Land Use Policy 72 (2018) 161–170

Contents lists available at ScienceDirect

Land Use Policy

journal homepage: www.elsevier.com/locate/landusepol

Mass appraisal of farmland using classical econometrics and spatial modeling

Marlene Salete Uberti*, Mauro Antonio Homem Antunes, Paula Debiasi, Wagner Tassinari

Federal Rural University of Rio de Janeiro, Seropédica, Rio de Janeiro, Brazil

ARTICLE INFO

Keywords:
Plan of generic values – PGV
Mass appraisal
Ordinary kriging
Surface of values

ABSTRACT

Mass appraisals of properties traditionally use classical linear regression models (CLRMs); however, there has been the need to model the data spatially. Such modeling of the geographic effects has been used mainly in appraisals of urban areas, but the values of the properties in rural areas are also affected by the geographic location. This paper aims to use spatial regression econometric models in a sample of rural properties to elaborate the plan of values for an area of the North Fluminense Region – RJ, Brazil. The proposed methodology is to investigate and model the effects caused by the spatial autocorrelation on the CLRMs, evaluate their performance comparing them with the spatial models and produce the plan of values through ordinary kriging. The utilized sample consisted of 113 observations and 25 samples of verification. The performance of the obtained surfaces of values was evaluated through the Root Mean Squared Error (RMSE). The results showed that the spatial autocorrelation can have its effects controlled by Spatial Regression Models, because the Spatial Error Model (CAR) allowed to model the spatial dependence present in the residuals. Using the metrics of Akaike information criterion (AIC), R^2 and likelihood function (LIK), the CAR model showed better fit in comparison to the CLRM. The results showed that the surface generated by the CAR model showed the best performance with the lowest RMSE. The combination of the methodologies of classical and spatial regressions and the use of geostatistical techniques were adequate to elaborate and obtain the plan of values for rural areas, to be used for various purposes, such as taxation, financing, expropriations, indemnities (in case of creation of conservation units or even in environmental disasters), among others.

1. Introduction

Brazil is a country with continental dimensions, territorial surface of 8,514,876.599 km^2 and the rural properties occupy 71.10% of this surface. Thus, the techniques of mass appraisal of values of the properties are of great importance for a series of applications. One of the main applications is in the determination of the Rural Land Tax – RLT. In addition, another tax in Brazil that depends on the correct appraisal of the property is the Tax on the Transfer of Real Estate – TTRE, which is of total responsibility of the Municipal Governments. There are also many other actions in rural properties that also need correct determination of the values, such as: financing, expropriations, indemnities (in case of creation of conservation units or even in environmental disasters), real estate buying and selling, land reform, etc. Each one of these actions determines values that do not always follow evaluative techniques, generating different values for each situation. As to the RLT, according to the Federal Revenue Secretariat – FRS (BRASIL, 2012), the collection in 2012 was equal to R$ 677 million[1]. Using the area estimated by the National Institute of Colonization and Land Reform – INCRA in the same year, 605,387,746.06 ha, the value in 2012 was 1.12 R$/ha, which is considered to be too small and reflects an inefficient taxation.

The elaboration of the Plan of Generic Values - PGVs can use the statistical techniques of Multiple Regression, through the homogenization of the values of a sample of properties collected in the real estate market. However, there has been the need to incorporate variables of geographic location in the regression models, in order to model spatial effects. This modeling of the spatial effects, through spatial econometrics, has been used mainly in the PGVs of urban areas (Trivelloni, 2005; Hornburg, 2009), and the values of properties in rural areas are also affected by these effects (Santos, 2014).

However, there are no studies in the literature on the spatial analysis of values of rural properties for the determination of which is the best technique to obtain the PVG. Thus, the present study aims to treat the data statistically through Classical and Spatial Regressions, generate the surface of values through Geostatistics and thus elaborate the plan

* Corresponding author.
E-mail addresses: msuberti@ufrrj.br (M.S. Uberti), mauroantunes@ufrrj.br (M.A.H. Antunes), paula@ufrrj.br (P. Debiasi), tassinari@ufrrj.br (W. Tassinari).

[1] In December 31, 2012, the exchange rate was 1USD = R$2.04

https://doi.org/10.1016/j.landusepol.2017.12.044
Received 21 October 2016; Received in revised form 13 December 2017; Accepted 13 December 2017

of values for rural properties in the area of the PROJIR – Project of Irrigation and Drainage of Sugarcane in the North Fluminense Region – RJ, Brazil. It is expected with this study to evaluate the techniques that can be used in plan of values elaboration and that can be used in other regions of the country.

With the obtained information, it will be possible to implement, for example, other mechanisms of reduction in the RLT value, besides the ones that already exist for those properties that maintain environmental preservation areas in the form of legal reserves, permanent preservation areas or private reserve of natural protection, making it an instrument of environmental policy, through which the RLT would work as a compensation paid by the society to the landowner who preserved the environment.

This study aims to generate knowledge that will contribute to the solution of problems related to the determination of prices of rural properties in Brazil, with possible application of the methodology also in other countries. Therefore, this study will address one of the issues that involve rural land management, the mass appraisal for the determination of the value, which has multiple applications, including territorial taxation.

2. Studied area

For the application of the proposed study, it was selected an area of the PROJIR that encompasses part of the municipalities of São João da Barra, Campos dos Goytacazes, Cardoso Moreira, São Francisco de Itabapoana, Conceição de Macabu, Carapebús and Quissamã, in the state of Rio de Janeiro – RJ, Brazil (Fig. 1).

The PROJIR area is approximately located between the coordinates 21°17'15"S / 40°59'40"W and 22°04'55"S / 41°45'01"W, approximately 280°17'15"W / 40°59'40"S and 22°04'55"W / 41°45'01"S, approximately 280 km away from the capital, Rio de Janeiro. This area was selected for this study because it has basic maps that will support the research and the evaluation of the results. In 1982, surveys regarding Basic Cartography, Registration of Rural Properties, Maps of Soils and pedological profiles and Land Aptitude for Irrigation, besides other pedological, geological and hydrogeological maps, were performed in this area, which corresponds to approximately 250,000.00 ha.

The North Fluminense Region is considered of traditional predominance of agricultural activity, because its flat relief, typical of lowland, and tropical climate with dry winter season, Aw, according to Köppen-Geiger classification (Kottek et al., 2006), caused this region to structure its economy on the sugarcane crop. The region is bathed by the Paraíba do Sul and Muriaé rivers, which provide the availability of water to the practice of irrigated agriculture.

With the arrival of Petrobrás in the region, from the 1970's on, there has been a series of social and environmental transformations, because, with the construction of industrial centers necessary for petroleum production and the natural increase in the population, there were changes in soil use, followed by appreciation of the properties, as historically occurs in these cases.

Petroleum exploitation in the Pre-Salt layer, accompanied by infrastructure works and large industrial projects due to the construction of the COMPERJ – Petrochemical Complex of Rio de Janeiro, caused transformations in all segments of the real estate industry. Another large investment that promised to boost the region was the construction of the CLIPA – Logistic-Industrial Complex of the Açu Port. Rodrigues et al (2010) detected considerable modifications in the dynamics of the Municipality of São João da Barra. One of the changes was the super-appreciation of the properties in the municipality, for both rental and sale.

However, some recent factors can alter such dynamics in the prices of the properties of the region. First: the reduction of the petroleum barrel price to half, which on one hand decreases the royalties and on

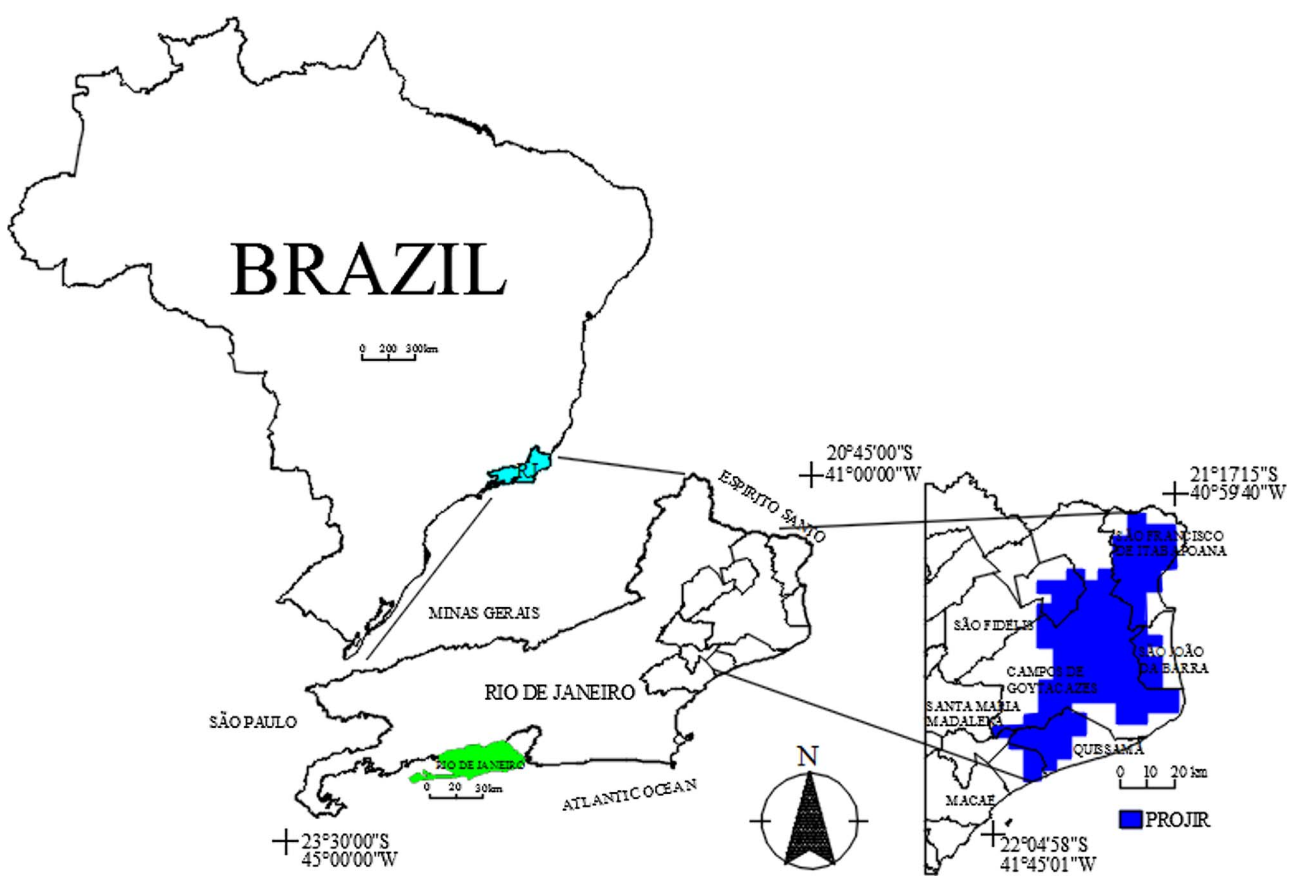

Fig. 1. Location of the study area (PROJIR project).

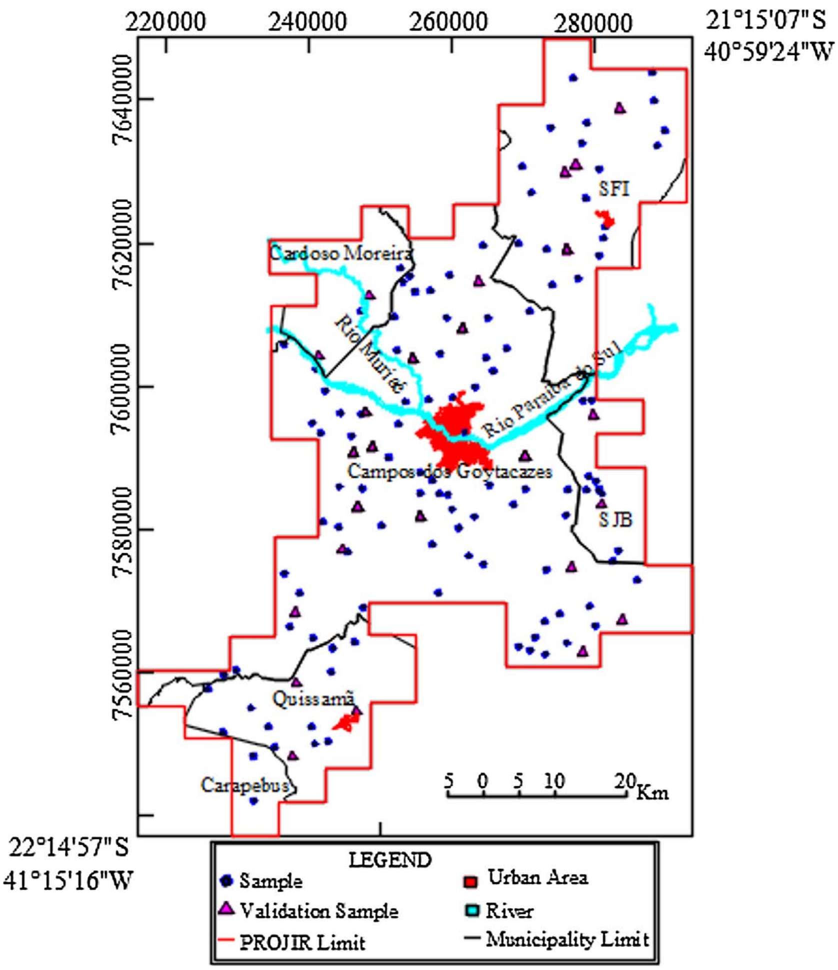

Fig. 2. Spatial distribution of the samples.

Table 1
Estimates of the parameters standard error, t value and p value of the model.

Variable	Coefficients	p-value	CI – 95% lower	CI – 95% upper
Intersection	8.97630	.0000[a]	8.59842	9.35417
AR	0.02695	.0043[a]	0.00862	0.04527
INV_DAC	1.74149	.1669	−0.74151	4.22449
INV_DS	2.01804	.0184[a]	0.347342	3.68875
DAM	0.02669	.0073[a]	0.007347	0.04603
POW_NVA	0.10171	.1063	−0.02217	0.22560
NL_TA	−0.11166	.0001[a]	−0.16869	−0.05464
TOPO	0.02574	.0125[a]	0.005785	0.045705
SIT	0.11642	.0000[a]	0.066038	0.166820
IRRIGA	0.00036	.0459[a]	6.7022×10^{-6}	0.000715
CROP	−0.03878	.0000[a]	−0.05454	−0.02303
R	0.8153			
R^2	0.6646			
AIC	68.567			
Log Likelihood – LIK	23.2835			

[a] Statistically significant (p-value < .05).

the other hand makes the Pre-Salt unviable. Second: the uncertainty caused by the financial problems of the companies from the X Group that control the Açu Port and that did not attract enough investment to make it sustainable. Third and not less relevant is the reduction in China's growth rate, which caused the decrease in the demand for *commodities*, including iron ore, which is the main product to be exported by the Açu Port.

These cited problems are already causing a decrease in the "real estate bubble" that had formed since the beginning of the Açu Port's construction, because properties that had been rented are being returned. With this, the supply of properties became greater than the demand and the values of sale and rental tend to decrease, as well as the real estate speculation.

3. Materials

The cartographic material (Plan of Rural Registration, Basic Cartography, Soil Classes and Aptitude of Agricultural Lands for Irrigation, 1: 10,000) was digitalized and only some parts of interest for the study were vectorized.

The environmental data, such as area of riparian forest, permanent preservation etc., were obtained using local images of the RapidEye satellite constellation, which was hired by the Federal Government for the Rural Environmental Registration – RER of the entire national territory (Ministério Do Meio Ambiente, 2013). These images are received orthorectified and are provided in values of radiance ($Wm^{-2} Sr^{-1} \mu m^{-1}$) multiplied by 100 (Antunes et al., 2014). Natural vegetation areas were determined through the Maxver Supervised Classification, using an acceptance threshold of 99%, because it leads to better results, generating thematic maps containing the classes of interest, which allow the identification and measurement of natural vegetation areas.

4. Methods

The samples of the properties were collected from June to August 2015, in the real estate market of the PROJIR region. In addition, the study also used data provided by INCRA – Regional of RJ, through the

Table 2
Diagnosis of spatial autocorrelation for LM.

Test	Value	p-value
LM (lag)	2.1709	.14064
Robust LM (lag)	0.8210	.36489
LM (error)	5.4681	.0193[a]
Robust LM (error)	4.1182	.04242[a]

[a] Statistically significant (p-value < .05).

Table 3
Estimates of the parameters standard error, z value and probability of the CAR model.

Variable	Coefficient	p-value	CI - 95% lower	CI – 95% upper
Intersection	8.95022	.0000[a]	8.59355	9.30690
AR	0.02452	.0031[a]	0.00825	0.04081
INV_DAC	1.90888	.1539	−0.71536	4.53313
INV_DS	2.28419	.0165[a]	0.41581	4.15259
DAM	0.01779	.0409[a]	0.00073	0.03486
POW_NVA	0.10601	.0525	−0.00118	0.21321
NL_TA	−0.09961	.0001[a]	−0.15158	−0.04765
TOPO	0.01675	.0950	−0.00291	0.03642
SIT	0.12469	.0000[a]	0.08006	0.16933
IRRIGA	0.00029	.0689	−0.00002	0.00062
CROP	−0.03698	.0000[a]	−0.05061	−0.02337
LAMBDA - λ	0.34543	.0007[a]	0.14532	0.54555
R^2	0.699215			
AIC	61.4325			
Log Likelihood - LIK	19.71626			

[a] Statistically significant (p-value < .05).

“Report Referring to the Spreadsheet of Reference Prices of the North Fluminense Region” made in February 2013 (INCRA, 2013). From this report, 32 data of properties were used, considering that the values were still updated with the real estate market of 2016.

At the end of the collection, among offers and transactions, 138 values were obtained in the different municipalities. The PROJIR area has 9685 rural properties; therefore, the obtained sample corresponds to approximately 1.42% of the total number of properties.

A random sample of these properties composed of 25 points was collected for the validation of the models and performance evaluation, which corresponds to 22.1% of the data. Ultimately, the work sample comprised 113 observations of market and their spatial distribution, as well as the points of verification (Fig. 2).

These data were subjected to an exploratory analysis and then to the Classical Multiple Linear Regression modeling (CLRM) and the verification if they satisfy the assumptions set by the Brazilian guidelines (ABNT, 2004), such as normality and autocorrelation of the residuals, multicollinearity of the independent variables, among others (Neter et al., 1990). In the following step, the matrix of spatial weights, also called neighborhood matrix, was determined in order to measure the spatial autocorrelation (Anselin, 1999). After specifying the spatial regression model, it was also validated.

The accuracy and parsimony of the CLRM and spatial regression model were evaluated by the following metrics: Akaike information criterion (AIC), coefficient of determination (R^2), likelihood function (LIK) and the standard error of the regression. In the next step involving the use of Geostatistics, the spatial smoothing technique, called Kriging, was applied (Yamamoto and Landim, 2013). This technique consists in the interpolation of the predicted values of each model in order to obtain a surface of values that will correspond to the plan of values. After defining the surface of values, the performance of the models was evaluated using the validation sample that did not participate in the modeling, through the RMSE value, to analyze the conformity of the models.

5. Results and discussion

The variables that could influence the prices of lands in the studied area were determined based on the tabulation of the market data. As dependent variable: Unit Value – UV (R$/ha) and, as independent variables: Access roads – AR, trafficability – TRAF, Distance to Appreciation Center – DAC (km), which was the Center of Campos de Goytacazes, Distance to the Sea – DS (km), Availability of water – AW, Rivers and/or creeks – RC, Dams – DAM, Riparian Forest Area – RFA (ha), Native Vegetation Area – NVA, Reforestation Area – RA (ha), Front – FRT, Depth – DEP, Back – BCK, Total Area – TA (ha), Form – FOR, Topography – TOPO, Situation – SIT, Direct Access to the Main Road – DAMR, Index of Irrigation – INIRRI, Irrigable Area – IRRIGA (ha), Crop – CROP.

In order to meet the recommendation of Medri (2011), an exploratory data analysis - EDA was performed to know the variables that will be used in the modeling, to correct preexisting problems and introduce the regression model that best fitted to the collected data.

Potential transformations were explored for both dependent and independent variables. For the dependent variable, logarithmic and

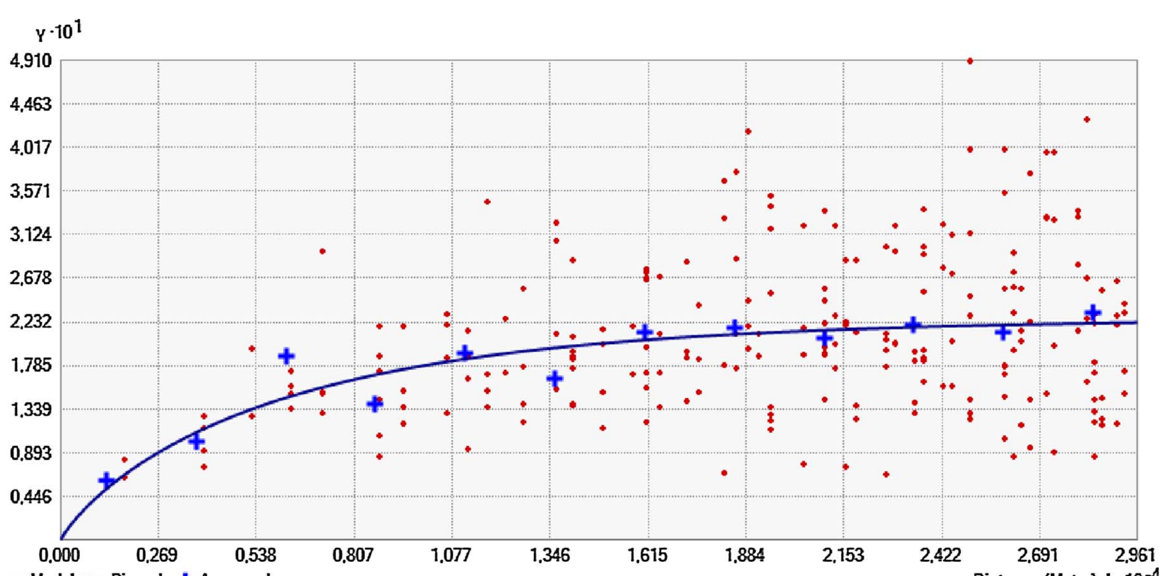

Fig. 3. Empirical semivariogram for the CLRM.

Table 4
Parameters of the theoretical semivariograms for the CLRM.

Theoretical Model	C_0-Nugget Effect	C = $C_0 + C_1$[a] – Sill	$\left(\frac{C_0}{C}x100\%\right)$	a – Range (km)	Spatial dependency
Spherical	0.06264	0.15290	40.97	18.13	Moderate
Exponential	0.01345	0.20793	6.47	18.41	Strong
Gaussian	0.08192	0.13146	62.32	14.14	Moderate

[a] C1 - structural variance.

square root transformations were tested and the results were compared with the variable without transformation, through the use of histograms, coefficients of asymmetry and kurtosis, and *Boxplot* and quantile-quantile plots. The EDA results indicated that the logarithmic transformation should be applied to the dependent variable, because, according to Dantas (2014), this transformation is preferred when one aims to fit models to data of real estate values, since the explained variables with values in the positive real interval guarantee that the variation range of the corresponding fitted values will also be positive.

It should be pointed out that the UV without transformation showed great amplitude of variation. This occurs due to the large difference of the market value of the properties located in the surroundings of the urban area of Campos dos Goytacazes, which is the most important municipality, in relation to the values practiced in the other municipalities of the PROJIR region.

For the independent variables, a matrix of partial correlations was constructed, based on the observation of the simple correlation coefficient between the independent variables, considered two by two. The matrix indicated that some variables showed multicollinearity, being removed from the modeling. The matrix indicated that some variables showed multicollinearity, being removed from the modeling. The criterion to eliminate the variables that presented multicollinearity was the correlation with the dependent variable, i.e., the dependent variable that presented the lowest correlation with the removed independent variable.

5.1. Classical linear regression model – CLRM

The result for the CLRM after various transformations iterations in the independent variables and some of the statistical tests are represented in Table 1. The variable AR was not transformed, and the values of 10, 5 and 1 were adopted for asphalt, other types of paving and dirt road, respectively. The variables DAM and CROP were not transformed and were treated as dichotomous, assuming the value of 10 for the presence of dams and sugarcane crop and 1 for the absence of dams and other types of crops. For the quantitative variables DAC and DS, inverse transformation was applied and the variable NVA was raised to the power of 0.143. For the quantitative variable TA, the natural logarithm was applied and there was no transformation for IRRIGA. The variables TOPO and SIT were not transformed; TOPO assumed values of 10 for flat, 5 for slightly rough and 1 for very rough and SIT assumed values of 10 for optimal, 8 for very good, 6 for good, 4 for regular, 2 for unfavorable and 1 for bad.

According to the results in Table 1, the signs and transformations are consistent and all coefficients were significant at the 0.05 level of significance (p-value) with the exception of the variables INV_DAC and POT_NVA. The ANOVA test was also significant at 0.01.

The independent variables are not multicollinear and 12 elements considered as outliers were removed from the sample; thus, 101 data were left. Were considered as outliers the data with standardized residuals outside the $[-2;+2]$ interval, as indicated by the *Boxplot* graphs. The normality of the residuals was verified through the *Boxplot* graph and the Jarque-Bera test (p-value = .56287).

The homoscedasticity of the residuals was tested by the Breusch–Pagan and Koenker-Bassett tests (p-value = .02880) and the results indicate that the residuals have a non-constant variance (significance level of .05). The dispersion diagram allowed to observe that only a few observations diverge from most of the data. Such instability, the heteroscedasticity, can be caused by the presence of spatial dependence in the residuals of the model.

5.2. Spatial regression

In order to begin the spatial regression analysis, one must test the residuals for spatial autocorrelation through the values of the Moran's I statistic, using the weighting matrix or matrix of spatial weights or neighborhood matrix. The objective is that this matrix captures the spatial autocorrelation present in the residuals of the CLRM determined by the Ordinary Least Squares (OLS) method. According to Demetriou (2016), the Moran's I is the most popular statistic to detect spatial autocorrelation.

The recommendation is that various tactics must be tested to construct the matrix of spatial weights, between continuity matrix,

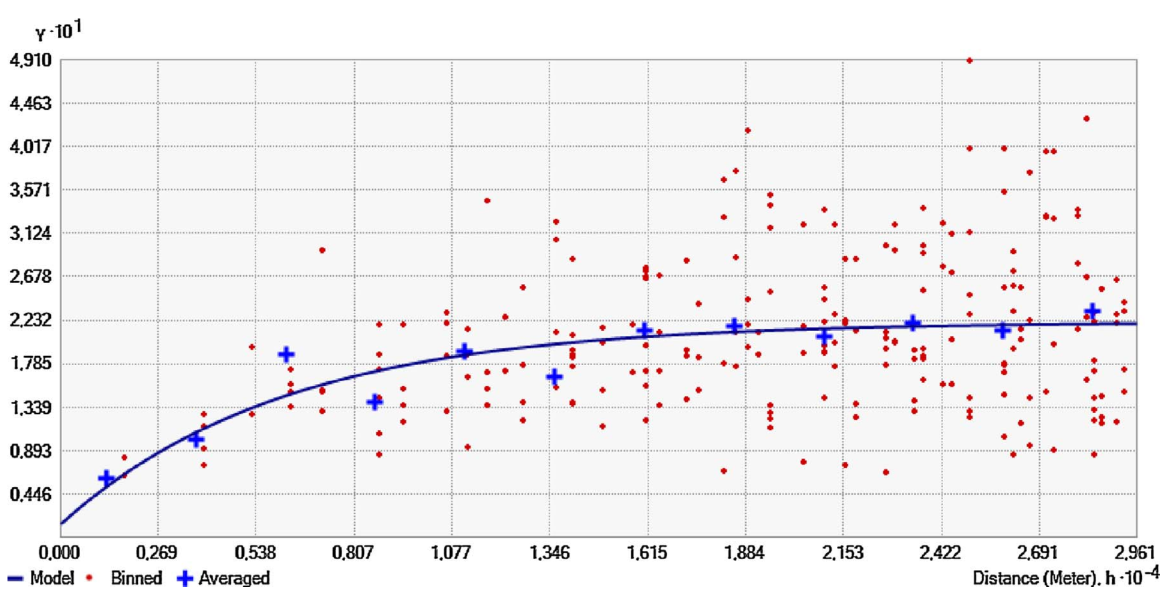

Fig. 4. Exponential theoretical semivariogram for the CLRM.

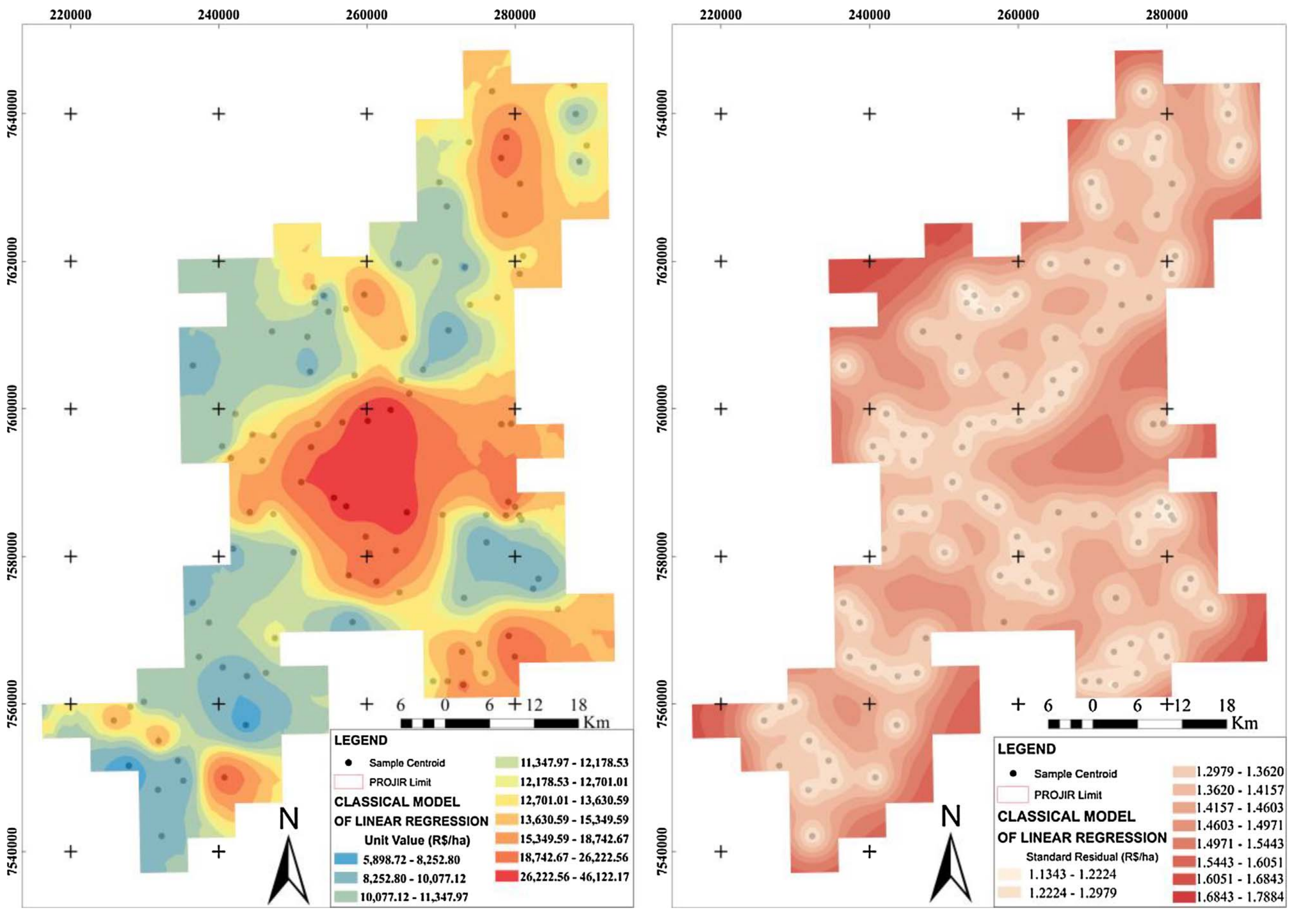

Fig. 5. Ordinary Kriging (left) and error prediction (right) for the CLRM.

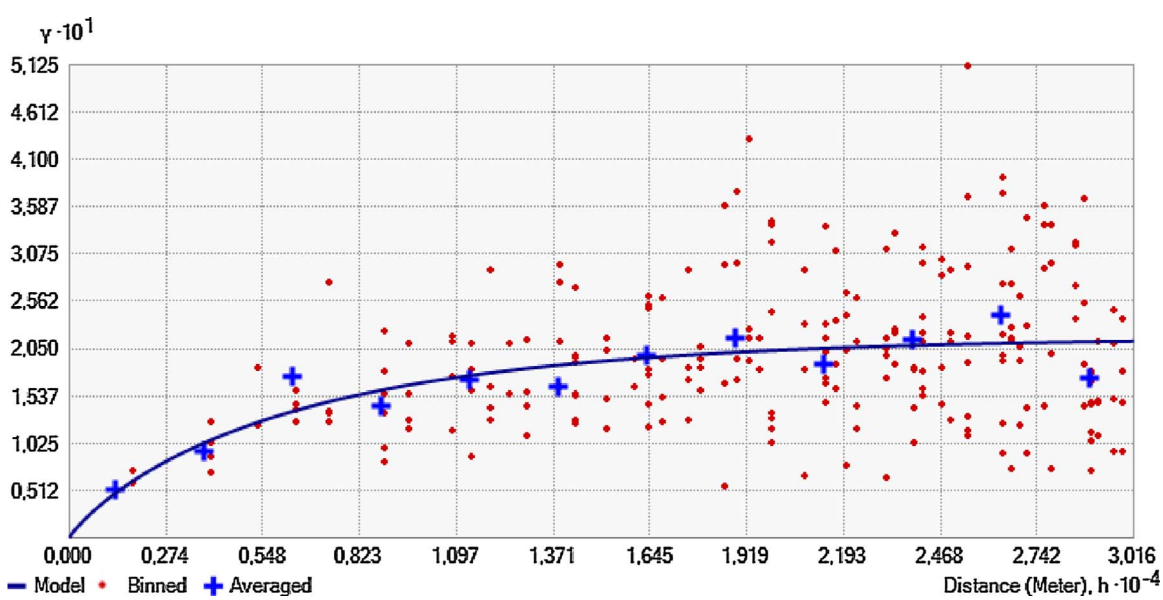

Fig. 6. Empirical semivariogram for the CAR.

geographic distance, inverse distance, closest neighbors, travel time, etc. In the case of properties appraisal, the matrix based on contiguity does not apply, because the samples are geographically separated and do not share any common limits.

Some authors use the information of range provided by the fit of a theoretical model to the semivariogram to create matrices of spatial weights based on the distance that can capture the autocorrelation existing in the data. However, Hornburg and Hochheim (2009) recommend caution in the use of the distance indicated by the semivariogram, because the obtained distance will not always be the best distance for the neighborhood matrix.

Santos (2014) fitted a theoretical model to the empirical semivariogram to define its range and used the distance to verify the spatial dependence. The author found that the distance determined by the

Table 5
Parameters of the theoretical semivariogram for the CAR.

Theoretical Model	C_0-Nugget Effect	C = $C_0 + C_1$[a] – Sill	$\left(\frac{C_0}{C}x100\%\right)$	a – Range (km)	Spatial dependency
Spherical	0.05924	0.14758	40.14	19.24	Moderate
Exponential	0.01070	0.20164	5.31	19.24	Strong
Gaussian	0.0793	0.12538	63.25	15.20	Moderate

[a] C_1 - structural variance.

semivariogram did not capture the largest part of the spatial correlation, thus reinforcing the claim of Hornburg and Hochheim (2009).

For the matrix to capture most of the spatial autocorrelation, various matrices of spatial weights were created based on two strategies: distance and neighborhood. Then, the Global Moran's I test was applied for the residuals of the value of the properties in the studied region with the software GeoDa (Anselin, 2005). All matrices based on the number of neighbors, in which 12–15 were tested, obtained Moran's I values lower than 0.1, demonstrating that this type of matrix did not capture spatial autocorrelation.

The results of the matrices based on the Euclidean distance from the centroids were more significant; the distance of 7 km captured the highest value of Global Moran's I (.18614) and the p-value was significant with less than .05 significance (p-value = .02).

The result of the Moran's I showed that there is spatial autocorrelation in the CLRM residuals; however, according to Almeida (2012), the main problem of this test, due to the diffuse nature, is that, although significant in statistical terms, it does not indicate which type of spatial autocorrelation is predominant, i.e., Moran's I does not determine whether it is in the term of the dependent variable or in the term of the error. In order to solve this problem of specification of the spatial regression models, it is recommended to use the focused tests.

The results were compared using the spatial regression model specification tests based on the Lagrange Multiplier of the dependent variable LM (lag) and of the error LM (error) and also on their robust versions, with the matrix based on the Euclidean distance of 7 km. The process of decision proposed by Anselin (2005) was used to specify the spatial regression model (dependent variable or error).

The results of the specification tests, calculated in the software GeoDa, based on the Lagrange Multiplier also in its robust form and its significance, can be observed in Table 2. It can be pointed out that both the LM (lag) and the robust LM (lag) were not significant, whereas the LM (error) and the robust LM (error) were significant at probability level less than .05.

Therefore, it is confirmed the existence of a spatial structure in the residuals of the model determined by the OLS method. The results show the existence of strong spatial autocorrelation in the residuals of the regression model, because both tests of the spatial error model were highly significant.

Since location variables (X and Y coordinates) were not used in the regression model, only some endogenous and exogenous variables of the properties, the absence of location factors in the model causes the residual to contain the effect of these non-specified variables. The term of the error showed structure of spatial dependence and this result was confirmed with the tests of spatial autocorrelation.

According to these results, the specification of the model indicated the construction of the spatial autoregressive econometric model, better known as *Spatial Error Model* (SEM or CAR). After defining the matrix of spatial weights, the CAR model was estimated considering the same previously described explanatory variables. The results obtained for the spatial error model – CAR using the GeoDa software are shown in Table 3.

The parameter LAMBDA - λ represents the term of spatial autocorrelation of the error and is significant, indicating the existence of spatial autocorrelation. It is observed that all variables are significant at .05 significance level (p-value), with the exception of the variables INV_DAC, POW_NVA, TOPO and IRRIGA.

Based on the Breusch-Pagan test, the residuals of the model do not have problems of heteroscedasticity (p-value = .16626), i.e., the coefficient of the spatial autoregressive error model was able to correct the heteroscedasticity present in the CLRM.

In order to select the best model, in the estimation through Maximum Likelihood (ML), the value of the likelihood function (LIK) was combined with the Akaike information criteria (AIC). Considering the AIC and LIK the CAR model obtained a better fit in comparison to the CLRM. In the LIK criterion, the higher the value, the better the fit, and again the CAR model was better than the CLRM.

It is essential to remember that, in the CLRM, no variables related to the location were used and, therefore, the CAR model was able to model the spatial dependence present in the residuals of this model, thus fulfilling its function. This result agrees with those of other studies, in which the use of spatial regression treats the spatial dependence and improves the classical model.

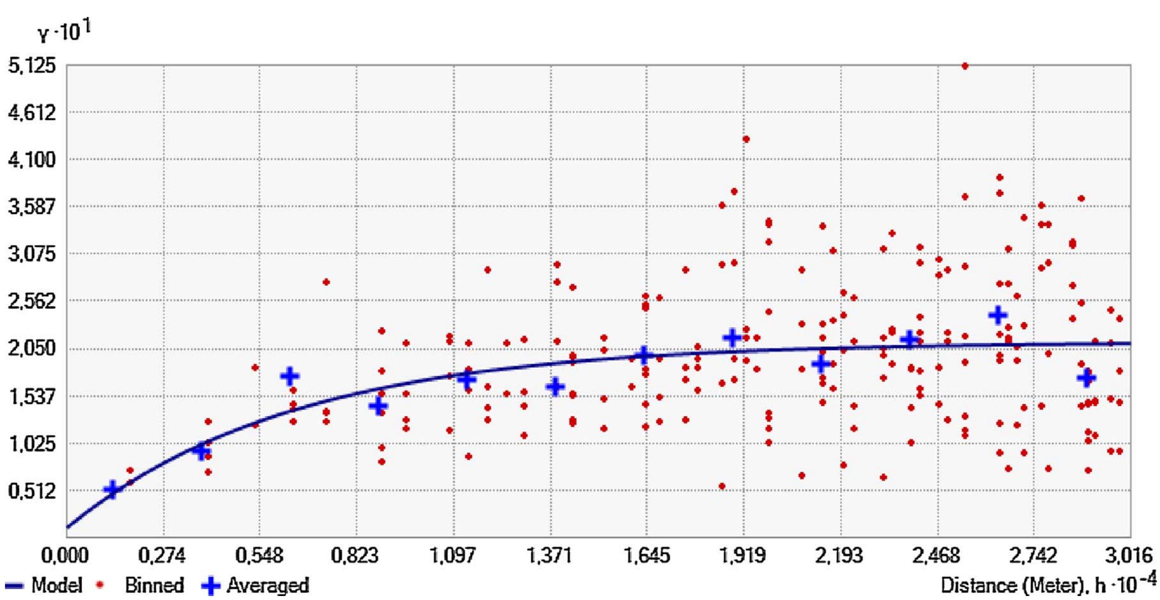

Fig. 7. Exponential theoretical semivariogram for the CAR.

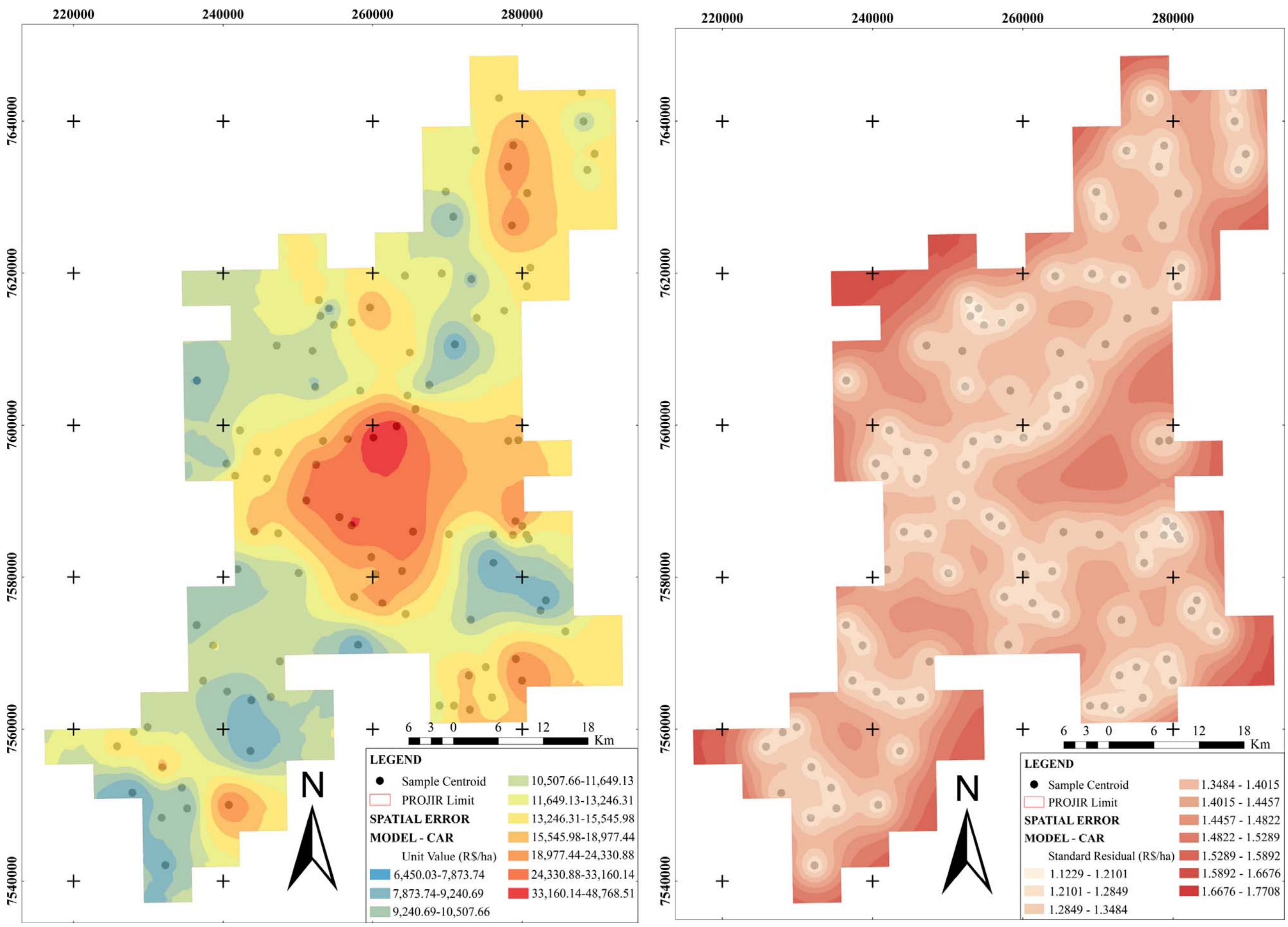

Fig. 8. Ordinary Kriging (left) and error prediction (right) for the CAR.

Table 6
RMSE values for each model after Kriging.

Map of Values - Kriging	RMSE (R$/ha)[a]
CLRM	16,216.64
CAR	16,132.02

[a] On July 11, 2016, the exchange rate was US$1.00 = R$3.30.

5.3. Elaboration of the plan of values through geostatistics

The surfaces of values in the PROJIR area will be elaborated using the set of values estimated by the CLRM and spatial regression model, through Geostatistics, using Ordinary Kriging and the software ArcGIS 10.2.1 (ESRI, 2013).

The empirical semivariogram for the CLRM was constructed with 12 points at a spacing of 2.46 km and the semivariance is increasing until approximately 22 km, when it stabilizes approaching a sill. The graphs of the omnidirectional semivariogram can be viewed in Fig. 3.

After the empirical semivariogram was obtained, the theoretical models were fitted to it in the ArGIS software. Table 4 shows the parameters obtained for the Spherical, Exponential and Gaussian models. The greatest range occurred with the Exponential model and the lowest range with the Gaussian model. The degree of spatial dependence was evaluated using the values proposed by Guimarães (2004): values lower than 25% characterize strong spatial dependence; between 25 and 75%, moderate dependence; and above 75%, weak dependence.

The results indicate that the exponential model is the best theoretical model, because it led to the lowest nugget effect and captured the highest spatial dependence. The evaluation of best fit was carried out through the cross validation, using the predicted MCRL and the values predicted by the theoretical model, which indicated the exponential model as the best fit. Fig. 4 shows the representation of this theoretical model.

The parameters obtained from the fitted theoretical model (exponential) were used to generate the surface of values for the studied area, and for the surface of prediction of the ordinary Kriging error, which can be visualized in Fig. 5.

On the left side of Fig. 5, where the surface of values is represented, it is noted that the Kriging was able to represent the variation of prices that occurs in the studied area. The highest values are related to the proximity to the urban areas of some municipalities of the PROJIR: São João da Barra, Campos dos Goytacazes and Quissamã. The area on the right side contains higher values, because of the proximity to the coastline and to the Açu Port, in São João da Barra. In the central part of the area, where the municipality of Campos dos Goytacazes is located, and above this region, there are high values, because of the proximity to the BR 101 highway, which crosses the municipality in the North direction.

The Kriging error prediction map (Fig. 5 on the right side) shows that the estimation error was lower close to the sample points and increased as the distance from the point increased. In the regions where the samples were scarce, especially on the edges, the error was higher, which demonstrates the importance and the necessity of the largest possible data set and a spatial coverage in the elaboration of the plan of values, to obtain a more accurate and reliable surface of values.

The empirical semivariogram was also determined for the predicted values of the CAR model, which had 12 points at a spacing of 2.51 km. According to Fig. 6, the semivariance is increasing until approximately

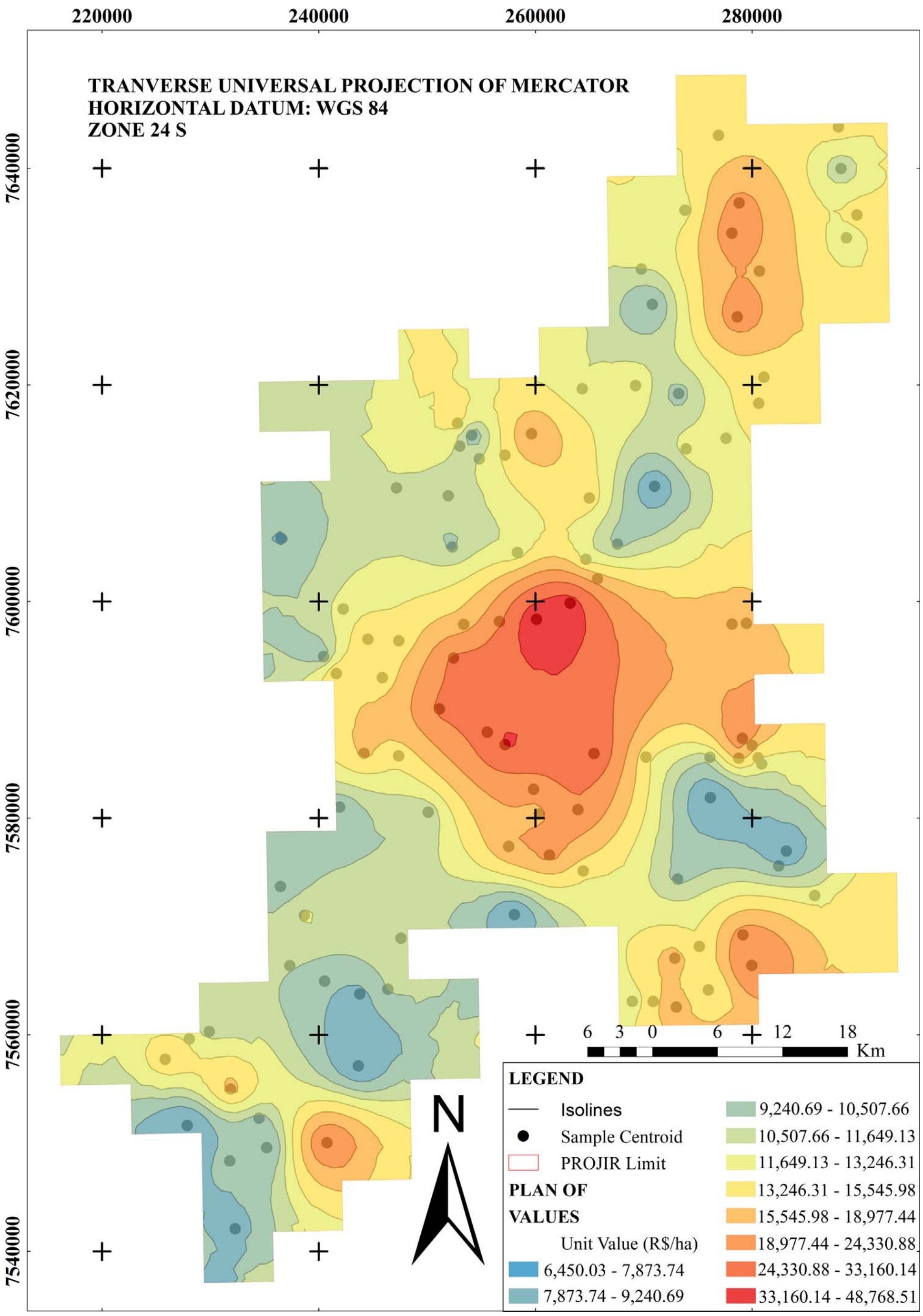

Fig. 9. Plan of Values of the PROJIR area.

22 km, when it stabilizes, characterizing a sill.

After the empirical semivariogram was obtained, the theoretical models were fitted to it. Table 5 shows the parameters obtained for the Spherical, Exponential and Gaussian models. The highest range occurred with the Exponential and Spherical models, while the lowest range occurred with the Gaussian model. The lowest nugget effect occurred in the exponential model, which showed strong spatial dependence. Therefore, we opted for the exponential model, due to its smallest nugget effect and for having captured the highest spatial dependence. The evaluation of best fit was carried out through the cross validation, using the predicted CAR and the values predicted by the theoretical model, which indicated the exponential model as the best fit. Fig. 7 shows the representation of this theoretical model.

The parameters obtained from the fitted theoretical model (exponential) were used to generate the surface of values for the studied area and for the surface of prediction of the ordinary Kriging error, which can be visualized in Fig. 8.

The left side of Fig. 8 represents the surface of values. It is noted how the Kriging allowed the representation of the price variation that occurs in the studied area, but in a more smoothed and less detailed way.

The Kriging error estimated map (Fig. 8 on the right side) also shows that the estimate was lower close to the sampling points and increased as the distance from these points increased. However, there was lower variation of the estimate errors in comparison to the Kriging performed using the CLRM values, probably because the spatial error model was able to treat and reduce such variation.

5.4. Performance evaluation

After the surface of values was obtained through ordinary Kriging, it was necessary to evaluate their performance, in order to define which of the models (CLRM and CAR) described with highest accuracy the market of lands in the PROJIR area. For that, the Root Mean Squared Error was used, as proposed by Marques et al. (2012).

$$RMSE = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(UV_{estimated} - UV_{field})^2}$$

UV_{field} is the actual unitary value of the verification sample, composed of the 25 observations. $UV_{estimated}$ was obtained from the values interpolated by Kriging on the surface of values for the same coordinates of the verification points. The results for each surface obtained in the previous step (Table 6) show that the Kriging performed using the CAR model led to the lowest RMSE; therefore, this is the model that will be used to obtain the final plan of values for the studied area.

RMSE values were high and a probable explanation is that the ordinary Kriging was performed only with the UV (dependent variable), i.e., the absence of the explanatory variables of each element (access roads, distance to appreciation center, distance to the sea, dams, native vegetation area, total area, situation, topography, irrigated area and crop) caused the unitary values to have a variation higher than that of the actual values.

Hence, the surface of values with the lowest RMSE was selected to represent the final plan of values of the PROJIR area, i.e., the surface generated through the Kriging of the CAR model. The plan of values is represented in Fig. 9, with the unitary values in R$/ha. The highest values occurred close to the urban areas of the municipalities of São João da Barra, Campos dos Goytacazes and Quissamã. There were higher values in the area on the right side, due to the proximity to the coastline and to the Açu Port, in São João da Barra, and in the central portion, above Campos dos Goytacazes, due to the proximity to the BR 101 highway, which crosses the municipality in the North direction.

The plan of values, as its name indicates, is a cartographic product that represents the generalized values of the properties of a certain area. Therefore, for the calculation of the value of each property, one must make a Rural Technical Registration to take into consideration the individualized physical and locational characteristics of the properties and, consequently, the RLT values.

6. Conclusions

The proposed methodology may be of great utility for the calculation and update of the Plans of Values for the rural areas and will also allow to analyze the sites where real estate appreciation occurs in the municipalities. Therefore, it may be used by municipal organs and state or federal public administration organs to obtain Plans of Values of rural areas for various purposes and applications.

The modeling through CLRM without the correct specification of variables related to the location proved to be inefficient, because the existence of spatial autocorrelation in the residuals of the least squares regression model was confirmed. The treatment of the spatial autocorrelation through the use of the spatial error model showed greater explanation capacity in relation to the least squares regression model for the utilized variables.

Regarding the use of Geostatistics to interpolate the values and generate the plan of values, the ordinary Kriging proved to be appropriate, allowing to generate values between the neighbors, considering that it is frequently difficult to obtain data in the field collection for the entire studied area, especially in rural areas.

Lastly, the combination of classical and spatial regression methodologies, and the combined use of geostatistical techniques proved to be adequate to elaborate and obtain the plan of values for rural areas.

Acknowledgements

We thank to the Brazilian Environmental Ministry (MMA) for furnishing the RapidEye images, to the Federal Rural University of Rio de Janeiro and to the PPGCTIA Doctor Degree Program.

References

ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2004. Avaliação de Bens Parte 3: Imóveis Rurais. NBR 14653-3. ABNT, Rio de Janeiro.
Almeida, E., 2012. Econometria espacial aplicada. Editora Alínea, Campinas. SP.
Anselin, L., 1999. Spatial Econometrics. Bruton Center. School of Social Sciences. University of Texas, Dallas. Richardson.
Anselin, L., 2005. Exploring Spatial Data with GeoDa™ : A Workbook. GeoDa 1.6.7 9 March 2015. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign.
Antunes, M.A.H., Debiasi, P., Siqueira, J.C.S., 2014. Avaliação espectral e geométrica das imagens RapidEye e seu potencial para o mapeamento e monitoramento agrícola e ambiental. Revista Brasileira de Cartografia 66, 101–109.
BRASIL. Secretaria da Receita Federal - SRF, 2012. Análise da Arrecadação das Receitas Federais – Dezembro de 2012. Disponível em: < http://www.receita.fazenda.gov.br/publico/arre/2012/Analisemensaldez12.pdf > Acesso em: 10 outubro 2013.
Dantas, R.A., 2014. Engenharia de avaliações: uma introdução à metodologia científica. Revisada de acordo com a nova versão da NBR 14.653-2:2011, 3ª edição. Pini, São Paulo.
Demetriou, D., 2016. The assessment of land valuation in land consolidation schemes: the need for a new land valuation framework. Land Use Policy 54, 487–498 Julho de 2016.
ESRI, 2013. ArcGIS Desktop: Release 10.2.1. Redlands, CA, Environmental Systems Research Institute.
Guimarães, E.C., 2004. Geoestatística Básica e Aplicada. Uberlândia. MG. .
Hornburg, R.A., 2009. Avaliação em massa com uso combinado da regressão espacial e da geoestatísitca: estudo de caso: Navegantes-SC. Dissertação (mestrado). Centro Tecnológico. UFSC, Florianópolis. SC.
Hornburg, R.A., Hochheim, N., 2009. Elaboração de Planta de Valores Genéricos com uso combinado de regressão espacial e Krigeagem. XV COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias. SP.
INCRA. Instituto de Colonização e Reforma Agrária, 2013. Relatório referente à planilha de preços referenciais da Região Norte Fluminense. RJ. .
Kottek, M., Grieser, J., Beck, C., Rudolf, B., Rubel, F., 2006. World Map of the Köppen-Geiger climate classification updated. Meteorologische Zeitschrif. 15, 259–263 Disponível em < http://koeppen-geiger.vu-wien.ac.at/pdf/Paper_2006.pdf > Acesso em 03/2015.
Marques, A.P.S., Junior, J.M., Amorim, A., Flores, E.F., 2012. Aplicação do interpolador Krigagem ordinária para a elaboração de planta de valores genéricos. Revista Brasileira de Cartografia 2 (64), 175–186.
Medri, W., 2011. Análise exploratória de dados. Curso de Especialização "Lato Sensu" em Estatística. Londrina. PR.
Ministério Do Meio Ambiente, 2013. http://www.mma.gov.br/informma/item/9319-mapa-recebe-imagens-de-satélite. Acessado em 06 de junho de.
Rodrigues, L.A., Palmeira, D.T., Lemos, L.M., 2010. Impactos sócio territoriais da implantação do Porto do Açu. XVI Encontro Nacional dos Geógrafos. ENG 2010. Porto Alegre. RS.
Neter, J., Wassermann, W., Kutner, M.H., 1990. Applied linear statistical models. Regression. analysis of variance. and experimental designs, 3ed. Richard D. Irwin, Burr Ridge (IL).
Santos, H.G., 2014. Efeitos espaciais em mercados de terras rurais: modelagem. validação e avaliação de desempenho. Dissertação (Mestrado). UFPE. Centro de Tecnologia e Geociências, Recife.
Trivelloni, C.A.P., 2005. Método para determinação do valor da localização com uso de técnicas inferenciais e Geoestatísticas na avaliação em massa de imóveis. Tese (Doutorado). UFSC. Centro Tecnológico, Florianópolis. SC.
Yamamoto, J.K., Landim, P.M.B., 2013. Geoestatística: conceitos e aplicações, 1. ed. Oficina de Textos, São Paulo 215 p.