

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

AMILTON LUCIANO GARCIA DA SILVA

**VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DA SOJA E DOS ATRIBUTOS  
DO SOLO: DIAGNÓSTICO DE INFLUÊNCIA LOCAL UTILIZANDO O MODELO  
GEOESTATÍSTICO *WAVE***

Cascavel – Paraná

2024

AMILTON LUCIANO GARCIA DA SILVA

**VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DA SOJA E DOS ATRIBUTOS  
DO SOLO: DIAGNÓSTICO DE INFLUÊNCIA LOCAL UTILIZANDO O MODELO  
GEOESTATÍSTICO *WAVE***

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, da Universidade Estadual do Oeste do Paraná (UNIOESTE) – *campus* Cascavel, em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola

Orientador: Prof. Dr. Miguel Angel Uribe Opazo

Coorientador: Prof. Dr. Gustavo Henrique Dalposso

Coorientadora: Profa. Dra. Luciana Pagliosa Carvalho Guedes

Cascavel – Paraná

2024

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Silva, Amilton Luciano Garcia da  
VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DA SOJA E DOS  
ATRIBUTOS DO SOLO: DIAGNÓSTICO DE INFLUÊNCIA LOCAL UTILIZANDO  
O MODELO GEOESTATÍSTICO WAVE / Amilton Luciano Garcia da  
Silva; orientador Miguel Angel Uribe Opazo; coorientador  
Gustavo Henrique Dalposso. -- Cascavel, 2024.  
119 p.

Dissertação (Mestrado Acadêmico Campus de Cascavel) --  
Universidade Estadual do Oeste do Paraná, Centro de Ciências  
Exatas e Tecnológicas, Programa de Pós-Graduação em  
Engenharia Agrícola, 2024.

1. Agricultura de Precisão. 2. Geoestatística. 3. Krigagem  
com drift externo. 4. Thin Plate Spline. I. Uribe Opazo,  
Miguel Angel, orient. II. Dalposso, Gustavo Henrique,  
coorient. III. Título.

Revisão de português, inglês e normas realizada por Me. Andreia Piechontcoski Uribe  
Opazo, em 08 de novembro de 2024.



Universidade Estadual do Oeste do Paraná  
Reitoria  
CNPJ 78.680.337/0001-84  
Rua Universitária, 1619, Jardim Universitário  
Tel: (45) 3220-3000 - www.unioeste.br  
CEP: 85819-110 - Cx. P.: 701  
Cascavel - PARANÁ



## AMILTON LUCIANO GARCIA DA SILVA

### VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DA SOJA E DOS ATRIBUTOS DO SOLO: DIAGNÓSTICO DE INFLUÊNCIA LOCAL UTILIZANDO O MODELO GEOESTATÍSTICO WAVE

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

Orientador - Miguel Angel Uribe Opazo

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

Elizabeth Giron Cima

Faculdade Assis Gurgacz (FAG)

Tamara Cantú Maltauro

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

Márcio Paulo de Oliveira

Universidade Tecnológica Federal do Paraná (UTFPR)

Cascavel, 23 de julho de 2024.

## BIOGRAFIA

Meu nome é Amilton Luciano Garcia da Silva, nasci em 1995, em Céu Azul, no Paraná. Desde pequeno, eu sempre gostei de números, formas, lógica e sonhava em mostrar a beleza da matemática para outras pessoas.

Ingressei na Universidade Tecnológica Federal do Paraná (UTFPR), *campus* Toledo, para cursar Licenciatura em Matemática. Lá, eu aprendi muito sobre os fundamentos e as aplicações da matemática, sobre as metodologias e as práticas de ensino. Meu trabalho de conclusão de curso foi sobre geoprocessamento e estatística espacial, duas áreas que me interessam muito e que têm muitas conexões com a matemática.

Em 2016, comecei a trabalhar como professor de matemática no setor privado e no setor público, na Secretaria de Educação do Estado do Paraná (SEED/PR).

Em 2021, concluí a Graduação em Matemática, realizando um dos meus sonhos. Agora, pretendo continuar me aperfeiçoando como professor e como matemático, buscando novos desafios e oportunidades para minha carreira.

Em 2022, dei mais um passo importante na minha formação acadêmica e profissional ao ingressar no Mestrado em Engenharia Agrícola, da Universidade Estadual do Oeste do Paraná (UNIOESTE), na linha de pesquisa de Geoprocessamento, estatística espacial e Agricultura de Precisão, estudando temas que estão relacionados com a minha graduação e com o meu trabalho de conclusão de curso.

Dedico este trabalho ao meu pai, Milton Ramiro da Silva (*in memoriam*), e a minha mãe, Maria Lucia Garcia. Obrigado pelo apoio incondicional em todos os momentos difíceis da minha trajetória acadêmica.

## **AGRADECIMENTOS**

A Deus, o meu primeiro e maior agradecimento, por me abençoar com a oportunidade de aprender e crescer durante a minha jornada acadêmica.

À minha família, o meu eterno e profundo agradecimento, por me apoiarem e me incentivarem em todos os momentos.

Ao meu orientador, Prof. Dr. Miguel Angel Uribe Opazo, e aos meus coorientadores, Prof. Dr. Gustavo Henrique Dalposso e Profa. Dra. Luciana Pagliosa Cavalho Guedes, o meu sincero e especial agradecimento, por me orientarem e me ensinarem com paciência e pela amizade desenvolvida durante este período. Agradeço, em especial, ao Prof. Dr. Miguel Angel Uribe Opazo, que foi mais que um orientador, foi um pai para mim, especialmente quando eu perdi meu pai na metade do Mestrado.

Aos meus colegas do Laboratório de Estatística Espacial (LEE) e do Laboratório de Estatística Aplicada (LEA), o meu grato e cordial agradecimento, por me ajudarem e me acompanharem na coleta dos dados em campo.

A todos os demais que fizeram parte desta etapa tão importante e feliz da minha vida, o meu carinhoso e afetuoso agradecimento. Que Deus ilumine a todos.

“Ouse coisas grandiosas”

## RESUMO

SILVA, Amilton Luciano Garcia da. **Variabilidade espacial da produtividade da soja e dos atributos do solo:** diagnóstico de influência local utilizando o Modelo Geoestatístico de *Wave* Orientador: Miguel Angel Uribe Opazo; Coorientador: Gustavo Henrique Dalposso. Coorientadora: Luciana Pagliosa Carvalho Guedes. 2024. 120f. Dissertação (Mestrado em Engenharia Agrícola) – Universidade Estadual do Oeste do Paraná, Cascavel – Paraná, 2024.

A soja é o principal produto de exportação do agronegócio brasileiro e uma das *commodities* de maior importância global. Este estudo aborda a importância da produtividade da soja e a necessidade de métodos avançados em Agricultura de Precisão (AP) para otimizar o retorno econômico e reduzir impactos ambientais. A geoestatística surge como um método estatístico essencial na AP, permitindo interpolações precisas para a construção de mapas de produtividade da soja e atributos químicos e físicos do solo, fundamentais para a tomada de decisões agrícolas. O Modelo Geoestatístico *Wave* destaca-se pela capacidade de caracterizar a dependência espacial nos dados, especialmente quando a semivariância indica o efeito *hole effect*. Neste trabalho, desenvolveu-se e aplicou-se novas técnicas de diagnóstico de influência local para dados espaciais, especificamente ao utilizar o Modelo Geoestatístico *Wave*. Essas técnicas permitiram a identificação de observações influentes, que podem distorcer a semivariância e comprometer a precisão da modelagem espacial. Por meio dessas técnicas, foi possível refinar as estimativas dos parâmetros espaciais e melhorar a qualidade dos mapas interpolados, fundamentais para a definição de zonas de manejo mais precisas, otimizando a aplicação de insumos agrícolas e maximizando ganhos econômicos (Artigo 1). Com base nos resultados, observou-se que o Modelo de *Wave* se destaca por sua capacidade de capturar padrões de dependência espacial mais complexos, como os relacionados a solos heterogêneos, que são frequentemente observados em áreas agrícolas. Além disso, a combinação entre o Método de Interpolação *Thin Plate Spline* (TPS) e a *krigagem* com *drift* externo mostraram-se como estratégias eficazes para melhorar as estimativas em locais não amostrados, aproveitando a informação de covariáveis consideradas fixas (Artigo 2). Ao utilizar essas técnicas, os produtores podem identificar áreas com maior potencial produtivo e aquelas que necessitam de intervenções específicas. Esse tipo de abordagem é fundamental para a otimização do uso de insumos, reduzindo desperdícios, minimizando impactos ambientais e maximizando o retorno econômico.

**Palavras-chave**, Agricultura de Precisão, Geoestatística, *Krigagem* com *drift* Externo, Máxima Verossimilhança, *Thin Plate Spline*.

## ABSTRACT

SILVA, Amilton Luciano Garcia da. **Spatial variability of soybean productivity and soil attributes:** local influence diagnosis using the Wave geostatistical model. Advisor: Miguel Angel Uribe Opazo; Co-advisor: Gustavo Henrique Dalposso. Co-advisor: Luciana Pagliosa Carvalho Guedes. 2024. 120p. Master's Thesis. (Master's in Agricultural Engineering) – Western Paraná State University, Cascavel – Paraná, 2024

Soybeans constitute Brazil's primary agribusiness export and rank among the world's most significant commodities. This study highlights soybean productivity and the growing need for advanced precision agriculture (PA) techniques to maximize economic returns while minimizing environmental impact. Geostatistics, as a core statistical approach in PA, enables accurate interpolations for mapping soybean productivity and soil chemical and physical attributes, which are critical for informed agricultural decision-making. The Wave geostatistical model is emphasized for its unique capacity to depict spatial dependence, particularly when semivariance suggests the presence of a hole effect. In this research, novel diagnostic techniques for local influence on spatial data were developed and tested specifically with the Wave model, allowing the identification of influential observations that may distort semivariance and compromise spatial model accuracy. These techniques refined spatial parameter estimates and enhanced the quality of interpolated maps, which are essential for delineating more precise management zones, optimizing input applications, and boosting economic returns (Article 1). Results demonstrated that the Wave model effectively captures complex spatial dependence patterns, notably those arising from heterogeneous soils commonly found in agricultural settings. Additionally, combining Thin Plate Spline (TPS) interpolation with kriging with external drift proved to be a robust approach to improving estimates in unsampled areas by leveraging data from fixed covariates (Article 2). Through these techniques, producers can identify zones with high productive potential and others needing targeted intervention. This approach is vital for optimizing input usage, reducing waste, lessening environmental impact, and enhancing economic profitability.

**Keywords:** External Drift Kriging, Geostatistics, Maximum Likelihood, Precision Agriculture, Thin Plate Spline.

## LISTA DE TABELAS

Tabela1 Limites de classes da RSP em MPa e graus de limitação ao crescimento das raízes. ....	8
Tabela 2 Classificação dos teores de nutrientes para o solo classificados como Latossolo Vermelho Distoférrico típico de textura argilosa.....	9
Tabela 3 Classificação do Índice de Dependência Espacial – SDI. ....	35

### Artigo 1

Tabela 1 Classificação do Índice de Dependência Espacial – SDI. ....	55
Tabela 2 Análise descritiva das variáveis resposta. ....	56
Tabela 3 Parâmetros estimados por máxima verossimilhança e erro padrão assintótico entre parênteses. ....	59
Tabela 4 Índice Kappa (K) e Kappa Ponderado (Kw) para a comparação entre os mapas obtidos com todos as observações e os mapas obtidos excluindo-se as observações influentes.....	63

### Artigo 2

Tabela 1 Análise descritiva das variáveis resposta e explicativas.....	81
Tabela 2 Parâmetros estimados por Máxima verossimilhança-ML do modelo linear espacial considerando a função de covariância Wave e erros padrão assintóticos entre parênteses.....	84
Tabela 3 Parâmetros estimados por ML do modelo linear espacial considerando a função de covariância Wave sem a observação #74 considerada influente e erros padrão assintóticos entre parênteses.....	86

## LISTA DE FIGURAS

Figura 1 Exemplo de modelo espacial ajustado ao semivariograma.....	18
Figura 2 Gráficos da função semivariância $\gamma(h)$ para os modelos: (a) exponencial, (b) gaussiano, (c) Matérn com parâmetro de suavização $k=1$ , (d) Matérn com parâmetro $k=2$ , (e) Matérn com parâmetro $k=3$ e (f) wave.xemplo de modelo espacial ajustado ao semivariograma.....	21

### Artigo 1

Figura 1 Localização da área agrícola com 102 pontos amostrais.....	51
Figura 2 Postplot (a) produtividade da soja ( $\text{Prod} [\text{t ha}^{-1}]$ ), (b) teor de cobre do solo ( $\text{Cu} [\text{mg dm}^3]$ ), (c) pH do solo (pH) e (d) resistência do solo à penetração na camada de 0.21 a 0.30 m de profundidade ( $\text{RSP}_{0,21-0,30\text{m}} [\text{MPa}]$ ) .....	56
Figura 3 Semivariância experimental (a) produtividade da soja, (b) Cobre, (c) pH, (d) resistência do solo à penetração na camada de 0,21 a 0,30 metros de profundidade.	58
Figura 4 Gráficos de $C_i$ e $ L_{\text{max}} $ em função da ordem das observações coletadas. (a) $C_i$ para a produtividade da soja; (b) $ L_{\text{max}} $ para a produtividade da soja; (c) $C_i$ .....	60
Figura 5 Mapa temático: (a) produtividade da soja com todas as observações; (b) produtividade da soja sem observação influente #85; (c) teor de cobre com todas as observações; (d) teor de cobre sem observação influente #36; (e) pH com todas as observações; (f) pH sem observação influente #99; (g) resistência do solo à penetração na camada de 0,21 a 0,30 metros de profundidade com todas as observações; (h) resistência do solo à penetração na camada de 0,21 a 0,30 metros de profundidade sem observação influente #46.....	62

### Artigo 2

Figura 1 Localização da área monitorada e o posicionamento dos 102 pontos amostrais.....	72
Figura 2 Gráficos boxplot para as variáveis em estudo: (a) produtividade da soja; (b) teor de cálcio; (c) teor de cobre; (d) teor de potássio; (e) teor de fósforo; (f) pH do solo e (g) resistência do solo à penetração na camada 0.21 a 0.30m de profundidade. Os	

valores destacados nos boxplots indicam os pontos amostrais identificados como outliers .....	81
Figura 3 Gráfico postplot da produtividade da soja.....	82
Figura 4 Semivariograma experimental da produtividade da soja para o ano-safra 2022/2023.....	83
Figura 5 Gráficos (a) $C_i$ e (b) $ L_{max} $ em função da ordem das observações $i$ para a identificação de pontos influentes.....	85
Figura 6 Mapa temático da: (a) produtividade da soja com todas as observações e (b) sem a observação #74 considerada influente.....	87

## Sumário

<b>1 INTRODUÇÃO</b> .....	<b>1</b>
<b>2 OBJETIVOS</b> .....	<b>4</b>
2.1 OBJETIVO GERAL .....	4
2.2 OBJETIVOS ESPECÍFICOS .....	4
<b>3 REVISÃO BIBLIOGRÁFICA</b> .....	<b>5</b>
3.1 SOJA.....	5
<b>3.1.1 Produtividade da soja na Mesorregião Oeste do Paraná</b> .....	<b>6</b>
3.2 ATRIBUTOS FÍSICOS DO SOLO .....	7
<b>3.2. 1 Resistência do Solo à Penetração (RSP)</b> .....	<b>7</b>
3.3 ATRIBUTOS QUÍMICOS DO SOLO .....	8
<b>3.3.1 Teor de Cálcio (Ca)</b> .....	<b>9</b>
<b>3.3.2 Teor de Cobre (Cu)</b> .....	<b>9</b>
<b>3.3.3 Teor de Fósforo (P)</b> .....	<b>10</b>
<b>3.3.4 Potencial Hidrogeniônico (pH)</b> .....	<b>10</b>
<b>3.3.5 Teor de Potássio (K)</b> .....	<b>11</b>
3.4 AGRICULTURA DE PRECISÃO – AP .....	12
3.5 GEOESTATÍSTICA .....	13
<b>3.5.1 Modelo Espacial Linear Gaussiano</b> .....	<b>14</b>
<b>3.5.2 Semivariograma</b> .....	<b>16</b>
<b>3.5.3 Isotropia</b> .....	<b>19</b>
<b>3.5.4 Modelos Espaciais</b> .....	<b>19</b>
<b>3.5.5 Estimação de parâmetros por Máxima Verossimilhança (ML)</b> .....	<b>22</b>
3.5.5.1 <i>Matriz de Informação Observada de Fisher</i> .....	23
3.5.5.2 <i>Matriz de Informação Esperada de Fisher</i> .....	24
<b>3.5.6 Krigagem Ordinária</b> .....	<b>25</b>
<b>3.5.7 Krigagem Com Drift Externo</b> .....	<b>29</b>
<b>3.5.8 Critérios de Seleção de Modelos</b> .....	<b>32</b>
3.5.8.1 <i>Validação Cruzada</i> .....	32
3.5.8.2 <i>Critérios de Informação de Akaike e Bayesiano de Schwarz</i> .....	33
3.6 ÍNDICE DE DEPENDÊNCIA ESPACIAL – SDI.....	34
<b>3.6.1 Diagnóstico de Influência Local</b> .....	<b>35</b>
3.6.1.1 <i>Perturbação de Zhu generalizada</i> .....	37

3.7 COMPARAÇÃO DE MAPAS .....	38
<b>4 REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>39</b>
<b>5 ARTIGOS .....</b>	<b>48</b>
5.1 ARTIGO 1 - VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DA SOJA E ATRIBUTOS DO SOLO: “ <i>HOLE EFFECT</i> ” E DIAGNÓSTICO LOCAL UTILIZANDO O MODELO DE <i>WAVE</i> .....	48
<b>5.1.1 Introdução .....</b>	<b>48</b>
<b>5.1.2 Material e métodos .....</b>	<b>50</b>
5.1.2.1 <i>Área de estudos de dados</i> .....	50
5.1.2.2 <i>Análise Geoestatística</i> .....	51
5.1.2.3 <i>Máxima Verossimilhança (ML)</i> .....	53
5.1.2.4 <i>Influência Local</i> .....	54
5.1.2.5 <i>Comparações</i> .....	54
5.1.2.6 <i>Recurso Computacionais</i> .....	55
<b>5.1.3 Resultados e Discussões .....</b>	<b>55</b>
5.1.3.1 <i>Análise Exploratória</i> .....	55
5.1.3.2 <i>Análise Espacial</i> .....	57
5.1.3.3 <i>Diagnóstico de Influência Local</i> .....	59
5.1.3.4 <i>Mapa Geoestatístico</i> .....	61
5.1.3.5 <i>Comparação de Mapas</i> .....	63
<b>5.1.4 Conclusões .....</b>	<b>64</b>
<b>5.1.5 Agradecimentos .....</b>	<b>64</b>
<b>REFERÊNCIAS.....</b>	<b>64</b>
5.2 ARTIGO 2 – ANÁLISE DA DEPENDÊNCIA ESPACIAL DA PRODUTIVIDADE DA SOJA UTILIZANDO COVARIÁVEIS E O MODELO ESPACIAL LINEAR GAUSSIA	69
<b>5.2.1 Introdução.....</b>	<b>69</b>
<b>5.2.2 Material e Métodos .....</b>	<b>72</b>
5.2.2.1 <i>Área de estudo</i> .....	72
5.2.2.2 <i>Análise Exploratória</i> .....	73
5.2.2.3 <i>Análise Geoestatística</i> .....	73
5.2.2.4 <i>Máxima Verossimilhança (ML)</i> .....	75
5.2.2.5 <i>Thin Plate Spline – TPS</i> .....	76
5.2.2.6 <i>Krigagem com Drift Externo</i> .....	78
5.2.2.7 <i>Influência Local</i> .....	78

5.2.2.8	<i>Comparações</i>	79
5.2.2.9	<i>Recursos Computacionais</i>	80
<b>5.2.3</b>	<b>Resultados e Discussões</b>	<b>80</b>
5.2.3.1	<i>Análise Exploratória</i>	80
5.2.3.2	<i>Análise Espacial</i>	82
5.2.3.3	<i>Mapa Geoestatístico</i>	84
5.2.3.4	<i>Diagnóstico de Influência Local</i>	84
5.2.3.5	<i>Comparação de Mapas</i>	87
5.2.3.6	<i>Modelo de Predição</i>	87
<b>5.2.4</b>	<b>Conclusões</b>	<b>88</b>
<b>5.2.5</b>	<b>Agradecimentos</b>	<b>89</b>
<b>5.2.6</b>	<b>Referências</b>	<b>89</b>
<b>6</b>	<b>CONSIDERAÇÕES FINAIS</b>	<b>96</b>
<b>APÊNDICE A</b>		<b>97</b>
<b>APÊNDICE B</b>		<b>99</b>

## 1 INTRODUÇÃO

O agronegócio é um dos principais setores da economia brasileira, sendo a soja o seu principal produto de exportação. Segundo o IBGE (2023), a produção nacional de soja deve atingir 147,5 milhões de toneladas em 2023, um aumento de 23,4% em relação a 2022. Diante da relevância dessa cultura, torna-se essencial a aplicação de métodos que aprimorem o entendimento da variabilidade espacial da produtividade da soja e da sua interação com o solo e o manejo.

A Agricultura de Precisão (AP) é uma abordagem inovadora que visa otimizar o manejo das culturas agrícolas, levando em consideração a variabilidade espacial dos fatores que afetam a produção. Essa abordagem traz diversos benefícios potenciais, tais como: melhoria da qualidade das colheitas, redução do impacto ambiental, aumento da sustentabilidade, garantia da segurança alimentar, promoção do desenvolvimento econômico e elevação da rentabilidade e da produtividade (ZAIN *et al.*, 2024).

Para implementar a AP, é necessário utilizar métodos que permitam descrever e modelar a variabilidade espacial dos fenômenos naturais, bem como estimar valores em locais não amostrados. A geoestatística é um dos principais métodos científicos para esse fim, pois oferece adaptações de técnicas clássicas da estatística para aproveitar a estrutura espacial dos dados (URIBE-OPAZO *et al.*, 2023).

A estrutura de dependência espacial descreve como os elementos de uma população se distribuem em um determinado espaço, por exemplo, de forma aleatória ou agrupada. A partir disso, os métodos geoestatísticos vêm sendo aproveitados para execuções de projetos e em análises do comportamento de nutrientes em geral, até mesmo do clima, que é um fator influente e importante quando se fala de cultivo (ALMEIDA *et al.*, 2011).

Na modelagem da estrutura de dependência espacial entre as amostras, é comum a utilização da função semivariância de Matheron (MATHERON, 1963) ajustando-se, tradicionalmente, os modelos Esférico, Exponencial, Gaussiano e família Matérn com diferentes parâmetros de suavização  $k > 0$ . Estes modelos caracterizam-se por apresentar em seu semivariograma um crescimento contínuo ao longo das distâncias em estudo (CRESSIE, 2015).

No entanto, em alguns casos, o semivariograma pode não apresentar um crescimento contínuo ao longo das distâncias em estudo, ou seja, não aumenta

continuamente com o aumento da distância, indicando que há valores altos cercados por valores baixos, gerando um movimento senoidal da função semivariância, sendo este efeito chamado de *hole effect*. Para esses casos, o Modelo de *Wave* é o mais indicado, pois consegue lidar de forma mais eficaz com a variação não monotônica do semivariograma (CARVALHO *et al.*, 2009; DALPOSSO; URIBE-OPAZO; OLIVEIRA, 2022).

O estudo da dependência espacial requer a consideração deste tipo de estrutura, que é essencial para a sua caracterização. A negligência dessa estrutura pode levar a interpretações equivocadas que não refletem os padrões de dependência espacial existentes. Portanto, é fundamental reconhecer que essas estruturas são fontes de informações valiosas sobre a variabilidade espacial do fenômeno (PYRCZ; DEUTSCH, 2003; DALPOSSO; URIBE-OPAZO; OLIVEIRA, 2022).

Além disso, a presença de observações influentes altera os resultados das análises de dependência espacial e, conseqüentemente, a construção dos mapas gerados pela *krigagem* (URIBE-OPAZO *et al.*, 2023).

Para avaliar se as observações podem causar distorções nos modelos de dependência espacial, Cook (1986) propôs causar pequenas perturbações no modelo ou nos dados. Esta técnica é conhecida na literatura como influência local que leva em consideração a diferenciação geométrica, e o afastamento da verossimilhança, que é empregado como procedimento para avaliar a influência das observações (URIBE-OPAZO; BORSSOI; GALEA 2012; DE BASTIANI *et al.*, 2015).

A Agricultura de Precisão (AP) tem se beneficiado de inúmeros estudos sobre diagnóstico de influência local, como os de Uribe-Opazo, Borssoi e Galea (2012), De Bastiani *et al.* (2015), Uribe-Opazo *et al.* (2021), Dalposso, Uribe-Opazo e De Bastiani (2021) e Uribe-Opazo *et al.* (2023) que contribuíram significativamente para o entendimento da influência local na modelagem de dados espaciais. Esses estudos forneceram subsídios importantes para a identificação de observações influentes, aprimorando a precisão dos modelos geoestatísticos tradicionais.

Entretanto, apesar desses avanços, ainda não há trabalhos que explorem o diagnóstico de influência local no contexto do modelo geoestatístico *Wave*, que tem se mostrado promissor na captura de padrões mais complexos de dependência espacial, como o efeito *hole effect*. O *Wave* é um modelo capaz de descrever estruturas espaciais não capturadas adequadamente por modelos tradicionais, sendo

essencial para a análise de variabilidade espacial em dados com comportamentos oscilatórios.

A ausência de estudos que apliquem o diagnóstico de influência local a esse modelo representa uma lacuna importante, visto que o desempenho do Modelo de *Wave* pode ser comprometido por observações influentes. Assim, este trabalho traz uma contribuição inédita, ao desenvolver e aplicar métodos de diagnóstico de influência local para o Modelo de *Wave*, permitindo a identificação e a remoção de dados que podem distorcer a dependência espacial.

Nesse sentido, a Seção 2 apresenta o objetivo geral e os objetivos específicos deste trabalho. Na Seção 3, é realizada uma revisão bibliográfica abrangente, com o propósito de estabelecer a fundamentação teórica necessária para o desenvolvimento dos artigos que compõem esta dissertação. Na Seção 4, são apresentadas as referências utilizadas. Por último, na Seção 5 são expostos os artigos que constituem os resultados da pesquisa realizada.

## 2 OBJETIVOS

### 2.1 OBJETIVO GERAL

Desenvolver e aplicar técnicas de diagnóstico de influência local ao empregar, no ajuste da função semivariância, o Modelo Geoestatístico de Wave e compará-lo com modelos tradicionais, como os Modelos Exponencial, Gaussiano e Família, mantendo os parâmetros de suavização,  $k=0,7$ ,  $k=1$  e  $k=2$ , na descrição da variabilidade espacial da produtividade da soja e dos atributos do solo.

### 2.2 OBJETIVOS ESPECÍFICOS

- a) Estudar a dependência espacial dos atributos do solo e da produtividade da soja do ano-safra 2022/2023;
- b) Desenvolver técnicas de diagnósticos de influência local na variável resposta utilizando o Modelo Geoestatístico *Wave*;
- c) Aplicar as técnicas de diagnóstico de influência local desenvolvidas para detectar observações influentes;
- d) Aplicar a técnica de interpolação *Thin Plate Spline* (TPS) para estimar o valor das covariáveis em locais não amostrados, consideradas como *drift*, para a *Krigagem* com *drift* externo da produtividade da soja;
- e) Construir mapas temáticos utilizando os métodos geoestatísticos do Modelo de *Wave*.

### 3 REVISÃO BIBLIOGRÁFICA

#### 3.1 SOJA

A soja (*Glycine max (L.) Merrill*) teve sua origem e processo de domesticação no nordeste da China há mais de 5.000 anos. O cultivo da cultura foi iniciado pelo imperador chinês, Shen-nung, que a utilizou como fonte de alimento, oferecendo uma alternativa para a redução do abate de animais. Em outras partes do mundo, a adaptação da soja ocorreu inicialmente locais com similaridade ao ambiente de origem, tendo início na Europa em 1712. Nos Estados Unidos, o primeiro relato sobre seu plantio ocorreu em 1765, contudo, a adaptação e expansão da cultura ocorreu de forma lenta (GAZZONI, 2018).

Com o passar do tempo, percebeu-se a importância de investir em pesquisas sobre a leguminosa, culminando no desenvolvimento tecnológico e no surgimento de cultivares de soja mais produtivos e resistentes a doenças e pragas, além de um sistema adequado para as diferentes condições de produção na América do Norte (DALL'AGNOL, 2016; GAZZONI, 2018).

No contexto brasileiro, a primeira tentativa de cultivo, em 1882, fracassou, uma vez que o material genético utilizado não era adequado para as condições climáticas do estado da Bahia. Posteriormente, em 1891, foram realizados testes com cultivares de soja no Instituto Agrônomo de Campinas, no estado de São Paulo, no entanto, sua produção comercial só se consolidou efetivamente na região do Rio Grande do Sul, na década de 1920 (DALL'AGNOL, 2016; GAZZONI, 2018).

A soja é uma planta herbácea, da classe *Dicotyledoneae*, ordem *Rosales*, família *Leguminosa* e, subfamília das *Papilionoideae*, gênero *Glycine L.* Seu desenvolvimento pode ser dividido em cinco fases: estabelecimento, vegetativo, florescimento, enchimento dos grãos e colheita. Em cada fase, a planta apresenta diferentes demandas de fatores ambientais, como concentração de gás carbônico, radiação solar, disponibilidade de água e nutrientes (BOOTE; JONES. HOOGENBOOM, 1998).

De acordo com os dados da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA, 2023), o Brasil é o segundo maior produtor de soja do mundo, com uma produção de 154,566 milhões de toneladas no ano-safra 2022/23, ocupando uma área

de 44,062 milhões de hectares, movimentando, dessa forma, bilhões de dólares por ano no mercado internacional.

Atualmente, é a cultura mais importante economicamente, tanto para o Brasil quanto para o estado do Paraná, pois, segundo o Instituto Brasileiro de Geografia e Estatística (2023), a produção de soja no estado pode chegar a 47,12 milhões de toneladas aproximadamente enquanto, a nível de Brasil, sua estimativa é de a 147,5 milhões de toneladas.

A cultura da soja também possui grande relevância social, uma vez que, segundo Dall'agnol (2016), seus grãos são utilizados tanto na alimentação humana quanto na de animais devido ao seu alto nível de proteína. Além do mais, seu óleo dispõe de diversas aplicações na indústria e na culinária, além de serem utilizados no processo de geração de energia.

Diante do apresentado, torna-se fundamental a compreensão do comportamento da soja sob diferentes condições a fim de que ocorra aumento de sua eficiência e sustentabilidade da produção nacional.

### **3.1.1 Produtividade da soja na Mesorregião Oeste do Paraná**

O estado do Paraná é o segundo maior produtor de soja no território brasileiro, sendo superado apenas pelo estado do Mato Grosso. Dentro do estado, a mesorregião oeste se destaca como um importante celeiro agrícola, apresentando-se como uma das principais áreas de semeadura antecipada em comparação com outras regiões, conforme apresentado pela Agência Estadual de Notícias (AEN, 2021) e pelo Departamento de Economia Rural Boletim Semanal (DERAL, 2022).

De acordo com a previsão divulgada pela Secretaria de Estado da Agricultura e do Abastecimento (SEAB, 2023), estima-se que a produção de soja no estado do Paraná alcance um recorde histórico de 22,45 milhões de toneladas em uma área de 5,76 milhões de hectares. Este volume representa um aumento significativo de 82% em relação ao ciclo anterior (2021/2022), tendo em vista que a produtividade média atual desta safra é de 3,845 t ha<sup>-1</sup> (AEN, 2023).

Ainda de acordo com a SEAB (2023), a cultura da soja ocupou 25% do território estadual sendo um produto de expressiva relevância no mercado internacional. Sua exportação é destinada a mais de 20 países, tendo atingido o valor de U\$\$ 6 bilhões, correspondendo a 17% do total exportado pelo Brasil (AEN, 2023).

Do ponto de vista social, a produção de soja na mesorregião oeste contribui para o desenvolvimento local, gerando renda e oportunidades de trabalho direto e indireto para os habitantes da região. Além disso, a cultura da soja tem um papel importante na sustentabilidade da agricultura familiar, representando uma participação de 37,4% da produção de soja da região (NUNES; MORAES; ROSSINI, 2020).

## 3.2 ATRIBUTOS FÍSICOS DO SOLO

### 3.2. 1 Resistência do Solo à Penetração (RSP)

O cultivo agrícola tem se beneficiado de tecnologias modernas e mecanismos avançados, com o objetivo de obter melhores e maiores colheitas, garantindo maior eficiência e sustentabilidade. No entanto, o uso intensivo de máquinas agrícolas e equipamentos pesados, associados a práticas inadequadas de manejo do solo, podem provocar a compactação do solo, que resulta em perda de produtividade, pois dificulta a penetração das raízes e compromete a disponibilidade de água e nutrientes essenciais para o desenvolvimento das plantas (LIMA *et al.*, 2010).

A compactação do solo é resultado da pressão aplicada ao solo, que provoca o aumento da densidade aparente ou a diminuição da porosidade do solo, sendo medida pela Resistência do Solo à Penetração (RSP), que é a força necessária para penetrar o solo (KUMI *et al.*, 2023). Essa resistência é influenciada pelo tamanho e peso das máquinas agrícolas, bem como pela umidade do solo, que torna o solo mais propenso à compactação, além de reduzir a capacidade de infiltração da água e o movimento de gases essenciais para a respiração das raízes (KELLER *et al.*, 2019; VANDERHASSELT *et al.*, 2023).

De acordo com De Bastiani, Uribe-Opazo e Dalposso (2012), a RSP é um indicador relevante da compactação do solo, pois afeta diretamente o crescimento das plantas. Colares *et al.* (2008) observaram que a presença de zonas compactadas no solo pode limitar o desenvolvimento da área foliar e do sistema radicular em um determinado ponto, fazendo com que as raízes busquem outras direções de crescimento.

Assim, Canarache (1990) determinou os limites da RSP e os graus de limitação ao crescimento das raízes, conforme a Tabela 1.

Tabela1: Limites de classes da RSP em MPa e graus de limitação ao crescimento das raízes.

Classificação da RSP	Limites (MPa)	Limites ao crescimento das raízes
Muito baixa	< 1,1	Sem limitação
Baixa	1,1 – 2,5	Pouca limitação
Média	2,6 – 5,0	Algumas limitações
Alta	5,1 – 10,0	Sérias limitações
Muito alta	10,1 – 15,0	Raízes praticamente não crescem
Extremamente alta	> 15,0	Raízes não crescem

Fonte: Canarache (1990).

### 3.3 ATRIBUTOS QUÍMICOS DO SOLO

A variabilidade espacial dos atributos químicos do solo é um aspecto importante para a AP, pois permite o delineamento de zonas de aplicação que visam a conservação do solo e a aplicação futura de insumos em taxa variada. Nesse sentido, Dalchiavon *et al.* (2012) afirmam que o conhecimento da distribuição espacial dos atributos químicos do solo é fundamental para o planejamento e a tomada de decisão na AP.

Abreu, Lopes e Santos (2007) destacam que, no Sistema de Plantio Direto (SPD), os macronutrientes primários, como o fósforo (P) e o potássio (K), são requeridos em maior quantidade pelas plantas, enquanto os macronutrientes secundários, como o cálcio (Ca) e o magnésio (Mg), são essenciais para o crescimento vegetal. Os micronutrientes, cobre (Cu) e zinco (Zn), se caracterizam por apresentarem teores inferiores no solo.

Portanto, o monitoramento da fertilidade do solo é uma prática indispensável para o manejo adequado da cultura e do solo no SPD, pois permite identificar deficiências nutricionais, planejar a reposição de nutrientes de forma precisa e minimizar os impactos ambientais associados ao uso excessivo de fertilizantes.

Neste estudo, foram avaliados os atributos químicos do solo teor cálcio (Ca), teor cobre (Cu), teor potássio (K), teor fósforo (P) e acidez do solo por meio do Potencial Hidrogeniônico (pH). Esses dados foram selecionados devido à sua importância para o desenvolvimento da cultura da soja.

Os teores desses atributos foram classificados segundo os critérios de Santos e Silva (2001), conforme a Tabela 2.

Tabela 2: Classificação dos teores de nutrientes para o solo classificado como Latossolo Vermelho Distoférrico *típico* de textura argilosa.

Atributos	Muito Baixo	Baixo	Médio	Alto	Muito Alto
Ca (cmolc/dm <sup>3</sup> )	-	< 2,00	2,00 – 4,00	> 4,00	-
Cu (mg/dm <sup>3</sup> )	-	< 0,80	0,80 - 1,70	> 1,70	-
K (cmolc/dm <sup>3</sup> )	-	≤ 0,10	0,11 – 0,20	0,21 – 0,30	> 0,30
P (mg/dm <sup>3</sup> )	-	≤ 3,00	3,01 – 6,00	6,01 – 9,00	> 9,00
pH	≤ 4,30	4,31 – 5,00	5,01 – 5,50	5,51 – 6,00	> 6,00

Ca: Cálcio; Cu: Cobre; K: Potássio; Fósforo (P); pH: Potencial Hidrogeniônico.

Fonte: Santos & Silva (2001).

### 3.3.1 Teor de Cálcio (Ca)

O teor de cálcio é um elemento essencial para a manutenção da integridade estrutural e funcional das plantas, pois participa de diversos processos fisiológicos e bioquímicos. A deficiência de Ca nas plantas provoca a redução dos tecidos meristemáticos e prejuízos no crescimento, além de causar deformações nas folhas novas (MALAVOLTA; VITTI; OLIVEIRA, 1997)

Na cultura da soja, a deficiência de Ca atrofia as raízes e retarda a emergência das primeiras folhas, que quando surgem apresentam enrugamento e colapso do pecíolo (SEDIYAMA, 2009; LANGE *et al.*, 2021). A deficiência de Ca e a toxicidade por alumínio têm sido apontadas como as principais barreiras químicas ao crescimento de raízes em subsolos ácidos (GRZEGOZEWSKI, 2012).

A calagem é a principal forma de fornecimento de Ca às lavouras, pois além de aumentar a disponibilidade desse nutriente no solo, também neutraliza o H<sup>+</sup>, diminui a toxicidade de alumínio e manganês e proporciona condições favoráveis ao crescimento das raízes (SOUSA & LOBATO, 2004; FIRMANO *et al.*, 2017).

### 3.3.2 Teor de Cobre (Cu)

O teor de cobre (Cu) é um micronutriente essencial para o crescimento e desenvolvimento das plantas, pois participa de diversos processos metabólicos, como a fotossíntese, a respiração, a síntese de lignina e a defesa contra estresses bióticos e abióticos (CANCIAN, 2018).

A deficiência de Cu pode causar redução na produtividade, no vigor vegetativo, na tolerância a pragas e doenças e na qualidade do produto colhido (FANCELLI, 2003). Pode ainda, ocorrer em solos com baixo teor de matéria orgânica, alto pH, alta disponibilidade de fósforo (P) ou ferro (Fe) e alta umidade. A correção da deficiência de Cu pode ser feita por meio da aplicação de fertilizantes contendo Cu, como sulfato de cobre, quelatos de cobre ou óxido de cobre, na forma de adubação foliar ou no solo (NOBILE & ANUNCIÃO, 2021).

### **3.3.3 Teor de Fósforo (P)**

O teor de fósforo (P), um dos três nutrientes primários, é indispensável para o crescimento das plantas e não pode ser substituído por nenhum outro elemento (LOPES, 1998).

A principal função do P é no armazenamento e utilização de energia, impulsionando o crescimento das raízes, a maturação dos frutos, e proporcionando uma melhor formação de grãos e frutos, além de conferir maior vigor às plantas (VILLAR, 2007). O P também é crucial na fotossíntese, respiração, divisão e crescimento celular (LOPES, 1998; MENDES, 2007).

A deficiência de fósforo se manifesta através de um desenvolvimento subnormal da planta (LOPES, 2007). As folhas mais velhas são afetadas antes das novas, apresentando uma coloração verde azulada. Além disso, há um atraso no crescimento e maturação, e os frutos podem se deformar ou apresentar grãos vazios (LOPES, 2007; VILLAR, 2007).

Estima-se que as plantas utilizem apenas 20 a 40% do P disponível, com 80 a 95% sendo fixado no solo (OLIVEIRA, 2007). Ainda, P é absorvido na forma aniônica, formando uma ligação covalente forte com o oxigênio, que se mantém mesmo após sua incorporação nos tecidos vegetais (MENDES, 2007). Para solos argilosos, especialmente na produção de soja, o teor médio de P varia entre 3,1 e 6,0 g/dm<sup>3</sup> (OLIVEIRA, 2007).

### **3.3.4 Potencial Hidrogeniônico (pH)**

De acordo com Malavolta, Vitti e Oliveira (1997), o pH (indicador de acidez do solo) é possivelmente o fator mais relevante no que diz respeito à utilização de

fertilizantes. Isso ocorre porque valores mais altos de pH auxiliam na disponibilidade e aproveitamento dos nutrientes essenciais para as plantas. Fagundes *et al.* (2018) ressalta que a faixa ideal de pH para a maioria das culturas agrícolas está entre 6 e 7.

A acidificação do solo, ainda que um fenômeno natural, é intensificada nos ambientes agrícolas modernos, especialmente pela aplicação de fertilizantes nitrogenados. Para combater esse desafio, a calagem se destaca como uma prática amplamente adotada. Ao aumentar o pH e os cátions básicos, ela não apenas corrige a acidez, mas também reduz a toxicidade causada pelo alumínio trocável, melhorando as condições para o crescimento das plantas (ZIGLIO; MIYAZAWA; PAVAN, 1999).

Segundo a Emater (1979), os solos são classificados em diferentes graus de acidez e alcalinidade: extremamente ácidos com pH igual ou inferior a 4,3; fortemente ácidos entre 4,4 e 5,3; moderadamente ácidos de 5,4 a 6,3; praticamente neutros entre 6,4 e 7,4; e, quando o pH supera 8,4, considerados alcalinos.

Nos casos de alta acidez, solos podem apresentar níveis perigosos de alumínio e manganês, afetando negativamente as plantas. A acidez excessiva, além de aumentar o alumínio tóxico, acarreta deficiência de cálcio e magnésio, elementos essenciais para o desenvolvimento saudável das raízes, limitando a capacidade das plantas de absorver água e nutrientes (MELLONI; PEREIRA; ALVARENGA, 2008; SILVA *et al.*, 2020).

O pH está diretamente relacionado a fatores relevantes para as plantas. Os efeitos do pH nas plantas incluem: abundância ou escassez de nutrientes, colaboração com elementos que podem ser tóxicos, atividade de microrganismos, colaboração com espécies invasoras e condições físicas do solo (NOVAES *et al.*, 2007). Portanto é crucial monitorar e gerenciar adequadamente o pH do solo para garantir o crescimento sustentável das plantas.

### **3.3.5 Teor de Potássio (K)**

O Teor de potássio é um nutriente essencial para o crescimento e desenvolvimento das plantas, pois participa de diversos processos metabólicos e fisiológicos, tais como a síntese de carboidratos, a ativação de enzimas, a abertura e fechamento dos estômatos, a transpiração e a eficiência do uso de água. Além disso, o K confere às plantas maior resistência a estresses abióticos, como geadas, secas,

salinidade do solo e doenças (NASCIMENTO *et al.*, 2017; GAUTAM *et al.*, 2022; BASAK *et al.*, 2023).

A disponibilidade adequada de K no solo é fundamental para aumentar a produtividade das culturas agrícolas (NASCIMENTO *et al.*, 2017).

A deficiência de K afeta negativamente o funcionamento dos estômatos, que se tornam mais lentos e permitem maior perda de água pelas folhas, agravando os efeitos do *déficit* hídrico. O teor de K também influencia o transporte de água das raízes para as partes aéreas, pois cria um gradiente de pressão osmótica que favorece a absorção de água pelas raízes (SERAFIM *et al.*, 2012).

As plantas deficientes em K apresentam menor capacidade de absorver água e sofrem maior estresse quando o conteúdo de água no solo fica abaixo do nível crítico para a cultura (SERAFIM *et al.*, 2012).

### 3.4 AGRICULTURA DE PRECISÃO – AP

O propósito central da Agricultura de Precisão (AP) é otimizar o manejo da variabilidade espacial, buscando maximizar o retorno econômico ao mesmo tempo em que minimiza os impactos ambientais (INAMASU *et al.*, 2011). Para alcançar esse objetivo, a AP se apoia na aplicação localizada de tecnologias, adaptadas às peculiaridades de cada área específica (GREGO; OLIVEIRA; VIEIRA, 2014).

Um dos pilares dessa abordagem é a criação de mapas detalhados dos sistemas de produção agrícola, que permitem identificar zonas homogêneas, facilitando o uso eficiente de insumos e práticas de manejo. Isso, por sua vez, auxilia o produtor na tomada de decisões mais assertivas e contribui para a redução de custos e aumento da sustentabilidade. Contudo, gerar esses mapas com precisão depende do uso de ferramentas como a geoestatística, que desempenha um papel fundamental ao fornecer informações detalhadas sobre a variabilidade espacial e ajudar na predição de atributos do solo e da cultura em áreas não amostradas.

Dessa forma, a geoestatística se torna um aliado crucial na formulação de decisões estratégicas e na gestão eficiente dos sistemas agrícolas, reduzindo impactos ambientais e promovendo a sustentabilidade dentro da AP (GREGO; OLIVEIRA; VIEIRA, 2014).

### 3..5 GEOESTATÍSTICA

A geoestatística surgiu na África do Sul, em 1951, impulsionada pela necessidade de métodos para estimar reservas minerais (CRESSIE, 2015). A iniciativa pioneira de Daniel G. Krige e H. S. Sichel estabeleceu o alicerce para o que viria a ser uma disciplina científica rigorosa. Na década de 1960, Matheron (1963) ampliou significativamente o escopo da geoestatística e formulou a Teoria das Variáveis Regionalizadas, que se tornou um marco na análise da dependência espacial e na interpolação de fenômenos naturais.

No Brasil, o desenvolvimento da geoestatística ganha impulso com o decreto-lei nº 764, de 15 de agosto de 1969 (BRASIL, 1969), que estabelece a Companhia de Pesquisa de Recursos Minerais (CPRM). Este marco legal, aliado ao decreto nº 66.522 de 30 de abril de 1970 (BRASIL, 1970), que define normas de assistência financeira à pesquisa mineral e ao setor de mineração, fortalece a pesquisa mineral no país e cria as bases para os avanços em geoestatística. Inicialmente voltada para atender às demandas do setor de mineração, a geoestatística foi fundamental na exploração e avaliação de recursos minerais, proporcionando metodologias inovadoras para analisar a variabilidade espacial dos recursos e otimizar processos de prospecção.

A partir dos anos 1980, com o avanço da computação, essas técnicas se expandiram para além da mineração e encontraram aplicações em outros campos, como questões ambientais (MEDEIROS; LUCIO; SILVA, 2017), análise espacial de crimes (BARRETO, 2016), mudanças climáticas (MOTA, 2021), e muitos outros setores que lidam com dados georreferenciados.

Dessa forma, a geoestatística evoluiu para se tornar uma ferramenta essencial para várias ciências aplicadas, permitindo a análise espacial detalhada e a criação de modelos preditivos que suportam tomadas de decisão mais precisas e informadas.

A metodologia geoestatística se destaca por sua capacidade de analisar a variabilidade espacial de variáveis de interesse por meio de modelos matemáticos e estatísticos sofisticados. Vieira (2000) ressalta a adequação da geoestatística para tal análise, dada sua consideração da dependência espacial entre os dados. Uma das suas aplicações mais notáveis é a criação de mapas detalhados que representam a distribuição espacial da variável em estudo (MCBRATNEY *et al.*, 2005).

Para empregar a geoestatística de forma eficaz, é essencial verificar a existência de dependência espacial entre os elementos amostrais. Isso implica que os valores da variável são afetados pela sua localização no espaço, resultando em uma variância não constante que é influenciada pelas distâncias entre os pontos amostrais. Instrumentos como o semivariograma, covariograma ou correlograma são utilizados para quantificar essa dependência espacial (URIBE-OPAZO; BORSSOI; GALEA, 2012).

Utilizando a função de dependência espacial definida, é possível estimar valores da variável em qualquer ponto do espaço por meio da *krigagem*. Esse método de interpolação é considerado ótimo por minimizar a variância do erro de estimativa e por não introduzir tendências nos valores estimados. A *krigagem* pondera os valores dos pontos amostrais próximos ao ponto de estimativa baseando-se no grau de dependência espacial entre eles (BURGESS & WEBSTER, 1980).

### 3.5.1 Modelo Espacial Linear Gaussiano

De acordo com Mardia e Marshall (1984), para modelar um conjunto de dados com dependência espacial, considera-se uma variável georreferenciada  $\{Z(\mathbf{s}_i), \mathbf{s}_i \in \mathcal{S}\}$ , com  $\mathbf{s}_i = (x_i, y_i)^\top$  sendo o vetor que representa uma determinada localização na área de estudo, em que  $\mathcal{S} \subset \mathcal{R}^2$ , sendo que  $\mathcal{R}^2$  é o espaço euclidiano bidimensional. Esse processo é composto por um conjunto de observações da variável aleatória sob estudo  $Z = (Z(\mathbf{s}_1), Z(\mathbf{s}_2), \dots, Z(\mathbf{s}_n))^\top$  em que  $\mathbf{s}_i, i = 1, \dots, n$  representa as localizações espaciais conhecidas e geradas pelo Modelo Espacial Linear Gaussiano dado conforme consta na Equação 1 (URIBE-OPAZO; BORSSOI; GALEA, 2012).

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \varepsilon(\mathbf{s}_i), \quad (\text{Eq. 1})$$

Para  $i = 1, \dots, n$ , em que,  $\mu(\mathbf{s}_i)$  é o termo determinístico e  $\varepsilon(\mathbf{s}_i)$  é o termo estocástico e dependem de localização em que  $Z(\mathbf{s}_i)$  é observada. Assume-se que os erros aleatórios  $\varepsilon(\mathbf{s}_i)$  possuem média zero, isto é,  $(E[\varepsilon(\mathbf{s}_i)] = 0)$ , e a variação entre as observações é dada por uma função de covariância  $Cov[\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_u)] = (\|\mathbf{s}_i - \mathbf{s}_u\|)$ , sendo  $\|\mathbf{s}_i - \mathbf{s}_u\|$  a distância euclidiana entre  $\mathbf{s}_i$  e  $\mathbf{s}_u$ .

Assume-se que quando o processo for estacionário de segunda ordem, então espacial  $\mu(\mathbf{s}_i) = \mu\mathbf{1}$ , onde  $\mu$  é constante e  $\mathbf{1}$  é o vetor unitário de dimensão  $n \times 1$ , com  $var(Z(\mathbf{s}_i)) = \sigma^2$  e  $cov(Z(\mathbf{s}_i), Z(\mathbf{s}_u)) = Cov[\varepsilon(\mathbf{s}_i), \varepsilon(\mathbf{s}_u)] = C(\|\mathbf{s}_i - \mathbf{s}_u\|)$ . Caso contrário, pode haver uma tendência em função tanto da coordenada X quanto da Y, ou até mesmo função de covariáveis (CRESSIE, 2015).

Considere-se o Modelo Espacial Linear Gaussiano expresso em notação matricial, conforme a Equação 2:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (Eq. 2)$$

Em que: o vetor de erros aleatórios  $\boldsymbol{\varepsilon}$ ,  $n \times 1$ , tem  $E(\boldsymbol{\varepsilon}) = \mathbf{0}$  e matriz de covariância  $\boldsymbol{\Sigma} = [(\sigma_{iu})]$ ,  $n \times n$ , que são as autocovariâncias entre a variável  $\mathbf{Z}$  em diferentes localizações, com elementos  $\sigma_{iu} = C(s_i, s_u)$ . Assume-se que  $\boldsymbol{\Sigma}$  é uma matriz não singular e simétrica e definida positiva, e que  $\mathbf{X}_{n \times (p+1)}$  é uma matriz cuja  $i$ -ésima linha é dada por  $\mathbf{x}_i^T = (1, x_{i1}, \dots, x_{ip})$ , sendo,  $x_{iu} = x_u(\mathbf{s}_i)$ , em que  $i = 1, \dots, n$  e  $u = 1, \dots, p$  representando o valor da  $u$ -ésima covariável na  $n$ -ésima posição;  $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$  é um vetor  $(p+1) \times 1$  dos parâmetros desconhecidos a serem estimados; e o vetor aleatório  $\mathbf{Z}$  segue uma distribuição gaussiana  $n$ -variada com vetor de médias  $\mu(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta}$  e matriz de covariância  $\boldsymbol{\Sigma}$ , ou seja,  $\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$  (CRESSIE, 2015; DE BASTIANI *et al.*, 2015, MALTAURO, 2022).

A modelagem espacial depende da estrutura da matriz de covariância  $\boldsymbol{\Sigma}$ , considerando-se de maneira particular a forma paramétrica dada por Mardia e Marshall (1984), conforme a Equação 3:

$$\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3), \quad (Eq. 3)$$

Em que,  $\varphi_1 \geq 0$  é o efeito pepita que é uma medida da descontinuidade do semivariograma para distâncias inferiores à menor distância entre dois pontos amostrados (ISAKS & SRIVASTAVA, 1989; CRESSIE, 2015), ou seja, é o valor da função semivariância quando a distância tende a zero;  $\mathbf{I}_n$  é a matriz identidade  $n \times n$ ;  $\varphi_2 \geq 0$  é a contribuição e o segmento gráfico neste intervalo representa a dependência espacial da variável de interesse, e é obtida pela diferença entre o patamar e o efeito pepita (SEIDEL & OLIVEIRA, 2016);  $\varphi_3 \geq 0$  é a função de alcance ( $a$ ) do modelo, isto

é,  $a = g(\varphi_3)$  em que este alcance indica a distância máxima em que os pontos amostrais apresentam correlação espacial significativa (SEIDEL & OLIVEIRA, 2016);  $\mathbf{R}$  é a matriz cujos elementos estão em função de  $\varphi_3$ , isto é,  $\mathbf{R} = \mathbf{R}(\varphi_3) = [(r_{iu})]$ , matriz  $n \times n$ , simétrica, em que  $r_{iu}$  depende apenas da Distância Euclidiana entre  $\mathbf{s}_i$  e  $\mathbf{s}_u$  ( $h_{iu} = \|\mathbf{s}_i - \mathbf{s}_u\|$ ), com elementos da diagonal  $r_{iu} = 1$ , para  $i = u$ ,  $r_{iu} = \varphi_2^{-1}C(\mathbf{s}_i, \mathbf{s}_u)$  para  $\varphi_2 \neq 0$ ; e  $r_{iu} = 0$  para  $\varphi_2 = 0$ ,  $i \neq u$  ( $i, u = 1, \dots, n$ ) (URIBE-OPAZO, BORSSOI, GALEA, 2012; DE BASTIANI *et al.*, 2015, MALTAURO *et al.*, 2023).

### 3.5.2 Semivariograma

O semivariograma é um gráfico que representa a função semivariância  $\gamma(h)$  em função da distância  $h$ . Sendo utilizado para a análise da estrutura de dependência espacial entre os elementos amostrais dentro de um espaço paramétrico. A estrutura de dependência espacial obtida a partir do ajuste de um modelo teórico é empregada nas estimativas de valores para locais não amostrados por meio da aplicação da técnica de *krigagem*.

A estimação da estrutura de dependência espacial é realizada por meio do ajuste de um modelo teórico à função semivariância amostral (ou à função covariância ou função de autocorrelação espacial). As estimativas dos parâmetros do modelo ajustado são obtidas utilizando métodos estatísticos a partir dos dados amostrais.

A função variância  $2\gamma(s, s + h)$  é definida, como demonstrado na Equação 4:

$$2\gamma(s, s + h) = 2\gamma(h) = \mathbb{E} \left[ (Z(s) - Z(s + h))^2 \right], \quad (\text{Eq. 4})$$

Assim, define-se a função semivariância  $\gamma(h)$ , dividindo os dois lados da equação por 2, resultando na Equação 5:

$$\gamma(h) = \frac{1}{2} \mathbb{E} \left[ (Z(s) - Z(s + h))^2 \right], \quad (\text{Eq. 5})$$

A Equação 6 apresenta a relação entre a função semivariância  $\gamma(h)$  e a função covariância  $C(h)$ :

$$\gamma(h) = C(0) - C(h), \quad (\text{Eq. 6})$$

A partir da relação entre a função semivariância  $\gamma(h)$  e a função covariância  $C(h)$ , Equação 6, têm-se que para  $C(0) \neq 0$  a relação  $\frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{C(0)}$ , assim, a função de autocorrelação espacial  $\rho(h)$ , a Equação 7 é:

$$\rho(h) = 1 - \frac{\gamma(h)}{C(0)}, \quad (\text{Eq. 7})$$

Portanto, se a hipótese de estacionariedade de segunda ordem puder ser satisfeita, a covariância  $C(h)$  e a função semivariância  $\gamma(h)$  pode caracterizar a dependência espacial (ISAACS & SRIVASTAVA, 1989).

Um estimador da função semivariância  $\hat{\gamma}(h)$ , como ilustrado na Equação 5, é muito conhecido é o estimador de Matheron (1963), Equação 8:

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

Em que,  $N(h)$  é o número de pares de valores  $Z(s_i)$ ,  $Z(s_i + h)$  separados pela distância  $h$ .

A literatura também apresenta outros estimadores da função semivariância como o Estimador de Cressie e Hawkins (CRESSIE & HAWKINS, 1980), o Estimador de Semivariância Relativa (SRIVASTAVA & PARKER, 1989), os Estimadores New1 e New2 (LI & LAKE, 1994) e o estimador de Genton (GENTON, 1998).

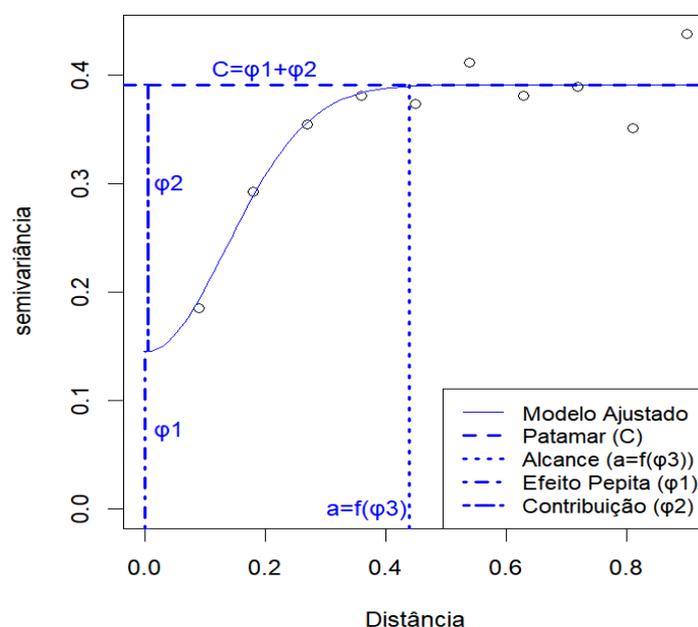
A Figura 1 ilustra os parâmetros do semivariograma, usados para medir a variabilidade espacial de uma determinada variável e definir sua estrutura de dependência espacial. Um dos parâmetros-chave é o efeito pepita, que reflete a variação em pequenas escalas (menores que a distância mínima entre dois pontos amostrados), ou seja, discrepâncias que a malha de amostragem não consegue captar.

No gráfico do semivariograma, o efeito pepita é representado pelo valor que o semivariograma assume quando  $h = 0$ . Já o alcance corresponde à distância a partir da qual a dependência espacial se estabiliza e a variância se mantém constante. No

gráfico, esse é o ponto em que o modelo ajustado se aproxima de um valor assintótico, indicando a distância máxima até onde as amostras ainda exibem correlação. A partir desse ponto, considera-se que a variabilidade entre as amostras não é mais afetada pela dependência espacial, o que significa que as amostras se tornam independentes (ISAAKS & SRIVASTAVA, 1989).

Conforme a distância  $h$  entre os pontos de amostragem aumenta, o semivariograma normalmente apresenta um crescimento na variância das diferenças entre os valores. Isso acontece porque, em distâncias maiores, as amostras tendem a ser mais independentes entre si, resultando em uma maior variabilidade.

Figura 1 Exemplo de modelo espacial ajustado ao semivariograma.



Fonte: O autor (2024).

Na geoestatística, a construção de um semivariograma amostral exige decisões criteriosas. Clark (1979) sugere calcular a semivariância até uma distância máxima entre pontos, o “*cutoff*”, que pode ser  $\frac{1}{4}$ ,  $\frac{1}{3}$  ou  $\frac{1}{2}$  dessa distância. A seleção de um “*cutoff*” além desses valores é possível com uma justificativa lógica.

A definição do número de *lags* ( $h$ ) e dos pontos amostrais para estimar as semivariância experimentais é essencial. Em amostragens regulares, os próprios intervalos entre amostras podem servir como *lags*. Journel e Huijbregts (1978) e consideram um mínimo de 30 pares para o cálculo da função semivariância, enquanto Webster e Oliver (2007) recomendam ao menos 100 pares.

Um semivariograma constante em todos os valores de ( $h$ ) indica um efeito pepita puro ou ausência de dependência espacial, sugerindo que o alcance ( $a$ ) é menor que o menor espaçamento entre amostras, resultando em dados não correlacionados. A normalidade dos dados não implica independência entre as amostras, que pode ser verificada pela autocorrelação (URIBE-OPAZO; BORSSOI; GALEA, 2012).

### 3.5.3 Isotropia

A isotropia é identificada quando os semivariogramas de diferentes direções, geralmente  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ , mostram um padrão uniforme de continuidade espacial. Isso permite o uso de um semivariograma omnidirecional para análise, simplificando a estimativa em locais não amostrados com uma área de influência circular.

Já a anisotropia ocorre quando a variabilidade não é uniforme em todas as direções, revelada por semivariogramas distintos para as diferentes direções. Clark (1979), Wackernagel (2003) e Guedes *et al.* (2008) descrevem como identificar anisotropia e suas direções privilegiadas.

Existem três tipos principais de anisotropia: geométrica, zonal e combinada. A geométrica apresenta alcances distintos, mas patamares consistentes em todas as direções; a zonal mostra o mesmo modelo com patamares diferentes; e a combinada mescla características das duas anteriores (ISAAKS & SRIVASTAVA, 1989; GUEDES *et al.*, 2008).

### 3.5.4 Modelos Espaciais

Nesta seção, serão apresentados alguns modelos espaciais teóricos para a função semivariância, função de covariância e função de autocorrelação, propostos por Isaaks e Srivastava (1989), Cressie (1993), Deutch e Journel (1999) e Olea (2006). Esses modelos teóricos são empregados para a avaliação da estrutura da dependência espacial.

Conforme apontado por (WEBSTER & OLIVER, 2007) os modelos amplamente adotados são: Exponencial, Gaussiano e Família Matérn com diferentes parâmetros de suavização  $k > 0$ . Esses modelos pressupõem que a variável em estudo apresenta

um padrão de continuidade espacial uniforme em todas as direções (isotropia). Portanto espera-se que o semivariograma exiba uma característica estrutural consistente em todas as direções.

A seguir são apresentados os modelos espaciais: exponencial, gaussiano, família Matérn e Wave, válidos em  $\mathbb{R}$ ,  $\mathbb{R}^2$  e  $\mathbb{R}^3$  (ISAAKS & SRIVASTAVA, 1989).

O primeiro é o Modelo exponencial (Equação 9), que apresenta comportamento aproximadamente linear na origem e atinge o patamar assintoticamente, com alcance prático definido como a distância na qual o valor do modelo é 95% de  $\varphi_2$ , sendo o alcance dado por  $a = 3\varphi_3$ , conforme a Figura 2a.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - \exp\left(-\frac{h}{\varphi_3}\right) \right], & h > 0 \end{cases} \quad (\text{Eq. 9})$$

Já o Modelo Gaussiano (Equação 10) apresenta comportamento parabólico na origem e é utilizado para modelar um fenômeno extremamente contínuo. Também atinge o patamar apenas assintoticamente e o alcance é dado por  $a = \sqrt{3}\varphi_3$ , conforme exposto na figura Figura 2b.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - \exp\left(-\frac{h}{\varphi_3}\right) \right]^2, & h > 0 \end{cases} \quad (\text{Eq. 10})$$

O Modelo família Matérn (Equação 11), proposto por Matérn (1986), é uma função positiva definida em termos de modelo teórico de semivariância, conforme as Figura 2c, 2d e 2e.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[ 1 - \left( 2^{k-1} \Gamma(k) \right)^{-1} \left( \frac{h}{\varphi_3} \right)^k K_k \left( \frac{h}{\varphi_3} \right) \right]^2, & h > 0 \end{cases} \quad (\text{Eq. 11})$$

Em que:  $k > 0$ ,  $\Gamma(k) = \int_0^\infty x^{k-1} e^{-x} dx$  é a função gama e  $K_k(u) = \frac{1}{2} \int_0^\infty x^{k-1} e^{\left(-\frac{1}{2}\right)u(x+x^{-1})} dx$  é a função de Bessel modificada de terceiro tipo de ordem  $k$  (GRADSHTEYN & RYZHIK, 2000). Na família Matérn, o parâmetro de ordem  $k$  é um

parâmetro de forma que controla o grau de suavidade do processo estocástico subjacente  $Z$ . Quando  $k$  é menor o processo apresenta menos suavidade, enquanto valores maiores de  $k$  implicam em maior suavidade.

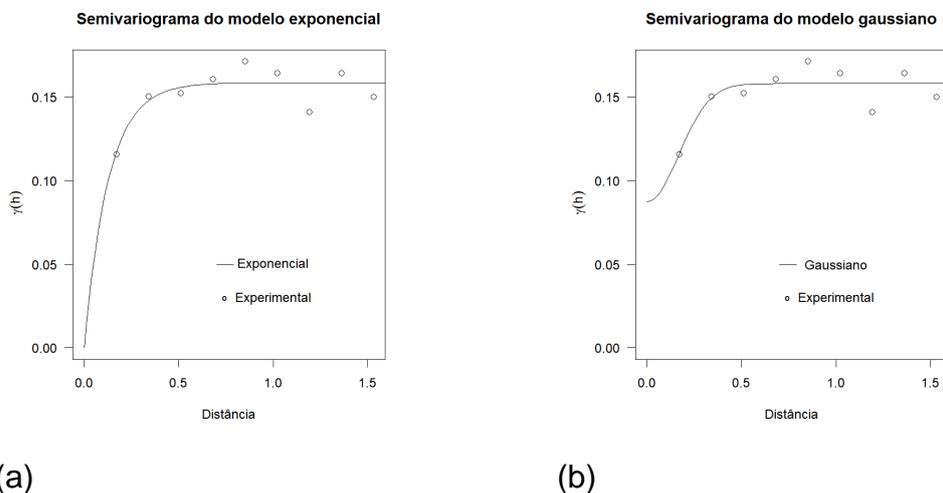
Os modelos Exponencial e Gaussiano são casos especiais da família Matérn. No caso do modelo exponencial, o valor de  $k$  é 0,5, já no Modelo Gaussiano, o valor de  $k \rightarrow \infty$ .

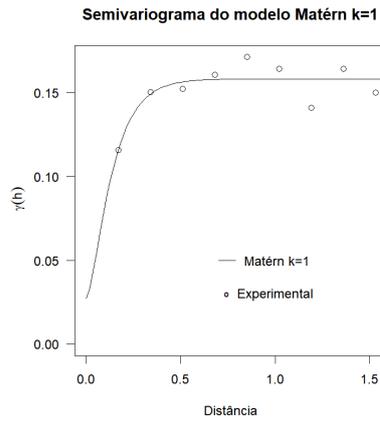
O Modelo de *Wave* é utilizado em geoestatística para fenômenos com alta variabilidade espacial, como por exemplo a precipitação (ALMEIDA *et al.*, 2011). Esse modelo de semivariograma não-monotônico, conhecido como *hole effect* (CARVALHO; VIEIRA; GREGO, 2009), reflete descontinuidades espaciais nas quais altos valores estão cercados por baixos valores e vice-versa (CHILÈS & DELFINER, 1999).

O Modelo de *Wave*, conforme a Equação 12, (OLEA, 2006) caracteriza-se por uma estrutura oscilante sem patamar definido, podendo apresentar comportamento isotrópico ou anisotrópico. As oscilações do Modelo de *Wave* diminuem com o aumento da distância e atingem periodicamente a contribuição quando  $h = \pi * \varphi_3$  (Mota, 2021).

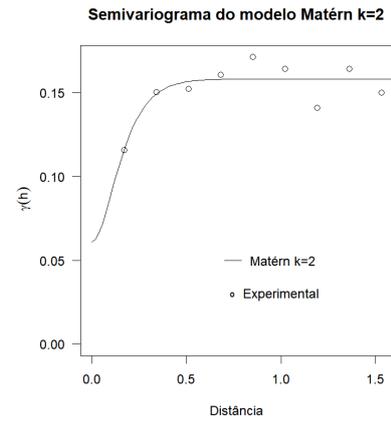
$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left( 1 - \frac{\varphi_3}{h} \sin\left(\frac{h}{\varphi_3}\right) \right), & h > 0 \end{cases} \quad (Eq. 12)$$

Figura 2 Gráficos da função semivariância  $\gamma(h)$  para os modelos: (a) exponencial, (b) gaussiano, (c) Matérn com parâmetro de suavização  $k=1$ , (d) Matérn com parâmetro  $k=2$ , (e) Matérn com parâmetro  $k=3$  e (f) *wave*.

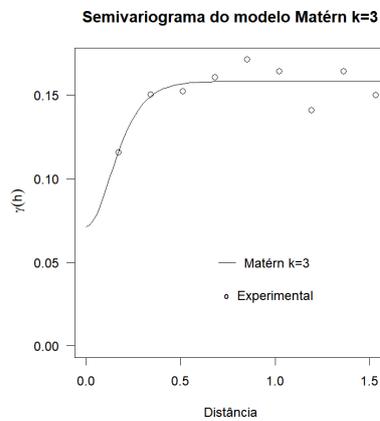




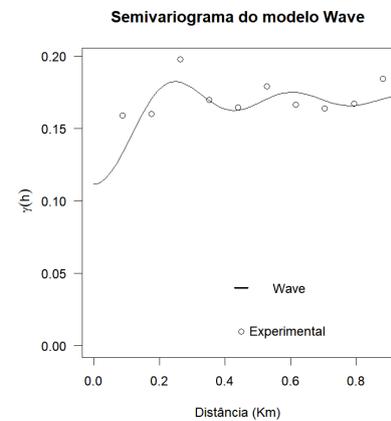
(c)



(d)



(e)



(f)

Fonte: O autor (2024).

### 3.5.5 Estimação de parâmetros por Máxima Verossimilhança (ML)

O método da máxima verossimilhança (ML) consiste em encontrar valores dos parâmetros dos modelos ajustados ao semivariograma experimental, que tornam mais provável a ocorrência dos dados, ou seja, que maximizam a função de verossimilhança (URIBE-OPAZO; BORSSOI; GALEA, 2012).

Este conceito foi introduzido por Mardia e Marshall (1984), considerando que, seja  $\mathbf{Z} = Z(s_1), \dots, Z(s_n)$ , representado por um modelo espacial linear gaussiano (Equação 13). Em nosso caso inicial considera-se  $\mathbf{Z} \sim N_n(\beta\mathbf{1}, \Sigma)$ , isso é,  $\mathbf{Z}$  tem distribuição normal  $n$ -variada com vetor de médias  $\beta\mathbf{1}$ , sendo  $\beta$  um escalar e matriz de covariância  $\Sigma$ .

Os parâmetros desconhecidos do modelo  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\varphi}^\top)^\top$ ,  $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^\top$  podem ser estimados maximizando o logaritmo da função verossimilhança. O logaritmo da função verossimilhança é dado pela Equação 13 (MARDIA; MARSHALL, 1984):

$$l(\boldsymbol{\theta}) = -\left(\frac{n}{2}\right) \log(2\pi) - \left(\frac{1}{2}\right) \log|\boldsymbol{\Sigma}| - \left(\frac{1}{2}\right) (\mathbf{Z} - \boldsymbol{\beta}\mathbf{1})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\beta}\mathbf{1}), \quad (\text{Eq. 13})$$

Em que:  $\delta = (\mathbf{Z} - \boldsymbol{\beta}\mathbf{1})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \boldsymbol{\beta}\mathbf{1})$  é conhecida como distância de Mahalanobis, que é uma medida da dissimilaridade entre os dados observados e os valores esperados pelo modelo, ponderado pela matriz de covariância  $\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3)$ .

Por meio da maximização da função escore, que é a derivada da função de verossimilhança em relação aos parâmetros igualada a zero, são geradas as estimativas de máxima verossimilhança para o vetor de parâmetros  $\boldsymbol{\theta}$  do modelo, isto é,  $l(\hat{\boldsymbol{\theta}}) = \max l(\boldsymbol{\theta})$ ,  $\boldsymbol{\theta} \in \theta$ , em que  $\theta$  é o espaço paramétrico e  $\hat{\boldsymbol{\theta}}$  é o estimador da máxima verossimilhança de  $\boldsymbol{\theta}$ .

### 3.5.5.1 Matriz de Informação Observada de Fisher

A matriz de informação observada, é uma matriz simétrica que mede a sensibilidade da função de log-verossimilhança aos valores dos parâmetros. Esta matriz contém a segunda derivada da função de verossimilhança em relação aos parâmetros do modelo, que indica a curvatura da função em torno do ponto de máximo.

A matriz de informação observada é definida como o negativo da segunda derivada da função de verossimilhança, avaliada no ponto de máximo, isto é,  $l(\boldsymbol{\theta}) = -\ddot{L}(\boldsymbol{\theta})$  e deve ser avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , ou seja, é avaliada no valor estimado dos parâmetros. Sua forma particionada pode ser encontrada a partir de  $-l(\boldsymbol{\theta}) = \ddot{L}(\boldsymbol{\theta})$ , sendo a Equação 14 (URIBE-OPAZO; BORSOI; GALEA, 2012):

$$\ddot{L}(\boldsymbol{\theta}) = \ddot{L}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{\partial^2 l(\boldsymbol{\theta})}{(\partial \boldsymbol{\theta} \partial \boldsymbol{\theta})^\top} = \begin{pmatrix} \ddot{L}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \ddot{L}_{\boldsymbol{\beta}\boldsymbol{\varphi}} \\ \ddot{L}_{\boldsymbol{\varphi}\boldsymbol{\beta}} & \ddot{L}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \end{pmatrix}, \quad (\text{Eq. 14})$$

Em que:

$$\ddot{L}_{\beta\beta} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}} = -(\mathbf{1}^\top \Sigma^{-1} \mathbf{1}); \quad (\text{Eq. 15})$$

$$\ddot{L}_{\beta\varphi} = \ddot{L}_{\varphi\beta^\top} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}^\top}, \quad (\text{Eq. 16})$$

Com elementos:

$$\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \varphi_j} = -\mathbf{1}^\top \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \boldsymbol{\varepsilon}, \quad (\text{Eq. 17})$$

Sendo,  $j = 1, \dots, q$ ;  $\boldsymbol{\varepsilon} = (\mathbf{Z} - \beta \mathbf{1})$ , e

$$\ddot{L}_{\varphi\varphi} = \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^\top}, \quad (\text{Eq. 18})$$

Com elementos:

$$\begin{aligned} \frac{\partial^2 l(\boldsymbol{\theta})}{\partial \varphi_i \partial \varphi_j} = & \frac{1}{2} \text{tr} \left\{ \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} \right) \right\} \\ & + \frac{1}{2} \boldsymbol{\varepsilon}^\top \Sigma^{-1} \left\{ \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \right\} \Sigma^{-1} \boldsymbol{\varepsilon}, \end{aligned} \quad (\text{Eq. 19})$$

Com,  $i, j = 1, \dots, q$ .

### 3.5.5.2 Matriz de Informação Esperada de Fisher

A matriz de informação esperada de Fisher é uma matriz que contém a esperança da segunda derivada da função de verossimilhança em relação aos parâmetros do modelo, calculada a partir da distribuição conjunta dos dados, ou seja, usando a expectativa da segunda derivada da função de verossimilhança.

Os erros padrões assintóticos podem ser calculados ao se inverter a matriz de informação de Fisher, esta matriz para o modelo espacial linear gaussiano é dada pela Equação 20 (LANGE; CAVALLI; PEREIRA, 1989; MITCHELL, 1989; URIBE-OPAZO; BORSOI; GALEA, 2012).

$$K(\boldsymbol{\theta}) = E[I(\boldsymbol{\theta})] = \begin{pmatrix} E[-\ddot{L}_{\beta\beta}] & E[-\ddot{L}_{\beta\varphi}] \\ E[-\ddot{L}_{\varphi\beta}] & E[-\ddot{L}_{\varphi\varphi}] \end{pmatrix} = \begin{pmatrix} K(\boldsymbol{\beta}) & 0 \\ 0 & K(\boldsymbol{\varphi}) \end{pmatrix}, \quad (\text{Eq. 20})$$

Em que:

$$K(\boldsymbol{\beta}) = E[-\ddot{L}_{\beta\beta}] = E \left[ -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^\top} \right] = \mathbf{1}^\top \Sigma^{-1} \mathbf{1}; \quad (\text{Eq. 21})$$

$$K(\boldsymbol{\varphi}) = E[-\ddot{L}_{\varphi\varphi}] = E \left[ -\frac{\partial^2 l(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^\top} \right], \quad (\text{Eq. 22})$$

A qual tem como elementos,

$$k_{ij}(\varphi) = \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \right], \quad (\text{Eq. 23})$$

Com  $i, j = 1, 2, 3$ .

### 3.5.6 Krigagem Ordinária

Em seus estudos, Elsaaks e Srivastava (1989) descrevem que, na geoestatística, uma estimativa para um local não amostrado  $Z(s_0)$ , pode ser obtida através de uma combinação linear ponderada de valores medidos  $Z(s_i)$ , onde cada valor é associado a um peso  $\lambda_i$ ,  $i = 1, \dots, n$ .

A *krigagem* ordinária é uma técnica de estimativa espacial, ideal para prever valores de uma variável distribuída geograficamente, com base nos valores medidos em pontos próximos. O diferencial desse método, em comparação a outras formas de interpolação numérica, está na sua precisão, pois utiliza o semivariograma — uma ferramenta poderosa que não apenas mede a distância entre os pontos, mas também leva em consideração fatores complexos como o efeito pepita (variabilidade em pequena escala), o alcance (distância máxima de correlação) e a presença de anisotropia (variação que depende da direção) (YAMAMOTO & LANDIM, 2013).

Esses elementos permitem à *krigagem* ajustar as estimativas com uma sensibilidade única às condições reais do terreno. Outro ponto-chave é o sistema de pesos que atribui mais importância às amostras mais próximas do ponto a ser estimado, garantindo que a estimativa seja influenciada principalmente pelos valores adjacentes mais relevantes.

Esse método se destaca por fornecer previsões mais precisas do que outras técnicas de interpolação, como a média ponderada ou o inverso da distância, resultando em uma melhor compreensão da variabilidade espacial (YAMAMOTO & LANDIM, 2013). A *krigagem* ordinária é dada pela Equação 24:

$$\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i), \quad (\text{Eq. 24})$$

Em que,  $\hat{Z}(s_0)$  é o valor estimado no ponto  $(s_0)$  não amostrado,  $n$ : número de valores medidos  $Z(s_i)$  das variáveis em pontos amostrados e  $\lambda_i$ : pesos associados a cada valor  $Z(s_i)$  medido.

Esse estimador, segundo Webster e Oliver (2007) é considerado como o melhor estimador pois produz estimativas não viciadas e com variâncias mínimas. Tais qualidades, são garantidas uma vez que os pesos  $\lambda_i$  são determinados impondo-se que a presença do erro seja zero e a variância do erro seja mínima.

Seja o erro  $\varepsilon(s_0) = Z(s_0) - \hat{Z}(s_0)$  com esperança  $\mathbb{E} [Z(s_0) - \hat{Z}(s_0)]$ , esse erro será nulo se:

$$\sum_{i=1}^n \lambda_i = 1, \quad (\text{Eq. 25})$$

A condição definida na Equação (25) é necessária para garantir que o estimador seja não tendencioso. Além disso, na *krigagem* a variância ( $\sigma_E^2$ ) do erro  $\varepsilon(s_0) = \hat{Z}(s_0) - Z(s_0)$ , deve ser mínima, ou seja:

$$\sigma_E^2 = \text{Var}[\hat{Z}(s_0) - Z(s_0)] = \mathbb{E} \left[ \left( \hat{Z}(s_0) - Z(s_0) \right)^2 \right] - \left\{ \mathbb{E} \left[ \hat{Z}(s_0) - Z(s_0) \right] \right\}^2,$$

Em que o primeiro termo corresponde ao erro quadrático e o segundo termo ao viés.

Aplicando a Equação 24, têm-se:

$$\sigma_E^2 = \mathbb{E} \left[ \left( \hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0) \right)^2 \right] = \mathbb{E} \left[ \left( \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0) \right)^2 \right],$$

Decompondo o termo quadrático:

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

Mas, como  $\mathbb{E} [Z(\mathbf{s}_i) Z(\mathbf{s}_j)] = Cov = [Z(\mathbf{s}_i) Z(\mathbf{s}_j)]$  e  $\mathbb{E} [(Z(\mathbf{s}_0))^2] = Var[Z(\mathbf{s}_0)]$ ,

têm-se:

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

Que é equivalente a:

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

Sendo:  $2\mu(1 - \sum_{i=1}^n \lambda_i) = 0$ , em que  $\mu$  é um multiplicador de Lagrange.

A minimização da  $\sigma_E^2$  consiste em calcular as  $n + 1$  derivadas parciais com relação a  $\lambda_1, \dots, \lambda_n$  e  $\mu$  e igualar a zero, obtendo um sistema de  $n+1$  equações e  $n+1$  incógnitas cuja solução resulta nos  $n$  pesos  $\lambda_i$  que cumprem a condição de não enviesamento  $\sum_{i=1}^n \lambda_i = 1$  e minimizam a variância de estimação  $Var[\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)] - 2\mu(1 - \sum_{i=1}^n \lambda_i)$ .

Assim, têm-se que:

$$2 \sum_{j=1}^n \lambda_j Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] - 2Cov[Z(\mathbf{s}_i), Z(\mathbf{s}_0)] - 2\mu = 0,$$

Para  $i = 1, \dots, n$  e  $2(1 - \sum_{i=1}^n \lambda_i) = 0$ .

O sistema de *krigagem* de  $n+1$  equações, cuja solução fornece  $n$  pesos  $\lambda_i$  é:

$$\sum_{j=1}^n \lambda_j \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_j)] - \mu = \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_0)],$$

Para  $i = 1, \dots, n$ , e  $\sum_{i=1}^n \lambda_i = 1$ , logo:

$$\sigma_E^2 = \text{Var}[Z(\mathbf{s}_0)] - \sum_{i=1}^n \lambda_i \text{Cov}[Z(\mathbf{s}_i), Z(\mathbf{s}_0)] - \mu,$$

Como,  $\gamma(h) = C(0) - C(h)$ , tem-se as condições que garantem a não tendenciosidade e variância mínima, são soluções do sistema de equações, dadas na Equação 26, para  $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$  que é a distância euclidiana entre  $\mathbf{s}_i$  e  $\mathbf{s}_j$ :

$$\sum_{j=1}^n \lambda_j \gamma(h_{ij}) + \mu = \gamma(h_{0i}), \quad (\text{Eq. 26})$$

Para  $i = 1, \dots, n$ , e  $\sum_{i=1}^n \lambda_i = 1$ .

De acordo com Webster e Oliver (2007) a solução da Equação (25) garante que o estimador de *krigagem* é o melhor estimador linear não tendencioso de variância mínima, com sigla em inglês, "BLUE" (*Best Linear Unbiased Estimator*).

O sistema expresso na Equação 26, construído com  $n+1$  equações e  $n+1$  incógnitas, é conhecido como sistema de *krigagem* ordinária. A incógnita  $\mu$  é um multiplicador de Lagrange, introduzido ao minimizar a variância do erro, e tem como forma matricial  $\gamma\lambda = A$ :

$$\gamma = \begin{bmatrix} \gamma(h_{11}) & \gamma(h_{12}) & \cdots & \gamma(h_{1n}) & 1 \\ \gamma(h_{21}) & \gamma(h_{22}) & \cdots & \gamma(h_{2n}) & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma(h_{n1}) & \gamma(h_{n2}) & \cdots & \gamma(h_{nn}) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}_{(n+1) \times (n+1)}, \lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \mu \end{bmatrix}_{(n+1) \times 1}, A = \begin{bmatrix} \gamma(h_{10}) \\ \gamma(h_{20}) \\ \vdots \\ \gamma(h_{n0}) \\ 1 \end{bmatrix}_{(n+1) \times 1}.$$

Os pesos  $\lambda_i$  podem, assim, ser obtidos mediante o produto da matriz inversa de  $\gamma$  pela matriz  $A$ , expresso pela Equação 27:

$$\boldsymbol{\lambda} = \boldsymbol{\gamma}^{-1} \boldsymbol{A}. \quad (\text{Eq. 27})$$

Uma vez obtidos os pesos  $\lambda_i$  e o valor  $\mu$ , pode-se também calcular a variância do erro  $\sigma_E^2$  (CRESSIE, 2015), tal como consta na Equação 28:

$$\sigma_E^2 = \sum_{i=1}^n \lambda_i \gamma(h_{i0}) - \mu. \quad (\text{Eq. 28})$$

Vale ressaltar que os componentes da matriz  $\boldsymbol{A}$  demandam conhecimento acerca da semivariância correspondente a valores de  $h$ , os quais não são necessariamente equivalentes aos *lags* do semivariograma experimental. Dessa forma, torna-se imperativa a adequação de um modelo teórico ao semivariograma experimental.

### 3.5.7 Krigagem Com Drift Externo

A *Krigagem Ordinária (Krigagem Comum)*, parte do princípio de que o processo é estacionário, ou seja, a média é constante em todo o espaço amostral. No entanto, quando informações auxiliares estão disponíveis em todos os pontos das grades, deve-se utilizar a *krigagem com drift* externo. Esta abordagem permite incorporar um modelo de tendência ou *drift* na estimação, o que é particularmente útil quando a variável de interesse está correlacionada com outras variáveis conhecidas (HENGL; HEUVELINK; STEIN, 2003; HENGL, 2009).

O objetivo da *krigagem* com drift externo é encontrar  $\lambda_i, i = 1, \dots, n$  tal que  $\hat{Z}(s_0) = \sum_{i=1}^n \lambda_i Z(s_i)$ , em que,  $s_0$  é um ponto não amostrado. Com erro de predição não enviesado e com variância mínima.

Para a condição de não enviesamento supõe-se que, conforme apresentado na Equação 29:

$$X_{0j} = \sum_{i=1}^n \lambda_i X_{ij}, \quad j = 0, \dots, p \quad (\text{Eq. 29})$$

Em que:

$$\begin{aligned}
 E\left(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)\right) &= \left(\sum_{i=1}^n \lambda_i E(Z(\mathbf{s}_i))\right) - E(Z(\mathbf{s}_0)) \\
 &= \left(\sum_{i=1}^n \lambda_i \sum_{j=0}^p \boldsymbol{\beta}_j \mathbf{X}_{ij}\right) - \sum_{j=0}^p \boldsymbol{\beta}_j \mathbf{X}_{0j} \\
 &= \sum_{j=0}^p \boldsymbol{\beta}_j \left(\sum_{i=1}^n \lambda_i \mathbf{X}_{ij} - \mathbf{X}_{0j}\right).
 \end{aligned}$$

As condições que minimizam a variância dos erros, define-se  $Z(\mathbf{s}) = \mu(\mathbf{s}) + \boldsymbol{\varepsilon}(\mathbf{s})$ , sendo que a  $E(\boldsymbol{\varepsilon}(\mathbf{s})) = 0$ , a covariância entre  $Z(\mathbf{s}_i)$  e  $Z(\mathbf{s}_j)$ , depende apenas da distância entre  $\mathbf{s}_i$  e  $\mathbf{s}_j$ , isto é,  $Cov(\mathbf{s}_i, \mathbf{s}_j) = C(h_{ij})$  em que  $h_{ij}$  é a distância euclidiana entre  $\mathbf{s}_i$  e  $\mathbf{s}_j$  (GARCIA-PAPANI *et al.*, 2018; SHEMMER, 2021).

Desse modo, a variância da estimação pode ser escrita em função dos resíduos  $\boldsymbol{\varepsilon}(\mathbf{s})$ , em que:

$$\begin{aligned}
 \hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0) &= \left(\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i)\right) - Z(\mathbf{s}_0) \\
 &= \left(\sum_{i=1}^n \lambda_i (\mu(\mathbf{s}_i) + \boldsymbol{\varepsilon}(\mathbf{s}_i))\right) - (\mu(\mathbf{s}_0) + \boldsymbol{\varepsilon}(\mathbf{s}_0)) \\
 &= \left(\sum_{i=1}^n \lambda_i (\mu(\mathbf{s}_i)) - \mu(\mathbf{s}_0)\right) + \left(\sum_{i=1}^n \lambda_i (\boldsymbol{\varepsilon}(\mathbf{s}_i) - \boldsymbol{\varepsilon}(\mathbf{s}_0))\right) \\
 &= \left(\sum_{i=1}^n \lambda_i \boldsymbol{\varepsilon}(\mathbf{s}_i)\right) - \boldsymbol{\varepsilon}(\mathbf{s}_0) \\
 &= \hat{\boldsymbol{\varepsilon}}(\mathbf{s}_0) - \boldsymbol{\varepsilon}(\mathbf{s}_0).
 \end{aligned}$$

Diante disso, a variância estimada é dada por:

$$\begin{aligned}
 Var(\hat{Z}(\mathbf{s}_0) - Z(\mathbf{s}_0)) &= E((\boldsymbol{\varepsilon}(\mathbf{s}_0) - \boldsymbol{\varepsilon}(\mathbf{s}_0))^2) \\
 &= C(0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(\mathbf{s}_i, \mathbf{s}_j) - 2 \sum_{i=1}^n C(\mathbf{s}_0, \mathbf{s}_i)
 \end{aligned} \tag{Eq. 30}$$

O processo de minimização na *krigagem* com *drift* externo é análogo ao da *Krigagem Ordinária*, onde deriva-se a Equação (29) em relação aos pesos e a iguala a zero para encontrar o mínimo. Não menos importante, pretende-se ainda que sejam cumpridas as condições estabelecidas pelas  $p + 1$  equações definidas na Equação 29.

Dessa forma, acrescentam-se  $p + 1$  termos nulos na Equação 29, correspondentes a estas condições. Assim, tem-se que:

$$\text{Var}(\hat{Z}(s_0) - Z(s_0)) = C(0) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C(s_i, s_j) - 2 \sum_{i=1}^n C(s_0, s_i) + 2 \sum_{j=0}^p n_j \sum_{i=1}^n \lambda_i X_{ij} - X_{0j},$$

em que,  $n_j$  são parâmetros de Lagrange. Agora, igualando a zero as  $n + p + 1$  derivadas em relação à  $\lambda_i, i = 1, \dots, n$  e  $n_j, j = 0, \dots, p$ , obtém-se o sistema de equações lineares:

$$\mathbf{S} = \begin{cases} \sum_{k=1}^n \lambda_k C(s_i, s_k) + \sum_{k=0}^p n_k X_{ik} = C(s_i, s_0), & i = 1, \dots, n \\ \sum_{k=1}^n \lambda_k X_{kj} = X_{0j}, & j = 0, \dots, p \end{cases}$$

Que, na representação matricial é  $\mathbf{S} : \mathbf{C} \boldsymbol{\lambda} = \mathbf{A}$ , em que  $\mathbf{C}$  é a matriz de ordem  $(n + p + 1) \times (n + p + 1)$ , dada por:

$$\mathbf{C} = \begin{pmatrix} C(s_2, s_1) & \dots & C(s_2, s_1) 1 & X_{11} & \dots & X_{1p} \\ C(s_2, s_1) & \dots & C(s_2, s_1) 1 & X_{21} & \dots & X_{2p} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ C(s_n, s_1) & \dots & C(s_n, s_n) 1 & X_{n1} & \dots & X_{np} \\ 1 & \dots & 1 & 0 & 0 & \dots & 0 \\ X_{11} & \dots & X_{n1} & 0 & 0 & \dots & 0 \\ \vdots & \dots & \vdots & \vdots & \vdots & \dots & \vdots \\ X_{p1} & \dots & X_{np} & 0 & 0 & \dots & 0 \end{pmatrix}_{(n+p+1) \times (n+p+1)},$$

Em que,  $\lambda$  é o vetor  $(n + p + 1) \times 1$ , dado por  $\lambda = (\lambda_1, \dots, \lambda_n, n_0, \dots, n_p)^\top$  e  $A$  é o vetor  $(n + p + 1) \times 1$ , dado por  $A = (C(s_1, s_0), \dots, C(s_1, s_n), 1, X_{01}, \dots, X_{0p})$ . Dessa forma, os pesos  $\lambda_i, i = 1, \dots, n$  são obtidos por  $\lambda = C^{-1}A$ .

### 3.5.8 Critérios de Seleção de Modelos

De acordo com Cressie (2015) após a seleção dos modelos teóricos de correlação espacial, torna-se necessário avaliar sua eficácia na interpolação de valor, a fim de possibilitar a construção de mapas temáticos com estimativas confiáveis.

A seguir são apresentados os critérios utilizados neste trabalho.

#### 3.5.8.1 Validação Cruzada

De acordo com Isaaks e Srivastava (1989), a Validação Cruzada é uma técnica fundamental em geoestatística para avaliar os erros de estimação, permitindo comparar valores previstos com valores amostrados. Esse método envolve a exclusão temporária de uma amostra específica  $Z(s_i)$ , para estimar o valor nessa localização utilizando *krigagem*, denotado como  $\hat{Z}(s_i)$ , com as demais amostras. Esse processo simula uma estimação sem realizar amostragem naquela localização. Após a estimativa,  $\hat{Z}(s_i)$  é comparado com o valor inicialmente removido, repetindo o procedimento para todas as amostras (FARACO *et al.*, 2008).

É esperado que os erros de estimação (Equação 31) apresentem média nula, variância constante e distribuição normal de probabilidade.

$$\varepsilon_i = \varepsilon(s_i) = Z(s_i) - \hat{Z}(s_i), \quad (\text{Eq. 31})$$

Em que,  $\hat{Z}(s_i)$  é o valor estimado por *krigagem* sem a  $i$ -ésima observação  $Z(s_i)$ .

Medidas como erro médio ( $EM$ ), erro médio reduzido ( $\overline{ER}$ ), desvio padrão dos erros médios ( $S_{EM}$ ), desvio padrão dos erros reduzidos ( $S_{ER}$ ) e erro absoluto ( $EA$ ) são usadas para comparação (MCBRATNEY & WEBSTER, 1986; CRESSIE, 2015).

O erro médio ( $EM$ ) (Equação (31)) por Validação Cruzada é dado pela Equação 32:

$$EM = \frac{1}{n} \sum_{i=1}^n (Z(s_i) - \hat{Z}(s_i)), \quad (\text{Eq. 32})$$

Em que,  $n$  é o número de dados (valores amostrados);  $Z(\mathbf{s}_i)$  é o valor observado no ponto  $(\mathbf{s}_i)$ ;  $\hat{Z}(\mathbf{s}_i)$  é o valor estimado por *krigagem* no ponto  $(\mathbf{s}_i)$ , sem considerar a observação  $Z(\mathbf{s}_i)$ .

O erro médio reduzido ( $\overline{ER}$ ) é definido pela Equação 33:

$$\overline{ER} = \frac{1}{n} \sum_{i=1}^n ER_i = \frac{1}{n} \sum_{i=1}^n \left( \frac{Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i)}{\sigma(\hat{Z}(\mathbf{s}_i))} \right), \quad (\text{Eq. 33})$$

Em que,  $\sigma(\hat{Z}(\mathbf{s}_i))$  é o desvio padrão da *krigagem* no ponto  $(\mathbf{s}_i)$ , sem considerar a observação  $Z(\mathbf{s}_i)$ .

O desvio padrão dos erros reduzidos ( $S_{ER}$ ) é dado conforme a Equação 34:

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i)|}{\sigma(\hat{Z}(\mathbf{s}_i))}}, \quad (\text{Eq. 34})$$

De acordo com Faraco *et al.* (2008), os critérios para a escolha do melhor modelo ajustado incluem valores próximos a zero para o erro médio ( $EM$ ) e erro médio reduzido ( $\overline{ER}$ ), o menor valor para o desvio padrão do erro médio ( $S_{EM}$ ) e valor próximo a um para o desvio padrão do erro médio reduzido ( $S_{ER}$ ). O erro absoluto, por sua vez, é uma medida da magnitude dos erros na unidade da variável, onde se procura o menor valor para o erro absoluto.

Conhecendo os valores medidos e preditos por *krigagem* ordinária  $Z(\mathbf{s}_i)$  e  $\hat{Z}(\mathbf{s}_i)$ , respectivamente, define-se o erro absoluto  $EA$ , conforme Equação (35) (FARACO *et al.*, 2008).

$$EA = \sum_{i=1}^n |Z(\mathbf{s}_i) - \hat{Z}(\mathbf{s}_i)|. \quad (\text{Eq. 35})$$

### 3.5.8.2 Critérios de Informação de Akaike e Bayesiano de Schwarz

De acordo com AKAIKE (1973), o critério de informação de Akaike (AIC) visa demonstrar que, quando dois modelos têm níveis de ajuste similares aos dados, o

modelo mais simples tende a ter melhor desempenho na predição de novos dados. O AIC promove a simplicidade, penalizando a complexidade excessiva, e é definido pela Equação 36:

$$AIC = -2l(\boldsymbol{\theta}) + 2r, \quad (\text{Eq. 36})$$

Em que,  $l(\boldsymbol{\theta})$  é o logaritmo da função verossimilhança avaliado em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ ; em que  $\hat{\boldsymbol{\theta}} = (\hat{\beta}, \hat{\varphi}_1, \hat{\varphi}_2, \hat{\varphi}_3)^\top$ ,  $r$  é o número de parâmetros do modelo ajustado.

O AIC relaciona a função de verossimilhança a uma medida de informação perdida quando um modelo representa uma realidade aproximada, penalizando a adição de parâmetros, levando em consideração a parcimônia. A escolha do melhor modelo é feita pelo menor valor de AIC.

O critério de informação bayesiano de Schwarz (BIC) (SCHWARZ, 1978) é uma técnica para comparar e selecionar modelos estatísticos, baseada na verossimilhança dos dados e na complexidade dos modelos. O BIC é definido conforme a Equação 37:

$$BIC = -2l(\boldsymbol{\theta}) + r \log(n), \quad (\text{Eq. 37})$$

Em que,  $l(\boldsymbol{\theta})$  é o logaritmo da função verossimilhança avaliado em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , que é o estimador de máxima verossimilhança do modelo;  $r$  é o número de parâmetros do modelo ajustado e  $n$  é o número de dados.

O BIC favorece modelos mais simples e parcimoniosos. O modelo com o menor valor de BIC é considerado o mais adequado para os dados.

### 3.6 ÍNDICE DE DEPENDÊNCIA ESPACIAL – SDI

A mensuração do grau de dependência espacial dos modelos ajustados é obtida utilizando o Índice de Dependência Espacial *SDI* para diversos modelos (Equação (38)) desenvolvido por Seidel e Oliveira (2016).

$$SDI = MF \left( \frac{\varphi_2}{\varphi_1 + \varphi_2} \right) \min \left\{ 1; \left( \frac{a}{0,5MD} \right) \right\} 100, \quad (\text{Eq. 38})$$

Em que,  $MF$  é o fator do modelo específico para cada modelo do semivariograma,  $\varphi_1$  é o efeito pepita,  $\varphi_2$  é a contribuição,  $a$  é o alcance e  $MD$  é a distância máxima calculada entre todos os pares de pontos amostrados.

Na Tabela 3, é apresentada a classificação do Índice de Dependência Espacial (SDI) (SEIDEL & OLIVEIRA, 2016; NETO *et al.*, 2018) considerando o modelo de função semivariância *Wave*.

Tabela 3: Classificação do Índice de Dependência Espacial – SDI

Modelo	MF	Fraco	Moderado	Forte
<i>Wave</i>	0,58900	SDI ≤11%	11% <SDI ≤24%	SDI >24 %

Fonte: (SEIDEL; OLIVEIRA, 2016; NETO *et al.*, 2018)

### 3.6.1 Diagnóstico de Influência Local

O objetivo do estudo de diagnóstico em geoestatística é identificar observações que podem afetar a escolha do modelo geoestatístico, a estimação de parâmetros de dependência espacial, a inferência de parâmetros por máxima verossimilhança (ML) e a interpolação por *krigagem*. Esses pontos são chamados de observações influentes, pois exercem uma influência desproporcional nas estimativas do modelo ou na significância dos parâmetros. As técnicas de diagnóstico de influência local consideram o efeito conjunto de observações nas estimativas do modelo.

De acordo com Cook (1986), um caso que apresenta influência local também pode apresentar influência global, mas o contrário não é necessariamente verdadeiro. O método de influência local proposto por Cook consiste em avaliar, por meio de uma medida apropriada de influência, a robustez das estimativas fornecidas pelo modelo mediante pequenas perturbações aplicadas no modelo ou aos dados. Isso verifica a existência de pontos que, sobre pequenas perturbações no modelo, podem causar grandes distorções nos resultados dos estimadores de ML, sem a necessidade da eliminação do conjunto de dados.

Para avaliar a influência local, Cook (1986) utilizou uma medida de desvio da verossimilhança, conforme a Equação 39:

$$LD(\omega) = 2\{L(\hat{\theta}) - L(\hat{\theta}|\omega)\}, \quad (Eq. 39)$$

Em que,  $LD(\boldsymbol{\omega}) \geq 0$  e  $\hat{\boldsymbol{\theta}}$  e  $\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}$  são as estimativas de máxima verossimilhança dos modelos não perturbado  $L(\boldsymbol{\theta})$  e perturbado  $L(\boldsymbol{\theta}|\boldsymbol{\omega})$ , respectivamente. Note que existe  $\boldsymbol{\omega}_0$  tal que  $LD(\boldsymbol{\omega}_0) = 0$ .

A avaliação dessa medida envolve a construção de um gráfico de  $LD(\boldsymbol{\omega})$  em função de  $\boldsymbol{\omega}$ . Esse gráfico fornece informações cruciais sobre a influência do esquema de perturbação utilizado (GALEA; PAULA; URIBE-OPAZO, 2003).

A análise do gráfico ocorre por meio de uma superfície geométrica  $(q + 1)$ -dimensional formada pelo vetor:

$$\alpha(\boldsymbol{\omega}) = \begin{pmatrix} \boldsymbol{\omega} \\ LD(\boldsymbol{\omega}) \end{pmatrix}, \quad (\text{Eq. 40})$$

Quando  $\boldsymbol{\omega}$  varia em  $\Omega$ , denomina-se essa superfície como gráfico de influência.

O estudo de influência local se concentra na análise de como a superfície  $\alpha(\boldsymbol{\omega})$  se desvia do seu plano tangente  $T_0$  ao redor de  $\boldsymbol{\omega}_0(T_0)$ . Essa análise envolve o estudo das curvaturas das seções normais da superfície  $\alpha(\boldsymbol{\omega})$  em  $\boldsymbol{\omega}_0$ , que são conhecidas como curvaturas normais (COOK, 1986).

As seções normais são obtidas através das interseções entre a superfície  $\alpha(\boldsymbol{\omega})$  e os planos que contém o vetor normal ao seu plano tangente em  $\boldsymbol{\omega}_0$  (COOK, 1986). Essa abordagem permite uma compreensão mais profunda da influência local e das perturbações na análise dos dados.

Para gerar a curvatura normal, têm-se o vetor  $\boldsymbol{l}$  de comprimento unitário em uma direção qualquer em  $\mathbb{R}^q$ , e uma reta em  $\Omega \subset \mathbb{R}^q$  que passa por  $\boldsymbol{\omega}_0$ . Assim, temos (COOK, 1986):

$$\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a\boldsymbol{l}, \quad (\text{Eq. 41})$$

Com  $a \in \mathbb{R}$ , dessa forma, têm-se uma linha projetada sobre a superfície  $\alpha(\boldsymbol{\omega})$  que passa pelo ponto  $\alpha(\boldsymbol{\omega}_0)$ . Essa linha pode também ser obtida pelo gráfico  $LD(\boldsymbol{\omega}_0 + a\boldsymbol{l})$  versus  $a \in \mathbb{R}$ .

A curvatura da normal da linha projetada, indicada por  $C_l$ , é definida como sendo a curvatura de  $(a, LD\{\boldsymbol{\omega}(a)\})$ , sendo  $a = 0$ , em que  $\boldsymbol{\omega}(a) = \boldsymbol{\omega}_0 + a\boldsymbol{l}$ , ou seja,  $C_l$  é uma curvatura normal da superfície  $\alpha(\boldsymbol{\omega})$  em torno de  $\boldsymbol{\omega}_0$  e do vetor unitário  $\boldsymbol{l}$ . Recomenda-se considerar a direção  $\boldsymbol{l}_{\max}$  correspondente à maior curvatura  $C_l$ .

Cook (1986) propôs analisar a influência local pelo uso da curvatura normal na direção do vetor unitário  $\mathbf{l}$ , dado por:

$$C_l(\boldsymbol{\theta}) = 2|\mathbf{l}^\top \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta} \mathbf{l}|, \quad (\text{Eq. 42})$$

Em que:  $\|\mathbf{l}\| = 1$ ;  $\mathbf{L}$ : é a matriz de informação observada avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ ;  $\boldsymbol{\Delta}$ : é a matriz de ordem  $(p + q) \times n$  dada por  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\varphi^\top)^\top$ , avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e em  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ .

### 3.6.1.1 Perturbação de Zhu generalizada

De Bastiani *et al.* (2015) propuseram uma perturbação generalizada  $\mathbf{Z}_\omega = \mathbf{Z} + \mathbf{A}_\omega$ , na variável resposta utilizando a matriz  $\mathbf{A}$  de dimensão  $n \times n$ , que não depende de  $\boldsymbol{\beta}$  e nem de  $\boldsymbol{\omega}$ , tal que a matriz de informação esperada de Fisher para  $Z_\omega$  com respeito ao vetor perturbado  $\boldsymbol{\omega}$  é  $G(\boldsymbol{\omega}_0) = c\mathbf{A}^\top \boldsymbol{\Sigma}^{-1} \mathbf{A}$ , com  $c$  uma constante positiva. Em geral  $\mathbf{A}^\top \boldsymbol{\Sigma}^{-1} \mathbf{A} \neq \mathbf{I}_n$ , porém, se  $\mathbf{A} = \boldsymbol{\Sigma}^{\frac{1}{2}}$ , então  $G(\boldsymbol{\omega}_0) = c\mathbf{I}_n$  sendo, a perturbação apropriada  $\mathbf{Z}_\omega = \mathbf{Z} + \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\omega}$ .

Neste estudo, a matriz  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^\top, \boldsymbol{\Delta}_\varphi^\top)^\top$  avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e em  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ , é da forma:

$$\begin{aligned} \boldsymbol{\Delta}_\beta &= -\mathbf{X}^\top \boldsymbol{\Sigma}^{-1} \mathbf{A}, \\ \boldsymbol{\Delta}_\varphi &= \frac{\partial^2 l(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\omega}^\top}, \end{aligned}$$

Que tem como elementos:

$$\Delta_{\varphi j} = \boldsymbol{\varepsilon}^\top \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{A}}{\partial \varphi_j} - \mathbf{D}_j \mathbf{A} \right], \text{ para } j = 1, \dots, q,$$

$$\text{Com } \boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{\omega}_0) = (\mathbf{Z} - \boldsymbol{\beta} \mathbf{1}), \text{ e } \mathbf{D}_j = \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1}.$$

Considere a matriz:

$$\mathbf{B} = \boldsymbol{\Delta}^\top \mathbf{L}^{-1} \boldsymbol{\Delta}, \quad (\text{Eq. 43})$$

Em que  $\Delta = (\Delta_{\beta}^{\top}, \Delta_{\phi}^{\top})^{\top}$  e  $L$  é a matriz de informação observada, avaliada em  $\theta = \hat{\theta}$ , e o elemento  $C_i = 2|b_{ii}|$ , para  $i = 1, \dots, n$ , em que  $b_{ii}$  é o elemento da diagonal principal da matriz  $B$ .

A partir da matriz  $B$  pode-se utilizar o gráfico de  $C_i$  versus  $i$  (ordem dos dados) como técnica de diagnóstico para avaliar a existência de observações influentes.

Seja  $L_{max}$  o autovetor, normalizado, associado ao maior autovalor, em módulo, da matriz  $B$ . Segundo Cook (1986) o gráfico dos elementos  $|L_{max}|$  versus  $i$  (ordem dos dados) pode revelar os elementos que, sob pequenas perturbações, exerce maior influência em  $LD(\omega)$ , na vizinhança de  $\omega_0$ .

### 3.7 COMPARAÇÃO DE MAPAS

O critério de similaridade entre mapas interpolados é essencial para avaliar sua qualidade. O índice Kappa  $\hat{K}$  (COHEN, 1960) compara a concordância entre dois mapas por meio da Equação 44:

$$\hat{K} = \frac{n \sum_{i=1}^k n_{ii} - \sum_{i=1}^k (n_{i\#} n_{\#i})}{n^2 - \sum_{i=1}^k (n_{i\#} n_{\#i})}, \quad (Eq. 44)$$

Em que  $n_{i\#} = \sum_{j=1}^k n_{ij}$  é quantidade de área da linha  $i$ ,  $n_{\#i} = \sum_{u=1}^n n_{ui}$  é a quantidade de área da coluna  $i$ ,  $n_{ii}$  é a quantidade de área da linha  $i$  e coluna  $i$ .

No entanto, o Índice Kappa  $\hat{K}$  trata todas as discordâncias igualmente, sem considerar suas intensidades. Em contraste o índice Kappa Ponderado ( $\hat{K}_w$ ) (Equação 45) (COHEN, 1968) atribui pesos ( $w_{ij}$ ) às discordâncias com base na magnitude das diferenças entre as classes dos mapas.

$$\hat{K}_w = \frac{\sum_{i=1}^r \sum_{j=1}^r w_{ij} P_{ij} - \sum_{i=1}^r \sum_{j=1}^r w_{ij} P_{i\#} P_{\#j}}{1 - \sum_{i=1}^r \sum_{j=1}^r w_{ij} P_{i\#} P_{\#j}}, \quad (Eq. 45)$$

Em que,  $w_{ij} = 1 - |i - j| / (r - 1)$  são os pesos lineares (CICCHETTI & ALLISON, 1971) para uma matriz de contingência com  $r$  classes, para  $i, j = 1, \dots, r$ ;  $P_{ij}$  são as proporções conjuntas, ou seja, a contagem de *pixels* em cada célula da

matriz de contingência dividida por  $N$ , o número total de *pixels* do mapa, para  $i, j = 1, \dots, r$ ;  $P_{i\#} = n_{i\#}/N$  são as proporções marginais, em que  $n_{i\#}$  representa a soma dos *pixels* da linha  $i$  da matriz de contingência,  $i = 1, \dots, r$ .  $P_{\#j} = n_{\#j}/N$  são as proporções marginais, em que  $n_{\#j}$  representa a soma dos *pixels* da coluna  $j$  da matriz de contingência,  $j = 1, \dots, r$ .

A interpretação da magnitude do índice Kappa Ponderado  $\hat{K}_w$  é equivalente à do índice Kappa  $\hat{K}$  (FLEISS; LEVIN; PAIK, 2003) em que,  $\hat{K}_w \geq 0,75$  indica uma alta similaridade,  $0,4 < \hat{K}_w < 0,75$  indica moderada similaridade e  $\hat{K}_w \leq 0,4$  indica uma fraca ou insuficiente similaridade.

#### 4 REFERÊNCIAS BIBLIOGRÁFICAS

ABREU, C. A.; LOPES, A. S.; SANTOS, G. C. G. Micronutrientes. In: NOVAIS, R. F.; ALVAREZ, V. H.; FONTES, R. L. F.; CANTARUTTI, R. B.; NEVES, J. C. L. **Fertilidade do solo**. Viçosa, MG, Sociedade Brasileira de Ciência do Solo, p. 645-736, 2007.

AEN – AGÊNCIA ESTADUAL DE NOTÍCIAS. Com colheita avançada, relatório confirma maior safra de soja do Paraná. **Governo do Estado do Paraná**, 2021. Disponível em: <https://www.aen.pr.gov.br/Noticia/Com-colheita-avancada-relatorio-confirma-maior-safra-de-soja-da-historia-do-Parana>. Acesso em: 13 jul. 2023.

AEN – AGÊNCIA ESTADUAL DE NOTÍCIAS. Soja ocupa um quarto do território estadual e é exportada para mais de 20 países. **Governo do Estado do Paraná**, 2023. Disponível em: <https://www.aen.pr.gov.br/Noticia/Soja-ocupa-um-quarto-do-territorio-estadual-e-e-exportada-para-mais-de-20-paises>. Acesso em: 13 jul. 2023.

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. **Journal of the Royal Statistical Society**, v. 51, p. 469-483, 1973.

ALMEIDA, A. Q. de; RIBEIRO, A.; PAIVA, Y. G.; RASCON, N. J. L.; LIMA, E. P. Geoestatística no estudo de modelagem temporal da precipitação. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 15, n. 4, p. 354-358, 2011.

BARRETO, L. N. Determinantes da criminalidade no município de Vitória: um estudo utilizando a Teoria econômica da escolha racional. 2016. 37 f. **Dissertação (Mestrado – Programa de Pós-graduação em Segurança Pública)** – Universidade de Vila Velha, Vila Velha, 2016. Disponível em: <https://repositorio.uvv.br/handle/123456789/636>>. Acesso em: 17 nov. 2023.

BASAK, B. B.; SARKAR, B.; MAITY, A.; CHARI, M. S.; BANERJEE, A.; BISWAS, D. R. Low-grade silicate minerals as value-added natural potash fertilizer in deeply weathered tropical soil. **Geoderma**, v. 433, mai. 2023. <https://doi.org/10.1016/j.geoderma.2023.116433>

BOOTE, K.J.; JONES, J. W.; HOOGENBOOM, G. Simulation of crop growth: CROPGRO model. **Agricultural systems modeling and simulation**, v. 18, p. 651-692, 1998.

BRASIL. Presidência da República. **Decreto-Lei nº 764, de 15 de agosto de 1969**. Autoriza a constituição da sociedade por ações Companhia de Pesquisa de Recursos Minerais - CPRM e dá outras providências. Brasília, 1969. Disponível em: <https://legislacao.presidencia.gov.br/atos/?tipo=DEL&numero=764&ano=1969&ato=5c3cXSU9EMjRVTb70>. Acesso em: 23 jun. 2024.

BRASIL. Presidência da República. Decreto nº 66.522 de 30 de abril de 1970. Regulamenta os arts 24 e 25 do decreto-lei 764, de 15/08/1969. Brasília, 1970. Disponível em: <https://legislacao.presidencia.gov.br/atos/?tipo=DEC&numero=66522&ano=1970&ato=ea1QTVE9UMjRVT111>. Acesso em: 23 jun. 2024.

BURGESS, T.M.; WEBSTER, R. Ótima em terpolação e mapeamento isarítmico de solo propriedades. I. O semivariograma e a *krigagem* pontual. **The Journal of Soil Science, Oxford**, 31: 315-331, 1980.

CANARACHE, A. PENETR - A generalized semi-empirical model estimating soil resistance to penetration. **Soil and Tillage Research**, v. 16, n. 1, p. 51-70, 1990.

CANCIAN, M. Aplicação de cobre na cultura da soja em solos com altos teores de fósforo. 2018. **Dissertação (Mestrado - Programa de Pós-graduação em Ciência do Solo)** - Universidade Federal de Santa Maria, Santa Maria, 2018. Disponível em: <https://repositorio.ufsm.br/handle/1/13823>. Acesso em: 12 out. 2023.

CARVALHO, J. R. P. de; VIEIRA, S. R.; GREGO, C. R. Comparação de métodos para ajuste de modelos de semivariograma da precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n. 4, p. 443-448, 2009.

CHICHETTI, D. V.; ALLISON, T. A new procedure for assessing reability of scoring EEG sleep recordings. **American Journal of Electroneurodiagnostic Technology**, v. 11, n. 3, p. 101-110.

CHILÈS, J.; DELFINER, P. **Geostatistics: Modeling Spatial Uncertainty**. Wiley, New York, p. 695, 1999. <https://doi.org/10.1002/9780470316993>

CLARK, I. **Practical Geostatistics**. Applied Science Publishers LTD, London, 1979.

COHEN, J. Coefficient of areement for nominal scales. **Educational and Psychological Measurement**, v.20, n.1, p.37-46, 1960.

COHEN, J. Weighted kappa: nominal scale agrément whit provision for scaled disagreement or partial credit. **Psychological Bulletin**, v. 70, n. 4, p. 213-220, 1968.

COLARES, GL; REINERT, DJ; REICHERT, JM; KAISER, D. Compactação de um Latossolo induzida pelo tráfico de máquinas e sua relação com o crescimento e

produtividade de feijão e trigo. **Revista Brasileira de Ciência do Solo**, Viçosa-MG, v.32, n.3, p. 933-942, 2008.

COOK, R. D. Assessment of Local Influence. **Journal of the Royal Statistical Society**, v. 48, n.2, p. 133-169, 1986.

CRESSIE, N. A. C; HAWKINS, D. M. Robust estimation oh the variogram. **Mathematical Geology**, v. 12, n. 2, 1980.

CRESSIE, N. A. C. **Statistics for Spatial Data**. John Wiley & Sons, Hoboken, 2015.

DALL'AGNOL, A. **A Embrapa Soja no contexto do desenvolvimento da soja no Brasil: histórico e contribuições**. Brasília, DF: Embrapa, 2016.

DALCHIAVON, F. C.; CARVALHO, M. P.; ANDREOTTI, M.; MONTANARI, R. Variabilidade espacial de atributos da fertilidade de um Latossolo Vermelho Distoférico sob Sistema de Plantio Direto. **Revista Ciência Agronômica**, n. 43, n. 3, p. 453-461, 2012.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; DE BASTIANI, F. Spatial-temporal Analysis of Soybean Productivity Using Geostatistical Methods. **Journal of Agricultural Studies**, v. 9, n. 2, 2021.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; OLIVEIRA, M. P. Hole effect na variabilidade espacial da produtividade da soja. In: LI Congresso Brasileiro de Engenharia Agrícola. **Anais [...]**, 2022.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; OLIVEIRA, M. P. Comparison between Matheron and Genton semivariance function estimators in spatial modeling of soybean yield. **Australian Journal of Crop Science**, v. 16, n. 7, 2022.

DERAL – Departamento de Economia Rural. Boletim Semanal – 09/2022. **Governo do Estado do Paraná**, 2022.

DEUTCH, C. V.; JOURNAL, A. G. Geostatistical Software Library and User's Guide. **Geological Magazine**, v. 136, n. 1, p. 83-108, 1999.

DE BASTIANI, F.; URIBE-OPAZO, M. A.; DALPOSSO, G. H. Comparison of maps of spatial variability of soil resistance to penetration constructed with and without covariables using a spatial linear model. **Engenharia Agrícola**, Jaboticabal, v. 32, n. 2, p. 393-404, 2012.

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

Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA. Soja em números (safra 2022/23). Disponível em: < <https://www.embrapa.br/soja/cultivos/soja1/dados-economicos>>. Acesso em: 16 nov. 2023.

ELAOU, A.; HASSEN, H. B.; JALEL, R.; SALAH, N. B.; MASMOUDI, A.; MASMOUDI, A. Machine learning approach for predicting soil penetration resistance under different moisture conditions. **Journal of Terramechanics**, v. 110, p. 39-45, 2023. <https://doi-org.ez89.periodicos.capes.gov.br/10.1016/j.jterra.2023.08.002>

FAGUNDES, R. S.; URIBE-OPAZO, M. A.; GUEDES, L. P. C.; GALEA, M. Spatial linear modeling: soybean yield variability as a function of soil chemical properties. **Revista Brasileira de Ciência do Solo**, v. 42, p. e0170030, 2018.

FANCELLI, A. L. Influência da nutrição de plantas na ocorrência de doenças e pragas. In: FANCELLI, A. L.; DOURADO NETO, D. (Ed.), *Feijão irrigado – Tecnologia de produtividade*, Piracicaba: ESALQ/USP, p. 1-29, 2003.

FARACO, M.A.; URIBE-OPAZO, M. A.; SILVA, E. A. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do Solo**, v. 32, n. 2, 2008.

FIRMANO, R. F.; MELO, T. R.; JORDÃO, L. T.; FREGONEZI, G. A. F.; BARZAN, R. R. Fontes de cálcio e magnésio nos atributos químicos de um latossolo e na produtividade da soja. **Scientia Agraria Paranaensis**, v. 16, n. 1, p. 27-31, jan./mar., 2017.

FLEISS, J.; LEVIN, B.; PAIK, M. The measurement of interrater agreement. **Statistical Methods for Rates and Proportions**, 3ed, cap. 18, Wiley Online Library: Hoboken, NJ, USA, 2003.

GALEA, M.; PAULA, G. A.; URIBE-OPAZO, M. A. On influence diagnostic in univariate elliptical linear regression models. **Springer**, v. 44, p. 23-45, 2003.

GARCIA-PAPANI, F., LEIVA, V., URIBE-OPAZO, M. A., AYKROYD, R. G. Birnbaum-Saunders spatial regression models: Diagnostics and application to chemical data. **Chemometrics and Intelligent Laboratory Systems**, v. 177, p. 114-128, 2018.

GAZZONI, D. L. A soja no Brasil é movida por inovações tecnológicas. **Ciência e Cultura**, v. 70, n. 3, p. 16-18, 2018.

GAUTAM, S. S.; GONDWAL, M.; SONI, R.; GAUTAM, B. P. S. **Trends of Applied Microbiology for Sustainable Economy: Developments in Applied Microbiology and Biotechnology**. Elsevier, 2022. <https://doi-org.ez89.periodicos.capes.gov.br/10.1016/C2020-0-02717-9>

GENTON, M. G. Variogram Fitting by Generalized Least Squares Using an Explicit Formula for the Covariance Structure. **Mathematical Geology**, v. 30, n. 4, 1998.

GRADSHTEYN, I. S.; RYZHIK, I. M. **Table of integrals, series, and products**. Academic Press, 7ed., 2000.

GREGO, C. R.; OLIVEIRA, R. P.; VIEIRA, S. R. Geoestatística aplicada a Agricultura de Precisão. In: BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V. de;

BASSOI, L. H.; INAMASU, R. Y. (Ed.). **Agricultura de Precisão: resultados de um novo olhar**. Brasília, DF: Embrapa, cap. 5, p. 74-83, 2014.

GRZEGOZEWSKI, D.M. Influência Local para Modelos Geoestatísticos Utilizando a Produtividade da Soja e Atributos Químicos do Solo. 2012. 121 f. **Dissertação (Mestrado - Programa de Pós-graduação em Engenharia Agrícola)** - Universidade Estadual do Oeste do Paraná, Cascavel, 2012. Disponível em: <https://tede.unioeste.br/bitstream/tede/2878/1/Denise.pdf>. Acesso em: 9 out. 2023.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; JOHANN, J. A.; SOUZA, E. G. de. Anisotropia no estudo da variabilidade espacial de algumas variáveis químicas do solo. **Rev Bras Ciênc Solo**, v. 32, n. 6, 2217-2226, 2008.

HENGL, T., HEUVELINK, G. B., STEIN, A. **Comparison of kriging with external drift and regression kriging**, Enschede, Netherlands: ITC, 2003.

HENGL, T. **A practical guide to geostatistical mapping**. [S.l.]: Hengl, v. 52, 2009.

INAMASU, R. Y.; BERNARDI, A. C. de C.; VAZ, C. M. P.; NAIME, J. de M.; QUEIROS, L. R.; RESENDE, A. V. de; VILELA, M. de F.; JORGE, L. A. de C.; BASSOI, L. H.; PEREZ, N. B.; FRAGALLE, E. P. **Agricultura de Precisão: um novo olhar**. São Carlos, SP: Embrapa Instrumentação, 2011.

Instituto Brasileiro de Geografia Estatística – IBGE. Em julho, IBGE prevê safra de 308,9 milhões de toneladas para 2023. Disponível em: < <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/37607-em-julho-ibge-preve-safra-de-308-9-milhoes-de-toneladas-para-2023>>. Acesso em: 27 jun. 2024.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied Geostatistics**. New York: Oxford University Press, 1989.

JOURNAL, A. G.; HUIJBREGTS, C. H. J. **Mining geostatistics**. Academic Press, New York, p. 600, 1978.

KELLER, T. M.; COLOMBI, T. R.; HORN, D. O. Historical increase in agricultural Machinery weights enhanced soil stress levels and adversely affected soil functioning. **Soil and Tillage Research**, v. 194, 2019.

KUMI, F.; OBOUR, P. B.; ARTHUR, E.; MOORE, S. E.; ASARE, P. A.; ASIEDU, J.; ANGNUURENG, D. B.; ATIAH, K.; AMOAH, K. K.; AMPONSAH, S. K.; DORVLO, S. Y.; BANAFU, S.; ADU, M. O. Quantifying root-induced soil strength, measured as soil penetration resistance, from different crop plants and soil types. **Soil and Tillage Research**, v. 233, 2023. <https://doi-org.ez89.periodicos.capes.gov.br/10.1016/j.still.2023.105811>

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

LANGE, A.; CAVALLI, E.; PEREIRA, C. S.; CHAPLA, M. V.; FREDDI, O. S. Relações cálcio:magnésio e características químicas do solo sob cultivo de soja e milho. **Pesquisas Agrárias e Ambientais**, Sinop, v. 9, n. 3, p. 294-301, 2021.

LIMA, C. L. R.; REINERT, D. J.; REICHERT, J. M.; SUZUKI, L. E. A. S. Produtividade de culturas e resistência à penetração de Argissolo Vermelho sob diferentes manejos. **Pesquisa Agropecuária Brasileira**, Brasília, v. 45, n. 1, p. 89-94, 2010.

LI, D.; LAKE, L. W. A moving window semivariance estimator. *Water Resources Research*, **Austin**, v. 30, n. 5, p. 1479-1489, 1994.

LOPES, A. S. **Manual internacional de fertilidade do solo**. 2.ed. rev. amp.; Piracicaba: Potafos, 1998.

MALAVOLTA, E.; VITTI, G. C.; OLIVEIRA, S. A. **Avaliação do estudo nutricional das plantas: princípios e aplicação**. 2 ed. Piracicaba: POTAFOS, p. 319, 1997.

MALTAURO, T. C. Estudo de algoritmos de otimização multivariados para a determinação de configuração amostral e tamanho amostral na análise da variabilidade espacial. 2022, 66 f. **Tese (Doutorado – Programa de Pós-graduação em Engenharia Agrícola)** – Universidade Estadual do Oeste do Paraná, Cascavel, 2022. Disponível em: <<https://tede.unioeste.br/bitstream/tede/6336/5/Tamara%20Cant%C3%BA%20Maltauro.pdf>>. Acesso em: 27 jun. 2024.

MALTAURO, T. C.; GUEDES, L. P. C.; CANTON, L. E. D.; URIBE-OPAZO, M. A. Spatial multivariate optimization for a sampling redesign with a reduced sample size of soil chemical properties. **Revista Brasileira De Ciencia Do Solo**, v. 47, p. e0220072, 2023.

MARDIA, K. V.; MARSHAL, R. J. Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. **Biometrika**, v. 71, n.1, p. 135-146, 1984.

MATHERON, G. Principles of geostatistics. **Economic Geology**, v. 58, n. 8, p. 1246-1266, 1963.

MATÉRN, B. Spatial Variation. **Lecture notes in Statistics**, 2ed. New York: Springer, 1986.

MCBRATNEY, A. B.; WEBSTER, R. Choosing functions for semi-variograms of soil Properties and fitting them to sampling estimates. **Journal Soil Science**, v. 37, p. 617-639, 1986.

MCBRATNEY, A.B.; WHELAN, B.M.; ANCEV, T.; BOUMA, J. Future directions of Precision Agriculture. **Precision Agriculture**, v. 6, n. 1, p. 1-17, 2005. <http://dx.doi.org/10.1007/s11119-005-0681-8>.

MEDEIROS, F. J.; LUCIO, P. S.; SILVA, H. J. F. Análise de métodos de *krigagem* na estimativa da precipitação no estado do Rio Grande do Norte. **Revista Brasileira de Geografia Física**, v. 10, n. 5, p. 1668-1676, 2017.

MELLONI, R.; PEREIRA, E. G. M.; ALVARENGA, M. I. N. Indicadores da qualidade do solo. **Informe Agropecuário**, Belo Horizonte, v. 29, n. 244, p. 17-29, 2008.

MOTA, V. C. Geoestatística bidimensional como alternativa à modelagem de series temporais sob cenários de mudanças climáticas. **Pubvet**, v. 15, n. 2, p. 1-9, 2021. <https://doi.org/10.31533/pubvet.v15n02a745.1-9>

MENDES, A. M. S. **Introdução a Fertilidade do Solo**. Salvador, BA: Superintendência Federal de Agricultura, Pecuária e Abastecimento do Estado da Bahia – SFA -BA/SDC/MAPA, 2007.

MITCHELL, A. F. S. The information matrix, skewness tensor and  $\alpha$ -connections for the general multivariate elliptic distribution, **Annals of the Institute of Statistical Mathematics**, v. 41, p. 289-304, 1989.

NASCIMENTO do. J. A. M.; SOUTO, J. S.; CAVALCANTE, L. F.; MEDEIROS, S. A. S.; PEREIRA, W. E. Produção de melancia em solo adubado com esterco bovino e potássio. **Revista Brasileira de Ciências Agrárias**, v. 12, n. 2, p. 122-127, 2017.

NETO, E. A.; BARBOSA, I. C.; SEIDEL, E. J.; OLIVEIRA, M. S. Spatial dependence index for cubic, pentaspherical and Wave semivariogram models. **Bulletin of Geodetic Sciences**, v. 24, n. 1, p. 142–151, 2018. <https://doi.org/10.1590/s1982-21702018000100010>.

NOBILE, F. O.; ANUNCIAÇÃO, M. G. Dinâmica do cobre no solo. Campo & Negócios, 2021. Disponível em: < <https://revistacampoenegocios.com.br/dinamica-do-cobre-no-solo/>>. Acesso em: 12 out. 2023.

NOVAES, R. F.; ALVAREZ, V. H.; BARROS, N. F.; FONTES, R. L. F.; CANTARUTTI, R. B.; NEVES, J. C. L. (eds.) Fertilidade do solo: fatores que influenciam o crescimento e desenvolvimento das plantas. Viçosa. **Sociedade Brasileira de Ciência do Solo**, p. 375-470, 2007.

NUNES, P. A.; MORAES, M. L.; ROSSONI, E. A. Eficiência da agricultura familiar nos municípios paranaenses. **Economia Ensaios**, Uberlândia, v. 34, n. 2, jan./jun. 2020.

OLEA, R.A. A six-step practical approach to semivariogram modeling. **Stochastic Environmental Research and Risk Assessment**, v. 20, n. 5, p. 307-318, 2006.

OLIVEIRA, E. F. **Treinamento**: fertilidade do solo e nutrição das plantas. Cascavel: Cooperativa Central de Pesquisa Agrícola – COODETEC, 2007.

PYRCZ, M.; DEUTSCH, C. The whole effect story on the hole effect. **Geostatistical Association of Australasia Newsletter**, v. 18, p. 3-5, 2003.

SANTOS, D. R.; SILVA, L. S. **Fertilidade do solo e nutrição de plantas**. Santa Maria: Ed. UFSC, 2001.

SANTOS, H.G.; JACOMINE, P.K.T.; ANJOS, L.H.C.; OLIVEIRA, V.A.; LUMBRERAS, J.F.; COELHO, M.R.; ALMEIDA, J.A.; ARAÚJO FILHO, J.C.; OLIVEIRA, J.B.; CUNHA, T.J.F. **Sistema brasileiro de classificação de solos**. 5. ed. rev. ampl. Brasília, DF: Embrapa; 2018.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

SEDIYAMA, T. **Tecnologias de produção e usos da soja**. Londrina: Mecenaz, 2009.

SEIDEL, E. J.; OLIVEIRA, M. S. A. Classification for a geostatistical index of spatial dependence. **Revista Brasileira de Ciência do Solo**, v. 40, 2016.

SERAFIM, M. E.; ONO, F. B.; ZEVIANI, W. M.; NOVELINO, J. O.; SILVA, J. V. Umidade do solo e doses de potássio na cultura da soja. **Revista Ciência Agrônômica**, v. 43, n. 2, p. 222-227, 2012.

SHEMMER, R. C. Inferência e diagnóstico no modelo espacial linear t-student reparametrizado: aplicações a dados adricolas. 2021. 150 f. **Dissertação (Mestrado - Programa de Pós-graduação em Engenharia Agrícola)** - Universidade Estadual do Oeste do Paraná, Cascavel, 2021. Disponível em: [https://tede.unioeste.br/bitstream/tede/5451/5/Rosangela\\_Schemmer2021.pdf](https://tede.unioeste.br/bitstream/tede/5451/5/Rosangela_Schemmer2021.pdf). Acesso em: 14 out. 2024.

SILVA, M. O.; VELOSO, C. L.; NASCIMENTO, D. L.; OLIVEIRA, J.; PEREIRA, D. F.; COSTA, K. D. S. Indicadores químicos e físicos de qualidade do solo. **Brazilian Journal of Development**, Curitiba, v. 6, n. 7, p. 47838-47855, jun. 2020.

SOUSA, D. M. G.; LOBATO, E. **Correção do solo e adubação**. 2.ed. Planaltina: Embrapa Cerrados, p. 81-96, 2004.

SRIVASTAVA, R. M.; PARKER, H. M. Robust measures of spatial continuity. In: **Geoestatics**, on the 3rd. Geostatistical Congress, Holland: Armstrong, p. 295-308, 1989.

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., ASSUMPÇÃO, R. A. B. Influence diagnostics on a reparameterized t-Student spatial linear model. **Spatial Statistics**, p. 100481, 2021.

URIBE-OPAZO, M. A.; DALPOSSO, G. H.; GALEA, M.; JOHANN, J. A.; DE BASTIANI, F.; MOYANO, E. N. C.; GRZEGOZEWSKI, D. Spatial variability of wheat yield using the gaussian spatial linear model. **Australian Journal of Crop Science**, v. 17, p. 179-189, 2023.

VANDERHASSELT, A.; COLL, C.; D'HOSE, T.; CORNELIS, W. How time characteristics of subsoilers affect fuel consumption, penetration resistance and potato yield of a sandy loam soil. **Soil and Tillage Research**, v. 228, 2023. <https://doi-org.ez89.periodicos.capes.gov.br/10.1016/j.still.2022.105631>

VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F.; ALVAREZ, V. H.; SCHAEFER, G. R. (Ed.). Tópicos em ciência do solo. Viçosa: **Sociedade Brasileira de Ciência do Solo**. v. 1. p. 1-54, 2000.

VILLAR, M. L. P. **Manual de interpretação de análise de plantas e solos e recomendação de adubação**. Mato Grosso: Empresa Mato-Grossense de Pesquisa, Assistência e Extensão Rural - EMPAER-MT, 2007.

WACKERNAGEL, H. **Multivariate Geostatistics: an introduction with applications**. Germany Springer, 2003.

WEBSTER, R; OLIVER, M. A. **Geostatistics for environmental scientists**. 2 ed. Chichester: John Wiley & Sons, p. 312, 2007.

ZAIN, M.; MA, H.; RAHMAN, S. U.; NURUZZAMAN, M.; CHAUDHARY, S.; AZEEM, I.; MEHMOOD, F.; DUAN, A.; SUN, C. Nanotechnology in precision agriculture: Advancing towards sustainable crop production. **Plant Physiology and Biochemistry**, v. 206, p. 108244, 2024.

ZIGLIO, C. M.; MIYAZAWA, M.; PAVAN, M. A. Formas orgânicas e inorgânicas de mobilização do cálcio no solo. **Brazilian Archives of Biology and Technology**, v. 42, n. 2, 1999. <https://doi.org/10.1590/S1516-89131999000200016>

## 5 ARTIGOS

### 5.1 ARTIGO 1 - VARIABILIDADE ESPACIAL DA PRODUTIVIDADE DA SOJA E ATRIBUTOS DO SOLO: “*HOLE EFFECT*” E DIAGNÓSTICO LOCAL UTILIZANDO O MODELO DE *WAVE*

#### **Variabilidade espacial da produtividade da soja e atributos do solo: *hole effect* e diagnóstico local utilizando o Modelo de *Wave*<sup>1</sup>**

##### **Resumo**

A soja é uma das commodities de maior importância comercial em escala global. Dessa forma, estudos em Agricultura de Precisão (AP) são essenciais para maximizar o retorno econômico e minimizar os impactos ambientais. Nesse contexto, a realização de interpolações para mapear a produtividade da soja e os atributos do solo desempenha um papel crucial na tomada de decisões e devem representar uma estimativa eficiente das condições reais do campo. Entre os métodos existentes no estudo da AP, destaca-se a geoestatística, que utiliza informações sobre a estrutura de dependência espacial presente nos dados para gerar mapas interpolados com variância mínima. No entanto, a presença de algumas observações atípicas ou influentes podem distorcer a dependência espacial, questionando a eficiência dos mapas interpolados. Diante desse cenário, o intuito deste estudo foi desenvolver um diagnóstico de influência local na variável resposta, produtividade da soja, resistência do solo à penetração, teor de cobre e pH do solo, utilizando o modelo geoestatístico *Wave*, que caracteriza a dependência espacial presente nos dados, no caso da semivariância apresentar o efeito *hole effect*. Os resultados mostraram que a remoção de observações influentes provocou alterações nas estimativas dos parâmetros que definem a estrutura de dependência espacial e na construção dos mapas interpolados. Isso impacta a criação de zonas de manejo e/ou zonas de aplicação, considerando a lucratividade e os atributos do solo. Portanto, reforça-se a relevância do estudo da influência local em dados espaciais.

**Palavras-chave:** Análise de dados espaciais. Diagnóstico de influência local; Geoestatística; *krigagem*. Mapa de produtividade. Máxima verossimilhança.

##### **5.1.1 Introdução**

A Agricultura de Precisão (AP) é uma abordagem inovadora que visa otimizar o manejo das culturas agrícolas, levando em conta a variabilidade espacial e/ou temporal dos fatores que afetam a produção. Essa abordagem traz diversos benefícios potenciais, tais como: melhoria da qualidade das colheitas, redução do impacto ambiental, aumento da sustentabilidade, garantia da segurança alimentar,

---

<sup>1</sup>Artigo submetido ao periódico internacional Australian Journal of Crop Science, com classificação B2 no QUALIS/CAPES na área de Ciências Agrárias. Aceito em: 28/06/2024.

promoção do desenvolvimento econômico e elevação da rentabilidade e da produtividade (ZAIN *et al.*, 2024).

Para implementar a AP, é necessário utilizar métodos que permitam descrever e modelar a variabilidade espacial dos fenômenos naturais, bem como estimar valores em locais não amostrados. A geoestatística é um dos principais métodos científicos para esse fim, pois oferece adaptações de técnicas clássicas da estatística para aproveitar a continuidade espacial dos dados (URIBE-OPAZO *et al.*, 2023). Nesse contexto, a *krigagem* é um método de interpolação, considerado o melhor interpolador linear espacial de variância mínima (ISAACS & SRIVASTAVA, 1989).

A *krigagem* requer o ajuste de modelos teóricos ao semivariograma experimental, que é uma função que mede o grau de dependência espacial de uma variável georreferenciada (URIBE-OPAZO *et al.*, 2012). Os modelos mais comuns utilizados em geoestatística são os modelos Exponencial, Esférico, Gaussiano e da família Matérn (CRESSIE, 2015), que são monotônicos e crescentes, ou seja, aumentam continuamente com o aumento da distância entre pontos.

No entanto, em alguns casos, o semivariograma pode apresentar um comportamento não monotônico, ou seja, não aumenta continuamente com o aumento da distância, indicando que há valores altos cercados por valores baixos, gerando um movimento senoidal da função semivariância, sendo este efeito chamado de *hole effect*. Para esses casos, o Modelo de *Wave* é o mais indicado, pois consegue lidar de forma mais eficaz com a variação não monotônica do semivariograma (CARVALHO; VIEIRA; GREGO, 2009; DALPOSSO *et al.*, 2022).

Segundo Dalposso *et al.* (2022), negligenciar essas estruturas não monotônicas na análise pode resultar em interpretações errôneas, pois o modelo adotado pode não reproduzir os padrões de variabilidade espacial observados. Além disso, a presença de observações influentes pode alterar consideravelmente os resultados das análises de dependência espacial e, conseqüentemente, a construção dos mapas gerados pela *krigagem*.

Para avaliar se as observações podem causar distorções nos modelos de dependência espacial, Cook (1986) propôs causar pequenas perturbações no modelo ou nos dados. Esta técnica é conhecida na literatura como influência local que leva em consideração a diferenciação geométrica, e o afastamento da verossimilhança, que é empregado como procedimento para avaliar a influência das observações (URIBE-OPAZO *et al.*, 2012; DE BASTIANI *et al.*, 2015).

A AP tem se beneficiado de diversos estudos de análise de influência local, como os propostos por Uribe-Opazo *et al.* (2012, 2021, 2023), De Bastiani *et al.* (2015) e Dalposso *et al.* (2021). Esses estudos contribuem para o avanço do conhecimento na área. Porém, ainda não há trabalhos que abordem o diagnóstico de influência local utilizando o Modelo Geoestatístico de *Wave*.

Nesse contexto, o objetivo do estudo foi desenvolver e aplicar técnicas de diagnóstico da influência local na variável resposta ao empregar o Modelo de *Wave* para descrever a variabilidade espacial da produtividade da soja, determinante para a rentabilidade da cultura; teor de cobre no solo, que influencia o crescimento das plantas; pH do solo, que afeta a disponibilidade de nutrientes; e resistência à penetração do solo, que influencia o desenvolvimento das raízes. As técnicas de diagnóstico de influência local permitiram identificar observações influentes, destacando quais podem afetar as estimativas dos parâmetros, os valores previstos pelo modelo e a construção de mapas por meio da *krigagem*.

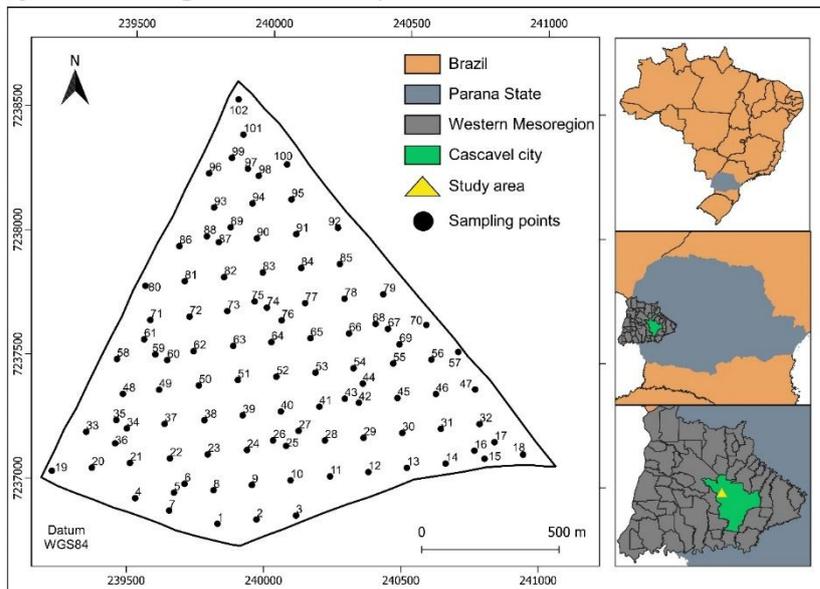
## 5.1.2 Material e métodos

### 5.1.2.1 Área de estudos de dados

Os dados de produtividade da soja (Prod) [ $t\ ha^{-1}$ ], teor de cobre do solo (Cu) [ $mg\ dm^{-3}$ ], pH do solo (pH) e resistência do solo à penetração na camada 0,21 a 0,30 m ( $RSP_{(0,21-0,30m)}$ ) [MPa] foram selecionados devido à sua simetria. Esses dados foram coletados durante o ano-safra 2022/2023 em uma área comercial de 167,35 ha localizada na região Oeste do Paraná, Brasil, próximo à cidade de Cascavel, com as seguintes coordenadas geográficas: latitude  $24^{\circ}57'18''S$ , longitude  $53^{\circ}34'29''W$ , e altitude média de 750 m (Figura 1). O clima regional é mesotérmico e temperado superúmido, tipo climático Cfa (Köppen), e seu solo é classificado como típico Latossolo Vermelho distrófico (SANTOS *et al.*, 2018).

Os 102 pontos amostrais foram definidos usando uma abordagem de amostragem lattice plus close pairs (DIGGLE & LOPHAVEN, 2006). Este tipo de amostragem combina uma grade regular de pontos com espaçamento  $\Delta$  e mais  $m$  pontos extras adicionais, que são colocados aleatoriamente dentro de um círculo de raio  $\delta$  em torno de um ponto da grade, também escolhido aleatoriamente (DIGGLE & RIBEIRO JR., 2007; CHIPETA *et al.*, 2017). Esses pontos foram coletados manualmente e rastreados com um dispositivo GPS.

Figura 1 Localização da área agrícola com 102 pontos amostrais.



Fonte: Os autores (2024).

### 5.1.2.2 Análise Geoestatística

Para modelar a estrutura de dependência espacial de uma variável regionalizada, foi considerado um processo estocástico  $\mathbf{Z} = \{Z(\mathbf{s}), \mathbf{s} \in S\}$  em que  $\mathbf{s} = (x, y)^T$  é um vetor representando um local específico na área de estudo  $S \subset \mathbb{R}^2$ , em que  $\mathbb{R}^2$  é o espaço euclidiano bidimensional. Supõe-se que os dados  $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$  constituem um processo estocástico estacionário de segunda ordem e isotrópico, coletados em locais conhecidos  $(\mathbf{s}_1, \dots, \mathbf{s}_n)$ , e modelado pelo conjunto  $\mathbf{Z} = \mu\mathbf{1} + \boldsymbol{\varepsilon}$ , em que  $\mu$  representa a média do processo constante,  $\mathbf{1}$  representa um vetor de uns de  $n \times 1$  e  $\boldsymbol{\varepsilon} = (\varepsilon(\mathbf{s}_1), \dots, \varepsilon(\mathbf{s}_n))^T$  representa o vetor dos erros aleatórios,  $n \times 1$ , com distribuição normal n-variada, em que  $E(\boldsymbol{\varepsilon}) = \mathbf{0}$  e matriz de covariância  $\boldsymbol{\Sigma}$ ,  $n \times n$ , sendo  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}[(\sigma_{ij})] = C(Z(\mathbf{s}_i), Z(\mathbf{s}_j))$ ,  $i, j = 1, \dots, n$ .

Na forma paramétrica a matriz de covariância  $\boldsymbol{\Sigma}$  pode ser definida como  $\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3)$ , em que  $\mathbf{I}_n$  é a matriz identidade de dimensão  $n \times n$ ,  $\varphi_1 \geq 0$  é o parâmetro denominado efeito pepita,  $\varphi_2 \geq 0$  é o parâmetro de contribuição,  $\varphi_3 \geq 0$  é o parâmetro que define o alcance ( $a = g(\varphi_3)$ ) do modelo, e  $\mathbf{R}(\varphi_3) = [(r_{ij})]$  é uma matriz simétrica e definida positiva. Os elementos  $r_{ij}$ ,  $i, j = 1, \dots, n$  representam a associação entre os pontos  $\mathbf{s}_i$  e  $\mathbf{s}_j$ , onde  $r_{ij} = 1$  se  $i = j$  e  $r_{ij} = 0$  se  $i \neq j$  e  $\varphi_2 = 0$ ; e  $r_{ij} = \varphi_2^{-1} C(h_{ij})$  se  $i \neq j$  e  $\varphi_2^{-1} \neq 0$ , onde  $C(h_{ij}) = C(Z(\mathbf{s}_i), Z(\mathbf{s}_j))$  é a função

de covariância teórica e  $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$  é a distância euclidiana entre  $\mathbf{s}_i$  e  $\mathbf{s}_j$  (URIBE-OPAZO *et al.*, 2012; DE BASTIANI *et al.*, 2015).

Para o estudo da estrutura de dependência espacial foram construídos semivariogramas omnidirecionais, empregando os estimadores de Função Semivariância de Matheron, conforme a Equação 1 (CRESSIE, 2015).

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

Em que,  $\hat{\gamma}(h)$  é o estimador da função semivariância de Matheron  $\gamma(h) = \frac{1}{2}E[(Z(\mathbf{s}) - Z(\mathbf{s} + h))^2]$ ;  $N(h)$  é o número de pares de valores amostrados em localizações separadas pela distância  $h$ ;  $Z(\mathbf{s}_i + h)$  é o valor da variável  $Z$  no ponto  $\mathbf{s}_i + h$ ;  $Z(\mathbf{s}_i)$  é o valor da variável  $Z$  no ponto  $\mathbf{s}_i$ ;

O Modelo de Wave (OLEA, 2006), conforme a Equação 2, foi empregado para modelar as estruturas de dependência espacial, sendo que os parâmetros foram estimados pelo método de máxima verossimilhança (ML). Além do Modelo de Wave, os modelos Exponencial, Gaussiano e família Matérn também foram empregados na análise (URIBE-OPAZO *et al.*, 2023).

$$\gamma(h) = \begin{cases} \varphi_1 + \varphi_2 \left(1 - \frac{\varphi_3}{h} \text{sen} \left(\frac{h}{\varphi_3}\right)\right), & h > 0 \\ 0, & h = 0 \end{cases}, \quad (Eq. 2)$$

Para o Modelo de Wave, as derivadas de primeira e segunda ordem da estrutura de covariância  $\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \left(\frac{\varphi_3}{h} \sin \left(\frac{h}{\varphi_3}\right)\right)$  são, conforme a Equação 3):

$$\begin{aligned} \frac{\partial \Sigma}{\partial \varphi_1} &= \mathbf{I}_n; \quad \frac{\partial \Sigma}{\partial \varphi_2} = \mathbf{R}(\varphi_3); \quad \frac{\partial \Sigma}{\partial \varphi_3} = \varphi_2 \left[ \frac{1}{h} \sin \left(\frac{h}{\varphi_3}\right) - \frac{1}{\varphi_3} \cos \left(\frac{h}{\varphi_3}\right) \right]; \quad \frac{\partial^2 \Sigma}{\partial \varphi_1^2} = 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_2} = \\ 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_3} &= 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_2^2} = 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_2 \partial \varphi_3} = \frac{1}{h} \sin \left(\frac{h}{\varphi_3}\right) - \frac{1}{\varphi_3} \cos \left(\frac{h}{\varphi_3}\right); \quad \frac{\partial^2 \Sigma}{\partial \varphi_3^2} = \\ -\frac{\varphi_2 h}{\varphi_3^3} \sin \left(\frac{h}{\varphi_3}\right) & \end{aligned} \quad (Eq. 3)$$

### 5.1.2.3 Máxima Verossimilhança (ML)

Seja um modelo  $Z \sim N_n(\mu\mathbf{1}, \Sigma)$  com média  $\mu\mathbf{1}$  e matriz de covariância  $\Sigma = \varphi_1\mathbf{I}_n + \varphi_2\mathbf{R}(\varphi_3)$ , que depende apenas do vetor de parâmetros  $\boldsymbol{\varphi}$  (DIGGLE & RIBEIRO JR, 2007). O logaritmo da função verossimilhança é definido pela Equação 4, na qual o vetor de parâmetros desconhecidos do modelo  $\boldsymbol{\theta} = (\mu, \boldsymbol{\varphi}^T)^T$  e  $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$  pode ser estimado, maximizando o logaritmo da função verossimilhança (MARDIA & MARSHAL, 1984):

$$\mathcal{L}(\boldsymbol{\theta}) = -\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (\mathbf{Z} - \mu\mathbf{1})^T \Sigma^{-1} (\mathbf{Z} - \mu\mathbf{1}), \quad (4)$$

As funções escore de  $\boldsymbol{\theta} = (\mu, \boldsymbol{\varphi}^T)^T$  são:  $U(\mu) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \mu} = \mathbf{1}^T \Sigma^{-1} (\mathbf{Z} - \mu\mathbf{1})$  e  $U(\boldsymbol{\varphi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}}$  que tem como  $j$ -ésimo elementos  $\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \varphi_j} = -\frac{1}{2} \text{tr} \left\{ \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \right\} + \frac{1}{2} \boldsymbol{\varepsilon}^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \boldsymbol{\varepsilon}$ ,  $j=1,2,3$  e  $\boldsymbol{\varepsilon} = (\mathbf{Z} - \mu\mathbf{1})$ .

Para estimar os parâmetros  $\mu$  e  $\boldsymbol{\varphi}$  resolve-se os sistemas de equações  $U(\mu) = 0$  e  $U(\boldsymbol{\varphi}) = \mathbf{0}$ .

A matriz de informação observada é definida por  $\mathcal{L}(\boldsymbol{\theta}) = -L(\boldsymbol{\theta})$  e deve ser avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , sua forma particionada é  $L(\boldsymbol{\theta}) = \begin{pmatrix} L_{\mu\mu} & L_{\mu\boldsymbol{\varphi}} \\ L_{\boldsymbol{\varphi}\mu} & L_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \end{pmatrix}$  em que,  $L_{\mu\mu} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial \mu^T} = -\mathbf{1}_n^T \Sigma^{-1} \mathbf{1}_n$ ;  $L_{\mu\boldsymbol{\varphi}} = L_{\boldsymbol{\varphi}\mu}^T = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial \boldsymbol{\varphi}^T}$ , com elementos  $\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial \varphi_j} = -\mathbf{1}_n^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \boldsymbol{\varepsilon}$ ,  $j = 1, \dots, q$ ; sendo,  $\boldsymbol{\varepsilon} = (\mathbf{Z} - \mu\mathbf{1})$ , e  $L_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T}$ , com elementos  $\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \varphi_i \partial \varphi_j} = \frac{1}{2} \text{tr} \left\{ \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} \right) \right\} + \frac{1}{2} \boldsymbol{\varepsilon}^T \Sigma^{-1} \left\{ \frac{\partial^2 \Sigma}{\partial \varphi_i \partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} - \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \right\} \Sigma^{-1} \boldsymbol{\varepsilon}$ ,  $i, j = 1, 2, 3$ .

A matriz de informação esperada é denominada como  $\mathbf{K}(\boldsymbol{\theta}) = E[-L(\boldsymbol{\theta})]$ , sendo uma matriz diagonal da seguinte forma  $\mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}(\mu) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\boldsymbol{\varphi}) \end{pmatrix}$ , em que  $\mathbf{K}(\mu) = E[-L_{\mu\mu}] = E \left[ -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \mu \partial \mu^T} \right] = \mathbf{1}_n^T \Sigma^{-1} \mathbf{1}_n$  e  $\mathbf{K}(\boldsymbol{\varphi}) = E[-L_{\boldsymbol{\varphi}\boldsymbol{\varphi}}] = E \left[ -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T} \right]$ , que tem como elementos  $k_{ij}(\boldsymbol{\varphi}) = \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \right]$  com  $i, j = 1, \dots, q$  (URIBE-OPAZO *et al.*, 2012).

Em especial para o Modelo de Wave, têm-se:  $\mathbf{K}(\mu) = \mathbf{1}_n^T \Sigma^{-1} \mathbf{1}_n$ ;  $k_{11}(\boldsymbol{\varphi}) = \frac{1}{2} \text{tr}[\Sigma^{-1} \mathbf{I}_n \Sigma^{-1} \mathbf{I}_n]$ ;  $k_{12}(\boldsymbol{\varphi}) = \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \mathbf{I}_n \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_2} \right] = k_{21}(\boldsymbol{\varphi})$ ;  $k_{13}(\boldsymbol{\varphi}) = \frac{1}{2} \text{tr} \left[ \Sigma^{-1} \mathbf{I}_n \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_3} \right]$ ;

$$k_{22}(\varphi) = \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right]; \quad k_{23}(\varphi) = \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right] = k_{32}(\varphi); \quad k_{33}(\varphi) = \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right],$$

em que as derivadas  $\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i}$  são definidas na Equação 3.

Os erros padrão assintóticos podem ser calculados invertendo a matriz de informação  $\mathbf{K}(\boldsymbol{\theta})$  e calculando a raiz quadrada dos elementos da diagonal principal  $\mathbf{K}(\boldsymbol{\theta})^{-1}$  (URIBE-OPAZO *et al.*, 2012;2023).

#### 5.1.2.4 Infkuença Local

De Bastiani *et al.* (2015) propuseram uma perturbação generalizada na variável resposta  $\mathbf{Z}_\omega = \mathbf{Z} + \boldsymbol{\Sigma}^{\frac{1}{2}} \omega$  em que  $\omega$  é um vetor que pertence ao espaço de perturbações  $\Omega$ . O logaritmo da função verossimilhança perturbado para o modelo normal é  $\mathcal{L}(\boldsymbol{\theta}_\omega) = \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{Z}_\omega - \boldsymbol{\mu}\mathbf{1})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z}_\omega - \boldsymbol{\mu}\mathbf{1})$ .

Nesse estudo, a matriz  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\mu^T, \boldsymbol{\Delta}_\varphi^T)^T$  avaliadas em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e em  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ , é da forma  $\boldsymbol{\Delta}_\mu = -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}}$ ;  $\boldsymbol{\Delta}_\varphi = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi \partial \boldsymbol{\omega}^T}$ , que tem como elementos  $\Delta_{\varphi_j} = \boldsymbol{\varepsilon}^T \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{A}}{\partial \varphi_j} - \mathbf{D}_j \mathbf{A} \right]$ , para  $j = 1, \dots, q$ , em que  $\mathbf{A} = \boldsymbol{\Sigma}^{\frac{1}{2}}$ ,  $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{\omega}_0) = (\mathbf{Z} - \boldsymbol{\mu}\mathbf{1})$ , e  $\mathbf{D}_j = \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1}$ .

Considere a matriz  $\mathbf{B} = \boldsymbol{\Delta}^T \mathbf{L}^{-1} \boldsymbol{\Delta}$  em que,  $\mathbf{L}$  é a matriz de informação observada, o elemento  $C_i = 2|b_{ii}|$ , para  $i = 1, \dots, n$ , em que  $b_{ii}$  é o elemento da diagonal principal da matriz  $\mathbf{B}$ . A partir da matriz  $\mathbf{B}$  pode-se utilizar o gráfico de  $C_i$  em função de  $i$  (ordem das observações) como técnica de diagnóstico para avaliar a existência de observações influentes.

Seja  $L_{max}$  o autovetor, normalizado, associado ao maior autovalor, em módulo, da matriz  $\mathbf{B}$ . Segundo Cook (1986) o gráfico dos elementos  $|L_{max}|$  em função de  $i$  (ordem das observações) pode revelar os elementos que, sob pequenas perturbações, exerce maior influência no afastamento da verossimilhança.

#### 5.1.2.5 Comparações

A medida do grau de dependência espacial foi obtida pelo Índice de Dependência Espacial (SDI) (SEIDEL & OLIVEIRA, 2014), classificado conforme a Tabela 1 (NETO *et al.*, 2018). Essa classificação é essencial para a interpretação dos resultados e a tomada de decisões com base na análise de dependência espacial nos modelos ajustados.

Tabela 1: Classificação do Índice de Dependência Espacial - SDI.

Modelo	MF	Fraco	Moderado	Forte
Wave	0,58900	SDI ≤ 11%	11% < SDI ≤ 24%	SDI > 24%

$SDI = MF \left( \frac{\varphi_2}{\varphi_1 + \varphi_2} \right) \min \left\{ 1; \left( \frac{a}{0,5MD} \right) \right\} 100$ , MF: fator específico do modelo,  $\varphi_1$ : efeito pepita,  $\varphi_2$ : contribuição,  $a$ : alcance, MD: é a distância máxima entre dois pontos amostrados.

Fonte: (SEIDEL & OLIVEIRA, 2016; NETO *et al.*, 2018)

A similaridade entre os mapas interpolados foi avaliada pelas estimativas dos Índices de Acurácia Kappa ( $\hat{K}$ ) e Kappa Ponderado ( $\hat{K}_w$ ) (COHEN, 1960; 1968). O índice Kappa Ponderado ( $\hat{K}_w$ ) se baseia em pesos  $w_{ij}$  para quantificar as diferenças. A classificação do índice Kappa Ponderado é equivalente ao índice Kappa (FLEISS *et al.*, 2003).

#### 5.1.2.6 Recurso Computacionais

Todas as análises foram realizadas utilizando o software R (R CORE TEAM, 2024). As semivariâncias foram calculadas utilizando as funções do pacote geoR, o qual também foi empregado para ajustar os modelos e produzir os mapas temáticos (RIBEIRO JR & DIGGLE, 2001). O pacote vcd foi utilizado para calcular o índice Kappa Ponderado (MEYER *et al.*, 2021).

### 5.1.3 Resultados e Discussões

#### 5.1.3.1 Análise Exploratória

Durante o período do ano-safra 2022/2023, a produtividade média da soja na área monitorada foi de  $1,533 \text{ t ha}^{-1}$ , considerada baixa em comparação com a produtividade estadual ( $3,860 \text{ t ha}^{-1}$ ) e nacional ( $3,507 \text{ t ha}^{-1}$ ) no mesmo período (CONAB, 2023). Esses resultados são inferiores aos encontrados por Dalposso *et al.* (2021) que conduziram estudos sobre a produtividade da soja na mesma cidade onde o experimento foi realizado.

Os resultados também indicam que o teor médio de cobre no solo foi classificado como baixo ( $0,69 \text{ mg dm}^{-3}$ ), enquanto o valor médio do pH (5,82) foi considerado bom, (SANTOS & SILVA, 2001). Esse resultado é superior ao encontrado por Dalposso *et al.* (2018) que realizaram experimentos sobre o teor do pH no solo na mesma cidade onde o estudo foi realizado. Quanto à  $RSP_{0,21-0,30m}$ , a média foi de 2,19

MPa, o que indica um baixo nível de compactação do solo, havendo pouca limitação ao desenvolvimento radicular, conforme os critérios de Canarache (1990).

A produtividade da soja e o teor de cobre no solo apresentaram o maior valor para o coeficiente de variação, sendo considerados heterogêneos ( $CV \geq 30\%$ ). Resultados semelhantes aos encontrados por Maltauro *et al.* (2023) que efetuaram um ensaio na mesma região onde a pesquisa foi realizada. Nenhuma das variáveis em estudo apresentou outlier.

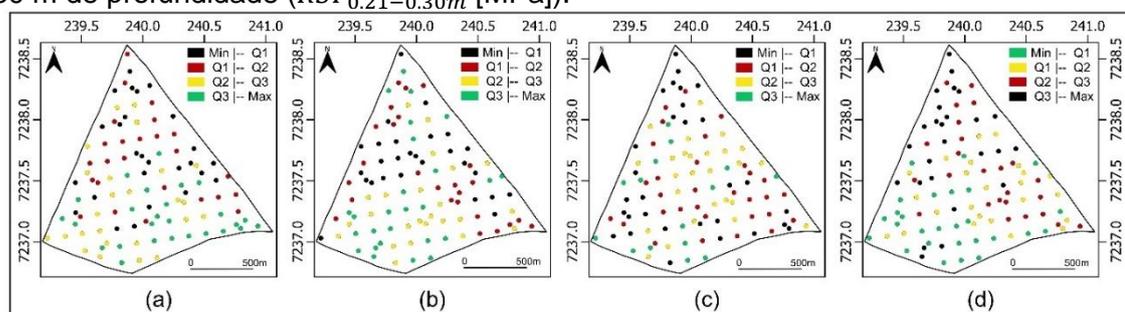
Tabela 2: Análise descritiva das variáveis resposta.

Estatísticas	Prod	Cu	pH	$RSP_{0,21-0,30m}$
Mínimo	0,330	0	4,60	1,23
1º Quartil	1,045	0,38	5,50	1,89
Mediana	1,442	0,68	5,90	2,22
Média	1,533	0,69	5,82	2,19
3º Quartil	1,988	1,02	6,20	2,47
Máximo	2,909	1,73	6,70	2,96
SD	0,607	0,43	0,43	0,38
CV (%)	39,59	62,69	7,40	17,39
$\tilde{u}_3$	0,38	0,21	-0,49	-0,20
Kur	-0,69	-0,84	0,04	-0,77

SD: desvio padrão; CV: coeficiente de variação;  $\tilde{u}_3$ : Coeficiente de assimetria; Kur: coeficiente de kurtose; Prod: produtividade da soja no ano-safra 2022/2023; Cu: cobre;  $RSP_{0,21-0,30m}$ : resistência do solo à penetração na camada 0,21 a 0,30m de profundidade. Fonte: Os autores (2024).

A distribuição espacial dos valores amostrais foi analisada por meio do Gráfico *post plot* (Figura 2). Esse gráfico permite verificar a presença de agrupamentos e tendências direcionais de valores amostrais altos ou baixos.

Figura 2 Postplot (a) produtividade da soja (Prod [ $t ha^{-1}$ ]), (b) teor de cobre do solo (Cu [ $mg dm^{-3}$ ]), (c) pH do solo (pH) e (d) resistência do solo à penetração na camada de 0.21 a 0.30 m de profundidade ( $RSP_{0,21-0,30m}$  [MPa]).



Fonte: Os autores (2024).

A Figura 2a mostra que, em geral, os maiores valores de produtividade da soja ( $t\ ha^{-1}$ ) se localizam na região central e sul da área monitorada. A região norte e central, por sua vez, apresenta uma predominância de amostras nas duas classes inferiores do gráfico (valor mínimo até a mediana), o que indica uma região de menor produtividade.

A Figura 2b revela que os maiores valores de Cobre estão concentrados na região sudoeste, que se destaca como a mais rica nesse nutriente. A Figura 2c indica que a maioria dos pontos estão entre o valor mínimo (4,6) e terceiro quartil (6,2), ou seja, estão próximos ao neutro, considerado adequado para a cultura. Por último, a Figura 2d indica que os maiores valores da  $RSP_{0,21-0,30m}$  estão localizados na região central e norte, sugerindo maior limitação ao crescimento radicular.

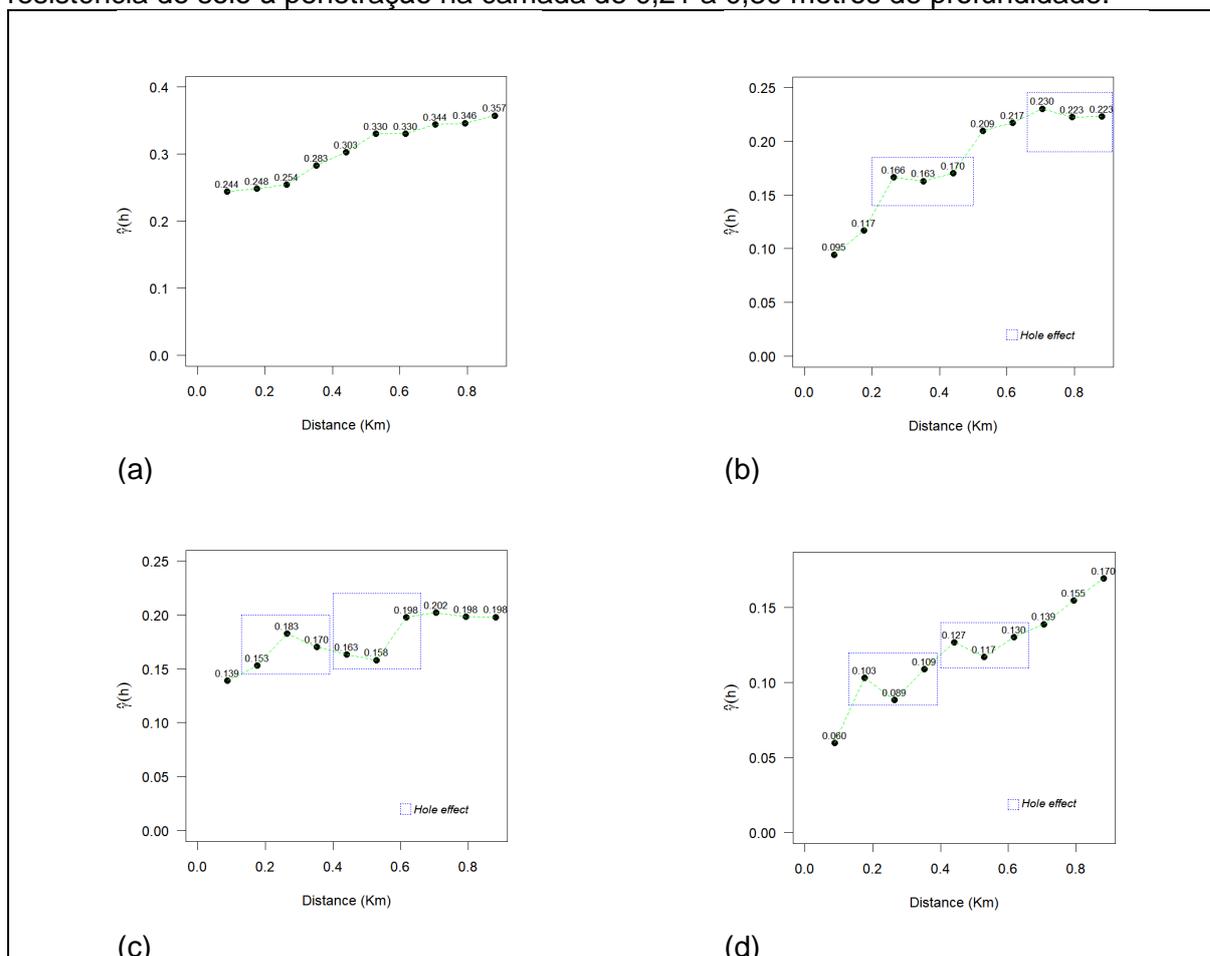
### 5.1.3.2 Análise Espacial

Para a análise da dependência espacial, foram consideradas 11 defasagens até a distância de 880 metros (ponto de corte de 50%) (CLARK, 1979). O semivariograma foi verificado nas direções  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ , confirmando que os dados são isotrópicos (GUEDES *et al.*, 2013).

A análise do semivariograma experimental revelou um comportamento oscilatório das semivariâncias, característica comum do *hole effect*. Esse efeito se manifesta quando valores elevados são sistematicamente cercados por valores baixos, e vice-versa, criando um padrão de alternância (Figura 3).

Essa característica já foi relatada em outros trabalhos sobre Agricultura de Precisão (SANTRA *et al.*, 2008; AMIRINEJAD *et al.*, 2011; DALPOSSO *et al.*, 2022) e foi melhor ajustado pelo Modelo geoestatístico *Wave*. Esse modelo é particularmente adequado para dados com periodicidade, que geram um "*hole effect*" devido às correlações positivas e negativas entre regiões distantes, fornecendo assim uma representação mais precisa da variabilidade espacial presente nos dados (MAHDI *et al.*, 2020).

Figura 3 Semivariância experimental (a) produtividade da soja, (b) Cobre, (c) pH, (d) resistência do solo à penetração na camada de 0,21 a 0,30 metros de profundidade.



Fonte: Os autores (2024).

Utilizando os critérios de Validação Cruzada (FARACO *et al.*, 2008), critério de informação de Akaike - AIC (AKAIKE, 1973) e critério de informação bayesiano de Schwarz - BIC (SCHWARZ, 1978), o Modelo de *Wave* foi o que obteve o melhor desempenho para todas as variáveis em estudo (Tabela 3).

Com base nos resultados do Índice de Dependência Espacial – SDI (Tabela 3), é evidente que a produtividade da soja está associada a uma forte dependência espacial entre as observações (SDI > 24%). Além disso, observa-se uma relação de moderada dependência espacial entre as observações para o Cu, pH e  $RSP_{0,21-0,30m}$  ( $11\% < SDI \leq 24\%$ ).

Esses resultados destacam a influência da distribuição espacial dessas variáveis nas características agrônomicas e na produtividade da cultura da soja, fornecendo informações valiosas para a compreensão dos padrões de variabilidade no ambiente de cultivo.

Tabela 3: Parâmetros estimados por máxima verossimilhança e erros padrão assintóticos entre parênteses.

Modelo de Wave*	Prod	Prod#85	Cu	Cu#36	pH	pH#99	RSP <sub>0,21-0,30m</sub>	RSP <sub>0,21-0,30m</sub> #46
$\hat{\mu}$	1,794 (0,356)	1,777 (0,355)	0,686 (0,080)	0,687 (0,081)	5,837 (0,049)	5,846 (0,062)	2,198 (0,113)	2,195 (0,113)
$\hat{\phi}_1$	0,229 (0,191)	0,229 (0,204)	0,117 (0,040)	0,118 (0,040)	0,109 (0,028)	0,154 (0,028)	0,095 (0,034)	0,096 (0,034)
$\hat{\phi}_2$	0,481 (0,391)	0,501 (0,402)	0,061 (0,033)	0,061 (0,033)	0,077 (0,029)	0,030 (0,021)	0,049 (0,039)	0,048 (0,039)
$\hat{\phi}_3$	0,489 (0,135)	0,476 (0,127)	0,166 (0,024)	0,166 (0,024)	0,068 (0,004)	0,146 (0,031)	0,357 (0,111)	0,354 (0,110)
$\hat{a}(km)$	1,462	1,424	0,498	0,497	0,610	0,438	1,068	1,060
<i>SDI</i>	39,90	40,42	11,38	11,36	16,84	4,76	20,04	19,83
<i>Class</i>	forte	forte	mod	mod	mod	fraco	mod	mod

\*: parâmetros do Modelo de Wave;  $\hat{\mu}$ : média;  $\hat{\phi}_1$ : efeito pepita;  $\hat{\phi}_2$ : contribuição;  $\hat{\phi}_3$ : função de alcance;  $\hat{a}$ : alcance; *SDI*: Índice de Dependência Espacial; *Class*: classificação; Prod: produtividade da soja com todas as observações; Prod#85: produtividade da soja sem a observação influente #85; Cu: cobre com todas as observações; Cu#36: cobre sem a observação influente #36; pH#99: pH sem a observação influente #99;  $RSP_{0,21-0,30m}$ : resistência do solo à penetração na camada 0,21 a 0,30m de profundidade com todas as observações;  $RSP_{0,21-0,30m}$ #46: resistência do solo à penetração na camada 0,21 a 0,30m de profundidade sem a observação influente #46; mod: moderado.

Fonte: Os autores (2024)

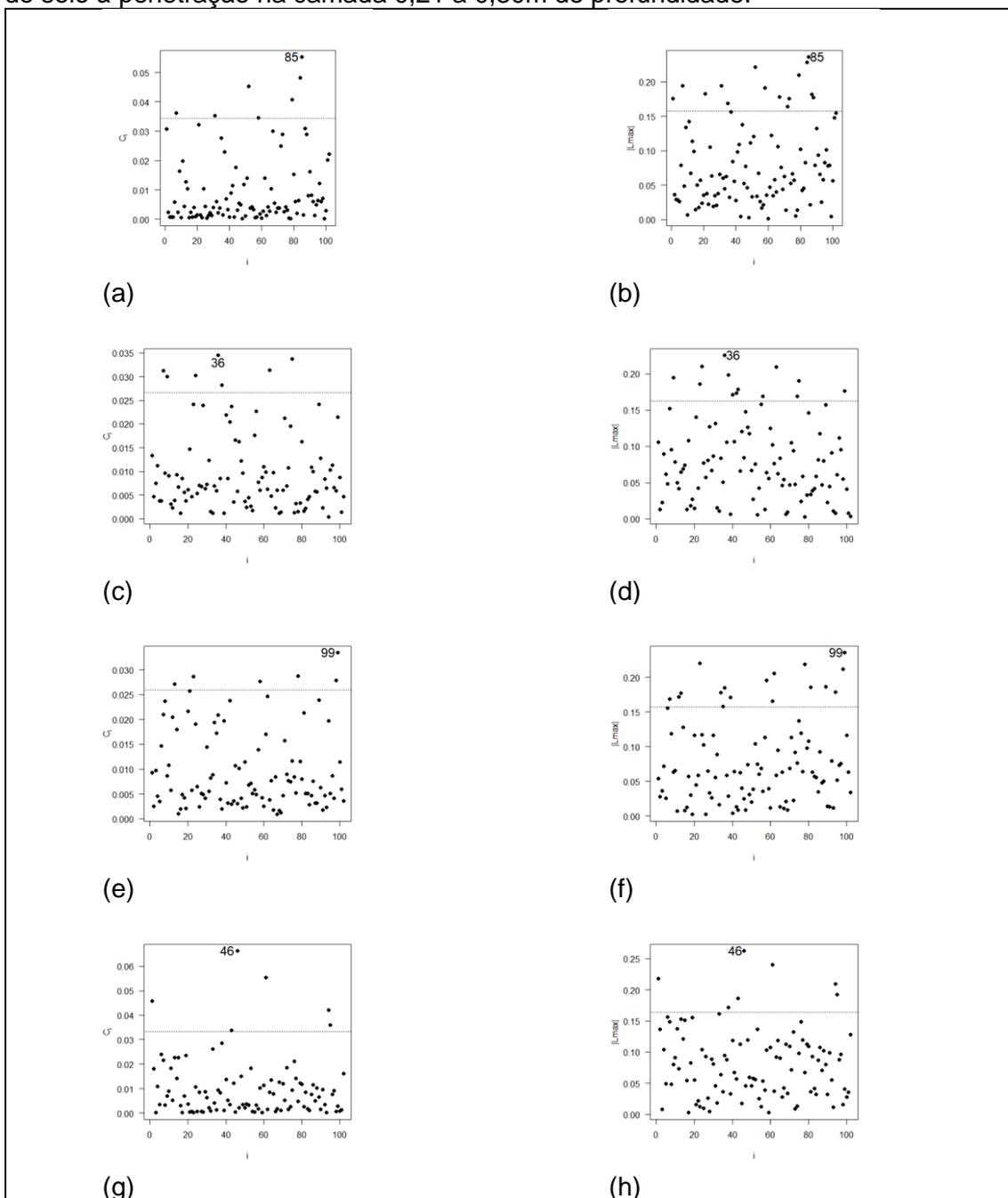
### 5.1.3.3 Diagnóstico de Influência Local

Os resultados da análise de influência local, (Figura 4), revelam as observações que têm uma influência significativa no afastamento da verossimilhança. As observações influentes são: #85 para a produtividade da soja; #36 para o teor de cobre no solo; #99 para o pH do solo; e #46 para a resistência do solo à penetração.

Segundo apontam Leiva *et al.* (2020) e Uribe-Opazo *et al.* (2023), é importante notar que em estatística espacial, um ponto influente não é necessariamente um outlier e vice-versa. Essa distinção ressalta a complexidade das relações espaciais e a necessidade de considerar cuidadosamente a influência de pontos individuais nos resultados da análise.

Para avaliar o efeito das observações influentes na análise da estrutura de dependência espacial e na elaboração do mapa temático, realizou-se a exclusão das observações influentes do banco de dados para cada variável, seguida de uma nova Análise de Variabilidade Espacial. Essa abordagem metodológica permite uma compreensão mais aprofundada da distribuição espacial das variáveis, considerando adequadamente a influência de observações individuais nos resultados da análise.

Figura 4 Gráficos de  $C_i$  e  $|L_{max}|$  em função da ordem das observações coletadas. (a)  $C_i$  para a produtividade da soja; (b)  $|L_{max}|$  para a produtividade da soja; (c)  $C_i$  para o teor de cobre; (d)  $|L_{max}|$  para o teor de cobre; (e)  $C_i$  para o pH; (f)  $|L_{max}|$  para o pH; (g)  $C_i$  para a resistência do solo à penetração na camada 0,21 a 0,30m de profundidade; (h)  $|L_{max}|$  para a resistência do solo à penetração na camada 0,21 a 0,30m de profundidade.



Fonte: Os autores (2024)

Considerando esse novo contexto, sem a presença de observações influentes, os resultados estão detalhados na Tabela 3. De acordo com os critérios de Validação Cruzada (FARACO *et al.*, 2008), o Critério de Informação de Akaike – AIC (AKAIKE, 1973) e o Critério de Informação Bayesiano – BIC (SCHWARZ, 1978), para a

produtividade da soja sem observação #85 (Prod#85), também se selecionou o Modelo de *Wave* para descrever a variabilidade espacial contínua, que apresentou valores semelhantes para os erros padrão assintóticos.

Além disso, há uma redução da faixa de dependência espacial em 38 metros, mantendo ainda uma forte dependência espacial de acordo com o Índice de Dependência Espacial – *SDI*. Dessa forma, novamente, utilizou-se o Modelo de *Wave* para as demais variáveis em estudo, sem as observações influentes, para descrever a variabilidade espacial contínua.

Porém, observou-se alterações nos parâmetros estimados pelo modelo. No caso do cobre sem observação #36 (Cu#36), os valores são semelhantes aos erros padrão assintóticos. Com relação ao pH sem observação #99 (pH#99), houve uma redução substancial de 172 metros na faixa de dependência espacial, resultando na transição da classificação moderada para fraca segundo o Índice de Dependência Espacial – *SDI* (NETO *et al.*, 2020).

Por fim, para a  $RSP_{(0,21-0,30m)}$  sem observação #46 ( $RSP_{(0,21-0,30m)}$ #46), foi observada uma redução no raio de dependência espacial de 8 metros, mantendo uma dependência espacial moderada de acordo com a Dependência Espacial Índice – *SDI* (NETO *et al.*, 2020).

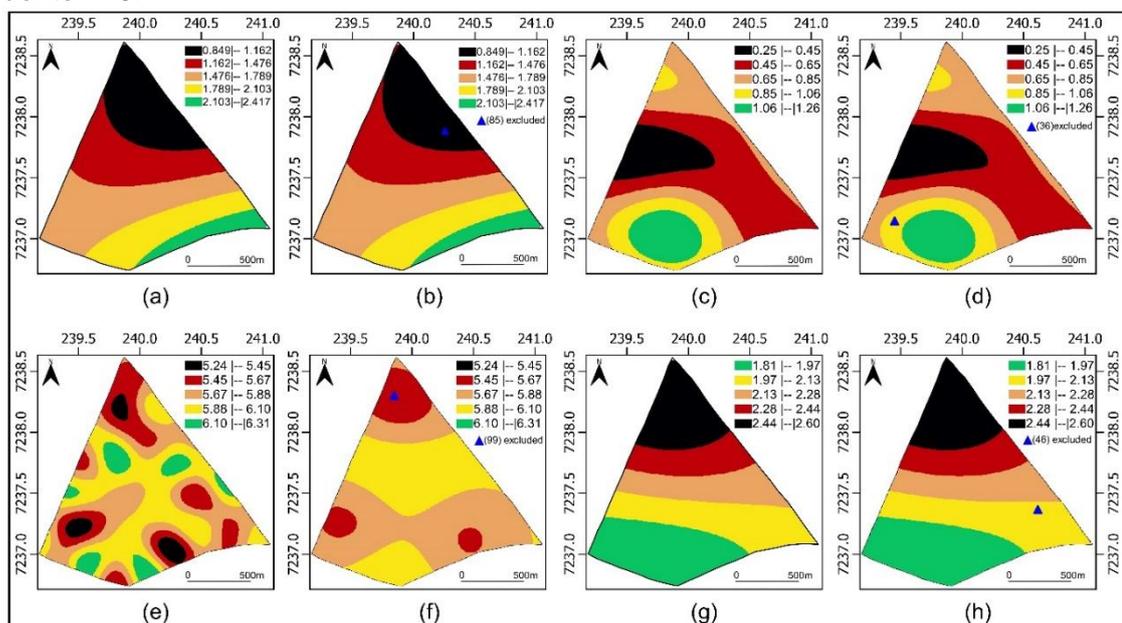
#### 5.1.3.4 Mapa Geoestatístico

Com base na interpolação por *krigagem* ordinária e nas informações do Modelo de *Wave* que melhor se ajustou a função semivariância em todas as variáveis em estudo, foi gerado o mapa da produtividade da soja ( $t\ ha^{-1}$ ), do teor de cobre ( $mg\ dm^{-3}$ ), do pH do solo e da  $RSP_{0,21-0,30m}$  (MPa) (Figura 5).

Ao comparar os mapas com todas as observações (Figura 5a) com aqueles sem a observação considerada influente (Figura 5b), observou-se que a distribuição espacial da produtividade da soja no ciclo 2022/2023 foi semelhante, segundo aos Índices de Acurácia Kappa e Kappa Ponderado ( $\hat{K}_w \geq 0,75$ ), conforme demonstrado na Tabela 4.

A análise comparativa entre o mapa temático do Cu considerando todos os pontos amostrais (Figura 5c) e o mapa excluindo a observação influente #36 (Cu#36) (Figura 5 d) nota-se uma semelhança na distribuição espacial dos valores de cobre na área de estudo.

Figura 5 Mapa temático: (a) produtividade da soja com todas as observações; (b) produtividade da soja sem observação influente #85; (c) teor de cobre com todas as observações; (d) teor de cobre sem observação influente #36; (e) pH com todas as observações; (f) pH sem observação influente #99; (g) resistência do solo à penetração na camada de 0,21 a 0,30 metros de profundidade com todas as observações; (h) resistência do solo à penetração na camada de 0,21 a 0,30 metros de profundidade sem observação influente #46.



Fonte: Os autores (2024).

O teor de Cu é um elemento essencial na fase reprodutiva das plantas e consequentemente na determinação da produtividade e qualidade das culturas, além de ser importante para a resistência da soja a doenças (HANSCH; MENDEL, 2009). Neste sentido, é crucial garantir uma análise mais eficiente da variabilidade espacial para que as plantas recebam uma quantidade adequada de cobre na sua dieta nutricional para um desenvolvimento sustentável e uma produção agrícola satisfatória.

Ao comparar os mapas de pH considerando todos os pontos (Figura 5e) e sem a observação #99 (pH#99) (Figura 5f), que é considerado influente, observam-se diferenças nas áreas de estudo, com diminuição da área na extremidade superior da classe de 6,10 a 6,31. A principal alteração ocorre na redução da área correspondente à classe de teor de pH médio, valores abaixo de 5,50, conforme classificação de Santos e Silva (2001).

Segundo Fagundes *et al.* (2018), para o cultivo da soja é importante que o pH do solo permaneça dentro da faixa ideal de 5,7 a 7,0 para garantir o bom desenvolvimento das plantas e uma produção satisfatória. Vale ressaltar que o pH do solo é influenciado por diversos fatores, como adição de fertilizante, presença de

matéria orgânica e tipo de solo. Neste sentido, é crucial monitorizar e ajustar o pH do solo para garantir uma nutrição adequada das plantas e uma elevada produtividade agrícola.

Ao comparar os dois mapas para a  $RSP_{(0,21-0,30m)}$ , com todas as observações (Figura 5 g) e sem a observação influente #46 ( $RSP_{(0,21-0,30m)}\#46$ ) (Figura 5 h), nota-se semelhança na distribuição espacial dos valores de  $RSP_{(0,21-0,30m)}$  na área de estudo. A compactação do solo é um fenômeno que ocorre quando uma força externa é aplicada à superfície do solo, alterando sua estrutura e propriedades (KELLER *et al.*, 2019; VANDERHASSELT *et al.*, 2023).

Neste estudo, constatou-se que a área mais afetada pela compactação do solo foi a região norte da propriedade, onde o terreno apresenta leve declive. Esta região também apresenta maior tráfego de máquinas, principalmente durante manobras de retorno, aumentando a pressão sobre o solo.

#### 5.1.3.5 Comparação de Mapas

Com a utilização das estimativas dos Índices de Acurácia Kappa ( $\hat{K}$ ) e Kappa Ponderado ( $\hat{K}_w$ ) adota-se uma abordagem não subjetiva para a análise comparativa de mapas, permitindo a quantificação das diferenças observadas visualmente nos mapas. Para a maioria das variáveis em estudo, observa-se uma alta similaridade entre os mapas que consideram todas as observações e aqueles que descartam as observações influentes ( $\hat{K}_w \geq 0.75$ ), como demonstrado na Tabela 4.

Tabela 4: Índice Kappa ( $\hat{K}$ ) e Kappa Ponderado ( $\hat{K}_w$ ) para a comparação entre os mapas obtidos com todas as observações e os mapas obtidos excluindo-se as observações influentes.

Comparação	$\hat{K}$	$\hat{K}_w$
Prod × Prod#85	0.967	0.991
Cu × Cu#36	0.992	0.998
pH × pH#99	0.245	0.452
$RSP_{0,21-0,30m} \times RSP_{0,21-0,30m}\#46$	0.980	0.996

Classificação:  $\hat{K}_w \geq 0.75$  indica alta similaridade entre os mapas;  $0.4 < \hat{K}_w < 0.75$  indica moderada similaridade;  $\hat{K}_w \leq 0.4$  indica baixa similaridade; Prod: produtividade da soja com todas as observações; Prod#85: produtividade da soja sem a observação influente #85; Cu: cobre com todas as observações; Cu#36: cobre sem a observação influente #36;  $RSP_{0,21-0,30m}$ : resistência do solo à penetração na camada 0,21 a 0,30 metros de profundidade com todas as observações;  $RSP_{0,21-0,30m}\#46$ : resistência do solo à penetração na camada 0,21 a 0,30 metros de profundidade sem a observação influente #46.

Fonte: Os autores (2024).

#### 5.1.4 Conclusões

O Modelo de *Wave* se destacou em relação aos demais modelos, sendo o mais adequado para caracterizar a estrutura de dependência espacial da produtividade da soja e dos atributos do solo para o ano-safra em estudo. Os mapas temáticos gerados não só permitiram a predição de variáveis em locais não amostrados dentro da área de estudo, bem como a criação de zonas de aplicação, considerando a produtividade e os atributos do solo. Essa abordagem permite a aplicação localizada de insumos, visando maximizar o retorno econômico e minimizar os impactos ambientais, resultando em um manejo agrícola eficiente.

Além disso, a identificação e remoção de observações influentes provocaram alterações nas estimativas dos parâmetros que definem a estrutura de dependência espacial, impactando diretamente nas características dos setores dos mapas de produtividade da soja e atributos do solo. Ressalta-se, ainda, a importância da realização de estudos de diagnóstico de influência local em análises geoestatísticas para garantir a robustez e precisão dos resultados obtidos.

Digno de nota é a potencial aplicabilidade da metodologia apresentada neste trabalho para estudar a produtividade de outras culturas e diferentes atributos do solo em diversas áreas e em diferentes períodos, ampliando assim seu alcance e relevância no contexto da Agricultura de Precisão e do manejo sustentável e racional de recursos naturais.

#### 5.1.5 Agradecimentos

Este trabalho foi realizado com apoio do Conselho de Desenvolvimento Científico e Tecnológico (CNPq); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Programa de Pós-Graduação em Engenharia Agrícola da Universidade Estadual do Oeste do Paraná (PGEAGRI-UNIOESTE); Universidade Tecnológica Federal do Paraná (UTFPR) e Laboratório de Estatística Espacial (LEE-UNIOESTE).

### REFERÊNCIAS

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. **Journal of the Royal Statistical Society**, v. 51, p. 469-483, 1973.

AMIRINEJAD, A. A.; KAMBLE, K. AGGARWAL, P.; CHAKRABORTY, D.; PRADHAN, S.; MITTAL, R. B.; Assessment and mapping of spatial variation of soil physical health in a farm. **Geoderma**, v. 160, p. 292-303, 2011.

CANARACHE, A. A generalized semi-empirical model estimating soil resistance to penetration. **Soil and Tillage Research**, v. 16, n. 1, p. 51-70, 1990.

CARVALHO, J. R. P. DE; VIEIRA, S. R.; GRECO, C. R. Comparação de métodos para ajuste de modelos de semivariograma da precipitação pluvial anual. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v. 13, n. 4, p. 443-448, 2009.

CHIPETA, M. G.; TERLOUW, D. J.; PHIRI, K. S.; DIGGLE, P. J. Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. **Environmetrics**, v. 28, n. 1, p. e2425, 2017.

CLARK, I. **Practical Geostatistics**. Applied Science Publishers LTD, London, 1979.

COHEN, J. A coefficient of agreement for nominal scales. **Educ Psychol Meas**, v. 20, n. 1, p. 37-46, 1960.

COHEN, J. Weighted kappa: nominal scale agrément with provision for scale disagreement or partial credit. **Psychol. Bull.**, v.70, n. 4, p. 213-220, 1968.

Conab – Companhia Nacional de Abastecimento. Séries históricas: soja Brasil – série histórica de produtividade – safras 1976/1977 a 2022/2023, 2022 Disponível em: <<https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras?start=30>>. Acesso em: 5 jun. 2024.

COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society**, v. 48, n. 2, p. 133–169, 1986.

CRESSIE, N. **Statistics for spatial data**. New York: John Wiley and Sons. P. 928, 2015.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; JOHANN, J. A.; GALEA, M.; DE BASTIANI, F. Gaussian spatial model of soybean yield using bootstrap methods. **Engenharia Agrícola**, v. 38, n. 1, p. 110-116, 2018.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; DE BASTIANI, F. Spatial-temporal Analysis of Soybean Productivity Using Geostatistical Methods. **Journal of Agricultural Studies**, v. 9, n. 2, p. 283-303, 2021.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; OLIVEIRA, M. P. Comparison between Matheron and Genton semivariance function estimators in spatial modeling of soybean yield. **Australian Journal of Crop Science**, v. 16, n. 7, p. 916-921, 2022.

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

- DIGGLE, P.; LOPHAVEN, S. Bayesian geostatistical design. **Scandinavian Journal of Statistics**, v. 33, n. 1, p. 53-64, 2006.
- DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. **Model based Geoestatistics**. New York: Springer, 2007.
- FAGUNDES, R. S.; URIBE-OPAZO, M. A.; GUEDES, L. P. C.; GALEA, M. Slash spatial linear modeling: soybean yield variability as a function of soil chemical properties. **Revista Brasileira de Ciência do Solo**, v. 42, p. e0170030, 2018.
- FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para a elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do solo**, v. 32, n. 2, p. 463-476, 2008.
- FLEISS, J. L.; LEVIN, B.; MYUNGHEE, C. P. **Statistical methods for rates and proportions**. 3rd edn. Nova Jersey: John Wiley. p. 800, 2003.
- GUEDES, L. P. C.; URIBE-OPAZO, M. A.; RIBEIRO JR, P. J. Influence of incorporating geometric anisotropy on the construction of thematic maps of simulated data and chemical attributes of soil. **Chilean Journal of Agricultural Research**, v. 73, n. 4, p. 414–423, 2013.
- HANSCH, R.; MENDEL, R. R. Physiological functions of mineral micronutrientes (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). **Current Opinion in Plant Biology**, v. 12, p. 259-266, 2009.
- ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, p. 560, 1989.
- KELLER, T.; SANDIN, M.; COLOMBI, T.; HORN, R.; OR, D. Historical increase in agricultural Machinery weights enhanced soil stress levels and adversely affected soil functioning. **Soil and Tillage Research**, v. 194, p. 104293, 2019.
- LEIVA, V.; SÁNCHEZ, L.; GALEA, M.; SAULO, H. Global and local diagnostic analytics for a geostatistical model based on a new approach to quantile regression. **Stochastic Environmental Research and Risk Assessment**, v. 34, p. 1457–1471, 2020.
- MAHDI, E.; ABUZOID, A. H.; ATTA, A. M. A. Empirical variogram for achieving the best valid variogram. **Commun Stat Appl Methods**, v. 27, p. 547-568, 2020.
- MALTAURO, T. C.; GUEDES, L. P. C.; CANTON, L. E. D.; URIBE-OPAZO, M. A. Spatial multivariate optimization for a sampling redesign with a reduced sample size of soil chemical properties. **Revista Brasileira de Ciencia do Solo**, v. 47, p. e0220072, 2023.

MARDIA, K. V.; MARSHAL, R. J. Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. **Biometrika**, v. 71, n.1, p. 135-146, 1984.

MEYER, D.; ZEILEIS, A.; HORNIK, K. vcd: Visualizing Categorical Data. **R package** version v. 1, p.4-9, 2021. <<https://CRAN.R-project.org/package=vcd>>

NETO, E. A.; BARBOSA, I. C.; SEIDEL, E. J.; OLIVEIRA, M. S.; Spatial dependence index for cubic, pentaspherical and wave semivariogram models. **Bulletin of Geodetic Sciences**, v. 24, n. 1, p. 142-151, 2018.

OLEA, R. A. A six-step practical approach to semivariogram modeling. **Stoch Environ Res Risk Assess**, v. 20, n. 5, p. 307-318, 2006.

RIBEIRO JR, P. J.; DIGGLE, P. J. geoR: A Package For Geostatistical Analysis. **R News**, v. 1, n. 2, p. 15–18, 2001.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2024. <https://www.R-project.org>.

SANTOS, D. R.; SILVA, L. S. **Fertilidade do solo e nutrição de plantas**. Santa Maria: Universidade Federal de Santa Catarina, p. 20, 2001.

SANTOS, H. G.; JACOMINI, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J.; COELHO, M. R.; ALMEIDA, J. A.; ARAÚJO, FILHO, J. C.; OLIVEIRA, J. B.; CUNHA, T. J. F. **Sistema Brasileiro de classificação de solos**. Brasília, DF: Embrapa, 2018.

SANTRA, P.; CHOPRA, U. K.; CHAKRABORTY, D. Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm. **Curr Sci.**, v. 95, n. 7, p. 937-945, 2008.

SEIDEL, E. J.; OLIVEIRA, M. S. Novo índice geoestatístico para a mensuração da dependência espacial. **Revista Brasileira de Ciência do Solo**, v. 38, p. 699-705, 2014.

SCHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

URIBE-OPAZO, M. A.; BORSSOI, J.; 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.; ASSUMPÇÃO, R. A. B. Influence diagnostics on a reparametrized *t*-Student spatial linear model. **Spatial Statistics**, v. 41, p. 100481, 2021.

URIBE-OPAZO, M. A.; DALPOSSO, G. H.; GALEA, M.; JOHANN, J. A.; DE BASTIANI, F.; MOYANO, E. N. C.; GRZEGOZEWSKI, D. M. Spatial variability of wheat yield using the gaussian spatial linear model. **Australian Journal of Crop Science**, v. 17, n. 2, p. 179-189, 2023.

VANDERHASSELT, A.; COOL, S.; D'HOUSE, T.; CORNELIS, W. How tine characteristics of subsoilers affect fuel consumption, penetration resistance and potato yield of a sandy loam soil. **Soil and Tillage Research**, v. 228, p. 105631, 2023.

ZAIN, M.; MA, H.; RAHMAN, S. U.; NURUZZAMAN, M.; CHAUDHARY, S.; AZEEM, I.; MEHMOOD, F.; DUAN, A.; SUN, C. Nanotechnology in precision agriculture: Advancing towards sustainable crop production. **Plant Physiology and Biochemistry**, v. 206, p. 108244, 2024.

## 5.2 ARTIGO 2 – ANÁLISE DA DEPENDÊNCIA ESPACIAL DA PRODUTIVIDADE DA SOJA UTILIZANDO COVARIÁVEIS E O MODELO ESPACIAL LINEAR GAUSSIANO<sup>2</sup>

### ANÁLISE DA DEPENDÊNCIA ESPACIAL DA PRODUTIVIDADE DA SOJA UTILIZANDO COVARIÁVEIS E O MODELO ESPACIAL LINEAR GAUSSIANO

#### Resumo

A interação entre a produtividade da soja e os atributos físicos e químicos do solo é crucial para desenvolver estratégias que aumentem o retorno econômico, reduzam os custos e minimizem os impactos ambientais. Este estudo, realizado em uma área comercial de 172,03 ha durante o ano-safra de 2022/2023, utilizou como covariáveis o teor de cálcio (Ca), cobre (Cu), acidez do solo (pH), potássio (K), fósforo (P) e a resistência do solo à penetração (RSP) para explicar a produtividade da soja (Prod). A metodologia combinou o método de interpolação *Thin Plate Spline* (TPS) para interpolar os atributos químicos e físicos do solo, considerados como covariáveis fixas e usados como *drift* no Modelo de *Krigagem* com *drift* externo, aplicado para interpolar a produtividade da soja. Essa integração de métodos se mostrou altamente eficaz para aprimorar as estimativas em áreas não amostradas, oferecendo uma visão espacial detalhada e precisa. Além disso, novas técnicas de diagnóstico de influência local foram desenvolvidas para dados espaciais utilizando o modelo geoestatístico *Wave*, que é particularmente útil quando a semivariância apresenta o efeito *hole effect*. Os resultados mostraram que a remoção de observações localmente influentes altera significativamente as predições do Modelo de *Wave*, ressaltando a importância de excluir tais observações para garantir estimativas confiáveis dos parâmetros de dependência espacial. A significância dos parâmetros das covariáveis no modelo final de variabilidade foi avaliada por meio do teste de razão de verossimilhança sobre o vetor de parâmetros estimados ( $\hat{\beta}$ ), permitindo identificar as covariáveis que compõem o Modelo de variabilidade espacial. O mapa de produtividade da soja, gerado a partir dessas covariáveis, oferece informações valiosas para a definição de zonas de manejo, o que não apenas potencializa a lucratividade ao melhorar a eficiência no uso de insumos, mas também promove uma agricultura mais sustentável e ambientalmente responsável.

**Palavras-chave:** geoestatística; influência local; *krigagem* com *drift* externo; Modelo *Wave*; máxima verossimilhança; *Thin Plate Spline*.

#### 5.2.1 Introdução

A cultura da soja (*Glycine max L.*) é fundamental para a agricultura global, com impactos econômicos e ambientais significativos (TEHULIE *et al.*, 2021). A eficiência e sustentabilidade dessa cultura dependem do entendimento da variabilidade espacial dos atributos químicos e físicos do solo que afetam diretamente a produtividade da soja. Tais atributos são importantes, visto que a soja é a base para produtos

---

<sup>2</sup> Artigo submetido ao periódico internacional Australian Journal of Crop Science, com classificação B2 no QUALIS/CAPES na área de Ciências Agrárias.

essenciais à saúde humana (CHI *et al.*, 2021), à nutrição animal (MONTEIRO *et al.*, 2021) e à produção de biodiesel (ZHU *et al.*, 2021).

Diante da relevância estratégica da soja, pesquisas em Agricultura de Precisão (AP) são cruciais para otimizar ganhos econômicos e minimizar os impactos ambientais. A AP, como metodologia inovadora, ajusta o manejo agrícola à variabilidade espacial dos fatores que afetam a produção, possibilitando benefícios como melhoria na qualidade das colheitas, redução ecológica, fortalecimento da sustentabilidade, segurança alimentar, estímulo ao crescimento econômico e aumento da lucratividade e produtividade (ZAIN *et al.*, 2024).

A implementação efetiva da AP requer o uso de métodos capazes de descrever e modelar a variabilidade espacial dos fenômenos naturais, bem como estimar valores em locais não amostrados. Nesse contexto, a geoestatística emerge como um método científico primordial, adaptando técnicas estatísticas tradicionais para capitalizar a estrutura de dependência espacial presente nos dados amostrados (URIBE-OPAZO *et al.*, 2021; 2023).

A estimativa dos parâmetros relacionados à produtividade das culturas e dos atributos do solo são de extrema importância para a adoção de práticas de manejo diferenciadas em áreas heterogêneas, sendo uma componente chave no planejamento agrícola para a instalação e manejo eficiente das culturas (SILVA *et al.*, 2008). A interpolação espacial de dados é uma prática recorrente em estudos que correlacionam a produtividade da soja com os atributos do solo, e diversas publicações têm abordado essa temática (MAURO *et al.*, 2018; DALPOSSO *et al.*, 2021; DANESH *et al.*, 2022; PEREIRA *et al.*, 2022).

Entre as diversas técnicas de interpolação espacial, a *krigagem* destaca-se frequentemente na AP pela sua eficácia (KILIÇ *et al.*, 2022; MAITI & MITRA, 2022). Originada nos trabalhos pioneiros do engenheiro de minas Daniel G. Krige (KRIGE, 1951) e refinada pelo matemático e geólogo Georges Matheron (1962), a *krigagem* baseia-se na teoria das variáveis regionalizadas, um pilar da geoestatística. Este método é conhecido como o melhor estimador linear não-tendencioso de mínima variância (*Best Linear Unbiased Estimated* - BLUE) (ISAAKS & SRIVASTAVA, 1989).

Alternativamente, os métodos de Função de Base Radial (RBF), como o *Thin Plate Spline* (TPS), desenvolvido por Duchon (1976), oferecem uma abordagem distinta para interpolações espaciais, gerando superfícies suaves que se assemelham à flexão de uma placa metálica fina sobre pontos fixos (BUNN *et al.*, 2020). O TPS

tem sido aplicado em uma variedade de estudo, incluindo mapeamento de precipitação (KATIPOĞLU, 2022), análise das condições ambientais do ar (KÜÇÜKTOPCU; CEMEK, 2022) e identificação de poluentes em locais contaminados (QIAO *et al.*, 2019).

Curiosamente, o TPS é considerado um caso particular de *krigagem*, compartilhando uma estrutura computacional semelhante (HUTCHINSON, 1995), mas distinguindo-se em aspectos teóricos e práticos. Uma das principais diferenças está na independência do TPS em relação à estimativa preliminar da estrutura de covariância (HUTCHINSON & GESSLER, 1994). Além disso, o TPS se destaca pela sua flexibilidade ao não depender de uma distribuição de probabilidade específica das variáveis, nem de um número mínimo de amostras, uma característica especialmente vantajosa para dados esparsos (HE & LI, 2024). Em contraste, a *krigagem* tradicional recomenda pelo menos 30 pares de pontos para o cálculo da semivariância (JOURNEL; HUIJBREGTS, 1978).

Apesar de sua comprovada eficácia em diversos campos, o uso do TPS para interpolação de covariáveis na Agricultura de Precisão (AP) permanece relativamente inexplorado. Este trabalho visa preencher essa lacuna ao demonstrar o potencial do TPS na estimativa de valores de covariáveis, fornecendo uma base mais detalhada e precisa para a geração de mapas de produtividade de soja.

Nesse contexto, a metodologia adotada integra as vantagens do TPS e da *krigagem* com *drift* externo, aplicando o TPS para interpolar covariáveis químicas e físicas do solo, consideradas como variáveis fixas, enquanto a *krigagem* com *drift* externo estima a produtividade da soja. As estimativas de covariáveis obtidas pelo TPS para cada pixel do mapa (de 5 x 5 metros) são usadas como *drift* no Modelo de *Krigagem*, oferecendo uma abordagem espacial integrada. Esse método contribui para um manejo agrícola mais eficiente, promovendo a otimização dos recursos e a sustentabilidade na produção de soja.

A combinação entre TPS e *krigagem* com *drift* externo representa um avanço significativo na precisão e confiabilidade dos dados espaciais, abrindo novas perspectivas para futuras pesquisas e práticas agrícolas mais eficientes e sustentáveis.

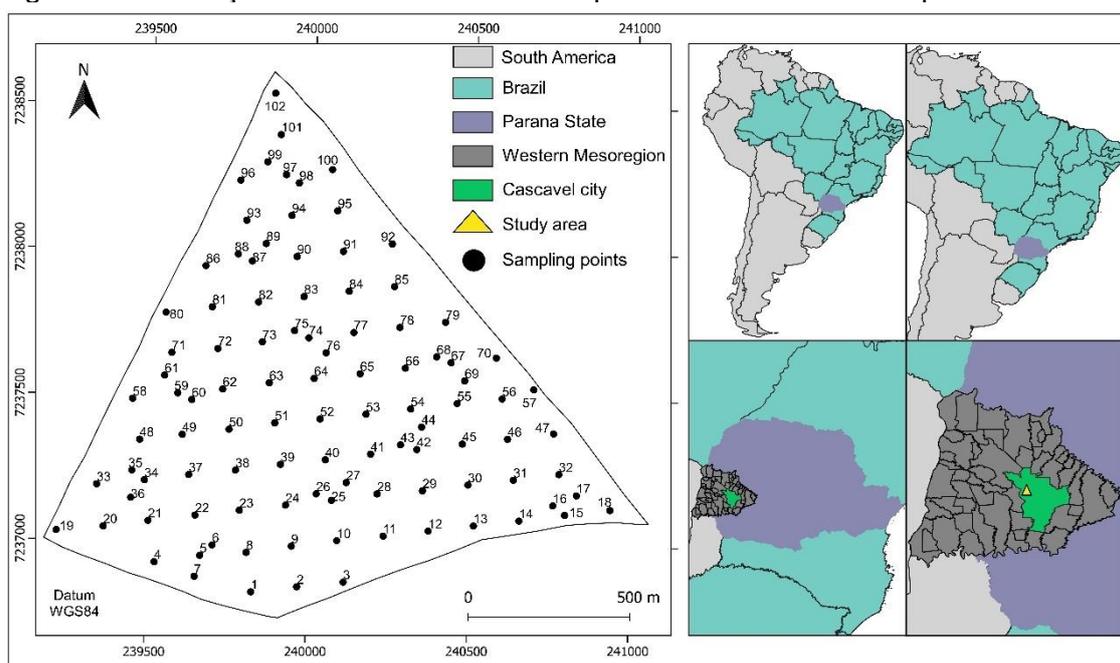
## 5.2.2 Material e Métodos

### 5.2.2.1 Área de estudo

Os dados de produtividade da soja (Prod) [ $t\ ha^{-1}$ ], do teor de Cálcio no solo (Ca) [ $cmolc\ dm^{-3}$ ], teor de Cobre do solo (Cu) [ $mg\ dm^{-3}$ ], teor de Potássio (K) [ $cmolc\ dm^{-3}$ ], teor de Fósforo (P) [ $mg\ dm^{-3}$ ], pH do solo (pH) e a resistência do solo à penetração na camada 0.21 a 0.30 m ( $RSP_{(0.21\ a\ 0.30m)}$ ) [MPa], foram selecionados devido à sua importância para o desenvolvimento da soja. Esses fatores são cruciais, pois a disponibilidade de nutrientes, o equilíbrio da acidez do solo e a resistência do solo à penetração podem afetar diretamente o crescimento radicular, a absorção de água e nutrientes, além de impactar a produtividade da cultura (HANSCH & MENDEL, 2009; FAGUNDES *et al.*, 2018; KELLER *et al.*, 2019; VANDERHASSELT *et al.*, 2023).

Os dados foram coletados durante o ano-safra 2022/2023 em uma área comercial de 172,03 ha localizada na região oeste do Paraná, Brasil, no município de Cascavel, com coordenadas geográficas aproximadas de 24°57'18"S para latitude, 53°34'29"W para longitude e altitude média de 750 m (Figura 1). O clima regional é mesotérmico e super húmido temperado, tipo climático Cfa (Köppen) e seu solo é classificado como Latossolo Vermelho Distroférico *típico de contextura argilosa* (SANTOS *et al.*, 2018).

Figura 1 Localização da área monitorada e o posicionamento dos 102 pontos amostrais.



Fonte: Os autores (2024).

Os 102 pontos amostrais foram definidos por meio de uma amostragem *lattice plus close pairs* (DIGGLE & RIBEIRO JR., 2007). Esse tipo de amostragem mescla uma grade regular de pontos com espaçamento  $\Delta$  (141 metros), a qual são acrescentados  $m$  pontos extras distribuídos aleatoriamente dentro de um círculo de raio  $\delta$ , sendo um com 75 metros e outro com 5 metros, ao redor de um ponto da grade também escolhidos de forma aleatória (DIGGLE & RIBEIRO JR., 2007; CHIPETA *et al.*, 2017). Essa abordagem cria uma cobertura espacial mais densa e uniforme, atenuando o efeito pepita, o que aumenta a precisão das estimativas espaciais. Os dados foram coletados manualmente e georreferenciados por GPS, assegurando alta fidelidade nas posições amostrais.

#### 5.2.2.2 Análise Exploratória

Realizou-se uma análise exploratória, na qual foram calculadas medidas de tendência central, variabilidade e assimetria. Para visualizar as distribuições e identificar possíveis valores discrepantes (outliers) nos dados amostrais, foram elaborados gráficos boxplots. Além disso, para observar o comportamento espacial dos pontos amostrais, foram construídos gráficos postplots.

As tendências direcionais, que representam uma associação linear entre os valores dos atributos químicos do solo e as coordenadas dos eixos X e Y, foram avaliadas pelo coeficiente de correlação linear de Pearson ( $r$ ), considerando-se que valores absolutos acima de 0,30 indicam tendência direcional, conforme recomendação de Callegari-Jaques (2003).

#### 5.2.2.3 Análise Geoestatística

Para modelar a estrutura de dependência espacial de uma variável regionalizada, considerou-se um processo estocástico  $\mathbf{Z} = \{\mathbf{Z}(\mathbf{s}), \mathbf{s} \in \mathcal{S}\}$  em que  $\mathbf{s} = (x, y)^T$  representa um local específico na área de estudo  $\mathcal{S} \subset \mathbb{R}^2$ , onde  $\mathbb{R}^2$  é o espaço euclidiano bidimensional. Supõe-se que os dados  $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$  constituem um processo estocástico estacionário de segunda ordem e isotrópico, coletados em locais conhecidos  $(\mathbf{s}_1, \dots, \mathbf{s}_n)$ .

Esse processo é modelado pelo conjunto  $\mathbf{Z} = \boldsymbol{\mu}(\mathbf{s}) + \boldsymbol{\varepsilon}(\mathbf{s})$ , onde o termo determinístico  $\boldsymbol{\mu}(\mathbf{s})$  é um vetor  $n \times 1$  das médias do processo  $\mathbf{Z}(\mathbf{s})$ ,  $\boldsymbol{\varepsilon} = (\boldsymbol{\varepsilon}(\mathbf{s}_1), \dots, \boldsymbol{\varepsilon}(\mathbf{s}_n))^T$  representa o vetor de erros aleatórios,  $n \times 1$ , com distribuição

normal  $n$ -variada, em que  $E[\varepsilon(\mathbf{s})] = \mathbf{0}$  e matriz de covariância  $\Sigma$ , de dimensão  $n \times n$ , é definida como  $\Sigma = \Sigma[(\sigma_{ij})] = \mathbf{C}(\mathbf{s}_i, \mathbf{s}_j), i, j = 1, \dots, n$  (URIBE-OPAZO *et al.*, 2012; 2023).

O vetor de médias  $\boldsymbol{\mu}(\mathbf{s})$  pode ser representado por um modelo linear espacial da forma  $\boldsymbol{\mu}(\mathbf{s}) = \mathbf{X}\boldsymbol{\beta}$ , onde  $\mathbf{X} = \mathbf{X}(\mathbf{s}) = [1, x_{j1}(\mathbf{s}), \dots, x_{jp}(\mathbf{s})]$  é uma matriz de dimensões  $n \times (p + 1)$  de variáveis explicativas, para  $j = 1, \dots, p$  e  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^\top$  é um vetor  $(p + 1) \times 1$  de parâmetros desconhecidos que necessitam ser estimados (URIBE-OPAZO *et al.*, 2012; 2023), que fornecem uma base sólida para a compreensão da variabilidade espacial em contextos agrícolas.

A matriz de covariância  $\Sigma$  é simétrica e definida positiva, com elementos  $\mathbf{C}(s_i, s_j)$  que dependem da distância Euclidiana  $d_{ij} = \|s_i - s_j\|$  entre os pontos  $s_i$  e  $s_j$ , sendo por vezes denotado por  $\mathbf{C}(d_{ij})$  ou  $\mathbf{C}(d)$ . A estrutura da matriz  $\Sigma$  é influenciada pelos parâmetros  $\boldsymbol{\varphi} = (\varphi_1, \dots, \varphi_s)^\top$  conforme estabelecido pela Equação 1 (URIBE-OPAZO *et al.*, 2012; 2023):

$$\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3), \quad (\text{Eq. 1})$$

Em que:  $\varphi_1 \geq 0$  é conhecido como efeito pepita;  $\varphi_2 \geq 0$  como contribuição;  $\mathbf{R} = \mathbf{R}(\varphi_3) = [(r_{ij})]$  é uma matriz simétrica  $n \times n$ , em função de  $\varphi_3 > 0$ , com elementos na diagonal  $r_{ii} = 1, j = 1, \dots, n$ ;  $r_{ij} = \varphi_2^{-1} \mathbf{C}(s_i, s_j)$  para  $\varphi_2 \neq 0$  e  $r_{ij} = 0$  para  $\varphi_2 = 0, i \neq j = 1, \dots, n$  onde  $r_{ij}$  é dependente de  $d_{ij}$ ;  $\varphi_3$  é determinado pelo Modelo de Alcance ( $a = g(\varphi_3)$ ) (URIBE-OPAZO *et al.*, 2012; 2023).

Para investigar a estrutura de dependência espacial, foram construídos semivariogramas omnidirecionais utilizando os estimadores de Função Semivariância de Matheron, como exposto na Equação 2 (CRESSIE, 2015).

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [(Z(s_i) - Z(s_i + h))^2], \quad (\text{Eq. 2})$$

Em que:  $\hat{\gamma}(h)$  é o estimador da função semivariância de Matheron;  $N(h)$  é o número de pares de valores amostrados em localizações separadas pela distância  $h$ ;

$Z(s_i + h)$  são os valores da variável  $Z$  nos pontos  $s_i + h$ ;  $Z(s_i)$  são os valores da variável  $Z$  nos pontos  $s_i$ .

O Modelo de *Wave*, conforme descrito por Olea (2006) e demonstrado na Equação 3, foi empregado para modelar as estruturas de dependência espacial, com os parâmetros estimados pelo método de máxima verossimilhança (ML). Além do Modelo de *Wave*, os modelos Exponencial, Gaussiano e Família Matérn, com valores de  $k = 0,7; 1,0$  e  $2,0$ , foram utilizados na análise (URIBE-OPAZO *et al.*, 2023). Os melhores ajustes foram escolhidos pelos critérios de Validação Cruzada (FARACO *et al.*, 2008), Critério de Informação de Akaike – AIC (AKAIKE, 1973) e Critério de Informação Bayesiano de Shwarz – BIC (SHWARZ, 1978).

$$\gamma(h) = \begin{cases} \varphi_1 + \varphi_2 \left( 1 - \frac{\varphi_3}{h} \operatorname{sen} \left( \frac{h}{\varphi_3} \right) \right), & h > 0 \\ 0, & h = 0 \end{cases}, \quad (\text{Eq. 3})$$

Para o Modelo de *Wave*, as derivadas de primeira e segunda ordem da estrutura de covariância  $\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \left( \frac{\varphi_3}{h} \sin \left( \frac{h}{\varphi_3} \right) \right)$  são, como exposto na Equação 4:

$$\begin{aligned} \frac{\partial \Sigma}{\partial \varphi_1} &= \mathbf{I}_n; \quad \frac{\partial \Sigma}{\partial \varphi_2} = \mathbf{R}(\varphi_3); \quad \frac{\partial \Sigma}{\partial \varphi_3} = \varphi_2 \left[ \frac{1}{h} \sin \left( \frac{h}{\varphi_3} \right) - \frac{1}{\varphi_3} \cos \left( \frac{h}{\varphi_3} \right) \right]; \quad \frac{\partial^2 \Sigma}{\partial \varphi_1^2} = 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_2} = \\ 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_3} &= 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_2^2} = 0; \quad \frac{\partial^2 \Sigma}{\partial \varphi_2 \partial \varphi_3} = \frac{1}{h} \sin \left( \frac{h}{\varphi_3} \right) - \frac{1}{\varphi_3} \cos \left( \frac{h}{\varphi_3} \right); \quad \frac{\partial^2 \Sigma}{\partial \varphi_3^2} = \\ -\frac{\varphi_2 h}{\varphi_3^3} \sin \left( \frac{h}{\varphi_3} \right) & \end{aligned} \quad (\text{Eq. 4})$$

#### 5.2.2.4 Máxima Verossimilhança (ML)

Em um contexto de modelagem geoestatística, considera-se um modelo  $\mathbf{Z} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \Sigma)$  com média  $\mathbf{X}\boldsymbol{\beta}$  e matriz de covariância  $\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3)$ , que depende exclusivamente do vetor de parâmetros  $\boldsymbol{\varphi}$  (DIGGLE; RIBEIRO, 2007). O logaritmo da função verossimilhança, essencial para a inferência estatística em modelos espaciais, é definido pela Equação (5).

O vetor de parâmetros desconhecidos  $\boldsymbol{\theta} = (\boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top)$  com  $\boldsymbol{\beta} = (\beta_0, \dots, \beta_p)^\top$  e  $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^\top$  pode ser estimado maximizando o logaritmo da função verossimilhança

(MARDIA; MARSHAL, 1984). Essa abordagem fornece um método robusto para a estimação de parâmetros em modelos geoestatísticos aplicados.

$$\mathcal{L}(\boldsymbol{\theta}) = -\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}), \quad (\text{Eq. 5})$$

As funções escores de  $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)$  são:  $\mathbf{U}(\boldsymbol{\beta}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$  e  $\mathbf{U}(\boldsymbol{\varphi}) = \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}}$  que tem como  $j$ -ésimo elemento  $\frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \varphi_j} = -\frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right\} + \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$ ,  $j = 1, 2, 3$  e  $\boldsymbol{\varepsilon} = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$ .

Para estimar os parâmetros  $\boldsymbol{\beta}$  e  $\boldsymbol{\varphi}$  resolve-se os sistemas de equações  $\mathbf{U}(\boldsymbol{\beta}) = \mathbf{0}$  e  $\mathbf{U}(\boldsymbol{\varphi}) = \mathbf{0}$ .

A matriz de informação observada é definida por  $\mathcal{L}(\boldsymbol{\theta}) = -\mathbf{L}(\boldsymbol{\theta})$  e deve ser avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$ , sua forma particionada é  $\mathbf{L}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\beta}} & \mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\varphi}} \\ \mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\beta}} & \mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \end{pmatrix}$  em que,  $\mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\beta}} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = -\mathbf{X}_n^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_n$ ;  $\mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\varphi}} = \mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\beta}}^T = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}^T}$ , com elementos  $\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \beta \partial \varphi_j} = -\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$ ,  $j = 1, \dots, q$ ; sendo,  $\boldsymbol{\varepsilon} = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$ , e  $\mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T}$ , com elementos  $\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \varphi_i \partial \varphi_j} = \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} \right) \right\} + \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left\{ \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} - \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right\} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$ ,  $i, j = 1, 2, 3$ .

A matriz de informação esperada é da forma  $\mathbf{K}(\boldsymbol{\theta}) = E[-\mathbf{L}(\boldsymbol{\theta})]$ , é uma matriz diagonal de forma  $\mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\boldsymbol{\varphi}) \end{pmatrix}$ , em que  $\mathbf{K}(\boldsymbol{\beta}) = E[-\mathbf{L}_{\boldsymbol{\beta}\boldsymbol{\beta}}] = E \left[ -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right] = \mathbf{X}_n^T \boldsymbol{\Sigma}^{-1} \mathbf{X}_n$  e  $\mathbf{K}(\boldsymbol{\varphi}) = E[-\mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}] = E \left[ -\frac{\partial^2 \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T} \right]$ , que tem como elementos  $k_{ij}(\boldsymbol{\varphi}) = \frac{1}{2} \text{tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right]$  com  $i, j = 1, \dots, q$  (URIBE-OPAZO *et al.*, 2012).

Os erros padrão assintóticos podem ser calculados invertendo a matriz de informação  $\mathbf{K}(\boldsymbol{\theta})$  e calculando a raiz quadrada dos elementos da diagonal principal  $\mathbf{K}(\boldsymbol{\theta})^{-1}$  (URIBE-OPAZO *et al.*, 2012; 2023).

### 5.2.2.5 Thin Plate Spline – TPS

A técnica *Thin Plate Spline* (TPS) oferece uma abordagem única para a interpolação espacial. Para descrever as propriedades matemáticas do TPS e torná-

las mais plausíveis, analogias mecânicas são utilizadas (KELLER & BORKOWSKI, 2019). Imagine uma placa metálica flexível e fina, fixada em pontos correspondentes às amostras do terreno agrícola. Ao deformar essa placa para que passe em cada ponto amostral, o TPS modela a superfície de maneira a minimizar a energia total de curvatura, resultando em uma superfície suave e contínua (BALEK & MIZERA, 2013; KELLER & BORKOWSKI, 2019), que captura de forma mais precisa as variações espaciais.

Essa metodologia se destaca pela suavidade e robustez, sendo amplamente aplicada na interpolação de dados georreferenciados. O princípio básico do TPS está fundamentado na dependência espacial dos dados: primeiro, ajusta-se uma função espacial que interpola os pontos amostrais e minimiza a energia de flexão. Assim, os valores de qualquer ponto da superfície podem ser estimados por essa função ajustada (BINDHU; NARASIMHAN; SUDHEER, 2013; CHEN *et al.*, 2014).

Segundo Bookstein (1989), a placa metálica é representada por uma função  $z = f(x, y)$  e a energia de curvatura de em um ponto de coordenadas  $(x, y)$ , é quantificada pela integral dupla expressa na Equação 6:

$$I_f = \iint_{-\infty}^{\infty} \left[ \left( \frac{\partial^2 z}{\partial x^2} \right)^2 + 2 \left( \frac{\partial^2 z}{\partial x \partial y} \right)^2 + \left( \frac{\partial^2 z}{\partial y^2} \right)^2 \right] dx dy, \quad (\text{Eq. 6})$$

A Função *Spline* que minimiza essa energia de curvatura é descrita pela Equação 7:

$$f(x, y) = a + bx + cy + \sum_{i=1}^n w_i U(r), \quad (\text{Eq. 7})$$

Em que:  $a, b, c$  são os coeficientes da Spline, que representam o plano que melhor se ajusta aos dados;  $w_i$  ( $i = 1, \dots, n$ ) são os pesos atribuídos a cada ponto de controle; e  $U$  é a função de base radial que confere suavidade à superfície interpolada. A distância  $r$  é calculada entre o ponto de interesse  $(x, y)$  e cada ponto de controle.

A essência da TPS está na função de base radial  $U$ , definida pela Equação 8:

$$U(r) = r^2 \log(r), \quad (\text{Eq. 8})$$

Esta função indica que uma maior distância, resulta em um maior valor da função  $U(r)$ , e uma menor distância, menor valor de  $U(r)$ , por isso ocorre a suavização. Isto permite que o TPS ajuste a superfície interpolada de forma flexível e precisa às características complexas do terreno agrícola.

Este método é usado para interpolar os atributos do solo (cálcio, cobre, potássio, fósforo, pH e resistência do solo à penetração) considerados neste estudo como covariáveis fixas da produtividade da soja.

#### 5.2.2.6 Krigagem com Drift Externo

A *krigagem* com deriva externa ou *krigagem* com *drift* externo (KDE) proposta por Wackernagel (1995), é uma metodologia que visa obter a interpolação de mapas temáticos levando em consideração a presença de fatores externos que influenciam a variável de interesse. A principal diferença entre a KDE e a *krigagem* ordinária (*krigagem* comum) reside na consideração da média espacial variável. Enquanto a *krigagem* ordinária assume uma média constante, a KDE incorpora informações de variáveis auxiliares para modelar o processo médio.

O modelo KDE é expresso como:  $Y(s) = \beta \mathbf{1} + \varepsilon(s)$ , onde:  $Y(s)$  representa o valor da variável de interesse no local  $s$ ;  $\beta \mathbf{1}$  é o termo de deriva externa, que varia espacialmente com base nas variáveis auxiliares e  $\varepsilon(s)$  é o termo do erro (WEBSTER & OLIVER, 2007).<sup>5</sup>

Na KDE, as predições são realizadas de maneira semelhante à *krigagem* ordinária, mas com a extensão da matriz de covariância residual incorporando os preditores auxiliares (WEBSTER & OLIVER, 2007). De acordo com Hengl *et al.* (2003), as predições em um novo local não amostrado  $s_0$  são feitas por  $\hat{Y}_{KED}(s_0) = \delta_0^T Y$  em que,  $\delta_0$  é o vetor dos pesos KED e  $Y$  é o vetor das  $n$  observações iniciais.

#### 5.2.2.7 Influência Local

De Bastiani *et al.* (2015) desenvolveram a perturbação Zhu generalizada na variável resposta, definida como  $Z_\omega = Z + \Sigma^{-\frac{1}{2}} \omega$ , em que  $\omega$  é um vetor que pertence ao espaço de perturbações  $\Omega$ .

A função de log-verossimilhança perturbada para o modelo normal é expressa por:

$$\mathcal{L}(\boldsymbol{\theta}_\omega) = \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega}) = -\left(\frac{n}{2}\right) \log(2\pi) - \frac{1}{2} \log|\boldsymbol{\Sigma}| - \frac{1}{2} (\mathbf{Z}_\omega - \mathbf{X}\boldsymbol{\beta})^T \boldsymbol{\Sigma}^{-1} (\mathbf{Z}_\omega - \mathbf{X}\boldsymbol{\beta}).$$

Neste contexto, a matriz  $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_\beta^T, \boldsymbol{\Delta}_\varphi^T)^T$  avaliada em  $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$  e em  $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ , assume a forma  $\boldsymbol{\Delta}_\beta = -\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{\frac{1}{2}}$ ;  $\boldsymbol{\Delta}_\varphi = \frac{\partial^2 \mathcal{L}(\boldsymbol{\theta}|\boldsymbol{\omega})}{\partial \varphi \partial \boldsymbol{\omega}^T}$ , que tem como elementos  $\Delta_{\varphi_j} = \boldsymbol{\varepsilon}^T \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \mathbf{A}}{\partial \varphi_j} - \mathbf{D}_j \mathbf{A} \right]$ , para  $j = 1, \dots, q$ , em que  $\mathbf{A} = \boldsymbol{\Sigma}^{\frac{1}{2}}$ ,  $\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon}(\boldsymbol{\omega}_0) = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$ , e  $\mathbf{D}_j = \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1}$  (URIBE-OPAZO; BORSSOI; GALEA, 2012).

Considerando a matriz  $\mathbf{B} = \boldsymbol{\Delta}^T \mathbf{L}^{-1} \boldsymbol{\Delta}$  em que,  $\mathbf{L}$  é a matriz de informação observada, o elemento  $C_i = 2|b_{ii}|$ , para  $i = 1, \dots, n$ , onde  $b_{ii}$  é o elemento da diagonal principal da matriz  $\mathbf{B}$ . A partir da matriz  $\mathbf{B}$  pode-se utilizar o gráfico de  $C_i$  em função de  $i$  (ordem das observações) como técnica de diagnóstico para avaliar a existência de observações influentes.

Seja  $L_{max}$  o autovetor, normalizado, associado ao maior autovalor, em módulo, da matriz  $\mathbf{B}$ . Segundo Cook (1986) o gráfico dos elementos  $|L_{max}|$  em função de  $i$  (ordem das observações) pode revelar os elementos que, sob pequenas perturbações, exercem maior influência no afastamento da verossimilhança.

### 5.2.2.8 Comparações

A avaliação do grau de dependência espacial do modelo ajustado foi realizada utilizando o Índice de Dependência Espacial – SDI, conforme proposto por Seidel e Oliveira (2014),  $SDI = 0.589 \left( \frac{\varphi_2}{\varphi_1 + \varphi_2} \right) \min \left\{ 1; \left( \frac{\alpha}{0.5MD} \right) \right\} 100$ , em que,  $\varphi_1$  é o efeito pepita,  $\varphi_2$  é a contribuição,  $\alpha$  é o alcance e  $MD$  é a distância máxima entre dois pontos amostrados. A classificação deste índice, segue a categorização apresentada por Neto *et al.* (2018), que para o Modelo de Wave é: fraca se  $SDI \leq 11\%$ ; moderada se  $11\% < SDI \leq 24\%$  e forte se  $SDI > 24\%$ . Tal classificação é essencial para a interpretação analítica dos resultados e fundamental para a tomada de decisões informadas, baseadas na compreensão da dependência espacial evidenciada pelos modelos geoestatísticos ajustados.

A congruência entre os mapas interpolados foi mensurada por meio dos Índices de Acurácia Kappa ( $\hat{K}$ ) e Kappa Ponderado ( $\hat{K}_w$ ), originários dos trabalhos de Cohen (1960;1968). O índice Kappa Ponderado, que incorpora pesos ( $w_{ij}$ ) para a quantificação das discrepâncias entre os mapas, oferece uma métrica refinada para a

avaliação da similaridade. A classificação do índice Kappa Ponderado segue os mesmos critérios do índice Kappa, como estabelecido por Fleiss *et al.* (2003), se  $\hat{K}_w \geq 0.75$  indica alta similaridade, se  $0.4 < \hat{K}_w < 0.75$  indica moderada similaridade e se  $\hat{K}_w \leq 0.4$  indica baixa similaridade entre os mapas interpolados.

#### 5.2.2.9 Recursos Computacionais

Todas as análises foram realizadas utilizando o software R (R CORE TEAM, 2024). As semivariâncias foram calculadas utilizando as funções do pacote geoR, o qual também foi empregado para ajustar os modelos e gerar os mapas temáticos (RIBEIRO JR & DIGGLE, 2001). Para interpolar as covariáveis consideradas fixas utilizou-se o método TPS do pacote fields (FURRER *et al.*, 2009) e para calcular o índice Kappa Ponderado utilizou-se o pacote vcd (MEYER *et al.*, 2021).

### 5.2.3 Resultados e Discussões

#### 5.2.3.1 Análise Exploratória

Durante o período do ano-safra 2022/2023, a produtividade média da soja na área monitorada foi de  $1,533 \text{ t ha}^{-1}$ , considerada baixa em relação à produtividade estadual ( $3,860 \text{ t ha}^{-1}$ ) e nacional ( $3,507 \text{ t ha}^{-1}$ ) no mesmo período (CONAB, 2023). Esses resultados são inferiores aos encontrados por Dalposso *et al.* (2022), para a safra 2021/2023, que conduziram experimentos sobre a produtividade da soja na mesma cidade onde o experimento foi realizado.

Os resultados indicam que o teor médio de potássio (K) no solo foi de  $0,76 \text{ cmolc dm}^{-3}$  e o do fósforo foi de  $19,14 \text{ mg dm}^{-3}$ , ambos considerados muito altos, enquanto o valor médio do pH do solo (5,82) foi classificado como bom (SANTOS & SILVA, 2001). Esse resultado é superior ao encontrado por Dalposso *et al.* (2018) que realizaram experimentos sobre o teor do pH no solo na mesma região onde o estudo foi realizado. Quanto à  $RSP_{0,21-0,30m}$ , a média foi de 2,19 MPa, indicando um baixo nível de compactação do solo, havendo pouca limitação ao desenvolvimento radicular, conforme os critérios de Canarache (1990).

A produtividade da soja (Prod) e teor de Cu no solo apresentaram os maiores valores para o coeficiente de variação (39,59% e 62,69%) respectivamente. Maltauro

*et al.* (2023) encontraram resultados semelhantes a estes em um ensaio efetuado na mesma região onde este experimento foi realizado.

Tabela 1: Análise descritiva das variáveis resposta e explicativas.

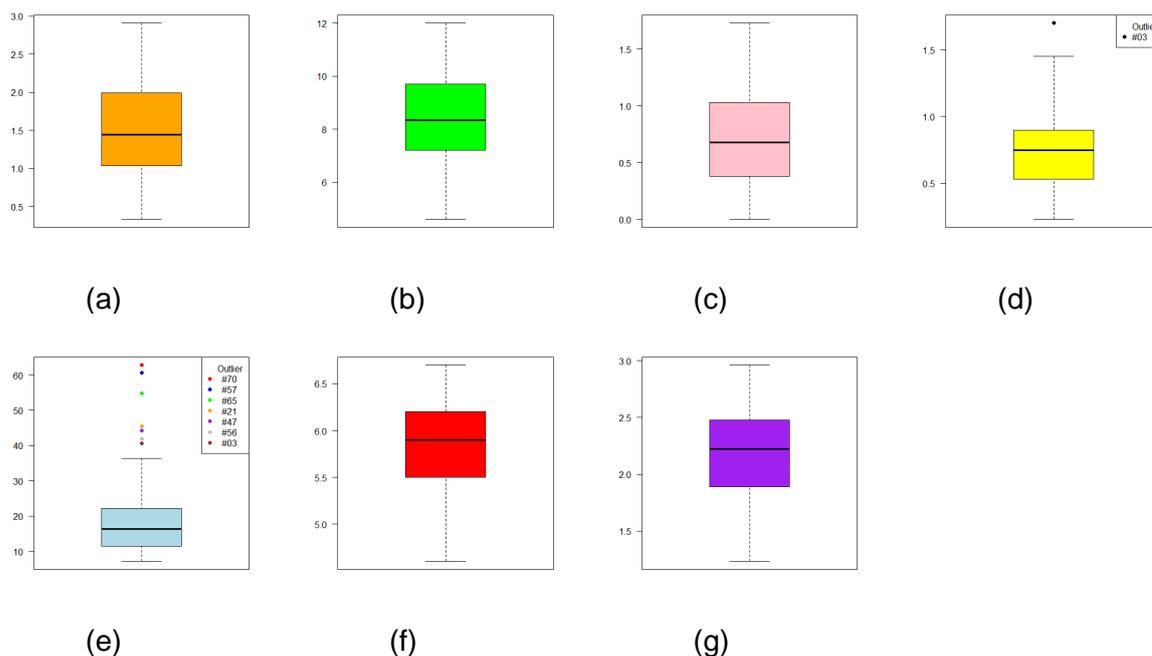
Estadística	Prod	Ca	Cu	K	P	pH	RSP0,21-0,30m
Mínimo	0,331	4,60	0	0,23	7,13	4,60	1,23
1º Quartil	1,045	7,20	0,38	0,53	12,39	5,50	1,89
Mediana	1,443	8,35	0,68	0,75	16,31	5,90	2,22
Média	1,534	8,46	0,69	0,76	19,14	5,82	2,19
3º Quartil	1,989	9,70	1,02	0,190	21,97	6,20	2,47
Máximo	2,909	12,00	1,73	1,70	62,79	6,70	2,96
CV(%)	39,59	19,60	62,69	37,44	57,79	7,40	17,39

Prod: Produtividade da soja (variável resposta); Ca: teor de Cálcio; Cu: teor de Cobre; K: teor de Potássio; P: teor de Fósforo; pH: pH do solo; RSP: Resistência do solo à penetração na camada 0,21 a 0,30m de profundidade; CV: Coeficiente de Variação.

Fonte: Os autores (2024).

Para visualizar a distribuição dos valores amostrais das variáveis em estudo, foram elaborados gráficos boxplot (Figura 2). As variáveis explicativas teor de potássio (K) (Figura 2(d)) e de fósforo (P) (Figura 2(e)) apresentaram outliers, indicando variações extremas que merecem atenção no contexto espacial.

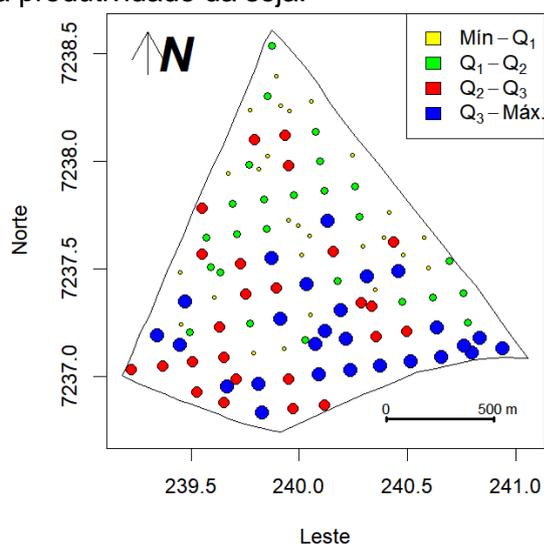
Figura 2 Gráficos boxplot para as variáveis em estudo: (a) produtividade da soja; (b) teor de cálcio; (c) teor de cobre; (d) teor de potássio; (e) teor de fósforo; (f) pH do solo e (g) resistência do solo à penetração na camada 0,21 a 0,30m de profundidade. Os valores destacados nos boxplots indicam os pontos amostrais identificados como outliers.



Fonte: Os autores (2024).

Complementando essa análise, a distribuição espacial dos valores amostrais foi representada por meio de Gráficos *Postplot* (Figura 3), permitindo observar a presença de agrupamentos e padrões direcionais, com valores altos tendendo a se agrupar ao sul e valores baixos ao norte. Essa tendência espacial é confirmada pelo teste de correlação de Pearson ( $r$ ), com um valor em módulo de 0,54 em relação ao eixo Y, o que indica uma associação linear moderada na direção norte-sul.

Figura 3 Gráfico postplot da produtividade da soja.



Fonte: Os autores (2024).

### 5.2.3.2 Análise Espacial

Considerando o modelo espacial linear, a produtividade da soja (Prod) em função das variáveis explicativas: Ca; Cu; K; P; pH e ( $RSP_{(0,21-0,30m)}$ ) é descrita como (Equação 9):

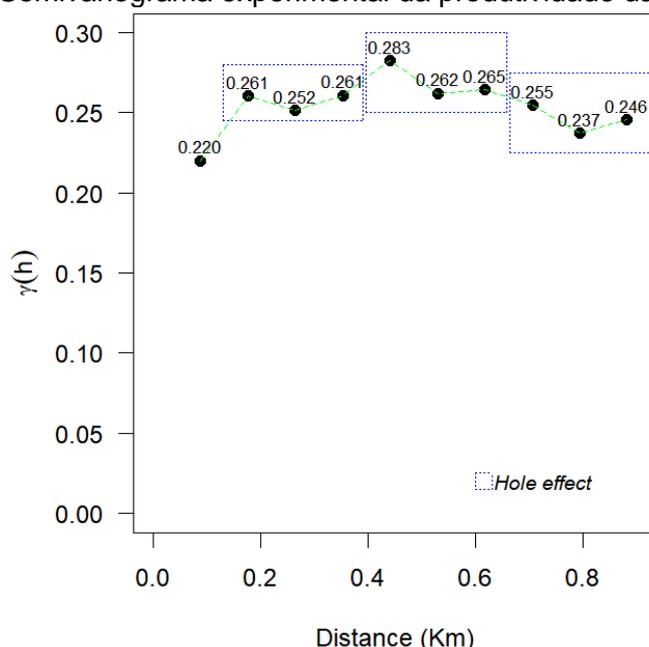
$$\begin{aligned} Prod(\mathbf{s}_i) = & \beta_0 + \beta_1 Ca(\mathbf{s}_i) + \beta_2 Cu(\mathbf{s}_i) + \beta_3 K(\mathbf{s}_i) + \beta_4 P(\mathbf{s}_i) + \beta_5 pH(\mathbf{s}_i) \\ & + \beta_6 RSP_{(0,21-0,30m)}(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \end{aligned} \quad (Eq. 9)$$

Com estrutura de covariância  $\Sigma = \varphi_1 I_n + \varphi_2 R(\varphi_3)$ .

Para a análise da dependência espacial, considerou-se 11 defasagens até a distância de 880 metros (ponto de corte de 50%) como recomenda (CLARK, 1979). Seguindo a recomendação de Guedes *et al.* (2013) o semivariograma foi verificado nas direções  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$  e  $135^\circ$ , confirmando que os dados são isotrópicos.

A análise do semivariograma experimental da produtividade da soja indicou um padrão oscilante nas semivariâncias, demonstrando aumentos e diminuições alternados, o que é indicativo do *hole effect* (Figura 4). Esse fenômeno é conhecido na AP e foi observado em estudos anteriores, tais como os conduzidos por Santra *et al.* (2008), Amirinejad *et al.* (2011) e Dalposso *et al.* (2022), sendo que o Modelo Geoestatístico de *Wave* se mostrou o mais adequado para o ajuste desses dados.

Figura 4 Semivariograma experimental da produtividade da soja para o ano-safra 2022/2023.



Fonte: Os autores (2024).

O Modelo de *Wave* é particularmente apropriado para dados com periodicidade, pois captura o “*hole effect*” resultante das correlações positivas e negativas entre regiões distantes, conforme descrito por Mahdi *et al.* (2020) e reiterado por Dalposso *et al.* (2022).

Após a aplicação do Critério de Informação de Akaike – AIC (AKAIKE, 1973), Critério de Informação Bayesiano de Schwarz – BIC (SHWARZ, 1978) e os critérios de Validação Cruzada (FARACO *et al.*, 2008), o Modelo de *Wave* se destacou como o mais eficaz na descrição da variabilidade espacial da produtividade da soja.

O desempenho superior do Modelo de *Wave* deve-se à sua capacidade de modelar a produtividade da soja em função das seguintes variáveis explicativas: teor de Ca, teor de Cu, teor de P, teor de K, pH do solo e  $RSP_{(0,21 a 0,30m)}$ .

Fundamentado nos resultados de dependência espacial do Modelo de *Wave* (Tabela 2) pode-se afirmar que a produtividade da soja em função das variáveis

explicativas está associada a uma forte dependência espacial entre as observações com SDI > 24%.

Esses resultados (Tabela 2) fornecem informações valiosas para a compreensão dos padrões de variabilidade espacial da produtividade da soja e dos atributos do solo. As covariáveis teor de K e teor de P nesta análise não foram significativas a 5 % de probabilidades no modelo espacial linear segundo o teste da razão de verossimilhança (URIBE-OPAZO *et al.* 2023).

Tabela 2: Parâmetros estimados por Máxima verossimilhança-ML do modelo linear espacial considerando a função de covariância Wave e erros padrão assintóticos entre parênteses.

Estimativas	Valor do parâmetro	Estimativas	Valor do parâmetro	Estimativas	Valor do parâmetro
$\hat{\beta}_0$	3,243* (0,766)	$\hat{\beta}_4$	-0,0005 <sup>ns</sup> (0,005)	$\hat{\varphi}_2$	0,219 (0,189)
$\hat{\beta}_1$	0,113* (0,046)	$\hat{\beta}_5$	-0,433* (0,156)	$\hat{\varphi}_3$	0,544 (0,026)
$\hat{\beta}_2$	-0,108* (0,138)	$\hat{\beta}_6$	-0,032* (0,152)	$\hat{a}(km)$	1,630
$\hat{\beta}_3$	0,144 <sup>ns</sup> (0,194)	$\hat{\varphi}_1$	0,217 (0,087)	SDI	26,56
				Class	Forte

$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5$  e  $\hat{\beta}_6$  denotam os parâmetros do modelo espacial linear gaussiano associados às covariáveis em estudo; \*: significativo ao nível de 5% de probabilidade; <sup>ns</sup>: não significativo ao nível de 5% de probabilidade;  $\hat{\varphi}_1$ : efeito pepita;  $\hat{\varphi}_2$ : contribuição;  $\hat{\varphi}_3$ : função de alcance;  $\hat{a}(km)$ : alcance; SDI: Índice de Dependência Espacial (%); Class: classificação.

Fonte: Os autores (2024).

### 5.2.3.3 Mapa Geoestatístico

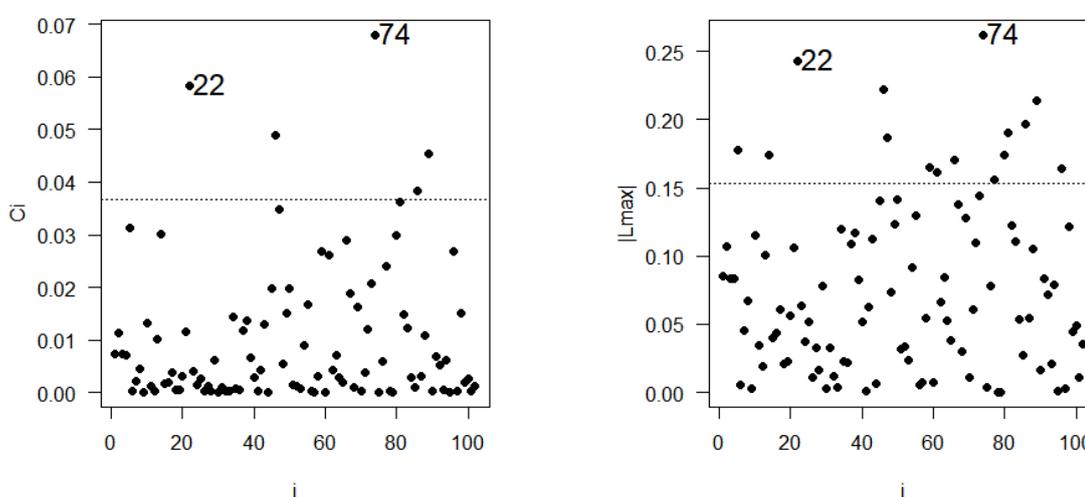
Com base na interpolação por *krigagem* com *drift* externo e nas informações obtidas do modelo geoestatístico Wave, que melhor se ajustou ao semivariograma, foi gerado o mapa da variável resposta produtividade da soja ( $t\ ha^{-1}$ ) em função das variáveis explicativas teor de cálcio (Ca), teor de cobre (Cu), teor de fósforo (P), teor de potássio (K), pH do solo (pH) e resistência do solo à penetração na camada 0,21 a 0,30m de profundidade ( $RSP_{0,21-0,30m}$ ) (Figura 6 (a)).

### 5.2.3.4 Diagnóstico de Influência Local

Os resultados da análise de influência local, ilustrados na Figura 5, revela a observação que possui influência significativa no afastamento da verossimilhança. De acordo com os gráficos  $C_i$  e  $|L_{max}|$  em função da ordem das observações, a observação #74 é a que possui maior influência no afastamento da verossimilhança.

Leiva *et al.* (2020) e Uribe-Opazo *et al.* (2023) apontam que em estatística espacial um ponto influente não é necessariamente um *outlier* e o contrário também é verdadeiro. Essa distinção revela a complexidade das relações espaciais e a necessidade de considerar cuidadosamente a influência de pontos individuais nos resultados da análise.

Figura 5 Gráficos (a)  $C_i$  e (b)  $|L_{max}|$  em função da ordem das observações  $i$  para a identificação de pontos influentes.



(a)

(b)

Fonte: Os autores (2024).

Com o objetivo de avaliar o efeito da observação influente na análise da estrutura de dependência espacial e na elaboração do mapa temático, realizou-se a exclusão da observação influente #74 do banco de dados, seguida de uma nova Análise de Variabilidade Espacial. Essa abordagem metodológica permite uma compreensão mais aprofundada da distribuição espacial das variáveis, considerando adequadamente a influência de observações individuais nos resultados da análise.

Considerando esse novo cenário, sem a presença da observação influente #74, foram ajustados, novamente, os modelos geoestatísticos Exponencial, Gaussiano, Wave e Matérn com parâmetro de suavização  $k=0,7, 1,0$  e  $2,0$ .

O modelo geoestatístico Wave continua sendo o modelo que melhor descreve a variabilidade espacial da produtividade da soja sem a presença do ponto influente #74 (Prod#74) em função das variáveis explicativas Ca, Cu, P, K, pH do solo e

$RSP_{0,21-0,30m}$  (Tabela 3). O valor da amostra retirada (ponto influente #74) foi de  $0,615 t ha^{-1}$ .

Embora a retirada do ponto influente não tenha alterado o modelo selecionado para descrever a variabilidade espacial, observa-se uma redução acentuada no raio de dependência espacial passando de 1,630 metros para 555 metros (redução de 1,075 metros), havendo uma transição na classificação do Índice de Dependência Espacial – SDI de forte para fraco ( $SDI \leq 11\%$ ), conforme a Tabela 3.

Tabela 3: Parâmetros estimados por ML do modelo linear espacial considerando a função de covariância Wave sem a observação #74 considerada influente e erros padrão assintóticos entre parênteses.

Estimativas	Valor do parâmetro	Estimativas	Valor do parâmetro	Estimativas	Valor do parâmetro
$\hat{\beta}_0$	3,380* (0,789)	$\hat{\beta}_4$	0,0027 <sup>ns</sup> (0,005)	$\hat{\varphi}_2$	0,029 (0,034)
$\hat{\beta}_1$	0,202* (0,040)	$\hat{\beta}_5$	-0,531* (0,157)	$\hat{\varphi}_3$	0,062 (0)
$\hat{\beta}_2$	-0,084* (0,137)	$\hat{\beta}_6$	-0,263* (0,143)	$\hat{a}(km)$	0,555
$\hat{\beta}_3$	0,165* (0,202)	$\hat{\varphi}_1$	0,219 (0,035)	SDI	4,38
				Class	Fraco

$\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\beta}_4, \hat{\beta}_5$  e  $\hat{\beta}_6$  denotam os parâmetros do modelo espacial linear gaussiano associados às covariáveis em estudo; \*: significativo ao nível de 5% de probabilidade; <sup>ns</sup>: não significativo ao nível de 5% de probabilidade;  $\hat{\varphi}_1$ : efeito pepita;  $\hat{\varphi}_2$ : contribuição;  $\hat{\varphi}_3$ : função de alcance;  $\hat{a}(km)$ : alcance; SDI: Índice de Dependência Espacial; Class: classificação.

Fonte: Os autores (2024)

Observa-se que, após a exclusão do ponto influente #74, apenas a covariável teor de P não alcançou significância ao nível de 5% de probabilidade no modelo espacial linear, evidenciando a influência deste ponto sobre o ajuste do modelo.

Embora o teor de P não seja estatisticamente significativo de forma isolada, ele demonstra significância conjunta, o que justifica sua permanência na formulação final do modelo espacial. Essa observação ressalta a complexidade e a interdependência das variáveis espaciais, ressaltando a importância de uma análise cuidadosa e criteriosa para assegurar a precisão e a robustez do Modelo desenvolvido.

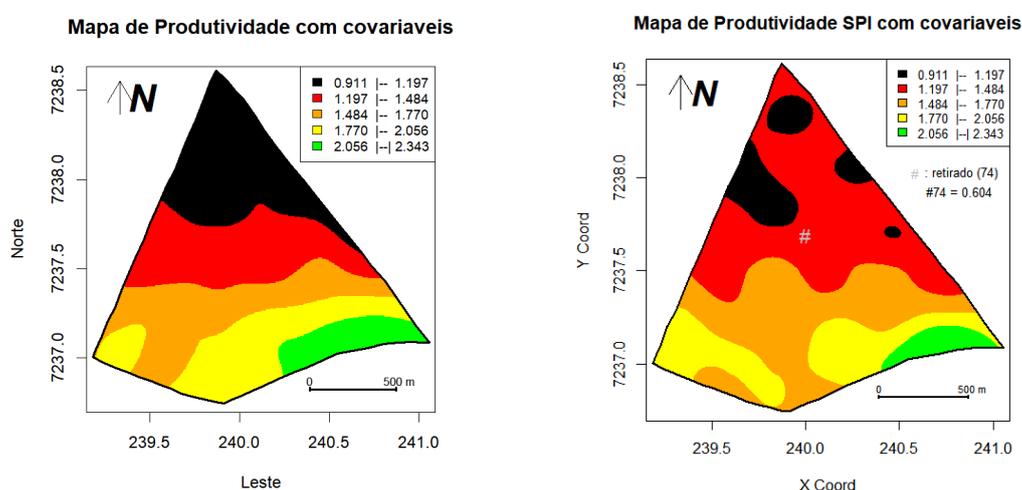
O teor de Ca é um dos componentes importantes para a integridade e funcionalidade da planta (MALAVOLTA; VITTI; OLIVEIRA, 1997); o teor de Cu no solo é relevante para o crescimento das plantas (CANCIAN, 2018); o teor de P é indispensável para o desenvolvimento vegetal (MENDES, 2007); o teor de K é fundamental para a capacidade de absorver água (SERAFIM, 2012); o pH é crucial

para o desenvolvimento da planta (NOVAES, 2007), e a RSP influencia na produtividade da soja e no desenvolvimento das raízes.

### 5.2.3.5 Comparação de Mapas

Ao comparar os mapas que incluem todas as observações (Figura 6a) com aquele sem a observação #74 considerada influente (Figura 6b), constatou-se que a distribuição espacial da produtividade da soja para o ano safra 2022/2023 é moderada ( $0,4 < \hat{K} \leq 0,75$ ) segundo o Índice de Acurácia Kappa ( $\hat{K}=0,49$ ).

Figura 6 Mapa temático da: (a) produtividade da soja com todas as observações e (b) sem a observação #74 considerada influente



(a)

(b)

Fonte: Os autores (2024).

As discrepâncias observadas entre o mapa com todas as observações e o mapa sem o ponto #74 reforçam a ideia de que a exclusão de observações influentes pode alterar consideravelmente os níveis de produtividade da soja. E enfatiza-se a importância de avaliar a influência das observações como parte integral da análise geoestatística.

### 5.2.3.6 Modelo de Predição

Utilizando as informações do Modelo de Wave para a predição dos valores da produtividade em função das variáveis explicativas Ca, Cu, P, K, pH do solo e  $RSP_{(0.21 \text{ a } 0.30m)}$ , os valores indicam que o modelo espacial linear é dado por (Equação 10):

$$\widehat{Prod}_{\#74}(s_i) = 3,380 + 0,202Ca(s_i) - 0,084Cu(s_i) + 0,165K(s_i) + 0,002P(s_i) - 0,531pH(s_i) - 0,263RSP_{(0,21 a 0,30m)}(s_i) \quad (Eq. 10)$$

Com estrutura de covariância para o Modelo de Wave  $\hat{\Sigma} = 0,219I_n + 0,029R(0,062)$ , sendo  $R(0,062) = \frac{0,029}{h} \text{sen}\left(\frac{h}{0,062}\right)$  e alcance  $a = 555\text{m}$  que é o raio de dependência espacial.

A estrutura possui dependência espacial fraca ( $SDI \leq 11\%$ ) e seu raio de dependência espacial é 66% menor que o modelo gerado com todos os pontos amostrais coletados.

#### 5.2.4 Conclusões

O Modelo de Wave foi o mais adequado para caracterizar a estrutura de dependência espacial da produtividade da soja considerando as variáveis explicativas teor de Ca, teor de Cu, teor de P, teor de K, pH do solo e  $RSP_{(0,21 a 0,30m)}$ . O Modelo de predição gerou mapas que não só estimaram a produtividade da soja em locais não amostrados, mas também viabilizaram a criação de zonas de manejo, baseadas na resposta da produtividade e nas variáveis explicativas. Tal abordagem permite intervenções localizadas que visam maximizar o retorno econômico e minimizar os impactos ambientais, resultando em uma gestão agrícola mais eficiente.

Foi constatado que a metodologia aplicada se mostrou eficaz na identificação de pontos influentes, cuja exclusão alterou as estimativas dos parâmetros que definem a estrutura de dependência espacial, impactando diretamente as características dos setores do mapa da produtividade da soja. Ressalta-se a importância da realização da metodologia de diagnóstico de influência local em análises geoestatísticas, visando assegurar a robustez e a precisão dos resultados.

Destaca-se que a metodologia descrita neste estudo possui uma aplicabilidade ampla, não se limitando apenas à produtividade da soja. Ela pode ser empregada em diversas culturas agrícolas e para a análise de diferentes atributos do solo, independentemente da região ou período avaliado. Essa versatilidade aumenta consideravelmente a relevância desta pesquisa no contexto da AP e do manejo sustentável de recursos naturais.

### 5.2.5 Agradecimentos

Este trabalho foi realizado com apoio do Conselho de Desenvolvimento Científico e Tecnológico (CNPq); Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES); Programa de Pós-Graduação em Engenharia Agrícola da Universidade Estadual do Oeste do Paraná (PGEAGRI-UNIOESTE); Universidade Tecnológica Federal do Paraná (UTFPR) e Laboratório de Estatística Espacial (LEE-UNIOESTE).

### 5.2.6 Referências

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. **Journal of the Royal Statistical Society**, v. 51, p. 469-483, 1973.

AMIRINEJAD, A. A.; KAMBLE, K.; AGGARWAL, P.; CHAKRABORTY, D.; PRADHAN, S.; MITTAL, R. B. Assessment and mapping of spatial variation of soil physical health in a farm. **Geoderma**, v.160, p. 292-303, 2011.

BALEK, V.; MIZERA, I. Mechanical models in nonparametric regression. **From Probability to Statistics and Back: High-Dimensional Models and Processes - A Festschrift in Honor of Jon A. Wellner**, v. 9, p. 5-20, 2013.

BINDHU, V, M.; NARASIMHAN, B.; SUDHEER, K. P. Development and verification of a non-linear disaggregation method (NL-DisTrad) to downscale MODIS land surface temperature to the spatial scale of Landsat thermal data to estimate evapotranspiration. **Remote Sensing**, v. 135, p. 118-129, 2013.

BOOKSTEIN, F. L. Principal warps: thin-plate splines and the decomposition of deformations. **IEEE Transactions on Pattern Analysis and Machine Intelligence**, v. 11, n. 6, p. 567-585, 1989.

BUNN, M.; LESHCHINSKY, B.; OLSEN, M.J. Estimates of three-dimensional rupture surface geometry of deepseated landslides using landslide inventories and high-resolution topographic data. **Geomorphology**, v. 367, p. 107332, 2020.

Callegari-Jacques, S. M. Bioestatística: Princípios e aplicações. Porto Alegre: Artmed. 2003.

CANARACHE, A. A generalized semi-empirical model estimating soil resistance to penetration. **Soil and Tillage Research**, v. 16, n.1, p. 51-70, 1990.

CANCIAN, M. Aplicação de cobre na cultura da soja em solos com altos teores de fósforo. 2018. **Dissertação (Mestrado - Programa de Pós-graduação em Ciência do Solo)** - Universidade Federal de Santa Maria, Santa Maria, 2018. Disponível em: <https://repositorio.ufsm.br/handle/1/13823>. Acesso em: 12 out. 2023.

CHEN, X.; LI, W.; CHEN, J.; RAO, Y.; YAMAGUCHI, Y. A Combination of TsHARP and Thin Plate Spline Interpolation for Spatial Sharpening of Thermal Imagery. **Remote Sensing**, v. 6, n. 4, p. 2845-2863, 2014.

CHI, J.; SONG, S.; ZHANG, H.; LIU, Y.; ZHAO, H.; DONG, L. Research on the Mechanism of Soybean Resistance to Phytophthora Infection Using Machine Learning Methods. **Frontiers in Genetics**, v. 1, n. 2, p. 634635, 2021.

CHIPETA, M. G.; TERLOUW, D. J.; PHIRI, K. S.; DIGGLE, P. J. Inhibitory geostatistical designs for spatial prediction taking account of uncertain covariance structure. **Environmetrics**, v. 28, n. 1, p. e2425, 2017.

CLARK, I. **Practical Geostatistics**. Applied Science Publishers LTD, London, 1979.

COHEN, J. A. coefficient of agreement for nominal scales. **Educ Psychol Meas**, v. 20, n. 1, p. 37-46, 1960.

COHEN, J. A. Weighted kappa: nominal scale agrément with provision for scale disagreement or partial credit. **Psychol. Bull**, v. 70, n. 4, p. 213-220, 1968.

Conab – Companhia Nacional de Abastecimento. Séries históricas: soja Brasil – série histórica de produtividade – safras 1976/1977 a 2022/2023. Disponível em: <<https://www.conab.gov.br/info-agro/safras/serie-historica-das-safras?start=30>>. Acesso em: 27 mai. 2024.

COOK, R. D. Assessment of local influence. **Journal of the Royal Statistical Society**, v. 48, n. 2, p. 133–169, 1986.

CRESSIE, N. **Statistics for spatial data**. New York: John Wiley and Sons, p. 928, 2015.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; JOHANN, J. A.; GALEA, M.; DE BASTIANI, F. Gaussian spatial linear of model soybean yield using bootstrap methods. **Engenharia Agrícola**, v. 38, n. 1, p. 110-116, 2018.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; DE BASTIANI, F. Spatial-temporal Analysis of Soybean Productivity Using Geostatistical Methods. **Journal of Agricultural Studies**, v. 9, n. 2, p. 283-303, 2021.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; OLIVEIRA, M. P. Comparison between Matheron and Genton semivariance function estimators in spatial modeling of soybean yield. **Australian Journal of Crop Science**, v. 16, n.7, p. 916-921, 2022.

DANESH, M.; TAGHIPOUR, F.; EMADI, S.M.; SEPANLOU, M.G. The interpolation methods and neural network to estimate the spatial variability of soil organic matter affected by land use type. **Geocarto International**, 2022. DOI: 10.1080/10106049.2022.2048905

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

DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. **Model based Geoestatistics**. New York: Springer, 2007.

DUCHON, J. **Splines minimizing rotation-invariant semi-norms in Sobolev Spaces: Constructive theory of functions of several variables**. Oberwolfach, W. Schempp and K. Zeller, Springer, Berlin-Heidelberg, 1976.

FAGUNDES, R. S.; URIBE-OPAZO, M. A.; GUEDES, L. P. C.; GALEA, M. Slash spatial linear modeling: soybean yield variability as a function of soil Chemical properties. **Revista Brasileira de Ciência do Solo**, v. 42, p. e0170030, 2018.

FARACO, M. A.; URIBE-OPAZO, M. A.; SILVA, E. A. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para a elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do solo**, v. 32, n.2, p. 463-476, 2008.

FLEISS, J. L.; LEVIN, B.; MYUNGHEE, C. P. **Statistical methods for rates and proportions**. 3 ed. Nova Jersey: John Wiley, p. 800, 2003.

FURRER, R.; NYCHKA, D.; SAIN, S. Package “fields”. **R package version**, v. 13, n.3, 2009.

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

HANSCH, R.; MENDEL, R. R. Physiological functions of mineral micronutrientes (Cu, Zn, Mn, Fe, Ni, Mo, B, Cl). **Current Opinion in Plant Biology**, v. 12, p. 259-266, 2009.

HENGL, T., GEUVELINK, G. B. M., & STEIN, A. Comparison of kriging with external *drift* and regression kriging. ITC. **Technical note**, 2003.  
[https://webapps.itc.utwente.nl/librarywww/papers\\_2003/misca/hengl\\_comparison.pdf](https://webapps.itc.utwente.nl/librarywww/papers_2003/misca/hengl_comparison.pdf).  
Acessado em: 29 mai. 2024.

HUTCHINSON, M. F. Stochastic space-time weather models from ground-based data. **Agricultural and Forest Meteorology**, v. 73, p. 237-264, 1995.

HUTCHINSON, M.F.; GESSLER, P.E. Splines more than just a smooth interpolator. *Geoderma*, v.62, p. 45-67, 1994.

ISAAKS, E. H.; SRIVASTAVA, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, p. 560, 1989.

KATIPOĞLU, O. M. Spatial analysis of seasonal precipitation using various interpolation methods in the Euphrates basin, Turkey. **Acta Geophysica**, v.70, p. 859–878, 2022.

KELLER, T.; SANDIN, M.; COLOMBI, T.; HORN, R.; OR, D. Historical increase in agricultural Machinery weights enhanced soil stress levels and adversely affected soil functioning. **Soil and Tillage Research**, v. 194, p. 104293, 2019.

KELLER, W.; BORKOWSKI, A. Thin Plate Spline interpolation. **Journal of Geodesy**, v. 93, p. 1251-1269, 2019.

KILIÇ, M.; GÜNDOĞAN, R.; GÜNAL, H.; CEMEK, B. Accuracy Assessment of Kriging, artificial neural network, and a hybrid approach integrating spatial and terrain data in estimating and mapping of soil organic carbon. **Plos One**, v. 17, n. 5, p. 0268658, 2022.

KRIGE, D.G. A statistical approach to some basic mine valuation problems on the Witwatersrand. **J Chemical, Metallurgical and Mining Soc.** South Africa, n. 52, p. 119– 139, 1951.

KÜÇÜKTOPCU, E.; CEMEK, B. A comparison of deterministic and stochastic models for predicting air and litter properties in a broiler building. **International Journal of Environmental Science and Technology**, 2022. DOI: 10.1007/s13762-022-03978-3

LEIVA, V.; SÁNCHEZ, L.; GALEA, M.; SAULO, H. Global and local diagnostic analytics for a geostatistical model based on a new approach to quantile regression. **Stochastic Environmental Research and Risk Assessment**, v. 34, p. 1457–1471, 2020.

MAHDI, E.; ABUZAIID, A. H.; ATTA, A. M. A. Empirical variogram for achieving the best valid variogram. **Commun Stat Appl Methods**, v. 27, p. 547-568, 2020.

MAITI, P.; MITRA, D. Ordinary kriging interpolation for indoor 3D REM. **Journal of Ambient Intelligence and Humanized Computing**, 2022. DOI: 10.1007/s12652-022-03784-2

MALAVOLTA, E.; VITTI, G. C.; OLIVEIRA, S. A. **Avaliação do estudo nutricional das plantas: princípios e aplicação**. 2 ed. Piracicaba: POTAFOS, p. 319, 1997.

MALTAURO, T. C.; GUEDES, L. P. C.; URIBE-OPAZO, M. A.; CANTON, L. E. D. Spatial multivariate optimization for a sampling redesign with a reduced sample size of soil chemical properties. **Revista Brasileira de Ciência do Solo**, v. 47, p. e0220072, 2023.

MARDIA, K. V.; MARSHAL, R. J. Maximum Likelihood Estimation of Models for Residual Covariance in Spatial Regression. **Biometrika**, v. 71, n. 1, p. 135-146, 1984.

MATHERON, G. **Traité de géostatistique appliquée, vol. I:** Memoires du Bureau de Recherches Géologiques et Minières, no. 14, Editions Technip, Paris, p. 333, 1962.

MAURO, G.D.; CIPRIOTTI, P.A.; GALLO, S.; ROTUNDO, J.L. Environmental and management variables explain soybean yield gap variability in Central Argentina. **European Journal of Agronomy**, v.99, p. 186-194, 2018.

Mendes, A. M. S. Introdução a fertilidade do solo. Superintendência Federal de Agricultura, Pecuária e Abastecimento do Estado da Bahia – SFA -BA/SDC/MAPA; 2007.

MEYER, D.; ZEILEIS, A.; HORNIK, K.; GERBER, F.; FRIENDLY, M. Package “vcd”. Visualizing Categorical Data. **R package version**, v. 1, p. 4-8, 2021.

MONTEIRO, A.; MIRANDA, C.; TRINDADE, H. Mediterranean Lupines as an Alternative Protein Source to Soybean. **Biology and Life Sciences Forum**, v. 3, n. 1, p. 38, 2021.

NETO, E. A.; BARBOSA, I. C.; SEIDEL, E. J.; OLIVEIRA, M. S. Spatial dependence index for cubic, pentaspherical and wave semivariogram models. **Bulletin of Geodetic Sciences**, v. 24, n. 1, p. 142-151, 2018.

NOVAES, R. F.; ALVAREZ, V. H.; BARROS, N. F.; FONTES, R. L. F.; CANTARUTTI, R. B.; NEVES, J. C. L. (eds.) Fertilidade do solo: fatores que influenciam o crescimento e desenvolvimento das plantas. Viçosa. **Sociedade Brasileira de Ciência do Solo**, p. 375-470, 2007.

OLEA, R. A. A six-step practical approach to semivariogram modeling. **Stoch Environ Res Risk Assess**, v.20, n.5, p. 307-318, 2006.

PEREIRA, G.W., VALENTE, D.S.M., DE QUEIROZ, D.M.; SANTOS, N.T.; FERNANDES-FILHO, E.I. Soil mapping for precision agriculture using support vector machines combined with inverse distance weighting. **Precision Agriculture**, 2022. DOI: 10.1007/s11119-022-09880-9.

QIAO, P.; LI, P.; CHENG, Y.; WEI, W.; YANG, S.; CHEN, T. Comparison of common spatial interpolation methods for analyzing pollutant spatial distributions at contaminated sites. **Environ Geochem Health**, v.41, p. 2709–2730, 2019.

RIBEIRO JUNIOR, P. J.; DIGGLE, P. J. geoR: A Package For Geostatistical Analysis. **R News**. V. 1, n. 2, p. 15–18, 2001.

R CORE TEAM R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria, 2024. <https://www.R-project.org>.

SANTOS, D. R.; SILVA, L. S. **Fertilidade do solo e nutrição de plantas**. Santa Maria: Universidade Federal de Santa Catarina, p. 20, 2001.

SANTOS, H. G.; JACOMINI, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. A.; LUMBRERAS, J. F.; COELHO, M. R.; ALMEIDA, J. A.; ARAÚJO FILHO, J. C.; OLIVEIRA, J. B.; CUNHA, T. J. F. **Sistema Brasileiro de classificação de solos**. Brasília, DF: Embrapa, 2018.

SANTRA, P.; CHOPRA, U. K.; CHAKRABORTY, D. Spatial variability of soil properties and its application in predicting surface map of hydraulic parameters in an agricultural farm. **Curr Sci**, v. 95, n. 7, p. 937-945, 2008.

SERAFIM, M. E.; ONO, F. B.; ZEVIANI, W. M.; NOVELINO, J. O.; SILVA, J. V. Umidade do solo e doses de potássio na cultura da soja. **Revista Ciência Agrônômica**, v. 43, n. 2, p. 222-227, 2012.

SHWARZ, G. Estimating the dimension of a model. **The Annals of Statistics**, v. 6, n. 2, p. 461-464, 1978.

SEIDEL, E. J.; OLIVEIRA, M. S. Novo índice geoestatístico para a mensuração da dependência espacial. **Revista Brasileira de Ciência do Solo**, v. 38, p. 699-705, 2014.

SILVA, S.A.; SOUZA LIMA, J.S.; SOUZA, G.S.; OLIVEIRA, R.B. Avaliação de interpoladores estatísticos e determinísticos na estimativa de atributos do solo em Agricultura de Precisão. **Idesia**, v. 26, n. 2, p. 75-81, 2008.

TEHULIE, N. S.; MISGAN, T.; AWOKE, T. Review on weeds and weed controlling methods in soybean (*Glycine max L.*). **Journal of Current Research in Food Science**, v. 2, n. 1, p. 1-6, 2021.

URIBE-OPAZO, M. A.; BORSSOI, J.; 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.; ASSUMPÇÃO, R. A. B. Influence diagnostics on a reparametrized *t*-Student spatial linear model. **Spatial Statistics**, v. 41, p. 100481, 2021.

URIBE-OPAZO, M. A.; DALPOSSO, G. H.; GALEA, M.; JOHANN, J. A.; DE BASTIANI, F.; MOYANO, E. N. C.; GRZEGOZEWSKI, D. M. Spatial variability of wheat yield using the gaussian spatial linear model. **AUSTRALIAN JOURNAL OF CROP SCIENCE**, v. 17, n. 2, p. 179-189, 2023.

VANDERHASSELT, A.; COOL, S.; D'HOUSE, T.; CORNELIS, W. How tine characteristics of subsoilers affect fuel consumption, penetration resistance and potato yield of a sandy loam soil. **Soil and Tillage Research**, v. 228, p. 105631, 2023.

WACKERNAGEL, H. **Multivariate geostatistics**. Berlin, GER: Springer Verlag, 1995.

WEBSTER, R.; OLIVER, M. A. **Geostatistics for environmental scientists**. 2 ed. Chichester: John Wiley & Sons, p. 312, 2007.

ZAIN, M.; MA, H.; RAHMAN, S. U.; NURUZZAMAN, M.; CHAUDHARY, S.; AZEEM, I.; MEHMOOD, F.; DUAN, A.; SUN, C. Nanotechnology in precision agriculture: Advancing towards sustainable crop production. **Plant Physiology and Biochemistry**, v. 206, p. 108244, 2024.

ZHU, Z.; LIU, Y.; CONG, W.; ZHAO, X.; JANAUN, J.; WEI, T. Soybean biodiesel production using synergistic CaO/Ag nano catalyst: Process optimization, kinetic study, and economic evaluation. **Industrial Crops & Products**, v. 166, p. 113479, 2021.

## 6 CONSIDERAÇÕES FINAIS

Com base nas análises realizadas, o Modelo de *Wave* se destacou em relação aos demais modelos na caracterização da estrutura de dependência espacial da produtividade da soja e dos atributos químicos e físicos do solo para o ano-safra em estudo.

Os mapas gerados por esse modelo não apenas permitiram a predição das variáveis em locais não amostrados dentro da área de estudo, mas também viabilizaram a criação de zonas de manejo, considerando tanto a produtividade da soja quanto os atributos do solo. Essa abordagem possibilita a aplicação de insumos de forma localizada, visando maximizar o retorno econômico e minimizar os impactos ambientais, resultando em um manejo agrícola eficiente.

Além disso, a identificação e remoção de observações influentes tiveram um impacto direto nas estimativas dos parâmetros que definem a estrutura de dependência espacial. A realização de estudos de diagnóstico de influência local em análises geoestatísticas é essencial para garantir a robustez e precisão dos resultados obtidos.

É importante ressaltar que a metodologia proposta neste estudo pode ser aplicada não apenas à produtividade da soja, mas também à outras culturas agrícolas e diferentes atributos do solo, em diferentes regiões e períodos. Essa abrangência aumenta a relevância da pesquisa no contexto da Agricultura de Precisão e do manejo sustentável de recursos naturais.

## APÊNDICE A

### MODELO DE WAVE

Função semivariância

$$\gamma(h) = \begin{cases} \varphi_1 + \varphi_2 \left( 1 - \frac{\varphi_3}{h} \text{sen} \left( \frac{h}{\varphi_3} \right) \right), & h > 0 \\ 0, & h = 0 \end{cases}$$

Função de covariância

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \text{sen} \left( \frac{h}{\varphi_3} \right), & h > 0 \end{cases}$$

Para processo estacionário  $\gamma(h) = C(0) - C(h)$ .

Seja  $\Sigma = \varphi_1 I_n + \varphi_2 R(\varphi_3)$  a matriz de covariância.  $\Sigma = [(\sigma_{ij})] = \text{Cov}[z(\mathbf{s}_i), z(\mathbf{s}_j)] = C(h_{ij})$  em que,  $h_{ij}$  é a distância euclidiana entre  $Z(\mathbf{s}_i)$  e  $Z(\mathbf{s}_j)$  onde,  $\mathbf{s}_i, \mathbf{s}_j \in \mathcal{S} \subset \mathbb{R}^2$ .

$R[(r_{ij})]$  é uma matriz simétrica cujos elementos

$$r_{ij} = \begin{cases} 1, & h_{ij} = 0 \\ \frac{\varphi_3}{h_{ij}} \text{sen} \left( \frac{h_{ij}}{\varphi_3} \right), & h_{ij} > 0 \end{cases}$$

$r_{ij} = r(\varphi_3)$  é uma função de  $\varphi_3$ .

As derivadas de primeira e segunda ordem de  $\Sigma$  em relação a  $\varphi_i$  são:

$$\frac{\partial \Sigma}{\partial \varphi_1} = I_n$$

$$\frac{\partial \Sigma}{\partial \varphi_2} = R(\varphi_3) = \frac{\varphi_3}{h} \text{sen} \left( \frac{h}{\varphi_3} \right)$$

$$\frac{\partial \Sigma}{\partial \varphi_3} = \varphi_2 \frac{\partial R(\varphi_3)}{\partial \varphi_3} = \left[ \varphi_2 \left( \frac{1}{h} \text{sen} \left( \frac{h}{\varphi_3} \right) - \frac{1}{\varphi_3} \cos \left( \frac{h}{\varphi_3} \right) \right) \right]$$

$$\frac{\partial^2 \Sigma}{\partial \varphi_1^2} = \frac{\partial}{\partial \varphi_1} (I_n) = 0$$

$$\frac{\partial^2 \Sigma}{\partial \varphi_2^2} = \frac{\partial}{\partial \varphi_2} \left( \frac{\varphi_3}{h} \text{sen} \left( \frac{h}{\varphi_3} \right) \right) = 0$$

$$\frac{\partial^2 \Sigma}{\partial \varphi_3^2} = -\frac{\varphi_2 h}{\varphi_3^3} \text{sen} \left( \frac{h}{\varphi_3} \right)$$

$$\frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_2} = \frac{\partial}{\partial \varphi_1} \left( \frac{\varphi_3}{h} \operatorname{sen} \left( \frac{h}{\varphi_3} \right) \right) = 0$$

$$\frac{\partial^2 \Sigma}{\partial \varphi_1 \partial \varphi_3} = \frac{\partial}{\partial \varphi_1} \left( \varphi_2 \left( \frac{1}{h} \operatorname{sen} \left( \frac{h}{\varphi_3} \right) - \frac{1}{\varphi_3} \cos \left( \frac{h}{\varphi_3} \right) \right) \right) = 0$$

$$\begin{aligned} \frac{\partial^2 \Sigma}{\partial \varphi_2 \partial \varphi_3} &= \frac{\partial}{\partial \varphi_2} \left( \varphi_2 \left( \frac{1}{h} \operatorname{sen} \left( \frac{h}{\varphi_3} \right) - \frac{1}{\varphi_3} \cos \left( \frac{h}{\varphi_3} \right) \right) \right) \\ &= \varphi_2 \left[ \left( \frac{1}{h} \right) \frac{\partial}{\partial \varphi_3} \left( \frac{h}{\varphi_3} \right) \cos \left( \frac{h}{\varphi_3} \right) - \frac{\partial}{\partial \varphi_3} \left( \frac{1}{\varphi_3} \right) \cos \left( \frac{h}{\varphi_3} \right) - \left( \frac{1}{\varphi_3} \right) \frac{\partial}{\partial \varphi_3} \cos \left( \frac{h}{\varphi_3} \right) \right] \end{aligned}$$

## APÊNDICE B

### MÁXIMA VEROSSIMILHANÇA

Seja  $Z = (Z(s_1), Z(s_2), \dots, Z(s_n))^T$  um processo estocástico gaussiano  $Z \sim N_n(\mu, \Sigma)$  em que  $E(Z) = \mu = \beta \mathbf{1}_n$ ,  $\mathbf{1}_n$  é vetor de uns e  $Var(Z) = \Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3)$ . Como se tem um ensaio de tamanho  $n$  a função densidade de probabilidade é a função verossimilhança, dada por:

$$l(\boldsymbol{\theta}) = f_Z(Z(s_1), \dots, Z(s_n)) = \left(\frac{l}{2\pi}\right)^{\frac{n}{2}} \frac{1}{|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(\mathbf{Z} - \beta \mathbf{1}_n)^T \Sigma^{-1} (\mathbf{Z} - \beta \mathbf{1}_n)}$$

O logaritmo da função verossimilhança é dado por:

$$L(\boldsymbol{\theta}) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma| - \frac{1}{2} (\mathbf{Z} - \beta \mathbf{1}_n)^T \Sigma^{-1} (\mathbf{Z} - \beta \mathbf{1}_n)$$

em que o parâmetro  $\boldsymbol{\theta} = (\beta, \varphi_1, \varphi_2, \varphi_3)^T$  são desconhecidos a serem estimados.

A Função Escore de  $\boldsymbol{\theta}$  é  $U(\boldsymbol{\theta})$  em que,  $U(\boldsymbol{\theta}) = (U(\beta), U(\boldsymbol{\varphi}))^T$ .

$$U(\beta) = \frac{\partial L(\boldsymbol{\theta})}{\partial \beta} = \mathbf{1}_n^T \Sigma^{-1} (\mathbf{Z} - \beta \mathbf{1}_n)$$

$$U(\boldsymbol{\varphi}) = \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\varphi}} = (U(\varphi_1), U(\varphi_2), U(\varphi_3))^T$$

De forma geral

$$U(\varphi_i) = -\frac{1}{2} \left\{ Tr \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_i} \right) - \boldsymbol{\varepsilon}^T \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \varphi_i} \right) \Sigma^{-1} \boldsymbol{\varepsilon} \right\}$$

sendo  $\boldsymbol{\varepsilon} = (\mathbf{Z} - \beta \mathbf{1}_n)$ , para  $i = 1, 2, 3$ .

Assim, para o caso do Modelo de Wave

$$U(\varphi_1) = -\frac{1}{2} \{ Tr(\Sigma^{-1}) - \boldsymbol{\varepsilon}^T \Sigma^{-1} \Sigma^{-1} \boldsymbol{\varepsilon} \}$$

$$U(\varphi_2) = -\frac{1}{2} \left\{ Tr \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_2} \right) - \boldsymbol{\varepsilon}^T \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \varphi_2} \right) \Sigma^{-1} \boldsymbol{\varepsilon} \right\}$$

$$U(\varphi_3) = -\frac{1}{2} \left\{ Tr \left( \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_3} \right) - \boldsymbol{\varepsilon}^T \Sigma^{-1} \left( \frac{\partial \Sigma}{\partial \varphi_3} \right) \Sigma^{-1} \boldsymbol{\varepsilon} \right\}$$

em que,  $\frac{\partial \Sigma}{\partial \varphi_2} = \frac{\varphi_3}{h} \text{sen} \left( \frac{h}{\varphi_3} \right)$  e  $\frac{\partial \Sigma}{\partial \varphi_3} = \left[ \varphi_2 \left( \frac{1}{h} \text{sen} \left( \frac{h}{\varphi_3} \right) - \frac{1}{\varphi_3} \cos \left( \frac{h}{\varphi_3} \right) \right) \right]$ .

A Matriz de Informação Observada (Hensiana) é:

$$L = \begin{pmatrix} L_{\beta\beta} & L_{\beta\boldsymbol{\varphi}} \\ L_{\boldsymbol{\varphi}\beta} & L_{\boldsymbol{\varphi}\boldsymbol{\varphi}} \end{pmatrix}$$

em que:

$$L_{\beta\beta} = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\beta} = -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \mathbf{1}_n$$

e

$$L_{\beta\varphi} = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\varphi} = \left( \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\varphi_1}, \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\varphi_2}, \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\varphi_3} \right)^T$$

com

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\varphi_1} = \frac{\partial}{\partial\beta} \left[ -\frac{1}{2} \{ \text{Tr}(\boldsymbol{\Sigma}^{-1}) - \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \} \right] = \frac{\partial}{\partial\beta} (U(\varphi_1)) = -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\varphi_2} = \frac{\partial}{\partial\beta} \left[ -\frac{1}{2} \left\{ \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) - \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right\} \right] =$$

$$\frac{\partial}{\partial\beta} (U(\varphi_2)) = -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$$

$$\frac{\partial^2 L(\boldsymbol{\theta})}{\partial\beta\partial\varphi_3} = \frac{\partial}{\partial\beta} \left[ -\frac{1}{2} \left\{ \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) - \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right\} \right] =$$

$$\frac{\partial}{\partial\beta} (U(\varphi_3)) = -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$$

$$L_{\varphi\varphi} = \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\varphi\partial\varphi^T} = \left[ \left( \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\varphi_i\partial\varphi_j} \right) \right]_{3 \times 3}$$

cujos elementos são:

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\varphi_i\partial\varphi_j} &= \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right) - \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} \right] \\ &+ \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left[ \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} - \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right) - \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right) \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \end{aligned}$$

Deste modo, simplificando com os resultados obtidos acima, tem-se para o Modelo de Wave:

$$\begin{aligned} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\varphi_1\partial\varphi_1} &= \frac{1}{2} \text{Tr}[\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1}] - \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} [\boldsymbol{\Sigma}^{-1} + \mathbf{I}_n] \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\varphi_1\partial\varphi_2} &= \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \right] - \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left[ \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) + \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\varphi_1\partial\varphi_3} &= \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \right] + \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left[ -\boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) - \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial\varphi_2\partial\varphi_2} &= \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \right] \\ &- \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left[ \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) + \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \end{aligned}$$

$$\begin{aligned}\frac{\partial^2 L(\boldsymbol{\theta})}{\partial \varphi_2 \partial \varphi_3} &= \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \right] \\ &\quad + \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left[ \left( \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_2 \partial \varphi_3} \right) - \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) - \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \varphi_3 \partial \varphi_3} &= \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) - \left( \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_3 \partial \varphi_3} \right) \right] \\ &\quad + \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left[ \left( \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_3 \partial \varphi_3} \right) - \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \boldsymbol{\Sigma}^{-1} \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) - \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \left( \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) \right] \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}\end{aligned}$$

A Matriz de Informação Esperada de Fisher é

$$\mathbf{K}(\boldsymbol{\theta}) = E[L] = \begin{bmatrix} -E[L_{\beta\beta}] & -E[L_{\beta\varphi}] \\ -E[L_{\varphi\beta}] & -E[L_{\varphi\varphi}] \end{bmatrix}$$

$$E[L_{\beta\beta}] = E[-\mathbf{1}^T \boldsymbol{\Sigma}^{-1} \mathbf{1}_n] = -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \mathbf{1}_n$$

$$E[L_{\beta\varphi}] = E \begin{bmatrix} \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta \partial \varphi_1} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta \partial \varphi_2} \\ \frac{\partial^2 L(\boldsymbol{\theta})}{\partial \beta \partial \varphi_3} \end{bmatrix} = E \begin{bmatrix} E[-\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}] \\ E \left[ -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right] \\ E \left[ -\mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon} \right] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

$$E[L_{\beta\varphi}] = (0,0,0)^T = E[L_{\varphi\beta}]$$

$$E[L_{\varphi\varphi}] = \left[ -\frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right) \right] = [k_{ij}(\varphi)]$$

Assim,

$$\mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} k(\beta) & 0 \\ 0 & k(\varphi) \end{pmatrix}$$

$$k(\beta) = \mathbf{1}_n^T \boldsymbol{\Sigma}^{-1} \mathbf{1}_n$$

$$k(\varphi) = [(k_{ij}(\varphi))]$$

$$k_{ij}(\varphi) = \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right)$$

$$k_{11}(\varphi) = \frac{1}{2} \text{Tr}(\boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1}) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_1} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_1} \right]$$

$$k_{12}(\varphi) = \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_1} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right]$$

$$k_{13}(\varphi) = \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_1} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right]$$

$$\begin{aligned}
k_{21}(\varphi) &= \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_1} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \right] \\
k_{22}(\varphi) &= \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right] \\
k_{23}(\varphi) &= \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right] \\
k_{31}(\varphi) &= \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_1} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \right] \\
k_{32}(\varphi) &= \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_2} \right] \\
k_{33}(\varphi) &= \frac{1}{2} \text{Tr} \left( \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right) = \frac{1}{2} \text{Tr} \left[ \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_3} \right]
\end{aligned}$$