
**UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS GEODÉSICAS E
TECNOLOGIAS DA GEOINFORMAÇÃO**



**EFEITOS ESPACIAIS EM MERCADOS DE TERRAS
RURAS: MODELAGEM, VALIDAÇÃO E AVALIAÇÃO
DE DESEMPENHO.**

HÉLDER GRAMACHO DOS SANTOS

Dissertação de Mestrado

VIRTUS IMPAVIDA

Recife, 2014

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS GEODÉSICAS E
TECNOLOGIAS DA GEOINFORMAÇÃO

Hélder Gramacho dos Santos

EFEITOS ESPACIAIS EM MERCADOS DE TERRAS
RURAIS: MODELAGEM, VALIDAÇÃO E AVALIAÇÃO
DE DESEMPENHO.

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Geodésicas e Tecnologias da Geoinformação, do Centro de Tecnologia e Geociências da Universidade Federal de Pernambuco, como parte dos requisitos para obtenção do grau de Mestre em Ciências Geodésicas e Tecnologias da Geoinformação, área de concentração Cartografia e Sistemas de Geoinformação defendida e aprovada no dia 09/12/2014.

Orientador: Prof. Dr. José Luiz Portugal

Recife, 2014

Catálogo na fonte
Bibliotecário Vimário Carvalho da Silva, CRB-4 / 1204

A635e Santos, Hélder Gramacho dos.
Efeitos espaciais em mercados de terras rurais: modelagem, validação e avaliação de desempenho. / Hélder Gramacho dos Santos. - Recife: O Autor, 2014.
182 folhas, il., color., planta., gráfs. e tabs.

Orientador: Prof^o. Dr. José Luiz Portugal.

Dissertação (Mestrado) – Universidade Federal de Pernambuco. CTG. Programa de Pós-graduação em Ciências Geodésicas e Tecnologias da Geoinformação. 2014.
Inclui Referências bibliográficas, apêndices e anexos.

1. Ciências Geodésicas. 2. Geoestatística. 3. Econometria. 4. Imóveis rurais. 5. Plantas. I. Portugal, José Luiz (orientador). II. Título.

526.1 CDD (22. ed.)

UFPE
BCTG/2015-04

**“EFEITOS ESPACIAIS EM MERCADOS DE TERRAS RURAIS:
MODELAGEM, VALIDAÇÃO E AVALIAÇÃO DE DESEMPENHO”**

POR

HÉLDER GRAMACHO DOS SANTOS

Dissertação defendida e aprovada em 09/12/2014.

Banca Examinadora:

Prof. Dr. JOSÉ LUIZ PORTUGAL

Departamento de Engenharia Cartográfica - Universidade Federal de Pernambuco.

Prof. Dra. ANDREA FLÁVIA TENÓRIO CARNEIRO

Departamento de Engenharia Cartográfica - Universidade Federal de Pernambuco.

Prof. Dr. WAYNER VIEIRA DE SOUZA

Fundação Oswaldo Cruz – Centro de Pesquisas Aggeu Magalhães.

Aos meus pais Adilton, Aliciene e avós por todo amor e dedicação na construção da nossa família.

À minha esposa Giovanna, pelo companheirismo, amor, por sempre acreditar nos meus sonhos e mais que isso, se empenhar para que eles se tornem realidade.

Aos meus queridos irmãos Dante e Wellington pelo incentivo.

À minha sogra Teresa e a todos os seus familiares pela confiança.

Aos meus queridos sobrinhos Lorenzo e Hellen, Lavínia e Laís pelos momentos de alegria.

AGRADECIMENTOS

Agradeço:

Primeiramente a Deus por ter proporcionado a saúde, a resiliência e a determinação para superar todas as dificuldades.

Ao Instituto Nacional de Colonização e Reforma Agrária - INCRA pela concessão de licença que proporcionou as condições necessárias para a dedicação integral ao desenvolvimento deste trabalho.

Ao Programa de Pós-Graduação em Ciências Geodésicas e Tecnologias da Geoinformação por proporcionar as condições ao nosso desenvolvimento. A todos os seus professores pela dedicação na transmissão dos conhecimentos e a todos os seus funcionários pela atenção e dedicação.

Ao Prof. Dr. José Luiz Portugal pela orientação, pelas conversas nas quais conhecimentos fundamentais ao desenvolvimento do trabalho foram transmitidos e pela confiança no nosso trabalho.

Ao amigo Eng. Agr. José Antonio Moura e Silva e família pela ajuda fundamental durante a fase de coleta de dados, pelo incentivo e confiança.

Ao Eng. Agrônomo Jadson Fraga Araújo pela ajuda durante a fase de coleta de dados.

Aos Professores Dr. Rubens Dantas e Dr^a. Andrea F. Tenório Carneiro pelas importantes contribuições ao longo do desenvolvimento do trabalho.

Aos meus orientadores de iniciação científica Dr^a. Cecília Ritzinger e Dr. José Fernandes de Melo Filho pelos conhecimentos transmitidos anos atrás mas que ainda hoje se mostram valiosos.

À amiga Ana Itamara pelas conversas, pelo incentivo, e pelas alegrias compartilhadas.

Aos colegas de mestrado pelos momentos de estudo e de alegrias compartilhados ao longo dos últimos meses.

“Labor improbus omnia vincit”
Virgílio

*“A maior recompensa pelo nosso trabalho
não é o que nos pagam por ele, mas
aquilo em que ele nos transforma.”*
John Ruskin

RESUMO

A geoestatística e a econometria espacial são técnicas que tem sido utilizadas com resultados satisfatórios na modelagem de efeitos espaciais presentes nos mercados imobiliários. No Brasil, os estudos tem se concentrado na avaliação destes efeitos em mercados de imóveis urbanos. Apesar dos imóveis rurais também estarem submetidos a estes efeitos, nestes mercados predominam os modelos clássicos de regressão linear. Além disso, a valoração cadastral rural dos municípios Brasileiros encontra-se por demais atrasada quando comparada com os países da Europa e América do Sul. Neste contexto, o objetivo deste trabalho foi avaliar a combinação de técnicas da econometria espacial e da geoestatística na modelagem dos efeitos espaciais em mercados de terras rurais e geração de PVG, propondo uma metodologia que seja aplicável à realidade dos municípios brasileiros. A proposta metodológica consistiu em investigar o efeito causado pela autocorrelação espacial sobre os modelos clássicos de regressão linear (MCRL), modelar estes efeitos por meio da econometria espacial e da geoestatística, avaliar o desempenho dos MCRL comparando-os com os modelos espaciais e produzir a PVG por meio da krigagem. A área de estudos foi o município de Petrolina-PE, onde foram coletadas 104 amostras de mercado georreferenciadas. A amostra de trabalho consistiu de 84 observações. A amostra de verificação consistiu de 20 observações. Os resultados mostraram que a autocorrelação espacial pode ter seus efeitos controlados tanto pela econometria tradicional quanto pela econometria espacial. No primeiro caso, a inserção de variáveis de localização foi suficiente para evitar o surgimento de autocorrelação nos resíduos. No segundo caso, o modelo espacial autorregressivo (SAR) foi capaz de controlar a autocorrelação presente nos resíduos do MCRL. Os modelos tiveram a qualidade do ajuste avaliada e além disso, tiveram o seu desempenho avaliado por meio de pontos de verificação que não fizeram parte da modelagem. O modelo econométrico tradicional com variáveis relacionadas à localização apresentou melhor ajuste bem como foi melhor na avaliação de desempenho. A avaliação de desempenho mostrou que todos os modelos apresentaram um coeficiente de dispersão elevado, provavelmente devido ao fato dos modelos buscarem retratar a realidade de mercado tanto da região de sequeiro quanto da região irrigada. Foram geradas superfícies de preços a partir das 84 observações originais bem como a partir dos valores preditos pelos modelos. A técnica utilizada foi a krigagem ordinária. As superfícies de preços obtidas tiveram seu desempenho avaliado por meio dos pontos de verificação. Os resultados mostraram que a superfície gerada por meio dos dados originais foi a que obteve o melhor desempenho. Em função da escala da base cartográfica disponível para geração das plantas de valores estas são indicadas para informar as faixas de preços nas diferentes regiões do município. Para avaliações dos imóveis o indicado é utilizar o modelo. A combinação de metodologias se mostrou aplicável à realidade dos mercados de terras rurais e permitiu a obtenção de modelos representativos da realidade destes mercados bem como a elaboração da PVG.

Palavras-chave: Plantas de valores genéricos. Econometria espacial. Geoestatística.

ABSTRACT

Geostatistical and Spatial Econometrics techniques have been used with satisfactory results to modeling spatial effects present in real estate markets. In Brazil, studies have focused on the assessment of these effects in urban real estate markets. Despite rural properties also being subjected to spatial effects, in rural land markets classical linear regression models are predominant. In addition, the rural cadastral valuation of Brazilian municipalities is delayed when compared with countries in Europe and South America. In this context, the aim of this study was to evaluate combination of spatial econometrics and geostatistical techniques in modeling spatial effects in rural land markets and generation of PVG, proposing an appropriate methodology to reality of Brazilian municipalities. Methodology was to investigate consequences of spatial autocorrelation in classical linear regression models (MCRL), modeling these effects through spatial econometrics and geostatistics techniques, evaluate the performance of MCRL comparing them with spatial models and produce PVG by kriging. Study area was Petrolina municipality, were collected 104 georeferenced data market. Work sample consisted of 84 observations. Verification sample consisted of 20 observations. Results showed that spatial autocorrelation can have its effects controlled both by traditional econometrics as spatial econometrics. In first case, insertion of location variables was sufficient to prevent the emergence of autocorrelation. In second case, spatial autorregressive model (SAR) was able to control the autocorrelation in residuals MCRL. Models were evaluated on the quality of fit and in addition, performances were evaluated through checkpoints that were not part of modeling. Traditional econometric model with variables related to location showed a better adjustment and was better in performance assessment. Performance evaluation showed that all models had a high dispersion coefficient, probably because the models tried to represent the market reality, both semi-arid and irrigated area. Price surfaces were generated from 84 original data and by predicted values model's. Technique used was ordinary kriging. Price surfaces obtained had their performance evaluated through checkpoints. Results showed that surface created from the original data was that had the best performance. Base map in scale 1:250.000 available for creation of PVG causes these are indicated to inform price ranges in different regions of municipality. For appraisals of rural properties is indicated use model. Combination of methodologies proved applicable to rural land markets and allowed to obtain representative models and development of PVG.

Key words: Plant of values. Spatial econometrics. Geoestatistics.

LISTA DE FIGURAS

Figura 1 - Planta de Valores Genéricos dos imóveis rurais do município de Santa Rita do Pardo-MS.	34
Figura 2 - Processo a-espacial. Fonte: Almeida, (2012).....	60
Figura 3 - Processo de defasagem espacial. Fonte: Almeida (2012).....	61
Figura 4 - Processo de erro autorregressivo espacial. Fonte: Almeida, (2012).....	62
Figura 5 - Localização do município de Petrolina-PE no estado de Pernambuco.....	77
Figura 6 - Área de estudos, município de Petrolina e seus 4 distritos: Petrolina, Curral Queimado, Cristália e Rajada.....	78
Figura 7 - Fluxograma geral das atividades desenvolvidas.....	81
Figura 8 - Fluxograma de detalhamento das atividades de modelagem do mercado de terras mostrados na Figura 1.....	83
Figura 9 - Histograma de frequência absoluta e distribuição espacial dos dados classificados de acordo com o ano da sua ocorrência.....	91
Figura 10 - Distribuição espacial da amostra de trabalho e da amostra de verificação.	92
Figura 11 - Obtenção da distância euclidiana dos imóveis em relação aos polos de influência: a) Cidade de Petrolina; b) Rio São Francisco.....	96
Figura 12 - Fluxograma das etapas relacionadas à análise exploratória.....	97
Figura 13 - Gráfico do logaritmo da função verossimilhança estimada versus valores de λ ajustados para a variável dependente.....	98
Figura 14 - Histogramas da variável dependente (da esquerda para direita): na escala original; transformação logarítmica; transformação potência com λ obtido pelo método da máxima verossimilhança.....	99
Figura 15 - Gráficos Boxplots para a variável dependente (da esquerda para direita): na escala original; transformação logarítmica; transformação potência com λ obtido pelo método da máxima verossimilhança.....	100
Figura 16 - Gráficos Quantil-Quantil para a variável dependente (da esquerda para direita): na escala original; transformação logarítmica; transformação potência com λ obtido pelo método da máxima verossimilhança.....	100
Figura 17 - Distribuição espacial da variável dependente, valor do imóvel por hectare classificado em quartis.....	101
Figura 18 - Superfície de Tendência de 1º Grau ajustada à variável dependente... ..	102
Figura 19 - Histograma e Boxplot para os resíduos do modelo 1.....	112
Figura 20 - Gráfico de probabilidade normal dos resíduos e diagrama dos resíduos	

padronizados externamente versus os valores preditos.....	113
Figura 21 - Gráficos de pontos de influência dos resíduos para o Modelo 1.....	114
Figura 22 - Histograma e Boxplot para os resíduos do modelo 2.....	115
Figura 23 - Gráfico de probabilidade normal dos resíduos e diagrama dos resíduos padronizados externamente versus os valores preditos para o modelo 2.....	116
Figura 24 - Gráficos de pontos de influência dos resíduos para o Modelo 2.....	116
Figura 25 - Distribuição espacial dos resíduos do modelo 1.1 (esquerda) e 2.1 (direita), divididos em quartis.....	120
Figura 26 - Superfície de 1º Grau ajustada aos resíduos do modelo 2.1.....	122
Figura 27 - Semivariogramas experimentais dos resíduos do modelo 1.1 (esquerda) e 2.1 (direita), a distância é dada em quilômetros.....	123
Figura 28 - Semivariograma teórico, modelo esférico, ajustado aos resíduos do modelo 2.1, a distância é dada em quilômetros.....	125
Figura 29 - Exemplos de arranjos espaciais de matrizes de pesos baseadas nos 5 vizinhos mais próximos (esquerda) e na distância de 10 km (direita).....	127
Figura 30 - Arranjo espacial da matriz baseada em 2 vizinhos mais próximos e boxplot das distâncias entre os vizinhos desta matriz.....	129
Figura 31 - Gráfico de dispersão e mapa de clusters LISA para os valores do VTI/ha.	130
Figura 32 - Distribuição espacial das observações que compõem a amostra de verificação utilizada na avaliação de desempenho dos modelos.....	136
Figura 33 - Semivariograma teórico, modelo exponencial, ajustado para 84 observações dos dados originais, a distância é dada em quilômetros.....	140
Figura 34 - krigagem ordinária (esquerda) e desvio padrão (direita) para as 84 observações coletadas em campo.....	142
Figura 35 - Semivariograma teórico, modelo gaussiano, ajustado para 82 observações dos valores preditos pelo Modelo 1.1, a distância é dada em quilômetros.....	144
Figura 36 - krigagem ordinária (esquerda) e desvio padrão (direita) para as 82 observações preditas pelo Modelo 1.1.....	145
Figura 37 - Semivariograma teórico, modelo esférico, ajustado para 81 observações dos valores preditos pelo Modelo SAR, a distância é dada em quilômetros.....	147
Figura 38 - krigagem ordinária (esquerda) e desvio padrão (direita) para as 81 observações preditas pelo Modelo SAR.....	148
Figura 39 - Semivariogramas direcionais (0°, 45°, 90° e 135°) para os valores	

preditos pelo modelo 1.1 (esquerda) e SAR (direita).....	149
Figura 40 - Planta de Valores Genéricos do município de Petrolina-PE.....	152
Figuras 41 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.....	169
Figura 42 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.....	170
Figura 43 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.....	171
Figura 44 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.....	172

LISTA DE TABELAS E QUADRO

Tabela 1 - Trabalhos recentes envolvendo avaliação de imóveis rurais.....	75
Tabela 2 - Compilação das diferentes metodologias propostas para a elaboração de plantas de valores.....	76
Tabela 3 - Estabelecimentos e área total, por classes de atividade econômica no município de Petrolina-PE no ano de 2006.....	79
Tabela 4 - Condição do produtor em relação às terras segundo os estabelecimentos no município de Petrolina-PE no ano de 2006.....	80
Tabela 5 - Arrecadação anual do ITR pelo município de Petrolina-PE.....	80
Tabela 6 - Informações das principais culturas agrícolas do município de Petrolina em 2012.....	81
Tabela 7 - Estabelecimentos com área irrigada, por método utilizado, segundo os estabelecimentos e a área no município de Petrolina-PE em 2006.....	81
Tabela 8 - Efetivo dos três principais rebanhos do município de Petrolina em 2011.	82
Tabela 9 - Pontuação para fins de classificação das avaliações quanto ao grau de fundamentação.....	94
Tabela 10 - Coeficientes de assimetria e curtose para a variável dependente.....	100
Tabela 11 - Testes de normalidade para a variável dependente na escala original, transformação logarítmica e utilizando a transformação potência.....	102
Tabela 12 - Resultados gerais do ajuste de uma superfície de tendência de 1º grau à variável dependente.....	103
Tabela 13 - Coeficientes das variáveis independentes da superfície de tendência de 1º grau.....	103
Tabela 14 - Estatística descritiva para a variável dependente sem transformação, transformação logarítmica e utilizando a transformação potência.....	104
Tabela 15 - Quadro resumo das transformações das variáveis independentes.....	105
Tabela 16: Matriz de Correlações para as variáveis envolvidas na modelagem (continua).....	107
Tabela 17 - Resultados Gerais da Regressão do modelo 1.....	109
Tabela 18 - Estimativas dos parâmetros, erro-padrão, valor t e o valor p do modelo 1.	110
Tabela 19 - Resultados Gerais da Regressão do modelo 2.....	110
Tabela 20 - Estimativas dos parâmetros e erros-padrão, valor t e o valor p do modelo 2.....	110
Tabela 21 - Principais modelos construídos com respectivos testes estatísticos.....	111

Tabela 22 - Fatores de Inflação da Variância para os modelos 1 e 2.....	112
Tabela 23 - Variação percentual das estimativas dos coeficientes de regressão e valor do R ² ajustado de acordo com a retirada de observações atípicas.....	115
Tabela 24 - Variação percentual das estimativas dos coeficientes de regressão e valor do R ² ajustado de acordo com a retirada de observações atípicas.....	118
Tabela 25 - Resultados Gerais da Regressão do modelo 1.1.....	118
Tabela 26 - Estimativas dos parâmetros, erros-padrão, valor t e valor p do modelo 1.1.....	119
Tabela 27 - Resultados Gerais da Regressão do modelo 2.1.....	119
Tabela 28 - Estimativas dos parâmetros, erros-padrão, valor t e valor p do modelo 2.1.....	119
Tabela 29 - Principais modelos construídos com respectivos testes estatísticos....	120
Tabela 30 - Ajuste de uma superfície de tendência de 1º grau aos resíduos do modelo 1.1.....	122
Tabela 31 - Coeficientes da superfície de tendência de 1º grau aos resíduos do modelo 1.1.....	122
Tabela 32 - Ajuste de uma superfície de tendência de 1º grau aos resíduos do modelo 2.1.....	122
Tabela 33 - Coeficientes da superfície de tendência de 1º grau ajustada aos resíduos do modelo 2.1.....	122
Tabela 34 - Parâmetros do semivariograma experimental.....	124
Tabela 35 - Parâmetros dos semivariogramas teóricos ajustados pelo método dos mínimos quadrados aos resíduos do modelo 2.1.	125
Tabela 36 - Parâmetros para escolha dos modelos ajustados.....	126
Tabela 37 - Teste I de Moran e significância para as diversas matrizes de ponderação espacial testadas para o VTI/ha.....	129
Tabela 38: Teste I de Moran e significância para as matrizes de ponderação espacial testadas sobre os resíduos do modelo 2.1.....	131
Tabela 39 - Resultados do teste de especificação da regressão espacial utilizando o multiplicador de Lagrange.....	132
Tabela 40 - Variáveis independentes do modelo espacial autorregressivo SAR.	133
Tabela 41 - Modelo espacial autorregressivo.....	133
Tabela 42 - Comparativo entre os modelos por meio do Critério de Informação de Akaike e Schwartz.....	134
Tabela 43 - Valores propostos pela IAAO para avaliação de desempenho dos	

modelos.....	138
Tabela 44 - Parâmetros de avaliação de desempenho dos modelos.....	138
Tabela 45 - Parâmetros do semivariograma experimental para 84 observações dos dados de campo.....	140
Tabela 46 - Parâmetros dos semivariogramas teóricos ajustados pelo método dos mínimos quadrados para 84 observações dos dados originais.....	141
Tabela 47 - Parâmetros para escolha dos modelos ajustados.....	141
Tabela 48 - Parâmetros do semivariograma experimental para 82 observações dos valores preditos pelo Modelo 1.1.....	143
Tabela 49 - Parâmetros dos semivariogramas teóricos ajustados pelo método dos mínimos quadrados aos valores preditos pelo Modelo 1.1.	144
Tabela 50 - Parâmetros para escolha dos modelos ajustados.....	144
Tabela 51 - Parâmetros do semivariograma experimental para 81 observações dos valores preditos pelo Modelo SAR.....	146
Tabela 52 - Parâmetros dos semivariogramas teóricos ajustados para 81 observações dos valores preditos pelo Modelo SAR.....	147
Tabela 53 - Parâmetros para escolha dos modelos ajustados.....	147
Tabela 54 - Estimativa da Raiz do Erro Médio Quadrático no VTI/ha dos vinte pontos de verificação.....	151
Tabela 55 - Resultado do Teste de Tukey de comparação de múltiplas médias com 95% de confiança.....	151
Tabela 56 - Índice de valoração da produção vegetal em função do tempo de vida útil das culturas perenes comumente cultivadas na microrregião de Petrolina.....	181
Quadro 1 - Modelos de Valoração Cadastral Rural na Europa e América Latina.....	28

LISTA DE SIGLAS E ABREVIATURAS

ABNT	- Associação Brasileira de Normas Técnicas
AIC	- Critério de Informação de Akaike
CNIR	- Cadastro Nacional de Imóveis Rurais
CD	- Coeficiente de Dispersão
CTM	- Cadastro Territorial Multifinalitário
EPR	- Efeito Pepita Relativo
GAMLSS	- Generalized Additive Models for Location, Scale and Shape
IAAO	- International Association of Assessing Officers
INCRA	- Instituto Nacional de Colonização e Reforma Agrária
LISA	- Indicador Local de Autocorrelação Espacial
LIK	- Função de Verossimilhança
ITR	- Imposto Territorial Rural
ITBI	- Imposto sobre a Transmissão de Bens Imóveis
MCRL	- Modelo Clássico de Regressão Linear
ML	- Multiplicadores de Lagrange
MQO	- Mínimos Quadrados Ordinários
MMQ	- Método dos Mínimos Quadrados
MV	- Máxima Verossimilhança
NBR	- Norma Brasileira Regulamentadora
OLS	- Ordinary Least Squares
PVG	- Planta de Valores Genéricos
PRD	- Price Related Differential
QMRes	- Quadrado Médio de Resíduos
REMQ	- Raiz Quadrada do Erro Médio Quadrático
RESET	- Regression Specification Error Test
RFB	- Receita Federal do Brasil
SAR	- Spatial Autoregressive
SEM	- Spatial Error Model
SQE	- Soma de Quadrado dos Erros
SQres	- Soma de Quadrado dos Resíduos
SC	- Critério de Informação de Schwartz
TSA	- Análise de Superfície de Tendência
VTN	- Valor da Terra Nua
VTI	- Valor Total do Imóvel

SUMÁRIO

1	INTRODUÇÃO.....	19
1.1	OBJETIVOS.....	22
1.1.1	Objetivo Geral.....	22
1.1.2	Objetivos Específicos.....	23
1.2	ESTRUTURA DO TRABALHO.....	23
2	FUNDAMENTAÇÃO TEÓRICA.....	24
2.1	CONCEITOS GERAIS.....	24
2.2	MÉTODOS DE AVALIAÇÃO IMOBILIÁRIA.....	25
2.3	SITUAÇÃO ATUAL DA VALORAÇÃO CADASTRAL RURAL.....	27
2.4	VARIÁVEIS INFLUENCIANTES NA VALORAÇÃO CADASTRAL RURAL....	29
2.5	PLANTAS DE VALORES GENÉRICOS.....	33
2.6	CADASTRO TERRITORIAL MULTIFINALITÁRIO.....	36
2.7	IMPOSTOS INCIDENTES SOBRE OS IMÓVEIS RURAIS NO BRASIL.....	38
2.8	MODELO CLÁSSICO DE REGRESSÃO LINEAR (MCRL).....	40
2.9	ESTIMATIVA POR MÍNIMOS QUADRADOS.....	41
2.10	INTERVALO DE CONFIANÇA PARA PREDIÇÃO DE UMA NOVA OBSERVAÇÃO.....	44
2.11	PRESSUPOSTOS DO MCRL.....	45
2.12	DIAGNÓSTICOS NOS MCRL.....	46
2.12.1	Matriz Chapéu ou Hat Matrix.....	47
2.12.2	Análise dos Resíduos.....	48
2.12.3	Distância de Cook.....	50
2.13	CONSEQUÊNCIAS DOS EFEITOS ESPACIAIS SOBRE OS PRESSUPOSTOS DO MCRL.....	51
2.14	MATRIZES DE PONDERAÇÃO ESPACIAL.....	54
2.15	ESPECIFICAÇÃO DA DEPENDÊNCIA ESPACIAL.....	56
2.15.1	Teste ML (defasagem).....	57
2.15.2	Teste ML (erro).....	58
2.15.3	Especificação de Modelos Espaciais.....	58
2.16	MODELOS ESPACIAIS.....	59
2.16.1	Modelo Espacial Autorregressivo ou Spatial Autoregressive SAR.....	60
2.16.2	Modelo Espacial do Erro ou Spatial Error Model SEM.....	62
2.17	GEOESTATÍSTICA.....	63

2.18	SEMIVARIOGRAMA.....	64
2.18.1	Krigagem.....	66
2.19	AVALIAÇÃO DE DESEMPENHO.....	69
2.19.1	Medidas de Qualidade do Ajuste.....	69
2.19.2	Avaliação de Desempenho de Avaliações em Massa.....	69
2.20	TRABALHOS RECENTES SOBRE VALORAÇÃO CADASTRAL E PLANTAS DE VALORES GENÉRICOS.....	72
3	METODOLOGIA.....	76
3.1	CARACTERIZAÇÃO DA ÁREA DE ESTUDOS.....	76
3.2	PROCEDIMENTOS METODOLÓGICOS.....	81
3.2.1	Estudo de Mercado: Identificação das Variáveis Formadoras de Valor..	81
3.2.2	Coleta de Dados: Ofertas e Transações.....	82
3.2.3	Modelagem do Mercado de Terras.....	83
3.2.4	Análise e Validação dos Resultados.....	85
3.2.5	Avaliação da Qualidade dos Produtos Gerados.....	86
3.3	RECURSOS TECNOLÓGICOS.....	86
4	RESULTADOS E DISCUSSÃO.....	88
4.1	ESTUDO DE MERCADO: IDENTIFICAÇÃO DAS VARIÁVEIS FORMADORAS DE VALOR.....	88
4.2	COLETA DE DADOS: OFERTAS E NEGÓCIOS REALIZADOS.....	89
4.3	MODELAGEM DO MERCADO DE TERRAS.....	93
4.4	ANÁLISE EXPLORATÓRIA.....	96
4.4.1	Variável Dependente.....	97
4.4.2	Variáveis Independentes.....	103
4.5	MODELO CLÁSSICO DE REGRESSÃO LINEAR MÚLTIPLA.....	108
4.5.1	Análise da Multicolinearidade.....	111
4.5.2	Análise Gráfica dos Resíduos.....	112
4.5.3	Modelos Após a Retirada das Observações Atípicas.....	117
4.6	ANÁLISE GEOESTATÍSTICA DOS MODELOS.....	120
4.7	ANÁLISE ECONOMETRICA ESPACIAL.....	126
4.8	AVALIAÇÃO DE DESEMPENHO DOS MODELOS.....	133
4.8.1	Qualidade do Ajuste.....	133
4.8.2	Avaliação de Desempenho.....	135
4.9	ELABORAÇÃO DA PLANTA DE VALORES GENÉRICOS.....	138
4.9.1	Modelagem Geoestatística do Conjunto de Dados Campo.....	138

4.9.2	Modelagem Geoestatística dos Valores Preditos pelo Modelo 1.1.....	142
4.9.3	Modelagem Geoestatística dos Valores Preditos pelo Modelo SAR.....	145
4.9.4	Anisotropia.....	148
4.9.5	Avaliação do Desempenho das Krigagens Ordinárias.....	149
5	CONCLUSÕES E RECOMENDAÇÕES.....	153
5.1	CONCLUSÕES.....	153
5.2	RECOMENDAÇÕES PARA TRABALHOS FUTUROS.....	157
	REFERÊNCIAS.....	159
	APÊNDICE 01.....	166
	APÊNDICE 02.....	169
	APÊNDICE 03.....	173
	ANEXO 01.....	178

1 INTRODUÇÃO

Segundo estimativas da Organização das Nações Unidas (UNITED NATIONS, 2012), em 2050 o mundo terá entre 9 e 11 bilhões de pessoas, um cenário no qual a necessidade de produzir e distribuir alimentos torna-se uma questão estratégica.

Além da crise dos alimentos, a recessão econômica que provoca a busca por ativos seguros e as metas para maior utilização de biocombustíveis tem provocado uma “*Land rush*” termo que tem sido utilizado para se referir à aquisição em larga escala de terras, especialmente na África, Ásia e América Latina. Muitas vezes estas terras são adquiridas por meios ilegais sendo designado pelo termo “*Land grabbing*” (UNEP, 2014).

Esta corrida pode ser comprovada por meio do projeto Land Matrix (www.landmatrix.org) que é um Observatório de Valores Mundial e monitora aquisições de terra em larga escala. No período entre 2000 e 2010 foram identificados cerca de 203 milhões de hectares transacionados, admitindo-se que muitas das transações não são identificadas (ANSEEUW et al., 2012).

Barros (2012) analisando o Brasil em relação a importantes produtores mundiais de alimentos, lista como principais vantagens nacionais os fatos que diante da área disponível, não mais que 2% são ocupados com lavouras; o aumento da produção não está condicionada a um aumento do desmatamento; existe disponibilidade de água e outros insumos necessários à produção; existe disponibilidade de empreendedores e trabalhadores; e há constante oferta de inovações aos produtores, o que tem proporcionado aumento de produtividade.

A agropecuária situa o Brasil em posição de destaque no cenário mundial. No ano de 2013 o Produto Interno Bruto (PIB) do Agronegócio foi de R\$1.092.238.000.000,00 e representou cerca de 22,54% do PIB Brasileiro (CEPEA, 2014).

No ano de 2013, de acordo com o IBGE, (2013) que realizou a estimativa da safra agrícola nacional, o Brasil bateu novos recordes de produção e de área colhida, atingindo respectivamente 188,2 milhões de toneladas em cerca de 52,8 milhões de hectares.

Estes números fornecem uma ideia da riqueza gerada pelas propriedades

agrícolas no Brasil que segundo estimativas do INCRA, (2012) constituem 5.498.505 imóveis.

Diante destes números, pode-se perceber que o papel da Engenharia de Avaliações no Brasil, particularmente no que se refere à avaliação de imóveis rurais é desafiador, seja pela quantidade de imóveis, seja pelo tamanho da riqueza produzida pelas propriedades que fazem parte do agronegócio.

Para diversas instituições, tanto públicas quanto privadas, a obtenção do valor de mercado de imóveis rurais é uma atividade estratégica. Como exemplo, pode-se citar a garantia em financiamentos de atividades agrícolas, desapropriações em grandes obras de infraestrutura, desintrusão de unidades de conservação mediante pagamento de indenizações, imobiliárias especializadas em imóveis rurais, fundos de investimento em propriedades agrícolas, elaboração de laudos de avaliação patrimonial de ativos agrícolas para auditorias, consultorias em agronegócio que fornecem relatório dos mercados de terras, a obtenção de imóveis para programas de reforma agrária, financiamento de programas de crédito fundiário, alimentação de Sistemas de Preços de Terras (este último atualmente em funcionamento na Receita Federal e no Ministério do Desenvolvimento Agrário).

Cada um dos responsáveis por estas atividades obtém o valor de mercado dos imóveis utilizando-se de diferentes técnicas de avaliação de propriedades rurais. Estas técnicas estão adaptadas às finalidades às quais são destinadas às informações, entretanto quanto mais importantes são as decisões que serão tomadas a partir das informações sobre o valor dos imóveis, maior a exigência com relação à precisão com a qual este valor é estimado.

A valoração cadastral rural no Brasil tem recebido pouca atenção das instituições governamentais. Os impostos que incidem sobre a terra, Imposto Territorial Rural (ITR) e Imposto sobre a Transmissão de Bens Imóveis (ITBI), têm pouca representatividade em volume arrecadado para os que possuem competência para arrecadá-los. Este fato tem sido usado como justificativa para não aprimorar a cobrança e a fiscalização destes.

Entretanto, essa realidade vem se alterando com a realização de convênios entre a Receita Federal do Brasil (RFB) e os municípios, e os resultados dessa descentralização para vários municípios tem resultado em incrementos significativos

de arrecadação.

A base de cálculo do ITBI é o valor venal do imóvel enquanto que a do ITR é o valor da terra nua; embora sejam diferentes, ambas devem ser baseadas no valor de mercado do imóvel rural. A obtenção de tal valor exige a aplicação das técnicas da Engenharia de Avaliações e, que além disso, sejam modelados efeitos relacionados à localização espacial dos imóveis os quais afetam a sua valoração.

Mesmo com essa possibilidade, ainda se observa a ausência de critérios técnicos, e a não utilização de métodos de avaliação precisos e que sejam transparentes para os contribuintes. Neste sentido, o Brasil encontra-se atrasado tanto com relação aos vizinhos da América do Sul quanto em relação aos países da Europeus.

O ITR possui uma função extrafiscal importante que é evitar a especulação com as terras, fazendo com que estas sejam inseridas nos sistemas produtivos. Entretanto, a falta de fiscalização deste imposto produz situações como a relatada por Silva e Barreto (2014) na qual estimativas indicam que no ano de 2010 haviam cerca de 12 milhões de hectares de pastagens mal utilizadas na amazônia. Segundo os mesmos autores o potencial de arrecadação do ITR no Pará seria de 133 vezes maior que frente ao que é arrecadado atualmente.

Silva e Barreto (2014) apontam entre outras soluções para que haja uma maior efetividade na cobrança e fiscalização do ITR que *“os órgãos públicos envolvidos (Receita Federal e prefeituras) providenciem, anualmente, a coleta ou aquisição de preços de mercado de terras para que sirvam de referência para a fiscalização, e tais dados devem ser disponibilizados facilmente ao público”* e além disso, propõem a utilização de mapas para a fiscalização do ITR: *“Os órgãos fiscalizadores devem submeter as declarações do ITR a uma análise do tipo malha fina com base em informações geográficas e dos preços de terra”*.

Diversos estudos têm mostrado que a incorporação dos chamados efeitos espaciais, mais especificamente a autocorrelação e a heterogeneidade espacial, aos modelos tem trazido maior precisão, o que se traduz em uma maior justiça fiscal. Entretanto, estes estudos utilizando a econometria espacial se concentram em áreas urbanas embora os mercados de terras rurais também possam ser afetados por estes efeitos.

A Planta de Valores Genéricos (PVG) é um documento gráfico que representa a distribuição espacial dos valores médios dos imóveis em uma determinada região. É um importante instrumento da tributação imobiliária por possibilitar uma maior justiça fiscal. Entretanto, na grande maioria dos municípios sua utilização está restrita à área urbana. Existe a necessidade de trabalhos mostrando a importância da aplicação desta ferramenta em mercados de terras rurais.

Alguns estudos têm demonstrado que a utilização do interpolador geoestatístico conhecido como krigagem ordinária permite a obtenção de PVG mais eficientes devido ao fato que as PVG elaboradas por tal interpolador permitem que os valores das parcelas possam ser descritos em função das características da região, bem como definidas as zonas de transição dentro de uma mesma parcela (MARQUES et al., 2012). Ressalta-se que também neste caso os estudos estão restritos às áreas urbanas e a realização de estudos deste tipo em áreas rurais pode trazer benefícios à valoração cadastral rural sendo inclusive uma recomendação dos referidos autores a realização de novos testes, considerando-se novas áreas de estudo e um número maior de elementos amostrais.

A combinação de técnicas pode então permitir a obtenção de dois produtos principais a saber: o modelo de valoração cadastral rural e a PVG municipal, obtidos respectivamente pela econometria e pela geoestatística.

Diante deste cenário, este trabalho propôs a elaboração de um modelo de valoração cadastral rural que leve em consideração a distribuição espacial dos dados de mercado e seja livre dos efeitos espaciais em seus resíduos. Para atingir este objetivo foi utilizada a combinação de técnicas da econometria espacial e da geoestatística, aprimorando os modelos de regressão múltipla. Entre os resultados, foram obtidos o modelo de valoração cadastral dos imóveis rurais e a Planta de Valores Genéricos (PVG) das terras do município onde foi realizado o estudo.

1.1 OBJETIVOS

1.1.1 Objetivo Geral

Avaliar a combinação de técnicas da econometria espacial e da geoestatística na modelagem dos efeitos espaciais em mercados de terras rurais e geração de

PVG, propondo uma metodologia que seja aplicável à realidade dos municípios brasileiros.

1.1.2 Objetivos Específicos

1-Analisar a realidade atual da valoração cadastral rural no Brasil em relação aos demais países, bem como identificar as principais técnicas envolvidas nestes processos;

2-Obter o modelo de valoração cadastral rural livre dos efeitos espaciais por meio da combinação das técnicas da econometria espacial e da geoestatística;

3-Elaborar a PVG utilizando o interpolador geoestatístico krigagem ordinária;

4-Validar e avaliar o desempenho dos produtos obtidos.

1.2 ESTRUTURA DO TRABALHO

A organização da dissertação foi feita em cinco capítulos. O primeiro capítulo é a introdução onde é mostrada uma visão geral da problemática que envolve o trabalho, a justificativa para a sua realização e os objetivos a serem alcançados.

O segundo capítulo trata da fundamentação teórica, onde são apresentados os principais temas que embasam os conhecimentos que serão utilizados ao longo do desenvolvimento deste estudo. Simultaneamente foram abordados os mais recentes trabalhos nos respectivos subitens.

O terceiro capítulo refere-se à metodologia, no qual apresenta-se a região de estudo, e são descritas as etapas, as técnicas e os meios necessários para alcançar os objetivos propostos.

No quarto capítulo são apresentados os resultados obtidos para cada uma das etapas propostas. A partir dos resultados apresentados, são realizadas análises que subsidiam a tomada de decisões que auxiliam no desenvolvimento das etapas subsequentes.

O quinto capítulo foi reservado para as principais conclusões em relação aos objetivos inicialmente propostos, bem como as recomendações aos trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Partindo do pressuposto que atualmente na sociedade da informação nenhuma pesquisa parte do zero, Lakatos e Marconi (1991) destacam a importância da revisão bibliográfica. O objetivo desta etapa é a não-duplicação de esforços, a não "descoberta" de ideias já expressas, a não-inclusão de "lugares-comuns" no trabalho.

Neste sentido, a revisão bibliográfica contemplou a busca no Portal de Periódicos da Capes. Esta é uma biblioteca virtual que reúne e disponibiliza para instituições de ensino e pesquisa no Brasil o melhor da produção científica internacional. Atualmente, são disponibilizados mais de 35 mil periódicos com texto completo, 130 bases referenciais, 11 bases dedicadas exclusivamente a patentes, além de livros, enciclopédias e obras de referência, normas técnicas, estatísticas e conteúdo audiovisual (CAPES, 2014). A UFPE faz parte das instituições que participam e disponibilizam acesso a essa base de dados para os pesquisadores e alunos.

A revisão envolveu as publicações dos últimos cinco anos referentes aos seguintes temas: valoração cadastral rural, econometria espacial aplicada, geoestatística aplicada, avaliação em massa de imóveis rurais, avaliação da qualidade em avaliação em massa, plantas de valores genéricos.

Os artigos encontrados subsidiaram a elaboração, no sentido em que, como afirmam Lakatos e Marconi (1991), a citação das principais conclusões a que outros autores chegaram permite salientar a contribuição da pesquisa realizada, demonstrar contradições ou refirmar comportamentos ou atitudes. Para estes autores tanto a confirmação, em dada comunidade, de resultados obtidos em outra sociedade, quanto a enumeração das discrepâncias são de grande importância.

2.1 CONCEITOS GERAIS

A Engenharia de Avaliações pode ser definida como um conjunto de conhecimentos técnico-científicos especializados, aplicados com o objetivo de identificar o valor de um bem, seus custos, frutos e direitos, bem como determinar indicadores da viabilidade de sua utilização econômica, para uma determinada finalidade, situação e data (ABNT, 2001).

Para fins de avaliações, um imóvel rural pode ser definido como qualquer imóvel que, independente de sua localização, possua vocação para exploração animal ou vegetal (ABNT, 2001). A lei 8629 de 25/02/1993 conceitua imóvel rural de forma um pouco mais detalhada “O prédio rústico de área contínua, qualquer que seja a sua localização que se destine ou possa se destinar à exploração agrícola, pecuária, extrativa vegetal, florestal ou agroindustrial.”

Dantas (2005), ao explicar sobre os diferentes tipos de valor que podem ser aplicados a um determinado bem, refere-se a duas escolas que conceituam o termo valor: a primeira associa o valor de um bem à finalidade de sua avaliação, o que possibilita o surgimento de diversos valores, sendo por isso denominada de plurivalente; a segunda escola define que independente do fim ao qual se destine, a avaliação, o valor do bem é único para determinado instante. Esta proposta é denominada univalente.

Na avaliação de imóveis rurais, o valor que se deseja determinar é o valor de mercado, definido pela ABNT (2001) como “Quantia mais provável pela qual se negociaria voluntariamente e conscientemente um bem, numa data de referência, dentro das condições do mercado vigente.”

Diz-se que um mercado é de concorrência perfeita quando existem muitos vendedores, muitos compradores e um equilíbrio na oferta de bens em relação ao potencial do mercado. Além disso, os vendedores e compradores não devem ter o poder de interferir individualmente ou em conjunto nos preços (DANTAS, 2005). Neste sentido, os mercados de terras agrícolas são imperfeitos, seja pela terra ser um bem escasso, seja pelo fato dela frequentemente estar concentrada em poder de poucos vendedores. Também faz parte da realidade dos mercados de terras a situação na qual determinados grupos de compradores ao se introduzirem nas diferentes regiões, serem capazes de modificar a realidade dos preços praticados. Isso é demonstrado em Feichtinger e Salhofer (2013), onde verifica-se que a participação de autoridades públicas na compra e venda de terras provoca alteração dos preços das mesmas no estado alemão da Bavaria.

2.2 MÉTODOS DE AVALIAÇÃO IMOBILIÁRIA

Para se alcançar o valor de mercado, diferentes métodos podem ser

utilizados, estes podem ser subdivididos em métodos para identificar o valor de um bem, de seus frutos e direitos e métodos para identificar o custo de um bem (ABNT, 2001).

Dantas (2005) observa que os métodos podem ser subdivididos em métodos diretos e indiretos. No primeiro, o resultado da sua aplicação já permite a obtenção do resultado da avaliação; já o segundo ocorre quando é preciso utilizar um método direto em alguma das etapas para se avaliar o bem.

Os métodos para identificar o valor de um bem, de seus frutos e direitos dividem-se em (ABNT, 2001):

1. Método comparativo direto de dados de mercado: objetivam identificar o valor de mercado do bem por meio de tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra. Este é o método recomendado para avaliação de terras nuas pela ABNT (2001); Dantas (2005) afirma que sempre que houverem dados de mercado suficientes este deve ser o método escolhido.
2. Método involutivo: identifica o valor de mercado de um 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.
3. Método evolutivo: identifica o valor do bem pelo somatório dos valores de seus componentes. Caso a finalidade seja a identificação do valor de mercado, deve ser considerado o fator de comercialização.
4. Método de capitalização da renda: existem bens que embora sejam capazes de proporcionar uma geração de renda não são comumente encontrados para comercialização, inviabilizando desta forma o emprego de métodos diretos. Como exemplo, se pode considerar culturas de ciclo longo que produzem safras anuais como cacau, café, laranja. Este método é o recomendado pela ABNT (2001) para avaliação de produções vegetais.

O método de capitalização da renda identifica o valor de um bem, com base na capitalização presente da sua renda líquida prevista durante a vida útil da cultura

ou duração do empreendimento. Lima (2005) afirma que este método se baseia no princípio da antecipação refletido no valor presente dos direitos sobre futuros benefícios intrínsecos da propriedade. O referido método demonstra assim, o relacionamento entre os ganhos líquidos anuais de um investimento numa propriedade e o valor de venda deste capital.

Quanto aos métodos para identificar o custo de um bem estes se dividem em:

1. Método comparativo direto de custo: identifica o custo do bem por meio do tratamento técnico dos atributos dos elementos comparáveis, constituintes da amostra.
2. Método da quantificação de custo: identifica o custo do bem ou de suas partes por meio de orçamentos sintéticos ou analíticos, a partir da quantidade de serviços e respectivos custos diretos e indiretos.

A grande maioria dos trabalhos que buscam desenvolver modelos de avaliação em massa utilizam predominantemente o método comparativo direto de dados de mercado por meio da construção de modelos de regressão.

2.3 SITUAÇÃO ATUAL DA VALORAÇÃO CADASTRAL RURAL

Segundo Averbeck (2003), a avaliação em massa pode ser definida como *“um processo de definição de modelos matemáticos, obtidos a partir da realidade dos valores locais, testados e validados estatisticamente e aplicados na avaliação de uma quantidade de imóveis de uma população”*. O autor destaca ainda que para a realização de avaliações em massa é necessário realizar um estudo do universo dos imóveis que compõem a população, as tipologias destes, seus quantitativos, como estão distribuídos espacialmente, bem como de suas características físicas e locais que permitem diferenciá-los.

Bueno, Barrós e Maseda (2012) afirmam que o cadastro evoluiu de um propósito puramente fiscal para se tornar a base de um Sistema de Informação Territorial com o objetivo de exercer uma gestão adequada do território. Entretanto, argumentam que o primeiro propósito não foi deixado de lado, haja vista que esta é uma das principais fontes de financiamento dos governos.

No Quadro 1 é possível verificar os diferentes critérios e métodos utilizados

por países da Europa e América Latina apresentados por Bueno, Barrós e Maseda (2012) para realizar a valoração de imóveis rurais. Nela percebe-se que existe um predomínio da utilização do critério de capitalização da renda tanto nos países Europeus quanto da América Latina. Entretanto, Alcázar e Ariza (2004) argumentam que a renda embora seja um dos principais componentes formadores dos preços, por si só não é capaz de explicar o valor de uma propriedade e por isso, outras variáveis que também influenciam na formação dos preços deveriam ser consideradas.

Quadro 1 - Modelos de Valoração Cadastral na Europa e América Latina.

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 (evidências, com segmentação do território)
República Eslovaca	Capitalização da Renda	Autodeclaração
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	Autoavaliação
México	Capitalização da Renda	Comparação no mercado
Nicarágua	Capitalização da Renda	Comparação no mercado

Fonte: Adaptado de Bueno, Barrós e Maseda 2012.

Na Argentina, a valoração cadastral rural vem evoluindo e incorporando as mais recentes ferramentas tecnológicas disponíveis (BENIGNI, 2008; BENIGNI, 2010; BONILLA, ROBLEDO, LÓPEZ, 2010). Para exemplificar, estão sendo desenvolvidos sistemas nos quais as variáveis influenciadoras do mercado são obtidas a partir de bases de dados disponibilizadas em SIG; imagens de satélite de

alta resolução são utilizadas para realizar estimativas de áreas plantadas; são utilizados índices de vegetação para determinar o rendimento potencial das culturas; realiza-se a aquisição de informações georreferenciadas em campo para inserção no banco de dados; são obtidas séries históricas de preços das diferentes atividades para não penalizar o produtor em função das oscilações típicas dos preços agrícolas. Tais sistemas vão na direção de obterem o valor das parcelas de forma automática, e oferecerem maior objetividade, precisão e facilidade de atualização. Como consequência, há uma maior equidade da carga tributária.

De acordo com Steiner et al. (2008), no Brasil a avaliação de imóveis com a finalidade realizar a cobrança de impostos, venda, garantia em financiamentos ou outros, na maioria das vezes, não utiliza qualquer procedimento científico de forma sistemática, baseando-se na experiência pessoal dos profissionais.

Com relação à obtenção da base de cálculo dos impostos que incidem sobre os imóveis rurais a mesma se dá sem a utilização da Engenharia de Avaliações. O ITR é autodeclaratório, enquanto que para a cobrança do ITBI os municípios utilizam como base de cálculo os valores declarados no cartório (geralmente subdeclarados), ou então utilizam tabelas de valores por região, estabelecidos sem critérios técnicos e que reconhecidamente não representam a realidade dos mercados pois frequentemente estão desatualizadas. Tais fatos revelam o atraso do Brasil com relação ao tema da valoração cadastral rural.

De acordo com Oliveira, (2010) nos países desenvolvidos a cobrança de impostos sobre a terra é plenamente estabelecida e o valor arrecadado em relação ao PIB é significativo. Num exemplo, esse mesmo autor, apresenta o percentual de arrecadação de alguns países em relação ao PIB para o período de 2002 a 2005: Estados Unidos: 3,09%; Canadá: 3,05%; Argentina: 0,82%; enquanto que no Brasil essa arrecadação é de 0,46%.

2.4 VARIÁVEIS INFLUENCIANTES NA VALORAÇÃO CADASTRAL RURAL

Alcázar e Ariza (2004) afirmam que a valoração cadastral rural torna-se mais complexa que a urbana. Isto se explica, segundo esses mesmos autores, pela maior complexidade na obtenção de variáveis que possibilitem a elaboração de modelos

econométricos consistentes.

O valor das terras nas regiões de um município está sujeito à influência de diversas variáveis. Lima (2005) aponta que as principais são: capacidade de uso das terras, situação, dimensão, potencial de irrigação, pluviosidade, porcentagem de área com cobertura florística natural, número de açudes, porcentagem de área que já foi aberta, dentre outras que podem existir em mercados específicos.

Alcázar e Ariza (2004) relacionam diversas variáveis que podem contribuir para a construção de modelos de valoração cadastral rural como a classificação agroecológica dos solos, aptidão agrícola das terras, zoneamento agrícola das culturas, clima, declividade, precipitação, textura do solo superficial, profundidade efetiva do perfil, drenagem, nível de fertilidade, acidez, ocorrência de geadas e erosão entre outras.

Entretanto, afirmam que o cruzamento de diferentes mapas e dados georreferenciados pode ser interessante caso haja disponibilidade de tais informações no nível de detalhe (escala) que possibilite a identificação e valoração de todas e de cada uma das parcelas ou unidades produtivas nas quais as propriedades podem se dividir. Caso contrário utilizar variáveis obtidas em escalas pequenas para avaliar parcelas cadastrais (levantadas em escalas grandes) desvirtuaria os modelos (ALCÁZAR; ARIZA, 2004).

Dubin (1992) considera que o principal fator determinante do preço de um imóvel é a sua localização. Michael et al. (2006) argumenta que propriedades com características similares e próximas apresentam um valor de mercado semelhante, atribuindo este fato a um “*valor de localização*” produzido pela imobilidade, sendo que a semelhança de valores tende a ter um comportamento inversamente proporcional à distância de separação entre os imóveis.

Resener e Hochheim (2006) atribuem à localização o mesmo grau de importância que às suas variáveis intrínsecas. Além disso, se referem à influência recíproca que os imóveis próximos exercem, chamando-a de “efeito de vizinhança”. Afirmam também que os imóveis rurais tem seu valor de localização influenciado não apenas pela distância da cidade mais próxima, mas por outros fatores de localização que atuam simultaneamente como a distância à agroindústria, posição em relação aos eixos de desenvolvimento e presença de outras cidades no entorno.

Resener e Hochheim (2006) também se referem à relação entre a localização e a pedologia, pois os tipos de solos existentes influenciam na implantação de determinadas culturas e com elas a infraestrutura necessária ao seu desenvolvimento. Assim, o valor de localização acaba por refletir o valor do solo em função da sua pedologia.

A influência que a localização exerce sobre o valor das terras dos imóveis rurais criou um desafio para a Engenharia de Avaliações pois, como o Valor da Terra Nua (VTN) é algo que apresenta dependência espacial, não é possível tratá-lo com as técnicas da estatística clássica. Então, tem-se o desafio de como incorporar os efeitos provocados pela localização espacial dos imóveis no estudo das variáveis que influenciam os valores das terras. É assim que no Brasil, alguns autores passam a estudar a autocorrelação espacial ou dependência espacial (DANTAS et al., 2001; DANTAS et al., 2003 (a); DANTAS et al., 2003 (b); DANTAS et al., 2006; HORNBURG; HOCHHEIM, 2009; MICHAEL et al., 2006) e sua influência nos preços de imóveis urbanos, passando a incorporar aos modelos inferenciais variáveis que reflitam esta influência. Essa metodologia ficou conhecida por Avaliação por Inferência Espacial.

No que se refere a imóveis rurais, poucos trabalhos se dedicaram ao tratamento da autocorrelação espacial. Assim, vale destacar o trabalho de Resener e Hochheim (2006) que buscaram avaliar a viabilidade de utilização da regressão linear na avaliação em massa de imóveis afetados por empreendimentos hidrelétricos e utilizaram a técnica da Análise de Superfície de Tendência (TSA) para tentar solucionar os efeitos da autocorrelação espacial. Esses autores concluíram, ao final do estudo, que ao comparar os valores de indenização dos imóveis rurais obtidos por meio da homogeneização, com os que foram analisados através da TSA houve uma tendência de supervalorização dos imóveis afetados pelo empreendimento, pois a homogeneização não considerou as diferenças de valores regionais, bem como também não considerou a área sem exploração econômica.

Outro trabalho com imóveis rurais foi realizado por Machado (2006), que utilizou a combinação da Regressão Múltipla com a TSA para gerar a PVG do município de Chapecó-SC. Utilizou uma amostra de 47 imóveis rurais, para obter um modelo no qual as principais variáveis influenciadoras foram: área, acesso, classes de terras, e um polinômio de tendência de quarto grau. O modelo obtido permitiu a

obtenção de uma malha de valores calculados para a região de abrangência da amostra, a qual permitiu a elaboração da PVG.

Lourival et al. (2008) utilizaram modelos de regressão múltipla para analisar o comportamento das variáveis que influenciam o valor da terra no Pantanal Sul-matogrossense. Neste trabalho propuseram um método alternativo para determinar a equivalência da terra para fins de compensação de áreas de reserva legal. Utilizaram uma amostra de 106 transações entre os anos de 2000 a 2003 e 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 em explicar as mudanças nos valores de terras.

Na região estudada, o desenvolvimento de atividades que degradam a biodiversidade, como a intensificação da agricultura, aumentam substancialmente o valor das propriedades rurais. Por outro lado, as variáveis favoráveis à biodiversidade (por exemplo, presença de reservas) influenciaram para diminuir o valor de mercado das propriedades. Quanto aos mecanismos de compensação, que atualmente funcionam com base no tamanho da área ou seja, 1 ha compensado para 1 ha degradado, os modelos obtidos mostraram que 1 ha de propriedade totalmente desenvolvido na zona de maior valorização econômica tem um valor de equivalência de até 14 ha na planície de inundação do Pantanal. Isto faz com que, valendo-se da legislação atual, desenvolvedores possam comprar terras mais baratas em regiões adjacentes para compensar a destruição de habitat em uma região com maior potencial agrícola e, conseqüentemente, maior o valor da terra. Isso levou os autores a proporem a utilização da modelagem para conhecer a assimétrica equivalência do valor da terra e dessa forma colaborar para a preservação ambiental de áreas ameaçadas pelo avanço das atividades agrícolas (LOURIVAL et al., 2008).

Silva, Nali e Marote (2009) utilizaram um modelo de regressão múltipla e a

estimativa dos parâmetros pelo método dos mínimos quadrados para avaliar imóveis rurais no Submédio São Francisco. Na amostragem, foram utilizados apenas dados de imóveis utilizados com agricultura irrigada ou com potencial para esta atividade. Seus resultados indicaram a influência positiva da variação da data de transação ou da coleta de informação de oferta, do nível de infraestrutura, da elasticidade da oferta, da percentagem da área irrigável, e da área equivalente de produção vegetal, além de influência negativa para a dimensão da área.

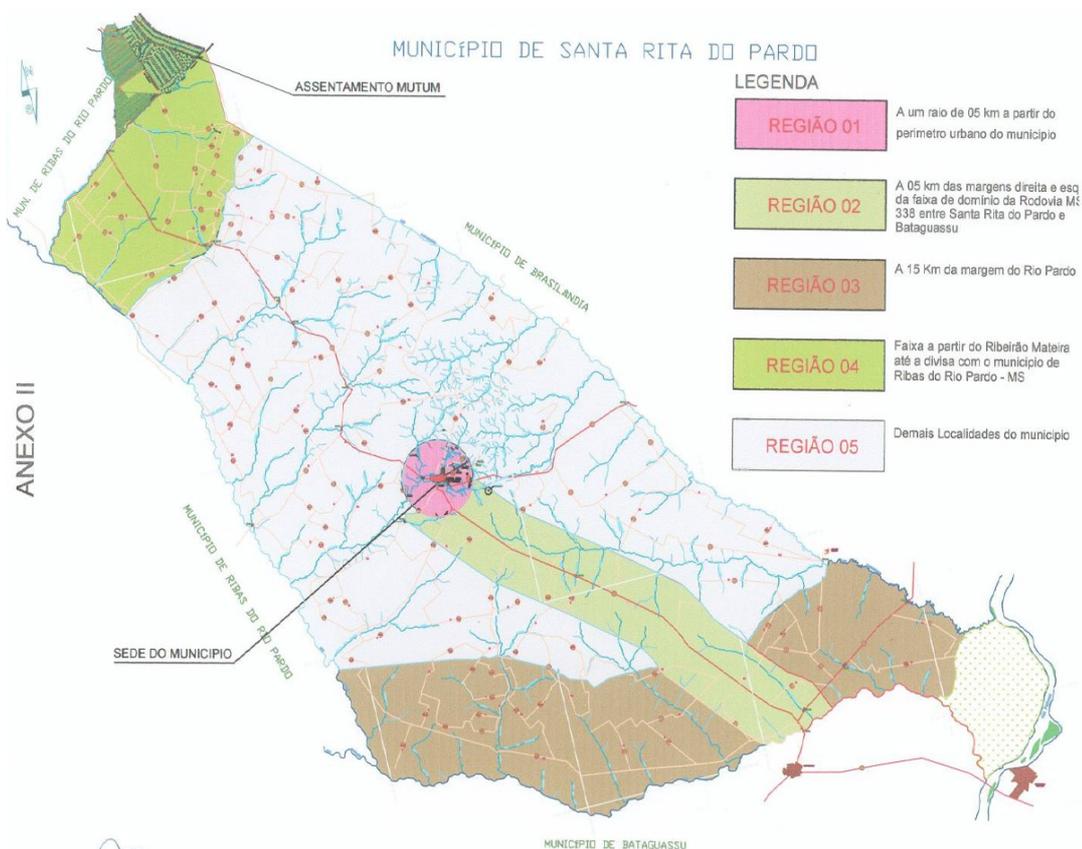
Trivelloni e Hochheim (2006) se referem a duas escolas quando o assunto é o tratamento espacial de dados do mercado imobiliário: a primeira utiliza técnicas de econometria espacial para geração de modelos de regressão espacial e a segunda utiliza as técnicas geoestatísticas. Em seu trabalho, esses autores utilizaram a regressão espacial para gerar uma malha densa de pontos que permitiu que a krigagem fosse aplicada na interpolação para elaborar a PVG sem gerar erros na estimação dos valores unitários. Dessa forma, demonstrou-se que a conjugação de ambas metodologias possibilita superar algumas das deficiências e limitações destes métodos, potencializando ao mesmo tempo as suas vantagens e alcançando resultados satisfatórios.

2.5 PLANTAS DE VALORES GENÉRICOS

Entre os instrumentos utilizados pelas administrações municipais para avaliações em massa com fins tributários destaca-se a PVG. Segundo Michael et al. (2006) *“A principal função da PVG é permitir a definição de uma política de tributação imobiliária que seja justa e tenha equidade. A Planta de Valores Genéricos consiste em um documento gráfico que representa a distribuição espacial dos valores médios dos imóveis em cada região da cidade, normalmente apresentados por face de quadra”*. Como se pode perceber nesta definição, a grande maioria dos trabalhos envolvendo PVG se referem a imóveis urbanos, mas existe potencial para aplicação desta técnica aos mercados de terras rurais.

Entretanto, a elaboração de plantas de valores pela metodologia tradicional pode gerar resultados tendenciosos, com consequências danosas tanto para os consumidores como para a administração municipal Dantas et al. (2003).

Na Figura 1 é mostrado o exemplo da PVG de imóveis rurais do município de Santa Rita do Pardo-MS. Como pode-se observar, a planta é composta por um mapa das regiões e pela tabela de preços das terras por hectare. Esta representação difere da realidade da maioria dos municípios brasileiros. O referido município é conveniado com a RFB para fiscalização do ITR desde novembro de 2008 (BRASIL, 2014).



REGIÃO	LOCALIZAÇÃO	VALOR/HA
Região 01	Raio de 05km(cinco quilômetros) a partir do perímetro urbano do Município;	R\$ 6.661,57(seis mil, seiscentos e onze reais e cinquenta e sete centavos)
Região 02	05Km(cinco quilômetros) das margens direita e esquerda da faixa de domínio da Rodovia MS 338 entre Santa Rita do Pardo – MS, e Bataguassu – MS;	R\$ 4.959,00(quatro mil novecentos e cinquenta e nove reais)
Região 03	15Km(quinze quilômetros) a partir da margem do rio pardo;	R\$ 4.545,45(quatro mil, quinhentos e quarenta e cinco reais e quarenta e cinco centavos)
Região 04	Faixa a partir do Ribeirão Mateira, até a divisa com o Município de Ribas do Rio Pardo – MS;	R\$ 2.479,39(dois mil, quatrocentos e setenta e nove reais e trinta e nove centavos)
Região 05	Demais localidades do Município;	R\$ 4.132,29(quatro mil, cento e trinta e dois reais e vinte e nove centavos)

Figura 1 - Planta de Valores Genéricos dos imóveis rurais do município de Santa Rita do Pardo-MS.

Fonte: <http://www.cmsantaritadopardo.ms.gov.br/autografos/2009/19.pdf>

A elaboração de PVG vem incorporando as técnicas da Inferência Espacial e da krigagem (DANTAS et al., 2003; HORNBURG, 2009; TRIVELLONI e HOCHHEIM, 2006), o que tem permitido estudar a influência que a localização espacial de um determinado imóvel exerce sobre a formação dos preços, além da influência que este imóvel exerce sobre os seus vizinhos. Com isso, é possível alcançar uma melhor precisão nos trabalhos de valoração de imóveis e a justa cobrança dos impostos, beneficiando o município e sua população. Tais trabalhos tem sido desenvolvido em áreas urbanas.

Dantas et al. (2003), ao utilizar a inferência espacial na elaboração de PVG, mostrou que tal metodologia pode ter grande utilidade, uma vez que pode encontrar valores mais justos para os imóveis, identificando-se com facilidade as zonas infra ou supra valoradas de uma cidade. Isto é, as zonas beneficiadas ou prejudicadas fiscalmente, o que pode resultar em maior arrecadação de impostos e bem-estar social.

Liporoni (2007) afirma que a utilização dos Sistemas de Informações Geográficas (SIG) na elaboração de plantas de valores agrega um poderoso ferramental, por permitirem a realização de análises espaciais, construção de entidades espaciais, modelagem de terrenos, criação de superfícies e isolinhas de valores, e a atribuição de características às entidades espaciais. Além disso, os SIG disponibilizam funções geoestatísticas, permitindo enriquecer a qualidade das informações e testar a consistência dos dados coletados em campo, possibilitando simulações de plotagem e análise de seus resultados.

As estimativas geoestatísticas possibilitam a realização de interpolação por krigagem ordinária. Esta técnica permite pontuar alguns aspectos referentes à espacialidade dos dados, tais como a disposição dos elementos na área, a proximidade entre eles e o tamanho da amostra, isso porque são fatores que podem influenciar diretamente na definição dos parâmetros da modelagem e, portanto, na estimativa dos valores que compõem a PVG (MARQUES et al., 2012).

Neste sentido, de acordo com Marques et al. (2012), a utilização da krigagem ordinária, acompanhado de métodos de classificação de dados, pode ser uma nova abordagem para a produção de PVG, a qual além de auxiliar no planejamento urbano, também promove a justiça fiscal e social na cidade.

Na maioria dos municípios brasileiros as PVG se referem apenas aos imóveis urbanos. Quanto aos imóveis rurais, 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 Eng. Agrônomo. Frequentemente tais valores são estabelecidos com base na experiência dos profissionais que atuam na região, e não são utilizadas técnicas de avaliação preconizadas pela NBR 14653-1 e 3.

2.6 CADASTRO TERRITORIAL MULTIFINALITÁRIO

A Portaria n.º 511 do Ministério das Cidades publicada no Diário Oficial da União em 8 de dezembro de 2009, orienta a implementação do cadastro através das Diretrizes para a Criação, Instituição e Atualização do Cadastro Territorial Multifinalitário (CTM) nos Municípios Brasileiros.

O CTM, conceito introduzido pela referida portaria pode ser definido da seguinte forma:

Art. 1º O Cadastro Territorial Multifinalitário (CTM), quando adotado pelos Municípios brasileiros, será o inventário territorial oficial e sistemático do município e será embasado no levantamento dos limites de cada parcela, que recebe uma identificação numérica inequívoca.

Em seguida no Artigo 2º a portaria traz a definição de parcela como sendo:

Art. 2º A parcela cadastral é a menor unidade do cadastro, definida como uma parte contígua da superfície terrestre com regime jurídico único.

Carneiro, Erba e Augusto (2012) expõem uma problemática existente com relação à diferentes definições existentes para a unidade imobiliária rural. A primeira oriunda do Direito Civil e utilizada no registro imobiliário, considera a matrícula como tal unidade; enquanto que a segunda, originária do Direito Agrário e utilizada pelo Instituto Nacional de Colonização e Reforma Agrária (INCRA), considera a unidade econômica rural a qual engloba áreas registradas e áreas de posse.

Entretanto, por meio de alguns exemplos Carneiro, Erba e Augusto (2012) demonstram que a utilização do conceito da parcela como definido pela Portaria

511/2009 permite o atendimento às necessidades do registro imobiliário, do INCRA, da representação de áreas de restrições ambientais, bem como outras de áreas de interesse, tituladas ou apenas de posse, garantindo assim o atendimento aos diferentes usuários das informações do CTM.

O caráter multifinalitário do cadastro visa atender às necessidades sociais, ambientais, econômicas, da Administração Pública e de segurança jurídica da sociedade. Neste sentido, as informações disponíveis devem ser compartilhadas e assim melhorar a gestão das instituições. No caso da valoração cadastral rural, esta pode ser bastante beneficiada com as informações armazenadas no CTM como os limites das parcelas, informações dos proprietários, as benfeitorias existentes, bem como as informações sobre a utilização dos solo. Em contrapartida a valoração cadastral rural pode fornecer para as instituições e os cidadãos informações precisas a respeito da valor das parcelas.

Alcázar (2012) indica algumas das vantagens proporcionadas por uma norma de valoração cadastral rural: a gestão cadastral terá uma base de dados que reflita a realidade do mercado de terras; as autoridades locais terão à sua disposição a base de cálculo para calcular os impostos sobre a propriedade; diversas instituições e os cidadãos terão informação confiável para utilizar como parâmetros úteis para transferências privadas de capitais, desapropriações, estudos estatísticos, fins sociais e hipotecas.

Paixão, Nichols e Carneiro (2012) listam diversos benefícios oriundos da implementação do CTM, dentre os quais destaca-se que a avaliação para os impostos sobre a propriedade imobiliária pode alcançar maior precisão, inclusive pela inserção de variáveis obtidas no próprio sistema cadastral, como características físicas das propriedades, outras relacionadas à localização e infraestrutura disponível. A consequência é que os impostos passam a ser aplicados de forma mais justa.

Reydon e Oliveira (2012) ao defenderem que o Brasil crie mecanismos administrativos e políticos que controlem a propriedade da terra por meio da governança e da regulação, apontam para a adoção de boas práticas de governança territorial. Esta pode ser entendida como o estabelecimento da sustentabilidade como pré-requisito em todo o gerenciamento territorial e a implementação de mecanismos para que se institucionalize tal mercado de forma contínua e em

sincronismo com o desenvolvimento econômico sustentável.

Segundo Reydon e Oliveira (2012), o CTM seria o principal instrumento para o estabelecimento da governança territorial que seja eficiente e equitativa, e que além disso, promova a função social do uso e ocupação do solo, colaborando para uma maior justiça social.

2.7 IMPOSTOS INCIDENTES SOBRE OS IMÓVEIS RURAIS NO BRASIL

O ITR, de acordo com o Código Tributário Nacional (BRASIL, 1966) é de competência federal e tem como fato gerador a propriedade, o domicílio útil ou a posse de imóvel rural; sua base de cálculo é o Valor da Terra Nua, assim entendida como o valor da terra desprovida de suas benfeitorias; tem como contribuinte o detentor imóvel seja ele proprietário ou posseiro. Embora seja de competência federal, a Constituição de 1988 (BRASIL, 1988) permite que sejam firmados convênios entre a União, intermediada pela Secretaria da Receita Federal, e os municípios aos quais são delegadas a fiscalização e cobrança do ITR.

De acordo com Confederação Nacional dos Municípios (2012), os resultados alcançados por municípios que aderiram ao convênio, como Ribas do Rio Pardo (MS), que aumentou sua arrecadação do ITR em 412% passando de R\$983.954,26 em 2008 para R\$4.057.514,11 em 2011 e Jataí (GO) que aumentou sua arrecadação em 674% passando de R\$350.790,18 em 2008 para R\$2.367.481,73 em 2011, mostram ser vantajoso o convênio com a união.

Entretanto, informação da Receita Federal do Brasil – RFB (BRASIL, 2014) revela que apenas 35% dos 5565 municípios firmaram convênio. No estado de Pernambuco existem apenas 8 municípios conveniados. Uma taxa de adesão que pode ser considerada baixa para um convênio que tem tamanho potencial de arrecadação. Conforme o Artigo 153 § 4º III da Constituição (BRASIL, 1988), o estabelecimento do convênio não prevê qualquer possibilidade de redução do imposto ou outra forma de renúncia fiscal. Assim, os municípios devem manter atualizados o VTN do seu território de modo que a base de cálculo do referido imposto esteja atualizada e reflita os preços de mercado praticados.

Oliveira (2010) realizou um estudo sobre a viabilidade da descentralização da

cobrança do ITR em dois municípios do estado de São Paulo. Realizou simulações utilizando um valor médio de alíquota do ITR e valor médio de mercado para o preço da terra por hectare. Seus resultados mostram que a descentralização da cobrança do ITR permite ampliar significativamente a sua arrecadação, a qual para os dois municípios estudados foi em torno de 340%. O autor também afirma que o CTM tem importante papel nesta descentralização e entre os resultados esperados tem-se a manutenção das informações fidedignas com a situação de fato e também o consequente aumento da arrecadação do ITR. A implementação da descentralização deve ser tratada em função das características dos municípios, entretanto, acreditam os autores que municípios com população acima de 100.000 habitantes possuem infraestrutura tanto física quanto de recursos humanos para gerir isoladamente seu perímetro. Com a implementação da cobrança do ITR espera-se proporcionar maior equidade tributária, melhor governança das terras e do uso e ocupação do solo.

O imposto sobre a transmissão de bens imóveis e de direitos a eles relativos (ITBI) de acordo com Art.156, II, da Constituição Federal (BRASIL, 1988) é de competência municipal, sua função é predominantemente fiscal. Este tem como fato gerador a transmissão, “inter vivos”, a qualquer título, de propriedade ou domínio útil de bens imóveis; quando há a transmissão a qualquer título de direitos reais sobre imóveis, exceto os direitos reais de garantia; ou quando há a cessão de direitos relativos às transmissões acima mencionadas.

O contribuinte do imposto, segundo o disposto no artigo 42 do Código Tributário Nacional (Lei nº 5.172, de 25 de Outubro de 1966), é qualquer das partes na operação tributada, como dispuser a lei, que, no caso, será lei municipal. A alíquota utilizada é fixada em Lei ordinária do município competente, os municípios costumam adotar a alíquota de 2% sobre o valor venal. A base de cálculo é o valor venal dos bens ou direitos transmitidos à época da operação que significa o valor de venda, de mercado e que deveria estar representado na PVG, de acordo com Averbeck (2007).

O grande número de transações realizadas nos mercados imobiliários dos grandes centros urbanos faz com que nestes, a importância do ITBI em relação às receitas correntes municipais seja menor que 3%, enquanto que nos pequenos municípios seja em por volta de 1% (AVERBECK, 2007).

Averbeck (2007) afirma ser comum a subdeclaração dos valores do ITBI para

fazer valer os valores da PVG do município (geralmente desatualizadas) e com isso conseguir sonegar parte do ITBI e do Imposto de Renda. Aponta que entre as providências para evitar a subdeclaração está a atualização do cadastro e da planta de valores municipal.

Neste sentido, tem-se reforçada a proposta do Cadastro Territorial Multifinalitário, a qual abrange levantamento sistemático do território municipal, no qual as menores unidades seriam as parcelas cadastrais. Esta delimitação possibilitaria a obtenção de informações detalhadas tanto sobre os imóveis urbanos quanto dos imóveis rurais, facilitando a obtenção de variáveis que auxiliem no processo de avaliação dos imóveis e com isso contribuindo para uma maior justiça fiscal.

Os próximos tópicos abordarão os principais modelos econométricos utilizados ao longo do trabalho. Será apresentada também a modelagem da dependência espacial por meio da econometria espacial e da geoestatística. Por fim, serão apresentadas as formas de avaliar o desempenho dos modelos.

2.8 MODELO CLÁSSICO DE REGRESSÃO LINEAR (MCRL)

A regressão linear pode ser definida como sendo o estabelecimento de uma relação funcional entre duas ou mais variáveis envolvidas para a descrição de um fenômeno. As variáveis que se pretende determinar são denominadas de dependentes e as que são responsáveis pela sua formação são as variáveis independentes.

A análise de um gráfico de dispersão pode sugerir uma relação funcional entre as variáveis, como por exemplo, uma reta, uma exponencial, etc. Surge neste caso, o modelo estatístico denominado de regressão linear simples.

O modelo de regressão linear simples pode ser representado pela função linear:

$$Y_i = \beta_0 + \beta_1 X_1 + \varepsilon_i, i = 1, \dots, m \quad (1)$$

onde:

Y_1, \dots, Y_m - é a variável dependente ou variável resposta;

X_i - é a variável independente ou variável explicativa;

β_0, β_1 - são parâmetros populacionais;

$\varepsilon_1, \dots, \varepsilon_n$ - são os erros aleatórios do modelo;

Uma generalização dessa regressão é conhecida como regressão múltipla. Na engenharia de avaliações geralmente os preços são formados por um conjunto de variáveis, por isso, o modelo de regressão múltipla é o mais frequentemente utilizado.

O modelo de regressão linear múltipla é dado por:

$$\begin{aligned}
 Y_i &= \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_{p-1} X_{ip-1} + \varepsilon_i, \quad i=1, \dots, m \\
 &= \beta_0 + \sum_{j=1}^{p-1} (\beta_j X_{ij}) + \varepsilon_i,
 \end{aligned} \tag{2}$$

Onde:

Y_1, \dots, Y_m - é a variável dependente ou variável resposta;

β_0, \dots, β_p - são parâmetros populacionais;

X_{i1}, \dots, X_{ip} - são as variáveis independentes ou variáveis explicativas;

$\varepsilon_1, \dots, \varepsilon_n$ - são os erros aleatórios do modelo;

Embora existam diversos métodos para determinação dos parâmetros da regressão, o mais frequentemente utilizado é o Método dos Mínimos Quadrados (MMQ), ou Mínimos Quadrados Ordinários (MQO) ou OLS (Ordinary Least Squares). Trata-se uma técnica de otimização matemática que procura encontrar o melhor ajuste para um conjunto de dados tentando minimizar a soma dos quadrados das diferenças entre o valor estimado e os dados observados (tais diferenças são chamadas resíduos).

Neste ponto, é importante esclarecer que a diferença entre os termos erros e resíduos é discutida no item 2.12.2.

2.9 ESTIMATIVA POR MÍNIMOS QUADRADOS

No processo de estimação por mínimos quadrados dos coeficientes de um modelo de regressão o que se busca é a determinação dos coeficientes que

minimizam a soma do quadrado dos erros. O modelo de regressão linear múltipla dado pela Equação 3 produz a seguinte soma de quadrado dos erros (SQE) (TACONELI, 2010):

$$SQE = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(Y_i - \beta_0 - \sum_{j=1}^{p-1} \beta_j X_{ij} \right)^2 \quad (3)$$

Os estimadores de mínimos quadrados para os coeficientes do modelo são os valores dos β 's que minimizam SQE, e devem satisfazer às seguintes condições:

$$\frac{\partial SQE}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^n \left(Y_i - \hat{\beta}_0 - \sum_{j=1}^{p-1} \hat{\beta}_j X_{ij} \right) = 0 \quad (4)$$

$$\frac{\partial SQE}{\partial \hat{\beta}_j} = -2 \sum_{i=1}^n \left(Y_i - \hat{\beta}_0 - \sum_{j=1}^{p-1} \hat{\beta}_j X_{ij} \right) X_{ij} = 0, j=1, 2, \dots, p-1 \quad (5)$$

Matricialmente, a soma de quadrado dos erros do modelo de regressão linear múltipla pode ser representada por:

$$SQE = \sum_{i=1}^n \varepsilon_i^2 = \varepsilon' \varepsilon = (Y - X \beta)' (Y - X \beta) = \quad (6)$$

$$= (Y' Y - \beta' X' Y - Y' X \beta + \beta' X' X \beta) = \quad (7)$$

$$= (Y' Y - 2 \beta' X' Y + \beta' X' X \beta) = \quad (8)$$

O vetor de estimadores de mínimos quadrados de β é o vetor $\hat{\beta}$ tal que SQE é mínima. Utilizando-se da álgebra matricial, obtém-se:

$$\hat{\beta} = (X' X)^{-1} X' Y \quad (9)$$

Assim sendo, o estimador de mínimos quadrados para β é dado por $\hat{\beta} = (X' X)^{-1} X' Y$, desde que $(X' X)^{-1}$ exista. A condição para que a matriz $(X' X)^{-1}$ exista é de que nenhuma das colunas de X seja obtida a partir da combinação linear das demais. O estimador de mínimos quadrados obtido $\hat{\beta}$ tem as seguintes

propriedades (TACONELI, 2010):

1. $\hat{\beta}$ minimiza a soma de quadrado dos erros independente de qualquer propriedade quanto à distribuição dos erros;
2. Os elementos de $\hat{\beta}$, que são combinações lineares das observações Y_1, Y_2, \dots, Y_n , são estimadores não viciados dos elementos de β , tendo variância mínima dentre os estimadores não viciados que são combinações lineares dos Y_i 's, independente das propriedades da distribuição dos erros;
3. Sob as pressuposições de que os erros são independentes, normalmente distribuídos com média zero e variância constante, ou seja, $\varepsilon \sim Normal(0, \sigma^2)$, então $\hat{\beta}$, além de estimador de mínimos quadrados, é o estimador de máxima verossimilhança de β .

Sendo $\hat{\beta}$ o vetor de estimadores de mínimos quadrados para os coeficientes do modelo de regressão linear múltipla. A média da variável aleatória Y para um particular vetor de valores das variáveis explicativas $X_h = (1, X_{1h}, X_{2h}, \dots, X_{p-1h})$ é estimada por:

$$\hat{Y}_h = X_h \hat{\beta}. \quad (10)$$

Sabendo-se que X é a matriz do modelo. O vetor de valores ajustados, pelo modelo de regressão obtido, para os valores da variável resposta Y , é calculado por:

$$\hat{Y} = X \hat{\beta}, \quad (11)$$

O vetor de resíduos pode ser obtido a pela diferença do vetor de valores observados de Y para o vetor de valores ajustados, ou seja,

$$e = Y - \hat{Y} \quad (12)$$

O estimador de σ^2 é fundamentado na soma de quadrado dos resíduos (SQres):

$$SQ_{Res} = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2 = e' e = \quad (13)$$

$$= (Y - X \hat{\beta})' (Y - X \hat{\beta}) = Y' Y - \hat{\beta}' X' Y = \quad (14)$$

A soma de quadrado dos resíduos têm $n-p$ graus de liberdade associados a ela, uma vez que p parâmetros são estimados no modelo de regressão. Assim, define-se como quadrado médio de resíduos (QM_{Res}):

$$QM_{Res} = \frac{SQ_{Res}}{n-p} \quad (15)$$

O valor esperado do QM_{Res} é σ^2 , ou seja, QM_{Res} é um estimador não viciado da σ^2 .

2.10 INTERVALO DE CONFIANÇA PARA PREDIÇÃO DE UMA NOVA OBSERVAÇÃO

A predição de uma nova observação Y_h^* refere-se à previsão de um resultado individual de Y para um valor X_h de X . Taconeli (2010) afirma que apesar do preditor pontual ser igual ao estimador da resposta média $Y_h^* = \hat{Y}_h = \beta_0 + \beta_1 X_h$, deve-se incorporar, na determinação da variância do preditor, a variabilidade das respostas em torno da resposta média. A variância de Y_h^* pode ser decomposta em dois componentes: a variância da distribuição amostral de \hat{Y}_h e a variância da distribuição de Y . O estimador usual para esta variância é dado por:

$$\hat{Var}(Y_h^*) = \hat{Var}(\hat{Y}_h) + QM_{Res}, \quad (16)$$

o qual pode ser representado da seguinte forma:

$$\widehat{Var}(Y_h^*) = \left[1 + \frac{1}{n} + \frac{(X_h - \bar{X})^2}{\sum (X_i - \bar{X})^2} \right] QM_{Res} \quad (17)$$

Os limites de um intervalo de predição com $100(1-\alpha)\%$ de confiança tem os seguintes limites:

$$IC(Y_h, 1-\alpha) = \left[\hat{Y}_h^* \pm t_{(1-\alpha/2; n-2)} \hat{DP}(Y_h^*) \right] \quad (18)$$

2.11 PRESSUPOSTOS DO MCRL

No processo de estimação dos parâmetros, objetiva-se que o estimador MQO seja o melhor estimador linear não viesado (best linear unbiased estimator – BLUE). Para que este objetivo seja alcançado, foram definidos uma série de cinco hipóteses conhecidas como Hipóteses de Gauss-Markov a saber (ALMEIDA, 2012):

1) A relação funcional entre a variável dependente Y e as variáveis explicativas X é linear nos parâmetros através das observações.

2) As variáveis explicativas não devem ser colineares perfeitamente, além disso, uma variável explicativa não deve ser expressa por valores constantes, sem variação nos dados.

3) A esperança matemática entre as variáveis independentes e os termos do erro é zero $E(\varepsilon_i|X)=0$. Para que esta hipótese não seja violada algumas condições devem ser atendidas:

3.1) Na especificação, nenhuma variável independente relevante pode ficar de fora do modelo.

3.2) Os erros e as variáveis explicativas não devem ter nenhuma relação de dependência linear ou não linear entre quaisquer pares de regiões.

3.3) Assume-se que não existe erro no processo de medição da variável dependente e das variáveis independentes.

4) Os erros são variáveis aleatórias nos quais a variância do erro condicional às variáveis explicativas é constante: $Var(\varepsilon|X)=\sigma^2$, hipótese conhecida como da homocedasticidade.

5) Os erros são não correlacionados, ou seja, o termo do erro observado na unidade de corte transversal i não depende do erro observado na unidade j :

$$\text{Corr}(\varepsilon_i \varepsilon_j | x) = \text{Corr}(\varepsilon_i \varepsilon_j) = 0, \text{ para } i \neq j \quad (19)$$

De acordo com Almeida (2012), tais hipóteses provam o teorema de Gauss-Markov: as hipóteses de linearidade, não colinearidade e da média condicional zero são necessárias para se provar a não-tendenciosidade do estimador MQO. Somando-se a estas, as hipóteses da homocedasticidade e da independência dos erros, prova-se que o estimador MQO é o mais eficiente entre todos os estimadores lineares não-tendenciosos. Com estas cinco últimas hipóteses, consegue-se demonstrar que a matriz de variância e covariância das estimativas MQO é representada por uma matriz diagonal, com termos não nulos constantes em sua diagonal principal.

Segundo Dantas (2003), a não-tendenciosidade significa que a média de todas as possíveis médias de amostras extraídas do mercado coincide com o verdadeiro valor de mercado; a eficiência está associada à dispersão destas possíveis médias estimadas em torno da verdadeira média, sendo que na comparação entre diversos estimadores não-tendenciosos, o estimador eficiente é aquele que apresentar a menor variância; enquanto que a propriedade da consistência indica que na medida em que a amostra cresce, a sua média se aproxima do verdadeiro valor de mercado.

As hipóteses de Gauss-Markov tratam-se de um subconjunto das hipóteses do MCRL, o qual ainda conta com uma sexta hipótese, a saber (ALMEIDA, 2012):

6) Os termos do erro aleatório seguem uma distribuição normal com média zero e variância constante: $\varepsilon \approx \text{Normal}(0, \sigma^2 I_n)$. Assumir esta hipótese pressupõe o atendimento das hipóteses 4 e 5.

2.12 DIAGNÓSTICOS NOS MCRL

A partir da obtenção do modelo de regressão ajustado, é necessário verificar se não ocorreram problemas em seu ajuste.

Três tipos problemas podem afetar tais modelos: presença de *outliers*,

observações que possuem um grande resíduo em relação às demais observações; violação dos pressupostos para os erros ou para as estruturas das médias; e a ocorrência de observações que se retiradas da análise provocam variações consideráveis seja no conjunto de valores ajustados seja nos parâmetros estimados, conhecidas por observações influentes.

Diagnóstico em modelos de regressão se refere ao conjunto de técnicas gráficas e análise dos resíduos utilizadas para verificar os pressupostos de adequação e validade do modelo.

2.12.1 Matriz Chapéu ou *Hat Matrix*

Em modelos lineares ajustados por mínimos quadrados, é muitas vezes útil determinar o grau de influência ou de alavancagem que cada valor de um valor observado y_j pode ter sobre cada valor ajustado \hat{y}_i . A relação entre o valor observado y_i e o seu valor ajustado \hat{y}_i é particularmente simples de interpretar, e pode revelar valores extremos multivariados entre as variáveis independentes que poderiam ser difíceis de detectar. A informação desejada está disponível na matriz chapéu, a qual mostra cada valor ajustado \hat{y}_i como uma combinação linear dos valores observados y_j (HOAGLIN e WELSCH, 1978).

Se $y = X\beta + \varepsilon$, onde X é $n \times (k+1)$, de posto $k+1 < n$, então o valor de $\hat{\beta} = [\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_k]'$ que minimiza a soma do quadrado dos resíduos é igual a $\hat{\beta} = (X'X)^{-1} X'y$, chamado de estimador de mínimos quadrados. O vetor de valores preditos, $\hat{y} = X\hat{\beta}$, pode ser descrito como:

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1} X'y \quad (20)$$

$$\hat{y} = Hy \quad (21)$$

Onde $H = X(X'X)^{-1} X'$, $m \times n$, é chamada *hat matrix* ou a matriz chapéu

em y porque ela transforma y (o vetor de respostas) em \hat{y} (vetor de valores ajustados ou seja, coloca o chapéu em y). Trata-se de uma matriz simétrica e idempotente ou seja, a partir da multiplicação de H por X obtém-se X :

$$HX = X(X'X)^{-1}X'X = X \quad (22)$$

Os elementos da diagonal principal desta matriz, designados de h_{ii} (efeito alavanca ou *leverage*) medem o quão distante a observação y_i está das demais $n-1$ observações no espaço definido pelas variáveis explicativas do modelo. Vale ressaltar que h_{ii} dependem apenas dos valores das variáveis explicativas relacionadas à matriz X , ou seja, não possui relação com os elementos do vetor das variáveis explicadas y .

A ocorrência deste tipo de situação está associada à existência de observações cujos valores das variáveis explicativas se distanciam dos valores das demais observações que compõem a amostra. Estas observações tem a capacidade de atrair as superfícies ajustadas para próximo de si funcionando como o ponto de apoio de uma alavanca.

Portanto, o elemento h_{ii} representa uma medida de alavanca da i -ésima observação, então se h_{ii} for grande o valor da variável explicativa associado a i -ésima observação será atípico, ou seja, estará distante do valor médio da variável explicativa, o que poderá ter influência no cálculo dos coeficientes da regressão. Ressalta-se entretanto que um ponto de alavanca não necessariamente será um *outlier*.

2.12.2 Análise dos Resíduos

A análise dos resíduos está entre as técnicas utilizadas para realização de diagnóstico de problemas nos modelos. Os erros verdadeiros ε_i são desconhecidos num modelo de regressão $\varepsilon_i = Y_i - E(Y_i)$, e atendem ao pressuposto de que

$$\varepsilon \sim Normal(0, \sigma^2).$$

No diagnóstico são utilizados os resíduos ou vetor de resíduos $e_i = Y_i - \hat{Y}_i$

que podem ser considerados os erros observados. Os resíduos apresentam as seguintes propriedades:

$$\text{Média: } \bar{e} = \frac{\sum e_i}{n} = 0 \quad (23)$$

$$\text{Variância: } \frac{\sum (e_i - \bar{e})^2}{n-2} = \frac{\sum e_i^2}{n-2} = \frac{SQ_{Res}}{n-2} = QM_{Res} \quad (24)$$

Caso o modelo esteja adequado, o Quadrado Médio dos Resíduos (QM_{Res}) é um estimador não tendencioso da variância do erro (σ^2).

O vetor de resíduos também pode ser expresso da seguinte forma:

$$e = Y - \hat{Y} = Y - HY = (I - H)Y \quad (25)$$

Pode-se mostrar que $\text{Var}(e) = (I - H)\sigma^2$. Assim, $\text{Var}(e_i) = \sigma^2(1 - h_{ii})$, ou seja, $\text{Var}(e_i)$ é dada pelo i -ésimo termo da diagonal de $(I - H)\sigma^2$. Pode-se estimar a variância substituindo-se σ^2 pelo QM_{Res} . Os termos fora da diagonal de $(I - H)\sigma^2$ correspondem às covariâncias dos resíduos.

Diferentemente dos erros, os resíduos não são independentes e em virtude disso não podem ser utilizados em testes formais baseados em amostras aleatórias com observações independentes. Além disso, os resíduos também não são identicamente distribuídos (uma vez que as suas variâncias não são iguais), e dependem de σ^2 , a variância desconhecida dos erros aleatórios. Em virtude destes fatos, é comum serem utilizados no diagnóstico dos resíduos as suas formas padronizadas.

Normalizando o resíduo pela distribuição normal obtém-se:

$$\frac{E_i}{\sqrt{\sigma^2(1 - h_{ii})}} \cap N(0,1) \quad (26)$$

Substituindo a variância desconhecida σ^2 pelo seu estimador QM_{Res} , são obtidos aquilo que se designam resíduos padronizados ou resíduos internamente padronizados:

$$R_i = \frac{E_i}{\sqrt{QM_{Res}(1-h_{ii})}} \quad (27)$$

Caso o modelo de regressão esteja correto, todos os resíduos terão a mesma variância e serão adequados para a verificação da normalidade e homocedasticidade dos erros. Observações com valores absolutos maiores que 2 para os resíduos padronizados podem ser considerados pontos discrepantes ou mal ajustados. Os resíduos padronizados não apresentam distribuição t-Student pois o resíduo de cada observação não é independente da variância estimada. Para superar esta dificuldade pode-se substituir a variância estimada pelo erro quadrático correspondente ao modelo sem a i-ésima observação.

Os resíduos externamente Studentizados, ou resíduos Studentizados podem ser obtidos a partir de:

$$T_i = \frac{Y_i - \hat{Y}_i}{\sqrt{QM_{Res(-i)}(1-h_{ii})}} \quad (28)$$

O índice i indica que a i-ésima observação foi excluída. Estes resíduos são utilizados para verificar se existem diferenças significativas entre os valores observados com e sem a i-ésima observação.

2.12.3 Distância de Cook

A estatística de Cook pode ser definida como uma medida para avaliar qual é o nível de influência isolada de uma observação suspeita sobre as estimativas de uma regressão. Em termos computacionais, a Distância de Cook mede a distância entre os coeficientes estimados com e sem a observação suspeita (ALMEIDA, 2012)

Formalmente, distância de Cook é uma medida de distância quadrática entre o estimador de mínimos quadrados $\hat{\beta}$ obtido com base nos n elementos amostrais e o estimador de mínimos quadrados $\hat{\beta}_{(i)}$, calculado sem considerar os resultados verificados para o i-ésimo elemento (TACONELI, 2010):

$$D_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' X' X (\hat{\beta}_{(i)} - \hat{\beta})}{p QM_{Res}} \quad (29)$$

Pontos com valores elevados para D_i influenciam fortemente a estimação por mínimos quadrados de $\hat{\beta}$. Usualmente considera-se pontos para os quais $D_i > 1$ como influentes. Uma forma equivalente de D_i é apresentada na sequência:

$$D_i = \left(\frac{e_i}{\sqrt{QM_{Res} h_{ii}}} \right)^2 \left(\frac{h_{ii}}{1 - h_{ii}} \right) \frac{1}{p}, \quad (30)$$

Nas equações acima:

e_i é o resíduo bruto (a diferença entre o valor observado e o valor ajustado pelo modelo proposto);

QM_{Res} é o erro médio quadrático do modelo de regressão;

h_{ii} é o i -ésimo elemento da diagonal da matriz de projeção $X(X'X)^{-1}X'$;

p é o número de parâmetros ajustados no modelo.

2.13 CONSEQUÊNCIAS DOS EFEITOS ESPACIAIS SOBRE OS PRESSUPOSTOS DO MCRL

Os fenômenos que acontecem no espaço estão sujeitos ao princípio anunciado por Tobler (1970) por meio da Primeira Lei da Geografia: *“tudo está relacionado a tudo mais, mas as coisas mais próximas estão mais relacionadas entre si do que as coisas mais distantes”*. Neste sentido, dados distribuídos no espaço tendem a apresentar dois efeitos espaciais na sua estrutura conhecidos como dependência espacial e heterogeneidade espacial.

A dependência ou autocorrelação espacial, de acordo com Anselin e Bera, (1998) manifesta-se quando ocorre uma coincidência entre valores semelhantes ocorrendo em locais semelhantes. Ela pode ocorrer de duas diferentes formas: quando há o agrupamento de altos ou baixos valores ela é denominada de autocorrelação positiva, e quando ocorre de um valor alto (baixo) ser rodeado por valores baixos (altos), a autocorrelação é denominada como negativa.

A autocorrelação espacial é sinônimo da dependência espacial e pode afetar tanto a variável dependente, quanto os erros. Formalmente, a autocorrelação espacial pode ser representada pela seguinte condição de momento:

$$Cov(y_i) = E(y_i y_j) - E(y_i) \cdot E(y_j) \neq 0, \text{ para } i \neq j \quad (31)$$

Em que y_i e y_j são observações de uma variável aleatória nas localizações i e j respectivamente, i e j podem ser pontos, tais como localização de estabelecimentos ou áreas metropolitanas – medidas em latitudes e longitudes – ou unidades de área, tal como estados, municípios ou setores censitários (ANSELIN e BERA, 1998).

O segundo efeito é específico da localização espacial, devido a isso é chamado de heterogeneidade espacial. Seu significado é de que o comportamento econômico não é estável através do espaço, podendo gerar padrões espaciais característicos sob a forma de agrupamentos ao longo do conjunto de dados, e variar com a unidade (VIEIRA, 2009). Isto representaria, econometricamente, a necessidade de parâmetros que variam entre as diferentes localidades, ou parâmetros aleatórios, ou, ainda, outras formas de mudança estrutural. Essas heterogeneidades podem ser vistas como erros de mensuração oriundos de variáveis ignoradas, ou erros de especificação do modelo, que geram heterocedasticidade e até mesmo erros de especificação (FIGUEIREDO, TEIXEIRA e BONJOUR, 2009).

A dependência espacial e a heterogeneidade espacial podem ocorrer de forma imbricada, o que torna mais difícil a tarefa de realizar a especificação dos modelos econométricos espaciais (ALMEIDA, 2012). Na presença de efeitos espaciais, as consequências sobre os pressupostos do MCRL são apresentados a seguir (ALMEIDA, 2012):

1. A hipótese da linearidade continua válida no MCRL na presença de efeitos espaciais. Entretanto, na presença de heterogeneidade espacial, a especificação do modelo linear às vezes se altera, a fim de permitir que os coeficientes sejam variáveis conforme a localização do fenômeno a ser estudado.

2. Com relação à hipótese de não colinearidade perfeita, esta também continua válida. Entretanto, mesmo que não se observe a colinearidade perfeita, em função da utilização de amostras de tamanho médio ou pequeno, corre-se o risco da

ocorrência do problema da multicolinearidade imperfeita (ocorre quando r_{ij} varia de 0 a 1 ou de 0 a -1, o que é mais frequente).

3. No que diz respeito à hipótese de média condicional zero, esta hipótese pode ser violada de diversas formas, seja pela omissão de relevantes variáveis defasadas espacialmente, seja pelo não tratamento adequado da endogeneidade espacial, e também devido à existência de erros de medida nas defasagens espaciais. Estes problemas acabam por viesar o estimador MQO.

4. A hipótese da homocedasticidade também é afetada pela presença da dependência espacial. Entretanto, pode ser ampliado pela imbricação que pode ocorrer entre a dependência espacial e heterogeneidade espacial, provocando a manifestação da heterocedasticidade.

5. Com relação à hipótese de independência dos erros, quando se trabalha com dados em corte transversal que são unidades geográficas, a amostragem aleatória, a qual garante esta independência pode ser perdida resultando que o erro de uma região i pode estar correlacionado com o erro de outra região j , conseqüentemente o estimador MQO torna-se ineficiente, e além disso esta dependência pode acarretar viés e inconsistência do estimador MQO.

6. Finalmente a hipótese de que os erros seguem uma distribuição normal com média zero e variância constante, também será violada caso as hipóteses de média condicional zero e independência dos erro sejam violadas. A violação desta hipótese implicará que os erros-padrão das estimativas bem como os testes t e F percam a sua validade.

De acordo com Florencio (2012), a solução para estimação segura da equação de preços hedônicos na presença de efeitos espaciais, deve ser feita com base em modelos espaciais, levando em consideração todas as interações espaciais possíveis entre os dados e servindo como uma proxy para variáveis locais não consideradas explicitamente no modelo.

Pode-se perceber nos últimos anos, que diversos trabalhos tem estudado a combinação de metodologias com o objetivo de incorporar os efeitos espaciais, especificamente a autocorreção espacial (MICHAEL, 2004; TRIVELLONI, 2005; MACHADO, 2006; HORNBERG, 2009; DALAQUA, AMORIN e FLORES, 2010; DANTAS, MAGALHÃES e VERGOLINO, 2007; HORNBERG, 2009; DANTAS,

MAGALHÃES e VERGOLINO 2010; FERMO, 2010; FLORENCIO, 2012). Entretanto, com exceção do trabalho de Machado (2006), todos os trabalhos citados foram desenvolvidos em áreas urbanas, o que revela a necessidade de avaliar como a combinação de metodologias também pode vir a obter modelos de valoração mais eficientes também em áreas rurais.

2.14 MATRIZES DE PONDERAÇÃO ESPACIAL

Considerando o modelo clássico de regressão linear:

$$Y = \beta X + \varepsilon \quad (32)$$

A matriz de variância-covariância dos erros, $cov[\varepsilon\varepsilon']$, expressa a covariância espacial quando os elementos fora da diagonal principal são diferentes de zero de acordo com uma dada estrutura espacial ou ordenamento espacial, o qual especifica aqueles pares de localizações $i-j$ (com $i \neq j$) cujas covariâncias não são zero, ou $E[\varepsilon_i\varepsilon_j] \neq 0$ (ANSELIN, 2003).

A estrutura espacial pode ser obtida de duas formas, pelo método direto (abordagem da geoestatística) ou pelo método indireto (utilização de matrizes de ponderação espacial). O método direto consiste em especificar a covariância como uma função da distância que separa dois pares de localização quaisquer, para isto, utiliza-se uma função decrescente para a distância e um parâmetro espacial que assegure a definição de uma matriz de variância-covariância positiva (ANSELIN, 2003). Alternativamente, o método indireto requer a especificação de um processo estocástico espacial que relaciona o valor de uma variável aleatória num local, com os valores desta variável nas localidades vizinhas. Em vez de ligar todos os pares de observação por meio de uma função de decaimento distância, os vizinhos para cada local individual são especificados por meio de uma chamada matriz de ponderação espacial (ANSELIN, 2003).

A matriz de ponderação espacial é uma matriz quadrada de dimensão n por n , na qual os pesos espaciais w_{ij} representam o grau de conexão entre as regiões segundo algum critério de proximidade, mostrando a influência da região j sobre a região i . Este grau de conexão pode ser classificado de acordo com critérios geográficos ou socioeconômicos. O objetivo é que essa matriz consiga capturar

grande parte da autocorrelação espacial contida no fenômeno estudado (ALMEIDA, 2012).

No que se refere aos critérios geográficos, pode-se utilizar dois critérios: a contiguidade e a distância geográfica. A contiguidade baseia-se na ideia de que duas regiões que possuam uma fronteira física comum possuem maior interação espacial, sendo assim, formalmente: $w_{ij}=1$ se i e j são contíguos ou $w_{ij}=0$ se i e j não são contíguos. A distância geográfica baseia-se na ideia de que regiões próximas geograficamente possuem maior interação espacial. Utilizando este critério, uma das matrizes mais conhecidas é a matriz de k vizinhos mais próximos $w_{ij}(k)$, na qual k é a distância geográfica medida em quilômetros, sendo assim: $w_{ij}(k)=1$ se $d_{ij} \leq d_i(k)$ ou $w_{ij}(k)=0$ se $d_{ij} > d_i(k)$ (ALMEIDA, 2012).

Diante das diversas possibilidades para criação das matrizes de ponderação espacial, é comum surgir a dúvida a respeito de qual seria a melhor matriz para capturar a autocorrelação existente nos dados de uma determinada região. O conhecimento a respeito do fenômeno que está sendo estudado pode ajudar, sendo essencial a realização de testes com diversas matrizes possíveis.

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_ξ e W_ε , essas variáveis serão medidas com erro sistemático, ocasionando viés na estimação, tornando-a inconsistente (ALMEIDA, 2012).

Baumont (2004) propõe o seguinte procedimento para identificar a melhor matriz de ponderação espacial: a partir de um modelo de regressão clássico o vetor de resíduos é extraído; em seguida constrói-se diversas matrizes de ponderação espacial a partir de diferentes critérios; o próximo passo é realizar o teste I de Moran Global para cada uma das matrizes criadas, sendo que a matriz a ser escolhida será aquela que apresentar o maior valor para o I de Moran, desde que este seja significativo.

A matriz w pode ser normalizada, o que consiste na divisão de cada elemento da matriz pelo total de suas linhas, com isso, a soma de cada linha da matriz resultará 1. Embora este procedimento provoque assimetria da matriz, significando

que $w_{ij} \neq w_{ji}$, duas vantagens justificam a sua utilização: a primeira é que o somatório das suas linhas e colunas resultará no tamanho da amostra e a segunda é que a matriz normalizada permite a interpretação de média dos valores da variável nos vizinhos para a defasagem espacial (ALMEIDA, 2012).

2.15 ESPECIFICAÇÃO DA DEPENDÊNCIA ESPACIAL

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 (Wy); como defasagem nas variáveis explicativas (Wx); ou então como defasagem no termo do erro ($W\xi$ e/ou $W\varepsilon$). 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 de econométrico espacial (VIEIRA, 2009; ALMEIDA, 2012).

Existe um conjunto de testes que auxiliam na escolha das defasagens espaciais a serem colocadas na regressão. Almeida (2012) afirma que os testes podem ser divididos em duas categorias: testes difusos e testes focados. Os testes difusos são aqueles cuja hipótese nula verifica se os resíduos são independentes espacialmente contra a hipótese alternativa de que os resíduos são dependentes espacialmente. Entretanto, a hipótese alternativa dos testes difusos não se refere a um modelo econométrico espacial específico. Os Testes focados se caracterizam por empreenderem uma especificação explícita do processo estocástico gerador do erro. Com isso, estes últimos, conseguem fornecer uma indicação na hipótese alternativa do tipo predominante da autocorrelação espacial. Como exemplo de testes difusos temos o Teste I de Moran e o Teste de Kalejian-Robinson (KR). Já como exemplos de testes focados tem-se os baseados nos Multiplicadores de Lagrange (ML), ML para defasagem, ML para o erro, ML robusto para defasagem, ML robusto para o erro.

Dantas (2003) alerta que a validade destes testes exige a aceitação das hipóteses de normalidade e homocedasticidade dos resíduos de MQO, obtidos pelo modelo.

Em função do maior número de detalhes nas informações fornecidas pelos testes focados serão apresentados de forma resumida o Teste ML para defasagem, e o Teste ML para o erro.

2.15.1 Teste ML (defasagem)

O objetivo deste teste é detectar a defasagem espacial da variável dependente, sendo baseado no Multiplicador de Lagrange. Este teste segue a distribuição qui-quadrado com um grau de liberdade. Por se tratar de um teste assintótico, a estatística ML (defasagem) é mais apropriada para grandes amostras (ALMEIDA, 2012). Apresenta relativa facilidade computacional devido ao fato de que, sob a hipótese nula, pode ser calculado com base nos resíduos da regressão estimada por MQO e pode ser obtido por:

$$ML(\text{defasagem}) = \frac{\left(\frac{e'Wy}{s^2} \right)}{\left\{ \frac{(WX\hat{\beta})'MWX\hat{\beta}}{s^2} + tr[W'W + W^2] \right\}} \quad (33)$$

Onde:

e = vetor de resíduos de mínimos quadrados;

W = matriz de pesos espaciais;

y = vetor de observações na variável dependente;

$s^2 = (e'e)/n$ = estimativa de máxima verossimilhança da variância do modelo $Y = X\beta + \varepsilon$;

X = matriz das variáveis independentes;

$\hat{\beta}$ = vetor de parâmetros estimados via mínimos quadrados ordinários;

e' = transposta de e ;

$M = I - X(X'X)^{-1}X'$;

tr = operador denominado traço da matriz.

Sua hipótese nula é de que o parâmetro do erro autorregressivo espacial seja nulo. Portanto, caso a estatística do teste, seja superior ao ponto crítico da distribuição qui-quadrado com um grau de liberdade, para um determinado nível de significância α , rejeita-se a hipótese de não autocorrelação espacial nos resíduos do modelo clássico de regressão a este nível.

2.15.2 Teste ML (erro)

Trata-se de um teste do tipo multiplicador de Lagrange para detectar a defasagem espacial nos termos do erro. Segue uma distribuição qui-quadrado com um grau de liberdade. Para sua implementação é necessário apenas os resíduos do modelo clássico de regressão estimado por MQO o que representa uma vantagem na sua utilização (ALMEIDA, 2012). A fórmula do teste é dada por:

$$ML(erro) = \frac{\left\{ \frac{e' W e}{s^2} \right\}^2}{tr[W' W + W^2]} \quad (34)$$

onde:

e = vetor de resíduos de mínimos quadrados;

e' = transposta de e;

W = matriz de pesos espaciais;

s² = (e'e)/n = estimativa de máxima verossimilhança da variância do modelo Y=Xβ+ ε;

n=número de dados da amostra;

tr = operador denominado traço da matriz;

Sua hipótese nula é de que não existe autocorrelação espacial no termo do erro. Portanto, caso a estatística do teste seja superior ao ponto crítico da distribuição qui-quadrado com um grau de liberdade, para um determinado nível de significância α, rejeita-se a hipótese de não autocorrelação espacial nos resíduos do modelo clássico de regressão a este nível.

2.15.3 Especificação de Modelos Espaciais

Almeida (2012) descreve o procedimento conhecido como híbrido para 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, utilizando a combinação dos testes de ML tradicionais e os ML nas suas versões robustas.

O procedimento segue os seguintes passos:

1. Estima-se o modelo clássico de regressão linear (MCRL) por MQO;

2. 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);

3. 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;

4. 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;

5. 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 como o mais apropriado.

2.16 MODELOS ESPACIAIS

O modelo clássico de regressão representa um processo no qual não se leva em consideração a influência dos efeitos espaciais e é dado por:

$$Y = \beta X + \varepsilon, \varepsilon \sim Normal(0, \sigma^2 I_n) \quad (35)$$

em que Y é um vetor n por 1 de observações sobre a variável dependente, X é uma matriz n por k de observações sobre as variáveis explicativas exógenas (mais a constante) com um vetor associado k por 1 de coeficientes de regressão β e ε é um vetor n por 1 de termos de erro aleatório, com média zero e variância constante.

A representação esquemática da Figura 2 mostra que entre as regiões vizinhas i e j existe uma relação de contiguidade, entretanto, não existe interação espacial entre as regiões. Devido a este fato a Equação 35 não incorpora nenhuma defasagem espacial.

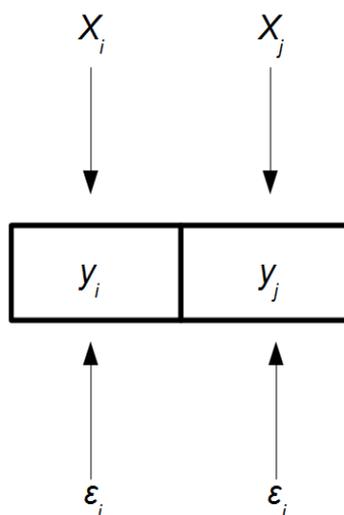


Figura 2 - Processo a-espacial. Fonte: Almeida, (2012).

O modelo econométrico-espacial envolve a incorporação de defasagens espaciais ao modelo expresso em 35 com o objetivo de se controlar a dependência espacial.

As defasagens espaciais podem assumir a forma de defasagem na variável dependente (Wy), defasagens nas variáveis independentes (Wx) e/ou defasagem nos termos de erro ($W\xi$ ou $W\varepsilon$). A seguir serão mostrados os modelos que incorporam as variáveis defasadas espacialmente referentes à variável dependente e ao termo do erro.

2.16.1 Modelo Espacial Autorregressivo ou *Spatial Autoregressive 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).

A Figura 3 mostra que além da relação de contiguidade entre as regiões i e j , estas também apresentam interação espacial por meio da variável dependente.

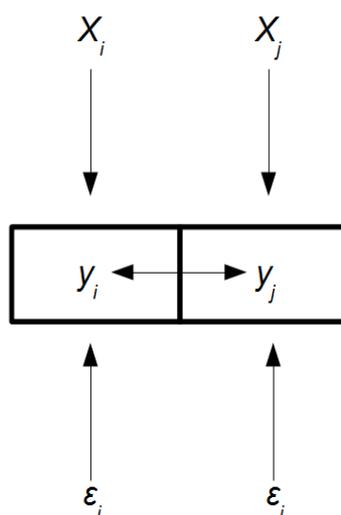


Figura 3 - Processo de defasagem espacial. Fonte: Almeida (2012)

O modelo SAR (*Spatial Autoregressive*) puro pode ser expresso do seguinte modo:

$$y = \rho W y + \varepsilon \quad (36)$$

no qual $W y$ 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). Caso este não seja estatisticamente significativo, significa que o coeficiente é zero, indicando a inexistência de autocorrelação espacial.

Quando o mesmo inclui o conjunto de variáveis explicativas exógenas X , obtém-se o modelo misto do modelo SAR:

$$y = \rho W y + X \beta + \varepsilon \quad (37)$$

no qual X é uma matriz de variáveis exógenas, permanecendo os demais termos como mostrado anteriormente.

O modelo SAR misto é especificado para que o valor da variável dependente observado numa determinada região seja determinado pela média dos valores da variável dependente observados na vizinhança ($W y$), pelos valores das variáveis explicativas exógenas (X) e ainda sejam influenciados aleatoriamente por um termo de erro (ε).

Na sua forma reduzida o modelo SAR é representado por:

$$y = (I_n - \rho W)^{-1} X \beta + (I_n - \rho W)^{-1} \varepsilon, \varepsilon \sim \text{Normal}(0, \sigma^2 I) \quad (38)$$

Sob o pressuposto da normalidade conjunta dos termos do erro, a função log-verossimilhança que permite a estimação do modelo de defasagem espacial assume a seguinte forma:

$$\ln L = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) + \frac{1}{2} \ln |I_n - \rho W| - \frac{(y - \rho W y - X \beta)' (y - \rho W y - X \beta)}{2\sigma^2} \quad (39)$$

2.16.2 Modelo Espacial do Erro ou *Spatial Error Model SEM*

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).

Observando a Figura 4 o erro da região i está correlacionado com o erro na região j .

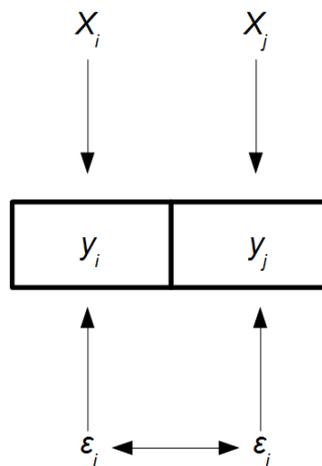


Figura 4 - Processo de erro autorregressivo espacial. Fonte: Almeida, (2012).

A dependência espacial do modelo SEM manifesta-se no termo do erro ao invés da variável dependente.

A expressão formal do modelo SEM, contendo o erro espacial autorregressivo de primeira ordem é representada por:

$$y = X\beta + \xi \quad (40)$$

Onde:

$$\xi = \lambda W \xi + \varepsilon \quad \text{com } \varepsilon \sim \text{Normal}(0, \sigma^2 I_n) \quad (41)$$

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

A forma reduzida do modelo pode ser representada por:

$$\xi = X\beta + (I_n - \lambda W)^{-1} \varepsilon \quad (42)$$

Sob o pressuposto da normalidade, a função log-verossimilhança para o modelo do erro autorregressivo espacial assume a seguinte forma:

$$\ln L = -\frac{n}{2} \ln \pi - \frac{n}{2} \ln \sigma^2 + \ln |I_n - \lambda W| - \frac{(y - X\beta)' (I_n - \lambda W)' (I_n - \lambda W) (y - X\beta)}{2\sigma^2} \quad (43)$$

2.17 GEOESTATÍSTICA

De acordo com Matheron (1971), quando um fenômeno se espalha no espaço e apresenta uma determinada estrutura espacial, diz-se que é regionalizado, sendo que a Geoestatística é a aplicação da teoria das variáveis regionalizadas para a estimativa de fenômenos de diversas naturezas.

Conforme explica Landim (2006) a *“Geoestatística proporciona a estimativa das variáveis regionalizadas usando informações e relações a partir de um conjunto discreto de amostras, com a avaliação dos erros de estimativa, para estabelecer o grau de segurança em previsões e os padrões ótimos de amostragem, que assegurem que um erro máximo de estimativa não seja excedido.”*

No estudo do comportamento das variáveis regionalizadas, duas são as ferramentas principais dos métodos geoestatísticos: o semivariograma e a krigagem.

2.18 SEMIVARIOGRAMA

Para Matheron (1971) as variáveis regionalizadas mostram dois aspectos contraditórios: um aspecto aleatório (marcada irregularidade e com variações imprevisíveis de um ponto a outro); um aspecto estruturado (deve refletir em certa medida, as características estruturais do fenômeno regionalizado).

Camargo, Fucks e Câmara (2004) explicam que a variação espacial de uma variável regionalizada 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.

De acordo com Dantas (2005) a função que estabelece a relação entre os preços e as características influenciadoras sobre os imóveis tem caráter aleatório (para um mesmo elemento do domínio podem estar associados vários elementos da imagem); desta forma, o valor de mercado pode ser considerado como uma variável aleatória, cuja estimação pontual é feita pela média dos preços nele praticados.

De acordo com Camargo, Fucks e Câmara (2004) são dois os principais efeitos que definem a estrutura espacial dos dados: Efeito de 1ª ordem é o valor esperado, ou seja, a média do processo no espaço; Efeito de segunda ordem é a covariância entre as áreas S_i e S_j . Diz-se que um fenômeno é estacionário quando os efeitos de primeira e segunda ordem são constantes em toda a área de estudo, ou seja, não há tendência.

De acordo com Yamamoto e Landim (2013) na Geoestatística a covariância mede a relação entre valores da mesma variável, obtidos em pontos separados por uma distância h , conforme uma determinada direção. Quando a alteração da direção não provoca alteração da covariância diz-se que o fenômeno é isotrópico.

A covariância de uma variável regionalizada para pontos separados por uma distância h pode ser calculada como:

$$C(h) = E \{ [Z(x+h) - m][Z(x) - m] \} \quad (44)$$

na qual h representa um vetor entre dois pontos x_1 e x_2 no espaço tridimensional. Caso $h=0$ a covariância torna-se igual à variância da variável

regionalizada.

De acordo com Yamamoto e Landim (2013) existem fenômenos nos quais embora não seja possível assumir a estacionaridade de 2ª ordem, é possível determinar um variograma. Nestes casos adota-se a hipótese intrínseca da geoestatística.

Afirmam também que a hipótese intrínseca não requer a existência de uma média constante e variância finita para a função aleatória $Z(x)$, mas apenas que os incrementos da função aleatória $[Z(x+h)-Z(x)]$ sejam estacionários de 2ª ordem, isto possibilita realizar a inferência estatística (YAMAMOTO E LANDIM, 2013).

A função variograma é definida como a variância do incremento $[Z(x+h)-Z(x)]$:

$$\gamma(h) = \frac{1}{2} E \{ [Z(x+h) - Z(x)]^2 \} \quad (45)$$

ou ainda:

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^n [Z(x+h) - Z(x)]^2 \quad (46)$$

O semivariograma mostra a medida do grau de dependência espacial entre amostras ao longo de um suporte específico e, para sua construção, são usados simplesmente as diferenças ao quadrado dos valores obtidos, assumindo-se uma estacionaridade nos incrementos (LANDIM, 2006).

A partir do semivariograma experimental é realizado o ajuste de um ou mais modelos teóricos, e dentre os principais modelos cita-se o modelo esférico, exponencial e gaussiano. Os parâmetros envolvidos nos modelos teóricos são: alcance (a), patamar (C), efeito pepita (C0) e contribuição (C1). O alcance é a distância a partir da qual as amostras tornam-se independentes, não ocorrendo mais dependência espacial. O efeito pepita é o valor da função na origem ou descontinuidade do semivariograma, para distâncias menores do que a menor distância entre as amostras. A contribuição representa a diferença entre o patamar e o efeito pepita e refere-se ao valor da semivariância em que não ocorrerá mais dependência espacial. O patamar, por sua vez, consiste no valor onde o semivariograma se estabiliza e as amostras se tornam independentes (ISAKS e

SRIVASTAVA, 1989).

Os semivariogramas expressam o comportamento espacial da variável regionalizada e mostram (LANDIM, 2006):

a) o tamanho da zona de influência em torno de uma amostra, pois toda amostra cuja distância ao ponto a ser estimado for menor ou igual ao alcance, fornece informações sobre o ponto;

b) a anisotropia, quando os semivariogramas mostram diferentes comportamentos para diferentes direções de linhas de amostragem e de estudo da variável; neste caso a anisotropia pode ser geométrica quando o alcance varia de acordo com as diversas direções consideradas, mantendo constante o patamar e zonal quando o alcance permanece constante e o patamar varia conforme for modificada a direção;

c) continuidade, pela forma do semivariograma, em que para $h \cong 0$, $\gamma(h)$ já apresenta algum valor. Essa situação é conhecida como efeito pepita (*nugget effect*) e é representada por C_0 . O efeito pepita pode ser atribuído a erros de medição ou ao fato de que os dados não foram coletados a intervalos suficientemente pequenos para mostrar o comportamento espacial subjacente do fenômeno em estudo.

A determinação de um variograma possibilita a realização do procedimento de estimativa geoestatística conhecido como Krigagem.

2.18.1 Krigagem

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.

Marques et al. (2012) afirmam que a utilização de modelos de regressão múltipla na elaboração de PVG resulta na criação de superfícies discretas. Entretanto, para que se possa mapear as áreas de transição de valores seria ideal realizar estimação de superfícies de ajuste obtidas por interpolação, com destaque

para o interpolador krigagem, o qual permite a partir de dados irregularmente espaçados, gerar representações temáticas isarítmicas.

O estimador da krigagem ordinária é:

$$Z^*_{ko}(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (47)$$

Os pesos ótimos são calculados sob duas condições de restrição (YAMAMOTO e LANDIM, 2013): a) que o estimador não seja viesado; e b) que a variância de estimativa seja mínima.

O não viés da estimativa é obtido quando o erro, diferença entre o valor real e o valor calculado, é igual a zero, em média:

$$E[Z^*_{ko}(x_0) - Z(x_0)] = 0 \quad (48)$$

Desenvolvendo a expressão da esperança do erro, chega-se à condição de não viés:

$$\sum_{i=1}^n \lambda_i = 1 \quad (49)$$

A variância de estimativa ou a variância do erro de estimativa é calculada como:

$$\sigma_E^2 = Var[Z(x_0) - Z^*_{ko}(x_0)] \quad (50)$$

As equações 48 e 50 envolvem uma grandeza desconhecida, o valor $Z(x_0)$, que é o valor real de um ponto não amostrado. A solução para esse problema é baseado em um modelo probabilístico, de tal forma que os valores desconhecidos são considerados realizações de um processo aleatório, assim como são os valores da variável aleatória $\{Z(x_i), 1=1, n\}$ (YAMAMOTO e LANDIM, 2013).

A minimização da variância do erro de estimativa parte do desenvolvimento da Eq. 48, conforme:

$$\sigma_E^2 = E[(Z(x_0) - Z^*_{ko}(x_0))^2] - (E[Z(x_0) - Z^*_{ko}(x_0)])^2 \quad (51)$$

Expandindo-se cada termo do lado direito dessa equação, chega-se à seguinte expressão:

$$\sigma_E^2 = C(0) - 2 \sum_i \lambda_i C(x_0 - x_i) + \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) \quad (52)$$

Essa é a expressão da variância de estimativa, em termos da função covariância que é conhecida. Para encontrar os pesos ótimos, deve-se minimizar a variância do erro de estimativa sob a condição de não viés ou de restrição (Eq. 48). Para encontrar o ponto de mínimo, utiliza-se a técnica dos multiplicadores de Lagrange, da qual se obtém a lagrangiana, isto é, a função das coordenadas generalizadas (YAMAMOTO e LANDIM, 2013):

$$L(\lambda_1, \lambda_2, \dots, \lambda_n, \mu) = C(0) - 2 \sum_i \lambda_i C(x_0 - x_i) + \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) - 2\mu \left(\sum_j \lambda_j - 1 \right) \quad (53)$$

na qual $L(\lambda_1, \lambda_2, \dots, \lambda_n, \mu)$ é a Lagrangiana e μ é o multiplicador de Lagrange.

Fazendo cada uma das derivadas parciais da lagrangiana iguais a zero e também derivando a lagrangiana em relação a μ , chega-se ao sistema de equações de Krigagem ordinária (YAMAMOTO e LANDIM, 2013).

$$\begin{cases} \sum_{j=1}^n \lambda_j C(x_i - x_j) - \mu = C(x_i - x_0) \text{ para } i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (54)$$

A variância de krigagem em termos da função covariância, é igual a:

$$\sigma_{KO}^2 = C(0) - \sum_{i=1}^n \lambda_i + C(x_0 - x_i) + \mu \quad (55)$$

Marques et al. (2012) afirmam que a krigagem pode proporcionar mais proximidade às situações reais enfrentadas pelo CTM. Isto se explica pelo fato de que as PVG elaboradas por tal interpolador permitem que os valores das parcelas possam ser descritos em função das características da região, bem como definidas as zonas de transição dentro de uma mesma parcela. Estes autores avaliaram de forma positiva a combinação da Krigagem ordinária com métodos de classificação

para geração de PVG e recomendaram a realização de novos testes, considerando-se novas áreas de estudo e um número maior de elementos amostrais, com o objetivo de realizar a validação desta técnica.

2.19 AVALIAÇÃO DE DESEMPENHO

2.19.1 Medidas de Qualidade do Ajuste

O ajuste de modelos de regressão tem como indicadores da qualidade deste ajuste dois parâmetros conhecidos como R^2 e R^2 ajustado, entretanto, na presença de erros autocorrelacionados e/ou heterocedásticos estes perdem a sua utilidade no cumprimento desta função.

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 e SC utilizam em suas fórmulas de cálculo, o valor da função de verossimilhança:

$$AIC = -2LIK + 2k \quad (56)$$

$$SC = -2LIK + k \cdot \ln(n) \quad (57)$$

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.

2.19.2 Avaliação de Desempenho de Avaliações em Massa

Após a determinação de um modelo é preciso avaliar a habilidade deste em estimar valores para um conjunto de dados não empregados na sua construção, este processo pode ser chamado de análise de desempenho e de acordo com

Cesare (2007) quanto menor a variabilidade das estimativas em relação ao preço de venda dos imóveis, maior é a precisão do trabalho desenvolvido. A partir dos resultados obtidos nesta análise, pode-se realizar ajustes globais ou parciais nos modelos.

Pode-se definir qualidade como o “grau em que um conjunto de características inerentes atendem a requisitos” (necessidades ou expectativas estabelecidas, geralmente implícitas ou obrigatórias) (ISO 9000:2000). Para aplicações envolvendo avaliações, no Brasil, os requisitos são estabelecidos pela NBR 14653-1 e 3 (ABNT, 2001; ABNT, 2004), e quando se trata de avaliação em massa de imóveis, a avaliação da qualidade de tais modelos se baseia na análise dos parâmetros estabelecidos pela *International Association of Assessing Officers* (IAAO) por meio da *Standard on Ratio Studies*.

Marques et al. (2012) e Averbeck (2003) enfatizam a importância de realizar a avaliação da representatividade do modelo. Além disso, alertam que nesta avaliação se deve utilizar pontos de verificação que não fizeram parte do conjunto amostral. Dessa forma, pode-se analisar a capacidade do modelo em estimar valores para a variável em análise em locais aleatórios da área de estudo.

De acordo com Hornburg (2009), os modelos de avaliação em massa devem ser avaliados na sua performance para que possam ser considerados de boa qualidade. Nesta avaliação, as medidas de performance se referem a pelo menos dois tipos de qualidade: medidas de performance global da avaliação e medidas relacionadas com a equidade e uniformidade.

As medidas de performance global da avaliação se referem ao quociente ou “ratios” (r) entre os valores avaliados (va) e os preços de venda (pv).

$$r = \frac{va}{pv} \quad (58)$$

Quanto mais próximos estiverem os valores avaliados dos valores de mercado, melhor é a qualidade das avaliações. Tais medidas fornecem informações sobre o grau de alcance dos objetivos ou dos requerimentos legais.

Ao ser realizada a divisão do somatório dos valores avaliados pelo somatório dos preços de venda será obtida a Média Ponderada das Razões.

A partir da obtenção das razões de avaliação, caso sejam identificados valores extremos estes devem ser retirados da análise de acordo com recomendação da IAAO (2013).

A mediana das razões de avaliação é uma medida que é menos afetada pela presença de valores extremos quando comparada com outras medidas de tendência central. Devido a este fato, ela geralmente é a medida preferida para avaliar o nível geral das avaliações, determinar prioridades de reavaliação ou avaliar a necessidade de reavaliação (IAAO, 2013).

A uniformidade das avaliações se refere ao grau em que as propriedades são avaliadas em igual percentual dos valores de mercado (IAAO, 2013). O Coeficiente de Dispersão (CD) é a medida mais útil da variabilidade ou uniformidade e mede o percentual médio em que as razões de avaliação se desviam da mediana das razões de avaliação. O CD é referido como uma medida de dispersão horizontal. Ele proporciona informação percentual sobre a uniformidade da avaliação dos imóveis em toda a área de estudo.

O CD é obtido da seguinte maneira:

- 1) diminui-se de cada razão de avaliação a mediana de todas as razões;
- 2) é tomado o valor absoluto das diferenças anteriores;
- 3) calcula-se a média dos valores absolutos;
- 4) divide-se a média anterior pela mediana das razões de avaliação;
- 5) multiplica-se por 100.

Um outro índice especializado pode ser definido para medir a equidade vertical da avaliação. Ele é chamado de Diferencial Relativo ao Preço ou PRD (*Price Related Differential*). Pode se obtido a partir da divisão da Média das razões de avaliação pela Média ponderada das razões.

É utilizado para detectar diferenças sistemáticas na forma como são avaliados os imóveis de alto e de baixo valor (HORNBERG, 2009). Caso seus valores sejam maiores que 1 é um sinal de regressividade, ou seja, os imóveis de baixo valor estão sendo superavaliados quando comparados com os imóveis de alto valor em termos relativos; caso seus valores sejam menores que 1 é um sinal de progressividade, ou seja, os imóveis de alto valor estão sendo superavaliados quando comparados com os imóveis de baixo valor em termos relativos.

2.20 TRABALHOS RECENTES SOBRE VALORAÇÃO CADASTRAL E PLANTAS DE VALORES GENÉRICOS

Entre os recentes trabalhos envolvendo modelos espaciais destaca-se Florencio (2012) que propõe uma metodologia para a realização de avaliações em massa a partir do uso de modelos espaciais integrados à classe de modelos aditivos generalizados para posição, escala e forma, mais conhecidos como GAMLSS (*Generalized Additive Models for Location, Scale and Shape*). No seu trabalho utilizou dados de terrenos urbanos na cidade de Aracaju-SE, e os resultados mostraram que o uso dos modelos de defasagem espacial incorporados à estrutura de regressão GAMLSS nas avaliações em massa pode fornecer ajustes superiores àqueles obtidos por métodos tradicionais.

Sunak e Madlener (2012) avaliaram o impacto provocado pela implantação de projetos de energia eólica na região da Renânia do Norte (Vestefália, Alemanha) sobre os preços das terras. Utilizaram um modelo de precificação hedônico, em conjunto com efeitos espaciais fixos *viewshed* (análise na qual o resultado é uma região formada por todos os pontos visíveis na paisagem a partir de um dado ponto ou local) e um modelo de Regressão Geograficamente Ponderada para lidar com a heterogeneidade espacial. Seus resultados mostraram que a proximidade, medida pela distância inversa para a turbina eólica mais próxima, provoca impactos negativos significativos sobre os valores de propriedades vizinhas. Tendo em vista que atualmente no Brasil, sobretudo no Nordeste, encontram-se em implantação diversos projetos de geração de energia eólica, avaliar qual seria o impacto destas projetos sobre os preços das terras torna-se importante.

Feichtinger e Salhofer (2013) realizaram uma análise espacial dos preços de terras na Bavária (Região da Alemanha). Utilizaram um conjunto extenso de dados de mercado composto por 7300 negócios realizados dos anos de 2001 e 2007 com o objetivo de identificar os fatores que influenciam os preços das terras agrícolas. Estes autores aplicaram o modelo espacial geral, o qual combina o modelo de defasagem espacial e um modelo de erro espacial, em conjunto com outras variáveis explicativas. Os mesmos verificaram a existência de forte influência de fatores agrícolas, como a produtividade da terra, variáveis que descrevem a estrutura regional do mercado de terras e de fatores não-agrícolas, tais como a

pressão urbana sobre os preços das terras agrícolas. Na região estudada os resultados mostraram que a influência de autoridades públicas como compradores e vendedores provocava aumentos dos preços. Encontraram também influência positiva sobre os preços provocada pelo pagamentos de apoios governamentais aos produtores. Foram observadas fortes relações espaciais no conjunto dos dados e de acordo com os autores, quando os preços dos vizinhos aumentam 1%, provocam o aumento do preço de um área específica na ordem de 0,24%.

Na Tabela 1 são apresentados os principais trabalhos envolvendo imóveis rurais ao longo dos últimos anos, as técnicas envolvidas e os principais resultados alcançados.

Com o objetivo de nortear a definição da metodologia, foi realizada uma compilação dos principais trabalhos que envolveram a elaboração de plantas de valores genéricos ou a geração de superfícies de preços recentemente no Brasil. O resultado está apresentado na Tabela 2.

De forma geral, podemos afirmar que nesta área os trabalhos se concentraram em áreas urbanas, sendo apenas 2 em áreas rurais. Tanto em áreas urbanas quanto nas rurais há pouca disponibilidade de dados de mercado, isto pode ser um reflexo da falta do CTM implantado e atualizado. A exceção é o trabalho de Dantas, Portugal e Prado, (2006) que contou com número de observações superior a 1000. A combinação de técnicas para obtenção de modelos e para a geração de superfícies interpoladas é o mais comum nos trabalhos entretanto, enquanto que nos primeiros trabalhos predominava a utilização da análise de superfície de tendência, os trabalhos mais recentes vem utilizando a krigagem como técnica de interpolação.

Um fato importante é que possivelmente devido à escassez de dados de mercado, poucos trabalhos realizaram a validação dos modelos. Os poucos que fizeram utilizaram os mesmos dados de mercado usados na modelagem. A exceção são os trabalhos de Marques et al. (2012), Malaman et al. (2012) e Malaman et al. (2014) os quais reservaram pontos de validação que não fizeram parte da modelagem e recomendam que trabalhos futuros procedam da mesma forma. Dois trabalhos realizaram a avaliação de desempenho com base na *Standard on Ratio Studies* (IAAO, 2013). Três trabalhos utilizaram a Raiz Quadrada do Erro Médio Quadrático para avaliar o desempenho dos modelos e das PVG geradas.

Tabela 1 - Trabalhos recentes envolvendo avaliação de imóveis rurais.

Autor	Ano	Técnica	Resultados
Resener	2005	Análise de Superfície de Tendência (TSA)	Realizaram avaliação em massa de imóveis atingidos por inundação.
Machado	2006	Regressão Múltipla + TSA para gerar a PVG	Elaboraram a PVG de Chapecó-SC.
Lourival et al.	2008	Regressão múltipla	Analisaram o comportamento das variáveis que influenciam o valor da terra no Pantanal Sul-mato-grossense.
Silva, Nali e Marote	2009	Regressão múltipla	Avaliaram imóveis rurais no Submédio São Francisco.
Oliveira	2010	Simulação de Preços	Avaliaram a viabilidade da realização do convênio entre RFB e municípios.
Sunak e Madlener	2012	Regressão múltipla + Análise de <i>intervisibilidade</i> + Regressão Geograficamente Ponderada (heterogeneidade espacial)	Avaliaram o impacto provocado pela implantação de projetos de energia eólica na região da Renânia do Norte (Vestefália, Alemanha) sobre os preços das terras
Feichtinger e Salhofer	2013	Regressão múltipla + Modelo espacial geral (SAR + SEM)	Identificaram os fatores que influenciam os preços das terras agrícolas na Bavária, utilizaram um conjunto extenso de dados de mercado composto por 7300 negócios realizados dos anos de 2001 e 2007.

Tabela 2 - Compilação das diferentes metodologias propostas para a elaboração de plantas de valores.

Autores	Tipo	Nº obs.	Metodologia	Validação /Nº obs.	Detalhes
Dantas, Sá e Portugal, (2003)	Urbana/ apartamentos	147	Modelo Clássico de Regressão + Geoestatística.	Não/0	Usa a superfície gerada pela krigagem dos resíduos para corrigir os valores obtidos pelo MCRL.
Michael, (2004)	Urbana/ apartamentos	186	MCRL+Superfície de tendência até 4ª ordem.	Não/0	Utiliza imóvel paradigma para gerar a PVG para as faces de quadra.
Resener, (2005)	rural	57	MCRL+Superfície de tendência até 4ª ordem.	Não/0	Compara valores obtidos pelo modelo com valores de indenização, identificando supervalorização dos mesmos.
Trivelloni, (2005)	Urbana/ diversos	249 ofertas	Regressão Espacial SEM+krigagem por blocos.	Sim/243	Realiza avaliação de desempenho (norma IAAO) com os dados utilizados na modelagem.
Dantas, Portugal e Prado, (2006)	Urbana/ terrenos	1055	Modelo Clássico de Regressão + Geoestatística.	Não/0	Usa a superfície gerada pela krigagem dos resíduos para corrigir os valores do MCRL e gerar a Planta de Valores Espacial
Machado, (2006)	rural	47	MCRL+Superfície de tendência até 4ª ordem.	Não/0	Utiliza um imóvel paradigma para gerar uma malha densa de 62000 pontos a partir da qual é gerada a PVG.
Hornburg, (2009)	Urbana/ terrenos	39 ofertas	Regressão Espacial SAR+krigagem ordinária.	Sim/39	Utiliza 3 imóveis paradigmas para gerar uma malha (não informa nº de pontos) e gerar a PVG. Realiza avaliação de desempenho (norma IAAO) com os dados utilizados na modelagem.
Dalaqua et al. (2010)	Urbana/	123	Análise de agrupamento + Modelo Clássico de Regressão	Não/0	Propõe a utilização da análise de <i>clusters</i> para identificar grupos homogêneos e gerar modelos a partir destes. Utiliza métodos indiretos para obter dados em regiões onde não há informação.
Marques et al. (2012)	Urbana/ terrenos	71	Modelo Clássico de Regressão + Geoestatística.	Sim/6	Utiliza a Raiz Quadrada do Erro Médio Quadrático (REMQ) para validar a planta de valores.
Malaman et al. (2012)	Urbana/ terrenos	71	Modelo Clássico de Regressão	Sim/6	Compara 6 modelos, utilizando a REMQ.
Malaman et al. (2014)	Urbana/ terrenos	71	Camparação dos MCRL com a Lógica Fuzzy	Sim/6	Utiliza a REMQ para validar os modelos. Propõe que a avaliação do terreno seja feita pelo somatório dos pixel por ele delimitado.

3 METODOLOGIA

De acordo com Lakatos e Marconi (1991) a especificação da metodologia da pesquisa é a que abrange maior número de itens, pois responde, a um só tempo, às questões Como?, com que?, onde? e quanto?

Neste sentido, serão descritos os principais procedimentos a serem desenvolvidos com o objetivo de responder a estas questões e alcançar os objetivos propostos nos itens 1.1.1 e 1.1.2. Assim sendo, primeiramente será apresentada o município onde foram coletados os dados; em seguida serão descritos os procedimentos metodológicos adotados, sendo apresentados em fluxogramas com o objetivo de facilitar a compreensão das etapas; finalmente serão informados quais os recursos tecnológicos utilizados para execução das atividades propostas.

A proposta metodológica deste trabalho trata-se da utilização de um método baseado na combinação da econometria espacial e da geoestatística para realizar a modelagem dos efeitos espaciais em mercados de terras rurais, elaborar plantas de valores genéricos e avaliar o desempenho dos produtos gerados.

3.1 CARACTERIZAÇÃO DA ÁREA DE ESTUDOS

O projeto foi desenvolvido no município de Petrolina-PE, cuja sede se localiza nas coordenadas 09°23'55" Sul, 40°30'03" Oeste, distante a cerca de 714,4km da capital Recife. Faz parte da região de Desenvolvimento do Sertão Pernambucano e limita-se com os municípios de Dormentes (N); estado da Bahia (S), Lagoa Grande (L) estado da Bahia e Afrânio (O).

O município é banhado pelas águas do Rio São Francisco, o qual proporciona a disponibilidade hídrica para a prática da agricultura irrigada. Em virtude desta característica, nele desenvolveu-se um importante polo de fruticultura que de acordo com ranking elaborado pelo IBGE (2013) coloca o município na primeira colocação entre os municípios produtores de frutas do país. Nele, o valor da produção frutícola somou R\$ 918 milhões, ou seja, 3,9% do valor total da produção frutícola nacional em 2013.

A escolha do município se deve ao fato de que para realização da modelagem dos mercados de terras é necessário contar com abundância de informações a respeito de transações e ofertas. Sendo assim, buscou-se um município que

possuísse uma agricultura dinâmica e tivesse um mercado de terras consolidado, a fim de que a pesquisa de mercado realizada proporcionasse uma quantidade de dados de mercado suficiente para a realização das análises. Por atender aos referidos critérios o município de Petrolina foi selecionado.

As principais vias de acesso ao município são as BR 232/110 e BR316/428/122 e a rodovia estadual PE360.

O município é composto por quatro distritos Petrolina, Cristália, Curral Queimado e Rajada.

A estimativa da população é de cerca de 326.017 habitantes e a taxa de urbanização do município é de cerca de 74,57%.

O clima é Tropical Semiárido, com chuvas de verão. O período chuvoso se inicia em novembro com término em abril. A temperatura média anual é de 26°C.

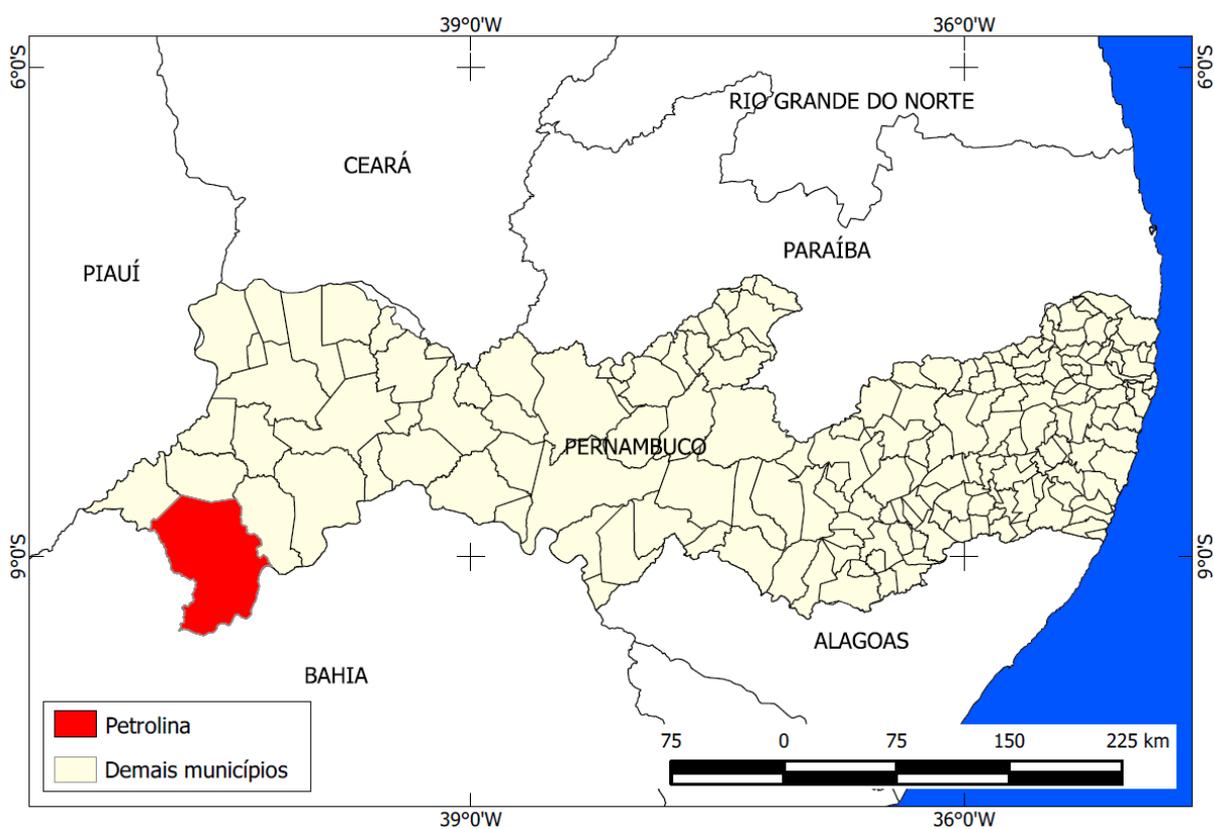


Figura 5 - Localização do município de Petrolina-PE no estado de Pernambuco.

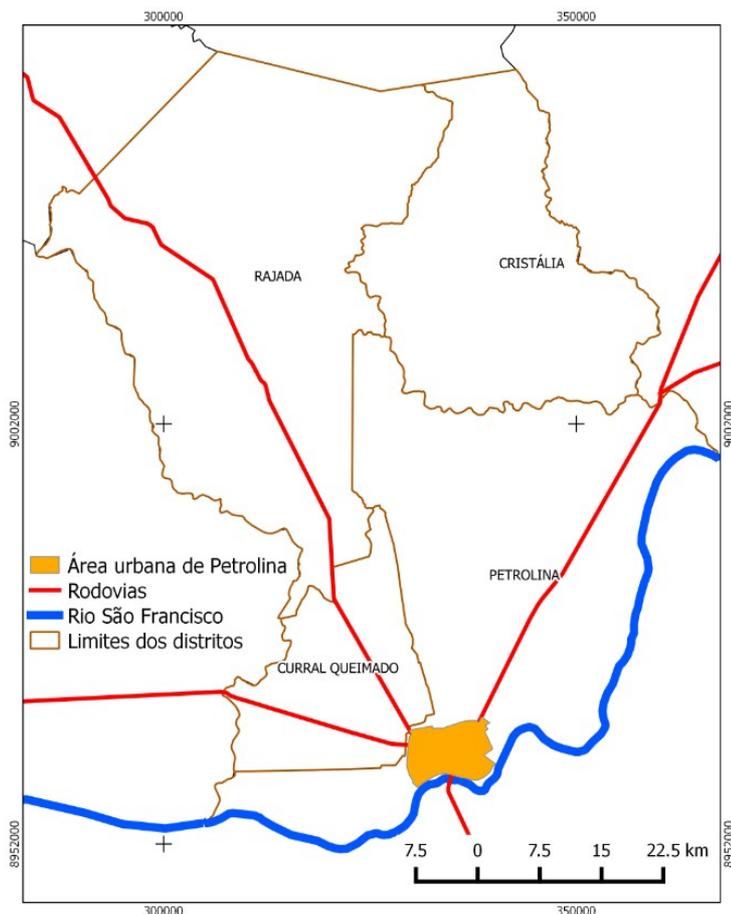


Figura 6 - Área de estudos, município de Petrolina e seus 4 distritos: Petrolina, Curral Queimado, Cristália e Rajada.

De acordo com EMBRAPA, (2001) os solos presentes no município são Areias Quartzosas, Latossolos Amarelos, Planossolos, Podzólicos Amarelos e Vermelho-Amarelos, e Brunos não Cálcicos.

Na Tabela 3 é possível observar que a maior parte dos estabelecimentos estão ocupados com as atividades de Lavouras temporárias, Horticultura e floricultura enquanto que a atividade que ocupa a maior parte da área é a pecuária e criação de outros animais.

Tabela 3 - Estabelecimentos e área total, por classes de atividade econômica no município de Petrolina-PE no ano de 2006.

Classes	Estabelecimentos	Área (ha)
Total	5.372	147.649
Lavouras temporárias	1.142	30.736
Lavouras permanentes	92	619
Horticultura e floricultura	1.975	32.301

Produção de sementes, mudas e outras formas de propagação vegetal	5	547
Pecuária e criação de outros animais	1.847	73.216
Produção florestal - florestas plantadas	284	9.917
Produção florestal - florestas nativas	15	97
Pesca	8	22
Aquicultura	4	194

Fonte: IBGE, censo agropecuário de 2006.

Na Tabela 4 pode-se ver que em cerca de 85% dos estabelecimentos agropecuários a condição do produtor é de proprietário das terras, e cerca de 7% são Assentados sem titulação definitiva.

Tabela 4 - Condição do produtor em relação às terras segundo os estabelecimentos no município de Petrolina-PE no ano de 2006.

Condição	Nº estabelecimentos
Nº de Estabelecimentos agropecuários	5372
Proprietário	4575
Assentado sem titulação definitiva	395
Arrendatário	83
Parceiro	19
Ocupante	220
Produtor sem-área	80

Fonte: IBGE, censo agropecuário de 2006.

O município de Petrolina realizou convênio com a RFB para realizar a cobrança e fiscalização do ITR em 18/04/2013. A arrecadação do ITR no município entre os anos de 2009 a 2013 pode ser visualizada na Tabela 5.

Tabela 5 - Arrecadação anual do ITR pelo município de Petrolina-PE.

Ano	Arrecadação do ITR (R\$)
2009	196.596,43
2010	150.837,45
2011	223.960,47
2012	177.825,95
2013	275.070,57

Fonte: http://www3.tesouro.gov.br/estados_municipios/municipios_novosite.asp

As culturas agrícolas com maior valor da produção no município de Petrolina podem ser observadas na Tabela 6, juntas foram responsáveis por mais de 600 milhões de reais no ano de 2012.

Tabela 6 - Informações das principais culturas agrícolas do município de Petrolina em 2012

Principais produtos das lavouras temporárias e permanentes	Área plantada ou destinada à colheita (ha)	Área colhida (ha)	Quantidade produzida (t)	Rendimento médio (kg/ha)	Valor (1 000 R\$)
Manga	7900	7900	173800	22000	129.020,00
Uva	4650	4650	153450	33000	352.122,00
Goiaba	2230	2230	73600	33004	85.648,00
Banana	1820	1820	38000	20879	27.835,00
Coco-da-baía	1620	1620	48600	30000	13.666,00
Mandioca	480	235	2057	8753	1.210,00
Maracujá	200	200	3200	16000	5.528,00
Melancia	180	180	3520	19556	1.097,00
Cebola	80	80	2000	25000	2.351,00
Feijão (em grão)	3280	80	93	1163	312,00
Cana-de-açúcar	115	75	2325	31000	372,00
Limão	70	70	1190	17000	1.097,00
Mamão	68	68	1088	16000	707,00
Melão	65	65	1430	22000	1.047,00
Tomate	60	60	2400	40000	2.532,00
Batata-doce	30	30	378	12600	293,00
Total					624.837,00

As culturas permanentes como uva, manga e goiaba, são produzidas utilizando irrigação, na Tabela 7, pode-se verificar que a maioria das propriedades utiliza sistemas de irrigação por aspersão ou sistemas localizados (mais eficientes).

Tabela 7 - Estabelecimentos com área irrigada, por método utilizado, segundo os estabelecimentos e a área no município de Petrolina-PE em 2006.

Método de irrigação	Estabelecimentos	Área (ha)
Total	2481	18582
Inundação	115	389
Sulcos	348	1201
Aspersão (pivô central)	1	-
Aspersão (outros métodos de aspersão)	1402	7277
Localizado (gotejamento, microaspersão, etc.)	994	9493
Outros métodos de irrigação e/ou molhação	38	87

Fonte: IBGE, censo agropecuário de 2006.

Com relação à pecuária, na Tabela 8 pode-se ver que a principal atividade pecuária no município é a criação de caprinos e ovinos o que explica o grande

quantitativo de área utilizado para pecuária como foi mostrado na Tabela 3.

Tabela 8 - Efetivo dos três principais rebanhos do município de Petrolina em 2011.

Culturas	Nº de cabeças
Caprino	140.000
Ovino	100.100
Bovino	23.600

3.2 PROCEDIMENTOS METODOLÓGICOS

O fluxograma da Figura 7 apresenta, de forma resumida, as etapas do desenvolvimento deste estudo.

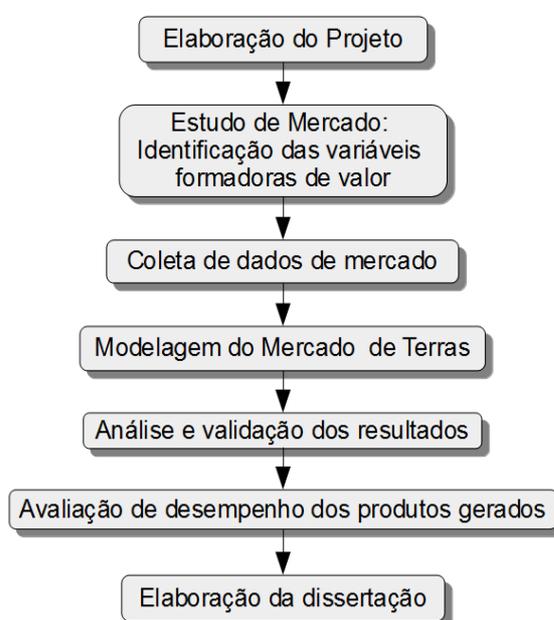


Figura 7 - Fluxograma geral das atividades desenvolvidas.

3.2.1 Estudo de Mercado: Identificação das Variáveis Formadoras de Valor

Esta etapa foi uma das mais importantes do trabalho, devido ao fato que dela dependia a etapa de especificação das variáveis influenciadoras do valor dos imóveis. Foi realizada por meio de pesquisa no município e região do entorno, com o objetivo de conhecer detalhadamente as características físicas, as vocações da região, os principais sistemas produtivos, os rendimentos alcançados e as tendências mercadológicas que atuam na região e que influenciem o valor dos imóveis.

As entrevistas com técnicos, produtores e profissionais de órgãos de pesquisa

e extensão rural também foram fundamentais para obter detalhes sobre as características mencionadas, além de auxiliar na etapa seguinte que foi a coleta de dados de mercado.

Como técnica de entrevista foi utilizada o tipo Padronizada ou Estruturada, a qual, de acordo com Lakatos e Marconi (2003), pode ser definida como aquela na qual o entrevistador segue um roteiro previamente estabelecido e as perguntas feitas ao indivíduo são predeterminadas. Ainda segundo os mesmos autores, o motivo da padronização é obter, dos entrevistados, respostas às mesmas perguntas, permitindo "que todas elas sejam comparadas com o mesmo conjunto de perguntas, e que as diferenças devem refletir diferenças entre os respondentes e não diferenças nas perguntas".

3.2.2 Coleta de Dados: Ofertas e Transações

A coleta de dados envolveu a pesquisa de dados de mercado oriundos de transações e ofertas. Assim, dois requisitos deveriam ser atendidos por estes dados: devem estar atualizados e georreferenciados. A atualização refere-se à necessidade de que os dados obtidos reflitam a realidade de mercado da região estudada, tendo em vista que uma das principais aplicações dos produtos gerados com este projeto é a avaliação em massa de imóveis rurais. O georreferenciamento dos dados de mercado tem a ver com a utilização das técnicas de modelagem espacial, as quais necessitam das coordenadas dos dados envolvidos na modelagem para processamento e obtenção dos resultados.

A pesquisa dos dados de mercado foi realizada no mês de março de 2014. Envolveu a identificação das informações dos imóveis transacionados ou ofertados. Fontes potenciais foram os agentes imobiliários, classificados de jornais e/ou internet, consultas ao Cartório de Registro de Imóveis, declarações de ITBI junto à prefeitura, entre outras.

Após identificação, os imóveis foram visitados com o objetivo de obter informações a respeito das principais variáveis identificadas como relevantes no estudo de mercado, bem como realizar entrevista com o proprietário visando obter a informação mais verdadeira possível a respeito do preço praticado na negociação. As informações foram registradas em fichas em papel e posteriormente

armazenados em planilhas eletrônicas. Além disso, foi coletada uma coordenada em cada elemento de mercado visitado.

3.2.3 Modelagem do Mercado de Terras

Optou-se por detalhar a etapa de Modelagem do Mercado de Terras, por ser esta a etapa na qual são aplicados as técnicas da Econometria Espacial e da Geoestatística. Assim, na Figura 8 pode-se visualizar um fluxograma das fases que envolvem esta etapa, as quais seguem explicadas nos próximos parágrafos.

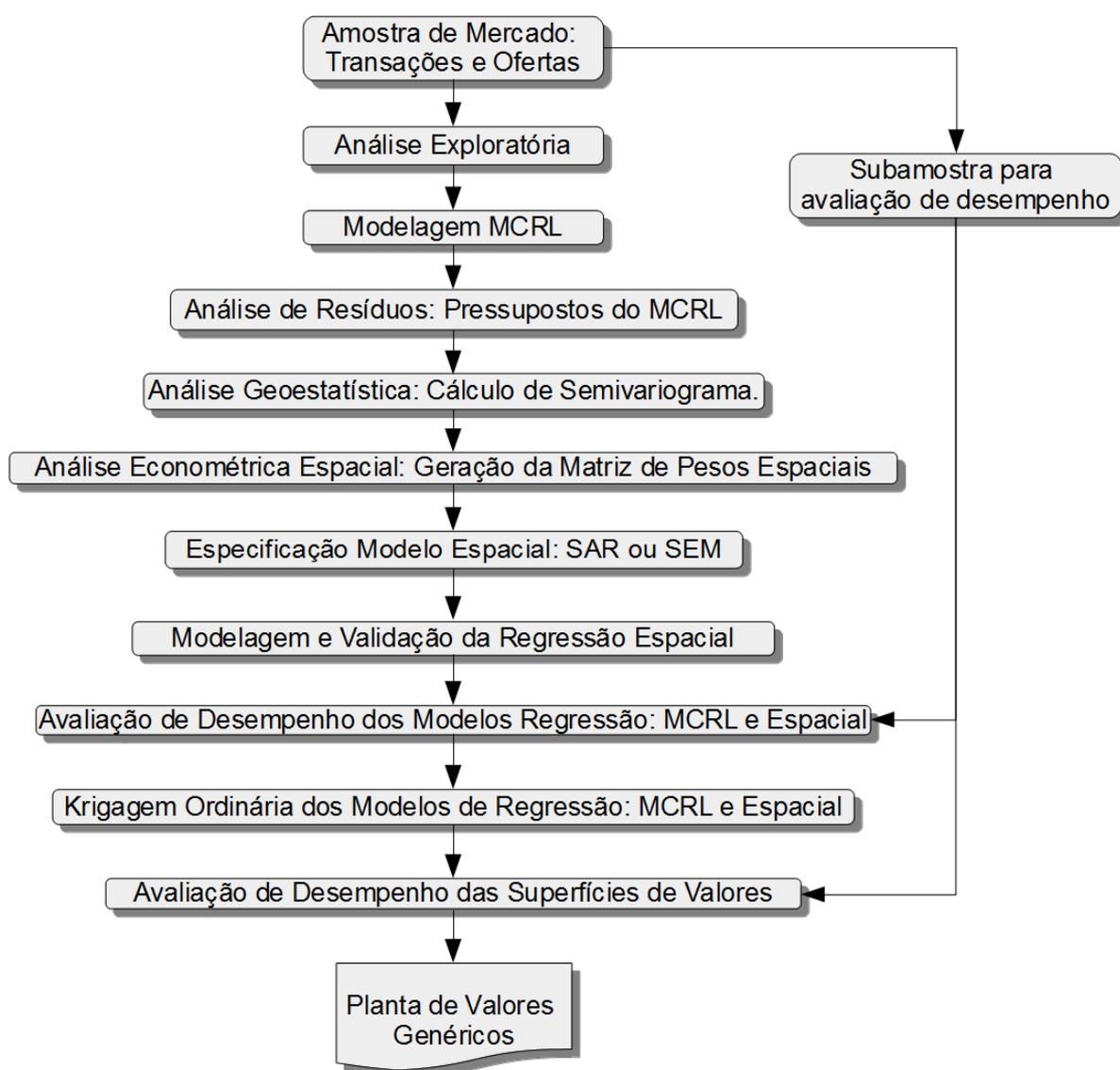


Figura 8 - Fluxograma de detalhamento das atividades de modelagem do mercado de terras mostrados na Figura 1.

No trabalho foram utilizadas fontes primárias, compostas por observações de

mercado, transações e ofertas, atuais e georreferenciadas. Estes dados primeiramente foram submetidos a uma análise exploratória e em seguida foi realizada a Modelagem Clássica por Regressão Linear Múltipla. Nesta etapa, as hipóteses inicialmente formuladas a respeito das variáveis influenciadoras do mercado podem ser comprovadas ou rejeitadas.

É importante que nesta fase não haja problemas na especificação do modelo, pois a ausência de variáveis importantes pode induzir problemas como a autocorrelação espacial nas fases seguintes. Além disso, verificou-se a não violação dos pressupostos do MCRL. O produto desta etapa foi o Modelo Clássico do Mercado de Terras de Petrolina devidamente validado.

Na etapa seguinte, levando-se em consideração que o modelo foi corretamente especificado, espera-se que os resíduos estejam apenas com os erros aleatórios e com os erros devido à estrutura espacial. Nesta etapa, foram utilizados os resíduos extraídos do modelo obtido na etapa anterior.

Estes foram submetidos à análise Geoestatística com o objetivo de verificar a existência da autocorrelação espacial nos mesmos. Para isso, foi construído um semivariograma experimental e a este foram ajustados semivariogramas teóricos (esférico, exponencial e gaussiano). A partir do semivariograma ajustado, foram obtidos os parâmetros do seu ajustamento: efeito pepita, patamar e alcance. Caso se verifique a ocorrência de efeito pepita puro (ausência de dependência espacial) pode-se afirmar que o MCRL obtido na etapa anterior é um modelo adequado. O produto desta etapa foi o semivariograma ajustado com os seus parâmetros.

Dos parâmetros obtidos pelo ajuste do semivariograma o mais importante é o alcance, pois ele indica o raio no qual todos os dados presentes estão autocorrelacionados. Esta informação é valiosa pois nesta etapa, tendo esta informação como dado de entrada e o arquivo digital no formato vetorial de pontos, será construída a matriz de ponderação espacial. Esta será uma matriz de proximidade geográfica baseada na distância, a qual será a informação do alcance fornecida pelo semivariograma obtido na etapa anterior. O produto desta etapa foi a matriz de ponderação espacial baseada na distância.

Utilizando a matriz de ponderação espacial como dado de entrada, pode-se realizar a etapa da Econometria Espacial propriamente dita. A referida matriz

permitirá identificar se a autocorrelação espacial está presente na variável dependente (Modelo SAR) ou nos erros (Modelo SEM). Após a especificação do modelo este será submetido aos testes para verificar a não violação dos seus pressupostos, etapa chamada de validação. O produto desta etapa será o modelo de regressão espacial (MRE) validado.

Os MCRL e o MRE terão a qualidade do seu ajuste pelos critérios de informação de Akaike (AIC) e de Schwartz (SC). O objetivo nesta etapa é comparar o ajuste dos diferentes modelos utilizando um mesmo critério e assim definir qual destes seria o melhor.

Tendo como entradas o MCRL e o MRE, é realizada a etapa de avaliação de desempenho dos modelos. Nesta etapa utiliza-se os critérios para modelos de avaliação em massa estabelecidos pela IAAO. Ambos os modelos foram avaliados na eficiência em avaliar uma amostra de dados que não fez parte da sua modelagem. São utilizados dados de mercado especialmente coletados para esta avaliação. Pode-se calcular os indicadores estabelecidos na norma da IAAO, analisado-se a conformidade dos modelos. O produto desta etapa foram os modelos avaliados no seu desempenho.

Após a realização das etapas citadas até então, foram utilizadas as 84 observações originais bem como os valores preditos pelos modelos para gerar superfícies de preços por meio da krigagem. Trata-se de uma nova etapa envolvendo a Geoestatística entretanto, agora o objetivo é interpolar tais pontos de modo a obter uma superfície de valores que corresponderá à PVG. Também nesta etapa foram comparados os MCRL e MRE. O produto final será a Planta de Valores Genéricos das terras rurais do município de Petrolina-PE.

3.2.4 Análise e Validação dos Resultados

Esta etapa compreende a realização da análise exploratória dos dados, elaboração de diagramas de dispersão das variáveis utilizadas, construção de matrizes de correlação entre as variáveis, cálculo das medidas de posição e dispersão, construção de gráficos box-plot, realização de testes de significância do modelo e dos parâmetros, testes de normalidade, verificação da existência de valores extremos e pontos influenciadores.

O objetivo destas análises é verificar se os modelos não violam os pressupostos estabelecidos para que os mesmos consigam realizar avaliações não tendenciosas, eficientes e consistentes.

3.2.5 Avaliação de Desempenho dos Produtos Gerados

A avaliação do desempenho do modelo proposto se dará a partir de uma amostra composta por dados de mercado (negócios realizados e ofertas) que não fizeram parte dos dados utilizados na elaboração do modelo.

A avaliação se deu de acordo com os critérios estabelecidos pela IAAO, a qual define os procedimentos metodológicos para o cálculo dos parâmetros e também informa os limites aceitáveis para cada um dos parâmetros a serem avaliados. Uma vez calculados os parâmetros, estes foram comparados com os valores de referência e avaliada a conformidade dos modelos obtidos.

3.3 RECURSOS TECNOLÓGICOS

O equipamento computacional empregado deve ser capaz de armazenar os dados da pesquisa com segurança, e além disso, deve ser capaz de suportar a operação dos programas destinados à análise estatística, geoestatística, operações de geoprocessamento, planilhas eletrônicas e editores de texto. Para atender a este fim foi utilizado o Notebook Dell Inspiron 15R SE-4570 com processador Intel Core i7, memória RAM de 8GB, HD de 1TB, Placa de Vídeo AMD Radeon HD 7730M 2GB, tela de LED 15", sistema operacional Windows 8.

Tendo em vista que os usuários potenciais dos produtos gerados são os municípios, foram utilizados softwares livres e gratuitos para realizar todas as análises bem como gerar todos os produtos. Espera-se com isso facilitar a adoção das metodologias propostas devido ao custo mais acessível.

As planilhas eletrônicas bem como os editores de texto utilizados foram do conjunto de aplicativos para escritório livres e multiplataforma conhecidos como Apache OpenOffice, atualmente é distribuído para Microsoft Windows, Unix, Solaris, Linux e Mac OS X, mantida pela Apache Software Foundation. O conjunto usa o formato ODF (OpenDocument) — formato homologado como ISO/IEC 26300 e NBR ISO/IEC 26300 — e é também compatível com os formatos do Microsoft Office, além

de outros formatos.

As análises econométrico espaciais bem como as análises geoestatísticas foram realizadas no programa R (R DEVELOPMENT CORE TEAM, 2013). O Ambiente R ou simplesmente R é uma linguagem de programação orientada a objetos criada em 1996 por Ross Ihaka e Robert Gentleman (1966). Apresenta código fonte aberto, o que significa que pode ser modificado ou implementado com novos procedimentos desenvolvidos por qualquer usuário a qualquer momento. Possibilita a realização de operações matemáticas simples, a manipulação de vetores e matrizes, bem como a análise e manipulação de dados, com testes paramétricos e não paramétricos, modelagem linear e não linear, análise de séries temporais, análise de sobrevivência, simulação e estatística espacial, entre outros, além de apresentar facilidade na elaboração de diversos tipos de gráficos, no qual o usuário tem pleno controle sobre o gráfico criado. Com isso, consolida-se como uma importante ferramenta estatística e uma alternativa aos programas comerciais. Neste trabalho os principais pacotes utilizados foram *mapproj* (BIVAND; KOH, 2014), *sp* (PEBESMA; BIVAND, 2005), *geoR* (RIBEIRO; DIGGLE, 2001), *rgdal* (BIVAND; KEITT; ROWLINGSON, 2014) e *spdep* (BIVAND, 2014).

Houve necessidade de trabalhar com dados georreferenciados, além disso, a PVG pode ser produzida em um Sistema de Informações Geográficas. Para estas ações será utilizado o QGIS (NANNI et al., 2012), o qual é um software SIG livre com uma interface gráfica simples, escrito em C++ e Python e baseado nas bibliotecas Qt4. Suporta numerosos formatos vetoriais, raster, e bases de dados, além de fornecer uma ampla gama de funções de geoprocessamento raster e vetorial. Trata-se de uma aplicação multiplataforma que funciona em todas as principais versões do Unix, GNU/Linux, bem como Mac OsX e MS Windows. O projeto QGIS é o resultado do trabalho voluntário de um grupo de desenvolvedores, tradutores, autores de documentação e pessoas que ajudam no processo de lançamento de novas versões, identificando e divulgando as falhas do programa. É livremente distribuído com a licença GPL (GNU General Public License) e é um projeto oficial da Open Source Geospatial Foundation (OSGeo). Atualmente encontra-se na versão 2.4.

4 RESULTADOS E DISCUSSÃO

Nesta seção são apresentados os principais resultados alcançados a partir das diligências a campo para coleta da amostra necessária à realização do estudo, processamento e análise dos dados de acordo com a metodologia proposta.

4.1 ESTUDO DE MERCADO: IDENTIFICAÇÃO DAS VARIÁVEIS FORMADORAS DE VALOR

Esta etapa está entre as mais importantes do trabalho, devido ao fato que dela depende a etapa de especificação das variáveis influenciadoras do valor dos imóveis.

Durante o período de 10 a 28/03/2014 foi realizada visita no município de Petrolina e região do entorno, com o objetivo de conhecer detalhadamente as características físicas, as vocações da região, os principais sistemas produtivos, os rendimentos alcançados e as tendências mercadológicas que atuem na região e possam influenciar o valor dos imóveis.

As entrevistas com técnicos, engenheiros de avaliações, profissionais de instituições financeiras, produtores e engenheiros agrônomos de órgãos de pesquisa e extensão rural também foram fundamentais para obter detalhes sobre as características mencionadas, além disso, auxiliaram na etapa seguinte, a coleta de dados de mercado.

Este contato com os atores do mercado de terras da região em estudo permitiu a elaboração de um Formulário de Pesquisa de Campo com uma série de itens que visavam identificar as principais questões relacionadas ao referido mercado. Este formulário orientou a realização da entrevista estruturada. O formulário foi disponibilizado no APÊNDICE 01 e dele constam questões que vão desde a identificação do proprietário e informante, passando pela caracterização do imóvel e das atividades produtivas nele praticadas até a identificação dos detalhes da negociação, preço e forma de pagamento.

4.2 COLETA DE DADOS: OFERTAS E NEGÓCIOS REALIZADOS.

A etapa da coleta de dados de mercado foi realizada concomitantemente com a etapa do estudo de mercado devido à necessidade de otimizar a utilização do tempo e dos recursos financeiros. Como ponto de partida, foi obtida uma lista dos imóveis levados a registro nos últimos dois anos no município de Petrolina. A partir desta, cerca de 90 transações foram a princípio identificadas. Esta lista serviria apenas para identificar comprador, vendedor, seus respectivos endereços e a localização do imóvel, pois os valores informados das transações poderiam estar subdeclarados e necessitariam de confirmação.

A partir desta lista, procedeu-se o trabalho de identificação dos endereços e busca de telefones de contato. Petrolina possui uma agricultura empresarial na qual boa parte dos proprietários de imóveis rurais são também donos de empresas ou profissionais liberais na sede do município, de forma que a maneira mais eficiente de identificá-los foi na cidade, e não no campo, como é mais frequente nas pesquisas com imóveis rurais.

Muitos dados da lista consistiam de endereços incompletos ou incorretos o que dificultou sobremaneira a identificação dos proprietários. Além disso, muitas vezes quando os endereços eram identificados o proprietário já havia se mudado daquela residência. Estes fatos, por si só acabaram por reduzir o número de imóveis a pesquisar.

Após a identificação e a localização dos proprietários, foram realizados os primeiros contatos e entrevistas, sendo que nesta etapa destaca-se a recomendação de Dantas (2005) de que é essencial *“o entendimento da psicologia da natureza humana no que se refere aos sentimentos e às atitudes diante de dificuldades e reações, bem como a observação perspicaz, a compreensão e habilidade de orientação”*.

As entrevistas consistiam primeiramente da explicação sobre a motivação para a realização do trabalho, e da garantia de que os nomes do proprietário e da propriedade não seriam revelados. A partir daí, utilizava-se o formulário disponibilizado no **APÊNDICE 01** o qual orientou a realização da entrevista estruturada. As informações obtidas durante a entrevista eram tabuladas nas

planilhas eletrônicas ao final do dia.

Alguns fatos apenas são possíveis de serem percebidos durante a realização das entrevistas. Um exemplo é que mesmo o imóvel tendo sido levado a registro nos últimos dois anos não significa que a transação efetivamente também houvesse ocorrido no referido período. Foram frequentes os casos de imóveis negociados a mais de dez anos e que só haviam sido registrados recentemente.

Outro exemplo é que muitas vezes, registrar o imóvel como uma compra e venda pode ser a maneira mais fácil de regularizar a divisão dos bens de uma herança por exemplo. Sendo assim, também foram frequentes os casos de imóveis registrados como compra e venda mas que não consistiam de uma transação efetivamente realizada.

Tanto na primeira quanto na segunda situações descritas, as informações obtidas eram descartadas por fugirem do objetivo do trabalho, entretanto, reduziam significativamente a quantidade de imóveis relevantes para o trabalho.

Fato curioso é que algumas vezes o imóvel só era levado a registro para que uma venda fosse realizada. Dessa forma, foi identificado o caso de um determinado imóvel cuja transação já havia ocorrido a mais de dez anos, mas que o proprietário registrou para efetuar a venda sendo que a venda recente sequer havia sido levada a registro. Nestes casos, o trabalho não era em vão, pelo contrário, havia ali uma informação recente do mercado a ser aproveitada.

A coleta dos dados de ofertas consistiu na visita às principais regiões do município em busca de imóveis à venda, seguida de contato com os proprietários e realização de entrevistas. Nestes casos não era revelado a finalidade do trabalho, pelo contrário, simulava-se o interesse de um potencial comprador, com o objetivo de captar o valor mais provável pelo qual o imóvel seria negociado.

Também foram obtidas informações fornecidas por Engenheiros de Avaliações que atuam no mercado de Terras de Petrolina, e que ajudaram a complementar a pesquisa. Obviamente algumas das informações obtidas não eram as mesmas que haviam sido coletadas a partir da realização de entrevistas, o que não inviabilizava sua utilização no trabalho. O ponto positivo destas informações é que tratavam-se de dados coletados por profissionais que são profundos conhecedores da realidade de mercado da região.

Ao final dos trabalhos foram obtidos, entre ofertas e transações, 110 dados de mercado, nas diferentes regiões do município de Petrolina-PE. De acordo com informações do Sistema Nacional de Cadastro Rural-SNCR o município de Petrolina possui 10.005 imóveis rurais, sendo assim, a amostra obtida corresponde a 1% do total de imóveis.

Após a tabulação foi verificado que os dados coletados durante a pesquisa compreenderam o período de tempo situado entre os anos de 2005 e 2014. Com o objetivo de trabalhar com uma amplitude de ano da transação mais estreita foram retirados da amostra 6 observações cujo ano da transação era inferior a 2009. Na Figura 9 são exibidos o histograma e a distribuição espacial dos dados de acordo com o ano da transação, após a retirada dos dados mencionados.

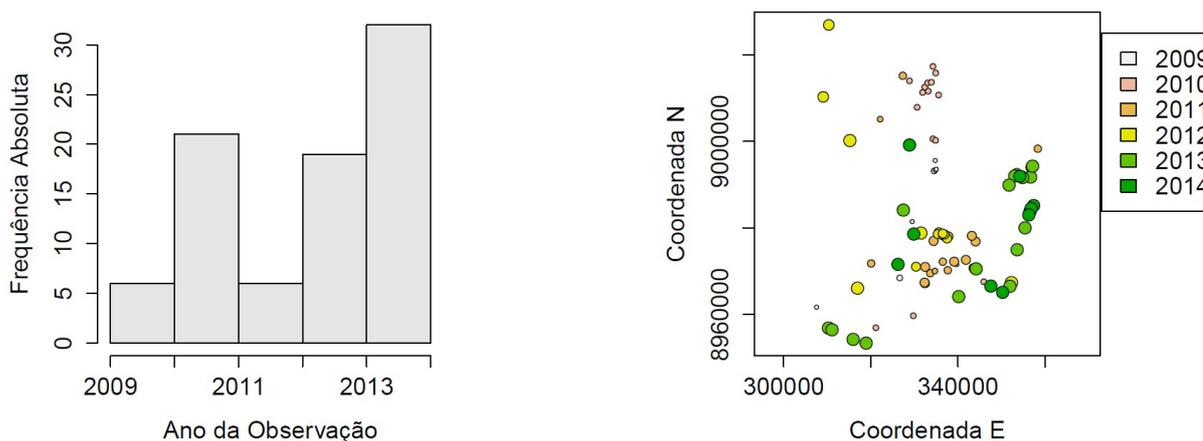


Figura 9 - Histograma de frequência absoluta e distribuição espacial dos dados classificados de acordo com o ano da sua ocorrência.

Após a retirada dos dados anteriores a 2009 restaram 104 observações. Destas foi retirada uma amostra aleatória composta por 20 pontos de verificação a serem utilizados na avaliação de desempenho. Por fim a amostra de trabalho contou com 84 observações de mercado. A distribuição espacial da amostra de trabalho e dos pontos de verificação pode ser visualizada na Figura 10.

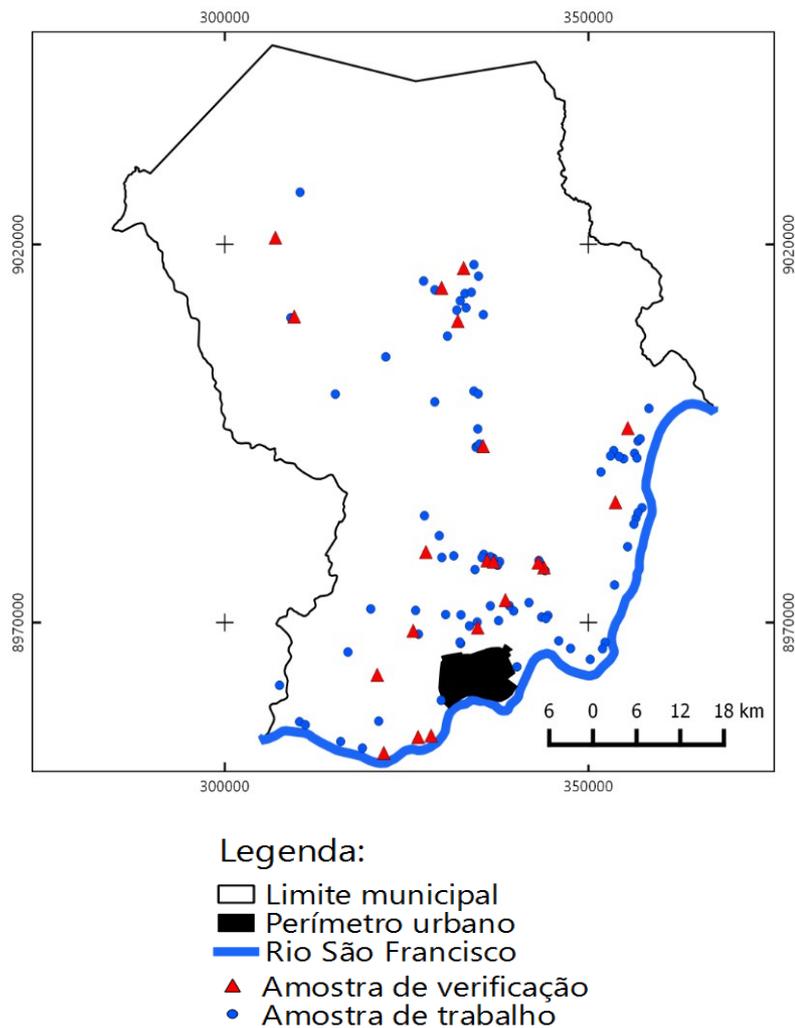


Figura 10 - Distribuição espacial da amostra de trabalho e da amostra de verificação.

De acordo com ABNT (2004), quando no processo de avaliação é utilizado o método comparativo direto de dados de mercado, a definição do tamanho adequado da amostra se baseia no número de variáveis (k) envolvidas na construção do modelo de regressão linear. A partir disso, no caso dos imóveis rurais são definidas pontuações (Tabela 9) que juntamente com outros fatores definem o grau de fundamentação alcançado pelo trabalho. Por exemplo: um modelo que possua 6 variáveis precisa ter pelo menos 21 dados de mercado para que receba a pontuação máxima de 18 pontos.

Tabela 9 - Pontuação para fins de classificação das avaliações quanto ao grau de fundamentação.

Especificação das avaliações de imóveis rurais.	Condição	Pontos	Condição	Pontos
Número de dados de mercado efetivamente utilizados.	$\geq 3(k+1)$ e no mínimo 5	18	≥ 5	9

4.3 MODELAGEM DO MERCADO DE TERRAS

A partir da tabulação das informações obtidas por meio da pesquisa de campo, e do contato com os profissionais que atuam na região, foram geradas hipóteses a respeito das diversas variáveis que poderiam ajudar a explicar a variação dos preços das terras no mercado em estudo.

Variável Dependente:

Valor Total do imóvel por hectare (VTI): Obtido a partir do valor da transação (no caso de negócios realizados) ou valor ofertado (no caso de imóveis em oferta) dividido pelo quantitativo da área total do imóvel em hectares.

Variáveis Independentes:

Coordenadas UTM da localização do imóvel: Foram utilizadas as coordenadas UTM da localização do imóvel na forma de polinômio de 1º grau. As coordenadas foram transformadas por translação de eixos com a origem em um ponto arbitrário localizado a sudoeste da área de estudos. A realização deste procedimento é justificada pelo objetivo de não obter desvios negativos. Em seguida, os desvios foram divididos por 1000 com o objetivo de facilitar o processamento dos dados pois alguns algoritmos costumam apresentar problemas quando utiliza-se matrizes com coordenadas de altos valores. Sendo assim, o resultado são duas variáveis:

01-Desvios N (DN)

02-Desvios E (DE)

03-Natureza do evento (NE): Os elementos de pesquisa caracterizados como transações receberam o valor zero e os que foram caracterizados como ofertas receberam o valor 1.

04-Situação (SIT): Foi representada por meio de uma variável *dummy* com o objetivo de abranger três situações específicas a saber: Áreas de Sequeiro, corresponde ao valor 1; Projetos de Irrigação da CODEVASF corresponde ao valor 2; Margens do Rio do São Francisco corresponde ao valor 3.

05-Área total (AT): Área do imóvel fornecida pelo informante durante entrevista, medida em hectares.

06-Área irrigada (AI): Quantitativo de área irrigada em hectares fornecida pelo informante durante entrevista.

07-Percentual de área irrigada (PAI): Relação entre a área do imóvel e o quantitativo de área efetivamente irrigada.

08-Limite confrontante com o rio (LCR): Quantitativo em metros de margem de rio que cada imóvel possui, fornecido pelo informante do imóvel.

09-Distância da sede (DS): Distância da sede do município de Petrolina em relação ao imóvel, medida em quilômetros, percorrendo estradas e rodovias.

10-Distância sem pavimentação (DSP): Distância situada em parte do percurso que dá acesso ao imóvel a qual não possui pavimentação asfáltica, medida em quilômetros.

11-Disponibilidade hídrica (DH): Variável dicotômica a qual caracteriza a disponibilidade de recursos hídricos para o desenvolvimento de atividades agrícolas. Áreas de sequeiro recebem o valor 0 e áreas que possuem água disponível recebem o valor 1. Esta variável agrupa as áreas de margens de rio e áreas de lotes irrigados num mesmo grupo, o objetivo foi verificar se existe diferença entre os preços das áreas de sequeiro com relação às demais.

12-Aptidão (AP): Variável *dummy* que busca caracterizar o tipo de exploração do imóvel. Foi codificada da seguinte forma: Imóveis com potencial para pecuária=1; Imóveis com potencial para pecuária e agricultura=2; Imóveis com potencial para agricultura=3; Imóveis com potencial para chácaras de lazer ou implantação condomínios=4.

13-Quantitativo de benfeitorias (QB): Variável *dummy* criada a partir da variável

qualitativa de mesmo nome e que foi codificada da seguinte forma: Imóveis sem benfeitorias=1; Imóveis que possuem quantidade insuficiente de benfeitorias para o pleno desenvolvimento das atividades=2; Imóveis que possuem quantidade adequada de benfeitorias para o desenvolvimento das suas atividades=3.

14-Estado das benfeitorias (EB): Variável *dummy* criada a partir da variável qualitativa de mesmo nome e que foi codificada da seguinte forma: Imóveis que não possuem benfeitorias=1; Imóveis cujas benfeitorias estão sem condições de uso=2; Imóveis cujas benfeitorias necessitam de reparos=3; Imóveis cujas benfeitorias não necessitam de reparos=4.

15-Índice de benfeitorias (IB): Variável *dummy* criada a partir da multiplicação das variáveis Quantitativo de benfeitorias e Estado das benfeitorias e que pode variar de zero a doze.

16-Índice de Produção Vegetal (IPV): Área equivalente de produção vegetal obtida pelo produto da área cultivada em hectare acrescida da unidade e o índice de valoração da produção vegetal. O valor econômico das produções vegetais foi calculado através do método da capitalização da renda para cada ano de vida útil da cultura. A metodologia se baseou no trabalho de SILVA, NALI e MAROTE (2009), e maiores detalhes foram disponibilizados no **ANEXO 01**.

17-Ano da transação (ANO): foi codificada como uma variável *dummy* associada da seguinte forma: 2009=1; 2010=2; 2011=3; 2012=4; 2013=5; 2014=6.

Distância aos polos de influência:

Os SIG possuem ferramentas que possibilitam a medição de distâncias a um determinado polo de influência. Estas ferramentas calculam a distância euclidiana de diversos imóveis a um determinado polo e foram utilizadas neste trabalho para criar variáveis baseadas na distância aos polos de influência do município de Petrolina.

Durante o trabalho foram identificados os seguintes polos de influência:

18-Aeroporto de Petrolina (DAP): Feição pontual (Coordenadas E:328.474,94m; N:8.964.094,369m Fuso: 24);

19-Rio São Francisco (DRSF): Feição Linear;

20-Área urbana de Petrolina (DC): Considerado o centróide da área urbana. Feição Pontual (Coordenadas E:334.705,39m, N:8.962.848,28m; Fuso: 24);

21-Rodovias BR-122, BR-235, BR-407 (DR): Feições lineares;

22-Mercado do Produtor de Juazeiro (DMP): Feição Pontual (Coordenadas E:336.391,42m, N:8.955.322,57m; Fuso: 24);

Na Figura 11 é possível visualizar dois exemplos de como estas variáveis foram criadas:

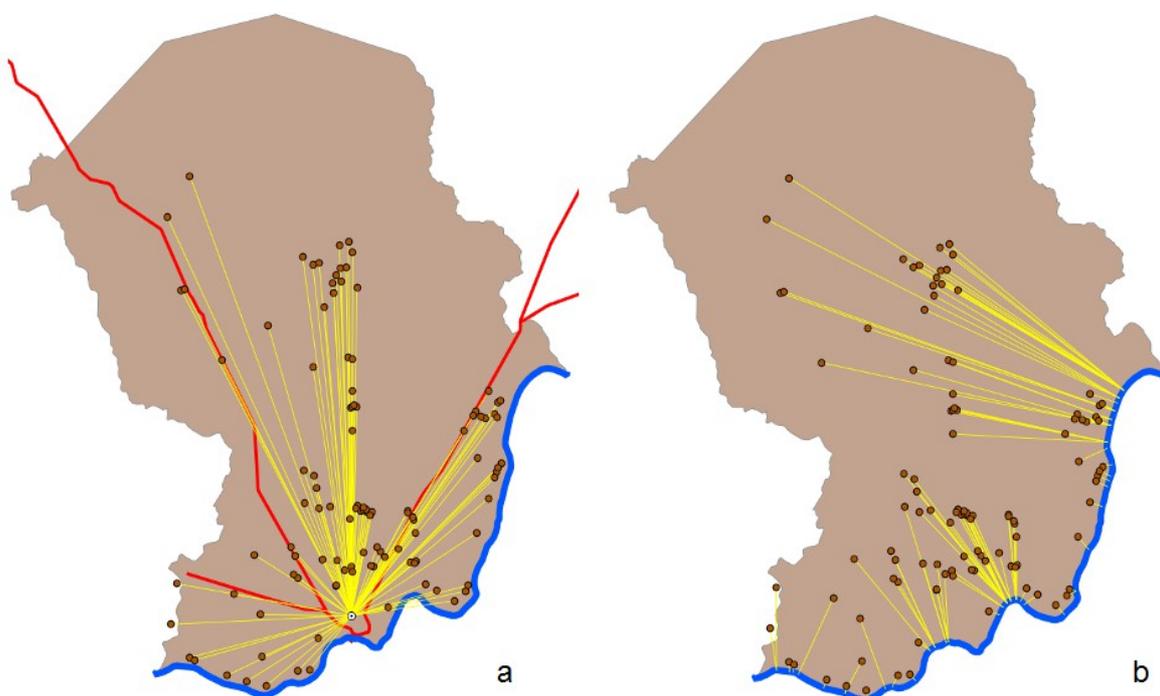


Figura 11 - Obtenção da distância euclidiana dos imóveis em relação aos polos de influência: a) Cidade de Petrolina; b) Rio São Francisco.

4.4 ANÁLISE EXPLORATÓRIA

A análise exploratória, neste trabalho, pode ser resumida em três etapas representadas na Figura 12. Esta análise, como o próprio nome sugere, visa conhecer detalhadamente as variáveis que serão utilizadas posteriormente na modelagem com o objetivo de corrigir problemas preexistentes e também de não introduzir problemas nas etapas subsequentes.

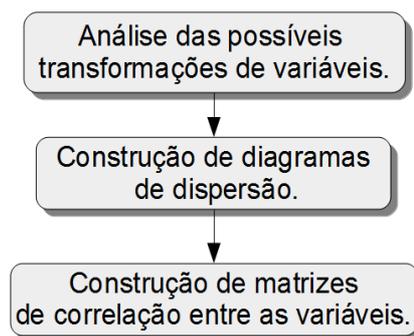


Figura 12 - Fluxograma das etapas relacionadas à análise exploratória.

Na primeira etapa, foram analisadas quais as melhores transformações poderiam ser efetuadas tanto na variável dependente quanto nas variáveis independentes. Esta análise compreendeu a realização da transformação de Box e Cox (1964), testes de normalidade, construção gráficos como o histogramas, box-plots e o gráfico quantil-quantil (para avaliação da normalidade), verificação de tendência nos dados por meio do ajuste de superfície de tendência, e finalmente o cálculo de medidas de posição e dispersão para as variáveis em análise.

Na segunda etapa, foram construídos diagramas de dispersão entre a variável dependente e as variáveis independentes, com o objetivo de investigar qual o tipo de relação entre as mesmas.

Na terceira etapa foi construída a matriz de correlação entre as variáveis envolvidas na modelagem com o objetivo de identificar variáveis com alta correlação para que elas não sejam utilizadas conjuntamente na modelagem, evitando assim sérios problemas de multicolinearidade.

4.4.1 Variável Dependente

Foi realizado o estudo com o objetivo de verificar qual a melhor transformação para variável dependente e para as variáveis independentes utilizando como ferramenta a transformação de Box e Cox (1964).

Por meio de métodos gráficos é possível plotar os valores de \hat{L} (logaritmo da função verossimilhança estimada) versus valores de $\hat{\lambda}$ ajustados para a variável

dependente e analisar qual a melhor transformação. Da análise do referido gráfico é possível identificar o \hat{L}_{\max} com o seu correspondente λ , o qual fornece a melhor escala para a variável Y_i .

Na Figura 13 é possível verificar que o \hat{L}_{\max} está associado ao $\lambda=0,1589818$ o qual corresponde a uma transformação potência; também é possível perceber como o $\lambda=0$ que corresponde à transformação logarítmica está relativamente distante do ponto máximo da função verossimilhança.

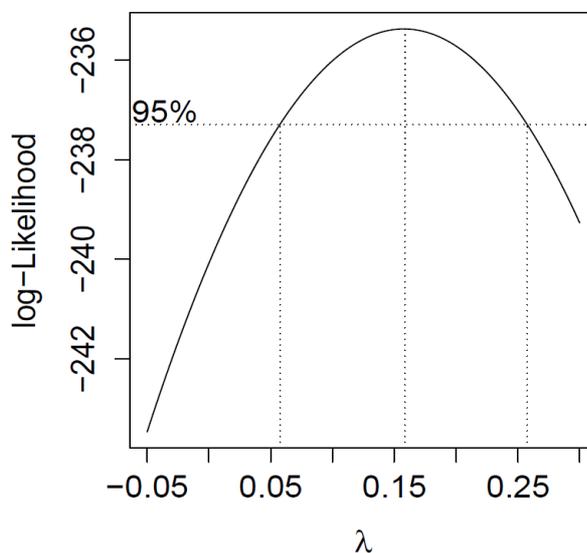


Figura 13 - Gráfico do logaritmo da função verossimilhança estimada versus valores de λ ajustados para a variável dependente.

Com o objetivo de 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 potência como pode ser visto na Figura 14.

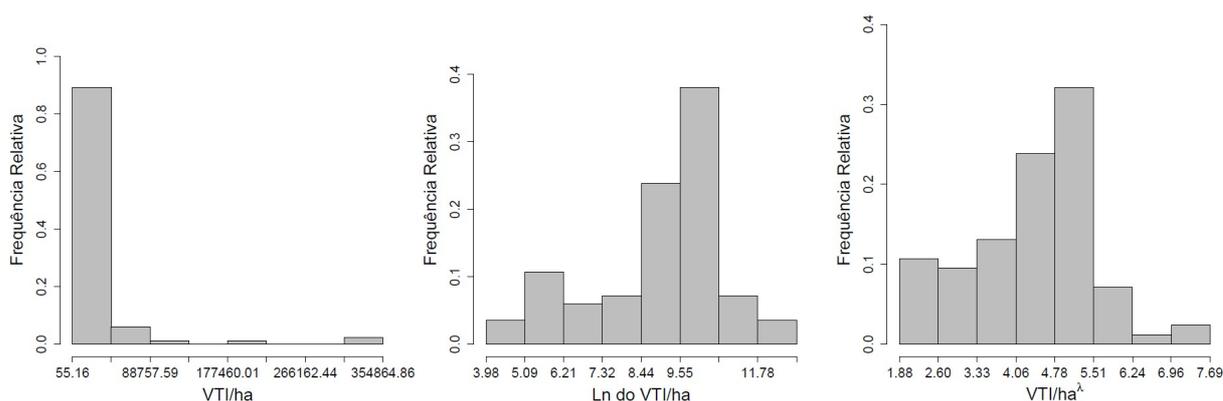


Figura 14 - Histogramas da variável dependente (da esquerda para direita): na escala original; transformação logarítmica; transformação potência com λ obtido pelo método da máxima verossimilhança.

A análise dos histogramas nos permite afirmar que ambas as transformações foram capazes de corrigir a forte assimetria que os dados originais apresentavam; entretanto, comparando visualmente os histogramas dos dados transformados observa-se que as diferenças são a princípio pequenas. Foram então calculados os coeficientes de assimetria e curtose para as três situações em análise e os resultados podem ser observados na Tabela 10.

Tabela 10 - Coeficientes de assimetria e curtose para a variável dependente.

	Assimetria	Curtose
VTI/ha	4,58	25,25
Ln do VTI/ha	-0,74	2,92
VTI/ha^λ	-0,02	3,07

Levando em consideração os parâmetros de que uma distribuição é simétrica se o Coeficiente de Assimetria $|As| < 0,15$; apresenta assimetria moderada se $0,15 \leq |As| < 1,0$; e apresenta assimetria forte se $|As| \geq 1,0$; pode-se perceber que ambas transformações foram capazes de corrigir a forte assimetria que os dados originais apresentavam, sendo que a transformação potência conseguiu tornar a distribuição simétrica. O coeficiente de curtose mais próximo do valor zero permite afirmar que os dados obtidos via transformação logarítmica são mais próximos de uma curva mesocúrtica.

Complementarmente, na Figura 15 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, e agora passaram a fazer parte dos limites definidos nos boxplots em ambas transformações.

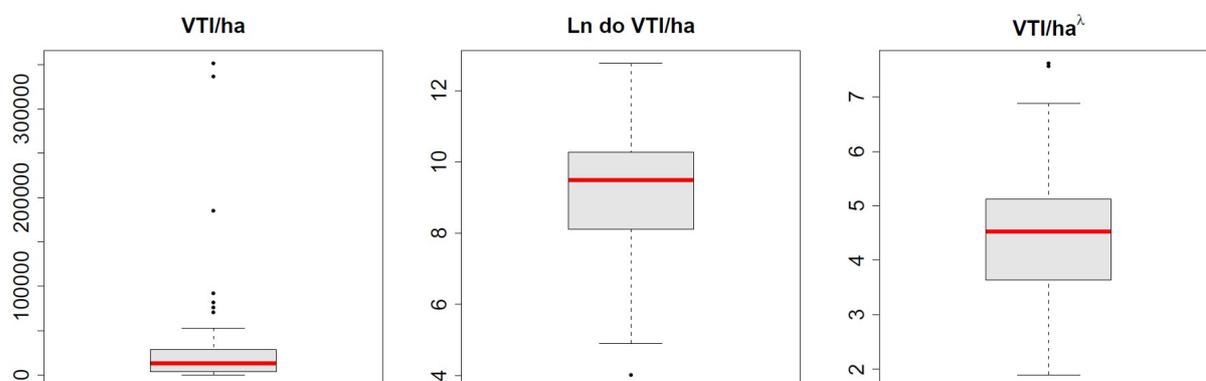


Figura 15 - Gráficos Boxplots para a variável dependente (da esquerda para direita): na escala original; transformação logarítmica; transformação potência com λ obtido pelo método da máxima verossimilhança.

Na Figura 16 estão representados os gráficos que plotam os quantis teóricos da distribuição Normal padrão versus os quantis 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.

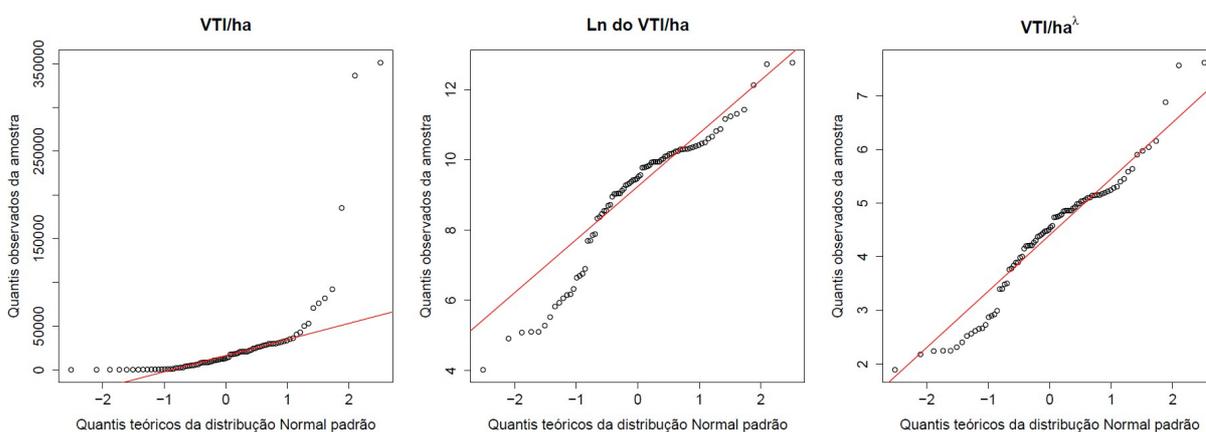


Figura 16 - Gráficos Quantil-Quantil para a variável dependente (da esquerda para direita): na escala original; transformação logarítmica; transformação potência com λ obtido pelo método da máxima verossimilhança.

Na Tabela 11 encontram-se os resultados de três testes de normalidade para as três situações em análise. Pode-se notar que mesmo os dados na escala original já apresentavam distribuição Normal, o que se manteve após a realização de ambas

as transformações.

Tabela 11 - Testes de normalidade para a variável dependente na escala original, transformação logarítmica e utilizando a transformação potência.

	Shapiro-Wilk	Kolmogorov-Smirnov	Chi-quadrado
VTI/ha sem transformação	W = 0,4404 p-value = 2,67E-016	D = 0,3107 p-value < 1,74E-022	P = 120,5714 p-value < 1,02E-021
Ln do VTI/ha	W = 0,9223 p-value = 8,30E-005	D = 0,1585 p-value = 1,90E-005	P = 37,7142 p-value = 1,96E-005
VTI/ha[^]	W = 0,9584 p-value = 0,0083	D = 0,0989 p-value = 0,0410	P = 19,7142 p-value = 0,0197

Na Figura 17 é possível verificar a distribuição espacial dos dados de mercado e a intensidade do valor da variável dependente no município de Petrolina-PE. Percebe-se claramente que os maiores valores do hectare na região concentram-se na porção sul do município, região onde localiza-se o Rio São Francisco e o Projeto de Irrigação Senador Nilo Coelho, enquanto que na região norte do município estão localizados os imóveis que praticam atividades de sequeiro e conseqüentemente ocorre um menor valor por hectare.

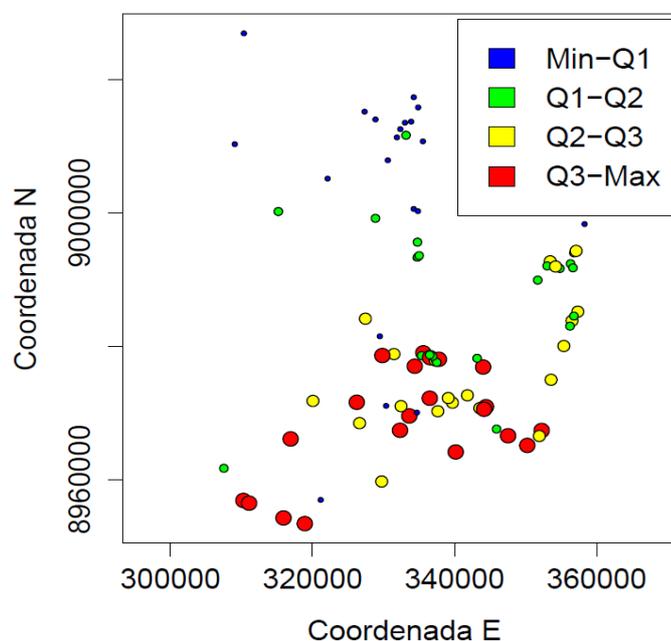


Figura 17 - Distribuição espacial da variável dependente, valor do imóvel por hectare classificado em quartis.

Diante do que foi mostrado na Figura 17 foi realizado o ajuste de uma

superfície de tendência de 1º grau aos valores da variável dependente. A significância global do modelo mostra que os dados apresentam tendência (Tabela 12). Isto também é confirmado pelo nível de significância da componente norte, a qual foi altamente significativa (Tabela 13). Na Figura 18 é mostrado a superfície de 1º grau ajustada. Esta informação será importante, na etapa de construção dos modelos, pois a tendência deverá ser removida.

Tabela 12 - Resultados gerais do ajuste de uma superfície de tendência de 1º grau à variável dependente.

Discriminação	Valor
Desvio padrão	50517,86
Estatística Fc	10,87
Probabilidade associada a Fc	3,88E-005
Coefficiente de determinação	0,21
Coefficiente de determinação ajustado	0,19
Observações	84,00

Tabela 13 - Coeficientes das variáveis independentes da superfície de tendência de 1º grau.

	Coeficientes	Desvio padrão	Estatística t	Nível de significância
intercepto	12.798.142,17	2.836.048,90	4,51	2,14E-005
Coordenada Leste	-0,55	0,43	-1,27	2,07E-001
Coordenada Norte	-1,40	0,32	-4,44	2,87E-005

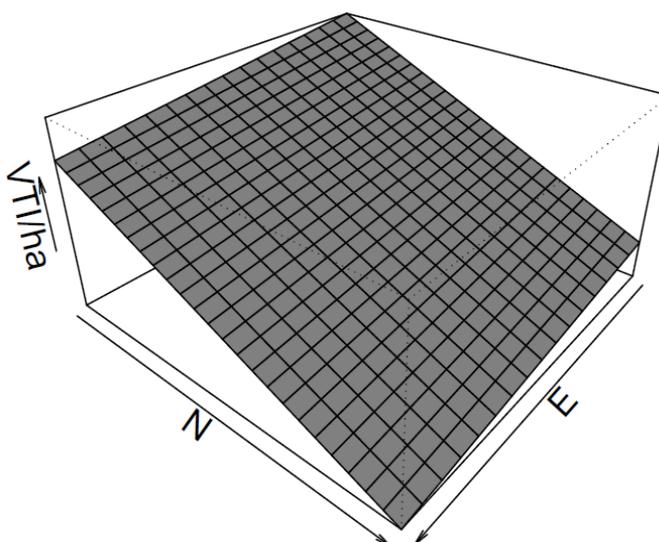


Figura 18 - Superfície de Tendência de 1º Grau ajustada à variável dependente.

A análise exploratória da variável dependente se completa com os dados da estatística descritiva para as três situações em análise, a qual pode ser visualizada

na Tabela 14. Destaca-se que os dados na escala original apresentam uma grande amplitude de variação. Este fato se deve à grande diferença do valor de mercado das propriedades localizadas em regiões de sequeiro e sem possibilidade de prática da irrigação quando comparadas com propriedades localizadas em perímetros irrigados e ou nas margens do Rio São Francisco. Quando existem dados de mercado suficientes uma alternativa para reduzir a grande amplitude é subdividir em regiões mais homogêneas.

Tabela 14 - Estatística descritiva para a variável dependente sem transformação, transformação logarítmica e utilizando a transformação potência.

	VTI/ha	Ln(VTI/ha)	VTI/ha ^λ
Mínimo (R\$/ha)	55,72	4,02	1,89
1º Quartil (R\$/ha)	3.785,36	8,22	3,70
Mediana (R\$/ha)	13.303,57	9,50	4,52
Média (R\$/ha)	27.935,67	9,01	4,37
3º Quartil (R\$/ha)	28.641,67	10,26	5,11
Máximo (R\$/ha)	351.351,35	12,77	7,62
Amplitude (R\$/ha)	351.295,63	8,75	5,72
Assimetria	4,58	-0,74	-0,02
Curtose	25,25	2,92	3,07
Desvio padrão (R\$/ha)	56.207,64	1,91	1,20
Variância (R\$²)	3.159.298.459,88	3,66	1,45
Coeficiente de variação (%)	201,20	21,23	27,56

4.4.2 Variáveis Independentes

O estudo por meio da transformação de Box-Cox para encontrar a melhor escala para a variável dependente também foi realizado para as variáveis independentes. Os resultados encontram-se resumidos na Tabela 15. Destaca-se entretanto a observação de Dantas, (2005) de que a transformação de variáveis “*não é mero exercício de matemática*”, o modelo obtido deve estar coerente para explicar com fidelidade a realidade em estudo.

Tabela 15 - Quadro resumo das transformações das variáveis independentes

Variável Independente	Sigla	Tipo	Transformação	λ
Desvios N.	DN	contínua	não	1
Desvios E.	DE	contínua	não	1
Situação	SIT	<i>dummy</i>	não	1
Natureza do evento	NE	dicotômica	não	1
Limite confrontante com o rio	LMR	contínua	não	1
Área total do imóvel (ha)	AT	contínua	Logarítmica	0
Área irrigada (ha)	AI	contínua	Logarítmica	0
Percentual de área irrigada	PAI	contínua	Potência	0,1569
Distância Petrolina (estradas)	DS	contínua	Potência	0,2930
Distância sem pavimentação	DSP	contínua	Logarítmica	0
Disponibilidade hídrica	DH	dicotômica	não	1
Distância à rodovia	DR	contínua	Potência	0,3837
Distância ao Rio São Francisco	DRSF	contínua	Potência	0,1703
Distancia euclidiana à Petrolina	DC	contínua	Potência	0,1484
Distância ao Mercado do Produtor	DMP	contínua	Potência	0,0356
Distância ao aeroporto	DAP	contínua	Potência	0,2167
Quantidade de Benfeitorias	QB	<i>dummy</i>	não	1
Estado das Benfeitorias	EB	<i>dummy</i>	não	1
Índice de Benfeitorias	IB	<i>dummy</i>	não	1
Índice de Produção Vegetal	IPV	contínua	Logarítmica	0
Aptidão	AP	<i>dummy</i>	não	1
Ano da transação	ANO	<i>dummy</i>	não	1

Durante a análise exploratória também é importante verificar a relação dos preços em relação às variáveis independentes nas suas escalas originais. Isto pode ser verificado por meio do diagrama de dispersão, o qual fornece uma ideia do tipo de relacionamento entre as duas variáveis. No **APÊNDICE 2** estão disponibilizados os diagramas de dispersão para as variáveis em análise.

Analisando-se a dispersão dos pontos nas referidas figuras, percebe-se que a grande maioria não apresenta tendência linear, portanto será necessário, durante a modelagem, buscar uma transformação que linearize os dados e permita a obtenção de um modelo coerente com a realidade do mercado.

A construção da matriz de correlações dois a dois para as variáveis sob

análise é um importante passo para evitar problemas na construção dos modelos. Variáveis independentes que apresentem alta correlação não devem ser utilizadas conjuntamente, sob pena de provocar problemas de multicolinearidade.

Na Tabela 16 encontra-se a matriz de correlações para as variáveis estudadas. Nela é possível perceber que a variável Desvios N apresenta forte correlação com Distancia sede (rodovias), Distância ao Aeroporto, Distância ao mercado do produtor e Distância à sede do município (distância euclidiana). Isto deve-se ao fato de que todas estas variáveis estão medindo uma variação de distâncias num mesmo sentido (norte-sul).

A variável Disponibilidade Hídrica apresentou forte correlação com Aptidão. Isto se deve ao fato de serem ambas variáveis dicotômicas nas quais as melhores aptidões estão relacionadas com a disponibilidade de água.

A variável Índice de Benfeitorias apresentou forte correlação com as variáveis Quantidade de Benfeitorias e Estado das Benfeitorias. Isto se deve ao fato de que o índice de benfeitorias foi obtido a partir da multiplicação das duas outras variáveis. Deve-se atentar para isto no momento da construção dos modelos e não utilizar tais variáveis conjuntamente.

Tabela 16: Matriz de Correlações para as variáveis envolvidas na modelagem (continua).

	VTI/ha	Desvios N	Desvios E	Natureza do evento	Área total (ha)	Área irrig. (ha)	Perc. Area Irrigada	Limite Conf. com o rio.	Distância sede
VTI/ha	1								
Desvios N	-0,44	1							
Desvios E	-0,14	0,04	1						
Natureza do evento	0,36	-0,21	0,25	1					
Situação	0,43	-0,71	0,34	0,43					
Área total (ha)	-0,19	0,44	-0,34	-0,25	1				
Ano	0,35	-0,30	0,23	0,78	-0,30	1			
Área irrig. (ha)	-0,06	-0,04	0,15	-0,15	0,16	-0,05	1		
Perc. Area Irrigada	-0,03	-0,25	0,20	-0,25	-0,31	0,03	0,33	1	
Limite conf. com o rio	0,47	-0,32	0,03	0,43	-0,06	0,35	-0,08	-0,18	1
Distância sede	-0,32	0,90	-0,08	-0,08	0,51	-0,25	-0,08	-0,42	-0,18
Distância chão	-0,24	0,60	-0,29	-0,36	0,62	-0,51	-0,06	-0,22	-0,19
Disponibilidade hídrica	0,25	-0,59	0,23	0,34	-0,26	0,49	0,21	0,39	0,15
Quant. Benf.	-0,07	-0,23	0,01	-0,10	-0,10	0,07	0,06	0,39	0,02
Estado benf.	-0,03	-0,26	-0,02	-0,07	-0,09	0,08	0,25	0,42	0,05
Índice benf.	-0,03	-0,29	-0,04	-0,05	-0,10	0,15	0,14	0,42	0,06
Aptidão	0,42	-0,63	0,21	0,48	-0,44	0,54	0,06	0,27	0,26
Índ. Prod. Veg.	0,00	-0,21	-0,15	-0,22	0,11	-0,15	0,29	0,18	-0,09
Dist. Rodovias	-0,09	0,51	-0,22	-0,17	0,33	-0,40	-0,25	-0,47	0,05
Dist. Rio S. Franc.	-0,33	0,77	-0,53	-0,37	0,50	-0,37	-0,16	-0,30	-0,29
Dist. aeroporto	-0,29	0,90	0,15	0,03	0,44	-0,11	-0,04	-0,34	-0,13
Dist. Petrolina	-0,29	0,90	-0,06	-0,02	0,51	-0,16	-0,04	-0,39	-0,16
Dist. Merc. Produtor	-0,34	0,95	-0,07	-0,08	0,50	-0,21	-0,04	-0,35	-0,22

Matriz de Correlações para as variáveis envolvidas na modelagem (conclusão).

	Distância sede	Distância chão	Disp. hídrica	Quant. Benf.	Estado benf.	Índice benf.	Aptidão	Índ. Prod. Veg.	Dist. Rodovias	Dist. Rio S. Franc.	Dist. aeroporto	Dist. Petrolina	Dist. Merc. Produtor
VTI/ha													
Desvios N													
Desvios E													
Natureza do evento													
Situação													
Área total (ha)													
Ano													
Área irrig. (ha)													
Perc. Area Irrigada													
Lim. conf. com o rio													
Distância sede	1												
Distância chão	0,66	1											
Disp. hídrica	-0,57	-0,51	1										
Quant. Benf.	-0,28	-0,08	0,43	1									
Estado benf.	-0,32	-0,05	0,40	0,73	1								
Índice benf.	-0,35	-0,09	0,47	0,93	0,89	1							
Aptidão	-0,56	-0,54	0,81	0,37	0,32	0,40	1						
Índ. Prod. Veg.	-0,21	0,06	0,21	0,38	0,27	0,39	0,16	1					
Dist. Rodovias	0,66	0,63	-0,60	-0,29	-0,29	-0,32	-0,48	-0,14	1				
Dist. Rio S. Franc.	0,69	0,59	-0,59	-0,13	-0,14	-0,14	-0,62	-0,06	0,50	1			
Dist. aeroporto	0,92	0,53	-0,47	-0,29	-0,35	-0,37	-0,50	-0,28	0,51	0,56	1		
Dist. Petrolina	0,95	0,59	-0,52	-0,30	-0,36	-0,38	-0,54	-0,24	0,56	0,66	0,97	1	
Dist. Merc. Produtor	0,96	0,61	-0,55	-0,28	-0,34	-0,35	-0,58	-0,22	0,55	0,72	0,96	0,99	1

4.5 MODELO CLÁSSICO DE REGRESSÃO LINEAR MÚLTIPLA

Nesta etapa, foram construídos alguns modelos para identificar qual melhor representava a realidade do mercado em estudo. De acordo com Dantas (2005), a transformação logarítmica é a preferida quando se procura ajustar modelos a dados de valores imobiliários. As razões para sua utilização se devem ao fato 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; além disso, a transformação logarítmica na variável explicada torna o modelo multiplicativo, característica sugerida pelas normas brasileiras de avaliações.

A partir desta etapa serão apresentados dois modelos: no Modelo 1 foram utilizadas além das variáveis relacionadas às características intrínsecas dos imóveis as variáveis relacionadas à localização, com isso espera-se modelar a dependência espacial e evitar que ela ocorra nos resíduos do modelo; no Modelo 2 foram utilizadas apenas as variáveis relacionadas às características intrínsecas dos imóveis sem variáveis relacionadas à localização.

Os resultados gerais do Modelo 1 são mostrados na Tabela 17. Este contou com 8 variáveis, o teste de significância global se mostrou altamente significativo. Apresentou um R^2 ajustado de 0,86, ou seja, 86% da variação de preços que ocorre no mercado em estudo pode ser explicada por este modelo.

Houve a transformação logarítmica da variável dependente bem como das variáveis independentes Área Total do Imóvel e Índice de Produção Vegetal, as demais permaneceram nas escalas originais (Tabela 18). Considerando a significância de 1% todos os coeficientes foram significativos, com exceção da variável Desvios E.

Tabela 17 - Resultados Gerais da Regressão do modelo 1.

Discriminação	Valor
Desvio padrão	0,72
Estatística Fc	64,82
Probabilidade associada a Fc	< 2.2e-16
Coefficiente de determinação	0,87
Coefficiente de determinação ajustado	0,86
Observações	84,00

Tabela 18 - Estimativas dos parâmetros, erro-padrão, valor t e o valor p do modelo 1.

	Estimativa	Erro padrão	valor t	valor p	IC(β_i , 95%)	
					2,50%	97,50%
(Intercept)	8,7591	0,5277	16,5993	7,96E-27	7,7079	9,8103
Desvios N	-0,0198	0,0068	-2,9222	0,0046	-0,0333	-0,0063
Desvios E	0,0019	0,0066	0,2887	0,7736	-0,0113	0,0151
Natureza do evento	0,6750	0,22	3,0811	0,0029	0,2386	1,1114
Ln da Área total (ha)	-0,5181	0,0765	-6,7719	2,47E-09	-0,6706	-0,3657
Estado das benfeitorias	0,2497	0,0800	3,1201	0,0026	0,0903	0,4091
Limite conf. com o rio	0,0071	0,0023	3,1598	0,0023	0,0026	0,0116
Disponibilidade hídrica	1,5253	0,2924	5,2168	1,56E-06	0,9428	2,1077
Ln do Índ. Prod. Veg.	0,1582	0,0446	3,5431	0,0007	0,0692	0,2471

Os resultados gerais do Modelo 2 são mostrados na Tabela 19. O modelo contou com 6 variáveis, foi realizada transformação logarítmica da variável dependente bem como das variáveis independentes Área Total do Imóvel, Área Irrigada e do Índice de Produção Vegetal, as demais permaneceram nas suas escalas originais. Este apresentou um R² ajustado de 0,77, ou seja, 77% da variação de preços que ocorre no mercado em estudo pode ser explicada por este modelo.

Tabela 19 - Resultados Gerais da Regressão do modelo 2.

Discriminação	Valor
Desvio padrão	0,91
Estatística Fc	48,03
Probabilidade associada a Fc	< 2.2e-16
Coeficiente de determinação	0,79
Coeficiente de determinação ajustado	0,77
Observações	84,00

Quanto aos coeficientes, todos se mostraram significativos ao nível de significância de 5% (Tabela 20).

Tabela 20 - Estimativas dos parâmetros e erros-padrão, valor t e o valor p do modelo 2.

	Estimativa	Erro padrão	valor t	valor p	IC(β_i , 95%)	
					2,50%	97,50%
(Intercept)	9,3180	0,4096	22,7480	3,93E-35	8,3846	10,1189
Natureza do evento	1,0717	0,2382	4,4989	2,49E-05	0,6168	1,6490
Ln Área total (ha)	-0,8176	0,0725	-11,2809	9,65E-18	-0,9523	-0,6325
Ln Área irrig. (ha)	0,2368	0,1137	2,0831	4,07E-02	-0,0091	0,4954
Limite conf. com o rio	0,0128	0,0030	4,3125	4,91E-05	0,0055	0,0160
Estado benef.	0,3448	0,0968	3,5603	6,52E-04	0,1297	0,5417
Ln Índ. Prod. Veg.	0,2199	0,0583	3,7721	3,24E-04	0,0952	0,3533

Na Tabela 21, são apresentadas as formas funcionais dos dois modelos

juntamente com os resultados de alguns testes estatísticos.

Tabela 21 - Principais modelos construídos com respectivos testes estatísticos.

Modelos	Forma Funcional	Considerações
Modelo 1	$\ln(VTI) = \beta_0 + \beta_1 DN + \beta_2 DE + \beta_3 NE + \beta_4 \ln(AT) + \beta_5 EB + \beta_6 LCR + \beta_7 DH + \beta_8 \ln(IPV) + \varepsilon$	
	Teste de Bonferonni	Obs. 76 é outlier, valor p=0,041171
	Teste de Jarque Bera	valor p=0,3186
	Teste de RESET	valor p= 0,5448
	Teste de Goldfeld-Quandt	valor p=0,0159
	Teste de Koenker	valor p=0,0934
	Teste de Breusch-Pagau	valor p=0,01831
Modelo 2	$\ln(VTI) = \beta_0 + \beta_1 NE + \beta_2 \ln(AT) + \beta_3 \ln(AI) + \beta_4 LCR + \beta_5 EB + \beta_6 \ln(IPV) + \varepsilon$	
	Teste de Bonferonni	Não foram identificados outliers.
	Teste de Jarque Bera	valor p=0,5081
	Teste de RESET	valor p=0,06606
	Teste de Goldfeld-Quandt	valor p=0,05178
	Teste de Koenker	valor p=0,03879
	Teste de Breusch-Pagau	valor p=0,01024

No modelo 1, o teste de Bonferonni para detecção de *outliers* identificou que a observação 76 possui esta característica, posteriormente na análise gráfica tal informação será confirmada. O resultado do Teste de Jarque Bera com valor p igual a 0,3186 indica que ao nível de significância de 5% aceita-se a hipótese nula de que os resíduos seguem uma distribuição normal. O teste RESET com valor p igual a 0,5448 indica que ao nível de significância de 5% os parâmetros do modelo são lineares, e portanto o modelo está bem especificado. Os testes de Goldfeld-Quandt, Koenker e de Breusch-Pagau indicam que os resíduos do modelo são heterocedásticos, ou seja apresentam problemas de variância dos erros não constante.

Com relação ao Modelo 2 o teste de Bonferonni para detecção de *outliers* não identificou nenhum ponto com esta característica. O resultado do Teste de Jarque Bera com valor p igual a 0,07169 indica que ao nível de significância de 5% aceita-se a hipótese nula de que os resíduos seguem uma distribuição normal. O teste

RESET com valor p igual a 0,06606 indica que ao nível de significância de 5% os parâmetros do modelo não são lineares, havendo então indícios de erros de especificação do modelo. Os testes de Goldfeld-Quandt e de Breusch-Pagau indicam que os resíduos do modelo são heterocedásticos, ou seja, apresentam problemas de variância dos erros não constante.

4.5.1 Análise da Multicolinearidade

A inclusão de variáveis com alta correlação no modelo pode gerar problemas na estimação. Quando isso ocorre a matriz $X'X$ usada na fórmula do estimador de mínimos quadrados pode ficar próxima de singular. Em consequência disso, ocorre um significativo aumento da variância associada ao estimador, o que se denomina de “inflação da variância”. Os Fatores de Inflação da Variância buscam quantificar a intensidade deste problema. Quando seus valores são iguais a 1 não há inflação, até 5 não há problemas, entre 5 e 10 são preocupantes e acima de 10 são problemáticos.

A Tabela 22 apresenta os Fatores de Inflação da Variância para os modelos 1 e 2. Observa-se que em ambos os modelos o FIV encontra-se abaixo de 5, indicando não haver problemas de multicolinearidade nos mesmos. Isto reflete o cuidado de não inserir nos modelos variáveis que na matriz de correlações apresentavam alta correlação.

Tabela 22 - Fatores de Inflação da Variância para os modelos 1 e 2.

Variáveis	Fator de Inflação da Variância	
	Modelo 1	Modelo 2
Desvios N	2,29	-
Desvios E	1,19	-
Natureza do evento	1,68	1,45
Ln da Área total (ha)	1,69	1,15
Estado das benfeitorias	1,37	1,41
Limite conf. com o rio	1,54	1,31
Disponibilidade hídrica	2,26	-
Ln do Índ. Prod. Veg.	1,57	2,03
Ln da Área irrig. (ha)	-	2,32

4.5.2 Análise Gráfica dos Resíduos

Nesta etapa foi realizada a construção de gráficos a partir dos resíduos dos modelos 1 e 2 com o objetivo de verificar o atendimento aos pressupostos do modelo clássico de regressão.

Na Figura 19 são mostrados o histograma e o *boxplot*. Neles é possível perceber que os resíduos padronizados externamente estão contidos fora do intervalo -2 a 2, o *boxplot* mostra que existem dois pontos que ultrapassam o limite de 1,5 vezes a distância interquartílica são eles os pontos 62 (*outlier* superior) e 76 (*outlier* inferior). Vale ressaltar que o Teste de Bonferroni para detecção de *outliers* identificou apenas o ponto 76 com estas características entretanto, pode se tratar de um ponto influenciante, o que será analisado na próxima etapa.

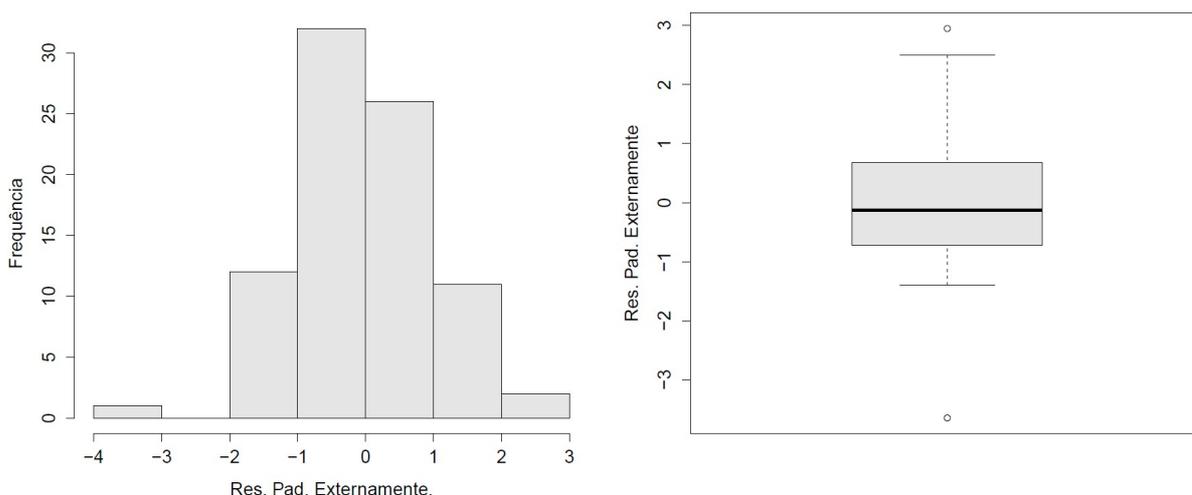


Figura 19 - Histograma e Boxplot para os resíduos do modelo 1

Na Figura 20 são mostrados o gráfico de probabilidade normal dos resíduos e o diagrama dos resíduos padronizados externamente versus os valores preditos. No primeiro percebe-se que os resíduos estão regularmente contidos dentro do intervalo delimitado pelas linhas tracejadas e próximos da linha de inclinação igual a 1 que define a distribuição normal, confirmando o que foi verificado por meio do Teste de Jarque Bera no qual os resíduos apresentaram distribuição normal.

O segundo gráfico também vem a confirmar os resultados dos testes estatísticos realizados anteriormente pois os Testes de Goldfeld-Quandt, Koenker e

de Breusch-Pagau apontaram problemas de heterocedasticidade, o que se confirma pela distribuição dos pontos no referido gráfico, onde a observação 76 está dispersa do conjunto de pontos. Além disso, nos extremos (faixa de 5 a 8 e de 11 a 12) os resíduos se comportam com uma maior dispersão que quando comparados com a região intermediária (faixa de 8 a 10), na qual apresenta uma menor dispersão.

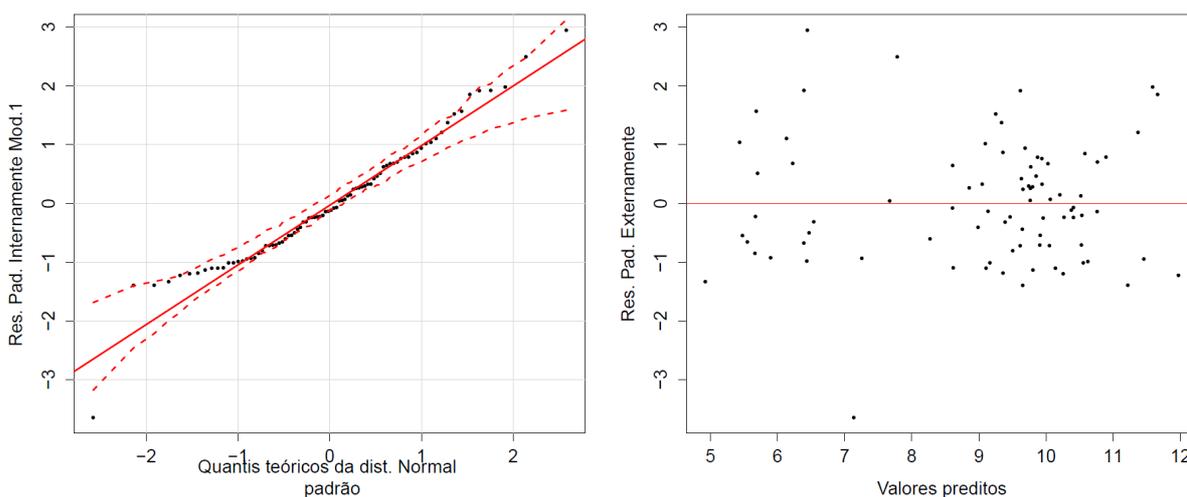


Figura 20 - Gráfico de probabilidade normal dos resíduos e diagrama dos resíduos padronizados externamente versus os valores preditos.

Na Figura 21 é mostrado um gráfico bastante útil para identificar observações influentes, conhecido como gráfico de pontos de influência. Este mostra simultaneamente três parâmetros utilizados para identificar este tipo de observação: os resíduos padronizados externamente (eixo y), os pontos de alavanca (eixo x) e a Distância de Cook representada pelo tamanho de cada circunferência. Linhas de referência verticais são desenhadas em duas e três vezes o valor médio dos elementos da diagonal principal h_{ii} da matriz H; linhas de referência horizontais são desenhadas em -2, 0 e 2 na escala dos resíduos padronizados externamente. Nele destacam-se o ponto 76, o qual possui uma Distância de Cook elevada, com resíduo padronizado externamente também elevado; bem como o ponto 89, que possui um valor do ponto de alavanca elevado.

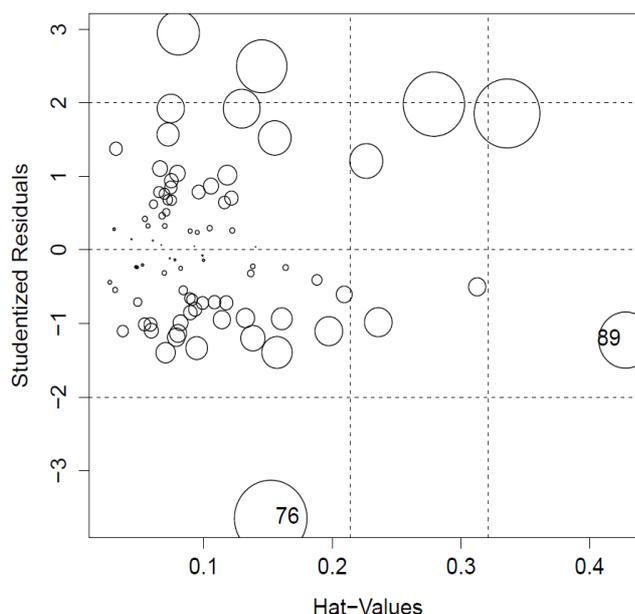


Figura 21 - Gráficos de pontos de influência dos resíduos para o Modelo 1.

A análise gráfica mostrou que as observações 62, 76 e 89 podem influenciar a estimação dos coeficientes. Sendo assim, buscou-se quantificar o impacto que as observações *outliers*, pontos de influência ou ainda de alavanca, provocariam sobre a estimativa dos coeficientes, bem como sobre o R^2 ajustado no modelo 1. Para isso, foi construída a Tabela 23 a qual mostra a variação percentual provocada pela retirada das supostas observações influentes na estimativa dos coeficientes e do R^2 .

Tabela 23 - Variação percentual das estimativas dos coeficientes de regressão e valor do R^2 ajustado de acordo com a retirada de observações atípicas.

Obs. Retirada	β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8	R^2 ajust.
62	-0,71	9,87	-0,63	1,08	-4,65	-6,92	-4,42	8,83	-1,74	1,67
76	6,56	35,57	-40,07	-6,50	-1,71	-18,55	15,81	-6,86	-6,79	3,31
89	0,89	-1,55	-91,13	-5,20	1,55	2,86	27,44	0,93	-4,50	0,02
62 e 76	5,73	37,17	47,03	-1,54	-3,20	-21,35	-10,41	-7,76	-2,51	3,25
62 e 89	0,16	8,32	-89,48	-4,00	-3,12	-4,10	22,35	9,70	-6,11	1,70
76 e 89	7,46	27,05	-40,06	-7,73	2,78	-12,39	20,45	-15,74	-5,28	1,69
62, 76 e 89	6,56	35,57	-40,07	-6,50	-1,71	-18,55	15,81	-6,86	-6,79	3,31

Percebe-se que as observações que provocam maior influência individualmente sobre os coeficientes do modelo bem como sobre o R^2 são as observações 62 e 76. A retirada das duas observações conjuntamente é a que provoca maior influência sobre os coeficientes. Sendo assim optou-se por retirar as observações 62 e 76 do modelo gerando assim o modelo 1.1 que será apresentado

adiante.

Na Figura 22 são apresentados o histograma e o *Boxplot* para os resíduos do modelo 2. Neles pode-se observar que os resíduos padronizados externamente estão fora do intervalo entre -2 e 2. No *Boxplot* pode-se observar que dois valores são considerados *outliers*, são eles as observações 35 (*outlier superior*) e 76 (*outlier inferior*). O teste de Bonferroni também não identificou *outliers* nos resíduos do modelo 2.

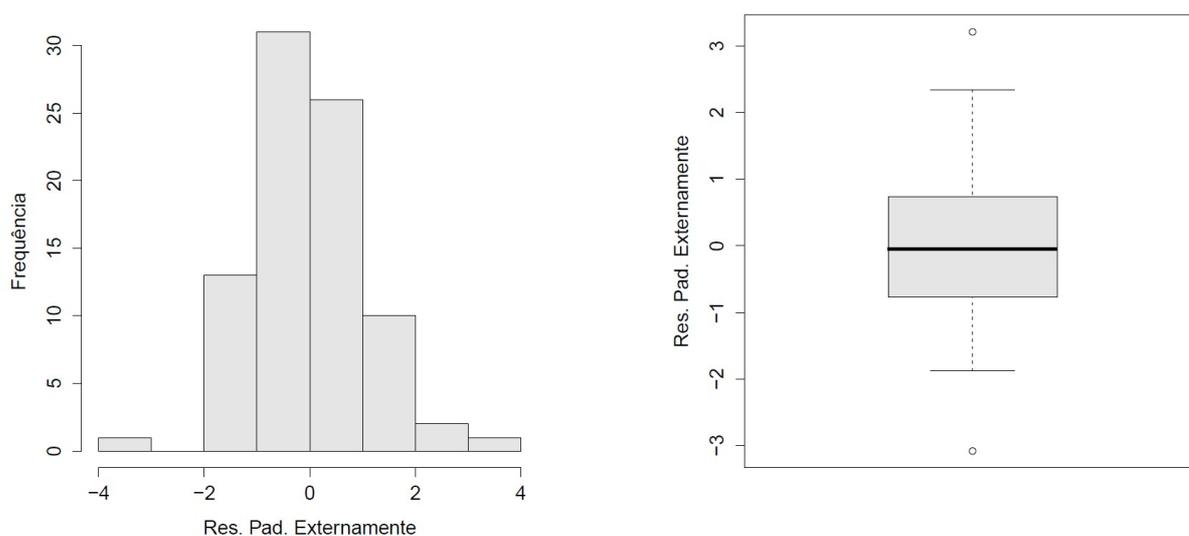


Figura 22 - Histograma e Boxplot para os resíduos do modelo 2.

A seguir são apresentados o gráfico de probabilidade normal e o gráfico dos valores preditos versus os resíduos padronizados externamente Figura 23. No primeiro, o resultado do teste de Normalidade de Jarque Bera se confirma, pois os dados estão dentro dos limites da distribuição normal. No segundo gráfico, pode-se visualizar que também ocorrem problemas de variância não contante para os resíduos, o que já havia sido indicado pelos testes de Goldfeld-Quandt, Koenker e de Breusch-Pagau. Além disso, chama a atenção dois pontos afastados do conjunto de pontos, são as observação 35 e 76 citadas anteriormente.

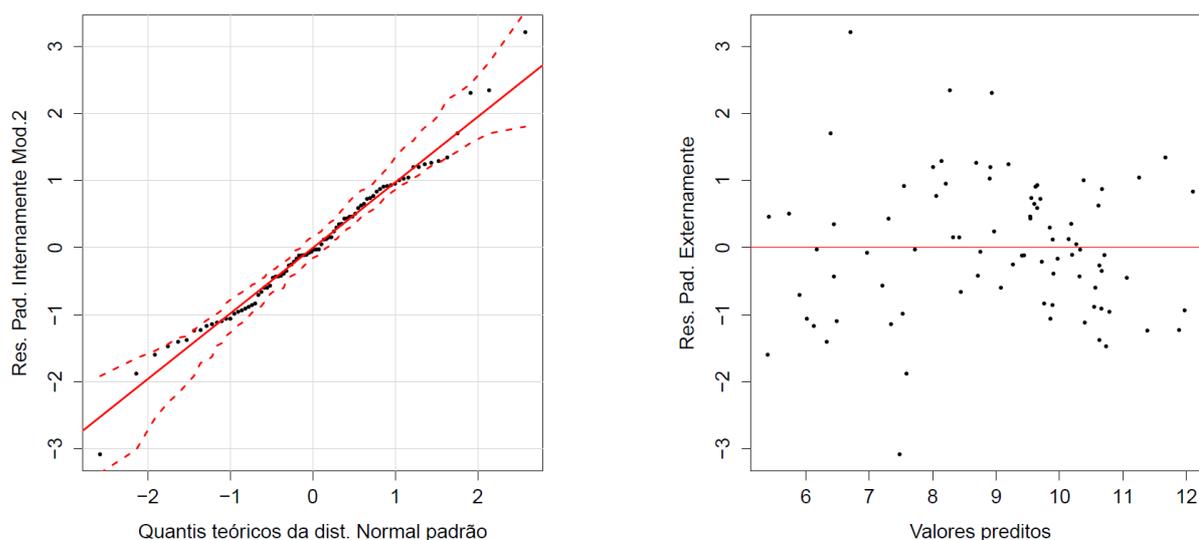


Figura 23 - Gráfico de probabilidade normal dos resíduos e diagrama dos resíduos padronizados externamente versus os valores preditos para o modelo 2.

No gráfico dos pontos de influência (Figura 24), três observações foram identificadas com estas características, são elas os pontos 35, 76 e 89. A observação 35 apresenta uma distância de Cook elevada e também valores do resíduo padronizado internamente elevados. A observação 76 apresenta Distância de Cook elevada e resíduos padronizados internamente fora do limite -2. A observação 89 apresentou valor do ponto de alavanca elevados, bem como a Distância de Cook.

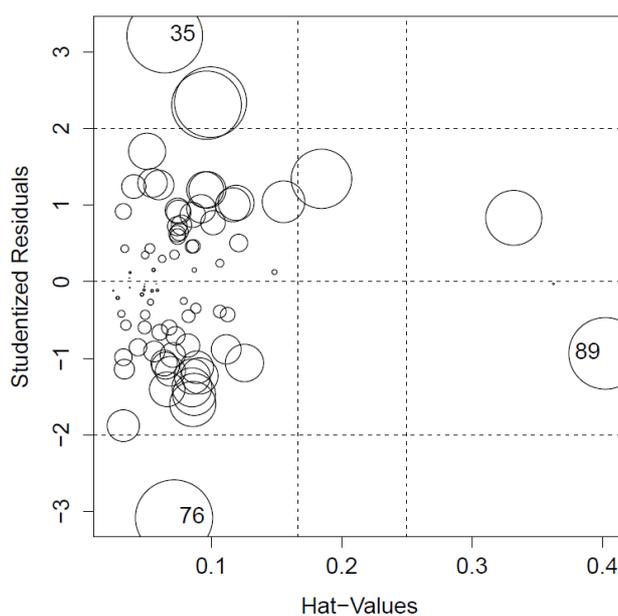


Figura 24 - Gráficos de pontos de influência dos resíduos para o Modelo 2.

Analogamente ao que foi realizado no modelo 1, buscou-se quantificar o impacto que a retirada de observações influenciadoras do modelo provoca na estimação dos coeficientes do mesmo (Tabela 24). Neste caso, percebe-se que as 3 observações (35, 76 e 89) provocam mudanças significativas tanto individualmente quanto em conjunto. Por isso, foram retiradas as 3 observações do modelo gerando assim o modelo 2.1, que será apresentado a seguir.

Tabela 24 - Variação percentual das estimativas dos coeficientes de regressão e valor do R² ajustado de acordo com a retirada de observações atípicas.

Obs. Retirada	β_0	β_1	β_2	β_3	β_4	β_5	β_6	R ² ajust.
35	-2,61	5,86	0,05	0,00	-1,37	17,58	0,38	3,47
76	3,38	-7,70	2,60	-0,57	2,59	-17,88	-0,07	1,69
89	0,03	-3,68	0,70	-2,03	17,23	2,74	-2,17	-0,22
35 e 76	0,70	-1,68	2,47	-0,53	1,13	-0,19	0,29	4,92
35 e 89	-2,60	2,08	0,77	-2,10	16,44	20,46	-1,86	3,35
76 e 89	3,40	-11,32	3,29	-2,57	19,61	-15,15	-2,21	1,51
35, 76 e 89	0,72	-5,40	3,18	-2,60	18,71	2,69	-1,92	4,82

4.5.3 Modelos Após a Retirada das Observações Atípicas

Na Tabela 25 são mostrados os resultados gerais do modelo 1.1. De acordo com o teste de significância Global do Modelo este é altamente significativo. Possui um coeficiente de determinação ajustado de 0,89 o que significa dizer que o modelo consegue explicar 89% de toda a variação ocorrida no mercado em estudo.

Quanto aos coeficientes, percebe-se que após a retirada dos pontos influenciadores todas as variáveis continuaram a ser significativas (Tabela 26).

Tabela 25 - Resultados Gerais da Regressão do modelo 1.1

Discriminação	Valor
Desvio padrão	0,63
Estatística Fc	81,36
Probabilidade associada a Fc	< 2.2e-16
Coeficiente de determinação	0,90
Coeficiente de determinação ajustado	0,89
Observações	82,00

Tabela 26 - Estimativas dos parâmetros, erros-padrão, valor t e valor p do modelo 1.1

	Estimativa	Erro padrão	valor t	valor p	IC(β_j ,95%)	
					2,50%	97,50%
(Intercept)	9,2609	0,4883	18,9657	6,28E-30	8,2877	10,2341
Desvios N	-0,0271	0,0062	-4,4085	3,51E-05	-0,0394	-0,0149
Desvios E	0,0028	0,0058	0,4825	6,31E-01	-0,0088	0,0145
Natureza do evento	0,6646	0,1926	3,4499	9,00E-04	0,2807	1,0485
Ln da Área total (ha)	-0,5015	0,0677	-7,4074	1,81E-10	-0,6365	-0,3666
Estado das benfeitorias	0,1964	0,0712	2,7578	7,30E-03	0,0545	0,3383
Limite conf. com o rio	0,0064	0,0020	3,2110	2,00E-03	0,0024	0,0104
Disponibilidade hídrica	1,4070	0,2692	5,2273	1,57E-06	0,8705	1,9434
Ln do Índ. Prod. Veg.	0,1542	0,0392	3,9285	2,00E-04	0,0760	0,2324

Na Tabela 27 são mostrados os resultados gerais do modelo 2.1. De acordo com o teste de significância Global do Modelo este é altamente significativo. O modelo possui um coeficiente de determinação ajustado de 0,81 o que significa que o modelo consegue explicar 81% de toda a variação ocorrida no mercado em estudo.

Quanto aos coeficientes do modelo percebe-se que após a retirada dos pontos influenciantes todas as variáveis continuam a ser significativas (Tabela 28).

Tabela 27 - Resultados Gerais da Regressão do modelo 2.1

Discriminação	Valor
Desvio padrão	0,82
Estatística Fc	57,82
Probabilidade associada a Fc	< 2.2e-16
Coeficiente de determinação	0,82
Coeficiente de determinação ajustado	0,81
Observações	81,00

Tabela 28 - Estimativas dos parâmetros, erros-padrão, valor t e valor p do modelo 2.1

	Estimativa	Erro padrão	valor t	valor p	IC(β_j ,95%)	
					2,50%	97,50%
(Intercept)	9,3180	0,4096	22,7480	3,93E-35	8,5018	10,1342
Natureza do evento	1,0717	0,2382	4,4989	2,49E-05	0,5970	1,5463
Ln Área total (ha)	-0,8176	0,0725	-11,2809	9,65E-18	-0,9620	-0,6732
Ln Área irrig. (ha)	0,2368	0,1137	2,0831	4,07E-02	0,0103	0,4633
Limite conf. com o rio	0,0128	0,0030	4,3125	4,91E-05	0,0069	0,0187
Estado benf.	0,3448	0,0968	3,5603	6,52E-04	0,1518	0,5377
Ln Índ. Prod. Veg.	0,2199	0,0583	3,7721	3,24E-04	0,1037	0,3361

A Tabela 29 apresenta novamente os modelos e os resultados dos principais

testes realizados após a retirada das observações influenciadas, sendo a partir de então chamados de Modelo 1.1 e Modelo 2.1. Ambos continuaram apresentando resíduos com distribuição normal. O teste RESET indica que o Modelo 1.1 não apresenta problemas de especificação enquanto que no Modelo 2 o valor $p=0,0197$ indica que há problemas na especificação do modelo. Provavelmente a falta de variáveis relacionadas à localização faz com que o modelo apresente estes resultados. Os Testes para detecção de problemas de Heterocedasticidade (Goldfeld-Quandt, Koenker e Breusch-Pagau) indicaram que o Modelo 1.1 é homocedástico enquanto que no Modelo 2.1 apenas o Teste de Breusch-Pagau indicou haver homocedasticidade contrariando os demais que indicaram que o modelo é heterocedástico.

A heterocedasticidade observada no Modelo 2.1 pode ser provocada pela presença de dependência espacial nos resíduos do modelo. A análise geoestatística por meio do ajuste de um semivariograma será capaz confirmar esta suposição.

Tabela 29 - Principais modelos construídos com respectivos testes estatísticos.

Modelos	Forma Funcional	Considerações	
Modelo 1.1	$\ln(VTI) = \beta_0 + \beta_1 DN + \beta_2 DE + \beta_3 NE + \beta_4 \ln(AT) + \beta_5 EB + \beta_6 CMR + \beta_7 DH + \beta_8 \ln(IPV) + \varepsilon$	Teste de Bonferonni	Não foram identificados <i>outliers</i> .
		Teste de Jarque Bera	valor $p=0,2033$
		Teste de RESET	valor $p= 0,4857$
		Teste de Goldfeld-Quandt	valor $p=0,1783$
		Teste de Koenker	valor $p=0,2344$
		Teste de Breusch-Pagau	valor $p=0,4605$
		Modelo 2.1	$\ln(VTI) = \beta_0 + \beta_1 NE + \beta_2 \ln(AT) + \beta_3 \ln(AI) + \beta_4 CMR + \beta_5 EB + \beta_6 \ln(IPV) + \varepsilon$
Teste de Jarque Bera	valor $p=0,5682$		
Teste de RESET	valor $p=0,0197$		
Teste de Goldfeld-Quandt	valor $p=0,00278$		
Teste de Koenker	valor $p=0,01938$		
Teste de Breusch-Pagau	valor $p=0,06691$		

4.6 ANÁLISE GEOESTATÍSTICA DOS MODELOS

Os resíduos dos modelos 1.1 e 2.1 foram espacializados e classificados em quartis, o que pode ser visto na Figura 25. Nota-se que no modelo 1.1 não há concentração dos resíduos maiores ou menores em determinadas regiões, ou seja, os resíduos mostram estar livres de tendência, é isso que se espera dos resíduos oriundos de um modelo de regressão. Entretanto, no modelo 2.1 percebe-se que os resíduos maiores estão concentrados à sudoeste da área de estudos, o que não é desejável quando se trata de modelos de regressão linear.

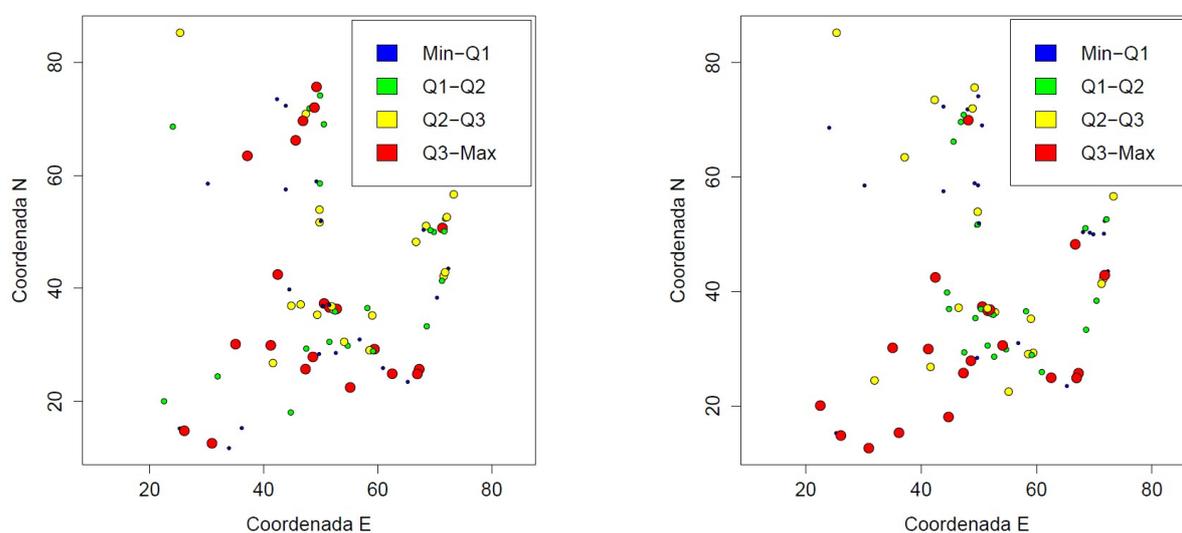


Figura 25 - Distribuição espacial dos resíduos do modelo 1.1 (esquerda) e 2.1 (direita), divididos em quartis.

Para verificar se há tendência nos resíduos buscou-se ajustar uma superfície de 1º grau aos resíduos do modelo 1.1 e 2.1. O teste que avalia a significância global do modelo se mostrou não significativo para o Modelo 1.1 (Tabela 30) enquanto que para o Modelo 2.1 foi altamente significativo e o R^2 ajustado foi de 0,15 (Tabela 32) confirmando o que foi mostrado na Figura 25(direita). Na Figura 26 é possível visualizar a superfície de resposta de 1º Grau ajustada aos resíduos do modelo 2.1. Na Tabela 33 são mostrados os coeficientes da superfície de 1º grau ajustada aos resíduos do Modelo 2.1 onde pode-se confirmar que a coordenada Norte é altamente significativa. Esta constatação é importante pois na próxima etapa, quando do cálculo de semivariogramas experimentais aos resíduos do modelo a tendência deverá ser removida.

Tabela 30 - Ajuste de uma superfície de tendência de 1º grau aos resíduos do modelo 1.1.

Discriminação	Valor
Desvio padrão	0,58
Estatística Fc	4,25E-08
Probabilidade associada a Fc	1,00
Coefficiente de determinação	1,07E-009
Coefficiente de determinação ajustado	-2,53E-02
Observações	82,000

Tabela 31 - Coeficientes da superfície de tendência de 1º grau aos resíduos do modelo 1.1.

	Coeficientes	Desvio padrão	Estatística t	Nível de significância
(Intercept)	-0,0069	33,2083	-0,0002	0,9998
Coordenada Leste	0,0000	0,0000	0,0002	0,9998
Coordenada Norte	0,0000	0,0000	0,0002	0,9998

Tabela 32 - Ajuste de uma superfície de tendência de 1º grau aos resíduos do modelo 2.1.

Discriminação	Valor
Desvio padrão	0,7226
Estatística Fc	8,3854
Probabilidade associada a Fc	5,03E-04
Coefficiente de determinação	0,1770
Coefficiente de determinação ajustado	0,1559
Observações	81,0000

Tabela 33 - Coeficientes da superfície de tendência de 1º grau ajustada aos resíduos do modelo 2.1.

	Coeficientes	Desvio padrão	Estatística t	Nível de significância
(Intercept)	167,3268	41,5808	4,0241	0,0001
Coordenada Leste	0,0000	0,0000	-0,9818	0,3292
Coordenada Norte	0,0000	0,0000	-3,9800	0,0002

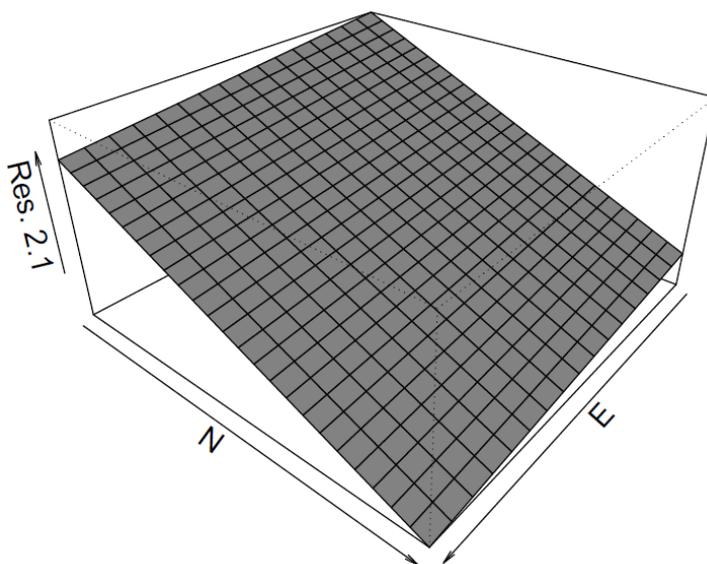


Figura 26 - Superfície de 1º Grau ajustada aos resíduos do modelo 2.1.

O próximo passo da análise é verificar a existência de dependência espacial nos resíduos do modelo. Para isto foi utilizado o semivariograma como ferramenta exploratória.

Na Figura 27 pode-se visualizar os semivariogramas experimentais para os modelos 1.1 e 2.1. Estes foram construídos considerando 60% da máxima distância entre os pontos, que correspondeu a 44km, com a exigência de que cada ponto fosse calculado a partir de no mínimo 32 pares de pontos.

Com relação ao semivariograma do modelo 1.1 pode-se afirmar que os resíduos não possuem dependência espacial. Com relação ao modelo 2.1, a semivariância é mínima próximo à distância 14km, a partir de onde ela se torna crescente até um ponto no qual ela se estabiliza que é conhecido como patamar. A partir de agora o trabalho será continuado com o modelo 2.1 objetivando realizar a modelagem da dependência espacial nos resíduos. O próximo passo será ajustar um modelo teórico ao semivariograma experimental de modo que seja possível definir o alcance do semivariograma que será utilizado nas etapas subsequentes.

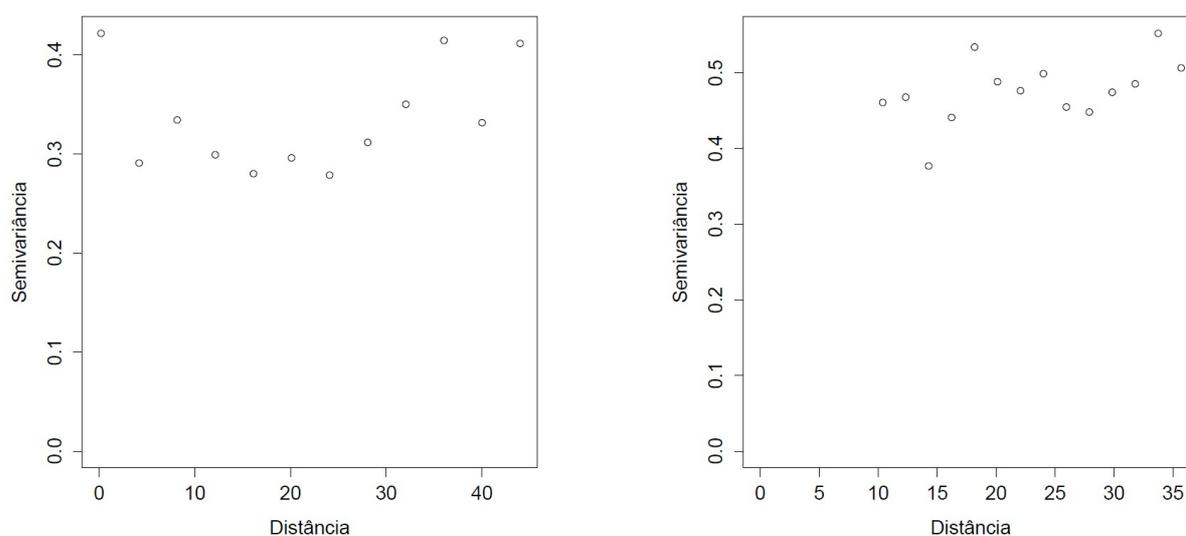


Figura 27 - Semivariogramas experimentais dos resíduos do modelo 1.1 (esquerda) e 2.1 (direita), a distância é dada em quilômetros.

Na Tabela 34 são mostrados os parâmetros resultantes do cálculo do semivariograma experimental para o modelo 2.1. Foram calculados 14 pontos distanciados de 1,9km e pode-se perceber que a semivariância é crescente até determinado ponto, quando então ela se estabiliza, o que caracteriza o patamar. Vale salientar que o mínimo de pares de pontos com os quais cada ponto do semivariograma foi calculado foi de 110 pares, bem acima do mínimo recomendado.

Tabela 34 - Parâmetros do semivariograma experimental.

	Distância (km)	Semivariância	Autocorrelação	Núm. de pares.
1	10,500	0,460	0,621	572
2	12,438	0,455	0,649	127
3	14,377	0,394	0,540	128
4	16,315	0,437	0,638	102
5	18,254	0,531	0,728	155
6	20,192	0,494	0,701	149
7	22,131	0,486	0,703	165
8	24,069	0,489	0,699	172
9	26,008	0,479	0,666	164
10	27,946	0,419	0,586	133
11	29,885	0,477	0,673	177
12	31,823	0,486	0,653	135
13	33,762	0,552	0,759	141
14	35,700	0,507	0,607	110

Após a obtenção do semivariograma experimental foi possível realizar o ajuste de modelos teóricos ao mesmo. A Tabela 35 mostra os parâmetros obtidos para os modelos Esférico e Exponencial.

O maior alcance ocorreu com o modelo Exponencial e o menor alcance aconteceu com o modelo Esférico. Para avaliar o grau de dependência espacial, foi

utilizada o Efeito Pepita Relativo $EPR = \left(\frac{\varphi_1}{\varphi_1 + \varphi_2} \right)$, proposto por CAMBARDELLA *et al.* (1994), em que: φ_1 é o valor da variância quando h tende a zero (efeito pepita) e $(\varphi_1 + \varphi_2)$ é a maior variância (patamar). Valores da relação inferiores a 25% caracterizam forte dependência espacial, entre 25 e 75% moderada e acima de 75%, fraca dependência espacial.

Para os resíduos do modelo 2.1 em todos os modelos teóricos ficou caracterizada moderada dependência espacial. O modelo que apresentou o menor alcance apresentou a menor dependência espacial, enquanto que o modelo que apresentou o maior alcance apresentou a maior dependência espacial.

Tabela 35 - Parâmetros dos semivariogramas teóricos ajustados pelo método dos mínimos quadrados aos resíduos do modelo 2.1.

Modelo	φ_1	φ_2	$\varphi_1 + \varphi_2$	$\varphi_3 = g(a)$	a	EPR
Esférico	0,3556	0,1328	0,4883	25,0723	25,0723	0,7282
Exponencial	0,3400	0,1651	0,5051	11,8817	35,5945	0,6731

φ_1 : efeito pepita; φ_2 : contribuição; $\varphi_3 = g(a)$: função do alcance; **a**: alcance (km) ; $EPR = \varphi_1 / (\varphi_1 + \varphi_2)$: efeito pepita relativo.

A proximidade entre os valores dos parâmetros fornecidos pela Tabela 35 não permite a tomada de decisão sobre qual o melhor modelo teórico. A técnica da validação cruzada é bastante utilizada por oferecer critérios para a escolha do melhor modelo ajustado. Segundo Faraco (2008), o melhor modelo ajustado será aquele que apresentar os menores Erro Médio e Erro Médio Reduzido, Desvio padrão do erro médio menor possível, Desvio padrão do erro reduzido mais próximo de 1 e Erro Absoluto menor possível.

Na Tabela 36 o modelo que atende a todos os critérios supracitados é o modelo Esférico, o qual pode ser visualizado na Figura 28.

Tabela 36 - Parâmetros para escolha dos modelos ajustados.

Modelos	Erro Médio	Erro Médio Reduzido	Desvio padrão do erro médio	Desvio padrão do erro reduzido	Erro absoluto
Esférico	0,0055	0,7799	0,0038	1,1793	51,3339
Exponencial	0,0053	0,7834	0,0036	1,1878	51,4113

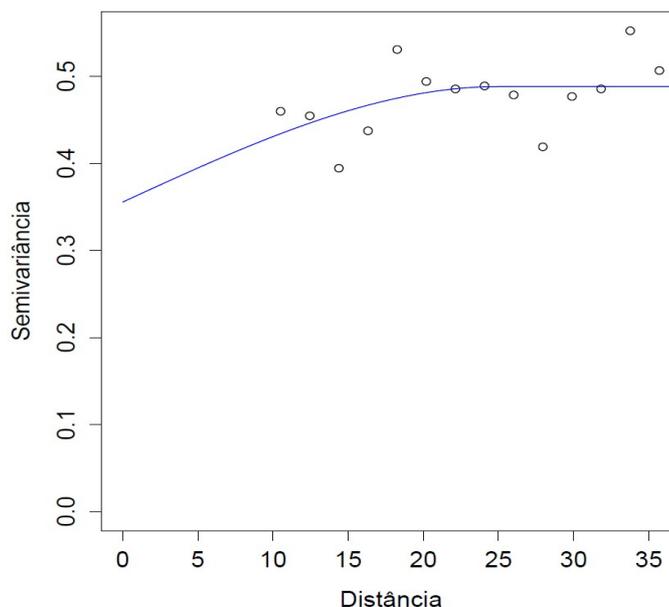


Figura 28 - Semivariograma teórico, modelo esférico, ajustado aos resíduos do modelo 2.1, a distância é dada em quilômetros.

Os parâmetros associados ao melhor modelo Exponencial por MMQ são os seguintes:

Efeito pepita ($\varphi_1 = 0,35$): Representa as variações locais de pequena escala, como erros de amostragem, é o valor de semivariância para distâncias menores que a mínima distância amostrada.

Contribuição ($\varphi_2 = 0,16$): Representa as diferenças espaciais entre os valores de uma variável tomada em dois pontos separados por distâncias cada vez maiores.

Patamar ($\varphi_1 + \varphi_2 = 0,50$): Para os resíduos do Modelo 2.1 na área em estudo, os valores do semivariograma experimental, estabilizam-se a partir desse valor (0,50). O patamar é uma estimativa a variância dos dados, se calcularmos a variância dos resíduos encontra-se $S^2 = 0,61$, percebe-se que os valores são próximos.

Alcance ($a = 25$): Indica que a partir dessa distância (25,07 quilômetros), um ponto

amostral não apresenta mais influência sobre o ponto vizinho.

Efeito pepita relativo (EPR= 0,72): Indica que os resíduos do modelo 2.1 na área em estudo apresenta moderada dependência espacial, segundo a classificação de Cambardella et al. (1994).

Com a determinação do modelo teórico que obteve o melhor ajuste, é possível utilizar a informação fornecida pelo alcance do semivariograma na construção das matrizes de pesos espaciais e assim pode-se iniciar a análise econométrica espacial.

4.7 ANÁLISE ECONOMETRICA ESPACIAL

O primeiro passo para iniciar a análise econométrica espacial é construir uma representação matemática que consiga melhor representar o arranjo espacial dos dados na região, isto pode ser conseguido com a matriz de pesos espaciais. O objetivo é que esta matriz capture a autocorrelação espacial presente nos resíduos do modelo de regressão.

Diferentes estratégias podem ser adotadas, como matrizes baseadas na vizinhança ou então numa distância pré-determinada, exemplos podem ser vistos na Figura 29. Uma das dificuldades na criação das matrizes é a definição de qual a melhor distância para capturar a maior parte da autocorrelação espacial. Neste trabalho, uma das vantagens de integrar a geoestatística com a econometria espacial se baseia em utilizar 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.

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. Por isso, recomenda-se testar várias distâncias bem como diferentes estratégias de vizinhança (k vizinhos por exemplo) para ter certeza de qual delas é capaz de capturar a maior parte da autocorrelação espacial.

Sendo assim, foram criadas algumas matrizes de pesos espaciais com base

nas duas estratégias supracitadas. Após a criação das matrizes foi realizado o Teste I de Moran Global para os valores dos imóveis na região de estudo. A matriz que apresentasse o maior valor para o I de Moran Global seria a que conseguiu capturar a maior parte de autocorrelação espacial presente nos dados.

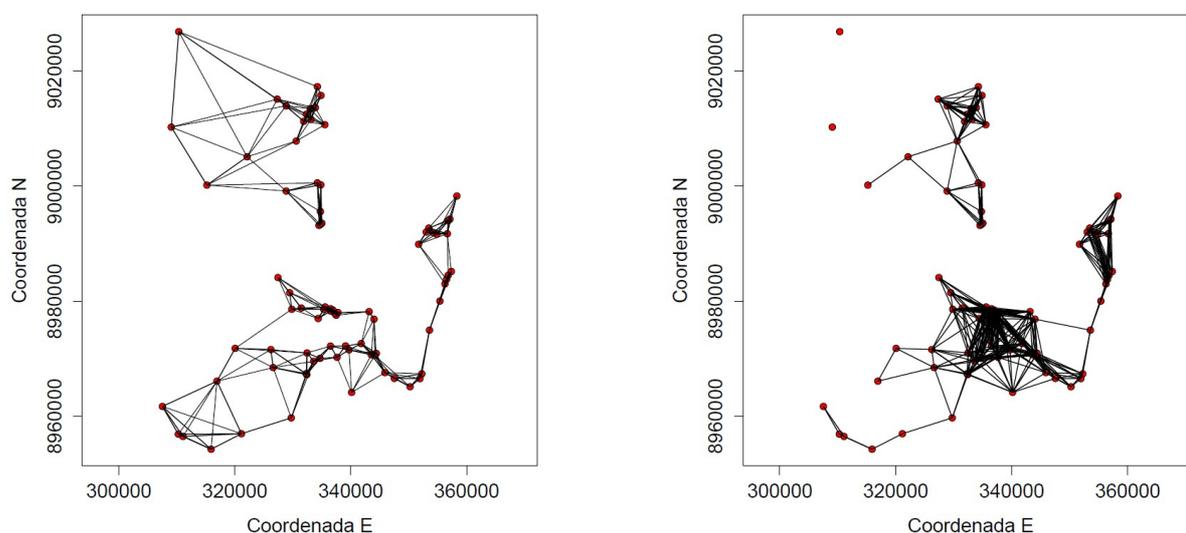


Figura 29 - Exemplos de arranjos espaciais de matrizes de pesos baseadas nos 5 vizinhos mais próximos (esquerda) e na distância de 10 km (direita).

Na Tabela 37, são apresentados os resultados do teste I de Moran Global para as diferentes configurações de matrizes de pesos testadas. Para as matrizes baseadas na vizinhança todos os testes se mostraram significativos a 10% de significância. Entretanto, para as matrizes baseadas na distância os resultados significativos só ocorreram a partir da distância de 10km. Isto se deve ao fato de que para distâncias curtas 1, 2 e 5km ocorreu a formação de pontos que não possuem vizinhança alguma (ilhas) o que provoca o baixo desempenho do teste. Vale salientar que no arranjo baseado na vizinhança não ocorre a formação de ilhas. Foram incluídas as distâncias definidas pelo alcance dos dois modelos teóricos de semivariogramas ajustados na Tabela 37, apesar da validação cruzada ter apontado que o modelo com o melhor ajuste foi o modelo esférico.

As matrizes que conseguiram capturar a maior parte da autocorrelação espacial presente nos dados foram as matrizes baseadas na vizinhança. Destas, as que obtiveram o maior I de Moran foram as baseadas em 1 e 2 vizinhos mais próximos, seguida pela de 10km e na quarta colocação está a matriz baseada na vizinhança com 3 vizinhos. No geral, as matrizes que utilizaram a distância indicada

pelo alcance do semivariograma capturaram pouca autocorrelação sendo que as distâncias indicadas pelos modelos esférico (25km) e exponencial (35km) foram as que conseguiram capturar o menor valor da autocorrelação.

Tabela 37 - Teste I de Moran e significância para as diversas matrizes de ponderação espacial testadas para o VTI/ha.

Matrizes	I de Moran	valor p
Baseadas em k vizinhos		
1	0,233	7,66E-002
2	0,223	1,73E-002
3	0,165	2,75E-002
4	0,119	6,08E-002
5	0,114	4,10E-002
8	0,120	6,42E-003
10	0,118	2,21E-003
15	0,110	2,36E-004
Baseadas na distância		
1km	0,030	7,68E-001
2km	0,040	6,99E-001
5km	0,069	3,49E-001
10km	0,182	2,38E-004
15km	0,118	4,74E-004
20km	0,089	1,28E-003
24km	0,076	1,02E-004
25km	0,071	1,01E-004
35km	0,057	4,54E-009

Na Figura 30 é mostrado o arranjo espacial da matriz de 2 vizinhos mais próximos e o Boxplot da distância entre os vizinhos. Tomando como base a matriz de 2 vizinhos mais próximos, a qual obteve o segundo melhor desempenho, nela a distância mínima entre vizinhos foi de 193m e a máxima foi de 8,1km, sendo que a mediana das distâncias foi 1,9km. Isto significa que a maior dependência espacial entre vizinhos ocorre em distâncias curtas, entre 800 e 3,7km (a caixa do Boxplot). Sendo assim, a distância indicada pelos semivariogramas não está adequada para capturar a autocorrelação na região em análise. Utilizando a distância de 35km por exemplo seria como afirmar que imóveis da região de sequeiro seriam também vizinhos de imóveis localizados em projetos de irrigação, o que não faz sentido na realidade do mercado em estudo.

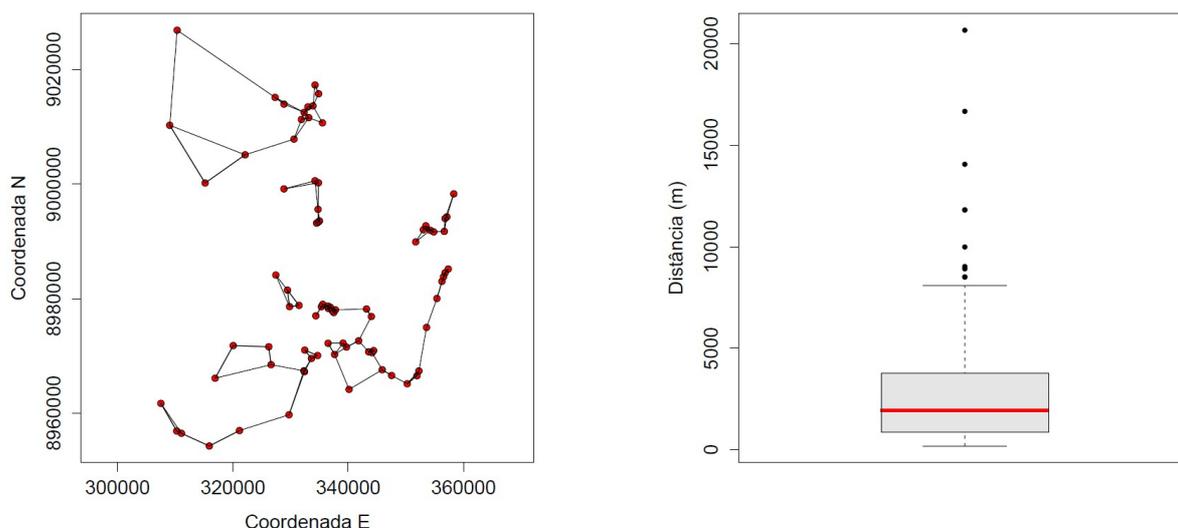


Figura 30 - Arranjo espacial da matriz baseada em 2 vizinhos mais próximos e boxplot das distâncias entre os vizinhos desta matriz.

Na Figura 31 são apresentados o diagrama de dispersão de Moran e Mapa de Clusters fornecido pelo Indicador Local de Autocorrelação Espacial-LISA. Neles pode-se perceber a existência outliers espaciais de alto valor (HL) (observações 82 e 103) e de baixo valor (LH) (observações 75 e 78).

Também percebe-se claramente a existência de um Cluster de alto valor (HH) que pode ser considerado como um ponto de alavanca ou seja, um ponto que é capaz de influenciar fortemente a estimativa dos coeficientes (observação 100); basta observar como a reta de regressão de Wz contra z , a qual fornece o valor do coeficiente I de Moran, é atraída para referida observação.

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 em virtude que como mostrado na análise dos resíduos, tais observações não foram identificadas com as referidas características.

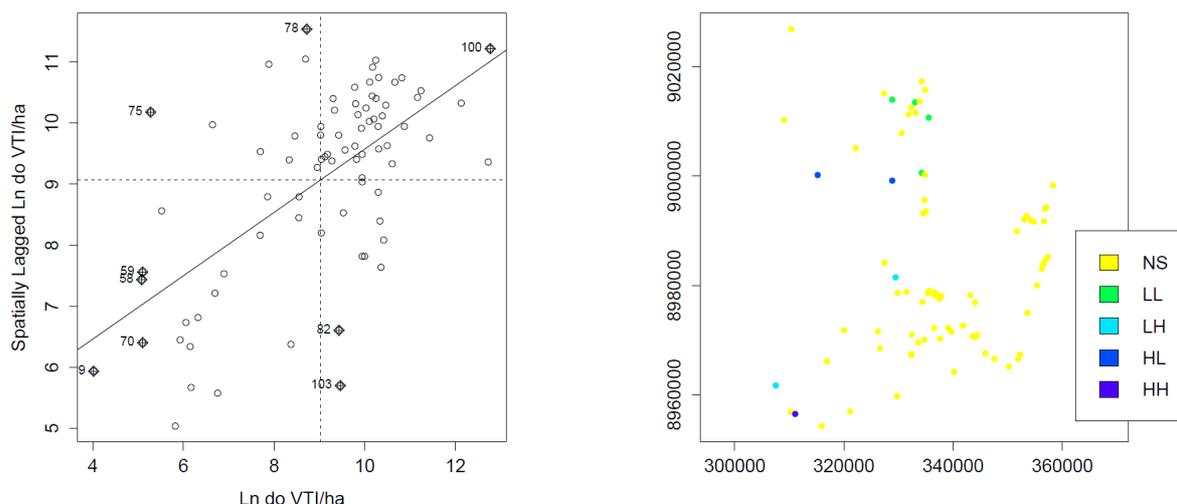


Figura 31 - Gráfico de dispersão e mapa de clusters LISA para os valores do VTI/ha.

O próximo passo foi utilizar as matrizes de pesos espaciais que obtiveram o melhor desempenho para realizar o Teste I de Moran Global sobre os resíduos do modelo 2.1. A Tabela 38 mostra o resultados das 4 melhores matrizes baseadas na vizinhança e na distância.

Tabela 38: Teste I de Moran e significância para as matrizes de ponderação espacial testadas sobre os resíduos do modelo 2.1.

Matrizes	I de Moran	valor p
Baseadas em k vizinhos		
6	0,1638	0,0002703
8	0,1610	1,31E-005
9	0,1535	5,96E-006
4	0,1533	0,0064222
Baseadas na distância		
15km	0,1407	2,66E-007
20km	0,1315	2,30E-008
24km	0,1250	2,80E-014
25km	0,1150	2,96E-014

O resultados mostrados na Tabela 38 mostram que para todas as matrizes apresentadas o teste I de Moran foi significativo, ou seja, os resíduos do modelo apresentam autocorrelação espacial. As matrizes que melhor conseguiram capturar a autocorrelação foi a de 6 vizinhos seguida pela de 8 vizinhos mais próximos.

Apesar destes resultados o Teste I de Moran trata-se de um teste difuso. Para a especificação dos modelos de regressão espacial recomenda-se a utilização de

testes focados pois estes são capazes de identificar a autocorrelação espacial, e indicar se ela ocorre predominantemente na variável dependente ou nos termos do erro.

Com o objetivo de comparar os resultados acima expostos, 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. Nesta etapa, foi utilizado o processo de decisão para especificação de modelo de regressão espacial proposto por Anselin (2005).

Foi utilizada a matriz baseada nos 8 vizinhos mais próximos, a qual ficou entre as melhores matrizes na capacidade de capturar a autocorrelação espacial pelo teste I de Moran mostrado na Tabela 38.

Na Tabela 39 são mostrados os resultados dos testes de especificação baseados no Multiplicador de Lagrange e sua significância. Pode-se observar que o LM (lag) e o LM (erro) foram ambos significativos, devido a isso foi realizada a forma robusta de ambos os testes as quais mostraram que apenas o LM (lag) robusto foi significativo a 1% enquanto que LM (erro) robusto foi não significativo.

Tabela 39 - Resultados do teste de especificação da regressão espacial utilizando o multiplicador de Lagrange.

Teste	Valor	valor p
LM (lag)	21,03	4,52E-06
LM (erro)	9,65	0,0019
LM (lag) robusto	13,69	0,0002
LM (erro) robusto	2,31	0,1284

Diante deste resultado, a especificação do modelo indicou ser necessário a construção do modelo econométrico espacial autorregressivo mais conhecido como Spatial Autorregressive Model (SAR). O resultado do modelo SAR é mostrado na Tabela 40.

De acordo com Trivelloni (2005), o equivalente ao valor t de *Student* para o método de mínimos quadrados corresponde à estatística z nos modelos de regressão por máxima verossimilhança. As probabilidades indicam o grau de significância de cada variável de forma análoga ao da regressão por mínimos quadrados. Observa-se que todas as variáveis são significativas.

A variável ρ representa o coeficiente autorregressivo espacial e se mostra fortemente significativo, indicando a existência de forte autocorrelação espacial. Isto reforça a importância da aplicação dos testes focados para a adequada especificação dos modelos de regressão espacial. O fato de ρ ser positivo indica que quando o valor da variável dependente nos 8 vizinhos mais próximos é elevado o valor do imóvel também é elevado.

Tabela 40 - Variáveis independentes do modelo espacial autorregressivo SAR.

	Coefficientes	Desvio padrão	Estatística z	valor p
(Intercept)	6,5630	0,7187	9,1315	0,00E+000
Natureza do Evento	0,8959	0,2094	4,2782	1,88E-005
Ln Área total (ha)	-0,6973	0,0675	-10,3286	0,00E+000
Ln Área irrig. (ha)	0,1317	0,1020	1,2917	0,1964440
Limite conf. com o rio	0,0105	0,0026	4,1019	4,10E-005
Estado benf.	0,3231	0,0827	3,9076	9,32E-005
Ln Índ. Prod. Veg.	0,1759	0,0499	3,5266	0,0004210
ρ	0,3039	0,0690	4,4061	1,05E-005

Na Tabela 41 é mostrado o modelo com os resultados de alguns testes estatísticos. De acordo com o Teste de Jarque Bera os resíduos do modelo apresentam distribuição normal. E considerando o teste de Breusch-Pagau os resíduos do modelo não apresentam problemas de heterocedasticidade, ou seja, o coeficiente autorregressivo espacial foi capaz de corrigir a heterocedasticidade que o modelo 2.1 apresentava como mostrado na Tabela 29.

Tabela 41 - Modelo espacial autorregressivo.

Modelo	Forma Funcional	Considerações
SAR	$\text{LnVTI} = \beta_0 + \rho W_1 \text{VTI}/\text{ha} + \beta_1 \text{NE} + \beta_2 \text{LnAT} + \beta_3 \text{LnAIR}$ $+ \beta_4 \text{LCR} + \beta_5 \text{EB} + \beta_6 \text{LnIPV} + \varepsilon$	
	Teste de Jarque Bera	valor p=0,4957
	Teste de Breusch-Pagau	valor p=0,05315

4.8 AVALIAÇÃO DE DESEMPENHO DOS MODELOS

4.8.1 Qualidade do Ajuste

Na Tabela 42 é mostrado um comparativo entre os parâmetros de escolha dos melhores modelos de regressão. Nele percebe-se que inicialmente o modelo 1 foi melhor que o modelo 2; após a retirada de observações influenciantes o modelo 1.1 continuou melhor que o 2.1; e finalmente o modelo autorregressivo espacial foi melhor que o modelo 2.1 do qual foi originado.

Tabela 42 - Comparativo entre os modelos por meio do Critério de Informação de Akaike e Schwartz.

Modelos	AIC	SC
1	190,26	217,00
2	231,60	251,05
1.1	167,09	191,40
2.1	205,95	225,10
SAR	190,25	211,80

É importante lembrar que o modelo 2.1 não possuía variáveis relacionadas à localização. Percebe-se então que o modelo SAR conseguiu modelar a dependência espacial presente nos resíduos do referido modelo, cumprindo assim o seu papel.

O fato do modelo 1.1 ter obtido um ajuste melhor que o modelo SAR significa que tanto os MCRL utilizando variáveis relacionadas à localização quanto os modelos de regressão espacial podem ser opções para modelar a dependência espacial presente nos mercados de terras rurais.

Outro fato importante é que nem sempre as variáveis relacionadas à localização podem estar disponíveis (como no caso do modelo 1.1) ou mesmo que estejam disponíveis podem não conseguir capturar toda a autocorrelação presente na região. Nestes casos os modelos de regressão espacial são um importante recurso à disposição dos profissionais, por isso torna-se importante saber utilizá-los.

A seguir serão apresentadas as interpretações dos modelos que obtiveram os melhores ajustes: Modelo 1.1 e o modelo SAR.

Em condições *ceteris paribus* a interpretação do modelo 1.1 é mostrada a seguir:

$$VTI/ha = 10.518,71 * 0,9732334^{DN} * 1,002821^{DE} * 1,943728^{NE} \\ AT^{-0,5015347} * 1,216986^{EB} * 1,006412^{LCR} * 4,083526^{DH} * IPV^{0,154184259} * \varepsilon$$

O valor médio do hectare no município de Petrolina é de R\$ 10.518,71;

A cada 1km em direção ao norte o valor do hectare diminui 2,67%;

A cada 1km em direção a oeste o valor do hectare aumenta 0,28%;

Os dados das ofertas são 94,37% maiores que os dados de negócios realizados;

O aumento da área em 10% provoca redução no valor do hectare em 5,01%;

A cada aumento do estado de conservação das benfeitorias o valor do hectare aumenta 21,69%;

O aumento de 1 metro no limite confrontante com o rio provoca aumento de 0,64% no valor do hectare;

A disponibilidade hídrica para irrigação provoca aumento de 308,35% no valor do hectare;

O aumento do Índice de Produção Vegetal em 10% provoca aumento de 1,54% no valor do hectare.

Em condições *ceteris paribus* a interpretação do modelo SAR é mostrada a seguir:

$$VTI/ha = 708,4143 * 1,355106 W_1 VTI/ha * 2,449472^{NE} * AT^{-0,6973065} * AIR^{0,1317411} * \\ 1,010587^{CMR} * 1,381345^{EB} * IPV^{0,1758903} * \varepsilon$$

O valor médio do hectare no município de Petrolina é de R\$ 708,41. Chama a atenção que este valor seja bastante diferente do obtido no modelo 1.1. Isto se deve ao fato de que neste modelo tal valor também é influenciado pelo coeficiente autorregressivo espacial $\rho=1,355106$ o qual é positivo, indicando que quando o valor do VTI/ha dos 8 vizinhos mais próximos $W_1 VTI/ha$ aumenta, o valor do VTI/ha do

imóvel avaliado também aumenta;

Os dados das ofertas são 144% maiores que os dados de negócios realizados;

O aumento da área em 10% provoca redução no valor do hectare em 6,97%;

O aumento de 10% na área irrigada provoca aumento no valor do hectare em torno de 1,31%;

O aumento de 1 metro no limite confrontante com o rio provoca aumento de 1,05% no valor do hectare;

O aumento de uma unidade no Estado de Conservação das benfeitorias aumenta 38,13% no valor do hectare;

O aumento do Índice de Produção Vegetal em 10% provoca aumento de 1,75% no valor do hectare.

Outro ponto que merece explicação é o coeficiente que a variável Natureza do Evento apresentou em ambos os modelos. No Modelo 1.1 as ofertas são 87% maiores que os negócios realizados e no Modelo SAR as ofertas são 144% maiores que os negócios realizados. Isto se deve ao fato que os dados caracterizados como ofertas estão concentrados às margens do Rio São Francisco, e tratam-se de imóveis em transição de uma aptidão rural para a criação de chácaras de lazer ou ainda a implantação de condomínios nas margens do rio. Devido a isso, existe uma grande expectativa de valorização dos preços dos imóveis situados nesta região. Isto faz com que a diferença entre os imóveis ofertados quando comparados com os negócios realizados que estão distribuídos em toda a área de estudos chegue aos níveis inicialmente comentados. Seria possível controlar isto com a coleta de dados oriundos de ofertas de forma mais distribuída nas demais regiões do município, desta forma o coeficiente refletiria melhor a influência desta variável em toda a região analisada.

4.8.2 Avaliação de Desempenho

Tendo em vista o que foi explicado na metodologia, assim que a amostra foi obtida foram selecionados de forma aleatória 20 elementos de negócios realizados

para compor a subamostra que foi utilizada na avaliação da qualidade dos modelos. A importância deste procedimento se refere ao fato de garantir que a avaliação do desempenho do modelo seja realizada utilizando-se dados que não foram utilizados na modelagem, e assim verificar o funcionamento do mesmo numa situação mais próxima da sua aplicação prática.

A subamostra utilizada correspondeu a 23,8% do total de observações obtidas. Na Figura 32 é possível verificar a distribuição espacial da subamostra. Nela pode-se verificar que a subamostra contempla todas as regiões onde foram obtidos dados de mercado.

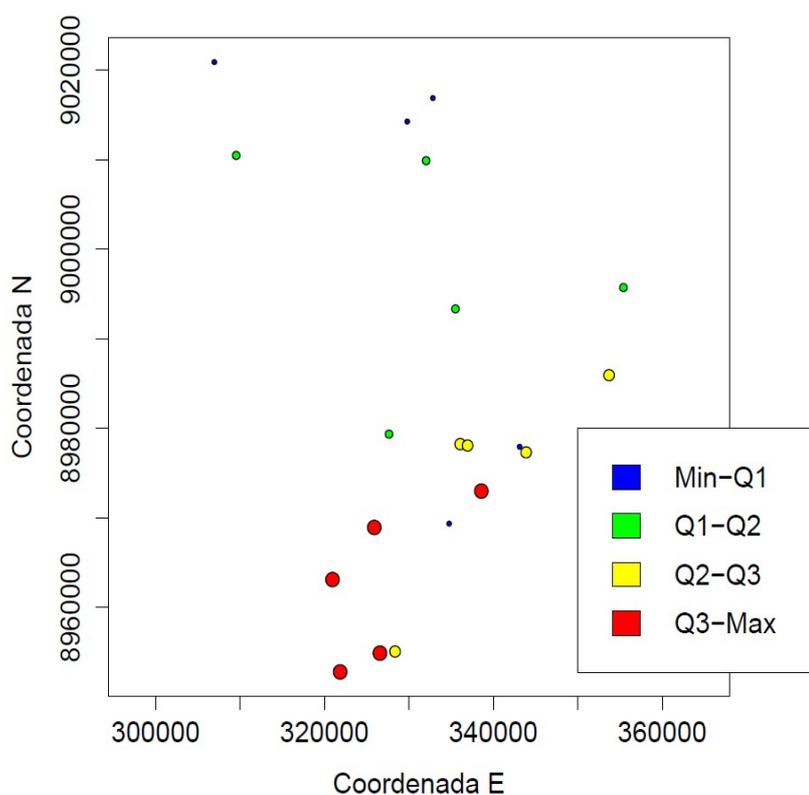


Figura 32 - Distribuição espacial das observações que compõem a amostra de verificação utilizada na avaliação de desempenho dos modelos.

Esta amostra foi utilizada para o cálculo dos parâmetros definidos pela *Standard on Ratio Studies* (IAAO, 2013) a saber: Mediana das razões, Coeficiente de Dispersão (CD) 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 43.

Tabela 43 - Valores propostos pela IAAO para avaliação de desempenho dos modelos.

Estatística	Valores recomendados
Mediana das razões da avaliação	Entre 0,9 e 1,1
Coefficiente de Dispersão (CD)	<20,0%
Diferencial Relativo ao Preço (PRD)	Entre 0,98 e 1,03

Os resultados podem ser visualizados na Tabela 44. Com relação à Mediana das Razões os modelos 1.1 e SAR ficaram dentro dos limites estabelecidos pela norma indicando que o nível geral das avaliações está adequado enquanto que o Modelo 2.1 ficou acima dos limites de referência.

Com relação ao CD todos os modelos se mostraram com uma dispersão elevada quando se analisa toda área de estudo. Provavelmente isto se deve ao fato dos modelos contemplarem regiões muito diferentes como a região de sequeiro e a região dos projetos de irrigação. Apesar disso, o modelo 1.1 foi o que apresentou o menor CD ou seja, foi mais eficiente para controlar a dispersão dos dados.

Com relação ao PRD observa-se que todos os modelos se mostraram com valores pouco acima mas próximos dos níveis de referência, indicando haver pouca regressividade, sendo que o modelo 1.1 foi o que obteve o valor mais próximo dos limites de referência.

Tabela 44 - Parâmetros de avaliação de desempenho dos modelos.

	Mediana das Razões	CD (%)	PRD
Modelo 1.1	1,1	50,1	1,12
Modelo 2.1	1,4	97,2	1,59
Modelo SAR	1,0	79,9	1,43

Avalia-se que o Modelo 1.1 apresentou o melhor desempenho entre os três analisados, em seguida o Modelo SAR foi o que apresentou o segundo melhor desempenho o que corrobora as informações obtidas na avaliação da qualidade do ajuste mostrada na Tabela 42.

A importância da realização da avaliação de desempenho é que ela é capaz de indicar quais aspectos do modelo devem receber maior atenção e assim por meio das correções é possível construir modelos mais eficientes. Nos modelos em análise, o principal aspecto a ser aperfeiçoado é o coeficiente de dispersão elevado.

Provavelmente com a subdivisão do município em duas regiões, por exemplo região de sequeiro e região dos projetos de irrigação, e a construção de modelos específicos para cada região, seja possível enquadrar os modelos dentro dos limites definidos pela norma. A indisponibilidade de uma quantidade maior de dados na região de sequeiro impossibilitou a adoção desta providência.

Na próxima etapa estes modelos serão utilizados para gerar superfícies de preços que também serão submetidas à uma avaliação de desempenho, quando então será possível obter a PVG.

4.9 ELABORAÇÃO DA PLANTA DE VALORES GENÉRICOS

Nesta etapa serão utilizados o conjunto de dados obtidos em campo bem como o conjunto dos valores preditos pelos Modelos 1.1 e SAR para elaborar as superfícies de preços do município de Petrolina. Vale ressaltar que estes conjuntos contam com diferentes números de dados a saber: conjunto original 84 observações; Modelo 1.1 82 observações e Modelo SAR 81 observações.

Em seguida, as superfícies de preços obtidas serão avaliadas na sua capacidade de fornecer informações a respeito do valor do hectare na região de estudos. Para isto, serão utilizados o conjunto de 20 observações de verificação, da mesma forma como na avaliação de desempenho dos modelos.

Nesta etapa foi utilizada novamente a geoestatística. Primeiramente será realizada a modelagem dos semivariogramas teóricos os quais fornecerão os parâmetros necessários ao interpolador da Krigagem Ordinária possibilitando assim elaborar a superfície de preços de terras.

4.9.1 Modelagem Geoestatística do Conjunto de Dados Campo.

O conjunto de dados de campo conta com as 84 observações originais que subsidiaram a elaboração dos modelos. Primeiramente foi construído o semivariograma experimental cujos parâmetros podem ser visualizados na Tabela 45. O semivariograma experimental contou com 26 pontos num espaçamento de aproximadamente 1,1km e considerando 60% da distância máxima entre pontos o que correspondeu a 44km. Todos os pontos foram calculados a partir de no mínimo

51 pares de pontos, bem acima do mínimo recomendado que é 32 pares.

Tabela 45 - Parâmetros do semivariograma experimental para 84 observações dos dados de campo.

	Distância (km)	Semivariância	Autocorrelação	Núm. de pares.
1	16,00	1,69	3,29	942,00
2	17,10	0,67	0,87	65,00
3	18,19	1,27	2,84	103,00
4	19,29	1,06	2,10	88,00
5	20,38	1,28	2,69	89,00
6	21,48	1,15	1,98	102,00
7	22,58	2,12	3,81	94,00
8	23,67	1,63	3,56	125,00
9	24,77	1,26	2,85	82,00
10	25,86	1,48	2,05	102,00
11	26,96	1,89	3,15	93,00
12	28,06	1,70	3,33	86,00
13	29,15	1,87	2,93	95,00
14	30,25	1,66	2,80	120,00
15	31,34	2,01	3,70	83,00
16	32,44	1,85	3,45	90,00
17	33,54	1,56	2,49	88,00
18	34,63	1,41	2,74	70,00
19	35,73	1,98	3,46	63,00
20	36,82	1,60	2,53	72,00
21	37,92	0,99	1,35	58,00
22	39,02	1,42	1,88	57,00
23	40,11	1,53	3,28	60,00
24	41,21	1,57	2,55	47,00
25	42,30	1,44	2,27	51,00
26	43,40	1,62	2,45	64,00

Após o cálculo do semivariograma experimental foram ajustados modelos teóricos de semivariogramas cujos parâmetros de ajuste encontram-se apresentados na Tabela 46. O alcance variou de 30 a 32km e o Efeito Pepita Relativo (EPR) indicou que há forte dependência espacial.

Tabela 46 - Parâmetros dos semivariogramas teóricos ajustados pelo método dos mínimos quadrados para 84 observações dos dados originais.

Modelo	φ_1	φ_2	$\varphi_1+\varphi_2$	$\varphi_3=g(a)$	a	EPR
Esférico	0	1,62	1,62	30,46	30,46	0
Exponencial	0	1,67	1,67	10,72	32,11	0

φ_1 : efeito pepita; φ_2 : contribuição; $\varphi_3=g(a)$: função do alcance; a: alcance (km); $EPR=\varphi_1/(\varphi_1+\varphi_2)$: efeito pepita relativo.

Dando prosseguimento à análise, foi realizada a validação cruzada dos modelos teóricos anteriormente ajustados cujos resultados encontram-se expostos na Tabela 47. A análise dos parâmetros permite afirmar que o modelo exponencial foi o que obteve o melhor ajuste. Na Figura 33 encontra-se a representação do referido modelo teórico.

Tabela 47 - Parâmetros para escolha dos modelos ajustados.

Modelos	Erro Médio	Erro Médio Reduzido	Desvio padrão do erro médio	Desvio padrão do erro reduzido	Erro absoluto
Esférico	0,0060	1,6014	-0,0141	4,0054	100,8695
Exponencial	0,0070	1,5858	-0,0079	2,8873	99,1512

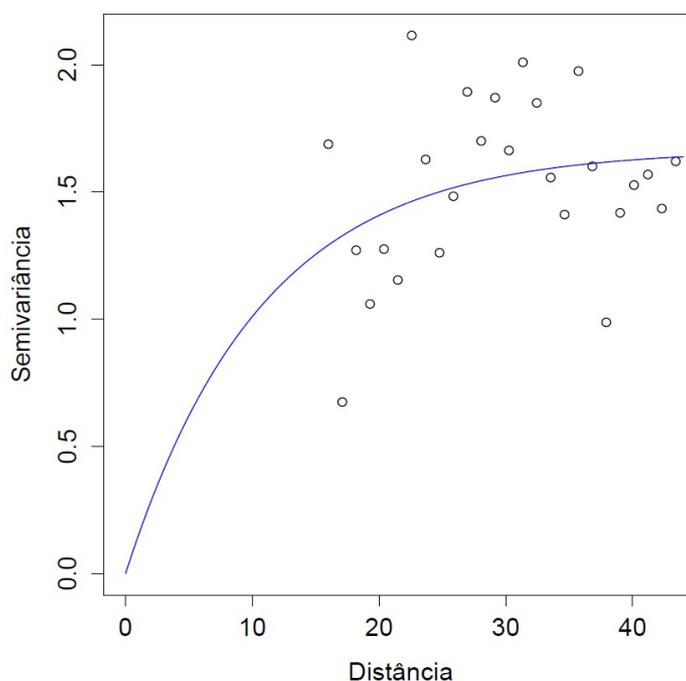


Figura 33 - Semivariograma teórico, modelo exponencial, ajustado para 84 observações dos dados originais, a distância é dada em quilômetros.

A partir do melhor modelo teórico ajustado, seus parâmetros de ajuste foram utilizados para gerar a superfície de preços de terras do município de Petrolina, bem como a superfície do desvio padrão da krigagem ordinária, os quais podem ser visualizados na Figura 34. Da observação da figura do lado esquerdo percebe-se como a krigagem conseguiu representar a variação de preços que ocorre no município de Petrolina com os valores mais altos situados na parte sul, próximo ao Rio São Francisco e à medida que se avança para o norte ocorre uma significativa redução dos preços a qual é reflexo da indisponibilidade hídrica para o desenvolvimento da agricultura irrigada.

O mapa do desvio padrão da krigagem mostra que o erro de estimativa foi menor nas proximidades de onde houve pontos amostrados e cresce à medida que afasta-se destes pontos. Como na região de sequeiro houve maior dificuldade de obter observações de mercado, refletindo-se numa menor densidade de amostragem, observa-se que é nesta região que também o erro de estimativa é maior. Isto reforça a importância de que haja uma boa espacialização dos dados amostrais com o objetivo de diminuir o erro de estimativa em toda a região de abrangência da superfície de preços.

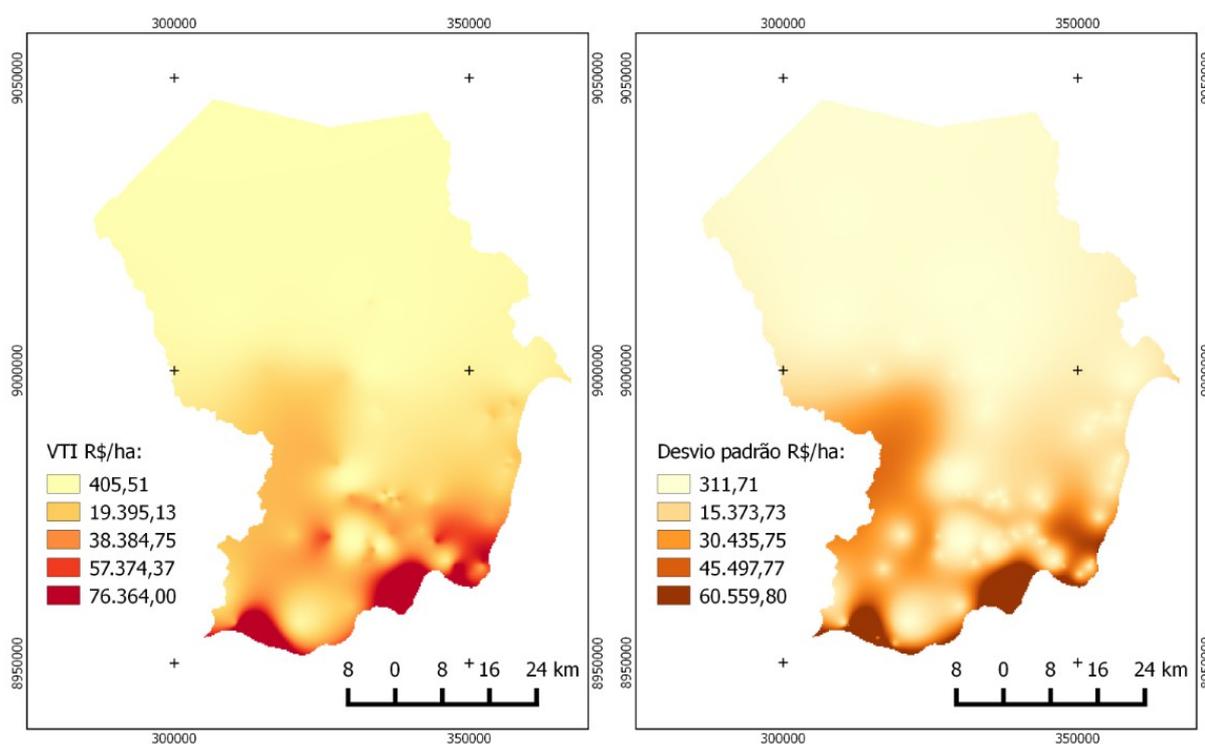


Figura 34 - krigagem ordinária (esquerda) e desvio padrão (direita) para as 84 observações coletadas em campo.

4.9.2 Modelagem Geoestatística dos Valores Preditos pelo Modelo 1.1.

Este conjunto de dados refere-se aos valores preditos pelo Modelo 1.1, o qual conta com 82 observações. Primeiramente foi construído o semivariograma experimental cujos parâmetros podem ser visualizados na Tabela 48. O semivariograma experimental contou com 09 pontos num espaçamento de aproximadamente 5,4km e considerando 60% da distância máxima entre pontos o que correspondeu a 44km. Todos os pontos foram calculados a partir de no mínimo 99 pares de pontos, bem acima do mínimo recomendado que é 32 pares.

Tabela 48 - Parâmetros do semivariograma experimental para 82 observações dos valores preditos pelo Modelo 1.1.

	Distância (km)	Semivariância	Autocorrelação	Núm. de pares.
1	0,3	0,53403	1,14587	99
2	5,7	0,67654	1,35611	292
3	11,1	0,76899	1,45651	329
4	16,5	0,75610	1,51904	369

5	21,9	0,94954	1,60328	476
6	27,3	1,15573	1,65945	445
7	32,7	1,00356	1,33124	410
8	38,1	0,99898	1,57858	280
9	43,5	1,02946	1,96044	148

Após o cálculo do semivariograma experimental foram ajustados modelos teóricos de semivariogramas cujos parâmetros de ajuste encontram-se apresentados na Tabela 49. O alcance variou de 29 a 56km e o Efeito Pepita Relativo (EPR) indicou que todos os modelos apresentaram moderada dependência espacial.

Tabela 49 - Parâmetros dos semivariogramas teóricos ajustados pelo método dos mínimos quadrados aos valores preditos pelo Modelo 1.1.

Modelo	φ_1	φ_2	$\varphi_1+\varphi_2$	$\varphi_3=g(a)$	a	EPR
Exponencial	0,5187	0,5146	1,0333	33,3234	33,3234	0,50
Esférico	0,5106	0,6144	1,1250	18,9606	56,8010	0,45
Gaussiano	0,5728	0,4725	1,0453	17,0733	29,5507	0,55

φ_1 : efeito pepita; φ_2 : contribuição; $\varphi_3=g(a)$: função do alcance; **a**: alcance; $EPR=\varphi_1/(\varphi_1+\varphi_2)$: efeito pepita relativo.

Em seguida, foi realizada a validação cruzada dos modelos teóricos anteriormente ajustados, cujos resultados encontram-se expostos na Tabela 50. A análise dos parâmetros permite afirmar que o modelo gaussiano foi o que obteve o melhor ajuste. Na Figura 35 encontra-se a representação do referido modelo teórico.

Tabela 50 - Parâmetros para escolha dos modelos ajustados.

Modelos	Erro Médio	Erro Médio Reduzido	Desvio padrão do erro médio	Desvio padrão do erro reduzido	Erro absoluto
Exponencial	0,0053	1,0027	0,0025	1,1665	60,0795
Esférico	0,0047	1,0087	0,0021	1,1862	60,2816
Gaussiano	0,0046	1,0036	0,0019	1,1964	59,7164

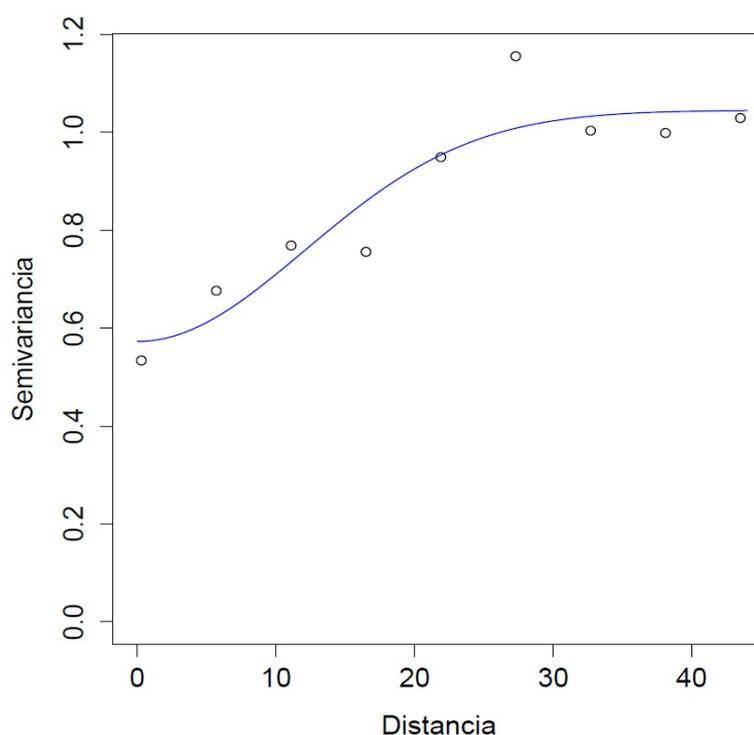


Figura 35 - Semivariograma teórico, modelo gaussiano, ajustado para 82 observações dos valores preditos pelo Modelo 1.1, a distância é dada em quilômetros.

Os parâmetros obtidos a partir do melhor modelo teórico ajustado foram utilizados para gerar a superfície de preços de terras do município de Petrolina, bem como a superfície do desvio padrão da krigagem ordinária, os quais podem ser visualizados na Figura 36. Da observação da figura do lado esquerdo percebe-se que krigagem também conseguiu representar a variação de preços que ocorre no município de Petrolina entretanto, de forma menos detalhada que na krigagem dos dados de campo.

O mapa do desvio padrão da krigagem também mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrados e cresce à medida que se afasta destes pontos. Entretanto, neste caso parece haver uma menor variação dos valores do erro de estimativa quando comparado com a krigagem realizada a partir dos dados originais.

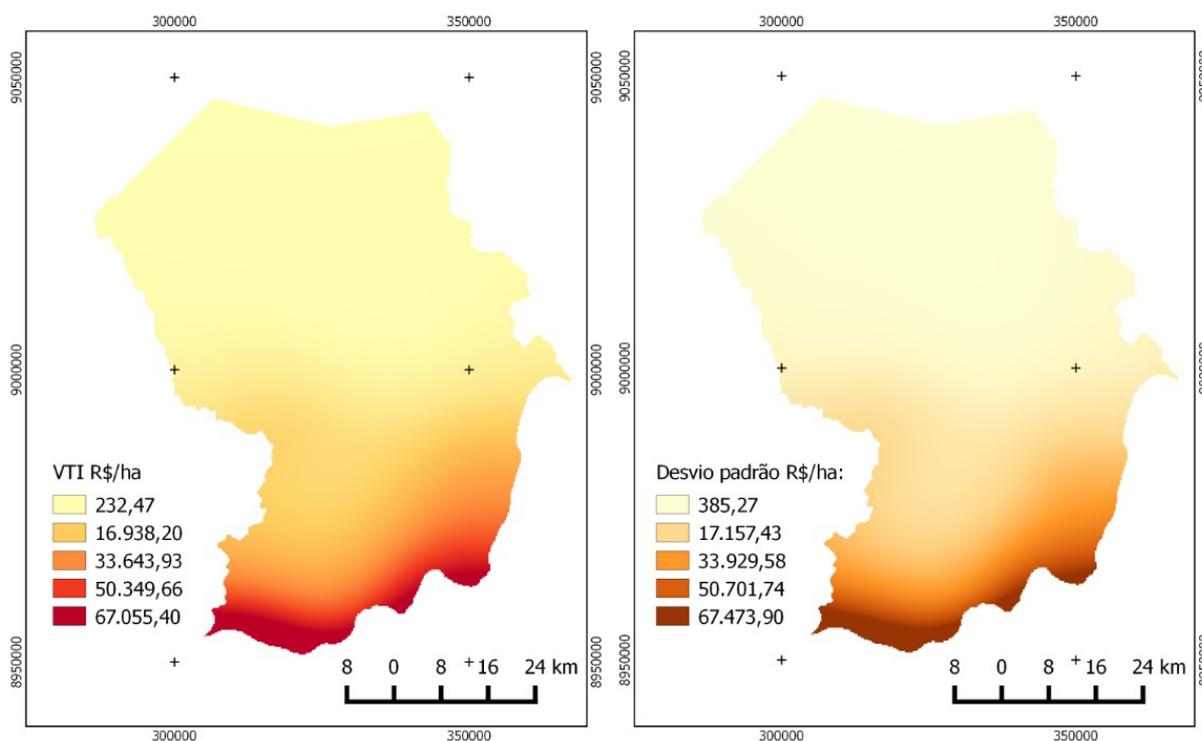


Figura 36 - krigagem ordinária (esquerda) e desvio padrão (direita) para as 82 observações previstas pelo Modelo 1.1.

4.9.3 Modelagem Geoestatística dos Valores Preditos pelo Modelo SAR.

Finalmente, o último conjunto de dados refere-se aos valores preditos pelo Modelo SAR, o qual contou com 81 observações. Primeiramente, foi construído o semivariograma experimental cujos parâmetros podem ser visualizados na Tabela 51. O semivariograma experimental contou com 13 pontos num espaçamento de aproximadamente 3,4km e considerando 60% da distância máxima entre pontos o que correspondeu a 44km. Todos os pontos foram calculados a partir de no mínimo 69 pares de pontos, acima do mínimo recomendado que é 32 pares.

Tabela 51 - Parâmetros do semivariograma experimental para 81 observações dos valores preditos pelo Modelo SAR.

	Distância (km)	Semivariância	Autocorrelação	Núm. de pares.
1	0,6000	0,3622	0,6999	69
2	4,0462	0,6702	1,1561	131
3	7,4923	0,8138	1,1625	253
4	10,9385	0,7366	1,4334	202
5	14,3846	0,9482	1,4054	217

6	17,8308	0,8458	1,5212	240
7	21,2769	1,0227	1,4916	278
8	24,7231	1,1343	1,7146	299
9	28,1692	1,1245	1,6333	264
10	31,6154	1,1306	1,5144	269
11	35,0615	0,9766	1,2363	211
12	38,5077	1,1068	1,6967	180
13	41,9538	1,0741	2,2620	164

Após o cálculo do semivariograma experimental foram ajustados modelos teóricos de semivariogramas cujos parâmetros de ajuste encontram-se apresentados na Tabela 52. O alcance variou de 22 a 31km e o Efeito Pepita Relativo (EPR) indicou que há moderada dependência espacial em todos os modelos.

Tabela 52 - Parâmetros dos semivariogramas teóricos ajustados para 81 observações dos valores preditos pelo Modelo SAR.

Modelo	φ_1	φ_2	$\varphi_1+\varphi_2$	$\varphi_3=g(a)$	a	EPR
Exponencial	0,450	0,636	1,086	28,203	28,203	0,414
Esférico	0,360	0,748	1,108	10,378	31,091	0,325
Gaussiano	0,517	0,563	1,08	13,194	22,836	0,479

φ_1 : efeito pepita; φ_2 : contribuição; $\varphi_3=g(a)$: função do alcance; a: alcance (km); $EPR=\varphi_1/(\varphi_1+\varphi_2)$: efeito pepita relativo.

Em seguida, foi realizada a validação cruzada dos modelos teóricos anteriormente ajustados cujos resultados encontram-se expostos na Tabela 53. A análise dos parâmetros permite afirmar que o modelo esférico foi o que obteve o melhor ajuste. Na Figura 37 encontra-se a representação do referido modelo teórico.

Tabela 53 - Parâmetros para escolha dos modelos ajustados.

Modelos	Erro Médio	Erro Médio Reduzido	Desvio padrão do erro médio	Desvio padrão do erro reduzido	Erro absoluto
Exponencial	-0,0065	1,3034	-0,0035	1,5609	77,3442
Esférico	-0,0097	1,3032	-0,0056	1,5662	76,8965
Gaussiano	-0,0115	1,3229	-0,0068	1,6035	78,2261

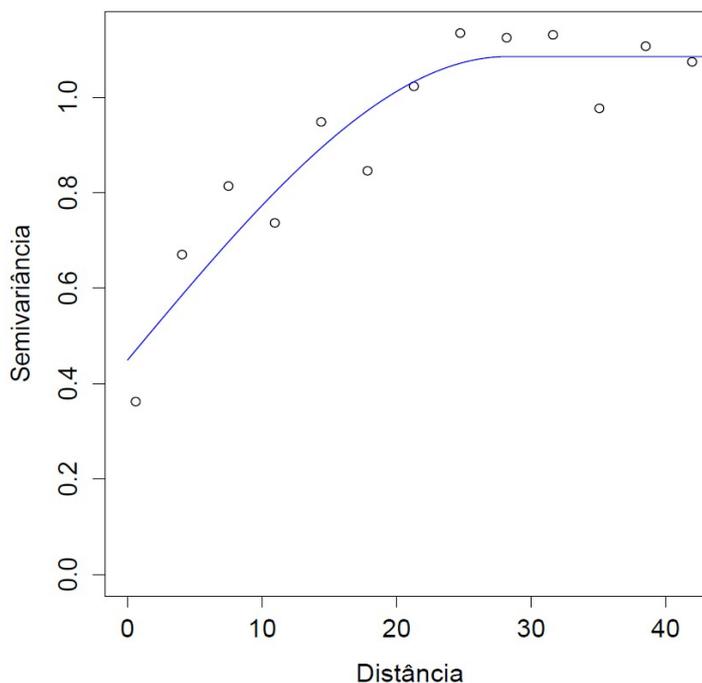


Figura 37 - Semivariograma teórico, modelo esférico, ajustado para 81 observações dos valores preditos pelo Modelo SAR, a distância é dada em quilômetros.

Os parâmetros obtidos a partir do melhor modelo teórico ajustado foram utilizados para gerar a superfície de preços de terras do município de Petrolina, bem como a superfície do desvio padrão da krigagem ordinária, os quais podem ser visualizados na Figura 38. Da observação da figura do lado esquerdo percebe-se que krigagem também conseguiu representar a variação de preços que ocorre no município de Petrolina entretanto, semelhantemente ao que foi observado na krigagem realizada a partir dos valores preditos pelo Modelo 1.1 observa-se que a superfície é menos detalhada quando comparada com a krigagem obtida a partir dos dados de campo.

O mapa do desvio padrão da krigagem também nos mostra que o erro de estimativa foi menor nas proximidades dos pontos amostrados e cresce à medida que afasta-se destes pontos Figura 38. Entretanto, neste caso a superfície do erro de estimativa foi a que se apresentou com aparência mais suavizada quando comparado com os outros resultados.

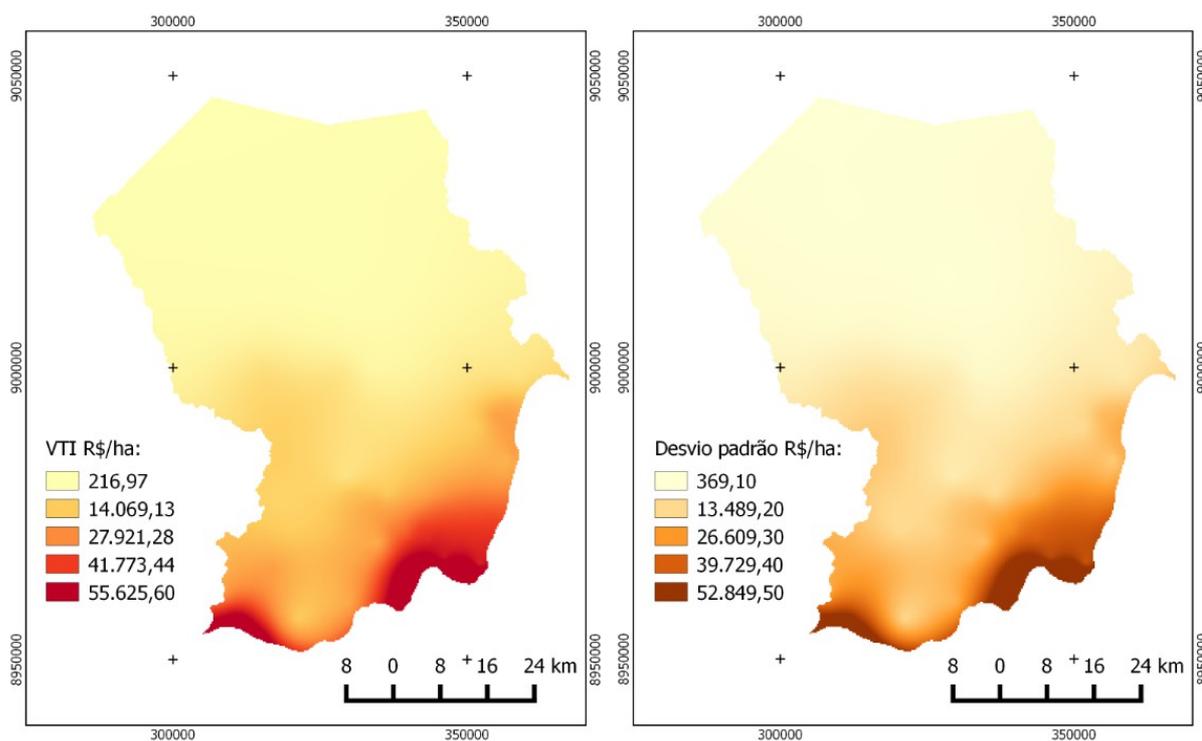


Figura 38 - krigagem ordinária (esquerda) e desvio padrão (direita) para as 81 observações preditas pelo Modelo SAR.

Na próxima etapa será realizada a avaliação de desempenho das superfícies de preços obtidas quando então poderemos chegar a conclusões mais acertadas a respeito de qual delas melhor representa a realidade do mercado em estudo.

4.9.4 Anisotropia

Para análise da anisotropia foram construídos variogramas direcionais nas quatro principais direções recomendadas a saber: 0° , 45° , 90° e 135° . Na Figura 39 estão dois exemplos destes variogramas, o do lado esquerdo para os dados oriundos dos valores preditos pelo modelo 1.1 e do lado direito para os valores preditos pelo modelo SAR. Nestes 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.

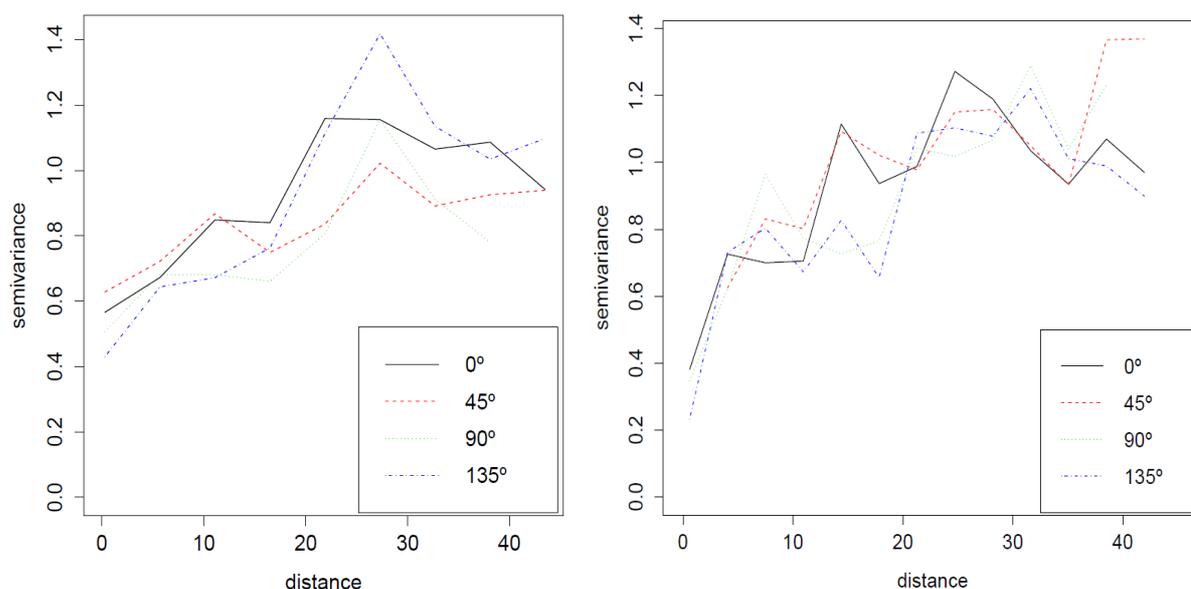


Figura 39 - Semivariogramas direcionais (0° , 45° , 90° e 135°) para os valores preditos pelo modelo 1.1 (esquerda) e SAR (direita).

4.9.5 Avaliação do Desempenho das Krigagens Ordinárias.

Após a obtenção das superfícies de preços é preciso avaliar qual delas melhor conseguiu representar a realidade do mercado de terras do município de Petrolina. Para isto será utilizada a Raiz Quadrada do Erro Médio Quadrático (REMQ) da mesma forma como proposto por Marques et al. (2012) por meio da Equação 59.

$$REMQ = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (VTI/ha_{estimado} - VTI/ha_{real})^2} \quad (59)$$

O valor do VTI/ha_{real} foi obtido a partir da amostra de verificação, a qual é composta por 20 observações e que foi utilizada na avaliação de desempenho dos modelos. O valor do $VTI/ha_{estimado}$ foi obtido a partir dos valores interpolados pela krigagem nas superfícies de preços para as mesmas coordenadas dos pontos de verificação. O resultado para cada uma das superfícies de preços mostradas na etapa anterior estão disponíveis na Tabela 54.

Tabela 54 - Estimativa da Raiz do Erro Médio Quadrático no VTI/ha dos vinte pontos de verificação.

Superfícies de preços	REMQ R\$/ha
Dados originais (84 obs.)	18.301,40
Modelo 1.1 (82 obs.)	28.481,54
Modelo SAR (81 obs.)	22.798,67

Os resultados mostram que a Krigagem realizada a partir dos dados originais foi a que proporcionou o menor valor da REMQ dentre todas as superfícies obtidas. Foi realizado o teste DHS (*Honest Significant Differences*) de Tukey para a média dos resíduos com o objetivo de verificar se os resultados obtidos são estatisticamente diferentes. Os resultados podem ser vistos na Tabela 55 e mostram que as diferenças observadas para a REMQ em relação às superfícies de preços são significativas estatisticamente apenas para a superfície gerada a partir Modelo 1.1 em relação em relação às demais. Entre a superfície de preços gerada a partir dos dados originais e a gerada a partir do modelo SAR as diferenças entre as médias se devem ao acaso.

Tabela 55 - Resultado do Teste de Tukey de comparação de múltiplas médias com 95% de confiança.

Modelos comparados	Diferença entre as médias	Mínimo	Máximo	Valor p
Modelo 1.1-Dados originais	8679,65	927,67	16431,64	0,03
Modelo SAR-Dados originais	227,27	-7524,72	7979,26	1,00
Modelo SAR-Modelo 1.1	-8452,38	-16204,37	-700,40	0,03

Ressalta-se que os valores obtidos para a REMQ são elevados e mostram a limitação da utilização das PVG para a realização de avaliações pontuais. Isto se deve ao fato que a base cartográfica disponível para a elaboração da PVG estava na escala 1:250.000. Sendo assim, a principal indicação das PVG rurais é de informar as faixas de preços nas diferentes regiões do município.

Além disso, destaca-se que a PVG foi gerada a partir da interpolação apenas da variável dependente nas diferentes localizações. Entretanto, os modelos mostram que o valor da variável dependente é controlado tanto pelas suas coordenadas quanto por uma série de variáveis como a disponibilidade hídrica, a área, o índice de produção vegetal entre outras. Isto reforça a ideia de que quando for necessário realizar a avaliação de um imóvel especificamente o ideal é utilizar o modelo, e a

PVG será capaz de fornecer uma ideia geral dos preços numa determinada região. Trata-se portanto de dois produtos com diferentes aplicações.

Sendo assim, foi escolhida a superfície que apresentou a menor REMQ para representar a PVG do município de Petrolina a qual pode ser vista na Figura 40.

Da mesma forma como observado por Marques et al. (2012) percebe-se que a interpolação por krigagem ordinária permitiu a observação de fatores que podem influenciar diretamente na definição dos parâmetros da modelagem como tais como a disposição dos elementos na área, a proximidade entre eles e o tamanho da amostra, permitindo assim uma melhor estimativa dos valores que compõem a PVG.

A disponibilização das PVG num ambiente SIG permite a sobreposição de diferentes camadas de informações como as principais localidades municipais, rodovias e principalmente os limites dos imóveis, quando disponíveis. Se pensarmos nos dois principais cadastros, o Cadastro Nacional de Imóveis Rurais e o Cadastro Ambiental Rural, a sobreposição destas informações à PVG se torna um importante instrumento de planejamento e de fiscalização em vários níveis de atuação.

Outro aspecto importante se refere ao fato que a PVG é um instrumento de justiça fiscal pois permite que a avaliação dos imóveis seja realizada com base nas variações locais dos preços nas diferentes regiões. Isto evita que haja superavaliações o que onera os produtores, e também possibilita aumento de arrecadação para o município pois evita que os imóveis sejam subavaliados.

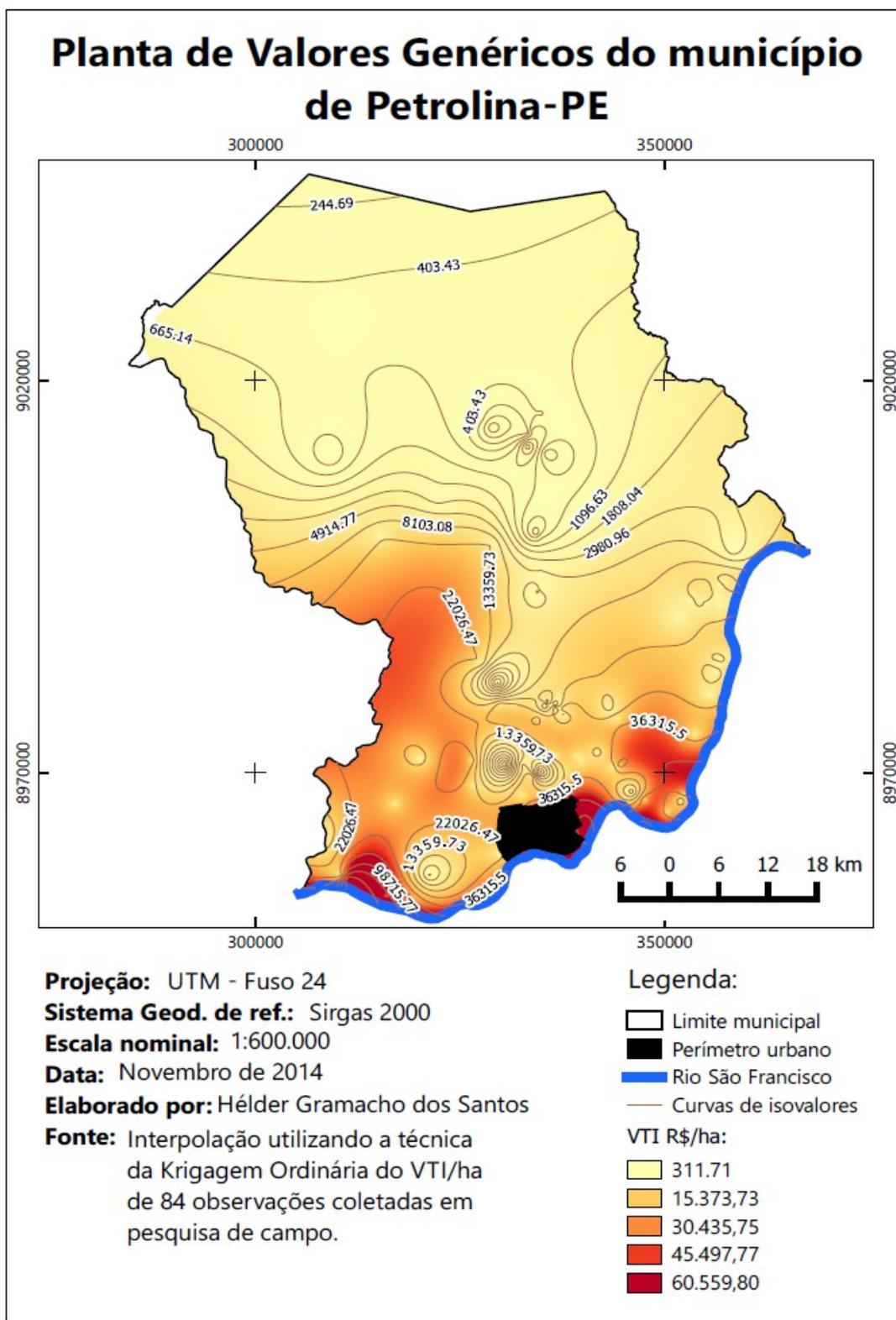


Figura 40 - Planta de Valores Genéricos do município de Petrolina-PE.

5 CONCLUSÕES E RECOMENDAÇÕES

Levando-se em consideração os objetivos específicos determinados no início do trabalho e os resultados alcançados ao longo do seu desenvolvimento teceremos comentários a respeito das principais conclusões obtidas neste estudo.

5.1 CONCLUSÕES

A análise da literatura científica atual permitiu compreender que o Brasil encontra-se atrasado com relação à valoração cadastral rural, tanto quando comparado com a realidade dos vizinhos sul-americanos quanto com os países europeus, com destaque para a Espanha, um dos países com maior desenvolvimento nesta área.

Entre as questões que vem recebendo a atenção dos pesquisadores destaca-se a modelagem dos dados de mercado que encontram-se georreferenciados. Nestes casos, a Econometria tradicional pode tornar-se ineficiente quando afetada pela autocorrelação espacial, pois suas estimativas tornam-se tendenciosas. Diante disso, a Econometria Espacial e a Geoestatística vem sendo utilizadas como alternativas na modelagem dos efeitos espaciais em diferentes mercados imobiliários.

Outro importante aspecto proporcionado pela análise dos trabalhos recentes foi a possibilidade de seguir as técnicas que vem sendo utilizadas com sucesso bem como testar novas possibilidades baseadas nas recomendações para os trabalhos futuros ofertadas pelas dissertações e teses. Dessa forma, o trabalho foi aprimorado, alcançando resultados interessantes e contribuindo para o avanço do conhecimento nos mercados de terras rurais. Como exemplo foi possível obter um modelo de valoração cadastral que leva em consideração a localização dos imóveis bem como gerar a PVG do município de Petrolina.

Adentrando na parte prática do trabalho, foram mostradas as principais dificuldades na obtenção dos dados de mercado quando se trabalha com imóveis rurais. Tais problemas poderiam ser muito reduzidos caso os municípios contassem com o Cadastro Territorial Multifinalitário implantado, atualizado e em funcionamento. Entretanto, este fato não pode ser usado como desculpa para não avançar no tema

da valoração cadastral rural pois foram explicitadas as formas como foi possível obter os dados de mercado necessários à realização do trabalho com tempo e recursos financeiros restritos.

Entende-se que o município é que possui as melhores condições para coletar tais dados. Contando com planejamento, é possível coletar os dados que possibilitarão a construção das PVG nos próprios setores de tributos, informações estas que podem ser validadas por uma equipe de campo com relativa facilidade, frente ao que é possível atualmente. Sugere-se a coleta de informações como as fornecidas na ficha de pesquisa disponível no **APÊNDICE 1**, adaptadas às diferentes realidades.

Avançando um pouco mais, foi mostrada a importância de realizar a análise exploratória dos dados antes de iniciar a modelagem propriamente dita. O estudo das melhores transformações para as diferentes variáveis obtidas, a construção de diagramas de dispersão bem como da matriz de correlações permitiu a construção de modelos mais consistentes e evitou o surgimento de problemas durante a etapa de validação dos pressupostos dos modelos. Os dados bem como as regiões de estudos já contam com problemas intrínsecos às suas realidades, a análise exploratória evita que o pesquisador introduza mais problemas à situação analisada.

Com relação à modelagem propriamente dita, foi dado início a esta etapa a partir de dois modelos construídos com base na econometria tradicional, um contando com variáveis relacionadas à localização e outro no qual elas estavam ausentes. No primeiro, mostrou-se que a introdução de variáveis relacionadas à localização permite controlar o surgimento de autocorrelação espacial nos resíduos do modelo, o que invalidaria os pressupostos do MCRL. No segundo modelo, foi mostrado que a falta destas variáveis provoca o surgimento da autocorrelação espacial e deu início à etapa da modelagem dos efeitos espaciais por meio da geoestatística e da econometria espacial.

Ao final da modelagem foram obtidos dois modelos livres da autocorrelação espacial nos seus resíduos: um obtido por meio da econometria tradicional e outro obtido pela econometria espacial. Estes resultados mostram que ambas as técnicas, quando bem utilizadas, podem ser eficientes para controlar o problema da autocorrelação espacial. Entretanto, a análise da qualidade do ajuste mostrou que o modelo econométrico tradicional obteve um resultado melhor no AIC e SC quando

comparado com o modelo econométrico espacial. Ressalta-se porém que o modelo econométrico espacial cumpriu o seu papel pois controlou a autocorrelação espacial nos resíduos do modelo sem variáveis relacionadas ao espaço e quando comparados, melhorou o ajuste com relação ao modelo do qual foi originado.

A avaliação de desempenho de todos os produtos gerados representa uma importante contribuição para os mercados de terras rurais. Além da avaliação da qualidade do ajuste, também foi realizada a avaliação de desempenho dos modelos com base na metodologia estabelecida pela IAAO, (2013) e neste aspecto, merece destaque a utilização de pontos de validação do modelo que não fizeram parte da construção dos mesmos, seguindo a recomendação de diversos autores. Além disso, a utilização de 20 pontos de verificação, o equivalente a 23,8% das observações de trabalho 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.

A avaliação de desempenho analisou o nível geral das avaliações, a dispersão em toda área de estudos e questões de progressividade e regressividade por meio dos respectivos parâmetros Mediana das razões, Coeficiente de Dispersão e o Diferencial Relativo do Preço. No geral, os modelos apresentaram bom nível geral das avaliações e pequena regressividade entretanto, todos eles ficaram fora dos parâmetros estabelecidos pela *Standard on Ratio Studies* (IAAO, 2013) no que se refere ao CD, apresentando portanto elevada dispersão.

Acredita-se que este resultado se deva ao fato do trabalho ter utilizado informações coletadas em todo o município, o que trouxe para os modelos realidades bastantes diferentes como a dos imóveis localizados em zonas de sequeiro e dos imóveis localizados na região em que há água disponível para irrigação, o que tornou a modelagem mais complexa. Ressalta-se que não existem trabalhos com mercados de terras rurais que realizaram a avaliação de desempenho com a referida norma. Os poucos trabalhos que a utilizaram trabalharam com imóveis urbanos, mas mesmo assim utilizaram os próprios pontos da modelagem na avaliação de desempenho, ou ainda quando utilizaram pontos que não fizeram parte da modelagem utilizaram uma quantidade relativamente pequena. Portanto, falta base de comparação dos resultados deste trabalho com outros envolvendo imóveis rurais, para que se possa avaliar melhor estes resultados.

Posteriormente à obtenção dos modelos, foi realizada a modelagem geoestatística dos dados originais coletados em campo bem como dos pontos obtidos a partir dos valores preditos pelos modelos e finalizando com a krigagem de tais valores. Estas interpolações deram origem às superfícies de preços, as quais também foram submetidas à avaliação de desempenho utilizando os dados de validação utilizados na etapa anterior.

A avaliação de desempenho das superfícies de preços utilizando o RMSE foi importante para indicar a aplicação da PVG rural. Como a base cartográfica digital obtida para a elaboração da PVG estava na escala 1:250.000 observou-se um RMSE elevado o que torna a PVG imprópria para avaliações pontuais mas adequada para uma indicação geral das faixas de preços nas diferentes regiões do município. Quando for possível obter bases cartográficas digitais em escalas maiores espera-se um melhor desempenho das PVG rurais.

Diante disso, mostra-se que a metodologia proposta possibilitou a obtenção de um modelo bem como a elaboração da PVG, ambos representativos da realidade dos preços de terras do município de Petrolina. Tais produtos possuem diferentes aplicações, o modelo é indicado para avaliações em massa, pois utiliza as principais variáveis que influenciam os preços das terras no município; enquanto que a PVG serve como indicação geral dos preços nas regiões.

A metodologia proposta se mostrou aplicável à realidade dos mercados de terras rurais, e possível de ser implementada pelos municípios brasileiros. Com planejamento e uma equipe relativamente pequena é possível coletar os dados necessários para que os municípios passem a elaborar suas PVG. O investimento maior é com a contratação do serviço de elaboração propriamente dito, o qual exige profissional experiente.

Neste trabalho, a variável dependente utilizada foi o Valor Total do Imóvel/ha entretanto, tal metodologia pode ser aplicada para obter estimativas também do Valor da Terra Nua. Sendo assim, os municípios podem obter a base de cálculo para a cobrança tanto do ITBI quanto do ITR. Ressalta-se que para os município conveniados com a RFB, uma das obrigações é manter atualizados os preços das terras. Acredita-se que a utilização dos valores atualizados destas bases de cálculos pode significar incrementos de arrecadação importantes para o município e a utilização da metodologia proposta confere transparência e justiça fiscal para os

contribuintes.

A utilização dos SIG possibilitou a obtenção de variáveis relacionadas à localização que podem ser úteis na modelagem dos efeitos espaciais nos mercados de terras rurais. Os SIG permitem aplicar a ideia de que os imóveis estão sujeitos não apenas a um polo de influência mas sim a diversos polos. Por meio de algoritmos específicos é possível obter a distância euclidiana entre os imóveis armazenados em um banco de dados e os diferentes polos de influência. Além disso, os SIG contam com ferramentas que permitem a construção dos semivariogramas e a realização da krigagem, possibilitando a elaboração das PVG obtidas neste trabalho. Sendo assim, o ambiente de análises espaciais proporcionado pelo SIG tem muito a contribuir para a evolução dos sistemas de valoração cadastral rural.

A combinação de metodologias proposta neste trabalho mostrou-se 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. A combinação das metodologias tornou possível a obtenção dos produtos definidos nos objetivos do trabalho.

Todos os programas utilizados neste trabalho encontram-se disponíveis aos profissionais e instituições interessados por se tratarem de programas livres e gratuitos. Isto mostra que, apesar de existirem programas comerciais capazes desenvolver estas análises, a limitação de acesso a estas técnicas não é financeira. Espera-se que isto facilite a difusão destas metodologias devido ao fato que o investimento maior necessário pode ser concentrado na capacitação para utilização de forma correta destas ferramentas.

5.2 RECOMENDAÇÕES PARA TRABALHOS FUTUROS

Neste trabalho, em virtude dos dados que estavam disponíveis optou-se por trabalhar com um modelo para representar todo o município. Sugere-se que quando ocorram diferenças muito grandes entre regiões de um mesmo município, desde que existam dados suficientes deve-se buscar construir modelos específicos para as

diferentes regiões, com isso espera-se que ocorra uma menor dispersão nos resultados da avaliação de desempenho dos modelos.

Cada vez mais os países buscam a obtenção de variáveis para construção dos modelos de valoração dos imóveis rurais utilizando-se do sensoriamento remoto. Sugere-se a utilização desta técnica para obtenção de variáveis e a avaliação destas na melhoria dos trabalhos de valoração cadastral rural.

Uma vez obtidos o modelos de valoração cadastral rural e a Planta de Valores Genéricos sugere-se o desenvolvimento de trabalhos que possibilitem a utilização destas ferramentas integradas num ambiente SIG de modo que seja possível a avaliação das parcelas de forma cada vez mais automatizada.

Trabalhos recentes na avaliação em massa de imóveis urbanos apontam na direção da utilização de uma classe de modelos de regressão denominada de modelos aditivos generalizados para posição, escala e forma, em inglês Generalized Additive Models for Location, Scale and Shape, GAMLSS. Sugere-se avaliar a possibilidade de utilização destes modelos na avaliação em massa de imóveis rurais.

REFERÊNCIAS

- ALCÁZAR M. G. M.; ARIZA L. F. J. **Situación actual de la valoración catastral rústica en España: propuesta de un modelo alternativo.** Revista CT/Catastro. 54: 85-111. 2004.
- ALCÁZAR M. G. M. **Proposal For Rustic Cadastral Valuation.** In: FIG Working Week 2012, Rome, 6-10 May 2012 – Knowing to manage the territory, protect the environment, evaluate the cultural heritage. Rome: FIG. 2012. ISBN: 97887-90907-98-3. 24 p.
- ALMEIDA, E. **Econometria espacial aplicada.** Campinas, SP: Editora Alínea, 2012.
- ANSELIN, L. **The moran scatterplot as an ESDA tool to asses local instability in sapatial association.** In: Spatial Analytical Perspectives on GIS in Environmental and Socio-Economic Sciences. London: Taylor and Francis, 1996, p. 111-125.
- 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.
- ANSELIN, L. **Spatial externalities, spatial multipliers, and spatial econometrics.** International regional science review, v. 26, n. 2, p. 153-166, 2003.
- ANSELIN, L. **GeoDa™ 0.9 User's Guide.** Spatial Analysis Laboratory, Department of Agricultural and Consumer Economics, University of Illinois, Urbana-Champaign. Center for Spatially Integrated Social Science, 2003.
- ANSELIN, L. **Exploring Spatial Data with Geoda™: A Workbook.** Spatial Analysis Laboratory and Center for Spatially Integrated Social Science (CSISS). Department of Geography, University of Illinois, Urbana-Champaign, 2005.
- ANSEEUW, W. et al. **Land Rights and the Rush for Land: Findings of the Global Commercial Pressures on Land Research Project.** Rome: International Land Coalition, 2012. 84 p.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-1: avaliações de bens – parte 1: procedimentos gerais.** 1. ed .Rio de Janeiro: ABNT, 2001.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 14653-3: avaliações de bens – parte 3: imóveis rurais.** 1. ed. Rio de Janeiro: ABNT, 2004.
- AVERBECK, C. E. **Os sistemas de cadastro e planta de valores no município: prejuízos da desatualização.** 2003, 200 f. Dissertação (Mestrado em Engenharia Civil) – Curso de Pós-Graduação em Engenharia Civil, Universidade Federal de Santa Catarina – UFSC, Florianópolis, SC, 2003.
- AVERBECK, C. E.. **Os municípios e o ITBI: imposto sobre a transmissão de bens imóveis e de direitos a eles relativos.** In: CUNHA, E. M. P.; CESARE, C. M. de.Financiamento das Cidades: Instrumentos Fiscais e de Política Urbana – SEMINÁRIOS NACIONAIS. Brasília: Ministério das Cidades, 2007. Cap. 3. p. 79-84.
- BARROS, J. R. M. de. **O Brasil e a agricultura mundial.** O Estado de S. Paulo, São

Paulo, v. 5, 2012.

BAUMONT C. **Spatial Effects in Housing Price Models; Do Housing Prices Capitalize Urban Development Policies In The Agglomeration Of Dijon (1999)?**. Université de Bourgogne, 2004. Disponível em: <http://leg.u-bourgogne.fr/documents-de-travail/e2004-04.pdf>. Acessado em: novembro de 2013.

BENIGNI, M. **Los bienes rurales y el catastro**. Revista de la Red de Expertos Iberoamericanos en Catastro. n. 2, 2008.

BENIGNI, M. **Valuación y Descripción de Parcelas Rústicas. Nuevo Diseño de Valuación**. Revista de la Red de Expertos Iberoamericanos en Catastro, n. 7, 2010.

BIVAND, R.; LEWIN-KOH N. (2014). **Maptools: Tools for reading and handling spatial objects**. R package version 0.8-30. <http://CRAN.R-project.org/package=maptools>.

BIVAND, R.; KEITT T.; ROWLINGSON B. (2014). **rgdal: Bindings for the Geospatial Data Abstraction Library**. R package version 0.8-16. <http://CRAN.R-project.org/package=rgdal>.

BIVAND, R. (2014). **spdep: Spatial dependence: weighting schemes, statistics and models**. R package version 0.5-74. <http://CRAN.R-project.org/package=spdep>.

BONILLA, D. E.; ROBLEDO, E. C.; LÓPEZ, Y. M. A. **Nueva Metodología para el Avalúo de Inmuebles Rurales. Valor Económico por Parcela Rural en San Luis. Análisis para Optimizar su Aplicación**. Revista de la Red de Expertos Iberoamericanos en Catastro, n. 7, 2010.

BRASIL. Lei nº 5.172, de 26 de outubro de 1966. **Dispõe sobre o Sistema Tributário Nacional e institui normas gerais de direito tributário aplicáveis a União, Estados e Municípios**. Código Tributário Nacional. Disponível em: http://www.planalto.gov.br/ccivil_03/leis/l5172.htm. Acesso em: 16 set. 2012.

BRASIL. Constituição (1988). **Constituição nº 8, de 05 de outubro de 1988. Constituição da República Federativa do Brasil**. Disponível em: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Acesso em: 08 maio 2013.

BRASIL, RECEITA FEDERAL DO. **ITR Convênio - Consulta de Optantes pelo Convênio**. Disponível em: <http://www.receita.fazenda.gov.br/Aplicacoes/ATSDR/TermoITR/default.asp>. Acessado em 05 out. 2014.

BUENO, F. R.; BARRÓS, D. M.; MASEDA, R. C. **Situación de la valoración catastral rural**. CT/Catastro. n. 75, p. 41-54, 2012.

CAMBARDELLA, C. A. et al. **Field-scale variability of soil properties in central Iowa soils**. Soil Science Society of America Journal, v. 58, n. 5, p. 1501-1511, 1994.

CAMARGO, E. C. G.; FUCKS, S. D.; CÂMARA, G. **Análise espacial de superfícies**. In: DRUCK, S. et. al. (eds). **Análise espacial de dados geográficos**. Brasília, EMBRAPA, p. 79-122, 2004.

CAPES. **Portal de Periodicos Capes: Institucional**. Disponível: http://www.periodicos.capes.gov.br/index.php?option=com_painstitucional&mn=69. Acessado em: jan. 2014.

CARNEIRO, A. F. T.; ERBA, D. A.; AUGUSTO, E. A. A. **Cadastro multifinalitário 3d: conceitos e perspectivas de implantação no Brasil**. Revista Brasileira de Cartografia, n. 64 ESP. 1, 2012.

CEPEA. **PIB do Agronegócio - Dados de 1994 a 2011**. Centro de Estudos Avançados em Economia Aplicada – ESALQ/USP. Disponível em: <http://www.cepea.esalq.usp.br/pib/other/Pib_Cepea_1994_2011.xls>. Acesso em: 21 jan. 2014.

CESARE, C. M. de. **A tributação sobre a propriedade imobiliária e o IPTU: fundamentação, caracterização e desafios**. In: CUNHA, E. M. P.; CESARE, C. M. de. Financiamento das Cidades: Instrumentos Fiscais e de Política Urbana – SEMINÁRIOS NACIONAIS. Brasília: Ministério das Cidades, 2007. Cap. 2. p. 45-78.

CONFEDERAÇÃO NACIONAL DOS MUNICÍPIOS (CNM). **Informativo CNM Finanças: Imposto Territorial Rural e os municípios**. Maio de 2012. Disponível em: <http://www.cnm.org.br/index.php?option=com_docman&task=doc_download&gid=562&Itemid=4>. Acesso em: 19 set. 2012.

DALAQUA, R. R.; AMORIM, A.; FLORES, E. **Utilização de métodos combinados de avaliação imobiliária para a elaboração da planta de valores genéricos**. Boletim de Ciências Geodésicas, v. 16, n. 2, p. 232-252, 2010.

DANTAS, R. et al. **Uma nova metodologia para avaliação de imóveis utilizando regressão espacial**. In: Anais do XI Congresso Brasileiro de Engenharia de Avaliações e Perícias. Guarapari/ES. 2001.

DANTAS, R. A. **Modelos Espaciais Aplicados ao Mercado Habitacional – Um Estudo de Caso para a Cidade do Recife**. 2003. 133 f. Tese de doutorado – Programa de Pós-Graduação em Economia, Universidade Federal de Pernambuco, Recife, 2003.

DANTAS, R. A.; MAGALHÃES, A. M.; VERGOLINO, J. R. O. **Modelos Espaciais Aplicados ao Mercado de Apartamentos do Recife**. In: Anais do XII Congresso Brasileiro de Engenharia de Avaliações e Perícias. Belo Horizonte/MG. 2003.

DANTAS, R. A.; SÁ, L. A. C. M.; PORTUGAL, J. L. **Elaboração de Planta de Valores sob a Ótica da Inferência Espacial**. In: Anais do XII Congresso Brasileiro de Engenharia de Avaliações e Perícias. Belo Horizonte/MG. 2003.

DANTAS, R. A. (2005). **Engenharia de Avaliações: uma introdução à metodologia científica**, Pini, São Paulo.

DANTAS R. A.; PORTUGAL J. L.; PRADO J. F. **Avaliação de cidades por inferência espacial: um estudo de caso para a cidade de Aracaju**. In: Anais do XIII Congresso Brasileiro de Engenharia de Avaliações e Perícias. Fortaleza/CE, 2006.

DANTAS, R. A.; MAGALHÃES, A. M.; VERGOLINO, J. R. de O. **Avaliação de imóveis: a importância dos vizinhos no caso de Recife**. Economia Aplicada, v. 11, n. 2, p. 231-251, 2007.

DANTAS, R. A.; MAGALHÃES, A. M.; VERGOLINO, J. R. de O. **Um modelo espacial de demanda habitacional para a cidade do Recife**. Estudos Econômicos (São Paulo), v. 40, n. 4, p. 891-916, 2010.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (eds) "**Análise Espacial de Dados Geográficos**". Brasília, EMBRAPA, 2004.

DUBIN, R. A. **Spatial autocorrelation and neighborhood quality**. *Regional Science and Urban Economics*, v. 22, n. 3, p. 433-452, 1992.

EMBRAPA. **Mapa Exploratório-Reconhecimento de solos do município de Petrolina, PE**. EMBRAPA SOLOS UEP-RECIFE, 2001.

FARACO, M. A. et al. **Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja**. *Revista Brasileira de Ciência do Solo*, 32, n. 2, p. 463-476, 2008.

FLORAX.; GRAAF, T. **The performance of diagnostics tests for spatial dependence in linear regression models: a meta-analysis of simulation studies**. In: ANSELIN, L.; FLORAX, R. J. G. M.; REY, S. J. (ed.). *Advances in spatial econometrics*. Nova York: Springer, 2004. p. 29-65.

FERMO, G. O. **Influência da ferramenta de georreferenciamento na avaliação de apartamentos usando regressão espacial**. 2010. 151 f. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Civil, Florianópolis, 2010.

FEICHTINGER, P.; SALHOFER, K. **A Spatial Analysis of Agricultural Land Prices in Bavaria**. *Factor Markets: Comparative Analysis of Factor Markets for Agriculture across the Member States*. Factor Markets Working Paper N°. 50/June 2013. Disponível em: <<http://www.factormarkets.eu/content/spatial-analysis-agricultural-land-prices-bavaria>>. Acesso em 20 de dez. 2013.

FIGUEIREDO, A. M. R.; TEIXEIRA, E. C.; BONJOUR, S. C. DE M. **Introdução à Econometria Espacial na Agricultura**. DESAFIO: *Revista Economia e Administração*, Campo Grande, MS, v. 10, n. 21, p. 16-28, [s.d.] mai./ago. 2009.

FLORENCIO, L. A. **Engenharia de avaliações com base em modelos GAMLSS**. 2010. 125 f. Dissertação (mestrado) - Universidade Federal de Pernambuco. CCEN. Estatística, 2010.

FLORENCIO, L. de A. **Modelos Espaciais para Avaliações de Imóveis em Massa com Base em Modelos de Regressão GAMLSS**. In: *Anais da 12ª Conferência Internacional da Latin American Real Estate Society*. Centro Britânico, São Paulo – Brasil, 2012.

FLORENCIO, L. A.; CRIBARI-NETO, F.; OSPINA, R. **Real estate appraisal of land lots using GAMLSS models**. *Chilean Journal of Statistics* 3, 75–91, 2012.

GOOVAERTS, P. **Geostatistics for Natural Resources Evaluation**. New York: Oxford University Press, 1997. 483 p.

HOAGLIN, DAVID C.; WELSCH, ROY E. **The hat matrix in regression and ANOVA**. *The American Statistician*, v. 32, n. 1, p. 17-22, 1978.

HORNBURG, R. A.; HOCHHEIM, N. **Elaboração de planta de valores genéricos com uso combinado de regressão espacial e krigagem**. In: *XV Congresso Brasileiro de Engenharia de Avaliações e Perícias*. São Paulo, 2009.

HORNBURG, R. A. **Avaliação em massa com uso combinado da regressão**

espacial e da geoestatística. 2009. 224 f. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico, Programa de Pós-Graduação em Engenharia Civil, Florianópolis, 2009.

INSTITUTO BRASILEIRO de GEOGRAFIA e ESTATÍSTICA - IBGE. **Indicadores IBGE: Estatística da Produção Agrícola.** Dezembro de 2013. Instituto Brasileiro de Geografia e Estatística. Disponível em: [ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_\[mensal\]/Comentarios/lspa_201312comentarios.pdf](ftp://ftp.ibge.gov.br/Producao_Agricola/Levantamento_Sistematico_da_Producao_Agricola_[mensal]/Comentarios/lspa_201312comentarios.pdf). Acesso em 21 jan. 2014.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA-IBGE. **Produção Agrícola Municipal.**, Rio de Janeiro, v. 40, p.1-102, 2013.

INSTITUTO BRASILEIRO de GEOGRAFIA e ESTATÍSTICA - IBGE. **Censo agropecuário: 2006: segunda apuração: Brasil, grandes regiões e unidades da federação, mesorregiões, microrregiões geográficas e municípios.** Rio de Janeiro, IBGE, 2012.

INCRA. **Relação total de imóveis rurais Brasil (abril 2012).** Instituto Nacional de Colonização e Reforma Agrária. 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: 28 jan. 2014.

ISAAKS, E. H. e SRIVASTAVA, R. M. **An introduction to applied geostatistics.** New York: Oxford University Press, 1989. 561 p.

INTERNATIONAL ASSOCIATION of ASSESSING OFFICERS (IAAO). **Standard on ratio studies.** Kansas City: IAAO, 2013.

JOURNEL, A. G. e HUIJBREGTS, J. C. H. **Mining geostatistics.** New York: Academic Press, 1978. 600 f.

LANDIM, P. M. B. **Sobre Geoestatística e mapas.** Terrae Didatica, v. 2, p. 19-33, 2006.

LAKATOS, E. M.; MARCONI, M. de A. **Metodologia científica.** São Paulo: Atlas, 1991.

LAKATOS, E. M.; MARCONI, M. A. **Fundamentos de pesquisa metodológica científica.** São Paulo: Atlas, 2003.

IHAKA, R.; GENTLEMAN, R. **"R: A Language for Data Analysis and Graphics".** Journal of Computational and Graphical Statistics 5 (3): 299–314. 1966.

LIMA, M. R. C. **Avaliação de Propriedades rurais: manual básico.** – 2. ed. rev. e atual. – São Paulo: Liv. e Ed. Universitária de Direito, 2005.

LIPORONI, A. S. **Avaliações em massa com ênfase em planta de valores (Capítulo 21).** In: IBAPE-SP. **Engenharia de Avaliações.** 1ª Edição. São Paulo. Editora Pini, 2007.

LOURIVAL, R.; CALEMAN, S. M. D. Q.; VILLAR, G. I. M.; RIBEIRO, A. R. e 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.

MACHADO, E. F. **Elaboração de planta de valores genéricos para zona rural com uso de superfície de tendência.** 2006. 105 f. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia Civil. 2006.

MARQUES, A. P. da S. et al. **Aplicação do interpolador krigagem ordinária para a elaboração de planta de valores genéricos.** Revista Brasileira de Cartografia, [s. L.], v. 2, n. 64, p.175-186, abr. 2012.

MATHERON, G. **The Theory of regionalized variables and its applications.** Paris School of Mines Publication, p. 211, 1971.

MICHAEL, R. **Avaliação em massa de imóveis com uso de inferência estatística e análise de superfície de tendência.** 2004. 91 f. Dissertação (Mestrado) - Curso de Pós-graduação em Engenharia Civil, Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2004.

MICHAEL, R.; HOCHHEIM, N.; TRIVELLONI, C. A. P. **Avaliação em massa de imóveis com uso de inferência estatística e análise de superfície de tendência.** In: Anais do XIII Congresso Brasileiro de Engenharia de Avaliações e Perícias. Fortaleza, Brasil, 2006.

NANNI A. S. et al.. **Quantum GIS – Guia do Usuário, Versão 1.7.4 'Wroclaw'.** Acesso em: janeiro de 2014. Disponível em: <http://qgisbrasil.org>. 291p., il.

OLIVEIRA, T. A. M.. **Imposto Territorial Rural: um Estudo Econômico sobre a Descentralização da Cobrança.** 2010. 135 f. Tese (Doutorado) - Curso de Pós Graduação em Desenvolvimento Econômico, Instituto de Economia, Unicamp, Campinas, 2010.

PAIXÃO, S. K. S.; NICHOLS, S. e CARNEIRO, A.F.T. **Cadastro Territorial Multifinalitário: dados e problemas de implementação do convencional ao 3D e 4D.** Boletim de Ciências Geodésicas, v. 18, n. 1, p. 3-21, 2012.

PEBESMA, E.J., R.S. BIVAND, 2005. **Classes and methods for spatial data in R.** R News 5 (2), <http://cran.r-project.org/doc/Rnews/>.

RIBEIRO P. J. JR & DIGGLE P. J. **geoR: a package for geostatistical analysis.** R-NEWS, 1(2):15-18. June, 2001.

R Development Core Team (2013). **R: A language and environment for statistical computing.** R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.

RESENER, M. C. **Avaliação de imóveis rurais por inferência estatística e análise de superfície de tendência.** 2005. 186 f. Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia Civil. 2005.

RESENER, M.; HOCHHEIM, N. **Avaliação de imóveis rurais por inferência estatística e análise de superfície de tendência.** In: Anais do XIII Congresso Brasileiro de Engenharia de Avaliações e Perícias. Fortaleza, Brasil, 2006.

REYDON, B. P.; OLIVEIRA, T. A. M. **A descentralização fiscal do itr através do cadastro territorial multifinalitário.** In: Anais do IV Simpósio Brasileiro de Ciências

Geodésicas e Tecnologias da Geoinformação. Recife – PE, 2012.

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**. In: XV Congresso Brasileiro de Engenharia de Avaliações e Perícias, 2009, São Paulo, 2009.

SILVA, D.; BARRETO, P. **O potencial do Imposto Territorial Rural contra o desmatamento especulativo na Amazônia**. Belém: Imazon, 2014. 48 p.

STEINER, M. T. A.; NETO, A. C.; BRAULIO, S. N.; ALVES, V. **Métodos estatísticos multivariados aplicados à engenharia de avaliações**. *Gestão e Produção*, São Carlos, v. 15, n. 1, p. 23-32, jan.-abr. 2008.

SUNAK, Y.; MADLENER, R. **The Impact of Wind Farms on Property Values: A Geographically Weighted Hedonic Pricing Model** (May 1, 2012). FCN Working Paper No. 3/2012 (revised March 2013). Disponível em: <http://ssrn.com/abstract=2114216> or <http://dx.doi.org/10.2139/ssrn.2114216> Acesso em: 20 jan. 2014.

TACONELI, C. A. **Análise de Regressão Linear**. Universidade Federal do Paraná. Departamento de Estatística. Curitiba, 2010. Disponível em: <<http://people.ufpr.br/~niveam/ce071/Apostila.pdf>>. Acessado em: agosto de 2014.

THE APACHE SOFTWARE FOUNDATION. **Apache OpenOffice 4.0.1**. <http://www.openoffice.org>.

TRIVELLONI, C. A. P. **Método para determinação do valor da localização com uso de técnicas inferenciais e geoestatísticas**. 2005. 172 f. Tese (Doutorado) - Curso de Pós-graduação em Engenharia Civil, Engenharia Civil, Universidade Federal de Santa Catarina, Florianópolis, 2005.

TRIVELLONI, C. A. P. e HOCHHEIM, N. **O valor de localização dos imóveis: determinação por métodos de análise espacial**. In: Anais do XIII Congresso Brasileiro de Engenharia de Avaliações e Perícias. Fortaleza, Brasil, 2006.

TOBLER, W. R. **A computer movie simulating urban growth in the Detroit region**. *Economic geography*, v. 46, p. 234-240, 1970.

VIEIRA, R. S. **Crescimento econômico no estado de São Paulo: uma análise espacial** [online]. São Paulo: Editora UNESP; São Paulo: Cultura Acadêmica, 2009. 103 p. ISBN 978-85-7983-013-6. Available from SciELO Books <<http://books.scielo.org>>.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística. Conceitos e Aplicações**. 1. ed. São Paulo: Oficina de Textos, 2013. 215 p.

UNITED NATIONS. Department of Economic and Social Affairs. **World population prospects: the 2012 revision: total population – both sexes**. Disponível em: <<http://esa.un.org/unpd/wpp/Excel-Data/population.htm>>. Acesso em: 02 out. 2013.

UNITED NATION ENVIRONMENT PROGRAM-UNEP. **Assessing Global Land Use: Balancing Consumption with Sustainable Supply**. Bringezu S., Schütz H., Pengue W., O'Brien M., Garcia F., Sims R., Howarth R., Kauppi L., Swilling M., and Herrick J. 2014.

APÊNDICE 01

Formulário de pesquisa de campo

1- **Dados do entrevistado** – *Comprador* ___ *Vendedor* ___

1.1. Nome : _____

1.2. Profissão: _____

1.3. Escolaridade:

Nenhuma ___ *1º grau incompl.* ___ *1º grau compl.* ___ *2º grau incompl.* ___
2º grau compl. ___ *Superior incomp.* ___ *Superior compl.* ___

2- Imóvel:

2.1. Denominação: _____

2.2. Coordenadas de localização: N _____; E _____ WGS84

2.3. Área (ha): _____

2.4. Acesso: _____

2.5. Quantos quilômetros? _____

2.6. Energia elétrica instalada:

Monofásica ___ *Trifásica* ___ *Não possui* ___

2.7. Energia disponível na rede pública a – 10km:

Monofásica ___ *Trifásica* ___ *Não possui* ___

2.8. Disponibilidade hídrica:

Rio perene ___ *Rio temporário* ___ *Riacho* ___ *Açude* ___ *Poço tubular* _____

Poço amazonas ou cacimba ___ *Não havia* ___ *Outra* _____

2.9. Qual o potencial de uso da água no imóvel?

Para irrigação ___ *Para os animais* ___ *Indisponível* ___

2.10. Aproveitamento e distribuição da água no imóvel:

Toda a área do imóvel ___ *Parte do imóvel* ___ *Próximo a fonte* ___

2.11 – Sistema de produção principal:

Fruticultura irrigada ___ *Olericultura irrigada* ___ *pastagem irrigada* ___ *Agricultura de subsistência* ___ *Pecuária caprina/ovina* ___ *Pecuária bovina/leiteira* ___ *Pecuária bovina/corte* ___ *Pecuária mista* ___ *Outra:* _____

2.12. Área destinada ao sistema produtivo (%): _____

2.13. Infraestrutura: *Adequada* __ *Muita* __ *Insuficiente* __

2.14. Descrição das benfeitorias: _____

2.15. Estado geral das benfeitorias:

Não necessita de reparos __ *Necessita de reparos* __ *Sem condições de uso* __

2.16. Outro sistema produtivo, qual: _____

2.17. Área destinada (%): _____

2.18. Infraestrutura:

Adequada __ *Muita* __ *Insuficiente* __; Estado de conservação: *Não neces. de reparos* __ *Necessitava de reparos* __ *Sem condições* __

2.18. Possuía título de propriedade: *Sim* __ *não* __

2.19. O imóvel possuía APP? *Sim* __ *Não* __ Reserva legal? *Sim* __ *Não* __

2.20. Reserva legal: *Averbada* __ *Não averbada e demarcada* __ *Não possuía* __

2.21. Estado de preservação da APP e RL:

Preservada __ *Parcialmente preservada* __ *Degradada* __

2.22. Qualidade da terra do imóvel (%):

Baixios _____; Chapadas: _____ Terra ruim: _____ Outras: _____

2.23. Relevo geral: *Plano* __ *Suave* __ *Ondulado* __ *Montanhoso* __

2.24. Aptidão natural da terra:

Plantio __ *Pastagem* __ *Silvicultura* __ *Preservação* __ *Outros* _____

2.25. Textura do solo: *Argiloso* __ *Arenoso* __ *Média* __

2.26. Culturas perenes ou pastagens:

área, tipo e idade: _____

3. Preço do imóvel:

3.1. Data da compra: ____ / ____ / ____

3.2. Valor pago (R\$): _____

3.3. Forma de pagamento: À vista __ A prazo (Nº de vezes _____)

3.4. Valor das benfeitorias (R\$): _____

3.5. Valor proporcional das benfeitorias (%): _____

3.6. Na época da compra o imóvel estava:

Abaixo do valor de mercado ___ *Valor de mercado* ___ *Acima do valor de mercado* ___

4. Valor de venda:

4.1. Quanto vale o imóvel, hoje? (R\$): _____

4.2. Quanto você acha que pagariam hoje (R\$): _____

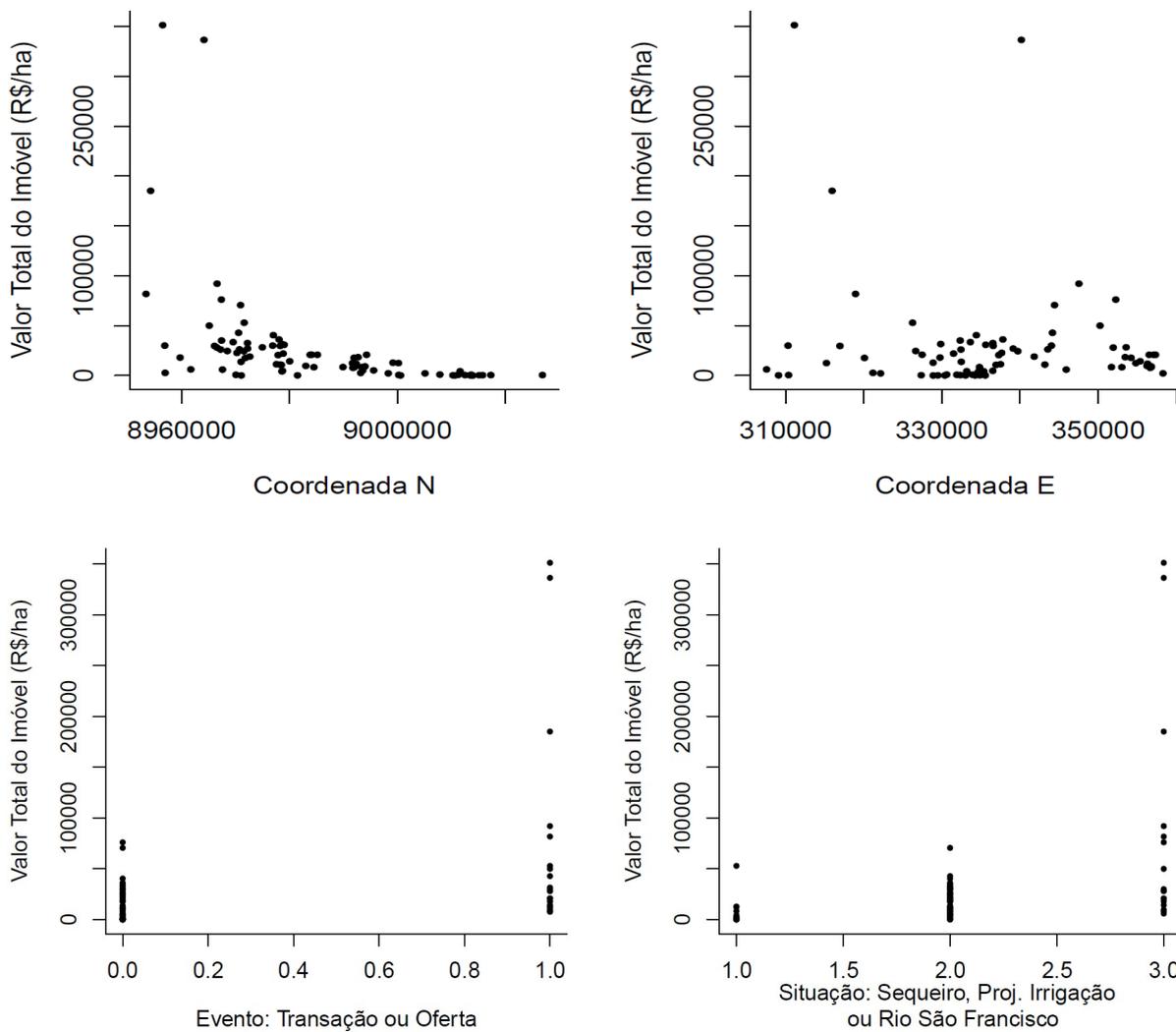
5. Telefone do entrevistado: _____

6. E-mail do entrevistado: _____

7. Data da entrevista: ____ / ____ / ____ Hora: _____

APÊNDICE 02

Gráficos de dispersão entre a variável dependente e as variáveis independentes.



Figuras 41 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.

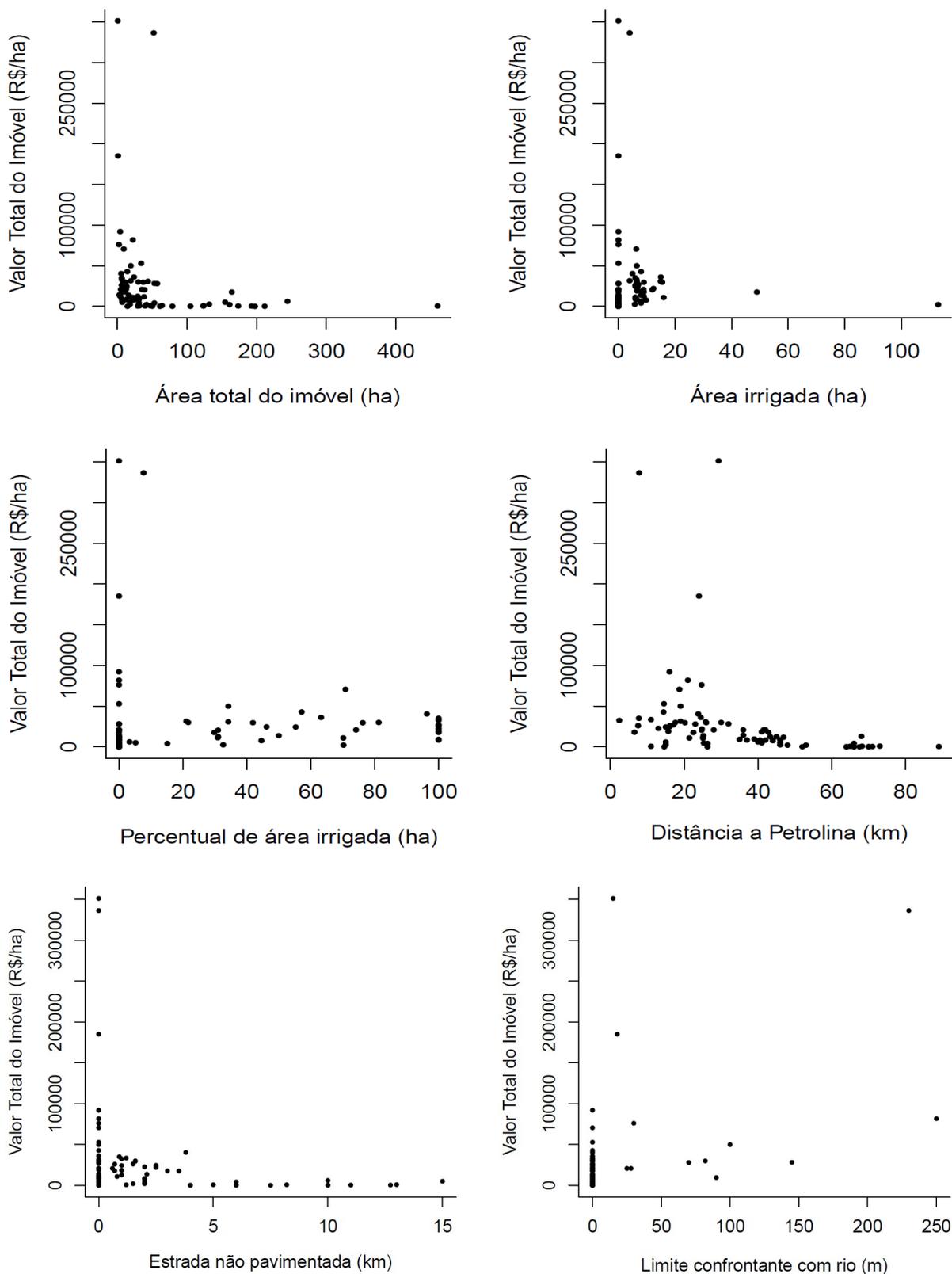


Figura 42 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.

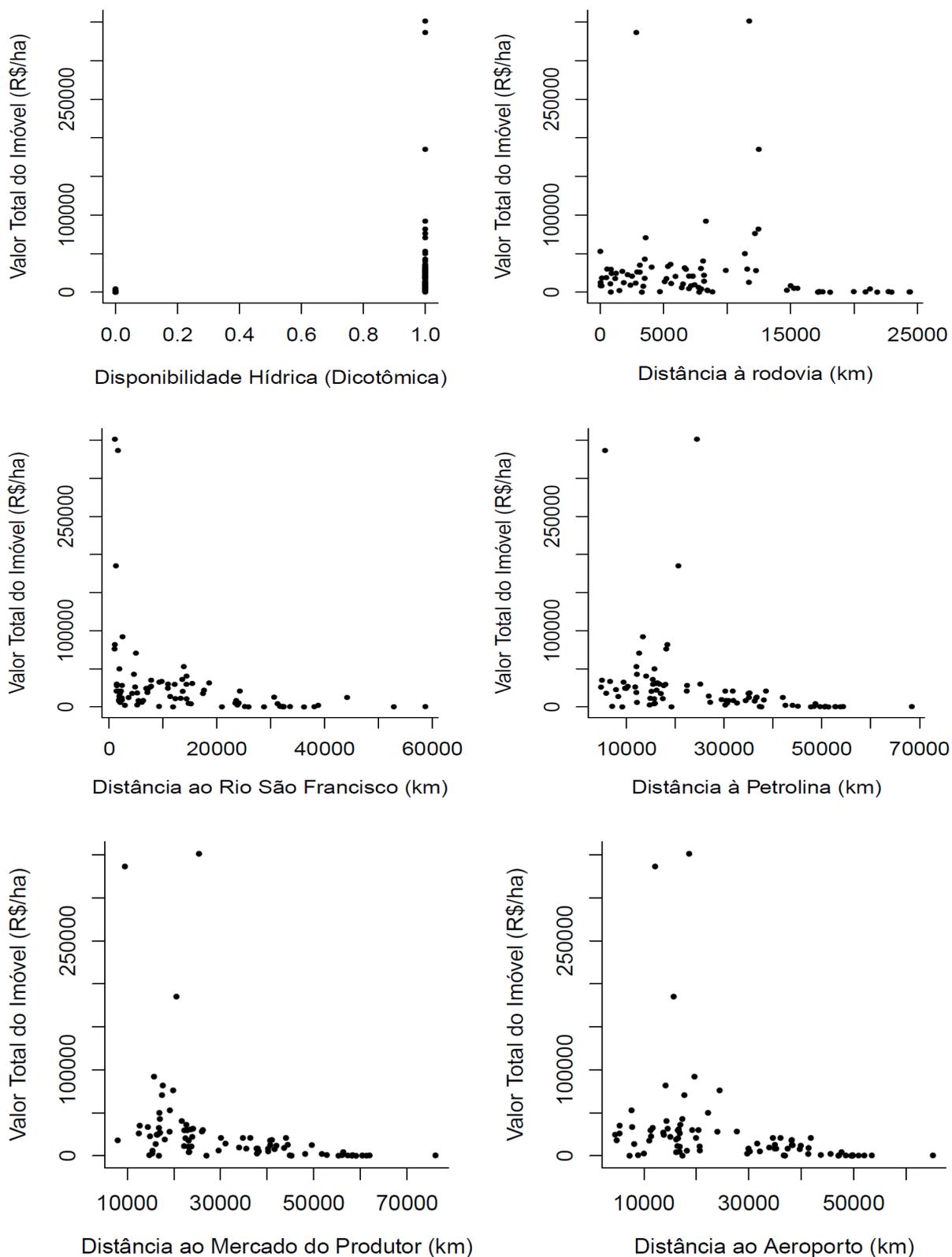


Figura 43 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.

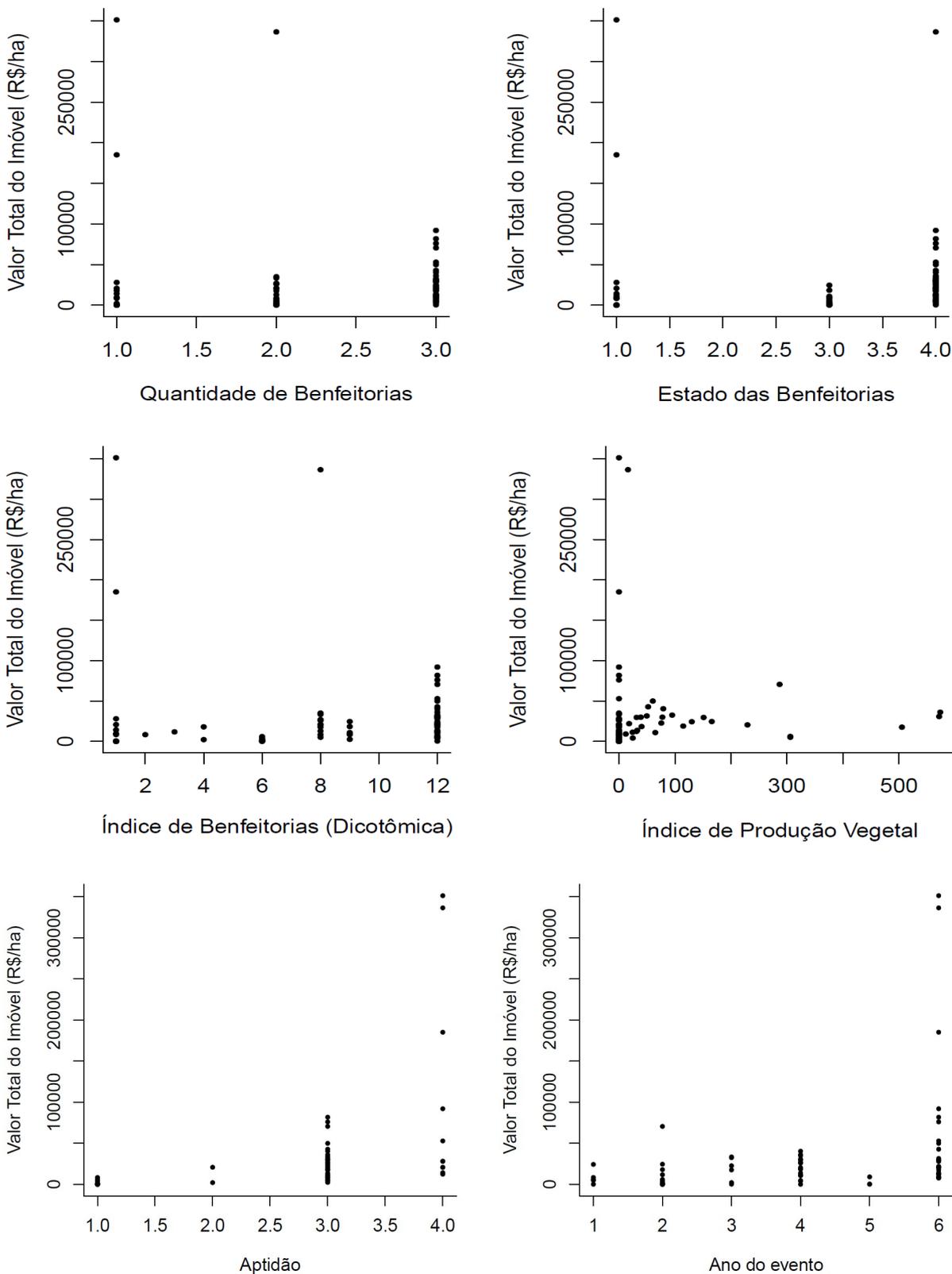


Figura 44 - Gráficos de dispersão entre a variável dependente e as variáveis independentes.

APÊNDICE 03

A seguir estão disponibilizadas as 84 observações utilizadas na amostra de trabalho bem como as 20 observações utilizadas como pontos de verificação. As variáveis estão identificadas por siglas e codificadas como mostrado no item 4.3.

1-Amostra de trabalho composta por 84 observações.

N	VTI	DN	DE	NE	SIT	AT	AI	LCR	DS	DSP	DH	AP	QB	EB	IB	IPV	ANO	DAP	DR	DRSF	DC	DMP
1	250	68,61	24,07	0	1	50	0	0	52	0	0	1	1	1	1	0	5	49,84	0,84	52,84	53,80	61,33
2	24390,24	29,86	54,72	0	2	12,3	6,8	0	15	1	1	3	3	4	12	130,1	1	13,79	0,90	6,90	9,96	16,45
3	337,28	73,48	42,34	0	1	192,72	0	0	64	4	1	1	1	1	1	0	3	50,96	17,25	38,09	52,71	60,44
4	478,26	85,22	25,35	0	1	460	0	0	89	12,73	1	1	3	4	12	0	5	65,15	8,84	58,69	68,40	76,10
5	70627,8	29,25	59,42	0	2	8,92	6,32	0	18,7	0	1	3	3	4	12	286,89	2	17,71	3,58	4,99	12,61	17,45
6	24615,38	26,79	41,63	0	2	13	6	0	16	2,5	1	3	3	3	9	165,82	2	4,50	1,25	10,91	9,73	16,33
7	2584,27	51,51	49,54	0	2	17,8	5,8	0	46	2	1	3	3	3	9	0	2	29,71	14,70	23,82	30,23	37,83
8	22727,27	28,58	52,65	0	2	6,6	6,6	0	13	2	1	3	3	4	12	75,67	3	11,36	2,16	7,15	7,89	14,89
9	30000	35,21	59,03	0	2	8	6,5	0	17,6	1,6	1	3	3	4	12	39,11	4	20,41	0,55	10,94	16,78	22,77
10	10964,91	36,52	58,18	0	2	22,8	16	0	21,4	0,8	1	3	3	3	9	64,87	4	20,62	0,80	12,30	17,46	23,76
11	9166,67	52,32	71,81	0	2	6	6	0	35	0	1	3	3	3	9	12,11	5	41,42	2,38	2,39	38,14	43,64
12	29661,02	24,43	31,93	0	2	11,8	9	0	20,2	1,6	1	3	3	4	12	151,18	6	11,27	0,85	12,17	18,00	22,25
13	22016,13	37,15	46,48	0	2	12,4	12,4	0	24,8	2,5	1	3	3	4	12	18,16	6	15,03	8,17	17,62	16,18	23,93
14	18018,99	18,04	44,75	0	3	11,1	0	0	6,5	0,7	1	3	1	4	4	0	2	4,79	3,50	2,01	5,88	7,96
15	18404,91	51,05	68,47	0	2	8,15	8,15	0	41	1	1	3	3	3	9	40,41	4	38,22	0,14	5,21	35,21	41,02
16	8461,54	51,69	49,8	0	2	6,5	6,5	0	46	2	1	3	3	3	9	0	1	29,94	14,99	23,60	30,41	37,99
17	5147,06	51,9	50,03	0	2	6,8	0	0	46	2	1	3	2	4	8	0	1	30,20	15,25	23,42	30,62	38,20
18	5973,63	25,9	60,92	0	3	6,7	0	0	15	0	1	1	2	3	6	0	2	18,19	6,42	2,07	12,17	15,42
19	5176,52	53,92	49,79	0	2	154,54	8	0	41	15	1	3	3	4	12	306,01	1	32,12	15,53	24,06	32,64	40,22
20	2658	15,28	36,16	0	1	131,68	0	0	15	0	1	3	2	3	6	0	2	9,99	8,43	5,23	14,77	15,36
21	11842,11	50,69	71,34	0	2	38	0	0	46,9	0	1	3	3	1	3	0	2	39,93	2,80	2,35	36,54	41,97
22	40462,43	35,33	49,4	0	2	5,19	5	0	23,8	3,8	1	3	3	4	12	79,18	4	14,30	8,10	14,38	14,05	21,69
23	17656,01	30,13	35,07	0	2	164,25	48,92	0	22,5	3,5	1	3	3	4	12	505,43	3	11,02	5,22	17,43	17,05	23,17
24	26043,41	25,56	47,42	0	2	5,99	5,99	0	7,5	0,7	1	3	2	4	8	0	4	5,33	3,12	7,63	4,82	12,49
25	35117,06	25,72	47,33	0	2	5,98	5,98	0	7,7	0,9	1	3	2	4	8	0	4	5,35	3,12	7,81	5,01	12,68
26	33492,82	27,87	48,63	0	2	6,27	6,27	0	11	1,2	1	3	2	4	8	0	3	7,74	5,34	9,76	6,67	14,42
27	27027,03	30,55	54,12	0	2	7,03	7,03	0	17,1	0	1	3	2	4	8	0	4	13,69	1,74	7,80	10,29	17,03
28	32558,14	30,53	51,52	0	2	6,45	6,45	0	2,4	1	1	3	3	4	12	95,02	3	11,66	4,05	9,32	9,43	16,79

N	VTI	DN	DE	NE	SIT	AT	AI	LCR	DS	DSP	DH	AP	QB	EB	IB	IPV	ANO	DAP	DR	DRSF	DC	DMP
29	13750	29,34	47,47	0	2	16	8	0	25,2	2,1	1	3	3	4	12	32,71	4	8,13	5,08	11,34	8,35	16,10
30	18996,96	30,96	56,82	0	2	6,58	6,58	0	15,7	0	1	3	3	4	12	114,76	4	16,16	0,49	7,12	12,04	18,05
31	26234,57	29,04	58,56	0	2	6,48	6,48	0	16,2	1,5	1	3	2	4	8	0	4	16,84	2,91	4,83	11,81	16,88
32	30836,95	37,33	50,6	0	2	43,73	14,95	0	25,8	0	1	3	3	4	12	571,86	4	16,63	7,93	15,43	16,08	23,61
33	4161,15	36,92	50,36	0	2	52,87	8	0	26,34	0	1	3	3	4	12	24,65	4	16,15	7,96	15,18	15,65	23,21
34	29729,73	36,61	51,56	0	2	37	15,5	0	25,96	0	1	3	3	4	12	31,49	4	16,47	6,75	14,35	15,45	22,88
35	20487,99	36,2	52,28	0	2	38,73	12	0	24,71	0	1	3	3	4	12	229,51	4	16,50	5,93	13,66	15,14	22,48
36	36172,01	36,37	52,8	0	2	23,72	15	0	24,49	0	1	3	3	4	12	573,78	4	16,94	5,54	13,58	15,41	22,68
37	11328,53	35,9	52,51	0	2	19,42	6	0	25,07	0	1	3	3	4	12	24,22	4	16,39	5,59	13,29	14,89	22,19
38	10694,95	36,82	51,92	0	2	22,16	0	0	25,1	0	1	3	3	4	12	0	4	16,83	6,52	14,37	15,70	23,09
39	4721,27	37,03	51,51	0	2	17,58	0	0	25,3	0	1	3	3	4	12	0	4	16,80	6,99	14,75	15,85	23,29
40	988,57	66,19	45,61	0	1	63,8	0	0	73	13	0	1	2	3	6	0	2	43,75	17,25	31,81	45,09	52,78
41	160,91	58,91	49,25	0	1	211,44	0	0	66	6	0	1	2	3	6	0	2	36,92	17,14	25,88	37,63	45,23
42	164,01	69,03	50,55	0	1	78,87	0	0	70	10	0	1	1	1	1	0	2	47,12	22,94	28,76	47,76	55,31
43	4319,83	69,94	48,17	0	1	30,61	0	0	66	6	0	1	2	3	6	0	2	47,70	21,24	31,27	48,68	56,30
44	375,64	74,12	49,89	0	1	29,11	0	0	70	10	0	1	2	3	6	0	2	52,05	24,34	32,15	52,84	60,41
45	556,36	58,56	49,86	0	1	39,61	0	0	66	6	0	1	2	3	6	0	2	36,68	17,52	25,20	37,28	44,86
46	427,11	70,86	47,37	0	1	123,35	0	0	66	6	0	1	2	3	6	0	2	48,53	20,89	32,43	49,63	57,27
47	467,95	75,64	49,26	0	1	173,49	0	0	71	11	0	1	2	3	6	0	2	53,49	24,39	33,58	54,36	61,95
48	809,67	69,63	46,89	0	1	47,03	0	0	65	5	0	1	2	3	6	0	2	47,26	19,96	32,23	48,42	56,08
49	55,72	72,31	43,87	0	1	197,54	0	0	64	4	0	1	1	1	1	0	2	49,80	18,10	36,18	51,36	59,07
50	164,01	71,83	48,03	0	1	61	0	0	67,5	7,5	0	1	1	1	1	0	2	49,56	21,81	32,37	50,57	58,20
51	857,19	71,99	48,9	0	1	123,43	0	0	68,2	8,2	0	1	2	3	6	0	2	49,82	22,69	31,72	50,72	58,32
52	2193,4	63,45	37,14	0	1	41,59	0	0	53	2	0	1	2	3	6	0	2	41,35	8,46	38,81	43,99	51,73
53	766,95	28,38	49,71	0	2	31,41	0	0	11	1,2	0	1	2	3	6	0	2	8,88	4,70	9,28	7,10	14,75
54	196,12	39,81	44,49	0	2	104,75	0	0	26,3	0	0	1	1	1	1	0	1	17,35	7,80	20,94	19,23	26,99
55	135,5	29,38	45,37	0	2	14,31	0	0	14,6	0	0	1	1	1	1	0	4	7,25	3,28	11,87	9,16	16,79
56	6141,06	20,03	22,53	0	2	244,26	8	0	40	10	1	3	3	4	12	306,01	1	20,66	7,75	6,09	27,15	29,59

N	VTI	DN	DE	NE	SIT	AT	AI	LCR	DS	DSP	DH	AP	QB	EB	IB	IPV	ANO	DAP	DR	DRSF	DC	DMP
57	76142,13	25,7	67,27	0	3	1,97	0	30	24,7	0	1	3	3	4	12	0	6	24,43	12,19	1,03	18,16	19,84
58	12500	58,52	30,2	1	1	4	0	0	45	0	1	4	3	4	12	0	6	38,23	0,04	44,17	42,01	49,57
59	30000	15,21	25,28	1	3	30	6,5	82	30	0	1	3	3	4	12	77,56	6	19,21	11,57	1,42	25,12	26,20
60	14285,71	38,36	70,39	1	3	2,8	0	0	36	0	1	4	1	1	1	0	6	31,60	8,18	1,91	26,86	31,07
61	20800	42,14	71,54	1	3	6,25	0	25	41,5	0	1	4	1	1	1	0	6	34,60	7,27	1,84	30,24	34,80
62	28301,89	33,29	68,59	1	3	53	0	145	32	0	1	4	3	4	12	0	6	27,73	9,89	2,41	22,43	26,01
63	81818,18	11,73	33,93	1	3	22	0	250	21	0	1	3	3	4	12	0	6	14,12	12,46	1,07	18,40	17,63
64	28070,18	24,86	66,94	1	3	57	0	70	23	0	1	4	1	1	1	0	6	24,01	12,27	1,44	17,65	19,08
65	42857,14	28,88	59,16	1	2	14	8	0	14,4	0	1	3	3	4	12	52,15	6	17,33	3,51	4,61	12,17	17,00
66	7727,27	50,08	71,67	1	2	22	9,8	0	44	0	1	3	3	4	12	0	6	39,74	3,39	1,87	36,25	41,60
67	8333,33	50,36	68,04	1	2	12	0	0	42	2	1	3	2	4	8	0	6	37,42	0,11	5,47	34,40	40,22
68	336538,46	22,47	55,17	1	3	52	4	230	7,8	0	1	4	2	4	8	16,15	6	12,13	2,84	1,65	5,64	9,50
69	8469,06	48,23	66,71	1	1	30,7	0	0	37	0	1	1	1	1	1	0	6	34,95	0,04	6,34	31,90	37,73
70	20833,33	52,61	72,12	1	2	12	8,9	0	42	0	1	3	3	4	12	0	6	41,85	2,51	2,22	38,55	44,04
71	12413,79	49,98	69,84	1	2	29	9	0	43,32	0	1	3	3	4	12	31,29	6	38,37	1,86	3,63	35,09	40,65
72	185185,19	12,59	30,93	1	3	0,65	0	18	24	0	1	4	1	1	1	0	6	15,66	12,49	1,27	20,65	20,54
73	20857,14	42,45	42,46	1	2	35	0	0	28	0,6	1	2	2	4	8	0	6	19,94	6,99	24,27	22,36	30,09
74	351351,35	14,8	26,09	1	3	0,37	0	15	29,3	0	1	4	1	1	1	0	6	18,63	11,73	1,08	24,44	25,38
75	52941,18	29,92	41,24	1	1	34	0	0	14,5	0	1	4	3	4	12	0	6	7,62	0,01	13,87	12,06	19,14
76	17777,78	50,27	69,23	1	2	9	9	0	42,82	3	1	3	2	4	8	0	6	38,16	1,18	4,28	34,98	40,64
77	12857,14	57,5	43,87	1	1	3,5	0	0	68	1	1	4	2	4	8	0	6	34,99	11,71	30,64	36,67	44,41
78	20833,33	43,5	72,36	1	3	4,8	0	28	36	0	1	4	2	4	8	0	6	36,05	7,30	1,38	31,77	36,39
79	31578,95	36,93	44,84	1	2	19	4	0	19	0	1	3	3	4	12	49,67	6	14,52	6,65	18,57	16,37	24,12
80	8382,23	42,84	71,81	1	3	23,86	0	0	40,5	0	1	3	2	1	2	0	6	35,22	7,15	1,85	30,92	35,53
81	9677,42	41,34	71,26	1	3	31	0	90	39	0	1	3	1	1	1	0	6	33,92	7,43	1,86	29,48	33,98
82	92105,26	24,9	62,54	1	3	3,8	0	0	16	0	1	4	3	4	12	0	6	19,64	8,32	2,50	13,39	15,75
83	50000	23,46	65,25	1	3	19	6,5	100	19	0	1	3	3	4	12	60,46	6	22,23	11,39	1,92	15,75	16,89
84	2205,99	56,63	73,32	0	2	160,92	113	0	48	1,5	1	2	1	4	4	0	3	45,61	1,51	2,94	42,54	48,15

2-Amostra de verificação composta por 20 observações.

N	VTI	DN	DE	NE	SIT	AT	AI	LCR	DS	DSP	DH	AP	QB	EB	IB	IPV	ANO	DAP	DR	DRSF	DC	DMP
1	1163,07	68,79	24,55	0	1	417	0	0	52	0	1	1	3	3	9	144,9	3	49,83	0,32	52,44	53,73	61,27
2	435,9	79,22	21,96	0	1	19,5	0	0	69,8	2,8	0	1	1	1	1	0	3	60,49	1,97	58,76	64,22	71,82
3	6217,62	51,66	50,52	0	2	19,3	6,8	0	46	2	1	3	3	4	12	107,69	5	30,08	15,68	22,89	30,39	37,93
4	7500	54,04	70,41	0	2	12	10	0	45	3	1	3	3	3	9	0	4	41,75	0,31	4,29	38,78	44,55
5	16000	13,21	41,58	0	3	16	0	120	12	0	1	1	2	2	4	0	1	9,43	9,08	1,79	11,42	9,87
6	591,84	37,66	42,65	0	2	19,6	0	0	23,6	0	1	2	1	1	1	0	4	15,15	5,13	20,59	17,81	25,49
7	11111,11	35,63	58,89	0	2	22,5	5,5	0	18	2	1	3	3	3	9	6,46	5	20,57	0,23	11,37	17,06	23,13
8	7591,09	44,25	68,71	0	2	9,88	6	0	34	4	1	3	1	1	1	0	5	33,63	3,77	4,56	29,84	35,07
9	31666,67	31,3	53,58	0	2	6	6	0	18,5	1	1	3	3	4	12	8,28	4	13,72	2,56	8,71	10,76	17,70
10	9075,91	36,55	51,08	0	2	30,3	8	0	25,97	0	1	3	3	4	12	0	4	16,17	7,16	14,52	15,34	22,82
11	10263,43	36,4	51,95	0	2	29,23	0	0	25,28	0	1	3	3	4	12	0	4	16,49	6,31	13,99	15,29	22,67
12	2824,63	68,21	47,04	0	1	36,33	0	0	66	6	0	1	2	3	6	0	2	45,86	19,44	31,41	47,00	54,65
13	164,54	75,2	47,84	0	1	62,76	0	0	70	10	0	1	1	1	1	0	2	52,90	22,92	34,43	53,95	61,57
14	164	72,59	44,81	0	1	197,9	0	0	68,5	8,5	0	1	1	1	1	0	2	50,10	19,07	35,51	51,53	59,23
15	230,95	36,26	58,12	0	2	25,1	0	0	21,3	0,7	0	1	1	1	1	0	4	20,40	0,74	12,05	17,20	23,49
16	242,71	27,67	49,77	0	2	12,26	0	0	10,3	0,5	0	1	1	1	1	0	4	8,47	4,32	8,84	6,38	14,03
17	14838,02	13,39	43,37	0	3	121,31	9,72	75	10,7	0	1	3	3	4	12	371,81	3	9,14	7,56	1,26	10,09	8,08
18	15638,96	21,42	35,96	1	2	22,38	5	0	16,7	1,6	1	3	3	4	12	142,88	4	7,17	2,59	10,75	13,70	17,29
19	100000	27,24	40,91	1	2	10	7,5	0	15,7	1,3	1	3	3	4	12	217,04	6	5,18	1,65	11,67	10,58	17,13
20	64285,71	11,11	36,87	1	3	14	0	55	18	0	1	4	3	4	12	0	6	12,97	12,27	1,39	16,34	14,81

ANEXO 01

Obtenção do Índice de Produção Vegetal

A variável Índice de Produção Vegetal foi obtida a partir da metodologia desenvolvida por SILVA, NALI e MAROTE (2009). No referido estudo foi utilizado o cálculo do valor econômico das culturas para obter o rendimento médio das principais culturas agrícolas praticadas no Vale do São Francisco.

A avaliação de culturas é pautada na NBR 14653-3 (ABNT, 2004), a referida norma estabelece que:

10.3 Produções vegetais

Na avaliação em separado das produções vegetais deve-se observar o descrito em 10.3.1 a 10.3.3.

10.3.1 Deve ser empregado o método da capitalização da renda para a identificação do valor econômico.

A definição do referido método na mesma NBR é:

8.2.2 No caso de avaliação de produções vegetais, devem ser observados os procedimentos de 8.2.2.1 a 8.2.2.3.

8.2.2.1 Os rendimentos líquidos esperados devem ser considerados a partir da data de referência da avaliação até o final da vida útil da produção vegetal.

8.2.2.2 Na determinação da renda líquida, deve-se considerar a receita bruta, deduzidos os custos diretos e indiretos, inclusive o custo da terra nua, os impostos e o custo de erradicação, se houver.

8.2.2.3 No cálculo do custo da terra nua, pode-se utilizar o custo de oportunidade sobre o capital que ela representa ou o valor de seu arrendamento.

Voltando a produções vegetais:

10.3.2 No caso de culturas de ciclo longo no primeiro ano de implantação, recomenda-se utilizar, alternativamente ao método da capitalização da renda, o custo de implantação, incluídos os custos diretos e indiretos.

Para obtenção dos rendimento líquidos esperados foram levantados na região os custos de produção, a vida útil, os rendimentos médios, bem como os preços médios pagos ao produtor para as principais culturas agrícolas cultivadas no vale do São Francisco. No custo de produção optou-se por não utilizar o custo da terra nua.

De posse destas informações o valor econômico das produções vegetais foi calculado através do método da capitalização da renda para cada ano de vida útil da cultura, conforme expressão:

$$VE = RL \times Fa (1 - r)$$

$$Fa = \frac{(1+i)^n - 1}{(1+i)^n \times i}$$

onde,

Ve é o valor econômico;

RL é a renda líquida obtida para uma safra média, renda bruta deduzida dos custos diretos e indiretos;

r é a taxa de risco do empreendimento;

Fa é o fator de valor presente para pagamento único para uma série uniforme de rendimentos futuros para **n** períodos de capitalização; e,

i é a taxa de capitalização anual.

De acordo com SILVA, NALI e MAROTE (2009) as variáveis envolvidas na estimativa do valor econômico podem sofrer variação em função das tendências de mercado, como aumentos dos preços de insumos e serviços, variação de preços de venda da produção, quedas na produção, alterações de câmbio, além de turbulências como crises econômicas internacionais. Para minimizar os efeitos acima, ao menor valor econômico foi atribuído a unidade e de forma proporcional aos demais valores relativos às culturas e ao tempo de vida útil (índice de valoração da produção vegetal) (Tabela 56).

Tabela 56 - Índice de valoração da produção vegetal em função do tempo de vida útil das culturas perenes comumente cultivadas na microrregião de Petrolina.

	Vida útil (anos)								
	17	16	15	14	13	12	11	10	9
Banana									
Acerola					16,038	15,465	14,825	14,107	13,303
Coco	8,541	8,366	8,170	7,951	7,706	7,431	7,123	6,778	6,392
Goiaba					28,221	27,214	26,086	24,823	23,409
Mamão									
Manga	7,974	7,811	7,628	7,424	7,194	6,938	6,650	6,328	5,968
Maracujá									
Uva de mesa					68,163	65,731	63,007	59,957	56,540
Uva s/ semente					38,694	37,314	35,768	34,036	32,096

Fonte: SILVA, NALI e MAROTE (2009)

	Vida útil (anos)							
	8	7	6	5	4	3	2	1
Banana					5,236	4,140	2,913	1,539
Acerola	12,403	11,394	10,265	9,000	7,583	5,997	4,220	2,229
Coco	5,959	5,475	4,932	4,324	3,644	2,881	2,027	1,071
Goiaba	21,824	20,050	18,063	15,837	13,344	10,552	7,425	3,923
Mamão						2,789	1,962	1,037
Manga	5,564	5,111	4,605	4,037	3,402	2,690	1,893	1,000
Maracujá						3,688	2,595	1,371
Uva de mesa	52,714	48,428	43,628	38,252	32,231	25,487	17,934	9,474
Uva s/ semente	29,924	27,491	24,766	21,715	18,296	14,468	10,181	5,378

Fonte: SILVA, NALI e MAROTE (2009)