VALOREM
Revista da Sociedade Brasileira de Engenharia de Avaliações

MASS APPRAISAL FOR PROPERTY TAX PURPOSES: APPROACH USING GLOBAL AND LOCAL SPATIAL REGRESSION

Avaliação em massa de imóveis para fins de imposto predial: abordagem utilizando regressão espacial global e local

Carlos Augusto Zilli
http://orcid.org/0000-0002-4926-1186
Instituto Federal de Santa Catarina (IFSC), Florianópolis, Brasil
carloszilli@gmail.com

Norberto Hochheim
http://orcid.org/0000-0003-4670-4831
Universidade Federal de Santa Catarina (UFSC), Florianópolis, Brasil
hochheim@gmail.com

ABSTRACT

Property tax is a vital revenue source for Brazilian municipalities, but the current system based on outdated assessment values results in unequal treatment of taxpayers. To promote fairness, regular property appraisals are necessary, especially in the rapidly evolving Brazilian real estate market. A key challenge in this process is developing a model that effectively addresses spatial effects. This study explores the application of global and local spatial regression models to tackle spatial dependence and heterogeneity in mass property appraisals, aiming to ensure fair property taxation. Analyzing urban properties, we evaluate the performance of the conditional autoregressive model (CAR) as a global regression model and geographically weighted regression (GWR) as a local regression model. Predictions from both models are combined with ordinary kriging interpolation to generate a comprehensive Property Value Map (PVM) representing the study area. Our findings show that the GWR modeling surpasses the CAR model in accuracy and its ability to account for spatial dependence and heterogeneity. This highlights GWR as a viable alternative for establishing a fair and equitable basis for calculating real estate taxes. This research significantly contributes to advancing mass appraisal techniques, ensuring property tax collection aligns with the dynamic nature of the Brazilian real estate market. The study's insights provide valuable guidance for policymakers and practitioners in improving property tax systems, ultimately benefiting municipalities and taxpayers alike.

Keywords: Mass appraisal of properties; Property tax; Property values map; Geographically weighted regression; Spatial econometrics.

RESUMO

O IPTU é uma fonte de receita essencial para os municípios brasileiros, mas o sistema atual, baseado em valores de avaliação desatualizados, resulta em um tratamento desigual dos contribuintes. Para promover a equidade, são necessárias avaliações regulares de imóveis, especialmente no mercado imobiliário brasileiro, que evolui rapidamente. Um desafio chave neste processo é desenvolver um modelo que trate eficazmente os efeitos espaciais. Este estudo explora a aplicação de modelos de regressão espacial global e local para enfrentar a dependência e heterogeneidade espaciais em avaliações em massa de imóveis, visando garantir uma tributação justa do imóvel. Analisando imóveis urbanos, avaliou-se o desempenho do modelo autoregressivo condicional (CAR) como um modelo de regressão global e o modelo de regressão geograficamente ponderada (GWR) como um modelo de regressão local. As previsões de ambos os modelos são combinadas com a interpolação por krigagem ordinária para gerar uma Planta de Valores Genéricos (PVG) abrangente, representando a área de estudo. Nossos achados demonstram que a modelagem GWR supera o modelo CAR em precisão e sua capacidade de contabilizar a dependência e heterogeneidade espacial. Isso destaca o GWR como uma alternativa viável para estabelecer uma base justa e equitativa para o cálculo de impostos sobre imóveis. Esta pesquisa contribui significativamente para o avanço das técnicas de avaliação em massa de imóveis, assegurando que a coleta do IPTU esteja alinhada com a natureza dinâmica do mercado imobiliário brasileiro. As percepções do estudo fornecem orientações valiosas para formuladores de políticas e profissionais na melhoria dos sistemas de tributação de propriedades, beneficiando, em última análise, tanto municípios quanto contribuintes.

Palavras-Chave: Avaliação em massa de imóveis; Imposto predial, Planta de valores genéricos; Regressão geograficamente ponderada; Econometria.

Preenchimento dos Editores

INFORMAÇÕES SOBRE O ARTIGO

Submetido em	20/08/2023
Publicado em	25/09/2023

Comitê Científico Interinstitucional
Editor-Responsável: Carlos Augusto Zilli
Avaliado pelo Sistema Double Blind Review (SEER/OJS – Versão 3)

1. INTRODUCTION

The current scarcity of fiscal resources, coupled with the increasing social demands arising from urban development, has necessitated significant efforts by public administrations to maintain budgetary and financial balance in Brazilian municipalities. The Urban Land and Property Tax (IPTU), a municipal tax, serves as a crucial source of financial revenue for urban management in these municipalities.

According to the Brazilian Federal Constitution (1988), IPTU is a municipal tax calculated based on the venal value of urban properties, which can have differentiated tax rates depending on the location and use of the property. The venal value of the property, representing the immediate selling price, is determined by summing the value of the land with the value of the corresponding buildings.

The Urban Land and Property Tax (IPTU) is an important source of collection of financial resources for the urban management of Brazilian municipalities. According to IMF (2017), in Brazil, IPTU collection indicators have varied between 0.40% and 0.45% of GDP since 2000. In 2016, IPTU revenues reached 0.50% of GDP, placing the country ahead of Mexico (0.2%) and Argentina (0.4%), but behind Colombia (0.8%) and Africa (1.3%).

The assessed value of the property, on which the IPTU is levied, needs to be correctly determined and periodically updated through mass property appraisal systems. The IPTU is often viewed as a regressive tax, as in the USA (McMillen, Singh, 2020).

Mass appraisal, according to Liporoni (2013), aims to determine property values on a large scale in order to maintain fair proportionality of these values to each other, in view of their location and the specific characteristics of each property, using standardized procedures and statistical tests.

The mass valuation of real estate has become increasingly important due to the large share of the real estate market in economic measures, which has become one of the countries' development indicators (Yilmazer and Kocaman, 2020).

Among the techniques commonly used for mass property valuation is classical linear regression (e.g. Uberti et al., 2018; Faria Filho et al., 2019; Benjamin et al., 2020). Geostatistics has already been tried with success in mass appraisals (Hornburg & Hochheim, 2017; Theodoro et al., 2019; Duarte, 2019) and, in some cases, non-parametric regressions have been applied successfully (e.g. Filho et al., 2005). Among machine learning methods, the most commonly used are artificial neural networks (McCluskey et al., 1999; Verikas et al., 2002; Pelli Neto, 2006; Selim, 2009) and, more recently, tree-based learning (Oliveira, 2020; Yilmazer & Kocaman, 2020).

However, in the midst of using these techniques, there has been a need to incorporate location variables into the models in order to model the so-called spatial effects.

According to Dantas (2003), the incomplete or inadequate treatment of spatial effects can generate serious problems in estimating the model, since, in the presence of spatial dependence on residues, the estimated parameters are inefficient and, in this case, the significance tests and inferred confidence intervals are no longer valid and the decisions become misleading.

An efficient way to incorporate spatial effects is through the use of spatial regression models (Anselin, 2005). The spatial regression models are divided into global and local.

The spatial error model (CAR), in which the spatial autocorrelation is attributed to the error term u, and the spatial lag model (SAR), in which the spatial autocorrelation is attributed to the explained variable, area among the global regression models.

Among the local regression modeling, there is geographically weighted regression (GWR), which makes it possible to adjust a regression model for each point in the data set, thus weighting all other observations, as a function of the distance to the respective point (Brunsdon et al., 1996).

Consequently, when using the GWR model, the advantage is the possibility of varying the model parameters according to its geographical location, incorporating cases of spatial heterogeneity, whereas a global regression model may not adequately represent local variations. Faced with this situation, this study proposes to apply the techniques of global and local spatial regression in the treatment of the effects resulting from spatial dependence and heterogeneity, aiming at obtaining more efficient and accurate models, which consider the spatial effects present in the data of estate market.

225 market data from the three neighborhoods under study were used, 190 for the work sample and 35 for the validation partition, collected between March and April 2020, available in Zilli (2020). The predictions were assessed using quality and performance indicators (AIC, RMSE, COD, PRD). The motivation and justification for the development of this study stem from the urgent need to have, at the same time, a more

accurate and fair method of mass valuation of properties for tax purposes. With the development of this study, we intend to generate knowledge that will contribute to the solution of problems related to the mass evaluation of real estate, such as, for example, the generation of generic value maps for the city halls.

The objective of this research is to investigate the use of the global model of spatial error or spatial lag (CAR or SAR) and the geographically weighted local model (GWR) as a tool for mass appraisal of urban properties, aiming to generate a Property Value Map (PVM) for apartments in the Centro, Trindade, and Agronômica neighborhoods in Florianópolis, Santa Catarina, Brazil. It is expected that the use of these models can correct the spatial effects of autocorrelation and heterogeneity present in the market data.

In the first part, the study area and the description of the variables used in this study will be presented. Then, the method and a review of the global spatial regression model and the local spatial regression model are provided. The criteria for evaluating the models are described, and the results of this research are presented. Finally, a Property Value Map is constructed, and the conclusions are provided.

2. MATERIALS AND METHODS

2.1. STUDIED AREA AND DATA

The neighborhoods of Agronômica, Centro and Trindade, in the city of Florianópolis, state of Santa Catarina, southern Brazil, are the studied area of this research. This area can be seen in figure 1.

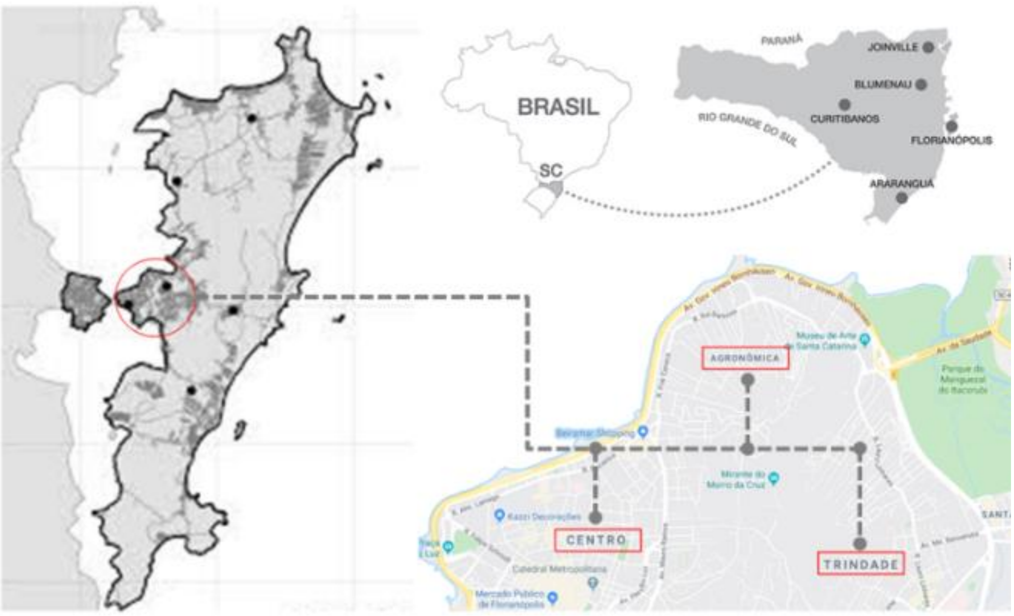

Figure 1: Location of the neighborhoods analyzed in Florianópolis/Brazil.

The Center (Centro) is a valued neighborhood in Florianópolis. Despite being in the West of the Island of Santa Catarina, Centro is geographically in the center of the urban area of Greater Florianópolis and is the most dense and vertical area of the Island of Santa Catarina. The best known avenue in the Center is Avenida Beira-Mar Norte which, in addition to a bike path and boardwalk, has some of the most expensive apartments in the city. Avenida Beira-Mar Norte extends to part of Agronômica.

The Agronômica is also a much valued neighborhood, exhibiting relatively new and high standard properties. Trindade is considered a university district because it hosts the main campus of the Federal University of Santa Catarina (UFSC); in this neighborhood, in general, older properties are found.

As material for this research, market data collected between March and April 2020 was used. The data were treated in the software R 3.5.3, where exploratory analysis, graph generation, and statistical tests were performed. For the modeling of the global spatial regression (CAR), the software GeoDa (Anselin, 2005) was used, and for the local spatial regression (GWR), the software GWR4 (Nakaya et al., 2014) was utilized. The Surfer 15 software was also employed to model the surfaces of values.

For this research, 225 market data were used, of which 190 were used as a training partition and 35 as a validation partition, in order to analyze the modeling performance. The data were distributed according to table 1. All data can be seen, spatially, in figure 2.

Table 1: Distribution, by neighborhood, of the working and validation partitions.

Neighborhood	Training Partition	Validation Partition
Centro	90	17
Agronômica	45	6
Trindade	55	12
Total	190	35

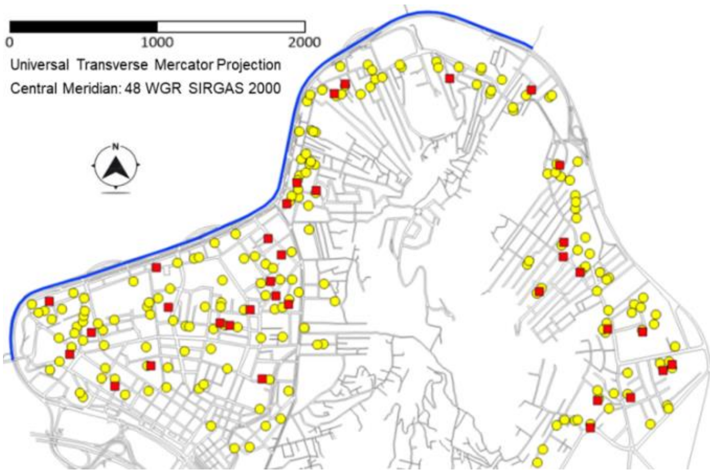

Figure 2: Spatialization of market data used in this study.

In figure 2, the yellow dots represent the 190 market data from the training partition and the red squares represent the 35 data from the validation partition. The blue line (Beira-Mar Avenue) represents a supposed valuation pole.

For this study, the price per squared meter was considered as the dependent variable, measured in R$/m²; the explanatory variables included: floor area (FA), in m²; distance to Beira Mar Avenue (DA), in meters; number of bedrooms (NB), bathrooms (BT) and parking spaces (NG), in units; existence of a swimming pool (SP), a dichotomous variable; and construction quality (CQ), of allocated codes (low standard as 1, medium as 2 and high as 3).

2.2 METHOD

In this study, the use of global and local spatial regression was investigated to address cases of spatial dependence and heterogeneity present in mass property appraisals. For this purpose, the spatial error regression (CAR) methods were employed for global modeling, and geographically weighted regression (GWR) was utilized for local modeling of market data. Additionally, ordinary kriging was performed on the predicted values obtained from the previous models to generate a representative value surface for the study area, following the stages outlined below:

2.2.1. Global Spatial Regression Model (CAR)

The global spatial regression models capture the spatial structure through a single parameter that is added to the traditional regression model. According to Anselin (2005), the spatial dependency can be incorporated in the classical regression models in two ways: as an additional regressor in the form of a spatially outdated dependent variable (Wy), or a spatially outdated structure in the regression error (Wε).

The first is known as the spatial lag model (spatial auto regressive - SAR or spatial lag model) and the second as the spatial error model (conditional auto regressive - CAR or spatial error model). For this work, the spatial error model proved to be more adherent to the real estate data of the area under study.

The spatial error model (CAR) considers that the errors associated with any observation are an average of the errors of the neighboring regions plus a component of random error. In this case, the effects of spatial autocorrelation are associated with the error term ε and the model can be expressed by equation (1).

$$Y = X\beta + \varepsilon \qquad \text{com} \qquad \varepsilon = \lambda \cdot W\varepsilon + \xi \qquad (1)$$

Where Y is the vector of the dependent variable, λ is the spatial autocorrelation coefficient that accompanies the lag $W\varepsilon$, W is the spatial neighborhood matrix or spatial weighting matrix, X is the matrix of the observations in the independent variables of the data, β is the vector of model parameters and ξ is the vector of random errors of the model with a mean of zero and a constant variance. The null hypothesis for the non-existence of autocorrelation is that the value of $\lambda = 0$, that is, the error term is not spatially correlated.

According to Almeida (2012), the spatial error model is also characterized by being a model with global reach. Its intuitive meaning is that the spatial pattern manifested in the error term is given by non-modeled effects due to the lack of adequate measure that, in turn, are not randomly distributed in space, but, on the contrary, are spatially autocorrelated.

In the context of evaluation engineering, this results in venal values that are not correlated with each other, so that there is only the correlation in the error term ε. This model occurs, therefore, when the error term of one location is correlated with the error values of other locations in the neighborhood.

2.2.2. Local Spatial Regression Model (GWR)

In regression analyses, when the spatial autocorrelation pattern varies in space, the global spatial regression model can ignore some significant local variations. In such cases, it is recommended to use a local regression model in which the parameters vary in space depending on their geographical position.

For this research, the geographic weighted regression technique (GWR), proposed by Brunsdon et al., (1996), will be applied as a method to model the spatial non-stationarity (hetorogeneity), a condition in which a global spatial regression model would not be able to adequately explain the relationships between some variables defined in a given geographic region.

In the GWR, a point-based calibration is performed around each regression point in which the closest observations should have a greater impact on the local set of coefficients than the more distant observations (Fotheringham et al., 2005).

GWR is a simple and effective technique, which extends the classical linear regression structure to explore spatial non-stationarity. It allows different relationships to exist at different points in space, so that local rather than global parameters are estimated. The model can be expressed by (2).

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^{p} \beta_j(u_i, v_i) \cdot x_{ij} + \varepsilon_i \qquad i = 1,\ 2,\ ...,\ n \tag{2}$$

Where yi is the dependent variable at location i, x_{ij} is the j-th independent variable at i, $\beta_0(u_i, v_i)$ is the intercept parameter at i, $\beta_j(u_i, v_i)$ the local regression coefficient for the k-th independent variable in i, (u_i, v_i) is the coordinate of the location i and ε_i is the random error in i. The geographically weighted regression model allows parameter estimates to vary in space and, therefore, spatial non-stationarity can be captured.

To estimate the parameters of the GWR equation, it is important to choose the criterion to decide on the weighting matrix. Weights are usually obtained using a spatial kernel function. There are two types that are frequently used, namely: fixed kernel and adaptive kernel.

In a fixed kernel function, an optimal spatial kernel (bandwidth) is calculated and applied over the study area. The weighting function most commonly used, and which brought the best results for this study, is the fixed Gaussian function given by (3).

$$W_{ij} = \exp[-(d_{ij} / \beta)^2] \qquad j = 1,\ 2,\ ...,\ n \tag{3}$$

Where W_{ij} is the weight attributed to observation j; d_{ij} is the distance between observation j and the regression point i; and β is the bandwidth, a key parameter for controlling the magnitude of the distance reduction. The term β represents the bandwidth, a parameter that controls the variance of the weighting function and determines the weight decay speed with distance (Silva, 2019).

The appropriate value for the bandwidth can be found by minimizing the cross-validation (CV) process or the Akaike information criterion (AIC). AIC approaches are preferred because they are responsible for the model's parsimony, that is, a trade-off between forecast accuracy and complexity.

Finally, considering that some explanatory variables can be stationary (global) and others non-stationary (local) in the studied area, Brunsdon et al. (1999) proposed a mixed model of geographically weighted regression (MGWR) in which some coefficients of the model (2) are assumed to be constant and others may vary in the studied region.

2.2.3. Selection criteria for spatial models

There are quite widespread criteria in the literature that can be adopted for the selection of models, and these take into account the complexity of the model in the selection criterion.

The Akaike information criterion (AIC) and the Bayesian information criterion (BIC) are frequently used for the selection of models, the best model being the one with the lowest value for these statistics. In this case, the Akaike information criterion (AIC) and the Bayesian information criterion (BIC) incorporate the value of the maximum likelihood function $f(x_n|\theta)$.

In his research, Akaike (1974) showed that the model's bias is given asymptotically by p, where p is the number of parameters to be estimated in the model, and defined his information criterion as given by equation (4).

$$AIC = -2\ln f(x_n \mid \theta) + 2p \tag{4}$$

Regarding the Bayesian information criterion (BIC), proposed by Schwarz (1978), with $f(x_n|\theta)$ being a statistical model estimated using the maximum likelihood method, then the Bayesian information criterion is given by equation (5).

$$BIC = -2\ln f(x_n \mid \theta) + p\ln(n) \tag{5}$$

Where $f(x_n|\theta)$ is the chosen model, p is the number of parameters to be estimated and n is the number of observations in the sample.

2.2.4. Mass appraisal performance measures

To verify the quality of the modeling, the square root of the mean square error was calculated by equation (6).

$$RMSE = \sqrt{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / n} \tag{6}$$

Where y_i corresponds to the observed unit value, $\hat{y}_i$ corresponds to the fitted unit value and n is the number of observations.

It is also possible to verify the quality of the models through the dispersion of the absolute relative errors of the values predicted by the models, according to equation (7).

$$ERRO_{REL} = \mid y_i - \hat{y}_i \mid / y_i \tag{7}$$

Where y_i corresponds to the observed unit value and $\hat{y}_i$ corresponds to the adjusted unit value. The absolute relative errors in this study generate a surface of easy interpretation and they playfully reflect the precision of the fitted model.

To analyze the performance of a mass evaluation, used for the generation of PVMs, it is advisable to verify if it fulfills the conditions of the International Association of Assessing Officers (IAAO, 2013). The performance of mass appraisals consists in comparing the values obtained in the appraisal with the values practiced in the market, that is, the market value of the properties. The performance parameters for mass assessments recommended by the IAAO standard (2013) are:

1. Median of the Valuation Ratios (R_{ATIO}): It is a measure of global performance and, for its calculation, given an observation i whose sales value is P_v and whose value calculated by the model is P_c, it is called valuation ratio (R_i) the ratio $R_i = P_c/P_v$. The median of all R_i's is then taken.

2. Median Dispersion Coefficient (COD): It is the average deviation expressed in percentage terms from the level at which each property was assessed in relation to the median of the assessed value divided by the market value. The coefficient measures the variability (degree of uniformity) of the assessments.

3. Price-related Differential (PRD): It is an indicator used to measure the vertical equity of the assessment and serves to verify systematic differences in the way that properties of higher values and properties of lower values are assessed. To obtain the price related differential, the average of the valuation ratios must be divided by the weighted average of the valuation ratios.

Table 2 shows the limits recommended by the IAAO (2013) for mass appraisals.

Table 2: Limits recommended by the IAAO (2013) for mass appraisals.

Indicator	Recommendation [1]
Median of Valuation Ratios (R_{atio})	$0.90 < R_{atio} < 1.10$
Coefficient of Dispersion of the Median (COD)	COD ≤ 15%
Price Related Differential (PRD)	0.98 ≤ PRD ≤ 1.03

[1] Note: The parameters in table 2 are defined for apartment type properties, in urban and heterogeneous residential areas.

It is desirable that the mass assessment presents parameters calculated within the limits established by the International Association of Assessing Officers (IAAO, 2013). In order to obtain sufficient precision in a mass appraisal process, the determination of the correct independent variables and the selection of the appropriate prediction model must be carried out properly.

3. RESULTS AND DISCUSSIONS

The spatial regression analysis process for the collected data includes two stages: exploratory analysis and modeling. In a simplified way, the exploratory analysis of the data allows to describe the distributions of the variables and the patterns of their spatial association, providing indicators about possible transformations to be carried out and the decision making for the modeling stage. In this work, classical descriptive statistics and Moran's global autocorrelation index were calculated to detect the presence of spatial components.

3.1. EXPLORATORY DATA ANALYSIS

The exploratory data analysis was performed on both the explained variable and the explanatory variables, using the graphs Box-Cox transformation, dispersion, moments, correlations and histograms as statistical tools. In figures 3a and 3b, we may see the frequency histograms for the dependent variable UV in normal scale and transformed into logarithmic.

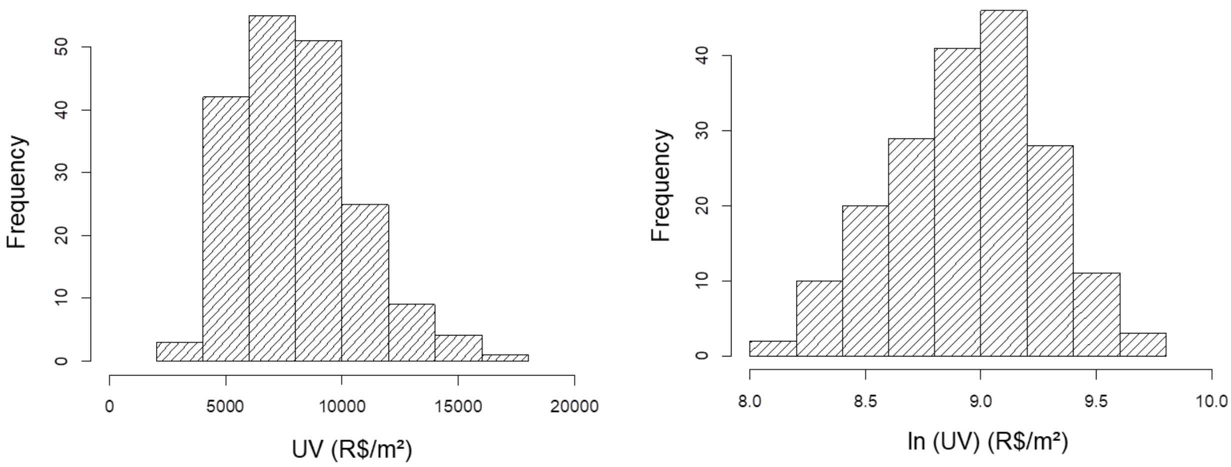

Figure 3: Frequency 3a) histogram of UV and 3b) histogram of ln (UV).

It can be seen that in the original scale (UV) the data presented a slight pattern of positive asymmetry with a platicurtic curve, with the asymmetric data to the right. Already, when the unit value variable goes through the logarithmic transformation (ln UV), there is a correction, resulting in a slightly negative and platicurtic asymmetry curve, with a flatter distribution function.

In order to confirm whether this transformation would be adequate to the data, a Box-Cox analysis was performed. The data in this study showed the value $\lambda = 0.15$, indicating that among the possible transformations, the natural logarithm will provide results very close to the ones obtained with the Box-Cox transformation.

It was also verified that the zero is contained in the 95% confidence interval for the values of λ, confirming that the logarithmic transformation - ln (UV) is the appropriate one for these data. Therefore, the natural logarithm transformation was adopted in the unit value variable (ln UV) for modeling market data.

Influencing points were checked using the Cook Distance method. Regarding the outliers, the variation of $\pm$ 3.0 standard deviations around the average was considered, since it is a considerably large sample, for real estate data.

Thus, two data were eliminated from the work sample, namely: the influencing point AP_86 (7,556.00) and the outlier AP_115 (6,627.00). The regression models built in this study were determined without these two points and the entire sample was reorganized for use in the following stages.

Regarding the explanatory variables, the logarithmic transformation was carried out in the floor area (FA) and distance to Beira-Mar Avenue (DA), and the other variables remained on the original scale.

For spatial modeling, the best classical regression model (OLS) was initially built, which could explain the real estate market in the most reliable way possible. Several simulations were carried out, with several transformations in the independent variables, and a model was obtained that did not violate any of the basic assumptions of classical regression.

To verify whether the data presented spatial dependence, an experimental semivariogram was generated to determine the spatial weights matrix. The semivariogram was built on the logarithm of the unit values and adjusted in software R in its version 3.5.3. The gstat package and its respective geostatistical libraries were used in the construction of the semivariogram. The curve becomes constant at 525 m. This is the distance that best captures the spatial autocorrelation of the real estate data in the sample.

After defining the matrix of spatial weights, the next step is to define the best model to fit. In this stage of the study, the Moran's I index was calculated and tests of the classic and robust Lagrange multipliers (LM) were also performed on the spatial weighting matrix with d = 525 m, calculated using the GeoDa software. The values obtained are shown in table 3.

Table 3: Calculation of Moran's I statistics and Lagrange Multiplier test.

Statistics	**Value**	**p-Value**
Moran's I (err)	6.9834	0.000
LM (lag)	7.3667	0.006
LM_{ROBUST} (lag)	0.1318	0.716
LM (err)	16.348	0.002
LM_{ROBUST} (err)	9.1136	0.000

The value of Moran's test I confirmed the existence of spatial dependence on the residuals of classical regression. The specification tests of the Lagrange Multipliers showed that, at the 5.0% level of significance, both the LM (lag) and the LM (err) became significant, showing spatial dependence on the dependent variable and in the classical regression model's error terms. In this case, the robust criterion will define the modeling that best fits the data, being, thus, only the significant LM_{ROBUST} (error). So, for this study, the spatial error modeling will be adopted (Spatial Error Model – CAR).

3.2. GLOBAL SPATIAL REGRESSION RESULTS (CAR)

Having defined the spatial error model (CAR) as the spatial regression model that would best fit the data of the properties in the area under study and establishing the spatial weight matrix (d = 525 m), it is now time to calculate the model coefficients and their respective significance using the z statistics.

The spatial regression equation for the error is presented in (8):

$$\ln(UV) = \beta_0 + \beta_1 \times \ln(FA) + \beta_2 \times \ln(DA) + \beta_3 \times (NB) + \beta_4 \times (BT) + \beta_5 \times (NG) + \beta_6 \times (SP) + \beta_7 \times (CQ) + \lambda \times (W_u) \qquad (8)$$

In the spatial regression error model, the lambda (λ) represents the term of spatial error autocorrelation and is significant at the level of 5.0%. It is also possible to observe that all variables in the spatial regression error model were significant at the level of 5.0%, as shown in table 4.

Table 4: Statistics related to the parameters of the CAR model.

Variable[1]	**Coefficients**	z_{calc}	**Significance**
Intercepto	10.6470	38.1924	0.00000
ln (FA)	– 0.43255	7.81141	0.00000
ln (DA)	– 0.11770	5.35113	0.00000
NB	0.05794	2.76309	0.00573
BT	0.03583	2.12263	0.03379
NG	0.18520	8.92755	0.00000
SP	0.09129	3.47056	0.00052
CQ	0.20171	12.3279	0.00000
λ (lâmbda)	0.56916	4.57072	0.00000

[1] Note: FA: floor area; DA: distance to Beira Mar Avenue; NB: number of bedrooms; BT: number of bathrooms; NG: number of parking spaces; SP: swimming pool; CQ: construction quality; λ: term of spatial autocorrelation.

It is possible to observe that all regressors are significant and important for the formation of the model, confirmed by the low and close to zero significance. The signs of the regressors of the spatial regression error model confirm the expectation of the local real estate market, and are therefore consistent.

Finally, the Breusch-Pagan test for the residuals of the spatial regression error model indicated that there are no problems of heteroscedasticity. With a BP value of 6.57 and a p-value of 0.4740, it can be said, at the 5.0% level of significance, that the autoregressive spatial error model is homocedastic.

3.3. LOCAL SPATIAL REGRESSION RESULTS (GWR)

The GWR model was adjusted with the same transformations and the same 190 market data used in the spatial regression of the error. To choose the best weighting function, whether fixed or variable, and the optimal bandwidth, the software GWR4 by Fotheringham et al. (2005) was employed, as described in the Methodology. Multiple simulations were conducted, which were separated according to selection criteria defined by the user in the GWR4 software interface.

The type of weighting function was evaluated between fixed or variable Gaussian and fixed or variable bi-squared. The software also allows you to choose the stopping criterion and check the automatic checkbox for the best bandwidth. It then generates several models for different bandwidths, based on the selected weighting function and finally presents the one with the lowest value for the selected criterion (AIC, AICc and CV).

Twelve models were generated and the chosen one would be the one that provided the lowest value in the Akaike information criterion (AIC), a parameter that will be used to make the comparison between the models generated in this study. The model that generated the lowest value in the AIC criterion was the fixed Gaussian weighting function with bandwidth β = 278.42 m.

As seen, there is still the possibility that, among the seven explanatory variables considered in this study, one of them presents spatial stationarity, that is, it does not have a relevant spatial variability along the geographic plane.

In general, the geographic variability for each coefficient can be tested by comparing models. To test the geographical variability of the k-th variable coefficient, a model comparison is performed between the adjusted GWR and a test model in which only the k-th coefficient is fixed, while other coefficients are maintained as in the geographically weighted adjusted model.

In this case, if the adjusted GWR is better than the test GWR model, using the AICc as a comparison criterion, we can conclude that the k-th coefficient certainly varies in space. This test routine repeats this comparison between models for each geographically variable coefficient.

For this test, as described by Leung et al. (2000), a positive DIFF Criterion, especially greater than 2.0, suggests that the local variable can be assumed as global. In the test, only the pool variable (PS) showed a positive result and greater than 2.0. In this study, this variable will be considered as global. The model, containing average coefficients for the local variables, and fixed coefficient for the pool variable, is in (9):

$$\ln (UV) = \beta_0 + \beta_1 \times \ln (FA) + \beta_2 \times \ln (DA) + \beta_3 \times (NB) + \beta_4 \times (BT) + \beta_5 \times (NG) + \beta_6 \times (SP) + \beta_7 \times (CQ) \qquad (9)$$

In the GWR model, the transformations adopted in the CAR model remained. It is noteworthy that the average coefficients were used because the GWR modeling, being a local regression, generates a regression equation for each sample point, weighted by the distance from its neighbors. The GWR model generated a report containing 190 regressions with the parameter estimates, the calculation of the t statistics and the standard error for each variable, as well as the predicted value, the local determination coefficient, the Cook distance and the influence, for each observation of the sample. Table 5 shows the statistics of the GWR model:

Table 5: Statistics related to the parameters of the GWR model.

Variable	Average Coefficients	Median Coefficients	DIFF Criterion	Type
Intercepto	10.73166	10.73542	– 681.00	Local
ln (FA)	– 0.51700	– 0.50740	– 490.75	Local
ln (DA)	– 0.08873	– 0.10434	– 685.68	Local
NB	0.06497	0.06628	– 6.0723	Local
BT	0.06173	0.06011	– 8.8288	Local
NG	0.19071	0.19394	– 8.5496	Local
SP	0.09732	0.09732	2.2167	Global
CQ	0.20735	0.22325	– 13.272	Local

The signs of the regressors of the geographically weighted regression model (GWR) were the same as the signs of the spatial error model (CAR) and confirm the expectation of the local real estate market, being, therefore, coherent.

3.4. GOODNESS OF FIT AND PERFORMANCE

In this study, the quality of the fit of the models was analyzed using the Akaike (AIC), Bayesian (BIC) and Loglikelihood (LIK) information criteria, the determination coefficient (R^2) and the standard error of the regression. Table 6 shows the indicators.

Table 6: Quality indicators of the modeling performed.

Statistics	CAR Model	GWR Model
Determination Coefficient (R^2)	0.8169	0.9239
Log Likelihood (LIK)	97.348	182.005
Schwarz Criterion (BIC)	– 152.80	– 187.12
Akaike Criterion (AIC)	– 178.70	– 209.78
Standard Error of Regression (Se)	0.1425	0.1303

It is possible to notice that the GWR model proved to be superior to the CAR model in all the fit quality indicators in table 6.

It was also verified how the absolute relative errors of the values predicted by the models were dispersed. Then, linear interpolation was performed to generate the surface of gradients, containing the absolute relative errors estimated for each modeling. These interpolations can be seen in figures 4a and 4b.

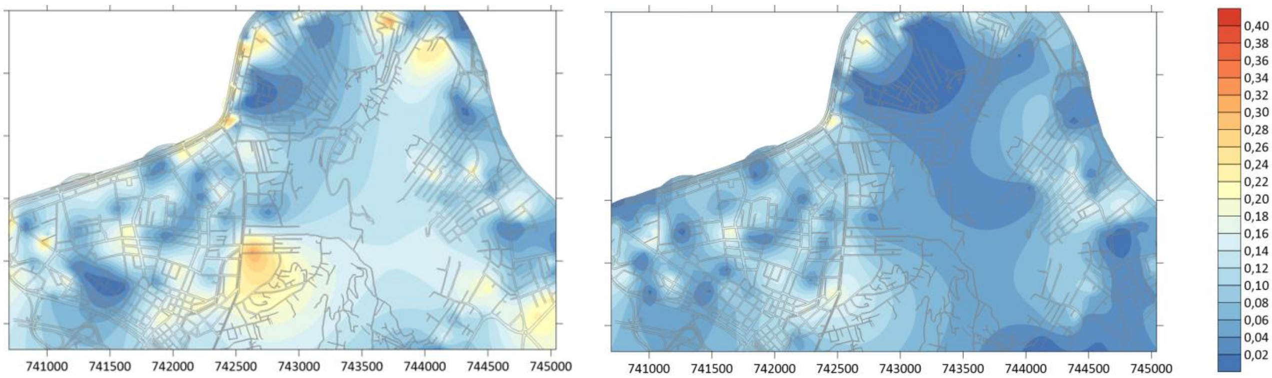

Figure 4: Surface of relative absolute errors for 4a) CAR model and 4b) GWR model.

Figures 4a and 4b show that the surface of relative errors for the GWR modeling was more uniform, with errors of lesser magnitude, evidenced by the dominance of the blue color.

As a way of visualizing the predictive power of the models, a dispersion graph of the observed versus adjusted values was constructed for each of the models, according to figures 5a and 5b.

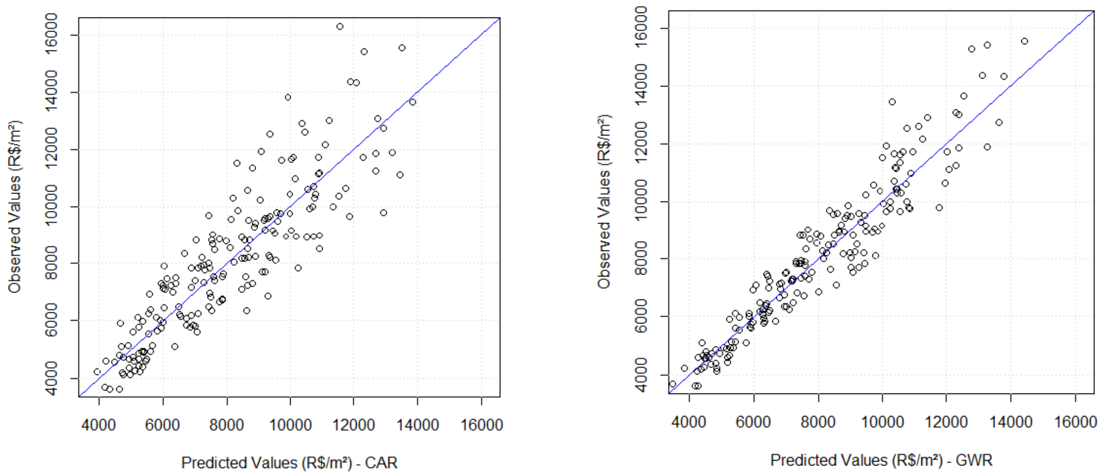

Figure 5: Dispersion diagram for 8a) CAR model and 8b) GWR model.

It is possible to notice that the GWR model was more adjusted, with less dispersed points, indicating that it is an adjustment with greater predictive power of unit values than the CAR model. To analyze the performance of the modeling, we used the metrics established by the IAAO (2013). The limits established by the standard for the properties under study are shown in table 2.

The data from the validation partition contained 17 properties in the Centro neighborhood, 6 properties in the Agronômica neighborhood and 12 properties in the Trindade neighborhood, as shown in table 1. The validation partition was analyzed to verify the presence of some discrepant data that could impair the results. It was found that the data AV_14, whose unit value of the property is R$ 17,500.00, is an outlier and, for this reason, was not used in the calculations.

To calculate the performance of the geographically weighted regression, the prediction of the unit values of the validation partition was performed directly in the GWR4 software. The GWR4 software provides, for each of the validation partitions, based on the selected kernel and bandwidth, a prediction equation, with the respective t-statistic, standard error and local R^2 values.

The results of the three indicators recommended by the IAAO (International Association of Assessing Officers) are shown in table 7.

Table 7: Performance indicators of the modeling performed.

Statistics	**CAR Model**	**GWR Model**
Median of the Assessment Ratios (R_{ATIO})	0.968	0.957
Coefficient of Dispersion of the Median (COD)	12.22%	10.57%
Price Related Differential (PRD)	1.017	1.015

It can be observed that the two models presented values within the limits recommended by the standard. However, the GWR modeling proved to be superior, indicating greater uniformity in the assessment, as evidenced by the lower value in the COD. As for the PRD, the results indicate that there was no progressivity or regressivity in the predictions. Nevertheless, the GWR model exhibited slightly superior performance in this indicator.

In addition, the square root of the mean square error (RMSE) was also calculated for the unit values predicted by the validation partition. The results are shown in table 8.

Table 8: Root mean squared error (RMSE) for the validation partition.

Statistics	**CAR Model**	**GWR Model**
$RMSE_{\text{VALIDATION PARTITION}}$	1,412.27	1,321.25

Notes: The currency exchange rate in 05/2023 is US$ 1.00 = R$ 5.05

In this case, the GWR modeling, which is a local spatial regression, again presented lower values for the RMSE. This phenomenon corroborates the idea that the value of a given property is strongly affected by the value of the properties in its surroundings.

The results of this topic indicate that, although all models have presented parameters within the recommendations of the IAAO, the GWR model proved to be superior to the CAR.

3.5. ELABORATION OF THE PROPERTY VALUE MAPS

To generate the surface that will compose the Property Value Map (PVM), it is necessary, initially, to homogenize the values predicted by the regression models, by means of a paradigm property representative of the area under study. The paradigm property used in this study is shown in table 9.

Table 9: Paradigm property used for homogenization of properties.

	FA	NB	BT	NG	SP	CQ
Paradigm	110	3	2	1	0	2

The homogenization of the values predicted by the CAR model was performed on the global regression of table 4. In the GWR model, the homogenization was performed on the regression of the point itself.

After the homogenization and calculation of the new unit values for the properties were carried out, the theoretical semivariograms were defined, which will support the creation of the unit value surfaces for the modelings under study.

With the aid of the R software, the values of lump effect, contribution and platform were defined. Exponential, spherical and Gaussian modeling was performed and, through cross-validation, it can be seen that the exponential semivariogram showed lower RMSE and stronger spatial dependence in both models.

For the construction of the surfaces of values adjusted by the two models, the parameters obtained in the semivariograms were used in the Surfer 15 software. The interpolation by ordinary kriging was performed, and for each model, it was possible to obtain a Property Value Map (PVM) for apartments, representative of the study area. These PVMs are shown in figures 6a and 6b.

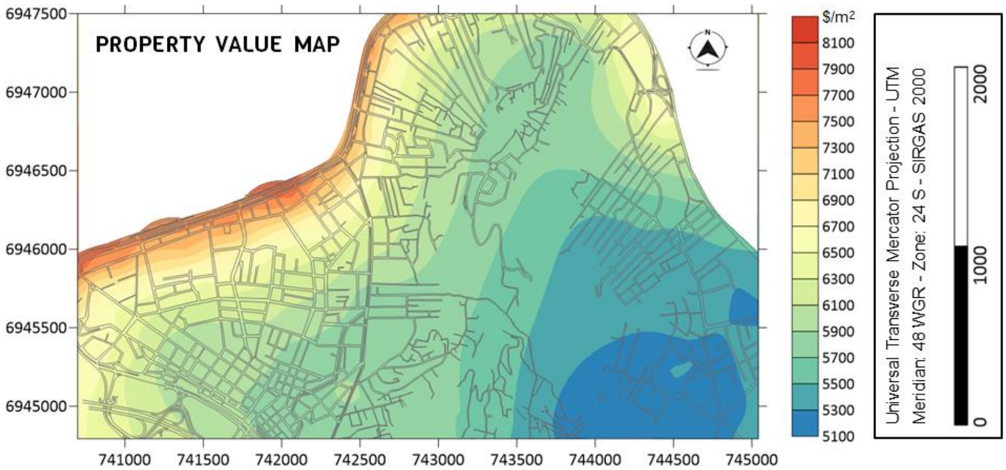

Figure 6a: Property Value Map (PVM) using kriging of the CAR model.

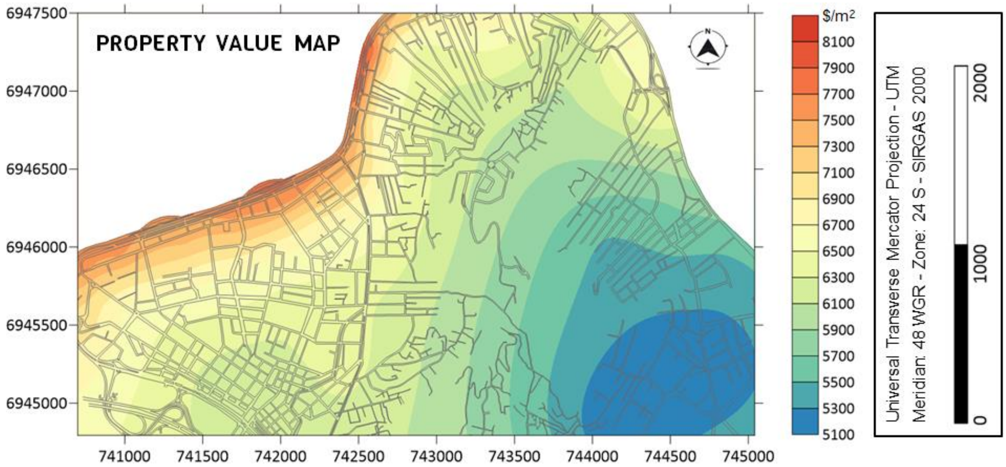

Figure 6b: Property Value Map (PVM) using kriging of the GWR model.

It can be seen, by the warmer regions spread over a good part of the extension of the Centro neighborhood, that the GWR surface was able to explain in a way closer to reality the unit value of the properties of this neighborhood, considering that, in general, the Centro has the most valued apartments in Florianópolis.

The same occurs with the kriging of the Agronômica neighborhood, where the Beira-Mar Avenue strip concentrates relatively new, high-standard properties and, consequently, valued in Florianópolis.

For the Trindade neighborhood, it can be observed that the two models were presented in a similar way, showing properties with homogenized unit value ranging from R$ 5,100.00 to R$ 6,300.00 per m². The Trindade neighborhood, in relation to the other two neighborhoods under study, has properties with the lowest unit value.

Therefore, the variable location of the properties was captured much more efficiently by GWR modeling. In general, the valuation of the neighborhoods under study was best captured in the PVM obtained by the kriging of GWR modeling.

The same occurs with the kriging of the Agronômica neighborhood, where the Beira-Mar Avenue strip concentrates relatively new, high-standard properties and, consequently, valued in Florianópolis.

For the Trindade neighborhood, it can be observed that the two models were presented in a similar way, showing properties with homogenized unit values ranging from R$ 5,100.00 to R$ 6,300.00 per m². The Trindade neighborhood, in relation to the other two neighborhoods under study, has properties with the lowest unit value.

Therefore, the variable location of the properties was captured much more efficiently by GWR modeling. In general, the valuation of the neighborhoods under study was best captured in the PVM obtained through kriging in GWR modeling.

3.6. QUALITY AND PERFORMANCE OF KRIGING

To evaluate the performance of the interpolations performed, we calculated the RMSE of the values predicted by the krigings for the 34 validation partitions. The results are shown in table 10.

Table 10: Root mean square error (RMSE) for the validation partition.

Statistics	**CAR Model**	**GWR Model**
$\text{RMSE}_{\text{VALIDATION PARTITION}}$	318.36	303.77

We may observe that the GWR modeling presented a lower RMSE, indicating that this model was able to make predictions closer to the expected. Again, the results demonstrate the superiority of GWR modeling in predicting property values.

Knowing that the lowest homogenized unit value observed is R$ 5,096.34 and that the RMSE of the validation partition is lower than this value, we find that the adjusted model is consistent.

Finally, the IAAO indicators were calculated for the values predicted by the krigings for the validation partition. The results of the performance indicators are shown in table 11.

Table 11: Performance indicators of the krigings carried out.

Statistics	**CAR Model**	**GWR Model**
Median of the Assessment Ratios (R_{ATIO})	0.995	0.995
Coefficient of Dispersion of the Median (COD)	4.84%	3.41%
Price Related Differential (PRD)	1.002	1.001

The results for both models were within the limits established by the IAAO, however, the GWR model presented a slightly superior performance in the COD and PRD indicators.

Considering the results in the RMSE, R_{ATIO}, COD and PRD indicators, we conclude that the GWR model is the one which would best represent a property value map (PVM) for the area under study.

4. CONCLUSIONS

Through the completion of this study, significant advancements were made in the field of mass valuation of urban properties by incorporating local spatial modeling through geographically weighted regression (GWR) and comparing it with global spatial modeling using conditional autoregressive (CAR) techniques.

The research showcases the novelty and compelling results of this approach, highlighting its potential for future analyses.

The findings of this study unequivocally demonstrate that GWR outperforms CAR in terms of overall performance and quality indicators. The GWR model exhibited superior adjustment to both the training partition data and the validation partition data, as evidenced by prediction diagrams and various indicators of quality and performance.

Our results indicate that the GWR model provides more accurate estimations compared to the CAR model, which assumes stationarity across the dataset. Notably, the GWR model improved the R^2 from 0.8169 to an impressive 0.9239 when compared to the CAR model. Additionally, the GWR modeling showed a substantial reduction in both the AIC and BIC values, specifically 31.08 and 34.32, respectively, in relation to the CAR model.

Furthermore, the GWR model displayed a lower standard error and a surface of absolute relative errors with smaller magnitudes and a more uniform distribution. These characteristics are highly desirable for generating a Property Value Map (PVM) as they contribute to a more accurate and consistent representation of property values.

In terms of predictive performance, both models satisfied the recommendations of the International Association of Assessing Officers (IAAO). However, the GWR model surpassed the CAR model by yielding lower values for the coefficient of dispersion (COD) and price-related differential (PRD). The COD value of 10.57% obtained by the GWR model indicates greater uniformity and reduced horizontal dispersion of the evaluation ratios, while the PRD value of 1.015, close to the unit (1.000), signifies an absence of progressivity or regressivity in the assessments.

The GWR modeling not only provided lower square root of the mean square error (RMSE) but also exhibited smaller absolute relative errors in predicting the unit values for the validation partition. These results further enhance the desirability of the GWR model for mass appraisals.

Additionally, the GWR-generated Property Value Map (PVM) demonstrated greater efficiency in representing the unit values of apartments in the studied neighborhoods, yielding results that closely aligned with the expected values.

Moreover, the GWR model's kriging technique yielded a lower RMSE value (R$ 303.77/m²) compared to the global model (R$ 318.36/m²). This reinforces the superiority of the GWR approach in minimizing errors and capturing spatial heterogeneity.

In summary, our study has unveiled the significant benefits of employing a local spatial regression model, specifically the GWR model, in the context of urban property valuation. The GWR model not only effectively addresses spatial dependence but also successfully tackles the challenge of spatial heterogeneity in the data. As a result, it enables more accurate value predictions, reduces distortions in Property Value Maps (PVMs), and promotes fairness and equity in taxation—a crucial step towards achieving fiscal justice.

By showcasing the novelty and remarkable outcomes of the research, the foundation has been laid for future analyses and advancements in the field of mass valuation of urban properties. The implications of the findings extend beyond this study and offer valuable insights for researchers, policymakers, and practitioners seeking to enhance the accuracy and fairness of property valuation systems.

REFERÊNCIAS

AKAIKE, H. A new look at the statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, n. 6, p. 716-723, 1994.

ALMEIDA, E. **Econometria espacial aplicada**. São Paulo: Alínea, 2012.

ANSELIN, L. **Spatial econometrics**: Methods and models. Kluwer Academic Publishers, 1988.

ANSELIN, L. **Exploring spatial data with GeoDaTM:** A workbook. Urbana Champaign: Spatial Analysis Laboratory, University of Illinois, GeoDa 1.6.7 9, 2005.

BRASIL. **Constituição da República Federativa do Brasil**. Brasília: DF, 1988.

BENJAMIN, J.; GUTTERY, R.; SIRMANS, C. Mass appraisal: an introduction to multiple regression analysis for real estate valuation. **Journal of Real Estate Practice and Education**, v. 23, n. 1, p. 65-77, 2020.

BRUNSDON, C.; FOTHERINGHAM, A. S.; CHARLTON, M. E. Geographically weighted regression: a method for exploring spatial nonstationarity. **Geographical Analysis**, v. 28, n. 4, p. 281-298, 1996.

BRUNSDON, C.; FOTHERINGHAM, A. S.; CHARLTON, M. E. Some notes on parametric significance tests for geographically weighted regression. **Journal of Regional Science**, v. 39, n. 3, p. 497-524, 1999.

CÂMARA, G. et al. Análise espacial de áreas. In: FUKS, S. D.; CARVALHO, M. S.; CÂMARA, G.; INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (Eds.). **Análise espacial de dados geográficos**. São José dos Campos, Brasil: Instituto Nacional de Pesquisas Espaciais, 2002. p. 133-164.

DANTAS, R. A. **Modelos espaciais aplicados ao mercado habitacional**: um estudo de caso para cidade do Recife. Recife: Universidade Federal de Pernambuco, 2003. (Doutorado em Economia).

DUARTE, D. C. O. **Análise multicritério e geoestatística aplicadas na avaliação em massa de imóveis urbanos**. Viçosa: Universidade Federal de Viçosa, 2019. (Doutorado em Engenharia Civil).

FARIA FILHO, R. F.; GONÇALVES, R. M. L.; LUIZ, H. T. G. Statistical models for generating the plants of generic values: an application in a small municipality. **Urbe - Revista Brasileira de Gestão Urbana**, v. 11, p. 29-44, 2019.

FILHO, C. M.; BIN, O. Estimation of hedonic price functions via additive nonparametric regression. **Empirical Economics**, v. 30, n. 1, p. 93-114, 2005.

FOTHERINGHAM, A. S.; CHARLTON, M.; BRUNSDON, C.; NAKAYA, T. Model selection problems in Geographically Weighted Regression. In: **Proceedings of the Geocomputation** 2005 Conference, Ann Arbor, MI, USA, 2005.

FOTHERINGHAM, A. S.; CHARLTON, M.; BRUNSDON, C. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. **Environment and Planning:** Economy and Space, v. 30, n. 11, p. 1905-1927, 1998.

HORNBURG, R. A.; HOCHHEIM, N. Avaliação em massa de imóveis usando geoestatística e krigagem bayesiana: um estudo de em Balneário Camboriú/SC. **REEC - Revista Eletrônica de Engenharia Civil**, v. 13, n. 1, 2017.

IAAO (International Association of Assessing Officers). **Standard on Ratio Studies**. Kansas City, Missouri, USA: IAAO, 2013.

IMF (International Monetary Fund). **Government finance statistics**. 2017.

LEUNG, Y.; MEI, C. L.; ZHANG, W. X. Statistical tests for spatial nonstationarity based on the geographically weighted regression model**. Environment and Planning**, v. 32, n. 1, p. 9-32, 2000.

LIPORONI, A. S. Avaliação de massa com ênfase em planta de valores. In: ALONSO, N. R. P. (Ed.). **Engenharia de avaliações**. 2. ed. São Paulo: Pini, 2013. p. 469-500.

MCCLUKEY, W. J.; ANAND, S. The application of intelligent hybrid techniques for the mass appraisal of residential properties. **Journal of Property Investment and Finance**, v. 17, n. 3, p. 218-238, 1999.

MCMILLEN, D.; SINGH, R. Assessment regressivity and property taxation. **The Journal of Real Estate Finance and Economics**, v. 60, p. 155-169, 2020.

NAKAYA, T.; CHARLTON, M.; BRUNSDON, C.; LEWIS, P.; YAO, J.; FOTHERINGHAM, A. S. GWR4 - **User manual Windows application for Geographically Weighted Regression modeling**. Department of Geography, Ritsumeikan University, 2014.

OLIVEIRA, A. A. F. **Avaliação em massa com modelos de aprendizado de máquina aplicados aos terrenos urbanos do município de Fortaleza**. Fortaleza: Universidade Federal do Ceará, 2020. (Mestrado em Economia do Setor Público).

PELLI NETO, A. **Redes neurais artificiais aplicadas às avaliações em massa:** estudo de caso para a cidade de Belo Horizonte/MG. Belo Horizonte: Universidade Federal de Minas Gerais, 2006. (Mestrado em Engenharia Elétrica).

SCHWARZ, G. Estimating the dimensional of a model. **Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

SELIM, H. Determinants of house prices in Turkey: hedonic regression versus artificial neural network. **Expert Systems with Applications**, v. 36, p. 2843-2852, 2009.

SILVA, A. C. C. **Análise do problema da unidade de área modificável pela regressão geograficamente ponderada**. Distrito Federal: Universidade de Brasília, 2019. (Mestrado em Estatística).

THEODORO, L. T. C.; UBERTI, M. S.; ANTUNES, M. A. H.; DEBIASI, P. Avaliação em massa de imóveis rurais através da regressão clássica e da geoestatística. **Revista Brasileira de Cartografia**, v. 71, n. 2, p. 459-485, 2019.

UBERTI, M. S.; ANTUNES, M. A. H.; DEBIASI, P.; TASSINARI, W. Mass appraisal of farmland using classical econometrics and spatial modeling. **Land Use Policy**, v. 72, p. 161-170, 2018.

VERIKAS, A.; LIPNICKAS, A.; MALMQVIST, K. Selecting neural networks for a committee decision. **International Journal of Neural Systems**, v. 12, n. 5, p. 351-362, 2002.

YILMAZER, S.; KOCAMAN, S. A mass appraisal assessment study using machine learning based on multiple regression and random forest. **Land Use Policy**, v. 99, 104982, 2020.

ZILLI, C. A. **Regressão geograficamente ponderada aplicada na avaliação em massa de imóveis urbanos**. Florianópolis: Universidade Federal de Santa Catarina, 2020. (Mestrado em Engenharia de Transportes e Gestão Territorial).