

AVALIAÇÃO EM MASSA COM MODELO ADITIVOS GENERALIZADOS E GEOESTATÍSTICA

Real Estate Mass Appraisal Using Generalized Additive Models and Geostatistics

Antônio Augusto Ferreira de Oliveira
Secretaria Municipal das Finanças de Fortaleza
augusto.oliveira@sefin.fortaleza.ce.gov.br

Terezinha Ferreira de Oliveira
Universidade Federal do Pará – UFPA
tfo@ufpa.br

Sandro Ricardo Vasconcelos Bandeira
Secretaria Municipal das Finanças de Fortaleza
sandro.bandeira@sefin.fortaleza.ce.gov.br

Resumo:

Este trabalho apresenta uma metodologia para avaliação em massa utilizando Modelos Aditivos Generalizados e geoestatística para dados de terrenos urbanos do município de Fortaleza. O Modelo Aditivo Generalizado confirmou algumas hipóteses do comportamento dos preços de mercado imobiliário como aumento em média no valor unitário quando o lote está dentro de um condomínio fechado, com rede de água disponível e quando localizado em via pavimentada. Com este modelo, também foi revelada a não linearidade dos preços unitários com a área, testada, índice de aproveitamento, distância às vias principais, densidade de comercialização, renda e valor base de IPTU da atual PGV. Entretanto, a interação das coordenadas do terreno do modelo não foi suficiente para eliminar a autocorrelação espacial dos resíduos, de modo que foi feito o ajuste de um variograma esférico sobre estes, com posterior interpolação por krigagem ordinária, obtendo-se uma superfície de resíduos a serem somados à predição final. A partir deste procedimento, houve uma melhora em todas as medidas de desempenho utilizadas neste trabalho, principalmente quanto ao COD que passou de 34,40% a 13,06% após a interpolação, o que atende plenamente as recomendações do IAAO para modelos de avaliação em massa. Conclui-se que a metodologia proposta, utilizando Modelos Aditivos Generalizados e interpolação dos resíduos com técnicas de geoestatística, é bastante promissora na avaliação em massa, mas deve ser usada com bastante parcimônia, pois os erros de medição na coleta de dados de mercado podem enviesar a predição final.

Palavras-chave: Modelos aditivos generalizados (GAMs); geoestatística; krigagem, ordinária; avaliação em massa.

Abstract

This work presents a methodology for real estate mass appraisal using Generalized Additive Models and geostatistical techniques for urban land data in the city of Fortaleza. The Generalized Additive Model confirmed some hypotheses of the behavior of real estate market prices such as an average increase in the unit value of condominium lands, with an available water network and when located on a paved road. With this model, the non-linearity of unit prices with respect to area, frontage, legal utilization rate, distance to the main roads, marketing density, income and land tax assessment were also revealed. However, the interaction of the model's land coordinates was not enough to eliminate the spatial autocorrelation of the errors. For that reason a spherical variogram was applied, with subsequent interpolation by ordinary kriging, obtaining a surface of residues to be added to the final prediction. From this methodology, there was an improvement in all the performance measures used in this work. In particular, the coefficient of dispersion (COD), went from 34.40% to 13.06% after the interpolation, which fully meets the recommendations of the IAAO for real estate mass appraisal. We conclude

that the proposed methodology using Generalized Additive Models and interpolation of residues with geostatistical techniques is very promising in real estate mass appraisal. However, as measurement errors in the collection of market data can skew the final prediction, it should be used with caution.

Keywords: Generalized additive models; geostatistics; ordinary kriging; real estate mass appraisal.

1. INTRODUÇÃO

O objetivo da avaliação em massa para fins fiscais é determinar o valor de mercado para uma grande quantidade de imóveis de uma região para tributação imobiliária. Desta forma, as estimativas geradas por estas devem ser precisas para resultar em avaliações uniformes e equânimes.

As abordagens tradicionais de avaliação de imóveis se baseiam em modelo de preços hedônicos, os quais geralmente fazem uso da regressão linear múltipla pelos mínimos quadrados ordinários (MQO). Os pressupostos de utilização dos modelos MQO são muitos, e suas violações nas avaliações de imóveis urbanos causam vieses e inconsistências nos estimadores MQO, o que acarreta a má especificação da forma funcional do modelo. No mesmo diapasão, autocorrelação espacial entre as observações e/ou resíduos do modelo pode(m) trazer sérias consequências ao entendimento do fenômeno (ANSELIN; REY, 2014; WOOLDRIDGE, 2016).

Outrossim, é comum nos dados de mercado imobiliário se observar a assimetria à direita dos preços observados e, por isso, o uso de transformação logarítmica naqueles como forma de se estabelecerem relações supostamente lineares entre a variável resposta e seus preditores. Entretanto, estas transformações muitas vezes são ineficientes, pois não retratam de maneira eficaz o relacionamento subjacente.

Uma alternativa a essa problemática seria a utilização dos Modelos Aditivos Generalizados (GAM) com uma distribuição que mais se assemelha a forma funcional dos dados de mercado imobiliário, como por exemplo a distribuição gama e função de ligação logarítmica (VEIE; PANDURO, 2013) e posterior tratamento da autocorrelação espacial dos resíduos obtidos com técnicas de geoestatística para proposição de plantas de valores genéricos (PGV) mais precisas a serem utilizadas na tributação do imposto predial e territorial urbano (IPTU) e do imposto de transmissão *inter vivos* (ITBI).

Nesse contexto, este trabalho parte de uma amostra composta por 1.924 dados de terrenos urbanos colhidos no mercado imobiliário de Fortaleza, bem como das avaliações e declarações para lançamento ITBI efetuadas na Secretaria das Finanças de Fortaleza (SEFIN) para aplicar a metodologia supracitada. Escolheu-se a tipologia “terrenos” para este estudo dado sua dificuldade de modelagem, seja pela ausência confiável de dados disponíveis, bem como por essa tipologia ser necessária na tributação de cada imóvel do cadastro municipal, independentemente de ser edificado ou não, no uso do método evolutivo, usualmente empregado na maioria dos municípios brasileiros na tributação do IPTU.

2. MATERIAS E MÉTODOS

2.1. Área de estudo e descrição dos dados

O Município de Fortaleza é a capital do estado do Ceará, com população estimada em 2,6 milhões de pessoas, segundo projeção do IBGE para o ano 2018, sendo a quinta mais populosa cidade do Brasil. Possui área territorial de 314,93 km², o que torna a mais adensada entre todas as capitais com 7.786,44 hab./km², segundo o censo de 2010.

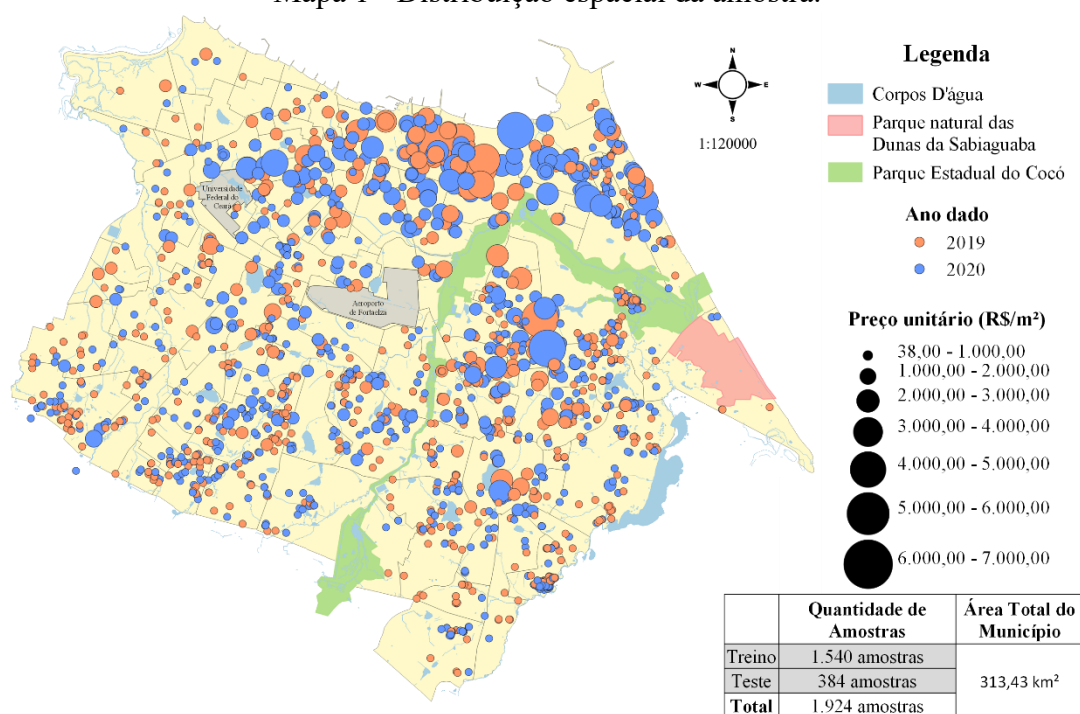
De acordo com o cadastro imobiliário municipal da SEFIN (CIM), a quantidade total de

lotes é de aproximadamente 385 mil, dentro os quais cerca de 74 mil são não edificadas (terrenos vazios). O número de inscrições municipais sujeitas à tributação do IPTU é de, aproximadamente, 779 mil. Muito embora o número de imóveis territoriais represente menos que 10% do total de inscrições, salienta-se que a tributação do IPTU em Fortaleza se dá pelo método evolutivo, onde a composição do valor venal, base de cálculo desse imposto, se dá pela soma do valor da parte territorial e da parte edificada depreciada.

Os 1.924 dados para composição de amostra para esse trabalho foram disponibilizados pela SEFIN no período compreendido de julho de 2019 a julho de 2020. A amostra abrange 111 bairros do município, do total de 121, e pode ter sua distribuição espacial visualizada através do

Mapa 1:

Mapa 1 - Distribuição espacial da amostra.



Fonte: elaboração própria.

A variável explicada foi o preço unitário do terreno em R\$/m², representando a divisão do preço observado (R\$) dividido pela sua área (m²). As variáveis explicativas utilizadas para as modelagens encontram-se na Tabela 1:

Tabela 1 – Descrição das variáveis testadas/utilizadas no modelo.

Variáveis	Descrição	Mín.	Média	Máx.
<i>vunit</i>	Valor unitário do terreno em R\$/m².	38,40	865,06	7.000,00
Variáveis estruturais				
<i>numfr</i>	Número de frentes.	1,00	1,43	9,00
<i>test</i>	Comprimento da testada principal em m.	3,00	28,21	880,00
<i>area</i>	Área do terreno em m².	23,04	2.206,57	330.000,00
Variáveis espaciais e de localização				

Variáveis	Descrição	Mín.	Média	Máx.
<i>xcen</i>	Centroide X do lote relativo ao centroide do próprio município e dividido por 1.000 (<i>datum</i> SIRGAS 2000, UTM fuso 24S).	- 1,16	0,09	1,26
<i>ycen</i>	Centroide Y do lote relativo ao centroide do próprio município e dividido por 1.000 (<i>datum</i> SIRGAS 2000, UTM fuso 24S).	- 1,11	- 0,09	0,92
<i>distvp</i>	Distância em m à via principal mais próxima.	4,46	279,67	1.532,50
<i>dscm</i>	Densidade de comercialização de lotes que estão no mesmo trecho de logradouro do lote corrente.	0,01	0,23	1,00
<i>vbt</i>	Valor de referência do terreno do IPTU.	3,02	56,51	1.532,69
<i>renda</i>	Renda do responsável krigeada IBGE 2010.	0,40	3,46	18,89
<i>idhed</i>	IDH, dimensão educação por bairro.	0,65	0,80	0,95
<i>iaeq</i>	Índice de aproveitamento equivalente.	0,01	1,66	3,00
<i>lotcnd</i>	0- <i>fora</i> de condomínio; 1- <i>em</i> condomínio.	0		1
<i>esq</i>	0- <i>não</i> é esquina; 1- <i>é</i> esquina.	0		1
<i>agua</i>	0- <i>sem</i> água; 1- <i>com</i> água.	0		1
<i>esg</i>	0- <i>sem</i> esgoto; 1- <i>com</i> esgoto.	0		1
<i>pav</i>	0- <i>sem</i> pavimentação asfáltica (ou concreto) no trecho de logradouro; 1- <i>com</i> pavimentação asfáltica (ou concreto) no trecho de logradouro.	0		1
Variáveis de informação				
<i>a2020</i>	0- <i>informação</i> do ano de 2019; 1- <i>informação</i> do ano de 2020.	0		1
<i>Tr</i>	0- <i>Não</i> é transação; 1- <i>Transação</i> .	0		1
<i>of</i>	0- <i>Não</i> é oferta; 1- <i>Oferta</i> .	0		1

Fonte: elaboração própria.

2.2. Modelos aditivos generalizados (GAM)

Os modelos aditivos generalizados (GAM) são extensões semi-paramétricas dos modelos lineares generalizados (GLM) que utilizam uma função de ligação para estabelecer uma relação entre a média da variável resposta e uma função de “suavização” para as variáveis explicativas (HASTIE; TIBSHIRANI; FRIEDMAN, 2008). A média condicional $\mu(X)$ da resposta Y está relacionada a uma função aditiva dos preditores por meio de uma função de ligação g definida na equação (1):

$$g[\mu(X)] = E(Y|X_1, X_2, \dots, X_n) = \alpha + f_1(X_1) + f_2(X_2) + \dots + f_n(X_n) \quad (1)$$

onde X_1, X_2, \dots, X_n representam as variáveis explicativas e Y a variável resposta. Os f_j são funções não paramétricas que podem ser ajustadas por meio de um conjunto de *splines* bases (v.g. de suavização cúbico ou um *kernel* mais suave). Formalmente, essas funções de suavização são dadas pela soma de k *splines* bases (b_h na equação 2) multiplicados pelos seus respectivos coeficientes a serem estimados (β_h na equação 2) (VEIE; PANDURO, 2013):

$$f = \sum_{h=1}^k \beta_h b_h \quad (2)$$

A aditividade desse tipo de modelo facilita a interpretação e permite estabelecer as relações não-lineares e não monótonas entre a variável resposta e o conjunto de variáveis explicativas (ibidem). Para as variáveis contínuas, foram utilizadas funções de suavização, permitindo a observação de suas relações com o preço unitário (penúltimo termo da equação

3), conforme se verá na seção 3. Quanto às variáveis dicotômicas (segundo termo da equação 3), permaneceram na sua forma original, permitindo captar seu efeito marginal diretamente.

Ressalta-se também a utilização da função de ligação na forma logarítmica para o GAM, já que esta é muito utilizada nas análises hedônicas. A incorporação das coordenadas UTM no modelo, a fim de captar o efeito de localização no comportamento dos preços observados, seguiu a metodologia de Veie e Panduro (2013), ou seja, através da introdução da função de suavização com a interação das coordenadas de localização espacial. Assim o modelo GAM final teve a seguinte forma:

$$\ln(v_{unit}) = \beta_0 + \beta_i D_i + f_1(xcen, ycen, k) + f_j(X_j, k) + \varepsilon_i \quad (3)$$

, onde D_i são o conjunto de variáveis explicativas *dummies*, e X_j o conjunto de variáveis explicativas (e contínuas) que serão suavizadas pela soma de um número k de *splines* bases.

Convém ainda observar, que a especificação de um modelo GAM requer a escolha de uma família de distribuição de funções exponenciais (gama, gaussiana, gaussiana inversa, poisson etc.), tendo sido escolhido a distribuição gama pelos preços de imóveis serem sempre positivos com variância crescente com o preço (ibidem). Foi utilizado o pacote “mgcv” do R (WOOD, 2010) para seleção otimizada do nível de suavização dos preditores (valor de k) a partir da minimização do Critério de Informação Bayesiano – BIC, bem como os Graus de liberdade Efetivos (EDF). O método para estimação preditores e dos coeficientes dos fatores fixos do modelo foi o *REstricted Maximum Likelihood* (REML) (ibidem).

O modelo GAM foi estimado sobre os 80% dados e procedeu-se a predição sobre os restantes 20%. As medidas de desempenho para a escolha do modelo e predição final utilizadas neste trabalho foram as recomendadas pelos padrões da *International Association of Assessing Officers* (IAAO, 2010), bem como outras métricas já consagradas na literatura, quais sejam: o nível de avaliação mediano, coeficiente de dispersão (COD), média percentual absoluta do erro (MAPE), raiz quadrada da média dos erros ao quadrado (RMSE) e coeficiente de determinação R^2 .

Não obstante a utilização de variáveis predictoras de localização, os resíduos podem estar autocorrelacionados espacialmente, situação em que as estimativas geradas pelo modelo GAM não foram suficientes para captar toda a dependência espacial dos preços observados. Assim, testou-se essa hipótese [de dependência espacial] com a quantificação desta através do índice global I de Moran (Moran’s I) a seguir descrito.

2.3 Verificação da autocorrelação espacial dos resíduos

É sabido que a localização é um dos principais atributos para a explicação dos preços do mercado imobiliário, valendo a famosa expressão *location, location and location are everything in real estate*. Dessa maneira, é comum que os preços e/ou os erros da modelagem avaliatória sejam espacialmente autocorrelacionados. Não obstante a introdução de variáveis de localização e vizinhança nos modelos preditivos, essa dependência espacial pode ser feita por meio do índice global de Moran I (ARBIA, 2014). O índice global de Moran I pode ser entendido como o coeficiente angular da reta de regressão linear obtida por MQO da variável independente de interesse (no presente caso, dos resíduos do modelo GAM) com a variável dependente “defasada” desses mesmos resíduos. Essa última (W_y) é obtida pelo produto vetorial de uma matriz de vizinhança de peso ponderado (W) sobre esses erros. Em outras palavras, W_y representa o erro ponderado dos vizinhos sobre a observação em questão. Destarte, o índice global de Moran I é o β_1 da equação 4 e sua estatística está representada na equação 5.

$$W_y = \beta_0 + \beta_1 y + \varepsilon \quad (4)$$

$$I = n \frac{\sum_{i=1}^n \sum_{j=1}^n w_{ij} (y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (5)$$

onde n representa o número de observações, y_i é a variável de interesse e w_{ij} é o elemento da matriz de vizinhança espacial entre as observações i e j .

2.4. Matriz de vizinhança

Neste trabalho, optou-se por uma matriz de vizinhança pelo inverso da distância entre os centroides dos lotes dos terrenos, conforme recomendado na literatura (c.f. DANTAS, 2003 e ALMEIDA, 2012)¹. Entretanto, as relações de vizinhança não se deram com todos os dados, mas foram restringidas por um “raio de contágio espacial” de tal forma que todos os dados contivessem pelo menos 1 vizinho.

2.5 Variograma e Krigagem ordinária

O semivariograma é uma função utilizada para medir as relações de dependência espacial que existem na amostra de dados (ANDRIOTTI, 2009). Ele é construído pela medição das covariâncias que existem entre os dados quando tomados dois a dois e dentro de uma distância h (ibidem). Por definição, é dado pela equação abaixo:

$$\gamma(h) = \frac{1}{2} E\{[Z(x_i + h) - Z(x_i)]^2\} \quad (6)$$

, onde h é representa um vetor entre dois pontos no espaço. À medida que h aumenta, $\gamma(h)$ também aumenta até um valor máximo no qual se estabiliza, o que se denomina de patamar (*sill*). A distância na qual $\gamma(h)$ atinge o patamar é chamada de alcance (*range*) e representa a distância limite da dependência espacial. Medições localizadas a distâncias maiores que o alcance possuem distribuição espacial aleatória e por isso são independentes entre si. Construída a matriz de vizinhança pelo inverso da distância entre os vizinhos de cada dado amostral², calculou-se o índice global de Moran I . Refutou-se a hipótese nula de não dependência espacial dos resíduos, justificando, desta feita, um tratamento espacial deles por algum critério. Escolheu-se interpolá-los por krigagem³ e depois reincorporá-lo à predição inicial, a fim de minimizar os erros decorrentes de possíveis variáveis de localização não utilizadas na modelagem inicial.

A técnica de Krigagem Ordinária (KO) considera que os dados espacialmente distribuídos podem ser avaliados em qualquer local a partir de uma pequena amostras, e as predições em um determinado ponto e posição $\hat{z}(s_0)$ de um local não observado s_0 são obtidas

¹ As matrizes de vizinhança baseadas em inverso da distância reproduzem mais fielmente a “Primeira Lei da Geografia” de Waldo Tobler: “todas as coisas estão relacionadas com todas as outras, mas coisas próximas estão mais relacionadas do que coisas distantes” (TOBLER, 1970).

² Nem sempre é possível a criação de matrizes de vizinhança com o raio estipulado pelo valor do alcance do variograma. Nas situações em que o alcance é menor que a distância mínima para que todos os dados possuam pelo menos 1 (um) vizinho, é um exemplo.

³ A técnica de krigagem foi desenvolvida formalmente (matematicamente) pelo francês Georges Matheron, considerado o fundador da geoestatística, o qual homenageou o nome da técnica de “krigagem” em consideração ao engenheiro de minas sul-africano, Daniel Kriging, o primeiro a utilizar a técnica de maneira experimental.

a partir dos pesos médios λ_i de determinadas observações s_i (WEBSTER e OLIVER, 2007):

$$\hat{z}(s_0) = \sum_{i=1}^n \lambda_i \times z(s_i) \quad (7)$$

Os pesos λ_i (com soma igual a unidade) são escolhidos para que a variância do erro (σ^2) entre os valores observados s_i e os amostrados (s_0) seja minimizada (equações 8).

$$\sigma^2 = \sum_{i=1}^n \lambda_i + \gamma(s_i, s_0) + \varphi, \text{ onde } \sum_{i=1}^n \lambda_i = 1 \quad (8)$$

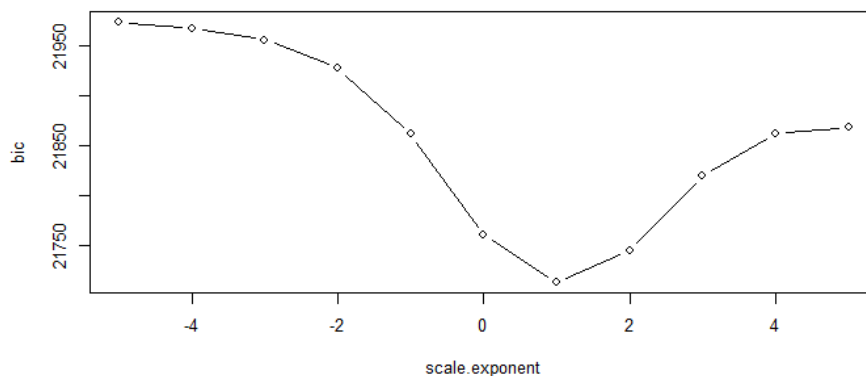
O φ é o multiplicador de Lagrange que otimiza a variância da krigagem e depende da estrutura de autocorrelação espacial dos dados $\gamma(s_i, s_0)$ obtida na equação (6).

3. RESULTADOS E DISCUSSÕES

3.1. Modelos aditivos generalizados (GAM)

Como supracitado, a seleção otimizada do nível de suavização dos preditores (valor de k) foi obtida a partir da minimização do Critério de Informação Bayesiano – BIC. A Figura 1 mostra a variação do BIC com os expoentes de suavização das *splines*.

Figura 1 – Variação do BIC com o nível de suavização dos preditores.

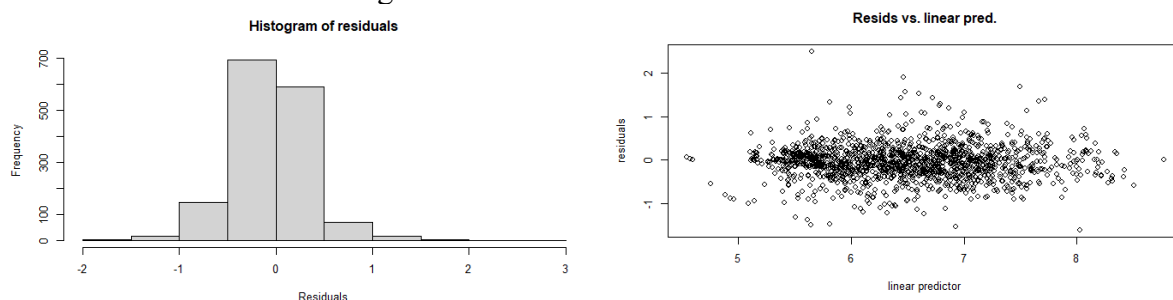


Fonte: elaboração própria.

A equação (8) mostra o modelo GAM com 77,30% de explicação da *deviance* com os fatores fixos e as funções suavizadas que foram significativas. A Tabela 2 mostra os valores dos coeficientes dos fatores fixos e suas estatísticas. Os resíduos do modelo seguem, aproximadamente, uma distribuição normal (ver Figura 2).

$$\begin{aligned} \ln(vunit) = & 6,12 + 0,44lotcnd + 0,07agua + 0,09pav + 0,05a2020 \\ & + 0,20tr + 0,28of + f_1(xcen,ycen) + f_2(distvp) \\ & + f_3(dscom) + f_4(vbt) + f_5(renda) + f_6(taeq) + f_7(area) \\ & + f_8(test) \end{aligned} \quad (9)$$

Figura 2 – Resíduos do modelo GAM.



Fonte: elaboração própria.

Tabela 2 - Estimativa, erro padrão e valores de p dos fatores fixos

Parâmetro	Estimativa	Erro-padrão	p-value
β_0	6,12	0,04	0,000
β_1 -lotend	0,44	0,09	0,000
β_2 -agua	0,07	0,03	0,040
β_3 -pav	0,09	0,03	0,002
β_4 -a2020	0,05	0,02	0,031
β_5 -tr	0,20	0,05	0,000
β_6 -of	0,28	0,03	0,000

Fonte: elaboração própria.

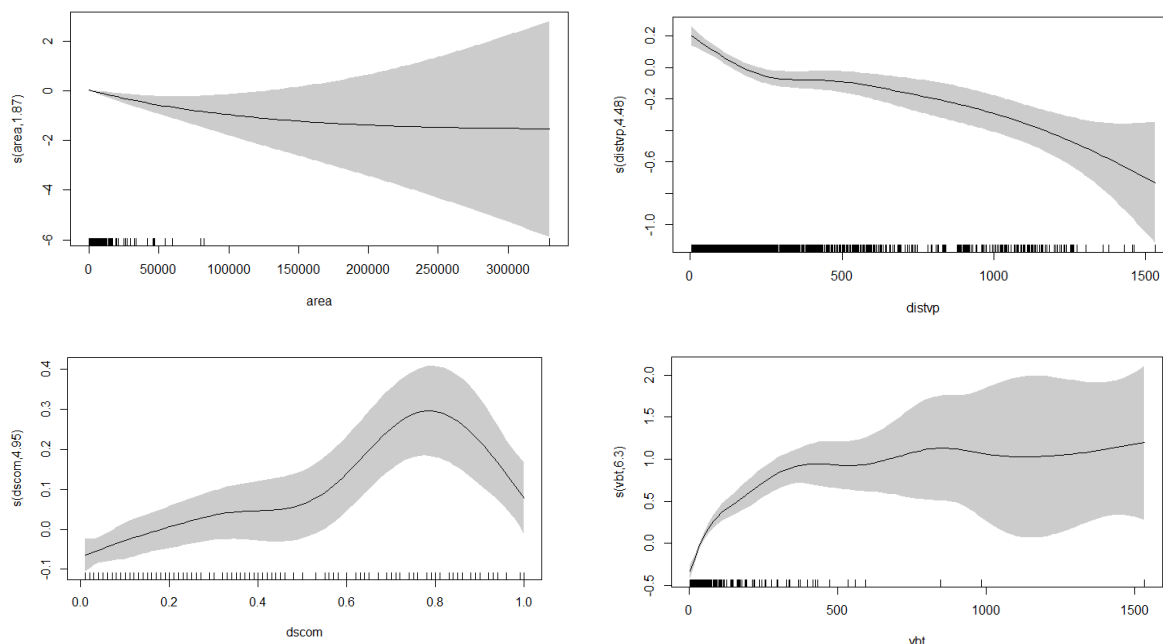
Na Tabela 2, as estimativas dos fatores fixos do modelo, representadas pelas variáveis *dummies* (0-sem e 1-com), foram todas positivas, conforme o esperado, sendo o seu efeito marginal calculado pela expressão (para função de ligação logarítmica):

$$\text{efeito marginal} = (e^{\beta_i} - 1) \times 100 \quad (10)$$

Assim foi possível observar (em média): que há um aumento no valor unitário de 55,32% quando o lote está dentro de um condomínio fechado; quando o terreno tem rede de água disponível valor unitário aumenta 7,17% em relação aqueles que não a tem; terrenos localizados em via pavimentadas são 9,69% mais valorizados; os valores dos preços colhidos em 2020 são 21,66% mais elevados que os de 2019; os preços de efetiva transação são 21,66% mais elevados que as avaliações de ITBI e os referentes aos dados de oferta são 32,41% mais elevados que essas mesmas avaliações. Todos os crescimentos apresentados confirmam as hipóteses iniciais de comportamento dos preços de mercado imobiliário.

A Figura 3 demonstra os relacionamentos não lineares das variáveis área (*area*), distâncias à via principal (*distvp*), densidade de comercialização (*dscm*) e o valor base de terreno da PGV de 2014 (*vbt*) e suas respectivas funções de suavização, permitindo algumas considerações relevantes.

Figura 3 – Demonstração das relações não lineares das variáveis *area*, *distvp*, *dscom* e *vbt*.

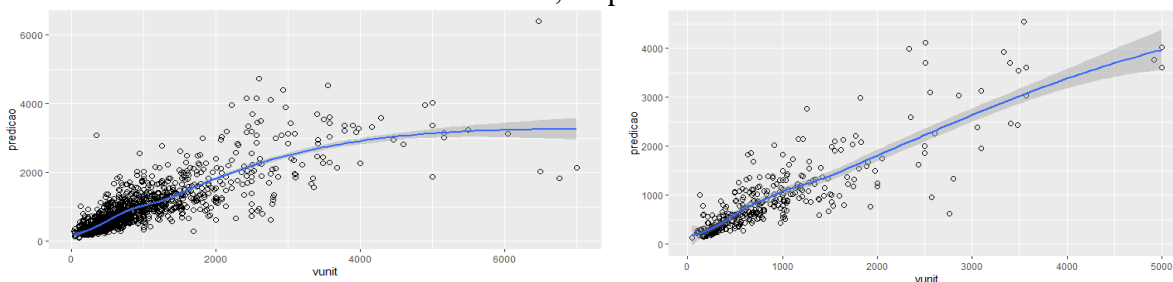


Fonte: elaboração própria.

Observa-se que o valor unitário do terreno decresce com área (princípio da utilidade marginal decrescente); o valor unitário decresce rapidamente até a uma distância aproximada de 400m à via principal e depois passa a ter um decaimento mais suave; o valor unitário cresce até 80% de imóveis comerciais no trecho do logradouro, passando decair a partir daí até 100% e o valor base de terreno da PGV influencia expressivamente os valores unitários de terreno até R\$ 400/m², se estabilizando a partir daí.

As previsões finais do modelo GAM se encontram na Figura 4, onde se observa um maior erro para valores unitários observados superiores a R\$ 2.500/m².

Figura 4 – Dispersão do observado (*vunit*) vs. predição do modelo GAM nas amostras de treino e teste, respectivamente.

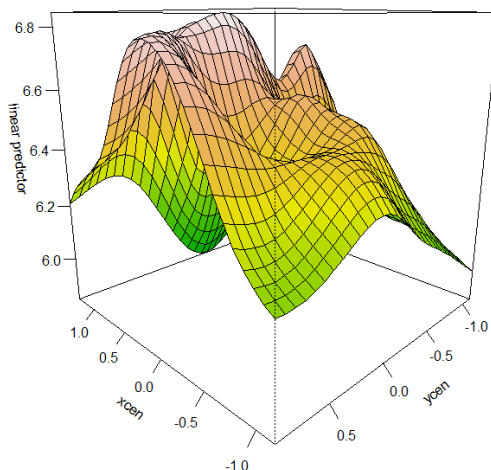


Fonte: elaboração própria.

A Figura 5 apresenta um mapa tridimensional das previsões do modelo GAM mantendo-se todas as demais variáveis fixas, com exceção das coordenadas centralizadas, *xcen* e *ycen*. Verifica-se, através dela, o comportamento das previsões ao longo da superfície do município, indicando aumento de preços com valores elevados de *ycen* (aproximação do litoral norte da Av. Beira Mar) e decaimento dos mesmos nas direções leste e oeste, sendo, neste último [oeste], observado um maior decréscimo (o litoral leste com a “Praia do Futuro” tem preços unitários

observados maiores).

Figura 5 – Mapa de contorno das predições do modelo GAM com variação apenas das variáveis x_{cen} e y_{cen} .

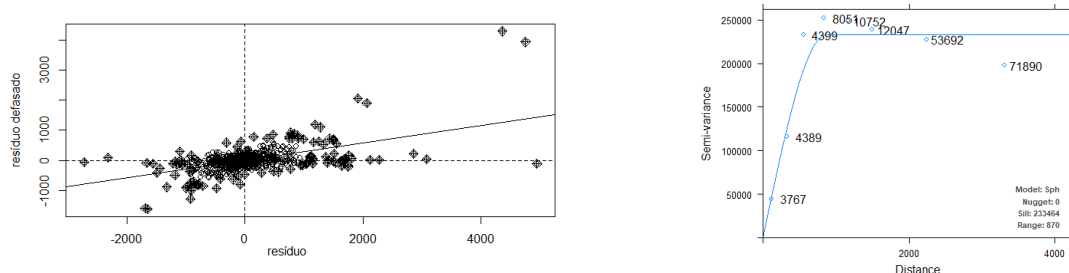


Fonte: elaboração própria.

3.2 Índice de Moran's I dos resíduos da regressão GAM e Krigagem Ordinária.

Como afirmado anteriormente, construiu-se uma matriz de vizinhança com a distância de 1.542m, a qual é a distância mínima para que todos os dados possam ter pelo menos um vizinho. Rejeitou-se a hipótese nula de não autocorrelação espacial dos resíduos a 1% ($pvalue \sim 0$), com a estatística de Moran's I igual a 0,2906 (autocorrelação positiva). O diagrama de dispersão de Moran, se encontra na Figura 6 (esquerda). Em razão disso, verifica-se a necessidade de tratamento destes resíduos para melhorar a performance da predição final. Para tanto, ajustou-se um variograma teórico esférico com pepita=0, patamar de 233464 m e alcance de 870m, conforme Figura 6 (direita) e promoveu-se a interpolação daqueles resíduos por krigagem ordinária.

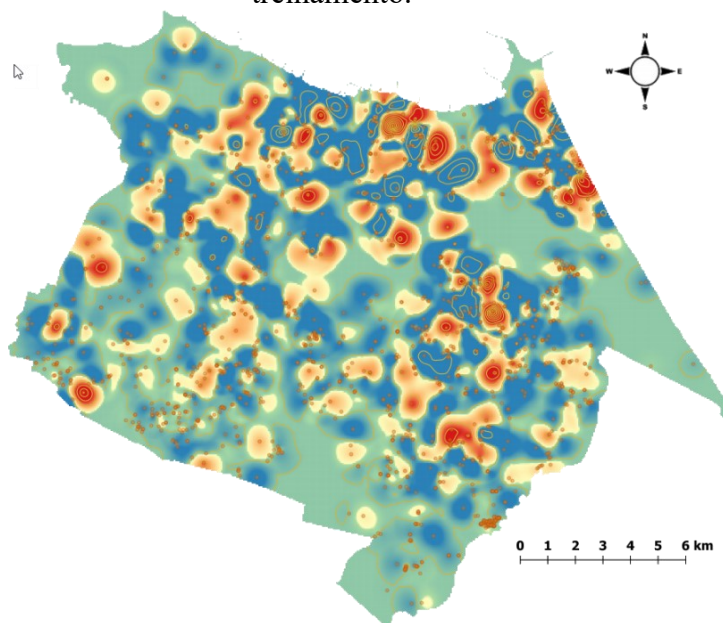
Figura 6 – Diagrama de dispersão de Moran's I e variograma teórico.



Fonte: elaboração própria.

A Figura 7 mostra o mapa de resíduos do modelo GAM (amostra de treinamento) interpolados por krigagem ordinária. A partir dessa interpolação, a predição final foi calculada por meio da soma do seu valor inicial de predição com o resíduo interpolado, tanto para a amostra de treinamento como amostra de teste.

Figura 7 - Mapa de krigagem ordinária dos resíduos do modelo GAM sobre dados de treinamento.



Fonte: elaboração própria.

Por sua vez, os indicadores de performance destas predições foram calculados para cada um dos modelos e estão apresentados na

Tabela 3.

Tabela 3 – Comparativo de desempenho dos modelos GAM com e sem interpolação dos resíduos por krigagem ordinária (KO).

Métrica de Performance	GAM		GAM KO	
	Treino	Teste	Treino	Teste
Nível de avaliação	1,05	1,06	1,00	1,00
COD	35,76%	34,40%	7,18%	13,06%
RMSE	474,35	413,14	112,04	204,48
MAE	262,08	253,93	50,43	90,48
MAPE	37,70%	36,71%	7,19%	13,10%
R ²	0,68	0,73	0,98	0,93

Fonte: elaboração própria.

A

Tabela 3 mostra uma sensível melhora dos indicadores de performance sobre as predições de teste quando somadas aos seus respectivos resíduos interpolados. O coeficiente de dispersão (COD) passou para 13,06% o que atende as recomendações do IAAO (máximo de 30%) tendo 93% da variabilidade dos preços observados explicados pelas variáveis do modelo.

4. CONCLUSÃO

Neste trabalho, utilizou-se uma metodologia aplicada a uma amostra de 1.924 dados de terrenos urbanos de Fortaleza obtidos no mercado imobiliário, das avaliações e declarações

para lançamento ITBI.

Inicialmente foi conduzido um Modelo Aditivo Generalizado (GAM) com o qual se confirmou algumas hipóteses do comportamento dos preços de mercado imobiliário como aumento em média no valor unitário de 55,32% quando o lote está dentro de um condomínio fechado; de 7,17% quando tem uma rede de água disponível e 9,69% quando está localizado em via pavimentada. Observou-se que a área, testada, índice de aproveitamento, distância à via principais, densidade de comercialização, renda e valor base de IPTU da atual PGV não guardam relação linear com os preços unitários, cada uma tendo uma determinada curva de efeito marginal.

Apesar de ter sido incorporado no modelo GAM um termo de interação das coordenadas do terreno, esta não foi suficiente para eliminar a autocorrelação espacial dos resíduos. Assim, foi feito o ajuste de um variograma esférico sobre estes resíduos e promovendo-se a interpolação destes pela krigagem ordinária, obtendo-se uma superfície de resíduos a serem somados à predição final.

Com esse procedimento, houve uma melhora substancial em todas as medidas de desempenho utilizadas neste trabalho como o COD que de 34,40% passou a 13,06%, o que atende as recomendações do IAAO (máximo de 30%). O MAE, de 253,93, passou a 90,48 e R² de 73% passou a 93%.

A metodologia proposta utilizando Modelos Aditivos Generalizados e geoestatística para dados de terrenos urbanos do município é bastante promissora para a utilização na avaliação em massa desta natureza, mas deve ser usada com bastante acuidade, pois os erros de medição na coleta de dados podem enviesar a predição final.

Referências

ANDRIOTTI, José Leonardo Silva. **Fundamentos de Estatística e Geoestatística**. São Leopoldo: Editora Unisinos, 2009. 165 p.

ANSELIN, Luc; REY, Sergio J. **Modern Spatial Econometrics in Practice: a guide to GeoDa, GeoDaSpace and PySAL**. Geoda Press LLC. Chicago, IL, 2014. 368 p.

ARBIA, Giuseppe. 2014. **A Primer for Spatial Econometrics With Applications in R**. London : Palgrave Macmillan, 2014. 230 p.

DANTAS, Rubens A. **Modelos Espaciais aplicados ao mercado habitacional** – um estudo de caso para a cidade de Recife. Recife, 2003 (Tese – Universidade Federal de Pernambuco – UFPE).

GOLGHER, André B. **Introdução à econometria espacial**. Jundiaí: Paco Editorial, 2015, 384 p.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, Jerome. **The elements of statistical learning: data mining, inference, and prediction**. 2. ed. Standford: Springer, 2008. 764 p.

INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS (IAAO). 2010. **Standards on Ratio Studies**. International Association of Assessing Officers: Kansas City, Missouri, US.

TOBLER, W. R.. **A Computer Movie Simulating Urban Growth in the Detroit Region**. Economic Geography, [S.L.], v. 46, p. 234-234, jun. 1970. JSTOR.

Anais do COBRAC 2020 - Florianópolis – SC – Brasil - UFSC – de 09 a 12 de novembro 2020

<http://dx.doi.org/10.2307/143141>.

TEAM, R. C.: 2012, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.

VEIE, Kathrine Lausted, PANDURO, Toke Emil. **An alternative to the standard spatial econometric approaches in hedonic house price models**, 2013. Disponível em: www.ifro.ku.dk/english/publications/foi_series/working_papers/. Acesso: 15 de agosto de 2020.

WOOD, Simon N.. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. **Journal Of The Royal Statistical Society: Series B (Statistical Methodology)**, [S.L.], v. 73, n. 1, p. 3-36, 14 set. 2010. Wiley. <http://dx.doi.org/10.1111/j.1467-9868.2010.00749.x>.

WOOD, Simon N.. **Generalized Additive Models: an introduction with r**. London: CRC Press, 2006. 397 p.

WOOLRIDGE, Jeffrey M. **Introdução à econometria: uma abordagem moderna**. 6. ed. São Paulo: Cengage Learning, 2016. 848 p.