UNIVERSIDADE FEDERAL DE VIÇOSA

Otimização das decisões experimentais em cana-de-açúcar por meio de análise da variabilidade do solo, interpolação e agrupamento

Raphael Henrique Teixeira da Silva
Doctor Scientiae

VIÇOSA - MINAS GERAIS
2025

RAPHAEL HENRIQUE TEIXEIRA DA SILVA

Otimização das decisões experimentais em cana-de-açúcar por meio de análise da variabilidade do solo, interpolação e agrupamento

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

Orientador: Luiz Alexandre Peternelli

Coorientadores: Marcio H. P. Barbosa
Matheus de P. Ferreira

VIÇOSA - MINAS GERAIS
2025

Ficha catalográfica elaborada pela Biblioteca Central da Universidade Federal de Viçosa - Campus Viçosa

T
S586o
2025

Silva, Raphael Henrique Teixeira da, 1991-
Otimização das decisões experimentais em cana-de-açúcar por meio de análise da variabilidade do solo, interpolação e agrupamento / Raphael Henrique Teixeira da Silva. – Viçosa, MG, 2025.
1 tese eletrônica (78 f.): il. (algumas color.).

Inclui apêndices.
Orientador: Luiz Alexandre Peternelli.
Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Estatística, 2025.
Referências bibliográficas: f. 64-68.
DOI: https://doi.org/10.47328/ufvbbt.2025.813
Modo de acesso: World Wide Web.

1. Geologia - Métodos estatísticos. 2. Fertilidade do solo. 3. Decisão estatística. 4. Química do solo. I. Peternelli, Luiz Alexandre, 1966-. II. Universidade Federal de Viçosa. Departamento de Estatística. Programa de Pós-Graduação em Estatística Aplicada e Biometria. III. Título.

CDD 22. ed. 551.028

Bibliotecário(a) responsável: Bruna Silva CRB-6/2552

RAPHAEL HENRIQUE TEIXEIRA DA SILVA

Otimização das decisões experimentais em cana-de-açúcar por meio de análise da variabilidade do solo, interpolação e agrupamento

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

APROVADA: 13 de outubro de 2025.

Assentimento:

Raphael Henrique Teixeira da Silva
Autor

Luiz Alexandre Peternelli
Orientador

Essa tese foi assinada digitalmente pelo autor em 11/12/2025 às 15:12:07 e pelo orientador em 13/12/2025 às 18:29:23. As assinaturas têm validade legal, conforme o disposto na Medida Provisória 2.200-2/2001 e na Resolução nº 37/2012 do CONARQ. Para conferir a autenticidade, acesse https://siadoc.ufv.br/validar-documento. No campo 'Código de registro', informe o código **THH6.RNYM.XM5W** e clique no botão 'Validar documento'.

Dedico este trabalho à minha filha, Sophia, que desde o nascimento, é exemplo de força e coragem.

AGRADECIMENTOS

A Deus, por me conduzir nos momentos mais difíceis.
Aos meus pais, pelo suporte concedido durante a vida. Em especial a minha mãe Cristiane, pelo amor e apoio nos momentos difíceis.
À minha esposa, Sara, pelo amor e carinho.
À minha filha, Sophia, por trazer alegria, me fazer sorrir e me dar forças para acreditar em mim, mesmo quando pensei em desistir.
Ao meu amigo Átila, pelas valiosas contribuições e apoio na construção deste trabalho.
Aos colegas de trabalho da FUPAC – Ponte Nova, que em diversos momentos me auxiliaram nas rotinas profissionais, contribuindo para que eu alcançasse o objetivo de concluir o doutorado.
Aos colegas Matheus e Samanta, que colaboraram com as rotinas e análises deste estudo.
Ao professor Peternelli, pela humanidade, compreensão e apoio. Serei eternamente grato por sua confiança.
À Universidade Federal de Viçosa, pela oportunidade de realizar a pós-graduação.
Este trabalho foi realizado com o apoio das seguintes agências de pesquisa brasileiras: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001, Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

Conheça todas as teorias, domine todas as técnicas, mas ao tocar uma alma humana, seja apenas outra alma humana.
(Carl Gustav Jung).

RESUMO

SILVA, Raphael Henrique Teixeira da, D.Sc., Universidade Federal de Viçosa, outubro de 2025. **Otimização das decisões experimentais em cana-de-açúcar por meio de análise da variabilidade do solo, interpolação e agrupamento**. Orientador: Luiz Alexandre Peternelli. Coorientadores: Marcio Henrique Pereira Barbosa e Matheus de Paula Ferreira.

A fertilidade do solo influencia diretamente o desenvolvimento das plantas. A análise das variáveis químicas do solo é fundamental para embasar a tomada de decisão em estudos experimentais com cana-de-açúcar, pois cada variável contribui para características desejáveis, como a produtividade, fator essencial para essa cultura. Este estudo investigou a variabilidade do solo em uma quadra experimental de melhoramento genético de cana-de-açúcar, aplicando métodos geoestatísticos e a interpolação pelo inverso da distância para otimizar a tomada de decisões e reduzir custos no planejamento experimental. Para isso, foram coletadas amostras em 60 pontos de uma quadra da Unidade de Ensino, Pesquisa e Extensão em Melhoramento Genético de Cana-de-Açucar (UEPE), localizado em Oratórios – MG. A partir da análise das características químicas do solo, com foco nos teores de fósforo e potássio, e do ajuste do modelo de variograma, realizou-se a interpolação por meio da krigagem simples e pelo inverso da distância para a construção de mapas de distribuição espacial. Foi utilizada a técnica de agrupamento k-means para a segmentação das regiões conforme a concentração de fósforo e potássio. Foram identificadas as áreas homogêneas. Com os resultados obtidos, concluiu-se que é possível tomar decisões mais precisas, que podem economizar recursos financeiros no planejamento experimental, visto que o pesquisador já tem o conhecimento prévio dos locais onde possuem maior ou menor concentração de fósforo e potássio. E também, pela confirmação da viabilidade da redução de 20% dos pontos amostrais, reduzindo tempo de coleta de amostras e custos de análises. Pode-se otimizar a área experimental, uma vez que é possível alocar futuros experimentos de acordo com a região de fertilidade mais adequada.

Palavras-chave: Geoestatística; Delineamentos experimentais; Fertilidade do solo; Redução amostral

ABSTRACT

SILVA, Raphael Henrique Teixeira da, D.Sc., Universidade Federal de Viçosa, October, 2025. **Optimizing experimental decisions in sugarcane through soil variability analysis, interpolation and clustering**. Adviser: Luiz Alexandre Peternelli. Co-advisers: Marcio Henrique Pereira Barbosa and Matheus de Paula Ferreira.

Soil fertility directly influences plant development. Analyzing soil chemical variables is essential for decision-making in experimental studies with sugarcane, as each variable contributes to desirable characteristics, such as productivity, a key factor for this crop. This study investigated soil variability in an experimental sugarcane breeding plot, applying geostatistical methods and inverse-distance interpolation to optimize decision-making and reduce experimental planning costs. To this end, samples were collected from 60 points in a plot at the Sugarcane Genetic Breeding Teaching, Research, and Extension Unit (UEPE), located in Oratórios, Minas Gerais. Based on the analysis of soil chemical characteristics, focusing on phosphorus and potassium levels, and the adjustment of the variogram model, interpolation was performed using simple kriging and inverse-distance interpolation to construct spatial distribution maps. The k-means clustering technique was used to segment regions based on phosphorus and potassium concentrations. Homogeneous areas were identified. The results indicate that more accurate decisions can be made, potentially saving money in experimental planning, as researchers already have prior knowledge of the locations with higher or lower phosphorus and potassium concentrations. Furthermore, the feasibility of reducing sampling points by 20% can be confirmed, reducing sample collection time and analysis costs. The experimental area can be optimized, allowing future experiments to be allocated according to the most suitable fertility region.

Keywords: Geostatistics; Experimental designs; Soil fertility; Sample reduction.

SUMÁRIO

1. INTRODUÇÃO 10
2. REVISÃO DE LITERATURA 12
2.1. Geoestatística 12
2.2. Modelagem geoestatística: variograma experimental e modelos teóricos 13
2.3. Krigagem 16
2.3.1. Preditor 17
2.3.2. Tipos de métodos de krigagem 18
2.3.3. Krigagem simples 18
2.4. Validação cruzada 20
2.5. Interpolação pelo inverso da distância (IDW) 20
2.6. Algoritmo *k-means* para agrupamentos 22
2.7. Fertilidade do solo 23
2.7.1. Nutrientes do solo para cana-de-açúcar 24
2.8. Amostragem otimizada: Conditioned Latin Hypercube Sampling ($cLHS$) 26
3. MATERIAL E MÉTODOS 28
3.1. Coleta das amostras de solo 28
3.2. Análise exploratória dos dados originais 30
3.3. Análise geoestatística dos dados originais 31
3.4. Interpolação e agrupamento 32
3.5. Avaliação dos valores reduzidos de pontos de amostragem 33
3.6. Avaliação da distribuição sistemática e aleatória dos pontos de amostragem 34
3.7. Descrição da abordagem analítica para a correlação entre dados agroindustriais da cana-de-açúcar e a fertilidade do solo 34
3.8. Resumo da metodologia 36
4. RESULTADOS E DISCUSSÃO 38
4.1. Análise de fertilidade do solo e estatística descritiva 38
4.2. Histogramas e boxplot 41
4.3. Interpolação espacial 42
4.3.1. Análise geoestatística 42
4.3.2. Mapas obtidos via krigagem simples e IDW 43
4.3.3. Análise de agrupamento 45
4.4. Estudo da viabilidade ou não de redução de pontos amostrais 50
4.5. Estudo comparativo da análise sistemática versus aleatória 53

4.6. Estudo analítico entre os resíduos dos delineamentos experimentais e a fertilidade do solo 56
5. CONCLUSÕES 63
REFERÊNCIAS BIBLIOGRÁFICAS 64
APÊNDICE A – Tabela com os dados amostrais provenientes do resultado da análise de fertilidade do solo, para os 60 pontos amostrados 69
APÊNDICE B – Histogramas e boxplot 72
APÊNDICE C – Semivariogramas experimentais obtidos e a curva que representa o modelo teórico ajustado para fósforo e potássio 74
APÊNDICE D – Semivariogramas experimentais obtidos para os resíduos de brix e TCH em DIC e DBC. 75

1. INTRODUÇÃO

A cultura da cana-de-açúcar prospera em climas tropicais e subtropicais (Toppa et al., 2010), encontrando no Brasil condições ideais de cultivo. No contexto brasileiro, sua importância é notável, destacando-se tanto na cadeia de alimentação humana e animal quanto, crucialmente, como uma fonte de energia renovável (Godoi et al., 2004). Para sustentar esta produção e atender às crescentes demandas, a fertilização é essencial; a cana-de-açúcar é a terceira cultura que mais consome fertilizantes no Brasil, em termos de volume (USDA, 2024). Este esforço de intensificação é amparado pela necessidade global de garantir a segurança alimentar, sendo imprescindível dobrar os níveis de produção agrícola globalmente até o ano de 2050 (Formaggio; Sanches, 2017).

Neste cenário de expansão e necessidade energética, o Brasil é um grande produtor de cana-de-açúcar. Sua produção mais que dobrou nas últimas décadas para atender às demandas globais por bioenergia, visando a redução da dependência do petróleo bruto e a mitigação das alterações climáticas (Bordonal et al. 2018). Além disso, a implementação de tecnologias inovadoras no setor agrícola e industrial tem demonstrado que esses avanços contribuíram para uma maior sustentabilidade e aceitação dos produtos derivados da cana-de-açúcar (como etanol, açúcar e bioeletricidade) no mercado global (Bordonal et al., 2018).

Estudos de melhoramento de cana-de-açúcar se fazem necessários, uma vez que a perspectiva futura é de alta produção para diversas aplicações. O solo é um componente crucial nos estudos de melhoramento de cana-de-açúcar, pois afeta vários aspectos do crescimento, desenvolvimento e produtividade da planta (Silva et al., 2024). Compreender as propriedades do solo permite que os pesquisadores desenvolvam variedades de cana-de-açúcar adaptadas às condições específicas do solo, contribuindo para um cultivo mais eficiente e sustentável (Fernandes et al., 2023).

A utilização de informações provenientes do solo pode contribuir na qualidade da seleção de genótipos, uma vez que os nutrientes (propriedades químicas) exercem influência direta no desenvolvimento da cana-de-açúcar. Contudo, a absorção e o aproveitamento desses nutrientes são intrinsecamente ligados à saúde física do solo.

Nesse sentido, as propriedades físicas do solo podem sofrer alterações significativas devido às atividades agrícolas da cana-de-açúcar. Penatti (2013) aponta que a prática de manejo que utiliza implementos pesados, como arados, grades e subsoladores, é uma das causas dessas mudanças. O autor também salienta que o fato de o solo permanecer descoberto durante grande parte do tempo contribui para o adiantamento do processo erosivo e, consequentemente, para a degradação de suas propriedades físicas e químicas.

Ao realizar o estudo da variabilidade do solo, é possível identificar a quantidade de nutrientes em uma determinada região e expressar a sua fertilidade. Esse tipo de estudo também é fundamental para orientar o processo de amostragem e análises laboratoriais, garantindo resultados mais representativos e confiáveis (Penatti, 2013).

O conhecimento das características do solo transcende o ajuste de médias fenotípicas em programas de melhoramento, sendo fundamental para a otimização do manejo e a eficiência operacional da cultura. O principal benefício reside na capacidade de regionalizar e mapear as zonas homogêneas (Ferreira et al., 2024).

Ao quantificar e delimitar a variabilidade espacial do solo, é possível reduzir significativamente o número de pontos amostrais necessários para caracterização, o que viabiliza a redução de custos e de tempo de operação (Formaggio; Sanches, 2017). Dessa forma, a amostragem passa a ser mais precisa e representativa, permitindo uma aplicação localizada e racional de insumos. Isso garante que o manejo cultural e a aplicação de fertilizantes sejam ajustados de forma específica para cada região homogênea, mitigando os efeitos de variações ambientais não controladas e maximizando o potencial produtivo da cana-de-açúcar.

O objetivo deste trabalho foi propor o desenvolvimento e a validação de uma metodologia, por meio de métodos estatísticos, que utilize a informação da variabilidade do solo para otimizar a alocação dos pontos de amostragem, tornando-a aplicável a qualquer experimento de campo com o intuito de reduzir custos com análises laboratoriais de solo, reduzir tempo de coleta e análise, alocar experimentos de forma correta e melhorar a precisão experimental.

2. REVISÃO DE LITERATURA

2.1. Geoestatística

A geoestatística é uma subárea da Estatística Espacial utilizada na análise e previsão de valores relacionados a fenômenos espaciais ou espaço-temporais e é de suma importância para compreender a distribuição espacial das variáveis agrícolas. Para isso, as técnicas geoestatísticas possibilitam a compreensão de fenômenos espaciais que, à primeira vista, podem parecer apenas valores numéricos aleatórios das variáveis em estudo (Bernardi et al., 2014).

Segundo Goovaerts (1997), uma das características distintivas da geoestatística é a sua abordagem para lidar com a autocorrelação espacial dos dados, ou seja, a dependência entre as observações em diferentes localidades. Isso significa que, ao contrário da estatística tradicional, que assume independência entre as observações, a geoestatística reconhece e incorpora a estrutura espacial dos dados em suas análises.

A geoestatística utiliza ferramentas como o variograma, que descreve a variabilidade espacial dos dados, e a krigagem, que é uma técnica de interpolação espacial baseada na estrutura de dependência espacial modelada pelo variograma (Vieira, 2000). Essas ferramentas são fundamentais para a elaboração de mapas de previsão, estimativa de valores em locais não amostrados e tomada de decisões em relação ao gerenciamento de recursos naturais, planejamento urbano, agricultura de precisão, entre outras aplicações (Kumar et al., 2020).

Rodrigues et al. (2020) apresentaram uma revisão da literatura sobre geoestatística e seu potencial na agricultura, mostrando um resumo das abordagens e alguns usos práticos. Os autores concluíram que a geoestatística tem sido usada para fins agrícolas e produzido resultados interessantes a Agricultura 4.0, que é a quarta fase de evolução da agricultura, caracterizada pela integração de tecnologias digitais e de dados em toda a cadeia produtiva, com o objetivo de aumentar a eficiência, a produtividade e a sustentabilidade. O principal objetivo da geoestatística é compreender e modelar a variabilidade espacial de fenômenos naturais e humanos, permitindo previsões e tomadas de

decisão mais precisas em relação à distribuição e comportamento desses fenômenos ao longo do espaço (Chiles e Delfiner, 2012).

2.2. Modelagem geoestatística: variograma experimental e modelos teóricos

De acordo com Vieira (2000), o semivariograma é uma representação gráfica (conjunto de pontos) que descreve como a variância de certo atributo (ou variável) varia em relação ao vetor de distâncias $\boldsymbol{h}$, que separa dois pontos aleatórios amostrados. Isto é, pela definição teórica, o variograma é uma espécie de variância. É a medida da variabilidade entre duas variáveis Z, que distam $\boldsymbol{h}$, caso todas as malhas de amostragem fossem observadas.

O semivariograma é estimado por meio do semivariograma experimental (Cressie, 1985), conforme Equação 1.

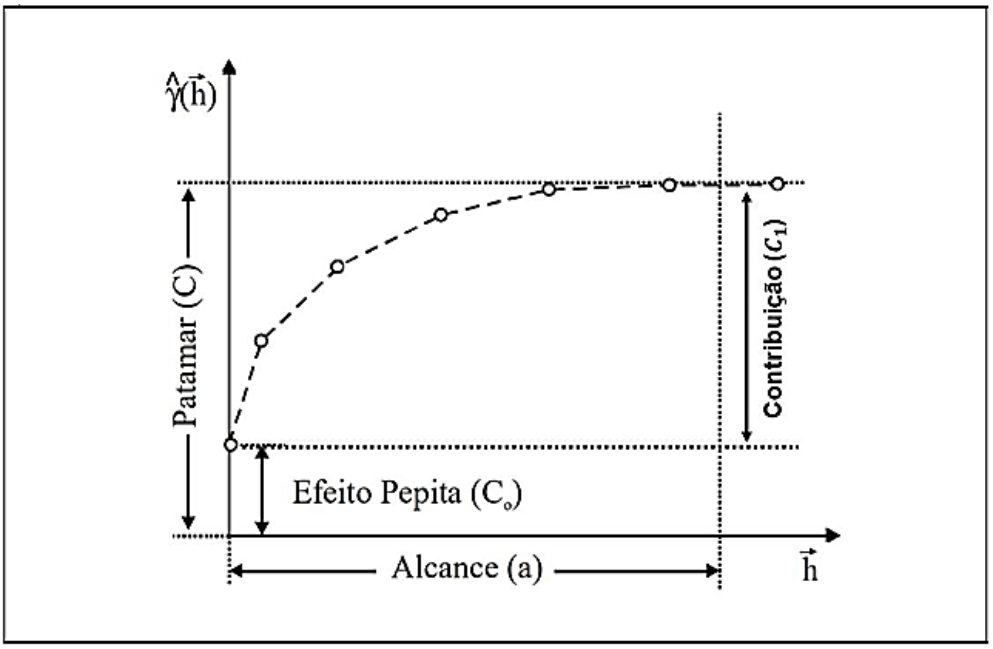
$$\hat{y}(\boldsymbol{h}) = \frac{1}{2N(\boldsymbol{h})}\sum_{i=1}^{N(\boldsymbol{h})}[Z(x_i) - Z(x_i + \boldsymbol{h})]^2 \quad (1)$$

em que $N(\boldsymbol{h})$ é o número de pares de valores medidos $Z(x_i)$ e $Z(x_i + \boldsymbol{h})$, separados por $\boldsymbol{h}$.

Calculando as semivariâncias estimadas para cada $\boldsymbol{h}$, é possível descobrir o comportamento espacial, em relação à distância da variável Z em estudo. Isso pode ser visualizado em forma de gráfico, que também recebe o nome de variograma (Yamamoto; Landim, 2015).

A principal característica desejável do variograma é representar, da forma mais fiel possível, a estrutura espacial da variável em estudo (Vieira, 2000), como ilustrado na Figura 1.

Figura 1 – Elementos essenciais a serem considerados em um variograma com patamar.



Fonte: Druck et. al., 2004, p. 80.

De acordo com Yamamoto e Landim (2013) e Vieira (2000), os parâmetros do variograma, demonstrados pela Figura 1 são explicados como:

a) Alcance (a): é a distância dentro da qual as amostras coletadas encontram-se correlacionadas espacialmente. Isto é, indica até que distância há correlação espacial entre os valores amostrados, o que restringiria as técnicas de estatística inferencial baseadas em independência. Pode ser definido também como a distância a partir da qual o variograma alcança seu patamar;

b) Efeito pepita (C_0): é o valor que o variograma atinge quando a distância tende a zero. Inicialmente, à medida que as distâncias diminuem, é esperado que o variograma diminua até atingir zero. Entretanto, na prática, isso nem sempre ocorre, indicando a presença do efeito pepita. O efeito pepita reflete um fenômeno que não é totalmente conhecido, por falta de informação, mas não necessariamente um fenômeno espacial aleatório;

c) Contribuição (C_1): é a diferença entre o patamar (C) e o efeito pepita (C_0). O fator representa a contribuição atribuída ao fenômeno espacial em análise;

d) Patamar ($C = C_0 + C_1$): é o valor do variograma correspondente a seu alcance (a) e coincide com a variância a priori do conjunto de dados analisados no estudo. Deste ponto em diante, considera-se que não existe mais

dependência entre as amostras. Em outras palavras, representa a contribuição atribuída ao fenômeno espacial em estudo que somada à contribuição aleatória expressa pelo efeito pepita, seria responsável por levar o variograma ao seu valor assintótico.

A partir da Figura 1, ajusta-se um variograma teórico (ou modelo teórico) ao variograma experimental que melhor descreva o desempenho dos dados no espaço. Segundo Bernadi et al. (2014) o ajuste de um modelo ao variograma é fundamental nas aplicações da geoestatística, pois os cálculos nessa área dependem do modelo do variograma para cada distância específica.

Espera-se de um variograma que quanto menor o $\boldsymbol{h}$, menores deverão ser as diferenças $[Z(x_i) - Z(x_i + \boldsymbol{h})]$, ou seja, as observações mais próximas devem ser mais parecidas do que as mais-- distantes, caso haja dependência espacial. Por definição, $\hat{\gamma}(\mathbf{0}) = 0$, mas na prática quando $\boldsymbol{h} \to \mathbf{0}$, $\hat{\gamma}(\mathbf{0}) = k$, chamado efeito pepita (C_0). O C_0 indica que distâncias menores do que o menor $\boldsymbol{h}$ deveriam ser praticadas na amostragem, ou ainda, houve falta de precisão. A partir de um certo $\boldsymbol{h}$, a semivariância estabiliza, chegando à máxima dissimilaridade (patamar) e é aproximadamente igual à variância dos dados. Quando o patamar é atingido, o respectivo $\boldsymbol{h}$ é chamado de alcance e indica a distância limite da dependência espacial (Yamamoto; Landim, 2013).

Apesar de haver diversos modelos teóricos disponíveis para ajustar aos dados de variogramas, três são comumente empregados na modelagem: esférico (Equação 2), exponencial (Equação 3) e gaussiano (Equação 4). Segundo Yamamoto e Landim (2013) as discrepâncias entre os modelos teóricos derivam de variações na concavidade, inclinação e comportamento assintótico.

$$\gamma(h) = \begin{cases} C_0 + C_1\left[\frac{3}{2}\left(\frac{h}{a}\right) - \frac{1}{2}\left(\frac{h}{a}\right)^3\right], & se\ 0 < h < a \\ C_0 + C_1, & se\ h \geq a \end{cases} \quad (2)$$

$$\gamma(h) = C_0 + C_1\left[1 - exp\left(-\left(\frac{h}{a}\right)\right)\right], \quad (3)$$

$$\gamma(h) = C_0 + C_1\left[1 - exp\left(-\left(\frac{h}{a}\right)^2\right)\right], \quad (4)$$

Como os três modelos podem ser escolhidos para uma mesma base de dados, critérios devem ser adotados para definir aqueles que apresentam melhor

qualidade e apresentam o melhor ajuste. A literatura científica cita os indicadores da validação cruzada como os mais relevantes na tarefa para avaliar a qualidade do ajuste do variograma (Vieira et al., 2010). Compara-se cada valor amostral com a estimativa obtida por krigagem com base no modelo de variograma ajustado ao retirar-se o ponto amostral em questão.

A partir da metodologia de validação cruzada, é possível aferir a qualidade do modelo geoestatístico com base em indicadores específicos. Neste contexto, Vieira (2000) estabelece que o modelo ideal deve atender a critérios como: a média dos erros, cujo valor desejado é próximo de zero; e a média dos erros padronizados, que deve apresentar um valor próximo de um. Além disso, o erro quadrático médio deve apresentar o menor valor possível relativo à amplitude dos dados, atestando a mínima dispersão e, consequentemente, o melhor ajuste do modelo.

2.3. Krigagem

Se uma variável exibe dependência espacial e os pontos os quais deseja-se realizar previsões estão dentro da área de influência de locais conhecidos, é possível obter estimativas bastante precisas de seu valor. Esse é o princípio subjacente do método de interpolação conhecido como krigagem, em homenagem a Daniel G. Krige, pioneiro no desenvolvimento dessa técnica (McKillup; Dyar, 2010).

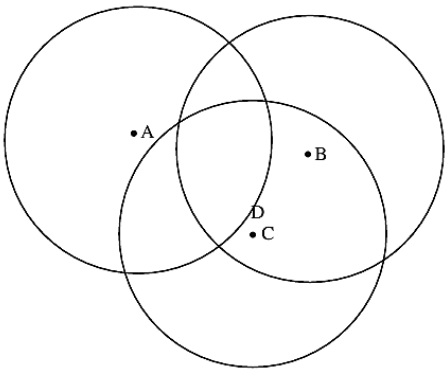
A krigagem é um método geoestatístico de interpolação, que vai além das simples médias ou inversos de distância. O processo envolve a resolução de um complexo conjunto de equações simultâneas para estimar o valor em um ponto desconhecido. Isso é alcançado buscando o melhor ajuste geral aos pontos vizinhos, ponderados pelo variograma teórico previamente ajustado aos dados (McKillup; Dyar, 2010).

Em termos práticos, a krigagem utiliza os valores amostrados em pontos específicos dentro da malha de estudo para realizar a interpolação. Um dos maiores diferenciais entre a krigagem e outros métodos de interpolação é que a mesma não possui viés e tem a finalidade de minimizar a variância dos erros de predição (Yamamoto e Landim, 2013). Também é conhecida como BLUP, do inglês, *Best Linear Unbiased Predictor*, ou seja, o melhor preditor linear não

viesado. Possui essa nomenclatura, pois suas predições são calculadas através da combinação linear das amostras obtidas e é não viesada porque a média dos erros de predição é nula. É o melhor porque dentre todos os preditores lineares não viesados é o que minimiza a variância dos erros de predição (Isaaks; Srivastava, 1989).

A Figura 2 apresenta o princípio do método de krigagem, em que o valor da variável não é conhecido para o ponto D, mas D está dentro das regiões de influência dos pontos A, B e C, faixa onde a variável é conhecida. Portanto, D pode ser predito por interpolação dos dados obtidos dos pontos A, B e C.

Figura 2 – O valor de uma variável é conhecido para os locais A, B e C e precisa ser estimado para o local D.



Fonte: McKillup; Dyar, 2010, p. 379.

2.3.1. Preditor

Considere que é desejado predizer valores para qualquer local, x_0, em que não se têm valores medidos já conhecidos e que a estimativa deva ser uma combinação linear dos valores medidos conforme a Equação 5.

$$\hat{Z}(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i), \quad (5)$$

em que $\hat{Z}(x_0)$ é o valor predito em x_0, λ_i são os pesos associados a cada valor previamente amostrado e $Z(x_i)$ são os valores amostrados $Z_1, Z_2, \cdots Z_n$ nas posições $1,2,\cdots,n$ respectivamente. O que torna a krigagem um interpolador ótimo é a forma como os pesos estão distribuídos (Vieira, 2000).

2.3.2. Tipos de métodos de krigagem

De acordo com Santos (2010) se destacam na literatura especializada a krigagem simples, a krigagem ordinária e a krigagem universal. O autor apresentou uma breve descrição abordando os pontos mais importantes de cada tipo de krigagem. A krigagem simples parte do pressuposto de que a função média do processo estocástico é conhecida e, portanto, não é estimada a partir das realizações. A krigagem ordinária, sendo a mais utilizada, baseia-se apenas no variograma, e a função média não é conhecida, sendo assim, é estimada a partir dos dados. A krigagem universal é aplicável quando a variável regionalizada não é estacionária. Nesse caso, a tendência é removida ajustando-se polinômios de baixo grau, e o restante do procedimento analítico consiste em uma análise dos resíduos.

Para a realização deste trabalho optou-se em utilizar a krigagem simples, por ser a mais básica e por assumir que a média é conhecida e constante em todo o espaço. Além disso, é útil quando a média do processo é bem definida e não varia significativamente em toda a área de estudo. Dentre as variantes dos preditores de krigagem pode-se destacar a simples, visto que de acordo com Santos et al. (2011) apesar do conhecimento da média ser um fator limitante em sua utilização, a krigagem simples apresentou-se como um preditor mais preciso do que a krigagem universal.

2.3.3. Krigagem simples

As krigagens simples são processos estacionários que consideram a média local como uma constante igual à média populacional. De modo geral, a krigagem pode ser compreendida como pertencente à família de algoritmos de regressão de mínimos quadrados generalizados, conforme descrito por Goovaerts (1997). Sua técnica de estimação fundamenta-se no estimador definido em (6).

$$\boldsymbol{Z}^*(u) - \boldsymbol{m}(u) = \sum_{\alpha=1}^{n(u)} \boldsymbol{\lambda_\alpha}(u)[\boldsymbol{Z}(u_\alpha) - \boldsymbol{m}(u_\alpha)] \quad (6)$$

em que $n(u)$ é o número de pares de pontos amostrais, $\lambda_\alpha(\boldsymbol{u})$ são os pesos definidos para os dados $z(\boldsymbol{u}_\alpha)$ quando considerados como realizações da variável

aleatória $Z(\boldsymbol{u}_\alpha)$. Os valores esperados para as variáveis aleatórias $Z(\boldsymbol{u}_\alpha)$ e $Z(\boldsymbol{u})$ são iguais a $m(\boldsymbol{u}_\alpha)$ e $m(\boldsymbol{u})$, respectivamente. Com essa abordagem, é viável calcular erros de estimação para a diferença entre as variáveis $Z^*(\boldsymbol{u}) - Z(\boldsymbol{u})$.

O processo de estimação da krigagem parte da diferença $Z^*(\boldsymbol{u}) - Z(\boldsymbol{u})$, no qual minimiza-se a estimativa da variância do erro de $\boldsymbol{u}$ representado em (7).

$$\boldsymbol{\sigma_E^2}(u) = \boldsymbol{Var}\{\boldsymbol{Z}^*(u) - \boldsymbol{Z}(u)\} \tag{7}$$

Observe que cada preditor de krigagem distinto implica em um modelo diferente. Uma abordagem comum, portanto, é adicionar um componente ao componente de tendência para representar $Z(\boldsymbol{u})$.

$$\boldsymbol{Z}(u) = \boldsymbol{R}(u) + \boldsymbol{\mu} \tag{8}$$

em que o componente $R(\boldsymbol{u})$ é modelado por meio de uma função que apresenta estacionaridade, média nula e covariância $C(\boldsymbol{h})$. Portanto,

$$\boldsymbol{E}[\boldsymbol{R}(u)] = \boldsymbol{0} \tag{9}$$

$$\boldsymbol{COV}\big(\boldsymbol{R}(u), \boldsymbol{R}(u+h)\big) = \boldsymbol{E}(\boldsymbol{R}(u).\boldsymbol{R}(u+h)) \tag{10}$$

Observa-se que a esperança de Z no local $\boldsymbol{u}$ corresponde ao valor médio de $\boldsymbol{u}$ analisado de forma local.

De acordo com Yamamoto e Landim (2013), para uma variável e um ponto de interesse, é possível obter uma estimava dessa variável na sua respectiva localização $\boldsymbol{u}$, conforme a Equação 11.

$$\boldsymbol{Z}^*_{\boldsymbol{x_0}} = \sum_{\boldsymbol{i=1}}^{\boldsymbol{n}} \boldsymbol{\mu} + \boldsymbol{\lambda_i Z_{x_i}} \tag{11}$$

em que μ é média de atributo, $Z^*_{x_0}$ é o valor estimado do atributo na localização x_0, λ_i é o peso da amostra na localização x_i, sendo i o indexador da amostra x_i de localização conhecida.

O preditor linear na krigagem simples, denotado por $Z^*_{SK}(\boldsymbol{u})$ é dado pela equação (12).

$$\boldsymbol{Z_{SK}^{*}}(u) = \sum_{\alpha=1}^{\boldsymbol{n(u)}} \boldsymbol{\lambda_{\alpha}^{SK}}(u)\boldsymbol{Z}(u_{\alpha}) + \left[\mathbf{1} - \sum_{\alpha=1}^{\boldsymbol{n(u)}} \boldsymbol{\lambda_{\alpha}^{SK}}(u)\right]\boldsymbol{m} \qquad (12)$$

em que $\lambda_{\alpha}^{SK}(\boldsymbol{u})$ são determinados de tal maneira que a variância do erro seja mínima $\sigma_E^2(\boldsymbol{u}) = Var\{Z_{SK}^{*}(\boldsymbol{u}) - Z(\boldsymbol{u})\}$ mediante a condição de não tendenciosidade. Já a variância de krigagem é dada pela equação (13).

$$\boldsymbol{\sigma_{KS}^{2}} = \boldsymbol{C}(\mathbf{0}) - \sum_{\alpha=1}^{\boldsymbol{n(u)}} \boldsymbol{\lambda_{\alpha}^{SK}}(u)\ \boldsymbol{C}(u - u_{\alpha}) \geq \mathbf{0} \qquad (13)$$

2.4. Validação cruzada

Segundo Vieira (2010), após realizar a modelagem da dependência espacial e a interpolação, é fundamental validar o modelo geoestatístico. A validação cruzada emerge como uma ferramenta indispensável para essa análise. Essa abordagem permite avaliar a incerteza associada à medição prévia dos dados e tem como meta predizer a precisão do modelo para amostras futuras. Nesse processo, os dados amostrais são retirados um a um e estimados pelo modelo utilizando os demais pontos amostrais. Consequentemente, torna-se viável comparar os valores amostrais com os estimados pelo modelo e obter uma medida estatística do erro de estimação (Isaaks e Srivastrava, 1989).

Entre as estatísticas de erro, Ferreira (2018) destaca a média do erro de estimação, média do erro de estimação padronizado, raiz quadrada do erro quadrático médio, média do desvio padrão e raiz quadrada da média do erro de estimação padronizado ao quadrado. No processo de autovalidação, é essencial que a média dos erros e dos erros padronizados sejam nulas, que a variância dos erros seja finita e que a variância dos erros padronizados seja unitária (Vieira, 2000).

2.5. Interpolação pelo inverso da distância (IDW)

A estimação ponderada pelo inverso da distância (IDW) é um dos métodos de interpolação não estocásticos mais utilizado para situações em que não há dependência espacial (PEREIRA, 2021). Um ponto não amostrado é estimado de forma determinística por meio de uma média ponderada, conforme a Equação 14.

$$\hat{z}_i = \frac{\sum_{j=1}^{n} w_{ij} z_j}{\sum_{j=1}^{n} w_{ij}} \quad (14)$$

em que $\hat{z}_i$ é a estimativa para um local não amostrado, w_{ij} são os pesos de estimação e z_j é o *j-ésimo* valor amostrado.

Os pesos são definidos como o inverso da distância euclidiana entre o *i-ésimo* ponto a ser estimado e o *j-ésimo* amostrado, elevada a uma potência *k* (Equação 15).

$$w_{ij} = \frac{1}{d_{ij}^{k}} \quad (15)$$

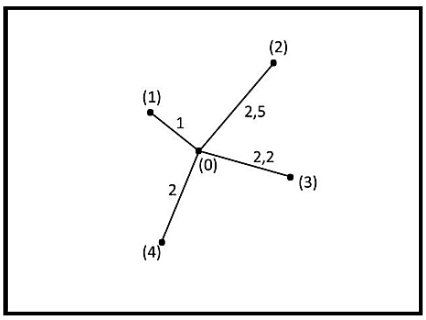
em que d_{ij} representa a distância euclidiana. Considere dois pontos $P = (x_i, y_i)$ e $Q = (x_j, y_j)$. A distância euclidiana entre os pontos P e Q pode ser calculada pela Equação 16.

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (16)$$

em que $x_i - x_j$ é a diferença entre as abcissas dos pontos P e Q e $y_i - y_j$ representa a diferença entre ordenadas dos pontos P e Q.

Silva (2022) apresentou um exemplo de aplicação da técnica IDW para quatro pontos amostrados e determinou o valor no ponto central, não amostrado e identificado por 0 (Figura 3). Para o exemplo proposto pelo autor, foi considerado o expoente $k = 2$ (estimação pelo inverso do quadrado da distância).

Figura 3 – Exemplo de aplicação da técnica de estimação por IDW, em que o ponto central (0) será estimado com base nos quatro pontos amostrados, em que 1, 2, 2,2 e 2,5 representam as distâncias dos pontos (1), (2), (3) e (4), respectivamente, em relação ao ponto (0).



Fonte: Silva, 2022, p. 26.

Silva (2022) verificou por meio do exemplo apresentado que o peso w_{ij} é inversamente proporcional à distância euclidiana e portanto, o ponto amostrado 1, separado por 1 unidade de comprimento do ponto 0, terá um efeito maior sobre a estimação do valor desse ponto em relação ao ponto 2 que está mais distante.

2.6. Algoritmo *k-means* para agrupamentos

Segundo James et al. (2013), o objetivo das técnicas de agrupamento é agrupar observações em um conjunto de dados, buscando dividi-las em grupos distintos, garantindo que as observações dentro de cada grupo sejam bastante semelhantes entre si, enquanto as observações em grupos diferentes, apresentem diferenças significativas entre elas. Para concretizar essa ideia, é essencial definir o que torna duas ou mais observações semelhantes ou diferentes. Essa definição geralmente depende do domínio específico e deve ser baseada no conhecimento detalhado dos dados em análise.

James et al. (2013) consideram que um dos algoritmos mais renomados e amplamente utilizados nessa prática é o *k-means* (KM). O método é classificado como não supervisionado, permitindo ao pesquisador definir o número k de grupos (*clusters*) desejados. Nesse método, os grupos não se sobrepõem, ou seja, cada ponto pertence a apenas um dos k grupos. Além disso, não há resíduo, garantindo que todos os pontos sejam agrupados em um dos *clusters* resultantes. As Equações 17 e 18 retratam essa característica no contexto matemático.

$$C_1 \cup C_2 \cup \cdots \cup C_k = \{1,2,\cdots n\} \tag{17}$$

$$C_k \cap C_{k'} = \emptyset \tag{18}$$

em que C_k representa o *k-ésimo* agrupamento, e $\{1,2,\cdots n\}$ o conjunto total de pontos.

O princípio fundamental na determinação do melhor agrupamento é a minimização da variação dentro de cada grupo, indicada por $W(C_k)$, que é calculada com base na soma dos quadrados das distâncias euclidianas entre os pares de pontos do grupo, dividida pelo número total de pontos no grupo (Equação 19). Desta forma, se cada grupo possuir o valor mínimo para variação

interna, a soma da variação de todos os grupos também deverá ser mínima (James et al., 2013).

$$W(C_k) = \frac{1}{|C_k|}\sum_{i,i' \in C_k}\sum_{j=1}^{p}\left(x_{ij} - x_{i'j}\right)^2 \quad (19)$$

em que *i* e *i'* são os pontos no interior do *k-ésimo* agrupamento e *p* o número de variáveis consideradas.

Após definir a variação dentro de cada grupo, busca-se minimizar a Expressão 20, alocando cada uma das *n* observações a um dos *k* grupos utilizando o algoritmo KM.

$$mínimo\left\{\sum_{k=1}^{K}\frac{1}{|C_k|}\sum_{i,i' \in C_k}\sum_{j=1}^{p}\left(x_{ij} - x_{i'j}\right)^2\right\} \quad (20)$$

O algoritmo KM consiste em dois passos elementares. O primeiro é classificar aleatoriamente em *k* grupos o conjunto das n observações. A partir desta etapa, deve-se calcular *k* centroides, dados pela média dos pontos pertencentes a cada um dos grupos. Nessa fase calcula-se a distância euclidiana entre os centroides e os pontos, reclassificando-os com base na proximidade com cada um dos centroides previamente obtidos. O cálculo dos centroides e a reclassificação são reproduzidos até o momento em que a variação dentro dos grupos para de mudar. Ou seja, a Expressão 20 atinge uma mínimo local. Portanto, James et al. (2013) recomendam a realização de várias iterações do método, analisando os valores de variação total obtidos em cada repetição. Ao final, deve-se selecionar o agrupamento que apresenta o menor valor de variação total entre todas as iterações realizadas.

2.7. Fertilidade do solo

Segundo Vieira (2000), a variedade de características encontradas na composição do solo resulta da interação entre os processos de formação, práticas de manejo e cultivo, destacando-se a importância de sua distribuição espacial. O autor considera que ao realizar declarações sobre as propriedades do solo, é crucial considerar a variação do atributo em questão, uma vez que

quanto mais heterogêneo o solo, menos confiáveis se tornam as afirmações. De acordo com Grego e Vieira (2005), para melhor compreender a dinâmica do solo, métodos geoestatísticos são utilizados para obter informações elementares para o manejo agrícola a partir da variabilidade espacial das suas características.

Segundo Kumar et al. (2020) a representação espacial das variáveis do solo é uma das opções viáveis para melhorar a fertilidade através do manejo de seus nutrientes específicos. O trabalho dos autores foi conduzido no bloco Jhandutta (distrito de Bilaspur), Himachal Pradesh na Índia. O objetivo foi avaliar a variabilidade espacial de alguns nutrientes selecionados do solo. A partir de métodos geoestatísticos, concluíram que a distribuição espacial dos nutrientes do solo é fortemente influenciada por fatores ambientais e naturais. O mapa de distribuição espacial obtido delineou as áreas que necessitavam de gestão de nutrientes específica para o local, para uma agricultura sustentável.

2.7.1. Nutrientes do solo para cana-de-açúcar

A cana-de-açúcar possui significativa importância econômica, além disso, é amplamente cultivada em várias regiões do país, consequentemente encontrando-se em solos com diferentes propriedades físico-químicas, muitas vezes distantes dos ideais. A cultura demonstra traços de rusticidade e capacidade de adaptação a uma variedade de solos, incluindo aqueles com propriedades físico-químicas mais desfavoráveis. No entanto, para alcançar rendimentos elevados, o produtor deve selecionar cuidadosamente o tipo de solo mais adequado às necessidades da cultura (EMBRAPA, 2022).

Para exemplificar, Arcoverde et al. (2023) realizaram um trabalho com o intuito de avaliar a variabilidade e a dependência espacial dos atributos físicos de Latossolos sob plantio direto e preparo reduzido no cultivo da cana-de-açúcar. Utilizando métodos geoestatísticos foi verificado dependência espacial em diferentes profundidades. Além disso, as correlações lineares evidenciaram a relação entre os atributos físicos do solo e as condições de manejo, e os padrões de variabilidade foram representados por meio de cartogramas de isolinhas.

Além das características físicas, variáveis como fósforo (P), potássio (K), cálcio (Ca), enxofre (S), nitrogênio (N) e magnésio (Mg) estão no grupo dos

macronutrientes (aqueles altamente consumidos pelas plantas) (Oliveira et al., 2005). Cobre (Cu), zinco (Zn), ferro (Fe), boro (B) manganês (Mn), cloro (Cl) e molibdênio (Mo) são micronutrientes (aqueles menos exigidos pelas plantas). Esses são os atributos importantes à planta em algum estágio do desenvolvimento (Oliveira et al., 2005).

Outros elementos presentes no solo, como o alumínio (Al), podem ser prejudiciais ao desenvolvimento das plantas (Brito, 2014). O autor ressalta que, apesar de sua toxicidade e dos problemas que causa na produção agrícola, ainda há muito a ser discutido sobre o verdadeiro impacto do alumínio tóxico do solo na cana-de-açúcar. Na forma em que é encontrada no solo, o hidróxido de alumínio (um átomo de Al cercado por seis hidroxilas) é altamente reativo e, ao hidrolisar durante o processo de calagem do solo, libera íons de hidrogênio, tornando o pH do solo mais ácido. Assim, o potencial do alumínio em acidificar o solo durante o processo de hidrólise é denominado acidez potencial e é definido pela soma de H + Al.

O pH exerce uma influência significativa no desenvolvimento das plantas, pois está diretamente relacionado à concentração, solubilidade, disponibilidade e capacidade de retenção de nutrientes essenciais. Além disso, serve como um indicador de possíveis problemas de fertilidade do solo e até mesmo de toxicidade, devido à sua associação com o desenvolvimento das culturas (Lima et al., 2016).

Segundo Lima et al. (2016), quimicamente, o potencial hidrogeniônico (pH) está intrinsicamente ligado à atividade iônica e a todas as reações que ocorrem no solo. Já a soma de bases trocáveis (SB) expressa a quantidade total de cátions permutáveis, enquanto a capacidade de troca catiônica (CTC) traduz a habilidade do solo em reter cátions. Ronquim (2010) destaca que a capacidade de troca iônica mede a liberação dos nutrientes presentes no solo, mantendo-o fértil e controlando a toxicidade resultante do uso de fertilizantes. Quando a capacidade de troca catiônica está relacionada aos cátions de cálcio, magnésio e potássio, o solo favorece a planta nutricionalmente; no entanto, cátions tóxicos de hidrogênio e alumínio empobrecem o solo. O índice de saturação por bases (V) indica a fertilidade do solo, enquanto o índice de saturação de alumínio (M) revela a proporção de alumínio solúvel em relação aos teores de alumínio e de bases trocáveis.

Conforme destacado por Vitti et al. (2005), a deficiência de nutrientes pode ser prontamente identificada nas plantas de cana-de-açúcar. Sintomas de carência de fósforo, cálcio, potássio e magnésio podem ser observados nas folhas, manifestando-se através de colorações e manchas específicas, podendo evoluir para necrose. A falta de potássio resulta em redução do teor de açúcar na cana, enquanto a deficiência de fósforo compromete o crescimento e o número de perfilhos (Epstein e Bloom, 2006). O fósforo desempenha um papel crucial no armazenamento e na transferência de energia, atuando como componente dos ácidos nucleicos, fitina e fosfolipídios, o que o torna fundamental para o metabolismo da planta (Epstein e Bloom, 2006). Além disso, ele melhora a eficiência no uso da água pela planta, facilita a utilização e assimilação de outros nutrientes, conferindo à planta maior resistência a doenças e à escassez de água (Penatti, 2013).

Com base nos elementos e nas propriedades do solo essenciais ao desenvolvimento da cana-de-açúcar e na influência que as condições do solo exercem sobre o rendimento da cultura, Cambardella et al. (1994) destacam a importância de compreender a variabilidade espacial das características do solo para aprimorar a qualidade das práticas de manejo e avaliar os efeitos da agricultura sobre o meio ambiente. Essa compreensão é vital, visto que, conforme demonstram Mangabeira et al. (2005), solos com deficiência de fertilidade nutricional estão frequentemente relacionados a áreas de baixa produtividade.

2.8. Amostragem otimizada: Conditioned Latin Hypercube Sampling ($cLHS$)

A complexidade e a elevada variabilidade espacial dos atributos do solo, destacadas anteriormente (Cambardella et al., 1994), tornam imperativo o uso de métodos eficientes para o seu mapeamento e manejo preciso. A avaliação detalhada desses atributos em grandes áreas requer a coleta de um número significativo de amostras, o que implica um custo elevado e um grande esforço logístico.

Diante da necessidade de obter informações representativas da área de estudo, mas de forma economicamente viável, é fundamental empregar técnicas

de amostragem otimizadas. Tais métodos visam selecionar um número reduzido de pontos de coleta que consigam capturar o máximo da variabilidade presente na área de interesse. O método *Conditioned Latin Hypercube Sampling* (cLHS) surge como uma alternativa para atender essa necessidade no planejamento amostral.

A segmentação do conjunto de dados pode ser realizada através de diversos métodos de amostragem. O método *Latin Hypercube Sampling* (LHS), desenvolvido por McKay, Beckman e Conover (1979), possibilita a obtenção de amostras de variáveis a partir de suas distribuições multivariadas, conforme destacado por Minasny e McBratney (2006). O LHS se baseia no conceito de quadrado latino, no qual apenas uma amostra é alocada em cada linha e coluna, estendendo essa lógica para múltiplas dimensões.

Nesse método, considerando um conjunto de *k* variáveis $(X_1, \cdots X_k)$ o intervalo de variação de cada variável é dividido em *n* estratos de igual probabilidade. O principal objetivo do LHS é assegurar que cada variável seja devidamente representada em seus respectivos estratos, promovendo uma amostragem eficiente e representativa da distribuição dos dados analisados.

Um dos desafios do método está na possibilidade de que as amostras geradas não existam no ambiente real de estudo. Visando contornar essa limitação Minasny e McBratney (2006) propuseram um método condicionado, que consiste em uma estratégia de amostragem adequada para situações em que se dispõe de informações prévias sobre a área investigada.

Além disso, os autores ressaltam que essa metodologia pode ser aplicada tanto em variáveis contínuas quanto categóricas, permitindo uma representação satisfatória do solo e de sua variabilidade, o que aumenta as chances de reproduzir adequadamente os relacionamentos presentes, quando existentes.

3. MATERIAL E MÉTODOS

3.1. Coleta das amostras de solo

O conjunto de dados foi obtido a partir da coleta de 60 pontos amostrais de solos provenientes da quadra sete da Unidade de Ensino, Pesquisa e Extensão em Melhoramento Genético de Cana-de-Açucar (UEPE), localizado em Oratórios – MG. A quadra utilizada possui 2,13 ha. Para coletar as amostras foi proposto um grid quadrado 25 × 25m com um total de 33 pontos centrados. As demais amostras foram aleatorizadas por meio de sorteio no interior do polígono utilizando o software QGIS (Figura 4).

A área de estudo apresentou como classe pedológica predominante um Latossolo Vermelho-Amarelo (LVA), caracterizado por elevada profundidade, boa drenagem e estrutura granular, condições que favorecem o desenvolvimento radicular, mas demandam manejo específico para manter a fertilidade. O sistema de manejo adotado incluiu correção e enriquecimento do solo, destacando-se a aplicação de agrosilício na dose de 3 t/ha.

Durante o plantio foi realizado adubação de base com a formulação 06-30-24 na quantidade de 500 kg/ha, oferecendo fósforo e potássio em níveis adequados para o estabelecimento inicial da cultura. No manejo químico, utilizou-se herbicida pré-emergente, especificamente, Combine® na dose de 2 L/ha. Para a adubação de soqueira, foi empregada a formulação 20-05-20, também na dose de 500 kg/ha, assegurando reposição de nutrientes após o corte e manutenção da produtividade ao longo do ciclo da cultura. Essas práticas refletem um manejo integrado que busca equilibrar fertilidade, controle de plantas daninhas e sustentabilidade produtiva na área.

Figura 4 **–** Área experimental em Oratórios (MG), identificada pela linha vermelha. Os pontos amarelos indicam os pontos sistemáticos (33 pontos centrados em grade regular de 25m × 25m). Os pontos azuis indicam os pontos aleatórios.



Fonte: Elaborado pelo autor.

A coleta das amostras de solo ocorreu no mês de fevereiro de 2023. Em cada local indicado na Figura 4, uma amostra de solo foi obtida na camada de 0-20 cm de profundidade (Figura 5). Essas amostras foram devidamente armazenadas e encaminhadas a um laboratório de análises de solos no município de Viçosa - MG, onde foram realizadas as análises.

Figura 5 – Homogeneização das amostras na área experimental em Oratórios (MG) após a retirada do solo na cama de 0-20 cm de profundidade.



Fonte: Preparado pelo autor.

A Tabela 1 indica as variáveis químicas de solo, com as respectivas unidades de medida, que foram consideradas no estudo.

Tabela 1 – Variáveis de solo analisadas, suas siglas e unidades/escala de medida.

Variável	Sigla	Unidade/escala de medida
Boro	B	mg/dm³
Cálcio	Ca	$cmol_C/dm^3$
Cobre	Cu	mg/dm³
Ferro	Fe	mg/dm³
Acidez potencial	H+Al	$cmol_C/dm^3$
Potássio	K	mg/dm³
Magnésio	Mg	$cmol_C/dm^3$
Manganês	Mn	mg/dm³
Matéria orgânica	MO	dag/kg
Fósforo	P	mg/dm³
Fósforo remanescente	P-rem	mg/L
Potencial hidrogeniônico	pH	0-14
Índice de saturação de bases	V	%
Zinco	Zn	mg/dm³

3.2. Análise exploratória dos dados originais

Com o intuito de capturar os aspectos fundamentais das variáveis em análise, foram computadas algumas medidas de estatística descritiva, pois as medidas de posição e dispersão desempenham um papel crucial na compreensão da distribuição das variáveis em questão (Neto, 2002). Sendo assim, foram determinadas a média aritmética, a mediana, o desvio padrão, os valores máximo e mínimo, o coeficiente de variação, a assimetria e a curtose.

Para melhor visualizar os resultados obtidos com a estatística descritiva, foram construídos histogramas e boxplots para cada uma das variáveis analisadas. O boxplot foi empregado para identificar e remover *outliers*, seguindo a sugestão de Devore (2006), uma vez que, de acordo com Isaaks e Strivastava (1989), *outliers* têm um impacto significativo no variograma e na interpolação geoestatística.

A partir dos resultados foi feita a comparação das estatísticas descritivas de cada variável com as respectivas faixas de interpretação da disponibilidade dos micronutrientes com base em Raij et al. (1996). O objetivo foi comparar as estatísticas descritivas, obtidas a partir dos resultados das análises laboratoriais

das variáveis de fertilidade do solo, com o que é recomendado pela literatura. A partir dessa comparação foi possível apontar quais variáveis da Tabela 1 seriam consideradas para a realização da etapa descrita na seção 3.3.

Dessa forma, apenas as variáveis que apresentaram teores fora das faixas recomendadas por Raij et al. (1996) foram consideradas para a análise de variabilidade do solo. Isso se justifica pelo impacto direto que a deficiência ou excesso desses nutrientes pode ter sobre o desempenho da cultura da cana-de-açúcar, comprometendo seu desenvolvimento e produtividade. Portanto, a análise proposta neste estudo concentrou-se apenas nas variáveis que demandaram atenção em relação à fertilidade do solo.

3.3. Análise geoestatística dos dados originais

Esta etapa consistiu em aplicar métodos geoestatísticos a fim de identificar e modelar a dependência espacial das variáveis do estudo selecionadas na etapa 3.2. Nesta etapa foi utilizado o pacote base do software R (R Development Core Team, 2024) e o pacote de contribuição geoR versão 1.8.1 (Ribeiro Júnior et al., 2020).

As variáveis foram analisadas individualmente. Foi observado se havia correlação espacial por meio do variograma experimental *omnidirecional* utilizando a função variog() do geoR. Foi possível identificar se as variáveis demonstraram ou não dependência espacial, ou seja, quando o variograma revela uma estrutura de efeito pepita puro, conclui-se que não há dependência espacial (Yamamoto; Landim, 2015).

É importante dizer que para a variável que não apresentou dependência espacial, foi utilizado o método IDW (Monteiro et al., 2004), por meio da função idw(). Por outro lado, uma vez confirmada a dependência espacial, foram ajustados modelos teóricos conhecidos e descritos na literatura. Neste trabalho foram considerados os modelos gaussiano, exponencial e esférico, empregando-se a função variofit() disponível no pacote geoR. Para o ajuste dos modelos foram avaliados três métodos de estimação: mínimos quadrados ordinários (OLS), máxima verossimilhança (ML), máxima verossimilhança restrita (REML) e dos mínimos quadrados ponderados (WLS) (Cressie, 1985).

O método de validação cruzada foi empregado para avaliar a qualidade do ajuste por meio da função xvalid() do pacote geoR. De acordo com Isaaks e Srivastava (1989), a seleção do modelo ajustado deve ser baseada nos resultados mais satisfatórios obtidos na validação. Dessa forma, os seguintes aspectos da validação cruzada foram observados e utilizados para avaliar a qualidade do ajuste (Guimarães, 2013; Hernández, 2021):

- Coeficiente angular da regressão entre valores estimados e observados deve ser estaticamente igual a um e o intercepto de regressão dever ser nulo teoricamente;
- Média do erro da estimação, teoricamente deve ser nula, ou seja, deve ser estatisticamente igual a zero;
- Média do erro padronizado teoricamente deve ser nula, ou seja, deve ser estatisticamente igual a zero;
- Variância do erro da estimação, teoricamente deve ser finita;
- Variância do erro de estimação padronizado, teoricamente deve ser unitária, ou seja, deve ser estatisticamente igual a um.

Após o ajuste do modelo, foi calculado o Índice de Dependência Espacial (IDE) proposto por Cambardella et al. (1994) a fim de determinar o grau da intensidade da dependência espacial. O IDE é calculado como expresso na Equação 21.

$$IDE = \frac{C_1}{C_1 + C_0} \tag{21}$$

em que C_1 é a contribuição e C_0 o efeito pepita, parâmetros estes obtidos no ajuste do modelo ao semivariograma experimental. Classifica-se a correlação espacial fraca se o IDE < 0,25; moderada se o IDE for ≥ 0,25 e < 0,75; e forte, se IDE ≥ 0,75 (Cambardella et al.,1994).

3.4. Interpolação e agrupamento

A partir dos dados das características químicas do solo e do modelo de variograma ajustado, iniciou-se à fase de interpolação, realizada por meio do método de krigagem. Cada ponto da grade apresentada na Figura 4 foi

associado a um par ordenado de coordenadas de longitude e latitude, que serviu de base para a posterior interpolação por krigagem.

Para realizar a interpolação utilizando o método de krigagem, foi empregada a função krig.conv() disponível no pacote geoR. Esta função requer como argumentos a variável original georreferenciada (sendo analisada uma variável por vez), uma grade que especifica os locais onde a interpolação será realizada e, por fim, o modelo teórico do variograma previamente ajustado para a variável em questão. O tipo de krigagem escolhida foi a simples (Santos et al., 2011). Durante a interpolação, a função krig.conv( ) usou todos os pontos vizinhos nos cálculos das estimativas, não sendo utilizada uma borda ou então um número máximo de vizinhos para delimitar a krigagem.

Por fim, foi realizado o agrupamento das regiões utilizando o método *k-means*, por meio da função kmeans(), identificando as áreas mais e menos favoráveis com base na concentração das variáveis de interesse disponíveis. Essas informações servem como referência para futuros trabalhos, permitindo a distribuição mais correta das plantações de acordo com o objetivo pretendido do pesquisador ou produtor.

Antes de realizar o agrupamento, os dados das variáveis selecionadas foram padronizados utilizando a função scale() do software R. Essa função realiza o escore Z (*Z-score*), transformando cada variável de modo que ela passe a ter média zero e desvio padrão unitário. Segundo James et al. (2013) a padronização dos dados é crucial devido ao efeito de escala, que ocorre quando variáveis possuem unidades de medida ou amplitudes muito diferentes. Portanto, a padronização teve como finalidade de garantir a formação dos grupos, pelo agrupamento, de forma equilibrada, e que os resultados das análises fossem interpretáveis.

3.5. Avaliação dos valores reduzidos de pontos de amostragem

Após realizar os passos mencionados nas seções 3.1 até 3.4, considerando todos os pontos amostrados, foi proposto o estudo considerando um número reduzidos de pontos de amostragem, com o intuito de reduzir custos operacionais tanto na fase de coleta, como na etapa de análise das amostras de solo.

O plano de ação adotado para este estudo foi a redução do número de amostras, em incrementos de 10% do total, até chegar no valor mínimo de pontos em que não ocorria efeito pepita puro no ajuste do semivariograma (Isaaks; Srivastava, 1989) (Tabela 2). O estudo foi realizado para as variáveis selecionadas conforme os critérios estabelecidos na seção 3.1. A seleção dos pontos removidos foi realizada de forma aleatória, a partir do banco de dados original, utilizando o método *Conditioned Latin Hypercube Sampling* apresentado na seção 2.8.

Após a obtenção do novo conjunto de dados reduzido, foram seguidos os passos das seções 3.2 e 3.3 com o intuito de concluir se reduzindo o número de pontos coletados afetaria as conclusões dos resultados obtidos.

Tabela 2 – Tabela com as informações do número de pontos reduzidos.

Percentual de redução (%)	Número de pontos para análise
10	54
20	48

3.6. Avaliação da distribuição sistemática e aleatória dos pontos de amostragem

De acordo com Yamamoto e Landim (2013, p. 21), em análises geoestatísticas, a amostragem sistemática fornece os melhores resultados. Para verificar esta afirmação, foi proposto o estudo comparativo da distribuição sistemática versus a aleatória dos pontos de amostragem obtidos neste trabalho, indicando qual a melhor estratégia de análise. Para isso, foram aplicados os passos das seções 3.2 e 3.3 para a grade regular (pontos em amarelo da Figura 4) e para os pontos aleatórios (pontos em azul da Figura 4), das variáveis pré-selecionadas de acordo com os critérios da seção 3.1.

3.7. Descrição da abordagem analítica para a correlação entre dados agroindustriais da cana-de-açúcar e a fertilidade do solo

A última etapa do trabalho foi realizar uma análise espacial das variáveis brix e toneladas de colmos por hectare (TCH) com o objetivo de investigar padrões de comportamento espacial dessas características na área

experimental utilizada para o cultivo da cana-de-açúcar. Essa avaliação considerou dados provenientes de um experimento de seleção de genótipos de cana-de-açúcar conduzido na área. Contudo, para garantir que as variáveis analisadas refletissem as condições ambientais predominantes, foram consideradas exclusivamente as parcelas testemunhas desse experimento, que representam a ausência de tratamento genético específico.

A área experimental objeto de estudo dessa etapa é de 0,23 ha. O experimento foi alocado em uma grade regular retangular. Foram consideradas duas testemunhas distribuídas no experimento, mas com cada testemunhas aparecendo em sete parcelas dentro de cada bloco.

A variável brix refere-se à porcentagem de sólidos solúveis presentes no caldo da cana-de-açúcar, sendo um indicador direto da concentração de sacarose (Varella et al., 2012), principal componente de interesse industrial para a produção de açúcar e etanol. Já a TCH representa a produtividade agrícola da cana, expressa em toneladas por hectare, e está diretamente relacionada ao volume total de biomassa colhida, sendo uma métrica crucial para avaliação de rendimento agronômico (Fernandes et al., 2003).

Inicialmente, foram coletados os dados de brix e TCH ao final do ciclo da safra, abrangendo as parcelas com plantas testemunhas. A partir desses dados, foram calculados os resíduos experimentais com base no Delineamento Inteiramente Casualizado (DIC) e o Delineamento em Blocos Casualizados (DBC). Os resíduos foram utilizados como base para análises subsequentes.

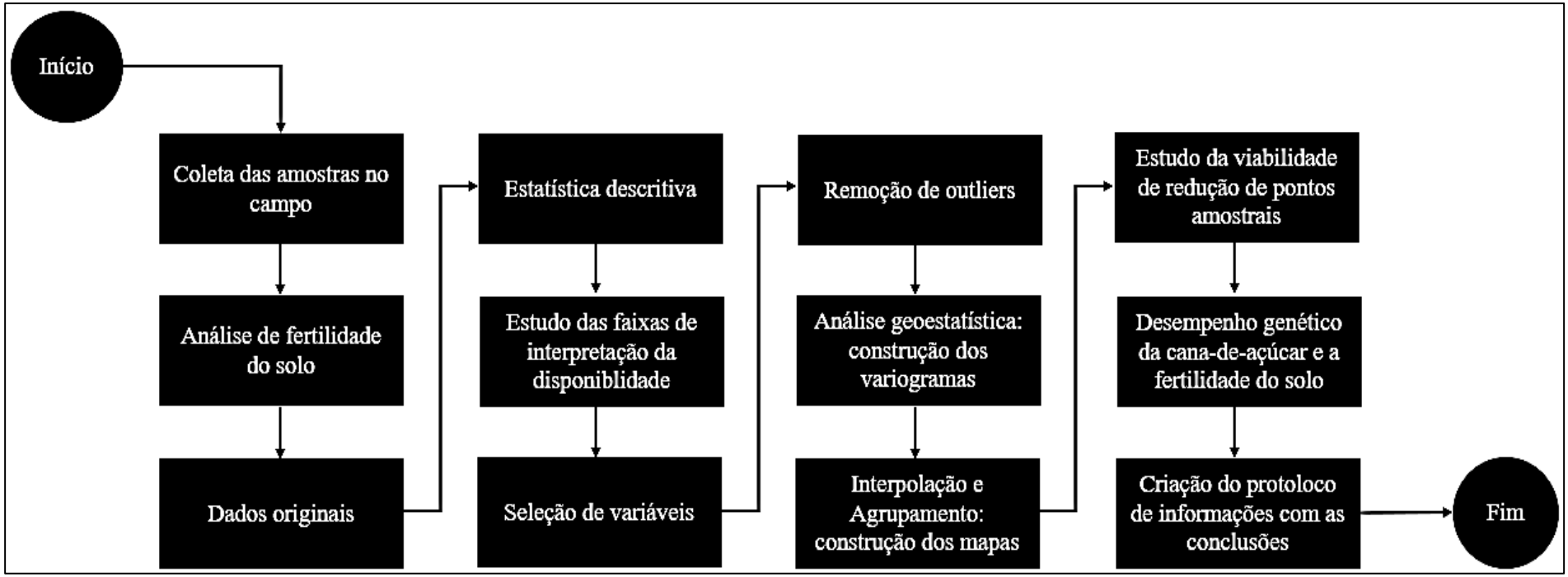
Para estudar o comportamento espacial dessas variáveis na área de plantio, foi empregada a técnica de Interpolação Ponderada pelo Inverso da Distância (IDW), por não terem apresentado dependência espacial.

Dessa forma, foi possível compreender como TCH e Brix se distribuíram espacialmente na área experimental, utilizando os resultados dos resíduos como indicador de variação não controlada pelo experimento, possivelmente influenciada por fatores ambientais, como as condições do solo.

3.8. Resumo da metodologia

O fluxograma apresentado na Figura 6 ilustra de forma resumida todas as etapas metodológicas deste trabalho que foram descritas na seção 3.1 até a seção 3.7.

Figura 6 – Fluxograma representativo da metodologia e dos processos envolvidos nas análises empregadas neste trabalho.



Fonte: Elaborado pelo autor.

4. RESULTADOS E DISCUSSÃO

4.1. Análise de fertilidade do solo e estatística descritiva

Os dados amostrais provenientes do resultado da análise de fertilidade do solo, para os 60 pontos amostrados, das variáveis da Tabela 1, estão dispostos no Apêndice A. A Tabela 3 inclui estimativas para média aritmética, valor mínimo, valor máximo, desvio-padrão, coeficiente de variação, curtose e assimetria de todas as variáveis.

De acordo com a Tabela 3 e os níveis adequados das concentrações de fertilizantes recomendadas por Raij et al. (1996), os teores de micronutrientes em solos (Cu, Fe, Mn e Zn) estão dentro da faixa de classificação como adequados (Cu > 0,8; Fe > 30,0; Mn > 5,0 e Zn > 1,0). Já a variável B está dentro da faixa classificado como médio (0,1 a 0,3). Não foi possível verificar a saturação por alumínio, pois o H não foi separado do Al. É possível avaliar que os valores de potássio e fósforo estão desfavoráveis para algumas parcelas/área, sendo o valor médio para a variável K igual a 0,08, classificado pelos autores como baixo e para PR (fósforo relativo) igual a 42,55, considerado como muito baixo.

De acordo com Warrick e Nielsen (1980), quanto ao coeficiente de variação, em relação aos cátions Ca^{+2} e Mg^{+2}, MO e V, todos apresentaram valores considerados médios (12% < CV < 60%). Ou seja, essas propriedades do solo não são uniformes na área de estudo. Este nível de variabilidade é considerado comum para atributos químicos em áreas agrícolas sob manejo. De acordo Ageitec (2020), o valor recomendado de pH para a cultura de cana-de-açúcar é de 6,5. O pH médio obtido na área em estudo foi de 6,2. O CV do pH é menor que 12%. Outro ponto que merece destaque é que o solo objeto de estudo deste trabalho é caracterizado como fértil, pois possui V médio igual a 57,74. Solos são caracterizados como distróficos ou inférteis apenas quando V < 50 (Prezotti e Guarçoni, 2013).

As variáveis K e P foram selecionadas para realização da análise geoestatística. Segundo Marin (2020) a cana-de-açúcar é capaz de se desenvolver em solos com condições físicas desfavoráveis, porém somente

atinge altas produtividades em solos férteis. O acúmulo de nutrientes pela cana-de-açúcar é bastante elevado, sendo o K o elemento mais absorvido pela cultura, e o P o quinto elemento mais absorvido, considerando a seguinte comparação proposta por Salviano et al. (2017): K > N > Ca > S > P > Mg.

Tabela 3 - Análise descritiva para as variáveis químicas de solo obtidas a partir dos 60 pontos amostrais na área de estudo.

	Média aritmética	Valor Mínimo	Valor Máximo	Desvio- padrão	Coeficiente de Variação	Curtose	Assimetria
B (mg/dm³)	0,13	0,10	0,20	0,04	34,93	2,33	1,15
Ca^{2+} ($cmol_C/dm^3$)	2,41	1,55	3,63	0,52	21,47	2,42	0,40
Cu (mg/dm³)	1,66	1,01	2,03	0,30	17,81	2,04	0,18
Fe (mg/dm³)	67,70	39,80	117,40	16,08	24,82	3,36	0,69
H+Al ($cmol_C/dm^3$)	2,41	0,66	5,28	0,99	41,16	2,06	0,50
K ($cmol/dm^3$)	0,08	0,03	0,16	0,03	37,28	2,33	0,37
Mg^{2+} ($cmol_C/dm^3$)	0,65	0,42	0,92	0,12	19,31	2,23	0,11
Mn (mg/dm³)	64,23	39,40	96,50	13,67	21,28	2,58	0,39
MO (dag/kg)	1,97	1,01	3,01	0,53	26,81	2,22	0,40
P (mg/dm³)	7,20	3,02	24,04	3,74	51,94	15,71	3,29
P-rem (mg/L)	29,27	18,06	37,30	5,08	17,34	2,12	-0,38
PR (%)	42,55	22,90	71,10	10,65	25,03	2,87	0,51
pH (H_2O)	6,20	5,03	7,03	0,39	6,29	3,01	0,28
V (%)	57,74	33,6	85,30	10,01	17,49	3,06	0,23
Zn (mg/dm³)	2,07	1,00	4,10	0,66	31,98	3,27	0,63

B: Boro; Ca_{+2}: Cálcio na forma de cátion; Cu: Cobre; Fe: Ferro; H+Al: Acidez potencial; K: Potássio; Mg^{+2}: Magnésio na forma de cátion; Mn: Manganês; MO: Matéria orgânica; P: Fósforo; P-rem: Fósforo remanescente; PR: fósforo relativo; pH: Potencial hidrogeniônico; V: Índice de saturação de bases; Zn: Zinco

O potássio é um elemento essencial encontrado em todas as células vivas. No solo, ele se apresenta na forma de cátion (K^+) e é absorvido pelas plantas nessa mesma forma. Uma das funções vitais do potássio é regular o fluxo de água dentro das células vegetais através das membranas celulares, além de desempenhar um papel fundamental na regulação de diversos processos químicos e enzimáticos (Reetz, 2017).

Korndorfer et al. (1998) afirmam que o fósforo possui a capacidade de melhorar a eficiência da planta na utilização da água, bem como na absorção e aproveitamento de outros nutrientes, seja proveniente do solo ou de adubos. Essa propriedade contribui para fortalecer a resistência da planta contra diversas doenças, além de aumentar sua capacidade de suportar baixas temperaturas e períodos de escassez de umidade. Embora o fósforo seja absorvido em pequenas quantidades pela cana-de-açúcar, se comparado ao potássio, exerce função-chave no metabolismo da cultura.

Portanto, optou-se por analisar somente K e P nas análise seguintes.

4.2. Histogramas e boxplot

Com o objetivo de melhor observar a distribuição das variáveis selecionadas para esse estudo, foi elaborado um histograma e um boxplot para fósforo e potássio do solo avaliado (Apêndice B). Além de destacar aspectos relevantes das distribuições de frequências, o boxplot foi empregado na detecção de outliers. No boxplot (Apêndice B), foram identificados dois outliers para fósforo e um para potássio. Embora não tenham sido identificados erros de digitação, há a possibilidade de problemas na coleta ou análise laboratorial, por isso, optou-se por excluir esses valores das análises geoestatísticas. Essa decisão foi tomada com o intuito de evitar distorções na modelagem do semivariograma e, consequentemente, nos mapas de interpolação, visando garantir maior robustez e representatividade dos resultados (Isaaks e Strivastava,1989).

4.3. Interpolação espacial

4.3.1. Análise geoestatística

A análise geoestatística das amostras do solo, por meio da modelagem de semivariogramas, identificou dependência espacial apenas para a variável potássio, visto que o semivariograma se apresentou crescente com a distância, chegando a um patamar a partir do valor do alcance (Apêndice C). Para a variável fósforo foi proposto o estudo por meio do interpolador ponderado pelo inverso da distância (IDW), que será abordado na seção 4.3.2, por não ter apresentado dependência espacial (Apêndice C). Cambardella et al. (1994) destacam que a alta variabilidade em pequenas distâncias (e o consequente baixo índice de dependência espacial) em nutrientes como o P está frequentemente associada a variações aleatórias e/ou às práticas de manejo da área, o que justifica o P não ter apresentado dependência espacial.

O modelo esférico apresentou melhor ajuste para a variável potássio. Os parâmetros da validação cruzada, que serviram de base para a seleção do modelo teórico estão apresentados na Tabela 4.

Tabela 4 - Parâmetros estimados para os melhores valores da validação cruzada.

Variável	$\hat{\mu}$	$\hat{\mu}_Z$	$\hat{S}_Z^2$	$\hat{\beta}_1$	Modelo
K	<0,001	0,0020	1,029	0,873	Esférico
K	0,022	0,0055	1,203	0,750	Exponencial
K	0,018	0,0080	1,115	0,720	Gaussiano

K: Potássio; $\hat{\mu}$: média do erro de estimação; $\hat{\mu}_Z$: média do erro de estimação padronizado; $\hat{S}_Z^2$: variância do erro de estimação padronizado; $\hat{\beta}_1$: inclinação da reta de regressão linear.

Observa-se pela Tabela 4 que os valores dos parâmetros da validação-cruzada foram adequados. A média do erro de estimação é muito próxima de zero. Isso implica que não ocorreram valores estimados muito acima ou muito abaixo dos valores amostrados. A média do erro padronizado e a variância do erro padronizado ficaram próximos de 0 e 1, respectivamente. A inclinação da reta de regressão linear entre os valores amostrados e estimados foi próxima de 1. O teste de hipótese para o

coeficiente de regressão linear (H_0: $\beta_1 = 1$), realizado por meio da função linearHypothesis() no R, apresentou p-valor = 0,559, indicando que a inclinação estimada não difere estatisticamente de 1. Esse resultado, aliado ao valor médio do erro de estimação próxima a zero, indica que os valores preditos de K são mais próximos aos observados e indica que o modelo esférico é adequado.

Apresenta-se na Tabela 5 um resumo contendo os parâmetros do ajuste do modelo teórico selecionado.

Tabela 5 - Parâmetros geoestatísticos estimados e respectivos métodos de estimação.

Variável	C_0	C_1	Patamar	Alcance	Modelo	Método
K	0,00045	0,00053	0,00098	95	Esférico	WLS

K: potássio; C_0: efeito pepita; C_1: contribuição, WLS: Mínimos quadrados ponderados.

De acordo com o exposto na Tabela 5, a variável K apresentou efeito pepita igual a 0,00045. O alcance estimado para a dependência espacial foi de 95 metros. Este valor equivale a 38% da maior dimensão (250 m) da área experimental. Em termos práticos, a dependência espacial do potássio limita a precisão da interpolação em regiões mais afastadas entre si. Isso significa que, ao estimar o teor de K em um local específico utilizando a krigagem simples, a estimativa será precisa e robusta se o ponto de interesse estiver a menos de 95 metros de uma amostra coletada. Analisando o índice de dependência espacial (IDE), nota-se que o valor de IDE = 0,54 para a variável K resulta em uma dependência espacial moderada.

4.3.2. Mapas obtidos via krigagem simples e IDW

Após a caracterização da dependência espacial do potássio, por meio do modelo esférico (melhor ajuste dos modelos de variograma), a interpolação foi realizada usando krigagem simples (Figura 6-A). Para a variável fósforo, em que não foi identificada dependência espacial, conforme apresentado na seção 4.1, foi utilizado o interpolador ponderado pelo inverso da distância (IDW) (Figura 6-B).

A análise visual dos mapas de interpolação da Figura 6 demonstra claramente a variabilidade espacial dos atributos do solo na quadra experimental. A Figura 6-A, referente ao Potássio, interpolado por Krigagem, exibe grandes manchas contínuas, revelando um padrão estruturado de variabilidade, com as maiores concentrações

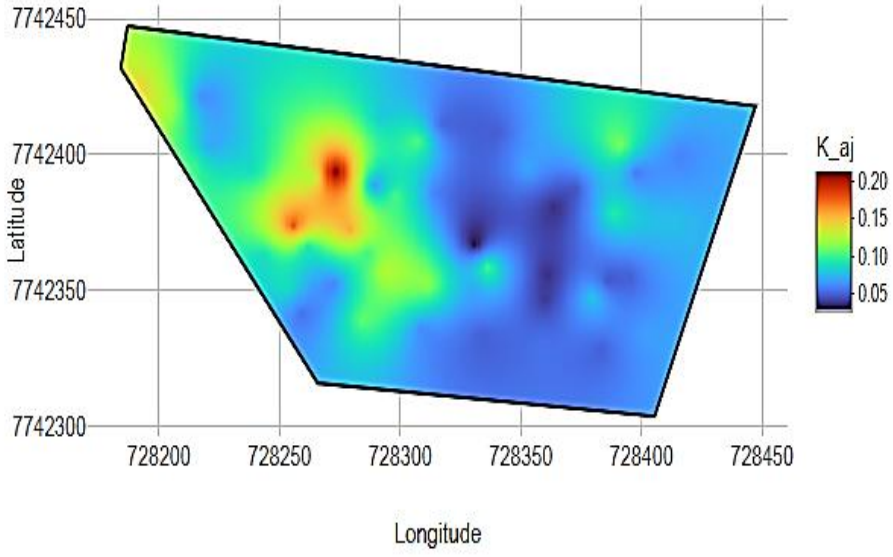
(manchas vermelhas) localizadas predominantemente na região à esquerda da Figura, e as menores (manchas azuis escuras) dominando à direita. Por outro lado, a Figura 6-B, que representa o Fósforo por IDW, apresenta uma distribuição mais dispersa, sugerindo uma heterogeneidade em curtas distâncias, característica comum para este nutriente em áreas sob manejo.

A prática de adubação na área experimental sempre foi realizada com base em recomendações médias ou padronizadas para a cultura da cana-de-açúcar, ou seguindo doses uniformes. Contudo, essa abordagem foi executada sem o mapeamento prévio da variabilidade espacial da fertilidade do solo, resultando em uma aplicação que não considera as oscilações de K e P, conforme evidenciado nos mapas da Figura 6. Essa adubação uniforme sobre um solo heterogêneo leva inevitavelmente à subdosagem em áreas de baixa fertilidade e à superdosagem, acúmulo, em áreas de alta fertilidade, comprometendo a homogeneidade e, consequentemente, a precisão dos resultados dos ensaios de melhoramento genético.

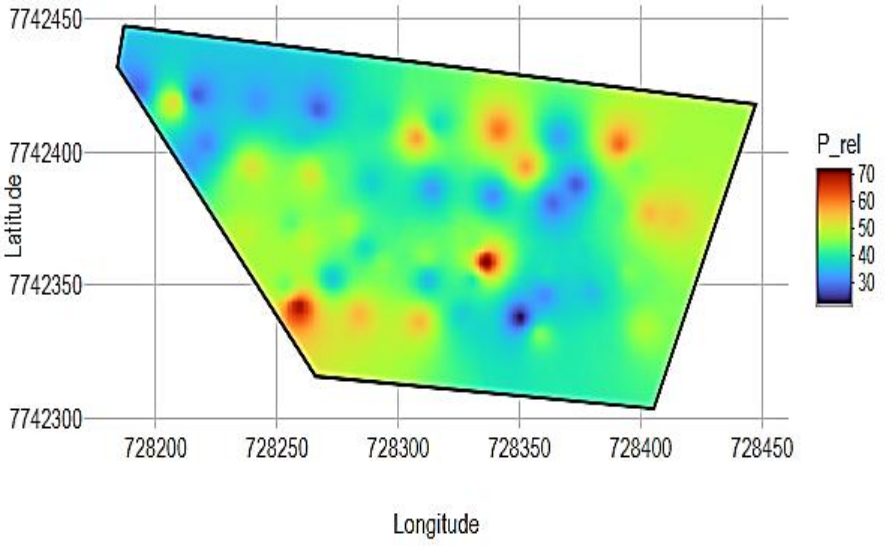
Para superar as limitações da adubação sem conhecimento prévio da fertilidade, a Figura 6, com a visualização espacial da variabilidade, é o ponto de partida. Os mapas gerados fornecem o embasamento para o desenvolvimento de zonas de manejo específicas ou para o delineamento de parcelas homogêneas. Em vez de aplicar uma dose única em toda a área, os dados da Figura 6 permitirão a criação de um plano de amostragem mais inteligente ou de um mapa de prescrição de fertilizantes, garantindo que as áreas deficientes recebam a devida correção e que a precisão experimental seja restabelecida.

Figura 6 **–** Mapa interpolado da variável Potássio (K) em cmol/dm³ pela krigagem simples (A). Mapa estimado via IDW para a variável Fósforo (P) em mg/dm³ (B).

A



B



Fonte: elaborado pelo autor.

4.3.3. Análise de agrupamento

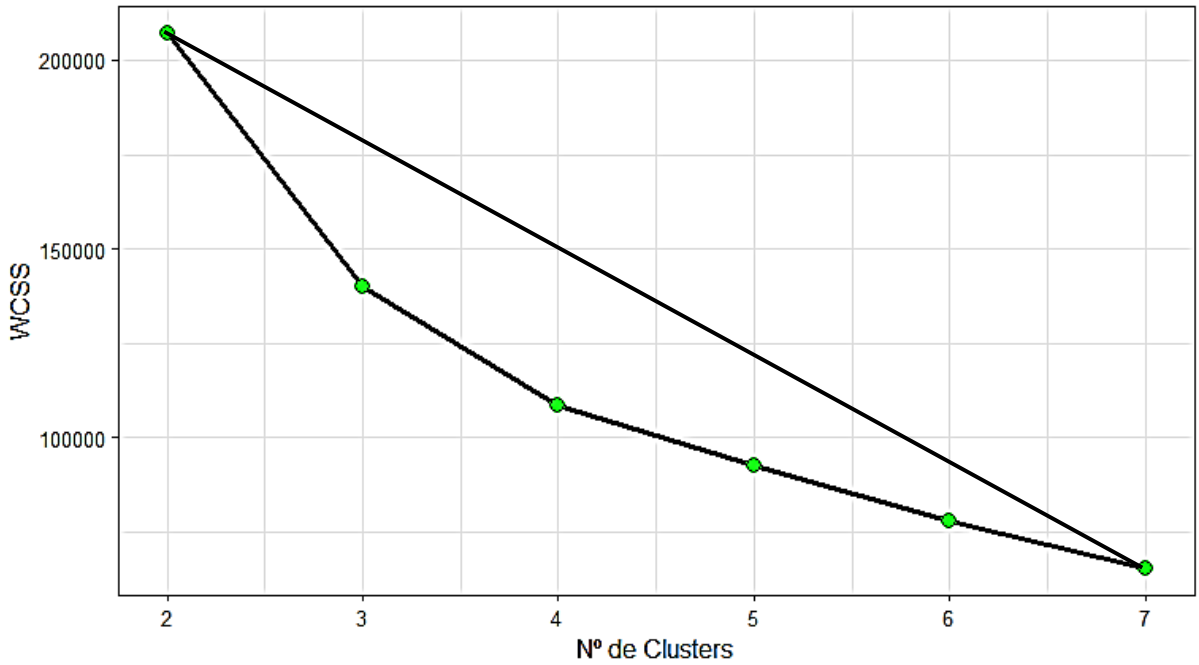
Antes de iniciar a fase de agrupamento, os dados das concentrações das variáveis fósforo e potássio foram padronizados. Para determinar os blocos experimentais ideais, com base nessas variáveis, foi utilizado o algoritmo de agrupamento *k-means*.

Dividir uma área experimental em um maior número de clusters, ou agrupamentos (*k*) aumenta a uniformidade dentro de cada cluster. No entanto, isso resulta

em uma redução do tamanho de cada cluster, o que limita tanto o número de tratamentos que podem ser testados quanto o tamanho das unidades experimentais.

Apresenta-se na Figura 7 o resultado da aplicação do Método do Cotovelo, uma técnica utilizada para auxiliar na determinação do número ótimo de clusters (agrupamentos) em análises de agrupamento. Com base na Figura 7, o número de clusters mais viável foi igual a três, pois é o ponto em que, a partir dele, se forem adicionados mais clusters, resulta em uma redução mínima no WCSS (*Within-Cluster Sum of Squares* - Soma dos Quadrados dentro do Cluster), sugerindo que esses clusters extras não são necessários (Figura 7) (Kodinariya e Makwana, 2013).

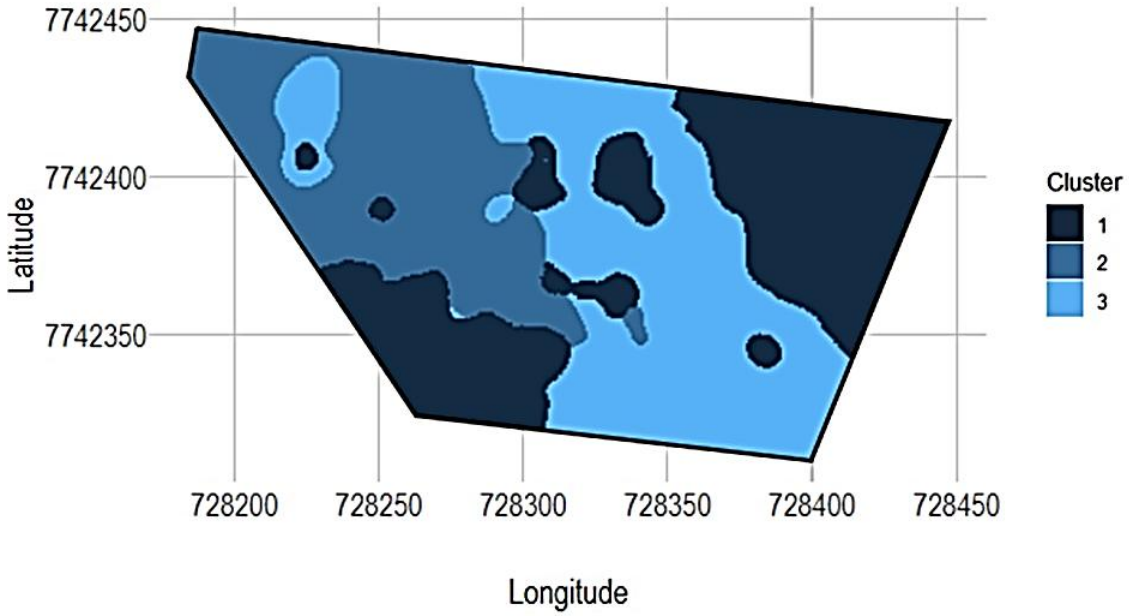
Figura 7 **–** Método do cotovelo para determinação do número de *clusters* mais viável.



Fonte: elaborado pelo autor.

Considerando que, nesse estudo, o número ideal de clusters seja três, a Figura 8 mostra os três grupos formados a partir do algoritmo KM após considerar K e P conjuntamente, conforme sugerido por Silva et al. (2024). A Figura 8 temos a representação dos clusters com maior uniformidade em relação às variáveis K e P. A ilustração destaca a quantidade de clusters, diferenciados por cores e formas.

Figura 8 **–** Blocos agrupados via algoritmo *k-means* em função de três clusters.



Fonte: elaborado pelo autor.

Analisando as Figuras 6-A, 6-B e 8 simultaneamente, foi possível identificar que a região do *cluster* um é equivalente às regiões com maiores concentrações de fósforo (média de 46,300 mg/dm³ – Tabela 6). A região do *cluster* dois representa as regiões de maiores concentrações de potássio (média de 0,1050 cmol/dm3 – Tabela 6). O *cluster* três indica regiões com os valores intermediários de concentração de fósforo e potássio (Tabela 6).

Com base nos resultados apresentados na Tabela 6, o teste t de Welch indicou, para todas as comparações entre os clusters, p-valores próximos de zero, demonstrando que as diferenças observadas entre as médias são estatisticamente significativas ao nível de 5%. Assim, há evidência robusta de que as médias das concentrações de K e P diferem entre todos os pares de clusters analisados, reforçando que cada grupo apresenta um comportamento distinto em relação a P e K.

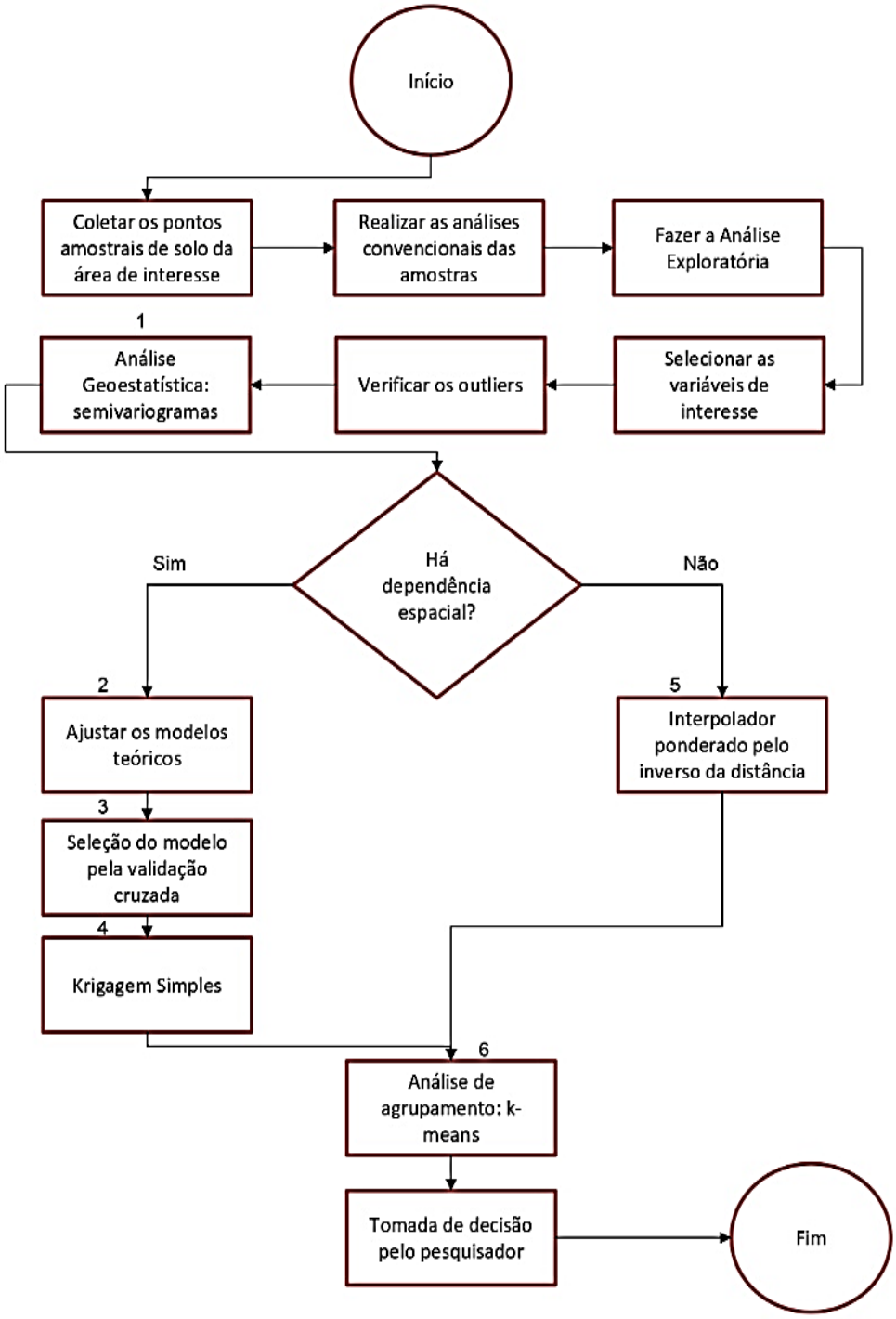
Tabela 6 - Valores médios das concentrações de cada variável para cada cluster e o p-valor do teste t de Welch para avaliar se as médias entre os grupos são estatisticamente diferentes.

	Cluster 1	Cluster 2	Cluster 3	p-valor Cluster 1 x 2	p-valor Cluster 1 x 3	p-valor Cluster 2 x 3
K (cmol/dm³)	0,0788	0,1050	0,0651	$< 2{,}2\times10^{-16}$	$< 2{,}2\times10^{-16}$	$< 2{,}2\times10^{-16}$
P (mg/dm³)	46,300	38,600	41,100	$< 2{,}2\times10^{-16}$	$< 2{,}2\times10^{-16}$	$< 2{,}2\times10^{-16}$

Os resultados obtidos fornecem informações para a correção de fertilidade de forma regionalizada, caso seja de interesse do pesquisador. Essa condição pode ser alterada a depender de tratos culturais que venham a ser realizados na área experimental em experimentos vigentes ou futuros, caso nenhuma interferência ocorra pelos pesquisadores, e que altere os valores obtidos. Além disso, o pesquisador consegue tomar decisões para realizar locações de experimentos no campo para formar blocos homogêneos.

A Figura 9 apresenta os passos necessários para implementar os métodos propostos no trabalho. Além dos passos, a contribuição do trabalho são as rotinas computacionais no software R (Quadro 1) para o pesquisador implementar em qualquer área experimental de qualquer tipo de cultivo de interesse.

Figura 9 **–** Diagrama com os passos necessários para implementar os métodos estatísticos propostos e a numeração do passo a passo das rotinas necessárias no software estatístico R apresentado no Quadro 1.



1 – 4: métodos estatísticos e geoestatísticos empregados, incluindo a análise de semivariogramas, a seleção de modelos por validação cruzada e krigagem simples. 5: alternativa de Interpolação Ponderada pelo Inverso da Distância (IDW) na ausência de dependência espacial. 6: análise de agrupamento *k-means*.

Fonte: elaborado pelo autor.

Quadro1 **–** Resumo das rotinas computacionais do software estatístico R para implantação da metodologia proposta.

#Krigagem pacote GeoR read.geodata () #ler os dados como geodados, identificando as coordenadas x, y e a varíavel de interesse variog() #semivariograma experimental (Passo 1 da Figura 6) eyefit() #ajuste visual do modelo de semivariograma teórico #cadastro dos parâmetros iniciais (sill, range e nugget) e do modelo teórico apropriado #(esférico, exponencial ou gaussiano) variofit() #Estimação do semivariograma teórico (Passo 2 da Figura 6) xvalid() #Verificaçaõ da qualidade do ajuste - cross validação (*leave - one - out*) #(Passo 3 da Figura 6) krige.conv() #interpolação via krigagem simples (Passo 4 da Figura 6) ggplot() #representação gráfica #pacote ggplot2 **#IDW** (Passo 5 da Figura 6) read.table() - utils #leitura do conjunto de dados idw() - phylin #interpolação via inverso da potência das distâncias (p=2) ggplot() #representação gráfica #pacote ggplot2 **#Kmeans** (Passo 6 da Figura 6) read.table() - utils #leitura do conjunto de dados scale() - raster #padronização das variáveis kmeans() - stats #clusterização por *Kmeans* ggplot() #representação gráfica #pacote ggplot2

Fonte: elaborado pelos autores.

4.4. Estudo da viabilidade ou não de redução de pontos amostrais

Na análise apresentada, considerando todos os pontos amostrados, a variável K obteve IDE = 0,54 classificada como dependência espacial moderada. Por outro lado, no caso da variável P, não foi identificado dependência espacial. Portanto, a variável K foi escolhida para o estudo da viabilidade ou não, de redução do número de pontos amostrais. A não escolha da análise para a variável P se justifica por sua ausência de estrutura espacial. O objetivo deste subcapítulo foi validar a eficácia dos métodos geoestatísticos em manter a acurácia do mapeamento com menor esforço amostral. Uma vez que o P não apresentou dependência espacial, a interpolação, mesmo por métodos determinísticos como o IDW, resultaria em estimativas cuja precisão não seria otimizada pela distância real de dependência.

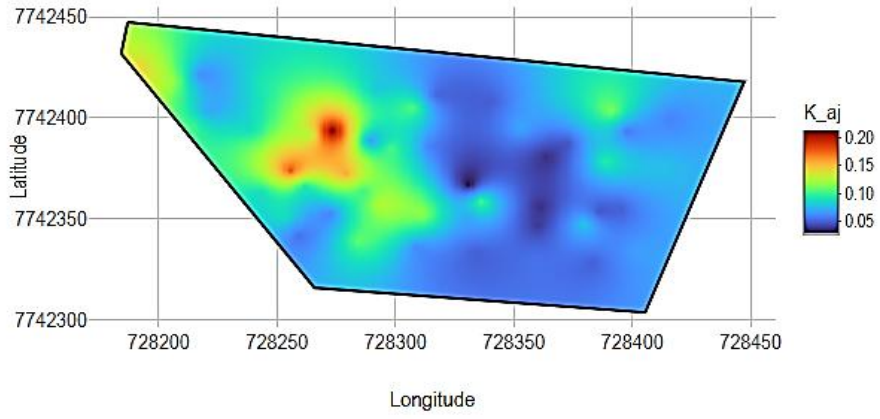
A escolha dos percentuais de redução (10% e 20%) baseou-se em dois critérios. Primeiro, reduções superiores a 20% comprometeram a estrutura espacial da variável potássio, resultando em efeito pepita puro no ajuste do semivariograma, conforme descrito por Isaaks e Srivastava (1989). Segundo, esses percentuais estão de acordo com recomendações da literatura, que indicam ser possível otimizar amostragens mantendo entre 70% e 90% dos pontos originais, desde que respeitadas

as características da área e da variável avaliada (Vieira, 2000; Cambardella et al., 1994).

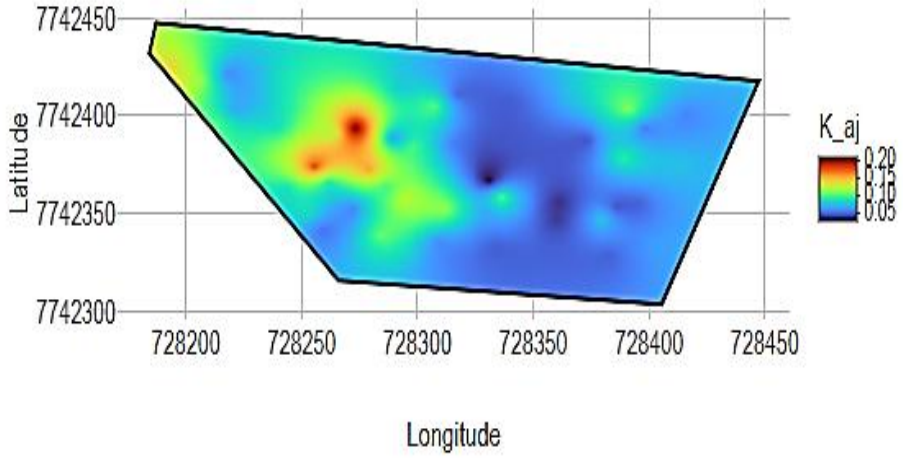
Comparando o mapa original (Figura 10-A), com o mapa sob redução de 10% (Figura 10-B), percebe-se uma pequena variação, quase imperceptível, nos teores de K. Comparando o mapa original (Figura 10-A) com o mapa sob redução de 20% (Figura 10-C), foi observado uma modificação do contraste espacial: as regiões que apresentavam maiores concentrações de K no mapa original passaram a indicar teores ainda mais elevados, enquanto as regiões com menores concentrações passaram a apresentar valores ainda mais baixos.

Figura 10 **–** Mapas interpolados da variável Potássio (K), em cmol/dm^3, pela krigagem simples, para a malha original (A), com a redução de 10% dos pontos, totalizando 54 pontos para a krigagem (B) e considerando a redução de 20% dos pontos, totalizando 48 pontos para a krigagem (C).

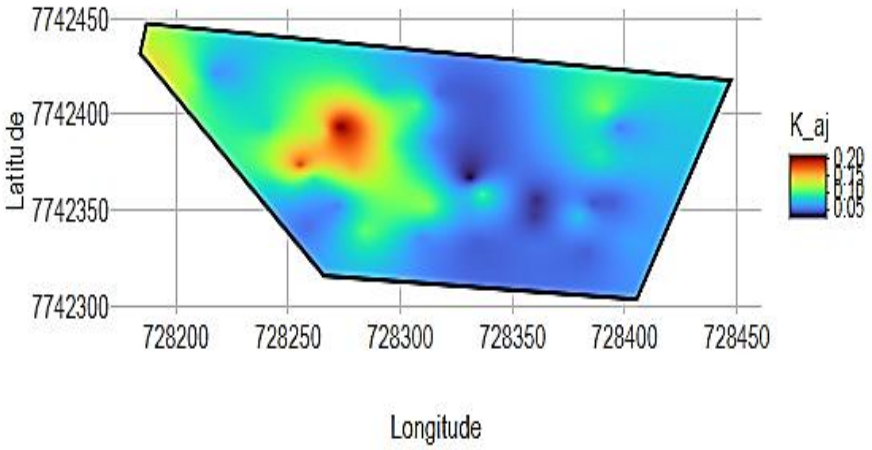
A



B



C



Fonte: elaborado pelo autor.

De acordo com os resultados obtidos na Figura 10, é possível observar que a diminuição do número de pontos tende a provocar uma perda na variabilidade espacial. Isso se justifica, pois os mapas com os pontos reduzidos, apresentaram manchas mais homogêneas, maiores e menos detalhadas (Figura 10 B e C), o que configura o efeito de suavização. Apesar dessa limitação, observou-se que o padrão geral de distribuição do micronutriente K no solo foi preservado nas malhas reduzidas.

Assim, as áreas originalmente caracterizadas por maiores concentrações mantiveram níveis elevados, enquanto as regiões com menores teores continuaram apresentando valores inferiores. Portanto, neste estudo, pode ser reduzido o número de pontos em 20%, reduzindo custos e tempos com coleta e processamento de amostras do solo. Os resultados da Tabela 7 corroboram esta conclusão, pois analisando as estatísticas descritivas para os três cenários propostos, houve diferenças mínimas entre os resultados.

Tabela 7 – Estatísticas descritivas da variável Potássio (K), para a malha original, com a redução de 10% dos pontos, totalizando 54 pontos e considerando a redução de 20% dos pontos, totalizando 48 pontos .

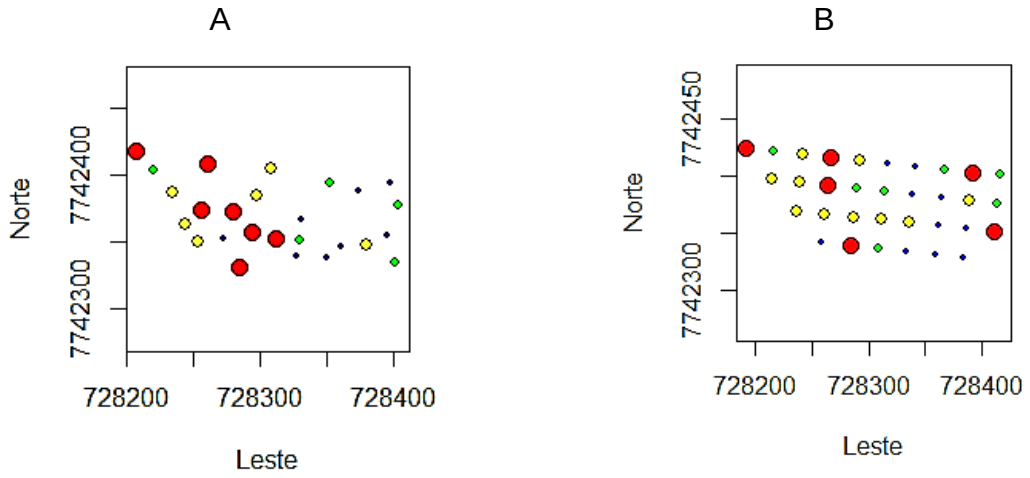
	Média	Mínimo	Máximo	Variância	Desvio padrão	Coeficiente de variação	Curtose	Assimetria
Malha original	0,0827	0,03	0,21	0,0013	0,04	43,65	4,84	1,16
Redução de 10% (54 pontos)	0,0840	0,03	0,21	0,0013	0,04	43,45	4,76	1,15
Redução de 20% (48 pontos)	0,0841	0,03	0,21	0,0014	0,04	44,82	4,65	1,15

4.5. Estudo comparativo da análise sistemática versus aleatória

Segundo Yamamoto e Landim (2015) em geoestatística, costuma-se definir um plano de amostragem sistemática utilizando uma malha regular, quadrada ou retangular, onde cada nó, ou centro, representa um ponto a ser amostrado. A aleatoriedade do processo é introduzida ao sortear o primeiro ponto de coleta da amostra. Esta etapa do trabalho teve como objetivo avaliar a distribuição sistemática e aleatória dos pontos de amostragem, com o intuito de verificar se há diferenças significativas entre as distribuições de ponto.

Para realizar o estudo comparativo, foi utilizada a variável K, visto que somente esta apresentou dependência espacial com todo o conjunto de dados. A Figura 11 apresenta a análise exploratória espacial representando os pontos distribuídos de forma aleatória (A) e pelo grid sistemático (B). As cores e os tamanhos diferentes dos pontos na Figura 11 representam as classes de valores da variável K do solo em estudo, conforme categorização realizada por quantis. O objetivo foi visualizar a distribuição espacial e a distribuição espacial da variável antes da modelagem geoestatística. Os pontos maiores e de cores quentes (vermelho/amarelo) indicam os quantis de valores mais altos. Os pontos menores e de cores frias (verde/azul) indicam os quantis de valores mais baixos.

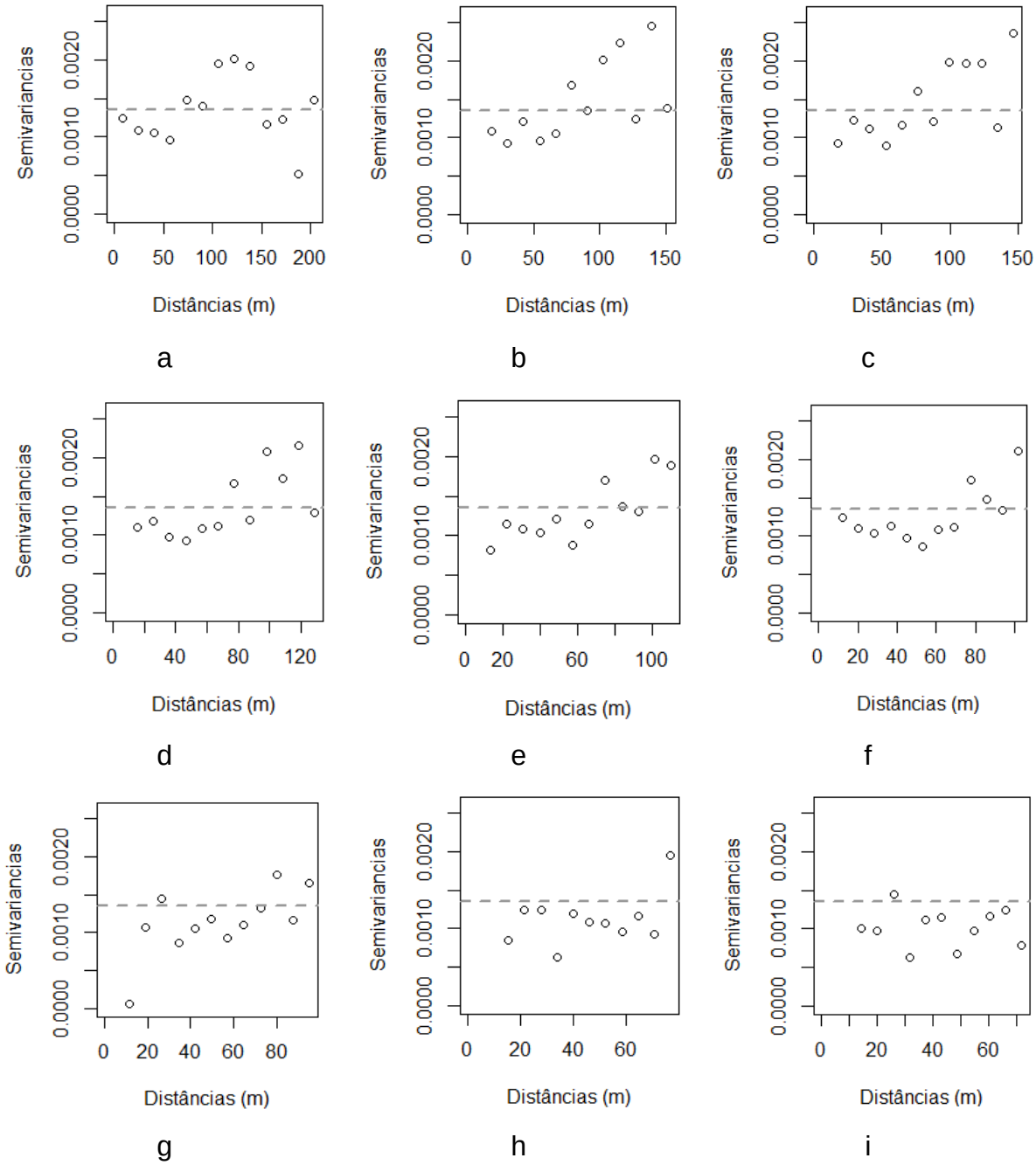
Figura 11 – Análise exploratória espacial por quantil, representando os pontos distribuídos de forma aleatória (A) e pelo grid (B). Os círculos vermelhos grandes denotam o 4º quartil (valores mais altos), os losangos amarelos e verdes representam os quartis intermediários (3º e 2º), e os círculos azuis pequenos indicam o 1º quartil (valores mais baixos).



Fonte: elaborado pelo autor.

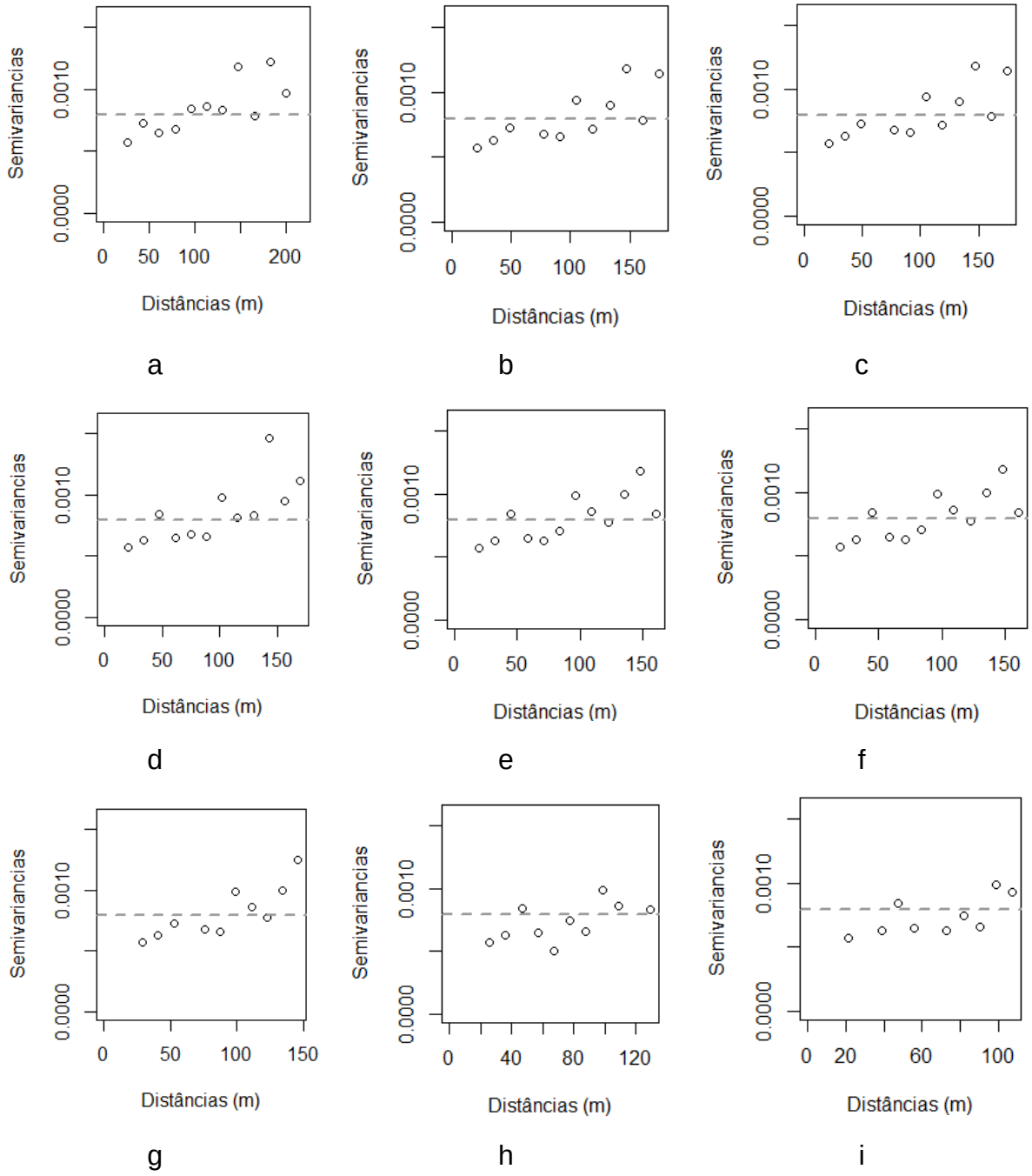
Analisando as Figura 12 e 13, percebeu-se que, para todas as tentativas de estimações do semivariograma experimental, para os pontos aleatórios e para o grid, não foi detectada dependência espacial. Os semivariogramas apresentaram, em sua maioria, o comportamento de efeito pepita puro. Portanto, considerando apenas os pontos aleatórios e o grid de forma separada, não foi possível realizar a krigagem para a variável K. Esse resultado indica que, não é viável realizar a análise considerando apenas os pontos do grid ou considerando apenas os pontos aleatórios. É importante dizer que a redução de ponto acarretou neste resultado.

Figura 12 Estimações do semivariograma experimental considerando diferentes distâncias para os pontos do grid da variável K. (a) 212 m, (b) 159 m, (c) 153 m, (d) 135 m, (e) 115 m, (f) 106 m, (g) 100 m, (h) 80 m e (i) 75 m.



Fonte: elaborado pelo autor.

Figura 13 – Estimações do semivariograma experimental considerando diferentes distâncias para os pontos do grid da variável K. (a) 230 m, (b) 190 m, (c) 185 m, (d) 180 m, (e) 175 m, (f) 170 m, (g) 153 m, (h) 135 m e (i) 115 m.



Fonte: elaborado pelo autor.

4.6. Estudo analítico entre os resíduos dos delineamentos experimentais e a fertilidade do solo

Para fazer a análise foi utilizada uma blocagem, assumindo oito blocos a priori da quadra sete. Os mapas gerados por interpolação IDW a partir dos resíduos de Brix

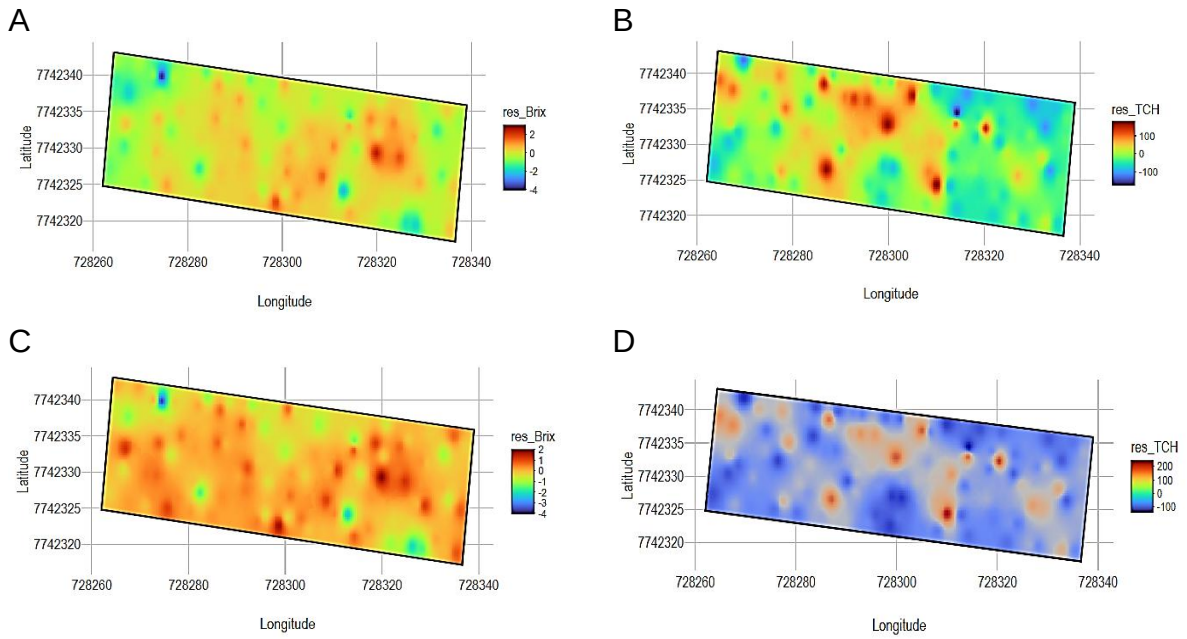
e TCH revelaram zonas distintas de variação dentro da área experimental, tanto para o delineamento DIC quanto para o DBC.

No caso do Brix, os mapas IDW demonstraram áreas com resíduos positivos (acima do esperado pelo modelo) e negativos (abaixo do esperado), sugerindo que fatores não controlados pelo experimento, como a disponibilidade de nutrientes ou condições físico-químicas do solo, podem ter influenciado a planta (Figura 14 A e C).

Para o TCH, a variação residual é maior quando comparado à variação do Brix. Observa-se um núcleo com resíduos positivos em uma faixa central da área e valores negativos nas extremidades (Figura 14 B e D).

Analisando a Figura 14 (A e B), os mapas IDW para os resíduos no DIC mostram uma estrutura com gradientes visíveis, o que viola o pressuposto de aleatoriedade dos resíduos no DIC. Ou seja, os mapas tendem a mostrar gradientes mais amplos (variação de baixa frequência), coerentes com a ausência de controle explícito de variação entre blocos. Em especial, o resíduo do TCH (Figura 14 B) mostra zonas vermelhas e azuis bem marcadas, o que indica presença de efeito ambiental não capturado pelo modelo DIC. No DBC, parte do gradiente é absorvida pelo fator bloco; o que resta nos resíduos tende a se organizar em manchas menores. Em outras palavras, o DBC reduz (não elimina) a estrutura espacial residual, sobretudo para TCH (Figura 14 C e D).

Figura 14 – Mapas interpolados via IDW para os resíduos das variáveis brix (A) e TCH (B) em DIC. E, mapas estimados via IDW para os resíduos das variáveis brix (C) e TCH (D) em DBC.

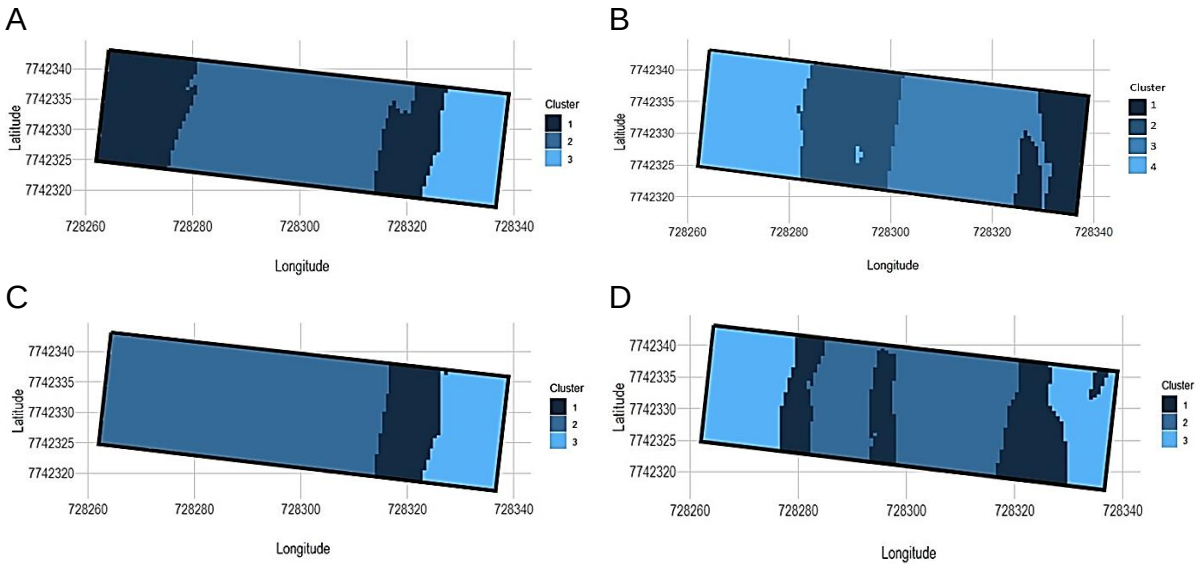


Fonte: elaborado pelo autor.

Os mapas com a divisão de clusters realizados por *k-means* (Figura 15) segmentam a área em classes espaciais de resíduos. Para Brix (Figura 15 A e C), aparecem três manchas principais relativamente estáveis em DIC e DBC, coerentes com a baixa amplitude residual. Para TCH, as classes são mais contrastantes e revelam com nitidez a faixa central com resíduos positivos e bordas negativas (o mesmo padrão observado na Figura 14).

No DIC (Figura 15 A e B), os clusters mostram divisões longitudinais acentuadas, especialmente em TCH analisado em DIC, com quatro regiões bem separadas. No DBC (Figura 15 C e D), o cluster para brix mostra regiões bem separadas de forma longitudinal, já o cluster para TCH aparece mais disperso e sem padrão de separação nítido, sugerindo que os resíduos ficaram mais aleatórios.

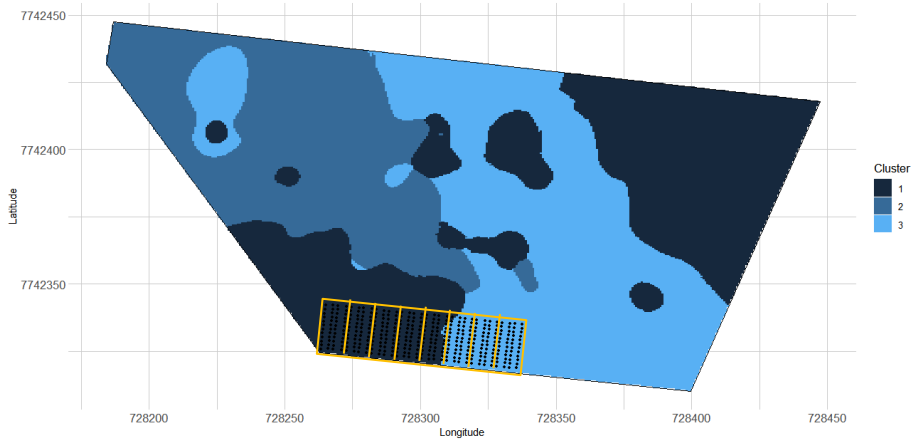
Figura 15 – Clusters agrupados via algoritmo k-means para os resíduos das variáveis brix (A) e TCH (B) em DIC. E, clusters agrupados via algoritmo k-means para os resíduos das variáveis brix (C) e TCH (D) em DBC.



Fonte: elaborado pelo autor.

De acordo com a Figura 16, a área se organiza em três zonas de fertilidade (baixa, intermediária e alta). A faixa amarela (experimento de melhoramento genético) cruza mais de uma zona. Os blocos definidos com base em fósforo e potássio via k-means mostram um padrão espacial diferenciado, com duas regiões principais distintas.

Figura 16 – Clusters agrupados via algoritmo k-means em função de três grupos, na quadra 7, para fósforo e potássio, com a faixa que representa o experimento de genética e melhoramento proposto para estudo comparativo, em amarelo.



Fonte: elaborado pelo autor.

Na Tabela 8, os valores de Brix (tanto em DIC quanto em DBC) apresentaram medianas e médias muito próximas entre clusters (≈ 20,5 a 21,5) e amplitude reduzida (mínimo ≈ 16,5 e máximo ≈ 22,8) (Tabela 8). Pequenas diferenças entre quartis, sugerindo baixa dispersão intra-cluster. Na Figura 17, ao observar os resíduos de Brix em blocos, nota-se variação entre −0,79 e +0,43, ou seja, diferenças muito sutis. Alternância de blocos positivos e negativos, mas em magnitudes pequenas. Portanto, os resultados da Tabela 8 confirmam o que a Figura 17 sugere visualmente: o Brix apresenta baixa sensibilidade à heterogeneidade espacial do solo. Embora a fertilidade (P e K) esteja estratificada, os resíduos de Brix não acompanham fortemente as divisões, indicando que essa variável da cana-de-açúcar foi relativamente estável, independentemente da variação de fertilidade.

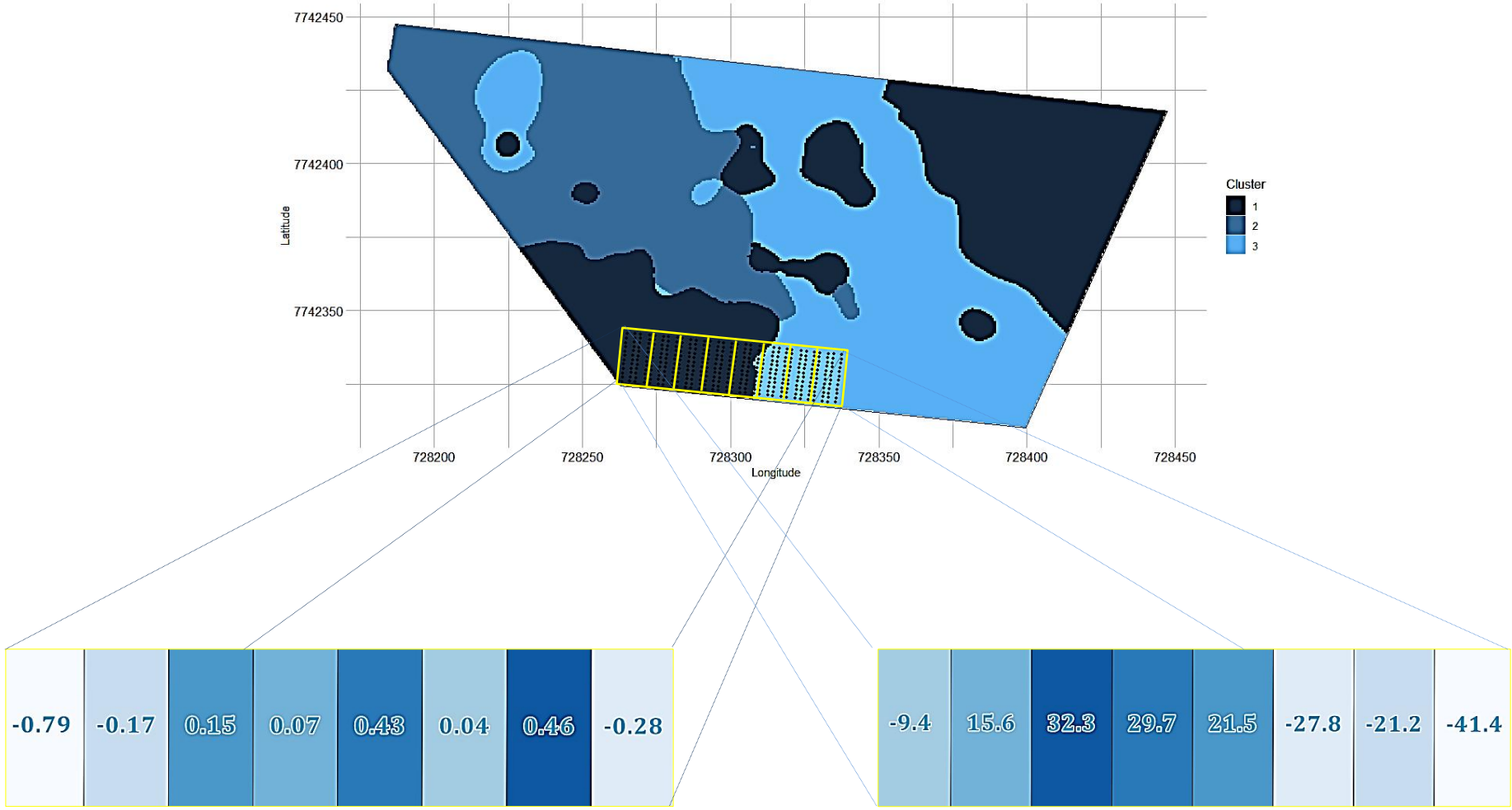
Em relação ao TCH, a Tabela 8 apresentou amplitude muito maior entre clusters (mínimos de 42–65 t ha^{-1} e máximos de 291–366 t ha^{-1}) e medianas diferenciadas entre classes (Cluster 1 ≈ 175 t ha^{-1} no DIC vs. Cluster 2 ≈ 127 t ha^{-1}). Há indícios de que os clusters de resíduos representam zonas de produtividade real distintas. Na Figura 17, a distribuição dos resíduos de TCH reforça essa leitura: valores variam de −41,4 a +32,3 t ha^{-1}, diferenças expressivas dentro da mesma área. Os blocos com resíduos positivos se concentram no setor central, sobreposto a uma das zonas de maior fertilidade. Os blocos com resíduos negativos predominam em bordas coincidentes com classes de menor fertilidade. Os padrões de dispersão observados na Tabela 8 se manifestam espacialmente na Figura 17.

Tabela 8 - Análise descritiva para os clusters agrupados via algoritmo k-means para os resíduos das variáveis brix e TCH em DIC. E, para os clusters agrupados via algoritmo k-means para os resíduos das variáveis brix e TCH em DBC.

				Brix analisado em DIC		
	Valor Mínimo	**1º quartil**	**Mediana**	**Média**	**3º quartil**	**Valor Máximo**
Cluster 1	18,06	20,45	21,17	20,96	21,57	22,18
Cluster 2	18,46	19,80	20,87	20,56	21,32	22,14
Cluster 3	16,67	20,18	20,89	20,74	21,54	22,81
				Brix analisado em DBC		
	Valor Mínimo	**1º quartil**	**Mediana**	**Média**	**3º quartil**	**Valor Máximo**
Cluster 1	16,67	20,20	20,94	20,76	21,46	22,18
Cluster 2	18,46	19,80	20,87	20,56	21,32	22,14
Cluster 3	20,22	20,93	21,55	21,45	21,66	22,81
				TCH analisado em DIC		
	Valor Mínimo	**1º quartil**	**Mediana**	**Média**	**3º quartil**	**Valor Máximo**
Cluster 1	65,05	119,79	175,71	172,05	197,50	291,43
Cluster 2	42,86	92,22	127,99	133,97	161,65	251,79
Cluster 3	16,07	109,75	134,42	158,59	192,05	364,02
Cluster 4	66,07	149,73	186,52	202,92	250,09	366,43
				TCH analisado em DBC		
	Valor Mínimo	**1º quartil**	**Mediana**	**Média**	**3º quartil**	**Valor Máximo**
Cluster 1	16,07	124,38	174,82	186,90	243,21	366,43
Cluster 2	42,86	110,27	161,38	157,11	196,07	291,43
Cluster 3	84,82	120,74	151,56	158,85	175,89	269,29

Figura 17 – Clusters agrupados via algoritmo k-means em função de três grupos, na quadra 7, para fósforo e potássio, com a faixa que representa o experimento de genética e melhoramento proposto para estudo comparativo, em amarelo e a representação da média dos resíduos de brix e TCH na forma de blocos.

.



Resíduos Brix

Resíduos TCH

5. CONCLUSÕES

Os resultados obtidos apresentaram informações da variabilidade do solo para as variáveis fósforo e potássio, de forma que o pesquisador consiga realizar a correção de fertilidade, de forma regionalizada, numa quadra experimental.

Com o agrupamento realizado percebeu-se que é possível tomar decisões mais precisas, que podem economizar recursos financeiros no planejamento experimental, visto que o pesquisador já tem o conhecimento prévio dos locais onde se concentram mais ou menos fósforo e potássio. Além disso, pode-se otimizar a área experimental, uma vez que é possível alocar futuros experimentos de acordo com a região de fertilidade mais adequada.

O estudo também incentiva a aplicação dos métodos estatísticos considerados para a experimentação, pois permitem reduzir o número de análises de solo, uma vez que as informações de fertilização são conhecidas e regionalizadas. Além disso, com a redução de 20% do número de pontos de amostragem foi possível concluir que é possível reduzir custos com as análises laboratoriais do solo e reduzir tempo de coleta.

Embora o foco tenha sido a cana-de-açúcar, o embasamento teórico e prático proposto pode ser aplicado a qualquer cultura. Recomenda-se que futuras pesquisas utilizem as ferramentas estatísticas indicadas para outras áreas experimentais, reforçando a otimização no planejamento.

REFERÊNCIAS BIBLIOGRÁFICAS

ARCOVERDE, S. N. S.; SOUZA, C. M. A.; NAGAHAMA, H. J.; CORTEZ, J. W.; NASCIMENTO, J. M. Spatial variability of soil physical attributes under conservation management systems for sugarcane cultivation. **Revista Engenharia na Agricultura**, v.31, p.127-139, 2023.

AGEITEC, Agência Embrapa de Informação Tecnológica. *Árvore do conhecimento – cana-de-açúcar*. Embrapa, disponível em: < https://www.agencia.cnptia.embrapa.br/gestor/cana-deacucar/arvore/CONTAG01_18_3112006152934.html> acesso em 30/09/2024.

BERNARDI, A. C. de C.; NAIME, J. de M.; RESENDE, A. V.; BASSOI, L. H.; INAMASU, R. Y. **Agricultura de precisão: resultados de um novo olhar**. Embrapa Instrumentação-Livro técnico (INFOTECA-E), 2014.

BORDONAL, R. O.; CARVALHO, J. L. N.; LAL, R.; FIGUEIREDO, E. B.; OLIVEIRA, B. G.; SCALA JR., N. L. Sustainability of sugarcane production in Brazil. A review. **Agronomy for Sustainable Development**, 38:13, 2018.

BRITO, C. M. **Variabilidade genética em cana-de-açúcar na tolerância ao alumínio**. 2014. 24p. Tese (Doutorado em Genética e melhoramento) – Universidade Federal de Viçosa, Viçosa, MG.

CAMBARDELLA, C.A.; MOORMAN, T. B.; NOVAK, J. M.; PARKIN, T. B.; KARLEN, D. L.; TURCO, R. F.; KANOPKA, A. E. Field-scale variability of soil properties in central Iowa soils. **Soil science society of America jornal,** 58.5:1501-1511, 1994.

CHILÈS, J. P.; DELFINER, P. **Geostatistics: Modeling Spatial Uncertainty**. 2º ed. New York: John Wiley & Sons, 2012, 741 p.

CRESSIE, Noel. Fitting variogram models by weighted least squares. **Journal of the international Association for mathematical Geology**, v. 17, n. 5, p. 563-586, 1985.

DEVORE, J. L. **Probabilidade e estatística: para engenharia e ciências**. 6. ed. São Paulo: Cengage Learning, 2006. 708 p.

DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.V.M. (eds). **Análise Espacial de Dados Geográficos**. Brasília, EMBRAPA, 2004.

EMBRAPA - Empresa Brasileira de Pesquisa Agropecuária. **Manual de métodos de análise de solo**. 2. ed. Rio de Janeiro: Embrapa Solos. 230 p, 2022.

EPSTEIN, E.; BLOOM, A. J. **Nutrição Mineral. Princípios e perspectivas**. 2.ed., Londrina: Editora Planta, 2006.

FERNANDES, A. M.; QUEIROZ, A. C.; PEREIRA, J. C.; LANA, R. P.; BARBOSA, M. H. P.; FONSECA, D. M.; DETMANN, E.; CABRAL, L. S.; PEREIRA, E. S.; VITTOR, A. Composição químico-bromatológica de variedades de cana-de-açúcar (Saccharum

ssp L.) com diferentes ciclos de produção (precoce e intermediário) em três idades de corte. **Revista Brasileira de Zootecnia**, v. 32, n.4, p.977-985, 2003.

FERREIRA, D. F. **Análise multivariada**. 3. Ed. Lavras: UFLA, 2018. p. 394.

FERREIRA, M. P. New insights into adjustment for spatial dependence of soil attributes and use of aerial images in the initial selection stage of sugarcane families. **Scientia Agraria**, v. 20, n. 1, 2024.

FORMAGGIO, A. R; SANCHES, I. D. **Sensoriamento Remoto em agricultura**. 1. ed. São Paulo: Oficina de Textos, 2017. 284 p.

GREGO, C. R.; VIEIRA, S. R. Variabilidade espacial de propriedades físicas do solo em uma parcela experimental. **Sociedade Brasileira de Ciência do Solo**, v. 29, n. 2, p. 169-177, 2005.

GOOVAERTS, P. **Geostatistics for natural resources evaluation**. Oxford University Press, New York, 1997.

GUIMARÃES, W. D. **Geoestatística para o mapeamento da variabilidade espacial de atributos física do solo**. 2013. 60 p. Tese (Doutorado em Engenharia Civil) Universidade Federal de Viçosa, Viçosa, 2013.

HERNÁNDEZ, M. M. **Análise geoestatística de multivariada para definição de zonas de manejo em cana-de-açúcar (Saccharum officinarum) na Guatemala**. 2021. 60 p. Dissertação (Mestrado em Estatística Aplicada e Biometria) Universidade Federal de Viçosa, Viçosa, 2021.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Applied geostatistics**. Nova York: Oxford Univ. Press, p .561, 1989.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. An Introduction to Statistical Learning with Applications in R. New York: Springer, 2013.

JÚNIOR, R. et al. **geoR: Analysis of Geostatistical Data**. R package version 1.8-1. p. 155, 2020.

KODINARIYA, T. M.; MAKWANA, P. R. “Review on determining number of cluster in K-means clustering,” **International Journal**, vol. 1, no. 6, pp. 90–95, 2013.

KORNDÖRFER, G. H.; FARIA, R. J.; MARTINS, M. Efeito do fósforo na produção da cana-de-ano e cana-soca cultivada em solo de cerrado. **Pesquisa Agropecuária Brasileira**, Brasília, v. 33, n. 10, p. 1667-1673, 1998.

KUMAR, P.; KUMAR, P.; SHUKLA, A. K. Spatial Modeling of Some Selected Soil Nutrients Using Geostatistical Approach for Jhandutta Block (Bilaspur District), Himachal Pradesh, India. **Agricultural Research**, 2020.

LIMA, C. G. R.; CARVALHO, M. DE P. E; SOUZA, A.; COSTA, N. R.; MONTANARI, R. Correlação entre componentes da produtividade da cana-de-açúcar com ph de um

argissolo vermelho distrófico do noroeste paulista. **Revista Engenharia na Agricultura - REVENG**, v. 24, n. 2, p. 120-130, 30 jun. 2016.

MANGABEIRA, J. A. de C.; VALLADARES, G. S.; BATISTELLA, M.; GOMES, E. G. Relação espacial entre produtividade agrícola e fertilidade do solo com uso de imagens de satélite e técnicas de geoprocessamento. **Anais XII Simpósio Brasileiro de Sensoriamento Remoto, Goiânia: SBSR**, p. 193- 195, 2005.

MARIN, F. R. **Árvore do conhecimento cana-de-açúcar: características**. Agência Embrapa de Informação Tecnológica (Ageitec), [s.d.] Disponível em:,< http://www.agencia.cnptia.embrapa.br/gestor/canadeacucar/arvore/CONTAG01_20_3112006152934.html>. Data de acesso: 24 de abril de 2024.

MCKAY, M. D.; BECKMAN, R. J.; CONOVER, W. J. Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. **Technometrics**, v. 21, n. 2, p. 239–245, maio 1979.

MCKILLUP, S. & DARBY DYAR, M. 2010. **Geostatistics Explained: An Introductory Guide for Earth Scientists**. 396pp, Cambridge University Press, 2010.

MINASNY, B.; MCBRATNEY, A. B. A conditioned Latin hypercube method for sampling in the presence of ancillary information. **Computers & Geosciences**, v. 32, n. 9, p. 1378–1388, nov. 2006.

NETO, P. L. O. C. **Estatística**. 2. ed. São Paulo: Edgar Blücher Ltda, 2002. 265 p.

OLIVEIRA, I. P; COSTA K. P.; RODRIGUES, C.; MACEDO, R. F.; MOREIRA, F. P. e SANTO, K. J. G. Manutenção e correção da fertilidade do solo para inserção do cerrado no processo produtivo. **Revista Eletrônica Montes Velos**, Goiás, ISSN 1808-8597, 2005, Vol 1(1), p. 50-64, 2005.

OLIVER, M. A.; WEBSTER, R. A tutorial guide to geostatistics: Computing and modelling variograms and kriging. **Catena**, v. 113, n. 0, p. 56-69, 2014.

PENATTI, P. C. **Adubação da cana-de-açúcar: 30 anos de experiência**. São Paulo: Ottoni Editora, 2013.

PEREIRA, R. G. **Características morfológicas e teores de cafeína e teobromina entre ervais nativos sombreados e a pleno sol**. 2021. 62 p. Dissertação (Mestrado em Ciências Florestais) Universidade Estadual do Centro-Oeste, Irati, 2021.

PREZOTTI, L.C.; GUARÇONI, A. **Guia de interpretação de análise de solo e foliar**. Vitória: Incaper, 2013. 104 p.

R Development Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2019. ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org>.

RAIJ, B. V.; CANTARELLA, H.; QUAGGIO, J. A.; FURLANI, A. A. M. C. **Recomendações de adubações e calagem para o estado de São Paulo**. Instituto Agronômico, 1996. 285 p. (IAC Boletim Técnico 100).

REETZ. H. F. Jr. **Fertilizantes e o seu uso eficiente**. Tradução: Alfredo Scheid Lopes – São Paulo: Associação Nacional para Difusão de Adubos (ANDA), 2017. Disponível em: < https://www.ufla.br/dcom/wp-content/uploads/2018/03/Fertilizantes-e-seu-usoeficiente-WEB-Word-Ouubro-2017x-1.pdf > Acesso em: 17 abr. 2024.

RODRIGUES, M. S.; CASTRIGNANÒ, A.; BELMONTE, A.; SILVA, K. A.; LESSA, B. F. T. Geostatistics and its potential in Agriculture 4.0. **Revista Ciência Agronômica**, v. 51, 2020.

RONQUIM, C.C. 2010. **Conceitos de Fertilidade do Solo e Manejo Adequado para as Regiões Tropicais**. Campinas: Embrapa Monitoramento por Satélite, 26p, 2010.

ROSSETTO, R.; SPIRONELLO, A.; CANTARELLA, H.; QUAGGIO, J. A. Calagem para a cana-de-açúcar e sua interação com a adubação potássica. **Bragantia**, Campinas, v.63, n.1, p.105-119, 2004.

SALVIANO, A. M.; MOURA, M. S. B.; SILVA, T. G. F.; CARMO, F. A.; BRANDÃO, E. O. Acúmulo e exportação de macronutrientes pela cana-de-açúcar irrigada no semiárido brasileiro. **Embrapa Semiárido - Revista Científica Intelletto**. Venda Nova do Imigrante – ES. p. 16-27, v.2, n.2, 2017.

SANTOS, G. R. **Hierarquização geométrica dos preditores geoestatísticos**. Tese (Doutorado em Estatística e Experimentação Agropecuária) – Setor de Ciências Exatas, Universidade Federal de Lavras, Lavras. 2010. 200p.

SANTOS, G. R.; OLIVEIRA, M. S.; LOUZADA, J. M.; SANTOS, A. M. R. T. Krigagem simples versus krigagem universal: qual o preditor mais preciso? **Energia na Agricultura**, 26, 2011.

SILVA, G.H. da; PEREIRA, K.D.; CARNEIRO, A.P.S.; FERREIRA, M. de P.; SANTOS, G.R. dos; PETERNELLI, L.A. Geostatistics and multivariate analysis to determine experimental blocks for sugarcane. **Pesquisa Agropecuária Brasileira**, v.59, e03373, 2024.

SILVA, G. H. **Determinação de blocos experimentais utilizando geoestatística, componentes principais e técnicas de agrupamento**. 2022. 69 p. Dissertação (Mestrado em Estatística Aplicada e Biometria) Universidade Federal de Viçosa, Viçosa, 2022.

VARELLA, C. A. A., BARROS, V. R., & BAESSO, M. M. (2012). MAPEAMENTO DA VARIABILIDADE ESPACIAL DO BRIX EM CANA-DE-AÇÚCAR. **Revista Engenharia Na Agricultura** - *REVENG*, *20*(2), 112–117, 2012.

VIEIRA, S. R. **Uso de geoestatística em estudos de variabilidade espacial de propriedades do solo**. In: NOVAIS, R. F.; ALVAREZ V, V. H.; SCHAEFER, C. E. G.

R. (Org.). Tópicos em Ciência do Solo. Viçosa: Sociedade Brasileira de Ciência do Solo, 2000. v. 1, p. 1-54.

VIEIRA, S.R.; CARVALHO, J.R.P. de; PAZ GONZÁLEZ, A. Jack knifing for semivariogram validation. **Bragantia**, v.69, p.97-105, 2010.

VITTI G. C; QUEIROZ F. E. C; OTTO R.; QUINTINO T. A; **Nutrição e adubação da cana-de-açúcar**; Departamento de Solos e Nutrição de Plantas – ESALQ/USP – Piracicaba – SP, 2005.

YAMAMOTO, J.K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações**. 1ª ed. São Paulo: Oficina de Textos. 215p., 2013.

WARRICK, A. W.; NIELSEN, D. R. Spatial variability of soil physical properties in the field. In: HILLED, D. (ed). **Applications of soil physics**. New York: Academic, 1980.

ZIMBACK, C.R.L. **Análise espacial de atributos químicos de solos para fins de mapeamento da fertilidade**. Botucatu, Universidade Estadual Paulista, 2001. 114p. Tese (livre-Docência).

APÊNDICE A – Tabela com os dados amostrais provenientes do resultado da análise de fertilidade do solo, para os 60 pontos amostrados

ID	B	Ca	Cu	Fe	H_Al	K	Mg	Mn	MO	P	P_rem	PR	pH	V	Zn
1	0,1	3,63	2,0	64,4	1,32	0,05	0,87	64,4	2,0	7,7	18,6	68,7	7,3	77,5	3,4
2	0,2	3,15	1,8	61,8	1,98	0,11	0,80	96,4	2,0	8,0	26,6	55,6	7,0	67,2	3,4
3	0,2	2,96	2,3	69,0	3,63	0,06	0,72	84,4	2,3	8,3	28,1	55,3	6,3	50,7	4,1
4	0,1	3,02	1,3	48,9	0,66	0,05	0,77	68,8	1,9	5,2	25,8	37,0	6,9	85,3	2,2
5	0,2	2,32	1,2	39,8	1,65	0,05	0,63	60,9	1,5	7,4	31,7	44,8	6,5	64,5	2,0
6	0,1	1,88	2,3	91,7	1,65	0,05	0,48	75,7	1,2	6,3	29,8	40,1	6,2	59,4	2,1
7	0,2	1,64	2,1	70,9	4,29	0,11	0,42	52,2	1,6	24,4	34,9	136,2	5,3	33,6	3,3
8	0,1	1,79	1,7	72,8	1,98	0,05	0,44	63,7	1,3	6,8	33,8	39,0	5,9	53,5	2,0
9	0,1	2,05	1,5	117,4	1,32	0,04	0,52	80,4	1,2	6,8	33,8	39,0	6,7	66,4	1,3
10	0,1	2,27	1,1	43,6	2,31	0,10	0,62	52,3	1,9	12,4	33,8	71,1	5,9	56,4	1,6
11	0,1	2,69	1,6	67,1	0,99	0,10	0,72	73,8	2,0	6,8	28,1	45,3	6,3	78,0	1,7
12	0,1	2,35	2,0	80,5	2,31	0,10	0,59	60,8	1,6	5,7	28,1	38,0	6,3	56,8	1,9
13	0,1	2,84	1,6	49,4	3,30	0,09	0,72	56,6	2,3	7,1	26,6	49,3	6,2	52,5	2,6
14	0,2	3,51	2,0	64,3	2,97	0,10	0,92	80,6	2,7	6,8	25,1	49,3	6,3	60,4	2,5
15	0,1	2,27	1,6	77,8	1,32	0,08	0,61	85,1	1,5	10,3	37,3	54,3	6,7	69,2	1,7
16	0,1	1,99	1,5	79,8	1,65	0,09	0,60	68,3	1,6	7,7	36,0	41,9	6,2	61,9	1,5
17	0,1	1,74	1,7	92,1	0,99	0,04	0,49	55,8	1,1	5,2	36,0	28,3	6,4	69,6	1,4
18	0,1	2,02	1,3	62,7	1,65	0,05	0,55	58,0	1,2	5,4	33,8	31,0	6,6	61,4	1,1
19	0,1	2,19	1,3	42,4	1,98	0,06	0,63	43,9	1,7	5,2	30,8	32,2	6,2	59,3	1,7
20	0,1	2,46	1,4	52,1	2,97	0,06	0,66	54,0	2,0	5,7	29,8	36,3	6,1	51,7	1,6
21	0,1	2,83	1,9	63,1	3,30	0,13	0,78	80,3	2,6	7,4	26,6	51,4	5,9	53,1	2,8
22	0,1	3,14	2,0	73,4	3,96	0,09	0,77	63,2	2,8	6,5	21,9	52,0	5,9	50,3	2,6
23	0,2	2,52	1,6	70,5	5,28	0,10	0,58	39,4	3,1	3,9	21,3	31,8	5,6	37,7	2,2
24	0,2	1,72	1,6	71,5	1,32	0,06	0,57	53,1	1,2	8,6	36,0	46,8	6,7	64,0	1,0

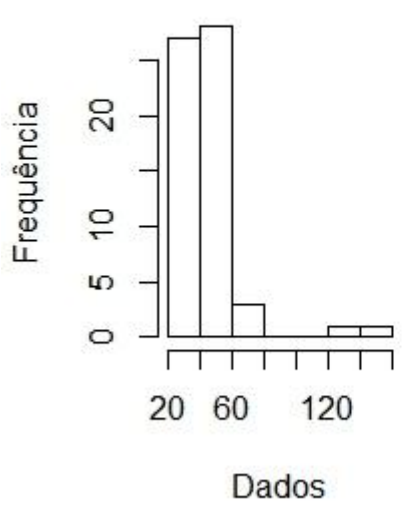
25	0,1	2,17	2,0	91,1	1,65	0,11	0,65	70,5	1,5	10,3	32,7	60,7	6,6	64,0	2,1
26	0,1	2,07	1,3	62,9	1,98	0,07	0,57	55,9	2,3	5,7	33,8	32,7	6,4	57,8	1,5
27	0,1	1,77	1,6	84,5	1,65	0,05	0,49	58,3	1,7	9,9	31,7	59,9	6,1	58,3	1,1
28	0,1	2,16	2,0	80,4	1,32	0,05	0,62	75,3	1,9	6,0	29,8	38,2	6,3	68,2	2,1
29	0,1	1,72	1,8	74,9	2,97	0,08	0,50	54,7	2,1	5,7	28,1	38,0	5,5	43,6	2,4
30	0,1	1,88	1,8	67,9	2,97	0,11	0,53	60,0	2,1	3,6	23,1	27,7	5,7	45,9	1,9
31	0,2	2,66	1,6	60,5	3,96	0,09	0,67	72,8	3,0	3,9	21,3	31,8	5,9	46,3	2,1
32	0,1	2,62	1,6	57,9	3,63	0,06	0,71	45,0	3,0	3,2	19,6	27,6	6,0	48,3	1,8
33	0,2	2,37	1,7	99,5	3,63	0,14	0,74	44,9	2,6	3,9	25,1	28,3	6,0	47,2	2,1
34	0,2	2,67	1,6	68,9	3,63	0,12	0,78	41,8	2,8	6,3	20,7	52,4	5,9	49,6	2,8
35	0,2	2,73	2,0	67,0	3,63	0,07	0,76	45,6	2,7	3,4	18,6	30,3	5,9	49,5	2,8
36	0,2	2,69	1,4	51,4	2,64	0,12	0,75	66,0	2,3	4,9	25,1	35,6	6,1	57,4	2,7
37	0,2	2,73	1,6	50,6	2,97	0,21	0,84	71,6	2,3	6,3	27,3	42,9	6,0	56,0	3,0
38	0,1	2,32	1,4	50,1	2,31	0,10	0,67	60,3	2,0	6,5	31,7	39,3	6,1	57,2	2,2
39	0,1	2,70	1,6	78,6	1,32	0,11	0,72	91,2	1,6	8,6	28,1	57,3	6,6	72,7	2,1
40	0,1	1,55	1,5	67,3	2,97	0,07	0,42	57,1	1,5	10,6	36,0	57,6	5,7	40,7	2,0
41	0,1	2,14	1,2	49,7	2,64	0,05	0,58	43,5	1,6	4,9	34,9	27,4	6,2	51,2	1,3
42	0,1	1,80	1,4	56,5	1,98	0,06	0,60	57,0	1,6	8,0	34,9	44,7	6,1	55,4	1,0
43	0,1	1,93	1,8	98,3	1,32	0,08	0,55	73,7	1,3	9,9	34,9	55,3	6,1	66,0	1,4
44	0,1	2,06	2,1	79,9	1,65	0,05	0,52	80,0	1,5	8,0	34,9	44,7	6,5	61,4	2,8
45	0,1	2,58	1,7	80,3	1,32	0,07	0,65	82,6	1,5	8,3	33,8	47,6	6,8	71,4	1,9
46	0,1	1,77	1,5	63,7	1,65	0,08	0,47	56,8	1,5	6,0	33,8	34,4	6,0	58,4	1,3
47	0,1	1,90	1,2	71,3	2,31	0,04	0,46	49,8	1,5	5,7	36,0	31,0	6,0	51,0	1,2
48	0,1	2,50	1,1	49,2	1,65	0,05	0,57	63,6	1,5	3,6	29,8	22,9	6,7	65,4	1,3
49	0,1	2,96	2,0	109,8	1,98	0,06	0,73	96,5	2,0	5,4	28,1	36,0	6,6	65,4	2,0
50	0,1	2,41	1,6	72,7	2,64	0,07	0,60	65,7	2,0	6,5	31,7	39,3	5,9	53,8	2,0
51	0,1	1,77	1,5	69,5	1,98	0,03	0,47	57,3	1,6	7,7	33,8	44,2	6,2	53,4	1,3
52	0,1	2,49	1,4	44,9	3,63	0,12	0,67	48,9	2,1	5,7	31,7	34,5	5,9	47,5	1,9

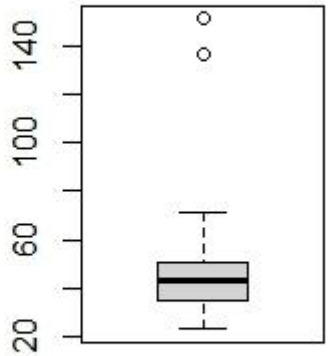
53	0,1	1,94	1,4	43,4	2,97	0,13	0,58	54,4	1,9	7,4	31,7	44,8	5,5	47,2	1,8
54	0,1	3,21	1,8	51,2	2,31	0,13	0,82	72,7	2,3	24,4	30,8	151,1	6,3	64,3	2,9
55	0,2	2,35	2,2	89,8	3,30	0,16	0,67	82,1	2,1	6,8	26,6	47,2	5,9	49,1	2,3
56	0,1	2,46	1,7	59,7	1,32	0,06	0,67	61,0	2,0	4,7	23,8	35,4	6,5	70,7	2,0
57	0,1	3,01	1,8	61,9	3,30	0,09	0,80	61,6	2,1	5,7	23,1	43,9	6,3	54,2	2,0
58	0,2	2,82	1,8	61,9	2,97	0,18	0,74	77,7	2,7	6,3	27,3	42,9	5,8	55,7	2,7
59	0,1	3,56	1,6	51,4	1,98	0,09	0,89	66,7	2,8	6,0	22,5	47,1	6,7	69,6	1,8
60	0,1	3,12	2,1	56,3	3,96	0,10	0,81	59,0	2,8	6,3	25,1	45,7	6,1	50,4	2,8

APÊNDICE B – Histogramas e boxplot

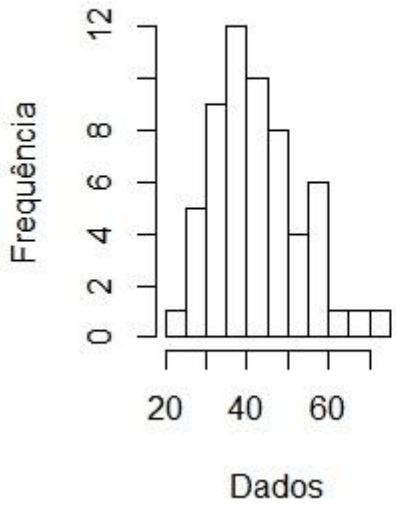
Figura B-1 – Quadro representativo com o histograma e boxplot para fósforo e potássio, com e sem outliers.

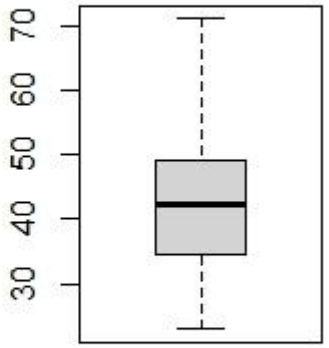
Fósforo (P) com outliers



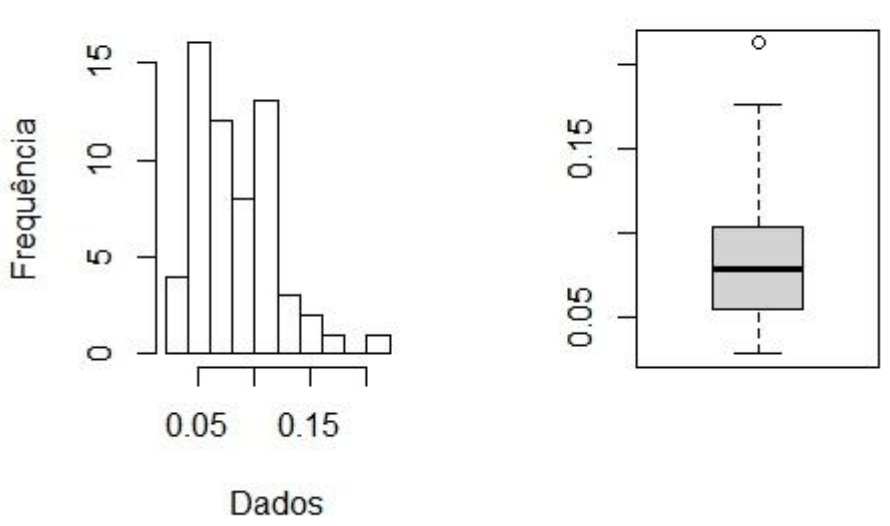


Fósforo (P) sem outliers

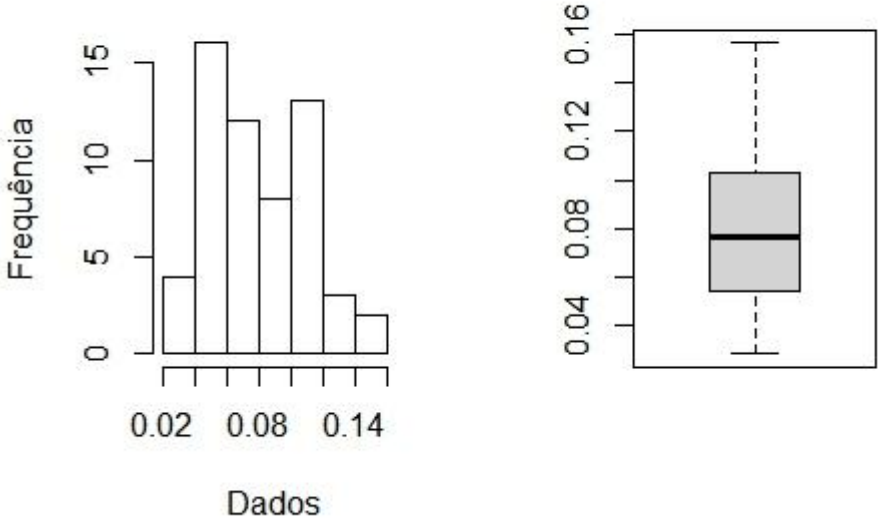




Potássio (K) com outlier



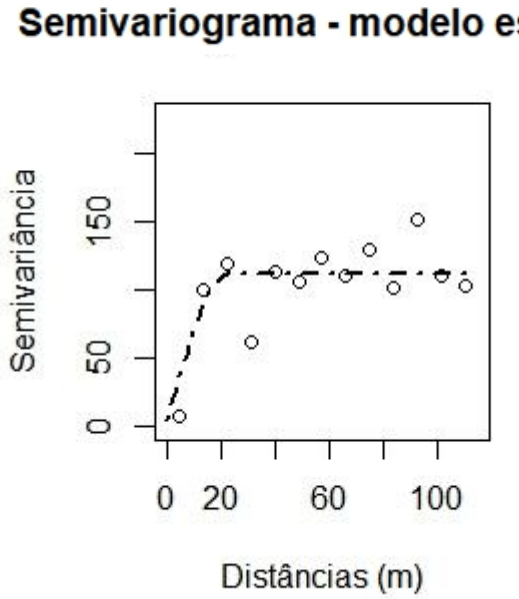
Potássio (K) sem outliers



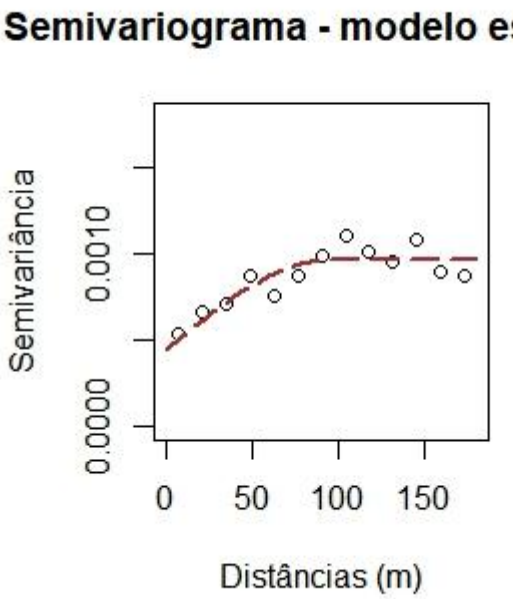
APÊNDICE C – Semivariogramas experimentais obtidos e a curva que representa o modelo teórico ajustado para fósforo e potássio

Figura C-1 - Semivariogramas experimental omnidirecional e curva de modelo teórico ajustada para os atributos fósforo e potássio.

Fósforo (P)



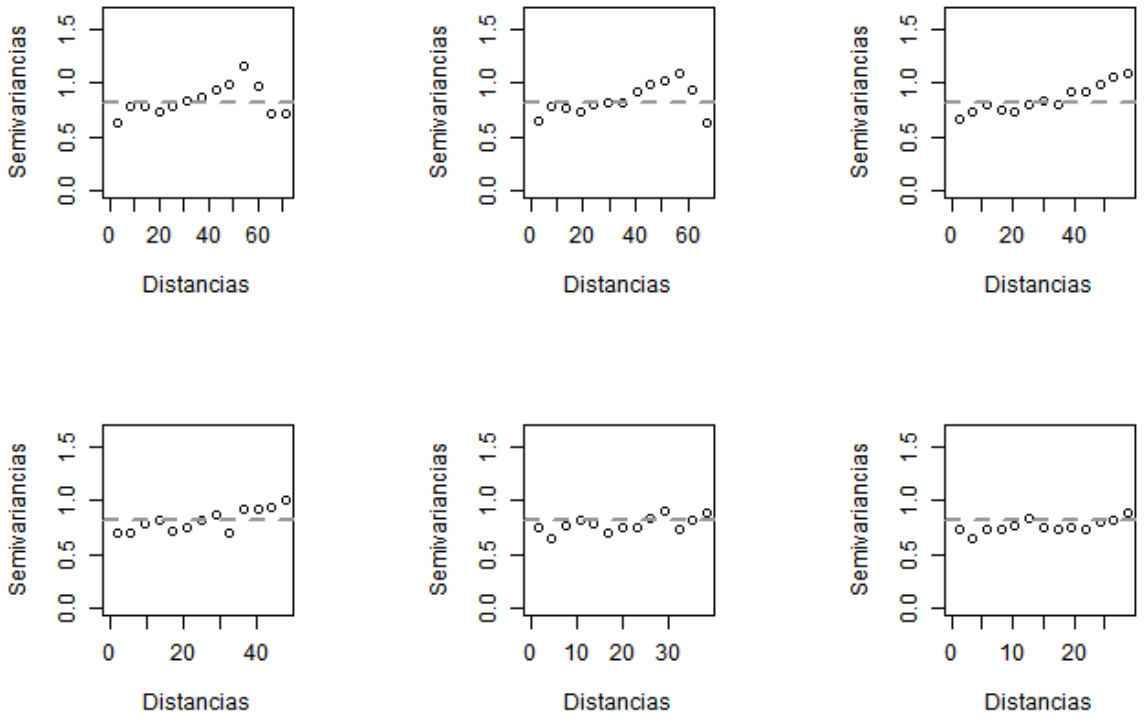
Potássio (K)



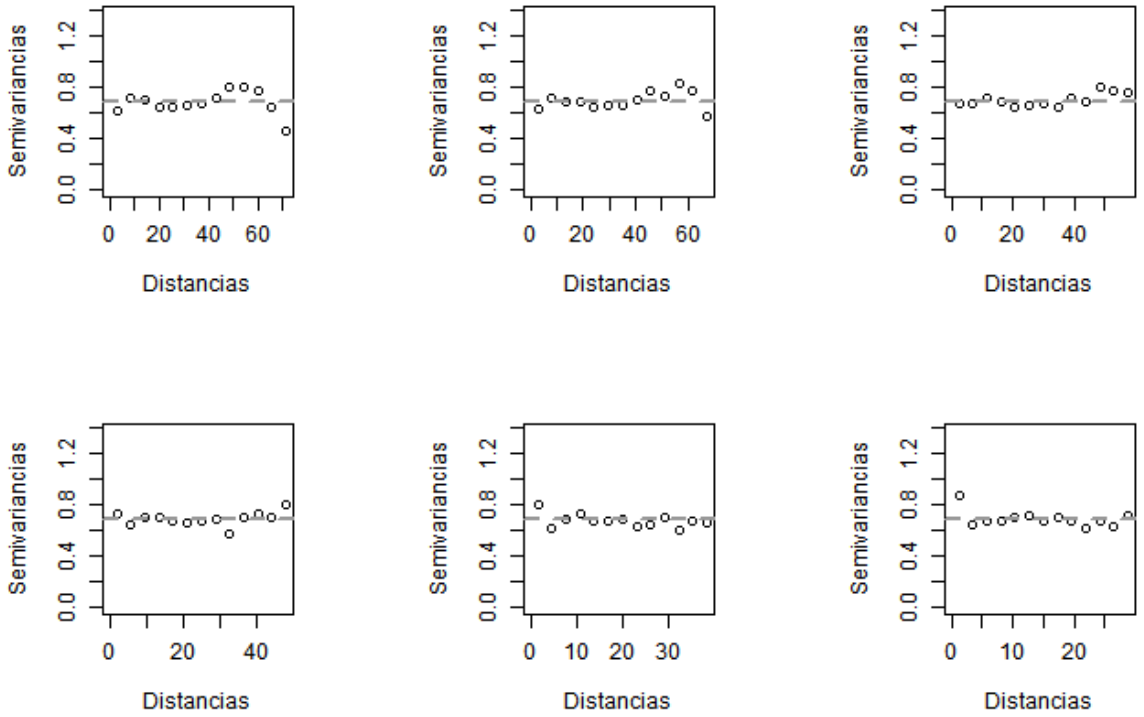
APÊNDICE D – Semivariogramas experimentais obtidos para os resíduos de brix e TCH em DIC e DBC.

Figura D-1 - Semivariogramas experimental omnidirecional com diferentes alcances para os resíduos de brix em DIC (A), em DBC (B) e para os resíduos de TCH em DIC (C) e em DBC (D).

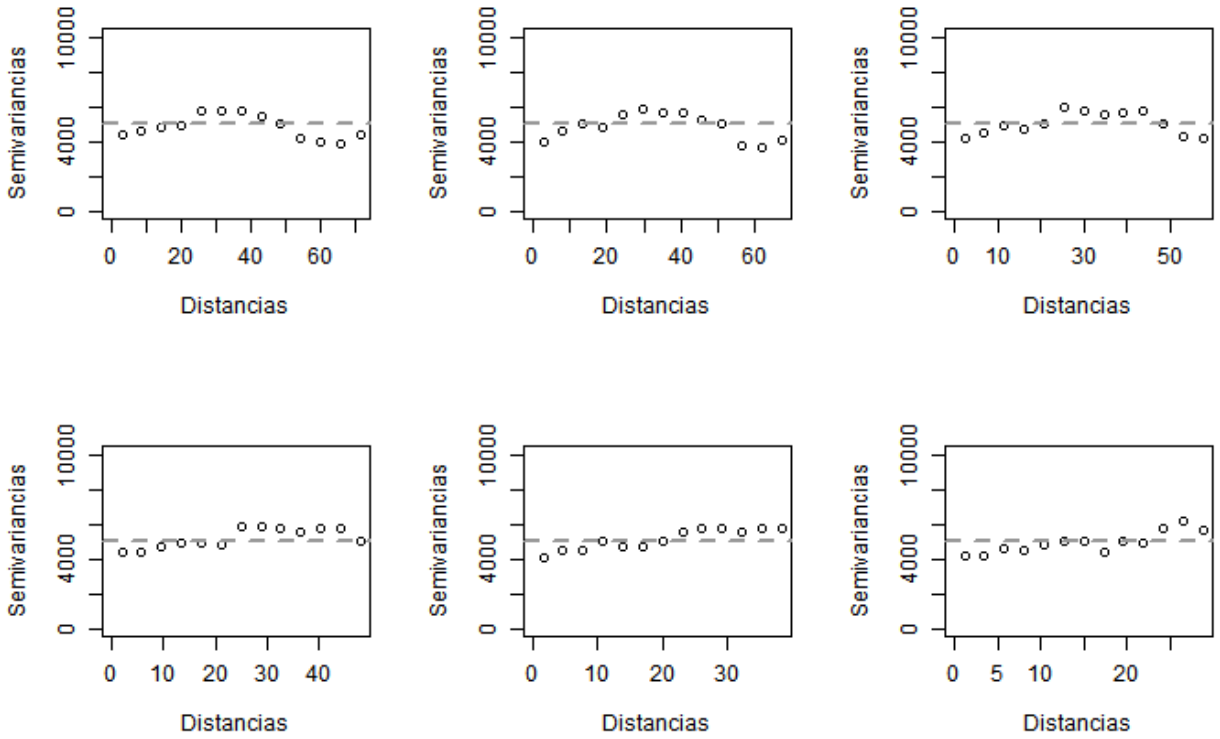
A



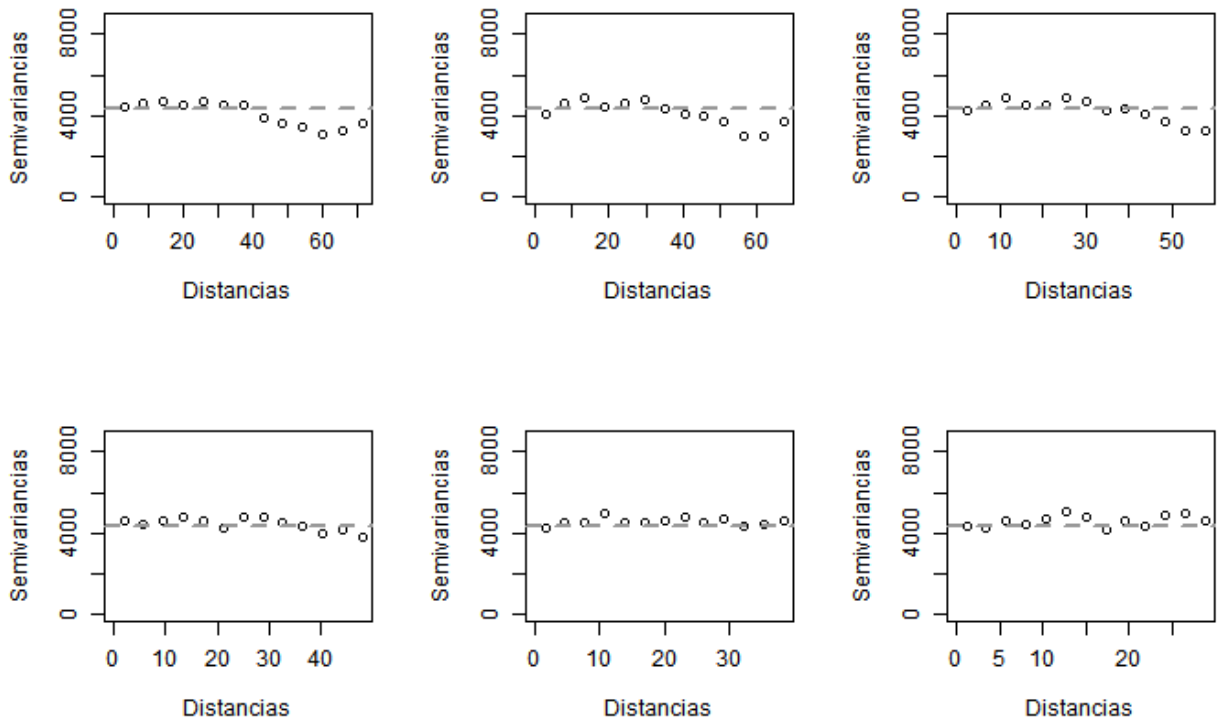
B


Semivariancias
0.0 0.4 0.8 1.2
0 20 40 60
Distancias
Semivariancias
0.0 0.4 0.8 1.2
0 20 40 60
Distancias
Semivariancias
0.0 0.4 0.8 1.2
0 20 40
Distancias
Semivariancias
0.0 0.4 0.8 1.2
0 20 40
Distancias
Semivariancias
0.0 0.4 0.8 1.2
0 10 20 30
Distancias
Semivariancias
0.0 0.4 0.8 1.2
0 10 20
Distancias

C


Semivariancias
0
4000
10000
0
20
40
60
Distancias
Semivariancias
0
4000
10000
0
20
40
60
Distancias
Semivariancias
0
4000
10000
0
10
30
50
Distancias
Semivariancias
0
4000
10000
0
10
20
30
40
Distancias
Semivariancias
0
4000
10000
0
10
20
30
Distancias
Semivariancias
0
4000
10000
0
5
10
20
Distancias

D


Semivariancias
0
4000
8000
0
20
40
60
Distancias
Semivariancias
0
4000
8000
0
20
40
60
Distancias
Semivariancias
0
4000
8000
0
10
30
50
Distancias
Semivariancias
0
4000
8000
0
10
20
30
40
Distancias
Semivariancias
0
4000
8000
0
10
20
30
Distancias
Semivariancias
0
4000
8000
0
5
10
20
Distancias