

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

**CORRELAÇÃO ESPACIAL BIVARIADA PARA O REDIMENSIONAMENTO DO TAMANHO  
AMOSTRAL EFETIVO**

LETÍCIA ELLEN DAL' CANTON

CASCAVEL-PARANÁ  
2022

LETÍCIA ELLEN DAL' CANTON

**CORRELAÇÃO ESPACIAL BIVARIADA PARA O REDIMENSIONAMENTO  
DO TAMANHO AMOSTRAL EFETIVO**

Tese apresentada ao Programa de Pós-graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Doutora em Engenharia Agrícola, área de concentração em Sistemas Biológicos e Agroindustriais e linha de pesquisa em Geoprocessamento, Estatística Espacial e Agricultura de Precisão.

Orientador: Dra. Luciana Pagliosa Carvalho Guedes  
Coorientador: Dr. Miguel Angel Uribe Opazo

CASCADEL-PARANÁ  
2022

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste<sup>1</sup>.

Dal' Canton, Leticia Ellen

Correlação espacial bivariada para o redimensionamento do tamanho amostral efetivo / Leticia Ellen Dal' Canton; orientadora Luciana Pagliosa Carvalho Guedes; coorientador Miguel Angel Uribe-Opazo. -- Cascavel, 2022.

86 p.

Tese (Doutorado Campus de Cascavel) -- Universidade Estadual do Oeste do Paraná, Centro de Ciências Exatas e Tecnológicas, Programa de Pós-Graduação em Engenharia Agrícola, 2022.

1. Geoestatística. 2. Processos Gaussianos. 3. Redução amostral. 4. Simulação. I. Guedes, Luciana Pagliosa Carvalho, orient. II. Uribe-Opazo, Miguel Angel, coorient. III. Título.

---

<sup>1</sup> Revisão de Língua Portuguesa, Língua Inglesa e Normas: Prof. Dr. José Carlos da Costa.

# LETÍCIA ELLEN DAL' CANTON

## CORRELAÇÃO ESPACIAL BIVARIADA PARA O REDIMENSIONAMENTO DO TAMANHO AMOSTRAL EFETIVO

Tese apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Doutora 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)



Miguel Angel Uribe Opazo

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



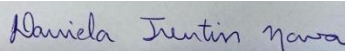
Fernanda De Bastiani

Universidade Federal de Pernambuco (UFPE)



Elizabeth Giron Cima

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



Daniela Trentin Nava

Universidade Tecnológica Federal do Paraná (UTFPR)



Márcio Paulo de Oliveira

Universidade Tecnológica Federal do Paraná (UTFPR)

Cascavel, 29 de agosto de 2022.

## 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. Mestre em Engenharia Agrícola pela UNIOESTE em 2018, tendo pesquisado sobre redimensionamento amostral utilizando a proposta do Tamanho Amostral Efetivo, considerando as distribuições de probabilidade normal e *t-student*. Em março de 2018, ingressou por essa mesma instituição de ensino no curso de doutorado 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 nos momentos mais difíceis.

Aos meus pais, que apesar de não terem tido a oportunidade de estudar, sempre priorizaram minha vida acadêmica, sem poupar esforços. Me orgulho de vocês todos os dias e os amo muito!

À Luana, por ter ouvido meus desabafos e entendido cada momento de estresse, ansiedade e nervosismo. Obrigada pelo seu companheirismo, pelos abraços em momentos oportunos e pelo seu amor diário. Te amo, vida!

Encerrando a sessão família, agradeço aos meus *pets* (sim!), Chulepa, Neguinha e Nina. Embora não tenham conhecimento disso, o amor gratuito e espontâneo deles me ajudou em cada dia difícil. Durante a execução deste trabalho, a Neguinha e a Nina nos deixaram e não puderam me ver ser doutora. Mas saibam que o amor de vocês me trouxe até aqui. As amarei para todo sempre!

Aos meus amigos, que sempre estiveram comigo, cada um ajudando da sua forma. Obrigada, de todo coração, essa tese tem um pouco de cada um de vocês. Agradeço, em especial, à Naisa e a Sabrina e, também, à Tamara e ao Weverton.

Aos amigos que a vivência acadêmica me trouxe: Darlon Vasata, Elizabeth Cima, Marília Favalesso, Professora Ana Tereza Bittencourt, Rosangela Schemmer e, é claro, à melhor consultora de solos de todo o Brasil, Natália Pereira.

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

Ao meu coorientador, Dr. Miguel Angel Uribe Opazo, pela disposição em contribuir (e muito) com o trabalho e por toda preocupação e incentivo dados.

Aos professores que compuseram a banca pelas contribuições dadas. Além dos mais sinceros agradecimentos à Unioeste e ao Programa de Pós-Graduação em Engenharia Agrícola - PGEAGRI.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pelo suporte financeiro.

Ao Laboratório de Estatística Espacial (LEE) e ao Laboratório de Estatística Aplicada (LEA), pelos dados agrícolas fornecidos.

DAL' CANTON, Letícia Ellen. **Correlação espacial bivariada para o redimensionamento do tamanho amostral efetivo**. Orientadora: Dra. Luciana Pagliosa Carvalho Guedes; Coorientador: Dr. Miguel Angel Uribe Opazo. 2022. 80 p. Tese (Doutorado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel – Paraná, 2022.

## RESUMO

Determinar a variabilidade espacial de atributos do solo e da planta é fundamental para o gerenciamento de áreas agrícolas. Essa prática exige que seja realizado um planejamento amostral apropriado que oportunize coletar o menor número possível de amostras georreferenciadas, visando manter a qualidade na amostragem e poupar recursos financeiros e operacionais. Além disso, estudos geoestatísticos sobre a distribuição espacial de processos estocásticos, em que as observações são georreferenciadas, podem evidenciar estruturas espaciais não totalmente independentes entre os atributos. Nesses casos, é recomendado calcular a associação entre as variáveis por meio de uma métrica de correlação espacial bivariada, além de utilizar um modelo geoestatístico bivariado. Sendo assim, no primeiro artigo deste trabalho foi analisada a correlação espacial bivariada, considerando variáveis com diferentes estruturas de dependência espacial, para tal, foi calculado o índice de Lee bivariado. Os resultados mostraram que o raio de dependência espacial comum a ambas às variáveis foi o parâmetro que mais influenciou o valor do índice de Lee, sendo que quanto maior o valor desse parâmetro mais elevada foi a correlação espacial bivariada. No segundo artigo, com base no pressuposto da existência de correlação espacial entre pares de variáveis, o estudo teve como objetivo redimensionar o tamanho amostral, a partir do cálculo do tamanho amostral efetivo bivariado ( $ESS_{bi}$ ), utilizando o modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum (BGCCM). Porquanto, a maioria das propostas na literatura para o  $ESS_{bi}$  utiliza alternativas em suas metodologias que visam evitar o uso de estrutura de correlação espacial entre as variáveis ou consideram o modelo espacial bivariado de correionalização (BCRM). A diferença é que na estrutura do BGCCM, além da existência de um campo aleatório Gaussiano comum compartilhado pelo par de variáveis, existe também um campo aleatório Gaussiano associado à cada variável individualmente. Enquanto no BCRM, necessariamente, a estrutura de correlação espacial individual de uma das variáveis é desconsiderada do modelo. Para verificar a viabilidade teórica da proposta do  $ESS_{bi}$ , todas as propriedades que incidem sobre a metodologia univariada foram verificadas para a bivariada, utilizando estudos de simulação ou de forma algébrica. O  $ESS_{bi}$  também foi aplicado a um conjunto de dados reais de matéria orgânica (MO) e soma de bases (SB), coletados em uma área agrícola com plantio de soja, no qual se constatou que 60% das observações amostrais do par MO-SB continham informações espacialmente duplicadas. Além disso, o redimensionamento amostral obtido para o conjunto de dados reais se mostrou praticável, em termos da qualidade obtida na predição espacial.

**Palavras-chave:** geoestatística, índice de Lee, processos Gaussianos, redução amostral, simulação.

DAL' CANTON, Letícia Ellen. **Bivariate spatial correlation for effective sample size redimensioning**. Advisor: Dr. Luciana Pagliosa Carvalho Guedes; Co-advisor: Dr. Miguel Angel Uribe Opazo. 2022. 80 p. Dissertation (PhD in Agricultural Engineering) – Western Paraná State University, Cascavel – Paraná, 2022.

## ABSTRACT

Determining the spatial variability of soil and plant attributes is critical for farmland management. This practice requires that an appropriate sample planning be carried out in order to collect as few georeferenced samples as possible, with a view to maintaining sampling quality and saving financial and operational resources. Furthermore, geostatistical studies involving the spatial distribution of stochastic processes, in which observations are georeferenced, may show spatial structures that are not totally independent between the attributes. In these cases, it is recommended to calculate the association between the variables by resorting to a bivariate spatial correlation metric, in addition to utilizing a bivariate geostatistical model. Therefore, in the first scientific paper of this research, the bivariate spatial correlation was analyzed considering variables with different structures of spatial dependence and, for this, the bivariate Lee index was calculated. The results showed that the radius of spatial dependence common to both variables was the parameter that most influenced the Lee index value, whereas the higher the value of this parameter, the greater the bivariate spatial correlation. Subsequently, in the second paper, based on the assumption of the existence of spatial correlation between pairs of variables, the study aimed to redimension the sample size by calculating the bivariate effective sample size ( $ESS_{bi}$ ), utilizing the bivariate Gaussian common component model (BGCCM). This is because most of the proposals in the literature for the  $ESS_{bi}$  adopt alternatives in their methodologies that aim to avoid the usage of a spatial correlation structure between the variables or consider the bivariate coregionalization model (BCRM). The difference is that in the BGCCM structure, in addition to having a common Gaussian random field shared by the pair of variables, there is also a Gaussian random field associated with each variable individually. Whilst in BCRM, the individual spatial correlation structure of one of the variables is necessarily disregarded from the model. In order to verify the theoretical feasibility of the  $ESS_{bi}$  proposal, all properties that affect the univariate methodology were verified for the bivariate one, performing simulation studies or algebraically. The  $ESS_{bi}$  was also applied to a real data set of organic matter (OM) and sum of bases (SB), collected in an agricultural area with soybean plantation, in which it was found that 60% of the sample observations of the OM-SB pair contained spatially duplicated information. Moreover, the sample redimensioning obtained for the real data set proved to be feasible in terms of the quality obtained in the spatial prediction.

**Keywords:** geostatistics, Lee index, Gaussian processes, sample reduction, simulation.



# Sumário

	<b>Lista de ilustrações</b>	<b>x</b>
	<b>Lista de tabelas</b>	<b>xi</b>
<b>1</b>	<b>INTRODUÇÃO</b>	<b>10</b>
<b>2</b>	<b>OBJETIVOS</b>	<b>12</b>
2.1	Objetivo geral	12
2.2	Objetivos específicos	12
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>13</b>
3.1	Geoestatística	13
3.2	Modelo espacial Gaussiano	15
3.2.1	Modelo espacial Gaussiano univariado	15
3.2.2	Modelo espacial Gaussiano bivariado	16
3.3	Modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum	17
3.4	Estimação dos parâmetros do modelo por máxima verossimilhança	18
3.5	Crêterios de validação do modelo e predição espacial	20
3.6	Matriz de ponderação espacial e correlação espacial bivariada	22
3.6.1	Matriz de peso por linha $W$	22
3.6.2	Matriz de peso global $C$	22
3.6.3	Índice de Moran bivariado	23
3.6.4	Índice de Lee	23
3.7	Amostragem espacial	24
3.8	Tamanho amostral efetivo	27
3.8.1	Tamanho amostral efetivo univariado	27
3.8.2	Tamanho amostral efetivo bivariado	28
	<b>Referências</b>	<b>31</b>
<b>4</b>	<b>ARTIGOS</b>	<b>36</b>
4.1	Artigo 1: Correlação espacial bivariada entre atributos do solo e produtividade da soja em uma áreia agrícola	36
4.1.1	Introdução	36
4.1.2	Material e métodos	38
4.1.3	Resultados	43
4.1.4	Discussão	49
4.1.5	Conclusão	51
	<b>Referências</b>	<b>52</b>

<b>4.2</b>	<b>Artigo 2: Tamanho Amostral Efetivo utilizando o modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum . . . . .</b>	<b>56</b>
4.2.1	Introdução . . . . .	56
4.2.2	Material e métodos . . . . .	58
4.2.3	Resultados . . . . .	63
4.2.4	Discussão . . . . .	69
4.2.5	Conclusão . . . . .	71
	<b>Referências . . . . .</b>	<b>72</b>
<b>5</b>	<b>CONSIDERAÇÕES FINAIS . . . . .</b>	<b>75</b>
<b>6</b>	<b>APÊNDICES . . . . .</b>	<b>76</b>
6.1	Apêndice A. Matriz de informação de Fisher observada bivariada . . . . .	76
6.2	Apêndice B. Derivadas para obtenção da matriz de informação de Fisher observada bivariada . . . . .	78

# Lista de ilustrações

Figura 3.1 – Configurações amostrais tradicionais: (a) aleatória simples; (b) sistemática; (c) estratificada; (d) por <i>cluster</i> . As circunferências em vermelho indicam as observações selecionadas para amostragem. Nas configurações sem tais circunferências, todos os pontos representam observações amostrais. . . . .	25
Figura 3.2 – Configurações amostrais aprimoradas: (a) <i>lattice plus close pairs</i> ; (b) <i>lattice plus in fill</i> ; (c) <i>multi-stage</i> ; (d) <i>multi-phase</i> ; (e) <i>generalized random tessellation</i> . As circunferências em vermelho indicam as observações selecionadas para amostragem, com exceção de (d), em que se faz uma pré-seleção (em vermelho) e uma seleção final (em amarelo). Nas configurações sem tais circunferências, todos os pontos representam observações amostrais. . . . .	26
Figura 4.1 – (a) Esquema metodológico usado nos estudos de simulação; (b) Área experimental com a localização dos pontos amostrais usando coordenadas UTM. $\mu$ : vetor de médias; $\sigma$ : vetor de dispersão; $\varphi$ : vetor de alcance; $E1, E2, E3$ : ensaios 1 a 3; $\mathbf{W}$ e $\mathbf{C}$ : matrizes de ponderação espacial. . . . .	41
Figura 4.2 – Correlograma (vermelho) e gráfico de envelope (azul) para os nove pares de variáveis simuladas nos ensaios (a) $E1$ e (b) $E2$ , considerando a matriz de pesos $\mathbf{W}$ e a distância euclidiana (DE). . . . .	44
Figura 4.3 – Correlograma (vermelho) e gráfico de envelope (azul) para os nove pares de variáveis simuladas, considerando a matriz de pesos $\mathbf{W}$ e utilizando: (a) a distância euclidiana (DE) no ensaio $E3$ e (b) o inverso da distância euclidiana (IDE). . . . .	45
Figura 4.4 – Correlograma (vermelho) e gráfico de envelope (azul) para os nove pares de variáveis simuladas, considerando: (a) matriz de pesos $\mathbf{C}$ e distância euclidiana (DE), e (b) matriz de pesos $\mathbf{C}$ e o inverso da distância euclidiana (IDE). . . . .	46
Figura 4.5 – Correlograma (vermelho) e gráfico de envelope (azul) para os pares de atributos (a) MO-P, (b) MO-SB, (c) Prod-Ca, (d) Prod-Mg e (e) Prod-SB, considerando a matriz de pesos $\mathbf{W}$ e a distância euclidiana. . . . .	49
Figura 4.6 – (a) Cenários e valores utilizados para o vetor de alcance $\varphi = (\varphi_0, \varphi_1, \varphi_2)^T$ em cada ensaio; (b) Esquema metodológico utilizado nos estudos de simulação. $\mu = (\mu_1, \mu_2)^T$ : vetor de médias; $\sigma = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$ : vetor de dispersão; $E1, \dots, E12$ : ensaios 1 a 12. . . . .	61
Figura 4.7 – (a) Esquema metodológico utilizado nos estudos de dados reais dos atributos do solo matéria orgânica (MO) e soma de bases (SB); (b) Área experimental e esquema amostral com as localizações dos pontos amostrais em coordenadas UTM. . . . .	62
Figura 4.8 – Mapa de variabilidade espacial da matéria orgânica tendo a soma de bases como covariável: (a) Configuração amostral original com 102 pontos amostrais (em preto), (b) Configuração amostral reduzida com 42 pontos amostrais (em vermelho). Valores estimados dos índices Exatidão Global (EG) e índice de concordância Tau (Tau). . . . .	69

# Lista de tabelas

Tabela 4.1 – Valores estimados para o alcance prático ( $a$ ) do semivariograma cruzado e para o alcance prático comum ( $a_0$ ) do BGCCM, considerando os três ensaios. Entre parênteses estão os valores simulados em cada ensaio. . . . .	43
Tabela 4.2 – Estatística descritiva dos atributos: cálcio (Ca, $\text{cmol}_c \text{ dm}^{-3}$ ), fósforo (P, $\text{mg dm}^{-3}$ ), magnésio (Mg, $\text{cmol}_c \text{ dm}^{-3}$ ), matéria orgânica (MO, $\text{g dm}^{-3}$ ), produtividade da soja (Prod, $\text{t ha}^{-1}$ ) e soma de bases (SB, $\text{cmol}_c \text{ dm}^{-3}$ ). . . . .	47
Tabela 4.3 – Valores estimados para os parâmetros do semivariograma cruzado, considerando o melhor modelo geoestatístico escolhido por validação cruzada, para os atributos: cálcio (Ca, $\text{cmol}_c \text{ dm}^{-3}$ ), fósforo (P, $\text{mg dm}^{-3}$ ), magnésio (Mg, $\text{cmol}_c \text{ dm}^{-3}$ ), matéria orgânica (MO, $\text{g dm}^{-3}$ ), produtividade da soja (Prod, $\text{t ha}^{-1}$ ) e soma de bases (SB, $\text{cmol}_c \text{ dm}^{-3}$ ). . . . .	47
Tabela 4.4 – Valores estimados para os parâmetros do modelo bivariado (BGCCM), considerando o melhor modelo geoestatístico escolhido por validação cruzada, para os atributos: cálcio (Ca, $\text{cmol}_c \text{ dm}^{-3}$ ), fósforo (P, $\text{mg dm}^{-3}$ ), magnésio (Mg, $\text{cmol}_c \text{ dm}^{-3}$ ), matéria orgânica (MO, $\text{g dm}^{-3}$ ), produtividade da soja (Prod, $\text{t ha}^{-1}$ ) e soma de bases (SB, $\text{cmol}_c \text{ dm}^{-3}$ ). . . . .	48
Tabela 4.5 – Valores médios estimados para os parâmetros de dispersão ( $\widehat{\sigma}_{01}$ , $\widehat{\sigma}_1$ , $\widehat{\sigma}_{02}$ , $\widehat{\sigma}_2$ ) e de alcance ( $\widehat{\varphi}_0$ , $\widehat{\varphi}_1$ , $\widehat{\varphi}_2$ ), e para o tamanho amostral efetivo bivariado ( $\widehat{ESS}_{bi}$ ), nos cenários 1 a 3. Entre parênteses estão o erro absoluto médio (EAM) das estimativas dos parâmetros e o desvio-padrão (DP) do $\widehat{ESS}_{bi}$ . . . . .	64
Tabela 4.6 – Valores médios estimados para os parâmetros de dispersão ( $\widehat{\sigma}_{01}$ , $\widehat{\sigma}_1$ , $\widehat{\sigma}_{02}$ , $\widehat{\sigma}_2$ ) e de alcance ( $\widehat{\varphi}_0$ , $\widehat{\varphi}_1$ , $\widehat{\varphi}_2$ ), e para o tamanho amostral efetivo bivariado ( $\widehat{ESS}_{bi}$ ) no cenário 4. Entre parênteses estão o erro absoluto médio (EAM) das estimativas dos parâmetros e o desvio-padrão (DP) do $\widehat{ESS}_{bi}$ . . . . .	65
Tabela 4.7 – Valores médios estimados para os parâmetros de dispersão ( $\widehat{\sigma}_{01}$ , $\widehat{\sigma}_1$ , $\widehat{\sigma}_{02}$ , $\widehat{\sigma}_2$ ) e para o tamanho amostral efetivo bivariado ( $\widehat{ESS}_{bi}$ ) no cenário 5. Entre parênteses estão o erro absoluto médio (EAM) das estimativas dos parâmetros e o desvio-padrão (DP) do $\widehat{ESS}_{bi}$ . . . . .	66
Tabela 4.8 – Medidas da análise exploratória (mínimo, máximo, média, DP e CV) dos teores de matéria orgânica (MO, $\text{g dm}^{-3}$ ) e soma de bases (SB, $\text{cmol}_c \text{ dm}^{-3}$ ), utilizando a configuração amostral original ( $n = 102$ ) e reduzida ( $n^* = 42$ ). . . . .	68
Tabela 4.9 – Medidas da análise geoestatística contendo os parâmetros estimados do BGCCM para o par matéria orgânica (MO, $\text{g dm}^{-3}$ ) e soma de bases (SB, $\text{cmol}_c \text{ dm}^{-3}$ ), utilizando a configuração amostral original ( $n = 102$ ) e reduzida ( $n^* = 42$ ). . . . .	68

# 1 Introdução

Ao estudar o cultivo de grãos é fundamental conhecer a dispersão e a variabilidade espacial dos atributos físicos e químicos do solo na área, os quais têm influência direta na produtividade das culturas. Para compreender a distribuição espacial de um fenômeno, são coletados dados provenientes de amostras georreferenciadas, os quais são utilizados para modelar o fenômeno. No contexto da Agricultura de Precisão, a obtenção de informações georreferenciadas é essencial para o gerenciamento agrônomico das lavouras, pois viabiliza o mapeamento da variação da fertilidade do solo nas áreas cultivadas e a definição de intervenções de manejo localizado (MOLIN; AMARAL; COLAÇO, 2015).

A modelagem geoestatística utiliza, além dos valores mensurados de um atributo de interesse, as localizações espaciais em que foram mensurados esses valores e que podem ser expressas por um sistema de coordenadas. Desse modo, as técnicas geoestatísticas possibilitam assimilar a correlação entre as observações das variáveis de interesse, presumindo-se que amostras mais próximas possuem valores mais semelhantes da variável em estudo (CRESSIE, 2015).

Entretanto, em se tratando de pesquisas no âmbito agrícola, geralmente, o interesse não está em uma variável apenas, mas em um conjunto de atributos vinculados ao solo e à planta. Se esses atributos forem independentes, deve-se propor um modelo geoestatístico para cada um. Caso contrário, se existirem evidências de correlação espacial entre algumas das variáveis, preconiza-se que modelos geoestatísticos multivariados sejam utilizados considerando a especificação de distribuições de probabilidade conjuntas (DIGGLE; RIBEIRO JR., 2007).

Para analisar a existência de correlação espacial entre pares de variáveis é fundamental empregar metodologias que considerem simultaneamente informações referentes ao valor e à localização geográfica das observações das variáveis (CIMA et al., 2018). O índice de Lee é uma das métricas existentes na literatura que possibilitam identificar padrões e comportamentos que sugerem associação espacial entre um par de atributos (LEE, 2001).

Para variáveis espacialmente correlacionadas, são encontradas na literatura diversas abordagens para modelos geoestatísticos multivariados (WACKERNAGEL, 2013). Todavia, modelos multivariados tendem a apresentar uma estrutura mais complexa e com maior número de parâmetros a serem estimados, haja vista a necessidade em identificar as estruturas de dependência espacial de cada variável e entre as variáveis. Como neste trabalho serão considerados pares de atributos espacialmente correlacionados, foi utilizado um modelo espacial bivariado. Trata-se do *Bivariate Gaussian Common Component Model* - BGCCM (Modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum), que considera tanto a correlação espacial entre as duas variáveis quanto a associação espacial individual delas (DIGGLE; RIBEIRO JR., 2007).

O uso de metodologias de redimensionamento amostral, aliadas à abordagem bivariada, pode potencializar a redução nos impactos financeiros com o monitoramento de atributos do solo na área agrícola. Dentre essas metodologias, podem ser citados, desde os algoritmos de otimização, especialmente os que envolvem inteligência artificial na busca da melhor configuração amostral (GUEDES et al., 2011, 2016; GUEDES; URIBE-OPAZO; RIBEIRO JR., 2014; WADOUX et al., 2017; MALTAURO; GUEDES; URIBE-OPAZO, 2019), até o uso do tamanho amostral efetivo (GRIFFITH,

2013; ACOSTA; OSORIO; VALLEJOS, 2016; VALLEJOS; ACOSTA, 2021).

O tamanho amostral efetivo univariado (ESS) propicia a redução no número de amostras para uma variável com dependência espacial. Considera-se, nessa metodologia, que alguns pontos amostrais coletados na área agrícola podem estar altamente correlacionados entre si. Fato que gera custos desnecessários com coleta e análises laboratoriais, pois as observações espacialmente autocorrelacionadas agregam informações repetidas quanto à variabilidade espacial do atributo estudado (VALLEJOS; OSORIO, 2014).

O caso univariado do ESS foi generalizado para o multivariado em modelos espaciais gaussianos no artigo de Vallejos e Osorio (2014), sendo este uma versão ponderada do univariado para evitar uma estrutura de correlação espacial entre as variáveis. Posteriormente, Vallejos e Acosta (2021) propuseram outro método multivariado para estimar o ESS, utilizando o modelo espacial bivariado de correionalização (BCRM) para introduzir uma estrutura de correlação espacial entre as variáveis. O BCRM propõe uma estrutura paramétrica em que se tem um parâmetro de correlação espacial comum a duas variáveis e outro parâmetro de correlação espacial exclusivo a uma das variáveis. Dessa forma, necessariamente a correlação espacial da outra variável é desconsiderada do modelo, diferentemente do BGCCM.

Assim, os objetivos centrais estabelecidos para este trabalho foram analisar a correlação espacial entre pares de variáveis por meio do cálculo do índice de Lee e redefinir o tamanho amostral de pares de variáveis espacialmente correlacionadas, utilizando o tamanho amostral efetivo bivariado e o modelo BGCCM.

Para tal, a tese foi elaborada segundo a seguinte estrutura: na primeira parte, tem-se introdução, justificativa e objetivos. Posteriormente, é apresentada uma revisão de literatura com os principais tópicos utilizados no desenvolvimento do trabalho. Os resultados são apresentados na forma de dois artigos científicos. O primeiro artigo trata do uso do índice de Lee para calcular a correlação espacial bivariada, considerando variáveis com diferentes estruturas de dependência espacial. O segundo artigo explora o uso do tamanho amostral efetivo bivariado para obtenção de amostras independentes para pares de variáveis, modeladas pelo BGCCM e, assim, redimensionar o número de amostras coletadas em uma área agrícola. Por fim, tem-se os capítulos com as considerações finais relativas à pesquisa e os Apêndices.

## 2 Objetivos

### 2.1 Objetivo geral

Estimar a correlação espacial bivariada através do índice de Lee e utilizar o BGCCM e o ESS bivariado para redefinir o tamanho amostral em estudos da Geoestatística que envolvem a variabilidade espacial de pares de variáveis georreferenciadas.

### 2.2 Objetivos específicos

- Utilizar o índice de Lee para calcular a correlação espacial bivariada entre diferentes pares de atributos do solo e produtividade da soja;
- Analisar pares de variáveis simuladas com diferentes valores para o alcance prático comum do modelo espacial BGCCM e explorar o índice de Lee para o cálculo da correlação espacial bivariada;
- Calcular o índice de Lee explorando diferentes matrizes de ponderação espacial e métricas para o cálculo da distância entre os pares de pontos;
- Desenvolver uma alternativa à proposta de Vallejos e Osorio (2014) e Vallejos e Acosta (2021) para a estimação do tamanho amostral efetivo bivariado ( $ESS_{bi}$ ), tendo como base a matriz Informação de Fisher e adotando o BGCCM;
- Verificar se as propriedades do ESS univariado também são válidas para o  $ESS_{bi}$  utilizando dados simulados;
- Avaliar a metodologia do  $ESS_{bi}$  utilizando dados experimentais de atributos químicos do solo;
- Observar a influência dos parâmetros que definem a estrutura dependência espacial do modelo BGCCM sob os valores estimados do  $ESS_{bi}$ .

## 3 Revisão Bibliográfica

### 3.1 Geoestatística

A origem da Geoestatística data da década de 1950, quando Daniel Krige, engenheiro de minas da África do Sul, verificou não poderia estimar adequadamente a concentração mineral se não considerasse a configuração espacial das amostras e concluiu que havia maior precisão na estimativa quando se utilizava amostras vizinhas. Passou-se então a considerar que para determinar a variabilidade espacial dos minérios, era preciso ponderar a localização geográfica e a dependência espacial das amostras. Essa concepção deu origem à Teoria das Variáveis Regionalizadas, considerada a base da Geoestatística (LANDIM; YAMAMOTO, 2013).

O uso de técnicas geoestatísticas objetiva descrever espacialmente variáveis de interesse, especialmente atributos do solo e da planta, os quais agregam importantes informações para o gerenciamento localizado do solo realizado pela Agricultura de Precisão (SANCHES; MAGALHÃES; FRANCO, 2019). O comportamento espacial, principalmente, de atributos físico-químicos do solo, está diretamente vinculado ao estudo de sua distribuição espacial na área agrícola experimental e à análise do grau de dependência espacial das amostras coletadas em campo.

A partir de um conjunto de elementos amostrais georreferenciados, pode-se analisar os parâmetros que descrevem a estrutura de dependência espacial do atributo do solo, sendo eles: efeito pepita ( $\tau^2$ ), contribuição ( $\sigma^2$ ) e um parâmetro que é uma função do alcance ( $\varphi$ ) (CRESSIE, 2015).

O efeito pepita associa em uma única parcela, além da variabilidade em escala menor que a amostrada, possíveis erros de medição a campo. A contribuição é um parâmetro que exibe a variância de dispersão, pois representa as diferenças espaciais dos valores da variável em dois pontos. A função de alcance possibilita que seja estimado o alcance prático ( $a = g(\varphi)$ ), o qual indica a amplitude dentro da qual as amostras apresentam autocorrelação espacial (LANDIM; YAMAMOTO, 2013). Utilizando esses parâmetros, é possível utilizar índices para quantificar o grau de dependência espacial das observações coletadas de um atributo, tal como o índice proposto por Cambardella et al. (1994), um dos mais utilizados na literatura.

No estudo da variabilidade espacial dos atributos, considera-se que o valor em um ponto amostral ( $x_1$ ) está relacionado com o valor obtido em outro ponto ( $x_2$ ), sendo razoável supor que a influência no valor mensurado é inversamente proporcional a distância euclidiana entre os pontos (SOARES, 2014). Sendo assim, a cada par de pontos amostrais ( $x_1$  e  $x_2$ ) tem-se uma realização diferente da variável sob estudo, o que torna possível estabelecer o cálculo da covariância ( $C$ ) em função da distância euclidiana ( $h$ ) (CRESSIE, 2015).

A função covariância espacial ( $C(h)$ ) é estimada a partir de um modelo teórico. Entre os principais modelos teóricos apresentados na literatura estão o exponencial, o Gaussiano e os da família Matérn. Os modelos exponencial e Gaussiano serão definidos abaixo segundo Soares (2014) e os da família Matérn de acordo com Matérn (2013).



- **Modelo exponencial**

Como o alcance prático ( $a$ ) é definido a partir de uma função de  $\varphi$  ( $g(\varphi) = a$ ) e cada modelo teórico apresenta um valor estimado distinto para  $\varphi$ , por conseguinte, a forma de estimar o alcance prático depende do modelo. Para o exponencial, o alcance prático é dado por  $a = 3\varphi$ , enquanto a função de covariância é definida como:

$$C(h) = \begin{cases} \tau^2 + \sigma^2, & \text{se } h = 0 \\ \sigma^2 e^{-\left(\frac{h}{\varphi}\right)}, & \text{se } h > 0. \end{cases} \quad (3.1)$$

- **Modelo Gaussiano**

O alcance prático do modelo Gaussiano é igual a  $a = \sqrt{3}\varphi$ . Sua função de covariância espacial é definida por:

$$C(h) = \begin{cases} \tau^2 + \sigma^2, & \text{se } h = 0 \\ \sigma^2 \left[ 1 - e^{-\left(\frac{h}{\varphi}\right)^2} \right], & \text{se } h > 0. \end{cases} \quad (3.2)$$

- **Modelo da família Matérn**

A função de covariância para a família Matérn é dada por:

$$C(h) = \begin{cases} \tau^2 + \sigma^2, & \text{se } h = 0 \\ \sigma^2 \left[ (2^{\kappa-1} \Gamma(\kappa))^{-1} \left(\frac{h}{\varphi}\right)^{\kappa} K_{\kappa} \left(\frac{h}{\varphi}\right) \right], & \text{se } h > 0. \end{cases} \quad (3.3)$$

Na função de covariância da família Matérn, têm-se as funções auxiliares Gama ( $\Gamma$ ) e Bessel ( $K_{\kappa}$ ). A função Gama ( $\Gamma$ ) é expressa por:

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

Considerando  $\kappa > 0$ , a função de Bessel ( $K_{\kappa}$ ) do terceiro tipo de ordem  $\kappa$ , é definida como:

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

Segundo Diggle e Ribeiro Jr. (2007), devido à impossibilidade de se considerar a covariância para todos os valores do parâmetro de forma  $\kappa$ , geralmente é fixado um valor para  $\kappa > 0$ , de acordo com o contexto da aplicação, ou utiliza-se um conjunto limitado de valores para este parâmetro.

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,  $3\varphi$ ,  $3,45\varphi$ ,  $4\varphi$ ,  $4,75\varphi$ ,  $5,37\varphi$  e  $5,92\varphi$  quando se considera  $\kappa = 0,5$ ,  $\kappa = 0,7$ ,  $\kappa = 1$ ,  $\kappa = 1,5$ ,  $\kappa = 2$  e  $\kappa = 2,5$ , respectivamente (DIGGLE; RIBEIRO JR., 2007).

## 3.2 Modelo espacial Gaussiano

A modelagem espacial tem como objetivo descrever o comportamento de uma ou mais variáveis que não podem ser caracterizadas exclusivamente por seus valores médios.

Em pesquisas realizadas no âmbito agrícola, geralmente, estuda-se mais de um atributo físico-químico do solo. Se o estudo desses atributos é feito de forma individual, propondo-se um modelo espacial para cada um deles, diz-se que o modelo espacial é univariado (URIBE-OPAZO; BORSSOI; GALEA, 2012) (Seção 3.2.1). Por outro lado, existindo evidências de correlação entre os atributos estudados, pode-se utilizar modelos espaciais multivariados, que consideram a correlação entre as variáveis através da especificação de distribuições de probabilidades conjuntas (DIGGLE; RIBEIRO JR., 2007). Neste trabalho, será apresentado o modelo espacial Gaussiano bivariado, que agrega duas variáveis normalmente distribuídas e correlacionadas entre si (Seção 3.2.2).

### 3.2.1 Modelo espacial Gaussiano univariado

Considere uma área de interesse  $S \subset \mathbb{R}^r$ , em que  $\mathbb{R}^r$  é um espaço euclidiano  $r$ -dimensional,  $r \geq 1$ . Suponha que em  $S$  exista um campo aleatório Gaussiano  $Z$  no qual se deseja realizar uma amostragem de  $n$  pontos georreferenciados, observando o valor de algum atributo físico-químico do solo, denotado por  $Y$  (CHIPETA et al., 2017).

Seja  $\mathbf{Y} = (Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n))^T$  o vetor aleatório de dimensão  $n \times 1$ , das  $n$  observações da variável aleatória  $Y$ . Considere que  $Y(\mathbf{s}_1), \dots, Y(\mathbf{s}_n)$  são observações da variável aleatória sob estudo nas  $\mathbf{s}_m$  localizações espaciais amostradas,  $m = 1, \dots, n$ . Cada observação  $Y(\mathbf{s}_m)$  é gerada, sob o pressuposto de ter distribuição gaussiana de probabilidade, pelo seguinte modelo espacial linear (DIGGLE; RIBEIRO JR., 2007):

$$Y(\mathbf{s}_m) = \mu(\mathbf{s}_m) + Z(\mathbf{s}_m) + \epsilon(\mathbf{s}_m), \quad m = 1, \dots, n, \quad (3.4)$$

em que  $\mu(\mathbf{s}_m)$  é o termo determinístico,  $Z(\mathbf{s}_m)$  é um campo aleatório Gaussiano e  $\epsilon(\mathbf{s}_m)$  é o termo estocástico, e todos dependem da localização espacial  $\mathbf{s}_m$  em que a variável  $Y$  é observada.

Para algumas funções conhecidas de  $\mathbf{s}_m$ ,  $d_1(\mathbf{s}_m), \dots, d_p(\mathbf{s}_m)$ , tem-se que a média do processo estocástico é dada por (URIBE-OPAZO; BORSSOI; GALEA, 2012):

$$\mu(\mathbf{s}_m) = \sum_{j=1}^p d_j(\mathbf{s}_m) \beta_j, \quad (3.5)$$

em que  $d_j(\mathbf{s}_m)$ ,  $m = 1, \dots, n$ ,  $j = 1, \dots, p$ , representa o valor da  $j$ -ésima covariável tomada na  $n$ -ésima posição e  $\beta_j$  é um  $j$ -ésimo parâmetro desconhecido a ser estimado. Considerando os  $n$  pontos amostrais, pode-se escrever o vetor de médias  $\boldsymbol{\mu}$ , de ordem  $n \times 1$ , dado por:  $\boldsymbol{\mu} = \mathbf{D}\boldsymbol{\beta}$ , em que  $\mathbf{D}$ , com dimensão  $n \times p$ , é denominada matriz de delineamento, na qual a  $m$ -ésima linha é dada por  $\mathbf{d}_m^T = (d_1(\mathbf{s}_m), \dots, d_p(\mathbf{s}_m))$ ; e o vetor  $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ , de ordem  $p \times 1$ , agrega os parâmetros associados às  $p$  covariáveis (URIBE-OPAZO et al. 2021).

Em particular, quando se considera um processo estocástico estacionário (SOARES, 2014), tem-se que  $\mathbf{D} = \mathbf{1}$  e  $\boldsymbol{\beta} = \mu$ , sendo  $\mathbf{1}$  um vetor de uns. Logo,  $\boldsymbol{\mu} = \mu\mathbf{1}$ . Isto significa que para qualquer localização  $\mathbf{s}_m$  a média é constante igual a  $\mu$ .

O campo aleatório Gaussiano  $Z$  (Equação 3.4) tem vetor de médias nulo e matriz de covariância  $\Sigma_Z$ ,  $n \times n$ , que é dada por  $\Sigma_Z = [(c_{ml})]$ ,  $m, l = 1, \dots, n$ , em que  $c_{ml} = C(Z(\mathbf{s}_m), Z(\mathbf{s}_l))$  é a covariância entre os valores da variável georreferenciada nos pontos  $\mathbf{s}_m$  e  $\mathbf{s}_l$ .

Vale observar que  $C(Z(\mathbf{s}_m), Z(\mathbf{s}_l)) = \sigma^2 \rho(Z(\mathbf{s}_m), Z(\mathbf{s}_l))$ , em que  $\rho(Z(\mathbf{s}_m), Z(\mathbf{s}_l))$  é a correlação entre os valores da variável georreferenciada nos pontos  $\mathbf{s}_m$  e  $\mathbf{s}_l$ . Desta forma,  $\Sigma_Z = \sigma^2 \mathbf{R}(\varphi)$ , sendo  $\sigma^2$  e  $\varphi$ , respectivamente, os parâmetros contribuição e função de alcance do modelo geoestatístico, e  $\mathbf{R} = \mathbf{R}(\varphi) = [(r_{ml})]$  é uma matriz simétrica,  $n \times n$ , cujos elementos da diagonal são dados por  $r_{mm} = 1$ ,  $m = 1, \dots, n$ , enquanto os demais elementos são denotados por  $r_{ml} = (1/\sigma^2)c_{ml}$ , para  $m \neq l = 1, \dots, n$  e  $\sigma^2 \neq 0$ , e  $r_{ml} = 0$  para  $m \neq l = 1, \dots, n$  e  $\sigma^2 = 0$  (URIBE-OPAZO; BORSSOI; GALEA, 2012).

Além disso, o vetor dos  $n$  erros aleatórios ( $\epsilon$ ) (Equação 3.4) é caracterizado por  $\epsilon \sim N_n(\mathbf{0}, \Sigma_e)$ , sendo  $\Sigma_e = \tau^2 \mathbf{I}$ , em que  $\mathbf{I}$  é uma matriz identidade de dimensão  $n \times n$ .

Sendo assim,  $\mathbf{Y} \sim N_n(\mathbf{D}\beta, \Sigma_Y)$ . Tal simbologia indica que o vetor aleatório  $\mathbf{Y}$  tem distribuição gaussiana, com vetor de médias  $\mathbf{D}\beta$  e matriz de covariância  $\Sigma_Y = \Sigma_z + \Sigma_e = \sigma^2 \mathbf{R} + \tau^2 \mathbf{I}$ . Logo, o vetor de parâmetros a estimar é dado por  $\theta = (\beta, \sigma^2, \varphi, \tau^2)^T$ .

### 3.2.2 Modelo espacial Gaussiano bivariado

Diferentemente do proposto na seção anterior (3.2.1), nos casos em que existem evidências estatísticas de correlação espacial entre dois atributos sob estudo, tem-se interesse em modelar e descrever o padrão espacial dessas variáveis, considerando um modelo espacial Gaussiano bivariado

Sendo assim, dada a correlação espacial entre dois atributos, suponha que, associados a eles, existam dois campos aleatórios Gaussianos que podem ser modelados por (FONSECA, 2008):

$$Y_k(\mathbf{s}_m) = \mu_k(\mathbf{s}_m) + Z_k(\mathbf{s}_m) + \epsilon_k(\mathbf{s}_m), \quad k = 1, 2 \quad \text{e} \quad m = 1, \dots, n_k, \quad (3.6)$$

em que  $Y_k$  representa a  $k$ -ésima variável de interesse e, portanto, considerando todos os pontos amostrais: tem-se que  $\mathbf{Y}_{bi}$  é um vetor aleatório, de dimensão  $n_k \times 2$ , dos valores observados nos  $n_k$  pontos do campo aleatório Gaussiano  $Z_k$ . O campo aleatório  $Z_k$  tem vetor de médias nulo, de ordem  $n_k \times 1$  e matriz de covariância  $\Sigma_k$ , de ordem  $n_k \times n_k$ . Além disso,  $\mu_k$ , de dimensão  $n_k \times 1$ , representa o termo determinístico e pode ser reescrito de forma similar a Equação 3.5, apresentada no modelo univariado. Segundo Diggle e Ribeiro Jr. (2007), o termo estocástico  $\epsilon_k$ , de ordem  $n_k \times 1$ , pode ser desprezado. Assim a Equação 3.6 é reescrita e considerada em sua forma mais simples.

Como se está considerando dois atributos, tem-se que  $\mathbf{Y}_{bi} = (\mathbf{Y}_1, \mathbf{Y}_2)^T$ , donde  $\mathbf{Y}_{bi}$  possui uma distribuição gaussiana  $n$ -variada, com  $n = n_1 + n_2$ , em que  $n_1$  e  $n_2$  são os tamanhos amostrais de  $\mathbf{Y}_1$  e  $\mathbf{Y}_2$ , respectivamente. Além disso,  $\mathbf{Y}_{bi}$  tem vetor de médias  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)^T$  e matriz de covariâncias positiva definida, dada por (FONSECA, 2008):

$$\Sigma_Y = \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_2 \end{pmatrix}, \quad (3.7)$$

em que  $\Sigma_k$ , de dimensão  $n_k \times n_k$ , com  $k = 1, 2$ , é a matriz de covariância do  $k$ -ésimo atributo; e  $\Sigma_{1,2}$ , de ordem  $n_1 \times n_2$ , é a matriz com as covariâncias cruzadas entre atributos  $Y_1$  e  $Y_2$ .

A literatura apresenta algumas propostas para estudo dessas situações em que dois atributos têm correlação espacial, como o modelo bivariado de co-regionalização (GELFAND et al., 2004) e o modelo bivariado com componente de correlação parcialmente comum (DIGGLE; RIBEIRO JR., 2007). Este último será empregado no presente estudo e descrito na sequência.

### 3.3 Modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum

Embora inúmeros trabalhos sobre modelos espaciais bivariados sejam encontrados na literatura (BOGNOLA et al., 2008; MARTINS; BONAT; RIBEIRO JR., 2016; GIORGI; SCHLÜTER; DIGGLE, 2018), estão em menor número dos que utilizam o componente de correlação parcialmente comum. Uma das principais motivações pode estar vinculada às adaptações necessárias ao modelo em casos de correlações negativas (RIGHETTO, 2012). Isso faz do *Bivariate Gaussian Common Component Model* - BGCCM (modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum) um campo a ser explorado.

Na proposta do BGCCM, desenvolvida por Diggle e Ribeiro Jr. (2007), os autores sugerem que o campo aleatório latente  $(Z_k)$  (Equação 3.6) seja particionado da seguinte forma:

$$Z_k = \sigma_{0k}S_0 + \sigma_k S_k, \quad k = 1, 2, \quad (3.8)$$

em que  $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$  é um vetor de parâmetros de dispersão estimados vinculados ao modelo (Equação 3.6);  $S_0, S_1$  e  $S_2$  são campos aleatórios Gaussianos mutuamente independentes, que apresentam vetores de médias nulos, variâncias unitárias e cujas correlações são determinadas por funções de correlação  $\rho$ . O campo aleatório  $S_0$  é comum às variáveis  $Y_1$  e  $Y_2$ , enquanto  $S_1$  e  $S_2$  estão associados a cada variável individualmente (RIGHETTO, 2012). Sendo assim, o modelo BGCCM exhibe uma estrutura de covariância construída a partir de três funções de correlação válidas para  $S_0, S_1$  e  $S_2$ , as quais serão denotadas respectivamente por  $\rho_0, \rho_1$  e  $\rho_2$ .

Considerando a decomposição do campo aleatório latente, pode-se reescrever a Equação 3.6 como (FONSECA, 2008):

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}S_0 + \sigma_1 S_1 \\ Y_2 = \mu_2 + \sigma_{02}S_0 + \sigma_2 S_2, \end{cases} \quad (3.9)$$

em que:  $\mu_1, \mu_2$  são as médias das variáveis  $Y_1$  e  $Y_2$ , respectivamente;  $\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2$  são os parâmetros de dispersão;  $S_0, S_1$  e  $S_2$  são campos aleatórios Gaussianos.

Suponha que  $Y_k(\mathbf{s}_m)$  e  $Y_k(\mathbf{s}_l)$  sejam observações do atributo  $Y_k$  mensuradas nas localizações  $\mathbf{s}_m$  e  $\mathbf{s}_l$ , as quais estão separadas por uma distância euclidiana  $h = h_{ml}$ , com  $m, l = 1, \dots, n_k$  e  $k = 1, 2$ . Desta forma, os elementos da matriz de covariâncias  $\Sigma_Y$ , são dados por:

$$C(h) = \sigma_{0k}^2 \rho_0(h) + \sigma_k^2 \rho_k(h), \quad (3.10)$$

em que:  $C$  representa a função de covariância em relação a distância  $h$ ;  $\rho_0, \rho_k$  são as funções de correlação em  $S_0$  e  $S_k$ , respectivamente,  $k = 1, 2$ .

Utilizando a Equação 3.10, a matriz de covariâncias para o vetor  $\mathbf{Y}_{bi} = (Y_k(\mathbf{s}_m), Y_k(\mathbf{s}_l))^T$ , com  $k = 1, 2$  e  $m, l = 1, \dots, n_k$ , é dada por (FONSECA, 2008):

$$\Sigma_Y = \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_2 \end{pmatrix}, \quad (3.11)$$

sendo  $\Sigma_k$  a matriz de covariância da variável  $Y_k$ ,  $n_k \times n_k$ ,  $k = 1, 2$ , dada por:

$$\Sigma_k = \begin{pmatrix} \sigma_{0k}^2 + \sigma_k^2 & \sigma_{0k}^2 \rho_0(h) + \sigma_k^2 \rho_k(h) \\ \sigma_{0k}^2 \rho_0(h) + \sigma_k^2 \rho_k(h) & \sigma_{0k}^2 + \sigma_k^2 \end{pmatrix}, \quad (3.12)$$

e  $\Sigma_{1,2}$ ,  $n_1 \times n_2$ , é a matriz de covariância cruzada entre as variáveis  $Y_1$  e  $Y_2$ , dada por:

$$\Sigma_{1,2} = \begin{pmatrix} \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) \\ \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} \end{pmatrix}, \quad (3.13)$$

sendo:  $\rho_0$  e  $\rho_k$  funções de correlação;  $\sigma_{0k}$  e  $\sigma_k$  parâmetros de dispersão em  $S_0$  e  $S_k$  para  $k = 1, 2$ .

Logo, a distribuição de probabilidade do vetor  $\mathbf{Y}_{bi}$  depende da estimação do vetor de parâmetros  $\theta = (\beta^*, \sigma^*, \varphi^*)^T$ , em que o vetor  $\beta^*$  está associado a média  $\mu$ , o vetor  $\sigma^*$  à dispersão, e  $\varphi^* = (\varphi_0, \varphi_1, \varphi_2)^T$  à função de alcance  $\varphi_t$ , a qual está vinculada ao modelo geoestatístico escolhido para a função de correlação espacial  $\rho_t$ . O  $t$ -ésimo índice está relacionado ao campo aleatório  $S_t$ ,  $t = 0, 1, 2$ .

Segundo Righetto (2012), a estimação dos parâmetros de  $\theta$  (Seção 3.4) e a predição espacial (Seção 3.5) seguem os mesmos moldes de técnicas geoestatísticas univariadas. Neste trabalho, será utilizado o método de máxima verossimilhança para estimação dos parâmetros e, para tal, necessita-se da função de máxima verossimilhança.

### 3.4 Estimação dos parâmetros do modelo por máxima verossimilhança

Considerando um modelo geoestatístico univariado isotrópico (SOARES, 2014), o conjunto das  $n$  observações de  $\mathbf{Y}$  pode ser descrito com a seguinte distribuição de probabilidade (DIGGLE; RIBEIRO JR., 2007):

$$\mathbf{Y} \sim N(\mathbf{D}\beta, \sigma^2\mathbf{R} + \tau^2\mathbf{I}),$$

em que  $\mathbf{D}\beta$  é o vetor de médias (ver Equação 3.5),  $\Sigma = \sigma^2\mathbf{R} + \tau^2\mathbf{I}$  é a matriz de covariância, análoga à definida na Seção 3.2.1,  $\sigma^2$  é o parâmetro de dispersão do modelo,  $\mathbf{R}$  é uma matriz de correlação, de dimensão  $n \times n$ ,  $\tau^2$  é o efeito pepita, isto é, o parâmetro que corresponde a variância associada ao erro e  $\mathbf{I}$  é uma matriz identidade  $n \times n$ .

Sendo assim, deve-se obter o vetor  $\theta = (\beta, \sigma^2, \varphi, \tau^2)^T$  que maximize a função de verossimilhança. O resultado é o logaritmo da função de verossimilhança dado por (DIGGLE; RIBEIRO JR., 2007):

$$l(\theta) = -\frac{1}{2} \left[ \ln(|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|) + n \ln(2\pi) + (\mathbf{Y} - \mathbf{D}\beta)^T (\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1} (\mathbf{Y} - \mathbf{D}\beta) \right], \quad (3.14)$$

como  $\Sigma = \sigma^2 \mathbf{R} + \tau^2 \mathbf{I}$ , a Equação 3.14 pode ser reescrita da seguinte forma:

$$l(\theta) = -\frac{1}{2} \left[ \ln(|\Sigma|) + n \ln(2\pi) + (\mathbf{Y} - \mathbf{D}\beta)^T \Sigma^{-1} (\mathbf{Y} - \mathbf{D}\beta) \right].$$

Esse raciocínio pode ser estendido para o modelo bivariado descrito na Equação 3.6. Cantú (2015) descreve o logaritmo da função de verossimilhança para esse modelo:

$$l(\theta) = -\frac{1}{2} \left[ \ln(|\Sigma_Y|) + n \ln(2\pi) + (\mathbf{Y}_{bi} - \mathbf{D}\beta)^T \Sigma_Y^{-1} (\mathbf{Y}_{bi} - \mathbf{D}\beta) \right]. \quad (3.15)$$

Precisa-se então definir como é dada a matriz de covariâncias  $\Sigma_Y$  considerando o modelo BGCCM. Para tal, utilizando-se propriedades de covariância, as matrizes apresentadas nas Equações 3.12 e 3.13 serão reescritas como um bloco de matrizes formada pela matriz de covariância individual de cada atributo (Equações 3.16 e 3.17) e pela matriz de covariância cruzada (Equação 3.18) (CANTÚ, 2015):

$$\Sigma_1 = \sigma_{01}^2 \mathbf{R}_0 + \sigma_1^2 \mathbf{R}_1, \quad (3.16)$$

$$\Sigma_2 = \sigma_{02}^2 \mathbf{R}_0 + \sigma_2^2 \mathbf{R}_2, \quad (3.17)$$

$$\Sigma_{1,2} = \sigma_{01} \sigma_{02} \mathbf{R}_0, \quad (3.18)$$

em que:  $\mathbf{R}_k$  é a matriz de correlação de  $Y_k$  e depende da função de correlação associada ao campo aleatório  $S_k$ ,  $k = 1, 2$ ;  $\mathbf{R}_0$  é a matriz das correlações cruzadas e depende da função de correlação associada ao campo aleatório  $S_0$ ;  $\sigma_{0k}$ ,  $\sigma_k$  são os parâmetros de dispersão do  $k$ -ésimo atributo em  $S_0$  e  $S_k$ , respectivamente.

Considerando as reparametrizações  $\sigma = \sigma_{01}$ ,  $\eta = \frac{\sigma_{02}}{\sigma_{01}}$ ,  $\tau_1 = \frac{\sigma_1}{\sigma_{01}}$  e  $\tau_2 = \frac{\sigma_2}{\sigma_{02}}$ , propostas por Diggle e Ribeiro Jr. (2007), pode-se reescrever as Equações 3.16, 3.17 e 3.18, da seguinte forma (CANTÚ, 2015):

$$\begin{aligned} \Sigma_1 &= \sigma_{01}^2 \mathbf{R}_0 + \sigma_1^2 \mathbf{R}_1 \\ &= \sigma^2 \mathbf{R}_0 + \tau_1^2 \sigma_{01}^2 \mathbf{R}_1 \\ &= \sigma^2 \mathbf{R}_0 + \tau_1^2 \sigma^2 \mathbf{R}_1 \\ &= \sigma^2 (\mathbf{R}_0 + \tau_1^2 \mathbf{R}_1) \\ &= \sigma^2 \mathbf{V}_1, \end{aligned} \quad (3.19)$$

sendo  $v_1 = \rho_0(h) + \tau_1^2 \rho_1(h)$  elementos da matriz  $\mathbf{V}_1$ , de dimensão  $n_1 \times n_1$ .

$$\begin{aligned} \Sigma_2 &= \sigma_{02}^2 \mathbf{R}_0 + \sigma_2^2 \mathbf{R}_2 \\ &= \eta^2 \sigma^2 \mathbf{R}_0 + \tau_2^2 \sigma_{02}^2 \mathbf{R}_2 \\ &= \eta^2 \sigma^2 \mathbf{R}_0 + \tau_2^2 \eta^2 \sigma^2 \mathbf{R}_2 \\ &= \eta^2 \sigma^2 (\mathbf{R}_0 + \tau_2^2 \mathbf{R}_2) \\ &= \eta^2 \sigma^2 \mathbf{V}_2, \end{aligned} \quad (3.20)$$

em que  $v_2 = \rho_0(h) + \tau_2^2 \rho_2(h)$  são os elementos da matriz  $\mathbf{V}_2$ ,  $n_2 \times n_2$ .

$$\begin{aligned}
\Sigma_{1,2} &= \sigma_{01}\sigma_{02}\mathbf{R}_0 \\
&= \sigma\eta\sigma_{01}\mathbf{R}_0 \\
&= \sigma^2\eta\mathbf{R}_0 \\
&= \sigma^2\eta\mathbf{V}_{1,2}.
\end{aligned} \tag{3.21}$$

sendo  $v_{1,2} = \rho_0(h)$  elementos da matriz  $\mathbf{V}_{1,2}$ ,  $n_1 \times n_2$ .

Dessa forma, após as reparametrizações e considerando-se o modelo BGCCM, a matriz das covariâncias  $\Sigma_Y$  (Equação 3.7) é descrita por Cantú (2015) como:

$$\Sigma_Y = \sigma^2 \begin{pmatrix} \mathbf{V}_1 & \eta\mathbf{V}_{1,2} \\ \eta\mathbf{V}_{1,2}^T & \eta^2\mathbf{V}_2 \end{pmatrix} = \sigma^2\mathbf{V}. \tag{3.22}$$

Sendo assim, ao considerar uma reparametrização, os parâmetros a serem estimados pelo processo de maximização do logaritmo da função de verossimilhança (Equação 3.15), são:  $\eta$ ,  $\tau_1$ ,  $\tau_2$ ,  $\varphi_0$ ,  $\varphi_1$ ,  $\varphi_2$ , cujas estimativas são obtidas por meio de algoritmos computacionais; e  $\beta$  e  $\sigma^2$ , que possuem formas analíticas fechadas, têm expressões que são verificadas diferenciando-se a função descrita na Equação 3.15, em relação aos respectivos parâmetros, obtendo-se:

$$\hat{\beta} = (\mathbf{D}^T \Sigma_Y^{-1} \mathbf{D})^{-1} (\mathbf{Y}^T \Sigma_Y^{-1} \mathbf{D}), \tag{3.23}$$

$$\hat{\sigma}^2 = \frac{(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{n}, \tag{3.24}$$

em que, da Equação 3.22, tem-se  $\mathbf{V} = \frac{1}{\sigma^2} \Sigma_Y$ .

A diferenciação do logaritmo da função de verossimilhança, em relação aos parâmetros  $\beta$  e  $\sigma^2$ , que culmina nas expressões 3.23 e 3.24, são apresentados no Apêndice B (Seção 6.2).

### 3.5 Critérios de validação do modelo e predição espacial

Neste trabalho, o ajuste de um modelo teórico à função de covariância foi avaliado segundo os critérios de Validação Cruzada (VC). Na aplicação da VC para modelos espaciais gaussianos, 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 o valores previstos com os amostrados (ISAAKS; SRIVASTAVA, 1989).

Na Validação Cruzada, Faraco et al. (2008) sugerem os seguintes critérios para avaliação dos modelos: erro médio, desvio padrão do erro médio, erro médio reduzido, desvio padrão do erro médio reduzido e erro absoluto.

O Erro Médio (EM), o Desvio Padrão do Erro Médio (DPEM) e o Erro Absoluto (EA) são dados, respectivamente, por:

$$EM = \frac{1}{n} \sum_{i=1}^n \left( Y_k(\mathbf{s}_i) - \hat{Y}_k(\mathbf{s}_i) \right), \quad (3.25)$$

$$DPEM = \sqrt{\frac{1}{n} \sum_{i=1}^n |Y_k(\mathbf{s}_i) - \hat{Y}_k(\mathbf{s}_i)|}, \quad (3.26)$$

$$EA = \sum_{i=1}^n |Y_k(\mathbf{s}_i) - \hat{Y}_k(\mathbf{s}_i)|, \quad (3.27)$$

em que  $n$  é o número de observações realizadas;  $Y_k(\mathbf{s}_i)$  é o valor da  $k$ -ésima variável aleatória observada no ponto  $\mathbf{s}_i$ ,  $k = 1, 2$ ; e  $\hat{Y}_k(\mathbf{s}_i)$  é o valor de  $Y_k$  no ponto  $\mathbf{s}_i$  predito por krigagem, sem considerar a observação  $Y_k(\mathbf{s}_i)$ .

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

$$EMR = \frac{1}{n} \sum_{i=1}^n \frac{(Y_k(\mathbf{s}_i) - \hat{Y}_k(\mathbf{s}_i))}{\ddot{\sigma}(\hat{Y}_k(\mathbf{s}_i))}, \quad (3.28)$$

em que  $\ddot{\sigma}(\hat{Z}(\mathbf{s}_i))$  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 (DPEMR), obtido a partir da seguinte equação:

$$DPEMR = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|Y_k(\mathbf{s}_i) - \hat{Y}_k(\mathbf{s}_i)|}{\ddot{\sigma}(\hat{Y}_k(\mathbf{s}_i))}}. \quad (3.29)$$

Considerando-se a condição de estacionariedade, o valor populacional do EMR é zero e do DPEMR é igual a 1. Logo, deseja-se que o EM, o EMR e o DPEM sejam o mais próximos de zero possível. O DPEMR 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).

No que se refere à interpolação espacial, dentre as técnicas apresentadas na literatura (CRESSIE, 2015), a krigagem foi escolhida para ser empregada neste trabalho. Landim e Yamamoto (2013) apresentam vários métodos de krigagem, sendo a krigagem ordinária a mais utilizada na prática e que apresenta o seguinte estimador:

$$\hat{Y}_k(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Y_k(\mathbf{s}_i), \quad (3.30)$$

em que:  $\hat{Y}_k(\mathbf{s}_0)$  é o valor estimado da  $k$ -ésima variável georreferenciada,  $k = 1, 2$ , no local  $\mathbf{s}_0$  não amostrado;  $n$  é o número de observações realizadas e  $\lambda_i$  são os pesos associados a cada valor  $Y_k(\mathbf{s}_i)$  mensurado,  $i = 1, \dots, n$ , os quais são baseados nas distâncias entre os pontos vizinhos e destes em relação ao predito. Os cálculos matriciais envolvendo os pesos podem ser encontrados em Landim e Yamamoto (2013).

O estimador definido em na Equação 3.30 é tido como *Best Linear Unbiased Estimator* - BLUE, ou seja, é o melhor estimador por fornecer predições não-viesadas e com variâncias mínimas.



Tais qualidades são garantidas, pois os pesos  $\lambda_m$  são determinados impondo-se que a esperança do erro seja zero (não-viesada) e a variância do erro seja mínima (SOARES, 2014).

### 3.6 Matriz de ponderação espacial e correlação espacial bivariada

Uma matriz de ponderação espacial é uma matriz quadrada, de dimensão  $n \times n$ , sendo  $n$  o número de observações, em que os pesos espaciais  $w_{ij}$  representam o grau de conexão entre as regiões, segundo algum critério de proximidade, mostrando a influência da localização  $j$  sobre a localização  $i$ . Desta forma, a matriz de pesos apresenta uma espécie de ponderação da influência que as localizações exercem entre si (ALMEIDA, 2013).

A distância geográfica pode ser utilizada para estabelecer o grau de conexão expresso nas matrizes de pesos espaciais. Neste trabalho foram utilizadas as matrizes de ponderação espacial **W** (BIVAND; WONG, 2018) e **C** (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2013), ambas de dimensão  $n \times n$ . A distância geográfica, por sua vez, depende de uma determinada métrica. Para o desenvolvimento desta tese foram utilizadas as métricas de distância euclidiana e do inverso da distância euclidiana.

#### 3.6.1 Matriz de peso por linha **W**

A matriz de ponderação espacial  $\mathbf{W} = [(w_{ij})]$  é padronizada por linha, de modo que a soma dos pesos de cada linha seja igual a 1. Para a distância euclidiana, o cálculo do peso  $w_{ij}$  é vinculado ao número de vizinhos, sendo dado por  $w_{ij} = \frac{1}{v_i}$ , em que  $v_i$  é o número total de vizinhos na  $i$ -ésima linha, considerando as colunas  $j$  que possuem vizinhos. Para as localizações que não possuem vizinhos, é atribuído peso  $w_{ij} = 0$ . São considerados vizinhos as localizações  $i$  e  $j$  cuja distância é inferior a distância de corte. Para o inverso da distância euclidiana, o peso é vinculado apenas as distâncias, sendo dado por  $w_{ij} = \frac{1}{\sum_{j=1}^n 1/d_j} \frac{1}{d_{ij}}$ , em que  $\sum_{j=1}^n 1/d_j$  representa a soma do inverso das distâncias entre os vizinhos na  $i$ -ésima linha considerando as  $n$  colunas, e  $\frac{1}{d_{ij}}$  é o inverso da distância do elemento pertencente a linha  $i$  e coluna  $j$ ,  $i \neq j = 1, \dots, n$  (BIVAND; WONG, 2018). Para  $i = j$ ,  $w_{ij} = 0$ .

#### 3.6.2 Matriz de peso global **C**

A matriz de peso  $\mathbf{C} = [(c_{ij})]$  é globalmente padronizada, fazendo com que a soma dos pesos seja  $n$ , e alterando o modo com que os pesos  $c_{ij}$  são obtidos. Para a distância euclidiana, a ponderação é dada por  $c_{ij} = \frac{1}{v}$ , em que  $v$  é o número total de vizinhos considerando os  $n$  pontos. Enquanto para o inverso da distância euclidiana, o peso é calculado por  $c_{ij} = \frac{1}{\sum_{i,j=1}^n 1/d_{ij}} \frac{1}{d_{ij}}$ , em que  $\sum_{i,j=1}^n 1/d_{ij}$  representa a soma do inverso das distâncias considerando todos os vizinhos (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2013). Para  $i = j$ ,  $c_{ij} = 0$ .

Utilizando uma matriz de ponderação espacial, é possível estabelecer a correlação espacial entre duas variáveis georreferenciadas em  $n$  pontos amostrais. Dentre as expressões apresentadas na literatura para esse fim, pode-se destacar os índices de Moran bivariado (ANSELIN; SYABRI; SMIRNOV, 2002; ALMEIDA, 2013) e de Lee (LEE, 2001).

### 3.6.3 Índice de Moran bivariado

A maioria dos trabalhos existentes na literatura, que utilizam a correlação espacial bivariada, fazem uso do índice de Moran bivariado (CIMA et al., 2018; WIECZOREK et al. 2018; ADEGBOYE et al., 2021; QUEIROZ; MEDRONHO, 2021; RIBEIRO et al., 2021). O uso da correlação espacial bivariada aborda desde o agronegócio, em como associar a produção agrícola e a capacidade de armazenamento de grãos; a saúde pública, em que se pode associar número de casos de COVID-19 e o fluxo de viagens ou a transmissão de doenças e regiões com diferentes indicadores socioeconômicos; o nicho ambiental, no qual se correlaciona a quantidade de metais no solo com impactos ambientais para avaliar risco ecológico.

O índice de Moran bivariado possui duas principais expressões, descritas por Anselin, Syabri e Smirnov (2002) (Equação 3.31) e por Almeida (2013) (Equação 3.32):

$$I_{xy} = \frac{\sum_{i=1}^n \sum_{j=1}^n (x_i - \bar{x})(y_j - \bar{y})w_{ij}}{\sum_{i=1}^n \sum_{j=1}^n w_{ij} \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n} \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}}, \quad (3.31)$$

$$I_{x_1x_2} = \frac{n}{\sum_{i=1}^n \sum_{j=1}^n w_{ij}} \frac{\sum_{i=1}^n \sum_{j=1}^n (x_{1i} - \bar{x}_1)(x_{2j} - \bar{x}_2)w_{ij}}{\sqrt{\frac{\sum_{i=1}^n (x_{1i} - \bar{x}_1)^2}{n}}}, \quad (3.32)$$

em que:  $n$  é o número de pontos amostrais georreferenciados;  $x_i$  ( $x_{1i}$ ) e  $y_{i,j}$  ( $x_{2j}$ ) são os valores dos atributos  $X$  ( $X_1$ ) e  $Y$  ( $X_2$ ) nas  $i, j$ -ésimas observações,  $i, j = 1, \dots, n$ ;  $\bar{x}$  ( $\bar{x}_1$ ) e  $\bar{y}$  ( $\bar{x}_2$ ) são os valores médios dos atributos  $X$  ( $X_1$ ) e  $Y$  ( $X_2$ ) na área em estudo;  $w_{ij}$  é um elemento da matriz de pesos espaciais.

Todavia, na proposta desenvolvida por Almeida (2013), apenas a variância de uma das variáveis é considerada, de modo que um atributo deve ser covariável do outro, pois trocar a ordem em que as variáveis são selecionadas influencia diretamente no valor obtido para o índice de Moran bivariado.

### 3.6.4 Índice de Lee

A proposta desenvolvida por Lee (2001) se assemelha a de Anselin, Syabri e Smirnov (2002) e utiliza a variância de ambas as variáveis. Logo, a ordem em que os atributos são tomados não afeta o valor da correlação espacial entre eles. A expressão do índice de Lee para o cálculo da correlação espacial bivariada é dada por:

$$L = \frac{n}{\sum_{i=1}^n (\sum_{j=1}^n w_{ij})^2} \frac{\sum_{i=1}^n [(\sum_{j=1}^n w_{ij}(x_i - \bar{x}))(\sum_{j=1}^n w_{ij}(y_j - \bar{y}))]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.33)$$

sendo a descrição dos elementos dessa expressão análoga à da Equação 3.31. Os valores de  $L$  variam entre -1 e 1.

Entretanto, comparada às propostas de Anselin, Syabri e Smirnov (2002) e Almeida (2013), a metodologia desenvolvida por Lee (2001) é pouco explorada na literatura. Realizando-se um mapeamento sistemático de literatura, nas principais bases de dados (*ScienceDirect*, *Scopus*, *Web of Science* e *Wiley*), ao longo de mais de uma década e excluindo as duplicidades, foi obtido um total de 229 trabalhos que utilizaram o índice de Moran bivariado, contra apenas 25 que empregaram o índice de Lee, sendo, portanto, dez vezes menos aplicado. Foi encontrado apenas um estudo com o índice de Lee no contexto agrícola (GASO et al., 2021), que utilizou essa métrica apenas para validar um modelo de previsão de produtividade da soja. Sendo assim, não foi um estudo exploratório sobre o índice de Lee.

Desse modo, o desenvolvimento desta tese pode contribuir para divulgação do índice de Lee no meio científico, especialmente no ambiente agrícola.

### 3.7 Amostragem espacial

Como já apontado brevemente em seções anteriores, a análise da variabilidade espacial de variáveis georreferenciadas é fundamental para o gerenciamento localizado do solo, realizado por técnicas de Agricultura de Precisão e isso requer a coleta de alguns pontos amostrais na área. Dada a inviabilidade de se analisar toda a população, a discretização da área agrícola deve agregar uma grade amostral representativa (ZHANG et al., 2018), pois a eficiência dos mapas de variabilidade espacial de atributos do solo e da planta, que serão utilizados como tomada de decisão no gerenciamento agrícola, dependem, especialmente, da qualidade da amostragem realizada e sua capacidade em representar a realidade de campo (SANCHES; MAGALHÃES; FRANCO, 2019).

Faz parte do planejamento amostral, deliberar sobre o tamanho, o formato e o relevo da região sob estudo (GAO et al., 2016). Além disso, deve-se estabelecer o número de pontos que serão coletados e a distribuição espacial destes na área (MIOT, 2011). O número de pontos e a configuração amostral dependem, principalmente, dos recursos financeiros e dos equipamentos disponíveis.

A literatura apresenta diversas configurações amostrais, dentre as quais estão as tradicionais, como: a aleatória simples (Figura 3.1a), a sistemática (Figura 3.1b), a estratificada (Figura 3.1c) e a por clusterização (Figura 3.1d) (SINGH; MASUKU, 2014).

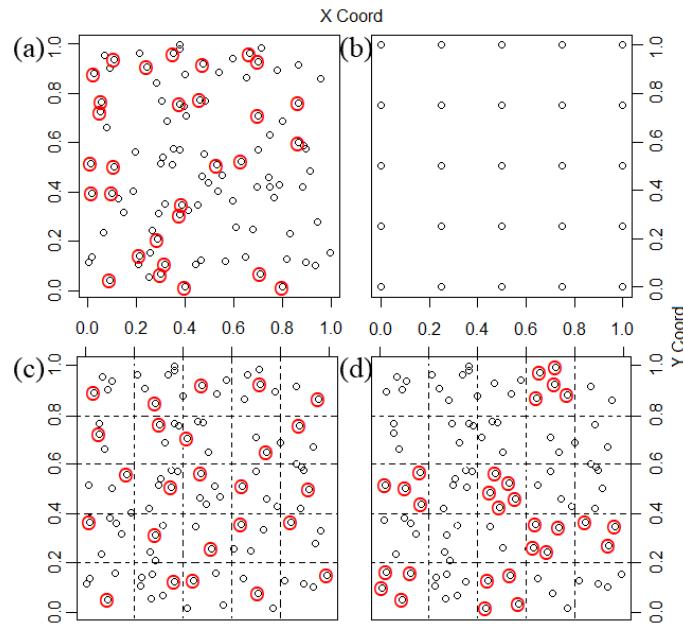


Figura 3.1 – Configurações amostrais tradicionais: (a) aleatória simples; (b) sistemática; (c) estratificada; (d) por *cluster*. As circunferências em vermelho indicam as observações selecionadas para amostragem. Nas configurações sem tais circunferências, todos os pontos representam observações amostrais.

Há também outras configurações, cuja base são as tradicionais, mas que foram aprimoradas e têm se mostrado mais vantajosas em termos da representatividade espacial demonstrada. Como exemplo, tem-se: a *lattice plus close pairs* (Figura 3.2a) e a *lattice plus in fill* (Figura 3.2b), que têm como base a amostragem aleatória (CHIPETA et al., 2017). Há, também, àquelas fundamentadas na estratificação da área, como a *multi-stage* (Figura 3.2c), a *multi-phase* (Figura 3.2d) e a *generalized random tessellation* (Figura 3.2e) (BENEDETTI; PIERSIMONI; POSTIGLIONE, 2015). As rotinas para confecção das malhas amostrais (c), (d) e (e), exibidas na Figura 3.2, estão disponibilizadas em Benedetti, Piersimoni e Postiglione (2015).

Em suma, na amostragem *lattice plus close pairs*, tem-se os pontos da grade regular, cuja distância é  $\delta$ , aos quais são adicionados aleatoriamente uma quantidade pré-fixada de pontos, dentro de uma circunferência de raio  $\alpha\delta$ , cujo centro é sempre um ponto da grade regular tomado ao acaso. É necessário que  $\alpha\delta < \Delta$ ,  $0 < \alpha < 0,5$ , para que o ponto acrescentado esteja mais próximo a um ponto da grade regular do que seu vizinho também pertencente à grade regular. A amostragem *lattice plus in fill* assemelha-se a *close pairs*, a diferença é que são adicionados à grade regular alguns esquemas amostrais regulares menores. Na *multi-stage* a área é estratificada e, posteriormente, toma-se aleatoriamente um determinado número de estratos, dentro dos quais todos os pontos são selecionados para amostragem. Na *multi-phase*, a área também é estratificada, mas considera-se todos os estratos, nos quais faz-se uma pré-seleção de uma quantidade  $p$  de pontos, de modo aleatório. E dentre estes  $p$  pontos, sorteiam-se  $t$  em cada estrato para serem amostrados. Finalmente, na *generalized random tessellation*, a estratificação é feita com base no polígono de Voronoi, sendo uma amostra selecionada aleatoriamente em cada polígono.

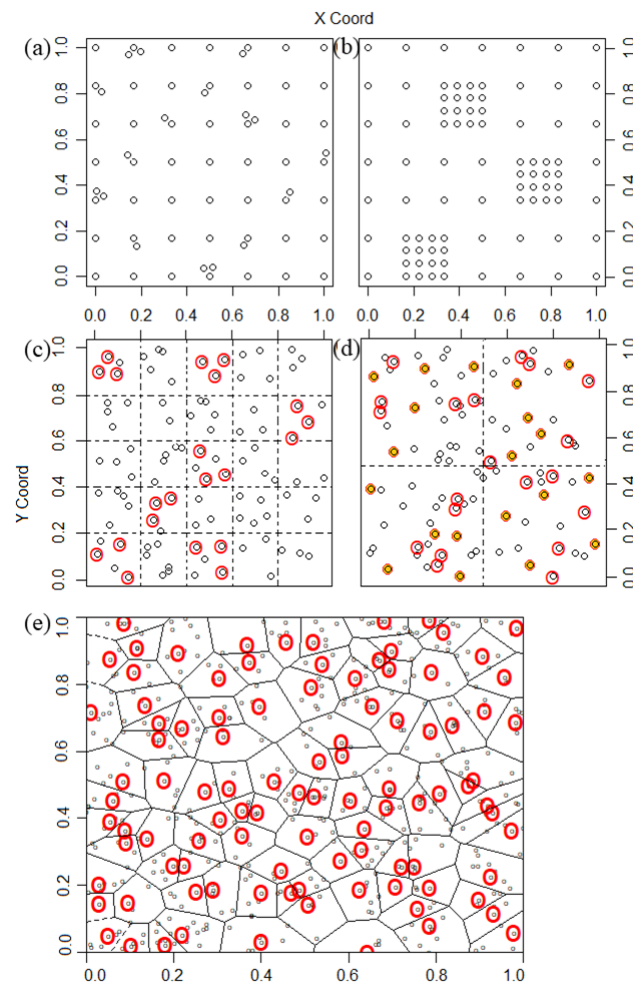


Figura 3.2 – Configurações amostrais aprimoradas: (a) *lattice plus close pairs*; (b) *lattice plus in fill*; (c) *multi-stage*; (d) *multi-phase*; (e) *generalized random tessellation*. As circunferências em vermelho indicam as observações selecionadas para amostragem, com exceção de (d), em que se faz uma pré-seleção (em vermelho) e uma seleção final (em amarelo). Nas configurações sem tais circunferências, todos os pontos representam observações amostrais.

Em relação ao tamanho amostral, vários trabalhos têm sido realizados no âmbito da densidade amostral (CAON; GENÚ, 2013; VAN BUSSEL et al., 2016). Para esses pesquisadores, o intuito é confrontar diferentes intensidades de pontos amostrais e analisar os efeitos da redução no número de observações aliada ao aumento na distância entre os pontos. Ainda sob o viés do redimensionamento amostral, pode-se citar os algoritmos de otimização, como os de inteligência artificial (TILLÉ, 2011; WADOUX et al., 2017) e também o uso do tamanho amostral efetivo (ESS) (ACOSTA; VALLEJOS, 2018; ACOSTA; VALLEJOS; GRIFFITH, 2018). Essas diferentes metodologias de redimensionamento amostral têm um objetivo comum: avaliar qual o menor tamanho amostral possível para a área, de modo que a qualidade das predições espaciais e, conseqüentemente, dos mapas seja minimamente afetada. A busca pela redução amostral é justificada pelos altos custos vinculados às análises laboratoriais das amostras.

## 3.8 Tamanho amostral efetivo

### 3.8.1 Tamanho amostral efetivo univariado

A proposta do *Effective Sample Size* - ESS (Tamanho amostral efetivo) considera que alguns pontos amostrais coletados estão altamente correlacionados entre si, proporcionando informações repetidas em relação à variabilidade espacial. Sendo assim, o ESS utiliza os efeitos da autocorrelação espacial entre as observações para estimar um novo tamanho amostral (GRIFFITH, 2005). Para calcular o valor do tamanho amostral efetivo, é necessário que já tenha sido realizado pelo menos um levantamento amostral na área desejada. Dessa forma, são conhecidos previamente o número de pontos amostrais e a disposição destes no campo, além da estrutura de dependência espacial da variável georreferenciada (VALLEJOS; OSORIO, 2014).

A expressão do tamanho amostral efetivo univariado tem como base a matriz de informação de Fisher sobre a média. Quando se considera um processo estocástico estacionário, em que o conjunto de observações do atributo segue uma distribuição normal, a expressão da matriz de informação de Fisher sobre a média tem a seguinte expressão (URIBE-OPAZO; BORSSOI; GALEA, 2012):

$$\mathbf{K}(\beta) = \mathbf{1}^T \Sigma^{-1} \mathbf{1}. \quad (3.34)$$

Embasado em 3.34 e considerando um vetor aleatório  $\mathbf{Y}$  observado nas  $n$  localizações amostrais, com distribuição normal de probabilidade e sem a presença de covariáveis, a proposta de Vallejos e Osorio (2014) para a estimação do tamanho amostral efetivo univariado é dada por:

$$ESS = \mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}, \quad (3.35)$$

em que  $\mathbf{1}$  é o vetor unitário  $n \times 1$ , sendo  $n$  o número de pontos amostrais da grade amostral original ( $n \geq 1$ ) e  $\mathbf{R}$  é a matriz de correlação espacial,  $n \times n$ , inversível, cuja expressão é dada por:

$$\mathbf{R}_{ml} = \begin{cases} 1, & \text{se } m = l \\ \frac{\hat{\sigma}^2 \rho(h)}{\hat{\sigma}^2 + \hat{\tau}^2}, & \text{se } m \neq l, \end{cases} \quad (3.36)$$

em que  $\mathbf{R}_{ml}$ ,  $m, l = 1, \dots, n$ , é o valor estimado da correlação espacial existente entre o  $m$ -ésimo e o  $l$ -ésimo ponto amostral,  $\rho(h)$  é a função de correlação, cujo cálculo depende do modelo geoestatístico e da distância euclidiana  $h$  entre as observações (CRESSIE, 2015) e  $\hat{\sigma}^2$  e  $\hat{\tau}^2$  são os valores estimados dos parâmetros contribuição e efeito pepita, respectivamente.

Sobre os valores estimados do tamanho amostral efetivo univariado, incidem as seguintes propriedades (VALLEJOS; OSORIO, 2014):

- (i)  $ESS$  diminui à medida que a correlação espacial aumenta;
- (ii)  $ESS = 1$  quando se tem uma perfeita correlação espacial entre todos os pares de observações;
- (iii)  $ESS = n$  quando não há correlação espacial entre os pares de observações, em que  $n$  é o número de observações georreferenciadas de uma variável;
- (iv)  $ESS$  cresce conforme  $n$  aumenta;

(v)  $ESS \geq 1$ .

Vários trabalhos foram desenvolvidos utilizando a abordagem univariada proposta por Vallejos e Osorio (2014). Acosta, Vallejos e Osorio (2016) quantificaram o ESS univariado para um modelo de regressão, incluindo correlações espaciais e seriais existentes. Acosta e Vallejos (2018) utilizaram o método para estimar o ESS univariado, considerando modelos espaciais de regressão. Canton et al. (2021) fizeram uso da distribuição *t-student* e do algoritmo EM para modelar diferentes profundidades de resistência do solo à penetração (RSP) com observações discrepantes e calcularam o ESS univariado para cada faixa de profundidade.

Feitas essas considerações em relação ao ESS univariado, será explorada uma metodologia apresentada na literatura para o tamanho amostral efetivo bivariado.

### 3.8.2 Tamanho amostral efetivo bivariado

No contexto agrícola, geralmente, são considerados vários atributos do solo no estudo da variabilidade espacial. Assim, considerando-se o gerenciamento regionalizado do solo, torna-se impraticável obter um redimensionamento amostral diferente para cada variável. O tamanho amostral efetivo bivariado possibilita estimar um novo número de pontos amostrais, a partir da correlação espacial entre observações amostrais de duas variáveis. Logo, o novo tamanho amostral é constituído apenas por observações espacialmente independentes, evitando que sejam coletadas informações duplicadas.

A proposta multivariada que será apresentada foi desenvolvida por Vallejos e Osorio (2014), com base no trabalho conduzido por Griffith (2005). Entretanto, com base na proposta desta pesquisa, será considerado apenas o caso bivariado. Considere  $\{X(\mathbf{s}_m); \mathbf{s}_m \in S \subset \mathbb{R}^2\}$  e  $\{Y(\mathbf{s}_m); \mathbf{s}_m \in S \subset \mathbb{R}^2\}$ ,  $m = 1, \dots, n$ , dois campos aleatórios representando as observações dos atributos de interesse nas  $\mathbf{s}_m$  localizações amostradas. A metodologia para estimação do tamanho amostral efetivo bivariado ( $p = 2$ ) é dada por:

$$ESS_p = \frac{\mathbf{1}_{(1)}^T ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \Psi_d^{1/2})^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \Psi_d^{1/2}) \mathbf{1}_{(1)}}{\mathbf{1}_{(2)}^T \mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d \mathbf{1}_{(2)}}, \quad (3.37)$$

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  $n \times 1$ ;  $\mathbf{1}_{(2)}$  é um vetor unitário  $p \times 1$ ;  $\Phi_d$ , de dimensão  $p \times p$  é uma matriz diagonal contendo o desvio padrão da  $k$ -ésima variável,  $\sigma_k$ ,  $k = 1, \dots, p$ ;  $\mathbf{I}$ , de dimensão  $n \times n$ , é a matriz identidade;  $\Psi_d$ , de dimensão  $np \times np$ , é uma matriz bloco diagonal contendo as matrizes  $\mathbf{R}_k$ , de ordem  $n \times n$ , de correlação espacial de cada  $k$ -ésima variável,  $k = 1, \dots, p$  (Equação 3.36).

Ao efetuar os cálculos matriciais de 3.37, considerando  $p = 2$ , tem-se que:

$$\mathbf{A}_d \Phi_d = \begin{pmatrix} \alpha & 0 \\ 0 & 1 - \alpha \end{pmatrix} \begin{pmatrix} \sigma_X & 0 \\ 0 & \sigma_Y \end{pmatrix} = \begin{pmatrix} \alpha \sigma_X & 0 \\ 0 & (1 - \alpha) \sigma_Y \end{pmatrix},$$

em que  $\alpha$  e  $(1 - \alpha)$ ,  $0 \leq \alpha \leq 1$ , são os coeficientes dos autovetores padronizados dos componentes principais, que atuam como pesos aplicados nos atributos  $X$  e  $Y$ , respectivamente.

Então,

$$\begin{aligned} (\mathbf{A}_d \Phi_d) \mathbf{R} (\Phi_d \mathbf{A}_d) &= \begin{pmatrix} \alpha \ddot{\sigma}_X & 0 \\ 0 & (1 - \alpha) \ddot{\sigma}_Y \end{pmatrix} \begin{pmatrix} 1 & \rho_{XY} \\ \rho_{XY} & 1 \end{pmatrix} \begin{pmatrix} \ddot{\sigma}_X \alpha & 0 \\ 0 & \ddot{\sigma}_Y (1 - \alpha) \end{pmatrix} \\ &= \begin{pmatrix} \alpha^2 \ddot{\sigma}_X^2 & \alpha(1 - \alpha) \ddot{\sigma}_Y \ddot{\sigma}_X \rho_{XY} \\ \alpha(1 - \alpha) \ddot{\sigma}_X \ddot{\sigma}_Y \rho_{XY} & (1 - \alpha)^2 \ddot{\sigma}_Y^2 \end{pmatrix}. \end{aligned}$$

Dessa forma,

$$\mathbf{1}_{(2)}^T \mathbf{A}_d \Phi_d \mathbf{R} \Phi_d \mathbf{A}_d \mathbf{1}_{(2)} = \alpha^2 \ddot{\sigma}_X^2 + (1 - \alpha)^2 \ddot{\sigma}_Y^2 + 2\alpha(1 - \alpha) \rho_{XY} \ddot{\sigma}_X \ddot{\sigma}_Y. \quad (3.38)$$

Sendo:

$$\Psi_d = \begin{pmatrix} \mathbf{R}_X^{-1/2} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_Y^{-1/2} \end{pmatrix}, \quad (3.39)$$

então,

$$\begin{aligned} ((\mathbf{A}_d \Phi_d) \otimes \mathbf{I}) \Psi_d^{1/2}{}^T &= \left[ \begin{pmatrix} \alpha \ddot{\sigma}_X & \mathbf{0} \\ \mathbf{0} & (1 - \alpha) \ddot{\sigma}_Y \end{pmatrix} \otimes \mathbf{I} \right] \begin{pmatrix} (\mathbf{R}_X^{-1/2})^T & \mathbf{0} \\ \mathbf{0} & (\mathbf{R}_Y^{-1/2})^T \end{pmatrix} \\ &= \begin{pmatrix} \alpha \ddot{\sigma}_X (\mathbf{R}_X^{-1/2})^T & \mathbf{0} \\ \mathbf{0} & (1 - \alpha) \ddot{\sigma}_Y (\mathbf{R}_Y^{-1/2})^T \end{pmatrix}, \end{aligned}$$

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

Tem-se que:

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

Assim,

$$\begin{aligned} &((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \Psi_d^{1/2}{}^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \Psi_d^{1/2}{}^T) = \\ &= \begin{pmatrix} \alpha \ddot{\sigma}_X (\mathbf{R}_X^{-1/2})^T & \mathbf{0} \\ \mathbf{0} & (1 - \alpha) \ddot{\sigma}_Y (\mathbf{R}_Y^{-1/2})^T \end{pmatrix} \begin{pmatrix} \mathbf{I} & \rho_{XY} \mathbf{I} \\ \rho_{XY} \mathbf{I} & \mathbf{I} \end{pmatrix} \\ &\quad \begin{pmatrix} \alpha \ddot{\sigma}_X (\mathbf{R}_X^{-1/2}) & \mathbf{0} \\ \mathbf{0} & (1 - \alpha) \ddot{\sigma}_Y (\mathbf{R}_Y^{-1/2}) \end{pmatrix} \\ &= \begin{pmatrix} \alpha^2 \ddot{\sigma}_X^2 (\mathbf{R}_X^{-1}) & \alpha(1 - \alpha) \ddot{\sigma}_X \ddot{\sigma}_Y \rho_{XY} (\mathbf{R}_X^{-1/2})^T (\mathbf{R}_Y^{-1/2}) \\ \alpha(1 - \alpha) \ddot{\sigma}_X \ddot{\sigma}_Y \rho_{XY} (\mathbf{R}_X^{-1/2})^T (\mathbf{R}_Y^{-1/2}) & (1 - \alpha)^2 \ddot{\sigma}_Y^2 (\mathbf{R}_Y^{-1}) \end{pmatrix}. \end{aligned}$$

Donde, tem-se que:

$$\begin{aligned} &\mathbf{1}_{(1)}^T ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \Psi_d^{1/2}{}^T (\mathbf{R} \otimes \mathbf{I}) ((\mathbf{A}_d \Phi_d \otimes \mathbf{I}) \Psi_d^{1/2}{}^T) \mathbf{1}_{(1)} = \\ &\alpha^2 \ddot{\sigma}_X^2 \mathbf{1}_{(1)}^T \mathbf{R}_X^{-1} \mathbf{1}_{(1)} + (1 - \alpha)^2 \ddot{\sigma}_Y^2 \mathbf{1}_{(1)}^T \mathbf{R}_Y^{-1} \mathbf{1}_{(1)} + 2\alpha(1 - \alpha) \rho_{XY} \ddot{\sigma}_X \ddot{\sigma}_Y \mathbf{1}_{(1)}^T \mathbf{R}_X^{-1/2} \mathbf{R}_Y^{-1/2} \mathbf{1}_{(1)}, \quad (3.40) \end{aligned}$$

sendo  $(\mathbf{R}_X^{-1/2})^T = \mathbf{R}_X^{-1/2}$ , devido à simetria da matriz  $\mathbf{R}$ .



Quando as expressões 3.38 e 3.40 são substituídas em 3.37, tem-se uma versão ponderada do tamanho amostral efetivo bivariado:

$$ESS_{XY} = \frac{\alpha^2 \ddot{\sigma}_X^2 \mathbf{1}_{(1)}^T \mathbf{R}_X^{-1} \mathbf{1}_{(1)} + (1 - \alpha)^2 \ddot{\sigma}_Y^2 \mathbf{1}_{(1)}^T \mathbf{R}_Y^{-1} \mathbf{1}_{(1)} + 2\alpha(1 - \alpha) \rho_{XY} \ddot{\sigma}_X \ddot{\sigma}_Y \mathbf{1}_{(1)}^T \mathbf{R}_X^{-1/2} \mathbf{R}_Y^{-1/2} \mathbf{1}_{(1)}}{\alpha^2 \ddot{\sigma}_X^2 + (1 - \alpha)^2 \ddot{\sigma}_Y^2 + 2\alpha(1 - \alpha) \rho_{XY} \ddot{\sigma}_X \ddot{\sigma}_Y}.$$

Dessa forma, Vallejos e Osorio (2014) e Griffith (2005), autores referência na metodologia do tamanho amostral efetivo em processos espaciais, sugerem um versão ponderada da expressão 3.35 para evitar o uso da estrutura de correlação espacial entre as variáveis. Esse fato é citado pelos próprios autores e pode ser observado na matriz de correlação espacial (Equação 3.39), a qual não considera a existência de correlação espacial entre as variáveis, sendo utilizada apenas uma correlação linear (Equação 3.38). Além disso, apesar de Vallejos e Osorio (2014) relatarem a associação entre a equação do tamanho amostral efetivo univariado e a matriz de informação de Fisher sobre a média, não remetem a esse vínculo no caso bivariado.

Em pesquisa conduzida por Canton, Guedes e Uribe-Opazo (2021), foi realizada uma redução amostral multivariada, utilizando nove atributos físico-químicos do solo em uma área agrícola com plantio de soja. Foi considerada nesse estudo a versão ponderada do ESS multivariado (Equação 3.37), logo não se têm uma estrutura de correlação espacial entre as variáveis.

De acordo com Griffith (2005), para se correlacionar espacialmente duas variáveis sem considerar ponderações na estimação do tamanho amostral efetivo bivariado, deve se analisar duas fontes de informação: a correlação entre os dois atributos e a associação espacial dentro de cada um. Com base nisso, Vallejos e Acosta (2021) propuseram um método multivariado para estimar o ESS, dividindo os atributos em grupos de dois e utilizando o *bivariate coregionalization model* - BCRM (modelo espacial bivariado de coregionalização) para calcular o ESS bivariado. Posteriormente, o ESS multivariado foi estimado fazendo a média dos valores do ESS bivariado calculados.

O BCRM propõe uma estrutura paramétrica em que se tem um parâmetro de correlação espacial comum a duas variáveis e outro parâmetro de correlação espacial exclusivo a uma das variáveis. Todavia, dessa forma, necessariamente a correlação espacial da outra variável é desconsiderada do modelo. Por outro lado, o *bivariate Gaussian common component model* – BGCCM (modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum) considera tanto a correlação espacial entre as duas variáveis quanto a associação espacial individual de ambas e não apenas de uma.

Sendo assim, considerando o modelo espacial BGCCM e tendo como base as propostas de Vallejos e Osorio (2014) e Vallejos e Acosta (2021), a expressão para a estimação do tamanho amostral efetivo bivariado ( $ESS_{bi}$ ) utilizada neste trabalho foi calculada da seguinte forma:

$$ESS_{bi} = \mathbf{1}^T \mathbf{R}_{(bi)}^{-1} \mathbf{1}, \quad (3.41)$$

em que:  $\mathbf{1}$  é o vetor unitário de dimensão  $2n \times 1$ , em que  $n$  é o número de observações somadas das duas variáveis na grade amostral original ( $n \geq 1$ );  $\mathbf{R}_{(bi)}$  é a matriz de correlação espacial bivariada,  $2n \times 2n$ , obtida a partir das Equações 3.12 e 3.13.

## Referências

- ACOSTA, J.; OSORIO, F.; VALLEJOS, R. Effective sample size for line transect sampling models with an application to marine macroalgae. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 21, n. 3, p. 407-425, 2016.
- ACOSTA, J.; VALLEJOS, R. Effective sample size for spatial regression models. **Electronic Journal of Statistics**, v. 12, p. 3147-3180, 2018.
- ACOSTA, J.; VALLEJOS, R.; GRIFFITH, D. On the effective geographic sample size. **Journal of Statistical Computation and Simulation**, v. 88, n. 10, p. 1958-1975, 2018.
- ADEGBOYE, O. A.; ADEKUNLE, A. I.; PAK, A.; GAYAWAN, E.; LEUNG, D. H.; ROJAS, D. P.; ELFAKI, F.; MCBRYDE, E. S.; EISEN, D. P. Change in outbreak epicentre and its impact on the importation risks of COVID-19 progression: A modelling study. **Travel Medicine and Infectious Disease**, v. 40, p. 101988, 2021.
- ALMEIDA, E. **Econometria Espacial Aplicada**. 2. ed. Campinas: Alinea, 2013. 498p.
- ANSELIN, L.; SYABRI, I.; SMIRNOV, O. Visualizing multivariate spatial correlation with dynamically linked windows. In: ANSELIN, L.; REY, S. (Eds.), **New tools for spatial data analysis: proceedings of the specialist meeting**. Santa Barbara: Center for Spatially Integrated Social Science (CSISS), 2002. CD-ROM.
- BENEDETTI, R.; PIERSIMONI, F.; POSTIGLIONE, P. **Sampling Spatial Units for Agricultural Surveys**. Berlin: Springer, 2015. 325p.
- BIVAND, R. S.; PEBESMA, E. J.; GÓMEZ-RUBIO, V. **Applied spatial data analysis with R**. 2. ed. New York: Springer, 2013. 405p.
- BIVAND, R. S.; WONG, D. W. Comparing implementations of global and local indicators of spatial association. **Test**, v. 27, p. 716-748, 2018.
- BOGNOLA, I. A.; RIBEIRO JR., P. J.; SILVA, E. A. A.; LINGNAU, C.; HIGA, A. R. Modelagem uni e bivariada da variabilidade espacial de rendimento de *Pinus taeda* L. **Floresta**, v. 38, n. 2, p. 373-385, 2008.
- CANTON, L. E. D.; GUEDES, L. P. C.; URIBE-OPAZO, M. A. Reduction of sample size in the soil physical-chemical attributes using the multivariate Effective Sample Size. **Journal of Agricultural Studies**, v. 9, n. 1, p. 357-376, 2021.
- CANTON, L. E. D.; GUEDES, L. P. C.; URIBE-OPAZO, M. A.; ASSUMPÇÃO, R. A. B.; MALTAURO, T.

C. Sampling redesign of soil penetration resistance in spatial t-Student models. **Spanish Journal of Agricultural Research**, v. 19, n. 1, p. e0202, 2021.

CANTÚ, J. G. **Influência de configurações amostrais na qualidade de estimação espacial sob o uso de modelos espaciais bivariados**. 2015. 95 f. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Centro de Ciências Exatas e Tecnológicas, Cascavel, 2015.

CAON, D.; GENÚ, A. M. Mapeamento de atributos químicos em diferentes densidades amostrais e influência na adubação e calagem. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 17, n. 6, p. 629-639, 2013.

CHIPETA, M.; TERLOUW, D.; PHIRI, K.; DIGGLE, P. Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. **Environmetrics**, v. 24, n. 1, p. e2425, 2017.

CIMA, E. G.; URIBE-OPAZO, M. A.; JOHANN, J. A.; ROCHA JR., W. F. D.; DALPOSSO, G. H. Analysis of spatial autocorrelation of grain production and agricultural storage in Paraná. **Engenharia Agrícola**, v. 38, p. 395-402, 2018.

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

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

FONSECA, B. H. F. **Um estudo sobre estimação e predição em modelos geoestatísticos bivariados**. 2008. 74 f. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo, 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.

GASO, D. V.; WIT, A.; BERGER, A. G.; KOOISTRA, L. Predicting within-field soybean yield variability by coupling Sentinel-2 leaf area index with a crop growth model. **Agricultural and Forest Meteorology**, v. 308-309, p. 108553, 2021.

GELFAND, A.E.; SCHMIDT, A.M.; BANERJEE S.; SIRMANS, C.F. Nonstationary multivariate process modeling through spatially varying coregionalization. **Test**, v. 13, n. 2, p. 263-312, 2004.

GIORGI, E.; SCHLÜTER, D. K., DIGGLE, P. J. Bivariate geostatistical modelling of the relationship between *Loa loa* prevalence and intensity of infection. **Environmetrics**, v. 29, n. 5-6, p. e2447, 2018.

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.

GRIFFITH, D. A. Establishing qualitative geographic sample size in the presence of spatial autocorrelation. **Annals of the Association of American Geographers**, v. 103, n. 5, p. 1107-1122, 2013.

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. 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.

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

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

LEE, S. I. Developing a bivariate spatial association measure: an integration of Pearson's  $r$  and Moran's  $I$ . **Journal of Geographical Systems**, v. 3, p. 369-385, 2001.

MALTAURO, T. C.; GUEDES, L. P. C.; URIBE-OPAZO, M. A. Reduction of sample size in the analysis of spatial variability of nonstationary soil chemical attributes. **Engenharia Agrícola**, v. 39, p.56-65, 2019.

MARTINS, A. B. T.; BONAT, W. H., RIBEIRO JR., P. J. Likelihood analysis for a class of spatial geostatistical compositional models. **Spatial Statistics**, v. 17, p. 121-130, 2016.

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

MIOT, H. A. Tamanho da amostra em estudos clínicos e experimentais. **Jornal Vascular Brasileiro**, v. 10, n. 4, p. 275-278, 2011.

MOLIN, J. P.; AMARAL, L. R.; COLAÇO, A. F. **Agricultura de Precisão**. São Paulo: Oficina de Textos, 2015. 224p.

QUEIROZ, E. R. S.; MEDRONHO, R. A. Spatial analysis of the incidence of Dengue, Zika and Chikungunya and socioeconomic determinants in the city of Rio de Janeiro, Brazil. **Epidemiology & Infection**, v. 149, p. e188, 2021.

RIBEIRO, C. J.; SANTOS, A. D.; LIMA, S. V.; SILVA, E. R.; RIBEIRO, B. V.; DUQUE, A. M.; PEIXOTO, M. V. S.; SANTOS, P. L.; OLIVEIRA, I. M.; LIPSCOMB, M. W.; ARAUJO, K. C. G. M.; MOURA, T. R. Space-time risk cluster of visceral leishmaniasis in Brazilian endemic region with high social vulnerability: an ecological time series study. **PLoS Neglected Tropical Diseases**, v. 15, n. 1, p. e0009006, 2021.

RIGHETTO, A. J. **Avaliação de modelos geoestatísticos multivariados**. 2012. 92 f. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo, 2012.

SANCHES, G. M.; MAGALHÃES, P. S. G.; FRANCO, H. C. J. Site-specific assessment of spatial and temporal variability of sugarcane yield related to soil attributes. **Geoderma**, v. 334, p. 90-98, 2019.

SINGH, A. S.; MASUKU, M. B. Sampling techniques e determination of sample size in applied statistics research: An overview. **International Journal of Economics, Commerce and Management**, v. 2, n. 11, p. 1-22, 2014.

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

TILLÉ, Y. **Sampling Algorithms**. Berlin: Springer, 2011. 216p.

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.; DE BASTIANI, F.; GALEA, M.; SCHEMMER, R. C.; ASSUMPCÃO, R. A. B. Influence diagnostics on a reparameterized t-Student spatial linear model. **Spatial Statistics**, v. 41, p. 100481, 2021.

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

VALLEJOS, R.; ACOSTA, J. 2021. The effective sample size for multivariate spatial processes with an application to soil contamination. **Natural Resource Modeling**, v. 34, n. 4, p. e12322, 2021.

VAN BUSSEL L. G.; EWERT F.; ZHAO G.; HOFFMANN H.; ENDERS A.; WALLACH D.; ASSENG S.; BAIGORRIA G. A.; BASSO B.; BIERNATH C.; CAMMARANO D.; CHRYSANTHACOPOULOSH J.; CONSTANTIN J.; ELLIOTT J.; GLOTTER M.; HEINLEIN F.; KERSEBAUM K. C.; KLEIN C.; NENDEL C.; PRIESACK E.; RAYNAL H.; ROMERO C. C.; RÖTTER R. P.; SPECKA X.; TAO F. Spatial sampling

of weather data for regional crop yield simulations. **Agricultural and Forest Meteorology**, v. 220, p. 101-115, 2016.

WACKERNAGEL, H. **Multivariate Geostatistics: An Introduction with Applications**. Heidelberg: Springer, 2013. 257p.

WADOUX, A. M. C.; BRUS, D. J.; RICO-RAMIREZ, M. A.; HEUVELINK, G. B. Sampling design optimisation for rainfall prediction using a non-stationary geostatistical model. **Advances in Water Resources**, v. 107, p. 126-138, 2017.

WIECZOREK, J.; BARAN, A.; URBANSKI, K.; MAZUREK, R.; KLIMOWICZ-PAWLAS, A. Assessment of the pollution and ecological risk of lead and cadmium in soils. **Environmental Geochemistry and Health**, v. 40, n. 6, p. 2325-2342, 2018.

ZHANG, Z.; SUN, Y.; YU, D.; MAO, P.; XU, L. Influence of sampling point discretization on the regional variability of soil organic carbon in the red soil region, China. **Sustainability**, v. 10, p. 3603, 2018.

## 4 Artigos

### 4.1 Artigo 1: Correlação espacial bivariada entre atributos do solo e produtividade da soja em uma área agrícola

#### Resumo<sup>1</sup>

Utilizar a correlação espacial bivariada possibilita identificar padrões e comportamentos que sugerem uma associação espacial entre dois atributos do solo, viabilizando assim um melhor gerenciamento do solo e um uso mais eficiente dos recursos. O objetivo principal deste trabalho foi analisar a correlação espacial bivariada, considerando variáveis com diferentes estruturas de dependência espacial, calculando o índice de Lee bivariado. Para modelar e descrever o padrão espacial de duas variáveis correlacionadas espacialmente, foi utilizado o modelo bivariado com componente de correlação parcialmente comum (BGCCM). Foram associadas as variáveis produtividade da soja e matéria orgânica com atributos químicos do solo, além de nove pares de variáveis simuladas, cujo intuito foi reproduzir um rol de possibilidades presentes nos dados agrícolas. Observou-se tanto nos resultados dos dados simulados quanto dos dados reais, que quanto maior o alcance prático comum, maior o valor do índice de Lee, indicando uma correlação espacial bivariada mais elevada. Além disso, quanto menor a distância entre os pares de pontos vizinhos, maior é o valor do índice de Lee. A produtividade da soja apresentou correlação espacial direta com a soma de bases, e com os teores de cálcio e magnésio. Enquanto a matéria orgânica teve correlação espacial direta com a soma de bases e inversa com o teor de fósforo. Estudos em solos agrícolas cuja capacidade de troca catiônica é baixa, é fundamental analisar a correlação espacial bivariada da produtividade da soja e matéria orgânica com os atributos químicos do solo. Assim, é viável identificar padrões e comportamentos que sugiram uma associação espacial entre tais pares de atributos, possibilitando tomadas de decisão na área agrícola.

**Palavras-chave:** BGCCM, correlograma, geoestatística, índice de Lee, semivariograma cruzado.

#### 4.1.1 Introdução

O desafio atual do setor agrícola é manter um ritmo crescente de aumento de produção, preferencialmente sem aumentar a área plantada, o que implica um aumento da produtividade. Uma das possibilidades para o incremento da produtividade agrícola é utilizar o avanço tecnológico para

<sup>1</sup> Artigo aceito na revista *Australian Journal of Crop Science*.

aperfeiçoar o conhecimento das exigências nutricionais de cada cultura, proporcionando o correto uso de insumos na propriedade agrícola (DEISS et al., 2020).

Estudos demonstram a importância de se investigar a existência de uma relação da quantidade de nutrientes do solo com a produtividade da soja, visando estabelecer um melhor gerenciamento da produção desta commodity agrícola (MALVEZI et al., 2019; DEUS et al., 2020). Isso porque o excesso ou a falta de macro e micronutrientes no solo pode alterar as fases de crescimento e desenvolvimento da planta, afetando o grão e, conseqüentemente, a produtividade da soja (MENGEL; KIRKBY, 2001; MALAVOLTA, 2006; BARBOSA et al., 2016). Em solos agrícolas cuja capacidade de troca catiônica (CTC) é baixa, a matéria orgânica tem um papel relevante, pois em quantidades adequadas melhora as condições físicas e químicas do solo, além de auxiliar na disponibilidade de nutrientes para as plantas, contribuindo, assim, com o aumento da fertilidade (CUNHA; MENDES; GIONGO 2015; SIQUEIRA-NETO et al., 2021).

Dessa forma, torna-se imprescindível avaliar os dados de produtividade da soja e da matéria orgânica em relação aos atributos químicos e físicos do solo, empregando para tal a estatística espacial de áreas, que considera simultaneamente as informações referentes ao valor e à localização geográfica das variáveis. Utilizando a correlação espacial bivariada, é possível identificar padrões e comportamentos que sugiram uma associação espacial entre dois atributos do solo, viabilizando as tomadas de decisão (CIMA et al., 2018).

Dentre os trabalhos existentes na literatura e que utilizam a correlação espacial bivariada, observa-se o uso de duas expressões principais (MATKAN et al., 2013; CIMA et al., 2018). Uma é descrita por Lee (2001) e Anselin, Syabri e Smirnov (2002), de forma bem similar, enquanto a outra é apresentada por Almeida (2013). Na proposta de Almeida (2013), apenas a variância de uma das variáveis é considerada no denominador da razão que expressa essa medida, de modo que um atributo deve ser covariável do outro. Dessa forma, trocar a ordem em que as variáveis são selecionadas influencia diretamente no valor da correlação espacial bivariada. A proposta obtida por Lee (2001) e Anselin, Syabri e Smirnov (2002) utiliza a variância de ambas as variáveis, de modo que a ordem em que os atributos são tomados não afeta o valor da correlação espacial entre eles.

Entretanto, comparado ambas as propostas, de Anselin, Syabri e Smirnov (2002) e Almeida (2013), as quais são denominadas de índice de Moran bivariado, a metodologia desenvolvida por Lee (2001) é pouco explorada na literatura. Realizando um mapeamento sistemático de literatura nas principais bases de dados (*ScienceDirect*, *Scopus*, *Web of Science* e *Wiley*), ao longo de mais de uma década e excluindo as duplicidades, foi obtido um total de 229 trabalhos científicos que utilizaram o índice de Moran bivariado, contra apenas 25 que empregaram o índice de Lee. Foi encontrado um único estudo utilizando o índice de Lee no contexto agrícola (GASO et al., 2021), o qual aplicou essa métrica apenas para validar um modelo de previsão de produtividade da soja. Não sendo, portanto, um estudo exploratório sobre o índice de Lee.

Sendo assim, foram estabelecidos como objetivos para este trabalho: a) analisar ensaios considerando pares de variáveis simuladas com diferentes valores para o alcance prático comum e explorar o índice de Lee para o cálculo da correlação espacial bivariada; b) calcular o índice de Lee explorando diferentes matrizes de ponderação espacial e métricas para o cálculo da distância entre os pares de pontos; c) calcular a correlação espacial bivariada utilizando o índice de Lee para diferentes pares de atributos: matéria orgânica (MO) e fósforo (P), MO e soma de bases (SB), produtividade da soja (Prod) e cálcio (Ca), Prod e magnésio (Mg) e Prod e SB. Para o ajuste dos



dados foram utilizados o modelo espacial bivariado com componente de correlação parcialmente comum (BGCCM) e o semivariograma cruzado, os quais tiveram seus resultados comparados.

#### 4.1.2 Material e métodos

Para o cálculo do índice de Lee foram considerados pares de variáveis correlacionadas espacialmente, seguindo o modelo bivariado BGCCM. Além disso, foram utilizadas duas matrizes de ponderação espacial (**W** e **C**) e duas métricas para o cálculo da distância entre as observações vizinhas (distância euclidiana e inverso da distância euclidiana).

#### Modelo espacial Gaussiano bivariado

Nos casos em que existem evidências estatísticas de correlação espacial entre dois atributos, pode-se modelar e descrever o padrão espacial dessas variáveis, considerando um modelo espacial Gaussiano bivariado (FONSECA, 2008). Foi empregado neste estudo o modelo bivariado com componente de correlação parcialmente comum - BGCCM (DIGGLE; RIBEIRO JR., 2007).

Na proposta do BGCCM, sugere-se que existem dois campos aleatórios Gaussianos que podem ser modelados como (FONSECA, 2008):

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}S_0 + \sigma_1S_1 \\ Y_2 = \mu_2 + \sigma_{02}S_0 + \sigma_2S_2 \end{cases}, \quad (4.1)$$

em que  $\mu_1, \mu_2$  são as médias das variáveis  $Y_1$  e  $Y_2$ , respectivamente;  $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$  é o vetor de parâmetros de dispersão vinculados ao modelo geoestatístico bivariado;  $S_0, S_1$  e  $S_2$  que são campos aleatórios Gaussianos mutuamente independentes. O campo aleatório  $S_0$  é comum às variáveis  $Y_1$  e  $Y_2$ , enquanto  $S_1$  e  $S_2$  estão associados a cada variável individualmente (RIGHETTO, 2012). Sendo assim, o modelo BGCCM exibe uma estrutura de covariância construída a partir de três funções de correlação válidas para  $S_0, S_1$  e  $S_2$ , as quais serão denotadas por  $\rho_0, \rho_1$  e  $\rho_2$ , respectivamente.

Suponha que  $Y_k(s_m)$  e  $Y_k(s_l)$  sejam observações do atributo  $Y_k$  mensuradas nas localizações  $s_m$  e  $s_l$ , as quais estão separadas por uma distância euclidiana  $h = h_{ml}$ , com  $m, l = 1, \dots, n_k$  e  $k = 1, 2$ . Desta forma, tem-se  $Y = (Y_1, Y_2)^T$ , em que o vetor  $Y$  possui uma distribuição gaussiana  $n$ -variada, com  $n = n_1 + n_2$ , sendo  $n_1$  e  $n_2$  os tamanhos amostrais de  $Y_1$  e  $Y_2$ , respectivamente. Neste trabalho, foi considerado que  $n_1 = n_2$ . Além disso,  $Y$  tem vetor de médias  $\mu = (\mu_1, \mu_2)^T$  e matriz de covariâncias positiva definida  $\Sigma_Y$ , dada por (FONSECA, 2008):

$$\Sigma_Y = \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_2 \end{pmatrix}, \quad (4.2)$$

em que  $\Sigma_k$  é a matriz de covariância da variável  $Y_k$ , com dimensão  $n_k \times n_k$ ,  $k = 1, 2$ ;  $\Sigma_{1,2}$ ,  $n_1 \times n_2$ , é a matriz com as covariâncias cruzadas entre as variáveis  $Y_1$  e  $Y_2$ ; os elementos da matriz de covariância  $\Sigma_Y$  são dados por  $C(h) = \sigma_{0k}^2\rho_0(h) + \sigma_k^2\rho_k(h)$ , em que  $C$  representa a

função de covariância em relação a distância  $h$ , e  $\rho_0, \rho_k$  são as funções de correlação em  $S_0$  e  $S_k$ , respectivamente, para  $k = 1, 2$ .

Logo, a distribuição de probabilidade do vetor  $\mathbf{Y}$  depende da estimação do vetor de parâmetros  $\boldsymbol{\theta}^* = (\boldsymbol{\mu}^*, \boldsymbol{\sigma}^*, \boldsymbol{\varphi}^*)^T$ , em que o vetor  $\boldsymbol{\mu}^* = (\mu_1, \mu_2)^T$  está associado à média, o vetor  $\boldsymbol{\sigma}^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$  é o vetor de dispersão, e o vetor  $\boldsymbol{\varphi}^* = (\varphi_0, \varphi_1, \varphi_2)^T$  está associado a função de alcance  $\varphi_t$ , vinculada ao modelo geoestatístico escolhido para a função de correlação espacial  $\rho_t$ , em que  $t = 0, 1, 2$  está relacionado ao campo aleatório  $S_t$ . Segundo Righetto (2012), a estimação dos parâmetros de  $\boldsymbol{\theta}^*$  seguem os mesmos critérios de técnicas geoestatísticas univariadas. Assim, neste trabalho, foi utilizado o método de máxima verossimilhança para estimação dos parâmetros (CRESSIE, 2015).

### Matriz de ponderação espacial e correlação espacial bivariada

Uma matriz de ponderação espacial é uma matriz quadrada, de dimensão  $n \times n$ , em que os pesos espaciais  $w_{ij}$  representam o grau de conexão entre as regiões segundo algum critério de proximidade, mostrando a influência da localização  $j$  sobre a localização  $i$ ,  $i, j = 1, \dots, n$ . Dessa forma, a matriz de pesos apresenta uma espécie de ponderação da influência que as localizações exercem entre si (ALMEIDA, 2013).

Para estabelecer o grau de conexão expresso nas matrizes de pesos espaciais foi considerada a distância geográfica, que por sua vez, depende de uma determinada métrica. Neste trabalho, foram utilizadas as matrizes de ponderação espacial  $\mathbf{W}$  e  $\mathbf{C}$ , ambas de dimensão  $n \times n$ , em que  $n$  é o número de observações, e como métricas a distância euclidiana (DE) e o inverso da distância euclidiana (IDE).

A matriz de peso  $\mathbf{W} = [(w_{ij})]$  é padronizada por linha, de modo que a soma dos pesos de cada linha seja igual a 1. Para a DE,  $w_{ij} = \frac{1}{v_i}$ , em que  $v_i$  é o número total de vizinhos na  $i$ -ésima linha, considerando as colunas  $j$  que possuem vizinhos. Para as localizações que não possuem vizinhos, é atribuído peso  $w_{ij} = 0$ . São considerados vizinhos as localizações  $i$  e  $j$  cuja distância é inferior a distância de corte. Neste trabalho, a distância de corte variou de 200 a 1750 metros para os dados simulados, e de 140 a 1750 metros para os dados reais. Para o IDE,  $w_{ij} = \frac{1}{\sum_{j=1}^n 1/d_j} \frac{1}{d_{ij}}$ , em que  $\sum_{j=1}^n 1/d_j$  representa a soma do inverso das distâncias entre os vizinhos na  $i$ -ésima linha considerando as  $n$  colunas, e  $\frac{1}{d_{ij}}$  é o inverso da distância do elemento pertencente a linha  $i$  e coluna  $j$ ,  $i \neq j = 1, \dots, n$  (BIVAND; WONG, 2018). Para  $i = j$  o peso é zero.

A matriz de peso  $\mathbf{C} = [(c_{ij})]$  é globalmente padronizada, fazendo com que a soma dos pesos seja  $n$ , e alterando o modo com que os pesos  $c_{ij}$  são obtidos. Para a DE,  $c_{ij} = \frac{1}{v}$ , em que  $v$  é o número total de vizinhos considerando os  $n$  pontos. Enquanto para o IDE,  $c_{ij} = \frac{1}{\sum_{i,j=1}^n 1/d_{ij}} \frac{1}{d_{ij}}$ , em que  $\sum_{i,j=1}^n 1/d_{ij}$  representa a soma do inverso das distâncias considerando todos os vizinhos (BIVAND; PEBESMA; GÓMEZ-RUBIO, 2013). Para as localizações que não apresentam vizinhos,  $c_{ij} = 0$ .

Utilizando uma matriz de ponderação espacial, é possível estabelecer a correlação espacial entre duas variáveis georreferenciadas em  $n$  pontos amostrais. A expressão para o cálculo da correlação espacial bivariada, desenvolvida por Lee (2001), é dada por:

$$L = \frac{n}{\sum_{i=1}^n (\sum_{j=1}^n w_{ij})^2} \frac{\sum_{i=1}^n [(\sum_{j=1}^n w_{ij}(x_i - \bar{x}))(\sum_{j=1}^n w_{ij}(y_j - \bar{y}))]}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (4.3)$$

em que  $n$  é o número de pontos amostrais georreferenciados;  $x_i$ ,  $y_i$  e  $y_j$  são os valores dos atributos  $X$  e  $Y$  nas  $i, j$ -ésimas observações,  $i, j = 1, \dots, n$ ;  $\bar{x}$  e  $\bar{y}$  são os valores médios dos atributos  $X$  e  $Y$  na área em estudo; e  $w_{ij}$  é um elemento da matriz de pesos espaciais.

### Descrição das simulações

O estudo de simulação teve como intuito reproduzir situações que poderiam ocorrer em diferentes áreas agrícolas experimentais com quaisquer pares de atributos, agregando assim um conhecimento prático-teórico acerca da correlação espacial bivariada entre atributos do solo com estrutura de dependência espacial.

Foram construídos três ensaios ( $E1$ ,  $E2$ ,  $E3$ ) considerando diferentes estruturas de dependência espacial (Figura 4.1), e tendo como base o modelo bivariado com componente de correlação parcialmente comum (BGCCM). A elaboração destes ensaios para o estudo de simulação ocorreu considerando dois objetivos. O primeiro objetivo consistiu em analisar a variação do índice de Lee, em relação ao aumento dos valores dos parâmetros de alcance simulado ( $\varphi$ ), utilizando para isso, os três ensaios que diferenciam entre si quanto ao valor do vetor de alcance.

O segundo objetivo consistiu em avaliar o desempenho do índice de Lee considerando diferentes matrizes de ponderação espacial, combinadas a métricas distintas para o cálculo das distâncias entre as observações. O ensaio  $E3$  foi descartado neste objetivo pois as variáveis tiveram raios superiores ao *cutoff* na etapa anterior (matriz  $\mathbf{W}$  com a DE). Devido à similaridade observada entre os ensaios  $E1$  e  $E2$ , na primeira etapa, quanto aos valores do índice de Lee, foram realizados testes para ambos os ensaios com as matrizes  $\mathbf{W}$  e  $\mathbf{C}$ . Entretanto, no ensaio  $E2$ , para a maioria das variáveis o raio de dependência espacial foi superior a 1000 m, extrapolando o ponto de corte (*cutoff*). Desse modo, foi considerado na segunda etapa do trabalho apenas o ensaio  $E1$ .

Os procedimentos realizados para avaliar cada um dos objetivos foram detalhados na Figura 4.1a.

Para cada ensaio, foram geradas nove simulações, com duas variáveis, contendo 102 pontos amostrais cada, utilizando um experimento de Monte Carlo, a partir da decomposição de Cholesky da matriz de covariância  $\Sigma_Y$  (GRESSIE, 2015). As demais características a respeito das simulações são apresentadas no esquema metodológico da Figura 1a.

Seguindo o esquema da Figura 4.1a, para cada um dos nove pares de variáveis simulados de cada ensaio, foi estimado o alcance prático considerando o modelo bivariado BGCCM, a partir da construção do semivariograma cruzado, no qual o alcance prático representa a distância máxima de dependência espacial entre duas variáveis (GRESSIE, 2015). Na sequência, o cálculo da correlação espacial bivariada, utilizando o índice de Lee, foi subdividido em duas etapas que correspondem aos dois objetivos citados anteriormente (Figura 4.1a).

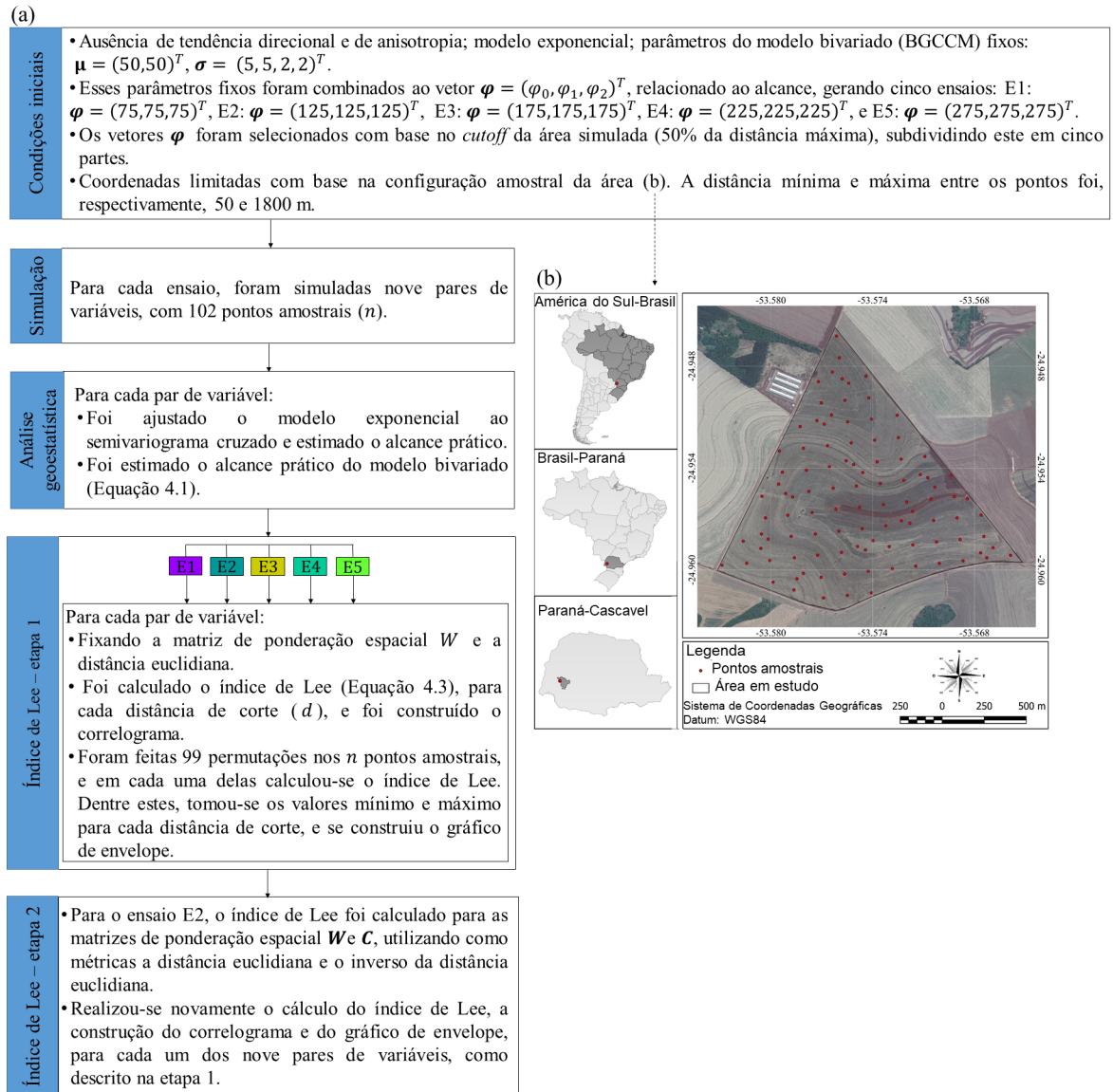


Figura 4.1 – (a) Esquema metodológico usado nos estudos de simulação; (b) Área experimental com a localização dos pontos amostrais usando coordenadas UTM.  $\mu$ : vetor de médias;  $\sigma$ : vetor de dispersão;  $\varphi$ : vetor de alcance; E1, E2, E3: ensaios 1 a 3;  $W$  e  $C$ : matrizes de ponderação espacial.

Em ambas as etapas, o correlograma foi elaborado para investigar o comportamento da correlação espacial bivariada de Lee. Para analisar a significância estatística dos valores que compõem o correlograma, foi utilizado o gráfico de envelope simulado, gerado a partir de permutações dos dados simulados nas coordenadas da grade amostral. Dessa forma, para obtenção do valor do índice de Lee, mantêm-se inalteradas as coordenadas geográficas e permutam-se os  $n$  valores do par de variável. Segundo Diggle e Ribeiro Junior (2007), o princípio da realização das permutações é tentar quebrar a estrutura de dependência espacial dos dados, gerando uma espécie de dados independentes.

Obtido o envelope de simulação a interpretação é a seguinte: o correlograma não é estatisticamente significativo se a linha correspondente aos valores do índice de Lee estiver totalmente contida entre os limites superior e inferior do envelope simulado. Caso contrário, se a linha estiver

acima do limite superior para alguma distância de corte  $d$ , o correlograma é considerado estatisticamente significativo em favor da dependência espacial positiva; se a linha estiver abaixo do limite inferior para alguma distância, o correlograma é estatisticamente significativo, porém indica uma dependência espacial negativa entre as variáveis (DIGGLE; RIBEIRO JR., 2007; COSTA; SCALON, 2015). Além disso, a distância na qual o correlograma adentra o envelope é o raio de dependência espacial do par de variáveis em questão (COSTA; SCALON, 2015). O raio de dependência espacial foi comparado aos alcances práticos obtidos, ajustando-se o semivariograma cruzado e utilizando o modelo bivariado.

### Descrição da área em estudo e dos dados reais

Os dados foram coletados no ano-safra 2016-2017, em área comercial de produção de grãos com 167,35 hectares, cultivada com soja, em que o plantio direto é realizado desde 1994 (Figura 4.1b). A área está localizada no município de Cascavel, região Oeste do Paraná, Brasil, com coordenadas geográficas aproximadas de 24,95° Sul para latitude, 53,37° Oeste para longitude, e 650 m de altitude média. O solo é classificado como Latossolo Vermelho Distroférico, o clima da região é temperado mesotérmico e superúmido, com tipo climático Cfa (Köppen).

Nessa área, foi utilizada uma amostragem *lattice plus close pairs* composta por 102 pontos amostrais, que compreende tanto uma grade regular cuja distância mínima entre os pontos foi de 141 m, como também 19 localidades que foram adicionadas aleatoriamente na grade regular e que apresentam distâncias menores, com algumas observações (de 50 e 75 m) (Figura 4.1b). Todos os dados coletados foram georreferenciados por um aparelho receptor de sinal com Sistema Global de Navegação por Satélite (GNSS), sob um sistema espacial de coordenadas Universal Transversa de Mercator (UTM).

Os atributos do solo utilizados nesse estudo apresentaram dependência espacial, foram isotrópicos e sem tendência direcional, sendo eles: teor de cálcio (Ca,  $\text{cmol}_c \text{ dm}^{-3}$ ); teor de fósforo (P,  $\text{mg dm}^{-3}$ ); teor de magnésio (Mg,  $\text{cmol}_c \text{ dm}^{-3}$ ); teor de matéria orgânica (MO,  $\text{g dm}^{-3}$ ); soma de bases (SB,  $\text{cmol}_c \text{ dm}^{-3}$ ), que representa a soma de cátions trocáveis ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  e  $\text{K}^+$ ) no solo; e produtividade da soja na área ( $\text{t ha}^{-1}$ ).

Foram considerados como pares de variáveis a produtividade da soja (Prod) com os atributos cálcio (Ca), magnésio (Mg) e soma de bases (SB), além da matéria orgânica (MO) com os atributos fósforo (P) e soma de bases (SB). Esses pares foram escolhidos por terem uma associação agronomicamente interessante, devido à contribuição no crescimento e desenvolvimento da planta e, conseqüentemente, dos grãos (MENGEL; KIRKBY, 2001; MALAVOLTA, 2006; DALCHIAVON et al., 2017).

Para o conjunto de dados reais, foram realizadas as mesmas análises de estatística espacial (estimação do modelo espacial BGCCM, semivariograma cruzado e cálculo do índice Lee), feitas nos dados simulados. Para o conjunto de dados reais, foram ajustados modelos geoestatísticos para cada par de variáveis (DIGGLE; RIBEIRO JR., 2007), para os quais a qualidade das estimativas obtidas foram avaliadas utilizando os critérios de validação cruzada (FARACO et al., 2008).

## Recursos computacionais

As rotinas para o cálculo do índice de Lee e para demais análises estatísticas e geostatísticas foram desenvolvidas utilizando o *software* R (R DEVELOPMENT CORE TEAM, 2022), considerando os pacotes *geoR* (RIBEIRO JR.; DIGGLE, 2001) e *spdep* (BIVAND, 2020).

### 4.1.3 Resultados

#### Dados simulados utilizando matriz de peso $W$ e função de distância euclidiana

O modelo BGCCM apresentou os maiores valores médios estimados para o alcance prático comum ( $a_0$ ), como também a menor dispersão (desvio padrão-DP) dos valores estimados desse parâmetro, quando se comparou com as estimativas obtidas para o semivariograma cruzado ( $a$ ) (Tabela 4.1).

Foi observado que, conforme o alcance prático comum simulado pelo modelo exponencial ( $a_0 = 3\varphi_0$ ) aumentou ( $E1$ : 375,  $E2$ : 525, e  $E3$ : 825 m), os valores médios do alcance prático estimados pelo semivariograma cruzado ( $a$ ) e pelo modelo BGCCM ( $a_0$ ) também aumentaram e foram muito próximos aos alcances práticos simulados (Tabela 4.1), indicando que ambas as estimativas para o raio de dependência espacial foram satisfatórias. Verificou-se ainda que, no último ensaio ( $E3$ ), os valores médios estimados para os alcances práticos do semivariograma cruzado e do modelo bivariado foram bem similares ao *cutoff* da área simulada (875 m), atendendo ao intuito deste ensaio.

Tabela 4.1 – Valores estimados para o alcance prático ( $a$ ) do semivariograma cruzado e para o alcance prático comum ( $a_0$ ) do BGCCM, considerando os três ensaios. Entre parênteses estão os valores simulados em cada ensaio.

	Semivariograma Cruzado			Modelo bivariado (BGCCM)		
	$E1$ (375 m)	$E2$ (525 m)	$E3$ (825 m)	$E1$ (375 m)	$E2$ (525 m)	$E3$ (825 m)
$S1$	378,75	577,32	856,83	391,32	533,55	851,70
$S2$	372,81	485,70	804,06	385,56	541,80	845,13
$S3$	332,28	571,98	816,96	378,15	532,38	812,82
$S4$	383,34	392,19	740,40	400,80	548,22	853,14
$S5$	413,04	537,27	812,79	389,10	528,93	845,34
$S6$	435,99	519,78	834,15	381,33	570,66	827,94
$S7$	434,31	522,27	905,10	378,36	535,80	842,13
$S8$	387,99	590,34	770,07	380,07	532,26	836,10
$S9$	334,86	400,86	755,52	383,97	531,78	846,24
$\bar{X}$	375,56	510,86	810,65	385,41	539,49	843,39
DP	60,26	72,62	51,62	7,37	13,11	7,69

$E1, E2, E3$ : ensaios 1 a 3;  $S1, \dots, S9$ : os nove pares de variáveis simuladas; DP: desvio padrão;  $\bar{X}$ : média. Os alcances práticos são dados em metros.

Para os nove pares de variáveis simuladas, considerando os cinco ensaios, os valores do

índice de Lee variaram entre 0,19 e 0,23 ( $E1$ ; Figura 4.2a), 0,26 e 0,33 ( $E2$ ; Figura 4.2b) e 0,37 e 0,46 ( $E3$ ; Figura 4.3a). Sendo assim, em todos os ensaios, constatou-se uma correlação espacial bivariada positiva em regiões vizinhas (LEE, 2001). Além disso, é possível notar que conforme os alcances simulados aumentaram, os valores nos quais o índice de Lee variou também aumentaram.

Para os ensaios  $E1$  e  $E2$ , o raio de dependência espacial variou entre 305 e 380 m (Figura 4.2a) e entre 505 e 780 m (Figura 4.2b), aproximando-se dos valores médios estimados para os alcances práticos do semivariograma cruzado (375,56 e 510,86 m, respectivamente; Tabela 4.1) e do modelo bivariado (385,41e 539,49 m, respectivamente; Tabela 4.1) nestes ensaios. Para o ensaio  $E3$ , o raio de dependência espacial foi a partir de 1105 m (Figura 4.3a), sendo, portanto, superior ao *cutoff* da área simulada. Dessa forma, houve uma superestimação nos raios de dependência espacial, observados nos correlogramas, em relação aos valores médios estimados para os alcances práticos do semivariograma cruzado e do modelo bivariado (Tabela 4.1).

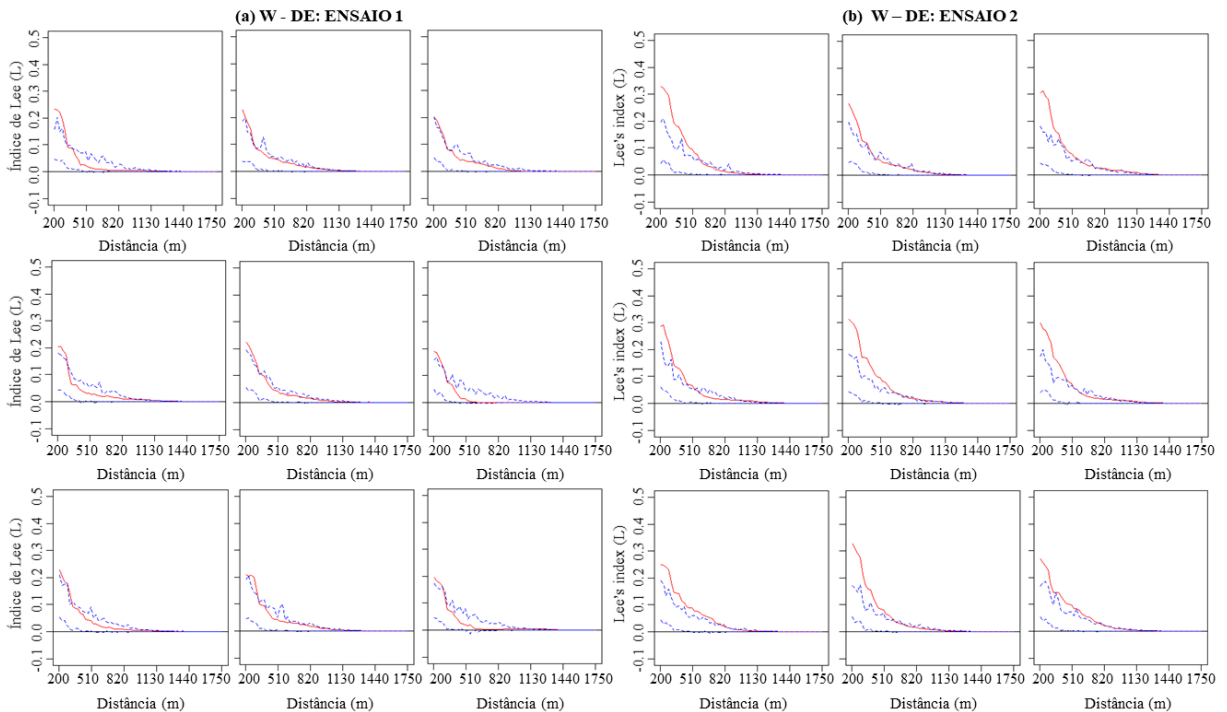


Figura 4.2 – Correlograma (vermelho) e gráfico de envelope (azul) para os nove pares de variáveis simuladas nos ensaios (a)  $E1$  e (b)  $E2$ , considerando a matriz de pesos  $\mathbf{W}$  e a distância euclidiana (DE).

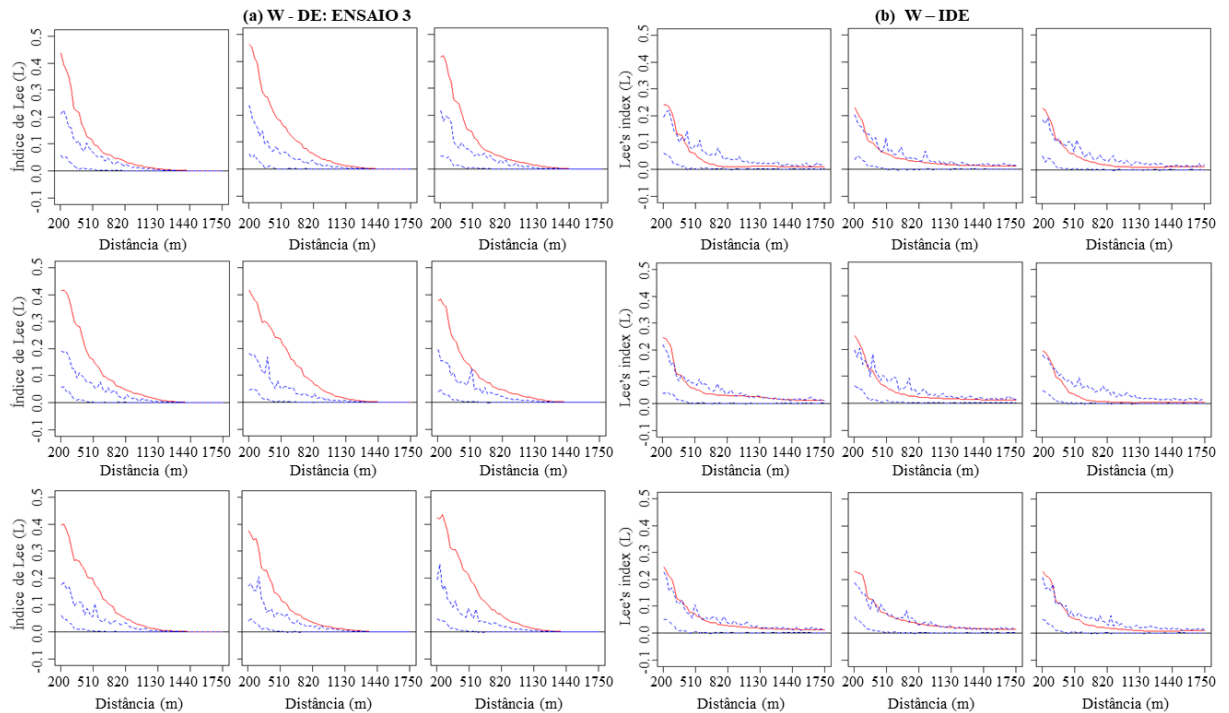


Figura 4.3 – Correlograma (vermelho) e gráfico de envelope (azul) para os nove pares de variáveis simuladas, considerando a matriz de pesos  $\mathbf{W}$  e utilizando: (a) a distância euclidiana (DE) no ensaio  $E3$  e (b) o inverso da distância euclidiana (IDE).

### Índice de Lee em dados simulados com diferentes matrizes e métricas de distância

Verificou-se que independente da matriz de ponderação espacial e da métrica utilizada, os correlogramas foram estatisticamente significativos para todos os pares de variáveis simuladas (Figuras 4.3b e 4.4). Foi constatada uma correlação espacial bivariada positiva em localizações vizinhas para todos os pares de variáveis.

Considerando a matriz de ponderação espacial  $\mathbf{W}$  e o inverso da distância euclidiana (IDE), os nove pares de variáveis simuladas apresentaram valores do índice de Lee variando entre 0,20 e 0,25. O raio de dependência espacial variou entre 330 e 505 m (Figura 4.3b). Comparando estes resultados com o ensaio  $E1$  (Tabela 4.1; Figura 4.2a), no qual as simulações foram realizadas apenas alterando a métrica (distância euclidiana - DE), foi verificado que os valores do raio de dependência espacial foram menores utilizando a DE (305 a 380 m), entretanto os valores do índice de Lee foram semelhantes (entre 0,19 e 0,23).



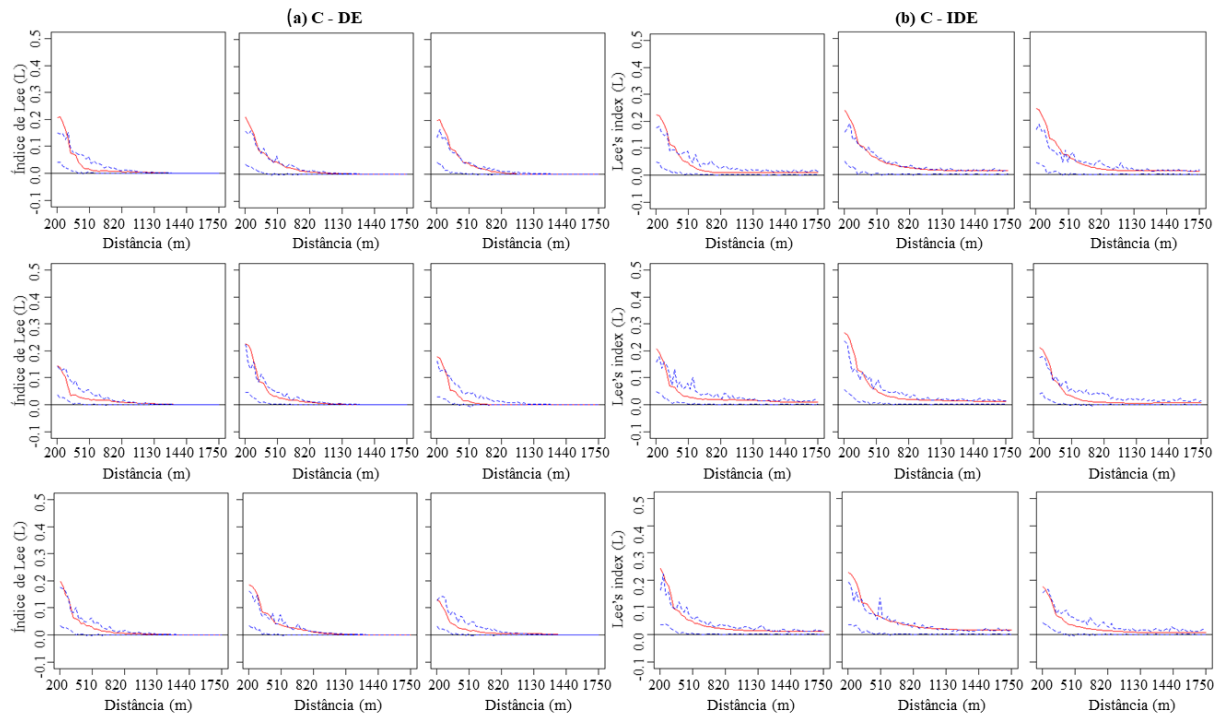


Figura 4.4 – Correlograma (vermelho) e gráfico de envelope (azul) para os nove pares de variáveis simuladas, considerando: (a) matriz de pesos  $\mathbf{C}$  e distância euclidiana (DE), e (b) matriz de pesos  $\mathbf{C}$  e o inverso da distância euclidiana (IDE).

Por outro lado, utilizando a matriz de pesos  $\mathbf{C}$  e a DE, os valores do índice de Lee variaram entre 0,13 e 0,23, com raio de dependência espacial entre 230 e 430 m (Figura 4.4a), contrastando este caso com o ensaio  $E1$ , no qual se alterou a matriz de pesos (para  $\mathbf{W}$ ), verificou-se uma semelhança no valor máximo obtido para o raio de dependência espacial (380 m) (Figuras 4.2a e 4.4a). Entretanto, a maioria dos valores máximos do índice Lee foram menores e com uma maior amplitude na variação de seus valores.

Por fim, mantendo-se a matriz de pesos ( $\mathbf{C}$ ) e alterando-se a métrica para o IDE, obtiveram-se valores do índice de Lee entre 0,18 e 0,27 e raio de dependência espacial entre 255 e 530 m (Figura 4.4b). Comparando as matrizes de pesos  $\mathbf{C}$  e  $\mathbf{W}$ , e mantendo o IDE como métrica, verificou-se que os valores máximos do índice de Lee foram similares (Figuras 4.3b e 4.4b). Todavia, utilizando a matriz de pesos  $\mathbf{C}$ , o raio de dependência espacial exibiu maior amplitude ( $\mathbf{W}$ : entre 330 e 505 m).

O tempo computacional foi um fator relevante. Isto porque, para a função de distância euclidiana, em ambas as matrizes de peso, o tempo computacional foi semelhante, entre 120 e 140 segundos por par de variáveis simuladas. Utilizando o inverso da distância euclidiana, embora as matrizes de peso  $\mathbf{C}$  e  $\mathbf{W}$  tenham sido similares, o tempo computacional foi cerca de sete vezes maior. Para realizar o cálculo do índice de Lee para as 63 classes de distâncias entre vizinhos por par de variáveis simuladas, considerando 99 permutações por classe, utilizou-se uma máquina com processador Intel®Core™i5-8265U, sistema operacional de 64 bits, e memória física (RAM) instalada de 8 GB.

## Aplicação da metodologia aos atributos do solo e produtividade da soja

Considerando os atributos químicos do solo para o Paraná, em média, os teores médios de cálcio (Ca; 4,03  $\text{cmol}_c \text{dm}^{-3}$ ), fósforo (P; 19,53  $\text{mg dm}^{-3}$ ), magnésio (Mg; 1,73  $\text{cmol}_c \text{dm}^{-3}$ ), matéria orgânica (MO; 24,93  $\text{g dm}^{-3}$ ) e soma de bases (SB; 6,05  $\text{cmol}_c \text{dm}^{-3}$ ) são considerados altos ou muito altos (SBCS-NEPAR, 2017). Enquanto a produtividade média da soja (Prod; 3,12  $\text{t ha}^{-1}$ ) no ano-safra 2016-2017 foi abaixo das médias estadual (3,72  $\text{t ha}^{-1}$ ) e nacional (3,36  $\text{t ha}^{-1}$ ) (CONAB, 2017). A análise exploratória evidenciou pelo coeficiente de variação (CV) que a dispersão dos atributos do solo, como também da produtividade, varia entre média ( $10\% \leq CV \leq 20\%$ ), alta ( $20\% < CV \leq 30\%$ ) e heterogênea ( $CV > 30\%$ ) (PIMENTEL GOMES; GARCIA, 2002) (Tabela 4.2).

Tabela 4.2 – Estatística descritiva dos atributos: cálcio (Ca,  $\text{cmol}_c \text{dm}^{-3}$ ), fósforo (P,  $\text{mg dm}^{-3}$ ), magnésio (Mg,  $\text{cmol}_c \text{dm}^{-3}$ ), matéria orgânica (MO,  $\text{g dm}^{-3}$ ), produtividade da soja (Prod,  $\text{t ha}^{-1}$ ) e soma de bases (SB,  $\text{cmol}_c \text{dm}^{-3}$ ).

Atributos	Mínimo	Máximo	Média	DP	CV
Ca	1,40	6,00	4,03	0,85	21,27
P	4,62	56,48	19,53	10,69	54,74
Mg	0,40	4,20	1,73	0,73	42,09
MO	13,40	89,80	42,14	10,51	24,93
Prod	1,51	4,20	3,12	0,54	17,33
SB	2,55	9,65	6,05	1,38	22,82

DP: desvio padrão;  $CV = 100 \frac{DP}{Média}$ : coeficiente de variação (%).

No semivariograma cruzado, as covariáveis da produtividade da soja (Ca, Mg e SB) foram melhores ajustadas pelos modelos Gaussiano e exponencial (Tabela 4.3). Enquanto as covariáveis da matéria orgânica (P e SB) tiveram seus dados ajustados pelo modelo Matérn com  $\kappa=2,5$  (Tabela 4.3). Todos os pares de atributos apresentaram forte dependência espacial (CAMBARDELLA et al., 1994). Pode-se ainda observar que o patamar ( $\varphi_1 + \varphi_2$ ) do par MO e P é negativo (Tabela 4.3). Como em um semivariograma cruzado o patamar se aproxima do valor da covariância entre as duas variáveis, isso indica que a correlação entre MO e P é inversa (RIGUETTO, 2012).

Tabela 4.3 – Valores estimados para os parâmetros do semivariograma cruzado, considerando o melhor modelo geoestatístico escolhido por validação cruzada, para os atributos: cálcio (Ca,  $\text{cmol}_c \text{dm}^{-3}$ ), fósforo (P,  $\text{mg dm}^{-3}$ ), magnésio (Mg,  $\text{cmol}_c \text{dm}^{-3}$ ), matéria orgânica (MO,  $\text{g dm}^{-3}$ ), produtividade da soja (Prod,  $\text{t ha}^{-1}$ ) e soma de bases (SB,  $\text{cmol}_c \text{dm}^{-3}$ ).

Atributos	Modelo geoestatístico	$\widehat{\varphi}_1$	$\widehat{\varphi}_2$	$\widehat{\varphi}_3$	$\widehat{a}$	$\widehat{EPR}$
MO e P	Matérn $\kappa=2,5$	-0,0041	-0,0337	178,43	1056,30	10,84
MO e SB	Matérn $\kappa=2,5$	0,1819	5,1839	61,19	362,24	3,38
Prod e Ca	Gaussiano	0,0125	0,1454	159,85	276,86	7,90
Prod e Mg	Exponencial	-0,0576	0,1857	161,62	484,86	23,67
Prod e SB	Exponencial	-0,1616	0,5268	106,55	319,65	23,47

$\widehat{\varphi}_1, \widehat{\varphi}_2, \widehat{\varphi}_3, \widehat{a}$ : são os valores estimados dos parâmetros efeito pepita, contribuição, função de alcance, e alcance prático (metros), respectivamente;  $\widehat{EPR} = 100 \frac{\widehat{\varphi}_1}{\widehat{\varphi}_1 + \widehat{\varphi}_2}$ : Efeito Pepita Relativo (%).

Para o modelo bivariado, as covariáveis da matéria orgânica (P e SB) foram melhores ajustadas, respectivamente, pelos modelos Gaussiano e exponencial (Tabela 4.4). Em relação à produtividade da soja, a maioria das covariáveis foram melhores ajustadas pelo modelo Matérn

com  $\kappa=2,5$ , segundo os critérios de validação cruzada (Tabela 4.4). Dentre os modelos ajustados, o Matérn com  $\kappa=2,5$  é o que possui a maior constante multiplicada à função de alcance, sendo aproximadamente  $5,92\varphi_t$ ,  $t = 0, 1, 2$  (DIGGLE; RIBEIRO JR., 2007). Consequentemente, os raios de dependência espacial foram altos (superiores a 1000 m; Tabela 4.4).

As funções de alcance, tanto do componente comum ( $\varphi_0$ ), quanto às vinculadas a cada variável ( $\varphi_1, \varphi_2$ ), foram semelhantes entre si nos pares de atributos, com exceção do  $\varphi_0$  para MO e SB (Tabela 4.4). Os parâmetros de dispersão ( $\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2$ ) foram baixos para a maioria dos pares de atributos analisados (Tabela 4.4).

Tabela 4.4 – Valores estimados para os parâmetros do modelo bivariado (BGCCM), considerando o melhor modelo geoestatístico escolhido por validação cruzada, para os atributos: cálcio (Ca,  $\text{cmol}_c \text{ dm}^{-3}$ ), fósforo (P,  $\text{mg dm}^{-3}$ ), magnésio (Mg,  $\text{cmol}_c \text{ dm}^{-3}$ ), matéria orgânica (MO,  $\text{g dm}^{-3}$ ), produtividade da soja (Prod,  $\text{t ha}^{-1}$ ) e soma de bases (SB,  $\text{cmol}_c \text{ dm}^{-3}$ ).

Atributos		MO-P	MO-SB	Prod-Ca	Prod-Mg	Prod-SB
Modelo geoestatístico		Gaussiano	Exponencial	Matérn $\kappa=2,5$	Gaussiano	Matérn $\kappa=2,5$
Parâmetros	$\widehat{\mu}_1$	45,81	43,95	-0,96	3,16	3,50
	$\widehat{\mu}_2$	17,52	6,14	5,95	1,94	6,40
	$\widehat{\sigma}_{01}$	$-1,30 \cdot 10^{-11}$	3,14	$1,54 \cdot 10^{-13}$	$2,12 \cdot 10^{-14}$	$-5,65 \cdot 10^{-14}$
	$\widehat{\sigma}_1$	$-1,90 \cdot 10^{-10}$	$1,91 \cdot 10^2$	$3,46 \cdot 10^{-11}$	$5,84 \cdot 10^{-12}$	$-3,36 \cdot 10^{-12}$
	$\widehat{\sigma}_{02}$	$-1,55 \cdot 10^{-13}$	1,87	$5,70 \cdot 10^{-14}$	$1,98 \cdot 10^{-12}$	$-3,62 \cdot 10^{-12}$
	$\widehat{\sigma}_2$	$-1,91 \cdot 10^{-10}$	$5,78 \cdot 10^{-7}$	$3,18 \cdot 10^{-12}$	$5,92 \cdot 10^{-12}$	$-2,20 \cdot 10^{-11}$
	$\widehat{\varphi}_0$	179,41	79,43	179,53	190,95	186,60
	$\widehat{a}_0$	310,53	237,97	1062,60	330,50	1104,45
	$\widehat{\varphi}_1$	270,46	269,28	267,82	259,49	275,55
	$\widehat{a}_1$	468,13	806,70	1585,16	449,13	1630,90
	$\widehat{\varphi}_2$	275,66	296,02	267,84	247,56	269,10
	$\widehat{a}_2$	477,12	886,81	1585,25	428,28	1592,71

$\widehat{\varphi}_1, \widehat{\varphi}_2, \widehat{\varphi}_3, \widehat{a}$ : são os valores estimados dos parâmetros efeito pepita, contribuição, função de alcance, e alcance prático (metros), respectivamente;  $\overline{EPR} = 100 \frac{\widehat{\varphi}_1}{\widehat{\varphi}_1 + \widehat{\varphi}_2}$ : Efeito Pepita Relativo (%).

Para todos os pares de atributos foi obtido um maior valor do índice de Lee, em módulo, na menor distância entre vizinhos utilizada (140 m) (Figura 4.5). Após essa distância, os pares de atributos que tiveram correlações espaciais bivariadas positivas, apresentaram uma queda abrupta para o índice de Lee, exibindo um comportamento decrescente para o correlograma, o qual vai se atenuando com o aumento da distância, até se estabilizar próximo a zero. O correlograma do par MO e P, apresentou um índice de Lee negativo, o que indica uma correlação espacial bivariada negativa (Figura 4.5a). Sendo assim, o comportamento do correlograma foi inverso, sendo crescente com o aumento das distâncias entre os vizinhos, até se estabilizar perto de zero.

Considerando o raio de dependência espacial identificado nos correlogramas (Figura 4.5), notou-se que os pares de atributos com maiores valores do índice de Lee, em módulo, exibiram os maiores raios. Sendo que os maiores valores de raios de dependência espacial e do índice Lee foram para os pares MO-SB (465 m; L: 0,36), Prod-SB (440 m; L: 0,35) e Prod-Mg (415 m; 0,34). Já os pares Prod-Ca e MO-P, tiveram o mesmo raio de dependência espacial (315 m) e apresentaram valores menores para o índice de Lee, sendo respectivamente iguais a 0,26 e -0,27.

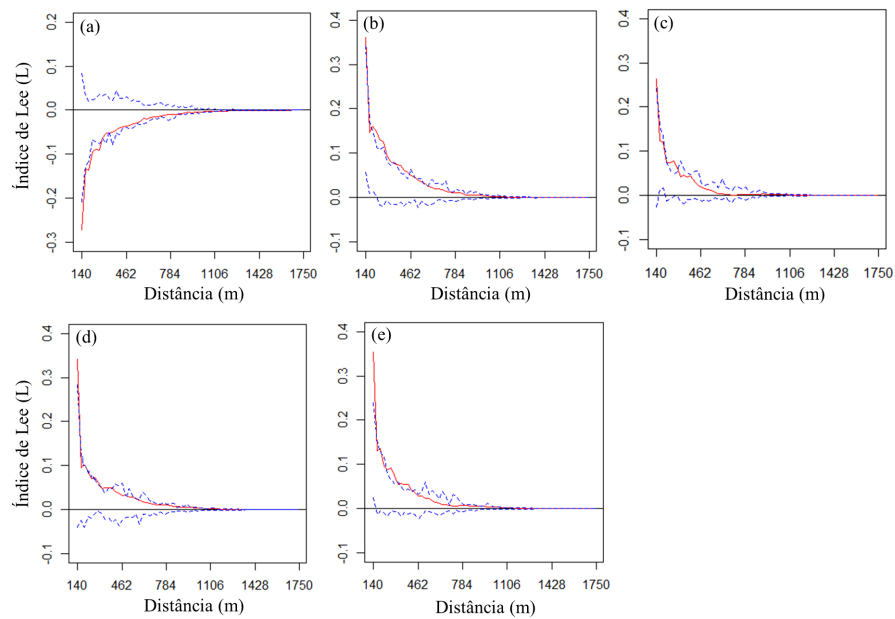


Figura 4.5 – Correlograma (vermelho) e gráfico de envelope (azul) para os pares de atributos (a) MO-P, (b) MO-SB, (c) Prod-Ca, (d) Prod-Mg e (e) Prod-SB, considerando a matriz de pesos  $W$  e a distância euclidiana.

#### 4.1.4 Discussão

##### Dados simulados

Em todos os ensaios realizados com a matriz de ponderação espacial  $W$  e a função de distância euclidiana, o índice de Lee apresentou um comportamento decrescente em função da distância entre os vizinhos, aproximando-se de zero conforme o aumento da distância (Figuras 4.2a, 4.2b e 4.3a). Assim, os maiores valores para o índice de Lee foram obtidos com as menores distâncias entre os vizinhos. O mesmo resultado foi observado nos trabalhos de Liu, Xie e Xia (2013) e Costa e Scalco (2015), nos quais se utilizou o índice de Moran univariado para analisar a autocorrelação espacial de diferentes atributos.

O decréscimo foi mais acentuado na primeira classe de distâncias (entre 200 e 510 m), sendo atenuado na segunda classe (entre 510 e 820 m), a partir da qual começa a se estabilizar e tender a zero. Em uma análise de correlograma, o decréscimo ao longo do aumento das distâncias entre os vizinhos, até a estabilização de sua curva, indica a estacionariedade do processo estocástico (SOUZA et al., 2007). Além disso, foi verificado que o aumento no alcance prático simulado, implicou decréscimos mais lentos do correlograma. Isso pode ser observado, comparando-se, principalmente, os ensaios  $E1$  e  $E2$  (Figuras 4.2a e 4.2b) com o ensaio  $E3$  (Figura 4.3a). O maior valor do índice de Lee (entre 0,37 e 0,46) e o maior raio de dependência espacial (superior a 1105 m) foi obtido no ensaio  $E3$ .

Conforme houve um aumento no alcance prático simulado, houve também uma maior diferença entre os valores do raio de dependência espacial no correlograma (Figuras 4.2a, 4.2b e 4.3a) com os valores médios do raio de dependência espacial estimado pelo semivariograma cruzado (Tabela 4.1). Essa diferença foi menor para os ensaios  $E1$  e  $E2$ , e maior para o ensaio  $E3$ . Isso pode ser explicado devido à diferença entre as metodologias do correlograma e do semivariograma cruzado (LIU et al., 2013).

Comparando as métricas do inverso da distância euclidiana e da distância euclidiana, mantendo-se a matriz de pesos  $\mathbf{C}$  fixa, foram obtidos maiores raios de dependência espacial utilizando a métrica do inverso da distância euclidiana (Figura 4.4). Além disso, embora em ambos os casos a diferença entre o menor e o maior valor máximo do índice de Lee tenha sido em torno de 0,10, os valores da correlação espacial bivariada foram maiores quando se utilizou o inverso da distância euclidiana (Figura 4.4). As conclusões a respeito do raio de dependência espacial e dos valores da correlação espacial bivariada, considerando o inverso da distância euclidiana, também foram obtidas ao fixar a matriz de pesos  $\mathbf{W}$ .

No geral, o comportamento do correlograma considerando o inverso da distância foi distinto da outra métrica, em ambas as matrizes de pesos utilizadas, pois nas maiores distâncias o valor do índice Lee não se aproximou de zero.

### **Aplicação da metodologia aos atributos do solo e produtividade da soja**

Como os valores dos raios de dependência espacial ( $a$  e  $a_0$ ) foram influenciados pela escolha do modelo, as comparações foram realizadas em relação as funções de alcance, contrastando as obtidas no semivariograma cruzado ( $\varphi_3$ ), com aquelas estimadas pelo modelo bivariado ( $\varphi_0$ ). Para todos os pares de atributos (Prod-Ca, Prod-Mg, Prod-SB, MO-P, MO-SB), os valores da função de alcance foram maiores para o modelo BGCCM (Tabelas 4.3 e 4.4). Esse fato também foi observado nos estudos de simulação realizados neste trabalho e por Cantú (2015), considerando outros pares de atributos do solo analisados na mesma área experimental nos anos-safra 2010-2011 e 2013-2014.

Sobre o comportamento do correlograma ser decrescente e tendendo a zero com o aumento da distância entre vizinhos, isso também foi observado nos ensaios simulados neste estudo (Figuras 4.2a, 4.2b e 4.3a). Entretanto, diferentemente das simulações, os pares formados pela produtividade da soja e matéria orgânica com suas covariáveis, apresentaram correlogramas que se mantiveram sobrepostos ao limite superior do gráfico de envelope até adentrar completamente nele na distância identificada como raio de dependência espacial.

Os valores do índice de Lee mostraram uma correlação espacial bivariada negativa entre MO e P, e positiva para os demais pares de atributos (Figura 4.5), corroborando com o que indicou os valores estimados do patamar no semivariograma cruzado, o qual foi negativo apenas para o par MO e P (Tabela 4.3). A correlação espacial bivariada entre todos os pares indicou que quanto mais próximos são os vizinhos (menor a distância), mais forte foi a correlação espacial bivariada (positiva ou negativa) (DALCHIAVON et al., 2017).

A correlação espacial positiva entre MO e SB é esperada, pois, em solos tropicais, o teor de MO é fundamental para elevar a capacidade de troca catiônica (CTC) do solo. Logo, solos com maior

CTC tendem a apresentar maiores valores de SB, ou seja, maior retenção de cátions na camada arável do solo (RAMOS et al., 2018). SB representa a soma de cátions trocáveis ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  e  $\text{K}^{+}$ ) no solo, portanto, a correlação espacial positiva entre Prod e SB é também refletida nas correlações positivas entre Prod e teores de Ca e Mg.

A correlação espacial inversa entre MO e P pode ser explicado devido ao fato que em locais com maior quantidade de MO, menor é o teor de P adsorvido na fração argilosa do latossolo (FINK et al., 2016). Sendo assim, em regiões com maior quantidade de MO, houve maior disponibilidade de P na solução do solo. Como P é um elemento de baixa mobilidade e que raramente é perdido por lixiviação (MAGGI et al., 2011), possivelmente foi perdido por erosão ou absorvido em maior quantidade pelas plantas, resultando em vizinhanças com um teor menor de P no solo.

#### 4.1.5 Conclusão

Ao comparar as diferentes matrizes de ponderação espacial na análise de dados simulados, verificou-se similaridade entre os valores do índice de Lee. Por outro lado, as métricas do inverso da distância euclidiana e da distância euclidiana apresentaram resultados distintos, tendo sido obtidos maiores do raio de dependência espacial e da correlação espacial bivariada utilizando o inverso da distância euclidiana, que também demandou maior tempo computacional. Logo, a métrica para calcular a distância entre os pares de pontos se mostrou mais relevante na estimação do raio de dependência espacial e do valor do índice de Lee do que a matriz de pesos.

Conforme o alcance prático comum simulado aumentou, os valores médios do alcance prático estimados pelo semivariograma cruzado e pelo modelo BGCCM também aumentaram e foram satisfatórios, pois se aproximaram dos alcances práticos simulados. Além disso, o valor da correlação espacial bivariada, calculado pelo índice de Lee, foi maior, de acordo com o aumento do alcance prático simulado. A mesma relação entre os valores do índice de Lee e do raio de dependência espacial foi constatada para os pares de atributos analisados nos dados reais, de modo que os estudos práticos corroboraram os simulados.

A produtividade da soja apresentou correlação espacial positiva com a soma de bases e com os teores de cálcio e magnésio, indicando que, nas regiões com maior produtividade da soja, houve maior disponibilidade desses atributos. Enquanto a matéria orgânica teve correlação espacial positiva com a soma de bases e negativa com o fósforo, devido, respectivamente, à CTC do solo da região e à absorção pelas plantas ou erosão.

## Referências

- ALMEIDA, E. **Econometria Espacial Aplicada**. 2. ed. Campinas: Alinea, 2013. 498p.
- ANSELIN, L.; SYABRI, I.; SMIRNOV, O. Visualizing multivariate spatial correlation with dynamically linked windows. In: ANSELIN, L.; REY, S. (Eds.), **New tools for spatial data analysis: proceedings of the specialist meeting**. Santa Barbara: Center for Spatially Integrated Social Science (CSISS), 2002. CD-ROM.
- BARBOSA, J. M.; REZENDE, C. F. A.; LEANDRO, W. M.; RATKE, R. F.; FLORES, R. A.; SILVA, A. R. Effects of micronutrients application on soybean yield. **Australian Journal of Crop Science**, v. 10, n. 8, p. 1092-1097, 2016.
- BIVAND, R. S.; PEBESMA, E. J.; GÓMEZ-RUBIO, V. **Applied spatial data analysis with R**. 2. ed. New York: Springer, 2013. 405p.
- BIVAND, R. S.; WONG, D. W. Comparing implementations of global and local indicators of spatial association. **Test**, v. 27, p. 716-748, 2018. Acesso em: 01 jul. 2021.
- BIVAND, R. S. **spdep**: Spatial Dependence: Weighting Schemes, Statistics and Models. R package version 1.1.5, 2020. <http://CRAN.R-project.org/package=spdep>.
- 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.
- CANTÚ, J. G. **Influência de configurações amostrais na qualidade de estimação espacial sob o uso de modelos espaciais bivariados**. 2015. 112 f. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel, 2015.
- CIMA, E. G.; URIBE-OPAZO, M. A.; JOHANN, J. A.; ROCHA JR., W. F. D.; DALPOSSO, G. H. Analysis of spatial autocorrelation of grain production and agricultural storage in Paraná. **Engenharia Agrícola**, v. 38, p. 395-402, 2018.
- COMPANHIA NACIONAL DE ABASTECIMENTO - CONAB. **Acompanhamento da Safra Brasileira de Grãos**. CONAB, Brasília, 2017. Disponível em: <https://www.conab.gov.br/info-agro/safras/graos/boletim-da-safra-de-graos>. Acesso em: 02 Ago. 2017.
- COSTA, F. M.; SCALON, J. D. Análise da estrutura de dependência espacial em dados de áreas. **Brazilian Journal of Biometrics**, v. 33, n. 2, p. 197-215, 2015.
- CRESSIE, N. A. C. **Statistics for Spatial Data**. rev. ed. New York: John Wiley & Sons, 2015. 928p.

CUNHA, T. J. F.; MENDES, M. A. S.; GIONGO, V. Matéria orgânica do solo. In: NUNES, R. R.; REZENDE, M. O. O. (eds), **Recurso solo: propriedades e usos**. Cubo: São Carlos, 2015. p. 273-293.

DALCHIAVON, F. C.; RODRIGUES, A. R.; LIMA, E. S.; LOVERA, L. H.; MONTANARI, R. Variabilidade espacial de atributos químicos do solo cultivado com soja sob plantio direto. **Revista de Ciências Agroveterinárias**, v. 16, n. 2, p. 144-154, 2017.

DEISS, L.; KLEINA, G. B.; MORAES, A.; FRANZLUEBBERS, A. J.; MOTTA, A. C.; DIECKOW, J.; SANDINI, I. E.; ANGHINONI, I.; CARVALHO, P. C. Soil chemical properties under no-tillage as affected by agricultural trophic complexity. **European Journal of Soil Science**, v. 71, n. 6, p. 1090-1105, 2020.

DEUS, A. C. F.; BÜLL, L. T.; GUPPY, C. N.; SANTOS, S. D. M. C.; MOREIRA, L. L. Q. Effects of lime and steel slag application on soil fertility and soybean yield under a no till-system. **Soil and Tillage Research**, v. 196, p. 104422, 2020.

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.

FINK, J. R.; INDA, A. V.; TIECHER, T.; BARRON, V. Iron oxides and organic matter on soil phosphorus availability. **Ciência e Agrotecnologia**, v. 40, n. 4, p. 369-379, 2016.

FONSECA, B. H. F. **Um estudo sobre estimação e predição em modelos geoestatísticos bivariados**. 2008. 74 f. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo, 2008.

GASO, D. V.; WIT, A.; BERGER, A. G.; KOOISTRA, L. Predicting within-field soybean yield variability by coupling Sentinel-2 leaf area index with a crop growth model. **Agricultural and Forest Meteorology**, v. 308-309, p. 108553, 2021.

LEE, S. I. Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I. **Journal of Geographical Systems**, v. 3, p. 369-385, 2001.

LIU, Q.; XIE, W. J.; XIA, J. B. Using semivariogram and Moran's I techniques to evaluate spatial distribution of soil micronutrients. **Communications in Soil Science and Plant Analysis**, v. 44, n. 7, p. 1182-1192, 2013.

MAGGI, C. F.; FREITAS, P. S.; SAMPAIO, S. C.; DIETER, J. Lixiviação de nutrientes em solo cultivado



com aplicação de água residuária de suinocultura. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 2, p. 170-177, 2011.

MALAVOLTA, E. **Manual de nutrição de plantas**. São Paulo: Ceres, 2006. 638p.

MALVEZI, K. E. D.; JÚNIOR, L. A. Z.; GUIMARAES, E. C.; VIEIRA, S. R.; PEREIRA, N. Soil chemical attributes variability under tillage and no-tillage in a long-term experiment in southern Brazil. **Bioscience Journal**, v. 35, n. 2, p. 467-476, 2019.

MATKAN, A. A.; SHAHRI, M.; MIRZAIE, M. Bivariate Moran's I and LISA to explore the crash risky locations in urban areas. **N-Aerus**, v. 14, p. 1-12, 2013.

MENGEL, K.; KIRKBY, E. **Principles of plant nutrition**. 5th ed. Boston: Kluwer Academic Publishers, 2001. 849p.

PIMENTEL GOMES, F.; GARCIA, C. H. **Estatística Aplicada a Experimentos Agronômicos e Florestais**. Piracicaba: FEALQ, 2002. 309p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R foundation for Statistical Computing, Vienna, Austria, 2022.

RAMOS, F. T.; DORES, E. F. C.; WEBER, O. L. S.; BEBER, D. C.; CAMPELO JR., J. H.; MAIA, J. C. S. Soil organic matter doubles the cation exchange capacity of tropical soil under no-till farming in Brazil. **Journal of the Science of Food and Agriculture**, v. 98, n. 9, p. 3595-3602, 2018.

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

RIGHETTO, A. J. **Avaliação de modelos geoestatísticos multivariados**. 2012. 92 f. Dissertação (Mestrado em Estatística e Experimentação Agronômica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo, 2012.

SIQUEIRA-NETO, M.; POPIN, G. V.; PICCOLO, M. C.; CORBEELS, M.; SCOPEL, E.; CAMARGO, P. B.; BERNOUX, M. Impacts of land use and cropland management on soil organic matter and greenhouse gas emissions in the Brazilian Cerrado. **European Journal of Soil Science**, v. 72, n. 3, p. 1431-1446, 2021.

SOCIEDADE BRASILEIRA DE CIÊNCIA DO SOLO - SBCS. NÚCLEO ESTADUAL PARANÁ - NEPAR. **Manual de adubação e calagem para o estado do Paraná**. SBCS-NEPAR, Curitiba, 2017. Disponível em: <https://sbcs-nepar.org.br/produto/manual-de-adubacao-e-calagem-para-o-estado-do-parana-2a-edicao/>. Acesso em: 09 Set. 2017.

SOUZA, W. V.; CARVALHO, M. S.; CRUZ, O. G.; RAGONI, V. Análise Espacial de Dados de Áreas.

In: SANTOS, S. M.; SOUZA, W. V. (eds), **Introdução à estatística espacial para a saúde pública**.  
Ministério da Saúde: Brasília, 2007.

## 4.2 Artigo 2: Tamanho Amostral Efetivo utilizando o modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum

### Resumo<sup>2</sup>

Sabe-se que conforme aumenta a autocorrelação espacial entre os dados georreferenciados, maior é a quantidade de informações duplicadas contidas nesses dados. Partindo dessa premissa, o *Effective Sample Size* – ESS (Tamanho amostral efetivo) utiliza os efeitos da autocorrelação espacial existente entre as observações georreferenciadas para calcular quantos desses pontos amostrais são independentes. O número de observações independentes representa então o novo tamanho amostral. O objetivo desse artigo foi desenvolver uma metodologia bivariada para o ESS ( $ESS_{bi}$ ), considerando o *Bivariate Gaussian Common Component Model* – BGCCM (Modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum), que considera tanto a correlação espacial entre as duas variáveis quanto a associação espacial individual. Para verificar a viabilidade teórica da proposta do  $ESS_{bi}$ , todas as propriedades que incidem sobre a metodologia univariada foram verificadas para o  $ESS_{bi}$ , utilizando estudos de simulação ou de forma algébrica. Os cenários de simulação desenvolvidos possibilitaram ainda verificar uma influência do parâmetro de alcance comum do BGCCM nos valores estimados do  $ESS_{bi}$ . O  $ESS_{bi}$  foi aplicado a um conjunto de dados reais de matéria orgânica (MO) e soma de bases (SB), coletados no solo de uma área agrícola com plantio da soja. Foi verificado que 60% das observações amostrais do par MO-SB continham informações espacialmente duplicadas. A configuração amostral reduzida se mostrou eficiente, pois ao comparar os mapas original e reduzido de MO, tendo SB como covariável, observou-se que, visualmente, o tamanho amostral reduzido manteve as regiões pertencentes a cada uma das cinco classes. Isso foi corroborado pelo índice de concordância Tau, que evidenciou moderada acurácia entre os mapas.

**Palavras-chave:** BGCCM, geoestatística, redimensionamento amostral, simulação.

#### 4.2.1 Introdução

Estabelecer um planejamento amostral é um dos primeiros passos dentro de um estudo espacial. Isso inclui analisar a área em que serão coletadas as amostras (clima, vegetação, relevo), determinar a configuração amostral, o espaçamento entre as observações e o número de observações que serão coletadas (BENEDETTI; PIERSIMONI; POSTIGLIONE, 2015).

Suponha que  $n$  seja o tamanho amostral que atualmente é coletado em uma área agrícola, por exemplo, e que estas amostras sejam utilizadas para construir e analisar mapas temáticos de variáveis espacialmente correlacionadas. Um problema que vários pesquisadores veem discutindo e

<sup>2</sup> Artigo submetido à revista *Spatial Statistics*.

buscando soluções é: qual seria o número equivalente de observações independentes contidas em  $n$ ? (GRIFFITH, 2005, 2008; GRIFFITH; PAELINCK, 2011; VALLEJOS; OSORIO, 2014; VALLEJOS et al., 2020; VALLEJOS; ACOSTA, 2021). Esse número é denominado de tamanho amostral efetivo (ESS). Logo, o ESS representa a estimativa de um novo tamanho amostral, considerando os efeitos da autocorrelação espacial entre as observações georreferenciadas para reduzir o número de amostras coletadas. Em termos práticos, na agricultura, utiliza-se a informação sobre a estrutura de dependência espacial do atributo do solo, coletado em uma área agrícola, para calcular o ESS, determinando ao produtor rural quantas amostras devem ser coletadas e enviadas ao laboratório para análise química ou física.

A definição formal do ESS para variáveis georreferenciadas foi publicada por Griffith (2005), cuja estimação do tamanho amostral reduzido deriva do fator de inflação da variância, com aplicações desse método encontradas em Griffith (2008) e Watson (2021). No trabalho de Vallejos e Osorio (2014) foi proposto o cálculo do ESS univariado considerando a matriz de informação de Fisher. Nesse mesmo artigo, o caso univariado foi estendido para o bivariado, que era uma versão ponderada do primeiro para evitar uma estrutura de correlação espacial entre as variáveis. A partir da premissa do ESS univariado proposto em Vallejos e Osorio (2014), outros trabalhos foram publicados: Acosta, Osorio e Vallejos (2016) quantificaram o ESS univariado para um modelo de regressão, incluindo correlações espaciais e seriais existentes. Acosta e Vallejos (2018) utilizaram o método para estimar o ESS univariado, considerando modelos espaciais de regressão. Canton et al. (2021) fizeram uso da distribuição *t-student* e do algoritmo EM para modelar diferentes profundidades de resistência do solo à penetração das raízes com observações discrepantes e calcularam o ESS univariado para cada faixa de profundidade. O valor máximo do ESS foi considerado no referido trabalho para obtenção de uma única redução amostral. Canton, Guedes e Uribe-Opazo (2021) fizeram uma redução amostral utilizando o ESS multivariado, entretanto o método considerou apenas uma versão ponderada do ESS, para evitar o uso de uma correlação espacial entre as variáveis. Vallejos e Acosta (2021) propuseram um método multivariado para estimar o ESS, dividindo os atributos em grupos de dois e utilizando o *bivariate coregionalization model* – BCRM (modelo espacial bivariado de coregionalização). O BCRM propõe uma estrutura paramétrica em que se tem um parâmetro de correlação espacial comum a duas variáveis e outro parâmetro de correlação espacial exclusivo a uma das variáveis. Dessa forma, necessariamente a correlação espacial da outra variável é desconsiderada do modelo.

O intuito deste trabalho foi preencher essa lacuna e utilizar o modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum (BGCCM) para estimar o ESS bivariado. O BGCCM considera tanto a correlação espacial entre as duas variáveis quanto a associação espacial individual de ambas e não apenas de uma.

O trabalho se dividiu em duas etapas. Primeiramente, foi utilizado um estudo de simulação que foi subdividido em cenários distintos entre si pelos parâmetros do BGCCM, considerados para (a) verificar se as propriedades do ESS univariado também são válidas para o ESS bivariado ( $ESS_{bi}$ ) e (b) analisar a influência dos parâmetros do BGCCM no comportamento do  $ESS_{bi}$ . Na sequência, foi apresentada uma aplicação no contexto agrícola utilizando duas variáveis: matéria orgânica (MO) e soma de bases (SB), cuja relação espacial é agronomicamente importante (MENGEL; KIRKBY, 2001). A escolha de MO e SB não foi ao acaso, pois a MO aumenta a solubilidade e a disponibilidade de cátions (SB) benéficos para as plantas, tornando o solo mais fértil (CUNHA; MENDES; GIONGO,

2015). Os dados foram coletados em uma região produtora da soja, com o intuito de (a) analisar se existe informação duplicada entre as observações coletadas dessas variáveis nessa área e, em caso afirmativo, redimensionar o tamanho amostral e (b) verificar se a metodologia proposta mantém a acurácia na identificação da variabilidade espacial de MO tendo SB como covariável.

Este artigo foi estruturado da seguinte forma: na Seção 2, foi descrito o modelo BGCCM e seus parâmetros, o tamanho amostral efetivo univariado (ESS) e o método do ESS bivariado utilizado. Ainda nessa seção, foram descritos os estudos de simulação e a aplicação em dados reais. Na Seção 3, foram verificadas as propriedades que incidem sobre o ESS bivariado, além de aplicar a metodologia a um par de atributos químicos do solo de uma área agrícola. Na Seção 4, foram discutidos os resultados apresentados na seção anterior. As considerações finais do trabalho foram feitas na Seção 5.

#### 4.2.2 Material e métodos

##### Modelo espacial Gaussiano bivariado

Nos casos em que existem evidências estatísticas de correlação espacial entre dois atributos, pode-se modelar e descrever o padrão espacial destas variáveis considerando um modelo espacial Gaussiano bivariado (FONSECA, 2008). Dentre as propostas apresentadas na literatura para estudar situações em que dois atributos têm correlação espacial, foi empregado neste estudo o modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum - BGCCM (DIGGLE; RIBEIRO JR., 2007).

Na proposta do BGCCM, sugere-se que existem dois campos aleatórios Gaussianos que podem ser modelados como (FONSECA, 2008):

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}S_0 + \sigma_1S_1 \\ Y_2 = \mu_2 + \sigma_{02}S_0 + \sigma_2S_2 \end{cases}, \quad (4.4)$$

em que:  $\mu_1, \mu_2$  são as médias das variáveis  $Y_1$  e  $Y_2$ , respectivamente;  $\sigma = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$  é o vetor de parâmetros de dispersão vinculados ao modelo geoestatístico bivariado;  $S_0, S_1$  e  $S_2$  são campos aleatórios Gaussianos mutuamente independentes. O campo aleatório  $S_0$  é comum às variáveis  $Y_1$  e  $Y_2$ , enquanto  $S_1$  e  $S_2$  estão associados a cada variável individualmente (RIGUETTO, 2012). Sendo assim, o modelo BGCCM exibe uma estrutura de covariância construída a partir de três funções de correlação válidas para  $S_0, S_1$  e  $S_2$ , denotadas por  $\rho_0, \rho_1$  e  $\rho_2$ , respectivamente.

Suponha que  $Y_k(s_m)$  e  $Y_k(s_l)$  sejam observações do atributo  $Y_k$  mensuradas nas localizações  $s_m$  e  $s_l$ , as quais estão separadas por uma distância euclidiana  $h = h_{ml}$ , com  $m, l = 1, \dots, n_k$  e  $k = 1, 2$ , com  $n_1 = n_2$ . Desta forma, tem-se  $\mathbf{Y} = (Y_1, Y_2)^T$ , em que o vetor  $\mathbf{Y}$  possui uma distribuição gaussiana  $n$ -variada, com  $n = n_1 + n_2$ , sendo  $n_1$  e  $n_2$  os tamanhos amostrais de  $Y_1$  e  $Y_2$ , respectivamente. Além disso,  $\mathbf{Y}$  tem vetor de médias  $\boldsymbol{\mu} = (\mu_1, \mu_2)^T$  e matriz de covariâncias positiva definida  $\Sigma_{\mathbf{Y}}$ , dada por (FONSECA, 2008):

$$\Sigma_Y = \begin{pmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2} & \Sigma_2 \end{pmatrix},$$

em que  $\Sigma_k$  é a matriz de covariância da variável  $Y_k$ ,  $n_k \times n_k$ ,  $k = 1, 2$ , dada por:

$$\Sigma_k = \begin{pmatrix} \sigma_{0k}^2 + \sigma_k^2 & \sigma_{0k}^2 \rho_0(h) + \sigma_k^2 \rho_k(h) \\ \sigma_{0k}^2 \rho_0(h) + \sigma_k^2 \rho_k(h) & \sigma_{0k}^2 + \sigma_k^2 \end{pmatrix}, \quad (4.5)$$

e  $\Sigma_{1,2}$ ,  $n_1 \times n_2$ , é a matriz de covariância cruzada entre as variáveis  $Y_1$  e  $Y_2$ , dada por:

$$\Sigma_{1,2} = \begin{pmatrix} \sigma_{01}\sigma_{02} & \sigma_{01}\sigma_{02}\rho_0(h) \\ \sigma_{01}\sigma_{02}\rho_0(h) & \sigma_{01}\sigma_{02} \end{pmatrix}, \quad (4.6)$$

sendo  $\rho_0$  and  $\rho_k$  as funções de correlação, e  $\sigma_{0k}$  e  $\sigma_k$  os parâmetros de dispersão em  $S_0$  e  $S_k$ , para  $k = 1, 2$ .

Dessa forma, a distribuição de probabilidade do vetor  $Y$  depende da estimação do vetor de parâmetros  $\theta = (\mu, \sigma, \varphi)^T$ , em que o vetor  $\mu = (\mu_1, \mu_2)^T$  está associado à média, o vetor  $\sigma = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$  é o vetor de dispersão e o vetor  $\varphi = (\varphi_0, \varphi_1, \varphi_2)^T$  representa a função de alcance  $\varphi_t$ , vinculada ao modelo geoestatístico escolhido para a função de correlação espacial  $\rho_t$ , em que  $t = 0, 1, 2$  está relacionado ao campo aleatório  $S_t$ . Segundo Righetto (2012), a estimação dos parâmetros de  $\theta$  seguem os mesmos critérios de técnicas geoestatísticas univariadas. Assim, foi utilizado neste trabalho o método de máxima verossimilhança para a estimação dos parâmetros (CRESSIE, 2015).

### Tamanho amostral efetivo univariado e bivariado

A proposta do tamanho amostral efetivo univariado (ESS) considera que alguns pontos amostrais coletados estão altamente correlacionados entre si, proporcionando informações duplicadas em relação à variabilidade espacial. Sendo assim, o ESS utiliza os efeitos da autocorrelação espacial entre as observações para estimar um novo tamanho amostral (GRIFFITH, 2005).

O ESS tem como base a matriz de informação de Fisher sobre a média e considera um processo estocástico estacionário cujo conjunto de observações do atributo  $Y_1$  segue uma distribuição normal. A expressão para estimar o ESS é dada por (VALLEJOS; OSORIO, 2014):

$$ESS = \mathbf{1}^T \mathbf{R}^{-1} \mathbf{1}, \quad (4.7)$$

sendo:  $\mathbf{1}$  é o vetor unitário,  $n_1 \times 1$ , em que  $n_1$  é o número de observações de  $Y_1$ ;  $\mathbf{R}$  é a matriz de correlação espacial, inversível,  $n_1 \times n_1$ .

Sobre os valores estimados do ESS incidem as seguintes propriedades (VALLEJOS; OSORIO, 2014):

- i.  $ESS$  diminui à medida que a correlação espacial aumenta;
- ii.  $ESS = 1$  quando há uma perfeita correlação espacial entre todos os pares de observações;
- iii.  $ESS = n_1$  quando não há correlação espacial entre os pares de observações, sendo  $n_1$  o número de observações georreferenciadas da variável;

- iv.  $ESS$  cresce conforme  $n_1$  aumenta;
- v.  $ESS \geq 1$ .

Na intenção de correlacionar espacialmente duas variáveis, é necessário analisar duas fontes de informação repetidas: a correlação espacial entre os dois atributos e a associação espacial dentro de cada um (Griffith, 2005). Sendo assim, a expressão para a estimação do tamanho amostral efetivo bivariado ( $ESS_{bi}$ ), utilizada neste trabalho, contém, além da estrutura de associação espacial individual dos atributos, a estrutura de correlação espacial entre eles. Sendo assim, as variáveis foram modeladas e tiveram seu comportamento espacial descrito pelo modelo BGCCM (Equação 4.4)

Logo, tendo como base as propostas de Vallejos e Osorio (2014) e Vallejos, Osorio e Bevilacqua (2021) para o ESS univariado (Equação 4.7) e bivariado com o modelo BCRM, e considerando os parâmetros estimados pelo BGCCM, o valor do  $ESS_{bi}$  foi calculado pela seguinte expressão:

$$ESS_{bi} = \mathbf{1}^T \mathbf{R}_{(bi)}^{-1} \mathbf{1}, \quad (4.8)$$

sendo que  $\mathbf{1}$  é o vetor unitário,  $2n \times 1$ , em que  $n$  é o total de observações somados os dois atributos e  $\mathbf{R}_{(bi)}$  é a matriz de correlação espacial bivariada,  $2n \times 2n$ , obtida a partir das Equações 4.5 e 4.6.

### Descrição do estudo de simulação

As quatro primeiras propriedades descritas para o ESS univariado foram validadas para o caso bivariado utilizando estudos de simulação. Apenas a propriedade (v) foi verificada algebricamente. Além disso, as simulações também agregam conhecimento prático-teórico sobre o tamanho amostral efetivo bivariado, reproduzindo situações que poderiam ocorrer em diferentes áreas agrícolas experimentais com quaisquer pares de atributos.

Para verificar a propriedade (i), foi elaborado o cenário 1, para o qual foram construídos cinco ensaios ( $E1, \dots, E5$ ), considerando diferentes estruturas de dependência espacial e tendo como base o modelo BGCCM (Figura 4.6a).

Para o cenário 1, além das simulações, também foi calculado o  $ESS_{bi}$  com os valores dos parâmetros simulados para o modelo BGCCM em cada ensaio. O objetivo foi comparar estes valores do  $ESS_{bi}$  com os valores médios do  $(ESS_{bi})_p$ ,  $p = 1, \dots, 100$ , obtidos ao estimar os parâmetros. Esse comparativo foi realizado apenas para esse cenário, pois os valores de alcance simulados nos respectivos ensaios não extrapolavam o *cutoff* da área e não eram inferiores à distância mínima entre os pontos.

Os cenários 2 e 3 contaram com três ensaios cada, nos quais variaram-se os valores dos alcances do modelo BGCCM, para que isso refletisse na correlação espacial entre os pares de pontos e, assim, fosse possível verificar as propriedades (ii) e (iii), respectivamente (Figura 4.6a).

Para a propriedade (iv) foi elaborado o cenário 4, utilizando o cenário 1 como referência, pois ele contém o tamanho amostral original ( $n=102$  pontos amostrais). O cenário 4 foi subdividido em três situações com diferentes tamanhos amostrais (75%, 50% e 25% de  $n$ ), cada uma delas com cinco ensaios (Figura 4.6a). Foram utilizados cinco ensaios para que fosse possível observar o

comportamento dos valores médios do  $ESS_{bi}$  diante do aumento dos valores dos alcances para os diferentes tamanhos amostrais.

(a)

Cenário 1 – (i)	Cenário 2 – (ii)	Cenário 3 – (iii)	Cenário 4 – (iv)	Cenário 5 – $\sigma$
E1: $\varphi = (75,75,75)^T$	E1: $\varphi = (600,600,600)^T$	E1: $\varphi = (10,10,10)^T$	<div style="display: flex; justify-content: space-around; font-size: small;"> <span>n = 77</span> <span>n = 51</span> <span>n = 26</span> </div>	E1: $\varphi = (125,75,75)^T$
E2: $\varphi = (125,125,125)^T$	E2: $\varphi = (1200,1200,1200)^T$	E2: $\varphi = (30,30,30)^T$		E2: $\varphi = (75,125,75)^T$
E3: $\varphi = (175,175,175)^T$	E3: $\varphi = (2000,2000,2000)^T$	E3: $\varphi = (50,50,50)^T$		E3: $\varphi = (75,75,125)^T$
E4: $\varphi = (225,225,225)^T$				E4: $\varphi = (175,125,125)^T$
E5: $\varphi = (275,275,275)^T$				E5: $\varphi = (125,175,125)^T$
			E1: $\varphi = (75,75,75)^T$	E6: $\varphi = (125,125,175)^T$
			E2: $\varphi = (125,125,125)^T$	E7: $\varphi = (225,175,175)^T$
			E3: $\varphi = (175,175,175)^T$	E8: $\varphi = (175,225,175)^T$
			E4: $\varphi = (225,225,225)^T$	E9: $\varphi = (175,175,225)^T$
			E5: $\varphi = (275,275,275)^T$	E10: $\varphi = (275,225,225)^T$
				E11: $\varphi = (225,275,225)^T$
				E12: $\varphi = (225,225,275)^T$

(b)

Condições iniciais	<ul style="list-style-type: none"> <li>• Ausência de tendência direcional e de anisotropia; modelo exponencial;</li> <li>• Valores dos parâmetros utilizados como chute inicial do modelo BGCCM: <math>\mu = (50,50)^T</math> e <math>\sigma = (2,5,2,5)^T</math>;</li> <li>• Esses parâmetros foram combinados aos valores de alcance do vetor <math>\varphi</math> segundo cada ensaio e cenário;</li> <li>• Os vetores <math>\varphi</math> dos cenários 1, 4 e 5 foram selecionados com base no <i>cutoff</i> da área simulada (50% da distância máxima), subdividindo este em cinco partes;</li> <li>• Coordenadas limitadas com base na configuração amostral da área (Figura 4.7b). A distância mínima e máxima entre os pontos foi, respectivamente, 50 e 1800 m.</li> </ul>
Simulação e estimação de parâmetros	<p>Para cada ensaio, foram simuladas duas variáveis, contendo 102 pontos amostrais em cada.</p> <p>Os valores simulados dos vetores de parâmetros <math>\mu</math>, <math>\sigma</math> e <math>\varphi</math> foram utilizados para estimar o vetor de parâmetros <math>\theta_p</math>, <math>p = 1, \dots, 100</math>, do modelo BGCCM.</p>
Estimação do $ESS_{bi}$	<p>O vetor <math>\hat{\theta}_p</math> foi utilizado para estimar <math>(ESS_{bi})_p</math> (Equação 4.8), <math>p = 1, \dots, 100</math>.</p>

Figura 4.6 – (a) Cenários e valores utilizados para o vetor de alcance  $\varphi = (\varphi_0, \varphi_1, \varphi_2)^T$  em cada ensaio; (b) Esquema metodológico utilizado nos estudos de simulação.  $\mu = (\mu_1, \mu_2)^T$ : vetor de médias;  $\sigma = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^T$ : vetor de dispersão; E1, ..., E12: ensaios 1 a 12.

O cenário 5 foi construído para analisar a influência dos parâmetros de dispersão ( $\sigma$ ) na estimação dos valores do  $ESS_{bi}$ . E, também, verificar se existe relação entre os alcances simulados e os valores médios estimados do vetor  $\sigma$ . Para tal, foram construídos 12 ensaios, com os valores do parâmetro de alcance variando individualmente. Diferentemente dos demais cenários, neste apenas os parâmetros de dispersão foram estimados pelo modelo BGCCM, os parâmetros de alcance foram mantidos fixos (Figura 4.6a). Os valores de alcance foram variados individualmente para que fosse possível analisar a influência do aumento de cada parâmetro de alcance ( $\varphi_0$ ,  $\varphi_1$  e  $\varphi_2$ ) na estimação do vetor  $\sigma$  e foram fixados para garantir que apenas um alcance variasse por vez, mantendo os outros dois iguais.

Em todos os ensaios de cada cenário foram calculados os valores do erro absoluto médio



(EAM) para avaliar a qualidade das estimativas dos parâmetros.

Para realizar as simulações foi utilizado um experimento de Monte Carlo a partir da decomposição de Cholesky da matriz de covariância  $\Sigma_Y$  (CRESSIE, 2015). As demais características a respeito das simulações foram apresentadas no esquema metodológico Figura 4.6b.

### Descrição do estudo de dados reais

Foram utilizados os atributos teor de matéria orgânica (MO,  $\text{g dm}^{-3}$ ) e soma de bases (SB,  $\text{cmol}_c \text{ dm}^{-3}$ ), que representa a soma de cátions trocáveis ( $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  e  $\text{K}^{+}$ ) no solo (Figura 4.7a). Estes atributos do solo apresentaram dependência espacial, isotropia e ausência de tendência direcional. Os dados experimentais pertencem ao Laboratório de Estatística Espacial (LEE) da Universidade Estadual do Oeste do Paraná - UNIOESTE, *campus* de Cascavel.

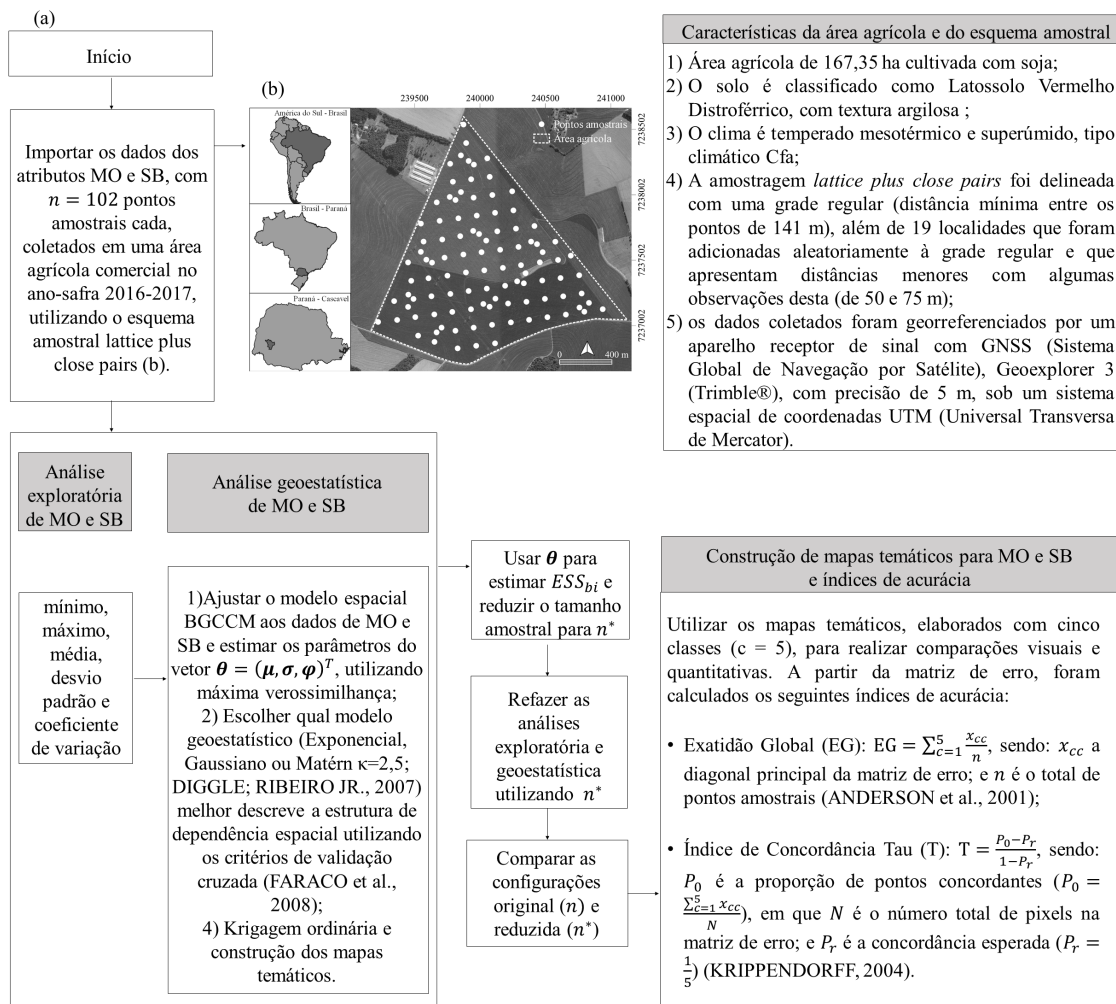


Figura 4.7 – (a) Esquema metodológico utilizado nos estudos de dados reais dos atributos do solo matéria orgânica (MO) e soma de bases (SB); (b) Área experimental e esquema amostral com as localizações dos pontos amostrais em coordenadas UTM.

As amostras de solo (Figura 4.7b) foram obtidas considerando, para cada ponto, a coleta de cinco subamostras, de 0 a 20 cm de profundidade nas proximidades dos pontos. Em seguida,

foram misturadas e colocadas em sacos plásticos, com aproximadamente 500 g, para a composição da amostra representativa da parcela. As amostras foram encaminhadas ao laboratório de análise do solo da Universidade Tecnológica Federal do Paraná, no qual os teores de cálcio (Ca), magnésio (Mg) e potássio (K) foram determinados de acordo com metodologia descrita em Pavan (1992). A soma de bases (SB) consiste na soma dos teores desses cátions e o teor de matéria orgânica do solo (MO) foi determinado de acordo com o método descrito por Walkley e Black (1934).

Foram construídos mapas temáticos para o par de atributos e calculados os índices de acurácia Exatidão Global (ANDERSON et al., 2001) e Tau (KRIPPENDORFF, 2004) para analisar a similaridade da variabilidade espacial apresentado pelo mapa, com tamanho amostral reduzido em comparação ao original (Figura 4.7a). A descrição e a sequência de etapas realizadas no estudo de dados reais são apresentadas no esquema metodológico da Figura 4.7a.

## Recursos computacionais

As rotinas para o cálculo do  $ESS_{bi}$  e para demais análises estatísticas e geoestatísticas foram desenvolvidas utilizando o *software* R (R DEVELOPMENT CORE TEAM, 2022) e o pacote *geoR* (RIBEIRO JR.; DIGGLE, 2001).

### 4.2.3 Resultados

#### Dados simulados: propriedades (i) a (iv) do tamanho amostral efetivo bivariado

A Tabela 4.5 apresenta os valores médios estimados dos parâmetros dos vetores de dispersão ( $\sigma$ ) e de alcance ( $\varphi$ ), como também do tamanho amostral efetivo bivariado ( $ESS_{bi}$ ) para os cenários 1 a 3. Além disso, também é exibido o desvio padrão do  $ESS_{bi}$  e o erro absoluto médio dos parâmetros considerando as 102 estimações. No cenário 1, conforme aumentaram os valores médios estimados de alcance ( $E1$  até  $E5$ ), os valores médios estimados do  $ESS_{bi}$  diminuíram (de 41,57 para 7,84). Os valores médios estimados para os parâmetros de dispersão, diminuíram com o aumento do alcance, algo que o cenário 2 também evidenciou. Em geral, os parâmetros de dispersão individuais ( $\sigma_1, \sigma_2$ ) foram, em média, maiores que os comuns ( $\sigma_{01}, \sigma_{02}$ ), especialmente a menores distâncias (Tabela 4.5).

Para o cenário 1, com intuito de comprovar a fidedignidade das estimativas médias do  $ESS_{bi}$  obtidas com os parâmetros estimados pelo BGCCM, foram utilizados os parâmetros simulados dos ensaios  $E1$  a  $E5$  para estimar o valor “real” do  $ESS_{bi}$  ( $R1$  a  $R5$ ). Os valores reais foram bastante próximos às médias do  $ESS_{bi}$  estimado ( $R1$ : 44,61;  $R2$ : 22,64;  $R3$ : 14,24;  $R4$ : 10,26;  $R5$ : 7,98). Além disso, os desvios-padrão dos valores estimados do  $ESS_{bi}$  foram baixos, variando de 12,95 a 0,17 (Tabela 4.5).

Nos cenários 1 e 2, o aumento acentuado dos valores médios estimados de alcance, fizeram o valor médio do  $\widehat{ESS}_{bi}$  cair e se aproximar de 1 (de 41,57 a 1,62; Tabela 4.5). Por outro

lado, no cenário 3, diminuir intensamente os valores médios estimados de alcance ( $E3$  até  $E1$ ), proporcionou um aumento no valor médio do  $\widehat{ESS}_{bi}$  (de 68,44 a 101,60; Tabela 4.5), fazendo este tender a 102, que é o tamanho amostral original ( $n$ ). Os desvios-padrão do  $\widehat{ESS}_{bi}$  nos cenários 2 e 3 variaram de 24,82 a 0,01, diminuindo de acordo com o aumento do alcance (Tabela 4.5).

O erro absoluto médio (EAM) nos cenários 1 a 3 variou de 1,57 a 5,00 nas estimativas dos parâmetros de dispersão, e de 5,78 a 171,37 nos parâmetros de alcance estimados (Tabela 4.5). Vale ressaltar que os erros são proporcionais à escala dos parâmetros simulados, dessa forma os parâmetros de dispersão têm EAM menor que os de alcance.

Tabela 4.5 – Valores médios estimados para os parâmetros de dispersão ( $\widehat{\sigma}_{01}$ ,  $\widehat{\sigma}_1$ ,  $\widehat{\sigma}_{02}$ ,  $\widehat{\sigma}_2$ ) e de alcance ( $\widehat{\varphi}_0$ ,  $\widehat{\varphi}_1$ ,  $\widehat{\varphi}_2$ ), e para o tamanho amostral efetivo bivariado ( $\widehat{ESS}_{bi}$ ), nos cenários 1 a 3. Entre parênteses estão o erro absoluto médio (EAM) das estimativas dos parâmetros e o desvio-padrão (DP) do  $\widehat{ESS}_{bi}$ .

		$\widehat{\sigma}_{01}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_{02}$	$\widehat{\sigma}_2$	$\widehat{\varphi}_0$	$\widehat{\varphi}_1$	$\widehat{\varphi}_2$	$\widehat{ESS}_{bi}$
Cenário 1	$E1$	0,47 (1,66)	6,19 (1,60)	2,58 (2,29)	4,29 (2,40)	92,35 (27,59)	68,97 (18,58)	74,50 (39,63)	41,57 (12,95)
	$E2$	0,46 (1,76)	4,79 (2,28)	1,33 (2,13)	4,27 (2,77)	137,05 (32,28)	118,65 (28,25)	120,47 (26,61)	22,73 (8,37)
	$E3$	0,10 (1,90)	1,79 (3,80)	0,38 (1,93)	1,93 (4,02)	186,79 (19,60)	170,75 (17,37)	178,63 (13,46)	13,76 (2,80)
	$E4$	0,01 (1,98)	0,37 (4,63)	0,17 (2,02)	0,55 (4,78)	233,12 (15,39)	221,01 (11,20)	232,02 (11,14)	10,47 (3,66)
	$E5$	$2,40 \cdot 10^{-11}$ ( $\approx 2,00$ )	$1,95 \cdot 10^{-10}$ ( $\approx 5,00$ )	$5,72 \cdot 10^{-11}$ ( $\approx 2,00$ )	$1,70 \cdot 10^{-10}$ ( $\approx 5,00$ )	285,98 (11,24)	278,75 (5,78)	279,28 (7,05)	7,84 (0,17)
Cenário 2	$E1$	$2,11 \cdot 10^{-11}$ ( $\approx 2,00$ )	$3,19 \cdot 10^{-10}$ ( $\approx 5,00$ )	$2,28 \cdot 10^{-10}$ ( $\approx 2,00$ )	$4,54 \cdot 10^{-10}$ ( $\approx 5,00$ )	629,67 (30,58)	607,81 (12,21)	608,75 (10,94)	3,50 (0,04)
	$E2$	$4,16 \cdot 10^{-12}$ ( $\approx 2,00$ )	$1,83 \cdot 10^{-10}$ ( $\approx 5,00$ )	$2,17 \cdot 10^{-10}$ ( $\approx 2,00$ )	$2,81 \cdot 10^{-10}$ ( $\approx 5,00$ )	1278,03 (79,63)	1208,10 (14,33)	1213,74 (17,08)	2,10 (0,02)
	$E3$	$1,55 \cdot 10^{-10}$ ( $\approx 2,00$ )	$1,87 \cdot 10^{-9}$ ( $\approx 5,00$ )	$1,34 \cdot 10^{-10}$ ( $\approx 2,00$ )	$1,59 \cdot 10^{-9}$ ( $\approx 5,00$ )	2171,38 (171,37)	2004,29 (20,99)	1987,34 (37,28)	1,62 (0,01)
Cenário 3	$E1$	1,14 (1,49)	5,83 (1,59)	0,63 (1,65)	6,27 (1,65)	6,75 (7,60)	12,77 (9,79)	14,63 (12,01)	101,60 (24,82)
	$E2$	1,40 (1,57)	5,43 (1,62)	1,35 (2,20)	5,63 (2,35)	32,50 (16,53)	29,43 (12,45)	29,86 (14,21)	87,63 (14,63)
	$E3$	0,58 (1,66)	6,15 (1,62)	2,95 (2,65)	3,96 (2,75)	54,89 (18,62)	46,43 (17,76)	46,58 (23,72)	68,44 (0,95)

$E1, \dots, E5$ : ensaios 1 a 5. “ $\approx$ ” foi utilizado para simbolizar valor aproximado.

A Tabela 4.6 exibe os valores médios estimados dos parâmetros de dispersão, de alcance e o  $ESS_{bi}$  para o cenário 4, para os três diferentes tamanhos amostrais. Além dos mesmos resultados observados nos cenários anteriores (Tabela 4.5), quanto à tendência inversa entre o vetor de alcance e o tamanho amostral efetivo bivariado, houve também um aumento no valor médio estimado do  $ESS_{bi}$  e no seu desvio-padrão (DP) conforme pontos foram adicionados à grade amostral: no  $E1$  foi de 17,32 para 37,29 (DP de 4,44 para 15,14); no  $E2$  de 13,89 para 21,24 (DP de 4,33 para 8,29); no  $E3$  de 10,11 para 13,76 (DP de 2,76 para 6,05); no  $E4$  de 7,87 para 13,36 (DP de 1,71 para 4,09); e por fim em  $E5$  de 6,49 para 7,48, no qual o DP não aumentou, devido à proximidade dos valores do  $\widehat{ESS}_{bi}$  (Tabela 4.6).

Tabela 4.6 – Valores médios estimados para os parâmetros de dispersão ( $\widehat{\sigma}_{01}$ ,  $\widehat{\sigma}_1$ ,  $\widehat{\sigma}_{02}$ ,  $\widehat{\sigma}_2$ ) e de alcance ( $\widehat{\varphi}_0$ ,  $\widehat{\varphi}_1$ ,  $\widehat{\varphi}_2$ ), e para o tamanho amostral efetivo bivariado ( $\widehat{ESS}_{bi}$ ) no cenário 4. Entre parênteses estão o erro absoluto médio (EAM) das estimativas dos parâmetros e o desvio-padrão (DP) do  $\widehat{ESS}_{bi}$ .

		$\widehat{\sigma}_{01}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_{02}$	$\widehat{\sigma}_2$	$\widehat{\varphi}_0$	$\widehat{\varphi}_1$	$\widehat{\varphi}_2$	$\widehat{ESS}_{bi}$	
Cenário 4	$n=77$ (75%)	E1	0,62 (1,75)	5,92 (1,66)	2,46 (2,31)	4,40 (2,54)	97,23 (31,01)	65,68 (23,74)	66,01 (34,59)	37,29 (15,14)
		E2	0,52 (1,88)	4,86 (2,11)	2,12 (2,35)	3,70 (2,90)	170,52 (62,89)	113,90 (36,66)	110,08 (37,20)	21,24 (8,29)
		E3	0,16 (1,91)	2,44 (3,33)	0,97 (2,22)	2,15 (3,69)	194,16 (36,29)	164,61 (25,51)	168,25 (30,24)	13,76 (6,05)
		E4	0,05 (1,97)	0,87 (4,35)	0,53 (2,08)	0,90 (4,49)	232,28 (21,83)	217,20 (18,97)	230,40 (14,37)	10,36 (4,09)
		E5	0,02 (1,97)	0,24 (4,78)	0,08 (2,02)	0,33 (4,85)	289,19 (15,77)	275,68 (11,37)	278,30 (9,47)	7,48 (0,38)
	$n=51$ (50%)	E1	0,46 (1,63)	5,83 (1,78)	2,82 (2,58)	3,91 (2,68)	101,36 (36,02)	60,82 (30,88)	67,28 (40,64)	28,70 (10,65)
		E2	0,37 (1,76)	4,97 (2,04)	2,44 (2,54)	3,54 (2,81)	157,02 (52,68)	105,98 (33,94)	116,42 (59,92)	18,49 (7,72)
		E3	0,22 (1,88)	3,30 (2,88)	1,26 (2,26)	2,96 (3,32)	197,72 (40,72)	158,68 (36,87)	165,81 (41,87)	12,95 (4,79)
		E4	0,04 (1,95)	1,61 (3,81)	0,78 (2,05)	1,66 (4,23)	242,50 (38,88)	209,03 (25,06)	230,76 (33,36)	9,61 (3,91)
		E5	0,02 (1,98)	1,02 (4,16)	0,62 (2,21)	1,10 (4,24)	306,30 (44,91)	257,65 (28,52)	283,61 (30,98)	7,15 (1,75)
	$n=26$ (25%)	E1	0,86 (1,69)	5,44 (2,05)	2,68 (2,63)	4,02 (2,51)	110,17 (45,52)	53,43 (39,03)	65,70 (41,70)	17,32 (4,44)
		E2	0,76 (1,77)	4,35 (2,13)	2,40 (2,38)	3,51 (2,95)	145,37 (50,08)	102,18 (49,69)	106,91 (58,19)	13,89 (4,33)
		E3	0,43 (1,91)	3,60 (2,67)	2,12 (2,62)	2,44 (3,45)	205,45 (55,01)	149,13 (52,66)	178,98 (84,78)	10,11 (2,76)
		E4	0,12 (1,88)	3,07 (2,93)	1,25 (2,16)	2,36 (3,34)	255,67 (48,83)	190,27 (60,70)	187,42 (54,76)	7,87 (1,71)
		E5	0,11 (1,91)	2,53 (3,29)	1,23 (2,31)	2,04 (3,52)	303,92 (59,50)	236,79 (62,18)	255,68 (61,41)	6,49 (1,96)

$E1, \dots, E5$ : ensaios 1 a 5.

A Tabela 4.6 evidencia ainda que os valores estimados dos parâmetros de dispersão e alcance tiveram a tendência de aumentar e se distanciar dos valores reais conforme se reduziu o tamanho amostral, especialmente em ensaios cujo valor simulado para o alcance foi mais elevado ( $E4$  e  $E5$ ). Fato ressaltado pelo aumento nos valores do EAM (Tabela 4.6).

A Tabela 4.7 mostra que os valores médios estimados dos parâmetros de dispersão diminuíram, conforme o alcance comum ( $\varphi_0$ ) foi aumentado ( $E1$ ,  $E4$ ,  $E7$  e  $E10$ ), distanciando-se dos valores simulados e com EAM maiores. Já  $\varphi_1$  ( $\varphi_2$ ) quando aumentado, mantendo  $\varphi_0$  fixo, fez com que todos os valores médios estimados dos parâmetros de dispersão comum  $\sigma_{01}$  ( $\sigma_{02}$ ) e a maioria dos individuais  $\sigma_1$  ( $\sigma_2$ ) tivessem um pequeno aumento. Como exemplo, pode-se observar os ensaios  $E11$  e  $E12$ :  $\varphi_0$  é fixo em 225 m; e  $\varphi_1$  vale 225 m e 275 m, em  $E12$  e  $E11$ , respectivamente. Em  $E12$ ,  $\sigma_{01}$  ( $\sigma_1$ ) é igual a 0,003 (0,49). Enquanto em  $E11$  é igual a 0,01 (0,52). Já  $\varphi_2$  vale 225 m e 275 m, em  $E11$  e  $E12$ , respectivamente. Em  $E11$ ,  $\sigma_{02}$  ( $\sigma_2$ ) é 0,33 (0,49). Enquanto em  $E12$  é igual a 0,34 (0,53).

No cenário 5, independente dos valores dos parâmetros de dispersão, ao aumentar  $\varphi_0$  ( $E1$ : 1125 m para  $E10$ : 275 m), o valor médio estimado do  $ESS_{bi}$  diminuiu (de 27,95 para 8,84), assim

como o DP (de 2,74 para 0,27) (Tabela 4.7). Ao fixar o alcance comum e aumentar os alcances individuais, o valor médio estimado do  $ESS_{bi}$  também diminuiu (de 45,30 para 10,42), porém não houve distinção entre os valores do  $ESS_{bi}$  quando se aumentou individualmente  $\varphi_1$  ou  $\varphi_2$  (E2-E3; E5-E6; E8-E9; E11-E12) (Tabela 4.7).

Tabela 4.7 – Valores médios estimados para os parâmetros de dispersão ( $\widehat{\sigma}_{01}$ ,  $\widehat{\sigma}_1$ ,  $\widehat{\sigma}_{02}$ ,  $\widehat{\sigma}_2$ ) e para o tamanho amostral efetivo bivariado ( $\widehat{ESS}_{bi}$ ) no cenário 5. Entre parênteses estão o erro absoluto médio (EAM) das estimativas dos parâmetros e o desvio-padrão (DP) do  $\widehat{ESS}_{bi}$ .

		$\widehat{\sigma}_{01}$	$\widehat{\sigma}_1$	$\widehat{\sigma}_{02}$	$\widehat{\sigma}_2$	$\widehat{ESS}_{bi}$
Cenário 5	E1	0,38 (1,71)	5,68 (1,85)	1,78 (2,17)	4,52 (2,43)	27,95 (2,74)
	E2	0,29 (1,70)	6,39 (1,66)	2,16 (2,13)	4,62 (2,08)	45,30 (2,30·10 <sup>-13</sup> )
	E3	0,25 (1,76)	6,37 (1,41)	2,73 (1,93)	4,24 (1,94)	45,30 (2,30·10 <sup>-12</sup> )
	E4	0,25 (1,85)	2,76 (3,07)	1,08 (2,25)	2,54 (3,53)	16,99 (0,95)
	E5	0,35 (1,78)	3,80 (2,81)	0,81 (2,05)	3,59 (3,12)	22,87 (1,24·10 <sup>-13</sup> )
	E6	0,24 (1,76)	5,52 (2,00)	1,03 (1,76)	5,07 (2,31)	22,87 (8,14·10 <sup>-13</sup> )
	E7	0,14 (2,01)	1,23 (4,18)	0,46 (2,03)	1,34 (4,33)	11,61 (0,48)
	E8	0,09 (1,90)	1,62 (3,93)	0,36 (1,99)	1,72 (4,20)	14,45 (2,51·10 <sup>-13</sup> )
	E9	0,05 (1,94)	2,20 (3,68)	0,61 (2,01)	2,15 (3,92)	14,45 (2,28·10 <sup>-13</sup> )
	E10	0,02 (1,97)	0,36 (4,67)	0,28 (2,09)	0,50 (4,84)	8,84 (0,27)
	E11	0,01 (1,98)	0,52 (4,73)	0,33 (2,10)	0,49 (4,68)	10,42 (9,03·10 <sup>-13</sup> )
	E12	0,003 (1,99)	0,49 (4,61)	0,34 (2,05)	0,53 (4,73)	10,42 (9,33·10 <sup>-13</sup> )

E1, ..., E12: ensaios 1 a 12.

### Verificação algébrica da propriedade (v) do tamanho amostral efetivo bivariado

Na sequência será realizada a verificação algébrica de que  $ESS_{bi} \geq 1$ .

Da Equação 4.8, tem-se que  $ESS_{bi} = \mathbf{1}^T \mathbf{R}_{(bi)}^{-1} \mathbf{1}$ , sendo  $\mathbf{1}$  um vetor de uns,  $2n \times 1$ , e  $\mathbf{R}_{(bi)}$  uma matriz de correlação espacial bivariada,  $2n \times 2n$ , dada por:

$$\mathbf{R}_k = \begin{cases} 1, & \text{se } i = j \\ \frac{a_{ij}}{\sigma_{0k}^2 + \sigma_k^2}, & \text{se } i \neq j \end{cases}, \text{ e}$$

$$\mathbf{R}_{1,2} = \begin{cases} 1, & \text{se } i = j \\ \frac{b_{ij}}{\sigma_{01} + \sigma_{02}}, & \text{se } i \neq j \end{cases},$$

para  $i, j = 1, \dots, n; k = 1, 2$ . Sendo  $a_{ij}$  os elementos da matriz  $\Sigma_k$  relacionados a  $Y_k$  (Equação 4.5), e  $b_{ij}$  são os elementos da matriz  $\Sigma_{1,2}$  relacionados a  $Y_{1,2}$  (Equação 4.6), em que:

$$\mathbf{R}_{(bi)} = \begin{pmatrix} \mathbf{R}_1 & \mathbf{R}_{1,2} \\ \mathbf{R}_{1,2}^T & \mathbf{R}_2 \end{pmatrix},$$

em que:  $\mathbf{R}_k$  e  $\mathbf{R}_{1,2}$  tem dimensão  $n \times n$ .

Denota-se por  $\mathbf{R}_{(bi)} = [r_{uv}]$ , uma matriz  $2n \times 2n$ , com  $u, v = 1, \dots, 2n$  e  $n \in N$ . Sabe-se que  $4n^2 \leq 4n^2n$ . Usando a desigualdade de Cauchy-Schwartz para matrizes baseado em Vallejos e Osorio (2014), tem-se que:

$$4n^2 \leq (\mathbf{1}^T \mathbf{R}_{(bi)} \mathbf{1}) (\mathbf{1}^T \mathbf{R}_{(bi)}^{-1} \mathbf{1}), \quad (4.9)$$

pois,  $MAX(\mathbf{1}^T \mathbf{R}_{(bi)} \mathbf{1}) = 4n^2$ , e  $MAX(\mathbf{1}^T \mathbf{R}_{(bi)}^{-1} \mathbf{1}) = n$ .

Como  $0 \leq r_{uv} \leq 1$ , tem-se que:

$$\mathbf{1}^T \mathbf{R}_{(bi)} \mathbf{1} = 4n + \sum_{v=1}^{2n} \sum_{u \neq v} r_{uv} \leq 4n + 4n(n-1).$$

Logo,  $\mathbf{1}^T \mathbf{R}_{(bi)} \mathbf{1} \leq 4n + 4n^2 - 4n = 4n^2$ . E então,  $\mathbf{1}^T \mathbf{R}_{(bi)} \mathbf{1} \leq 4n^2$ .

Voltando na Equação 4.9, se  $\mathbf{1}^T \mathbf{R}_{(bi)} \mathbf{1} \leq 4n^2$ , então tem-se que  $\mathbf{1}^T \mathbf{R}_{(bi)}^{-1} \mathbf{1} \geq 1$  para que a desigualdade  $\mathbf{1}^T \mathbf{R}_{(bi)} \mathbf{1} \leq 4n^2$  seja satisfeita. E assim:

$$1 \leq \mathbf{1}^T \mathbf{R}_{(bi)}^{-1} \mathbf{1} = ESS_{bi},$$

demonstrando que  $ESS_{bi} \geq 1$ .

### Aplicação aos dados de matéria orgânica e soma de bases em área agrícola cultivada com soja

O valor estimado de  $ESS_{bi}$  obtido para o par MO-SB foi de 42 pontos amostrais ( $n^*$ ). Isso representou uma redução de aproximadamente 60% no total de observações amostrais originais ( $n = 102$ ).

Nas Tabelas 4.8 e 4.9 são apresentadas as medidas de estatística descritiva dos atributos MO e SB, e os valores estimados dos parâmetros do modelo BGCCM para o par MO-SB, considerando o tamanho da amostra original  $n$  e o tamanho da amostra reduzido  $n^*$ .

Considerando a configuração amostral original ( $n$ ), os teores de MO variaram de 13,40 a 89,80 g dm<sup>-3</sup>, enquanto os teores de SB ficaram entre 2,55 e 9,65 cmol<sub>c</sub> dm<sup>-3</sup> (Tabela 4.8). Em média, os teores de MO (42,14 g dm<sup>-3</sup>) e de SB (6,05 cmol<sub>c</sub> dm<sup>-3</sup>) foram considerados, respectivamente, muito altos e altos para o solo da região (SBCS-NEPAR, 2017) (Tabela 4.8). Os coeficientes de variação (CV) de MO (24,93%) e SB (22,83%) foram próximos entre si e considerados altos, segundo classificação de Pimentel Gomes e Garcia (2002) (Tabela 4.8). Para a configuração amostral reduzida ( $n^*$ ), os teores médios de MO (41,87 g dm<sup>-3</sup>) e de SB (6,32 cmol<sub>c</sub> dm<sup>-3</sup>), assim como o CV de MO (21,96%) e de SB (24,03%), mantiveram suas classificações (PIMENTEL GOMES; GARCIA, 2002; SBCS-NEPAR, 2017).

Tabela 4.8 – Medidas da análise exploratória (mínimo, máximo, média, DP e CV) dos teores de matéria orgânica (MO, g dm<sup>-3</sup>) e soma de bases (SB, cmol<sub>c</sub> dm<sup>-3</sup>), utilizando a configuração amostral original ( $n = 102$ ) e reduzida ( $n^* = 42$ ).

Configuração amostral	Atributos	Mínimo	Máximo	Média	DP	CV (%)
$n$	MO	13,40	89,80	42,14	10,51	24,93
	SB	2,55	9,65	6,05	1,38	22,83
$n^*$	MO	22,78	57,63	41,87	9,19	21,96
	SB	2,55	9,65	6,32	1,52	24,03

DP: desvio padrão; CV =  $100 \frac{DP}{Média}$ : coeficiente de variação (%).

Para o BGCCM, segundo os critérios de validação cruzada (FARACO et al., 2008), o par MO-SB foi melhor ajustado pelo modelo exponencial (para  $n$ ) e pelo modelo Matérn com  $\kappa = 2,5$  (para  $n^*$ ) (Tabela 4.9). Para os modelos exponencial e Matérn com  $\kappa = 2,5$ , os raios de dependência espacial ( $a_t$ ) são dados por  $3\varphi_t$  e  $5,92\varphi_t$ ,  $t = 0, 1, 2$  (DIGGLE; RIBEIRO JR., 2007). Utilizando a totalidade de observações amostrais, os valores estimados do raio de dependência espacial vinculados a MO ( $a_M$ ) e SB ( $a_S$ ) foram semelhantes entre si: 806,70 m e 886,81 m, respectivamente (Tabela 4.9). Já o raio de dependência espacial comum ( $a_0$ ) teve um valor estimado menor, igual a 237,97 m (Tabela 4.9).

Com a redução no número de pontos amostrais,  $a_M$  (1430,16 m) e  $a_S$  (1430,51 m) foram basicamente iguais, enquanto  $a_0$  teve um valor estimado menor, 961,15 m (Tabela 4.9). Como para o modelo Matérn a constante multiplicada à função de alcance é maior, conseqüentemente os raios de dependência espacial foram mais altos com a configuração reduzida comparada à original.

Com  $n$  pontos amostrais, o valor estimado do parâmetro de dispersão individual associado a SB ( $\sigma_S = 5,78 \cdot 10^{-7}$ ) foi menor comparado ao de MO ( $\sigma_M = 1,91 \cdot 10^2$ ) (Tabela 4.9). O mesmo ocorreu com o parâmetro de dispersão comum de SB ( $\sigma_{0S} = 1,87$ ), tendo sido inferior ao valor de MO ( $\sigma_{0M} = 3,14$ ) (Tabela 4.9). Utilizando  $n^*$  observações amostrais, todos os valores estimados para os parâmetros de dispersão foram inferiores aos da configuração original, tendo variado entre  $1,65 \cdot 10^{-10}$  e  $4,88 \cdot 10^{-13}$  (Tabela 4.9).

Tabela 4.9 – Medidas da análise geoestatística contendo os parâmetros estimados do BGCCM para o par matéria orgânica (MO, g dm<sup>-3</sup>) e soma de bases (SB, cmol<sub>c</sub> dm<sup>-3</sup>), utilizando a configuração amostral original ( $n = 102$ ) e reduzida ( $n^* = 42$ ).

Modelo geoestatístico	$\widehat{\mu}_M$	$\widehat{\mu}_S$	$\widehat{\sigma}_{0M}$	$\widehat{\sigma}_M$	$\widehat{\sigma}_{0S}$	$\widehat{\sigma}_S$	$\widehat{\varphi}_0$ ( $\widehat{a}_0$ )	$\widehat{\varphi}_M$ ( $\widehat{a}_M$ )	$\widehat{\varphi}_S$ ( $\widehat{a}_S$ )
Exponencial	43,95	6,14	3,14	$1,91 \cdot 10^2$	1,87	$5,78 \cdot 10^{-7}$	79,43 (237,97)	269,28 (806,70)	296,02 (886,81)
Matérn 2,5	51,22	2,81	$4,88 \cdot 10^{-13}$	$7,94 \cdot 10^{-11}$	$1,65 \cdot 10^{-10}$	$1,96 \cdot 10^{-11}$	162,39 (961,15)	241,63 (1430,16)	241,69 (1430,51)

$\widehat{\mu}_M$  ( $\widehat{\mu}_S$ ),  $\widehat{\sigma}_M$  ( $\widehat{\sigma}_S$ ),  $\widehat{\varphi}_M$  ( $\widehat{\varphi}_S$ ) e  $\widehat{a}_M$  ( $\widehat{a}_S$ ): são os valores estimados dos parâmetros média, dispersão, função de alcance e alcance prático do teor de MO (SB);  $\widehat{\sigma}_{0M}$  ( $\widehat{\sigma}_{0S}$ ),  $\widehat{\varphi}_0$  e  $\widehat{a}_0$ : são os valores estimados dos parâmetros de dispersão de MO (SB), função de alcance e alcance prático associados ao campo aleatório comum.

A Figura 4.8 apresenta os mapas de variabilidade espacial de MO tendo SB como covariável obtidos com a configuração amostral original (a) e reduzida (b). A Exatidão Global foi de 79,61%, sendo próximo ao nível de acurácia tido como adequado ( $EG \geq 85\%$ ; ANDERSON et al., 2001). O

índice Tau foi de 74,51%, o que indica moderada acurácia entre as classes dos mapas temáticos ( $67\% \leq \text{Tau} < 80\%$ ; KRIPPENDORFF, 2004).

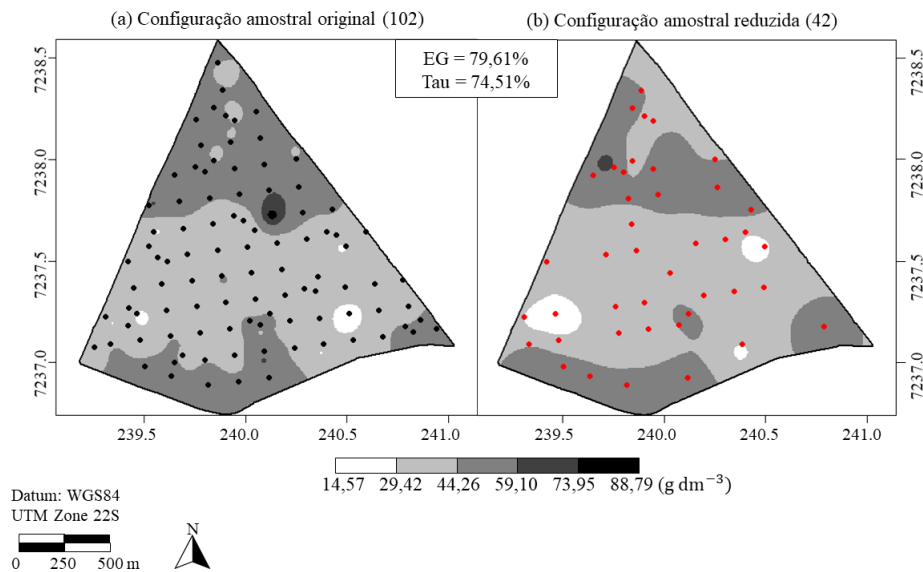


Figura 4.8 – Mapa de variabilidade espacial da matéria orgânica tendo a soma de bases como covariável: (a) Configuração amostral original com 102 pontos amostrais (em preto), (b) Configuração amostral reduzida com 42 pontos amostrais (em vermelho). Valores estimados dos índices Exatidão Global (EG) e índice de concordância Tau (Tau).

#### 4.2.4 Discussão

No cenário 1 (Tabela 4.5) dos estudos de simulação foi verificado que, ao aumentar os valores médios estimados de alcance, os valores médios estimados do  $ESS_{bi}$  diminuíram. Esses resultados atendem a propriedade (i), pois aumentar o raio de dependência espacial corresponde a aumentar a correlação espacial entre as observações.

No cenário 2 (Tabela 4.5), foi constatado que aumentar abruptamente o alcance, simulando uma perfeita correlação espacial entre os pares de observações, fez com que os valores médios estimados do  $ESS_{bi}$  tendessem a 1, satisfazendo a propriedade (ii). Por outro lado, no cenário 3 (Tabela 4.5), ao diminuir a intensidade do alcance, sendo  $E1$  e  $E2$  com alcances menores que a menor distância entre a maioria dos pares de pontos e  $E3$  próximo ao limiar da menor distância, reproduziu-se a ausência de correlação espacial, tanto entre as variáveis ( $\varphi_0$ ) quanto individualmente ( $\varphi_1$ ,  $\varphi_2$ ), e foi verificado que os valores médios estimados do  $ESS_{bi}$  aumentaram e tenderam à quantidade total de observações ( $n$ ), o que verifica a propriedade (iii).

No cenário 4 (Tabela 4.6), ao utilizar 75% (77 pontos), 50% (51 pontos) e 25% (26 pontos) do tamanho amostral original (102 pontos), observou-se que o valor estimado do  $ESS_{bi}$  aumentou, de acordo com o aumento no número de pontos na grade amostral. Isso mostra que o  $ESS_{bi}$  cresce à medida que o tamanho amostral aumenta, o que satisfaz a propriedade (iv). E, por fim, a validade da propriedade (v) foi demonstrada algebricamente.



Com o cenário 5 (Tabela 4.7) foi possível constatar a influência do parâmetro de alcance comum ( $\varphi_0$ ) no valor médio estimado do  $ESS_{bi}$ . Pois, ao comparar os ensaios  $E1$  a  $E4$  do cenário 1, quando os valores simulados para parâmetros de alcance foram de 75 a 225 m, com os ensaios  $E2-E3$ ,  $E5-E6$ ,  $E8-E9$ , e  $E11-E12$  do cenário 5, em que os valores de  $\varphi_0$  foram fixados de 75 a 225 m, observou-se que os valores estimados do  $ESS_{bi}$  são muito similares respectivamente. Além disso, verificou-se que entre esses pares de ensaios do cenário 5 não houve diferença no valor estimado do  $ESS_{bi}$ .

Portanto, foi verificado que as cinco propriedades que incidem sobre o ESS univariado (VALLEJOS; OSORIO, 2014) também são válidas para a metodologia bivariada do  $ESS_{bi}$  utilizada neste trabalho. Sobre os erros nas estimativas dos parâmetros, ficou evidente que o alcance é um parâmetro influente, pois os valores do EAM foram maiores conforme se aumentou o alcance simulado, tendo em vista que as estimativas médias dos parâmetros se distanciaram do valor real. Os parâmetros de dispersão, especialmente o  $\sigma_{01}$ , se mostraram mais sensíveis e variaram mais, apresentando erros proporcionalmente mais elevados que os parâmetros de alcance. Já o valor estimado do  $ESS_{bi}$  variou menos conforme se aumentou o alcance simulado, pois os valores do desvio-padrão diminuíram. Fato análogo foi observado por Vallejos e Osorio (2014) no caso univariado do ESS.

No geral, considerando a influência do alcance na estimação do  $ESS_{bi}$  e que os erros na estimação dos parâmetros de alcance, dada sua magnitude, não foram altos, pode-se concluir que houve acurácia nas estimativas do modelo. Sendo assim, a expressão do  $ESS_{bi}$  foi aplicada a um conjunto de dados reais contendo o par MO e SB para que se pudesse analisar sua aplicabilidade e viabilidade na prática.

A MO é uma das variáveis mais importante em relação a fertilidade do solo. Porquanto, em solos nos quais a capacidade de troca catiônica (CTC) é muito baixa, como é o caso do solo da região (latossolo), esse é o único atributo que faz aumentar a CTC (MENGEL; KIRKBY, 2001). Quanto maior a MO, maior também será a CTC, e mais elevada será a capacidade do solo em reter os cátions: cálcio ( $Ca^{2+}$ ), magnésio ( $Mg^{2+}$ ) e potássio ( $K^+$ ) que, somados, resultam em SB. Como produto final das análises geoestatísticas, foi obtido o mapa da variabilidade espacial de MO tendo SB como covariável.

Embora a Exatidão Global não tenha atingido o nível de acurácia considerado adequado para o mapa de variabilidade espacial com configuração amostral reduzida comparado ao original (Figura 4.8), o índice tem a desvantagem de considerar apenas a diagonal principal da matriz de confusão, analisando, assim, apenas os *pixels* corretamente classificados (ANDERSON et al., 2001). Enquanto o índice Tau, que mostrou acurácia moderada entre os mapas (Figura 4.8), considera, além da diagonal principal, também a possibilidade de ocorrência de cada classe, fornecendo assim uma medida quantitativa precisa sobre a acurácia de classificação (MA; REDMOND, 1995).

Visualmente é possível perceber a semelhança entre os mapas (Figura 4.8). A terceira (entre 29,42 e 44,26  $g\ dm^{-3}$ ) e quarta classes (entre 44,26 e 59,10  $g\ dm^{-3}$ ) concentraram a maior parte da área em ambos os mapas, sendo que os contornos que representam a variabilidade espacial no sul, no centro e mesmo ao norte da área agrícola foram mantidos no mapa reduzido. É possível observar, inclusive, que no extremo norte do mapa original, têm-se cinco regiões circulares pertencentes à terceira classe (Figura 4.8a). Com a redução no número de pontos amostrais, essas regiões foram suavizadas e agrupadas (Figura 4.8b), de modo semelhante ao que se observa nos

trabalhos de Kestring et al. (2015) e Maltauro et al. (2021) que analisaram diferentes tamanhos amostrais em áreas agrícolas. O mesmo ocorreu com outras regiões menores no centro e ao sul da área, entretanto, apesar de se tornarem regiões maiores no mapa reduzido, mantiveram o padrão espacial observado no mapa original.

Em termos práticos, isso implica que o teor de matéria orgânica no solo poderia ser analisado com apenas 40% dos pontos amostrais que, atualmente, são coletados na área agrícola sem perder a qualidade na representação da variabilidade espacial. Para o produtor, isso resulta em continuar realizando o manejo e a correção do solo nos locais adequados, mas com menor investimento financeiro em análises de solo, que reduzirá na mesma proporção que a quantidade de pontos amostrais.

Os estudos utilizando dados reais de atributos químicos e físicos do solo, desenvolvidos por Griffith (2005, 2008), Canton et al. (2021) e Vallejos e Acosta (2021), nos quais o ESS univariado foi calculado, obtiveram reduções amostrais que variaram entre 30% e 61%. Os trabalhos de Canton, Guedes e Uribe-Opazo (2021) e Vallejos e Acosta (2021), nos quais foram estimados valores multivariados do ESS, as reduções amostrais variaram entre 21% e 60%. Todavia, embora a redução máxima obtida nesses estudos tenha sido corroborada por este trabalho (60%), a estrutura de correlação entre duas variáveis foi construída utilizando uma matriz de pesos espaciais ou semivariograma cruzado. Em nenhum deles foi utilizado o modelo BGCCM, que possibilita modelar a estrutura de dependência espacial de modo bivariado, agregando a correlação espacial entre as variáveis e dentro de cada variável, característica que, segundo Griffith (2005), deve ser considerada no tratamento conjunto de duas variáveis espacialmente correlacionadas.

#### 4.2.5 Conclusão

Os estudos de simulação foram satisfatórios em verificar que as propriedades que acometem o ESS univariado, também são válidas para o  $ESS_{bi}$ , além de possibilitar conclusões adicionais acerca da influência do alcance comum ( $\varphi_0$ ) sobre o valor do  $ESS_{bi}$ .

Ao aplicar a metodologia no par de atributos do solo MO-SB, verificou-se que 60% das observações do tamanho amostral original carregavam informações duplicadas. E, após o redimensionamento amostral, a variabilidade espacial do mapa de MO tendo SB como covariável apresentou grande semelhança visual, além de exibir um índice de concordância Tau que assegurou uma similaridade moderada entre os mapas com tamanho amostral original e reduzido.

Os estudos simulados e a aplicação a dados reais da metodologia do  $ESS_{bi}$ , utilizando o BGCCM, evidenciaram sua relevância prática e também teórica. Pois, além de manter todas as características e propriedades da metodologia univariada, fornece uma melhor viabilidade prática, por se tratar de uma proposta que reduz a quantidade de amostras de solo que vão ao laboratório para duas variáveis simultaneamente, mantendo a qualidade dos mapas temáticos obtidos. Sendo assim, é possível diminuir gastos em monitoramentos de solos, se forem utilizados os modelos geoestatísticos bivariados para problemas com atributos químicos espacialmente correlacionados.

## Referências

- ACOSTA, J.; OSORIO, F.; VALLEJOS, R. Effective sample size for line transect sampling models with an application to marine macroalgae. **Journal of Agricultural, Biological, and Environmental Statistics**, v. 21, n. 3, p. 407-425, 2016.
- ACOSTA, J.; VALLEJOS, R. Effective sample size for spatial regression models. **Electronic Journal of Statistics**, v. 12, p. 3147-3180, 2018.
- 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.
- BENEDETTI, R.; PIERSIMONI, F.; POSTIGLIONE, P. **Sampling Spatial Units fo Agricultural Surveys**. Berlin: Springer, 2015. 325p.
- CANTON, L. E. D.; GUEDES, L. P. C.; URIBE-OPAZO, M. A. Reduction of sample size in the soil physical-chemical attributes using the multivariate Effective Sample Size. **Journal of Agricultural Studies**, v. 9, n. 1, p. 357-376, 2021a.
- CANTON, L. E. D.; GUEDES, L. P. C.; URIBE-OPAZO, M. A.; ASSUMPÇÃO, R. A. B.; MALTAURO, T. C. Sampling redesign of soil penetration resistance in spatial t-Student models. **Spanish Journal of Agricultural Research**, v. 19, n. 1, p. e0202, 2021b.
- CRESSIE, N. A. C. **Statistics for Spatial Data**. rev. ed. New York: John Wiley & Sons, 2015. 928p.
- CUNHA, T. J. F.; MENDES, M. A. S.; GIONGO, V. Matéria orgânica do solo. In: NUNES, R. R.; REZENDE, M. O. O. (eds), **Recurso solo: propriedades e usos**. Cubo: São Carlos, 2015. p. 273-293.
- 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.
- FONSECA, B. H. F. **Um estudo sobre estimação e predição em modelos geoestatísticos bivariados**. 2008. 74 f. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo, 2008.
- 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.

GRIFFITH, D. A. Geographic sampling of urban soils for contaminant mapping: how many samples and from where. **Environmental Geochemistry Health**, v. 30, p. 495-509, 2008.

GRIFFITH, D. A.; PAELINCK, H. P. **Non-Standard Spatial Statistics and Spatial Econometrics**. Berlin: Springer, 2011. 264p.

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. 456p.

MA, Z.; REDMOND, R. L. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering & Remote Sensing**, v. 61, n. 4, p. 453-439, 1995.

MALTAURO, T. C.; GUEDES, L. P. C.; URIBE-OPAZO, M. A.; CANTON, L. E. D. A genetic algorithm for resizing and sampling reduction of non-stationary soil chemical attributes optimizing spatial prediction. **Spanish Journal of Agricultural Research**, v. 19, n. 4, p. e0210-e0210, 2021.

MENGEL, K.; KIRKBY, E. **Principles of plant nutrition**. 5th ed. Boston: Kluwer Academic Publishers, 2001. 849p.

PAVAN, M. A.; BLOCH, M. F.; ZEMPULSKI, H. D.; MIYAZAWA, M.; ZOCOLER, D. C. **Manual de análise química de solo e controle de qualidade**. Londrina: Instituto Agrônomo do Paraná, 1992. 38p.

PIMENTEL GOMES, F.; GARCIA, C. H. **Estatística Aplicada a Experimentos Agronômicos e Florestais**. Piracicaba: FEALQ, 2002. 309p.

R DEVELOPMENT CORE TEAM. **R: A language and environment for statistical computing**. R foundation for Statistical Computing, Vienna, Austria, 2022.

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

RIGHETTO, A. J. **Avaliação de modelos geoestatísticos multivariados**. 2012. 92 f. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Universidade de São Paulo, Escola Superior de Agricultura "Luiz de Queiroz", São Paulo, 2012.

SOCIEDADE BRASILEIRA DE CIÊNCIA DO SOLO - SBCS. NÚCLEO ESTADUAL PARANÁ -

NEPAR. **Manual de adubação e calagem para o estado do Paraná**. SBCS-NEPAR, Curitiba, 2017. Disponível em: <https://sbcs-nepar.org.br/produto/manual-de-adubacao-e-calagem-para-o-estado-do-parana-2a-edicao/>. Acesso em: 09 Set. 2017.

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

VALLEJOS, R.; OSORIO, F.; BEVILACQUA, M. **Spatial relationships between two georeferenced variables**: With applications in R. Cham: Springer, 2020. 194p.

VALLEJOS, R.; ACOSTA, J. The effective sample size for multivariate spatial processes with an application to soil contamination. **Natural Resource Modeling**, v. 34, n. 4, p. e12322, 2021.

WALKLEY, A.; BLACK, I. A. An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. **Soil Science**, v. 37, p. 29-38, 1934.

WATSON, S. I. Efficient design of geographically-defined clusters with spatial autocorrelation. **Journal of Applied Statistics**, p. 1-19, 2021.

## 5 Considerações Finais

Dado o fato que os resultados deste trabalho foram apresentados em forma de artigos científicos, e que as conclusões foram apresentadas em cada um deles, serão sintetizadas a seguir as principais contribuições constatadas no desenvolvimento desta tese:

- Embora pouco explorado na literatura, especialmente no âmbito agrícola, o índice de Lee se mostrou eficiente como métrica para o cálculo da correlação espacial bivariada utilizando pares de atributos do solo, o que possibilitou investigar a existência de uma relação espacial da quantidade de nutrientes do solo com a produtividade da soja. Em termos de Agricultura de Precisão e gerenciamento localizado do solo, identificar padrões e comportamentos que sugiram uma associação espacial entre dois atributos do solo viabiliza tomadas de decisão mais assertivas e eficazes.
- Quanto ao tamanho amostral efetivo bivariado, os estudos simulados e a aplicação à dados reais da metodologia do  $ESS_{bi}$ , evidenciaram sua relevância prática e também teórica. Pois, além de manter todas as características e propriedades da metodologia univariada, também fornece uma melhor viabilidade prática, por se tratar de uma proposta que reduz a quantidade de amostras de solo que vão à laboratório para duas variáveis simultaneamente, mantendo a qualidade dos mapas temáticos obtidos. Sendo assim, é possível diminuir gastos em monitoramento de solo se forem utilizados os modelos geoestatísticos bivariados para atributos químicos espacialmente correlacionados.

O estudo realizado abre um amplo leque para trabalhos futuros, dos quais pode-se destacar dois principais. O primeiro deles seria estruturar a matriz de informação de Fisher observada bivariada, utilizando para tal as derivadas calculadas nos Apêndices desta tese. Outro estudo seria generalizar o tamanho amostral efetivo para multivariado, obter a sua expressão e verificar todas as propriedades que incidem sobre o método univariado.

## 6 Apêndices

### 6.1 Apêndice A. Matriz de informação de Fisher observada bivariada

Do processo de maximização e considerando as reparametrizações (Apêndice B), obtém-se o vetor de parâmetros estimados descritos por  $\hat{\theta} = (\hat{\beta}, \hat{\sigma}^2, \hat{\eta}, \hat{\tau}_1, \hat{\tau}_2, \hat{\varphi}_0, \hat{\varphi}_1, \hat{\varphi}_2)^T$ . O cálculo da segunda derivada do logaritmo da função de verossimilhança maximizada (Equação 3.15) em relação ao parâmetro  $\hat{\theta}$ , fornece uma matriz Hessiana numérica. Sendo a matriz de informação de Fisher observada ( $\mathbf{I}$ ) para este parâmetro definida como o negativo da matriz Hessiana.

Na sequência, será apresentada a matriz de informação de Fisher observada para o parâmetro  $\hat{\theta}^* = (\hat{\eta}, \hat{\tau}_1, \hat{\tau}_2, \hat{\varphi}_0, \hat{\varphi}_1, \hat{\varphi}_2)^T$ :

$$\mathbf{I}(\hat{\theta}^*) = -\frac{\partial^2 l(\hat{\theta})}{\partial \hat{\theta}^* \partial \hat{\theta}^{*T}} = - \begin{pmatrix} \frac{\partial^2 l(\hat{\theta})}{\partial \eta^2} & \frac{\partial^2 l(\hat{\theta})}{\partial \eta \partial \tau_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \eta \partial \tau_2} & \frac{\partial^2 l(\hat{\theta})}{\partial \eta \partial \varphi_0} & \frac{\partial^2 l(\hat{\theta})}{\partial \eta \partial \varphi_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \eta \partial \varphi_2} \\ \frac{\partial^2 l(\hat{\theta})}{\partial \tau_1 \partial \eta} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_1^2} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_1 \partial \tau_2} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_1 \partial \varphi_0} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_1 \partial \varphi_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_1 \partial \varphi_2} \\ \frac{\partial^2 l(\hat{\theta})}{\partial \tau_2 \partial \eta} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_2 \partial \tau_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_2^2} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_2 \partial \varphi_0} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_2 \partial \varphi_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \tau_2 \partial \varphi_2} \\ \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_0 \partial \eta} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_0 \partial \tau_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_0 \partial \tau_2} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_0^2} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_0 \partial \varphi_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_0 \partial \varphi_2} \\ \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_1 \partial \eta} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_1 \partial \tau_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_1 \partial \tau_2} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_1 \partial \varphi_0} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_1^2} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_1 \partial \varphi_2} \\ \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_2 \partial \eta} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_2 \partial \tau_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_2 \partial \tau_2} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_2 \partial \varphi_0} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_2 \partial \varphi_1} & \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_2^2} \end{pmatrix}. \quad (6.1)$$

Já as matrizes de informação de Fisher observada para os parâmetros com forma fechada,  $\hat{\beta}$  e  $\hat{\sigma}^2$ , são dadas, respectivamente, por:

$$\mathbf{I}(\hat{\beta}) = -\frac{\partial^2 l(\hat{\theta})}{\partial \beta \partial \beta}, \quad (6.2)$$

$$\mathbf{I}(\hat{\sigma}^2) = -\frac{\partial^2 l(\hat{\theta})}{\partial \sigma^2 \partial \sigma^2}. \quad (6.3)$$

Valendo-se de resultados de diferenciação de matrizes para modelos espaciais Gaussianos (WAND, 2002)<sup>1</sup>, foram obtidas as expressões para as derivadas descritas em 6.1, 6.2 e 6.3. No que segue, considere  $\frac{\partial \Sigma_Y}{\partial \hat{\theta}^*} = \left[ \frac{\partial \sigma_{ml}}{\partial \hat{\theta}^*} \right] = \left[ \frac{\partial \sigma_{ml}}{\partial \eta}, \frac{\partial \sigma_{ml}}{\partial \tau_1}, \frac{\partial \sigma_{ml}}{\partial \tau_2}, \frac{\partial \sigma_{ml}}{\partial \varphi_0}, \frac{\partial \sigma_{ml}}{\partial \varphi_1}, \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right]$  e  $m, l = 1, \dots, n$ , em que  $\sigma_{ml}$  representa os elementos da matriz  $\Sigma_Y$ .

$$\begin{aligned} \frac{\partial^2 l(\hat{\theta})}{\partial \eta^2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] + \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \eta} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \Sigma_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \eta} \right] \Sigma_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned} \quad (6.4)$$

<sup>1</sup> WAND, M. P. Vector differential calculus in statistics. *The American Statistician*, v. 56, n. 1, p. 55-62, 2002.

sendo  $tr(\cdot)$  o operador traço.

$$\begin{aligned} \frac{\partial^2 l(\hat{\theta})}{\partial \eta \partial \tau_q} = \frac{\partial^2 l(\hat{\theta})}{\partial \tau_q \partial \eta} &= -\frac{1}{2} \left[ tr \left[ \left( -\Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] + \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \tau_q} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \tau_q} \right] \Sigma_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned} \quad (6.5)$$

com  $q = 1, 2$ .

$$\begin{aligned} \frac{\partial^2 l(\hat{\theta})}{\partial \eta \partial \varphi_k} = \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_k \partial \eta} &= -\frac{1}{2} \left[ tr \left[ \left( -\Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] + \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_k} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] \Sigma_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_k} \right] \Sigma_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned} \quad (6.6)$$

donde  $k = 0, 1, 2$ .

$$\begin{aligned} \frac{\partial^2 l(\hat{\theta})}{\partial \tau_q^2} &= -\frac{1}{2} \left[ tr \left[ \left( -\Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] + \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_q \partial \tau_q} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_q \partial \tau_q} \right] \Sigma_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned} \quad (6.7)$$

em que  $q = 1, 2$ .

$$\begin{aligned} \frac{\partial^2 l(\hat{\theta})}{\partial \tau_q \partial \tau_r} = \frac{\partial^2 l(\hat{\theta})}{\partial \tau_r \partial \tau_q} &= -\frac{1}{2} \left[ tr \left[ \left( -\Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_r} \right] \right] \right. \\ &\left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_r} \right] \Sigma_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned} \quad (6.8)$$

em que  $q, r = 1, 2$  com  $q \neq r$ .

$$\begin{aligned} \frac{\partial^2 l(\hat{\theta})}{\partial \tau_q \partial \varphi_k} = \frac{\partial^2 l(\hat{\theta})}{\partial \varphi_k \partial \tau_q} &= -\frac{1}{2} \left[ tr \left[ \left( -\Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] + \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_q \partial \varphi_k} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_q} \right] \Sigma_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] \Sigma_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \Sigma_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_q \partial \varphi_k} \right] \Sigma_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned} \quad (6.9)$$

sendo  $q = 1, 2$  e  $k = 0, 1, 2$ .



$$\begin{aligned}
\frac{\partial^2 l(\hat{\boldsymbol{\theta}})}{\partial \varphi_k^2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_k \partial \varphi_k} \right] \right] \right. \\
&+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\
&\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_k \partial \varphi_k} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \tag{6.10}
\end{aligned}$$

em que  $k = 0, 1, 2$ .

$$\begin{aligned}
\frac{\partial^2 l(\hat{\boldsymbol{\theta}})}{\partial \varphi_k \partial \varphi_p} = \frac{\partial^2 l(\hat{\boldsymbol{\theta}})}{\partial \varphi_p \partial \varphi_k} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_p} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_k \partial \varphi_p} \right] \right] \right. \\
&+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_k} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_p} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\
&\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_k \partial \varphi_p} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \tag{6.11}
\end{aligned}$$

onde  $k, p = 0, 1, 2$  e  $k \neq p$ .

Findadas as derivadas parciais referentes ao parâmetro  $\hat{\boldsymbol{\theta}}^*$ , serão apresentadas na sequência as expressões para a segunda derivada do logaritmo da função de verossimilhança em relação a  $\hat{\boldsymbol{\beta}}$  (Equação 6.2) e  $\hat{\sigma}^2$  (Equação 6.3):

$$\frac{\partial^2 l(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} = -(\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}), \tag{6.12}$$

$$\frac{\partial^2 l(\hat{\boldsymbol{\theta}})}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2(\sigma^2)^2} - \frac{(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{(\sigma^2)^3}. \tag{6.13}$$

O detalhamento de todos os processos de derivação matricial, bem como os elementos resultantes são apresentados no Apêndice B (Seção 6.2).

Por fim, a matriz de informação de Fisher esperada ( $\mathbf{K}$ ) é obtida calculando-se a esperança matemática da matriz de informação observada. Sendo assim, para os parâmetros  $\hat{\boldsymbol{\theta}}^*$ ,  $\hat{\boldsymbol{\beta}}$  e  $\hat{\sigma}^2$  tem-se as respectivas matrizes de informação de Fisher esperada, as quais assumem a forma bloco diagonal (LANGE, LITTLE; TAYLOR, 1989)<sup>2</sup>:

$$\mathbf{K}(\hat{\boldsymbol{\theta}}^*) = E[\mathbf{I}(\hat{\boldsymbol{\theta}}^*)], \quad \mathbf{K}(\hat{\boldsymbol{\beta}}) = E[\mathbf{I}(\hat{\boldsymbol{\beta}})], \quad \mathbf{K}(\hat{\sigma}^2) = E[\mathbf{I}(\hat{\sigma}^2)].$$

## 6.2 Apêndice B. Derivadas para obtenção da matriz de informação de Fisher observada bivariada

Conforme já descrito na Seção 3.4, considerando o modelo espacial Gaussiano bivariado com componente de correlação parcialmente comum, o logaritmo função de verossimilhança é dado

<sup>2</sup> 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.

por:

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \left[ \ln(|\boldsymbol{\Sigma}_Y|) + n \ln(2\pi) + (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta})^T \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \mathbf{D}\boldsymbol{\beta}) \right],$$

sendo  $\boldsymbol{\Sigma}_Y$  a matriz de covariância reparametrizada dada por:

$$\boldsymbol{\Sigma}_Y = \sigma^2 \begin{pmatrix} \mathbf{V}_1 & \eta \mathbf{V}_{1,2} \\ \eta \mathbf{V}_{1,2}^T & \eta^2 \mathbf{V}_2 \end{pmatrix},$$

em que  $\boldsymbol{\mu} = \mathbf{D}\boldsymbol{\beta}$  é o vetor de médias;  $v_1 = \rho_0(h) + \tau_1^2 \rho_1(h)$  são os elementos da matriz  $\mathbf{V}_1$ ;  $v_2 = \rho_0(h) + \tau_2^2 \rho_2(h)$  são os elementos da matriz  $\mathbf{V}_2$ ; e  $v_{1,2} = \rho_0(h)$  são os elementos da matriz  $\mathbf{V}_{1,2}$ .

Para a diferenciação do logaritmo da função de verossimilhança em relação aos parâmetros  $\boldsymbol{\theta}$ ,  $\boldsymbol{\beta}$  e  $\sigma$ , foram utilizados resultados de derivadas de matrizes e álgebra matricial. As principais propriedades empregadas estão listadas a seguir (HARVILLE, 1997)<sup>3</sup>, em que  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  representam matrizes quadradas,  $\alpha$  um parâmetro e  $tr(\cdot)$  o operador traço:

**P1.**  $\frac{\partial \ln|\mathbf{A}|}{\partial \alpha} = tr(\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha});$

**P2.**  $\frac{\partial \mathbf{A}^{-1}}{\partial \alpha} = -\mathbf{A}^{-1} \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{A}^{-1};$

**P3.**  $\frac{\partial \mathbf{A}^T \mathbf{B} \mathbf{A}}{\partial \mathbf{A}} = (\mathbf{B} + \mathbf{B}^T) \mathbf{A}$ . Se  $\mathbf{B}$  for simétrica, a última igualdade é reescrita como  $2\mathbf{B}\mathbf{A}$ ;

**P4.**  $\frac{\partial}{\partial \alpha} (\mathbf{A}\mathbf{B}) = \frac{\partial \mathbf{A}}{\partial \alpha} \mathbf{B} + \mathbf{A} \frac{\partial \mathbf{B}}{\partial \alpha};$

**P5.**  $\frac{\partial \mathbf{A}}{\partial \alpha} = \left[ \frac{\partial a_{ml}}{\partial \alpha} \right]$ , em que  $a_{ml}$  são elementos da matriz  $\mathbf{A}$ ;

**P6.**  $tr(\mathbf{A}\mathbf{B}) = tr(\mathbf{B}\mathbf{A});$

**P7.**  $tr(\mathbf{A} + \mathbf{B}) = tr(\mathbf{A}) + tr(\mathbf{B}).$

Na sequência serão exibidas detalhadamente as segundas derivadas do logaritmo da função de verossimilhança apresentadas no Apêndice A (Seção 6.1), nas Equações 6.11 a 6.13, as quais compõe a matriz de informação de Fisher observada para o parâmetro  $\boldsymbol{\theta}^* = (\eta, \tau_1, \tau_2, \varphi_0, \varphi_1, \varphi_2)^T$ :

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta^2} = \frac{\partial}{\partial \eta} \left( \frac{\partial l(\boldsymbol{\theta})}{\partial \eta} \right),$$

em que:

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \eta} = -\frac{1}{2} \left[ tr \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \right) - [(\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu})] \right], \quad (6.14)$$

sendo,

$$\left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] = \begin{pmatrix} 0 & \sigma^2 \mathbf{V}_{1,2} \\ \sigma^2 \mathbf{V}_{1,2}^T & 2\sigma^2 \eta \mathbf{V}_2 \end{pmatrix}.$$

As derivadas de primeira ordem do logaritmo da função de verossimilhança em relação aos demais parâmetros são análogas a Equação 6.14. Desta forma, para os outros parâmetros esta expressão serão ocultados.

<sup>3</sup> HARVILLE, D. A. **Matrix algebra from a statistician's perspective**. New York: Springer, 1997. 634p.

Posto isso, tem-se que:

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta^2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \eta} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

donde  $m, l = 1, \dots, n$  e

$$\left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \eta} \right] = \begin{pmatrix} 0 & 0 \\ 0 & 2\sigma^2 \mathbf{V}_2 \end{pmatrix}.$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta \partial \tau_1} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau_1 \partial \eta} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right] \right] \right. \\ &\left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right] = \begin{pmatrix} \sigma^2 \mathbf{V}'_1 & 0 \\ 0 & 0 \end{pmatrix},$$

em que cada elemento  $v'_1$  da matriz  $\mathbf{V}'_1$  é dado por  $\mathbf{V}'_1 = \left[ \frac{\partial v_1}{\partial \tau_1} \right] = 2\tau_1 \rho_1(h)$ .

Pode-se observar que, pela regra do produto (P4), a expressão para  $\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta \partial \tau_1}$  tem ainda o termo  $\left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \tau_1} \right]$ . Ocorre que este termo é uma matriz nula, pois os elementos da matriz  $\left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right]$  não estão em função do parâmetro  $\eta$ . Logo ao derivar em relação a  $\eta$ , os elementos resultantes são zero. Nos casos posteriores em que o termo proveniente da regra do produto não for exibido, deve-se subentender que se trata de uma matriz nula, pelas mesmas justificativas aqui apresentadas.

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta \partial \tau_2} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau_2 \partial \eta} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_2} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \tau_2} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_2} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \tau_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\left[ \frac{\partial \sigma_{ml}}{\partial \tau_2} \right] = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \eta^2 \mathbf{V}'_2 \end{pmatrix} \quad \text{e} \quad \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \tau_2} \right] = \begin{pmatrix} 0 & 0 \\ 0 & 2\sigma^2 \eta \mathbf{V}'_2 \end{pmatrix},$$

em que cada elemento  $v'_2$  da matriz  $\mathbf{V}'_2$  é dado por  $\mathbf{V}'_2 = \left[ \frac{\partial v_2}{\partial \tau_2} \right] = 2\tau_2 \rho_2(h)$ .

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta \partial \varphi_0} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_0 \partial \eta} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_0} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\begin{bmatrix} \frac{\partial \sigma_{ml}}{\partial \varphi_0} \end{bmatrix} = \begin{pmatrix} \sigma^2 \mathbf{V}'_1 & \sigma^2 \eta \mathbf{V}'_{1,2} \\ \sigma^2 \eta \mathbf{V}'_{1,2}{}^T & \sigma^2 \eta^2 \mathbf{V}'_2 \end{pmatrix} \quad \text{e} \quad \begin{bmatrix} \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_0} \end{bmatrix} = \begin{pmatrix} 0 & \sigma^2 \mathbf{V}'_{1,2} \\ \sigma^2 \mathbf{V}'_{1,2}{}^T & 2\sigma^2 \eta \mathbf{V}'_2 \end{pmatrix},$$

cujos elementos são dados por:

$$\begin{aligned} \mathbf{V}'_1 &= \left[ \frac{\partial v_1}{\partial \varphi_0} \right] = \frac{\partial \rho_0(h)}{\partial \varphi_0}, \\ \mathbf{V}'_{1,2} &= \mathbf{V}'_{1,2}{}^T = \left[ \frac{\partial v_{1,2}}{\partial \varphi_0} \right] = \left[ \frac{\partial v_{1,2}{}^T}{\partial \varphi_0} \right] = \frac{\partial \rho_0(h)}{\partial \varphi_0}, \\ \mathbf{V}'_2 &= \left[ \frac{\partial v_2}{\partial \varphi_0} \right] = \frac{\partial \rho_0(h)}{\partial \varphi_0}. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta \partial \varphi_1} &= \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_1 \partial \eta} = -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \right] \right. \\ &\quad \left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\begin{bmatrix} \frac{\partial \sigma_{ml}}{\partial \varphi_1} \end{bmatrix} = \begin{pmatrix} \sigma^2 \mathbf{V}'_1 & 0 \\ 0 & 0 \end{pmatrix},$$

em que cada elemento  $v'_1$  da matriz  $\mathbf{V}'_1$  é dado por  $\mathbf{V}'_1 = \left[ \frac{\partial v_1}{\partial \varphi_1} \right] = \tau_1^2 \frac{\partial \rho_1(h)}{\partial \varphi_1}$ .

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \eta \partial \varphi_2} &= \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_2 \partial \eta} = -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_2} \right] \right] \right. \\ &\quad \left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \eta} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right. \\ &\quad \left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\begin{bmatrix} \frac{\partial \sigma_{ml}}{\partial \varphi_2} \end{bmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \eta^2 \mathbf{V}'_2 \end{pmatrix} \quad \text{e} \quad \begin{bmatrix} \frac{\partial^2 \sigma_{ml}}{\partial \eta \partial \varphi_2} \end{bmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 2\sigma^2 \eta \mathbf{V}'_2 \end{pmatrix},$$

em que cada elemento  $v'_2$  da matriz  $\mathbf{V}'_2$  é dado por  $\mathbf{V}'_2 = \left[ \frac{\partial v_2}{\partial \varphi_2} \right] = \tau_2^2 \frac{\partial \rho_2(h)}{\partial \varphi_2}$ .

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau_1^2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_1 \partial \tau_1} \right] \right] \right. \\ &\quad \left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \tau_1} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right. \\ &\quad \left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_1 \partial \tau_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\begin{bmatrix} \frac{\partial^2 \sigma_{ml}}{\partial \tau_1 \partial \tau_1} \end{bmatrix} = \begin{pmatrix} \sigma^2 \mathbf{V}''_1 & 0 \\ 0 & 0 \end{pmatrix},$$

em que  $\mathbf{V}''_1 = \left[ \frac{\partial^2 v_1}{\partial \tau_1 \partial \tau_1} \right] = 2\rho_1(h)$ .



$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau_2 \partial \varphi_1} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_1 \partial \tau_2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \right] \right. \\ &\quad \left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right]. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \tau_2 \partial \varphi_2} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_2 \partial \tau_2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_2 \partial \varphi_2} \right] \right] \right. \\ &\quad + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \tau_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\quad \left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_2 \partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\left[ \frac{\partial^2 \sigma_{ml}}{\partial \tau_2 \partial \varphi_2} \right] = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \eta^2 \mathbf{V}_2'' \end{pmatrix},$$

donde  $\mathbf{V}_2'' = \left[ \frac{\partial^2 v_2}{\partial \tau_2 \partial \varphi_2} \right] = 2\tau_2 \frac{\partial \rho_2(h)}{\partial \varphi_2}$ .

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_0^2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_0 \partial \varphi_0} \right] \right] \right. \\ &\quad + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\quad \left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_0 \partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_0 \partial \varphi_0} \right] = \begin{pmatrix} \sigma^2 \mathbf{V}_1'' & \sigma^2 \eta \mathbf{V}_{1,2}'' \\ \sigma^2 \eta \mathbf{V}_{1,2}''^T & \sigma^2 \eta^2 \mathbf{V}_2'' \end{pmatrix},$$

cujos elementos são dados por:

$$\begin{aligned} \mathbf{V}_1'' &= \left[ \frac{\partial^2 v_1}{\partial \varphi_0 \partial \varphi_0} \right] = \frac{\partial^2 \rho_0(h)}{\partial \varphi_0 \partial \varphi_0}, \\ \mathbf{V}_{1,2}'' = \mathbf{V}_{1,2}''^T &= \left[ \frac{\partial^2 v_{1,2}}{\partial \varphi_0 \partial \varphi_0} \right] = \left[ \frac{\partial^2 v_{1,2}^T}{\partial \varphi_0 \partial \varphi_0} \right] = \frac{\partial^2 \rho_0(h)}{\partial \varphi_0 \partial \varphi_0}, \\ \mathbf{V}_2'' &= \left[ \frac{\partial^2 v_2}{\partial \varphi_0 \partial \varphi_0} \right] = \frac{\partial^2 \rho_0(h)}{\partial \varphi_0 \partial \varphi_0}. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_0 \partial \varphi_1} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_1 \partial \varphi_0} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \right] \right. \\ &\quad \left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right]. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_0 \partial \varphi_2} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_2 \partial \varphi_0} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \right] \right. \\ &\quad \left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_0} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right]. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_1^2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_1 \partial \varphi_1} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_1 \partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_1 \partial \varphi_1} \right] = \begin{pmatrix} \sigma^2 \mathbf{V}_1'' & 0 \\ 0 & 0 \end{pmatrix},$$

donde,

$$\mathbf{V}_1'' = \left[ \frac{\partial^2 v_1}{\partial \varphi_1 \partial \varphi_1} \right] = \tau_1^2 \frac{\partial^2 \rho_1(h)}{\partial \varphi_1 \partial \varphi_1}.$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_1 \partial \varphi_2} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_2 \partial \varphi_1} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \right] \right. \\ &\left. + 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_1} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \right]. \end{aligned}$$

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_2^2} &= -\frac{1}{2} \left[ \text{tr} \left[ \left( -\boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] + \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_2 \partial \varphi_2} \right] \right] \right. \\ &+ 2 \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) \left[ \frac{\partial \sigma_{ml}}{\partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right] \\ &\left. - \left[ (\mathbf{Y} - \boldsymbol{\mu})^T \left( \boldsymbol{\Sigma}_Y^{-1} \left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_2 \partial \varphi_2} \right] \boldsymbol{\Sigma}_Y^{-1} \right) (\mathbf{Y} - \boldsymbol{\mu}) \right] \right], \end{aligned}$$

sendo:

$$\left[ \frac{\partial^2 \sigma_{ml}}{\partial \varphi_2 \partial \varphi_2} \right] = \begin{pmatrix} 0 & 0 \\ 0 & \sigma^2 \eta^2 \mathbf{V}_2'' \end{pmatrix},$$

em que,

$$\mathbf{V}_2'' = \left[ \frac{\partial^2 v_2}{\partial \varphi_2 \partial \varphi_2} \right] = \tau_2^2 \frac{\partial^2 \rho_2(h)}{\partial \varphi_2 \partial \varphi_2}.$$

Por fim, serão apresentadas as segundas derivadas do logaritmo da função de verossimilhança que compõe as matrizes de informação de Fisher observada para os parâmetros  $\boldsymbol{\beta}$  e  $\sigma$  (Equações 6.12 e 6.13, respectivamente).

Da maximização do logaritmo da função de verossimilhança, a expressão  $(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu})$  é a única que tem relação com o vetor de médias  $\boldsymbol{\mu}$ , em que  $\boldsymbol{\mu} = \mathbf{D}\boldsymbol{\beta}$  (ver Seção 3.2.1). E mais, utilizando manipulação algébrica, tem-se a seguinte igualdade:

$$(\mathbf{Y} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}_Y^{-1} (\mathbf{Y} - \boldsymbol{\mu}) = (\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y}) - 2(\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta} + \boldsymbol{\beta}^T (\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta}.$$

Desta forma,

$$\begin{aligned}
\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\mu}} = \frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} &= -\frac{1}{2} \left( \frac{\partial [(\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{Y}) - 2(\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta} + \boldsymbol{\beta}^T (\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta}]}{\partial \boldsymbol{\beta}} \right) \\
&= -\frac{1}{2} \left[ -2(\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}) + [(\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}) + (\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})^T] \boldsymbol{\beta} \right] \\
&= -\frac{1}{2} \left[ -2(\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}) + 2(\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}) \boldsymbol{\beta} \right] \\
&= (\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}) - (\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta}.
\end{aligned} \tag{6.15}$$

Igualando a derivada  $\frac{\partial l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}}$  a zero, vem que:

$$\begin{aligned}
(\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}) - (\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta} &= 0 \\
(\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta} &= (\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}),
\end{aligned}$$

donde obtém-se a forma analítica fechada para o parâmetro  $\boldsymbol{\beta}$ :

$$\boldsymbol{\beta} = (\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})^{-1} (\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}).$$

Da Equação 6.15, tem-se que a segunda derivada do logaritmo da função de verossimilhança em relação a  $\boldsymbol{\beta}$  é dada por:

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}^2} = \frac{\partial [(\mathbf{Y}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}) - (\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D})\boldsymbol{\beta}]}{\partial \boldsymbol{\beta}} = -(\mathbf{D}^T \boldsymbol{\Sigma}_Y^{-1} \mathbf{D}).$$

Por fim, em relação ao parâmetro  $\sigma^2$ , considerando que  $\boldsymbol{\Sigma}_Y = \sigma^2 \mathbf{V}$  (Equação 3.22), pode-se reescrever a função de máxima verossimilhança como:

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \left[ \ln(|\sigma^2 \mathbf{V}|) + n \ln(2\pi) + (\mathbf{Y} - \boldsymbol{\mu})^T (\sigma^2 \mathbf{V})^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right],$$

em que  $\ln(|\sigma^2 \mathbf{V}|) = \ln((\sigma^2)^n |\mathbf{V}|)$ , e utilizando as devidas propriedades do logaritmo natural, se obtém a seguinte igualdade:

$$l(\boldsymbol{\theta}) = -\frac{1}{2} \left[ n \ln(\sigma^2) + \ln(|\mathbf{V}|) + n \ln(2\pi) + (\mathbf{Y} - \boldsymbol{\mu})^T \frac{1}{\sigma^2} \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \right],$$

para  $\sigma \neq 0$ .

E deste modo,

$$\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma^2} = -\frac{1}{2} \left[ \frac{n}{\sigma^2} + (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \left( -\frac{1}{(\sigma^2)^2} \right) \right]. \tag{6.16}$$

Igualando a derivada  $\frac{\partial l(\boldsymbol{\theta})}{\partial \sigma^2}$  a zero, tem-se que:

$$\begin{aligned}
-\frac{1}{2} \left[ \frac{n}{\sigma^2} + (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu}) \left( -\frac{1}{(\sigma^2)^2} \right) \right] &= 0 \\
\frac{\sigma^2 n - (\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{(\sigma^2)^2} &= 0,
\end{aligned}$$

donde obtém-se a forma analítica fechada para o parâmetro  $\sigma^2$ :

$$\sigma^2 = \frac{(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{n}.$$

Da Equação 6.16, tem-se que a segunda derivada do logaritmo da função de máxima verossimilhança em relação a  $\sigma^2$  é igual a:

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial (\sigma^2)^2} = \frac{n}{2(\sigma^2)^2} - \frac{(\mathbf{Y} - \boldsymbol{\mu})^T \mathbf{V}^{-1} (\mathbf{Y} - \boldsymbol{\mu})}{(\sigma^2)^3}.$$