

ÉSIO DE CASTRO PAES

**PREDIÇÃO E ESPACIALIZAÇÃO DE ELEMENTOS POTENCIALMENTE
TÓXICOS EM SOLOS NO ENTORNO DE PILHA DE REJEITO DE MINERAÇÃO:
RISCOS À SAÚDE HUMANA E FITORREMEDIAÇÃO**

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

Orientadora: Emanuelle Mercês Barros Soares

Coorientador: Elpídio Inácio Fernandes Filho

**VIÇOSA - MINAS GERAIS
2022**

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

T

P126p
2022

Paes, Ésio de Castro, 1992-

Predição e espacialização de elementos potencialmente tóxicos em solos no entorno de pilha de rejeito de mineração: riscos à saúde humana e fitorremediação / Ésio de Castro Paes. – viçosa, MG, 2022.

1 tese eletrônica (167 f.): il. (algumas color.).

Orientador: Emanuelle Mercês Barros Soares.

Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Solos, 2022.

Inclui bibliografia.

DOI: <https://doi.org/10.47328/ufvbbt.2022.071>

Modo de acesso: World Wide Web.

1. Solos - Poluição. 2. Algoritmos computacionais. 3. Aprendizado do computador. 4. Fitorremediação. I. Soares, Emanuelle Mercês Barros, 1979-. II. Universidade Federal de Viçosa. Departamento de Solos. Programa de Pós-Graduação em Solos e Nutrição de Plantas. III. Título.

CDD 22. ed. 628.168

Bibliotecário(a) responsável: Alice Regina Pinto CRB6 2523

ÉSIO DE CASTRO PAES

**PREDIÇÃO E ESPACIALIZAÇÃO DE ELEMENTOS POTENCIALMENTE
TÓXICOS EM SOLOS NO ENTORNO DE PILHA DE REJEITO DE MINERAÇÃO:
RISCOS À SAÚDE HUMANA E FITORREMEDIAÇÃO**

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

APROVADA: 11 de fevereiro de 2022.

Assentimento:



Ésio de Castro Paes
Autor



Emanuelle Mercês Barros Soares
Orientadora

*Dedico esta tese aos meus pais,
Manoel e Eleonora.*

AGRADECIMENTOS

Agradeço primeiramente a Deus, por ter me dado força nos momentos de dificuldade nessa caminhada e agora permitindo o título de Doutorado, em Solos e Nutrição de Plantas.

Aos meus pais, Manoel Nelson e Eleonora Maria, pelo amor e incentivo durante toda minha jornada. Aos meus irmãos Elton e Manoel pela amizade e palavras de conforto nos momentos de dificuldades. Aos meus avós, Nelson, Godson, Lindaura e Lurdes pelo amor e por sempre torcerem por mim.

A minha namorada e melhor amiga Iara Oliveira pela compreensão, amor e ajuda na execução do trabalho.

Agradeço a minha orientadora Professora. Dra. Emanuelle Mercês, pela compreensão e companheirismo nos momentos que precisei, pelos ensinamentos durante esses quatro anos de convivência. e pela grande contribuição na minha formação profissional. Ao meu Coorientador, Professor Elpídio, pelos ensinamentos e apoio nos momentos de dificuldades. Ao meu amigo Dr. Gustavo, pela ajuda e ensinamentos no desenvolvimento desse trabalho.

Quero agradecer imensamente ao Prof. Dr. Maurício Fontes por todo apoio durante a realização das análises de laboratório. Ao Prof. Dr. José João L. L. de Souza pela realização da análise de florescência de raios-X.

Agradeço a prefeitura de Boquira-Ba pelo apoio e auxílio na execução do projeto. Aos meus amigos Luís (Secretário de Agricultura) e Zezinho, funcionários da prefeitura de Boquira pela ajuda, apoio e dedicação no desenvolvimento das coletas das amostras de solos.

À Universidade Federal de Viçosa, pela estrutura e aos Técnicos Marcos, Janílson, Carlinhos, João Milagres, Humberto, Paulo, Fabiane, Maurício e Júlio, por todo apoio durante a condução desse trabalho, sou eternamente grato a todos vocês.

Aos meus colegas de Pós-Graduação, Aymbiré, Patrícia, Matheus, Rafael, Klever, Saulo e David, pelo companheirismo e conhecimentos adquiridos durante essa caminhada. Ao amigo Ricardo pelos ensinamentos e ajuda nos momentos de dificuldades.

Ao Programa Solos e Nutrição de Plantas – UFV, pela oportunidade e apoio durante a minha trajetória no doutorado e a Capes pela concessão da bolsa.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

RESUMO

PAES, Ésio de Castro, D.Sc., Universidade Federal de Viçosa, fevereiro de 2022. **Predição e espacialização de elementos potencialmente tóxicos em solos no entorno de pilha de rejeitos de mineração: riscos à saúde humana e fitorremediação.** Orientadora: Emanuelle Mercês Barros Soares. Coorientador: Elpídio Inácio Fernandes Filho.

A atividade de mineração é muito importante para o crescimento de um País, porém a mesma gera uma grande quantidade de rejeitos, que, quando dispostos de maneira inadequada podem se dispersar no meio ambiente, aumentando os níveis de elementos potencialmente tóxicos (PTEs) nos solos. Assim, determinar os níveis e a distribuição espacial dos PTEs é essencial para traçar estratégias mitigadoras e remediar áreas contaminadas. Dessa forma, o objetivo desse estudo foi avaliar os níveis e a extensão da contaminação de elementos potencialmente tóxicos nos solos próximos a pilha de rejeito de mineração, bem como avaliar o potencial risco à saúde da população (Capítulo 1), desenvolver metodologia rápida e eficiente para monitorar os níveis de contaminantes em áreas impactadas pela atividade de mineração (Capítulo 2) e identificar espécies vegetais capazes de remediar áreas contaminadas (Capítulo 3). O estudo foi desenvolvido no Município de Boquira-Ba, no qual foram coletadas 120 amostras compostas de solos em uma malha regular, na profundidade de 0-10 cm. Os resultados demonstraram que os uso de algoritmos de aprendizagem de máquinas é uma alternativa eficiente para espacializar áreas contaminadas pela atividade de mineração, onde predomina fontes difusas de contaminantes. Observou-se também que as áreas próximas as pilhas de rejeito se encontram altamente contaminadas podendo ser prejudicial ao meio ambiente e a saúde humana. No segundo capítulo foi desenvolvida uma estrutura metodológica para monitorar áreas contaminadas com PTEs, por meio do uso de algoritmos de aprendizagem de máquinas. Os resultados demonstraram que o uso de covariáveis como as concentrações totais dos elementos, medidas por fluorescência de raio-X (XRF), texturais e morfométricas são úteis para prever os teores de PTEs nos solos. Por fim, foi possível observar que as espécies vegetais *P. juliflora*, *A. peregrina* toleram altas concentrações de Pb nos solos, e são capazes de estabilizar esse metal no sistema radicular. As demais espécies *M. oleífera* e *U. ruziziensis* também cresceram nos solos contaminados, porém ocorreram reduções nos valores dos parâmetros de crescimentos.

Palavras-chave: Áreas contaminadas. Algoritmos de aprendizagem de máquinas. Fitorremediação

ABSTRACT

PAES, Ésio de Castro, D.Sc., Universidade Federal de Viçosa, February, 2022. **Prediction and spatialization of potentially toxic elements in soils around a mining tailings pile: risks to human health and phytoremediation.** Advisor: Emanuelle Mercês Barros Soares. Co-advisor: Elpídio Inácio Fernandes Filho.

Mining activity is very important for the growth of a country, but it generates a large amount of tailings, which, when improperly disposed of, can be dispersed in the environment, increasing the levels of potentially toxic elements (PTEs) in the soil. Thus, determining the levels and spatial distribution of PTEs is essential to design mitigating strategies and remediate contaminated areas. Thus, the objective of this study was to assess the levels and extent of contamination of potentially toxic elements in soils near mining tailings piles, as well as to assess potential risks to the health of the population (Chapter 1), develop a rapid and efficient methodology to monitor contaminant levels in areas impacted by mining activity (Chapter 2) and identify plant species capable of remediate contaminated areas (Chapter 3). The study was developed in the city of Boquira-Ba, in which 120 samples were collected in a regular mesh, at a depth of 0-10 cm. The results showed that the use of machine learning algorithms is an efficient alternative to spatialize areas contaminated by mining activity, where diffuse sources of contaminants predominate. It was also observed that the areas close to the tailings piles are highly contaminated and can be harmful to the environment and human health. In the second chapter, a methodological framework was developed to monitor areas contaminated with PTEs, through the use of machine learning algorithms. Our results demonstrated that the use of covariates such as X-ray fluorescence (XRF), textural and morphometric measurements are useful to predict PTEs levels in soils. Finally, it was possible to observe that the plant species *P. juliflora*, *A. peregrina* tolerate high concentrations of Pb in soils, and are able to stabilize this metal in the root system. The other species *M. oleifera* and *U. ruziziensis* also grew in contaminated soils, however there were reductions in the values of growth parameters.

Keywords: Contaminated areas. Machine learning algorithms. Phytoremediation

SUMÁRIO

INTRODUÇÃO GERAL	10
REFERÊNCIAS	12
CAPITULO 1	15
Modelos de aprendizagem de máquinas para mapeamento de elementos potencialmente tóxicos e avaliação de risco a saúde humana em solos afetados pela atividade de mineração	15
RESUMO	15
Machine learning models for mapping potentially toxic elements and assessing human health risk in soils affected by mining activity.....	16
ABSTRACT	16
1-0- INTRODUÇÃO.....	17
2.0- MATERIAIS E MÉTODOS	19
2.1- Área de estudo	19
2.2- Amostragens de solo.....	21
2.3- Caracterizações física e química.....	22
2.4- Determinações dos teores de PTEs, elementos principais e controle de qualidade	23
2.5- Obtenção de covariáveis ambientais para espacialização da previsão dos PTEs	23
2.5.1- Coordenadas geográficas dos pontos amostrais.....	23
2.5.2- Variáveis obtidas por meio do MDE	24
2.5.3- Criação da covariável de localização da pilha de rejeito e canal drenagem (PDC)	26
2.6- Criação da função de pedotransferências por aprendizagem de máquinas	27
2.6.1- Remoção de covariáveis e treinamento dos algoritmos de aprendizagem de máquinas.....	27
2.6.2- Previsão dos mapas finais e incerteza dos mapas de teores PTEs nos solos	31
2.7- Avaliação ambiental de PTEs em solos.....	31
2.8- Avaliação de risco para a saúde humana	33
3.0- RESULTADOS	34
3.1- Remoção de Covariáveis	34
3.2- Propriedades químicas e textura dos solos estudados	34
3.3- Teores de PTEs nos solos estudados	35
3.4- Avaliação do desempenho dos modelos preditivos.....	36
3.5- Importância das covariáveis	38
3.6- Previsão dos mapas finais e incerteza dos PTEs e elementos principais nos solos.....	40
3.7- Avaliação ambiental de PTEs em solos.....	42
3.8- Avaliação de risco a saúde humana.....	44

4- DISCUSSÃO.....	45
4.1- Propriedades químicas e textura dos solos estudados	45
4.2- Desempenho dos algoritmos de aprendizagem de máquinas na previsão espacial dos teores de elementos potencialmente tóxicos e elementos principais nos solos	46
4.3- Importância das covariáveis	47
4.4- Distribuição espacial dos teores de PTEs e elementos principais nos solos	48
4.5- Avaliação da poluição do solo e risco a saúde humana.....	50
5.0 – CONCLUSÕES	52
6.0- AGRADECIMENTOS.....	53
7.0- REFERÊNCIAS	53
APÊNDICE	67
CAPITULO 2	70
Modelagem para previsão dos teores de elementos potencialmente tóxicos a partir de dados morfométricos, sensores proximais, físicos e químicos dos solos em áreas sob influência de mineração.....	70
RESUMO	70
Predictive modeling of contents of potentially toxic elements using morphometric data, proximal sensing, and chemical and physical properties of soils under mining influence	71
ABSTRACT	71
1.0- INTRODUCTION	72
2.0 - MATERIALS AND METHODS	74
2.1- Study site and soil sampling	74
2.2- Determination of potentially toxic elements contents and quality control.....	76
2.3- Obtaining covariates to predict the contents of potentially toxic elements.....	77
2.4- Creating pedotransfer function using machine learning.....	80
2.4.1- Removal of covariates and training data for the machine learning algorithm	80
2.4.2- Assessing algorithm performance.....	85
3.0- RESULTS	86
3.1- Contents of potentially toxic elements in the studied soils	86
3.2- Assessment of the precision of predictive models.....	87
3.3- Importance of covariates	93
4.0. DISCUSSION	96
4.1. Assessment of the precision of predictive models.....	96
4.2. Importance of covariates.....	97
5.0- CONCLUSIONS	100
6.0- ACKNOWLEDGEMENTS	100

7.0- REFERENCES	101
APÊNDICE	113
CAPÍTULO 3	117
Potencial de espécies vegetais adaptadas as condições do semiárido para fitorremediação de solos contaminados.....	117
RESUMO	117
Potential of plant species adapted to semiarid conditions for phytoremediation of contaminated soils.....	118
ABSTRACT	118
1.0- INTRODUÇÃO.....	119
2.0- MATERIAIS E MÉTODOS	121
2.1- Local do estudo.....	121
2.2- Delineamento experimental.....	121
2.3- Espécies utilizadas.....	122
2.4- Montagem e condução do experimento.....	123
2.5- Variáveis analisadas	124
2.6- Análise estatística	125
3.0- RESULTADOS	126
4.0- DISCUSSÃO.....	143
5.0- CONCLUSÕES.....	150
6.0- AGRADECIMENTOS.....	150
7.0- REFERÊNCIAS	150
APÊNDICE	162
CONSIDERAÇÕES FINAIS	167

INTRODUÇÃO GERAL

Crucial para a sustentação da vida na Terra, o solo exerce diversas funções ecossistêmicas essenciais. Além disso, é o principal compartimento transformador de resíduos e substâncias químicas produzidas pela humanidade. No entanto, o desenvolvimento urbano e industrial associado ao aumento da produção dos bens de consumo, tem resultado na geração de grande quantidade de substâncias e resíduos químicos, excedendo os limites dos solos de sorver tais poluentes comprometendo assim a qualidade ambiental.

Diversas atividades antrópicas têm contribuído para elevar os níveis de contaminantes nos solos e a mineração é uma das mais importantes, pois além de produzir fontes diretas de contaminantes, as pilhas de rejeito e escórias podem ser fontes contínuas dos mesmos após o fim das atividades de mineração (Krishna et al., 2013; Carvalho, 2017; Qu et al., 2018). Dentre os principais contaminantes que podem ser gerados, podemos destacar os elementos potencialmente tóxicos (PTEs). Esses elementos apresentam alta toxicidade mesmo em pequenas concentrações, não são biodegradáveis e podem acumular e serem transferidos na cadeia trófica, promovendo sérios problemas ambientais e a saúde humana (Yang et al., 2021).

No Sudoeste do estado da Bahia, Brasil, a atividade de mineração de chumbo (Pb) e zinco (Zn) realizada entre as décadas 1950 e 1990 produziu mais de 894 dm³ de rejeitos de mineração, que até os dias atuais estão servindo de fontes de poluentes para os solos (Santos et al., 2020). Por estar disposta de forma inadequada, a pilha de rejeito vem sofrendo com processos erosivos, tanto eólico quanto pluvial dispersando grande quantidade de elementos potencialmente tóxicos no ambiente. Um agravante que ocorre é a proximidade das pilhas de rejeitos de mineração com o centro urbano da cidade de Boquira-Ba.

Dessa forma, é necessário determinar as concentrações de PTEs e entender as formas geoquímicas e a extensão da área atingida pela contaminação para buscar estratégias que visem mitigar e remediar essas áreas contaminadas. A legislação brasileira dispõe sobre critérios e metodologias analíticas que devem ser adotadas para determinação de substâncias inorgânicas nos solos. São métodos propostos pela Agência de proteção Ambiental dos Estados Unidos, USEPA 3050b e 3051a ou suas atualizações (Conama, 2009). Porém, apesar de precisos, esses métodos são realizados por uma digestão ácida, seguida da dosagem por meio de equipamentos como (ICP) ou absorção atômica (AA), os quais apresentam custos elevados, são laboriosos e geram resíduos químicos que necessitam de tratamento especial (Usepa, 1998; McGladdery et al., 2018; Jin et al., 2020; Weindorf e Chakraborty, 2020).

Dessa forma, o uso de algoritmos de aprendizagem de máquinas tem se tornado uma alternativa eficiente e promissora para prever e espacializar teores de PTEs nos solos, além de identificar potenciais fontes de contaminantes (Sergeev et al., 2019; Tziachris et al., 2019; Zhang et al., 2020). O uso de algoritmos de aprendizagem de máquinas possibilita integrar múltiplas variáveis melhorando a predição. O uso de variáveis obtidas por meio de modelos digitais de elevação (DEM) e medidas por meio de fluorescência de raio-X, por exemplo, têm contribuído para as previsões de PTEs nos solos, pois, podem influenciar direta e indiretamente no comportamento do fenômeno em estudo e são de fáceis aquisições.

Depois de identificar as áreas contaminadas, é necessária a adoção de medidas remediadoras que visem reestabelecer as funções ecossistêmicas dos solos. Diversas técnicas de remediação têm sido utilizadas para estabilizar e recuperar áreas contaminadas por PTEs, como as técnicas físicas, químicas e biológicas (Kulshreshtha et al., 2014; Dhaliwal et al., 2020), porém, as técnicas físicas e químicas apresentam custos elevados e promovem danos a estrutura do solo (Dhaliwal et al., 2020). Já as técnicas de remediação biológica, que se dividem em Fitorremediação e Biorremediação, são eficientes em extrair e estabilizar metais, sustentáveis e podem ser utilizadas em grandes áreas devido ao baixo custo (Mukhopadhyay e Maiti 2010; Mota et al., 2016; Bhatti et al., 2018; Dhaliwal et al., 2020).

A fitorremediação consiste no uso de espécies vegetais capazes de absorver e estabilizar metais nos solos (Mukhopadhyay e Maiti 2010; Dhaliwal et al., 2020). Apesar de muito vantajosa, essa técnica apresenta algumas limitações, como encontrar espécies vegetais com capacidade de absorver PTEs, produzir grande volume de biomassa e serem adaptadas as condições edafoclimáticas específicas de cada ambiente contaminado (Bhatti et al., 2018; Dhaliwal et al., 2020; Egendorf et al., 2020). Dessa forma, estudos buscando identificar espécies vegetais com tais características é de fundamental importância para que essa técnica seja difundida nos programas de remediação de solos contaminados.

Diante disso, o objetivo desse estudo foi avaliar os níveis e a extensão da contaminação de elementos potencialmente tóxicos nos solos próximos a pilha de rejeito de mineração da cidade de Boquira-Ba, bem como avaliar os potenciais riscos à saúde da população (Capítulo 1), desenvolver uma metodologia rápida e eficiente para monitorar os níveis de contaminantes em áreas impactadas pela atividade de mineração (Capítulo 2) e identificar espécies vegetais capazes de remediar áreas contaminadas (Capítulo 3).

REFERÊNCIAS

Bhatti, S. S., Kumar, V., Sambyal, V., Singh, J., & Nagpal, A. K. (2018). Comparative analysis of tissue compartmentalized heavy metal uptake by common forage crop: a field experiment. *Catena*, 160, 185-193.

Carvalho, F. P. (2017). Mining industry and sustainable development: time for change. *Food and Energy Security*, 6(2), 61-77.

Conama - National Council for the Environment, Brazil, 2009. Resolution n° 420/2009. Available online at. <http://www.mma.gov.br/port/conama/legiabre.cfm?codlegi=620>, Accessed date: 16 January 2012 (in Portuguese).

Dhaliwal, S. S., Singh, J., Taneja, P. K., & Mandal, A. (2020). Remediation techniques for removal of heavy metals from the soil contaminated through different sources: a review. *Environmental Science and Pollution Research*, 27(2), 1319-1333.

Egendorf, S. P.; Groffman, P.; Moore, G.; Cheng, Z. (2020) The limits of lead (Pb) phytoextraction and possibilities of phytostabilization in contaminated soil: a critical review. *International journal of phytoremediation*, 22(9), 916–930.

Jin, M., Yuan, H., Liu, B., Peng, J., Xu, L., & Yang, D. (2020). Review of the distribution and detection methods of heavy metals in the environment. *Analytical Methods*, 12(48), 5747-5766.

Krishna, A. K., Mohan, K. R., Murthy, N. N., Periasamy, V., Bipinkumar, G., Manohar, K., & Rao, S. S. (2013). Assessment of heavy metal contamination in soils around chromite mining areas, Nuggihalli, Karnataka, India. *Environmental earth sciences*, 70(2), 699-708.

Kulshreshtha, A., Agrawal, R., Barar, M., & Saxena, S. (2014). A review on bioremediation of heavy metals in contaminated water. *IOSR Journal of Environmental Science, Toxicology and Food Technology*, 8(7), 44-50.

McGladdery, C., Weindorf, D. C., Chakraborty, S., Li, B., Paulette, L., Podar, D., ... & Duda, B. (2018). Elemental assessment of vegetation via portable X-ray fluorescence (PXRF) spectrometry. *Journal of Environmental Management*, 210, 210-225.

Mota, F. A. C.; Sargentini Júnior, E.; Santana, G. P. (2016). Potencial de fitorremediação da *Alocasia macrorrhiza* para Co, Cu, Ni e Zn. *Scientia Amazonia*, 5(3), 11-20.

Mukhopadhyay, S., & Maiti, S. K. (2010). Phytoremediation of metal mine waste. *Applied Ecology and Environmental Research*, 8(3), 207-222.

Qu, C., Wang, S., Ding, L., Zhang, M., Wang, D., & Giesy, J. P. (2018). Spatial distribution, risk and potential sources of lead in soils in the vicinity of a historic industrial site. *Chemosphere*, 205, 244-252.

Santos, N. L., dos Anjos, J. Â. S. A., & Klammler, H. (2020). Exposição da Zona Urbana de Boquira, estado da Bahia, aos Metais Tóxicos Associados à Bacia de Rejeito de Mineração Abandonada. *Anuário do Instituto de Geociências*, 43(3), 280-291.

Sergeev, A. P., Buevich, A. G., Baglaeva, E. M., & Shichkin, A. V. (2019). Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals. *Catena*, 174, 425-435.

Tziachris, P., Aschonitis, V., Chatzistathis, T., & Papadopoulou, M. (2019). Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. *Catena*, 174, 206-216.

USEPA. (1998). METHOD 3051a - Microwave assisted acid digestion of sediments, sludges, soils and oils. United States Environmental Protection Agency. <https://www.epa.gov/sites/production/files/2015-12/documents/3051a.pdf>.

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

Yang, S., Taylor, D., Yang, D., He, M., Liu, X., & Xu, J. (2021). A synthesis framework using machine learning and spatial bivariate analysis to identify drivers and hotspots of heavy metal pollution of agricultural soils. *Environmental Pollution*, 117611.

Zhang, C., & Yang, Y. (2020). Modeling the spatial variations in anthropogenic factors of soil heavy metal accumulation by geographically weighted logistic regression. *Science of the Total Environment*, 717, 137096.

CAPITULO 1

Modelos de aprendizagem de máquinas para mapeamento de elementos potencialmente tóxicos e avaliação de risco a saúde humana em solos afetados pela atividade de mineração

RESUMO

Os passivos ambientais deixados pela mineração, quando dispostos de maneira inadequada podem se dispersar no meio ambiente, aumentando os níveis de elementos potencialmente tóxicos (PTEs) nos solos. Assim, determinar os níveis e a distribuição espacial dos PTEs é essencial para traçar estratégias mitigadoras e remediar áreas contaminadas. Nessas áreas, previsões eficientes de PTEs são difíceis devido a elevada variabilidade espacial e fontes múltiplas de contaminação. Diante disso, o uso de algoritmos de aprendizagem de máquinas é uma alternativa para otimizar essas previsões. O objetivo do estudo foi modelar a distribuição espacial dos PTEs em solos ao entorno da pilha de rejeito de mineração e avaliar os riscos dessa para saúde humana. Nesse estudo, foi usado um conjunto total de 116 amostras de solos, coletadas na profundidade de 0-10 cm. O banco de dados de covariáveis foi composto por variáveis morfométricas derivadas de um Modelo Digital de Elevação com resolução espacial de 30 m e a criação de uma covariável de localização da pilha de rejeito. Os modelos testados foram o Cubist (CUB), Random Forest (RF), Suport Vector Machine Radial Sigma (SVM) e Linear Model (Lm). Com base no melhor modelo de previsão foram criados os mapas de distribuição espacial, os índices de poluição e de risco a saúde humana. O modelo SVM apresentou o melhor desempenho na previsão espacial dos teores de Pb e Zn, enquanto que para o Ba e Al o melhor desempenho foi obtido utilizando o algoritmo RF. Já para o Fe e Cr, o modelo CUB apresentou maior desempenho. Por meio do cálculo do fator de enriquecimento e do índice de geoacumulação de Pb e Zn é possível constatar que os solos próximos da pilha de rejeito estão extremamente contaminados, e apresentam riscos à saúde humana, principalmente para crianças.

Palavras-chaves: Índices de contaminação. Modelagem. Solos contaminados. risco cancerígeno

Machine learning models for mapping potentially toxic elements and assessing human health risk in soils affected by mining activity

ABSTRACT

Environmental liabilities left by mining, when improperly disposed of, can disperse into the environment, increasing the levels of potentially toxic elements (PTEs) in soils. Thus, determining the levels and spatial distribution of PTEs is essential to design mitigation strategies and remediate contaminated areas. In these areas, efficient prediction of PTEs is difficult due to high spatial variability and multiple sources of contamination. Therefore, the use of machine learning algorithms is an alternative to optimize these predictions. The aim of the study was to model the spatial distribution of PTEs in soils around the tailings pile and to assess its risks to human health. In this study, a total set of 116 soil samples, collected at a depth of 0-10 cm, was used. The covariate database was composed of morphometric variables derived from a Digital Elevation Model with a spatial resolution of 30 m and the creation of a tailing pile location covariate. The models tested were Cubist (CUB), Random Forest (RF), Support Vector Machine Radial Sigma (SVM) and Linear Model (Lm). Based on the best prediction model, maps of spatial distribution, pollution and risk to human health were created. The SVM model presented the best performance in the spatial prediction of Pb and Zn contents, while for Ba and Al the best performance was obtained using the RF algorithm. As for Fe and Cr, the CUB model presented greater performance. The enrichment factor and the geoaccumulation index of Pb and Zn show that the soils close to the tailings piles are extremely contaminated, and present risks to human health, especially for children.

Keywords: Contamination indices. Modeling. Contaminated soils. carcinogenic risk

1-0- INTRODUÇÃO

A mineração é uma atividade econômica e industrial que se baseia nas operações de exploração, extração e beneficiamento de minérios presentes no subsolo (Csavina et al., 2012; Carvalho, 2017). É de fundamental importância para o crescimento e desenvolvimento de um país, pois fornece a maior parte da matéria prima necessária para a construção civil e confecção de componentes eletrônicos (Measham e Zhang, 2009). Por outro lado, a atividade de mineração oferece potenciais riscos ao meio ambiente, na medida em que fornece fontes de contaminantes, mesmo após o encerramento das atividades de exploração mineral (Krishna et al., 2013), já que áreas próximas podem ser contaminadas por bacias e pilhas de rejeitos e escórias (Carvalho, 2017; Qu et al., 2018).

Dentre os principais contaminantes provenientes das atividades de minerações, podemos destacar os metais pesados, que são elementos potencialmente tóxicos, não biodegradáveis e com alta capacidade de se bioacumular e serem transferidos na cadeia trófica, acarretando sérios problemas ambientais e para saúde humana (Yang et al., 2021). Quando dispostos de maneira inadequada, as fontes de contaminação deixadas dentro e no entorno de minas metalíferas, estão sujeitas a processos erosivos, que favorece a dispersão de sedimentos contaminados em áreas circunvizinhas (Antoszczyszyn e Michalska, 2016).

Geralmente, PTEs ligados a sedimentos tendem a se dispersar pela ação eólica e pluvial, sendo carreados na forma de material particulado ou em solução aquosa por longas distâncias (Csavina et al., 2012). A extensão e o nível de contaminação desses sedimentos transportados dependem das características físicas e geoquímicas intrínsecas do material (Krishna, et al., 2013).

Entender as formas geoquímicas e a extensão da contaminação dos PTEs é de fundamental importância para traçar estratégias que visam mitigar essas áreas contaminadas, especialmente quando as operações de mineração ocorrem próximas a centros urbanos e áreas agrícolas (Ali e Khan, 2018). Além do acúmulo via cadeia trófica, a transferência dos PTEs para os seres humanos pode ocorrer diretamente através da inalação, contato dérmico, ingestão de partículas de solo e consumo de água contaminada (Zheng et al. 2010; Al osman et al. 2019). Em contato com o organismo humano, os PTEs geram diversos impactos negativos à saúde, como retardo mental, distúrbios neurocognitivos e comportamental em crianças, problemas respiratórios, cardiovasculares e carcinogênicos (Al osman et al. 2019; Mohammadyan et al., 2019).

Em diversas regiões do mundo há relatos de minas metalíferas e siderurgias em atividade e abandonadas, que são fontes efetivas de PTEs para as áreas circunvizinhas (Lima et al., 2018; Qu et al., 2018; Taylor et al., 2019). A exemplo disso tem-se a cidade de Boquira, localizada no centro-oeste do estado da Bahia, que durante as décadas de 1950 e 1990 foi sede de uma das mais relevantes minas de exploração de metais do Brasil, especialmente de chumbo e zinco, período no qual a cidade viveu grande prosperidade (Cunha et al., 2016). Porém, depois de encerradas as atividades de exploração, um grande volume de rejeito depositado de forma inadequada foi abandonado muito próximo à cidade e de áreas agricultáveis, sendo potencial risco de contaminação para o entorno. De acordo com Andrade et al. (2017) cerca de 300 estabelecimentos rurais estão dispostos em um raio de 5 km da bacia de rejeito. Nesse cenário, determinar a distribuição espacial desses contaminantes, com o objetivo de avaliar o risco humano e ecológico é de suma importância, visto que dados espaciais podem auxiliar na definição de áreas de risco elevado, possibilitando tomadas de decisão específicas para cada situação (Maas et al., 2010).

Muitos estudos têm utilizado técnicas geoestatísticas para espacializar os teores de PTEs nos solos, com o método da Krigagem, que consiste na estimativa de valores de variáveis distribuídas no espaço a partir de valores adjacentes, considerados como interdependentes pela análise variográfica (Cattle et al., 2002; Calderón, 2009; Maas et al., 2010; Lin et al., 2010; Souza et al., 2015). No entanto, esse método apresenta algumas limitações, como a necessidade de dependência espacial entre os pontos analisados, além de produzir um mapa espacial interpolado único, que geralmente suaviza os detalhes da incerteza nos resultados da previsão espacial (Chilès e Delfiner, 1999; Gharaibeh et al., 2020; Abbas et al 2021).

Assim, modelos preditivos, incluindo algoritmos de aprendizagem de máquina, como Random Forests, Support Vector Machines-SVM, Linear Model, Cubist e Redes Neurais têm sido utilizados para estimar a distribuição espacial de PTEs em solos, principalmente devido a sua ampla aplicabilidade, previsões precisas e eficientes e alta capacidade de relacionar múltiplas variáveis (Sergeev et al., 2019; Tziachris et al., 2019; Zhang et al., 2020a). O uso de variáveis auxiliares, como os modelos digitais de elevação (DEM) também têm contribuído para as previsões de PTEs nos solos, pois, podem influenciar direta e indiretamente no comportamento do fenômeno em estudo.

Dessa forma, os objetivos do presente estudo foram determinar as concentrações de chumbo (Pb), Zinco (Zn), Bário (Ba), Cromo (Cr), Ferro (Fe) e Alumínio (Al) (i), modelar a distribuição espacial dos teores desses PTEs em solos no entorno da pilha de rejeito de mineração (ii), determinar quais variáveis contribuem para melhor desempenho das previsões

(iii) selecionar o melhor modelo para previsão dos teores de PTEs nos solos (iv), avaliar os níveis de contaminação dos solos por meio do fator de enriquecimento e do índice de geoacumulação (v) e avaliar o risco à saúde humana devido à exposição a solos contaminados por meio dos índices de risco (vi). A hipótese é que os passivos ambientais deixados pela atividade de mineração vêm contribuindo para elevar os teores de PTEs, causando potenciais riscos à saúde humana e a todo ecossistema.

2.0- MATERIAIS E MÉTODOS

2.1- Área de estudo

A área de estudo está localizada na bacia do Rio São Francisco, no município de Boquira-BA, com população de 22.042 habitantes (Figura 1). O município abrange uma área de 1.570,1 km² e situa-se na porção Centro-Oeste do estado da Bahia, entre a Serra do Espinhaço e o vale do Rio Paramirim, (Santos, 2014).

O clima varia entre sub-úmido a seco, com precipitação média anual de 894,8 mm, e duas estações bem definidas (inverno e verão), com chuvas concentradas nos meses de outubro a abril (SEI, 2011). A vegetação do município é de Caatinga Arbórea Aberta, sem palmeiras, com zonas de contato entre Cerrado-Caatinga-Floresta Estacional (SEI, 2011).

As principais classes de solos ocorrentes na área de estudo estão associadas a duas unidades geológicas, o vale do Paramirim e o domínio da Serra do Espinhaço. No domínio do vale do Paramirim as principais classes de solos são: os Latossolos Vermelhos Eutróficos que estão associados às rochas vulcanossedimentares básicas e ultrabásicas; Latossolos Vermelho-Amarelos Distróficos que estão associados aos depósitos detríticos neogênicos, de natureza argilo-arenosa e areno-argilosa; Neossolos Litólicos Eutróficos ocorrentes em áreas planas de gnaisses, migmatitos e granitos; e Neossolos Litólicos com apenas os horizontes A e C ou somente o horizonte A sobre a rocha matriz (RADAMBRASIL, 1982). No domínio da Serra do Espinhaço as classes de solos observadas são; os Neossolos Litólicos que estão associados a terrenos de quartzitos, filitos e metarenitos, diferindo dos solos litólicos observados no vale do Paramirim por sua menor saturação em bases e maior quantidade de alumínio trocável (RADAMBRASIL, 1982). Os Latossolos Vermelho-Amarelos que estão associados às coberturas detríticas areno-argilosas, apresentam características semelhantes aos solos formados sob as coberturas detríticas do vale do Paramirim, diferindo apenas pela maior quantidade de alumínio trocável, que denomina a unidade (RADAMBRASIL, 1982).

A cidade de Boquira constituiu um importante núcleo mineral durante a década de 1970, cessando as atividades em 1992. As jazidas de minério desta região tinham como principais componentes metálicos Pb, Zn e prata (Ag) (Arcanjo et al., 2005). O Pb estava presente principalmente em minerais como a cerussita, smithsonita, goethita e anglesita e com proporções menores nas piromorfita, hemimorfita, hidrozincita, crisocola, bornita, covellita, malaquita e azurita os quais estavam associadas as formações férricas (Alves et al., 2017).

Como consequência do longo período de atividade de mineração, grandes volumes de resíduos foram gerados e depositados em uma pilha de rejeito muito próxima do centro da cidade. Além disso, as áreas mineradas não foram recuperadas e também ocorrem diversos pontos com deposição de escórias de mineração.

De acordo com Santos (2014), a pilha de rejeito por estar disposta de forma inadequada, vem sofrendo danos promovidos pela erosão pluvial, a qual foi intensificada pela retirada da vegetação e ausência de barreiras de contenção. Além disso, o autor relata que a inclinação e a instabilidade dos taludes da pilha, permitem que o rejeito se movimente por gravidade chegando a áreas localizadas fora dos limites da pilha. Além da contaminação de áreas a jusante da pilha de rejeito, provocada pela ação pluvial, ocorre também a dispersão atmosférica do material fino constituinte da pilha de rejeito, a qual alcança longas distâncias, podendo contaminar solos e principalmente os bairros mais próximos. Um agravante que também ocorreu na área de estudo foi a instalação do lixão municipal em cima da pilha de rejeito, o qual também pode contribuir para dispersão de PTEs na área.

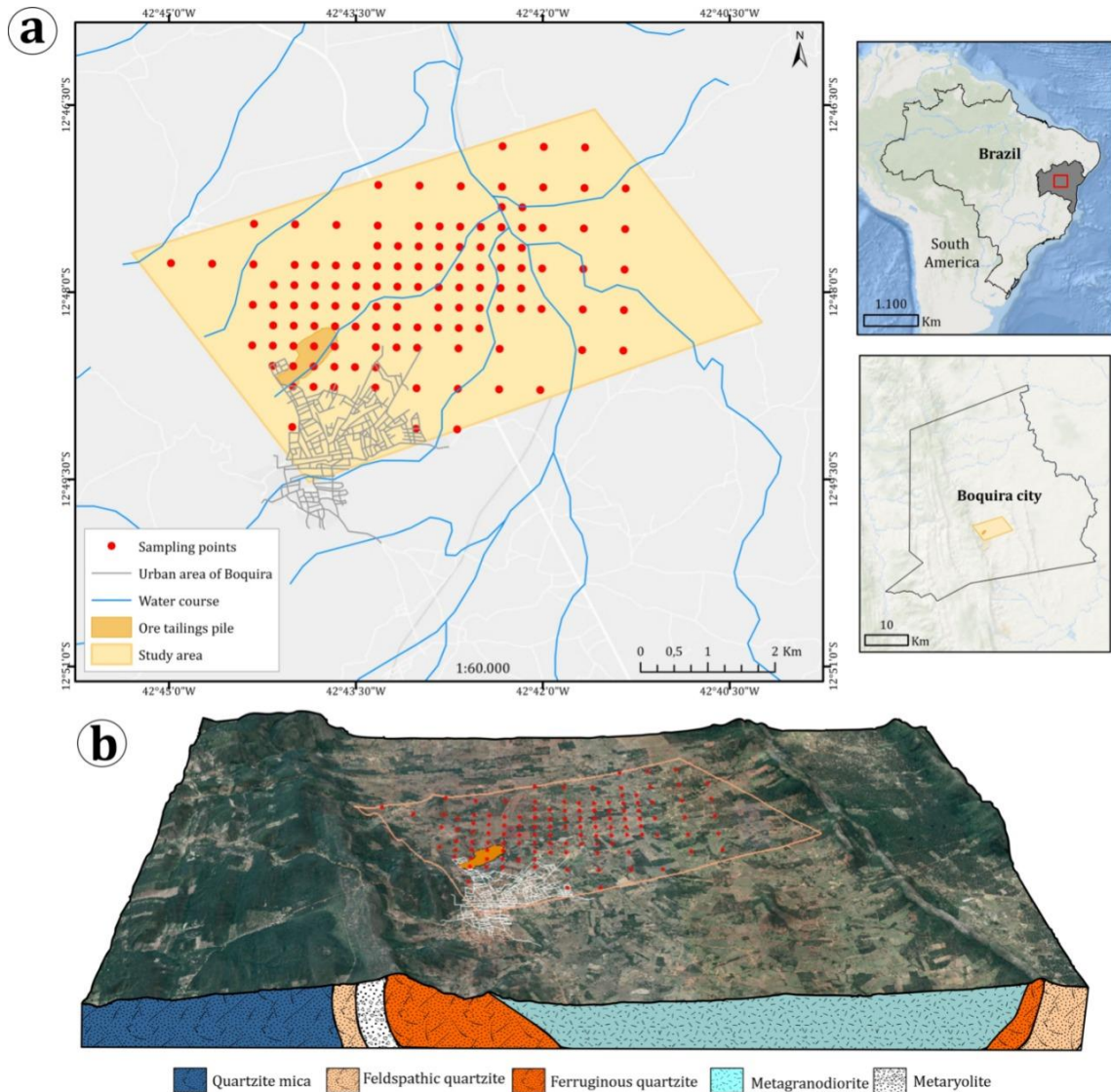


Figura 1. a) Localização da região do município de Boquira, com detalhe da pilha de rejeito, pontos de amostragem e área urbana b) Seção transversal dos litotipos e topografia associados a região do município de Boquira-Ba.

2.2- Amostragens de solo

Os pontos foram sistematicamente distribuídos ao entorno da pilha de rejeito de mineração englobando áreas agricultáveis e parte da cidade de Boquira-Ba, com base em uma grade regular de 300 em 300 m nas áreas mais próximas e de 300 por 450 m nas áreas mais distantes da pilha de rejeito. O grid de amostragem foi pré-definido em laboratório, com base na disposição da pilha de rejeito. Nesse grid foram determinadas as coordenadas geográficas em cada nó da grade na qual foi realizada a amostragem de solo. Em cada nó da grade amostral foram coletados 05 subamostras para obter uma amostra composta representativa. As subamostras foram coletadas no centro e em cada direção cardinal (N, S, L, O) a uma distância de cerca de 3 a 5 m em torno do ponto central (Abrahão e Marques, 2013). Como na área de

estudo ocorrem desde classes de solos com horizontes muito rasos a solos com horizontes espessos, a profundidade de coleta das amostras foi padronizada de 0,0 - 0,10 m. Para a coleta das amostras foram utilizados equipamentos de aço inox, a fim de evitar possíveis contaminações (Abrahão e Marques, 2013). A área total amostrada foi de 4,0 km de largura e 7,0 km de comprimento, totalizando 116 amostras compostas em uma área de aproximadamente 28,0 km². As amostras foram secas ao ar e destorroadas em peneira de malha de 2 mm, com posterior armazenamento em sacos de plástico até análise.

2.3- Caracterizações física e química

As amostras de solo secas ao ar e passadas em peneira de 2 mm foram submetidas a análises químicas e físicas de acordo com Teixeira et al. (2017). O pH foi determinado em H₂O, na proporção de 1:2,5 usando um medidor de pH de bancada digital (pHmetro). O P, Na⁺ e K⁺ foram extraídos com o extrator Melich-1 (HCl 0,05 mol L⁻¹ + H₂SO₄ 0,0125 mol L⁻¹) e determinados por fotocolorimetria e fotometria de chama, respectivamente. O Ca²⁺ e Mg²⁺ trocáveis foram extraídos com KCl 1 mol L⁻¹ e determinados por espectrofotometria de absorção atômica. Acidez trocável (M^{x+}) foi extraída com KCl 1 mol L⁻¹ e quantificada por titulometria com hidróxido de sódio 0,025 mol L⁻¹. A H+Al foi extraído utilizando o extrator Ca(OAc)₂ 0,5 mol L⁻¹ tamponado a pH 7,0 e quantificado por titulação com NaOH 0,025 mol L⁻¹. Os teores de Fe, Zn, Cu e Mn disponíveis foram extraídos utilizando o extrator Melich-1 e dosados por espectrofotometria de absorção atômica. O teor de C-org foi determinado via oxidação com dicromato (0,167 mol L⁻¹ de K₂Cr₂O₇) em meio ácido (Walkley e Black), utilizando uma fonte externa de calor para maximizar a oxidação (Yeomans e Bremner, 1988). A partir dos valores de K⁺, Na⁺, Ca²⁺, Mg²⁺, Al³⁺ e H+Al foram calculados os índices soma de bases (SB), a capacidade de troca de cátions efetiva (t) e potencial (T), a porcentagens de saturação por bases (V) e por alumínio (m).

A determinação granulométrica do solo foi realizada em amostras secas e peneiradas (2 mm) pelo método de pipeta (Gee e Bauder, 1986), utilizando solução de NaOH 0,1 mol L⁻¹ como dispersante, e agitando a 50 rpm por 16 h. As frações areia grossa (2 - 0,212 mm) e areia fina (0,212 - 0,063 mm) foram obtidas por peneiramento, a argila por sedimentação e o silte por diferença da fração total em relação a areia mais argila. A densidade de partícula foi determinada usando o método do balão volumétrico (Teixeira et al. 2017).

2.4- Determinações dos teores de PTEs, elementos principais e controle de qualidade

Amostras de solo foram maceradas em almofariz de ágata e peneiradas em malha de 200 mesh. Em seguida foram pesados 0,5000 g de amostras e digeridas em forno de micro-ondas (Anton Paar, Multiwave 3000) com uma mistura de 9 ml de HNO_3 e 3 ml de HCl de acordo com o método 3051a, digestão ácida assistida por micro-ondas de sedimentos, lamas, solos e óleos (Usepa 1998). As soluções foram preparadas utilizando água ultrapura (18,2 $\text{M}\Omega\cdot\text{cm}$ a 25°C - Millipore Direct-Q System) e ácidos de alta pureza (Merck PA®). O procedimento de digestão foi realizado em triplicata para cada amostra de solo.

A dosagem dos elementos Ba, Cr, Zn, Pb, Fe e Al foram determinadas por espectrometria de emissão óptica com plasma indutivamente acoplado (ICP-OES). Visando o controle de qualidade das análises, as curvas de calibração foram preparadas utilizando padrões de alta qualidade (Merck PA®). Além disso, um ponto de verificação foi preparado, sendo que a cada conjunto de 10 leituras o mesmo era dosado e quando apresentava diferença acima de 15 % às curvas de calibração eram reanalisadas e recalibradas.

Os limites de detecção (DL) e limites de quantificação (QL) foram calculados por meio das equações $\text{DL} = 3,3s/b1$ e $\text{QL} = 10s/b1$ (ANVISA, 2003), onde s é o desvio de referência da leitura de 10 brancos e $b1$ é o coeficiente angular da reta obtido para cada elemento. Os DL e QL são apresentados na Tabela S1. A análise de controle de qualidade foi realizada com materiais certificados SRM 2710a - Montana Soil I e SS-1 Contaminated Soil - EnviroMAT™ & AgroMAT™. As taxas de recuperação do SRM 2710a foram de 65 (Ba), 72 (Cr), 87 (Fe), 87 (Zn), 88 (Al) e 90 % (Pb), enquanto para o SS-1 foram de 84 (Ba), 88 (Cr), 96 (Zn), 99 (Pb), 101 (Al) e 109 % (Fe).

2.5- Obtenção de covariáveis ambientais para espacialização da previsão dos PTEs

2.5.1- Coordenadas geográficas dos pontos amostrais

A amostragem dos solos foi realizada com base em um grid amostral de 300 em 300 m nas áreas mais próximas a pilha de rejeito e de 300 por 450 m nas áreas mais distantes. O grid foi pré-definido em laboratório. As coordenadas (X e Y) utilizadas para a realização da amostragem dos solos também foi utilizada como covariáveis para espacialização dos teores de PTEs nos solos. O sistema de coordenadas projetadas SIRGAS 2000/UTM zone 23S (EPSG:31983) foi usado como padrão durante o processo de espacialização dos PTEs.

2.5.2- Variáveis obtidas por meio do MDE

As variáveis topográficas foram derivadas a partir do modelo digital de elevação (DEM) SRTM - Shuttle Radar Topography Mission, com resolução espacial de aproximadamente 30 m (30 segundos de arco). Para a obtenção das variáveis, foi utilizado o *software* R (R Core Team, 2021), utilizando os pacotes “rsaga” (Brenning, 2008), “raster” (Robert, 2019) e “rgrass7” (Bivand et al., 2019; Sena et al., 2020). Os atributos gerados com base no DEM são descritos na tabela 1.

Tabela 1. Variáveis morfométricas obtidas gerada por meio de um modelo de elevação digital

Atributo de terreno	Abreviação	Breve descrição
Convergence index	CI	Índice de convergência / divergência em relação ao escoamento
Cross sectional curvature	CSC	Mede a curvatura perpendicular à direção do declive
Digital elevation model	DEM	Representa a elevação em cada célula do modelo
Diurnal anisotropic heating	DAH	Medição contínua de energia dependente da exposição
Flow line curvature	FLC	Representa a projeção de uma linha gradiente em um plano horizontal
General curvature	GC	A combinação das curvaturas do plano e do perfil
Geomorphons	GEO	Geometria associada usando abordagem de visão de máquina
Hill	HI	Sombreamento analítico de colina
Hill index	HIINDEX	Sombreamento de colina de índice analítico
Longitudinal curvature	LC	Mede a curvatura na direção do declive
Mass balance index	MBI	Índice de equilíbrio entre erosão e deposição
Maximal curvature	MAXC	Curvatura máxima na seção normal local
Morphometric Protection Index	MPI	Medida de exposição / proteção de um ponto do relevo circundante
Mid-slope position	MSP	Representa a distância do topo ao vale, variando de 0 a 1

Minimal curvature	MINC	Curvatura mínima para seção normal local
Multiresolution index of ridge top flatness	MRRTF	Indica posições planas em áreas de alta altitude
Multiresolution index of valley bottom flatness	MRVBF	Indica superfícies planas no fundo do vale
Normalized height	NH	Distância vertical entre a base e o cume da inclinação normalizada
Plan curvature	PLANC	Descrita como a curvatura da linha de contorno hipotética que passa por uma célula específica
Profile curvature	PROC	Descreve a curvatura da superfície na direção da inclinação mais íngreme
Real surface area	RSA	Cálculo real da área da célula
Slope	S	Representa inclinação angular local
Slope height	SH	Distância vertical entre a base e o cume da encosta
Slope Index	SI	Representa um índice de inclinação angular local
Solrad Diffuse1	SolDiffuse1	Insolação Difusa para o mês de janeiro
Solrad Diffuse2	SolDiffuse2	Insolação Difusa para o mês de julho
Solrad dur 1	SolDur1	Duração da insolação para o mês de janeiro
Solrad dur 2	SolDur2	Duração da Insolação para o mês de julho
Solrad Direct1	SolDiret1	Insolação Direta do mês de janeiro
Solrad Direct2	SolDiret2	Insolação Direta do mês de julho
Solrad Ration1	SolRation1	Razão entre a Insolação Direta e a Insolação Difusa do mês de Janeiro
Solrad Ration2	SolRation2	Razão entre a Insolação direta e a Insolação difusa do mês de julho
Solrad Sunrise1	SolSunrise1	Horário médio do nascer do sol do mês de janeiro
Solrad Sunrise2	SolSunrise2	Horário médio do nascer do sol do mês julho
Solrad Sunset1	SolSunset1	Tempo médio do pôr do sol do mês de janeiro
Solrad Sunset2	SolSunset2	Tempo médio do pôr do sol do mês julho
Solrad total1	SolTotal1	Insolação total para o mês de janeiro

Solrad total2	SolTotal2	Insolação total para o mês de julho
Standardized height	STANH	Distância vertical entre a base e o índice de inclinação padronizado
Surface specific points	SSP	Indica diferenças entre pontos de deslocamento de superfície específicos
Tangencial curvature	TANC	Medido no plano normal em uma direção perpendicular ao gradiente
Terrain ruggedness index	TRI	Índice quantitativo de heterogeneidade topográfica
Terrain surface convexity	TSC	Razão entre o número de células que têm curvatura positiva e o número de todas as células válidas dentro de um raio de pesquisa especificado
Terrain surface texture	TST	Divide a textura da superfície em 8, 12 ou 16 classes
Total curvature	TC	Medida geral da curvatura da superfície
Topographic position index	TPI	Diferença entre a elevação do ponto a com a elevação circundante
Valley depth	VD	Cálculo da distância vertical ao nível da base de drenagem
Valley	VA	Vale nebuloso de cálculo usando a abordagem Cartola
Valley Index	VA	Cálculo do índice de vale difuso usando a abordagem Cartola
Vector ruggedness measure	VRM	Mede a variação da rugosidade do terreno
Topographic wetness index	TWI	Descreve a tendência de cada célula em acumular água em função do relevo

2.5.3- Criação da covariável de localização da pilha de rejeito e do canal drenagem (PDC)

A pilha de rejeito de Pb foi depositada no canal de drenagem principal, e devido a processos erosivos ocorreu transporte de elevada quantidade de rejeito ao longo do canal, a jusante da pilha, o que elevou drasticamente os teores de PTEs em relação as demais áreas. Para melhorar a predição dos teores de PTEs, foi criada uma variável de localização da pilha de rejeito juntamente com o canal de drenagem principal, classificando a mesma em 1 e as demais áreas como 0 (PDC). Essa covariável foi utilizada para separar a área em duas classes, a

primeira onde se encontra a pilha de rejeito e o canal de drenagem principal, na qual os teores de PTEs são muito elevados devido a deposição direta de rejeito e a segunda área onde os teores de PTEs são inferiores pois não houve deposição direta de rejeito (Figura 2).

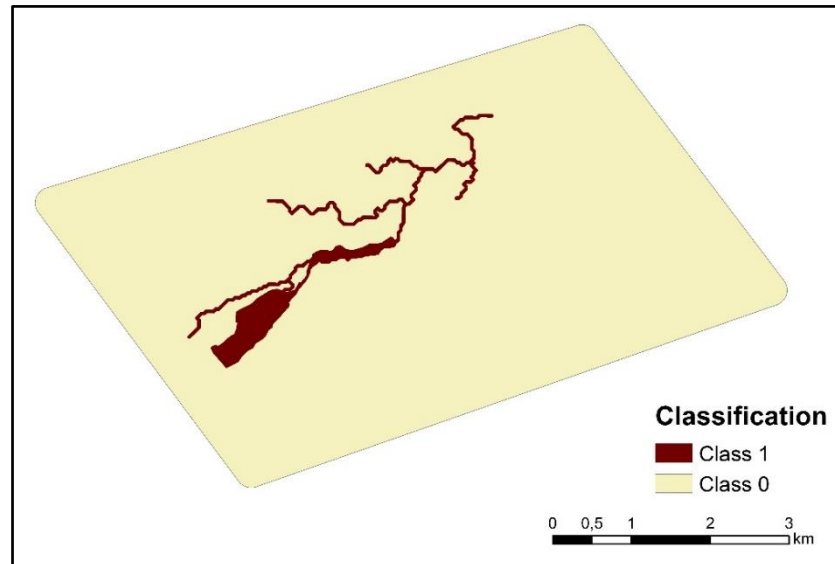


Figura 2. Covariável de localização. Área classificada em 1 (pilha de jeito e canal de drenagem principal) e 0 (demais área amostrada).

2.6- Criação da função de pedotransferências por aprendizagem de máquinas

As funções de pedotransferências são funções preditivas de determinadas propriedades dos solos com base em atributos facilmente obtidos e com custos inferiores (Budiman et al., 2002; McBratney et al., 2002; Kotlar et al., 2020). Atualmente com a grande disponibilidade de dados, e também da facilidade de obtenção deles por meio de sensores proximais e imagens de satélites, a criação de funções de pedotransferências tornou-se uma alternativa eficiente para prever diversas propriedades dos solos.

2.6.1- Remoção de covariáveis e treinamento dos algoritmos de aprendizagem de máquinas

O processo de remoção de covariáveis têm por objetivo o desenvolvimento de um modelo/função que demande um menor custo computacional, além de obedecer ao princípio da parcimônia (Gomes et al., 2019; Souza et al., 2020, Ferreira et al., 2021; Reis et al., 2021). A remoção das covariáveis foi realizada em três etapas: remoção pela variância (I), por correlação (II) e pela importância na predição para cada algoritmo (III)

A remoção pela variância (I) tem como fundamento retirar as covariáveis que apresentam muito baixa variância, pois elas não contribuem para o modelo de predição, e ainda podem afetar a performance de alguns modelos (Hujoel et al., 2018, Kern et al., 2017; Armitage e Ober, 2010). Para a remoção das covariáveis de muito baixa variância foi utilizada a função *nearZeroVar* disponível no pacote *Caret* (Kuhn, 2008).

A remoção de covariáveis por correlação (II) tem como princípio retirar covariáveis altamente correlacionadas e com isso proporcionar um menor custo computacional e aumentar a parcimônia dos modelos, sem perdas de performance. Para a remoção das covariáveis altamente correlacionadas, primeiramente calcula-se a correlação de Spearman para o conjunto de todas as covariáveis. Em seguida as duplas de covariáveis que apresentarem correlação superior ou igual a 95 % são avaliadas e removidas aquelas que apresentarem maior valor da soma absoluta de correlação com as demais variáveis. As covariáveis altamente correlacionadas foram eliminadas usando a função *findcorrelation*, disponível no pacote *Caret do software R* (Kuhn et al., 2020).

Após eliminar as covariáveis de muito baixa variância e covariáveis altamente correlacionadas, o conjunto total de dados foram separados aleatoriamente em 75 % para treinamento e 25 % para teste (*repeated hold out test*). Esse processo foi realizado 100 vezes usando a seleção aleatória de amostras de treinamento e teste (Shortridge et al., 2016; Gomes et al., 2019; Souza et al., 2020; Reis et al., 2021). O processo de repetição múltipla é importante, pois determina a variabilidade da predição, uma vez que diferentes grupos de conjuntos de dados de treinamento e teste geram modelos diferentes que podem gerar diferentes valores de métricas de precisão (Kuhn e Johnson, 2013).

A remoção das covariáveis pela importância da covariável para cada modelo avaliado (III) tem como principal objetivo o aumento da parcimônia do modelo determinando um conjunto ideal de covariáveis para o modelo final gerado por um algoritmo. A eliminação das covariáveis é definido pelo nível da importância da covariável na predição do elemento, sendo realizada usando a ferramenta Recursive Feature Elimination – RFE do pacote *Caret* (Kuhn et al., 2013). O RFE é um método de remoção/seleção de covariáveis iterativo do tipo Backforward, que define o conjunto ideal de covariáveis avaliando diferentes subgrupos com diferente número de covariáveis (“subsets”). Esses são definidos pelo operador e também pelo conjunto completo, sempre avaliando do maior conjunto (todas as covariáveis - “set”) até o menor conjunto (Ferreira et al., 2021). Por apresentar características distintas, é necessário a execução do RFE em cada algoritmo que se deseja avaliar. Os subgrupos definidos foram compostos por 5, 20, 25 e 30 covariáveis, e também foi testado o conjunto completo de

covariáveis que passaram na etapa de eliminação por variância muito baixa e por correlação (38). Os modelos testados foram: Cubist (CUB), Random Forest (RF), Suport Vector Machine Radial Sigma (SVM) e Linear Model (Lm). Cada algoritmo apresenta uma metodologia de cálculo de importância e as funções usadas estão listadas e consideradas “funções de apoio” do RFE (Kuhn e Johnson, 2013).

O treinamento para cada elemento foi feito usando o conjunto ideal de covariáveis para cada algoritmo escolhidos na fase 3. A otimização do modelo final foi feita usando o método validação cruzada repetida com 10 grupos (repeatedcv -10 folds) e testando 5 valores de hiperparâmetros internos de cada algoritmo avaliado. O parâmetro de performance para escolha do valor ótimo dos hiperparâmetros foi o MAE. Os hiperparâmetros de cada algoritmo estão descritos no manual do pacote *Caret* no capítulo 6, (“Models described” available at <https://topepo.github.io/caret/train-models-by-tag.html>). O modelo final gerado para cada algoritmo foi usado para predição do mapa dos teores dos elementos avaliados na área de estudo (Figura 3).

O conjunto de dados da amostra de teste (25 % da amostra total) para cada algoritmo foi usado para determinar a qualidade da predição do modelo usando métricas de performance. As métricas de performance avaliados neste trabalho foram o erro médio absoluto - MAE (EQ.1), Raiz do Erro Quadrático Médio – RMSE (EQ.2) e coeficiente de correlação - R^2 (EQ. 3).

$$MAE = \frac{\sum_{i=1}^N |P_i - O_i|}{n} \quad (\text{EQ. 1})$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{\frac{1}{2}} \quad (\text{EQ. 2})$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - \overline{om})^2}{\sum_{i=1}^n (O_{mi} - \overline{om})^2} \quad (\text{EQ. 3})$$

Em que: P_i = valores das amostras previstas; O_i = valores das amostras observadas; om = média dos valores observados; n = número de amostras.

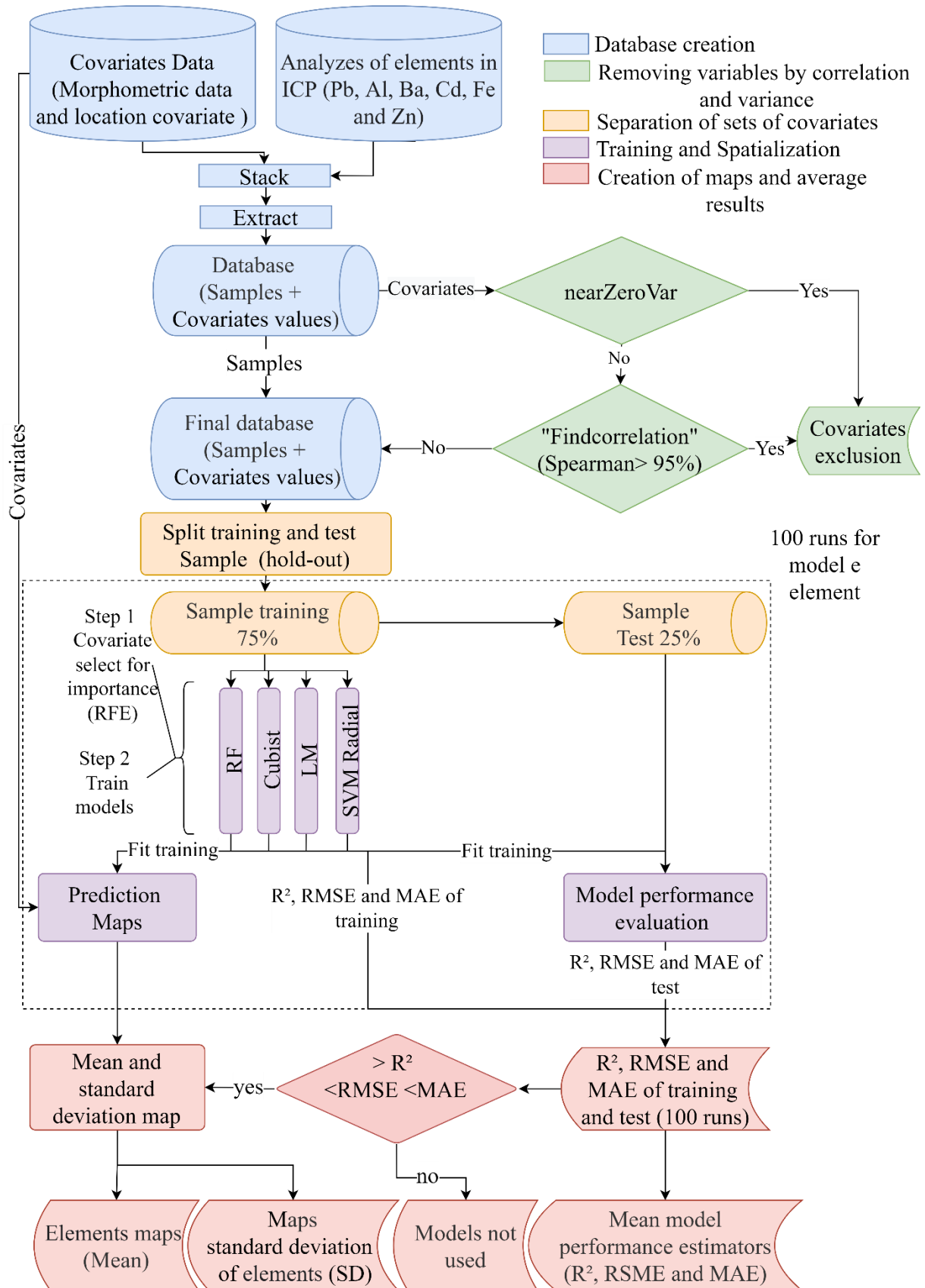


Figura 3. Fluxograma metodológico mostrando a sequência de metodologias aplicadas para a previsão espacial de elementos potencialmente tóxicos (PTEs) em solos.

Com objetivo de comparar os resultados gerados para cada elemento, foi calculado o RMSE e MAE para o modelo nulo (RMSE_NULL, MAE_NULL) (EQ.4 e EQ 5). Esse modelo é considerado um dos mais simples que pode ser ajustado aos dados, pois considera utilizar o valor médio quantificado pelas amostras coletadas. O modelo nulo simula outros modelos, porém, sempre retorna o mais simples possível para o conjunto de treinamento (Mello et al., 2022). Quando um algoritmo prediz pior que o modelo nulo, ele deve ser desconsiderado, sendo eliminado antes mesmo de uma comparação com os resultados de outros algoritmos para o mesmo atributo.

$$NULL_RMSE = \left[\frac{1}{N} \sum_{i=1}^N \left(Om_t - Oi \right)^2 \right]^{\frac{1}{2}} \quad (EQ.4)$$

$$NULL_MAE = \frac{\sum_{i=1}^N |Om_t - Oi|}{n} \quad (EQ.5)$$

Em que: Om_t = média das amostras de treinamento; Oi = amostra de teste; n = número de amostras (loop).

2.6.2- Previsão dos mapas finais e incerteza dos mapas de teores PTEs nos solos

O processo de seleção pela importância (RFE), treinamento, predição das amostras teste e predição dos mapas, foi repetido 100 vezes. Esta técnica visa avaliar a variabilidade dos modelos em prever os teores dos elementos, visto que a utilização de conjunto de treinamento diferente gerará modelos que resultam em distintos resultados de precisão e mapas (Gomes, et al, 2019; Siqueira et al, 2021; Dias et al, 2021).

Para elaborar os mapas de previsão dos teores de PTEs nos solos foi selecionado entre os cinco algoritmos avaliados, aquele que apresentou o melhor desempenho com base no valor médio do MAE das 100 execuções. Com base nos 100 mapas previstos dos teores PTEs, foi calculado a média e desvio padrão das previsões para cada pixel da área de estudo e foram gerados mapas de média e desvio respectivamente. Esses mapas foram utilizados para discussão dos resultados. A importância das covariáveis também foi calculada usando os resultados das 100 execuções, sendo calculada a média da importância da covariável para cada modelo/elemento.

2.7- Avaliação ambiental de PTEs em solos

A toxicidade dos PTEs presentes nos solos em estudo foi avaliada de acordo com os valores norteadores para solos da Resolução N° 420/2009 do Conselho Nacional do Meio

Ambiente, o qual possui valores de referência para prevenção e investigação de solos agrícolas, urbanos e industriais e pelos valores de referência de qualidade de solos para o estado de Minas Gerais (COPAM, 2011; Tabela S2).

Para distinguir o possível impacto antrópico dos solos em comparação com solos naturais, foi calculado o fator de enriquecimento (EF) (Yan et al. 2018) e o índice de geoacumulação (Igeo) (Mueller, 1979) de acordo com as equações a seguir:

$$\text{Eq.1} \quad \text{EF} = \frac{\text{CN/CR}}{\text{BN/BR}}$$

$$\text{Eq.2} \quad \text{Igeo} = \log_2 \frac{\text{CN}}{1,5 * \text{BR}}$$

Em que, *CN* é a concentração do elemento *N* no solo investigado e *BN* é o valor de background geoquímico do elemento *N*. *CR* e *BR* são a concentração do elemento de referência *R* no solo investigado e os valores regionais de background, respectivamente. Normalmente o *BN* utilizado é o valor da média mundial dos solos, porém, valores universais podem distorcer da realidade regional, visto que cada solo apresenta materiais de origem com composição elementar diferentes. Portanto, os valores de referência de qualidade considerados como a quantidade do elemento naturalmente presente no solo regional são mais adequados para utilização de valores de background regionais. Como o estado da Bahia ainda não apresenta valores de referência de qualidade utilizou-se os definidos por Minas Gerais, estado brasileiro com maiores quantidades de amostras para elaboração dos VRQs e com características semelhantes aos solos da área de estudo (COPAM, 2011; Souza et al. 2015). Como elemento de referência, foi utilizado o Fe visto que é o elemento presente no solo regional, está presente nas frações mais finas, estável no solo, sua concentração natural nos solos é mais uniforme e tem comportamento geoquímico análogo aos PTEs (Arcanjo et al., 2005; Costa et al., 2019). O grau de contaminação do solo determinado por EF e Igeo foi classificado usando os critérios de classificação estabelecidos por Wu et al. (2014; Tabela 2).

Tabela 2: Classificação de diferentes modelos de avaliação de contaminação do solo.

Class	EF ¹	Igeo ¹	Descrição das Classes
1	<1	<0	Natural
2	1 - 5	0 - 1	Contaminação moderada
3	5 - 20	1 - 3	Contaminação considerável
4	20 - 40	3 - 5	Alta contaminação
5	>40	>5	Contaminação extrema

¹Wu et al. (2014).

Visando à avaliação da distribuição espacial dos índices Igeo e FE os cálculos acima foram realizados com base nos valores previsto nos mapas de distribuição obtido pelo melhor modelo de predição para cada elemento estudado.

2.8- Avaliação de risco para a saúde humana

A avaliação de risco à saúde foi usada para estimar os efeitos da exposição de PTEs aos seres humanos. Para quantificar o risco não carcinogênico e o risco carcinogênico da exposição de adultos (A) e crianças (C) a solos contaminados por PTEs, duas principais vias de exposição foram determinadas (ingestão e dérmica) de acordo com o método da Agência de Proteção Ambiental dos Estados Unidos (USEPA 1989, 1998, 2004, 2007). As formulas usadas para avaliar a exposição por meio da exposição à ingestão (ADD_{ingestão}) e dérmico (ADD_{dérmico}) foram:

$$\text{Eq. 3} \quad \text{ADD}_{\text{ingestão}} = \frac{C \cdot IR \cdot Ef \cdot ED(C, A)}{Bw(C, A) \cdot At(C, A)} \cdot 10^{-6}$$

$$\text{Eq. 4} \quad \text{ADD}_{\text{dermal}} = \frac{C \cdot SA(C, A) \cdot SL(C, A) \cdot ABS \cdot Ef \cdot ED(C, A)}{Bw(C, A) \cdot At(C, A)} \cdot 10^{-6}$$

Os valores dos parâmetros utilizados nas equações acima estão descritos na Tabela (S3). A avaliação do quociente de perigo (risco não cancerígeno - HQ) para as pessoas que vivem na área de estudo foi determinada pela razão entre a dose de exposição ambiental e a dose de referência correspondente de cada metal em estudo (RfD) (USEPA 2007; USEPA 2002).

Em seguida, o índice de perigo (HI), definido como o total de potenciais efeitos não cancerígenos apresentados pelas vias de exposição, foi obtido por meio da soma do HQ_{ingestão} com HQ_{dérmico} de cada elemento. O HQ e o HI foram determinados de acordo com as Eq. (3) e (4).

$$\text{Eq. 5} \quad \text{HQ} = \frac{\text{ADD}}{\text{RfD}}$$

$$\text{Eq. 6} \quad \text{HI} = \text{HQ}_{\text{ingestão}} + \text{HQ}_{\text{dérmico}}$$

Para a determinação do potencial risco cancerígeno (CR), utilizaram-se as equações descritas abaixo de acordo com USEPA (2004):

$$\text{Eq. 7} \quad \text{CR}_{\text{ingestão}} = \text{ADD}_{\text{ingestão}} \cdot \text{SF}$$

$$\text{Eq. 8} \quad \text{CR}_{\text{dermico}} = \text{ADD}_{\text{dermico}} \cdot \text{UR}$$

Em que o SF (mg kg⁻¹ dia⁻¹) é definido como o fator de inclinação do elemento carcinogênico e UR (µg m⁻³) como o risco da unidade de inalação (OEHHA 2009; USEPA

1989). Para a obtenção do risco carcinogênico total (CR_{total}), as vias cancerígenas (CR_{ingestão} + CR_{dérmico}) foram somados. Dentre os elementos avaliados neste estudo, apenas Cr e Pb são considerados contaminantes carcinogênicos (OEHHA, 2009; USEPA 1989).

Para espacialização dos índices HI e CR os cálculos acima foram realizados com base nos valores previsto nos mapas de distribuição obtido pelo melhor modelo de predição para cada elemento estudado.

3.0- RESULTADOS

3.1- Remoção de Covariáveis

O processo de remoção de covariáveis objetiva o desenvolvimento de um modelo/função que demande um menor custo computacional, obedecendo o princípio da parcimônia (Gomes et al., 2019; Souza et al., 2020, Ferreira et al., 2021; Reis et al., 2021). Na primeira etapa denominada remoção pela variância foram excluídas quatro covariáveis, enquanto que na remoção por correlação foram removidas 12 covariáveis. Do total de 54 covariáveis apenas 38 chegaram na etapa de remoção pela importância (Tabela S4). Na terceira etapa de remoção pela importância os modelos utilizados removeram grupos distintos de covariáveis, visto que utilizam grupos de amostras diferentes em cada execução, gerando conjuntos de covariáveis ideais para predição dos teores espacializados de PTEs. As covariáveis selecionadas na terceira etapa para predição são apresentadas no tópico importância das covariáveis.

3.2- Propriedades químicas e textura dos solos estudados

As propriedades químicas e físicas dos solos da área estudada apresentaram uma heterogeneidade elevada, o que era esperado devido à área englobar uma grande diversidade de solos, formações geológicas e diferentes usos e manejo (Tabela 3). Os solos apresentaram uma classificação textural variando desde a classe areia a argilo arenosa. Para o pH observou-se solos desde muito ácido a solos com alcalinidade elevada. De acordo com os macronutrientes (P, K, Ca e Mg), os solos foram classificados quanto a fertilidade desde baixas a muito elevada, mesmo comportamento ocorreu para os micronutrientes. Os teores de alumínio trocável variaram de muito baixo a médio. Os teores de carbono orgânico variaram de muito baixo a bom, o que era esperado devido às características climáticas e vegetativas da região de estudo (Ribeiro, 1999).

A acidez potencial (H+Al) apresentou valores médios de 2,11 cmol_c kg⁻¹, característicos de solos semiáridos, com baixas perdas por lixiviação de cátion básicos.

Tabela 3: Estatística descritiva das propriedades físicas e químicas do solo da área estudada no município de Boquira-Bahia

	pH	P	K	Na ⁺	Ca ²⁺	Mg ²⁺	M ^{x+}	H+Al ³⁺	CECe	CECp	V
	H ₂ O	mg kg ⁻¹	----- cmol _c Kg ⁻¹ -----				----- % -----				
Máximo	8,13	392,46	3,16	0,01	12,50	1,95	0,86	5,74	14,97	16,54	100,00
Media	6,47	17,74	0,60	0,00	3,17	0,62	0,04	2,18	4,44	6,58	63,47
Mediana	6,44	6,78	0,42	0,00	2,60	0,50	0,00	2,11	3,63	5,73	65,62
Mínimo	4,48	0,05	0,06	0,00	0,38	0,11	0,00	0,00	0,96	2,31	12,86
D. Padrão	0,72	46,88	0,54	0,00	2,15	0,39	0,12	1,08	2,77	2,74	19,29
	M	COT	Cu	Fe	Mn	Zn	PD	Areia grossa	Areia fina	Silte	Argila
	----- % -----	----- mg kg ⁻¹ -----				g cm ⁻³	----- % -----				
Máximo	54,68	3,36	42,90	821,10	732,00	11947,50	3,41	65,23	54,88	31,83	43,18
Media	1,98	1,12	2,77	82,82	120,41	329,75	2,63	35,15	35,34	9,39	20,12
Mediana	0,00	0,96	1,27	25,27	88,60	8,44	2,59	34,00	34,33	7,74	19,13
Mínimo	0,00	0,15	0,28	6,91	11,62	1,45	2,22	17,28	17,67	1,79	4,28
D. Padrão	6,93	0,54	5,63	172,77	126,14	1567,99	0,20	9,25	7,23	5,38	8,06

PD: Particle density; M^{x+}: Exchangeable acidity

3.3- Teores de PTEs nos solos estudados

Os teores de elementos potencialmente tóxicos e elementos principais nos solos da área estudada demonstraram uma heterogeneidade elevada (Tabela 4). Os teores de Pb e Zn apresentaram valores desde muito elevados (22467,78; 9583,70 mg kg⁻¹) a extremamente baixos (0,02; 0,00 mg kg⁻¹). Os teores de Ba e Cr foram baixos na maioria das amostras coletadas, com valores médios de 39,60 e 7,03 mg kg⁻¹, respectivamente. Para os elementos principais, ferro e alumínio os teores médios foram de 31116,77 e 14523,28 mg kg⁻¹.

Tabela 4: Estatística descritiva dos níveis de elementos potencialmente tóxicos e elementos principais em amostras de solo da área estudada no município de Boquira-Bahia.

	Pb	Zn	Ba	Cr	Fe	Al
	----- mg kg ⁻¹ -----					
Máximo	22467,78	9583,70	223,27	35,10	182460,06	33551,11
Media	751,90	300,91	39,60	7,03	31116,77	14523,28
Mediana	42,56	14,27	24,28	4,64	19831,40	13747,27
Mínimo	0,02	0,00	0,00	0,00	0,30	1411,69
D. Padrão	883,39	794,38	43,40	9,05	26137,63	6418,86

A correlação de Pearson foi usada para descrever as relações entre os elementos estudados e algumas propriedades químicas e físicas dos solos. Correlações significativas e positivas foram observadas para Pb/Zn, Pb/Cd e Pb/Fe. O Pb e Zn também apresentaram correlação positiva, porém fraca com FS, PD e Silte. O Al apresentou forte correlação positiva com os teores de argila, enquanto que o Fe a maior correlação foi com o silte e Zn. O Cd correlacionou-se positivamente com os teores de Zn e FS (Figura 4).

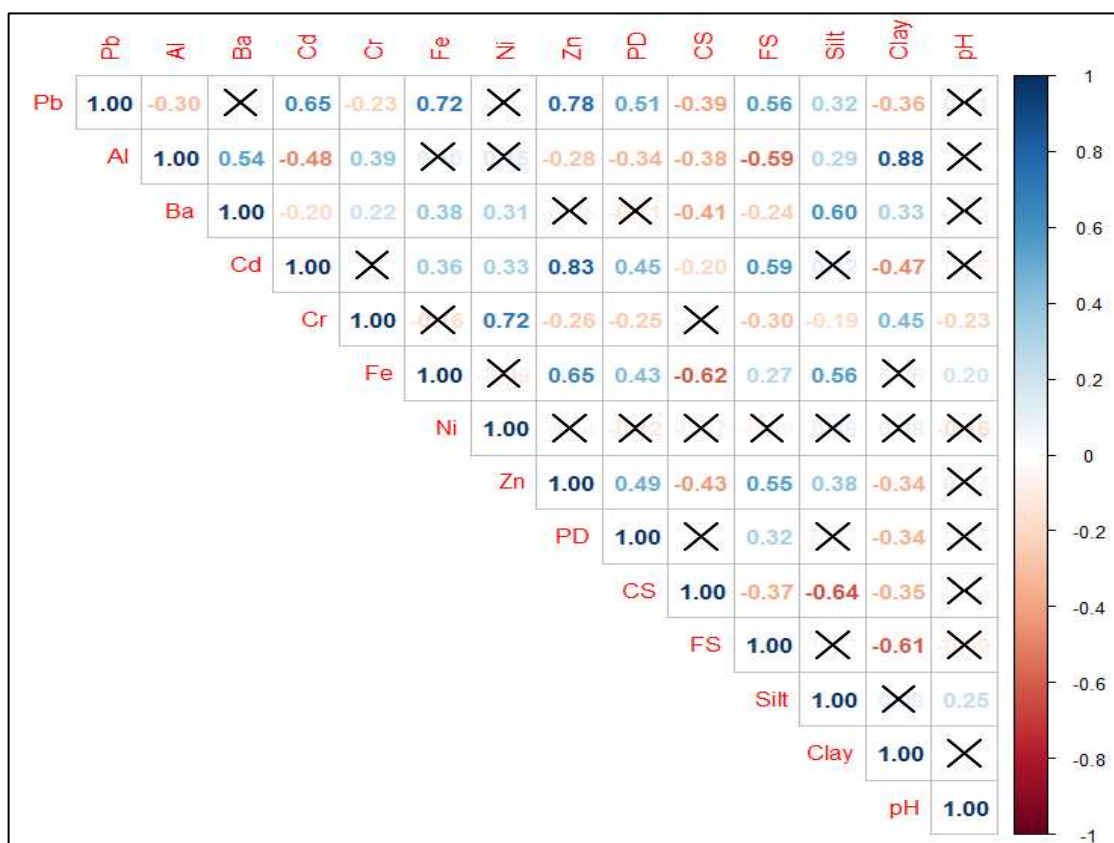


Figura 4. Matriz de correlação de Pearson para concentrações de elementos potencialmente tóxicos e atributos químicos do solo.

PD – Densidade de partículas; CS - areia grossa; FS - areia fina; Silt – silte; Clay – argila; X – correlação não significativa.

3.4- Avaliação do desempenho dos modelos preditivos

Os modelos testados apresentaram boa performance na predição da maioria dos teores de PTEs e elementos principais nos solos. O modelo SVM, CUB e RF obtiveram maiores desempenhos para todas as variáveis testadas (Tabela 5). A maioria dos modelos testados apresentaram valores de MAE e RMSE inferiores ao modelo nulo (Null_Model), dessa forma, sendo mais indicado do que o uso dos valores médios.

O algoritmo SVM apresentou o melhor desempenho para prever os teores de Pb e Zn, com menores valores de MAE e RMSE e maiores valores de R^2 . O segundo melhor desempenho foi obtido pelo CUB, atrás apenas do SVM. O algoritmo RF apresentou desempenho moderado, e o LM apresentou o pior desempenho, inclusive inferior ao null_model para o Pb.

Para os teores de Ba, o maior desempenho foi obtido pelo algoritmo RF, com menores valores de MAE e RMSE e maior valor de R^2 . O segundo algoritmo em desempenho foi CUB seguido do SVM. O pior desempenho ocorreu no algoritmo LM (Tabela 5).

Tabela 5: Desempenho dos modelos Cubist (CUB), Modelo Linear (LM), Florestas Aleatórias (RF), e Máquina de Vetor de Suporte (SVM) na predição dos teores de elementos potencialmente tóxicos e elementos principais (Pb, Zn, Ba, Cr, Fe e Al), avaliado pelo erro médio absoluto (MAE), raiz do erro quadrático médio (RMSE) e R-quadrado (R^2) do conjunto de dados de teste

	Chumbo	Zinco	Bário	Cromo	Ferro	Alumínio
	R^2					
RF	0,48 ± 0,3	0,48 ± 0,3	0,32 ± 0,1	0,18 ± 0,1	0,67 ± 0,1	0,27 ± 0,1
LM	0,43 ± 0,2	0,51 ± 0,2	0,24 ± 0,2	0,10 ± 0,1	0,47 ± 0,2	0,18 ± 0,1
SVM	0,56 ± 0,2	0,52 ± 0,2	0,28 ± 0,1	0,19 ± 0,1	0,67 ± 0,1	0,17 ± 0,1
Cubist	0,52 ± 0,3	0,50 ± 0,3	0,34 ± 0,2	0,19 ± 0,1	0,69 ± 0,2	0,29 ± 0,2
	MAE (mg kg ⁻¹)					
RF	846,24 ± 358,9	398,72 ± 181,1	24,47 ± 3,6	6,09 ± 0,8	12436,02 ± 2569,2	4432,19 ± 510,1
LM	1294,60 ± 304,5	465,25 ± 149,9	29,00 ± 5,2	6,83 ± 0,9	18237,06 ± 3935,3	5018,35 ± 678,4
SVM	744,71 ± 406,3	306,36 ± 168,2	24,55 ± 3,7	6,09 ± 0,9	12706,79 ± 2884,8	4861,23 ± 562,7
Cubist	831,11 ± 398,2	379,73 ± 206,9	24,51 ± 4,3	5,97 ± 0,8	11693,23 ± 2856,3	4554,40 ± 705,8
Null_model	1232,70 ± 309,9	492,03 ± 133,3	31,12 ± 3,0	6,95 ± 0,5	23633,89 ± 3092,5	5219,46 ± 390,4
	RMSE (mg kg ⁻¹)					
RF	2483,25 ± 1030,6	1218,39 ± 547,2	35,58 ± 6,2	8,42 ± 1,3	21028,02 ± 5084,8	5745,57 ± 758,5
LM	2438,87 ± 911,4	1028,01 ± 457,2	38,80 ± 7,3	9,08 ± 1,3	27223,39 ± 6760,5	6366,35 ± 1017,6
SVM	2309,79 ± 1448,4	946,49 ± 673,4	36,74 ± 7,4	8,42 ± 1,2	21805,91 ± 6029,4	6310,39 ± 853,5
Cubist	2745,26 ± 1177,9	1307,07 ± 644,1	35,90 ± 8,3	8,46 ± 1,2	20339,40 ± 5687,5	5914,43 ± 969,1
Null_model	2640,07 ± 1379,9	1063,70 ± 657,3	41,62 ± 7,2	9,00 ± 1,1	34088,40 ± 7195,3	6569,71 ± 594,9

Para o Cr, o algoritmo que apresentou maior desempenho foi o CUB, com menores valores de MAE e RMSE e maior R^2 . O segundo algoritmo em desempenho foram o SVM e RF com mesmos valores de MAE e RMSE, porém o SVM apresentou maior valor de R^2 , quando comparado com o RF. Assim como nos demais elementos potencialmente tóxicos o algoritmo LM apresentou o pior desempenho (Tabela 5).

Para os teores de Fe, o maior desempenho foi obtido no algoritmo CUB, com menores valores de MAE e RMSE e maiores valores de R^2 . O segundo algoritmo em desempenho foi o RF, seguido do SVM. O pior desempenho foi obtido no LM. Já para os teores de Al e melhor

desempenho ocorreu no algoritmo RF. O segundo algoritmo em desempenho foi o CUB. O algoritmo SVM apresentou desempenho moderado e o LM o pior desempenho (Tabela 5).

3.5- Importância das covariáveis

Os algoritmos de aprendizagem de máquina selecionaram covariáveis distintas na previsão de PTEs nos solos. Para os teores de Pb, a covariável mais importante foi a PDC com importância de 100 %. Outras variáveis exercem impactos fracos (NH, X, FLC e DEM) ou desprezíveis (SolDiret1 e STANH) na predição de Pb nos solos. Para a predição dos teores de Zn a covariável mais importante também foi a PDC, com importância de 100 %, seguida das covariáveis DAH e TSC com importância próxima de 50 e 25 % respectivamente (Figura 5).

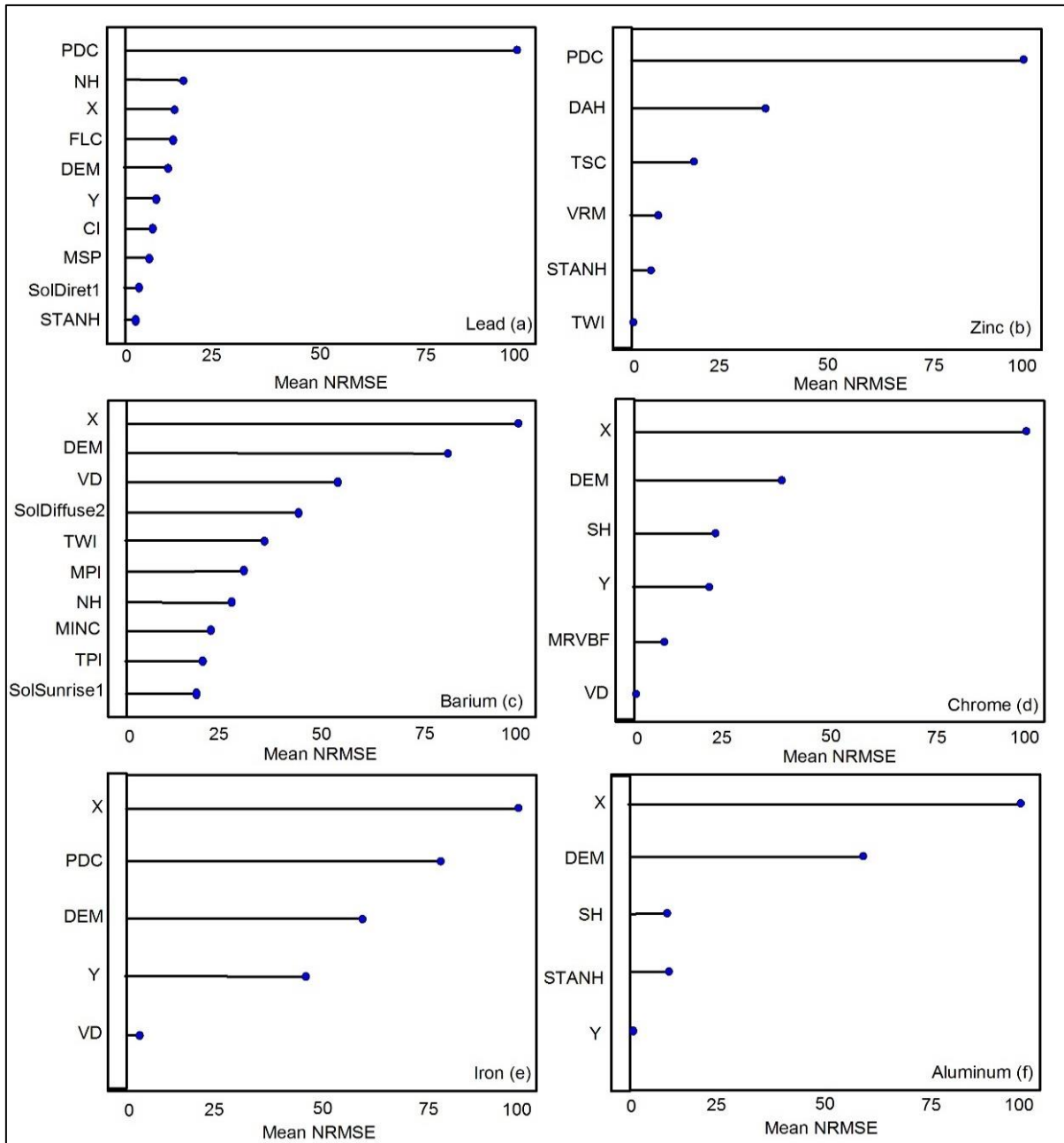


Figura 5. Importância das covariáveis na previsão dos teores de Pb, Zn, Ba, Cr, Fe e Al utilizando o algoritmo de melhor desempenho para cada elemento.

Para os demais elementos potencialmente tóxicos (Ba e Cr) e elementos principais (Fe e Al) a longitude (X), foi a covariável que mais contribuiu na previsão, com importância de 100 %. Para os elementos Ba e Al covariável DEM, apresentou importância próximo a 75 % e para os elementos Cr e Fe importância próximo a 50 %. Para os teores de Ba outras covariáveis também apresentaram importância superior a 25 %, como a VD, SolDiffuse2, TWI, MPI e NH. Para os teores de Fe as covariável PDC e Y (latitude) também apresentaram importância elevada, com valores próximo a 75 e 50 %, respectivamente (Figura 5).

3.6- Previsão dos mapas finais e incerteza dos PTEs e elementos principais nos solos

Os algoritmos que apresentaram melhor desempenho para cada elemento foram utilizados para elaborar os mapas de previsão dos teores de PTEs da área de estudo. Os maiores teores de Pb ocorrem na pilha de rejeito e nos canais de drenagem a jusante da pilha, local onde o processo de erosão pluvial favorece o carreamento de grande quantidade de material, principalmente partículas mais finas com elevados teores de Pb (Figura 6A). Nestes locais os teores de Pb variaram de 650 a 4000 mg kg⁻¹. O sudoeste e noroeste da pilha de rejeito também apresentaram teores elevados de Pb, com valores entre 300 a 1600 mg kg⁻¹. As regiões com teores elevados de Pb, englobam alguns bairros e um distrito da cidade de Boquira-BA. Nas demais áreas de estudos os teores se encontram abaixo de 72 mg kg⁻¹. Por meio dos mapas de previsão dos teores de Zn, também é possível observar que os maiores teores se encontram na pilha de rejeito e nos canais de drenagem. O rejeito minerado e depositado na pilha é constituído principalmente de Pb e Zn. Nas demais regiões da área de estudo os teores de Zn se encontram em níveis baixos (Figura 6C).

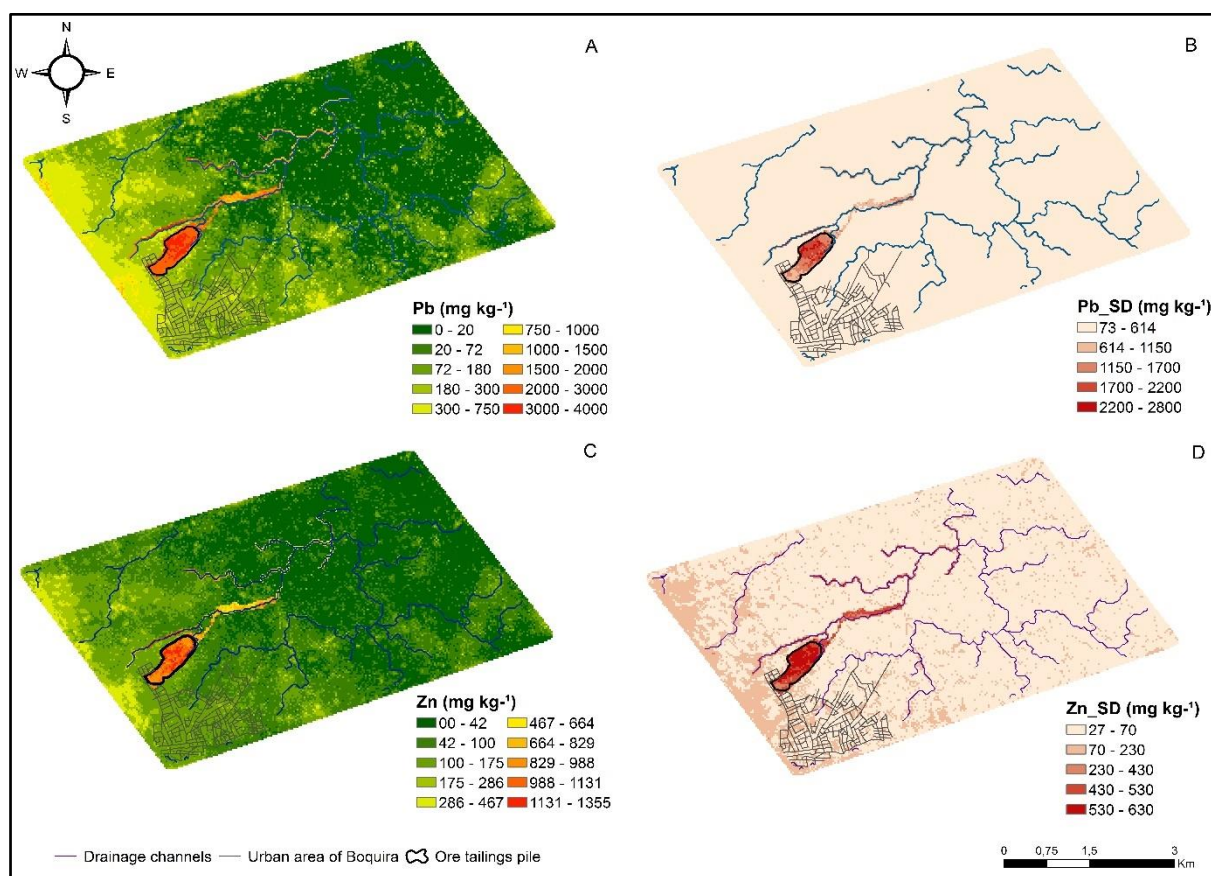


Figura 6. Teores de elementos potencialmente tóxicos nos solos, obtido a partir de 100 execuções usando os algoritmos de melhor desempenho para cada elemento. Teores de Pb e Zn previsto usando o algoritmo SVM (A, C) e SD usando o algoritmo SVM (B, D).

Os teores de Ba apresentaram valores abaixo do valor de prevenção (150 mg kg^{-1}) estabelecido pela resolução do Conama (2009) em toda área de estudo. Os maiores teores ocorreram na faixa inicial da área (oeste - sudoeste), onde se encontra parte da Serra do Espinhaço, local onde foram extraídos os minérios de Pb e Zn. Os teores de Ba foram diminuindo gradativamente ao se distanciar da Serra do Espinhaço (Figura 7C). Os teores previstos de Cr na área de estudo variaram de 0,3 a $24,5 \text{ mg kg}^{-1}$, também abaixo dos valores de prevenção estabelecidos pela resolução do Conama (2009) e dos valores de referência de qualidade do estado de Minas Gerais (COPAM, 2011). Os maiores teores ocorreram a noroeste e leste da pilha de rejeito (Figura 7A).

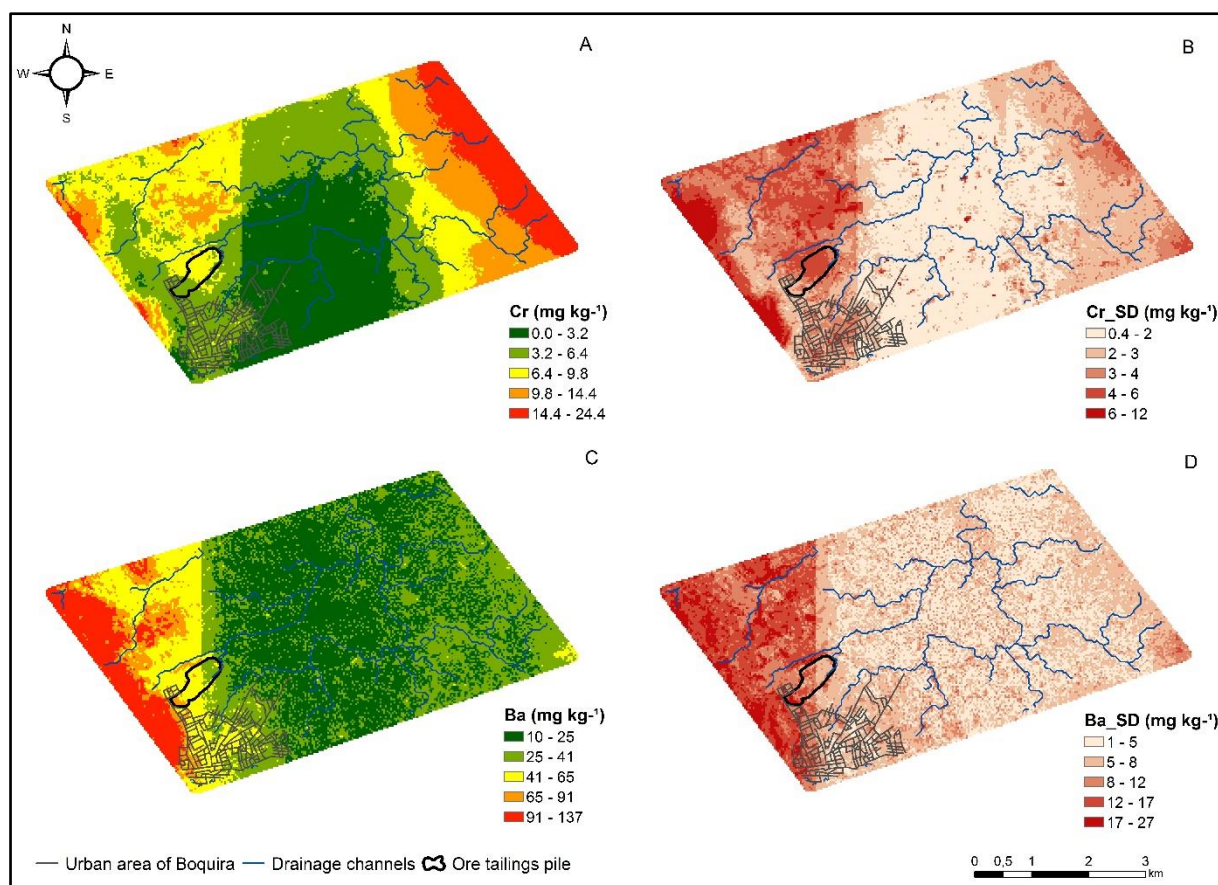


Figura 7. Teores de elementos potencialmente tóxicos nos solos, obtido a partir de 100 execuções usando os algoritmos de melhor desempenho para cada elemento. Teores e desvio padrão de Cr (A, B) e Ba (C, D) previsto usando o algoritmo Cubist e RF, respectivamente.

Com relação aos teores de Fe os maiores teores ocorrem na pilha de rejeito e próximo da Serra do Espinhaço. Nos canais de drenagem os teores também se encontram elevados em relação às demais regiões da área estudada (Figura 8A). Por meio dos mapas de previsão dos teores de Al é possível observar que as maiores concentrações se encontram a oeste da pilha de rejeito, com valores superiores a 20000 mg kg^{-1} (Figura 8C).

Os valores de desvio padrão para a maioria dos elementos estudados foram elevados, principalmente nos locais onde os teores dos elementos foram altos (Figura 6B, 6D, 7B, 7D, 8B e 8D).

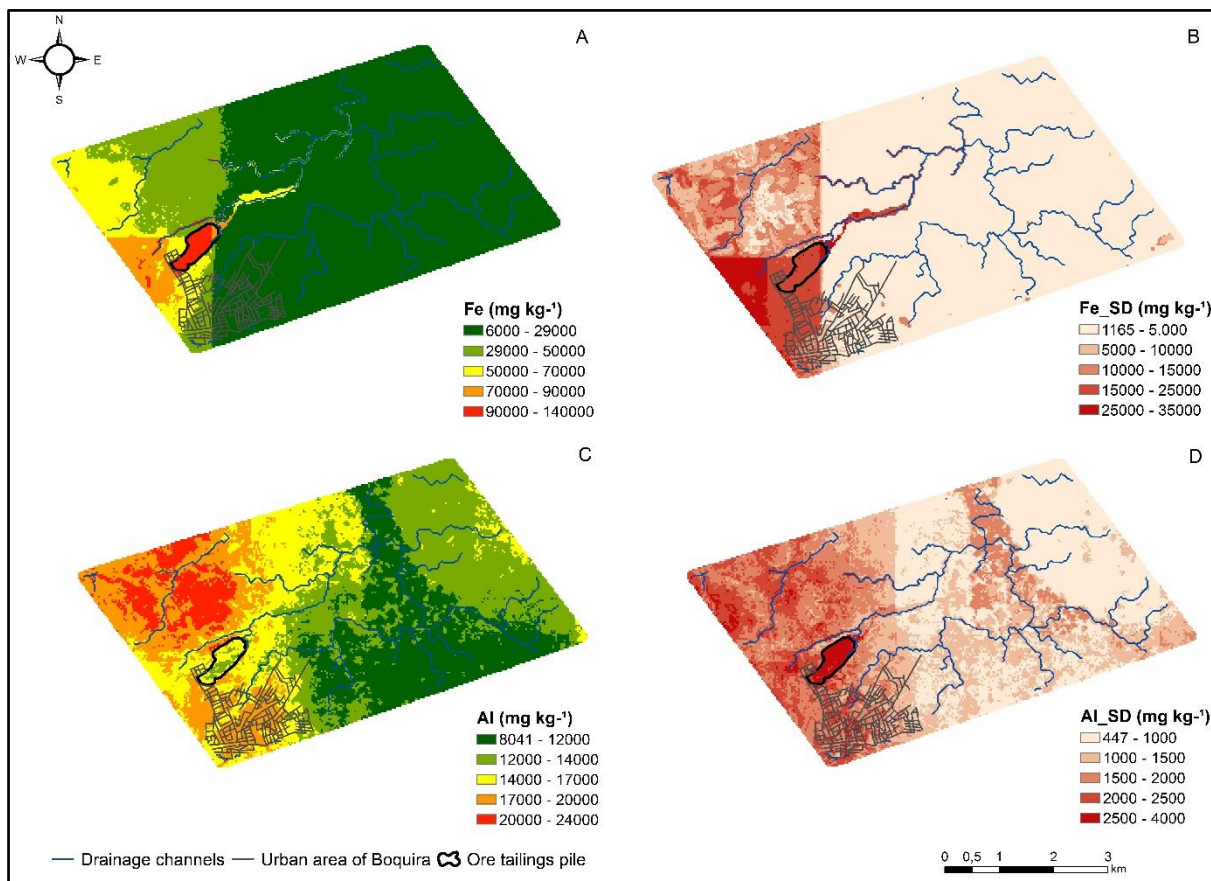


Figura 8. Teores de elementos principais nos solos, obtido a partir de 100 execuções usando os algoritmos de melhor desempenho para cada elemento. Teores e desvio padrão de Fe (A, B) e Al (C, D) previsto usando o algoritmo Cubist e RF, respectivamente.

3.7- Avaliação ambiental de PTEs em solos

Os mapas de distribuição do fator de enriquecimento (FE) são apresentados na Figura 9. A maior parte da área estudada foi classificada como natural ou moderadamente contaminada ($FE < 5$) para os PTEs Cd e Ba (Figura 9B e 9D). Para o Zn é possível observar no mapa de distribuição, desde regiões naturais a altamente contaminada. Porém a maior parte da área de estudo se encontra moderadamente contaminada a consideravelmente contaminada, sendo que essas áreas englobam toda área da cidade estudada e o lado norte da pilha de rejeito (Figura 9C). A área da pilha de rejeito e no canal de drenagem se encontram extremamente contaminada com Pb ($FE > 40$). No entorno da pilha e ao longo da serra do espinhaço os solos também se

encontram consideravelmente ($FE = 5 - 20$) a altamente contaminados ($FE = 20 - 40$; Figura 9A).

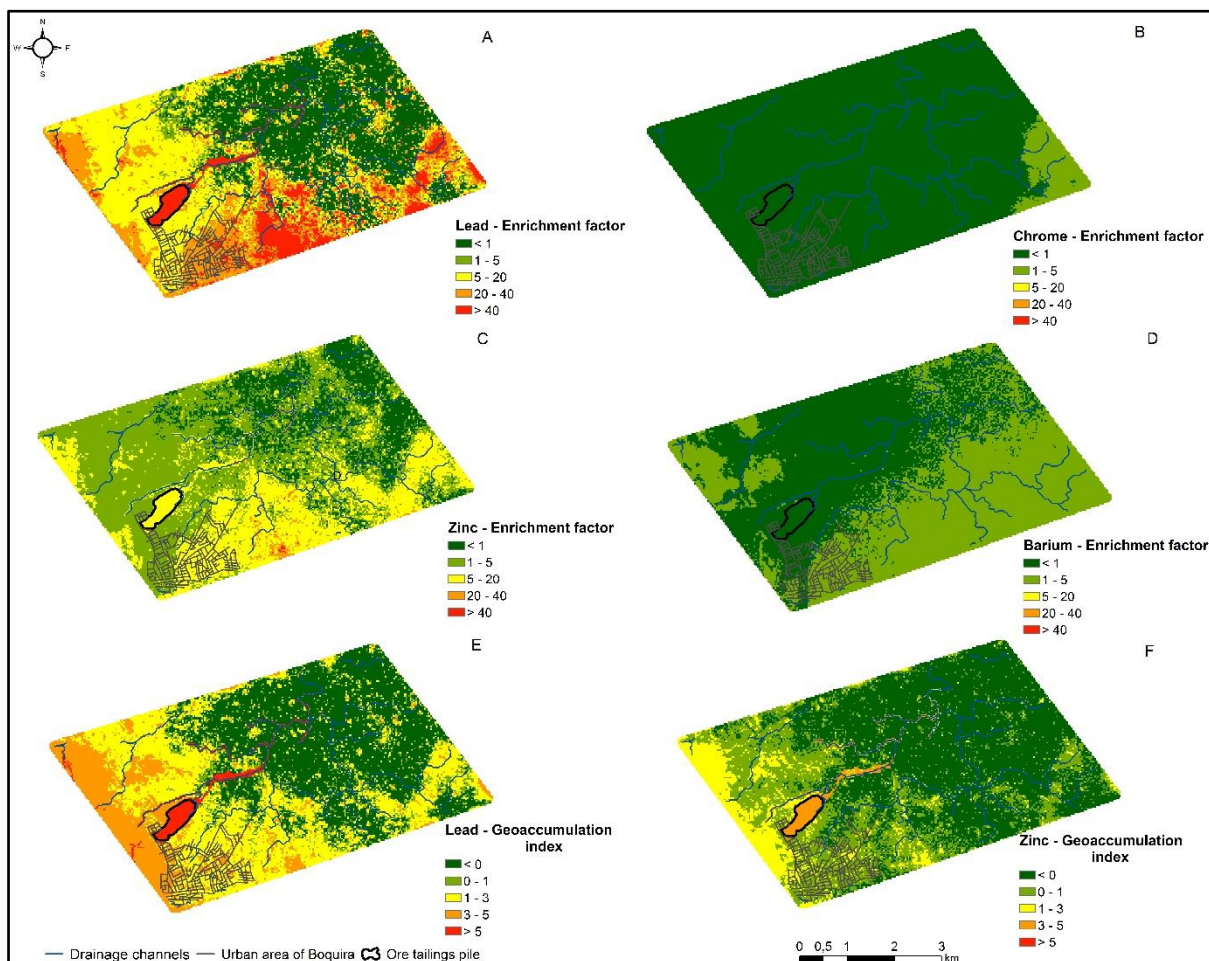


Figure 9. Fator de enriquecimento (A, B, C e D) e índice de geoacumulação (E, F) para os elementos potencialmente tóxicos em amostras de solo da cidade de Boquira-BA

Com exceção do Cr e Ba, os valores de Igeo para os demais PTEs apresentaram valores positivos, sugerindo que alguns pontos da área estudada apresentam algum nível de contaminação. Para o Zn a maior parte da área estudada foi classificada como natural, no entanto, houveram regiões classificadas como moderadamente contaminadas, consideradamente contaminadas e altamente contaminadas. As áreas altamente contaminadas estão na pilha de rejeito e no canal de drenagem principal, enquanto que as áreas consideravelmente contaminadas ocorrem ao longo da serra do espinhaço e a noroeste da pilha (Figura 9F). Diferente dos outros PTEs, menos de 50 % da área estudada foi classificada como natural para o Pb ($n = 45$). Os solos próximos a pilha de rejeito se encontram altamente

contaminados. Já o local onde a pilha de rejeito foi depositada e o canal de drenagem a jusante da pilha se encontram extremamente contaminados (Figura 9E).

3.8- Avaliação de risco a saúde humana

Os valores de HI foram inferiores a 1, tanto para crianças quanto para adultos para a maioria dos PTEs estudados. A exceção ocorreu apenas para o Pb, cujo os valores de HI foram superiores a 1 para crianças (Figura 10). Valores de HI menores que 1,0 são considerados seguros, porém quando superiores, são considerados risco não cancerígeno para a saúde humana (USEPA 2004).

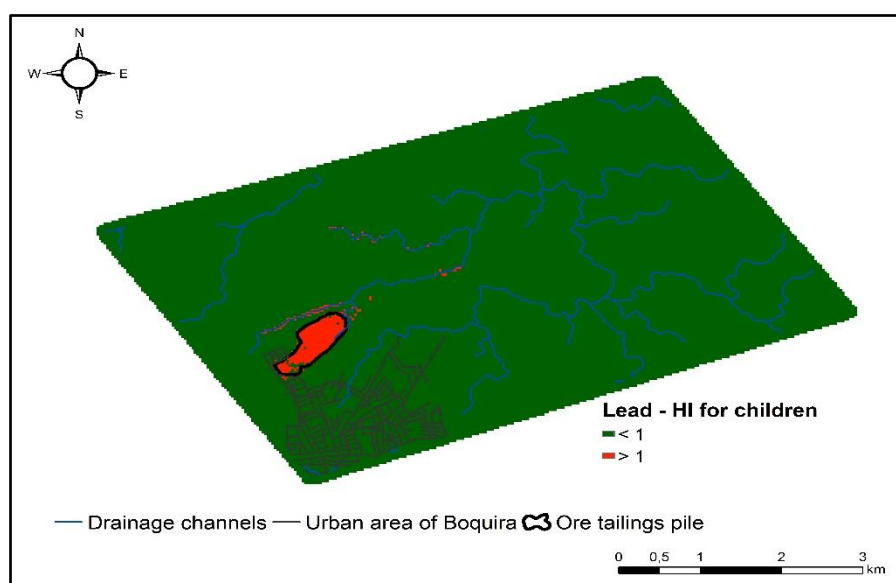


Figure 10. Risco não cancerígeno de Pb para crianças em amostras de solos da cidade de Boquira-BA

O risco carcinogênico associado ao Pb em adultos apresentou valores inferiores ao estabelecido como limite de segurança ($CR_{total} > 1 \times 10^{-6}$) para maior parte da área avaliada, com exceção apenas da pilha de rejeito e alguns pontos no canal de drenagem principal (Figura 11). Já o risco carcinogênico associado ao Pb em crianças apresentou valores bem superiores aos limites de segurança, principalmente na pilha de rejeito e no canal principal de drenagem (11B). Além disso, em toda área próxima a Serra do Espinhaço e no entorno da pilha, englobando alguns bairros da cidade e o distrito de tiros, ocorreram valores superiores a 1×10^{-6} . Para o risco carcinogênico associado ao Cr, em adultos os valores foram inferiores a 1×10^{-6} , já para crianças a maior parte da área apresentou valores superiores aos limites estabelecidos como de segurança (11C, 11D).

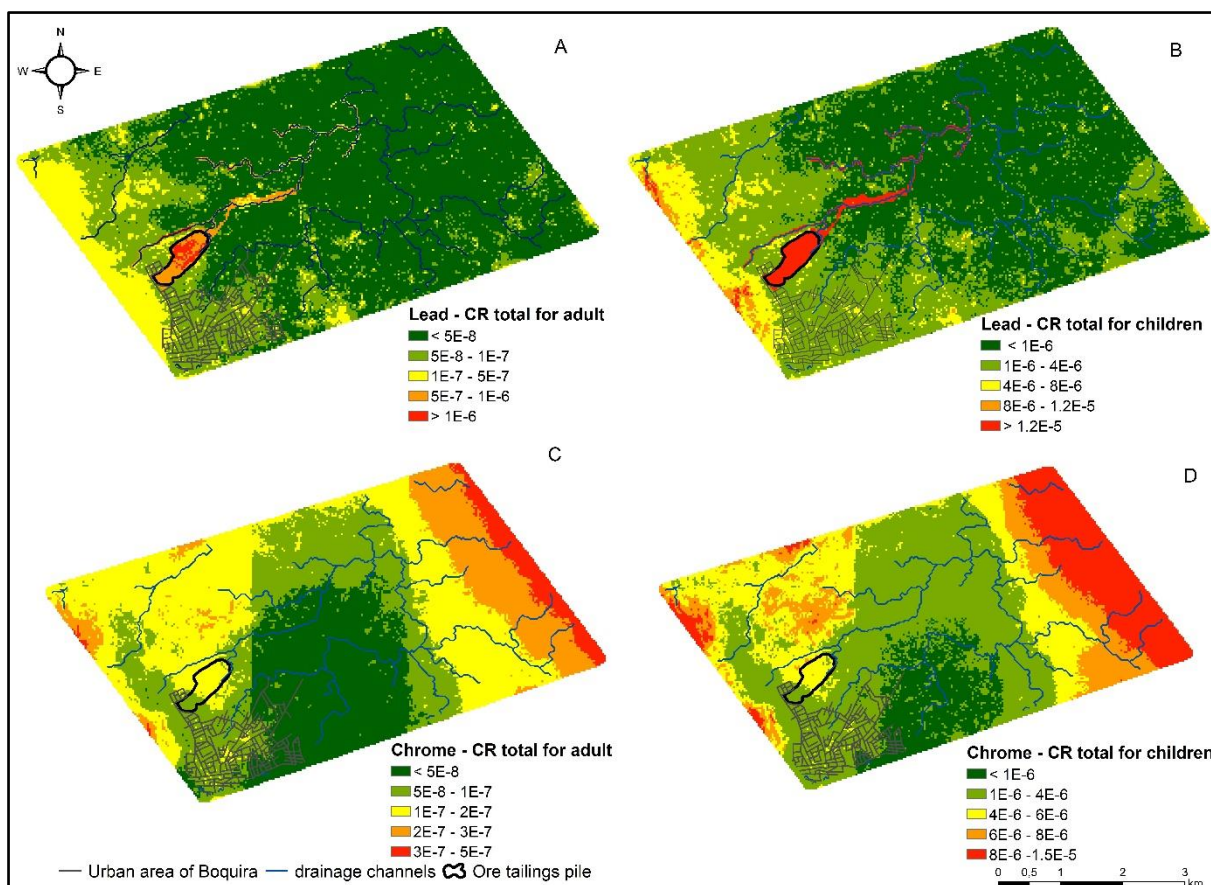


Figura 10. Risco de câncer em adultos e crianças por elementos potencialmente tóxicos em amostras de solo da cidade de Boquira-BA

4- DISCUSSÃO

4.1- Propriedades químicas e textura dos solos estudados

A predominância de pH próximo a neutralidade nas amostras de solos estudadas, concorda com a composição mineralógica das rochas metabásicas e com o clima semiárido da área. Nesse cenário, a ação dos agentes de intemperismo tem baixa atuação, o que favorece a permanências de cátions de dissolução básica, promovendo maior fertilidade nos solos (Carvalho et al., 1982; Carvalho et al. 1997; Oliveira et al., 2018; Câmara et al., 2021; Nascimento et al., 2021; Sousa et al., 2020). Por consequência dos altos teores de bases trocáveis e pH elevado, os teores de alumínio trocáveis são muito baixos nesses solos. Os teores de carbono orgânico total (COT) foram baixos, o que era esperado, visto que a área de estudo está inserida no Bioma Caatinga, onde apresenta vegetação com poucas folhas, aportando pouca matéria orgânica aos solos (Ribeiro et al., 1999; Gomes et al., 2019).

A grande variação nos teores disponíveis de micronutrientes (Zn, Cu, Fe e Mn), pode estar relacionada a proximidade das amostras em relação a pilha de rejeito, onde têm-se altos

teores desses elementos, já os baixos valores podem estar relacionados ao pH próximo a neutralidade, condição essa, que reduz a disponibilidade de micronutrientes catiônicos (Santos et al., 2020; Gondal et al., 2021). Além disso, a atuação de diferentes processos pedogenéticos e a diversidade geológica da área possibilitou a formação de diferentes classes de solos como os Argissolos Vermelho-Amarelo Eutrófico, Latossolos Vermelho Eutrófico, Latossolos Vermelho-Amarelo Distrófico, Neossolos Litólicos Eutróficos, os quais apresentam deferentes características químicas e físicas (RADAMBRASIL,1982).

Apesar de apresentar uma elevada variabilidade nos teores de areia, silte e argila, é possível verificar uma predominância da fração areia em relação as demais, isso ocorre devido a geologia da área, predominantemente xistosa, como a presença da formação ferrífera bandada (fácies óxido, silicato e carbonato), associada com mica xisto e de metaquartzarenito, de rochas granitoides (Biotita monzogranito médio a grosso), e depósitos lateríticos (Espourteille e Fleischer, 1980; Arcanjo et al., 2005; Guimarães et al., 2014). Além disso, o baixo intemperismo do semiárido favorece a ocorrência de frações mais grosseiras, como areia e silte.

4.2- Desempenho dos algoritmos de aprendizagem de máquinas na previsão espacial dos teores de elementos potencialmente tóxicos e elementos principais nos solos

Para solos urbanos impactados ou não pela atividade de mineração, previsões eficientes de PTEs são difíceis devido a elevada variabilidade espacial e fontes múltiplas de contaminação (Wei e Yang, 2010; Zhang et al., 2021). No entanto, os resultados do presente estudo demonstraram que os algoritmos de aprendizagem de máquinas são alternativas promissoras para espacializar os teores de elementos potencialmente tóxicos e elementos principais, em ambientes impactados pela atividade de mineração.

Os modelos testados apresentaram diferentes desempenhos na predição espacial dos teores de elementos principais e elementos potencialmente tóxicos nos solos. Para os PTEs, Pb e Zn o maior desempenho foi obtido utilizando o modelo SVM, enquanto que para o Ba e Al o maior desempenho foi obtido utilizando o RF. Resultados semelhantes foram encontrados por Zhang et al. (2020b) utilizando diferentes algoritmos de aprendizagem de máquinas para previsão espacial dos teores de PTEs nos solos em área de urbanização na China, neste caso os autores verificaram alto desempenho dos modelos SVM e RF. O alto desempenho do modelo SVM também foi observado na predição dos teores disponíveis de N, P e K em solos localizados no território da República do Tataristão (Sahabiev et al., 2021) e na predição de Fe e Ni em águas do rio Shur contaminadas por atividade de mineração a jusante da mina Sarcheshmeh

(Gholami et al., 2011). Já Zhang et al., (2021) verificaram resultados satisfatórios na predição dos teores de PTEs utilizando o algoritmo RF.

Para os elementos Cr e Fe, o maior desempenho foi obtido utilizando o algoritmo CUB. Na literatura existem poucos trabalhos utilizando esse modelo na previsão de PTEs nos solos (Peng et al., 2016), porém há diversos estudos demonstrando alta performance do modelo CUB na predição de outras propriedades dos solos (Butler et al., 2018; John et al., 2021; Zhao et al., 2021).

Apesar dos algoritmos de aprendizagem de máquinas demonstrarem bom desempenho na predição de PTEs em áreas impactadas pela atividade de mineração, uma melhoria no desempenho poderia ser alcançada com a identificação de outras covariáveis que podem estar influenciando na contaminação dos solos, como por exemplo, a velocidade e a direção dos ventos.

4.3- Importância das covariáveis

Para os elementos associados a atividade de mineração, foi possível verificar que a covariável PDC foi a que mais influenciou na predição. Isso ocorre devido à mesma estar relacionada a pilha de rejeito e ao canal de drenagem principal que possuem altas concentrações de PTEs, e possivelmente são a principal fonte de contaminação de Pb e Zn na área estudada.

A pilha de rejeito de mineração por estar depositada de forma inadequada vem sofrendo com processos erosivos, promovidos pela ação hídrica, a qual foi intensificada pela retirada da vegetação e ausência de barreiras de contenção. Além disso, a inclinação e a instabilidade dos taludes da pilha, permitem que o rejeito se movimente por gravidade chegando a áreas localizadas fora dos seus limites (Santos, 2014, Santos et al., 2020). A falta de cobertura vegetal em torno da pilha, associada a granulometria do rejeito de mineração (fração silte a areia fina), favorece o carreamento de partículas de rejeito pela ação do vento, contaminando os solos. Estudos têm demonstrado que os casos de deposição atmosférica de poeiras contaminadas, oriundas de fontes pontuais de contaminação, como pilhas de rejeitos e áreas de fundição, a distância da fonte, direção e velocidade dos ventos, frequência e intensidade das precipitações, rugosidade e interceptação pelo ambiente podem influenciar nas taxas de deposição e na distribuição espacial dos teores de PTEs nos solos (Rawlins et al., 2006; Fritsch et al., 2010). Reza et al., (2014) verificaram que as áreas de depósitos de resíduos de mineração de carvão são as principais fontes de contaminação de áreas cultivadas próximo as minas.

Para os elementos Ba e Cr, as covariáveis longitude (X) e DEM foram a que mais contribuíram para predição desses elementos. Tal resultado pode estar relacionado as

características geológicas da área de estudo, pois, a oeste da área está localizada a Serra do Espinhaço, que possui rochas associadas as formações férlicas bandadas (fácies óxido, silicato e carbonato), além de estar localizada em cotas mais elevadas, de forma que a influência das rochas associadas as formações férlicas diminui quando se desloca para leste, e aumenta a influência de rochas granitoides e depósitos lateríticos na formação dos solos (Guimarães et al., 2014).

As variáveis DEM e X também foram as mais importantes na predição dos teores de Fe e Al, o que pode estar relacionado à disposição geológica das rochas na área, como ocorreu para Ba e Cr. Para o Fe, as variáveis PDC e Y (latitude) também foram importantes na predição, pois, as rochas mineradas para extração de Pb e Zn estavam associadas as formações férlicas, dessa forma o rejeito depositado no canal de drenagem também possui altos teores de Fe. Além disso, como citado anteriormente, a pilha sofre com processos erosivos devido a ação eólica, contaminando áreas ao entorno, principalmente a nordeste da mesma. Um fator que se agrava devido a estação seca ser caracterizada por ventos de elevada intensidade, com alta capacidade de dispersão de partículas (Santo et al., 2020). Ademais, os elementos presentes nas pilhas de rejeitos estão associados a partículas que correspondem a frações granulométricas finas (areia fina e silte), facilitando o transporte atmosférico a longas distâncias (Alves et al., 2018).

4.4- Distribuição espacial dos teores de PTEs e elementos principais nos solos

Diversos estudos demonstram que distribuição espacial dos teores de PTEs e elementos principais são influenciados pelas características geológicas e pedogenéticas dos solos da área de estudo e de passivos ambientais oriundos de atividades de mineração deixados no ambiente (Reimann e Garrett 2005; Biondi et al., 2011; Gloaguen e Passe, 2017; Santos et al., 2017). No presente estudo a PDC foi a covariável mais importante para evidenciar a distribuição espacial dos teores de Pb, Zn e Fe. O carreamento de material proveniente da erosão hídrica da pilha de rejeito está elevando os teores de PTEs nos canais de drenagem a jusante da pilha. Nesses locais os teores de Pb e Zn estão muito acima dos valores de investigação para áreas industriais, determinados pela resolução do Conama (2009). Em Santo Amaro da Purificação-BA, por exemplo, as escórias e a bacia de rejeito de mineração de Pb e Zn, vem sofrendo com processos erosivos hídricos, dispersando contaminantes no rio Subaé, demonstrando que esse tipo de erosão contribui significativamente para dispersar contaminantes no ambiente (Hatje e Barros 2012; Silva et al., 2017).

Do lado sul da pilha de rejeito, onde se concentra grande parte da área urbana do município de Boquira, e toda área ao lado norte e nordeste, onde se concentram alguns bairros

e um distrito da cidade (Tiros), além de áreas utilizadas para agricultura e pecuária, os teores de Pb se encontram elevados, com teores variando de 300 a 1000 mg kg⁻¹, acima dos valores de investigação para áreas urbanas e agrícolas estabelecidas pela resolução do Conama (2009). As características do material da pilha de rejeito, associado a elevada intensidade dos ventos, estão contribuindo para dispersar partículas de poeira que contém elevados teores de PTEs. De acordo com Santos et al. (2020) a intensidade dos ventos durante o período seco é superior a 18 m/s, com capacidade de dispersar partículas de poeira contaminadas com PTEs. Cunha et al. (2016) destacam que os ventos são constantes na área da pilha de rejeito, os quais têm provocado ressuspensão do material fino ali depositado, disseminando sedimentos de poeiras contaminadas nas ruas e residências da cidade de Boquira. Em algumas residências, os teores de Pb e Zn na poeira foram acima de 9,0 e 5,0 g kg⁻¹, respectivamente (Cunha et al., 2016). De acordo com Csavina et al. (2012) os PTEs podem ser transportados rapidamente e por longas distâncias pela poeira atmosférica e aerossóis. Qu et al. (2018) destacam que as concentrações de Pb nos solos foram inversamente proporcionais às distâncias de locais industriais na China, onde historicamente era produzido óxido de chumbo. Conforme demonstrado, devido à proximidade com a área urbana e agrícola e os riscos que os PTEs trazem aos seres humanos e ao meio ambiente, medidas devem ser tomadas para impedir que esse material contaminado continue sendo disseminado.

Os teores de Ba e Cr estão abaixo do limite de prevenção estabelecido pela resolução do Conama (2009) em toda área estudada. Quando comparados com os VRQs do estado de Minas Gerais, os teores de Cr estão abaixo, enquanto que, os teores de Ba encontram-se um pouco acima em alguns pontos da área estudada (COPAM, 2011). Para o Ba, as áreas com maiores teores, se encontram próximo a Serra do Espinhaço. A geologia da Serra do Espinhaço, constituída por rochas máficas como as formações férricas que são ricas em diversos elementos, têm contribuído para elevar os teores de metais nos solos (Espourteille e Fleischer, 1980; Alves et al., 2017). Além disso, como a área de estudo está localizada na região semiárida do Brasil a influência litológica sobressai à influência pedológica na concentração de metais no solo, fazendo com que os teores de metais dos solos se correlacionem altamente com o material de origem (Santos et al., 2017).

Para o Cr, os maiores teores (24 mg kg⁻¹) ocorrem a leste da área estudada, associado a solos com maiores proporções de areia. Diversos estudos têm demonstrado diferenças na distribuição de PTEs nas frações granulométricas dos solos, onde Pb, Zn, Cu e Cd se acumulam na fração argila, enquanto que Ni, Co e Cr, podem ser enriquecidos nas frações argila ou areia, dependendo dos tipos de material de origem (Ajmone-Marsan et al., 2008; Acosta et al., 2011).

Os teores de Fe variaram de 6,0 a 140,0 g kg⁻¹ em toda área de estudo, com maiores teores na pilha de rejeito e canal de drenagem principal e ao longo da Serra do Espinhaço, assim como os elementos Cr e Ba. Além disso, o minério de Pb e Zn estava associado a formações férreas e com processo de extração e beneficiamento desses elementos, assim pode ter ocorrido uma concentração de Fe nos rejeitos de mineração. Apesar do Fe não representar sérios riscos à saúde humana, tem função essencial como nutriente para as plantas e sua presença está sempre associada a ocorrência de PTEs, então, seu estudo se torna essencial em trabalhos de monitoramento ambiental (Biondi et al., 2011; Santos et al., 2017).

A distribuição espacial dos teores de Al também demonstrou uma heterogeneidade muito grande na área de estudo, com valores variando de 8,0 a 24,0 g kg⁻¹. Essa variação pode estar correlacionada a ocorrência de rochas ricas desse elemento em alguns pontos da área estudada, visto que nessa região também ocorre exploração de minerais quartzitos como os aluminofosfatos e aluminoborosilicatos (Evangelista e Danderfer Filho, 2012). Porém, os teores de Al determinados em toda área foram inferiores à média da crosta terrestre (Rudnick e Fountain, 1995). Isso pode estar relacionado ao método de abertura de amostra utilizado, pois a metodologia EPA 3051A não extrai elementos associados aos silicatos (USEPA, 2007).

4.5- Avaliação da poluição do solo e risco a saúde humana

O FE indicou que os solos na área de estudo se enquadram na classe natural para os elementos Ba e Cr. Já o Igeo para os mesmos PTEs mostrou maior sensibilidade enquadrando os solos desde natural a moderadamente contaminados (Wu et al., 2014). Ferreira et al. (2021) também encontraram discrepância nos resultados de FE e Igeo, porém, em seu estudo o FE se mostrou mais sensível em detectar alterações nos teores de PTEs nos solos. De acordo com Zhang e Liu (2002), valores de FE entre 0,05 e 1,5 indicam que os teores de PTEs são inteiramente derivados de materiais crustais ou processos naturais, enquanto valores de FE maiores que 1,5 sugerem que as fontes são mais prováveis de serem antropogênicas.

Os resultados de FE e Igeo demonstram que os solos da área de estudo se encontram nas classes desde natural a extremamente contaminados com Pb e Zn (Wu et al., 2014). Na área da pilha de rejeito e em seu entorno, que englobam grande parte da cidade e áreas agrícolas ao norte da pilha, os solos estão extremamente contaminados. As concentrações elementares na classe extremamente contaminado podem ser cem vezes maiores do que o valor de fundo geoquímico (Bhuiyan et al., 2010). O alto FE e Igeo evidencia que grande parte da área estudada está altamente contaminada, o que pode ser correlacionado com a atividade de mineração passada e pós-abandono de rejeitos de mineração. Krishna et al. (2012) também

atribuíram o alto FE para PTEs obtido em seu estudo à atividade de mineração e aos passivos ambientais deixados. De acordo com Shahmoradi et al. (2020), os índices de poluição (Igeo e FE) indicaram que os sedimentos do rio Aqyazi estavam moderadamente contaminados por Cu e extremamente contaminados por Cd, também sugerindo influência da atividade de mineração na deposição de Cu e Cd no Rio. É importante destacar que a geologia de parte da área estudada é composta de rochas básicas e ultrabásicas associadas a depósitos de Pb e Zn, os quais podem elevar o background geoquímico desses elementos nos solos estudados, podendo fazer com que ocorra uma superestimação do FE e Igeo. Como o estado da Bahia não dispõe de valores de referência de qualidade foi necessário utilizar o background geoquímico do estado de Minas Gerais, porém, apesar de apresentar características semelhantes com a área de estudo, podem não refletir o mesmo nível de fundo.

Os valores de índice de perigo (HI) para todos os PTEs avaliados foram inferiores a 1, com exceção do HI de Pb para crianças, pois na área da pilha de rejeito os valores foram superiores a 1, indicando a possibilidade de risco não carcinogênico, ou seja, diversos problemas de saúde, exceto câncer. De acordo com a USEPA (2004), valores de índice de perigo (HI) superiores a 1 podem indicar a existência de efeitos adversos à saúde humana. Além da pilha de rejeito, neste local se encontra o lixão da cidade, onde, diversas crianças e adultos trabalham com reciclagem, sendo expostos diariamente aos contaminantes.

Os valores de CR em adultos foram considerados seguros para o elemento Cr. Por outro lado, para o Pb os valores de CR ultrapassaram o nível de segurança, principalmente na área da pilha de rejeito (1×10^{-6}). Quando avaliado o CR em criança é possível verificar que tanto para o Pb quanto o Cr os valores ultrapassaram os níveis de segurança em parte da área estudada. As áreas com maiores riscos carcinogênicos para o Pb ocorreram na pilha de rejeito e seu entorno, que engloba grande parte da cidade e áreas agricultáveis circunvizinhas, indicando a contribuição da pilha de rejeito como grande responsável pela contaminação. Valores de CR acima de 1×10^{-4} são considerados inaceitáveis, CR entre 1×10^{-6} e 1×10^{-4} são vistos como aceitáveis, enquanto os riscos abaixo de 1×10^{-6} não são vistos como efeitos significativos à saúde (Fryer et al., 2006; USEPA, 2007; EPA, 2016). Dessa forma, os resultados encontrados no presente estudo servem de alerta para as autoridades buscarem alternativas para evitar que os passivos ambientais deixados pós mineração continuem contaminando os solos do município, elevando os riscos à saúde da população e de todo o meio ambiente.

O monitoramento dos teores de PTEs nos solos é indispensável para prevenção do risco a saúde humana, pois, o contato com esses elementos pode ocasionar diversos problemas de saúde, como retardo mental, distúrbios neurucognitivos e comportamental em crianças,

problemas respiratórios, cardiovasculares, renais, gastrointestinais, reprodutivos e carcinogênicos, tanto em crianças quanto em adultos (Matta e Gjyli 2016; Pan et al. 2018; Al osman et al. 2019; Mohammadyan et al., 2019).

De maneira geral, observamos no presente estudo que a área avaliada apresenta elevados teores de elementos potencialmente tóxicos, possivelmente devido a atividade de mineração e dos passivos ambientais deixados. Porém, como a área é constituída de rochas ricas em metais, novos estudos são necessários para separar de forma efetiva a contribuição da geologia e da atividade de mineração nos teores de metais dos solos.

5.0 – CONCLUSÕES

Este estudo demonstra que os algoritmos de aprendizagem de máquinas podem ser aplicados com sucesso para mapear os teores de elementos potencialmente tóxicos e elementos principais nos solos de áreas urbanas e agrícolas impactadas pela atividade de mineração.

O modelo SVM apresentou o melhor desempenho na previsão espacial dos teores de Pb e Zn nos solos, quando comparados com os demais algoritmos de aprendizagem de máquina testados. Para os elementos Ba e Al o melhor desempenho foi obtido utilizando o algoritmo RF, enquanto que, para o Fe e Cr, o modelo CUB apresentou maior desempenho.

A covariável mais importante que influenciaram a distribuição de Pb e Zn na área estudada foi a pilha de rejeito, enquanto que para os demais elementos Fe, Al, Ba e Cr a longitude e o modelo de elevação digital foram as covariáveis mais importantes.

Os solos da área avaliada apresentaram teores de Pb e Zn muito acima dos valores de prevenção estabelecidos pela resolução do Conama (2009), principalmente nas áreas próximas da pilha de rejeito, indicando que a mesma é a principal fonte de contaminação. Já próximo a Serra do Espinhaço a geologia da área também pode contribuir para elevar os teores de PTEs.

O fator de enriquecimento (EF) e o índice de geoacumulação (Igeo) dos elementos potencialmente tóxicos Pb e Zn mostram que os solos na área de estudo são naturais a extremamente contaminados, enquanto Ba e Cr estão geralmente dentro dos níveis de fundo. Assim como na distribuição espacial, os solos com maiores FE e Igeo se encontram ao entorno da pilha de rejeito, indicando ser a pilha a principal fonte de contaminação da área. Contudo é importante enfatizar que novos estudos buscando determinar o background geoquímico desses solos são necessários para evitar que ocorra superestimação ou subestimação dos índices de contaminação.

O risco não cancerígeno (HI) da maioria dos elementos potencialmente tóxicos foi consideravelmente baixo para crianças e adultos, com exceção do Pb que obteve valores de HI maiores que 1 para crianças na área da pilha de rejeito. Já os riscos cancerígenos (CR) de Cr e Pb em crianças ultrapassaram o limite de segurança em áreas ao entorno da pilha de rejeito, indicando que a longa exposição a solos contaminados aumenta a probabilidade de ocorrência de câncer. As áreas ao entorno da pilha correspondem a grande parte da cidade, o que torna o problema ainda maior devido à maior densidade populacional.

Os resultados demonstram que os passivos ambientais deixados pós mineração na cidade de Boquira vêm contribuindo para elevar os teores de PTEs nos solos e que o monitoramento constante deve ser realizado para avaliar o efeito da contaminação na saúde da população. Além disso, novos estudos devem ser realizados, visando buscar alternativas para impedir que os passivos ambientais continuem sendo fonte de contaminação, bem como buscar alternativas para remediar áreas já contaminadas.

6.0- AGRADECIMENTOS

Esta Pesquisa foi desenvolvida com o auxílio do cluster da UFV (Universidade Federal de Viçosa). Este estudo foi parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Programa Nacional de Cooperação Acadêmica-PROCAD 2013 - Código Financeiro 001.

7.0- REFERÊNCIAS

Abbas, T., Akmal, M., Aziz, I., Iqbal, M., & Ahmed, H. (2021). Risk assessment and GIS-based mapping of heavy metals in the secondary rock deposits derived soils of Islamabad, Pakistan. *Environmental Earth Sciences*, 80(3), 1-9.

Abrahão, W. A. P., & Marques, J. J. (2013). Manual de coleta de solos para valores de referência de qualidade no estado de Minas Gerais. Belo Horizonte: Fundação Estadual do Meio Ambiente.

Acosta, J. A., Martínez-Martínez, S., Faz, A., & Arocena, J. (2011). Accumulations of major and trace elements in particle size fractions of soils on eight different parent materials. *Geoderma*, 161(1-2), 30-42.

Ajmone-Marsan, F., Biasioli, M., Kralj, T., Grčman, H., Davidson, C. M., Hursthouse, A. S., ... & Rodrigues, S. (2008). Metals in particle-size fractions of the soils of five European cities. *Environmental pollution*, 152(1), 73-81.

Al osman, M., Yang, F. & Massey, I.Y. Exposure routes and health effects of heavy metals on children. *Biometals* 32, 563–573 (2019). <https://doi.org/10.1007/s10534-019-00193-5>.

Ali, H. & Khan, E. (2018). Trophic transfer, bioaccumulation, and biomagnification of non-essential hazardous heavy metals and metalloids in food chains/ webs concepts and implications for wildlife and human health. *Human and Ecological Risk Assessment: An International Journal*.

Alves, F.E.A.; Bertolino, L.C. & Mendes, J. (2018). Mineralogical Characterization of Lead Mine Tailing in Boquira, Bahia State, Brazil. *Anuário do Instituto de Geociências – UFRJ*, 40(3): 14-23.

Andrade, A. A. X.; Soares, E. M. B.; Cunha, D. A.; de Oliveira, M. L. R. (2017). Riscos e incertezas: a realidade pós-extração do minério de chumbo em Boquira, BA. *Interações*, 18(1), 103-117.

Antoszczyszyn, T., & Michalska, A. (2016). The potential risk of environmental contamination by mercury contained in Polish coal mining waste. *Journal of sustainable mining*, 15(4), 191-196.

Anvisa, 2003. Agência Nacional de Vigilância Sanitária - Resolução Específica (RE) no 899 de 29 de maio de 2003. Determina a publicação do “Guia para validação de métodos analíticos e bioanalíticos”. *Diário Oficial da União*, 2.6.2003. Ministerio da Saúde - Brasil.

Arcanjo, J. B. A.; Martins, A. M.; Loureiro, H. C. Delgado, I. M.; Souza, J. D.; NeveS, J. P.; Oliveira, J. E.; Teixeira, L. R.; Varela, P. H.; Gomes, R. D.; Santos, R. A.; Melo, R. C. Projeto Vale do Paramirim: geologia e recursos minerais. Salvador – BA: CPRM, 2005.

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics*, 5(6), 465-473.

- Bhuiyan, M. A., Parvez, L., Islam, M. A., Dampare, S. B., & Suzuki, S. (2010). Heavy metal pollution of coal mine-affected agricultural soils in the northern part of Bangladesh. *Journal of hazardous materials*, 173(1-3), 384-392.
- Biondi, C. M., Nascimento, C. W. A. D., Fabricio Neta, A. D. B., & Ribeiro, M. R. (2011). Teores de Fe, Mn, Zn, Cu, Ni e Co em solos de referência de Pernambuco. *Revista Brasileira de Ciência do Solo*, 35, 1057-1066.
- Bivand, R., Krug, R., Neteler, M., Jeworutzki, S., 2019. rgrass7: Interface between GRASS 7 geographical information system and R. R package version 0.2–1. <https://CRAN.Rproject.org/package=rgrass7>, Accessed date: 15 January 2019.
- Brenning, A., 2008. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In: Böhner, J., Blaschke, T., Montanarella, L. (Eds.), *SAGA – Second Out (= Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie)*, 19, 23–32.
- Budiman, M., McBratney, A. B., Mendonça-Santos, M. D. L., & Santos, H. D. (2003). Revisão sobre funções de pedotransferência (PTFs) e novos métodos de predição de classes e atributos do solo (Vol. 45). Rio de Janeiro: Embrapa Solos.
- Butler, B. M., O'Rourke, S. M., & Hillier, S. (2018). Using rule-based regression models to predict and interpret soil properties from X-ray powder diffraction data. *Geoderma*, 329, 43-53.
- Calderón, G. F. A. (2009). Spatial regression analysis vs. kriging methods for spatial estimation. *International Advances in Economic Research*, 15(1), 44-58.
- Câmara, E. R. G., dos Santos, J. C. B., de Araujo Filho, J. C., Schulze, S. M. B. B., Correa, M. M., Ferreira, T. O., ... & de Souza Junior, V. S. (2021). Parent rock–pedogenesis relationship: How the weathering of metamorphic rocks influences the genesis of Planosols and Luvisols under a semiarid climate in NE Brazil. *Geoderma*, 385, 114878.

Carvalho, F. P. (2017). Mining industry and sustainable development: time for change. *Food and Energy Security*, 6(2), 61-77.

Carvalho, I. G., Iyer, S. S., Tassinari, C. C., & Misi, A. (1997). Lead-and sulfur-isotope investigations of the Boquira sediment-hosted sulfide deposit, Brazil. *International Geology Review*, 39(2), 97-106.

Carvalho, I. G., Zantop, H., & Torquato, J. R. (1982). Geologic setting and genetic interpretation of the Boquira Pb-Zn deposits, Bahia State, Brazil. *Revista Brasileira de Geociências*, 12(1-3), 414-425.

Cattle, J. A., McBratney, A. B., & Minasny, B. (2002). Kriging method evaluation for assessing the spatial distribution of urban soil lead contamination. *Journal of Environmental Quality*, 31(5), 1576-1588.

Chiles, J. and Delfiner, P. (1999) *Geostatistics: Modeling Spatial Uncertainty*. Wiley, New York, 695 p. <https://doi.org/10.1002/9780470316993>

CONAMA. (2009). Resolução CONAMA n.420. Conselho Nacional do Meio Ambiente. <http://www.mma.gov.br/port/conama/legiabre.cfm?codlegi=620>.

Costa, S. S. L., Alves, J. C., Almeida, T. S., Ribeiro, V. S., Bascuñan, V. L. A. F., Maranhão, T. A., ... & Araujo, R. G. O. (2019). Seasonality of airborne trace element sources in Aracaju, Northeastern, Brazil. *Journal of environmental management*, 247, 19-28.

COPAM. (2010). Conselho Estadual de Política Ambiental. Deliberação Normativa Conjunta COPAM/CERH nº 02, de 08 de setembro de 2010. Republicação – Diário do Executivo – “Minas Gerais” – 29/12/2010.

Csavina, J., Field, J., Taylor, M. P., Gao, S., Landázuri, A., Betterton, E. A., & Sáez, A. E. (2012). A review on the importance of metals and metalloids in atmospheric dust and aerosol from mining operations. *Science of the Total Environment*, 433, 58-73.

Cunha, F.G.; Viglio, E.P.; Anjos, J.A.S.A. & Loureiro, T.B. 2016. Estudos geoquímicos no município de Boquira – Estado da Bahia. Serviço Geológico do Brasil – CPRM. Rio de Janeiro. 44p.

Dias, R. L. S., da Silva, D. D., Fernandes-Filho, E. I., do Amaral, C. H., dos Santos, E. P., Marques, J. F., & Veloso, G. V. (2021). Machine learning models applied to TSS estimation in a reservoir using multispectral sensor onboard to RPA. *Ecological Informatics*, 65, 101414.

EPA. (2016). Risk Assessment Regional Screening Levels (RSLs) - Generic Tables Regional Screening Levels (RSLs). Regional Screening Levels (RSLs), (November 2015), 2015–2016. Retrieved Dec 24, 2019 from <https://www.epa.gov/risk/regional-screening-levels-rsls-generic-tables>.

Espourteille, F.; Fleischer, R. A mina de Boquira. In: INDA, H. A. V.: DUARTE, F. B. *Geologia e Recursos Minerais do Estado da Bahia: textos básicos*. Salvador: SME, 1980, v. 3, p. 104-125.

Evangelista, H. J., & Danderfer Filho, A. (2012). Quartzito azul com dumortierita e fosfatos de alumínio do Espinhaço setentrional, Bahia: mineralogia e petrogênese. *Revista Brasileira de Geociências*, 42(2), 363-372.

Ferreira, M., S., Fontes, M. P. F., Pacheco, A. A., Ker, J. C., & Lima, H. N. (2021). Health risks of potentially toxic trace elements in urban soils of Manaus city, Amazon, Brazil. *Environmental Geochemistry and Health*, 1-21.

Ferreira, R. G., da Silva, D. D., Elesbon, A. A. A., dos Santos, G. R., Veloso, G. V., de Souza Fraga, M., & Fernandes-Filho, E. I. (2021). Geostatistical modeling and traditional approaches for streamflow regionalization in a Brazilian Southeast watershed. *Journal of South American Earth Sciences*, 108, 103355.

Fryer, M., Collins, C. D., Ferrier, H., Colvile, R. N., & Nieuwenhuijsen, M. J. (2006). Human exposure modelling for chemical risk assessment: a review of current approaches and research and policy implications. *Environmental science & policy*, 9(3), 261-274.

Fritsch, C., Giraudoux, P., Cœurduassier, M., Douay, F., Raoul, F., Pruvot, C., ... & Scheifler, R. (2010). Spatial distribution of metals in smelter-impacted soils of woody habitats: Influence of landscape and soil properties, and risk for wildlife. *Chemosphere*, 81(2), 141-155.

Gharaibeh, M. A., Marschner, B., Heinze, S., & Moos, N. (2020). Spatial distribution of metals in soils under agriculture in the Jordan Valley. *Geoderma Regional*, 20, e00245.

Gee, G.W., Bauder, J.W., 1986. Particle-size analysis. In: Klute, A. (Ed.), *Methods of soil analysis: Physical and Mineralogical Methods Part 1*. American Society of Agronomy, Madison.

Gholami, R., Kamkar-Rouhani, A., Ardejani, F. D., & Maleki, S. (2011). Prediction of toxic metals concentration using artificial intelligence techniques. *Applied Water Science*, 1(3), 125-134.

Gloaguen, T. V., & Passe, J. J. (2017). Importance of lithology in defining natural background concentrations of Cr, Cu, Ni, Pb and Zn in sedimentary soils, northeastern Brazil. *Chemosphere*, 186, 31-42.

Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G., & Fernandes Filho, E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337-350.

Gondal, A. H., Hussain, I., Ijaz, A. B., Zafar, A., Ch, B. I., Zafar, H., ... & Tariq, M. (2021). Influence of soil pH and microbes on mineral solubility and plant nutrition: A review. *International Journal of Agriculture and Biological Sciences*, 5(1), 71-81.

Guimarães, J. T., Di Salvio, L. P. P., Medeiros, K. O. P. D., Neves, J. P. D., Pereira, L. H. M., Macêdo, L. D. L., & Fraga, L. M. B. (2014). Carta geológica: folha Boquira-SC. 23-XBV.

Hatje V, Barros F (2012) Overview of the 20th century impact of trace metal contamination in the estuaries of Todos os Santos Bay: past, present and future scenarios. *Mar Pollut Bull* 64:2603–2614.

Hujoel, I. A., Murphree Jr, D. H., Van Dyke, C. T., Choung, R. S., Sharma, A., Murray, J. A., & Rubio-Tapia, A. (2018). Machine learning in detection of undiagnosed celiac disease. *Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association*, 16(8), 1354.

John, K., Kebonye, N. M., Agyeman, P. C., & Ahado, S. K. (2021). Comparison of Cubist models for soil organic carbon prediction via portable XRF measured data. *Environmental Monitoring and Assessment*, 193(4), 1-15.

Kern, A. N., Addison, P., Oommen, T., Salazar, S. E., & Coffman, R. A. (2017). Machine learning based predictive modeling of debris flow probability following wildfire in the intermountain Western United States. *Mathematical Geosciences*, 49(6), 717-735.

Kotlar, A. M., van Lier, Q. D. J., & de Souza Brito, E. (2020). Pedotransfer functions for water contents at specific pressure heads of silty soils from Amazon rainforest. *Geoderma*, 361, 114098.

Krishna, A. K., Mohan, K. R., Murthy, N. N., Periasamy, V., Bipinkumar, G., Manohar, K., & Rao, S. S. (2013). Assessment of heavy metal contamination in soils around chromite mining areas, Nuggihalli, Karnataka, India. *Environmental earth sciences*, 70(2), 699-708.

Kuhn, M. (2008). Building predictive models in R using the caret package *J. Stat. Software*, 10.18637/jss.v028.i05.

Kuhn, M., Johnson, K., (2013). *Applied Predictive Modeling*. Springer, New York, NY.

Lima, L. D. A., Bernardez, L. A., dos Santos, M. G., & Souza, R. C. (2018).

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., (2020). *Caret: classification and regression training*.

Lima, L. D. A., Bernardez, L. A., dos Santos, M. G., & Souza, R. C. (2018). Remediation of clay soils contaminated with potentially toxic elements: The Santo Amaro lead smelter, Brazil, case. *Soil and Sediment Contamination: An International Journal*, 27(7), 573-591.

Lin, Y. P., Cheng, B. Y., Shyu, G. S., & Chang, T. K. (2010). Combining a finite mixture distribution model with indicator kriging to delineate and map the spatial patterns of soil heavy metal pollution in Chunghua County, central Taiwan. *Environmental Pollution*, 158(1), 235-244.

Maas, S., Scheifler, R., Benslama, M., Crini, N., Lucot, E., Brahmia, Z., ... & Giraudoux, P. (2010). Spatial distribution of heavy metal concentrations in urban, suburban and agricultural soils in a Mediterranean city of Algeria. *Environmental pollution*, 158(6), 2294-2301.

Matta, G., & Gjyli, L. (2016). Mercury, lead and arsenic: Impact on environment and human health. *Journal of Chemical and Pharmaceutical Sciences*, 9(2), 718–725.

McBratney, A. B., Minasny, B., Cattle, S. R., & Vervoort, R. W. (2002). From pedotransfer functions to soil inference systems. *Geoderma*, 109(1-2), 41-73.

Measham, T. G., & Zhang, A. (2019). Social licence, gender and mining: Moral conviction and perceived economic importance. *Resources Policy*, 61, 363-368.

Mello, D. C. D., Veloso, G. V., Lana, M. G. D., Mello, F. A. D. O., Poppiel, R. R., Cabrero, D. R. O., ... & Demattê, J. A. M. (2021). A new methodological framework by geophysical sensors combinations associated with machine learning algorithms to understand soil attributes. *Geoscientific Model Development Discussions*, 1-42.

Mohammadyan, M.; Moosazadeh, M.; Borji, A.; Khanjani, N. & Moghadam, S. R. Exposure to lead and its effect on sleep quality and digestive problems in soldering workers. *Environ Monit Assess*, v. 191, n. 3, 2019.

Mueller, G. (1979). Schwermetalle in den sedimenten des Rheins -Veränderungen seit 1971. *UMSCH. WISSENSCH. TECHN.*, 79(24), 778–783.

Nascimento, CWA, da Silva, FBV, Neta, ADBF, Biondi, CM, da Silva Lins, SA, de Almeida Júnior, AB, & Preston, W. (2021). As interações geopedologia-clima governam a distribuição

espacial do selênio nos solos: Um estudo de caso no nordeste do Brasil. *Geoderma*, 399, 115119.

OEHHA. (2009). Technical Support Document for Cancer Potency Factors. Office of Environmental Health Hazard Assessment. Retrieved Dec 23, 2019 from <https://oehha.ca.gov/air/crn/technical-support-document-cancer-potency-factors-2009>.

Oliveira, D. P., Sartor, L. R., Júnior, V. S., Correa, M. M., Romero, R. E., Andrade, G. R. P., & Ferreira, T. O. (2018). Weathering and clay formation in semi-arid calcareous soils from Northeastern Brazil. *Catena*, 162, 325-332.

Pan, S., Lin, L., Zeng, F., Zhang, J., Dong, G., Yang, B., ... & Ma, H. (2018). Effects of lead, cadmium, arsenic, and mercury co-exposure on children's intelligence quotient in an industrialized area of southern China. *Environmental pollution*, 235, 47-54.

Peng, Y., Kheir, R. B., Adhikari, K., Malinowski, R., Greve, M. B., Knadel, M., & Greve, M. H. (2016). Digital mapping of toxic metals in Qatari soils using remote sensing and ancillary data. *Remote Sensing*, 8(12), 1003.

PROJETO RADAMBRASIL. Folha SD-23 Brasília: geologia, geomorfologia, pedologia, vegetação, uso potencial da terra. Rio de Janeiro, 1982. 655 p

Qu, C., Wang, S., Ding, L., Zhang, M., Wang, D., & Giesy, J. P. (2018). Spatial distribution, risk and potential sources of lead in soils in the vicinity of a historic industrial site. *Chemosphere*, 205, 244-252.

Rawlins, B. G., Lark, R. M., Webster, R., & O'Donnell, K. E. (2006). The use of soil survey data to determine the magnitude and extent of historic metal deposition related to atmospheric smelter emissions across Humberside, UK. *Environmental Pollution*, 143(3), 416-426.

Reimann, C., & Garrett, R. G. (2005). Geochemical background—concept and reality. *Science of the total environment*, 350(1-3), 12-27.

Reis, G. B., da Silva, D. D., Fernandes Filho, E. I., Moreira, M. C., Veloso, G. V., de Souza Fraga, M., & Pinheiro, S. A. R. (2021). Effect of environmental covariable selection in the hydrological modeling using machine learning models to predict daily streamflow. *Journal of Environmental Management*, 290, 112625.

Reza, S. K., Baruah, U., Singh, S. K., & Das, T. H. (2015). Geostatistical and multivariate analysis of soil heavy metal contamination near coal mining area, Northeastern India. *Environmental earth sciences*, 73(9), 5425-5433.

Ribeiro, A. C., Guimarães, P. T., & Venegaz, V. H. A. (1999). 5ª Aproximação-Recomendações para o uso de corretivos e fertilizantes em Minas Gerais. Editora UFV, Viçosa, Minas Gerais.

Robert, J.H., 2019. Raster: geographic data analysis and modeling. R package version 2.8–19. <http://CRAN.R-project.org/package=raster>, Accessed date: 15 January 2019.

Rudnick, R.L., Fountain, D.M., (1995). Nature and composition of the continental crust: a lower crustal perspective.

Sahabiev, I., Smirnova, E., & Giniyatullin, K. (2021). Spatial Prediction of Agrochemical Properties on the Scale of a Single Field Using Machine Learning Methods Based on Remote Sensing Data. *Agronomy*, 11(11), 2266.

Santos, L. M. R., Gloaguen, T. V., de Souza Fadigas, F., Chaves, J. M., & Martins, T. M. O. (2017). Metal accumulation in soils derived from volcano-sedimentary rocks, Rio Itapicuru Greenstone Belt, northeastern Brazil. *Science of The Total Environment*, 601, 1762-1774.

Santos, N. L., dos Anjos, J. Â. S. A., & Klammler, H. (2020). Exposição da Zona Urbana de Boquira, estado da Bahia, aos Metais Tóxicos Associados à Bacia de Rejeito de Mineração Abandonada. *Anuário do Instituto de Geociências*, 43(3), 280-291.

Santos, P. H. R. Avaliação preliminar dos impactos ambientais da mineração de PBZN de Boquira, Bahia. 2014. Monografia. Universidade Federal da Bahia, Salvador – BA, 2014.

Sena, N. C., Veloso, G. V., Fernandes-Filho, E. I., Francelino, M. R., & Schaefer, C. E. G. (2020). Analysis of terrain attributes in different spatial resolutions for digital soil mapping application in southeastern Brazil. *Geoderma Regional*, 21, e00268.

Sergeev, A. P., Buevich, A. G., Baglaeva, E. M., & Shichkin, A. V. (2019). Combining spatial autocorrelation with machine learning increases prediction accuracy of soil heavy metals. *Catena*, 174, 425-435.

Shahmoradi, B., Hajimirzaei, S., Amanollahi, J., Wantalla, K., Maleki, A., Lee, S. M., & Shim, M. J. (2020). Influence of iron mining activity on heavy metal contamination in the sediments of the Aqyazi River, Iran. *Environmental monitoring and assessment*, 192(8), 1-10.

Shortridge, J.E., Guikema, S.D., Zaitchik, B.F., (2016). Machine learning methods for empirical streamflow simulation: a comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrol. Earth Syst. Sci.* 20, 2611–2628.

Silva, G. S., Gloaguen, T. V., Couto, C. F., & Motta, P. N. S. D. (2017). Persistence and mobility of metals in an estuarine environment 25 years after closure of a lead smelter, Bahia State, Brazil. *Environmental earth sciences*, 76(16), 1-12.

Siqueira, R. G., Veloso, G. V., Fernandes-Filho, E. I., Francelino, M. R., Schaefer, C. E. G., & Corrêa, G. R. Evaluation of machine learning algorithms to classify and map landforms in Antarctica. *Earth Surface Processes and Landforms*.

Sousa, J. E. S., dos Santos, J. C. B., Corrêa, M. M., do Nascimento, A. F., Schulze, S. M. B. B., Ferreira, T. O., ... & de Souza Júnior, V. S. (2020). Mineralogy and genesis of Planosols under a semi-arid climate, Borborema Plateau, NE Brazil. *Catena*, 184, 104260.

Souza, C. M. P., Figueredo, N. A., Costa, L. M., Veloso, G. V., Almeida, M. I. S., & Ferreira, E. J. (2020). Machine learning algorithm in the prediction of geomorphic indices for appraisal

the influence of landscape structure on fluvial systems, Southeastern-Brazil. *Revista Brasileira de Geomorfologia*, 21(2).

Souza, J. J. L. L., Abrahão, W. A. P., de Mello, J. W. V., da Silva, J., da Costa, L. M., & de Oliveira, T. S. (2015). Geochemistry and spatial variability of metal (loid) concentrations in soils of the state of Minas Gerais, Brazil. *Science of The Total Environment*, 505, 338-349.

Superintendência de estudos econômicos e sociais da Bahia (SEI). Estatística dos municípios baianos. Salvador, BA, v. 19, p. 182, 2011. Disponível em: <https://www.sei.ba.gov.br/index.php?option=com_content&view=article&id=2441&Itemid=284>. Acesso em: Maio. 2021.

Taylor, M. P., Isley, C. F., & Glover, J. (2019). Prevalence of childhood lead poisoning and respiratory disease associated with lead smelter emissions. *Environment international*, 127, 340-352.

Teixeira, P. C., Donagemma, G. K., Fontana, A., & Teixeira, W. G. (2017). Manual de métodos de análise de solo. Brasília: Embrapa, 573.

Tziachris, P., Aschonitis, V., Chatzistathis, T., & Papadopoulou, M. (2019). Assessment of spatial hybrid methods for predicting soil organic matter using DEM derivatives and soil parameters. *Catena*, 174, 206-216.

USEPA. (1989). Risk Assessment Guidance for Superfund Volume I Human Health Evaluation Manual (Part A). Office of Emergency and Remedial Response, 1(540/R/99/005), 1–291. Doi: EPA/540/1-89/002

USEPA. (1998). METHOD 3051a - Microwave assisted acid digestion of sediments, sludges, soils and oils. United States Environmental Protection Agency. <https://www.epa.gov/sites/production/files/2015-12/documents/3051a.pdf>.

USEPA. (2002). Supplemental Guidance for Developing Soil Screening. U.S. Environmental Protection Agency, (December), 106.

USEPA. (2004). Risk assessment guidance for superfund (RAGS). Volume I. Human health evaluation manual (HHEM). Part E. Supplemental guidance for dermal risk assessment
USEPA, 2004. Risk assessment guidance for superfund (RAGS). Volume I. Human health evaluation manual (H. US Epa, 1(540/R/99/005), 1–156. Doi: EPA/540/1-89/002.

USEPA. 2007. “Method 3051A (SW-846): Microwave Assisted Acid Digestion of Sediments, Sludges, and Oils,” Revision 1. Washington, DC.

Walkley, A., & Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science*, 37, 29-38.

Wei, B., & Yang, L. (2010). Uma revisão das contaminações por metais pesados em solos urbanos, poeiras de estradas urbanas e solos agrícolas da China. *Jornal de microquímica*, 94 (2), 99-107.

Wu, J., Teng, Y., Lu, S., Wang, Y., & Jiao, X. (2014). Evaluation of soil contamination indices in a mining area of Jiangxi. China. *PLoS ONE*, 9(11), 1–14.

Yan, G., Mao, L., Liu, S., Mao, Y., Ye, H., Huang, T., ... & Chen, L. (2018). Enrichment and sources of trace metals in roadside soils in Shanghai, China: a case study of two urban/rural roads. *Science of the Total Environment*, 631, 942-950.

Yang, S., Taylor, D., Yang, D., He, M., Liu, X., & Xu, J. (2021). A synthesis framework using machine learning and spatial bivariate analysis to identify drivers and hotspots of heavy metal pollution of agricultural soils. *Environmental Pollution*, 117611.

Yeomans, J. C.; Bremner, J. M. A rapid and precise method for routine determination of carbon in soil. *Commun. in Soil Sci. Plant. Anal.*, v.19, p. 1467-1476, 1988.

Zhang, C., & Yang, Y. (2020a). Modeling the spatial variations in anthropogenic factors of soil heavy metal accumulation by geographically weighted logistic regression. *Science of the Total Environment*, 717, 137096.

- Zhang, H., Yin, A., Yang, X., Fan, M., Shao, S., Wu, J., ... & Gao, C. (2021). Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils. *Ecological Indicators*, 122, 107233.
- Zhang, H., Yin, S., Chen, Y., Shao, S., Wu, J., Fan, M., ... & Gao, C. (2020b). Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. *Journal of Cleaner Production*, 273, 122858.
- Zhang, J., & Liu, C. L. (2002). Riverine composition and estuarine geochemistry of particulate metals in China - weathering features, anthropogenic impact and chemical fluxes. *Estuarine, coastal and shelf science*, 54(6), 1051-1070.
- Zhao, D., Arshad, M., Li, N., & Triantafilis, J. (2021). Predicting soil physical and chemical properties using vis-NIR in Australian cotton areas. *Catena*, 196, 104938.
- Zheng, N., Liu, J., Wang, Q., & Liang, Z. (2010). Heavy metals exposure of children from stairway and sidewalk dust in the smelting district, northeast of China. *Atmospheric environment*, 44(27), 3239-3245.

APÊNDICE

Table S1: Detection limited (DL) and quantification (QL) for elements determination procedure by ICP-OES.

	Pb	Al	Ba	Cr	Fe	Zn
	----- mg L ⁻¹ -----					
SD	3.5E-2	3.5E+1	1.1E-1	1.2E-1	8.2E+1	1.2E-1
LRS	2.3E+3	2.3E+4	4.8E+4	2.5E+4	1.4E+4	1.7E+4
LOD	4.8E-5	5.0E-4	7.0E-6	1.6E-5	1.9E-3	2.4E-5
LOQ	1.0E-4	1.5E-3	2.2E-5	4.9E-5	5.9E-3	7.2E-5

Detection Limit (DL): (3.3 x Blank Standard Deviation (SD))/ Linear regression slope (LRS); Detection quantification (QL): (10 x Blank Standard Deviation)/ Linear regression slope.

Table S2: Valores orientadores de metais nos solos

Elementos	Prevenção	Investigação			VRQ
		Agrícola	Residencial	Industrial	Minas Gerais
		mg kg ⁻¹			
Lead	72	180	300	750	19, 5
Chrome	75	150	300	400	75,00
Zinc	300	450	1000	2000	46,50
Barium	150	300	500	750	93,00
Iron		38 -81*			

*Teores médios de ferro (g kg⁻¹) para solos de Minas Gerais de acordo com Souza et al., (2015).

Table S3. Parameter values for computing human health risk assessment

Parameter	Symbol	Units	Value	Reference
Element pseudo-concentration	C	mg kg ⁻¹	-	–
Body weight children	Bw (C)	kg	15	IBGE (2010)
Body weight adult	Bw (A)	kg	70	IBGE (2010)
Averaging time children	At (C)	day	2190	<u>USEPA (2007)</u>
Averaging time adult	At (A)	day	9125	<u>USEPA (2007)</u>
Exposed skin surface area children	SA (C)	m ²	2800	<u>USEPA (2007)</u>
Exposed skin surface area adult	SA (A)	m ²	3300	<u>USEPA (2007)</u>
Dermal absorption factor	ABS	–	0.001	<u>USEPA (2007)</u>
Exposure frequency	EF	day year ⁻¹	250	<u>USEPA (2007)</u>
Exposure duration children	ED (C)	year	6	<u>USEPA (2007)</u>
Exposure duration adult	ED (A)	year	25	<u>USEPA (2007)</u>
Skin adherence factor children	SL (C)	cm ⁻² day ⁻¹	0.2	<u>USEPA (2007)</u>
Skin adherence factor adult	SL (A)	cm ⁻² day ⁻¹	0.2	<u>USEPA (2007)</u>
Ingestion rate	IR (C)	mg day ⁻¹	200	<u>USEPA (2007)</u>
Ingestion rate	IR (A)	mg day ⁻¹	100	<u>USEPA (2007)</u>
Corresponding reference dose	RfD	Pb:0.0036; Zn:0.30; Ba:0.20; Cr:0.003		<u>USEPA (2007)</u>

Table S4: Covariates used and removed in the variance and correlation removal steps.

Todas as Variáveis utilizadas	Covariáveis removidas pela variância	Covariáveis removidas por correlação
Convergence index, Cross sectional curvature, Digital elevation model, Diurnal anisotropic heating Flow line curvature, General curvature, Hill index, Geomorphons, Hill, Longitudinal curvature, Slope, Mass balance index, Maximal curvature, Morphometric Protection Index, Mid-slope position, Minimal curvature, Multiresolution index of ridge top flatness, Multiresolution index of valley bottom flatness, Normalized height, Plan curvature, Profile curvature, Real surface area, Slope height, Slope Index, Solrad Diffuse1, Solrad Diffuse2, Solrad dur 1, Solrad dur 2, Solrad Direct1, Solrad Direct2, Solrad Ration1, Solrad, Ration2, Solrad Sunrise1, Solrad Sunrise2, Solrad, Sunset1, Solrad Sunset2, Solrad total1, Solrad total2, Standardized height, Surface specific points, Tangencial curvature, Terrain ruggedness index, Terrain surface convexity, Terrain surface texture, Total curvature, Topographic position index, Valley depth, Valley, Valley Index, Vector ruggedness measure, Topographic wetness index, X (longitude), Y (latitude), PDC (Pilha canal de drenagem).	hill, hill_idx, slope_idx, valey_idx, valley	curv_general, solrad_diffuse1, curv_cross_secational, curv_tangencial, curv_longitudinal, euc_a_min, realsurfacearea, solrad_total1, solrad_ration2, solrad_direct2, solrad_total2, solrad_ration1

CAPITULO 2

Modelagem para previsão dos teores de elementos potencialmente tóxicos a partir de dados morfométricos, sensores proximais, físicos e químicos dos solos em áreas sob influência de mineração

RESUMO

Diversas atividades antrópicas vêm contribuindo para elevar os níveis de elementos potencialmente tóxicos nos solos no mundo, com destaque para a mineração. Essa atividade pode liberar elevada quantidade de fontes diretas de contaminantes no ambiente, e mesmo após a exaustão das minas, os passivos ambientais gerados podem continuar proporcionando riscos de contaminação. Os elementos potencialmente tóxicos (PTEs), quando presente no ambiente, podem entrar na cadeia trófica promovendo sérios riscos à saúde humana e ao ecossistema. Diversos métodos têm sido utilizados para determinar os teores desses elementos nos solos, porém a maioria são laboriosos, apresentam custos elevados e geram resíduos. Neste estudo, aplicamos uma estrutura metodológica para otimizar a previsão dos teores de elementos potencialmente tóxicos nos solos. Usamos um conjunto total de 120 amostras de solos, coletada na profundidade de 0-10 cm. O banco de dados de covariáveis é composto por variáveis medidas por sensores proximais, características físicas e químicas dos solos e morfométricas derivadas de um modelo de elevação digital com resolução espacial de 30 m. Foram testados cinco algoritmos de aprendizado de máquina: Florestas Aleatórias, Cubistas, Modelo Linear, Máquina de Vetor de Suporte e K Nearest Neighbor. No geral o algoritmo Cubist produziu melhores resultados na previsão dos teores de Pb, Zn, Ba e Fe em relação aos demais modelos testados. Para os teores de Al, o Máquina de Vetor de Suporte produziu a melhor previsão. Para os teores de Cr, todos os modelos apresentaram baixa predição. As covariáveis mais importantes na previsão dos teores disponíveis de elementos potencialmente tóxicos variaram de acordo ao metal estudado, no entanto as medidas por fluorescência de raios-X, texturais e morfométricas se destacaram para todos os elementos. A estrutura metodologia aqui reportada representa uma alternativa para previsão rápida e de baixo custo de PTEs nos solos, além de eficiente e econômica para o monitoramento de áreas potencialmente contaminadas e também para obtenção de valores orientados de qualidade de solos.

Palavras-chaves: Avaliação e monitoramento de solos contaminados. solos contaminados. algoritmos de aprendizado de máquina

Predictive modeling of contents of potentially toxic elements using morphometric data, proximal sensing, and chemical and physical properties of soils under mining influence

ABSTRACT

Several anthropic activities, especially mining, have contributed to the exacerbation of contents of potentially toxic elements in soils around the world. Mines can release a large amount of direct sources of contaminants into the environment, and even after the mines are no longer being exploited, the environmental liabilities generated may continue to provide contamination risks. Potentially toxic elements (PTEs), when present in the environment, can enter the food chain, promoting serious risks to human health and the ecosystem. Several methods have been used to determine the contents of PTEs in soils, but most are laborious, costly and generate waste. In this study, we use a methodological framework to optimize the prediction of levels of PTEs in soils. We used a total set of 120 soil samples, collected at a depth of 0–10 cm. The covariate database is composed of variables measured by proximal sensors, physical and chemical soil characteristics, and morphometric data derived from a DEM with a spatial resolution of 30 m. Five machine learning algorithms were tested: Random Forests, Cubist, Linear Model, Support Vector Machine and K Nearest Neighbor. In general, the Cubist algorithm produced better results in predicting the contents of Pb, Zn, Ba and Fe compared to the other tested models. For the Al contents, the Support Vector Machine produced the best prediction. For the Cr contents, all models showed low predictive power. The most important covariates in predicting the contents of PTEs varied according to the studied element. However, x-ray fluorescence measurements, textural and morphometric variables stood out for all elements. The methodology structure reported in this study represents an alternative for fast, low-cost prediction of PTEs in soils, in addition to being efficient and economical for monitoring potentially contaminated areas and obtaining quality reference values for soils.

Keywords: Assessment and monitoring of contaminated soils. contaminated soils. machine learning algorithms

1.0- INTRODUCTION

Heavy metals are natural elements commonly found in the environment; however, when at harmful concentrations, they may pose serious risks to ecosystems and human health (Liu et al., 2018; Al osman et al., 2019; Mohammadyan et al., 2019). Toxicity from excessive ingestion of potentially toxic elements (PTEs) may cause hepatotoxicity, respiratory, cardiovascular and carcinogenic problems, and immunologic, neurocognitive and behavioral damage, especially in children (Al osman et al., 2019; Renu et al., 2021).

Potentially toxic elements enter the soil mainly through natural weathering processes of rocks containing high levels of heavy metals, so that erosion and leaching spread these elements into the environment (Kabata-Pendias, 2011; Alloway, 2013; Fernandes et al., 2018; Vaziri et al., 2021). However, environments under the anthropogenic influence, such as urban, agricultural and mining sites, have contributed to the exacerbation of heavy metal levels, especially in soils, water, and sediments (Liu et al., 2018; Gujre et al., 2021; Ferreira, M. et al., 2021). Of the anthropogenic activities, mining plays a leading role as it can release a high amount of direct sources of contaminants. Even after the mines are no longer being exploited, they may continue to exhibit contamination risks as PTEs are highly persistent and have high bioaccumulation potential (Alloway, 2013; Vaziri et al., 2021).

Several reports in the world have revealed that metal mines and factories, both in activity and abandoned, may be sources of PTEs for soils, waters, and sediments (Bourhane-Eddine et al., 2013; Lima et al., 2018; Jiang et al., 2019; Pan et al., 2021). In the city of Boquira, located in the center-west region of the state of Bahia, Brazil, one of the most important lead and zinc mines in the world was exploited from the 50s to the 90s. As the mine was abandoned, a volume of 894,000 m³ of tailings was deposited inappropriately (Santos et al., 2020). As an aggravating factor, the environmental liabilities left behind are very close to the city's downtown and even include some neighborhoods. Therefore, it is necessary to measure the extent of the risks that these environmental liabilities can pose to humans and the biodiversity of the region, in addition to being against the current law (Santos et al., 2019; Santos et al., 2020).

Brazilian legislation establishes guidelines for the environmental management of contaminated areas to prevent soil and water contamination (Conama, 2009), and provides subsidies for the total or partial deactivation of potentially polluting projects and activities, and removal of environmental liabilities (INEMA Normative Instruction N°2 – 2021). Of the guidelines established in these regulations, we highlight the importance of preliminary and confirmatory assessment of the presence of toxic substances for subsequent intervention in the

area, and the determination of quality reference values (QRVs), which reflect the natural concentration of a given substance in the soil, without anthropic interference. The proper analytical determination of PTEs is the only way, standardized by law, to identify contaminated environments.

The National Council for the Environment (Conama) also provides analytical methodologies that must be used to extract inorganic substances from soils. They cite methods proposed by the United States Environmental Protection Agency, USEPA 3050b and 3051a or their updates, which are methodologies that quantify the potentially bioavailable metal fraction (Conama, 2009). Using these methods is more suitable for soil contamination studies than total extractions, as the determined levels represent the maximum amount of potentially bioavailable metals. Determining the concentrations of heavy metals in soils and their bioavailability is essential to decision making and mitigating possible risks to public health and the environment. Although accurate, the main methods consist in acid digestion followed by measurements using techniques such as spectrometry (ICP) or atomic absorption (AA), which are costly, laborious and generate chemical residues that require special treatment (Usepa, 1998; McGladdery et al., 2018; Jin et al., 2020; Weindorf and Chakraborty, 2020).

Given these limitations, traditional methods to determine PTE contents are limited to specific areas, or, when performed over large areas, the sampling density is low, resulting in inaccurate spatial variations (Shi et al., 2014). Such methods have high analysis costs and need a high sampling density due to the great diversity of soils and geochemical differentiation of the earth's crust; as a result, most Brazilian states do not have yet their respective QRVs, and, when they do, the values were obtained at low sample density (Cetesb, 2009; Copam, 2010; CPRH, 2014). Therefore, fast, accurate and ecologically sustainable methodologies are needed to determine the contents of PTEs in the soil.

Proximal sensors such as x-ray fluorescence spectrometry (XRF) and magnetic susceptibility (MS) have shown to be promising, non-destructive, fast and low-cost technologies for soil characterization (Weindorf and Chakraborty, 2020). XRF enables a multi-elemental analysis with results comparable to those from laboratory analysis for most elements, which can be directly used to assess soil parameters, including total PTE contents (Radu and Diamond, 2009; Fujimori and Takigami, 2014) or as a proxy to predict other soil attributes (Sharma et al., 2015; Andrade et al., 2020; Benedet et al., 2020). Magnetic methods are sensitive in detecting ferrimagnetic minerals, which are considered micromarkers of soil properties, and, similarly to XFR, can be used as a proxy to predict other soil attributes (Ferreira et al., 1994; Peluco et al., 2013; Ayoubi et al., 2019; Cervi et al., 2019; Li et al., 2021). With the ease of

obtaining data through proximal sensors, the development of pedotransfer functions (PTFs) may be widely used to estimate a specific soil attribute. The functions of PTFs can predict a soil attribute based on other measurements (McBratney et al., 2002; Kotlar et al., 2020). Furthermore, soil morphometric variables, obtained using digital elevation models, and satellite images can be incorporated in PTFs, thereby improving the prediction of a given attribute, as soil properties are associated with positions in the landscape.

A wide variety of theoretical frameworks has been conceived for the development of PTFs, from simple query techniques, such as linear regression, to current advanced methods that include neural networks and machine learning algorithms (Peng et al., 2016; Gomes et al., 2019; Andrade et al., 2019; al., 2020; Zhang et al., 2020). The use of machine learning algorithms is a powerful tool in creating pedotransfer functions for predicting various soil properties (Peng et al., 2016; Gomes et al., 2019; Souza et al., 2019; Kim et al., 2019; Kim et al., 2021). Thus, the objectives of the present study were (i) to test the performance of different machine learning algorithms in predicting the contents of Ba, Cr, Zn, Pb, Fe and Al in areas under the influence of mine tailings; (ii) to define the best predictive model for each PTE; and (iii) to determine and select which variables contribute to an improved predictive performance. Our main hypothesis is that the association of variables obtained by proximal sensing, remote sensing and morphometric data allows a better prediction of PTEs by different machine learning models and promotes the better performance of the models.

2.0 - MATERIALS AND METHODS

2.1- Study site and soil sampling

The study area is located in the São Francisco River basin, in the municipality of Boquira, southwest of the state of Bahia, Brazil (12°49'23" S, 42°43'50"W, and altitude of 641 m), covering an area of approximately 2800 ha (Figure 1). The climate varies from sub-humid to dry, with two well-defined seasons (winter and summer) It has an average annual rainfall of 894.8 mm, concentrated between October and April (Sei, 2011). The vegetation is characterized by Open Arboreal Caatinga, without palm trees, and Cerrado-Caatinga-Seasonal Forests. Of the soils present in the area, Latosols (Oxysol), Luvisols and Neosols stand out (Sei, 2011).

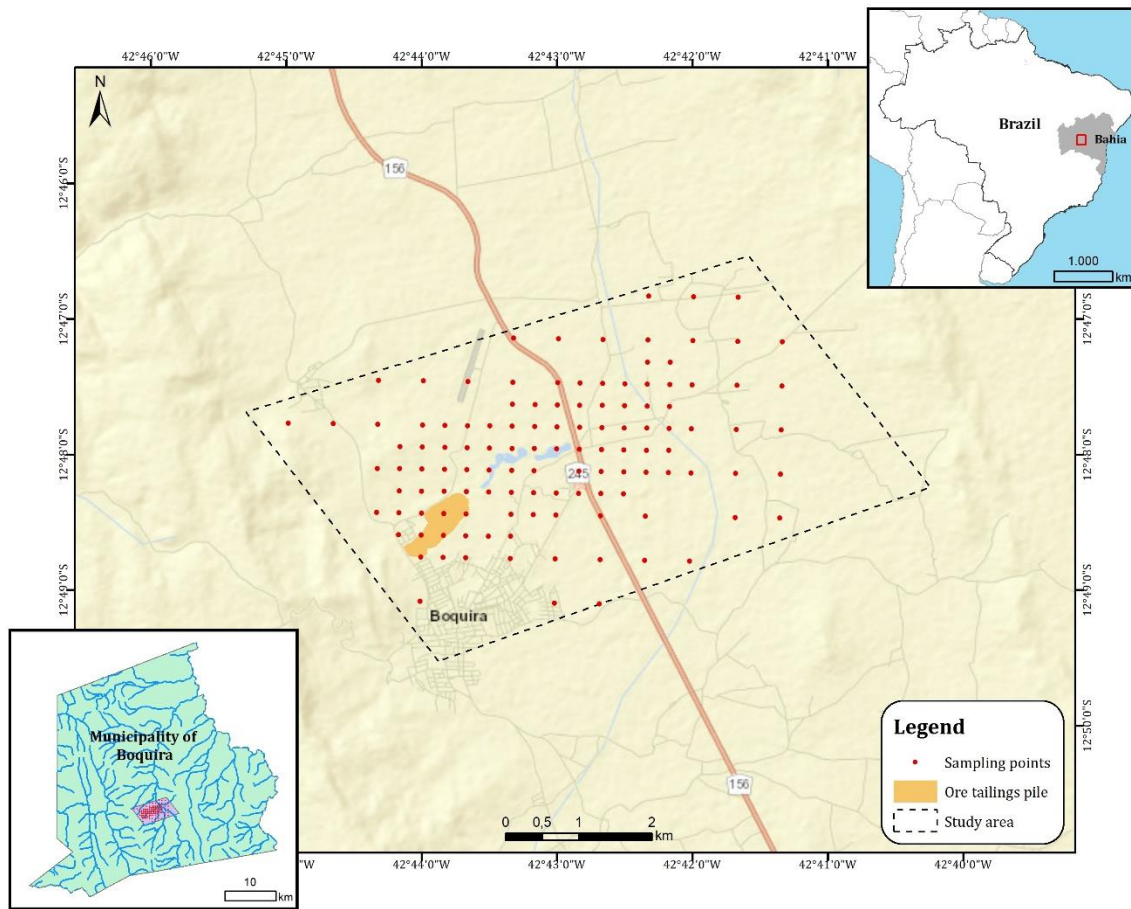


Fig 1. Map of the study site.

The area's geology is composed of a sequence of terrigenous sedimentary rocks, apparently without volcanic contribution, metamorphized in epidote-amphibolite facies consisting of ironstone formations, carbonates, quartzite, and schist. According to Spourteille and Fleischer (1980), these rocks occur in the form of lenticular bodies that interdigitate and graduate from each other. Iron formations are composed of oxide facies, silicates and carbonates at various proportions, or by quartz-hematite sub-facies, quartz-magnetite, silicate-magnetite and carbonate-silicate, with the silicate-magnetite facie associated with the mineralization of Pb and Zn (Spour and Fleischer, 1980; Rocha, 1985). The study region was chosen because the city of Boquira had, from 50s to 90s, one of the largest and most important lead mine in the world, which left environmental liabilities in the area, thus resulting in the occurrence of a wide variation in metal contents (Cunha et al., 2016; Santos et al., 2020).

The sampling points were systematically distributed on an area encompassing agricultural lands and part of the city of Boquira, based on a regular 300 m x 300 m grid in areas close to the environmental liabilities and 300 m x 450 m in distant areas. The total sampled area was 4.0 km wide and 7.0 km long, totaling 120 composite samples of approximately 28.0

km² (Figure 1). The geographical coordinates were determined at each node of the grid at which soil samples were collected. At each node, five subsamples were collected to obtain a representative composite sample. The subsamples were collected in the center and in each cardinal direction (N, S, E, and W) at a distance of about 3 to 5 meters from the central point and at a depth of 0.0 – 0.10 m. Sampling was carried out using stainless steel equipment to avoid any contamination. The samples were dried in the air and large clusters were gently broken, sieved through 2 mm sieves, and stored for further analysis.

2.2- Determination of potentially toxic elements contents and quality control

Soil aggregates were crushed in an agate mortar and sieved through a 2 mm mesh. Then, 0.500 g of sample were weighed and digested in a microwave oven (Anton Paar, MultiWave 3000) with a digestive mixture composed of 09 mL of HNO₃ and 03 mL of HCl according to the EPA 3051A method (USEPA, 1998). All solutions were prepared using ultrapure water (18.2 mΩ.cm at 25 °C – Millipore Direct-Q System) and high purity acids (Merck Pa®). The digestion procedure was carried out in triplicate for each soil sample. The quantification of the Ba, Cr, Zn, Pb, Fe and Al elements were performed by inductively coupled plasma emission spectrometry (ICP-OES, Optimal Perkinelmer 8300).

Quality control analysis was performed using SRM 2710a – Montana Soil I and SS-1 Contaminated Soil - EnviroMAT™ & AgroMAT™ certified materials. The recovery rates of SRM 2710a ranged from 65 to 87% and 87 to 88% for potentially toxic elements concentrations and main element concentrations, respectively. The SS-1 recovery rates ranged from 84 to 109% for the concentrations of all studied elements (Table 1).

Tabela 1. Recovery rates of the reference material

	Pb	Al	Ba	Cr	Fe	Zn
	----- % -----					
SRM 2710 ^a	90	88	65	72	87	87
SS-1	99	101	84	88	109	96

SS-1: Contaminated Soil - EnviroMAT™ & AgroMAT™ Certified Reference Materials (CRM);
2710a - Montana Soil I - NIST Reference Materials.

Limits of detection (LOD) and quantification (LOQ) were calculated using the equations $LOD = 3.3s/b1$ and $LOQ = 10s/b1$ (ANVISA, 2003), where s is the reference deviation of the reading of 10 blanks and b1 is the angular coefficient of the line obtained for each element. The limits of detection for the studied elements were: 4.8×10^{-5} (Pb); 7.0×10^{-6} (Ba); 1.6×10^{-5} (Cr);

2.4×10^{-5} (Zn); 1.9×10^{-3} (Fe) and 5.4×10^{-4} (Al) (Table 2). For the sampling points that were below the practicable quantification limit (PQL), PQL/2 was considered for the statistical treatments (Conama, 2009).

Table 2. Limits of detection (LOD) and quantification (LOQ) for elements determination procedure by ICP-OES

	Pb	Al	Ba	Cr	Fe	Zn
	----- mg L ⁻¹ -----					
SD	3.5E-2	3.5E+1	1.1E-1	1.2E-1	8.2E+1	1.2E-1
LRS	2.3E+3	2.3E+4	4.8E+4	2.5E+4	1.4E+4	1.7E+4
LOD	4.8E-5	5.0E-4	7.0E-6	1.6E-5	1.9E-3	2.4E-5
LOQ	1.0E-4	1.5E-3	2.2E-5	4.9E-5	5.9E-3	7.2E-5

Detection Limit (DL): $(3.3 \times \text{Blank Standard Deviation (SD)}) / \text{Linear regression slope (LRS)}$; Detection quantification (QL): $(10 \times \text{Blank Standard Deviation}) / \text{Linear regression slope}$.

2.3- Obtaining covariates to predict the contents of potentially toxic elements

To obtain the chemically and physical variables of the soils, the air-dried samples were analyzed according to Teixeira et al. (2017). The pH was determined in H₂O, at a ratio of 1:2.5, using a digital benchtop pH meter. Phosphorus, Na⁺ and K⁺ were determined using Melich-1 extractor (0.05 mol L⁻¹ HCl + 0.0125 mol L⁻¹ H₂SO₄), photocolormetry and flame photometry, respectively. Exchangeable Ca²⁺ and Mg²⁺ were extracted with 1 mol L⁻¹ KCl and determined by atomic absorption spectrophotometry. Exchangeable acidity (M^{x+}) was extracted using 1 mol L⁻¹ KCl and quantified by titration with 0.025 mol L⁻¹ NaOH. H+Al was extracted using a 0.5 mol L⁻¹ Ca(OAc)₂ extractor buffered at pH 7.0 and quantified by titration with 0.025 mol L⁻¹ NaOH. Available Fe, Zn, Cu and Mn contents were extracted using Melich-1 extractor (0.05 mol L⁻¹ HCl + 0.0125 mol L⁻¹ H₂SO₄) and measured by atomic absorption spectrophotometry. Organic C (C_{org}) content was determined via oxidation with dichromate (0.167 mol L⁻¹ K₂Cr₂O₇) in an acidic medium (Walkley and Black, 1934), using an external heat source to maximize oxidation (Yeomans and Bremner, 1988). Using the values of K⁺, Na⁺, Ca²⁺, Mg²⁺, Al³⁺ and H+Al, the sum of bases (SB), effective (CECe) and potential (CECp) cation exchange capacity, and base saturation (V) and aluminum (m) percentages were calculated.

Granulometric analysis of the soil was performed on air-dried, sieved samples (2 mm) using the pipette method (Teixeira et al. 2017): 0.1 mol L⁻¹ NaOH solution was used as a dispersant, and stirred at 50 rpm for 16 h. Coarse sand (2 – 0.212 mm) and fine sand (0.212 – 0.063 mm) fractions were obtained by sieving, while silt and clay fractions were by

sedimentation. Particle density was determined using the volumetric flask method (Teixeira et al., 2017).

The analysis of soil samples to obtain the elementary variables was performed using micro energy-dispersive (μ -EDX, 1300 Shimadzu) x-ray fluorescence (XRF), following the methodology proposed by Alves et al. (2015). From each sample, sub-samples were taken to form pressed pastilles. The soil samples were macerated using an agate mortar and passed through 75 μ m sieves. To build pastilles, 1 g of soil was weighed and pressed at 5 t cm⁻² for 2 min in a hydraulic press, with a stainless steel pastillator having a usable area of 1.755 cm² (14.95 mm diameter pastille). The device was previously calibrated using the fundamental parameters technique, together with a Montana Soil II reference sample (NIST 2711a) to obtain the sensitivity coefficient for each chemical element. The elements were determined using an average of 4800 reading points.

Air-dried, fine-textured soil samples were used to determine the magnetic susceptibility (MS) of the soil using the Bartington MS2 System (Bartington Instruments LTD, Oxford, England), coupled with an MS2B sensor. The magnetic susceptibility was measured at low (klf) (0.47 kHz) and high frequency (khf) (4.7 kHz), thus indicating the volumetric (dimensionless) magnetic susceptibility. To measure the magnetic susceptibility, the samples were placed in a 10-cm³ container and then weighed. The mass (m), in grams, was used to calculate the magnetic susceptibility per mass unit ($X_{lf} = 10 \times \text{klf} / m$) expressed in (10⁻⁸ m³ kg⁻¹). Frequency dependence on magnetic susceptibility $X_{fdf} = 100 \times [(X_{lf} - X_{hf}) / X_{hf}]$, expresses in percentage (Dearing, 1999; Kozhevnikov et al., 2014).

The topographic variables were derived from a digital elevation model (DEM) of a radar image from the Shuttle Radar Topography Mission – SRTM, at a spatial resolution of approximately 30 m (30 arc seconds). To obtain the variables, R software (R Core Team, 2021) was employed, using the “rsaga” (Brenning, 2008), “raster” (Robert, 2019) and “rgrass7” (Bivand et al., 2019) packages. The attributes generated based on the DEM are described in table 3.

Table 3. Variables of the terrain generated from the digital elevation model

Terrain attribute	Abbreviation	Brief description
Convergence index	CI	Convergence/divergence index in relation to runoff
Cross sectional curvature	CSC	Measures the curvature perpendicular to the down slope direction

Digital elevation model	DEM	Represents the elevation in each model cell
Diurnal anisotropic heating	DAH	Continuous measurement of exposure dependent energy
Flow line curvature	FLC	Represents the projection of a gradient line to a horizontal plane
General curvature	GC	The combination of both plan and profile curvatures
Geomorphons	GEO	Associated geometry using machine vision approach
Hill	HI	Analytical hill shading
Hill index	HIINDEX	Analytical index hill shading
Longitudinal curvature	LC	Measures the curvature in the down slope direction
Mass balance index	MBI	Balance index between erosion and deposition
Maximal curvature	MAXC	Maximum curvature in local normal section
Morphometric Protection Index	MPI	Measure of exposure/protection of a point from the surrounding relief
Mid-slope position	MSP	Represents the distance from the top to the valley, ranging from 0 to 1
Minimal curvature	MINC	Minimum curvature for local normal section
Multiresolution index of ridge top flatness	MRRTF	Indicates flat positions in high altitude areas
Multiresolution index of valley bottom flatness	MRVBF	Indicates flat surfaces at bottom of valley
Normalized height	NH	Vertical distance between base and ridge of normalized slope
Plan curvature	PLANC	Described as the curvature of the hypothetical contour line passing through a specific cell
Profile curvature	PROC	Describes surface curvature in the direction of the steepest incline
Real surface area	RSA	Actual calculation of cell area
Slope	S	Represents local angular slope
Slope height	SH	Vertical distance between base and ridge of slope
Slope Index	SI	Represents a local angular slope index
Solrad Diffuse1	SolDiffuse1	Insolation Diffuse for the month of January
Solrad Diffuse2	SolDiffuse2	Insolation Diffuse for the month of July
Solrad dur 1	SolDur1	Insolation Duration for the month of January
Solrad dur 2	SolDur2	Insolation Duration for the month of July
Solrad Direct1	SolDiret1	Insolation Direct of month January
Solrad Direct2	SolDiret2	Insolation Direct of the month of July
Solrad Ration1	SolRation1	Ratio between direct Insolation and diffuse Insolation of the month of January
Solrad Ration2	SolRation2	Ratio between direct Insolation and diffuse Insolation of the month of July
Solrad Sunrise1	SolSunrise1	Mean sunrise time of month January

Solrad Sunrise2	SolSunrise2	Mean sunrise time of month July
Solrad Sunset1	SolSunset1	Mean sunset time of month January
Solrad Sunset2	SolSunset2	Mean sunset time of month July
Solrad total1	SolTotal1	Total Insolation for the month of January
Solrad total2	SolTotal2	Total Insolation for the month of July
Standardized height	STANH	Vertical distance between base and standardized slope index
Surface specific points	SSP	Indicates differences between specific surface shift points
Tangencial curvature	TANC	Measured in the normal plane in a direction perpendicular to the gradient
Terrain ruggedness index	TRI	Quantitative index of topography heterogeneity
Terrain surface convexity	TSC	Ratio of the number of cells that have positive curvature to the number of all valid cells within a specified search radius
Terrain surface texture	TST	Splits surface texture into 8, 12, or 16 classes
Total curvature	TC	General measure of surface curvature
Topographic position index	TPI	Difference between the a point elevation with surrounding elevation
Valley depth	VD	Calculation of vertical distance at drainage base level
Valley	VA	Calculation fuzzy valley using the Top Hat approach
Valley Index	VA	Calculation fuzzy valley index using the Top Hat approach
Vector ruggedness measure	VRM	Measures the variation in terrain roughness
Topographic wetness index	TWI	Describes the tendency of each cell to accumulate water as a function of relief

Another variable used to predict the levels of PETs in the soils was the Euclidean distance to the main drainage channel (EUDC). EUDC was derived as the Euclidean distance, from each sampling point to the closest drainage channel. The distance to mines has been considered as an important factor in the distribution of heavy metals carried by wind or rain (Santos et al., 2019; Santos et al., 2020).

2.4- Creating pedotransfer function using machine learning

2.4.1- Removal of covariates and training data for the machine learning algorithm

The removal of covariates is a process that focuses on the creation of a pedotransfer model/function with fewer covariates, aiming at the principle of parsimony and lower computational cost (Gomes et al., 2019; Souza et al., 2020, Ferreira et al., 2020, Ferreira et al., al., 2021; Reis et al., 2021). The covariate removal processes can be separated into three steps:

removal by variance (1), removal by correlation (2), and removal by the predictive importance of each algorithm (3).

Removal by variance (1) is based on the fact that covariates with very low variance do not contribute to the predictive capability of a model and they only increase the computational cost while possibly affecting the performance of some algorithms (Hujoel et al., 2018, Kern et al., 2017; Armitage and Ober, 2010). The following covariates with a variance of zero were removed using the `nearZeroVar` function available on `Caret` (Kuhn, 2008): Hill, Hill index, Slope index, Valley index and Valley (Table 3).

The removal of covariates by correlation (2) consists in removing highly correlated covariates. Removing one of two highly correlated covariates reduces the computational cost, increases the parsimony of the final model and does not entail a loss in the model's performance. Dast et al. (2018) point out the importance of the removal by correlation as it allows the gain in importance of low-importance covariates when the model is created without applying the removal. Removal is done by calculating Spearman's correlation values for the full set of covariates. Covariate pairs that correlate greater than the set limit ($\geq 95\%$) are evaluated and the one that has the highest value of the absolute sum of correlation with all other covariates is removed. The "findcorrelation" function, available in the package `Caret` on R software, was used (Kuhn et al., 2020) to remove the following highly correlated covariates: exchangeable acidity, CECe, CO_2O_3 , cross-sectional curvature, longitudinal curvature, mass balance index, plan curvature, real surface area, Solrad Diffuse1, Solrad Direct1, Solrad Direct2, Solrad total1, and Solrad total2 (Table 3). The covariates that went through this phase joined the samples and underwent the separation processes of the training and test sets.

The training separation procedure was the nested leave-one-out cross-validation ("nested" LOOCV) (Ferreira, et al., 2021; Vergara et al., 2018; Zhang et al., 2020). The Nested-LOOCV works with double-loop training and testing procedures (Li et al., 2020, Nguyen et al., 2021). The inner loop included the removal by importance (3) step and the training of the final model. The outer loop was responsible for the assessment of the performance of the final model (Ferreira et al., 2021). The inner loop is composed of the sample set where one sample is removed ($n-2$), and the outer loop is the remaining sample. Inner loops are run n times, with n representing the set of samples. All samples go through the outer loop, where predicted values are calculated. The predicted values are paired with the observed values for the samples, thereby allowing the evaluation of the model's performance in the final prediction containing the set of all samples in the outer loop. We chose this method because of the low number of samples (120). Nested-LOOCV is a suitable method for small sample sets (up to 150 samples), where

the training and test sets generated by the hold-out method are too small for adequate modeling. Nested-LOOCV is more often used in biological than in medical studies. (Hu et al., 2020; Shbou et al., 2020; Harari et al., 2020; Nguyen et al., 2021).

The third step, called removal by importance (3), focused on reducing the computational cost and increasing the parsimony of the model by determining the ideal set of covariates for the final model generated by a base algorithm. This step was performed using the Recursive Feature Elimination (RFE) method of the package Caret (Kuhn et al., 2013). The RFE is a Backforward-type interactive covariate removal/selection method, which chooses the ideal set by evaluating different subgroups with different number of covariates defined by the operator and the complete set, always evaluating from the largest (all covariates) to the smallest set (Ferreira et al., 2021). Each algorithm has a unique selection characteristic, requiring the RFE to be executed for each algorithm under evaluation. The defined subgroups were composed of 5, 6, 7 ... 20, 25, and 30 covariates. In addition, the complete set of covariates that underwent the removal by correlation step was also tested (80). The RFE was executed and optimized by the leave-one-out cross-validation (LOOCV) and by testing five values of internal hyperparameters of each tested algorithm (Tunnelenght). The ideal subset was selected based on the mean absolute error (MAE). The tested models were: Cubist (CUB), Random Forest (RF), Support Vector Machine Radial Sigma (SVM), Linear Model (Lm), and K Nearest Neighbor (KNN). The hyperparameters of each algorithm are described in the Caret package manual in chapter 6, ("Models described" available at <https://topepo.github.io/caret/train-models-by-tag.html>). Each algorithm has an importance calculation methodology and the functions are listed and considered "support functions" of the RFE (Kuhn and Johnson, 2013). For the RF algorithm, the support function "*rfFuncs*" was used, while for the other algorithms, the support function "*caretFuncs*" was used. The ideal set of covariates for each algorithm was used in the final model of the inner loop.

The models tested in this study represent different families of machine learning algorithms. They have different characteristics for predicting PTE contents and main elements in soils, particularly for the training dataset, which can result in using different covariates for each predictive model.

Random Forest is a representative algorithm of decision trees. It has been developed for use in both classification and regression. This model is based on 'ensemble learning', which generates a large set of random decision trees during model training, which are merged into a single prediction (Breiman, 2001). Each created tree is trained using the bootstrap technique, in which samples from the original training set are randomly removed, with replacements, thereby

generating different subsets, while 1/3 of the observations of the total set is left out (out-of-bag data). Based on the out-of-bag dataset, the estimate of the training error and the importance of the predictors are calculated. Three parameters are needed to train the models: the number of trees (ntree), the minimum number of data points in each terminal node (nodesize) and the number of predictors experienced in each tree (mtry). The 'mtry' is the parameter that most affects the performance of the model (Breiman, 2002), as it is the only parameter in this study used to optimize the model in the Caret package. The other two parameters, 'nodesize' and 'ntree', are kept constant at 5 and 1000, respectively.

Cubist is a rule-based model that is an extension of the M5 model (Quinlan, 1992). As in the RF model, Cubist uses a combination of predictions generated from simple trees to obtain the final prediction because its structure consists of a conditional component, acting as a decision tree together with several linear regression models (Kuhn, 2017). The model recursively partitions the data into subsets that internally have similar characteristics in relation to the predicted variable. A regression model is then fitted based on subset data defined by the model rules. The Cubist model's advantage in relation to RF is its ability to estimate values that are not within the range of values of the observed variable. The parameters optimized by 'caret' in the Cubist model are committees and neighbors. Committees are used to create a sequential rule system, similar to boosting. The values of this parameter are greater than one and lower than 100. Neighbors are used to correct the internal rules of the model.

Support vector machines (SVM) are a set of supervised learning techniques used for classification and regression (Cortes and Vapnik, 1995). They construct a hyper-plane capable of separating training data with the greatest possible margin, leading to a lower generalization error (Ließ et al. al., 2016). The margin is constructed as the space between the decision boundary and the first sample points on each side. Points that fall outside the margin are allowed, but cost weights are assigned to them, which is important in reducing the impacts of outliers (Smola and Scholkopf, 2004). When the dataset is not easily separated, it is necessary to use non-linear functions to transform the data into a higher dimensional space allowing the model to make flexible decisions. Nonlinear regression is often obtained by applying the so-called Kernel trick, as it enables finding a hyper-plane on the upper dimensional space without increasing the computational cost. The best known Kernel function is the radial basis, whose generalization performance depends on the balance between the 'sigma' and 'cost' hyperparameters. The 'sigma' controls the degree of nonlinearity of the model while the 'cost' controls the compensation between margin and training errors (Kuhn and Johnson, 2013).

The K-Nearest Neighbor prediction method is one of the most frequently used algorithms in machine learning due to its simplicity and good performance, representing a non-parametric alternative for both classification and regression problems (Kuhn and Johnson, 2013). The prediction is performed using the distance between the current sample in relation to its K closest neighbors in the training set. The K refers to the number of samples neighboring the current sample. The K must be chosen precisely, as small and large K values ‘overfit’ and ‘underfit’ the data, respectively. KNN is an extension of the k-nearest neighbor algorithm that also takes into account the individual distances of the nearest neighbors, but uses kernel functions to weight neighbors according to their distance. The main tuning hyperparameter used for optimization is ‘kmax’, which refers to the number of k-neighbors considered (Kuhn and Johnson, 2013; Bocklitz et al., 2016).

Linear regression models are a supervised machine learning algorithm used to estimate continuous responses and result in an interpretable description of relationships between predictors and the variable to be predicted. The purpose of linear regression is to find relationships and dependencies between variables (Nasteski, 2017).

Training the model for predicting each element was performed using the ideal set of covariates for each algorithm chosen in step 3. The optimization of the final model was done using the LOOCV method and by testing five values of internal hyperparameters of each tested algorithm. The performance parameter for choosing the optimal value of the hyperparameters was the mean absolute error (MAE). The final model was used to calculate the value of the test samples (outer loop), stored with the observed value of the sample. The selection process for the importance, training and prediction of test samples was repeated 120 times (total number of samples), where the set of predicted and observed values were created and the performance of the algorithms to predict each element was evaluated (Figure 2).

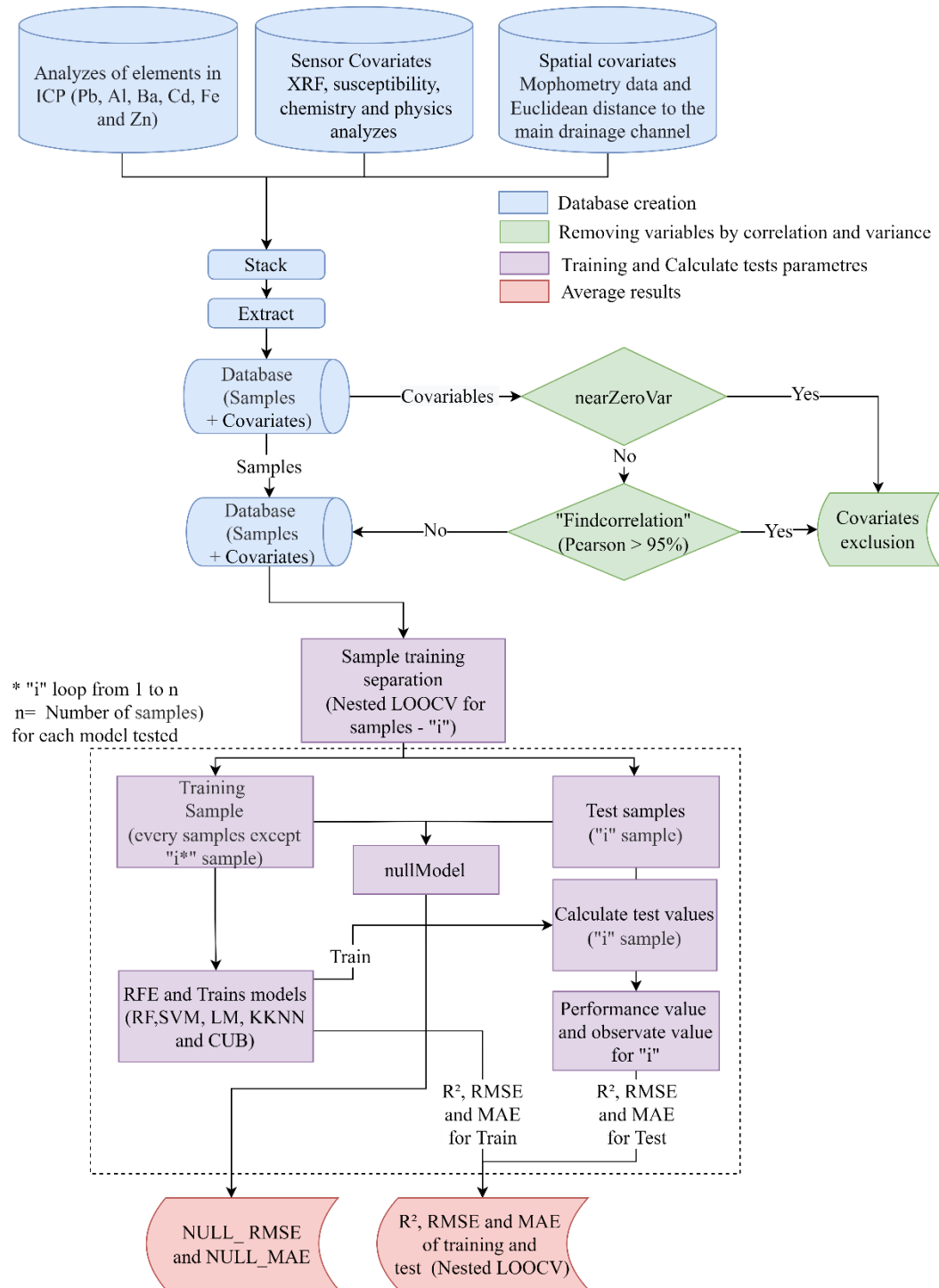


Fig 2. Methodological flowchart showing the sequence of methodologies applied for potentially toxic element (PTEs) contents prediction in soils.

2.4.2- Assessing algorithm performance

The dataset generated in the outer loop containing the predicted and observed values for each algorithm was used to determine the best-performing algorithm for each element. The performance parameters evaluated in this work were the mean absolute error – MAE (EQ.1), Root Mean Squared Error – RMSE (EQ.2) and coefficient of determination - R^2 (EQ. 3).

$$MAE = \frac{\sum_{i=1}^N |P_i - O_i|}{n} \quad (\text{EQ. 1})$$

$$RMSE = \left[\frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right]^{\frac{1}{2}} \quad (\text{EQ. 2})$$

$$R^2 = \frac{\sum_{i=1}^n (P_i - Om)^2}{\sum_{i=1}^n (Om_i - Om)^2} \quad (\text{EQ. 3})$$

Where: P_i = predicted samples, O_i = observed samples, om = mean of observed values, n = number of samples.

For comparison purposes, using the results generated for each element, the RMSE and MAE were calculated for the null model (RMSE_NULL, MAE_NULL) (EQ.4 and EQ 5). The null model uses the average value of all data for the modeled soil attribute (EQ. 4). An algorithm that predicts worse than the null model does must be removed even before the comparison with the results of other algorithms for the same attribute.

$$NULL_RMSE = \left[\frac{1}{N} \sum_{i=1}^N (Omt - O_i)^2 \right]^{\frac{1}{2}} \quad (\text{EQ.4})$$

$$NULL_MAE = \frac{\sum_{i=1}^N |Omt - O_i|}{n} \quad (\text{EQ.5})$$

Where: Omt = the mean of the training samples, O_i = the validation sample, n = number of samples (loop).

The two best algorithms for each element were also evaluated by predicted versus observed graphs (1x1); this graph can evaluate regions where the final model under- or overestimates when comparing the predictions made in the test samples with the actual data obtained from the laboratory analysis using field-collected samples. All processes and graphics were performed in R software version 4.01 (R Core Team, 2021).

3.0- RESULTS

3.1- Contents of potentially toxic elements in the studied soils

Contents of Pb, Zn, Cd, Ba, Fe and Al of the soil samples collected from the studied area revealed highly heterogeneous values, with a coefficient of variation above 400% (Table 4). The contents of Pb and Zn had values from very high, 22467.780 and 9583.699 mg kg⁻¹, to extremely low, 0.0201 and 0.0036 mg kg⁻¹, respectively. Ba and Cd levels were low in most collected samples, with mean values of 41.09 and 1.79 mg kg⁻¹, respectively. For the main elements, Fe and Al, the average contents were 31452.03 and 14505.78 mg kg⁻¹, respectively.

Table 4. Descriptive statistics of the levels of potentially toxic elements and main elements in soil samples from the area studied in the municipality of Boquira, Bahia, Brazil.

	Pb	Al	Ba	Cd	Fe	Zn
	----- mg kg ⁻¹ -----					
Maximum	22467.78	33551.10	223.27	103.37	182460.06	9583.69
Mean	735.27	14505.78	41.09	1.79	31452.03	292.19
Median	44.09	13747.26	24.52	0.002	19831.39	14.26
Minimum	0.02	1411.69	0.00	0.00	0.30	0.00
Standard Deviation	2990.11	6719.93	44.95	11.61	35890.08	1274.62
Coefficient of variation	406.67	46.33	109.40	646.13	114.11	436.23

3.2- Assessment of the precision of predictive models

The models showed good performance in predicting contents of PTEs in the soils extracted to the EPA 3051A method (Table 5). The CUB model had the highest performance for most tested variables, followed by the SVM model (Table S1). All tested models had MAE and RMSE values lower than those of the null model (Null_Model). All models are, therefore, more suitable to use than the use of average values (Figure S1, S2 and S3).

For the prediction of semi-total Pb contents, the CUB algorithm showed better performance, with lower MAE and RMSE values and with higher R². The R² values for training and testing of CUB algorithm were 0.795 and 0.988, respectively. Lm model ranked second in performance, which also had low MAE and RMSE values, though higher than those of CUB. The R² values of Lm model were 0.685 for training and 0.983 for testing. When comparing Lm's validation R² against CUB's, we see a loss in performance of only 0.51%. The remaining algorithms, RF, KNN and SVM, exhibited moderate performances, very close to each other's (Table 5).

For the prediction of Zn contents, the performance of the models was similar to that for Pb. CUB was the best-performing model, with lower MAE and RMSE values. CUB model had also the highest R², with 0.801 and 0.949 for training and testing, respectively (Table 5). Although the RF algorithm had a higher R² (0.774) for training compared to Lm, it presented lower R² (0.705) for testing and higher MAE and RMSE values. The other KNN and SVM showed moderate performance.

Despite the lower performance compared to Pb and Zn, the best prediction of the Ba contents was also obtained with the CUB algorithm, with lower MAE and RMSE values and

higher R² for training (0.552) and testing (0.433), respectively. Conversely, for Pb and Zn elements, the Lm model showed low prediction, with R² of 0.286 for testing and higher MAE and RMSE values. The second model in performance was the SVM, with MAE values close to those of CUB. The KKNN algorithm showed moderate performance in predicting the Ba contents. The RF algorithm had a high R² (0.502) for training, but the R² for testing was low and the MAE and RMSE values were high (Table 5).

In general, all models showed low predictive performance for soil Cr contents, with R² values for testing below 20%. Despite the low predictive power, the KKNN, SVM and CUB models exhibited the highest R² values, 0.183, 0.164 and 0.123 respectively, and lower MAE and RMSE values. The RF and Lm models, despite showing MAE close to those of KKNN, SVM and CUB models, had R² values lower than 10% (Table 5).

Table 5. Performance of Cubist (CUB), Linear (Lm), Random Forests (RF), K Nearest Neighbor (KKNN) and Support Vector Machine (SVM) models for the prediction of contents of potentially toxic elements (Pb, Zn, Ba, and Cr), assessed by mean absolute error (MAE), root mean squared error (RMSE) and R-squared (R²) of the training and test data sets.

		Lead		Zinc		Barium		Chrome	
		Training	Test	Training	Test	Training	Test	Training	Test
		mg kg ⁻¹							
MAE	CUB	120.97	114.07	76.33	86.37	21.24	22.77	5.31	6.39
	Lm	182.82	222.41	114.99	124.54	26.25	27.13	6.41	6.86
	KKNN	408.06	429.41	196.92	231.84	22.67	23.72	5.30	5.65
	SVM	390.99	410.64	225.40	242.06	22.73	23.28	5.26	5.85
	RF	342.70	380.57	155.47	167.09	22.58	24.60	5.87	6.60
	Null_Model	-	1216.27	-	494.44	-	32.93	-	6.83
		mg kg ⁻¹							
RMSE	CUB	264.03	341.13	193.11	311.78	30.820	34.35	7.21	8.63
	Lm	294.38	383.87	202.50	324.02	35.752	38.27	8.53	9.34
	KKNN	1040.92	1469.68	538.67	958.67	32.575	34.41	7.59	8.22
	SVM	854.26	1591.81	527.16	926.35	33.599	35.51	7.24	8.09
	RF	901.82	1627.06	427.32	697.68	32.026	37.25	7.59	8.61
	Null_Model	-	3002.65	-	1279.70	-	45.14	-	8.91
R ²	CUB	0.795	0.988	0.801	0.95	0.55	0.43	0.37	0.13
	Lm	0.685	0.983	0.586	0.94	0.38	0.28	0.20	0.04
	KKNN	0.592	0.778	0.601	0.47	0.50	0.42	0.31	0.16
	SVM	0.639	0.741	0.615	0.51	0.49	0.38	0.36	0.18
	RF	0.755	0.706	0.774	0.70	0.50	0.31	0.31	0.08

For Fe contents, the CUB model showed the highest performance, with the lowest MAE and RMSE, and R^2 values of 0.835 for testing, followed by the RF and SVM models. Lm had the highest MAE and RMSE values and lowest R^2 . The KNN model exhibited moderate performance (Table 6).

For the prediction of Al contents, all models had performance above 70%. The SVM algorithm had R^2 of 0.837 for training and 0.795 for testing and the lowest values of MAE and RMSE. The second RF algorithm, despite presenting R^2 for testing of 0.801, superior to that of SVM, the MAE and RMSE values were superior. The other models showed moderate performances (Table 6).

Table 6. Performance of Cubist (CUB), Linear (Lm), Random Forest (RF), K Nearest Neighbor (KNN) and Support Vector Machine (SVM) models in the prediction of Fe and Al contents, evaluated by absolute mean error (MAE), root mean square error (RMSE) and R-squared (R^2) of the training and test dataset.

		Iron		Aluminum	
		Training	Test	Training	Test
		mg kg ⁻¹			
MAE	CUB	5179.49	6161.42	2098.20	2276.76
	Lm	9530.62	12301.50	2145.20	2469.43
	KNN	7919.84	7994.82	2417.70	2579.80
	SVM	6943.44	7101.24	2089.88	2171.26
	RF	6463.89	6779.97	2171.70	2173.91
	Null_Model	-	24600.97	-	5368.04
		mg kg ⁻¹			
RMSE	CUB	9357.34	14859.67	2870.44	3232.34
	Lm	14821.50	24307.58	2829.29	3294.18
	KNN	13168.93	15851.95	3313.69	3626.83
	SVM	11630.66	13669.80	2809.74	3033.94
	RF	11339.24	145887.39	2910.29	2984.87
	Null_Model	-	36040.57	-	6748.11
R^2	CUB	0.90	0.83	0.84	0.77
	Lm	0.80	0.65	0.83	0.76
	KNN	0.85	0.80	0.78	0.71
	SVM	0.88	0.85	0.84	0.79
	RF	0.88	0.84	0.82	0.80

Metal contents determined in the laboratory and the values predicted by the best-performing models are shown in Figure 3, 4 and 5 for visual comparison of their predictive

capabilities. For Pb contents, the CUB model predicted values closer to those determined in the laboratory. The CUB algorithm could predict high and low values determined in the laboratory, with only a few scattered points. A similar behavior was observed for Lm model; however, it revealed a greater data dispersion compared to the CUB model, mainly for the low values of Pb.

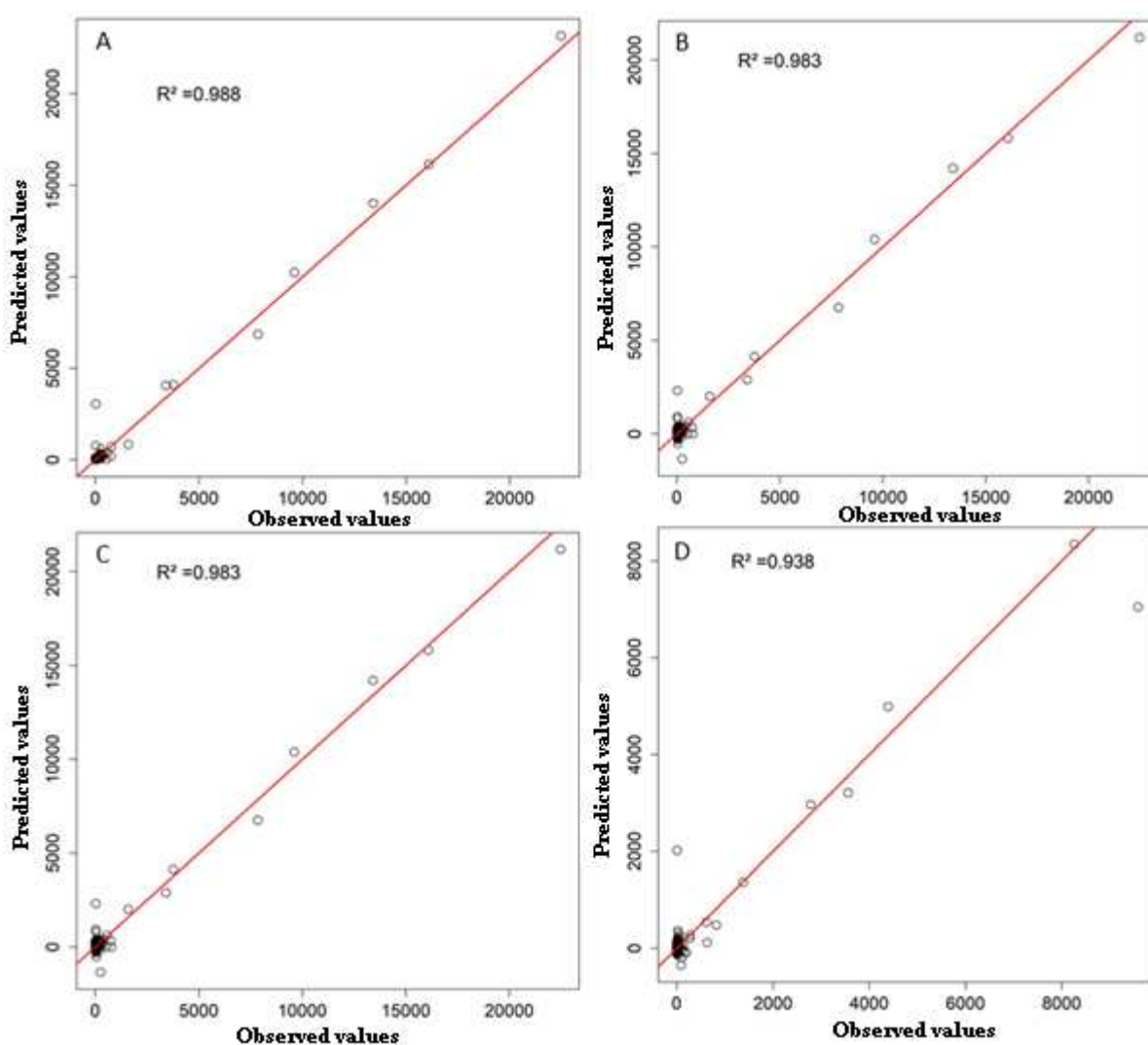


Fig 3. Predicted values versus laboratory-determined values for potentially toxic element contents (mg kg^{-1}). Predicted versus observed values of Pb using Cubist (A) and Linear Model (B) algorithms. Predicted versus observed values of Zn using Cubist (C) and Linear Model (D) algorithms.

At both high and low measured values, the CUB model over- and underestimates Zn levels, respectively. The same behavior occurred for Lm, but with greater dispersion (Figure 3).

According to figure 4, both CUB and SVM models overestimate and underestimate Ba contents; thus, a greater underestimation of the high levels of Ba was verified when using the

SVM. Graphs showing predicted versus observed values of Cr contents show a high scattering of values in relation to the 1:1 line (Figure 4).

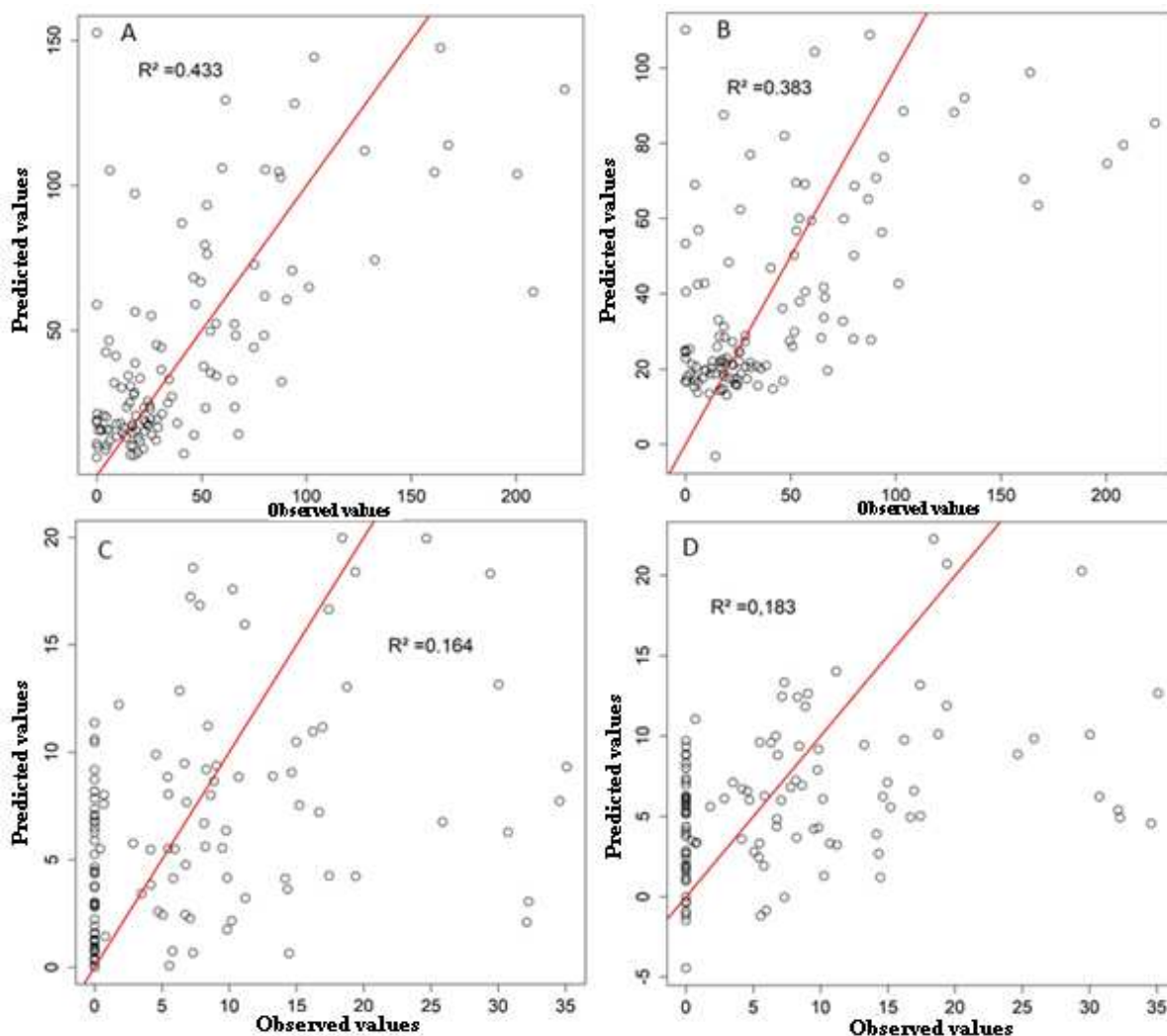


Fig 4. Predicted values versus laboratory determined values for potentially toxic element contents (mg kg^{-1}). Predicted versus observed values of Ba using the CUB (A) and SVM (B) algorithms. Predicted versus observed values of Cr using the KNN(C) and SVM (D) algorithms.

For Fe contents, the best models were CUB and RF, which showed similar behavior, both overestimating Fe contents at the smallest measured values. Trend analysis of measured versus predicted values of Al revealed that most points were closely distributed around the 1:1 line for the SVM and RF algorithms. For the prediction of extremes, both models underestimated the maximum values and overestimated the minimum values, with smaller deviations for the SVM model (Figure 5).

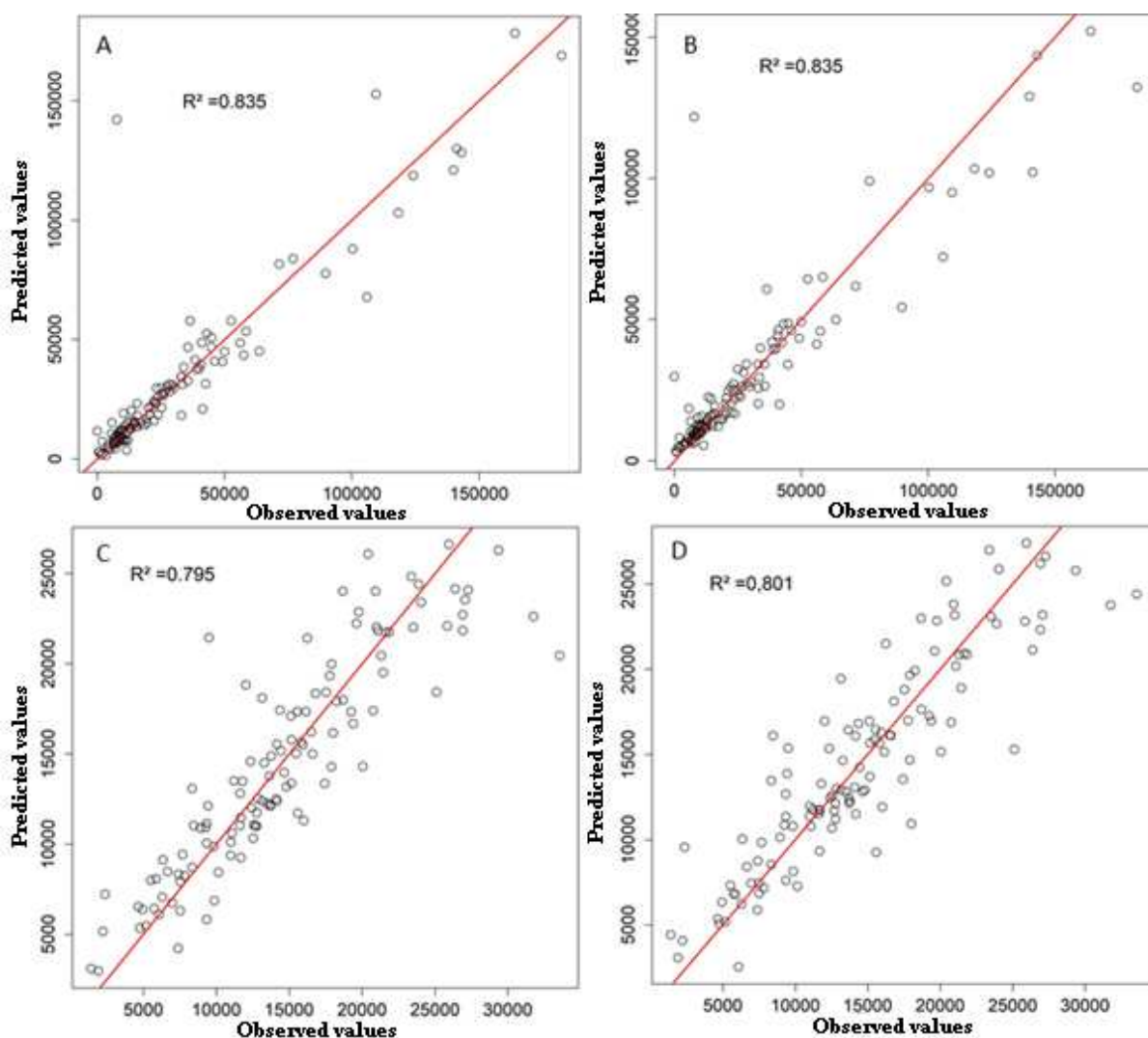


Fig 5. Predicted values versus laboratory determined values for potentially toxic element contents (mg kg^{-1}). Predicted versus observed values of Fe using the CUB (A) and RF (B) algorithms. Predicted versus observed values of Al using the SVM (C) and RF (D) algorithms.

3.3- Importance of covariates

Machine learning models selected distinct groups of covariates to predict each studied chemical element using RFE selection. Figures 6, 7 and 8 depict the average importance of each environmental covariate, considering the two best-performing models for each predicted chemical element. The main covariates selected in the CUB algorithm for available Pb contents were some variables measured by XRF, such as PbO, with an importance of 100%, ZnO and Fe_2O_3 , with importance between 25 and 50%, in addition to other oxides such as SO_3 and MnO. Texture variables were also important, especially silt and coarse sand. Similarly to the CUB model, the Lm model's main covariates for the prediction of Pb contents were measured by XRF, mainly PbO and morphometric data.

For Zn, the most important group of covariates for the CUB algorithm were data measured by XRF, morphometric and textural data such as coarse sand, fine sand, and silt. ZnO had an importance of 100%, followed by MnO and SO₃ with importance close to 50%. The most important covariates for the Lm model were measured by XRF, mainly ZnO, SO₃, ThO₂, PbO, MgO and TiO₂, in addition to the TOC contents (Figure 6). The other covariates, such as those related to fertility and magnetic susceptibility, showed intermediate to low importance for the prediction of Pb and Zn.

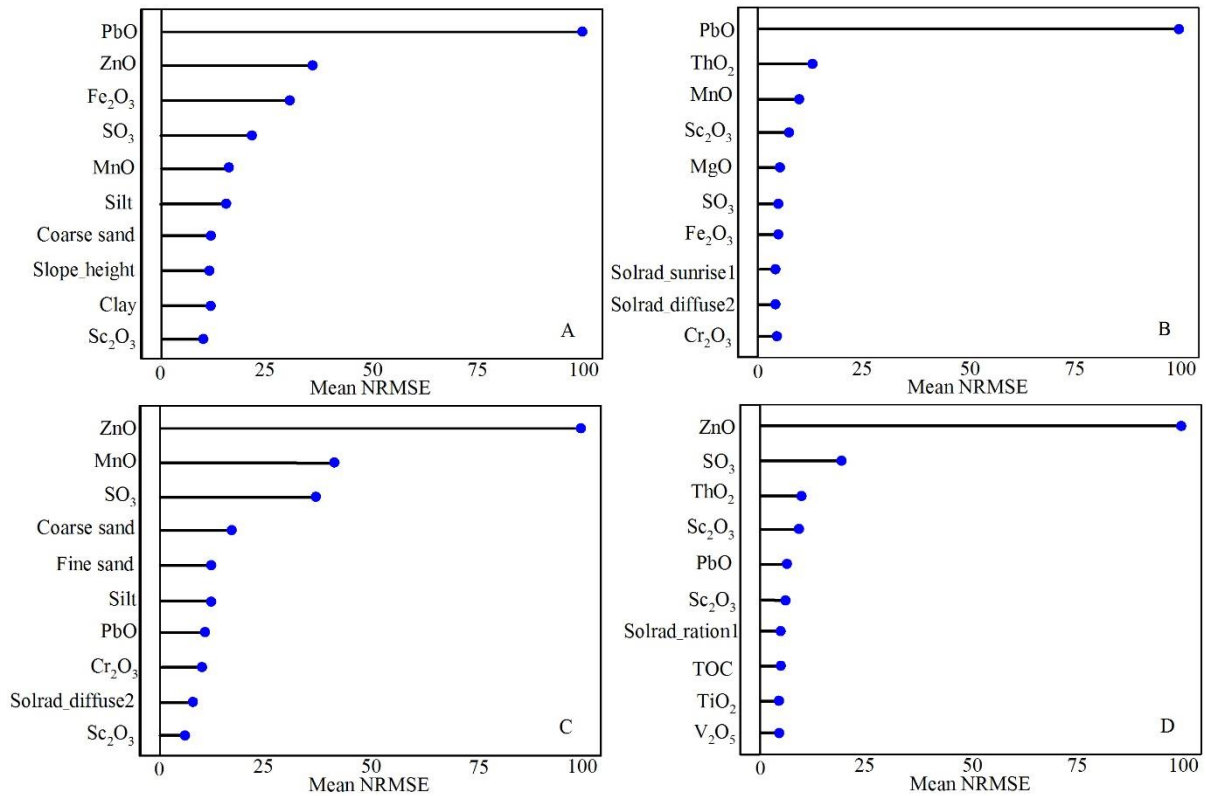


Fig 6. Importance of the main covariates in the prediction of Pb contents for Cubist (A) and Linear (B) algorithms and in the prediction of Zn contents using the CUB (C) and Lm (D) algorithms.

In the prediction of Ba contents using the CUB model, the most important variables were those related to soil texture, mainly silt and coarse sand, with importance of 100 and 25%, respectively. Morphometric covariates such as DEM, with importance above 80%, and Fe₂O₃ measured by XRF, with importance close to 75%, were also important. The EUDC covariate also showed importance above 25% in the prediction with CUB model. For the second best-performing model for Ba, the most important covariates were morphometric, textural, data measured by XRF, and TOC contents. Most variables related to these characteristics had importance greater than 50%. For Cr, the best performance was obtained with KKN and SVM

models, showing the groups of textural, morphometric and elemental covariates obtained by XRF as the most important. Clay contents assumed an importance of 100%, followed by Cl and K_2O contents, with importance above 80 and 65%, respectively. The EUDC covariate was also important to predict the contents of Cr, ranking among the top 10 in the KNN and SVM models (Figure 7).

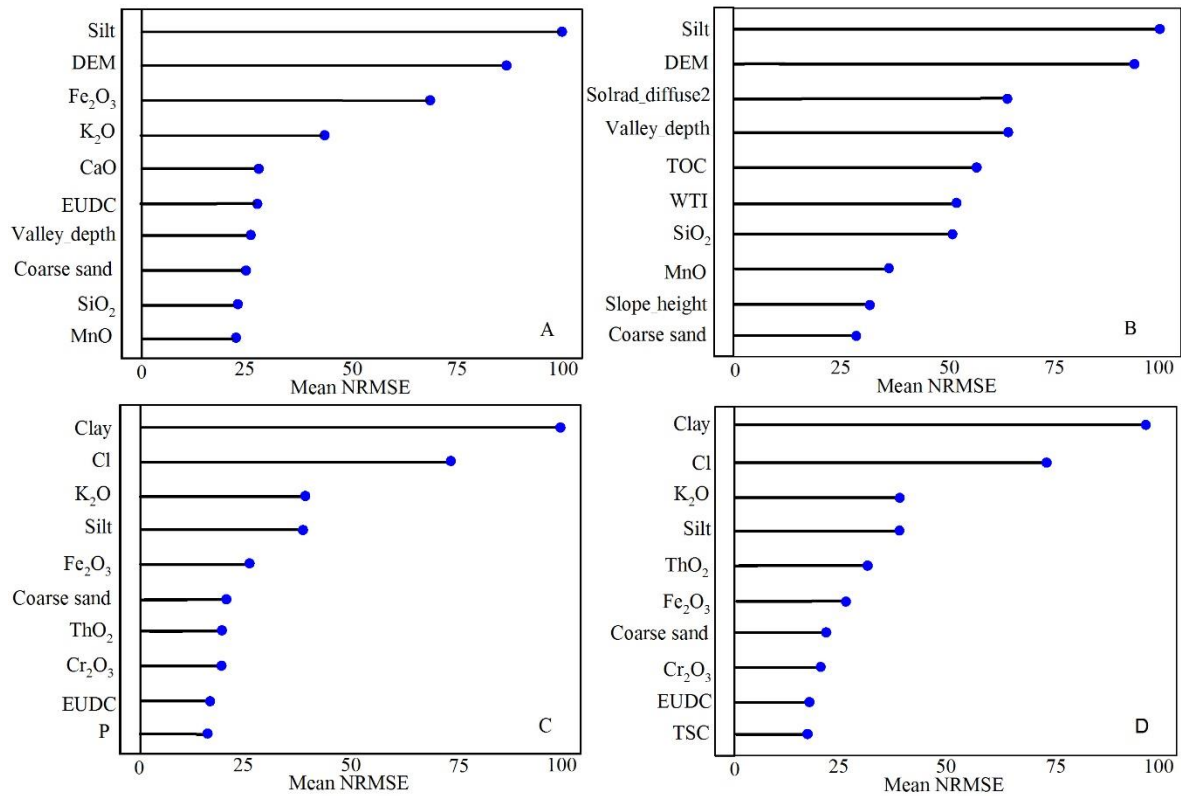


Fig 7. Importance of the main covariates for the prediction of the Ba contents for the Cubist (A) and RF (B) algorithms and for the Cr contents using the KNN (C) and SVM (D) algorithms.

In the prediction of Fe contents using the CUB model, the most important group of covariates for the prediction were those related to XRF and texture, mainly clay contents. Fe_2O_3 had importance of 100%, followed by clay contents with values above 50%. As for RF, the second best-performing model, the most important variable was measured by XRF, with Fe_2O_3 being 100% important. For the prediction of Al, the textural, morphometric and XRF covariates were the most important, both in the SVM and in the RF algorithms, with clay contents being the most important covariate in both models, followed by Al_2O_3 and TiO_2 (Figure 8).

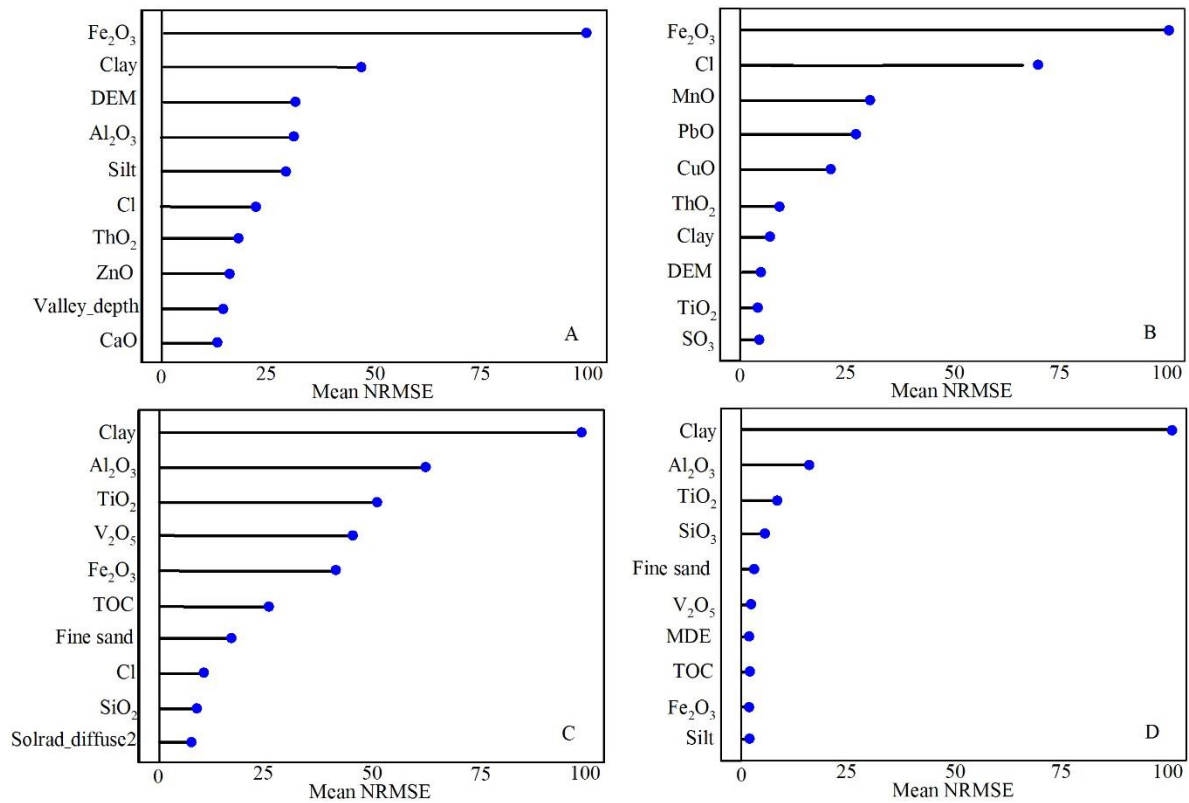


Fig 8. Importance of the main covariates for the prediction of Fe contents using Cubist (A) and RF (B) algorithms and for prediction of Al contents using the SVM (C) and RF (D) algorithms.

4-0. DISCUSSION

4.1. Assessment of the precision of predictive models

This study presents a comparison between the performance of different machine learning algorithms in predicting the concentrations of main elements and potentially toxic elements in soils, based on a set of predictors composed of morphometric variables, data measured by x-ray fluorescence, magnetic susceptibility, soil fertility, and soil physical properties.

The results indicate that the tested models showed different accuracy in predicting the contents of main elements (Fe and Al) and potentially toxic elements (Pb, Zn, Cr, Ba) in soils. For Pb and Zn contents, the CUB and Lm models showed the best performance. Despite having a lower R^2 than that of the Cubist model, the linear model is also indicated for the prediction of these elements in soils as it is a simpler and easier to understand the algorithm, and the linear structure allows us to see clearly the relationship between the variables (Ferreira et al., 2021).

The results also confirm that the CUB model is a powerful machine learning algorithm for the prediction of Ba and Fe contents in soils, as well as for Pb and Zn. The high performance of the CUB model was also verified by Peng et al. (2016) in the prediction and spatialization of

the contents of potentially toxic metals, As, Cr, Ni, Pb, Cu and Zn, in areas under the influence of mining and industry activities in Qatar.

In the literature, although several works using modeling to predict the total contents of PTEs in soils have been published, most of them use only linear models and the Random Forest algorithms (Ma et al., 2016; Zhang et al., 2020; Zhang et al., 2021). In spite of that, other models such as the CUB, SVM and KNN, may exhibit superior performance, as verified in the present study. Furthermore, in some similar studies, the RFE selection is applied only to RF model, which makes the prediction biased towards RF model (Menezes et al., 2020). For other soil properties, such as organic carbon content, CUB model has shown a high predictive capacity (Kidd et al., 2015; John et al., 2021).

The prediction of Cr and Al contents, unlike Pb, Zn, Ba and Fe, was superior when using the KNN and SVM models, respectively. This demonstrates the importance of testing different models to predict soil attributes, as the complexity, chemical forms and interactions of heavy metals with other soil properties (Kabata-Pendias, 2011; Alloway, 2013) can favor the prediction of a certain model at the expense of others. For example, Li et al. (2021) have shown that the SVM model is a powerful algorithm for the rapid prediction of heavy metal pollution in lake sediments.

The low performance of the predictive models for Ba ($R^2 = 0.433$) and Cr ($R^2 = 0.188$) contents when compared to the predictive models for the other elements may be associated with the presence of many samples with contents below the detection limit of the instrument (Ba = 10 and Cr = 50). In this situation, Brazilian law recommends using half of the practical detection limit (Conama, 2009), which may have harmed the performance of the models, as they have samples with the same contents and different characteristics (covariate).

4.2. Importance of covariates

In this study, the models selected distinct groups of covariates to predict each PTE, and x-ray fluorescence measurements were important for most of them. The contribution of covariates related to x-ray fluorescence to the predictive power of the models is mainly due to the close correlation between the levels determined by XRF and the levels obtained by inductively coupled plasma (ICP) or atomic absorption (AA), following acid digestion (Radu and Diamond, 2009; Fujimori and Takigami, 2014). Because XRF determines a wide range of chemical elements and these elements often have a geochemical affinity with each other, they contribute to the increase in efficiency of the models. This was evident, especially for the contents of Pb, Zn and Fe, which, based on the geology of the area, occur in association with

elements in a similar rock composition (Rocha, 1985; Espourteille and Fleischer, 1980). Furthermore, one of the predominant forms of Pb found in the environment is in association with Fe/Mn oxides (Kabata-Pendias, 2011; Alloway, 2013; Qu et al., 2018). Research has also demonstrated the importance of iron oxides in retaining potentially toxic elements in soils (Gupta et al., 2021; Mohammadian et al., 2021). Many studies have used XRF elementary data to predict different soil properties, as they are easily acquired (Xu et al., 2019; Andrade et al., 2020; Benedet et al., 2021), and because they have a close correlation with soil properties, which facilitates the development of predictive models.

As XRF can obtain the total content of many elements (Reidinger et al., 2012; Ribeiro et al., 2017) and the PTEs contents are a fraction of this total, other covariates should be incorporated into the models to improve their reliability and performance. As verified in the present study, the textural and morphometric data were important for the prediction in all tested models, mainly CUB and SVM.

In general, PTEs such as Pb and Zn are associated with the finest soil fractions; however, sand and silt fractions had great importance in predicting the contents of these elements in soils. This may be explained by the type of source material of the contamination, which has a coarser particle size, close to silt and fine sand fractions. In addition, the main contamination route in the study area is rain and wind transport of material from the tailing pile to adjacent areas, which also corroborates the importance of morphometric covariates in predicting these elements. Qu et al. (2018) found that, in areas close to licensed factories, concentrations of potentially toxic metals are high, mainly due to the action of weathering agents. Moreover, the geology of the area is formed by metal-rich rocks and the study area is located in a semi-arid environment, which does not favor the weathering of the soils and contributes to the presence of these elements in the coarsest fractions. Silva et al. (2020), using modeling to predict the content of elements determined via sulfuric acid attack, verified that the incorporation of soil texture variables improves the predictive performance of models, mainly because XRF analysis provides results of the total chemical composition of the sample and most potentially toxic metals are associated with the finest soil fractions. Peng et al. (2016) also reported that the use of several sets of environmental covariates (geopedological, geomorphological, anthropogenic), satellite images and spectral indices are promising in predicting the contents of potentially toxic metals in soils when using the Cubist model.

For the contents of Ba, we verified that the importance of the covariate K₂O (XRF) is mainly due to the association of these elements with geochemical processes occurring in soils. Furthermore, a high carbonate concentration favors Ba precipitation, which justifies the

importance of the CaO covariate in the prediction. Barium can also be found concentrated in Mn concretions and is specifically adsorbed on MnO and TiO (Alloway, 2013; Peana et al. 2021). The importance of the EUDC covariate is mainly due to the high concentration of Ba in the tailing pile located in the main stream channel, which decreases with increasing distance. Because it undergoes erosive processes, with a predominance of rainfall erosion, the tailing pile distributes the contaminated material along the channel drains, which supports the importance of using morphometric covariates to predict Ba in soils. Yang et al. (2021) used the RF model to identify factors and critical points of pollution caused by PTEs on agricultural soils, and demonstrated the importance of the distance of contamination sources in the concentrations of these elements in soils. In the prediction of Ba contents, the granulometric characteristic of the soil, mainly coarse sand and silt fractions, is also associated with environmental liabilities left by the mining activity, as the tailings deposited in the area consist of ground rock, with most particle size belonging to either sand or silt fractions.

Chromium occurs in soils mostly in the ionized form $\text{Cr}(\text{OH})_2^+$ and due to its positive charge, it binds to the clay and organic fraction of the soil, which predominately have negative charges (Alloway, 2013). Thus, Cr content has a positive correlation with the clay contents of the soils. In addition, Cr is usually associated with groups of minerals such as pyroxene, amphibolite, mica, as well as with the element iron (Kabata-Pendias and Mukherjee, 2007), demonstrating the importance of the covariate clay and F_2O_3 for the prediction of contents of Cr. The Fe_2O_2 and clay contents were important in predicting the Fe and Al contents in soils, contributing 100%. This is mainly due to the mineralogical composition of most tropical soils having a predominance of iron and aluminum oxides in the clay fraction (Schaefer et al., 2008; Silva L. et al., 2020). The parent material is also rich in these oxides.

In general, we can observe that machine learning algorithms performed well for the prediction of contents of all PTEs under study, except for Ba and Cr. Of the models, Cubist performed best in predicting four of the studied elements, followed by SVM and KNN models. The covariates measured by XRF were important for the prediction of all the studied models, but the use of other covariates, mainly the textural and morphometric ones, also helped in the prediction of contents of PTEs in the soils, and in some cases, it was more important than the XRF measurements. Given the wide availability of morphometric data and the easy-to-obtain variables measured by proximal and textural soil sensors, the methodology proposed in this study is an efficient and fast alternative for predicting metal contents in soils, especially for evaluation of potentially contaminated areas due to anthropic activities. In addition, the

methodology has the potential to obtain soil quality reference values (QRVs) quickly and uncostly.

5.0- CONCLUSIONS

This study demonstrates that using machine learning algorithm approaches can successfully predict the contents of potentially toxic elements and main elements in soils, using different groups of covariates.

Cubist model had the best performance in predicting Pb, Zn, Ba and Fe contents, while for the Al and Cr contents, the best performance was obtained using SVM and KNN models, respectively.

The most important covariates in the prediction of PTE contents varied according to the studied element. For Pb contents, the most important covariates in the prediction were those measured by XRF (PbO, ZnO, Fe₂O₃ and MnO). For the prediction of Zn contents, measurements were made in XRF (ZnO, MnO and SO₃), and textural (coarse sand, fine sand, and silt). In the prediction of Ba contents, textural (silt), morphometric (DEM), and XRF (Fe₂O₃, K₂O, and CaO) covariates were the ones that most influenced the prediction. As for Zn, the most important covariates to predict the Cr contents were textural (clay, silt) and XRF (Cl, K₂O, and Fe₂O) measurements. The most important covariates that influenced the prediction of Fe contents were those measured by XRF (Fe₂O₃, Al₂O₃, Cl), textural (clay, silt), and morphometric (DEM). As for Al, the best covariates were textural (clay) and XRF (Al₂O₃, TiO₂, V₂O₅ and Fe₂O₃).

As the traditional methods for the determination of PTEs in soils are laborious, costly and generate waste, the present study presents the use of machine learning algorithms as an alternative for fast and low-cost prediction of PTEs in soils. Moreover, the approach is efficient and cost-effective for monitoring potentially contaminated areas and allows obtaining guided soil quality values.

6.0- ACKNOWLEDGEMENTS

This Research was developed with the help of the UFV cluster (Federal University of Viçosa). This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Programa Nacional de Cooperação Acadêmica-PROCAD 2013 - Finance Code 001.

7.0- REFERENCES

Al osman M. Yang, F., & Massey, I. Y. (2019). Exposure routes and health effects of heavy metals on children. *Biometals*, 32(4), 563-573.

Alloway, B. J. *Heavy Metals in Soils: trace metals and metalloids in soils and their Bioavailability*. 3rd ed. New York: Springer, 2013.

Alves, E. E. N., Rocha, P. A., Siebeneichler, E. A., Costa, L. M., Vergütz, L., & Schaefer, C. E. G. R. (2015). Determinação da massa por área mínima de amostras de solo e vegetal para análise no μ -EDX. III Simpósio Mineiro de Ciência do Solo.

Anvisa, 2003. Agência Nacional de Vigilância Sanitária - Resolução Específica (RE) no 899 de 29 de maio de 2003. Determina a publicação do “Guia para validação de métodos analíticos e bioanalíticos”. Diário Oficial da União, 2.6.2003. Ministerio da Saúde - Brasil.

Armitage, D. W., & Ober, H. K. (2010). A comparison of supervised learning techniques in the classification of bat echolocation calls. *Ecological Informatics*, 5(6), 465-473.

Andrade, R., Silva, S. H. G., Weindorf, D. C., Chakraborty, S., Faria, W. M., Mesquita, L. F., ... & Curi, N. (2020). Assessing models for prediction of some soil chemical properties from portable X-ray fluorescence (pXRF) spectrometry data in Brazilian Coastal Plains. *Geoderma*, 357, 113957.

Ayoubi, S., Adman, V., & Yousefifard, M. (2019). Use of magnetic susceptibility to assess metals concentration in soils developed on a range of parent materials. *Ecotoxicology and environmental safety*, 168, 138-145.

Benedet, L., Acuna-Guzman, S. F., Faria, W. M., Silva, S. H. G., Mancini, M., dos Santos Teixeira, A. F., Pierangeli, L. M. P., Acerbi Júnior, F. W., Gomide, L. R., Pádua Júnior, A. L., Souza, I. A., Menezes, M. D., Marques, J. J., Guilherme, L. R. G., & Curi, N. (2021). Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms. *Catena*, 197, 105003.

Bivand, R., Krug, R., Neteler, M., Jeworutzki, S., 2019. rgrass7: Interface between GRASS 7 geographical information system and R. R package version 0.2–1. <https://CRAN.Rproject.org/package=rgdal>, Accessed date: 15 January 2019.

Bocklitz, T. W., Salah, F. S., Vogler, N., Heuke, S., Chernavskaia, O., Schmidt, C., ... & Popp, J. (2016). Pseudo-HE images derived from CARS/TPEF/SHG multimodal imaging in combination with Raman-spectroscopy as a pathological screening tool. *BMC cancer*, 16(1), 1-11.

Bourhane-Eddine, B., Victor, F., Amel, D., Souad, T., & Lotfi, A. (2013). What factors determine trace metal contamination in Lake Tonga (Algeria)?. *Environmental monitoring and assessment*, 185(12), 9905-9915.

Breiman, L. (2001). Random forests. *Machine learning*, 45(1), 5-32.

Breiman, L. (2002). Manual on setting up, using, and understanding random forests v3. 1. Statistics Department University of California Berkeley, CA, USA, 1(58).

Brenning, A., 2008. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In: Böhner, J., Blaschke, T., Montanarella, L. (Eds.), *SAGA – Seconds Out (= Hamburger Beiträge zur Physischen Geographie und Landschaftsökologie)*, 19, 23–32.

Cervi, E. C., Maher, B., Poliseli, P. C., de Souza Junior, I. G., & da Costa, A. C. S. (2019). Magnetic susceptibility as a pedogenic proxy for grouping of geochemical transects in landscapes. *Journal of Applied Geophysics*, 169, 109-117.

Cetesb-Environmental Agency of the State of Sao Paulo, 2005. Report on Establishment of Guiding Values for Soils and Groundwater of the State of Sao Paulo, São Paulo, Brazil. pp. 247 (in Portuguese).

Conama - National Council for the Environment, Brazil, 2009. Resolution n° 420/2009. Available online at. <http://www.mma.gov.br/port/conama/legiabre.cfm?codlegi=620>, Accessed date: 16 January 2012 (in Portuguese).

COPAM - Conselho Estadual de Política Ambiental. Deliberação Normativa Conjunta COPAM/CERH nº 02, de 08 de setembro de 2010. Republicação – Diário do Executivo – “Minas Gerais” – 29/12/2010.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.

CPRH - Companhia Pernambucana de Recursos Hídrico. Deliberação Normativa nº 07, de 07 de julho de 2014. Diário do Executivo – “Pernambuco” – 31/12/2014.

Cunha, F. G., Viglio, E. P., Anjos, J. A. S. A. & Loureiro, T. B. 2016. Estudos geoquímicos no município de Boquira – Estado da Bahia. Serviço Geológico do Brasil – CPRM. Rio de Janeiro. 44p.

Darst, B. F., Malecki, K. C., & Engelman, C. D. (2018). Using recursive feature elimination in random forest to account for correlated variables in high dimensional data. *BMC genetics*, 19(1), 1-6.

Dearing JA. (1999). Environmental magnetic susceptibility: using the Bartington MS2 system. 2nd ed. Kenilworth: Chi Publishing.

Espourteille, F.; Fleischer, R. (1980). A mina de Boquira. In: INDA, H. A. V.: DUARTE, F. B. *Geologia e Recursos Minerais do Estado da Bahia: textos básicos*. Salvador: SME, 3, 104-125.

Fernandes, A. R., de Souza, E. S., de Souza Braz, A. M., Birani, S. M., & Alleoni, L. R. F. (2018). Quality reference values and background concentrations of potentially toxic elements in soils from the Eastern Amazon, Brazil. *Journal of Geochemical Exploration*, 190, 453-463.

Ferreira, M., S., Fontes, M. P. F., Pacheco, A. A., Ker, J. C., & Lima, H. N. (2021). Health risks of potentially toxic trace elements in urban soils of Manaus city, Amazon, Brazil. *Environmental Geochemistry and Health*, 1-21.

Ferreira, R. G., da Silva, D. D., Elesbon, A. A. A., dos Santos, G. R., Veloso, G. V., de Souza Fraga, M., & Fernandes-Filho, E. I. (2021). Geostatistical modeling and traditional approaches

for streamflow regionalization in a Brazilian Southeast watershed. *Journal of South American Earth Sciences*, 108, 103355.

Ferreira, S. A. D., Santana, D. P., Fabris, J. D., Curi, N., Nunes Filho, E., & Coey, J. M. D. (1994). Relações entre magnetização, elementos traços e litologia de duas sequências de solos do Estado de Minas Gerais. *Embrapa Milho e Sorgo-Artigo em periódico indexado (ALICE)*.

Fujimori, T., & Takigami, H. (2014). Pollution distribution of heavy metals in surface soil at an informal electronic-waste recycling site. *Environmental geochemistry and health*, 36(1), 159-168.

Gomes, L. C., Faria, R. M., de Souza, E., Veloso, G. V., Schaefer, C. E. G., & Fernandes Filho, E. I. (2019). Modelling and mapping soil organic carbon stocks in Brazil. *Geoderma*, 340, 337-350.

Gujre, N., Mitra, S., Soni, A., Agnihotri, R., Rangan, L., Rene, E. R., & Sharma, M. P. (2021). Speciation, contamination, ecological and human health risks assessment of heavy metals in soils dumped with municipal solid wastes. *Chemosphere*, 262, 128013.

Gupta, K., Joshi, P., Gusain, R., & Khatri, O. P. (2021). Recent advances in adsorptive removal of heavy metal and metalloid ions by metal oxide-based nanomaterials. *Coordination Chemistry Reviews*, 445, 214100.

Harari, Y., O'Brien, M. K., Lieber, R. L., & Jayaraman, A. (2020). Inpatient stroke rehabilitation: prediction of clinical outcomes using a machine-learning approach. *Journal of neuroengineering and rehabilitation*, 17(1), 1-10.

Hu, J., Zhao, Y., Li, M., Liu, J., Wang, F., Weng, Q., ... & Cao, D. (2020). Machine learning-based radiomics analysis in predicting the meningioma grade using multiparametric MRI. *European Journal of Radiology*, 131, 109251.

Hujoel, I. A., Murphree Jr, D. H., Van Dyke, C. T., Choung, R. S., Sharma, A., Murray, J. A., & Rubio-Tapia, A. (2018). Machine learning in detection of undiagnosed celiac

disease. *Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association*, 16(8), 1354.

INEMA - Instituto do Meio Ambiente e Recursos Hídricos - Instrução Normativa - Nº 2 de 18 de junho 2021- Norma Estadual - Bahia - Publicado no DOE em 19 jun 2021.

Jiang, F., Ren, B., Hursthouse, A., Deng, R., & Wang, Z. (2019). Distribution, source identification, and ecological-health risks of potentially toxic elements (PTEs) in soil of thallium mine area (southwestern Guizhou, China). *Environmental Science and Pollution Research*, 26(16), 16556-16567.

Jin, M., Yuan, H., Liu, B., Peng, J., Xu, L., & Yang, D. (2020). Review of the distribution and detection methods of heavy metals in the environment. *Analytical Methods*, 12(48), 5747-5766.

John, K., Kebonye, N. M., Agyeman, P. C., & Ahado, S. K. (2021). Comparison of Cubist models for soil organic carbon prediction via portable XRF measured data. *Environmental Monitoring and Assessment*, 193(4), 1-15.

Kabata-Pendias, A. Trace elements in soils and plants. 4th ed. Taylor and Francis Group, 2011.

Kabata-Pendias, Mukherjee, 2007. Trace elements from soil to human. Springer Science & Business Media.

Kern, A. N., Addison, P., Oommen, T., Salazar, S. E., & Coffman, R. A. (2017). Machine learning based predictive modeling of debris flow probability following wildfire in the intermountain Western United States. *Mathematical Geosciences*, 49(6), 717-735.

Kidd, D., Webb, M., Malone, B., Minasny, B., & McBratney, A. (2015). Eighty-metre resolution 3D soil-attribute maps for Tasmania, Australia. *Soil Research*, 53(8), 932-955.

Kim, S., Karahan, G., Sharma, M., & Pachepsky, Y. (2021). Estimating parameters of empirical infiltration models from the global dataset using machine learning. *International Agrophysics*, 35(1), 73-81.

Kozhevnikov, N. O., Kamnev, Y. K., & Kazansky, A. Y. (2014). Error analysis of frequency-dependent magnetic susceptibility measurements: magnetic viscosity studies with the Bartington MS2 system. *Russian Geology and Geophysics*, 55(4), 508-514.

Kotlar, A. M., van Lier, Q. D. J., & de Souza Brito, E. (2020). Pedotransfer functions for water contents at specific pressure heads of silty soils from Amazon rainforest. *Geoderma*, 361, 114098.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of statistical software*, 28(1), 1-26.

Kuhn, M. & Johnson, K. (2013). *Applied Predictive Modeling*. Springer, New York, NY.

Kuhn, M. (2017) Classification and regression with random forest R package version 4.6–12. <http://CRAN.R-project.org/package=randomForest>

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., ... & Benesty, M. (2019). R Package Caret: Classification and Regression Training. *R Packag.* version 6.0, 84.

Ließ, M., Schmidt, J., & Glaser, B. (2016). Improving the spatial prediction of soil organic carbon stocks in a complex tropical mountain landscape by methodological specifications in machine learning approaches. *PLoS One*, 11(4), e0153673.

Li, X., Yang, Y., Yang, J., Fan, Y., Qian, X., & Li, H. (2021). Rapid diagnosis of heavy metal pollution in lake sediments based on environmental magnetism and machine learning. *Journal of Hazardous Materials*, 416, 126163.

Li, W., Hu, X., Long, X., Tang, L., Chen, J., Wang, F., & Zhang, D. (2020). EEG responses to emotional videos can quantitatively predict big-five personality traits. *Neurocomputing*, 415, 368-381.

Lima, L. D. A., Bernardez, L. A., dos Santos, M. G., & Souza, R. C. (2018). Remediation of clay soils contaminated with potentially toxic elements: The Santo Amaro lead smelter, Brazil, case. *Soil and Sediment Contamination: An International Journal*, 27(7), 573-591.

Liu, J., Liu, Y. J., Liu, Y., Liu, Z., & Zhang, A. N. (2018). Quantitative contributions of the major sources of heavy metals in soils to ecosystem and human health risks: A case study of Yulin, China. *Ecotoxicology and environmental safety*, 164, 261-269.

Ma, W., Tan, K., & Du, P. (2016). Predicting soil heavy metal based on Random Forest model. In 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS) (4331-4334). IEEE.

McBratney, A. B., Minasny, B., Cattle, S. R., & Vervoort, R. W. (2002). From pedotransfer functions to soil inference systems. *Geoderma*, 109(1-2), 41-73.

McGladdery, C., Weindorf, D. C., Chakraborty, S., Li, B., Paulette, L., Podar, D., ... & Duda, B. (2018). Elemental assessment of vegetation via portable X-ray fluorescence (PXRF) spectrometry. *Journal of Environmental Management*, 210, 210-225.

Menezes, M. D., Bispo, F. H. A., Faria, W. M., Gonçalves, M. G. M., Curi, N., & Guilherme, L. R. G. (2020). Modeling arsenic content in Brazilian soils: What is relevant?. *Science of The Total Environment*, 712, 136511.

Mohammadian, S., Krok, B., Fritzsche, A., Bianco, C., Tosco, T., Cagigal, Mata, B., Gonzalez, V., Diez-Ortiz, M., Ramos, V., Montalvo, D., Smolders, E., Sethi, R., & Meckenstock, R. U. (2021). Field-scale demonstration of in situ immobilization of heavy metals by injecting iron oxide nanoparticle adsorption barriers in groundwater. *Journal of Contaminant Hydrology*, 237, 103741.

Mohammadyan, Mahmoud, Moosazadeh, Mahmood, Borji, Abasalt, Khanjani, Narges, Moghadam, Somayeh Rahimi, 2019. Exposure to lead and its effect on sleep quality and digestive problems in soldering workers. *Environmental monitoring and assessment* 191 (3), 1–9.

Nasteski, V. (2017). An overview of the supervised machine learning methods. *Horizons. b*, 4, 51-62.

Nguyen, K. P., Raval, V., Treacher, A., Mellema, C., Yu, F. F., Pinho, M. C., Subramaniam, R. M., Dewey Jr, R. B., & Montillo, A. A. (2021). Predicting Parkinson's disease trajectory using clinical and neuroimaging baseline measures. *Parkinsonism & related disorders*, 85, 44-51.

Pan, L., Guan, X., Liu, B., Chen, Y., Pei, Y., Pan, J., Zhang, Y., & Hao, Z. (2021). Pollution Characteristics, Distribution and Ecological Risk of Potentially Toxic Elements in Soils from an Abandoned Coal Mine Area in Southwestern China. *Minerals*, 11(3), 330.

Peluco, R. G., Marques Júnior, J., Siqueira, D. S., Cortez, L. A., & Pereira, G. T. (2013). Magnetic susceptibility in the prediction of soil attributes in two sugarcane harvesting management systems. *Engenharia Agrícola*, 33, 1134-1143.

Peana, M., Medici, S., Dadar, M., Zoroddu, M. A., Pelucelli, A., Chasapis, C. T., & Bjørklund, G. (2021). Environmental barium: potential exposure and health-hazards. *Archives of Toxicology*, 1-8.

Peng, Y., Kheir, R. B., Adhikari, K., Malinowski, R., Greve, M. B., Knadel, M., & Greve, M. H. (2016). Digital mapping of toxic metals in Qatari soils using remote sensing and ancillary data. *Remote Sensing*, 8(12), 1003.

Qu, C., Wang, S., Ding, L., Zhang, M., Wang, D., & Giesy, J. P. (2018). Spatial distribution, risk and potential sources of lead in soils in the vicinity of a historic industrial site. *Chemosphere*, 205, 244-252.

Quinlan, J. R. (1992, November). Learning with continuous classes. In *5th Australian joint conference on artificial intelligence*, 92, 343-348.

R DEVELOPMENT CORE TEAM et al. The R project for statistical computing. 2021.

Radu, T., & Diamond, D. (2009). Comparison of soil pollution concentrations determined using AAS and portable XRF techniques. *Journal of Hazardous Materials*, 171(1-3), 1168-1171.

Reidinger, S., Ramsey, M. H., & Hartley, S. E. (2012). Rapid and accurate analyses of silicon and phosphorus in plants using a portable X-ray fluorescence spectrometer. *New Phytologist*, 195(3), 699-706.

Reis, G. B., da Silva, D. D., Fernandes Filho, E. I., Moreira, M. C., Veloso, G. V., de Souza Fraga, M., & Pinheiro, S. A. R. (2021). Effect of environmental covariable selection in the hydrological modeling using machine learning models to predict daily streamflow. *Journal of Environmental Management*, 290, 112625.

Renu, K., Chakraborty, R., Myakala, H., Kote, R., Famurewa, A. C., Madhyastha, H., Vellingiri, B., George, A., & Gopalakrishnan, A. V. (2021). Molecular mechanism of heavy metals (Lead, Chromium, Arsenic, Mercury, Nickel and Cadmium) induced hepatotoxicity—A review. *Chemosphere*, 129735.

Ribeiro, B. T., Silva, S. H. G., Silva, E. A., & Guilherme, L. R. G. (2017). Portable X-ray fluorescence (pXRF) applications in tropical Soil Science. *Ciência e Agrotecnologia*, 41, 245-254.

Robert, J.H., 2019. Raster: geographic data analysis and modeling. R package version 2.8–19. <http://CRAN.R-project.org/package=raster>, Accessed date: 15 January 2019.

Rocha, G. M. F. Caracterização faciológica da formação ferrífera de Boquira, encaixante damineralização de Pb/Zn. 1985. Dissertação de mestrado. Instituto de Geociências da Universidade Federal da Bahia, Salvador – BA, 1985.

Santos, N. L., dos Anjos, J. Â. S. A., & Klammler, H. (2020). Exposição da Zona Urbana de Boquira, estado da Bahia, aos Metais Tóxicos Associados à Bacia de Rejeito de Mineração Abandonada. *Anuário do Instituto de Geociências*, 43(3), 280-291.

Santos, N. L., Nery, V. S., & Assumpção, H. C. P. (2019). Caminhos do Chumbo na Bahia: Histórico e Perspectivas de Exploração Mineral no Município de Boquira/BA. *Revista Brasileira de Geografia Física*, 12(07), 2603-2616.

Schaefer, C., Fabris, J., & Ker, J. (2008). Minerals in the clay fraction of Brazilian Latosols (Oxisols): A review. *Clay Minerals*, 43(1), 137-154. doi:10.1180/claymin.2008.043.1.11

Sharma, A., Weindorf, D. C., Wang, D., & Chakraborty, S. (2015). Characterizing soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). *Geoderma*, 239, 130-134.

Shboul, Z. A., Chen, J., & Iftekharuddin, K. M. (2020). Prediction of molecular mutations in diffuse low-grade gliomas using MR imaging features. *Scientific reports*, 10(1), 1-13.

Shi, T., Liu, H., Wang, J., Chen, Y., Fei, T., & Wu, G. (2014). Monitoring arsenic contamination in agricultural soils with reflectance spectroscopy of rice plants. *Environmental science & technology*, 48(11), 6264-6272.

Silva, L. S., Júnior, J. M., Barrón, V., Gomes, R. P., Teixeira, D. D. B., Siqueira, D. S., & Vasconcelos, V. (2020). Spatial variability of iron oxides in soils from Brazilian sandstone and basalt. *Catena*, 185, 104258.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, 14(3), 199-222.

Souza, C. M. P., Figueredo, N. A., Costa, L. M., Veloso, G. V., Almeida, M. I. S., & Ferreira, E. J. (2020). Machine learning algorithm in the prediction of geomorphic indices for appraisal the influence of landscape structure on fluvial systems, Southeastern-Brazil. *Revista Brasileira de Geomorfologia*, 21(2).

Souza, C. M. P. D., Thomazini, A., Schaefer, C. E. G. R., Veloso, G. V., Moreira, G. M., & Fernandes, E. I. (2018). Multivariate analysis and machine learning in properties of Ultisols (Argissolos) of Brazilian amazon. *Revista Brasileira de Ciência do Solo*, 42.

Superintendência de estudos econômicos e sociais da Bahia (SEI). Estatística dos municípios baianos. Salvador, BA, v. 19, p. 182, 2011. Disponível

em:<https://www.sei.ba.gov.br/index.php?option=com_content&view=article&id=2441&Itemid=284 >. Acesso em: Maio. 2021.

Teixeira, P. C., Donagemma, G. K., Fontana, A., & Teixeira, W. G. (2017). Manual de métodos de análise de solo. Brasília: Embrapa, 573.

USEPA. (1998). METHOD 3051a - Microwave assisted acid digestion of sediments, sludges, soils and oils. United States Environmental Protection Agency.<https://www.epa.gov/sites/production/files/2015-12/documents/3051a.pdf>.

Vaziri, A., Nazarpour, A., Ghanavati, N., Babainejad, T., & Watts, M. J. (2021). An integrated approach for spatial distribution of potentially toxic elements (Cu, Pb and Zn) in topsoil. *Scientific Reports*, 11(1), 1-16.

Vergara, V. M., Mayer, A. R., Kiehl, K. A., & Calhoun, V. D. (2018). Dynamic functional network connectivity discriminates mild traumatic brain injury through machine learning. *NeuroImage: Clinical*, 19, 30-37.

Walkley, A., & Black, I. A. (1934). An examination of the Degtjareff method for determining soil organic matter and a proposed modification of the chromic acid titration method. *Soil Science*, 37, 29-38.

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

Xu, D., Chen, S., Rossel, R. V., Biswas, A., Li, S., Zhou, Y., & Shi, Z. (2019). X-ray fluorescence and visible near infrared sensor fusion for predicting soil chromium content. *Geoderma*, 352, 61-69.

Yeomans, J. C., & Bremner, J. M. (1988). A rapid and precise method for routine determination of organic carbon in soil. *Communications in soil science and plant analysis*, 19(13), 1467-1476.

Yang, S., Taylor, D., Yang, D., He, M., Liu, X., & Xu, J. (2021). A synthesis framework using machine learning and spatial bivariate analysis to identify drivers and hotspots of heavy metal pollution of agricultural soils. *Environmental Pollution*, 117611.

Zhang, H., Yin, S., Chen, Y., Shao, S., Wu, J., Fan, M., Furong, C., & Gao, C. (2020). Machine learning-based source identification and spatial prediction of heavy metals in soil in a rapid urbanization area, eastern China. *Journal of Cleaner Production*, 273, 122858.

Zhang, Y., Liu, S., & Yu, X. (2020). Longitudinal structural MRI analysis and classification in Alzheimer's disease and mild cognitive impairment. *International Journal of Imaging Systems and Technology*, 30(2), 421-433.

Zhang, H., Yin, A., Yang, X., Fan, M., Shao, S., Wu, J., Wu, P., Zhang, M., & Gao, C. (2021). Use of machine-learning and receptor models for prediction and source apportionment of heavy metals in coastal reclaimed soils. *Ecological Indicators*, 122, 107233.

APÊNDICE

Table S1. Summary of the best models for predicting the contents of potentially toxic elements and main elements in soils

Elements	Models	MAE	RMSE	R ²
		----- mg kg ⁻¹ -----		
Lead	Cubist	114.066	341.130	0.988
Zinc	Cubist	86.366	311.774	0.949
Barium	Cubist	22.773	34.348	0.433
Chrome	K Nearest Neighbor	5.651	8.218	0.164
Iron	Cubist	6161.419	14859.673	0.835
Aluminum	Support Vector Machine	2171.262	3033.944	0.795

Figure S1 – Performance of Cubist (CUB), Linear (Lm), Random Forests (RF), K Nearest Neighbor (KKNN) and Support Vector Machine (SVM) models for the prediction of contents of potentially toxic elements (Pb and Zn), assessed by mean absolute error (MAE) and R-squared (R^2) of the training and test data sets.

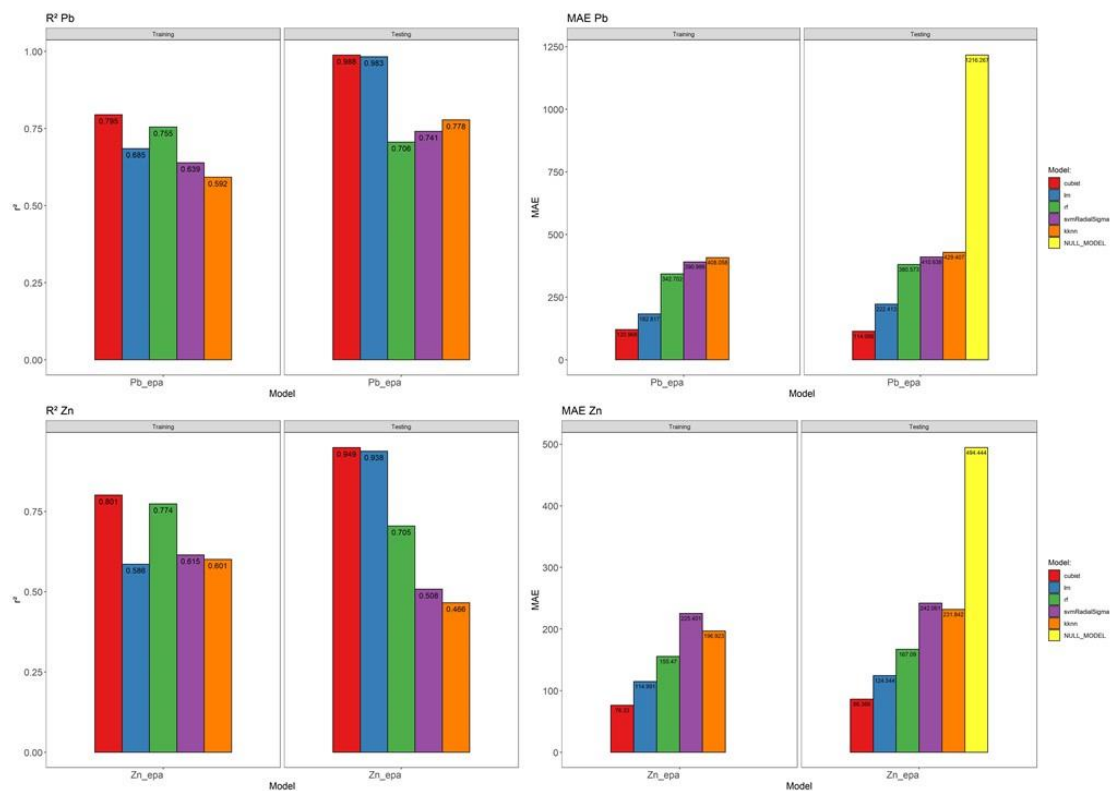


Figure S2 – Performance of Cubist (CUB), Linear (Lm), Random Forests (RF), K Nearest Neighbor (KKNN) and Support Vector Machine (SVM) models for the prediction of contents of potentially toxic elements (Ba and Cr), assessed by mean absolute error (MAE) and R-squared (R^2) of the training and test data sets.

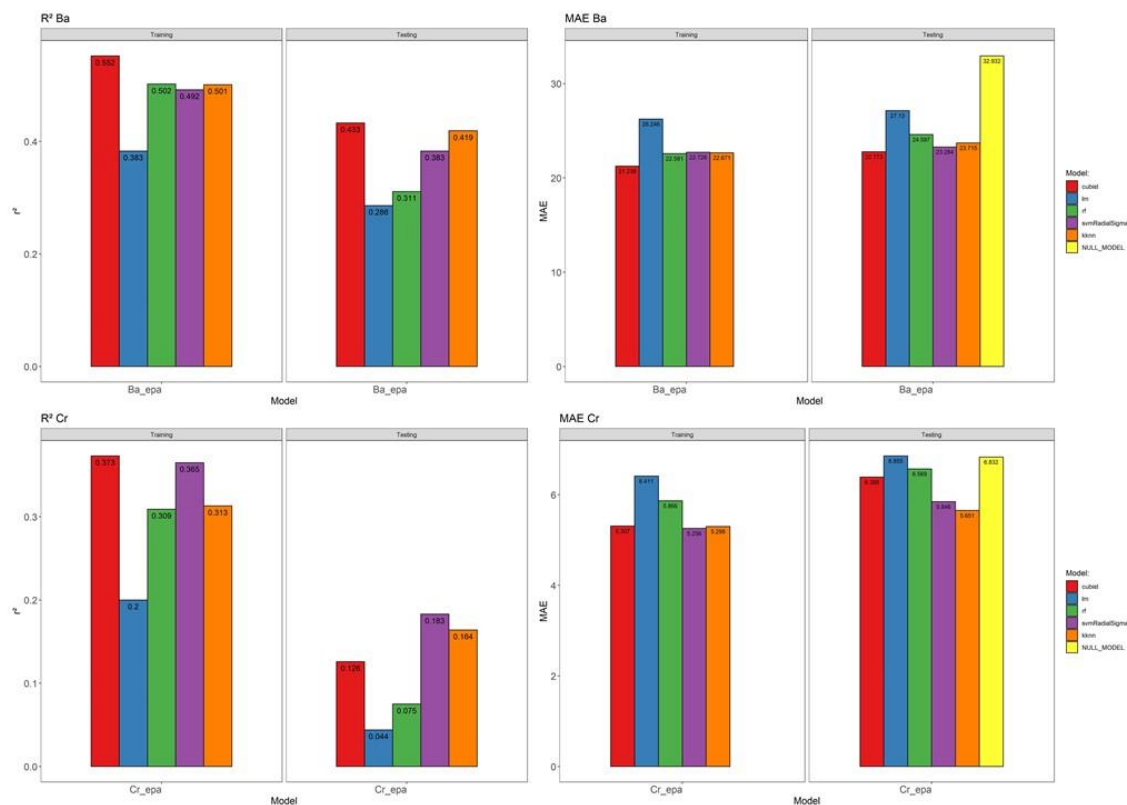
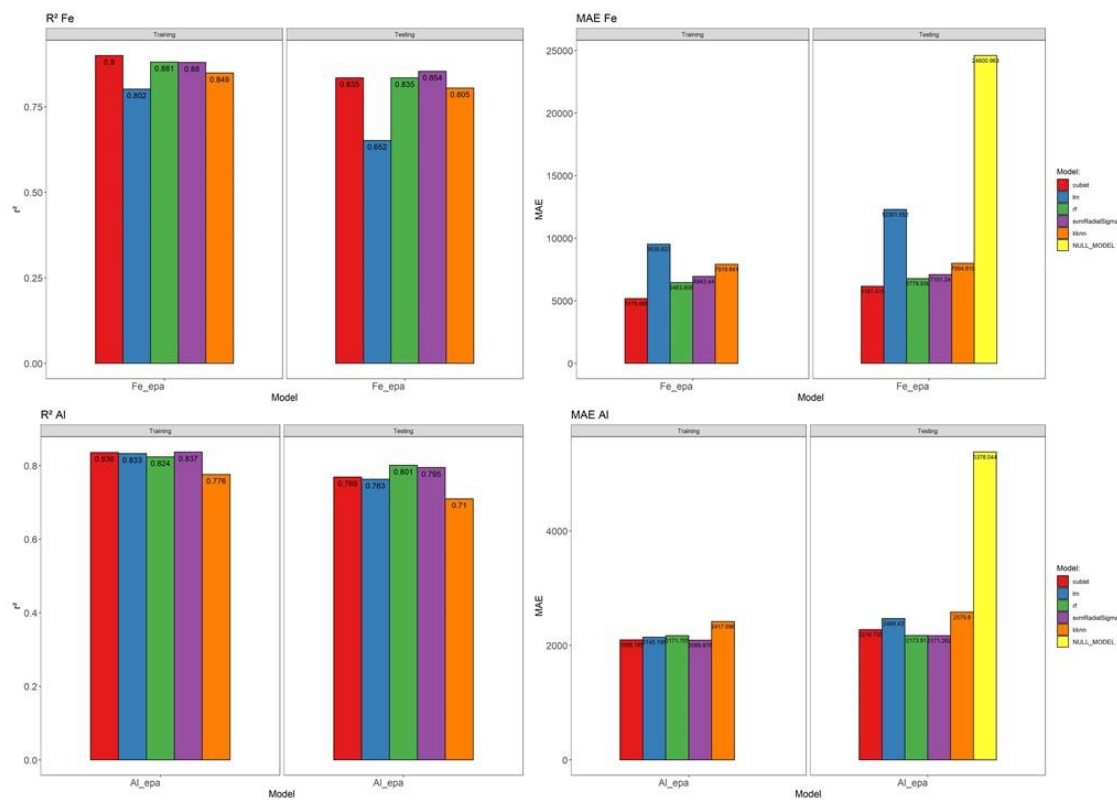


Figure S3 – Performance of Cubist (CUB), Linear (Lm), Random Forests (RF), K Nearest Neighbor (KKNN) and Support Vector Machine (SVM) models for the prediction of contents of main elements (Fe and Al), assessed by mean absolute error (MAE) and R-squared (R^2) of the training and test data sets.



CAPÍTULO 3

Potencial de espécies vegetais adaptadas as condições do semiárido para fitorremediação de solos contaminados

RESUMO

A atividade de mineração é muito importante para o crescimento de um país, porém, gera impactos nos solos e produz uma grande quantidade de rejeitos que se depositados de forma inadequada, podem ser fonte de contaminação. Os elementos potencialmente tóxicos estão entre os contaminantes prejudiciais ao meio ambiente e a saúde humana. Na cidade de Boquira-Ba, a quantidade de rejeito produzido e depositado em pilhas e bacias vem sofrendo processos erosivos, dispersando Pb e Zn nos solos. As técnicas de remediação físicas e químicas utilizadas em áreas contaminadas apresentam custo elevado e promovem danos a estrutura dos solos. Já as biológicas, como a fitorremediação, são de baixo custo e aplicáveis a extensas áreas. No entanto, uma limitação da fitorremediação é encontrar plantas com capacidade de estabilizar e/ou absorver metais e serem adaptadas as condições edafoclimáticas das áreas contaminadas. Assim, o objetivo do trabalho foi avaliar a capacidade de espécies vegetais adaptadas a condições do semiárido a crescerem em solos contaminados com Pb. O experimento foi conduzido em casa de vegetação em um esquema fatorial 4 x 5, representado pela combinação de quatro espécies vegetais (*M. oleífera*, *P. juliflora*, *A. peregrina*, *U. ruziziensis*) e cinco níveis de Pb (0,0; 0,52; 1,05; 2,10 e 4,20 g kg⁻¹), em delineamento em blocos casualizados. Aos 70 dias após a semeadura foram realizadas medições para avaliar a capacidade das plantas de sobreviverem e crescerem em solos contaminados, absorver e translocar o Pb para a parte aérea e sua influência na absorção de nutrientes pelas plantas. Todas as espécies cresceram em todas as doses de Pb, porém apenas a *P. juliflora* e *A. peregrina* não apresentaram reduções significativas na maioria das variáveis de crescimento. A *U. ruziziensis* apesar de apresentar reduções nas variáveis de crescimento, foi a espécie com maior produção de massa de matéria seca tanto da parte aérea e raiz, além de acumular maiores teores de Pb. Todas as espécies apresentaram capacidade de acumular o Pb nas raízes e baixa capacidade de translocar para a parte aérea. O aumento das doses de Pb influenciou nos teores de Ca da parte aérea de todas as espécies estudadas. Diante disso, concluímos que as espécies *P. juliflora* e *A. peregrina* e *U. ruziziensis* são mais indicadas para cultivos em solos com altos teores de Pb.

Palavras chave: Contaminação. espécies arbóreas. metais pesados. remediação

Potential of plant species adapted to semiarid conditions for phytoremediation of contaminated soils

ABSTRACT

The mining activity is very important for the growth of a country, however, it generates impacts on the soil and produces a large number of tailings that, if deposited inappropriately, can be a source of contamination. Heavy metals are among the contaminants harmful to the environment and human health. In the city of Boquira-Ba, the number of tailings produced and deposited in piles and basins has been suffering erosive processes, dispersing Pb and Zn in the soils. The physical and chemical remediation techniques used in contaminated areas are costly and damage the soil structure. The biological ones, such as phytoremediation, are of low cost and applicable to extensive areas. However, a limitation of phytoremediation is finding plants with the capacity to stabilize and/or absorb metals and adapt to the edaphoclimatic conditions of the contaminated areas. Thus, the objective of this work was to evaluate the capacity of plant species adapted to semiarid conditions to grow in soils contaminated with Pb. The experiment was carried out in a protected environment with a 4 x 5 factorial arrangement, represented by the combination of four plant species (*M. oleifera*, *P. juliflora*, *A. peregrina*, *U. ruziziensis*) and five levels of Pb (0.0; 0.52; 1.05; 2.10 and 4.20 g kg⁻¹), in a block design randomized. At 70 days after sowing, measurements were taken to assess the ability of plants to survive and grow in contaminated soil, to absorb and translocate Pb to the shoot and its influence on the uptake of nutrients by plants. All species grew at all levels of Pb, but only *P. juliflora* and *A. peregrina* did not show significant reductions in most growth variables. *U. ruziziensis*, despite showing reductions in growth variables, was the species with the highest production of dry matter mass in both shoots and roots, in addition to accumulating higher levels of Pb. All species showed the ability to accumulate Pb in the roots and low capacity to translocate it to the shoot. The increase in Pb doses influenced the Ca contents of the aerial part of all studied species. Therefore, we conclude that the species *U. ruziziensis*, *P. juliflora* and *A. peregrina* are more suitable for cultivation in soils with high levels of Pb.

Keywords: Contamination. tree species. heavy metals. remediation

1.0- INTRODUÇÃO

Os elementos potencialmente tóxicos (PTEs) são de ocorrência natural na crosta terrestre. Porém, as atividades antropogênicas têm contribuído para aumentar seus níveis no ambiente. Dentre as atividades antropogênicas podemos destacar a mineração, que apesar de ser muito importante para o desenvolvimento industrial do País, é nociva para o meio ambiente, visto que, mesmo após a exploração dos minérios, os rejeitos e escórias de mineração se depositados de forma inadequada podem ser fontes de contaminação (Miralles et al., 2006; Lima et al., 2018; Yan et al., 2018; Jiang et al., 2019; Dhaliwal et al., 2020; Pan et al., 2021).

A atividade de mineração está sujeita ao licenciamento ambiental, por se configurar uma atividade efetiva ou potencialmente poluidoras, conforme o artigo 10 da Lei 6.938/81, texto da Política Nacional do Meio Ambiente (PNMA). Pelo caráter potencialmente lesivo ao meio ambiente, tal atividade implica a responsabilidade do minerador pela recuperação ambiental das áreas degradadas, principalmente após a desativação das minas, conforme previsto no Decreto Nº 9.406/18. O não cumprimento desse dispositivo configura crime ambiental pela Lei de Crimes Ambientais 9.605/98.

Diversas técnicas de remediação têm sido utilizadas para estabilizar e recuperar áreas contaminadas por PTEs oriundos da atividade de mineração, como as técnicas físicas, químicas e biológicas (Kulshreshtha et al., 2014; Dhaliwal et al., 2020). As técnicas físicas e químicas apresentam custos elevados e promovem danos a estrutura do solo (Dhaliwal et al., 2020). Já as técnicas de remediação biológica se dividem em biorremediação, que consiste na remediação dos elementos tóxicos por meio do uso de microrganismos e a fitorremediação, que utiliza espécies vegetais capazes de absorver tais elementos do solo (Mukhopadhyay e Maiti 2010; Dhaliwal et al., 2020). Tanto a fitorremediação quanto a biorremediação podem ser utilizadas como estratégia biotecnológica para extrair, sequestrar ou descontaminar ambientes terrestres ou aquáticos poluídos com PTEs.

A fitorremediação tem sido considerada uma das técnicas de remediação mais vantajosas, quando se trata de PTEs. Essa técnica é considerada promissora na descontaminação, devido a sua simples execução, por ser sustentável e apresentar custo relativamente baixo (Mota et al., 2016; Bhatti et al., 2018). A fitorremediação tem sido estudada e difundida principalmente na Europa e nos Estados Unidos, sendo pouco explorada no Brasil. No estado de São Paulo, segundo o relatório de dezembro de 2013 elaborado pela CETESB, das 2.674 áreas nas quais a remediação foi realizada, apenas em três foi utilizada a fitorremediação.

Para que as técnicas de fitorremediação tenham sucesso, conhecer as características do ambiente contaminado e as espécies de plantas mais adequadas para cada situação são fundamentais. Plantas com alta capacidade de absorver PTEs em suas raízes (Fitoestabilização) e/ou translocá-lo para a parte aérea (Fitoextração), com rápido crescimento e alta produção de biomassa são ideais para programas de fitorremediação (Bhatti et al., 2018; Egendorf et al., 2020). Além dessas características, as plantas têm que se adaptar as condições edafoclimáticas da região onde se encontra a área contaminada. Dessa forma, estudos buscando espécies com características para uso em programas de fitorremediação são necessários, pois apesar de terem muitos estudos demonstrando a capacidade de algumas espécies em estabilizar e extrair metais nos solos, muitos dessas apresentam pouca produção de biomassa, além de serem adaptadas a condições edafoclimáticas específicas (Dhaliwal et al., 2020). O uso de plantas nativas para fitorremediação pode ser uma opção útil, pois essas plantas são mais adaptadas às condições ambientais da região do que plantas introduzidas de outros ambientes (Bech et al., 2012).

No Sudoeste do estado da Bahia a atividade de mineração de chumbo (Pb) e zinco (Zn) produziu uma grande quantidade de rejeito de mineração (894 dm³), os quais estão servindo de fontes de poluentes para os solos (Santos et al., 2020). Por estar disposta de forma inadequada, a pilha de rejeito vem sofrendo com processos erosivos, tanto eólico quanto pluvial, os quais estão dispersando grande quantidade de PTEs no ambiente. Quando presente nos solos esses PTEs podem entrar na cadeia trófica, prejudicando todo o ecossistema e a saúde humana (Yang et al., 2021). Dessa forma, a fitorremediação pode ser uma técnica eficiente para estabilizar os PTEs na pilha de rejeito, evitando, que os mesmos se espalhem no ambiente.

Diante disso, o presente estudo baseia-se na hipótese de que espécies nativas e espécies adaptadas a condições do semiárido, presentes próximas de áreas contaminadas podem ser utilizadas para estabilizar os PTEs nas pilhas de rejeito. Dessa forma, os objetivos do presente estudo foram (i) avaliar a capacidade de espécies vegetais ocorrentes nas áreas próximos dos depósitos de rejeito de mineração de Pb, de crescerem e se desenvolverem em solos com altas concentrações de Pb, (ii) avaliar a capacidade de algumas espécies vegetais adaptadas a condições do semiárido a crescerem e desenvolverem em solos contaminados com Pb, (iii) avaliar a capacidade das plantas de absorver e estabilizar Pb nas raízes (iv) avaliar a capacidade das plantas de translocar PTEs para a parte aérea e (v) verificar a influência do Pb na absorção de nutrientes pelas plantas.

2.0- MATERIAIS E MÉTODOS

2.1- Local do estudo

O experimento foi conduzido nos meses de fevereiro a abril de 2021 em casa de vegetação do Departamento de Solos da Universidade Federal de Viçosa, localizado no município de Viçosa-MG com coordenadas geográficas 20°45' S e 42°50' O e 672 m de altitude (Alencar et al., 2011). Segundo Köppen, o clima é do tipo temperado quente, com verão chuvoso e inverno frio e secos (Cwb), com temperatura média de 20 °C (Lorenzon et al., 2013). Durante a condução do experimento a temperatura apresentou máxima de 38 °C e mínima de 11 °C (Figura 1).

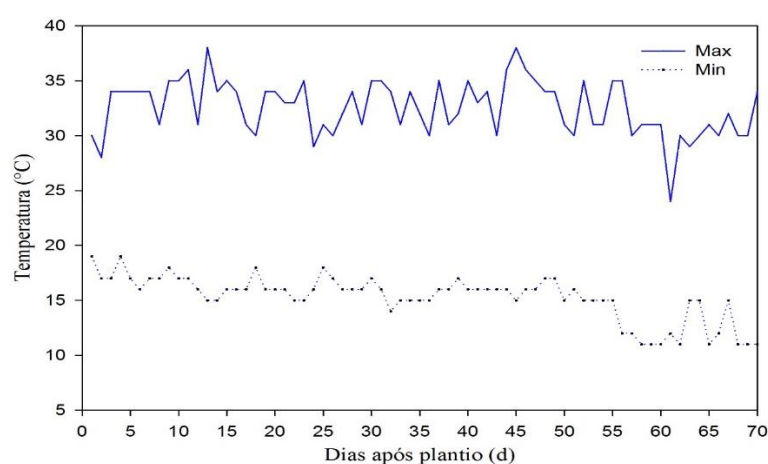


Figura 1. Temperatura mínima e máxima diária durante a condução do experimento em casa de vegetação.

2.2- Delineamento experimental

Os tratamentos foram esquematizados em um fatorial 4 x 5, representado pela combinação de quatro espécies vegetais (*Moringa oleífera*, *Prosopis juliflora*, *Anadenanthera peregrina*, *Urochloa ruziziensis*) e cinco níveis de Pb (0,0; 0,52; 1,05; 2,10 e 4,20 g kg⁻¹), em delineamento em blocos casualizados com quatro repetições.

Os níveis de Pb foram obtidos por meio da mistura de rejeito de mineração com solos (Tabela 1). As amostras de solos utilizados para diluir a concentração de Pb foi de um Latossolo Vermelho distrófico da região de Sete Lagoas-MG (Tabela S1).

Tabela 1: Mistura do rejeito com solo para obtenção dos níveis de chumbo nos solos

Doses de Chumbo					
Solo (%)	100	90	80	60	20
Rejeito (%)	0	10	20	40	80
Concentração de Pb (g kg ⁻¹)	0	0,52	1,05	2,10	4,20

O rejeito foi coletado no Município de Boquira-Ba, onde ocorreu atividade de mineração por mais de 30 anos. As jazidas de minério dessa área tinham como principais componentes metálicos o chumbo e zinco (Tabela 2). Durante o longo período de atividade de mineração, grandes volumes de resíduos foram gerados e depositados em uma pilha de rejeitos próximo do centro da cidade. Essa pilha por estar disposta de forma inadequada vem sofrendo com processos erosivos, principalmente pluvial e eólico dispersando grande quantidade de material contaminado nas áreas urbanas e agrícolas ao entorno da pilha de rejeito. De acordo com Andrade et al., (2017), no entorno da bacia de rejeito tem-se mais de 300 propriedades rurais.

Tabela 2: Caracterização química e física do rejeito utilizado no experimento

	Ba	Cd	Cr	Ni	Pb	Zn	Al	Fe	Ca	K	P	Ti	Silte	Argila	Areia
	 mg kg ⁻¹			 g kg ⁻¹							%				
Média	22,2	23,9	2,47	3,0	5,3	2,41	3,5	172,7	15,9	0,0	0,3	0,1	17,0	3,8	79,1
SD	1,6	4,1	1,13	1,1	0,5	0,16	0,2	2,5	1,3	0,0	0,0	0,0	4,5	3,2	3,5
CV %	7,4	17,2	45,9	37,7	9,2	6,5	5,1	1,5	8,3	17,6	4,6	0,5	26,7	83,0	4,5

* Caracterização química realizada pela metodologia EPA 3051A e caracterização química pelo método da pipeta

2.3- Espécies utilizadas

As espécies foram selecionadas devido a capacidade de se adaptar as condições edafoclimáticas do semiárido brasileiro e/ou predomínio das mesmas no ambiente próximo aos depósitos de rejeitos de mineração de Pb. As espécies selecionadas foram: *M. oleífera*, *P. juliflora*, *A. peregrina*, *U. ruziziensis*. As sementes de *M. oleífera*, *P. juliflora* e *A. peregrina*, foram coletadas em diversas plantas matrizes localizadas no município de Guanambi situado no sudoeste da Bahia, Brasil. Já as sementes de *U. ruziziensis* foram adquiridas no comércio de Viçosa-MG.

A *M. oleífera* L. é uma leguminosa com adaptação em diversas condições edafoclimáticas, sendo bastante utilizada em diversos países devido ao seu uso industrial, medicinal e alimentício, e principalmente ao seu alto potencial produtivo em solos com baixa fertilidade natural (Ferreira et al., 2008; Kasolo et al., 2010; Mayo et al., 2011; Fernandes et al., 2015; Gopalakrishnan et al., 2016; Souza et al., 2018; Christophe et al., 2019). A *M. oleífera* L. também é muito utilizada como adubação verde, reflorestamento e para recuperação de áreas degradadas (Mridha, 2015). Além disso, alguns estudos têm demonstrado que plantas de *M. oleífera* L. apresentam grande potencial para fitorremediar áreas contaminadas com PTEs (George et al., 2016; Ogundiran et al., 2018).

U. ruziziensis é uma gramínea conhecida popularmente como capim Brachiaria, muito utilizada em pastagens, adaptada às condições edafoclimáticas das regiões semiáridas, apresenta sistema radicular fasciculado bem desenvolvido, muito importante para a estruturação e estabilização de solos. Diversos estudos tem demonstrado que as gramíneas apresentam potencial para estabilizar solos em áreas de mineração (Maiti e Maiti, 2015; Kumar et al 2017; Zanchi et al., 2021)

P. juliflora DC. (Mimosoideae) popularmente conhecida como algaroba, é uma espécie leguminosa de ocorrência natural no México, América Central e Norte da América do Sul, (Ribask et al., 2009). É uma espécie que apresenta crescimento rápido com alta resistência à seca, e, portanto, se adapta às condições do semiárido brasileiro. Além disso, é adaptada a diversas condições edáficas (Jayaram e Prasad, 2009). Diversos estudos também têm demonstrado o potencial dessa planta na recuperação de áreas impactadas pela atividade de mineração (Gabriel e Patten, 1994; Senthilkumar et al., 2005).

A. peregrina (L.) também é uma leguminosa arbórea, conhecida popularmente como angico, encontrada em praticamente todas as regiões do Brasil, adaptada a diversas condições edafoclimáticas, são tolerantes a solos encharcados ou secos, rasos e compactados e com baixa fertilidade natural. É uma espécie muito utilizada para recuperação de áreas degradadas (Araújo et al., 2006). Estudos têm demonstrado que a capacidade de associação da *A. peregrina* (L.) com microrganismos é uma estratégia eficiente para adaptação das mesmas em solos contaminados com elementos potencialmente tóxicos, tendo potencial para programas de fitorremediação (Gomes et al. 2013; Gomes et al., 2020).

2.4- Montagem e condução do experimento

O solo e o rejeito foram secos ao ar, posteriormente peneirado em malha de 4 mm e homogeneizado manualmente com o auxílio de uma enxada de acordo com as proporções de cada tratamento. Em seguida as parcelas experimentais representadas por vasos de 2,5 kg foram pesadas preenchidas manualmente com os materiais previamente misturados.

Com uma antecedência de cerca de 30 d do início do experimento, os solos utilizados na mistura com rejeito foram incubados com calcário dolomítico comercial. A dose de calcário aplicada foi calculada com base na análise química do solo, utilizando o método de correção da acidez proposto por Teixeira (2017). Antes do plantio foi realizada adubação em todo volume da parcela de acordo a recomendação de Novais et al. (1991) para experimentos em casa de vegetação.

Foram semeadas 10 sementes em cada unidade experimental e após a germinação foi realizado o desbaste, mantendo apenas quatro plantas para as espécies arbóreas e três plantas para a gramínea. As plantas foram cultivadas por 70 d, mantendo-se os tratamentos com umidade próxima de 90 % da capacidade de campo por meio de irrigações diárias.

2.5- Variáveis analisadas

Decorrido 70 d após a semeadura foram avaliadas as seguintes variáveis: altura de plantas (ALP), utilizando régua graduada, com os resultados expressos em centímetros; diâmetro do caule (DC), com o auxílio de um paquímetro digital, com resultados expressos em mm; massa de matéria seca da parte aérea ($mMSPA$), massa de matéria seca das raízes ($mMSR$) e massa de matéria seca total ($mMST$), que foram acondicionadas em sacos de papel e secas individualmente em estufa de circulação forçada de ar, a temperatura de 60 °C, por 72 horas; O Índice de Qualidade de Dickson (IQD) (Dickson et al., 1960) foi determinado pela equação:

$$IQD = mMST \text{ (g)} / \left[\left(\frac{ALP(cm)}{DC \text{ (mm)}} \right) + \left(\frac{mMSR(g)}{mMSPA \text{ (g)}} \right) \right] \quad \text{eq. 1}$$

Para avaliação das trocas gasosas foram mensuradas as seguintes variáveis com auxílio de um analisador de gás infravermelho (IRGA): condutância estomática (g_s , $\text{mmol H}_2\text{O m}^{-2} \text{s}^{-1}$), fotossíntese líquida (A , $\mu\text{mol CO}_2 \text{ m}^{-2} \text{s}^{-1}$), transpiração (E , $\text{mmol vapor d'água m}^{-2} \text{s}^{-1}$), concentração interna de carbono (C_i , $\mu\text{mol CO}_2 \text{ mol}^{-1} \text{ ar}$) e eficiência de carboxilação (A/C_i).

Para a coleta dos padrões de fluorescência foi utilizado o instrumento MultispeQ v2.0, por meio do protocolo Photosynthesis Rides, com auxílio do software PhotosynQ. Os parâmetros mensurados foram: radiação fotossinteticamente ativa (PAR), Teor de clorofila (SPAD), Eficiência do fotossistema II (PhiII), coeficiente de extinção não-fotoquímica (PhiNPQ) (Dissipação de energia na forma de calor), energia da luz perdida para processos não regulados que podem danificar a fotoquímica (PhiNO), parâmetros que descrevem o acúmulo de prótons no tilacóide e seu fluxo através da ATP sintase, que converte ADP em ATP (ECSt, vH+, gH+). As avaliações desses parâmetros foram realizados nas folhas mais jovens totalmente expandida. Os parâmetros pertinentes a fluorescência foram obtidos sob radiação fotossinteticamente ativa (PAR) natural ($\sim 850\text{-}1100 \mu\text{mol m}^{-2} \text{s}^{-1}$ de fótons), temperatura e vapor de água atmosférico.

As avaliações morfométricas do sistema radicular foram realizadas utilizando-se o sistema informatizado WinRhizo (WinRhizo Pro, Regent Inc. Instr., Canada), em que foi

mensurado as variáveis, volume, comprimento, diâmetro médio e área superficial, número de garfos e cruzamentos de raízes, além do comprimento de raízes por classes de diâmetro: raízes muito finas ($\varnothing$ inferior a 0,5 mm), raízes finas ($> 0,5 \varnothing < 2,0$ mm) e raízes grossas ($\varnothing > 2,0$ mm) (Magalhães et al., 2011).

Também foram determinados os teores de Ca, Mg, K, S, P, Fe, Cu, Mn, B, Zn e Pb na parte aérea e raízes das plantas. Para isso, a massa de matéria seca da parte aérea e da raiz foram moídas em moinho tipo Willey com peneira de 20 *mesh*. Em seguida foram pesadas 0,5000 g de amostras e digeridas por meio da digestão nitroperclórica. A dosagem dos elementos foi realizada por espectrometria de emissão óptica com plasma indutivamente acoplado (ICP-OES, PerkinElmer Optima 8300). Após a determinação dos teores dos nutrientes, foi calculado os conteúdos de Pb e Zn na parte aérea raiz e total.

Para avaliar a capacidade das plantas de acumular PTES do solo e de translocá-los das raízes para a parte aérea, o fator de acumulação (FA) e fator de translocação (FT) foram calculados de acordo as equações a seguir.

$$FA = \frac{tPb \text{ total na planta}}{tPb \text{ no solo}} \quad \text{eq. 2}$$

$$FT = \frac{tPb \text{ na parte aérea}}{tPb \text{ na raiz}} \quad \text{eq. 3}$$

2.6- Análise estatística

As análises estatísticas foram realizadas por meio do Programa Sisvar 5.7 (Ferreira, 2011). Os dados foram testados quanto à normalidade pelo teste de Shapiro-Wilk ($p > 0,05$) e homogeneidade das variâncias pelo teste de Bartlett ($p > 0,05$). Todos os dados foram submetidos à análise de variância (ANOVA) e, quando detectadas diferenças significativas, por meio do teste F, os mesmos foram submetidos à análise de regressão. Para a seleção das equações, foram considerados a significância dos parâmetros, a significância biológica dos modelos e o coeficiente de determinação (R^2). Além disso utilizou-se a análise multivariada visando entender o comportamento dos padrões de fluorescência da clorofila das plantas avaliadas em função das doses crescentes de Pb nos tratamentos. Para isso, aplicou-se a Análise de Componentes Principais (PCA) com o objetivo de reduzir o número de variáveis sem perda de informação (Sena et al., 2002; Savegnago et al., 2011). As análises multivariadas foram realizadas utilizando o software R versão 4.1 (R CORE TEAM, 2021) com o pacote FactoMineR (Lê et al., 2008).

3.0- RESULTADOS

Todas as espécies estudadas cresceram durante o período de teste com os rejeitos da mina de Pb e Zn de Boquirá-Ba. Para a massa de matéria seca da parte aérea, raiz e total houve interação significativa entre os níveis de Pb e as espécies estudadas, demonstrando que cada espécie apresenta comportamento diferenciado mediante os níveis de Pb. Já para ALP não foi observado significância, indicando que os fatores agem de forma independente sobre tal variável (Figura 2).

Para as plantas de *U. ruziziensis* o melhor ajuste do modelo foi o quadrático para a massa de matéria seca da parte aérea, raiz e total, em que, nas menores proporções do rejeito, e consequentemente menores doses de Pb, ocorreu um estímulo no crescimento das plantas, porém com o aumento das proporções houve uma redução. Para a *P. juliflora* verificou-se comportamento semelhante a *U. ruziziensis* para massa de matéria seca da parte aérea e total, porém para a massa de material seca da raiz não houve influência das doses de Pb. O aumento das doses de Pb, proporcionado pela adição de rejeito, provocou redução na massa de matéria seca da parte aérea, raiz e total em plantas de *M. oleífera*. Por outro lado, o aumento das doses de Pb, proporcionado pela adição dos rejeitos de mineração, não influenciaram na massa de matéria seca da parte aérea, raiz e total das plantas de *A. peregrina*. Apesar de ocorrer reduções nas variáveis de crescimento, as plantas de *U. ruziziensis* foram as que mais produziram massa tanto de matéria seca da parte aérea quanto de raízes.

Para altura de plantas houve um incremento médio de 3,73 cm, independente da espécie estudada até a dose de $1,7 \text{ g kg}^{-1}$ de Pb, porém a partir dessa dose houve redução na variável (Figura S1, S2, S3 e S4).

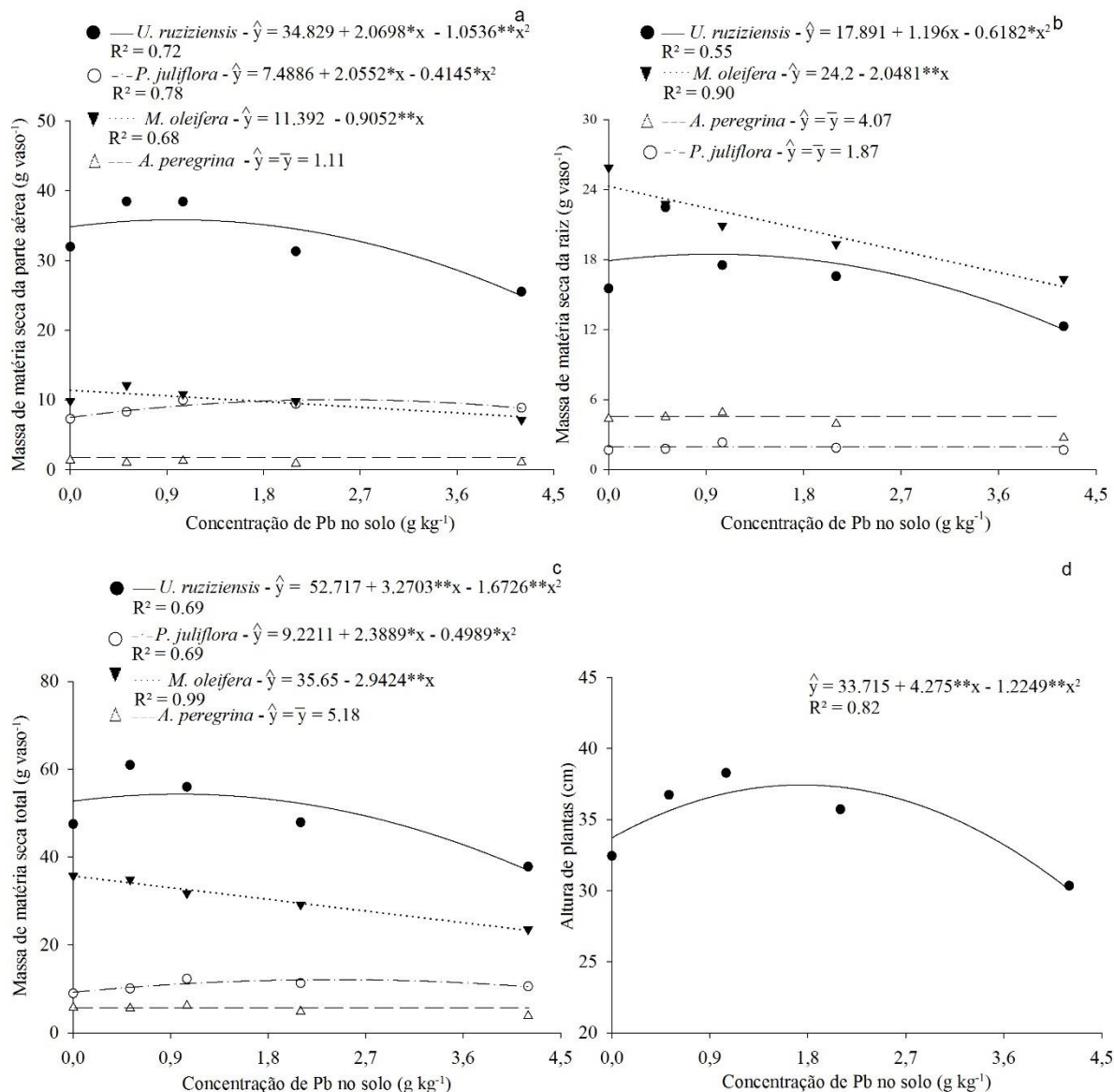


Figura 2. Massa de matéria seca da parte aérea (a), massa de matéria seca da raiz (b), massa de matéria seca total e altura de plantas de *M. oleífera*, *P. juliflora*, *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

As variáveis diâmetro do caule (DC) e o Índice de qualidade de Dickson (IQD) foram determinados apenas para as espécies arbóreas. Observou-se que as doses de Pb não influenciaram no DC das espécies de *A. peregrina* e *P. juliflora*, enquanto que, para a *M. oleífera* houve comportamento quadrático, com ponto de inflexão em 1,67 g kg⁻¹ (Figura 3a). O aumento das doses de Pb não influenciaram no IQD nas plantas de *A. peregrina* e *P. juliflora*, enquanto que, para as plantas de *M. oleífera* ocorreu uma redução linear com decréscimo de 48 % (Figura 3b).

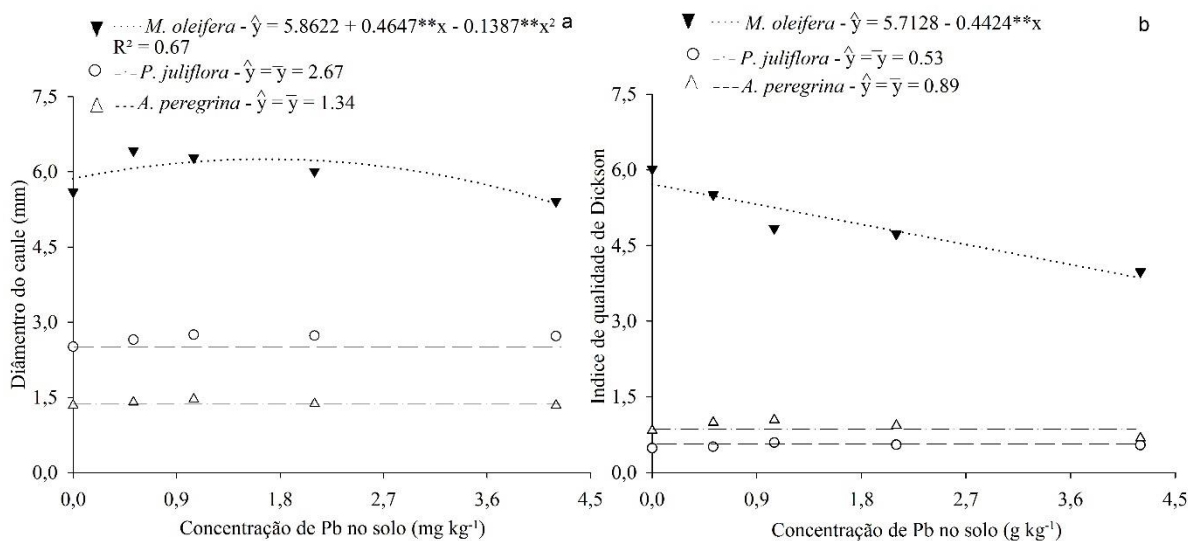


Figura 3. Diâmetro do caule (a) e índice de qualidade de Dickson (b), de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* em função das doses de Pb.

Assim como ocorreu na massa de matéria seca da raiz, as doses crescentes de Pb, reduziram linearmente o comprimento e área superficial das raízes de *M. oleífera* e da *U. ruziziensis*, enquanto que, para *P. juliflora* e *A. peregrina* não houve influência significativa a 5 % de probabilidade (Figura S5, S6, S7 e S8). Para o volume de raiz o comportamento foi semelhante ao comprimento e área superficial das raízes para a maioria das espécies, com exceção apenas da *U. ruziziensis*, em que ocorreu um ajuste quadrático, com aumento no volume de raiz até a dose de 1,65 g kg de Pb. As doses crescentes de Pb nos solos também promoveu um aumento linear no diâmetro médio das raízes de *M. oleífera*. Nas espécies *P. juliflora* e *A. peregrina* o modelo que melhor se ajustou aos dados foi o quadrático decrescente com ponto de inflexão na dose de 1,65 e 2,16 g kg⁻¹ de Pb, respectivamente. E na *U. ruziziensis* o modelo que melhor se ajustou aos dados foi o quadrático crescente com maior valor de diâmetro médio na dose de 2,34 g kg⁻¹ de Pb nos solos (Figura 4).

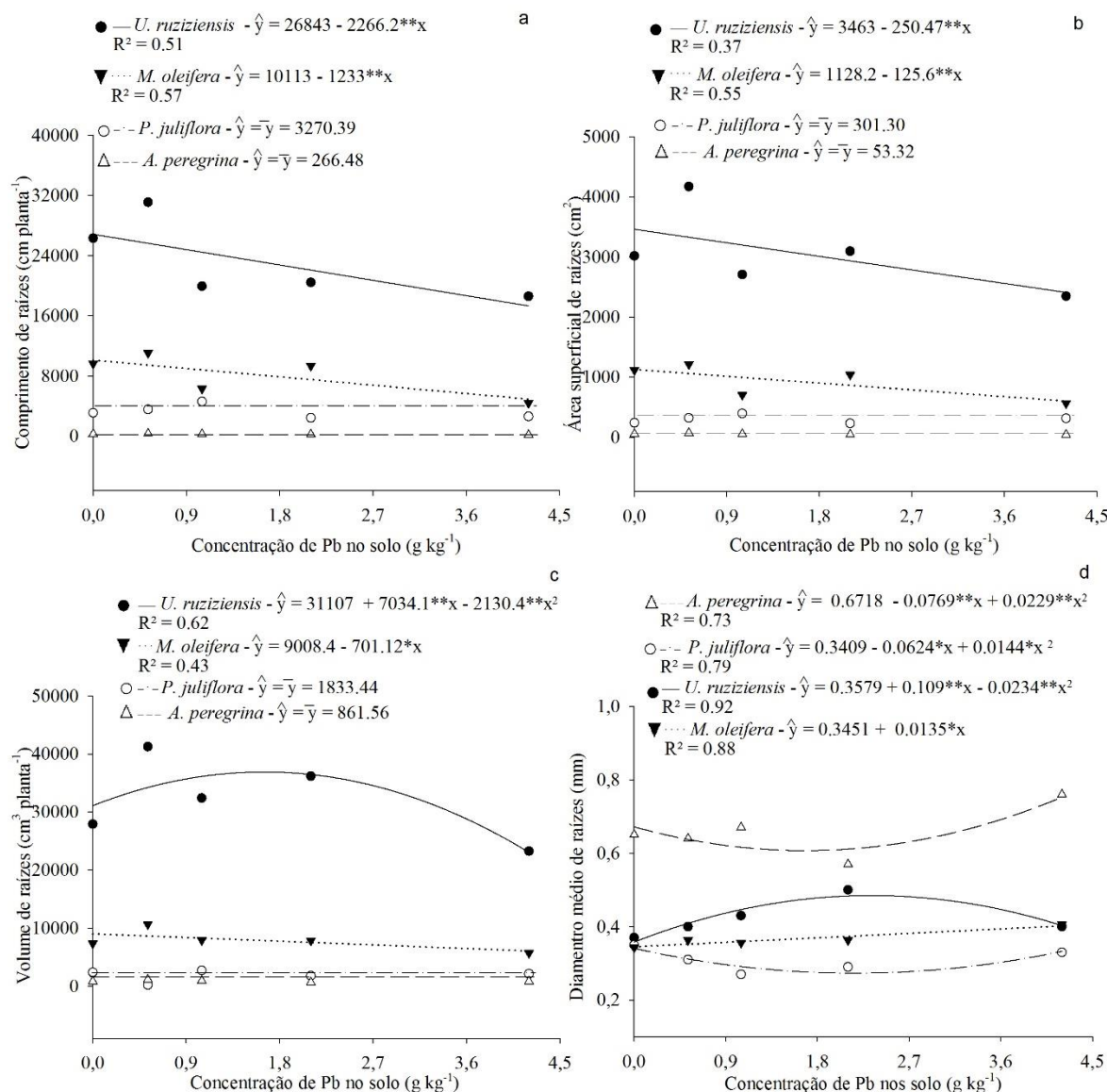


Figura 4. Comprimento de raiz (a), área superficial da raiz (b), volume de raiz (c) e diâmetro médio de raiz (d) de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

O aumento das doses de Pb nos solos influenciou negativamente no comprimento de raízes muito finas nas plantas de *M. oleífera* e *U. ruziziensis*, enquanto que para as espécies *P. juliflora* e *A. peregrina*, não houve influência significativa a 5 % de probabilidade (Figura 5a). Já nas raízes finas e raízes grossas, houve influência significativa apenas para a *U. ruziziensis* com ajuste quadrático (Figura 5b e 5c). O número de cruzamentos e garfos não foram influenciados pelas doses crescentes de Pb nos solos para as plantas de *P. juliflora* e *A. peregrina*, enquanto que na *M. oleífera* ocorreu uma redução linear (Figura 5d e 5e). O número de cruzamentos apresentou um comportamento quadrático decrescente, enquanto que o número de garfos aumentou até a dose de 1 g kg⁻¹ de Pb com posterior decréscimo nas plantas de *U.*

ruziziensis (Figura 5e). Para todas as variáveis morfométricas do sistema radicular ocorreram reduções com o aumento da dose de Pb nas plantas de *U. ruziziensis*, mesmo assim, apresentou valores para essas variáveis superiores as demais espécies em todas as doses de Pb.

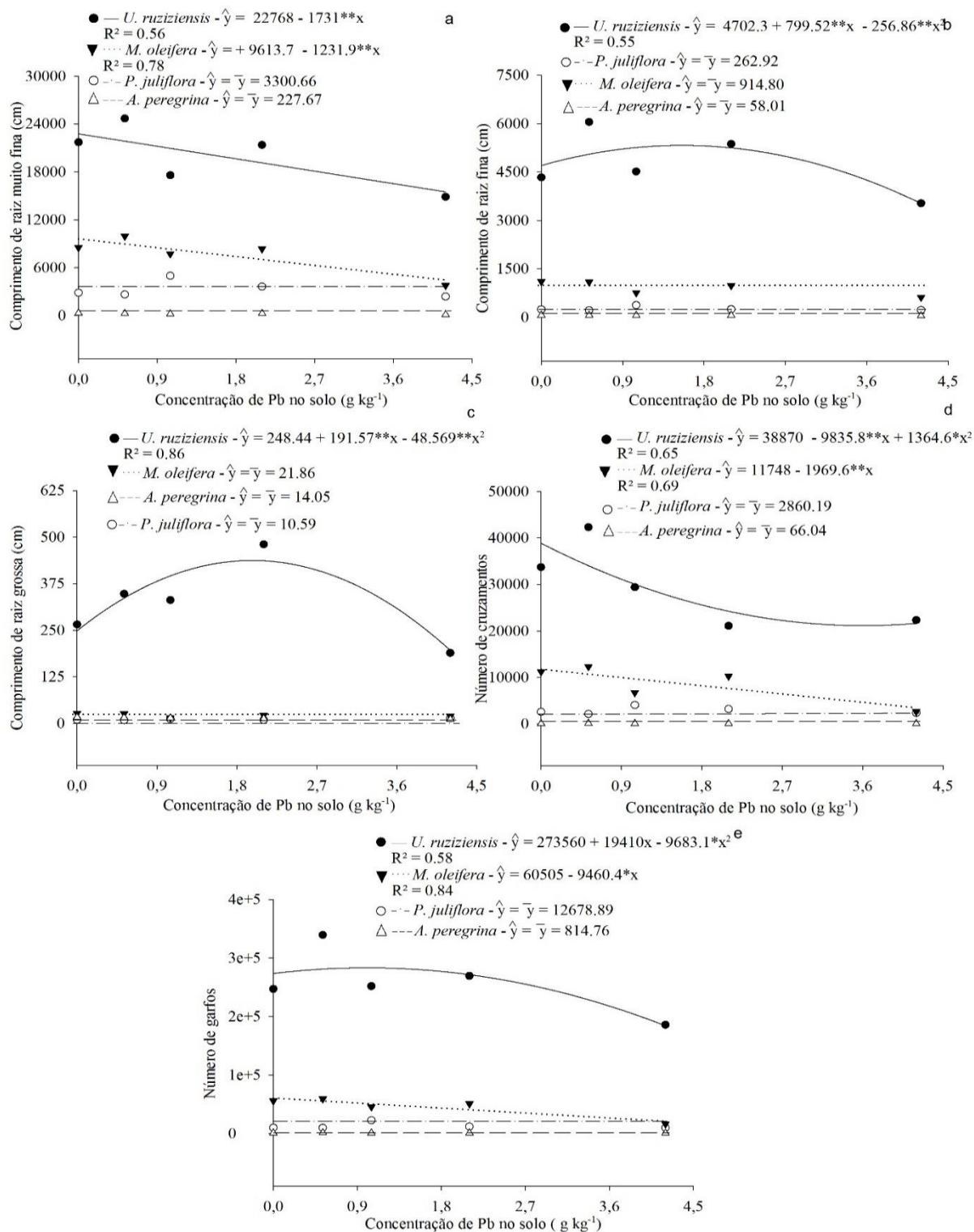


Figura 5. Comprimento de raiz muito fina (a), Comprimento de raiz fina (b), Comprimento de raiz grossa (c), número de cruzamentos (d) e número de garfos (d) de *M. oleifera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

As variáveis condutância estomática (g_s), fotossíntese líquida (A) e eficiência de carboxilação (A/C_i) apresentaram comportamento diferenciado em cada espécie estudada. A concentração interna de CO_2 (C_i) se manteve constante em todas as espécies mesmo com o aumento das doses de Pb. Já para transpiração (E), foi observado um decréscimo exponencial com aumento das doses de Pb, independentemente das espécies estudadas. O aumento das doses de Pb não influenciaram na condutância estomática das plantas de *U. ruziziensis* e *A. peregrina*, enquanto que para as plantas de *P. juliflora* e *M. oleífera* houve uma redução exponencial nessa variável. Também foi observado redução na fotossíntese líquida em todas as espécies estudadas, com o aumento das doses de Pb, com redução mais acentuada para as plantas de *M. oleífera* e *P. juliflora*. Para a *U. ruziziensis* e *M. oleífera* houve redução linear na eficiência de carboxilação com o aumento das doses de Pb. Já na *P. juliflora* também ocorreu redução, porém de forma exponencial com o aumento das doses de Pb. Para o *A. peregrina* as doses de Pb não influenciaram na eficiência de carboxilação (Figura 6).

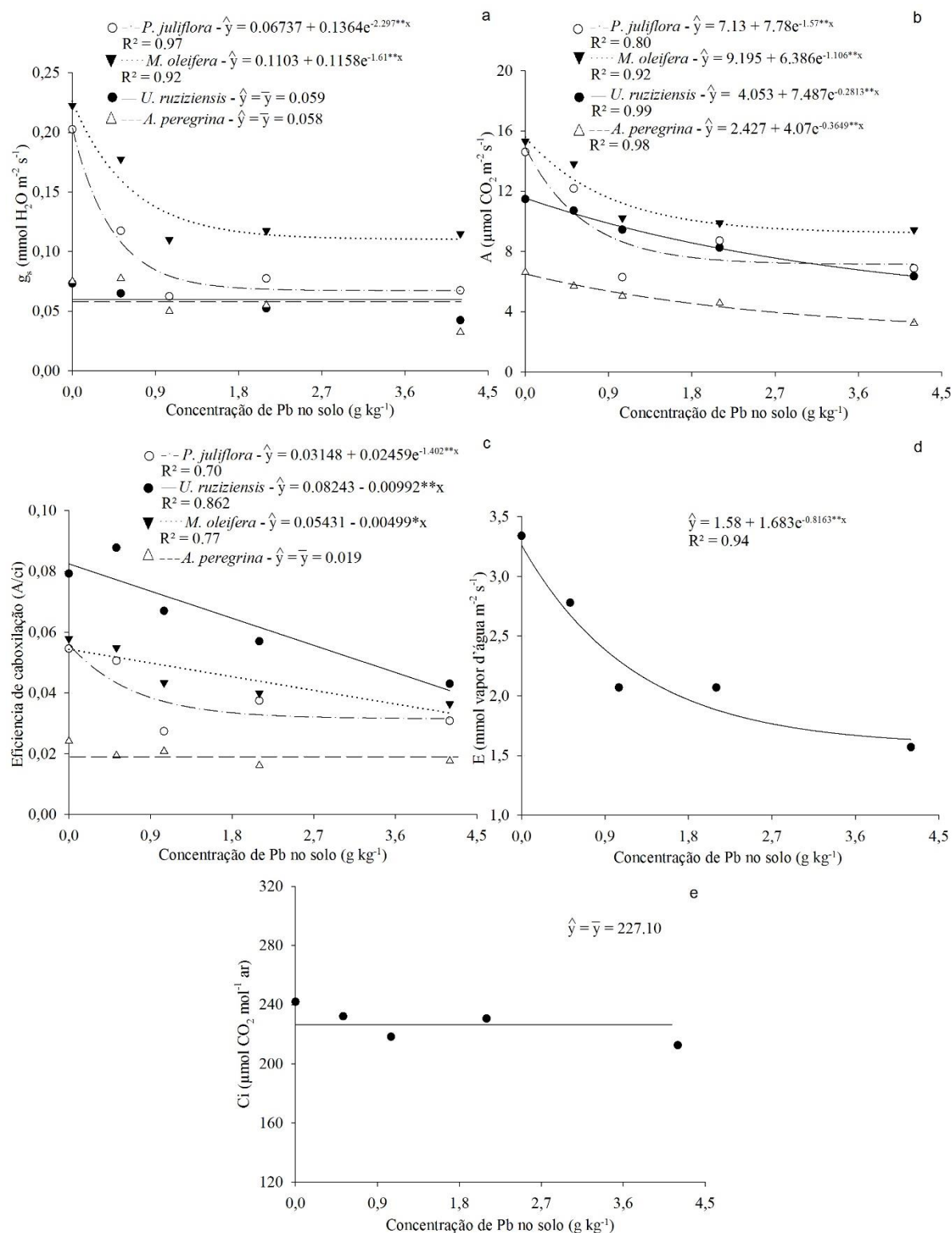


Figura 6. Condutância estomática (a), fotossíntese líquida (b), eficiência de carboxilação (c), transpiração (d) e concentração interna de CO_2 (e) de plantas de *M. oleifera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

Os teores de Pb na parte aérea aumentaram linearmente nas plantas de *P. juliflora*, *U. ruziziensis* e *A. peregrina* com o aumento das doses de Pb nos solos, com teores máximos de

32, 27 e 11 mg kg⁻¹, respectivamente. Já nas plantas de *M. oleífera*, o comportamento dos teores de Pb na parte aérea mediante as doses crescentes de Pb no solo, apresentaram um ajuste quadrático, com maiores teores (17,37 mg kg⁻¹) na dose de 2,94 g kg⁻¹ de Pb (Figura 7a). Para os teores de Pb nas raízes, verificou-se um comportamento linear crescente em função das doses de Pb nos solos, para todas as espécies estudadas, sendo a *U. ruziziensis* a espécie com maior capacidade de acumular Pb nas raízes com teores próximos a 1400 mg kg⁻¹ na maior dose de Pb nos solos (Figura 7b).

Em relação ao conteúdo de Pb na raiz verificou-se um ajuste quadrático para as espécies *U. ruziziensis* e *M. oleífera*, com acúmulos de 17,62 e 6,48 g kg⁻¹ de Pb nas doses de 4,2 e 2,94 g kg⁻¹ de Pb, respectivamente. Para as espécies *A. peregrina* e *P. juliflora*, o aumento das doses de Pb nos solos não influenciaram o conteúdo de Pb nas raízes (Figura 7c). O conteúdo de Pb na parte aérea da *U. ruziziensis*, *M. oleífera* e *A. peregrina* apresentou comportamento semelhante ao conteúdo de Pb nas raízes. Já para a *P. juliflora* o conteúdo de Pb nas raízes apresentou comportamento linear crescente (Figura 7d). O conteúdo total de Pb nas plantas de *U. ruziziensis* e *M. oleífera* apresentaram comportamento semelhante ao conteúdo de Pb nas raízes e parte aérea, enquanto que para a demais espécies o comportamento foi linear crescente (Figura 7e).

Os teores de Zn na parte aérea também aumentaram linearmente com as doses crescentes de Pb nos solos, visto que, os níveis de Pb foram obtidos por meio da diluição de rejeito de mineração com solos, sendo que o rejeito também contém elevados teores de Zn (Figura 8a). Para os teores de Zn nas raízes, as espécies *U. ruziziensis* e *A. peregrina* apresentaram incremento linear com o aumento das doses de rejeito. Já para a *M. oleífera* e *P. juliflora* o comportamento foi quadrático, com maiores teores nas doses de 3,68 e 3,06 g kg⁻¹ de Pb, respectivamente (Figura 8b).

O conteúdo de Zn na parte aérea, raiz e total foram lineares com o aumento das doses de Pb nos solos nas espécies de *A. peregrina* e *P. juliflora* (Figura 8c, 8d, 8e), já para a *U. ruziziensis*, o acúmulo de Zn apresentou comportamento quadrático, com máximo acúmulo na dose 4,2 g kg⁻¹ de Pb (Figura 8c, 8d, 8e). Na *M. oleífera*, o conteúdo de Zn na parte aérea foi linear, porém na raiz e o conteúdo total, o comportamento foi quadrático, com maior conteúdo na dose de 2,94 g kg⁻¹ de Pb (Figura 8c, 8d, 8e).

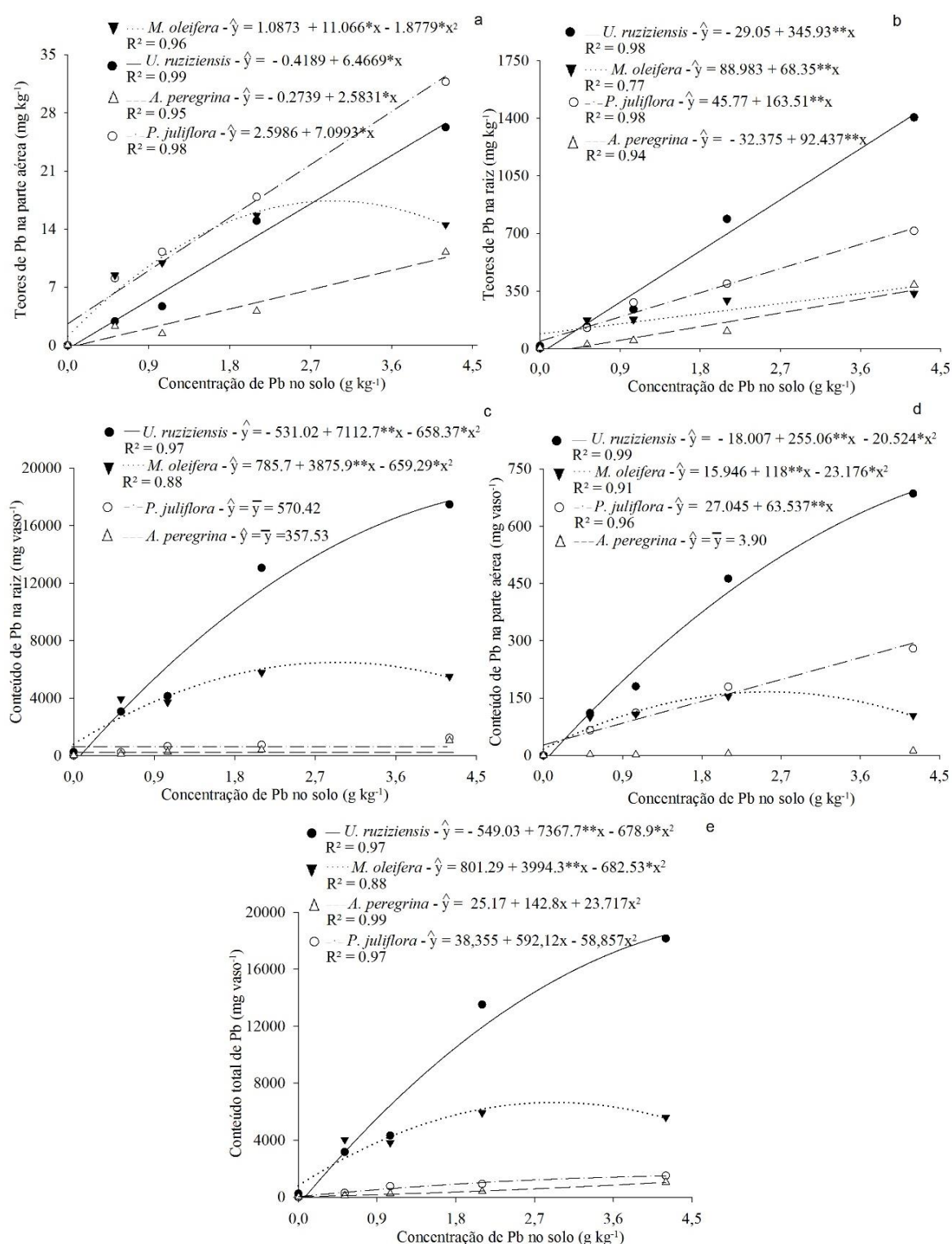


Figura 7. Teores de Pb na parte aérea (a), teores de Pb na raiz (b), conteúdo de Pb na parte aérea (c), raiz (d) e total de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

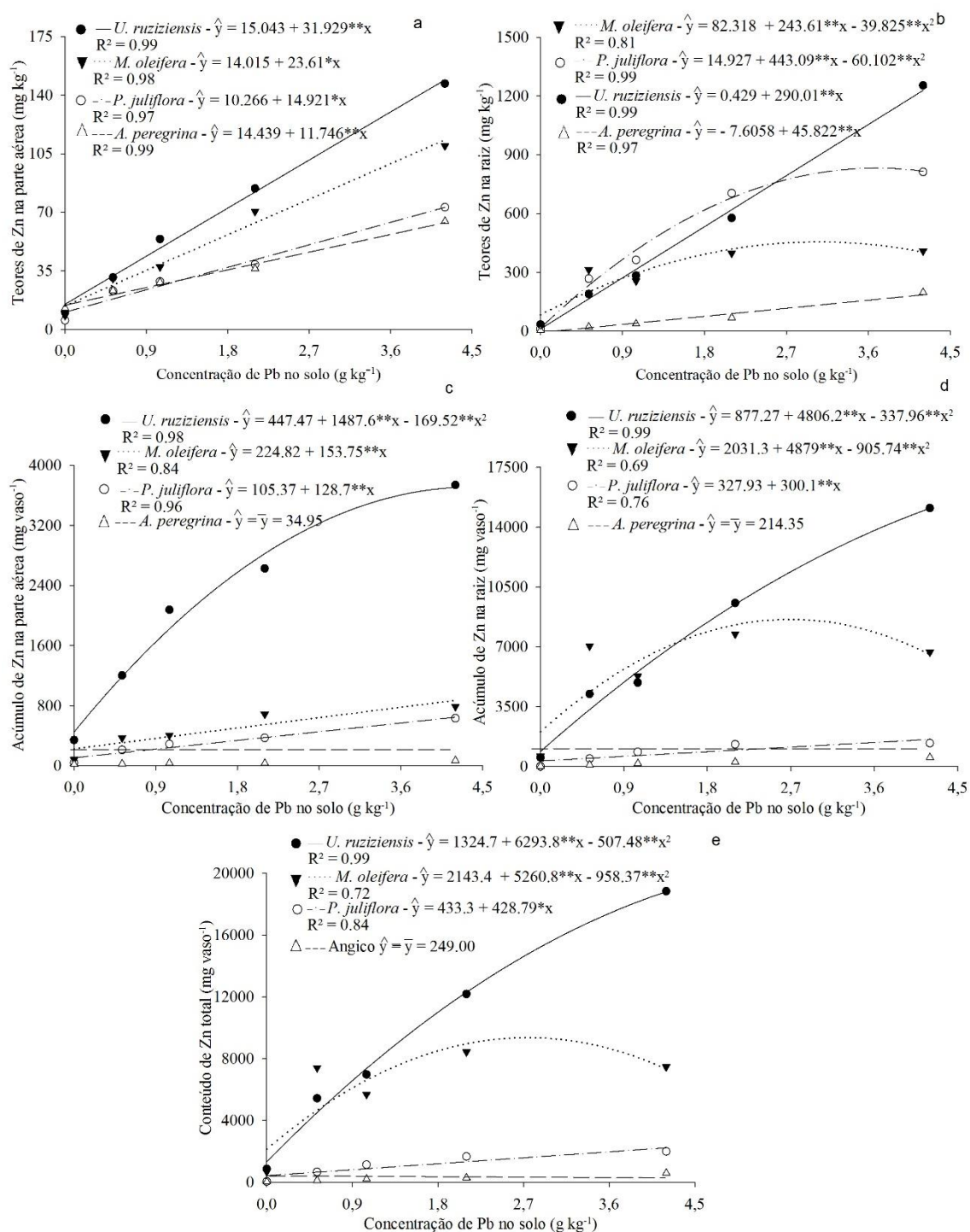


Figura 8. Teores de Zn na parte aérea (a), teores de Pb na raiz (b), conteúdo de Zn na parte aérea (c), raiz (d) e total de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

Os coeficientes de transferência calculados, fator de acumulação (FA) e translocação (FT) foram inferiores a 1 em todos os teores de Pb nos solos. Além disso, em todas as espécies estudadas os maiores fatores de translocação foram observados na dose de 0,52 g kg⁻¹ de Pb

nos solos, sendo a espécie *P. juliflora* a que apresentou maior FT seguida da *A. peregrina* (Figura 9b). Para o FA também se verifica que na dose de 0,52 g kg de Pb que ocorreu maiores valores para a maioria das espécies estudadas (Figura 9a).

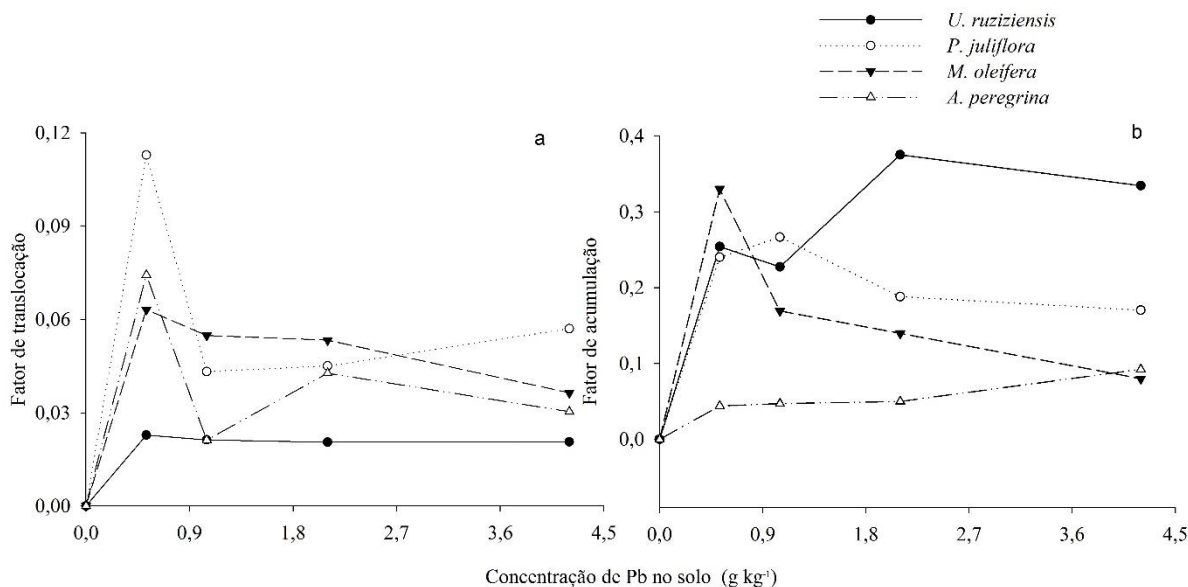


Figura 9. Fator de translocação (a) e fator de acumulação (b) em plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

A absorção de Pb pelas plantas, promovida pela adição de doses crescentes de rejeito contaminado, influenciou na absorção de macro e micronutrientes pelas plantas, principalmente nos elementos catiônicos. Para a *U. ruziziensis*, observou-se que as doses crescentes de Pb nos solos reduziu os teores de ferro (Fe) na parte aérea, comportamento inverso ocorreu para as plantas de *A. peregrina*. Já nas plantas de *M. oleífera* e *P. juliflora*, as doses crescentes de Pb não influenciaram nos teores de Fe da parte aérea (Figura 10a).

Os teores foliares de boro (B) na parte aérea cresceu linearmente com o aumento das doses de Pb, nas plantas de *A. peregrina*. Por outro lado, para as plantas de *M. oleífera* e *P. juliflora* observou-se comportamento quadrática, porém decrescendo com o aumento da dose, e em seguida, um novo aumento. Já nas plantas de *U. ruziziensis* o aumento das doses de Pb não influenciaram nos teores de B na parte aérea (Figura 10b).

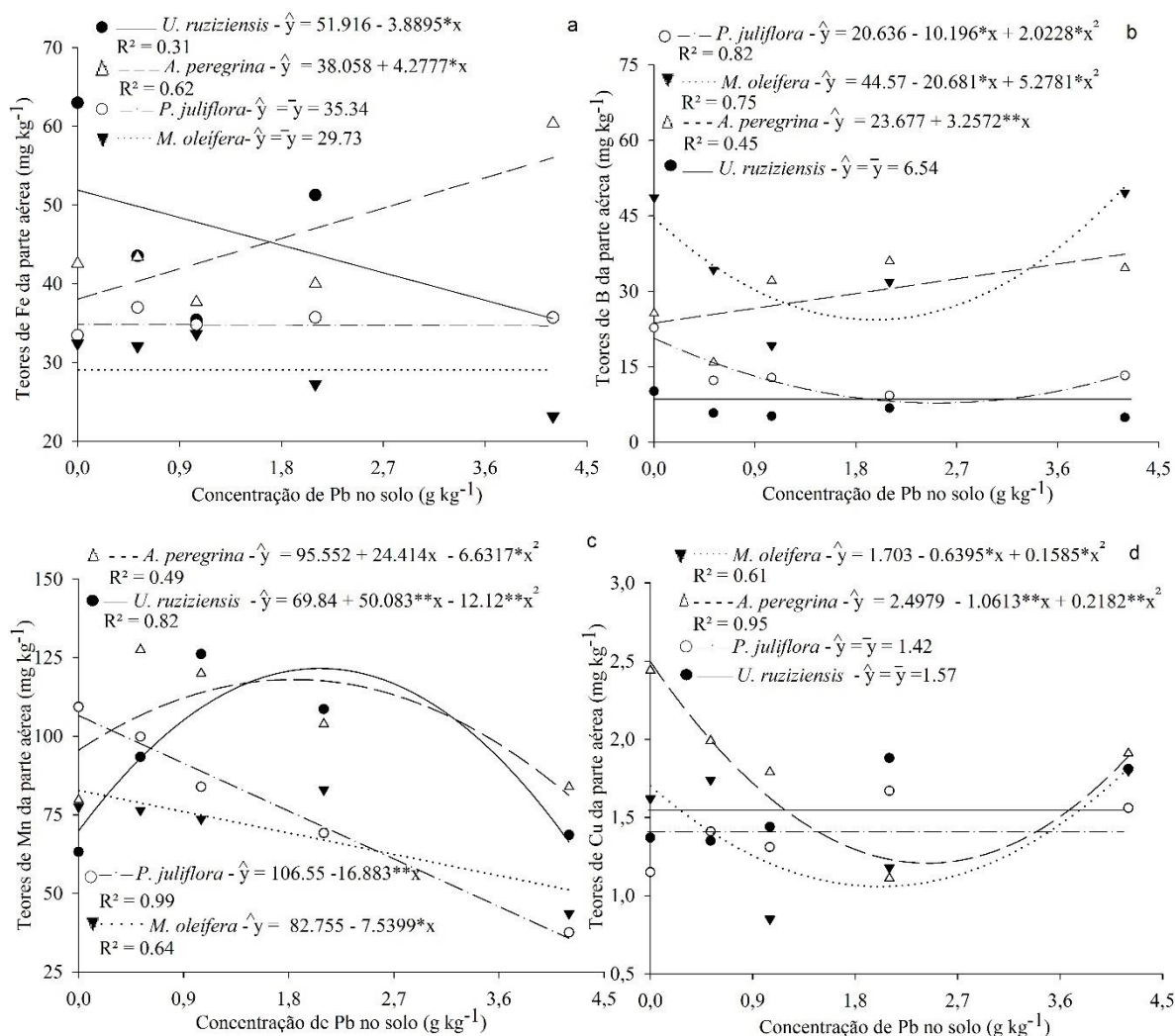


Figura 10. Teores de Fe (a), B (b), Mn (c) e Cu (d) na parte aérea de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

Para os teores de manganês (Mn) na parte aérea, observou-se comportamento quadrático para as plantas de *U. ruziziensis* e *A. peregrina* com maiores teores nas doses de 1,21 e 1,84 g kg⁻¹ de Pb, a partir dessa dose houve novamente um decréscimo. Nas plantas de *P. juliflora* e *M. oleífera* ocorreu um decréscimo linear nos teores de Mn da parte aérea com o aumento das doses de Pb (Figura 10c).

Para os teores de cobre (Cu) na parte aérea verificou-se que nas plantas de *M. oleífera* e *A. peregrina* o comportamento foi quadrático, porém decrescendo com o aumento da dose, e em seguida, um novo aumento. Já nas plantas de *U. ruziziensis* e *P. juliflora* o aumento das doses de Pb, não influenciaram nos teores de Cu na parte aérea (Figura 10d).

As respostas das espécies estudadas mediante aplicação de doses crescentes de Pb nos solos para os teores de magnésio (Mg) na parte aérea foram diferentes. Na *M. oleífera* e *U.*

ruzizensis, o comportamento dos teores de Mg da parte aérea se ajustou ao modelo quadrático, com máximo teores nas doses de 1,81 e 1,57 g kg⁻¹ de Pb. Já nas plantas de *P. juliflora*, ocorreu uma redução linear no teor de Mg na parte aérea com o aumento das doses de Pb, enquanto que nas plantas de *A. peregrina*, o aumento das doses de Pb não influenciou os teores de Mg da parte aérea (Figura 11a).

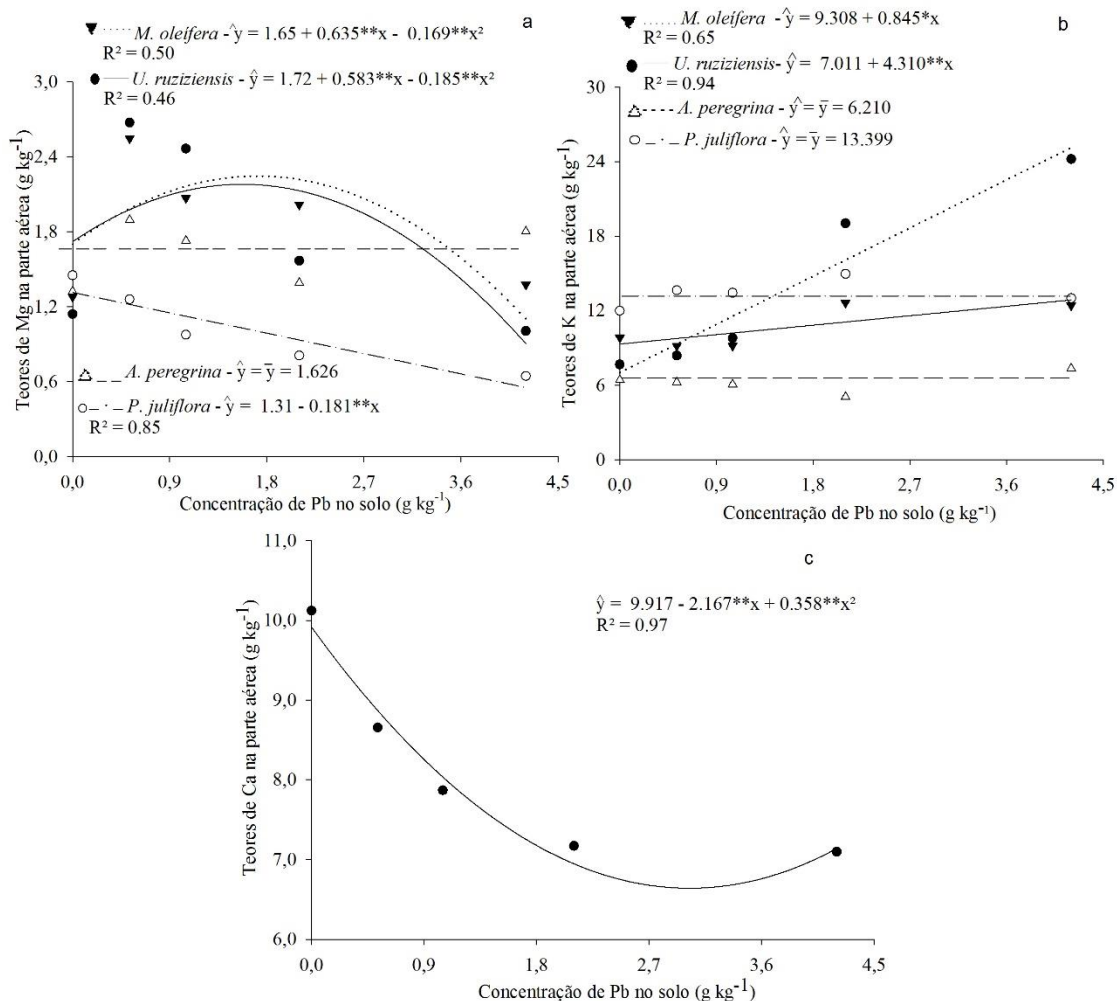


Figura 11. Teores de Mg (a), Ca (b) e K (c) na parte aérea de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruzizensis* em função das doses de Pb.

Para os teores de potássio (K) da parte aérea verificou-se que nas plantas de *P. juliflora* e *A. peregrina* as doses crescentes de Pb não influenciaram, enquanto que, para as plantas de *M. oleífera* e *U. ruzizensis* ocorreu um aumento linear (Figura 11b). Para os teores de cálcio (Ca), o comportamento de todas as espécies estudadas mediante aumento dos teores de Pb nos solos foram similares, observando um decréscimo nos teores desse nutriente até a dose de 3,1 g kg⁻¹ de Pb, a partir dessa dose houve uma tendência de estabilização (Figura 11c).

Com relação ao fósforo (P) e enxofre (S), observou-se que as doses crescentes de Pb nos solos, não influenciaram nos teores desses elementos na parte aérea em todas as espécies estudadas, se mantendo com média de 1,15 e 2,02 g kg⁻¹, respectivamente (Figura 12a e 12b).

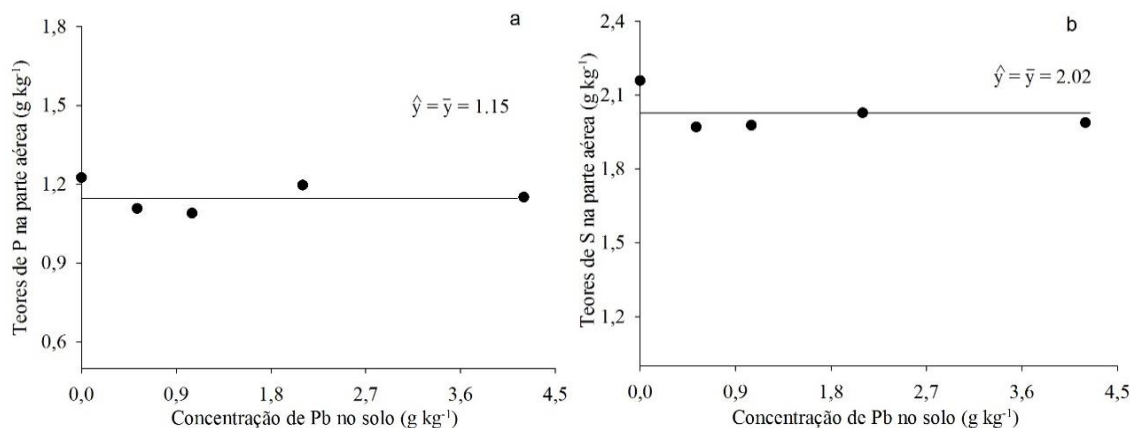


Figura 12. Teores de P (a) e S (b) na parte aérea de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

Assim como na parte aérea das plantas, as doses crescentes de Pb nos solos influenciaram nos teores de nutrientes nas raízes das espécies estudadas. Observou-se que para a *U. ruziziensis*, o aumento das doses crescentes de Pb, promoveu um aumento linear nos teores de Fe, diferentemente das plantas de *M. oleífera* que ocorreu um decréscimo linear. Para as plantas de *P. juliflora* o comportamento foi quadrático, com maiores teores na dose de 1,75 g kg⁻¹ de Pb, com decréscimo acentuado a partir dessa dose (Figura 13a). Para os teores de Mn, verificou-se comportamento quadrático nas plantas de *U. ruziziensis* e *P. juliflora*, com máximo teores nas doses de 3,57 e 2,17 g kg⁻¹ de Pb, respectivamente. Já nas plantas de *M. oleífera* o comportamento dos teores de Mn foi semelhante ao Fe (Figura 13b). Em relação aos teores de Cu nas raízes, observou-se crescimento linear para as plantas de *U. ruziziensis*, enquanto que nas plantas de *M. oleífera* ocorreu um decréscimo linear. Já nas plantas de *P. juliflora* o comportamento dos teores de Cu, foram similares ao de Fe e Mn (Figura 13d). Para os teores de B, apenas as plantas de *P. juliflora* apresentaram teores crescente em função das doses de Pb (Figura 13c). As doses crescentes de Pb nos solos não influenciaram nos teores de Fe, Mn, Cu e B das plantas de *A. peregrina* (Figura 13).

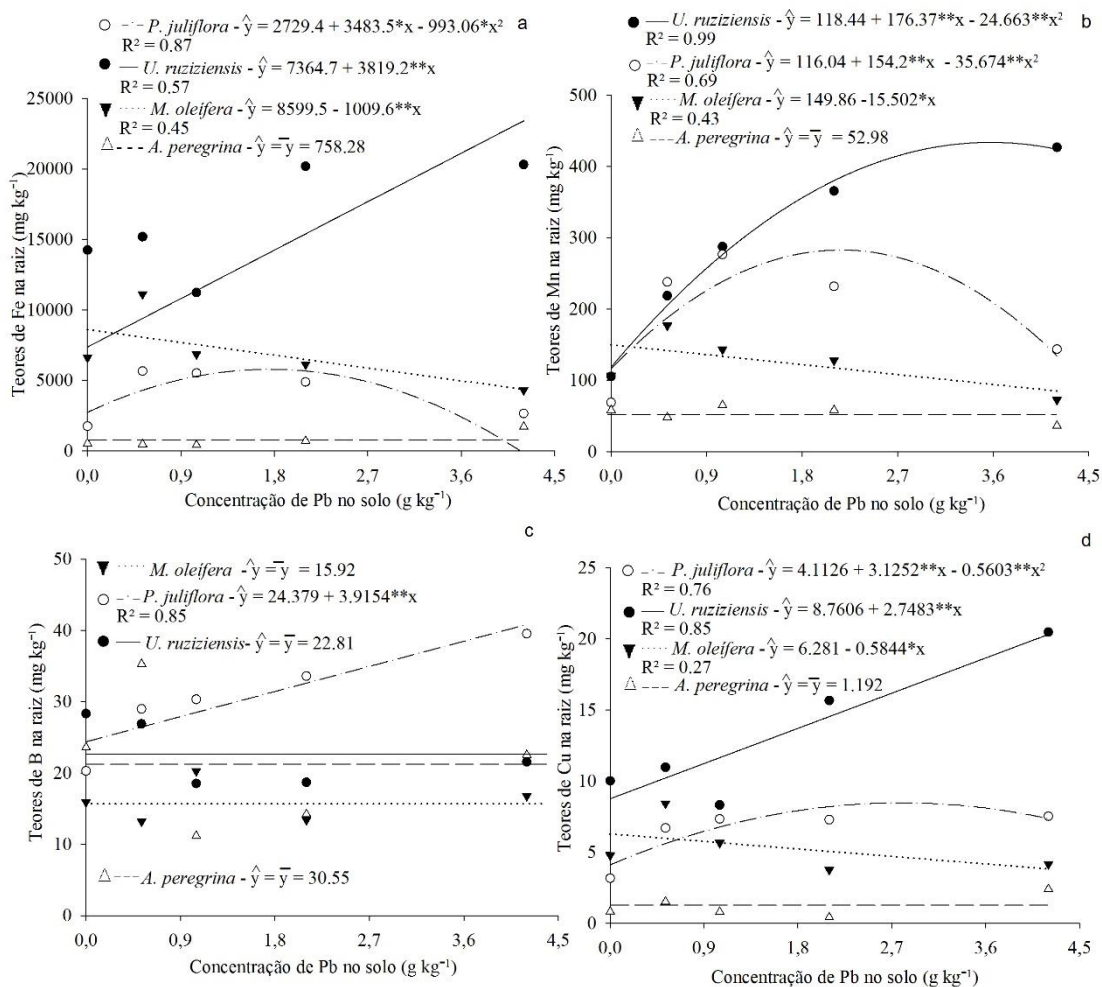


Figura 13. Teores de Fe (a), B (b), Mn (c) e Cu (d) nas raízes de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

Para os teores de Mg nas raízes de *P. juliflora* observou-se um comportamento quadrático crescente, até a dose de 3,15 g kg⁻¹ de Pb. Já nas plantas de *U. ruziziensis* e *M. oleífera* o comportamento foi o inverso (Figura 14a). Para as plantas de *A. peregrina*, as doses crescentes de Pb nos solos, não influenciaram nos teores de Mg e K nas raízes (Figura 14a, 14b). Por outro lado, nas plantas de *P. juliflora* e *U. ruziziensis* os teores de K foram crescentes em função das doses de Pb (Figura 13b). Já nas plantas de *M. oleífera* houve um ajuste quadrático, com máximo teores na dose de 3,3 g kg⁻¹ de Pb (Figura 14b). Para os teores de Ca o comportamento para todas espécies estudadas foi semelhante e crescentes em função das doses de Pb aplicada (Figura 14c).

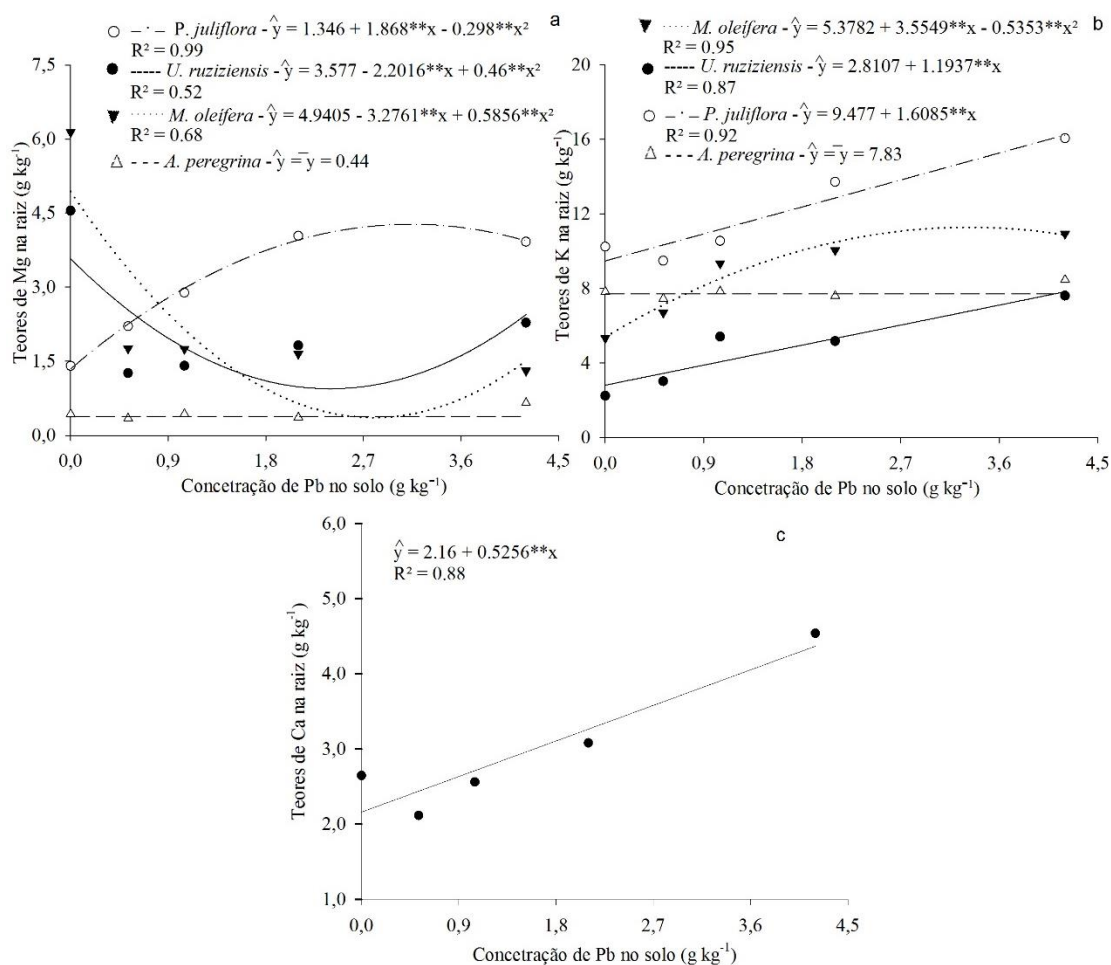


Figura 14. Teores de Mg (a), K (b) e Fe (c) nas raízes de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

Para os teores de P e S, observou-se um comportamento quadrático crescente para as plantas de *P. juliflora*, com máximo teores nas doses de 2,6 e 2,9 g kg⁻¹ de Pb. Já para as plantas de *M. oleífera* e *U. ruziziensis* os teores de S foram crescentes em função do aumento das doses de Pb, diferentemente do que ocorreu para o P, em que nas plantas de *M. oleífera* houve um decréscimo linear, e nas plantas de *U. ruziziensis* as doses de Pb não influenciaram os teores desse elemento. Nas plantas de *A. peregrina* as doses crescentes de Pb no solo, não influenciaram nos teores de P da raiz, porém, para os teores de S o comportamento foi quadrático decrescente, com ponto de inflexão na dose de 1,31 g kg⁻¹ de Pb (Figura 15a e 15b).

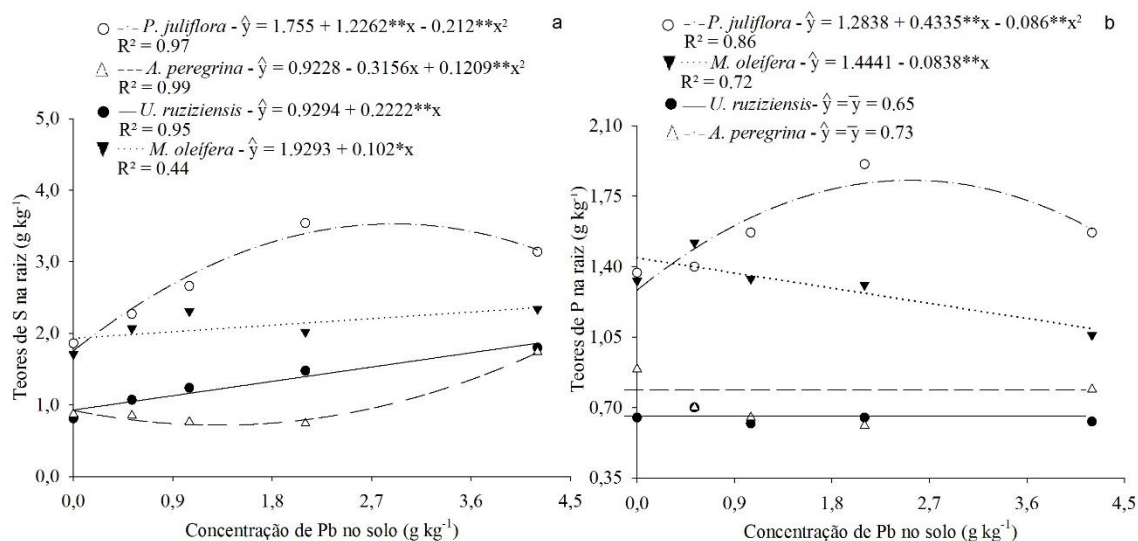


Figura 15. Teores de S (a) e Fe (b) nas raízes de plantas de *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses de Pb.

Para avaliar o comportamento dos padrões de fluorescência das espécies estudadas mediante as doses crescentes de Pb foi realizada uma análise de componentes principais, que mostrou o grau de correlação dos parâmetros avaliados (Figura 16). Os dois primeiros componentes principais explicaram mais de 65 % da variação total dos dados. As variáveis que mais contribuíram para a separação dos grupos foram o coeficiente de extinção não-fotoquímica, eficiência do fotossistema II, energia da luz perdida para processos não regulados que podem danificar a fotoquímica e intensidade luminosa. Observa-se pela análise de componentes a formação de dois grupos distintos, de forma que a separação ocorreu principalmente em função das doses de Pb, pois, verifica-se pontos de intercessão entre todos as espécies no centro da componente. Para as plantas de *U. ruziziensis* alguns tratamentos se dispersaram formando um grupo distinto em função das variáveis radiação fotossintética ativa e intensidade luminosa. A separação ocorreu principalmente nas doses intermediárias de Pb (0,52 a 2,1 g kg⁻¹). Mesmo comportamento ocorreu nas plantas de *M. oleífera*, em que a variável teor de clorofila contribui para separar os tratamentos com dose intermediária de Pb. Para as plantas de *A. peregrina* e *P. juliflora*, verifica-se que as doses crescentes de Pb não influenciaram nos padrões de fluorescência.

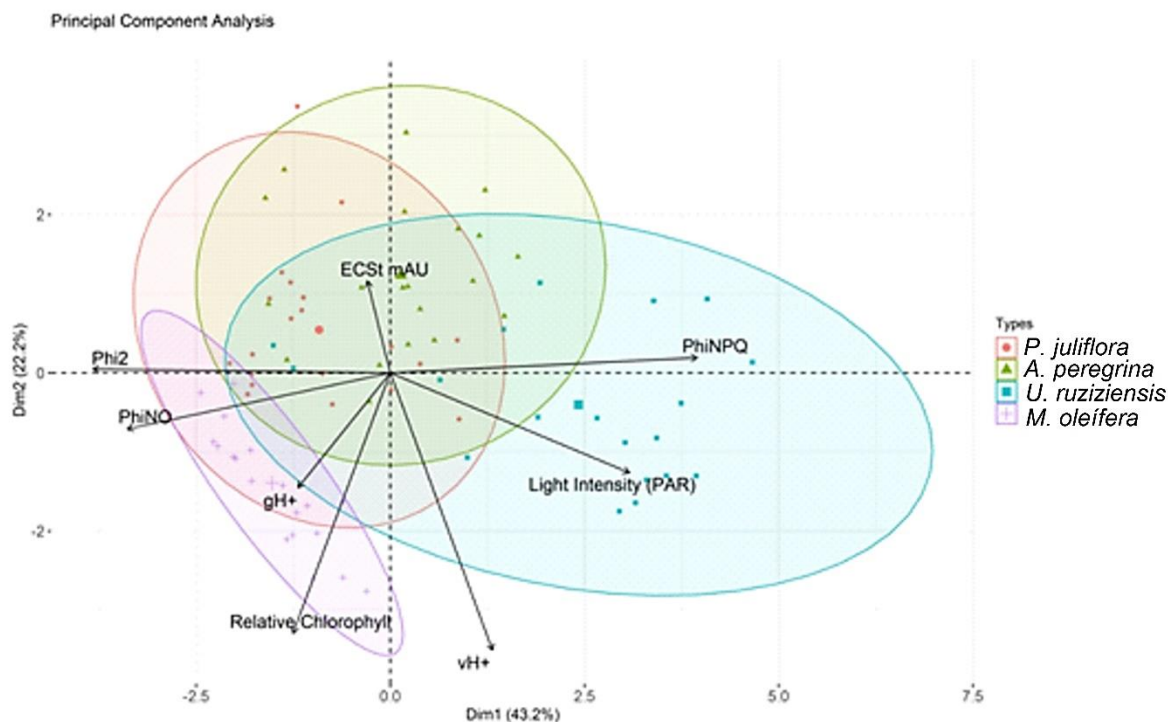


Figura 16. Análise de componentes principais das espécies *M. oleífera*, *P. juliflora* e *A. peregrina* e *U. ruziziensis* em função das doses crescentes de Pb nos solos. Radiação fotossinteticamente ativa (PAR), teor de clorofila relativa (SPAD), eficiência do fotossistema II (PhiII), coeficiente de extinção não-fotoquímica (PhiNPQ), energia da luz perdida para processos não regulados que podem danificar a fotoquímica (PhiNO) e parâmetros que descrevem o acúmulo de prótons no tilacóide e seu fluxo através da ATP sintase, que converte ADP em ATP (ECSt, vH+, gH+).

4.0- DISCUSSÃO

A produção de massa de matéria seca vegetal é característica intrínseca de cada espécie vegetal, dessa forma não é uma variável comparável. No entanto, investigar a produção de biomassa pode ser útil para demonstrar os efeitos da toxicidade do Pb no crescimento de cada espécie (Amorim et al., 2021). Todas as espécies estudadas apresentaram capacidade de crescerem e se desenvolverem em solos contaminados com doses crescentes de Pb. Porém, às espécies *M. oleífera* e *U. ruziziensis* tiveram o crescimento prejudicado quando comparadas com as espécies *P. juliflora* e *A. peregrina*. Em termos relativos calculados entre a dose 0,0 e 4,2 g kg⁻¹ de Pb, a *M. oleífera* e *U. ruziziensis* reduziram 53 e 42 % na massa de matéria seca total, enquanto para a *A. peregrina* não houve redução significativa. Já as plantas de *P. juliflora* ocorreu um incremento de 11 % na massa de matéria seca total, demonstrado ter potencial para estabilizar solos altamente impactados por atividades de mineração. Apesar da redução das variáveis de crescimento das plantas de *U. ruziziensis*, as mesmas foram a que produziu maior

volume de biomassa, o que é uma característica desejável em plantas fitorremediadoras (Bhatti et al., 2018; Engendorf et al., 2020).

O comportamento das plantas de *U. ruziziensis* demonstram que nas menores doses de Pb ocorreram estímulos no crescimento, verificado pelo comportamento quadrático dos gráficos de todas as variáveis da parte aérea e do volume de raízes. O maior crescimento das plantas em condições de estresse químico pode estar relacionado a estratégias de sobrevivência das plantas em ambientes contaminados, visto que, em concentrações de PTEs abaixo do nível considerado tóxico pode ocorrer um estímulo no crescimento das plantas como mecanismo de defesa (Baligar et al., 1998). Além disso, alguns estudos têm observado o fenômeno hormese em muitos organismos em respostas a diferentes fatores de estresse como os provocados por PTEs em plantas (Morkunas et al., 2018). O fenômeno hormese está relacionado aos efeitos hormonais dos PTEs quando presentes em pequenas concentrações e efeitos tóxicos em concentrações elevadas nas plantas (Calabrese e Baldwin, 1998). Efeito benéfico do Pb (1 mg L⁻¹) também foi verificado em plantas de *Euruca sativa* L., cultivadas em solução nutritiva apresentando incremento de 60 % na produção de raízes e 43 % nos brotos, quando comparadas com o tratamento sem aplicação de Pb (Cannata et al., 2013).

A *M. oleífera* conseguiu crescer em solos contaminados com Pb, porém ocorreu uma redução linear nos parâmetros de crescimento com o aumento das doses de Pb, demonstrando ser a espécie mais sensível em relação as demais espécies estudadas. Apesar de ser sensível essa planta conseguiu acumular maiores teores de Pb quando comparadas com as demais leguminosas, tendo, portanto, potencial para ser uma espécie fitorremediadora. Ogundiran et al. (2018) ao comparar os valores dos parâmetros de crescimento de *M. oleífera* cultivadas em solo controle, em relação as cultivadas em solos contaminados com Pb, verificaram que a contaminação pelo Pb inibiu a altura da planta, circunferência do caule e número de folhas. Porém, verificaram que as plantas de *M. oleífera* toleraram concentração de até 8300 mg kg⁻¹ de Pb nos solos.

As plantas de *P. juliflora* e *A. peregrina* demonstraram se adaptar bem em condições de estresse químico provocado pela presença de doses crescentes de Pb nos solos. Isso pode ser verificado tanto pelas variáveis mensuradas da parte aérea, quanto das raízes. Para o IQD, em que avalia a qualidade das mudas e a capacidade de sobrevivência, foi possível verificar que as doses crescentes de Pb não influenciaram as espécies *P. juliflora* e *A. peregrina*, enquanto para a *M. oleífera* ocorreu um decréscimo linear nessa variável. Diferentemente do presente estudo, Usman et al. (2019) verificaram efeito tóxico nas plantas de *P. juliflora* nos tratamentos com 100 mg L⁻¹ de Pb, apresentando sintomas como clorose foliar, murcha e escurecimento das

raízes, além de redução no crescimento radicular das plantas. De acordo com Khan et al. (2015), a capacidade das plantas de *P. juliflora* se adaptarem aos solos contaminados estar relacionado a presença de microrganismos tolerantes a PTEs em sua rizosfera. Já para a espécie *A. peregrina* estudos têm demonstrado boa capacidade de adaptação em solos contaminados, porém em sua maioria relacionados a contaminação por arsênio (Gomes et al., 2012; Gomes et al., 2013), sendo necessário aprofundar pesquisas com os demais PTEs.

O Pb é um elemento pouco móvel nos solos, tendo sua adsorção regulada pelo pH, capacidade de troca de cátions, exsudação radiculares e dentre outras propriedades físico-químicas, porém, quando presente em solução é facilmente absorvido pelas plantas (Cullen e McAlister, 2017; Kumar e Prasad, 2018). Após o cultivo das espécies o pH dos solos permaneceram em sua maioria acima de 6, o que promove redução na disponibilidade do Pb, e consequentemente, diminuindo a absorção pelas plantas (Tabela S2). A absorção do Pb dos solos pelas raízes das plantas ocorre através da membrana plasmática, principalmente pelos canais de cálcio, devido a semelhança desse metal com os alcalinos terrosos (Kabata-Pendias e Pendias, 2001). Em baixas concentrações de Pb, o movimento dentro da raiz ocorre via apoplasto e acumula-se próximo da endoderme, a qual funciona como uma barreira impedindo a translocação do metal para a parte aérea das plantas (Kumar e Prasad, 2018). Dessa forma, as raízes das plantas é o principal órgão afetado pelo Pb. Isso pode ser verificado, principalmente nas plantas de *M. oleífera* e *U. ruziziensis*, que apresentaram reduções no comprimento, área superficial, volume, comprimento e no número de cruzamentos e garfos das raízes, além de promover um aumento no diâmetro médio das raízes. Elevados teores de Pb interferem nos processos de divisão celular do meristema da raiz causando redução no alongamento do sistema radicular das plantas (Baligar et al., 1998; Eun et al., 2000). Além disso, o Pb acumulado no sistema radicular pode causar um aumento do meristema apical provocando o espessamento da parede celular, aumento no tamanho do vacúolo e escurecimento do sistema radicular (Probst et al., 2009; Rucińska-Sobkowiak et al., 2013). De acordo com Kumar e Prasad (2018), o Pb nas raízes pode precipitar extracelularmente, ligando-se aos locais de troca iônica nas paredes celulares.

Em altas concentrações, o Pb pode danificar a parede celular das raízes e ser translocado para a parte aérea fazendo com que ocorra um aumento nos teores do metal (Romeiro et al., 2007; Usman et al., 2019). Além disso, diversas proteínas transportadoras como as ATPases de metais pesados do tipo CPx têm influência tanto no transporte de micronutrientes, quanto de elementos potencialmente tóxicos através das membranas celulares. Esses transportadores utilizam ATP com a finalidade de bombear uma variedade de substratos através das membranas

celulares (Williams et al., 2000; Yang et al., 2005). Quando presente na parte aérea, o Pb pode promover clorose foliar, modificações nas atividades enzimáticas, inibição ou redução da fotossíntese, modificação de características anatômicas, afeta a estrutura e a permeabilidade da membrana, induz o aumento do número de estômatos, entre outros (Probst et al., 2009; Moraes et al., 2014; Hatamian et al., 2020). Alguns desses sintomas foram verificados, como por exemplo, a redução da atividade fotossintética e a transpiração de todas as espécies estudadas. Também foi verificado redução da condutância estomática e da eficiência de carboxilação das espécies de *M. oleífera* e *U. ruziziensis*. Além disso, foi verificado que o Pb interfere nos padrões de fluorescência da clorofila de todas as espécies de plantas. Apesar da redução nas variáveis que medem as trocas gasosas das plantas de *M. oleífera*, os valores ainda estão de acordo ao encontrado na literatura (Vasconcelos, 2013). A redução das variáveis da atividade fotossintética das plantas, está relacionada aos danos causados pelo Pb aos cloroplastos (Liu et al., 2008; Gupta et al., 2011; Moraes et al., 2014; Hatamian et al., 2020; Giannakoula et al., 2021). Como a fotossíntese é responsável pela assimilação de CO₂, qualquer redução nessa variável provoca diminuição no desenvolvimento das plantas. O Pb pode provocar inibição na síntese de clorofila, no transporte de elétrons para os fotossistemas e consequentemente na atividade dos fotossistemas, principalmente o II (Kumar e Prasad, 2015; Piwowarczyk et al., 2018).

Os teores de Pb nas raízes das plantas aumentaram linearmente com o aumento das doses nos solos, com exceção da *M. oleífera* e foram muito superiores aos teores de Pb na parte aérea. Esses resultados confirmam a capacidade das plantas estudadas em acumular esse metal no sistema radicular. Amorim et al. (2021) também verificaram maiores teores de Pb nas raízes de *U. brizantha*, em relação a parte aérea, porém os teores foram muito inferiores ao encontrado no presente estudo. Já em plantas de *M. oleífera*, Ogundiran et al. (2018) encontraram teores de Pb nas raízes e nos brotos muito superiores aos encontrados no presente estudo, com valores variando de 930 a 2100 mg kg⁻¹ e 420-1120 mg kg⁻¹, respectivamente. Corroborando com nosso estudo, Usman et al. (2019) também verificaram maiores teores de Pb nas raízes das plantas de *P. juliflora*, quando comparados com a parte aérea.

O fator de translocação (FT) (< 1) também confirma a capacidade de todas as plantas estudadas de acumular o Pb no sistema radicular, impedindo seu movimento para a parte aérea. O FT menor que 1 indica baixa eficiência de transferência do Pb, sugerindo maior acúmulo de metais nas raízes do que na parte aérea. (Maiti e Jaiswal, 2008). Em todas as espécies estudadas foi possível verificar que o FT apresentou maior valor na menor dose de Pb aplicada, e que as plantas de *P. juliflora* e *A. peregrina* apresentaram maiores valores de FT. Dentre os

mecanismos de defesa podemos citar a adsorção à parede celular, compartimentação no vacúolos, aumento do efluxo ativo ou indução de níveis mais elevados de quelatos metálicos como complexos proteicos (metalotioneínas, fitoquelatinas e nicotinaminas), complexos orgânicos (citratos) e inorgânicos (sulfuretos) (Yang et al., 2005; Rascio e Navari-izzo, 2011; Gupta et al., 2013; Corso e Torre, 2020). Os flavonoides por exemplo têm como principal função a eliminação das espécies reativas de oxigênio e a quelação de metais (Corso e Torre, 2020).

Os resultados alcançados corroboram com inúmeros estudos, os quais indicam que o FT para o Pb na maioria das espécies avaliadas é menor que 1, podendo estar relacionado a restrição da translocação desse metal para a parte aérea, e a capacidade de ser acumulado no sistema radicular, visto que, apenas 5 % do Pb absorvido pelas raízes é translocado para a parte aérea (Probst et al., 2009; Zhou et al., 2016; Bhatti et al., 2018; Chandra et al., 2018). Amorim et al. (2021) também encontraram valores de FT inferiores a 1 em *U. brizantha*. Além de apresentar altos teores de Pb, a produção elevada de massa de raiz pela *U. ruziziensis* torna essa planta uma alternativa interessante para estabilizar solos contaminados, com capacidade de acumular na raiz um conteúdo de até 17,6 g kg⁻¹ de Pb por planta. As gramíneas utilizam suas raízes para limitar a mobilidade de Pb nos solos, evitando que o mesmo se espalhe para áreas circunvizinhas (Egendorf et al., 2020). Usman et al. (2019) também encontraram valores de FT inferiores a 1 em plantas de *P. juliflora*, porém, diferentemente do presente estudo, os maiores valores de FT foram encontrados nas maiores doses de Pb.

Diversas estratégias são adotadas pelas plantas para impedir a translocação do Pb para a parte aérea, um exemplo é estria de Caspary presente na endoderme que atua como impedimento desse movimento (Kumar e Prasad 2018). Além disso, estudos tem demonstrado que para minimizar o estresse provocado pelo Pb, esse metal é sequestrado principalmente na parede celular, espaços intercelulares e vacúolos das células das raízes (Pourrut et al., 2011; Gupta et al., 2013; Corso e Torre, 2020). Plantas de *Brachiaria decumbens* apresenta como mecanismo de tolerância ao Pb a capacidade de sequestrá-lo nas paredes celulares deixando-o insolúvel (Kopittke et al., 2008). Alguns constituintes da parede celular como as pectinas e hemiceluloses apresentam potencial na adaptação das plantas aos elementos potencialmente tóxicos, pois, esses compostos são capazes de se ligar a íons e formar complexos estáveis com cátions divalentes e trivalentes (Corso e Torre, 2020). Em uma revisão de literatura, Egendorf et al. (2020) verificaram que 95 % do Pb na planta tende a se acumular no vacúolo e parede celular, e que no vacúolo o metal forma complexos, desintoxicando a planta.

Apesar de acumular altos teores de Pb no sistema radicular, todas as plantas avaliadas apresentaram um fator de bioacumulação (FA) inferiores 1, o que indica baixa de extrair

PTEs quando comparados com os teores presentes nos solos. Plantas com essas características são denominadas excludentes (Baker, 1981). Apesar de apresentar FA inferior a 1, as plantas de *U. ruziziensis* conseguiu acumular teores totais de Pb próximo de 1400 mg kg^{-1} . Estudos tem demonstrado que plantas com alta produção de massa de matéria seca vegetal e com capacidade de acumular teores de metais acima de 1000 mg kg^{-1} são ideais para programas de fitorremediação (Vamerali et al., 2010; Malik et al., 2010). Fator de acumulação menor que 1 em plantas de *U. ruziziensis* está relacionado as concentrações muito elevadas de Pb nos tratamentos ($4,2 \text{ g kg}^{-1}$ de Pb). Dentre esses mecanismos podemos destacar a tentativa de impedir que o Pb seja absorvido pelas raízes por meio de mudanças no pH da rizosfera e a liberação de exsudados radiculares os quais podem complexar com Pb diminuindo sua mobilidade (Bluskov et al., 2005; Khan et al., 2016). Amorim et al. (2021) também encontraram valores de FA inferiores a 1 em *U. brizantha*. Além disso, os mesmos autores verificaram que na rizosfera das plantas ocorreram mudanças na especiação do Pb, diminuindo a proporção de Pb ligados aos argilominerais e aumentando o Pb complexado a matéria orgânica, limitando a absorção desse metal.

A liberação de compostos pelo sistema radicular das plantas, como por exemplo, o O_2 pode promover a oxidação de íons Pb^{2+} , reduzindo sua mobilidade, assim como a liberação de compostos orgânicos de baixo peso molecular que também pode modificar a mobilidade de Pb, por meio da formação de complexos (Bluskov et al., 2005; Wang et al., 2010; Khan et al., 2016). Diferentemente do determinado no presente estudo, Usman et al. (2019) em experimento com solução nutritiva verificaram FA superior a 1 em plantas de *P. juliflora* indicando a grande capacidade dessas plantas em acumular metais.

As doses crescentes de Pb nos solos também influenciaram no balanço e equilíbrio nutricional de todas as espécies de plantas estudadas. Na parte aérea das plantas foi possível verificar que os teores de Ca diminuíram com o aumento das doses de Pb em todas as espécies estudadas. Nas maiores doses de Pb, também ocorreu redução nos teores de Mn e Mg para todas as espécies estudadas, com exceção apenas do Mg para a *A. peregrina*. Diversos estudos têm demonstrado que a absorção de Pb pelas plantas ocorrem por canais catiônicos, como os canais de Ca. Kim et al. (2002) verificaram efeito antagônico do Pb na absorção de Ca e Mg. Huang e Cunningham (1996) observaram que o Pb promoveu a redução em 40 % nos teores de Ca na parte aérea de plantas de milho, como também reduziu os teores de Mg, K, Fe, Mn e Zn. De acordo com Kabata-Pendias (2011) e Lamhamdi et al. (2013) a interferência do Pb com o Ca é de importância metabólica, visto que o Pb pode mimetizar o comportamento fisiológico do Ca e, portanto, inibir algumas enzimas alterando o metabolismo das plantas.

Em relação aos teores de K na parte aérea não houve diferença significativo com o aumento das doses de Pb para as espécies *A. peregrina* e *P. juliflora*, enquanto para as demais espécies ocorreu um aumento linear. Paiva et al. (2003) verificaram redução nos teores de K em plantas de *Tabebuia impetiginosa* com o aumento das doses de Pb. Para os teores de Zn na parte aérea, o comportamento foi semelhante ao Pb, visto que o rejeito utilizado nos tratamentos também apresentava elevadas concentrações de Zn.

Para os teores dos elementos P e S na parte aérea foi possível verificar que não houve diferença significativa com o aumento das doses de Pb nos solos. Isso pode estar relacionado as características desse elemento, visto que são aniônicos, enquanto que o Pb é catiônico, podendo não ocorrer competição entre os sítios de absorção. Huang e Cunningham (1996) também verificaram que a concentração de Pb (20 μmol) não influenciaram os teores de P na parte aérea de plantas de *Zea mays* e *Ambrosia artemisiifolia*.

Para os teores de nutrientes nas raízes foi possível verificar uma tendência de aumento para todos os nutrientes à medida que aumentou as doses de Pb nos solos. A exceção foi para as plantas de *M. oleífera* em que o comportamento foi o oposto para a maioria dos teores de elementos nas raízes. O aumento nos teores pode estar relacionado a características do material utilizado para obtenção dos tratamentos, visto que são ricos em diversos nutrientes o que possivelmente competiu com Pb nos sítios de absorção.

Os rejeitos de mineração usado nos tratamentos é oriundo da pilha de rejeito de mineração de Pb da cidade de Boquira-Ba. Essa pilha de rejeito está disposta muito próximo da cidade e de áreas agricultáveis e vêm sofrendo com processos erosivos (erosão eólica e hídrica) contaminando o entorno da cidade. Diante disso, o presente estudo traz como alternativa o uso de espécies com potencial para estabilizar a pilha de rejeito evitando a dispersão dos contaminantes. Apesar da *U. ruziziensis* ter seus parâmetros de crescimento afetados nas maiores doses, a mesma possui um sistema radicular abundante e apresenta crescimento rápido, sendo capaz de estabilizar rapidamente elevados teores de metais. Além disso, as espécies *A. peregrine* e *P. juliflora* não foram tão afetadas pelos altos teores de Pb e por serem leguminosas com capacidade de fixar nitrogênio podem também estabilizar os metais na área da pilha de rejeito, pois essas espécies favorece a comunidade microbiológica dos solos. Essas espécies por estarem presente na região de estudo são altamente adaptadas as condições edafoclimáticas locais.

5.0- CONCLUSÕES

A utilização de espécies vegetais para remediação de áreas contaminadas é uma alternativa eficiente e de baixo custo. No presente estudo ficou demonstrado o potencial das espécies *M. oleífera*, *U. ruziziensis*, *A. peregrine* e *P. juliflora* em crescerem e se desenvolverem em ambientes altamente contaminados com Pb.

Todas as espécies estudadas apresentaram fatores de translocação e de bioacumulação inferiores a 1, sendo, portanto, consideradas plantas excludentes, com baixa capacidade de transferência de metais da raiz para a parte aérea e do solo para planta.

Apesar das plantas de *U. ruziziensis* apresentar FA menor que 1, as mesmas conseguiram acumular teores de Pb acima de 1400 mg kg⁻¹, sendo, portanto, uma espécie com potencial para fitorremediação.

Dentre as espécies estudadas a *M. oleífera* foi a mais afetada pelas doses crescentes de Pb nos solos, promovendo reduções em praticamente todas as variáveis biométricas e trocas gasosas avaliadas. A *A. peregrine* e *P. juliflora* foram as espécies que apresentaram os menores efeitos tóxicos relacionados as doses crescentes de Pb, sendo prontamente recomenda para cultivos nesses ambientes

As plantas de *U. ruziziensis* conseguiram crescer em solos contaminados com Pb e por apresentar um sistema radicular desenvolvido, possuem capacidade de acumular teores acima de 1400 mg kg⁻¹ de Pb, sendo também recomendada para estabilizar solos contaminados com Pb.

De forma geral, a aplicação de doses crescentes de Pb provocaram reduções nos teores de nutrientes catiônicos na parte aérea das plantas, principalmente Ca e Mg, enquanto que os teores de nutrientes aniônicos como o P e S não foram afetados.

6.0- AGRADECIMENTOS

Este estudo foi parcialmente financiado pela Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Programa Nacional de Cooperação Acadêmica-PROCAD 2013 - Código Financeiro 001.

7.0- REFERÊNCIAS

Alencar, L. P. D., Sedyama, G. C., Mantovani, E. C., & Martinez, M. A. (2011). Tendências recentes nos elementos do clima e suas implicações na evapotranspiração da cultura do milho em Viçosa-MG. Engenharia Agrícola, 31, 631-642.

Amorim, H. C., Hurtarte, L. C., Vergütz, L., Silva, I. R., Costa, O. D., Pacheco, A. A., & Fontes, M. P. (2021). Lead speciation and availability affected by plants in a contaminated soil. *Chemosphere*, 285, 131468.

Andrade, Á. A. X. D., Soares, E. M. B., Cunha, D. A. D., & Oliveira, M. L. R. D. (2017). Riscos e incertezas: a realidade pós-extração do minério de chumbo em Boquira, BA. *Interações (Campo Grande)*, 18, 103-117.

Araújo, F. S., Martins, S. V., Neto, J. A. A. M., Lani, J. L., & Pires, I. E. (2006). Estrutura da vegetação arbustivo-arbórea colonizadora de uma área degradada por mineração de caulim, Brás Pires, MG. *Revista Árvore*, 30(1), 107-116.

Baker, A. J. (1981). Accumulators and excluders-strategies in the response of plants to heavy metals. *Journal of plant nutrition*, 3(1-4), 643-654.

Baligar, V. C.; Fageria, N. K.; Elrashidi, M. A. (1998). Toxicology and nutrient constraints on root growth. *Hortscience*, Alexandria, 33(6), 960-965

Bech, J., Duran, P., Roca, N., Poma, W., Sánchez, I., Barceló, J., ... & Poschenrieder, C. (2012). Shoot accumulation of several trace elements in native plant species from contaminated soils in the Peruvian Andes. *Journal of Geochemical Exploration*, 113, 106-111.

Bhatti, S. S., Kumar, V., Sambyal, V., Singh, J., & Nagpal, A. K. (2018). Comparative analysis of tissue compartmentalized heavy metal uptake by common forage crop: a field experiment. *Catena*, 160, 185-193.

Bluskov, S., Arocena, J. M., Omotoso, O. O., & Young, J. P. (2005). Uptake, distribution, and speciation of chromium in *Brassica juncea*. *International Journal of Phytoremediation*, 7(2), 153-165.

BRASIL. Lei Federal Nº 9.605, de 12 de fevereiro de 1998. Dispõe sobre as sanções penais e administrativas derivadas de condutas e atividades lesivas ao meio ambiente, e dá outras providências. Disponível em http://www.planalto.gov.br/ccivil_03/leis/19605.htm.

BRASIL. Decreto Nº 9.406, DE 12 de junho de 2018. Regulamenta o Decreto-Lei nº 227, de 28 de fevereiro de 1967, a Lei nº 6.567, de 24 de setembro de 1978, a Lei nº 7.805, de 18 de julho de 1989, e a Lei nº 13.575, de 26 de dezembro de 2017. Disponível em http://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/decreto/D9406.htm. Acesso em janeiro de 2022.

BRASIL. Lei nº 6938 de 1981: Dispõe sobre a política nacional do Meio Ambiente. 1981. Disponível em http://www.planalto.gov.br/ccivil_03/leis/l6938.htm. Acesso em janeiro de 2022.

Calabrese, E. J., & Baldwin, L. A (1998). O conceito de hormesis pode ser generalizado para a carcinogênese? *Regulatory Toxicology and Pharmacology* , 28 (3), 230-241.

Cannata, M. G., Carvalho, R., Bertoli, A. C., Augusto, A. S., Bastos, A. R. R., Carvalho, J. G., & Freitas, M. P. (2013). Effects of cadmium and lead on plant growth and content of heavy metals in arugula cultivated in nutritive solution. *Communications in soil science and plant analysis*, 44(5), 952-961.

CETESB - COMPANHIA DE TECNOLOGIA DE SANEAMENTO AMBIENTAL. (2013). Diretoria de controle e licenciamento ambiental. Texto explicativo – Relação de áreas contaminadas e reabilitadas no Estado de São Paulo.

Chandra, R., Kumar, V., Tripathi, S., & Sharma, P. (2018). Heavy metal phytoextraction potential of native weeds and grasses from endocrine-disrupting chemicals rich complex distillery sludge and their histological observations during in-situ phytoremediation. *Ecological Engineering*, 111, 143-156.

Christophe, H. L., Albert, N., Martin, Y., & Mbaiguinam, M. (2019). Effect of organic fertilizers rate on plant survival and mineral properties of *Moringa oleifera* under greenhouse conditions. *International Journal of Recycling of Organic Waste in Agriculture*, 8(1), 123-130.

Corso, M.; Torre, V. S. G. L. Biomolecular approaches to understanding metal tolerance and hyperaccumulation in plants, *Metallomics*, v. 12, p. 840-859, 2020.

Cullen, J. T., & McAlister, J. (2017). 2. Biogeochemistry of Lead. Its Release to the Environment and Chemical Speciation. Lead: Its effects on environment and health, 21-48.

Dhaliwal, S. S., Singh, J., Taneja, P. K., & Mandal, A. (2020). Remediation techniques for removal of heavy metals from the soil contaminated through different sources: a review. *Environmental Science and Pollution Research*, 27(2), 1319-1333.

Dickson, A.; Leaf, A. L.; Hosner, J. F. (1960). Quality appraisal of white spruce and white pine seedling stock in nurseries. *The Forest Chronicle*, v.36, n.1, p.10-13.

Egendorf, S. P.; Groffman, P.; Moore, G.; Cheng, Z. The limits of lead (Pb) phytoextraction and possibilities of phytostabilization in contaminated soil: a critical review. *International journal of phytoremediation*, 22(9), 916–930, 2020.

Eun, S. O. Youn, H. S. Lee, Y. (2000). Lead disturbs microtubule organization in the root meristem of *Zea mays*. *Physiologia Plantarum*, Copenhagen, 110, 357-365.

Fernandes, D. M, Sousa, R. M. F., Oliveira, A., Morais, S. A. L., Richter EM, Muñoz RAA (2015) *Moringa oleifera*: A potential source for production of biodiesel and antioxidant additives. *Fuel* 146:75-80.

Ferreira, D. F. (2011). Sisvar: a computer statistical analysis system. *Ciência e Agrotecnologia*, v.35, n.6, p. 1039-1042.

Ferreira, P. M. P., Farias, D. F., Oliveira, J. T. D. A., & Carvalho, A. D. F. U. (2008). *Moringa oleifera*: bioactive compounds and nutritional potential. *Revista de Nutrição*, 21, 431-437.

Gabriel, I. E., & Patten, T. (1994). Distribution of copper smelter emissions in southeastern Arizona—using honey mesquite as a bioindicator. *Water, Air, and Soil Pollution*, 72(1), 67-87.

George, K. S., Revathi, K. B., Deepa, N., Sheregar, C. P., Ashwini, T. S., & Das, S. (2016). A study on the potential of Moringa leaf and bark extract in bioremediation of heavy metals from water collected from various lakes in Bangalore. *Procedia Environmental Sciences*, 35, 869-880.

Giannakoula, A., Therios, I., & Chatzissavvidis, C. (2021). Effect of lead and copper on photosynthetic apparatus in citrus (*Citrus aurantium* L.) plants. The role of antioxidants in oxidative damage as a response to heavy metal stress. *Plants*, 10(1), 155.

Gomes, M. P., Carvalho, M., Carvalho, G. S., Marques, T. C. L. L. S. M., Garcia, Q. S., Guilherme, L. R. G., & Soares, A. M. (2013). Phosphorus improves arsenic phytoremediation by *Anadenanthera peregrina* by alleviating induced oxidative stress. *International journal of phytoremediation*, 15(7), 633-646.

Gomes, M. P., Marques, R. Z., Nascentes, C. C., & Scotti, M. R. (2020). Synergistic effects between arbuscular mycorrhizal fungi and rhizobium isolated from As-contaminated soils on the As-phytoremediation capacity of the tropical woody legume *Anadenanthera peregrina*. *International Journal of Phytoremediation*, 22(13), 1362-1371.

Gomes, M. P., Moreira Duarte, D., Silva Miranda, P. L., Carvalho Barreto, L., Matheus, M. T., & Garcia, Q. S. (2012). The effects of arsenic on the growth and nutritional status of *Anadenanthera peregrina*, a Brazilian savanna tree. *Journal of Plant Nutrition and Soil Science*, 175(3), 466-473.

Gopalakrishnan, L., Doriya, K., & Kumar, D. S. (2016). *Moringa oleifera*: A review on nutritive importance and its medicinal application. *Food science and human wellness*, 5(2), 49-56.

Gupta, D. K., Huang, H. G., & Corpas, F. J. (2013). Lead tolerance in plants: strategies for phytoremediation. *Environmental Science and Pollution Research*, 20(4), 2150-2161.

Gupta, D. K., Nicoloso, F. T., Schetinger, M. R., Rossato, L. V., Huang, H. G., Srivastava, S., & Yang, X. E. (2011). Lead induced responses of *Pfaffia glomerata*, an economically important Brazilian medicinal plant, under in vitro culture conditions. *Bulletin of environmental contamination and toxicology*, 86(3), 272-277.

Hatamian, M., Nejad, A. R., Kafi, M., Souri, M. K., & Shahbazi, K. (2020). Interaction of lead and cadmium on growth and leaf morphophysiological characteristics of European hackberry (*Celtis australis*) seedlings. *Chemical and Biological Technologies in Agriculture*, 7(1), 1-8.

Huang, J. W.; Cunningham, S. D. (1996). Lead phytoextraction: species variation in lead uptake and translocation. *The New Phytologist*, 134(1), 5-84.

Jayaram, K., & Prasad, M. N. V. (2009). Removal of Pb (II) from aqueous solution by seed powder of *Prosopis juliflora* DC. *Journal of hazardous materials*, 169(1-3), 991-997.

Jiang, F., Ren, B., Hursthouse, A., Deng, R., & Wang, Z. (2019). Distribution, source identification, and ecological-health risks of potentially toxic elements (PTEs) in soil of thallium mine area (southwestern Guizhou, China). *Environmental Science and Pollution Research*, 26(16), 16556-16567.

Kabata-Pendias, (2011) Trace elements in soils and plants - 4th ed. by Taylor and Francis Group, LLC.

Kabata-Pendias, A.; Pendias, H. (2001). Trace elements in soil and plants. 3rd ed. Boca Raton: CRC, 331 p.

Kasolo, J. N., Bimenya, G. S., Ojok, L., Ochieng, J., & Ogwal-Okeng, J. W. (2010). Phytochemicals and uses of *Moringa oleifera* leaves in Ugandan rural communities. *Journal of Medicinal Plants Research*, 4(9), 753-757.

Khan, I., Iqbal, M., Ashraf, M. Y., Ashraf, M. A., & Ali, S. (2016). Organic chelants-mediated enhanced lead (Pb) uptake and accumulation is associated with higher activity of enzymatic antioxidants in spinach (*Spinacea oleracea* L.). *Journal of hazardous materials*, 317, 352-361.

Khan, M. U., Sessitsch, A., Harris, M., Fatima, K., Imran, A., Arslan, M., ... & Afzal, M. (2015). Cr-resistant rhizo- and endophytic bacteria associated with *Prosopis juliflora* and their potential as phytoremediation enhancing agents in metal-degraded soils. *Frontiers in Plant Science*, 5, 755.

Kim, Y. Y., Yang, Y. Y., & Lee, Y. (2002). Pb and Cd uptake in rice roots. *Physiologia Plantarum*, 116(3), 368-372.

Kopittke, P. M., Asher, C. J., Blamey, F. P. C., Auchterlonie, G. J., Guo, Y. N., & Menzies, N. W. (2008). Localization and chemical speciation of Pb in roots of signal grass (*Brachiaria decumbens*) and Rhodes grass (*Chloris gayana*). *Environmental science & technology*, 42(12), 4595-4599.

Kulshreshtha, A., Agrawal, R., Barar, M., & Saxena, S. (2014). A review on bioremediation of heavy metals in contaminated water. *IOSR Journal of Environmental Science, Toxicology and Food Technology*, 8(7), 44-50.

Kumar, A., Maiti, S. K., Prasad, M. N. V., & Singh, R. S. (2017). Grasses and legumes facilitate phytoremediation of metalliferous soils in the vicinity of an abandoned chromite–asbestos mine. *Journal of soils and sediments*, 17(5), 1358-1368.

Kumar, A., & Prasad, M. N. V. (2015). Lead-induced toxicity and interference in chlorophyll fluorescence in *Talinum triangulare* grown hydroponically. *Photosynthetica*, 53(1), 66-71.

Lamhamdi, M., El Galiou, O., Bakrim, A., Nóvoa-Muñoz, J. C., Arias-Estévez, M., Aarab, A., & Lafont, R. (2013). Effect of lead stress on mineral content and growth of wheat (*Triticum aestivum*) and spinach (*Spinacia oleracea*) seedlings. *Saudi journal of biological sciences*, 20(1), 29-36.

Lê, S. J.; Josse, F. (2008). *Husson Facto MineR: An R Package for multivariate analysis* J. Stat. Softw., 25, pp. 1-18, 10.18637/jss.v025.i01.

Lima, L. D. A., Bernardez, L. A., dos Santos, M. G., & Souza, R. C. (2018). Remediation of clay soils contaminated with potentially toxic elements: The Santo Amaro lead smelter, Brazil, case. *Soil and Sediment Contamination: An International Journal*, 27(7), 573-591.

- Liu, D., Islam, E., Li, T., Yang, X., Jin, X., & Mahmood, Q. (2008). Comparison of synthetic chelators and low molecular weight organic acids in enhancing phytoextraction of heavy metals by two ecotypes of *Sedum alfredii* Hance. *Journal of Hazardous Materials*, 153(1-2), 114-122.
- Lorenzon, A. S., Dias, H. C. T., & Leite, H. G. (2013). Precipitação efetiva e interceptação da chuva em um fragmento florestal com diferentes estágios de regeneração. *Revista Árvore*, 37, 619-627.
- Magalhães, P. C.; Souza, T. C.; Cantão, F. R. O. (2011). Early evaluation of root morphology of maize genotypes under phosphorus deficiency. *Plant Soil Environ.*, 57(3): 135-138.
- Maiti, S. K., & Jaiswal, S. (2008). Bioaccumulation and translocation of metals in the natural vegetation growing on fly ash lagoons: a field study from Santaldih thermal power plant, West Bengal, India. *Environmental monitoring and assessment*, 136(1), 355-370.
- Maiti, S. K., & Maiti, D. (2015). Ecological restoration of waste dumps by topsoil blanketing, coir-matting and seeding with grass–legume mixture. *Ecological Engineering*, 77, 74-84.
- Malik, R. N., Husain, S. Z., & Nazir, I. (2010). Heavy metal contamination and accumulation in soil and wild plant species from industrial area of Islamabad, Pakistan. *Pak J Bot*, 42(1), 291-301.
- Moyo, B., Masika, P. J., Hugo, A., & Muchenje, V. (2011). Nutritional characterization of Moringa (*Moringa oleifera* Lam.) leaves. *African Journal of Biotechnology*, 10(60), 12925-12933.
- Miralles, J., Veron, A. J., Radakovitch, O., Deschamps, P., Tremblay, T., & Hamelin, B. (2006). Atmospheric lead fallout over the last century recorded in Gulf of Lions sediments (Mediterranean Sea). *Marine Pollution Bulletin*, 52(11), 1364-1371.
- Moraes, C. L., Marini, P., Fernando, J. A., de Moraes, D. M., de Castro, L. A. S., & Lopes, N. F. (2014). Alterações fisiológicas e ultraestruturais de plântulas de tomate induzidas por chumbo. *Iheringia. Série Botânica.*, 69(2), 313-322.

- Mota, F. A. C.; Sargentini Júnior, E.; Santana, G. P. (2016). Potencial de fitorremediação da *Alocasia macrorrhiza* para Co, Cu, Ni e Zn. *Scientia Amazonia*, 5(3), 11-20.
- Morkunas, I., Woźniak, A., Mai, V. C., Rucińska-Sobkowiak, R., & Jeandet, P. (2018). The role of heavy metals in plant response to biotic stress. *Molecules*, 23(9), 2320.
- Mridha, M. A. U. (2015). Prospects of moringa cultivation in Saudi Arabia. *J. Appl. Environ. Biol. Sci*, 5(3), 39-46.
- Mukhopadhyay, S., & Maiti, S. K. (2010). Phytoremediation of metal mine waste. *Applied Ecology and Environmental Research*, 8(3), 207-222.
- Novais, R. F.; Neves, J. C. L.; Nairam, F. B. Ensaios em ambiente controlado. In: Oliveira, A. J.; Garrido, W. E.; Araújo, J. D.; Lourenço, S. Métodos de pesquisa em fertilidade do solo. Embrapa; 1991; p. 189-254.
- Ogundiran, M. B., Mekwunyei, N. S., & Adejumo, S. A. (2018). Compost and biochar assisted phytoremediation potentials of *Moringa oleifera* for remediation of lead contaminated soil. *Journal of Environmental Chemical Engineering*, 6(2), 2206-2213.
- Paiva, H. N. D., Carvalho, J. G. D., Siqueira, J. O., Fernandes, A. R., & Miranda, J. R. P. D. (2003). Influência de doses crescentes de chumbo sobre o teor e o conteúdo de nutrientes e Pb em mudas de ipê-roxo (*Tabebuia impetiginosa* (Mart.) Standl.). *Revista Árvore*, 27, 151-158.
- Pan, L., Guan, X., Liu, B., Chen, Y., Pei, Y., Pan, J., Zhang, Y., & Hao, Z. (2021). Pollution Characteristics, Distribution and Ecological Risk of Potentially Toxic Elements in Soils from an Abandoned Coal Mine Area in Southwestern China. *Minerals*, 11(3), 330.
- Piowarczyk, B., Tokarz, K., Muszyńska, E., Makowski, W., Jędrzejczyk, R., Gajewski, Z., & Hanus-Fajerska, E. (2018). The acclimatization strategies of kidney vetch (*Anthyllis vulneraria* L.) to Pb toxicity. *Environmental Science and Pollution Research*, 25(20), 19739-19752.

Pourrut, B.; Shahid, M.; Dumat, C.; WInterton, P.; Pinelli, E. (2011) Lead uptake, toxicity, and detoxification in plants. *Rev Environ Contam Toxicol*, 213, 113–136.

Probst, A., Liu, H., Fanjul, M., Liao, B., & Hollande, E. (2009). Resposta de *Vicia faba* L. à toxicidade de metais em substratos de rejeitos de minas: mudanças geoquímicas e morfológicas em folhas e raízes. *Botânica Ambiental e Experimental*, 66 (2), 297-308.

R. Core Team. R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria (2021) <https://www.R-project.org/>

Rascio, N. & Navari-Izzo, F. (2011). Heavy metal hyperaccumulating plants: How and why do they do it? And what makes them so interesting? *Plant Science*, 180(2), 169-181.

Ribaski, J., Drumond, M. A., de Oliveira, V. R., & Nascimento, C. D. S. (2009). Algaroba (*Prosopis juliflora*): árvore de uso múltiplo para a região semiárida brasileira. Embrapa Florestas-Comunicado Técnico (INFOTECA-E).

Romeiro, S., Lagôa, A. M. M. A., Furlani, P. R., Abreu, C. A. D., & Pereira, B. F. F. (2007). Lead uptake and potential for fitoremediation of *Canavalia ensiformes* L. *Bragantia*, 66(2), 327-334.

Rucińska-Sobkowiak, R., Nowaczyk, G., Krzesłowska, M., Rabęda, I., & Jurga, S. (2013). Water status and water diffusion transport in lupine roots exposed to lead. *Environmental and experimental botany*, 87, 100-109.

Santos, N. L., dos Anjos, J. Â. S. A., & Klammler, H. (2020). Exposição da Zona Urbana de Boquira, estado da Bahia, aos Metais Tóxicos Associados à Bacia de Rejeito de Mineração Abandonada. *Anuário do Instituto de Geociências*, 43(3), 280-291.

Savegnago, R.P. Caetano, S.L. Ramos, S.B. Nascimento, G.B. Schmidt, G.S. Ledur, M.C. Munari. D.P. (2011). Estimates of genetic parameters, and cluster and principal components analyses of breeding values related to egg production traits in a White Leghorn population *Poult. Sci.*, 90, pp. 2174-2188, 10.3382/ps.2011-01474.

Sena, M. M. Frighetto, R. T. S. Valarini, P. J. Tokeshi, H. Poppi R. J (2002). Discrimination of management effects on soil parameters by using principal component analysis: a multivariate analysis case study, *Soil Tillage Res.*, 67. pp. 171-181, 10.1016/S0167-1987(02)00063-6.

Senthilkumar, P., Prince, WS, Sivakumar, S., & Subbhuraam, CV (2005). *Prosopis juliflora* — uma solução verde para descontaminar solos contaminados por metais pesados (Cu e Cd). *Chemosphere*, 60 (10), 1493-1496.

Souza, F. M. D., Pereira, W. E., Dantas, J. S., Nóbrega, J. S., Lima, E. C. S., & Sá, F. V. D. S. (2018). Initial growth of *Moringa oleifera* Lam. as a function of poultry litter doses and granulometry. *Pesquisa Agropecuária Tropical*, 48, 399-406.

Teixeira, W. G. Novos métodos de recomendação da necessidade de calagem e especiação de cálcio por espectroscopia XANES em solos de Minas Gerais [Tese]. Viçosa: Universidade Federal de Viçosa; 2017.

Usman, K., Abu-Dieyeh, M. H., & Al-Ghouti, M. A. (2019). Evaluating the invasive plant, *Prosopis juliflora* in the two initial growth stages as a potential candidate for heavy metal phytostabilization in metalliferous soil. *Environmental Pollutants and Bioavailability*, 31(1), 145-155.

Vamerali, T., Bandiera, M., & Mosca, G. (2010). Field crops for phytoremediation of metal-contaminated land. A review. *Environmental Chemistry Letters*, 8(1), 1-17.

Vasconcelos, M. C. (2013). *Moringa oleifera* Lam.: aspectos morfométricos, fisiológicos e cultivo em gradiente de espaçamento. Dissertação (Mestrado em Agroecossistemas). Universidade Federal de Sergipe. <https://ri.ufs.br/handle/riufs/6543>

Wang, Y., Xie, Y., Li, W., Wang, Z., & Giammar, D. E. (2010). Formation of lead (IV) oxides from lead (II) compounds. *Environmental science & technology*, 44(23), 8950-8956.

Williams, L. E.; Pittman, J. K.; Hall, J. L. (2000). Emerging mechanisms for heavy metal transport in plants. *Biochimica et Biophysica Acta*, 1465, 104-126.

Yan, X., Liu, M., Zhong, J., Guo, J., & Wu, W. (2018). How human activities affect heavy metal contamination of soil and sediment in a long-term reclaimed area of the Liaohe River Delta, North China. *Sustainability*, 10(2), 338.

Yang, S., Taylor, D., Yang, D., He, M., Liu, X., & Xu, J. (2021). A synthesis framework using machine learning and spatial bivariate analysis to identify drivers and hotspots of heavy metal pollution of agricultural soils. *Environmental Pollution*, 117611.

Yang, X.; Feng Y. HE, Z.; Stoffella P. J. (2005). Molecular mechanisms of heavy metal hyperaccumulation and phytoremediation. *Journal of Trace Elements in Medicine and Biology*, 18(4), 339-35327.

Zanchi, C. S., Batista, E. R., Silva, A. O., Barbosa, M. V., Pinto, F. A., dos Santos, J. V., & Carneiro, M. A. C. (2021). Recovering soils affected by iron mining tailing using herbaceous species with mycorrhizal inoculation. *Water, Air, & Soil Pollution*, 232(3), 1-13.

Zhou, C., Huang, M., Li, Y., Luo, J., & ping Cai, L. (2016). Changes in subcellular distribution and antioxidant compounds involved in Pb accumulation and detoxification in *Neyraudia reynaudiana*. *Environmental Science and Pollution Research*, 23(21), 21794-21804.

APÊNDICE

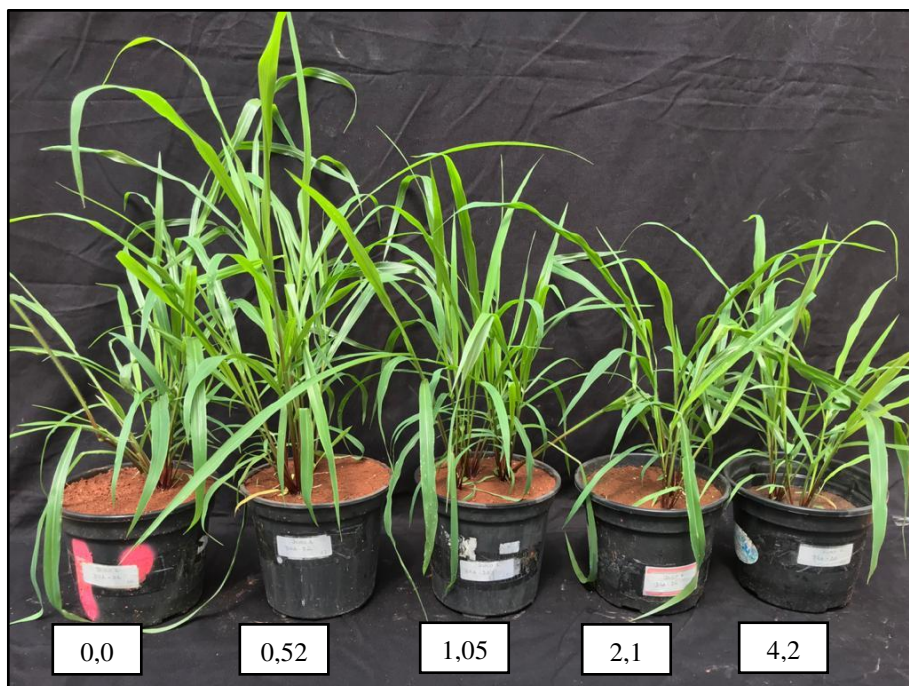


Figura S1: Plantas de *U. ruziziensis* cultivadas com doses crescentes de Pb (g kg⁻¹) nos solos aos 45 dias após o plantio.

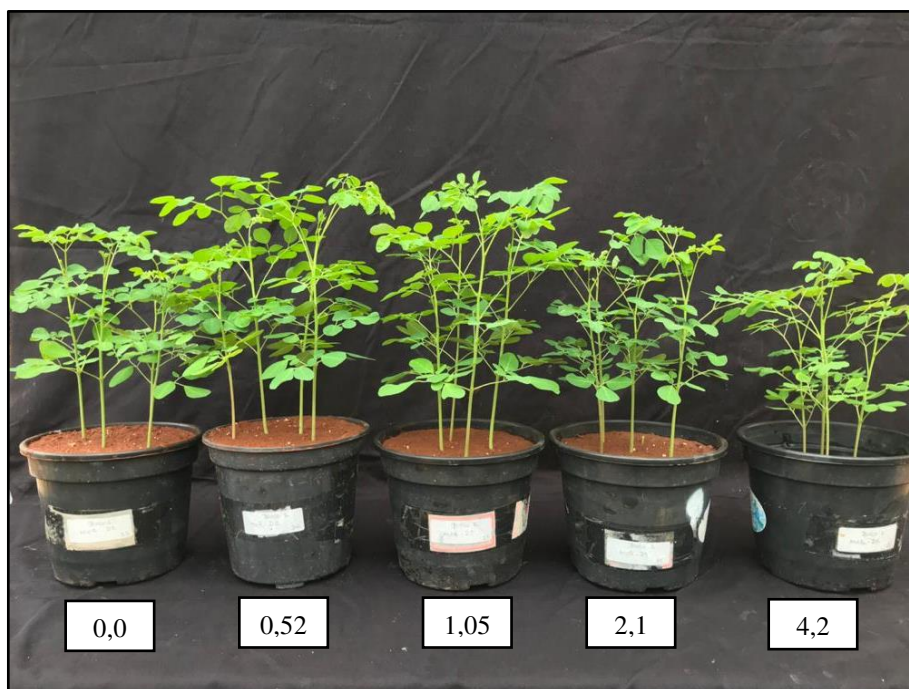


Figura S2: Plantas de *M. oleífera* cultivadas com doses crescentes de Pb (g kg⁻¹) nos solos aos 45 dias após o plantio.

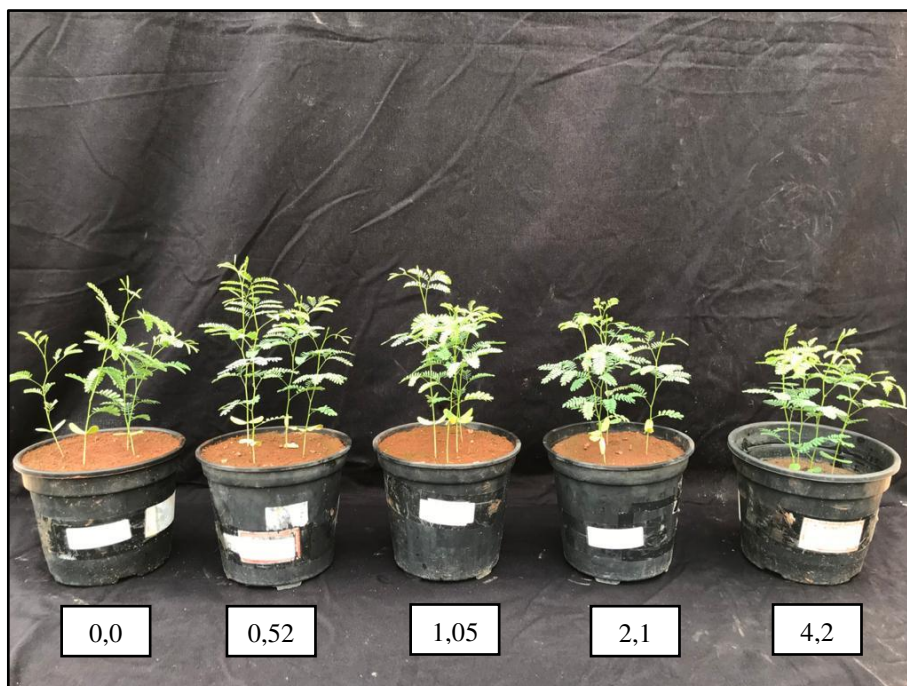


Figura S3: Plantas de *P. juliflora* cultivadas com doses crescentes de Pb (g kg^{-1}) nos solos aos 45 dias após o plantio.

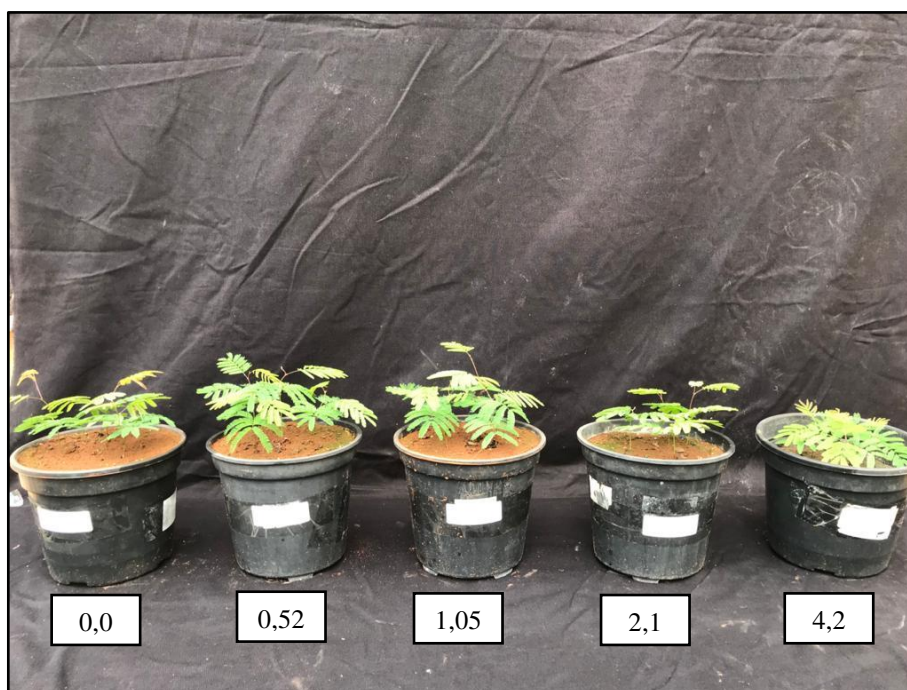


Figura S4: Plantas de *A. peregrine* cultivadas com doses crescentes de Pb (g kg^{-1}) nos solos aos 45 dias após o plantio



Figura S5: Raízes de plantas de *A. peregrine* cultivadas com doses crescentes de Pb (g kg^{-1}) nos solos

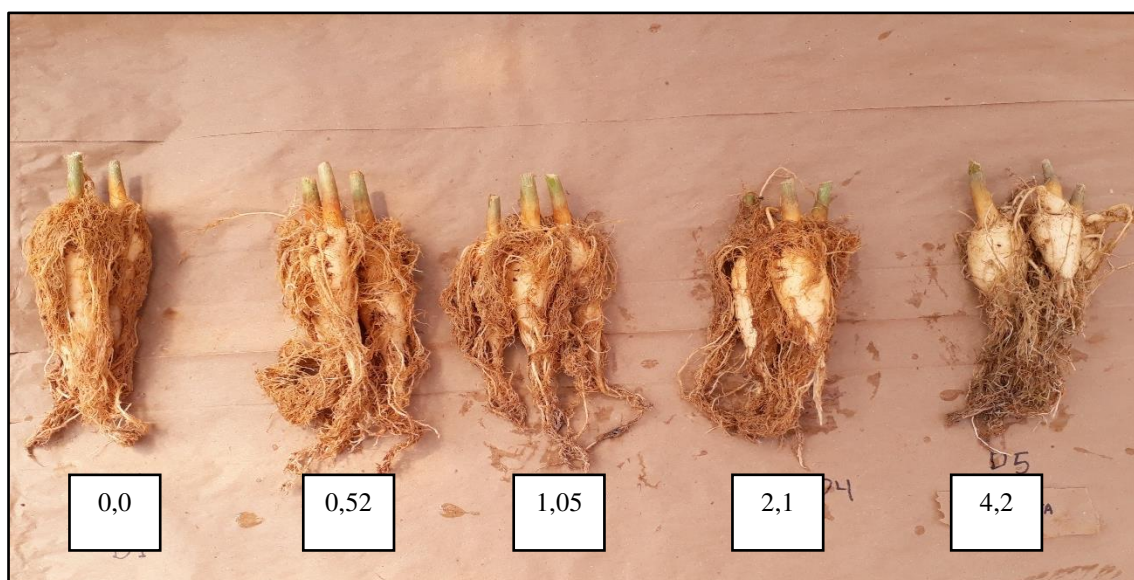


Figura S6: Raízes de plantas de *M. oleífera* cultivadas com doses crescentes de Pb (g kg^{-1}) nos solos

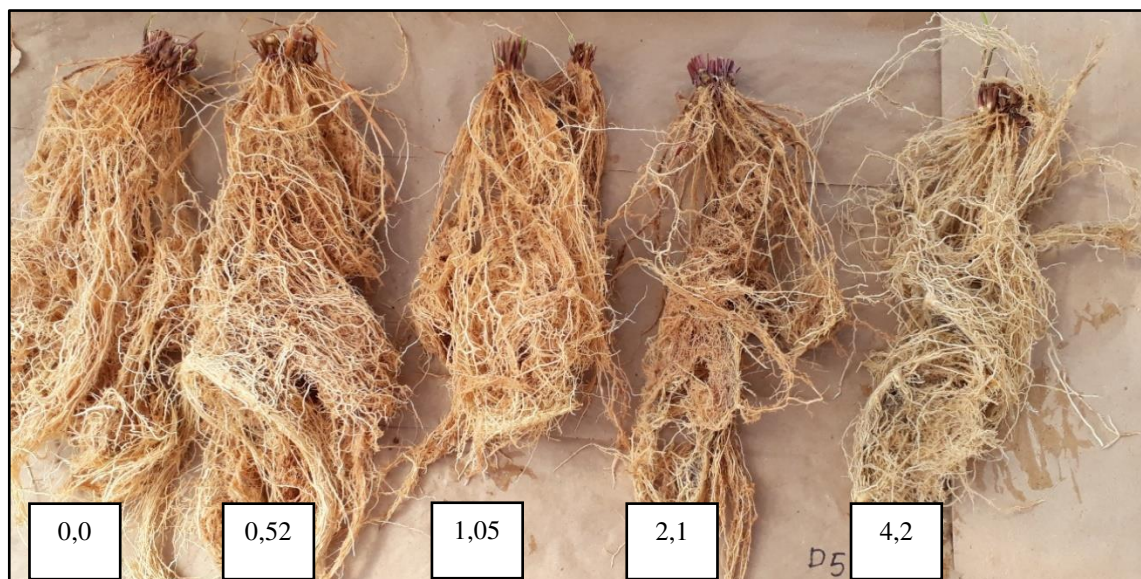


Figura S7: Raízes de plantas de *U. ruziziensis* cultivadas com doses crescentes de Pb (g kg⁻¹) nos solos



Figura S8: Raízes de plantas de *P. juliflora* cultivadas com doses crescentes de Pb (g kg⁻¹) nos solos

Tabela S1. Caracterização física e química do Latossolo Vermelho distrófico utilizado na diluição dos rejeitos para obtenção dos níveis de Pb.

pH	P	K	Na ⁺	Ca ²⁺	Mg ²⁺	M ^{x+}	H+Al ³⁺	CECe	CECp	V
H ₂ O	mg kg ⁻¹		----- cmol _c Kg ⁻¹ -----				-----			-- % --
5,25	0,3	6	-	0,58	0,04	1,26	7,7	1,90	8,34	7,7
M	MO	Cu	Fe	Mn	Zn	PD	Areia grossa	Areia fina	Silte	Argila
----- % -----	----- mg kg ⁻¹ -----				g cm ⁻³ -----		----- mg kg ⁻¹ -----			
66,3	3,36	0,67	26,1	13,6	33	-	80	30	80	810

Tabela S2. Valores médios de pH em cada tratamento após o cultivo das espécies.

	<i>P. juliflora</i>						<i>A. peregrina</i>				
Dose de Pb (mg kg ⁻¹)	0,00	0,53	1,05	2,10	4,20	0,00	0,53	1,05	2,10	4,20	
pH em H ₂ O	6,08	6,86	7,10	7,30	7,40	5,86	6,64	6,92	7,11	7,33	
	<i>M. oleifera</i>						<i>U. ruziziensis</i>				
Dose de Pb (mg kg ⁻¹)	0,00	0,53	1,05	2,10	4,20	0,00	0,53	1,05	2,10	4,20	
pH em H ₂ O	6,14	6,83	7,06	7,28	7,25	6,21	6,88	7,29	7,49	7,23	

CONSIDERAÇÕES FINAIS

O uso de algoritmos de aprendizagem de máquinas apresentou alto desempenho para predição e espacialização de elementos principais e elementos potencialmente tóxicos nos solos. No presente estudo comprovou-se que a área estudada na cidade de Boquira Bahia encontra-se altamente contaminada com Pb e Zn, principalmente nas áreas próximas da cidade e da pilha de rejeito de mineração, indicando ser a pilha de rejeito a principal fonte de contaminação. Além disso, os altos teores de elementos potencialmente tóxicos, principalmente o Pb, pode promover riscos à saúde da população de Boquira-Ba, como verificado pelos índices de risco não cancerígeno (HI) e riscos cancerígenos (CR) para crianças.

Nosso estudo também, desenvolveu uma estrutura metodológica integrando múltiplas covariáveis por meio de algoritmos de aprendizagem de máquinas para predizer os teores de elementos potencialmente tóxicos nos solos. A estrutura metodológica desenvolvida se mostrou rápida, eficiente e de baixo custo, além de ser ecologicamente sustentável, quando comparadas com as metodologias tradicionais, as quais geram resíduos químicos. Essa estrutura metodológica é uma alternativa eficiente principalmente em áreas contaminadas pela atividade de mineração, onde o monitoramento dos teores de elemento potencialmente tóxicos nos solos deve ser constante.

Por fim, nosso estudo também comprovou que as espécies vegetais *M. oleífera*, *U. ruziziensis*, *A. peregrine* e *P. juliflora* conseguiram crescerem em solos com altos teores de Pb. Entretanto apenas as espécies *A. peregrine* e *P. juliflora* foram as que apresentaram os menores efeitos tóxicos, sendo prontamente recomenda para cultivos nesses ambientes. Enquanto que as plantas de *U. ruziziensis* foi a que conseguiu acumular maiores teores de Pb nas raízes (1400 mg kg⁻¹). Apesar das espécies vegetais conseguirem crescerem nos solos com teores de Pb, as mesmas apresentaram baixa capacidade de transferência de Pb do solo para raiz (FA) e da raiz para a parte aérea (FT).