MASS APPRAISAL OF APARTMENT THROUGH GEOGRAPHICALLY WEIGHTED REGRESSION

Caio Flávio Martinez Fontoura Júnior[1] - ORCID: 0000-0001-8948-351X

Marlene Saleti Uberti [2] - ORCID: 0000-0002-5643-0304

Vilma Mayumi Tachibana[3] - ORCID: 0000-0002-8804-6163

[1] Universidade Estadual Paulista - UNESP, Faculdade de Ciência e Tecnologia, Departamento de Cartográfica, Presidente Prudente - SP, Brasil.
E-mail: caio.fontoura@unesp.br

[2] Universidade Federal Rural do Rio de Janeiro - UFRRJ, Instituto de Tecnologia, Departamento de Engenharia, Seropédica – RJ, Brasil.
E-mail: msuberti@hotmail.com

[3] Universidade Estadual Paulista - UNESP, Faculdade de Ciência e Tecnologia, Departamento de Estatística, Presidente Prudente - SP, Brasil.
E-mail: vilma.tachibana@unesp.br

Abstract:

Housing Market appraisal studies generally apply classic regression models, whose parameters are globally estimated. However, the use of the Geographically Weighted Regression (GWR) model, allows the parameters to be locally estimated, increasing its precision. The aim of this article is to apply the GWR model to a sample of 82 apartments, in order to create a plan of values of some districts of the West Zone of Rio de Janeiro city, Brazil. With the proposed methodology, GWR and kernel estimator, it is possible to generate a surface of values. The performance of the surface of values was assessed with (i) cross-validation between the kernel functions, with the Root-Mean Square Standardized (RMSS) error; and with (ii) the GWR adjustment factors to determine the ideal bandwidth. The contribution of generating a surface of values with geographical location via kernel estimator lies on supporting apartment pricing, such as in calculating the venal value of apartments of the West Zone of Rio de Janeiro city, besides being applied in IPTU- Imposto sobre Propriedade Predial e Territorial (The Urban Real Estate Property Tax) and ITBI - Imposto de Transmissão de Bens Imóveis (Tax on the Transfer of Real Estate) and ITBI collection.

Keywords: Plan of generic value; Mass appraisal; Kernel interpolator; Surface of value; Geostatistics.

1. Introduction

For market evaluation, local factors and real conditions must be considered in the elaboration of statistical models. These considerations help to justify the values in a given set of properties, which highlights that the same variables do not always influence a property in the same way, as there are factors that weigh more or less the value of the property and thus contribute to each property obtain their characteristics that will not necessarily be the same (Hornburg, 2015).

One how Brazilian municipalities obtain it as a source of resources is through tax collection by charging fees and taxes such as IPTU - Imposto sobre Propriedade Predial e Territorial (The Urban Real Estate Property Tax) and ITBI - Imposto de Transmissão de Bens Imóveis (Tax on the Transfer of Real Estate). These taxes are calculated according to municipal laws, based on the property's market value, and are acquired based on local real estate behavior. This process is known as the Plan of Generic Values – PGV (Paiva and Antunes, 2017).

The PGV is defined as a cartographic or graphic document that represents the spatial distribution of the average values of the properties in each locality of the municipality, typically presented by square face. Its main attribute is the definition of a real estate tax policy that is fair and equal (Averbeck, 2003; Malaman and Amorim, 2017).

There are two distinct statistical procedures for the treatment and modeling of spatial geographic features in market data: the spatial regression and Geostatistics methods. Through the use of econometric methods, there is a great difficulty in the search for models for evaluation, taking into account the geographical location that causes both the appreciation and the devaluation of properties (Hornburg, 2015).

The property valuation process for tax purposes has a direct relationship with the Multipurpose Cadastre (MC). According to Art 28 of Ordinance MCid 511 of 12/07/2009, the MC plus other thematic registries, provide information for the evaluation of properties for tax, extra-tax and any other purposes involving values of urban and rural properties (Ministério das Cidades, 2009).

Thus, in this work, the mass evaluation of urban properties, in particular apartments in the West Zone of Rio de Janeiro (Camorim, Barra da Tijuca, Recreio and Jacarepaguá), which uses the direct comparative method of market data according to NBR 14.653: 2 (ABNT, 2011).

In this context, a Plan of Generic Values was developed for the urban area of the quoted neighborhoods of the West Zone in Rio de Janeiro, based on data collection in the field, which used the cartographic base of IBGE (Brazilian Institute of Geography and Statistics) and the use of the Kernel estimator, to interpolate the estimated values of the Geographically Weighted Regression (GWR), to obtain the digital database of the real estate market in the study area.

2. Theoretical Referential

2.1. Plan of Generic Values

According to ABNT (2011), NBR 14.653-2 defines the Plan of Generic Values as "the graphical representation or listing of the generic values of square meter of land or property on the same date". Santos (2014) defines the Plan of Generic Values as: "a graphic document that represents the spatial distribution of average real estate values in a given region. It is an important instrument of real estate taxation as it allows for greater tax justice".

According to Santos (2014), the use of PGV is restricted to the urban area in most municipalities, but it can also be used in rural areas. And Nadolny (2016), shows that PGV is of great importance for public agencies, taxpayers of real estate taxes, technicians in the field of Evaluation Engineering and the community in general.

For Oliveira (2015), the PGV serves as a database for evaluating land in a city, which through it manages to formulate the calculations to be able to determine the venal values of urban properties, to collect the IPTU and can also be used to limit the minimum collection of the ITBI. In addition to the tax part, the PGV is also an important document to serve as a support to direct the municipal planning, which helps to reflect the real estate valuation and devaluation indexes.

According to Lemes (2016), in addition to researching the current value of the square meter of buildings in the real estate market, certain factors must be taken into account regarding the area's infrastructure and the geographic location of the property, which directly influence depreciation and appreciation of its value, for the elaboration of the PGV, which are: security, access routes, availability of public service, neighborhood and possible environmental risks, unhealthy factors and the possibility of future undertakings.

When evaluating properties for the preparation of a value plan, the basic characteristics of the city's population of properties must be known, so that the model to be used is capable of evaluating all properties, so that in the end it determines the individual value of each property. For the municipalities, it is important to know the individual value of the property to collaborate in the extra-fiscal objectives and determine the policy and tax collection. And the best way to obtain the best result is through the mass evaluation of the properties (Averbeck, 2003). Mass appraisal or collective appraisal can be defined as "the process of defining mathematical models, obtained from the reality of local values, tested and statistically validated and applied in the evaluation of several properties in a population" (Averbeck, 2003).

The assessments that have the tax purpose are carried out in bulk of the properties, as it is impossible to carry out individually for each property. In which it consists of obtaining the values that use common data and homogenized evaluation processes for all properties located within a specific perimeter and within a determined area (Dib; Júnior and Meireles, 2008).

According to Oliveira (2015), for this procedure to be carried out, some variables must be taken into account, in which these variables are the most prominent, and also, the specific characteristics that belong to the region of study. Among these variables, some have great influences on the composition of the properties and are cited in the bibliography as the distance to poles of valorization and trade, the zone of use and occupation of the soil, topography, means of transport available, paving, pedology and quality of the local infrastructure.

There are some methodologies for the execution of PGV, among them are the classic methodology elaborated through the homogenization of the comparative elements through the application of factors and the inferential methodology that obtains the rule for building the price of real estate from the reality of the area. For a city with a population of more than twenty thousand inhabitants, there is a need for a specialist in the area with a capacity in mass evaluation, to meet the demand. And the PGV defines the basic value of the properties, usually per unit area (R\$/m²), which it uses for individual assessment (Averbeck, 2003).

According to Rampazzo (2012), over time the PGV becomes outdated, which is common in Brazilian municipalities, as the difficulty of revisions is due to the high operational cost and the delay in the formation of research, evaluations, and conclusions of the data. It is through the collection of taxes and the technical registration of the municipality that the city administration seeks to research, analyze, and determine a better-quality strategy for society.

For example, the city of Curitiba is considered a model city in terms of planning, however, in its updating of the Valuation Plan, it did not use the comparative method of market data as required by NBR 14653: 2 - Evaluation of Urban Real Estate. The consequences of not updating by the method meant that the public administration caused much lower taxation than normal, also influencing the planning behavior and the direction of improvements in the regions of the city (Rampazzo, 2012).

According to the Ministério das cidades (2009), Ordinance 511, for each size of the municipality, cycles of four to eight years are recommended for updating the PGV. However, according to Nadolny (2016), the PGV update must be done annually, it is not enough to be updated only when there is a drop in its collection, as the work has to be constant, aiming to seek new data in the market. Since the greatest difficulty for this update is not due to technical, bureaucratic or financial reasons, but to political reasons, there is resistance from the city to negotiate with the city council for the approval of the PGV under the law.

2.2. Geographically Weighted Regression

For Menezes (2009), the models of continuous local variations, have the idea that for each observed point the adjustment of the regression model should be performed, considering the other observations concerning the distance from the point, this method is called Geographically Weighted Regression. At first, this process was designed to address extreme heterogeneity.

According to Almeida (2012), in the nineties, there was the development of Geographically Weighted Regression from the set of works by Fotheringham, Brundson, and Charlton, being part of the econometric approach to provide a local version of the linear regression analysis, acquiring with the aid of sub-samples of observations weighted by geographical distance.

Initially proposed by Brunsdon, Fotheringham and Charlton (1996) geographically weighted linear regression is a method that aims to explore spatial non-stationarity, the same being a condition in which the global spatial regression model cannot correctly explain the relationships between some sets of variables defined in a given geographical location (Carvalho et al., 2006).

The first procedure to be developed for a GWR is the mapping of possible blocks of similar values in the area in which the study is taking place is called Clusters (Carvalho et al., 2006).

According to Uberti (2016), Geographically Weighted Regression is: "a technique that extends the traditional regression method, allowing the parameters to be estimated locally, so that the model coefficients, instead of being global estimates, are specific to a point i (location)".

The location of each point in the sample is measured according to its proximity and similarity, that is, the points that are closest to a specific location have a greater influence than the points that are far away and the model is calibrated by the method of weighted least squares (Uberti, 2016). Another important factor in the GWR methodology is the bandwidth, which can be considered as a smoothing parameter. After all, the greater the bandwidth, the greater the coefficients are since it encompasses more observations around the calibration point and the smaller the bandwidth the answer will be greater heterogeneity because the calibration point has fewer observations.

The fixed spatial kernel function is responsible for calibrating the model for n sub-sample around the regression point i. Each sub-sample is defined by the spatial kernel function. And calibration does not only occur at regression points that are part of the data sample but can also be performed at any point defined in space by the coordinates (Almeida, 2011).

The estimator of GWR, uses the method of weighted least squares with weight in the location, being used by the spatial kernel function (Uberti, 2016). It results in a set of parameters adjusted for each point of the analyzed geographical location (Carvalho et al., 2006).

2.3. Adjustment Factors of GWR

According to Emiliano (2009), the Akaike information criterion (AIC) was developed by Hirotugu Akaike, under the name of "an information criterion" in 1971, which can be understood as a relative measure of the fit quality of a statistical model estimated. Where AIC is a criterion capable of evaluating the quality of the fit of the parametric model, estimated by the maximum likelihood method.

Akaike (1974) defined his information criterion as (Equation 1):

$$\text{AIC} = -2 \log \text{ (maximum likelihood)} + 2 \text{ (k)} \quad (1)$$

Where k is the number of parameters to be estimated in the model.

Bozdogan (1987) suggested corrections for the AIC, called another corrected Akaike information criterion (AIC_c) such as (Equation 2):

$$AIC = -2 \log L (\text{maximum likelihood}) + 2k + 2 \frac{k(k+1)}{n-k-1} \quad (2)$$

For Almeida (2012), when the bandwidth is not known, it can be acquired by the technique of the cross-validation optimizer (CV). Another way to determine the best bandwidth is by using the lower values of the AIC, which indicates the most suitable bandwidth for determining the data model. The AIC can be used in different areas, one of the main areas applied is statistics, used mainly in classical regression and geostatistics (Emiliano, 2009).

2.4. Kernel Interpolator

When the subject portrays the punctual processes, it is necessary to estimate the expected amount of case per unit area, that is, to estimate the intensity. There are several ways to estimate the intensity using different procedures that can be calculated using interpolators such as Kriging, trend surface, local regression models and kernel estimator (Santos, Santos and Santo, 2012).

Uberti et al. (2017) used three regression models (classical, spatial and GWR), performed the geostatistical modeling of the values estimated by the models and ending with the interpolation by the ordinary Kriging methods and the Kernel estimator, comparing which would be the best among them. According to the analyses, both the ordinary Kriging and the Kernel can be used to generate surfaces of values. To determine which is the best to be used, the author performed the calculations of RMSE (Root Mean Square Error), median of the reasons, the values of the dispersion coefficient and the referential relative to the price, relative errors and the value of the absolute errors of these relative errors and concluded that the surface generated from the Kernel estimator of the GWR model was the one that had the best performance.

There are several functions of the Kernel estimators, among them are Gaussian or Normal, Quadratic, triangular, negative and uniform exponential. Each having a different form of processing (Santos, Santos and Santo, 2012).

According to Monteiro et al. (2004), the Kernel estimator is an intensity estimator, in which it has the function of counting each point within the area of influence, being weighted by the distance of each one to the geographic location of interest.

In the context of geo-technologies, the term Kernel means nucleus, which refers to a statistical process for estimating density curves, and for each observation, we weight through the distance with the core value, the central value (Monteiro et al., 2004).

According to Monteiro et al. (2004), the class of estimators in the literature is described as non-parametric

density estimators and defines these estimators as: "generalize the idea of the local moving average, assuming that the density of the phenomenon varies locally smoothly, without "peaks" or "discontinuities".

The main objective of the Kernel estimator is to create a smooth surface that can achieve the closest representation to the reality of natural and socioeconomic phenomena (Monteiro et al., 2004).

The main parameters of the kernel interpolator are:

a) the radius of influence that determine the point to be interpolated in the neighborhood; and,

b) the function of estimating with properties that are "favorable" for the smoothing of the phenomenon.

The disadvantage is the strong dependence on the search radius and the exaggerated smooth surface, in which in some cases it can hide important variations in the location (Monteiro et al., 2004).

Among the various applications of the kernel estimator, they can be used to map and estimate the distribution of points in the study area, and it is also used for the location and densification of forest fires (Weber and Wollmann, 2016).

For Weber and Wollmann (2016), among the techniques for estimating the density applied to the set of point events, which results in an intensity contour map (Kernel Map) estimated in cases across the study region.

3. Methodologic

3.1. Studied Area

For the realization of the values plan for a given municipality, the first step is the choice of the area in which the study will be carried out. The locality of interest chosen was the region of Barra da Tijuca, Camorim, Jacarepaguá and Recreio dos Bandeirantes, part of the West Zone of the City of Rio de Janeiro, located between 650000 meters / 690000 meters south latitude and 7450000 meters / 7470000 meters west longitude, as it represents the area of interest on the location map in Figure 1.

The neighborhoods, Barra da Tijuca and Jacarepaguá, are considered upscale neighborhoods in the West Zone of the municipality of Rio de Janeiro and considered the most populous in 2010, being Barra da Tijuca (135 900 inhab) and Jacarepaguá (157 300 inhab). Since the 70s, Barra da Tijuca has attracted large real estate investments and consequently increases pollution. The properties in this region grew by 47% in 10 years, and also, have been the target of migration from other neighborhoods in the city due to the financial, gastronomic, hotel and entertainment center of the capital of Rio de Janeiro (G1, 2011).

The estimated area of these four neighborhoods is approximately 163.47 Km^2 and the sum of the population of the Camorim, Barra da Tijuca, Jacarepaguá and Recreio dos Bandeirantes neighborhoods is equivalent to 377,410 thousand inhabitants.
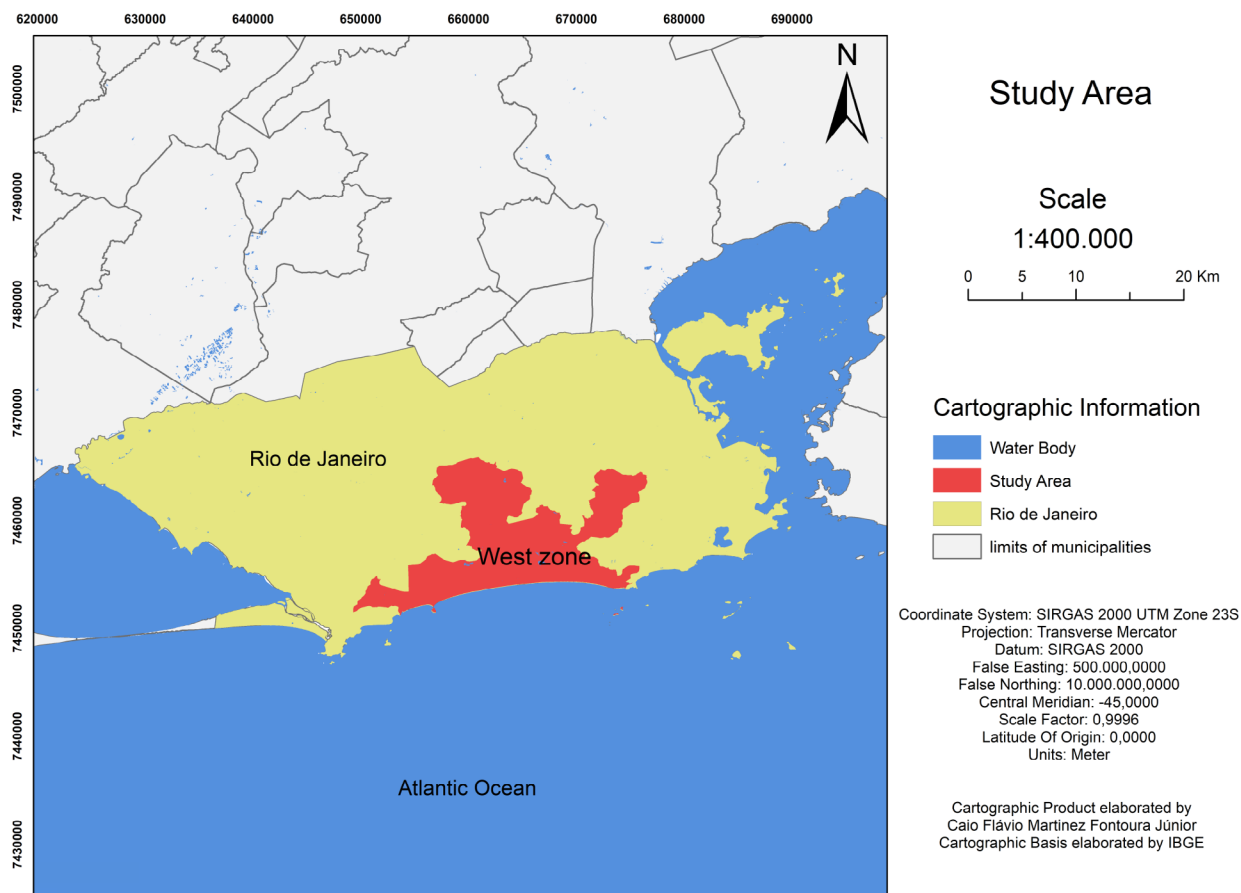
Figure 1: Location Map of the West Zone of Rio de Janeiro

For the creation of the Value Plan of the region of interest, it was structured according to the summarized steps of Figure 2.
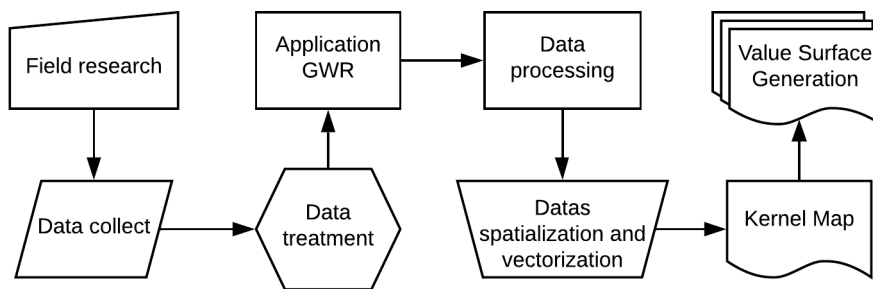

Figure 2: Paper phases to creation of municipality PGV

3.2. Field Research

Field data collection took place from August to September 2018, on the websites of real estate agencies in the city of Rio de Janeiro, where information about the apartments was obtained, such as area, distance from the beach, garage number, number of rooms and the value of the property, obtaining a total of 82 samples.

3.3. Data Processing

After data acquisition, the data was processed using the INFER software (ÁRIA, 2003), in which variables that influenced the real estate market in the West Zone of Rio de Janeiro were used, such as number of rooms and garage, distance from the beach and area of the property, which is all quantitative variables.

After determining the multiple regression model, the samples and variables that proved to be significant were inserted into the GWR4 software (Nakaya et al., 2014) to find a regression model for each geographical position, being tested for the four types of kernel estimators among them fixed Gaussian, fixed Bi-square, adaptive Bi-square, and adaptive Gaussian, and carrying out the analysis using selection criteria such as AIC_c, AIC, and CV.

3.4. Kernel Mapping

With the values estimated by the GWR the spatial distribution of the data, began the creation of the surface of values, through Geostatistics, through the Kernel estimator. The Kernel estimator is an interpolator, which can estimate the intensity of the eventuality across the region, even in locations where the process has not created any real occurrences (Santos, Santos and Santo, 2012). Through the analysis of Geostatistics, six surfaces were generated using the kernel functions, among them: Exponential, Gaussian, Quartic, Epanechinikov, Polynomial and Constant. Thus, the analysis was carried out by cross-validation the previously mentioned kernel functions.

4. Results

The multiple regression and statistical tests were performed to estimate the best model that represents the real estate market in the region of the West Zone of City of Rio de Janeiro. The 82 observations, that were acquired by the real estate websites of the city of Rio de Janeiro, were used.

4.1. Classic Regression Model

To find the ideal model that represents the real estate market in this region (West Zone), several transformations in the independent variables were tested, such as exponential transformation, logarithm, inverse, and no transformation.

Through the software INFER and considering the unit value without transformation as a dependent variable (Equation 3).

$$[VU] = 7610,3 + 22,578 * [AREA(m^2)] - 192,82 * [Disntace\ from\ de\ beach\ (km^2)] - 92,795 * EXP[\text{Number of Bedrooms}] + 271,65 * EXP[\text{Garage}] \tag{3}$$

The description of the independent variables used, and their respective transformations are Area, without transformation; distance from the beach, without transformation; Number of rooms, the exponential transformation was applied, and Number of Garages were also applied to exponential transformation. Table 1 shows the calculated t and its significance for each regressor. The tabular t was considered two-tailed with a significance level of 10% and n-k-1 degree of freedom (77) equal to 1.6649. All regressors have significance, as the t table in the module is less than the t calculated.

Table 1: Estimation of parameters, value t and the significance

	Coefficient	Standard deviation	Value t	Value-p
Area (m^2)	22,578	13,9038	1,802	$7,6 \times 10^{-2}$
Distance from the beach (Km^2)	-192,82	37,3633	-5,663	$2,5 \times 10^{-7}$
Number of bedrooms	-92,795	48,6744	-2,654	1×10^{-2}
Garage	271,65	147,6463	2,653	1×10^{-2}
R		0,6177		
R^2		0,3816		
Regression standard error		1417,17		
Fcalc		11,88		

Analyzing the transformations and signs, it is concluded that the model is coherent, therefore acceptable, because as the area of the property increases the unit value of the property also increases, in the same way, that, when the number of garage increases, the greater it becomes the unit value of the property. On the other hand, as the observations are more distant from the beach and with a smaller number of rooms (the value of the square meter of the samples with fewer rooms, the unit values are higher), the unit value of the property reduces, that is, the closer to the beach and the larger the square footage of the apartment, the more valuable they become.

The regression had four independent variables, with 82 observations and presented a moderate correlation (0.6177). The analysis of variance known as ANOVA, which consists of using the Snedecor distribution with a significance level of 1.0%, obtains the Ftab = 3.573. In which the Fcalc = 11.88 is considered, it is concluded that there is regression and the significance of the model is equal to 1.0×10^{-5}%.

4.2. Multicollinearity and Residue Analysis

Through the correlation matrix, it can be identified if the model does not present Multicollinearity between the independent variables, as can be seen in Table 2. The NBR 14.653:2 (ABNT, 2011), determines that the maximum correlation coefficient is 0.8, however, for this work, the maximum limit of 0.7 was adopted, which makes the analysis more careful.

Table 2: Matrix of partial correlations

	Area (m^2)	Distance from the beach (Km^2)	EXP (Number of bedrooms)	EXP (Garage)
Area (m^2)	1			
Distance from the beach (Km^2)	-0,1768	1		
EXP(Number of bedrooms)	0,1615	0,3204	1	
EXP(Garage)	0,3389	0,2428	0,6761	1

Through the sampling data, the presence of divergent elements was verified, where the suspicious data are evaluated as outliers when the points present standardized residues greater than two in the module.

In this situation, three points were detected considering outliers by the criterion of two standard deviations, which are the sample data 5, 31 and 73, but they were not eliminated, as they are close to the limit value. The verification of the existence of outliers is performed by plotting the standardized residues versus the values

estimated by the regression. The regression model found was determined using 82 samples. Figure 3 represents the outlier indication graph.
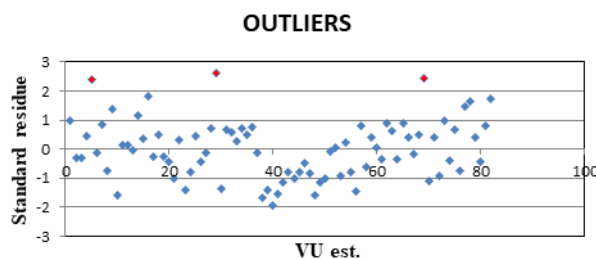

Figure 3: Graphic of indicators of outlier

The regression model showed an approximately normal distribution of residues within the values stipulated by the technical standard (NBR 14.653: 2) as shown in Table 3.

Table 3: Distribution of standardized residual

Interval	Gauss distribution	% residue in the ranging
-1; +1	68,3 %	74,39 %
-1,64; +1,64	89,9 %	91,46 %
-1,96; +1,96	95,0 %	96,34 %

According to the Kolmogorov-Smirnov test, at a significance level of 1%, it is accepted the alternative hypothesis that there is normality in the distribution of waste, with the largest difference obtained of 0.0611 and the critical value of 0.1800.

4.3. Geographically Weighted Regression

The GWR is a method that allows parameters to be determined regionally, where model coefficients are estimated for a specific point instead of being estimated globally as the traditional regression methodology.

The mathematical functions of the kernel can be fixed or adaptive, in which the geographic information can be through the UTM or geographic coordinates. In this work the UTM coordinates of the properties acquired by Google Earth Pro were used. Concerning the bandwidth, it is essential to estimate what is the ideal bandwidth to determine these parameters, using the width that results in the lowest (AIC).

In this phase, the GWR4 software was used to find the ideal bandwidth and also the results of the geographically weighted regression. Among the functions of the kernel, the following types of the kernel were tested: Fixed Gaussian, Fixed bi-square, Adaptive bi-square and Adaptive Gaussian. E to find the ideal bandwidth in the GWR4 software, the analysis was performed using the following criteria: AIC (Akaike), AICc (Corrected Akaike) and CV (Cross-Validation).

Tests were performed with all types of the kernel and with the following criteria: AIC, AICc, CV. Table 4 shows that the lowest value for the AIC criterion was in the fixed-width Gaussian Kernel type with a value of 1419.358.

Table 4: Results of the types of kernel estimators with the AIC, AICc and CV criteria

Criteria types	AICc		AIC		CV	
Types of Kernel	Ideal bandwidth	Minimum value	Ideal bandwidth	Minimum value	Ideal bandwidth	Minimum value
Fixed Gaussian	3846,602	1427,63	1750,5	1419,358	2003,21	2002094,215
Fixed bi-square	10284,128	1428,149	4490,645	1419,876	9866,973	2032789,236
Adaptive Gaussian	77	1425,533	57	1419,905	57	1945319,244
Adaptive bi-square	50	1427,812	50	1425,849	50	2037913,827

4.4. Data Spatialization

Figure 4 illustrates the specialized properties by urban areas of a part of the West Zone of Rio de Janeiro, with the samples being concentrated in greater quantities in the neighborhoods of Barra da Tijuca and Jacarepaguá. And some properties located in isolated areas, two properties in the Camorim neighborhood and one property in Recreio.
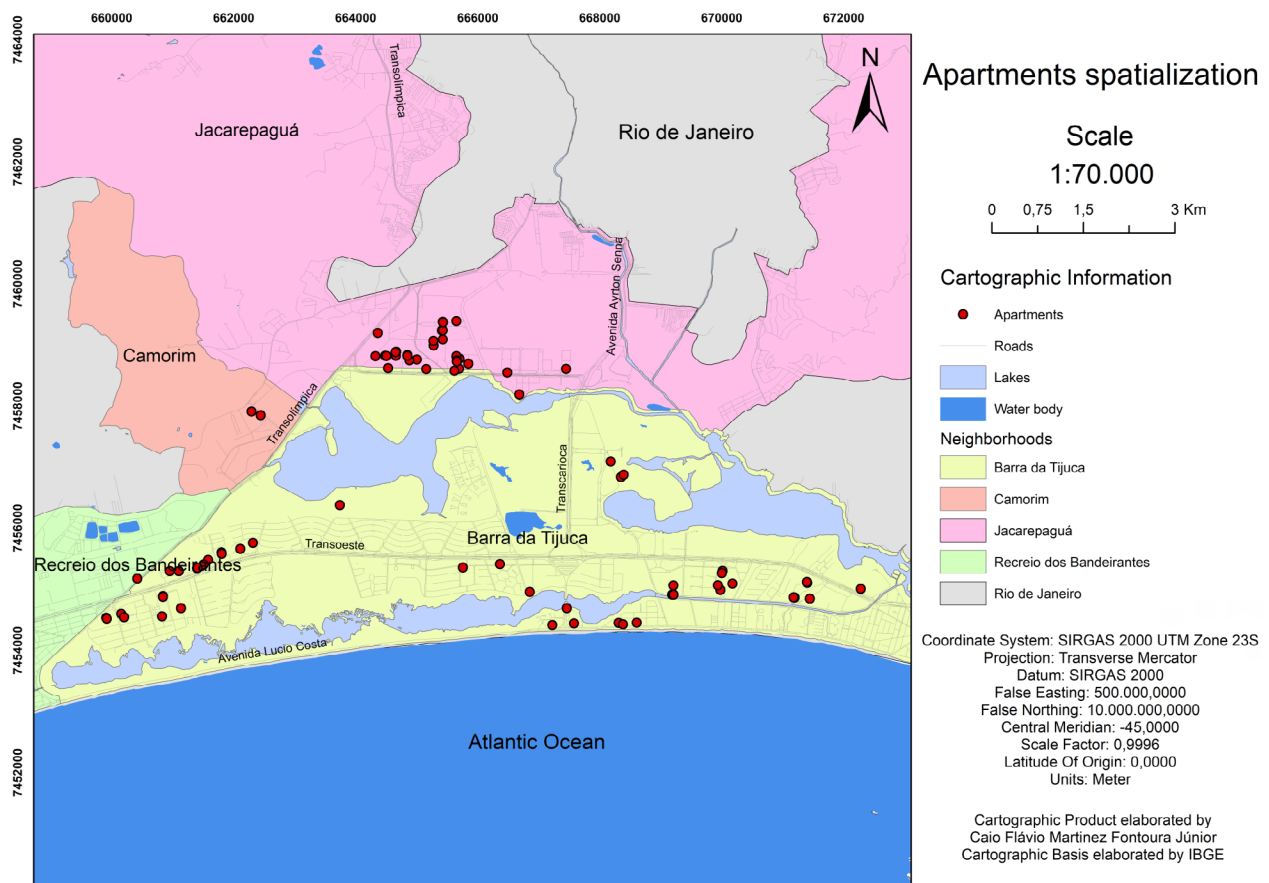
Figure 4: Spatialization map

4.5. Surface of Values

Through the values foreseen in the geographically weighted regression model, using the unit values of each property according to each geographical location, the surface of values was determined, which uses the software ArcGIS 10.3 (ESRI, 2014). Table 5 shows the results of the cross-validation of the Exponential, Gaussian, Epanechnikov, Quarctic, 5th order Polynomial functions and the Constant, in which it can be concluded that the Gaussian function obtained the best results in the Cross-validation.

Table 5: Cross Validation

Function	Mean Error	RMS	Mean Error standardized	RMS standardized	Average Standard standardized
Exponential	0,3434247	724,6689	-0,0057738	1,030538	705,3281
Gaussian	-24,75185	749,2781	-0,03912382	1,003673	747,5501
Epanechnikov	-27,80263	751,19	-0,04968608	1,0383	713,3344
Quartic	-17,76333	758,9594	-0,03408838	1,026706	731,7085
5th order Polynomial	-17,16267	761,1284	-0,03419014	1,023274	737,6035
Constant	-42,47294	461,8615	-0,06898561	1,030968	725,3268

The error values are very close between the functions, however, for the Gaussian function, the standardized RMS value was closer to 1 (1.003673).

In this way, six value surfaces were generated, where the greatest concentration of samples is located in the Barra da Tijuca and Jacarepaguá regions, by different estimators as shown in Figure 5. The surfaces are very similar, with some variations in the spots corresponding to each value range. The use of cross-validation allowed the determination of which surface to consider as the most representative for the region under study.
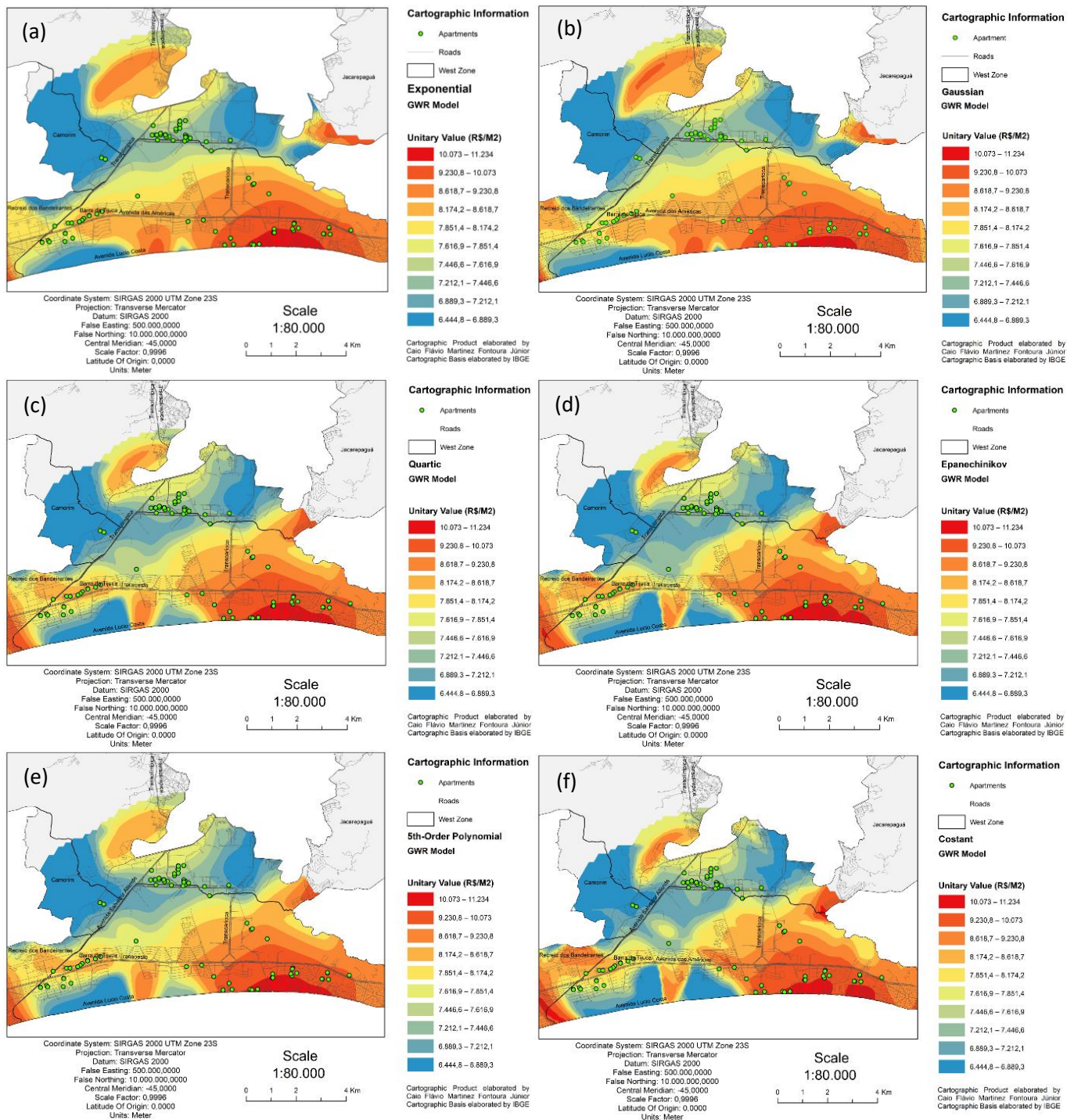
Figure 5: Surface of Values map by the Exponential (a), Gaussian (b), Quartic (c), *Epanechinikov* (d), 5th order Polynomial (e) and Constant (f) estimator.

5. Conclusion

Acquiring real estate data in a part of the West Zone of Rio de Janeiro to constitute a sample set was a complicated task, given that the real estate acquired on real estate websites did not have the geographical location, only the street name and some the number with the zip code of the apartments, in this way a detailed work was carried out through Google Earth Pro for the acquisition of these coordinates (UTM), because for this work the

geographic location is extremely important.

In this way, the comparative method of market data and Geographical Weighted Regression (GWR) was used to generate the surface of values using the INFER 32 and GWR4 software. And the surfaces of predicted values of the Exponential, Gaussian, Epanechnikov, Quartic, 5th order Polynomial and Constant functions were generated by the ArcGIS software. As the errors between the functions are very similar, the way to choose the most suitable surface is by analyzing the standardized RMS that is closest to 1, in this situation the surface generated by the Gaussian function of the Kernel was the one that was closest to 1 (1.003673).

In this work, the region where properties are most valued is located in Barra da Tijuca, considering that the closer to the beach, the more expensive the square meter (m^2) becomes and also taking into account the size of the apartment. A large concentration was located in the Jacarepaguá region, but with lower values, due to its location far from the beach.

The relevance of generating a surface of values using geographic location through the Kernel estimator can assist in the indication of apartment values and can be used in the calculation of the venal value of apartments in the West Zone of the city of Rio de Janeiro, in addition to being applied mainly in the collection of Urban Property Tax (IPTU in Brazil) and Property Transfer Tax (ITBI in Brazil).

Most Brazilian municipalities use the Cost Quantification Method through the values of the Basic Unit Cost - BUC, which does not reflect the market value for this segment of urban properties (apartments). The use of BUC by the city halls in the calculation of real estate taxes generates losses to the public coffers, which would be solved with the use of the statistical techniques covered in this work. In addition to the losses in the collection, the use of the value of BUC generates distortions in the values, since the properties-built value/devalue according to fluctuations in the real estate market and the values of BUC do not follow these variations.

As future work for that same region, other variables such as cultural equipment (shopping malls, theaters, and cinemas) and urban equipment (circulation and transportation, infrastructure, public safety, education, and health) must be taken into account to help surface other types of values of real estate.

ACKNOWLEDGMENT

AUTHOR´S CONTRIBUTION

Researched and selected theoretical references C.F.M.F.J.; Data collection: C.F.M.F.J.; Data processing: C.F.M.F.J.; Preparation of draft manuscript: C.F.M.F.J.; Review of the proposed method: C.F.M.F.J., M.S.U. and V.M.T; Supervision: M.S.U. and V.M.T.; Discussion of the results, data analysis and paper revision and refinement: all authors.

REFERENCES

ABNT (Associação Brasileira de Normas Técnicas). 2011. *Avaliação de Bens: Imóveis Urbanos/*NBR 14.653-2. 2. ed. Rio de Janeiro: ABNT, pp.1.

ÁRIA, Informática. 2004. INFER versão 32, manual de operação. Belo Horizonte. MG.

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19 (6), pp.716-723.

Almeida, E. 2011. Lei de Verdoorn local para a agricultura. *Revista Econômica do Nordeste*, 42(1), pp. 87-104.

Almeida, E. 2012. *Econometria Espacial Aplicada*. Rio de Janeiro: ALINEA.

Averbeck, C. 2003. *Os Sistemas de Cadastro e Planta de Valores no Município: Prejuízo da Desatualização*. MSc – Universidade Federal de Santa Catarina.

Bozdogan, H. 1987. Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions. *Psychometrika,* 52(3), pp.345-370.

Brunsdon, C; Fotheringham, A. S. Charlton, M. E. 1996. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geographical analysis,* 28 (4), pp. 281-298.

Carvalho, L. et al. 2006. Regressão Linear Geograficamente Ponderada em Ambiente SIG. *Revista transporte*, 14 (2), pp.18-26

Emiliano, P. C. 2009. *Fundamentos e Aplicações dos Critérios de informação: Akaike e Bayesiano*. MSc Universidade Federal de Lavras.

ESRI. 2014. ArcGIS Desktop: Release 10.3. Redlands, CA: Environmental Systems Research Institute.

Dib, M.; Júnior, R. and Meireles, S. 2008. *A Importância da Engenharia de Avaliações de Imóveis para o Tribunal de Contas do Estado do Rio de Janeiro*. Specialist. Pontifícia Universidade Católica do Rio de Janeiro.

G1. 2011. Bairros na Zona Oeste do Rio crescem até 150% em uma década, diz IBGE. G1, [online] 2011. Available at: <http://g1.globo.com/brasil/noticia/2011/07/ibge-bairros-na-zona-oeste-do-rio-crescem-ate-150.html>.   [Accessed 14 September 2018].

Hornburg, R. A. 2015. Avaliação em Massa de Imóveis usando estatística Bayesiana. PhD. Universidade Federal de Santa Catarina.

Lemes, A. 2016. A Planta de Valores de Manaus e sua interferência na Avaliação de Imóveis Urbanos. *Revista Especialize On-line IPOG*. 1(11).

Malaman, C. S. and Amorim, A. 2017. Método para Determinação de Valores na Avaliação Imobiliária: Comparação entre 0 Modelo de Regressão Linear e Lógica Fuzzy. *Boletim de Ciências Geodésicas,* 23(1), pp. 87-100.

Menezes, P. 2009. *Análise da Equidade no Acesso à Energia Elétrica Utilizando Técnicas de Regressão Espacial: Estudo de Caso no Estado do Amazonas, Brasil*. Dissertação (Mestrado em Estudos Populacionais e Pesquisas Sociais da Escola Nacional de Ciências Estatísticas) – Escola Nacional de Ciências Estatísticas.

MINISTÉRIO DAS CIDADES. 2009. Gabinete do Ministro. Portaria n.º 511, de 7 de dezembro de 2009. Diretrizes para a criação, instituição e atualização do cadastro territorial Multifinalitário (CTM) nos municípios brasileiros. Diário oficial da união, Brasília, DF.

Monteiro, A. M. V. et al. 2004. *Análise espacial de dados geográficos.* Brasília: Embrapa.

Nakaya, T. et al. 2014. GWR4 - User manual Windows application for Geographically Weighted Regression modelling. Department of Geography, Ritsumeikan University.

Nadolny, M. 2016. A Importância da Planta Genérica de Valores na Tributação Municipal. *Revista Técnico-Científica do CREA-PR*. (4), pp. 1-12.

Oliveira, D.; Silva, C. R. 2015. Modelagem de Planta Genérica de Valores para a Cidade de Monte Carmelo/Mg. In: Meireles, T. A. V.; Silva, C. R.; Santil, F. L. P. Geotecnologias aplicadas ao mapeamento. Maringá: PGE editora, Ch. 2.

Paiva, C. A. and Antunes, A. F. B. 2017. Geração de Planta de Valores Genéricos a Partir do Cadastro Territorial Urbano. *Revista Brasileira de Cartografia*, 69(3), pp. 505-518.

Rampazzo, D. 2012. Consequências da desatualização da Planta de Valores Genéricos de Curitiba. *IPOG - Revista Especialize On Line.*

Santos, B.; Santos, R. and Santo, S. 2012. Análise Espacial Aplicada à Expansão de Condomínios Fechados na Cidade de Feira de Santana (Ba). In: *IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação,* Recife – PE, 06-09 may 2012.

Santos, H. G. 2014. *Efeitos Espaciais Em Mercados De Terras Rurais: Modelagem, Validação E Avaliação De Desempenho.* MSc. Universidade Federal de Pernambuco.

Souza, N. et al. 2013. Aplicação do Estimador de Densidade *kernel* em Unidades de Conservação na Bacia do Rio São Francisco para análise de focos de desmatamento e focos de calor. In: *Anais XVI Simpósio Brasileiro de Sensoriamento Remoto – (SBSR),* Foz do Iguaçu/PR, 13-18 april 2013.

Uberti, M. 2016. *Avaliação em Massa de Imóveis Rurais através de Modelagem Clássica, Espacial e Geoestatística*. PhD. Universidade Federal Rural do Rio de Janeiro.

Uberti, M. S. et al. 2017. Mass appraisal of farmland using classical econometrics and spatial modeling. *Land use Policy*, 72, pp. 161-170.

WEBER, A. and WOLLMANN, C. 2016. Mapeamento dos incêndios residências na área urbana de Santa Maria, RS, Brasil utilizando o estimador de densidade Kernel. *Investigaciones Geográficas*. (51), pp. 49-60.