

Conditionally Autoregressive Models for House Price Data: Insights From a Comparative Simulation Study

Indira Puteri Kinasih^{1,2}, Haziq Jamil^{1*}, Elvynna Leong^{1*}

¹Mathematical Sciences, Faculty of Science, Universiti Brunei Darussalam, Universiti Brunei Darussalam, Jalan Tungku Link, Bandar Seri Begawan, BE 1410, Brunei.

²Mathematics Education, Faculty of Education, Universitas Islam Negeri Mataram.

*Corresponding author(s). E-mail(s): haziq.jamil@ubd.edu.bn;
elvynna.leong@ubd.edu.bn;

Contributing authors: indiraputeri@uinmataram.ac.id;

Abstract

The modelling of property prices has been extensively studied in econometrics, with widely used approaches including generalised linear regression and geographically weighted regression. These models commonly address local spatial correlations observed in property price data. However, despite its potential to capture spatial effects, the Conditional Autoregressive (CAR) model remains underutilised for this purpose. This study examines the robustness and predictive power of the CAR model, comparing it with established spatial models across three different datasets generation. An illustrative case study on Lombok house price data is also included. Simulation results showed that the CAR model demonstrates a distinct advantage, achieving lower bias and variability compared to other spatial regression models, effectively capturing neighbourhood-based spatial relationships, and exhibiting strong predictive power across various scenarios. In the Lombok case study, the CAR model outperformed other models, providing more precise estimates for property-related factors such as land size and built-up area. The results confirm that CAR's

spatial framework enables a nuanced analysis of property values across regions, enhancing accuracy in predictive models. This study also reveals the distinct strengths and limitations of each model, offering insights into their predictive accuracy and applicability across diverse real estate contexts.

Keywords: Conditional auto-regressive, simultaneous auto-regressive, geographically weighted regression, Property prices, Neighbourhood structure

1 Introduction

Real estate valuation represents a complex challenge, requiring a nuanced understanding of spatial dynamics and interdependencies within property markets. Traditional valuation methods often overlook these spatial dimensions, resulting in incomplete predictions and less effective policy interventions (Pagourtzi et al. (2003), Droj, Kwartnik-Pruc, and Droj (2024), McCord et al. (2014)). In response, spatial regression models have emerged as powerful tools to address these shortcomings by explicitly incorporating spatial relationships into the analysis of property prices (Stewart Fotheringham and Park (2018), Yang et al. (2019), Soltani et al. (2021)).

Among the various spatial regression models, Geographically Weighted Regression (GWR) is particularly prominent in property price research. Known for its capacity to capture local variations in property prices, GWR addresses spatial heterogeneity by allowing coefficients to vary across different locations. Significant studies, such as those by Sisman and Aydinoglu (2022), Soltani et al. (2021), Brunsdon, Fotheringham, and Charlton (1996), Yu (2007), and Lu, Charlton, and Fotheringham (2011), have highlighted its effectiveness in revealing spatially varying relationships between property values and various explanatory factors, including structural and neighbourhood characteristics, locational attributes, and socio-economic variables.

Spatial Autoregressive (SAR) models account for spatial dependence in the dependent variable by including a spatial lag parameter (ρ), which measures the influence of neighboring values. This makes SAR well-suited for data where values are directly impacted by nearby areas or points, such as house prices influenced by surrounding properties. However, SAR relies on a global spatial structure, which can limit its ability to capture localised patterns. Additionally, according to Golgher and Voss (2016) and LeSage and Pace (2014) the interdependence created by the spatial lag term introduces feedback effects, where values influence each other in a looping manner, making the interpretation of coefficients more complex.

Conditional autoregressive (CAR) models, in contrast, are designed to handle more localised spatial dependencies effectively, compare to SAR. They assume values at a location are conditionally dependent on neighboring areas, defined through an adjacency matrix \mathbf{W} . This approach models spatial dependence via a spatial random effect, ϕ , which captures the influence of neighboring areas on the value at a specific location (Banerjee, Carlin, and Gelfand (2014)). CAR models are particularly suited for

areal data, such as aggregated district-level statistics, where spatial autocorrelation is prominent. Unlike GWR, CAR does not require exact point coordinates and provides better handling of spatial heterogeneity than SAR. Its ability to model finer localised effects through ϕ makes CAR a preferred choice for applications like district-level property price analysis, public health studies, or crime mapping (De Oliveira (2012)).

This study is motivated by the need to address the spatial intricacies that characterise real estate markets. Properties located near each other often share similar price trends, influenced by common amenities, and neighbourhood attributes. Incorporating spatial dependencies into property valuation allows spatial regression models to provide a deeper insight into market dynamics than traditional methods. It explores the theoretical basis, methodological structures, and practical uses of the GWR, SAR, and CAR models within house price modelling. Through a comparative analysis of these models, we aim to clarify their respective advantages, limitations, and appropriateness for enhancing the accuracy and detail of property market assessments, with a particular focus on the distinct context of Lombok, Indonesia.

The paper begins with a literature review on spatial regression techniques for modeling property prices. Section 3 outlines the theoretical foundations of spatial autoregressive models, including GLM, GWR, SAR, CAR, and their multilevel variants. Section 4 provides a comparative analysis of these models across three artificial study regions, evaluating their predictive accuracy and robustness. Section 5 applies these models to Lombok house price data, demonstrating their efficacy in capturing spatial patterns and improving real estate market insights.

2 Related works

Recent studies have leveraged GWR to investigate the spatial heterogeneity of housing market determinants. For example, Wang and Chen (2020) applied GWR to examine the impact of local built-environment factors on home prices across different phases of the housing market cycle. Their findings highlighted significant spatial variations, underscoring the critical role of local context in influencing home prices. Similarly, Lu, Charlton, and Fotheringham (2011) employed GWR with a non-Euclidean distance metric to analyze London house price data. By incorporating this alternative metric, they better captured the spatially varying relationships between house prices and their determinants, emphasizing the importance of considering different distance metrics in spatial analyses. In their 2018 study, Stewart Fotheringham and Park (2018) explored spatial and temporal variations in the determinants of apartment prices in Seoul, South Korea, over a decade. Utilizing a hedonic price model with a spatio-temporal lag, calibrated through GWR, they revealed consistent spatial variations and strong spatial lag effects on house prices over time. Additionally, Cellmer, Cichulska, and Belej (2020) applied GWR to assess the spatial variations in determinants of the housing market in Poland. The study identified spatial patterns of autocorrelation in average housing prices and market activity, illustrating the value of GWR in examining localized effects of socio-demographic and environmental factors on the housing market. Collectively, these studies demonstrate the efficacy of GWR and its extensions in capturing the

nuanced spatial and temporal variations in housing market data, providing valuable insights for urban planning and policy-making.

In parallel, numerous studies have demonstrated the versatility and applicability of SAR models across various fields. For instance, Fix, Cooley, and Thibaud (2021) explored the use of SAR models for analyzing spatial extremes, highlighting their ability to capture spatial dependence structures in extreme values. Sarlas and Axhausen (2015) showcased the effectiveness of SAR models in transportation research by predicting localized speed variations, illustrating their utility in improving traffic flow predictions. Beguería and Pueyo (2009) emphasized the advantages of SAR models in addressing spatial autocorrelation by comparing them with generalized least squares models, providing more accurate predictions in ecological and biogeographical studies.

In the realm of real estate, SAR models have been particularly valuable. Trojanek and Gluszak (2018) used SAR models to analyze the spatial and temporal effects of subway availability on property prices in Warsaw. Their study revealed significant price premiums for properties located near subway stations, underscoring the importance of transportation infrastructure on real estate values. Bottero et al. (2017) applied SAR models to investigate the relationship between buildings' energy performance and real estate market value. Their findings illustrated the model's utility in linking environmental attributes to economic outcomes in the property market. Furthermore, Cellmer, Kobylińska, and Belej (2019) demonstrated the practical relevance of SAR models in urban planning and property valuation by employing hierarchical SAR models to develop detailed land value maps in urban areas, underscoring the model's effectiveness in capturing the complexities of urban land valuation.

Despite their limited use in property price modelling, CAR models offer a promising approach for addressing spatial autocorrelation and generating reliable estimates, making them well-suited for future studies in house price analysis. Lee (2013) introduced the R package CARBayes, which supports Bayesian spatial modelling using conditional autoregressive priors. By including house price data as a case study, Lee demonstrated the flexibility of CAR models in capturing localised spatial dependencies relevant to real estate valuation. Wall (2004) provided a comparative analysis of the spatial structures in CAR and SAR models, highlighting CAR's strength in handling localised spatial interactions through an adjacency-based framework. While their work focused on educational and environmental applications, the insights are directly applicable to real estate modelling. Similarly, Ver Hoef et al. (2018) examined CAR models in ecological studies, emphasizing their robustness in managing spatial heterogeneity and clarifying their structural differences from SAR models. By addressing the limitations of SAR's global spatial framework and excelling in the detection of localised patterns, CAR models emerge as a robust and adaptable methodology for accurately modelling spatial dynamics in property prices, advancing the precision of real estate valuation techniques.

3 Study region

In the spatial modelling, the study region \mathcal{S} is typically divided into K distinct non-overlapping geographic units, denoted as \mathcal{S}_k for $k = 1, 2, \dots, K$. Each geographic unit \mathcal{S}_k is associated with a target variable y_k and a set of explanatory factors represented as a vector \mathbf{x}_k .

In this study, the artificial study region was constructed to facilitate model simulations and analysis. This synthetic region serves as a spatial representation. The dataset consists of a simple feature (**sf**) collection. It spans a bounding box from longitude 116.0355 to 116.5 and latitude -8.500079 to -8.000057, using the WGS 84 coordinate reference system (CRS). The geometry type is polygon, indicating that each feature is a polygonal area, with 216 polygons in total. Each polygon is defined by two attributes: **area**, representing the size in square units, and **geometry**, specifying the polygon's boundaries with longitude and latitude coordinates. The artificial study region is illustrated in Figure 1.

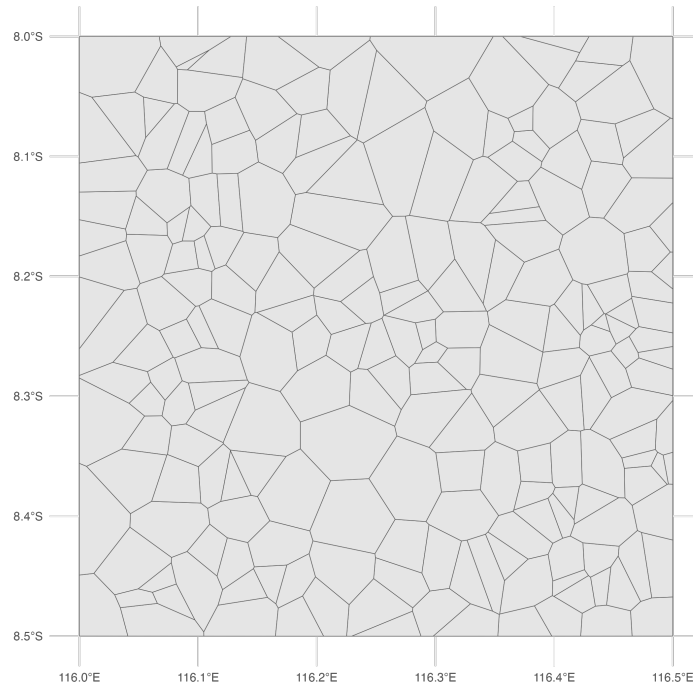


Figure 1: Artificial study region used for simulation analysis, consisting of 216 non-overlapping polygonal areas. Each area represents a distinct spatial unit for modelling and analysis, providing a controlled environment to evaluate spatial models' performance

Source: [Article Notebook](#)

Another specific ingredient in spatial modelling is the existence of W matrix, also known as the spatial weights matrix. It encodes the spatial relationships between areas in a study region. The weight matrix visually shows how different regions are connected to each other, indicating their neighbourhood relationships (Morris et al. (2019)). To demonstrate this concept, let's take a simple example of a map with four regions (R_1, R_2, R_3, R_4), as visualised in Figure 2.



Figure 2: Illustration of neighbourhood structure. The figure depicts a simplified spatial configuration where each numbered area represents a distinct spatial unit, demonstrating how neighbouring relationships can be defined for spatial modelling purposes.

Source: [Article Notebook](#)

This straightforward neighbourhood structure creates a single connected component with the following relationships: ($R_1 \sim R_2$, $R_2 \sim R_3$, $R_3 \sim R_4$). Therefore the weight matrix W is

	R_1	R_2	R_3	R_4
R_1	0	1	0	0
R_2	1	0	1	0
R_3	0	1	0	1
R_4	0	0	1	0

Later on, in the modelling stage, we will also need to construct a diagonal matrix D . It is an $N \times N$ matrix where each diagonal element d_{ii} denotes the number of neighbours of region R_i , while all non-diagonal elements are zero. This matrix is essential in spatial econometric models because it captures the local neighbourhood structure of each region. In the context of Geographically Weighted Regression (GWR), Spatial Autoregressive (SAR), and Conditional Autoregressive (CAR) models, D plays a pivotal role in defining spatial relationships and dependencies among different regions.

For instance, in SAR and CAR models, the diagonal matrix D aids in incorporating spatial lag and error components by appropriately weighting the influence of neighbouring regions (Wall (2004), Ver Hoef et al. (2018)). In GWR, this matrix assists in locally calibrating the model by reflecting the density and connectivity of regions (Brunsdon, Fotheringham, and Charlton (1996), Stewart Fotheringham and Park (2018)). Here, D is represented as:

	R_1	R_2	R_3	R_4
R_1	1	0	0	0
R_2	0	2	0	0
R_3	0	0	2	0
R_4	0	0	0	1

This matrix format indicates that region R_1 has one neighbour, regions R_2 and R_3 each have two neighbours, and region R_4 has one neighbour. By integrating this matrix into our modelling framework, we ensure that the spatial structure and interactions among regions are accurately represented, thereby enhancing the robustness and precision of our spatial analysis.

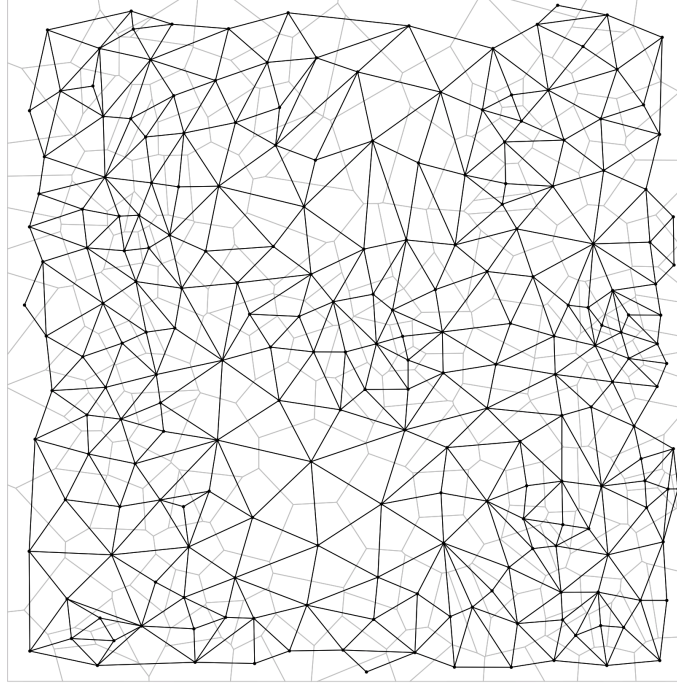


Figure 3: Neighbourhood structure of the artificial region. This figure illustrates the spatial connectivity between polygons in the artificial study region, where each line represents a defined neighbour relationship based on spatial adjacency

Source: [Article Notebook](#)

In relation to our artificial study region, Figure 3 illustrates the connections between areas, represented by vertices or centroids for each area and nodes connecting them to each other. This visualisation highlights the spatial adjacency structure within the region.

4 Spatial regression model

When discussing spatial regression, it's crucial to comprehend the basic notion of linear regression. In classical linear regression, the relationship between the dependent variable y and the x_1, x_2, \dots, x_p independent variables is expressed as

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \varepsilon \quad (1)$$

This formulation assumes a global relationship between the variables, where the coefficients $\beta_1, \beta_2, \dots, \beta_p$ are constant across the entire study area. In many spatial datasets, relationships between variables may exhibit spatial variation. For example, in the case of property pricing, a consistent rates of change assumption may not hold true universally. For example, the house price increase for an extra bedroom is often thought to be fixed across all locations. However, local customs or knowledge may actually dictate these rates, rather than a universal utility assigned to each commodity. For instance, in neighbourhoods with families, where extra space is highly valued, the perceived value of an additional bedroom may be greater compared to areas with singles or elderly couples, for whom extra space may not be as desirable.

4.1 Geographically Weighted Regression (GWR)

Brunsdon, Fotheringham, and Charlton (1996) developed GWR which is one such technique that extends classical linear regression by allowing coefficients to vary spatially. It allows for the estimation of local relationships between a response variable and predictor variables. It is particularly useful for exploring spatial non-stationarity and identifying spatially varying relationships across different locations. The GWR model can be expressed as:

$$y_i(s) = \beta_0(s) + \sum_{k=1}^p \beta_k(s) x_{ik}(s) + \varepsilon(s), i = 1, \dots, n \quad (2)$$

Equation 2 represents a spatial regression model where $\mathbf{y}(s)$ is the dependent variable at location s . The term $\beta_0(s)$ is the spatially varying intercept, while $\sum_{k=1}^p \beta_k(s) \mathbf{x}_k(s)$ represents the spatially varying coefficients for the independent variables $\mathbf{x}_k(s)$. $\varepsilon(s)$ denotes the error term, capturing unexplained variation at location s .

Using a weighted least squares approach to calibrate regression models allows different weights to be assigned to different observations, influencing the estimated parameters. In ordinary least squares, the goal is to minimize the sum of squared differences between predicted and actual y values. Weighted least squares, however, apply a weighting factor w to each squared difference, making some prediction inaccuracies more significant. If w is a diagonal matrix of weights, the estimated coefficients are given by Equation 3:

$$\beta(s) = (x^T W(s) x)^{-1} x^T W(s) y \quad (3)$$

This method allows GWR to address spatial heterogeneity by emphasizing observations near the location of interest, thereby improving the accuracy and relevance of local model estimates.

The estimation of GWR parameters involves fitting a separate regression equation for each location in the study area. Various estimation techniques can be employed, including ordinary least squares (OLS), weighted least squares (WLS), and maximum likelihood estimation (MLE). These techniques aim to optimize the model parameters to minimize the differences between the observed and predicted values of the dependent variable.

4.2 Simultaneously Autoregressive (SAR) Models

While GWR focuses on capturing localised spatial heterogeneity by allowing coefficients to vary across space, SAR take a different approach by explicitly modelling spatial dependencies through a spatial lag. The SAR model is a spatial econometric model used to analyse spatial dependencies and relationships among observations in a geographic space (Anselin and Griffith (1988)). It is widely employed in various fields, such as regional economics, environmental studies, and urban planning. SAR is a type of spatial autoregressive model involving a simultaneous equation framework to capture spatial interactions.

The general form of SAR model can be expressed as follows

$$y(s) = \rho \sum_{s'} w(s, s') y(s') + \sum_{k=1}^p \beta_k x_k(s) + \varepsilon(s) \quad (4)$$

where \mathbf{Y} is the vector of observed values for the dependent variable, W is the spatial weights matrix, ρ is the spatial autoregressive coefficient, X is the matrix of observed values for exogenous variables, β is the vector of coefficients, and $\varepsilon \sim N(0, \sigma^2)$ is the vector of error terms. Estimation of the SAR model parameters is typically done using statistical techniques such as maximum likelihood estimation (MLE) or generalised method of moments (GMM). The joint distribution of \mathbf{Y} can be written as

$$\mathbf{Y} \sim \mathcal{N} \left((I - \rho W)^{-1} X \beta, \sigma^2 (I - \rho W)^{-1} (I - \rho W^T)^{-1} \right) \quad (5)$$

Extensions of the SAR model include the Spatial Lag Model, the Spatial Error Model, and the Spatial Durbin Model (Elhorst et al. (2014)), each incorporating distinct assumptions regarding the spatial configuration of the errors (Elhorst, Lacombe, and Piras (2012); Anselin (2013)). These extensions provide flexibility to account for different types of spatial dependencies and can be chosen based on the specific spatial relationships and hypotheses under investigation.

The SAR model, with its various specifications, provides a flexible framework to account for spatial dependencies and explore the spatial dynamics of observed phenomena. Researchers often choose between these models based on the nature of the

spatial relationships in their data and the specific hypotheses they aim to test. The SAR model is a valuable tool for understanding spatial interdependence and making informed policy and planning decisions in diverse spatial contexts

4.3 Conditional Autoregressive (CAR) Models

CAR models are a class of spatial statistical models used to analyze spatially structured data. The general formulation of a CAR model can be expressed as:

$$y(s) = \sum_{k=1}^p \beta_k x_k(s) + \varepsilon(s) + \phi(s) \quad (6)$$

Here, $\varepsilon \sim N(0, \nu^2)$ while ϕ is a specific component in CAR model that has a role as spatial effect. It is also common to mention ϕ as a CAR priors. It is a type of Gaussian Markov random field (Rue and Held (2005)), capture spatial autocorrelation by ensuring that values at nearby locations are more similar than those further apart (Lee (2013)). This can be expressed in a general term

$$\phi \sim N(0, \tau^2 Q^{-1}) \quad (7)$$

where Q is a precision matrix that may be singular (intrinsic model). Q controls the spatial autocorrelation structure of the random effects, and is based on a non-negative symmetric $n \times n$ neighbourhood or weight matrix W .

Together with the spatial weights matrix W , the prior information are crucial components of CAR models. The choice of W determines the spatial structure of the model, while the priors for the variance parameters and the spatial random effects influence the model's ability to capture spatial dependencies. Commonly used priors for the variance parameters include inverse-gamma distributions, which provide flexibility and can be tuned to reflect prior beliefs about the scale of variability in the data. The prevailing approach typically involves a binary representation using geographical adjacency, where $w_{ki} = 1$ if areal units (S_k, S_i) have a mutual boundary (denoted $k \sim i$, and is zero otherwise). This specification forces (ϕ_k, ϕ_i) relating to geographically adjacent areas (that is $w_{ki} = 1$) to be correlated. On the other hand, random effects associated with areas that are not adjacent are independent of each other, provided we know the values of the other random effects.

A CAR prior was introduced by Leroux, Lei, and Breslow (2000) and Stern and Cressie (1999) to model diverse levels of spatial autocorrelation. In this type of prior, a single collection of random effects is utilised and its primary objective is to model spatial data, with a specific focus on dealing with spatial relationships and auto-correlation between data points. This model is especially proficient in the task of smoothing data and detecting spatial patterns within data set. The random effect ϕ_k structured as follow:

$$\phi_k | \phi_{-K}, W, \rho, \tau^2 \sim N \left(\frac{\rho \sum_{i=1}^K w_{ki} \phi_i}{A}, \frac{\tau^2}{A} \right) \quad (8)$$

where $A = \rho \sum_{i=1}^K w_{ki} + 1 - \rho$. Note that when $\rho = 1$, the prior forms an intrinsic CAR prior (Besag, York, and Mollié 1991), indicating full spatial dependency. Conversely, when $\rho = 0$, $A = 1$ it means that there will be no W matrix role in the model, and it will become a common random effect. In other words, the model reduces to a generalised linear model.

When handling i observations within each area k , Equation 9 closely resembles Equation 2 in the GWR model, with the exception of the ϕ component. It is usually called multilevel CAR models

$$y_i(s) = \sum_{k=1}^p \beta_k x_{ki}(s) + \varepsilon_i(s) + \phi(s) \quad (9)$$

The Bayesian approach to CAR models entails defining prior distributions for all model parameters, such as the regression coefficients β , variance parameters ν^2 and τ^2 , and spatial random effects ϕ_k . Subsequently, Bayesian inference techniques like Markov Chain Monte Carlo (MCMC) are employed to derive posterior distributions of these parameters. This method facilitates the integration of prior knowledge and offers a versatile framework for quantifying uncertainty.

In general, Table 1 below summarises the key elements utilised in the GWR, SAR, and CAR models:

Table 1: Comparison of Model Elements Across GWR, SAR, and CAR

Elements	<i>GWR</i>	<i>SAR</i>	<i>CAR</i>
Number of Observations (N)	✓	✓	✓
Covariate Matrix (X)	✓	✓	✓
Spatial Weights Matrix (W)	×	✓	✓
Spatial Coordinates (coords)	✓	✓	✓
Covariate Coefficients (β)	✓	✓	✓
Error Term variance (σ^2)	✓	✓	✓
Spatial Autocorrelation (ρ)	×	✓	✓
Spatial Structure component (ϕ)	×	×	✓
Spatial structure variance (τ^2)	×	×	✓
Additional scale (ν^2)	×	×	✓
Estimation method	MLE	MLE/GMM	MCMC

Source: [Article Notebook](#)

5 Comparative Analysis of GLM, SAR, and CAR

This section focuses on generating spatially autocorrelated property price data, exploring two distinct data generation frameworks: one using a CAR specification and another without spatial autocorrelation (non-CAR). The analysis is conducted within an artificially constructed study area using key property covariates: land size, building size, number of bedrooms, and number of bathrooms. The generated datasets are then employed to examine spatial dependencies in property prices and are analysed using a suite of models including GLM, SAR, and CAR.

5.1 The experiment design

The experiment begins by creating a dataset based on the defined study region. The idea is to generate covariates that hypothetically can explain price of a property. It basically consist of property structural characteristics such as land size, building size, number of bedrooms, and number of bathrooms. To represent real-world data, each polygon contains multiple observations.

To generate the covariates, the first step involves creating area-level data with a specific spatial pattern, referred to as the housing density for the study region. The spatial pattern implies that certain areas or polygons will have a higher housing density than others. In this case, the central horizontal region of the study area is set to has a higher housing density, mirroring the housing distribution found on the island of Lombok. Although this density will not be directly utilized in the model simulation, it serves as a critical foundation for determining the spatial distribution of observation points across each polygon. In areas with higher density, a greater number of observation points are generated, which aligns with the notion that densely populated regions typically have more residential developments, thereby increasing the number of properties available for sale. Across the 216 polygons, a total of 4,314 observation points were generated. The number of observations per polygon varies from 5 to 56 points, reflecting the density in each area. Figure 4 illustrate how each polygon has multiple observations and they are distributed according to a specific pattern.

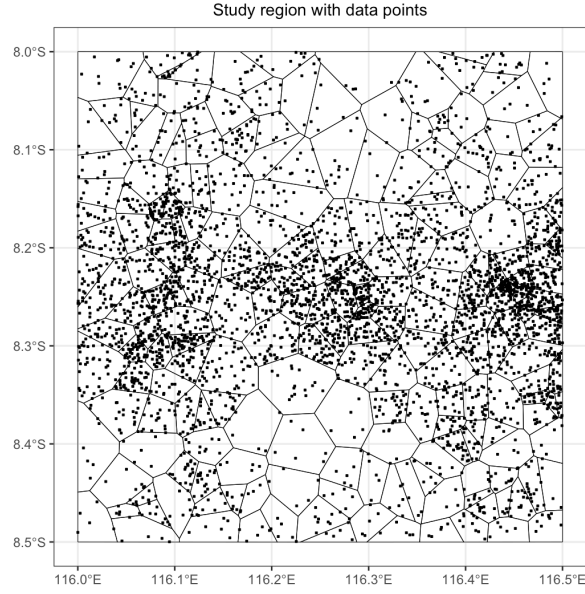


Figure 4: Artificial study region with observation points. The points represent data locations within each polygon, simulating real-world spatial observations for the purpose of model analysis and validation.

Source: [Article Notebook](#)

Following this, covariates were generated using `mvrnorm()` for each observation point, creating a dataset for further analysis. Table 2 provides a brief slice of the covariates' structure.

Table 2: Slice of generated dataset as covariates for simulations

ID	land_size	building_size	#bedrooms	#bathrooms
1	270.48	108.55	4	2
1	295.99	100.13	4	2
2	319.52	126.31	3	1
2	325.01	120.84	4	3
2	294.15	144.05	4	3
3	222.18	81.40	3	2
4	414.23	150.15	2	2
4	278.75	101.41	2	1

Source: [Article Notebook](#)

By employing artificially generated covariates and an artificially constructed study region, we developed three distinct property price datasets, each based on a different

spatial model: CAR, a combination of CAR and SAR, and GWR. The true parameter values for this purpose are $\beta = [1 \ 0.9 \ 0.7 \ 0.5 \ 0.3]$, $\nu^2 = 0.2$, $\tau^2 = 0.6$ and $\rho = 0.6$. Whereas beta values for GWR dataset is generated using a beta function, which is a function of coordinate points. The matrix \mathbf{W} was derived from the neighbourhood structure of an artificial study region using the functions `poly2nb()` and `nb2mat()`. Data generation for *SAR + CAR* employs a slightly different strategy. In addition to being generated per area simultaneously using the joint distribution of \mathbf{Y} as in Equation 5, information about the spatial effect ϕ is also added at the end of the process. Therefore, we refer to it as *SAR + CAR* data.

Each data generation resulting two types of datasets: area-level and point-level. These datasets were then modelled using six linear regression approaches, comprising both classical and spatial regression models: GLM, SAR, and CAR for area-level data, and GWR, GLMM, and multilevel CAR for point-level data. Simulations were conducted $sim = 1000$ times. From each simulation, coefficients, parameters, and fitted values were extracted, allowing for the computation of bias and RMSE values for robustness and prediction power analysis.

5.2 Robustness and power prediction analysis

The bias values show that the models exhibit varying levels of accuracy across the different parameters and datasets. The coefficients β_k for $k = 0, 1, 2, 3, 4$ represent the values associated with the intercept, land size, building size, number of bedrooms, and number of bathrooms, respectively. For β_0 , the GLM and CAR models display relatively small biases with narrow standard deviations, indicating better consistency, while the SAR model demonstrates larger bias values, suggesting potential issues with accurately estimating the intercept. For higher parameters like β_2 and β_3 , SAR generally has larger bias and variability, especially in the *SAR + CAR* dataset, suggesting that the combined spatial structure poses challenges for SAR's estimation. It is important to note that the Figure 5 utilises area-level data. In this context, the CAR model consistently demonstrates the smallest bias compared to the other two models.

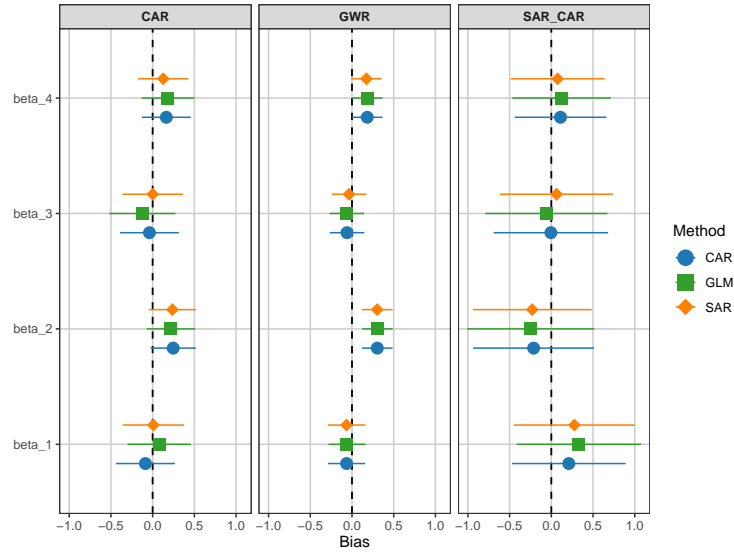


Figure 5: Bias values by method and dataset type for area-level data. This figure compares the bias estimates of different methods (CAR, GLM, SAR) across datasets generated under CAR, GWR, and SAR+CAR frameworks. Each method's performance is visualised for four coefficients ($\beta_1, \beta_2, \beta_3, \beta_4$), highlighting variations in bias and the effectiveness of each model in addressing spatial dependencies.

Source: [Article Notebook](#)

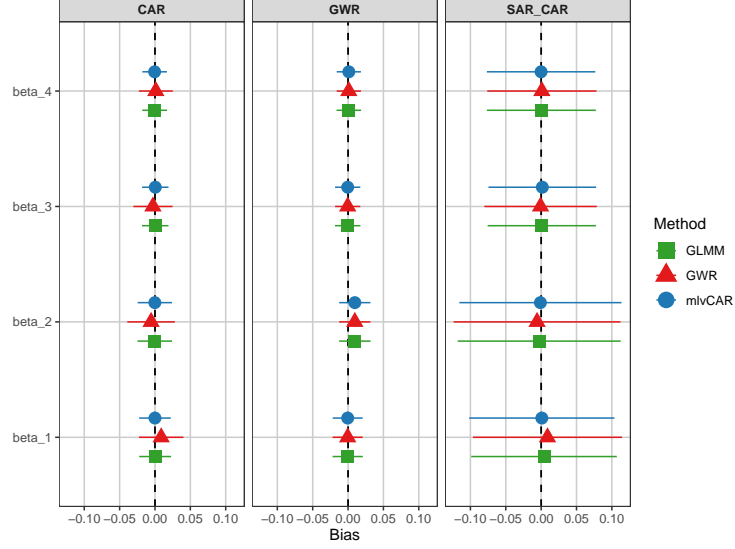


Figure 6: Bias values by method and dataset type for point-level data. This figure compares the bias estimates of different methods (GLMM, GWR, and mlvCAR) across datasets generated under CAR, GWR, and SAR+CAR frameworks. Each method's performance is visualised for four coefficients (β_1 , β_2 , β_3 , β_4), highlighting variations in bias and the effectiveness of each model in addressing spatial dependencies.

Source: [Article Notebook](#)

As shown in the second plot of Figure 6, this analysis utilises point-level data. Overall, the mlvCAR model demonstrates a consistently lower bias across most parameters and scenarios. In the *SAR + CAR* scenario, where spatial complexity is most pronounced, mlvCAR achieves notably low bias, particularly for parameters β_1 through β_4 , indicating its suitability for handling mixed spatial structures. making it a solid choice for a range of spatial modelling tasks.

The GWR model, which provides highly localised estimates, displays slightly higher bias for several parameters compared to mlvCAR, especially for β_1 , β_2 , and β_3 within the *CAR* scenario. This suggests GWR's greater sensitivity to spatial patterns without incorporating neighbourhood-based effects. In contrast, GLMM, which lacks explicit spatial components, shows larger bias variability, particularly in scenarios with pronounced spatial dependencies, highlighting its limitations for purely spatial analyses.

Table 3: RMSE values across different datasets

Dataset	Area-level data			Point-level data		
	<i>GLM</i>	<i>SAR</i>	<i>CAR</i>	<i>GLMM</i>	<i>GWR</i>	<i>mlvCAR</i>
<i>CAR</i>	4.09	3.85	1.81	14.04	21.13	9.94
<i>SAR + CAR</i>	11.88	11.68	3.88	47.67	48.87	38.97
<i>GWR</i>	2.43	2.42	2.28	18.31	18.33	14.87

Source: [Article Notebook](#)

Table 3 provide insights into model accuracy at different spatial levels. For area-level data, the models GLM, SAR, and CAR are compared, with the CAR model consistently achieving the lowest RMSE. This suggests that CAR is more effective at capturing spatial dependencies at the area level. For point-level data, the models GLMM, GWR, and mlvCAR are evaluated, with mlvCAR showing the lowest RMSE across all datasets. This indicates that mlvCAR is particularly well-suited for handling detailed spatial variations at the point level. Overall, these results highlight the strengths of CAR and mlvCAR in reducing error within their respective spatial contexts, underscoring their potential utility in spatial modeling applications.

6 Applications in Lombok house prices dataset

In this section, we apply the simulated models to property price data on Lombok Island, Indonesia. Lombok is located in the eastern part of Indonesia, specifically between latitudes -8.9° and -8.1° , and longitudes 115.9° and 116.6° .

area, number of bedrooms, number of bathrooms, floors, property type/category, and furnishing status.

Further, we conducted several preprocessing steps. First, we filtered the dataset by removing irrelevant variables, including floors, furnishing status, and category. It was then subsequently filtered to retain only properties with plausible characteristics: land area between 90 and 800 square meters, built-up size between 70 and 600 square meters, a maximum of 6 bedrooms, and a maximum 5 bathrooms. These thresholds were applied to ensure the data reflect realistic and context-appropriate housing characteristics in Lombok, based on common residential patterns and local housing norms. This step was essential to enhance the validity of the analysis by excluding outliers or potentially erroneous entries. Following this procedure, the dataset was refined to comprise a total of 576 observations. A summary of the key variables is presented in Table 4.

Table 4: Summary statistics of variables in the property price dataset

Variable	Mean	Median	Q_1	Q_3	Min	Max	SD	NA
Prices (million IDR)	1983.00	1400	800.0	2500.0	170	22500	2056.81	26
Land-size (sqm)	299.60	255	147.2	407.2	90	800	179.39	24
Built-up area (sqm)	184.80	150	100.0	210.0	70	600	109.20	24
No. of bedroom	3.24	3	2.0	4.0	1	6	1.19	24
No. of bathroom	2.55	2	2.0	3.0	1	5	1.05	24

Source: [Article Notebook](#)

Next, we merged the cleaned dataset with spatial administrative data by matching sub-district and district names. This process resulted in a dataset containing 598 entries. Compared to the previous dataset, which consisted of 576 observations, this indicates that 22 sub-districts did not have any property data available—meaning that property sales in those areas were not recorded on the online property listing platform used as the data source.

To address missing values, we employed the `mice` package (Van Buuren and Groothuis-Oudshoorn (2011)) to perform multiple imputation with parameters set to $maxit = 10$, $m = 10$, and a fixed random seed to ensure reproducibility. The imputed values were then reintegrated into the main dataset, replacing the original missing entries. This process resulted in a final dataset that was complete and ready for subsequent analysis using the selected spatial models.

In addition to the point-level dataset (`ppd`), which contains individual property listings, the imputed dataset was also aggregated to the sub-district level by calculating the mean values of key variables. This aggregation produced an area-level dataset (`ppd_kc`), allowing the analysis to be conducted at both the individual and administrative levels. These two levels of data granularity provide complementary perspectives for evaluating the performance of spatial models in capturing local property market dynamics.

Table 5: Point-level Model Comparison

	GLMM	GWR	mlvCAR
Coefficients			
Intercept	7.09 [6.93, 7.24]*	7.25 [7.25, 7.26]*	7.08 [6.99, 7.17]*
Land-size	0.26 [0.21, 0.32]*	0.31 [0.30, 0.32]*	0.26 [0.21, 0.32]*
Built-up area	0.16 [0.10, 0.22]*	0.19 [0.19, 0.20]*	0.16 [0.10, 0.21]*
No. of bedroom	-0.03 [-0.09, 0.01]	-0.15 [-0.17, -0.15]*	-0.03 [-0.09, 0.02]
No. of bathroom	0.19 [0.13, 0.25]*	0.24 [0.23, 0.26]*	0.19 [0.13, 0.25]*
Spatial parameter			
ρ	-	-	0.48 [0.07, 0.90]*
Variances			
ν^2	0.22 [0.19, 0.25]*	0.33 [0.29, 0.37]*	0.22 [0.20, 0.25]*
τ^2	0.20 [0.11, 0.34]*	-	0.31 [0.14, 0.58]*
Model fit criterion			
AIC	803.38	-645.104	786.95
DIC	-	-	850.39
WAIC	838.3	-	851.14
LMPL	-421.31	-	-427.21
Log-likelihood	-400.69	-520.97	-385.47

6.2 Model Implementation & Analysis

The cleaned and completed Lombok dataset was then model. The tables below present the results of the data fitting, offering a comparative view of each model’s effectiveness and accuracy in capturing the dataset’s characteristics.

$$prices \sim land_{std} + building_{std} + beds_{std} + baths_{std} \quad (10)$$

Standardization of the predictors ensures that the resulting coefficients are on a comparable scale, facilitating meaningful interpretation and comparison across different modeling approaches. The results presented in the following tables provide a comprehensive evaluation of each model’s performance in capturing both the structural and spatial characteristics of the property market in Lombok. This comparison allows us to assess the extent to which spatially explicit models (e.g., GWR, CAR, SAR) offer improvements over conventional methods in explaining price variability across regions. — Let me know if you’d like to include metrics such as R^2 , AIC, or Moran’s I in the paragraph as well.

We conducted model fitting on the area-level data, with detailed results provided in the Appendix (Table 9). However, this method does not provide optimal parameter estimates because data aggregation can reduce variability and obscure spatial details essential for capturing local effects. To address this, we used point-level data to apply three different models: GLMM, GWR, and multilevel CAR.

Source: [Article Notebook](#)

Here, Table 5 compares parameter estimates and fit criteria for three models: GLMM, GWR, and multilevel CAR. Significant coefficients are marked with an asterisk (*). Key determinants like land-size and built-up area have positive, significant effects across models, while the number of bedrooms shows only minor, varying effects, indicating model-specific differences in capturing this variable’s influence. Only the multilevel CAR model includes the spatial parameter ρ , with a significant estimate of 0.48, showing the model’s ability to account for spatial dependence. Variance components ν^2 (residual) and τ^2 (spatial) are also detailed, with both variance terms significant. Fit criteria (AIC, WAIC, and Log-likelihood) suggest mlvCAR and GWR may better capture local spatial variation, with mlvCAR balancing spatial structure and precision.

7 Results and Discussion

The findings from simulation consistently indicate that both the CAR and mlvCAR model outperform the other models in capturing spatial dependencies. This is evidenced by lower RMSE values and reduced bias across various dataset scenarios, highlighting the CAR model’s robustness. In datasets with strong spatial structures, such as those generated under *CAR* and *SAR + CAR* conditions, the CAR model achieves substantially lower RMSE values compared to GLM and SAR, suggesting its effectiveness in handling localized spatial variations.

In the context of house price modelling, it is common to encounter multiple sales observations within a single area, yet precise coordinates for each sale are often unavailable. Obtaining precise geographic coordinates in real-world datasets is often challenging due to limitations in data collection, privacy concerns, or the aggregation of data into broader administrative units. This limitation renders point-level approaches, such as GWR, less effective. With multiple observations available for each area, models like mlvCAR offer a more practical and accurate solution, effectively capturing spatial dependencies at an area level without requiring precise point-level locations.

The results underscore the importance of using models that explicitly account for spatial dependencies in fields like property valuation, where spatial heterogeneity plays a significant role. The superior performance of the CAR model, particularly in datasets with strong spatial structure, suggests that it may be a more reliable choice for applications where local spatial variations are critical. The findings align with previous studies indicating that CAR models are better suited for managing spatial autocorrelation and heterogeneity, especially when spatial dependencies are strong.

Additionally, the limitations of point-level models such as GWR in the absence of precise location data highlight a practical challenge in spatial modeling. When only area-level data is available, the use of models like mlvCAR becomes essential, as it allows for spatial analysis without the need for detailed coordinates, thus expanding the applicability of spatial models in real-world settings. These insights suggest potential for further exploration of the CAR and mlvCAR models in various domains, especially in urban planning, real estate, and other fields where spatial relationships

impact outcomes. Future research could examine the use of CAR and mlvCAR models with different spatial resolutions or apply these models to other datasets with varying levels of spatial dependency to further validate their effectiveness.

Supplementary information

Not applicable

Acknowledgments

Not applicable

Declarations

- Funding
- Competing interests

No, I declare that the authors have no competing interests as defined by Springer, or other interests that might be perceived to influence the results and/or discussion reported in this paper.

- Ethics approval
- Consent to participate
- Consent for publication
- Availability of data and materials
- Code availability
- Authors' contributions

∴ {.appendices latex=true}

8 Simple Feature Data Slice

Table 6: a slice of a simple feature collection with 6 polygon features and 1 attribute field, area(sqm), representing polygon sizes in square units. The dataset, using the WGS 84 CRS, spans longitudes 116.0355–116.5 and latitudes -8.500079—8.000057. The geometry field defines each polygon’s spatial boundaries, representing artificial study regions for spatial modeling in this research

area	geometry
17616519	POLYGON ((116.2296 -8.50007...
11485356	POLYGON ((116.0778 -8.32479...

[!h]

Table 7: Mean \pm SD values for bias across different area-level datasets and parameters. This table presents the mean and standard deviation (SD) of bias values across area-level datasets (CAR, SAR + CAR, GWR) and spatial models (GLM, SAR, CAR), with $nsim = 1000$. This is also act as numerical representation of fig. 5 (top panel)

Dataset	Mean \pm SD		
	GLM	SAR	CAR
β_1			
CAR	$7.80e-02 \pm 0.1900$	$9.00e-03 \pm 0.1900$	$-8.70e-02 \pm 0.1800$
SAR + CAR	$3.30e-01 \pm 0.3800$	$2.77e-01 \pm 0.3700$	$2.10e-01 \pm 0.3500$
GWR	$-6.00e-02 \pm 0.1100$	$-6.50e-02 \pm 0.1100$	$-6.50e-02 \pm 0.1100$
β_2			
CAR	$2.19e-01 \pm 0.1500$	$2.37e-01 \pm 0.1400$	$2.46e-01 \pm 0.1400$
SAR + CAR	$-2.47e-01 \pm 0.3900$	$-2.29e-01 \pm 0.3600$	$-2.12e-01 \pm 0.3700$
GWR	$3.05e-01 \pm 0.0900$	$3.02e-01 \pm 0.0900$	$3.04e-01 \pm 0.0900$
β_3			
CAR	$-1.24e-01 \pm 0.2000$	$0.00e+00 \pm 0.1900$	$-3.80e-02 \pm 0.1800$
SAR + CAR	$-5.80e-02 \pm 0.3700$	$6.30e-02 \pm 0.3500$	$-5.00e-03 \pm 0.3500$
GWR	$-6.20e-02 \pm 0.1100$	$-3.50e-02 \pm 0.1100$	$-5.90e-02 \pm 0.1100$
β_4			
CAR	$1.83e-01 \pm 0.1600$	$1.27e-01 \pm 0.1500$	$1.64e-01 \pm 0.1500$
SAR + CAR	$1.22e-01 \pm 0.3000$	$7.60e-02 \pm 0.2900$	$1.10e-01 \pm 0.2800$
GWR	$1.82e-01 \pm 0.0900$	$1.73e-01 \pm 0.0900$	$1.83e-01 \pm 0.0900$

4755629	POLYGON ((116.399 -8.380894...
15959450	POLYGON ((116.3567 -8.00006...
12727929	POLYGON ((116.14 -8.292773,...
8848979	POLYGON ((116.5 -8.210459, ...
7248031	POLYGON ((116.4308 -8.45091...
16906811	POLYGON ((116.1352 -8.26696...
6346890	POLYGON ((116.0723 -8.30335...
7620652	POLYGON ((116.4541 -8.43327...
10987363	POLYGON ((116.4982 -8.06795...
10411472	POLYGON ((116.1071 -8.07289...
20452118	POLYGON ((116.4233 -8.12744...
19613075	POLYGON ((116 -8.5, 116 -8....
9252319	POLYGON ((116.4541 -8.43327...

Source: [Article Notebook](#)

9 Bias values for Area-level Data

Source: [Article Notebook](#)

[!h]

Table 8: Mean \pm SD values for bias across different point-level datasets and parameters. This table presents the mean and standard deviation (SD) of bias values across point-level datasets (CAR, SAR + CAR, GWR) and spatial models (GLM, SAR, CAR), with $nsim = 1000$. This is also act as numerical representation of fig. 5 (bottom panel)

Dataset	Mean \pm SD		
	GLMM	GWR	mlvCAR
β_1			
CAR	$2.80e-04 \pm 0.0115$	$8.77e-03 \pm 0.0160$	$-2.00e-05 \pm 0.0114$
SAR + CAR	$4.06e-03 \pm 0.0525$	$9.00e-03 \pm 0.0538$	$9.80e-04 \pm 0.0524$
GWR	$-5.00e-04 \pm 0.0109$	$-5.20e-04 \pm 0.0108$	$-5.40e-04 \pm 0.0108$
β_2			
CAR	$-3.10e-04 \pm 0.0125$	$-5.50e-03 \pm 0.0172$	$-1.70e-04 \pm 0.0124$
SAR + CAR	$-2.81e-03 \pm 0.0588$	$-5.72e-03 \pm 0.0602$	$-1.02e-03 \pm 0.0585$
GWR	$9.54e-03 \pm 0.0113$	$9.55e-03 \pm 0.0113$	$9.56e-03 \pm 0.0113$
β_3			
CAR	$3.80e-04 \pm 0.0096$	$-2.86e-03 \pm 0.0142$	$4.50e-04 \pm 0.0095$
SAR + CAR	$9.00e-04 \pm 0.0391$	$-8.00e-04 \pm 0.0405$	$1.67e-03 \pm 0.0389$
GWR	$-4.90e-04 \pm 0.0092$	$-5.00e-04 \pm 0.0091$	$-5.20e-04 \pm 0.0091$
β_4			
CAR	$-4.30e-04 \pm 0.0089$	$1.18e-03 \pm 0.0122$	$-5.10e-04 \pm 0.0089$
SAR + CAR	$3.10e-04 \pm 0.0393$	$1.03e-03 \pm 0.0394$	$-8.00e-05 \pm 0.0391$
GWR	$1.09e-03 \pm 0.0089$	$1.05e-03 \pm 0.0088$	$1.07e-03 \pm 0.0087$

Source: [Article Notebook](#)

10 Model Comparison for area-level data

Source: [Article Notebook](#)

[!h]

Table 9: This table compares the performance of GLM, SAR, and CAR models on area-level data. It reports parameter estimates (with 95% confidence intervals), spatial parameters (ρ), variances (ν^2 , τ^2), and model fit criteria (AIC, DIC, WAIC, LMPL, log-likelihood). The results highlight each model's capability to capture spatial dependencies and variability

	GLM	SAR	CAR
Coefficients			
Intercept	7.19 [7.02, 7.35]*	6.92 [6.48, 7.37]*	7.19 [7.03, 7.35]*
Land-size	0.45 [0.24, 0.66]*	0.45 [0.25, 0.64]*	0.45 [0.23, 0.65]*
Built-up area	0.01 [-0.20, 0.23]	0.01 [-0.18, 0.22]	0.02 [-0.20, 0.23]
No. of bedroom	-0.20 [-0.43, 0.02]	-0.23 [-0.45, -0.01]*	-0.21 [-0.44, 0.02]
No. of bathroom	0.54 [0.33, 0.76]*	0.56 [0.36, 0.77]*	0.55 [0.34, 0.76]*
Spatial parameter			
ρ	-	0.009[-0.005, 0.023]	0.39 [0.01, 0.91]*
Variances			
ν^2	0.31 [0.21, 0.45]*	0.27 [0.19, 0.41]*	0.28 [0.01, 0.44]*
τ^2	0.20 [0.11, 0.34]*	-	0.04 [0.01, 0.91]*
Model fit criterion			
AIC	-	-	-
DIC	94.73	94.73	82.52
WAIC	95.64	-	87.43
LMPL	-48.04	-	-46.71
Log-likelihood	-41.49	-41.36	-37.29

11 Example of beta-values in GWR model

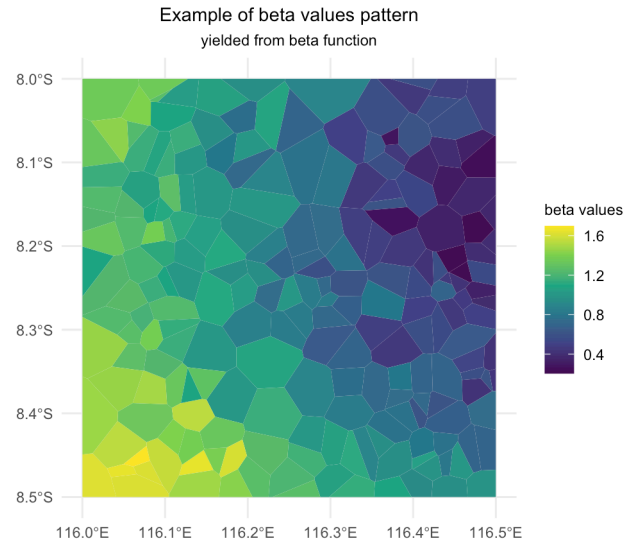


Figure 8: Spatial variation of β values generated using a custom beta function. The color gradient represents the magnitude of β values across the artificial study region, highlighting localised patterns and heterogeneity. The visualisation demonstrates the spatially varying coefficient structure modeled in the study

Source: [Article Notebook](#)

References

- Anselin, Luc. 2013. *Spatial Econometrics: Methods and Models*. Vol. 4. Springer Science & Business Media.
- Anselin, Luc, and Daniel A Griffith. 1988. "Do Spatial Effects Really Matter in Regression Analysis?" *Papers in Regional Science* 65 (1): 11–34.
- Banerjee, Sudipto, Bradley P Carlin, and Alan E Gelfand. 2014. *Hierarchical Modeling and Analysis for Spatial Data*. CRC press.
- Beguiría, Santiago, and Yolanda Pueyo. 2009. "A Comparison of Simultaneous Autoregressive and Generalized Least Squares Models for Dealing with Spatial Autocorrelation." *Global Ecology and Biogeography* 18 (3): 273–79.
- Besag, Julian, Jeremy York, and Annie Mollié. 1991. "Bayesian Image Restoration, with Two Applications in Spatial Statistics." *Annals of the Institute of Statistical Mathematics* 43 (March): 1–20. <https://doi.org/10.1007/BF00116466>.
- Bottero, Marta, Marina Bravi, Giulio Mondini, and Antonio Talarico. 2017. "Buildings Energy Performance and Real Estate Market Value: An Application of the Spatial

- Auto Regressive (SAR) Model.” *Appraisal: From Theory to Practice: Results of SIEV 2015*, 221–30.
- Brunsdon, Chris, A Stewart Fotheringham, and Martin E Charlton. 1996. “Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity.” *Geographical Analysis* 28 (4): 281–98.
- Cellmer, Radosław, Aneta Cichulska, and Mirosław Belej. 2020. “Spatial Analysis of Housing Prices and Market Activity with the Geographically Weighted Regression.” *ISPRS International Journal of Geo-Information* 9 (6): 380. <https://doi.org/10.3390/ijgi9060380>.
- Cellmer, Radosław, Katarzyna Kobylńska, and Mirosław Belej. 2019. “Application of Hierarchical Spatial Autoregressive Models to Develop Land Value Maps in Urbanized Areas.” *ISPRS International Journal of Geo-Information* 8 (4): 195.
- De Oliveira, Victor. 2012. “Bayesian Analysis of Conditional Autoregressive Models.” *Annals of the Institute of Statistical Mathematics* 64: 107–33.
- Droj, Gabriela, Anita Kwartnik-Pruc, and Laurențiu Droj. 2024. “A Comprehensive Overview Regarding the Impact of GIS on Property Valuation.” *ISPRS International Journal of Geo-Information* 13 (6): 175.
- Elhorst, J Paul et al. 2014. *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*. Vol. 479. Springer.
- Elhorst, J Paul, Donald J Lacombe, and Gianfranco Piras. 2012. “On Model Specification and Parameter Space Definitions in Higher Order Spatial Econometric Models.” *Regional Science and Urban Economics* 42 (1-2): 211–20.
- Fix, Miranda J, Daniel S Cooley, and Emeric Thibaud. 2021. “Simultaneous Autoregressive Models for Spatial Extremes.” *Environmetrics* 32 (2): e2656.
- Golgher, André Braz, and Paul R Voss. 2016. “How to Interpret the Coefficients of Spatial Models: Spillovers, Direct and Indirect Effects.” *Spatial Demography* 4: 175–205.
- Lee, Duncan. 2013. “CARBayes: An r Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors.” *Journal of Statistical Software* 55 (13): 1–24. <https://doi.org/10.18637/jss.v055.i13>.
- Leroux, Brian G, Xingye Lei, and Norman Breslow. 2000. “Estimation of Disease Rates in Small Areas: A New Mixed Model for Spatial Dependence.” In *Statistical Models in Epidemiology, the Environment, and Clinical Trials*, 179–91.
- LeSage, James P., and R. Kelley Pace. 2014. “Interpreting Spatial Econometric Models.” In *Handbook of Regional Science*, edited by Manfred M. Fischer and Peter Nijkamp, 1535–52. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-23430-9_91.
- Lu, Binbin, Martin Charlton, and A Stewart Fotheringham. 2011. “Geographically Weighted Regression Using a Non-Euclidean Distance Metric with a Study on London House Price Data.” *Procedia Environmental Sciences* 7: 92–97.
- McCord, Michael, Peadar T Davis, Martin Haran, David McIlhatton, and John McCord. 2014. “Understanding Rental Prices in the UK: A Comparative Application of Spatial Modelling Approaches.” *International Journal of Housing Markets and Analysis* 7 (1): 98–128.

- Morris, Mitzi, Katherine Wheeler-Martin, Dan Simpson, Stephen J Mooney, Andrew Gelman, and Charles DiMaggio. 2019. "Bayesian Hierarchical Spatial Models: Implementing the Besag York Mollié Model in Stan." *Spatial and Spatio-Temporal Epidemiology* 31 (November): 100301. <https://doi.org/10.1016/j.sste.2019.100301>.
- Pagourtzi, Elli, Vassilis Assimakopoulos, Thomas Hatzichristos, and Nick French. 2003. "Real Estate Appraisal: A Review of Valuation Methods." *Journal of Property Investment & Finance* 21 (4): 383–401.
- Rue, Havard, and Leonhard Held. 2005. *Gaussian Markov Random Fields: Theory and Applications*. Chapman; Hall/CRC.
- Sarlas, Georgios, and Kay W Axhausen. 2015. "Localized Speed Prediction with the Use of Spatial Simultaneous Autoregressive Models." In *TRB 94th Annual Meeting Compendium of Papers*. National Academy of Sciences.
- Sisman, Suleyman, and Arif Cagdas Aydinoglu. 2022. "A Modelling Approach with Geographically Weighted Regression Methods for Determining Geographic Variation and Influencing Factors in Housing Price: A Case in Istanbul." *Land Use Policy* 119 (August): 106183. <https://doi.org/10.1016/j.landusepol.2022.106183>.
- Soltani, Ali, Christopher James Pettit, Mohammad Heydari, and Fatemeh Aghaei. 2021. "Housing Price Variations Using Spatio-Temporal Data Mining Techniques." *Journal of Housing and the Built Environment* 36 (3): 1–29. <https://doi.org/10.1007/s10901-020-09811-y>.
- Stern, Hal, and Noel A Cressie. 1999. "Inference for Extremes in Disease Mapping." *John Wiley and Sons* 11: 63–84.
- Stewart Fotheringham, A, and Bumsu Park. 2018. "Localized Spatiotemporal Effects in the Determinants of Property Prices: A Case Study of Seoul." *Applied Spatial Analysis and Policy* 11 (August): 581–98. <https://doi.org/10.1007/s12061-017-9232-8>.
- Trojanek, Radoslaw, and Michal Gluszak. 2018. "Spatial and Time Effect of Subway on Property Prices." *Journal of Housing and the Built Environment* 33 (October): 359–84. <https://doi.org/10.1007/s10901-017-9569-y>.
- Van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "Mice: Multivariate Imputation by Chained Equations in r." *Journal of Statistical Software* 45: 1–67.
- Ver Hoef, Jay M, Erin E Peterson, Mevin B Hooten, Ephraim M Hanks, and Marie-Josée Fortin. 2018. "Spatial Autoregressive Models for Statistical Inference from Ecological Data." *Ecological Monographs* 88 (1): 36–59.
- Wall, Melanie M. 2004. "A Close Look at the Spatial Structure Implied by the CAR and SAR Models." *Journal of Statistical Planning and Inference* 121 (2): 311–24.
- Wang, Chih-Hao, and Na Chen. 2020. "A Geographically Weighted Regression Approach to Investigating Local Built-Environment Effects on Home Prices in the Housing Downturn, Recovery, and Subsequent Increases." *Journal of Housing and the Built Environment* 35: 1283–1302.
- Yang, Linchuan, Jiangping Zhou, Oliver F Shyr, et al. 2019. "Does Bus Accessibility Affect Property Prices?" *Cities* 84 (January): 56–65. <https://doi.org/10.1016/j.cities.2018.07.005>.

Yu, Danlin. 2007. "Modeling Owner-Occupied Single-Family House Values in the City of Milwaukee: A Geographically Weighted Regression Approach." *GIScience & Remote Sensing* 44 (3): 267–82.