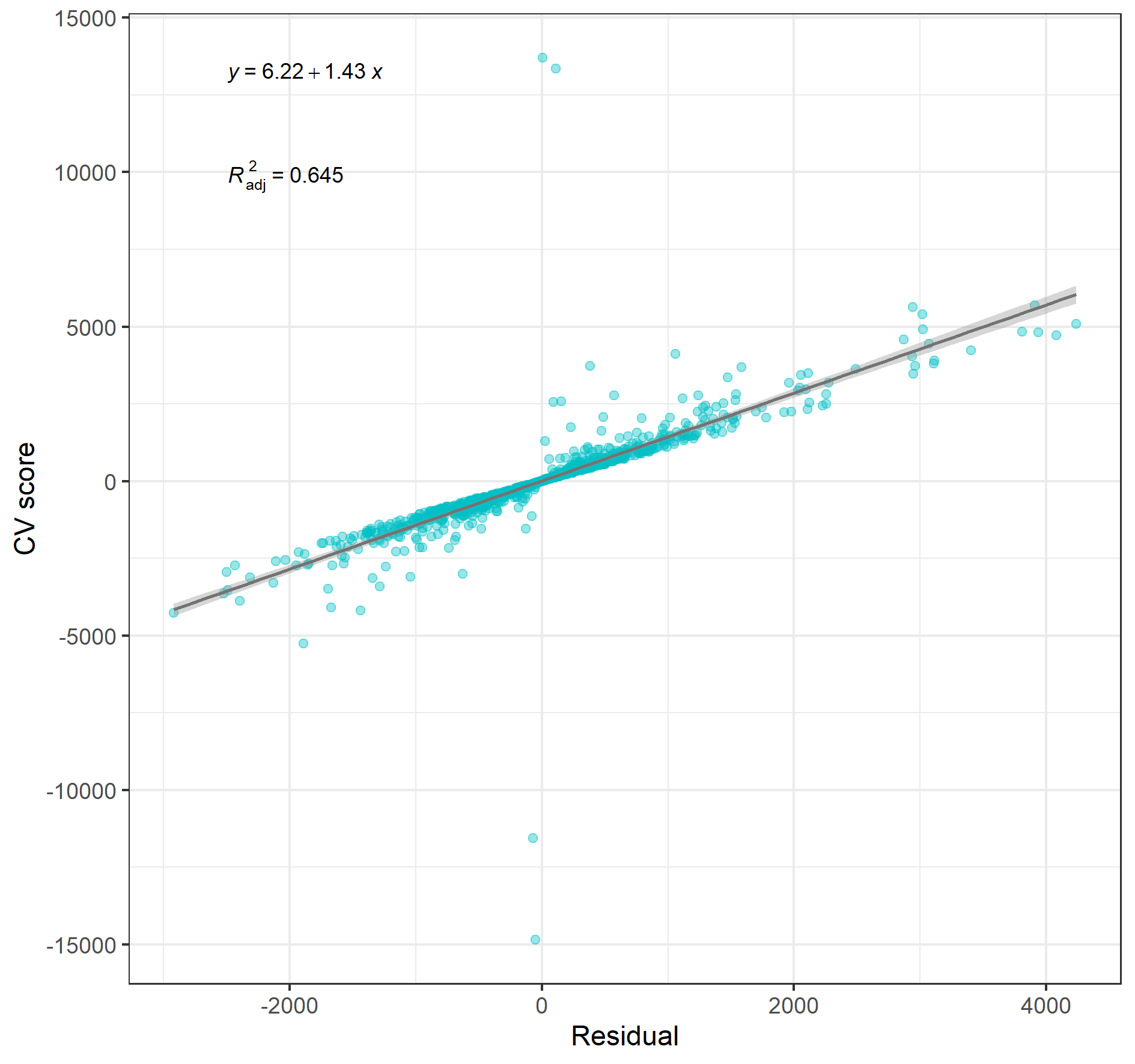
## 3.1 Regression model estimations and fit performance

Model fit diagnostics for LR, SLM, GWR and MGWR, for 2015 and 2019, respectively, are presented in Table 2. The adj. R2, residual sum of squares (RSS) and AICc values are complemented with the optimally-found kernel bandwidths; one for GWR, six for MGWR. Note that the AICc is not applicable for the SLM model and the general form of Akaike Information Citerion (AIC) (Akaike 1973) is provided for all the models. MGWR bandwidths provide insight into the scale of each of the five community-level house price relationships, plus the intercept term. Thus, relationships with *Green\_Rate* and *Den\_POI* were relatively global or stationary in 2015 (as bandwidths relatively large), while those with *GPD\_per\_Land* and *Pop\_Den* were relatively local or non-stationary (as bandwidths relatively small). In 2019, all five relationships were considered local, where the spatial scale of house price relationships with *Green\_Rate* and *Den\_POI* changed the most from that found in 2015. The bandwidth for the intercept term was small for both years. The single bandwidth of GWR (45 nearest neighbours) was not reflective of any bandwidths estimated by MGWR.

Table 2 Diagnostic information of LR, SLM, GWR and MGWR model estimations

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2015 | LR | Adjusted R2 | 0.114 | | | | | |
| AICc | 18267.87 | | | | | |
| AIC | 18267.75 | | | | | |
| RSS | 7856055677 | | | | | |
| SLM | Adjusted R2 | 0.213 | | | | | |
| AIC | 16511 | | | | | |
| RSS | 1005346118 | | | | | |
| GWR | Bandwidth | 45 (number of nearest neighbors) | | | | | |
| Adjusted R2 | 0.868 | | | | | |
| AICc | 16757.53 | | | | | |
| AIC | 16240.47 | | | | | |
| RSS | 744307992 | | | | | |
| MGWR | Coefficient |  |  |  |  |  |  |
| Bandwidth | 14 | 398 | 22 | 145 | 328 | 22 |
| Adjusted R2 | 0.941 | | | | | |
| AICc | 16108.75 | | | | | |
| AIC | 15342.66 | | | | | |
| RSS | 274926241 | | | | | |
| 2019 | LR | Adjusted R2 | 0.094 | | | | | |
| AICc | 18715.25 | | | | | |
| AIC | 17823.47 | | | | | |
| RSS | 12436195146 | | | | | |
| SLM | Adjusted R2 | 0.272 | | | | | |
| AIC | 17162 | | | | | |
| RSS | 1958635540 | | | | | |
| GWR | Bandwidth | 47 (number of nearest neighbors) | | | | | |
| Adjusted R2 | 0.807 | | | | | |
| AICc | 17539.42 | | | | | |
| AIC | 17042.59 | | | | | |
| RSS | 1707951694 | | | | | |
| MGWR | Coefficient |  |  |  |  |  |  |
| Bandwidth | 12 | 50 | 24 | 107 | 27 | 24 |
| Adjusted R2 | 0.919 | | | | | |
| AICc | 17101.01 | | | | | |
| AIC | 15954.91 | | | | | |
| RSS | 474525718 | | | | | |

In terms of model fit, both MGWR (adj. *R*2: 0.94 and 0.97, for 2015 and 2019, respectively) and GWR (adj. *R*2: 0.87 and 0.81) represented an improvement over LR (adj. *R*2: 0.11 and 0.09) and SLM (adj. *R*2: 0.21 and 0.27). Similar, improvements for GWR and for MGWR were found using AICc and AIC, with MGWR providing the largest reductions in AICc or AIC from that found for LR and SLM, for both years. Clearly, models with concerning spatial dependence and heterogeneity shew advantages over the non-spatial LR model, even though LR and SLM are both global models. MGWR provided the greatest improvement in model fit over LR, indicating house price relationships to not only be non-stationary but to also operate at different spatial scales across the CBA of Wuhan, for both 2015 and 2019. In this respect, the coefficients estimated by MGWR were considered more representative of the community-level house price process than those found for GWR and for LR, and in both study years (Lu *et al.* 2018).

Observe the dramatic increases in adj. *R*2 values from around 0.1 (LR models) to more than 0.8 (GWR and MGWR models), which should give rise to the potential overfitting issues. To check whether the local models are affected or not, we adopted the leave-one-out cross-validation (CV) approach (Farber and Páez 2007), and produced the scatterplots between location-wise residuals and CV scores from the 2015 and 2019 GWR models. As shown in Figures 4a and 4b, the CV scores are highly correlated with the residuals from the GWR model estimations, which implies that the improvements are mostly evoked by the spatially heterogeneous concerns instead of extremely over-fits from the localized model estimations.

(a) Results for 2015 (b) Results for 2019

Figure 4 Scatter plots between residuals and CV scores from the 2015 and 2019 GWR models

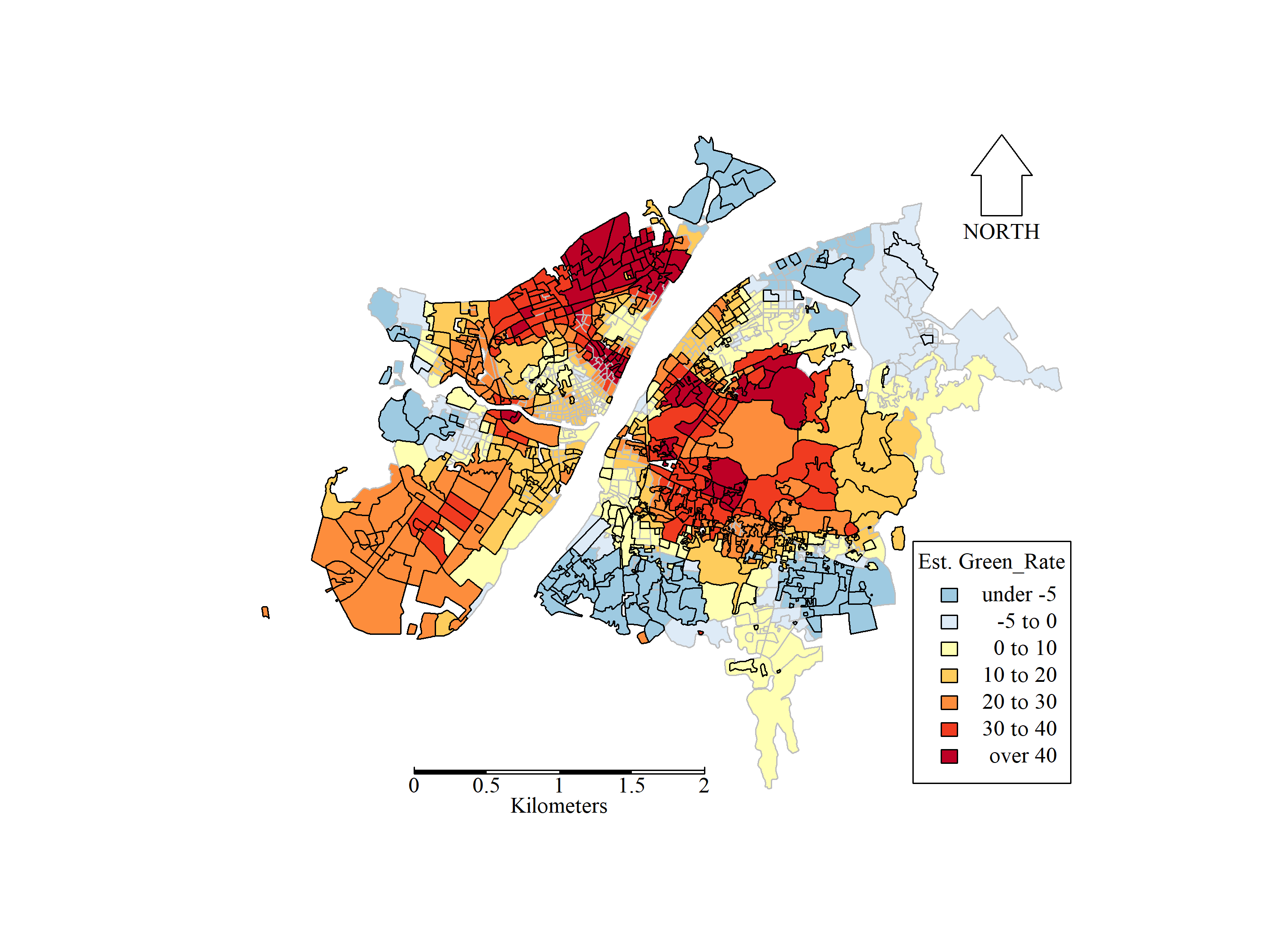
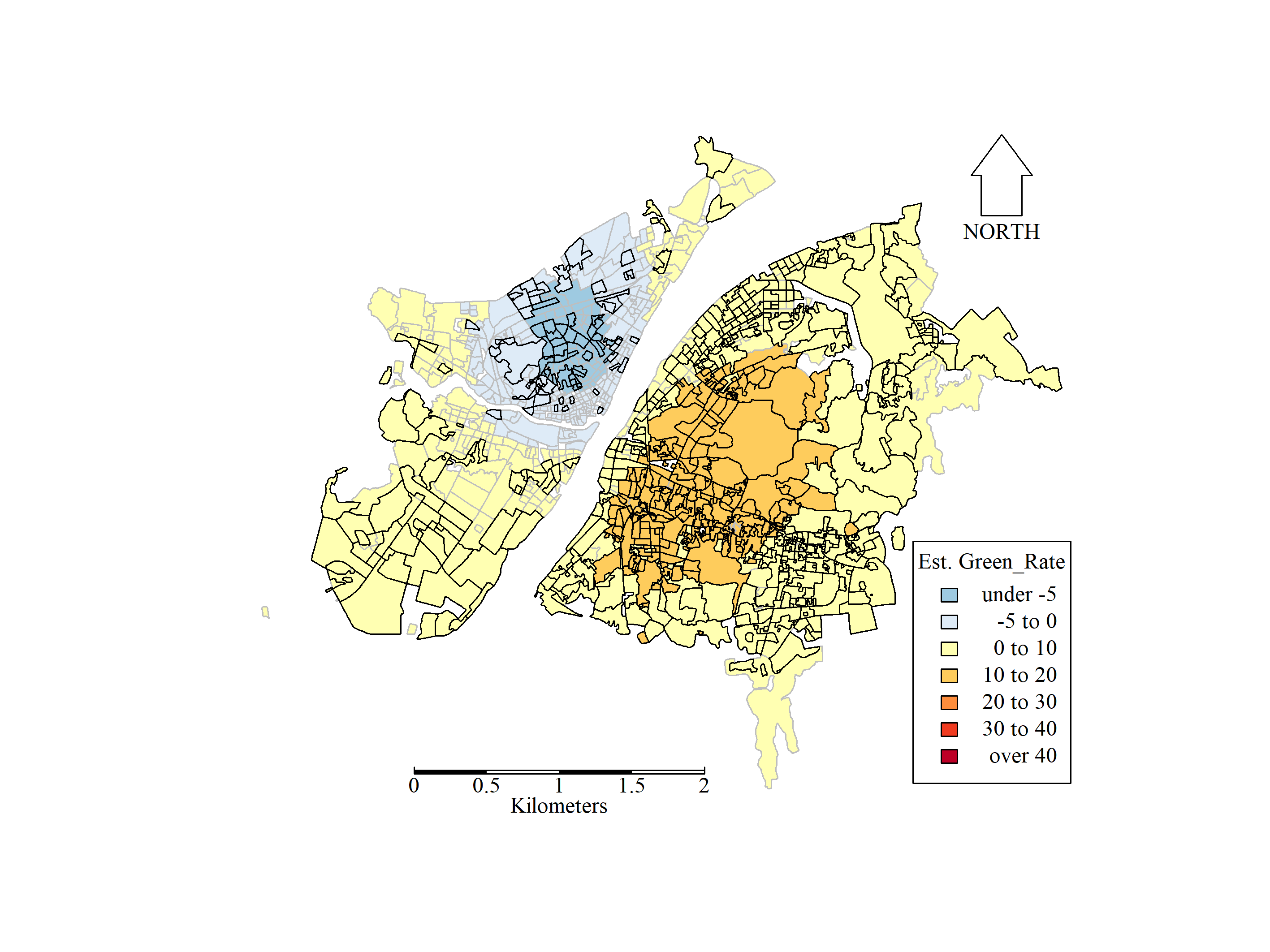
## 3.2 Geographic distribution of MGWR coefficient estimates

Given the model fit results, above, only the pair-wise coefficient estimates from the 2015 and 2019 MGWR estimations need to be presented in detail (Figures 5 to 10). The mapped coefficient estimates are shown as significantly different from zero, at the 5% level, by highlighting the border of the corresponding community-level areal unit. Corresponding LR and SLM results, for stationary coefficient estimates, are given in Table 3 to provide context and to demonstrate value in the MGWR analyses in terms of the spatial, scale-dependent drivers of house price variation over the two years.

Table 3 Coefficient estimates of the LR and SLM model estimations (Significant correlations are indicated at the 5% (marked by \*), 1% (marked by \*\*) and 0.1% (marked by \*\*\*) level)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Model | | *Intercept* | *Green\_Rate* | *GDP\_per\_Land* | *Rev\_per\_Land* | *Den\_POI* | *Pop\_Den* |
| 2015 | LR | 12973.43\*\*\* | 58.37\*\*\* | 49.58\*\*\* | 5.3 | -21.72\*\*\* | 317.66\*\*\* |
| SLM | 808.58\*\*\* | 9.17\*\*\* | 7.03\*\*\* | -0.29 | -2.01 | 15.81 |
| 2019 | LR | 16575.79\*\*\* | 22.30\* | 39.36\*\*\* | 12.92\* | -6.47 | 268.78\*\*\* |
| SLM | 750.11\*\*\* | 13.36\*\*\* | 5.25\* | 0.32 | 4.08 | 21.97 |

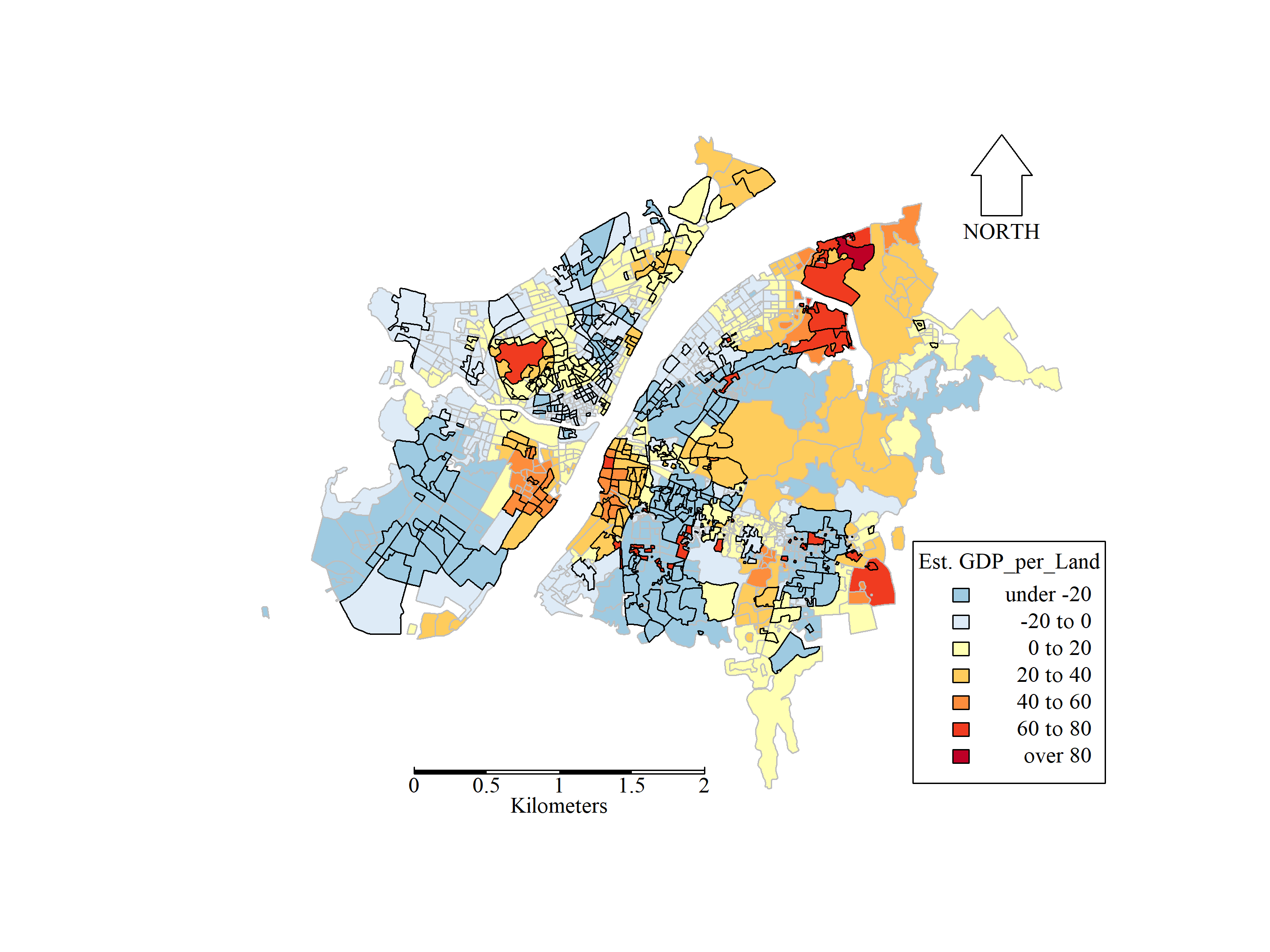
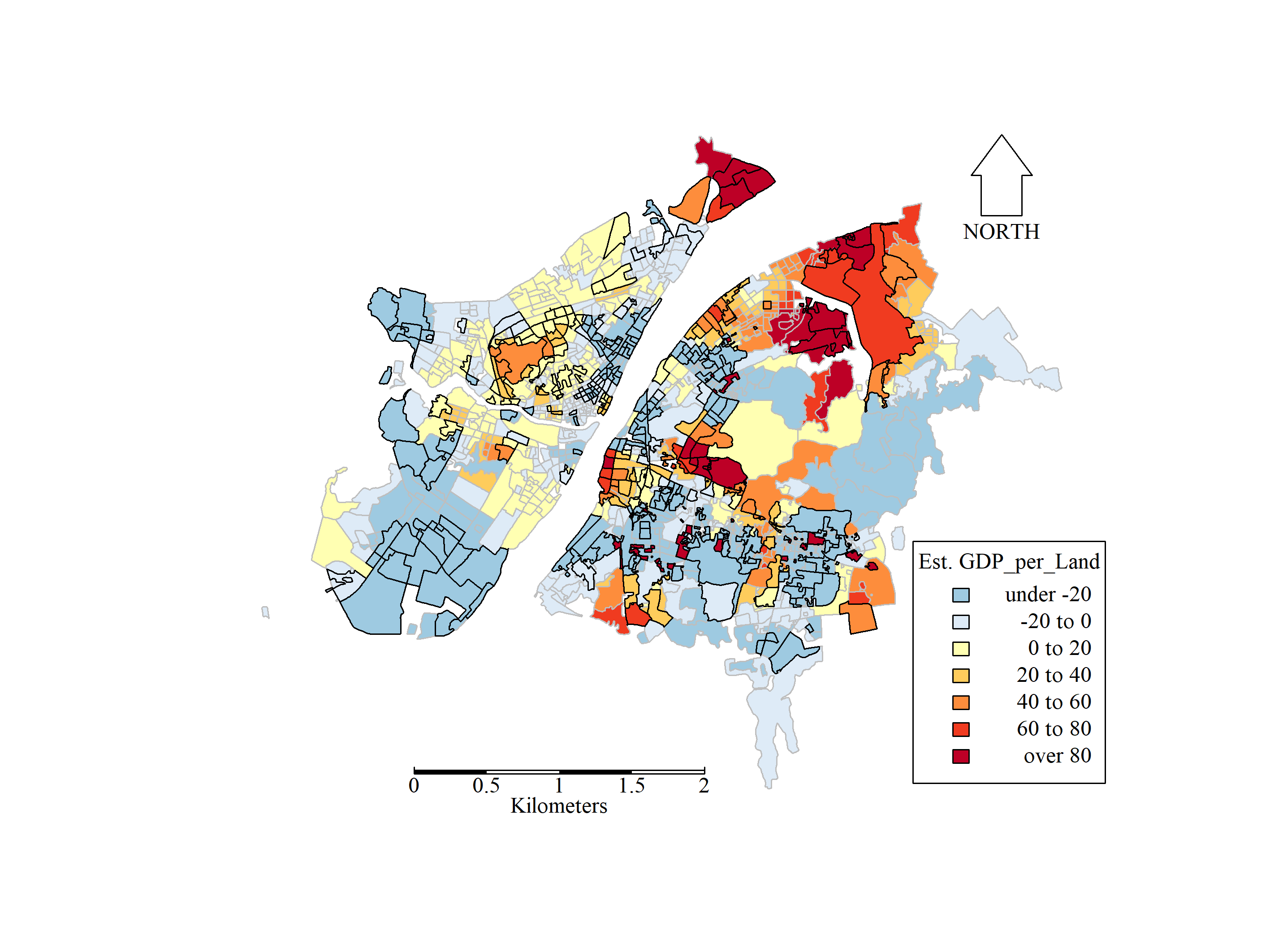
In Figure 5, the mapped coefficients for *Green\_Rate* in 2015 were very different to those estimated in 2019. This was a consequence of *Green\_Rate’s* relationship to house price being broadly the same across the CBA in 2015 (as the bandwidth is large at 398), but locally varying in 2019 (as bandwidth is small at 50). Furthermore, the coefficients were typically small (and close to zero) in 2015, while for some areas, relatively large in 2019. The local significance of this relationship also varied between years. Differences could be attributed to the 13th five-year (2016-2020) plan for the economic and social develoment of Wuhan, where the municipal government lanched the ‘green net’ project to increase green spaces for improving urban ecological quality. This initiative was specifically along the shorelines of Yangtze River, Han River, East Lake, South Lake – the same areas where *Green\_Rate* had impacted on (increased) house price.



(a) Coefficients for 2015 (b) Coefficients for 2019

**Figure 5**. Coefficient estimates for *Green\_Rate* for 2015 and 2019 MGWR models. Estimates are shown as significantly different from zero, at the 5% level, by highlighting the border of the corresponding areal units.

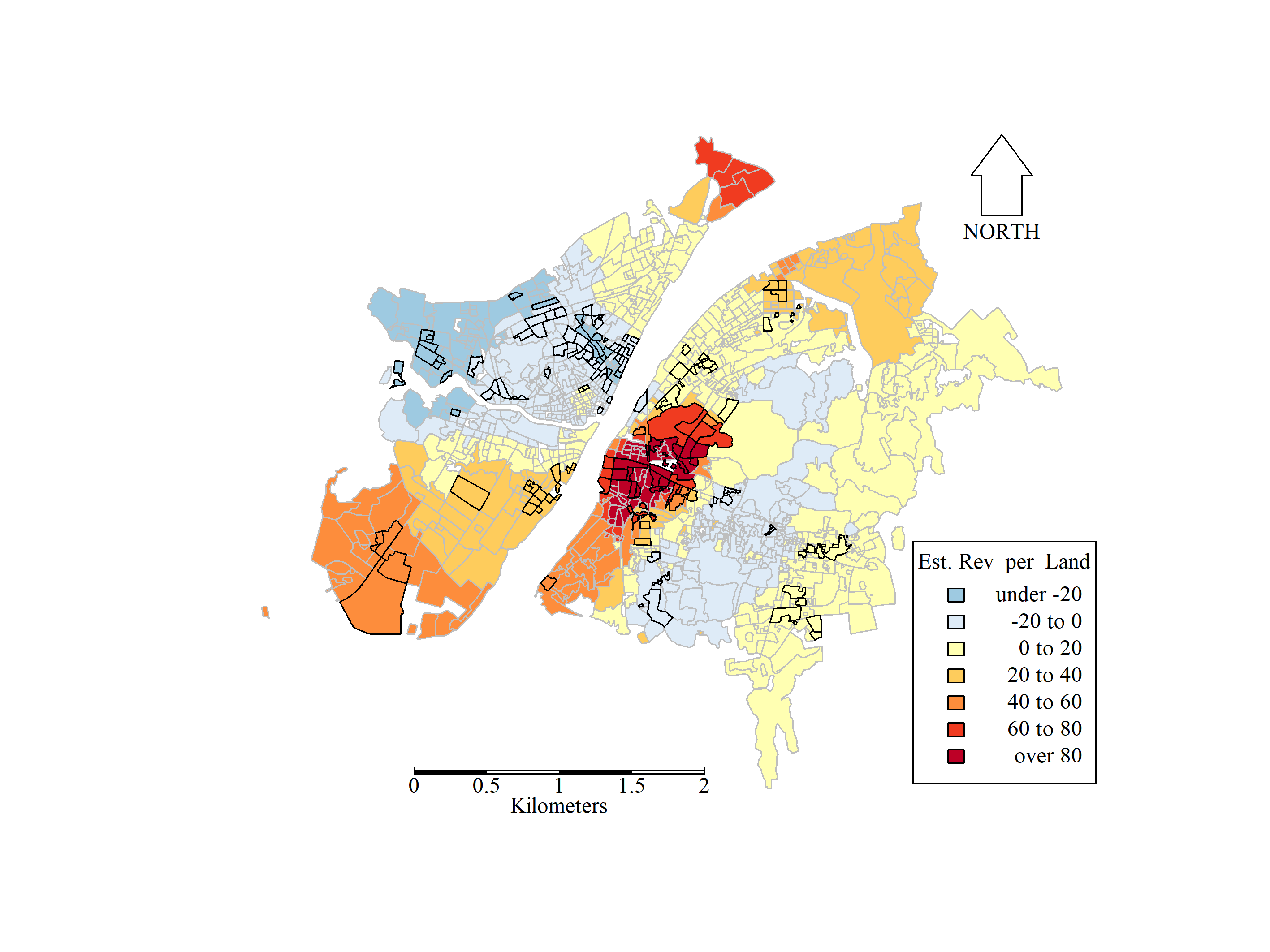
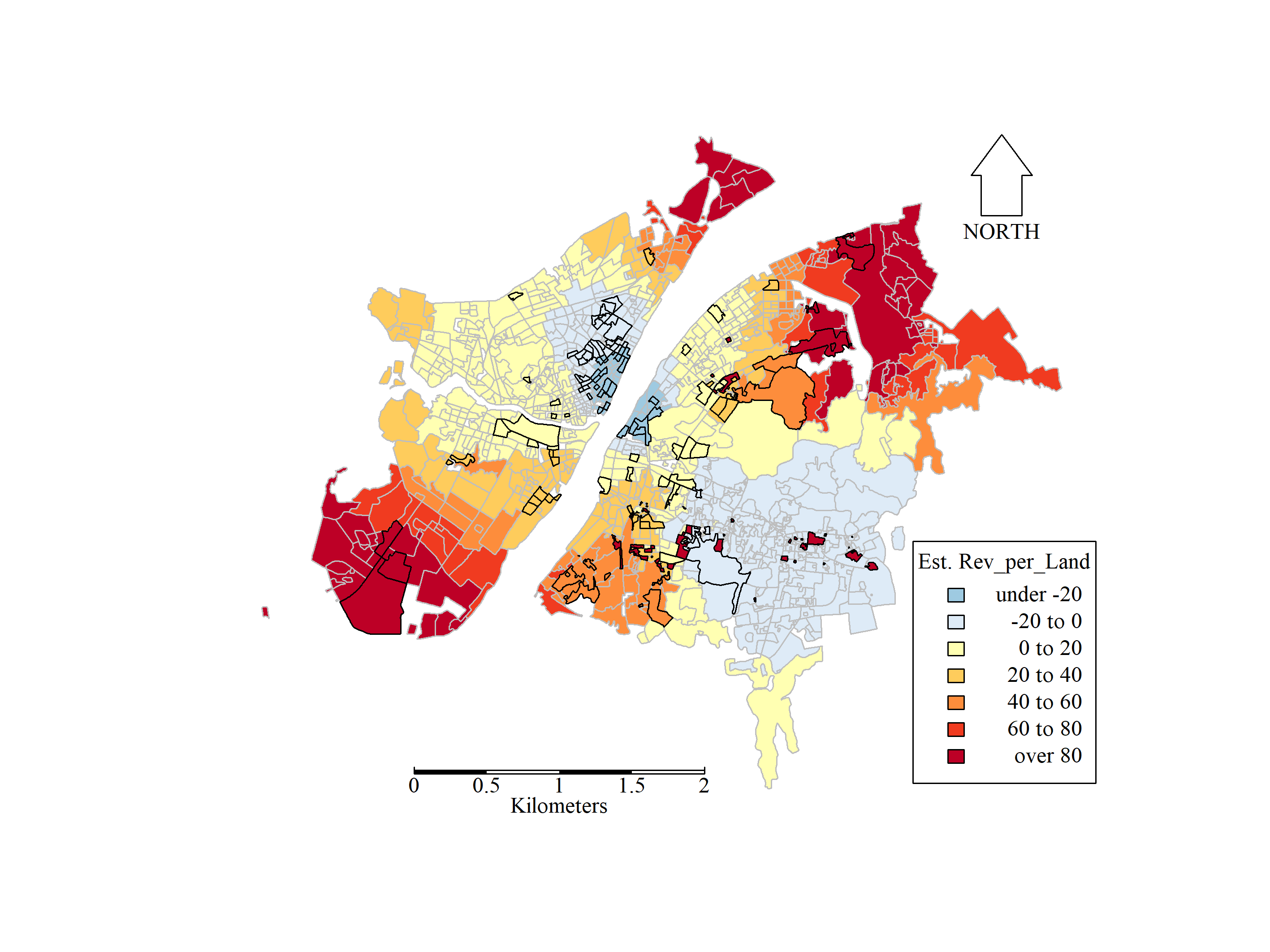
In Figure 6, the coefficients for *GDP\_per\_Land* are mapped, whose patterns tended to be driven by areas of industry and business. Clear localised patterns in this factor’s relationship to house price were found, as would be expected with small bandwidths of 22 and 24 nearest neighbours, for 2015 and 2019, respectively. However, these patterns changed between years. Interestingly, areas of both negative and positive coefficients were evident. For residential communities, relationships tended to be signigicantly negative, particularly in the old quarters of Wuhan (e.g. Jiangan, Jianghan and Wuchang districts) and in the marginal areas (e.g. Caidian and Hongshan districts), for both years. Relationships in the East Lake High-Tech Development Zone changed from negative in 2015 to postive in 2019, as a result of an influx of high-tech companies (e.g. Huawei and Xiaomi), which resulted in coupled increases in *GDP\_per\_Land* with house price.

.

(a) Coefficients for 2015 (b) Coefficients for 2019

**Figure 6**. Coefficient estimates for *GDP\_per\_Land* for 2015 and 2019 MGWR models. Estimates are shown as significantly different from zero, at the 5% level, by highlighting the border of the corresponding areal units.

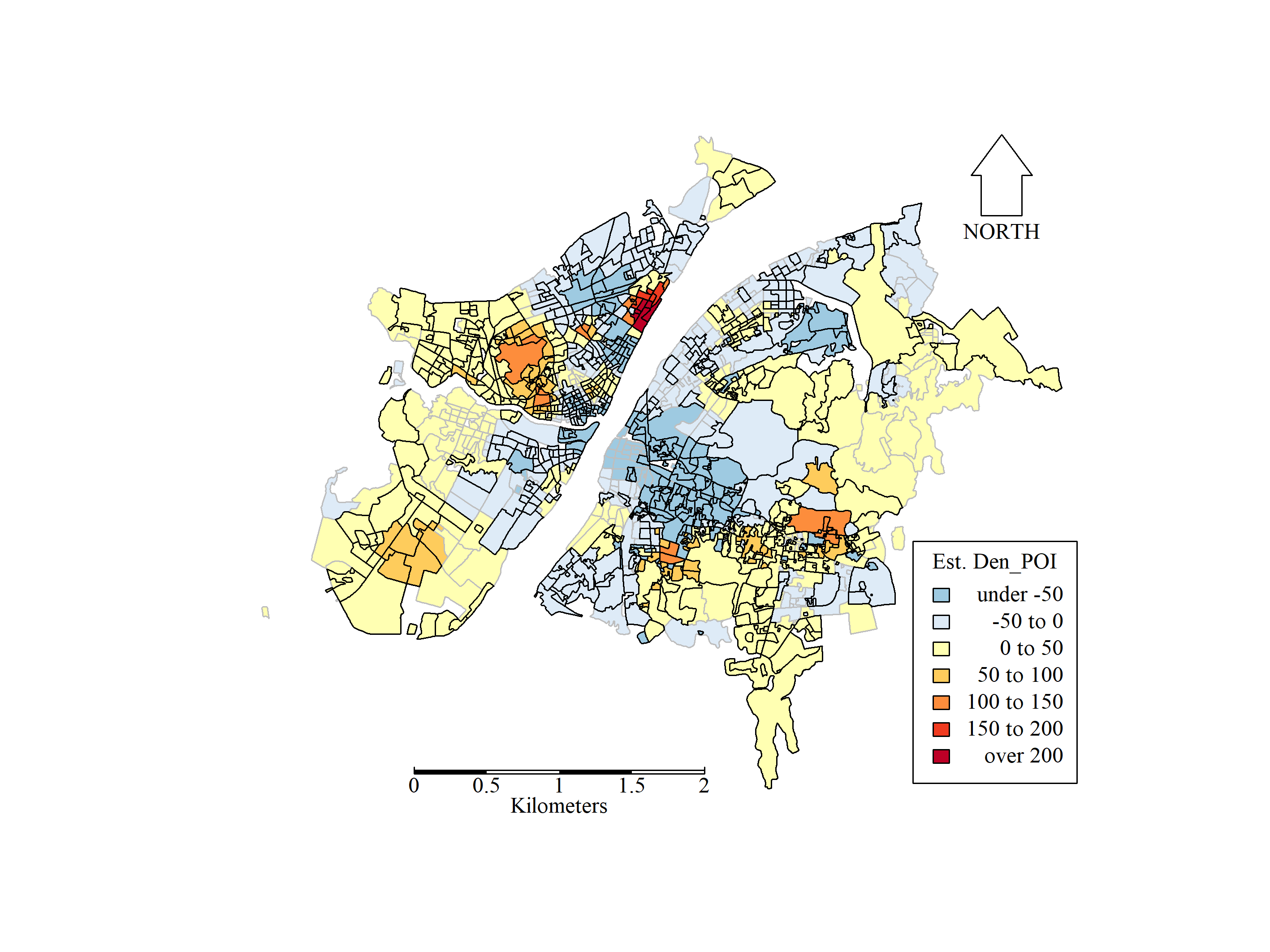
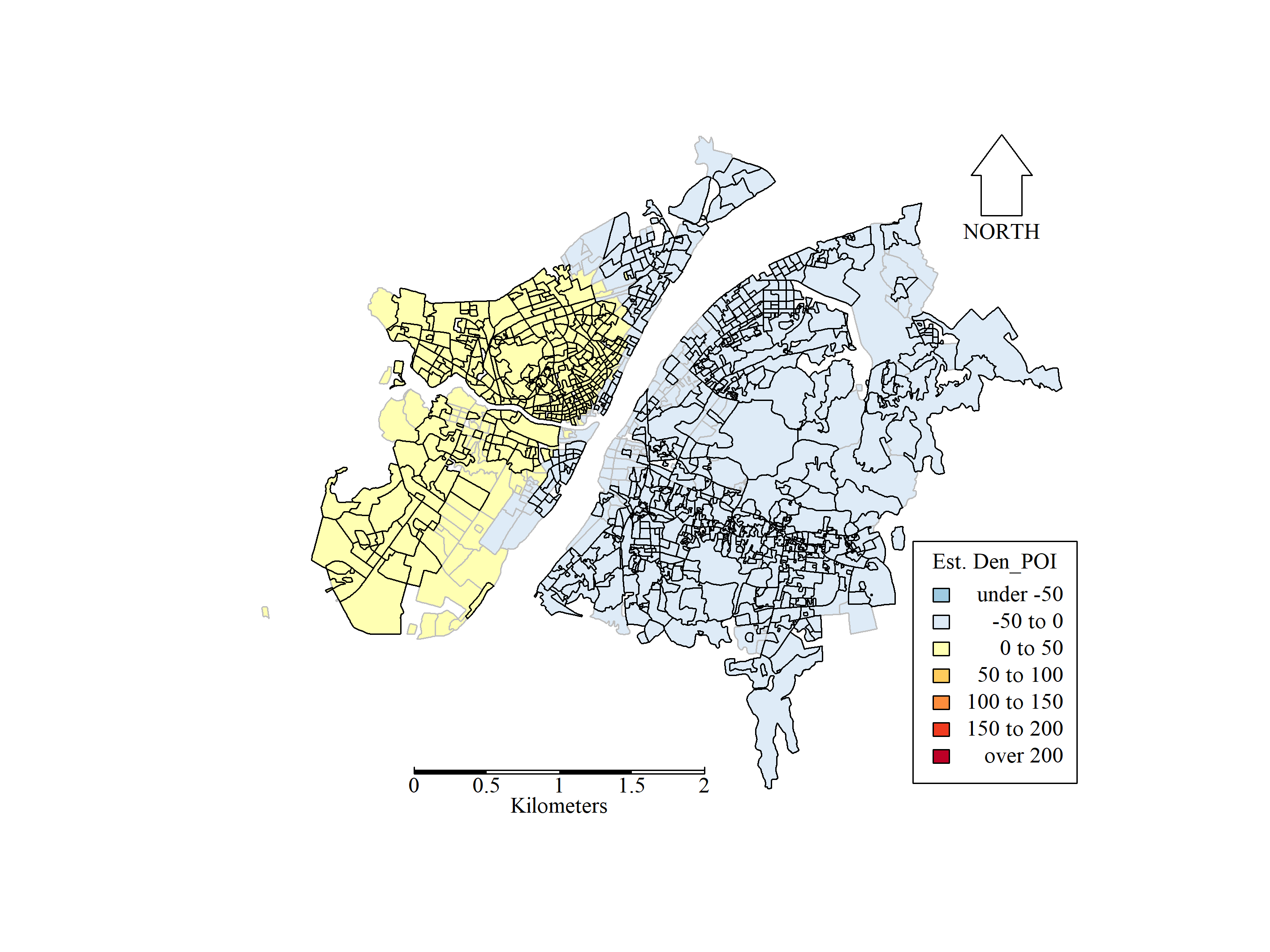
In Figure 7, the coefficients for *Rev\_per\_Land* are mapped, where more broad scale patterns were found for this factor’s relationship to house price. Relatively large bandwidths of 145 and 107, for 2015 and 2019, respectively, reflect these broad spatial trends. *Rev\_per\_Land* measures total financial revenues, inclduing tax revenue, budget receipt and public resource auctions. It reflects, to some extent, the ability of a local authority to invest in public affairs and facilities, and thus would be expected to have a postive relationship to house price. However, this relationship was not always positive and spatial patterns changed dramatically between the two years. Further, significant relationships tended to only occur in small spatial clusters. For example in 2015, significant positive coefficients could be found in the Qingshan district, which is characterised by two industrial clusters, i.e. Wuhan Iron & Steel and the Dongfeng Motor Group. For 2019, signicant postive coefficients could be observed in the adjacent area of Wuchang and Hongshan districts. These districts are characterised by universities, including Wuhan University, Huazhong University of Science and Technology, Central China Normal University, Wuhan University of Technology. House price increased rapidly in these districts through 2015 to 2019 due to extensive urban renovation projects.



(a) Coefficients for 2015 (b) Coefficients for 2019

**Figure 7**. Coefficient estimates for *Rev\_per\_Land* for 2015 and 2019 MGWR models. Estimates are shown as significantly different from zero, at the 5% level, by highlighting the border of the corresponding areal units.

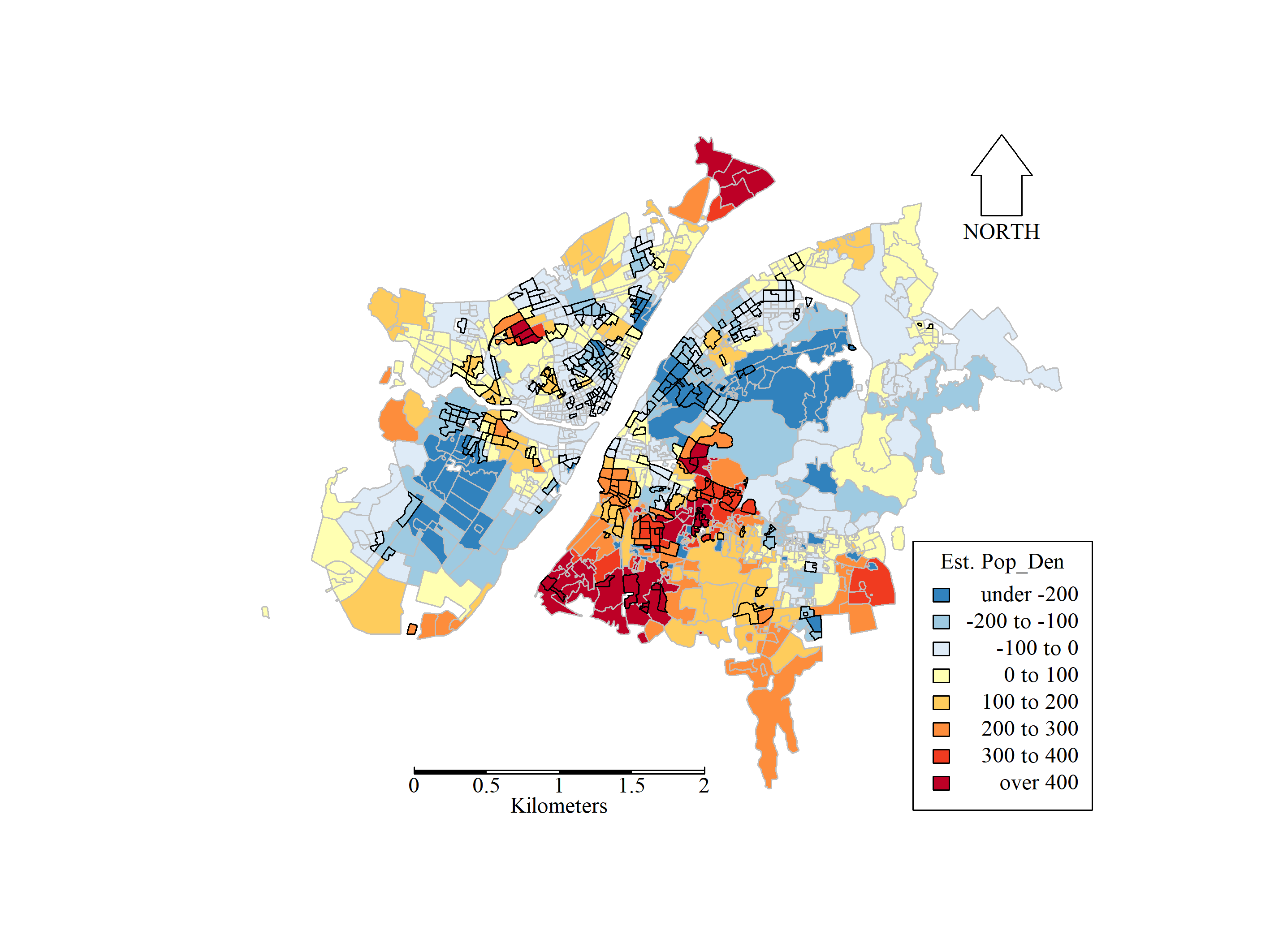
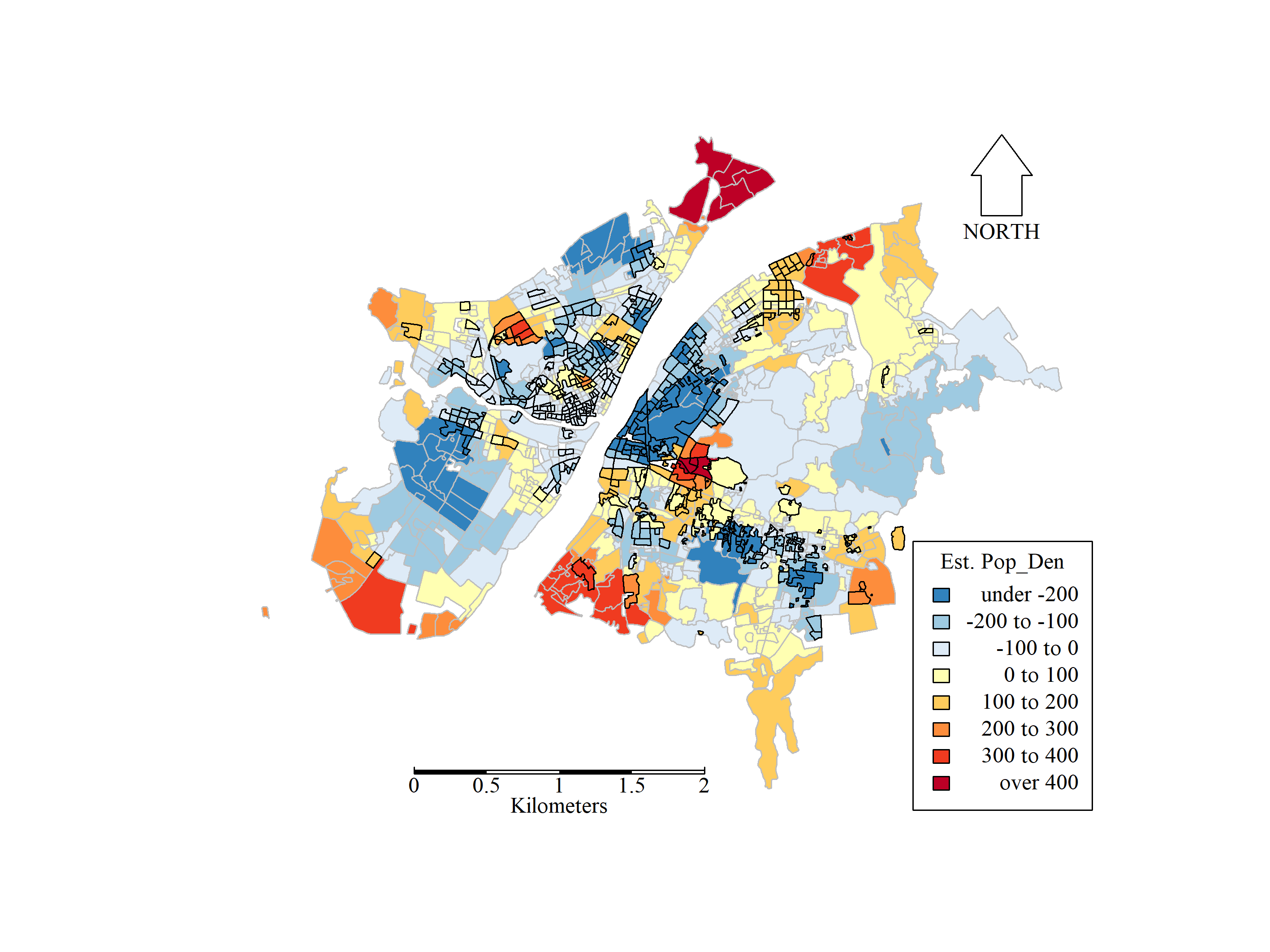
In Figure 8, the mapped coefficients for *Den\_POI* in 2015 are clearly different to those estimated in 2019. This was a consequence of *Den\_POI*’s relationship to house price being broadly the same across the CBA in 2015 (as the bandwidth is large at 328), but locally varying in 2019 (as bandwidth is only 27). In 2015, *Den\_POI*’s relationship to house price was weak with an east-west trend of negative to positive coefficient signs (divided by the Yangtze river), while in 2019 areas of strong relationships were apparent. In 2019, significant negative coefficients appeared in the old quarters of Jianghan, Wuchang, Hanyang and Qingshan districts; while significant positive coefficients appeared in new town and renovated areas (e.g., the East Lake High-Tech Development Zone) of Jiangan, Hongshan and Zhuankou districts. Thus, house price had a strong relationship to urban vitality in these areas.



(a) Coefficients for 2015 (b) Coefficients for 2019

**Figure 8**. Coefficient estimates for *Den\_POI* for 2015 and 2019 MGWR models. Estimates are shown as significantly different from zero, at the 5% level, by highlighting the border of the corresponding areal units.

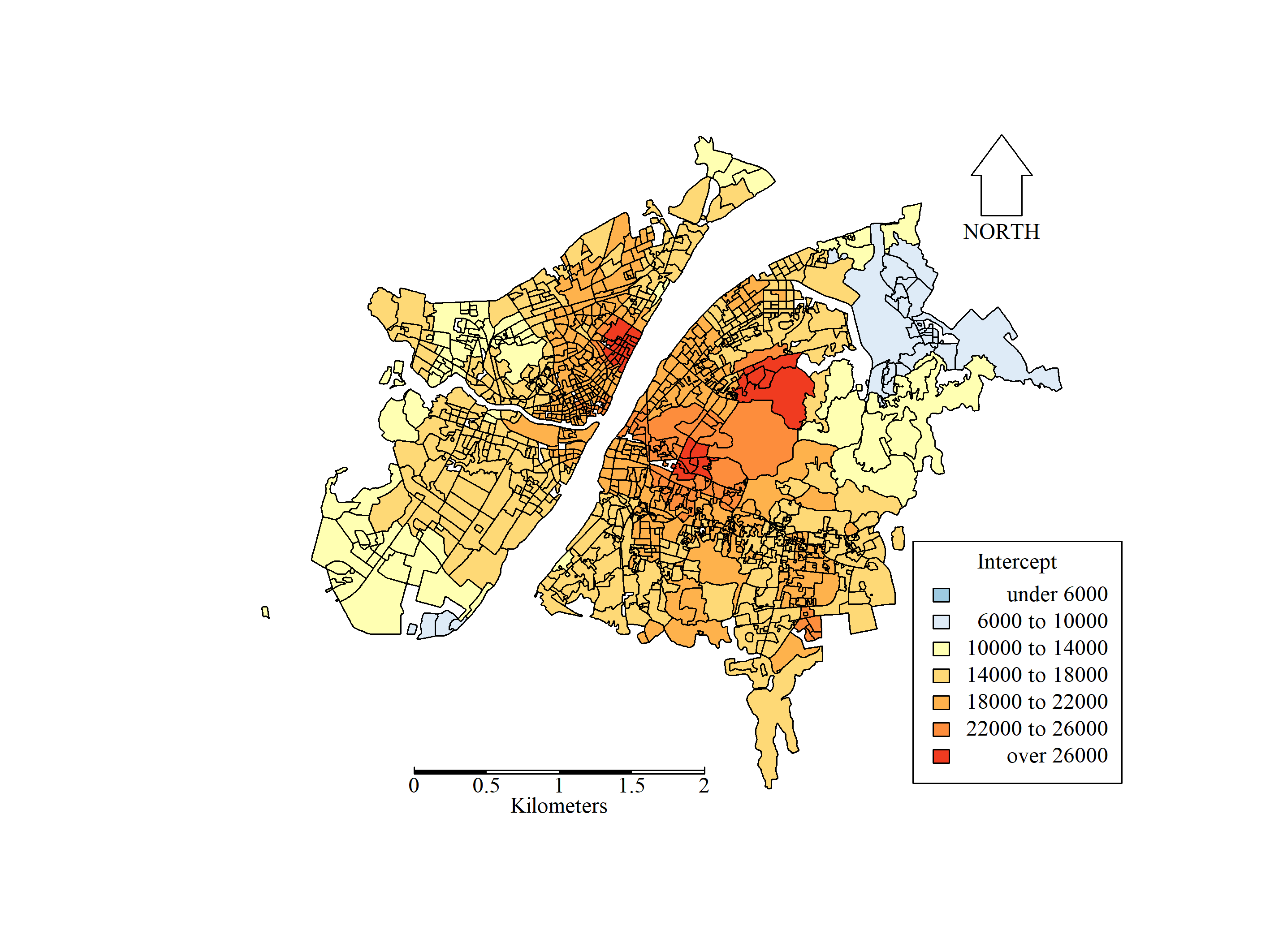
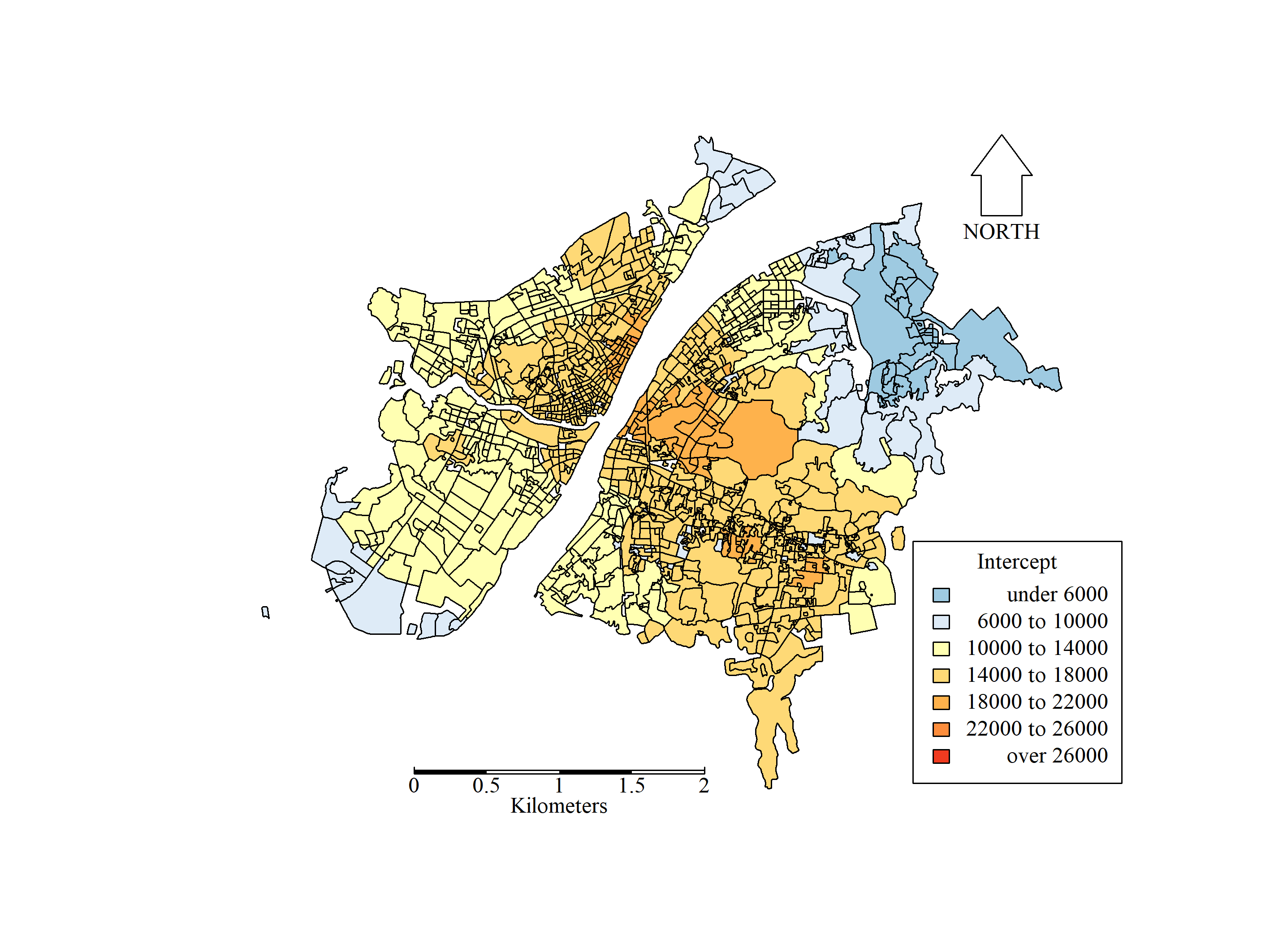
In Figure 9, coefficients for *Pop\_Den* are mapped. Clear localised relationships to house price were found, as would be expected with bandwidths of only 22 and 24, for 2015 and 2019, respectively. Patterns were similar between years, where significant negative coefficients arose in the old quarters of Wuchang, Hankou and Hanyang regions; while significant positve coefficients arose in new town and renovated areas, including East Lake High-Tech Development Zone and Changjiang New City. For the old quarters (houses typically built in the 1990s), residential concentrations exert pressures on resources, public facilities and environment, and so tend to drive house prices down.



(a) Coefficients for 2015 (b) Coefficients for 2019

**Figure 9**. Coefficient estimates for *Pop\_Den* for 2015 and 2019 MGWR models. Estimates are shown as significantly different from zero, at the 5% level, by highlighting the border of the corresponding areal units.

Estimates of the intercept term from MGWR, represent the impact of location itself on house price, where the intercept maps of Figure 10 are highly similar to the average house prices maps of Figure 2. This similarity was confirmed by strong positive correlation coefficients (*r*) between the *Intercept* estimates and house price (*Avg\_HP*) of *r* = 0.954 and *r* = 0.842, in 2015 and 2019, respectively. Thus, given the stronger correlation, location could be considered more important to house price variation in 2015, than that found in 2019.



(a) Interceptfor 2015 (b) Intercept for 2019

**Figure 10**. Estimates of *Intercept* for 2015 and 2019 MGWR models. Estimates are shown as significantly different from zero, at the 5% level, by highlighting the border of the corresponding areal units.

# 4 Discussion

The housing market in China is typically driven by habitation and asset portfolio (Liu and Xiong 2020), where the structural attributes of the house itself (e.g., floor area, the number of rooms, built date, house type; e.g., Lu et al. 2014a) indirectly relate to the socio-economic attributes of the community (e.g., public services, household incomes, job opportunities, occupations of the inhabitants), which for this study were represented, at the community-level, by the following five variables: *Pop\_Den, Green\_Rate, GDP\_per\_Land, Rev\_per\_Land, Den\_POI*.

For the five-year study period, and to this present day, China has undertook rapid urbanization, urban expansion and renewal, whose effects are evident at a daily resolution, particularly in ‘Tier 1’ and ‘Tier 2’ cities (Xu *et al.* 2019). However, these processes are not fixed across a given city, where they will vary spatially and temporally, as a consequence of socio-economic drivers, that in turn dynamically influence the house price market. For Wuhan, a mega city with a relatively large urban area and with a population of more than 12 million, such housing market dynamics were clearly evident through this study’s GWR and MGWR estimations and outputs.

The MGWR outputs indicated that relationships between community-level house price and expected drivers would vary at different spatial scales, and where these scales could change from that found in 2015 to that found in 2019. House price’s relationship to green space (*Green\_Rate*) and to places of interest (*Den\_POI*) moved from a broadly stationary, constant relationship in 2015, to a strongly non-stationary, locally-varying relationship in 2019. The significance of these relationships also varied between years. Differences could be attributed to the municipal government’s ‘green net’ project for increasing green spaces (and in turn, places of interest) for improving quality of urban living. Observe that data for places of interest are volunteered geographic information (VGI), so their reliability may vary across years.

House price’s relationship to GDP (*GDP\_per\_Land*) were strongly localised in both years, but where the spatial patterns in these relationships, differered between years. House price’s relationship to population density (*Pop\_Den*) were similarly strongly local in both years, but where the spatial patterns in these relationships were broadly similar between years. House price’s relationship to revenue (*Rev\_per\_Land*) were only moderately local in both years, but again the spatial patterns in these relationships, differered between years. Oberserve that the original spatial units of the explanatory variables may have an influnce on MGWR outputs, where *GDP\_per\_Land* and *Rev\_per\_Land* were initially reported at the district-level but assigned to each community using downscaling techniques for this study.

As only two time periods were considered (2015 and 2019), then repeated applications of a spatial model was considered appropriate. However future work with datasets that are also rich in the temporal dimension could consider the single application of a multiscale geographically and temporally weighted regression (Wu *et al.* 2019, Zhang *et al.* 2021).

# 5 Conclusions

Geocoded data provides opportunities to spatially explore housing markets where typically, individual property sales have been studied. However, difficulties arise when a city has a large number of apartments with the same geo-location. This situation is particularly true for most Chinese cities, where analyses need to be conducted at some aggregated scale. Here the community-level scale was considered ideal, as it’s a common administrative unit. Having decided on this core spatial unit, and following previous house price studies, a spatial regression model that captures non-stationary relationship between house price and drivers of house was considered appropriate. In this respect, multiscale geographically weighted regression and simpler models (linear regression, spatial lag model and basic geographically weighted regression) were applied to investigate two community-level housing market datasets, one from 2015, the other from 2019, for the city of Wuhan. Analyses were used to uncover diversities and similarities in the relationships between house price and socio-economic data in the context of increasingly volatile markets, moving from 2015 to 2019. The superiorities in the spatial regression models, SLM, GWR and MGWR over non-spatial LR largely corroborate the importance of concerning spatial dependence and heterogeneity in modelling house prices at a aggregated spatial unit, i.e. community-level in this study. Among them, the MGWR model provided the best performance by adopting individually optimized bandwidths for each coefficient (plus the intercept term) estimates. This technique could be a preferable choice for similar cases in spatial econometrics.

# References

Adair, A., Berry, J., Mcgreal, S., 1998. Assessing influences upon the housing market in northern ireland. Journal of Property Research15 (2), 121-134.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Petrov, B.N., Csaki, F. eds. 2nd international symposium on information theory*.* Akadémiai Kiadó, Budapest, Hungary, pp. 267-281.

Bajari, P., Chan, P., Krueger, D., Miller, D., 2010. A dynamic model of housing demand: Estimation and policy implications. NBER WORKING PAPER SERIES*.* NATIONAL BUREAU OF ECONOMIC RESEARCH, Cambridge.

Bian, L., Tang, Y., Shen, Z., 2021. Chinese urban planning and construction Springer, Cham, Switzerland AG.

Bitter, C., Mulligan, G., Dall'erba, S., 2007. Incorporating spatial variation in housing attribute prices: A comparison of geographically weighted regression and the spatial expansion method. Journal of Geographical Systems9 (1), 7–27.

Bivand, R., Piras, G., 2015. Comparing implementations of estimation methods for spatial econometrics. Journal of Statistical Software63 (18), 1 - 36.

Brunsdon, C., Fotheringham, A.S., Charlton, M., 1999. Some notes on parametric significance tests for geographically weighted regression. Journal of Regional Science39 (3), 497-524.

Brunsdon, C., Fotheringham, A.S., Charlton, M.E., 1996. Geographically weighted regression: A method for exploring spatial nonstationarity. Geographical Analysis28 (4), 281-298.

Can, A., 1992. Specification and estimation of hedonic housing price models. Regional Science and Urban Economics22 (3), 453-474.

Case, B., Clapp, J., Dubin, R., Rodriguez, M., 2004. Modeling spatial and temporal house price patterns: A comparison of four models. The Journal of Real Estate Finance and Economics29 (2), 167-191.

Chen, D.-R., Truong, K., 2012. Using multilevel modeling and geographically weighted regression to identify spatial variations in the relationship between place-level disadvantages and obesity in taiwan. Applied Geography32 (2), 737-745.

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., Lü, Y., Murakami, D., Nakaya, T., Wang, Y., Harris, P., 2022. A route map for successful applications of geographically weighted regression. Geographical Analysisn/a (n/a).

Credit, K., 2022. Spatial models or random forest? Evaluating the use of spatially explicit machine learning methods to predict employment density around new transit stations in los angeles. Geographical Analysis54 (1), 58-83.

Deng, M., Yang, W., Liu, Q., 2017. Geographically weighted extreme learning machine: A method for space–time prediction. Geographical Analysis49 (4), 433-450.

Elhorst, J.P., 2010. Applied spatial econometrics: Raising the bar. Spatial Economic Analysis5 (1), 9-28.

Esmaeili, M., Vancheri, A., Giordano, P., Year. Mathematical and computational modeling of housing market dynamics. In: Proceedings of the Systems Conference, 2010 4th Annual IEEE, pp. 29-34.

Farber, S., Páez, A., 2007. A systematic investigation of cross-validation in gwr model estimation: Empirical analysis and monte carlo simulations. Journal of Geographical Systems9 (4), 371-396-396.

Fik, T.J., Ling, D.C., Mulligan, G.F., 2003. Modeling spatial variation in housing prices: A variable interaction approach. Real Estate Economics31 (4), 623-646.

Fotheringham, A.S., Yang, W., Kang, W., 2017. Multiscale geographically weighted regression (mgwr). Annals of the American Association of Geographers107 (6), 1247-1265.

Gao, X., Asami, Y., 2005. Influence of spatial features on land and housing prices. Tsinghua Science & Technology10 (3), 344-353.

Garriga, C., Hedlund, A., Tang, Y., Wang, P., 2021. Rural-urban migration and house prices in china. Regional Science and Urban Economics91, 103613.

Gasparėnienė, L., Remeikienė, R., Skuka, A., 2016. Assessment of the impact of macroeconomic factors on housing price level: Lithuanian case. Intellectual Economics10 (2), 122-127.

Gauvin, L., Vignes, A., Nadal, J.-P., 2011. Modeling urban housing market dynamics: Can the socio-spatial segregation preserve some social diversity? Physics and Society.

Gelfand, A.E., Banerjee, S., Sirmans, C.F., Tu, Y., Eng Ong, S., 2007. Multilevel modeling using spatial processes: Application to the singapore housing market. Computational Statistics &amp; Data Analysis51 (7), 3567-3579.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., Harris, P., 2015. Gwmodel: An r package for exploring spatial heterogeneity using geographically weighted models. Journal of Statistical Software63 (17), 1-50.

Hagenauer, J., Omrani, H., Helbich, M., 2019. Assessing the performance of 38 machine learning models: The case of land consumption rates in bavaria, germany. International Journal of Geographical Information Science33 (7), 1399-1419.

Harris, P., 2019. A simulation study on specifying a regression model for spatial data: Choosing between autocorrelation and heterogeneity effects. Geographical Analysis51 (2), 151-181.

Hastie, T., Tibshirani, R., 1986. Generalized additive models. Statistical Science1 (3), 297-310.

Holly, S., Pesaran, M.H., Yamagata, T., 2006. A spatio-temporal model of house prices in the us. Faculty of Economics, University of Cambridge.

Hu, Y., Lu, B., Ge, Y., Dong, G., 2022. Uncovering spatial heterogeneity in real estate prices via combined hierarchical linear model and geographically weighted regression. Environment and Planning B: Urban Analytics and City Scienceonline, 1-26.

Hui, E.C.-M., Chen, T., Lang, W., Ou, Y., 2021. Urban community regeneration and community vitality revitalization through participatory planning in china. Cities110, 103072.

Jiang, Y., Wang, Y., 2021. Price dynamics of china’s housing market and government intervention. Applied Economics53 (10), 1212-1224.

Jim, C.Y., Chen, W.Y., 2006. Impacts of urban environmental elements on residential housing prices in guangzhou (china). Landscape and Urban Planning78 (4), 422-434.

Kang, Y., Zhang, F., Peng, W., Gao, S., Rao, J., Duarte, F., Ratti, C., 2021. Understanding house price appreciation using multi-source big geo-data and machine learning. Land Use Policy111, 104919.

Kestens, Y., Thériault, M., Rosiers, F.D., 2006. Heterogeneity in hedonic modelling of house prices: Looking at buyers’ household profiles. Journal of Geographical Systems8 (1), 61-96.

Kirby, A.M., 1976. Housing market studies: A critical review. Transactions of the Institute of British Geographers1 (1), 2-9.

Lau, M.H.M., Wei, X., 2018. Housing size and housing market dynamics: The case of micro-flats in hong kong. Land Use Policy78, 278-286.

Leishman, C., Bramley, G., 2005. A local housing market model with spatial interaction and land-use planning controls. Environment and Planning A37 (9), 1637-1649.

Leong, Y.-Y., Yue, J.C., 2017. A modification to geographically weighted regression. International Journal of Health Geographics16 (1), 11.

Limsombunchai, V., 2004. House price prediction: Hedonic price model vs. Artificial neural network. American Journal of Applied Sciences1, 193-201.

Liu, C., Xiong, W., 2020. China's real estate market. In: Amstad, M., Sun, G., Xiong, W. eds. The handbook of china's financial system*.* Princeton University Press, Princeton, New Jersey.

Lu, B., Brunsdon, C., Charlton, M., Harris, P., 2017. Geographically weighted regression with parameter-specific distance metrics. International Journal of Geographical Information Science31 (5), 982-998.

Lu, B., Brunsdon, C., Charlton, M., Harris, P., 2019. A response to ‘a comment on geographically weighted regression with parameter-specific distance metrics’. International Journal of Geographical Information Science33 (7), 1300-1312.

Lu, B., Charlton, M., Harris, P., Fotheringham, A.S., 2014a. Geographically weighted regression with a non-euclidean distance metric: A case study using hedonic house price data. International Journal of Geographical Information Science28 (4), 660-681.

Lu, B., Harris, P., Charlton, M., Brunsdon, C., 2014b. The gwmodel r package: Further topics for exploring spatial heterogeneity using geographically weighted models. Geo-spatial Information Science17 (2), 85-101.

Lu, B., Yang, W., Ge, Y., Harris, P., 2018. Improvements to the calibration of a geographically weighted regression with parameter-specific distance metrics and bandwidths. Computers, Environment and Urban Systems71, 41-57.

Madsen, J.B., 2011. A q model of house prices Department of Economics, Monash University, Melbourne, Australia.

Maurice, J.R., 1999. Irish house prices - will the roof cave in? The Economic and Social Review30 (4), 343-362.

Meen, D., Meen, G., 2003. Social behaviour as a basis for modelling the urban housing market: A review. Urban Studies40 (5-6), 917-935.

Nagaraja, C.H., Brown, L.D., Zhao, L.H., 2011. An autoregressive approach to house price modeling. Annals of Applied Statistics5 (1), 124-149.

Nanthakumaran, N., Watkins, C., Orr, A., 2000. Understanding property market dynamics: Insights from modelling the supply-side adjustment mechanism. Environment and Planning A32 (4), 655-671.

Orford, S., 2000. Modelling spatial structures in local housing market dynamics: A multilevel perspective. Urban Studies37 (9), 1643-1671.

Ou, Y., Zheng, S., Nam, K.-M., 2021. Impacts of air pollution on urban housing prices in china. Journal of Housing and the Built Environment.

Páez, A., Fei, L., Farber, S., 2008. Moving window approaches for hedonic price estimation: An empirical comparison of modelling techniques. Urban Studies45 (8), 1565-1581.

Qian, X., Qiu, S., Zhang, G., 2021. The impact of covid-19 on housing price: Evidence from china. Finance Research Letters43, 101944.

Simlai, P., 2014. Estimation of variance of housing prices using spatial conditional heteroskedasticity (sarch) model with an application to boston housing price data. The Quarterly Review of Economics and Finance54 (1), 17-30.

Wang, X.-R., Hui, E.C.-M., Sun, J.-X., 2017. Population migration, urbanization and housing prices: Evidence from the cities in china. Habitat International66, 49-56.

Wei, Y., Cao, Y., 2017. Forecasting house prices using dynamic model averaging approach: Evidence from china. Economic Modelling61, 147-155.

Wu, C., Ren, F., Hu, W., Du, Q., 2019. Multiscale geographically and temporally weighted regression: Exploring the spatiotemporal determinants of housing prices. International Journal of Geographical Information Science33 (3), 489-511.

Xu, G., Jiao, L., Yuan, M., Dong, T., Zhang, B., Du, C., 2019. How does urban population density decline over time? An exponential model for chinese cities with international comparisons. Landscape and Urban Planning183, 59-67.

Yang, W., 2014. An extension of geographically weighted regression with flexible bandwidths. St Andrews.

Yao, J., Fotheringham, A.S., 2016. Local spatiotemporal modeling of house prices: A mixed model approach. The Professional Geographer68 (2), 189-201.

Yazdani, M., 2021. Machine learning, deep learning, and hedonic methods for real estate price prediction. Colorado University at Boulder, arXiv, pp. 1-33.

You, J., Ding, X., Niño-Zarazúa, M., Wang, S., 2021. The intergenerational impact of house prices on education: Evidence from china. Journal of Housing Economics54, 101788.

Yu, D., 2007. Modeling owner-occupied single-family house values in the city of milwaukee: A geographically weighted regression approach. GIScience & Remote Sensing44 (3), 267-282.

Zhang, L., Hui, E.C.-M., Wen, H., 2015. Housing price–volume dynamics under the regulation policy: Difference between chinese coastal and inland cities. Habitat International47, 29-40.

Zhang, Z., Li, J., Fung, T., Yu, H., Mei, C., Leung, Y., Zhou, Y., 2021. Multiscale geographically and temporally weighted regression with a unilateral temporal weighting scheme and its application in the analysis of spatiotemporal characteristics of house prices in beijing. International Journal of Geographical Information Science35 (11), 2262-2286.