

Decision tree e geoestatística na redução do número de análises de micronutrientes do solo

Luciano G. Batista^{1†}, Samantha G. Oliveira¹, Nerilson T. Santos¹, Luiz Alexandre Peternelli¹,
Marcelo M. Costa², Matheus de P. Ferreira¹

¹ Universidade Federal de Viçosa; Centro de Ciências Exatas; Departamento de Estatística; Viçosa – Minas Gerais, Brasil.

² Universidade Federal de Jataí; Centro de Ciências Agrárias; Departamento de Agronomia; Jataí – Goiás, Brasil.

Resumo: Para realizar a interpolação por krigagem, é importante que cada ponto num semivariograma seja obtido com base no mínimo da combinação de 30 pares de pontos. Além disso, alguns autores alegam que é necessário ter pelo menos 100 amostras para fazer a interpolação. Sendo assim, o processo de amostragem se torna caro para o produtor rural. Como alternativa de contornar este problema de amostragem, foi utilizado metodologias de machine learning, em especial, decision tree. O objetivo principal deste trabalho foi avaliar o uso da metodologia de decision tree na redução do adensamento amostral para atributos do solo visando a realização da krigagem ordinária com tamanho amostral reduzido. Para isso, foram realizadas 50 amostragens pelo algoritmo Latin Hypercube Sampling (LHS), com malhas contendo 82, 112 e 127 pontos amostrados e os valores faltantes foram preditos com decision tree, até completar 150 pontos e logo em seguida foi realizado a krigagem ordinária para as malhas MR_{127} , MR_{112} e MR_{82} , que foi gerado pela combinação das 50 predições por decision tree e avaliados os valores da Raiz Quadrada do Erro Médio (RMSE) e Média do Erro Absoluto (MAE), denominados $RMSE_Krig$ e MAE_Krig . Foi percebido que há uma redução nestas estatísticas ao passo que reduz a quantidade de amostras originais. Ao fazer o mapa de atributos das malhas reduzidas é percebido que o padrão de concentração de micronutrientes dos solos nas malhas reduzidas segue semelhante ao padrão original, ou seja, zonas com maiores concentrações ainda continuam com concentrações elevadas e regiões com menores concentrações continuam com concentrações menores. Sendo assim, o uso das decision trees se mostraram eficientes em preservar o padrão de concentração dos micronutrientes do solo.

Palavras-chave: Adensamento amostral; Aprendizado estatístico; Krigagem ordinária.

Decision tree and geostatistics in reducing the number of soil micronutrient analyses

Abstract: To perform kriging interpolation, it is crucial that each point in the semivariogram is derived from a minimum combination of 30 point pairs. Additionally, some authors suggest that at least 100 samples are required for interpolation. As a result, the sampling process becomes expensive for farmers. To address this sampling issue, machine learning methodologies, specifically decision trees, were employed. The primary objective of this study was to evaluate the use of decision tree methodology to reduce sample density for soil attributes, with the goal of performing ordinary kriging using a reduced sample size. To achieve this, 50 samples were taken using the Latin Hypercube Sampling (LHS) algorithm, with grids containing 82, 112, and 127 sampled points. Missing values were predicted using decision trees until a total of 150 points was reached. Ordinary kriging was then performed for the MR_{127} , MR_{112} , and MR_{82} grids, which were generated by combining the 50 predictions from the decision tree. The Root Mean Square Error (RMSE) and Mean Absolute Error (MAE), referred to as $RMSE_Krig$ and MAE_Krig , were evaluated. It was observed that these statistics decreased as the number of original samples decreased. When mapping the attributes of the reduced grids, it was found that the pattern of micronutrient concentrations in the soils of the reduced grids closely resembled the original pattern. That is, areas with higher concentrations maintained high levels, while regions with lower concentrations continued to show lower values. Therefore, the use of decision trees proved to be effective in preserving the micronutrient concentration pattern in the soil.

Keywords: Sample density; Statistical learning; Ordinary kriging.

[†] Autor correspondente: luciano.batista@ufv.br

Manuscrito recebido em: 20/10/2024

Manuscrito revisado em: 08/01/2025

Manuscrito aceito em: 08/01/2025

Introdução

Estudar o padrão de variabilidade espacial de atributos químicos e físicos de propriedades do solo é tema importante para a agricultura de precisão, que consiste em aumentar a eficiência da propriedade rural, através do mapeamento e monitoramento da atividade agrícola (Reghini; Cavichioli, 2020). A agricultura de precisão desempenha um papel crucial no manejo sustentável da produção agrícola, contribuindo para a redução de impactos ambientais e a preservação dos recursos naturais, ao utilizar insumos na quantidade apropriada e nos locais adequados, minimizando os danos causados ao meio ambiente e prejudicando o mínimo possível as reservas naturais (De Oliveira; Schneider, 2024).

Para medir e analisar a variabilidade espacial das propriedades do solo, a geoestatística se destaca como uma ferramenta fundamental. A geoestatística é um ramo da estatística dedicada ao estudo de variáveis regionalizadas com dependência espacial, em que utiliza como um de seus principais instrumentos o semivariograma, uma função matemática que modela a continuidade espacial (Yamamoto; Landim, 2013).

Há algumas limitações quanto ao uso da geoestatística para prever a produção da lavoura dentre elas se destaca a necessidade de um grande número de pontos de amostragem. Mendes *et al.* (2020) e Cherubin *et al.* (2015) alertam que o uso de malhas que utilizam poucos pontos de amostragem deve ser evitado para não afetar a acurácia do poder preditivo do modelo. O que torna o processo de amostragem e análise do solo mais caro para o produtor (Mendes *et al.*, 2020).

Como alternativa para contornar esta dificuldade do grande número de amostras de solos, pode-se utilizar técnicas de *machine learning* para mapeamento do solo em atividades agrícolas (Mendes *et al.*, 2020). As técnicas de *machine learning* têm sido muito utilizadas, pois suas metodologias são capazes de lidar com problemas complexos e não lineares (Pantazi *et al.*, 2016; Tantalaki; Souravlas; Roumeliotis, 2019), que é uma das complicações de previsões de rendimento das culturas, que estão interligadas com a distribuição dos nutrientes no solo.

Alguns trabalhos recentes, como Pereira *et al.* (2022) e Martins *et al.* (2023), utilizaram técnicas de *machine learning* como método de interpolação e comparado com a krigagem ordinária, que é um método de interpolação na geoestatística. Em suas pesquisas, a krigagem ordinária ainda continuou sobressaindo em comparação às técnicas usadas para interpolação com *machine learning*.

As pesquisas realizadas por Martins *et al.* (2023) e Pereira *et al.* (2022) têm mostrado que a krigagem ordinária ainda é um bom método de interpolação e que o *machine learning* é um grande aliado para redução do adensamento amostral. A justificativa para o desenvolvimento deste artigo baseia-se em combinar os dois métodos, pois até então há poucos trabalhos seguindo esta vertente.

Na análise de solos, os laboratórios fornecem ao produtor a medição de um conjunto padrão de atributos, cobrando um valor fixo por essa análise de rotina. A avaliação de atributos adicionais, porém, implica custos extras para cada nova análise solicitada. Nesse contexto, este artigo tem como objetivo avaliar a aplicação da técnica de *decision tree* como uma estratégia para reduzir o número de amostras adicionais exigidas, possibilitando a realização da krigagem ordinária com um tamanho amostral menor e, conseqüentemente, diminuindo os custos para o produtor.

Materiais e métodos

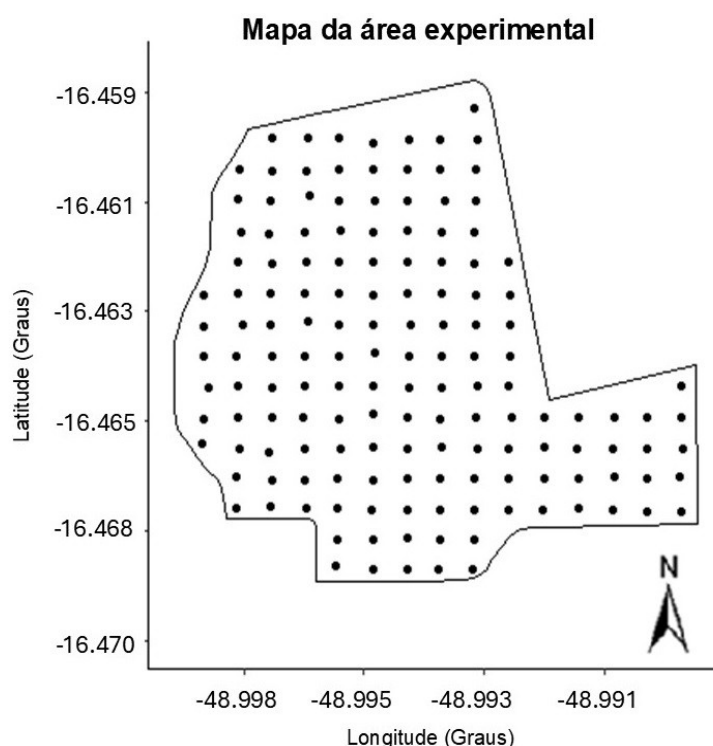
Banco de dados e área de estudos

Os dados utilizados neste estudo foram gentilmente fornecidos pelo professor Marcelo Marques Costa e incluem informações sobre os atributos químicos e físicos do solo. Esses dados

foram coletados em um talhão da fazenda "Andréia ou Sozinha", localizada em Goianópolis, no Centro-Oeste do Brasil. A região possui solo classificado como Latossolo Vermelho Amarelo.

Foram coletadas 150 amostras, dita "compostas", coletadas usando um trado de rosca com a profundidade de 0 a 0,20 m em uma área de 75 hectares. Cada uma dessas amostras compostas foi obtida a partir 10 subamostras do solo aleatoriamente localizadas em um raio de 5m em torno de cada um dos pontos de amostragem apresentados na Figura 1.

Figure 1: Map of the 75 ha experimental area containing the 150 sampling points



Source: from the authors (2024).

Para análise química do solo, foram mensurados os seguintes componentes: pH, acidez potencial (H + Al), teores de cálcio (Ca), magnésio (Mg), potássio (K), alumínio (Al), fósforo (P), fósforo remanescente (Prem), matéria orgânica, soma de bases (SB), capacidade de troca catiônica (CTC) efetiva (t) e CTC total, saturação por bases (V) e os teores de zinco (Zn), ferro (Fe), manganês (Mn) e cobre (Cu). Já, para análises físicas, foram mensurados os teores de argila, silte e areia total.

Quando realiza a análise de solos, os laboratórios entregam para o produtor a mensuração do teor de um conjunto de atributos que são considerados de rotina. O produtor paga por um valor fixo para a mensuração destes atributos. Caso o cliente queira analisar mais algum atributo, é cobrado separadamente para cada atributo.

Os atributos de rotina são: pH-H₂O, P, K, Ca, Mg, Al, H+Al, P-rem, SB, t, T e V. Para desenvolvimento deste artigo, foi utilizado os atributos de rotina para predizer o teor dos micronutrientes Fe, Zn, Mn e Cu, que também são atributos muito importantes para correção do solo.

Como SB, t, T e V são combinações de outros atributos, foi decidido não as utilizar. Sendo assim, foram utilizados apenas atributos que não são função de outros. Em resumo, foram utilizados os atributos pH-H₂O, P, K, Ca, Mg, H+Al e P-rem para predizer Fe, Zn, Mn e Cu.

Redução do adensamento amostral

A malha dita original é composta pelos 150 pontos amostrados. Neste estudo foram avaliadas as reduções da densidade da malha em 15%, 25% e 45% para os micronutrientes. O que correspondem as malhas reduzidas contendo 127, 112 e 82 pontos originais amostrados. Daqui em diante essas malhas são identificadas como MR_{127} , MR_{112} e MR_{82} .

Estas malhas reduzidas foram obtidas a partir da seleção de pontos da malha original utilizando o algoritmo *Latin Hypercube Sampling (LHS)* proposto por Mickey et al. (1979). O *LHS* se baseia no conceito de um quadrado latino, em que cada linha e cada coluna contém exatamente uma amostra (Minasny; McBratney, 2006).

Para diminuir o efeito da variabilidade das predições foram geradas 50 malhas, para cada uma das reduções, ou seja, foram geradas as seguintes malhas MR_{ij} tal que $i=127, 112$ e 82 e $j=1, \dots, 50$.

Construção da decision tree

Tomando como referência a malha original, o método *decision tree* foi utilizado para prever os teores dos micronutrientes dos pontos faltantes de cada uma das grades reduzidas MR_{ij} . Vale ressaltar que pontos faltantes são aqueles pontos que foram removidos da malha original para gerar a malha reduzida MR_{ij} .

Por exemplo, a malha MR_{127} contém 127 pontos originais e 23 pontos preditos com *decision tree*, totalizando 150 amostras. Os atributos foram retirados preservando as coordenadas geográficas, o que facilitou preencher os atributos faltantes na localização correta.

O critério de parada das *decision trees*, diz o ponto em que a divisão de um nó deve ser parada. Para este estudo, foi utilizado como critério de parada até o nó ter pelo menos uma observação. Este critério adotado, certamente está criando árvores que tenha um ajuste excessivo (*overfitting*) no conjunto de treinamento, mas este fato não é tão preocupante, pois foi realizado a poda de todas as árvores e, como é sabido, a poda diminui a variabilidade da árvore, diminuindo assim o *overfitting* (Izbicki; Santos, 2020, James et al., 2013).

O parâmetro de corte utilizado foi o menor valor de CP (*complexity parameter*), que é baseado no *xerror* (erro cruzado). O erro cruzado foi determinado pela comparação das predições do modelo com os valores reais no conjunto de teste.

Para o treinamento do modelo da *decision tree* foram utilizadas as amostras selecionadas pelo algoritmo *LHS* e para o conjunto de teste foi utilizado o complementar do conjunto de treinamento.

Como o algoritmo *LHS* foi utilizado 50 vezes para cada redução de um atributo, então a *decision tree* foi implementada 50 vezes para todas as malhas reduzidas MR_{ij} , $i=127, 112$ e 82 e $j=1, \dots, 50$, para cada um dos micronutrientes. Sendo assim, foi criada uma planilha em formato .xlsx, para guardar os valores preditos por *decision tree* com os originais, oriundos das 50 amostragens feitas pelo *LHS*.

Interpolação por krigagem

Os modelos de semivariograma exponencial, esférico e gaussiano foram selecionados porque, de acordo com Soares (2000), esses modelos geralmente abrangem a maioria das situações de dispersão de fenômenos espaciais nas ciências da terra e ambientais. Os semivariogramas foram ajustados de acordo com o método dos mínimos quadrados, *ordinary least squares (OLS)*, tanto para as malhas originais quanto para as malhas reduzidas MR_{ij} , $i=127, 112$ e 82 e $j=1, \dots, 50$. O

semivariograma escolhido foi àquele que apresentou o menor valor de erro quadrático médio (MSE). Quando o método *OLS* indicava que o melhor modelo de semivariograma era o efeito pepita puro, o processo de ajuste do semivariograma foi interrompido.

O modelo de semivariograma ajustado foi utilizado para realizar a interpolação por krigagem ordinária para a média destas 50 amostragens que foram armazenadas na planilha .xlsx mencionada anteriormente. Essas interpolações geraram os mapas temáticos de cada uma das malhas MR_{127} , MR_{112} e MR_{82} para cada um dos micronutrientes.

Para a krigagem nas malhas originais e nas malhas reduzidas, foram adotados uma vizinhança de busca com um mínimo de 8 e um máximo de 10 pontos ao entorno do ponto que se quer prever. Os mapas interpolados foram gerados com uma resolução de 10.000 pixels dentro da área estudada.

A média da validação cruzada das interpolações da krigagem combinada com *decision tree*, foi utilizada para comparar com as estatísticas da validação cruzada da krigagem ordinária com a grade original. Essas comparações permitiram conhecer o quanto é possível reduzir o adensamento amostral para se obter mapas de interpolação associando a krigagem ordinária com *decision tree*.

Software utilizado e principais pacotes

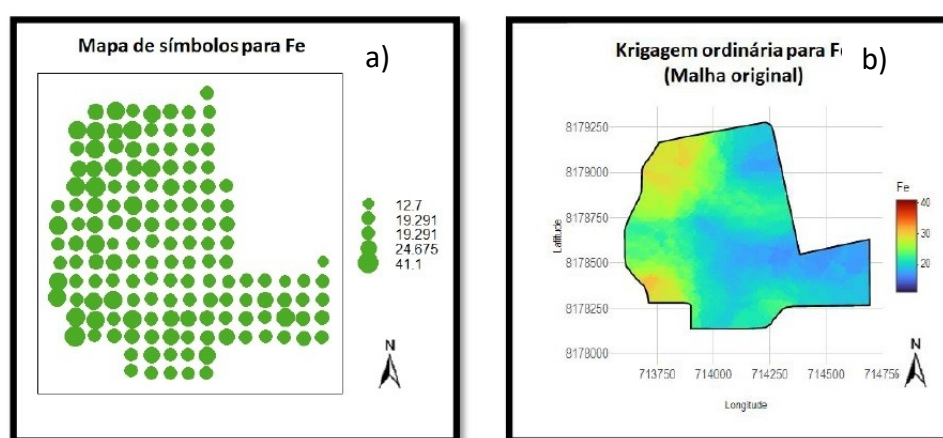
Para a análise dos dados, foi utilizado o software R (R DEVELOPMENT CORE TEAM, 2024) versão 4.3.1 e o pacote *geoR* package, versão 1.8.1, para fazer a interpolação por krigagem ordinária.

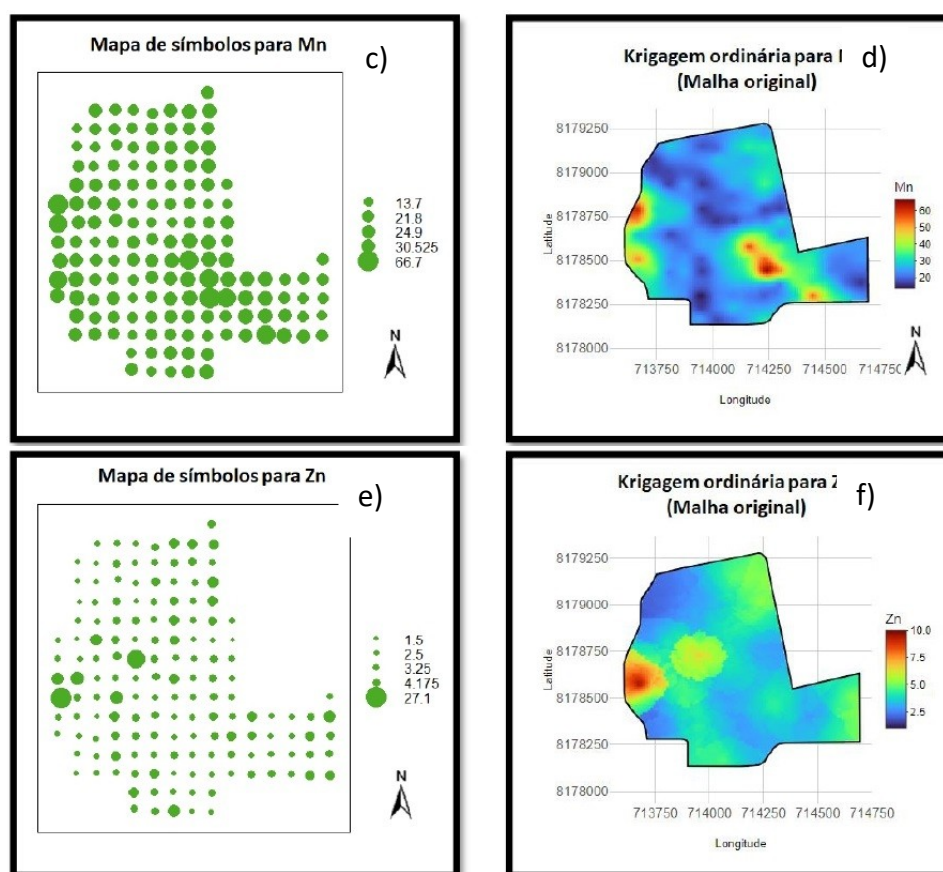
Resultados e discussões

Krigagem ordinária da malha original para os micronutrientes

Antes de iniciar o processo de redução do adensamento amostral para os micronutrientes, foi realizado a interpolação para a grade original destes atributos, no qual obtivemos os mapas apresentados na Figura 2.

Figure 2: Symbol maps and maps interpolated by ordinary kriging for the original micronutrient grid. Legend: (a) symbol map for Fe, (b) interpolated map for Fe, (c) symbol map for Mn, (d) interpolated map for Mn, (e) symbol map for Zn, (f) interpolated map for Zn.





Source: from the authors (2024).

A Figura 2 (a), (c) e (e), representam o mapa de símbolos dos atributos Fe, Mn e Zn, respectivamente, no qual as dimensões das bolhas indicam o seu grau de concentração. Quanto maior o diâmetro da bolha, maior a concentração do atributo no ponto e vice-versa.

Os parâmetros e os modelos de semivariograma que melhor se ajustaram aos dados da malha original são apresentados na Tabela 1.

Table 1: Parameters and semivariogram models for the original grid

Atributo	Semivariograma	Alcance (a)	Efeito pepita (C_0)	Patamar (C_0+C)	IDE (%)
Fe	Modelo esférico	249,47	5,15	30,03	83,75
Mn	Modelo esférico	187,11	19,93	93,94	78,78
Zn	Modelo esférico	343,02	4,36	7,82	44,25

Source: from the authors (2024).

De acordo com a Tabela 1 o modelo de semivariograma que melhor se ajustou aos atributos Fe, Mn e Zn foi o modelo esférico, o que corrobora com pesquisas de Vieira (2000), Corá *et al.* (2004), Souza *et al.* (2003), Grego; Vieira (2005), Cambardella *et al.* (1994), Boyer *et al.* (1996),

Albuquerque *et al.* (1996), Tsegaye; Hill, (1998) e Paz-González *et al.* (2000), que afirmam que este modelo é o de maior ocorrência para atributos químicos do solo.

O Índice de Dependência Espacial (IDE) apresentado na última coluna da Tabela 1, é medido por: $IDE = (\text{Patamar} - \text{Efeito pepita})/\text{Patamar}$. De acordo com Biondi *et al.* (1994), se $IDE < 0,25$ a dependência espacial é fraca, se $0,25 \leq IDE \leq 0,75$ a dependência espacial é moderada e caso $IDE \geq 0,75$ a dependência espacial é forte. Sendo assim, os atributos Fe e Mn apresentam dependência espacial forte e Zn apresenta dependência espacial moderada.

O alcance (*a*) indica a distância em que já não há dependência espacial, ou seja, para distâncias maiores que (*a*), não há correlação espacial (PYRCZ; DEUTSCH, 2014). De acordo com a Tabela 1, distâncias superiores à 249,47; 187,11 e 343,02 não correlação espacial para os atributos Fe, Mn e Zn, respectivamente.

Na Figura 2 (b), (d) e (f), são apresentados os mapas interpolados por krigagem ordinária para a malha original dos atributos Fe, Mn e Zn. Nestes mapas pode-se observar as zonas com maiores e menores concentrações de Fe, Mn e Zn e que estão em consonância com os mapas de símbolos para estes micronutrientes.

De acordo com a Figura 2 (b), é observado que nas regiões noroeste e sudoeste, estão as maiores concentrações de Fe, destacado nas cores vermelha e amarela. Para o atributo Mn (Figura 2 (d)), é observado que as maiores concentrações de Mn estão dispersas em algumas regiões da área em estudo, como pode ser observado nas cores vermelha da legenda do mapa. Além disso, também pode ser observado que nesta área, há pequenas concentrações de Mn em sua totalidade. Já para o Zn, observa-se que na região oeste estão as maiores concentrações e nas regiões nordeste, leste e na parte central, estão os de concentrações medianas.

Krigagem ordinária para as malhas reduzidas dos micronutrientes

Para a krigagem ordinária foram escolhidas as malhas MR_{127} , MR_{112} e MR_{82} que representam reduções de 15%, 25% e 45%, respectivamente, nos pontos amostrados. Optou-se por valores de redução em até 45%, pois quando a redução amostral era maior, alguns atributos não mostraram dependência espacial, indicado pela ocorrência de efeito pepita puro no ajuste do semivariograma (Isaaks; Srivastava, 1989).

Os mapas interpolados por krigagem ordinária para os atributos Fe, Mn e Zn e para as malhas MR_{127} , MR_{112} e MR_{82} são apresentados nas Figuras 3, 4 e 5.

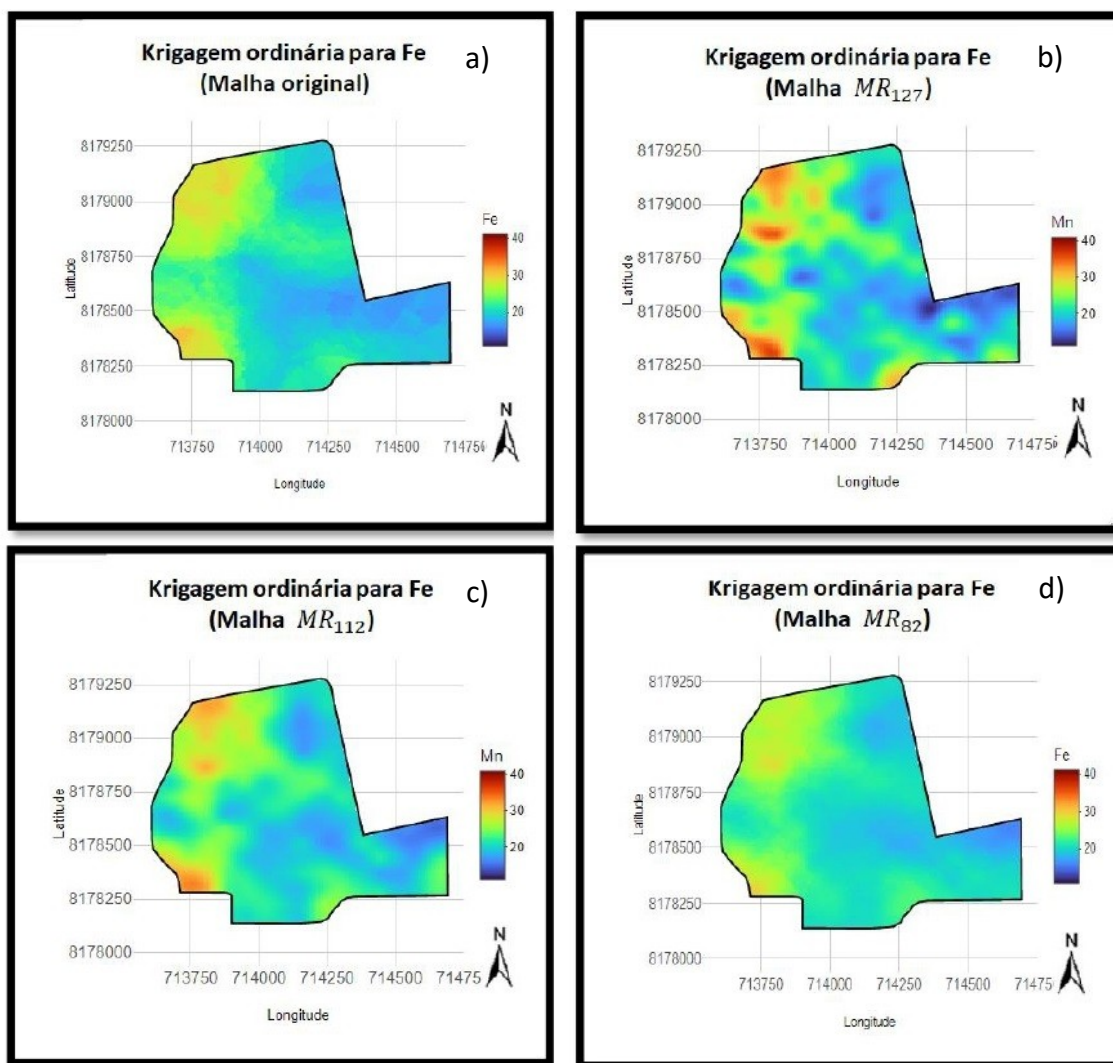
Observando as Figuras 3, 4 e 5, percebe-se que, para as malhas MR_{127} , MR_{112} e MR_{82} , os padrões de concentração dos micronutrientes Fe, Mn e Zn permanecem semelhantes ao padrão visto na malha original. Regiões com maiores concentrações ainda mantêm níveis elevados, enquanto aquelas com menores concentrações continuam a apresentar índices reduzidos.

Para a redução de 45%, malha MR_{82} (Figura 5 (d)) percebe-se que a concentração de Zn na parte central, mudou bruscamente em comparativo com a malha original. Alguns pontos, ainda permaneceram com concentrações altas, mas a área da concentração deste atributo não permaneceu a mesma.

É sabido que numa *decision tree*, os atributos que caem em uma folha recebem valores iguais para todas as observações (Izbicki; Santos, 2020; James *et al.*, 2013). Por exemplo, se uma *decision tree*, após a poda, tiver uma única folha para uma malha MR_{82} , então 68 amostras foram preditas com valores iguais. Com esses 68 valores iguais, o conjunto de dados tem pouca variabilidade, o que o torna mais homogêneo. Ao realizar a krigagem ordinária nas malhas reduzidas há uma suavização nas manchas das concentrações do atributo, como pode ser observado

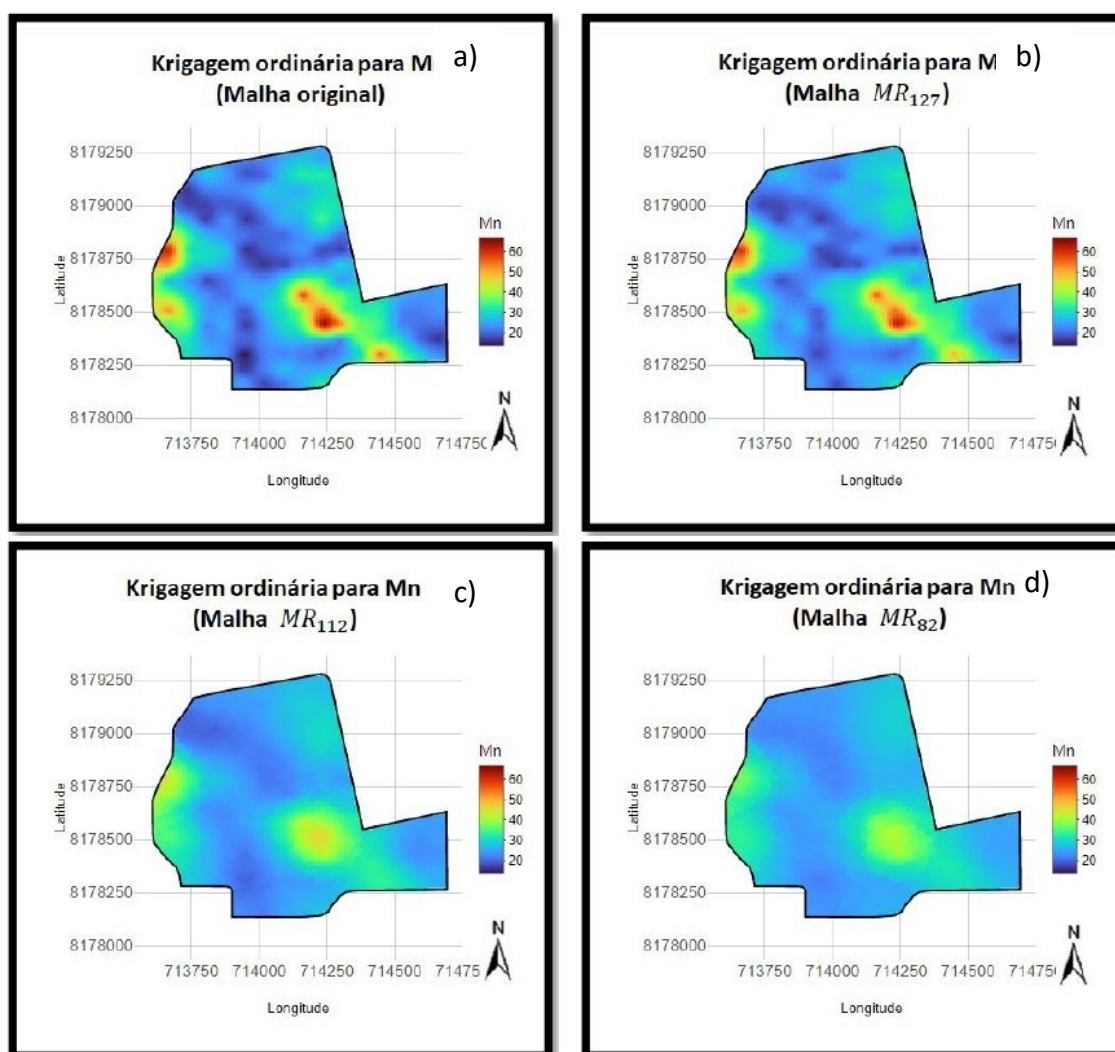
nas Figuras 3, 4 e 5, devido as malhas reduzidas conterem pontos com valores muito semelhantes entre si.

Figure 3: a) Interpolated map of the Fe attribute for the original grid. (b) Interpolated map of Fe for the MR_{127} grid. (c) Interpolated map of Fe for the MR_{112} grid. (d) Interpolated map of Fe for the MR_{82} grid.



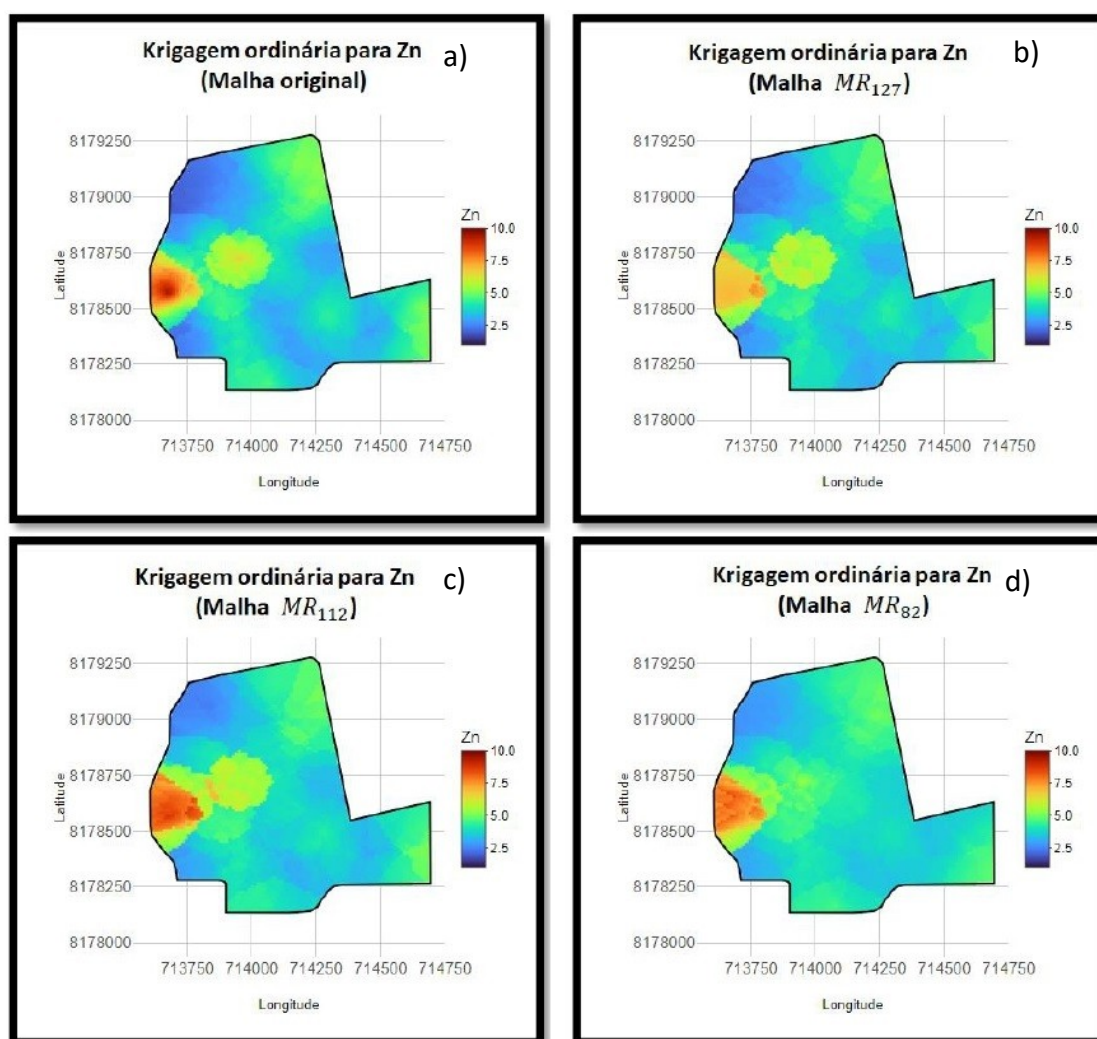
Source: from the authors (2024).

Figure 4: a) Interpolated map of the Mn attribute for the original grid. (b) Interpolated map of Mn for the MR_{127} grid. (c) Interpolated map of Mn for the MR_{112} grid. (d) Interpolated map of Mn for the MR_{82} grid.



Source: from authors (2023).

Figure 5: a) Interpolated map of the Zn attribute for the original grid. (b) Interpolated map of Zn for the MR_{127} grid. (c) Interpolated map of Zn for the MR_{112} grid. (d) Interpolated map of Zn for the MR_{82} grid.



Source: from the authors (2024).

Recentes estudos como os de Pereira *et al.* (2022) e Martins *et al.* (2023), têm explorado o emprego de técnicas de *machine learning* como um método de interpolação utilizando todo o conjunto de dados e comparado com krigagem ordinária. Segundo eles, a krigagem ordinária ainda demonstrou superioridade em comparativo as técnicas de *machine learning*. Este trabalho segue uma vertente diferente, que é utilizar de forma mista as duas metodologias, *machine learning* e krigagem ordinária, para aproveitar um pouco de cada método e não usar as técnicas de *machine learning* como substituto da krigagem ordinária.

Estudos como os de Yang *et al.* (2020) e Qu *et al.* (2024) avaliaram o desempenho de algumas metodologias estatísticas para mapeamento do solo, inclusive *machine learning*, sob diferentes condições amostrais, e concluíram que ainda são necessários uma grande quantidade de amostras para se obter predições mais precisas sobre a distribuição de micronutrientes dos solos.

A Tabela 2 exibe os valores da Raiz Quadrada do Erro Médio (RMSE) e Média do Erro Absoluto (MAE) da krigagem ordinária para o conjunto de dados original e os valores de RMSE e MAE (RMSE_Krig e MAE_Krig) das malhas MR_{127} , MR_{112} e MR_{82} para os atributos Fe, Mn e Zn.

Table 2: Table with the cross-validation statistics, RMSE and MAE, for ordinary kriging on the original grid and reduced grids for the Fe, Mn, and Zn attributes.

Atributo	Malhas	RMSE_Krig	MAE_Krig
Fe	Original	4,24	3,13
	MR_{127}	6,66	5,26
	MR_{112}	6,57	5,15
	MR_{82}	6,07	4,59
Mn	Original	5,90	4,09
	MR_{127}	10,92	7,91
	MR_{112}	9,50	6,64
	MR_{82}	8,37	5,43
Zn	Original	3,00	1,88
	MR_{127}	3,72	1,99
	MR_{112}	3,38	1,94
	MR_{82}	2,98	1,58

Legend: RMSE_Krig: Root Mean Square Error for the MR_{127} , MR_{112} , and MR_{82} grids; MAE_Krig: Mean Absolute Error for the MR_{127} , MR_{112} , and MR_{82} grids.

Source: from the authors (2024).

Ao observar a Tabela 2, nota-se também que ao passo que aumenta a quantidade de amostras preditas com *decision tree*, os valores de RMSE_Krig e MAE_Krig diminuem. De acordo com El-Sayed Ewis (2012), quando estamos escolhendo o melhor método estatístico, optamos por aqueles que tem o menor RMSE ou o menor MAE. Com isso, a redução das estatísticas de validação, RMSE_Krig e MAE_Krig, indica estar havendo uma melhoria nas predições. Precisamos ter cautela quanto a interpretação destas estatísticas, pois as *decision trees* geram valores iguais para todas as observações que caem na mesma folha e isto pode estar afetando a interpolação na hora da krigagem ordinária.

James et al. (2013) e Izbicki; Santos (2020), afirmam que as árvores de decisão são de fácil interpretação. Eles alertam também que as árvores de decisão têm algumas limitações, como propensão ao *overfitting* em conjuntos de dados pequenos e sensibilidade a pequenas variações nos

dados de entrada. Apesar dessas limitações, sua versatilidade e facilidade de interpretação continuam a torná-las uma escolha popular em muitas aplicações de *machine learning* (James et al., 2013).

Conclusão

Ao fazer o mapa de atributos das malhas reduzidas é percebido que o padrão de concentração de micronutrientes dos solos nas malhas reduzidas segue semelhante ao padrão original, ou seja, zonas com maiores concentrações ainda continuam com concentrações elevadas e regiões com menores concentrações continuam com concentrações menores. Sendo assim, o uso das *decision trees* se mostraram eficientes em preservar o padrão de concentração dos micronutrientes do solo.

Referências

ALBUQUERQUE, J. A.; REINERT, D. J.; FIORIN, J. E. **Variabilidade de solo e planta em Podzólico Vermelho-Amarelo**. *Revista Brasileira de Ciência do Solo*, v. 20, n. 1, p. 151-157, 1996.

BOYER, D. G.; WRIGHT, R. J.; FELDHAKE, C. M.; BLIGH, D. P. **Relações de variabilidade espacial do solo em um ambiente de solo ácido de declive acentuado**. *Soil Science*, v. 161, n. 5, p. 278-287, 1996.

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

CHERUBIN, M. R.; SANTI, A. L.; EITELWEIN, M. T.; AMADO, T. J. C.; SIMON, D. H.; DAMIAN, J. M. **Dimensão da malha amostral para caracterização da variabilidade espacial de fósforo e potássio em Latossolo Vermelho**. *Pesquisa Agropecuária Brasileira*, v. 50, p. 182-177, 2015.

CORÁ, J. E.; FERRARESE, C. P.; OLIVEIRA, F. S.; CORÁ, J. A.; BORGES, J. C.; LIMA, J. R.; HERRMANN, R.; CRUZ, L. M. **Variabilidade espacial de atributos do solo para adoção do sistema de agricultura de precisão na cultura de cana-de-açúcar**. *Revista Brasileira de Ciência do Solo*, Viçosa, v. 28, n. 6, p. 1013-1021, 2004. Disponível em: <https://doi.org/10.1590/S0100-08232004000600010>. Acesso em: 01 nov. 2023.

DAHIKAR, S. S.; RODE, D. S. V. **Agricultural crop yield prediction using artificial neural network approach**. *International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering*, v. 2, n. 1, p. 823-826, 2014. Disponível em: <https://www.ijireeice.com/upload/2014/ijireeice-2844.pdf>. Acesso em: 23 out. 2023.

DE OLIVEIRA, S.; SCHNEIDER, PW. **Viabilidade econômico-financeira da agricultura de precisão no cultivo do tabaco em uma propriedade rural localizada em sobradinho/rs**. *revista de contabilidade dom alberto*, v. 13, n. 25, p. 147-176, 2024.

- EL-SAYED EWIS, O. **Improving the prediction accuracy of soil mapping through geostatistics.** *International Journal of Geosciences*, v. 2012, 2012.
- FATORGIS. **Agricultura de precisão: A tecnologia de GIS/GPS chega às fazendas.** Curitiba, 1998. Disponível em: <http://www.fatorgis.com>. Acesso em: 21 out. 2023.
- GREGO, C.R.; VIEIRA, S.R. **Variabilidade espacial de propriedades físicas do solo em uma parcela experimental.** *Revista Brasileira de Ciência do Solo*, Viçosa, v. 29, n. 2, p. 169-177, 2005. Disponível em: <https://www.scielo.br/j/rbcs/article/view/32799>. Acesso em: 01 nov. 2023. doi: 10.1590/S0100-08232005000200002.
- IZBICKI, R.; DOS SANTOS, T. M. **Aprendizado de máquina: uma abordagem estatística.** Rafael Izbicki, 2020.
- ISAACS, E. H.; SRIVASTAVA, R. M. **Applied geostatistics.** New York: Oxford University Press, 1989.
- JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R.; TAYLOR, J. **An introduction to statistical learning: With applications in R.** Springer Nature, 2013.
- MANZATTO, C. V.; BHERING, S. B.; SIMÕES, M. **Agricultura de precisão: propostas e ações da Embrapa Solos.** EMBRAPA Solos, 1999. Disponível em: <http://www.cnps.embrapa.br/search/pesqs/proj01/proj01.html>. Acesso em: 02 fev. 2023.
- MENDES, W. S.; DEMATTÊ, J. A. M.; BARROS, A. S.; SALAZAR, D. F. U.; AMORIM, M. T. **A. Geostatistics or machine learning for mapping soil attributes and agricultural practices.** *Revista Ceres*, v. 67, n. 4, p. 330–336, 2020. Disponível em: <https://doi.org/10.1590/0034-737X202067040010>. Acesso em: 29 jul. 2023.
- OLIVER, M.A. **Exploring soil spatial variation geostatistically.** In: EUROPEAN CONFERENCE ON PRECISION AGRICULTURE, 2, 1999. Odense. Proceedings. Silsoe: Sheffield, 1999. p. 03-18.
- PANTAZI, X. E.; MOSHOU, D.; ALEXANDRIDIS, T.; WHETTON, R. L.; MOUAZEN, A. M. **Wheat yield prediction using machine learning and advanced sensing techniques.** *Computers and Electronics in Agriculture*, v. 121, p. 57–65, 2016.
- PAZ-GONZÁLEZ, A.; SÁNCHEZ, J. S.; FRUTOS, M. A.; MORENO, J. A.; ALBA, L. **The effect of cultivation on the spatial variability of selected properties of an umbric horizon.** *Geoderma*, v. 97, p. 272–292, 2000.
- PEREIRA, G. W.; VALENTE, D. S. M.; DE QUEIROZ, D. M.; SANTOS, N. T.; FERNANDES-FILHO, E. I. **Soil mapping for precision agriculture using support vector machines combined with inverse distance weighting.** *Precision Agriculture*, v. 23, p. 1189-1204, 2022. Disponível em: <https://doi.org/10.1007/s11119-022-09880-9>. Acesso em: 01 nov. 2023.
- PYRCZ, M.; DEUTSCH, C. **Geostatistical reservoir modeling.** 2. ed. New York, NY: Oxford University Press, 2014.

QU, L.; LU, H.; TIAN, Z.; SCHOORL, J. M.; HUANG, B.; LIANG, Y. **Spatial prediction of soil sand content at various sampling density based on geostatistical and machine learning algorithms in plain areas.** *Catena*, v. 234, p. 107572, 2024.

REGHINI, F.; L; CAVICHIOLI, Fábio Alexandre. Utilização de geoprocessamento na agricultura de precisão. **Revista Interface Tecnológica**, v. 17, n. 1, p. 329-339, 2020.

SOARES, A. **Geoestatística para as Ciências da Terra e do Ambiente.** Lisboa: IST Press, 2000.

SOUZA, C. K.; MARQUES JÚNIOR, J.; MARTINS FILHO, M. V.; PEREIRA, G. T. **Influência do relevo na variação anisotrópica dos atributos químicos e granulométricos de um Latossolo em Jaboticabal, SP.** *Engenharia Agrícola*, v. 23, n. 3, p. 486-495, 2003.

TANTALAKI, N.; SOURAVLAS, S.; ROUMELIOTIS, M. **Data-driven decision making in precision agriculture: The rise of big data in agricultural systems.** *Journal of Agricultural & Food Information*, v. 20, n. 4, p. 344-380, 2019.

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

YAMAMOTO, J. K.; LANDIM, P. M. B. **Geoestatística: conceitos e aplicações.** São Paulo: Oficina de Textos, 2013. Disponível em: <https://www.ofitexto.com.br>. Acesso em: 24 fev. 2023.

YANG, L.; LI, X.; SHI, J.; SHEN, F.; QI, F.; GAO, B.; ZHOU, C. **Evaluation of conditioned Latin hypercube sampling for soil mapping based on a machine learning method.** *Geoderma*, v. 369, p. 114337, 2020.