

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ  
*CAMPUS* DE CASCAVEL  
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

**TAMANHO AMOSTRAL EFETIVO NO ESTUDO DA VARIABILIDADE ESPACIAL DE VARIÁVEIS  
GEORREFERENCIADAS USANDO AS DISTRIBUIÇÕES NORMAL E T-STUDENT**

LETÍCIA ELLEN DAL' CANTON

CASCAVEL-PARANÁ  
2018

Dados Internacionais de Catalogação-na-Publicação (CIP)  
(Sistema de Bibliotecas - UNIOESTE)

D138t Dal' Canton, Letícia Ellen.  
Tamanho amostral efetivo no estudo da variabilidade espacial de variáveis georreferenciadas usando as distribuições normal e t-Student / Letícia Ellen Dal' Canton. --- Cascavel (PR), 2018. xi, 95 f.: il.

Orientador: Dra. Luciana Pagliosa Carvalho Guedes.  
Co-orientadora: Dra. Rosângela Aparecida Botinha Assumpção.  
Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual do Oeste do Paraná, Campus de Cascavel, 2018, Programa de Pós-Graduação em Engenharia Agrícola.  
Inclui Bibliografia

1. Agricultura de precisão. 2. Amostragem do solo. 3. Dependência espacial. 4. Distribuição t-Student. 5. Geoestatística. I. Guedes, Luciana Pagliosa Carvalho. II. Assumpção, Rosângela Aparecida Botinha. III. Universidade Estadual do Oeste do Paraná. IV. Título.

CDD 631.4

Rosângela A. A. Silva – CRB 9ª/1810



**unioeste**

Universidade Estadual do Oeste do Paraná

Campus de Cascavel CNPJ 78680337/0002-65  
Rua Universitária, 2069 - Jardim Universitário - Cx. P. 000711 - CEP 85819-110  
Fone:(45) 3220-3000 - Fax:(45) 3324-4566 - Cascavel - Paraná



**PARANÁ**  
GOVERNO DO ESTADO

## LETÍCIA ELLEN DAL' CANTON

Tamanho amostral efetivo no estudo da variabilidade espacial de variáveis georreferenciadas usando as distribuições normal e t-Student

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestra em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão, APROVADO(A) pela seguinte banca examinadora:

Orientador(a) - Luciana Pagliosa Carvalho Guedes

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)

Manuel Jesus Galea Rojas

Pontifícia Universidade Católica do Chile (PUC-Chile)

Miguel Angel Uribe Opazo

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)

Cascavel, 6 de fevereiro de 2018

LETÍCIA ELLEN DAL' CANTON

**TAMANHO AMOSTRAL EFETIVO NO ESTUDO DA VARIABILIDADE  
ESPACIAL DE VARIÁVEIS GEORREFERENCIADAS USANDO AS  
DISTRIBUIÇÕES NORMAL E T-STUDENT**

Dissertação apresentada ao Programa de Pós-graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área concentração Sistemas Biológicos e Agroindustriais.

Orientador: Dra. Luciana Pagliosa Carvalho Guedes  
Coorientador: Dra. Rosângela Aparecida Botinha Assumpção

CASCADEL-PARANÁ  
2018

## **BIOGRAFIA**

Letícia Ellen Dal' Canton, nascida em Corbélia-PR no dia 18 de Janeiro de 1993, graduada em Matemática pela Universidade Estadual do Oeste do Paraná (UNIOESTE) no ano de 2014. Em março de 2016, ingressou por essa mesma instituição de ensino no mestrado do Programa de Pós-Graduação Stricto Sensu em Engenharia Agrícola, com área de concentração em Sistemas Biológicos e Agroindustriais e linha de pesquisa em Estatística Espacial.

## AGRADECIMENTOS

A Deus, por ter iluminado meu caminho e me dado forças para prosseguir; sem Ele eu não teria conseguido superar os obstáculos e os períodos de angústias.

Aos meus pais, por sempre terem priorizado meus estudos e por acreditarem em mim a todo momento. Obrigada pelo apoio e carinho, amo vocês!

À Luana, por ter ouvido meus desabaços e entendido cada momento de estresse, ansiedade e nervosismo. Obrigada pelo seu companheirismo e amor. Te amo!

À minha orientadora, Dra. Luciana Pagliosa Carvalho Guedes, pelos ensinamentos repassados, além do apoio e confiança depositada.

À minha coorientadora, Dra. Rosângela Aparecida Botinha Assumpção, por ter aceitado o convite e pela disposição em contribuir com o trabalho.

Aos professores que compuseram a banca pelas contribuições dadas.

Aos meus amigos, pelas horas de estudo compartilhadas e pelos momentos de descontração. Obrigada por todas as conversas aleatórias e sem propósito, pois elas tornavam o dia mais divertido.

Ao Darlon, Fabrício e à Rosângela, que sempre estiveram dispostos a me ajudar com os problemas no LaTeX ou na rotina computacional.

Aos membros do projeto de extensão “Qual é sua pergunta?”, especialmente à professora Dra. Ana Tereza Bittencourt, pela oportunidade concedida e pelo apoio.

A esta Universidade e ao Programa de Pós-Graduação em Engenharia Agrícola - PGEAGRI, bem como ao seu corpo docente.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo suporte financeiro.

## RESUMO

A concorrência de mercado impõe ao agronegócio brasileiro que se produza mais a custos cada vez menores. Para tal, uma das alternativas é a utilização da Agricultura de Precisão (AP), que possibilita identificar a variabilidade espacial das propriedades físico-químicas do solo, de modo a conhecer melhor a área agrícola e, conseqüentemente, elevar o nível de produtividade das culturas. Independente do emprego da AP, é necessário conhecer a variabilidade espacial de uma variável na área agrícola e tal demanda exige planejamento amostral adequado que viabilize coletar o mínimo possível de pontos amostrais para evitar custos demasiados e manter a qualidade na amostragem. Um dos objetivos desse trabalho é reduzir o número de pontos amostrais coletados a partir do cálculo do tamanho amostral efetivo (ESS). Foi estimado o valor do ESS univariado e multivariado para variáveis georreferenciadas com distribuição normal de probabilidade utilizando-se duas metodologias: a de Griffith e a de Vallejos e Osorio. O estudo foi realizado com dados simulados, variando os valores do efeito pepita e alcance atribuídos às variáveis, e com atributos físico-químicos do solo. Nem sempre as variáveis têm distribuição normal de probabilidade, devido principalmente à presença de pontos discrepantes. Desta forma, estimou-se o valor do tamanho amostral efetivo univariado para processos estocásticos estacionários e isotrópicos, considerando-se que a estrutura de covariância apresentava distribuição de probabilidade t-Student. Diante dos resultados multivariados provenientes de variáveis com distribuição de probabilidade normal, constatou-se uma redução no número de pontos amostrais que variou entre 48% e 93%. Tanto no caso uni quanto multivariado, o valor estimado do ESS foi menor pelo método de Griffith, indicando que essa proposta viabiliza maior redução no tamanho amostral. Os resultados univariados derivados dos atributos com distribuição t-Student mostraram redução entre 40% e 95% no número de pontos amostrais. Tal variação na redução do tamanho amostral é justificada pelos diferentes valores dos parâmetros de dependência espacial apresentados pelas variáveis. Verificou-se ainda que o raio de dependência espacial foi o parâmetro que exerceu maior influência no valor estimado do ESS uni e multivariado, sendo que quanto maior seu valor, menor o tamanho amostral efetivo e, conseqüentemente, maior a redução no tamanho amostral.

**Palavras-chave:** agricultura de precisão, amostragem do solo, dependência espacial, geoestatística, redimensionamento amostral.

## ABSTRACT

Trading competition has demanded to the Brazilian agribusiness greater production at lower costs. Thus, the Precision Agriculture (AP) comes to light as an alternative, which can identify the spatial variability of physical and chemical soil properties, in order to better know the agricultural area and, consequently, raise crop yield standard. Regardless of the PA's use, knowing the spatial variability of a variable in the agricultural area requires adequate sampling planning that allows collecting as few sample points as possible, avoiding too many costs and maintaining quality in sampling. Regardless of the PA's management, it is required to know the spatial variability of a variable in an agricultural area and this also asks for an adequate sample planning that makes possible the collection of the least number of sampling points, in order to avoid too much cost and to keep quality in sampling. So, this trial aimed at reducing the number of sample points collected by calculating the effective sample size (ESS). The univariate and multivariate ESS value was estimated for georeferenced variables with normal probability distribution using two methodologies: Griffith and Vallejos and Osorio. This study was carried out with simulated data, varying the values of the nugget effect as well as the range attributed to the variables, and with physical-chemical attributes of a soil. Therefore, variables do not always have normal probability distribution, mainly due to the presence of discrepant points. Thus, the value of univariate effective sample size for stationary and isotropic stochastic processes was estimated, considering that the covariance structure had a t-Student probability distribution. According to the multivariate results from variables with normal probability distribution, there was a decrease in the number of sample points from 48% to 93%. In both univariate and multivariate cases, the estimated ESS was lower by the Griffith method, indicating that this suggestion can make feasible a larger decrease in sample size. Univariate results derived from attributes with Student's t-distribution showed a decrease from 40% to 95% in the number of sample points. Such variation in the sample size is justified by the different values of the spatial dependence parameters presented by the variables. It was also recorded that the radius of spatial dependence was the parameter with the greatest influence on the estimated value of uni and multivariate ESS, and the higher its value, because the smaller the effective sample size, the larger is the decrease in the sample size.

**Keywords:** precision agriculture, soil sampling, spatial dependence, geostatistics, sample resizing.

# Sumário

	<b>Lista de ilustrações</b>	<b>viii</b>
	<b>Lista de tabelas</b>	<b>x</b>
<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>2</b>	<b>OBJETIVOS</b>	<b>12</b>
<b>2.1</b>	<b>Objetivo geral</b>	<b>12</b>
<b>2.2</b>	<b>Objetivos específicos</b>	<b>12</b>
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>13</b>
<b>3.1</b>	<b>Distribuições de probabilidade da família de contornos elípticos</b>	<b>13</b>
3.1.1	Distribuição t-Student generalizada	14
<b>3.2</b>	<b>Agricultura de Precisão e sua aplicabilidade na Geoestatística</b>	<b>15</b>
<b>3.3</b>	<b>Geoestatística</b>	<b>15</b>
3.3.1	Origem da Geoestatística	15
3.3.2	Variável aleatória, regionalizada e processo estocástico	16
3.3.3	Estacionariedade do processo estocástico	16
3.3.4	Modelos espaciais lineares	18
3.3.5	Funções semivariância e covariância	19
<b>3.4</b>	<b>Gráfico semivariograma</b>	<b>21</b>
<b>3.5</b>	<b>Parâmetros de dependência espacial</b>	<b>24</b>
<b>3.6</b>	<b>Modelos teóricos</b>	<b>25</b>
3.6.1	Modelo exponencial	26
3.6.2	Modelo gaussiano	26
3.6.3	Modelo da família Matérn	27
<b>3.7</b>	<b>Modelos espaciais lineares com distribuição de probabilidade da família de contornos elípticos</b>	<b>28</b>
3.7.1	Função score	29
3.7.2	Matriz de informação observada	29
3.7.3	Matriz de informação de Fisher e erro padrão das estimativas dos parâmetros	30
<b>3.8</b>	<b>Métodos de estimação de parâmetros no ajuste de modelos à função semi-variância</b>	<b>31</b>
3.8.1	Método de Máxima Verossimilhança	31
<b>3.9</b>	<b>Critérios de seleção do melhor modelo teórico ajustado</b>	<b>32</b>
<b>3.10</b>	<b>Krigagem</b>	<b>34</b>
3.10.1	Krigagem ordinária	35
<b>3.11</b>	<b>Tamanho amostral efetivo</b>	<b>37</b>
3.11.1	Tamanho amostral efetivo univariado	37

3.11.2	Tamanho amostral efetivo multivariado . . . . .	42
<b>3.12</b>	<b>Referências . . . . .</b>	<b>44</b>
<b>4</b>	<b>APÊNDICE . . . . .</b>	<b>49</b>
<b>4.1</b>	<b>Apêndice A. Tamanho amostral efetivo bivariado . . . . .</b>	<b>49</b>
<b>5</b>	<b>ARTIGOS . . . . .</b>	<b>51</b>
<b>5.1</b>	<b>ARTIGO 1 . . . . .</b>	<b>51</b>
5.1.1	Introdução . . . . .	51
5.1.2	Material e Métodos . . . . .	53
5.1.2.1	Descrição das simulações . . . . .	53
5.1.2.2	Descrição dos dados experimentais . . . . .	54
5.1.2.3	Cálculo do tamanho amostral efetivo (ESS) . . . . .	55
5.1.3	Resultados e discussões . . . . .	57
5.1.3.1	Estudo de simulação . . . . .	57
5.1.3.2	Aplicação da metodologia à variáveis físico-químicas do solo . . . . .	60
5.1.4	Conclusões . . . . .	70
5.1.5	Referências . . . . .	70
<b>5.2</b>	<b>ARTIGO 2 . . . . .</b>	<b>74</b>
5.2.1	Introdução . . . . .	74
5.2.2	Material e Métodos . . . . .	76
5.2.2.1	Descrição das simulações . . . . .	76
5.2.2.2	Descrição dos dados experimentais . . . . .	77
5.2.2.3	Análises estatísticas e geoestatísticas e cálculo do tamanho amostral efetivo . . . . .	78
5.2.3	Resultados e discussões . . . . .	81
5.2.3.1	Estudo de simulação . . . . .	81
5.2.3.2	Aplicação da metodologia na resistência do solo à penetração . . . . .	82
5.2.4	Conclusões . . . . .	89
5.2.5	Referências . . . . .	90
<b>6</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>95</b>

# Lista de ilustrações

Figura 1 – Semivariograma experimental. . . . .	22
Figura 2 – Esquema para localização de pontos em grades irregulares para o cálculo da função semivariância. Fonte: Landim, 2007, p. 26, adaptada. . . . .	23
Figura 3 – Direções utilizadas para análise de tendência direcional. Fonte: De Bastiani, 2012, p. 10 adaptada. . . . .	24
Figura 4 – Mapa da área experimental com os pontos amostrados. . . . .	55
Figura 5 – Gráfico com a média e desvio padrão dos valores estimados do tamanho amostral efetivo, em todas as simulações, de acordo com as metodologias de Griffith (2005) ( $ESS_{GU}$ ) (A) e Vallejos e Osorio (2014) ( $ESS_{VOU}$ ) (B) para cada variável. . . . .	57
Figura 6 – Gráficos boxplot dos valores estimados do ESS multivariado nas 100 simulações em cada cenário, de acordo com as metodologias de Griffith (2005) ( $\widehat{ESS}_{GM}$ ) (A) e Vallejos e Osorio (2014) ( $\widehat{ESS}_{VOM}$ ) (B). O símbolo “+” representa o valor médio estimado do ESS multivariado em cada cenário. . . . .	60
Figura 7 – Gráfico post-plot dos atributos: (A) cálcio [ $cmolc\ dm^{-3}$ ], (B) carbono [ $g\ dm^{-3}$ ], (C) cobre [ $mg\ dm^{-3}$ ], (D) ferro [ $mg\ dm^{-3}$ ], (E) $H+Al^3$ [ $cmolc\ dm^{-3}$ ], (F) manganês [ $mg\ dm^{-3}$ ], (G) zinco [ $mg\ dm^{-3}$ ]; e resistência do solo à penetração (RSP) [MPa] nas camadas: (H) 0-10 cm, (I) 11-20 cm, (J) 21-30 cm, (K) 31-40 cm, (L) fósforo [ $mg\ dm^{-3}$ ], (M) magnésio [ $cmolc\ dm^{-3}$ ], (N) pH, (O) potássio [ $cmolc\ dm^{-3}$ ]. As circunferências destacam espacialmente os pontos discrepantes. Min: mínimo, Q1: 1º quartil, Q3: 3º quartil, Max: máximo. . . . .	62
Figura 8 – Mapa dos valores estimados dos atributos: (A) cálcio [ $cmolc\ dm^{-3}$ ], (B) carbono [ $g\ dm^{-3}$ ], (C) cobre [ $mg\ dm^{-3}$ ], (D) ferro [ $mg\ dm^{-3}$ ], (E) $H+Al^3$ [ $cmolc\ dm^{-3}$ ], (F) manganês [ $mg\ dm^{-3}$ ], (G) zinco [ $mg\ dm^{-3}$ ]; e resistência do solo à penetração (RSP) [MPa] nas camadas: (H) 0-10 cm, (I) 11-20 cm, (J) 21-30 cm, (K) 31-40 cm, (L) alumínio [ $cmolc\ dm^{-3}$ ], (M) fósforo [ $mg\ dm^{-3}$ ], (N) magnésio [ $cmolc\ dm^{-3}$ ], (O) pH, (P) potássio [ $cmolc\ dm^{-3}$ ]. . . . .	65
Figura 9 – Diagramas de dispersão dos valores estimados do tamanho amostral efetivo em função dos valores estimados de alcance prático e efeito pepita relativo (EPR) de acordo com (A) método de Griffith ( $\widehat{ESS}_{GU}$ ) e (B) método de Vallejos e Osorio ( $\widehat{ESS}_{VOU}$ ), para cada um dos seguintes atributos: C: carbono, Ca: cálcio, Cu: cobre, Fe: ferro, $H+Al^3$ : ácido potencial total, Mn: manganês, Zn: zinco, e 0-10, 11-20, 21-30, 31-40: resistência do solo à penetração nas respectivas profundidades (em cm). . . . .	67
Figura 10 – Área experimental com a localização dos pontos amostrados. . . . .	78
Figura 11 – Gráfico com a média e desvio padrão dos valores estimados do tamanho amostral efetivo ( $ESS_t$ ) para cada variável, considerando-se todas as simulações . . . . .	81
Figura 12 – Gráfico post-plot da resistência do solo à penetração (RSP) [MPa] nas quatro camadas de profundidade. As circunferências destacam espacialmente os pontos discrepantes. Min: mínimo, Q1: 1º quartil, Q3: 3º quartil, Max: máximo. . . . .	83

Figura 13 – Mapa dos valores estimados da resistência do solo à penetração das raízes (RSP) [MPa] nas quatro camadas de profundidade. . . . .	85
Figura 14 – Gráfico post-plot da resistência do solo à penetração (RSP) [MPa] nas quatro camadas de profundidade, considerando a configuração amostral reduzida. As circunferências destacam espacialmente os pontos discrepantes. Min: mínimo, Q1: 1º quartil, Q3: 3º quartil, Max: máximo. . . . .	87
Figura 15 – Mapa dos valores estimados da resistência do solo à penetração das raízes (RSP) [MPa] nas quatro camadas de profundidade, considerando a malha amostral original e reduzida com os mesmos intervalos de classe. Valores estimados de EG (Exatidão Global) e T (índice Tau). . . . .	88

# Lista de tabelas

Tabela 1 – Algumas distribuições pertencentes à classe de contornos elípticos, em que $c$ é a constante normalizadora. . . . .	14
Tabela 2 – Valores de $V_{g(u)}$ para algumas distribuições pertencentes à classe elíptica. . . . .	29
Tabela 3 – Expressões de $d_g$ e $f_g$ para as distribuições Normal e t-Student. . . . .	31
Tabela 4 – Valores dos parâmetros das variáveis que compõe cada cenário. . . . .	54
Tabela 5 – Estatísticas descritivas dos valores estimados do ESS multivariado nas 100 simulações em cada cenário para as metodologias de Griffith (2005) (G) e Vallejos e Osorio (2014) (VO). . . . .	59
Tabela 6 – Estatística descritiva dos atributos físico-químicos do solo. . . . .	61
Tabela 7 – Valores estimados dos parâmetros do melhor modelo geoestatístico ajustado para cada atributo. . . . .	64
Tabela 8 – Valores estimados do tamanho amostral efetivo de cada atributo segundo as metodologias de Griffith (2005) ( $\widehat{ESS}_{GU}$ ) e Vallejos e Osorio (2014) ( $\widehat{ESS}_{VOU}$ ). . . . .	66
Tabela 9 – Valores estimados do tamanho amostral efetivo multivariado pelos métodos de Griffith (2005) ( $\widehat{ESS}_{GM}$ ) e Vallejos e Osorio (2014) ( $\widehat{ESS}_{VOM}$ ). . . . .	68
Tabela 10 – Valores dos parâmetros que definem a estrutura de dependência espacial das variáveis simuladas, efeito pepita e função de alcance, além do raio e intensidade de dependência espacial de cada variável. . . . .	77
Tabela 11 – Estatística descritiva das quatro camadas de profundidade da resistência do solo à penetração (RSP). . . . .	83
Tabela 12 – Valores estimados dos parâmetros que definem a estrutura de dependência espacial em cada camada de profundidade da resistência do solo à penetração a partir dos melhores valores dos parâmetros de forma $\kappa$ e $\nu$ . . . . .	84
Tabela 13 – Valor estimado do tamanho amostral efetivo univariado em cada camada de profundidade da resistência do solo à penetração ( $\widehat{ESS}_t$ ). . . . .	85
Tabela 14 – Estatística descritiva das quatro camadas de profundidade da resistência dos solo à penetração (RSP), de acordo com a malha amostral reduzida. . . . .	86
Tabela 15 – Valores estimados dos parâmetros que definem a estrutura de dependência espacial em cada camada de profundidade da resistência do solo à penetração, com a malha amostral reduzida, de acordo com os melhores valores dos parâmetros de forma $\kappa$ e $\nu$ . . . . .	87

# 1 Introdução

Estudar determinado fenômeno espacial, o qual representa o conjunto de todos os valores possíveis de uma variável de interesse, ou mesmo comparar o comportamento de variáveis em parcelas da população, requer o estabelecimento de um planejamento amostral. E a primeira etapa deste planejamento é determinar o número de amostras. Isto porque não é acessível e nem financeiramente viável estudar toda a população, além do que não é necessário, desde que se tenha uma amostra representativa que viabilize realizar inferências em relação à população-alvo (FONTELLES *et al.*, 2009).

O planejamento amostral da pesquisa deve conter, além do número de amostras que serão coletadas, a técnica de amostragem que será adotada. Durante a realização de uma pesquisa estatística, é fundamental que o planejamento amostral seja acertadamente elaborado, pois possíveis problemas podem comprometer tanto a análise final dos dados, como a interpretação dos resultados (MIOT, 2011).

As técnicas geoestatísticas utilizam procedimentos estatísticos aplicados a problemas cujos dados são provenientes de fenômenos naturais espacialmente distribuídos e autocorrelacionados. Sendo assim, além do valor obtido para a variável de interesse, considera-se sua posição no espaço. O objetivo é caracterizar essa variável a partir do estudo de sua distribuição e variabilidade espacial, para determinação das incertezas associadas a esse estudo (LANDIM e YAMAMOTO, 2013).

Quando se deseja estudar dados agrícolas, é fundamental conhecer a dispersão e variabilidade espacial dos atributos físicos e químicos do solo, os quais influenciam diretamente o nível de produtividade das culturas (BERNARDI *et al.*, 2014b). E sabe-se que produzir mais e reduzir custos são necessidades impostas ao agronegócio brasileiro pela concorrência de mercado.

A redução dos custos da pesquisa demanda reanalisar o planejamento amostral e avaliar se o número inicial de amostras coletadas na área sob estudo poderia ser reduzido sem perdas consideráveis na qualidade das inferências realizadas. Deste modo, uma das metodologias a ser aplicada é o cálculo do tamanho amostral efetivo - ESS (em inglês, *Effective Sample Size*) (GRIFFITH, 2005). Essa proposta considera que a existência de elevada correlação espacial entre as amostras proporciona coleta de informações redundantes, de modo que o número de pontos amostrais poderia ser reduzido em coletas posteriores. Logo, para que se possa estimar o valor do ESS, são necessárias informações oriundas de uma coleta prévia de pontos amostrais realizada na área agrícola.

Outro tópico relevante para o processo de amostragem é a distribuição probabilística dos dados. Existem na literatura diversos estudos acerca do planejamento amostral, entretanto, a maior parte deles assume que os dados têm distribuição normal de probabilidade. E, geralmente, quando a suposição de normalidade não é satisfeita, adota-se como alternativa a transformação de variáveis (BOX e COX, 1964). Contudo, interpretar os resultados provenientes de dados transformados pode ser complicado. Além do que, a modelagem de dados transformados é altamente influenciada por observações atípicas (AZALLINI e CAPITANIO, 1999). Desta forma, torna-se importante o estudo de outras distribuições de probabilidade que sejam capazes de modelar os diferentes graus de assimetria e curtose apresentados por fenômenos espaciais cada vez mais complexos. Os modelos

esaciais lineares com distribuição da família de contornos elípticos, tal como a t-Student, vêm sendo recorrentemente utilizados para tal, pois estende a classe dos erros para outras distribuições de probabilidade além da normal e assim reduz-se a influência de pontos atípicos ao incorporar parâmetros que melhor acomodam estas observações. Além disso, as distribuições elípticas ainda preservam a estrutura de simetria da distribuição normal (CYSNEIROS, PAULA e GALEA, 2005); (DE BASTIANI *et al.*, 2015).

Posto isto, o objetivo geral deste trabalho foi obter uma redução no tamanho amostral utilizado no estudo da variabilidade espacial de variáveis georreferenciadas, considerando-se para tal atributos com distribuição de probabilidade normal (Artigo 1) e t-Student (Artigo 2).

## 2 Objetivos

### 2.1 Objetivo geral

Redefinir o tamanho amostral em estudos que envolvem a variabilidade espacial de variáveis georreferenciadas, de modo a reduzir o número de pontos amostrais considerando-se os dados simulados e os atributos físicos e químicos do solo, com distribuição de probabilidade normal e t-Student, com base na metodologia do tamanho amostral efetivo (ESS).

### 2.2 Objetivos específicos

- Avaliar a influência que os parâmetros de dependência espacial da estrutura de covariância exercem sobre os valores estimados do ESS univariado e multivariado;
- Identificar se a adição de variáveis proporciona aumento no valor estimado do ESS multivariado;
- Analisar se os valores estimados do ESS univariado e multivariado são similares entre os métodos de Griffith (2005) e Vallejos e Osorio (2014).

## 3 Revisão Bibliográfica

### 3.1 Distribuições de probabilidade da família de contornos elípticos

A suposição de normalidade é muito utilizada para a modelagem de conjuntos de dados cujas variáveis aleatórias são contínuas e simétricas. Quando esta suposição não é satisfeita, pode-se aplicar uma transformação ao conjunto de dados com o intuito de obter, pelo menos, a simetria. Embora transformar os dados para atingir a normalidade seja uma alternativa, tal como a proposta de Box e Cox (1964), esse tipo de modelagem é altamente influenciada por observações atípicas. Ademais, trabalhar com dados transformados dificulta a interpretação dos resultados, sendo preferível, portanto, utilizar os dados originais (AZALLINI e CAPITANIO, 1999).

A ausência de normalidade em alguns conjuntos de dados ocorre porque a distribuição de probabilidade é sensível a observações atípicas e erros cuja distribuição apresenta caudas mais pesadas que a normal. Sendo assim, torna-se interessante estudar outras distribuições de probabilidade que sejam capazes de modelar diferentes graus de assimetria e curtose. A família de distribuições de contornos elípticos vêm sendo recorrentemente utilizada para tal, tendo em vista que preservam a estrutura de simetria da distribuição normal, e podem ser úteis na construção de modelos quando a suposição de normalidade não é satisfeita (CYSNEIROS, PAULA e GALEA, 2005). Alguns exemplos de distribuições de contornos elípticos são t-Student, Slash e Exponencial Potência, dentre as quais, apenas a primeira será exposta no presente trabalho, além da normal.

Abaixo será apresentada a definição que caracteriza quando um vetor aleatório exibe uma distribuição da família de contornos elípticos (DE BASTIANI *et al.*, 2015):

**Definição 1** O vetor aleatório  $\mathbf{Z} \in \mathbb{R}^r$  ( $r \geq 2$ ) segue uma distribuição de contornos elípticos se sua função característica é dada pela seguinte expressão:

$$E[\exp(i\mathbf{t}^T \mathbf{Z})] = \exp(i\mathbf{t}^T \boldsymbol{\mu})\psi(\mathbf{t}^T \boldsymbol{\Sigma} \mathbf{t}),$$

em que  $E[\cdot]$  denota o valor esperado;  $i^2 = -1$ ;  $\mathbf{t} = (t_1, \dots, t_n)^T \in \mathbb{R}^n$ ;  $\boldsymbol{\mu}$ ,  $n \times 1$ , representa o vetor de parâmetros de posição;  $\boldsymbol{\Sigma}$ ,  $n \times n$ , denota a matriz escala, que é simétrica e semidefinida positiva; e  $\psi : \mathbb{R}^n \rightarrow \mathbb{R}$  é uma função geradora de características.

Se  $\mathbf{Z}$  tem distribuição de contornos elípticos cuja função característica apresenta-se conforme a definição acima, denota-se  $\mathbf{Z} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \psi)$ , indicando que  $\mathbf{Z}$  tem distribuição elíptica  $n$ -variada.

Não é regra que o vetor aleatório  $\mathbf{Z}$  deva possuir uma função densidade, mas, caso tenha, é dada pela seguinte expressão (DE BASTIANI *et al.*, 2015):

$$f(\mathbf{Z}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = |\boldsymbol{\Sigma}|^{-1/2} g(u),$$

em que

$$u = (\mathbf{Z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\mu}), \quad (3.1)$$

é a distância de Mahalanobis.

Tem-se ainda que  $g(\cdot)$  é uma função escalar contínua e diferenciável, denominada função geradora de densidades, sendo que  $g : \mathbb{R} \rightarrow [0, \infty)$  satisfaz a condição  $\int_0^\infty u^{n-1} g(u^2) du < \infty$  (DE BASTIANI *et al.*, 2015). E mais, diz-se que  $f$  é uma função densidade de probabilidade conjunta. A notação para designar que  $\mathbf{Z}$  tem distribuição elíptica  $n$ -variada passa a ser  $\mathbf{Z} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}; g)$ .

Na sequência, serão apresentadas algumas distribuições pertencentes à família de contornos elípticos e suas respectivas funções densidade conjunta.

### 3.1.1 Distribuição t-Student generalizada

Seja  $\mathbf{Z}$  um vetor aleatório que exibe distribuição t-Student generalizada  $n$ -variada com parâmetro de posição  $\boldsymbol{\mu}$ , matriz escala  $\boldsymbol{\Sigma}$  e parâmetro de forma  $v, \gamma > 0$ , então escreve-se  $\mathbf{Z} \sim Gt_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, v, \gamma)$ . A função densidade de  $\mathbf{Z}$  é dada por (DE BASTIANI, 2012):

$$f(\mathbf{z}) = \frac{\Gamma(\frac{n+v}{2})}{\Gamma(\frac{v}{2})(\pi\gamma)^{n/2}} |\boldsymbol{\Sigma}|^{-1/2} \left(1 + \frac{u}{\gamma}\right)^{-(n+v)/2},$$

em que  $\Gamma(\cdot)$  é a função gama, tal que  $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t} dt$ ;  $u$  é a distância de Mahalanobis (Equação 3.1) e  $n$  é o número de observações realizadas no processo de amostragem.

Considera-se a distribuição t-Student  $n$ -variada, representada por  $t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, v)$ , como um caso particular da generalizada  $n$ -variada quando  $v = \gamma$ , e mais, converge para a distribuição normal quando  $v \rightarrow \infty$ .

A Tabela 1 apresenta um resumo com as funções geradoras de densidades de algumas distribuições elípticas. Outras distribuições pertencentes a esta classe, bem como suas principais características, podem ser encontradas nos trabalhos desenvolvidos por Osorio, Paula e Galea (2007) e De Bastiani *et al.* (2015).

Tabela 1 – Algumas distribuições pertencentes à classe de contornos elípticos, em que  $c$  é a constante normalizadora.

Distribuição	$g(u)$	Condição
Normal	$c \exp(-u/2)$	
t-Student	$v^{-n/2} (1 + u/v)^{-(n+v)/2}$	$v > 0$
t-Student Generalizada	$\gamma^{-n/2} (1 + u/\gamma)^{-(n+v)/2}$	$v, \gamma > 0$

Serão descritas abaixo duas propriedades apresentadas pelas distribuições pertencentes à família de contornos elípticos. Partindo-se do pressuposto que a esperança matemática e a variância do vetor aleatório existam, tem-se que (CYSNEIROS, PAULA e GALEA, 2005):

(i)  $E[\mathbf{Z}] = \boldsymbol{\mu}$ ;

(ii)  $Var(\mathbf{Z}) = \xi \boldsymbol{\Sigma}$ , em que a constante  $\xi > 0$  é dada por:

$$\xi = -2 \frac{\partial \psi(u)}{\partial u} \Big|_{u=0}.$$

Sendo assim, a matriz escala  $\boldsymbol{\Sigma}$  coincide com a matriz de covariância a menos de uma constante  $\xi$ , sendo esta expressada de forma singular para cada distribuição da classe de contornos elípticos. Para a distribuição t-Student generalizada, a constante  $\xi$  é dada por  $\gamma/(v-2)$ , em que  $\gamma > 0$  e  $v > 2$ , sendo  $\gamma$  e  $v$  os parâmetros de forma da distribuição. Para a distribuição t-Student,  $\xi$

é escrita como  $v/(v - 2)$ , em que  $v > 2$  denota os graus de liberdade da distribuição. Já no caso da distribuição normal, tem-se que  $\xi$  é igual a 1.

## 3.2 Agricultura de Precisão e sua aplicabilidade na Geoestatística

A economia globalizada exige uma crescente eficiência em todos seus setores para que se possa sobressair à concorrência imposta. O mesmo vale para a agricultura, considerando-se que produzir mais e reduzir custos são necessidades impostas pelo agronegócio. Visando maximizar a produção, os agricultores passaram a aplicar insumos nas áreas agrícolas, os quais são calculados segundo os valores médios de fertilidade. Em decorrência deste fato, áreas férteis e produtivas são adubadas em excesso, enquanto solos nutricionalmente pobres e com baixa produtividade não são suficientemente corrigidos. Em suma, a uniformização dos tratamentos gera elevação de custos e pode causar sérios impactos ambientais, como contaminação dos recursos hídricos locais, agravada pela aplicação excessiva de agroquímicos (MOLIN, 2001).

A proposta da Agricultura de Precisão (AP) é realizar o gerenciamento da produção agrícola de maneira localizada de acordo com a variabilidade espacial (BERNARDI *et al.*, 2014a). Desta forma, a construção de mapas que expressem a variabilidade de características importantes da produção agrícola é de suma importância para implementação da AP. Enquanto mapas relacionados à produtividade possibilitam localizar regiões com baixos valores, os mapas dos atributos químicos e físicos do solo permitem a identificação de possíveis causas para tal produtividade insatisfatória. E para que se possa mapear e analisar a variabilidade espacial em uma área agrícola, a técnica mais utilizada é a Geoestatística (KESTRING *et al.*, 2015).

## 3.3 Geoestatística

### 3.3.1 Origem da Geoestatística

Por volta de 1951, Daniel Krige, engenheiro de minas na África do Sul, observou ganhos na precisão da estimativa com dados de concentração mineral quando analisou amostras vizinhas. A partir de então, pesquisadores passaram a ponderar que para encontrar sentido nas variações das concentrações dos minérios, como o ouro, era preciso levar em conta as distâncias entre as amostras. Ou seja, deveria-se considerar a localização geográfica e a dependência espacial das amostras, concepção esta que deu origem à Teoria das Variáveis Regionalizadas, considerada a base da Geoestatística. Na década de 60, Georges Matheron apresentou importantes contribuições que permitiram formalizar essa teoria, por isto é considerado o pai da Geoestatística (BERNARDI *et al.*, 2014a).

Até o início dos anos 60, antes do desenvolvimento da Teoria das Variáveis Regionalizadas, a análise dos dados coletados em determinada região era realizada supondo sua aleatoriedade, para que assim se pudesse usufruir de técnicas da Estatística Clássica, tal como a Análise de Variância. Entretanto, a hipótese de aleatoriedade não pode simplesmente ser considerada, ela precisa ser provada. Isto é, para que os dados sejam considerados aleatórios, é necessário verificar que não existe correlação entre amostras vizinhas (VIEIRA, 2000). Sendo assim, a Geoestatística difere da Estatística Clássica por considerar que os valores de uma variável sob estudo estão, de

alguma forma, associados a sua distribuição espaço-temporal. Logo, espera-se que, na média, as observações mensuradas a pequenas distâncias sejam mais similares do que àquelas tomadas a distâncias maiores (ISAAKS e SRIVASTAVA, 1989).

A Geoestatística distingue-se de outras técnicas de estatística espacial especialmente por dois motivos: a localização espacial da variável é considerada fixa e sua distribuição é contínua no espaço (RODRIGUES *et al.*, 2014). Desta forma, objetiva-se, com o uso de técnicas geoestatísticas, caracterizar espacialmente uma variável de interesse pelo estudo de sua distribuição e variabilidade espaciais (LANDIM e YAMAMOTO, 2013).

### 3.3.2 Variável aleatória, regionalizada e processo estocástico

Considera-se *a priori* uma área de interesse  $S$  para a qual se tenha um conjunto de valores mensurados  $\{z(\mathbf{s}_i); i = 1, \dots, n\}$ , em que  $\mathbf{s}_i$  caracteriza uma posição no espaço ou no tempo e é representado por pares de coordenadas. As localizações espaciais são apresentadas como vetores no espaço  $r$ -dimensional ( $\mathbb{R}^r$ ); em particular, no espaço bidimensional, tem-se que  $\mathbf{S} = (x, y)^T$  representa um ponto em  $\mathbb{R}^2$ .

Seja  $\mathbf{s}_i$  uma posição arbitrária dentre àquelas consideradas para amostragem  $i = 1, \dots, n$ , em que  $n$  é o número de amostras. Então,  $z(\mathbf{s}_i)$  é o valor de uma particular realização de uma variável aleatória  $Z$  no ponto  $\mathbf{s}_i$ . Considerando-se as  $n$  observações realizadas nas localizações  $\mathbf{s}_1, \dots, \mathbf{s}_n$  na região  $S \subset \mathbb{R}^r$  ( $r \geq 1$ ), tem-se uma realização de um conjunto de variáveis aleatórias  $Z(\mathbf{s}) = \{Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)\}$ , o qual é denominado Função Aleatória ou Processo Estocástico. Assim, o conjunto de valores amostrados  $z(\mathbf{s}) = \{z(\mathbf{s}_1), \dots, z(\mathbf{s}_n)\}$  é uma realização aleatória e na prática representa a variável regionalizada (SOARES, 2014).

Vale ressaltar que a variável aleatória em um ponto não amostrado  $\mathbf{s}_0$  é também considerada parte do processo estocástico, portanto goza das mesmas propriedades que as variáveis aleatórias em pontos amostrados (SOARES, 2014). Tal fato será importante para a estimativa espacial dos valores de uma variável georreferenciada em localizações não amostradas (krigagem).

### 3.3.3 Estacionariedade do processo estocástico

Para modelar um conjunto de dados que apresente estrutura de dependência espacial, considera-se um processo estocástico  $\mathbf{Z} = \{Z(\mathbf{s}_i); \mathbf{s}_i \in S\}$ , em que  $S$  é um subconjunto de  $\mathbb{R}^r$ , o espaço euclidiano  $r$ -dimensional. Tem-se ainda que  $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$  são observações da variável aleatória  $Z$  sob estudo nas  $\mathbf{s}_i$  localizações espaciais amostradas,  $i = 1, \dots, n$ , em que  $\mathbf{s}_i$  é um vetor  $r$ -dimensional da localização das coordenadas espaciais (URIBE-OPAZO, BORSSOI e GALEA, 2012).

A estacionariedade do conjunto de variáveis aleatórias considera que todas as amostras pertencem a populações com mesmo momento estatístico, ou seja, os pontos amostrais não dependem da posição que ocupam, e sim da distância entre si. As hipóteses de estacionariedade visam reduzir os parâmetros dos quais depende o modelo de inferência geoestatística. Além disso, para que se possa estimar o valor do atributo de interesse em uma localização não amostrada, é necessário introduzir a restrição de que a variável regionalizada seja estacionária pelo menos de ordem 2 (SOARES, 2014). A existência da estacionariedade possibilita que um experimento seja repetido mesmo que as amostras sejam coletadas em locais diferentes, tendo em vista que estas são

consideradas pertencentes a populações de mesmo momento estatístico (ISAAKS e SRIVASTAVA, 1989).

- **Estacionariedade de ordem 2:**

Para ser considerado estacionário de ordem 2, um processo estocástico deve cumprir as seguintes condições:

(i) A esperança matemática de  $Z(\mathbf{s}_i)$  existe e não depende da posição  $\mathbf{s}_i$ . Isto é,  $E[Z(\mathbf{s}_i)] = \mu$  é constante,  $\forall \mathbf{s}_i \in S \subset \mathbb{R}^r$ . Esta condição é denominada estacionariedade de primeira ordem;

(ii) Para cada par de variáveis aleatórias  $Z(\mathbf{s}_i, Z(\mathbf{s}_i + \mathbf{h}))$ , a covariância existe e depende somente de  $\mathbf{h}$ . Ou seja,

$$\begin{aligned} Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_i + \mathbf{h})] &= E[Z(\mathbf{s}_i)Z(\mathbf{s}_i + \mathbf{h})] - E[Z(\mathbf{s}_i)]E[Z(\mathbf{s}_i + \mathbf{h})] \\ &= E[Z(\mathbf{s}_i)Z(\mathbf{s}_i + \mathbf{h})] - \mu^2, \forall \mathbf{s}_i \in S \\ &= Cov(\mathbf{h}), \end{aligned}$$

em que  $\mathbf{h}$  é o vetor distância entre duas localizações desconhecidas, tal que  $\|\mathbf{h}\| = h$ .

Observa-se na condição (ii) que a variância de uma variável aleatória é um caso particular da covariância quando  $h = 0$ .

Tais condições são tidas como os dois momentos da estacionariedade de ordem 2. Um momento de ordem  $k$  de uma variável aleatória  $X$  é dado por  $\mu_k = E[X^k] < \infty$ .

Dizer que a variância é estacionária de ordem 2, ou seja, que (ii) se cumpre, é uma hipótese muito forte, tendo em vista que muitos fenômenos espaciais apresentam uma capacidade infinita de dispersão e, por conseguinte, a variância não é finita (estacionária). Implicando assim, por lógica matemática, que a covariância também não é estacionária e, portanto, o processo estocástico não é estacionário de ordem 2 (LANDIM e YAMAMOTO, 2013). Para tais situações, adota-se uma hipótese menos restritiva sobre o processo estocástico, a qual é descrita na sequência.

- **Hipótese Intrínseca:**

Também conhecida como Fracamente Estacionária, essa hipótese verifica as condições de estacionariedade de ordem 2 sobre o incremento  $[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]$  e não propriamente sobre  $Z(\mathbf{s})$ , sendo que  $\mathbf{h}$  representa o vetor entre dois pontos no espaço. A hipótese intrínseca pode ser considerada uma estacionariedade local (JOURNEL e HUIJBREGTS, 1978).

O processo estocástico é intrínseco se (LANDIM e YAMAMOTO, 2013):

(i)  $E[Z(\mathbf{s}_i)] = \mu$  é constante,  $\forall \mathbf{s}_i \in S \subset \mathbb{R}^r$ ;

(ii)  $Var[Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i)] = \{E[Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i)]\}^2, \forall \mathbf{s}_i \in S$ , isto é, o incremento tem variância finita.

Uma ponderação importante a ser feita é que um processo estocástico estacionário de ordem 2 verifica, automaticamente, a hipótese intrínseca. A volta não é necessariamente verdadeira.

### 3.3.4 Modelos espaciais lineares

Seja  $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$  um vetor aleatório de dimensão  $n \times 1$ , das  $n$  observações da variável aleatória  $Z$ . Considere que  $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$  são observações da variável aleatória sob estudo nas  $\mathbf{s}_i$  localizações espaciais amostradas,  $i = 1, \dots, n$ . Cada observação  $Z(\mathbf{s}_i)$  é gerada pelo seguinte modelo linear (URIBE-OPAZO, BORSSOI e GALEA, 2012):

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad i = 1, \dots, n \quad (3.2)$$

em que  $\mu(\mathbf{s}_i)$  é o termo determinístico e  $\epsilon(\mathbf{s}_i)$  é o termo estocástico, e ambos dependem da localização espacial  $\mathbf{s}_i$  em que a variável  $Z$  é observada. Assume-se que a média do erro estocástico é zero, isto é,  $E[\epsilon(\mathbf{s}_i)] = 0$ , e que a variação entre os pontos no espaço é determinada por uma função de covariância  $Cov[\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)] = C(\mathbf{s}_i, \mathbf{s}_j)$ , em que  $i, j = 1, \dots, n$ .

Para algumas funções conhecidas de  $\mathbf{s}_i$ ,  $x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i)$ , tem-se que a média do processo estocástico é dada por:

$$\mu(\mathbf{s}_i) = \sum_{j=1}^p x_j(\mathbf{s}_i) \beta_j,$$

em que:  $x_j(\mathbf{s}_i)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, p$ , representa o valor da  $j$ -ésima covariável tomadas na  $n$ -ésima posição, e  $\beta_j$  é um  $j$ -ésimo parâmetro desconhecido a ser estimado. E mais, cada família de funções covariância  $C(\mathbf{s}_i, \mathbf{s}_j)$  é especificada por um vetor de parâmetros  $\varphi = (\varphi_1, \dots, \varphi_q)^T$ . Para os estudos realizados, será considerado  $q = 3$ , pois está se assumindo um processo isotrópico.

É possível reescrever  $\mathbf{Z}$  e  $\boldsymbol{\mu}$  matricialmente da seguinte forma (URIBE-OPAZO, BORSSOI e GALEA, 2012):

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.3)$$

$$\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta},$$

em que:  $\mathbf{X}$ , de dimensão  $n \times p$ , é denominada matriz de delineamento, na qual a  $i$ -ésima linha é dada por  $\mathbf{x}_i^T = (x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i))$ ;  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  e  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$ .

Em particular, quando se considera um processo estocástico estacionário (de primeira ordem), tem-se que  $\mathbf{X} = \mathbf{1}$  e  $\boldsymbol{\beta} = \mu$ , ou seja,  $\boldsymbol{\mu} = \mu \mathbf{1}$ . O que significa que para qualquer localização  $\mathbf{s}_i$ , a média é constante igual a  $\mu$ .

A esperança do vetor de erros aleatórios  $\boldsymbol{\epsilon}$  é dada por  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ . Já a matriz escala de  $\boldsymbol{\epsilon}$  é dada por  $\boldsymbol{\Sigma} = [(\sigma_{ij})]$ ,  $i, j = 1, \dots, n$ , em que  $\sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$  é a covariância entre os valores da variável georreferenciada nos pontos  $\mathbf{s}_i$  e  $\mathbf{s}_j$ . Assume-se que  $\boldsymbol{\Sigma}$  é inversível e  $\mathbf{X}$  tem posto coluna completo, ou seja, o número de colunas linearmente independentes é o maior valor possível (ASSUMPÇÃO, URIBE-OPAZO e GALEA, 2014).

Pode-se ainda escrever a matriz escala utilizando-se a forma paramétrica (DE BASTIANI *et al.*, 2015):

$$\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3), \quad (3.4)$$

sendo que:  $\mathbf{R}(\varphi_3) = [(r_{ij})]$  é uma matriz  $n \times n$  simétrica, cujos elementos da diagonal são iguais a  $r_{ii} = 1$ ,  $i = 1, \dots, n$ ; enquanto os demais elementos são denotados por  $r_{ij} = (1/\varphi_2)\sigma_{ij}$ , para  $i \neq j = 1, \dots, n$  e  $\varphi_2 \neq 0$ , e  $r_{ij} = 0$  para  $i \neq j = 1, \dots, n$  e  $\varphi_2 = 0$ .

### 3.3.5 Funções semivariância e covariância

A variância associada a média é dada por (JOURNEL e HUIJBREGTS, 1978):

$$Var[Z(\mathbf{s})] = E[Z(\mathbf{s})^2] - \{E[Z(\mathbf{s})]\}^2 = E\{[Z(\mathbf{s}) - \mu]^2\}. \quad (3.5)$$

Nas deduções que seguem, será considerado  $h = \|\mathbf{h}\|$ , ou seja, somente a norma do vetor  $\mathbf{h}$ , em que  $h$  representa a distância euclidiana entre os pontos.

De (3.5), tem-se ainda que:

$$\begin{aligned} Var[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})] &= E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\} - \{E[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]\}^2 \\ &= E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\} - \{E[Z(\mathbf{s}+\mathbf{h})] - E[Z(\mathbf{s})]\}^2 \\ &= E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\} - \{\mu - \mu\}^2, \end{aligned}$$

pois  $E[Z(\mathbf{s}+\mathbf{h})] = E[Z(\mathbf{s})] = \mu$ .

Donde,

$$Var[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})] = E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\}. \quad (3.6)$$

A função semivariância é definida por (CRESSIE, 2015):

$$\gamma(h) = \frac{1}{2}E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\}, \quad (3.7)$$

Considerando-se a igualdade obtida na expressão 3.6, a semivariância definida na Equação 3.7 pode ser escrita equivalentemente da seguinte forma:

$$\gamma(h) = \frac{1}{2}Var[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]. \quad (3.8)$$

A função covariância é dada por (SOARES, 2014):

$$Cov(h) = E[Z(\mathbf{s})Z(\mathbf{s}+\mathbf{h})] - E[Z(\mathbf{s})]E[Z(\mathbf{s}+\mathbf{h})], \quad (3.9)$$

ou ainda, da condição (i) de estacionariedade de ordem 2,

$$Cov(h) = E[Z(\mathbf{s})Z(\mathbf{s}+\mathbf{h})] - \mu^2. \quad (3.10)$$

Considerando-se que  $h = 0$ , tem-se ainda que:

$$\begin{aligned} E[Z(\mathbf{s})Z(\mathbf{s})] - \mu^2 &= E[Z(\mathbf{s})^2] - \mu^2 \\ &= E[Z(\mathbf{s})^2] - 2\mu^2 + \mu^2 \\ &= E\{Z(\mathbf{s})^2 - 2(E[Z(\mathbf{s})])^2 + (E[Z(\mathbf{s})])^2\} \\ &= E\{[Z(\mathbf{s}) - E[Z(\mathbf{s})]]^2\} \\ &= E\{[Z(\mathbf{s}) - E[Z(\mathbf{s})]][Z(\mathbf{s}) - E[Z(\mathbf{s})]]\} \end{aligned} \quad (3.11)$$

Logo, das expressões 3.10 e 3.11, vem que:

$$Cov(h) = E\{[Z(\mathbf{s}+\mathbf{h}) - E[Z(\mathbf{s})]][Z(\mathbf{s}) - E[Z(\mathbf{s}+\mathbf{h})]]\}.$$

Caso o processo estocástico cumpra com as condições impostas pela estacionariedade de ordem 2, pode-se considerar que as funções semivariância (Equação 3.7) e covariância (Equação 3.9) são equivalentes para caracterizar a dependência espacial (VIEIRA, 2000).

Na sequência, serão apresentados alguns resultados que relacionam as funções variância e covariância. O intuito é utilizá-los para obter uma equação que relaciona as funções semivariância e covariância.

- (i)  $Var[Z(\mathbf{s})] = Cov(0)$ ;
- (ii)  $Cov(\mathbf{s}, \mathbf{s}+\mathbf{h}) = Cov(h)$ ;
- (iii)  $E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\} = 2[Cov(0) - Cov(h)]$ .

(i) Para verificar esse item, basta considerar  $h = 0$ , isto é:

$$\begin{aligned} Var[Z(\mathbf{s})] &= Cov[Z(\mathbf{s}), Z(\mathbf{s})] \\ &= Cov[Z(\mathbf{s}), Z(\mathbf{s}+\mathbf{0})] \\ &= Cov(0). \end{aligned}$$

Analogamente verifica-se que  $Var[Z(\mathbf{s}+\mathbf{h})] = Cov(0)$ .

(ii) Este item nos diz basicamente que a covariância entre dois pontos depende somente da distância  $h$  entre eles e não da posição que ocupam. Logo, se essa igualdade é verificada, tem-se que a covariância é estacionária, implicando na estacionariedade das funções variância e semivariância (LANDIM e YAMAMOTO, 2013). A veracidade desse item é constatada quando se utilizam a definição de covariância (Equação 3.9) e a condição (i) de estacionariedade de ordem 2:

$$\begin{aligned} Cov(\mathbf{s}, \mathbf{s}+\mathbf{h}) &= E[Z(\mathbf{s})Z(\mathbf{s}+\mathbf{h})] - E[Z(\mathbf{s})]E[Z(\mathbf{s}+\mathbf{h})] \\ &= E[Z(\mathbf{s})Z(\mathbf{s}+\mathbf{h})] - \mu^2 \\ &= Cov(h), \end{aligned}$$

O item (iii) é verificado quando são utilizados a igualdade 3.5 e os itens (i) e (ii), donde vem que:

$$\begin{aligned} E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\} &= E\{Z(\mathbf{s}+\mathbf{h})^2 - 2Z(\mathbf{s}+\mathbf{h})Z(\mathbf{s}) + Z(\mathbf{s})^2\} \\ &= E\{Z(\mathbf{s}+\mathbf{h})^2 - 2Z(\mathbf{s}+\mathbf{h})Z(\mathbf{s}) + Z(\mathbf{s})^2 + 2\mu^2 - 2\mu^2\} \\ &= E[Z(\mathbf{s}+\mathbf{h})^2] - \mu^2 - 2(E[Z(\mathbf{s}+\mathbf{h})Z(\mathbf{s})] - \mu^2) + E[Z(\mathbf{s})^2] - \mu^2 \\ &= Cov(0) - 2Cov(h) + Cov(0) \\ &= 2[Cov(0) - Cov(h)]. \end{aligned}$$

Considerando-se a expressão 3.7 e o item (iii), vem que

$$2\gamma(h) = E\{[Z(\mathbf{s}+\mathbf{h}) - Z(\mathbf{s})]^2\} = 2[Cov(0) - Cov(h)].$$

Logo, segue que,

$$\gamma(h) = Cov(0) - Cov(h), \tag{3.12}$$

sendo essa uma equação que relaciona as funções semivariância e a covariância.

A função semivariância tem as seguintes propriedades (CRESSIE, 2015):

- (i)  $\gamma(0) = 0$ ;
- (ii)  $\gamma(h) = \gamma(-h) \geq 0$ , ou seja, a função semivariância é par e sempre produz valores positivos;
- (iii)  $\lim_{\|h\| \rightarrow \infty} \frac{\gamma(h)}{\|h\|^2} = 0$ , isto é,  $\gamma(h)$  cresce mais lentamente que  $h^2$ .

A partir da relação 3.12, pode-se obter a função correlação  $\rho$ , da seguinte forma:

$$Cov(h) = Cov(0) - \gamma(h),$$

supondo  $Cov(0) \neq 0$ , vem que (JOURNEL e HUIJBREGTS, 1978):

$$\frac{Cov(h)}{Cov(0)} = 1 - \frac{\gamma(h)}{Cov(0)} = \rho(h). \quad (3.13)$$

Para processos estocásticos estacionários, a função correlação traz o mesmo tipo de informação acerca da dependência espacial que as funções semivariância e covariância, embora a função correlação apresente uma limitação para valor máximo apresentado, sendo este 1 (ISAAKS e SRIVASTAVA, 1989).

Quanto menor a distância  $h$  entre os pontos, maior será a covariância e menor a variância entre eles. Isto é, quando a distância se faz infinitamente pequena ( $h \rightarrow 0$ ), a covariância tende a ser máxima e a variância mínima, justamente porque a dependência entre os pontos é inversamente proporcional à distância. Desta forma, se os dados forem estacionários, isto é, não houver descontinuidade, a semivariância varia de zero (quando  $h = 0$ ) até um valor igual à variância dos dados; enquanto a covariância, para  $h = 0$  é igual a variância dos dados, e tende a zero conforme a distância aumenta. Graficamente, a curva que representa as funções semivariância e covariância *versus* a distância têm a mesma forma, exceto por ser invertida (LANDIM, 2007).

Em relação à função semivariância, existe na literatura uma gama de estimadores (TEIXEIRA e SCALON, 2013), todavia, o mais comumente utilizado é o estimador clássico de Matheron, dado por (CRESSIE, 2015):

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(\mathbf{s}_i + \mathbf{h}) - Z(\mathbf{s}_i)]^2, \quad (3.14)$$

em que  $\hat{\gamma}(h)$  é o valor estimado da semivariância considerando-se a distância  $h$ ,  $[Z(\mathbf{s}_i), Z(\mathbf{s}_i + \mathbf{h})]$  são os pares amostrados que distam  $h$  entre si e  $N(h)$  é o número de pares de pontos tal que isso ocorre.

### 3.4 Gráfico semivariograma

Considerando o que foi discorrido até então, a pergunta agora poderia ser: como caracterizar e prever o comportamento espacial da variável regionalizada  $Z(\mathbf{s}_i), i = 1, \dots, n$ , mensurada nos  $n$  pontos amostrais dispostos na área de interesse  $S$ ?

O semivariograma permite estudar o grau de dependência espacial entre as amostras coletadas ao longo da área sob estudo. Há, portanto, os seguintes semivariogramas: verdadeiro, experimental e teórico. O semivariograma verdadeiro é aquele real do fenômeno, sempre desconhecido. O semivariograma experimental (Figura 1) é o único conhecido, pois é obtido considerando-se os pontos observados na área. O semivariograma teórico apresenta um modelo ajustado aos pontos do experimental e será utilizado como referência na predição dos locais não amostrados (GUERRA, 1988).

Dizer que um semivariograma experimental apresenta características próximas do "ideal" significa que, intuitivamente, espera-se que observações geograficamente mais próximas tenham

comportamento mais semelhante do que àquelas mais distantes entre si (LANDIM e YAMAMOTO, 2013).

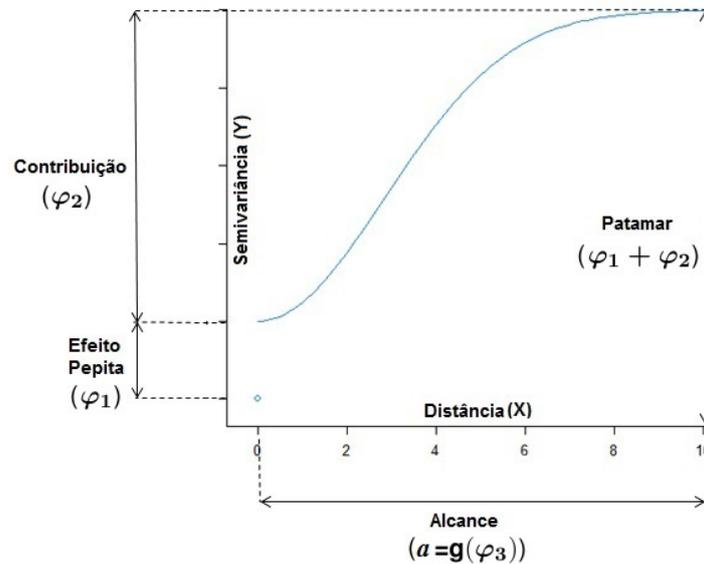


Figura 1 – Semivariograma experimental.

Na Seção 3.3.3, foi afirmado que um processo estocástico cuja hipótese intrínseca é verificada, não necessariamente também é estacionário de ordem 2. Isto porque os gráficos semivariogramas de processos estocásticos de ordem 2 são sempre limitados (ou seja, com patamar). Entretanto, alguns semivariogramas que satisfazem a hipótese intrínseca não são limitados, logo, não verificam uma das condições da estacionaridade de ordem 2. De modo que, um processo estocástico estacionário de ordem  $K > 0$ , será estacionário de todas as ordens menores que  $K$  (ANDRIOTTI, 2013).

Para se construir um semivariograma experimental (o qual ao longo do texto será denominado apenas de semivariograma), é preciso dispor de um conjunto de valores observados, fazendo com que o gráfico seja bastante sensível em relação à distribuição dos pontos amostrais, o qual pode ser regular ou irregular (LANDIM e YAMAMOTO, 2013).

Para grades regulares, o estudo é feito considerando-se uma direção ao longo de uma linha, ou uma série de linhas paralelas, utilizando-se as  $N$  possíveis diferenças entre os pares de pontos para a distância  $h$  tomada (CRESSIE, 2015).

Já para dados irregularmente distribuídos na grade, não é possível, *a priori*, encontrar o mínimo de pares de amostras recomendado com exatamente a mesma distância, para que se possa calcular a semivariância em determinada direção. Segundo Landim (2007), essa problemática pode ser contornada definindo-se uma tolerância  $h'$  para a distância  $h$  entre os pares amostrados, bem como uma tolerância angular  $\alpha'$  para a direção  $\alpha$  considerada. Sendo assim, para o cálculo do semivariograma a partir de uma grade irregular, ao longo de uma direção  $\alpha$  definida, tomam-se as amostras que se encontram no ângulo  $\alpha \pm \alpha'$ . Na sequência, recorre-se àquelas pares de amostras que distam  $h \pm h'$ ,  $2h \pm h'$ ,  $\dots$ ,  $mh \pm h'$ , em que  $h$  é a distância básica e  $mh$  é um múltiplo de  $h$ . Tal procedimento, exibido na Figura 2, deve permitir cobrir toda a área de interesse.

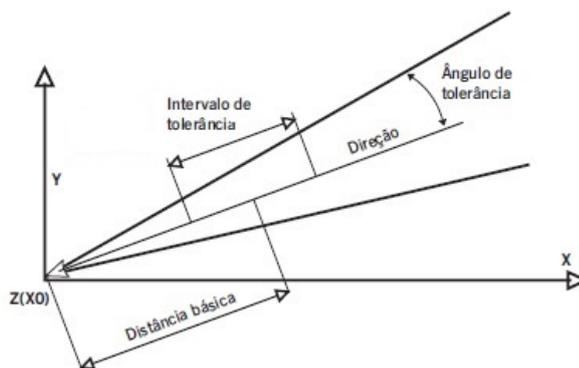


Figura 2 – Esquema para localização de pontos em grades irregulares para o cálculo da função semivariância. Fonte: Landim, 2007, p. 26, adaptada.

Além dessa questão da disposição dos pontos na grade amostral e da escolha do estimador para função semivariância, é preciso definir um corte (*cutoff*) representando a distância limite até onde o semivariograma será construído. Clark (1979) recomenda um *cutoff* de 50% da distância máxima entre os pontos amostrados.

Um semivariograma é dito omnidirecional quando é obtido utilizando-se as informações de todas as direções. Enquanto um semivariograma direcional é calculado com as informações provenientes de uma única direção.

A análise da dependência espacial da variável regionalizada pode ser feita considerando os valores amostrais em dois pontos, tendo em vista que a influência de um ponto sobre outro é inversamente proporcional à distância que os separa (JOURNEL e HUIJBREGTS, 1978). O gráfico de envelope para semivariogramas omnidirecionais possibilita verificar, de forma exploratória, a presença de dependência espacial da variável sob estudo. Um ponto do semivariograma fora do envelope indica que essa variável apresenta certa dependência espacial. Para analisar a continuidade espacial nesse gráfico, é preciso observar se há um aumento na semivariância conforme também se aumenta a distância. Caso isso ocorra, têm-se indícios de continuidade espacial. Além disso, como a região do envelope representa a região sem dependência espacial, verificar a existência de ponto(s) fora dessa região é outro indicativo de continuidade espacial (DIGGLE e RIBEIRO JR, 2007).

De acordo com Soares (2014), a continuidade espacial de alguns fenômenos pode variar de acordo com a direção considerada. Geralmente, estudam-se semivariogramas nas direções  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$  (Figura 3).

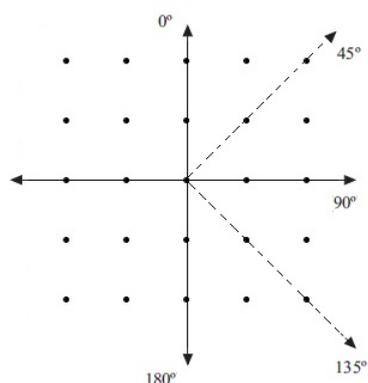


Figura 3 – Direções utilizadas para análise de tendência direcional. Fonte: De Bastiani, 2012, p. 10 adaptada.

Caso o semivariograma seja o mesmo para qualquer direção tomada, tem-se isotropia, indicando que a função semivariância depende do módulo do vetor  $h$  e não da direção, isto é, depende somente da distância  $h$  considerada; caso se constate uma direção privilegiada, diz-se que o fenômeno é anisotrópico, existindo, portanto, uma tendência direcional (SOARES, 2014).

Após discorrer sobre o semivariograma e considerar que esse possibilita analisar a dependência espacial de uma variável que se está estudando, serão apresentados na sequência os parâmetros que caracterizam a estrutura de dependência espacial, os quais foram exibidos na Equação 3.4 e visualizados na Figura 1.

### 3.5 Parâmetros de dependência espacial

- **Efeito Pepita** (*nugget effect*,  $\varphi_1$ ):

Este parâmetro revela a descontinuidade na origem, causada pela variância aleatória. O valor da função semivariância no ponto  $h = 0$  é, teoricamente, nulo, pois a variância de um valor amostrado com ele mesmo é zero (TEIXEIRA e SCALON, 2013). Entretanto, próximo à origem, a curva do gráfico semivariograma geralmente apresenta uma descontinuidade. Isso ocorre devido a variações encontradas a distâncias menores do que a menor distância entre as amostras; além de também estar associada a erros de medição ou análise aos quais a pesquisa está vinculada. Contudo, é impossível quantificar se a maior parcela de descontinuidade provém dos erros ou da variabilidade em pequena escala, logo, ambas estão associadas à uma única parcela representada por este parâmetro (ISAAKS e SRIVASTAVA, 1989).

- **Contribuição** (*sill*,  $\varphi_2$ ):

É a variância de dispersão, pois representa as diferenças espaciais dos valores de uma mesma variável, tomadas em dois pontos cada vez mais distantes (DE BASTIANI, 2012).

- **Alcance** (*range*,  $a$ ):

Representa a amplitude dentro da qual as amostras apresentam autocorrelação espacial, isto é, marca a distância a partir da qual um ponto da variável aleatória sob estudo não tem mais

influência sobre o ponto vizinho, indicando que, após essa distância, as amostras tornam-se espacialmente independentes. O alcance é obtido como função do parâmetro  $\varphi_3$ ,  $a = g(\varphi_3)$  (LANDIM e YAMAMOTO, 2013).

Para Vieira (2000), o alcance é a “linha divisória” entre o uso da Estatística Clássica e da Geoestatística. Isto porque amostras separadas por distâncias maiores que  $a$  apresentam uma distribuição aleatória e, por conseguinte, são estatisticamente independentes entre si, de modo que é cabível o uso da Estatística Clássica para estudá-las.

- **Patamar** (*still*,  $\varphi = \varphi_1 + \varphi_2$ ):

É a variância máxima da variável sob estudo, pois a partir deste ponto a variância da diferença entre os pares de amostras passa a ser invariante com a distância. Sendo assim, o patamar é o valor máximo atingido no semivariograma, a partir do qual a curva se estabiliza (TEIXEIRA e SCALON, 2013).

Importante: quando o efeito pepita for, aproximadamente, igual ao patamar, tem-se o efeito pepita puro, indicando que os dados apresentam uma distribuição aleatória, não recebendo, portanto, influência espacial (LANDIN e YAMAMOTO, 2013).

Quando são utilizados tais parâmetros, torna-se viável verificar o quão aleatórios são os dados, tendo em vista que o gráfico semivariograma possibilita apenas uma avaliação visual da dependência espacial, mas não fornece um valor numérico que permita mensurar o grau de aleatoriedade dos dados. Dentre os índices de dependência espacial encontrados na literatura, tais como os propostos por Guerra (1988), Cambardella *et al.* (1994) e Seidel e Oliveira (2014), optou-se, neste trabalho, por utilizar o de Cambardella *et al.* (1994).

Cambardella *et al.* (1994) sugerem o uso do efeito pepita relativo (*EPR*), o qual é calculado para cada modelo geoestatístico ajustado, e dado por:

$$EPR(\%) = \left( \frac{\varphi_1}{\varphi} \right) 100,$$

onde  $\varphi_1$  e  $\varphi$  representam o efeito pepita e o patamar, respectivamente.

Tem-se a seguinte classificação para o *EPR*:

- $EPR \leq 0,25$ , indica uma forte dependência espacial;
- $0,25 < EPR < 0,75$ , indica uma média dependência espacial;
- $EPR \geq 0,75$ , indica uma fraca dependência espacial.

### 3.6 Modelos teóricos

O gráfico do semivariograma experimental obtido a partir do cálculo da função semivariância fornece uma série de pontos discretos sobre os quais se deseja ajustar uma função contínua. É fundamental que o modelo ajustado represente a tendência da semivariância em relação à distância  $h$ . Entretanto, o procedimento de ajuste não é direto nem automático e, dependendo do ajuste obtido, pode-se ou não redefinir o modelo teórico até encontrar algum considerado satisfatório (CAMARGO, 1998).

Journal e Huijbregts (1978) alertam para o fato de que o modelo teórico deve cumprir a condição de positividade definida condicional. Isso significa que o modelo que satisfizer essa condição garante que  $\gamma(h) > 0$  e  $\gamma(-h) = \gamma(h)$ , para qualquer  $h$  (VIEIRA, 2000).

Tem-se uma capacidade infinita de dispersão para fenômenos não-estacionários, logo a variância não é finita e, conseqüentemente, a função semivariância cresce indefinidamente com a distância, sem se estabilizar, de modo que o gráfico semivariograma não apresenta patamar (SOARES, 2014). Sendo assim, os modelos teóricos utilizados na geoestatística para definir a estrutura de variabilidade espacial são divididos em modelos não transitivos, que não possuem patamar, e transitivos, que apresentam patamar (LANDIM e YAMAMOTO, 2013). Os principais modelos transitivos são: esférico, exponencial, gaussiano, família Matérn e exponencial potência, os quais serão utilizados no desenvolvimento desse trabalho.

Os modelos exponencial e gaussiano serão definidos abaixo segundo Soares (2014) e Cressie (2015), e os modelos da família Matérn de acordo com Matérn (2013). No que segue,  $h$  representa a distância,  $\varphi_1$  o efeito pepita,  $\varphi_2$  a contribuição,  $\varphi_3$  a função do alcance e  $a$  o alcance prático.

### 3.6.1 Modelo exponencial

Este modelo, válido no espaço  $r$ -dimensional, para  $r = 1, 2, 3$ , apresenta comportamento aproximadamente linear na origem, porém, a curva que o representa cresce mais rápido e atinge o patamar assintoticamente. O alcance prático do modelo exponencial é dado por  $a = 3\varphi_3$ , e sua função semivariância é expressada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - e^{-\left(\frac{h}{\varphi_3}\right)} \right], & \text{se } h > 0. \end{cases} \quad (3.15)$$

Define-se a função covariância por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & \text{se } h = 0 \\ \varphi_2 e^{-\left(\frac{h}{\varphi_3}\right)}, & \text{se } h > 0, \end{cases} \quad (3.16)$$

enquanto a função de correlação é dada por:

$$\rho(h) = \begin{cases} 1, & \text{se } h = 0 \\ e^{-\left(\frac{h}{\varphi_3}\right)}, & \text{se } h > 0. \end{cases} \quad (3.17)$$

### 3.6.2 Modelo gaussiano

O modelo gaussiano, caracterizado por apresentar um comportamento parabólico próximo à origem, assim como o exponencial, também atinge o patamar assintoticamente, além de ser mais contínuo que este último e também que o esférico. O alcance prático do modelo gaussiano é igual a  $a = \sqrt{3}\varphi_3$  e, assim como os demais, é válido para todo espaço  $r$ -dimensional,  $r = 1, 2, 3$ . Sua função semivariância é dada por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - e^{-\left(\frac{h}{\varphi_3}\right)^2} \right], & \text{se } h > 0. \end{cases} \quad (3.18)$$

A função covariância é expressa por:

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & \text{se } h = 0 \\ \varphi_2 \left[ 1 - e^{-\left(\frac{h}{\varphi_3}\right)^2} \right], & \text{se } h > 0. \end{cases} \quad (3.19)$$

E a função de correlação é definida como:

$$\rho(h) = \begin{cases} 1, & \text{se } h = 0 \\ 1 - e^{-\left(\frac{h}{\varphi_3}\right)^2}, & \text{se } h > 0. \end{cases} \quad (3.20)$$

### 3.6.3 Modelo da família Matérn

A função semivariograma é definida por:

$$\gamma(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - (2^{\kappa-1} \Gamma(\kappa))^{-1} \left(\frac{h}{\varphi_3}\right)^{\kappa} K_{\kappa} \left(\frac{h}{\varphi_3}\right) \right], & \text{se } h > 0. \end{cases} \quad (3.21)$$

Tem-se que  $K_{\kappa}$  é a função de Bessel modificada do terceiro tipo de ordem  $\kappa > 0$ , em que  $\kappa$  é o parâmetro de forma, que é dada por:

$$K_{\kappa}(u) = \frac{1}{2} \int_0^{\infty} x^{\kappa-1} e^{-\frac{1}{2}u(x+\frac{1}{x})} dx,$$

para  $\varphi_3, \kappa > 0$ .

O parâmetro de forma  $\kappa$  determina a suavização analítica do processo  $Z(s)$ , em que este processo é  $(\kappa - 1)$  vezes diferenciável, sendo  $(\kappa)$  o menor inteiro maior ou igual a  $\kappa$ . Devido à impossibilidade de se considerar a semivariância para todos os valores de  $\kappa$ , é fixado um valor para  $\kappa$  de acordo com o contexto da aplicação, ou toma-se um conjunto limitado de valores para o mesmo (DE BASTIANI, 2012).

A função Gama ( $\Gamma$ ) é expressa por:

$$\Gamma(\kappa) = \int_0^{\infty} x^{\kappa-1} e^{-x} dx, \quad \kappa > 0.$$

Magalhães (2015) elenca algumas propriedades dessa função:

(i)  $\Gamma(\kappa + 1) = \kappa \Gamma(\kappa)$ ,  $\kappa > 0$ ;

(ii)  $\Gamma(n) = (n - 1)!$ ,  $n \in \mathbb{Z}^+$ ;

(iii)  $\Gamma\left(\frac{1}{2}\right) = \sqrt{\pi}$ .

Tem-se que o modelo exponencial é um caso particular da família Matérn, e coincidem quando  $\kappa = 0,5$ . E mais, conforme se aumenta o valor de  $\kappa$ , o comportamento do modelo se aproxima do gaussiano (WEBSTER e OLIVER, 1990). Nos modelos da família Matérn, conforme se aumenta o valor de  $\kappa$ , o valor do alcance também aumenta. Por exemplo, o alcance prático é, aproximadamente,  $3\varphi_3$ ;  $3,45\varphi_3$ ;  $4\varphi_3$ ;  $4,75\varphi_3$ ;  $5,37\varphi_3$  e  $5,92\varphi_3$  quando se considera  $\kappa = 0,5$ ;  $\kappa = 0,7$ ;  $\kappa = 1$ ;  $\kappa = 1,5$ ;  $\kappa = 2$  e  $\kappa = 2,5$ , nesta ordem (DE BASTIANI, 2012).

As funções covariância de correlação dos modelos da família Matérn são dadas, respectivamente, por:

$$C(h) = \begin{cases} 0, & \text{se } h = 0 \\ \varphi_2 \left[ (2^{\kappa-1} \Gamma(\kappa))^{-1} \left( \frac{h}{\varphi_3} \right)^\kappa K_\kappa \left( \frac{h}{\varphi_3} \right) \right], & \text{se } h > 0. \end{cases} \quad (3.22)$$

$$\rho(h) = \begin{cases} 1, & \text{se } h = 0 \\ (2^{\kappa-1} \Gamma(\kappa))^{-1} \left( \frac{h}{\varphi_3} \right)^\kappa K_\kappa \left( \frac{h}{\varphi_3} \right), & \text{se } h > 0. \end{cases} \quad (3.23)$$

De Bastiani *et al.* (2015) fizeram um interessante comparativo entre o comportamento das funções covariância dos modelos esférico, exponencial, gaussiano, exponencial potência ( $p = 1.5$ ) e Matérn ( $\kappa = 0.7$ ), no qual se verifica que o modelo esférico apresenta um alcance prático bem menor que os demais modelos. Em outro comparativo considerou-se a função covariância do modelo da família Matérn com diferentes valores do parâmetro  $\kappa$ , e se constatou que quando se mantém o parâmetro  $\varphi_3$  fixo e se aumenta o valor de  $\kappa$ , o valor de  $\frac{a}{\varphi_3}$  também aumenta.

### 3.7 Modelos espaciais lineares com distribuição de probabilidade da família de contornos elípticos

Considerando-se a crescente complexidade dos fenômenos estudados pela Geoestatística, torna-se fundamental o uso de modelos que possam englobar as características intrínsecas a esses. Os modelos espaciais lineares com distribuição de contornos elípticos representam uma alternativa interessante para explicar a estrutura de variabilidade espacial, além de ter uma versatilidade que permite estender a classe dos erros a outras distribuições além da normal, de modo a melhor acomodar as observações atípicas.

Suponha que o processo estocástico  $\{Z(\mathbf{s}_i); \mathbf{s}_i \in S \subset \mathbb{R}^r\}$  é estacionário (Seção 3.3.4), em que  $\mathbb{R}^r$  é um espaço euclidiano  $r$ -dimensional,  $r \geq 1$ . Considere também que o vetor aleatório  $\mathbf{Z}$  tem distribuição elíptica  $n$ -variada,  $\mathbf{Z} \sim El_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, g(u))$ , sendo o vetor de médias  $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ , de ordem  $n \times 1$ , em que  $\mathbf{X}$  é a matriz de delineamento  $n \times p$  (Seção 3.3.4),  $g(\cdot)$  é a função geradora de densidades e  $u$  a distância de Mahalanobis (Seção 3.1). Tem-se que os parâmetros desconhecidos do modelo,  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$ , com  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  e  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_q)^T$ , podem ser estimados maximizando o logaritmo da função de verossimilhança, o qual é expresso por (DE BASTIANI *et al.*, 2015):

$$\mathcal{L}(\boldsymbol{\theta}) = -\frac{1}{2} \log |\boldsymbol{\Sigma}| + \log(g(u)), \quad (3.24)$$

em que  $\boldsymbol{\Sigma}$  é a matriz escala (Equação 3.4) e  $u$  é a distância de Mahalanobis (Equação 3.1).

Ao se partir do pressuposto que a função geradora de densidades  $g(\cdot)$  (Tabela 1) é contínua e diferenciável, define-se (DE BASTIANI *et al.*, 2015):

$$V_{g(u)} = \frac{\partial \log[g(u)]}{\partial u} = \frac{g'(u)}{g(u)} \quad \text{e} \quad V'_{g(u)} = \frac{\partial V_{g(u)}}{\partial u},$$

sendo que o cálculo de tais expressões depende da distribuição de contorno elíptico com a qual se está modelando.

Pode-se interpretar  $V_{g(u)}$  como um peso. Sendo que, para a distribuição t-Student,  $V_{g(u)}$  é inversamente proporcional à distância de Mahalanobis ( $u$ ). Então, observações mais distantes ( $u$  maior) receberão menores valores de  $V_{g(u)}$  e produzirão estimativas de parâmetros mais robustas para observações atípicas do que a distribuição normal (OSORIO, PAULA e GALEA, 2007).

A Tabela 2 sintetiza os valores de  $V_{g(u)}$  para algumas distribuições da família de contornos elípticos (OSORIO, PAULA e GALEA, 2007); (DE BASTIANI *et al.*, 2015).

Tabela 2 – Valores de  $V_{g(u)}$  para algumas distribuições pertencentes à classe elíptica.

Distribuição	$V_{g(u)}$	Condição
Normal	$-\frac{1}{2}$	
t-Student	$-\frac{1}{2}(v+n)/(v+u)$	$v > 0$
t-Student Generalizada	$-\frac{1}{2}(v+n)/(\gamma+u)$	$v, \gamma > 0$

### 3.7.1 Função escore

Estatisticamente, a função escore possui uma considerável importância no que se refere à inferência, como na aproximação do erro em uma estimativa obtida pelo método da máxima verossimilhança e no processo computacional durante as estimativas realizadas pelo método. Além disso, pode-se também recorrer a essa função para formulação de intervalos de confiança. A função escore apresenta os seguintes componentes (MAGNUS e NEUDECKER, 2007):

$$U(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = -2V_{g(u)} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}$$

e

$$U(\boldsymbol{\varphi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}},$$

sendo este último com os seguintes elementos:

$$\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \varphi_j} = -\frac{1}{2} \text{tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right) - V_{g(u)} \boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon},$$

em que  $j = 1, 2, 3$  e  $\boldsymbol{\epsilon} = \mathbf{Z} - \mathbf{X}\boldsymbol{\beta}$ . Para verificar o cálculo das derivadas parciais  $\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j}$ , pode-se consultar o trabalho de Uribe-Opazo, Borssoi e Galea (2012).

Considerando-se as distribuições da família de contornos elípticos e que o parâmetro  $\boldsymbol{\beta}$  tem forma fechada, então a estimativa de  $\boldsymbol{\beta}$  é dada por (DE BASTIANI *et al.*, 2015):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{Z},$$

em que  $\hat{\boldsymbol{\Sigma}} = \boldsymbol{\Sigma}(\hat{\boldsymbol{\varphi}})$ .

Já o parâmetro  $\boldsymbol{\varphi}$  não tem forma fechada, sendo portanto estimado pela maximização do logaritmo da função de verossimilhança (Equação 3.24).

### 3.7.2 Matriz de informação observada

Objetiva-se com a matriz de informação observada a obtenção da matriz de informação de Fisher, a qual possibilita calcular os erros padrões das estimativas dos parâmetros.

Define-se a matriz de informação observada por  $\mathbf{I}(\boldsymbol{\theta}) = -\mathbf{L}(\boldsymbol{\theta})$ , a qual deve ser avaliada considerando  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , em que  $\hat{\boldsymbol{\theta}} = (\hat{\boldsymbol{\beta}}^T, \hat{\boldsymbol{\varphi}}^T)^T$ . Sendo que (URIBE-OPAZO, BORSSOI e GALEA,

2012):

$$\mathbf{L}(\boldsymbol{\theta}) = \mathbf{L}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{L}_{\beta\beta} & \mathbf{L}_{\beta\varphi} \\ \mathbf{L}_{\varphi\beta} & \mathbf{L}_{\varphi\varphi} \end{pmatrix}. \quad (3.25)$$

Considerando-se a diferenciação de matrizes para modelos espaciais lineares com distribuição de probabilidade elíptica, tem-se o seguinte (DE BASTIANI, 2012):

$$\mathbf{L}_{\beta\beta} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = 2\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \left( V_{g(u)} \boldsymbol{\Sigma} + 2V'_{g(u)} \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \right) \boldsymbol{\Sigma}^{-1} \mathbf{X}. \quad (3.26)$$

$$\mathbf{L}_{\beta\varphi} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}^T}, \quad (3.27)$$

em que:

$$\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \varphi_j} = 2\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \left( V'_{g(u)} \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \boldsymbol{\Sigma} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} + V_{g(u)} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}, \quad j = 1, 2, 3.$$

$$\mathbf{L}_{\varphi\beta} = \mathbf{L}_{\beta\varphi}^T. \quad (3.28)$$

$$\mathbf{L}_{\varphi\varphi} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T}, \quad (3.29)$$

sendo que:

$$\begin{aligned} \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \varphi_i \partial \varphi_j} &= \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} \right) \right] + \boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1} \left[ V'_{g(u)} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon} \boldsymbol{\epsilon}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} + \right. \\ &\quad \left. V_{g(u)} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} + \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\epsilon}, \quad i, j = 1, 2, 3. \end{aligned}$$

### 3.7.3 Matriz de informação de Fisher e erro padrão das estimativas dos parâmetros

Os erros padrões das estimativas dos parâmetros podem ser obtidos pelo cálculo da matriz de informação de Fisher  $[\mathbf{K}(\boldsymbol{\theta})]$ , que é a esperança matemática da matriz de informação observada (Equação 3.25), e é dada na forma bloco diagonal, do seguinte modo (DE BASTIANI *et al.*, 2015):

$$\mathbf{K}(\boldsymbol{\theta}) = E[\mathbf{I}(\boldsymbol{\theta})] = \begin{pmatrix} E[-\mathbf{L}_{\beta\beta}] & E[-\mathbf{L}_{\beta\varphi}] \\ E[-\mathbf{L}_{\varphi\beta}] & E[-\mathbf{L}_{\varphi\varphi}] \end{pmatrix} = \begin{pmatrix} \mathbf{K}(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\boldsymbol{\varphi}) \end{pmatrix}, \quad (3.30)$$

donde:

$$\mathbf{K}(\boldsymbol{\beta}) = \frac{4d_g}{n} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}; \quad (3.31)$$

e o elemento  $k_{ij}$  da matriz  $\mathbf{K}(\boldsymbol{\varphi})$  é dado por:

$$k_{ij}(\boldsymbol{\varphi}) = \frac{b_{ij}}{4} \left( \frac{4f_g}{n(n+2)} - 1 \right) + \frac{2f_g}{n(n+2)} \text{tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right),$$

em que  $n$  é o número de observações realizadas na amostragem,  $d_g = E[UV_{g(U)}^2]$ ,  $f_g = E[U^2V_{g(U)}^2]$ , sendo  $U = \|\mathbf{Z}\|^2$ ,  $\mathbf{Z} \sim El_n(\mathbf{0}, \mathbf{I}_n; g(u))$ ,  $b_{ij} = \text{tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right) \text{tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right)$ , em que  $i, j = 1, 2, 3$  e  $\text{tr}(\cdot)$  é o operador traço.

Observa-se que tanto  $d_g$  quanto  $f_g$  estão em função de  $V_{g(U)}$  que, por sua vez, depende de qual distribuição da família de contornos elípticos o vetor aleatório tem. Sendo assim, serão

apresentadas as expressões de  $d_g$  e  $f_g$  (Tabela 3) para as distribuições normal e t-Student, as quais possuem formas fechadas para o cálculo dessas expressões (DE BASTIANI *et al.*, 2015).

Tabela 3 – Expressões de  $d_g$  e  $f_g$  para as distribuições Normal e t-Student.

Distribuição	$d_g$	$f_g$
Normal	$\frac{n}{4}$	$\frac{n(n+2)}{4}$
t-Student	$\frac{n}{4} \frac{v+n}{v+n+2}$	$\frac{n(n+2)}{4} \frac{v+n}{v+n+2}$

### 3.8 Métodos de estimação de parâmetros no ajuste de modelos à função semivariância

Ajustar um modelo ao semivariograma experimental não é um processo automático. É comum na Geoestatística a utilização do método empírico, em que um modelo é ajustado visualmente aos pontos do semivariograma experimental até que se obtenha o mais adequado. Entretanto, tal ajuste é bastante subjetivo e depende da experiência do pesquisador (MELLO *et al.*, 2005). Sendo assim, opta-se por métodos em que se possa utilizar um critério que permita escolher o melhor modelo ajustado ao semivariograma experimental (Seção 3.9), e para tal, é necessário, primeiramente, estimar os parâmetros de dependência espacial,  $\varphi = (\varphi_1, \varphi_2, \varphi_3)^T$ , para cada modelo. Os métodos de estimação são divididos em dois grupos: aquele em que o ajuste dos modelos é realizado com base no semivariograma experimental (Mínimos Quadrados Ordinários e Ponderados) e aquele em que baseia-se o ajuste diretamente nos dados (Máxima Verossimilhança). Neste trabalho, as estimações dos parâmetros de dependência espacial serão realizadas por este último método.

#### 3.8.1 Método de Máxima Verossimilhança

A ideia do método de Máxima Verossimilhança (ML) é obter, a partir de uma amostra, o estimador “mais verossímil” dos parâmetros de um certo modelo probabilístico. Logo, esse procedimento é baseado nos dados e não no semivariograma experimental.

Seja uma amostra aleatória  $(z_1, \dots, z_n)$  proveniente de uma população com uma função densidade de probabilidade  $f(\mathbf{z}, \theta)$ , que depende do vetor de parâmetros  $\theta$ , tem uma função densidade de probabilidade (f.d.p.) conjunta dada por (MAGALHÃES, 2015):

$$\prod_{i=1}^n f(z_i, \theta),$$

em que  $\theta$  é um parâmetro fixo e  $z_i$  é uma variável aleatória.

Antes da retirada da amostra, cada observação é uma variável aleatória cuja f.d.p. é igual à função densidade da população. Entretanto, uma vez extraída a amostra,  $z_i$  passa a ser fixo, enquanto o vetor de parâmetros  $\theta$  torna-se variável. Para dada amostra  $(z_1, \dots, z_n)$ , a função densidade de probabilidade conjunta, que considera um vetor de parâmetros desconhecidos  $\theta$ , é denominada função de verossimilhança. Objetiva-se estimar o vetor  $\theta$  que maximize esta função, fazendo com que a probabilidade de se obter a amostra já retirada seja a maior possível (MARDIA e MARSHALL, 1984).

Assim, com o intuito de estimar o vetor de parâmetros desconhecidos  $\theta = (\beta^T, \varphi^T)^T$  do modelo espacial linear (Equações 3.2 e 3.4), pode-se utilizar o método de máxima verossimilhança. Para que se obtenham os parâmetros estimados por esse método, é necessário considerar a distribuição de probabilidade da variável sob estudo.

Suponha inicialmente que o vetor aleatório  $\mathbf{Z}$  tem distribuição normal multivariada, isto é,  $\mathbf{Z} \sim N_n(\mathbf{X}\beta, \Sigma)$ , em que  $\mu = \mathbf{X}\beta$  e  $\Sigma$  é a matriz escala simétrica e definida positiva dada pela Equação 3.4.

A função de verossimilhança é dada por (URIBE-OPAZO, BORSSOI e GALEA, 2012):

$$L(\theta) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left[ -\frac{1}{2} (\mathbf{Z} - \mathbf{X}\beta)^T \Sigma^{-1} (\mathbf{Z} - \mathbf{X}\beta) \right].$$

Devido à monotonicidade da função logaritmo, tem-se que maximizar  $L(\theta)$  é equivalente a maximizar o logaritmo da função verossimilhança. Portanto, estima-se o vetor de parâmetros de dependência espacial ( $\theta$ ) pelo método ML a partir da seguinte função:

$$l(\theta) = \log[L(\theta)] = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log |\Sigma| - \frac{1}{2} (\mathbf{Z} - \mathbf{X}\beta)^T \Sigma^{-1} (\mathbf{Z} - \mathbf{X}\beta). \quad (3.32)$$

Pode-se perceber que do ponto de vista matemático, para se implementar o procedimento da máxima verossimilhança para maximizar a função  $l(\theta)$  (Equação 3.32), deve-se igualar a zero as derivadas parciais dessa função com respeito a cada parâmetro do vetor  $\theta$  e resolver o sistema de equações obtido. Todavia, na prática, frequentemente não é possível obter uma solução analítica considerando o método ML. Dessa forma, surge a necessidade de utilizar algum critério de otimização numérica, tal como Newton-Raphson, pelo qual obtêm-se os parâmetros de interesse considerando-se os processos iterativos.

O método de Máxima Verossimilhança nem sempre apresenta solução analítica, e mais, o estimador de ML pode não existir ou não estar definido para algum processo estocástico  $\mathbf{Z}$ , além de não ser necessariamente único (OLIVEIRA, SOUZA e MELO, 2010). Apesar disso, o método ML é constantemente utilizado na literatura devido às boas propriedades dos estimadores. Considerando-se um número suficientemente grande de amostras, os estimadores de máxima verossimilhança são, em geral, consistentes, apresentam distribuição assintótica normal, além de serem assintoticamente eficientes e possuírem a propriedade da invariância (MARDIA e MARSHALL, 1984).

Supondo agora que o vetor aleatório  $\mathbf{Z}$  tem distribuição t-Student,  $\mathbf{Z} \sim t_n(\mu, \Sigma, \nu)$ , então a função de verossimilhança considera um vetor de dados completos, e obtêm-se a função de verossimilhança completa (ASSUMPÇÃO, URIBE-OPAZO e GALEA, 2014). Para maximização dessa função e conseqüente estimação do parâmetro  $\theta = (\beta^T, \varphi^T)^T$ , é necessário fazer considerar um processo iterativo, no qual o grau de liberdade  $\nu$  é fixado. No presente trabalho foi utilizado o algoritmo EM (Expectation Maximization) para maximizar a função de verossimilhança completa e obter o estimador do parâmetro  $\theta$  (ASSUMPÇÃO, URIBE-OPAZO e GALEA, 2014); (SCHEMMER *et al.*, 2017).

### 3.9 Critérios de seleção do melhor modelo teórico ajustado

O ajuste de um modelo teórico à função semivariância visa depreender se ele é o melhor possível. Para tal, serão apresentados os critérios de Validação Cruzada, Akaike e Traço.

Na aplicação da Validação Cruzada (VC) para modelos espaciais lineares ajustados a dados com distribuição normal de probabilidade, o método de estimação é testado nos locais amostrados. O valor da amostra em certa localização é temporariamente descartado do conjunto de dados amostrais, em seguida o valor na mesma localização é estimado por krigagem usando as amostras restantes. Uma vez que a estimação é calculada, pode-se compará-la ao valor da amostra que foi inicialmente removida do conjunto de dados. O procedimento é repetido para todas as amostras disponíveis, e assim permite comparar os valores previstos com os amostrados (ISAAKS e SRIVASTAVA, 1989).

Na Validação Cruzada, consideram-se os seguintes critérios para avaliação dos modelos (FARACO *et al.*, 2008):

O Erro Médio (EMD), o Desvio Padrão do Erro Médio (DPM) e o Erro Absoluto são dados, respectivamente, por:

$$EMD = \frac{1}{n} \sum_{i=1}^n \left( Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i) \right), \quad (3.33)$$

$$DPM = \sqrt{\frac{1}{n} \sum_{i=1}^n |Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i)|}, \quad (3.34)$$

$$EA = \sum_{i=1}^n |Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i)|, \quad (3.35)$$

em que  $n$  é o número de observações realizadas;  $Z(\mathbf{s}_i)$  é o valor da variável aleatória observada no ponto  $\mathbf{s}_i$ ; e  $\hat{Z}(\mathbf{s}_i)$  é o valor de  $Z$  no ponto  $\mathbf{s}_i$  predito por krigagem, sem considerar a observação  $Z(\mathbf{s}_i)$ .

O Erro Médio Reduzido (EMR), obtido pela seguinte equação:

$$EMR = \frac{1}{n} \sum_{i=1}^n \frac{(Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i))}{\sigma(\hat{Z}(\mathbf{s}_i))}, \quad (3.36)$$

em que  $\sigma(\hat{Z}(\mathbf{s}_i))$  é o desvio padrão de  $\hat{Z}(\mathbf{s}_i)$ , que é o valor de  $Z$  no ponto  $\mathbf{s}_i$  predito por krigagem, sem considerar a observação  $Z(\mathbf{s}_i)$ .

O Desvio Padrão do Erro Médio Reduzido (DPMR), obtido a partir da seguinte equação:

$$DPMR = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i)|}{\sigma(\hat{Z}(\mathbf{s}_i))}}. \quad (3.37)$$

Considerando-se a condição de estacionariedade, o valor populacional do EMR é zero e do DPMR é igual a 1. Logo, deseja-se que o EMD, o EMR e o DPM sejam mais próximos de zero possível; o DPMR deve ser adjacente a 1 e o EA apresentar o menor valor. O modelo que exibe os melhores resultados na maior parte dos critérios é o que apresenta o melhor ajuste (FARACO *et al.*, 2008). Entretanto, eventualmente, dois ou mais modelos podem “empatar” ao se considerar a Validação Cruzada, desse modo, torna-se necessário utilizar outro critério de seleção para decidir, de fato, qual o melhor modelo ajustado.

O critério de informação de Akaike (AKAIKE, 1974), o qual baseia-se no teorema de mesmo nome, parte do pressuposto que se dois ou mais modelos apresentam resultados igualmente satisfatórios, então pode-se esperar um melhor desempenho do modelo mais simples na predição de

novos dados, pois este apresenta um número menor de parâmetros a serem estimados e, portanto, está menos suscetível a erros. Considerando-se o princípio da parcimônia, o critério de Akaike (AIC) impõe uma penalidade para a complexidade do modelo e busca aquele que apresente bom ajuste com o mínimo de parâmetros possíveis. O critério de Akaike é calculado da seguinte forma:

$$AIC = -2l(\boldsymbol{\theta}) + 2k,$$

em que:  $l$  representa o valor da função de máxima verossimilhança (Equação 3.32) e  $k$  é o número de parâmetros a serem estimados. Quanto menor o valor do AIC, melhor o modelo.

O fato do critério de Akaike estar vinculado à função de máxima verossimilhança, implica que o mesmo é influenciado pelo número de observações realizadas ( $n$ ). Quando se tem  $\frac{n}{k} < 40$ , Bozdogan (1987) sugere uma correção para o AIC, denominado critério de Akaike de Segunda Ordem (CAIC), cujo valor tende para AIC conforme  $n$  cresce. O CAIC é obtido da seguinte forma:

$$CAIC = -2l + 2k + \frac{2k(k+1)}{n-k-1},$$

e escolhe-se o modelo que apresentar o menor valor de CAIC.

Supondo-se agora o vetor aleatório  $\mathbf{Z}$  segue uma distribuição t-Student multivariada, De Bastiani *et al.* (2015) e Schemmer *et al.* (2017) sugerem a utilização da validação cruzada (Equação 3.38) para selecionar o grau de liberdade  $v$  do modelo espacial linear com distribuição t-Student.

$$VC(v) = \frac{1}{n} \left[ \sum_{i=1}^n \left( \frac{z(s_i) - \hat{z}_i(s_i)}{h_{ii}} \right)^2 \right], \quad (3.38)$$

em que:  $\hat{z}_i(s_i) = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}_i$ , donde  $x_i^T = (x_{i1}, \dots, x_{ip})$  é a  $i$ -ésima linha da matriz de delineamento  $\mathbf{X}$  e representa a predição na localização  $s_i$  sem considerar a observação nessa posição;  $\hat{\boldsymbol{\beta}}_i$  é o estimador de máxima verossimilhança para  $\boldsymbol{\beta}$ , sem considerar a  $i$ -ésima observação;  $h_{ii}$  é o  $i$ -ésimo elemento da diagonal da matriz projeção  $\mathbf{H} = \mathbf{X}(\mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1}$ , para  $i = 1, \dots, n$ .

De Bastiani *et al.* (2015) apresentam o método do Traço, proposto por Kano, Berkane e Bentler (1993), utilizado para seleção do melhor valor para o grau de liberdade. Esse método consiste em calcular o traço da matriz de covariância assintótica da média estimada ( $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ ), da seguinte forma:

$$TR(v) = \left( \frac{v+n+2}{v+n} \right) tr[(\mathbf{X}^T \mathbf{X})(\mathbf{X}^T \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X})^{-1}], \quad (3.39)$$

em que  $tr(\cdot)$  é o operador traço.

Em ambos os métodos, o melhor valor para o grau de liberdade ( $v$ ) é determinado pelos menores valores da Validação Cruzada (Equação 3.38) e do Traço (Equação 3.39).

### 3.10 Krigagem

A krigagem é uma técnica de interpolação utilizada para estimar valores de determinado atributo sob estudo em locais não amostrados, a partir de valores amostrais vizinhos, para predição espacial.

A krigagem distingue-se de outros métodos de interpolação pela maneira como os pesos são atribuídos às diferentes amostras. No caso de interpolação linear simples, por exemplo, os pesos

são todos iguais a  $1/n$  (em que  $n$  é o número de amostras). Outro exemplo é a interpolação baseada no inverso do quadrado das distâncias, na qual os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observados. Já na krigagem, os pesos são determinados a partir de uma análise espacial, a qual baseia-se na função semivariância estimada (CIGAGNA *et al.*, 2015).

A partir de um rol de valores da variável aleatória  $Z$  mensurados em  $n$  pontos, pode-se realizar uma estimativa a partir de uma combinação linear ponderada dos  $n$  valores observados, na qual associa-se a cada valor medido um peso  $\lambda_i$  (ISAACS e SRIVASTAVA, 1989). Com respeito ao número de observações a serem usadas nesta estimativa, os autores sugerem que se o fenômeno apresenta grande continuidade, então apenas os pontos mais próximos são suficientes. Por outro lado, quanto mais descontínuo for o processo sob estudo, mais as estimativas devem incluir dados distantes.

Webster e Oliver (2007) apresentam vários métodos de krigagem para realização da predição espacial, dentre eles está a krigagem ordinária, que é a mais utilizada na prática.

### 3.10.1 Krigagem ordinária

A interpolação por krigagem ordinária é realizada utilizando-se o seguinte estimador (WEBSTER e OLIVER, 2007):

$$\hat{Z}(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad (3.40)$$

onde:  $\hat{Z}(\mathbf{s}_0)$  é o valor estimado da variável georreferenciada no local  $\mathbf{s}_0$  não amostrado,  $n$  é o número de observações realizadas,  $\lambda_i$  representa o peso associado a cada valor  $Z(\mathbf{s}_i)$  mensurado,  $i = 1, \dots, n$ .

O estimador definido na Equação (3.40) é tido como BLUE (*Best Linear Unbiased Estimator*), ou seja, é o melhor estimador por fornecer predições não-viesadas e com variâncias mínimas. Tais qualidades são garantidas tendo em vista que os pesos  $\lambda_i$  são determinados impondo-se que a esperança do erro seja zero (não-viesada) (Equação 3.41) e a variância do erro seja mínima (Equação 3.43) (ISAACS e SRIVASTAVA, 1978).

#### **Não-viesada:**

Assegurar a condição de não-viesamento, é garantir que (SOARES, 2014):

$$E[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 0. \quad (3.41)$$

Ou ainda, da Equação 3.40:

$$E\left[\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right] - E[Z(\mathbf{s}_0)] = 0,$$

sendo que a soma dos ponderadores deve ser unitária, isto é (WEBSTER e OLIVER, 2007):

$$\sum_{i=1}^n \lambda_i = 1. \quad (3.42)$$

### Variância Mínima:

No caso do estimador da krigagem, dizer que este apresenta variância mínima, é verificar que (SOARES, 2014):

$$Var[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)] = E[(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0))^2] - \{E[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)]\}^2. \quad (3.43)$$

Na expressão 3.43, o primeiro termo corresponde ao erro quadrático médio e o segundo termo é referente ao viés. Como o preditor da krigagem é não-viesado (Equação 3.41), este segundo termo é zero. Logo, vem que:

$$\begin{aligned} Var[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)] &= E\left[\left(\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0)\right)^2\right] \\ &= E\left[\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Z(\mathbf{s}_i) Z(\mathbf{s}_j)\right] - 2E\left[\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right] + E[(Z(\mathbf{s}_0))^2] \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] - 2 \sum_{i=1}^n \lambda_i Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_0)] + Var[Z(\mathbf{s}_0)], \end{aligned}$$

sendo que a última expressão é equivalente a:

$$\sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] - 2 \sum_{i=1}^n \lambda_i Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_0)] + Var[Z(\mathbf{s}_0)] - 2\alpha \left(1 - \sum_{i=1}^n \lambda_i\right),$$

pois, da igualdade 3.42, tem-se que  $2\alpha \left(1 - \sum_{i=1}^n \lambda_i\right) = 0$ , donde  $\alpha$  é o parâmetro de Lagrange.

Isaaks e Srivastava (1989) complementam que:

$$Var[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 2 \sum_{i=1}^n \lambda_i \gamma(\mathbf{s}_{i0}) - \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \gamma(\mathbf{s}_{ij}),$$

em que  $s_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ .

É extremamente difícil minimizar esta última expressão, logo, a utilização do parâmetro de Lagrange  $\alpha$  transforma o problema de minimização com restrição em um problema sem restrição. Dessa forma, os pesos  $\lambda_i$  são obtidos a partir do seguinte sistema de equações chamado de sistema de krigagem ordinária e dado por (ISAACS e SRIVASTAVA, 1989):

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma_{ij} + \alpha = \gamma_{i0} \\ \sum_{i=1}^n \lambda_i = 1, \end{cases}$$

em que  $\gamma_{ij} = \gamma(\mathbf{s}_{ij})$ , sendo  $s_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ .

A minimização da variância (Equação 3.43) constitui-se em calcular as  $n + 1$  derivadas parciais em relação a  $\lambda_1, \dots, \lambda_n, \alpha$  e igualar a zero. Dessa forma, obtém-se um sistema com  $n + 1$  equações e  $n + 1$  incógnitas, cuja solução resulta nos  $n$  pesos  $\lambda_i$ , tal que a soma desses é 1 e minimizam a variância da estimação (SOARES, 2014).

As estimativas dos pesos  $\lambda_i$  possibilita prever o valor da variável georreferenciada na localização não amostrada  $\mathbf{s}_0$

Os cálculos das derivadas parciais e o sistema de equações obtido com elas podem ser encontrados em Soares (2014).

O sistema de krigagem ordinária pode ser escrito na forma matricial por:

$$\gamma\lambda = \mathbf{A},$$

isto é, apresenta a seguinte solução:

$$\lambda = \gamma^{-1}\mathbf{A},$$

em que:

$$\gamma = \begin{pmatrix} \gamma(\mathbf{s}_{11}) & \gamma(\mathbf{s}_{12}) & \dots & \gamma(\mathbf{s}_{1n}) & 1 \\ \gamma(\mathbf{s}_{21}) & \gamma(\mathbf{s}_{22}) & \dots & \gamma(\mathbf{s}_{2n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(\mathbf{s}_{n1}) & \gamma(\mathbf{s}_{n2}) & \dots & \gamma(\mathbf{s}_{nn}) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{pmatrix}, \mathbf{A} = \begin{pmatrix} \gamma(\mathbf{s}_{10}) \\ \gamma(\mathbf{s}_{20}) \\ \vdots \\ \gamma(\mathbf{s}_{n0}) \\ 1 \end{pmatrix} \text{ e } \lambda = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \alpha \end{pmatrix}.$$

A krigagem possibilita, além de realizar predições do valor da variável regionalizada em um ponto, associar a essa predição uma medida de confiabilidade, denominada variância da krigagem e dada por (CRESSIE, 2015):

$$\sigma_0^2 = \sum_{i=1}^n \lambda_i \gamma_{i0} + \alpha. \quad (3.44)$$

Assim como seu interpolador, a variância da krigagem depende somente das coordenadas dos pontos amostrados e das semivariâncias. Logo, a variância será máxima em locais mais longínquos de valores medidos em relação ao local de predição.

### 3.11 Tamanho amostral efetivo

É necessário observar o valor das variáveis georreferenciadas em alguns pontos da área para que se possa obter informações sobre essas. Isto é, deve-se realizar uma amostragem das variáveis de interesse, tendo em vista que é impossível ou inviável estudar toda a população (WANG *et al.*, 2012). Entretanto, ao considerar  $n$  pontos para serem amostrados, alguns deles podem, eventualmente, estar altamente correlacionados entre si, de modo a fornecer informações repetidas. Sendo assim, qual seria o novo tamanho da amostra considerando-se os efeitos da autocorrelação espacial? Para responder a tal questionamento, será apresentada na sequência uma teoria denominada tamanho amostral efetivo - ESS (em inglês, *Effective Sample Size*), a partir das abordagens de Griffith (2005) e Vallejos e Osorio (2014). Em ambas metodologias, o intuito é utilizar as observações da variável coletadas nos  $n$  pontos amostrados para obter o valor estimado do ESS, e empregar o novo número de amostras para estimar a média amostral dessa variável da maneira mais precisa possível.

#### 3.11.1 Tamanho amostral efetivo univariado

O problema geral abordado por essa teoria é: dada uma grade amostral com  $n$  pontos em um espaço  $r$ -dimensional, qual o tamanho amostral efetivo (ESS) associado a esses pontos? Segundo Vallejos e Osorio (2014), se as observações são independentes e uma média local é estimada, então o ESS é  $n$ . Caso haja dependência espacial entre as amostras, o ESS deve ser

menor que  $n$ . Por outro lado, dada uma autocorrelação espacial perfeita e positiva entre as amostras, intuitivamente, o ESS deveria ser 1.

Considere o vetor das observações da variável aleatória  $Z$  nos  $n$  pontos amostrados como  $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$ , em que cada  $Z(\mathbf{s}_i)$  é obtido conforme o modelo expressado na Equação 3.2. A princípio, será assumido um processo estacionário, isto é, com média constante, logo  $E[\mathbf{Z}] = \mu \mathbf{1}$ , em que  $\mu \in \mathbb{R}$  e  $\mathbf{1}$  é um vetor unitário  $n \times 1$ .

Seja  $\Sigma$  a matriz escala  $n \times n$  (Equação 3.4), que está em função de  $\boldsymbol{\theta} = (\mu \mathbf{1}, \boldsymbol{\varphi}^T)^T$  tal que  $\boldsymbol{\theta} \in \mathbb{R}^q$  é o  $q$ -ésimo vetor de parâmetros desconhecidos. Considerando  $q = 3$ , então  $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$ , em que  $\varphi_1$  é o efeito pepita,  $\varphi_2$  é a contribuição e  $\varphi_3$  é o parâmetro da função que define o alcance.

Como  $\mathbf{Z}$  apresenta distribuição multivariada com segundo momento finito, isto é, com média constante  $\mu$ , então o tamanho amostral efetivo pode ser caracterizado pela matriz de correlação  $\mathbf{R}(\boldsymbol{\varphi}) = [(\sigma_{ij} / \sqrt{\sigma_{ii}\sigma_{jj}})]$ , para  $i, j = 1, \dots, n$ . Ou ainda, reescreve-se a partir da notação matricial,  $\mathbf{R}(\boldsymbol{\varphi}) = \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Sigma}(\boldsymbol{\varphi}) \boldsymbol{\Upsilon}^{-1}$ , em que  $\boldsymbol{\Upsilon} = \text{diag}(\sigma_{11}^{1/2}, \dots, \sigma_{nn}^{1/2})$  (VALLEJOS e OSORIO, 2014).

Antes de definir a expressão para quantificar o tamanho amostral efetivo univariado, é necessário frisar que se considera um processo estocástico estacionário e sem presença de covariáveis, logo a matriz de delineamento é unitária, isto é,  $\mathbf{X} = \mathbf{1}$ .

A seguir, será formalizado o conceito de tamanho amostral efetivo univariado (VALLEJOS e OSORIO, 2014):

**Definição 2** *Seja  $\{Z(\mathbf{s}_i); \mathbf{s}_i \in S \subset \mathbb{R}^r\}$  um campo aleatório tal que para  $D = \{\mathbf{s}_1, \dots, \mathbf{s}_n\} \subset S$ , o vetor aleatório  $\mathbf{Z} = [Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)]^T \sim N(\mu \mathbf{1}, \mathbf{R}(\boldsymbol{\varphi}))$ , em que  $\mathbf{R}$  é uma matriz inversível. Então, o tamanho amostral efetivo de  $\mathbf{Z}$  é dado por:*

$$\text{ESS} = \text{ESS}(n, \mathbf{R}(\boldsymbol{\varphi}, D), r) = \mathbf{1}^T \mathbf{R}(\boldsymbol{\varphi})^{-1} \mathbf{1}. \quad (3.45)$$

É possível verificar que a expressão que define o ESS (Equação 3.45) decorre da matriz de informação de Fisher (Equação 3.30), considerando o caso particular em que  $\mathbf{Z}$  apresenta distribuição Normal.

De fato, quando se considera uma distribuição normal, tem-se que  $\mathbf{X} = \mathbf{1}$ ,  $d_g = \frac{n}{4}$  (Tabela 3), posto que a matriz bloco diagonal para a média, a qual compõe a matriz de informação de Fisher, é dada pela Equação 3.31, então tem-se que:

$$\mathbf{K}(\boldsymbol{\beta}) = \frac{4}{n} \frac{n}{4} \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1} = \mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}, \quad (3.46)$$

sendo que, neste trabalho,  $\boldsymbol{\beta} = \mu$ , dado que não será considerado covariáveis.

Sabe-se ainda que  $\mathbf{R}(\boldsymbol{\varphi}) = \boldsymbol{\Upsilon}^{-1} \boldsymbol{\Sigma}(\boldsymbol{\varphi}) \boldsymbol{\Upsilon}^{-1}$ , em que  $\mathbf{R}$  e  $\boldsymbol{\Sigma}$  são, respectivamente, as matrizes de correlação e covariância, e  $\boldsymbol{\Upsilon} = \text{diag}(\sigma_{11}^{1/2}, \dots, \sigma_{nn}^{1/2})$ . Logo, para que as matrizes de correlação e covariância sejam iguais, conseqüentemente implicando na igualdade entre as Equações 3.45 e 3.46, é necessário que  $\boldsymbol{\Upsilon} = \mathbf{I}$ , ou seja, que a variância de cada  $n$ -ésima amostra seja unitária,  $\sigma_{11} = \dots = \sigma_{nn} = 1$ .

Portanto, as Equações 3.45 e 3.46 são iguais somente para o caso particular supracitado. Caso contrário, a Equação 3.45 provém da Equação 3.46, pois a matriz de correlação é obtida a partir de elementos da matriz de covariância.

Considere  $\{Z(\mathbf{s}); \mathbf{s} \in S \subset \mathbb{R}^r\}$  um processo estocástico estacionário, em que  $Z$  é a variável aleatória que se deseja estudar. Considere o erro estocástico  $\{\epsilon(\mathbf{s}); \mathbf{s} \in S \subset \mathbb{R}^r\}$ , de modo que  $Cov(\epsilon(\mathbf{h}), \epsilon(\mathbf{h} + \mathbf{s})) = \sigma(\mathbf{s}, \varphi)$ , sendo que  $\mathbf{s}, \mathbf{h} \in S$ . Tem-se então o modelo dado por  $\mathbf{Z} = \mu \mathbf{1} + \epsilon$ , em que  $\mu \in \mathbb{R}$  e  $\epsilon \sim N_n(\mathbf{0}, \Sigma(\theta))$ , este modelo espacial é o linear gaussiano (Equação 3.3) quando processo estocástico é estacionário, de modo que  $\mathbf{X}\beta = \mu \mathbf{1}$ .

Na sequência, será apresentada uma estimativa para o tamanho amostral efetivo univariado proposta por Vallejos e Osorio (2014), considerando-se, para tal, o método de máxima verossimilhança (ML). Tem-se que o parâmetro desconhecido  $\varphi$  é estimado de modo que se possa obter o tamanho amostral efetivo substituindo-se o valor de  $\varphi$  na matriz de correlação. Este parâmetro será estimado pelo método ML como definido na Equação 3.32. Pode-se então reescrever a expressão do tamanho amostral efetivo univariado (Definição 2) considerando sua estimativa:

$$\widehat{\text{ESS}} = \mathbf{1}^T \mathbf{R}(\hat{\varphi}_{ML})^{-1} \mathbf{1}, \quad (3.47)$$

em que:  $\mathbf{1}$  é um vetor unitário, de dimensão  $n \times 1$ , sendo  $n$  o número de pontos amostrais, e  $\mathbf{R}(\hat{\varphi}_{ML})$ , ou simplesmente  $\mathbf{R}(\hat{\varphi})$ , é a matriz de correlação espacial dos pontos amostrais, de dimensão  $n \times n$ , cujos elementos  $R(\hat{\varphi})_{ij}$ ,  $i, j = 1, \dots, n$ , são dados por (VALLEJOS e OSORIO, 2014):

$$R(\hat{\varphi})_{ij} = \begin{cases} 1, & \text{se } i = j \\ \frac{\hat{\varphi}_2 \rho(h)}{\hat{\varphi}_1 + \hat{\varphi}_2}, & \text{se } i \neq j, \end{cases} \quad (3.48)$$

em que:  $R(\hat{\varphi})_{ij}$ ,  $i, j = 1, \dots, n$ , é o valor estimado da correlação espacial existente entre o  $i$ -ésimo e o  $j$ -ésimo ponto amostral; e a expressão para calcular a função de correlação ( $\rho(h)$ ) depende do modelo geoestatístico utilizado, em que  $h$  é a distância euclidiana entre a  $i$ -ésima e a  $j$ -ésima observação. Tem-se ainda que  $\hat{\varphi}_2$  e  $\hat{\varphi}_1 + \hat{\varphi}_2$  são os valores estimados dos parâmetros contribuição e patamar, nesta ordem.

Pode-se observar que uma das suposições consideradas acima é que o erro estocástico tem distribuição normal de probabilidades. Entretanto, pode-se estender a definição do ESS tendo em vista o caso em que o erro apresenta uma distribuição elíptica. Para tal, Anderson *et al.* (1986) sugerem que a estimativa do parâmetro  $\varphi$  seja obtida da seguinte forma:

$$\tilde{\varphi}_{ML} = \frac{n}{2u_g} \hat{\varphi}_{ML},$$

em que:  $u_g$  é o máximo da função  $h(u) = u^{\frac{n}{4}} g(u)$ , lembrando que  $u$  é a distância de Mahalanobis e  $g(u)$  é a função geradora de densidades;  $n$  é o número de pontos amostrais; e  $\hat{\varphi}_{ML}$  é o valor estimado dos parâmetros que definem a estrutura de dependência espacial pelo método de máxima verossimilhança (Equação 3.32).

Vallejos e Osorio (2014) descrevem algumas propriedades do tamanho amostral efetivo. No que segue, consideram-se  $\mathbf{Z} \sim N_n(\mu \mathbf{1}, \mathbf{R}(\varphi))$  e um espaço  $r$ -dimensional fixado:

- (i) ESS é crescente em  $n$ , em que  $n$  é o número de pontos amostrais da grade;
- (ii)  $\text{ESS} \geq 1$ ;
- (iii)  $\text{ESS} = n$ , quando não há correlação espacial entre as observações;

- (iv)  $ESS = 1$ , quando tem-se uma perfeita correlação espacial entre todos os possíveis pares de observações;
- (v)  $ESS$  cresce conforme  $n$  aumenta;
- (vi)  $ESS$  diminui à medida que a correlação aumenta.

É natural pensar que o valor do tamanho amostral efetivo dependa da disposição espacial dos pontos observados, dado que a estimação do  $ESS$  está em função da matriz de correlação amostral (VALLEJOS e OSORIO, 2014). Logo, os autores identificaram que o valor do  $ESS$  também depende dos parâmetros que caracterizam a estrutura de dependência espacial, os quais estão vinculados ao modelo teórico ajustado.

Vallejos e Osorio (2014) realizaram simulações com o intuito de explorar o comportamento do  $ESS$  para diferentes modelos teóricos (esférico, exponencial, gaussiano e família Mátern com valores do parâmetro de forma  $\kappa = 0, 5; 1; 1, 5; 2, 5$ ). A partir dessas simulações, constatou-se que o  $ESS$  aumentou conforme aumento do parâmetro efeito pepita. Outro fato observado foi que em modelos nos quais verificou-se efeito pepita puro, o  $ESS$  foi igual ao número total de amostras ( $n$ ). O que era previsível, já que o efeito pepita puro reflete a ausência de dependência espacial, logo não há correlação espacial entre as amostras e, necessariamente, o tamanho efetivo da amostra deve ser igual a  $n$ . Além disso, outra constatação realizada pelos autores foi que o  $ESS$  diminuiu conforme o raio de dependência espacial aumentou, como já era esperado.

A definição de tamanho amostral efetivo univariado pode ainda ser generalizada considerando-se que o vetor  $\mathbf{Z}$  tem distribuição de probabilidade diferente da distribuição normal (VALLEJOS e OSORIO, 2014).

**Definição 3** Suponha que  $\mathbf{Z} \sim El_n(\mu\mathbf{1}, \mathbf{R}(\varphi), g(u))$ . O tamanho amostral efetivo ( $ESS_e$ ) associado a  $\mathbf{Z}$  é definido por:

$$ESS_e = 4E[||\mathbf{Y}||^2 V_g^2(||\mathbf{Y}||^2)] \frac{1}{n} \mathbf{1}^T \mathbf{R}(\varphi)^{-1} \mathbf{1} = \tau ESS, \quad (3.49)$$

em que  $V_{g(u)} = g'(u)/g(u)$ ,  $\mathbf{Y} = \mathbf{R}(\varphi)^{-1/2}(\mathbf{Z} - \mu\mathbf{1})$  e  $n$  é o número inicial de pontos amostrais.

A definição de  $ESS_e$  também decorre da matriz informação de Fisher, considerando-se para tal a distribuição da família de contornos elípticos, isto é, a Equação 3.49 é uma generalização da Equação 3.45.

Na sequência, será exibido o índice  $\tau$  apresentado na Definição 3 para as distribuições t-Student (LANGE, LITTLE e TAYLOR, 1989), a qual faz parte da família de contornos elípticos.

**Definição 4** Suponha que  $\mathbf{Z} \sim t_n(\mu\mathbf{1}, \mathbf{R}(\varphi), v)$ ,  $v > 2$ . Então, o tamanho amostral efetivo associado a  $\mathbf{Z}$  é dado por:

$$ESS_t = \frac{v + n}{v + n + 2} ESS. \quad (3.50)$$

Da Equação 3.50 pode-se observar que  $ESS_t \leq ESS$ . E mais: do item (ii) das propriedades do  $ESS$  acima explicitadas, para  $v > 2$  e  $n \geq 1$ , tem-se que  $ESS_t \geq \frac{3}{2}$ .

Outra metodologia existente para o cálculo do tamanho amostral efetivo univariado foi proposta por Griffith (2005). Neste método, suponha  $\mathbf{V}$  uma matriz  $n \times n$  que contém a estrutura de

covariância das  $n$  amostras georreferenciadas, tal que  $\mathbf{V}$  é uma matriz simétrica e positiva definida, associada ao vetor aleatório  $\boldsymbol{\epsilon} \sim N_n(\mathbf{0}, \sigma^2 \mathbf{V}^{-1})$ . Assim, a matriz de covariância de  $\boldsymbol{\epsilon}$  é dada por  $\boldsymbol{\Sigma} = \sigma^2 \mathbf{V}^{-1}$ .

E considere que:

$$\boldsymbol{\epsilon} = \mathbf{V}^{-1/2} \boldsymbol{\epsilon}^*, \quad (3.51)$$

Da regressão linear múltipla, sejam  $\mathbf{U} \sim N_n(\boldsymbol{\mu}_U, \mathbf{V}_U)$  e  $\mathbf{Y} = \mathbf{B}\mathbf{U}$ .

Donde vem que  $\mathbf{Y} \sim N_n(\boldsymbol{\mu}_Y, \mathbf{V}_Y)$ , sendo  $\boldsymbol{\mu}_Y = \mathbf{B}\boldsymbol{\mu}_U$  e  $\mathbf{V}_Y = \mathbf{B}\mathbf{V}_U\mathbf{B}^T$ .

Considerando as seguintes relações:  $\mathbf{Y} = \boldsymbol{\epsilon}$ ,  $\mathbf{V}^{1/2} = \mathbf{B}$ ,  $\mathbf{U} = \boldsymbol{\epsilon}^*$ ,  $\mathbf{V}_U = \sigma_{\epsilon^*}^2 \mathbf{I}$ ,  $\mathbf{V}_Y = \sigma^2 \mathbf{V}^{-1}$ , nota-se que  $\boldsymbol{\epsilon}^* \sim N_n(\mathbf{0}, \sigma_{\epsilon^*}^2 \mathbf{I})$ .

E mais: tem-se que  $\boldsymbol{\epsilon}$  e  $\boldsymbol{\epsilon}^*$  representam os vetores de erros correlacionados e não-correlacionados, nesta ordem, sendo que  $\sigma^2 = \sigma_{\epsilon^*}^2$ .

Como apresentado anteriormente (Equação 3.2), como o modelo espacial linear é dado por  $\mathbf{Z} = \boldsymbol{\mu} + \boldsymbol{\epsilon}$ . Tem-se que:

$$\mathbf{Z} = \boldsymbol{\mu} + \boldsymbol{\epsilon} = \boldsymbol{\mu} + \mathbf{V}^{-1/2} \boldsymbol{\epsilon}^*, \quad (3.52)$$

em que  $\mathbf{Z}$  é o vetor  $n \times 1$  dos valores do atributo sob estudo,  $\boldsymbol{\mu}$  representa a média populacional do atributo,  $\boldsymbol{\epsilon}$  e  $\boldsymbol{\epsilon}^*$  indicam, respectivamente, os  $n \times 1$  vetores dos erros correlacionados e não-correlacionados espacialmente.

Segundo Rencher e Schaalje (2008), o estimador da variância populacional obtido por máxima verossimilhança do atributo  $Z$ , vigorando a autocorrelação espacial (isto é,  $\boldsymbol{\Sigma} = Cov(\mathbf{Z}) = \sigma^2 \mathbf{I}_n$ ), é dado por:

$$\hat{\sigma}_Z^2 = \frac{(\mathbf{Z} - \boldsymbol{\mu}\mathbf{1})^T (\mathbf{Z} - \boldsymbol{\mu}\mathbf{1})}{n}.$$

E mais, a esperança desse estimador é dada por (GRIFFITH, 1987):

$$E[\hat{\sigma}_Z^2] = E[(\mathbf{Z} - \boldsymbol{\mu}\mathbf{1})^T (\mathbf{Z} - \boldsymbol{\mu}\mathbf{1})/n] = \sigma^2 \frac{TR(\mathbf{V}^{-1})}{n}, \quad (3.53)$$

em que  $TR(\cdot)$  é o operador traço. Segundo este autor,  $\frac{TR(\mathbf{V}^{-1})}{n}$  na Equação 3.53 representa o fator de inflação da variância (VIF).

Considerando-se que  $\sigma^2 = \sigma_{\epsilon^*}^2$ , tem-se que

$$E[\hat{\sigma}_Z^2] = \sigma_{\epsilon^*}^2 \frac{TR(\mathbf{V}^{-1})}{n}, \quad (3.54)$$

que é a notação matricial da esperança da variância estimada do atributo  $Z$ , ignorando a autocorrelação espacial.

Além disso, a variância da média amostral do atributo  $Z$ , também desconsiderando a autocorrelação espacial, é dada por:

$$Var(\bar{Z}) = \frac{\sigma_{\epsilon^*}^2 \mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}}{n^2}. \quad (3.55)$$

Após algumas manipulações algébricas com a Equação 3.55, tem-se que:

$$\begin{aligned}
 Var(\bar{Z}) &= \frac{\sigma_{\epsilon^*}^2 \mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}}{n^2} \\
 &= \frac{\sigma_{\epsilon^*}^2}{n} \frac{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}}{n} \\
 &= \frac{\frac{\sigma_{\epsilon^*}^2}{n}}{\frac{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}}{n}} \frac{TR(\mathbf{V}^{-1})}{TR(\mathbf{V}^{-1})} \\
 &= \frac{TR(\mathbf{V}^{-1}) \sigma_{\epsilon^*}^2}{\frac{TR(\mathbf{V}^{-1})}{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}} n} \\
 &= \frac{E[\hat{\sigma}_Z^2]}{\frac{TR(\mathbf{V}^{-1})}{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}} n},
 \end{aligned}$$

em que o denominador desta equação é a expressão definida por Griffith (2005) para calcular o tamanho amostral efetivo univariado, isto é:

$$ESS = \frac{TR(\mathbf{V}^{-1})}{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}} n, \quad (3.56)$$

em que:  $\mathbf{1}$  é um vetor unitário,  $\mathbf{V}^{-1} = \Sigma$ ,  $n$  é o número de pontos amostrais iniciais e  $TR(\cdot)$  é o operador traço.

A expressão 3.56 é para se obter o valor do ESS univariado, ou seja, quando se tem somente um atributo  $Z$  sob estudo.

### 3.11.2 Tamanho amostral efetivo multivariado

Na metodologia sugerida por Griffith (2005), é possível generalizar o conceito de tamanho amostral efetivo para uma situação multivariada, ou seja, envolvendo  $p$  atributos, em que  $p \geq 2$ . O caso bivariado é desenvolvido no Apêndice A (4.1).

O valor do ESS multivariado é dado por:

$$ESS_p = \frac{\mathbf{1}_{(1)}^T \mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d \mathbf{1}_{(1)}}{TR(\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d)} \frac{TR[\{(\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}\}^T (\mathbf{R} \otimes \mathbf{I}) \{(\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}\}]}{\mathbf{1}_{(2)}^T \{(\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}\}^T (\mathbf{R} \otimes \mathbf{I}) \{(\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}\} \mathbf{1}_{(2)}} n, \quad (3.57)$$

em que:  $TR(\cdot)$  é o operador traço;  $\otimes$  representa o produto de Kronecker;  $\mathbf{A}_d$ ,  $p \times p$ , é uma matriz diagonal contendo o coeficiente da combinação linear  $a_p$ ;  $\mathbf{1}_{(1)}$  é um vetor unitário  $p \times 1$ ;  $\mathbf{1}_{(2)}$  é um vetor unitário  $np \times 1$ ;  $\Phi_d$ ,  $p \times p$  é uma matriz diagonal contendo o desvio padrão da  $j$ -ésima variável  $\sigma_j$ ,  $j = 1, \dots, p$ ;  $\mathbf{V}_d$ ,  $np \times np$  é uma matriz bloco diagonal contendo as matrizes  $\mathbf{V}_p^{-1}$ ,  $n \times n$ , sendo estas as matrizes escala da cada  $j$ -ésima variável;  $\mathbf{R}$ ,  $p \times p$ , é a matriz de correlação entre os  $p$  atributos; e  $\mathbf{I}$ ,  $n \times n$ , é a matriz identidade.

Vale ressaltar que o valor obtido em  $ESS_p$  deve estar contido no conjunto de valores obtido para o ESS de cada atributo  $p$ , isto é, o ESS multivariado está contido no conjunto formado pelos ESS univariados.

Na sequência será verificado que, considerando  $p = 2$ , a Equação 3.57 se reduz à Equação 4.10 (que se encontra desenvolvida no Apêndice A).

De fato, tem-se que:

$$\mathbf{A}_d \Phi_d = \begin{pmatrix} \omega & 0 \\ 0 & 1 - \omega \end{pmatrix} \begin{pmatrix} \sigma_X & 0 \\ 0 & \sigma_Y \end{pmatrix} = \begin{pmatrix} \omega \sigma_X & 0 \\ 0 & (1 - \omega) \sigma_Y \end{pmatrix}. \quad (3.58)$$

$$\begin{aligned}
\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d &= \begin{pmatrix} \omega \sigma_X & 0 \\ 0 & (1-\omega) \sigma_Y \end{pmatrix} \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix} \begin{pmatrix} \sigma_X \omega & 0 \\ 0 & \sigma_Y (1-\omega) \end{pmatrix} \\
&= \begin{pmatrix} \omega^2 \sigma_X^2 & \omega(1-\omega) \sigma_Y \sigma_X \rho_{XY} \\ \omega(1-\omega) \sigma_X \sigma_Y \rho_{XY} & (1-\omega)^2 \sigma_Y^2 \end{pmatrix}. \tag{3.59}
\end{aligned}$$

Aplicando o operador traço na Equação 3.59, vem que:

$$\text{TR}(\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d) = \omega^2 \sigma_X^2 + (1-\omega)^2 \sigma_Y^2. \tag{3.60}$$

$$\mathbf{1}^T \mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d \mathbf{1} = \omega^2 \sigma_X^2 + (1-\omega)^2 \sigma_Y^2 + 2\omega(1-\omega) \rho_{XY} \sigma_X \sigma_Y. \tag{3.61}$$

$$\begin{aligned}
((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2})^T &= \left[ \begin{pmatrix} \omega \sigma_X & \mathbf{0} \\ \mathbf{0} & (1-\omega) \sigma_Y \end{pmatrix} \otimes \mathbf{I} \right] \begin{pmatrix} (\mathbf{V}_X^{-1/2})^T & \mathbf{0} \\ \mathbf{0} & (\mathbf{V}_Y^{-1/2})^T \end{pmatrix} \\
&= \begin{pmatrix} \omega \sigma_X (\mathbf{V}_X^{-1/2})^T & \mathbf{0} \\ \mathbf{0} & \omega \sigma_Y (\mathbf{V}_Y^{-1/2})^T \end{pmatrix}, \tag{3.62}
\end{aligned}$$

em que  $\mathbf{0}$  é uma matriz  $n \times n$ .

$$\mathbf{R} \otimes \mathbf{I} = \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix} \otimes \mathbf{I} = \begin{pmatrix} \mathbf{I} & \rho_{XY} \mathbf{I} \\ \rho_{XY} \mathbf{I} & \mathbf{I} \end{pmatrix}. \tag{3.63}$$

Das expressões 3.62 e 3.63, tem-se:

$$\begin{aligned}
&((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}) = \\
&= \begin{pmatrix} \omega \sigma_X (\mathbf{V}_X^{-1/2})^T & \mathbf{0} \\ \mathbf{0} & \omega \sigma_Y (\mathbf{V}_Y^{-1/2})^T \end{pmatrix} \begin{pmatrix} \mathbf{I} & \rho_{XY} \mathbf{I} \\ \rho_{XY} \mathbf{I} & \mathbf{I} \end{pmatrix} \times \begin{pmatrix} \omega \sigma_X (\mathbf{V}_X^{-1/2}) & \mathbf{0} \\ \mathbf{0} & \omega \sigma_Y (\mathbf{V}_Y^{-1/2}) \end{pmatrix} \\
&= \begin{pmatrix} \omega^2 \sigma_X^2 (\mathbf{V}_X^{-1}) & \omega(1-\omega) \sigma_X \sigma_Y \rho_{XY} (\mathbf{V}_X^{-1/2})^T (\mathbf{V}_Y^{-1/2}) \\ \omega(1-\omega) \sigma_X \sigma_Y \rho_{XY} (\mathbf{V}_X^{-1/2})^T (\mathbf{V}_Y^{-1/2}) & (1-\omega)^2 \sigma_Y^2 (\mathbf{V}_Y^{-1}) \end{pmatrix}. \tag{3.64}
\end{aligned}$$

Da expressão 3.64, vem que:

$$\begin{aligned}
\mathbf{1}^T ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}) \mathbf{1} &= \omega^2 \sigma_X^2 \mathbf{1}^T \mathbf{V}_X^{-1} \mathbf{1} + (1-\omega)^2 \sigma_Y^2 \mathbf{1}^T \mathbf{V}_Y^{-1} \mathbf{1} \\
&+ 2\omega(1-\omega) \rho_{XY} \sigma_X \sigma_Y \mathbf{1}^T (\mathbf{V}_X^{-1/2})^T \mathbf{V}_Y^{-1/2} \mathbf{1}. \tag{3.65}
\end{aligned}$$

Ao se aplicar o operador traço na Equação 3.64, tem-se:

$$\text{TR}[(\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2})] = \omega^2 \sigma_X^2 \text{TR}(\mathbf{V}_X^{-1}) + (1-\omega)^2 \sigma_Y^2 \text{TR}(\mathbf{V}_Y^{-1}). \tag{3.66}$$

Quando as expressões 3.60, 3.61, 3.65 e 3.66 são substituídas na Equação 3.57, verifica-se que, de fato, se obtém a Equação 4.10.

Vallejos e Osorio (2014) também apresentam uma proposta para o cálculo do tamanho amostral efetivo multivariado, tendo como base a metodologia de Griffith (2005).

O valor do ESS multivariado é dado pela seguinte expressão:

$$ESS_p = \frac{\mathbf{1}_{(1)}^T ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{W}_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{W}_d^{1/2}) \mathbf{1}_{(1)}}{\mathbf{1}_{(2)}^T \mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d \mathbf{1}_{(2)}}, \quad (3.67)$$

em que:  $\otimes$  representa o produto de Kronecker;  $\mathbf{R}$ , de dimensão  $p \times p$ , em que  $p$  é o número de variáveis consideradas, é a matriz de correlação linear entre as  $p$  variáveis;  $\mathbf{A}_d$ , de dimensão  $p \times p$ , é uma matriz diagonal contendo o coeficiente dos  $p$  autovetores padronizados da matriz de correlação linear;  $\mathbf{1}_{(1)}$  é um vetor unitário de dimensão  $p \times 1$ ;  $\mathbf{1}_{(2)}$  é um vetor unitário de dimensão  $np \times 1$ ;  $\Phi_d$ , de dimensão  $p \times p$  é uma matriz diagonal contendo o desvio padrão da  $m$ -ésima variável  $\sigma_m$ ,  $m = 1, \dots, p$ ;  $\mathbf{I}$ , de dimensão  $n \times n$ , é a matriz identidade; e  $\mathbf{W}_d$ , de dimensão  $np \times np$ , é uma matriz bloco diagonal contendo as matrizes  $\mathbf{R}_m(\varphi) = [r_{ij}(\varphi)]$ , de ordem  $n \times n$ , em que  $\mathbf{R}_p(\varphi)$  é a matriz de correlação espacial de cada  $m$ -ésima variável,  $m = 1, \dots, p$  (Equação 3.48).

### 3.12 Referências

AKAIKE, H. A new look at statistical model identification. **IEEE Transactions on Automatic Control**, v. 19, p. 716-723, 1974.

ANDERSON, T. W.; HSU, H; FANG, K-T. Maximum-likelihood estimates and likelihood-ratio criteria for multivariate elliptically contoured distributions. **Canadian Journal of Statistics**, v. 14, n. 1, p. 55-59, 1986.

ANDRIOTTI, J. L. S. **Fundamentos de Estatística e Geoestatística**. Porto Alegre: UNISINOS, 2013. 165p.

ASSUMPTÇÃO, R. A. B.; URIBE-OPAZO, M. A.; GALEA, M. Analysis of local influence in geostatistics using Student's t-distribution. **Journal of Applied Statistics**, v. 41, n. 11, p. 2323-2341, 2014.

AZZALINI, A.; CAPITANIO, A. Statistical applications of the multivariate skew normal distribution. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, v. 61, n. 3, p. 579-602, 1999.

BERNARDI, A. C. C.; NAIME, J. M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. Geoestatística aplicada a Agricultura de Precisão. In: GREGO, C. R.; OLIVEIRA, R. P.; VIEIRA, S. R. **Agricultura de Precisão: resultados de um novo olhar**. São Carlos: EMBRAPA, 2014a. cap. 5, p. 74-83.

BERNARDI, A. C. C.; NAIME, J. M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. Estudo da variabilidade espacial dos atributos químicos e físicos do solo em um agroecossistema de arroz irrigado para suporte à Agricultura de Precisão. In: CRUZ, L. E. C; FILIPPINI, J. M. A.; PARFITT, J. M.

B.; PILLON, C. N. **Agricultura de Precisão**: resultados de um novo olhar. São Carlos: EMBRAPA, 2014b. cap. 15, p. 209-216.

BOX, G. E. P; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 211-252, 1964.

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

BUSSAB, W. O.; MORETTIN, P. A. **Estatística Básica**. 8. ed. São Paulo: Saraiva, 2013. 548p.

CAMARGO, E. C. G. Geoestatística: fundamentos e aplicações. In: CÂMARA, G.; MEDEIROS, J.S. **Geoprocessamento para Projetos Ambientais**. São José dos Campos: INPE, 1998.

CAMBARDELLA, C. A.; MOORMAN, T.; PARKIN, T.; KARLEN, D.; NOVAK, J.; TURCO, R.; KONOPKA, A. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, v. 58, n. 5, p. 1501-1511, 1994.

CIGAGNA, C.; BONOTTO, D. M., STURARO, J. R.; CAMARGO, F. M. Geostatistical techniques applied to mapping limnological variables and quantify the uncertainty associated with estimates. **Acta Limnologica Brasiliensia**, v. 27, n. 4, p. 421-430, 2015.

CLARK, I. **Practical Geostatistics**. London: Applied Science Publishers, 1979. 141p.

CRESSIE, N. A. C. **Statistics for Spatial Data**. rev. ed. New York: John Wiley & Sons, 2015. 928p.

CYSNEIROS, F. J. A.; PAULA, G. A.; GALEA, M. Modelos simétricos aplicados. **Escola de Modelos de Regressão**, v. 9, 2005.

DE BASTIANI, F. **Influência local em modelos espaciais lineares com distribuição da família de contornos elípticos**. 2012. 137f. Dissertação (Mestrado em Estatística) - Universidade Federal de Pernambuco, Recife, 2012.

DE MELLO, J. M.; BATISTA, J. L. F.; RIBEIRO JR, P. J.; OLIVEIRA, M. S. Adjustment and selection of spatial models of semivariogram envisaging eucalyptus grandis volumetric estimates. **Scientia Forestalis**, n. 69, p. 23-37, 2005.

DE BASTIANI, F.; CYSNEIROS, A. F. J.; CYSNEIROS, A. H. M.; URIBE-OPAZO, M. A.; GALEA, M. Influence diagnostics in elliptical spatial linear models. **Test**, v. 24, n. 2, p. 322-340, 2015.

DIGGLE, P. J.; RIBEIRO JR, P. J. **Model-based Geostatistics**. New York: Springer Series in Statistics, 2007. 230p.

FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A.; JOHANN, J. A.; BORSSOI, J. A. Selection criteria of spatial variability models used in thematical maps of soil physical attributes and soybean yield. **Revista Brasileira de Ciência do Solo**, v. 32, n. 2, p. 463-476, 2008.

FONTELLES, M. J.; SIMÕES, M. G.; ALMEIDA, J. C.; FONTELLES, R. G. S. Metodologia da pesquisa científica: diretrizes para a elaboração de um protocolo de pesquisa. **Revista Paraense de Medicina**, v. 23, n. 3, p. 1-8, 2009.

GRIFFITH, D. A. **Spatial Autocorrelation: a primer**. Resource publication in geography. Association of American Geographers, Washington-DC, 1987.

GRIFFITH, D. A. Effective geographic sample size in the presence of spatial autocorrelation. **Annals of the Association of American Geographers**, v. 95, n. 4, p. 740-760, 2005.

GUERRA, P. A. G. **Geoestatística Operacional**. Brasília: Ministério de Minas e Energia, Departamento de Produção Mineral, 1988. 145p.

ISAAKS, E.; SRIVASTAVA, R. M. **Applied Geoestistics**. New York: Oxford University Press, 1989. 561p.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining Geoestistics**. London: Academic Press, 1978. 600p.

KANO, Y.; BERKANE, M.; BENTLER, P. M. Statistical inference based on pseudo–maximum likelihood estimators in elliptical populations. **Journal of the American Statistical Association**, v. 88, n. 421, p. 135-143, 1993.

KESTRING, F.; GUEDES, L. P. C.; DE BASTIANI, F.; URIBE–OPAZO, M. A. Thematic maps comparison of different sampling grids for soybean productivity. **Engenharia Agrícola**, v. 35, n. 4, p. 733-743, 2015.

LANDIM, P. M. B. Sobre geoestatística e mapas. **Terrae Didatica**, v. 2, n. 1, p. 19-33, 2006.

LANDIM, P. M. B.; YAMAMOTO, J. K. **Geoestatística: conceitos e aplicações**. São Paulo: Oficina de Textos, 2013. 215p.

LANGE, K. L.; LITTLE, R. J. A.; TAYLOR, J. M. G. Robust statistical modeling using the t distribution. **Journal of the American Statistical Association**, v. 84, n. 408, p. 881-896, 1989.

MAGALHÃES, M. N. **Probabilidade e Variáveis Aleatórias**. 3ed. São Paulo: EDUSP, 2015. 410p.

MAGNUS, J. R.; NEUDECKER, H. **Matrix differential calculus with applications in statistics and econometrics**. 3rd ed. Chichester: John Wiley & Sons, 2007. 450p.

MARDIA, K. V.; MARSHALL, R. J. Maximum likelihood estimation of models for residual covariance in spatial regression. **Biometrika**, v. 71, n. 1, p. 135-146, 1984.

MIOT, H. A. Tamanho da amostra em estudos clínicos e experimentais. **J. Vasc. Bras.**, v. 10, n. 4, p. 275-8, 2011.

MOLIN, J. P. **Agricultura de Precisão**: o gerenciamento da variabilidade. Piracicaba: Oficina de Textos, 2001. 83p.

OLIVEIRA JR, J. C.; PAULA SOUZA, L. C.; FREITAS MELO, V. Variabilidade de atributos físicos e químicos de solos da Formação Guabirotuba em diferentes unidades de amostragem. **Revista Brasileira de Ciência do Solo**, v. 34, n. 5, 2010.

OSORIO, F.; PAULA, G. A.; GALEA, M.. Assessment of local influence in elliptical linear models with longitudinal structure. **Computational Statistics & Data Analysis**, v. 51, n. 9, p. 4354-4368, 2007.

RENCHER, A. C.; SCHAALJE, G. B. **Linear Models in Statistics**. New York: John Wiley & Sons, 2008. 578p.

RODRIGUES, M. S.; RAMOS, R. R. D.; AZEVEDO, T. P.; PATROCÍNIO FILHO, A. P.; OLIVEIRA, L. G. Variabilidade espacial da resistência do solo à penetração em área capineira irrigada no semiárido. **Agropecuária Científica no Semiárido**, v. 10, n. 1, p. 161-166, 2014.

SCHEMMER, R. C.; URIBE-OPAZO, M. A.; GALEA, M.; ASSUMPÇÃO, R. A. B. Spatial variability of soybean yield through a reparametrized t-Student model. **Engenharia Agrícola**, v. 37, n. 4, p. 760-770, 2017.

SEIDEL, E. J.; OLIVEIRA, M. S. Novo índice geoestatístico para a mensuração da dependência espacial. **Revista Brasileira de Ciência do Solo**, v. 38, n. 3, 2014.

SOARES, A. **Geoestatística para Ciências da Terra e do Ambiente**. 3. ed. Lisboa: Press, 2014. 214p.

TEIXEIRA, M. B. R.; SCALON, J. D. Comparação entre estimadores de semivariância. **Rev. Bras. Biom.**, v. 31, n. 2, p. 248-269, 2013.

URIBE-OPAZO, M. A.; BORSSOI, J. A.; GALEA, M. Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics**, v. 39, n. 3, p. 615-630, 2012.

VALLEJOS, R.; OSORIO, F. Effective sample size of spatial process models. **Spatial Statistics**, v. 9, p. 66-92, 2014.

VIEIRA, S. R.. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.;

ALVAREZ, V.; SCHAEFER, C. E. G. R. **Tópicos em Ciência do Solo**. Viçosa: Sociedade Brasileira de Ciência do Solo, v. 1, p. 1-54, 2000.

WANG, J-F; STEIN, A; GAO, B-B; GE, Y. A review of spatial sampling. **Spatial Statistics**, v. 2, p. 1-14, 2012.

WEBSTER, R.; OLIVER, M. A. **Statistical Methods in Soil and Land Resource Survey**. Oxford: Oxford University Press, 1990. 316p.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for Environmental Scientists**. Chichester: John Wiley & Sons, 2007. 330p.

## 4 Apêndice

### 4.1 Apêndice A. Tamanho amostral efetivo bivariado

Na sequência, serão apresentadas algumas expressões referentes ao caso em que se têm dois atributos  $(X, Y)$  sob estudo, considerando para tal o modelo autoregressivo, as quais compõe a expressão para o cálculo do tamanho amostral efetivo quando se emprega modelos espaciais (GRIFFITH, 2005).

A partir da correlação entre os atributos  $X$  e  $Y$ ,  $\rho_{XY}$ , tem-se que (BUSSAB e MORETTIN, 2013):

$$\begin{aligned}
 \rho_{XY} &= \frac{E[(X - E[X])(Y - E[Y])]}{\sigma_X \sigma_Y} \\
 &= \frac{abE[(X - E[X])(Y - E[Y])]}{ab\sigma_X \sigma_Y} \\
 &= \frac{E[a(X - E[X])b(Y - E[Y])]}{a\sigma_X b\sigma_Y} \\
 &= \frac{E[(aX - aE[X])(bY - bE[Y])]}{a\sigma_X b\sigma_Y} \\
 &= \frac{Cov(aX, bY)}{a\sigma_X b\sigma_Y} = \rho_{aXbY}, \tag{4.1}
 \end{aligned}$$

em que  $\sigma_X$  e  $\sigma_Y$  denotam, respetivamente, as variâncias dos atributos  $X$  e  $Y$ , e  $a$  e  $b$  são constantes.

Logo,

$$\rho_{XY} = \rho_{aXbY} = \frac{Cov(aX, bY)}{a\sigma_X b\sigma_Y},$$

donde,

$$Cov(aX, bY) = ab\rho_{XY}\sigma_X\sigma_Y.$$

Ou ainda, considerando duas amostras aleatórias  $X_1, \dots, X_n$  e  $Y_1, \dots, Y_n$ , de modo que  $\bar{X} = \sum_{i=1}^n \frac{X_i}{n}$  e  $\bar{Y} = \sum_{i=1}^n \frac{Y_i}{n}$ , tem-se que:

$$Cov(a\bar{X}, b\bar{Y}) = \frac{ab\rho_{XY}\sigma_X\sigma_Y}{n}. \tag{4.2}$$

Tem-se ainda que a expressão abaixo representa a variância da soma das médias dos atributos  $X$  e  $Y$  (BUSSAB e MORETTIN, 2013):

$$\begin{aligned}
 Var(a\bar{X} + b\bar{Y}) &= Var(a\bar{X}) + Var(b\bar{Y}) + 2Cov(a\bar{X}, b\bar{Y}) \\
 &= a^2 \frac{\sigma_X^2}{n} + b^2 \frac{\sigma_Y^2}{n} + 2Cov(a\mu_X, b\mu_Y). \tag{4.3}
 \end{aligned}$$

E quando se substitui a Equação 4.2 na Equação 4.3, observa-se que:

$$Var(a\bar{X} + b\bar{Y}) = a^2 \frac{\sigma_X^2}{n} + b^2 \frac{\sigma_Y^2}{n} + 2ab\rho_{XY} \frac{\sigma_X\sigma_Y}{n}. \tag{4.4}$$

Considere  $a = \omega$  e  $b = 1 - \omega$ , em que  $\omega$  ( $0 \leq \omega \leq 1$ ) é o peso aplicado na média do atributo  $X$ , então a Equação 4.4 pode ser reescrita como:

$$\frac{\omega^2 \sigma_X^2 + (1 - \omega)^2 \sigma_Y^2 + 2\omega(1 - \omega)\rho_{XY}\sigma_X\sigma_Y}{n}. \tag{4.5}$$

Segundo Griffith (2005), a seguinte expressão envolve o fator de inflação da variância (VIF) de  $X$  e  $Y$ , sendo que as matrizes  $\mathbf{V}_X$  e  $\mathbf{V}_Y$  têm dimensão  $n \times n$ :

$$\frac{\omega^2 \sigma_X^2 \text{TR}(\mathbf{V}_X^{-1}) + (1 - \omega)^2 \sigma_Y^2 \text{TR}(\mathbf{V}_Y^{-1})}{n}, \quad (4.6)$$

ou ainda, quando se assume o modelo autorregressivo:

$$\frac{\omega^2 \sigma_X^2 \text{TR}\{[(\mathbf{I} - \rho_X \mathbf{W})^T (\mathbf{I} - \rho_X \mathbf{W})]^{-1}\} + (1 - \omega)^2 \sigma_Y^2 \text{TR}\{[(\mathbf{I} - \rho_Y \mathbf{W})^T (\mathbf{I} - \rho_Y \mathbf{W})]^{-1}\}}{n}$$

A próxima expressão representa a variabilidade da média amostral dos atributos  $X$  e  $Y$ , na presença de autocorrelação espacial não-nula:

$$\begin{aligned} & \omega^2 \sigma_X^2 \mathbf{1}^T [(\mathbf{I} - \rho_X \mathbf{W})^T (\mathbf{I} - \rho_X \mathbf{W})]^{-1} \mathbf{1} + (1 - \omega)^2 \sigma_Y^2 \mathbf{1}^T [(\mathbf{I} - \rho_Y \mathbf{W})^T (\mathbf{I} - \rho_Y \mathbf{W})]^{-1} \mathbf{1} + \\ & 2\omega(1 - \omega) \rho_{XY} \sigma_X \sigma_Y \mathbf{1}^T [(\mathbf{I} - \rho_X \mathbf{W})^T (\mathbf{I} - \rho_Y \mathbf{W})]^{-1} \mathbf{1}. \end{aligned} \quad (4.7)$$

Segundo Griffith (2005), após realizar uma análise mais rigorosa, percebeu-se que não era correto escrever o fator de inflação da variância (VIF) cruzado, isto é, considerando os atributos  $X$  e  $Y$  ao mesmo tempo, como feito na Equação 4.7. A partir da relação  $\mathbf{V}^{-1} = \mathbf{V}^{-1/2} (\mathbf{V}^{-1/2})^T$  (RENCHEER e SCHAALJE, 2008), substituiu-se  $[(\mathbf{I} - \rho_X \mathbf{W})^T (\mathbf{I} - \rho_Y \mathbf{W})]^{-1} \mathbf{1} = \mathbf{V}_{XY}^{-1}$  por  $(\mathbf{V}_X^T)^{-1/2} \mathbf{V}_Y^{-1/2}$ , isto é,

$$\omega^2 \sigma_X^2 \mathbf{1}^T \mathbf{V}_X^{-1} \mathbf{1} + \omega^2 \sigma_Y^2 \mathbf{1}^T \mathbf{V}_Y^{-1} \mathbf{1} + 2\omega(1 - \omega) \rho_{XY} \sigma_X \sigma_Y \mathbf{1}^T (\mathbf{V}_X^T)^{-1/2} \mathbf{V}_Y^{-1/2} \mathbf{1}. \quad (4.8)$$

Verificou-se também, após a referida análise, que a expressão (4.5) contém os termos de erro padrão para ambos os atributos, dados por  $\omega^2 \sigma_X^2 / n$  e  $\omega^2 \sigma_Y^2 / n$ , respectivamente. Assim, a Equação (4.5) foi reescrita da seguinte forma:

$$\frac{\omega^2 \sigma_X^2 + (1 - \omega)^2 \sigma_Y^2 + 2\omega(1 - \omega) \rho_{XY} \sigma_X \sigma_Y}{n(\omega^2 \sigma_X^2 / n + \omega^2 \sigma_Y^2 / n)}. \quad (4.9)$$

Tomando-se as expressões (4.6), (4.8) e (4.9) definidas acima, a estimação do tamanho amostral efetivo (ESS) para o caso bivariado e com amostras espacialmente correlacionadas é dada por (GRIFFITH, 2005):

$$\text{ESS}_{XY} = \frac{\frac{\omega^2 \sigma_X^2 + (1 - \omega)^2 \sigma_Y^2 + 2\omega(1 - \omega) \rho_{XY} \sigma_X \sigma_Y}{\omega^2 \sigma_X^2 + (1 - \omega)^2 \sigma_Y^2} [\omega^2 \sigma_X^2 \text{TR}(\mathbf{V}_X^{-1}) + (1 - \omega)^2 \sigma_Y^2 \text{TR}(\mathbf{V}_Y^{-1})]}{\omega^2 \sigma_X^2 \mathbf{1}^T \mathbf{V}_X^{-1} \mathbf{1} + \omega^2 \sigma_Y^2 \mathbf{1}^T \mathbf{V}_Y^{-1} \mathbf{1} + 2\omega(1 - \omega) \rho_{XY} \sigma_X \sigma_Y \mathbf{1}^T (\mathbf{V}_X^T)^{-1/2} \mathbf{V}_Y^{-1/2} \mathbf{1}} n, \quad (4.10)$$

em que  $\mathbf{V}_X$  e  $\mathbf{V}_Y$  representam, respetivamente, a matriz  $n \times n$  que contém a estrutura de covariância de  $X$  e  $Y$ . Se  $\mathbf{V}_X = \mathbf{V}_Y = \mathbf{I}$ , então o  $\text{ESS} = n$ . Caso  $\omega = 0$  ou  $\omega = 1$ , ou ainda  $\mathbf{V}_X = \mathbf{V}_Y$ , então a expressão se reduz à Equação 3.56.

Em outras palavras, o ESS bivariado é uma média ponderada do ESS univariado de cada atributo ( $X$  e  $Y$ ).

# 5 Artigos

## 5.1 ARTIGO 1

### **REDUÇÃO DO TAMANHO AMOSTRAL PELA METODOLOGIA DO TAMANHO AMOSTRAL EFETIVO MULTIVARIADO NO ESTUDO DA VARIABILIDADE ESPACIAL DE VARIÁVEIS GEORREFERENCIADAS COM DISTRIBUIÇÃO NORMAL DE PROBABILIDADE**

#### **Resumo**

Em estudos de variabilidade espacial e mapeamento de áreas agrícolas, o investimento financeiro com coleta e análise laboratorial é um fator crítico que deve ser considerado. Por conseguinte, pesquisas são desenvolvidas com intuito de analisar a viabilidade de reduzir o número de pontos amostrais a serem coletados em uma área agrícola. Uma das possibilidades é considerar os valores da variável georreferenciada em alguns pontos amostrais coletados estejam altamente correlacionados entre si, de modo a fornecer informações redundantes e, conseqüentemente, gastos desnecessários. Os objetivos desse trabalho foram reavaliar o número de pontos amostrais coletados em uma área agrícola e verificar a possibilidade de reduzir o tamanho amostral em experimentos futuros. Foram considerados dados simulados em que se atribuiu às variáveis diferentes valores de efeito pepita e alcance, e também dados práticos cujos atributos físicos e químicos possuíam variados raios e intensidades de dependência espacial. Utilizaram-se duas diferentes propostas para estimar o valor do tamanho amostral efetivo (ESS). Em ambas as metodologias, verificou-se pelos valores estimados do ESS univariado e multivariado redução no tamanho amostral. Constatou-se também que o alcance foi o parâmetro que exerceu maior influência sobre o ESS, sendo que quanto maior o raio de dependência espacial, menor foi o valor estimado do tamanho amostral efetivo, tanto nos estudos práticos quanto nos de simulação.

**Palavras-chave:** agricultura de precisão, dependência espacial, geoestatística.

#### 5.1.1 Introdução

A Agricultura de Precisão pode ser considerada um sistema de gestão que leva em conta a variabilidade espacial do campo e cujo o intuito é obter resultados favoráveis nos âmbitos social, econômico e ambiental (BERNARDI *et al.*, 2014). Entretanto, para o estudo da variabilidade espacial de atributos físico-químicos do solo seja uma realidade, é preciso observar o valor das variáveis em alguns pontos da área agrícola. Para tal, é realizado um planejamento amostral, ponderando-se o tamanho e o formato da região a ser estudada (DIGGLE e RIBEIRO JR, 2007) e as características do relevo (GAO *et al.*, 2016). É de suma importância que a amostragem realizada seja representativa da área, pois a eficiência da caracterização de cada atributo físico-químico do solo na área está diretamente vinculada à qualidade da amostragem realizada (CHERUBIN *et al.*, 2014). Vários trabalhos foram desenvolvidos com intuito de comparar, no âmbito agrícola, diferentes grades amostrais, ou seja, amostragens realizadas com distintas densidades e configurações amostrais, em relação à qualidade dos mapas temáticos obtidos, os quais representam a variabilidade espacial

das variáveis do solo (SOUZA *et al.*, 2014), (CHERUBIN *et al.*, 2014), (KESTRING *et al.*, 2015), (GUEDES *et al.*, 2016).

Haja vista os custos com a coleta e análise laboratorial das amostras serem fatores limitantes em estudos de variabilidade espacial e mapeamento de regiões agrícolas, são propostos, na literatura, métodos para otimizar o esquema de amostragem a partir da redução do número de pontos amostrais coletados, mantendo-se a representatividade da área (RUIZ-CÁRDENAS, FERREIRA e SCHMIDT, 2010), (GUEDES *et al.*, 2011, 2016), (GUEDES, URIBE-OPAZO e RIBEIRO JR, 2014).

Outra metodologia que possibilita reavaliar o número de pontos amostrais que devem ser coletados em uma área agrícola é o tamanho amostral efetivo - ESS (em inglês, *Effective Sample Size*). Essa proposta considera que alguns pontos amostrais coletados podem estar altamente correlacionados entre si, de modo a fornecer informações redundantes e, conseqüentemente, custos de coleta e análise desnecessários (GRIFFITH, 2005), (YANG *et al.*, 2011), (LIU, 2012), (VALLEJOS e OSORIO, 2014), (BARTOSZEK, 2016), (MARTINO, ELVIRA e LOUZADA, 2017). Dessa forma, a metodologia do tamanho amostral efetivo possibilita estimar o novo tamanho da amostra considerando os efeitos da autocorrelação espacial (VALLEJOS e OSORIO, 2014).

O problema abordado pela proposta do ESS, no contexto de estudos sobre a variabilidade espacial de variáveis georreferenciadas, é o seguinte: dada uma grade amostral com  $n$  pontos em um espaço bidimensional, no qual se tenha realizado previamente uma coleta de pontos amostrais, qual é o ESS associado a esses pontos? Se as observações de determinado atributo são independentes e uma média é estimada, então o valor do ESS é igual a  $n$ . Caso haja dependência espacial entre as amostras, então, necessariamente, o valor do ESS é menor que  $n$  (VALLEJOS e OSORIO, 2014). Logo, utilizam-se as observações da variável coletadas nos  $n$  pontos amostrados para se obter o valor estimado do ESS, e se empregar este novo número de amostras para estimar a média amostral dessa variável da maneira mais precisa possível.

Griffith (2005) e Vallejos e Osorio (2014) desenvolveram estudos em dados espaciais reais e simulados, em que o valor do tamanho amostral efetivo foi estimado para cada variável estudada. Entretanto, em estudos agrícolas geralmente se considera um conjunto de dados com várias variáveis georreferenciadas, portanto, é inviável calcular o valor do ESS para cada variável (univariado), pois, na prática, é necessário conhecer o número ideal de pontos amostrais a serem coletados em experimentos futuros considerando todos os atributos físico-químicos do solo, e não o melhor número de pontos amostrais para cada variável. Desta forma, pode-se calcular o valor do ESS multivariado, o qual fornece um único tamanho amostral considerando as características e os efeitos da autocorrelação espacial de todas as variáveis.

Com base nas metodologias multivariadas para a estimação do valor do ESS desenvolvidas, mas não aplicadas, nos trabalhos de Griffith (2005) e Vallejos e Osorio (2014), os objetivos desse trabalho foram determinar a redução no número de pontos amostrais para conjuntos de dados reais e simulados, considerando simultaneamente as informações oriundas de todas as variáveis (caso multivariado), e avaliar a influência que os parâmetros de dependência espacial das variáveis exercem no valor estimado do ESS.

## 5.1.2 Material e Métodos

### 5.1.2.1 Descrição das simulações

Seja um processo estocástico  $\{Y(\mathbf{s}_i), \mathbf{s}_i \in S \subset \mathbb{R}^2\}$ ,  $i = 1, \dots, n$ , estacionário de segunda ordem, em que  $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)$  são os valores observados da variável aleatória sob estudo nas  $\mathbf{s}_i$  localizações espaciais amostradas,  $i = 1, \dots, n$ . Assim,  $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$  é um vetor aleatório, de dimensão  $n \times 1$ , das  $n$  observações da variável aleatória  $Y$ . Suponha que  $\mathbf{Y} \sim N_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , isto é,  $\mathbf{Y}$  tem distribuição normal  $n$ -variada, sendo que  $\boldsymbol{\mu}$ , de dimensão  $n \times 1$ , é o vetor de médias,  $\boldsymbol{\Sigma}$ , de dimensão  $n \times n$ , é a matriz escala.

Cada observação  $Y(\mathbf{s}_i)$  é gerada pelo seguinte modelo linear espacial gaussiano (URIBE-OPAZO, BORSSOI e GALEA, 2012):

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad (5.1)$$

sendo  $\mu(\mathbf{s}_i)$  ( $i = 1, \dots, n$ ) o termo determinístico e  $\epsilon(\mathbf{s}_i)$  o termo estocástico, em que ambos dependem da localização espacial  $\mathbf{s}_i$  em que a variável  $Y$  é observada.

É possível reescrever  $\mathbf{Y}$  matricialmente da seguinte forma (URIBE-OPAZO, BORSSOI e GALEA, 2012):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (5.2)$$

em que:  $\mathbf{X}$ , de ordem  $n \times p$ , é denominada matriz de delineamento, na qual a  $i$ -ésima linha é dada por  $\mathbf{x}_i^T = (x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i))$ , sendo  $p$  o número de covariáveis;  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$  é um vetor de parâmetros desconhecidos a serem estimados, e  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  é o vetor de erros aleatórios, em que  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ .

A matriz escala de  $\boldsymbol{\epsilon}$  é dada por  $\boldsymbol{\Sigma} = [(\sigma_{ij})]$ , em que  $\sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$ ,  $i, j = 1, \dots, n$ , é a covariância entre os valores da variável georreferenciada nos pontos  $\mathbf{s}_i$  e  $\mathbf{s}_j$ . Assume-se que a matriz escala é inversível e pode-se reescrevê-la utilizando a forma paramétrica (DE BASTIANI *et al.*, 2015):

$$\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3), \quad (5.3)$$

sendo que:  $\varphi_1$  e  $\varphi_2$  são respectivamente os parâmetros efeito pepita e contribuição;  $\mathbf{I}_n$  é a matriz identidade de ordem  $n \times n$ ;  $\mathbf{R}(\varphi_3) = [(r_{ij})]$  é uma matriz  $n \times n$  simétrica, em que  $\varphi_3$  é um parâmetro que está em função do alcance ( $a$ ),  $g(\varphi_3) = a$ .

Considere  $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$ , o vetor  $n \times 1$  dos dados simulados, o qual representa uma função de variáveis aleatórias  $Y(\mathbf{s}_i)$ , com  $\mathbf{s}_i \in S \subset \mathbb{R}^2$  e  $i = 1, \dots, n$ . Supõe-se que o vetor de médias do processo estocástico  $\mathbf{Y}$  é estacionário de segunda ordem, isto é, que o valor esperado de cada variável aleatória do vetor  $\mathbf{Y}$  é igual a uma constante  $\mu$ ,  $E[Y(\mathbf{s}_i)] = \mu$ ,  $i = 1, \dots, n$ .

As simulações foram obtidas por um experimento de Monte Carlo para processos isotrópicos e sem tendência direcional (MULLER, 2008), e considerando a configuração amostral de uma área agrícola experimental (Figura 4), limitando-se as coordenadas e o alcance com base nesta área.

Foram simulados sete cenários contendo variáveis às quais se atribuiu diferentes alcances e intensidades de dependência espacial, sendo que algumas delas foram agregadas em mais de um cenário (Tabela 4). Realizaram-se 100 simulações em cada cenário, supondo que as variáveis seguem o modelo linear espacial gaussiano (Equação 5.1). Para cada variável, a média ( $\mu = 5$ ) e a contribuição ( $\varphi_2 = 1$ ) foram fixadas, e considerou-se o ajuste do modelo exponencial.

Tabela 4 – Valores dos parâmetros das variáveis que compõe cada cenário.

Cenários	nº de variáveis	Variáveis	$\varphi_1$	$\varphi_3$	$a$	EPR
1	4	V1	0,10	0,100	0,30	9%
		V2	0,50	0,100	0,30	33%
		V3	0,10	0,170	0,50	9%
		V4	0,50	0,170	0,50	33%
2	4	V1	0,10	0,100	0,30	9%
		V3	0,10	0,170	0,50	9%
		V5	0,50	0,030	0,10	33%
		V6	0,99	0,270	0,80	49%
3	4	V1	0,10	0,100	0,30	9%
		V6	0,99	0,270	0,80	49%
		V7	0,10	0,017	0,05	9%
		V8	0,10	0,023	0,07	9%
4	4	V9	0,03	0,400	1,20	3%
		V10	0,03	0,360	1,10	3%
		V11	0,05	0,330	1,00	5%
		V12	0,07	0,330	1,00	6%
5	4	V7	0,10	0,017	0,05	9%
		V8	0,10	0,023	0,07	9%
		V13	0,10	0,030	0,09	9%
		V14	0,10	0,036	0,11	9%
6	11	V15*	0,01	0,100	0,30	1%
7	15	V16**	0,10	0,030	0,10	9%

$\varphi_1$ : efeito pepita,  $\varphi_3$ : função do alcance,  $a$ : alcance prático (em km),  $EPR = \frac{\varphi_1}{\varphi_1 + \varphi_2}$ : efeito pepita relativo (em %);

\* O cenário 6 é composto ainda pelas variáveis V1 a V6 e V9 a V12;

\*\* O cenário 7 é composto também pelas variáveis V1 a V14.

### 5.1.2.2 Descrição dos dados experimentais

O conjunto de dados reais utilizados nessa pesquisa refere-se ao ano-safra 2015-2016 e pertence ao banco de dados dos projetos desenvolvidos por pesquisadores do grupo de pesquisa do Laboratório de Estatística Espacial (LEE) e do Laboratório de Estatística Aplicada (LEA) da Universidade Estadual do Oeste do Paraná - UNIOESTE, *Campus Cascavel*.

A coleta dos dados que foi efetuada em uma área comercial com 167,35 hectares, localizada na cidade de Cascavel-PR, com coordenadas geográficas aproximadas de latitude 24,95° Sul e 53,37° Oeste, e 650 metros de altitude média. O solo é classificado como Latossolo Vermelho Distrófico, com textura argilosa (EMBRAPA, 2013). O clima da região é identificado como temperado mesotérmico e superúmido, tipo climático Cfa (Koeppen) e a temperatura anual média é de 21°C (APARECIDO *et al.*, 2016).

Nesta área, foi realizada uma amostragem *lattice plus close pairs* composta por 102 pontos amostrais (DIGGLE e LOPHAVEN, 2006), (DIGGLE e RIBEIRO JR, 2007), (GUEDES *et al.*, 2011), em que a distância mínima entre pontos da grade regular desta amostragem foi de 141 m, e a distância máxima de aproximadamente 1800 m. Selecionaram-se aleatoriamente 19 localidades desta amostragem para que fossem adicionados pontos amostrais com distâncias menores de 50 e 75 m entre estes pontos e a observação pertencente à grade regular. As amostras foram georreferenciadas e localizadas com assessoria de um aparelho receptor de sinal com o sistema de posicionamento global (GPS) GEOEXPLORE 3, em um sistema espacial de coordenadas UTM. A

Figura 4 representa a área experimental e a grade amostral utilizada, que apresentam a localização espacial segundo as coordenadas geográficas (X,Y) dos pontos amostrais.

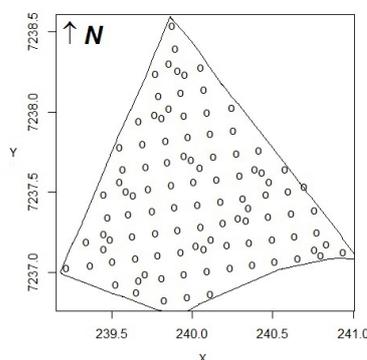


Figura 4 – Mapa da área experimental com os pontos amostrados.

A área foi cultivada com soja e os dados utilizados nesse estudo referem-se à resistência do solo à penetração (MPa) nas profundidades 0-10 cm (RSP 0-10 cm), 11-20 cm (RSP 11-20 cm), 21-30 cm (RSP 21-30 cm) e 31-40 cm (RSP 31-40 cm), além do pH e dos teores dos seguintes macro e micronutrientes do solo: ácido potencial total ( $\text{cmolc dm}^{-3}$ ), alumínio ( $\text{cmolc dm}^{-3}$ ), cálcio ( $\text{cmolc dm}^{-3}$ ), carbono ( $\text{g dm}^{-3}$ ), cobre ( $\text{mg dm}^{-3}$ ), ferro ( $\text{mg dm}^{-3}$ ), fósforo ( $\text{mg dm}^{-3}$ ), magnésio ( $\text{cmolc dm}^{-3}$ ), manganês ( $\text{cmolc dm}^{-3}$ ), potássio ( $\text{cmolc dm}^{-3}$ ) e zinco ( $\text{mg dm}^{-3}$ ).

O conjunto de dados sob estudo foi obtido pela realização de análise química de rotina, obtendo uma amostragem em cada ponto demarcado (Figura 4). Em cada um dos pontos, foram coletadas cinco subamostras de solo, de 0 a 0,2 m de profundidade nas proximidades dos pontos, as quais foram misturadas e colocadas em sacos plásticos, com aproximadamente 500 g, para composição da amostra representativa da parcela. As amostras foram encaminhadas ao laboratório de análise do solo da COODETEC (Cooperativa Central de Pesquisa Agrícola).

As análises exploratória e geoestatística foram realizadas para cada atributo químico e físico do solo. Na análise geoestatística, foram construídos gráficos post-plot, em que os valores mensurados de cada atributo foram classificados de acordo com os quartis. E, para estimar o modelo geoestatístico que define a estrutura de dependência espacial, foi utilizado o método de máxima verossimilhança (URIBE-OPAZO, BORSSOI e GALEA, 2012) com estrutura na matriz escala dos modelos exponencial, gaussiano (CRESSIE, 2015) e família Mátern com valores do parâmetro de forma  $\kappa = 1; 1,5$  e  $2,5$  (MATÉRN, 2013). O modelo que melhor se ajustou ao conjunto de dados foi determinado pelo critério de validação cruzada (FARACO *et al.*, 2008). Por fim, considerando-se a estrutura de variabilidade espacial do modelo estimado escolhido, realizou-se a predição espacial, na qual se estimou a semivariância em pontos não amostrados utilizando-se a krigagem ordinária (SOARES, 2014).

#### 5.1.2.3 Cálculo do tamanho amostral efetivo (ESS)

O valor do tamanho amostral efetivo (ESS) foi estimado para cada variável dos conjuntos de dados reais e simulados, considerando-se que os dados apresentavam distribuição normal de probabilidade e média constante. Em ambas as situações estimou-se o valor do ESS univariado pelos métodos de Griffith (2005) (Equação 5.4,  $\widehat{ESS}_{GU}$ ) e Vallejos e Osorio (2014) (Equação 5.5,

$\widehat{ESS}_{VOU}$ ). Vale ressaltar que as metodologias apresentam diferentes pressupostos, tanto em relação às heurísticas utilizadas para o cálculo do ESS, quanto à distribuição probabilística das variáveis, haja vista a proposta de Griffith ser restrita para distribuição normal de probabilidade, enquanto a de Vallejos e Osorio pode ser adaptada e abranger outras distribuições da classe elíptica.

$$\widehat{ESS}_{GU} = \frac{\text{TR}(\mathbf{V}^{-1})}{\mathbf{1}^T \mathbf{V}^{-1} \mathbf{1}} n, \quad (5.4)$$

$$\widehat{ESS}_{VOU} = \mathbf{1}^T \mathbf{R}(\hat{\varphi})^{-1} \mathbf{1}, \quad (5.5)$$

em que:  $\mathbf{1}$  é um vetor unitário, de dimensão  $n \times 1$ , sendo  $n$  o número de pontos amostrais;  $\mathbf{V}^{-1} = \hat{\Sigma}$  é a matriz escala, simétrica, de dimensão  $n \times n$ ;  $\text{TR}(\cdot)$  é o operador traço; e  $\mathbf{R}(\hat{\varphi})$  é a matriz de correlação espacial dos pontos amostrais, de dimensão  $n \times n$ , invertível, cujos elementos  $R(\hat{\varphi})_{ij}$ ,  $i, j = 1, \dots, n$ , são dados por (VALLEJOS e OSORIO, 2014):

$$R(\hat{\varphi})_{ij} = \begin{cases} 1, & \text{se } i = j \\ \frac{\hat{\varphi}_2 r_{ij}}{\hat{\varphi}_1 + \hat{\varphi}_2}, & \text{se } i \neq j, \end{cases} \quad (5.6)$$

em que:  $R(\hat{\varphi})_{ij}$ ,  $i, j = 1, \dots, n$ , é o valor estimado da correlação espacial existente entre o  $i$ -ésimo e o  $j$ -ésimo ponto amostral; e  $r_{ij}$  são os elementos da matriz  $\mathbf{R}(\varphi_3)$  cujo cálculo está vinculado ao modelo geoestatístico utilizado e a distância euclidiana entre as observações. Tem-se ainda que  $\hat{\varphi}_2$  e  $\hat{\varphi}_1 + \hat{\varphi}_2$  são os valores estimados dos parâmetros contribuição e patamar, nesta ordem.

O tamanho amostral efetivo multivariado foi estimado considerando-se as propostas de Griffith (Equação 5.7,  $\widehat{ESS}_{GM}$ ) e Vallejos e Osorio (Equação 5.8,  $\widehat{ESS}_{VOM}$ ):

$$\widehat{ESS}_{GM} = \frac{\mathbf{1}_{(1)}^T \mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d \mathbf{1}_{(1)}}{\text{TR}(\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d)} \frac{\text{TR}(((\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}))}{\mathbf{1}_{(2)}^T ((\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \otimes \mathbf{I}) \mathbf{V}_d^{1/2}) \mathbf{1}_{(2)}} n, \quad (5.7)$$

$$\widehat{ESS}_{VOM} = \frac{\mathbf{1}_{(1)}^T ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{W}_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \mathbf{W}_d^{1/2}) \mathbf{1}_{(1)}}{\mathbf{1}_{(2)}^T \mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d \mathbf{1}_{(2)}}, \quad (5.8)$$

em que:  $\text{TR}(\cdot)$  é o operador traço;  $\otimes$  representa o produto de Kronecker;  $\mathbf{R}$ , de dimensão  $p \times p$ , em que  $p$  é o número de variáveis consideradas, é a matriz de correlação linear entre as  $p$  variáveis;  $\mathbf{A}_d$ , de dimensão  $p \times p$ , é uma matriz diagonal contendo o coeficiente dos  $p$  autovetores padronizados da matriz de correlação linear;  $\mathbf{1}_{(1)}$  é um vetor unitário de dimensão  $p \times 1$ ;  $\mathbf{1}_{(2)}$  é um vetor unitário de dimensão  $np \times 1$ ;  $\Phi_d$ , de dimensão  $p \times p$  é uma matriz diagonal contendo o desvio padrão da  $m$ -ésima variável  $\sigma_m$ ,  $m = 1, \dots, p$ ;  $\mathbf{V}_d$ , de dimensão  $np \times np$ , é uma matriz bloco diagonal contendo as matrizes  $\mathbf{V}_m^{-1}$ , de ordem  $n \times n$ , sendo que  $\mathbf{V}_p$  é a matriz escala de cada  $m$ -ésima variável,  $m = 1, \dots, p$ ;  $\mathbf{I}$ , de dimensão  $n \times n$ , é a matriz identidade; e  $\mathbf{W}_d$ , de dimensão  $np \times np$ , é uma matriz bloco diagonal contendo as matrizes  $\mathbf{R}_m(\varphi) = [r_{ij}(\varphi)]$ , de ordem  $n \times n$ , em que  $\mathbf{R}_p(\varphi)$  é a matriz de correlação espacial de cada  $m$ -ésima variável,  $m = 1, \dots, p$ .

No conjunto de dados simulados, para se obter a estimação dos valores do tamanho amostral efetivo uni e multivariado, utilizou-se o método de máxima verossimilhança para estimar os parâmetros efeito pepita e alcance.

As estimativas dos valores do tamanho amostral efetivo uni e multivariado, considerando os atributos físico-químicos do solo, foram calculadas para as seguintes situações:

- **Situação 1:** Envolveu as variáveis que apresentaram raio de dependência espacial superior a 100 m. São elas: os teores de ácido potencial total, cálcio, carbono, cobre, ferro, manganês, zinco, além da resistência do solo à penetração nas camadas 0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm.
- **Situação 2:** Agregou-se ao conjunto de dados descrito na situação 1, os teores de alumínio, fósforo, magnésio, potássio, além do pH, os quais exibiram baixo raio de dependência espacial, o qual foi inferior a 20 m.

Os procedimentos para a realização das simulações, a estimação dos valores do tamanho amostral efetivo, bem como as análises estatísticas e geoestatísticas foram realizados no software R (R DEVELOPMENT CORE TEAM, 2017) utilizando o pacote geoR (RIBEIRO JR e DIGGLE, 2001).

### 5.1.3 Resultados e discussões

#### 5.1.3.1 Estudo de simulação

A Figura 5 apresenta os gráficos com as médias e desvios padrão dos valores estimados do tamanho amostral efetivo univariado, considerando-se as 100 simulações de cada variável simulada, segundo as metodologias de Griffith (2005) (A) e Vallejos e Osorio (2014) (B).

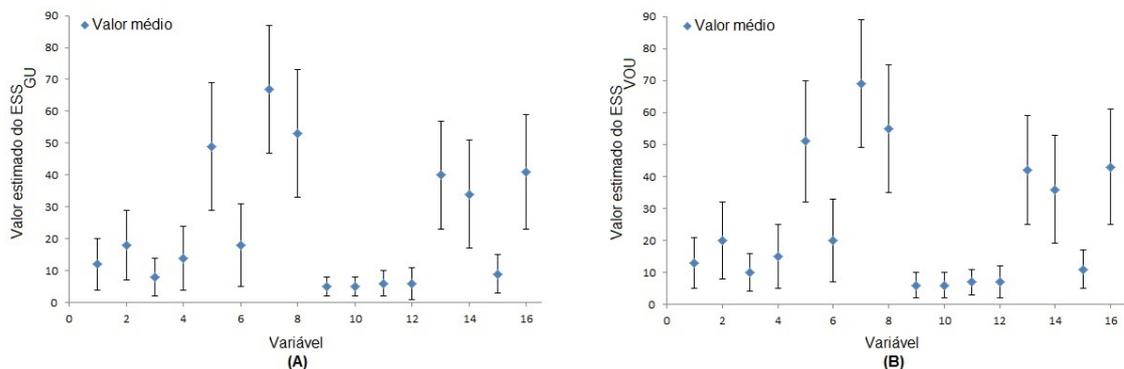


Figura 5 – Gráfico com a média e desvio padrão dos valores estimados do tamanho amostral efetivo, em todas as simulações, de acordo com as metodologias de Griffith (2005) ( $ESS_{GU}$ ) (A) e Vallejos e Osorio (2014) ( $ESS_{VOU}$ ) (B) para cada variável.

Em ambas metodologias utilizadas para a estimação do valor do tamanho amostral efetivo univariado, verificou-se que dois grupos de variáveis se destacaram por apresentar os maiores valores médios estimados do tamanho amostral efetivo, são eles: o grupo das variáveis V5 ( $\widehat{ESS}_{GU} = 49$  e  $\widehat{ESS}_{VOU} = 51$  pontos amostrais), V7 ( $\widehat{ESS}_{GU} = 67$  e  $\widehat{ESS}_{VOU} = 69$  pontos amostrais) e V8 ( $\widehat{ESS}_{GU} = 53$  e  $\widehat{ESS}_{VOU} = 55$  pontos amostrais), e o grupo das variáveis V13 ( $\widehat{ESS}_{GU} = 40$  e  $\widehat{ESS}_{VOU} = 42$  pontos amostrais), V14 ( $\widehat{ESS}_{GU} = 34$  e  $\widehat{ESS}_{VOU} = 36$  pontos amostrais) e V16 ( $\widehat{ESS}_{GU} = 41$  e  $\widehat{ESS}_{VOU} = 43$  pontos amostrais). Além disso, tanto pela proposta de Griffith (2005) quanto pela de Vallejos e Osorio (2014), estes dois grupos de variáveis exibiram elevados valores de desvio padrão, os quais variaram de 17 a 20 pontos amostrais (Figura 5 - A e B).

As variáveis V5 ( $a = 0,1$  m), V7 ( $a = 0,05$  m) e V8 ( $a = 0,07$  m) possuem os valores mais baixos do raio de dependência espacial. Mesma característica das variáveis V13 ( $a = 0,09$  m), V14

( $a=0,11$  m) e V16 ( $a=0,1$  m). Entretanto, V13 e V14 apresentaram alcances práticos pouco maiores quando comparados aos das variáveis V7 e V8. Isso resultou em valores médios estimados do tamanho amostral efetivo levemente maiores nestas últimas variáveis.

Notou-se, em ambas as metodologias utilizadas, que as variáveis V9 ( $\widehat{ESS}_{GU}=5$  e  $\widehat{ESS}_{VOU}=6$  pontos amostrais), V10 ( $\widehat{ESS}_{GU}=5$  e  $\widehat{ESS}_{VOU}=6$  pontos amostrais), V11 ( $\widehat{ESS}_{GU}=6$  e  $\widehat{ESS}_{VOU}=7$  pontos amostrais) e V12 ( $\widehat{ESS}_{GU}=6$  e  $\widehat{ESS}_{VOU}=7$  pontos amostrais) exibiram os menores valores médios estimados do tamanho amostral efetivo univariado (Figura 5 - A e B). Estas variáveis detêm os maiores valores de alcance prático, entre 1000 e 1200 m.

Em geral, quando se compara o grupo das variáveis V7, V8, V13 e V14 com o grupo das variáveis V9, V10, V11 e V12, pode-se ainda verificar uma grande diferença no valor médio estimado do ESS, em ambas as metodologias (Figura 5 - A e B). Embora todas as variáveis tenham forte dependência espacial,  $EPR=9\%$  (CAMPARDELLA *et al.*, 1994), o fato do alcance das variáveis V7, V8, V13 e V14 ser inferior a 100 m, enquanto das variáveis V9, V10, V11 e V12, varia de 1000 a 1200 m, influenciou na grande diferença do valor estimado do ESS obtido entre os grupos.

Sendo assim, o alcance foi um parâmetro que influenciou de maneira inversa no valor obtido do tamanho amostral efetivo univariado, pois quanto maior o raio de dependência espacial da variável, menor foi o número estimado de pontos amostrais. Nos estudos de simulação de Vallejos e Osorio (2014), foram obtidas conclusões similares em relação ao alcance, mesmo considerando-se uma situação distinta da presente neste trabalho quanto ao tamanho amostral inicial e tipo de amostragem utilizado.

Contrastando as variáveis V1 ( $\widehat{ESS}_{GU}=12$  e  $\widehat{ESS}_{VOU}=13$  pontos amostrais), V3 ( $\widehat{ESS}_{GU}=8$  e  $\widehat{ESS}_{VOU}=10$  pontos amostrais), V7 ( $\widehat{ESS}_{GU}=67$  e  $\widehat{ESS}_{VOU}=69$  pontos amostrais), V8 ( $\widehat{ESS}_{GU}=53$  e  $\widehat{ESS}_{VOU}=55$  pontos amostrais), V13 ( $\widehat{ESS}_{GU}=40$  e  $\widehat{ESS}_{VOU}=42$  pontos amostrais), V14 ( $\widehat{ESS}_{GU}=34$  e  $\widehat{ESS}_{VOU}=36$  pontos amostrais) e V16 ( $\widehat{ESS}_{GU}=41$  e  $\widehat{ESS}_{VOU}=43$  pontos amostrais), as quais apresentam efeito pepita igual a 0,1, com as variáveis V2 ( $\widehat{ESS}_{GU}=18$  e  $\widehat{ESS}_{VOU}=20$  pontos amostrais), V4 ( $\widehat{ESS}_{GU}=14$  e  $\widehat{ESS}_{VOU}=15$  pontos amostrais), V5 ( $\widehat{ESS}_{GU}=49$  e  $\widehat{ESS}_{VOU}=51$  pontos amostrais) e V6 ( $\widehat{ESS}_{GU}=18$  e  $\widehat{ESS}_{VOU}=20$  pontos amostrais), que têm efeito pepita variando entre 0,5 e 0,9, foi possível constatar que não há um padrão no valor médio estimado do ESS obtido entre as variáveis em cada grupo, e nem entre os grupos. Isso indica que a intensidade da dependência espacial não exerce uma relevante influência no valor médio estimado do ESS.

Em relação à redução no número de amostras, verificou-se que entre as metodologias abordadas não houve muita distinção no valor médio estimado do ESS univariado. No geral, as metodologias proporcionaram redução no número de pontos amostrais que variou de 32% a 95%. A elevada variabilidade foi ocasionada pela diferença nos valores dos parâmetros das variáveis, especialmente em relação ao alcance, que variou entre 50 e 1200 m.

As estatísticas descritivas dos valores estimados do ESS multivariado, de acordo com as metodologias de Griffith (2005) e Vallejos e Osorio (2014) para as 100 simulações em cada cenário, estão apresentadas na Tabela 5.

Tabela 5 – Estatísticas descritivas dos valores estimados do ESS multivariado nas 100 simulações em cada cenário para as metodologias de Griffith (2005) (G) e Vallejos e Osorio (2014) (VO).

	Mín.		1° Q		Med.		Média		3° Q		Máx.		Amp.		CV (%)	
	G	VO	G	VO	G	VO	G	VO	G	VO	G	VO	G	VO	G	VO
<b>Cen. 1</b>	2,30	19,11	4,93	27,88	8,29	31,80	9,85	33,03	12,11	37,15	37,84	52,68	35,54	33,57	69,50	21,67
<b>Cen. 2</b>	2,10	18,69	6,24	35,76	9,50	41,09	12,10	42,01	14,49	48,43	48,16	59,66	46,16	40,97	72,79	20,51
<b>Cen. 3</b>	1,42	33,81	9,13	54,74	17,44	61,27	22,55	62,08	29,16	69,88	84,78	96,69	83,36	62,88	82,46	18,22
<b>Cen. 4</b>	1,06	13,26	2,00	17,22	3,19	19,40	4,37	20,10	4,90	22,33	23,66	35,46	22,60	22,20	91,46	19,58
<b>Cen. 5</b>	7,12	47,39	26,62	62,52	38,09	68,52	43,34	68,31	52,20	73,95	102,00	87,51	94,88	40,12	54,17	12,98
<b>Cen. 6</b>	1,20	16,02	3,43	24,37	6,14	27,49	7,92	27,77	11,22	31,48	36,31	41,44	35,11	25,42	83,18	18,93
<b>Cen. 7</b>	1,56	19,63	3,74	27,77	5,51	33,49	7,96	32,53	10,27	36,75	34,03	55,19	32,47	35,56	77,68	17,86

Cen.: cenário, Mín.: mínimo, 1° Q: 1° quartil, Med.: mediana, 3° Q: 3° quartil, Máx.: máximo, Amp.: amplitude, CV: coeficiente de variação, G: metodologia de Griffith, VO: metodologia de Vallejos e Osorio.

No geral, os valores estimados do ESS multivariado, a partir da metodologia de Vallejos e Osorio (2014), foram maiores quando comparados aos obtidos via proposta de Griffith (2005). Isso porque embora o intuito da aplicação das metodologias seja reduzir o tamanho amostral ponderando os efeitos da autocorrelação espacial, as heurísticas das propostas são distintas.

Considerando-se o cenário 7, o qual agrega todas as variáveis (exceto a V15), observou-se uma redução média no número de pontos amostrais de 93% (via Griffith) e 68% (por Vallejos e Osorio) (Tabela 5). Mesmo que financeiramente seja mais interessante coletar uma quantidade menor de pontos amostrais, uma grande redução no número de pontos pode ocasionar uma perda na qualidade dos mapas temáticos, os quais representam a variabilidade espacial das variáveis em uma área agrícola.

Kestring *et al.* (2015) fizeram um estudo, compararam diferentes grades e densidades amostrais e verificaram que ao se reduzir muito o número de pontos amostrais (em 92%), os índices de acurácia, os quais comparam o mapa temático confeccionado com a quantidade inicial de pontos amostrais com aquele construído com número reduzido de pontos, foram bastante baixos. Isso indica que os mapas não são tão similares entre si, isto é, que o mapa construído com a quantidade reduzida de pontos amostrais é pouco representativo em relação às características da área agrícola.

Verificou-se que os cenários 1, 2, 6 e 7 são bastante semelhantes. A similaridade foi constatada em ambas as metodologias de acordo com os valores estimados do ESS multivariado, bem como suas estatísticas descritivas calculadas para estes cenários (Tabela 5 e Figura 6 - A e B).

Contrapondo os cenários 6 e 7, verificou-se uma diferença no valor médio estimado de, aproximadamente, cinco pontos amostrais (de 28 no cenário 6 para 33 no cenário 7) via método de Vallejos e Osorio. Todavia, pela proposta de Griffith, não houve diferença média no número de pontos amostrais entre estes cenários. Isso indica que as cinco variáveis com alcances inferiores a 110 metros que foram agregadas ao cenário 6, compondo-se assim o cenário 7, não influenciaram o valor estimado do ESS. Possivelmente, isto se deve ao fato de que o cenário 7 apresenta 10 variáveis com alcances variando de 300 a 1200 m, e apenas 5 variáveis com alcances entre 50 e 110 m. Deste modo, essa diferente proporção entre o número de variáveis com alcances altos e baixos fez com basicamente não houvesse diferença no valor médio estimado do ESS entre os cenários 6 e 7.

Entre os cenários 1 e 3 percebe-se, em ambos os métodos, um pequeno aumento nos valores estimados do ESS (Tabela 5 e Figura 6 - A e B). Considerando-se a proposta de Griffith, o valor médio estimado do ESS aumentou de 10 pontos amostrais no cenário 1 para 23 pontos

amostrais no cenário 3; e de 33 para 62 pontos amostrais, pelo método de Vallejos e Osorio (Tabela 5). O que distinguiu estes dois cenários foi, principalmente, o valor do raio de dependência espacial das variáveis que os compõe: no cenário 1, o alcance das variáveis georreferenciadas variou de 300 a 500 m, enquanto no cenário 3, duas das quatro variáveis têm alcance inferior a 100 m.

Para o método de Griffith, verificou-se que o cenário 5 apresentou os maiores valores de todas as estatísticas de posição quando comparado aos demais cenários. Essa constatação se repete ao se considerar a proposta de Vallejos e Osorio, exceto para o valor máximo (Tabela 5; Figura 6 - A e B). Quando comparado aos outros cenários, o cenário 3 ostentou os segundos maiores valores para a maioria das medidas de posição em ambas as metodologias utilizadas para a estimação do valor do ESS multivariado (Tabela 5; Figura 6 - A e B). Verificou-se ainda, nas duas propostas, que o cenário 4 apresentou os mais baixos valores em todas as estatísticas de posição em relação aos demais cenários (Tabela 5; Figura 6 - A e B). Os cenários 5 e 3 têm em comum variáveis com baixo valor do raio de dependência espacial, enquanto o cenário 4 agrega as variáveis com os maiores valores de alcance prático. Os resultados indicam que o raio de dependência espacial influencia o valor estimado do tamanho amostral efetivo: quanto maior o alcance, menor o valor estimado do ESS.

Embora os resultados obtidos nas metodologias estudadas tenham convergido quanto às conclusões descritas anteriormente, pode-se destacar que o método de Vallejos e Osorio apresentou menor variabilidade nos valores estimados do ESS multivariado comparado ao método de Griffith (Tabela 5).

A proposta de Vallejos e Osorio quando comparada à de Griffith apresentou, em cada cenário, um menor número de pontos discrepantes dos valores estimados do ESS multivariado nas 100 simulações. A metodologia de Griffith apresentou pontos discrepantes em todos os cenários, em um percentual que variou de 3 a 10%; enquanto, por Vallejos e Osorio, apenas os cenários 1, 3 e 4 exibiram valores atípicos: um ponto em cada cenário (Figura 6 - A e B).

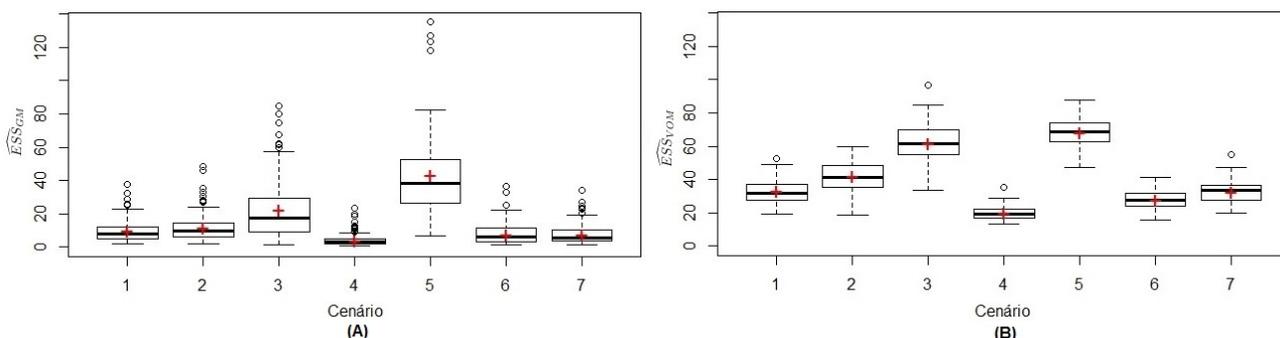


Figura 6 – Gráficos boxplot dos valores estimados do ESS multivariado nas 100 simulações em cada cenário, de acordo com as metodologias de Griffith (2005) ( $\widehat{ESS}_{GM}$ ) (A) e Vallejos e Osorio (2014) ( $\widehat{ESS}_{VOM}$ ) (B). O símbolo “+” representa o valor médio estimado do ESS multivariado em cada cenário.

### 5.1.3.2 Aplicação da metodologia à variáveis físico-químicas do solo

Na Tabela 6, são apresentados os valores das estatísticas descritivas referentes às variáveis físico-químicas do solo estudadas.

Os valores do coeficiente de variação (CV) dos atributos químicos do solo indicaram que apenas as variáveis pH e teor de carbono apresentaram, respectivamente, homogeneidade ( $CV < 10\%$ ) e média dispersão ( $10\% \leq CV \leq 20\%$ ) (PIMENTEL GOMES, 1985). Os dados dos demais atributos apresentaram alta dispersão ( $20\% < CV \leq 30\%$ ) ou heterogeneidade ( $CV > 30\%$ ), segundo classificação de Pimentel Gomes (1985), com destaque para o alto coeficiente de variação do teor de alumínio (144,80%) (Tabela 6).

A análise da diferença entre os valores da média e mediana do teor de alumínio, bem como da discrepância entre seus valores de mínimo e máximo e do alto valor do CV deste, permitem concluir que não há similaridade nos dados do referido atributo químico. O mesmo ocorre com os teores de fósforo e manganês; este último embora não tenha exibido alto valor para o CV, apresentou elevado valor de desvio padrão (Tabela 6).

Em relação à resistência do solo à penetração (RSP) nas quatro diferentes profundidades analisadas (0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm), tem-se que: o menor valor (0,75 MPa) e o maior valor (6,67 MPa) de resistência do solo à penetração foram apresentados na profundidade 0-10 cm. Os dados da RSP na camada 0-10 cm podem ser considerados heterogêneos ( $CV = 31,22\%$ ), enquanto na profundidade 11-20 cm tem-se alta dispersão ( $CV = 20,70\%$ ). Por outro lado, a RSP nas camadas 21-30 cm e 31-40 cm registrou-se média dispersão nos dados, com valores de CV iguais a 16,27% e 18,01%, respectivamente (PIMENTEL GOMES, 1985).

Tabela 6 – Estatística descritiva dos atributos físico-químicos do solo.

Atributos	Min.	1° Q	Med.	Média	3° Q	Máx.	DP	CV (%)
Cálcio (cmolc dm <sup>-3</sup> )	2,51	4,61	5,35	5,51	6,37	11,43	1,33	24,12
Carbono (g dm <sup>-3</sup> )	23,80	29,47	32,27	32,01	34,12	39,34	3,39	10,58
Cobre (mg dm <sup>-3</sup> )	1,51	3,33	3,78	3,82	4,35	5,97	0,91	23,79
Ferro (mg dm <sup>-3</sup> )	29,22	47,32	53,66	53,22	58,22	122,31	11,38	21,39
H+Al <sup>3+</sup> (cmolc dm <sup>-3</sup> )	2,95	6,69	8,36	8,05	9,53	14,08	2,01	25,01
Manganês (mg dm <sup>-3</sup> )	44,48	71,7	83,45	86,41	97,29	150,00	22,18	25,66
Zinco (mg dm <sup>-3</sup> )	1,85	3,57	4,68	4,97	5,84	13,84	2,03	40,92
RSP 0-10 cm (MPa)	0,75	2,49	2,99	3,12	3,67	6,67	0,97	31,22
RSP 11-20 cm (MPa)	1,78	2,62	2,94	3,01	3,34	5,53	0,62	20,70
RSP 21-30 cm (MPa)	1,50	1,83	1,96	2,03	2,16	3,86	0,33	16,27
RSP 31-40 cm (MPa)	1,48	1,84	2,01	2,08	2,27	3,95	0,37	18,01
Alumínio (cmolc dm <sup>-3</sup> )	0	0	0,08	0,15	0,20	1,12	0,22	144,80
Fósforo (mg dm <sup>-3</sup> )	6,13	14,42	18,46	21,88	25,90	117,10	14,56	66,54
Magnésio (cmolc dm <sup>-3</sup> )	0,87	1,57	1,93	2,05	2,30	5,50	0,76	37,21
pH	4,30	4,70	4,90	4,95	5,10	6,50	0,36	7,39
Potássio (cmolc dm <sup>-3</sup> )	0,08	0,27	0,35	0,37	0,47	0,78	0,14	37,53

Mín.: mínimo, 1° Q: 1° quartil, Med.: mediana, 3° Q: 3° quartil, Máx.: máximo, DP: desvio padrão, CV: coeficiente de variação.

Os pontos discrepantes de cada atributo são destacados espacialmente na área sob estudo nos gráficos post-plot (Figura 7).

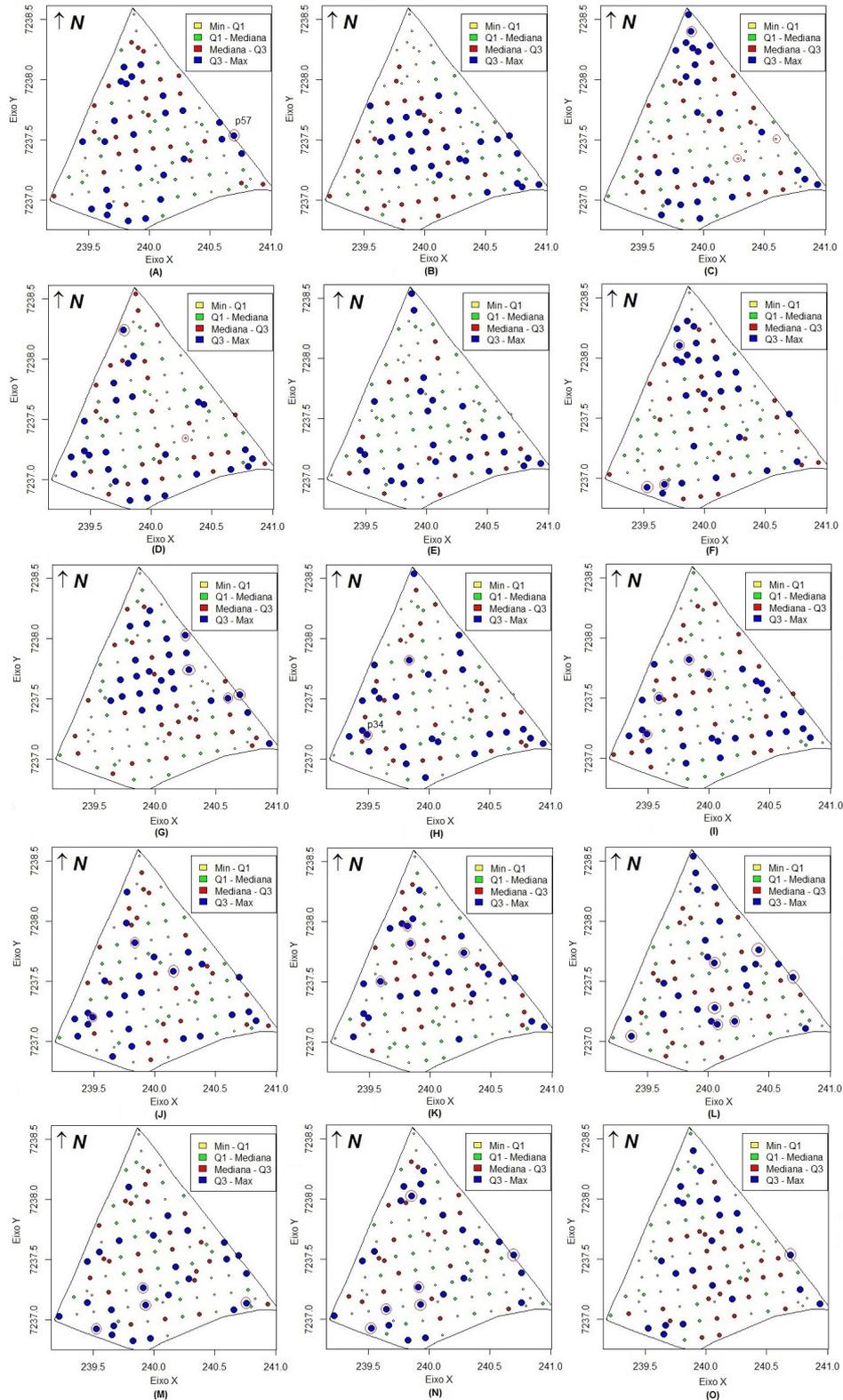


Figura 7 – Gráfico post-plot dos atributos: (A) cálcio [ $\text{cmolc dm}^{-3}$ ], (B) carbono [ $\text{g dm}^{-3}$ ], (C) cobre [ $\text{mg dm}^{-3}$ ], (D) ferro [ $\text{mg dm}^{-3}$ ], (E)  $\text{H+Al}^3$  [ $\text{cmolc dm}^{-3}$ ], (F) manganês [ $\text{mg dm}^{-3}$ ], (G) zinco [ $\text{mg dm}^{-3}$ ]; e resistência do solo à penetração (RSP) [MPa] nas camadas: (H) 0-10 cm, (I) 11-20 cm, (J) 21-30 cm, (K) 31-40 cm, (L) fósforo [ $\text{mg dm}^{-3}$ ], (M) magnésio [ $\text{cmolc dm}^{-3}$ ], (N) pH, (O) potássio [ $\text{cmolc dm}^{-3}$ ]. As circunferências destacam especialmente os pontos discrepantes. Min: mínimo, Q1: 1º quartil, Q3: 3º quartil, Max: máximo.

Com exceção do teor de carbono (Figura 7 - B) e do teor de  $H+Al^3$  (Figura 7 - E), todos os atributos químicos apresentaram pontos discrepantes com elevadas concentrações, localizados principalmente na região leste da área. O teor de fósforo (Figura 7 - L) foi o atributo que exibiu a maior quantidade de pontos discrepantes (sete). Verificou-se que o ponto amostral 57 (localizado espacialmente na Figura 7 - B) apresentou valores discrepantes simultaneamente para cinco atributos químicos: teor de cálcio (Figura 7 - A), teor de fósforo (Figura 7 - L), pH (Figura 7 - N), teor de potássio (Figura 7 - O) e teor de zinco (Figura 7 - G). E que o ponto amostral 34 (localizado espacialmente na Figura 7 - M) exibiu valores discrepantes em todas as camadas de RSP, com exceção da 31-40 cm. As camadas 11-20 cm (Figura 7 - I) e 31-40 cm (Figura 7 - K) de profundidade da RSP exibiram a maior quantidade de valores atípicos (quatro) localizados nas regiões central e oeste da área agrícola.

Destaca-se que devido à quantidade elevada de amostras mensuradas com teores nulos do alumínio, não foi possível obter o post-plot dessa variável. Pode-se verificar que os atributos teor de carbono (Figura 7 - B), teor de ferro (Figura 7 - D), teor de  $H+Al^3$  (Figura 7 - E), teor de manganês (Figura 7 - F) e teor de zinco (Figura 7 - G) apresentaram aparente aglomeração de pontos cujas concentrações dos referidos atributos são elevadas. Os teores de carbono (Figura 7 - B) e zinco (Figura 7 - G) apresentaram aglomeração no centro da área, o teor de manganês (Figura 7 - F) exibiu aglomeração na direção norte, o teor de ferro (Figura 7 - D) nas direções sul e oeste da área, e o teor de  $H+Al^3$  na direção sul (Figura 7 - E).

Entretanto, como não se trata de um agrupamento denso de pontos tal que a tendência direcional fique clara (Figura 7), foi calculado o coeficiente de correlação linear de Pearson dos atributos químicos e físicos do solo nas direções X e Y. Os valores calculados dos coeficientes de correlação linear de Pearson, para ambas as direções, variaram, em módulo, de 0,01 a 0,25. Portanto, segundo a classificação de Dancey e Reidy (2006), as variáveis não apresentaram tendência direcional, pois os coeficientes de correlação linear de Pearson foram inferiores a 0,4.

Para todas as variáveis, ajustou-se à função semivariância modelos geoestatísticos com fator de anisotropia. Assim, identificou-se que, nas variáveis teor de cobre, teor de potássio, teor de alumínio, teor de ferro, teor de ácido potencial total e resistência do solo à penetração nas camadas 0-10 cm e 31-40 cm de profundidade, os fatores de anisotropia estimados variaram entre 1 e 2, indicando presença de anisotropia geométrica na estrutura espacial dessas variáveis. Entretanto, segundo estudos desenvolvidos por Guedes, Uribe-Opazo e Ribeiro JR (2013), fatores de anisotropia geométrica estimados inferiores a 2 não são relevantes na estimativa dos parâmetros de dependência espacial. Deste modo, considerou-se nesse trabalho que a estrutura de dependência espacial é isotrópica para todos os atributos físico-químicos do solo.

A Tabela 7 exibe os valores estimados dos parâmetros de dependência espacial e o efeito pepita relativo (EPR) obtidos pelo método de máxima verossimilhança e considerando o melhor modelo ajustado para cada atributo pelos critérios de validação cruzada (VC), Akaike e Akaike de segunda ordem. Pode-se observar que as variáveis cobre e resistência do solo à penetração nas camadas 11-20 cm e 31-40 cm foram melhor ajustadas pelo modelo exponencial, enquanto o teor de magnésio e o pH foram melhor ajustados pelo modelo da família Matérn com parâmetro de forma  $\kappa = 1$  e 1,5, respectivamente. O gaussiano foi o melhor modelo ajustado aos demais atributos.

Segundo a classificação de Cambardella *et al.* (1994), todos os atributos, com exceção dos teores de ferro e  $H+Al^3$ , apresentaram moderada ou forte dependência espacial, pois os valores

estimados do efeito pepita relativo (EPR) foram inferiores a 0,75. Embora o ferro tenha apresentado fraca dependência espacial (EPR=0,80) (CAMBARDELLA *et al.*, 1994), o alcance prático foi o maior entre os atributos físico-químicos estudados, 968,5 m.

As variáveis teor de alumínio, teor de fósforo, teor de magnésio, pH e teor de potássio exibiram valor zero para o efeito pepita e, conseqüentemente, o EPR também foi nulo, indicando forte dependência espacial (CAMBARDELLA *et al.*, 1994). Entretanto, o raio de dependência espacial extremamente baixo (variando de 5,2 a 17,30 m), sendo inclusive menor que a maioria das distâncias entre os pares de pontos amostrados (cuja distância mínima foi de 50 m), é um indício de que, na prática, estas variáveis não apresentaram dependência espacial (efeito pepita puro).

Observou-se que a magnitude dos desvios padrão é similar à do respectivo parâmetro estimado, isto é, parâmetros cuja estimativa foi alta, tendem a exibir maiores valores de desvio padrão. Além disso, nas variáveis que apresentaram efeito pepita com valor estimado próximo a zero ou nulo, os valores estimados do desvio padrão deste parâmetro e da contribuição foram iguais ou muito semelhantes, corroborando com os resultados de Dalposso *et al.* (2018) e Fagundes *et al.* (2018).

Tabela 7 – Valores estimados dos parâmetros do melhor modelo geoestatístico ajustado para cada atributo.

Atributos (Modelos)	Valores estimados						$\widehat{EPR}$
	$\hat{\beta}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_1 + \hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{a} = g(\hat{\varphi}_3)$	
Cálcio (Gaus.)	5,53 (0,18)	1,29 (0,36)	0,48 (0,36)	1,77 -	0,16 (0,08)	284,10 -	0,73 -
Carbono (Gaus.)	31,86 (0,70)	8,02 (1,36)	3,39 (1,80)	11,41 -	0,31 (0,11)	538,70 -	0,70 -
Cobre (Exp.)	4,03 (0,27)	0,34 (0,14)	0,52 (0,22)	0,86 -	0,28 (0,19)	855,10 -	0,39 -
Ferro (Gaus.)	55,52 (2,92)	109,02 (16,39)	27,16 (20,28)	136,17 -	0,56 (0,28)	968,50 -	0,80 -
H+Al <sup>3</sup> (Gaus.)	8,03 (0,28)	3,44 (0,67)	0,57 (0,57)	4,02 -	0,22 (0,16)	384,10 -	0,85 -
Manganês (Gaus.)	86,78 (4,16)	268,79 (59,89)	226,14 (89,74)	494,93 -	0,21 (0,06)	367,3 -	0,54 -
Zinco (Gaus.)	5,10 (0,45)	1,59 (0,38)	3,05 (0,95)	4,64 -	0,22 (0,04)	367,70 -	0,34 -
RSP 0-10 cm (Gaus.)	3,11 (0,13)	0,64 (0,21)	0,29 (0,22)	0,93 -	0,15 (0,07)	255,50 -	0,68 -
RSP 11-20 cm (Exp.)	3,00 (0,07)	0,01 (1,16)	0,36 (1,17)	0,38 -	0,03 (0,06)	101,50 -	0,04 -
RSP 21-30 cm (Gaus.)	2,02 (0,04)	0,03 (0,08)	0,07 (0,08)	0,11 -	0,07 (0,04)	115,90 -	0,30 -
RSP 31-40 cm (Exp.)	2,07 (0,05)	0,06 (0,10)	0,07 (0,10)	0,14 -	0,07 (0,07)	221,50 -	0,45 -
Alumínio (Gaus.)	0,150 (0,021)	0 (0,003)	0,050 (0,003)	0,050 -	0,010 (1,160.10 <sup>-26</sup> )	17,300 -	0 -
Fósforo (Gaus.)	21,880 (1,430)	0 (14,700)	209,950 (14,700)	209,950 -	0,003 (1,490.10 <sup>-26</sup> )	5,200 -	0 -
Magnésio (Mat. $\kappa=1$ )	2,050 (0,080)	0 (0,040)	0,580 (0,040)	0,580 -	0,002 (5,770.10 <sup>-49</sup> )	6,700 -	0 -
pH (Mat. $\kappa=1,5$ )	4,950 (0,040)	0 (0,010)	0,130 (0,010)	0,130 -	0,002 (7,90.10 <sup>-41</sup> )	8,000 -	0 -
Potássio (Gaus.)	0,370 (0,014)	0 (0,001)	0,019 (0,001)	0,019 -	0,008 (1,910.10 <sup>-35</sup> )	13,800 -	0 -

Valores estimados de:  $\hat{\beta}$ : média,  $\hat{\varphi}_1$ : efeito pepita,  $\hat{\varphi}_2$ : contribuição,  $\hat{\varphi}_3$ : função do alcance,  $\hat{\varphi}_1 + \hat{\varphi}_2$ : patamar,  $\hat{a}$ : alcance prático (em metros),  $\widehat{EPR} = 100 \frac{\hat{\varphi}_1}{\hat{\varphi}_1 + \hat{\varphi}_2}$ : efeito pepita relativo (%). Modelos geoestatísticos: Gaus.: gaussiano, Exp.: exponencial, Mat.: família Matérn. Entre parênteses, são apresentados os valores estimados do desvio padrão de cada parâmetro.

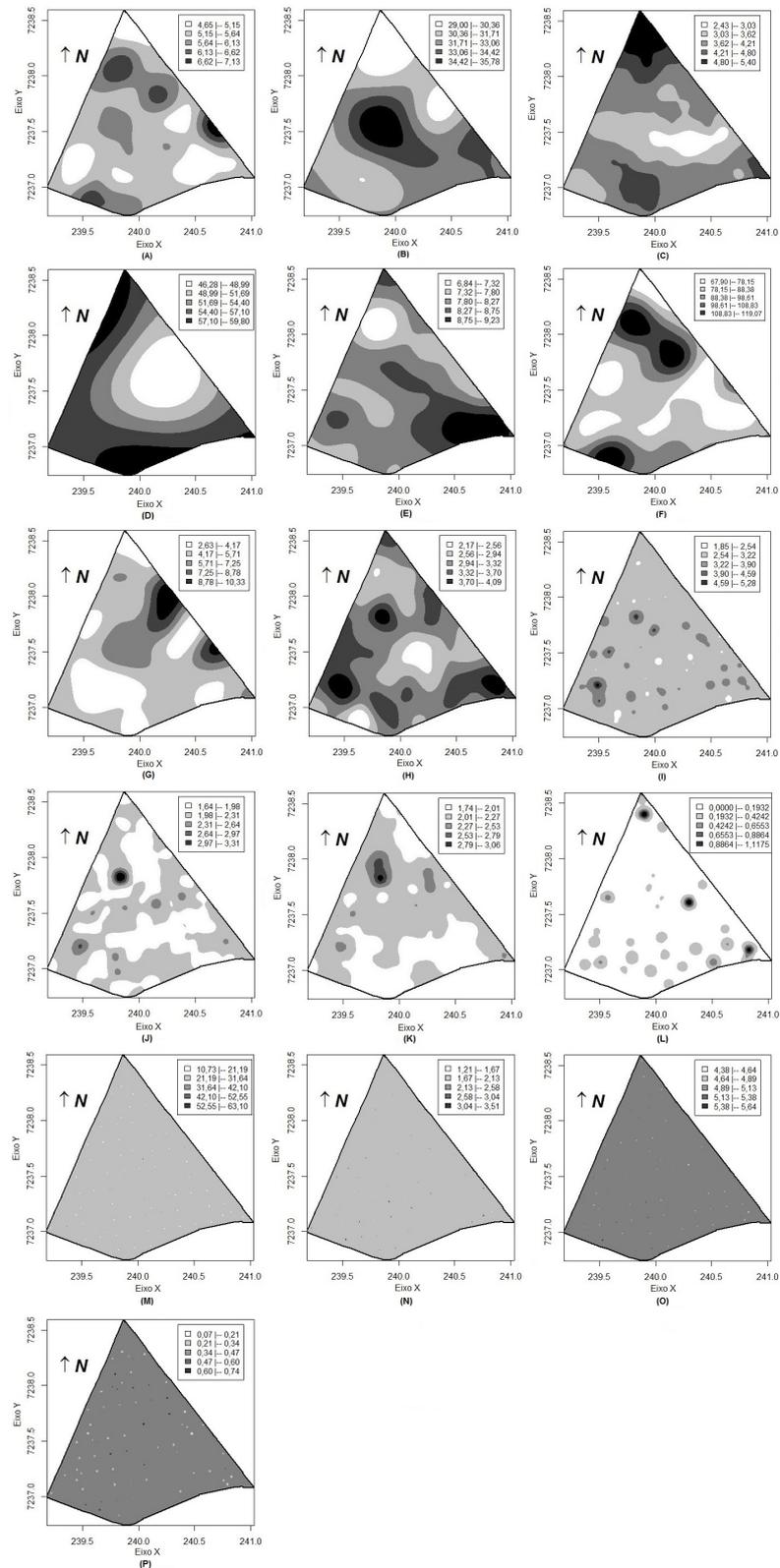


Figura 8 – Mapa dos valores estimados dos atributos: (A) cálcio [ $\text{cmolc dm}^{-3}$ ], (B) carbono [ $\text{g dm}^{-3}$ ], (C) cobre [ $\text{mg dm}^{-3}$ ], (D) ferro [ $\text{mg dm}^{-3}$ ], (E) H+Al<sup>3</sup> [ $\text{cmolc dm}^{-3}$ ], (F) manganês [ $\text{mg dm}^{-3}$ ], (G) zinco [ $\text{mg dm}^{-3}$ ]; e resistência do solo à penetração (RSP) [MPa] nas camadas: (H) 0-10 cm, (I) 11-20 cm, (J) 21-30 cm, (K) 31-40 cm, (L) alumínio [ $\text{cmolc dm}^{-3}$ ], (M) fósforo [ $\text{mg dm}^{-3}$ ], (N) magnésio [ $\text{cmolc dm}^{-3}$ ], (O) pH, (P) potássio [ $\text{cmolc dm}^{-3}$ ].

Verificou-se nos mapas temáticos dos atributos teor de alumínio (Figura 8 - L), teor de fósforo (Figura 8 - M), teor de magnésio (Figura 8 - N), pH (Figura 8 - O) e teor de potássio (Figura 8 - P), nos quais o alcance foi inferior à menor distância amostrada, o fenômeno *bull eyes effect* (MENEZES *et al.*, 2016), que consiste na formação de sub-regiões centradas nos pontos amostrais. Nestes mapas, toda área foi homogênea em relação ao valor do atributo, exceto pelas microrregiões circulares em torno das amostras. A resistência do solo à penetração (RSP) nas camadas 11-20 cm (Figura 8 - I), 21-30 cm (Figura 8 - J) e 31-40 cm (Figura 8 - K), cujo alcance prático variou entre 101,5 e 221,5 m (Tabela 7), também exibiram sub-regiões centradas em alguns pontos amostrais.

Em relação aos demais atributos, pode-se destacar a baixa e alta concentração do teor de carbono respectivamente ao norte e no centro (Figura 8 - B) da área. O mapa do teor de cobre (Figura 8 - C) evidenciou altas concentrações desse atributo ao norte da área agrícola estudada. Quanto aos teores de cálcio, manganês e zinco (Figura 8 - A, F e G), a maior proporção da área (51%, 41% e 50%, respectivamente) apresentou concentrações desses atributos que variam na segunda classe apresentada no mapa, a qual abrange o teor médio amostral (Tabela 6).

Verificou-se ainda que existem regiões com baixas concentrações do teor de ferro ao leste da área, e que estas aumentam conforme direciona-se ao oeste (Figura 8 - D). Em relação ao teor de ácido potencial total ( $H+Al^3$ ), notou-se que seus valores variam entre 7,32 e 8,27  $cmolc\ dm^{-3}$ , representadas respectivamente nas 2ª e 3ª classes no mapa, na maior proporção da área (24% e 47%, nesta ordem) (Figura 8 - E).

A RSP, na profundidade 0-10 cm (Figura 8 - H), não apresentou nenhuma classe que é claramente mais abrangente em termos de área que a outra. Enquanto na camada 31-40 cm (Figura 8 - K), observou-se que a maior proporção da área (60%) é representada pela segunda classe (que varia de 2,01 a 2,27 MPa), a qual inclui a média amostral.

Considerando-se os dados obtidos nos 102 pontos amostrais, estimou-se o tamanho amostral efetivo de cada atributo sob estudo (Tabela 8), pelas metodologias de Griffith (2005) (Equação 5.4) e Vallejos e Osorio (2014) (Equação 5.5).

Tabela 8 – Valores estimados do tamanho amostral efetivo de cada atributo segundo as metodologias de Griffith (2005) ( $\widehat{ESS}_{GU}$ ) e Vallejos e Osorio (2014) ( $\widehat{ESS}_{VOU}$ ).

Atributos (Modelos)	$\widehat{ESS}_{GU}$	$\widehat{ESS}_{VOU}$
Cálcio (Gaus.)	50,77 $\approx$ 51	52,32 $\approx$ 52
Carbono (Gaus.)	21,15 $\approx$ 21	23,40 $\approx$ 23
Cobre (Exp.)	9,74 $\approx$ 10	11,51 $\approx$ 11
Ferro (Gaus.)	13,66 $\approx$ 14	16,01 $\approx$ 16
$H+Al^3$ (Gaus.)	51,40 $\approx$ 52	52,63 $\approx$ 53
Manganês (Gaus.)	26,26 $\approx$ 26	28,58 $\approx$ 28
Zinco (Gaus.)	19,78 $\approx$ 20	22,42 $\approx$ 22
RSP 0-10 cm (Gaus.)	53,07 $\approx$ 53	54,77 $\approx$ 55
RSP 11-20 cm (Exp.)	87,76 $\approx$ 88	88,74 $\approx$ 89
RSP 21-30 cm (Gaus.)	84,90 $\approx$ 85	87,76 $\approx$ 88
RSP 31-40 cm (Exp.)	58,91 $\approx$ 59	60,40 $\approx$ 60
Alumínio (Gaus.)	102	102
Fósforo (Gaus.)	102	102
Magnésio (Mat. $\kappa = 1$ )	102	102
pH (Mat. $\kappa = 1,5$ )	102	102
Potássio (Gaus.)	102	102

Gaus.: modelo gaussiano; Exp.: modelo exponencial; Mat.  $\kappa=1$  e 1,5: modelos da família Matérn com parâmetro de forma  $\kappa=1$  e 1,5, respectivamente.

Verificou-se que para os atributos teor de alumínio, teor de fósforo, teor de magnésio, pH

e teor de potássio (os quais apresentaram baixo valor estimado do raio de dependência espacial, variando entre 5,2 e 17,3 m), o valor estimado do tamanho amostral efetivo univariado em ambas metodologias foi igual ao número inicial de pontos amostrais (102) (Tabela 8). Isso porque as duas metodologias baseiam-se na existência de informação redundante proveniente de amostras vizinhas altamente correlacionadas, mas como as observações destes atributos são espacialmente independentes entre si, fato constatado ao se analisar o baixo valor estimado do alcance (Tabela 7), não ocorreu redução no tamanho amostral.

Considerando-se os atributos em que se constatou redução no número de pontos amostrais, construiu-se a Figura 9, que apresenta um diagrama de dispersão dos valores estimados do ESS univariado (Tabela 8) em função dos parâmetros alcance e EPR estimados para cada variável (Tabela 7).

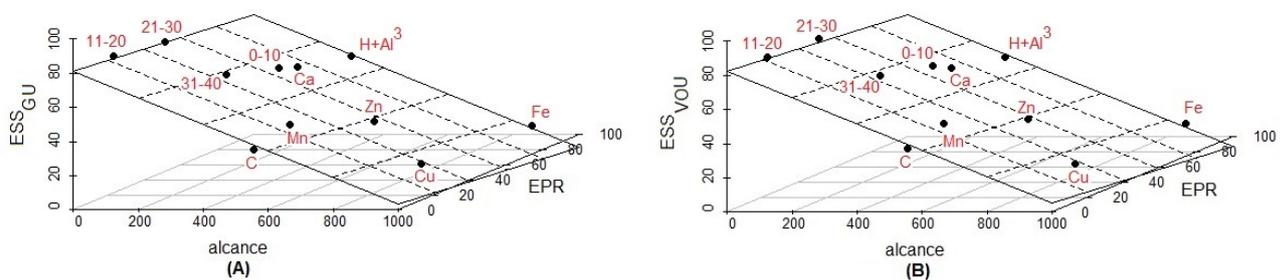


Figura 9 – Diagramas de dispersão dos valores estimados do tamanho amostral efetivo em função dos valores estimados de alcance prático e efeito pepita relativo (EPR) de acordo com (A) método de Griffith ( $\widehat{ESS}_{GU}$ ) e (B) método de Vallejos e Osorio ( $\widehat{ESS}_{VOU}$ ), para cada um dos seguintes atributos: C: carbono, Ca: cálcio, Cu: cobre, Fe: ferro,  $H+Al^3$ : ácido potencial total, Mn: manganês, Zn: zinco, e 0-10, 11-20, 21-30, 31-40: resistência do solo à penetração nas respectivas profundidades (em cm).

Em relação à redução no tamanho amostral, em ambas as metodologias, é possível observar que as variáveis se dividem em quatro grupos (Figura 9). O primeiro grupo, formado pelas variáveis teor de cobre ( $\widehat{ESS}_{GU} = 10$  e  $\widehat{ESS}_{VOU} = 11$  pontos amostrais) e teor de ferro ( $\widehat{ESS}_{GU} = 14$  e  $\widehat{ESS}_{VOU} = 16$  pontos amostrais), que apresentaram os menores valores estimados do ESS univariado. O segundo grupo agrega o teor de carbono ( $\widehat{ESS}_{GU} = 21$  e  $\widehat{ESS}_{VOU} = 23$  pontos amostrais), o teor de manganês ( $\widehat{ESS}_{GU} = 26$  e  $\widehat{ESS}_{VOU} = 28$  pontos amostrais) e o teor de zinco ( $\widehat{ESS}_{GU} = 20$  e  $\widehat{ESS}_{VOU} = 22$  pontos amostrais), e o terceiro grupo reúne os atributos teor de cálcio ( $\widehat{ESS}_{GU} = 51$  e  $\widehat{ESS}_{VOU} = 52$  pontos amostrais), teor de  $H+Al^3$  ( $\widehat{ESS}_{GU} = 52$  e  $\widehat{ESS}_{VOU} = 53$  pontos amostrais) e RSP nas camadas 0-10 cm ( $\widehat{ESS}_{GU} = 53$  e  $\widehat{ESS}_{VOU} = 55$  pontos amostrais) e 31-40 cm ( $\widehat{ESS}_{GU} = 59$  e  $\widehat{ESS}_{VOU} = 60$  pontos amostrais). Os grupos 2 e 3 exibiram valores estimados intermediários do ESS, sendo que, o segundo grupo foi o que apresentou os segundos menores valores estimados do ESS, enquanto no terceiro grupo as variáveis exibiram valores estimados do tamanho amostral efetivo iguais a aproximadamente 50% do número inicial de pontos amostrais (102). As variáveis da RSP nas profundidades 11-20 cm ( $\widehat{ESS}_{GU} = 88$  e  $\widehat{ESS}_{VOU} = 89$  pontos amostrais) e 21-30 cm ( $\widehat{ESS}_{GU} = 85$  e  $\widehat{ESS}_{VOU} = 88$  pontos amostrais) formaram um quarto grupo caracterizado por ter apresentado os maiores valores estimados do ESS.

O que diferencia tais grupos é, principalmente, o valor estimado do raio de dependência espacial. O primeiro grupo, o qual apresentou os menores valores estimados do ESS univariado,

exibiu os maiores alcances (538,7 e 855,1 m). Enquanto os mais baixos raios de dependência espacial (101,5 e 115,9 m) foram apresentados pelo quarto grupo, para o qual foram obtidos os maiores valores estimados do ESS univariado. No segundo e terceiro grupos, com valores intermediários do ESS univariado, apenas uma das variáveis apresentou alcance superior a 500 m. Nos demais atributos destes grupos os alcances práticos variaram entre 200 e 400 m, cujos valores, considerando a distância máxima entre pontos na área (cerca de 1760 m), são considerados baixos.

Como os alcances entre esses dois grupos foram próximos, a diferença no valor estimado do tamanho amostral efetivo univariado foi influenciada pela intensidade da dependência espacial. As variáveis do terceiro grupo apresentaram valores estimados maiores do ESS univariado comparados aos do segundo grupo e tiveram menor intensidade de dependência espacial (EPR maior).

No geral, verificou-se que o alcance influenciou inversamente o valor estimado do ESS univariado, isto é, quanto maior foi o raio de dependência espacial, menor foi o valor do tamanho amostral efetivo. Este resultado corrobora com os obtidos nos estudos de simulação do presente trabalho (Seção 5.1.3.1) e também com aqueles verificados nas simulações de Vallejos e Osorio (2014). Percebeu-se ainda que quando os valores dos alcances práticos foram próximos, a intensidade da dependência espacial exerceu influência direta, ou seja, quanto maior o EPR, maior o valor estimado do ESS univariado.

Por fim, assim como nas análises de dados simulados (Seção 5.1.3.1), os valores estimados do tamanho amostral efetivo para cada variável pelos métodos de Griffith (2005) e Vallejos e Osorio (2014) foram bem próximos.

Considerando-se que, na prática, a aplicabilidade do tamanho amostral efetivo é mais viável quando se agregam todas as variáveis físico-químicas do solo e obtém-se um único tamanho amostral reduzido para ser aplicado em experimentos futuros na área agrícola, são apresentados os valores estimados do tamanho amostral efetivo multivariado calculado pelas metodologias de Griffith (2005) e Vallejos e Osorio (2014) para duas situações multivariadas (Tabela 9).

Tabela 9 – Valores estimados do tamanho amostral efetivo multivariado pelos métodos de Griffith (2005) ( $\widehat{ESS}_{GM}$ ) e Vallejos e Osorio (2014) ( $\widehat{ESS}_{VOM}$ ).

	$\widehat{ESS}_{GM}$	$\widehat{ESS}_{VOM}$
<b>Situação 1</b>	29	53
<b>Situação 2</b>	31	54

Ao se comparar a situação 1 (na qual todos os atributos têm moderada/forte dependência espacial ou alcance alto) e a situação 2 (em que agregaram-se atributos com fraca dependência espacial e alcance baixo às variáveis da situação 1), ocorreu um pequeno aumento no número estimado de pontos amostrais, em ambos os métodos: de 29 para 31 pontos amostrais, via proposta de Griffith, e de 53 para 54 pontos amostrais, pelo método de Vallejos e Osorio.

Possivelmente o irrelevante aumento no número de pontos amostrais foi devido à diferente proporção de variáveis nas situações 1 e 2 com valores distintos do raio de dependência espacial. Isto porque a diferença entre os dois casos é que na situação 2 foram acrescentadas às 11 variáveis da situação 1, cujo alcance variou entre 101,5 e 968,5 m, apenas cinco atributos com alcance prático inferior a 20 metros.

Quando são comparados estes resultados com os obtidos nos estudos de simulação

(Seção 5.1.3.1), nos quais os cenários 6 e 7 apresentavam variáveis com características semelhantes respectivamente àquelas das situações 1 e 2, quanto ao raio de dependência espacial, os resultados foram semelhantes, ou seja, as conclusões obtidas nos estudos dos dados práticos corroboraram com aquelas verificadas nas simulações.

No contexto da Agricultura de Precisão, considerando-se a variabilidade espacial de variáveis físico-químicas do solo georreferenciadas, verificou-se que ambas as metodologias multivariadas propostas apresentaram redução no tamanho amostral multivariado. Pela metodologia de Vallejos e Osorio (2014), houve redução de 48 pontos amostrais (aproximadamente 50%) no número inicial de amostras (de 102 para 54), e uma diminuição de 71 pontos amostrais (cerca de 70%, de 102 para 31), pela proposta de Griffith (2005).

Dessa forma, em experimentos futuros, é possível descrever a estrutura de variabilidade espacial dos atributos físicos e químicos do solo na área agrícola estudada com um número menor de pontos amostrais. A princípio, para o pesquisador ou produtor rural, a metodologia de Griffith parece ser mais vantajosa para ser aplicada nos próximos levantamentos de dados na área agrícola, tendo em vista que proporciona maior redução no número de amostras a serem coletadas e, conseqüentemente, menores investimentos financeiros com as análises laboratoriais. No entanto, a redução acentuada no número de pontos amostrais, proporcionada pela proposta de Griffith, pode fazer com que os mapas temáticos construídos não sejam satisfatórios.

Em trabalho desenvolvido por Uribe-Opazo *et al.* (2006), verificou-se que um número insuficiente de amostras pode dificultar o ajuste de modelos aos pontos do semivariograma, pois maiores distâncias entre os pontos amostrais produzem estimativas incorretas das semivariâncias. Mesmo que os dados satisfaçam as condições de estacionariedade intrínseca e isotropia, uma quantidade muito reduzida de pontos amostrais pode ser insuficiente para que os semivariogramas experimentais evidenciem a estrutura de dependência espacial existente. Isto afeta a predição espacial e, conseqüentemente, a eficiência dos mapas obtidos.

Estudos práticos realizados por Coelho *et al.* (2009) com diferentes interpoladores constataram que uma redução de 75% no número de pontos amostrais não produz resultados aceitáveis nos mapas temáticos construídos. Já Souza *et al.* (2014) analisaram diferentes densidades amostrais e concluíram que uma redução acima de 73% no tamanho amostral não proporciona bons resultados na caracterização da variabilidade espacial de variáveis georreferenciadas. A pesquisa realizada por Kestring *et al.* (2015), na qual foram contrastadas distintas grades e densidades amostrais, evidenciou que mapas temáticos confeccionados com quantidade reduzida de pontos amostrais superior a 66%, não obtiveram índices de acurácia satisfatórios, corroborando com as constatações de Uribe-Opazo *et al.* (2006), Coelho *et al.* (2009) e Souza *et al.* (2014).

Sendo assim, a acentuada redução amostral é uma desvantagem do método de Griffith (2005) quando comparado ao de Vallejos e Osorio (2014), considerando que o mapa temático deve descrever o melhor possível a estrutura de variabilidade espacial das variáveis e, conseqüentemente, as características da área para que se possa realizar adequadamente o gerenciamento agrícola regionalizado (SOARES, 2014).

#### 5.1.4 Conclusões

Os resultados obtidos nos estudos dos atributos físicos e químicos do solo e dos dados simulados evidenciaram que, ao serem consideradas simultaneamente as características das variáveis (caso multivariado), houve redução no tamanho amostral. Ademais, para cada variável (caso univariado), também verificou-se redução no número de pontos amostrais. O valor estimado do ESS multivariado pelo método de Vallejos e Osorio foi maior em todos os contextos analisados, ou seja, a redução no tamanho amostral foi menor comparada à de Griffith. Isto porque mesmo que em ambas as propostas o intuito seja reduzir o tamanho amostral original com base nos efeitos da autocorrelação espacial, as heurísticas de cada metodologia são distintas.

O raio de dependência espacial foi o parâmetro que exerceu maior influência no valor do tamanho amostral efetivo. Em ambas as metodologias, quanto maior o valor do alcance, menor foi o valor estimado do ESS. Quando as variáveis apresentaram alcances próximos, a intensidade de dependência espacial influenciou o valor do ESS, portanto, quanto maior o EPR, isto é, menor a dependência espacial, maior foi o valor do ESS. Além disso, observou-se que a adição de variáveis, geralmente, proporcionou aumento no valor estimado do ESS multivariado.

A partir da ponderação dos resultados obtidos nas simulações e nos dados práticos, recomenda-se que seja considerado o número reduzido de pontos amostrais estimado pelo tamanho amostral efetivo multivariado em futuras coletadas de dados agrícolas para caracterização da variabilidade espacial de variáveis georreferenciadas. Entretanto, embora a proposta de Griffith seja vantajosa no sentido de proporcionar maior redução no tamanho amostral, não é regra que, para qualquer situação prática, a metodologia que sugere a maior redução no número de pontos amostrais é a mais eficiente quanto ao estudo da variabilidade espacial das variáveis georreferenciadas. Isto porque reduções muito acentuadas podem gerar mapas temáticos cuja representatividade da variabilidade espacial dos atributos na área não é satisfatória, prejudicando assim o gerenciamento agrícola regionalizado.

#### Agradecimentos

Os autores agradecem inicialmente aos membros do Laboratório de Estatística Espacial (LEE) e do Laboratório de Estatística Aplicada (LEA) pelo banco de dados disponibilizado, e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

#### 5.1.5 Referências

APARECIDO, L. E. O.; ROLIM, G. S.; RICHETTI, J.; SOUZA, P. S.; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, v. 40, n. 4, p. 405-417, 2016.

BERNARDI, A. C. C.; NAIME, J. M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. Agricultura de Precisão. In: INAMASU, R. Y.; BERNARDI, A. C. C. **Agricultura de Precisão**: resultados de um novo olhar. São Carlos: EMBRAPA, 2014. cap. 1, p. 21-33.

BARTOSZEK, K. Phylogenetic effective sample size. **Journal of Theoretical Biology**, v. 407, p. 371-386, 2016.

CAMBARDELLA, C. A.; MOORMAN, T.; PARKIN, T.; KARLEN, D.; NOVAK, J.; TURCO, R.; KONOPKA, A. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, v. 58, n. 5, p. 1501-1511, 1994.

CHERUBIN, M. R.; SANTI, A. L.; EITELWEIN, M. T.; MENEGOL, D. R.; DA 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.

COELHO, E. C.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NETO, R. P. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum. Agronomy**, v. 31, n. 1, p. 165-174, 2009.

CRESSIE, N. A. C. **Statistics for Spatial Data**. rev. ed. New York: John Wiley & Sons, 2015. 928p.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; JOHANN, J. A.; GALEA, M.; DE BASTIANI, F. Gaussian spatial linear model of soybean yield using bootstrap methods. **Engenharia Agrícola**, v. 38, p. 110-116, 2018.

DANCEY, C. P.; REIDY, J. **Estatística sem Matemática para Psicologia**. 5. ed. Porto Alegre: Penso, 2013. 608p.

DE BASTIANI, F.; CYSNEIROS, A. F. J.; CYSNEIROS, A. H. M.; URIBE-OPAZO, M. A.; GALEA, M. Influence diagnostics in elliptical spatial linear models. **Test**, v. 24, n. 2, p. 322-340, 2015.

DIGGLE, P.; LOPHAVEN, S. Bayesian geostatistical design. **Scandinavian Journal of Statistics**, v. 33, n. 1, p. 53-64, 2006.

DIGGLE, P. J.; RIBEIRO JR, P. J. **Model-based Geostatistics**. New York: Springer Series in Statistics, 2007. 230p.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 3ed. Brasília: Centro Nacional de Pesquisa de Solos, 2013. 306p.

FAGUNDES, R. S.; URIBE-OPAZO, M. A.; GUEDES, L. P. C.; GALEA, M. Slash spatial linear modeling: soybean yield variability as a function of soil chemical properties. **Revista Brasileira de Ciência do Solo**, v. 42, 2018.

FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A.; JOHANN, J. A.; BORSSOI, J. A. Selection criteria of spatial variability models used in thematical maps of soil physical attributes and soybean yield. **Revista Brasileira de Ciência do Solo**, v. 32, n. 2, p. 463-476, 2008.

GAO, X.; ZHAO, X.; WU, P.; BROCCA, L.; ZHANG, B. Effects of large gullies on catchment-scale soil moisture spatial behaviors: a case study on the Loess Plateau of China. **Geoderma**, v. 261, p. 1-10, 2016.

GRIFFITH, D. A. Effective geographic sample size in the presence of spatial autocorrelation. **Annals of the Association of American Geographers**, v. 95, n. 4, p. 740-760, 2005.

GUEDES, L. P. C.; RIBEIRO JR, P. J.; PIEDADE, S. M. S.; URIBE-OPAZO, M. A. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. **Chilean Journal of Statistics (ChJS)**, v. 2, n. 2, 2011.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; RIBEIRO JR, P. J. Influence of incorporating geometric anisotropy on the construction of thematic maps of simulated data and chemical attributes of soil. **Chilean Journal of Agricultural Research**, v. 73, n. 4, p. 414-423, 2013.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; RIBEIRO JR, P. J. Optimization of sample design sizes and shapes for regionalized variables using simulated annealing. **Ciencia e Investigação Agrária**, v. 41, n. 1, p. 33-48, 2014.

GUEDES, L. P. C.; RIBEIRO JR, P. J.; URIBE-OPAZO, M. A.; DE BASTIANI, F. Soybean yield maps using regular and optimized sample with different configurations by simulated annealing. **Engenharia Agrícola**, v. 36, n. 1, p. 114-125, 2016.

KESTRING, F.; GUEDES, L. P. C.; DE BASTIANI, F.; URIBE-OPAZO, M. A. Thematic maps comparison of different sampling grids for soybean productivity. **Engenharia Agrícola**, v. 35, n. 4, p. 733-743, 2015.

LANDIM, P. M. B.; YAMAMOTO, J. K. **Geoestatística**: conceitos e aplicações. São Paulo: Oficina de Textos, 2013. 215p.

LIU, G. F. A note on effective sample size for constructing confidence intervals for the difference of two proportions. **Pharmaceutical Statistics**, v. 11, n. 2, p. 163-169, 2012.

MARTINO, L.; ELVIRA, V.; LOUZADA, F. Effective sample size for importance sampling based on discrepancy measures. **Signal Processing**, v. 131, p. 386-401, 2017.

MATÉRN, B. **Spatial Variation**: lecture notes in statistics, v. 36. New York: Springer, 2013. 156p.

MENEZES, M. D.; SILVA, S. H. G.; MELLO, C. R.; OWENS, P. R.; CURTI, N. Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. **Scientia Agrícola**, v. 73, n. 3, p. 274-285, 2016.

MULLER, A. **Simulação Estocástica**: o método de Monte Carlo. 2008. 75f. Monografia (Bacharelado

em Estatística) - Universidade Federal do Paraná, Curitiba, 2008.

PIMENTEL GOMES, F. **Curso de Estatística Experimental**. São Paulo: Nobel, 1985. 467p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R foundation for Statistical Computing, Vienna, Austria, 2017.

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

RUIZCÁRDENAS, R.; FERREIRA, M. A. R.; SCHMIDT, A. M. Stochastic search algorithms for optimal design of monitoring networks. **Environmetrics**, v. 21, n. 1, p. 102-112, 2010.

SOARES, A. **Geoestatística para Ciências da Terra e do Ambiente**. 3. ed. Lisboa: Press, 2014. 214p.

SOUZA, Z. M.; SOUZA, G. S.; MARQUES JÚNIOR, 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, 2014.

URIBE-OPAZO, M. A.; BORSSOI, J. A.; GALEA, M. Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics**, v. 39, n. 3, p. 615-630, 2012.

URIBE-OPAZO, M. A.; JOHANN, J. A.; VILAS BOAS, M. A.; BORSSOI, J. A.; LUNKES, C. Fatores que influenciam na estimação de parâmetros no ajuste de modelos teóricos à semivariogramas experimentais no estudo da produtividade de soja. In: VII CONGRESSO LATINO AMERICANO E DO CARIBE DE ENGENHARIA AGRÍCOLA E V CONGRESSO INTERNACIONAL DE ENGENHARIA AGRÍCOLA, 2006, Chile. **Anais do VII CLIA e V CIACH**. Chile: 2006, p. 1-8.

VALLEJOS, R.; OSORIO, F. Effective sample size of spatial process models. **Spatial Statistics**, v. 9, p. 66-92, 2014.

YANG, Y.; REMMERS, E. F.; OGUNWOLE, C. B.; KASTNER, D. L.; GREGERSEN, P. K.; LI, W. Effective sample size: quick estimation of the effect of related samples in genetic case-control association analyses. **Computational Biology and Chemistry**, v. 35, n. 1, p. 40-49, 2011.

## 5.2 ARTIGO 2

### REDIMENSIONAMENTO AMOSTRAL PARA A ANÁLISE DA VARIABILIDADE ESPACIAL DA RESISTÊNCIA DO SOLO À PENETRAÇÃO COM DISTRIBUIÇÃO DE PROBABILIDADE T-STUDENT

#### Resumo

Modelar a dependência espacial a partir de elementos amostrais de variáveis georreferenciadas e caracterizar a variabilidade espacial destas podem ser tarefas árduas, principalmente quando são consideradas as diversas particularidades dos fenômenos, tal como a presença de pontos discrepantes. Nestes casos, recomenda-se, de acordo com a literatura, o uso de uma distribuição de probabilidade, tal como a t-Student, que agregue parâmetros capazes de reduzir a influência das observações atípicas. Além disso, considerando-se o tempo investido na coleta de dados no campo e os elevados gastos com as análises laboratoriais das amostras, diversas pesquisas tem sido desenvolvidas com o intuito de reduzir o número de pontos amostrais coletados. Portanto, este trabalho visa redefinir o número de pontos amostrais coletados em uma área agrícola, a partir da redução no tamanho amostral, para o estudo da variabilidade espacial da resistência do solo à penetração (RSP), nas camadas 0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm de profundidade, que apresentam observações discrepantes. Para isso, foi utilizada a metodologia do tamanho amostral efetivo (ESS) univariado, considerando-se que a RSP apresentou distribuição de probabilidade t-Student ( $ESS_t$ ). Utilizaram-se dados simulados para dar maior embasamento científico aos resultados do  $ESS_t$  univariado, obtidos com a RSP, e para tal atribuiu-se às variáveis simuladas valores diferenciados de efeito pepita e alcance. Constatou-se tanto para as variáveis simuladas quanto para as diferentes camadas de RSP uma redução de 40% a 95% no número de pontos amostrais. Essa variação na redução do tamanho amostral se deve aos diferentes valores apresentados pelos parâmetros que definem a estrutura de dependência espacial, principalmente, das variáveis simuladas. Com base nas camadas da RSP em que se identificou dependência espacial (0-10 cm e 31-40 cm de profundidade) e considerando-se o máximo valor estimado do  $ESS_t$  nestas camadas, selecionou-se uma nova grade com 60 pontos amostrais. Após a realização das análises estatísticas e geoestatísticas para as malhas original e reduzida, e comparação dos mapas temáticos de cada camada da RSP entre estas malhas, verificaram-se diferenças entre os mapas confeccionados com a configuração amostral original (102 pontos amostrais) e reduzida (60 pontos amostrais). Contudo, mesmo com essa dessemelhança entre as grades amostrais, o índice tau evidenciou que os mapas apresentaram de média a alta acurácia em todas as camadas da RSP, com exceção da camada 11-20 cm de profundidade.

**Palavras-chave:** dependência espacial, geoestatística, tamanho amostral efetivo, variáveis físicas do solo.

#### 5.2.1 Introdução

Haja vista a crescente necessidade de aumentar a produção de grãos ocasionada pela ampla concorrência no agronegócio e pela limitação na abertura de novas áreas de plantio, tem-se investido fortemente no desenvolvimento de técnicas que permitam elevar a produtividade das

culturas (JOHANN *et al.*, 2004). A Agricultura de Precisão (AP) é tida como alternativa, pois viabiliza o gerenciamento do manejo do solo visando corrigir atributos que comprometem o rendimento, e, desta forma, possibilitam o aumento do potencial produtivo, além de contribuírem com uma exploração agrícola mais sustentável (PATERNIANI, 2001; BERNARDI *et al.*, 2014).

Os fundamentos da AP baseiam-se na aplicação de fertilizantes e corretivos no solo a taxas variáveis, isto é, de acordo com as demandas locais e não de modo uniforme por toda a área (MOLIN, AMARAL e COLAÇO, 2015). Para tal, é indispensável o conhecimento adequado do ambiente agrícola em relação à variabilidade espacial dos atributos físicos e químicos do solo. Assim, propõem-se as melhores estratégias para o manejo sustentável do solo, visando à sustentabilidade agronômica e econômica desse (ARAÚJO *et al.*, 2014). Dentre os atributos do solo, as variáveis físicas desempenham importantes funções (WARRICK e NIELSEN, 1980). A resistência do solo à penetração é uma variável física particularmente relevante, pois está inversamente vinculada ao crescimento das raízes e à produtividade das culturas (RODRIGUES *et al.*, 2014). E para que se possa mapear as regiões agrícolas utilizando as informações acerca da variabilidade espacial das variáveis do solo, é necessário realizar uma amostragem destes atributos em alguns pontos georreferenciados da área. O objetivo é caracterizar toda a área agrícola a partir de uma pequena porção representativa desta, a qual é dada pelos pontos amostrais, tendo em vista que avaliar um campo de produção em sua totalidade é uma tarefa economicamente inviável (MOLIN, AMARAL e COLAÇO, 2015).

Haja vista os impactos financeiros causados ao produtor pelos elevados custos de coleta e análise laboratorial das amostras, na literatura, são propostos métodos que possibilitam reduzir o número de pontos amostrais a ser utilizado em experimentos futuros na área agrícola, sem que haja perda considerável no mapeamento dela (GRIFFITH, 2005; RUIZ-CÁRDENAS, FERREIRA e SCHMIDT, 2010; GUEDES *et al.*, 2011, 2016; GUEDES, URIBE-OPAZO e RIBEIRO JR, 2014). Uma das propostas é o tamanho amostral efetivo - ESS (*Effective Sample Size*, em inglês), tal metodologia considera que alguns pontos amostrais coletados podem estar altamente correlacionados entre si, proporcionando custos desnecessários com coleta e análises laboratoriais, pois tais pontos agregam informações repetidas em relação à variabilidade espacial (VALLEJOS E OSORIO, 2014). Isto é, o ESS possibilita estimar um novo tamanho amostral a partir dos efeitos da autocorrelação espacial. Logo, o principal objetivo é utilizar este novo número de amostras para estimar a média amostral do valor da variável físico-química da maneira mais precisa possível (GRIFFITH, 2005).

O método desenvolvido por Griffith (2005) para estimação do ESS univariado tem como premissa que as variáveis seguem uma distribuição normal de probabilidade. Com base na proposta de Griffith, Vallejos e Osorio (2014) sugeriram outra abordagem mais abrangente para calcular o valor estimado do ESS univariado, na qual pressupõe-se que as variáveis têm distribuição de probabilidade elíptica. A maior abrangência em relação à distribuição probabilística das variáveis é interessante porque, quando se estuda a variabilidade espacial de variáveis georreferenciadas, pode ocorrer que elas não apresentem distribuição normal de probabilidade, especialmente devido ao fato de tal distribuição ser sensível a pontos discrepantes (SCHEMMER *et al.*, 2017). Assim a família de contornos elípticos, da qual a distribuição normal é um caso particular, permite estender a classe dos erros a outras distribuições de probabilidade, de modo a melhor acomodar as observações atípicas (DE BASTIANI *et al.*, 2015).

Assim, este trabalho objetiva redefinir o número de pontos amostrais da variável resistência

do solo à penetração (RSP) coletados em uma área agrícola, utilizando a metodologia de Vallejos e Osorio (2014). Optou-se pela RSP, pois além da contribuição que o parecer da variabilidade espacial dessa variável representa em termos de melhoramento no manejo do solo (CARVALHO *et al.*, 2013; TAVARES *et al.*, 2014), ela também apresentou pontos discrepantes em todas as camadas de profundidade analisadas. Sendo assim, assumiu-se que as camadas da RSP apresentavam distribuição de probabilidade elíptica para estimar o valor do tamanho amostral efetivo univariado ( $ESS_t$ ). Para maior embasamento científico nas discussões dos resultados da RSP, estimou-se ainda o valor do  $ESS_t$  de dados simulados, cujas variáveis apresentavam diversos valores de efeito pepita e alcance, e explorou-se a influência que os parâmetros de dependência espacial das variáveis exercem no valor do  $ESS_t$  univariado obtido. Ao final, compararam-se os resultados da malha amostral original e reduzida a fim de se analisar o efeito da redução no tamanho amostral na descrição da estrutura de dependência espacial das variáveis georreferenciadas.

## 5.2.2 Material e Métodos

### 5.2.2.1 Descrição das simulações

Considere um processo estocástico  $\{Y(\mathbf{s}_i), \mathbf{s}_i \in S \subset \mathbb{R}^2\}$ ,  $i = 1, \dots, n$ , estacionário e isotrópico. Sendo que  $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$  é um vetor aleatório, de dimensão  $n \times 1$ , em que  $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)$  são os valores observados da variável aleatória sob estudo nas  $\mathbf{s}_i$  localizações espaciais amostradas,  $i = 1, \dots, n$ . Suponha que  $\mathbf{Y}$  tem distribuição t-Student  $n$ -variada, isto é,  $\mathbf{Y} \sim t_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, v)$ , em que  $\boldsymbol{\mu}$ , de dimensão  $n \times 1$ , é o vetor de médias;  $\boldsymbol{\Sigma}$ , de dimensão  $n \times n$ , é a matriz escala; e  $v$  é o grau de liberdade,  $v > 0$ , fixo (ASSUMPTÃO, URIBE-OPAZO e GALEA, 2014).

Cada observação  $Y(\mathbf{s}_i)$  é gerada pelo seguinte modelo espacial linear (GUEDES, URIBE-OPAZO e RIBEIRO JR, 2013):

$$Y(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i), \quad (5.9)$$

em que  $\mu(\mathbf{s}_i)$ ,  $i = 1, \dots, n$  é o termo determinístico, e  $\epsilon(\mathbf{s}_i)$  é o termo estocástico, sendo que ambos dependem da localização espacial  $\mathbf{s}_i$  em que a variável  $Y$  é observada.

Considera-se  $\mu(\mathbf{s}_i) = \mathbf{X}\boldsymbol{\beta}$ , sendo que:  $\mathbf{X}$ , de ordem  $n \times p$ , é denominada matriz de delineamento, na qual a  $i$ -ésima linha é composta por um vetor  $\mathbf{x}_i^T = ([x_m(\mathbf{s}_i)])$ , ( $i = 1, \dots, n; m = 1, \dots, p$ ), em que  $x_m(\mathbf{s}_i)$  representa o valor da  $m$ -ésima covariável tomada na  $i$ -ésima posição; e  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$  é um vetor de parâmetros desconhecidos a serem estimados (SCHEMMER *et al.*, 2017).

O vetor aleatório  $\mathbf{Y}$  pode ser reescrito matricialmente da seguinte forma (URIBE-OPAZO, BORSSOI E GALEA, 2012):

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (5.10)$$

em que  $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$  é o vetor de erros aleatórios que representa o termo estocástico, sendo que  $E(\boldsymbol{\epsilon}) = \mathbf{0}$ . Para o caso em que se considera a média constante, tem-se que  $\mathbf{X}\boldsymbol{\beta} = \mu\mathbf{1}$ , sendo  $\mathbf{1}$  um vetor unitário cuja dimensão depende da quantidade  $n$  de pontos amostrais e  $\mu$  o valor da média.

Assume-se que a matriz escala de  $\boldsymbol{\epsilon}$  é dada por  $\boldsymbol{\Sigma} = [(\sigma_{ij})]$ , em que  $\sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$ ,  $i, j = 1, \dots, n$ , é a covariância entre os valores da variável georreferenciada nos pontos  $\mathbf{s}_i$  e  $\mathbf{s}_j$ . Tem-se ainda que  $\boldsymbol{\Sigma}$  é inversível e pode ser escrita na sua forma paramétrica (DE BASTIANI *et al.*,

2015):

$$\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3), \quad (5.11)$$

sendo que:  $\varphi_1$  e  $\varphi_2$  são os parâmetros efeito pepita e contribuição, respectivamente;  $\mathbf{I}_n$  é a matriz identidade de ordem  $n \times n$ ;  $\mathbf{R}(\varphi_3) = [(r_{ij})]$  é uma matriz  $n \times n$  simétrica, em que  $\varphi_3$  é um parâmetro que está em função do alcance prático ( $a$ ), isto é,  $g(\varphi_3) = a$ .

Considerando-se o modelo espacial linear definido (Equações 5.9 e 5.10), as simulações foram obtidas a partir de um experimento de Monte Carlo, utilizando-se a decomposição de Cholesky, com base na configuração amostral da área agrícola comercial que foi considerada nessa pesquisa (Figura 10), limitando-se as coordenadas e o alcance com base nesta área.

Foram consideradas 11 variáveis com diferentes raios e intensidades de dependência espacial que foram obtidos pelos valores pré-definidos dos parâmetros (Tabela 10) que definem a estrutura de dependência espacial do modelo espacial linear (Equação 5.11). Realizaram-se 100 simulações para cada variável, com grau de liberdade fixado ( $v = 5$ ). Para cada variável, foram fixadas também a média ( $\mu = 5$ ) e a contribuição ( $\varphi_2 = 1$ ), e considerou-se o ajuste do modelo exponencial à função semivariância.

Tabela 10 – Valores dos parâmetros que definem a estrutura de dependência espacial das variáveis simuladas, efeito pepita e função de alcance, além do raio e intensidade de dependência espacial de cada variável.

Variáveis	$\varphi_1$	$\varphi_3$	$a$	EPR
V1	0,10	0,10	0,3	9%
V2	0,50	0,10	0,3	33%
V3	0,10	0,17	0,5	9%
V4	0,50	0,17	0,5	33%
V5	0,10	0,20	0,6	9%
V6	0,50	0,20	0,6	33%
V7	0,03	0,40	1,2	3%
V8	0,05	0,33	1,0	5%
V9	0,80	0,17	0,5	44%
V10	0,03	0,36	1,1	3%
V11	0,07	0,33	1,0	6%

$\varphi_1$ : efeito pepita,  $\varphi_3$ : função do alcance,  $a$ : alcance prático (em km), EPR =  $100 \frac{\varphi_1}{\varphi_1 + \varphi_2}$ : efeito pepita relativo (em %).

### 5.2.2.2 Descrição dos dados experimentais

A coleta dos dados foi realizada em uma área comercial com 167,35 hectares, cultivada com soja, localizada no município de Cascavel-PR, com coordenadas geográficas aproximadas de latitude 24,95° Sul e 53,37° Oeste, e 650 metros de altitude média. O clima da região é identificado como temperado mesotérmico e superúmido, tipo climático Cfa (Koeppen), com temperatura anual média de 21°C (APARECIDO *et al.*, 2016). O solo é classificado como Latossolo Vermelho Distroférico, com textura argilosa (EMBRAPA, 2013).

Nesta área, foi empreendida uma amostragem *lattice plus close pairs* composta por 102 pontos amostrais (DIGGLE e LOPHAVEN, 2006; DIGGLE e RIBEIRO JR, 2007; GUEDES *et al.*, 2011), em que a distância mínima entre pontos da grade regular contida nessa amostragem foi de 141 metros. Os 102 pontos amostrais da *lattice plus close pairs* compreendem tanto a grade regular quanto as 19 localidades que foram adicionadas aleatoriamente a esta, apresentando distâncias

menores de 50 e 75 m entre tais locais e a observação pertencente à grade regular (Figura 10). As amostras foram georreferenciadas e localizadas com auxílio de um aparelho receptor de sinal com o sistema de posicionamento global (GPS) GEOEXPLORE 3, em um sistema espacial de coordenadas UTM.

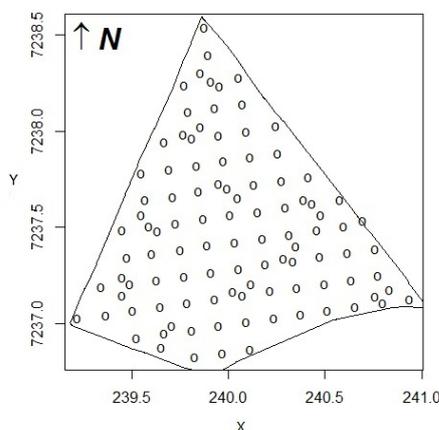


Figura 10 – Área experimental com a localização dos pontos amostrados.

Neste trabalho, foram utilizados os dados da resistência do solo à penetração das raízes (MPa) nas profundidades 0-10 cm (RSP 0-10 cm), 11-20 cm (RSP 11-20 cm), 21-30 cm (RSP 21-30 cm) e 31-40 cm (RSP 31-40 cm). Os dados experimentais deste atributo físico são referentes ao ano-safra 2015-2016 e pertencem ao banco de dados do Laboratório de Estatística Espacial (LEE) e do Laboratório de Estatística Aplicada (LEA) da Universidade Estadual do Oeste do Paraná - UNIOESTE, *Campus Cascavel*.

As determinações da resistência do solo à penetração foram realizadas utilizando o penetrógrafo da seguinte maneira: em cada ponto amostral foram efetuadas três leituras a cada centímetro, de 0 a 40 cm, cobrindo assim as quatro profundidades consideradas (0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm). Os dados obtidos foram transformados em MPa e, posteriormente, estabeleceu-se o valor da resistência do solo à penetração em cada camada fazendo a média aritmética das três medições.

### 5.2.2.3 Análises estatísticas e geoestatísticas e cálculo do tamanho amostral efetivo

A partir da malha inicial de 102 pontos, para cada camada da resistência do solo à penetração, realizaram-se as análises exploratória e geoestatística. Na análise geoestatística, inicialmente foram construídos gráficos post-plot, e os valores mensurados de cada atributo foram classificados de acordo com os quartis. Para identificar a estrutura de dependência espacial, foi ajustado à função semivariância o modelo teórico da família Matérn, atribuindo-se diferentes valores para o parâmetro de forma  $\kappa$  ( $\kappa = 0,5; 1,5; 2,5$ ), sendo que os parâmetros de dependência espacial foram estimados pelo método de máxima verossimilhança utilizando-se o algoritmo EM (ASSUMPÇÃO, URIBE-OPAZO E GALEA, 2014). Consideraram-se ainda para cada valor do parâmetro  $\kappa$  três diferentes graus de liberdade fixos,  $\nu = 3, 5$  e  $7$ . E, para determinar o melhor grau de liberdade ( $\nu$ ) em cada  $\kappa$ , foram utilizados os critérios de validação cruzada e traço (DE BASTIANI *et al.*, 2015). Posteriormente, tendo escolhido o melhor parâmetro de forma  $\nu$ , fez-se uso do erro padrão para determinar o melhor

modelo teórico ajustado a cada variável, isto é, o melhor parâmetro  $\kappa$  (SCHEMMER *et al.*, 2017). Por fim, de acordo com a estrutura de variabilidade espacial do modelo estimado escolhido, realizou-se a predição espacial, na qual se estimou o valor da resistência do solo à penetração nas quatro profundidades em pontos não amostrados, utilizando para tal a krigagem ordinária (SOARES, 2014), para construção dos mapas temáticos.

Na análise geostatística dos dados simulados foi realizado o ajuste do modelo exponencial à função semivariância e utilizou-se o método de máxima verossimilhança via algoritmo EM com grau de liberdade fixo ( $v = 5$ ) para estimar os parâmetros efeito pepita e alcance, haja vista os valores de média e contribuição terem sido fixados (ASSUMPÇÃO, URIBE-OPAZO e GALEA, 2014).

É necessário que já tenha sido realizado pelo menos um levantamento amostral na área agrícola para se calcular o valor do tamanho amostral efetivo. Desta forma, são conhecidos previamente o número de pontos amostrais e a disposição destes no campo, além da estrutura de dependência espacial da variável georreferenciada. Geralmente, quando não se tem conhecimento prévio da área e se está iniciando a coleta de dados, o tamanho amostral inicial pode ser determinado ponderando-se a dimensão da área, os custos relacionados à amostragem e os recursos disponíveis (FLEISCHER, BLOM e WEISZ, 1999; DOBBIE, HENDERSON e STEVENS, 2008). Em uma primeira coleta de pontos amostrais na área, pode-se empregar uma amostragem probabilística como critério para decidir o tamanho amostral. Desenvolvida por Cordy e Thompson (1995), essa amostragem considera um universo contínuo e espacialmente distribuído, sendo que tal universo pode representar uma região geográfica. Em suma, segundo essa metodologia, define-se uma função densidade de probabilidade com base em alguma configuração amostral (aleatória, estratificada, regular etc), e a partir dela se calcula a função densidade de inclusão que mensura a quantidade de pontos amostrais por unidade de área (CORDY e THOMPSON, 1995). Outra opção mais simples, porém menos rigorosa, a ser considerada para uma primeira amostragem de solo, é seguir umas das recomendações existentes na Agricultura de Precisão, na qual se sugere de 3 a 5 hectares por amostra (MOLIN, AMARAL e COLAÇO, 2015).

O valor do tamanho amostral efetivo ( $ESS_t$ ) foi estimado para cada variável simulada e para cada camada da resistência do solo à penetração (0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm de profundidade), utilizando-se a metodologia univariada proposta por Vallejos e Osorio (2014), e considerando-se que o vetor  $\mathbf{Y}$  (Equação 5.10), que representa a variável georreferenciada, tem distribuição de probabilidade t-Student. O valor estimado do  $ESS_t$  univariado é obtido pela seguinte equação (VALLEJOS e OSORIO, 2014):

$$\widehat{ESS}_t = \frac{v + n}{v + n + 2} \mathbf{1}^T \mathbf{R}(\hat{\varphi})^{-1} \mathbf{1}, \quad (5.12)$$

em que:  $n$  é o número de pontos amostrais na malha inicial ( $n \geq 1$ );  $v$  é o grau de liberdade ( $v > 2$ );  $\mathbf{1}$  é um vetor unitário, de dimensão  $n \times 1$ , e  $\mathbf{R}(\hat{\varphi})$  é a matriz de correlação espacial dos pontos amostrais, de dimensão  $n \times n$ , cujos elementos  $R(\hat{\varphi})_{ij}$ ,  $i, j = 1, \dots, n$ , são dados por (VALLEJOS e OSORIO, 2014):

$$R(\hat{\varphi})_{ij} = \begin{cases} 1, & \text{se } i = j \\ \frac{\hat{\varphi}_2 r_{ij}}{\hat{\varphi}_1 + \hat{\varphi}_2}, & \text{se } i \neq j, \end{cases} \quad (5.13)$$

em que:  $R(\hat{\varphi})_{ij}$ ,  $i, j = 1, \dots, n$ , é o valor estimado da correlação espacial existente entre o  $i$ -ésimo e o  $j$ -ésimo ponto amostral; e  $r_{ij}$  são os elementos da matriz  $\mathbf{R}(\varphi_3)$ , cujo cálculo está vinculado ao

modelo geoestatístico utilizado e à distância euclidiana entre as observações. Tem-se ainda que  $\hat{\varphi}_2$  e  $\hat{\varphi}_1 + \hat{\varphi}_2$  são os valores estimados dos parâmetros contribuição e patamar, nesta ordem.

O que difere a Equação 5.12 da equação utilizada para estimar o tamanho amostral efetivo univariado considerando vetores aleatórios com distribuição normal de probabilidade (ESS), é a constante  $\frac{v+n}{v+n+2}$ , a qual é proveniente da matriz de informação de Fisher para modelos espaciais lineares com distribuição t-Student (DE BASTIANI *et al.*, 2015). Como  $v > 2$  e  $n \geq 1$ , tem-se que  $v + n + 2 > v + n$ , então  $ESS_t$  é necessariamente menor que ESS.

Dados os valores estimados do tamanho amostral efetivo ( $ESS_t$ ) em cada camada da resistência do solo à penetração, o intuito foi redefinir um tamanho amostral reduzido único. Primeiramente, para essa escolha, foram considerados os valores estimados do  $ESS_t$  provenientes de variáveis com estrutura de dependência espacial. Isto porque, segundo Vallejos e Osorio (2014), atributos cujas observações são espacialmente independentes tendem a não apresentar redução no número de pontos amostrais. Em seguida, dentre as variáveis com dependência espacial, tomou-se o maior valor que foi estimado do  $ESS_t$  univariado. Este segundo critério foi determinado, pois para um maior número de pontos amostrais, a tendência é que se obtenham mapas temáticos mais representativos em relação à variabilidade espacial do atributo na área agrícola (KESTRING *et al.*, 2015).

Para verificar se o tamanho amostral reduzido foi adequado, dentre os 102 pontos da malha amostral original, selecionou-se uma amostra aleatória, a partir do novo tamanho amostral válido para todas as camadas da RSP. Para esta amostra de tamanho reduzido, realizou-se a análise exploratória e geoestatística, de forma análoga às análises realizadas para a malha original, e contrastaram-se os resultados obtidos nestas análises com aquelas em que utilizou-se o tamanho amostral original. A comparação se faz necessária porque a diminuição no número de amostras tende a proporcionar um aumento nas distâncias entre os pontos amostrais e, conseqüentemente, atribuem-se pesos menores às amostras vizinhas na predição espacial. Tal fato contribui para uma variância do erro mais elevada e mapas temáticos menos representativos quanto à variabilidade espacial da área agrícola (SOARES, 2014).

Para possibilitar o contraste entre os mapas temáticos confeccionados com o tamanho amostral original e reduzido das diferentes camadas de profundidade da RSP, foram classificados os 66935 pixels da grade utilizando o mesmo intervalo de classe nos mapas com malha original e reduzida. Optou-se por cinco classes, haja vista essa quantidade proporcionar uma identificação visual das regiões na área agrícola e viabilizar a aplicação localizada de insumos. Um grande número de intervalos resulta em regiões pequenas e dificulta o manejo localizado do solo (DAL POSSO *et al.*, 2012).

A comparação quantitativa entre os mapas temáticos com a configuração amostral original e reduzida foi realizada utilizando-se medidas de acurácia obtidas a partir da matriz de erros (GRINAND *et al.*, 2008), tendo em vista que a subjetividade das comparações visuais podem fazer com que importantes diferenças entre os mapas sejam ignoradas (DALPOSSO *et al.*, 2012). Definiu-se o mapa gerado com a quantidade original de pontos amostrais (102) como mapa referência, e aquele obtido com a malha reduzida foi definido como mapa modelo. A Exatidão Global (ANDERSON *et al.*, 2001) e o índice Tau (MA e REDMOND, 1995; KRIPPENDORFF, 2004) foram utilizadas como medidas de acurácia.

As análises estatísticas e geoestatísticas, a realização das simulações e a estimação dos

valores do tamanho amostral efetivo foram obtidas por meio do *software* R (R DEVELOPMENT CORE TEAM, 2017) utilizando o pacote *geoR* (RIBEIRO JR e DIGGLE, 2001).

### 5.2.3 Resultados e discussões

#### 5.2.3.1 Estudo de simulação

Considerando-se as 100 simulações de cada variável, são apresentados os gráficos com as médias e desvios padrão dos valores estimados do tamanho amostral efetivo univariado, segundo a metodologia de Vallejos e Osorio (2014) (Figura 11).

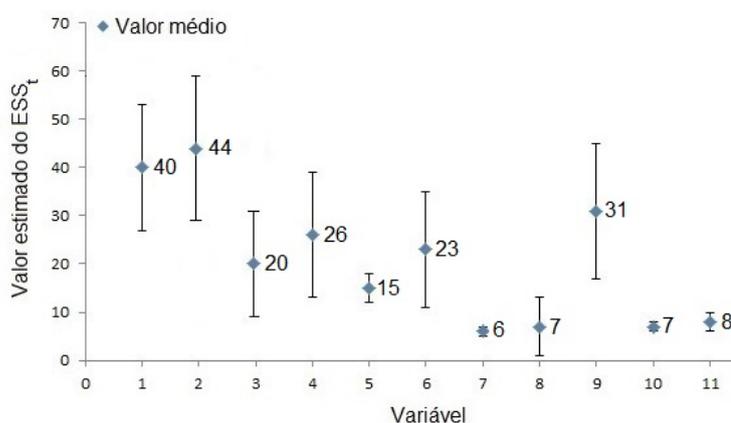


Figura 11 – Gráfico com a média e desvio padrão dos valores estimados do tamanho amostral efetivo ( $ESS_t$ ) para cada variável, considerando-se todas as simulações

Verificou-se que a variação do valor do efeito pepita não gerou relevante mudança no valor estimado do tamanho amostral efetivo (Figura 11). Isto porque, para cada par de variáveis com alcance prático fixo, a maior diferença no valor médio estimado do  $ESS_t$  foi de onze pontos amostrais, a qual foi obtida pela comparação das variáveis V3 e V9, cujo alcance prático foi fixado em 0,6 km e o valor do efeito pepita tem uma diferença de 0,7. Na prática, onze pontos amostrais não é uma diferença tão considerável, além disso, os elevados desvios padrão fazem com que, basicamente, não haja diferença estatística entre os valores estimados do  $ESS_t$  de V3 e V9. Os outros pares de variáveis os valores do efeito pepita foram diferenciados e manteve-se o alcance fixado (V1 e V2, V3 e V4, V4 e V9, V5 e V6, V8 e V11), nos quais a diferença entre os valores estimados do  $ESS_t$  foi ainda menor, observaram-se situações semelhantes quanto ao desvio padrão.

Contudo, os valores estimados do tamanho amostral efetivo univariado evidenciaram a existência de três grupos de variáveis (Figura 11). Os dois primeiros grupos apresentaram, respectivamente, os maiores e intermediários valores estimados do  $ESS_t$ , sendo eles, respectivamente: o grupo que agrega as variáveis V1 e V2, cujo valor médio estimado do  $ESS_t$  foi de 40 e 44 pontos amostrais, nesta ordem; e o grupo das variáveis V3, V4, V5, V6 e V9, em que o valor médio estimado do  $ESS_t$  variou entre 15 a 31 pontos amostrais. Estes dois grupos de variáveis exibiram também elevados valores de desvios padrão, variando de 11 a 14 pontos amostrais.

As variáveis simuladas V1 e V2 possuem o menor raio de dependência espacial ( $a=0,3$  km), e este é um alcance prático baixo quando comparado à distância máxima entre as coordenadas da área simulada (aproximadamente 1,8 km). As variáveis V3, V4, V6 e V9 exibiram raios de

dependência espacial um pouco maiores do que as do primeiro grupo (variando de 0,5 a 0,6 km), o que contribuiu para que os valores estimados do  $ESS_t$  univariado fossem menores comparados aos do grupo anterior.

O terceiro grupo, formado pelas variáveis V7, V8, V10 e V11 (em que o valor médio estimado do  $ESS_t$  variou entre 6 e 8 pontos amostrais), apresentou os menores valores médios estimados do tamanho amostral efetivo univariado (Figura 11). Nesse grupo, notou-se que embora tenha apresentado média próxima às demais variáveis, V8 exibiu maior valor de desvio padrão. Estas quatro variáveis têm em comum os maiores valores do raio de dependência espacial (entre 1,0 e 1,2 km).

Desta forma, o alcance prático foi o parâmetro que exerceu influência negativa nos valores estimados do tamanho amostral efetivo univariado, isto é, quanto maior foi o alcance prático da variável, menor foi o valor estimado do  $ESS_t$ . Embora tenham considerado configuração e tamanho amostral distintos e até mesmo outra distribuição probabilística (normal), os estudos de simulação de Vallejos e Osorio (2014) chegaram a conclusões semelhantes em relação à influência deste parâmetro na redução do número de pontos amostrais. No geral, a metodologia de Vallejos e Osorio para estimação do valor do  $ESS_t$  univariado proporcionou redução entre 57% e 95% no número de pontos amostrais. Isso implica que o valor estimado do  $ESS_t$  variou entre 6 e 44 pontos amostrais. A elevada diferença na redução obtida pode ser explicada pela discrepância ocorrida entre as variáveis, em relação aos valores dos parâmetros de dependência espacial, principalmente quanto ao alcance prático, cuja variação foi de 0,3 a 1,2 km.

Segundo estudos desenvolvidos por Coelho *et al.* (2009), Souza *et al.* (2014) e Kestring *et al.* (2015), mapas temáticos construídos com reduções superiores a 66% no número de pontos amostrais tendem a apresentar índices de acurácia não satisfatórios quando comparados aos mapas confeccionados com o tamanho amostral original. Isto é, o mapa temático obtido utilizando um conjunto amostral com número reduzido de pontos apresenta qualidade inferior na representação da variabilidade espacial das variáveis na área agrícola.

Teoricamente, existe na literatura uma quantidade mínima de pontos amostrais que possibilitam ajustar um modelo à função semivariância, descrevendo assim a estrutura de dependência espacial e, posteriormente, é possível realizar a krigagem e construir o mapa temático. Entretanto, não existe um consenso entre os autores sobre qual seria esse número mínimo de pontos. Journel e Huijbregts (1978), Guerra (1988) e Wollenhaupt, Mulla e Crawford (1997) consideram que 30 pares de pontos amostrais seriam suficientes, enquanto Webster e Oliver (1990) sugerem um número mínimo de 100 pares de pontos.

Estudos são encontrados na literatura utilizando-se um tamanho amostral inferior a 50 pontos amostrais para se caracterizar a variabilidade espacial de atributos do solo (CARVALHO *et al.*, 2013; ARAÚJO *et al.*, 2014; TAVARES *et al.*, 2014). Essa quantidade de pontos amostrais é similar à obtida no presente trabalho para a maioria das variáveis simuladas. Contudo, identificou-se nesses trabalhos que, geralmente, a área agrícola utilizada é inferior a deste estudo.

#### 5.2.3.2 Aplicação da metodologia na resistência do solo à penetração

Nas quatro diferentes profundidades analisadas da resistência do solo à penetração (RSP) (0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm), verificou-se que o menor valor (0,75 MPa) e o maior

valor (6,67 MPa) foram apresentados na camada 0-10 cm de profundidade, a qual também exibiu o maior valor médio (3,12 MPa), enquanto o menor valor médio (2,03 MPa) foi verificado na camada 21-30 cm de profundidade (Tabela 11).

Os valores do coeficiente de variação (CV) indicaram que nas duas maiores profundidades (21-30 cm e 31-40 cm) da resistência do solo à penetração houve média dispersão dos dados ( $10\% \leq CV \leq 20\%$ ) (PIMENTEL GOMES, 1985). Foi identificada alta dispersão na camada 11-20 cm de profundidade ( $20\% < CV \leq 30\%$ ) (PIMENTEL GOMES, 1985); enquanto na menor profundidade (0-10 cm), a RSP apresentou heterogeneidade ( $CV > 30\%$ ) (PIMENTEL GOMES, 1985) (Tabela 11).

Tabela 11 – Estatística descritiva das quatro camadas de profundidade da resistência do solo à penetração (RSP).

Variável	Mín.	Média	Máx.	DP	CV (%)
RSP 0-10 cm (MPa)	0,75	3,12	6,67	0,97	31,22
RSP 11-20 cm (MPa)	1,78	3,01	5,53	0,62	20,70
RSP 21-30 cm (MPa)	1,50	2,03	3,86	0,33	16,27
RSP 31-40 cm (MPa)	1,48	2,08	3,95	0,37	18,01

RSP: resistência do solo à penetração, Mín.: mínimo, Máx.: máximo, DP: desvio padrão, CV: coeficiente de variação.

As diferentes camadas de RSP apresentaram pontos discrepantes, destacados espacialmente no gráfico post-plot (Figura 12).

As camadas 11-20 cm (Figura 12 - B) e 31-40 cm (Figura 12 - D) de profundidade exibiram a maior quantidade de valores atípicos (quatro cada), localizados nas regiões central e oeste da área agrícola. Os pontos amostrais 82 e 34 (localizados espacialmente na Figura 12 - A) exibiram valores discrepantes em todas as camadas da resistência do solo à penetração das raízes, com exceção da 31-40 cm de profundidade, na qual o ponto 34 não foi considerado atípico.

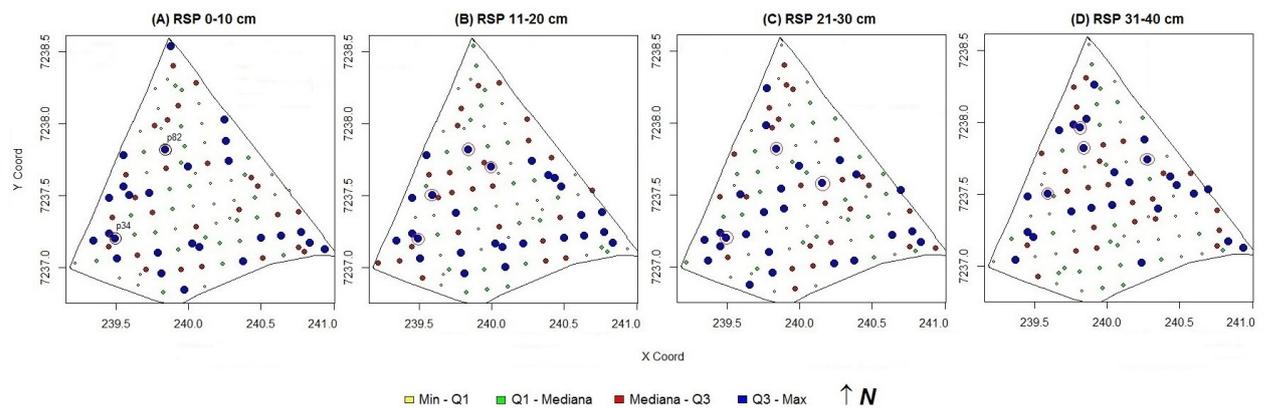


Figura 12 – Gráfico post-plot da resistência do solo à penetração (RSP) [MPa] nas quatro camadas de profundidade. As circunferências destacam espacialmente os pontos discrepantes. Min: mínimo, Q1: 1º quartil, Q3: 3º quartil, Max: máximo.

Para todas as camadas de profundidade da RSP, ajustou-se à função semivariância modelos geoestatísticos com fator de anisotropia. Identificou-se que nas camadas 0-10 cm e 31-40 cm de profundidade, os fatores de anisotropia estimados variaram entre 1 e 2. Mas, segundo estudos

desenvolvidos por Guedes, Uribe-Opazo e Ribeiro JR (2013), fatores de anisotropia geométrica estimados inferiores a 2 não são relevantes na estimativa dos parâmetros de dependência espacial. Deste modo, considerou-se nesse trabalho que a estrutura de dependência espacial é isotrópica para todas as camadas da RSP.

A Tabela 12 exibe um resumo dos melhores valores do grau de liberdade ( $\nu$ ) obtidos pelos critérios de validação cruzada e traço, e para cada um desses valores de  $\nu$ , o melhor valor para o parâmetro de forma  $\kappa$ , determinado pelo erro padrão, com os respectivos valores estimados dos parâmetros de dependência espacial. Os valores estimados do erro padrão dos parâmetros são destacados entre parênteses (Tabela 12). Para a resistência do solo à penetração nas camadas 0-10 cm e 11-20 cm de profundidade, verificou-se que, no geral, o menor erro padrão foi obtido pelo modelo da família Matérn com  $\kappa = 2,5$  e grau de liberdade  $\nu = 5$ . Nas duas maiores profundidades analisadas da RSP, constatou-se que, de maneira geral, os menores valores do erro padrão foram obtidos considerando o modelo da família Matérn com  $\kappa = 1,5$  e  $\nu = 5$  na camada 21-30 cm; e  $\kappa = 0,5$ , isto é, modelo exponencial, e  $\nu = 10$  na camada 31-40 cm.

Tabela 12 – Valores estimados dos parâmetros que definem a estrutura de dependência espacial em cada camada de profundidade da resistência do solo à penetração a partir dos melhores valores dos parâmetros de forma  $\kappa$  e  $\nu$ .

Variável	$\nu$	$\kappa$	$\hat{\beta}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\varphi}_1 + \hat{\varphi}_2$	$\hat{a} = g(\hat{\varphi}_3)$	$\widehat{EPR}$
RSP 0-10 cm	5	2,5	3,1188 (0,1362)	0,6246 (0,2338)	0,3072 (0,3667)	0,0491 (0,0107)	0,9318 -	291,50 -	67,03% -
RSP 11-20 cm	5	2,5	2,9991 (0,0570)	0,2842 (0,1086)	0,0209 (0,9388)	0,0305 (0,6620)	0,3051 -	180,50 -	93,14% -
RSP 21-30 cm	5	1,5	2,0179 (0,0297)	0,0819 (0,0425)	0,0056 (3,3146)	0,0231 (6,4972)	0,0875 -	110,20 -	93,60% -
RSP 31-40 cm	10	0,5	2,0659 (0,0484)	0,0634 (0,0165)	0,0755 (0,2477)	0,0741 (0,0483)	0,1389 -	222,50 -	45,67% -

$\nu$ : grau de liberdade,  $\kappa$ : parâmetro de forma do modelo da família Matérn, valores estimados de:  $\hat{\beta}$ : média,  $\hat{\varphi}_1$ : efeito pepita,  $\hat{\varphi}_2$ : contribuição,  $\hat{\varphi}_3$ : função do alcance,  $\hat{\varphi}_1 + \hat{\varphi}_2$ : patamar,  $\hat{a}$ : alcance prático (em metros),  $\widehat{EPR} = 100 \frac{\hat{\varphi}_1}{\hat{\varphi}_1 + \hat{\varphi}_2}$ : efeito pepita relativo (%). Os valores estimados do erro padrão de cada parâmetro são apresentados entre parênteses.

Devido aos baixos valores estimados do alcance prático (180,5 e 110,2 m) e a classificação de Cambardella *et al.* (1994) indicar fraca dependência espacial (EPR superior a 75%), tem-se que a estrutura de dependência espacial nas camadas intermediárias da RPS (11-20 cm e 21-30 cm de profundidade) apresenta efeito pepita puro (Tabela 12).

Considerando-se a distância máxima entre as localizações amostradas na área agrícola sob estudo (cerca de 1,8 km), pode-se admitir que os valores estimados do raio de dependência espacial apresentados pelas diferentes profundidades da RSP foram baixos (entre 110,2 e 291,5 m), especialmente nas camadas 11-20 cm e 21-30 cm, nas quais os valores estimados do alcance prático foram inferiores a 200 metros (Tabela 12).

Verificou-se no mapa temático da resistência do solo à penetração na camada 21-30 cm de profundidade (Figura 13 - C), a formação de algumas microrregiões circulares centradas nos pontos amostrais combinada à predominância de pixels em uma única classe, e estes fatores caracterizam o fenômeno *bull eyes effect* (MENEZES *et al.*, 2016). O fenômeno pode ter sido causado pelo baixo valor estimado do raio de dependência espacial apresentado nessa camada da RSP ( $\hat{a} = 110,2$  m).

Segundo a classificação desenvolvida por Canarache (1991), a qual estabelece uma relação

entre o intervalo de variação da resistência do solo à penetração (em MPa) e o grau de limitação ao crescimento das raízes, a variabilidade espacial na camada 21-30 cm de profundidade da RSP indica “pouca limitação ao crescimento das raízes”, pois os valores da RSP variaram entre 1,98 e 2,14 MPa. As camadas 0-10 cm (Figura 13 - A) e 11-20 cm (Figura 13 - B) de profundidade da RSP apresentaram condições mais restritivas e oferecem “algumas limitações” quanto ao crescimento das raízes na área agrícola (RSP variando entre 2,6 e 4,11 Mpa) (CANARACHE, 1991). A única exceção é a primeira classe no mapa da RSP na camada 0-10 cm de profundidade (Figura 13 - A), a qual representa cerca de 2% da área, onde existe pouca limitação ao crescimento das raízes (RSP inferior a 2,6 Mpa) (CANARACHE, 1991). Por fim, na camada 31-40 cm de profundidade da RSP (Figura 13 - D), observou-se que em mais de 95% da área (agregando a primeira, segunda e terceira classes) tem-se “pouca limitação” ao crescimento das raízes, de acordo com a classificação sugerida por Canarache (1991).

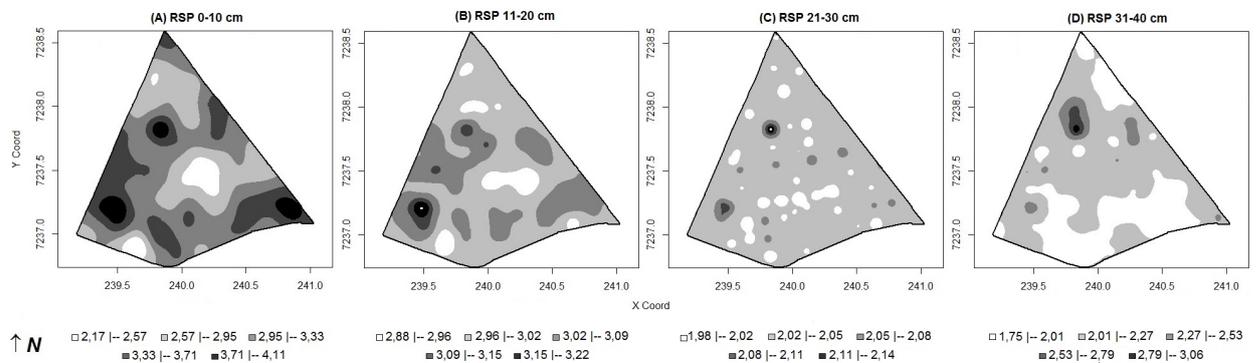


Figura 13 – Mapa dos valores estimados da resistência do solo à penetração das raízes (RSP) [MPa] nas quatro camadas de profundidade.

De acordo com os dados dos 102 pontos amostrais coletados bem como com as estimativas dos parâmetros de dependência espacial obtidas via algoritmo EM, foram estimados os valores do tamanho amostral efetivo univariado nas camadas 0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm de profundidade da resistência do solo à penetração (Tabela 13).

Tabela 13 – Valor estimado do tamanho amostral efetivo univariado em cada camada de profundidade da resistência do solo à penetração ( $\widehat{ESS}_t$ ).

Variável ( $\kappa; \nu$ )	$\widehat{ESS}_t$
RSP 0-10 cm ( $\kappa=2,5; \nu=5$ )	51
RSP 11-20 cm ( $\kappa=2,5; \nu=5$ )	95
RSP 21-30 cm ( $\kappa=1,5; \nu=5$ )	101
RSP 31-40 cm ( $\kappa=0,5; \nu=10$ )	60

$\kappa$ : parâmetro de forma do modelo da família Matérn;  $\nu$ : grau de liberdade.

Verificou-se que, nas camadas 11-20 cm e 21-30 cm de profundidade da resistência do solo à penetração, praticamente não houve redução no tamanho amostral (Tabela 13). Tal fato é justificável, principalmente devido à influência do alcance sobre o valor estimado do  $\widehat{ESS}_t$ , verificada tanto nos estudos de simulação do presente trabalho (Seção 5.2.3.1) quanto nos de Vallejos e Osorio

(2014). Haja vista as referidas camadas de RSP apresentarem um pequeno valor estimado do raio de dependência espacial (110,2 e 180,5 m), vinculado a uma fraca intensidade de dependência espacial observada pelos valores estimados do EPR superiores a 93% (Tabela 12).

Devido aos maiores valores estimados do alcance (291,5 e 222,5 m), as camadas 0-10 cm e 31-40 cm de profundidade da RSP tiveram reduções mais elevadas no número de pontos amostrais, variando entre 40% e 50% (Tabela 13). Embora utilizando atributos químicos do solo, além de diferentes circunstâncias quanto à configuração amostral e distribuição de probabilidade dos atributos, Griffith (2005) verificou redução no tamanho amostral similar em seus estudos de dados reais, a qual variou de 36% a 45%.

Sendo assim, tomando-se como referência as camadas de RSP em que se identificou dependência espacial (0-10 e 31-40 cm de profundidade) e o máximo valor estimado do tamanho amostral efetivo observado nestas (60 pontos), selecionou-se uma nova configuração com 60 pontos amostrais, escolhidos aleatoriamente, para o estudo da variabilidade espacial da resistência do solo à penetração. Foram encontrados na literatura trabalhos com o intuito de analisar a variabilidade espacial em diferentes profundidades da RSP, os quais foram desenvolvidos utilizando entre 49 e 60 pontos amostrais, tamanho amostral adjacente ao estudado nesta pesquisa, embora a magnitude da área agrícola mapeada tenha sido menor do que a deste trabalho (ROSALEN *et al.*, 2011; RODRIGUES *et al.*, 2014; TAVARES *et al.*, 2014). Além disso, deve-se ponderar que utilizar um número reduzido de pontos amostrais também está vinculado à variabilidade espacial de cada atributo do solo verificada em estudos anteriores realizados na área agrícola, pois variáveis cuja variabilidade espacial é sabidamente menor, tendem a exibir bons resultados em termos da descrição da variabilidade espacial com tamanho amostral reduzido (RODRIGUES, CORÁ e FERNANDES, 2012).

Dadas as observações das quatro diferentes camadas da resistência do solo à penetração (RSP) (0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm de profundidade) obtidas com a configuração amostral reduzida, verificou-se que o menor valor (1,49 MPa) e o maior valor (6,67 MPa) foram apresentados na camada 0-10 cm de profundidade, a qual também exibiu o maior valor médio (3,18 MPa), enquanto o menor valor médio (2,00 MPa) foi verificado na camada 31-40 cm de profundidade (Tabela 14).

Os valores do coeficiente de variação (CV) indicaram que nas duas maiores profundidades (21-30 cm e 31-40 cm) da resistência do solo à penetração houve média dispersão dos dados (PIMENTEL GOMES, 1985). Já nas camadas 0-10 cm e 11-20 cm de profundidade da RSP, foram identificadas heterogeneidade e alta dispersão, respectivamente (PIMENTEL GOMES, 1985) (Tabela 14).

Tabela 14 – Estatística descritiva das quatro camadas de profundidade da resistência dos solo à penetração (RSP), de acordo com a malha amostral reduzida.

Variável	Mín.	Média	Máx.	DP	CV (%)
RSP 0-10 cm (MPa)	1,49	3,18	6,67	1,04	32,69
RSP 11-20 cm (MPa)	1,78	3,05	5,53	0,67	22,12
RSP 21-30 cm (MPa)	1,50	2,04	3,86	0,37	18,20
RSP 31-40 cm (MPa)	1,56	2,00	3,95	0,42	19,83

RSP: resistência do solo à penetração, Mín.: mínimo, Máx.: máximo, DP: desvio padrão, CV: coeficiente de variação.

Todas as camadas de RSP exibiram pontos discrepantes (Figura 14), com destaque para a

camada 21-30 cm de profundidade (Figura 14 - B), que apresentou a maior quantidade de valores atípicos (quatro), localizados nas regiões central e oeste da área agrícola.

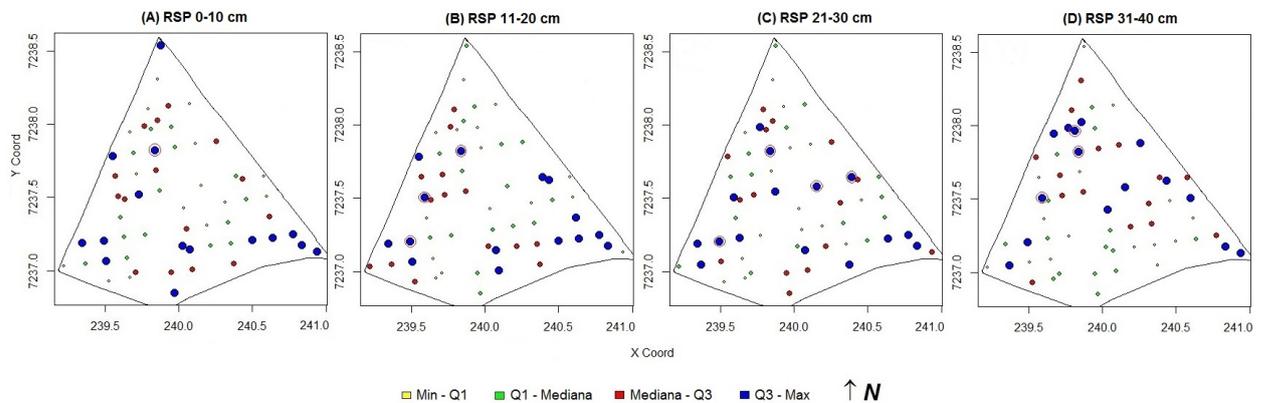


Figura 14 – Gráfico post-plot da resistência do solo à penetração (RSP) [MPa] nas quatro camadas de profundidade, considerando a configuração amostral reduzida. As circunferências destacam espacialmente os pontos discrepantes. Min: mínimo, Q1: 1º quartil, Q3: 3º quartil, Max: máximo.

A Tabela 15 exibe um resumo dos melhores graus de liberdade ( $\nu$ ) obtidos pelos critérios de validação cruzada e traço, e para cada um desses  $\nu$ , o melhor parâmetro de forma  $\kappa$ , determinado pelo erro padrão, com os respectivos valores estimados dos parâmetros de dependência espacial. Os erros padrão das estimativas destes parâmetros estão destacados entre parênteses. Para a resistência do solo à penetração nas camadas intermediárias, 11-20 cm e 21-30 cm de profundidade, verificou-se que, no geral, o menor erro padrão foi obtido pelo modelo da família Matérn com  $\kappa = 2,5$  e graus de liberdade respectivamente iguais a  $\nu = 10$  e  $\nu = 5$ . Nas camadas 0-10 cm e 31-40 cm de profundidade da RSP, constatou-se que, de maneira geral, os menores valores do erro padrão foram obtidos a partir do modelo da família Matérn com  $\kappa = 0,5$  e  $\nu = 10$ .

Tabela 15 – Valores estimados dos parâmetros que definem a estrutura de dependência espacial em cada camada de profundidade da resistência do solo à penetração, com a malha amostral reduzida, de acordo com os melhores valores dos parâmetros de forma  $\kappa$  e  $\nu$ .

Variável	$\nu$	$\kappa$	$\hat{\beta}$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	$\hat{\varphi}_3$	$\hat{\varphi}_1 + \hat{\varphi}_2$	$\hat{a} = g(\hat{\varphi}_3)$	$\widehat{\text{EPR}}$
RSP 0-10 cm	10	0,5	3,1751 (0,2412)	0,3446 (0,1275)	1,4262 (0,2446)	0,0910 (0,0027)	1,7709 -	273,20 -	19,46% -
RSP 11-20 cm	10	2,5	3,0446 (0,1090)	0,2356 (0,0672)	0,2558 (0,2468)	0,0353 (0,0051)	0,4914 -	209,10 -	47,94% -
RSP 21-30 cm	5	2,5	2,0382 (0,0430)	0,0706 (0,3243)	0,0356 (4,5560)	0,0132 (0,6775)	0,1063 -	78,30 -	66,50% -
RSP 31-40 cm	10	0,5	2,0973 (0,0807)	0,1096 (0,0338)	0,1596 (0,2609)	0,0716 (0,0313)	0,2692 -	214,90 -	40,71% -

$\nu$ : grau de liberdade,  $\kappa$ : parâmetro de forma do modelo da família Matérn, valores estimados de:  $\hat{\beta}$ : média,  $\hat{\varphi}_1$ : efeito pepita,  $\hat{\varphi}_2$ : contribuição,  $\hat{\varphi}_3$ : função do alcance,  $\hat{\varphi}_1 + \hat{\varphi}_2$ : patamar,  $\hat{a}$ : alcance prático (em metros),  $\widehat{\text{EPR}} = 100 \frac{\hat{\varphi}_1}{\hat{\varphi}_1 + \hat{\varphi}_2}$ : efeito pepita relativo (%). Os valores estimados do erro padrão de cada parâmetro são apresentados entre parênteses.

A análise do erro padrão dos parâmetros que definem a estrutura de dependência espacial (Tabela 15) mostrou que os valores estimados não foram tão maiores dos obtidos quando comparados à quantidade original de pontos amostrais, inclusive, em alguns casos, o erro padrão foi menor

utilizando a malha amostral reduzida (Tabela 12).

De acordo com a classificação de Cambardella *et al.* (1994), a camada 0-10 cm de profundidade da RSP foi a única que apresentou forte dependência espacial, pois o valor estimado do EPR foi inferior a 25% (Tabela 15). Já nas demais profundidades, a intensidade da dependência espacial foi moderada (CAMBARDELLA *et al.*, 1994), pois o valor estimado do EPR variou entre 25 e 75% (Tabela 15).

No geral, embora os parâmetros que definem a estrutura de dependência espacial tenham sido estimados com diferentes parâmetros  $\kappa$  no modelo da família Matérn, pode-se observar que os valores estimados do alcance prático foram relativamente baixos. Haja vista a distância máxima na área ser aproximadamente 1800 m, e os alcances variarem de 110,2 a 291,5 m na malha original, e de 78,3 a 273,2 m na grade amostral reduzida.

A Figura 15 apresenta os mapas temáticos gerados com a configuração amostral original e reduzida. Algumas classes em mapas com a configuração amostral original apresentaram quantidade de pixels nula, a partir da construção dos mapas temáticos com os mesmos intervalos de classe, como pode ser constatado visualmente nas camadas 0-10 cm, 11-20 cm e 21-30 cm de profundidade da RSP (Figura 15 - A, B e C, respectivamente). Além disso, o baixo valor estimado do alcance prático da camada 21-30 cm de profundidade da RSP ( $\hat{\alpha} = 78,30$  m) acarretou na formação de pequenas sub-regiões centradas nos pontos amostrais (Figura 15 - C). Tal fenômeno é conhecido como *bull eyes effect* (MENEZES *et al.*, 2016).

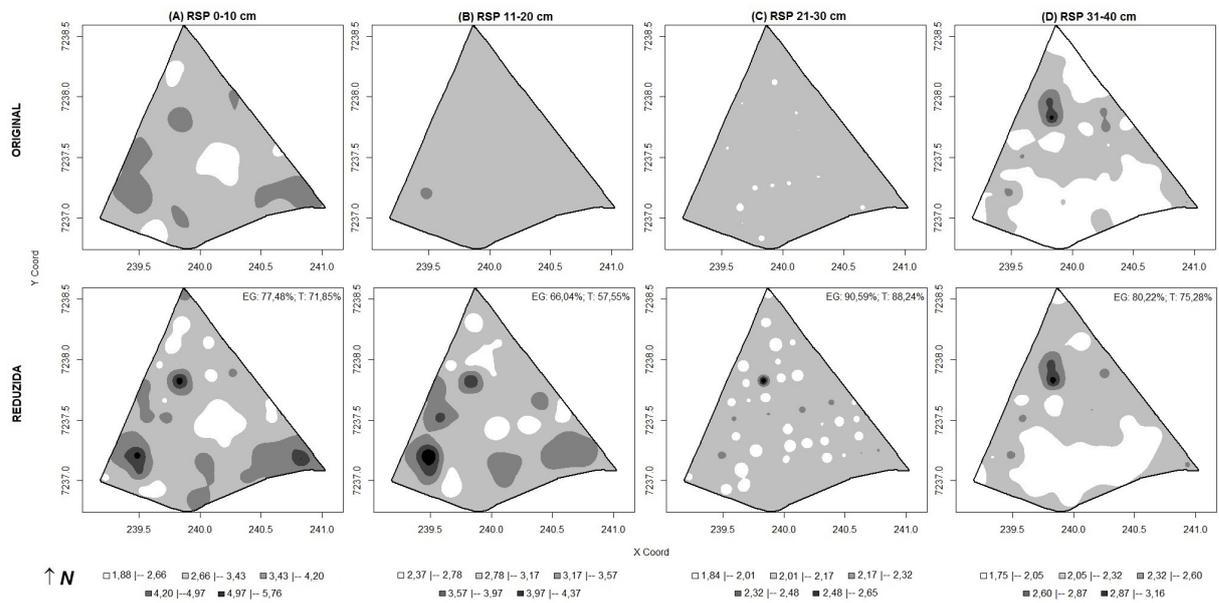


Figura 15 – Mapa dos valores estimados da resistência do solo à penetração das raízes (RSP) [MPa] nas quatro camadas de profundidade, considerando a malha amostral original e reduzida com os mesmos intervalos de classe. Valores estimados de EG (Exatidão Global) e T (índice Tau).

Visualmente, é possível notar que existem diferenças entre os mapas com a configuração amostral original e reduzida, as quais são mais notórias em algumas camadas da RSP, tal como na profundidade 11-20 cm (Figura 15).

De acordo com a classificação de Anderson *et al.* (2001), na maioria das camadas, existe

um baixo percentual de acertos entre os mapas referência e modelo, pois o valor estimado do índice de Exatidão Global (EG) foi inferior a 85%. Isso indica que um menor número de pixels foi classificado no mesmo intervalo de classe nos mapas referência e modelo, o que aponta diferenças entre os mapas com a malha amostral original e reduzida. A única exceção foi a camada 21-30 cm de profundidade da RSP (Figura 15 - C), na qual se verificou similaridade entre os mapas obtidos com a configuração amostral original e reduzida (EG superior a 85%).

O índice Tau (T), diferentemente da EG, contabiliza não apenas a proporção de pixels classificados no mesmo intervalo nos mapas referência e modelo, mas também aqueles cuja classificação não foi a mesma em ambos os mapas. Os mapas com grade original e reduzida das demais camadas apresentaram de média a alta acurácia ( $T \geq 0,67$ ; KRIPPENDORFF *et al.*, 2004), com exceção da camada 11-20 cm de profundidade da RSP.

Deve-se destacar que, na camada 21-30 cm de profundidade da RSP, os elevados valores dos índices de acurácia EG e Tau, referentes aos mapas com a configuração amostral original e reduzida, são devido à elevada quantidade de pixels em uma mesma classe em ambos os mapas. No mapa referência, 99,20% dos pixels pertencem ao intervalo  $[2,01; 2,17)$ , enquanto no mapa modelo essa classe agregou 90,38% dos pixels. O fato pode ser visualmente constatado na Figura 15 - C, pela homogeneidade dos mapas temáticos em uma mesma classe, o que levou a elevados índices de acurácia nessa camada.

Segundo a classificação de Canarache (1991), identificou-se que, na camada 0-10 cm de profundidade da RSP (Figura 15 - A), existem “pouca” ou “algumas limitações” no que tange ao crescimento das raízes em 99% da área agrícola. Há “sérias limitações” em apenas 1% do campo. Nas demais camadas da RSP, a área agrícola estudada apresentou “pouca” ou “algumas limitações” para o crescimento das raízes (Figura 15 - B, C e D) (CANARACHE, 1991).

#### 5.2.4 Conclusões

Nos estudos acerca do tamanho amostral efetivo ( $ESS_t$ ) realizados com as camadas 0-10 cm, 11-20 cm, 21-30 cm e 31-40 cm de profundidade da resistência do solo à penetração (RSP), nos quais se supôs que as variáveis georreferenciadas apresentam distribuição de probabilidade t-Student, foi constatado uma redução no tamanho amostral. Considerando o maior valor estimado do  $ESS_t$  nas camadas da RSP com dependência espacial, obteve-se uma grade amostral reduzida com 60 pontos amostrais.

Foi verificado que, no geral, com a malha reduzida, os valores estimados do erro padrão dos parâmetros que definem a estrutura de dependência espacial não foram tão distintos em comparação aos obtidos de acordo com a malha amostral original. Além disso, constatou-se que existem diferenças entre os mapas temáticos gerados com a configuração amostral original e reduzida, indicando que houve influência do tamanho amostral na caracterização da variabilidade espacial da RSP. Entretanto, mesmo com essa dessemelhança, o índice Tau apontou que, na maioria das camadas da RSP, com exceção da 11-20 cm de profundidade, os mapas confeccionados com menor número de pontos amostrais apresentaram de média a alta acurácia.

Desta forma, a metodologia do  $ESS_t$  univariado mostrou-se vantajosa quando foram ponderados os resultados obtidos na caracterização da variabilidade espacial da área agrícola e o menor custo no processo de amostragem vinculado à redução no tamanho amostral. Além

disso, o método proposto nesse trabalho utiliza um único tamanho amostral para todos os atributos, sem considerar técnicas multivariadas, mas apenas baseando-se nas variáveis com estrutura de dependência espacial e tomando-se dentre elas o máximo valor estimado do  $ESS_t$ .

## Agradecimentos

Os autores agradecem à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro, e aos membros do Laboratório de Estatística Espacial (LEE) e do Laboratório de Estatística Aplicada (LEA) pelo banco de dados fornecido.

## 5.2.5 Referências

ANDERSON, J. F.; HARDY, E. E.; ROACH J. T.; WITMER R. E. A land use and land cover classification system for use with remote sensor data, U.S. **Geological Survey Professional Paper 964**, U.S. Geologic Survey, Washington - DC, 2001, 41p.

APARECIDO, L. E. O.; ROLIM, G. S.; RICHETTI, J.; SOUZA, P. S.; JOHANN, J. A. Köppen, Thornthwaite and Camargo climate classifications for climatic zoning in the State of Paraná, Brazil. **Ciência e Agrotecnologia**, v. 40, n. 4, p. 405-417, 2016.

ARAÚJO, D. R.; MION, R. L.; SOMBRA, W. A.; ANDRADE, R. R.; AMORIM, M. Q. Variabilidade espacial de atributos físicos em solo submetido à diferentes tipos de uso e manejo. **Revista Caatinga**, v. 27, n. 2, p. 101-115, 2014.

ASSUMPÇÃO, R. A. B.; URIBE-OPAZO, M. A.; GALEA, M. Analysis of local influence in geostatistics using Student's t-distribution. **Journal of Applied Statistics**, v. 41, n. 11, p. 2323-2341, 2014.

BERNARDI, A. C. C.; NAIME, J. M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. Aplicações da agricultura de precisão em sistemas de produção de grãos no Brasil. In: RESENDE, A. V.; HURTADO, S. M. C.; VILELA, M. F.; CORAZZA, E. J.; SHIRATSUCHI, L. S. **Agricultura de Precisão: resultados de um novo olhar**. São Carlos: EMBRAPA, 2014. cap. 14, p. 194-208.

CAMBARDELLA, C. A.; MOORMAN, T.; PARKIN, T.; KARLEN, D.; NOVAK, J.; TURCO, R.; KONOPKA, A. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society of America Journal**, v. 58, n. 5, p. 1501-1511, 1994.

CANARACHE, A. Factors and indices regarding excessive compactness of agricultural soils. **Soil and Tillage Research**, v. 19, n. 2, p. 145-164, 1991.

CARVALHO, L. C. C.; SILVA, F. M.; ARAÚJO, G.; FERRAZ, S.; SILVA, F. C.; STRACIERI, J. Variabilidade espacial de atributos físicos do solo e características agronômicas da cultura do café. **Coffee Science**, v. 8, n. 3, p. 265-275, 2013.

COELHO, E. C.; SOUZA, E. G.; URIBE-OPAZO, M. A.; NETO, R. P. Influência da densidade amostral e do tipo de interpolador na elaboração de mapas temáticos. **Acta Scientiarum. Agronomy**, v. 31, n. 1, p. 165-174, 2009.

CORDY, C. B.; THOMPSON, C. M. An application of the deterministic variogram to design-based variance estimation. **Mathematical Geology**, v. 27, n. 2, p. 173-205, 1995.

DALPOSSO, G. H., URIBE-OPAZO, M. A., MERCANTE, E., JOHANN, J. A., BORSSOI, J. A. Comparison measures of maps generated by geostatistical methods. **Engenharia Agrícola**, v. 32, n. 1, p. 174-183, 2012.

DE BASTIANI, F.; CYSNEIROS, A. F. J.; CYSNEIROS, A. H. M.; URIBE-OPAZO, M. A.; GALEA, M. Influence diagnostics in elliptical spatial linear models. **Test**, v. 24, n. 2, p. 322-340, 2015.

DIGGLE, P.; LOPHAVEN, S. Bayesian geostatistical design. **Scandinavian Journal of Statistics**, v. 33, n. 1, p. 53-64, 2006.

DIGGLE, P. J.; RIBEIRO JR, P. J. **Model-based Geostatistics**. New York: Springer Series in Statistics, 2007. 230p.

DOBBIE, M. J.; HENDERSON, B. L.; STEVENS, D. L. Sparse sampling: spatial design for monitoring stream networks. **Statistics Surveys**, v. 2, p. 113-153, 2008.

EMBRAPA. **Sistema Brasileiro de Classificação de Solos**. 3. ed. Brasília: Centro Nacional de Pesquisa de Solos, 2013. 306p.

FLEISCHER, S. J.; BLOM, P. E.; WEISZ, R. Sampling in precision IPM: when the objective is a map. **Phytopathology**, v. 89, n. 11, p. 1112-1118, 1999.

GRIFFITH, D. A. Effective geographic sample size in the presence of spatial autocorrelation. **Annals of the Association of American Geographers**, v. 95, n. 4, p. 740-760, 2005.

GRINAND, C.; ARROUAYS D.; LAROCHE B.; MARTIN M. P. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. **GEODERMA**, v. 143, n. 1-2, p. 180-190, 2008.

GUEDES, L. P. C.; RIBEIRO JR, P. J.; PIEDADE, S. M. S.; URIBE-OPAZO, M. A. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. **Chilean Journal of Statistics (ChJS)**, v. 2, n. 2, 2011.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; RIBEIRO JR, P. J. Influence of incorporating geometric anisotropy on the construction of thematic maps of simulated data and chemical attributes of soil. **Chilean Journal of Agricultural Research**, v. 73, n. 4, p. 414-423, 2013.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; RIBEIRO JR, P. J. Optimization of sample design sizes and shapes for regionalized variables using simulated annealing. **Ciencia e Investigación Agraria**, v. 41, n. 1, p. 33-48, 2014.

GUEDES, L. P. C.; RIBEIRO JR, P. J.; URIBE-OPAZO, M. A.; DE BASTIANI, F. Soybean yield maps using regular and optimized sample with different configurations by simulated annealing. **Engenharia Agrícola**, v. 36, n. 1, p. 114-125, 2016.

GUERRA, P. A. G. **Geoestatística Operacional**. Brasília: Ministério de Minas e Energia, Departamento de Produção Mineral, 1988. 145p.

JOHANN, J. A.; URIBE-OPAZO, M. A.; SOUZA, E. G.; ROCHA, J. V. Spatial variability of the physical attributes and the productivity in a dystrophic Latosolo of Cascavel-PR, region. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 8, n. 2-3, p. 212-219, 2004.

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining Geostatistics**. London: Academic Press, 1978. 600p.

KESTRING, F.; GUEDES, L. P. C.; DE BASTIANI, F.; URIBE-OPAZO, M. A. Thematic maps comparison of different sampling grids for soybean productivity. **Engenharia Agrícola**, v. 35, n. 4, p. 733-743, 2015.

KRIPPENDORFF, K. **Content Analysis: an Introduction to its Methodology**. Beverly Hills: Sage Publications, 2004. 412p.

MA, Z. ; REDMOND, R. L. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering and Remote Sensing**, v. 61, n. 4, p. 453-439. 1995.

MENEZES, M. D.; SILVA, S. H. G.; MELLO, C. R.; OWENS, P. R.; CURI, N. Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. **Scientia Agrícola**, v. 73, n. 3, p. 274-285, 2016.

MOLIN, J. P.; AMARAL, L. R.; COLAÇO, A. F. **Agricultura de Precisão**. São Paulo: Oficina de Textos, 2015. 224p.

PATERNIANI, E. Agricultura sustentável nos trópicos. **Estudos Avançados**, v. 15, n. 43, p. 303-326, 2001.

PIMENTEL GOMES, F. **Curso de Estatística Experimental**. São Paulo: Nobel, 1985. 467p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R foundation for Statistical Computing, Vienna, Austria, 2017.

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

RODRIGUES, M. S.; CORÁ, J. E.; FERNANDES, C. Soil sampling intensity and spatial distribution pattern of soils attributes and corn yield in no-tillage system. **Engenharia Agrícola**, v. 32, n. 5, p. 852-865, 2012.

RODRIGUES, M. S.; RAMOS, R. R. D.; AZEVEDO, T. P.; PATROCÍNIO FILHO, A. P.; OLIVEIRA, L. G. Variabilidade espacial da resistência do solo à penetração em área capineira irrigada no semiárido. **Agropecuária Científica no Semiárido**, v. 10, n. 1, p. 161-166, 2014.

ROSALEN, D. L.; RODRIGUES, M. S.; CHIODEROLI, C. A.; BRANDÃO, F. J. C.; SIQUEIRA, D. S. GPS receivers for georeferencing of spatial variability of soil attributes. **Engenharia Agrícola**, v. 31, n. 6, p. 1162-1169, 2011.

RUIZCÁRDENAS, R.; FERREIRA, M. A. R.; SCHMIDT, A. M. Stochastic search algorithms for optimal design of monitoring networks. **Environmetrics**, v. 21, n. 1, p. 102-112, 2010.

SCHEMMER, R. C.; URIBE-OPAZO, M. A.; GALEA, M.; ASSUMPÇÃO, R. A. B. Spatial variability of soybean yield through a reparametrized t-Student model. **Engenharia Agrícola**, v. 37, n. 4, p. 760-770, 2017.

SOARES, A. **Geoestatística para Ciências da Terra e do Ambiente**. 3. ed. Lisboa: Press, 2014. 214p.

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

TAVARES, U. E.; MONTENEGRO, A. A. A.; ROLIM, M. M.; SILVA, J. S.; VICENTE, T. F. S.; ANDRADE, C. W. L. Variabilidade espacial da resistência à penetração e da umidade do solo em Neossolo Flúvico. **Water Resources and Irrigation Management**, v. 3, n. 2, p. 79-89, 2014.

URIBE-OPAZO, M. A.; BORSSOI, J. A.; GALEA, M. Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics**, v. 39, n. 3, p. 615-630, 2012.

VALLEJOS, R.; OSORIO, F. Effective sample size of spatial process models. **Spatial Statistics**, v. 9, p. 66-92, 2014.

WARRICK, A. W.; NIELSEN, D. R. Spatial variability of physical properties in the field. In: D. HILLEL, (Ed.). **Applications of Soil Physics**. New York: Academic Press, 1980. p.319-344.

WEBSTER, R.; OLIVER, M. A. Kriging: a method of interpolation for geographical information systems.

**International Journal of Geographical Information System**, v. 4, n. 3, p. 313-332, 1990.

WOLLENHAUPT, N. C.; MULLA, D. J.; CRAWFORD, C. A. G. Soil sampling and interpolation techniques for mapping spatial variability of soil properties. In: PIERCE, J. P.; SADLER, E. J. (Eds.). **The site-specific management for agricultural system**. Madison: ASA-CSSA-SSSA, 1997. p. 19-53.

## 6 Considerações Finais

Verificou-se com os resultados obtidos no presente estudo que, para variáveis com distribuição normal de probabilidade, os métodos utilizados para estimar o valor do tamanho amostral efetivo (ESS) uni e multivariado propiciaram redução no número de pontos amostrais. Já para atributos com distribuição t-Student, o método de Vallejos e Osorio (2014) proporcionou redução no tamanho amostral de cada variável.

Logo, tem-se que, na prática, é possível caracterizar a variabilidade espacial de uma área agrícola a partir das variáveis físico-químicas do solo com diferentes intensidades e raios de dependência espacial, e que seguem uma distribuição normal de probabilidade, com um único tamanho amostral reduzido obtido pelo cálculo do valor estimado do ESS multivariado. Já para variáveis georreferenciadas que apresentam distribuição de probabilidade t-Student, embora a metodologia utilizada possibilite apenas estimar o ESS univariado, considerou-se no presente trabalho o máximo valor estimado do ESS dentre as variáveis que exibiram dependência espacial. Tal fato possibilitou redimensionar a grade amostral e reduzir o número de pontos empregando um tamanho amostral único para todas as variáveis.

Por menor que seja a redução no tamanho amostral original obtida com o ESS, considerando os custos elevados de coleta e análise laboratorial, e dado que, teoricamente, a informação quanto à variabilidade espacial fornecida não seria afetada pelo tamanho amostral reduzido, pode-se reconhecer que redimensionar a grade amostral com um número reduzido de pontos amostrais seria economicamente interessante para o produtor rural.

Em trabalhos futuros, objetiva-se comparar as análises exploratórias e geoestatísticas realizadas com tamanho amostral original e reduzido, a partir de variáveis georreferenciadas com distribuição normal de probabilidade. Os principais intuits são comparar os mapas temáticos obtidos e verificar o nível de similaridade entre eles, de modo que se possa avaliar a influência da redução no número de pontos amostrais na qualidade da caracterização da variabilidade espacial das variáveis na área agrícola. Além disso, tem-se como objetivo desenvolver o método do tamanho amostral efetivo multivariado para variáveis com distribuição de probabilidade t-Student.