

CARLOS AUGUSTO ZILLI  
LUIZ FERNANDO PALIN DROUBI  
NORBERTO HOCHHEIM  
RAQUEL MARION DE FRANCESCHI

## **ANÁLISE DE MODELOS DE REGRESSÃO ESPACIAL GLOBAL E LOCAL PARA AVALIAÇÃO EM MASSA DE IMÓVEIS URBANOS**

Trabalho de Avaliação

Goiânia/GO  
2021

# ANÁLISE DE MODELOS DE REGRESSÃO ESPACIAL GLOBAL E LOCAL PARA AVALIAÇÃO EM MASSA DE IMÓVEIS URBANOS

## RESUMO

O uso da planta de valores genéricos (PVG) para fins de tributação imobiliária é uma eficiente ferramenta que os municípios podem dispor para formação da base de cálculo do imposto predial e territorial urbano (IPTU). A PVG é a representação gráfica dos valores genéricos do metro quadrado dos imóveis em uma mesma data, calculados por meio de avaliações em massa. Entre os métodos comumente utilizados para avaliações em massa está a regressão linear múltipla, que não considera efeitos espaciais em sua modelagem. Este estudo objetivou investigar o uso de modelos de regressão espacial global e local para o tratamento dos efeitos de dependência e heterogeneidade espaciais em avaliações em massa de imóveis urbanos. O estudo comparativo se deu no município de Santa Cruz do Sul (RS) e contou com uma base contendo 101 dados de terrenos. Os valores preditos pelas duas modelagens foram utilizados de forma combinada com o interpolador de krigagem ordinária para gerar uma planta de valores genéricos (PVG) representativa da área em estudo. Os resultados mostraram que tanto o modelo de regressão global (conditional autoregressive model - CAR) quanto o modelo de regressão local (geographically weighted regression – GWR) utilizados nesse estudo, se mostraram eficientes para tratar os efeitos espaciais, contudo os resultados indicaram superioridade da modelagem GWR em todos os indicadores de qualidade e desempenho (AIC, BIC, LIK, R<sup>2</sup>, RMSE, COD, PRD), demonstrando ser essa modelagem uma alternativa viável para geração de plantas de valores genéricos nos municípios brasileiros.

**Palavras-Chave:** Regressão Geograficamente Ponderada (GWR); Avaliação em Massa de Imóveis; Planta de Valores Genéricos (PVG); Engenharia de Avaliações.

## 1. INTRODUÇÃO

Conforme a Constituição Federal (1988), o IPTU é um imposto municipal calculado a partir do valor venal da propriedade urbana, que pode ter alíquotas diferenciadas, de acordo com a localização e utilização do imóvel. O valor venal de um imóvel é calculado por meio de um processo de avaliação imobiliária conhecido como avaliação em massa de imóveis.

Segundo Liporoni (2013), a avaliação em massa tem como objetivo a determinação, em larga escala, dos valores de imóveis, de forma a manter justa proporcionalidade destes valores entre si, em face da sua localização e das características específicas dos imóveis.

Entre as técnicas comumente empregadas para a avaliação em massa está a regressão linear múltipla. Contudo, objetivando modelar os chamados efeitos espaciais, tem-se verificado a necessidade de se incorporar variáveis de localização nos modelos de regressão. O uso da econometria espacial pode ser utilizado para modelagem dos efeitos espaciais sobre o valor de mercado dos imóveis na elaboração de planta de valores genéricos (PVG).

Conforme Anselin (2005), uma forma eficiente de se incorporar efeitos espaciais é por meio da utilização de modelos de regressão espacial, que podem ser divididos em globais e locais.

Entre as modelagens de regressões espaciais globais estão o modelo do erro espacial (CAR), em que a autocorrelação espacial é atribuída ao termo de erro  $u$ , e o modelo da defasagem espacial (SAR), em que a autocorrelação espacial é atribuída à variável explicada  $y$ .

Já entre as modelagens de regressões locais, está a regressão geograficamente ponderada (GWR), que possibilita o ajuste de um modelo de regressão para cada ponto do conjunto de dados, ponderando, desta forma, todas as demais observações, como função da distância ao respectivo ponto (Brunsdon *et al.*, 1996).

Sendo assim, este estudo se propõe a aplicar as técnicas de regressão espacial global e local no tratamento dos efeitos decorrentes da dependência e heterogeneidade espacial, visando a obtenção de modelos mais eficientes e precisos, que considerem os efeitos espaciais.

Objetiva-se, então, investigar como se dá o processo de avaliação em massa de imóveis utilizando a modelagem global do erro ou da defasagem espaciais (CAR ou SAR) e a modelagem local geograficamente ponderada (GWR) para a geração de uma planta de valores genéricos de terrenos do município de Santa Cruz do Sul.

Como base de dados para essa pesquisa utilizou-se 101 dados de mercado de terrenos, coletados em 2020, e se realizou a modelagem dos dados por meio de regressão espacial global (CAR/SAR) e local (GWR). Os resultados foram avaliados por meio de indicadores qualidade e desempenho (AIC, BIC, RMSE, COD, PRD).

O desenvolvimento deste trabalho decorre da necessidade iminente de se ter um método mais preciso e, ao mesmo tempo justo, de avaliação em massa de imóveis para fins fiscais, que são a base para definição de impostos como o IPTU.

## **2. AVALIAÇÃO EM MASSA DE IMÓVEIS PARA GERAÇÃO DE PVG's**

Segundo Uberti (2016), a avaliação em massa consiste em determinar os valores de todos os imóveis de uma região ou município. Uma importante aplicação das avaliações em massa de imóveis é na tributação imobiliária por meio dos valores constantes em uma planta de valores genéricos (PVG). A correta estimação dos valores constantes na PVG é fundamental para que se atinja a equidade (mesma relação para todos os imóveis entre o valor da avaliação e o valor de mercado) e justiça fiscal.

Nesse sentido, a principal função da planta de valores genéricos é permitir a definição de uma política de tributação imobiliária que seja acima de tudo justa e equânime. A PVG é um documento gráfico que representa a distribuição espacial dos valores médios dos imóveis em cada região do município, normalmente apresentados por face de quadra.

Para se analisar o desempenho de uma avaliação em massa, utilizada para a geração de PVG's, sugere-se verificar se esta cumpre as condições da *International Association of Assessing Officers* (IAAO, 2013), avaliando-se, dessa forma, a qualidade do ajuste.

### 3. REGRESSÃO ESPACIAL GLOBAL E LOCAL

O modelo de regressão é uma das ferramentas estatísticas mais conhecidas, no qual se utiliza o relacionamento existente entre duas ou mais variáveis, de modo que uma delas pode ser descrita ou o seu valor estimado a partir das demais. Os modelos de regressão são processos que envolvem, geralmente, uma variável resposta  $y$  e uma ou mais variáveis explicativas  $(x_1, x_2, \dots, x_p)$ , como na (Eq. 01).

$$y_i = \beta_0 + \sum_{j=1}^p \beta_j \cdot x_{ij} + \varepsilon_i \quad (\text{Eq. 01})$$

Onde,  $y_i$  é a  $i$ -ésima observação da variável resposta, com  $(i = 1, 2, \dots, n)$ ,  $x_{ij}$  é a  $j$ -ésima variável preditora para  $i$ -ésima observação, com  $j = (1, 2, \dots, p)$  e  $\varepsilon_i$  é o termo de erro relacionado a  $i$ -ésima observação. Se o modelo for uma representação realística da relação entre as variáveis  $y$  e  $x$ , então os valores de  $y$  em várias localidades estariam muito próximos daqueles preditos pelo modelo, com valores individuais dos erros  $y_i - E(y_i) = y_i - \hat{y}_i = \varepsilon_i$ , relativamente pequenos.

Na regressão tradicional, em que as posições geográficas não são consideradas, o procedimento de mínimos quadrados ordinários fornece um modelo de predição que resulta no menor valor da soma dos quadrados dos erros. Para obtenção das inferências, é necessário que os erros tenham uma distribuição normal, com média 0 e variância constante  $\sigma^2$ , e que sejam independentes entre si.

Esse modelo assume que os coeficientes da relação linear entre a variável explicada e as variáveis explicativas são constantes, independente da posição geográfica. Essa relação é válida também para pontos situados no mesmo *cluster*.

Segundo Almeida (2012), quando se aplica o método de mínimos quadrados ordinários para ajustar um modelo clássico de regressão linear em dados espaciais, com frequência, depara-se com o problema de não cumprimento da exigência dos erros independentes (que pode ser verificado por meio de testes de autocorrelação espacial nos erros) e da variância desses erros ser constante.

Completando, Câmara *et al.* (2002) diz que a inclusão de efeitos espaciais em modelos de regressão pode ocorrer de diferentes formas, destacando-se:

A) Os modelos com efeitos espaciais globais: partem do princípio de que o processo espacial subjacente aos dados analisados é estacionário. Neste grupo estão o modelo da defasagem espacial (SAR) e o modelo do erro espacial (CAR);

B) Os modelos com efeitos espaciais locais: ocorrem quando o processo espacial não é estacionário, os coeficientes de regressão precisam refletir a heterogeneidade espacial, sendo que uma das alternativas é modelar a tendência espacial de forma contínua, com parâmetros variando no espaço. Nesse grupo temos a abordagem chamada de regressão geograficamente ponderada (GWR).

### 3.1 REGRESSÃO ESPACIAL GLOBAL (CAR)

Conforme Druck *et al.* (2004), a inclusão explícita de efeitos espaciais em modelos de regressão pode ser feita de diferentes formas, sendo uma delas considerando-se modelos com efeitos espaciais globais. Tais modelos supõem que é possível capturar a estrutura da correlação espacial em um único parâmetro, que é adicionado ao modelo de regressão linear por mínimos quadrados ordinários (OLS).

Têm-se, considerando este contexto, duas alternativas: na primeira, a autocorrelação espacial ignorada é atribuída à variável dependente  $y$  e essa abordagem é denominada como modelo da defasagem espacial - SLM ou *spatial auto regressive* - SAR. Caso esse modelo não seja adequado, lança-se mão da segunda alternativa, que seria associar a autocorrelação espacial ao termo erro do modelo, considerando que os efeitos espaciais são ruídos que precisam ser removidos. Essa abordagem é denominada modelo do erro espacial - SEM ou *conditional auto regressive* - CAR.

Segundo Anselin (2005), a dependência espacial pode ser incorporada nos modelos clássicos de regressão de duas formas: como regressor adicional na forma de uma variável dependente espacialmente defasada, representada por  $Wy$ , ou uma estrutura espacialmente defasada no erro da regressão, representada por  $Wu$ .

O processo de especificação consiste na identificação de qual tipo de variável defasada, espacialmente, deverá ser introduzida ao modelo econométrico espacial. Os testes comumente utilizados para detectar a autocorrelação espacial são os testes I de Moran, o Multiplicador de Lagrange (LM) para defasagem espacial e, para o erro espacial, e o Multiplicador de Lagrange Robusto para ambos os casos.

Neste estudo, por ter se mostrado significativo nos testes de Lagrange, será apresentada a abordagem do modelo CAR. Nesta abordagem, a autocorrelação espacial é atribuída ao termo de erro  $u$ . Neste modelo, os erros associados com qualquer observação são uma média dos erros das regiões vizinhas mais uma componente de erro aleatório. O modelo do erro espacial é dado pela (Eq. 02).

$$y = X\beta + u \quad \text{com} \quad u = \lambda \cdot Wu + \varepsilon \quad (\text{Eq. 02})$$

Onde  $y$  é o vetor da variável dependente,  $\lambda$  é o coeficiente de autocorrelação espacial que acompanha a defasagem  $Wu$ ,  $W$  a matriz de vizinhança espacial ou matriz de ponderação espacial,  $X$  é a matriz das observações nas variáveis independentes dos dados,  $\beta$  é o vetor de parâmetros do modelo e  $\varepsilon$  é o vetor de erros aleatórios do modelo com média zero e variância constante.

O modelo do erro espacial, conforme Almeida (2012), caracteriza-se também por ser um modelo de alcance global. Seu significado intuitivo é que o padrão espacial manifestado no termo do erro é dado por efeitos não modelados por conta da falta de adequada medida que, por sua vez, não são distribuídos aleatoriamente no espaço, mas, ao contrário, estão espacialmente autocorrelacionadas.

Nesse modelo deve-se supor a inexistência de autocorrelação na variável explicada. No contexto da engenharia de avaliações, isso resulta em valores venais não correlacionados entre si, fazendo, então, com que haja apenas a correlação no termo de erro  $u$ . Esse modelo ocorre, portanto, quando o termo do erro de um local está correlacionado com os valores do erro de outros locais da vizinhança.

Para o diagnóstico de autocorrelação espacial, o teste do índice global de Moran é amplamente difundido e possibilita determinar se os resíduos do modelo de regressão clássica apresentam autocorrelação espacial, verificando se há a presença da componente espacial no erro. Essa estatística é determinada pela (Eq. 03).

$$I = \frac{n}{S} \cdot \frac{e'We}{e'e} \quad (\text{Eq. 03})$$

Onde  $n$  é o número de observações,  $e$  é o vetor de resíduos obtido por mínimos quadrados (variável padronizada),  $We$  é a matriz que representa os valores médios da variável de interesse padronizada nos vizinhos, definidos pela matriz de ponderação espacial  $W$ ;  $S$  é a soma dos elementos da matriz de pesos espaciais  $W$ .

### 3.2 REGRESSÃO ESPACIAL LOCAL (GWR)

O termo regressão geograficamente ponderada (GWR) foi apresentado por Fotheringham, Charlton e Brunson (2002) para descrever uma família de modelos de regressão em que os coeficientes podem variar espacialmente. O modelo GWR utiliza as coordenadas de cada ponto da amostra como um ponto alvo para uma forma de regressão de mínimos quadrados espacialmente ponderados.

O modelo GWR gera uma sequência de regressões lineares, estimadas para cada região, usando sub amostras de dados, ponderadas pela distância (Almeida, 2012). A ideia de atribuir pesos às observações individuais a partir de um ponto focal, chamado de ponto de regressão ou ponto de calibragem, embute o conceito de que a importância relativa é decrescente com a distância desse ponto.

O conceito de GWR evoluiu de métodos tradicionais de regressão linear, permitindo que as relações entre as variáveis variem espacialmente. Um modelo GWR básico para cada ponto de regressão pode ser escrito como em (Eq. 04).

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^p \beta_j(u_i, v_i) \cdot x_{ij} + \varepsilon_i \quad (\text{Eq. 04})$$

Onde  $y_i$  é a variável dependente no local  $i$ ,  $x_{ij}$  é o valor da  $j$ -ésima variável explicativa no local  $i$ ,  $\beta_0(u_i, v_i)$  é o parâmetro de intercepto no local  $i$ ,  $\beta_j(u_i, v_i)$  é o coeficiente de regressão local para a  $j$ -ésima variável explicativa no local  $i$ ,  $(u_i, v_i)$  é a coordenada do local  $i$  e  $\varepsilon_i$  é o erro aleatório no local  $i$ . Os  $\beta_j(u_i, v_i)$  são estimados

localmente em cada posição  $(u_i, v_i)$  pelo procedimento dos mínimos quadrados ponderados, no qual são utilizados pesos de decaimento em função da distância.

Considerando que certas variáveis explicativas que influenciam a variável resposta podem ser de natureza global, enquanto outras podem ser de natureza local, Brunson *et al.* (1999) propuseram um modelo misto de regressão geograficamente ponderada (MGWR) em que alguns coeficientes no modelo (Eq. 04) são assumidos como constantes e os demais podem variar na região estudada. Após reordenar as variáveis explicativas, um modelo MGWR é especificado como:

$$y_i = \beta_0(u_i, v_i) + \sum_{j=1}^q \beta_j \cdot x_{ij} + \sum_{j=q+1}^p \beta_j(u_i, v_i) \cdot x_{ij} + \varepsilon_i \quad (\text{Eq. 05})$$

Nesse caso, ao se tomar  $x_{i1} = 1$  ou  $x_{i,q+1} = 1$  para qualquer  $i$ , o modelo de regressão poderá envolver o intercepto constante ou uma variação espacial.

Conforme Fotheringham *et al.* (2005), em alguns casos os resultados podem sugerir que algumas relações são estacionárias no espaço, enquanto outras variam significativamente. Nesses casos, modelos GWR mistos, em que algumas relações variam espacialmente enquanto outras são mantidas constantes, são apropriados.

Quando um modelo misto do tipo MGWR é aplicado para se analisar um conjunto de dados do mundo real, deve-se primeiro determinar quais são os coeficientes que podem ser mantidos fixos e quais não. Para essa verificação, há o teste DIFF. Conforme Leung *et al.* (2000), um Critério DIFF positivo, especialmente maior ou igual a dois, sugere que o termo local é melhor para ser assumido global.

Segundo Brunson *et al.* (1999), em um modelo GWR a calibração de cada observação é realizada para cada local independentemente. Neste processo, o método dos mínimos quadrados ponderados é usado, e o cálculo da matriz para a estimação dos coeficientes de regressão pode ser expresso como em (Eq. 06).

$$\hat{\beta}(u_i, v_i) = (X^T W(u_i, v_i) X)^{-1} X^T W(u_i, v_i) y \quad (\text{Eq. 06})$$

Onde  $X$  é a matriz das variáveis explicativas com uma coluna de 1's para o intercepto,  $y$  é o vetor das variáveis dependentes,  $(\hat{\beta}(u_i, v_i) = (\beta_0(u_i, v_i), \dots, \beta_n(u_i, v_i)))$  é o vetor de  $n + 1$  coeficientes de regressão local,  $W(u_i, v_i)$ , é a matriz diagonal que denota a ponderação geográfica de cada observação para o ponto de regressão  $i$ .

Nesse caso, a matriz de ponderação  $W(u_i, v_i)$  é tomada como base na proximidade do ponto de regressão  $i$  aos pontos de dados em torno de  $i$  e calculado a partir de uma função de ponderação espacial ou kernel. Existem muitas funções de ponderação para se escolher, por exemplo, gaussiana, exponencial e bi-quadrada. Na prática, uma escolha contínua comum é a função de kernel gaussiana.

O kernel gaussiano pode ser do tipo fixo ou variável, que diferem entre si pela escolha da largura de banda  $\beta$ . Em geral, se for utilizado um kernel de banda fixa,  $d_{ij}$

refere-se a distância entre o ponto de regressão  $i$  e a observação  $j$ , o parâmetro  $\beta$  refere-se à suavização fixa. Já se for utilizado um kernel de banda variável, o parâmetro  $\beta_{i(k)}$  refere-se à suavização variável, com índice  $k$  representando o número de vizinhos mais próximos do ponto  $i$ . Essas equações estão na Tabela 1.

Tabela 1 – Funções de ponderação espacial abordadas

Função de Ponderação		Equação
Kernel Gaussiano Fixo	$w_{ij} = \exp^{[-(d_{ij} / \beta)^2]}$	$j = 1, 2, \dots, n.$
Kernel Gaussiano Variável	$w_{ij} = \exp^{[-(d_{ij} / \beta_{i(k)})^2]}$	$j = 1, 2, \dots, n.$

Fonte: os autores (2021).

Onde  $w_{ij}$  refere-se ao peso atribuído à observação  $j$  para a regressão de  $i$ . A largura de banda ideal, se fixa ou variável, pode ser encontrado minimizando a pontuação do CV (Cross Validation) ou do critério de informação de Akaike (AIC). As abordagens de AIC são preferidas porque são responsáveis pela parcimônia do modelo, ou seja, uma compensação entre a precisão da previsão e a complexidade.

#### 4. ÁREA DE ESTUDO, MATERIAIS E MÉTODO

Esta pesquisa tem como área de estudo o município de Santa Cruz do Sul, estado de Rio Grande do Sul. A área de estudo pode ser vista na figura 1.

Figura 1 – Localização da área de estudo.



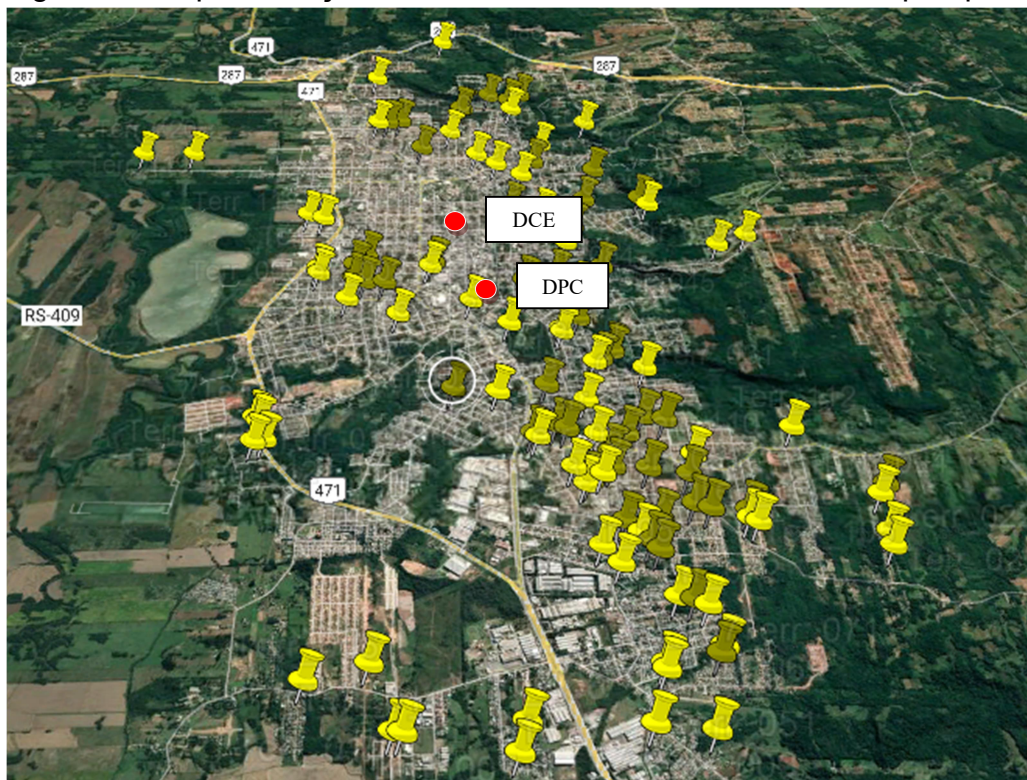
Fonte: os autores (2021).

Os dados foram tratados no software R 3.5.3, onde se realizou a análise exploratória, geração de gráficos e realização de testes estatísticos. Para a modelagem da regressão espacial global (CAR), utilizou-se o software GeoDa (Anselin, 2005) e para a regressão espacial local (GWR), utilizou-se o software



GWR4 (Nakaya *et al.*, 2014). Utilizou-se, ainda, o software Surfer 15 para modelagem das superfícies de valores. Na Figura 2 se vê a espacialização dos dados de mercado de Santa Cruz do Sul (RS), considerados nesta pesquisa.

Figura 2 – Espacialização dos dados de mercado utilizados na pesquisa.

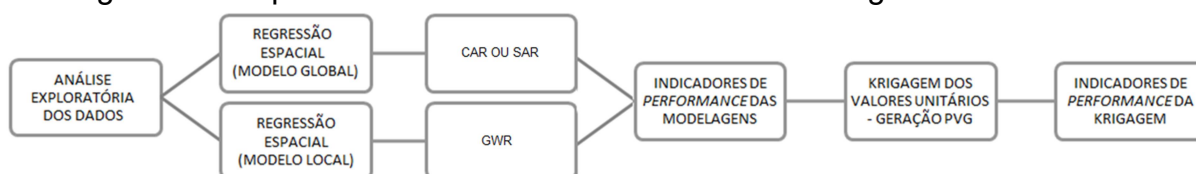


Fonte: adaptado de Hochheim e Franceschi (2020).

Para essa pesquisa, foram consideradas as variáveis valor unitário (VU), em R\$/m<sup>2</sup>; a área do terreno (ART), em m<sup>2</sup>; testada do terreno (TST), em metros; distância ao polo comercial da cidade, representado pela Praça Getúlio Vargas (DPC); distância ao Centro de Eventos, representado pelo parque da Oktoberfest (DCE), em metros; se está situado numa rua pavimentada ou não (PVT) e situação na quadra, informando se é terreno de esquina ou não (STU). Outras variáveis explicativas também foram testadas, mas elas não se mostraram significativas.

No Figura 3 é apresentada a sequência de etapas realizadas nesse estudo para se atingir o objetivo proposto. Em seguida, descreve-se as etapas propostas.

Figura 3 – Etapas do estudo contendo os modelos de regressão abordados.



Fonte: os autores (2021).

Inicialmente realizou-se a análise exploratória dos dados. Fez-se uma análise dos diagramas de dispersão, plotagem dos histogramas de frequências, verificação de possíveis transformações nas variáveis e testes de significância e de correlação, com o objetivo de verificar quais variáveis eram relevantes para a formação dos valores dos terrenos. São apresentados alguns resultados no tópico seguinte.

Para as modelagens espaciais, construiu-se, inicialmente, o melhor modelo de regressão clássica (OLS), que pudesse explicar o mercado imobiliário da forma mais fidedigna possível. Realizaram-se diversas simulações, com diversas transformações nas variáveis independentes e obteve-se um modelo que não violava nenhum dos pressupostos básicos da regressão clássica.

Para verificar se os dados apresentavam dependência espacial, gerou-se um semivariograma experimental para determinação da matriz de pesos espaciais e, a partir daí, a definição de qual o melhor modelo a se ajustar (CAR ou SAR). Nesta etapa do estudo, fez-se o cálculo do índice Moran's I e se realizou, também, os testes dos multiplicadores de Lagrange (LM) clássico e robusto. O semivariograma foi ajustado no software R na sua versão 3.5.3. Utilizou-se na construção do semivariograma o pacote gstat e suas respectivas bibliotecas geoestatísticas.

Finalizada esta etapa, fez-se a modelagem dos dados por regressão geograficamente ponderada. Inicialmente, definiu-se que o critério de seleção do modelo seria aquele que apresentaria o menor valor no Critério de Informação de Akaike (AIC). Utilizou-se o software *GWR4* criado por Fotheringham *et al.* (1998) para modelagem dos dados e testou-se, para todos os 2 tipos de função de ponderação (gaussiano fixo e variável), as mais diversas larguras de banda, por meio da função "*Golden Section Search*", que busca a melhor largura de banda por meio de critérios definidos pelo usuário. Testou-se, também, as variáveis que se apresentavam estacionárias, com comportamento global.

A *performance* das modelagens foi verificada por meio do Critério de Informação de Akaike (AIC), o Critério de Informação Bayesiano (BIC), o log *Likelihood* (LIK), o desvio padrão e o coeficiente de determinação ( $R^2$ ). Fez-se, também, o gráfico de predição e o cálculo da raiz quadrada do erro quadrático médio (RMSE) e do erro relativo das modelagens. Calculou-se, ainda, a mediana das razões de avaliação (Med R), o coeficiente de dispersão das razões de avaliação (COD) e o diferencial relativo ao preço (PRD) (IAAO, 2013).

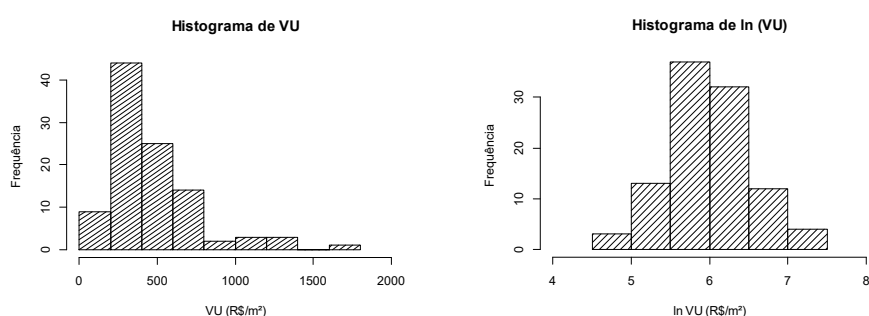
Para construção das superfícies de valores, definiu-se, inicialmente, o imóvel paradigmático representativo da área em estudo, seguido pela homogeneização dos valores preditos por cada uma das modelagens. Realizada a homogeneização, buscou-se construir o semivariograma experimental e, por meio do software R 3.5.3, calculou-se os parâmetros de alcance, efeito pepita e contribuição. Para etapa de construção das superfícies utilizou-se o software *Surfer 15*. Realizou-se, portanto, a krigagem ordinária dos valores homogeneizados preditos por cada modelo de regressão com base no semivariograma teórico escolhido. Construiu-se, então, o mapa de predição dos valores unitários de cada modelagem (OLS e GWR).

## 5. RESULTADOS

### 5.1. ANÁLISE EXPLORATÓRIA DOS DADOS

Realizou-se a análise exploratória dos dados tanto na variável explicada quanto nas variáveis explicativas, utilizando como ferramentas estatísticas os gráficos de *boxplot* e de dispersão, momentos, correlações e histogramas. Na Figura 4 é apresentado o histograma de frequências para a variável dependente em duas escalas: normal (VU) e logarítmica (ln VU).

Figura 4 – Histograma de frequências da variável valor unitário (VU e ln VU).



Fonte: os autores (2021).

Ao se realizar o histograma da variável transformada em logaritmo, pode-se verificar melhor correção da assimetria, tornando os dados mais próximos da curva normal. Segundo Dantas (2012), a transformação logarítmica é a preferida quando se procura ajustar modelos de dados imobiliários. Com base nesses resultados, adotou-se a transformação logarítmica na variável valor unitário (ln VU).

Fez-se a verificação de pontos influenciantes pelo método da Distância de Cook. Com relação aos outliers, considerou-se a variação de  $\pm 2,0$  desvios padrões em torno da média. Não foram detectados dados que violassem estas condições.

Com relação às variáveis explicativas, após diversas simulações, realizou-se a transformação logarítmica na área do terreno (ART), na testada (TST), na distância ao polo comercial (DPC) e distância ao centro de eventos (DCE) e as variáveis pavimentação (PVT) e situação (STU) permaneceram na escala original.

Para verificar se os dados apresentavam dependência espacial, gerou-se um semivariograma experimental para determinação da matriz de pesos espaciais. O semivariograma foi construído sobre o logaritmo dos valores unitários e ajustado no software R na sua versão 3.5.3. Utilizou-se na construção do semivariograma o pacote gstat e suas respectivas bibliotecas geoestatísticas. Constata-se que a curva se torna constante em 1772 m. Essa é a distância que captura melhor a autocorrelação espacial dos dados dos imóveis da amostra e será a adotada.

Definida a matriz de pesos espaciais deve-se seguir para definição de qual o melhor modelo a se ajustar. Nesta etapa do estudo, fez-se o cálculo do índice

Moran's I e se realizou, também, os testes dos multiplicadores de Lagrange (LM) clássico e robusto sobre a matriz de ponderação espacial com  $d = 1772m$ , calculados por meio do software GeoDa. Os valores obtidos são mostrados na Tabela 4.

Tabela 4 – Cálculo da estatística I de Moran e dos Multiplicadores de Lagrange.

<b>Estatística</b>	<b>Valor</b>	<b>p-Valor</b>
Moran's I (err)	5,675	0,000
LM (lag)	2,630	0,105
LM ROBUSTO (lag)	1,478	0,224
LM (err)	13,948	0,000
LM ROBUSTO (err)	12,795	0,000

Fonte: os autores (2021).

O valor do teste I de Moran confirmou a existência de dependência espacial nos resíduos da regressão clássica. Os testes de especificação dos Multiplicadores de Lagrange mostraram que, ao nível de significância de 5,0%, apenas o LM (err) se mostrou significativo, evidenciando dependência espacial nos termos de erro do modelo de regressão clássico. Nesse caso, para esse estudo, será adotada a modelagem do erro espacial (*Spatial Error Model – CAR*).

## 5.2. MODELO DE REGRESSÃO ESPACIAL GLOBAL (CAR)

Definido o modelo do erro espacial (CAR) como o modelo de regressão espacial que melhor se ajustaria aos dados dos terrenos da área em estudo e estabelecida a matriz de pesos espaciais ( $d = 1772 m$ ), calculou-se os coeficientes do modelo e as respectivas significâncias por meio da estatística z. A equação de regressão espacial do erro é apresentada em (Eq. 07).

$$\ln(VU) = \beta_0 + \beta_1 \times \ln(ART) + \beta_2 \times \ln(TST) + \beta_3 \times \ln(DPC) + \beta_4 \times \ln(DCE) + \beta_5 \times (PVT) + \beta_6 \times (STU) + \beta_7 \times (W_u) \quad (\text{Eq. 07})$$

No modelo de regressão espacial do erro a variável lambda ( $\lambda$ ) representa o termo de autocorrelação espacial do erro e se mostra significativa ao nível de 5,0%. Observa-se, ainda, que todas as variáveis do modelo de regressão espacial do erro se mostraram significativas ao nível de 5,0%, conforme Tabela 5.

Tabela 5 – Estatísticas relativas aos parâmetros do modelo CAR

<b>Variável</b>	<b>Coefficiente</b>	<b>Erro Padrão</b>	<b>Estatística z</b>	<b>Significância</b>
Intercepto	14,20270	0,68154	20,83910	0,00000
ln (ART)	- 0,72679	0,07828	9,28412	0,00000
ln (TST)	0,50736	0,12238	4,14594	0,00003
ln (DPC)	- 0,07174	0,01326	5,40859	0,00000

ln (DCE)	- 0,62139	0,07202	8,62805	0,00000
PVT	0,18744	0,06497	2,88476	0,00392
STU	0,17315	0,05977	2,89687	0,00377
LAMBDA $\lambda$	0,59728	0,15803	3,77956	0,00016

Fonte: os autores (2021).

Os sinais dos regressores confirmam a expectativa do mercado imobiliário local, tornando-os coerentes. Observa-se que todos são significativos e importantes para a formação do modelo, confirmado pelas significâncias próximas de zero.

Por fim, o teste de Breusch-Pagan para os resíduos do modelo de regressão espacial do erro indicaram que não há problemas de heterocedasticidade. Com um valor de BP = 3,92 e um  $p$ -valor = 0,6872 pode-se afirmar, ao nível de significância de 5,0%, que o modelo do erro espacial autorregressivo é homocedástico.

### 5.3. MODELO DE REGRESSÃO ESPACIAL LOCAL (GWR)

O modelo de regressão geograficamente ponderada (GWR) foi ajustado utilizando a base de 101 dados de mercado utilizadas na regressão do erro espacial, bem como as transformações realizadas nas variáveis. Para escolha da melhor função de ponderação, se fixa ou variável, e melhor largura de banda, foi utilizado, conforme descrito na metodologia, o *software* GWR4 e realizadas diversas simulações, separadas por critérios de escolha definido pelo usuário.

O modelo que gerou o menor valor no critério de AIC (critério escolhido) foi o de função de ponderação gaussiano fixo com largura de banda de  $\beta = 1145$  m.

O valor obtido foi AIC = + 2,93. Contudo, a variabilidade geográfica para cada coeficiente pode ser testada por comparação de modelos. Para testar do  $k$ -ésimo coeficiente variável, uma comparação de modelo é realizada entre o GWR ajustado e um modelo de teste no qual apenas o  $k$ -ésimo coeficiente é fixo, enquanto outros coeficientes são mantidos como no modelo geograficamente ponderado ajustado. Nesse caso, se o GWR ajustado for melhor que o modelo GWR de teste, utilizando como critério de comparação o AICc, pode-se concluir que  $k$ -ésimo coeficiente certamente varia no espaço. Essa rotina de teste, chamado de teste Diff, repete essa comparação entre modelos para cada coeficiente geograficamente variável.

Realizou-se, então, o teste Diff (Leung *et al.*, 2000) para verificar se alguma variável apresentava estacionariedade espacial, ou seja, não tinha variabilidade espacial relevante no plano geográfico. Os autores sugerem atenção aos valores Diff positivos, especialmente aquelas variáveis cujo critério seja superior a + 2,0.

Nesse estudo, a variável pavimento (PVT) apresentou Diff = + 3,24, superior a 2,0 e se percebeu que, quando se considerava essa variável como global, o AIC tinha um incremento de 3,24 unidades, chegando a AIC = - 0,31. Sendo assim, esta variável (PVT) será considerada estacionária (global) nesse estudo. O modelo, contendo os parâmetros médios para a regressão GWR, é visto em (Eq. 08):

$$\ln(VU) = \beta_0 + \beta_1 \times \ln(ART) + \beta_2 \times \ln(TST) + \beta_3 \times \ln(DPC) + \beta_4 \times \ln(DCE) + \beta_5 \times (PVT) + \beta_6 \times (STU) \quad (\text{Eq. 08})$$

No modelo de regressão geograficamente ponderada (GWR) anterior, como se pode observar, permaneceram as transformações adotadas na modelagem de regressão clássica. Ressalta-se que os coeficientes médios foram utilizados pois a modelagem GWR, por ser uma regressão local, gera uma equação de regressão para cada ponto amostral, ponderada pela distância dos seus vizinhos. Na Tabela 6 podem-se acompanhar algumas estatísticas para a modelagem GWR:

Tabela 6 – Estatísticas relativas aos parâmetros do modelo GWR

<b>Variável</b>	<b>Coeficiente Médio</b>	<b>Coeficiente Mediano</b>	<b>Erro Padrão Médio</b>	<b>Tipo de Coeficiente</b>
Intercepto	14,31144	14,11005	1,06186	Local
ln (ART)	- 0,61100	- 0,55638	0,16488	Local
ln (TST)	0,37249	0,39875	0,20078	Local
ln (DPC)	- 0,06980	- 0,07186	0,03375	Local
ln (DCE)	- 0,66679	- 0,65544	0,14992	Local
PVT	0,21112	0,17827	0,11369	Global
STU	0,20348	14,11005	1,06186	Local

Fonte: os autores (2021).

Os sinais dos regressores do modelo de regressão geograficamente ponderada (GWR) se apresentaram iguais aos sinais do modelo do erro espacial (CAR) e confirmam a expectativa do mercado imobiliário local, sendo, portanto, coerentes.

Para complementar esta análise, é apresentando os valores de mínimo, máximo, da mediana e dos 1º e 3º quartis relativos aos parâmetros obtidos pelos modelos gerados para cada um dos dados de terreno, conforme Tabela 7.

Tabela 7 – Estatísticas relativas aos parâmetros do modelo GWR

<b>Variável</b>	<b>Mínimo</b>	<b>1º Quartil</b>	<b>Mediana</b>	<b>3º Quartil</b>	<b>Máximo</b>
Intercepto	12,32517	13,77975	14,11005	14,66553	17,88631
ln (ART)	- 1,10288	- 0,72818	- 0,55638	- 0,49022	- 0,28506
ln (TST)	0,20192	0,24238	0,39875	0,46410	0,88314
ln (DPC)	- 0,16997	- 0,08565	- 0,07186	- 0,04209	- 0,02251
ln (DCE)	- 1,09937	- 0,72477	- 0,65544	- 0,58515	- 0,44014
STU	0,02594	0,11609	0,17827	0,32890	0,38876

Fonte: os autores (2021).

Observa-se, ainda, que os parâmetros permanecem com o mesmo sinal (positivo ou negativo), ao longo de toda a área em estudo, o que os torna coerentes.

#### 5.4. QUALIDADE E DESEMPENHO DAS MODELAGENS

Neste estudo, a qualidade das modelagens foi analisada por meio dos critérios de informação de Akaike (AIC), bayesiano (BIC) e o loglikelihood (LIK), o coeficiente de determinação ( $R^2$ ) e o erro padrão da regressão, conforme visto na Tabela 8.

Tabela 8 – Indicadores de qualidade das modelagens realizadas

<b>Indicador</b>	<b>CAR</b>	<b>GWR</b>
Coeficiente de Determinação ( $R^2$ )	0,8023	0,8820
Valor do log Likelihood (LIK)	1,1664	28,540
Critério de Informação de Schwarz (BIC)	29,972	15,382
Critério de Informação de Akaike (AIC)	11,668	- 0,3208
Erro Padrão da Regressão (Se)	0,2360	0,2245

Fonte: os autores (2021).

Percebe-se que o modelo GWR mostrou-se superior ao modelo CAR em todos os indicadores de qualidade de ajuste da Tabela 8. Calculou-se, também, a raiz quadrada do erro quadrático médio (RMSE) para os valores unitários preditos pela amostra de trabalho. Os resultados são apresentados na Tabela 9.

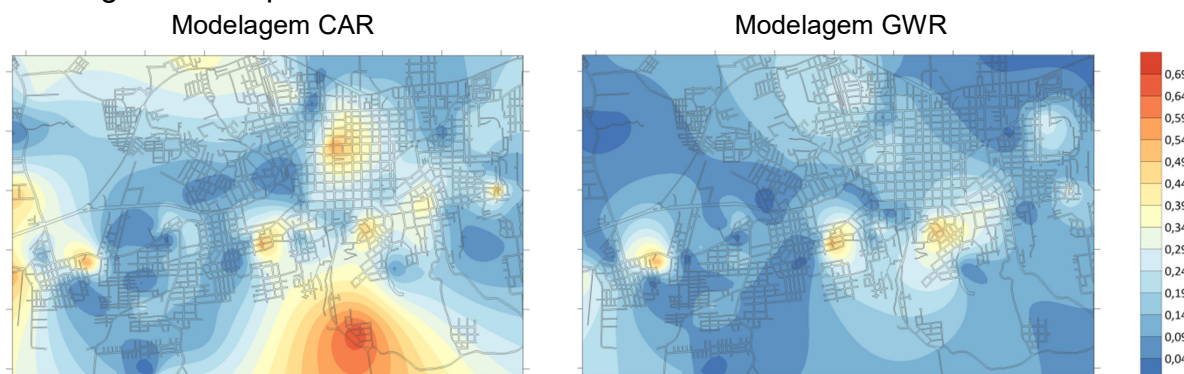
Tabela 9 – Valor do RMSE para os valores preditos pela amostra de trabalho.

<b>Indicador</b>	<b>CAR</b>	<b>GWR</b>
RMSE (R\$/m <sup>2</sup> )	131,68	100,40

Fonte: os autores (2021).

Percebe-se que a modelagem GWR, para esse indicador, apresentou um desempenho muito superior, com valor para o RMSE aproximadamente 23,75% menor que o valor do modelo CAR. Como forma de complementar a análise, realizou-se a interpolação linear para geração da superfície de gradientes, contendo os erros relativos absolutos estimados para cada modelagem, conforme Figura 5.

Figura 5 – Superfície de erros relativos absolutos da amostra de trabalho.

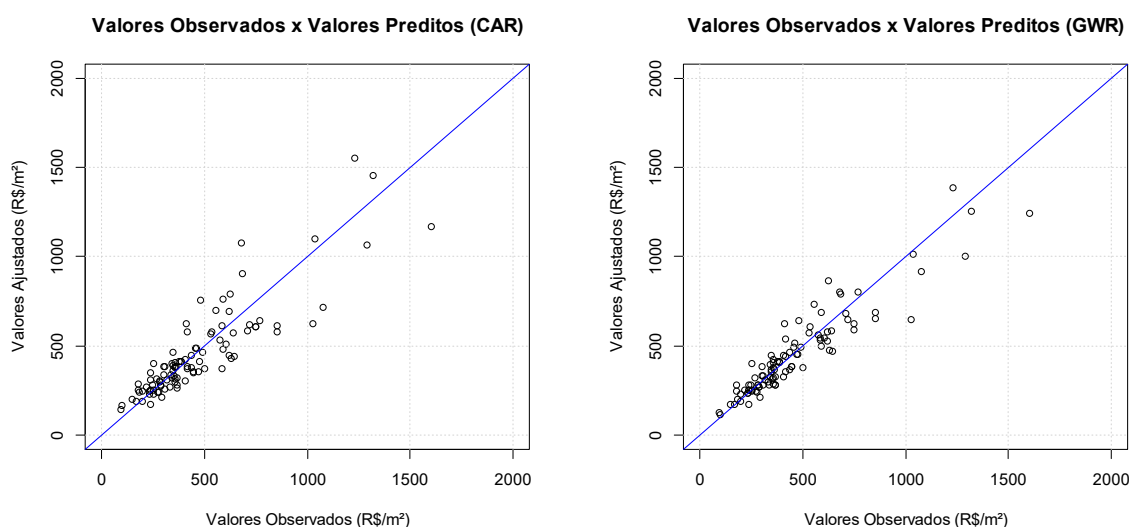


Fonte: Os autores (2020).

Verifica-se, assim, que a superfície de erros relativos absolutos para a modelagem geograficamente ponderada mostrou-se muito mais uniforme, com erros de menor magnitude, mais próximos de zero, evidenciado pela coloração azul.

Como forma de visualizar o poder de predição dos modelos, construiu-se o gráfico de dispersão dos valores observados versus valores ajustados para cada uma das modelagens, conforme visto na Figura 6.

Figura 6 – Diagrama de predição das modelagens CAR e GWR.



Fonte: os autores (2021).

Verifica-se, por meio da Figura 6, que a regressão geograficamente ponderada (GWR) apresentou-se mais ajustada, com pontos menos dispersos, indicando se tratar de um modelo que foi capaz de explicar de forma muito mais eficiente os valores unitários dos imóveis da área em estudo.

Para se analisar o desempenho das modelagens, utilizou-se as métricas estabelecidas pela norma IAAO (2013) sobre os valores preditos. Os limites estabelecidos pela norma, para os imóveis em estudo, são: **A)** Mediana das Razões de Avaliação (Med R)  $\rightarrow 0,90 < \text{Med R} < 1,10$ ; **B)** Coeficiente de Dispersão (COD)  $\rightarrow \text{COD} \leq 25\%$ ; **C)** Diferencial Relativo ao Preço (PRD)  $\rightarrow 0,98 \leq \text{PRD} \leq 1,03$ . Os resultados obtidos para as duas modelagens são apresentados na Tabela 10.

Tabela 10 – Indicadores de desempenho das modelagens realizadas.

Indicador	CAR	GWR
Med R	1,03	1,00
COD (%)	19,30%	14,73%
PRD	1,052	1,028

Fonte: os autores (2021).



Observa-se que as duas modelagens apresentaram nível de avaliação (Med R) e coeficiente de dispersão (COD) dentro dos limites estabelecidos pela norma, entretanto, a modelagem GWR mostrou-se superior, indicando maior uniformidade na avaliação, observada pelo menor valor no COD.

Para o indicador diferencial relativo ao preço (PRD), somente a modelagem GWR apresentou valores dentro dos limites estabelecidos pela norma, demonstrando que essa modelagem manteve a equidade vertical. Sendo assim, observa-se que o modelo GWR apresentou melhores resultados que o modelo CAR.

## 6. ELABORAÇÃO DA PLANTA DE VALORES GENÉRICOS

Para se gerar a superfície que irá compor a planta de valores imobiliários (PVG), deve-se, inicialmente, fazer a homogeneização dos valores preditos pelos modelos de regressão, por meio de um imóvel paradigma representativo da área em estudo. O imóvel paradigma usado neste estudo é mostrado na Tabela 11.

Tabela 11 – Imóvel paradigma utilizado para homogeneização dos imóveis.

	<b>ART</b>	<b>TST</b>	<b>PVT</b>	<b>STU</b>
Paradgma	110	3	2	1

Fonte: os autores (2021).

No processo de homogeneização, as variáveis da Tabela 11 foram mantidas fixas e as variáveis distância ao polo comercial (DPC) e distância ao centro de eventos (DCE) foram variando para cada imóvel. A homogeneização dos valores preditos pelo modelo CAR foi realizada sobre a regressão global da Tabela 5. No modelo GWR realizou-se a homogeneização sobre a regressão do próprio ponto.

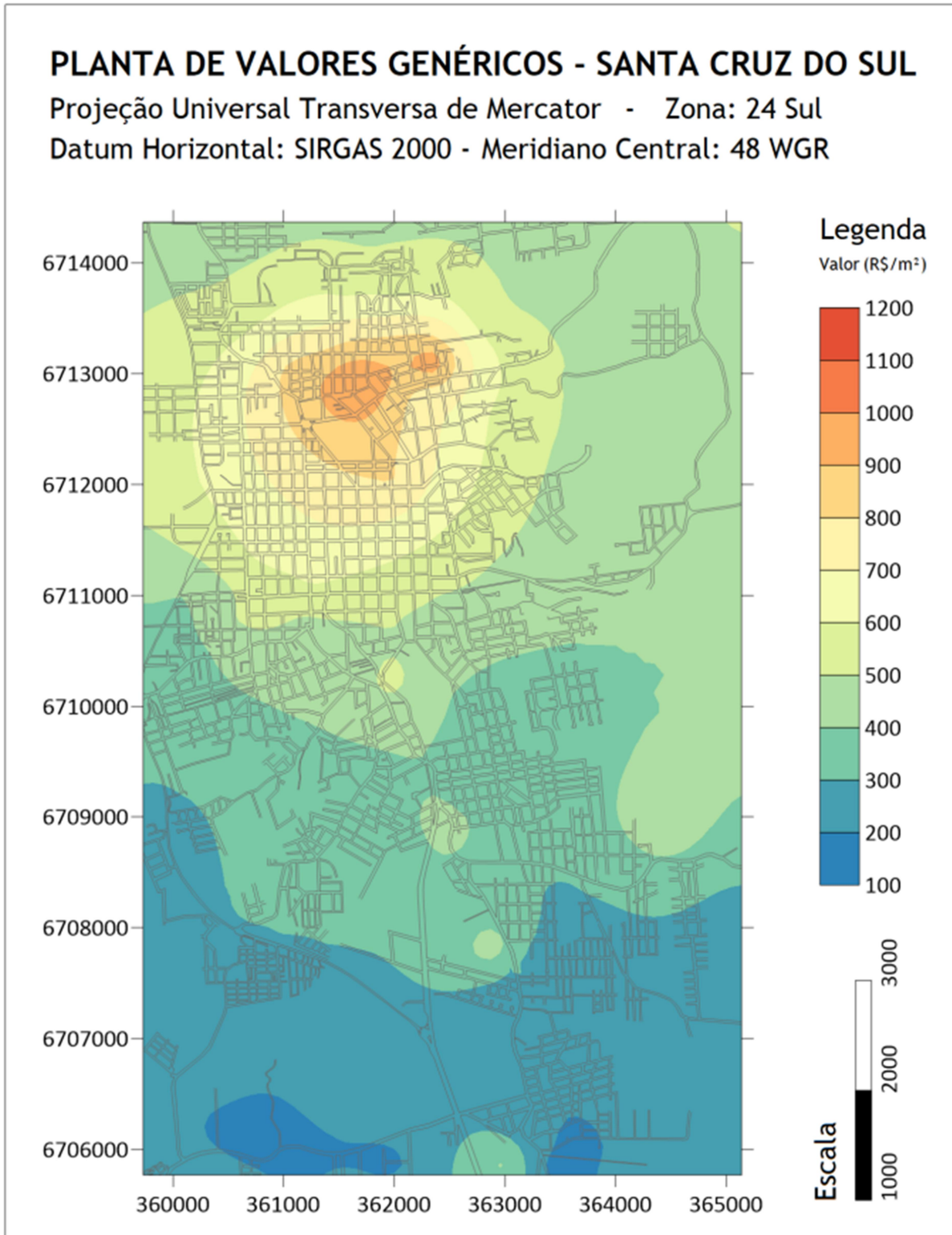
Realizada a homogeneização e o cálculo dos novos valores unitários para os imóveis, procedeu-se com a definição dos semivariogramas. Realizou-se a modelagem exponencial, esférica e gaussiana e por meio de validação cruzada, pode-se verificar que o semivariograma teórico exponencial foi o que apresentou menor RMSE, maior  $R^2$  e a mais forte dependência espacial nas duas modelagens. Nesse estudo, utilizaram-se os semivariogramas exponenciais.

Para a construção das superfícies de valores ajustados pelas duas modelagens, utilizou-se os parâmetros obtidos nos semivariogramas exponenciais e por meio do software *Surfer* 15. Realizou-se a interpolação por krigagem ordinária e pôde-se, então, obter uma planta de valores genéricos (PVG) dos terrenos representativa da área em estudo. Essas PVGs são apresentadas nas Figuras 7 e 8.

Pode-se constatar, pelas regiões mais quentes espalhadas nas proximidades do centro de eventos e da praça central, do município de Santa Cruz do Sul, que tanto a superfície CAR como a GWR foram capaz de explicar de maneira próxima da

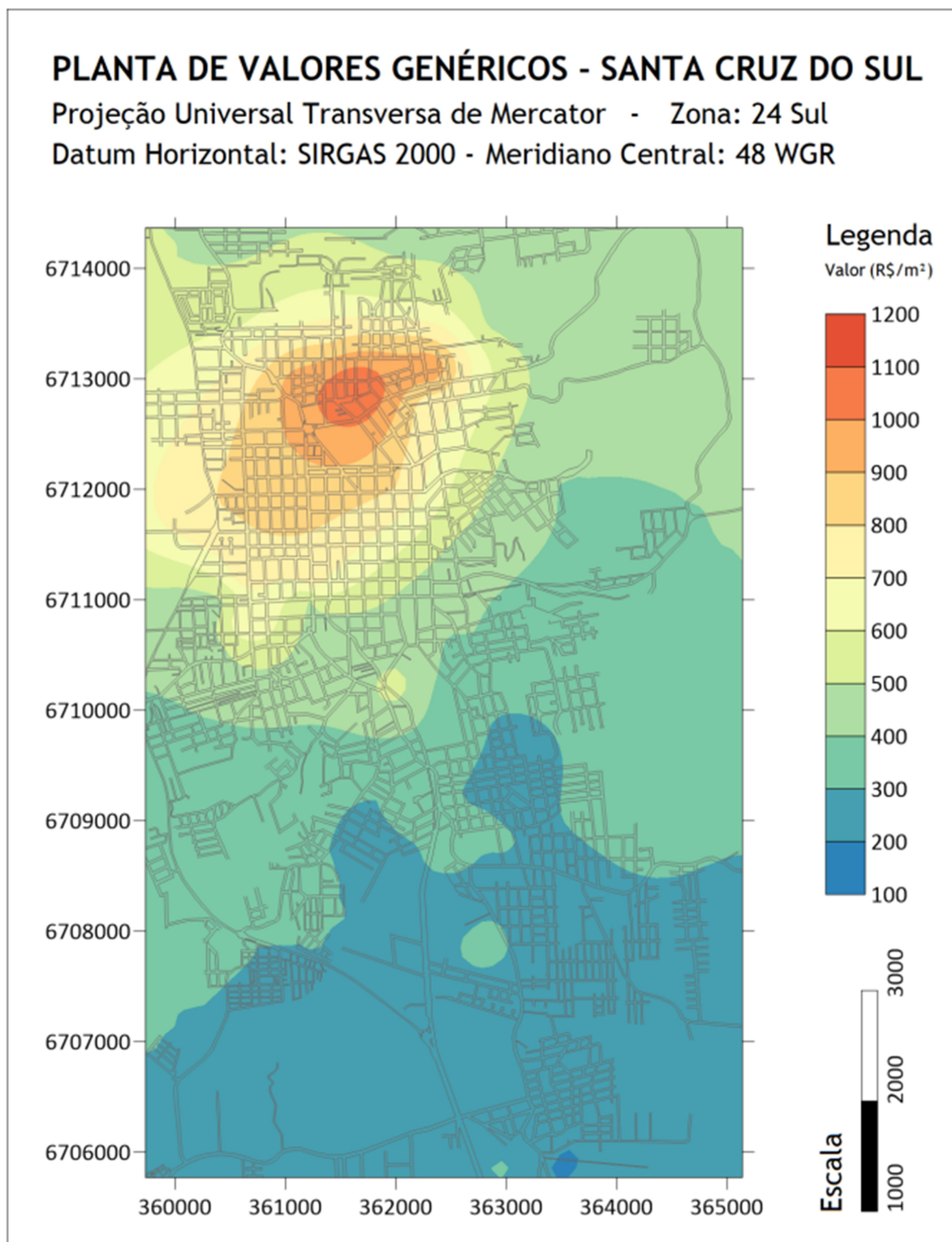
realidade o valor unitário dos imóveis, tendo em vista que, comumente, esta região do município de Santa Cruz do Sul apresenta os terrenos mais valorizados.

Figura 7 – Planta de Valores Genéricos usando krigagem do modelo (CAR)



Fonte: os autores (2021).

Figura 8 – Planta de Valores Genéricos usando krigagem do modelo (GWR)



Fonte: os autores (2021).

Observa-se, no entanto, que a variável localização dos imóveis foi captada de forma muito mais eficiente pela modelagem GWR. Em geral, essa valorização nos

terrenos do município de Santa Cruz do Sul foi melhor capturada pela planta de valores genéricos obtida pela krigagem da modelagem GWR.

Para se avaliar a performance das interpolações realizadas, calculou-se o RMSE dos valores preditos pelas krigagens para a amostra dos 101 terrenos em estudo. Os resultados são apresentados na Tabela 12.

Tabela 12 – Valor do RMSE para os valores preditos pela amostra de trabalho.

<b>Indicador</b>	<b>CAR</b>	<b>GWR</b>
RMSE (R\$/m <sup>2</sup> )	120,58	100,02

Fonte: os autores (2021).

Constata-se que a krigagem dos valores preditos pelo modelo GWR foi a que apresentou o menor valor de RMSE para a amostra de terrenos, indicando ser essa a superfície que apresenta o melhor ajuste dentre as duas modelagens.

## 7. CONCLUSÕES

Com o término deste estudo, pôde-se concluir que a modelagem espacial local (GWR) apresentou uma performance geral superior à modelagem espacial global (CAR) em todos indicadores de qualidade e desempenho analisados. O modelo GWR mostrou-se mais ajustado aos dados da amostra, como demonstrado pelos diagramas de predição e indicadores de qualidade e desempenho apresentados.

Com relação à qualidade, os resultados mostraram que o modelo local foi mais preciso que o modelo global, que considera todos os dados como estacionários. Comparando os modelos, observa-se que a modelagem GWR foi capaz de melhorar o R<sup>2</sup> da modelagem CAR de 0,8023 para 0,8820. O AIC e o BIC da GWR também mostram uma redução de 11,98 e 14,59, respectivamente, em relação à CAR.

O modelo local apresentou, ainda, menor erro padrão e uma superfície de erros relativos absolutos de menor magnitude e mais uniformes, o que é desejável para a geração de uma PVG. O RMSE da regressão local foi 23,75% menor que o global.

Com relação ao desempenho das modelagens em predizer os valores unitários da própria amostra, as duas modelagens apresentaram os parâmetros de nível e uniformidade da avaliação (Med R e COD) dentro das recomendações da IAAO, contudo, a modelagem local mostrou-se superior à global para o parâmetro que mede equidade vertical (PRD), ficando dentro dos limites estabelecidos pela norma.

Deve-se destacar que a modelagem local (GWR) apresentou o menor valor para o COD (14,73%), indicando maior uniformidade e menor dispersão horizontal das razões de avaliação, e valor de PRD (1,028) próximo à unidade (1,000), indicando inexistência de progressividade ou regressividade nas avaliações.

Verificou-se, também, que a planta de valores (PVG) gerada para a modelagem GWR foi capaz de representar de maneira mais eficiente, com resultados mais próximos do esperado, o valor unitário dos terrenos do município em estudo.

A performance das duas superfícies indica que a modelagem GWR se mostrou melhor. Com relação ao RMSE, a krigagem do modelo local se mostrou inferior (R\$ 100,02/m<sup>2</sup>) ao da modelagem global (R\$ 120,58/m<sup>2</sup>).

Os resultados obtidos neste estudo demonstram que a regressão espacial global (CAR) tratou bem o problema da dependência espacial presente nos dados coletados, contudo, a regressão espacial local (GWR) foi capaz de tratar de forma eficiente tanto a dependência quanto a heterogeneidade espacial dos dados.

Isto demonstra que o modelo de regressão local, neste caso em estudo o GWR, se bem ajustado, é capaz de fazer melhores previsões de valores para os imóveis urbanos, incorporando os efeitos espaciais que o modelo global ignora, diminuindo assim as distorções de valores presentes nas plantas de valores genéricos (PVG's), possibilitando, assim, uma tributação mais justa e equânime.

Frente a isso, o método proposto pode ser de grande utilidade para os órgãos municipais responsáveis pelo cálculo e pela atualização de produtos cartográficos, como por exemplo, as Plantas de Valores Genéricos.

## REFERÊNCIAS BIBLIOGRÁFICAS

ABNT. ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. NBR 14653-2: **Avaliação de Bens. Parte 2: Imóveis Urbanos.** Rio de Janeiro, 2011. 53 p.

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

ANSELIN, L., 2005. **Exploring Spatial Data with GeoDa™: A Workbook.** GeoDa 1.6.7 9. March 2015. Spatial Analysis Laboratory, University of Illinois, Urbana Champaign.

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

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

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

CÂMARA, G.; CARVALHO, M.S.; CRUZ, O.G.; CORREA, V. (2002). **Análise Espacial de Áreas.** Em: *Análise Espacial de Dados Geográficos*, eds. Fuks, S.D.; Carvalho, M.S.; Câmara, G.; Monteiro, A.M.V. – Divisão de Processamento de Imagens – Instituto Nacional Pesquisas Espaciais – São José dos Campos, Brasil.

DANTAS, Rubens Alves. **Engenharia de avaliações: uma introdução à metodologia científica.** 3. ed. São Paulo: Pini, 2012. 255 p.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (eds) "**Análise Espacial de Dados Geográficos**". Brasília, EMBRAPA, 2004 (ISBN: 85-7383-20-6).

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

FOTHERINGHAM AS, CHARLTON M, BRUNSDON C, NAKAYA T (2005) **Model selection problems in Geographically Weighted Regression**. Proc., Geocomputation 2005 Conference, Ann Arbor, Mi. EUA

GUJARATI, D. N; PORTER, D. C. **Econometria básica**. 5. ed. Porto Alegre: AMGH Bookman, 2011.

IAAO (International Association of Assessing Officers). **Standards on ratio studies**. Mirrouri: IAAO. 2013.

LEUNG Y.; MEI C. L.; ZHANG W. X. (2000). **Statistical tests for spatial nonstationarity based on the geographically weighted regression**. Environment and Planning, v. 32, 2000.

LIPORONI, A. S. **Avaliação em massa com ênfase em planta de valores**. In: ALONSO, Nelson Roberto Pereira (Coord). Engenharia de Avaliações. 2. ed. São Paulo: PINI, 2013.

MEI, Chang-Lin. **Geographically weighted regression technique for spatial data analysis**. School of Science. ResearchGate. 2004, 14 p.

UBERTI, M. S. **Avaliação em massa de imóveis rurais através de modelagem clássica, espacial e geoestatística**. 2016. 151 f. Tese (Doutorado em Ciência, Tecnologia e Inovação) – Universidade Federal Rural do Rio de Janeiro (UFRRJ), Rio de Janeiro, 2016.