

**SENSORES PROXIMAIS PARA ESTIMAÇÃO TEOR DE ARGILA DO SOLO
UTILIZANDO ALGORITMOS DE APRENDIZADO DE MÁQUINA****PROXIMAL SENSORS FOR ESTIMATING SOIL CLAY CONTENT USING
MACHINE LEARNING ALGORITHMS****SENSORES PROXIMALES PARA ESTIMAR EL CONTENIDO DE ARCILLA DEL
SUELO UTILIZANDO ALGORITMOS DE APRENDIZAJE AUTOMÁTICO**

Apresentação: Comunicação Oral

João de Deus Ferreira e Silva¹; José Marques Júnior²; Kathleen Fernandes³ Denise de Lima Dias Delarica⁴; Mara Regina Moitinho⁵

RESUMO

O manejo sustentável do solo nos dias atuais, necessita do entendimento da variação espacial das características intrínsecas ao mesmo. O presente estudo objetivou-se encontrar relações um sensor e a concentração de argila, possibilitando a estimativa e o mapeamento para consolidação de áreas de manejo específico. A área de utilizada para este estudo possui cerca de um total de 45% do estado de São Paulo (13 milhões de hectares). Foi utilizado um total de 226 amostras de solo para estudo da variabilidade espacial. Foram realizadas análises físicas (areia, silte e argila). Realizou a estatística descritiva e a estimativa dos atributos foi realizada por meio de algoritmos de aprendizado de máquina (Cubista e XGBoost). A variabilidade espacial foi realizada por interpolação usando krigagem ordinária. Foi observado que o material de origem basalto e arenito influenciaram significativamente nos atributos físicos. Foi observado que a sensor de fluorescência de raios-X apresentou os melhores modelos de estimativa. Em comparação aos modelos utilizados, pode-se notar que apresentaram valores de precisão e acurácia semelhantes os modelos. Os mapas de variabilidade espacial obtidos para os atributos observados e preditos para os algoritmos de aprendizado de máquina foram similares, indicando que o sensor neste estudo, pode ser utilizado para caracterização espacial dos atributos do solo. Conclui-se que, as medidas do sensor espectral associado aos algoritmos de aprendizado de máquina foram capazes de estimar com precisão satisfatória a teor de argila do solo, possibilitando o conhecimento do comportamento espacial, demonstrando o potencial dessa tecnologia proximal para a ciência do solo.

Palavras-chave: granulometria, magnetismo, espectros do solo, inteligência artificial, pedometria.

INTRODUÇÃO

O manejo sustentável do solo vai de encontro com o entendimento da variação espaço-temporal das características intrínsecas ao mesmo (George et al., 2024). As informações que

1 Doutorado em Agronomia (Produção Vegetal), UNESP- Jaboticabal, joaodeus.caf@gmail.com

2 Professor UNESP Jaboticabal, joaodeus.caf@gmail.com

3 Doutorado em Agronomia (Produção Vegetal), UNESP- Jaboticabal, joaodeus.caf@gmail.com

4 Doutorado em Agronomia (Produção Vegetal), UNESP- Jaboticabal, joaodeus.caf@gmail.com

5 Doutorado em Agronomia (Produção Vegetal), UNESP- Jaboticabal, joaodeus.caf@gmail.com

podem ser adquiridas sobre o solo são relevantes para implementação de políticas públicas de seu uso e manejo, visando alcançar níveis seguros da segurança alimentar, como proposto nos objetivos de desenvolvimento sustentável (ODS) (Filla et al., 2021).

Dentro os atributos estudados, a granulometria do solo se refere às proporções relativas de partículas de areia, silte e argila em um volume conhecido. Tal atributo pode impactar de formas diferentes no crescimento vegetativo, produção das culturas de importância agrícola, na infiltração e retenção da água e absorção de nutrientes pelas plantas (Zhao et al., 2022). Assim, o mapeamento desse atributo nos solos agrícolas é relevante para o auxílio na gestão apropriada. No entanto, os métodos tradicionais de análise da granulometria do solo são caracterizados como métodos que exigem alto gasto financeiro, são demorados e podem gerar reagentes laboratoriais que necessitam de descarte adequado (Matus et al., 2021).

Nos dias atuais, o desenvolvimento de técnicas alternativas com uso de sensores proximais tem sido um dos campos de investigação ascendente no ramo da ciência do solo (Peng et al., 2020; Padarian et al., 2020). Dentre essas técnicas, temos fluorescência de raios-X (FRX) que é uma técnica pedométrica considerada econômica, rápida, não destrutiva e ambientalmente sustentável (Yousefi et al., 2018). Os espectros obtidos pelo FRx, podem ser usados para estimar uma ampla gama de atributos do solo (Rossel et al., 2022).

A FRx possui potencial para analisar elementos químicos em muitos campos da ciência, tais como na geologia, ciências ambientais e áreas de explorações de minério (Tian et al., 2018; Yao et al., 2021). A FRx tem permitido avaliações multielementares do solo com resultados comparáveis aos obtidos em laboratório para diversos elementos presentes em sua composição (Weindorf; Chakraborty, 2020). A FRx é um sensor emergente, que ainda necessita de mais investigações acerca da relação das suas medidas com os atributos granulométricos do solo (Lemière, 2018).

A utilização dos métodos tradicionais de avaliação das frações granulométricas dificulta os estudos quando o foco é o monitoramento de grandes áreas de estudo (Liu et al., 2018). No entanto, esse impedimento pode ser contornado pelo progresso das técnicas pedométricas e uso dos sensores proximais, que podem ser utilizados para fornecer um grande volume de dados sobre uma área estudada, auxiliando na proteção do solo, a exemplo de estimativa de atributos cruciais como a concentração de argila, silte e areia (Hu et al., 2018; Chen et al., 2016).

Junto ao aumento da utilização de sensores, que geram dados de alta dimensão, nos últimos anos, as técnicas de aprendizado de máquina tornaram-se populares para a modelagem

desses dados, pois conseguem extrair padrões de dados não lineares (Behrens et al., 2022, Wadoux et al., 2022). Essas técnicas aplicadas atualmente têm demonstrado ser mais precisas que os modelos tradicionais, podendo realizar boas estimativas dos atributos do solo estudados (Ramirez et al. 2021). Portanto, os algoritmos têm sido utilizados na estimativa dos atributos físicos (granulométricos) por conseguir aumentar a acurácia dos modelos nas estimativas realizadas.

Por fim, cabe ressaltar que as pesquisas acerca da variabilidade dos atributos granulométricos em solos das regiões tropicais em áreas de grande dimensão como o Planalto Ocidental Paulista ainda são escassas. Surge com isso, a necessidade de pesquisas que abordem esse tema para entender quais os fatores que influenciam o comportamento desse atributo nos solos tropicais, visando gerar informações úteis por meio da delimitação de zonas específicas de manejo, tendo como intuito alcançar estratégias para manter a segurança alimentar e a produtividade agrícola regional. Assim, por meio do presente estudo temos como objetivos: avaliar as relações existentes entre as medidas obtidas pelo sensor FRx com a concentração de argila nos solos Planalto Ocidental Paulista, estimar a argila por meio da FRx nos solos, mapear áreas de manejo específico para a concentração de argila a fim de delimitar regiões com maior e menor concentração.

FUNDAMENTAÇÃO TEÓRICA

Um dos grandes entraves para o conhecimento detalhado nos estudos do solo desde o primórdio, são as etapas de coleta de amostras e análises laboratoriais, que trata-se de uma etapa árdua e onerosa (Mcbratney et al., 2002). Porém, nos últimos anos com o avanço tecnológico mundial, mais precisamente com o surgimento de sensores aplicados à ciência do solo, o conhecimento detalhado dos atributos do solo tem aumentado nos últimos anos (Marques et al., 2014). Esses sensores atuam como auxiliares aos métodos tradicionais, possuindo a vantagem de serem técnicas de rápida obtenção de resultados, mais baratos que as medidas tradicionais, além da simplicidade pela não destruição das amostras (Viscarra Rossel et al., 2009).

A suscetibilidade magnética do solo surge como exemplo de método alternativo aplicado para fins de estudos na ciência do solo, a qual é caracterizada como o grau de magnetização de um material em resposta a um campo magnético aplicado (Siqueira et al., 2010). A suscetibilidade é produzida por indução do campo magnético terrestre durante a cristalização do mineral presente na rocha ou no solo, e se origina das propriedades de rotação dos elétrons (Dearing et al., 1995).

São considerados 5 tipos básicos de comportamento magnético: diamagnetismo,

paramagnetismo, ferromagnetismo, ferrimagnetismo e antiferromagnetismo. Nos minerais diamagnéticos, os átomos que compõem a cela unitária do mineral, possuem as camadas eletrônicas preenchidas. Assim, o número de “spins” eletrônicos alinhados numa direção é igual ao número de “spins” eletrônicos na direção oposta (exemplo: quartzo e feldspato). Nos minerais paramagnéticos as camadas eletrônicas estão incompletas. A presença de um campo magnético externo faz com que os “spins” se alinhem, e mesmo após a retirada do campo magnético, alguns “spins” permanecem alinhados (exemplo: Olivina, piroxênio, granada, biotita e carbonatos de ferro e manganês).

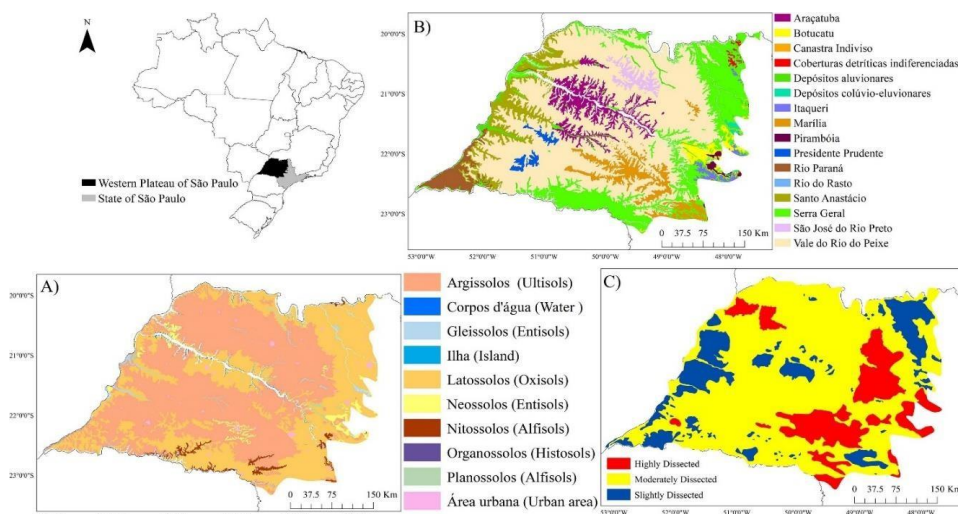
Os minerais ferromagnéticos são um caso especial de paramagnetismo. Após a retirada do campo magnético os “spins” permanecem alinhados, fazendo com que o mineral possua um grande valor de magnetização remanescente (exemplo: ferro e cobalto). Nos minerais ferrimagnéticos, os “spins” não estão emparelhados, assim prevalece o momento magnético do maior número de “spins” no mesmo sentido (exemplo: magnetita). Os minerais antiferromagnéticos não apresentam propriedades magnéticas (Thompson et al., 1986).

Atualmente, um dos campos da ciência de solo com forte aplicação das propriedades magnéticas tem sido investigações acerca dos atributos do solo (Karimi et al., 2011). Estudos incipientes indicam a existência da relação entre a suscetibilidade magnética e a granulometria do solo. A aplicabilidade da suscetibilidade magnética em investigações com atributos do solo é dada pela origem semelhante dos metais e minerais que conferem o magnetismo ao solo, sendo liberados no material de origem subjacente através do intemperismo e dos processos pedogenéticos (Wang et al., 2013).

METODOLOGIA

O estudo está desenvolvido no Planalto Ocidental Paulista (POP), que corresponde a aproximadamente 13 milhões de hectares (~ 48% do Estado de São Paulo) (Figura 1). De acordo com a classificação de Thornthwaite (1948) o clima é tropical com invernos secos (C2rA'a') no norte e noroeste; temperado úmido com verões quentes (B4rB4a) no Sul e temperado úmido com invernos secos e verões quentes (B2rB3 'a) no leste e sudeste. A vegetação natural é composta por espécies da Mata Atlântica a oeste e cerrados a leste e sudeste.

Figura 1: a) Mapa pedológico do Planalto Ocidental Paulista. (Fonte: Instituto Agrônomo de Campinas: Entre parênteses equivalência ao soil taxonomy); b) Mapa geológico do Planalto Ocidental Paulista (Fernandes et al., 2007, Fernandes, 2004); c) Mapa de dissecação do Planalto Ocidental Paulista.



Fonte: Própria (2024)

O esboço geológico é caracterizado principalmente por sedimentos arenosos, argilosos, cascalho, rochas vulcânicas de composição básica e sequências sedimentares, principalmente psamíticas, podendo incluir piroclásticas (Figura 1B). Dos 13 milhões de hectares de solos do Planalto, aproximadamente 2 milhões hectares são ocupados por basalto da formação Serra Geral (15,5%), 7,4 de milhões hectares pela Formação Vale do Rio do Peixe (57,1%) e 3,6 milhões de hectares pelas outras formações sedimentares (27,5%).

As formações geológicas do POP são divididas em dois Grupos: (I) Grupo Caiuá composto pelas formações Santo Anastácio e Rio Paraná no estado de São Paulo, correspondentes a depósitos de lençóis de areia, de clima seco, acumulados em extensas e monótonas planícies desérticas, marginais com grandes complexos de dunas do deserto Caiuá, que se estende para região norte do estado do Paraná. (II) O Grupo Bauru é composto pelas formações Uberaba, Vale do Rio do Peixe, Araçatuba, São José do Rio Preto, Presidente Prudente e Marília e com analcimitos Taiúva, rochas vulcânicas localmente intercaladas na sequência (Fernandes, 2004; Fernandes et al., 2007).

Os solos de maior ocorrência são classificados como Argissolo Vermelho-Amarelo, Latossolo Férrico, Neossolo Litólico, Nitossolo Vermelho e Gleissolo Háptico (Figura 1B) (Rossi et al., 2017). A Figura 1C apresenta a geomorfologia da área estudada.

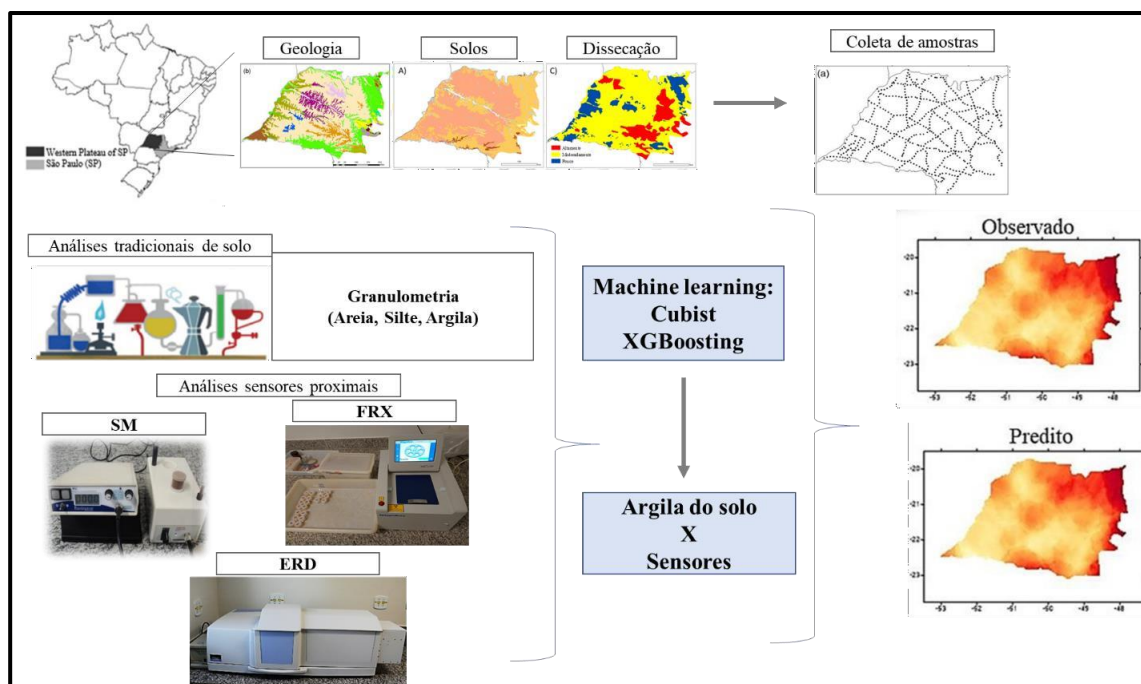
O nível de dissecação expressa a evolução solo-paisagem (Figura 1C), que é associada a pedogênese (formação do solo) e geomorfogênese (evolução da paisagem). As áreas geologicamente relacionadas aos solos originados de basalto estão relacionadas ao

compartimento pouco dissecado. Enquanto áreas dos solos originados de arenito estão relacionadas ao compartimento moderadamente dissecado. O predomínio na região é de áreas moderadamente dissecadas seguindo o comportamento geológico sedimentar do planalto (Figura 1C).

O POP representa a região agrícola produtora do estado de São Paulo, na qual os levantamentos agrários revelam que o estado de São Paulo é a unidade federativa com maior número de área colhida por hectares e rendimento médio, somando um total de 8.640 milhões de hectares de área colhida de um total de 86.820 milhões de hectares colhidos nacionalmente (IBGE, 2023). O POP é responsável pelo maior valor bruto de produção do país e a maior diversificação de culturas agrícolas do país representados principalmente pela produção de cana-de-açúcar, soja, milho, café, amendoim e espécies frutíferas como goiaba, limão, laranja e pastagem representando a principal região produtora nacional (IEA, 2023).

Planejamento amostral

Por meio do conhecimento prévio do mapa geológico, de solos e dissecação da paisagem (Figura 1) e seguindo o esquema da metodologia que foi adotada no presente estudo apresentado na Figura 2, foram coletadas 300 amostras de solo georreferenciadas do POP para a obtenção das medidas dos sensores próximos (SM, ERD, FRx) e atributos granulométricos do solo (Areia, Silte, Argila) na profundidade de 0,00-0,20 e 0,20-0,40 m, seguindo como base o arquivo de rodovias do estado fornecidas pelo Departamento de Estradas e Rodagem, sendo elaborado um plano de amostragem utilizando a ferramenta ETGeoWizards no ArcView 9.3. O espaçamento mínimo e máximo foi de 10 km a 60 km. As amostras foram coletadas ao longo das rodovias do estado, em locais representativos dos solos e da geologia regional. As coordenadas de cada ponto foram definidas previamente e inseridos em um GPS de navegação com precisão de 3 a 5 metros.

Figura 2. Esquema dos passos utilizados no desenvolvimento do projeto de pesquisa.

Fonte: Própria (2024)

Análises laboratoriais

Caracterização granulométrica

As amostras foram secas e passadas em peneira de 2 mm de abertura de malha para obtenção da terra fina seca ao ar (TFSA). A granulometria do solo foi determinada pelo método da pipeta utilizando solução de NaOH 0,1 mol L⁻¹ como dispersante químico e agitação mecânica em aparelho de baixa velocidade por 16 h (Figura 2) (Teixeira et al., 2017).

Espectrometria de fluorescência de raio-X (EFRx)

Para obtenção dos espectros da FRx, serão utilizadas 2g de TFSA peneirados e homogeneizados, no qual serão escaneadas no equipamento modelo NexQc da marca Rigaku (Applied rigaku Technologies, inc., TX USA) que possui um detector de desvio de silício com uma resolução < 145. As amostras serão irradiadas na atmosfera do ar, utilizando um tubo de raios-X Rh operado a 15 kV (Na - Sc) e 50 kV (Ti - U) por 200 e 250 s. Para obter uma excitação adequada para uma ampla gama de elemento, o tubo de raios-x Rh será operado a 15 kV e 50 kV. A voltagem de 15 kV favorece a fluorescência de elementos na faixa de Na-Sc e 50 kV é mais apropriado para elementos mais pesados na faixa Ti-Fe. Para fins de avaliação da detecção do conteúdo elementar da amostra pela EFRx, materiais do solo certificados pelo Serviço Geológico dos Estados Unidos (USGS): BCR-2, DNC-1, QLO-1a, SDC-1 e W-2A (Flanagan, 1970; Jochum et al., 2016) foram usados para calibrar o equipamento.

Análise de dados

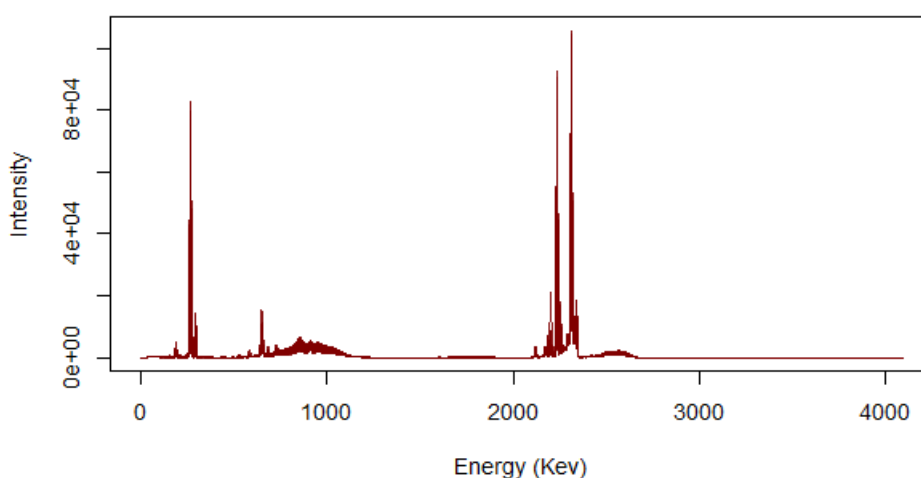
Estatística descritiva

Os atributos granulométricos (Areia, Silte, Argila) foram analisadas previamente por meio da estatística descritiva calculando-se os valores de medianas, máximos e mínimos para os níveis de dissecação da paisagem e materiais de origem. foram realizadas comparações entre os três níveis de dissecação da paisagem pelo teste Kruskal-Wallis e quando observada diferença entre os grupos, o teste post-hoc de Dunn a 5% será utilizado. As comparações entre os dois materiais de origem forma realizadas pelo teste de Mann-Whitney a 5%. Os dados forma processados por meio do software estatístico R Core Team (2022) com auxílio dos pacotes: rstatix (Kassambara) e corrplot (Wei; Simko, 2017).

Pré-processamento das curvas espectrais

Etapa 1: Os dados das duas curvas espectrais obtidas na voltagem de 15 KeV e 50 KeV espectrais foram concatenados (Figura 2). A voltagem de 15 kV favorece a fluorescência de elementos na faixa de Na-Sc e 50 kV é mais apropriada para elementos mais pesados na faixa Ti-Fe (Silva et al., 2021).

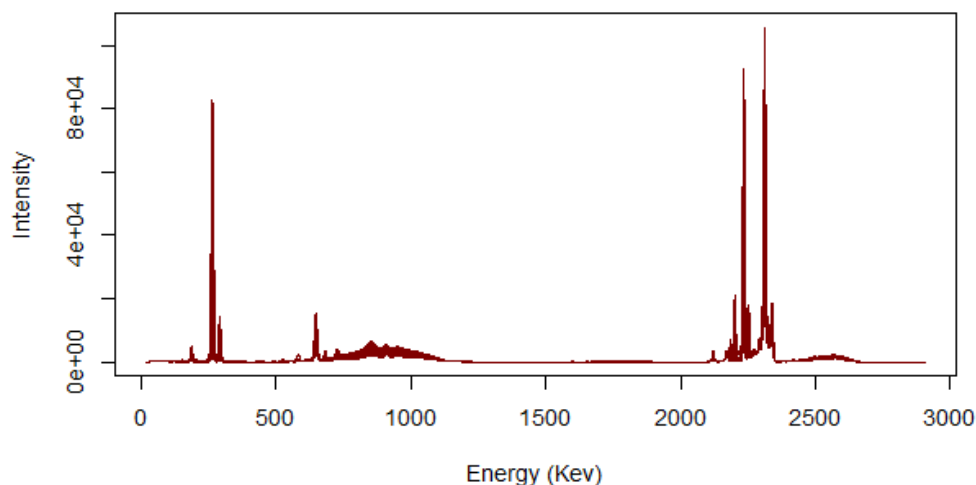
Figura 2: Espectros da espectrometria de fluorescência de raios-X concatenados.



Fonte: Própria (2024)

Etapa 2: Realizou-se uma análise contínua em todo o espectro obtido para a retirada de todas as colunas que apresentam somatório igual a zero e ruídos providos durante o período da análise (Figura 3) (Weindorf et al., 2020).

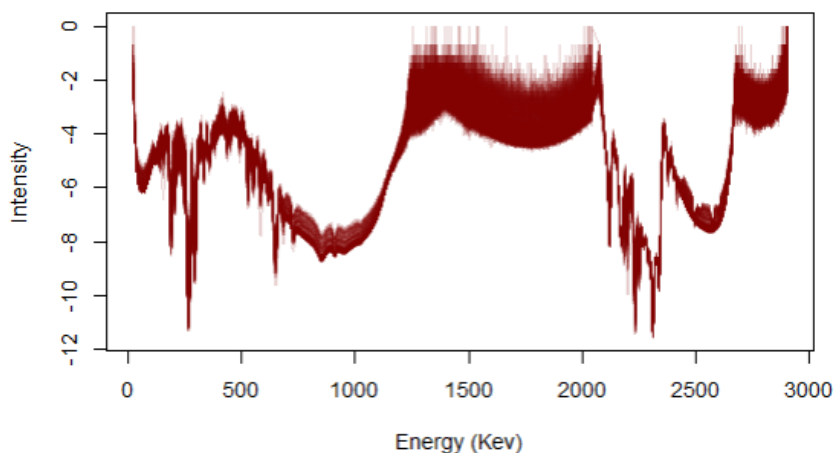
Figura 3: Espectros da espectrometria fluorescência de raios-X concatenados após remoção das colunas de baixo sinal espectral.



Fonte: Própria (2024)

Etapa 3: Os espectros de reflectância foram transformados em absorvância usando a fórmula $\log_{10}(1/R)$, onde R é a reflectância obtida, tornando evidente os picos dos elementos presentes. Quando os espectros são usados para extrair informações do solo por meio de modelagem, os mesmos são melhor expressos em termos de absorvância em vez de unidades de reflectância (Gobrecht et al., 2015) (Figura 4).

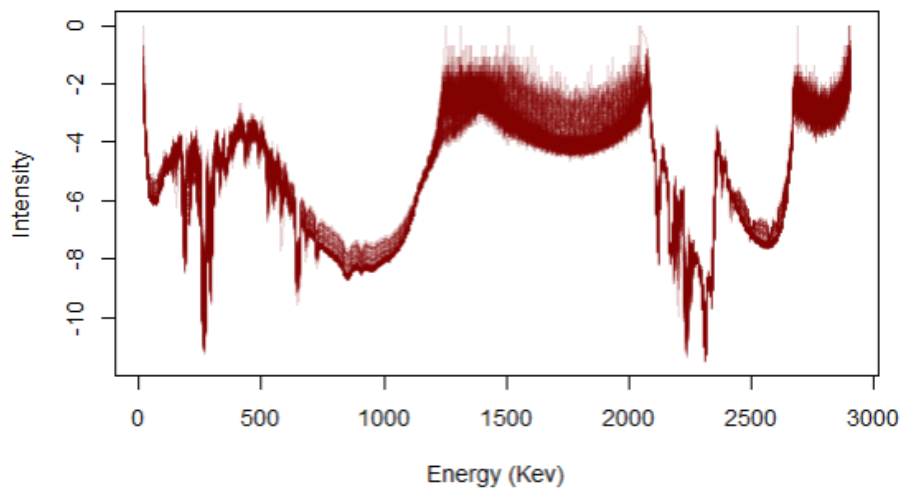
Figura 4: Espectros da espectrometria de fluorescência transformados em absorvância.



Fonte: Própria (2024)

Etapa 4: Após a transformação para absorvância, foi realizada redução da dimensão dos espectros, mantendo a absorção das características mais importantes. Para redução da dimensão foi usada a técnica da reamostragem e cálculo da média, para a reamostragem aqui realizada foi utilizada uma janela de tamanho 5, reduzindo as curvas espectrais para 578 informações (Figura 5).

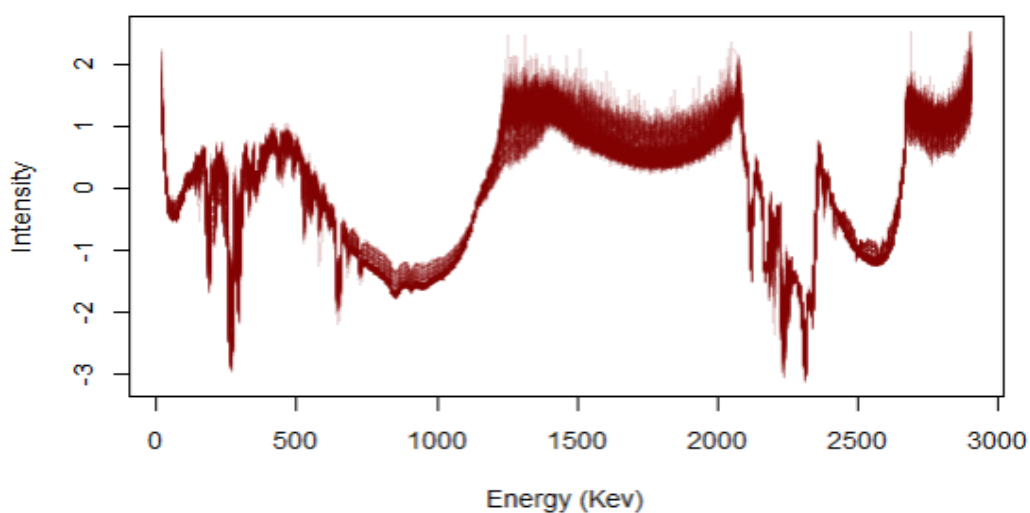
Figura 5: Espectros da espectrometria de fluorescência de raios-X após aplicação da técnica de reamostragem.



Fonte: Própria (2024)

Etapa 5: Foi realizada a transformação variável normal padrão (VNP) que corrige o espalhamento de luz (Barnes et al., 1989). Os espectros são normalizados ou dimensionados para média zero e variância unitária (Figura 6).

Figura 6: Espectros da espectrometria de fluorescência de raios-X após centralização.

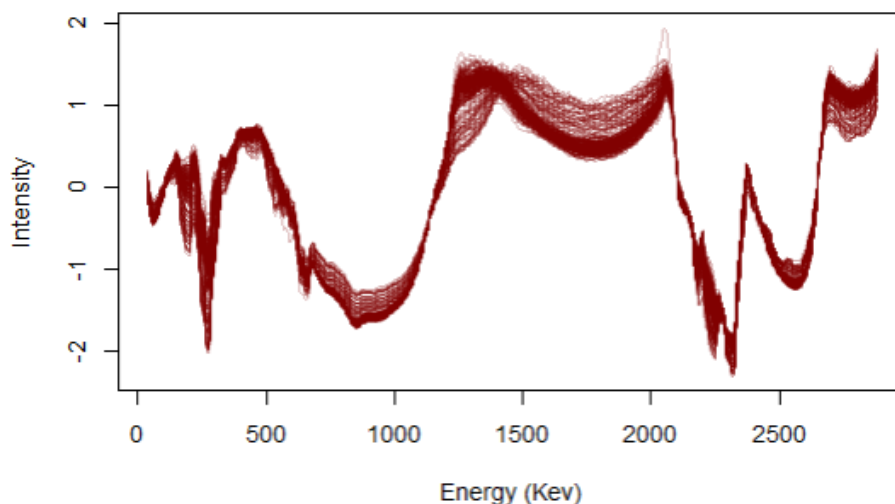


Fonte: Própria (2024)

Etapa 6: Para finalização do pré-processamento, foi realizado a operação de janela móvel, que cada valor de comprimento de onda é considerado como as médias dos comprimentos de onda vizinhos. Os espectros originais são suavizados, reduzindo o ruído que contém no espectro. O

espectro passa a ter um total de 569 informações (Figura 7).

Figura 7: Espectros de espectrometria de fluorescência de raios-X após remoção de ruído.



Fonte: Própria (2024)

Aprendizado de máquina

A estimativa do teor de argila utilizando os sensores SM, ERD e EFRx foi obtida pelas técnicas de aprendizado de máquina utilizando os algoritmos *Cubist* e *extreme gradient boosting* (XGboosting). O modelo cubista é baseado em regras que permite explorar relações não lineares em dados observados. É um algoritmo de modelo de árvore baseado na teoria M5 (Quinlan, 1992) e particiona as variações do preditor em diferentes subconjuntos de acordo com uma série de regras com cada regra representando um modelo linear multivariado do preditor (Kuhn et al., 2016).

O XGboosting é um algoritmo híbrido do modelo de árvore, aplicado a previsões de regressão e classificação. É um modelo ensemble que funciona com base na previsão de várias árvores, que é altamente eficaz e amplamente utilizado (Chen; Guestrin, 2016). Seu mecanismo de trabalho escalável é baseado em estágios com aumentos no número de árvores.

O banco de dados será padronizado a média nula e variância unitária, transformando todas as variáveis para mesma ordem de grandeza. Para realizar a predição o banco de dados será dividido em treino e teste. Assim, 70% dos dados ficaram para treinamento do modelo e 30% para o teste. Os modelos gerados serão avaliados por meio do coeficiente de determinação ajustado (R^2), erro absoluto médio (MAE) e raiz do erro quadrático médio (RMSE). Os modelos preditivos foram desenvolvidos usando o pacote Caret (Kuhn, 2008) por meio dos algoritmos: “Cubist” (Quinlan and Ross, 1992) e Xgbtree (Chen and Guestrin, 2016) utilizando

linguagem de programação R (R Development Core Team, 2020).

Análise Espacial

A análise da dependência espacial dos dados será realizada por meio da geoestatística (VIEIRA et al., 1983). O semivariograma experimental será calculado de acordo com Oliver e Webster (2014) por meio da seguinte equação:

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

Em que: $\hat{\gamma}(h)$ é a semivariância na distância h ; $N(h)$ é o número de pares envolvidos para calcular h ; $Z(x_i)$ é o valor atributo Z na posição x_i , e $Z(x_i + h)$ é o valor do atributo Z separado por uma distância h da posição x_i .

Os resultados dos variogramas experimentais serão submetidos ao ajuste de modelos teóricos: modelo esférico, exponencial e gaussiano. Após a modelagem e escolha do melhor modelo teórico ajustado aos dados, os valores correspondentes aos pontos não amostrados foram estimados pela técnica de interpolação da krigagem ordinária usando pacote gstat (Pebesma, 2001). Os mapas de variabilidade espacial dos atributos serão processados utilizando os seguintes pacotes: pacote sp (Pebesma e Bivand, 2005); pacote maptools (Bivand et al., 2023) e pacote raster (Hijmans et al., 2013). O grau de dependência espacial (GDE) será estimado a partir da razão entre o efeito pepita (C_0) e o patamar ($C_0 + C_1$). O GDE foi considerado forte, moderado ou fraco quando a razão $[C_0/(C_0+C_1) \times 100]$ foi de menor que 25%, 25-75% ou superior a 75%, respectivamente (Cambardella et al., 1994). Os processamentos dos dados geoestatísticos serão realizados utilizando linguagem de programação R (R Development Core Team, 2020).

Resultados e discussão

Caracterização descritiva da área de estudo.

A caracterização da estatística descritiva dos atributos granulométricos (Areia, Silte, Argila) estão presentes na Tabela 1. Os níveis de dissecação da paisagem para os solos de origem basáltica não apresentaram diferenças significativas entre si ($p < 0,05$) para os teores de areia, silte e argila. Para os solos de origem arenítica, os níveis de dissecação da paisagem também não apresentaram diferenças significativas, demonstrando que os atributos são influenciados pelos tipos de materiais de origem (Basalto e Arenito).

Tabela 1. Estatística descritiva para atributos granulométricos por metodologias convencionais para os solos do Planalto Ocidental Paulista (POP).

SOLOS ORIGINADOS DE BASALTO					
	Pouco ¹	Moderadamente ¹	Altamente ¹	Geral ²	Min-Max
Areia (g kg ⁻¹)	449,31 a	399,36 a	369,80 a	393,00 B	95,89-832,00
Silte (g kg ⁻¹)	121,12 a	165,70 a	183,15 a	154,00 A	32,00-472,48
Argila (g kg ⁻¹)	442,50 a	411,50 a	428,50 a	426,00 A	125,00-610,00
SOLOS ORIGINADOS DE ARENITO					
Areia (g kg ⁻¹)	787,00 a	780,29 a	777,70 a	780,00 A	182,11 - 926,98
Silte (g kg ⁻¹)	92,25 a	94,89 a	90,30 a	94,20 B	13,00 - 500,24
Argila (g kg ⁻¹)	103,25 a	110,00 a	115,50 a	110,00 A	24,00 - 628,00

¹Medianas seguidas pela mesma letra minúscula não diferem pelo teste de Dunn a 5% para os níveis de dissecação da paisagem, dentro da mesma geologia. ²Mediana, medianas seguidas pela mesma letra maiúscula na pelo teste de Mann-Whitney a 5% entre as geologias.

Fonte: Própria (2024)

Os solos originados de basalto apresentaram os maiores teores de argila (426 g kg⁻¹) e silte (154 g kg⁻¹). Já os solos de arenito apresentaram os maiores teores de areia em sua composição (780 g kg⁻¹) (Tabela 1). O comportamento aqui observado para a granulometria do solo ocorre devido aos solos originados de rochas ígneas (basalto) possuírem baixo teor de sílica e alto teor de minerais ferromagnesianos, que após a sua transformação por meio do intemperismo, resultam na formação de minerais secundários presentes na fração argila, como a caulinita, gibbsita e os óxidos de ferro (Schwertmann, 1988; Waroszewski et al., 2019; Fruett et al., 2022).

Estimativa dos atributos

As métricas e gráficos de desempenho da estimativa do teor de argila por meio de espectros de EFRx usando algoritmos de aprendizado de máquina cubista e XGBoost são apresentados na Tabela 2 e Figuras 2, respectivamente. Os dados para o período de teste dos modelos das estimativas apresentam os maiores valores de coeficientes de determinação (R^2) para as estimativas obtidas pelo algoritmo cubista. Para o algoritmo cubista os resultados obtidos foram: para o período de treinamento ($R^2 = 0,94$ e RMSE = 29,76) e teste ($R^2 = 0,73$ e RMSE = 80,48).

Para as estimativas obtidas pelo algoritmo XGBoosting os resultados foram: para o período de treinamento ($R^2 = 0,98$ e RMSE = 19,14) e teste ($R^2 = 0,65$ e RMSE = 90,75). Isso demonstra que os menores valores de RMSE apresentados para o modelo cubista para o teor

confirma o potencial desse modelo para o estudo e estimativa de atributos do solo.

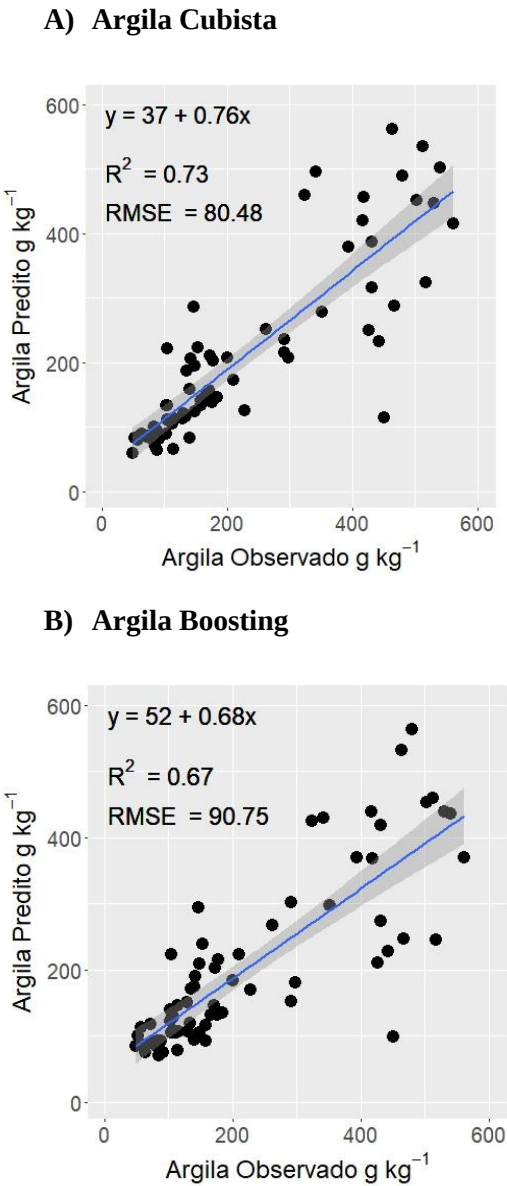
Tabela 2. Métricas de avaliação dos modelos cubista e extreme gradient boosting (XGBoost) para o teor de argila do solo.

Atributo	CUBISTA				XGBOOSTING			
	Treino		Teste		Treino		Teste	
	R ²	RMSE*	R ²	RMSE*	R ²	RMSE*	R ²	RMSE*
Argila	0,94	29,76	0,73	80,48	0,98	19,14	0,67	90,75

*RMSE - raiz do erro quadrático médio.

Fonte: Própria (2024)

Figura 3. Regressão do desempenho do modelo cubista e XGBoosting para a argila.



Fonte: Própria (2024)

Em solos subtropicais do Sul do Brasil, Fontenelli et al. (2021) estimando o teor de argila por meio dos espectros de EFRx usando regressão parcial dos mínimos quadrados (PLSR) encontraram valores de inferiores aos encontrados no presente estudo, com valores de R^2 em torno de 0,50 e valores de RMSE em torno de 5,51. Os resultados encontrados pelos autores foram menores que os encontrados no presente estudo, demonstrando que apesar da dimensão territorial do Planalto Ocidental Paulista os modelos desenvolvidos apresentam desempenho similar a áreas agrícolas de menor dimensão.

Mapeamento da argila obtida e estimada pelo aprendizado de máquina: Cubista e Boosting.

O modelo esférico foi o que melhor se ajustou aos variogramas experimentais (Tabela 3). Tal comportamento indica que variabilidade espacial da propriedade do solo estudada é menos errático em pequena escala (Khan et al., 2021). O modelo esférico é caracterizado por se ajustar as variáveis que possuem mudanças abruptas no espaço (Isaaks e Srivastava, 1989). As mudanças abruptas podem estar relacionadas ao material de origem (Teixeira et al., 2018), tipos de solo (Silveiro et al., 2019) ou a mineralogia do solo (Fernandes et al., 2020, Silva et al., 2020).

Tabela 5. Métricas geoestatísticas para argila na de na profundidade de 0,0–0,2 m.

	Atributo	Modelo	C₀	C1	A₀(km)	SQR	GDE
Observado	Argila (g kg ⁻¹)	Esférico	17091,01	6442,30	209,83	0,74	72,62
Cubista	Argila (g kg ⁻¹)	Esférico	13479,77	1222,66	161,69	0,86	73,02
XGBoosting	Argila (g kg ⁻¹)	Esférico	13704,69	806,04	152,63	0,95	73,30

*C₀ – Efeito pepita; C – Patamar; A₀ - Alcance; SQR – Soma dos quadrados dos resíduos;

Fonte: Própria (2024)

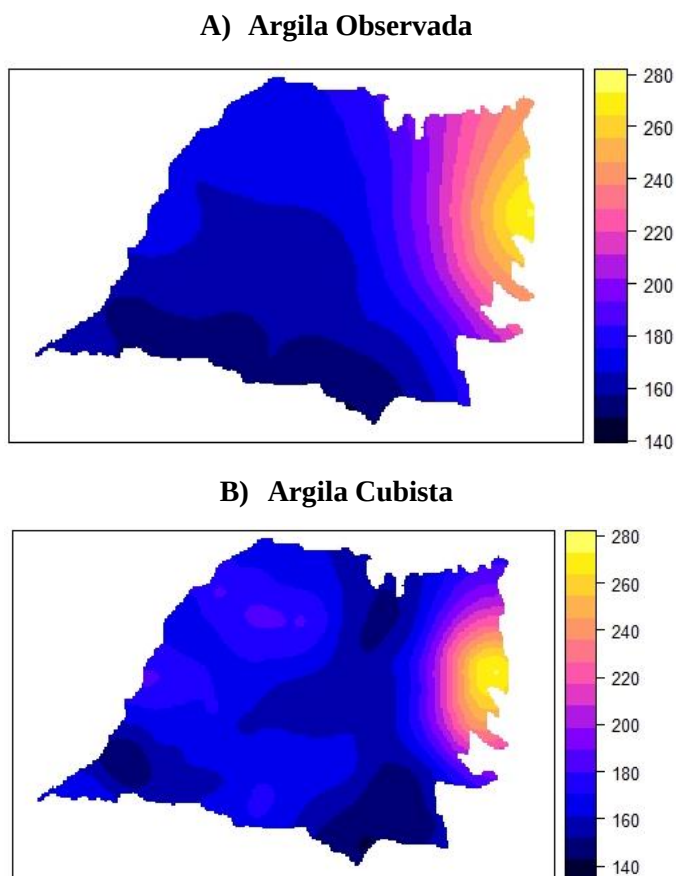
Os valores de alcance (A) obtidos em estudos geoestatísticos representam o intervalo máximo em que os valores do atributo apresentam dependência espacial (Isaaks e Srivastava, 1989), e podem ser usados para futuros estudos geoestatísticos em regiões com características semelhantes, indicando o intervalo espacial para coletas de amostras em estudos futuros (Siqueira et al., 2014, Marques Jr. et al., 2015). Os valores de alcance encontrados para o atributo do solo apresentaram ordem decrescente: Argila observada (209, 83 Km), argila estimada por o modelo cubista (161,69 Km), argila estimada pelo modelo XGBoosting (152,63 Km). Os valores de alcance em ordem decrescente demonstram que ocorre uma suavização dos resultados dos dados nas estimativas dos modelos, porém não se apresenta como fator limitante,

notando os valores de acurácia.

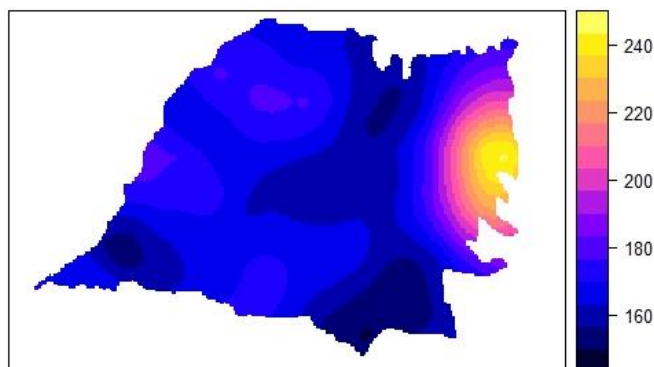
Os atributos do solo presentes nesse estudo apresentaram um grau de dependência espacial classificado como moderado (GDE entre 0,50 - 0,75). Segundo Isaaks e Srivastava (1989) a moderada dependência espacial encontrada para atributos do solo são atribuídas a fatores intrínsecos ao solo, como o material de origem do solo e aos seus processos de formação somado as ações de manejo antrópico do solo.

Os mapas de variabilidade espacial dos atributos químicos observados e estimados por meio das medidas da EFRx utilizando os algoritmos cubista e XGboosting são apresentados na Figura 11. Observa-se que os mapas do comportamento espacial semelhantes entre os dados observados e estimados, reforçando que a EFRx aliada as técnicas de aprendizado de máquina são ferramentas robustas para estudo desses atributos em áreas de grande dimensão. Pelegrino et al. (2021) reforçam os resultados aqui obtidos, já que os autores obtiveram resultados promissores nos estudos de variabilidade espacial do teor de argila em solos altamente intemperados do Brasil. Alder et al. (2020) avaliaram o potencial da EFRx para estimar o teor de argila e outras frações granulométricas obtendo valores de precisão com R^2 superior a 0,63.

Figura 4. Mapas de variabilidade espacial observados e estimados pelos modelos.



C) Argila Boosting



Os resultados demonstram a semelhança entre os mapas observados e os estimados por meio do sensor da EFRx, reforçam seu potencial para reconhecimento de zonas específicas de manejo em áreas de grande dimensão. Somado a isso, os resultados obtidos nesse estudo auxiliam na consolidação da EFRx como uma técnica pedométrica robusta para dá argila do solo, auxiliando na elaboração de mapas detalhados, mesmo sobre complexa variação pedológica e geológica, no caso do Planalto Ocidental Paulista.

A EFRx fornece resultados promissores que podem auxiliar no planejamento estratégico sustentável de uso e ocupação do solo regional, auxiliando na aplicação adequada de fertilizantes por meio do estabelecimento de zonas de manejo homogêneas.

Conclusões

As estimativas obtidas utilizando os espectros de fluorescência por meio dos algoritmos de aprendizado de máquina utilizados foram capazes de estimar com precisão satisfatória a teor de argila do solo.

Para os dois algoritmos utilizados na modelagem, o modelo cubista foi o que apresentou os maiores valores de precisão e acurácia quando comparados ao modelo XGBoosting.

Os mapas obtidos a partir dos valores estimados pelo sensor de EFRx apresentou comportamento espacial semelhante ao mapa observado para o atributo, demonstrando o potencial dessa tecnologia proximal para o mapeamento dos atributos do solo.

Referências Bibliográficas

AYOUBI, S., KARAMI, M. Pedotransfer functions for predicting heavy metals in natural soils using magnetic measures and soil properties. *Journal of Geochemical Exploration*, 197, 212-219, 2019.

BEHRENS, T., VISCARRA ROSSEL, R. A., RAMIREZ-LOPEZ, L., BAUMANN, P. Soil

spectroscopy with the gaussian pyramid scale space. *Geoderma*, 426, 116095, 2022.

BIVAND, R., LEWIN-KOH, N., PEBESMA, E., ARCHER, E., BADDELEY, A., BEARMAN, N., GOLICHER, D. 2023. Package ‘maptools’, 2023.

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*, 58:1501-1511, 199.

CHEN T, GUESTIN C. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* pp. 785-794, 2016.

COSTA, A.C.S., BIGHAM, J.M., RHOTON, F.E., TRAINA, S.J. Quantification and characterization of maghemite in soils derived from volcanic rocks in southern Brazil. *Clays Clay Minerals*, New York, v.47, p.466-473, 1999.

DEARING, J.A. Environmental magnetic susceptibility. Using the Bartington MS2 system. England: British Library, 104p, 1994.

FERNANDES, L. A., DE CASTRO, A. B., BASILICI, G. Seismites in continental sand sea deposits of the late Cretaceous Caiua Desert, Bauru Basin, Brazil. *Sedimentary Geology*, 199(1-2), 51-64, 2007.

FERNANDES, LA. Mapa litoestratigráfico da parte oriental da Bacia Bauru (PR, SP, MG), escala 1: 1.000. 000. *Boletim Paranaense de Geociências*, 55, 2004.

FILLA, V.A., COELHO, A.P., FERRONI, A.D., DE SOUZA BAHIA, A.S.R., JÚNIOR, J.M. Estimation of clay content by magnetic susceptibility in tropical soils using linear and nonlinear models. *Geoderma*, 403, 115371, 2021.

FLANAGAN, F.J. Sources of geochemical standards—II. *Geochimica et Cosmochimica Acta* 34(1), 121-125, 1970.

GEORGE, E.B., GOMEZ, C., KUMAR, N.D. Adapting Prediction Models to Bare Soil Fractional Cover for Extending Topsoil Clay Content Mapping Based on AVIRIS-NG Hyperspectral Data. *Remote Sensing*, 16(6), 1066, 2024.

HIJMANS, R.J.; VAN ETEN, J.; MATTIUSZI, M.; SUMNER, M.; GREENBERG, J.A.; LAMIGUEIRO, O.P.; SHORTRIDGE, A. Raster package in R. Version. <https://Mirro Rs. Sjtug. Sjt. Edu. Cn/Cran/Web/Packa Ges/Rast er/Raster. Pdf>, 2013.

HU, B., JIA, X., HU, J., XU, D., XIA, F., LI, Y. Assessment of heavy metal pollution and health risks in the soil-plant-human system in the Yangtze River Delta, China. *International*

journal of environmental research and public health, 14(9), 1042, 2017.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA – IBGE. Produção agropecuária. Disponível em: <https://www.ibge.gov.br/explica/producao-agropecuaria/>. Acesso em: 16/04/2024. Instituto de economia agrícola. Banco de dados: estatísticas de preços médios. São Paulo, EIA,2022, Disponível em: http://ciagri.iea.sp.gov.br/nia1/precos_medios.aspx?cod_sis=2.

JOCHUM KP, WEIS U, SCHWAGER B, STOLL B, WILSON SA, HAUG GH, ENZWEILER J. Reference values following ISO guidelines for frequently requested rock reference materials. *Geostandards and Geoanalytical Research* 40(3), 333-350, 2016.

KASSAMBARA, A. Package ‘rstatix’. R topics documented, 2020.

KUHN, M.; WESTON, S.; KEEFER, C.; COULTER, N Cubist: Rule-and Instance-Based Regression Modeling (Version 0.2. 2), 2018.

LEMIERE, B.A. review of pXRF (field portable X-ray fluorescence) applications for applied geochemistry. *Journal of Geochemical Exploration*, 188, 350-363, 2018.

LI, Y., CHEN, H., TENG, Y. Source apportionment and source-oriented risk assessment of heavy metals in the sediments of an urban river-lake system. *Science of the Total Environment*, 737, 140310, 2020.

LIU, P., ZHANG, Y., FENG, N., ZHU, M., & TIAN, J. Potentially toxic element (PTE) levels in maize, soil, and irrigation water and health risks through maize consumption in northern Ningxia, China. *BMC Public Health*, 20, 1-13, 2020.

MATUS, F.J. Fine silt and clay content is the main factor defining maximal C and N accumulations in soils: a meta-analysis. *Scientific Reports*, 11(1), 6438. 2021.

MCKEAGNE, J.A., DAY, J.H. Dithionite and oxalate-extractable Fe and Al as aids in differentiating various classes of soils. *Canadian Journal of Soil Science*, Ottawa, v.46, p.13–22, 1966.

MEHRA, O.P, JACKSON, M.L. Iron oxide removed from soils and clays by dithionitecitrate system buffered with sodium bicarbonate. *Clays and Clay Minerals*, 7(1), 1317-327, 1960.

PADARIAN, J., MCBRATNEY, A. B., MINASNY, B. 2020. Game theory interpretation of digital soil mapping convolutional neural networks. *SOIL Discussions*, 2020, 1-12, 2020.

PEBESMA, E, BIVAND, R.S.S. classes and methods for spatial data: the sp package. *R news*, 5(2), 9-13, 2005.

PEBESMA, E. J. Gstat user’s manual. Dept. of Physical Geography, Utrecht University, Utrecht, The Netherlands, 2001.



PENG, X., DAI, Q., DING, G., SHI, D., LI, C. 2020. Impact of vegetation restoration on soil properties in near-surface fissures located in karst rocky desertification regions. *Soil and Tillage Research*, 200, 104620.

PETERS, C.; DEKKERS, M. 2003. Selected room temperature magnetic parameters as a function of mineralogy concentration and grain size. *Physics and Chemistry of the Earth*, v.28, p.659- 667.

Quinlan JR (1992) Learning with continuous classes., *Proc. In: of the Fifth Australian*

R DEVELOPMENT CORE TEAM, R. R: A Language and Environment for Statistical Computing, 2020.

RAMÍREZ, P. B., CALDERÓN, F. J., HADDIX, M., LUGATO, E., COTRUFO, M. F. Using diffuse reflectance spectroscopy as a high throughput method for quantifying soil C and N and their distribution in particulate and mineral-associated organic matter fractions. *Frontiers in Environmental Science*, 9, 634472, 2021.

RAVANSARI, R., WILSON, S. C., TIGHE, M. Portable X-ray fluorescence for environmental assessment of soils: Not just a point and shoot method. *Environment international*, 134, 105250, 2020.

ROSSI, M. Mapa Pedológico Do Estado de São Paulo: Revisado e Ampliado. Instituto Florestal, 2019.

TEIXEIRA, P.C., DONAGEMMA, G.K., FONTANA, A. TEIXEIRA, W.G. 2017. Manual de métodos de análise de solo. Embrapa, 2017.

THORNTHWAITE, C.W. 1948. An approach toward a rational classification of climate. *Geographical review*, 38(1), 55-94, 1948.

TIAN, K., HUANG, B., XING, Z., HU, W. 2018. In situ investigation of heavy metals at trace concentrations in greenhouse soils via portable X-ray fluorescence spectroscopy. *Environmental Science and Pollution Research*, 25, 11011-11022, 2018.

VIEIRA, S.R.; HATFIELD, J.L.; NIELSEN, D.R.; BIGGAR, J.W. 1983. Geostatistical theory and application to variability of some agronomical properties. *Hilgardia* 51(3), 1-75, 1983.

VISCARRA ROSSEL, R. A., BEHRENS, T., BEN-DOR, E., CHABRILLAT, S., DEMATTE, J. A. M., GE, Y.; SHEN, Z. Diffuse reflectance spectroscopy for estimating soil properties: A technology for the 21st century. *European Journal of Soil Science*, 73(4), e13271, 2022.

WADOUX, A.M.J.C., MOLNAR, C. 2022. Beyond prediction: Methods for interpreting

complex models of soil variation. *Geoderma*, 422, 115953, 2022.

WEI T, SIMKO V, . In: R Package "corrplot": Visualization of a Correlation Matrix (Version 0.84). Available from: <https://github.com/taiyun/corrplot>, 2017.

WEINDORF, D.C., CHAKRABORTY, S. Portable X-ray fluorescence spectrometry analysis of soils. *Soil Science Society of America Journal*, 84(5), 1384-1392, 2020.

YAO, W., HU, C., YANG, X., SHUI, B. Spatial variations and potential risks of heavy metals in sediments of Yueqing Bay, China. *Marine Pollution Bulletin*, 173, 112983, 2021.

YOUSEFI, G., HOMAEE, M.; NOROUZI, A.A. Estimating soil heavy metals concentration at large scale using visible and near-infrared reflectance spectroscopy. *Environmental monitoring and assessment*, 190, 513, 2018.

ZHAO, D., WANG, J., ZHAO, X., TRIANTAFILIS, J. Clay content mapping and uncertainty estimation using weighted model averaging, 2022.

