

LIDIANE MARIA FERRAZ ROSA

**ESTUDOS SOBRE A INFLUÊNCIA DE AFIRMAÇÕES
POPULARES NA GEOESTATÍSTICA CLÁSSICA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.

VIÇOSA
MINAS GERAIS – BRASIL
2017

**Ficha catalográfica preparada pela Biblioteca Central da Universidade
Federal de Viçosa - Câmpus Viçosa**

T

R788e
2017
Rosa, Lidiane Maria Ferraz, 1978-
Estudos sobre a influência de afirmações populares na
Geoestatística clássica / Lidiane Maria Ferraz Rosa. – Viçosa,
MG, 2017.
xiii, 88f. : il. (algumas color.) ; 29 cm.

Orientador: Géron Rodrigues dos Santos.
Tese (doutorado) - Universidade Federal de Viçosa.
Inclui bibliografia.

1. Geoestatística. I. Universidade Federal de Viçosa.
Departamento de Estatística. Programa de Pós-graduação em
Estatística Aplicada e Biometria. II. Título.

CDD 22. ed. 519.5


LIDIANE MARIA FERRAZ ROSA

**ESTUDOS SOBRE A INFLUÊNCIA DE AFIRMAÇÕES POPULARES NA
GEOESTATÍSTICA CLÁSSICA**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Estatística Aplicada e Biometria, para obtenção do título de *Doctor Scientiae*.


APROVADA: 15 de fevereiro de 2017.


João Marcos Louzada


Nilcilene das Graças Medeiros


Leandro Roberto de Macedo


Elpídio Inácio Fernandes Filho


Gérson Rodrigues dos Santos
(Orientador)

*À Deus,
sem Ele nada seria possível.
Dedico*

*Aos meus pais Francisco (in memoriam) e Fátima,
ao meu esposo Carlos,
aos meus irmãos Luthiane e Leiliane
e ao meu sobrinho Caio*
OFEREÇO

AGRADECIMENTOS

À Deus por me fazer sentir sua presença em todos os momentos de minha vida.

Aos meus pais, por terem me mostrado a importância do estudo em minha vida e por me ensinarem valores como ética, honestidade, humildade, perseverança, solidariedade.

Ao meu esposo Carlos pelo amor, paciência, companheirismo, incentivo, compreensão pelas horas em que eu não pude estar presente. Muito obrigada!!!

Aos meus irmãos Leiliane e Luthiane, pela amizade e carinho.

Ao meu sobrinho Caio por me fazer tão feliz.

A Universidade Federal de Viçosa por tornar meu sonho realidade. Por toda infraestrutura disponibilizada e todo conhecimento adquirido.

Ao meu orientador, prof. Gérson, pela paciência, dedicação, confiança, atenção, conselhos e ensinamentos na área de Estatística, principalmente Geoestatística.

Ao professor João Marcos Louzada pelos ensinamentos, conselhos, amizade, paciência e dedicação.

À minha grande amiga e funcionária do departamento de Estatística Anita, pelas risadas e incentivo.

A todos os professores do Departamento de Estatística pelos ensinamentos.

Aos coautores professor Paulo César Emiliano, professor João Marcos Louzada professor Afonso Santos, pelas grandes contribuições neste trabalho.

Ao professor Afonso Santos por me socorrer nas horas de desespero com o ArcGis. Pela educação, humildade, conhecimento e disposição em me atender em sua sala.

Aos meus colegas de doutorado professora Camila Azevedo e professor Vinícius Santos pelos estudos em grupo e por estarem sempre prontos a me ajudar.

À Adriana Santos por me ajudar sempre que precisei.

Aos amigos Rosane, Danilo, André e Matheus pelo incentivo e força durante toda a caminhada.

À Capes pelo investimento.

A todos que, de alguma forma, contribuíram para realização deste trabalho.

SUMÁRIO

LISTA DE FIGURAS	vii
LISTA DE TABELAS	x
RESUMO	xii
ABSTRACT	xiii
INTRODUÇÃO GERAL	1
REFERÊNCIAS BIBLIOGRÁFICAS	6
CAPÍTULO 1 - QUAL A MELHOR CONFIGURAÇÃO PARA A AMOSTRAGEM GEOESTATÍSTICA?	8
RESUMO	8
ABSTRACT	10
1. INTRODUÇÃO	11
2. MATERIAL E MÉTODOS	13
2.1 Eficiência entre malhas amostrais	13
2.2 Experimento Prático	19
2.2.1 Área de Estudo	19
2.2.2 Métodos	20
3. RESULTADOS E DISCUSSÃO	24
4. CONCLUSÃO	31
REFERÊNCIAS BIBLIOGRÁFICAS	32
CAPÍTULO 2 - TESTE DE NORMALIDADE PARA GEOESTATÍSTICA: FAZER OU NÃO FAZER? EIS A QUESTÃO.....	34
RESUMO	34
ABSTRACT	35
1. INTRODUÇÃO	36
2. MATERIAL E MÉTODOS	38
2.1 Abordagem Teórica	38
2.2 Abordagem Prática	40
3. RESULTADOS E DISCUSSÃO	43
4. CONCLUSÃO	48

REFERÊNCIAS BIBLIOGRÁFICAS	49
CAPÍTULO 3 - AVALIAÇÃO DA INFLUÊNCIA DA SIMETRIA DOS DADOS EM UMA ANÁLISE GEOESTATÍSTICA	51
RESUMO	51
ABSTRACT	52
1. INTRODUÇÃO	53
2. MATERIAL E MÉTODOS	54
3. RESULTADOS E DISCUSSÃO	59
4. CONCLUSÃO	65
REFERÊNCIAS BIBLIOGRÁFICAS	66
CAPÍTULO 4 - MODELAGEM DIGITAL DE PROFUNDIDADE ATRAVÉS DE INTERPOLADORES GEOESTATÍSTICOS	68
RESUMO	68
ABSTRACT	69
1. INTRODUÇÃO	70
2. MATERIAL E MÉTODOS	72
2.1 Área de Estudo.....	72
3. RESULTADOS E DISCUSSÃO	78
4. CONCLUSÃO	83
REFERÊNCIAS BIBLIOGRÁFICAS	84
CONCLUSÕES GERAIS	88

LISTA DE FIGURAS

INTRODUÇÃO

Figura 1: Representação de um semivariograma com os principais parâmetros que são estimados em uma análise geoestatística.....	03
--	----

CAPÍTULO 1

Figura 1: Representação da malha retangular e os quatro pontos de estimação.....	15
--	----

Figura 2: Representação da malha triangular e os três pontos de estimação.....	17
--	----

Figura 3: Representação da malha hexagonal e os respectivos pontos de estimação.....	18
--	----

Figura 4: Representação da área de estudo compreendendo uma área de 5,7 km ² do município de Viçosa, Minas Gerais – Brasil.....	19
--	----

Figura 5: Representação da altimetria da área de estudo, próximo à cidade de Viçosa – MG, Brasil, com aproximadamente 230 mil pontos amostrais.	20
--	----

Figura 6: Representação por quartis das malhas amostrais hexagonal, retangular e triangular da subamostra 1 (primeira linha), subamostra 5 (segunda linha) e subamostra 10 (terceira linha)	24
---	----

Figura 7: Representação dos quartis das malhas amostrais hexagonal, retangular triangular, das subamostras , T1 (primeira linha); T4 (segunda linha) e T8 (terceira linha).....	25
---	----

CAPÍTULO 2

Figura 1: Representação da área (3ha) de estudo da variável altimetria no Campus da Universidade Federal de Viçosa – MG.....	41
--	----

Figura 2: Distribuição de frequências da variável altimetria da área (3ha) de estudo no campus da Universidade Federal de Viçosa – MG.....	44
Figura 3: Representação por quartis da variável altimetria da área (3ha) de estudo no campus da Universidade Federal de Viçosa – MG.....	45
Figura 4: Gráfico mostrando a tendência de segunda ordem presente nos dados da variável altimetria.....	45
Figura 5: Semivariograma experimental direcional e modelo ajustado das direções: N-S(0°), E-W (90°), SW-NE (45°) e NW-SE (135°).....	46
Figura 6: Mapa de distribuição espacial da variável altimetria gerado pela krigagem simples.....	47

CAPÍTULO 3

Figura 1: Representação por quartis da malha amostral do levantamento batimétrico de um represamento do Rio São Bartolomeu.....	54
Figura 2: Penetrômetro de cone de solo de múltiplas sondas (MPSCP) com cinco sondas e sistema GPS para medição de índices de cone em solo de argila arenosa Pacolet em Auburn, AL.....	55
Figura 3: Elevação do campo de amostragem do solo argiloso de Pacolet. As marcas indicam os pontos de amostragem para a medição do índice de cone.....	56
Figura 4: Histogramas dos dados originais (a) e das transformações cúbicas (b), Box-Cox ($\lambda = 1,6$) (c) e Box-Cox ($\lambda = 2,5$) (d) respectivamente	59
Figura 5: Variogramas dos dados originais (a), da transformação cúbica (b), transformação Box-Cox ($\lambda = 1,6$) (c) e da transformação Box-Cox ($\lambda = 2,5$) (d) respectivamente.....	62

Figura 6: Distribuição de frequência dos dados originais (a) e da transformação logarítmica (b) respectivamente.....	63
--	----

CAPÍTULO 4

Figura 1: Distribuição de frequências da variável batimetria do represamento do Rio São Bartolomeu.....	78
---	----

Figura 2: Semivariogramas ajustados e teóricos da krigagem simples, krigagem universal e krigagem regressão respectivamente do levantamento batimétrico do represamento do Rio São Bartolomeu.....	79
--	----

Figura 3: Semivariogramas da krigagem simples (a), krigagem universal (b) e krigagem regressão (c) respectivamente.....	82
---	----

LISTA DE TABELAS

CAPÍTULO 1

Tabela 1: Dados referentes à redução do tamanho amostral e elevação do RMS da malha hexagonal	26
Tabela 2: Dados referentes à redução do tamanho amostral e elevação do RMS da malha retangular	27
Tabela 3: Dados referentes à redução do tamanho amostral e elevação do RMS da malha triangular	28
Tabela 4: Apresentação dos dados obtidos após a realização da análise geoestatística para o conjunto de dados em que as áreas são diferentes para as malhas amostrais adotadas	30

CAPÍTULO 2

Tabela 1: Estatística descritiva para a variável altimetria da área de estudos no Campus da Universidade Federal de Viçosa – MG.....	43
Tabela 2: Apresentação das principais estatísticas da validação cruzada.....	47

CAPÍTULO 3

Tabela1: Valores dos coeficientes de assimetria dos dados originais e transformados do levantamento batimétrico de um represamento do Ribeirão São Bartolomeu – Viçosa – MG.....	60
Tabela 2: Valores da média e dos desvios padrão dos resíduos obtidos pela autovalidação.....	60
Tabela 3: Valores dos coeficientes da regressão e coeficiente de determinação obtidos pela autovalidação.....	61

Tabela 4: Valores da variância dos erros obtidos pela autovalidação.....	61
Tabela 5: Valores dos coeficientes de assimetria dos dados originais e transformados da variável índice de cone de pico.....	63
Tabela 6: Intervalo de confiança dos dados originais e transformados	63
Tabela 7: Variância dos dados originais e transformados	64

CAPÍTULO 4

Tabela 1: Estatística descritiva da batimetria do represamento do Rio São Bartolomeu.....	78
Tabela 2: Comparação entre os métodos de krigagem pela média e variância dos resíduos padronizados obtidos pela autovalidação.....	80
Tabela 3: Valores obtidos do MSE para os três métodos de krigagem.....	80

RESUMO

ROSA, Lidiane Maria Ferraz Rosa, D.Sc., Universidade Federal de Viçosa, fevereiro de 2017. **Estudos sobre a influência de afirmações populares na Geoestatística Clássica.** Orientador: Gérson Rodrigues dos Santos. Coorientadores: Carlos Henrique Osório Silva e Paulo César Emiliano.

O presente trabalho teve por objetivo estudar assuntos abordados na Geoestatística de modo a preencher lacunas metodológicas que surgem naturalmente em áreas recentes da ciência e/ou que surgem de afirmações populares. Para isso, foi necessário realizar estudos teórico-práticos afim de se obter mais evidências e comprovações. Foram utilizados em todos os estudos conjuntos de dados reais. No primeiro capítulo foi feito um estudo quanto à precisão entre malhas amostrais regulares (malha hexagonal, malha quadrangular e malha triangular). No segundo capítulo foi realizado um estudo sobre a questão do uso do teste de Normalidade em uma análise geoestatística. Nesse estudo foi mostrada toda fundamentação teórica do principal teste de normalidade, teste de *Shapiro-Wilk*, e estudos práticos para essa comprovação. No terceiro capítulo realizou-se um estudo sobre a assimetria e o comprometimento de uma análise geoestatística para dados com essa característica. Finalmente, no quarto capítulo foram comparados três interpoladores geoestatísticos (krigagem simples, krigagem universal e krigagem regressão) quanto à eficácia na representação de modelos digitais de profundidade (MDP).

ABSTRACT

ROSA, Lidiane Maria Ferraz Rosa, D.Sc., Universidade Federal de Viçosa, February, 2017. **Studies on the influence of popular affirmations on Classical Geostatistics.** Advisor: Gérson Rodrigues dos Santos. Co-advisors: Carlos Henrique Osório Silva and Paulo César Emiliano.

The present work had the objective of studying subjects covered in Geostatistics in order to fill methodological gaps that arise naturally in recent areas of science and / or that emerge from popular affirmations. For this, it was necessary to carry out theoretical-practical studies in order to obtain more evidences and proofs. Actual data sets were used in all studies. In the first chapter a study was made on the precision between regular sample meshes, among them, hexagonal mesh, quadrangular mesh and triangular mesh. In the second chapter a study was carried out on the question of using the Normality test in a geostatistical analysis. In this study, the theoretical basis of the main normality test, Shapiro-Wilk test, and practical studies for this verification were shown. Already in the third chapter was carried out a study the asymmetry and the commitment of a geostatistical analysis for data with this characteristic. Finally, in the fourth chapter three geostatistical interpolators (simple kriging, universal kriging and regression kriging) were compared for the effectiveness of the representation of digital depth models (MDP).

INTRODUÇÃO GERAL

A Estatística Clássica considera alguns pressupostos como a aleatoriedade das amostras, o que de certa forma deveria garantir a independência estatística entre as mesmas. Assim, pode-se apenas adotar a média amostral, entre outras medidas, como um resumo descritivo dos dados.

Contudo, mesmo adotando os princípios que norteiam a aleatoriedade estatística, a independência pode não ser atingida. Vieira (2000) cita vários trabalhos que encontraram este resultado, pois não consideravam a dependência espacial das amostras. São nestes casos que a Geoestatística tem suas principais aplicações. Segundo Cressie (1993) a Geoestatística leva em consideração que a posição dos dados amostrais tem influência no comportamento da variância da variável de interesse. Portanto, a posição geográfica desses dados é utilizada pela Geoestatística para caracterizar a variável em estudo.

A Geoestatística foi desenvolvida através de estudos do engenheiro de minas sul-africano Daniel Krige na década de 1950. Krige conseguiu constatar que havia sentido nas variâncias estimadas, ao levar em consideração a distância e a direção da amostragem (SANTOS et al., 2011). Georges Matheron, um engenheiro francês foi o primeiro a utilizar a Geoestatística e a noção de variograma na análise estatística dos dados de Krige. Segundo Matheron (1963) “a Geoestatística é a aplicação do formalismo das funções aleatórias ao reconhecimento e à estimação dos fenômenos naturais.”

O termo Geoestatística tem uma abrangência muito maior do que o proposto por Matheron (1971) e pode ser definido como uma subárea da estatística espacial que estuda as variáveis regionalizadas (YAMAMOTO e LANDIM, 2013).

Variável regionalizada é definida, segundo Matheron (1963), como uma função espacial numérica que varia de um local para outro tendo uma continuidade aparente, sendo que esta variabilidade não pode ser representada por simples funções matemáticas.

Segundo Vieira (2000) até os anos de 1960, as análises dos dados eram baseadas na hipótese de independência estatística, ou seja, distribuição espacial aleatória. Já para Guedes (2013) a hipótese de aleatoriedade dos dados não pode ser assumida antes que se prove a inexistência de correlação espacial entre os dados. Portanto, o conhecimento da variabilidade espacial permite a descrição da correlação espacial das variáveis.

Além de identificar a existência ou não da dependência espacial entre as observações, a metodologia geoestatística, pode ser utilizada para mapeamentos, orientações de futuras amostragens e modelagens (GOMES et al., 2007).

De acordo com Ferreira et al. (2013) uma das etapas mais importantes de uma análise geoestatística é o estudo do semivariograma, pois auxilia no estudo do comportamento das variáveis regionalizadas. De acordo com Santos et al. (2011) o semivariograma é uma importante ferramenta pois descreve quantitativamente a variação espacial, ou em outras palavras, mede o grau de semelhança entre amostras vizinhas. O esperado é que amostras próximas sejam mais semelhantes entre elas do que amostras mais afastadas.

Segundo Andriotti (2013) o semivariograma é, em termos matemáticos, a média do quadrado das diferenças entre todos os pares de pontos em uma área estudada, separados por um vetor h . A equação do semivariograma é dada na equação (1),

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

em que $z(x_i)$ é o valor observado da variável no ponto (x_i) e $z(x_i + h)$ é o valor observado no ponto $(x_i + h)$, $\hat{\gamma}(h)$ é o valor estimado da semivariância na distância h e $N(h)$ é o número de pares de pontos separados entre si por uma distância h . De acordo com Vieira (2000) o estimador de semivariância apresentado na equação (1) é baseado no método de momentos e foi proposto por Matheron em 1963.

Ao se calcular as semivariâncias estimadas para cada valor de h observa-se, em relação à distância h , o comportamento espacial da variável em estudo (CARMO et al., 2015). O gráfico das semivariâncias pode ser visualizado na Figura 1.

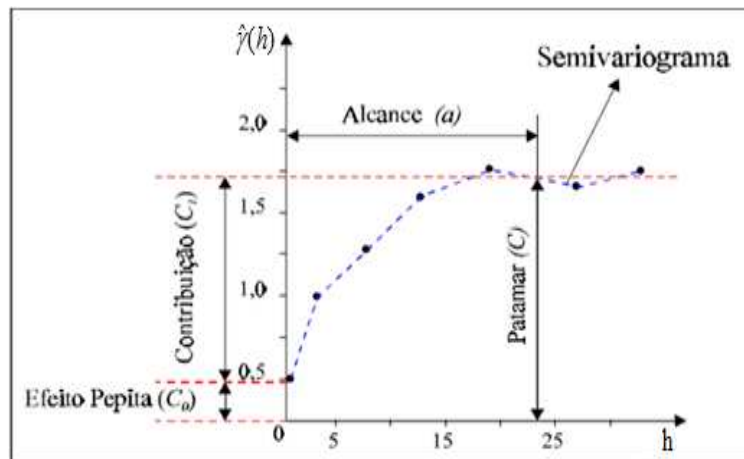


Figura 1 – Representação de um semivariograma com os principais parâmetros que são estimados em uma análise geoestatística
 Fonte: Câmara e Medeiros (1998)

Na Figura 1 são identificados os seguintes parâmetros para ajuste do semivariograma:

- Alcance: é a distância dentro da qual as amostras apresentam-se correlacionadas espacialmente.
- Patamar: é o valor da variância em que o semivariograma se estabiliza.
- Efeito pepita (C_0): é o valor da semivariância para a distância zero.
- Contribuição: é a diferença entre os valores do patamar (C) e do efeito pepita (C_0) e é uma forma de quantificar a variabilidade dos dados.

Segundo Yamamoto e Landim (2013) quando o comportamento do semivariograma não se alterar, conforme a direção, este é dito ser isotrópico, caso contrário, é dito ser anisotrópico. Quando detectada a presença de anisotropia, esta deve ser modelada através de um modelo direcional de semivariograma.

De acordo com Santos et al. (2011) após a obtenção do semivariograma experimental pode-se fazer o ajuste do mesmo através dos modelos teóricos. Os modelos teóricos mais utilizados são os modelos esférico, exponencial e gaussiano. Esse ajuste é uma etapa importante, logo não deve ser realizado automaticamente, pois a etapa de krigagem depende do modelo de semivariograma ajustado. Assim, se o modelo for ajustado incorretamente, a krigagem não apresentará a precisão esperada (VILELA, 2004).

Para avaliar a qualidade do ajuste do modelo teórico escolhido é utilizada a validação cruzada. De acordo com Isaaks e Srivastava (1989) escolhe-se o modelo teórico que apresentar os valores mais satisfatórios para as estatísticas dos resíduos.

Após a escolha do modelo é realizada a etapa da krigagem. A krigagem é o método de interpolação geoestatístico que pode ser definida como um processo de predição de valores de variáveis distribuídas no espaço a partir de valores adjacentes, enquanto considerados como interdependentes pelo variograma (LANDIM, 1998).

Segundo Vieira (2000) a krigagem fornece valores sem tendência e com variância mínima, o que lhe concedeu o status de melhor preditor linear não viesado ou simplesmente denominado de BLUP (Best Linear Unbiased Predictor). Através da krigagem é possível conhecer a variância de krigagem, isto é, a precisão associada a cada predição (CAMARGO, 1998).

Encontra-se na literatura diversos tipos de krigagem: krigagem simples, krigagem ordinária, krigagem universal, krigagem fatorial, entre outras. Dentre todas, a mais utilizada é a krigagem ordinária, pois, para esta não é necessário que se conheça a média estacionária. Porém, segundo Santos et al. (2011) a krigagem simples é a mais precisa dentre os preditores lineares.

Objetiva-se de uma forma geral com este trabalho elucidar algumas inadequações teóricas na análise geoestatística. Para isso, foram estudados alguns temas da Geoestatística que necessitavam de um aprofundamento. Estes estudos foram realizados de forma teórica e/ou prática.

Especificamente, pretende-se atingir esse objetivo geral através dos específicos:

- a) Avaliar três tipos de configurações de malhas regulares amostrais (triangular, retangular e hexagonal) e verificar qual a melhor configuração para a amostragem geoestatística.
- b) Mostrar através de estudos teórico e prático a inviabilidade da exigência de normalidade para uma análise geoestatística.
- c) Avaliar o quanto uma análise geoestatística fica comprometida ao se trabalhar com dados simétricos e dados assimétricos, partindo de uma assimetria natural dos dados e chegando numa simetria aceitável estatisticamente.
- d) Comparar três interpoladores geoestatísticos mais comumente utilizados (krigagem simples, krigagem universal e regressão krigagem) e verificar qual deles é o mais preciso.

Este trabalho está estruturado em seis seções, com os seguintes títulos e objetivos:

- Introdução Geral – apresenta uma introdução geral sobre a Geoestatística, os objetivos gerais e específicos.
- Capítulo 1 – Qual a melhor configuração para a amostragem geoestatística? Neste capítulo será estudado a amostragem Geoestatística com objetivo de detalhar essa amostragem. Esse estudo será em relação a 2 aspectos: primeiro, quando as malhas amostrais apresentam mesma áreas e diferentes densidades e; segundo, quando as malhas apresentam mesma densidade e diferentes áreas.
- Capítulo 2 – Teste de normalidade para Geoestatística: fazer ou não fazer? Eis a questão. Mostrar através de uma fundamentação teórica e prática sobre a a realização ou não do teste de normalidade para dados que apresentem dependência espacial. Para atingir tal objetivo será analisado um conjunto que não apresente normalidade e verificar se as pressuposições da Geoestatística são satisfeitas.
- Capítulo 3 – Avaliação da influência da simetria dos dados em uma análise geoestatística. Mostrar a fundamentação teórica envolvida na transformação dos dados e avaliar o comprometimento de uma análise geoestatística para dados assimétricos e dados simétricos. Para isso serão feitas diferentes transformações em um mesmo conjunto de dados.
- Capítulo 4 – Modelagem digital de profundidade através de interpoladores geoestatísticos. Estudar a diferença entre três interpoladores geoestatísticos. Para isso, serão estudados conjunto de dados com tendência e a comparação será feita utilizando indicadores comparativos de qualidade da Geoestatística.
- Conclusões Gerais – expõem as considerações finais e recomendações para trabalhos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

ANDRIOTTI, J. L. S. **Fundamentos de Estatística e Geoestatística**. São Leopoldo: Editora Unisinos, 2013.

CÂMARA, G.; MEDEIROS, J. S. **Geoprocessamento para projetos ambientais**. V1, Online Book, São José dos Campos, INPE, 1998.

CAMARGO, E. C. G. **Geoestatística: Fundamentos e Aplicações**. In: CAMARA, G.; MEDEIROS, J.S. **Geoprocessamento em Projetos Ambientais**. 2ª ed. São José dos Campos: INPE, 1998.

CARMO, E. J.; RODRIGUES, D. D.; SANTOS, G. R. Avaliação dos interpoladores krigagem e topo to raster para geração de modelos digitais de elevação a partir de um “as built”. **Boletim de Ciências Geodésicas**, v. 21, p.674-690, 2015.

CRESSIE, A. G. **Statistics for spatial data**. New York: J. Wiley, 1993.

FERREIRA, Í. O.; SANTOS, G. R.; RODRIGUES, D. D. Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. **Revista Brasileira de Cartografia** (Online), v. 65, n°5, p. 831/03-842, 2013.

GOMES, N.M.; SILVA, A.M.; MELLO, C.R.; FARIA, M.A.; OLIVEIRA, P.M. Métodos de ajuste e modelos de semivariograma aplicados ao estudo da variabilidade espacial de atributos físico-hídricos do solo. **Revista Brasileira de Ciência do Solo**, v.31, p.435-443, 2007.

GUEDES, I.C.L. Uso de métodos da estatística espacial em biometria e inventário florestal. [Tese], UFLA, 2013.

ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York, Oxford University Press, 1989.

LANDIM, P. M. B. **Análise estatística de dados geológicos**. São Paulo: Fundação Editora da Unesp, 1998.

MATHERON, G. Principles of geostatistics. **Economic Geology**, v. 58, p. 1246–1266. 1963.

MATHERON, G. (1971). The Theory of Regionalized Variables and Its Applications. **Les Cahiers du Centre de Morphologie Mathématique de Fontainebleau**, n° 5. Fontainebleau: Ecole Nationale Supérieure des Mines de Paris.

SANTOS, G. R.; [OLIVEIRA, M. S.](#); LOUZADA, J. M.; [SANTOS, A. M. R. T.](#) Krigagem Simples versus Krigagem universal: qual o preditor mais preciso? **Energia na Agricultura**, v.26, n°2, p.49-55, 2011.

VIEIRA, S.R. **Geoestatística em estudos de variabilidade espacial do solo**. Tópicos em Ciências do Solo 1(2000): 1-54.

VILELA, L. C. **Comparação entre as redes neurais artificiais e o método de interpolação krigagem aplicados à pesquisa agrônômica**. [Tese], UNESP. Faculdade de Ciências Agrônômicas, 2004.

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística – conceitos e aplicações**. São Paulo: Oficina de textos, 2013.

CAPÍTULO 1

QUAL A MELHOR CONFIGURAÇÃO PARA A AMOSTRAGEM GEOESTATÍSTICA?

RESUMO

A técnica para obtenção de uma amostra, ou seja, parte selecionada de uma população é definida como amostragem. Ao se realizar uma amostragem georreferenciada deve-se primeiramente fazer um planejamento e definir a forma como as amostras serão coletadas. A Geoestatística tem sido utilizada em diversos estudos amostrais e, portanto a configuração espacial deve ser considerada. O objetivo deste trabalho foi avaliar a diferença entre três tipos de configurações de malhas regulares amostrais (triangular, retangular e hexagonal) quanto a precisão preditiva, com o intuito de alcançar a melhor configuração para a amostragem de um campo aleatório de espaço de índice espacialmente contínuo, isto é, que atenda a metodologia Geoestatística. Para isso foi utilizada uma medida da variância de krigagem obtida na validação cruzada das interpolações visando à avaliação da precisão preditiva. Neste estudo foram utilizados dados de altimetria do município de Viçosa, considerando duas etapas: na primeira, os três polígonos apresentavam tamanhos amostrais diferentes e na segunda foi mantida o mesmo número de pontos para os polígonos. Para as configurações em que os polígonos apresentavam número de pontos diferentes, a hierarquia obtida foi que a malha hexagonal apresentou maior precisão, seguida da malha retangular e por fim a malha triangular. Porém, não foi possível detectar diferenças significativas, estatisticamente, quanto à precisão preditiva nas configurações em que os polígonos apresentavam mesmo número de pontos. Assim, os resultados indicam que foi possível a obtenção de uma hierarquia para as malhas amostrais, indicando qual a melhor configuração para a amostragem geoestatística. Porém, ficou claro que essa hierarquização depende de alguns fatores descritos no trabalho.

Palavras-chave: amostragem geoestatística, malhas regulares, malhas hexagonais, malhas retangulares, malhas triangulares.

ABSTRACT

The technique to obtain a sample of the part of a population is defined as sampling. When performing a georeferenced sampling it must first be made a planning and define the way that the samples will be collected. Geostatistics has been utilized in diverse sampling studies and because of that the spatial configuration must be considered. The aim of this paper was evaluate the difference between three types of regular sampling grids configurations (triangular, rectangular and hexagonal) according to the predictive accuracy, with the aim of reaching the better configuration for the geostatistics sampling. For that matter, it was utilized a measure of kriging variance obtained in the cross-validation of the interpolations aiming the evaluation of the predictive precision. In this study there was utilized altimetry dataset of the county of Viçosa, meaning that, it was performed in two stages in the study: in the first stage, the three polygons presented different sampling sizes and in the second stage it was maintained the same number of points for the polygons. For the configurations of which the polygons presented different number of points, the hierarchy obtained was that the hexagonal grid presented greater precision, followed by the rectangular grid and finally the triangular grid. However, it was not possible to detect significant differences, statistically, for accuracy predicted in the configurations in which the polygons presented the same number of points. Thus, the results indicate that it was possible the obtainance of a hierarchy for the sampling grids, indicating the best configuration for a geostatistics sampling. However, it was clear that this ranking depends on some factors described in the paper.

Keywords: geostatistics sampling, regular grids, hexagonal grids, rectangular grids, triangular grids.

1.INTRODUÇÃO

Ao se realizar um estudo, nem sempre é possível a observação de todos os elementos de uma população, sendo assim, é necessário que se faça uso de amostras. É importante que essas amostras sejam de qualidade, ou seja, que apresentem características bem próximas da população, o que comumente é chamado de representatividade (OLIVEIRA et al., 2014).

A amostragem é definida segundo Oliveira et al. (2014) como sendo o processo de coleta de parte de uma população. Alguns autores, como Bunchaft e Kellner (2000), Bolfarine e Bussab (2005) e Ferreira (2009) estudaram acerca da importância da amostragem e concluíram que a amostra deve descrever a população de origem.

Para o processo de coleta de dados georreferenciados deve-se buscar a mesma qualidade. Toda amostragem georreferenciada deve ser realizada segundo um planejamento e este definirá como serão coletadas as unidades amostrais. Essa amostragem poderá ser de forma aleatória ou sistemática (YAMAMOTO e LANDIM, 2013).

Ainda segundo Yamamoto e Landim (2013), a amostragem sistemática é aquela que apresenta os pontos regularmente espaçados formando uma malha ou *grid*. O tipo de malha regular mais utilizada é a retangular, devido a praticidade de planejamento e coleta (ROEL e TERRA, 2006; WEBSTER e LARK, 2012).

Flatman e Yfantis (1984) recomendam que, em estudos onde apresenta variabilidade espacial, a Geoestatística deve ser utilizada por apresentar maior eficiência analítica, frente à Estatística Clássica, na análise de dados.

Diversos autores como Diniz (2007), Cherubin et al. (2014) e Sousa et al. (2014) têm utilizado as técnicas geoestatísticas em seus estudos acerca de amostragem.

O objetivo deste trabalho foi avaliar três tipos de configurações de malhas regulares amostrais, a saber: triangular, quadrangular e hexagonal; com a finalidade de obter a melhor configuração para a amostragem geoestatística, utilizando como medida de precisão preditiva um tipo de variância de krigagem.

Especificamente, apresenta-se a demonstração do teorema apresentado por Yfantis et al. (1987), de uma forma mais completa, acerca de tais configurações, em que a relação entre os três polígonos se dá através da medida dos lados. Ademais, avalia-se, através de um estudo criterioso de um grande conjunto de dados, a diferença entre os três polígonos, quanto à precisão preditiva, considerando áreas iguais (consequentemente, diferentes tamanhos amostrais) e áreas diferentes (logo, com tamanhos amostrais iguais) para uma mesma região de estudo.

Sendo assim, o trabalho está organizado em duas partes: a primeira sendo teórica, em que a demonstração do teorema apresentado por Yfantis et al. (1987) foi realizada detalhadamente e a segunda, a parte prática, foi subdividida em dois estudos.

O primeiro estudo da parte prática consistiu em manter as malhas amostrais com os três polígonos apresentando a mesma área interna (consequentemente com tamanhos amostrais diferentes) e o segundo estudo apresentou áreas internas diferentes (mantendo assim o mesmo número de pontos).

2. MATERIAL E MÉTODOS

2.1. Eficiência entre malhas amostrais

Ao estudar a eficiência entre malhas amostrais, Yfantis et al. (1987) enunciaram um teorema em que mostram que a malha hexagonal possui erro quadrático médio máximo menor que as malhas retangular e triangular, respectivamente. O teorema diz que se um semivariograma é dado da forma:

$$\begin{aligned}\gamma(h) &= C_0 + C_1 f(h) & 0 < h \leq r \\ \gamma(h) &= C_0 + C_1 f(r) = \sigma^2 & h > r\end{aligned}\quad (1)$$

em que C_0 é o efeito pepita e $C_0 \geq 0$, C_1 é a contribuição e $C_1 > 0$ e r é o alcance, o erro quadrático médio máximo para a malha retangular é $\sigma_\varepsilon^2 = \frac{5}{4}[C_0 + C_1 f(r)] = \frac{5}{4}\sigma^2$. Isto ocorre quando o lado a do quadrado ou a distância entre amostras sucessivas é igual a $a = \sqrt{2}r$. Para a malha triangular o erro quadrático médio máximo é dado por $\sigma_\varepsilon^2 = \frac{4}{3}[C_0 + C_1 f(r)] = \frac{4}{3}\sigma^2$ quando o lado a do triângulo equilátero é igual a $a = \sqrt{3}r$. E com relação a malha hexagonal, o erro quadrático médio máximo é igual a $\sigma_\varepsilon^2 = \frac{7}{6}[C_0 + C_1 f(r)] = \frac{7}{6}\sigma^2$ quando o lado a do hexágono regular é igual a r (YFANTIS et al., 1987).

Para provar esse teorema, algumas considerações devem ser feitas.

Assume-se que os pontos amostrais possuem mesmo processo aleatório ($Z(X)$) no campo de interesse. Portanto, suponha que seja necessário estimar $Z(X_0)$. Então:

$$Z^*(X_0) = \sum_{i=1}^n a_i Z(X_i) \quad (2)$$

em que X_1, X_2, \dots, X_n são todos pontos da malha e a distância (a) é menor ou igual ao alcance r (YFANTIS et al., 1987).

Os pesos a_i podem ser obtidos resolvendo o seguinte sistema de equações:

$$A.B = C \quad (3)$$

em que:

$$A = \begin{bmatrix} 0 & \gamma(|X_1 - X_2|) & \gamma(|X_1 - X_3|) & \dots & \gamma(|X_1 - X_n|) & 1 \\ \gamma(|X_1 - X_2|) & 0 & \gamma(|X_2 - X_3|) & \dots & \gamma(|X_2 - X_n|) & 1 \\ \gamma(|X_1 - X_3|) & \gamma(|X_2 - X_3|) & 0 & \dots & \gamma(|X_3 - X_n|) & 1 \\ \dots & \dots & \dots & \dots & \dots & 1 \\ \gamma(|X_1 - X_n|) & \gamma(|X_2 - X_n|) & \gamma(|X_3 - X_n|) & \dots & 0 & 1 \\ 1 & 1 & 1 & \dots & 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ \dots \\ a_n \\ \lambda \end{bmatrix} \quad C = \begin{bmatrix} \gamma(|X_1 - X_0|) \\ \gamma(|X_2 - X_0|) \\ \gamma(|X_3 - X_0|) \\ \dots \\ \gamma(|X_n - X_0|) \\ 1 \end{bmatrix}$$

e λ é o multiplicador de Lagrange (YFANTIS et al.,1987).

A partir do sistema de equações (3) pode-se perceber que a solução dependerá apenas da malha amostral e do semivariograma. Se $Z(X_0)$ é o verdadeiro valor de X_0 , a variância do erro também é o erro quadrático médio, pois $Z^*(X_0)$ é um estimador não viesado de $Z(X_0)$, dado por:

$$\sigma_\varepsilon^2 = \text{var}[Z^*(X_0) - Z(X_0)] = 2 \sum_{i=1}^n a_i \gamma(X_i - X_0) - \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(X_i - X_j) \quad (4)$$

Verifica-se a partir das equações (3) e (4) que o erro quadrático médio está em função somente da malha amostral e do semivariograma.

De posse dessas considerações pode-se então provar o teorema enunciado por Yfantis et al. (1987). Tendo a malha retangular, com o lado do quadrado sendo a , maior que $\sqrt{\frac{2}{5}}r$ e menor ou igual a $\sqrt{2}r$, observa-se através da Figura 1 que o número de pontos envolvidos na estimação de $Z(X_0)$ são 4. São eles, X_1, X_2, X_3 e X_4 .

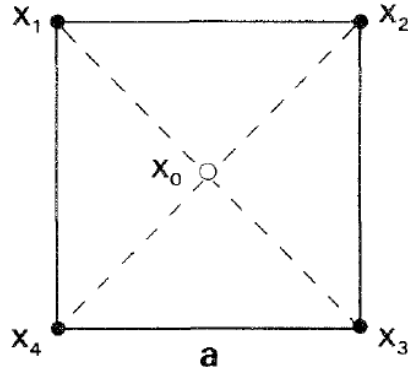


Figura 1 - Representação da malha retangular e os quatro pontos de estimação.
 Fonte: Yfantis et al. (1987).

A equação matricial para a malha retangular é:

$$\begin{bmatrix} 0 & \gamma(|X_1 - X_2|) & \gamma(|X_1 - X_3|) & \gamma(|X_1 - X_4|) & 1 \\ \gamma(|X_1 - X_2|) & 0 & \gamma(|X_2 - X_3|) & \gamma(|X_2 - X_4|) & 1 \\ \gamma(|X_1 - X_3|) & \gamma(|X_2 - X_3|) & 0 & \gamma(|X_3 - X_4|) & 1 \\ \gamma(|X_1 - X_4|) & \gamma(|X_2 - X_4|) & \gamma(|X_3 - X_4|) & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(|X_1 - X_0|) \\ \gamma(|X_2 - X_0|) \\ \gamma(|X_3 - X_0|) \\ \gamma(|X_4 - X_0|) \\ 1 \end{bmatrix}$$

Note que $\gamma(|X_1 - X_2|) = \gamma(|X_1 - X_4|) = \gamma(|X_2 - X_3|) = \gamma(|X_3 - X_4|) = c$ e $\gamma(|X_1 - X_3|) = \gamma(|X_2 - X_4|) = b$ tem-se:

$$\begin{bmatrix} 0 & c & b & c & 1 \\ c & 0 & c & b & 1 \\ b & c & 0 & c & 1 \\ c & b & c & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ a_4 \\ \lambda \end{bmatrix} = \begin{bmatrix} \gamma(|X_1 - X_0|) \\ \gamma(|X_2 - X_0|) \\ \gamma(|X_3 - X_0|) \\ \gamma(|X_4 - X_0|) \\ 1 \end{bmatrix}$$

Ao resolver esse sistema obtém-se $a_1 = a_2 = a_3 = a_4 = \frac{1}{4}$.

Como $\sigma_\varepsilon^2 = 2 \sum_{i=1}^n a_i \gamma(X_i - X_0) - \sum_{i=1}^n \sum_{j=1}^n a_i a_j \gamma(X_i - X_j)$ então

$$\begin{aligned}
\sigma_\varepsilon^2 &= 2\{[a_1\gamma(X_1 - X_0) + a_2\gamma(X_2 - X_0) + a_3\gamma(X_3 - X_0) + a_4\gamma(X_4 - X_0)] - \\
&- [a_1a_1\gamma(X_1 - X_1) + a_1a_2\gamma(X_1 - X_2) + a_1a_3\gamma(X_1 - X_3) + a_1a_4\gamma(X_1 - X_4) + \\
&+ a_2a_1\gamma(X_2 - X_1) + a_2a_2\gamma(X_2 - X_2) + a_2a_3\gamma(X_2 - X_3) + a_2a_4\gamma(X_2 - X_4) + \\
&+ a_3a_1\gamma(X_3 - X_1) + a_3a_3\gamma(X_3 - X_3) + a_3a_2\gamma(X_3 - X_2) + a_3a_4\gamma(X_3 - X_4) + \\
&+ a_4a_1\gamma(X_4 - X_1) + a_4a_2\gamma(X_4 - X_2) + a_4a_3\gamma(X_4 - X_3) + a_4a_4\gamma(X_4 - X_4)] \\
\sigma_\varepsilon^2 &= 2\left\{\left[\frac{1}{4}\gamma(h) + \frac{1}{4}\gamma(h) + \frac{1}{4}\gamma(h) + \frac{1}{4}\gamma(h)\right]\right\} - \left[\frac{1}{4}\cdot\frac{1}{4}c + \frac{1}{4}\cdot\frac{1}{4}b + \frac{1}{4}\cdot\frac{1}{4}c + \frac{1}{4}\cdot\frac{1}{4}c\right] \\
&+ \left[\frac{1}{4}\cdot\frac{1}{4}c + \frac{1}{4}\cdot\frac{1}{4}b + \frac{1}{4}\cdot\frac{1}{4}b + \frac{1}{4}\cdot\frac{1}{4}c + \frac{1}{4}\cdot\frac{1}{4}c + \frac{1}{4}\cdot\frac{1}{4}c + \frac{1}{4}\cdot\frac{1}{4}c + \frac{1}{4}\cdot\frac{1}{4}b + \frac{1}{4}\cdot\frac{1}{4}c\right] \\
\sigma_\varepsilon^2 &= 2\gamma(h) - \left[\frac{1}{16}\gamma(a)\cdot 8 + 4\cdot\frac{1}{16}\gamma(2h)\right] \\
\sigma_\varepsilon^2 &= 2\gamma(h) - \left[\frac{\gamma(a)}{2} + \frac{\gamma(2h)}{4}\right] \tag{5}
\end{aligned}$$

h é a distância entre X_0 e X_1 .

Substituindo a equação (5) na equação (1) obtém-se:

$$\sigma_\varepsilon^2 = 2\sigma^2 - \left[\frac{\sigma^2}{2} + \frac{\sigma^2}{4}\right] \tag{6}$$

Portanto, o erro quadrático médio obtido para a malha retangular é:

$$\sigma_\varepsilon^2 = \frac{5}{4}\sigma^2. \tag{7}$$

Em relação à malha triangular, quando o lado do triângulo equilátero é maior que $\frac{\sqrt{3}}{2}r$ e menor ou igual a $\sqrt{3}r$, os pontos X_1, X_2, X_3 distam de X_0 por um valor menor ou igual a r (Figura 2).

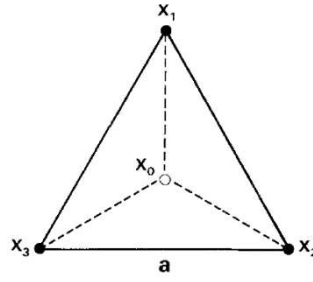


Figura 2 - Representação da malha triangular e os três pontos de estimação.
Fonte: Yfantis et al. (1987).

Ao resolver o sistema de equações dado na equação (4) obtém-se que

$$a_1 = a_2 = a_3 = \frac{1}{3}.$$

$$\begin{aligned} \sigma_\varepsilon^2 = & 2\{[a_1\gamma(X_1 - X_0) + a_2\gamma(X_2 - X_0) + a_3\gamma(X_3 - X_0)] - [a_1a_1\gamma(X_1 - X_1) + a_1a_2\gamma(X_1 - X_2) \\ & + a_1a_3\gamma(X_1 - X_3) + a_2a_1\gamma(X_2 - X_1) + a_2a_2\gamma(X_2 - X_2) + a_2a_3\gamma(X_2 - X_3) + a_3a_1\gamma(X_3 - X_1) \\ & + a_3a_2\gamma(X_3 - X_2) + a_3a_3\gamma(X_3 - X_3)] \} \end{aligned}$$

$$\sigma_\varepsilon^2 = 2 \left\{ \left[\frac{1}{3}\gamma(h) + \frac{1}{3}\gamma(h) + \frac{1}{3}\gamma(h) \right] \right\} - \left[\frac{1}{3} \cdot \frac{1}{3}c + \frac{1}{3} \cdot \frac{1}{3}c + \frac{1}{3} \cdot \frac{1}{3}c + \frac{1}{3} \cdot \frac{1}{3}c + \frac{1}{3} \cdot \frac{1}{3}c + \frac{1}{3} \cdot \frac{1}{3}c \right]$$

$$\sigma_\varepsilon^2 = 2 \frac{3}{3}\gamma(h) - \frac{6}{9}c$$

$$\sigma_\varepsilon^2 = 2 \frac{3}{3}\gamma(h) - \frac{6}{9}\gamma(a)$$

$$\sigma_\varepsilon^2 = 2\sigma^2 - \frac{2}{3}\sigma^2$$

Portanto, o erro quadrático médio máximo da malha triangular é

$$\sigma_\varepsilon^2 = \frac{4}{3}\sigma^2 \quad (8)$$

Para a malha hexagonal de lado a com $\frac{r}{2} \leq a \leq r$, tem-se a representação do hexágono regular com seus respectivos pontos amostrais $X_1, X_2, X_3, X_4, X_5, X_6$ (Figura 3).

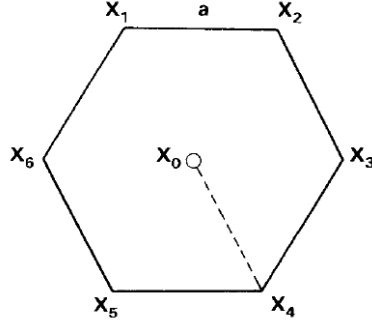


Figura 3 - Representação da malha hexagonal e os respectivos pontos de estimação.
Fonte: Yfantis et al., (1987).

Resolvendo o sistema de equações (3) encontra-se

$$a_1 = a_2 = a_3 = a_4 = a_5 = a_6 = \frac{1}{6}. \text{ Assim,}$$

$$\sigma_\varepsilon^2 = 2[a_1\gamma(X_1 - X_0) + a_2\gamma(X_2 - X_0) + a_3\gamma(X_3 - X_0) + a_4\gamma(X_4 - X_0) + a_5\gamma(X_5 - X_0) + a_6\gamma(X_6 - X_0)] - [a_1a_1\gamma(X_1 - X_1) + \dots + a_6a_6\gamma(X_6 - X_6)]$$

$$\sigma_\varepsilon^2 = 2\left[\frac{1}{6}\gamma(h) + \dots + \frac{1}{6}\gamma(h)\right] - \left[\frac{1}{6}\cdot\frac{1}{6}\gamma(a) + \frac{1}{6}\cdot\frac{1}{6}\gamma(b) + \frac{1}{6}\cdot\frac{1}{6}\gamma(2h) + \frac{1}{6}\cdot\frac{1}{6}\gamma(b) + \frac{1}{6}\cdot\frac{1}{6}\gamma(a)\right] + \left[\frac{1}{6}\cdot\frac{1}{6}\gamma(a) + \frac{1}{6}\cdot\frac{1}{6}\gamma(a) + \frac{1}{6}\cdot\frac{1}{6}\gamma(b) + \frac{1}{6}\cdot\frac{1}{6}\gamma(2h) + \frac{1}{6}\cdot\frac{1}{6}\gamma(b) + \frac{1}{6}\cdot\frac{1}{6}\gamma(b) + \dots + \frac{1}{6}\cdot\frac{1}{6}\gamma(a)\right]$$

$$\sigma_\varepsilon^2 = 2\frac{6}{6}\gamma(h) - \left[\frac{1}{36}\cdot 12\gamma(a) + \frac{1}{36}\cdot 12\gamma(b) + \frac{1}{36}\cdot 6\gamma(2h)\right]$$

$$\sigma_\varepsilon^2 = 2\sigma^2 - \frac{1}{3}\sigma^2 - \frac{1}{3}\sigma^2 - \frac{1}{6}\sigma^2$$

Portanto, o erro quadrático médio máximo da malha hexagonal é dado por

$$\sigma_\varepsilon^2 = \frac{7}{6}\sigma^2 \quad (9)$$

2.2. Experimento Prático

De posse do teorema apresentado foram feitos dois estudos com dados de altimetria para efeito de comparação em relação à precisão: o primeiro para malhas amostrais com mesma área para os polígonos (consequentemente, diferentes tamanhos amostrais) e o segundo para malhas amostrais com aproximadamente tamanhos amostrais iguais (consequentemente, áreas diferentes).

2.2.1. Área de Estudo

Visando avaliar outros dois aspectos não considerados por Yfantis et al. (1987), um grande conjunto de dados foi utilizado, a saber, um conjunto de dados, referente à variável altimetria de parte da Zona da Mata Mineira, no município de Viçosa – MG, Brasil, com 229414 pontos amostrais, em que o relevo é acidentado. A região de estudo apresenta altitude média de, aproximadamente, 721 metros e variância média de 1.473 metros quadrados.

Os dados utilizados neste trabalho são provenientes de uma área que compreende uma porção de 5,7 km² do município de Viçosa, situado na Zona da Mata Mineira no estado de Minas Gerais, Brasil. Esta região é delimitada pelas latitudes 20°45'53''S e 20°46'53''S e longitude 42°52'24''W e 42°53'49''W, de acordo com a Figura 4.

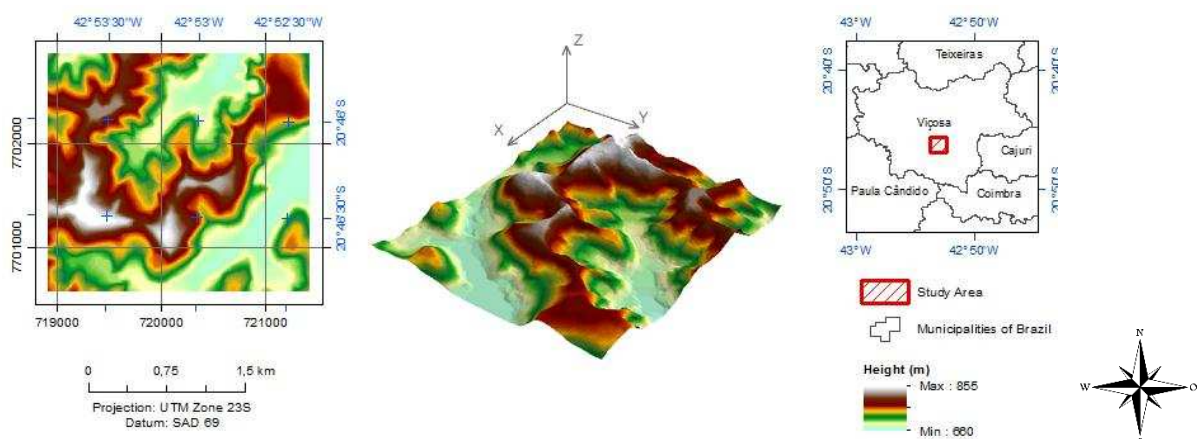


Figura 4 - Representação da área de estudo compreendendo uma área de 5,7 km² do município de Viçosa, Minas Gerais - Brasil

A base de dados utilizada compreende cerca de 230 mil pontos contendo informações altimétricas acerca da área de estudo espaçados em 5 metros, em ambas as direções X e Y. Os valores de altitudes tem amplitude de 660 a 855 metros (Figura 5), sendo estas informações provenientes de um processo restituição fotogramétrico a partir de um modelo estereoscópico obtido de imagens do satélite Ikonos II. A base de dados está representada na projeção Universal Transversa de Mercator (UTM) fuso 23 sul, sendo referenciado ao sistema geodésico SAD-69 (South American Datum of 1969).

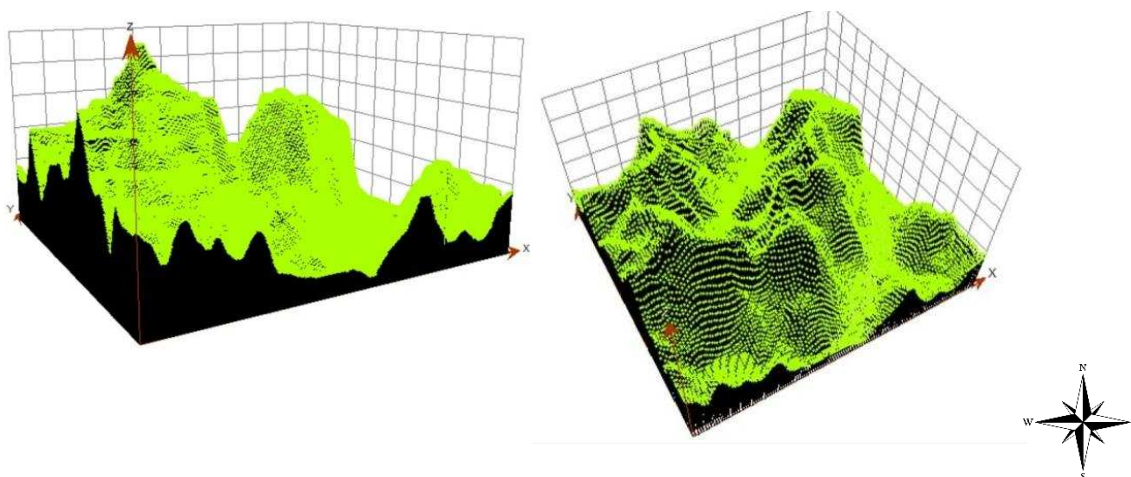


Figura 5 – Representação da altimetria da área de estudo, próximo à cidade de Viçosa – MG – Brasil, com aproximadamente 230 mil pontos amostrais.

2.2.2. Métodos

Inicialmente o estudo foi realizado com aproximadamente 230 mil pontos, dos quais foram selecionadas subamostras menores conforme especificado na coluna 3 da Tabela 1, para o primeiro estudo. Essas subamostras são representadas respectivamente por A_1, A_2, \dots, A_{10} com áreas iguais em m^2 a 1600, 1800, 2000, 2200, 2800, 3300, 4600, 6800, 13800 e 27600, respectivamente. Já no segundo estudo, foram selecionados 7380, 3654, 1836, 931, 460, 224, 110 e 56 pontos amostrais que serão representadas neste trabalho por subamostras T_1, \dots, T_8 sucessivamente.

A técnica geoestatística foi utilizada para identificar e caracterizar a dependência espacial dos dados altimétricos, a partir do ajustamento de funções teóricas aos dados dos semivariogramas experimentais, baseada na pressuposição de estacionariedade da hipótese intrínseca de acordo com a equação (10):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (10)$$

em que: $N(h)$ é o número de pares experimentais de observações $(Z(x_i), Z(x_i + h))$ separados por um vetor h .

Para o ajuste dos modelos teóricos aos semivariogramas experimentais foram estimados o efeito pepita (C_0), o patamar ($C_0 + C_1$) e o alcance (a). Estimou-se também a dependência espacial, dado pelo índice de dependência espacial (IDE), ou seja, $IDE = \frac{C_1}{C_0 + C_1}$ proposto por Zimback (2001) que considera dependência espacial fraca se $IDE < 25\%$, moderada se $25\% \leq IDE < 75\%$ e forte se $IDE \geq 75\%$. Para a estimação dos valores dos dados de altimetria em locais não amostrados e confecção dos mapas de krigagem, utilizou-se a krigagem simples, segundo Santos et al. (2011). O procedimento da análise geoestatística utilizado neste trabalho foi sugerido por Ferreira et al. (2013).

Para comparar a precisão entre as malhas foi considerada a raiz quadrada do erro médio (RMSE) obtido pela autovalidação dado pela equação (11), conforme sugere Vieira (2000).

$$RMSE = \sqrt{\frac{\sum_{i=1}^n [\hat{Y}(x_i) - Y(x_i)]^2}{n}} \quad (11)$$

em que, $\hat{Y}(x_i)$ é o valor estimado, $Y(x_i)$ é o valor observado e n o tamanho amostral.

Visando a comparação entre malhas amostrais com o mesmo número aproximado de pontos, foram realizados testes t de *Student* para duas médias, conforme recomenda Ferreira (2009). Além disso, foram testadas a normalidade e a homogeneidade de variâncias.

Segundo Ferreira (2009) após a realização dos testes supracitados, para a realização do teste t de *Student* utiliza-se a estatística do teste dada pela equação (12),

$$T = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{\sqrt{S_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}, \quad (12)$$

e o estimador da variância combinada, devido à homocedasticidade, dada pela equação (13).

$$S_c^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \quad (13)$$

Em que, \bar{X}_1 e S_1^2 são os estimadores das médias e da variância da população 1 e \bar{X}_2 e S_2^2 são os estimadores das médias e da variância da população 2, n_1 e n_2 são os tamanhos amostrais das populações 1 e 2 respectivamente, μ_1 e μ_2 são as médias das populações 1 e 2.

A estatística da equação (12) segue uma distribuição t de *Student* com $v = n_1 + n_2 - 2$ graus de liberdade.

Para verificar a independência foi utilizado o semivariograma, conforme recomendam Ferreira, et al. (2013). Quanto à normalidade, utilizou-se do teste qqplot, conforme recomenda Ferreira (2009).

O teste de homocedasticidade adotado foi o de Bartlett. Riboldi et al. (2014) citam que existem vários testes de igualdade de variâncias, mas para Conover et al. (1981) o teste que tem se tornado modelo é o teste de Bartlett (1937) e ele deve ser utilizado quando a distribuição dos dados é a normal. A estatística do teste é dada por:

$$B_o = \frac{q}{c} \quad (14)$$

em que

$$q = (N - k) * \ln s_p^2 - \sum_{i=1}^k [(n_i - 1) * \ln s_i^2]$$

$$c = 1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{N-k} \right),$$

$$s_p^2 = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) s_i^2 \quad \text{e} \quad s_i^2 = \sum_{j=1}^{n_i} \frac{(y_{ij} - \bar{y}_{i.})^2}{n_i - 1}.$$

Toda a parte de geoprocessamento do trabalho foi realizado no software ArcGis (ESRI, 2014) e a parte estatística do estudo (análise geoestatística e testes) foi realizada no software R (R Development Core Team, 2015), principalmente no pacote geoR (RIBEIRO JUNIOR E DIGGLE, 2001).

3. RESULTADOS E DISCUSSÃO

Serão apresentados e discutidos primeiramente os resultados obtidos para o estudo das malhas amostrais regulares considerando diferentes números de pontos.

Na Figura 6 tem-se a representação por quartis das amostras. Os pontos em azul indicam os valores com menor altitude, seguidos pelos pontos verdes, pontos amarelos e por último indicando maior altitude os pontos em vermelho. Nesta figura estão representadas as três malhas amostrais estudadas com mesma área interna dos polígonos. As malhas estão representadas respectivamente pelos polígonos hexágonos, quadrados e triângulos.

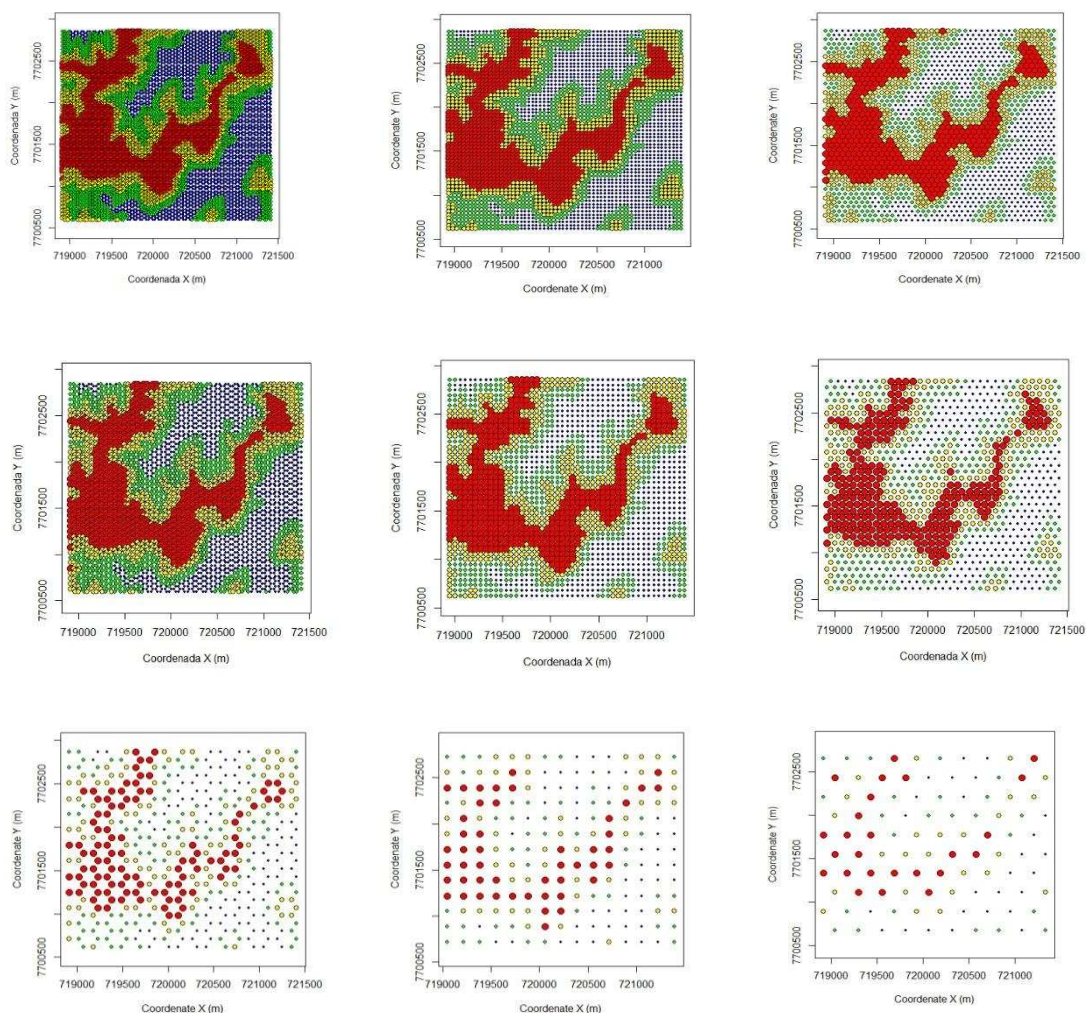


Figura 6 – Representação por quartis das malhas amostrais hexagonal, retangular e triangular de área 160 m^2 (primeira linha), área de 3300 m^2 (segunda linha) e área de 28000 m^2 (terceira linha).

Apesar da aparente conservação do agrupamento dos valores, o indicador de qualidade adotado (RMSE) neste trabalho mostra resultados bem diferentes, conforme resultados das análises geoestatísticas apresentados no gráfico abaixo.

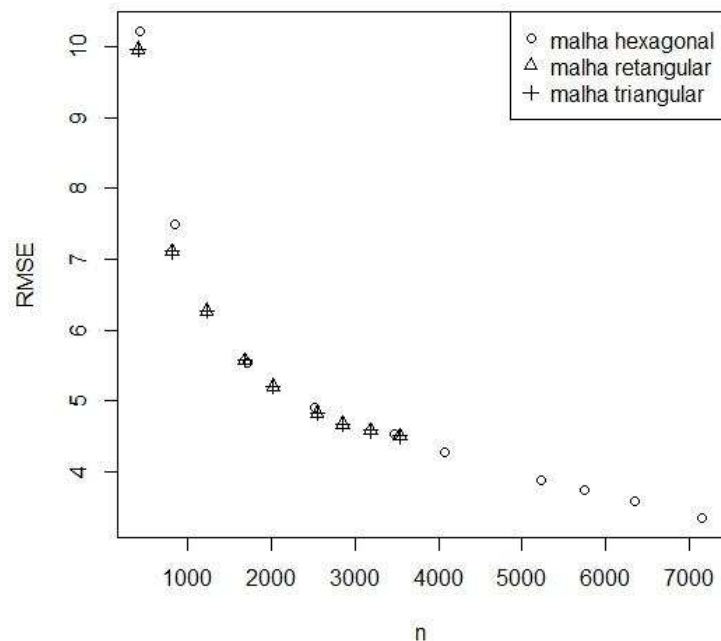


Figura 7 – Gráfico do tamanho amostral das malhas hexagonal, retangular e triangular versus o RMSE.

Através da Figura 7 é possível notar que quando as amostras possuem um número em torno de 1000 a 3000 pontos amostrais, o comportamento é bem parecido. O RMSE é bem próximo para as três malhas amostrais.

Oliveira (1991), ao estudar a hierarquia entre as malhas amostrais encontrou o mesmo resultado que Yfantis et al. (1987), em que a variância máxima de krigagem da malha triangular é superior a quadrangular. Oliveira (1991) não considerou a malha hexagonal em seus estudos pois a julgou inviável na amostragem espacial, pois não apresentava vantagens em termos de custo em comparação a malha quadrangular e triangular.

Um comportamento de grande importância para estudos dessa natureza é a perda de precisão que acontece na redução do tamanho amostral. Assim, apresentam-se na Tabela 1 resultados (em %) do que aconteceu com a redução do tamanho amostral relacionados com a redução da precisão do estudo para a malha hexagonal.

Tabela 1 - Dados referentes à redução do tamanho amostral e elevação do RMSE da malha hexagonal.

n	Redução (%) do tamanho amostral	RMSE	Acréscimo (%) do RMSE
7155	---	3,35	---
6350	11,25	3,58	6,87
5747	9,50	3,73	4,19
5233	8,94	3,88	4,02
4080	22,03	4,27	10,05
3478	14,75	4,53	6,09
2520	27,54	4,9	8,17
1716	31,90	5,53	12,86
846	50,70	7,5	35,62
429	49,29	10,22	36,27
Média	25,10		13,79

Destaca-se que nas Tabelas 1, 2 e 3, as reduções apresentadas foram calculadas sempre em relação ao valor anterior. Sendo assim, primeiramente pela Tabela 1, percebe-se que em média a redução do tamanho amostral foi de 25% e a elevação do RMSE, para a malha hexagonal foi em média de aproximadamente 14%.

De igual maneira para a malha regular retangular, conforme apresentado na Tabela 2, pode-se notar que, em média, a redução do tamanho amostral foi de aproximadamente 30%, enquanto a elevação do RMSE foi em média 22%.

Tabela 2 – Dados referentes a redução do tamanho amostral e elevação do RMS da malha retangular.

n	Redução (%) do tamanho amostral	RMSE	Acréscimo (%) do RMSE
3534	---	4,5	---
3186	9,85	4,58	1,78
2856	10,36	4,67	1,97
2544	10,92	4,83	3,43
2021	20,56	5,2	7,66
1677	17,02	5,57	7,12
1221	27,19	6,27	12,57
810	33,66	7,11	13,40
399	50,74	9,97	40,23
195	51,13	14,08	41,22
Média	30,49		21,50

Os resultados provenientes da malha regular triangular, conforme apresentado na Tabela 3, mostram que, em média, a redução do tamanho amostral foi de aproximadamente 31%, enquanto a elevação do RMSE foi de aproximadamente, 28%.

Comparando os resultados apresentados nas Tabelas 1, 2 e 3, nota-se que a redução média entre as três malhas foram, respectiva e aproximadamente, 25%, 30% e 31%. Enquanto que a elevação do RMSE foi de 14%, 22% e 28%. Assim, mesmo diante de percentuais quase similares de redução do tamanho amostral, a perda de precisão (percebida pela elevação do RMSE) se mostra em diferenças mais elevadas.

Tabela 3 – Dados referentes a redução do tamanho amostral e elevação do RMS da malha triangular.

n	Redução (%) do tamanho amostral	RMS	Acréscimo (%) do RMS
1826	---	5,36	---
1599	12,43	5,72	6,72
1462	8,57	5,93	3,67
1313	10,19	6,08	2,53
1039	20,87	6,89	13,32
870	16,27	7,38	7,12
637	26,78	8,56	16,00
430	32,50	10,26	19,86
217	49,53	14,1	37,43
110	49,31	21,11	49,72
Média	30,87		28,22

Pode-se perceber ainda com os resultados das Tabelas 1, 2 e 3 que mediante o RMSE, a malha hexagonal apresentou sempre o menor valor, seguida pela malha retangular e, finalmente, triangular. Esse resultado está em conformidade com o teorema de Yfantis et al. (1987) em que a malha amostral hexagonal apresentou teoricamente menor erro quadrático médio máximo em relação às outras duas malhas.

Portanto, fica evidente que, fixando as áreas dos três polígonos mais comuns de configuração amostral para malhas regulares, acontece uma hierarquização, quanto à precisão, de uma configuração em relação às demais, a saber, malha hexagonal, malha retangular e malha triangular, respectivamente. Pode-se perceber ainda, que tal hierarquização pode estar ocorrendo devido à diferença existente no tamanho amostral.

Em contrapartida, pode-se considerar uma densidade fixa de pontos amostrais para as malhas, conforme estudou teoricamente Oliveira (1991) e estabelecer outra hierarquia para as configurações das malhas. Este autor concluiu que a variância máxima de krigagem da malha triangular é superior à retangular.

Na Figura 8 as malhas amostrais estão divididas pelos tamanhos amostrais T_1 (7380), T_2 (3654), T_3 (1836), T_4 (931), T_5 (460), T_6 (224), T_7 (110) e T_8 (56) pontos amostrais e estão representadas pelos quartis. Os pontos que simbolizam as menores altitudes estão representados em azul, seguidos, respectivamente, pelos pontos verdes,

pontos amarelos e, finalmente, pelos pontos em vermelho, que simbolizam as maiores altitudes da região estudada.

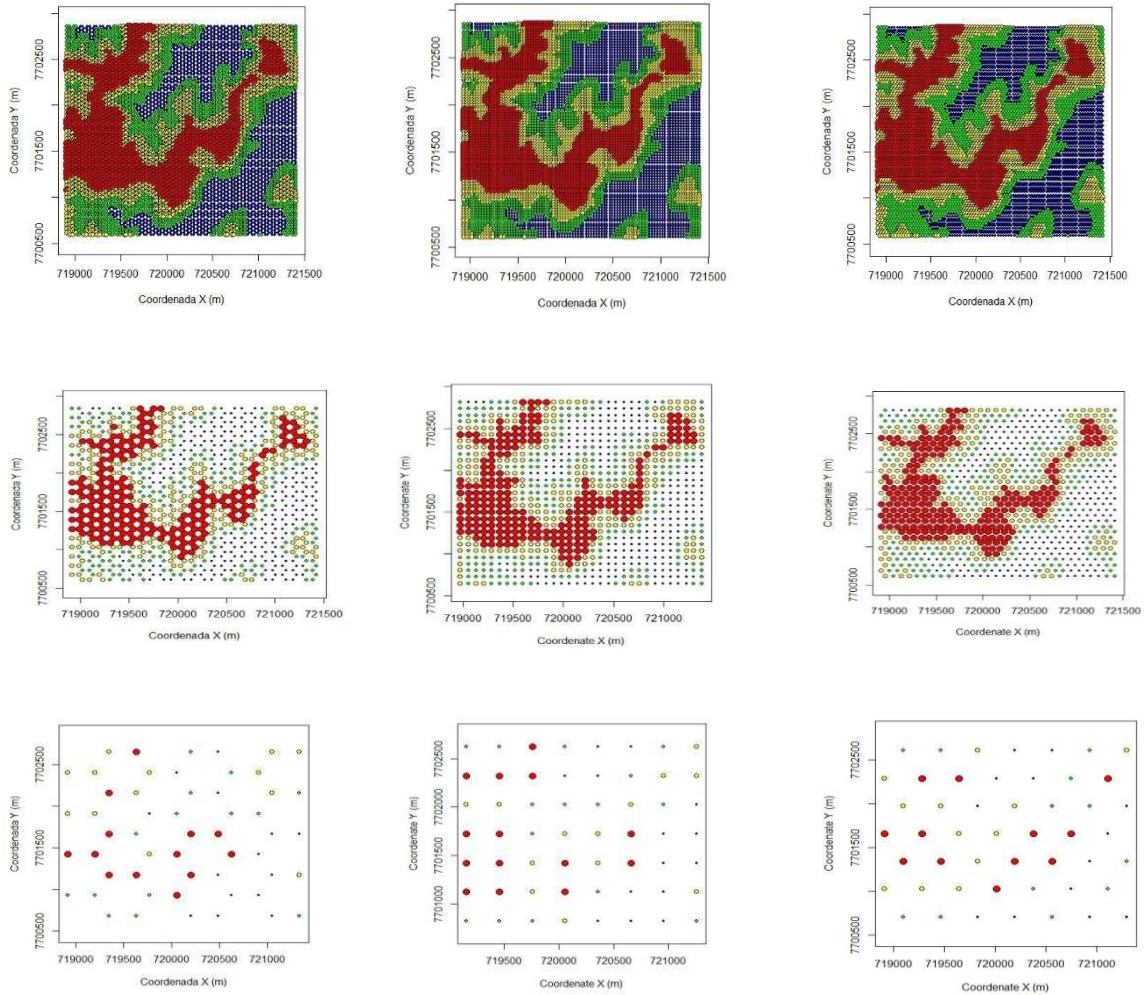


Figura 8 – Representação dos quartis das malhas amostrais hexagonal, retangular e triangulas, das subamostras T_1 com 7380 pontos amostrais (primeira linha), T_4 com 931 pontos amostrais (segunda linha) e T_8 com 56 pontos amostrais (terceira linha).

Assim, a Figura 8 é a representação gráfica das três configurações amostrais mais comuns quanto aos polígonos, cujas malhas amostrais regulares apresentam áreas internas diferentes e tamanhos amostrais iguais ou aproximadamente iguais.

Na Tabela 4 estão apresentados os dados relativos a análise geoestatística das malhas amostrais com diferentes áreas e números de pontos aproximados. Para este conjunto de dados foi realizado o teste t para duas médias para verificar se havia diferença ente os valores dos RMSE's das malhas amostrais. Essas malhas foram

comparadas duas a duas. Portanto, pela Tabela 4 pode-se verificar que estes valores não diferenciam pelo teste t.

Tabela 4 – Apresentação dos dados obtidos após a realização da análise geoestatística para o conjunto de dados em que as áreas são diferentes para as malhas amostrais adotadas.

		RMSE	IDE	Resultado			RMSE	IDE	Resultado
T ₁	H vs Q	3,28 vs 3,24	Forte	NS	T ₅	H vs Q	9,99 vs 9,53	Forte	NS
7380	H vs T	3,28 vs 3,32	Forte	NS	460	H vs T	9,99 vs 10,36	Forte	NS
pts	Q vs T	3,24 vs 3,32	Forte	NS	pts	Q vs T	9,53 vs 10,36	Forte	NS
T ₂	H vs Q	4,29 vs 4,39	Forte	NS	T ₆	H vs Q	13,33 vs 13,92	Forte	NS
3654	H vs T	4,29 vs 4,44	Forte	NS	224	H vs T	13,33 vs 14,2	Forte	NS
pts	Q vs T	4,39 vs 4,44	Forte	NS	pts	Q vs T	13,92 vs 14,2	Forte	NS
T ₃	H vs Q	5,47 vs 5,45	Forte	NS	T ₇	H vs Q	18,32 vs 17,45	Forte	NS
1836	H vs T	5,47 vs 5,51	Forte	NS	110	H vs T	18,32 vs 19,23	Forte	NS
pts	Q vs T	5,45 vs 5,47	Forte	NS	pts	Q vs T	17,45 vs 19,23	Forte	NS
T ₄	H vs Q	6,78 vs 6,87	Forte	NS	T ₈	H vs Q	24,47 vs 24,45	Forte	NS
931	H vs T	6,78 vs 7,22	Forte	NS	56	H vs T	24,47 vs 27,84	Forte	NS
pts	Q vs T	6,87 vs 7,22	Forte	NS	pts	Q vs T	24,45 vs 27,84	Forte	NS

Legenda: NS – não significativo; H- malha hexagonal; Q- malha retangular; T- malha triangular; T₁ a T₈ subamostras. IDE – índice de dependência espacial.

Com base nos testes realizados e apresentados na Tabela 4 pode-se afirmar que não há diferença estatística entre os RMSE das diferentes configurações amostrais quando as amostragens são feitas de tal forma que os três polígonos mais comuns (hexágonos, quadrados e triângulos) apresentam diferentes áreas e mesmo número de pontos amostrais para a área estudada.

Pode-se constatar ainda, conforme Oliveira et al. (2014) afirma, que vale a regra áurea dos levantamentos amostrais que adotam as metodologias estatísticas, ou seja, quanto maior o tamanho da amostra, mais se aproxima do tamanho populacional, e menor é o erro amostral (apresentado neste trabalho como RMSE). Tal constatação vem dos valores reduzidos de RMSE para tamanhos amostrais maiores e consequentes elevações deste indicador ao se reduzir os tamanhos das amostras.

4. CONCLUSÃO

Diante dos resultados obtidos, pode-se afirmar que é possível obter uma hierarquia, ou seja, pode-se apontar a melhor configuração para a amostragem geoestatística. Contudo, ficou evidente neste trabalho que tal hierarquização depende de alguns aspectos como área interna dos polígonos e número de pontos das malhas amostrais.

De modo geral, encontrou-se teoricamente que, conservando a mesma medida lateral para os três polígonos, a malha hexagonal apresenta maior precisão, seguida da malha retangular e, finalmente, a malha triangular.

Na prática, pode-se optar por polígonos de mesma área (tamanhos amostrais diferentes) e polígonos de áreas diferentes (tamanhos amostrais iguais ou aproximadamente iguais).

Adotando as configurações em que os polígonos apresentavam a mesma área, a maior precisão é dada pela malha hexagonal, seguida pela retangular e, posteriormente, a malha triangular.

Contudo, ao adotar-se as configurações em que os polígonos apresentavam mesmo número de pontos, não foi possível detectar diferenças significativas, estatisticamente, quanto à precisão preditiva.

REFERÊNCIAS BIBLIOGRÁFICAS

ArcGIS. Version 10.2.1. Redlands, CA: Esri. 2014.

BARTLETT, M.S. Properties of sufficiency and statistical tests. **Proceeding of the Royal Society of London**, Series A, v.160, n° 901, p. 268–282. 1937.

BOLFARINE, H; BUSSAB, W.O. **Elementos de amostragem**. Porto Alegre: Bookman. 2005.

BUNCHAFT, G.; KELLNER, S.R.O. **Estatística sem mistérios**. v.4, Petrópolis: Vozes. 2000.

CHERUBIN, M.R; SANTI, A.L.; EITELWEIN, M.T.; MENEGOL, D.R.; ROS, C.O.; PIAS, O.H.C.; BERGHETTI, J. Eficiência de malhas amostrais utilizadas na caracterização da variabilidade espacial de fósforo e potássio. **Ciência Rural**, v.44, n° 3, p.425-432. 2014.

CONOVER, W.J.; JOHNSON, M.E.; JOHNSON, M.M.A. Comparative study of tests for homogeneity of variances, with application to the outer continental shelf bidding data. **Technometrics**, v.23, n° 4, p.351-361. 1981.

DINIZ, F.D. **Métodos de amostragem e geoestatística aplicados ao inventário florestal**. [dissertação]. Lavras: Universidade Federal de Lavras. 2007.

FLATMAN, G.T.; YFANTIS, A.A. Geostatistical strategy for soil sampling: the survey and the census. **Environmental Monitoring and Assessment**, v.4, n°4, p.335-349. 1984.

FERREIRA, I.O.; SANTOS, G.R.; RODRIGUES, D.D. Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. **Revista Brasileira de Cartografia** (Online), v.65, n° 5, p.831-842. 2013.

FERREIRA, D.F. **Estatística básica**. 2ª ed. Lavras: UFLA. 2009.

OLIVEIRA, M.S. **Planos amostrais para variáveis espaciais utilizando Geoestatística**. [dissertação]. São Paulo: UNICAMP. 1991.

OLIVEIRA, M.S; BEARZOTI, E.; BOAS, F.L.V.; NOGUEIRA, D.A.; NICOLAU, LA; OLIVEIRA, H.S.S. **Introdução à estatística**. 2ª ed.Lavras: UFLA. 2014.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing. 2016.

Disponível em: <<http://www.r-project.org>>. Acesso em: 30 Apr. 2015.

RIBEIRO JUNIOR, P.J.; DIGGLE, P.J. **geoR: a package for geostatistical analysis**. R-NEWS, London, v. 1, n. 2, p. 15-18. 2001.

RIBOLDI J, BARBIAN, M.H.; KOLOWSKI, A.B.S, SELAU, L.P.R.; TORMAN, V.B.L. Precisão e poder de testes de homocedasticidade paramétricos e não-paramétricos avaliados por simulação. **Revista Brasileira de Biometria**, v.32, n°3, p.334-344. 2014.

ROEL, A.; TERRA, J. Muestreo de suelos y factores limitantes del rendimiento. Agricultura de precisión: Integrando conocimientos para una agricultura moderna y sustentable. PROCISUR /IICA. p. 65-80. 2006.

SANTOS, G.R.; OLIVEIRA, M.S.; LOUZADA, J.M.; SANTOS, A.M.R.T. Krigagem Simple versus Krigagem universal: qual o preditor mais preciso? **Energia na Agricultura**, v.26, n°2, p.49-55. 2011.

SOUSA, Z.M.; SOUZA, G.S.; JOSÉ, M.J.; PEREIRA, G.T. Número de amostras na análise geoestatística e na krigagem de mapas de atributos do solo. **Ciência Rural**. v.44, n°2, p.261-268. 2014.

VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R.F. et al. (Ed.). Tópicos em ciência do solo. **Sociedade Brasileira de Ciência do Solo**. Viçosa, v.1, p.1-53. 2000.

WEBSTER, R.; LARK, M. **Field Sampling for Environmental Science and Management**. 1st ed. London: Routledge. 2012.

YAMAMOTO, J.K.; LANDIM, P.M.B. **Geoestatística – conceitos e aplicações**. São Paulo: Oficina de textos, 2013.

YFANTIS, E.A.; FLATMAN, G.T.; BEHAR, J.V. Efficiency of kriging estimation for square, triangular and hexagonal grids. **Mathematical Geology**. v.19, n°3, p.183-205. 1987.

ZIMBACK, C.R.L. **Análise Espacial de atributos químicos de solos para fins de mapeamento da fertilidade**. [tese]. Faculdade de Ciências Agronômicas, Botucatu: FCA/UNESP; 2001.

CAPÍTULO 2

TESTE DE NORMALIDADE PARA GEOESTATÍSTICA: FAZER OU NÃO FAZER? EIS A QUESTÃO

RESUMO

A base da inferência estatística é identificar a distribuição de probabilidade da variável aleatória em estudo. Provavelmente a distribuição de probabilidade mais conhecida e utilizada é a de Gauss-Laplace, ou simplesmente, distribuição normal. Assim sendo, a verificação de adequação de dados amostrais a esta distribuição torna-se exigida em grande número devido a quantidade de metodologias que dependem da normalidade. Embora seja usual ou costumeiro se aplicar o teste de normalidade em dados de Geoestatística, objetiva-se com este trabalho mostrar, por meio de uma abordagem teórica, a inviabilidade de tal exigência, pois variáveis estudadas por esta metodologia apresentam um tipo de dependência, a espacial. Além disso, objetivou-se ainda mostrar, através de um estudo prático, que um conjunto de dados que não atenda à normalidade pode ser analisado pela Geoestatística, desde que sejam válidas às pressuposições exigidas pela metodologia.

Palavras-chave: Geoestatística, Normalidade, Dependência Espacial.

ABSTRACT

The basis of statistical inference is to identify the probability distribution of the random variable under study. Probably the best known and distribution likely used is Gaussian-Laplace, or simply normal distribution. Therefore, the appropriateness checking of sample data with this distribution becomes required in large numbers because the amount of methodologies that rely normal. However, despite the popularity of the requirement of normality tests for geostatistics, the objective of this study was to show, through a theoretical approach, the impossibility of such a requirement because variables studied by this method have a type of dependency, spatial. Moreover, the objective was to still show through a case study, a set of data that does not meet the normality can be analyzed by geostatistics, since it meets the assumptions required by the methodology.

Key-words: Geostatistics, Normality, Spatial Dependence

1.INTRODUÇÃO

Na Estatística Clássica, o conhecimento sobre as distribuições de probabilidade das variáveis aleatórias é de fundamental importância em estudos em que se deseja fazer inferência sobre parâmetros da população.

Segundo Ferreira (2009) a distribuição de probabilidade da variável aleatória em estudo será a base da inferência estatística. Entre todas as distribuições conhecidas, provavelmente a distribuição de probabilidade de Gauss-Laplace, ou simplesmente, distribuição normal, é a mais exigida e, conseqüentemente, verificada.

Torman (2012) acrescenta ainda que a distribuição normal é a distribuição mais testada na Estatística Clássica devido à quantidade de teoremas e/ou definições que dependem de tal pressuposição.

De acordo com Ferreira (2009) existem vários testes para a avaliação da normalidade dos dados. Este autor afirma ainda que a importância da realização de testes que chequem a normalidade está associada ao habitual uso dessa condição que alguns pesquisadores fazem e muitas vezes não a verificam.

Assim, pode-se perceber que para a Estatística Clássica, a violação dessa importante pressuposição pode induzir a equivocados resultados e/ou a inapropriadas conclusões.

A Geoestatística tem seus conceitos teóricos baseados em funções de variáveis aleatórias. Segundo esta teoria, também chamada de variáveis regionalizadas, uma variável aleatória indexada $Z(x_i)$ assume diferentes valores em função da localização dentro de uma região de estudo. Assim, uma função aleatória $Z(x_i)$ pode ser considerada um conjunto de variáveis aleatórias que podem assumir dependência entre elas (ISAACS e SRIVASTAVA, 1989).

Segundo Soares (2006) é impossível com uma única realização determinar estatísticas no ponto x_i da função aleatória, como por exemplo, a média e a variância. Uma solução seria assumir alguns graus de estacionariedade da função aleatória (YAMAMOTO e LANDIM, 2013).

De acordo com Santos et al. (2011) como em grande parte das aplicações da Geoestatística só há uma realização, então pressuposições de estacionariedade são

necessárias (estacionariedade de 1ª ordem, 2ª ordem e intrínseca, por exemplo). Para que se possa aplicar Geoestatística pelo menos uma das hipóteses de estacionariedade deve ser satisfeita, ou seja, a estacionariedade do variograma (VIEIRA, 2000).

De acordo com Cressie (1993), a normalidade dos dados não é uma exigência da Geoestatística, mas é conveniente apenas que, no gráfico de distribuição de frequências, geralmente utiliza-se o histograma, o atributo não apresente extremidades da distribuição muito alongadas, o que poderia indicar a presença de dados discrepantes (outliers), o que poderia comprometer as análises.

Landim (2003), abordando o mesmo tema, também cita que o uso do semivariograma para a estimativa por krigagem não exige que os dados tenham distribuição normal, mas a presença de distribuição assimétrica, com muitos valores anômalos, deve ser levada em conta, pois a krigagem é um estimador linear.

Dessa forma, objetiva-se com este trabalho apresentar uma abordagem teórica sobre a inviabilidade da exigência da normalidade para variáveis regionalizadas, ou seja, variáveis que apresentam dependência espacial entre si (em outras palavras, variáveis estudadas através da Geoestatística). Busca-se ainda, apresentar na parte prática do trabalho um estudo com um conjunto de dados que não atenda à normalidade e que ainda assim pode ser analisado usando a Geoestatística.

2. MATERIAL E MÉTODOS

2.1 Abordagem Teórica

Com o intuito de atingir os objetivos deste trabalho foi inicialmente descrito o teste de normalidade proposto por Shapiro e Wilk (1965) e foram apontadas as passagens da derivação da estatística W em que foi exigida a independência dos dados. Este teste foi o escolhido por ser o mais utilizado na literatura. Também foi verificado se, em alguma passagem, havia a possibilidade de dependência.

De acordo com Ferreira (2009) “o teste de Shapiro - Wilk é baseado em estatísticas de ordem da distribuição normal e de seus respectivos valores médios”. Segundo Mood et al. (1974) quando se tem $Y_1 \leq Y_2 \leq \dots \leq Y_n$ em que Y_i são os X_i dispostos em ordem de aumento de magnitude, então Y_i é dito ser a i-ésima estatística de ordem correspondente à amostra aleatória X_1, X_2, \dots, X_n .

De acordo com Mood et al. (1974), considerando as variáveis aleatórias X_1, X_2, \dots, X_n com função densidade $f_{X_1, X_2, \dots, X_n}(\cdot, \dots, \cdot)$ fatora-se esta função nas densidades marginais como segue: $f_{X_1, X_2, \dots, X_n}(x_1, x_2, \dots, x_n) = f(x_1) \cdot f(x_2) \cdot \dots \cdot f(x_n)$ em que $f(\cdot)$ é a densidade de cada x_i . Assim, X_1, X_2, \dots, X_n é dita ser uma amostra aleatória (a.a.) de tamanho n de uma população com densidade $f(\cdot)$.

Shapiro e Wilk (1965) fizeram a derivação da estatística W e em uma dessas passagens segue do teorema dos mínimos quadrados generalizados que a melhor estimativa linear não viesada (BLUE) de μ e σ são aquelas quantidades que minimizam a forma quadrática $(y - \mu 1 - \sigma m)' V^{-1} (y - \mu 1 - \sigma m)$, em que $1' = (1, 1, \dots, 1)$, $y' = (y_1, \dots, y_n)$ é o vetor das observações ordenadas, μ é a média desconhecida de uma distribuição normal, σ é o desvio padrão desconhecido de uma distribuição normal, $m' = (m_1, \dots, m_n)$ é o vetor de valores esperados da estatística de ordem da normal padrão e V é a matriz n x n de covariância correspondente.

Lembrando que o conceito de BLUE considera Y_1, Y_2, \dots, Y_n variáveis aleatórias que são observadas de forma que $E[Y_i] = \tau_i(\theta)$, em que $\tau_i(\cdot)$ são funções conhecidas

que contém parâmetros desconhecidos θ (θ pode ser um vetor de valores). “Para estimar algum θ_j em θ , será considerada somente a classe de estimadores que são funções lineares da variável aleatória Y_i . Nesta classe só serão consideradas subclasses de estimadores que são não viesados para θ_j . Se nesta classe restrita existir um estimador θ_j que tenha menor variância que qualquer outro estimador de θ_j nesta classe restrita, ele é definido ser o melhor estimador não viesado de θ_j (o termo melhor se refere à variância mínima)” (Mood, Graybill e Boes, 1974).

As estimativas de μ e σ estão representadas pelas equações (1) e (2), respectivamente,

$$\hat{\mu} = \frac{m'V^{-1}(m1' - 1m')V^{-1}y}{1'V^{-1}1m'V^{-1}m - (1'V^{-1}m)^2} \quad (1)$$

e

$$\hat{\sigma} = \frac{1'V^{-1}(1m' - m1')V^{-1}y}{1'V^{-1}1m'V^{-1}m - (1'V^{-1}m)^2} \quad (2)$$

Para distribuições simétricas tem-se que $1'V^{-1}m = 0$, e assim a equação (3) representa μ e σ .

$$\hat{\mu} = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y} \quad \text{e} \quad \hat{\sigma} = \frac{m'V^{-1}y}{m'V^{-1}m} \quad (3)$$

De acordo com Casella e Berger (2011) a média é o primeiro momento de uma variável aleatória X e a variância o segundo momento central. Considere um conjunto de variáveis aleatórias X_1, X_2, \dots, X_n . Define-se como o n -ésimo momento de X (μ'_n) como sendo (equação 4):

$$\mu'_n = E(X)^n \quad (4)$$

O n -ésimo momento central de X , μ_n é representado na equação (5):

$$\mu_n = E(X - \mu)^n \quad (5)$$

em que, $\mu = \mu' = E(X)$.

De acordo com Shapiro e Wilk (1965), a estatística do teste W para normalidade é dada pela equação (6):

$$W = \frac{R^4 \hat{\sigma}^2}{C^2 S^2} = \frac{b^2}{S^2} = \frac{(a' y)^2}{S^2} \quad (6)$$

em que, $S^2 = \sum_{i=1}^n (y_i - \bar{y})^2$ representa a estimativa não viesada simétrica usual de

$$(n-1)\sigma^2, R^2 = m'V^{-1}m, C^2 = m'V^{-1}V^{-1}m, a' = (a_1, \dots, a_n) = \frac{m'V^{-1}}{(m'V^{-1}V^{-1}m)^{1/2}}$$

$$\text{e } b = \frac{R^2 \hat{\sigma}}{C}.$$

Portanto, a equação (7) representa a estatística W para o teste de Shapiro-Wilk.

$$W = \frac{\left(\sum_{i=1}^n a_i y_i \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (7)$$

Shapiro e Wilk (1965) fazem considerações sobre o que acontece na falta de normalidade, porém nem se cogita a possibilidade da falta de independência.

2.2. Abordagem Prática

Visando atingir os objetivos deste trabalho foi utilizado uma base de dados referente à variável altimetria de uma área que compreende uma porção de cerca de 3 ha no campus da Universidade Federal de Viçosa, na cidade de Viçosa-MG, Brasil. A região de estudo tem o centróide determinado pela

latitude $20^{\circ}45'49,4''S$ e longitude $42^{\circ}51'59,7''W$. Os dados de altimetria desta base de dados são provenientes de levantamento fotogramétrico realizado por um Veículo Aéreo Não-Tripulado (VANT), sendo referenciado ao sistema geodésico SIRGAS 2000 e representado na projeção UTM (Universal Transversa de Mercator) fuso 23S (Figura 1).

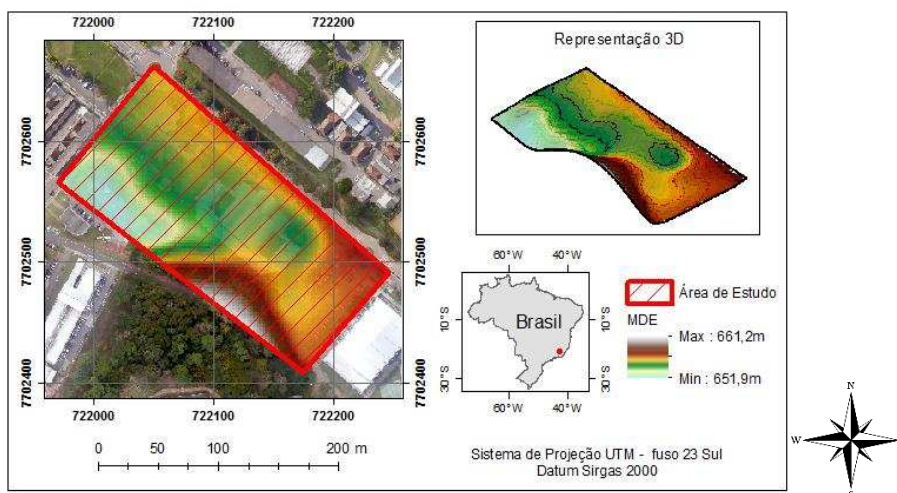


Figura 1 - Representação da área (3ha) de estudo da variável altimetria no Campus da Universidade Federal de Viçosa – MG.

Tais dados compreendem 4867 pontos com altitudes conhecidas, com um espaçamento de 2,5 m em ambas as direções X e Y. Os valores de altitudes apresentam valores mínimos de 651,9 m e valores máximos de 661,2 m. Desta base de dados foram selecionados 2021 pontos amostrais para o estudo.

Inicialmente foi realizada a análise exploratória dos dados que consistiu da obtenção da estimativa da média, variância, desvio-padrão, histograma, coeficiente de variação (CV), coeficiente de assimetria, curtose, valor máximo e mínimo. Esta análise foi seguida pela verificação da condição de normalidade dos dados utilizando o teste proposto por Shapiro-Wilk (1965).

Logo após foi realizada a análise exploratória espacial que consistiu em detectar a presença de outliers, através do gráfico de Box Plot e/ou tendência que podem vir a afetar a identificação e interpretação da dependência espacial através do semivariograma. Assim, realizada esta etapa procedeu-se a análise geoestatística verificando se o conjunto de dados apresentado atendia às pressuposições da mesma.

Na fase de verificação e caracterização da dependência espacial dos dados utilizou-se, em primeiro lugar, a Equação (8) para a obtenção do semivariograma experimental, segundo recomenda Ferreira et al. (2013).

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(x_i) - Z(x_i + h)]^2 \quad (8)$$

em que: $N(h)$ é o número de pares experimentais de observações $(Z(x_i), Z(x_i + h))$ separados por um vetor de distâncias \mathbf{h} .

De acordo com Santos et al. (2011), ao se obter o semivariograma experimental, este será ajustado através de modelos teóricos. O ajuste do semivariograma é uma importante fase, pois os parâmetros necessários para efetuar a krigagem dependem do modelo de semivariograma ajustado e, portanto, se no ajustamento do modelo não forem tomados alguns cuidados recomendados por Vieira (2000), a krigagem conterá erros de estimação e poderá não interpolar de maneira ótima (VILELA, 2004). Todo o procedimento da análise geoestatística utilizado neste trabalho foi sugerido por Ferreira et al. (2013).

Computacionalmente, a parte de geoprocessamento do trabalho foi realizada no software ArcGis (ESRI, 2014) e a parte estatística do estudo (análise geoestatística e testes) foi realizada no software R (R Development Core Team, 2015), principalmente no pacote geoR (RIBEIRO JUNIOR e DIGGLE, 2001).

3. RESULTADOS E DISCUSSÃO

Na Tabela 1, é apresentada a análise exploratória dos dados. Verifica-se que os dados apresentaram um comportamento homogêneo do ponto de vista estatístico, isto é, os valores das altitudes apresentaram baixa variação, o que pode ser comprovado pelo coeficiente de variação (CV), segundo a classificação proposta por Warrick e Nielsen (1980).

Tabela 1: Estatística descritiva para a variável altimetria da área de estudos no Campus da Universidade Federal de Viçosa - MG

Estatísticas	Estimativas
N	2021
Média (m)	655.2
Mediana (m)	655.2
Variância (m ²)	2.4
Desvio (m)	1.6
CV(%)	0.2
Assimetria	0.4
Curtose	4.4
Mínimo (m)	651.9
Máximo (m)	661.2

Pela análise estatística descritiva (Tabela 1), observa-se que os valores das estimativas de tendência central (média e mediana) foram iguais, inicialmente indicando que a distribuição é simétrica. Porém, de acordo com o coeficiente de assimetria obtido nas análises, este indica que a distribuição é assimétrica à direita, de acordo com a classificação de Ferreira (2009). Pode-se perceber ainda que a distribuição de frequências dos dados apresentou certo “achatamento”, muitas vezes classificado como um comportamento leptocúrtica (FERREIRA, 2009).

Assim, e em complementação às informações da Tabela 1, foi construído o histograma dos dados amostrados, conforme Figura 2.

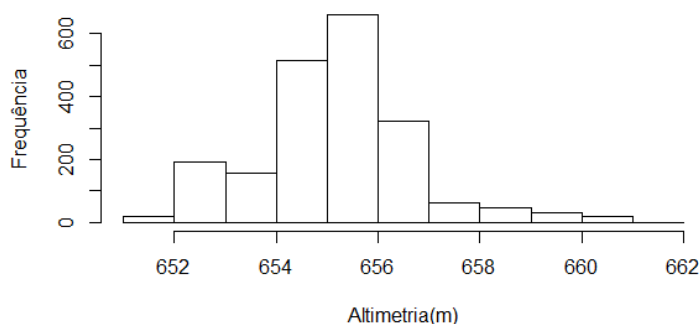


Figura 2 – Distribuição de frequências da variável altimetria da área (3ha) de estudo no campus da Universidade Federal de Viçosa – MG.

Através da Figura 2 e pela análise descritiva dos dados pode-se perceber que o histograma é assimétrico à direita. Portanto, nota-se, através de uma inspeção visual, que os dados não seguem uma distribuição normal. Para confirmar essa afirmação procedeu-se o teste estatístico de normalidade proposto por Shapiro e Wilk (1965).

Foi obtido um p-valor menor que 2.2×10^{-16} e verificado que o teste de Shapiro e Wilk (1965) rejeita a hipótese de normalidade para a variável altimetria nessa região de estudo.

O teste de normalidade realizado neste trabalho foi feito pró-forma, pois a pressuposição teórica, não foi comprovada.

O estudo descritivo dos dados não informa nada sobre a estrutura de dependência espacial. Portanto, é feita uma análise minuciosa acerca das características espaciais dos dados afim de se realizar uma correta interpolação pelo método da krigagem.

Na Figura 3 tem-se a representação por quartis das amostras. Os pontos em azul indicam os valores com menor altitude, seguidos pelos pontos verdes, pontos amarelos e por último indicando maior altitude os pontos em vermelho.

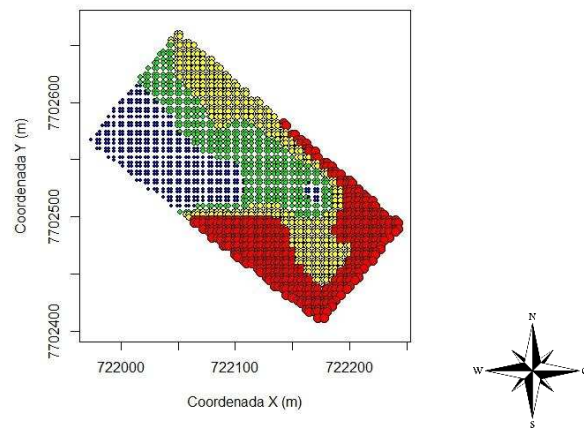


Figura 3 - Representação por quartis da variável altimetria da área (3ha) de estudo no campus da Universidade Federal de Viçosa - MG.

Dando continuidade à análise espacial procedeu-se a verificação de presença de tendência nos dados. Para isso, construiu-se o gráfico de tendência no Arcgis 10.2 (Figura 4).

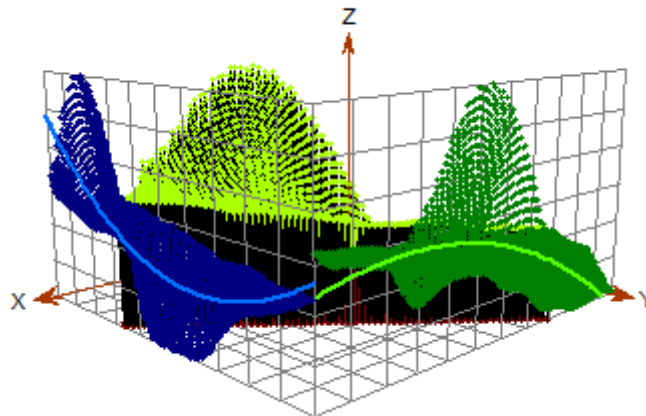


Figura 4 - Gráfico mostrando a tendência de segunda ordem presente nos dados da variável altimetria.

Analisando a Figura 4 pode-se observar a presença de tendência de segunda ordem nos dados de altimetria. Segundo Journel e Huijbrets (1978), Isaaks e Srivastava (1989), Ribeiro Júnior (1995) e Vieira (2000), a condição de não tendência é fundamental para se admitir a estacionaridade de primeira ordem. Porém, de acordo com Santos et al. (2011) não é necessário fazer krigagem universal para a retirada de

tendência dos dados, basta fazer a krigagem simples uma vez que esta gera menor variância de krigagem.

Ao ser comprovada esta dependência, percebe-se que não há porque realizar o teste de normalidade proposto por Shapiro e Wilk (1965) para estes dados, pois o teste parte da pressuposição de que os dados são independentes e isto não acontece.

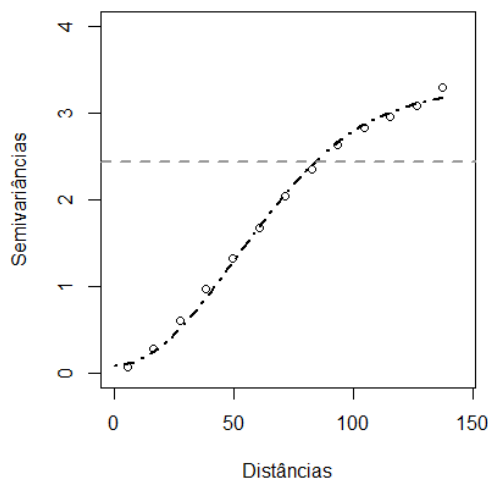


Figura 5 – Semivariograma experimental direcional e modelo ajustado das direções: N-S(0°), E-W (90°), SW-NE (45°) e NW-SE (135°)

Foi possível através da dependência espacial ajustar um modelo teórico aos dados. O modelo teórico ajustado foi o Gaussiano, com efeito pepita de $0,09 \text{ m}^2$, patamar de $3,1 \text{ m}^2$ e um alcance de $123,3 \text{ m}$. O efeito pepita próximo de zero pode ter sido conseguido devido ao tamanho amostral. Ao se analisar o semivariograma verificase que o patamar não coincide com a variância dos dados. Isso pode ter ocorrido devido a presença de tendência, e presença de anisotropia. Após detectada a presença de tendência o processo foi continuado conforme recomendação de Santos et al. (2011) que como citado acima sugere a realização da krigagem simples em um conjunto de dados com tendência por esta apresentar menor variância de krigagem em relação a krigagem universal. O melhor método de ajuste do modelo foi o OLS (Ordinary Least Square). Esse método foi escolhido dentre os outros métodos de ajuste por seguir as recomendações de Vieira (2000).

De acordo com Vieira (2000), é importante verificar na validação cruzada a regressão linear simples entre valores observados e preditos. As principais estatísticas

que devem ser avaliadas são: o coeficiente da regressão $\hat{\beta}_1$ (que deve estimar o parâmetro $\beta_1 = 1$) e o intercepto $\hat{\beta}_0$ (que deve estimar o parâmetro $\beta_0 = 0$). Além disso, este autor recomenda a avaliação do estimador R^2 (que deve estimar a unidade também).

Os resultados da validação cruzada, descrevem a precisão do modelo ajustado. Os principais estimadores da validação cruzada com as respectivas estatísticas, recomendados por Vieira (2000), Santos et al. (2011) e Ferreira et al. (2013), são apresentados na Tabela 2.

Tabela 2: Apresentação das principais estatísticas da validação cruzada.

$\hat{\beta}_1$	$\hat{\beta}_0$	R^2
1,0	-2,5	0,9924

De acordo com a Tabela 2 verifica-se que os valores de $\hat{\beta}_1$ e R^2 são bem próximos da unidade conforme recomendado por Vieira (2000). Isso significa que a função foi bem ajustada. O valor de $\hat{\beta}_0$ encontra-se um pouco afastado da condição ideal.

Por conhecimento da área, de fato o mapa gerado pela krigagem simples (Figura 6) representa o comportamento do relevo na região.

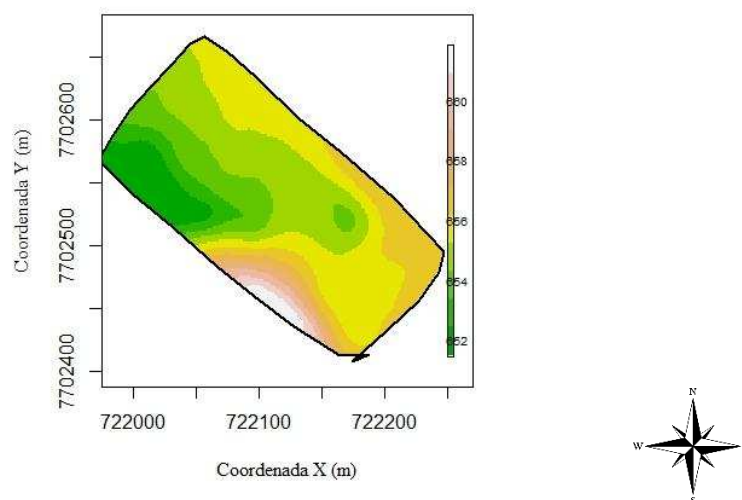


Figura 6 - Mapa de distribuição espacial da variável altimetria gerado pela krigagem simples.

4. CONCLUSÃO

Neste trabalho buscou-se apresentar os aspectos teóricos e práticos sobre a inviabilidade da exigência de normalidade para dados que apresentam dependência espacial entre si.

Como o teste de normalidade exige a independência entre os dados, comprova-se que o mesmo não interfere em uma análise Geoestatística.

De acordo com os estudos práticos realizados neste trabalho foi possível comprovar que é possível realizar uma análise Geoestatística em que um conjunto de dados que não atenda a normalidade, mas que atenda pressuposições dessa análise.

Percebeu-se ainda que há falhas quanto à existência de testes para verificar a independência, o que fica como sugestão para trabalhos futuros.

REFERÊNCIAS BIBLIOGRÁFICAS

ArcGIS. Version 10.2.1. Redlands, CA: Esri, 2014.

CASELLA, G.; BERGER, R.L. **Statistical Inference**. Duxbury. 2011.

CRESSIE, N.A.C. **Statistics for spatial data**. New York, J. Wiley. 1993.

FERREIRA, D.F. **Estatística Básica** – 2^a. ed. Lavras: Editora UFLA. 2009.

FERREIRA, I.O.; SANTOS, G.R.; RODRIGUES, D.D. Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. **Revista Brasileira de Cartografia** (Online), v.65, n° 5, p.831-842. 2013.

ISAAKS, E.H; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York, Oxford University Press, 1989.

JOURNEL, A.G.; HUIJHREGTS, C. **Mining geostatistics**. San Diego CA: Academic Press. 1978.

LANDIM, P.M.B. **GEOEAS: Introdução à análise geoestatística**. Rio Claro:UNESP/IGCE. 2003.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D.C. **Introduction to the theory of statistics**. New York - US: McGraw-Hill. 1974.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2010. Disponível em: <<http://www.r-project.org>>. Acesso em: 30 Apr. 2015.

RIBEIRO JUNIOR, P.J; DIGGLE, P.J. **geoR: a package for geostatistical analysis**. R-NEWS, London. v.2, p.15-18. 2001.

RIBEIRO JÚNIOR, P.J. **Métodos geoestatísticos no estudo da variabilidade espacial de parâmetros do polo**. 1995. 99f. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura "Luiz de Queiroz", Piracicaba 1995.

SANTOS, G.R.; OLIVEIRA, M.S.; LOUZADA, J.M.; SANTOS, A.M.R.T. Krigagem Simples versus Krigagem universal: qual o preditor mais preciso? **Energia na Agricultura**, v.26, n°2, p.49-55. 2011.

SHAPIRO, S.S; WILK, M.B. An analysis of variance test for normality (complete samples), **Biometrika**, v.52, p. 591–611. 1965.

SOARES, A. **Geoestatística para as ciências da terra e do ambiente**. 2.ed. Lisboa: IST Press, 2006.

TORMAN, V.B.L.; COSTER, R.; RIBOLDI, J. Normalidade de variáveis: métodos de verificação e comparação de alguns testes não-paramétricos por simulação. **Revista HCPA**, v.32, p. 227-234, 2012.

VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, RF et al. (Ed.). **Tópicos em ciência do solo**. Sociedade Brasileira de Ciência do Solo. Viçosa, v.1, p. 1-53. 2000.

VILELA, L.C. **Comparação entre as redes neurais artificiais e o método de interpolação krigagem aplicados à pesquisa agronômica**. Tese (Doutorado). UNESP. Faculdade de Ciências Agronômicas. Campus de Botucatu. 2004.

WARRICK, A.W; NIELSEN, D.R. **Spatial variability of soil physical properties in the field**. In: HILLEL, D., ed. Application of soil physics. New York, Academic Press, p.319-344. 1980.

YAMAMOTO, J.K.; LANDIM, P.M.B. **Geoestatística – conceitos e aplicações**. São Paulo: Oficina de textos, 2013.

CAPÍTULO 3

AVALIAÇÃO DA INFLUÊNCIA DA SIMETRIA DOS DADOS EM UMA ANÁLISE GEOESTATÍSTICA

RESUMO

A Geoestatística é utilizada para estudar a continuidade de um fenômeno e se caracteriza por utilizar toda a informação disponível, incluindo a posição geográfica dos dados amostrais, para caracterizar a variável em estudo. O objetivo deste trabalho foi de avaliar o comprometimento de uma análise geoestatística para dados assimétricos, adotando diferentes transformações de dados, partindo da assimetria natural dos dados chegando a uma simetria aceitável estatisticamente. Para isso foram utilizados dois diferentes conjuntos de dados, um relacionado a dados de altimetria e outro relacionado a variáveis de solo. Para o primeiro conjunto de dados verificou-se que as vantagens conseguidas na análise geoestatística, para os dados transformados, não foram significativas frente às desvantagens obtidas, entre estas, destaca-se o acréscimo significativo da variância de krigagem, que era de 0.02 para os dados originais e passou a 51.8 para a melhor transformação. Porém, para o segundo conjunto de dados ocorreu o contrário. A transformação logarítmica foi mais eficiente na análise geoestatística. Porém, observou-se que o coeficiente de variação não foi preservado e a variância dos dados originais era 30 vezes maior que a dos dados após a transformação.

Palavras-chave: Geoestatística, simetria, transformação.

ABSTRACT

Geostatistics is used to study the continuity of a given phenomenon and is characterized for utilizing all the available information, including the geographical position of the sampled data, in order to characterize a given variable in study. The objective of this paper is to evaluate the commitment of a geostatistical analysis for asymmetric data, adopting different data transformations, from the natural asymmetry of the data to a statistically acceptable symmetry. Two sets of different data were used, in which one group was related to the altimetry data, while the other was related to soil variables. For the former set, it was observed that the advantages obtained with the use of geostatistics, for the transformed data, were not significant in comparison to the disadvantages obtained. Among these, the significant increase, from 0.02, with the original data, to 51.8, according the best fit transformation, of the kriging variation stands out. However, it was observed that the variation coefficient was not preserved and the variance of the original data was 30 times higher than the data after the transformation.

Key-words: Geostatistics, symmetry, transformation.

1.INTRODUÇÃO

A Geoestatística é utilizada principalmente no estudo da continuidade espacial de um fenômeno. Este é um ramo da Estatística Espacial que se destaca por utilizar toda a informação disponível, inclusive a posição geográfica dos dados amostrais, visando caracterizar a dependência espacial da variável em estudo e interpolar através de um BLUP – Best Linear Unbiased Predictor (melhor preditor linear não viesado) sem viés e de variância mínima (VIEIRA, 2000; SANTOS et al., 2011 e YAMAMOTO e LANDIM, 2013).

A base teórica da Geoestatística vem da teoria das variáveis regionalizadas. De acordo com esta teoria, a diferença entre os valores de um atributo tomados em dois pontos distintos depende da distância entre eles, conforme Matheron (1963). Portanto, a diferença entre os valores tomados em pontos mais próximos deve ser menor do que a diferença entre dois pontos mais distantes. Assim, cada valor tem consigo uma forte dependência dos valores de sua vizinhança, ilustrando uma continuidade espacial (ISAAKS e SRIVASTAVA, 1989).

Contudo para a base teórica da Geoestatística, a normalidade dos dados não é uma exigência, porém é conveniente a utilização do histograma para representar a distribuição de frequência e o atributo não deve apresentar caudas muito alongadas, o que poderia representar a presença de dados discrepantes (outliers) e isso poderia comprometer as análises (CRESSIE, 1993).

Isaaks e Srivastava (1989) afirmam que mais importante que a normalidade dos dados na Geoestatística, é que os variogramas apresentem patamares bem definidos, permitindo aceitar a ocorrência da hipótese intrínseca. Mas, é comum encontrar na literatura especializada a busca pela simetria dos dados. Para tornar uma distribuição dos dados simétrica e para remover a tendência da variância é utilizada transformação dos dados (ANDRIOTTI, 2002).

Assim, o objetivo deste trabalho foi avaliar o comprometimento de uma análise geoestatística para dados assimétricos e simétricos. Para tanto, dois conjuntos de dados foram utilizados, sendo um deles obtido de um levantamento batimétrico, para a variável profundidade, de um represamento do Ribeirão São Bartolomeu, no Campus da Universidade Federal de Viçosa – MG/Brasil e outro, para a variável solo seco,

proveniente de um experimento conduzido em uma área perto do Campus da Universidade de Auburn – Alabama/USA.

2. MATERIAL E MÉTODOS

Os dados do primeiro conjunto foram coletados em dezembro de 2010 no levantamento batimétrico de um dos principais represamentos do Ribeirão São Bartolomeu, localizado no Campus da Universidade Federal de Viçosa (UFV), no município de Viçosa, estado de Minas Gerais, Brasil. A área de estudo possui aproximadamente 8800 m², 150 m de comprimento e 66 m de largura, representada na Figura 1.

Foram coletadas 577 amostras contendo as coordenadas planimétricas e as respectivas profundidades.

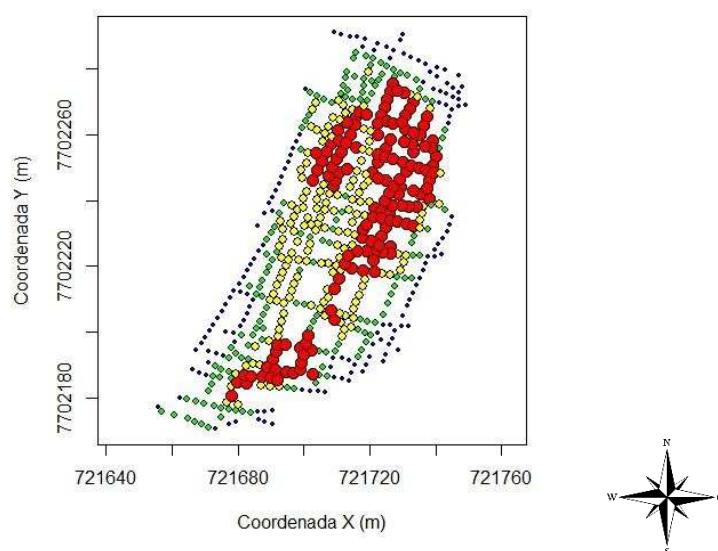


Figura 1 – Representação por quartis da malha amostral do levantamento batimétrico de um represamento do Rio São Bartolomeu

O segundo conjunto de dados foi obtido a partir de um experimento conduzido em uma área de 2 ha perto do campus e em uma estação experimental da Auburn University em Auburn, Alabama, no solo de Pacolet arenoso (fino, caulínico, térmico Typic Kanhapludults). A área de estudo está localizada na latitude 32° 21 '15 " N e longitude 85° 17' 30" W.

As medições dos índices de cone foram coletadas em duas repetições de 20 cm de distância em cada uma das células de 10 m x 10 m cobrindo uma área de 2 ha usando um Penetrômetro de Cone de Solo de Sonda Múltipla (MPSCP) montado em trator equipado com GPS para posicionamento de campo (Figura 2) (ASAE, 2002a, b, Raper et al., 1999).



Figura 2 - Penetrômetro de cone de solo de múltiplas sondas (MPSCP) com cinco sondas e sistema GPS para medição de índices de cone em solo de argila arenosa Pacolet em Auburn, AL

Os dados do índice de cone foram adquiridos a 25 Hz a uma profundidade de 60 cm. Os dados de elevação ao longo do campo (Figura 3) foram obtidos utilizando um receptor de GPS RTK de duas frequências, AgGPS® 214. As medidas foram obtidas em condições de umidade do solo de 11,25% (d.b.) e condições de umidade do solo seco de 9,83% (d.b.) na camada de arado superior (0-30cm), respectivamente. A amostragem dos dados do índice de cone em duas réplicas em 200 pontos de grade (400 posições dos dados de índice de cone) levou um dia de trabalho completo (8 horas). Dentro de cada condição de amostragem não houve eventos de precipitação que

causariam variações indesejáveis nos dados do índice de cone devido às variações de umidade do solo. As condições de umidade do solo para a camada de arado superior (0-30 cm) foram monitoradas após a primeira medição da condição de umidade do solo úmido até o solo secar na camada de arado superior (0-30 cm).

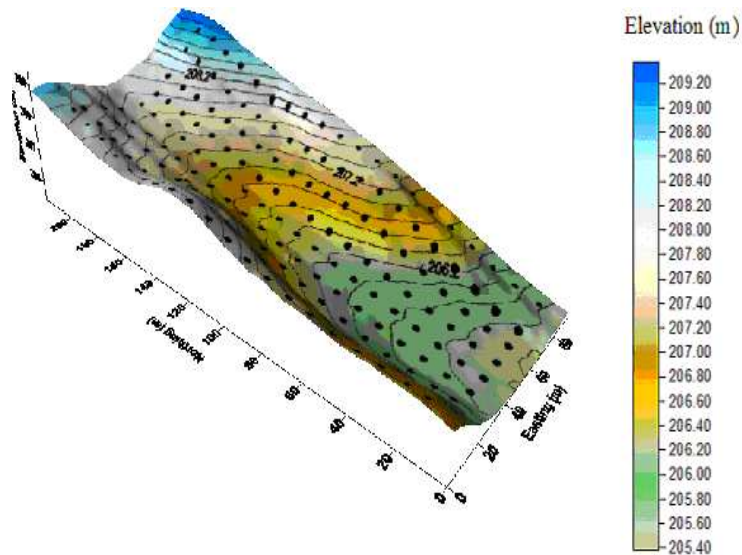


Figura 3 - O modelo digital de elevação representa o campo de amostragem do solo argiloso de Pacolet. As marcas indicam os pontos de amostragem para a medição da resistência à penetração do solo.

Com o intuito de atingir os objetivos, realizou-se uma análise descritiva dos dados e posteriormente uma análise geoestatística. Essas etapas foram realizadas tanto para os dados sem transformação como para os dados transformados.

Neste estudo foram utilizadas diferentes transformações, a cúbica, a de Box-Cox e a logarítmica. As equações (1), (2) e (3) representam essas transformações.

$$f(x) = x_i^3 \quad (1)$$

$$y^{(\lambda)} = \begin{cases} \frac{y^\lambda - 1}{\lambda}, & \lambda = 1,6 \text{ e } \lambda = 2,5 \\ \ln y, & \lambda = 0 \end{cases} \quad (2)$$

$$f(x) = \ln x \quad (3)$$

Foram escolhidas essas transformações, pois para os conjuntos de dados utilizados neste trabalho, estas foram as que mais se aproximaram da simetria. Para o primeiro conjunto de dados foram utilizadas as transformações, cúbica e de Box-Cox, sendo que para esta última foram utilizados dois valores diferentes para λ . Já para o segundo conjunto foi feita a transformação logarítmica.

Inicialmente foi realizada a análise exploratória dos dados que consistiu da obtenção da estimação da média, variância, desvio-padrão, histograma, coeficiente de variação (CV), coeficiente de assimetria, curtose, valor máximo e mínimo.

Em relação ao coeficiente de assimetria, este mostra o quanto uma variável está afastada em relação a um valor central, caracterizando como e quanto a distribuição de frequência se afasta da simetria (FERREIRA, 2009). Ainda, de acordo com Ferreira (2009) o coeficiente de assimetria é calculado com a seguinte fórmula,

$$\sqrt{b_1} = \frac{m_3}{(m_2)^{\frac{3}{2}}} \quad (4)$$

em que $\sqrt{b_1}$ é o estimador do coeficiente de assimetria populacional $\sqrt{\beta_1}$, m_2 é o segundo momento (variância) e m_3 é o terceiro momento (assimetria).

Populações cuja distribuição é simétrica, apresentam $\sqrt{\beta_1} = 0$. No caso das distribuições assimétricas, quando $\sqrt{\beta_1} > 0$ a distribuição é classificada como assimétrica à direita (assimetria positiva) e quando $\sqrt{\beta_1} < 0$ a distribuição é classificada como assimétrica à esquerda (assimetria negativa) (FERREIRA, 2009).

Após a análise exploratória dos dados foi realizada a análise exploratória espacial que consistiu em detectar a: presença de outliers, utilizando o box-plot, e análise de tendência, através do gráfico de tendência do pacote geoR (Ribeiro Júnior e Diggle, 2001) do software R (R Development Core Team, 2016), que podem afetar a identificação e interpretação da dependência espacial. Assim, realizada esta etapa procedeu-se a análise geoestatística verificando se os conjuntos de dados apresentados atendiam às pressuposições da mesma.

Para efeitos de comparação, foram utilizados alguns parâmetros de qualidade de ajuste estatístico, sendo a maioria obtida na autovalidação *leave-one-out* da Geoestatística. Os parâmetros foram os seguintes: coeficiente de assimetria (obtido na transformação dos dados), média e desvio padrão dos resíduos, variância do erro, coeficientes da regressão linear simples entre os valores observados e preditos, bem como o coeficiente de determinação da mesma. A autovalidação, de acordo com Mello et al. (2005), é o procedimento que consiste em estimar cada ponto observado supondo que ele não existe no processo de predição, levando em consideração a variabilidade espacial local expressa nas primeiras distâncias do variograma. Ao final geram-se n valores observados e seus correspondentes preditos (VIEIRA, 2000; SANTOS, 2010).

3. RESULTADOS E DISCUSSÃO

Primeiramente são apresentadas todas as análises para o conjunto de dados proveniente do levantamento batimétrico e em seguida as análises para o conjunto de dados do solo.

Na Figura 4 estão representados os histogramas relativos ao primeiro conjunto de dados. O histograma (a) é o que representa o conjunto de dados original, seguido pelo histograma (b) que é relativo à transformação cúbica e os histogramas (c) e (d) que são das transformações de Box-Cox com $\lambda = 1,6$ e $\lambda = 2,5$ respectivamente. Nota-se ainda que os histogramas (a) e (c) apresentam uma assimetria à esquerda, conforme classificação apresentada por Ferreira (2009).

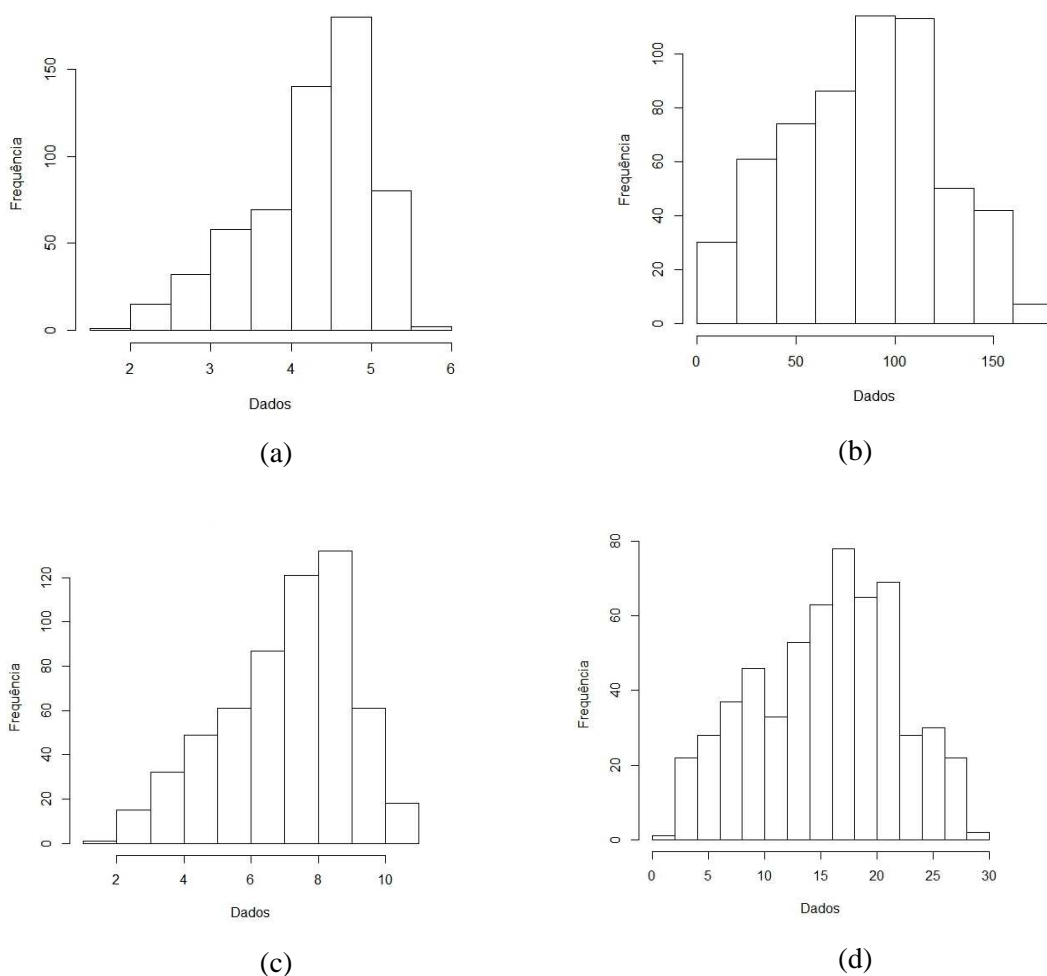


Figura 4 - Histograma dos dados originais (a) e das transformações cúbicas (b), Box-Cox ($\lambda = 1,6$) (c) e Box-Cox ($\lambda = 2,5$) (d), respectivamente.

Também, de acordo com a classificação de Ferreira (2009), os histogramas (b) e (d) aproximam-se da simetria dos dados. Isso pode ser confirmado ao se analisar os coeficientes de assimetria na Tabela 1.

Tabela1: Valores dos coeficientes de assimetria dos dados originais e transformados do levantamento batimétrico de um represamento do Ribeirão São Bartolomeu – Viçosa – MG.

Dados Originais	Transformação Cúbica	Transformação Box-Cox $\lambda = 1.6$	Transformação Box-Cox $\lambda = 2.5$
-0,72	-0,02	-0,52	-0,18

Pela Tabela 1 verifica-se que os coeficientes de assimetria foram negativos, indicando, mesmo que levemente, uma certa assimetria à esquerda, mesmo após a transformação dos dados. Contudo, pela estimativa do coeficiente de assimetria da transformação cúbica pode-se afirmar que a simetria foi atingida, conforme classificação de Ferreira (2009).

Avaliando os conjuntos de dados em relação à média dos resíduos padronizados que foram obtidos na autovalidação *leave-one-out*, pode-se perceber (Tabela 2) que a transformação de Box-Cox com $\lambda = 2,5$ (juntamente com a transformação cúbica), foi bem pequena. Segundo Vieira (2000) essa média deveria ser nula. Em relação ao desvio padrão todas as transformações apresentaram valores bem próximos. O esperado que o desvio padrão tenha valor unitário.

Tabela 2: Valores da média e dos desvios padrão dos resíduos obtidos pela autovalidação

	Média	Desvio Padrão
Dados originais	-0,002	0,419
Transformação Cúbica	0,001	0,438
Transformação Box-Cox $\lambda = 1.6$	-0,002	0,425
Transformação Box-Cox $\lambda = 2.5$	-0,001	0,400

Conforme recomenda Vieira (2000), a regressão linear simples entre valores observados e preditos deve apresentar R^2 (coeficiente de determinação da RLS entre dados observados e dados preditos) bem próximo da unidade, assim como o coeficiente da regressão “a”. O intercepto “b” deveria ser teoricamente nulo. Portanto de acordo com a Tabela 3 a estimativa do coeficiente da regressão foi bem próximo para todas as análises. Porém a estimativa do intercepto “b” se diferencia mais na transformação cúbica.

Tabela 3: Valores dos coeficientes da regressão e coeficiente de determinação obtidos pela autovalidação.

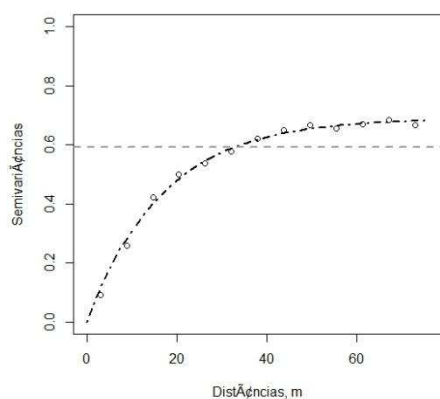
	a	b	R ²
Dados originais	1,038	-0,164	0,971
Transformação Cúbica	1,033	-2,784	0,966
Transformação Box-Cox $\lambda = 1,6$	1,037	-0,261	0,969
Transformação Box-Cox $\lambda = 2,5$	1,040	-0,623	0,967

Quanto à estimativa da variância do erro (Tabela 4), pode-se perceber o quanto a transformação foi ineficiente para este conjunto de dados, pois variou de 0,02 (dados originais) para 51,81 (transformação cúbica dos dados). Assim, fica evidente que apesar da pequena melhoria em alguns indicadores de qualidade de ajuste (estimativas de alguns parâmetros), por meio da variância do erro (um tipo de variância de krigagem segundo Yfantis et al., 1987), não se justifica a transformação de dados para esta variável em estudo.

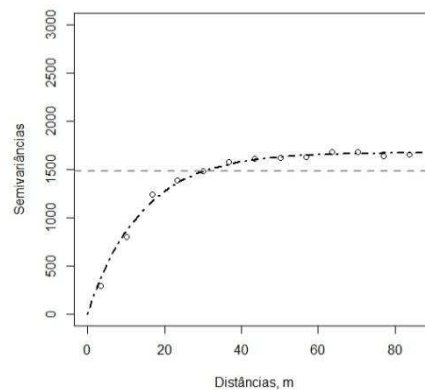
Tabela 4: Valores da variância dos erros obtidos pela autovalidação

Dados Originais	Transformação Cúbica	Transformação Box-Cox $\lambda = 1,6$	Transformação Box-Cox $\lambda = 2,5$
0,02	51,81	0,11	1,34

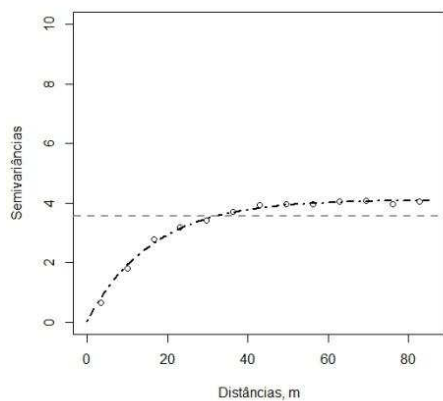
Na Figura 5 foram apresentados os variogramas experimentais obtidos para o conjunto de dados sem transformação e para as transformações. Além disso, foram apresentados os quatro variogramas com o modelo exponencial ajustado às semivariâncias estimadas.



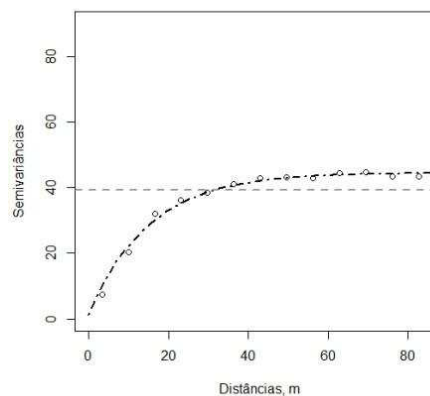
(a)



(b)



(c)



(d)

Figura 5 - Variogramas dos dados originais (a), da transformação cúbica (b), transformação Box-Cox ($\lambda = 1,6$) (c) e da transformação Box-Cox ($\lambda = 2,5$) (d), respectivamente.

Feitas as análises para o primeiro conjunto de dados apresenta-se então todos os resultados das análises relativas ao segundo conjunto de dados. Assim, os histogramas relativos estão representados na Figura 6.

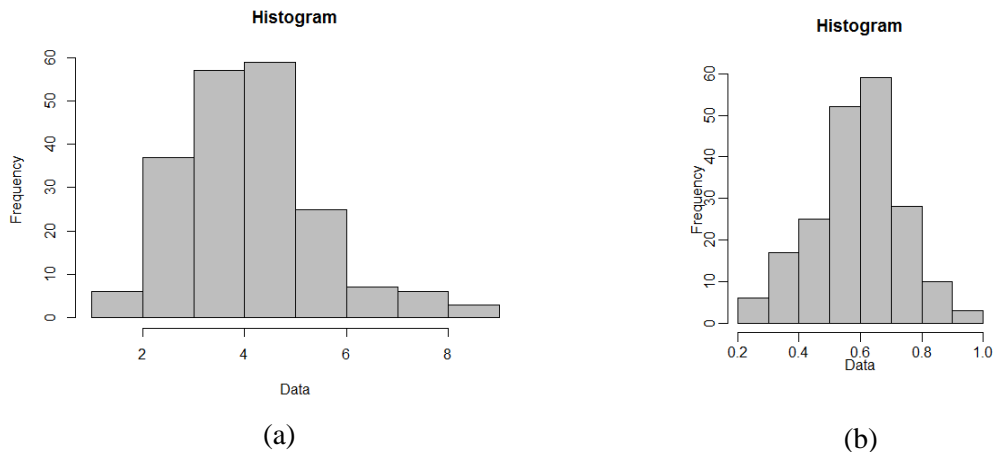


Figura 6 - Histograma dos dados originais (a) e da transformação logarítmica (b) respectivamente

Pela Figura 6 pode-se verificar que a simetria foi alcançada ao se realizar a transformação logarítmica, de acordo com a classificação de Ferreira (2009). Os valores dos coeficientes de assimetria estão apresentados na Tabela 5.

Tabela 5: Valores dos coeficientes de assimetria dos dados originais e transformados da variável resistência a penetração

Dados	Coefficiente de assimetria
Originais	0,77
Transformação log	-0,16

Verifica-se pela Tabela 5 que o coeficiente de assimetria que era de 0,77, após a transformação logarítmica passou a ser de -0,16. Assim, pela classificação de Ferreira (2009), o conjunto de dados passou a ser simétrico.

Mood et al. (1974) mostra que é possível comparar os dados originais e os dados transformados através do intervalo de confiança para os resíduos, verificando a intersecção entre estes. Os intervalos de confiança, para os dois conjuntos de dados, estão apresentados na Tabela 6.

Tabela 6: Intervalo de confiança dos resíduos dos dados originais e transformados da variável resistência à penetração

Dados	Intervalo de Confiança
Originais	$-2,72 \leq \mu \leq 2,72$
Transformação log	$-0,26 \leq \mu \leq 0,26$

Observa-se pela Tabela 6 que os dois intervalos de confiança possuem intersecção. Contudo fica evidente que o intervalo de confiança dos dados transformados ficou mais concentrado em torno da média esperada dos resíduos.

Foram obtidos também os coeficientes de variação para o conjunto de dados originais e o para os dados transformados que foram respectivamente de 32,90 e 24,23.

O coeficiente de variação fornece a variação dos dados obtidos em relação à média. Assim, quanto menor seu valor mais homogeneidade os dados apresentarão. Portanto houve uma maior variação quando foi realizada a transformação logarítmica. Assim, os desvios atingem 24,23% em relação à média (Gomes, 2000).

Na Tabela 7 são apresentados os valores das variâncias dos dados originais e transformados.

Tabela 7: Variância dos dados originais e transformados da variável de resistência a penetração

Dados	Var
Originais	0,59
Transformação log	0,02

De acordo com a tabela 7 observou-se que ocorre uma maior variabilidade no conjunto dos dados originais. Esta é 30 vezes maior que a variabilidade dos dados transformados.

Assim, notoriamente, a estrutura da média e variabilidade dos dados foi alterada devido o processo de transformação.

4. CONCLUSÃO

Verificou-se que para o primeiro conjunto de dados, a melhor transformação foi a cúbica, pois apresentou menor coeficiente de assimetria e a média dos resíduos mais próxima de zero em relação as outras transformações. Porém as vantagens conseguidas na análise geoestatística não foram significativas frente às desvantagens obtidas, pois, destaca-se o acréscimo significativo da variância da krigagem, que era para os dados originais de 0,0177 e passou a 51,8138 para a melhor transformação (cúbica).

Conclui-se que a transformação cúbica não atendeu a todos os critérios de decisão e, portanto, a falta de simetria nos dados não exerceu influência na modelagem geoestatística.

Em relação ao segundo conjunto de dados observou-se que a transformação logarítmica obteve melhores resultados em relação ao coeficiente de assimetria, a média dos resíduos que foi mais próxima de zero e o desvio padrão que foi o mais próximo de 1.

Ainda pode-se observar que em relação à análise geoestatística a transformação logarítmica continuou sendo a melhor.

Ao analisarmos o coeficiente de variação, o mesmo não é preservado e a variância dos dados originais é 30 vezes maior que os dados transformados.

Portanto, a sugestão é que não se faça a transformação, pois esta muda totalmente a estrutura dos dados.

REFERÊNCIAS BIBLIOGRÁFICAS

- ANDRIOTTI, J.L.S. Notas de Geoestatística. **Acta Geologica Leopoldensia** (Unisinus), v.XXV, n.55, p. 3 – 14. 2002.
- CRESSIE, N. **Statistics for spatial data**. New York: John Willey. 1993.
- FERREIRA, D.F. **Estatística Básica**. 2a ed. Lavras: Editora UFLA. 2009.
- GOMES, F.P. **Curso de estatística experimental**. 14.ed. Piracicaba: Degaspari. 2000.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **An introduction to applied geostatistics**. New York: Oxford University Press. 1989.
- MATHERON, G. Principles of geostatistics. **Economic Geology**. El Paso, v.58, p. 11246-66. 1963.
- MELLO, J.M.; BATISTA, J.L.F.; RIBEIRO JUNIOR, P. J.; OLIVEIRA, M. S. Ajuste e seleção de modelos espaciais de semivariograma visando à estimativa volumétrica de *Eucalyptus grandis*. **Scientia Forestalis**, Piracicaba, v.1, n.1, p. 25-37. 2005.
- MOOD, A.M.; GRAYBILL, F.A.; BOES, D.C. Introduction to the theory of statistics. New York - US: McGraw-Hill. 1974.
- RAPER, R.L., WASHINGTON, B.H., JARRELL, J.D. A Tractor - Mounted –Multiple Probe - Soil Cone - Penetrometer. *App. Eng. Agr.* v.15, nº 4, p.287-290. 1999.
- R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2016.
- RIBEIRO JUNIOR, P.P.; DIGGLE, P.P. GeoR: a package for geoestatistical analysis. **R News**, Rochester, v. 1, n. 2, p. 15-18. 2001.
- SANTOS, G.R. **Hierarquização geométrica dos preditores geoestatísticos**. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Setor de Ciências Exatas, Universidade Federal de Lavras, Lavras. 2010.
- SANTOS, G.R. dos ; OLIVEIRA, M.S. de ; LOUZADA, J.M. ; SANTOS, A.M.R.T. KRIGAGEM SIMPLES versus KRIGAGEM UNIVERSAL: QUAL O PREDITOR MAIS PRECISO?. **Energia na agricultura** (UNESP. Botucatu. CD-Rom), v. 26, p. 49-55. 2011.
- VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In. NOVAES, R.F.; ALVARES V.H.; SCHAEFER, C.E.G.R. **Tópicos em ciências do solo**. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, v.1. p.2-54. 2000.

YAMAMOTO, J.K.; LANDIM, P.M.B. **Geoestatística: conceitos e aplicações.** São Paulo: Oficina de Textos. 2013.

CAPÍTULO 4

MODELAGEM DIGITAL DE PROFUNDIDADE ATRAVÉS DE INTERPOLADORES GEOESTATÍSTICOS

RESUMO

O objetivo deste capítulo é comparar três interpoladores geoestatísticos e verificar qual deles é mais preciso para obtenção de Modelos Digitais de Profundidade (MDP). Estes modelos são obtidos através de levantamentos batimétricos que fornecem valores da profundidade representada de forma espacialmente contínua. Para este estudo foram escolhidos os métodos de krigagem simples, krigagem universal e krigagem regressão. Para a comparação entre esses métodos foi utilizado o erro quadrático médio (MSE) que é uma medida de precisão. Verificou-se através dessa medida que os valores de MSE para a krigagem simples e krigagem universal foram os menores e não foram diferentes estatisticamente. Sendo assim recomenda-se o uso da krigagem simples por mais precisa que a krigagem universal.

Palavras chave: Geoestatística, krigagem, precisão.

ABSTRACT

The objective of this chapter is to compare three geostatistical interpolators and verify which one is most accurate for obtaining Digital Depth Models (MDP). These models are obtained through bathymetric surveys that provide values of the represented depth in spatially continuous form. For this study were chosen the methods of simple kriging, universal kriging and kriging regression. For the comparison between these methods the mean square error (MSE) was used which is a measure of precision. It was verified through this measure that the MSE values for simple kriging and universal kriging were the smallest and were not statistically different. Therefore, it is recommended to use simple kriging as more precise than universal kriging.

Key words: Geostatistics, kriging, precision.

1.INTRODUÇÃO

Em diversos estudos, é de fundamental importância a representação do relevo submerso. As profundidades dos corpos aquáticos, obtidas através de levantamentos batimétricos, são indispensáveis para a representação destes através de Modelos Digitais de Profundidade (MDP) (FERREIRA et al., 2013).

A batimetria é a medição da profundidade de mares, lagos ou rios. A partir dessa medição é possível fazer a descrição topográfica de um curso abaixo do nível d'água (HIDROTEL, 2002).

De acordo com Álvares et al. (2001) os levantamentos batimétricos automatizados tem uma importante relevância na modelagem e gestão de recursos hídricos, pois o grau de assoreamento, cálculo de volumes de armazenamento, modelagem do relevo submerso, entre outros, podem ser estimados. O que é levado em conta na tomada de decisões para gerenciamento e utilização do recurso hídrico.

Segundo Camargo (1998) o produto obtido dos levantamentos batimétricos é uma malha de pontos amostrais. Para se obter o MDP, que represente a profundidade de forma espacialmente contínua, é preciso que se utilize interpoladores para estimar o valor da profundidade em locais não amostrados (FERREIRA et al., 2013).

Existem vários métodos de interpolação baseados na estatística clássica. Estes supõem que as amostras sejam independentes e sejam distribuídas espacialmente de forma aleatória (VIEIRA, 2000) desconsiderando a anisotropia e a continuidade do fenômeno observado (CAMARGO, 1998).

Dessa forma, de acordo com Ferreira et al. (2013) a Geoestatística (krigagem) tem sido amplamente aplicada no estudo da variabilidade espacial de fenômenos naturais.

Segundo Carmo et al., (2015) as características de autocorrelação das variáveis regionalizadas são levadas em consideração pela Geoestatística. Portanto, os dados amostrados são usados para estimar valores nos locais em que a variável não é conhecida (LANDIM, 2000).

De acordo com Vieira (1995), se é verificada a presença de tendência, esta deve ser removida e o semivariograma deve ser ajustado para os resíduos. Para tanto, no trabalho de Gonçalves et al. (2001) é citado o procedimento descrito por Davis (1986) para a remoção de tendência que consiste no seguinte: ajusta-se um polinômio de

primeiro ou segundo grau para os valores da variável, em função das coordenadas, e faz com que o resíduo seja obtido pela diferença entre o valor medido e o valor do polinômio em cada ponto. Já Journel (1969) e Matheron (1969) propuseram o método da krigagem universal para resolver problemas desse tipo.

Segundo Landim et al., (2002) “caso o modelo variográfico adotado não seja o correto os erros associados às estimativas por krigagem tornam-se maiores”.

A justificativa de se utilizar esse interpolador geoestatístico no trabalho é que de acordo com Ferreira et al., (2013) pelo fato dos levantamentos batimétricos apresentarem uma grande quantidade de pontos e pela variável estudada ser a profundidade, a batimetria apresenta características ideais para a aplicação do método de interpolação da Geoestatística (krigagem).

De acordo com o exposto acima, o presente trabalho tem por objetivo comparar três interpoladores geoestatísticos recomendados para a representação de Modelos Digitais de Profundidade. Entre estes interpoladores estão a krigagem universal, a krigagem-regressão, e a krigagem simples. Como indicadores comparativos de qualidade, serão adotadas as estatísticas: média e desvio padrão dos resíduos padronizados obtidos na validação cruzada, o erro quadrático médio e a variância de krigagem.

2. MATERIAL E MÉTODOS

2.1. Área de Estudo

O conjunto de dados utilizado neste estudo é de um levantamento batimétrico de um dos principais represamentos do Ribeirão São Bartolomeu, coletado em dezembro de 2010. Este represamento está localizado no campus da Universidade Federal de Viçosa (UFV), no município de Viçosa, em Minas Gerais. A área de estudo possui aproximadamente 8800 m², 150 m de comprimento e 66 m de largura.

Visando atingir o objetivo deste trabalho, serão realizadas análises geoestatísticas utilizando diferentes métodos de krigagem (krigagem universal, a krigagem - regressão e a krigagem simples) com a finalidade de verificar qual destes apresentará a melhor precisão.

De acordo com Hengl et al., (2007) existem na literatura geoestatística diferentes termos para técnicas muito parecidas. Estas seguem o modelo de krigagem universal (KU). Este modelo foi introduzido por Matheron (1969) e segundo Christensen (2001), para dados espaciais, este é o melhor preditor não viesado. Alguns autores como Deutsch e Journel (1998), Wackernagel (1998) e Papritz e Stein (1999) usam o termo krigagem universal quando a tendência pode ser modelada como uma função das coordenadas, sem um sistema de krigagem.

Porém de acordo com Wackernagel (1998) e Chiles and Delfiner (1999) quando a tendência for modelada externamente por uma função linear de algumas variáveis auxiliares, a krigagem é conhecida por krigagem com tendência externa (KED – kriging with External drift).

A krigagem universal também é conhecida como krigagem com tendência. De acordo com Andriotti (2013) esta é utilizada quando a média não for mais constante, ou seja, ocorre uma tendência nos dados, e quando o variograma ou a covariância não mais forem adequados para modelar a dependência espacial dos dados originais. Matheron (1969) definiu krigagem universal como sendo uma técnica que combina o modelo de regressão linear múltipla com o método de krigagem simples, usando as coordenadas geográficas como fatores externos. Neste caso a tendência é modelada em função das coordenadas geográficas (HENGL, 2009).

Segundos Santos (2011) a krigagem universal ocorre nos casos em que a variável regionalizada não seja estacionária. Assim, a remoção da tendência é feita no ajuste de polinômios de baixo grau, e então, o restante do procedimento analítico torna-se uma análise de resíduos.

O procedimento para o caso da tendência e dos resíduos serem estimados separados e somados é conhecido por krigagem-regressão (HENGL, 2007). Este procedimento foi sugerido por Ahmed e de Marsily (1987), porém o nome foi dado por Odeh et al., (1995). Já Goovaerts (1997) nomeou este procedimento como krigagem simples com médias locais variando.

Segundo Hengl (2009) a krigagem regressão é uma técnica de interpolação híbrida, ou seja, é a combinação de duas abordagens conceituais diferentes. Essa técnica é utilizada para modelar e mapear variabilidade espacial. De acordo com Odeh et al., (1995), Bishop e McBratney (2001), Hengl et al., (2007), Zhu e Lin (2010). A técnica de Krigagem-Regressão combina um modelo linear de regressão simples ou múltipla com krigagem dos resíduos da regressão.

Hengl (2009) tem preferência pela krigagem-regressão a krigagem universal por esta explicar a estimação da tendência e dos resíduos interpolados separadamente. Além disso, permite que estes dois componentes sejam interpretados em separados. Alguns estudos (Hengl et al., 2007; Knotters et al., 1995; Bishop and McBratney, 2001; Bourennane and King, 2003; Lloyd, 2005; Yemefack et al., 2005) mostram que esse tipo de técnica pode gerar previsões melhores que uma abordagem única.

As técnicas de krigagem-regressão e krigagem com tendência externa geram previsões dos resultados e variância de previsões iguais para o mesmo conjunto de dados, variáveis auxiliares, e método de ajuste da regressão, porém estas se diferem pelas etapas computacionais (HENGL, 2007).

Para verificar a equivalência da krigagem-regressão e a krigagem com tendência externa Hengl (2007) mostra que a previsão da krigagem com tendência externa é feita como na krigagem ordinária de acordo com a equação (2),

$$\hat{\mathbf{z}}_{\text{KED}}(\mathbf{s}_0) = \boldsymbol{\lambda}_{\text{KED}}^T \cdot \mathbf{z} \quad (2)$$

Segundo Wackernagel (1998) para encontrar os pesos da krigagem (λ_{KED}^T) deve-se resolver o seguinte sistema matricial (equação 3):

$$\begin{bmatrix} \mathbf{C} & \mathbf{q} \\ \mathbf{q}^T & \mathbf{0} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\lambda}_{KED} \\ \boldsymbol{\varphi} \end{bmatrix} = \begin{bmatrix} \mathbf{c}_0 \\ \mathbf{q}_0 \end{bmatrix}, \quad (3)$$

em que $\boldsymbol{\varphi}$ é o vetor multiplicador de Lagrange, \mathbf{C} é a matriz de covariância dos resíduos, \mathbf{q} é a matriz dos preditores dos locais amostrados, \mathbf{q}_0 é o vetor dos $p + 1$ preditores e \mathbf{c}_0 é o vetor das covariâncias dos resíduos nos locais não amostrados.

Resolvendo essa multiplicação matricial tem-se,

$$\begin{cases} \mathbf{C} \cdot \boldsymbol{\lambda}_{KED} + \mathbf{q} \cdot \boldsymbol{\varphi} = \mathbf{c}_0 \\ \mathbf{q}^T \cdot \boldsymbol{\lambda}_{KED} = \mathbf{q}_0 \end{cases} \quad (4)$$

Assim, isolando $\boldsymbol{\lambda}_{KED}$ na primeira linha do sistema de equações tem-se:

$$\boldsymbol{\lambda}_{KED} = \mathbf{c}_0 \cdot \mathbf{C}^{-1} - \mathbf{C}^{-1} \cdot \mathbf{q} \cdot \boldsymbol{\varphi} \quad (5)$$

Substituindo na segunda equação obtém-se:

$$\mathbf{q}^T \cdot \boldsymbol{\lambda}_{KED} = \mathbf{q}^T \cdot \mathbf{c}_0 \cdot \mathbf{C}^{-1} - \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q} \cdot \boldsymbol{\varphi} \quad (6)$$

Isolando $\boldsymbol{\varphi}$ tem-se:

$$\boldsymbol{\varphi} = (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0 - (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \boldsymbol{\lambda}_{KED} \quad (7)$$

Substituindo $\mathbf{q}^T \cdot \boldsymbol{\lambda}_{KED}$ por \mathbf{q}_0 encontra-se:

$$\boldsymbol{\varphi} = (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0 - (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}_0 \quad (8)$$

Logo, substituindo $\boldsymbol{\varphi}$ na equação 4 obtém-se:

$$\mathbf{C} \cdot \boldsymbol{\lambda}_{KED} + \mathbf{q} \cdot [(\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0 - (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}_0] = \mathbf{c}_0 \quad (9)$$

Isolando $\boldsymbol{\lambda}_{KED}$ tem-se:

$$\begin{aligned} \boldsymbol{\lambda}_{KED} &= \mathbf{C}^{-1} \cdot \mathbf{c}_0 - \mathbf{C}^{-1} \cdot \mathbf{q} \cdot [(\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0 - (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}_0] \\ \boldsymbol{\lambda}_{KED} &= \mathbf{C}^{-1} \cdot [\mathbf{c}_0 + \mathbf{q} \cdot (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot (\mathbf{q}_0 - \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0)] \end{aligned} \quad (10)$$

De acordo com Papritz e Stein (1999) essa equação mostra que os pesos da krigagem com tendência externa são iguais.

Passando para a Krigagem Regressão, para determinar o vetor dos coeficientes estimados da regressão pelo método dos mínimos quadrados generalizados parte-se da seguinte equação:

$$\hat{\boldsymbol{\beta}}_{GLS} = (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{z} \quad (11)$$

em que \mathbf{z} é o vetor dos valores medidos da variável resposta.

Os pesos para os resíduos é dado pela equação 12,

$$\boldsymbol{\lambda}_0^T = \mathbf{c}_0^T \cdot \mathbf{C}^{-1} \quad (12)$$

De acordo com Christensen (2001) uma vez que a tendência foi estimada, os resíduos podem ser interpolados com a krigagem e adicionados a tendência estimada. Essa soma é representada pela equação 13.

$$\hat{\mathbf{z}}(\mathbf{s}_0) = \mathbf{q}_0^T \cdot \hat{\boldsymbol{\beta}}_{GLS} + \boldsymbol{\lambda}_0^T \cdot \left(\mathbf{z} - \mathbf{q} \cdot \hat{\boldsymbol{\beta}}_{GLS} \right) \quad (13)$$

Substituindo as equações 10 e 11 na equação 12 obtém-se:

$$\hat{z}_{\text{RK}}(\mathbf{s}_0) = \begin{bmatrix} \mathbf{q}_0^T \cdot (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \mathbf{C}^{-1} + \mathbf{c}_0^T \cdot \mathbf{C}^{-1} \\ -\mathbf{c}_0^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q} \cdot (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T \cdot \mathbf{C}^{-1} \end{bmatrix} \cdot \mathbf{z}$$

$$\hat{z}_{\text{RK}}(\mathbf{s}_0) = \mathbf{C}^{-1} \cdot \left[\mathbf{c}_0^T + \mathbf{q}_0^T \cdot (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot \mathbf{q}^T - \mathbf{c}_0^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q} \cdot (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \right] \cdot \mathbf{z}$$

$$\hat{z}_{\text{RK}}(\mathbf{s}_0) = \mathbf{C}^{-1} \cdot \left[\mathbf{c}_0 + \mathbf{q} \cdot (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot (\mathbf{q}_0 - \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0) \right] \cdot \mathbf{z} \quad (14)$$

Substituindo a equação 10 na equação 2 tem-se:

$$\hat{z}_{\text{KED}}(\mathbf{s}_0) = \mathbf{C}^{-1} \cdot \left[\mathbf{c}_0 + \mathbf{q} \cdot (\mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{q})^{-1} \cdot (\mathbf{q}_0 - \mathbf{q}^T \cdot \mathbf{C}^{-1} \cdot \mathbf{c}_0) \right] \cdot \mathbf{z} \quad (15)$$

Observa-se que as equações 14 e 15 são iguais. Portanto, é provado dessa forma que a krigagem com tendência externa (KED) gera as mesmas previsões que a Krigagem Regressão quando são usados os mesmos conjuntos de dados.

Porém deve-se tomar alguns cuidados caso a krigagem escolhida seja a krigagem-regressão, pois a análise estatística desta segundo Hengl et al., (2009), é mais sofisticada que outras técnicas de krigagem e por isso demanda de computadores mais modernos para seu uso. Outro ponto levantado por Hengl et al., (2009) é que muitos usuários se confundem em qual metodologia escolher por existirem muitas opções.

Segundo Cressie (1993) a krigagem universal é equivalente à regressão linear quando executa-se o estimador de quadrados mínimos generalizados mais a krigagem simples de resíduos.

O último método de interpolação a ser estudado neste trabalho é a Krigagem Simples. Segundo Cressie, (1993) e Rao e Toutenburg (1999) este é o caso mais simples de previsão geoestatística que pode ser introduzido como preditor heterogeneamente linear (Ligas e Kulczycki, 2010). Nesta, a média da variável é considerada conhecida e o peso da média fornece um critério que permite testar a qualidade da krigagem (ANDRIOTTI, 2002).

De acordo com Santos et al., (2011) ao se comparar a variância de krigagem da krigagem simples com a krigagem universal a primeira apresenta menor valor em relação a segunda quando se tem um vetor de médias conhecido.

Todas as análises estatísticas foram realizadas no software estatístico R (R Core Team, 2016.) utilizando os pacotes geoR (RIBEIRO JR e DIGGLE, 2016) e gstat (PEBESMA, 2004).

Para fazer a comparação entre os métodos de krigagem foram calculados a média, o desvio-padrão dos resíduos obtidos pela autovalidação e a média dos erros ao quadrado (MSE). O MSE mede a acurância da predição e foi obtido através da equação (16):

$$MSE = \frac{1}{N} \sum_{j=1}^n \left[\hat{Z}(s_j) - Z^*(s_j) \right]^2 \quad (16)$$

em que N é o número de pontos da autovalidação.

O MSE foi utilizado para fazer a comparação entre os métodos, pois sua distribuição é conhecida. Como o erro está elevado ao quadrado então o MSE segue uma distribuição qui-quadrada.

3. RESULTADOS E DISCUSSÕES

A Tabela 1 mostra os resultados da análise exploratória dos dados batimétricos.

Tabela 1 – Estatística descritiva da batimetria do represamento do Rio São Bartolomeu.

Média	Variância	Desvio-Padrão	Assimetria	CV	Mín	Máx
-4,1813	1,1209	1,06	2,09	25,32	-5,540	0.000

Pela tabela 1 verifica-se que o coeficiente de assimetria obtido nas análises indica que esta distribuição é assimétrica à direita, de acordo com a classificação de Ferreira (2009).

Construiu-se o histograma dos dados amostrados para melhor visualização das informações da Tabela 1, de acordo com a Figura 1.

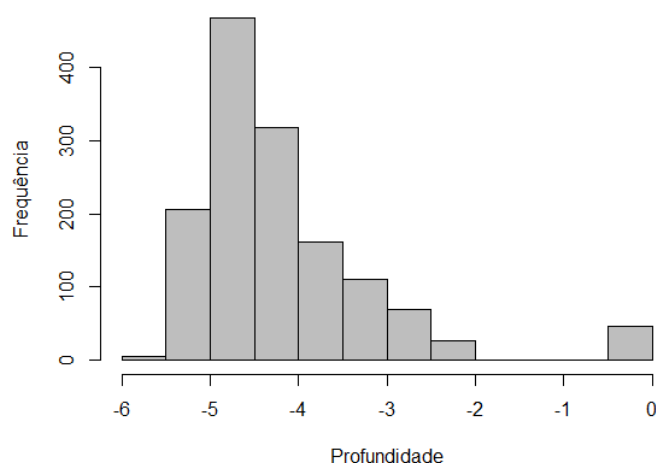
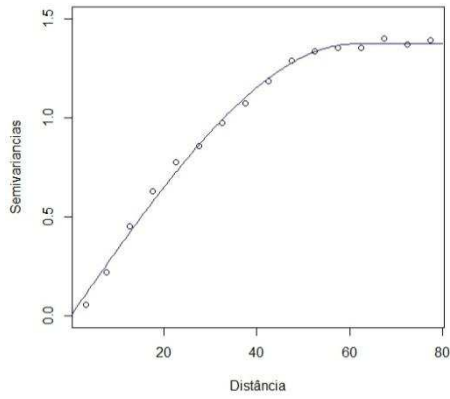


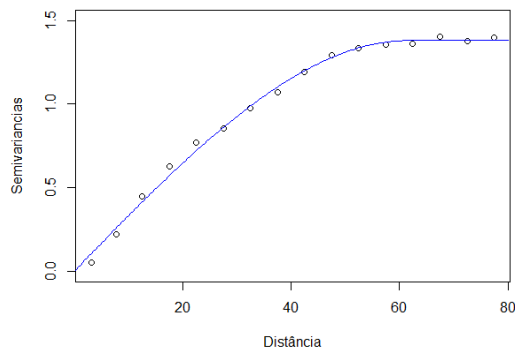
Figura 1 – Distribuição de frequências da variável batimetria do represamento do Rio São Bartolomeu.

Após o estudo descritivo foi realizada uma análise sobre as características espaciais dos dados, afim de, se realizar uma correta interpolação pelo método da krigagem.

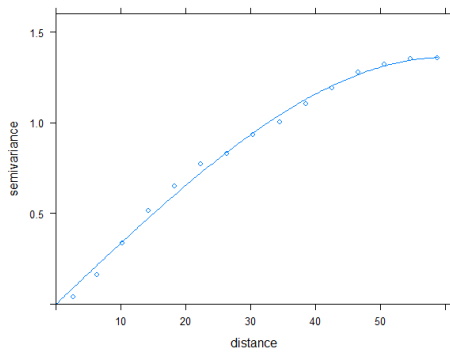
Foi observado através dos semivariogramas a comprovação da dependência espacial da variável batimetria. Na figura 2 são apresentados os semivariogramas obtidos para os três métodos de krigagem.



(a)



(b)



(c)

Figura 2 – Semivariograma ajustado e teórico da krigagem simples (a), krigagem universal (b) e krigagem regressão (c) respectivamente do levantamento batimétrico do represamento do Rio São Bartolomeu.

Para a três krigagens (simples, universal e regressão) o modelo ajustado foi o esférico.

Os resultados da autovalidação descrevem a precisão do modelo ajustado. Os principais estimadores da autovalidação com as respectivas estatísticas que são recomendadas por Vieira (2000), Santos et al., (2011) e Ferreira et al., (2013), são apresentados na Tabela 2.

Tabela 2- Comparação entre os métodos de krigagem pela média e variância dos resíduos padronizados obtidos pela autovalidação.

Krigagem	Média	Variância
Simples	0.0154	0.0361
Regressão	0.0219	0.0655
Universal	0.0137	0.0296

Segundo Vieira (2000) a média dos erros da autovalidação deve ser bem próxima de zero e a variância bem próxima de 1.

De acordo com Mood et al., (1974) o erro quadrático médio pode seguir uma distribuição qui-quadrado. Portanto, é indicada a realização de um teste de hipótese para verificar se há diferença significativa entre os métodos de krigagem a um nível de 5%.

Na Tabela 3 estão apresentados os valores do erro quadrático médio para os três diferentes métodos de krigagem. Observa-se que não houve diferença significativa a um nível arbitrário de 5% entre a krigagem simples e a krigagem universal. Porém a krigagem-regressão foi significativa a este mesmo nível (5%).

Tabela 3– Comparação entre os métodos de krigagem através do MSE.

Krigagem	MSE
Simples	0.0363
Regressão	0.0659*
Universal	0.0298

*Estatisticamente significativa a 5% pelo teste de qui-quadrado com (n-1) G.L.

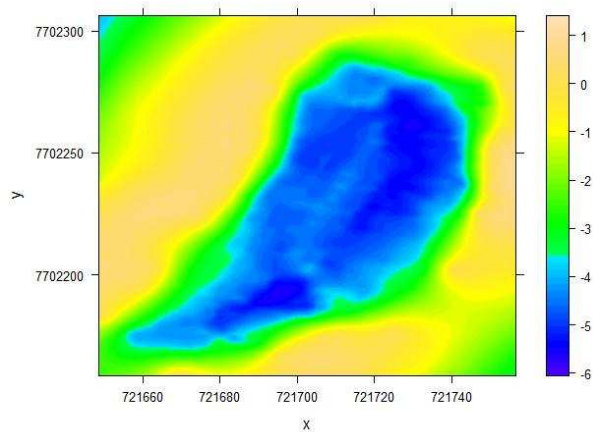
A krigagem-regressão pode ter apresentado o pior resultado dentre os três métodos, pois neste estudo as coordenadas foram utilizadas como variáveis auxiliares e, segundo Hengl (2009), esta é a melhor técnica de predição espacial para interpolar variáveis ambientais com a utilização de variáveis auxiliares, diferentes das coordenadas.

A diferença entre a krigagem-regressão e a krigagem universal era esperada, pois foi executado o estimador de mínimos quadrados ordinários mais krigagem

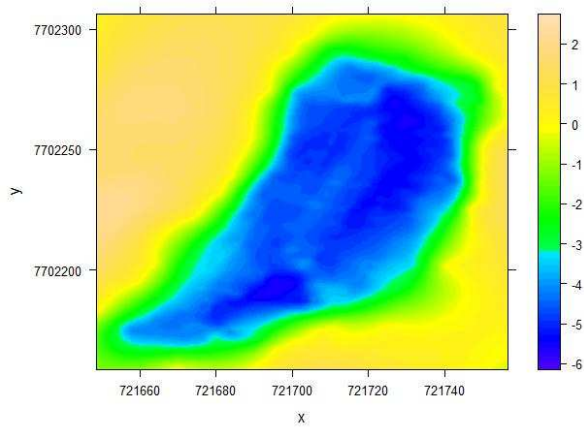
ordinária. Isso vem confirmar o que Cressie (1993) afirmou, que estas krigagens seriam iguais se o estimador de mínimos quadrados generalizados mais a krigagem simples fossem executados.

Como não houve diferença significativa entre a krigagem simples e a krigagem universal então a escolha deve ser feita por parcimônia. Segundo Santos et al., (2011) a krigagem simples apresenta menor variância de krigagem que a krigagem universal, então indica-se que a primeira seja utilizada, pois não é preciso fazer uso de covariáveis e, além disso, não há problemas de inversibilidade da matriz.

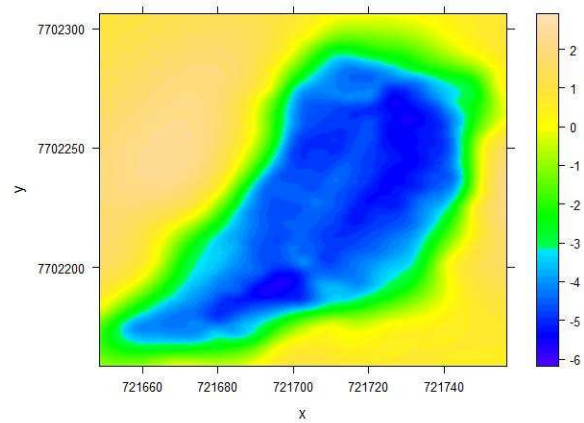
Ao realizar o teste de qui-quadrado, percebe-se que não há diferença significativa entre os valores. Isso pode ser verificado ao se fazer a comparação visual entre os mapas obtidos pelas krigagem simples, krigagem universal e krigagem regressão de acordo com a Figura 3.



(a)



(b)



(c)

Figura 3 – Mapas gerados pela krigagem simples (a), krigagem universal (b) e krigagem regressão (c), respectivamente.

4. CONCLUSÃO

Com este estudo pode-se verificar que por meio da medida de precisão (MSE) utilizada para comparação entre os métodos de krigagem, que não houve diferença significativa entre a krigagem simples e a krigagem universal, sendo assim sugere-se que se adote o método da krigagem simples pois, de acordo com a literatura esta é mais precisa que a krigagem universal.

Apesar de o estudo mostrar a krigagem-regressão como o pior método entre os três estudados ela tem sua importância. Porém quando se tem dados de batimetria, como é o caso do trabalho é melhor que se faça o simples.

REFERÊNCIAS BIBLIOGRÁFICAS

Ahmed, S., de Marsily, G., 1987. **Comparison of geostatistical methods for estimating transmissivity using data on transmissivity and specific capacity.** Water Resources Research n° 23, v.9, p.1717–1737. 1987.

ÁLVARES, M. T.et al. **Monitorização Batimétrica em Albufeiras.** Lisboa: Instituto da Água, Direção de Serviços de Recursos Hídricos. 2001.

ANDRIOTTI, J.L.S. **Fundamentos de Estatística e Geoestatística.** Editora Unisino: São Leopoldo. 2013.

ANDRIOTTI, J.L.S. Notas de Geoestatística. **Acta Geologica Leopoldensia.** XXV v.55, p. 3-14. 2002.

BISHOP, T.; MCBRATNEY, A. A comparison of prediction methods for the creation of field-extent soil property maps. **Geoderma**, v.103, p.149-160. 2001.

BOHLING, G. C&PE 940. Disponível em: <http://people.ku.edu/~gbohling/cpe940/Kriging.pdf>. Acessado em: 16-09-2016.

CAMARGO, E.C.G. **Geoestatística: Fundamento e Aplicações.** In: Camara, G.; Medeiros, J. S. Geoprocessamento em Projetos Ambientais. 2ª ed. São José dos Campos: INPE, 1998.

BOURENNANE, H.; KING, D. Using multiple external drifts to estimate a soil variable. **Geoderma**, v.114, p.1-18. 2003.

CARMO, E.J., RODRIGUES, D.D., SANTOS, G.R.dos. **Boletim de Ciências Geodésicas**, sec. Artigos, Curitiba, v. 21, no 4, p. 674-690. 2015.

CRESSIE, N.A.C. **Statistics for Spatial Data**, (Revised Edition) Wiley: New York. 1993.

CHRISTENSEN, R., 2001. Linear Models for Multivariate TimeSeries and Spatial Data, second ed. Springer, New York, 398pp.

DAVIS, J.C. **Statistics and data analysis in geology.** 2ªed. New York: John Wiley, 1986.

DEUTSCH, C., JOURNEL, A., 1998. GSLIB: Geostatistical Software and User's Guide, second ed. Oxford University Press, New York, 369pp.

EMÍDIO, Z. P. O; LANDIM, P. M. B. Análise de superfície de tendência aplicada à chuva, medida por radar meteorológico, nas regiões de assis e piracicaba, SP. **Geociências**, v.27, n.4, p.439 – 449. 2008.

FERREIRA, I.O.; SANTOS, G.R.; RODRIGUES, D.D. Estudo sobre a utilização adequada da krigagem na representação computacional de superfícies batimétricas. **Revista Brasileira de Cartografia** (Online), v.65, n° 5, p.831-842. 2013.

FERREIRA, D.F. **Estatística básica**. 2ª ed. Lavras: UFLA. 2009.

GOMES, O. M.; SANTOS, C. A. C.dos, SOUZA, F. de A. S.de; PAIVA, W.de; OLINDA, R. A. de. Análise comparativa da precipitação no estado da Paraíba utilizando modelos de regressão polinomial. *Revista Brasileira de Meteorologia*. v.30, n° 1, p. 47-58. 2015.

GONÇALVES, A.C.A.; FOLEGATTI, M.V.; MATA, J.D.V. Análises exploratória e geoestatística da variabilidade de propriedades físicas de um argissolo vermelho. **Acta Scientiarum**, Maringá, v.23, n.5, p.1149-1157. 2001.

Goovaerts, P. **Geostatistics for Natural Resources Evaluation**. Oxford University Press, New York, p.483. 1997.

HENGL, T.; HEUVELINK, G.B.M.; ROSSITER, D.G. About regression-kriging: From equations to case studies. **Computers & Geosciences**, v.33, 1301-1315. 2007.

HENGL, T. **A Practical Guide to Geostatistics Mapping**. 2009.

HIDROTEL. “Batimetria e Eco-Batimetria”. Disponível em: www.hidroTEL.com.br/bat_ecobat_topo. 2002.

JOURNAL, A.G. - **Etude sur l'Estimation d'une Variable Regionalisée. Application à la Cartographie Sous-Marine**: Service Central Hydrographique de la Marine, Paris. 1969.

KNOTTERS, M.; BRUS, D.; VOSHAAR, J. A comparison of kriging, co-kriging and kriging combined with regression for spatial interpolation of horizon depth with censored observations. **Geoderma**, v.67, p. 227-246. 1995.

LANDIM, P. M. B. Introdução aos métodos de estimação espacial para confecção de mapas. DGA, IGCE, UNESP/Rio Claro, Lab. **Geomatémática**, Texto Didático, 2, 2000.

LANDIM, P.M.B.; CORSI, A.C. **Cálculo de superfícies de tendência, por regressão polinomial**, pelo SURFER® 6: Texto didático 05. Rio Claro: UNESP, 11 p., 2001. Disponível em: <http://www.rc.unesp.br/igce/aplicada/DIDATICOS/LANDIM/trendsurfer.pdf>. Acesso em: 11fev2016.

LANDIM, P.M.B.; STURARO, J.R.; MONTEIRO, R.C. Krigagem ordinária para situações com tendência regionalizada. DGA,IGCE,UNESP/Rio Claro, Lab. **Geomatémática**, Texto Didático 06. 2002.

LIGAS, M., KULCZYCKI, M. Simple spatial prediction – least squares prediction, simple kriging, and conditional expectation of normal vector. **Geodesy and Cartography**. vol. 59, n° 2, 69-81. 2010.

LLOYD, C.D. Assessing the effect of integrating elevation data into the estimation of monthly precipitation in Great Britain. **Journal of Hydrology**. v.308, p.128-150. 2005.

MATHERON, G. **Le krigeage univesel. Cahiers du Centre de Morphologie Mathematique**, Ecole des Mines de Paris, Fontainebleau. vol. 1. 1969.

MOOD, A. M.; GRAYBILL, F. A.; BOES, D.C. **Introduction to the theory of statistics**. New York - US: McGraw-Hill. 1974.

ODEH, I.O.A., MCBRATNEY, A.B., CHITTLEBOROUGH, D.J., Further results on prediction of soil properties from terrain attributes: heterotopic cokriging and regression-kriging. **Geoderma**. v.67, p. 215-226. 1995.

OLIVEIRA, J.R.B.; SILVA, R.R. **Estudos Batimétricos dos lagos dos principais parques da cidade de Curitiba – Paraná**. Monografia (Graduação em Engenharia Civil). 2002.

Papritz, A., Stein, A. Spatial prediction by linear kriging. In: Stein, A., van der Meer, F., Gorte, B. (Eds.), **Spatial Statistics for Remote Sensing**. Kluwer Academic Publishers, Dodrecht, p. 83–113, 1999.

PEBESMA, E.J. Multivariable geostatistics in S: the gstat package. *Computers & Geosciences*, 30: 683-691. 2004.

R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2016.

RAO, C.R.; Toutenburg, H. **Linear Models: Least Squares and Alternatives**. New York: Springer – Verlag. 1999.

RIBEIRO JR., P.J.; DIGGLE, P.J. geoR: Analysis of Geostatistical Data. R package version 1.7-5.2. 2016.<https://CRAN.R-project.org/package=geoR>.

SANTOS, G.R. dos; OLIVEIRA, M.S. de; LOUZADA, J. M.; SANTOS, A.M.R.T. Krigagem Simples versus Krigagem universal: qual o preditor mais preciso? *Energia na Agricultura*, v.26, p. 49-55. 2011.

VIEIRA, S.R. **Uso da geoestatística em estudos de variabilidade espacial**. In: Curso de atualização em conservação do solo. Campinas: Instituto Agrônômico de Campinas, 1995.

VIEIRA, S. R. Variabilidade espacial de argila, silte e atributos químicos em parcela experimental de um Latossolo Roxo de Campinas (SP). **Bragantia**, v.56, nº 1, p.181-190. 1997.

VIEIRA, M.B. **Vetorização e análise de tendência de cartas de lineamentos geológicos**. Belo Horizonte. Dissertação (Mestrado em Ciências da Computação) – Universidade Federal de Minas Gerais. 1998.

VIEIRA, S.R. Geoestatística em estudos de variabilidade espacial do solo. In. NOVAES, R.F.; ALVAREZ V., V.H.; SCHAEFER, C.E.G.R. Tópicos em ciências do solo. Viçosa, MG: **Sociedade Brasileira de Ciência do Solo**. v.1. p.2-54. 2000.

Wackernagel, H., 1998. Multivariate Geostatistics: An Introduction with Applications, second ed. Springer, Berlin, 291pp.

YEMEFACK, M., ROSSITER, D.G., NJOMGANG, R. Multi-scale characterization of soil variability within an agricultural landscape mosaic system in southern Cameroon. **Geoderma**. v.125, p. 117-143. 2005.

ZHU, Q.; LIN, H.S. **Comparing ordinary kriging and regression kriging for soil properties in contrasting landscapes**. *Pedosphere*, v.20, p.594-606, 2010.

CONCLUSÕES GERAIS

Com a realização desta pesquisa pôde-se verificar, através dos resultados obtidos nos experimentos realizados, que:

- existe uma hierarquia entre as malhas amostrais regulares estudadas (hexagonal, quadrangular e triangular), sendo que a malha hexagonal é mais precisa em relação a malha quadrangular seguida pela malha triangular ao se adotar mesma área para os polígonos estudados. Porém ao fazer o estudo com áreas diferentes para os polígonos foi observado que não há diferença significativa estatisticamente entre as malhas. Ainda foi possível com esse estudo quantificar o aumento da imprecisão das malhas amostrais quando o número de pontos destas era diminuído.
- em conjuntos de dados que não atendam a normalidade é possível realizar uma análise Geoestatística atendendo as pressuposições da mesma.
- não é necessário que se faça transformação em um conjunto de dados assimétricos para alcançar a simetria para realizar uma análise Geoestatística, pois não houve melhoria nos indicadores de qualidade de ajuste. Sendo assim, sugere-se trabalhar com os dados originais.
- ao comparar três interpoladores geoestatísticos (krigagem simples, krigagem-regressão e krigagem universal) não há diferença entre a krigagem simples e a universal, sugerindo-se utilizar a krigagem simples pois como foi mostrado no trabalho esta é mais precisa que a krigagem universal.

Com base nas considerações acima, conclui-se que este trabalho atingiu os objetivos propostos, gerando resultados significativos para a realização de uma análise Geoestatística.