

UFRRJ

**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA,
TECNOLOGIA E INOVAÇÃO EM AGROPECUÁRIA**

TESE

**Avaliação em Massa de Imóveis Rurais Através de
Modelagem Clássica, Espacial e Geoestatística**

Marlene Salete Uberti

2016



**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA, TECNOLOGIA E
INOVAÇÃO EM AGROPECUÁRIA**

**AVALIAÇÃO EM MASSA DE IMÓVEIS RURAIS ATRAVÉS DE
MODELAGEM CLÁSSICA, ESPACIAL E GEOESTATÍSTICA**

MARLENE SALETE UBERTI

Sob a Orientação do Professor
Mauro Antonio Homem Antunes

e

Coorientação da Professora

Paula Debiasi

Tese submetida como requisito parcial para obtenção do grau de **Doutora**, no Programa de Pós-Graduação em Ciência, Tecnologia e Inovação em Agropecuária, Área de Concentração em Recursos Naturais e Proteção Ambiental.

Seropédica, RJ
Julho de 2016

333.332
U14a
T

Uberti, Marlene Salete, 1972-
Avaliação em massa de imóveis rurais através de modelagem clássica, espacial e geoestatística / Marlene Salete Uberti. – 2016.
129 f.: il.

Orientador: Mauro Antonio Homem Antunes.
Tese (doutorado) – Universidade Federal Rural do Rio de Janeiro, Curso de Pós-Graduação em Ciência, Tecnologia e Inovação Agropecuária, 2016.
Bibliografia: f. 116-123.

1. Bens imóveis - Avaliação - Teses. 2. Propriedade rural - Avaliação - Teses. 3. Geologia – Métodos estatísticos – Teses. 4. Modelos econométricos - Teses. I. Antunes, Mauro Antonio Homem, 1963- II. Universidade Federal Rural do Rio de Janeiro. Curso de Pós-Graduação em Ciência, Tecnologia e Inovação Agropecuária. III. Título.

É permitida a cópia parcial ou total desta Tese, desde que seja citada a fonte.

**UNIVERSIDADE FEDERAL RURAL DO RIO DE JANEIRO
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA, TECNOLOGIA E INOVAÇÃO
EM AGROPECUÁRIA**

MARLENE SALETE UBERTI

Tese submetida como requisito parcial para obtenção do grau de **Doutora**, no Programa de Pós-Graduação em Ciência, Tecnologia e Inovação em Agropecuária, Área de Concentração em Recursos Naturais e Proteção Ambiental.

TESE APROVADA EM 11/07/2016.

Mauro Antonio Homem Antunes. Ph.D. UFRRJ
(Orientador)

Lucia Helena Cunha dos Anjos. Ph.D. UFRRJ

Wagner Tassinari. Dr. UFRRJ

Norberto Hochheim. Ph.D. UFSC

Simone Feigelson Deutsch. Dra. UNIRIO

Dedico este trabalho à:
Meu querido Marcelo,
marido dedicado, sempre me apoiando, ouvindo e
ajudando a continuar;
Minha querida Mariana,
filha amada que teve a paciência de esperar a sua Mãe,
fazer este trabalho para sempre depois, brincar.....
Minha amada mãe Maria, primeira professora da vida,
você que sempre dedicou a mim seu amor.....

AGRADECIMENTOS

Em primeiro lugar a Deus, obrigada por esta conquista! Não há maior recompensa na vida que atingir qualquer meta através de nosso esforço. Acontece que nada disto seria possível sem seu apoio;

Ao meu marido Marcelo e minha querida Mariana;

A toda minha família, especialmente à minha querida mãezinha, que apesar de estarmos tão distantes, sempre serão a base da minha existência;

Ao meu orientador Mauro e orientadora Paula, colegas de departamento, por acreditarem e confiarem no meu trabalho;

Ao Departamento de Engenharia, aos vários colegas professores, por aprovar meu horário especial, sem o qual esta pesquisa não teria sido possível. Especialmente ao Tiago e Ricardo que no primeiro semestre de 2016 assumiram uma das minhas disciplinas, para que eu tivesse mais horário livre;

À colega Alessandra Carreiro Baptista, por me ajudar na tarefa de entender o ArcGIS;

A todos os colegas do Curso de Engenharia de Agrimensura e Cartográfica: João Bahia, Luiz Guimarães, Mauro, Marcelo, Juliana, Paula, Rosane, Alessandra, Tiago, Ricardo, Claudia, Alessandra Palmeiro e Wagner. E também ao Eng. Vitor (Chê) por me ajudar na digitalização das cartas.

Aos meus alunos, especialmente a futura Engenheira Lais Thuany Cardoso Theodoro, bolsista PIBIC, que me auxiliou com os dados da pesquisa;

Ao Dr. Carlos Magno Magalhães da Silva do INCRA/RJ por alguns dos dados de imóveis rurais utilizados neste trabalho;

Aos corretores de imóveis Rodrigo Guimarães, Pedro Paulo Chagas de Souza e Francisco Ribeiro pelo tempo dedicado em passar os dados de imóveis da área de estudo.

A Renata Gomes, secretária do programa de pós-graduação, sempre prestativa e atenciosa.

À UFRRJ, à coordenação e professores do PPGCTIA.

RESUMO

UBERTI, Marlene Salete. **Avaliação em massa de imóveis rurais através de modelagem clássica, espacial e geoestatística**. 2016. 129f. Tese (Doutorado em Ciência, Tecnologia e Inovação em Agropecuária). Pró-Reitoria de Pesquisa e Pós-Graduação. Universidade Federal Rural do Rio de Janeiro. Seropédica. RJ. 2016.

Nas avaliações em massa de imóveis tradicionalmente são utilizados os modelos clássicos de regressão linear (MCRL), entretanto tem-se verificado a necessidade de modelar os dados espacialmente. Esta modelagem dos efeitos espaciais vem sendo utilizada principalmente nas avaliações de áreas urbanas, sendo que os valores dos imóveis nas áreas rurais também são afetados pela localização geográfica. A inexistência de metodologias de avaliação em massa de imóveis rurais é um dos motivos da evasão da receita do imposto territorial rural (ITR), pois desde que foi criado em 1964, a arrecadação deste imposto é ineficiente e inexpressiva. O objetivo deste trabalho foi a utilização de modelos econométricos de regressão espacial na modelagem dos efeitos espaciais em uma amostra de imóveis rurais para a elaboração da Planta de Valores Genéricos (PVG) em uma área da Região Norte Fluminense, estado do Rio de Janeiro. A proposta metodológica consistiu em investigar e modelar os efeitos causados pela autocorrelação espacial sobre os MCRL, avaliar seus desempenhos comparando-os com os modelos espaciais e produzir a PVG por meio da Krigagem ordinária e do estimador Kernel. A amostra utilizada contou com 113 observações e 25 amostras de verificação. Para avaliar o desempenho das superfícies de valores obtidas foram utilizadas as amostras de verificação e calculados os valores da Raiz Quadrada do Erro Médio Quadrático (REMQ) e das métricas recomendadas pela *International Association of Assessing Officers* (IAAO). Os resultados mostraram que a autocorrelação espacial pode ter seus efeitos reduzidos pelo Modelo do Erro Espacialmente Correlacionado (*Conditional Auto Regressive - CAR*) e pela Regressão Geograficamente Ponderada (RGP). A superfície gerada pelo estimador Kernel, utilizando-se os valores preditos da amostra de verificação pelo modelo RGP foi a que obteve o melhor desempenho com menor REMQ e valores do coeficiente de dispersão (COD), da mediana das razões e do Diferencial Relativo ao Preço (*Price Related Differential - PRD*) próximos dos recomendados pela IAAO. A combinação das metodologias da regressão clássica e espacial, e a utilização de técnicas de Geoestatística se mostraram adequadas para a elaboração e obtenção da PVG para áreas rurais. A metodologia proposta se mostrou aplicável nos mercados de terras rurais, pois pode ser utilizada pelos municípios para obter modelos representativos da realidade destes mercados, bem como para elaborar a PVG das áreas rurais.

Palavras-chave: Planta de valores genéricos - PVG. Avaliação em massa. Geoestatística. Superfície de valores

ABSTRACT

UBERTI, Marlene Saete. **Mass appraisal of rural land through classical, spatial and geostatistics modeling**. 2016. 129p. Thesis (Doctorate in Science Science, Technology and Innovation in Agriculture). Pró - reitoria de Pesquisa e Pós-Graduação, Universidade Federal Rural do Rio de Janeiro, Seropédica, RJ, 2016.

Traditionally the Classical Linear Regression Models (CLRMs) have been used for mass appraisal of property bulk value, however, it has been noticed the need to take into account the data spatial variation. This modeling for the geographic effects has been used mainly in urban area appraisals, while farmland values are also affected by geographic location. The lack of methodologies for mass evaluation of farmland has led to tax evasion of farmland tax revenue (ITR), as it has been inefficiently and inexpressively collected since its enactment in 1964. The objective of this paper is to use econometrics models of spatial regression in farmland comparables to produce a map of standard ground value for the Northern Region of Rio de Janeiro State, Brazil. The proposed methodology includes the investigation and modeling the effects of spatial autocorrelation on the CLRMs, to evaluate their performance comparing them with the spatial models and to produce a map of standard ground value through ordinary Kriging and kernel estimator. The sample of comparables was comprised of 113 observations for model development and 25 observations for validation. To evaluate the performance of obtained maps of values were used the validation samples to calculate the Root Mean Square Error (RMSE) values and the metrics recommended by the International Association of Assessing Officers (IAAO). The results showed that the spatial autocorrelation can have its effect predicted by the Conditional AutoRegressive model (CAR) and by the Geographically Weighted Regression (GWR). By using the values predicted with the GWR model and the validation comparables, the Kernel estimator presented the best performance on map production, yielding the lowest RMSE and dispersion coefficients, median of ratios and Price Related Differential (PRD) close to IAAO recommended values. The combination of classical and spatial regression methodologies and the use of Geostatistics techniques showed to be suitable for obtaining maps of standard ground value for farmland areas. The proposed methodology has been show applicable to farmland sales market, as it can be used by municipalities to obtain representative models of real market values, as well as to produce farmland standard ground value maps.

Keywords: Standard ground value. Mass appraisal. Geostatistics. Ground value map

RESUMEN EXTENDIDO

UBERTI, Marlene Salette. **Avalúos masivos de inmuebles rurales usando modelaje clásica, espacial y geoestadística.** 2016. 129p. (Doctorado en Ciencia, Tecnología y Innovación en la Agricultura). Pró-reitoria de Pesquisa e Pós Graduação. Universidade Federal Rural de Rio de Janeiro, RJ, 2016.

1. Introducción

Lo Brasil es un país de dimensiones continentales, con superficie territorial de 8.514.876,599 km² y los inmuebles rurales ocupan 71,10% de esta superficie. Por lo tanto, las técnicas de determinación masiva de los valores de las propiedades son de gran importancia para una serie de aplicaciones. Una de las principales aplicaciones es en la determinación del Impuesto Territorial Rural - ITR. Además de esto, otro impuesto en Brasil, que depende de la correcta valoración del inmueble es el Impuesto sobre la Transmisión de Bienes Inmóviles - ITBI, que es la responsabilidad total de las autoridades municipales. Se puede citar aún diversas otras acciones en inmuebles rurales, que también necesitan la correcta determinación de los valores tales como: la financiación, la expropiación, la compensación (en el caso de la creación de unidades de conservación o incluso em desastres ambientales), la compra de bienes raíces y venta y para fines de reforma agraria. En cada una de estas acciones son determinados valores que ni siempre siguen las técnicas de evaluación, generando valores diferentes para cada situación. Com relación al ITR, según el Secretaría de la Receta Federal - SRF (Brasil, 2012), la recaudación el año de 2012 fue de R\$ 677 millones. Utilizándose el superficie estimada por el Instituto Nacional de Colonización y Reforma Agraria - INCRA en el mismo año, 605,387,746.06 hectáreas en 2012 tiene un valor de 1,12 R\$/ha, considerada una recaudación inexpressiva e ineficiente.

Em la elaboración de las Planta de Valores Genéricos - PVGs pueden ser utilizadas técnicas estadísticas de regresión múltiple, a través de la homogeneización de los valores de una muestra de inmuebles obtenida en el mercado inmobiliario. Sin embargo, se haverificado la necesidad de incorporar em los modelos de regresión variables de localización geográfica para modelar los efectos espaciales. Esta modelización de los efectos espaciales a través de la econometría espacial, está siendo utilizada principalmente en las PVGs de zonas urbanas (TRIVELLONI, 2005; HORNBERG, 2009), siendo que los valores de los inmuebles en las áreas rurales también son afectados por la localización geográfica (SANTOS, 2014).

Sin embargo, poca investigación se ha encontrado en la literatura sobre el análisis espacial de los valores de las propiedades rurales, para la determinación de cuál la mejor técnica para la obtención de PVG. Así, esta investigación tiene como objetivo procesar estadísticamente los datos a través de la regresión clásica, espacial y geográficamente ponderada, generar la superficie de valores a través de la geoestadística (kriging y kernel,) y así elaborar la PVG para los inmuebles rurales em el área del PROJIR – Proyecto de Irrigación y Drenaje de la Caña-de-Azúcar en el Norte Fluminense - RJ. También aplicar las normas internacionales de IAAO (*International Association of Assessing Officers*), para evaluar el desempeños de los valores generados por la superficie.

Con las informaciones obtenidas será posible implementar, por ejemplo, otros mecanismos de reducción en el valor del ITR, además de los que ya existen para aquellas propiedades que mantuvieran áreas de preservación del medio ambiente en forma de reserva legal, áreas de preservación permanente o reserva particular protección natural, haciéndolo un

instrumento de política ambiental, por medio del cual el ITR funcionaría como un resarcimiento efectuado por la sociedad para el propietario que preservara el medio ambiente.

Con el desarrollo de este trabajo tiene como objetivo generar conocimientos que contribuirán para la solución de los problemas relacionados con la determinación de los precios de las propiedades rurales en Brasil, con posible aplicación de la metodología también en los países del Mercosul. En este sentido, se abordará aquí una de las cuestiones que envuelven la gestión de la tierra rural, la evaluación masiva para la determinación del valor, que tiene múltiples aplicaciones, incluyendo los impuestos territoriales.

2. Material y Métodos

Para la aplicación del estudio propuesto fue seleccionado la área del PROJIR con aproximadamente 250.000 hectáreas, que abarca los municipios de São João da Barra, Campos dos Goytacazes, Cardoso Moreira, São Francisco de Itabapoana, Conceição de Macabu, Carapebús e Quissamã, no Estado de Rio de Janeiro – RJ.

La zona de PROJIR está situada aproximadamente entre las coordenadas 21°17'15"W / 40°59'40"S e 22°04'55"W / 41°45'01"S hasta unos 280 km de la capital, Rio de Janeiro. Esta zona fue elegida para este estudio porque contienen mapas de base (Planta catastral, cartografía básica, tipos de suelo y la aptitud de las tierras para la irrigación agrícola, en una escala de 1:10.000), que prestará apoyo a la investigación y evaluación de resultados.

El material cartográfico fue digitalizado, siendo que solamente algunas partes de interés del trabajo fueron vectorizadas. Para los datos ambientales como una zona de bosque, de preservación permanente y otros temas de interés ambiental fueron utilizadas las imágenes RapidEye de la región que el Gobierno Federal contrató para todo el territorio nacional para la realización del Catastro Ambiental Rural – CAR (Ministério do Meio Ambiente, 2013).

Para realizar los tests estadísticos fueron recolectados 138 inmóviles rurales, en los diferentes municipios del PROJIR, y de estos fue retirada una muestra compuesta por 25 puntos para la validación de los modelos obtenidos y en la evaluación de desempeño, lo que corresponde al 22,1% de los datos.

Estos datos fueron sometidos a un análisis exploratória y enseguida fue realizada el Modelado Clásico por Regresión Lineal Múltiple - MCRL y la verificación de la no violación de sus presupuestos, como normalidad y no autocorrelación de los residuos, multicolinealidad de variables independientes, entre otros. En el paso siguiente, fue determinada la matriz de pesos espaciales, también llamada matriz de vecindad, para medir la autocorrelação espacial. Después de la especificación del modelo de regresión espacial - MRE, este también fue validado. En seguida fue realizada la Regresión Geográficamente Ponderada - RGP, usando el modelo de MCRL definidas anteriormente y con las mismas variables explicativas o independientes. Para ejecutar la RGP fue necesario definir algunos parámetros como la anchura de banda óptima utilizándose algunos criterios como Akaike - AIC, Akaike Corregido - AICc o la Validación Cruzada - CV. Definida la anchura el modelo RGP fue obtenido siendo representado por sus valores medios, porque este tipo de modelo genera diferentes coeficientes para cada variable independiente y para cada una de las muestras utilizadas en el análisis.

La exactitud y la precisión de los modelos MCRL, MRE y RGP fueron evaluados por las siguientes métricas: el criterio de información de Akaike (AIC), por el coeficiente de determinación (R^2), por la función de verosimilitud (LIK) y por el error normalizado de la regresión. El objetivo fue comparar el ajuste de los diferentes modelos utilizando un mismo criterio y así definir cuál de estos ha logrado el mejor ajuste. Para evaluar el desempeño de los

modelos de regressão fue separado una muestra de validación que no fue utilizada en el modelado. Para la muestra de validación fue calculada los parámetros Mediana de las Razones, el Coeficiente de Dispersión - COD y Diferencial en el Precio - PRD y comparados con los valores de referencia recomendados por la norma internacional de la *International Association of Assessing Officers* – IAAO (IAAO, 2013a).

El siguiente paso que envuelve la utilización de la Geoestadística, fue realizada las técnicas de suavizado espacial llamada de Kriging Ordinario y Kernel. Tales técnicas consisten en la interpolación de los valores predichos de cada modelo con el fin de obtener una superficie de valores que corresponden a la PVG. Con la superficie de valores definida, fue realizada la evaluación de desempeño de los modelos utilizando nuevamente la muestra de validación que no era parte de la modelización. Fueron calculados los indicadores establecidos en la norma de la IAAO, la Raíz Cuadrada del Error Medio Cuadrático - REMQ el error relativo y el valor absoluto del error relativo analizándose la conformidad de los modelos.

3. Resultados y Discusión

A partir de la tabulación de los datos de mercado fueron determinadas las variables que podrían influir en los precios de las tierras en el área de estudio. Después de varias interacciones con las variables independientes, como la utilización sin transformación, con las transformaciones logarítmicas, inversa, raíz cuadrada y potencia, y la logarítmica (indicada por el análisis exploratoria de datos) en la variable dependiente (VU), el modelo fue como sigue:

$$\text{LN[VU]} = 8,9763 + 0,026953 * [\text{VA}] + 1,7415 [\text{INV_DPV}] + 2,0180 [\text{INV_DM}] + 0,026692 * [\text{AÇU}] + 0,10172 * [\text{POT_AVN}] - 0,11167 \times [\text{LN_AT}] + 0,025745 * [\text{TOPO}] + 0,11643 * [\text{SIT}] + 0,00036128 * [\text{AIRRIGA}] - 0,03879 \times [\text{CULT}]$$

La variable VA no sufrió transformación y fue adoptado los valores asfalto 10, otros tipos de pavimento 5 y carretera de tierra 1. Las variables AÇU y CULT no fueron transformadas, siendo tratadas como variables dicotómicas asumiendo valor 10 para la presencia de represas y cultura de caña-de-azúcar y valor 1 en la ausencia de represas y otros tipos de cultivos. En las variables cuantitativas DPV y DM fue aplicada a la transformación inversa y en la variable AVN fue elevado al exponente 0,143. En la variable cuantitativa AT fue aplicado el logaritmo natural y en la AIRRIGA no fue transformada. Las variables TOPO y SIT no fueron transformadas y asumen valores 10 para plana, 5 para poco accidentada y 1 para muy accidentada para TOPO y valores iguales a 10 para situación óptima, para muy buena, 6 para buena, 4 para regular, 2 para desfavorable y 1 para mala. Analizándose los presupuestos que determinan la aprobación o no de un MCRL el modelo está aprobado, pues todas las pruebas exigidas por las normas brasileñas fueron verificados, con la excepción de la homocedasticidad.

Para iniciar el análisis de la regressão espacial fueron creadas varias matrices de pesos espaciales con base en dos estrategias: distancia y vecindad. Los resultados de las matrices basadas en la distancia Euclidiana a partir de los centroides, se mostraron más significativos, siendo que la distancia de 7 km fue la matriz que capturó el mayor valor de I de Moran Global (0,18614). El p-valor se mostró significativo con menos del 10% de significación para esta matriz de 7 km.

El resultado del I de Moran mostró que existe autocorrelación espacial en los residuos del modelo de regresión múltiple, sin embargo el I de Moran no determina si está en la

variable dependiente o se está en el error. Para resolver este problema de especificación de los modelos de regresión espacial, fueron realizados las pruebas enfocadas basados en el Multiplicador de Lagrange de la variable dependiente LM (lag) y del error LM (error) y también en sus versiones robustas. Los resultados de las pruebas enfocadas indicaron que tanto el LM (lag) y el LM robusto (lag) no se mostraron significativos, ya el LM (error) y el LM robusto (error) fueron significativos a 5% de probabilidad. Delante de estos resultados, la especificación del modelo indicó la construcción del modelo econométrico espacial autorregresivo más conocido como *Spatial Error Model* (SEM o CAR). La variable LAMBDA - λ representa el autocorrelación espacial encontrada en el error y se mostró significativa, indicando la existencia de la autocorrelación espacial.

La regresión geográficamente ponderada es una técnica que extiende el método tradicional de regresión, permitiendo que los parámetros sean estimados localmente, de modo que los coeficientes del modelo, en lugar de las estimaciones globales, son específicas a un punto i (localización). Para la función matemática pueden ser utilizados el Kernel fijo o adaptativo, las informaciones de los puntos geográficos pueden ser las coordenadas (UTM o geográfica) de las muestras. Ahora para definir la anchura de banda es necesario determinar cuál es la anchura óptima y para determinar este parámetro fueron probadas varias anchuras y utilizada la anchura que resultar en un menor criterio de información *Akaike* (AIC).

Fueron probados todos los tipos de Kernel y los criterios AIC, AICc y CV, y el valor más bajo para el criterio AIC quedó en el tipo de Kernel bicuadrado de anchura fija con valor de 44,74. Con el tipo de Kernel bicuadrado de anchura fija definido y la respectiva anchura óptima se obtuvo los coeficientes locales para cada observación.

Para escoger el mejor modelo, la recomendación es que cuando de la utilización de la estimación por Máxima Verosimilitud (MV), sea realiza en combinación con el valor de la función de verossimilhança (LIK) con los criterios de información de *Akaike* (AIC). Teniendo en consideración el criterio AIC, el modelo RGP fue lo que obtuvo el menor valor, siendo, por lo tanto, el mejor ajustado. De la misma forma el coeficiente de determinación (R^2) del RGP, fue muy superior a los otros modelos y el error estándar de la regresión tenía una pequeña disminución. En el criterio LIK cuando mayor su valor mejor será el ajuste y en este caso el RGP fue mejor que los otros modelos.

El modelo RGP fue superior al MCRL y al CAR en todos los criterios y con esto concluimos en esta etapa que la regresión geográficamente ponderada dejó el modelado más ajustado y, por lo tanto, tuvo un desempeño superior a los otros dos modelos.

Para asegurar que la evaluación del desempeño de los modelos de regresión sea realizada utilizándose datos que no fueron utilizados en el modelado, fue separada una muestra de validación con 25 elementos, los cuales fueron obtenidos en el mismo periodo que los datos utilizados en la modelización de las regresiones. El objetivo aquí es verificar el funcionamiento de los mismos en una situación más próxima de su aplicación práctica.

La muestra de validación fue utilizada para el cálculo de los parámetros definidos por la IAAO: Mediana de las Razones, Coeficiente de Dispersión (COD) y Diferencial en el Precio (PRD). Los resultados fueron discutidos llevándose en consideración los parámetros de referencia para evaluación del desempeño definidos por la norma internacional. A pesar de la expectativa delante de los resultados del criterio de *Akaike* (AIC) de que el modelo RGP tendría el mejor ajustamiento, los valores del coeficiente de dispersión no confirmaron el resultado anterior.

Con relación a la Mediana de las Razones, todos los modelos están por debajo del límite mínimo de referencia establecidos por la norma de la IAAO. Con relación al COD todos los modelos estaban con los valores por encima de los recomendados por la norma

internacional, siendo que el modelo RGP fue lo que presentó el peor resultado. Probablemente esto ocurrió porque el área de estudio del PROJIR comprende más de un municipio, y los valores de los inmuebles en la mayoría de los casos, tiene comportamientos diferentes dentro de sus respectivos límites municipales, o sea, el análisis al nivel local (que es el caso del RGP), mientras que una sola área, generó una dispersión de los valores mayor que las otras técnicas. La solución para este problema sería analizar los modelos separando cada municipio y verificando aisladamente el comportamiento de los valores de los inmuebles rurales. Con relación al PRD los tres modelos están abajo de los límites de referencia, o sea, ocurrió progressividade en la evaluación.

Los tres modelos fueron utilizados para generar la superficie de valores a través de la Geoestadística, mediante el método Kriging Ordinario y el estimador Kernel. La superficie de los valores generados por el método Kriging y por el Kernel consiguieron representar la variación de precios que ocurre en el área de estudio. Los mayores valores están relacionados a la proximidad a las áreas urbanas de algunos municipios del PROJIR: São João da Barra, Campos dos Goytacazes e Quissamã. En el área a la derecha también ocurre valores más altos, debido a la proximidad con el litoral y con el Porto do Açú, en São João da Barra. En la parte central del área, donde está el municipio de Campos y por encima de esta región también ocurre valores altos, en función de la proximidad de la BR 101, que corta el municipio en la dirección norte. El mapa de la previsión del error mostró que el error de estimación fue más pequeña en las proximidades de los puntos de muestreo y crece a medida que se aleja de estos puntos. En las regiones donde las muestras fueron escasas, principalmente en los bordes, el error fue mayor lo que muestra la importancia y la necesidad, de la mayor cantidad posible de datos y una buena espacialização, para obtener una superficie de valores más precisa y confiable.

Después de la obtención de las superficies de valores mediante el método Kriging ordinario, fue necesaria una nueva evaluación de desempeño, para definir cuál de los modelos (MCRL, CAR o RGP) representó con más precisión el mercado de tierras en el área del PROJIR. Para la Raíz Cuadrada del Error Medio Cuadrático (REMQ) propuesto por Marques et al. (2012), muestran que tanto para el método Kriging como para el Kernel, realizado a partir de los valores previstos por el modelo RGP fue a que proporcionó el valor más bajo de la REMQ en comparación con otros modelos. Comparándose con los valores de Kriging ordinario con los valores del Kernel para el REMQ, se observa que el Kernel mejoró los valores, pues disminuyó en aproximadamente 4 % el valor de la raíz cuadrada del error medio cuadrático. Para el modelo RGP esta disminución fue de 4,47 %.

Siendo así, fue escogida la superficie de valores que presentó el menor REMQ, y menor error relativo para representar la PVG final del área de PROJIR, o sea, la superficie generada a partir del estimador Kernel del modelo RGP.

4. Conclusiones

La metodología propuesta posibilitó obtener el modelo de regressão, así como la elaboración de la PVG, ambos representantes de la realidad de los valores de mercado para inmuebles rurales en el área del PROJIR. Para la evaluación masiva de todos los inmuebles individualmente, el modelo de regressão RGP fue el más indicado, pues utiliza las características (variables) que influencian los valores de las tierras en el municipio. La PVG obtenida por la interpolación de los valores de muestreo debe ser utilizada para indicar el valor genérico de los inmuebles en las diferentes regiones del área del PROJIR.

La combinación de las metodologías propuestas fue eficiente para modelar los valores

de los mercados de tierras rurales. A través de la econometría espacial fue posible controlar la autocorrelación espacial existente en el MCRL y utilizar la regresión geográficamente ponderada y por medio de la Geoestadística, con el interpolador de Kriging ordinario y Kernel fue verificado que los mismos, pueden ser utilizados para generar superficies de valores para los modelos obtenidos. La combinación de metodologías ha permitido atender los objetivos del trabajo.

Para mejorar el desempeño del método Kriging ordinario es recomendada la utilización de la cokriging. La cokriging es un procedimiento geoestadístico según el cual diversas variables regionalizadas pueden ser estimadas en conjunto, con base en la correlación espacial entre ellos. Es una extensión multivariada del método de Kriging cuando para cada local amostrado se obtiene un vector de valores en lugar de un único valor. Con esto se espera utilizar los atributos estructurales y de localización de los inmuebles y no solamente la utilización de los valores unitarios en la interpolación para generar la superficie de las PVGs.

En la evaluación masiva de inmuebles urbanos trabajos más recientes utilizan una clase de modelos de regresión llamados de modelos aditivos generalizados, métodos de regresión Bayesianos y redes neurales artificiales, entre otros. Se sugiere probar estas técnicas en la evaluación masiva de inmuebles rurales para comprobar si los mismos mejoran el desempeño de los valores.

Palabras clave: Avalúo masivo. Kriging ordinario. Estimador Kernel. Superficie de valores. Regresión geográficamente ponderada.

LISTA DE ABREVISTURAS E SIGLAS

AIC - Critério de Informação Akaike
AICc - Critério de Informação Akaike corrigido
ABNT – Associação Nacional de Normas Técnicas
ANN – Rede Neural Artificial
BIC - Critério de Informação Bayesiano
CAR - Conditional Auto Regressive
COD – Coeficiente de Dispersão
CF – Constituição Federal
CTM – Cadastro Técnico Municipal
CTN – Código Tributário Nacional
CTR - Cadastro Técnico Rural
DITR - Declaração do Imposto sobre a Propriedade Territorial Rural
GWR - Geographically Weighted Regression
GU – Grau de Utilização
IAAO - International Association of Assessing Officers
INCRA - Instituto Nacional de Colonização e Reforma Agrária
ITBI - Imposto sobre a Transmissão de Bens Imóveis
ITR - Imposto Territorial Rural
IPTU - Imposto Predial e Territorial Urbano
IVDN - Índice de Vegetação por Diferença Normalizada
LIK - Função de Verossimilhança
LM - Multiplicador de Lagrange
LR - Likelihood Ratio
MCRL – Modelo Clássico de Regressão Linear
MMO - Mínimos Quadrados Ordinários
MMQ - Método dos Mínimos Quadrados
MRA - Multiple Regression Analysis
MV - Máxima Verossimilhança
NBR – Norma Brasileira
PRD - Price Related Differential
PVG - Planta de Valores Genéricos
PROJIR – Projeto de Irrigação e Drenagem da Cana-de-Açúcar na Região Norte Fluminense
RGP - Regressão Geograficamente Ponderada
SAR - Spatial Auto Regressive
SC – Critério de Schwartz
SEM - Spatial Error Model
SIPT - Sistema de Preços de Terras
SIG - Sistemas de Informação Geográfica
SRF - Secretaria da Receita Federal
STF - Superior Tribunal Federal
TSA - Superfície de Tendência
TVR - Teoria das Variáveis Regionalizadas
VTNt - Valor da Terra Nua Tributável

VTN - Valor da Terra Nua
VR - Variável Regionalizada
UVA - Use Value Programs
UTM - Universal Transversa de Mercator

LISTA DE FIGURAS

Figura 1. Diagrama de dispersão de Moran, outliers espaciais e pontos de alavancagem (CARVALHO, 2008).	24
Figura 2. Variável aleatória regionalizada $Z(x)$ (GENÚ, 2004).....	27
Figura 3(a) e (b). Componentes de uma variável regionalizada (CAMARGO et al., 2004). .	28
Figura 4. Amostragem em duas dimensões (CAMARGO et al., 2004).	30
Figura 5. Parâmetros do variograma (CAMARGO et al., 2004).	31
Figura 6. Representação gráfica de modelos transitivos normalizados (CAMARGO et al., 2004).	33
Figura 7. Representação gráfica de semivariogramas experimentais e modelos teóricos com efeito pepita (CAMARGO et al., 2004).....	34
Figura 8. À esquerda Anisotropia Geométrica e à direita Anisotropia Combinada (CAMARGO et al., 2004).	35
Figura 9. Kernel espacial (ALMEIDA, 2012).....	37
Figura 10. Estimador de intensidade para um padrão de pontos	41
Figura 11. Localização do PROJIR	47
Figura 12. Distribuição espacial das amostras	51
Figura 13. Fluxograma dos métodos estatísticos.....	54
Figura 14. Histograma da variável dependente: na escala original	59
Figura 15. Histograma da variável dependente: transformação logarítmica	59
Figura 16. Histograma da variável dependente: transformação raiz quadrada	60
Figura 17 a, b, c. Gráficos Boxplots para a variável dependente na escala original; transformação logarítmica; transformação raiz quadrada.....	61
Figura 18. Gráfico Quantil-Quantil da variável dependente na escala original	61
Figura 19. Gráfico Q-Q da variável dependente na transformação logarítmica.....	62
Figura 20. Gráfico Q-Q da variável dependente na transformação raiz quadrada	62
Figura 21. Gráfico de Indicação de outliers	71
Figura 22. Histograma e Box plot dos Resíduos Padronizados	72
Figura 23. Gráfico da Homocedasticidade	72
Figura 24. Gráfico da Autocorrelação	74
Figura 25. Diagrama de dispersão de Moran para os valores previstos do VU	75
Figura 26. Diagrama de dispersão de Moran para a variável dependente e os resíduos do modelo vazio.....	77
Figura 27. Semivariograma experimental para MCRL	84
Figura 28. Semivariograma teórico exponencial para MCRL	85
Figura 29. Mapa da Krigagem ordinária para o MCRL	86
Figura 30. Mapa da previsão do erro para o MCRL	87
Figura 31. Semivariograma experimental para o CAR	88
Figura 32. Semivariograma teórico exponencial para o CAR.....	89

Figura 33. Mapa da Krigagem ordinária para o CAR	90
Figura 34. Mapa da previsão do erro para o CAR.....	91
Figura 35. Semivariograma experimental para o RGP.....	92
Figura 36. Semivariograma teórico Gaussiano para o RGP.....	93
Figura 37. Mapa da Krigagem ordinária para o RGP.....	94
Figura 38. Mapa da previsão do erro para o RGP	95
Figura 39 a, b, c. Anisotropia para o MCRL, CAR e RGP	96
Figura 40. Validação cruzada para MCRL.....	99
Figura 41. Mapa do Kernel para o MCRL	100
Figura 42. Mapa da previsão do erro para o MCRL	101
Figura 43. Validação cruzada para o CAR.....	102
Figura 44. Mapa do Kernel para o CAR	103
Figura 45. Mapa da previsão do erro para o CAR.....	104
Figura 46. Validação cruzada para o RGP	105
Figura 47. Mapa do Kernel para o RGP.....	106
Figura 48. Mapa da previsão do erro para o RGP	107
Figura 49. Planta de Valores Genéricos da área do PROJIR	111

LISTA DE TABELAS

Tabela 1. Métodos de avaliação imobiliária em alguns países	6
Tabela 2. Alíquotas para apuração do ITR	13
Tabela 3. Quantidade de amostras em cada município	50
Tabela 4. Distribuição das áreas	51
Tabela 5. Resumo das variáveis independentes.....	58
Tabela 6. Coeficientes de assimetria e curtose para a variável dependente	60
Tabela 7. Estatística descritiva para a variável dependente sem transformação, transformação logarítmica e transformação raiz quadrada.....	63
Tabela 8. Matriz de correlações parciais para as variáveis independentes.....	64
Tabela 9. Estimativas dos parâmetros, erro-padrão, valor t e o valor p do modelo	68
Tabela 10. Resultados gerais do modelo de regressão	69
Tabela 11. Análise de variância - ANOVA.....	70
Tabela 12. Matriz de correlações parciais	70
Tabela 13. Distribuição dos Resíduos Normalizados.....	71
Tabela 14. Teste I de Moran e significância para as matrizes de ponderação espacial testadas sobre os resíduos.....	74
Tabela 15. Diagnóstico de autocorrelação espacial para o LM.....	76
Tabela 16. Teste I de Moran e significância para o modelo vazio	77
Tabela 17. Estimativas dos parâmetros, erro-padrão, valor z e probabilidade do modelo CAR	78
Tabela 18. Resultados dos tipos de Kernel com o critério AIC.....	79
Tabela 19. Resultados dos tipos de Kernel com o critério AICc.....	79
Tabela 20. Resultados dos tipos de Kernel com o critério CV.....	80
Tabela 21. Estimativa dos parâmetros RGP: coeficiente médio e erro padrão	80
Tabela 22. Comparativo entre os modelos por meio do AIC, R ² , LIK e erro padrão.....	81
Tabela 23. Valores recomendados pela IAAO para avaliações em massa de imóveis.....	82
Tabela 24. Resultados das métricas para avaliação de desempenho com as 25 observações .	82
Tabela 25. Resultados das métricas para avaliação de desempenho com a amostra saneada .	82
Tabela 26. Parâmetros para o semivariograma experimental modelo MCRL.....	84
Tabela 27. Parâmetros dos semivariograma teóricos para MCRL	85
Tabela 28. Valores da validação cruzada dos modelos teóricos para MCRL	85
Tabela 29. Parâmetros para o semivariograma experimental modelo CAR.....	88
Tabela 30. Parâmetros dos semivariograma teóricos para o CAR	89
Tabela 31. Valores da validação cruzada dos modelos teóricos para o CAR.....	89
Tabela 32. Parâmetros para o semivariograma experimental para o RGP	92
Tabela 33. Parâmetros dos semivariograma teóricos para o RGP.....	93
Tabela 34. Valores da validação cruzada dos modelos teóricos para o RGP	93

Tabela 35. Valores da REMQ para cada modelo após Krigagem	97
Tabela 36. Resultados das métricas para avaliação de desempenho pós Krigagem.....	97
Tabela 37. Valores da validação cruzada das funções Kernel para MCRL.....	98
Tabela 38. Valores da REMQ para cada modelo com Kernel.....	108
Tabela 39. Resultados das métricas para avaliação de desempenho pós Kernel.....	108
Tabela 40. Resíduo relativo e absoluto da Krigagem e Kernel	110

SUMÁRIO

1 INTRODUÇÃO	1
2 REVISÃO DE LITERATURA	5
2.1 Avaliação em Massa de Imóveis Rurais	5
2.1.1 Conceitos e métodos de avaliação	5
2.1.2 Variáveis formadoras do valor em imóveis rurais	7
2.1.3 Avaliação de desempenho em avaliações em massa	10
2.1.4 ITR: apuração e cálculo	11
2.2 Modelo Clássico de Regressão Linear – MCRL.....	14
2.3 Estatística Espacial.....	17
2.4 Econometria (regressão) Espacial	19
2.5 Geoestatística	26
2.5.1 Introdução	26
2.5.2 Teoria das variáveis regionalizadas	27
2.5.3 Determinação experimental do semivariograma	30
2.5.4 Modelos teóricos de semivariograma	32
2.5.5 Anisotropia	34
2.5.6 Krigagem	35
2.7 Medidas de Qualidade de Ajuste.....	40
2.8 O Interpolador Kernel	41
2.9 Exemplos de Aplicação em Avaliações de Imóveis	43
3 MATERIAL E MÉTODOS	46
3.1 Materiais.....	46
3.2 Área de Aplicação da Pesquisa	46
3.3 Obtenção dos Dados.....	49
3.3.1 Coleta dos dados de mercado imobiliário.....	49
3.3.2 Georreferenciamento e vetorização	52
3.3.3 Classificação das imagens <i>RapidEye</i>	52
3.4 Aplicação dos Métodos Estatísticos.....	53
4 RESULTADOS E DISCUSSÃO	56
4.1 Identificação das Variáveis Influenciantes.....	56
4.2 Análise e Preparação dos Dados	58
4.1.1 Variável dependente: valor unitário.....	59
4.1.2 Variáveis independentes	63
4.3 Modelo Clássico de Regressão Linear - MCRL	66
4.4 Análise da Regressão Espacial.....	73
4.5 Análise da Regressão Geograficamente Ponderada – RGP	78
4.6 Avaliação da Qualidade do Ajuste e Desempenho dos Modelos.....	80
4.7 Elaboração da PVG Através da Geoestatística	83

4.8	Elaboração da PVG Através do Estimador Kernel	97
4.9	PVG: Krigagem Ordinária ou Estimador Kernel	109
5	CONCLUSÕES	112
6	CONSIDERAÇÕES FINAIS	115
7	REFERÊNCIAS BIBLIOGRÁFICAS	116
8	ANEXOS	124

1 INTRODUÇÃO

O Brasil é um país de dimensões continentais, com superfície territorial de 8.514.876,599 km² e segundo o Instituto Nacional de Colonização e Reforma Agrária – INCRA (INCRA, 2012) a área rural ocupa 605.387.746,06 ha¹ (6.053.877,4606 km²), ou seja, os imóveis rurais ocupam 71,10 % da superfície territorial do País. Os números mencionados ilustram a importância da definição precisa e socialmente justa dos mecanismos e instrumentos de gestão do espaço rural, que entre outros, estão o Cadastro Técnico Rural - CTR e o Imposto Territorial Rural - ITR.

A principal característica imposta pela Lei Federal nº 9.393/96 (BRASIL, 1996), que regulamenta o imposto, foi a autodeclaração dos valores da terra e suas benfeitorias. O valor do imposto é apurado aplicando-se sobre o Valor da Terra Nua Tributável – VTNT, a alíquota de acordo com o grau de utilização e a área total do imóvel.

Esta forma de cálculo do imposto criada para detectar as propriedades improdutivas vem sendo criticada por muitos profissionais como juristas, agrônomos, avaliadores, etc. Os juristas consideram a progressividade da alíquota aplicada ao imposto sobre a propriedade territorial inconstitucional (MACHADO, 2004), sendo que o Superior Tribunal Federal - STF já julgou afirmando o ITR como um imposto constitucional. Os agrônomos consideram que a tabela de alíquotas fere o princípio da isonomia, pois atribui valores diferenciados de imposto para uma mesma capacidade de produção (ARANTES, 2005), ao passo que para os avaliadores na determinação do valor de imóveis rurais deve-se levar em conta a capacidade de uso do solo, as condições de acesso e a localização do mesmo, entre outros fatores (LIMA, 2011).

Segundo a Secretaria da Receita Federal - SRF (BRASIL, 2012), agente que arrecada atualmente o imposto, a arrecadação do ITR no ano de 2011 foi de R\$ 603 milhões e no ano de 2012 foi de R\$ 677.395.295,00. Utilizando-se a área estimada pelo INCRA em 2012 de 605.387.746,06 ha e a arrecadação do imposto no mesmo ano, teremos um valor de 1,12 R\$/ha. Com relação à carga tributária do Brasil no ano de 2012, o ITR teve a contribuição de 0,07% do total arrecadado.

A constituição de 1988 e a aprovação da Lei nº 11.250/2005 (BRASIL, 2005) permitiu a realização de convênios entre a SRF e os municípios, com o objetivo de que os próprios municípios realizem a apuração e a fiscalização do ITR. Apesar deste convênio trazer vantagens aos municípios, aumentando a arrecadação para 100% do imposto, após vários anos da entrada em vigor da legislação, apenas 37%² dos municípios aderiram ao mesmo. No estado do Rio de Janeiro o índice é de 30% de municípios conveniados e mais especificamente na Região Norte Fluminense os municípios de Campos dos Goytacazes, São Francisco de Itabapoana, Magé e Quissamã firmaram convênio com a SRF (situação em março de 2016). Um dos problemas é que os municípios enfrentam dificuldades para construir e manter atualizadas suas Plantas de Valores Genéricos - PVGs, o que reflete em inexpressiva arrecadação.

¹ O INCRA exclui 266.332 imóveis inconsistentes. Situação em abril de 2012.

² Segundo a Receita Federal somente 2.109 municípios optaram pelo convênio. Situação em março de 2016.

Os municípios conveniados devem informar anualmente a SRF, o Valor da Terra Nua por hectare (VTN/ha) para atualizar o Sistema de Preços de Terras – SIPT. Os proprietários de imóveis rurais declaram anualmente o valor dos seus imóveis que são comparados aos informados pelas prefeituras à SRF, sendo que se estiver em desacordo servirá de referência para o início de um processo de fiscalização por parte do Órgão responsável. **A pergunta é: como os municípios estão apurando o VTN/ha?**

Além do ITR, no Brasil, outro imposto depende da correta valoração do imóvel: o Imposto sobre a Transmissão de Bens Imóveis – ITBI, que é de total competência das Prefeituras Municipais. O ITBI tem como base de cálculo o valor venal e geralmente as prefeituras utilizam o valor declarado pelo comprador do imóvel ou o valor tabelado pela própria prefeitura (retirado da PVG), se este for maior que o declarado. Pode-se citar ainda diversas outras ações em imóveis rurais, que também precisam da correta determinação dos valores como: financiamentos, desapropriações, indenizações (em caso de criação de unidades de conservação ou até mesmo em desastres ambientais), mercado imobiliário de compra e venda, reforma agrária, etc. Em cada uma destas ações são determinados valores que nem sempre seguem técnicas avaliatórias, gerando valores diferentes para cada situação.

Apesar da evidente importância e da complexidade do assunto ainda são escassos os trabalhos nesta área, assim como a falta de ação governamental em dar soluções mais adequadas à questão do conhecimento sobre a área rural e da tributação de 71,10% do seu território, que poderia arrecadar muito mais com o ITR.

Citando o estudo de Silva e Barreto (2014), que analisaram o potencial do ITR na Amazônia, os autores estimaram que no estado do Pará, o potencial de arrecadação do ITR por hectare seria 133 vezes maior do que tem sido arrecadado (12,00 R\$/ha em vez de 0,09 R\$/ha).

A inexistência de metodologias de avaliação em massa de imóveis rurais é um dos motivos para a evasão de receita. O modelo atual de levantamento de informações cadastrais, do cálculo e de cobrança do ITR é ineficiente e até hoje nunca atendeu, desde sua criação, os objetivos da Lei 4.504/64, mais conhecida como Estatuto da Terra (BRASIL, 1964). Somente com a implantação do Cadastro Técnico Rural e a adequação dos mecanismos de cálculo e cobrança, o ITR poderá se tornar um mecanismo de gestão do espaço rural e de regularização fundiária.

Na elaboração das PVGs podem ser utilizadas as técnicas estatísticas da Regressão Múltipla, através da homogeneização dos valores de uma amostra de imóveis coletadas no mercado imobiliário. Entretanto tem-se verificado a necessidade de incorporar nos modelos de regressão variáveis de localização, para modelar os efeitos espaciais, com o objetivo de atingir maior precisão e justiça fiscal. Esta modelagem dos efeitos espaciais, através da econometria espacial, está sendo utilizada principalmente nas PVGs de áreas urbanas, sendo que os valores dos imóveis nas áreas rurais também podem ser afetados por estes efeitos.

Buscando dar respostas às críticas e aos questionamentos, o presente trabalho de pesquisa testa a hipótese de que a avaliação em massa de imóveis rurais, pode ser utilizada de maneira eficiente no cálculo do ITR. Baseando-se nas metodologias avaliatórias preconizadas pela NBR 14.653:3 - Avaliação de Bens: Imóveis Rurais (ABNT, 2004), pretende-se realizar um estudo para definir o método estatístico mais adequado para a elaboração da PVG para a área rural. Desta forma, os valores dos imóveis rurais deixariam de ser declaratórios, e com a utilização de métodos estatísticos na avaliação em massa alcançariam mais precisão e menos subjetividade.

O objetivo geral desta pesquisa é determinar e aplicar a metodologia estatística mais adequada para a elaboração de PVGs para áreas rurais, utilizando os dados e mapas do PROJIR - Projeto de Irrigação e Drenagem da Cana-de-Açúcar na Região Norte Fluminense – RJ.

Os objetivos específicos são:

- a) Utilizar no tratamento dos elementos amostrais, modelos de econometria: regressão clássica e regressão espacial, bem como a regressão linear geograficamente ponderada – RGP, para comparar os valores obtidos e definir a metodologia de avaliação em massa para áreas rurais;
- b) Avaliar os imóveis rurais da amostra, utilizando o método comparativo direto de dados de mercado, o método evolutivo e o método da quantificação do custo, conforme NBR 14.653:3 (ABNT, 2004);
- c) Aplicar a Geoestatística através da Krigagem ordinária e do Kernel, para interpolar os valores preditos da regressão clássica, espacial e RGP a fim de obter a superfície de valores para a área do PROJIR;
- d) Aplicar as normas internacionais da IAAO (*International Association of Assessing Officers*), para avaliar o desempenho da superfície de valores;
- e) Contribuir para o estudo dos processos e variáveis explicativas (atributos estruturais, de localização, acessibilidade e vizinhança, entre outros) da valorização do solo rural, dentro de um contexto local específico, a fim de acumular conhecimentos e reunir subsídios para estudos e aplicações futuros.

Como relevância do trabalho, tem-se que os valores das propriedades para fins de tributação sempre foram declaratórios, ou seja, o proprietário determina o valor da sua propriedade, por isso a avaliação em massa ou planta de valores genéricos, nunca foi realmente aplicada para a área rural. A maioria das pesquisas e trabalhos tem sido aplicados em áreas urbanas (FERREIRA, 2007; GONZÁLES, 2002; HORNBURG, 2009; TRIVELLONI, 2005), pois os valores da PVG são a base de cálculo de alguns impostos de responsabilidade das Prefeituras Municipais, sendo que para os imóveis rurais existem poucos estudos que promovam a aplicação destas técnicas.

Para a realização deste trabalho, será utilizado principalmente o Método Comparativo Direto de Dados de Mercado que consiste em identificar o valor do bem, por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes de uma amostra coletada no mercado imobiliário. Através da aplicação deste método, será testada a hipótese de que é possível obter uma planta de valores genéricos para a área rural, apesar da quantidade e variedade de informações que são necessárias na avaliação de um imóvel rural individualmente, quanto mais em uma avaliação em massa de um grande grupo de imóveis.

O método comparativo direto de dados de mercado além dos outros métodos, utilizados neste trabalho é definido pela NBR 14.653:1 (ABNT, 2001), norma técnica existente no Brasil e que determina quais são e em quais situações devem ser aplicados, contudo, estes não são utilizados por diversas razões, na avaliação dos imóveis rurais para tributação.

Tradicionalmente na avaliação individual de um imóvel rural, utiliza-se de variáveis como: capacidade de uso do solo, condições das estradas de acesso e distância à área urbana do município, dimensão (área), potencial de irrigação (área irrigada), pluviosidade (índice de precipitação), porcentagem de área com cobertura florística natural (mata nativa), número de açudes (ou represas, ou córregos), porcentagem de área que já foi aberta, etc.

Com as informações obtidas pelo cadastro será possível implantar, por exemplo, outros mecanismos de redução no valor do ITR, além dos que já existem, para aquelas propriedades que mantivessem áreas de preservação ambiental na forma de reserva legal, áreas de preservação permanente ou reserva particular de proteção natural, tornando-o um instrumento de política ambiental, por meio do qual o ITR funcionaria como um ressarcimento feito pela sociedade ao proprietário rural que preservasse o meio ambiente.

Em diversas pesquisas de avaliação em massa de imóveis urbanos, tem sido incorporadas às metodologias avaliatórias, técnicas que permitem identificar a influência que a dependência espacial exerce sobre a formação dos preços, bem como a influência que um imóvel exerce sobre os seus vizinhos (HORNBERG, 2009; TRIVELLONI, 2005). Entre estas técnicas está a Regressão Espacial para a geração de modelos que incorporam entre as suas variáveis a localização geográfica, a Geoestatística, que por interpolação via Krigagem permite identificar o preço dos imóveis em regiões não amostradas e assim obter a PVG e a regressão linear geograficamente ponderada – RGP (*Geographically Weighted Linear Regression* - GWR) que permite ajustar um modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto.

Através da aplicação destas técnicas estatísticas (regressão tradicional, regressão espacial, RGP e a Geoestatística para a superfície de valores) neste trabalho, queremos verificar qual técnica gera resultados mais confiáveis e precisos para a área rural do PROJIR no Norte Fluminense - RJ.

Considerando a atualidade do tema e os motivos expostos, este trabalho é primordial para o desenvolvimento de um dos instrumentos de política agrária no país, como é considerando o ITR. A aplicação de métodos de avaliação consagrados, para a determinação da base tributável do imposto, tornará o processo de cálculo mais transparente, principalmente nas questões de desapropriação, para construção de estradas, hidrelétricas, obras de infraestrutura, reforma agrária, etc.

Não constitui dúvida a necessidade de dar respostas adequadas para o tratamento de questões acima abordadas a curto prazo, assim como, adequar os instrumentos de gestão do espaço rural às suas características regionais.

Com o desenvolvimento deste trabalho pretende-se gerar conhecimentos que contribuirão para a solução dos problemas, relacionados à determinação de preços de imóveis rurais no Brasil, com possível aplicação da metodologia também nos países do Mercosul. Neste sentido, será abordado aqui um dos temas que envolvem a gestão territorial rural, a avaliação em massa para a determinação do valor, que tem múltiplas aplicações, incluindo a tributação territorial.

2 REVISÃO DE LITERATURA

2.1 Avaliação em Massa de Imóveis Rurais

2.1.1 Conceitos e métodos de avaliação

A norma brasileira que detalha as diretrizes e padrões específicos de procedimentos para a avaliação de imóveis rurais é a NBR 14.653:3 - Avaliação de bens Parte 3: Imóveis rurais (ABNT, 2004). De acordo com a norma, imóvel rural é o imóvel com vocação para exploração animal ou vegetal, qualquer que seja a sua localização.

O Estatuto da Terra - Lei nº 4.504/64 (BRASIL, 1964) define imóvel rural como o prédio rústico de área contínua, qualquer que seja a sua localização, que se destine ou possa se destinar à atividade agrária.

O valor a ser buscado na avaliação é o valor de mercado, ou valor venal, referente à venda, assim definido na NBR 14653:1 – Procedimentos gerais: “Valor de mercado: quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem, numa data de referência, dentro das condições do mercado vigente” (ABNT, 2001). A mesma norma classifica os seguintes métodos para a identificação do valor de um imóvel rural:

- a) Método Comparativo Direto de Dados de Mercado: Identifica o valor do bem por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra;
- b) Método da Capitalização da Renda: Identifica o valor do bem, com base na capitalização da renda líquida prevista, considerando-se cenários viáveis;
- c) Método Evolutivo: Identifica o valor do bem pelo somatório das parcelas componentes do mesmo;
- d) Método Involutivo: Identifica o valor de mercado do bem, alicerçado no seu aproveitamento eficiente, baseado em modelo de estudo de viabilidade técnico-econômica, mediante hipotético empreendimento compatível com as características do bem e com as condições do mercado no qual está inserido, considerando-se cenários viáveis para execução e comercialização do produto;
- e) Método Comparativo Direto de Custo: Identifica o custo do bem por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra;
- f) Método da Quantificação de Custo: Identifica o custo do bem ou de suas partes por meio de orçamentos sintéticos ou analíticos, a partir das quantidades de serviços e respectivos custos diretos e indiretos.

A norma recomenda a utilização do método comparativo de dados de mercado sempre que possível, mas em avaliações de imóveis rurais ocorre a conjugação dos métodos, pois são vários os componentes a serem avaliados, como o valor da terra, das benfeitorias reprodutivas e não-reprodutivas, das áreas de preservação ambiental, etc.

A avaliação em massa, também conhecida como Planta de Valores Genéricos - PVG, consiste em determinar os valores de todos os imóveis de uma região ou município. Uma importante aplicação das avaliações de imóveis é na tributação imobiliária. A correta estimação dos valores é fundamental para atingir a equidade (mesma relação para todos os imóveis entre o valor da avaliação e o valor de mercado) e para que ocorra justiça fiscal na tributação.

A avaliação em massa pode ser definida como uma avaliação sistemática de grandes

grupos de propriedades usando procedimentos padronizados, em vez de uma única propriedade. A avaliação precisa do valor de um conjunto predefinido de propriedades, ou uma propriedade particular, indiretamente, usando um modelo, para um determinado propósito prático, é o principal alvo dessas metodologias (KAUKO e D'AMATO, 2008).

As avaliações individuais e em massa diferem apenas em termos de escala, pois sua finalidade é a mesma: uma avaliação exata do valor de uma ou mais propriedades, respectivamente. Os métodos são essencialmente os mesmos, com algumas diferenças na análise de mercado e no controle de qualidade (McCUSKEY et al.,1997). Segundo estes autores, a avaliação em massa surgiu pela necessidade de uniformidade e coerência nas avaliações quando existe um grande número de imóveis para avaliar (geralmente centenas ou milhares).

Segundo a IAAO - *International Association of Assessing Officers* (IAAO, 2013a,b), são basicamente três os métodos de avaliação em massa utilizados nos Estados Unidos da América: sob a abordagem da renda que a propriedade produz; sob a abordagem do custo para substituir o imóvel por outro de igual utilidade, materiais atuais, design e padrões de construção idênticos e a abordagem da comparação de vendas que estima o valor de uma propriedade analisando estatisticamente os valores de venda de propriedades semelhantes.

Os métodos indicados pela norma americana têm os mesmos princípios que as normas brasileiras, ou seja, o Método da Capitalização da Renda, o Método da Quantificação do Custo e o Método Comparativo de Dados de Mercado, respectivamente.

Na Tabela 1 é possível verificar os diferentes critérios e métodos utilizados por países da Europa e América Latina para realizar a valoração de imóveis rurais.

Tabela 1. Métodos de avaliação imobiliária em alguns países

País	Critério	Método
Áustria	Capitalização da Renda	
República Checa	Capitalização da Renda	Comparação com preços de mercado; Valores de Renda; Análise de Regressão Múltipla
Itália	Capitalização da Renda	Comparação no mercado; Análise de regressão múltipla (com evidências de segmentação do território)
República Eslovaca	Capitalização da Renda	Auto declaração
Suécia	Valor de Mercado	Comparação no mercado; Capitalização; Custo de reprodução
Chipre	Isento	Comparação no mercado
Finlândia	Isento	Comparação no mercado; Custo de reprodução
Hungria	Capitalização da Renda	Comparação no mercado
Romênia	Mercado	Modelos matemáticos automatizados
Argentina: Província de Córdoba	Capitalização da Renda	Comparação no mercado
Colômbia	Capitalização da Renda	Comparação no mercado
Chile	Capitalização da Renda	Comparação no mercado
Guatemala	Valor de Mercado	Auto avaliação
México	Capitalização da Renda	Comparação no mercado
Nicarágua	Capitalização da Renda	Comparação no mercado

Fonte: Adaptado de Bueno et al. (2012)

Os métodos utilizados em alguns países, segundo Bueno et al. (2012) são: na Argentina não existe regra geral para a valoração cadastral, cada província tem o poder de gerar sua própria metodologia. Bolívia, Guatemala, República Eslovaca e República Checa utilizam a auto avaliação, que é o valor declarado pelo proprietário do imóvel. Em vários casos (Bélgica, França, Itália e algumas províncias da Argentina) é utilizado o método da renda como base para o estabelecimento do valor utilizado nos impostos. Alguns países, como a Romênia, Espanha, Chipre, Dinamarca e Suécia começaram a utilizar métodos automatizados para a avaliação em massa. Outra iniciativa foi a de estabelecer observações coletadas do mercado imobiliário, tais como Alemanha, Itália, Bélgica, Polônia, Áustria, Portugal, Hungria, Eslováquia, Eslovênia e Romênia.

Países como Chipre, Finlândia e Suécia não cobram impostos sobre terras agrícolas e florestais (no último caso). Em Malta não são tributados nenhum tipo de propriedade.

Nos Estados Unidos da América os solos de uso agrícola, são avaliados para fins tributários desde 1957, pelo valor de uso (*use value programs – UVA*). Ficou estabelecido em lei que os campos e pastos poderiam ser avaliados abaixo do valor de mercado, sempre e quando fossem “utilizados ativamente” para fins agrícolas. Os programas UVA trouxeram alguns benefícios como o retardamento da expansão urbana durante as últimas décadas. A redução na taxa de desenvolvimento do solo é um benefício ambiental, entretanto os programas públicos UVA, vêm com um custo social, pois para manter a arrecadação de impostos, as outras propriedades que não estavam inseridas no programa acabaram pagando mais impostos (ENGLAND, 2012; ANDERSON e ENGLAND, 2016).

A avaliação em massa é empregada para tributação, para desapropriações, ou para a identificação dos valores em outras funções da gestão urbana (GONZÁLES, 2002; GRIPP JR et al., 2006; HERRERA, 2001). Geralmente devem ser geradas estimativas para uma grande quantidade ou para todos os imóveis do cadastro municipal, o que limita a avaliação em massa ao uso dos atributos ou variáveis disponíveis na base de dados, em função de restrições de custo de coleta de informação (GONZÁLES, 2002).

Apesar da importância reconhecida de contar com PVGs atualizadas, para Trivelloni (2005), sua implementação apresenta importantes dificuldades técnicas, principalmente devido a problemas de metodologia aplicada, de disponibilidade de informações atualizadas e de acesso e domínio de ferramentas de análise e modelagem apropriadas.

Praticamente não são desenvolvidos estudos em avaliações em massa de imóveis rurais. Dos poucos existentes podemos citar Machado (2006), que usou superfície de tendência na elaboração de uma PVG para uma área rural. A autora observou melhores resultados com a aplicação da técnica de superfície de tendência, a partir da inclusão das variáveis de localização. Santos (2014) também aplica a avaliação em massa em imóveis rurais. Este autor avaliou a combinação de técnicas da econometria espacial e da Geoestatística na modelagem dos efeitos espaciais em mercados de terras rurais e geração de PVG, e propôs uma metodologia que seja aplicável à realidade dos municípios brasileiros com a utilização de *softwares* livres.

2.1.2 Variáveis formadoras do valor em imóveis rurais

A avaliação de imóveis rurais, seja uma avaliação individual ou em massa, se torna muito mais complexa que em imóveis urbanos, pois além da quantidade de variáveis que influenciam o valor dos imóveis rurais ser muito grande, a obtenção destas informações não é um trabalho simples.

Um grande número de variáveis deve ser utilizado para as avaliações individuais de

imóveis rurais, para que sejam representativas. Se o objetivo da avaliação é a venda da propriedade os trabalhos pesquisados desta área são unânimes em afirmar que valor do imóvel rural será identificado pelo somatório das suas parcelas, ou seja, o valor da terra, o valor de todas as benfeitorias, seja reprodutiva ou não-reprodutiva, o valor do potencial hídrico, a cobertura florística, entre outros.

Os componentes dos imóveis rurais, segundo Lima (2011), transformam-se em dezenas de variáveis, como:

- a) Terras: fertilidade natural, drenagem, pedregosidade, topografia, profundidade, permeabilidade, erosão e clima, que podem ou não serem condensadas na capacidade de uso, situação relativa (acessos, tipo de pavimento e distância), dimensões;
- b) Produções vegetais: cana de açúcar, laranja, soja, milho, algodão, manga, reflorestamentos, café, caju, etc.;
- c) Construções e instalações: estábulos, curral, galpões, casas, cercas, instalações, terreiros, etc.;
- d) Recursos naturais: disponibilidade de água para consumo, irrigação, preservação do recurso natural florestal, áreas de preservação permanente, etc.

As mesmas variáveis independentes são citadas por Arantes e Saldanha (2009) e também por Resener e Hochheim (2006): características físicas (área, classes de capacidade de uso, formato do imóvel, etc.), de situação (acesso, localização, distâncias a centros consumidores, etc.) e acrescentam as variáveis econômicas (oferta e transação, época, condições de pagamento). Para os autores as variáveis devem ser escolhidas com base em teorias existentes, conhecimentos adquiridos, senso comum e outros atributos que se revelem importantes no decorrer dos trabalhos.

As variáveis utilizadas para avaliação de imóveis rurais em alguns países da Europa e América Latina são: a) condições naturais: qualidade e uso do solo, elementos topográficos e formas do terreno, clima, recursos hídricos; b) condições econômicas de produção: sistemas de transporte, localização e distribuição das áreas agrícolas, distâncias aos centros das cidades consumidoras dos produtos, serviços e mercados, sistemas de irrigação (BUENO et al., 2012).

Para a realização de avaliações em massa de imóveis rurais, Machado (2006) utilizou as seguintes variáveis: área, localização, classificação das terras, nível de manejo tecnológico, distância ao perímetro urbano, acessibilidade, localização em coordenadas no sistema Universal Transversa de Mercator - UTM e benfeitorias para fins de avaliação da terra nua. Depois de testar todas as variáveis a autora substituiu a variável capacidade de uso do solo (recomendada pela NBR 14.653:3) pela classificação em terras de primeira, de segunda e de terceira, pois a classificação por capacidade de uso não foi significativa.

O valor de um imóvel rural está diretamente relacionado à sua capacidade de produção, ou seja, vale pelo que produz. Considerando que o objetivo da avaliação em massa é obter o valor dos imóveis para tributação e no caso das áreas rurais para obter o ITR as variáveis necessárias e que devem constar em um cadastro são as relacionadas ao valor da terra.

Atualmente para o cálculo do valor do ITR realizado pela SRF são utilizados: o VTN - valor da terra nua, a área total, a área tributável, o VTNt, a área efetivamente utilizada e o GU - grau de utilização. Ao analisar o que é utilizado hoje com as variáveis sugeridas pode-se perceber que o cadastro rural atualmente não fornece as informações necessárias para uma

avaliação em massa.

Nas avaliações de imóveis rurais, Machado (2006) recomenda como primeira providência a ser tomada pelo engenheiro avaliador a realização de um cadastro, que deverá conter: levantamento físico com qualificação das benfeitorias; classificação das terras quanto à aptidão agrícola; categoria de acessibilidade ao imóvel; nível de manejo tecnológico usado na exploração agrícola e a distância ao centro consumidor mais próximo.

Dois modelos de regressão foram analisados por Santos (2014), em um deles além das variáveis intrínsecas do imóvel foi utilizado variáveis de localização (X, Y) e no outro modelo somente variáveis relacionadas às características dos imóveis, tais como: natureza do evento (transação ou oferta), área total, área irrigada, limite confrontante com o rio, estado das benfeitorias e índice de produção vegetal.

Ao analisar o comportamento das variáveis que influenciam o valor da terra no Pantanal Sul-Matogrossense, Lourival et al. (2008), testaram a influência de diversas variáveis como o tamanho da propriedade, cobertura de vegetação (5 classes), uso atual do solo, percentual de área com soja, percentual de área com pasto, percentual de área com reserva, percentual de área não inundável, topografia (3 classes), percentual de vegetação nativa (madeira), município de localização, infraestrutura disponível, energia elétrica, água disponível, acessibilidade, título da terra, fertilidade do solo, tipo de solo, potencial turístico, distância da cidade. Seus resultados mostraram que o tamanho da amostra, a escala de análise e práticas de uso da terra, influenciam significativamente a capacidade dos modelos de regressão múltipla, em explicar as mudanças nos valores de terras.

Para quantificar o valor do serviço ambiental de áreas protegidas ambientalmente em imóveis rurais, Comarella (2013) mostrou que as externalidades positivas geradas foram muito significativas, quando se instituem a reserva legal e as áreas de preservação permanente. Este autor verificou que os mecanismos da engenharia de avaliação para expressar valor de mercado de bens ambientais, pode ser uma ferramenta para viabilizar os esforços conservacionistas sejam estes da iniciativa privada ou pública, pois evidencia o valor de mercado de forma objetiva.

Um modelo matemático para estimar o valor de imóveis rurais irrigáveis, foi obtido por Silva et al. (2009). O modelo se mostrou adequado para a estimativa do valor de imóveis irrigáveis com área inferior a 1.500 ha, com o propósito de avaliação por interesse social ou por outros motivos, sendo mais recomendado para avaliação de imóveis de forma massiva ou por motivos fiscais ou tributários.

Considerando as características físicas sugeridas pelos autores citados e que estão relacionadas ao valor da terra, as variáveis que devem constar no cadastro obtidas nas diversas etapas de levantamento cadastral são: área total, formato, dimensões, topografia, áreas de preservação permanente, áreas de reserva legal, hidrografia, classificação dos solos pela capacidade de uso, os acessos ao imóvel, tipo de pavimento e distância ao centro consumidor (ou ao centro da área urbana). Um dos sistemas existentes para classificação utilitária do potencial agrícola das terras, o da capacidade de uso, segundo Lepsch et al. (1991), exige mais alguns dados como: tipo de solo ou natureza do solo, declividade, erosão, drenagem e clima. Estes dados deverão ser interpretados e analisados em conjunto, para a determinação e separação das classes, subclasses e unidades de capacidade de uso das terras.

Todos estes dados devem constar em um cadastro técnico rural, para que seja possível a elaboração da PVG, o cálculo do valor venal para cada imóvel e seu correspondente valor de ITR. As outras variáveis como produção vegetal, as construções e instalações poderão constar em um cadastro técnico para compor o banco de dados. Entretanto para a elaboração da PVG e cálculo do valor do ITR estas variáveis não serão necessárias, uma vez que o imposto é

somente territorial.

2.1.3 Avaliação de desempenho em avaliações em massa

As avaliações em massa de imóveis devem cumprir algumas condições de desempenho para serem consideradas de boa qualidade. Estas condições evidentemente estão relacionadas com o grau de aderência que os valores calculados pelo modelo e os valores reais de mercado apresentam. Quanto mais próximos estejam os valores calculados pelo modelo dos valores observados no mercado melhor será a qualidade da avaliação. Geralmente as medidas de desempenho referem-se pelo menos a dois tipos de qualidade: medidas de desempenho global da avaliação e medidas relacionadas com a equidade e uniformidade (TRIVELLONI, 2005).

Segundo a *International Association of Assessing Officers* (IAAO, 2013a) o parâmetro mais usado como medida de desempenho global de uma avaliação é a mediana das razões de avaliação. Dada uma observação com valor de mercado (P_o) e valor calculado pelo modelo (P_c), chama-se razão de avaliação (R) ou *ratio* de avaliação ao quociente:

$$R = \frac{P_c}{P_o} \quad (1)$$

A mediana das razões é menos afetada por relações extremas do que as outras medidas de tendência central, e devido a estas propriedades, é a medida pela IAAO, para avaliar o nível geral da avaliação, determinando a necessidade ou não de uma reavaliação. Outro parâmetro que pode ser usado é a média ponderada das razões, que é obtida dividindo a soma dos valores calculados pela soma dos valores observados.

Em relação aos parâmetros de dispersão da avaliação, a medida mais comumente usada para avaliar a uniformidade de uma avaliação é o chamado Coeficiente de Dispersão ou COD. Segundo a IAAO, o COD é obtido da seguinte maneira:

- a) diminui-se de cada razão de avaliação a mediana de todas as razões;
- b) é tomado o valor absoluto das diferenças anteriores;
- c) calcula-se a média dos valores absolutos;
- d) divide-se a média anterior pela mediana das razões de avaliação;
- e) multiplica-se por 100 (IAAO, 2013a).

O COD é referido como uma medida de dispersão horizontal. Ele proporciona informação sobre a uniformidade da avaliação dos imóveis em toda a área de estudo. Como o COD é afetado sensivelmente por valores extremos das razões de avaliação, a amostra deve ser saneada dos elementos extremos. Segundo Davis (2001), quando esta medida começou a ser utilizada era calculada sobre o intervalo interquartil dos *ratios* da amostra (50% central dos dados), não sendo, portanto, necessário eliminar pontos atípicos. Posteriormente, foi utilizado um intervalo ampliado calculado multiplicando o intervalo interquartil por 1,5 ou por 3 para definir os valores limites de *ratios* a considerar na análise.

Outro índice especializado pode ser definido para medir a equidade vertical da avaliação. Ele é chamado de Diferencial Relativo ao Preço ou *Price Related Differential* - PRD e serve para detectar diferenças sistemáticas na forma que são avaliados os imóveis de alto e de baixo valor. Quando os imóveis de baixo valor são avaliados como uma

porcentagem maior do valor de mercado que os imóveis de alto valor, a avaliação é chamada regressiva. No caso contrário a avaliação é chamada progressiva. O PRD é calculado dividindo a média das razões de avaliação pela sua média ponderada (IAAO, 2013a).

Os valores recomendados para a mediana das razões de avaliação (IAAO, 2013a) estão no intervalo de 0,90 e 1,10. Quanto ao parâmetro PRD o intervalo recomendado é entre 0,98 e 1,03. Valores menores a 0,98 sugerem progressividade e valores maiores que 1,03 sugerem regressividade na avaliação. Tendo em vista garantir a equidade fiscal, o IAAO (2013a) estabelece a variabilidade máxima nas avaliações em massa para imóveis rurais, ou seja, o valor do COD deve ficar entre 5 a 20%, sendo aceitável o limite superior de até 25%.

Nas normas sobre avaliações brasileiras (ABNT, 2004), não há recomendações de limites similares para as métricas estatísticas, como nas normas americanas. Entretanto, De Cesare e Averbek (2010), por meio do Manual de apoio para Cadastro Técnico Municipal – CTM, recomendam como desempenho aceitável em avaliações em massa, o nível de avaliação com o percentual mínimo de 70%. Para uniformidade, o desvio máximo aceitável, de acordo com as diretrizes, é de 30%, porém a recomendação é que deve-se procurar permanentemente desvios menores.

As distorções de caráter avaliatório são responsáveis por gerar, além de outros problemas, a regressividade na cobrança do imposto na medida em que propriedades de alto valor são subavaliadas, em termos relativos, em comparação com avaliações de imóveis de baixo valor (DE CESARE, 2010). Por isso a importância da atualização cadastral, da avaliação em massa, além da avaliação de desempenho para verificar a qualidade dos valores utilizados no cálculo dos tributos.

A importância de realizar a avaliação da representatividade do modelo e a utilização de pontos de verificação é enfatizado por Marques et al. (2012). Pontos estes que não devem fazer parte do conjunto amostral, para dessa forma, analisar a capacidade do modelo em estimar valores para a variável em análise em locais aleatórios da área de estudo.

No Manual de apoio para Cadastro Técnico Municipal – CTM, De Cesare e Averbek (2010) também enfatiza a importância de realizar testes para avaliar o desempenho das avaliações, e que devem ser desenvolvidos com uma amostra de dados independente. Ou seja, uma amostra que não tenha sido empregada no desenvolvimento dos modelos de avaliação.

2.1.4 ITR: apuração e cálculo

O imposto que incide sobre a propriedade rural é o ITR e tem como fato gerador a propriedade, o domínio útil ou a posse de imóvel por natureza, como definido em lei, com localização fora da zona urbana do município. Para efeito do ITR, é considerado imóvel rural a área contínua, formada de uma ou mais parcelas de terras, localizada na zona rural do município (BRASIL, 1996). Considera-se área contínua a área total do prédio rústico, mesmo que fisicamente dividida por rua, estrada, rodovia, ferrovia ou por canal ou curso de água.

O imposto sobre a propriedade territorial rural, por força do que determina o artigo 153, inciso VI, da Constituição Federal - CF (BRASIL, 1988), combinado com o artigo 29, do Código Tributário Nacional - CTN (BRASIL, 1966), é de competência da União Federal.

Está sujeito à apuração do imposto todo imóvel rural que não se enquadre nas condições de imunidade ou isenção do ITR. Independentemente de ser isento, ou imune, o proprietário do imóvel rural, deve apresentar anualmente, em modelo aprovado pela SRF, a Declaração do Imposto sobre a Propriedade Territorial Rural - DITR correspondente a cada imóvel rural.

A Lei 9.393, de 19 de dezembro de 1996 (BRASIL, 1996) e suas alterações, que

regulamenta atualmente o imposto, determina que para a apuração anual do ITR, o contribuinte deve declarar os seguintes valores:

I – VTN - valor do imóvel, excluídos os valores relativos a:

- a) construções, instalações e benfeitorias;
- b) culturas permanentes e temporárias;
- c) pastagens cultivadas e melhoradas; e
- d) florestas plantadas.

II - Área tributável, a área total do imóvel, menos as áreas:

- a) de preservação permanente e de reserva legal;
- b) de interesse ecológico para a proteção dos ecossistemas;
- c) comprovadamente imprestáveis para qualquer exploração agrícola, pecuária, granjeira, aquícola ou florestal;
- d) sob regime de servidão florestal ou ambiental; e
- e) cobertas por florestas nativas, primárias ou secundárias em estágio médio ou avançado de regeneração.

III – VTN_t - valor da terra nua tributável, obtido pela multiplicação do VTN pelo quociente entre a área tributável e a área total;

$$VTN_t = VTN \frac{\text{Área}_T}{\text{Área}_T + \text{Área}_E} \quad (2)$$

IV - Área aproveitável, a que for passível de exploração agrícola, pecuária, granjeira, aquícola ou florestal, excluídas as áreas:

- a) ocupadas por benfeitorias úteis e necessárias; e
- b) de que tratam as áreas do item II.

V - Área efetivamente utilizada, a porção do imóvel que no ano anterior tenha:

- a) sido plantada com produtos vegetais;
- b) servido de pastagem, nativa ou plantada, observados índices de lotação por zona de pecuária;
- c) sido objeto de exploração extrativa, observados os índices de rendimento por produto e a legislação ambiental;
- d) servido para exploração de atividades granjeira e aquícola; e
- e) sido o objeto de implantação de projeto técnico.

VI - Grau de Utilização - GU, a relação percentual entre a área efetivamente utilizada e a área aproveitável.

$$GU = \frac{\text{Área}_U}{\text{Área}_A} \quad (3)$$

O valor do imposto será apurado aplicando-se sobre o Valor da Terra Nua Tributável a

alíquota correspondente, em função da área total do imóvel e o grau de utilização, conforme Tabela 2 (BRASIL, 1996). A alíquota do imposto varia de 0,03% até 20% em função da área do imóvel e do grau de sua utilização. Para imóvel com área de até 50 hectares, utilizada em mais de 80%, a alíquota é de apenas 0,03%. Essa alíquota aumenta na medida em que a área é maior, indo até a 0,45%, se o imóvel tem área superior a 5.000 hectares e mais de 80% da área utilizada. Cresce, também, a alíquota na medida em que diminui a proporção da área utilizada, em relação à área total do imóvel, de sorte que para um imóvel com área superior a 5.000 hectares, com até 30% utilizada, o imposto tem alíquota de 20%.

O contribuinte do ITR e, portanto, o responsável pelas informações, é o proprietário de imóvel rural, o titular de seu domínio útil ou o seu possuidor a qualquer título. Este processo de arrecadação, que parte da autodeclaração do contribuinte sobre o valor do imóvel é uma das explicações para seu fraco desempenho. Como a arrecadação é baixa e seu controle é caro (a União arrecada e é nos municípios que os imóveis estão), o ITR tem sido menosprezado pela União, com prejuízo também ao município, que atualmente recebe 50% do valor arrecadado em ITR em seu território (AVERBECK, 2003). Para o autor os investimentos em cadastro são tão importantes quanto na área urbana, em razão não só da arrecadação do ITR, mas com relação aos benefícios na questão ambiental, na reforma agrária e no planejamento agrícola.

Tabela 2. Alíquotas para apuração do ITR

ÁREA TOTAL DO IMÓVEL (em hectares)	GRAU DE UTILIZAÇÃO – GU (%)				
	ATÉ 30	MAIOR QUE 30 ATÉ 50	MAIOR QUE 50 ATÉ 65	MAIOR QUE 65 ATÉ 80	MAIOR QUE 80
Até 50	1,00	0,70	0,40	0,20	0,03
Maior que 50 até 200	2,00	1,40	0,80	0,40	0,07
Maior que 200 até 500	3,30	2,30	1,30	0,60	0,10
Maior que 500 até 1.000	4,70	3,30	1,90	0,85	0,15
Maior que 1.000 até 5.000	8,60	6,00	3,40	1,60	0,30
Acima de 5.000	20,00	12,00	6,40	3,00	0,45

A Lei 11.250, de 27/12/2005 (BRASIL, 2005), determinou a celebração de convênios entre a União e o Distrito Federal ou os Municípios que assim optarem, no intuito de delegar as atribuições de fiscalização, inclusive a de lançamento dos créditos tributários, e de cobrança do imposto, sempre observando a legislação federal de regência do ITR. É importante destacar que a respectiva opção não implica em redução do imposto ou qualquer outra forma de renúncia fiscal.

A partir desta lei, os municípios poderão optar em exercer as atividades de cobrança e fiscalização do tributo e permanecer com a totalidade do valor caso assumam essas funções administrativas. A partir da celebração destes convênios entre a SRF e os municípios, os próprios municípios podem realizar a apuração e a fiscalização do ITR, entretanto apesar deste convênio trazer vantagens aos municípios, aumentando a arrecadação pois ficam com 100% do imposto, após vários anos da entrada em vigor da legislação, apenas 37% (cerca de 2.109 até março de 2016) dos municípios aderiram ao mesmo. No Rio de Janeiro o índice é de 30% de municípios conveniados e mais especificamente na Região Norte Fluminense os municípios de Campos dos Goytacazes, São Francisco de Itabapoana, Magé e Quissamã

firmaram convênio com a SRF (situação em março de 2016). Um dos problemas é que os municípios enfrentam dificuldades para construir e manter atualizadas suas Plantas de Valores Genéricos, o que reflete em inexpressiva arrecadação.

Os municípios conveniados devem informar anualmente a SRF, o valor da terra nua por hectare para atualizar o Sistema de Preços de Terras – SIPT. Os proprietários de imóveis rurais declaram anualmente o valor dos seus imóveis que são comparados aos informados pelas prefeituras à SRF, sendo que se estiver em desacordo servirá de referência para o início de um processo de fiscalização por parte do órgão responsável.

Cabe salientar que a União está de certa forma, transferindo uma responsabilidade que é dela, aos municípios, sendo que pode ocorrer que os municípios não obtenham um efetivo incremento na captação de receita e ainda terão que criar mecanismos que tornem viável tal pretensão, aumentando seus custos com o desenvolvimento de sistemas e com a contratação e capacitação de novos fiscais.

Outra dificuldade é a fiscalização do valor da terra nua, que é a base de cálculo do imposto. A vastidão e heterogeneidade do território nacional, associada às deficiências da fiscalização, e a falta de profissionais qualificados para a elaboração e atualização das PVGs transformaram em prática corrente a subavaliação dos imóveis rurais para este fim.

Na avaliação dos imóveis rurais para tributação segundo Santos (2014), são utilizadas tabelas de preços por regiões. Nestas, o município é dividido em regiões e a cada uma é atribuído um valor específico, de acordo com o julgamento da comissão designada para elaboração, que normalmente conta com um técnico agrícola ou Engenheiro Agrônomo. Frequentemente tais valores são estabelecidos com base na experiência dos profissionais que atuam na região, e não são utilizadas técnicas de avaliação preconizadas pelas normas de avaliação de imóveis.

Acrescenta-se também o fato que o município, tem também a função de acompanhar o valor dos imóveis rurais, em razão da competência municipal sobre o ITBI. Para este imposto os municípios utilizam o valor do imóvel declarado em cartório, que normalmente são valores abaixo do valor de mercado, uma vez que o próprio contribuinte declara o valor negociado. Ou então são utilizadas tabelas de valores por região que não representam a realidade dos valores de mercado, pois além de serem elaboradas sem critérios técnicos geralmente estão desatualizadas.

Esta situação remete à importância de se aprofundar este tema, a busca de técnicas de avaliação em massa que estejam de acordo com as normas avaliatórias e que forneça aos municípios mecanismos de fiscalização e de arrecadação eficientes e que possam ser assimiladas com facilidade pelos mesmos.

2.2 Modelo Clássico de Regressão Linear – MCRL

A principal técnica utilizada nas avaliações em massa é a análise de regressão múltipla. Esta abordagem econométrica tradicional estendeu o uso da análise de regressão para determinar o valor, em função de vários atributos estruturais do imóvel (área total, número de dormitórios, padrão construtivo, entre outros), atributos de localização (acessibilidade e da vizinhança dos imóveis), bem como atributos relacionados às condições da transação (forma de pagamento e época da venda), os quais também podem afetar os preços praticados.

A maioria dos modelos de previsão dos preços da habitação tem utilizado nas análises a regressão múltipla (*Multiple Regression Analysis - MRA*) conforme Dantas (2014), Lai (2011), Kauko e d'Amato (2008), Lai (2006) e González et al. (2002).

O embasamento teórico vem dos modelos hedônicos de preços³, enquanto que a estimativa do valor é geralmente realizada através de análise de regressão, em uma visão econométrica. O analista deve estipular modelos com as hipóteses de relacionamento entre as variáveis, que devem ser testados segundo critérios estatísticos, verificando-se a validade destas hipóteses, ou seja, se os modelos são capazes de representar o segmento de mercado em questão. Para tanto, devem ser coletados dados de transações (evidências do mercado), analisando-se o ajuste dos modelos considerados a estes dados, dentro de um determinado grau de precisão. Os testes estatísticos permitem avaliar o próprio modelo e a importância individual das variáveis incluídas, indicando a qualidade geral do modelo formulado (GONZÁLES, 2002).

O modelo convencional assume o formato apresentado na equação 4 (NETER et al., 1990):

$$Y = \alpha_0 + \alpha_1.X_1 + \alpha_2.X_2 + \alpha_3.X_3 + \dots + \alpha_k.X_k + \varepsilon_\alpha = Y^h + \varepsilon_\alpha \quad (4)$$

Este formato é chamado de “modelo linear clássico”, no qual Y é a variável dependente ou explicada (geralmente o preço), X_1, \dots, X_k são as variáveis independentes ou explicativas (as características dos imóveis e da região), α_0 é o intercepto da equação, $\alpha_1, \dots, \alpha_k$ são os coeficientes parciais da regressão (preços hedônicos implícitos), ε_α é o termo de erro (desvio da estimativa) e Y^h é a estimativa para a variável dependente, calculada em função das variáveis explicativas incluídas (RAMANATHAN, 1998).

Os coeficientes α_i são estimados geralmente pelo Método dos Mínimos Quadrados - MMQ, que busca um conjunto de coeficientes que minimize o quadrado dos resíduos do modelo. O processo de análise de regressão exige que sejam atendidos alguns requisitos essenciais, chamados de pressupostos básicos, e ainda outras condições relacionadas, os quais precisam ser respeitados para que a análise seja válida, e as inferências (estimativas) possam ser realizadas com a equação determinada.

Os princípios a serem atendidos para garantia de validade dos modelos, e as consequências da violação destes pressupostos, são os seguintes, conforme Neter et al. (1990):

- a) o modelo é linear nos parâmetros: este pressuposto decorre da própria forma do modelo clássico (Equação 1) e a fuga, por não-linearidade da função, provoca tendências nos resíduos;
- b) os resíduos seguem a distribuição Normal: a suposição de normalidade dos resíduos simplifica a teoria de análise de regressão e é necessária para garantir a validade dos testes de hipóteses e a estimação de intervalos de confiança. Pequenas fugas não são importantes, pois os testes de hipóteses são baseados na distribuição t de Student, que não é muito sensível a estes desvios;
- c) os resíduos têm média nula: geralmente este pressuposto é garantido pela fixação conveniente do termo constante (α_0), mas deve ser verificado para evitar tendências nos resíduos;
- d) há homocedasticidade dos resíduos (a variância é constante): as consequências

³ A base teórica do método hedônico é a identificação de atributos ou características de um bem que tenham relacionamento com o valor da propriedade.

da heterocedasticidade são de que as estimativas dos parâmetros da regressão são ineficientes (ou seja, a variância não é mínima), as estimativas das variâncias são tendenciosas e os testes de hipóteses (t - Student, F - Snedecor) tendem a fornecer resultados incorretos;

- e) os resíduos são independentes entre si: ou seja, não há autocorrelação. Existindo correlação entre os resíduos, os estimadores de mínimos quadrados não são mais os melhores estimadores lineares não tendenciosos e os testes t e F indicam conclusões incorretas;
- f) não há colinearidade entre quaisquer variáveis independentes: a perfeita correlação entre duas ou mais variáveis, ou seja, uma é combinação linear de outra(s), implica a existência de diversos modelos com o mesmo grau de ajustamento, não sendo possível selecionar um dos modelos, o que impede a interpretação sobre os coeficientes. Na prática, é mais comum a existência de colinearidade em grau menor ($|r| < 1$), que provoca alteração nos coeficientes das variáveis afetadas, inclusive invertendo sinais;

O modelo deve ainda atender aos seguintes requisitos, decorrentes dos pressupostos e da forma de cálculo dos coeficientes (NETER et al., 1990):

- g) não existem observações espúrias: a existência de elementos claramente não adaptados ao modelo (chamados de outliers), provoca distorções nos coeficientes, quando estes são calculados pelo MMQ, pois um erro relativamente grande tem influência sensível nos coeficientes, mascarando os resultados;
- h) as variáveis importantes foram incluídas: o modelo especificado deve ser similar ao real, e a falta de variáveis importantes provoca tendências nos resíduos, por falta de explicação do fenômeno (variação da variável dependente), e
- i) a amostra de dados é suficientemente grande: ou seja, o número de observações é maior que o de coeficientes a serem estimados. Este requisito é necessário para que possam ser realizados os cálculos dos coeficientes.

Devido às peculiaridades do mercado imobiliário, e da dificuldade de obtenção de dados, há risco de ruptura de várias destas condições. A aplicação de regressão múltipla para avaliação de imóveis tem produzido muitas vezes sérios problemas, principalmente por causa da multicolinearidade nas variáveis independentes e a inclusão de amostras que são "outlier" e não-linearidade dos dados (DE CESARE, 1998; LAI, 2011). Outro problema com este método é que ele envolve o julgamento humano, para atribuir as transformações no modelo matemático e para relacionar o comportamento das variáveis (LAI, 2011).

A ruptura de algum dos pressupostos, para Gonzáles (2002), não impede a utilização desta técnica, apenas dificulta ou impede a generalização dos resultados. Se não estão presentes as condições básicas, a análise passa a ter um caráter determinístico, podendo ser vista como um ajustamento de curvas, e os resultados são restritos ao conjunto específico de dados considerado. Se todas as características do avaliando estiverem contidas dentro dos intervalos respectivos, não há prejuízos maiores, e também não impedem, em tese, o prosseguimento da análise. Porém, para avaliação de massa trata-se de um empecilho sério, em função da diversidade dos "avaliandos".

Para o autor, quase todas as condições podem ser razoavelmente garantidas, ou mesmo bem controladas, existindo técnicas estatísticas para corrigir eventuais problemas. Entretanto,

para a autocorrelação e para a definição da forma funcional as soluções paliativas não são suficientes para eliminar as dificuldades geradas.

A autocorrelação é o relacionamento entre elementos consecutivos. A forma mais comum é a correlação serial. Devido às características espaciais do mercado imobiliário, os modelos de regressão podem sofrer com um tipo de autocorrelação espacial, denominado geralmente de “autocorrelação espacial”, principalmente em análises baseadas em recortes espaciais ou no tempo (WYATT, 1996).

Como sempre referenciado nos textos de econometria espacial, a dependência espacial é, de certa forma, similar à dependência temporal da análise de séries temporais; porém a grande diferença é que a influência espacial se dá de forma recíproca (a unidade A influencia B e vice-versa), e não como nas séries temporais onde ocorre sem a reciprocidade ($t-1$ influencia t , mas t não influencia $t-1$).

Alguns processos espaciais, principalmente aqueles observados em aplicações ambientais, apresentam indexação no espaço e trazem como característica comum a continuidade, observando-se que seus valores variam de forma gradual numa determinada vizinhança (CAMARGO et al., 1999).

No caso do mercado imobiliário estes processos espaciais, principalmente quando se trata da avaliação em massa, estão presentes e causam a autocorrelação espacial. Para Gonzáles (2002) se a variável tem distribuição espacial, as variações nos dados ocorrem em todas as direções, o que dificulta a análise. A autocorrelação espacial é geral e está ligada a características peculiares dos dados, espacialmente distribuídos, que têm maior relação com os imóveis da vizinhança próxima do que com os mais distantes.

O tratamento espacial explícito não era comum, observa Anselin (1998), por questões metodológicas e operacionais. A questão metodológica envolve a dificuldade de reconhecer a natureza bi-dimensional da interação espacial (correlação espacial) e suas implicações na análise estatística. Já foi mostrado que ignorar este aspecto pode levar a estimativas tendenciosas ou ineficientes dos coeficientes ou inferências incorretas. Por outro lado, do ponto de vista operacional, a deficiência era devida à falta de *softwares* adequados, o que já não é mais problema, segundo o autor, devido à oferta de vários pacotes estatísticos com ferramentas espaciais e, principalmente, ao desenvolvimento dos Sistemas de Informação Geográfica - SIG.

O outro problema é a forma funcional do modelo matemático, ou seja, quais variáveis devem ser usadas, quais são importantes para que o modelo fique adequado e como devem ser aplicadas as transformações nas mesmas. Para Gonzáles (2002), não havendo indicações teóricas da forma de relacionamento das variáveis, a análise de uma função linear de regressão para verificar se ela é apropriada para os dados ou não geralmente é realizada por tentativas, através de gráficos de resíduos contra as variáveis dependentes ou independentes do modelo. Se houver uma forma definida nos resíduos, com tendências de crescimento ou curvaturas, pode ser que o modelo testado não seja o mais adequado, e deve-se tentar o ajuste de funções não lineares ou linearizar a função, por transformações nas variáveis, tais como logaritmos, inversas ou potências.

2.3 Estatística Espacial

A abordagem da estatística clássica, através da utilização da inferência estatística, utiliza nas amostras de imóveis técnicas bem estabelecidas, como média, variância, desvio padrão, testes de hipóteses e modelos lineares. Entretanto, estas técnicas não consideram adequadamente o espaço geográfico como variável relevante dos modelos estatísticos.

Quando a localização relativa entre os dados for uma variável de interesse, outros tipos de técnicas são necessárias para modelar estes fenômenos, as chamadas técnicas de estatística espacial ou análise espacial de dados geográficos.

A inclusão de efeitos espaciais em modelos de regressão, segundo Câmara et al. (2004), pode ser feita das diferentes formas:

- a) Modelos com efeitos espaciais globais: partem do princípio de que o processo espacial subjacente aos dados analisados é estacionário. Neste grupo estão o modelo espacial autoregressivo (“*Spatial Auto Regressive – SAR*”) e o modelo do erro espacial (“*Spatial Error Model*” - SEM ou “*Conditional Auto Regressive*” – CAR), e
- b) Modelos de regressão com efeitos espaciais locais: quando o processo espacial é não-estacionário, os coeficientes de regressão precisam refletir a heterogeneidade espacial. Para tanto, há duas grandes alternativas: modelar a tendência espacial de forma contínua, com parâmetros variantes no espaço (superfícies de tendência e regressão geograficamente ponderada - RGP); modelar a variação espacial de forma discreta, ao dividir o espaço em sub-regiões estacionárias, chamadas de regimes espaciais.

O objetivo da análise espacial é descrever os padrões existentes nos dados espaciais e estabelecer de forma quantitativa, os relacionamentos entre as diferentes variáveis. Na avaliação de imóveis o objetivo é mensurar as propriedades e os relacionamentos levando em conta a localização espacial, incorporando o espaço à análise que se deseja fazer.

O crescente interesse no desenvolvimento teórico e metodológico da econometria espacial surgiu para tentar resolver o problema da autocorrelação espacial dos dados, que as técnicas econométricas tradicionais falharam em resolver e também para utilizar de forma adequada a variável localização nas avaliações dos imóveis.

Existem muitas publicações e trabalhos desenvolvidos em avaliações imobiliárias, segundo Chica Olmo (2007), com a utilização do modelo de regressão hedônica. Características estruturais, características da vizinhança e acessibilidade são as variáveis explicativas utilizadas nas análises. Características estruturais são as características individuais do imóvel (idade, tamanho, número de banheiros, etc.), que podem ser atingidos pela localização da propriedade. As características da vizinhança e acessibilidade dependem diretamente da localização da propriedade.

Segundo o autor a localização é provavelmente a variável mais importante usada para explicar o preço do imóvel, pois o valor de mercado de um imóvel está diretamente relacionado com o valor de outros imóveis vizinhos.

Já a acessibilidade do imóvel está relacionada à distância a polos de valorização, como centros comerciais, shoppings, áreas de lazer, etc. ou de desvalorização. Neste caso a localização, medida em termos de distâncias a estes polos também vai influenciar no valor dos imóveis.

A estatística espacial, para Krempi (2004) traz resultados diferentes daqueles obtidos pela estatística clássica, sendo os primeiros geralmente mais robustos por incorporarem a dimensão espacial. Para sua análise são necessárias pelo menos as informações sobre a localização e os atributos, que são valores associados aos dados coletados em campo, e parte-se do pressuposto que os dados são espacialmente dependentes.

Nas avaliações de imóveis rurais assim como nos urbanos acontece a dependência espacial nos valores, pois já foi comprovado que os imóveis rurais mais próximos aos centros

consumidores de seus produtos ou mais próximos às vias de escoamento são mais valorizados que os mais distantes (LIMA, 2011; ARANTES e SALDANHA, 2009).

A localização influencia no valor dos imóveis rurais, também concluiu Machado (2006), pois à medida que aumenta a distância às rodovias principais, no município de Chapecó-SC o valor dos imóveis diminui, e o mesmo acontece com relação a proximidade da área urbana.

Há duas perspectivas na literatura que considerem o autocorrelação espacial no valor dos imóveis: econometria espacial e Geoestatística (CHICA OLMO, 1995; DUBIN, 2003).

O método de regressão espacial é o mais amplamente utilizado para obter modelos econométricos, enquanto os métodos da Krigagem são usados mais em Geoestatística.

No primeiro método, segundo Dubin (2003), a abordagem de matriz do peso, usa uma matriz de conectividade para modelar as relações espaciais. No segundo método, a abordagem Geoestatística, postula que a correlação entre observações é função da distância que separa suas localizações. Ambos os métodos foram utilizados com algum sucesso na literatura, no entanto, ainda há pouca justificativa teórica na escolha do modelo, e por isso é possível que o investigador estime um modelo mal especificado.

2.4 Econometria (regressão) Espacial

Em avaliações de imóveis, os métodos econométricos tradicionais têm sido aplicados extensamente, sendo o modelo linear geral de regressão múltipla o principal, desde que considere como variáveis explicativas todos os fatores influenciadores no valor dos imóveis. Já os métodos econométricos espaciais foram desenvolvidos mais recentemente e são aplicados numa proporção muito menor.

Na econometria espacial são usados os tratamentos das interações e estrutura espaciais de dados geograficamente distribuídos. Como dito anteriormente nas análises que utilizam dados onde o componente espacial é relevante, alguns dos pressupostos básicos podem não ser atendidos. Dividem-se, de modo geral, as peculiaridades desses dados em dois grandes grupos: heterogeneidade espacial e a autocorrelação espacial.

Esta metodologia ao considerar explicitamente os efeitos espaciais, apresenta quatro grandes áreas de interesse, segundo Anselin (1999): a) especificação formal da autocorrelação (ou dependência) e heterogeneidade espacial; b) estimação de modelos que incorporem os efeitos espaciais; c) elaboração de testes e diagnósticos da presença de efeitos espaciais; d) previsão (interpolação) espacial.

Em termos gerais, heterogeneidade espacial significa que o comportamento econômico não é estável através do espaço, e pode gerar padrões espaciais característicos sob a forma de agrupamentos ao longo do *set* de dados, e variar com a unidade (VIEIRA, 2009; TYSZLER, 2006). Dessa forma, os parâmetros variam e podem mudar a forma estrutural do modelo, podendo inclusive, gerar heterocedasticidade com possíveis erros de especificação.

A heterogeneidade geralmente ocorre junto com a autocorrelação espacial e as ferramentas econométricas tradicionais não são mais apropriadas neste caso (TRIVELLONI, 2005). O autor afirma que em amostragens transversais (sem variação no tempo) os fenômenos de heterogeneidade espacial e de autocorrelação espacial podem ser considerados equivalentes.

A autocorrelação espacial pode ser definida como a coincidência entre valores similares e similaridades locais. Assim, quando altos ou baixos valores para uma variável aleatória tendem a agrupar-se no espaço, temos o processo de autocorrelação espacial positiva. No entanto, pode acontecer também de as unidades espaciais serem circundadas por

unidades com valores significativamente distintos, ou seja, pode ocorrer que altos valores sejam acompanhados por vizinhos com valores baixos, ou vice-versa, processo que se denomina autocorrelação espacial negativa (ANSELIN e BERA, 1998).

Para Dantas et al. (2001) os efeitos espaciais compreendem a autocorrelação (ou dependência) e a heterogeneidade espacial. A heterogeneidade espacial pode ser considerada como uma instabilidade estrutural provocada por variâncias dos erros ou coeficientes do modelo não constantes; enquanto que a autocorrelação espacial, em dados de “*cross-section*” de unidades geográficas, ocorre quando uma observação associada a uma dada unidade geográfica i depende de outras observações em outras unidades geográficas $j \neq i$. Isto pode ser representado da seguinte forma:

$$y_i = f(y_j) \quad i, j=1, \dots, N \quad (5)$$

Nos modelos que consideram a componente espacial, a dependência espacial pode ser encontrada nas variáveis dependentes, no termo de erro (VIEIRA, 2009; TYSZLER, 2006), ou mesmo nas variáveis independentes (TYSZLER, 2006). Para Trivelloni (2005), a dependência espacial pode ser incorporada nos modelos clássicos de regressão de duas formas: como um regressor adicional na forma de uma variável dependente espacialmente defasada (Wy), ou uma estrutura espacialmente defasada no erro da regressão (Wu). O primeiro modelo é conhecido como Modelo de Defasagem Espacial ou da Variável Dependente Defasada e o segundo é o Modelo do Erro Espacial ou do Erro Espacialmente Correlacionado.

A dependência ou autocorrelação espacial pode ser incorporada aos modelos de regressão linear tradicional (ou clássico) da seguinte forma:

$$y = \rho W_1 y + X\beta + u \quad \text{com} \quad u = \lambda W_2 u + \varepsilon \quad (6)$$

Onde:

- y é um vetor N por 1 de variáveis dependentes;
- X é uma matriz N por k de variáveis explicativas (independentes);
- W_1 e W_2 são matrizes N por N de pesos espaciais, contendo relações de contiguidade de primeira ordem, padronizadas por linha ou funções de distância;
- ρ é o coeficiente de autocorrelação espacial do termo Wy ;
- λ é o coeficiente de autocorrelação espacial dos termos de erro (Wu).
- β é um vetor N por 1 de coeficientes estruturais, inclusive o termo de intercepto.

Este é o modelo geral pois o mesmo considera tanto as possibilidades de autocorrelação espacial: Modelo de Defasagem Espacial ou da Variável Dependente Defasada, quanto autocorrelação espacial dos termos de erros: Modelo do Erro Espacial ou do Erro Espacialmente Correlacionado (DANTAS et al., 2001).

2.4.1 Modelo de defasagem espacial ou *Spatial Auto Regressive* - SAR

Trata-se de um modelo de alcance global, o qual é caracterizado por hospedar a dependência espacial cujo alcance do transbordamento é global pela interferência do multiplicador espacial que faz com que um impacto sobre a variável dependente seja refletido

para todas as regiões da área de estudo (ALMEIDA, 2012).

Do modelo geral, da equação representada em (6), obtém-se o modelo regressivo-autorregressivo espacial ou “SAR *model*”, quando $W_2 = 0$, então:

$$y = \rho W_1 y + X\beta + \varepsilon \quad (7)$$

No qual W_1 é um vetor n por 1 de defasagens espaciais para a variável dependente, ρ é o coeficiente autorregressivo espacial. Este coeficiente é um indicador de autocorrelação positiva (negativa) caso ρ seja positivo (negativo). Em outros termos, um ρ positivo significa que um alto (baixo) valor de y nas regiões vizinhas aumenta (diminui) o valor de y na região i . Caso este não seja estatisticamente significativo, significa que o coeficiente é zero, indicando a inexistência de autocorrelação espacial.

2.4.2 Modelo do erro espacialmente correlacionado ou *Conditional Auto Regressive - CAR*

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

Da equação representada em (4), obtém-se o modelo do erro espacialmente correlacionado ou “CAR *model*”, quando $W_1 = 0$, então:

$$y = X\beta + u \quad \text{onde} \quad u = \lambda W_e + \varepsilon \quad (8)$$

No qual o coeficiente λ é o parâmetro do erro autorregressivo espacial que acompanha a defasagem W_e . Neste modelo, os erros associados com qualquer observação são uma média dos erros nas regiões vizinhas mais um componente de erro aleatório.

Estes dois modelos de estimação de regressão espacial não podem ser estimados diretamente via Mínimos Quadrados Ordinários (MQO), pois os estimadores assim gerados são viesados e inconsistentes, dado que existe viés de simultaneidade no modelo. O método de estimação de Máxima Verossimilhança (MV) é o mais indicado para estimar os dois modelos (DANTAS et al., 2001).

Os modelos de regressão espacial podem ser estimados utilizando vários métodos, segundo Anselin (1999), sendo os mais comumente usados o de máxima verossimilhança, contudo outros métodos também têm sido propostos, como os de variáveis instrumentais, mínimos quadrados espaciais, método dos momentos, método dos códigos, métodos Bayesianos, entre outros.

2.4.3 Matrizes de ponderação espacial

Nas amostragens transversais não há suficiente informação para estimar a matriz de covariâncias diretamente dos dados, por isso geralmente é necessário impor uma estrutura para essa covariância. Essa imposição se traduz na criação de um conceito de operador espacial de defasagem, que é uma média ponderada de uma variável aleatória em localizações vizinhas (ANSELIN, 1999). Para isso deve ser definido o conceito de vizinhança de cada observação. O operador espacial de defasagem opera por meio de uma matriz de pesos espaciais W que define a estrutura de covariância dos dados para uma determinada vizinhança

de cada observação. Por meio de pesos que representam a covariância entre dois elementos, o operador espacial calcula para cada observação a média ponderada da variável aleatória na vizinhança.

Para cada observação da variável aleatória y num ponto i o operador espacial Wy é obtido pela soma ponderada pelos pesos W_{ij} de todos os valores de y nos pontos j da vizinhança:

$$[W_y]_i = \sum_{j=1}^N W_{ij} W_j \quad (9)$$

Sendo que os pesos W_{ij} não nulos são os que definem a vizinhança de cada observação.

Por facilidade de interpretação, os pesos são padronizados por linha na matriz W de forma que a soma de pesos em cada linha é igual a 1. Consequentemente, a defasagem espacial da variável y no ponto i pode ser interpretada como sua média ponderada na vizinhança de i .

Os elementos que compõem a matriz W são não estocásticos e são externos ao modelo. Eles estão geralmente baseados nos conceitos de contiguidade dos dados ou na distância entre observações, baseados numa função decrescente com a distância entre observações. Esta estrutura de variação é definida pelo pesquisador, diferentemente de outros métodos, como os de estimação direta da covariância, que calculam a covariância a partir dos próprios dados. Esta falta de obrigatoriedade de uma justificativa teórica na seleção dos pesos espaciais e o fato de serem exógenos ao modelo e sem ambiguidades explica a popularidade dos pesos derivados geograficamente, segundo Anselin (1999).

As matrizes de pesos espaciais mais tradicionais são construídas a partir de atributos físicos e geográficos, como vizinhança, distâncias geográficas e tempo de deslocamento. De acordo com a distância geográfica, a construção da matriz W está baseada no ordenamento de um espaço cartesiano representado por latitudes e longitudes. Esse tipo de ordenamento permite calcular as distâncias de quaisquer pontos no espaço. Com relação à dependência espacial, pressupõe-se que o grau de dependência é negativamente relacionado com a distância (VIEIRA, 2009). Em outras palavras, assume-se que a intensidade da dependência espacial declina à medida que a distância entre as unidades aumenta.

Em avaliações de imóveis as matrizes de pesos espaciais são construídas em função da distância geográfica e/ou número de vizinhos mais próximos. A matriz baseada na contiguidade não se aplica, pois, as amostras estão separadas geograficamente, não possuindo limites em comum.

Almeida (2012) utilizou a mesma metodologia que Baumont (2004), que propõe o seguinte procedimento para a matriz de vizinhos mais próximos:

- a) Roda-se o modelo clássico de regressão linear por mínimos quadrados ordinários;
- b) Testam-se os resíduos para autocorrelação espacial por intermédio do valor da estatística I de Moran, usando L matrizes de k vizinhos mais próximos e variando L de $k=1$ a $k=20$; e
- c) Define-se k que tenha gerado o maior valor de I de Moran, significativo estatisticamente.

Outra matriz de ponderação espacial de natureza geográfica é aquela que define proximidade com base na distância inversa a partir dos pontos centroides dos polígonos. A

ideia intuitiva é quanto mais distantes duas amostras estão, menor será a dependência espacial entre elas.

A escolha da matriz de ponderação espacial é uma possível fonte de má especificação para o modelo. Caso se escolha uma matriz que capture pouca autocorrelação espacial, quando esta for utilizada como operador de defasagem espacial para a construção das variáveis W_y , W_x , W_e e W_ε , essas variáveis serão medidas com erro sistemático, ocasionando viés na estimação, tornando-a inconsistente (ALMEIDA, 2012).

Ainda, segundo Almeida (2012) é válido estimar os modelos econométricos usando mais de uma matriz de ponderação espacial e comparar os resultados, a fim de detectar discrepâncias significativas nos resultados, expondo a sensibilidade desses a escolha da matriz e checando assim sua robustez.

2.4.4 Testes de especificação

Quando se utiliza modelos econométricos espaciais o objetivo é remover a dependência espacial presente nos resíduos do modelo. Esta tarefa é realizada por meio da introdução de variáveis defasadas espacialmente do lado direito da regressão. Estas podem aparecer na forma de defasagem espacial na variável dependente (W_y); como defasagem nas variáveis explicativas (W_x); ou então como defasagem no termo do erro (W_e e/ou W_ε). Podem ocorrer de forma isolada ou em conjunto. A etapa de especificação consiste na identificação de qual tipo de variável defasada espacialmente deverá ser introduzida ao modelo econométrico espacial (VIEIRA, 2009; ALMEIDA, 2012).

Os principais testes usados para detectar autocorrelação espacial são: Moran I (erro); Multiplicador de Lagrange - LM (erro); LM (defasagem); LM robusto (defasagem); e Kelejian-Robinson (erro) (DANTAS et al., 2001; ANSELIN, 1999; VIEIRA, 2009).

Certamente o teste mais usado e citado para verificar a presença do componente espacial no erro é o I de Moran (TYSZLER, 2006). Algebricamente, essa estatística é dada por (ALMEIDA, 2012):

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} z_i z_j}{\sum_{i=1}^n z_i^2} \quad (10)$$

Ou matricialmente:

$$I = \frac{n}{S_0} \frac{\mathbf{z}' \mathbf{W} \mathbf{z}}{\mathbf{z}' \mathbf{z}} \quad (11)$$

Em que n é o número de regiões, z denota os valores da variável de interesse padronizada, W_z representa os valores médios da variável de interesse padronizada nos vizinhos, definidos segundo uma matriz de ponderação espacial W . Um elemento dessa matriz, referente à região i e à região j , é registrado como w_{ij} . S_0 é igual à operação $\sum \sum w_{ij}$, significando que todos os elementos da matriz de pesos espaciais W devem ser somados (ALMEIDA, 2012).

A estatística I tem sido usada como um teste para a presença de autocorrelação

especial residual, em linha com a estatística de Durbin-Watson para séries de tempo. Nesse caso, o teste I de Moran é aplicado sobre as estimativas dos erros de uma regressão feita por MQO, com a estatística I observada, comparada com uma distribuição aleatória aproximada por seus momentos, sob a hipótese nula de nenhuma correlação residual (VIEIRA, 2009).

O índice global de Moran (I) é uma das formas mais amplamente utilizadas de se medir a autocorrelação espacial. Essa estatística varia entre -1 e 1 , fornecendo uma medida geral da associação linear (espacial) entre os vetores no tempo t e a média ponderada dos valores da vizinhança, ou *lags* espaciais. Valores próximos de zero indicam inexistência de autocorrelação espacial significativa: quanto mais próximo do valor unitário, mais autocorrelacionados estará. Se o valor dessa estatística for positivo (negativo), a autocorrelação será positiva (negativa) (VIEIRA, 2009).

O diagrama de dispersão de Moran compara os valores normalizados do atributo em uma área com a média normalizada dos vizinhos, o que deriva um gráfico bidimensional dos valores normalizados pela média dos vizinhos. É uma forma de visualizar a dependência espacial e indicar os diferentes padrões espaciais presentes nos dados. O gráfico abaixo representa quatro quadrantes (AA, BB, BA e AB) como pode ser visualizado na Figura 1 que irão corresponder a quatro padrões de associação local espacial entre as regiões e seus vizinhos.

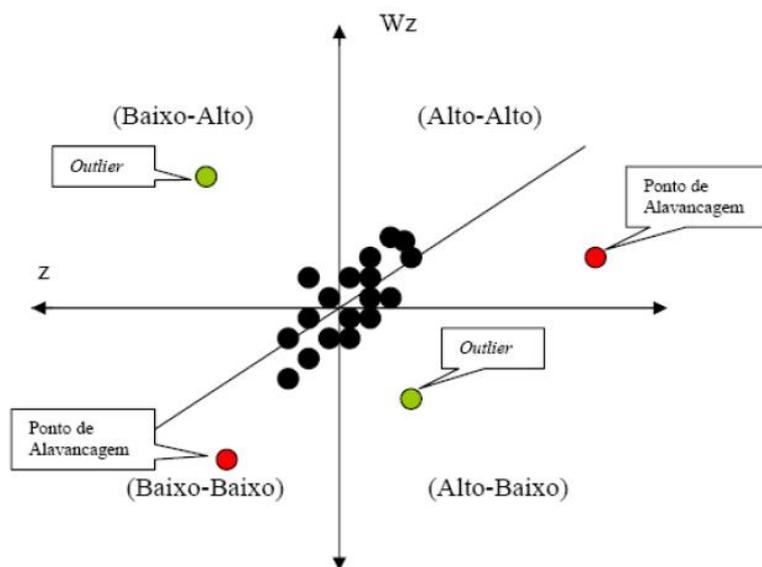


Figura 1. Diagrama de dispersão de Moran, outliers espaciais e pontos de alavancagem (CARVALHO, 2008).

O coeficiente I de Moran será a inclinação da curva de regressão de W_z (média dos vizinhos) contra Z (valores normalizados) e indicará o grau de ajustamento. O quadrante conhecido como alto-alto (AA), ou *high-high* – (HH), mostra regiões com altos valores para a variável, valores acima da média, assim como seus vizinhos. O quadrante geralmente chamado de baixo-baixo (BB) ou *low-low* – (LL), expressa localidades com baixos valores em relação aos atributos analisados, acompanhados por vizinhos que também apresentam baixos valores. O quadrante classificado como baixo-alto (BA) ou *low-high* – (LH), é constituído por baixos valores dos atributos na região estudada, cercada por vizinhos com altos valores. O último quadrante, é formado por regiões com altos valores para as variáveis

estudadas cercadas por regiões com baixos valores. Este é o quadrante alto-baixo (AB) ou *high-low* (HL). As regiões de *clusters* com valores similares ocorrem nos quadrantes AA e BB – e apresentam autocorrelação espacial positiva. As regiões identificadas pelos quadrantes BA e AB – apresentam, por sua vez, autocorrelação espacial negativa, ou seja, *clusters* com valores diferentes (VIEIRA, 2009).

A estatística I de Moran tem aspectos positivos e negativos evidenciados em alguns estudos. Tais estudos mostram que o teste de Moran apresenta um desempenho melhor que outros indicadores em termos de poder do teste para identificar correlações espaciais, porém não é capaz de diferenciar a correlação espacial nos erros da correlação espacial na variável dependente (TYSZLER, 2006).

O teste I de Moran possui poder contra outros problemas na regressão, além da autocorrelação espacial, como a má especificação do modelo, a heterocedasticidade e a ausência de normalidade nos resíduos. A significância do teste I pode ser devido não à presença de autocorrelação espacial, mas em virtude da presença desses outros problemas (ANSELIN, 2005).

Além da estatística I de Moran, aplicada aos resíduos de uma regressão linear, a presença de algum grau de dependência espacial pode ser verificada por meio de alguns testes específicos, entre eles, o teste de Wald, Razão de Verossimilhança (*Likelihood Ratio* – LR) e através de uma família de testes baseados no Multiplicador de Lagrange (*Lagrange Multiplier* – LM).

Os dois primeiros testes, Wald e LR, receberam mais atenção na literatura de econometria espacial até então. No entanto, a literatura focou os testes LM em diferentes formas, sendo este o teste indicado por Anselin (2003) para a escolha da especificação mais adequada.

Almeida (2012) descreve o procedimento conhecido como híbrido para a especificação de modelos espaciais proposto por Anselin (1996). Este procedimento objetiva discriminar entre o modelo de defasagem espacial autorregressivo SAR e o modelo de defasagem do erro SEM ou CAR, utilizando a combinação dos testes de ML tradicionais e os ML nas suas versões robustas.

O procedimento segue os seguintes passos:

- a) Estima-se o modelo clássico de regressão linear (MCRL) por MQO;
- b) Testa-se a hipótese de ausência de autocorrelação espacial devido a uma defasagem espacial ou a um erro espacial autorregressivo por meio das estatísticas ML (defasagem) e ML (erro);
- c) Caso ambos os testes não sejam significativos do ponto de vista estatístico, estima-se o modelo clássico como o modelo mais apropriado. Caso contrário, segue-se para o próximo passo;
- d) Se um deles for significativo, estima-se o modelo espacial indicado pela hipótese alternativa do teste. Por exemplo, se ML (defasagem) for significativo, deve-se, inicialmente, estimar o modelo SAR; e
- e) Caso ambos sejam significativos, estima-se o modelo apontado como o mais significativo pelas versões robustas desses testes ML robusto (defasagem) e ML robusto (erro). Por exemplo, se ML robusto (defasagem) for significativo, escolhe-se o modelo SAR como o mais apropriado. Caso o ML robusto (erro) seja o significativo, adota-se o modelo SEM ou CAR como o mais apropriado.

2.5 Geoestatística

2.5.1 Introdução

De uma forma geral, quando se utiliza dados em uma análise, sejam ambientais, socioeconômicos ou na avaliação de imóveis, estes dados estão disponíveis na forma de amostras pontuais, e para utilizá-los de forma efetiva em um ambiente de geoprocessamento, são necessários procedimentos de interpolação para obter as informações em áreas não amostradas.

Para gerar superfícies que aproximem o fenômeno estudado de forma realista, é necessário modelar sua variabilidade espacial. Os modelos que objetivam gerar superfícies a partir de procedimentos de interpolação, de forma geral, representam a variável em estudo como uma combinação da variabilidade em larga e pequena escala. Existem três grandes abordagens para a modelagem desta variabilidade espacial (CAMARGO et al., 2004):

- a) Modelos determinísticos de efeitos locais: cada ponto da superfície é estimado apenas a partir da interpolação das amostras mais próximas, utilizando funções como inverso do quadrado da distância. A suposição implícita é que predominam os efeitos puramente locais. Neste caso, não é feita qualquer hipótese estatística sobre a variabilidade espacial. Estes interpoladores são: por vizinho mais próximo; por média simples; por média ponderada; Kernel estimators, ou estimadores de densidade não-paramétricos.
- b) Modelos determinísticos de efeitos globais: a suposição implícita é que predominam os efeitos em larga escala e que os efeitos locais não são relevantes. É o caso das superfícies de tendência. A superfície é aproximada por um ajuste polinomial aos dados, através de um processo de regressão múltipla entre os valores do atributo e as localizações geográficas.
- c) Modelos estatísticos de efeitos locais e globais (Krigagem): cada ponto da superfície é estimado apenas a partir da interpolação das amostras mais próximas, utilizando um estimador estatístico. Esses procedimentos requerem que as variabilidades locais e globais sejam modeladas através de modelos de correlação espacial.

A Geoestatística envolve análise e inferência de fenômeno espacial e ou temporal, e teve sua origem na área de mineração com Daniel Krige, pioneiro no uso de médias móveis ponderadas (KRIGE, 1951). Krige utilizou esta ferramenta estatística para evitar a superestimação sistemática de reservas de mineração.

A Geoestatística é justamente um nome associado com uma classe de técnicas usada para analisar e inferir valores de uma variável distribuída no espaço e/ou no tempo. Tais valores são implicitamente assumidos ser correlacionados com outros, e o estudo de tal correlação é denominada de análise estrutural ou modelagem do variograma (CAMARGO et al., 2004). Resumidamente, os passos num estudo empregando técnicas da Geoestatística inclui: (a) análise exploratória dos dados, (b) análise estrutural (cálculo e modelagem do variograma) e (c) realização de inferências (Krigagem ou Simulação).

Krige (1951), trabalhando com dados de concentração de ouro, concluiu que somente a informação dada pela variância seria insuficiente para explicar o fenômeno em estudo. Para tal, seria necessário levar em consideração a distância entre as observações. A partir daí surge o conceito da Geoestatística, que leva em consideração a localização geográfica e a dependência espacial.

Para determinar qual das duas técnicas (estatística clássica ou Geoestatística) deve-se utilizar, faz-se uso do semivariograma, que expressa a relação espacial entre as amostras. Havendo dependência espacial pode-se utilizar a Geoestatística para modelar estes efeitos espaciais, por outro lado, se não for detectada a dependência espacial utiliza-se a estatística clássica (TRIVELLONI, 2005).

Verificada a dependência espacial pode-se estimar valores do fenômeno em estudo para os locais não amostrados, dentro da área em estudo, sem tendência e com variância mínima, por meio do método conhecido como interpolação por Krigagem (TRIVELLONI, 2005; CAMARGO et al., 1999). A Krigagem possibilita a construção de mapas de isolinhas, ficando definidas as classes dentro da área de estudo.

Inicialmente a Geoestatística era aplicada apenas para situações de geologia mineira, mas posteriormente estendeu-se para outros campos como climatologia, econometria espacial, geologia ambiental, entre outros. Atualmente, quase todas as versões de *softwares* desenvolvidos para manipulação de mapas, os Sistemas de Informações Geográficas apresentam módulos para a aplicação de Geoestatística.

Baseado nas observações de Krige, Matheron (1971) desenvolveu a teoria das variáveis regionalizadas. Uma variável regionalizada é uma função numérica com distribuição espacial, que varia de um ponto a outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função matemática simples.

Quando existe uma dependência entre uma variável aleatória e a sua localização espacial a variável pode ser chamada de regionalizada, segundo Chica Olmo (1994). A Teoria das Variáveis Regionalizadas (TVR) dedica-se à análise e resolução dos problemas de inferência espacial de variáveis que dependem da localização espacial onde são observadas.

2.5.2 Teoria das variáveis regionalizadas

Na teoria das variáveis regionalizadas, $Z(x)$ pode ser definida como uma variável aleatória que assume diferentes valores Z em função da posição x dentro de uma certa região S , e representa pares de coordenadas (x_i, y_i) (Figura 2). O ponto de referência para o sistema de coordenadas é arbitrário e fixado a critério do interessado. O conjunto de variáveis $Z(x)$ medidas em toda a área S pode ser considerada uma função aleatória $Z(x)$ uma vez que, são variáveis aleatórias, regionalizadas e assume-se que a dependência entre elas é especificada por algum mecanismo probabilístico (GENÚ, 2004).

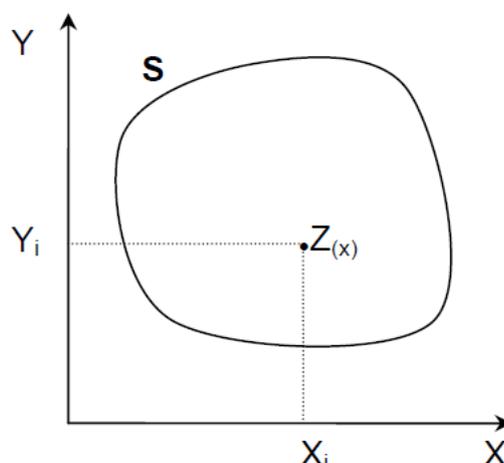


Figura 2. Variável aleatória regionalizada $Z(x)$ (GENÚ, 2004).

Segundo Camargo et al. (2004), a teoria das variáveis regionalizadas pressupõe que a variação de uma variável pode ser expressa pela soma de três componentes: a) uma componente estrutural, associada a um valor médio constante ou a uma tendência constante; b) uma componente aleatória, espacialmente correlacionada; e c) um ruído aleatório ou erro residual.

Se x representa uma posição em uma, duas ou três dimensões, então o valor da variável Z , em x , é dada por:

$$Z(x) = \mu(x) + \varepsilon'(x) + \varepsilon'' \quad (12)$$

Onde $\mu(x)$ é uma função determinística que descreve a componente estrutural de Z em x ; $\varepsilon'(x)$ é um termo estocástico, que varia localmente e depende espacialmente de $\mu(x)$; e ε'' é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

As Figuras 3(a) e 3(b) ilustram as três componentes principais da variação espacial. A Figura 3(a) apresenta uma componente determinística que varia abruptamente, enquanto a componente determinística na Figura 3(b) apresenta uma tendência constante.

Uma Variável Regionalizada (VR) é uma variável aleatória que se distribui no espaço e/ou no tempo e que apresenta autocorrelação espacial. As VR apresentam então uma dupla condição: aleatoriedade e estrutura espacial, segundo Trivelloni (2005). A utilização da TVR para realizar inferência estatística requer a introdução de algumas hipóteses para os dados, que são as de estacionariedade da variável aleatória.

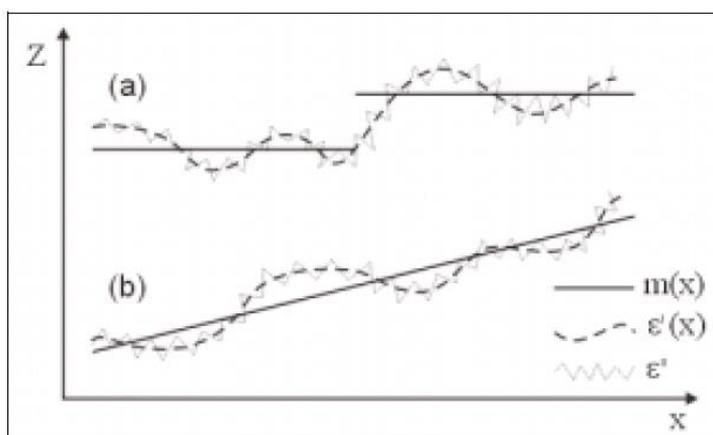


Figura 3(a) e (b). Componentes de uma variável regionalizada (CAMARGO et al., 2004).

Esta hipótese dá origem aos interpoladores de Krigagem ordinária. Na hipótese da Krigagem ordinária, $\mu(x)$ é constante e denotada por m . Deste modo, o valor esperado da função aleatória Z nas posições x e $x + h$ são iguais a m . Isto implica que o valor esperado da diferença entre os valores observados em x e $x + h$, separados por um vetor de distância h , é nulo:

$$E[Z(x) - Z(x+h)] = 0 \quad (13)$$

Admite-se também que o fenômeno considerado seja estacionário de segunda ordem,

isto é, a covariância entre dois pares quaisquer $Z(x)$ e $Z(x + h)$, separados por um vetor distância h , existe e depende somente de h . Então:

$$C(h) = COV[Z(x), Z(x+h)] = E[Z(x), Z(x+h)] - m^2 \quad (14)$$

Adicionalmente, a estacionariedade da covariância implica na estacionariedade da variância:

$$Var(Z(x)) = E[Z(x) - m]^2 = E[Z^2(x)] - 2E[Z(x)].m + m^2 \quad (15)$$

Ou ainda:

$$Var(Z(x)) = E[Z^2(x)] - 2m.m + m^2 = E[Z^2(x)] - m^2 = C(0) \quad (16)$$

Deste modo, verifica-se que as hipóteses de média constante e estacionariedade da covariância implicam que a determinação da função $C(h)$ é suficiente para caracterizar a variável regionalizada. Isto quer dizer que, com base na equação 12, a função $C(h)$ permite caracterizar o termo estocástico $\varepsilon'(x)$. Para determinar $C(h)$, utiliza-se uma função auxiliar, chamada de função variograma $2\gamma(h)$, definida por:

$$2\gamma(h) = E[Z(x) - Z(x+h)]^2 \quad (17)$$

Que pode ser desenvolvida em:

$$2\gamma(h) = E[Z^2(x) - 2Z(x).Z(x+h) - Z^2(x+h)] \quad (18)$$

Ou ainda:

$$2\gamma(h) = E[Z^2(x)] - 2E[Z(x).Z(x+h)] - E[Z^2(x+h)] \quad (19)$$

Da equação (14), obtém-se:

$$E[Z^2(x)] = E[Z^2(x+h)] = C(0) + m^2 \quad (20)$$

e da equação (13) obtém-se:

$$E[Z(x).Z(x+h)] = C(h) + m^2 \quad (21)$$

Substituindo as equações (20) e (21) na equação (19), obtém-se:

$$2\gamma(h) = 2C(0) - 2C(h) \quad \text{ou} \quad \gamma(h) = C(0) - C(h) \quad (22)$$

Em que $\gamma(h)$ representa o semivariograma, que é metade do variograma. A relação na

equação 22 indica que sob a hipótese de estacionariedade de 2ª ordem, a covariância e o semivariograma são formas alternativas de caracterizar a autocorrelação dos pares $Z(x)$ e $Z(x+h)$ separados pelo vetor h .

2.5.3 Determinação experimental do semivariograma

O semivariograma é uma ferramenta básica de suporte às técnicas de Krigagem, pois permite representar quantitativamente a variação de um fenômeno regionalizado no espaço. O semivariograma pode ser calculado experimentalmente, considerando o esquema de amostragem em duas dimensões mostrado na Figura 4, onde $z(x)$ denota o valor de uma posição cujos componentes são (x_1, y_1) , e $z(x+h)$ o valor da amostra numa posição cujos componentes são (x_2, y_2) , sendo h um vetor distância (módulo e direção) que separa os pontos (CAMARGO et al., 2004).

A determinação experimental do semivariograma, para cada valor de h , considera todos os pares de amostras $z(x)$ e $z(x+h)$, separadas pelo vetor distância h , a partir da equação:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \quad (23)$$

Onde:

$\gamma(h)$ é o semivariograma estimado e $N(h)$ é o número de pares de valores medidos, $z(x)$ e $z(x+h)$, separados pelo vetor h . Esta fórmula, entretanto, não é robusta. Podem existir situações em que variabilidade local não é constante e se modifica ao longo da área de estudo (heterocedasticidade).

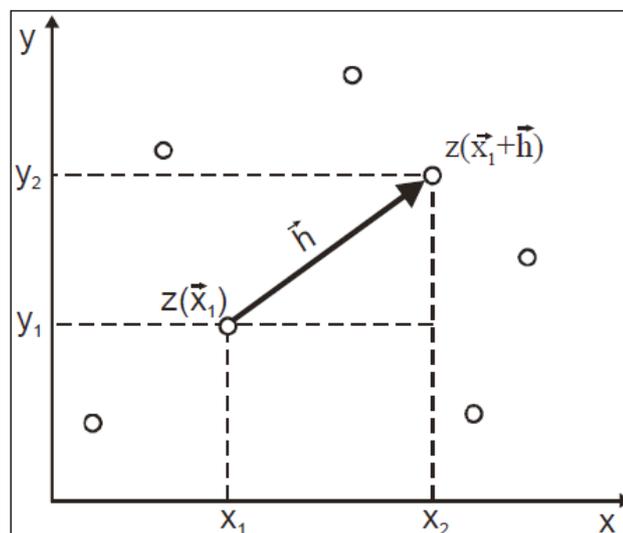


Figura 4. Amostragem em duas dimensões (CAMARGO et al., 2004).

Na prática, pode-se fazer a hipótese adicional de que o fenômeno é isotrópico (com comportamento igual em todas as direções). Neste caso, a determinação experimental do semivariograma depende apenas da distância entre as amostras e não da direção relativa entre elas. O tratamento da anisotropia (caso em que a estrutura espacial do fenômeno varia conforme a direção) é apresentado mais adiante.

As hipóteses de estacionariedade e média constante levam a postular um

comportamento idealizado para o semivariograma experimental, mostrado na Figura 5. Espera-se que observações mais próximas geograficamente tenham um comportamento mais semelhante entre si do que aquelas separadas por maiores distâncias. Assim, o valor absoluto da diferença entre duas amostras $z(x)$ e $z(x+h)$ deveria crescer à medida que aumenta a distância entre elas, até um valor na qual os efeitos locais não teriam mais influência.

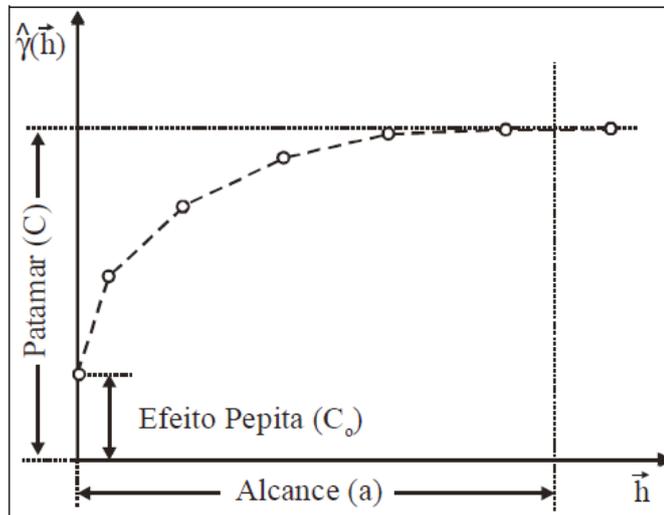


Figura 5. Parâmetros do variograma (CAMARGO et al., 2004).

Os parâmetros do semivariograma podem ser observados na Figura 5 e são assim definidos:

- a) Alcance (a): distância dentro da qual as amostras apresentam-se correlacionadas espacialmente.
- b) Patamar (C): é o valor do semivariograma correspondente ao seu alcance (a). Deste ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, porque a variância da diferença entre pares de amostras ($\text{Var}[Z(x) - Z(x+h)]$) torna-se aproximadamente constante.
- c) Efeito Pepita (C_0): idealmente, $\gamma(0)=0$. Entretanto, na prática, à medida que h tende para zero, $\gamma(h)$ se aproxima de um valor positivo chamado Efeito Pepita (C_0), que revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre as amostras. O efeito pepita é o valor da semivariância para a distância zero e representa a componente da variabilidade espacial que não pode ser relacionado com uma causa específica (variabilidade ao acaso). Parte desta descontinuidade pode ser também devida a erros de medição, sendo impossível quantificar se a maior contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem.

Quanto ao grau de dependência espacial da variável em estudo, Guimarães (2004) classifica como (sendo C_1 o valor da contribuição, ou seja, $C_0 + C_1 = C$):

- i) variável com forte dependência espacial – se o efeito pepita for menor ou igual a 25% do patamar $\left(\frac{C_0}{C_0 + C_1} < 0,25 \right)$;

- ii) variável com moderada dependência espacial – se o efeito pepita representar entre 25% e 75% do patamar $\left(0,25 \leq \frac{C_0}{C_0 + C_1} \leq 0,75 \right)$;
- iii) variável com fraca dependência espacial – se a relação entre efeito pepita e patamar estiver entre 75% e 100% $\left(0,75 < \frac{C_0}{C_0 + C_1} < 1,00 \right)$
- iv) variável independente espacialmente – se a relação entre efeito pepita e patamar for igual a 100%, neste caso temos o semivariograma com efeito pepita puro $\left(\frac{C_0}{C_0 + C_1} = 1,00 \right)$.

2.5.4 Modelos teóricos de semivariograma

O gráfico do semivariograma experimental, $\gamma(h)$, calculado através da Equação 21, é formado por uma série de valores, conforme ilustra a Figura 4, sobre os quais se objetiva ajustar uma função. É importante que o modelo ajustado represente a tendência de $\gamma(h)$ em relação a h . Deste modo, as estimativas obtidas partir da Krigagem serão mais exatas e, portanto mais confiáveis.

O procedimento de ajuste não é direto e automático, como no caso de uma regressão, por exemplo, mas sim iterativo, pois nesse processo o intérprete faz um primeiro ajuste e verifica a adequação do modelo teórico. Dependendo do ajuste obtido, pode ou não redefinir o modelo, até obter um que seja considerado satisfatório.

Os modelos aqui apresentados são considerados modelos básicos, denominados modelos isotrópicos. Estão divididos em dois tipos: modelos com patamar e modelos sem patamar. Modelos do primeiro tipo são referenciados na Geoestatística como modelos transitivos. Alguns dos modelos transitivos atingem o patamar (C) assintoticamente. Para tais modelos, o alcance (a) é arbitrariamente definido como a distância correspondente a 95% do patamar. Modelos do segundo tipo não atingem o patamar, e continuam aumentando enquanto a distância aumenta. Tais modelos são utilizados para modelar fenômenos que possuem capacidade infinita de dispersão. Os modelos transitivos mais utilizados são: modelo esférico (Sph), modelo exponencial (Exp) e modelo Gaussiano (Gau). Estes modelos estão apresentados na Figura 6 com o mesmo alcance (a).

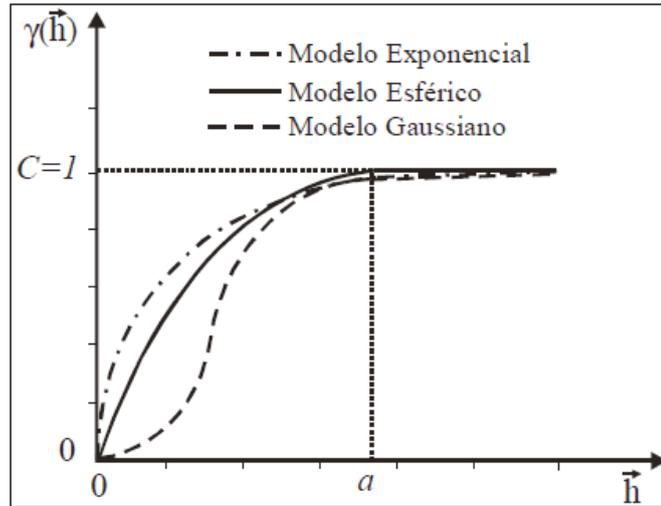


Figura 6. Representação gráfica de modelos transitivos normalizados (CAMARGO et al., 2004).

Em resumo, os semivariogramas dos modelos transitivos básicos são assim definidos:

Modelo Esférico: o modelo esférico é um dos modelos mais utilizados e está representado na Figura 6. A equação normalizada deste modelo é:

$$Sph(|h|) = \begin{cases} 0, & |h| = 0 \\ \left[\frac{3}{2} \left(\frac{|h|}{a} \right) - \frac{1}{2} \left(\frac{|h|}{a} \right)^3 \right], & 0 < |h| \leq a \\ 1, & |h| > a \end{cases} \quad (24)$$

Modelo Exponencial: um outro modelo bastante utilizado é o modelo exponencial, o qual é apresentado na Figura 6. A equação normalizada deste modelo é:

$$Exp(|h|) = \begin{cases} 0, & |h| = 0 \\ 1 - \exp\left(-\frac{|h|}{a}\right), & |h| \neq 0 \end{cases} \quad (25)$$

Este modelo atinge o patamar assintoticamente, com o alcance prático definido como a distância na qual o valor do modelo é 95% do patamar.

Modelo Gaussiano: o modelo Gaussiano é um modelo transitivo, muitas vezes usado para modelar fenômenos extremamente contínuos. Sua formulação é dada por:

$$Gau(|h|) = \begin{cases} 0, & |h| = 0 \\ 1 - \exp\left(-\frac{|h|^2}{a}\right), & |h| \neq 0 \end{cases} \quad (26)$$

Do mesmo modo que no modelo exponencial, o modelo Gaussiano atinge o patamar assintoticamente e o parâmetro a é definido como o alcance prático ou distância na qual o

valor do modelo é 95% do patamar. O que caracteriza este modelo é seu comportamento parabólico próximo à origem.

Até este ponto foram apresentados os principais modelos básicos normalizados, os quais são utilizados para ajustar o semivariograma experimental. Na prática, os semivariogramas experimentais possuem valores de efeito pepita (C_0) maior que zero e valores de patamar (C) maiores que a unidade, conforme ilustrado na Figura 7 e as equações 24, 25 e 26 sofrem modificações quando ocorre o efeito pepita (CAMARGO et al., 2004).

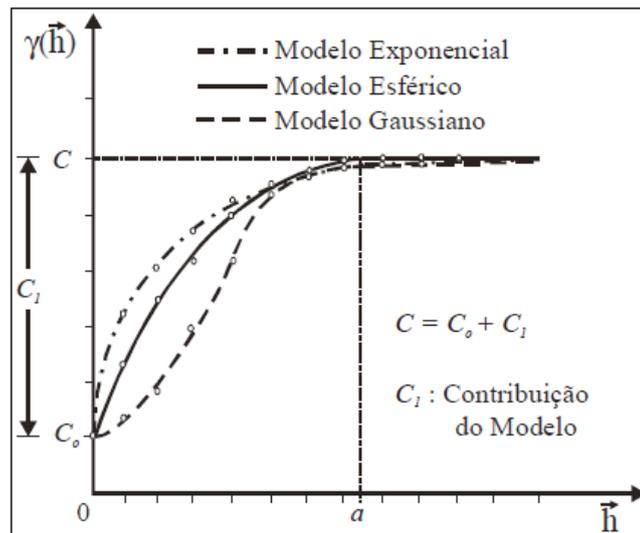


Figura 7. Representação gráfica de semivariogramas experimentais e modelos teóricos com efeito pepita (CAMARGO et al., 2004).

Segundo o autor, existem determinados fenômenos em que são necessários modelos mais complexos de semivariograma para explicar suas variações espaciais, tais modelos são combinações de modelos simples, denominados aninhados. Em muitos casos, os modelos aninhados são necessários para explicar a variação de fenômenos decorrentes da combinação de fatores independentes de formação, por exemplo, um modelo aninhado útil em estudos de mineração e pesquisa de solo é o duplo esférico.

2.5.5 Anisotropia

A anisotropia é uma característica muito frequente nos elementos da natureza, e acontece quando a variabilidade ou distribuição espacial dos elementos ocorre mais intensamente numa direção e menos intensamente em outra direção. Para resolver problemas com a anisotropia, é importante que o modelo proposto represente bem a variabilidade espacial da propriedade em estudo. Procedimentos determinísticos para este fim são limitados, porque não consideram a estrutura de autocorrelação espacial bem como a anisotropia presente. Modelos mais adequados para este objetivo vêm sendo propostos e são englobados pela Geoestatística.

Quando os semivariogramas experimentais direcionais apresentam diferenças acentuadas, a distribuição é denominada anisotrópica. Se a anisotropia é observada e é refletida pelo mesmo Patamar (C) com diferentes Alcances (a) do mesmo modelo, então ela é denominada Geométrica, conforme ilustra a Figura 8. Existe ainda um outro tipo de anisotropia em que os semivariogramas experimentais direcionais apresentam os mesmos Alcances (a) e diferentes Patamares (C). Neste caso, a anisotropia é denominada zonal. Como

a isotropia, a anisotropia zonal também é pouco presente nas variáveis ambientais. O mais comum é encontrar combinações da anisotropia Zonal e Geométrica, denominada anisotropia Combinada, conforme Figura 8. Na Figura 8, a_1 e a_2 estão relacionados às direções de menor e maior continuidade espacial da variável, respectivamente.

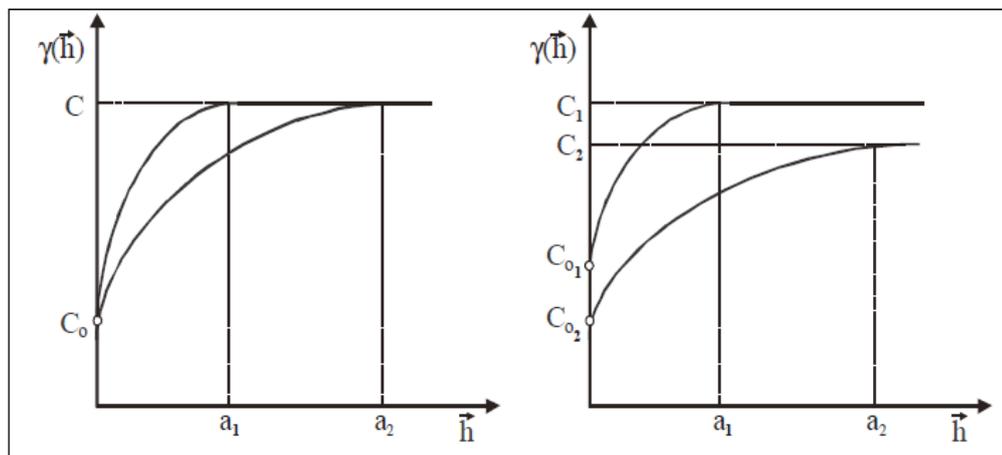


Figura 8. À esquerda Anisotropia Geométrica e à direita Anisotropia Combinada (CAMARGO et al., 2004).

Existem várias formas de detectar a anisotropia, por exemplo, calculando-se os semivariogramas experimentais direcionais em várias direções, desenhando todos num único gráfico e visualmente avaliando suas similaridades. Outra forma é através do esboço gráfico de uma elipse (conhecido também como diagrama da rosa), calculada através dos alcances obtidos em direções distintas.

A forma mais eficiente e direta de detectar a anisotropia é através do mapa de semivariograma, conhecido também como semivariograma de superfície, que é um gráfico, 2D, no qual obtém-se uma visão geral da variabilidade espacial da variável em estudo. Além disso, sobre o mapa de semivariograma é possível detectar rapidamente os eixos de anisotropia, isto é, as direções de maior e menor variabilidade espacial da variável em análise.

Segundo Yamamoto e Landim (2013) ao detectar a presença de anisotropias, elas devem ser modeladas, ou seja, ajustadas a um modelo teórico de variograma. Na fase de modelagem, bem como na sua utilização para fins de estimativa e simulação, deve-se considerar a correção da anisotropia.

2.5.6 Krigagem

O semivariograma é a ferramenta da Geoestatística que permite verificar e modelar a dependência espacial de uma variável. Uma aplicação imediata do semivariograma é a utilização das informações geradas por ele na interpolação, ou seja, na estimativa de dados e posterior mapeamento da variável. O interpolador que utiliza o semivariograma em sua modelagem é chamado de Krigagem (GUIMARÃES, 2004).

A Krigagem pode ser definida como um processo geoestatístico de estimativa de valores de variáveis distribuídas no espaço e/ou tempo, com base em valores adjacentes quando considerados interdependentes pela análise variográfica (YAMAMOTO e LANDIM, 2013). Ainda segundo os referidos autores quando comparada com os métodos tradicionais de estimativa por médias ponderadas ou por médias móveis, a principal vantagem da Krigagem é que apenas ela proporciona estimativas não tendenciosas e a mínima variância associada ao

valor estimado.

O termo Krigagem abrange uma família de algoritmos conhecidos como Krigagem simples, ordinária, universal, indicativa e com deriva externa. Segundo Yamamoto e Landim (2013), a Krigagem ordinária é o estimar mais utilizado, pela simplicidade e resultados que proporciona. É um método local de estimativa e, dessa forma, a estimativa em um ponto não amostrado resulta da combinação linear dos valores encontrados na sua vizinhança próxima.

Para Marques et al. (2012) a Planta de Valores Genéricos obtida pela Krigagem, pode ser usada para apontar o quanto um imóvel está sendo sub ou supervalorizado. E concluíram que a Krigagem é uma abordagem interessante para a produção de PVG, porque permite analisar áreas de transição de valores, dentro de uma mesma parcela e permite a partir de dados irregularmente espaçados, gerar representações temáticas isarítmicas que são, linhas traçadas de igual valor, também chamadas de isolinhas para unir pontos de igual valor da intensidade do fenômeno em estudo.

2.6 Regressão Linear Geograficamente Ponderada

A regressão geograficamente ponderada - RGP (ou do inglês, *Geographically Weighted Regression* - GWR) possibilita a modelagem espacial de dados não estacionários. Um processo espacial é dito estacionário se sua distribuição de probabilidade é invariante no espaço. Esta hipótese, que está presente no modelo de regressão global, é muito restritiva, pois somente em contextos muito particulares pode-se afirmar que um modelo único global representa adequadamente todas as partes da região de estudo.

A distribuição espacial dos valores de imóveis, seja urbano ou rural, pode-se afirmar que é um caso de dados não estacionários, pois para um mesmo tipo de imóvel, localizado no mesmo bairro os valores mudam, dependendo das características do seu entorno, ou seja os valores mudam no espaço geográfico.

Na técnica da regressão geograficamente ponderada ou regressão ponderada espacialmente, segundo Câmara et al. (2004), a ideia é ajustar um modelo de regressão a cada ponto observado, ponderando todas as demais observações como função da distância a este ponto. Desta forma, serão feitos tantos ajustes quantas observações existirem e o resultado será um conjunto de parâmetros, sendo que cada ponto considerado terá seus próprios coeficientes de ajuste. Estes parâmetros podem ser apresentados visualmente para identificar como se comportam espacialmente os relacionamentos entre variáveis.

A ideia básica da RGP é ajustar um modelo de regressão para cada ponto no conjunto de dados, ponderando as observações por uma função de distância a este ponto. Isto corresponde a considerar que pontos mais próximos ao ponto em estudo tenham maior influência nos parâmetros estimados da regressão do que observações obtidas em pontos mais distantes (CARVALHO et al., 2006).

Para Almeida (2012) a RGP constitui-se também num modo de se fazer interpolação espacial, prevendo a variável dependente para observações não amostradas.

A proposta inicial tinha como objetivo fornecer uma ferramenta útil de análise exploratória através da identificação das variações entre variáveis ao longo do espaço, o que corresponde à não estacionariedade espacial; entretanto, por auxiliar na explicação e previsão de fenômenos geográficos, este tipo de modelo pode ser classificado como confirmatório, desde que, seja avaliada com o devido cuidado sua premissa básica de influência do comportamento de uma determinada unidade geográfica sobre as suas vizinhas, e vice-versa (CARVALHO et al., 2006).

2.6.1 Concepção do modelo RGP

Considerando inicialmente um modelo clássico de regressão linear (MRCL) para dados espaciais, no qual uma variável dependente y é representada como uma combinação linear de variáveis explicativas x_k , pode-se obter uma estimativa de y para um ponto i a partir da equação de regressão:

$$y_i = \beta_0 + \sum_k \beta_k x_{ik} + \varepsilon_i \quad (27)$$

Em que o subscrito i indica regiões e β_k são coeficientes globais estimados, fornecendo uma resposta média, supostamente a mesma para qualquer região.

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

O modelo RGP é especificado como:

$$y_i = \beta_0(u_i, v_i) + \sum_k \beta_k(u_i, v_i) x_{ik} + \varepsilon_i \quad (28)$$

Em que (u_i, v_i) representa as coordenadas do ponto i no espaço e $\beta_k(u_i, v_i)$ é o coeficiente local no ponto i . O termo de erro aleatório segue distribuição normal com média zero e variância constante.

O modelo RGP pressupõe uma ponderação geográfica da estimação dos parâmetros do modelo (Equação 27): cada observação da amostra é ponderada de acordo com a sua distância para o ponto em que se baseia a regressão. As observações mais próximas possuem um peso maior, ao passo que as observações mais distantes, um peso menor (ALMEIDA, 2012).

Tal ponderação é feita pela função Kernel espacial. De forma prática, o Kernel usa a distância (d_{ij}) entre dois pontos geográficos representando duas regiões, e um parâmetro da largura da banda (b), para determinar um peso entre essas duas regiões, que é inversamente relacionado à distância geográfica (w_{ij}). A Figura 9, mostra como o Kernel espacial opera.

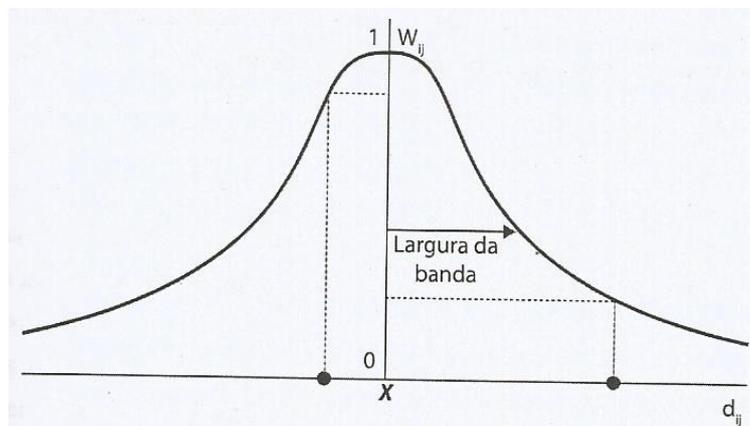


Figura 9. Kernel espacial (ALMEIDA, 2012).

Onde:

X – Ponto de regressão;

• - Observação;

W_{ij} – Peso da observação j no ponto de regressão i ;

D_{ij} – Distância entre o ponto de regressão i e a observação j .

O Kernel espacial permite fazer a calibragem do modelo para n sub amostras em torno do ponto de regressão i , formando “janelas móveis”. Cada sub amostra é definida pelo Kernel espacial. Com a especificação do Kernel espacial, é possível se obter a matriz de ponderação espacial local (ALMEIDA, 2012).

Existem diversas formas de se definir a função Kernel espacial a fim de se construir a matriz de ponderação espacial diagonal W . Os pesos a serem obtidos pela função do Kernel espacial podem ser binários, sendo a distância de varredura fixa, em um determinado valor. Ou é possível especificar pesos que são uma função contínua da distância d_{ij} por meio de funções como a função Kernel biquadrado de largura fixa, gaussiana fixa ou exponencial simples fixa.

Segundo Almeida (2012), o Kernel espacial de natureza fixa, gera desvantagem pois o número de observações dentro da sub amostra pode variar consideravelmente. Em algumas regiões podem ocorrer muitas observações (mais densa em dados) enquanto em outras a quantidade de dados serem escassos.

Para contornar este problema foi definida uma função de ponderação espacial adaptativa no sentido que se expanda em áreas em que as observações sejam escassas e encolha em áreas em que as observações sejam abundantes. A função é de natureza adaptativa pelo fato de a largura da banda se expandir em áreas menos densas de observações para abrigar k vizinhos, enquanto em áreas mais densas, a banda não precisa se expandir para que atinja os k vizinhos mais próximos.

Outro elemento necessário para ser inserido na função do Kernel espacial são as distâncias (d_{ij}) entre os pontos geográficos que representam as regiões. Esses pontos geográficos podem ser as coordenadas dos centroides de cada região. O último elemento é a largura de banda (b), que pode ser considerada como um parâmetro de suavização: quando maior a banda, maior é a suavização dos coeficientes locais; por sua vez quanto menor a banda, mais heterogeneidade nas repostas se terá, pois menos observações serão usadas ao redor do ponto de calibragem (ALMEIDA, 2012). Segundo o autor, quando a largura da banda não é conhecida, pode ser determinada pela adoção da técnica da otimizadora da validação cruzada (*cross validation - CV*). A outra forma alternativa para determinar a melhor largura de banda é através da minimização do critério de informação Akaike (AIC), ou seja, valores menores do critério AIC indicam a largura de banda mais indicada para a modelagem dos dados.

A RGP não é isenta de polêmicas, segundo Almeida (2012), pois a inferência estatística é um dos pontos fracos desta abordagem, ofuscando a apuração dos erros padrão dos coeficientes locais e a realização dos testes estatísticos do tipo t , além da possibilidade de ocorrer elevada colinearidade entre os coeficientes locais.

Com relação aos valores dos erros padrões dos coeficientes locais alguns *softwares*, como por exemplo, o ArcGIS 10.2.1 e o GWR4 têm implementados nos relatórios de saída estes valores calculados para cada coeficiente de todas as amostras. O GWR4 gera os valores de t calculados também para cada regressor de todos os dados amostrais, entretanto analisar cada um destes erros padrões e t calculados não é uma tarefa simples nem fácil.

2.6.2 Algumas aplicações do modelo RGP

A pesquisa bibliográfica por trabalhos de modelagem espacial utilizando a técnica RGP aplicada a dados de imóveis, mostrou que esta técnica não tem sido utilizada no Brasil. Entretanto na bibliografia internacional esta técnica já foi sugerida, sendo utilizada a algum tempo. Pode-se citar, por exemplo, Borst e McCluskey (2007) que compararam métodos geoestatísticos para avaliação em massa e uma das conclusões é que o desenvolvimento das técnicas de avaliação em massa vai na direção de modelos interativos híbridos. Para os autores é essencial a utilização em conjunto as abordagens tradicionais, como a regressão múltipla com elementos espacialmente derivados resultantes da técnica da regressão geograficamente ponderada.

McCluskey et al. (2013) aplicaram em uma amostra de 2.694 imóveis residenciais, a modelagem de dados, o modelo misto regressivo-autorregressivo espacial (SAR) e o modelo RGP em relação a uma rede neural artificial (ANN) e o modelo tradicional de regressão múltipla para analisar a precisão da avaliação em massa para fins de estimativa de preços. Os resultados mostraram que a abordagem da regressão geograficamente ponderada é superior em termos de explicação, de confiabilidade e de precisão. Já a rede neural artificial funcionou muito bem em termos de previsão, potência e precisão, superando, portanto, a tradicional regressão múltipla.

Entretanto os autores verificaram que o desempenho do modelo GWR é um pouco “*curates egg*”, com algumas propriedades com pouca precisão, enquanto em outras os valores são muito bons. Mas dado ao aumento do acesso a essas abordagens avançadas e seu poder preditivo ser inquestionável, afigura-se a utilidade das várias técnicas para maximizar a precisão preditiva no processo avaliatório. Isto pode ser particularmente importante quando se lida com grandes amostras típicas na avaliação de massa para fins de imposto sobre a propriedade.

Moore e Myers (2010) incorporaram um fator geoespacial, com coordenadas x, y obtidas em um SIGs nos modelos de avaliação automatizada e utilizaram a regressão geograficamente ponderada na especificação do modelo. Uma das conclusões dos autores é que com a utilização de coordenadas e a regressão geograficamente ponderada, ocorreu uma potencial melhora nos coeficientes de dispersão - CODs entre 20 a 25%.

O modelo GWR na forma tradicional, foi utilizado por Carta (2014), considerando a localização das habitações como fator de ponderação dos coeficientes de regressão, sendo que estes baseiam-se tradicionalmente em noções Euclidianas e bidimensionais do espaço. A proposta apresentada pelo autor ampliou a utilização do método GWR, estendendo a noção de espaço a lógicas que ultrapassam a perspectiva física e geográfica do território. Esta nova abordagem aumentou a capacidade explicativa dos modelos hedônicos, permitindo a inclusão de métricas de vizinhança não exclusivamente geográficas, tais como, contextos territoriais, características socioeconômicas, entre outros. A conclusão do autor, é que a aplicação desta metodologia e respetiva ferramenta ao mercado habitacional de Aveiro-Ílhavo permitiu obter melhores resultados quando comparado com os métodos tradicionais, como o MMQ e o GWR.

A utilização da regressão geograficamente ponderada em outras áreas de aplicação é muito mais frequente que nas avaliações de imóveis, sendo que as conclusões são de que a utilização da RGP sempre melhora os resultados comparado com a regressão tradicional. Pode-se citar o trabalho de Menezes et al. (2010) que utilizaram os Modelos de Regressão Linear Globais (MLRS) e Locais (RGP) em um estudo sobre renda e distância da fonte de energia elétrica no estado do Amazonas. Os autores concluíram que o modelo local (RGP)

teve um melhor desempenho que o modelo global.

A busca por modelos estatísticos que melhorem o desempenho nas avaliações de imóveis, principalmente quando se trata de avaliação imobiliária em massa para tributação (Imposto Predial e Territorial Urbano - IPTU, Imposto de Transmissão de Bens Imóveis - ITBI, Imposto Territorial Rural – ITR) é um grande desafio, pois avaliar um grande número de imóveis é mais complexo que uma avaliação individual.

Entretanto as buscas devem ser contínuas, pois o objetivo é sempre atingir a justiça fiscal. A utilização de modelos estatísticos como a regressão geograficamente ponderada pode trazer vantagens, como citado pelos autores acima, a RGP é superior a outros modelos em termos de poder de explicação, de confiabilidade e de precisão.

2.7 Medidas de Qualidade de Ajuste

Algumas medidas como o coeficiente de determinação (R^2) e o coeficiente de determinação ajustado (R^2_{ajust}), são utilizadas para avaliar a autocorrelação dos erros e a heterocedasticidade do modelo. Porém tais medidas não são mais recomendadas como um indicador de ajuste de regressão, na presença de erros autocorrelacionados e/ou heterocedásticos (ALMEIDA, 2012). Estas consequências são similares às que acontecem quando uma variável explicativa significativa é omitida do modelo de regressão. Quando se quer comparar um ajuste obtido por um modelo de regressão padrão, com um ajuste obtido por um dos modelos cuja especificação considera a autocorrelação espacial, uma medida como o R^2 não é mais confiável (CAMARGO et al., 2004).

O método mais usual de seleção de modelos de regressão baseia-se nos valores de máxima verossimilhança dos diferentes modelos, ponderando pela diferença no número de parâmetros estimados, segundo Camargo et al. (2004). Nos modelos com estrutura de dependência – espacial ou temporal - utilizam-se os critérios de informação onde a avaliação do ajuste é penalizada por uma função do número de parâmetros. É necessário ainda levar em conta o número de parâmetros independentes ao se incluir funções espaciais nos modelos. Para cada nova variável em modelo de regressão, acrescenta-se um parâmetro.

Almeida (2012) recomenda que quando da utilização da estimação por Máxima Verossimilhança (MV) que seja feita a combinação do valor da função de verossimilhança (LIK) com os critérios de informação de Akaike (AIC) e de Schwartz (SC) com o objetivo de escolher o melhor modelo.

Os critérios de informação AIC (Equação 29) e SC (Equação 30) utilizam em suas fórmulas de cálculo, o valor da função de verossimilhança (LIK):

$$AIC = -2LIK + 2k \quad (29)$$

$$SC = -2LIK + k \ln(n) \quad (30)$$

Onde:

k se refere ao número de variáveis utilizadas no modelo;

n é o número de observações utilizadas.

A interpretação da função de verossimilhança é de que quanto maior o valor dessa função, melhor é o modelo, de forma contrária a interpretação do AIC e SC é de que quanto menores os seus valores, melhor é a qualidade do modelo.

Diversos outros critérios de informação estão disponíveis, a maior parte dos quais são

variações do AIC, com mudanças na forma de penalização de parâmetros ou observações (CAMARGO et al., 2004).

Segundo Emiliano et al. (2009), outro critério de informação é o Akaike corrigido – AICc, com a seguinte correção para o AIC:

$$AIC_c = -2LIK + 2k + 2 \frac{k(k+1)}{n-k-1} \quad (31)$$

Para os autores que fizeram uma comparação dos três critérios, no caso de modelos normais, o uso do SC ou BIC (Critério de Informação Bayesiano) mostrou-se superior para amostras grandes, porém para amostras pequenas o BIC teve desempenho similar ao AIC e AICc. No caso de modelos de séries temporais, mesmo para amostras de tamanho pequeno (< 500), o desempenho dos critérios AIC e AICc foram similares e inferiores ao BIC, para todos os modelos simulados. Assim para os casos em que foram simulados, o uso do BIC para séries seria o mais indicado.

2.8 O Interpolador Kernel

Com base em Câmara e Carvalho (2004), uma alternativa simples para analisar o comportamento de padrões de pontos é estimar a intensidade pontual do processo em toda a região de estudo. Para isto, pode-se ajustar uma função bi-dimensional sobre os eventos considerados, compondo uma superfície cujo valor será proporcional à intensidade de amostras por unidade de área.

Para Santos et al. (2012), o estimador Kernel é um interpolador, que possibilita a estimação da intensidade do evento em toda a área, mesmo nas regiões onde o processo não tenha gerado nenhuma ocorrência real. É um estimador probabilístico não-paramétrico (não utiliza média e desvio padrão como parâmetro e não segue uma Distribuição Normal ou não tem elementos suficientes para afirmar que seja Normal). Uma característica importante é que se baseia na ordem (postos, ranks) dos dados. Sua única premissa básica é que ao se avaliar a ocorrência de eventos no espaço, tem que se considerar a localização dos eventos como aleatória.

Esse estimador realiza uma contagem de todos os pontos dentro de uma região de influência, ponderando-os pela distância de cada um à localização de interesse, representado conforme a Figura 10.

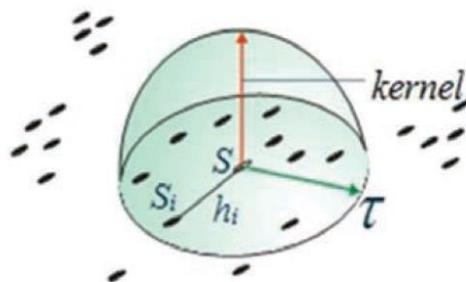


Figura 10. Estimador de intensidade para um padrão de pontos
Fonte: Silva e Nunes (2015)

A partir dos conceitos apresentados, suponha que S_1, \dots, S_n são localizações de n eventos observados em uma região A e que S representa uma localização genérica cujo valor se quer estimar. O estimador de intensidade é computado a partir dos m eventos (S_i, \dots, S_{i+m-1}) contidos num raio de tamanho τ em torno de S e da distância h entre a posição e a i -ésima amostra, a partir de funções cuja forma geral (32) é (CÂMARA e CARVALHO, 2004):

$$\hat{\lambda}_\tau(S) = \frac{1}{\tau^2} \sum_{i=1}^n k\left(\frac{h(S_i, S)}{\tau}\right), h(S_i, S) \leq \tau \quad (32)$$

Onde:

τ é o raio de influência;

h é a distância entre o ponto S e o ponto observado S_i .

O raio de influência define a área centrada no ponto de estimação “ S ” que indica quantos eventos “ S_i ” contribuem para a estimativa da função intensidade “ λ ”. Um raio muito pequeno irá gerar uma superfície muito descontínua; se for grande demais, a superfície poderá ficar muito suavizada, o resultado pode sugerir uma homogeneidade na região. Já “ h ”, representa a distância entre a localização em que desejamos calcular a função e o evento observado (CÂMARA e CARVALHO, 2004).

No caso da função de interpolação $k()$, é comum usar funções de terceira ou quarta ordem, como:

$$k(h) = \frac{3}{\pi} (1 - h^2) \quad (33)$$

Ou o Kernel gaussiano:

$$k(h) = \frac{1}{2\pi\tau} \exp\left(-\frac{h^2}{2\tau^2}\right) \quad (34)$$

Com o uso da função de quarta ordem (33), o estimador de intensidade pode ser expresso como:

$$\hat{\lambda}_\tau = \sum_{h_i \leq \tau} \frac{3}{\pi \tau^2} \left(1 - \frac{h_i^2}{\tau^2}\right)^2 \quad (35)$$

O estimador de intensidade é muito útil pois fornece uma visão geral da distribuição de primeira ordem dos eventos. Trata-se de um indicador de fácil uso e interpretação (CÂMARA e CARVALHO, 2004). O estimador Kernel possui dois parâmetros básicos que são:

- a) um raio de influência que define a vizinhança do ponto a ser interpolado;
- b) uma função de estimação com propriedades "convenientes" de suavização do fenômeno.

São vários os tipos de estimadores de Kernel, segundo Santos et al. (2012): o Gaussiano ou normal, quártico, triangular, exponencial negativo, uniforme. As diferentes funções do Kernel irão processar as informações de forma ligeiramente diferentes: a distribuição normal pesa os pontos dentro do círculo formado pelo raio de influência, de forma que pontos mais próximos são pesados mais intensamente comparados aos mais afastados. A distribuição uniforme pesa todos os pontos dentro do círculo igualmente. A função quártica dá mais peso aos pontos próximos do que aos distantes, mas o decréscimo é gradual. A função triangular pesa os pontos próximos mais do que pontos distantes dentro do círculo, mas o decréscimo é mais rápido. A função exponencial negativa pesa pontos próximos muito mais intensamente do que os pontos distantes. Para a autora o Kernel, como uma função de suavização, deve ser simétrica à origem, assim, uma das melhores soluções é aplicar um Kernel quártico. Existem ainda mais três funções: Epanechnikov, Polinomial de quinta ordem e constante, onde o parâmetro largura de banda ou raio de influência se aplica a todas as funções, exceto na função constante.

Os Kernel *estimators* é uma alternativa viável a métodos mais sofisticados de interpolação, pois não requerem a parametrização da estrutura de correlação espacial (como no caso da Geoestatística), segundo Camargo et al. (2004). As superfícies interpoladas são suaves e aproximam muitos fenômenos naturais e socioeconômicos. As desvantagens destes estimadores são a forte dependência no raio de busca e a excessiva suavização da superfície, que pode em alguns casos esconder variações locais importantes.

2.9 Exemplos de Aplicação em Avaliações de Imóveis

Existem duas escolas que propõem metodologias diferentes na abordagem do estudo da estrutura espacial nos dados, segundo Trivelloni (2005): por um lado os que propõem os métodos de regressão espacial para a modelagem dos efeitos espaciais presentes nos dados de mercado imobiliário e, por outro, os que propõem o uso de métodos geoestatísticos para a modelagem da estrutura espacial dos dados. A discussão sobre qual das duas abordagens é mais apropriada para os dados imobiliários tem recebido pouca atenção.

Os defensores da econometria espacial baseiam-se na extensão dos métodos econométricos ampla e tradicionalmente utilizados nos modelos de avaliação de imóveis, como uma justificativa natural para o uso dos modelos de regressão espacial. Os defensores dos métodos geoestatísticos baseiam-se nas vantagens deste método para tratar a variabilidade espacial, baseada na análise empírica da variabilidade dos dados, a possibilidade de encontrar o modelo de variação mais adequado e de modelagem da anisotropia. Ao mesmo tempo é questionada a falta de critérios adequados para a escolha da matriz de pesos W que determina os testes de autocorrelação espacial e a estrutura espacial dos dados nos modelos de regressão.

Anselin (1999) defende o critério de escolha baseado na análise do tipo de dados e o consequente modelo de dados a ser aplicado em cada caso: para dados com características de objetos espaciais, que são observações heterogêneas nas suas características intrínsecas e onde estas características participam e influenciam no processo de modelagem, defende que seja aplicada a metodologia da econometria espacial; enquanto que para dados com características de geocampos, ou variáveis homogêneas intrinsecamente e com variação contínua no espaço, deve ser aplicada a metodologia Geoestatística. Ainda, segundo este autor, o caso específico do valor dos imóveis parece apresentar características de ambos os tipos de modelos de dados,

deixando uma interrogação quanto ao melhor método para modelagem do valor.

Os métodos econométricos espaciais e os modelos geoestatísticos, são comparados por Dubin (2003), numa aplicação prática, que analisou os resultados em função do grau de ajustamento entre os valores observados e calculados por ambos os métodos e a análise dos erros obtidos sobre uma mesma amostra de dados aplicando as duas metodologias. O autor conclui que os modelos geoestatísticos apresentam superioridade na estimação por inferência estatística sobre os modelos de regressão espacial, sendo que ambos os métodos por sua vez são claramente superiores aos resultados obtidos por regressão múltipla por mínimos quadrados.

Dantas et al. (2001) aplicou modelos de econometria espacial, através da regressão espacial, na avaliação de imóveis na cidade de Recife-PE e concluíram que para se ter uma avaliação confiável do mercado imobiliário é fundamental que se verifiquem, além dos pressupostos básicos do modelo de regressão tradicionalmente adotado, a dependência espacial.

Uma amostra temporal de dados de imóveis na cidade de Granada para realizar uma análise da variabilidade espacial e temporal dos preços, foi utilizado por Chica Olmo et al. (2007), tanto por meio de análise de semivariogramas (ou variográfica) direta das amostras experimentais como uma análise de cruzamento variográfico. A análise variográfica concluiu que o preço dos imóveis são variáveis muito correlacionadas espacialmente e a estrutura da variabilidade mostra duas escalas de resultado: a variação espacial dos valores dos imóveis com relação à localização é influenciada por fatores micro, ou seja, pela vizinhança ao imóvel, bem como por fatores macro, em função das características do bairro, como acessibilidade ao centro, rede de transportes, densidade residencial, etc.

As duas metodologias (econometria espacial e Geoestatística) em amostra em bairros da cidade de São José-SC, foi aplicado por Trivelloni (2005). Os modelos de estatística espacial, em especial os de regressão espacial e os métodos geoestatísticos foram usados de forma combinada para estimar o valor da localização dos imóveis. Uma das conclusões é que as técnicas de econometria espacial e de Geoestatística, tradicionalmente utilizadas de forma separada e excludente, podem ser usadas em forma combinada na avaliação em massa de imóveis. Os resultados obtidos foram superiores aos conseguidos usando uma das técnicas de forma individual, possibilitando a análise de uma amostra heterogênea de dados e a determinação do valor da localização. O método proposto permitiu a elaboração de uma Planta de Valores Genéricos na área de estudo para todos os tipos de imóveis incluídos na amostra de mercado.

Santos (2014) também combinou as duas metodologias para modelar o mercado de terras rurais em Petrolina-PE. O autor concluiu que a combinação de metodologias se mostrou eficiente para modelar os efeitos espaciais nos mercados de terras rurais. Por meio da econometria espacial foi possível controlar a autocorrelação espacial existente no MCRL e por meio da Geoestatística, particularmente o interpolador da Krigagem ordinária foi possível obter as superfícies de preços para os modelos obtidos. Ressalta-se que o autor utilizou uma amostra de validação que não fez parte da modelagem e a mesma foi importante para que a avaliação de desempenho fosse capaz de mostrar a eficiência dos modelos bem como mostrar os pontos que necessitam ser aprimorados.

Gonzáles et al. (2002) apresentam uma comparação entre a abordagem tradicional de regressão, modelagem de erro, superfícies de resposta e superfície de tendência - TSA. Utilizaram também redes neurais artificiais - ANN para estimar alguns modelos, comparando seus resultados. Um estudo de caso foi realizado usando cerca de 30.000 dados de arrecadação do Imposto sobre Transmissão de Bens Imóveis - ITBI sobre a venda de

apartamento em Porto Alegre - RS. Os autores realizaram um estudo exploratório, na tentativa de contribuir para desenvolver formas alternativas de considerar a localização nas avaliações em massa. Os resultados indicaram que a TSA é uma boa alternativa a considerar o atributo localização. Os resultados são semelhantes às outras técnicas, mas a TSA tem mérito em diminuir a subjetividade e para ajustar o modelo em uma única etapa, com uma única ferramenta.

Na avaliação em massa de imóveis rurais, Machado (2006) empregou no tratamento estatístico dos dados a regressão múltipla, incluindo também a TSA, que leva em consideração as coordenadas UTM – Universal Transversa de Mercator dos imóveis, para modelar a variável localização. Os melhores resultados para os diversos testes estatísticos realizados, bem como a redução da autocorrelação espacial, foram obtidos com a aplicação do polinômio de quarta ordem.

Resener e Hochheim (2006) utilizou superfície de tendência em uma amostra de imóveis rurais para calcular valores de indenização da Usina Hidrelétrica de Barra Grande, entre os estados de Santa Catarina e Rio Grande do Sul. As modelagens foram realizadas considerando variáveis de uso e aptidão dos solos, além da localização. A conclusão mais evidente é a de que em amostras com grande amplitude espacial, como neste caso, a localização é mais importante do que as características intrínsecas de cada imóvel.

A Geoestatística foi utilizada por Macedo (2013) para determinar o preço da terra no estado de Goiás, levando em consideração variáveis de localização, socioeconômica e de infraestrutura nas análises realizadas por geoprocessamento. O autor realizou uma análise comparativa entre dois interpoladores geoestatísticos denominados Krigagem e cokrigagem e concluiu que o processo de interpolação por Krigagem se mostrou suficiente para a determinação do preço da terra. A determinação de pesos sobre variáveis ambientais, econômicas e sociais é possível, durante o processo de interpolação por cokrigagem, contudo, ocorreram variações distintas da influência dos pesos aplicados sobre as variáveis durante o processo de interpolação para a determinação do preço da terra para agricultura, pastagem ou cerrado. Os pesos definidos para as variáveis secundárias apresentaram uma maior relação com o preço da terra para pastagem.

3 MATERIAL E MÉTODOS

Neste item serão descritos os procedimentos utilizados nessa pesquisa. Sendo assim, inicialmente serão informados quais os recursos cartográficos e tecnológicos utilizados para execução das atividades, em seguida será apresentada a região do PROJIR onde foram coletados os dados, e finalmente os procedimentos metodológicos adotados, com a apresentação de fluxograma com as etapas estatísticas com o objetivo de facilitar a compreensão da metodologia.

Na proposta metodológica desta pesquisa utiliza-se os métodos estatísticos da regressão clássica, da regressão espacial e da RGP para realizar a modelagem do mercado de terras rurais, e da Geoestatística para gerar a superfície de valores (PVG), validar e avaliar o desempenho dos produtos gerados.

3.1 Materiais

Para trabalhar com regressão espacial ocorreu a necessidade de utilizar dados cartográficos georreferenciados, portanto foram utilizados os seguintes materiais:

- a) Planta de Cadastro Rural, na escala 1:10.000;
- b) Planta de Cartografia Básica, na escala 1:10.000;
- c) Planta de Classes de Solo e de Aptidão das Terras Agrícola para Irrigação, na escala 1:10.000;
- d) Cadastro dos imóveis rurais – banco de dados do Access;
- e) Imagens QuickBird e RapidEye;
- f) *Software* PCI Geomática 9.1 (PCI, 2003);
- g) *Software* livre GeoDa (ANSELIN, 2005);
- h) *Software* INFER 32 (ÁRIA, 2003);
- i) *Software* ArcGIS 10.2.1 (ESRI, 2013);
- j) *Software* Spring (SPRING, 1996);
- k) *Software* GWR4 (NAKAYA et al., 2014);

3.2 Área de Aplicação da Pesquisa

Para a aplicação do estudo proposto foi selecionada a área do PROJIR - Projeto de Irrigação e Drenagem da Cana-de-açúcar na Região Norte Fluminense, que abrange fundamentalmente, áreas dos municípios de São João da Barra, Campos dos Goytacazes, Cardoso Moreira, São Francisco de Itabapoana, Conceição de Macabu, Carapebús e Quissamã, no Estado do Rio de Janeiro – RJ (Figura 11).

A área do PROJIR se localiza aproximadamente entre as coordenadas 21°17'15"W / 40°59'40"S e 22°04'55"W / 41°45'01"S distante a cerca de 280 km da capital Rio de Janeiro. O perímetro do PROJIR delimitado para a realização do levantamento foi de aproximadamente 250.000,00 ha e correspondia nos anos 80 à região de maior concentração da lavoura canavieira. No ano de 1982 foram realizados os levantamentos de Cartografia Básica, Cadastro de Imóveis Rurais, Mapa de Solos e perfis pedológicos, de Aptidão de Terras para Irrigação, além de outros mapas pedológicos, geológicos e hidrogeológicos (PROJIR, 1982). Nesta mesma ocasião foi realizado um semi-cadastro dos imóveis rurais da área e que foi posteriormente digitalizado.

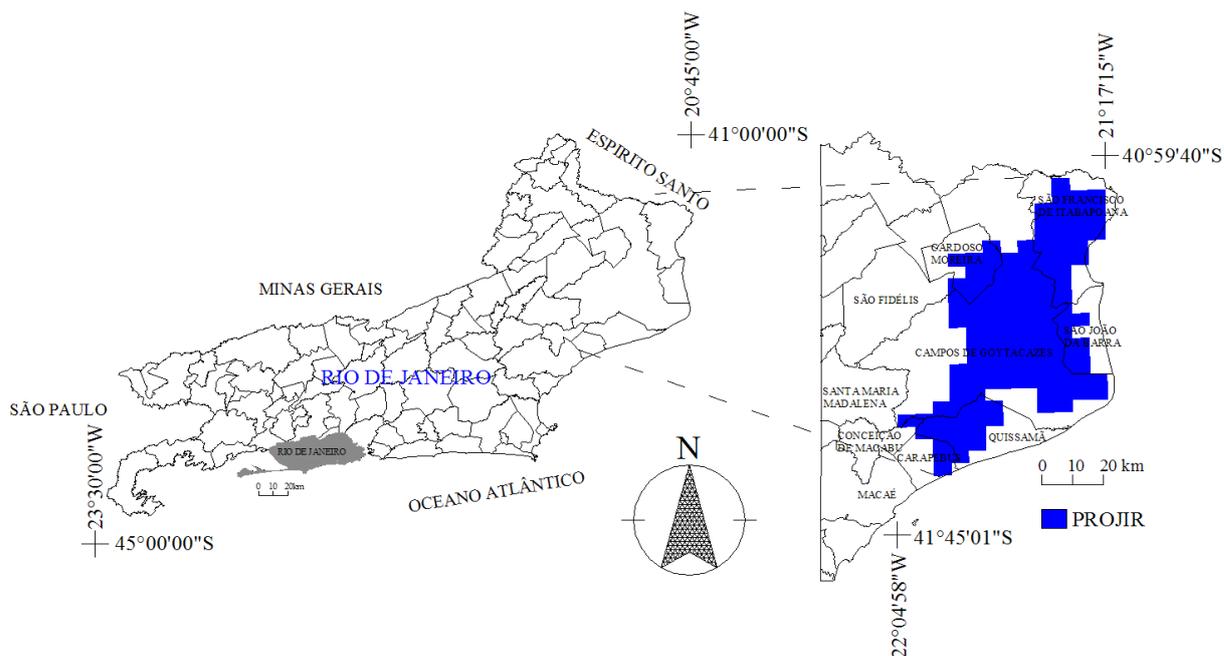


Figura 11. Localização do PROJIR

Este material cartográfico e banco de dados, gentilmente cedidos para esta pesquisa, está aos cuidados da biblioteca do Campus Leonel Miranda, localizado no município de Campos dos Goytacazes, RJ. Trata-se de um campus avançado da Universidade Federal Rural do Rio de Janeiro – UFRRJ, criado em 1991 com a transferência da estação experimental do antigo Programa Nacional de Melhoramento da Cana-de-açúcar - PLANALSUCAR para a universidade. O Campus deu continuidade à pesquisa no setor canavieiro e representa um importante centro de apoio ao ensino, à pesquisa e à extensão agropecuária nas regiões Norte e Noroeste Fluminense.

A área do PROJIR não corresponde na sua totalidade à região Norte Fluminense, pois além de abranger somente parte dos municípios citados, pertencem ainda ao Norte Fluminense os municípios de São Fidélis e Macaé.

A região Norte Fluminense é considerada de predomínio tradicional de atividade agrícola, pois seu relevo plano e suave-ondulado, e clima tropical com estação seca de inverno, sendo Aw conforme classificação de Köppen-Geiger (KOTTEK et al., 2006), fez com que esta região estruturasse sua economia sobre a lavoura de cana-de-açúcar. A região é banhada pelas águas dos rios Paraíba do Sul e Muriaé, o que proporciona a disponibilidade hídrica para a prática da agricultura irrigada.

Com a implantação de novas técnicas de produção açucareira e a modernização de estruturas de transporte, como a instalação de ferrovias, os antigos engenhos foram substituídos por engenhos centrais e usinas, que proporcionaram um melhor escoamento da produção agropecuária, recebimento de mercadorias de outras regiões e aumento da circulação de pessoas.

Para Souza et al. (2009) a região vivencia um processo de empobrecimento no campo, em parte devido às condições adversas do mercado de seu principal produto, a cana-de-açúcar. A produção de cana-de-açúcar na região retraiu-se por efeito dos sucessivos planos econômicos, da extinção do IAA - Instituto do Açúcar e do Alcool, da redução dos estímulos propiciados pelo Proálcool, o reduzido investimento em insumos na lavoura por parte das usinas que resulta em baixa produtividade regional, fator de desvantagem num contexto de

acirramento da competição com outras regiões.

A Região Norte Fluminense beneficiou-se a partir da década de 1970, com a implantação da Petrobras, em Macaé, como base de operação das atividades de prospecção e de produção para os recém-descobertos campos de petróleo na plataforma continental da Bacia de Campos (PIQUET, 2003).

A partir de então, segundo Forti (2013), observa-se na região, um apanhado de transformações sociais e ambientais. O novo ciclo voltado à produção petrolífera em que a região se insere, demandou uma mão de obra especializada forçando o crescimento do setor terciário, gerando a constituição de novas universidades e escolas técnicas.

Os municípios da região passam a receber *royalties* (e mais tarde as participações especiais) como compensação financeira. Como um agente de grande transformação regional, esse advento contribui para fragmentações das unidades territoriais municipais e até mesmo a criação de uma nova região, como aponta Cruz (2007) impondo assim uma nova configuração para a Região Norte Fluminense, que passou a ser composta pelos atuais municípios.

Para Forti (2013) em um curto período de tempo a região, passou por uma transição, onde o foco das atividades dos municípios mais dinâmicos muda para serviços e a extração mineral, fomentada pela atividade de exploração de petróleo e gás natural, que necessitam de mão de obra qualificada.

Ressalta-se que com a implantação destes novos polos industriais, necessários para a produção petrolífera, e o natural aumento da população surgiram mudanças no uso do solo. Áreas que eram de uso agrícola, são alteradas para usos industriais e urbanos e acompanhadas, como historicamente acontece nestes casos, da valorização dos imóveis.

Gomes (2012) analisou a evolução do valor do m² dos imóveis em Macaé e em Campos dos Goytacazes, entre 1981 e 2011, dois dos municípios do Norte Fluminense que recebem *royalties*. Em Macaé na região central do município os valores aumentaram de R\$/m² 131,35 em 1981 para R\$/m² 1.331,56 em 2011, ou seja, valorizaram neste período 913,74%. Em Campos, na região central os valores aumentaram de R\$/m² 169,11 para R\$/m² 538,41, um aumento de 218,37% (dados correntes a época da coleta de informações). Para o autor, neste período, ocorreram a ampliação e diversificação dos produtos imobiliários e a concentração de investimentos públicos em áreas específicas das cidades e também, o alargamento da distância entre os preços praticados em áreas consideradas de maior interesse imobiliário e aquelas habitadas ou destinadas à população de menor poder aquisitivo.

Para Nakamura (2012) os investimentos vultosos decorrentes da exploração de petróleo na camada Pré-Sal, veem acompanhado por obras de infraestrutura e de grandes empreendimentos industriais. A expectativa é a de que uma corrente de migração atinja algumas das cidades da região Norte Fluminense, em decorrência da construção do COMPERJ - Complexo Petroquímico do Rio de Janeiro, da Petrobras.

Estas transformações são encontradas em todos os segmentos da indústria imobiliária, mas o volume mais expressivo é no residencial, tanto para habitações de padrão econômico, quanto para moradias de médio e alto padrão. Com a instalação da nova refinaria da Petrobras, com perspectiva de gerar cerca de 170 mil empregos diretos e indiretos, os vazios rurais e urbanos, são objetos de novas implantações, voltados para o uso comercial e para novas áreas industriais.

Outro grande investimento que prometeu impulsionar a região foi a construção do CLIPA - Complexo Logístico-Industrial do Porto do Açú. Segundo Passos (2013) com a construção do CLIPA, o setor de construção civil na região Norte Fluminense e principalmente em São João da Barra, foi fortemente impactado, levando assim a um grande crescimento da mão de obra no setor.

Para Ribeiro (2010), o Complexo Portuário do Açú afetou a região, com os lançamentos de novos empreendimentos imobiliários, assim como os valores dos aluguéis na região que estão ainda mais caros do que já estavam, num processo acelerado de especulação imobiliária.

Com a construção do Complexo Portuário, Rodrigues et al. (2010) detectou consideráveis modificações na dinâmica do Município de São João da Barra. Uma das mudanças foi a supervalorização dos imóveis do município, levando em consideração tanto para locação como para venda. Foi constatado, através das entrevistas com os corretores das imobiliárias do município, que os preços para locação valorizaram 100% e os para venda, 50%.

Para o IPEA (2011), a capacidade instalada e em expansão da infraestrutura urbana, de formação e qualificação de mão de obra, e de logística e serviços de Campos e de São João da Barra, há possibilidades de ocorrência, induzida e espontânea, de alocação de parte dos investimentos em municípios vizinhos, como São Francisco do Itabapoana, São Fidélis e Cardoso Moreira, que podem se tornar municípios-dormitório, pelas condições mais acessíveis aos custos dos imóveis e dos aluguéis.

Conclui-se que com a implantação destes empreendimentos (indústria petrolífera e recentemente o Porto do Açú) ocorreu um forte êxodo rural e a migração de outras regiões do país em direção ao Norte Fluminense, e provocaram um crescimento populacional acompanhado de grande expansão urbana, especialmente em Macaé, Campos dos Goytacazes e São João da Barra.

Com a expansão urbana aconteceu a valorização das terras nestes municípios e sobrecarregam a infraestrutura urbana regional, levando a impactos ambientais como ocupação de áreas de risco, de proteção ambiental e permanente.

Contudo, alguns fatos recentes podem alterar esta dinâmica nos preços dos imóveis da região. Primeiro: a queda do preço do barril de petróleo para a metade, que por um lado diminui os royalties e por outro inviabiliza o Pré-Sal. Segundo: a incerteza ocasionada aos problemas financeiros das empresas do Grupo X: MMX, MPX, LLX, OGX e OSX que controlam o Porto do Açú e que não atraiu investimento suficiente para torná-lo sustentável. Terceiro e não menos relevante é a queda do nível de crescimento da China que fez com que reduzisse a demanda por *commodities*, incluindo-se o minério de ferro, que é o principal produto a ser exportado pelo Porto do Açú.

Estes problemas citados já estão provocando uma diminuição na “bolha imobiliária” que havia se formado desde o início da construção do Porto do Açú. Isso porque dezenas de imóveis que haviam sido alugados pelas empresas do Grupo X para seus gerentes, etc., estão sendo devolvidos. Com isso a oferta de imóveis ficou maior do que a procura e os valores de venda e de aluguéis tendem a diminuir, bem como a especulação imobiliária.

3.3 Obtenção dos Dados

3.3.1 Coleta dos dados de mercado imobiliário

Em agosto de 2014 foi realizada visita ao Campus Leonel Miranda, localizado no Município de Campos dos Goytacazes – RJ para conhecer a região do PROJIR e solicitar as plantas e a imagem *QuickBird* que foram utilizadas nesta pesquisa. Nesta ocasião em conversas com os técnicos e pesquisadores foram coletadas algumas informações como: características físicas, as vocações agrícolas da região e as tendências mercadológicas que atuam na região e que podem influenciar o valor dos imóveis.

Essas entrevistas com os engenheiros agrônomos do Campus Leonel foram

importantes na elaboração do formulário de pesquisa de campo utilizado para a coleta de dados. No formulário constam: a identificação do imóvel e do informante, as condições das vias de acesso e de trafegabilidade, as características físicas e ambientais dos imóveis, o uso atual, entre outras informações, e os valores e condições de pagamento, além da informação se o valor do imóvel é oferta ou é o valor final negociado.

As coletas dos dados utilizados como amostras dos imóveis foram realizadas no período de junho a agosto de 2015, junto aos corretores de imóveis, através de anúncios em páginas na internet, com peritos judiciais e nas imobiliárias que atuam na região do PROJIR. Desde o início surgiram dificuldades, na coleta de dados de imóveis junto às imobiliárias, nesta que é a etapa mais exaustiva do processo avaliatório, porém é uma das mais importantes. Através da ajuda de um perito avaliador que atua no município de Campos dos Goytacazes, a pesquisa teve continuidade, pois com muita boa vontade, este profissional auxiliou na coleta dos dados e na sua correta localização.

Além dos dados coletados no mercado imobiliário da região foram usados dados fornecidos pelo INCRA da Regional do Rio de Janeiro, através do “Relatório Referente à Planilha de Preços Referenciais da Região Norte Fluminense” realizado em fevereiro de 2013 (INCRA, 2013). Deste relatório foram aproveitados nesta pesquisa 32 dados de imóveis coletados pelos profissionais do INCRA, já que depois de analisados se observou que os valores ainda estavam atualizados de acordo com o mercado imobiliário em 2015.

Ao final da coleta dos dados obteve-se, entre ofertas e transações, 138 dados de mercado, nos diferentes municípios do PROJIR: São João da Barra, Campos dos Goytacazes, Cardoso Moreira, Conceição de Macabu, Carapebús e Quissamã. De acordo com o semi-cadastro realizado em 1982 a área do PROJIR possui 9.685 imóveis rurais, sendo assim, a amostra obtida corresponde a aproximadamente 1,42% do total de imóveis.

Destas foi retirada uma amostra composta por 25 pontos para a validação dos modelos obtidos e na avaliação de desempenho, que corresponde a 22,1% dos dados. Por fim a amostra de trabalho contou com 113 observações de mercado, que foram devidamente inseridas em planilha eletrônica. A distribuição espacial da amostra de trabalho e dos pontos de verificação pode ser visualizada na Figura 12.

As amostras de trabalho e de verificação utilizadas, tiveram a distribuição mostrada na Tabela 3 nos municípios do PROJIR. Todos os municípios tiveram amostras, com exceção de Carapebus que ficou sem amostra de verificação.

Tabela 3. Quantidade de amostras em cada município

Município	Amostra de trabalho	Amostra de verificação
São Francisco de Itabapoana	19	4
Cardoso Moreira	2	2
Campos dos Goytacazes	70	14
São João da Barra	9	2
Quissamã	11	3
Carapebus	2	0

A distribuição das amostras de trabalho com relação a área, varia de menos de 1 (um) hectare até acima de 1000 hectares, sendo o maior número de amostras de 10 a 30 hectares. Com relação a amostra de verificação existe pelo menos uma amostra para cada grupo de área, com exceção de 500 a 1000 hectares que ficou sem amostra (Tabela 4).

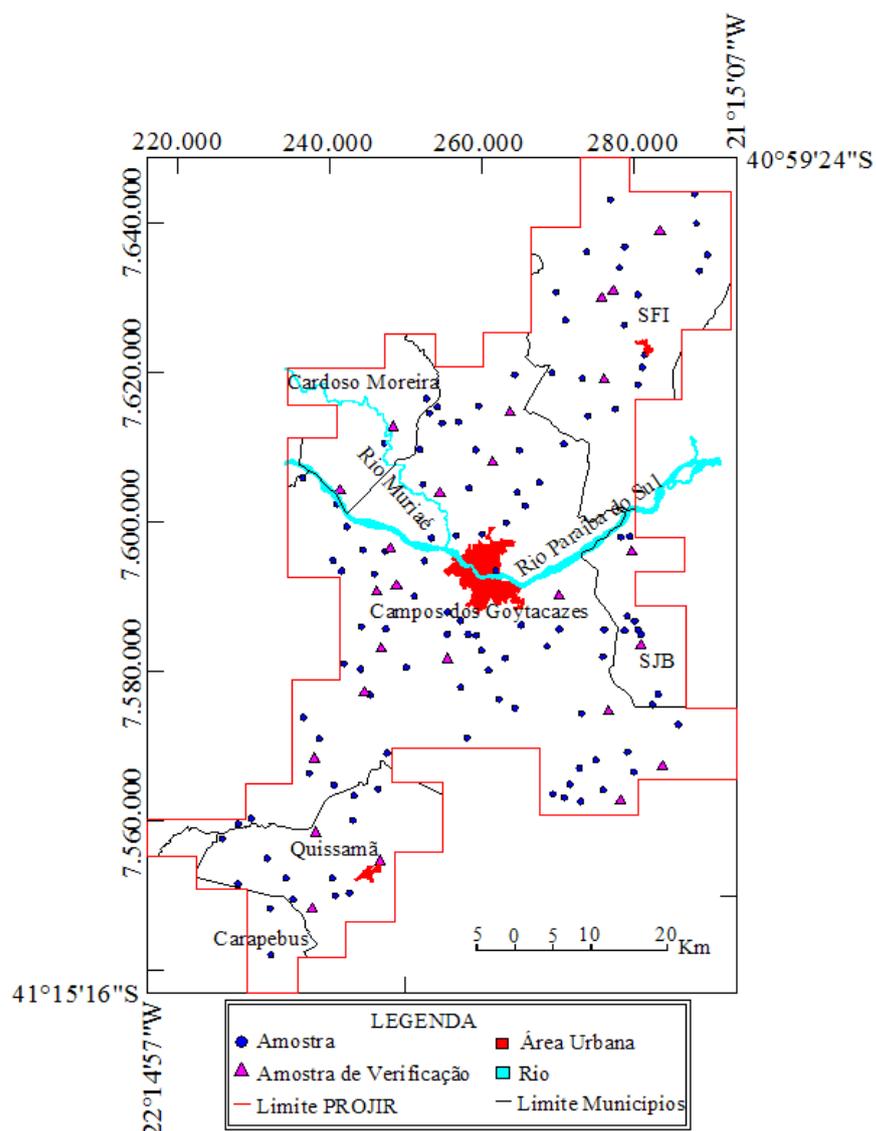


Figura 12. Distribuição espacial das amostras

Tabela 4. Distribuição das áreas

Área	Amostra de trabalho	Amostra de verificação
Até 10 ha	15	3
De 10 a 30 ha	30	2
De 30 a 50 ha	15	3
De 50 a 70 ha	5	3
De 70 a 100 ha	9	2
De 100 a 150 ha	9	2
De 150 a 250 ha	12	3
De 250 a 500 ha	7	6
De 500 a 1000 ha	6	0
Acima de 1000 ha	5	1

De acordo com NBR 14.653:3, ABNT (2004), quando no processo de avaliação é utilizado o método comparativo direto de dados de mercado e regressão múltipla, o tamanho adequado da amostra se baseia no número de variáveis (k) envolvidas na construção do modelo de regressão linear. Para atingir o grau III de fundamentação, com relação à quantidade de dados de mercado efetivamente utilizados devem ser no mínimo de 6 (k+1). Por exemplo, se em um modelo foram utilizadas 10 variáveis independentes a quantidade de dados deve ser maior ou igual a 66 dados.

3.3.2 Georreferenciamento e vetorização

As plantas de Cartografia Básica, Cadastro Rural e de Tipo de Solo/Aptidão das Terras para Irrigação cedidas pela biblioteca do Campus Leonel Miranda da UFRRJ foram digitalizadas no *Scanner HP Designjet T 2300* com resolução de 600 dpi e georreferenciadas no *software* PCI Geomática 9.1, dentro das especificações cartográficas de controle de qualidade e precisão. O equipamento utilizado pertence ao Curso de Engenharia de Agrimensura e Cartográfica, do Departamento de Engenharia da UFRRJ.

Posteriormente, o contorno das amostras de imóveis rurais utilizadas na pesquisa foi vetorizado no *software* ArcGIS 10.2.1. Além do contorno foram vetorizados os dados da cartografia básica e as linhas representativas do tipo do solo e aptidão para irrigação de cada amostra.

As plantas de Classes de Solo e de Aptidão das Terras Agrícola para Irrigação na escala 1:10.000, são uma grande exceção no Brasil, pois estes tipos de informação geralmente estão em escalas pequenas como 1:50.000 ou acima. Esta é uma das grandes dificuldades em trabalhar com áreas rurais, a falta de produtos cartográficos em escalas grandes.

3.3.3 Classificação das imagens *RapidEye*

Um dos objetivos desta pesquisa é a utilização de variáveis ambientais, e para tanto foram utilizadas as imagens *RapidEye* e uma imagem *QuickBird*. A imagem *QuickBird* com resolução espacial de 0,60 metros e ortorretificada, foi cedida para esta pesquisa pelo Campus Leonel Miranda, além das informações do semi-cadastro dos imóveis rurais digitalizados. Quanto às imagens *RapidEye*, com resolução espacial de 5 metros e ortorretificadas, foram utilizadas as imagens que o Governo Federal contratou para todo o território nacional para a realização do Cadastro Ambiental Rural – CAR (Ministério do Meio Ambiente, 2013).

Para maior autenticidade de seus dados, realizou-se a correção atmosférica nas imagens para obter a reflectância ao nível de superfície. Para cada imagem foram calculados os valores de ganho correspondente a cada banda, e em seguida foram transformadas para reflectância no topo da atmosfera, alterando o valor de ganho encontrado em cada banda da imagem.

Para avaliar a condição da cobertura vegetal, foi realizado o cálculo do NDVI - *Normalized Difference Vegetation Index* ou IVDN - Índice de Vegetação por Diferença Normalizada. Através desse índice pode-se fazer análises sobre a cobertura vegetal de determinada região.

Em seguida, de posse das imagens transformadas foi realizada a classificação das imagens com o objetivo de extrair as informações de interesse, relacionadas às áreas de APP, reflorestamento e mata ciliar. No processo de classificação, determinou-se as classes e os níveis de cinza referentes a cada classe; e através da classificação Supervisionada de Maxver, utilizando um limiar de aceitação de 99% por apresentar melhores resultados, obteve-se os mapas temáticos contendo as classes de interesse, sendo possível verificar a distribuição

geográfica de cada classe. Realizada a classificação, foi extraída as áreas de APP, mata ciliar e reflorestamento, para posteriormente testar se as mesmas influenciam o valor dos imóveis rurais.

3.4 Aplicação dos Métodos Estatísticos

Nesta pesquisa foram utilizados dados de mercado imobiliário composto por transações e ofertas, atuais e georreferenciados. Uma vez que os contornos dos imóveis foram vetorizados, os dados espaciais utilizados nesta pesquisa podem ser classificados como dados pontuais (*point pattern*) e também como dados de área (*Lattice*). Estes dados foram submetidos a uma análise exploratória e em seguida foi realizada a Modelagem Clássica por Regressão Linear Múltipla. Nesta etapa, as variáveis consideradas influenciadoras foram escolhidas e testadas. Se a variável se mostrasse importante esta era utilizada na modelagem, caso contrário, se comprovado através de testes estatísticos como Student e da matriz de correlação parcial, a sua não significância, então era rejeitada para a modelagem.

Esta etapa foi detalhada através do fluxograma da Figura 13 por se tratar da fase onde são aplicadas as técnicas da Econometria Espacial, da Regressão Geograficamente Ponderada e da Geoestatística.

Nesta fase foi definido e validado o MCRL do mercado de terras rurais da área do PROJIR, onde foi verificada a não violação dos pressupostos do MCRL. A análise exploratória da variável dependente e das variáveis independentes ajudou na escolha das variáveis mais importantes, bem como na transformação a ser aplicada na variável dependente, para não incorrer em problemas na especificação do modelo e nem induzir a autocorrelação espacial nas próximas fases. Para a definição do MCRL foi utilizado o *software* INFER 32, específico da área de Engenharia de Avaliações.

Na etapa seguinte, levando-se em consideração que o modelo foi corretamente especificado, foi determinada a matriz de pesos espaciais que obteve o maior valor de autocorrelação espacial. Nesta etapa, foram utilizados os resíduos extraídos do modelo obtido na etapa anterior no *software* Geoda. Utilizando como entrada de dados a matriz de pesos espacial, pode-se realizar a etapa da Econometria Espacial. A referida matriz permitiu identificar se a autocorrelação espacial está presente na variável dependente (Modelo SAR) ou nos erros (Modelo SEM ou CAR). Após a especificação do modelo de regressão espacial - MRE, este foi validado através dos testes para verificar a não violação dos seus pressupostos.

Na próxima etapa foi realizada a Regressão Geograficamente Ponderada - RGP, utilizando-se o modelo da MCRL definido anteriormente e com as mesmas variáveis explicativas ou independentes foi executado o RGP no *software* GWR4. Para executar o RGP foi necessário definir alguns parâmetros como a largura de banda ótima utilizando-se alguns critérios como AIC, AICc ou CV. Definida a largura o modelo RGP foi obtido sendo representado pelos seus valores médios, porque este tipo de modelo gera coeficientes diferentes para cada variável independente e para cada amostra utilizada na análise.

Os modelos MCRL, MRE e RGP tiveram a qualidade do seu ajuste verificados pelos critérios de informação de Akaike - AIC, pelo R^2 , pela LIK e pelo erro padrão da regressão. O objetivo foi comparar o ajuste dos diferentes modelos utilizando um mesmo critério e assim definir qual destes obtiveram o melhor ajustamento. Para avaliar o desempenho dos modelos de regressão foi separado uma amostra de validação que não foi utilizada na modelagem, que foram especialmente coletados para esta avaliação. Para a amostra de validação foi calculada os parâmetros Mediana das Razões, Coeficiente de Dispersão - COD e Diferencial Relativo ao Preço - PRD e comparados com os valores de referência recomendados pela IAAO.

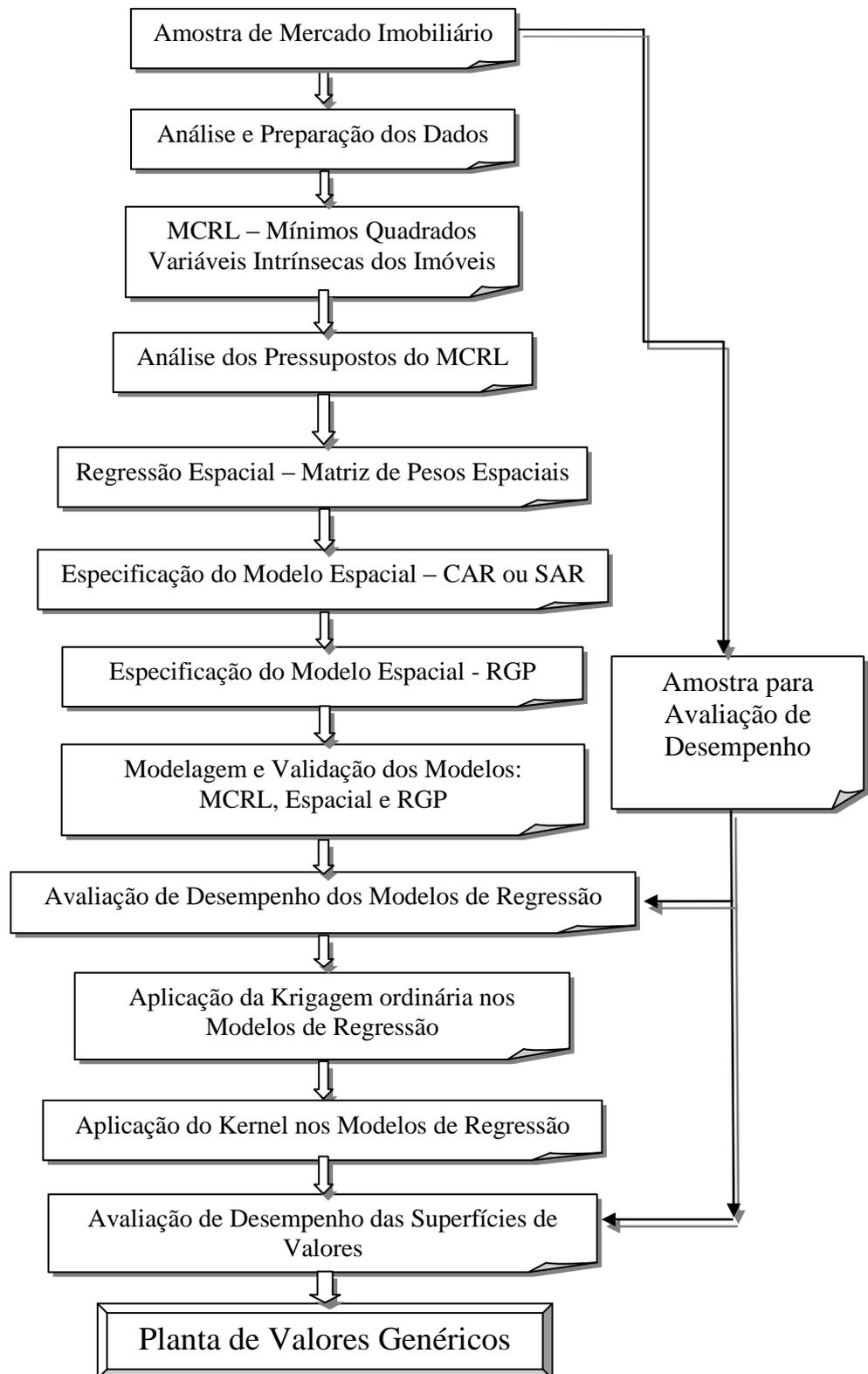


Figura 13. Fluxograma dos métodos estatísticos

Para a realização das etapas citadas até agora, foram utilizadas as observações originais bem como os valores preditos pelos modelos para gerar as superfícies de preços por meio da Krigagem no *software* ArcGIS. Esta etapa que envolve a Geoestatística, tem como objetivo interpolar os valores preditos de cada modelo e os pontos amostrais de modo a obter uma superfície de valores que corresponderá à PVG através da Krigagem ordinária.

Com o objetivo de gerar as superfícies de valores com outro estimador, foi utilizado no *software* ArcGIS, o estimador de intensidade Kernel que ajusta uma função bi-dimensional sobre os eventos considerados, compondo uma superfície cujo valor será proporcional à intensidade de amostras por unidade de área. O Kernel permite a estimação da intensidade do evento em toda a área, mesmo nas regiões onde o processo não tenha gerado nenhuma ocorrência real.

Com a superfície de valores definida pela Krigagem ordinária e pelo Kernel, que corresponde à PVG para cada modelo: o MCRL, o MRE e o RGP, foi realizada novamente a avaliação de desempenho dos modelos com os novos valores preditos. Para tanto foi utilizada novamente a amostra de dados que não fez parte na modelagem das regressões. Foram calculados os indicadores estabelecidos na norma da IAAO, a Raiz Quadrada do Erro Médio Quadrático – REMQ, o erro relativo e o valor absoluto do erro relativo analisando-se a conformidade dos modelos. O produto desta etapa foram os modelos da Krigagem e Kernel avaliados no seu desempenho. O produto final é a Planta de Valores Genéricos das terras rurais da área do PROJIR.

4 RESULTADOS E DISCUSSÃO

4.1 Identificação das Variáveis Influenciantes

A partir da tabulação dos dados de mercado e do contato com os profissionais que atuam na região, foram geradas as seguintes suposições a respeito das variáveis que poderiam influenciar os preços das terras na área de estudo.

a) Variável dependente:

Valor Unitário – VU (R\$/ha): como o objetivo da pesquisa é o valor da terra nua, o valor unitário foi obtido deduzindo-se o valor das benfeitorias de cada imóvel do seu valor total, através do método da quantificação do custo (ABNT, 2004). Esta variável foi previamente homogeneizada para retirar dos imóveis colocados em oferta, os acréscimos sistemáticos que geralmente são dados pelos ofertantes, através da aplicação do fator oferta de 0,90.

b) Variáveis independentes:

Coordenadas UTM (X,Y): coordenadas em metros que identificam o centroide da localização de cada amostra na planta;

Vias de acesso: variável que representa as condições das vias de acesso aos imóveis. Para Lima (2011) a qualidade dos acessos também exerce influência no valor da terra nua em imóveis rurais;

Trafegabilidade: variável que representa as condições de tráfego durante o ano;

Distância a polo de valorização (km): a distância ao mercado consumidor, segundo Lima (2011) também influencia o valor das terras. A área do PROJIR está abrangendo vários municípios, entretanto o município mais expressivo é Campos de Goytacazes, então foi considerada a distância em quilômetros do centro urbano de Campos até a amostra considerada;

Distância ao mar (km): distância da amostra até o litoral. A proximidade do litoral sempre exerce influência nos valores dos imóveis, e com a construção do Porto do Açú, a mesma foi considerável na região do PROJIR;

Disponibilidade de água: variável ambiental que caracteriza a disponibilidade de recursos hídricos para o desenvolvimento de atividades agrícolas;

Rios e/ou Córregos: variável ambiental que caracteriza a presença de divisa do imóvel com rios ou córregos;

Açudes: variável que caracteriza a presença de açudes no imóvel

Área de mata ciliar (ha): um dos objetivos da pesquisa é a utilização de variáveis ambientais no trabalho. Por isso através da classificação das imagens RapidEye e QuickBird foi mensurada a área de mata ciliar presente em cada amostra em hectares;

Área de vegetação nativa (ha): da mesma forma que o item anterior através das imagens de satélite, foi mensurada a área de vegetação nativa, ou seja, a cobertura florística natural, que pode ter uso de reserva legal ou área de proteção ambiental;

Área de reflorestamento (ha): área em hectares de reflorestamentos;

Frente1 (m): frente do imóvel à via principal, medida em metros;

Frente2 (m): frente dois do imóvel, medida em metros, caso exista;

Profundidade1 (m): profundidade do imóvel, medida em metros;

Profundidade² (m): profundidade do imóvel, medida em metros;

Fundos (m): medida dos fundos do imóvel;

Área total (ha): área total do imóvel em hectares;

Formato: variável a qual caracteriza o formato do imóvel. Aqui a hipótese é que imóveis com formato muito irregular dificultam um bom aproveitamento do mesmo, principalmente com relação aos deslocamentos das máquinas no imóvel;

Topografia: variável que representa a topografia presente no imóvel. Nos imóveis rurais a topografia é utilizada como fator limitante nas classificações de aptidão agrícola, entre outras. Quando o imóvel possui topografia acidentada (relevo forte ondulado a montanhoso) seu potencial de uso é muito reduzido;

Situação: variável que representa a situação do imóvel com relação à sua localização no município, levando em consideração as condições do acesso até o imóvel e a distância a vias principais;

Acesso direto a via principal: variável que caracteriza se imóvel tem acesso direto à via principal, ou se para chegar ao imóvel são necessárias servidões de passagens;

Índice de Irrigação: através das plantas de classes de aptidão das terras para irrigação na área do PROJIR, foi calculado um índice de ponderação de irrigação. Segundo Amaral (2011), a classe 1 de irrigação é a classe de referência e não possui fatores limitantes. Já as classes 2 a 6 apresentam um ou mais fatores limitantes que afetam os custos de produção, por exemplo, na classe 2 a produtividade média corresponde a 90% da classe de referência. Para calcular o índice de ponderação foram utilizados estes fatores de produção de cada classe com relação à sua área correspondente, dividindo pela área total do imóvel. Este índice é menor que uma unidade para todos os imóveis;

Área Irrigável (ha): quantitativo de área que pode ser irrigada em hectares, considerando uma ponderação com relação à produtividade das classes 1 a 4 de irrigação;

Cultura: variável que caracteriza o tipo de produção vegetal predominante na região, que é cana-de-açúcar.

A variável independente Capacidade de Uso do Solo não foi utilizada nesta pesquisa como recomenda a NBR 14.653:3, norma brasileira de avaliação de imóveis rurais. Isso se deve ao fato de que esse sistema de avaliação não foi utilizado na fase de levantamento de dados do PROJIR, publicado em 1982. Ainda, os dados de cada perfil teriam que ser atualizados de acordo com o novo Sistema Brasileiro de Classificação de Solos (SANTOS et al., 2013) e então interpretados, ponto a ponto quanto a capacidade de uso para definir as novas unidades de mapeamento. O que é inexecutável, já que estes perfis não estão organizados em meio digital e estão impressos em muitos volumes, tornando a procura morosa e quase impossível de ser realizada. Por outro lado, tendo em vista que o sistema de avaliação de terras para irrigação é ainda mais criterioso para definir o potencial das terras, o seu uso em substituição a capacidade de uso das terras permite a adequada avaliação dessa variável na área do PROJIR.

A Tabela 5 apresenta o resumo (nome, sigla, tipo, descrição) das variáveis que foram testadas no processo de homogeneização por regressão.

Tabela 5. Resumo das variáveis independentes

NOME	SIGLA	TIPO	DESCRIÇÃO
Coordenada X	X	Contínua	Localização X da amostra
Coordenada Y	Y	Contínua	Localização Y da amostra
Vias de acesso*	VA	Codificada	Asfalto = 10; outros tipos de pavimento = 5; estrada de chão = 1
Trafegabilidade*	TRAF	Codificada	Permanente = 10; ruim na estação chuvosa = 5; sempre ruim = 1
Distância a polo de valorização	DPV	Contínua	Em km até o centro urbano do município de Campos dos Goytacazes
Distancia ao mar	DM	Contínua	Em km até o mar
Disponibilidade de água	DA	Dicotômica	Sim=10; Não=1
Rios e/ou Córregos	RC	Dicotômica	Divisa a rios ou córregos: Sim=10; Não=1
Açudes	AÇU	Dicotômica	Sim=10; Não=1
Área de Mata Ciliar	AMC	Contínua	Área de mata ciliar em ha
Área de Vegetação Nativa	AVN	Contínua	Área de vegetação nativa em ha
Área com Reflorestamento	AR	Contínua	Área com reflorestamento em ha
Frente1	FRE1	Contínua	Medida da frente em metros
Frente2	FRE2	Contínua	Medida da segunda frente em metros
Profundidade1	PROF1	Contínua	Medida da primeira profundida em metros
Profundidade2	PROF2	Contínua	Medida da segunda profundida em metros
Fundos	FUND	Contínua	Medida dos fundos em metros
Área Total	AT	Contínua	Área total do imóvel em hectares
Formato	FOR	Dicotômica	Regular=10; Irregular=1
Topografia*	TOPO	Codificada	Plana = 10; pouco acidentada = 5; muito acidentada = 1
Situação*	SIT	Codificada	Ótima=10; muito boa=8; boa=6; regular=4; desfavorável=2; má=1
Acesso direto a via principal	ADVP	Dicotômica	Acesso direto =10; sem acesso direto = 1
Índice de Irrigação	INIRRI	Contínua	Índice ponderado de irrigação de acordo com as classes de cada imóvel
Área Irrigável	AIRRIGA	Contínua	Área Irrigável em hectares
Cultura	CULT	Dicotômica	Cana = 10; outras culturas = 1

* As variáveis foram codificadas por livre escolha da autora, considerando um intervalo maior entre os códigos. A NBR 14.653:3 (ABNT, 2004) recomenda o uso de códigos alocados na sequência natural (1, 2, 3, ...), a qual não foi utilizada nesta pesquisa.

4.2 Análise e Preparação dos Dados

Segundo Medri (2011), antes de qualquer análise é fundamental que se proceda ao exame dos dados relativos a uma variável, seja ela qualitativa ou quantitativa. Este procedimento é importante como um primeiro contato do analista com a distribuição, além de servir, também, para avaliar a existência de possíveis valores atípicos na distribuição.

Para atender a esta recomendação, foi realizada a análise exploratória nos dados (AED) para conhecer as variáveis que seriam utilizadas na modelagem para corrigir problemas preexistentes e introduzir o modelo de regressão mais ajustado aos dados coletados.

As análises foram realizadas tanto na variável dependente como nas independentes, utilizando os recursos estatísticos como: gráficos de dispersão, correlações parciais, histogramas, gráficos *boxplot*, gráficos Quantil-Quantil, etc.

4.1.1 Variável dependente: valor unitário

Esta análise na variável dependente tem como objetivo verificar se é necessário aplicar alguma transformação na mesma. Foram testadas as transformações logarítmica e raiz quadrada e comparados os resultados com a variável sem transformação, através da utilização de histogramas, coeficientes de assimetria e curtose e gráficos *boxplot*.

Para comparar o efeito provocado pelas transformações sobre os dados originais foram construídos histogramas da variável dependente sem transformação, com a transformação logarítmica e com a transformação da raiz quadrada como pode ser visto na Figura 14, 15 e 16. A utilização de tabelas/histogramas agrupados em classes é recomendada para variáveis quantitativas contínuas, que é o caso desta variável, pois assume muitos valores, o que torna as tabelas não agrupadas muito extensas e inúteis para analisar os resultados.

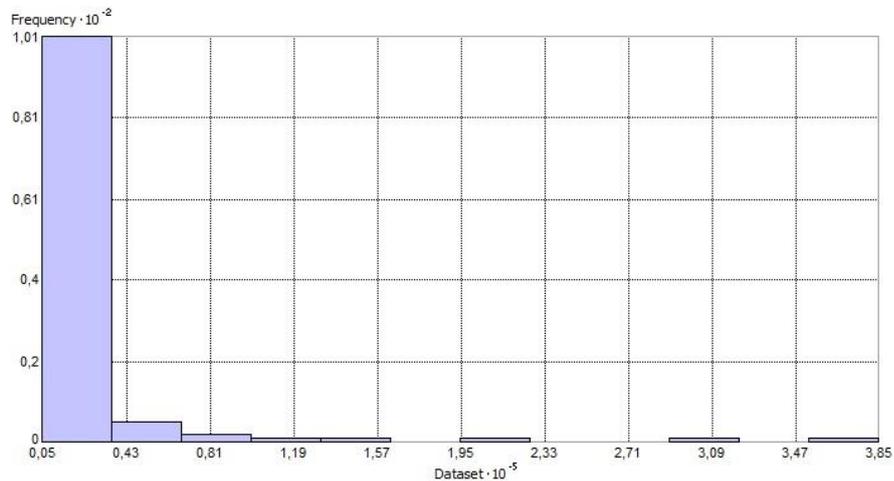


Figura 14. Histograma da variável dependente: na escala original

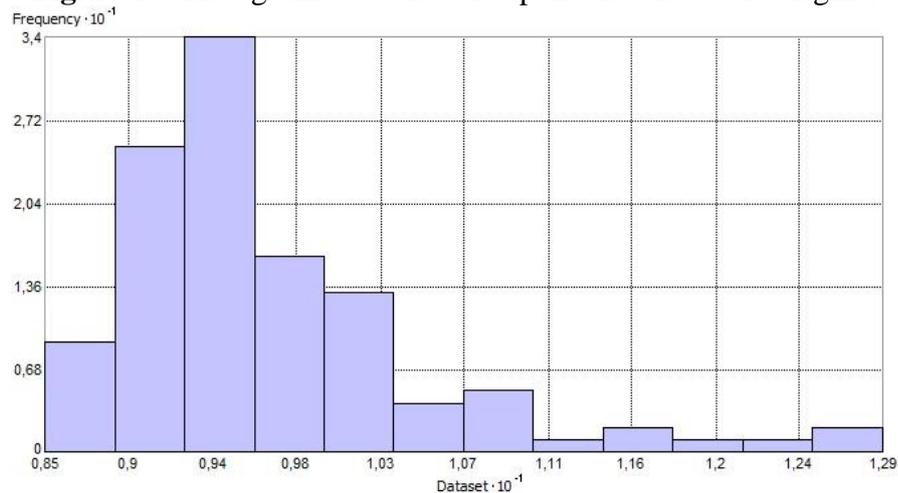


Figura 15. Histograma da variável dependente: transformação logarítmica

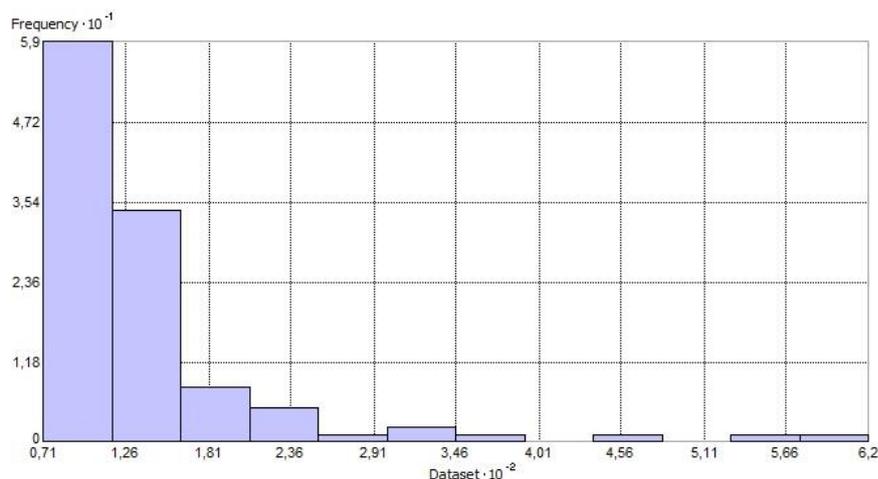


Figura 16. Histograma da variável dependente: transformação raiz quadrada

Através da análise dos histogramas podemos afirmar que a transformação logarítmica foi relativamente capaz de corrigir a forte assimetria nos dados originais. Entretanto, comparando os histogramas dos dados transformados observa-se que ainda está presente uma assimetria positiva (assimétrica para a direita). Foram então calculados os coeficientes de assimetria e curtose para as três situações em análise e os resultados podem ser observados na Tabela 6.

Tabela 6. Coeficientes de assimetria e curtose para a variável dependente

Variável Dependente	Assimetria - As	Curtose - C
VU/ha	5,218	30,568
LN VU/ha	1,740	3,704
(VU/ha) ^{0,5}	3,489	14,359

Segundo Medri (2011), para uma distribuição perfeitamente simétrica, o valor de As é zero, de modo geral, os valores As situam-se entre -3 e 3 . Se, $0,15 < As < 1$, a assimetria é considerada moderada; se $As > 1$, é forte. Para o coeficiente de curtose deve-se obter para uma curva relativamente normal, o valor de $C = 0,263$ que representa uma curva denominada “curva mesocúrtica”.

Os resultados não foram satisfatórios quando comparados ao esperado para os dois coeficientes, então será necessário a utilização dos gráficos *boxplot* para analisar se existem e quais são os *outliers* que estão afetando esta variável.

Na Figura 17 a, b, c é possível observar o gráfico *boxplot* para as três situações em análise. Nela, pode-se notar uma redução dos pontos que foram considerados *outliers* quando os dados estavam sem transformação, principalmente na transformação logarítmica, entretanto ainda existem alguns pontos que podem ser considerados discrepantes e que afetam a distribuição na parte superior dos quartis. Chama a atenção também a grande variação nos dados na escala original, pois o gráfico *boxplot* ficou achatado na parte inferior dos quartis.

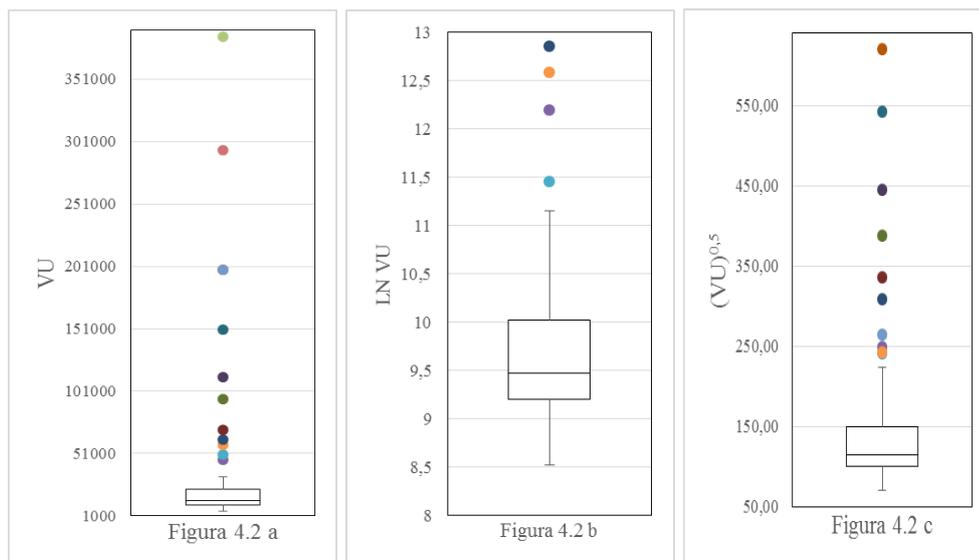


Figura 17 a, b, c. Gráficos Boxplots para a variável dependente na escala original; transformação logarítmica; transformação raiz quadrada

Na Figura 18, 19 e 20 estão representados os gráficos dos quantis teóricos da distribuição Normal padrão versus os valores observados da amostra. Neles é possível perceber como após a transformação, os dados passam a ter uma maior simetria com a linha reta que representa os quantis teóricos da distribuição Normal, sendo que a distribuição mais simétrica está na transformação logarítmica.

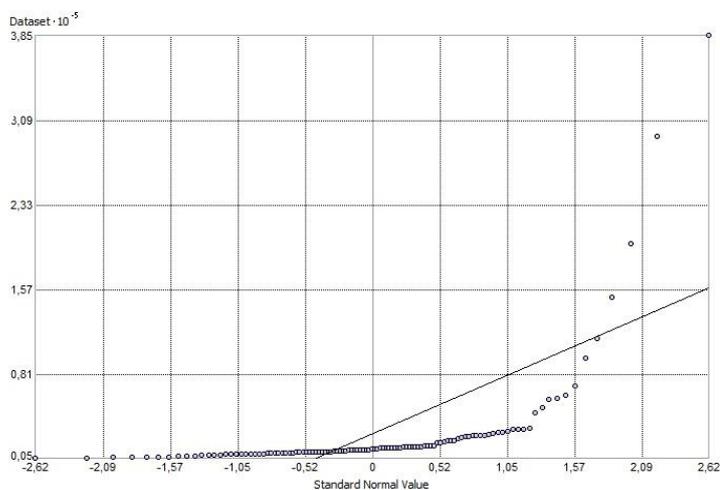


Figura 18. Gráfico Quantil-Quantil da variável dependente na escala original

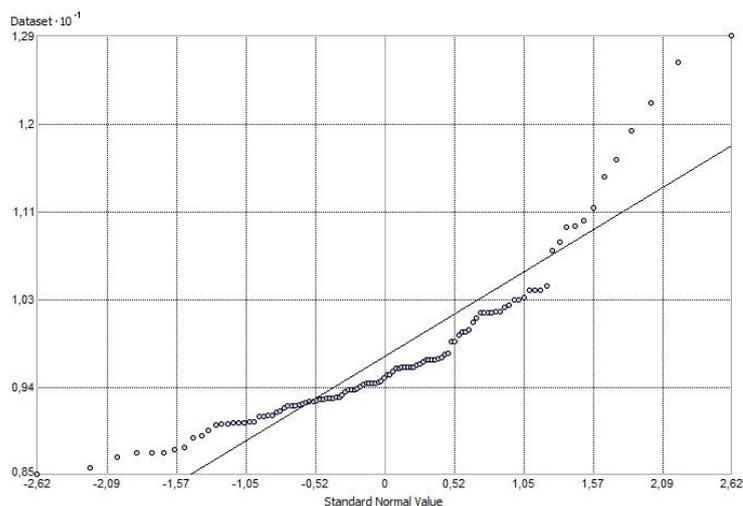


Figura 19. Gráfico Q-Q da variável dependente na transformação logarítmica

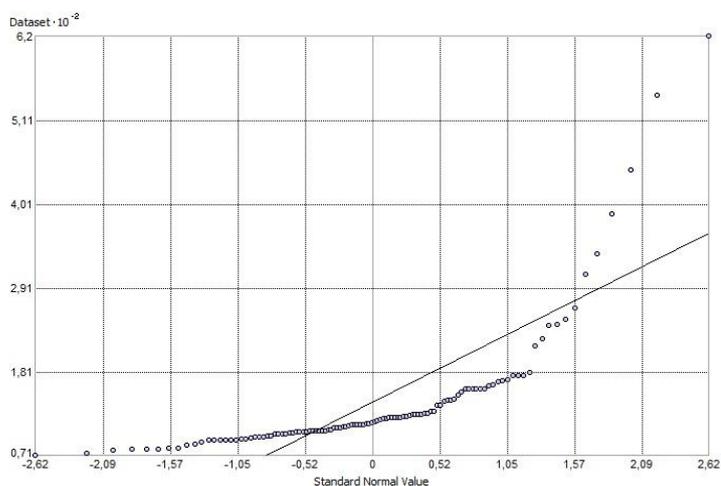


Figura 20. Gráfico Q-Q da variável dependente na transformação raiz quadrada

Para finalizar a análise da variável dependente na Tabela 7 estão os valores da estatística descritiva para as três situações: sem transformação, com transformação logarítmica e com a transformação raiz quadrada. Destaca-se que os dados sem transformação apresentam uma grande amplitude de variação. Isto acontece devido à grande diferença do valor de mercado das propriedades localizadas no entorno da área urbana de Campos dos Goytacazes, que é o município mais expressivo, quando comparados com os valores praticados nos outros municípios da região do PROJIR.

Na transformação logarítmica apesar dos valores de assimetria e curtose estarem acima do ideal, e terem pontos discrepantes, que são *outliers*, essa será a transformação utilizada na modelagem dos dados por Regressão Múltipla. Essa transformação indicada pela AED confirma o que autores como Dantas (2014) recomendam: a transformação logarítmica é a preferida quando se procura ajustar modelos a dados de valores imobiliários. É coerente sua utilização uma vez que as variáveis explicadas possuindo valores no campo dos reais positivos garantem que o campo de variação dos valores ajustados correspondentes também serão reais positivos. Outro aspecto importante é que a transformação logarítmica na variável explicada torna o modelo multiplicativo, bem como pode estabilizar a variância do modelo, característica sugerida pelas normas brasileiras de avaliações.

Tabela 7. Estatística descritiva para a variável dependente sem transformação, transformação logarítmica e transformação raiz quadrada

	VU/ha	LN VU/ha	(VU/ha) ^{0,5}
Mínimo (R\$/ha)	5.036,00	8,52	70,96
1º Quartil (R\$/ha)	9.933,00	9,20	99,66
Mediana (R\$/ha)	12.981,00	9,47	113,93
Média (R\$/ha)	26.736,68	9,68	140,13
3º Quartil (R\$/ha)	22.500,00	10,02	150,00
Máximo (R\$/ha)	385.000,00	12,86	620,48
Amplitude (R\$/ha)	379.964,00	4,34	549,52
Assimetria	5,218	1,740	3,489
Curtose	30,568	3,704	14,359
Desvio padrão (R\$/ha)	50.240,71	0,798	84,62
Variância da amostra	2.524.129.716,00	0,637	7.161,81
Coefficiente de variação (%)	187,90	8,24	60,38

Os dados que apresentaram os maiores valores discrepantes para a transformação logarítmica são 3, 31, 38 e 43 que coincidem com os maiores valores por hectare da amostra. Estas amostras são de imóveis com áreas pequenas e próximos a área urbana de Campos de Goytacazes.

4.1.2 Variáveis independentes

O estudo das transformações também foi realizado para as variáveis independentes, buscando sempre a coerência nas transformações e também para o modelo obtido, para que o modelo possa explicar com fidelidade a realidade em estudo.

A primeira análise realizada foi a construção da matriz de correlações parciais, pela observação do coeficiente de correlação simples entre as variáveis independentes, consideradas duas a duas. A multicolinearidade surge quando algumas ou todas as variáveis independentes de um modelo, estão de tal forma correlacionadas umas às outras, tornando-se difícil isolar suas influências separadamente. Os valores críticos adotados na análise foram maiores que 0,60, apesar da NBR 14.653:3 (ABNT, 2004) determinar atenção especial para resultados superiores a 0,80.

Na Tabela 8 encontra-se a matriz de correlações para as variáveis estudadas.

Tabela 8. Matriz de correlações parciais para as variáveis independentes

	VA	TRAF	DPV	DM	DA	RC	AÇU	AMC	AVN	AR	FRE1	FRE2	PROF1	PROF2	FUND	AT
VA	1															
TRAF	0,921	1														
DPV	-0,287	-0,288	1													
DM	0,241	0,231	-0,668	1												
DA	-0,035	0,006	0,291	-0,262	1											
RC	0,087	0,080	0,136	-0,130	0,716	1										
AÇU	-0,058	-0,020	0,070	-0,103	0,276	0,243	1									
AMC	-0,165	-0,133	0,161	0,035	-0,082	-0,085	0,064	1								
AVN	0,074	0,069	0,260	-0,064	0,041	0,087	-0,041	-0,067	1							
AR	0,159	0,145	-0,013	-0,011	-0,020	0,018	0,201	0,019	-0,038	1						
FRE1	0,130	0,132	0,042	0,071	0,218	0,189	0,280	0,037	0,585	0,032	1					
FRE2	0,108	0,107	0,087	-0,039	0,180	0,118	-0,037	0,004	0,532	0,041	0,452	1				
PROF1	0,099	0,118	-0,070	0,197	0,170	0,197	-0,037	0,198	0,143	-0,008	0,295	0,193	1			
PROF2	0,128	0,149	-0,109	0,213	0,202	0,199	-0,029	0,046	0,196	-0,009	0,282	0,227	0,908	1		
FUND	0,145	0,101	-0,019	0,035	0,252	0,247	0,143	0,137	0,380	0,084	0,643	0,526	0,255	0,222	1	
AT	0,184	0,178	-0,068	0,106	0,212	0,220	0,097	0,127	0,407	0,033	0,695	0,529	0,610	0,571	0,793	1
FOR	0,004	-0,005	0,156	-0,093	-0,004	-0,038	-0,091	-0,177	-0,068	-0,077	-0,287	-0,216	-0,106	-0,059	-0,395	-0,311
TOPO	0,244	0,218	-0,345	-0,063	-0,055	0,061	-0,181	-0,196	-0,091	0,065	-0,035	0,023	-0,190	-0,196	0,081	0,028
SIT	0,586	0,615	-0,491	0,312	-0,025	0,051	-0,052	-0,186	0,017	0,005	0,128	0,007	0,126	0,176	0,130	0,212
ADVP	0,818	0,756	-0,228	0,193	0,084	0,172	-0,019	-0,167	0,105	-0,095	0,208	0,182	0,177	0,224	0,222	0,265
INIRRI	0,317	0,294	-0,285	0,135	-0,129	-0,060	-0,010	0,035	-0,070	-0,029	-0,007	0,047	0,115	0,099	0,045	0,138
AIRRIGA	0,251	0,247	-0,132	0,124	0,191	0,222	0,077	0,053	0,338	0,026	0,625	0,492	0,559	0,526	0,699	0,951
CULT	0,149	0,172	-0,243	0,170	-0,123	-0,112	-0,014	-0,067	0,024	0,044	0,136	0,150	0,204	0,203	0,144	0,205

Continuação Tabela 8:

	FOR	TOPO	SIT	ADVP	INIRRI	AIRRIGA	CULT
VA							
TRAF							
DPV							
DM							
DA							
RC							
AÇU							
AMC							
AVN							
AR							
FRE1							
FRE2							
PROF1							
PROF2							
FUND							
AT							
FOR	1						
TOPO	-0,039	1					
SIT	0,028	0,204	1				
ADVP	-0,011	0,158	0,590	1			
INIRRI	-0,010	0,275	0,459	0,312	1		
AIRRIGA	-0,268	0,110	0,279	0,322	0,283	1	
CULT	-0,126	-0,048	0,159	0,212	0,250	0,241	1

VA – vias de acesso;
 TRAF – trafegabilidade;
 DPV - Distância a polo de valorização;
 DM – distância ao mar;
 DA – disponibilidade de água;
 RC – rios e córregos;
 AÇU – açudes;
 AMC – área de mata ciliar;
 AVN – área de vegetação nativa;
 AR – área com reflorestamento;
 FRE1 – frente 1;
 FRE2 – frente2;
 PROF1 – profundidade 1
 PROF2 – profundidade 2;
 FUND – fundos;
 AT – área total;
 FOR – formato;
 TOPO – topografia;
 SIT – situação;
 ADVP – acesso direto a via principal;
 INIRRI – índice de irrigação;
 AIRRIGA – área irrigável;
 CULT – cultura.

Pode-se observar na Tabela 8, que as variáveis vias de acesso (VA) e trafegabilidade (TRAF) estão fortemente correlacionadas, provavelmente porque quando as vias de acesso são por asfalto a trafegabilidade assume valor 10 (sem problema de trafegabilidade). A variável disponibilidade de água (DA) está correlacionada com a variável rios e/ou córregos (RC), pois quando existe disponibilidade o imóvel tem um dos seus limites em rios ou córregos.

As variáveis que representam as dimensões dos imóveis também estão correlacionadas. São as variáveis frente1 (FRE1) com fundos (FUND) e com área total (AT); as variáveis profundidade2 (PROF2) com área total (AT); as variáveis profundidade1 (PROF1) com área total (AT) e as variáveis fundos (FUND) com área total (AT). As variáveis acesso direto a via principal (ADVP) estão fortemente correlacionadas com as variáveis vias de acesso (VA) e com a variável trafegabilidade (TRAF), porque quando o imóvel tem acesso direto a via principal também tem suas vias de acesso por asfalto e trafegabilidade sem problemas.

As variáveis área total (AT) e área irrigável (AIRRIGA) apresentam uma correlação muito grande, pois são duas variáveis que representam a mesma grandeza. Apesar disso, as mesmas continuarão na análise, por serem consideradas muito importantes para a avaliação.

Observa-se que as variáveis trafegabilidade (TRAF), disponibilidade de água (DA), frente1 (FRE1), profundidade1 (PROF1); profundidade2 (PROF2), fundos (FUND) e acesso direto a via principal (ADVP) que apresentaram multicolinearidade não serão usadas na modelagem de regressão múltipla, com exceção da área total (AT) e área irrigável (AIRRIGA).

O passo seguinte foi a análise dos gráficos de dispersão (ANEXO 1) entre a variável dependente (VU) e as variáveis independentes nas suas escalas originais, ou seja, sem transformações. Esta análise é importante para verificar a relação dos valores das terras em relação às variáveis independentes fornecendo uma ideia do tipo de relacionamento entre as duas variáveis.

Analisando a dispersão dos pontos nos gráficos, observou-se que a maioria das variáveis independentes apresentam uma tendência não-linear, portanto durante a modelagem será necessário utilizar alguma transformação que linearize os dados para obter um modelo mais coerente com a realidade do mercado.

Analisando-se os gráficos das variáveis Área de Mata Ciliar (AMC), Área de Vegetação Nativa (AVN) e Área de Reflorestamento (AR) observa-se que existem amostras com valores de hectare mais altos, que não possuem áreas com estas variáveis ambientais (igual a zero).

Com relação a variável Índice de Irrigação (INIRRI) as amostras com os maiores valores de hectare são as que têm os maiores índices de irrigação, o que evidencia o quanto a avaliação do potencial de uso agrícola das terras, nesse caso avaliado através da aptidão de irrigação é importante na valorização dos imóveis rurais.

Observa-se também que as amostras com os maiores valores de hectare são as que possuem a relação com as seguintes variáveis independentes: Vias de Acesso (VA) asfalto, têm as menores Distância a Polo de Valorização (DPV), Topografia (TOPO) plana, Situação (SIT) ótima e têm Acesso Direto à Via Principal (ADVP).

4.3 Modelo Clássico de Regressão Linear - MCRL

Nesta etapa da pesquisa foram realizadas várias interações com as variáveis independentes, como a utilização sem transformação, com as transformações logarítmica,

inversa, raiz quadrada e potência, entre outras com o objetivo de encontrar o modelo de regressão múltipla que melhor representa o mercado imobiliário na região do PROJIR. Com relação à variável dependente (VU) foi testada a transformação logarítmica indicada na análise exploratória (item 4.2.1).

Santos (2014) utilizou dois modelos de regressão: um modelo com as variáveis independentes de localização (coordenadas e distâncias a polos de valorização) além das variáveis relacionadas às características dos imóveis e outro modelo somente com as variáveis relacionadas às características dos imóveis, sem as variáveis de localização espacial.

Trivelloni (2005), utilizou no modelo de regressão somente as características relacionadas aos imóveis, pois o objetivo do autor foi determinar o valor da localização nas avaliações em massa.

Como nessa pesquisa se busca determinar qual método estatístico tem o melhor desempenho nas avaliações em massa, foi utilizado no modelo de regressão clássico somente as variáveis relacionadas às características intrínsecas dos imóveis, e verificar a existência ou não da dependência espacial na regressão espacial e na regressão geograficamente ponderada - RGP.

O modelo hedônico geral, considerando o logaritmo natural do valor unitário, foi representado através da seguinte fórmula:

$$\ln VU = f(VA, TRAF, DVP, DM, DA, RC, AÇU, AMC, AVN, AR, FRE1, FRE2, PROF1, PROF2, FUND, AT, FOR, TOPO, SIT, ADVP, INIRRI, AIRRIGA, CULT)$$

O MCRL após várias interações e testes de transformações nas variáveis, com a eliminação de 12 amostras consideradas *outliers* e utilizando o *software* INFER 32, a equação de regressão, considerando o logaritmo do valor unitário como variável dependente do modelo, ficou da seguinte forma:

$$\ln[VU] = 8,9763 + 0,026953 * [VA] + 1,7415 [INV_DPV] + 2,0180 [INV_DM] + 0,026692 * [AÇU] + 0,10172 * [POT_AVN] - 0,11167 x [LN_AT] + 0,025745 * [TOPO] + 0,11643 * [SIT] + 0,00036128 * [AIRRIGA] - 0,03879 x [CULT]$$

O modelo para a variável dependente – VU é o seguinte:

$$[VU] = \text{EXP} (8,9763 + 0,026953 * [VA] + 1,7415 [INV_DPV] + 2,0180 [INV_DM] + 0,026692 * [AÇU] + 0,10172 * [POT_AVN] - 0,11167 x [LN_AT] + 0,025745 * [TOPO] + 0,11643 * [SIT] + 0,00036128 * [AIRRIGA] - 0,03879 x [CULT])$$

As variáveis influenciadoras e suas respectivas transformações estão descritas a seguir:
 VU – Valor Unitário (R\$/ha), onde foi aplicado o logaritmo natural;
 VA – Vias de acesso, sem transformação;
 INV_DPV – Distância a Polo de Valorização (km), onde foi aplicado a transformação inversa;
 INV_DM – Distância ao Mar (km), onde foi aplicado a transformação inversa;
 AÇU – Açudes, sem transformação;
 POT_AVN – Área de Vegetação Nativa (ha), onde foi elevado ao expoente 0,143;
 LN_AT – Área Total (ha) onde foi aplicado o logaritmo natural;
 TOPO – Topografia, sem transformação;
 SIT – Situação do Imóvel, sem transformação;

AIRRIGA – Área Irrigável, sem transformação;
 CULT – Cultura, sem transformação.

Em condições *ceteris paribus*, tomando antilogaritmos da equação de regressão clássica, a interpretação do modelo é mostrada a seguir:

$$VU = 7.913,331 * 1,02731^{VA} * 5,70584^{INV_DPV} * 7,52362^{INV_DM} * 1,02705^{AÇU} * 1,10707^{POT_AVN} * 0,894349^{LN_AT} * 1,02608^{TOPO} * 1,12348^{SIT} * 1,00036^{AIRRIGA} * 0,96195^{CULT}$$

4.3.1 Coerência e significância dos regressores

Houve a transformação logarítmica da variável dependente bem como na variável independente Área Total do Imóvel, nas variáveis Distância a Polo de Valorização. Na variável Distância ao Mar foi aplicada a transformação inversa, pois os valores de hectare são maiores quando esta variável possui menores distâncias. Na variável Área de Vegetação Nativa foi usada a transformação potência com o valor do expoente igual a 0,143. As demais permaneceram nas escalas originais. Estas transformações são necessárias para atender a linearidade da regressão, sendo que as mesmas não devem deixar o modelo incoerente.

Na Tabela 9 pode-se observar um resumo com os valores do erro-padrão, t calculado, o valor p e os intervalos de confiança de cada regressor. Todos os coeficientes foram significativos, sendo considerado o t tabelado ou crítico, unicaudal com significância de 10% igual a 1,291. Comparando-se o valor t de cada regressor com o t crítico, todos os regressores são significativos, pois tcalc em módulo é maior que o ttab.

Tabela 9. Estimativas dos parâmetros, erro-padrão, valor t e o valor p do modelo

	Coefficientes	Erro padrão	Valor t	valor-p	IC - 95% inf.	IC - 95% sup.
Interseção	8,97630	0,1902	47,1928	0,0000	8,59842	9,35417
VA	0,02695	0,0092	2,9222	0,0043	0,00862	0,04527
INV_DPV	1,74149	1,2498	1,3933	0,1669	-0,74151	4,22449
INV_DM	2,01804	0,8409	2,3997	0,0184	0,347342	3,68875
AÇU	0,02669	0,0097	2,7412	0,0073	0,007347	0,04603
POT_AVN	0,10171	0,0623	1,6310	0,1063	-0,02217	0,22560
LN_AT	-0,11166	0,0287	-3,8902	0,0001	-0,16869	-0,05464
TOPO	0,02574	0,0100	2,5624	0,0125	0,005785	0,045705
SIT	0,11642	0,0253	4,5902	0,0000	0,066038	0,166820
AIRRIGA	0,00036	0,0001	2,0242	0,0459	6,7022x10 ⁻⁶	0,000715
CULT	-0,03878	0,0079	-4,8907	0,0000	-0,05454	-0,02303

Analisando os sinais e as transformações podemos concluir que está coerente, pois as suposições formuladas no item 4.1 quanto as influências das variáveis foram confirmadas, com uma única exceção para a variável Cultura, pois esperava-se que os imóveis cultivados com cana-de-açúcar seriam mais valorizados, o que não foi confirmado pelo modelo de regressão. Esta variável mesmo assim deve continuar no modelo matemático.

De acordo com os coeficientes do modelo MCRL, o valor médio do hectare na região do PROJIR é de R\$ 7.013,33. O valor unitário aumenta quando o pavimento das vias de

acesso são asfalto com relação a via não pavimentada. O valor unitário diminui quando a distância ao centro de Campos de Goytacazes aumenta, sendo que a taxa de diminuição não é constante. O valor unitário diminui quando a distância ao mar aumenta. O valor unitário aumenta quando na propriedade existem açudes e áreas de vegetação nativa, diminui com o aumento da área, aumenta quando a topografia é plana e quando a sua situação em termos de localização no município é ótima, aumenta quando a área irrigada aumenta e diminui quando o produto cultivado no imóvel é a cana-de-açúcar.

A valorização que as variáveis ambientais exerceram no modelo, se relacionam com as mudanças na legislação florestal, que exige a preservação e recomposição das áreas de florestas e matas nativas. A criação do Cadastro Ambiental Rural, além de outros programas de incentivo à preservação do meio ambiente, está gerando mudanças no comportamento dos proprietários dos imóveis rurais. O modelo mostrou que a existência dessas áreas nas propriedades, valorizam financeiramente o imóvel rural.

Para a variável cultura – CULT, onde foi analisado a influência que a cana-de-açúcar exerce nos imóveis, o modelo mostrou que o valor do imóvel não aumenta quando este produto agrícola é cultivado. A cultura da cana-de-açúcar não se mostrou importante, pois as usinas implantadas na região Norte Fluminense, não realizaram investimentos na modernização tecnológica, nem na modernização empresarial (empresas familiares). Além disso, não acompanharam as melhorias introduzidas nas lavouras, como mudas de cana mais produtivas e resistentes a pragas, etc. (SOUZA et al., 2009). Diferente do que ocorreu em outras regiões do Brasil, que cultivam o mesmo produto. Também é importante salientar que a mão de obra escassa para esta cultura, também contribuiu para o declínio, pois a escassez ficou ainda mais evidente com a mudança desta atividade para a exploração de petróleo e gás natural, que exigiu pessoas mais qualificadas (FORTI, 2013).

4.3.2 Correlação e variância do modelo

Os resultados gerais do modelo são mostrados na Tabela 10. Este contou com 10 variáveis independentes e apresentou um coeficiente de correlação R de 0,8153. Já o valor do coeficiente de determinação R^2 evidencia que a equação encontrada explica 66,46% dos desvios, sendo 33,84% atribuídos a fatores aleatórios e a variáveis não consideradas no modelo.

Tabela 10. Resultados gerais do modelo de regressão

Resultados	Valor
Nº de elementos da amostra	101
Nº de variáveis independentes	10
Desvio padrão da regressão	0,3227
Coeficiente de correlação (R)	0,8153
Valor t calculado	13,36
Valor t tabelado (t crítico)	2,632 (para o nível de significância de 1,00 %)
Coeficiente de determinação (R^2)	0,6646
Coeficiente R^2 ajustado	0,6274

A análise da variância, também chamada ANOVA, foi realizada da forma indicada na Tabela 11, utilizando-se a distribuição de Snedecor. O F_{Tab} é igual a 2,524 para o nível de

significância de 1,0 %. Considerando que o F_{Calc} foi igual a 17,84 pode-se concluir que existe regressão e a significância do modelo é igual a $1,9 \times 10^{-15}\%$.

Tabela 11. Análise de variância - ANOVA

Fonte de erro	Soma dos quadrados	Graus de liberdade	Quadrados médios	F calculado
Regressão	18,5856	10	1,8585	17,84
Residual	9,3774	90	0,1041	
Total	27,9630	100	0,2796	

4.3.3 Análise da multicolinearidade

A presença da multicolinearidade pode ser constatada, na matriz de correlações pela observação do coeficiente de correlação simples entre as variáveis independentes, consideradas duas a duas. Pode-se observar na Tabela 12, que apenas com as variáveis AT e AIRRIGA, o coeficiente de correlação ficou maior que o limite adotado neste trabalho de 0,70.

Tabela 12. Matriz de correlações parciais

	VA	INV_ DPV	INV_ DM	AÇU	POT_ AVN	LN_ AT	TOPO	SIT	AIR RIGA	CULT
VA	1									
INV_ DPV	0,2889	1								
INV_ DM	-0,2460	-0,4493	1							
AÇU	-0,0614	-0,0364	0,0511	1						
POT_ AVN	0,1097	-0,0496	-0,0604	-0,0409	1					
LN_ AT	0,1012	-0,0381	-0,0697	0,1377	0,2288	1				
TOPO	0,1877	0,2624	-0,0085	-0,1937	-0,1230	-0,1526	1			
SIT	0,5610	0,6318	-0,3113	-0,0774	0,1098	0,1393	0,1340	1		
AIRRIGA	0,2964	0,1666	-0,1609	0,06967	0,2281	0,6528	0,1411	0,3569	1	
CULT	0,1640	0,1718	-0,0769	-0,0315	-0,0065	0,2101	-0,0205	0,1651	0,2571	1

Os resultados da matriz de correlação parcial mostram a importância da análise exploratória dos dados, pois evitou-se a utilização de variáveis correlacionadas entre si. Com a exceção do valor entre AT e AIRRIGA as outras variáveis obtiveram valores abaixo do valor crítico adotado.

4.3.4 Análise dos resíduos

No gráfico *Box plot* elaborado para a análise da variável dependente foram detectados alguns pontos que poderiam ser *outliers*, que são os dados 3, 31, 38 e 43. Um dos pressupostos da MCRL é que a presença de *outliers* deve ser analisada e estes devem ser retirados da amostra. A verificação de existência ou não de *outliers* é feita plotando-se os resíduos padronizados (ei) contra os valores estimados pela regressão. Considera-se elementos

suspeitos de serem *outliers*, os pontos que apresentarem resíduo padronizado superior a 2 (dois) em módulo. Os dados 1, 27, 30, 31, 38, 40, 43, 45, 50, 54, 69 e 87 foram detectados como pontos atípicos e retirados da amostra, ficando com 101 dados, que podem ser visualizados no ANEXO 2. Observa-se que os dados indicados pelo *Box plot* estão entre os que foram detectados como *outliers*, com exceção do dado 03.

O modelo de regressão encontrado foi determinado sem estes elementos. Pode-se observar pelo gráfico da Figura 21 que ainda existem pontos com resíduo padronizado superior a dois (dados 3, 51 e 84), entretanto optou-se por deixar estes dados na amostra, pois ficaram muito próximos do valor limite.

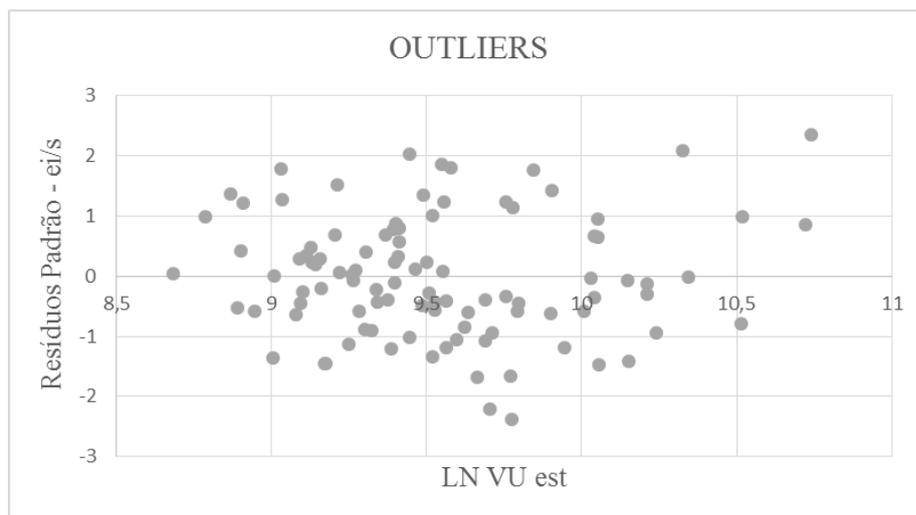


Figura 21. Gráfico de Indicação de outliers

A normalidade dos resíduos pode ser verificada comparando-se as frequências acumuladas dos resíduos padronizados observados na amostra, com as porcentagens esperadas para uma distribuição normal. Também pode ser verificado pelo histograma dos resíduos padronizados (ei/s , onde s é o desvio-padrão dos resíduos) ou pelo gráfico *Box plot*.

Pode-se concluir pela Tabela 13 que os resíduos possuem distribuição normal, pois os valores das porcentagens são muito próximos dos esperados. O histograma e *Box plot* (Figura 22) confirmam esta mesma afirmação, pois apresentam a distribuição simétrica no histograma e no *Box plot*, evidenciando que todos os valores estão dentro da distância interquartílica.

Tabela 13. Distribuição dos Resíduos Normalizados

Intervalo	Distribuição de Gauss	% de Resíduos no Intervalo
-1; +1	68,3 %	69,31 %
-1,64; +1,64	89,9 %	91,09 %
-1,96; +1,96	95,0 %	96,04 %

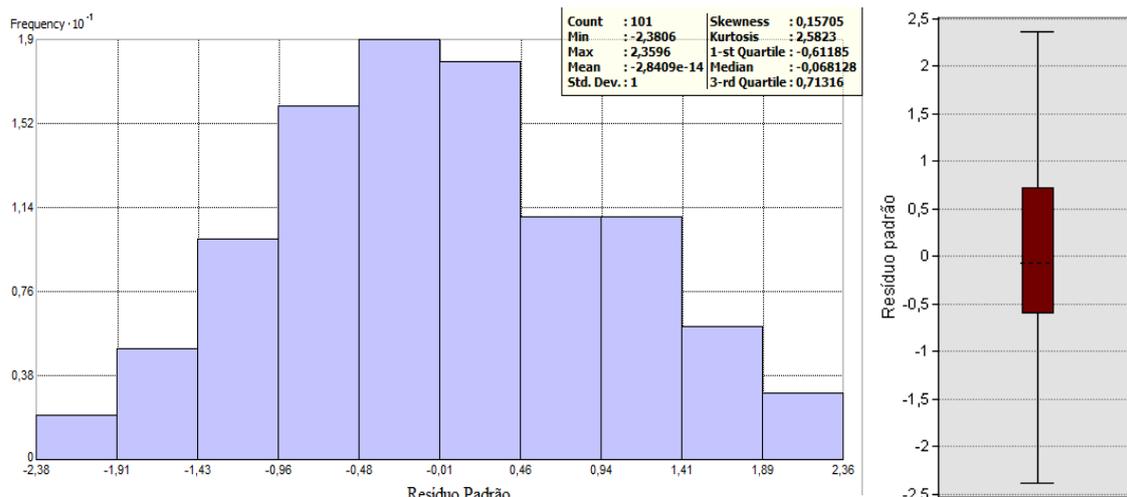


Figura 22. Histograma e Box plot dos Resíduos Padronizados

Outra forma de verificar a normalidade dos resíduos é pelo teste de Jarque-Bera, onde a hipótese nula é de que existe normalidade e a alternativa é de que não existe. O teste rejeita normalidade quando o valor for maior que 6,0 (com nível de significância igual a 5% e graus de liberdade igual a 2). Pelo teste de Jarque-Bera (valor = 1,1494 e p-valor = 0,56287), aceita-se a hipótese nula da existência de normalidade dos resíduos.

Um modelo é homocedástico quando os seus erros não explicados (resíduos) possuem variância constante e a verificação da homocedasticidade pode ser feita graficamente, plotando-se os resíduos versus os valores estimados pela reta de regressão. Na Figura 23 pode-se observar que o gráfico apresenta uma distribuição dos resíduos com pequena tendência, principalmente na parte inferior, indicando uma possível heterocedasticidade.

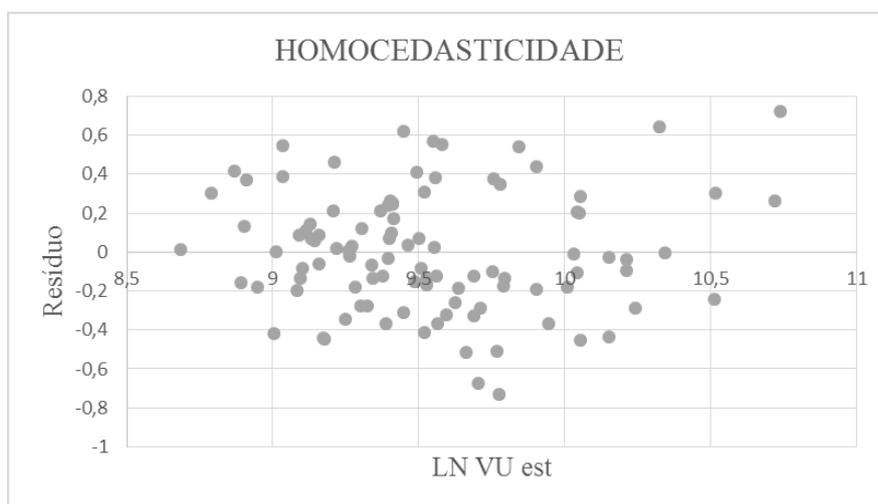


Figura 23. Gráfico da Homocedasticidade

O teste de Breush-Pagan considera como hipótese nula H_0 : a homocedasticidade, e a hipótese alternativa H_1 : heterocedasticidade, a um nível de significância de 5%. O teste de Breush-Pagan (valor = 15,8611 e p-valor = 0,10368) demonstrou que o modelo é homocedástico, pois a significância foi maior de 5%.

Entretanto o resultado do teste de Koenker-Bassett contraria o anterior (valor =

20,0481 e p-valor = 0,02880) a um nível de significância de 5%, aceita-se a hipótese de heterocedasticidade, ou seja, o modelo apresenta problemas de variância dos erros não constante. A heterocedasticidade observada no modelo pode ser provocada pela presença de dependência espacial nos resíduos do modelo.

Podemos concluir neste item que seguindo somente os pressupostos que determinam a aprovação ou não de um MCRL que o modelo está aprovado, pois analisamos todos os testes exigidos pelas normas brasileiras, com exceção da homocedasticidade. Nos próximos itens se verificou a existência da dependência espacial através da regressão espacial, e para que isso seja possível a amostra foi reenumerada, depois da retirada dos *outliers*, para evitar falhas nos programas GeoDa e ArcGIS.

4.4 Análise da Regressão Espacial

Para iniciar a análise da regressão espacial deve-se estabelecer a representação matemática que melhor representa o arranjo espacial dos dados na região. Isto pode ser conseguido testando os resíduos para autocorrelação espacial por intermédio do valor da estatística I de Moran, usando a matriz de ponderação ou matriz de pesos espaciais. O objetivo é que esta matriz capture a autocorrelação espacial presente nos resíduos do modelo de regressão linear determinada pelo MQO.

4.4.1 Determinação da matriz de pesos espaciais

A recomendação da literatura é que devem ser testadas diversas táticas na construção da matriz de pesos espaciais, também chamada de matriz de vizinhança, entre matriz de continuidade, distância geográfica, distância inversa, vizinhos mais próximos, tempo de viagem, etc. No caso de avaliação de imóveis a matriz baseada na contiguidade não se aplica, pois, as amostras estão separadas geograficamente, não possuindo limites em comum.

Alguns autores utilizaram a informação do alcance fornecida pelo ajuste de um modelo teórico ao semivariograma para criar matrizes de pesos espaciais baseadas na distância que consigam capturar a autocorrelação existente nos dados. Entretanto Hornburg e Hochheim (2009) recomendam cuidado quando da utilização da distância indicada pelo semivariograma pois nem sempre a distância obtida vai ser a melhor distância para a matriz de vizinhança.

Santos (2014) ajustou um modelo teórico ao semivariograma experimental para definir o alcance do semivariograma e utilizou esta distância para verificar a dependência espacial. O autor constatou que a distância determinada pelo semivariograma não capturou a maior parte da autocorrelação espacial, reforçando, portanto, a afirmação de Hornburg e Hochheim (2009).

Uma das dificuldades na criação das matrizes é a definição de qual a melhor distância ou quantos vizinhos mais próximos devem ser utilizados, para capturar a maior parte da autocorrelação espacial. A recomendação é testar várias matrizes para ter certeza de qual delas é capaz de capturar a maior parte da autocorrelação espacial.

Para atingir este objetivo foram criadas várias matrizes de pesos espaciais com base nas duas estratégias: distância e vizinhança. Em seguida foi realizado o teste I de Moran global para os resíduos do valor dos imóveis na região de estudo. A matriz que capturar a maior parte da autocorrelação espacial presente nos dados é a que apresenta o maior valor para o I de Moran Global. Essa estatística varia entre -1 e 1 , sendo que valores próximos de zero indicam inexistência de autocorrelação espacial significativa, e quanto mais próximo do valor unitário, mais auto correlacionados estão os dados.

Os resultados do teste I de Moran Global para as diferentes configurações de matrizes de pesos testadas estão representadas na Tabela 14 e foram calculadas com o *software* GeoDa. Todas as matrizes baseadas no número de vizinhos, onde foram testados de 2 a 15 vizinhos (a esquerda), obtiveram o valor de I de Moran menores que 0,1 demonstrando que este tipo de matriz não capturou autocorrelação espacial.

Os resultados das matrizes baseadas na distância Euclidiana a partir dos centroides, se mostraram mais significativos, sendo que a distância de 7 km foi a matriz que capturou o maior valor de I de Moran Global (0,18614). O p-valor se mostrou significativo com menos de 10% de significância para esta matriz de 7 km. No gráfico da Figura 24 pode-se observar os valores das matrizes da distância e o valor I de Moran.

Tabela 14. Teste I de Moran e significância para as matrizes de ponderação espacial testadas sobre os resíduos

<i>k</i> vizinhos	I de Moran	valor – p	Distância	I de Moran	valor – p
2	0,081502	0,12	3 km	0,09042	0,16
3	0,0536	0,24	4 km	0,12231	0,05
4	0,081502	0,09	5 km	0,07221	0,27
5	0,06049	0,19	6 km	0,09349	0,12
8	0,05924	0,06	6,5 km	0,10799	0,10
10	0,01590	0,21	7 km	0,18614	0,02
12	0,04461	0,08	7,3 km	0,17409	0,01
15	0,02425	0,15	7,5 km	0,16325	0,03
			8 km	0,15879	0,03
			9 km	0,09414	0,11
			10 km	0,06330	0,11
			11 km	0,04146	0,18
			13 km	0,05663	0,10
			15 km	0,04792	0,05
			20 km	0,01354	0,18

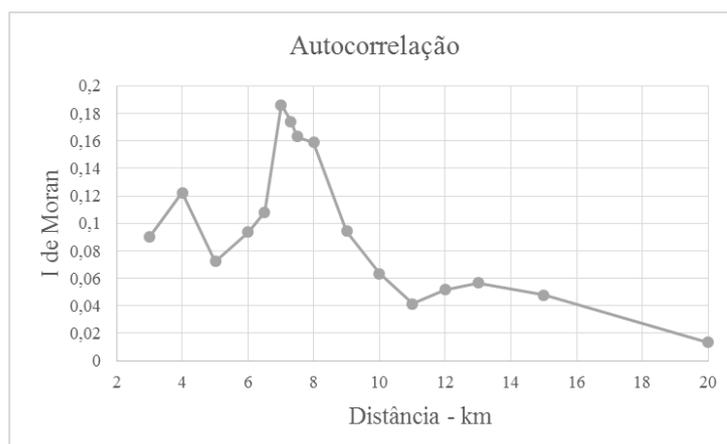


Figura 24. Gráfico da Autocorrelação

Como as matrizes da distância foram as que apresentaram a autocorrelação espacial, foram testadas também as matrizes baseadas na distância inversa e na distância inversa ao quadrado. Entretanto os resultados destes novos testes não se mostraram significativos, com valores de I de Moran muito próximos a zero, que indica inexistência de autocorrelação

espacial significativa.

O diagrama de dispersão de Moran é uma forma de visualizar a dependência espacial e indicar os diferentes padrões espaciais presentes nos dados. O gráfico representa quatro quadrantes AA, BB, AB e BA que irão corresponder a quatro padrões de associação local espacial entre as regiões e seus vizinhos.

O diagrama de dispersão de Moran para o VU previsto, gerado no programa Geoda, pode ser visualizado na Figura 25. Pode-se observar que os dados 2, 36 e 48 no quadrante AA que é o quadrante de alto valor, podem ser considerados como pontos de alavanca (pelo Geoda), ou seja, são pontos que podem influenciar fortemente a estimativa dos coeficientes. O Geoda indicou também a existência de *outliers* espaciais de alto valor (AB) (dados 4 e 66) e de baixo valor (BA) (dados 6 e 42).

Apesar da possibilidade de que estes *outliers* e o ponto de alavanca influenciem as estimativas, eles foram mantidos no conjunto dos dados para o prosseguimento das análises, pois somente o dado 42 (antigo número 51) foi identificado com esta característica (*outlier*) na análise dos resíduos da regressão múltipla.

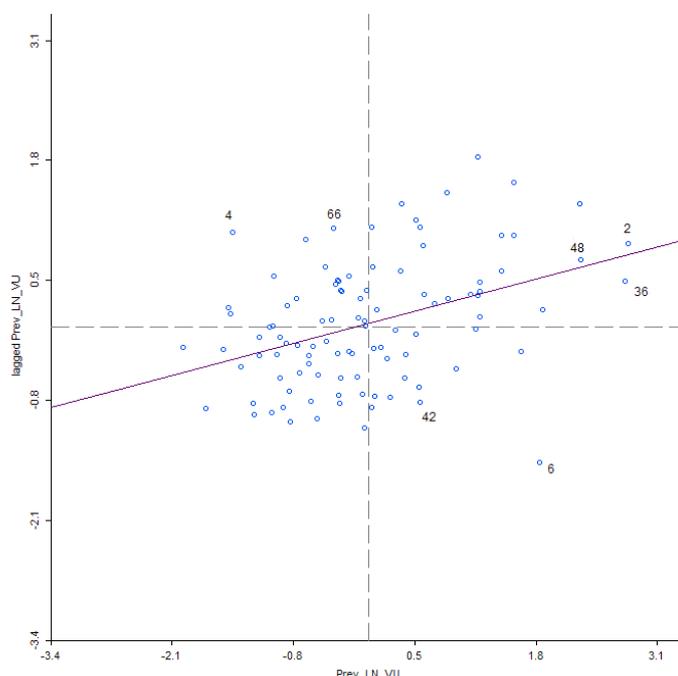


Figura 25. Diagrama de dispersão de Moran para os valores previstos do VU

4.4.2 Testes focados do Multiplicador de Lagrange

O resultado do I de Moran mostrou que existe autocorrelação espacial nos resíduos do modelo de regressão múltipla, entretanto segundo Almeida (2012), o principal problema deste teste, devido a sua natureza difusa, é que apesar de significativo em termos estatísticos não aponta para qual tipo de autocorrelação espacial é predominante, ou seja, o I de Moran não determina se está no termo da variável dependente ou se está no termo do erro.

Para resolver este problema de especificação dos modelos de regressão espacial, a recomendação é utilizar os testes focados, pois estes são capazes de identificar a autocorrelação espacial, e determinar se ela ocorre predominantemente na variável dependente ou nos termos do erro.

Para comparar os resultados, foram realizados os testes de especificação do modelo da

regressão espacial baseados no Multiplicador de Lagrange da variável dependente LM (lag) e do erro LM (erro) e também nas suas versões robustas. Para especificar o modelo de regressão espacial (variável dependente ou erro) foi utilizado o processo de decisão proposto por Anselin (2005).

Foi utilizada a matriz baseada na distância Euclidiana de 7 km, que ficou entre as melhores matrizes na capacidade de capturar a autocorrelação espacial pelo teste I de Moran mostrado na Tabela 14.

Os resultados dos testes de especificação, calculados no *software* GeoDa, baseados no Multiplicador de Lagrange e sua significância também na sua forma robusta, que podem ser observados na Tabela 15. Pode-se ressaltar que tanto o LM (lag) e o LM robusto (lag) não se mostraram significativos, já o LM (erro) e o LM robusto (erro) foram significativos a 5% de probabilidade.

Tabela 15. Diagnóstico de autocorrelação espacial para o LM

Teste	Valor	Valor - p
LM (lag)	2,1709	0,14064
LM robusto (lag)	0,8210	0,36489
LM (erro)	5,4681	0,0193
LM robusto (erro)	4,1182	0,04242

Desta forma é comprovada a existência de uma estrutura espacial nos resíduos do modelo de mínimos quadrados. Os resultados mostram a existência de forte autocorrelação espacial nos resíduos do modelo de regressão por mínimos quadrados, pois os dois testes do modelo espacial do erro mostram-se fortemente significativos.

Como no modelo de regressão não foram utilizadas variáveis de localização (coordenadas X e Y), somente variáveis endógenas e exógenas dos imóveis, este resultado era esperado, pois a ausência das variáveis relacionadas à posição geográfica no modelo, faz com que o resíduo carregue o efeito destas variáveis não especificadas. O termo do erro apresentou uma estrutura de dependência espacial e este resultado foi comprovado com os testes de autocorrelação espacial.

4.4.3 Teste com modelo de regressão vazio

Outra forma de detectar onde ocorre a autocorrelação espacial, se na variável dependente ou no erro é realizar a regressão através de um modelo vazio (MV), pois considera-se que a presença das variáveis explicativas (X_i) podem afetar ou mascarar a relação da variável dependente (Y_i) com os resíduos (ϵ_i).

Para tanto criou-se uma variável constante (CONST) onde foi atribuído valor igual a 1 para todos os 101 dados da amostra. Em seguida com a variável independente LN_VU e as covariáveis CONST e os resíduos do modelo da regressão (RESIDUOS) como independentes, foi realizada uma nova regressão no *software* GeoDa.

Em seguida foi calculado o I de Moran para os novos valores previstos de Y_i (Y_P_MV) e para os novos resíduos (RES_MV). Os resultados confirmaram o que foi detectado com o Multiplicador de Lagrange: a autocorrelação espacial é predominante no termo do erro, pois utilizando o artifício do modelo vazio, o I de Moran foi maior para os novos resíduos. Os resultados podem ser observados nos gráficos da Figura 26, onde estão os diagramas de dispersão de Moran para os valores previstos de Y_i (Y_P_MV) e os resíduos

(RES_MV) do modelo vazio. E na Tabela 16, com os valores de I do Moran e sua significância.

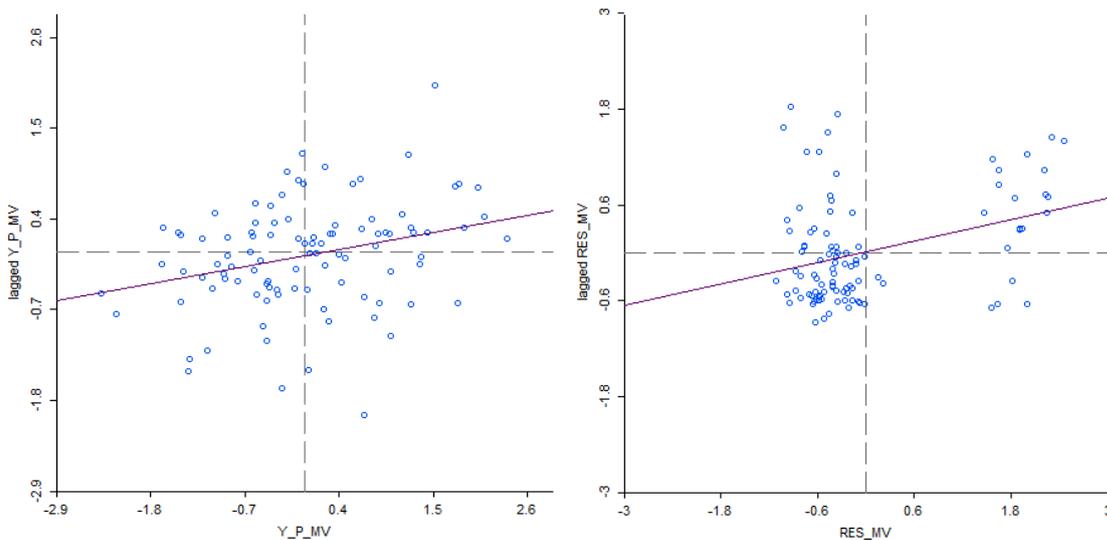


Figura 26. Diagrama de dispersão de Moran para a variável dependente e os resíduos do modelo vazio

Tabela 16. Teste I de Moran e significância para o modelo vazio

Termo	I de Moran	Valor-p
Y_P_MV	0,186143	0,01
RES_MV	0,222323	0,01

4.4.4 *Spatial Error Model - SEM ou CAR*

Diante destes resultados, a especificação do modelo indicou a construção do modelo econométrico espacial autorregressivo, conhecido como *Spatial Error Model* (SEM ou CAR).

Com a matriz de pesos espaciais - W definida, foi estimado o modelo espacial do erro considerando as mesmas variáveis explicativas anteriormente descritas. Os resultados para o modelo espacial do erro - CAR usando o *software* GeoDa são mostrados na Tabela 17.

A variável LAMBDA - λ representa o termo de autocorrelação espacial do erro e se mostra significativa, indicando a existência da autocorrelação espacial.

Segundo Trivelloni (2005), a estatística z corresponde ao equivalente para a regressão por máxima verossimilhança, ao valor t de *Student* para o método de mínimos quadrados. As probabilidades indicam o grau de significância de cada variável de forma análoga ao da regressão por MQO. Observa-se que, com exceção da variável distância a polo de valorização (INV_DPV), todas as variáveis são significativas, considerando 10% de probabilidade.

O novo modelo de regressão considerando agora o termo do erro LAMBDA ficou como segue abaixo:

$$\begin{aligned}
 \text{VU} = & \text{EXP} (8,950226 + 0,0245289 * [\text{VA}] + 1,908886 * [\text{INV_DPV}] + 2,284196 * \\
 & [\text{INV_DM}] + 0,017796 * [\text{AÇU}] + 0,1060113 * [\text{POT_AVN}] - 0,099618 * [\text{LN_AT}] + \\
 & 0,0167538 * [\text{TOPO}] + 0,1246987 * [\text{SIT}] + 0,0002978 * [\text{AIRRIGA}] - 0,0369859 * \\
 & [\text{CULT}] + 0,3454326 * \text{We})
 \end{aligned}$$

Tabela 17. Estimativas dos parâmetros, erro-padrão, valor z e probabilidade do modelo CAR

Variável	Coefficiente	Erro Padrão	Valor-z	Probabilidade
CONSTANT	8,9502260	0,1819787	49,1828300	0,0000000
VA	0,0245289	0,0083054	2,9533770	0,0031400
INV_DPV	1,9088860	1,3389010	1,4257110	0,1539500
INV_DM	2,2841960	0,9532604	2,3961930	0,0165700
AÇU	0,0177960	0,0087071	2,0438420	0,0409700
POT_AVN	0,1060113	0,0546910	1,9383690	0,0525800
LN_AT	-0,0996118	0,0265129	-3,7571110	0,0001700
TOPO	0,0167538	0,0100350	1,6695370	0,0950100
SIT	0,1246987	0,0227731	5,4757120	0,0000000
AIRRIGA	0,0002978	0,0001637	1,8190500	0,0689000
CULT	-0,0369859	0,0069492	-5,3223050	0,0000000
LAMBDA - λ	0,3454326	0,1020994	3,3832960	0,0007200

Considerando o teste de Breusch-Pagan os resíduos do modelo não apresentam problemas de heterocedasticidade (valor = 14,1499 e p-valor = 0,16626). Como a significância foi maior de 5%, o coeficiente do modelo do erro autorregressivo espacial foi capaz de corrigir a heterocedasticidade que estava presente no modelo de regressão clássico.

Em condições *ceteris paribus* tomando o antilogaritmo da equação de regressão, a interpretação do modelo CAR é mostrada a seguir:

$$VU = 7.709,634 * 1,02483^{VA} * 6,74557^{INV_DPV} * 9,81779^{INV_DM} * 1,01795^{AÇU} * 1,11183^{POT_AVN} * 0,90519^{LN_AT} * 1,01689^{TOPO} * 1,13281^{SIT} * 1,00029^{AIRRIGA} * 0,96369^{CULT} * 1,41260^{We} * \varepsilon$$

De acordo com o modelo CAR, o valor médio do hectare na região do PROJIR é de R\$ 7.709,63. O valor unitário aumenta quando o pavimento das vias de acesso são asfalto com relação à via não pavimentada. O valor unitário diminui quando a distância ao centro de Campos de Goytacazes aumenta, sendo que a taxa de diminuição não é constante. O valor unitário diminui quando a distância ao mar aumenta. O valor unitário aumenta quando na propriedade existem açudes e áreas de vegetal nativa, diminui com o aumento da área, aumenta quando a topografia é plana e quando a sua situação em termos de localização no município é ótima, aumenta quando a área irrigada aumenta e diminui quando o produto cultivado no imóvel é a cana-de-açúcar.

Comparando o modelo CAR com o modelo MCRL pode-se perceber que os coeficientes das variáveis distâncias a polo de valorização e distância ao mar, tiveram a maior alteração, pois ocorreu um aumento nos valores, indicando que as variáveis relacionadas à localização são muito importantes na avaliação de imóveis.

4.5 Análise da Regressão Geograficamente Ponderada – RGP

A regressão geograficamente ponderada é uma técnica que estende o método tradicional da regressão, permitindo que os parâmetros sejam estimados localmente, para que os coeficientes do modelo, em vez de serem estimativas globais, sejam específicos para um

ponto *i* (localização).

As localizações de cada um dos pontos são medidas conforme a sua proximidade e similaridade, isto é, observações que se encontrem mais perto de uma determinada localização tenham maior influência que pontos mais afastados e o modelo é calibrado utilizando o método dos mínimos quadrados ponderados. Além disso, os pesos variam conforme a localização das observações, existindo uma diferente calibração para cada observação. O estimador do GWR aplica, então, o método dos mínimos quadrados ponderados com pesos nas localizações, sendo estas ponderações calculadas pela função Kernel espacial.

A especificação do Kernel espacial depende da função matemática na qual serão colocadas as informações sobre as distâncias entre dois pontos geográficos e um parâmetro de largura de banda.

Para a função matemática podem ser utilizados o Kernel fixo ou adaptativo, as informações dos pontos geográficos podem ser as coordenadas (UTM ou geográfica) das amostras. Já para definir a largura de banda é necessário determinar qual é a largura ótima e para determinar este parâmetro foram testadas várias larguras e será utilizada a largura que resultar no menor critério de informação *Akaike* (AIC).

Nesta etapa foi utilizado o software GWR4 para procurar a largura ótima de banda e também os resultados da regressão geograficamente ponderada. O GWR4 tem implementado os seguintes tipos de Kernel: *Fixed Gaussian* (gaussiana de largura fixa), *Fixed bi-square* (biquadrado de largura fixa), *Adaptive bi-square* (biquadrado adaptativo) e *Adaptive Gaussian* (gaussiana adaptativo). O Kernel com largura fixa considera somente uma distância fixa, enquanto os adaptativos consideram a distância do vizinho mais próximo (*nearest neighbour*). Para determinar a largura de banda ótima no GWR4 tem-se os seguintes critérios: AIC (*Akaike*), AICc (*Akaike* corrigido) e CV (validação cruzada).

Foram testados todos os tipos de Kernel e os critérios AIC, AICc e CV, que estão nas Tabelas 18, 19 e 20, onde pode-se observar que o menor valor para o critério AIC ficou no tipo de Kernel biquadrado de largura fixa com valor de 44,74.

Tabela 18. Resultados dos tipos de Kernel com o critério AIC

Tipo de Kernel	Valor Mínimo do Critério AIC	Largura Ótima	Valor de AKAIKE
<i>Fixed Gaussian</i>	52,77	14.055,07	52,77
<i>Fixed bi-square</i>	44,74	28.110,14	44,74
<i>Adaptive Gaussian</i>	70,88	101,00	70,88
<i>Adaptive bi-square</i>	68,24	62,00	68,24

Tabela 19. Resultados dos tipos de Kernel com o critério AICc

Tipo de Kernel	Valor Mínimo do Critério AICc	Largura Ótima	Valor de AKAIKE
<i>Fixed Gaussian</i>	76,063	57.987,4	70,91
<i>Fixed bi-square</i>	76,711	115.974,8	71,20
<i>Adaptive Gaussian</i>	75,134	101,00	70,88
<i>Adaptive bi-square</i>	80,962	101,00	72,96

Tabela 20. Resultados dos tipos de Kernel com o critério CV

Tipo de Kernel	Valor Mínimo do Critério CV	Largura Ótima	Valor de AKAIKE
<i>Fixed Gaussian</i>	0,123	57.987,4	70,90
<i>Adaptive Gaussian</i>	0,121	101,00	70,88

Com o tipo de Kernel biquadrado de largura fixa definido e a respectiva largura ótima obtém-se os coeficientes locais para cada observação, ou seja, para cada elemento da amostra o GWR4 calcula um coeficiente diferente. Em termos de representação, os resultados obtidos para o modelo RGP com os valores médios dos coeficientes, são mostrados na Tabela 21. Chama a atenção o valor negativo no coeficiente da variável Distância ao Polo de Valorização e na Topografia, sendo que nos outros modelos o sinal foi positivo, bem como o erro padrão muito alto das variáveis relacionadas às distâncias (DPV e DM).

Tabela 21. Estimativa dos parâmetros RGP: coeficiente médio e erro padrão

Variável	Coefficiente Médio	Erro Padrão
Intercepto	9,418033	0,793629
VA	0,038124	0,038567
INV_DPV	-3,505919	14,28091
INV_DM	3,379248	8,735627
ACU	0,021735	0,031073
POT_AVN	0,035600	0,275032
LN_AT	-0,132732	0,070135
TOPO	-0,00685	0,045024
SIT	0,114595	0,09037
AIRRIGA	0,000698	0,000972
CULT	-0,045773	0,049869

O novo modelo considerando os coeficientes médios obtidos pela regressão geograficamente ponderada ficou como segue abaixo:

$$\text{VU} = \text{EXP} (9,418033 + 0,038124 * [\text{VA}] - 3,505919 * [\text{INV_DPV}] + 3,379248 * [\text{INV_DM}] + 0,021735 * [\text{AÇU}] + 0,035600 * [\text{POT_AVN}] - 0,132732 * [\text{LN_AT}] - 0,00685 * [\text{TOPO}] + 0,114595 * [\text{SIT}] + 0,000698 * [\text{AIRRIGA}] - 0,045773 * [\text{CULT}])$$

Para este modelo não faz muito sentido analisar os coeficientes, pois tanto os sinais como os valores mudam para cada observação da amostra.

4.6 Avaliação da Qualidade do Ajuste e Desempenho dos Modelos

4.6.1 Qualidade do ajuste

Para escolher o melhor modelo, a recomendação é que quando da utilização da estimação por Máxima Verossimilhança (MV), seja feita a combinação do valor da função de verossimilhança (LIK) com os critérios de informação de Akaike (AIC).

Na Tabela 22 é mostrado um comparativo entre os parâmetros de escolha dos melhores modelos de regressão. Levando-se em consideração o critério AIC, percebe-se que inicialmente o modelo CAR foi melhor que o modelo MCRL e o modelo RGP foi o que obteve o menor valor de AIC, sendo, portanto, o melhor ajustado. Da mesma forma o coeficiente de determinação (R^2) do RGP, foi muito superior aos outros modelos e o erro padrão da regressão teve uma pequena diminuição. No critério LIK quando maior o seu valor melhor será o ajuste e neste caso o RGP foi melhor que os outros modelos.

Tabela 22. Comparativo entre os modelos por meio do AIC, R^2 , LIK e erro padrão

Modelos	AIC	R^2	Log Likelihood - LIK	Erro padrão da regressão
MCRL	68,567	0,664650	23,2835	0,32279
CAR	61,4325	0,699215	19,71626	0,288576
RGP ou GWR	44,7437	0,885030	61,55423	0,284247

Com relação ao coeficiente de determinação, a literatura não recomenda sua utilização para comparar o ajuste de modelos clássicos de regressão com modelos de regressão espaciais, entretanto somente a título de ilustração foi demonstrado neste trabalho as mudanças de valores deste parâmetro.

É imprescindível lembrar que no MCRL não foram utilizadas variáveis relacionadas à posição geográfica e, portanto, percebe-se que o modelo CAR conseguiu modelar a dependência espacial presente nos resíduos do referido modelo, cumprindo assim a sua função. Este resultado corrobora com outros trabalhos desenvolvidos no Brasil, onde a utilização da regressão espacial trata a dependência espacial e melhora o modelo clássico.

O modelo RGP foi superior ao MCRL e ao CAR em todos os critérios e com isto podemos concluir nesta etapa que a regressão geograficamente ponderada deixou a modelagem mais ajustada e, portanto, teve um desempenho superior aos outros dois modelos.

4.6.2 Avaliação do desempenho

Para garantir que a avaliação do desempenho dos modelos de regressão seja realizada utilizando-se dados que não foram utilizados na modelagem, foi separada uma amostra de validação com 25 elementos (ANEXO 3), que foram obtidas no mesmo período que os dados utilizados na modelagem das regressões. O objetivo aqui é verificar o funcionamento dos mesmos numa situação mais próxima da sua aplicação prática.

A amostra de validação utilizada correspondeu a 22,1% do total de dados. Na Figura 12 é possível verificar a distribuição espacial da amostra de validação juntamente com as amostras da modelagem e pode-se verificar que a mesma contempla todas as regiões do PROJIR.

A amostra de validação foi utilizada para o cálculo dos parâmetros definidos pela *International Association of Assessing Officers* (IAAO, 2013a) a saber: Mediana das Razões, Coeficiente de Dispersão (COD) e Diferencial Relativo ao Preço (PRD). Os resultados serão discutidos levando-se em consideração os parâmetros de referência para avaliação do desempenho mostrados na Tabela 23.

Os resultados podem ser visualizados na Tabela 24, sendo que foram utilizados todos os dados da amostra de validação. Para o modelo RGP, que gera coeficientes diferentes para cada dado da amostra, o *software* GWR4 possui um módulo de cálculo para amostras de

verificação, que tornou possível o cálculo dos parâmetros da IAAO. Apesar da expectativa diante dos resultados do critério de Akaike (AIC) de que o modelo RGP teria o melhor ajustamento, os valores do coeficiente de dispersão não confirmaram o resultado anterior.

Tabela 23. Valores recomendados pela IAAO para avaliações em massa de imóveis

Métricas	Valores
Mediana das razões da avaliação	Entre 0,90 e 1,10
Coefficiente de Dispersão (COD) para imóveis rurais	Entre 5 a 20%
Diferencial Relativo ao Preço (PRD)	Entre 0,98 e 1,03

Tabela 24. Resultados das métricas para avaliação de desempenho com as 25 observações

Modelo	COD (%)	PRD	Mediana das Razões
MCRL	33,14	0,79	0,92
CAR	30,16	0,65	0,87
RGP ou GWR	38,72	0,74	0,86

Como o COD é afetado sensivelmente por valores extremos das razões de avaliação, a amostra foi saneada, multiplicando o intervalo interquartil por 1,5 para definir os valores limites de *ratios* a considerar na análise. Estes novos valores destas métricas podem ser visualizados na Tabela 25.

Com relação à Mediana das Razões, todos os modelos ficaram abaixo do limite mínimo de referência estabelecidos pela norma da IAAO. Com relação ao COD todos os modelos ficaram com os valores acima dos recomendados pela norma internacional, sendo que o modelo RGP foi o que apresentou o pior resultado. Provavelmente isso ocorreu porque a área de estudo do PROJIR abrange mais de um município, e os valores dos imóveis na maioria dos casos, tem comportamentos diferentes dentro dos seus respectivos limites municipais, ou seja, a análise ao nível local (que é o caso do RGP), considerando uma única área, gerou uma dispersão dos valores maior que as outras técnicas. A solução para este problema seria analisar os modelos separando cada município e verificando isoladamente o comportamento dos valores dos imóveis rurais. Com relação ao PRD os três modelos estão abaixo dos limites de referência, ou seja, ocorreu progressividade na avaliação.

Tabela 25. Resultados das métricas para avaliação de desempenho com a amostra saneada

Modelo	COD (%)	PRD	Mediana das Razões
MCRL	23,82	0,70	0,88
CAR	21,52	0,67	0,87
RGP ou GWR	28,78	0,65	0,83

Diante destes resultados, avalia-se que tanto o MCRL como o modelo CAR foram satisfatórios, apresentando uma boa avaliação de desempenho, sendo que o modelo CAR obteve em todos as métricas da IAAO, valores um pouco inferiores que o MCRL. Entretanto com relação ao RGP, os resultados do COD não foram condizentes com os resultados apresentados na Tabela 22, pois mostrava o RGP como o modelo com um melhor ajustamento, pelos critérios AIC e LIK.

A realização da avaliação de desempenho mostrou o quanto é importante a utilização de uma amostra que não fez parte do modelo, pois foi capaz de indicar quais aspectos devem ser aperfeiçoados para construir modelos mais eficientes. Nos modelos em análise, o principal aspecto a ser aprimorado é a realização dos modelos dividindo-se a área do PROJIR nos municípios que a compõe, entretanto, a quantidade de amostras muito reduzida para cada município, impossibilitou a realização desta análise separadamente. Outro aspecto que pode reduzir o valor do COD, em todos os modelos de regressão é aumentar a quantidade de dados amostrais a serem utilizados na modelagem para obter resultados melhores, pois em um conjunto de dados com centenas de características a modelagem se tornará mais ajustada.

Na próxima etapa os três modelos foram utilizados para gerar a superfície de valores através da Geoestatística, sendo que a amostra de validação foi usada novamente para avaliar o desempenho dos modelos, com os valores calculados para a Planta de Valores Genéricos.

4.7 Elaboração da PVG Através da Geoestatística

Nesta etapa para elaborar as superfícies de valores na área do PROJIR, foi utilizado o conjunto dos valores estimados pelos modelos MCRL, MRE e RGP, através da Geoestatística. Primeiramente foi realizada a modelagem dos semivariogramas experimentais, para posteriormente ajustar ao modelo teórico que fornecem os parâmetros (efeito pepita, alcance e patamar) necessários ao interpolador da Krigagem Ordinária, possibilitando assim a elaboração da superfície de valores das terras.

Em seguida, as superfícies de valores obtidas foram analisadas na sua capacidade de fornecer informações a respeito do valor do hectare na região e para isto, o conjunto de 25 observações de verificação foi utilizado, da mesma forma que na avaliação do desempenho dos modelos. Para esta etapa foi utilizado o *software* ArcGIS 10.2.1 através do módulo *Geostatistical Analyst*.

4.7.1 Superfície dos valores estimados com MCRL

O primeiro passo para a modelagem Geoestatística é a construção do semivariograma experimental, que contou com 12 pontos em um espaçamento de 2,46 quilômetros (Tabela 26) e pode-se perceber que a semivariância é crescente até aproximadamente 22 km, quando se estabiliza, caracterizando o patamar. Todos os pontos foram calculados a partir de no mínimo 52 pares de pontos, que ficou acima do número de pares recomendado por Landim (2006), para uma análise Geoestatística, sendo razoável entre 30 a 40 pares como um número mínimo de pontos amostrados. Os parâmetros podem ser observados na Tabela 26 e o gráfico do semivariograma omnidirecional na Figura 27.

Após a obtenção do semivariograma experimental foram ajustados os modelos teóricos ao mesmo, devido à necessidade de ajustar uma função matemática que descreva a variabilidade ou correlação espacial existente nos dados. A Tabela 27 mostra os parâmetros para os modelos Esférico, Exponencial e Gaussiano. O ArcGIS 10.2.1 tem outros modelos implementados, entretanto os três citados são os mais citados na literatura e segundo Yamamoto e Landim (2013) explicam a maioria dos fenômenos espaciais.

O maior alcance ocorreu com o modelo Exponencial e o menor alcance aconteceu com o modelo Gaussiano. Para avaliar o grau de dependência espacial, foram utilizados os valores propostos por Guimarães (2004): dependência espacial = $\left(\frac{C_0}{C}\right)$. Quando os valores da relação são inferiores a 25% caracterizam forte dependência espacial, entre 25 e 75%

moderada e acima de 75%, fraca dependência espacial.

O menor efeito pepita ocorreu no modelo exponencial, que apresentou dependência espacial forte. Diante disto optou-se em realizar a validação cruzada, que é bastante utilizada por oferecer critérios para a escolha do melhor modelo ajustado. A validação cruzada é realizada com os valores preditos pelo MCRL e os valores preditos pelo modelo teórico. O melhor modelo teórico ajustado será aquele que apresentar o Erro Médio, Erro Médio Quadrático (*Root Mean Square* - RMS) e Erro Médio Padronizado mais próximo possível de zero. Quando apresentar um RMS padronizado mais próximo possível do valor um e quando o erro padrão médio (*Average Standard Error*) apresentar valores mais próximos possível de zero.

Tabela 26. Parâmetros para o semivariograma experimental modelo MCRL

Ponto	Distância (km)	Semivariância * 10 ¹	Nº de Pares
1	1,23	0,6191	52
2	3,70	1,0185	156
3	6,17	1,8915	206
4	8,63	1,3986	236
5	11,10	1,9150	332
6	13,57	1,6584	322
7	16,04	2,1392	428
8	18,50	2,1854	466
9	20,97	2,0640	464
10	23,44	2,2122	450
11	25,90	2,1321	414
12	28,37	2,3248	452

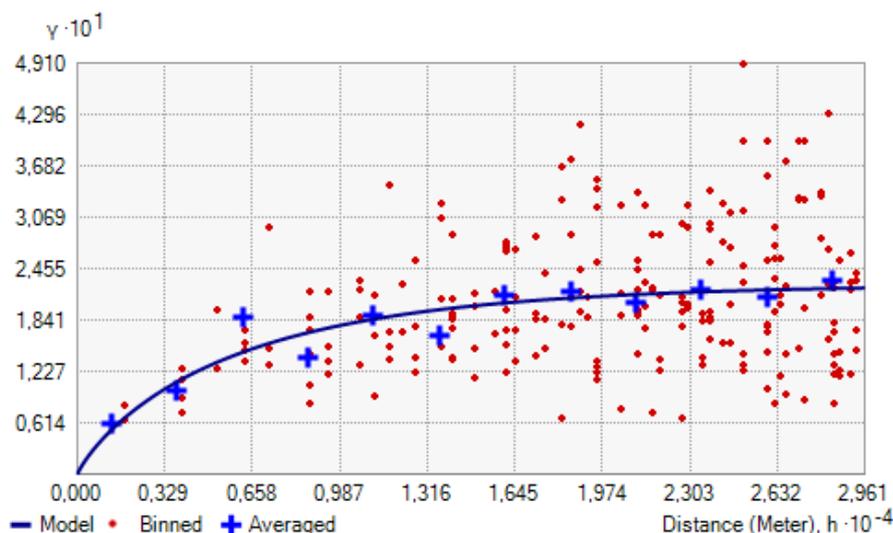


Figura 27. Semivariograma experimental para MCRL

Os resultados da validação cruzada dos modelos teóricos encontram-se expostos na Tabela 28. A análise dos parâmetros não permite afirmar qual dos modelos teóricos foi o que

obteve o melhor ajuste, como exemplo o método Gaussiano no RMS padronizado o valor está mais próximo de 1 (um), entretanto nos outros parâmetros os valores estão piores que nos métodos esférico e exponencial. Optou-se então pelo modelo exponencial por ter o menor efeito pepita e por ter capturado a maior dependência espacial. Na Figura 28 encontra-se a representação do referido modelo teórico.

Tabela 27. Parâmetros dos semivariograma teóricos para MCRL

Modelo Teórico	C ₀ -Efeito Pepita	C - Patamar	$\left(\frac{C_0}{C} \times 100\%\right)$	a – alcance (km)	Dependência Espacial
Esférico	0,06264	0,15290	40,97	18,13	Moderada
Exponencial	0,01345	0,20793	6,47	18,41	Forte
Gaussiano	0,08192	0,13146	62,32	14,14	Moderada

Tabela 28. Valores da validação cruzada dos modelos teóricos para MCRL

Modelo Teórico	Erro Médio	RMS	Erro Médio Padronizado	RMS Padronizado	Erro Padrão Médio
Esférico	0,005317	0,361453	0,007458	0,973770	0,365570
Exponencial	0,006967	0,360064	0,010514	0,981364	0,360260
Gaussiano	0,005759	0,368812	0,006129	0,994283	0,366009

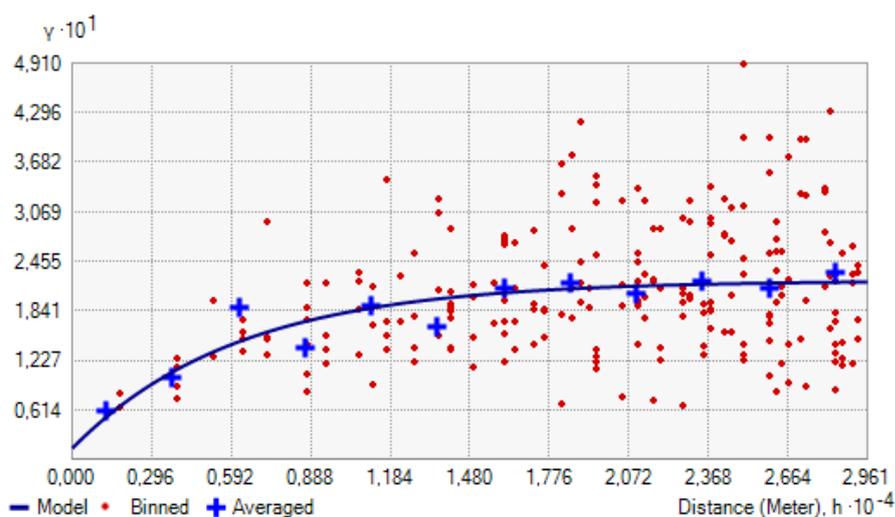


Figura 28. Semivariograma teórico exponencial para MCRL

Os parâmetros obtidos a partir do modelo teórico ajustado (exponencial) foram utilizados para gerar a superfície de valores para a área de estudo, e para a superfície de predição do erro da Krigagem ordinária, os quais podem ser visualizados na Figura 29 e 30.

Na Figura 29, onde está representada a superfície de valores percebe-se como a Krigagem conseguiu representar a variação de preços que ocorre na área de estudo. Os maiores valores estão relacionados à proximidade às áreas urbanas de alguns municípios do PROJIR: São João da Barra, Campos dos Goytacazes e Quissamã. Na área à direita também ocorre valores mais altos, devido à proximidade com o litoral e com o Porto do Açú, em São João da Barra. Na parte central da área, onde está o município de Campos e acima desta região também ocorre valores altos, em função da proximidade da BR 101, que corta o município na direção norte.

O mapa da previsão do erro da Krigagem (Figura 30) mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrais e cresce à medida que se afasta destes pontos. Nas regiões onde as amostras foram escassas, principalmente nas bordas, o erro foi maior o que mostra a importância e a necessidade, quando se trata da elaboração da PVG, da maior quantidade possível de dados e uma boa espacialização, para se obter uma superfície de valores mais precisa e confiável.

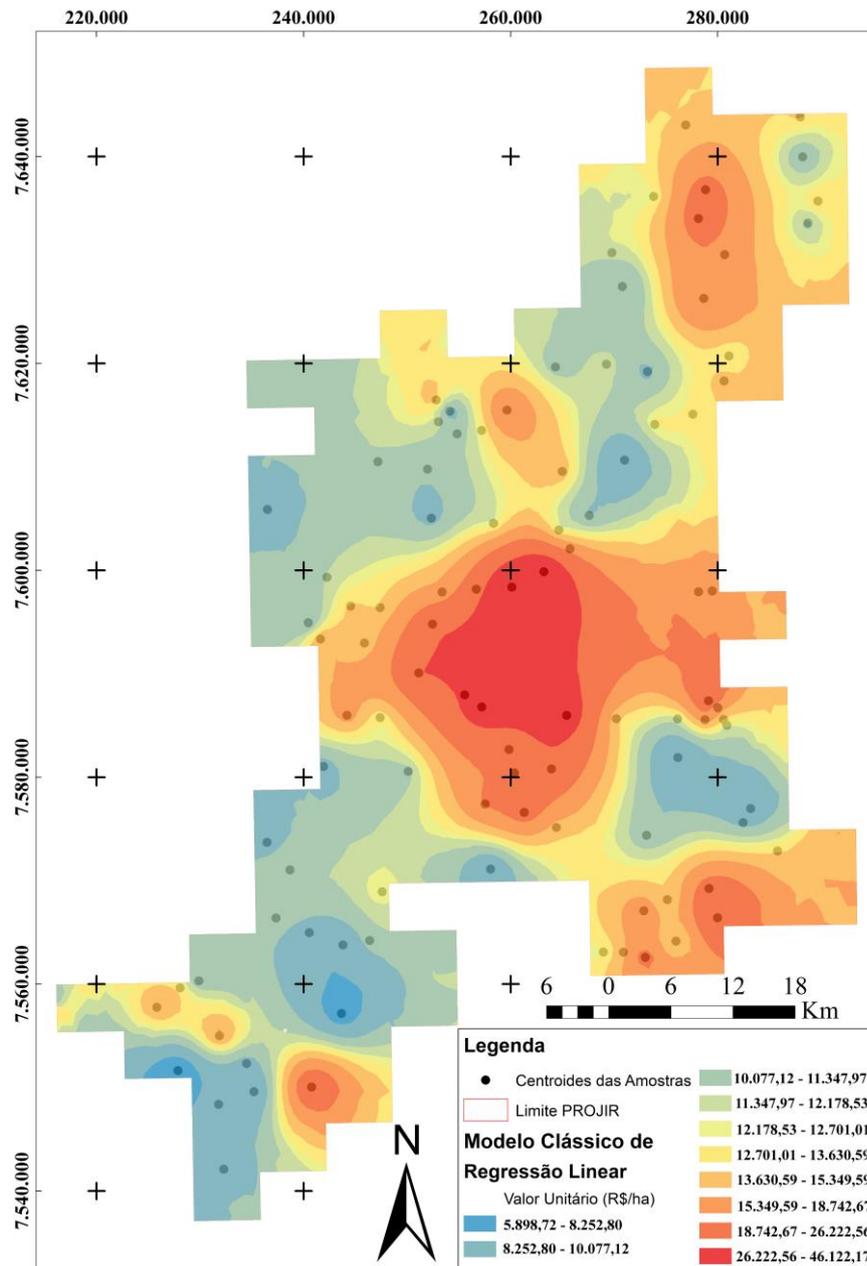


Figura 29. Mapa da Krigagem ordinária para o MCRL

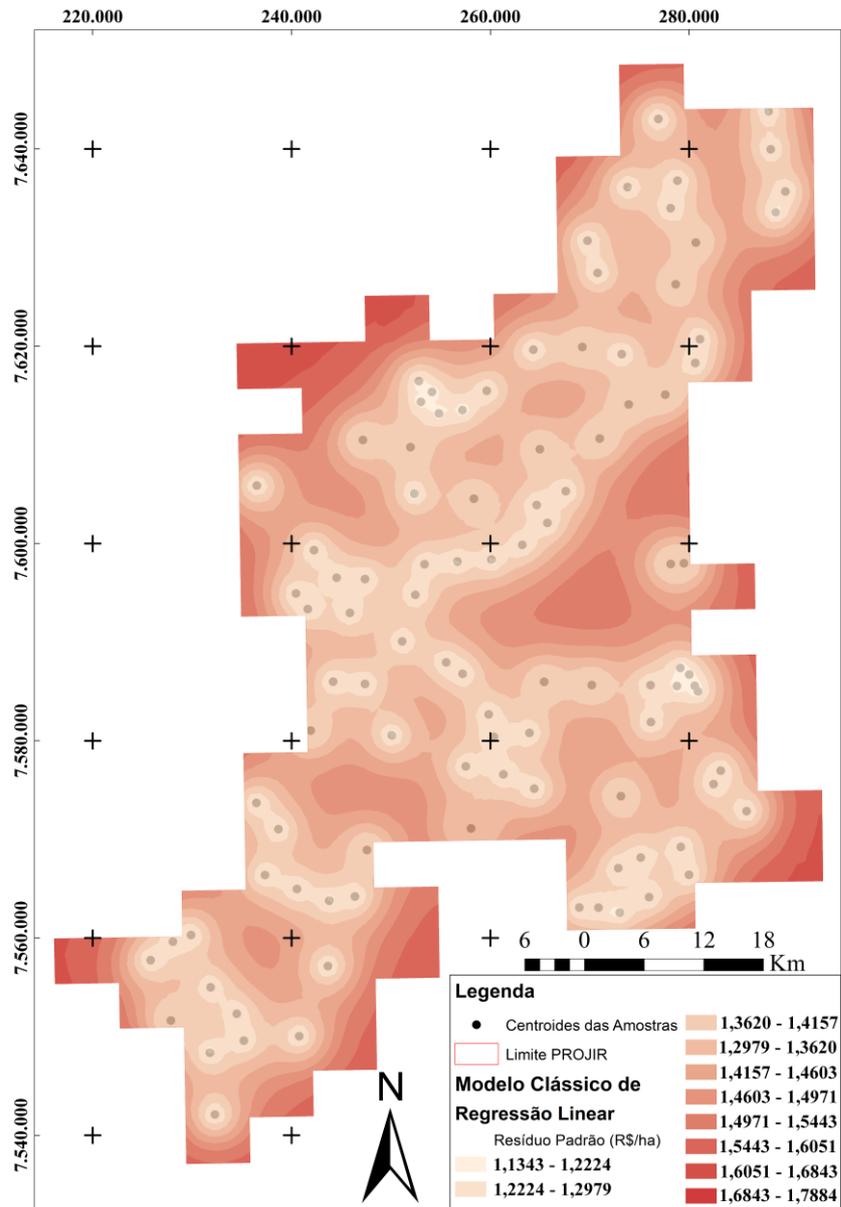


Figura 30. Mapa da previsão do erro para o MCRL

4.7.2 Superfície dos valores estimados com CAR

Da mesma forma que no item anterior, o primeiro passo é a construção do semivariograma experimental, que contou com 12 pontos em um espaçamento de 2,51 quilômetros (Tabela 29) e pode-se perceber que a semivariância é crescente até aproximadamente 22 km, quando se estabiliza, caracterizando o patamar. Todos os pontos foram calculados a partir de no mínimo 52 pares de pontos, que ficou acima do número de pares recomendado entre 30 a 40 pares como um número mínimo de pontos amostrados. Os parâmetros podem ser observados na Tabela 29 e o gráfico do semivariograma omnidirecional na Figura 31.

Após a obtenção do semivariograma experimental foram ajustados os modelos teóricos ao mesmo, devido à necessidade de ajustar uma função matemática que descreva a variabilidade ou correlação espacial existente nos dados. A Tabela 30 mostra os parâmetros

obtidos para os modelos Esférico, Exponencial e Gaussiano. O maior alcance ocorreu com o modelo Exponencial e Esférico e o menor alcance aconteceu com o modelo Gaussiano. O menor efeito pepita ocorreu no modelo exponencial, que apresentou dependência espacial forte. Diante disto optou-se em realizar a validação cruzada, que é bastante utilizada por oferecer critérios para a escolha do melhor modelo ajustado. A validação cruzada é realizada com os valores preditos pelo CAR os valores preditos pelo modelo teórico.

Tabela 29. Parâmetros para o semivariograma experimental modelo CAR

Ponto	Distância (km)	Semivariância * 10 ¹	Nº de Pares
1	1,26	0,5362	52
2	3,77	0,9391	162
3	6,28	1,7510	214
4	8,80	1,4306	242
5	11,31	1,7267	340
6	13,82	1,6452	336
7	16,34	1,9820	456
8	18,85	2,1713	458
9	21,36	1,8887	462
10	23,88	2,1551	482
11	26,39	2,4166	432
12	28,90	1,7425	424

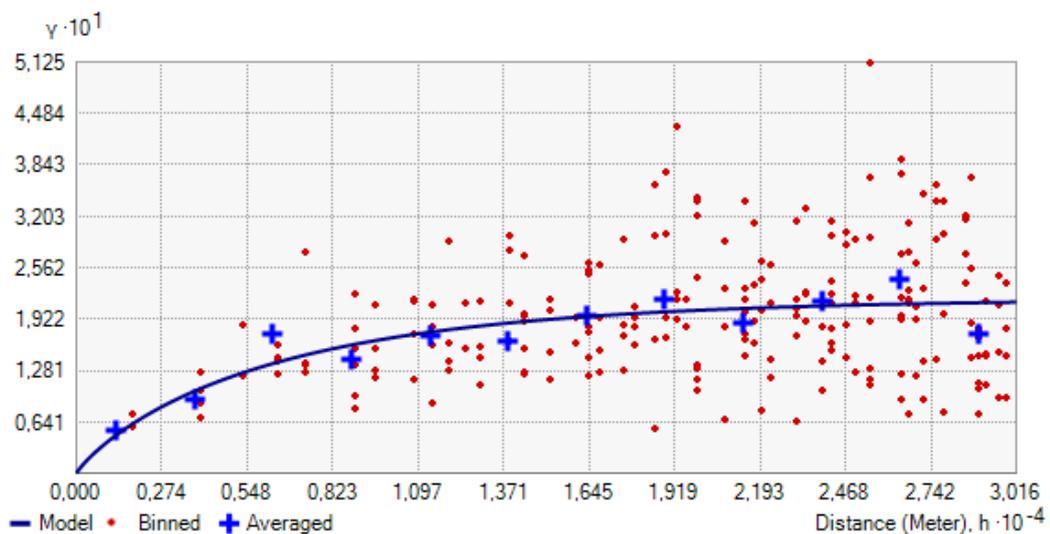


Figura 31. Semivariograma experimental para o CAR

Os resultados da validação cruzada dos modelos teóricos encontram-se expostos na Tabela 31. A análise dos parâmetros não permite afirmar qual dos modelos teóricos foi o que obteve o melhor ajuste, como exemplo o método Gaussiano no RMS padronizado o valor está mais próximo de 1 (um), entretanto nos outros parâmetros os valores estão piores que nos métodos esférico e exponencial. Optou-se então pelo modelo exponencial por ter o menor efeito pepita e por ter capturado a maior dependência espacial. Na Figura 32 encontra-se a

representação do referido modelo teórico.

Tabela 30. Parâmetros dos semivariograma teóricos para o CAR

Modelo Teórico	C ₀ -Efeito Pepita	C - Patamar	$\left(\frac{C_0}{C} \times 100\%\right)$	a – alcance (km)	Dependência Espacial
Esférico	0,05924	0,14758	40,14	19,24	Moderada
Exponencial	0,01070	0,20164	5,31	19,24	Forte
Gaussiano	0,0793	0,12538	63,25	15,20	Moderada

Tabela 31. Valores da validação cruzada dos modelos teóricos para o CAR

Modelo Teórico	Erro Médio	RMS	Erro Médio Padronizado	RMS Padronizado	Erro Padrão Médio
Esférico	0,005850	0,344599	0,008534	0,964327	0,352267
Exponencial	0,007820	0,343537	0,011466	0,978987	0,345445
Gaussiano	0,006149	0,350013	0,007698	0,979643	0,353338

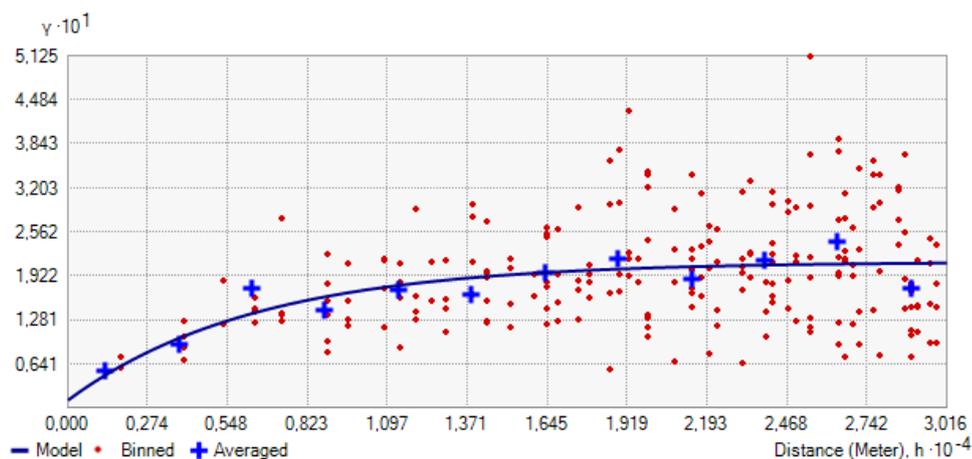


Figura 32. Semivariograma teórico exponencial para o CAR

Os parâmetros obtidos a partir do modelo teórico ajustado (exponencial) foram utilizados para gerar a superfície de valores para a área de estudo, e para a superfície de predição do erro da Krigagem ordinária, os quais podem ser visualizados na Figura 33 e 34.

A Figura 33, representa a superfície de valores e percebe-se como a Krigagem possibilitou a representação da variação de preços que ocorre na área de estudo, entretanto de uma forma mais suavizada e menos detalhada.

O mapa da previsão do erro da Krigagem (Figura 34) também mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrais e cresce à medida que se afasta destes pontos. Entretanto, ocorreu uma variação menor dos valores do erro de estimativa quando comparado com a Krigagem realizada a partir dos valores do MCRL, provavelmente porque o modelo espacial do erro conseguiu tratar e diminuir esta variação.

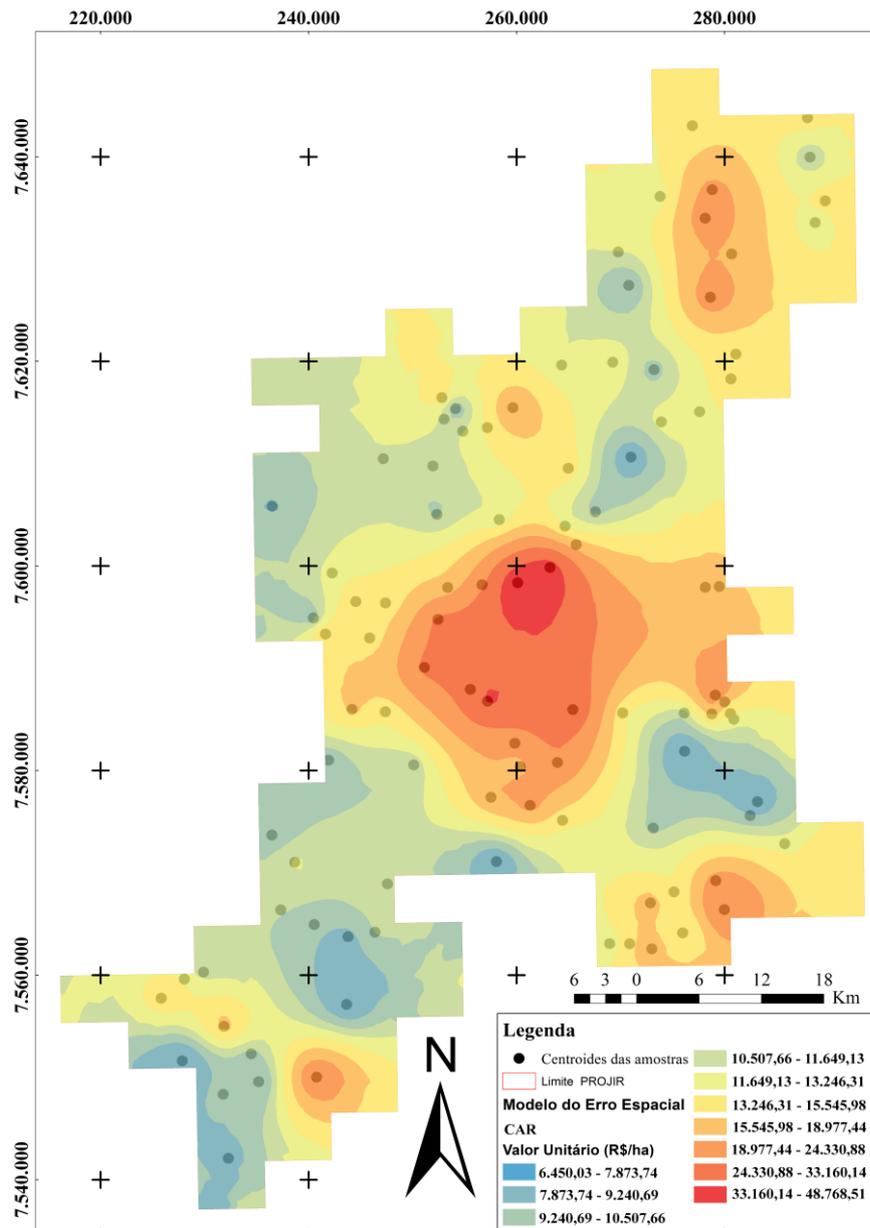


Figura 33. Mapa da Krigagem ordinária para o CAR

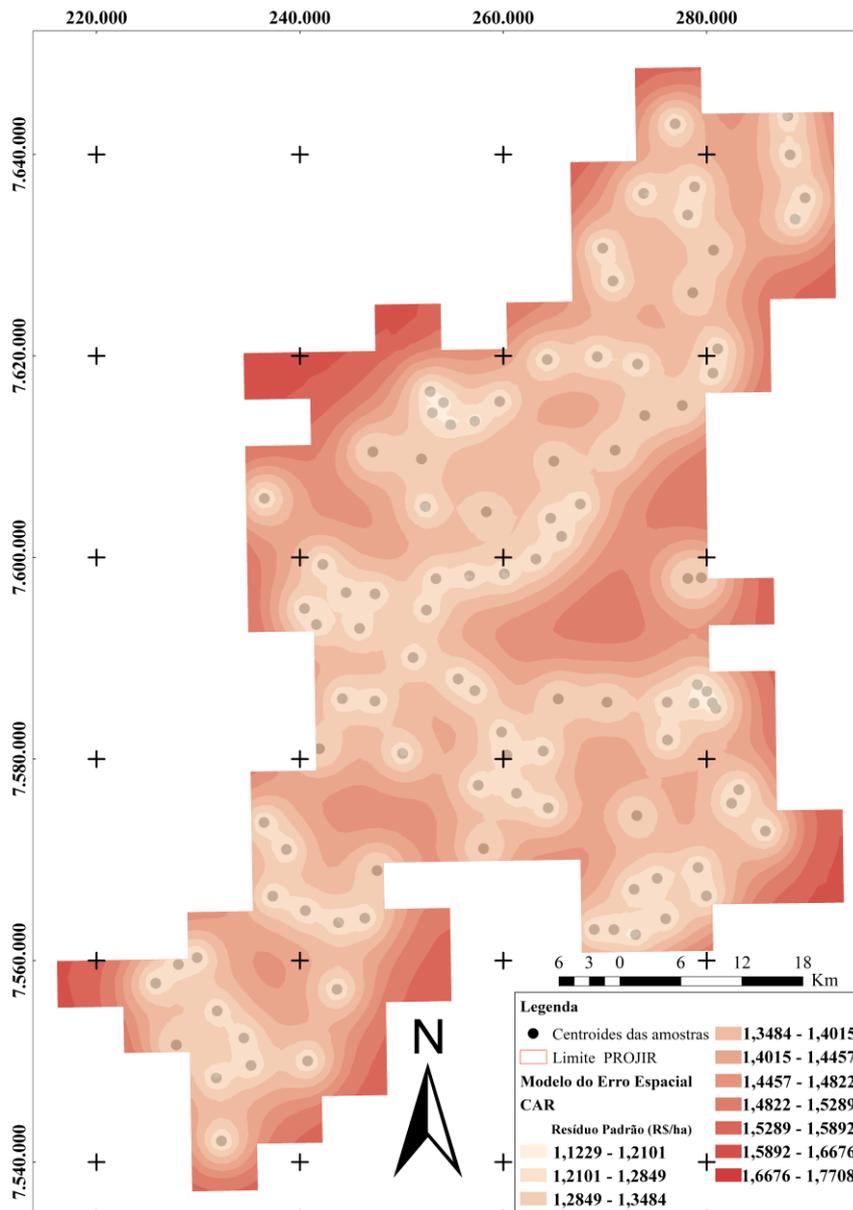


Figura 34. Mapa da previsão do erro para o CAR

4.7.3 Superfície dos valores estimados com RGP

Da mesma forma que no item anterior, o primeiro passo é a construção do semivariograma experimental, que contou com 12 pontos em um espaçamento de 2,16 quilômetros (Tabela 32) e pode-se perceber que a semivariância é crescente até aproximadamente 17 km, quando se estabiliza, caracterizando o patamar. Todos os pontos foram calculados a partir de no mínimo 36 pares de pontos, que ficou entre o intervalo do número de pares recomendado de 30 a 40 pares como um número mínimo de pontos amostrados. Os parâmetros podem ser observados na Tabela 32 e o gráfico do semivariograma omnidirecional na Figura 35.

Tabela 32. Parâmetros para o semivariograma experimental para o RGP

Ponto	Distância (km)	Semivariância * 10 ¹	Nº de Pares
1	1,08	0,5343	36
2	3,25	1,1184	118
3	5,42	1,7486	170
4	7,59	1,9567	198
5	9,75	2,0638	244
6	11,92	2,1235	294
7	14,09	2,1679	302
8	16,25	2,8092	386
9	18,42	2,6538	402
10	20,59	2,8526	406
11	22,76	2,7050	384
12	24,92	2,4999	404

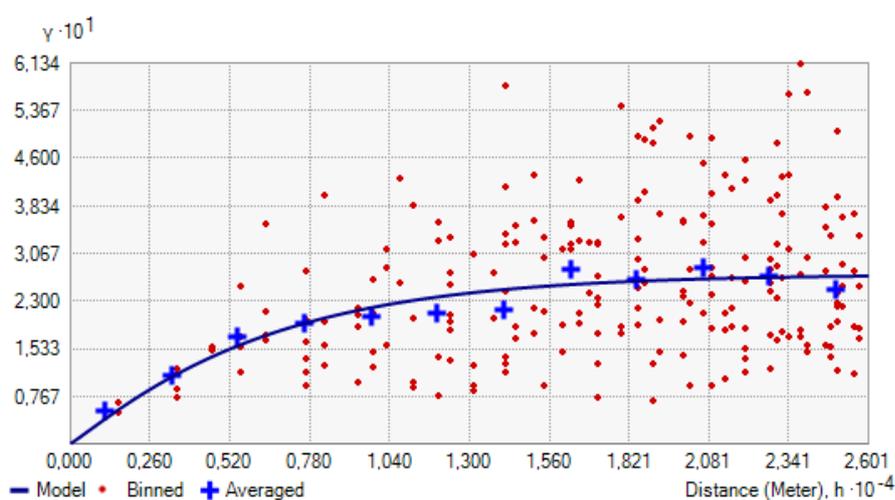


Figura 35. Semivariograma experimental para o RGP

Após a obtenção do semivariograma experimental foram ajustados os modelos teóricos ao mesmo. A Tabela 33 mostra os parâmetros obtidos para os modelos Esférico, Exponencial e Gaussiano. O maior alcance ocorreu com o modelo Exponencial e Esférico e o menor alcance aconteceu com o modelo Gaussiano. O modelo exponencial obteve efeito pepita puro e dependência espacial forte. Diante disto optou-se em realizar a validação cruzada, que foi realizada com os valores preditos pelo RGP e os valores preditos pelo modelo teórico em questão.

Tabela 33. Parâmetros dos semivariograma teóricos para o RGP

Modelo Teórico	C ₀ -Efeito Pepita	C - Patamar	$\left(\frac{C_0}{C} \times 100\%\right)$	a – alcance (km)	Dependência Espacial
Esférico	0,04640	0,21930	21,16	17,10	Forte
Exponencial	0	0,27879	0,00	19,25	Forte
Gaussiano	0,05624	0,20058	28,04	11,33	Forte

Os resultados da validação cruzada dos modelos teóricos estão na Tabela 34. A análise dos parâmetros permite afirmar que o modelo Gaussiano foi o que obteve o melhor ajuste, pois o RMS padronizado obteve valor igual a 1 (um). Apesar do modelo exponencial ter efeito pepita puro, todos tiveram dependência espacial forte. Na Figura 36 encontra-se a representação do referido modelo teórico.

Tabela 34. Valores da validação cruzada dos modelos teóricos para o RGP

Modelo Teórico	Erro Médio	RMS	Erro Médio Padronizado	RMS Padronizado	Erro Padrão Médio
Esférico	0,002979	0,384000	0,003512	0,988893	0,377036
Exponencial	0,004405	0,380459	0,006728	0,985713	0,378866
Gaussiano	0,006797	0,386440	0,006924	1,009823	0,371451

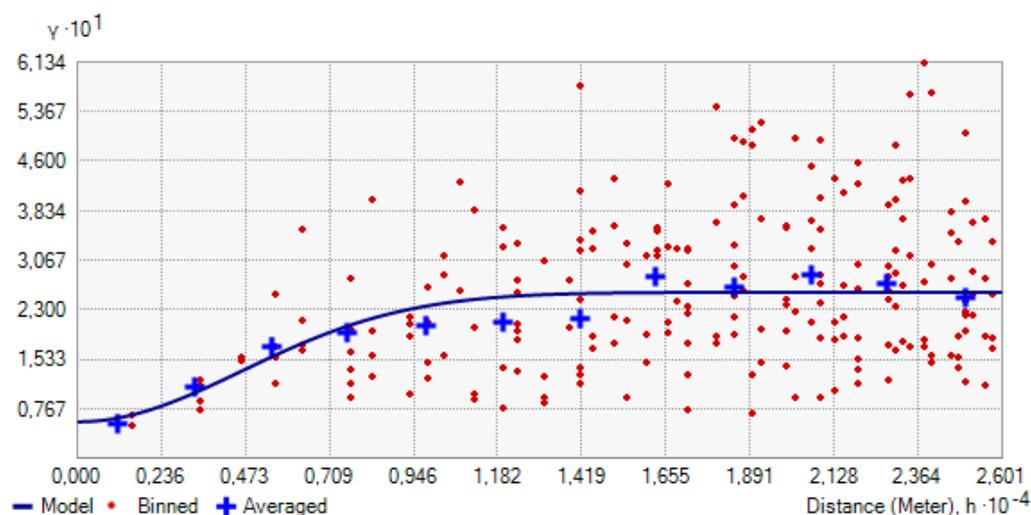


Figura 36. Semivariograma teórico Gaussiano para o RGP

Os parâmetros obtidos a partir do modelo teórico ajustado (Gaussiano) foram utilizados para gerar a superfície de valores para a área de estudo, e para a superfície de predição do erro da Krigagem ordinária, os quais podem ser visualizados na Figura 37 e 38.

Na Figura 37, onde está representada a superfície de valores percebe-se como a Krigagem também conseguiu representar a variação de preços que ocorre na área de estudo, sendo que ficou mais próxima da representação do MCRL, ou seja, o RGP conseguiu representar de uma forma mais detalhada que o CAR.

O mapa da previsão do erro da Krigagem (Figura 38) também mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrais e cresce à medida que se afasta destes pontos. Entretanto, ocorreu uma variação bem menor dos valores do erro de estimativa

quando comparado com a Krigagem realizada a partir dos valores do MCRL e CAR, além de também reduzir a variação nas bordas. Provavelmente isso ocorreu, porque o modelo de regressão geograficamente ponderada conseguiu tratar e diminuir esta variação muito mais que os dois modelos anteriores.

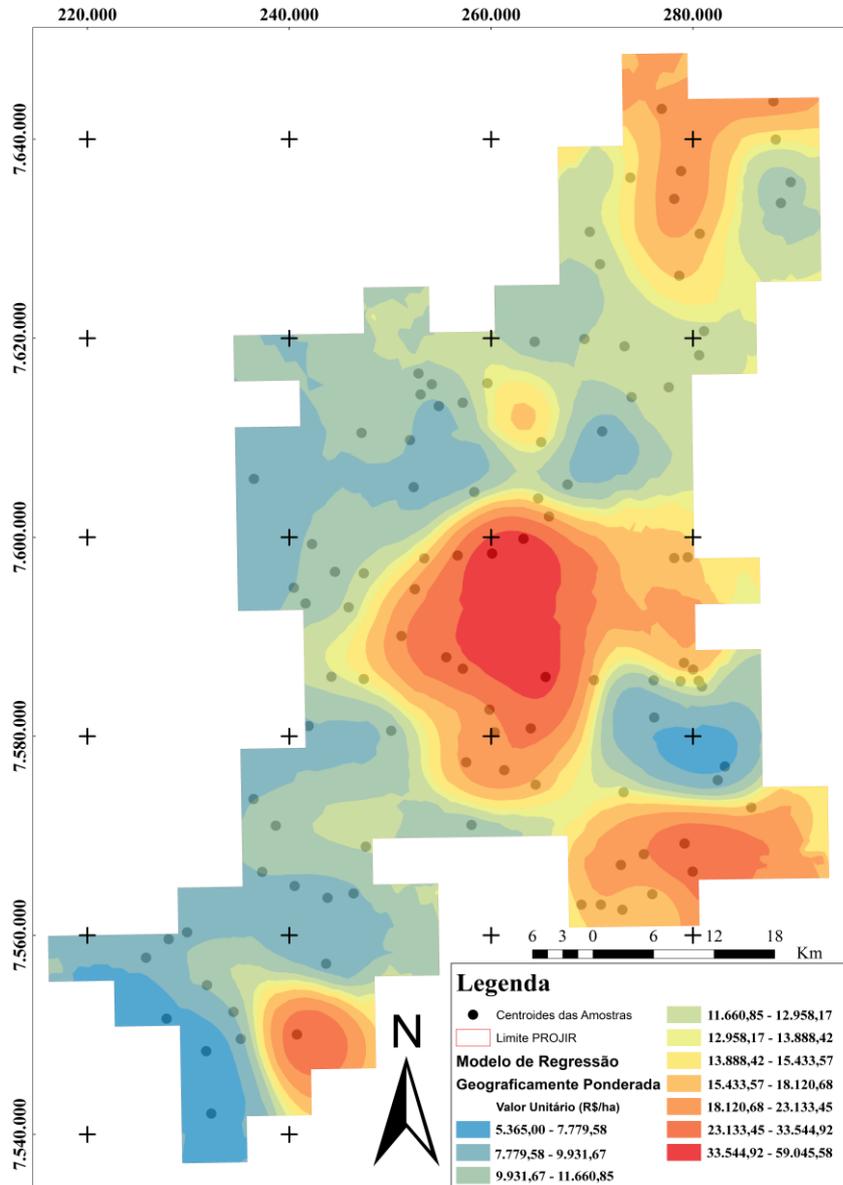


Figura 37. Mapa da Krigagem ordinária para o RGP

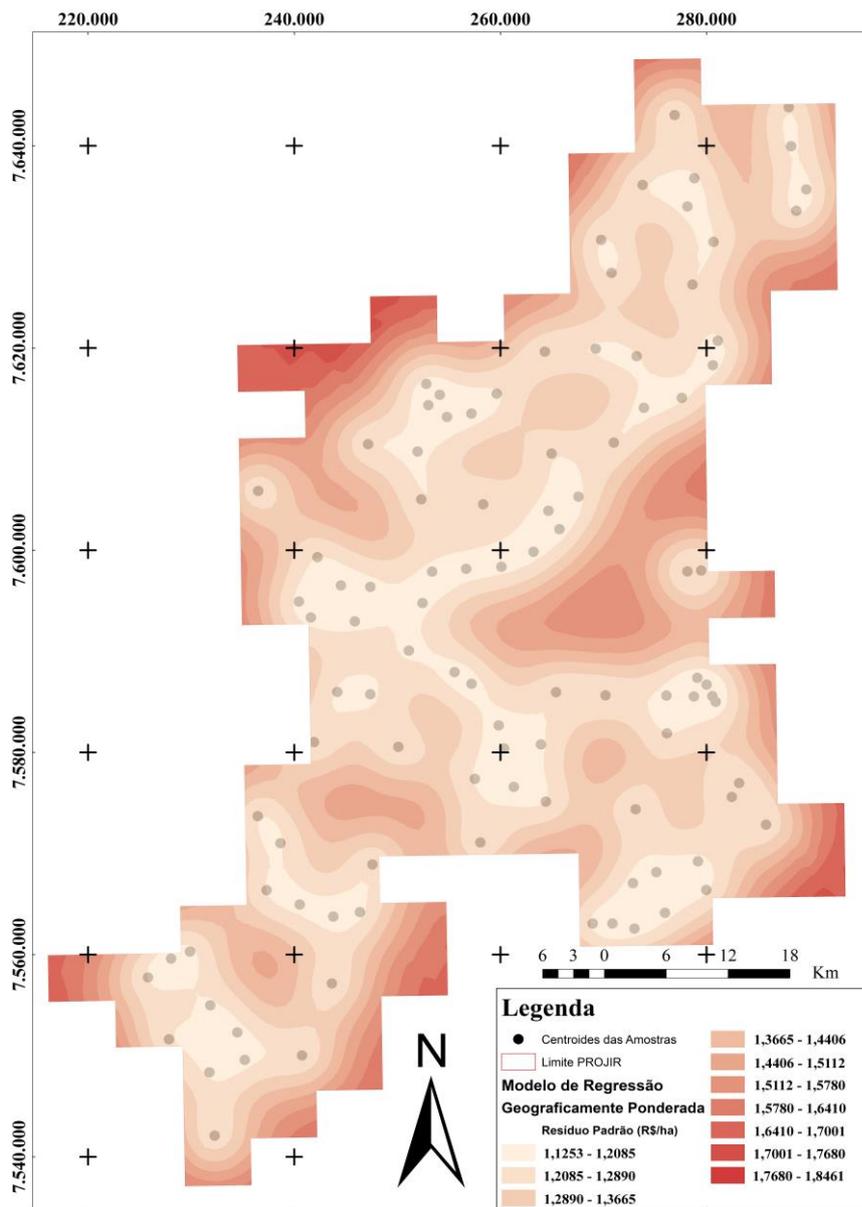


Figura 38. Mapa da previsão do erro para o RGP

4.7.4 Anisotropia

Para analisar os possíveis efeitos da anisotropia nos dados foram construídos semivariogramas direcionais nas quatro principais direções recomendadas: 0°, 45, 90° e 135°. Na Figura 39 a, b, c estão os variogramas direcionais para o MCRL, CAR e RGP. Para o modelo MCRL o semivariograma tem os mesmos valores para as direções 0° e 180° e nas direções 90° e 270° e para o modelo RGP acontece o mesmo nas direções 90° e 270°, o semivariograma tem os mesmos valores. Nos gráficos para os modelos CAR e RGP, observa-se que não existem grandes variações no eixo Y quando se analisa os semivariogramas para as diferentes direções traçadas, por isso considerou-se que a variação do fenômeno em estudo é isotrópica.

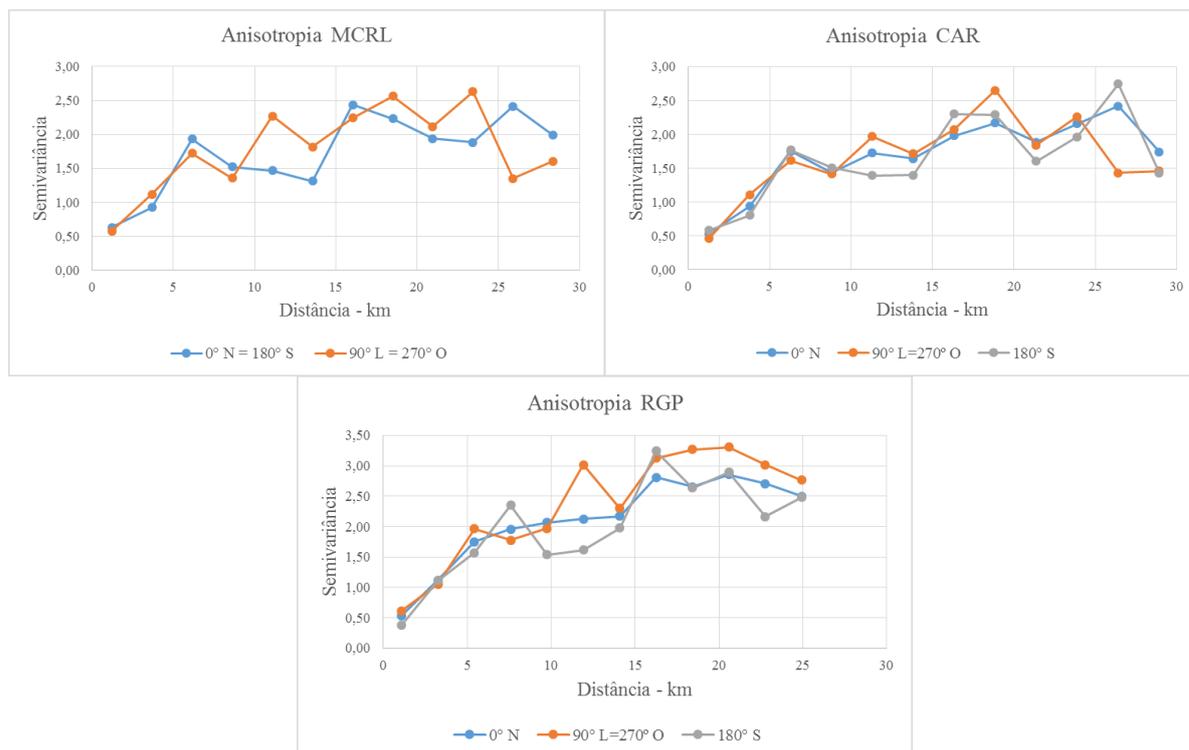


Figura 39 a, b, c. Anisotropia para o MCRL, CAR e RGP

4.7.5 Avaliação de desempenho da Geoestatística

Após a obtenção das superfícies de valores por Krigagem ordinária, foi necessária uma nova avaliação de desempenho, para definir qual dos modelos (MCRL, CAR ou RGP) representou com mais precisão o mercado de terras na área do PROJIR. Para isto foi utilizada a Raiz Quadrada do Erro Médio Quadrático (REMQ) como proposto por Marques et al. (2012) por meio da Equação 36. Além disso foi analisado novamente as métricas determinadas pela IAAO (COD, PRC e Mediana das razões).

$$REMQ = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (VU_{estimado} - VU_{campo})^2} \quad (36)$$

O valor do VU_{campo} é o valor unitário real da amostra de verificação, composta pelas 25 observações que já foram utilizadas na avaliação de desempenho dos modelos. O valor do $VU_{estimado}$ foi obtido a partir dos valores interpolados pela Krigagem nas superfícies de valores para as mesmas coordenadas dos pontos de verificação. Os resultados para cada superfície obtida na etapa anterior estão disponíveis na Tabela 35 e mostram que a Krigagem realizada a partir do modelo de regressão geograficamente ponderada foi a que proporcionou o menor valor da REMQ dentre todas as superfícies obtidas.

Tabela 35. Valores da REMQ para cada modelo após Krigagem

Superfície de Valores - Krigagem	REMQ (R\$/ha)
MCRL	16.216,64
CAR	16.132,02
RGP	15.981,62

Para adequar a análise às normas internacionais foram calculados os parâmetros determinados pela IAAO usando agora os valores preditos pela Krigagem ordinária e os valores de campo das 25 observações. Os resultados estão representados na Tabela 36, com os valores saneados pelo intervalo interquartil de 1,5. Os resultados mostram que a regressão geograficamente ponderada, após os valores serem saneados, obteve o menor valor para COD e para o PRD. Os parâmetros de referência para avaliação do desempenho mostrados na Tabela 23 e recomendados pela IAAO, foram utilizados para analisar os resultados alcançados pela Geoestatística. Para os modelos MCRL e CAR os valores saneados não mudam, pois, todos as amostras ficaram dentro do intervalo interquartil de 1,5.

Tabela 36. Resultados das métricas para avaliação de desempenho pós Krigagem

Modelo	COD (%)	PRD	Mediana das Razões	COD (%) Saneado	PRD Saneado
MCRL	36,16	0,72	1,35	36,16	0,72
CAR	36,73	0,73	1,34	36,73	0,73
RGP	39,56	0,73	1,33	34,88	0,69

Com relação ao COD os três modelos não atenderam um valor menor que 20%, ficando todos acima do recomendado. Entretanto após o saneamento da amostra o modelo RGP obteve o menor valor. Para a mediana das razões os três modelos não atenderam os valores recomendados, e o valor do PRD mostra a existência de progressividade na avaliação, ou seja, os imóveis de alto valor são avaliados com uma porcentagem maior do valor de mercado que os imóveis de baixo valor.

Os resultados mostram que, após a aplicação da Krigagem ordinária através da Geoestatística, o modelo que obteve os menores valores para o REMQ, COD, PRD e mediana das razões foi o RGP, sendo este, o modelo mais indicado para a PVG na área em estudo.

Cabe ressaltar que os valores das métricas da IAAO para os três modelos calculados após a Krigagem, em comparação com antes da Krigagem (Tabela 25) tiveram um desempenho inferior, ou seja, as métricas pioraram após a interpolação. Uma explicação para este problema, é que a Krigagem foi realizada com os VU como variável dependente e com as coordenadas UTM como variáveis independentes, enquanto que antes da interpolação os valores das amostras foram calculados levando-se em consideração as variáveis explicativas de cada elemento (vias de acesso, distância a polo de valorização, distância ao mar, açudes, área de vegetação nativa, área total, situação, topografia, área irrigada e cultura), que deixou os valores mais próximos do real e com variação menor.

4.8 Elaboração da PVG Através do Estimador Kernel

Nesta etapa para elaborar as superfícies de valores na área do PROJIR, foi utilizado o conjunto dos valores estimados pelos modelos MCRL, MRE e RGP, através do estimador

Kernel, utilizando o *software* ArcGIS 10.2.1 através do módulo *Geostatistical Analyst*. Para aplicar o estimador Kernel é necessário definir o raio de influência e a função de estimação. Foi testada a distância de 28 km determinada no item 4.5 como a largura ótima para a regressão geograficamente ponderada, que utiliza o estimador Kernel, entretanto as superfícies geradas se mostraram muito suavizadas.

Para contornar este problema e determinar qual raio de influência e função de estimação gerou os melhores resultados foi utilizada a validação cruzada, através da comparação dos parâmetros Erro Médio, RMS, Erro Médio Padronizado, RMS padronizado e Erro Padrão Médio (*Average Standard Error*).

4.8.1 Superfície dos valores estimados com MCRL

Para determinar a superfície de preços com os valores previstos pelo modelo MCRL no ArcGIS 10.2.1 foi utilizado o contorno dos imóveis como dados de área (*Lattice*), e os resultados da validação cruzada podem ser visualizados na Tabela 37. Em todas as funções, os valores dos erros ficaram muito próximos, o que não permite a conclusão de qual função é a mais apropriada, então foi utilizado o recurso da visualização das superfícies geradas. As funções exponencial, Gaussiano, quártico e constante deixaram a superfície muito suavizada, sendo que o mesmo não aconteceu com a Epanechnikov e com a polinomial 5ª ordem. Reduzindo a análise para estas duas funções, pode-se concluir que a função Epanechnikov obteve os melhores resultados na validação cruzada.

Tabela 37. Valores da validação cruzada das funções Kernel para MCRL

Função	Erro Médio	RMS	Erro Médio Padronizado	RMS Padronizado	Erro Padrão Médio
Exponencial	-0,01799	0,38623	-0,03754	1,03948	0,37766
Gaussiano	-0,02951	0,39539	-0,06252	1,00278	0,40169
Quártico	-0,02313	0,39302	-0,05062	1,01238	0,39351
Epanechnikov	0,01411	0,38519	0,04587	1,05611	0,35644
Polinomial 5ª Ordem	0,01560	0,38178	0,04813	1,08868	0,34335
Constante	-0,03483	0,40071	-0,07401	0,99397	0,40984

Na Figura 40 pode-se observar o gráfico dos valores preditos da função Epanechnikov e polinomial 5ª ordem, para a validação cruzada do ArcGIS. Como comentado anteriormente os valores dos erros são muito próximos, entretanto para a função Epanechnikov o valor do RMS padronizado ficou mais próximo de 1 (um).

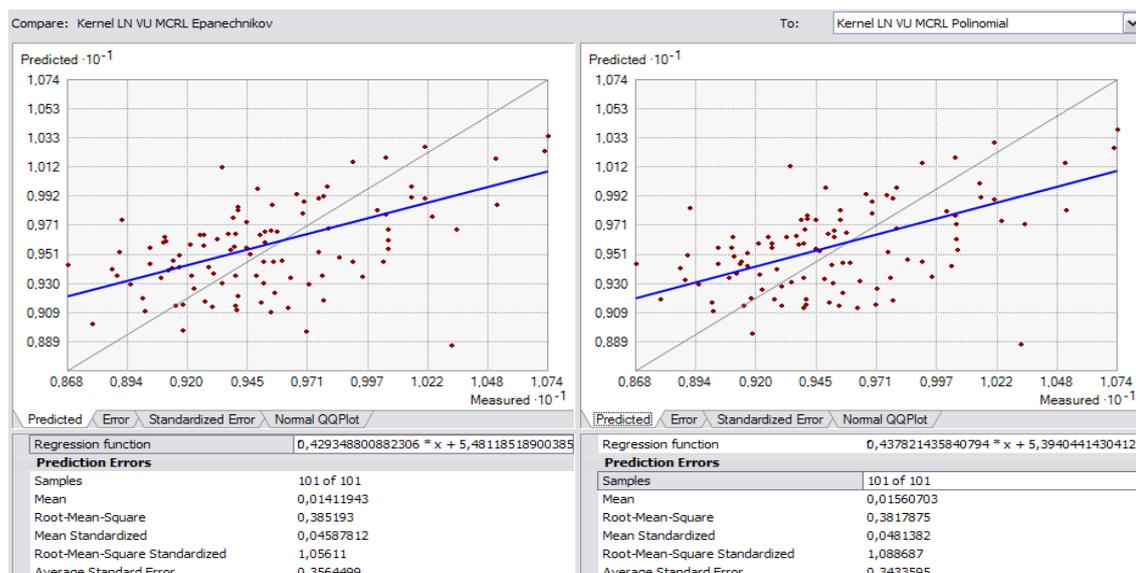


Figura 40. Validação cruzada para MCRL

Na Figura 41, está representada a superfície de valores percebe-se como o Kernel conseguiu representar a variação de preços que ocorre na área de estudo, com os valores previstos pelo MCRL. Em alguns locais das bordas da área, o Kernel não conseguiu projetar os valores, devido à falta de amostras. Os maiores valores estão relacionados à proximidade às áreas urbanas de alguns municípios do PROJIR: São João da Barra, Campos dos Goytacazes e Quissamã. Na área à direita também ocorrem valores mais altos, devido à proximidade com o litoral e com o Porto do Açu, em São João da Barra.

O mapa da previsão do erro do Kernel (Figura 42) mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrais e cresce à medida que se afasta destes pontos. Nas bordas o erro foi maior devido à ausência de amostras, o que evidencia a relevância, quando se trata da elaboração da PVG, da maior quantidade possível de dados e uma boa espacialização, para se obter uma superfície de valores mais precisa e confiável.

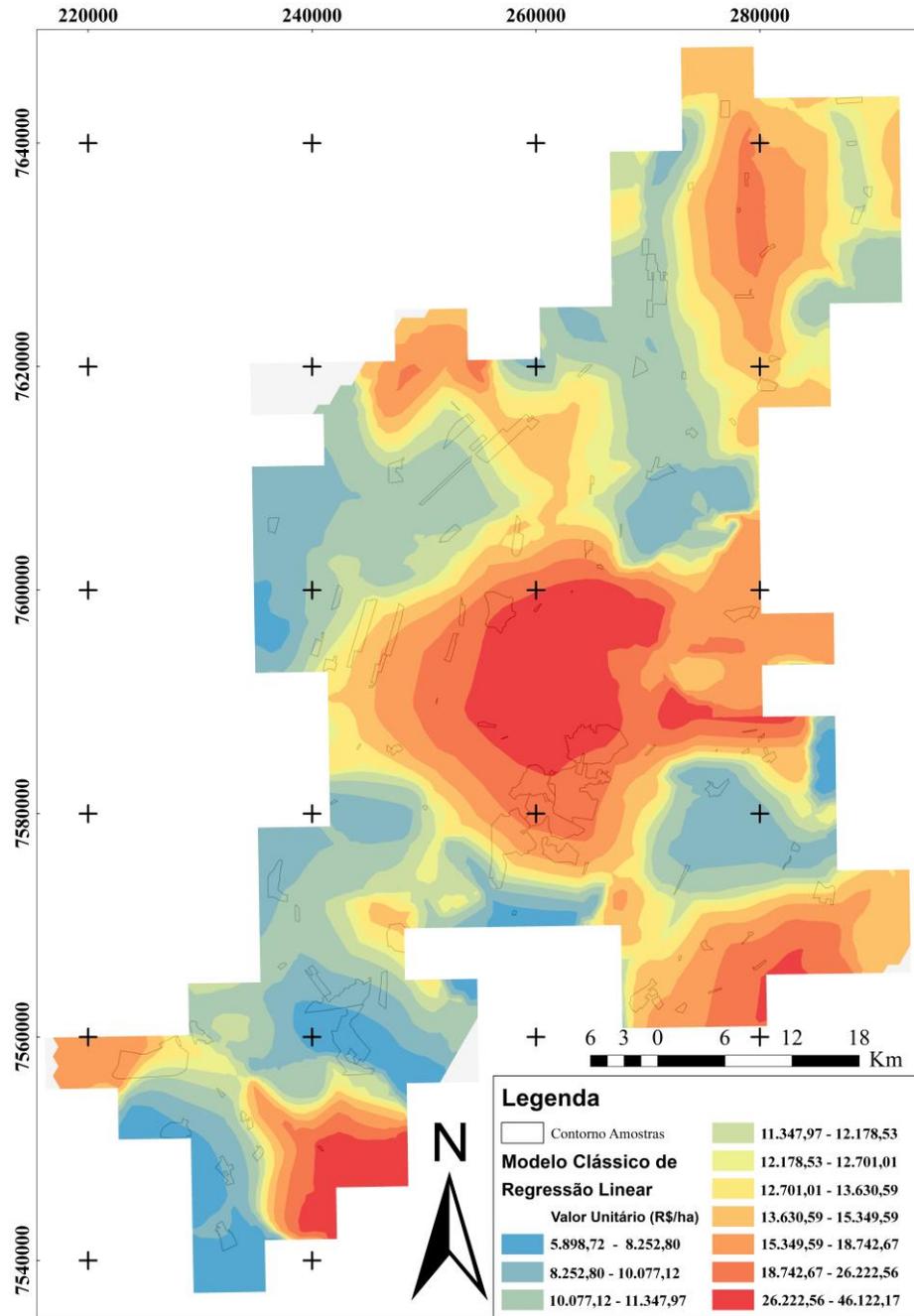


Figura 41. Mapa do Kernel para o MCRL

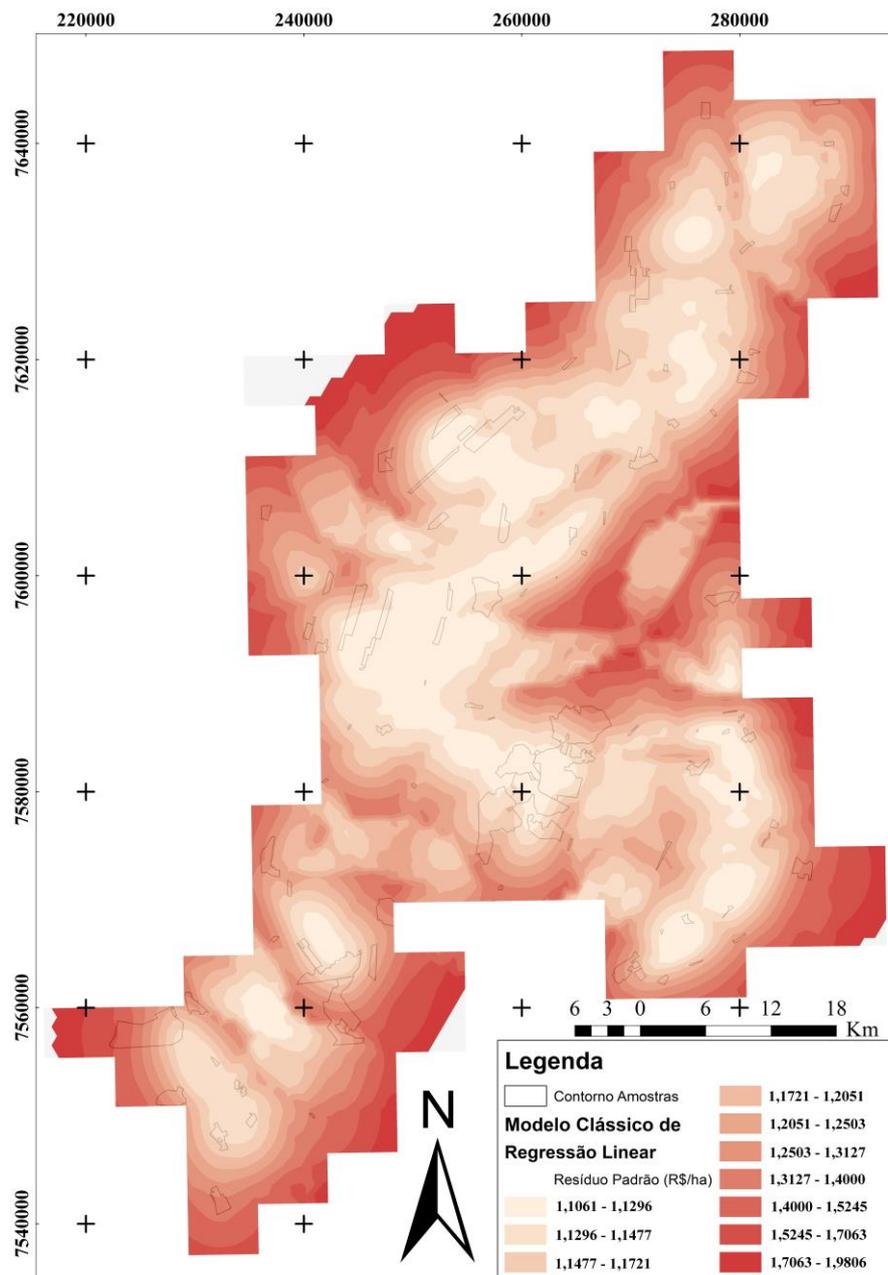


Figura 42. Mapa da previsão do erro para o MCRL

4.8.2 Superfície dos valores estimados com CAR

Para determinar a superfície de preços com os valores previstos pelo modelo CAR no ArcGIS 10.2.1 foi utilizado o contorno dos imóveis como dados de área (*Lattice*), e o resultado da validação cruzada pode ser visualizado na Figura 43. Da mesma forma que para o MCRL, os valores dos erros ficaram muito próximos em todas as funções, e novamente as funções exponencial, Gaussiana, quártico e constante deixaram a superfície muito suavizada, sendo que o mesmo não aconteceu com a Epanechnikov e com a polinomial 5ª ordem. Reduzindo a análise para estas duas funções, pode-se concluir que a função Epanechnikov obteve os melhores resultados na validação cruzada, pois para a função Epanechnikov o valor do RMS padronizado ficou mais próximo de 1 (um).

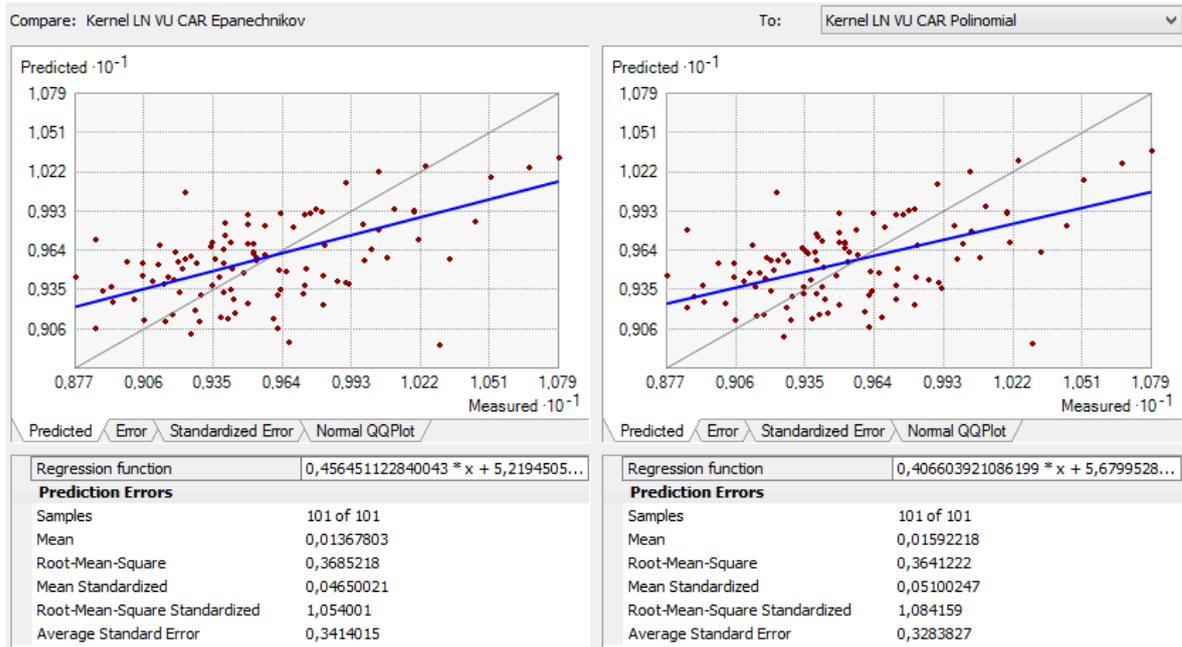


Figura 43. Validação cruzada para o CAR

Na Figura 44, onde está representada a superfície de valores com os valores preditos pelo CAR, percebe-se como o Kernel possibilitou a representação da variação de preços que ocorre na área de estudo, entretanto de uma forma mais suavizada e menos detalhada que o MCRL.

O mapa da previsão do erro do Kernel (Figura 45) também mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrais e cresce à medida que se afasta destes pontos. Entretanto existem mais áreas com menores valores de erro quando comparadas com o mapa de previsão do erro gerado pelo MCRL.

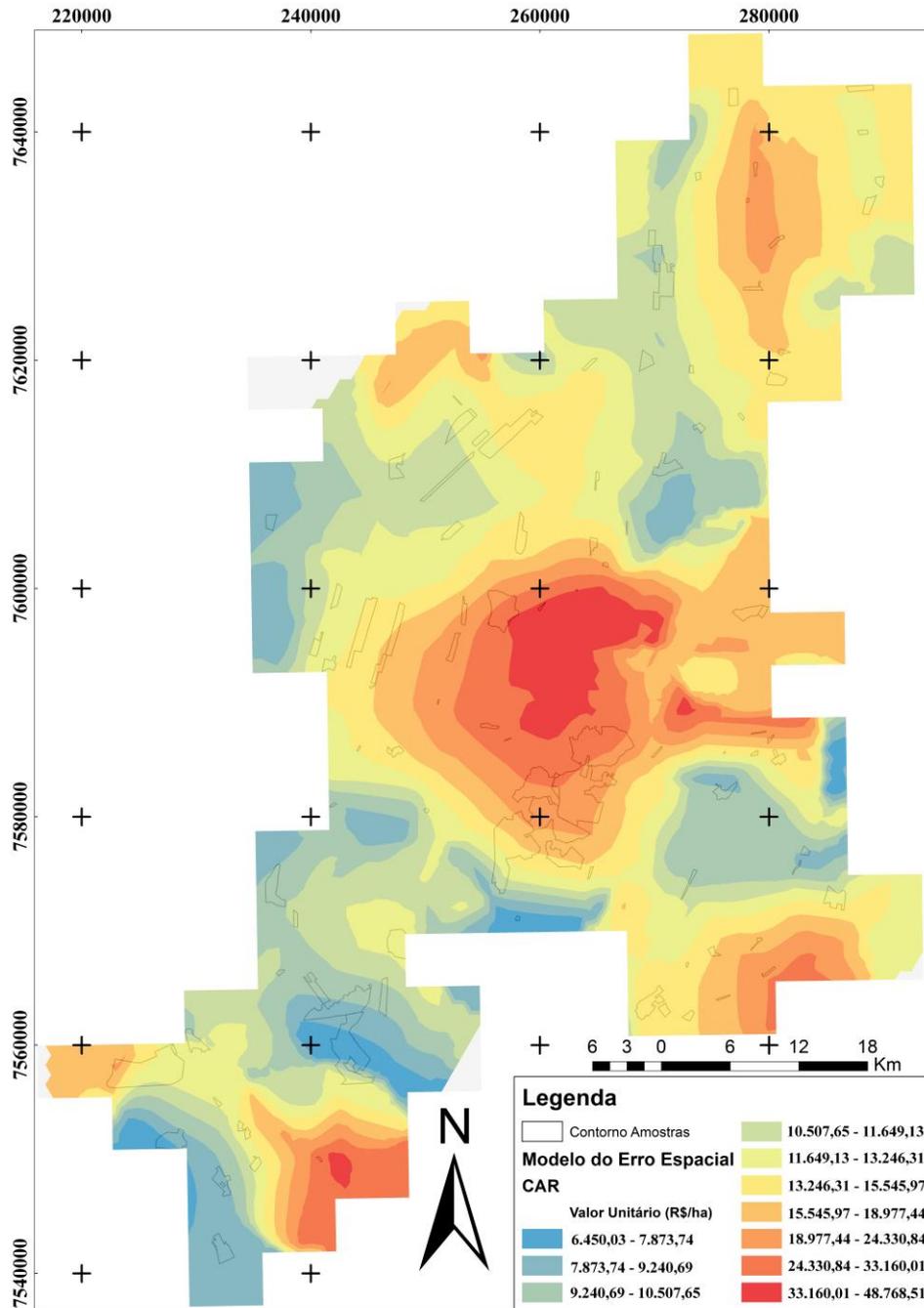


Figura 44. Mapa do Kernel para o CAR

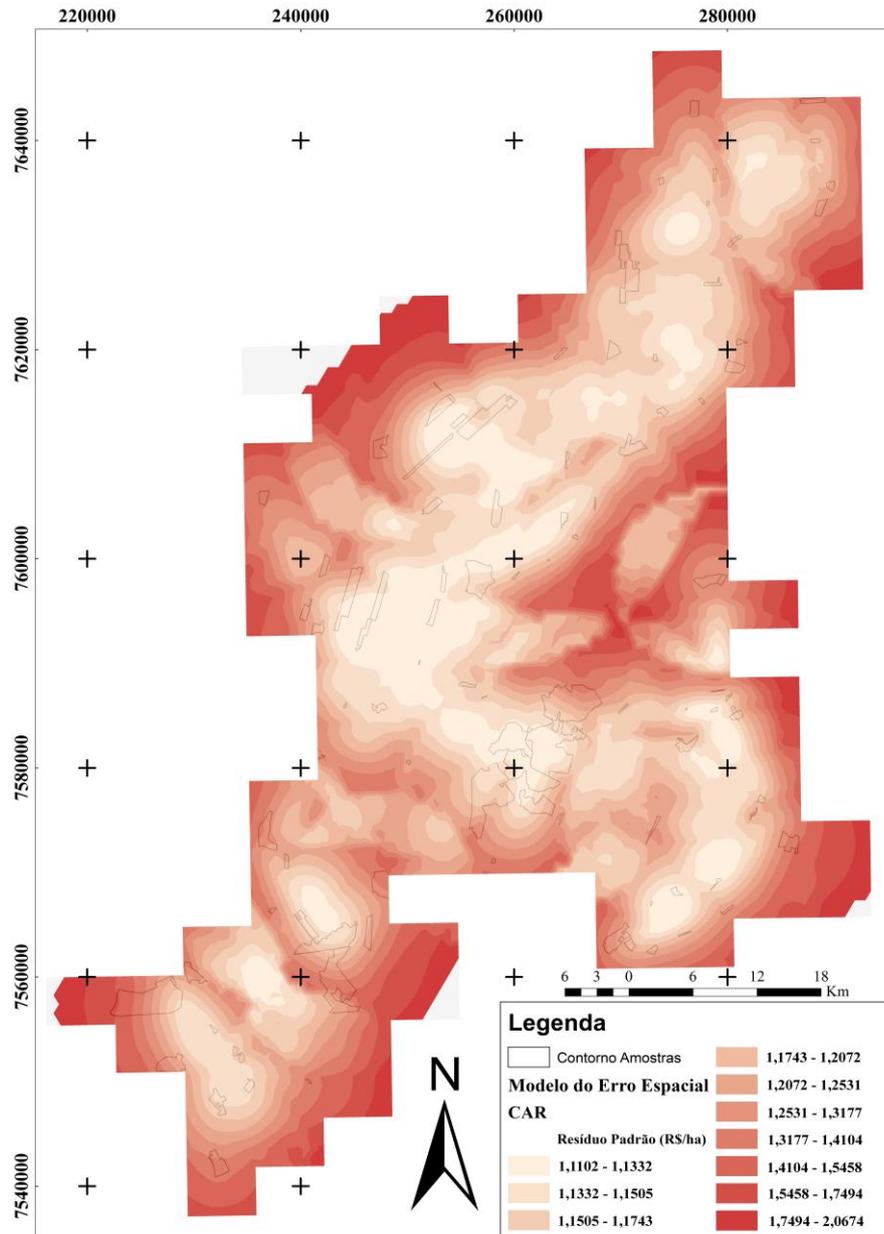


Figura 45. Mapa da previsão do erro para o CAR

4.8.3 Superfície dos valores estimados com RGP

Para determinar a superfície de preços com os valores previstos pelo modelo RGP no ArcGIS 10.2.1 foi utilizado o contorno dos imóveis como dados de área (*Lattice*), e o resultado da validação cruzada pode ser visualizado na Figura 46. Da mesma forma que para o MCRL e CAR, os valores dos erros ficaram muito próximos em todas as funções, e novamente as funções exponencial, Gaussiana, quártico e constante deixaram a superfície muito suavizada, sendo que o mesmo não aconteceu com a Epanechnikov e com a polinomial 5ª ordem. Reduzindo a análise para estas duas funções, pode-se concluir que a função Epanechnikov obteve os melhores resultados na validação cruzada, pois para a função Epanechnikov o valor do RMS padronizado ficou mais próximo de 1 (um).

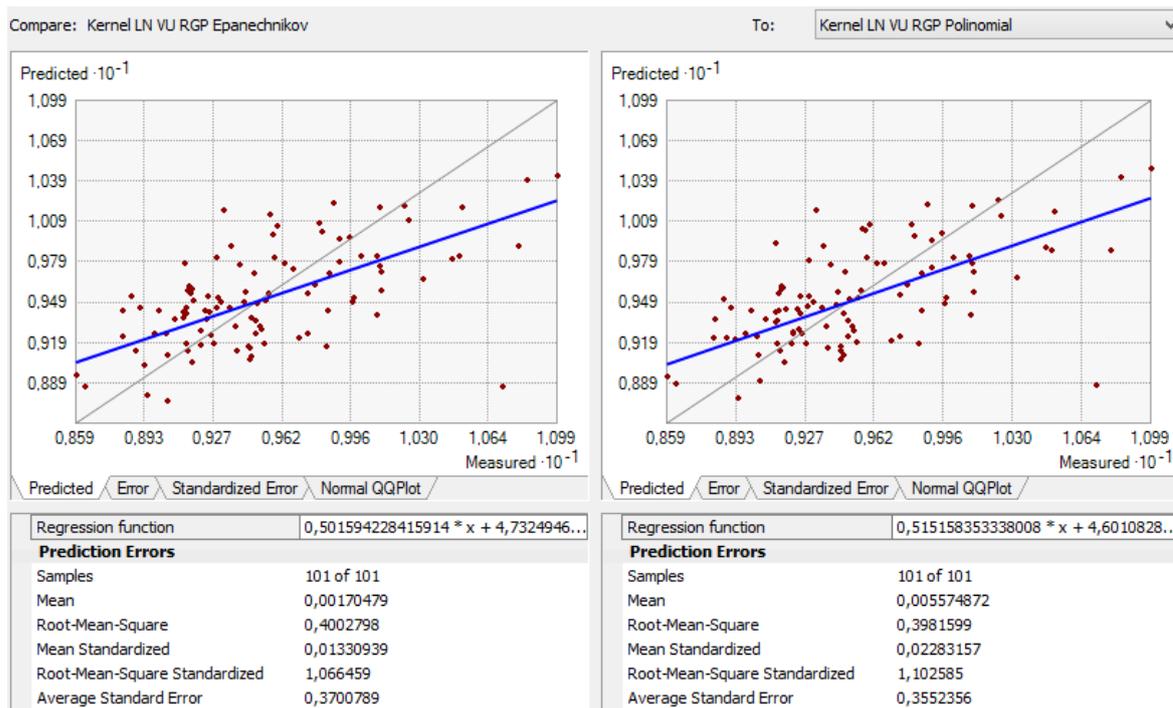


Figura 46. Validação cruzada para o RGP

Na Figura 47, onde está representada a superfície de valores percebe-se que o Kernel conseguiu representar a variação de preços que ocorre na área de estudo, com os valores previstos pelo RGP, representando de uma forma mais detalhada que o CAR.

O mapa da previsão do erro do Kernel (Figura 48) também mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrais e cresce à medida que se afasta destes pontos. Entretanto existem mais áreas com menores valores de erro (mais claras) quando comparadas com o mapa de previsão do erro gerado pelo MCRL e pelo CAR. Provavelmente isso ocorreu, porque o modelo de regressão geograficamente ponderada conseguiu tratar e diminuir esta variação muito mais que os dois modelos anteriores.

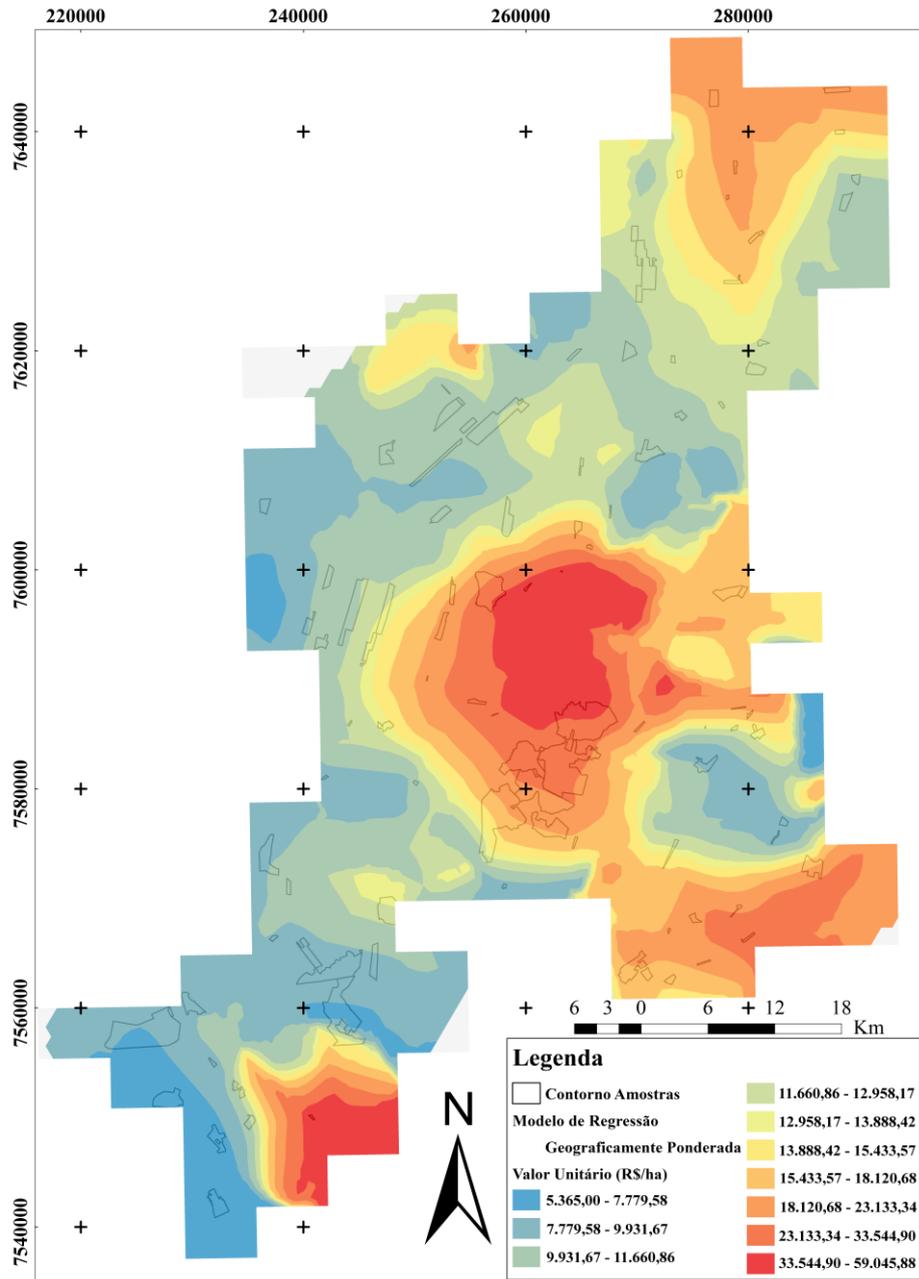


Figura 47. Mapa do Kernel para o RGP

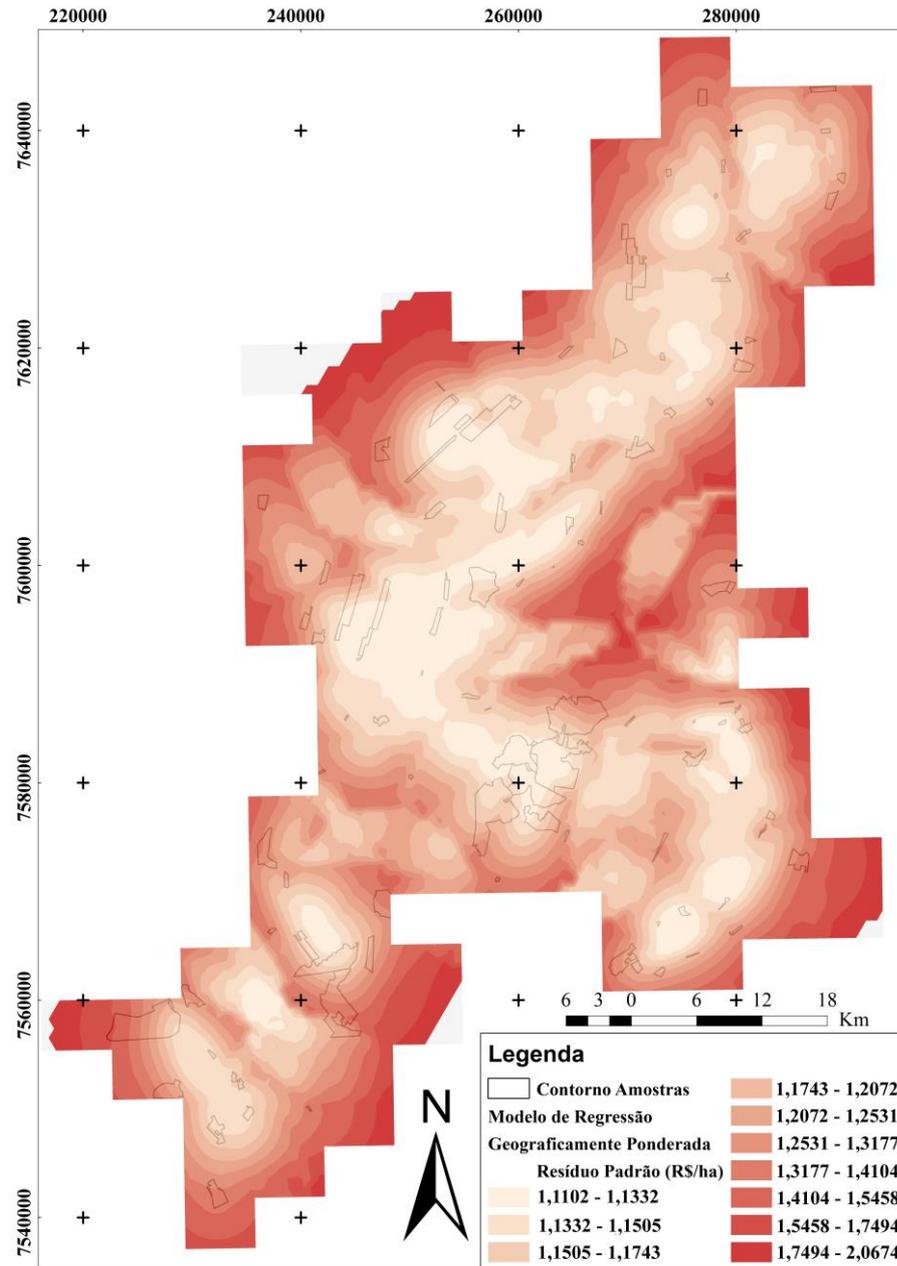


Figura 48. Mapa da previsão do erro para o RGP

4.8.4 Avaliação de desempenho do estimador Kernel

Após a obtenção das superfícies de valores através do estimador Kernel, foi necessária a avaliação de desempenho, para definir qual dos modelos (MCRL, CAR ou RGP) modelou com mais precisão o mercado de terras na área do PROJIR. Para isto foi utilizada novamente a Raiz Quadrada do Erro Médio Quadrático (REMQ) por meio da Equação 36. Além disso foram analisadas novamente as métricas determinadas pela IAAO (COD, PRC e Mediana das razões).

O valor do VU_{campo} é o valor unitário real da amostra de verificação, composta pelas 25 observações que já foram utilizadas na avaliação de desempenho dos modelos. O valor do VU_{estimado} foi obtido a partir dos valores interpolados pelo Kernel nas superfícies de valores

para as mesmas coordenadas dos pontos de verificação. Os resultados para cada superfície obtidos na etapa anterior (Tabela 38) mostram que o Kernel realizado a partir dos valores previstos pelo modelo RGP foi a que proporcionou o menor valor da REMQ em comparação com os outros modelos.

Comparando-se com os valores da Krigagem ordinária (Tabela 35) com os valores do Kernel para o REMQ, observa-se que o Kernel melhorou os valores da superfície de valores, pois diminuiu em aproximadamente 4 % o valor da raiz quadrada do erro médio quadrático. Para o modelo RGP esta diminuição foi de 4,47 %.

Tabela 38. Valores da REMQ para cada modelo com Kernel

Superfície de Valores - Kernel	REMQ (R\$/ha)	Variação % REMQ (Kernel/Krigagem)
MCRL	15.544,65	- 4,14
CAR	15.459,72	- 4,16
RGP	15.267,37	- 4,47

Para adequar a análise às normas internacionais foram calculados os parâmetros determinados pela IAAO usando agora os valores preditos pelo Kernel e os valores de campo das 25 observações. Os resultados estão representados na Tabela 39, com os valores saneados pelo intervalo interquartil de 1,5. Os resultados mostram que a regressão geograficamente ponderada, após os valores serem saneados, obteve o menor valor para COD e para o PRD. Os parâmetros de referência para avaliação do desempenho mostrados na Tabela 23 e recomendados pela IAAO, foram utilizados para analisar os resultados alcançados pelo Kernel.

Tabela 39. Resultados das métricas para avaliação de desempenho pós Kernel

Modelo	COD (%)	PRD	Mediana das Razões	COD (%) Saneado	PRD Saneado
MCRL	36,08	0,80	1,28	-	-
CAR	36,64	0,80	1,28	-	-
RGP	40,46	0,77	1,29	36,28	0,73

Com relação ao COD os três modelos não atenderam um valor menor que 20%, ficando todos acima do recomendado. Entretanto após o saneamento da amostra o modelo RGP obteve o menor valor. Para a mediana das razões os três modelos não atenderam os valores recomendados, e o valor do PRD mostra a existência de progressividade na avaliação, ou seja, os imóveis de alto valor são avaliados com uma porcentagem maior do valor de mercado que os imóveis de baixo valor.

Os resultados mostram que, após a aplicação do interpolador Kernel, o modelo que obteve os menores valores para o REMQ, COD, PRD e mediana das razões foi o RGP, sendo este, o modelo mais indicado para a PVG na área em estudo.

Assim como na Krigagem ordinária, para o Kernel os valores das métricas da IAAO para os três modelos após a interpolação tiveram um desempenho inferior comparando-se com os valores antes da interpolação. A explicação é a mesma para esta limitação quando utilizada a Krigagem, pois o interpolador Kernel somente utiliza os valores dos imóveis calculados pelos modelos de cada ponto amostral, ou seja, o valor de cada ponto corresponde

apenas à ocorrência do evento. Isto se deve ao fato de ser um estimador de intensidade e da quantidade de vizinhos que estão dentro do raio de influência, enquanto que antes da interpolação os valores unitários das amostras de verificação foram calculados levando-se em consideração as variáveis explicativas de cada elemento (vias de acesso, distância a polo de valorização, distância ao mar, açudes, área de vegetação nativa, área total, situação, topografia, área irrigada e cultura), que deixou os valores mais próximos do real e com uma menor dispersão.

4.9 PVG: Krigagem Ordinária ou Estimador Kernel

A PVG como o nome indica, é um produto cartográfico onde são representados os valores generalizados dos imóveis de uma determinada área. Portanto para o cálculo do valor de cada imóvel deve-se realizar um Cadastro Técnico Rural, para considerar as características físicas e locacionais individualizadas dos imóveis e conseqüentemente para o cálculo dos valores do ITR.

Na Figura 37 está representada a PVG, para o modelo RGP e elaborada através da Krigagem ordinária. Na Figura 47 está representada a PVG, para o modelo RGP e elaborada através do interpolador Kernel. Para ambas PVGs, os maiores valores aconteceram nas proximidades às áreas urbanas dos municípios de São João da Barra, Campos dos Goytacazes e Quissamã. Na área à direita também ocorrem valores mais altos, devido à proximidade com o litoral e com o Porto do Açú, em São João da Barra. Na parte central da área, acima de Campos também ocorre valores altos, em função da proximidade da BR 101, que corta o município na direção norte.

A interpolação por Krigagem e pelo Kernel permitiu a determinação dos valores entre os vizinhos, pois em áreas rurais dificilmente se obtém dados na coleta em campo para toda a área estudada. Além disso, a utilização da Krigagem/Kernel permitiu a detecção dos fatores que influenciam o resultado da interpolação, como a disposição das amostras, a proximidade e a quantidade de dados, pois cabe lembrar que por ser uma interpolação a falta de amostras ou a ocorrência de vizinhos muito afastados podem gerar erros na estimação dos valores unitários.

As técnicas de estimação pela Krigagem ordinária e pelo estimador Kernel mostraram-se eficientes em representar a distribuição espacial dos valores dos imóveis na área do PROJIR. Comparando-se os valores da REMQ e do COD para o modelo RGP, da Krigagem e do Kernel, percebe-se que: a) o valor do REMQ diminuiu 4,7 % utilizando o estimador Kernel; b) o valor do COD aumentou 1,6 % utilizando o estimador Kernel.

Apesar da melhora no valor do REMQ, o Kernel é um estimador de intensidade, ou seja, é uma função bidimensional que considera os eventos dentro de uma região, fazendo a contagem dos pontos ali contidos e indicando a área de maior ou menor concentração dos eventos analisados, enquanto a Krigagem requer a parametrização da estrutura de correlação espacial, através do semivariograma.

Para determinar qual das técnicas apresentou a melhor superfície de valores, foi calculado o erro relativo, definido como o quociente entre o resíduo e o valor observado para cada dado da amostra de validação, e o valor absoluto deste erro relativo, ambos em valores percentuais. Os resultados podem ser observados na Tabela 40 e mostra que o estimador Kernel obteve estimativas mais ajustadas aos dados de mercado, com resíduos relativos e absolutos menores em média e dispersão. O erro relativo mínimo diminuiu 5,99%, o erro máximo diminuiu 5,02%, a média e desvio padrão dos erros relativos absolutos diminuíram em 1,26% e 1,98% respectivamente.

Sendo assim, foi escolhida a superfície de valores que apresentou o menor REMQ, e menor erro relativo para representar a PVG final da área do PROJIR, ou seja, a superfície gerada a partir do estimador Kernel do modelo RGP, que pode ser visualizada na Figura 49.

Tabela 40. Resíduo relativo e absoluto da Krigagem e Kernel

Estimador	Erro Relativo - ER (ER=Resíduo/Valor Observado) em %					Valor Absoluto do ER em %	
	Mínimo	Quartil Inferior	Mediana	Quartil Superior	Máximo	Média	Desvio Padrão
Krigagem	-126,36	-26,58	10,88	25,37	80,05	36,49	30,05
Kernel	-120,37	-19,50	8,87	26,45	75,03	35,23	28,07

A utilização de uma amostra de verificação que não faz parte na determinação dos modelos de regressão, foi imprescindível pois permitiu a verificação do ajuste alcançado e indicou quais são mais adequados e tiveram melhor desempenho. As amostras de verificação ainda indicaram qual interpolador foi o mais ajustado, para os dados coletados na região em estudo.

A metodologia proposta mostra a importância de contar com dados cadastrais georreferenciados e atualizados. A implementação de Sistemas de Informações Geográficas (SIG) nas Prefeituras Municipais pode facilitar a aplicação do método na elaboração e atualização das Plantas de Valores Genéricos e em concordância com valores de mercado. Além disso, a maioria dos *softwares* de SIG contêm módulos de análise Geoestatística incorporados, facilitando a aplicação desta metodologia.

A PVG elaborada por métodos estatísticos é um instrumento de justiça fiscal, pois permite que a avaliação dos imóveis seja realizada com base nos valores de mercado imobiliário e nas variações locais dos preços nas diferentes regiões, evitando desta forma tanto as subavaliações que torna a arrecadação dos impostos deficiente, bem como valores acima do mercado que punem os proprietários.

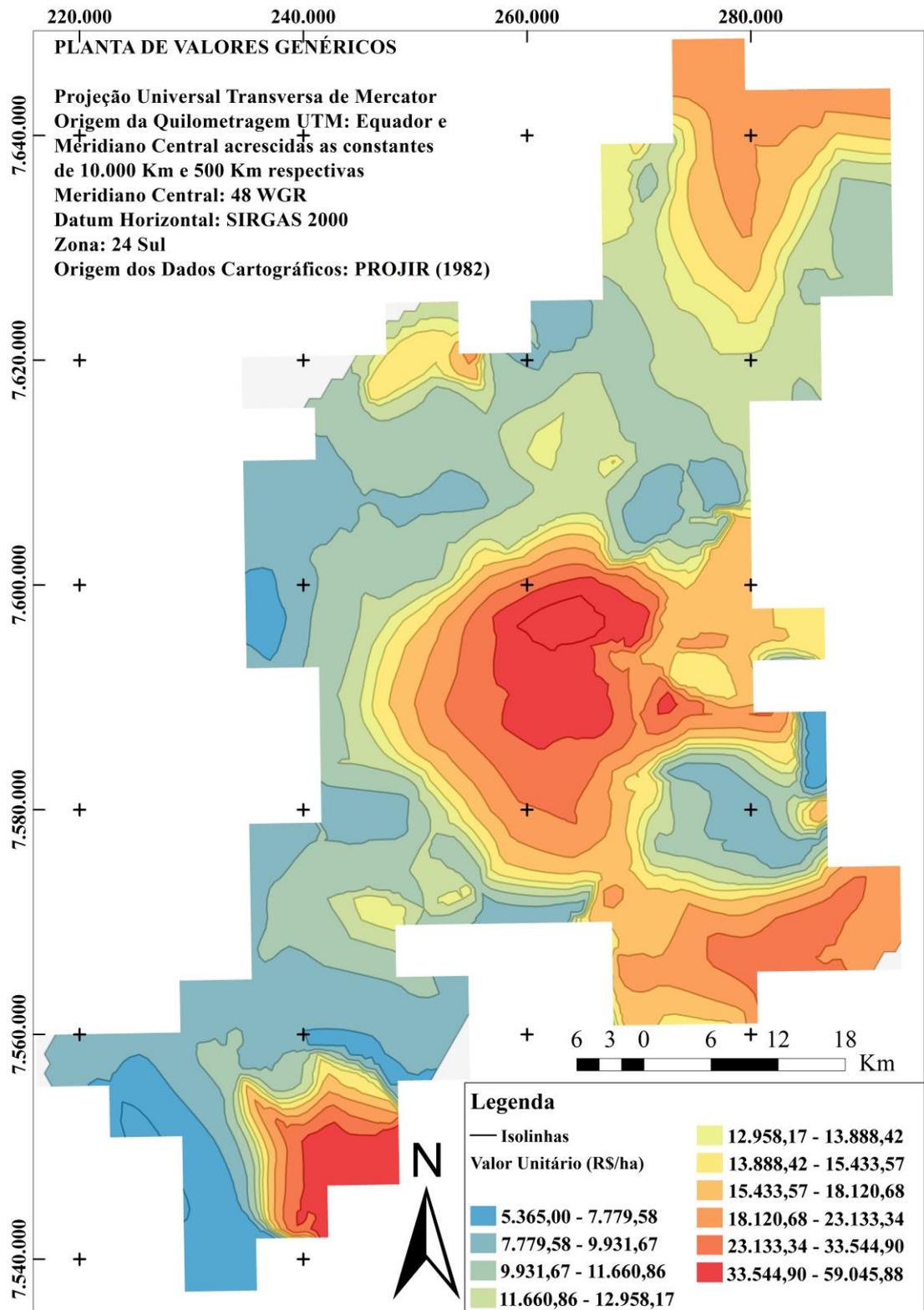


Figura 49. Planta de Valores Genéricos da área do PROJIR

5 CONCLUSÕES

Sobre as variáveis do modelo geral de regressão

A obtenção dos dados de mercado para realizar avaliações de imóveis rurais é uma tarefa difícil, pois a quantidade de dados nem sempre é satisfatória devido à escassez de transações para este tipo de imóvel. Se os municípios mantivessem um cadastro técnico rural georreferenciado e atualizado, estas dificuldades poderiam ser reduzidas, pois uma pesquisa no mercado imobiliário e uma consulta no sistema cadastral bastariam para compor uma amostra de imóveis. Para este trabalho, a etapa de coleta de dados de mercado, não se mostrou diferente dos problemas enfrentados em outros estudos de avaliação imobiliária, seja em áreas urbanas ou rurais. O acesso às informações de imóveis comercializados, através das imobiliárias e corretores nem sempre gera resultados e na maioria das prefeituras os valores praticados na cobrança do ITBI são desatualizados.

Após a análise dos dados que poderiam ser utilizados e da seleção de quais variáveis influenciariam, foi realizada a análise exploratória dos dados antes de iniciar a modelagem propriamente dita. Determinar quais transformações poderiam ser aplicadas para as diferentes variáveis obtidas, os diagramas de dispersão bem como a matriz de correlações, entre outras análises, permitiram a construção de modelos mais consistentes e evitou problemas durante a etapa de validação dos pressupostos dos modelos.

As variáveis utilizadas nos modelos de regressão se mostraram significativas e coerentes com as suposições estabelecidas no item 4.1, sendo que a variável Cultura foi a exceção, pois esperava-se que os imóveis cultivados com cana-de-açúcar seriam mais valorizados, o que não foi confirmado pelo modelo de regressão. No entanto percebe-se que este fato pode ser consequência do processo de decadência da cultura de cana-de-açúcar na região Norte Fluminense.

Com relação às variáveis relacionadas à presença de açudes e de áreas de vegetação nativa, os modelos demonstraram que a presença destas duas benfeitorias valoriza os imóveis na área de estudo. Este resultado mostra a importância da manutenção da vegetação nativa, seja na forma de reserva legal, áreas de preservação permanente, ou de reserva particular de proteção natural, para o valor dos imóveis rurais.

Sobre os modelos de regressão: MCRL, MRE e RGP

Com relação aos modelos de regressão foram construídos o modelo clássico, espacial e geograficamente ponderado. No MCRL não foram utilizadas variáveis relacionadas a localização geográfica das amostras e foi constatada a heterocedasticidade (variância dos erros não constante) provocada pela presença de dependência espacial nos resíduos do modelo. Partiu-se então para a modelagem dos efeitos espaciais através da econometria espacial.

Para determinar qual a matriz de pesos espaciais que capturasse a maior parte da autocorrelação espacial foram testadas várias matrizes de vizinhança, sendo que a mais significativa pelo teste I de Moran foi a matriz da distância Euclidiana com 7 km, para a área em estudo.

Através dos testes focados foi identificado que a autocorrelação espacial estava no termo do erro, ou seja, os resultados mostram a existência de forte autocorrelação espacial nos resíduos do modelo de regressão por mínimos quadrados, pois os dois testes do modelo

espacial do erro (LM e o LM robusto) apresentaram-se fortemente significativos.

Para detectar onde a autocorrelação espacial ocorre, se no termo da variável dependente ou no termo do erro, foi realizada a regressão com o modelo vazio. O teste com o modelo vazio, confirmou que ocorreu predominância de autocorrelação espacial no termo do erro, o mesmo resultado detectado com o Multiplicador de Lagrange.

Conclui-se que o modelo econométrico espacial cumpriu o seu papel, controlando a autocorrelação espacial nos resíduos do MCRL, pois a variável LAMBDA – λ , que representa o termo de autocorrelação espacial do erro, foi significativa estatisticamente.

Para o modelo RGP foram testados os Kernel fixo e adaptativo, sendo que o Kernel biquadrado de largura fixa, foi o que apresentou o menor valor para o critério *Akaike* (AIC), sendo esta a função matemática utilizada para gerar o modelo de regressão.

Os três modelos foram avaliados na qualidade de ajuste e desempenho. Pelos critérios *Akaike* (AIC), valor da função de verossimilhança (LIK), coeficiente de determinação (R^2) e erro padrão da regressão. O modelo RGP foi superior ao MCRL e ao CAR em todos os critérios e com isto podemos concluir que a regressão geograficamente ponderada deixou a modelagem mais ajustada e, portanto, foi superior aos outros dois modelos.

Para a avaliação de desempenho foi utilizada uma amostra de validação com 25 elementos, onde calculou-se para as mesmas as métricas definidas pela IAAO. A utilização de pontos de validação do modelo que não fizeram parte da construção dos mesmos merece ser destacada, pois são poucos os trabalhos na área de avaliações que utilizam amostras separadas para este fim.

A avaliação de desempenho analisou o nível geral das avaliações, através da mediana das razões, a dispersão em toda área de estudo pelo COD e questões de progressividade e regressividade, através do Diferencial Relativo do Preço. Com relação à Mediana das Razões, todos os modelos ficaram dentro dos limites de referência estabelecidos pela norma, indicando que o nível geral das avaliações está adequado. Com relação ao COD todos os modelos ficaram com os valores acima dos recomendados, sendo que o modelo RGP foi o que apresentou o pior resultado. Portanto, concluiu-se que isso ocorre porque a área de estudo abrange mais de um município e os valores dos imóveis na maioria dos casos, têm comportamentos diferentes dentro dos seus respectivos limites municipais, ou seja, a análise ao nível local (que é o caso do RGP), considerando uma única área, gerou uma dispersão dos valores maior que as outras técnicas. Com relação ao PRD os três modelos ficaram abaixo dos limites de referência, ou seja, ocorreu progressividade na avaliação.

A avaliação de desempenho e da qualidade do ajuste dos produtos gerados representa uma importante contribuição para os mercados de terras rurais. A utilização de 25 pontos de verificação foi importante para que a avaliação de desempenho fosse capaz de mostrar a eficiência dos modelos, bem como mostrar pontos que necessitam ser aperfeiçoados.

Sobre a Krigagem ordinária e Kernel

Após a obtenção dos três modelos de regressão, foi realizada a modelagem Geoestatística dos valores preditos pelos modelos e finalizando com a interpolação pela Krigagem ordinária e pelo Kernel. Estas interpolações deram origem às superfícies de valores, que também foram submetidas à avaliação de desempenho utilizando os 25 dados de validação utilizados na etapa anterior.

Os resultados mostraram que tanto a Krigagem ordinária como o Kernel podem ser utilizados para gerar a superfície de valores para a área do PROJIR. Foram calculados para os dois interpoladores os valores da REMQ, mediana das razões, do COD e o PRD para

determinar qual dos três modelos gerou a superfície de valores com mais precisão. Para os dois interpoladores o menor valor da REMQ indicou o modelo RGP como o mais ajustado, sendo que para o Kernel este índice ficou ainda menor que na Krigagem ordinária. Os valores das métricas da IAAO também indicaram o RGP como o melhor modelo. Entretanto, comparando-se os valores de antes e após a interpolação esses parâmetros aumentaram de valor, indicando um desempenho inferior. O desempenho da superfície de valores, indica que o produto gerado pelos interpoladores deve ser utilizado para indicação dos valores dos imóveis localizados nas diversas faixas, entretanto no cálculo do valor final dos imóveis o mais indicado é usar o modelo de regressão, que para esta área de estudo é o RGP.

Para finalmente determinar qual dos interpoladores foi o mais indicado para gerar a superfície de valores final para a área do PROJIR, foram calculados o erro relativo e o valor absoluto deste erro relativo. A superfície gerada a partir do estimador Kernel do modelo RGP foi a que obteve o melhor desempenho.

Sobre a avaliação em massa

A metodologia proposta possibilitou obter o modelo de regressão, bem como a elaboração da PVG, ambos representativos da realidade dos valores de mercado para imóveis rurais na área do PROJIR. Para a avaliação em massa de todos os imóveis individualmente, o modelo de regressão RGP foi o mais indicado, pois utiliza as características (variáveis) que influenciam os valores das terras no município. A PVG obtida pela interpolação dos valores amostrais deve ser utilizada para indicar o valor genérico dos imóveis nas diferentes regiões da área do PROJIR.

A metodologia utilizada pode ser implementada pelos municípios, pois com uma equipe para a coleta dos dados amostrais e com a contratação dos serviços de um engenheiro de avaliações, as prefeituras poderão elaborar as PVGs com relativa facilidade. Cabe lembrar que os municípios conveniados com a SRF, devem manter atualizados os valores da terra nua dos imóveis rurais e a utilização da metodologia proposta atribui transparência e justiça fiscal às PVGs.

Apesar de alguns dos *softwares* utilizados nessa pesquisa serem comerciais, eles ser facilmente substituídos por programas livres e gratuitos. Por exemplo o ArcGIS pode ser substituído pelo Quantum Gis (QGIS) e existem pacotes estatísticos como o R para fazer as análises necessárias. Com a utilização de *softwares* livres os investimentos das prefeituras ficariam somente na contratação de profissionais capacitados na elaboração e atualização das PVGs.

A combinação das metodologias propostas foi eficiente para modelar os valores dos mercados de terras rurais. Através da econometria espacial foi possível controlar a autocorrelação espacial existente no MCRL e utilizar a regressão geograficamente ponderada e por meio da Geoestatística, com o interpolador da Krigagem ordinária e Kernel foi verificado que os mesmos, podem ser utilizados para gerar superfícies de valores para os modelos obtidos. A combinação das metodologias tornou possível atender os objetivos do trabalho.

6 CONSIDERAÇÕES FINAIS

Sugere-se para trabalhos futuros utilizar também dados de ITBI, embora haja dificuldade de acesso destas informações junto às prefeituras municipais. Porém, devem ser tomados certos cuidados na atualização dos valores, pois na maioria das prefeituras os mesmos são menores que os praticados no mercado imobiliário.

Neste trabalho, em virtude dos dados cartográficos disponíveis optou-se por trabalhar com grande parte da Região Norte Fluminense, que abrangeu vários municípios. Sugere-se separar os municípios da área do PROJIR, pois ocorrem diferenças de valores muito grandes entre eles e muitas vezes dentro de um mesmo município. Sugere-se buscar a construção de modelos específicos para cada município e com isso, espera-se que ocorra uma menor dispersão nos resultados da avaliação de desempenho dos modelos.

Para melhorar o desempenho da Krigagem ordinária é recomendada a utilização da cokrigagem. A cokrigagem é um procedimento geoestatístico segundo o qual diversas variáveis regionalizadas podem ser estimadas em conjunto, com base na correlação espacial entre si. É uma extensão multivariada do método da Krigagem quando para cada local amostrado obtém-se um vetor de valores em lugar de um único valor. Com isso espera-se utilizar os atributos estruturais e de localização dos imóveis e não somente a utilização dos valores unitários na interpolação para gerar a superfície das PVGs.

Na avaliação em massa de imóveis urbanos trabalhos mais recentes utilizam uma classe de modelos de regressão denominada de modelos aditivos generalizados, métodos de regressão Bayesiano, redes neurais artificiais, etc. Sugere-se testar estas técnicas na avaliação em massa de imóveis rurais para verificar se os mesmos melhoram o desempenho dos valores.

7 REFERÊNCIAS BIBLIOGRÁFICAS

- ALMEIDA, E. **Econometria espacial aplicada**. Campinas. SP: Editora Alínea. 2012.
- AMARAL, F. C. S. **Sistema brasileiro de classificação de terras para irrigação: enfoque na região semiárida**. Editor: Fernando Cesar Saraiva do Amaral. Rio de Janeiro: Embrapa Solos. 2011.
- ANDERSON, J. E.; ENGLAND, R. W. **Estimación del índice entre uso y valor del suelo rural ¿Tiempo de reformas?** Land Lines: Revista Trimestral del Instituto Lincoln de Políticas de Suelo. V. 28. N. 01. Cambridge. 2016.
- ANSELIN, L. **The moran scatterplot as an ESDA tool to assess local instability in spatial association**. In: **Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences**. London: Taylor and Francis. 1996. p. 111-125.
- _____. **Gis research infrastructure for spatial analysis of real estate markets**. Journal of Housing Research. v.9. n.1. p.113-134. 1998.
- _____. **Spatial econometrics**. Bruton Center. School of Social Sciences. University of Texas. Dallas. Richardson. 1999.
- _____. **Spatial externalities, spatial multipliers, and spatial econometrics**. International Regional Science Review 26 (2). 2003.
- _____. GeoDa 1.6.7 9 March 2015. **Exploring spatial data with GeoDa™ : A Workbook**. Spatial Analysis Laboratory, University of Illinois, Urbana-Champaign, 2005.
- ANSELIN, L.; BERA. A. **Spatial dependence in linear regression models with an introduction to spatial econometrics**. In: ULAH A.; GILES D. E. (eds.). Handbook of Applied Economic Statistics. Marcel Dekker: New York. p.237-89. 1998.
- ARANTES, C. A. **ITR Imposto Territorial Rural e suas falhas**. Piracicaba. SP. 2005.
- ARANTES, C. A.; SALDANHA, M. S. **Avaliações de imóveis rurais: norma NBR 14.653-3: ABNT comentada**. São Paulo. Liv. E Ed. Universitária de Direito. 2009.
- ÁRIA, Informática. **INFER versão 32, manual de operação**. Belo Horizonte. MG. 2004.
- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Avaliação de Bens Parte 1: Procedimentos Gerais. NBR 14653-1**. Rio de Janeiro. ABNT. 2001.
- ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **Avaliação de Bens Parte 3: Imóveis Rurais. NBR 14653-3**. Rio de Janeiro. ABNT. 2004.
- AVERBECK, C. E. **Os sistemas de cadastro e plantas de valores no município: prejuízos da desatualização**. 2003. 202 f. Dissertação (Mestrado). Centro Tecnológico. Universidade Federal de Santa Catarina. Florianópolis. SC. 2003.
- BAUMONT, C. **Spatial effects in housing price models: do housing prices capitalize urban development policies In the agglomeration of Dijon (1999)**. Université de Bourgogne. France. 2004. Disponível em: < <http://leg.u-bourgogne.fr/documents-de-travail/e2004-04.pdf>>. Acesso em 10/2015.
- BORST, R.; McCLUSKEY, W. **Comparative evaluation of the comparable sales method**

with Geostatistical valuation models. Pacific Rim Property Research Journal. vol 13. nº 1. 2007.

BRASIL. Secretaria da Receita Federal - SRF. **Análise da arrecadação das receitas federais – dezembro de 2012.** DF. 2012. Disponível em: <<http://www.receita.fazenda.gov.br/publico/arre/2012/Analismensaldez12.pdf>> Acesso em: 11/2013.

BRASIL, **Lei nº 11.250 de 27 de dezembro de 2005.** Diário Oficial da União. Brasília. DF. 2005.

BRASIL. **Lei nº 9.393 de 19 de dezembro de 1996. Imposto Territorial Rural.** Diário Oficial da União. Brasília. DF 1996.

BRASIL. **Constituição da República Federativa do Brasil.** Diário Oficial da União. Brasília. DF. 1988. 292 p.

BRASIL. **Lei 5.172 de 25 de outubro de 1966. Código Tributário Nacional.** Diário Oficial da União. Brasília. DF. 1966.

BRASIL. **Lei nº 4.504 de 30 de novembro de 1964. Estatuto da Terra.** Diário Oficial da União. Brasília. DF. 1964.

BUENO, F. R.; BARRÓS. D. M.; MASEDA, R. C. **Situación de la valoración catastral rural.** CT/Catastro. n. 75. p. 41-54. 2012.

CÂMARA, G.; CARVALHO, **Análise espacial de eventos. In: Análise espacial de dados geográficos. Capítulo 2.** INPE. Instituto Nacional de Pesquisas Espaciais. 2004.

CÂMARA, G.; CARVALHO, M. S.; CRUZ. O.G.; CORREA. V. **Análise espacial de áreas. Análise espacial de dados geográficos. Capítulo 5.** INPE. Instituto Nacional de Pesquisas Espaciais. 2004.

CAMARGO, E. C. G.; MONTEIRO, A. M. V.; FELGUEIRAS. C. A.; FUKS. S. **Integração de Geoestatística e sistemas de informação geográfica: Uma necessidade.** In: 5º Congresso e Feira para Usuários de Geoprocessamento da América Latina (GIS Brasil 99). 1999. Salvador. Anais. (Em Cd-Rom). Curitiba: FatorGIS. 1999.

CAMARGO, E. C. G.; FUCKS, S. D.; CÂMARA, G. **Análise espacial de superfícies. Análise espacial de dados geográficos. Capítulo 3.** INPE. Instituto Nacional de Pesquisas Espaciais. 2004.

CARTA, L. M. C. **Sistema de apoio à decisão: Análise espacial aplicada ao mercado da habitação.** 2014. 101 f. Dissertação (Mestrado). Departamento de Eletrônica, Telecomunicações e Informática. Universidade de Aveiro. Aveiro. Portugal. 2014.

CARVALHO, L. E. X.; SILVA, H. N.; LOUREIRO, C. F. G.; MENESES, H. B. **Regressão linear geograficamente ponderada em ambiente SIG.** ANPET - Associação Nacional de Pesquisa e Ensino em Transportes. TRANSPORTES. v. XIV. n. 2. p. 18-26. Dezembro 2006.

CARVALHO, T. S. **A hipótese da curva de Kuznets ambiental global e o protocolo de Quioto.** 2008. 131 f. Dissertação (Mestrado). Faculdade de Economia e Administração. Universidade de Juiz de Fora. MG. 2008.

CHICA OLMO, J. **Teoría de las variables regionalizadas. Aplicación en economía espacial y valoración inmobiliaria.** Ed. Universidad de Granada. 1994.

_____. **Spatial estimation of housing prices and locational rents.** Urban Studies. 1995. 32:8. 1331-44.

_____. **Prediction of housing location price by a multivariate spatial method: Cokriging.** Journal of Real Estate Research 01/2007; 29(1):95-114

CHICA OLMO, J.; CANO GUERVOS, R.. CHICA OLMO, M. **Modelo hedónico espacio-temporal y análisis variográfico del precio de la vivienda.** GeoFocus (Artículos). nº 7. p. 56-72. España. 2007.

COMARELLA, E. A. **Avaliação de área de preservação permanente e reserva legal em imóvel rural, serviço ambiental e pagamento pelo uso da água.** XVII COBREAP. Florianópolis. SC. 2013.

CRUZ, J. L. V. **Os desafios do Norte e Noroeste Fluminenses frente aos grandes projetos estratégicos.** In: Revista Vértices, Instituto Federal Fluminense de Educação, Ciência e Tecnologia, v. 9, n.1. Campos dos Goytacazes, Essentia Editora, 2007.

DANTAS, R. A.; ROCHA, F. J. S.; MAGALHÃES, A. M.; LIMA, R. C. **Uma nova metodologia para avaliação de imóveis utilizando regressão espacial.** XI COBREAP. Guarapari. ES. 2001.

DANTAS, R. A.; PORTUGAL, J. L.; PRADO, J. F. **Avaliação de cidades por inferência espacial: um estudo de caso para a cidade de Aracaju.** XIII COBREAP. Fortaleza. CE. 2006.

DANTAS, R. A. **Engenharia de avaliações: uma introdução à metodologia científica. Revisada de acordo com a nova versão da NBR 14.653-2:2011.** 3ª edição. São Paulo. Pini. 2014.

DAVIS, P. **The IAAO standard on ratio studies: a framework for order and progress in assessment performance evaluation.** In: Seminário Internacional sobre Tributação Imobiliária, Anais. Porto Alegre, 2001.

DE CESARE, C. M. **An empirical analysis of equity in property taxation: A case study from Brazil.** 1998. PhD Thesis - University of Salford. Salford (UK).

_____. **Panorama do imposto sobre a propriedade imobiliária na América Latina.** Lincoln Institute of Land Policy. Working Paper. Cambridge. 2010.

DE CESARE, C. M.; AVERBECK, C. E. **Capítulo VI – da Avaliação de Imóveis.** In: CUNHA, E. M. P.. ERBA, D. A. Manual de Apoio – CTM: Diretrizes para a criação, instituição e atualização do cadastro territorial multifinalitário nos municípios brasileiros / Organizadores: Eglaisa Micheline Pontes Cunha e Diego Alfonso Erba – Brasília: Ministério das Cidades. 2010.

DUBIN, R. **Robustness of spatial autocorrelation specifications: some Monte Carlo evidence.** Journal of Regional Science. Vol. 43. 2. pp. 221-248. 2003.

EMILIANO, P. C.; VEIGA, E. P.; VIVANCO, M. J. F.; MENEZES, F. S. **Crítérios de informação de Akaike versus Bayesiano: Análise comparativa.** UNICAMP. SP. 2009. Disponível em <http://www.ime.unicamp.br/sinape/sites/default/files/Paulo%20C3%A9sar%20Emiliano.pdf>> Acesso em 02/2016.

ENGLAND, R. W. **Reconsideración del avalúo preferencial del suelo rural.** Lincoln

Institute of Land Policy. Working Paper. Cambridge. 2012.

ESRI. **ArcGIS Desktop: Release 10.2.1**. Redlands, CA: Environmental Systems Research Institute. 2013.

FERREIRA, J. F. **Proposta de tratamento da variável localização em modelos inferenciais de avaliação imobiliária para municípios médios**. 2007. 157 f. Dissertação (Mestrado). Centro de Ciências Exatas e de Tecnologia. Universidade Federal de São Carlos. São Carlos. SP. 2007.

FORTI, G. O. F. **Os municípios da região Norte Fluminense no período de 1990 a 2010: uma análise socioeconômica comparativa**. XIII Simpósio Nacional de Geografia Urbana - SIMPURB 2013. UERJ. RJ.

GENÚ, A. M. **Geoestatística multivariada**. Escola Superior De Agricultura “Luiz De Queiroz”. Universidade De São Paulo. SP. 2004. Disponível em <<http://www.lce.esalq.usp.br/tadeu/aline.pdf>>. Acesso em 12/2014.

GOMES, M. A. S. **A valorização fundiária nas cidades do petróleo: Campos dos Goytacazes e Macaé/RJ**. XVII Encontro Nacional de Geógrafos - XVII ENG. Belo Horizonte. MG. 2012.

GONZÁLEZ, M. A. S. **Aplicação de técnicas de descobrimento de conhecimento em base de dados e de inteligência artificial em avaliações de imóveis**. 2002. 300 f. Tese (Doutorado). Programa de Pós-Graduação em Engenharia Civil. Universidade Federal do Rio Grande do Sul. Porto Alegre. RS. 2002.

GONZÁLEZ, M. A. S.; SOIBELMAN, Lucio; FORMOSO, C. T. **A new approach to spatial analysis in CAMA**. 9th European Real Estate Society Conference (ERES 2002). Glasgow. Scotland. 4-7 June 2002.

GRIPP JR, J.; MARQUES, E. T.; GONÇALVES, R. P.; ANDRADE, R. J. O. **Avaliação de imóveis rurais**. In: COBRAC 2006. Florianópolis. SC. 2006.

GUIMARÃES, E. C. **Geoestatística básica e aplicada**. Uberlândia. MG. 2004.

HORNBURG, R. A. **Avaliação em massa com uso combinado da regressão espacial e da Geoestatística: estudo de caso: Navegantes-SC**. 2009. 79 f. Dissertação (Mestrado). Centro Tecnológico. Universidade Federal de Santa Catarina. Florianópolis. SC. 2009.

HORNBURG, R. A; HOCHHEIM, N. **Elaboração de Planta de Valores Genéricos com uso combinado de regressão espacial e Krigagem**. XV COBREAP - Congresso Brasileiro de Engenharia de Avaliações e Perícias. SP. 2009.

HERRERA, H.; ARGERICH, A. I.; GALERA, E. **La equidad en el impuesto inmobiliario**. Revista de Ciencia y Técnica. Vol. VII. N° 10. Ano 7. 2001

IAAO (INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS). **Standards on ratio studies**. Missouri: IAAO. 2013a.

IAAO (INTERNATIONAL ASSOCIATION OF ASSESSING OFFICERS). **Standard on mass appraisal of real property**. Missouri: IAAO. 2013b.

INCRA. Instituto de Colonização e Reforma Agrária. **Estatísticas cadastrais**. DF. 2012. Disponível em: <<http://www.incra.gov.br/index.php/estrutura-fundiaria/regularizacao-fundiaria/estatisticas-cadastrais/file/1250-relacao-total-de-imoveis-rurais-brasil-abril-2012>>. Acesso em: 11/2012.

INCRA. Instituto de Colonização e Reforma Agrária. **Relatório referente à planilha de preços referenciais da Região Norte Fluminense**. RJ. 2013.

IPEA - Instituto de Pesquisa Econômica Aplicada. **Grandes investimentos e a reestruturação do espaço regional no Norte Fluminense**. Anais do I Circuito de Debates Acadêmicos. Code 2011. Brasília. DF.

KAUKO, T.; D'AMATO, M. **Mass appraisal methods: an international perspective for property valuers**. Wiley-Blackwell; 1 edition. Hardcover. 2008.

KOTTEK, M.; GRIESER, J; BECK, C. RUDOLF, B; RUBEL, F. **World map of the Köppen-Geiger climate classification updated**. Meteorol. Z.. 15. 259-263. 2006. Disponível em http://koeppen-geiger.vu-wien.ac.at/pdf/Paper_2006.pdf. Acesso em 03/2015.

KRIGE, D. G. **A statistical approach to some basic mine evaluation problems on the Witwatersrand**. Johannesburg Chemistry Metallurgy Mining Society South African. 52 (6): 119-139. 1951. Disponível em <http://www.saimm.co.za/Conferences/DanieKrige/DGK01b.pdf>>. Acesso em 12/2014.

KREMPI, A. P. **Explorando recursos de estatística espacial para análise da acessibilidade na cidade de Bauru**. 2004. 94 f. Dissertação (Mestrado). Escola de Engenharia de São Carlos. Universidade de São Paulo. São Carlos. SP. 2004.

LAI, P. **Analysis of the mass appraisal model using Artificial Neural Network in Kaohsiung City**. 23rd Pan Pacific Congresso f Appraisers. Valuers and Counserlors. San Francisco. CA. September. 2006.

_____. **Analysis of the mass appraisal model using Artificial Neural Network in Kaohsiung City**. Journal of Modern Accounting and Auditing. October 2011. Vol. 7. N. 10. 1080-1089. New York. 2011.

LANDIM, P. M. B. **Sobre Geoestatística e mapas**. Terrae Didática. Disponível em <http://www.ige.unicamp.br/terraedidatica/>. UNESP. 2(1):19-33. 2006. Acesso em 02/2016.

LEPSCH, I. F.; BELLINAZZI JR., R.; BERTOLINI, D.; ESPÍNDOLA, C. R. **Manual para levantamento utilitário do meio físico e classificação de terras no Sistema de Capacidade de Uso. 4ª aproximação**. Campinas. SP: SOCIEDADE BRASILEIRA DE CIENCIA DO SOLO. 1991. 175p.

LIMA, M. R. C. **Avaliação de propriedades rurais: manual básico: a engenharia de avaliações rurais aplicadas às fazendas**. 3ª edição. São Paulo. Leud. 2011.

LOURIVAL, R.; CALEMAN, S. M. D. Q.; VILLAR, G. I. M.; RIBEIRO, A. R.; ELKIN, C. **Getting fourteen for the price of one! Understanding the factors that influence land value and how they affect biodiversity conservation in central Brazil**. Ecological Economics. v. 67. n. 1. p. 20-31. 2008.

MACEDO, F. C. **Mapeamento do preço da terra no Estado de Goiás com o uso de Geoestatística**. 2013. 129 p. Tese (Doutorado). Instituto de Estudos Socio-Ambientais. Universidade Federal de Goiás. Goiânia. 2013.

MACHADO, E. F. **Elaboração de Planta de Valores Genéricos para zona rural com uso de Superfície de Tendência**. 2006. 105 f. Dissertação (Mestrado). Centro Tecnológico). Universidade Federal de Santa Catarina. Florianópolis. SC. 2006.

MACHADO, H. B. **Curso de Direito Tributário**, 24ª Edição, Editora Malheiros, São Paulo,

2004.

MARQUES, A. P. S.; JUNIOR, J. M.; AMORIM, A. FLORES, E. F. **Aplicação do interpolador Krigagem ordinária para a elaboração de Planta de Valores Genéricos.** Revista Brasileira de Cartografia. v. 2. n. 64. p.175-186. abr. 2012.

MATHERON, G. **The theory of regionalized variables and its applications.** Paris. Les Cahiers du Centre de Morphologie Mathématique de Fontainebleu. 1971. 211p. Disponível em <http://cg.ensmp.fr/bibliotheque/public/MATHERON_Ouvrage_00167.pdf> Acesso em 12/2014.

McCLUSKEY, W. J.; McCORD, M.; DAVIS P.T.; HARAN, M.; McILHATTON D. **Prediction accuracy in mass appraisal: a comparison of modern approaches.** Journal of Property Research, 30:4, 239-265, 2013.

McCLUSKEY, W. J.; DEDDIS, W. G.; MANNIS, A.; McBURNEY, D.; BORST, R. A. **Interactive application of computer assisted mass appraisal and Geographic Information Systems.** Journal of Property Valuation and Investment. v.15. n.5. p.448-465. 1997.

McCLUSKEY, W. J.; ANAND, S. **The application of intelligent hybrid techniques for the mass appraisal of residential properties.** Journal of Property Investment and Finance. v.17. n.3. p.218-239. 1999.

MEDRI, W. **Análise exploratória de dados.** Curso de Especialização “Lato Sensu” em Estatística. Londrina. PR. 2011.

MENEZES, P. C. P.; STRAUCH, J. C. M.; DAMÁZIO, J. M.; AJARA, C. **Utilização de técnicas de regressão linear espacial no estudo sobre a relação entre renda e distância da fonte de energia elétrica no estado do Amazonas. Brasil.** VI Seminário Latino-Americano de Geografia Física. Universidade de Coimbra. 2010. Disponível em <<http://www.uc.pt/fluc/cegot/VISLAGF/actas/tema2/paulo>>. Acesso em 02/2016.

MINISTÉRIO DO MEIO AMBIENTE. <http://www.mma.gov.br/informma/item/9319-mapa-recebe-imagens-de-satélite>. Acessado em 07/2013.

MOORE, J.W.; MYERS, J. **Using geographic-attribute weighted regression for CAMA Modeling.** Journal of Property Tax Assessment & Administration. Vol. 7, nº 3. 2010.

NAKAYA, T.; CHARLTON, M.; LEWIS, P.; BRUNSDON, C.; YAO, J.; FOTHERINGHAM, S. **GWR4 - User manual Windows application for Geographically Weighted Regression modelling.** Department of Geography, Ritsumeikan University. Quioto, Japão.

NAKAMURA, J. **Em torno do petróleo.** Construção e Mercado. PINI. 2012. Disponível em <<http://construcaomercado.pini.com.br/negocios-incorporacao-construcao/127/artigo282610-1.aspx>>. Acesso em 02/215.

NETER, J.; WASSERMANN, W.; KUTNER, M. H. **Applied linear statistical models. Regression. analysis of variance. and experimental designs.** 3ed. Burr Ridge (IL): Richard D. Irwin. 1990.

PASSOS, D. C. P. **Impactos produtivos do projeto: Complexo Logístico-Industrial do Porto do Açú (RJ).** 2013. 80 f. Dissertação (Mestrado). Centro de Ciência e Tecnologia. Universidade Estadual do Norte Fluminense Darcy Ribeiro. Campos dos Goytacazes. RJ.

2013.

PCI. **Geomatica Software Solutions. reference manuals. version 9.1.** PCI Geomatics Inc. Canada. 2003.

PIQUET, R. P. S.(org). **Petróleo, royalties e região.** Rio de Janeiro: Garamond, 2003.

PROJIR. **Projeto de Irrigação e Drenagem da Cana-de-Açúcar na Região Norte Fluminense.** Instituto do Açúcar e do Alcool. Norte Fluminense. RJ. 1982

RAMANATHAN, R. **Introductory econometrics - With applications.** 4ed. Fort Worth (USA): Dryden. 1998.

RESENER, M. C.; HOCHHEIM, N. **Avaliação de imóveis rurais por inferência estatística e análise de superfície de tendência.** XIII COBREAP. 2006. Fortaleza. CE.

RIBEIRO, R. V. **Desafios ao desenvolvimento regional do Norte Fluminense.** Dissertação (Mestrado em Políticas Públicas, Estratégias e Desenvolvimento), Universidade Federal do Rio de Janeiro. Rio de Janeiro. RJ. 2010.

RODRIGUES, L. A.; PALMEIRA, D.T.; LEMOS, L. M. **Impactos sócio territoriais da implantação do Porto do Açú.** XVI Encontro Nacional dos Geógrafos. ENG 2010. Porto Alegre. RS. 2010.

SANTOS, H. G. **Efeitos espaciais em mercados de terras rurais: modelagem, validação e avaliação de desempenho.** 2014. 182 f. Dissertação (Mestrado). Universidade Federal de Pernambuco. Programa de Pós-graduação em Ciências Geodésicas e Tecnologias da Geoinformação. Recife. 2014.

SANTOS, H. G.; JACOMINE, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A.; CUNHA, T. J. F.; OLIVEIRA, J. B. **Sistema Brasileiro de Classificação de Solos.** 3ª. ed. Brasília, DF: Embrapa. 2013.

SANTOS, B. S.; SANTOS, R. L.; SANTO, S. M. **Análise espacial aplicada à expansão de condomínios fechados na cidade de Feira de Santana (BA).** IV Simpósio Brasileiro de Ciências Geodésicas e Tecnologias da Geoinformação. Recife. PE. 2012.

SILVA, A. H. C. V.; NUNES, F. G. **Contribuição do sistema de informação e técnicas geopaciais aplicadas ao estudo Geotectônico – Sismicidade em Goiás.** Atas das I Jornadas Lusófanos de Ciências e Tecnologias de Informação Geográfica. Sessão 2. Artigo 6. Universidade de Coimbra. Portugal. 2015.

SILVA, D.; BARRETO, P. **O potencial do Imposto Territorial Rural contra o desmatamento especulativo na Amazônia.** Belém: Imazon, 2014. 48 p.

SILVA, J. A. M.; NALI, L. R.; MAROTE, F. M. M. **Modelagem de dados por regressão linear múltipla para avaliação de imóveis rurais do submédio São Francisco.** XV COBREAP. São Paulo. SP. 2009.

SOUZA, P. M.; PONCIANO, N. J.; MATA, H. T. C.; BRITO, M. N.; GOLINSKI, J. **Padrão de desenvolvimento tecnológico dos municípios das Regiões Norte e Noroeste do Rio de Janeiro.** Rev. Econ. Sociol. Rural [online]. 2009, vol.47, n.4.

SPRING. **Integrating remote sensing and GIS by object-oriented data modelling.** Camara G., Souza R.C.M., Freitas, U.M., Garrido, J. Computers & Graphics, 20: (3) 395-403, May-Jun 1996.

TRIVELLONI, C. A. P. **Método para determinação do valor da localização com uso de técnicas inferenciais e Geoestatísticas na avaliação em massa de imóveis.** 2005. 172 f. Tese (Doutorado). Centro Tecnológico. Universidade Federal de Santa Catarina. Florianópolis. SC. 2005.

TYSZLER, M. **Econometria Espacial: discutindo medidas para a matriz de ponderação espacial.** 2006. 155 f. Dissertação (Mestrado). Escola de Administração de Empresas de São Paulo. Fundação Getúlio Vargas. São Paulo. SP. 2006.

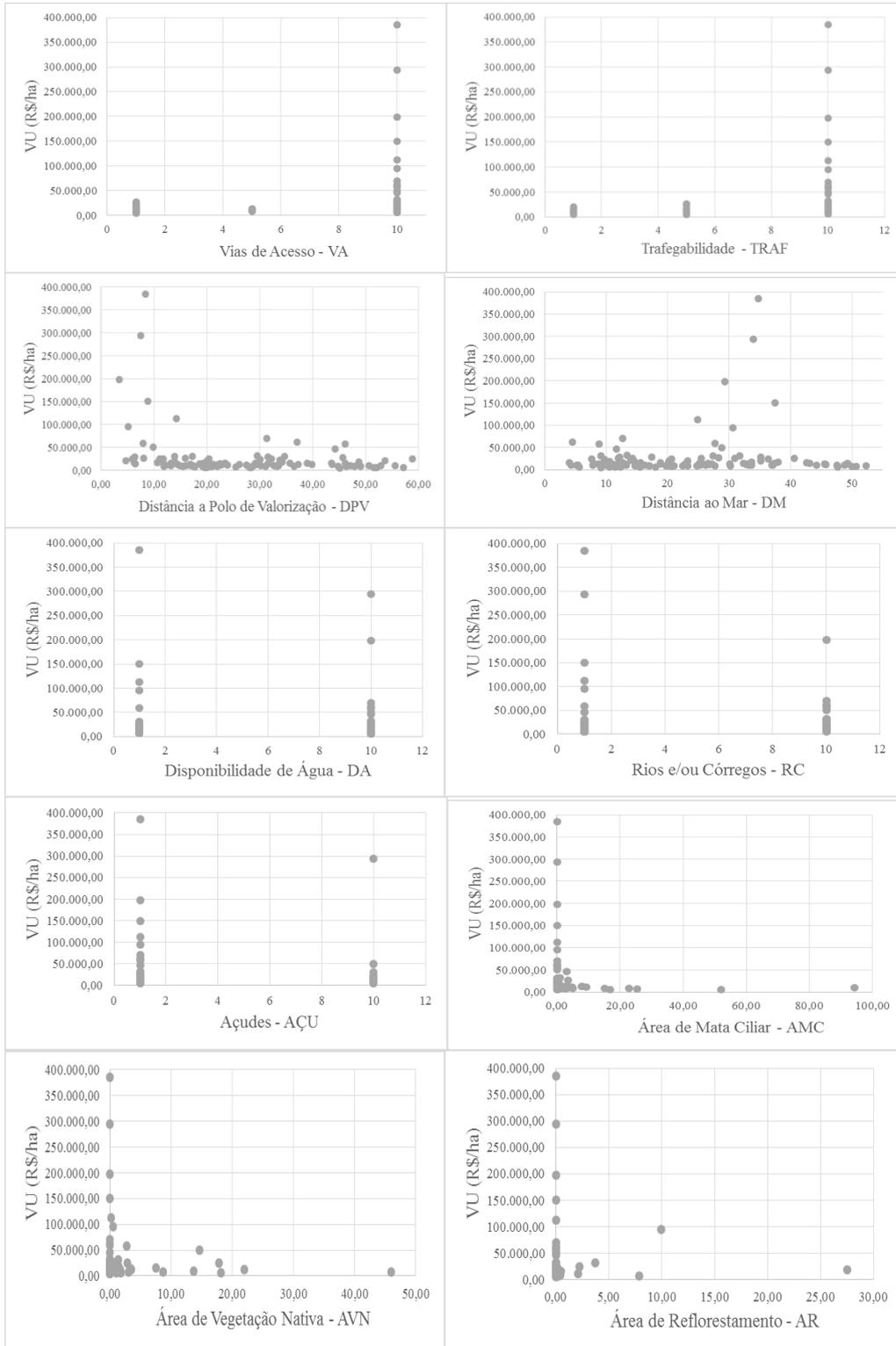
VIEIRA, R. S. **Crescimento econômico no estado de São Paulo: uma análise espacial [online].** São Paulo. Cultura Acadêmica, 2009.

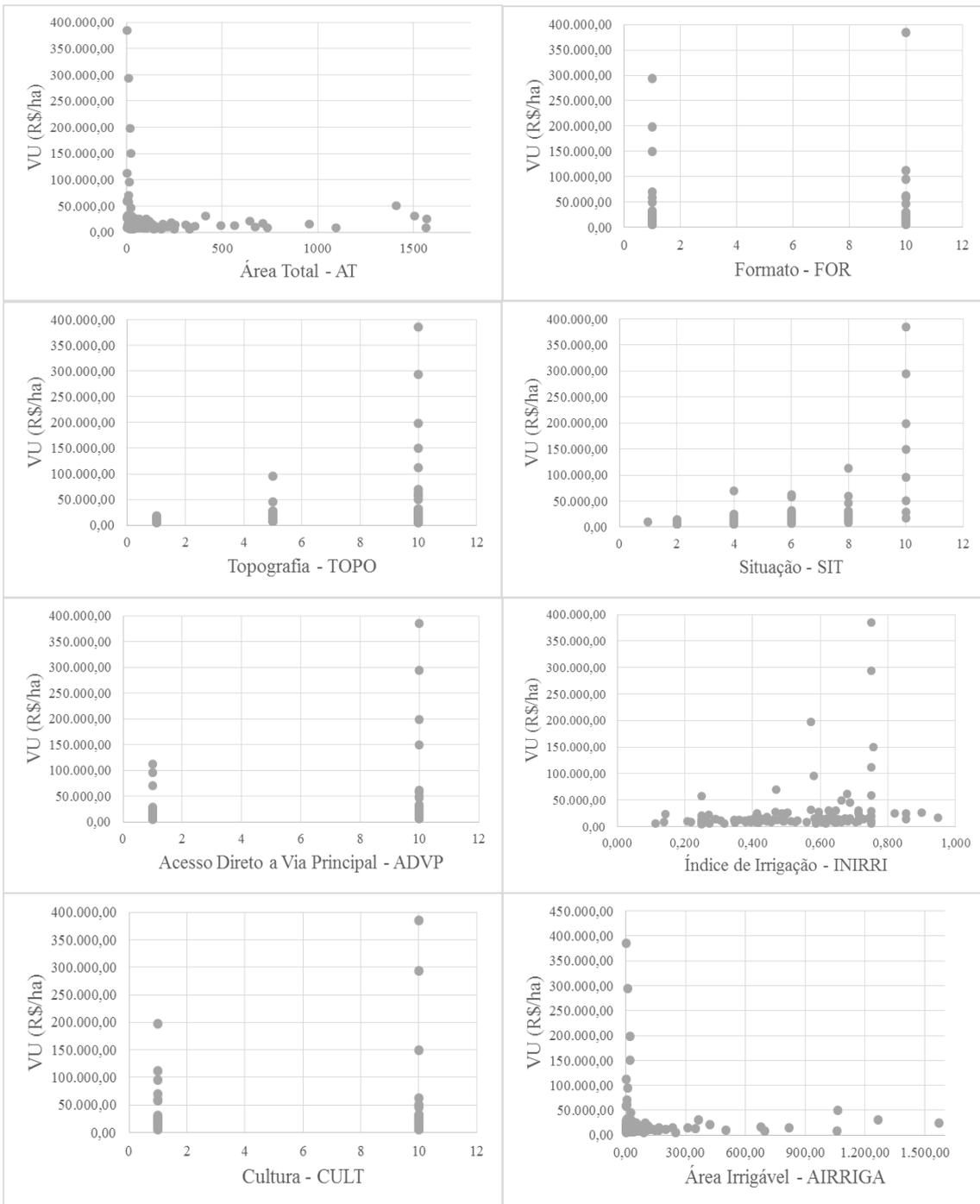
YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações.** 1. ed. São Paulo: Oficina de Textos. 2013. 215 p.

WYATT, P. **The development of a property information system for valuation using a Geographical Information System (GIS).** Journal of Property Research. v.13. p.317-336. 1996.

8 ANEXOS

ANEXO 1 – GRÁFICOS DE DISPERSÃO (Valor Unitário x Variáveis Independentes)





ANEXO 2 – AMOSTRA DE DADOS UTILIZADOS

Nº	VU (R\$)	X	Y	VA	DPV (km)	DM (km)	AÇU	AVN (ha)	AT (ha)	TOPO	SIT	AIRRIGA	CULT
1	61.950,00	281.377	7.622.297	10	37,10	4,46	1	0,0000	6,73	10	6	6,14	10
2	20.667,00	276.912	7.643.043	10	53,70	20,21	1	1,3463	116,37	5	4	56,91	1
3	95.000,00	260.119	7.598.370	10	5,16	30,6	1	0,5246	13,12	5	10	9,71	1
4	22.500,00	273.018	7.562.575	10	33,80	9,76	10	0,0000	47,62	10	4	20,26	1
5	10.743,00	258.066	7.571.104	1	22,04	26,17	10	0,0000	11,33	1	1	3,60	10
6	24.790,00	287.992	7.643.821	1	58,90	7,61	1	17,8027	102,84	5	4	95,38	1
7	58.000,00	240.770	7.550.038	10	46,20	8,86	1	2,7628	9,03	10	6	0,00	1
8	12.700,00	264.978	7.609.561	10	16,81	27,16	1	0,0000	29,50	5	4	21,84	1
9	12.400,00	273.928	7.614.092	1	26,20	12,9	10	1,7029	48,09	5	2	42,52	1
10	12.000,00	280.571	7.585.576	1	22,00	12,17	1	0,0000	17,11	10	6	0,00	10
11	13.400,00	277.609	7.615.082	1	29,10	8,7	1	0,0000	16,10	10	2	3,59	1
12	11.600,00	279.488	7.598.000	10	21,14	11,21	1	0,0000	48,03	10	6	48,03	10
13	25.000,00	265.741	7.602.093	10	11,12	25,46	1	0,0000	15,25	10	6	15,25	10
14	8.991,00	254.831	7.613.182	5	21,00	30,25	1	13,6577	79,33	1	6	0,00	10
15	31.000,00	279.989	7.566.409	10	34,69	9,13	1	0,0000	24,71	10	8	24,71	1
16	11.570,00	245.913	7.592.879	10	13,00	45,7	1	0,0000	120,12	10	6	120,12	10
17	8.500,00	276.123	7.585.631	1	19,20	17	1	0,0000	1,19	10	2	1,19	10
18	28.925,00	257.216	7.586.792	10	6,38	35,19	1	0,0000	3,04	10	10	3,04	10
19	11.680,00	253.023	7.614.360	1	22,61	32,96	10	0,0000	241,26	5	4	126,04	1
20	20.000,00	270.215	7.585.648	1	13,89	23,22	1	0,0000	18,44	10	6	18,44	10
21	7.770,00	240.454	7.594.930	5	17,72	50	1	1,7650	98,93	5	4	39,18	10
22	14.303,00	253.390	7.597.893	1	6,50	37,4	1	0,0000	77,84	10	8	65,92	10
23	11.652,00	252.822	7.616.464	1	23,85	32,87	1	0,4202	5,29	5	4	3,78	1
24	10.000,00	242.261	7.599.324	10	17,39	48,8	1	1,4397	99,41	1	6	0,00	10
25	9.006,00	228.057	7.559.619	1	44,84	23,25	1	0,0000	18,67	10	4	15,26	1
26	10.330,00	250.107	7.580.556	1	14,92	44,2	1	0,0000	11,02	10	4	11,02	10
27	150.000,00	255.524	7.584.999	10	8,84	37,48	1	0,0000	19,88	10	10	19,88	10
28	13.429,00	238.682	7.571.030	5	29,75	30,15	1	0,0000	46,46	5	4	46,46	1
29	12.300,00	229.895	7.560.300	1	43,70	22,87	1	21,9367	196,43	1	4	75,91	1
30	70.000,00	271.587	7.564.900	10	31,30	12,7	1	0,0000	7,88	10	4	4,93	1
31	385.000,00	258.255	7.584.961	10	8,39	34,77	1	0,0000	1,93	10	10	1,93	10
32	16.618,00	252.459	7.594.789	10	6,30	38	1	0,0000	106,72	10	10	106,72	10
33	9.933,00	252.344	7.605.044	5	13,17	37,05	1	0,0000	125,26	1	6	25,73	10
34	18.596,00	285.791	7.572.857	10	34,27	10,53	10	0,0000	234,72	10	4	111,79	10
35	10.528,00	244.190	7.585.989	10	15,98	47,6	1	0,0000	7,52	10	6	7,52	10
36	7.346,00	264.331	7.619.649	1	27,89	20,65	1	3,0141	39,29	1	6	34,23	10
37	28.209,00	278.149	7.634.007	10	45,70	12,12	1	0,0000	7,53	5	6	6,57	1
38	294.000,00	259.352	7.584.825	10	7,47	33,9	10	0,0000	8,27	10	10	8,27	10
39	32.314,00	272.873	7.567.076	10	29,51	13,38	1	0,0000	6,48	10	6	5,72	10
40	13.889,00	244.129	7.580.312	1	19,38	49,22	1	0,0000	16,46	1	2	2,67	10

Nº	VU (R\$)	X	Y	VA	DPV (km)	DM (km)	AÇU	AVN (ha)	AT (ha)	TOPO	SIT	AIRRIGA	CULT
41	26.206,00	251.135	7.590.081	10	8,00	40,56	1	0,0000	3,23	10	8	3,23	10
42	58.800,00	263.218	7.599.857	10	7,95	27,69	1	0,0000	0,94	10	8	0,94	1
43	198.120,00	261.898	7.593.556	10	3,45	29,28	1	0,0000	18,63	10	10	18,63	1
44	10.661,00	273.228	7.619.198	1	30,20	12,14	1	0,0000	20,44	5	4	12,56	10
45	112.557,00	268.614	7.583.401	10	14,28	24,86	1	0,1890	2,42	10	8	2,42	1
46	8.265,00	264.671	7.603.904	1	11,90	24,73	1	0,0000	62,45	10	6	62,45	10
47	15.322,00	273.815	7.636.139	1	46,20	18,81	10	0,0000	24,90	1	6	8,78	10
48	8.365,00	235.210	7.549.588	1	49,10	10,53	1	0,0000	69,72	1	4	40,23	10
49	5.365,00	232.291	7.542.098	1	57,20	5,62	10	0,0000	249,03	1	2	249,03	10
50	46.000,00	240.335	7.552.354	10	44,20	11,66	1	0,0000	22,42	5	8	22,42	10
51	8.500,00	231.870	7.555.000	1	46,40	17	1	8,6349	21,33	5	6	14,92	1
52	9.700,00	234.189	7.552.555	1	47,20	13,24	1	0,0000	70,39	1	6	30,94	10
53	15.800,00	247.389	7.585.758	1	13,27	42,59	1	1,0365	72,50	1	8	0,28	10
54	6.542,00	241.000	7.602.353	10	19,67	47,63	1	0,0000	42,98	10	6	39,15	10
55	14.462,00	259.650	7.615.489	1	22,41	26,62	10	0,0000	4,40	1	6	2,09	1
56	8.264,00	241.926	7.581.027	1	20,70	52,31	1	0,0000	11,64	1	2	7,37	1
57	18.598,00	278.826	7.636.796	10	48,70	14,6	10	0,0000	28,65	1	6	28,65	1
58	50.019,00	265.305	7.586.181	10	9,80	28,75	10	14,5862	1.410,02	10	10	1.061,29	10
59	31.000,00	257.543	7.577.397	10	13,88	31,74	1	0,0000	1.504,75	10	8	1.266,04	10
60	16.528,00	259.845	7.582.674	10	10,67	33,6	1	0,5549	711,00	10	8	678,53	10
61	15.000,00	261.242	7.579.616	10	13,61	33,68	1	0,0000	954,33	10	8	820,49	10
62	25.000,00	263.499	7.581.795	10	11,81	30,9	1	2,8079	1.572,51	10	8	1.572,51	10
63	30.990,00	262.129	7.576.476	10	17,28	27,41	10	1,3640	414,33	10	8	362,95	10
64	15.590,00	280.641	7.630.411	1	43,60	9,53	1	1,0128	47,03	5	6	35,13	1
65	10.595,00	288.719	7.633.583	1	50,70	4,27	1	0,0000	104,62	5	4	101,89	10
66	12.981,00	269.788	7.630.697	1	39,90	18,78	10	0,0000	78,28	1	4	32,83	1
67	20.991,00	279.116	7.587.384	10	20,10	13,93	1	0,2860	10,73	10	6	0,00	1
68	25.167,00	278.782	7.585.542	10	20,40	14,29	1	0,0000	16,71	10	6	7,92	10
69	26.859,00	259.191	7.609.617	1	15,88	28,3	1	0,0000	36,15	5	6	31,37	10
70	10.268,00	251.973	7.609.774	10	17,00	37	1	0,0000	153,85	1	6	43,15	10
71	14.330,00	247.397	7.596.389	10	11,84	43,12	1	0,0000	251,35	10	6	233,72	10
72	12.412,00	247.417	7.610.570	10	21,24	32,93	1	0,0000	191,45	5	6	103,37	10
73	10.100,00	236.767	7.573.721	1	28,80	33,6	1	0,0000	327,03	5	4	147,43	1
74	9.297,00	257.198	7.613.527	10	22,00	27,67	1	0,0000	673,95	1	8	500,14	10
75	14.563,00	269.515	7.563.593	10	31,50	12,85	1	3,3364	309,09	10	4	309,09	10
76	15.673,00	269.261	7.619.932	10	29,20	15,6	1	7,4956	131,50	1	6	54,77	10
77	10.456,00	258.349	7.604.540	1	12,00	33,2	10	0,0000	214,19	5	6	166,49	10
78	9.917,00	240.546	7.564.962	1	32,70	25,3	1	0,0000	181,51	5	4	17,25	1
79	10.970,00	237.329	7.566.384	10	33,40	26,8	1	0,0000	20,61	5	4	5,93	10
80	12.396,00	270.719	7.627.306	1	37,30	15,6	10	3,3400	566,37	1	4	198,47	10
81	11.363,00	247.592	7.568.925	1	27,50	26,1	10	0,0000	358,34	10	4	135,62	1
82	6.196,00	232.277	7.548.061	10	52,10	10,51	1	18,1520	182,05	1	4	4,40	10

Nº	VU (R\$)	X	Y	VA	DPV (km)	DM (km)	AÇU	AVN (ha)	AT (ha)	TOPO	SIT	AIRRIGA	CULT
83	5.966,00	227.862	7.551.631	1	51,60	15,5	1	0,0000	328,17	1	4	88,45	10
84	8.347,00	225.822	7.557.746	10	48,01	22,4	1	45,9787	1.566,02	10	6	1.059,73	10
85	7.222,00	236.511	7.605.875	10	25,50	50,7	1	0,0000	99,89	5	4	22,41	10
86	8.330,00	246.382	7.564.210	1	31,10	19,7	1	0,0000	170,39	10	4	6,51	1
87	5.036,00	242.643	7.550.326	1	45,20	9,1	1	0,0000	32,16	10	4	0,00	1
88	8.190,00	242.875	7.563.474	1	33,20	20,3	1	0,0000	734,83	5	4	26,80	1
89	8.897,00	243.467	7.559.375	1	36,50	14,7	1	0,0000	1.096,17	1	4	694,84	10
90	10.661,00	280.621	7.618.299	1	33,60	5,2	1	0,0000	148,38	10	4	1,78	1
91	15.650,00	281.214	7.620.723	1	35,70	4	1	0,0000	31,90	5	4	24,23	10
92	14.462,00	280.900	7.585.014	1	22,50	12,09	1	0,0000	69,68	10	6	0,00	10
93	9.420,00	280.120	7.586.802	1	21,30	12,52	1	0,0000	13,94	10	4	0,00	1
94	6.194,00	276.118	7.582.182	1	20,00	18	1	1,0749	140,85	10	2	21,58	10
95	6.194,00	282.463	7.575.600	1	28,40	11,5	1	1,7767	16,07	10	2	11,78	10
96	14.828,00	278.668	7.626.314	10	39,00	7,9	10	0,3114	46,94	1	8	43,93	10
97	12.560,00	244.543	7.596.527	10	14,50	45,5	1	3,4920	491,20	1	8	350,62	10
98	20.661,00	256.692	7.598.171	10	4,70	35,11	10	0,0000	645,31	5	8	420,14	10
99	15.661,00	278.157	7.597.918	10	19,38	12,69	1	0,0000	190,76	10	8	165,48	10
100	24.793,00	255.561	7.587.947	1	5,94	36,33	1	0,7342	47,80	10	8	47,80	1
101	12.396,00	241.632	7.593.356	10	16,90	49,3	1	0,0000	95,63	5	6	43,85	1
102	6.420,00	283.179	7.576.967	1	28,20	12,3	1	0,0000	39,67	10	2	0,00	10
103	10.661,00	275.975	7.564.147	1	33,51	9,6	10	0,0000	14,44	10	2	8,74	10
104	14.462,00	254.132	7.615.379	1	23,20	32,32	1	0,0000	22,48	5	4	16,56	10
105	28.332,00	279.053	7.569.091	10	31,53	17,36	1	0,0000	16,61	10	6	12,58	1
106	25.123,00	270.887	7.563.081	1	32,19	11,91	1	0,3465	64,35	10	4	50,32	1
107	14.462,00	264.416	7.575.143	10	18,60	25,2	1	0,6004	19,25	10	4	19,25	10
108	10.360,00	288.220	7.639.972	1	55,60	7,85	1	0,0000	29,40	5	4	27,66	10
109	9.890,00	289.694	7.635.696	1	52,90	5,4	1	0,0000	50,52	5	4	29,09	1
110	10.780,00	271.012	7.610.651	1	22,50	16,3	1	0,0000	227,34	5	4	103,84	10
111	8.960,00	267.592	7.605.304	1	15,50	21	1	0,0000	13,36	5	4	12,08	10
112	15.940,00	273.135	7.574.381	1	23,50	20,1	1	0,4201	49,37	10	4	49,37	10
113	23.560,00	275.146	7.568.159	1	30,13	20,6	1	0,0000	43,989	10	4	0,00	1

ANEXO 3 – AMOSTRA DE VALIDAÇÃO

Nº	VU (R\$)	X	Y	VA	DPV (km)	DM (km)	AÇU	AVN (ha)	AT (ha)	TOPO	SIT	AIRRIGA	CULT
1	12.400,00	238.021	7.558.415	1	40,1	18,44	1	0,0000	403,211	1	4	112,25	10
2	28.500,00	246.733	7.583.102	10	15,37	46,81	1	1,4685	118,112	1	8	81,17	1
3	43.390,00	261.447	7.608.099	5	15,97	31,59	1	0,3143	14,631	5	8	14,63	1
4	9.586,00	279.741	7.596.052	10	20,89	11	1	0,0000	21,128	10	6	21,13	10
5	32.258,00	276.643	7.574.723	10	26	18,9	1	0,0000	3,998	10	6	4,00	10
6	26.900,00	283.803	7.567.256	10	35,44	12,17	1	0,0000	34,39	10	6	0,00	1
7	22.727,00	248.810	7.591.499	10	9,8	43	1	0,0000	95,250	10	8	95,25	10
8	11.076,00	244.812	7.577.118	10	20,73	33,5	1	0,0000	64,417	5	4	26,83	10
9	16.528,00	278.231	7.562.852	1	35,28	7,65	1	0,0000	30,721	10	4	18,80	10
10	22.300,00	246.658	7.554.634	1	39,5	11,4	1	0,0000	7,245	10	6	0,00	1
11	16.528,00	263.744	7.614.664	10	22,36	22,6	10	0,0000	52,314	1	6	41,77	10
12	22.727,00	277.252	7.630.948	1	42,8	13,75	10	0,0000	33,097	5	4	22,44	10
13	77.551,00	237.750	7.548.329	10	48,8	8,55	1	0,7067	2,020	10	6	2,02	10
14	11.177,00	254.507	7.603.691	10	10,86	34,13	10	0,0000	258,655	1	6	131,88	1
15	12.396,00	246.216	7.590.611	10	11,8	44,6	10	0,0000	129,073	1	6	42,17	10
16	12.396,00	255.640	7.581.776	10	11,5	34,8	1	0,0000	255,467	10	6	204,61	10
17	10.429,00	269.369	7.590.197	10	11,34	22,8	1	8,0866	1903,113	10	8	1.860,64	10
18	13.429,00	248.364	7.612.551	1	21,79	39,83	1	2,2198	231,427	5	6	83,71	1
19	11.363,00	283.352	7.638.715	1	52,7	12,9	10	0,0000	181,079	1	6	174,46	10
20	10.330,00	275.844	7.629.954	1	41,5	11,64	10	0,0000	362,163	1	6	240,57	10
21	9.297,00	276.063	7.618.978	1	31,6	9,9	1	1,4925	275,614	5	4	234,20	10
22	10.330,00	280.902	7.583.489	1	23,2	18,1	1	0,0000	63,472	10	4	0,00	1
23	16.528,00	247.894	7.596.187	10	11,28	42,65	1	0,0000	233,707	10	6	209,91	10
24	10.330,00	241.395	7.604.064	10	20,5	47,5	1	28,9442	271,662	5	4	55,73	10
25	8.396,00	237.998	7.568.447	1	32	27,7	1	0,0000	95,311	5	4	26,12	1