AESTIMUM
CENTRO STUDI DI ESTIMO E DI ECONOMIA TERRITORIALE - Ce.S.E.T.

Firenze University Press
www.fupress.com/ceset

OPEN ACCESS

Citation: Gonzalez, M. A. S. & Erba, D. A. (2024). Alternative methods for measuring the influence of location in hedonic pricing models. *Aestimum* 85: 55-71. doi: 10.36253/aestim-15472

Received: December 1, 2023

Accepted: November 26, 2024

Published: February 14, 2025

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

ORCID:
MASG: 0000-0002-1975-0026
DAE: 0000-0003-1222-1401

Alternative methods for measuring the influence of location in hedonic pricing models

Marco Aurelio Stumpf Gonzalez[1,*], Diego Alfonso Erba[2]

[1] *Polytechnic School, Universidade do Vale do Rio dos Sinos, São Leopoldo, Brazil*
[2] *Independent researcher and consultant*
E-mail: mgonzalez@unisinos.br; diegoerba@gmail.com
*Corresponding author

Abstract. The effects of location play a crucial role in the real estate market, encompassing aspects of accessibility and neighborhood. However, these are elements that are not directly measurable. There are traditional ways to consider location, usually through subjective measures based on professional experience, through proxy variables. Understanding these elements is vital for estimating real estate values, whether for legal, commercial, or tax purposes. Furthermore, seeking more objective options is a relevant issue to broaden the justification of estimated values and to enable the development of mass appraisal models. This article proposes and evaluates alternative solutions based on statistics, machine learning, and geostatistics to estimate location. A study was conducted using market data from Novo Hamburgo, southern Brazil, verifying the feasibility of the options presented. Satisfactory statistical results demonstrate the viability of the proposed approach.

Keywords: location quality, hedonic modeling, Machine Learning, fuzzy logic, kriging.
JEL codes: O18, R33.

1. INTRODUCTION

Location is a crucial element in real estate market analysis. Its impact on property prices could be dissected into aspects related to accessibility and neighborhood quality. Accessibility is often considered in terms of the distance or travel time from the property to commercial areas and amenities. Typically, distances to the city center, shopping malls, supermarkets, as well as parks and other recreational areas are used. The challenge arises by considering simultaneously multiple points of interest. On the other hand, neighborhood effects are related to the quality of the surroundings, evaluated on different scales, either at a macro level (neighborhood or city part) or micro level (immediate surroundings within a neighborhood). However, in both cases, the effects are not directly observable (unlike built area or number of bedrooms, for example, which are elements of direct identification), and indirect measures (known as proxy variables) must be created for these effects (Anselin, 1998; Din et al., 2001; Dubin, 1988; Dubin, 1992; Dubin and Sung, 1987; Li et al., 2015; Li and Brown, 1980; Malpezzi, 2002; Smith et al., 1988).

There is no consensus on the most suitable measures for assessing accessibility and neighborhood quality. However, it seems clear that properties with similar characteristics located close to each other tend to share a similar location effect. It is reasonable to assume that the price of a property is influenced by the quality of its location, which is expected to vary continuously within urban areas. This "location value", resulting from the immobility of the product, decreases with the increase in distance between properties. These variations form almost continuous patterns rather than random fluctuations. Using appropriate tools such as mathematical surfaces or geostatistics, these patterns can be mapped from market data, generating a set of objective location variables (Ball, 1973; Can, 1990, 1998; Dubin, 1992; Gallimore et al., 1996; González et al., 2002; Li and Brown, 1980; McCluskey et al., 2000; Wyatt, 1996a).

A more objective approach is demanded by contemporary appraisal context conditions. On the one hand, there are facilities for obtaining larger market samples, considering the digital availability of data, web scraping, and big data. However, with larger samples, there is an increased need for objective criteria in defining variables to reduce the professionals' effort and enable teamwork. Furthermore, some applications require reducing the subjectivity of measures, such as judicial expertise and taxation, which generally need justification for the adopted solutions due to existing or potential disputes, respectively. Another crucial point is that it has become common for value schedules to be developed by hired professionals who may not have a deep knowledge of the city under study. Virtually, a consulting company in this area can develop value schedules in any city in Brazil or even abroad. The option to obtain location through market-driven mechanisms (data-driven) is relevant in this case.

The issue of objectively measuring location value does not present direct or trivial solutions, identifying a space for proposing alternatives to contribute to the understanding and development of pricing models with applications in individual appraisal, legal actions, and taxation. Following this approach, the main objective of this work is to present alternative solutions and compare them with traditional measures of location, such as distances to relevant points and location variables based on professional experience. More advanced alternatives for measuring location effects based on objective criteria are explored, using statistical techniques, machine learning, and geostatistics. A case study was developed in Novo Hamburgo, a southern Brazilian city, proposing and analyzing various hedonic price models, demonstrating the construction and use of alternative variables.

The paper is structured as follows. The literature review explores studies on real estate valuation, emphasizing methodologies, key variables, and advances in performance evaluation. The research method describes the steps for defining variables, assessing their effectiveness, and collecting data to examine value determinants. The results section details the generated variables, models at various complexity levels, and their significance in predicting real estate values. The discussion analyzes the findings, evaluates model performance, compares them with existing studies, and identifies limitations. Finally, the conclusions provide key insights, underscore contributions, and suggest future research directions in real estate valuation.

2. LITERATURE REVIEW

The importance of location in the real estate market is a well-known factor in literature. Although it is a highly relevant element, there are challenges in measuring the effects of location in the appraisal practice. Since accessibility and neighborhood do not have standardized and ready-to-use measures, proxy variables are employed. However, there are some limitations. Proxy variables for location, measuring accessibility and neighborhood quality, sometimes rely on subjective judgments and may not fully capture a property's location attributes. The lack of standardized metrics for these factors makes comparison and replication difficult across studies or appraisal practices. Limited availability of reliable data on factors such as traffic patterns, public services, and socio-economic conditions affects the accuracy of appraisal models. Dynamic changes in location due to urban development, infrastructure projects, or socio-economic shifts further challenge the relevance of location-based measures. These challenges call for improved methods to better quantify and integrate location effects in real estate valuation (Balchin and Kieve, 1986; Balchin et al., 1995; Derycke, 1971; Harvey, 1996, 2006; Lavender, 1990; Lefebvre, 1991; Muth, 1975; Robinson, 1979).

Accessibility is sometimes explored by considering distances or travel times to key points in the city. It's common to verify the effects of distances to city' commercial and historical center (Allen et al., 2015; Ball, 1973; Can, 1990; Can, 1998; D'Acci, 2019; Gallimore et al., 1996; McCluskey et al., 2000; Smith et al., 1988; Straszheim, 1987), public transport (Allen et al., 2015; Li et al., 2016; Swoboda et al., 2015; Welch et al., 2016; Wyatt, 1996a, 1996b), schools (Ball, 1973; Bartik and Smith, 1987; Boyle and Kiel 2001; Can, 1990; Can, 1998; Gallimore et al., 1996; González et al., 2002; Li et al., 2016), leisure centers (Bartik and Smith, 1987; Boyle and Kiel 2001; Can, 1990; Can, 1998), parks (Boyle and Kiel 2001; Din et al., 2001; Li et al., 2016), distances to main avenues or highways (Allen et al., 2015; Bartik and Smith, 1987; Straszheim, 1987; Swoboda et al., 2015), and other elements are also used.

Traditional models often consider the Central Business District (CBD) as a general attraction center. These models are based on the premise that the CBD concentrates trade, essential urban functions, and most jobs (Derycke, 1971; Muth, 1975). While this projection is generally suitable for small cities or studying parts of larger cities, this simplification can be exaggerated for other situations, as city growth tends to generate more complex structures with multiple attraction centers, resulting in a polycentric city. In fact, some empirical studies using the distance to the CBD as an accessibility measure find little statistical importance for this variable, suggesting alternative measures or considering multiple centers, such as the location of shopping malls (Allen et al., 2015; Ball, 1973; Bartik and Smith, 1987; Can, 1990; Dubin, 1992; Dubin and Sung, 1987; Smith et al., 1988; Straszheim, 1987; Wyatt, 1996a, 1996b).

There is a similar challenge in measuring neighborhood characteristics. The effects are equally important and difficult to measure. More specifically, some studies demonstrate the effects of various factors, such as the pattern of neighboring properties (built environment), land use intensity, education and income levels of local residents, air quality, noise level, availability of schools and public transportation, access to exclusive bike lanes, and ease and safety for pedestrians walking in the neighborhood, or negative externalities, such as proximity to factories, landfills, or even nuclear power plants (Ball, 1973; Boyle and Kiel, 2001; D'Acci, 2019; Din et al., 2001; Ding et al., 2000; Jud and Watts, 1981; Kain and Quigley, 1970; Lang and Jones, 1975; Li et al., 2015; Li et al., 2016; Swoboda et al., 2015; Welch et al., 2016). Some authors also addressed sustainability aspects, such as the value of ecosystems, the effect of green areas, or the distance of properties to water (Cohen et al., 2015; Sander and Haight, 2012; Saphores and Li, 2012).

In traditional practice, professionals often assess measures for each neighborhood based on experience and knowledge of the local market, which can be useful in some cases. However, this method faces limitations, such as the lack of systematic analysis and justification of results, dependence on personal assessment, and difficulties in periodic reassessment, which can result in duplicated efforts and lack of accuracy. For individual appraisals, this is a viable task since information is collected by seeking comparable properties in terms of quality and location, and differences are generally not significant. On the other hand, large-scale appraisals (mass appraisal), such as models for property taxation and market studies, present greater difficulties and are a complex task, given the large variations in building types and spatial price variations (Ball, 1973; Bartik and Smith, 1987; Boyle and Kiel, 2001; D'Amato and Kauko, 2017; Kauko and D'Amato, 2008; Smith et al., 1998; Vargas-Calderón and Camargo, 2022).

3. RESEARCH METHOD

It is observed that literature presents a set of options traditionally applied in pricing models, considering distances to commerce, schools, amenities, presence of noise or pollution sources, among others. In general, daily commerce (represented by CBD, supermarkets, and shopping malls) and free leisure elements such as parks receive more emphasis. The distance to the nearest element is considered.

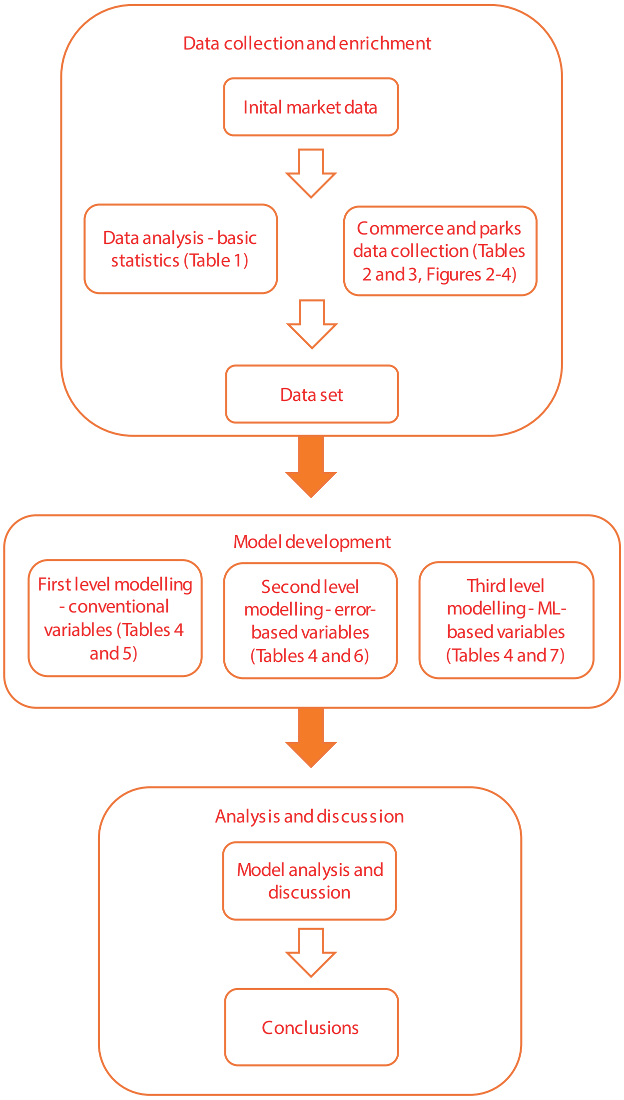
In a polycentric city, multiple attractions, such as commerce and leisure areas, influence property values through varying accessibility. These amenities and disamenities affect properties differently, depending on their proximity to these points. Distances and the mix of factors in different locations create unique impacts on real estate values. The traditional approach considers these effects through a set of variables, which complicates the analysis and reduces the degree of freedom of estimation. Statistical significance will probably not be achieved with this individualized analysis.

The assessment of location quality is often performed through a score for the neighborhood or part of it, based on the professional's experience, the average income of the region, and other parameters. However, this generates an aggregated measure with a low level of detail and requires frequent repetition of the process. In this case, the difficulty of justifying the assigned score is increased, and it is convenient to find ways to weigh the effects together.

This research employs a systematic methodology to address the research objectives. The process includes several key steps. Flowchart 1 represents visually the research methodology, highlighting each step and the interconnections between them to provide a clear outline of the approach.

3.1. Proposal of variables

The developed study is designed in three levels of complexity concerning the variables employed in the



Flowchart 1. Stages of the research methodology. Source: Authors.

models. Through a case study conducted in the city of Novo Hamburgo, southern Brazil, information was collected, and analyses were developed to examine the proposed measures.

a) Initial Level – Traditional Approach

The first level considers the most traditional analysis process, where conventional location measures are employed, with the neighborhood level being subjectively evaluated. It adopts straight-line distances to the nearest points of commerce and urban parks for each property in the sample. The Distance to CBD is related to the city's shopping mall, and the distances to commerce and parks indicate the distance from the nearest element to each property in the sample. The variables generated at this level are Neighborhood-A, Dist.CBD, Dist.Commerce, Dist.Park. Neighborhood-A represents the location variable determined based on the author's experience.

b) Intermediate Level – Statistics and kNN

The second level proposes a linear weighting model to coordinate accessibility measures and an error modeling mechanism (data-driven) to generate neighborhood measures. For accessibility, it is considered that there are multiple points of interest, such as various supermarkets or leisure points at similar distances, given the convenience of weighing the effects together. Therefore, it analyzes the simultaneous influence of these alternatives on the population. It proposes the analysis of a set of measures, generating a variable for commerce and another for parks, considering a weighting mechanism for the relative size of each option and the distances to the sample data. The linear weighting model is basically an equation. A relative weight (pre-determined) is adopted for each point of interest. As amenities, the variables Commerce and Parks were calculated, representing the weighted averages of the distances from supermarkets and parks to each property in the sample, respectively. The general scheme is presented in Equation (1):

$$Amenity(I_i) = \Sigma^a \left[\frac{Distance(I_i, a)}{weight\_a}\right] / A \qquad (1)$$

where $Amenity(I)$ is the weighted average measure of the attribute (in this case, supermarket or park) for property I_i; $Distance(I_i,a)$ is the Euclidean distance from property i to reference a, which has coordinates (x_i, y_i) and (x_a, y_a); $weight\_a$ is the relative weight of each alternative a, with $a = (1,\ldots, A)$, and A indicating the total number of points of interest. With the application of $weight\_a$, larger elements have a smaller distance, representing increased attractiveness.

For the neighborhood issue, the measurement variable at the neighborhood level was constructed from the residuals generated in a model that does not contain location-related variables, in a data-driven approach. Error modeling starts from a hedonic model without the inclusion of location-related variables. Consequently, location effects will be mixed with random errors, however, location effects should be spatially distributed, unlike random errors. In a second step, spatial analysis of errors should be developed, through trend surfaces or kriging, techniques that identify the trends of the studied attribute, isolating the effects of location into a new variable (D'Amato, 2017; Gallimore et al., 1996; Helbich et al., 2014; McCluskey et al., 2000; Ward et al., 1999).

The assumption is that location effects, as they were not explicitly considered, will be contained in the errors and can be isolated, filtering out random variations. At this intermediate level of complexity, the measure for the neighborhood was obtained by summing the standardized errors of the data for that neighborhood, followed by normalization to a scale (1-10), generating the variable Neighborhood-E.

A second neighborhood variable was determined at a micro level, defined pointwise for each property in the sample. A hedonic model was generated with basic variables, also including the neighborhood variable (Neighborhood-E). Following the same reasoning, the premise is adopted that internal neighborhood differences will be contained in the residuals (internal variability). Point estimates – neighborhood value for each property in the sample – were obtained by linear kNN (unweighted).

The k-Nearest Neighbors (kNN) algorithm is a robust and intuitive machine learning method used to solve classification and regression problems. It is a supervised learning method. By incorporating the concept of similarity, kNN calculates values for a new point considering its k nearest neighbors in the training data set. As it works with the average, random differences are filtered, obtaining the trend of neighborhood quality. The calculated variable was called Local-kNN, adopting the arithmetic mean of the 20 nearest neighbors.

c) Advanced Level – Machine Learning and Geostatistics

The third level follows the basic idea of the proposals of the second level but uses more complex techniques, introducing machine learning (fuzzy logic) and geostatistics (kriging). The weighting of commerce and parks distances was performed by fuzzy sets, with membership functions proportional to distances. The participation of each commerce or park element is calculated by the membership function, with the respective weight identified for each source.

A fuzzy system consists of a sum of the partial estimates of each considered effect, which are weighted according to a membership function. Unlike sets that follow classical logic, which have a binary membership definition, such as {0,1}, the membership functions of fuzzy sets assign fractional memberships, in a continuous interval [0,1]. In this approach, each element can belong to several sets with different participation, identified as any value in this interval. The sum of memberships of all elements in the set must reach 1, and at the same time, the sum of the participation of an element in different sets of the system also reaches 1 (Dubois and Prade, 1980; González, 2017; Nguyen and Walker, 2019).

In the case of location, the relationship of properties with neighboring properties occurs in all directions (360°), requiring an adaptation of the membership functions of fuzzy sets, normalizing values to achieve a unitary sum. The general influence is the weighted sum of effects in all directions. The participation of neighboring cases in the final values depends on the weighting scheme defined for fuzzy sets. Participation is more significant for closer units. A format based on the inverse of the distance to weigh cases is an interesting option, using 1/dk, usually k=1 (inverse function, 1/d), or k=2 (square of the distance, $1/d^2$). If adopted with k=0 (no weighting), the result is the unweighted kNN adopted at the intermediate level. Increasing k reinforces the membership values to neighboring points (weighing more strongly closer cases). Therefore, the importance of neighboring cases in the final value increases proportionally to the increase in k. In the studied case, the effects of the exponent were verified, obtaining better results with k=2 (González, 2017). More formally, a fuzzy system composed of D fuzzy sets (one for each attraction point) can be described as in Equation (2):

$$Distance(I_i) = \Sigma^d[\mu_d(I_i)] * Distance_d(I_i) \quad (2)$$

where $Distance(I_i)$ is the adjusted measure for property i; $\mu_d(I_i)$ is the function that calculates the membership of property i to each fuzzy set d; $Distance(I_i)$ is the calculated value for i using rule d. In the case of a function involving urban space, $\mu_d(I_i) = Distance(I_i,d)^{-k}/w$ is adopted, with $w = \Sigma^d[Distance(I_i,d)^{-k}]$, and w calculated to reach $\Sigma^{d,i}\mu_d(I_i) = 1$; $Distance(I_i,d)$ is the Euclidean (linear) distance from property i to the reference (supermarket or park) of rule d, which have coordinates (x_i, y_i) and (x_d, y_d); k is the exponent that gives the weight of the distance influence; and d=(1,…, D), with D representing the total number of reference points. The set of participations was normalized to reach 100% in all cases, using w. This scheme generated the variables Fuzzy-Commerce and Fuzzy-Parks.

For the neighborhood variable, a continuous surface was generated using kriging, from the residuals of an equation using only Neighborhood-E as a location measure. This technique allows smoothing the surface, to some extent, filtering out random errors and concentrating the result on the trends of the studied effect. A mean of the 20 nearest neighbors was also adopted, but now weighed by the inverse of the square distance. The calculation process using kriging is like weighted kNN, but kriging always uses the distance of each nearby information as a weight to consider spatial similarity. The weights were normalized by a mechanism similar to fuzzy sets (*w*), determining the variable Local-Kriging.

Kriging is a spatial weighting technique originally developed by Daniel Krige for use in mining and since then widely expanded for the study of any spatially distributed phenomenon. The basic premise is related to the so-called "first law of geography", introduced by Waldo Tobler, which essentially says, "everything is related to everything else, but things close are more related than things far away" (Tobler, 1970). This proposition is the basis of fundamental concepts of spatial dependence and spatial autocorrelation and is specifically used for the inverse distance weighting method for spatial interpolation and to support the theory of regionalized variables for kriging (Matheron, 1963; Miller and Kahn, 1962; Tobler, 1970).

3.2. Performance evaluation of studied measures

The work focuses on the proposition and testing of some alternative measures for location, with testing and comparison of the results with traditional measures. Each proposed variable must undergo an evaluation of its statistical performance to validate the obtained measure. For this purpose, hedonic price models can be used.

In the real estate market domain, it is essential to simultaneously consider the effects of various elements on prices. In this context, a real estate property is considered a "composite good", characterized by a set of attributes, each assuming different weights in explaining price variations. Hedonic price models involve the proposition and testing of a relationship between prices and the main attributes of properties (Goodman, 1978; Griliches, 1971; Lancaster, 1966; Lucena, 1985; Malpezzi, 2002; Rosen, 1974).

Given the complexities of the real estate market, specific conditions need to be met for price modeling. Hedonic models are constructed using a data set from the analyzed segment, resulting in equations suitable for property valuation or market condition analysis, usually using regression analysis. Regression analysis is a technique that associates independent variables with a dependent variable - in this case, the market price - generating a model. The goal is to establish a numerical model (in this case, an equation) (Gujarati, 2000). A general form for a hedonic price function is expressed in Equation (3):

$$Price = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \cdots + \beta_k x_k + \varepsilon \qquad (3)$$

where *Price* is the variable under study (response variable or dependent variable); $x_1, \ldots, x_k$ are the explanatory variables (for *k* independent attributes); $\beta_1, \ldots, \beta_k$ are the coefficients of the equation representing the relative importance of each of the attributes in explaining the dependent variable; β_0 is the constant or intercept of the equation; and ε is the error term.

The coefficients of the equation are interpreted as the contribution of one unit of each variable to the property price. In other words, βi is the weight or implicit price of that feature, measured in the same currency as the price when the equation is linear. The model's format is not clearly known beforehand, as it is determined through statistical analysis of the data, however, there are guides on literature about often-important attributes, such as size, age, location and other aspects. This data-driven approach allows for flexibility, enabling the identification of the most relevant variables and their relationships with property prices. The format evolves based on the data structure and the underlying patterns observed during the analysis.

The evaluation of regression models initially includes fundamental statistical parameters, including the coefficient of determination (R^2) and the model's significance level through an F-distribution-based variance test (Fisher-Snedecor F). The individual significance of variables is assessed through hypothesis tests based on the Student's t-distribution (Gujarati, 2000).

The value of each sample case is estimated through the adjusted model, and the differences between the collected market value and the estimated value generate residuals or errors. Error analysis is a crucial part of model evaluation. In addition to outlier analysis (individual case view), model residuals can be assessed using root mean square error (RMSE) and mean absolute error (MAE), common metrics used to evaluate predictive model accuracy (a collective, holistic view), particularly in the field of mass appraisal.

The analysis indicates the variables that should remain in the model under a certain significance level and their importance in explaining the price. Some conditions must be checked to ensure the quality of the generated model. Among the regression assumptions, the presence of homoscedasticity (constant variance of errors), normality of errors, and linearity of the relationship in Equation (3) should be analyzed (Gujarati, 2000).

Furthermore, considering the spatial nature of the market, addressing the issue of spatial correlation is crucial. The presence of spatial correlation may indicate trends in the model and reduce the accuracy of estimated values. Spatial correlation can be assessed using the Moran's I index (Anselin, 1998; Can, 1990; Can, 1998; Dubin, 1988; Dubin, 1992).

Finally, to avoid overfitting, it is common to develop the modeling stage with a cross-validation mechanism, typically using 80% of the sample data for training (model generation), reserving 20% for testing and model verification. The data are chosen through simple random sampling. The test verifies whether the model has the ability to generalize (in other words, if it can provide good estimates for cases not seen in the modeling stage).

3.3. Collected data

The study was conducted in Novo Hamburgo, southern Brazil (29°40'4" S; 51°07'5" W), a city located along the federal highway BR-116, about 45 km from the state capital. The research involved acquiring market data for apartments and reference information to assess accessibility on an urban scale. The city is 94 years old, and its urban space is distributed over an area of 223.6 km². It has approximately 247,000 inhabitants (1,105 inhabitants/km²), with a per capita GDP of R$ 37,500.00 (according to 2020 data).

The initial data was obtained from the Brazilian real estate website Viva Real, which is the country's real estate portal with one of the largest property listing databases, focusing on information including prices and precise locations. In some situations, the address was not disclosed in the advertisements, but it was possible to identify the building from photos of the facade or by the name of the condominium (a local peculiarity of referring to buildings by name instead of address). Information collection took place from January 2020 to December 2022, collecting all currently available properties. Since the data collected consists of listings and not sales data, one possible bias is the presence of some exorbitant pricing, but in a large sample this could be detected in conventional outlier analysis. Indeed, an initial analysis allowed the removal of data with discrepancies or lack of information, resulting in obtaining 963 apartment data, which were divided into a training sample with 80% for model generation (771 data) and a test sample with 20% for model evaluation (193 data).

The position of each data point was verified by identifying its coordinates (x_i, y_i). The classification into neighborhoods followed the legal definition of their boundaries (see Figure 1).

The information provided for each property includes traditional options such as private area, number of bedrooms, parking spaces, bathrooms, among others (Table 1). In cases of conflicting information between different advertisements, the latest information was adopted. The correlation of attributes with the price is an essential element, anticipating the expected relationship, although

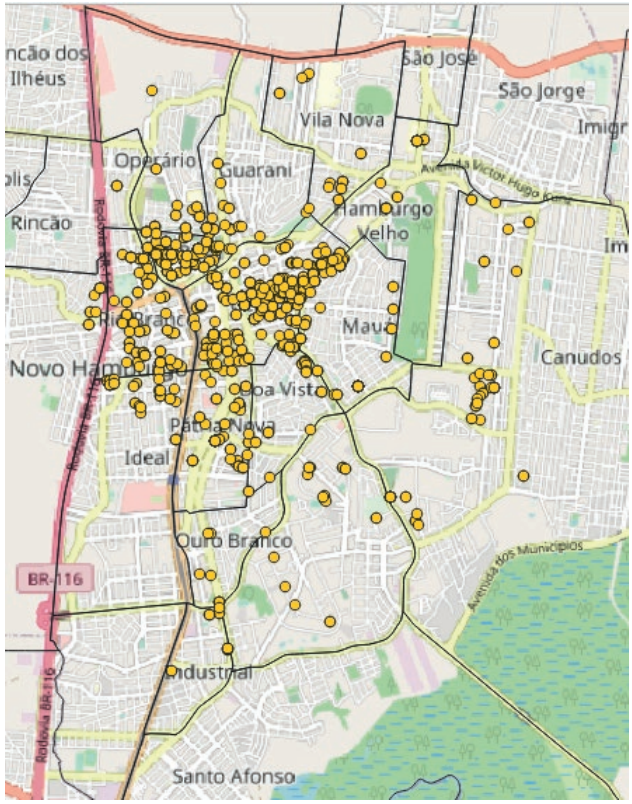


Figure 1. Collected market data and delimitation of city neighborhoods. Source: Data collection by the Authors; Neighborhood limits: Municipal Government of Novo Hamburgo.

the actual contribution is better assessed in hedonic models with multivariate analysis.

While some attributes are conventional and indicate in an objective way their contents, the number of elevators serves as a proxy variable for the construction standard. In this city, a building with two elevators tends to be of a higher standard, accompanied by amenities such as a party hall, swimming pool, or other common-use facilities. On the other hand, a building without an elevator tends to be older or of a lower standard, with a smaller shared area, and so forth.

For the evaluation of accessibility in this region, some reference points were considered. Regarding commerce, distances to the main shopping mall in the city were measured, representing the traditional center of the city (CBD). The central metro station is opposite the shopping mall, serving as an accessibility element and a representation of a relevant and recognized shopping space in the city. Notably, this part of the city does not have other significant points of interest. The major supermarkets were identified, assigning a relative weight based on their size (selling space). Table 2 presents the considered commercial elements.

Table 1. Characterization of Initial Variables.

Attribute	Description	Unity	Range	Average	Correlation with Price
Price	Price	BR Reals	115,000.00 – 3,800,000.00	567,777.21	-
Area	Private area	m^2	30.0-459.0	111.94	0.829
Bedroom	Number of bedrooms	-	1-4	2.44	0.568
Bathroom	Number of bathrooms	-	0-5	2.10	0.825
Parking	Number of parking spaces	-	1-5	1.43	0.796
Penthouse	Penthouse (1) regular (0)	-	0-1	0.082	0.234
Elevators	Number of elevators	-	0-2	0.856	0.309
Month	Information time, on a continuous scale: Month=1: Jan 2020; Month=36: Dec 2022	Month	1-36	18.81	-0.011

Source: Data collection by the Authors; the main source is https://www.vivareal.com.br/.

Table 2. Commerce elements.

#	Identification	Weight	Longitude	Latitude
-	Bourbon Shopping mall/CBD	4	487036.326	6716018.285
1	Hipermarket Bourbon	3	487264.408	6715143.973
2	Supermarket Carrefour	2	487255.919	6716323.870
3	Supermarket Atacadão	2	486800.125	6713588.738
4	Supermarket Rissul – Ave. Nações Unidas	2	486693.097	6717370.535
5	Supermarket Rissul –Bartolomeu Gusmão Str.	1	490319.794	6714939.050
6	Supermarket Rissul – Jamaica Str.	1	491141.607	6716357.547
7	Nacional supermarket – Hamburgo Velho	1	489205.314	6717119.572

Source: Data collection by the Authors.

Table 3. Urban Parks.

#	Identification	Weight	Longitude	Latitude
1	Parque do Trabalhador (Worker's park)	1	485206.205	6716795.575
2	Parque Floresta Imperial (Imperial Forest park)	1	487783.091	6713154.146
3	Parque Municipal Henrique Luis Roessler – "Parcão" ("Big park")	10	489410.252	6716359.260

Source: Data collection by the Authors.

The shopping mall was not included in the supermarket group because it does not offer this service; instead, it is composed of clothing stores, jewelry and accessories shops, musical instruments, electronic equipment, toys, among others.

The city's urban parks were identified, shown in Table 3.

Following findings from various published studies, it can be assumed that small-scale elements such as fruit stands, mini-markets, or squares are not decisive factors in the purchasing process and do not influence the prices charged, with more significant impact from supermarkets and urban parks.

Figure 2 indicates the position of the parks and supermarkets considered, as well as the shopping mall, to check the distribution of the elements.

4. RESULTS

The initial models assessed variables related to the property itself (size, characteristics, number of bedrooms) and upon this foundation, location variables were tested. After the initial exploration of the data and considering the spatially extensive sample with properties exhibiting significant variations, a semi-logarithmic

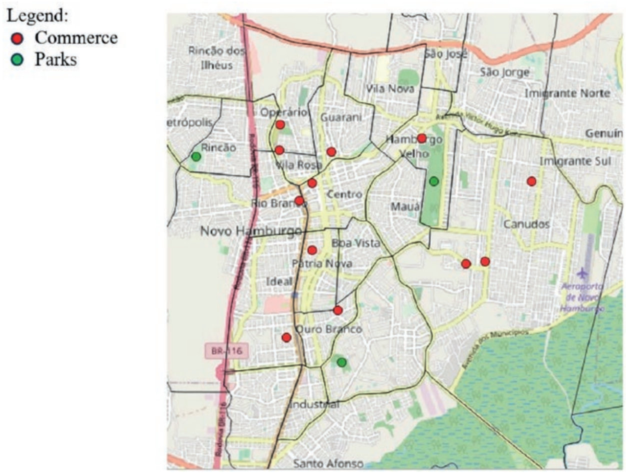


Figure 2. Position of the city's commerce and parks. Source: Data collection by the Authors.

model was adopted, presented in the Equation (4):

$$Price = exp(a_0 + a_1\text{Area} + a_2Bedrooms + a_3Parking + \cdots + a_k\{Location\}_k) + \varepsilon \quad (4)$$

where the basic variables are described in Table 1, and $\{Location\}$ represents one or more variables related to measuring the location effects, as per the level of analysis, presented in Table 4. Various models were examined, exploring different compositions of location variables. It was selected models with the best statistical performance, avoiding unnecessary repetition and proliferation of results.

4.1. Presentation of generated variables

Broad neighborhood variables (macro-location), at the urban scale of the neighborhood, were generated traditionally, based on the professional and research experience of the authors (Neighborhood-A). The second variable is based on the residuals of a model estimated without location variables. The sum of standardized residuals in each neighborhood was normalized, generating Neighborhood-E, representing a less subjective option for this attribute.

Figure 3 shows the distribution of neighborhood valuation for the sample data points. Although similar, there are differences between them. The values are the same for all data in the same neighborhood, in each case.

Local neighborhood variables, assessing intra-neighborhood variations, were based on the errors of the model including the objective neighborhood measure (Neighborhood-E). Two options were adopted. At the intermediate level, Local-kNN adopts the kNN option without weighting, and for the advanced level, a surface

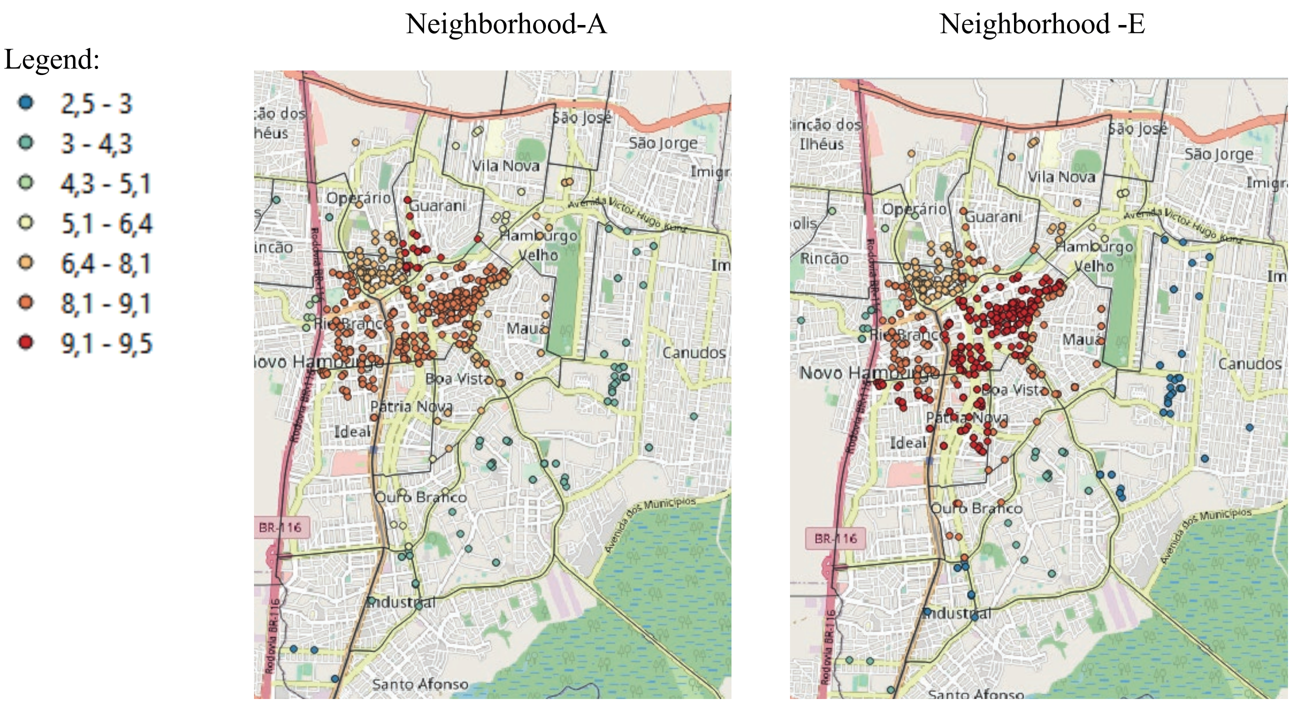


Figure 3. Distribution of neighborhood variables on the neighborhood scale. Source: Authors.

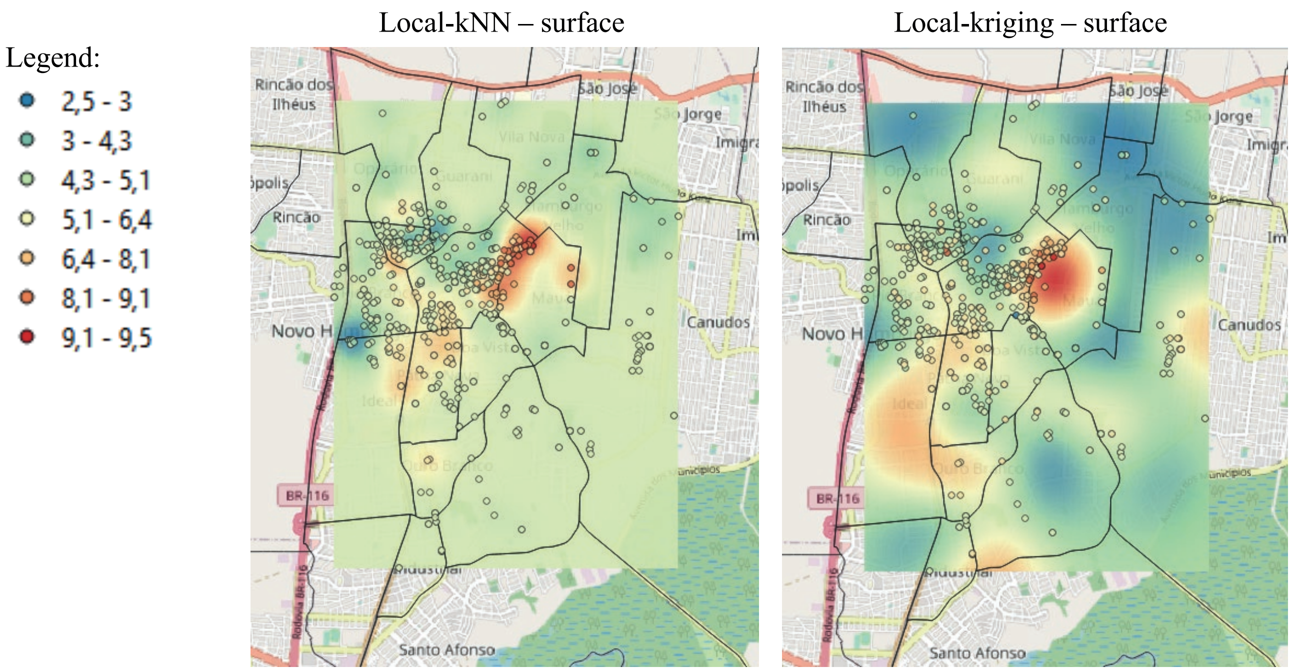


Figure 4. Distribution of intra-neighborhood variables. Source: Authors, using QGIS and Smart-Map plugin.

was calculated with kriging weighted by the inverse of the squared distance ($1/d^2$), subsequently interpolating the point values for each data point in the sample. The generated variable is Local-kriging. For both, the 20 nearest cases were used.

Figure 4 presents the spatial distribution of these variables. Point values and corresponding surfaces are indicated. There are natural differences between the distributions, considering the existence or absence of distance weighting. In the more peripheral areas, which also have less data availability, there is a prevalence of lower values which, in the case with kriging, are characterized by the dark blue color. On the other hand, the more valued region is in the same city area in both alternatives.

Regarding the variables that aim to measure the effects of accessibility, at the first level, distances to the CBD, for commerce and parks were calculated using the distances to all properties and choosing the nearest element (only one in this case).

For the second level, the weighting of the effects of supermarkets and parks was carried out considering the Euclidean distance (linear distance from each data point in the sample to the considered reference point) and the relative weight assigned to the element, whether it be a supermarket or park. For commerce, following the Equation (1) scheme, the function takes a form as in Equation (5):

$$Commerce = \Sigma^c \left[\frac{Distance_d(I_{i,c})}{weight_c}\right] / C \qquad (5)$$

where $c = (1, \ldots, C)$; $Distance_d(I_{i,c})$ is the Euclidean distance between property i and supermarket c; $weight_c$ is the relative size of supermarket c, and C=7 (see Table 2).

At the third level, Fuzzy variables for commerce and parks were calculated. For the fuzzy commerce model, Equations (2) and (5) transform into (Equation 6):

$$Fuzzy\text{-}commerce = \Sigma^c[weight_c * Distance_d(I_{i,c})^{-2}/\mathrm{w}] \qquad (6)$$

With $w = \Sigma^c[Distance_d(I_{i,c})^{-2}]$. The parameter w is the normalization element of the results. The other elements have the same meaning as in Equation (5). The calculation scheme considers the inverse of the square distance, a situation that provides better results than with the inverse of the distance.

In the case of Parks, the proposal follows the same format as Equations (5) and (6), but now considering the elements from Table 3.

In summary, the variables generated to measure the location effects are described in Table 4. The correlation of each variable with the price indicates the potential relationship, to be more precisely verified in the multivariate analysis.

Table 4. Description and averages for location variables in training and testing.

Level	Attribute	Description	Unity	Range	Average	Correlation with Price
Initial	Neighborhood-A	Defined as based on Author's experience	-	3.0-10	7.95	0.312
	Dist.CBD	Distance to shopping mall	km	0.2-6.2	1.182	-0.094
	Dist.Commerce	Shortest distance to supermarkets (Table 2)	km	0.1-1.85	0.690	-0.111
	Dist.Park	Shortest distance to urban parks (Table 3)	km	0.25-2.4	1.572	-0.201
Intermediate	Neighborhood -E	Defined as based on errors of a model with no location attributes	-	2.5-10	8.30	0.289
	Commerce	Weighted average of supermarket distances	km	1.3-3.0	1.721	-0.354
	Park	Weighted average of park distances	km	1.5-3.4	1.981	-0.069
	Local-kNN	Neighborhood calculated by kNN, 20 cases, no weighting	-	4.0-6.8	4.909	0.179
Advanced	Fuzzy-Commerce	Fuzzy weighting of supermarket distances	km	1.0-3.0	2.072	-0.187
	Fuzzy-Park	Fuzzy weighting of park distances	km	1.0-9.8	5.283	-0.354
	Local-kriging	Neighborhood calculated by kriging, 20 cases, with inverse squared distance	-	3.4-6.8	4.895	0.186

Source: Authors.

4.2. Models for the first level of complexity

The initial models adopted conventional attributes. Three models were generated. One uses only the Neighborhood-A variable (Model 1), another includes the three distance measures (Model 2), and the third includes the entire set (Model 3). Similarities between them are observed. The initially evaluated parameters are R^2 and F, which showed no restrictions. The determination coefficient of the models is suitable for a model with spatial coverage and significant variations in characteristics among the data, representing about 87 to 89% explanation for price variations. The calculated F-statistic indicates significance and is extremely low, close to zero (Table 5).

Variables were analyzed for significance with t' statistics. Generally, they were significant at the $\alpha = 0.01$ level. In cases where the achieved level was $\alpha = 0.05$, the coefficient was identified in the table (using an *). No variable exceeded this limit. The signs and coefficients of the variables are as expected, according to each one's contribution. They can be considered good results.

Being a semi-log model, the coefficients of continuous variables can be interpreted by their participation in price relative to a unit of the variable (this is the partial derivative of the equation). For example, in the case of the private area of the models presented in Table 5, a one-square-meter variation represents an increase of about 0.25% in price, considering a range near the variable's average.

The subjective neighborhood variable showed significance in both models in which it appears, with and without the distance' attributes. The coefficients are similar (0.0647 and 0.0722), indicating, respectively, that a 1-point increase in the variable represents about a 7% increase in the average price.

Location variables based on distance were significant, showing a negative coefficient, as expected. Without the Neighborhood-A variable, the coefficients of the distances indicate stronger effects, which is coherent, as in this case, these three variables represent all location effects. In Models 2 and 3, the weight of the proximity to the shopping mall is slightly higher than for the nearest supermarket, while the distance to parks has a much higher coefficient than these two (Table 5). Considering that the effort required to generate distance variables is reduced, it can be considered a positive result.

The verification with the test sample (20% of the data) indicated a slight increase in RMSE (1.7 to 3%), with a more significant effect on MAE (increase of 7 to 8%). There is no evidence of overfitting in this case. Considering the exploratory nature of the analysis, the results can be considered good.

Spatial correlation was assessed through the Moran's I coefficient (Table 5). The three models show similar results, not indicating the presence of spatial autocorrelation, with Moran's I values between 0.083 and 0.160. The second model indicates larger differences between the training and test samples, but both can be considered adequate.

Overall, the model with the four variables (Model 3) presents the best results, although with a slight difference from the others. All four variables are significant at the $\alpha=0.05$ level. RMSE and MAE measures are lower for both training and test samples. There are no indications of spatial correlation, and the determination coef-

Table 5. Result of models with traditional location variables (dependent variable: Ln(Price)).

Attributes		*Model 1*	*Model 2*	*Model 3*
Intersection		11.001020	11.495044	11.090280
Area		0.002438	0.002688	0.002395
Bedroom		0.141338	0.136166	0.147499
Bathroom		0.111389	0.131407	0.105304
Parking		0.254705	0.244094	0.247395
Penthouse		0.090689	0.055278*	0.099039
Elevators		0.237026	0.304255	0.253368
$Month^2$		$7.26*10^{-5}$*	$9.05*10^{-5}$	$7.69*10^{-5}$*
Neighborhood-A		0.064739	-	0.072231
Dist.CBD		-	-0.032437*	-0.012125*
Dist.Commerce		-	-0.029988*	-0.007198*
Dist.Park		-	-0.064838*	-0.084296
Training sample	R^2	0.873097	0.866564	0.895426
	F	~0	~0	~0
	RMSE	262,687.11	299,799.88	250,809.30
	MAE	349.66	361.96	345.83
	Moran's I	0.0857352	0.0752964	0.0827976
	N	770	770	770
Test sample	RMSE	270,632.17 (+3.0%)	303,811.29 (+1.3%)	254,972.07 (+1.7%)
	MAE	377.97 (+8.1%)	386.97 (+6.9%)	372.97 (+7.9%)
	Moran's I	0.1175520	0.1605040	0.0895514
	N	193	193	193

Source: Authors. Note: Variables significant at $\alpha = 0.01$, except *: $\alpha = 0.05$.

ficient indicates that almost 90% of price variations can be explained by the variables included in the model.

4.3. Models at the second level of complexity

Next, alternative models using weighted distances and neighborhood determined with error modeling are presented at two scales (Neighborhood-E and Local-kNN). Three models were generated, progressively incorporating location variables (Table 6). The first includes Neighborhood-E (Model 4), the second incorporates weighted distance variables (Model 5), while the third adds to these three the Local-kNN variable (Model 6).

The coefficients of the variables are similar from one model to another, and the overall results are also similar. The initial model evaluation parameters, R^2 and F, indicate that the models are suitable. The signs and values of the coefficients are consistent with expectations and the first-level models. Although the error level measured by RMSE and MAE, is slightly higher in the test models, it can be concluded that the models do not have problems in this issue.

The location variables show coefficients and signs consistent with expectations (positive for Neighborhood-E and Local-kNN, and negative for weighted distances). There is stability in the coefficients from one model to another. Based on these results, the models can be considered satisfactory.

The coefficients for Commerce in models 5 and 6 are higher than those of the initial models (2 and 3). Conversely, for Park, the coefficients are similar. However, a direct comparison cannot be made since weighted distances are involved here (Table 4 shows the differences in the means of these variables). If a direct comparison is desired, an alternative is to normalize the three measures to a common interval, such as [1-10], transforming them into indices but losing the physical reference of distance.

The results for the training and test samples are similar, ruling out the possibility of overfitting. For the second-level models, Moran's I do not indicate spatial correlation, but there are reasonably higher values for the test data.

Table 6. Result of models with alternative location variables (dependent variable: Ln(Price)).

Attributes		*Model 4*	*Model 5*	*Model 6*
Intersection		11.042742	11.508040	11.148211
Area		0.002385	0.002372	0.002384
Bedroom		0.142306	0.143130	0.144285
Bathroom		0.112631	0.108601	0.105705
Parking		0.257036	0.256736	0.256645
Penthouse		0.100017	0.108687	0.108715
Elevators		0.240248	0.239672	0.248087
Month²		$8.41*10^{-5}$	$8.21*10^{-5}$	$8.01*10^{-5}$
Neighborhood-E		0.055720	0.046151	0.045845
Commerce		-	-0.137610	-0.109840*
Park		-	-0.070760*	-0.067070*
Local-kNN		-	-	0.061709
Training sample	R^2	0.875252	0.876354	0.897412
	F	~0	~0	~0
	RMSE	261,157.49	261,750.97	257,747.22
	MAE	348.48	349.46	347.52
	Moran's I	0.0702365	0.0730797	0.0563623
	N	770	770	770
Test sample	RMSE	267,494.19 (+2.4%)	266,214.52 (+1.7%)	271,084.02 (+5.2%)
	MAE	376.14 (+7.9%)	376.98 (+8.2%)	378.02 (+8.8%)
	Moran's I	0.1121580	0.1223300	0.0839717
	N	193	193	193

Source: Authors. Note: Variables significant at α = 0.01, except *: α = 0.05.

4.4. Models for the third level of complexity

Subsequently, alternative models were developed with machine learning (distances with fuzzy logic) and geostatistics (neighborhood calculated with kriging). The resulting models are presented in Table 7 (models 7 and 8).

The basic parameters used for model evaluation (R^2, F, RMSE, MAE), as well as Moran's I analysis, indicate good results. The coefficients of the variables show signs and values consistent with the previous models. Additionally, the results for the training and test samples are similar, with a slight increase in error levels (3 to 4% for RMSE and about 7% for MAE). It can be concluded that the models are suitable by these criteria.

In general, models 7 and 8 show slightly better results than the initial and intermediate level models. For example, the determination coefficients exceeded 90% for these models.

5. DISCUSSION

The data sample is relatively diverse and poses challenges for generating a single model. It cannot be precisely classified as a mass appraisal, but the sample size is significant and allows for some insights for use with big data. In this context, the presented results can be deemed appropriate.

The produced hedonic models include a stable set of common variables with quite similar results among the models in terms of coefficient values and statistical significance. All variables are significant at levels often adopted in the cited literature (α = 0.01 or α = 0.05). This surpasses the requirements of the Brazilian property appraisal standard, which stipulates α = 0.10 as the minimum threshold for classifying evaluations in the highest quality grade of this standard. Thus, the presented models could even be used in professional activities in this sector (ABNT, 2011; Dantas, 2012; González, 2003).

In general, the presented models were similar in determination coefficient and error parameters (RMSE, MAE). Homoscedasticity, normality, and other conditioning analyses were not presented but were conducted, with results approving the models. Spatial correlation tests also indicate the good performance and suitability of the models. Results from the reserved sample test offer a relative assurance of no overfitting, meaning there is potential for generalization in the models. One

Table 7. Results of the models with advanced (macro and micro) location variables (dependent variable: Ln(Price)).

	Attributes	*Model 7*	*Model 8*
	Intersection	11.348717	11.019981
	Area	0.002357	0.002378
	Bedroom	0.142816	0.144896
	Bathroom	0.256539	0.256448
	Parking	0.112713	0.106633
	Penthouse	0.100957	0.100897
	Elevators	0.238955	0.250131
	Month²	$8.00*10^{-5}$	$7.19*10^{-5}$
	Neighborhood-E	0.053517	0.051674
	Fuzzy-Commerce	-0.073901*	-0.080903*
	Fuzzy-Park	-0.065351*	-0.071380*
	Local-kriging	-	0.075482
Training sample	R^2	0.902623	0.918551
	F	~0	~0
	RMSE	247,809.61	243,331.33
	MAE	318.02	316.54
	Moran's I	0.074836	0.0592868
	N	770	770
Test sample	RMSE	255,694.07 (3.2%)	253,286.94 (4.1%)
	MAE	340.92 (7.2%)	338.46 (6.9%)
	Moran's I	0.105896	0.0665885
	N	193	193

Source: Authors. Note: Variables significant at α = 0.01, except *: α = 0.05.

could say there is a statistical balance. Naturally, there is a dependence on the employed data, and the results are connected to a specific case, delimited in time and space.

The goal of the work was to demonstrate the use of techniques with an objective character and compare them with the traditional option, which is subjective. In this sense, the balance of results between models is promising, as unconventional techniques require fewer human resources and offer more reproducibility, ease of updating, and teamwork facilitation, besides expanding the possibility of justifying calculation parameters for taxation and other applications.

For example, comparing models containing the variable Neighborhood-A (models 1 and 3) and the variable Neighborhood-E (models 4 to 8) reveals minor differences. One can consider an advantage of the objective variable, which can be obtained and updated more quickly (actually within a few minutes of analysis) and independent of deep personal technical knowledge about the market context under study.

Although they require some processing time, the advantage of Local-kNN and Local-kriging variables is measuring neighborhood effects in more detail, considering existing variations within neighborhoods. These variables are generated for a broad space and can be used in different situations, with only periodic updating. In other words, the processing time is diluted by the reuse of generated numerical surfaces.

The variables used to measure accessibility, considering distances to trade elements and urban parks, were significant, with balanced coefficients and contributions to the models. In the presented case, weighted variables did not reveal significant contributions compared to the traditional option. Since they must be generated for each study sample, considering the processing time, their use should be evaluated case by case.

Looking ahead, the proposed methodology holds significant potential for adaptation and application in diverse real estate markets or geographical contexts. Its capacity to incorporate various layers of location variables, including those derived from machine learning and geostatistics, makes it flexible for different urban environments and market conditions. The use of objective and easily generated variables, such as accessibility and neighborhood quality, can facilitate the mass appraisal process in other regions, especially in areas where traditional data may be sparse or challenging to obtain. Furthermore, the methodology's robustness, demonstrated through solid statistical performance, suggests it could be applied to evaluate emerging real estate markets, offering industry professionals valuable insights into property pricing dynamics in evolving urban landscapes.

A sensitivity analysis could be developed aimed at checking the robustness of the data obtained by varying key input variables, model assumptions, and data sampling methods. By examining how price predictions change with these variations, the analysis ensures consistency and reliability across different techniques, confirming the methodology's applicability and generalization to real estate valuation.

There are some possible limitations on presented research. While the research demonstrates the effectiveness of the proposed methodology, its limitations include potential data constraints and the challenge of integrating diverse variables across different regions. Future developments could focus on testing it in different contexts (other cities or countries.

6. CONCLUSIONS

The location of a property is a crucial factor in the real estate market. In simple terms, the quality of location can be divided into two parts: accessibility (as a "macro" location, at the city or neighborhood level) and local neighborhood (a "micro" level, related to the quality of the immediate surroundings of each property).

Traditional measures have some limitations, and this study proposes alternatives. In summary, three sets of location variables were compared. At an initial level, traditional variables were employed. At the intermediate level, variables based on statistics and kNN were adopted, while the advanced level employed machine learning and geostatistics.

The comparison was based on a sample of over 960 cases, with good statistical performance for all presented models, from various perspectives. The balance of models with traditional variables with models developed with other techniques is considered an advantage for the more objective ones, which provide more detailed information for accessibility measured through distances and require less time to generate neighborhood variables. This suggests that the methodology not only produces consistent results but also yields well-qualified models that can be relevant for industry professionals.

Additionally, the statistical analysis revealed that models based on near-neighborhood variables (Local-kNN and Local-kriging), which have a higher degree of innovation, showed strong qualifications for statistical performance. These variables have a continuous spatial variation surface, a detail that is hardly obtainable without an objective data analysis.

Objectivity is important for promoting mass appraisal models, considering the effort required to generate variables that are not directly observed, such as location. In summary, the results indicate the viability of the methodology in using objective variables to measure accessibility and evaluate neighborhood quality, while emphasizing the ease of creating price models.

Ultimately, the methodology shows potential for adaptation to various real estate markets. Its flexibility, incorporating machine learning and geostatistics, allows it to be applied in different urban contexts with limited traditional data. The robust statistical performance suggests that it can assess emerging markets, offering valuable insights into price dynamics and supporting mass assessment efforts.

REFERENCES

Allen, M. T., Austin, G. W., & Swaleheen, M. (2015). Measuring highway impacts on house prices using spatial regression. *Journal of Sustainable Real Estate*, 7(1), 83–98. http://aresjournals.org/doi/abs/10.5555/1949-8276-7.1.83

Anselin, L. (1998). GIS research infrastructure for spatial analysis of real estate markets. *Journal of Housing Research*, 9(1), 113–134.

Associação Brasileira de Normas Técnicas - ABNT (2011). NBR 14653-2: avaliação de bens parte 2: avaliação de bens parte 2: imóveis urbanos. Rio de Janeiro.

Balchin, P. N., & Kieve, J. L. (1986). *Urban land economics*. 3 ed. London, MacMillan.

Balchin, P. N., Bull, G. H., & Kieve, J. L. (1995). *Urban land economics and public policy*. 5 ed. New York, Bloomsbury Publishing.

Ball, M. J. (1973). Recent empirical work on the determinants of relative house prices. *Urban Studies*, 10, 213–233.

Bartik, T. J., & Smith, V. K. (1987). Urban amenities and public policy. In Mills, E. S. (Ed). *Handbook of regional and urban economics*, v.2 (urban economics). Amsterdam, Elsevier, 1207–1254.

Boyle, M. A., & Kiel, K. A. (2001). A survey of house price hedonic studies of the impact of environmental externalities. *Journal of Real Estate Literature*, 9(2), 117–144.

Can, A. (1998). GIS and spatial analysis of housing and mortgage markets. *Journal of Housing Research*, 9(1), 61–86.

Can, A. (1990). The measurement of neighborhood dynamics in urban house prices. *Economic Geography*, 66(3), 254–272.

Cohen, J. P., Cromley, R. G., & Banach, K. T. (2015). Are homes near water bodies and wetlands worth more or less? An analysis of housing prices in one Connecticut town. *Growth and Change*, 46(1), 114–132. https://doi.org/10.1111/grow.12073

D'Amato, M. (2017). Location value response surface model as automated valuation methodology a case in Bari. In D'amato, M., & Kauko, T. (Eds). *Advances in Automated Valuation Modeling. Studies in Systems, Decision and Control*, vol. 86. Cham, Springer. https://doi.org/10.1007/978-3-319-49746-4_10

D'Acci, Luca. (2019). Quality of urban area, distance from city centre, and housing value. Case study on real estate values in Turin. *Cities*, 91, 71–92.

D'Amato, M., Kauko, T. (eds). (2017). *Advances in Automated Valuation Modeling. Studies in Systems, Decision and Control*, vol 86. Cham, Springer. https://doi.org/10.1007/978-3-319-49746-4_10

Dantas, R. A. (2012). *Engenharia de avaliações: uma introdução à metodologia científica*. 3 ed. São Paulo, PINI.

Derycke, P. H. (1971). *La economia urbana*. Madrid, IEAL.

Din, A., Hoesli, M., & Bender, A. (2001). Environmental variables and real estate prices. *Urban Studies*, 38(11), 1989–2000.

Ding, C., Simons, R., & Baku, E. (2000). The effect of residential investment on nearby property values: evidence from Cleveland, Ohio. *Journal of Real Estate Research*, 19(1/2), 23–48.

Dubin, R. A. (1988). Estimation of regression coefficients in the presence of spatially autocorrelated error terms. *Review of Economics and Statistics*, 70(3), 466–474.

Dubin, R. A. (1998). Predicting house prices using multiple listings data. *Journal of Real Estate Finance and Economics*, 17(1), 35–59.

Dubin, R. A. (1992). Spatial autocorrelation and neighbourhood quality. *Regional Science and Urban Economics*, 22(3), 433–452.

Dubin, R. A., & Sung, C. H. (1987). Spatial variation in the price of housing: rent gradients in nonmonocentric cities. *Urban Studies*, 24(3), 193–204.

Dubois, D. J., & Prade, H. (1980). *Fuzzy sets and systems: theory and applications*. New York, Academic Press.

Gallimore, P., Fletcher, M., & Carter, M. (1996). Modelling the influence of location on value, *Journal of Property Valuation and Investment*, 14(1), 6–19.

González, M. A. S. (2017). Automated valuation methods in real estate market - A two-level fuzzy system. In D'Amato, M., & Kauko, T. (Eds.). *Advances in Automated Valuation Modeling*. Studies in Systems, Decision and Control, vol 86. Cham, Springer International Publishing., 265–278.

González, M. A. S., Soibelman, L., & Formoso, C. T. (2002). A new approach to spatial analysis in CAMA. In *9th European Real Estate Society Conference (ERES 2002), June, 2002, Glasgow*. Glasgow: ERES.

González, M. A. S. (2003). *Metodologia de avaliação de imóveis*. Novo Hamburgo, SGE.

Goodman, A. C. (1978). Hedonic prices, price indices and housing markets. *Journal of Urban Economics*, 5(4), p. 471–484. https://doi.org/10.1016/0094-1190(78)90004-9

Griliches, Z. (1971). *Price indexes and quality change*. Cambridge, Harvard University Press.

Gujarati, D. N. (2000). *Econometria básica*. São Paulo, Makron Books.

Harvey, D. (2006). *Spaces of global capitalism: towards a theory of uneven geographical development*. London; New York, Verso.

Harvey, J. (1996). *Urban land economics*. 4 ed. London, MacMillan.

Helbich, M., Brunauer, W., Hagenauer, J., & Leitner, M. (2013). Data-driven regionalization of housing markets. *Annals of the Association of American Geographers*, 103(4), 871–889. https://doi.org/10.1080/00045608.2012.707587

Jud, G. D., & Watts, J. M. (1981). Schools and housing values. *Land Economics*, 57(3), 459–470.

Kain, J. F., & Quigley, J. M. (1970). Measuring the value of housing quality. *Journal of the American Statistical Association*, 65(330), 532–548.

Kauko, T., & D'Amato, M. (2008). *Mass Appraisal Methods. An international perspective for property valuers*. Oxford, Blackwell.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of Political Economy*, 74(2), 132–157. http://www.jstor.org/stable/1828835

Lang, J. R., & Jones, W. H. (1975). Hedonic property valuation models: are subjective measures of neighborhood amenities needed? *AREUEA Journal*, 7(4), 451–465.

Lavender, S. D. (1990). *Economics for builders and surveyors*. Essex, Longman.

Lefebvre, H. (1991). *The production of space*. Oxford, Basil Blackwell.

Li, M. M., & Brown, H. J. (1980). Micro-neighbourhood externalities and hedonic housing prices. *Land Economics*, 56(2), 125–141.

Li, W., Joh, K., Lee, C., Kim, J. H., Park, H., & Woo, A. (2015). Assessing benefits of neighborhood walkability to single-family property values: a spatial hedonic study in Austin, Texas. *Journal of Planning Education and Research*, 35(4), 471–488. https://doi.org/10.1177/0265813516663932

Lucena, J. M. P. D. (1985). *O mercado habitacional no Brasil*. Rio de Janeiro, Fundação Getúlio Vargas (Série Teses, n.9).

Malpezzi, S. (2002). *Hedonic pricing models: a selective and applied review*. The Center for Urban Land Economics Research, Madison, University of Wisconsin.

Matheron, G. (1963). Principles of geostatistics. *Economic Geology*, 58(8), 1246–1266.

Mccluskey, W. J., Deddis, W. G., Lamont, I. G., & Borst, R. A. (2000). The application of surface generated interpolation models for the prediction of residential property values. *Journal of Property Investment and Finance*, 18(2), 162–176.

Miller, R. L., & Kahn, J. S. (1962). *Statistical analysis in the geological sciences*. New York, John Wiley.

Muth, R. F. (1975). *Urban economic problems*. New York, Harper and Row.

Nguyen, H. T., & Walker, E. A. (2019). *A first course in fuzzy logic*. 4 ed. Boca Ratón, CRC Press, Taylor & Francis Book.

Robinson, R. (1979). *Housing economics and public policy*. London, MacMillan.

Rosen, S. (1974). Hedonic prices and implicit markets: product differentiation in pure competition. *Journal of Political Economy*, 82(1), 34–55.

Sander, H. A., & Haight, R. G. (2012). Estimating the economic value of cultural ecosystem services in an urbanizing area using hedonic pricing. *Journal of Environmental Management*, 113, 194–205. https://doi.org/10.1016/j.jenvman.2012.08.031

Saphores, J. D., & Li, W. (2012). Estimating the value of urban green areas: a hedonic pricing analysis of the single family housing market in Los Angeles, CA. *Landscape and Urban Planning*, 104(3), 373–387. https://doi.org/10.1016/j.landurbplan.2011.11.012

Smith, L. B., Rosen, K. T., & Fallis, G. (1988). Recent development in economic models of housing markets, *Journal of Economic Literature*, 26(1), 29–64.

Straszheim, M. (1987). The theory of urban residential location. In Mills, E. S. (Ed). *Handbook of regional and urban economics*, v.2 (urban economics). Amsterdam, Elsevier, 717–757.

Swoboda, A., Nega, T., & Timm, M. (2015). Hedonic analysis over time and space: the case of house prices and traffic noise. *Journal of Regional Science*, 55(4), 644–670. https://doi.org/10.1111/jors.12187

Tobler, W. R. (1970). A computer movie simulating urban growth in the Detroit region. *Economic Geography*, 46, 234–240.

Vargas-Calderón, V., & Camargo, J. E. (2022). Towards robust and speculation-reduction real estate pricing models based on a data-driven strategy. *Journal of the Operational Research Society*, 73(12), 2794–2807. https://doi.org/10.1080/01605682.2021.2023672

Ward, R. D., Weaver, J. R, & German, J. C. (1999). Improving CAMA models using geographic information systems/response surface analysis location factors. *Assessment Journal*, 6, 30–38.

Welch, T. F., Gehrke, S. R., & Wang, F. (2016). Long-term impact of network access to bike facilities and public transit stations on housing sales prices in Portland, Oregon. *Journal of Transport Geography*, 54, 264–272. http://dx.doi.org/10.1016/j.jtrangeo.2016.06.016

Wyatt, P. (1996a). The development of a property information system for valuation using a geographical information system (GIS). *Journal of Property Research*, 13(4), 317–336.

Wyatt, P. (1996b). Using a geographical information system for property valuation. *Journal of Property Valuation and Investment*, 14(1), 67–79.