ALVARO JOSÉ GOMES DE FARIA

PREDICTION OF SOIL AGRONOMIC AND GEOCHEMICAL ATTRIBUTES: COMPARISON OF THE PERFORMANCE OF SPECTRAL RANGES FROM PROXIMAL SENSORS

LAVRAS – MG
2022

ALVARO JOSÉ GOMES DE FARIA

**PREDICTION OF SOIL AGRONOMIC AND GEOCHEMICAL ATTRIBUTES:
COMPARISON OF THE PERFORMANCE OF SPECTRAL RANGES FROM
PROXIMAL SENSORS**

> Tese apresentada à Universidade Federal de
> Lavras, como parte das exigências do
> Programa de Pós-Graduação em Ciência do
> Solo, área de concentração em Recursos
> Ambientais e Uso da Terra, para obtenção do
> título de Doutor.

Prof. Dr. Sérgio Henrique Godinho Silva
Orientador

Prof. Dr. Leônidas Carrijo Azevedo Melo
Co - Orientador

**LAVRAS – MG
2022**

ALVARO JOSÉ GOMES DE FARIA

PREDICTION OF SOIL AGRONOMIC AND GEOCHEMICAL ATTRIBUTES: COMPARISON OF THE PERFORMANCE OF SPECTRAL RANGES FROM PROXIMAL SENSORS

PREDIÇÃO DE ATRIBUTOS AGRONÔMICOS E GEOQUÍMICOS DO SOLO: COMPARAÇÃO DO DESEMPENHO DE FAIXAS ESPECTRAIS DE SENSORES PROXIMAIS

> Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Ciência do Solo, área de concentração em Recursos Ambientais e Uso da Terra, para obtenção do título de Doutor
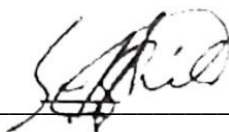
APROVADA em 20 de Maio de 2022.

Dr. Nilton Curi - UFLA

Dra. Geila Santos Carvalho - UFLA

Dr. Salvador Francisco Acunã Guzman – UPRM

Dr. Raul Roberto Poppiel – IG/UnB

_____
Dr. Sérgio Henrique Godinho Silva - UFLA
Orientador

_____
Dr. Leônidas Carrijo Azevedo Melo - UFLA
Co-Orientador

LAVRAS – MG
2022

Esta Tese é dedicada a toda minha família, pilares da minha formação como ser humano e profissional. Em especial a minha querida e amada avó Anedes Batista de Faria por toda confiança e incentivo imposto a mim durante toda a minha vida.

Dedico

AGRADECIMENTOS

ABSTRACT

Soil characterization provides a solid support for decision-making related to geochemical mapping, environmental monitoring, and food production. For that, quick, environmentally friendly, non-invasive, cost-effective, and reliable methods for soil chemical properties assessment are desirable. As such, this dissertation used proximal sensors like portable X-ray fluorescence (pXRF) spectrometry and Nix Pro^{TM} color sensor data to accurately predict soil properties in Brazil. The objectives were to: i) predict soil fertility properties in Brazilian Coastal Plains biome; ii) predict soil organic matter content via proximal sensors (pXRF and Nix Pro^{TM}); iii) predict elementary soil contents via USEPA 3051a through elementary data delivered by pXRF, evaluating samples preparation methods (field, post-field, air-dried fine earth, macerated, and macerated and sieved) and linear and non-linear regression methods. Four regression models - simple linear regression (SLR), stepwise multiple linear regressions (SMLR), support vector machine (SVM) with Linear Kernel and random forest (RF) - were tested for prediction of different soil agronomic attributes and assessment geochemical. The soil samples were collected in both surface and subsurface horizons in profiles of different soil classes, under several land uses, management practices, sampling sites, and with varying parent materials. Prediction models were built for surface, and subsurface horizons separately and combined for the following soil agronomic properties: pH (H$_2$O), sum of bases (SB); cation exchange capacity at pH 7.0 (CEC), and base saturation (BS) (*first chapter*). For soil organic matter (SOM) (*second chapter*) and 28 elements (*third and fourth chapters*) samples from surface and subsurface horizons were combined for building the prediction models. Samples were scanned with the Nix Pro^{TM} in the laboratory under both dry and moist conditions, while with pXRF only in dry condition also in laboratory. Samples were randomly separated into 70% for training and 30% for testing the prediction models. The performance of the prediction models was evaluated by the metrics: R^2, root mean square error (RMSE), normalized RMSE (NRMSE), mean absolute error (MAE), and residual prediction deviation (RPD). For soil agronomic properties, the results showed that SB was predicted with high accuracy (R^2 = 0.82, RMSE = 1.02 cmol$_c$ dm^{-3}, MAE = 1.17 and RPD = 2.3) using SVM models via pXRF data. Conversely, SOM was predicted with high accuracy using combined data from pXRF and Nix Pro™ (in moist soil samples) (R^2 = 0.73, RMSE = 1.09% and RPD = 2.00) via RF models. Prediction of elemental contents commonly determined by the USEPA 3051a method via pXRF data after scanning samples treated as air-dried fine earth (<2 mm) is indicated. Since, it can provide better predictions compared to other sample preparation procedures indicated above. Machine learning algorithms (SVM and RF) performed better than SLR and SMLR for the prediction of Al, Ca, Cr, Cu, Fe, Mn, Pb, Sr, Ti, V, Zn, Zr, Ba, Bi, Cd, Ce, Co, Mg and Tl in tropical soils, whose R² and RPD values ranged from 0.52 to 0.94 and 1.43 to 3.62, respectively, as well as the lowest values of RMSE and NRMSE values (0.28 to 0.70 mg kg^{-1}). The results reported in this dissertation represent alternative methods for reducing costs and time needed for assessing such soil properties data, supporting agronomic and environmental decision making.

Keywords: chemical attributes modeling. proximal sensors. machine learning algorithms. soil analysis. soil fertility. tropical soils. pedology. geochemistry.

RESUMO

A caracterização do solo fornece um suporte sólido para a tomada de decisões relacionadas ao mapeamento geoquímico, monitoramento ambiental e produção de alimentos. Para isso, métodos rápidos, ecologicamente corretos, não invasivos, econômicos e confiáveis para avaliação das propriedades químicas do solo são desejáveis. Como tal, esta tese utilizou sensores proximais como espectrometria de fluorescência de raios X portátil (pXRF) e dados do sensor de cor Nix Pro™ para prever de forma pontual as propriedades do solo no Brasil. Especificamente, os objetivos foram: i) predizer as propriedades de fertilidade do solo no bioma Planície Costeira Brasileira; ii) predizer o teor de matéria orgânica do solo por meio de sensores proximal (pXRF e Nix Pro™); iii) predizer conteúdos elementares do solo via método USEPA 3051a através de dados elementares entregues por pXRF, avaliando métodos de preparação de amostras (e.g., campo, pós-campo, terra fina seca ao ar, macerado e macerado e peneirado) e métodos de regressão linear e não linear. Quatro modelos de regressão - simple linear regression (SLR), stepwise multiple linear regressions (SMLR), support vector machine (SVM) com Linear Kernel and random forest (RF) - foram testados para predição de diferentes atributos agronômicos do solo e avaliação geoquímica. As amostras de solo foram coletadas em horizontes superficiais e subsuperficiais em perfis de diferentes classes de solo, sob diversos usos da terra, práticas de manejo, locais de amostragem e com diferentes materiais de origem. Modelos de predição foram construídos para horizontes superficiais e subsuperficiais separadamente e combinados para as seguintes propriedades agronômicas do solo: pH (H_2O), soma de bases (SB); capacidade de troca catiônica em pH 7,0 (CEC) e saturação por bases (BS) (*primeiro capítulo*). Para a matéria orgânica do solo (MOS) (*segundo capítulo*) e 28 elementos (*terceiro e quarto capítulos*) amostras de horizontes superficiais e subsuperficiais foram combinadas para construir os modelos de predição. As amostras foram escaneadas com o sensor Nix Pro™ em laboratório sob a condições secas e úmidas, enquanto com pXRF apenas em condições secas também em laboratório. As amostras foram separadas aleatoriamente em 70% para treinamento e 30% para teste dos modelos de predição. O desempenho dos modelos de predição foi avaliado pelas métricas: R^2, root mean square error (RMSE), normalized RMSE (NRMSE), mean absolute error (MAE) e residual prediction deviation (RPD). Para as propriedades agronômicas do solo, os resultados mostraram que a SB foi prevista com alta precisão (R^2 = 0,82, RMSE = 1,02 cmolc dm^{-3}, MAE = 1,17 e RPD = 2,3) usando modelos SVM via dados de pXRF. Por outro lado, a MOS foi prevista com alta precisão usando dados combinados de pXRF e Nix Pro™ (em amostras de solo úmido) (R^2 = 0,73, RMSE = 1,09% e RPD = 2,00) por meio de modelos de RF. A predição dos teores elementares comumente determinados pelo método USEPA 3051a via dados do pXRF após escaneamento de amostras tratadas como terra fina seca ao ar (<2 mm) é indicada. Uma vez que, pode fornecer as melhores predições em comparação com outros procedimentos de preparação de amostras indicas acima. Algoritmos de aprendizado de máquina (SVM e RF) tiveram um desempenho melhor que SLR e SMLR para a previsão de Al, Ca, Cr, Cu, Fe, Mn, Pb, Sr, Ti, V, Zn, Zr, Ba, Bi, Cd, Ce, Co, Mg e Tl em solos tropicais, cujos valores de R² e RPD variaram de 0,52 a 0,94 e 1,43 a 3,62, respectivamente, assim como os menores valores de RMSE e NRMSE (0,28 a 0,70 mg kg^{-1}). Os resultados relatados nesta tese representam métodos alternativos para redução de custos e tempo necessário para a avaliação desses dados de propriedades do solo, auxiliando na tomada de decisões agronômicas e ambientais.

Palavras-Chave: modelagem de atributos químicos. sensores proximais. algoritmos de aprendizado de máquina. análise do solo. fertilidade do solo. solos tropicais. pedologia. geoquímica.

SUMÁRIO

FIRST PART – GENERAL INTRODUCTION

1. GENERAL INTRODUCTION

For decades, several laboratory methodologies with different levels of complexity have been applied in tropical soil analyses for characterization of soil attributes, adding decision-making (CHEN; MA, 1998; SILVA; NASCIMENTO; BIONDI, 2014; TEIXEIRA et al., 2017; WEINDORF; CHAKRABORTY, 2020). Such laboratory methodologies are essential and have great relevance in the scientific environment. Conversely, depending on the number of soil samples, these analyses may be time consuming and expensive, require chemical reagents and generate hazardous chemical residues that must be treated to prevent environmental and human hazards (BENEDET et al., 2021; NASCIMENTO; TENUTA FILHO, 2010; SILVA et al., 2021).

Currently, portable proximal sensors such as X-ray fluorescence (pXRF) spectrometry and the Nix Pro^{TM} color sensor are being increasingly adopted for multiple applications in soil science and other fields of study as reliable and innovative alternatives to the traditional approaches (KUANG et al., 2012; STIGLITZ et al., 2016; WEINDORF; BAKR; ZHU, 2014). These sensors are easy to use, portable, and deliver results very rapidly, while analyses are inexpensive, require minimal sample preparation and deliver accurate and reliable data, besides being environmentally friendly for uses both in the field (*in situ*) and in the laboratory (*ex situ*) (SILVA et al., 2021; STIGLITZ et al., 2016). Literature shows that these sensors (pXRF and Nix Pro^{TM}) coupled with other sensors (e.g., diffuse reflectance spectroscopy from visible to infrared wavelengths – Vis-NIR DRS) have the main purpose of complementing the traditional laboratory analyses of soils, towards the approach called "*hybrid laboratory*", and not replacing them permanently (DEMATTÊ et al., 2019; VISCARRA ROSSEL et al., 2011; SILVA et al., 2021).

Among the proximal sensors recently adopted by Soil Science community, pXRF is considered one of the most important and popular innovations of the last decade to characterize soils and other materials (e.g., parental material, sediments, plant tissues, organic compounds, and water) by quantifying their total elemental content (BORGES et al., 2020; FARIA et al., 2021; FERREIRA et al., 2021; MANCINI et al., 2019; PEARSON et al., 2016; RAVANSARI; WILSON; TIGHE, 2020). For this to happen, this equipment emits high energy X-rays that hit the atoms of the targeted material, displacing electrons from inner to outer orbits, forming a vacancy therein (empty space). In order to fill this vacancy, electrons move to the outer orbits releasing energy in the form of fluorescence (secondary X-rays). Since each element has a characteristic fluorescence along with its intensity, pXRF identifies

and quantifies various elements simultaneously (between Mg and U in the Periodic Table) with a wide range (ppm to %) (WEINDORF et al., 2014; SILVA et al., 2021). Thus, data provided by this sensor can be used to successfully estimate many soil properties under agronomic (LIU et al., 2021; TEIXEIRA et al., 2020), pedological (ACREE et al., 2020; STOCKMANN et al., 2016), environmental (BORGES et al., 2020; ROUILLON; TAYLOR, 2016) and geochemical (O'ROURKE et al., 2016) approaches.

Unlike pXRF, studies with the Nix Pro™ color sensor in soils are even more recent worldwide (past 5 years), but it is already being considered a promising and cost-effective approach for obtaining a reliable soil color assessment compared to conventional methods (e.g., high-cost colorimeters and Munsell Soil Color Charts), once the determination of color becomes less subjective (MANCINI et al., 2020; MORITSUKA et al., 2019; STIGLITZ et al., 2016). This sensor is inexpensive, light-weighted, portable, equipped with a light-emitting diode (LED), extremely fast reading (1-2 seconds), can be connected to smartphones via Bluetooth®. It provides quantitative measurements of the light reflected from the samples in the 380 to 700 nm spectral range that is used to automatically calculate the soil color in various color systems (e.g., RGB, XYZ, CIELAB, CIELCH, and CMYK), which can be directly used for statistical analyses (STIGLITZ et al., 2016; STIGLITZ et al., 2017). Thus, in addition to the characterization of soil color, the data obtained by Nix Pro™ have been used to predict soil organic matter and/or soil organic carbon, as well as soil total nitrogen (MIKHAILOVA et al., 2017; STIGLITZ et al., 2017; MUKHOPADHYAY et al., 2020). With the advances of research, this sensor is already being used in the rapid and inexpensive assessment of total soil iron (JHA et al., 2021), as well as in the prediction of different soils orders and suborders in Brazilian tropical soils (ANDRADE et al., 2020).

Conversely, some factors can negatively influence the performance of these sensors and, therefore, deserve full attention before and during the analyses. Studies have reported that pXRF field data for many elements differ from those obtained under laboratory conditions, i.e, in air-dried fine earth (commonly used for scans with pXRF) (DIJAIR et al., 2020; SILVA et al., 2018; STOCKMANN et al., 2016). This mainly happens due to differences in moisture, organic matter, heterogeneity, particle size distribution, and soil structure (RAVANSARI; WILSON; TIGHE, 2020; SILVA et al., 2018; SILVA et al., 2021). This may constrain the use of field data and little has been done aimed at evaluating the effect of different sample preparation conditions on the correspondence between pXRF and traditional laboratory methods for a large number of chemical elements, especially in tropical soils.

For Nix Pro™, soil color analyses are mainly influenced by the moisture content of the samples, which consequently affects the prediction models (ANDRADE et al., 2020; RAEESI et al., 2019; STIGLITZ et al., 2017). Conversely, Stiglitz et al. (2016) observed that moisture did not cause strong changes in color of temperate soils analyzed via Nix Pro™. Thus, there is a need to further investigate the influence of moisture on Nix Pro™ analyses of tropical soils to standardize the best moisture condition that favors reliable prediction of soil attributes related to color. Despite the potential applications of this new sensor, there are few studies exploring it in tropical countries, where some very weathered-leached soils, compared with soils from temperate regions, tend to present colors that do not reflect the SOM content. This is a consequence of the presence of Fe-oxide minerals in these soils, mainly containing hematite in great amounts, a mineral that presents high pigmenting power (RESENDE et al., 2014). As such, further studies are required to assess the efficiency of this sensor under tropical conditions.

Studies involving the individual use of proximal sensors in the prediction of chemical, physical, and morphological attributes are being well reported worldwide, from temperate to tropical regions (BENEDET et al., 2020; TAVARES et al., 2021; CHATTERJEE et al., 2021). However, sometimes one sensor alone might be limited to provide accurate predictions for all targeted soil properties (LIMA et al., 2019; SILVA et al., 2017; SILVA et al., 2019). In this way, the combination of data from multiple sensors becomes a viable and promising alternative to provide robust soil information to accurately predict soil properties in a fast way and with high accuracy (ANDRADE et al., 2020; BENEDET et al. 2020; O'ROURKE et al., 2016; WAN et al., 2019, 2020). However, investigations on the combined use of the data delivered via pXRF and Nix Pro™ for the prediction of agronomic attributes and/or a more geochemical approach in tropical soils are still few and recent compared to soils from temperate regions.

Given the above, pXRF and Nix Pro™ offer a valuable alternative to traditional laboratory analyses; however, the data from these proximal sensors can be affected by different scanning conditions that attract the use of new approaches for the reliable prediction of agronomic and geochemical attributes in tropical soils. Thus, the objectives of this thesis were:

i) chapter one – to develop accurate prediction models for pH (H_2O), sum of bases (SB), cation exchange capacity (CEC) at pH 7.0 and base saturation (BS) from pXRF data acquired from highly variable soils representative of the Brazilian Coastal Plains biome;

ii) chapter two - (i) to predict SOM content through data obtained via Nix Pro™ and pXRF in tropical soils, also evaluating the efficiency of these sensors to provide data to be used separately or combined via three machine learning algorithms: stepwise multiple linear regression (SMLR), support vector machine (SVM) with linear kernel and random forest (RF) in two soil moisture conditions. In addition, the information on soil horizons from which the samples have been collected will be used as auxiliary input data for model generation;

iii) chapter three - (i) to evaluate the impact of different scanning conditions of soil samples on pXRF results, as follows: in the field (F), post-field (PF), air-dried, disaggregated and sieved at 2 mm (ADFE), ADFE followed by grinding (M), and M sieved at 150 μm (MS); the five scanning conditions were used to determine the total elemental contents of Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V and Zr; and (ii) to determine the relationship between the 12 elements obtained with pXRF in these five scanning conditions and the traditional acid-digestion-based method (USEPA 3051a/ICP-OES) via linear regression models and their validation.

iv) chapter four – (i) to characterize the elemental content of highly heterogeneous tropical soils with pXRF; and (ii) to develop and validate prediction models for 28 elements traditionally obtained by the USEPA 3051a method based on data generated by pXRF through machine learning algorithms (random forest - RF and support vector machine - SVM) and simple linear regressions (RLS) and multiple (stepwise multiple linear regression - SMLR).

Our hypothesis is that general and robust models can be generated using machine learning algorithms from data obtained via pXRF and Nix Pro™, combined or individually, capable of predicting attributes that indicate both soil agronomic and geochemical properties quickly and accurately, even considering highly heterogeneous tropical soils.

2. REFERENCES

ACREE, A. et al. Soil classification in Romanian catenas via advanced proximal sensors. **Geoderma**, v.377, p.114587, 2020.

ANDRADE, R. et al. Tropical soil order and suborder prediction combining optical and X-ray approaches. **Geoderma Regional**, v.23, p.e00331, 2020.

BENEDET, L. et al. Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms. **Catena**, v.197, p.105003, 2021.

BENEDET, L. et al. Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. **Geoderma**, v.376, p.114553, 2020.

BORGES, C. S. et al. Comparison of portable X-ray fluorescence spectrometry and laboratory-based methods to assess the soil elemental composition: applications for wetland soils. **Environmental Technology & Innovation**, v.19, p.100826, 2020.

CHATTERJEE, S. et al. Characterization of field-scale soil variation using a stepwise multi-sensor fusion approach and a cost-benefit analysis. **Catena**, v.201, p.105190, 2021.

CHEN, M.; MA L. Q. Comparison of four USEPA digestion methods for trace metal analysis using certified and florida soils. **Journal of Environmental Quality**, v.27, p.1294–1300, 1998.

DEMATTÊ J. A. et al. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. **Geoderma**, v.337, p.111-121, 2019.

DIJAIR, T. S. B. et al. Correcting field determination of elemental contents in soils via portable X-ray fluorescence spectrometry. **Ciência e Agrotecnologia**, v. 44, p. e002420, 2020.

FARIA, A. J. G. et al. Elemental analysis of biochar-based fertilizers via portable X-ray fluorescence spectrometry. **Environmental Technology & Innovation**, v.23, p.101788, 2021.

FERREIRA, G. W. D. et al. Assessment of iron-rich tailings via portable X-ray fluorescence spectrometry: the Mariana dam disaster, southeast Brazil. **Environmental Monitoring and Assessment**, v.193, p.1–19, 2021.

JHA, G. et al. Rapid and inexpensive assessment of soil total iron using Nix Pro color sensor. **Agricultural & Environmental Letters**, v. 6, n. 3, 2021.

KUANG B. et al. Sensing soil properties in the laboratory, in situ, and on-line. A review. **Advances in Agronomy**, v.114, p.155-223, 2012.

LIMA, T. M. et al. Elemental analysis of Cerrado agricultural soils via portable X-ray fluorescence spectrometry: Inferences for soil fertility assessment. **Geoderma**, v.353, p.264–272, 2019.

LIU, Y. et al. Prediction of multiple soil fertility parameters using VisNIR spectroscopy and pXRF spectrometry. **Soil Science Society of America Journal**, v.85, p.591–605, 2021.

MANCINI, M. et al. Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. **Geoderma**, v.337, p.718–728, 2019.

MANCINI, M. et al. From sensor data to Munsell color system: Machine learning algorithm applied to tropical soil color classification via Nix™ Pro sensor. **Geoderma**, v. 375, p. 114471, 2020.

MIKHAILOVA, E. A. et al. Predicting soil organic carbon and total nitrogen in the Russian Chernozem from depth and wireless color sensor measurements. **European Journal of Soil Science**, v.50, p.1414–1419, 2017.

MORITSUKA, N. et al. Comparison of visual and instrumental measurements of soil color with different low-cost colorimeters. **Journal of Soil Science and Plant Nutrition**, v.65, p.605–615, 2019.

MUKHOPADHYAY, S. et al. Assessment of heavy metal and soil organic carbon by portable X-ray fluorescence spectrometry and NixPro™ sensor in landfill soils of India. **Geoderma Regional**, v.20, p.e00249, 2020.

NASCIMENTO, E. S.; TENUTA FILHO, A. Chemical waste risk reduction and environmental impact generated by laboratory activities in research and teaching institutions. **Brazilian Journal of Pharmaceutical Sciences**, v.46, p.187–198, 2010.

O'ROURKE, S.M. et al. An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. **Geoderma**, v.279, p.31–44, 2016.

PEARSON, D. et al. Water analysis via portable X-ray fluorescence spectrometry. **Journal of Hydrology**, v.544, p.172–179, 2016.

RAEESI, M. et al. Prediction of soil organic matter using an inexpensive colour sensor in arid and semiarid areas of Iran. **Soil Research**, v.57, p.276–286, 2019.

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

RESENDE, M. et al. Pedologia: Base para distinção de ambientes (6th ed.). Editora UFLA, p. 56, 2014.

ROUILLON, M.; TAYLOR, M. P. Can field portable X-ray fluorescence (pXRF) produce high quality data for application in environmental contamination research?. **Environmental Pollution**, v.214, p.255–264, 2016.

SILVA, E. A. et. al. Advances in tropical soil characterization via portable X-ray fluorescence spectrometry. **Pedosphere**, v.29, p.468–482, 2019.

SILVA, S. H. G. et al. Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). **Ciência e Agrotecnologia**, v. 41, p. 648-664, 2017.

SILVA, S. H. G. et al. Tropical soils characterization at low cost and time using portable X-ray fluorescence spectrometer (pXRF): effects of different sample preparation methods. **Ciência e Agrotecnologia,** v.42, p.80–92, 2018.

SILVA, S. H. G. et al. pXRF in tropical soils: methodology, applications, achievements and challenges. **Advances in Agronomy**, v.167, p.62, 2021.

SILVA, Y. J. A. B.; NASCIMENTO, C. W. A.; BIONDI, C. M. Comparison of USEPA digestion methods to heavy metals in soil samples. **Environmental Monitoring and Assessment**, v.186, p.47–53, 2014.

STIGLITZ, R. et al. Evaluation of an inexpensive sensor to measure soil color. **Computers and Electronics in Agriculture**, v.121, p.141–148, 2016.

STIGLITZ, R. et al. Using an inexpensive color sensor for rapid assessment of soil organic carbon. **Geoderma**, v.286, p.98–103, 2017.

STOCKMANN, U. et al. Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. **Catena**, v. 139, p. 220–231, abr. 2016.

TAVARES, T. R. et al. Combined use of VIS-NIR and XRF sensors for tropical soil fertility analysis: Assessing different data fusion approaches. **Sensors** (Switzerland), v.21, n.1, p.1–23, 2021.

TEIXEIRA, A. F. S. et al. Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry. **Geoderma**, v.361, p.114132, 2020.

TEIXEIRA, P. C. et al. **Manual de métodos de análise do solo**. 3ª Edição Revisada e Ampliada, Brasília, DF, p.574, 2017.

VISCARRA ROSSEL, R. A. et al. Proximal soil sensing. An effective approach for soil measurements in space and time. **Advances in Agronomy**, v.113, p.243–291, 2011.

WAN, M. et al. Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals. **Process Safety and Environmental Protection**, v.132, p.73-81, 2019.

WAN, M. et al. Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. **Geoderma**, v. 363, p.14163, 2020.

WEINDORF, D. C.; CHAKRABORTY, S. Portable X-ray fluorescence spectrometry analysis of soils. **Soil Science Society of America Journal**, v.84, p.1384–1392, 2020.

WEINDORF, D. C.; BAKR, N.; ZHU, Y. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications. **Advance in Agronomy**, v.128, p.45, 2014.

SECOND PART – ARTICLES

ARTICLE 1 - Soils of the Brazilian Coastal Plains biome: prediction of chemical attributes via portable X-ray fluorescence (pXRF) spectrometry and robust prediction models.

Álvaro José Gomes de Faria[A], Sérgio Henrique Godinho Silva[A], Leônidas Carrijo Azevedo Melo[A], Renata Andrade[A], Marcelo Mancini[A], Luiz Felipe Mesquita[B], Anita Fernanda dos Santos Teixeira[A], Luiz Roberto Guimarães Guilherme[A], and Nilton Curi[A]

[A] Federal University of Lavras, Soil Science Department, PO Box 3037, 37200-000, Lavras – MG, Brazil.
[B] Suzano Papel e Celulose, Espírito Santo, ES, Brazil.

Abstract

Portable X-ray fluorescence (pXRF) spectrometry has been successfully used for soil attribute prediction. However, recent studies have shown that accurate predictions may vary according to soil type and environmental conditions, motivating investigations in different biomes. Hence, this work attempted to accurately predict soil pH, sum of bases (SB), cation exchange capacity (CEC) at pH 7.0 and base saturation (BS) using pXRF-obtained data with high variability and robust prediction models in the Brazilian Coastal Plains biome. A total of 285 soil samples were collected to generate prediction models for A (n = 123), B (n = 162) and A+B (n = 285) horizons through stepwise multiple linear regression, support vector machine with linear kernel (SVM) and random forest. Data were divided into calibration (75%) and validation (25%) sets. Accuracy of the predictions was assessed by coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and residual prediction deviation (RPD). The A+B horizons dataset had optimal performance, especially for SB predictions using SVM, achieving R^2 = 0.82, RMSE = 1.02 cmolc dm^{-3}, MAE = 1.17 cmolc dm^{-3} and RPD = 2.33. The most important predictor variable was Ca. Predictions using pXRF data were accurate especially for SB. Limitations of the predictions caused by soil classes and environmental conditions should be further investigated in other regions.

Keywords: modelling, soil analysis, soil fertility, tropical soils.

1. Introduction

The determination of soil chemical attributes is fundamental for agricultural management, pedological and geochemical mapping and for management of the environment (Zhang and Hartemink 2020). During past decades, soil chemical analyses have undergone much evolution, from the utilization of low precision colorimetric methods (Clark and Axley 1955; Fliermans and Brock 1973) to sophisticated methods like inductively coupled plasma mass spectroscopy and atomic absorption spectroscopy (AAS), which are accurate and now greatly utilised (Soltanpour *et al.* 1996; Weindorf and Chakraborty 2016). The standard methods for soil chemical analyses are essential for soil characterisation. However, depending on the required resolution of the maps to be created, the number of samples and the size of the study area, traditional analyses might be sluggish, expensive and demand large amounts of chemical reagents that need meticulous care concerning their disposal (Ribeiro *et al.* 2017).

The increasing use of proximal sensors worldwide offers a practical solution to overcome these issues (Viscarra Rossel *et al.* 2011; Kuang *et al.* 2012) by complementing or even substituting traditional methods (Viscarra Rossel *et al.* 2011). Similar to other prevailing proximal sensor techniques, portable X-ray fluorescence (pXRF) spectrometry has increasingly attracted soil scientists' attention in recent years (Mancini *et al.* 2019; Nawar *et al.* 2019; Zhang and Hartemink 2020; Andrade *et al.* 2020a, 2020b; Teixeira *et al.* 2020; Wan *et al.* 2020). The pXRF analysis has innumerable advantages over traditional laboratory methods, including (i) minimal sample preparation is required, (ii) chemical reagents are not necessary, (iii) allows in situ and ex situ analyses, (iv) fast, reliable and economic acquirement of data and (v) in a single scan, multiple attributes can be measured or predicted with high accuracy (Sarkhot *et al.* 2011; Viscarra Rossel *et al.* 2011; Weindorf *et al*. 2012; Wang *et al.* 2015).

Across the world, pXRF is becoming a reliable tool for diverse branches of science, including soil science, by delivering promising results, as reported by a myriad of studies regarding soil characterisation applied to agricultural and environmental assessment (Jang 2010; Zhu *et al.* 2011; Weindorf *et al.* 2012; Towett *et al.* 2016; Chakraborty *et al.* 2017; Hu *et al.* 2017; Pearson *et al.* 2017; Dijair *et al.* 2020). Data acquired by pXRF can be likewise utilised in soil attribute predictions with significant accuracy using statistical modelling techniques (Weindorf *et al.* 2013; Chakraborty *et al.* 2017; Hu *et al.* 2017; Rawal *et al.* 2019;

Xu *et al.* 2019; Silva *et al.* 2020; Zhang and Hartemink 2020). However, such efforts are yet incipient in Brazil compared with other countries, especially those of temperate regions, and recent studies have shown variations in results when assessing soils from different regions of Brazil (Silva *et al.* 2016, 2017; Ribeiro *et al.* 2018; Teixeira *et al.* 2018; Costa *et al.* 2019; Mancini *et al*. 2019; Andrade *et al.* 2020a, 2020b; Tavares *et al.* 2020; Teixeira *et al.* 2020). This means that the modelling must be performed at the regional or local level, motivating further investigations under these different environmental conditions, for example the Brazilian Coastal Plains region (~20 Mha in Brazil).

Thus, the objective of this work was to develop accurate prediction models for pH, sum of bases (SB), cation exchange capacity (CEC) at pH 7.0 and base saturation (BS) from pXRF data acquired from highly variable soils representative of the Brazilian Coastal Plains region. This study hypothesizes that pXRF-obtained data can be successfully used to rapidly and accurately predict fertility properties of soils from the Brazilian Coastal Plains.

2. Material and Methods

2.1 Study area

The soil samples were collected from the Brazilian Coastal Plains biome, in the states of Espírito Santo, Bahia and Minas Gerais (Fig. 1), representing 1 199 918 km^2 of the country. This biome covers ~20 Mha on the Brazilian coast. The parent materials are Tertiary sediments of the Barreiras Formation (Corrêa *et al.* 2008a; Gomes *et al.* 2017) and the predominant soil mineralogy is represented by quartz in the sand and silt fractions and kaolinite in the clay fraction (Carvalho Filho *et al.* 2013; Ker *et al.* 2017). This soil mineralogy is very uniform as a result of a very long period of sediment deposition on a flat landscape (Resende *et al.* 2011, 2019).

According to the Köppen classification system, the sampling region is classified as Aw (hot weather with summer rain) and Af (hot climate, without dry season), with altitude range of 20–320 m and annual rainfall of 970–1560 mm (Siqueira *et al.* 2004; Gomes *et al.* 2017). All the sampled areas were under eucalyptus cultivation. The agronomic management of this crop in all areas, in order to obtain high productivity, includes application of soil fertilisers and conditioners, according to the needs of the crop.
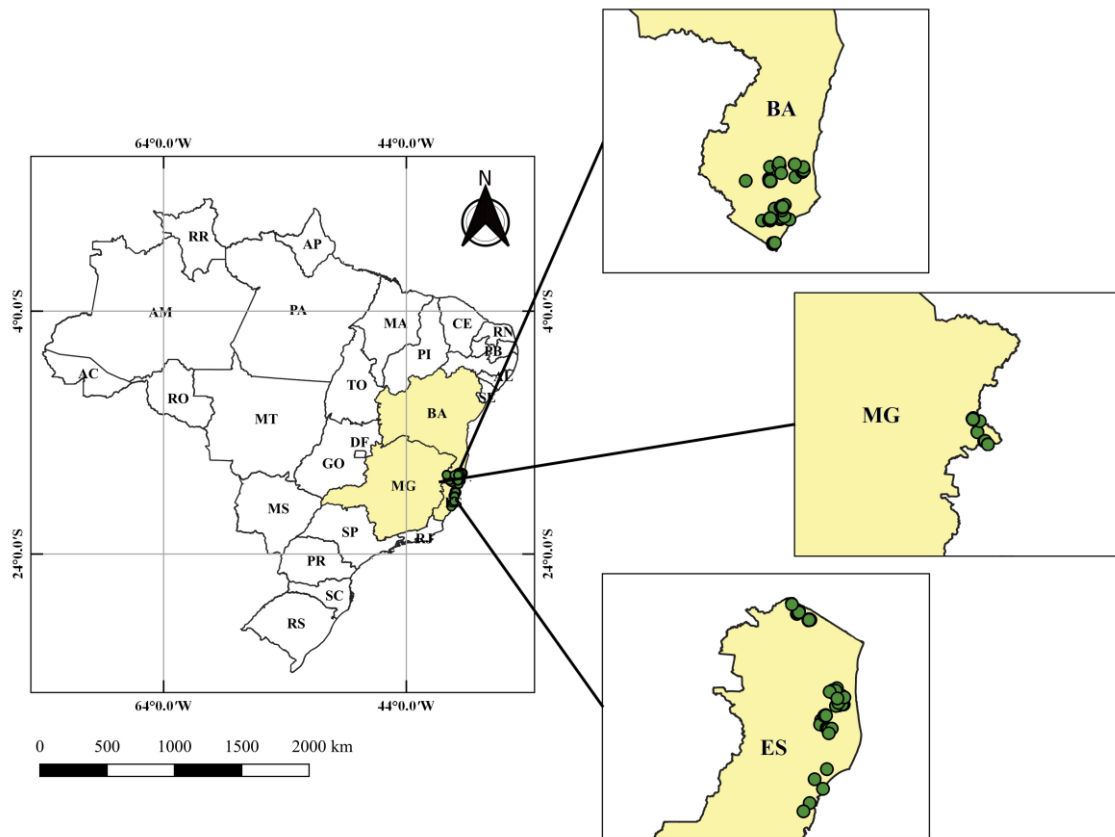
Fig. 1. Study area and soil sampling locations across the Brazilian Coastal Plains comprising the states of Minas Gerais (MG), Bahia (BA) and Espírito Santo (ES), representing 1 199 918 km2 of the country.

2.2 Soil sampling and laboratory analyses

Approximately 500 g of soil was collected from A (0–20 cm depth) and B (80–100 cm depth) horizons from 121 soil profiles comprising the most representative soil classes found in the Brazilian Coastal Plains region – Ultisols, Oxisols, Spodosols and Entisols, classified according to the US Soil Taxonomy (Soil Survey Staff 2014). In total, 285 samples were collected: 123 from A horizons and 162 from B horizons. Subsequently, samples underwent physical and chemical laboratory analyses and were classified accordingly, with the support of field soil morphology data.

Analytical determinations were performed in air-dried, ground samples, sieved through a 2-mm screen (air-dried fine earth). Next, the following chemical attributes were determined: pH in water (1 : 2.5, soil : water), exchangeable contents of Ca^{2+} and Mg^{2+} extracted by 1 mol L^{-1} KCl, available K extracted by Mehlich^{-1} solution and potential acidity ($H^+ + Al^{3+}$) after extraction with 0.5 mol L^{-1} calcium acetate at pH 7.0. These analyses

followed the recommendations for Brazilian soils (Teixeira *et al*. 2017). The pH analyses were performed using a Digimed® Double Channel pH meter model DM-23-DC. The contents of Ca^{2+} and Mg^{2+} were assessed by AAS (Perkin Elmer® model AAnalyst 800). Contents of K were determined by flame photometer and $H^{+} + Al^{3+}$ by titration with NaOH. Results of the aforementioned analyses were utilized to calculate the SB (cmolc dm^{-3}), CEC at pH 7.0 (cmolc dm^{-3}) and BS.

2.3 pXRF analyses

Analyses were performed with a Bruker® model S1 Titan LE pXRF. This equipment is composed of a Rh X-ray tube of 4 W, 15–50 keV, 5–100 mA, and a silicon drift detector (SDD) with a resolution of <145 eV, allowing for the detection of all elements ranging from Mg to U in the periodic table (Weindorf *et al.* 2014). The GeoChem software was utilized in Trace dual soil mode (60 s scan) and all 285 samples were scanned in triplicate and the mean obtained was used. The chemical elements and oxides utilized in this work follow: Al, Ca, Cl, Cu, Fe, K, Mn, Nb, P, S, Si, Sr, Ti, Zn and Zr.

To assess the equipment's accuracy, the reference materials 2710a (Montana soil I) and 2711a (Montana soil II), certified by the National Institute of Standards and Technology, and the standard sample provided by the pXRF manufacturer (check sample – CS) were analyzed. The results were used to calculate the recovery value for each element (Table 1), as follows: recovery % = 100 x (obtained content/content from reference samples). Zero indicates the absence of reference values for an element or no elemental content detected by pXRF.

Table 1. Recovery values for elemental contents obtained via pXRF calculated using reference samples certified by the National Institute of Standards and Technology (NIST) and provided by the pXRF manufacturer (check sample – CS).

Sample	Al	Ca	Cl	Cu	Fe	K	Mn	Nb	P	S	Si	Sr	Ti	Zn	Zr
								%							
2710a	34	80	0	84	68	68	69	0	128	0	43	102	93	101	102
2711a	65	46	0	71	70	50	60	0	465	0	48	93	68	85	0
CS	92	0	0	99	95	89	88	0	0	0	45	0	0	0	0

2.4 Statistical analyses

Prediction models were created for pH, SB, CEC and BS using datasets composed of A (n = 123), B (n = 162) and A+B (n = 285) horizons data using three different algorithms. Each dataset was randomly divided into calibration (75%) and validation (25%) sets. Hence, 88, 115 and 201 samples were used for modelling and 35, 47 and 84 samples were utilized for validation (for A, B and A+B horizons datasets respectively). In addition, the A+B dataset was created keeping samples from the A and B horizons of the same profile in the same dataset (modelling or validation). This process maintained true independence of the data during the validation procedure.

To create prediction models from pXRF data, all obtained elemental contents were centralized (centralized elemental content = (x - mean(x))/s.d.(x)), and scaled (scaled elemental content = x – mean(x)), where x is the original elemental content. Then, three methods were utilized: stepwise multiple linear regression (SMLR), support vector machine with linear kernel (SVM) and random forest (RF). Analyses were performed using R software (version 3.6.1) (R Core Team 2018) and the R package 'caret' (Kuhn 2008).

The SMLR is a semi-automatic model construction process that successively adds or removes variables based on the F-test or the t-test of its estimated coefficients (Khaledian *et al.* 2017; Mohamed *et al.* 2018). The SVM uses kernel functions (Vapnik 1995) to project data in a hyperspace where complex non-linear patterns can be represented (Karatzoglou *et al*. 2008). In the new hyperspace, SVM attempts to construct an ideal hyperplane that separates classes to predict data with minimum empirical risk (Chakraborty *et al.* 2019). The RF algorithm creates several trees with different predictor variables into each tree, generating a final model composed of several trees (a forest) (Breiman 2001; Liaw and Wiener 2002), and has been successfully utilized by many soil scientists in different parts of the world (Silva *et al.* 2017; Chakraborty *et al.* 2019; Rawal *et al.* 2019; Andrade *et al.* 2020a, 2020b; Teixeira *et al.* 2020). In the current study, the following parameters were used: number of trees (ntrees) = 500 and number of predictors (mtry) = 1/3 of the number of predictor variables.

Validation of the predictions was performed by calculating the coefficient of determination (R^2) (Eqn 1), the root mean square error (RMSE) (Eqn 2), the mean absolute error (MAE) (Eqn 3) and the residual prediction deviation (RPD) (Eqn 4) (Chang *et al.* 2001). According to Viscarra Rossel *et al.* (2010), the R^2 values for soil attribute predictions are considered very good (>0.81), good (0.61–0.8), regular (0.41–0.6) and bad (<0.4). Chang *et al.* (2001) defined three categories for RPD concerning the quality and reliability of models:

excellent models (RPD > 2), models with moderate prediction ability (1.4 < RPD < 1.4) and models with no prediction ability (RPD < 1.4). Thus, models achieving greater R^2 and RPD and lower RMSE and MAE when comparing observed and predicted data are more capable of predicting pH, SB, CEC and BS accurately from pXRF data.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(X_{means,i} - X_{predi,i}\right)^2}{\sum_{i=1}^{n} \left(X_{mean} - X_{predi,i}\right)^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{means,i} - X_{predi,i}\right)^2}{n}} \tag{2}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \left|X_{means,i} - X_{predi,i}\right| \tag{3}$$

$$RPD = \frac{SD}{RMSE} \tag{4}$$

Where X_{means} is measured values, X_{predi} is values predicted by the models, X_{mean} is the mean of the values obtained via conventional laboratory analyses and n is the number of samples.

Additionally, the variables' importance was calculated for the best prediction model per soil attribute to be predicted using the 'rminer' package (Cortez 2016). The importance of variables is an informative and robust indicative of the relative importance of predictor variables for the models (Liaw and Wiener 2002). The higher the importance, the higher the relevance of a variable for the model (González *et al.* 2015). Moreover, to help understand the variables' importance, Pearson's correlation was calculated and presented as a correlogram between the pXRF data and the evaluated soil attributes, using the 'corrplot' package (Wei *et al.* 2017) of R software.

3. Results and discussion

3.1 Descriptive statistics for soil chemical attributes

The descriptive analyses of soil pH, SB, CEC and BS obtained through conventional laboratory methods for A, B and A+B horizons datasets are presented in Fig. 2. In general, a high dispersion and variability of the attributes obtained in each dataset, especially the A+B,

confirmed the high variability of the studied soils (Entisols, Spodosols, Ultisols and Oxisols). Although all the samples were collected under eucalyptus plantations, which had been receiving applications of fertilisers and conditioners, such contrasting soils present different soil texture, organic matter content, porosity and intensity of leaching, among other properties (Duarte *et al.* 2000; Moreau *et al.* 2006; Lima Neto *et al.* 2010; Resende *et al.* 2011, 2019; Andrade *et al.* 2020a, 2020b). This fact explains the variability of the soil properties evaluated in this study.
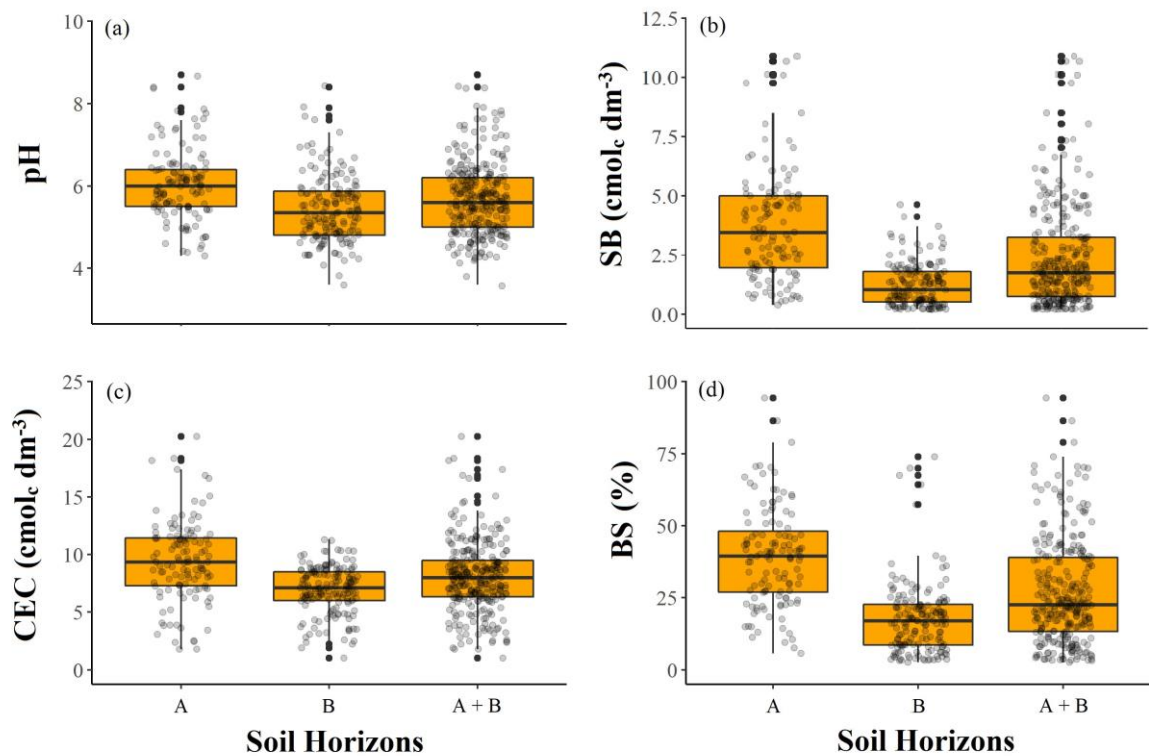
Fig. 2. Descriptive analyses of soil attributes (a) pH, (b) sum of bases (SB), (c) cation exchange capacity (CEC) at pH 7.0 and (d) base saturation (BS) for A, B and A+B horizons from soils of the Brazilian Coastal Plains, including the states of Espírito Santo, Bahia and Minas Gerais.

The A+B dataset comprises a relatively high quantity of data ($n = 285$) and, considering the observed variability (Fig. 2), might generate robust models that could be useful in conditions similar to those represented in these datasets. The A horizon samples had greater median values of soil pH, SB, CEC and BS than B horizon samples and greater dispersion and variability mainly for SB, CEC and BS, probably because of the greater

influence of fertilisers and conditioners in the surface horizon than in the subsurface horizon (Resende *et al.* 2014).

Symmetric distribution occurred for pH, SB and CEC, with mean values of 6.1, 5.4 and 5.7 for pH; 3.9, 1.2 and 2.4 cmolc dm^{-3} for SB; 9.5, 7.0 and 8.0 cmolc dm^{-3} for CEC for A, B and A+B horizons respectively. The BS data showed a negative asymmetric distribution for A and B horizons, and an asymmetric positive distribution for the A+B horizons. The mean values calculated for each horizon were 39.4, 17.6 and 27.0 respectively. Similar results were reported by Moreau *et al.* (2006), who obtained mean values of 5.0, 1.7, 6.8 and 28.1 for pH, SB, CEC and BS respectively, when studying six soil profiles of three distinct soil classes (Oxisol, Ultisol and Spodosol) under different land management practices in the Brazilian Coastal Plains. The soils of this region under natural conditions tend to be acidic and with low fertility, mainly in the subsurface horizon, with BS values commonly below 25% (Duarte *et al.* 2000; Moreau *et al.* 2006; Lima Neto *et al.* 2010; Resende *et al.* 2011, 2019), as also shown in Fig. 2. However, the soil fertility management contributed to the greater values determined for these soil properties in the areas of this study in comparison with the natural conditions.

Elemental contents obtained by pXRF also presented considerable variations (Table 2) due to the variability of soils and the changes caused by soil fertility management, which influenced the total elemental content obtained by pXRF (Table 2). For instance, the low mean contents of Fe (1.9% for A+B dataset) is related to the mineralogy of these soils dominated by kaolinite ($Al_2Si_2O_5(OH)_4$) in the clay fraction and quartz (SiO_2) in sand and silt fractions, as a consequence of the chemical composition of the Tertiary sediments that formed such soils (Resende *et al.* 2019). For comparison purposes, Pelegrino *et al.* (2019) reported mean Fe contents of 11.4% and 7.4% for soils derived from gabbro and gneiss respectively. Moreover, the low contents of nutrients such as Ca, K and P compared with soils from other Brazilian regions (Mancini *et al.* 2020), even under these cultivated areas that received applications of fertilisers and conditioners (Table 2), is related to the lack of nutrients also present in the crystalline structure of the minerals of these soils (Resende *et al.* 2011). These results agree with those of other authors who investigated soils of the Brazilian Costal Plains (Duarte *et al.* 2000; Lima *et al.* 2004; Moreau *et al.* 2006; Andrade *et al.* 2020a, 2020b).

Table 2. Descriptive statistics of elemental (mg kg^{-1}) contents obtained by pXRF from soil samples collected from the Brazilian Coastal Plains, including the states of Espírito Santo, Bahia and Minas Gerais.

CV, coefficient of variation; HA, A horizon; HB, B horizon; H A + B, A and B horizons combined

Parameters	Al			Ca			Cl			Cu		
	HA	HB	HA+B	HA	HB	HA+B	HA	HB	HA+B	HA	HB	HA+B
Minimum	1015.85	0.00	0.00	179.70	0.00	0.00	397.33	79.00	79.00	8.00	2.33	2.33
Maximum	98680.89	135900.10	135900.10	35546.25	3953.42	35546.25	1976.00	1478.33	1976.00	28.00	22.67	28.00
Mean	56573.24	77292.16	68350.31	4168.37	915.85	2319.56	803.78	734.37	764.32	13.76	11.16	12.28
Std. Dev.	17220.10	18279.61	20554.55	4687.42	725.02	3512.97	219.32	209.92	216.40	3.44	3.80	3.86
CV (%)[3]	30.44	23.65	30.07	112.45	79.16	151.45	27.29	28.58	28.31	24.98	34.01	31.44

Parameters	Fe			K			Mn			Nb		
	HA	HB	HA+B	HA	HB	HA+B	HA	HB	HA+B	HA	HB	HA+B
Minimum	466.33	311.67	311.67	424.69	343.90	343.90	0.00	0.00	0.00	0.00	0.00	0.00
Maximum	41742.33	54241.00	54241.00	1935.29	1441.15	1935.29	2127.33	1177.33	2127.33	38.67	253.33	253.33
Mean	13644.14	23209.89	19081.51	841.96	900.43	875.20	175.64	60.35	110.11	8.77	19.70	14.98
Std. Dev.	9025.86	12128.06	11870.47	254.04	196.69	224.71	296.27	127.74	224.12	6.72	19.69	16.39
CV (%)	66.15	52.25	62.21	30.17	21.84	25.68	168.68	211.66	203.54	76.69	99.97	109.42

Parameters	P			S			Si			Sr		
	HA	HB	HA+B	HA	HB	HA+B	HA	HB	HA+B	HA	HB	HA+B
Minimum	0.00	0.00	0.00	0.00	0.00	0.00	112170.44	114627.49	112170.44	0.00	0.00	0.00
Maximum	983.47	857.47	983.47	1312.67	565.67	1312.67	312834.89	357479.16	357479.16	79.67	49.67	79.67
Mean	255.37	33.14	129.05	181.90	7.82	82.95	215381.51	172390.26	190944.38	12.53	7.18	9.49
Std. Dev.	204.52	126.41	197.96	304.08	50.69	220.54	39481.22	41053.49	45607.74	11.60	6.55	9.44

| CV (%) | 80.09 | 381.52 | 153.40 | 167.17 | 648.17 | 265.87 | 18.33 | 23.81 | 23.89 | 92.58 | 91.13 | 99.48 |

Parameters	Ti			Zn			Zr		
	HA	HB	HA+ B	HA	HB	HA+B	HA	HB	HA+B
Minimum	1300.67	2013.67	1300.67	0.00	0.00	0.00	177.67	226.67	177.67
Maximum	16864.67	140707.33	140707.33	82.33333	37.67	82.33	1711.33	2245.33	2245.33
Mean	7627.99	11259.79	9692.38	13.40718	15.06	14.35	526.77	768.62	664.24
Std. Dev.	2717.50	10635.41	8398.98	9.713614	6.02	7.86	271.66	345.93	337.55
CV (%)	35.63	94.45	86.66	72.45083	39.99	54.78	51.57	45.01	50.82

3.2 Correlations between total elemental contents obtained by pXRF and soil attributes

The calculated correlations were low between the pXRF results and the soil properties evaluated (Fig. 3). The highest correlations were between pH and Mn (r = 0.42), Sr (r = 0.41), Ca (r = 0.40) and Cu (r = 0.32). Notably the correlation between pH and Ca, although low, is probably related to the application of limestone, a common agricultural practice in Brazilian soils to increase soil pH (Lopes and Guilherme 2016). Valadares *et al.* (1974) explains that, depending on the origin of the limestone, it may contain elements such as Mn as impurities in addition to Ca and Mg. Therefore, the application of this soil conditioner can increase the levels of not only Ca and Mg, but also Mn and other elements mainly in soil surficial layers. For instance, Sr, which was also highly correlated with pH, has been strongly correlated with Ca in other Brazilian agricultural soils (Teixeira *et al.* 2018). Regarding Ca, Silva *et al.* (2019) and Teixeira *et al.* (2020) found strong positive correlations between Ca determined by pXRF and pH for soils from different Brazilian regions. The same trend found for the soils of the Brazilian Costal Plains in our study is due to the lack of Ca-bearing minerals in these soils (Curi and Ker 2004), meaning that most of the Ca detected by pXRF is related to exchangeable Ca (Andrade *et al.* 2020a).
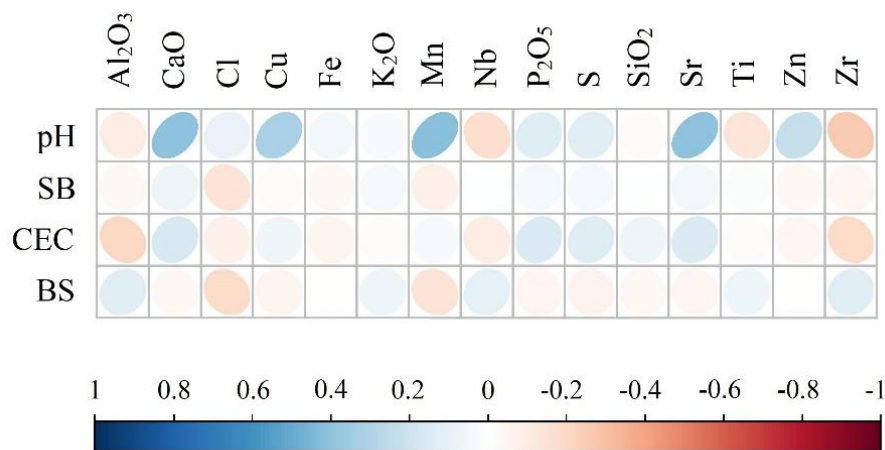
Fig. 3. Correlation between portable X-ray fluorescence spectrometry data and soil attributes (n = 285).

The elements used to determine SB, CEC and BS had weak correlations with SB, CEC and BS, even in the cases of K and Al, which are needed for calculating these attributes. These elements were present in the crystalline structure of the studied soils, i.e., their total content was not directly related to their exchangeable contents used for the calculations of SB,

CEC and BS. Similar weak correlations were found by Teixeira *et al.* (2020) between K and Al contents determined by pXRF and their exchangeable contents for Brazilian soils from various regions.

3.3 Predictions of soil chemical attributes

The R^2 values for all SMLR prediction models varied from 0.31 (pH) to 0.46 (SB) for A horizon data, from 0.20 (BS) to 0.43 (CEC) for B horizon data and from 0.29 (pH) to 0.47 (CEC) for the A+B data combined (Table 3). In general, the A+B dataset provided better R^2 values, mainly compared with the B horizon dataset.

Regarding CEC and SB using A+B data (Table 3), the equations generated by SMLR selected 10 (Al, Cu, Mn, Nb, P, S, Sr, Ti, Zn and Zr) and eight elements (Fe, Mn, P, S, Si, Sr, Zn and Zr) respectively, out of the 15 available for prediction of these attributes. Six out of these elements are nutrients for plants and were the most common in both equations: P, S, Fe, Mn, Zn and Cu. However, among them, only Al in its exchangeable form (Al^{3+}) is used for the determination of CEC through traditional laboratory methods. This reaffirms that prediction models based on pXRF-obtained data can create indirect relations between the elements and then predict soil chemical attributes, which makes pXRF a powerful tool for soil characterization, even if the direct correlations between pXRF variables and the soil attributes to be predicted are not strong (Fig. 3). Similar findings were reported by Sharma *et al.* (2014, 2015), Silva *et al.* (2017) and Teixeira *et al.* (2020) in predicting pH, SB, CEC and BS via pXRF. However, these indirect inferences may require that local specific prediction models be created instead of extrapolating the calibrated models to other regions with different conditions. This has been demonstrated by different studies achieving varying accuracies for the prediction of the same soil properties depending on the soils used for developing the prediction models (Zhu *et al.* 2011; Sharma *et al.* 2014, 2015; Silva *et al.* 2017; Rawal *et al.* 2019; Andrade *et al.* 2020a, 2020b; Benedet *et al.* 2020; Tavares *et al.* 2020).

From another perspective, the rapid quantification of multiple elements by pXRF and the robustness of prediction models for creating relations between such elements and the soil properties to be predicted raise new possibilities of investigating soil properties. For instance, Sr and Rb, elements less commonly used in soil-related studies than other elements (e.g., Fe, Si, Al, K, Ti and Ca) have been strongly correlated with exchangeable Ca^{2+} (Silva *et al.* 2018; Teixeira *et al.* 2018; Mancini *et al.* 2019) and clay content (Zhu *et al.* 2011; Silva *et al.* 2020) respectively, in various soils, helping their prediction.

Table 3. Equations obtained by SMLR to predict soil chemical attributes using A, B and A+B horizons' data obtained via pXRF from soils of the Brazilian Coastal Plains. SB, sum of bases (cmolc dm^{-3}); CEC, cation exchange capacity at pH 7.0 (cmolc dm^{-3}); BS, base saturation

Soil attributes	Equations	R²
-- Horizon A --		
pH	$6.04432 + 0.61486Ca + 0.15015K - 0.19529S$	0.31
SB	$3.8005 + 1.7816Ca + 0.5076Fe + 0.4702Nb + 0.3049S - 0.5318Zr$	0.46
CEC	$9.2595 + 0.5722Ca + 1.5720Sr + 1.0589Ti - 1.8911Zr$	0.38
BS	$39.435 + 3.543Al + 4.308Cu - 5.591Fe + 5.546Nb + 7.443Sr - 4.081Ti$	0.29
-- Horizon B --		
pH	$5.37826 + 0.15267Al + 0.21528Ca + 0.12677Cu + 0.18911Mn - 0.10099S + 0.16784Sr - 0.17460Ti$	0.27
SB	$1.22132 + 0.18046Ca + 0.28026Cu - 0.17974P + 0.50249Sr - 0.34782Ti - 0.20680Zr$	0.28
CEC	$6.9315 + 0.2476Ca - 0.6167Fe + 0.4056Mn - 1.3499Nb - 0.7192Si + 0.6686Sr + 0.2493Ti + 0.5082Zn$	0.43
BS	$17.5298 + 2.9937Ca + 5.0480Fe + 3.5637Nb - 2.7952P + 4.8709Si - 3.2811Ti$	0.20
-- Horizon A + B --		
pH	$5.70455 + 0.10940Cu + 0.18690Mn - 0.20665Nb - 0.14043P + 0.31032Sr$	0.29
SB	$2.4462 - 0.4781Fe + 0.3172Mn - 0.2489P + 0.4211S + 0.3058Si + 1.0874Sr + 0.4347Zn - 0.5185Zr$	0.46
CEC	$8.0109 - 0.3791Al - 0.4315Cu + 0.5152Mn - 1.2204Nb - 0.5692P + 0.4139S + 1.0471Sr + 0.3625Ti + 0.7056Zn - 0.6182Zr$	0.47
BS	$28.494 + 4.277Cu - 4.165Fe + 4.442Nb + 1.972S + 8.339Sr - 3.753Ti - 2.546Zr$	0.31

Validation results for all models built by the three studied algorithms (SMLR, SVM and RF) are presented in Table 4. In general, the models using the combined A+B dataset provided greater R^2 and RPD and lower RMSE and MAE for most evaluated soil attributes, especially SB (Table 4). Models created to predict pH using A and B horizon data separately had the lowest R^2 values, independently of the utilized algorithm. However, when the horizon data were combined, greater R^2 values were attained, although still lower than for the other attributes. For this soil property, considering all models, R^2 values ranged from 0.07 (SMLR, A horizon) to 0.40 (SVM, A+B horizons), and RPD values ranged between 0.98 (SMLR, A horizon) and 1.30 (SVM, A+B horizons) – models not considered reliable according to Chang *et al.* (2001) and Viscarra Rossel *et al.* (2010). In our study, R^2 values for soil pH predictions

were below those reported in the literature. Using pXRF-obtained data, Silva *et al.* (2017) concluded that models created by SMLR and RF could successfully predict soil pH by achieving R^2 values of 0.82 and 0.89 respectively, in soils from the Brazilian Cerrado biome. Sharma *et al.* (2014) also obtained satisfactory results for soil pH prediction, attaining maximum R^2 value of 0.83 when using SLMR to predict pH in soil samples from the USA. It is noteworthy that the latter two studies were conducted under different soil conditions.

The best values of R^2 and RPD were produced by models predicting SB, especially when using A+B data (Table 4). In this case, the best predictions were delivered by SVM with values of $R^2 = 0.82$, RMSE = 1.02 cmolc dm^{-3}, MAE = 1.17 cmolc dm^{-3} and RPD = 2.33 – classified as very good with excellent predictive capacity according to Chang et al. (2001) and Viscarra Rossel *et al.* (2010). Although existing studies have investigated the use of proximal sensors, including pXRF, for prediction of soil fertility attributes (Sharma *et al.* 2015; Silva *et al.* 2017; Rawal *et al.* 2019; Teixeira *et al.* 2020; Wan *et al.* 2020), very few evaluated SB prediction. Among them, the highest R^2 was 0.87 for the Cubist algorithm (Teixeira *et al.* 2020).

Similar behavior was observed for CEC and BS models (Table 4), with the best results for the A+B dataset. For both attributes, the prediction models created by the RF algorithm had the best R^2 results for CEC and BS (0.56 and 0.49 respectively) and intermediate values for RMSE (2.33 cmolc dm^{-3} and 12.68%), MAE (1.71 and 9.47) and RPD (1.38 and 1.33). These validation results showed that the models for CEC and BS prediction can be classified as regular with moderate prediction capability (Chang *et al.* 2001; Viscarra Rossel *et al.* 2010). Comparing these results with those obtained in other studies using pXRF for such predictions, for CEC, the values R^2 reported were between 0.20 (Teixeira et al. 2020) and 0.91 (Sharma et al. 2015) using RF and SMLR respectively. For BS, R^2 ranged from 0.28 (Teixeira et al. 2020) to 0.95 (Silva et al. 2017) for SMLR and RF respectively. However, the R^2 results of our study for SB, CEC and BS (Table 4) are intermediate compared to those in the literature, with the highest R^2 for SB through SVM ($R^2 = 0.82$). These results emphasise the importance of investigating the pXRF capability of predicting soil fertility properties for different soil conditions.

Table 4. Coefficient of determination (R^2), root mean square error (RMSE), mean absolute error (MAE) and residual prediction deviation (RPD) for the predictions of soil chemical attributes built via pXRF data obtained from soils of the Brazilian Coastal Plains. SB, sum of bases; CEC, cation exchange capacity at pH 7.0; BS, base saturation; SMLR, stepwise

multiple linear regression; SVM, support vector machine with linear kernel; RF, random forest

Soil Attributes	Horizon	Algorithm	Validation statistical parameters			
			R^2	RMSE	MAE	RPD
pH (H_2O)	A	SMLR	0.07	0.99	0.66	0.98
		SVM	0.09	0.95	0.69	1.02
		RF	0.22	0.85	0.63	1.14
	B	SMLR	0.30	0.85	0.70	1.18
		SVM	0.33	0.84	0.68	1.20
		RF	0.28	0.87	0.68	1.15
	A + B	SMLR	0.28	0.85	0.68	1.18
		SVM	0.40	0.77	0.60	1.30
		RF	0.26	0.87	0.67	1.16
SB ($cmol_c$ dm^{-3})	A	SMLR	0.43	2.07	1.51	1.32
		SVM	0.45	2.08	1.49	1.31
		RF	0.50	1.90	1.47	1.43
	B	SMLR	0.28	0.81	0.61	1.06
		SVM	0.29	0.79	0.64	1.08
		RF	0.27	0.73	0.59	1.18
	A + B	SMLR	0.73	1.29	1.04	1.85
		SVM	0.82	1.02	1.17	2.33
		RF	0.79	1.28	0.91	1.86
CEC ($cmol_c$ dm^{-3})	A	SMLR	0.34	3.18	2.42	1.22
		SVM	0.30	3.23	2.60	1.20
		RF	0.52	2.82	2.19	1.37
	B	SMLR	0.20	1.82	1.45	1.10
		SVM	0.24	1.75	1.32	1.14
		RF	0.38	1.57	1.07	1.28
	A + B	SMLR	0.42	2.45	1.90	1.31
		SVM	0.50	2.28	1.75	1.41
		RF	0.56	2.33	1.71	1.38
BS (%)	A	SMLR	0.08	16.27	12.92	1.00
		SVM	0.29	13.75	10.31	1.19
		RF	0.16	15.18	11.43	1.08
	B	SMLR	0.18	12.10	8.09	1.11
		SVM	0.18	12.22	7.70	1.10
		RF	0.16	12.24	7.73	1.10
	A + B	SMLR	0.35	14.50	10.60	1.16
		SVM	0.40	14.67	8.76	1.15
		RF	0.49	12.68	9.47	1.33

The variability in the accuracy of the predictions may be related to the peculiarities of the soils used in these different studies. Compared with the other studies, the soils of the Brazilian Coastal Plains present a mineralogy basically composed of large-sized kaolinite crystals and reduced amounts of iron oxides (Corrêa et al. 2008b), which may cause the diversity of results.

In general, the dataset comprising A+B data combined (i.e., the most diversified dataset) in tandem with SVM produced the best validation results, with high R^2 values and intermediate RMSE, MAE and RPD, especially for SB predictions (Table 4). Thus, this dataset will be further discussed. Figure 4 shows the scatter plots for pH, SB, CEC and BS predictions via the three studied algorithms using A+B data. The good performance of SVM has been previously documented (Were *et al.* 2015; Mancini *et al.* 2019; Silva *et al.* 2020; Wan *et al.* 2020). In our study, the worst performing model was SMLR (Table 4), as also reported in different studies developed in Brazil (Souza *et al.* 2016; Silva *et al.* 2017; Teixeira *et al.* 2020), including other types of regression (linear, polynomial and logarithmic) (Teixeira *et al.* 2018).
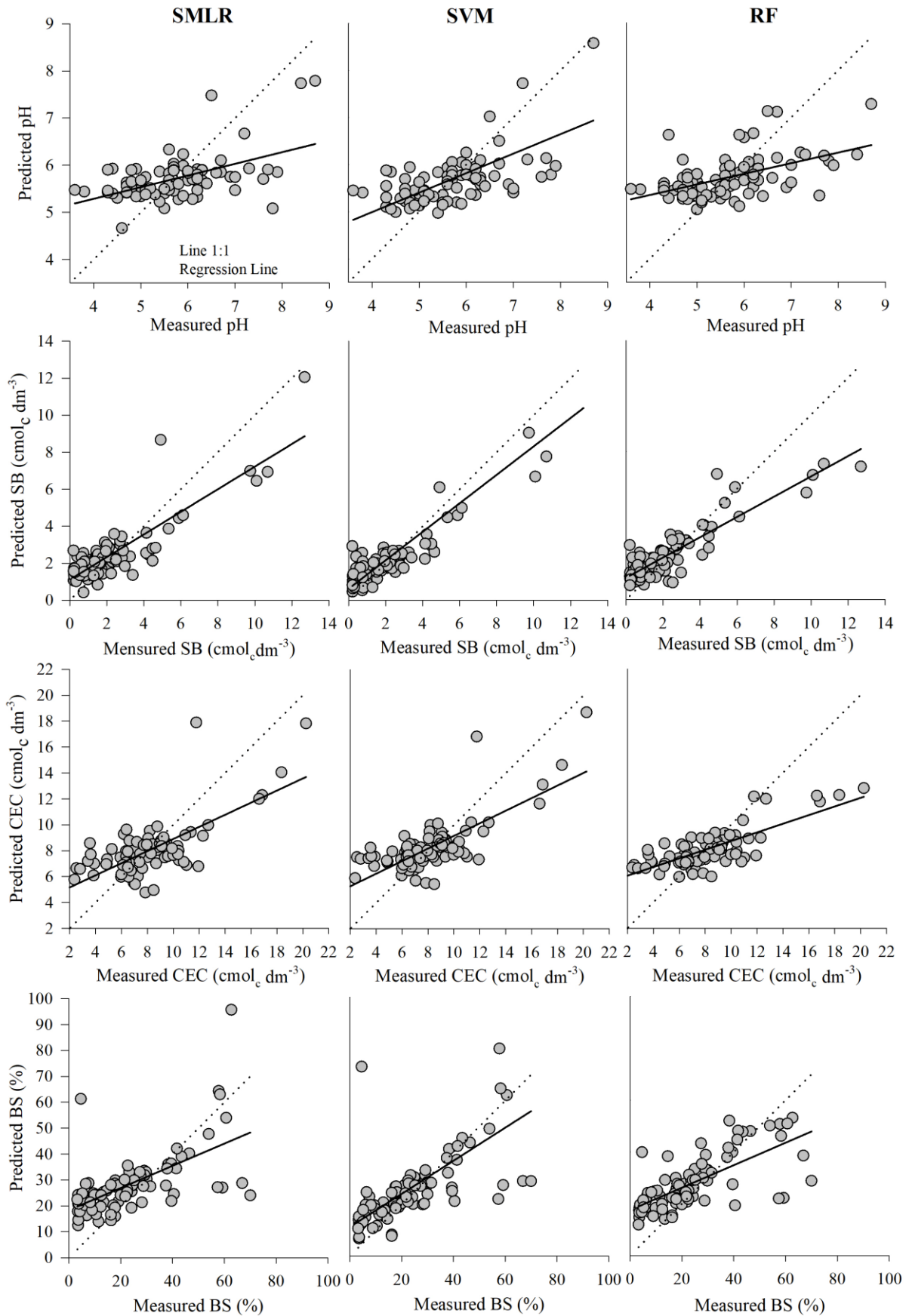
Fig. 4. Dispersion plots for pH, sum of bases (SB), cation exchange capacity (CEC) at pH 7.0 and base saturation (BS) predictions by the algorithms stepwise multiple linear regression

(SMLR), support vector machine (SVM) and random forest (RF) using data from A+B horizons combined obtained via pXRF from soils of the Brazilian Coastal Plains biome.

3.4 Importance of variables

The importance of variables for the best models for prediction of pH, SB, CEC and BS is shown in Fig. 5. In general, the most important variables for the different soil attributes prediction were those with the greatest correlations (Fig. 2) as discussed in *Correlations between total elemental contents obtained by pXRF and soil attributes*. According to the frequency that a variable was listed among the most important ones by the models, Ca was the most relevant for predictions. It was the most important variable for pH, SB and BS models and the second most important for CEC prediction (Mn was the most important in this case). As mentioned previously, there is a lack of Ca-bearing minerals in most soils of the Brazilian Coastal Plains (Curi and Ker 2004), meaning that most of the Ca detected by pXRF is related to the exchangeable Ca^{2+} used for the determination of SB, CEC and BS. Moreover, Ca is related to the limestone application on soils for increasing soil pH, thus, explaining its correlation with this soil property also observed in Fig. 2.

The obtained results agree with those of Andrade *et al.* (2020a, 2020b), who concluded that Ca followed by Si, Sr, S, K, Mn and P obtained by pXRF were the most important variables to predict total nitrogen, soil organic matter, CEC and exchangeable contents of Al^{3+}, Ca^{2+}, Mg^{2+}, K^+ and P in soils from the Brazilian Coastal Plains. In works developed in other Brazilian biomes, Ca obtained by pXRF was also noted as important in predictions of Al^{3+}, Ca^{2+}, K^+, Mg^{2+}, remaining P, pH, SB, CEC, BS and Al saturation (Silva *et al.* 2017; Teixeira *et al.* 2020). This highlights the importance of this variable to predict soil fertility attributes regardless of the studied biomes, because a lack of Ca-bearing minerals is not rare in most Brazilian soils, as previously discussed. The element Sr was as important in almost all models in the current work, confirming its importance for predictions. According to Gomes *et al.* (2017) and Mancini *et al.* (2019), Sr is stable in soils and might be related to the parent materials; Sr was also correlated with Ca content in Brazilian soils (Silva *et al.* 2018; Teixeira *et al.* 2018).
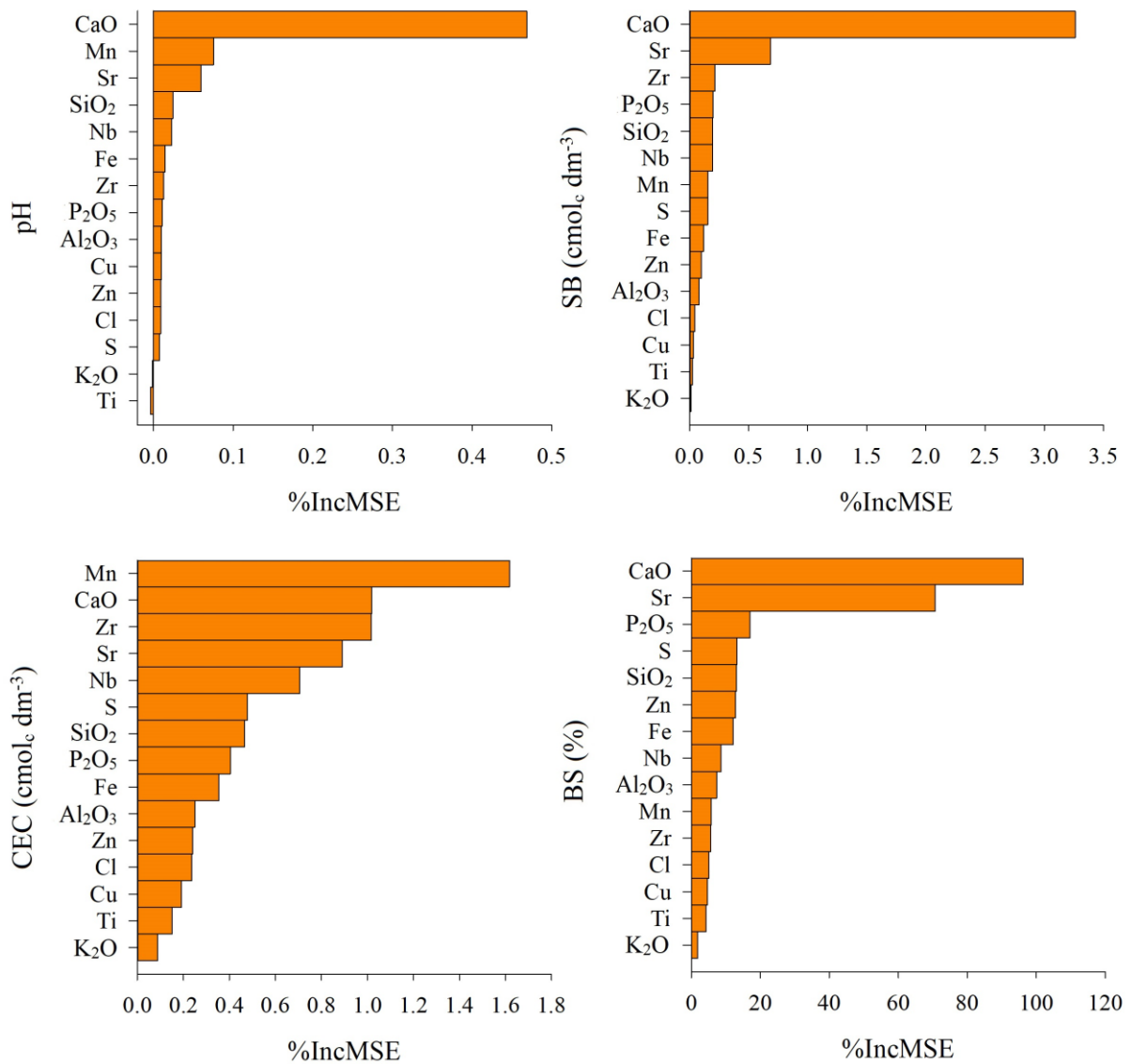
Fig. 5. Variables importance in decreasing order based on prediction results for soil chemical attributes via RF using A +B horizons data combined obtained by pXRF from soils of the Brazilian Coastal Plains. SB, sum of bases; CEC, cation exchange capacity; BS, base saturation.

The pXRF data have proved to be a promising tool in predicting soil fertility attributes for soils of the Brazilian Coastal Plains using A+B data combined, especially concerning SB. Among the studied models, SVM had superior performance and provided acceptable validation results for SB predictions. Under the current conditions, the data variability was not sufficient to deliver highly accurate predictions for pH, CEC and BS, particularly when using A and B horizon data separately.

Future studies employing combined data from different proximal sensors could help to improve predictive models for Brazilian Coastal Plains' soils, accounting for soil variability, sampling depth, land uses, parent material and environmental conditions. The number of studies exploring the combined use of proximal sensors in soil science applied to predict soil attributes is increasing (Wang *et al.* 2015; O'Rourke *et al.* 2016a, 2016b; Wan *et al.* 2019; Xu *et al.* 2019; Benedet *et al.* 2020), and existing results are very promising.

4. Conclusions

The SB was predicted with high accuracy ($R^2 = 0.82$, RMSE = 1.02 cmolc dm^{-3}, MAE = 1.17 cmolc dm^{-3} and RPD = 2.3) using SVM models via highly variable data obtained by pXRF, surpassing validation results of predictions delivered by SMLR and RF models. However, the models could not reliably predict pH, CEC and BS in soils of the Brazilian Coastal Plains. The most important predictor variable for pH, SB and SB was Ca, while for CEC, Mn was the most important variable followed by Ca. Further studies are encouraged in this line of research to improve the prediction capability for soils of the Brazilian Coastal Plains region by combining pXRF with different sensors. The use of proximal sensors is yet to be fully explored, but they are proving to offer novel, reliable soil assessment methods that are fast, accurate, low-cost and environmentally-friendly due to not generating residues.

5. Acknowledgments

References

Andrade R, Faria WM, Silva SHG, Chakraborty S, Weindorf D, Mesquita LF, Guilherme LRG, Curi N (2020a) Prediction of soil fertility via portable X-ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. *Geoderma* **357**, 113960. doi:10.1016/j.geoderma.2019.113960.

Andrade R, Silva SHG, Weindorf D, Chakraborty S, Faria WM, Mesquita LF, Guilherme LRG, Curi N (2020b) Geoderma Assessing models for prediction of some soil chemical

properties from portable X-ray fluorescence (pXRF) spectrometry data in Brazilian Coastal Plains. *Geoderma* **357**, 113957. doi:10.1016/j.geoderma.2019.113957.

Benedet L, Faria WM, Silva SHG, Mancini M, Demattê AM, Guilherme LRG, Curi N (2020) Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. *Geoderma* **376**, 114553. doi: 10.1016/j.geoderma.2020.114553.

Breiman L (2001) Random forests. *Machine Learning* **45**, 5–32.

Carvalho Filho A, Curi N, Fonseca S (2013) 'Avaliação informatizada e validada da aptidão silvicultural das terras dos tabuleiros costeiros Brasileiros para Eucalipto.' (Lavras, MG).

Chakraborty S, Li B, Weindorf DC, Deb S, Acree AP, Panda P (2019) Use of portable X-ray fluorescence spectrometry for classifying soils from different land use land cover systems in India. *Geoderma* **338**, 5–13. doi:10.1016/j.geoderma.2018.11.043.

Chakraborty S, Weindorf DC, Deb S, Li B, Paul S, Choudhury A, Ray DP (2017) Rapid assessment of regional soil arsenic pollution risk via diffuse reflectance spectroscopy. *Geoderma* **289**, 72–81. doi:10.1016/j.geoderma.2016.11.024.

Chang C, Laird DA, Mausbach MJ, Hurburgh CRJ (2001) Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal* **65**, 480–490.

Clark LJ, Axley JH (2005) Molybdenum Determination of Soils and Rocks with Dithiol. *Analytical Chemistry* **27**, 2000–2003. doi:10.1021/ac60108a046.

Corrêa MM, Ker JC, Barrón V, Fontes MPF, Torrent J, Curi N (2008a) Characterizing iron oxides from coastal and central plain soils. *Revista Brasileira de Ciência do Solo* **32**, 1017–1031. doi:10.1590/S0100-06832008000300011.

Corrêa MM, Ker JC, Barrón V, Torrent J, Fontes MPF, Curi N (2008b) Crystallographic properties of kaolinite soils from coastal tablelands, the Amazon and the great bay 'Reconcavo Baiano'. *Revista Brasileira de Ciência do Solo* **32**, 1857–1872. doi:10.1590/S0100-06832008000500007.

Cortez P, (2016) rminer: Data Mining Classification and Regression Methods. R package version 1.4.2. Available online at. https://CRAN.R-project.org/package=rminer/ (verified 23 Aug. 2018).

Costa YT, Ribeiro BT, Curi N, de Oliveira GC, Guilherme LRG (2019) Organic Matter Removal on Oxide Determination in Oxisols Via Portable X-ray Fluorescence. *Communications in Soil Science and Plant Analysis* **50**, 673–681. doi:10.1080/00103624.2019.1589479.

Curi N, Ker JC (2004) 'Levantamento pedológico de áreas da Aracruz Celulose S.A. nos estados do Espírito Santo, Bahia e Minas Gerais, e sua interpretação para o cultivo do eucalipto e para o ambiente em geral. (UFV, Lavras, MG; Viçosa, MG).

Dijair TSB, Silva FM, Teixeira AFS, Silva SHG, Guilherme LRG, Curi N (2020) Correcting field determination of elemental contents in soils via portable X-ray fluorescence spectrometry. *Ciência e Agrotecnologia* **44**, e002420. doi: 10.1590/1413-7054202044002420.

Duarte MN, Curi N, Pérez DV, Kämpf N, Claessen MEC (2000) Mineralogia, química e micromorfologia de solos de uma microbacia nos tabuleiros costeiros do Espírito Santo. *Pesquisa Agropecuaria Brasileira* **35**, 1237–1250. doi:10.1590/s0100-204x2000000600021.

Fliermans CB, Brock TD (1973) Assay of elemental sulfur in soil. *Soil Science* **115**, 120–122. doi:10.1097/00010694-197302000-00005.

Gomes JBV, Araújo Filho JC, Vidal-Torrado P, Cooper M, Silva EA, Curi N (2017) Cemented horizons and hardpans in the coastal tablelands of northeastern Brazil. *Revista Brasileira Ciência do Solo* 41.

González S, Herrera F, García S (2015) Monotonic random forest with an ensemble pruning mechanism based on the degree of monotonicity. *New Generation Computing* **33**, 367–388.

Hu B, Chen S, Hu J, Xia F, Xu J, Li Y, Shi Z (2017) Application of portable XRF and VNIR sensors for rapid assessment of soil heavy metal pollution. *PLoS ONE* **12**, 1–13. doi:10.1371/journal.pone.0172438.

Jang M (2010) Application of portable X-ray fluorescence (pXRF) for heavy metal analysis of soils in crop fields near abandoned mine sites. *Environmental Geochemistry and Health* **32**, 207–216. doi:10.1007/s10653-009-9276-z.

Karatzoglou A, Smola A, Hornik K (2008) Kernlab: Kernel-based Machine Learning Lab.

Ker JC, Schaefer CEGR, Romero RE, Corrêa MM (2017) Solos dos Tabuleiros Costeiros. 'Pedologia - Solos Dos Biomas Brasileiros, Sociedade Brasileira de Ciência do Solo'. pp. 467–492. (Viçosa, MG)

Khaledian Y, Brevik EC, Pereira P, Cerdà A, Fattah MA, Tazikeh H (2017) Modeling soil cation exchange capacity in multiple countries. *Catena* **158**, 194–200.

Kuang B, Mahmood HS, Quraishi MZ, Hoogmoed WB, Mouazen AM, van Henten EJ (2012) 'Sensing soil properties in the laboratory, in situ, and on-line. A review.' (Elsevier Inc.) doi:10.1016/B978-0-12-394275-3.00003-1.

Kuhn M (2008) Building Predictive Models in R Using the caret Package. *Journal of Statistical Software* **28**, 1–26. doi:10.18637/jss.v028.i05.

Liaw A, Wiener M (2002) Classification and regression by random forest. *R News* **2**, 18–22.

Lima Neto JA, Ribeiro MR, Corrêa MM, Souza-Júnior VS, Araújo Filho JC, Lima JFWF (2010) Atributos químicos, mineralógicos e micromorfológicos de horizontes coesos de latossolos e argissolos dos tabuleiros costeiros do estado de Alagoas. *Revista Brasileira de Ciencia do Solo* **34**, 473–486. doi:10.1590/s0100-06832010000200021.

Lima HV, Silva AP, Jacomine PTK, Romero RE, Libardi PL (2004) Identificação e caracterização de solos coesos no estado do Ceará. *Revista Brasileira de Ciencia do Solo* **28**, 467–476. doi:10.1590/s0100-06832004000300008.

Lopes AS, Guilherme LRG (2016) A career perspective on soil management in the Cerrado Region of Brazil. *Advances in Agronomy* **137**, 1–72.

Mancini M, Weindorf DC, Chakraborty S, Silva SHG, Teixeira AFS, Guilherme LRG, Curi N (2019) Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. *Geoderma* **337**, 718–728. doi:10.1016/j.geoderma.2018.10.026.

Mancini M, Silva SHG, Teixeira AFS, Guilherme LRG, Curi N (2020). Soil parent material prediction for Brazil via proximal soil sensing. *Geoderma* **22**, e00310. doi: 10.1016/j.geodrs.2020.e00310.

Mohamed ES, Saleh AM, Belal AB, Gad AA (2018) Application of near-infrared reflectance for quantitative assessment of soil properties. *Egyptian Journal of Remote Sensing and Space Science* **21**, 1–14. doi:10.1016/j.ejrs.2017.02.001.

Moreau AMSDS, Ker JC, Costa LM, Gomes FH (2006) Caracterização de solos de duas toposseqüências em tabuleiros costeiros do Sul da Bahia. *Revista Brasileira de Ciencia do Solo* **30**, 1007–1019. doi:10.1590/s0100-06832006000600010.

Nawar S, Delbecque N, Declercq Y, Smedt P, Finke P, Verdoodt A, Van Meirvenne M, Mouazen AM (2019) Can spectral analyses improve measurement of key soil fertility parameters with X-ray fluorescence spectrometry? *Geoderma* **350**, 29–39. doi:10.1016/j.geoderma.2019.05.002.

O'Rourke SM, Minasny B, Holden NM, McBratney AB (2016a) Synergistic Use of Vis-NIR, MIR, and XRF Spectroscopy for the Determination of Soil Geochemistry. *Soil Science Society of America Journal* **80**, 888. doi:10.2136/sssaj2015.10.0361.

O'Rourke SM, Stockmann U, Holden NM, McBratney AB, Minasny B (2016b) An assessment of model averaging to improve predictive power of portable vis-NIR and

XRF for the determination of agronomic soil properties. *Geoderma* **279**, 31–44. doi:10.1016/j.geoderma.2016.05.005.

Pearson D, Chakraborty S, Duda B, Li B, Weindorf DC, Deb S, Brevik E, Ray DP (2017) Water analysis via portable X-ray fluorescence spectrometry. *Journal of Hydrology* **544**, 172–179. doi:10.1016/j.jhydrol.2016.11.018.

Pelegrino MHP, Weindorf DC, Silva SHG, Menezes, MD, Poggere GC, Guilherme LRG, Curi N (2018) Synthesis of proximal sensing, terrain analysis, and parent material information for available micronutrient prediction in tropical soils. *Precision Agriculture* **20**, 746-766. doi: 10.1007/s11119-018-9608-z.

R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. https://www.r-project.org.

Rawal A, Chakraborty S, Li B, Lewis K, Godoy M, Paulette L, Weindorf DC (2019) Determination of base saturation percentage in agricultural soils via portable X-ray fluorescence spectrometer. *Geoderma* **338**, 375–382. doi:10.1016/j.geoderma.2018.12.032.

Resende M, Curi N, Ker JC, Rezende SB (2011) 'Mineralogy of Brazilian soils: interpretation and applications.' (Editora UFLA: Lavras)

Resende M, Curi N, Rezende SB, Corrêa GF, Ker JC (2014). Pedologia: Base para distinção de ambientes, 6th ed. Editora UFLA, Lavras.

Resende M, Curi N, Poggere GC, Barbosa JZ, Pozza AAA (2019) 'Pedologia, fertilidade, água e planta: Inter-relações e aplicações.' (Editora UFLA: Lavras)

Ribeiro BT, Silva SHG, Silva EA, Guilherme LRG (2017) Portable X-ray fluorescence (pXRF) applications in tropical Soil Science. *Ciência e Agrotecnologia* **41**, 245–254. doi:10.1590/1413-70542017413000117.

Ribeiro BT, Weindorf DC, Silva BM, Tassinari D, Amarante LC, Curi N, Guilherme LRG (2018) The influence of soil moisture on oxide determination in tropical soils via portable x-ray fluorescence. *Soil Science Society of America Journal* **82**, 632.

Sarkhot D V., Grunwald S, Ge Y, Morgan CLS (2011) Comparison and detection of total and available soil carbon fractions using visible/near infrared diffuse reflectance spectroscopy. *Geoderma* **164**, 22–32. doi:10.1016/j.geoderma.2011.05.006.

Sharma A, Weindorf DC, Man T, Aldabaa AAA, Chakraborty S (2014) Characterizing soils via portable X-ray fluorescence spectrometer: 3. Soil reaction (pH). *Geoderma* **232–234**, 141–147. doi:10.1016/j.geoderma.2014.05.005.

Sharma A, Weindorf DC, Wang DD, Chakraborty S (2015) Characterizing soils via portable

X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). *Geoderma* **239**, 130–134. doi:10.1016/j.geoderma.2014.10.001.

Silva SHG, Poggere GC, Menezes MD, Carvalho GS, Guilherme LRG, Curi N (2016) Proximal sensing and digital terrain models applied to digital soil mapping and modeling of Brazilian Latosols (Oxisols). *Remote Sensing* **8**, 614–635.

Silva SHG, Teixeira AFS, Menezes MD, Guilherme LRG, Moreira FMS, Curi N (2017) Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). *Ciência e Agrotecnologia* **41**, 648–664. doi:10.1590/1413-70542017416010317.

Silva SHG, Hartemink AE, Teixeira AFS, Inda AV, Guilherme LRG, Curi N (2018) Soil weathering analysis using a portable X-ray fluorescence (PXRF) spectrometer in an Inceptisol from the Brazilian Cerrado. *Applied Clay Science* **162**, 27–37. doi: 10.1016/j.clay.2018.05.028.

Silva FM, Weindorf DC, Silva SHG, Silva EA, Ribeiro BT, Guilherme LRG, Curi N (2019) Tropical Soil Toposequence Characterization via pXRF Spectrometry. *Soil Science Society of America Journal* **83**, 1153–1166. doi:10.2136/sssaj2018.12.0498.

Silva SHG, Weindorf DC, Pinto LC, Faria WM, Acerbi Junior FW, Gomide LR, Mello JM, Pádua Junior AL, Souza IA, Teixeira AFS, Guilherme LRG, Curi N (2020) Soil texture prediction in tropical soils: A portable X-ray fluorescence spectrometry approach. *Geoderma* **362**. doi:10.1016/j.geoderma.2019.114136.

Siqueira JDP, Lisboa RS, Ferreira AM, Souza MFR, Araújo E, Lisbão Júnior L, Siqueira MM (2004) Estudo ambiental para os programas de fomanto florestal da Aracruz Celulose S.A. e extensão florestal do Governo do Estado do Espírito Santo. *Floresta* 3–67.

Soil Survey Staff (2014) 'Keys to Soil Taxonomy.' (USDA, Washington, DC.).

Soltanpour PN, Johnson GW, Workman SM, Jones BJ, Miller RO (1996) Inductively coupled plasma emission spec- trometry and inductively coupled plasma-mass spectrometry. 'Methods of soil analysis. Part 3. Chemical methods.' pp. 91–139 doi:10.2136/ sssabookser5.3.c5.

Souza E, Fernandes Filho EI, Schaefer CEGR, Batjes NH, Santos GR, Pontes LM (2016) Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. *Scientia Agricola* **73**, 525–534. doi:10.1590/0103-9016-2015-0485.

Tavares TR, Molin JP, Nunes LC, Alves EEN, Melquiades FL, Carvalho HWP, Mouazen AM (2020) Effect of x-ray tube configuration on measurement of key soil fertility attributes

with XRF. *Remote Sensing* **12**, 963–983. doi: doi: 10.3390/rs12060963.

Teixeira AFS, Weindorf DC, Silva SHG, Guilherme LRG, Curi N (2018) Portable x-ray fluorescence (pXRF) spectrometry applied to the prediction of chemical attributes in inceptisols under different land use. *Ciencia e Agrotecnologia* **42**, 501–512. doi:10.1590/1413-70542018425017518.

Teixeira AFS, Pelegrino MHP, Faria WM, Silva SHG, Gonçalves MGM, Acerbi Júnior FW, Gomide LR, Pádua Júnior AL, Souza IA, Chakrabortyd S, Weindorfe DC, Guilherme LRG, Curi N (2020) Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry. *Geoderma* **361**, 114132. doi:10.1016/j.geoderma.2019.114132.

Teixeira PC, Donagema GK, Fontana A, Teixeira WG (2017) 'Manual de métodos de análise do solo.' (Brasília: Embrapa,).

Towett EK, Shepherd KD, Lee Drake B (2015) Plant elemental composition and portable X-ray fluorescence (pXRF) spectroscopy: Quantification under different analytical parameters. *X-Ray Spectrometry* **45**, 117–124. doi:10.1002/xrs.2678.

Valadares JMA, Bataglia OC, Furlani PR (1974) Estudo de materiais calcários usados como corretivo do solo no Estado de São Paulo. IQ - Determinação de Mo, Co, Cu, Zn,, Mn e Fe. *Bragantia* **33**, 147–152.

Vapnik V (1995) 'The Nature of Statistical Learning Theory Springer-Verlag.'

Viscarra Rossel RA, Adamchuk VI, Sudduth KA, McKenzie NJ, Lobsey C (2011) 'Proximal Soil Sensing. An Effective Approach for Soil Measurements in Space and Time.' (Elsevier Inc.) doi:10.1016/B978-0-12-386473-4.00010-5.

Viscarra Rossel RA, McBratney AB, Minasny B (2010) 'Proximal soil sensing.'

Wan M, Hu W, Qu M, Li W, Zhang C, Kang J, Hong Y, Chen Y, Huang B (2020) Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. *Geoderma* **363**, 114163. doi:10.1016/j.geoderma.2019.114163.

Wan M, Qu M, Hu W, Li W, Zhang C, Cheng H, Huang B (2019) Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals. *Process Safety and Environmental Protection*. doi:10.1016/j.psep.2019.09.025.

Wang D, Chakraborty S, Weindorf DC, Li B, Sharma A, Paul S, Ali MN (2015) Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. *Geoderma* **243**–**244**, 157–167. doi:10.1016/j.geoderma.2014.12.011.

Wei T, Simko V, Levy M, Xie Y, Jin Y, Zemla J (2017). Package "corrplot". https://cran.r-project.org/web/packages/corrplot/corrplot.pdf (verified 23 Jan. 2020).

Weindorf DC, Bakr N, Zhu Y (2014) 'Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications.' (Elsevier) doi:10.1016/B978-0-12-802139-2.00001-9.

Weindorf DC, Chakraborty S (2016) Portable X-ray Fluorescence Spectrometry Analysis of Soils. *Methods of Soil Analysis* **1**, 0. doi:10.2136/methods-soil.2015.0033.

Weindorf DC, Herrero J, Castañeda C, Bakr N, Swanhart S (2013) Direct Soil Gypsum Quantification via Portable X-Ray Fluorescence Spectrometry. *Soil Science Society of America Journal* **77**, 2071. doi:10.2136/sssaj2013.05.0170.

Weindorf DC, Zhu Y, McDaniel P, Valerio M, Lynn L, Michaelson G, Clark M, Ping CL (2012) Characterizing soils via portable x-ray fluorescence spectrometer: 2. Spodic and Albic horizons. *Geoderma* **189**–**190**, 268–277. doi:10.1016/j.geoderma.2012.06.034.

Were K, Bui DT, Dick ØB, Singh BR (2015) A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afromontane landscape. *Ecological Indicators* **52**, 394–403. doi:10.1016/j.ecolind.2014.12.028.

Xu D, Zhao R, Li S, Chen S, Jiang Q, Zhou L, Shi Z (2019) Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. *European Journal of Soil Science* **70**, 162–173. doi:10.1111/ejss.12729.

Zhang Y, Hartemink AE (2019) Data fusion of vis-NIR and PXRF spectra to predict soil physical and chemical properties. *European Journal of Soil Science* 1–18. doi:10.1111/ejss.12875.

Zhu Y, Weindorf DC, Zhang W (2011) Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. *Geoderma* **167**–**168**, 167–177. doi:10.1016/J.geoderma.2011.08.010.

ARTICLE 2 - Prediction of soil organic matter content by combining data from Nix Pro™ color sensor and portable X-ray fluorescence spectrometry in tropical soils

Alvaro José Gomes de Faria[a], Sérgio Henrique Godinho Silva[a], Renata Andrade[a], Marcelo Mancini[a], Leônidas Carrijo Azevedo Melo[a], David C. Weindorf[b], Luiz Roberto Guimarães Guilherme[a], Nilton Curi[a].

a Department of Soil Science, Federal University of Lavras – UFLA, Doutor Sylvio Menicucci Avenue, Lavras, Minas Gerais State 37200-900, Brazil.
b Department of Earth and Atmospheric Sciences, Central Michigan University, Mount Pleasant, MI 48858, USA.

Abstract

Soil organic matter (SOM) measurement is of great agricultural and environmental importance. Thus, the development of rapid, environmentally-friendly, economical and reliable assessment methods is challenging. Soil proximal sensors have become an important approach for SOM prediction worldwide, but require regional calibration. This work aimed to assess the efficiency of SOM content prediction using the Nix Pro™ color sensor and portable X-ray fluorescence (pXRF) spectrometry, either separately or combined. The type of soil horizon collected (A or B) was used as auxiliary input data. A total of 705 Brazilian variable soil samples were analyzed in the laboratory for SOM content and scanned by Nix Pro™ and pXRF. Via Nix Pro™, samples were analyzed both dry and moist since moisture changes their color. Prediction models were built using 70% of the data via the stepwise multiple linear regression (SMLR), support vector machine with linear kernel (SVM) and random forest (RF). Validation was performed with the remaining 30% of the data through the coefficient of determination (R^2), the root mean square error (RMSE) and the residual prediction deviation (RPD). SOM content was predicted with good accuracy (R^2 = 0.73, RMSE = 1.09% and RPD = 2.00) using the RF algorithm trained with combined data from the Nix Pro™ and pXRF sensors. Soil horizons and Ca content were the two most important predictor variables. The combination of data obtained by Nix Pro™ and pXRF yielded

accurate SOM predictions for a wide variety of Brazilian soils, in addition to being environmentally-friendly, without generating chemical waste.

Keywords: Proximal sensors, Random forest, pXRF, Soil color, Soil modeling, tropical soils.

1. Introduction

Soil organic matter (SOM) plays a central role in soil organic carbon (SOC) sequestration and mitigation of greenhouse gas emissions (Jia et al., 2017; Stanley et al., 2018). SOM is also of great importance for the soil structure, fertility, water availability, resistance to erosion, biological activity and diversity, which promote soil health and productivity (Dotto et al., 2018; Johnston et al., 2009; Munoz and Kravchenko, 2011; Wang et al., 2015). In tropical soils, SOM becomes even more important as it can contribute up to 90% of cation exchange capacity (CEC) in surface horizons (Raij, 1969; Silva et al., 1994; Verdade, 1956). Moreover, SOM is a primary indicator of soil quality and has a high demand for determination in routine soil testing (Kheir et al., 2010; Le Quere et al., 2018; Lorenz and Lal, 2016).

Classic wet or dry laboratory-based methods are widely adopted to determine SOM content (Wang et al., 2015), such as those conducted through the oxidation of the dichromate (Walkley and Black, 1934; Yeomans and Bremner, 1988) and dry combustion (Chatterjee et al., 2009; Wang et al., 2012). However, these methods are laborious, time-consuming and require chemical reagents generating wastes that can cause environmental impacts if not properly discarded (Benedet et al., 2021; Lal et al., 2001; McDowell et al., 2012; Nawar and Mouazen, 2019; Wang et al., 2015). These disadvantages are further exacerbated for the assessment of SOM variability in an area of interest due to the large number of samples needed (McCarty and Reeves, 2006; McDowell et al., 2012). Therefore, due to the importance of SOM determination for both agriculture and environmental aspects, there is an urgent need to investigate new methodologies that allow accurate, rapid, low cost, and chemical waste-free SOM analyses (Dotto et al., 2018; Mukhopadhyay et al., 2020; Silva et al., 2020; Wang et al., 2015).

In this context, proximal soil sensors have increasingly attracted the attention of scientists by offering a practical solution to estimate SOM content (Horta et al., 2015; Silva et al., 2017; Teixeira et al., 2020; Viscarra Rossel et al., 2011). Successful SOM and/or SOC predictions have been achieved across the globe via visible and near-infrared (Vis-NIR)

diffuse reflectance spectroscopy (DRS) (Brodský et al., 2013; Wang et al., 2015). However, this proximal sensor is expensive, results often require pre-processing (e.g., 1st order derivative transformation with Savitsky-Golay) and data analyses can be complex due to the large number of variables obtained (Stiglitz et al., 2017a, 2017b; Benedet et al., 2020a, 2020b). Other proximal sensors used for SOM and/or SOC predictions are the portable X-ray fluorescence (pXRF) spectrometer and the Nix Pro™ color sensor. They are less expensive and provide results easier to both interpret and use in prediction models (Dijair et al., 2020; Faria et al., 2020; Mikhailova et al., 2017; Mukhopadhyay et al., 2020; Raeesi et al., 2019; Silva et al., 2020; Stiglitz et al., 2017a; Teixeira et al., 2020). Besides, the adoption of both proximal sensors by the soil science community is more recent than Vis-NIR DRS, especially Nix Pro™ (Stiglitz et al., 2016), encouraging further studies on their efficacy for soil property predictions under variable soil conditions.

The Nix Pro™ sensor is already being considered a promising alternative for obtaining a reliable soil color assessment compared to conventional methods (other colorimeters and Munsell Soil Color Charts) (Mancini et al., 2020; Moritsuka et al., 2019; Stiglitz et al., 2017b; Stiglitz et al., 2016). This sensor is inexpensive, light-weighted, portable, equipped with a light-emitting diode (LED), can be connected to smartphones via Bluetooth and provides quantitative measurements of soil color in various color systems proving to be convenient for statistical analysis (Stiglitz et al., 2017b; Stiglitz et al., 2016). Thus, data obtained by Nix Pro™ have been used to predict SOM and/or SOC with significant accuracy, since dark soil colors are mainly caused by the presence of SOM (Mikhailova et al., 2017; Mukhopadhyay et al., 2020; Raeesi et al., 2019; Stiglitz et al., 2017a). Due to the correlations between SOM and total nitrogen, Nix Pro™ also proved successful in predicting nitrogen content (Mikhailova et al., 2017). However, some studies conducted in regions with contrasting climate types showed that the soil moisture directly interfered on soil color and, consequently, affected the SOM and/or SOC prediction models (Mukhopadhyay et al., 2020; Raeesi et al., 2019; Stiglitz et al., 2017a). Thus, there is a need to further investigate the influence of moisture on Nix Pro™ analyses to standardize the best moisture condition that favors SOM prediction. Despite the potential applications of this new sensor, there are rare studies exploring it in tropical countries where some very weathered-leached soils, compared with soils of temperate regions, tend to present colors that do not reflect the SOM content. This is a consequence of Fe-oxides in these soils, mainly containing hematite in great amounts, a mineral that presents high pigmenting power. Secondarily, the generally higher clay content and relatively lower SOM content in tropical soils than in temperate soils dilutes the SOM power to pigmentation

of these soils, considering they have more surface area to pigment. As such, further studies are required to assess the efficiency of this sensor under tropical conditions.

Conversely, pXRF has already been widely used in several areas because of its ability to identify and quantify the total elemental content in several matrices, including soil (Ribeiro et al., 2017; Silva et al., 2021; Teixeira et al., 2020; Weindorf et al., 2014). Briefly, pXRF detects the characteristic fluorescence energy released by the atoms of each element present in the analyzed material after they are hit by high energy X-rays (Weindorf et al., 2014). Subsequently, the quantification of the elements is directly related to the intensity of the fluorescence energy that is detected by the equipment (Ribeiro et al., 2017; Weindorf et al., 2014). Data obtained by pXRF have been used both for soil characterization and for predicting soil properties with significant accuracy in different regions of the world, including SOM (Andrade et al., 2020a, 2020b; Chakraborty et al., 2019; Duda et al., 2017; Faria et al., 2020; Rawal et al., 2019; Sharma et al., 2015; Silva et al., 2017, 2020; Tavares et al., 2020a; Tavares et al., 2020b; Wang et al., 2015).

Both Nix Pro™ and pXRF require minimal sample preparation and are considered non-destructive analyses, being fast and useful in laboratory or field studies, which facilitates their utilization compared with standard wet-chemistry laboratory methods (Dijair et al., 2020; Moritsuka et al., 2019; Ribeiro et al., 2017; Stiglitz et al., 2016; Weindorf et al., 2014; Silva et al., 2021). It is noteworthy that studies involving the individual use of these sensors to predict SOM are still quite incipient in the literature and more studies are needed in different regions of the world (Andrade et al., 2020a; Mikhailova et al., 2017; Mukhopadhyay et al., 2020; Raeesi et al., 2019; Silva et al., 2017; Stiglitz et al., 2017a; Tavares et al., 2020a; Wang et al., 2015). Besides that, due to the complex nature of soils, such predictions are not always successful (Wan et al., 2019, 2020). Thus, the combination of data from multiple sensors becomes a viable and promising alternative in providing robust models to improve the prediction of soil attributes (Wan et al., 2019, 2020). Finally, studies in the literature that explore data combination from multiple sensors in the generation of models for SOM prediction in temperate or tropical regions are still rare. Even so, Kagiliery et al. (2019) used such an approach (pXRF + Nix Pro™) in boosting the predictive accuracy of S content in lignite. Their combined predictive models were more accurate than either sensor in isolation.

Thus, this study aims to: (i) predict SOM content through data obtained by Nix Pro™ and pXRF spectrometer in tropical soils, (ii) evaluate the efficiency of these sensors to provide data to be used separately or combined via three machine learning algorithms: stepwise multiple linear regression (SMLR), support vector machine with linear kernel

(SVM) and random forest (RF) in two soil moisture conditions. In addition, the horizons from which the soil samples have been collected will be used as auxiliary input data for model generation. We hypothesized that the combination of data obtained by Nix Pro™ and pXRF will be able to predict SOM content quickly and accurately in varied tropical soils of Brazil, delivering results comparable with those reported in previous studies worldwide using Vis-NIR DRS.

2. Material and methods

2.1. Study area and field sampling

For this study, soil samples were collected in the Brazilian states of Minas Gerais (MG), Rio de Janeiro (RJ), São Paulo (SP) and Santa Catarina (SC) (Fig. 1), representing a total area covering ~11.4% (973,780 km^2) of the Brazilian territory. According to the Koppen classification system, the study area covers the following climate types: Aw (tropical climate with dry winter), Cf (subtropical humid climate without dry season) and Cw (subtropical humid climate with dry winter) (Alvares et al., 2013). The mean annual rainfall in the studied regions ranges between 1500 and 1900 mm, the mean annual temperature varies from 13 to 23 ºC and the altitude in those places varies between 600 and 1100 m amsl (above mean sea level) (Alvares et al., 2013).

A total of 705 soil samples representative of the main Brazilian soils (~500 g each) were collected from 313 soil profiles, encompassing 298 samples from the A or O horizon (most of them in the depth range from 0 to 20 cm) and 407 samples from the B horizon (or C horizon for soils without B horizon), mostly in the depth range between 80 and 100 cm (Fig. 1) (Santos et al., 2015). These samples encompassed eight soil classes at the order categorical level per the Brazilian Soil Classification System (Santos et al., 2018): Cambisols (Inceptisols), Latosols (Oxisols), Argisols (Ultisols), Nitosols (Ultisols), Neosols (Entisols), Plintosols (Oxisols – Plinthaquoxes), Gleisols (Entisolsa – Endoaquents), and Organosols (Histosols) - classes in parentheses represent the approximate correspondence with the Keys to US Soil Taxonomy (Soil Survey Staff, 2014). Together, these classes cover 89.6% (7,630,336 km^2) of the Brazilian territory (Anjos et al., 2012; Santos et al., 2011).

Sampled soils are derived from several parent materials including granite, gneiss, gabbro, alluvial and colluvial sediments, slate, sandstone, quartzite, phyllite, basalt, itabirite, mica-schist, amphibolite, charnockite, limestone and tuffite, all of them identified during

fieldwork (Silva et al., 2020; Teixeira et al., 2020). Land use encompasses native vegetation and areas cultivated with coffee, eucalyptus and pasture, whose agronomic management includes the application of limestone and/or gypsum prior to mineral fertilizers, according to the needs of each crop.
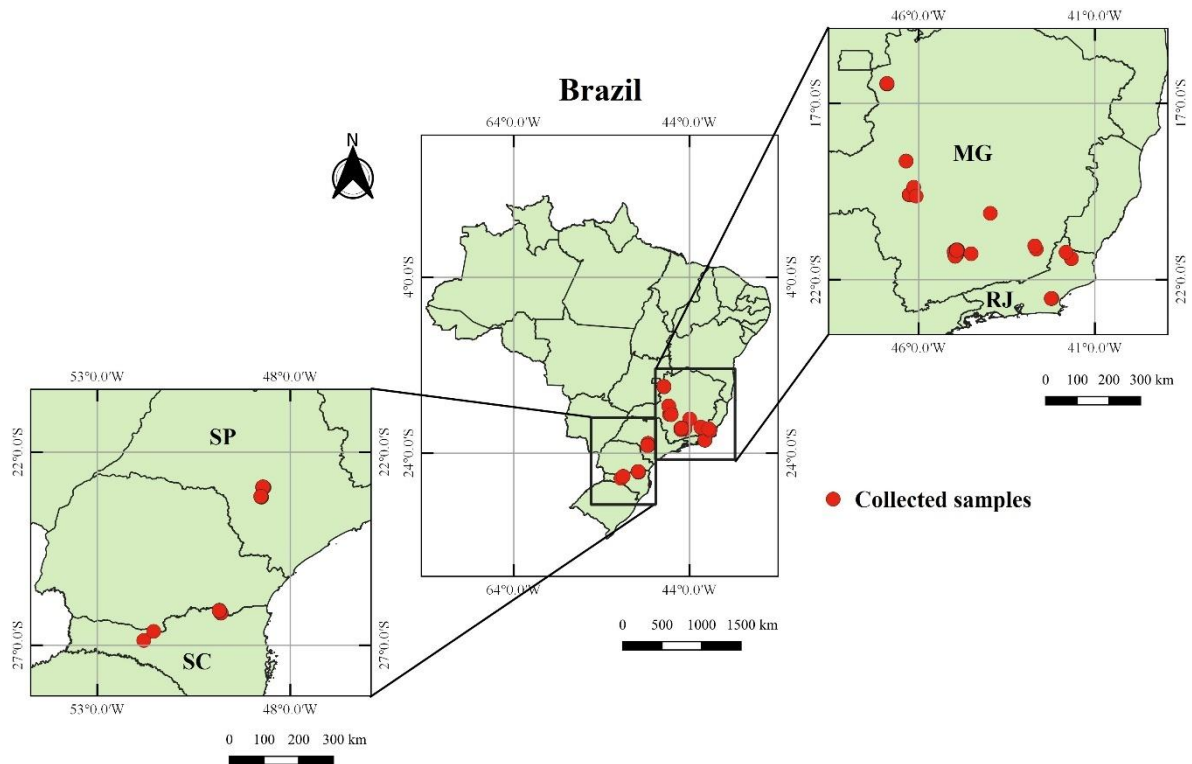
Fig. 1. Map of Brazilian states and locations where soil samples were collected. MG - Minas Gerais; RJ - Rio de Janeiro; SP – São Paulo; and SC - Santa Catarina.

2.2. Laboratory analyses

In the laboratory, all soil samples were air-dried, disaggregated and passed through a 2-mm sieve (air-dried fine earth - ADFE). SOC contents (in percentage) of the samples were determined by dichromate acid oxidation followed by titration with ammonium ferrous sulfate (Walkley and Black, 1934; Yeomans and Bremner, 1988). Then, SOM contents (in percentage) were calculated by multiplying the SOC content by the van Bemmelen factor of 1.724 (% SOM = SOC x 1.724) (Howard, 1965; Broadbent, 1953). Next, SOM contents were used to calibrate the prediction models based on proximal sensors data.

For the analyses via the Nix Pro™ sensor (Hamilton, Ontario, Canada) a subsample of ADFE was selected to determine soil color following the methods previously described by

Stiglitz et al. (2016, 2017a). First, a layer of approximately 3 cm of soil from each sample was spread over a clean white paper and leveled to prevent external light from entering the sensor aperture during the sample scanning via Nix Pro™. An unleveled surface of the sample would not entirely block the external light on the sensor's aperture during scanning, thus, influencing the Nix Pro™ results (Stiglitz et al., 2017a, 2017b; Mukhopadhyay et al., 2020). Subsequently, soils were scanned for color determination under both dry and moist conditions. For the moist condition, a dropper with distilled water was used to moisten the soil surface until no more color change occurred. After each analysis via Nix Pro™, proper care was taken to clean the dust off the sensor's aperture area. The results of color scanning were obtained and expressed in dimensionless magnitudes in terms of RGB (Red, Green and Blue), XYZ (Luminance - Y, Cartesian representation of colors - X and Z), CIELAB (Lightness - L*, Redness - A*, and Yellowness - B*), CIELCH (Lightness - L*, Chroma - C*, and hue angle in degrees - H*) and CMYK (Cyan, Magenta, Yellow, and Black). The L variable is the same for the CIELAB and CIELCH systems, only one of them was used for subsequent analyses. Thus, five color systems were included in the models since they were indispensable in predicting SOM through analyses carried out previously. Additional information on these five color systems can be obtained from Viscarra Rossel et al. (2006) and Mancini et al. (2020).

Another soil subsample of the ADFE was selected for scanning via a portable X-ray fluorescence (pXRF) spectrometer (model Bruker® S1 Titan LE) to determine the total elemental content of the samples. This equipment consists of an Rh X-ray tube of 4 W, 15–50 KeV, 5–100 μA, and a silicon drift detector (SDD) with a resolution of <145 eV, which allows the detection of numerous elements between Mg and U on the Periodic Table of Elements (Weindorf et al., 2014). The 705 soil samples were scanned for 60 s, in triplicate and the mean obtained was used, using the Trace (dual soil) mode integrated with the Geochem software, according to Weindorf and Chakraborty (2016). Thus, the following elements were used to develop the prediction models: Al, Ag, Al, As, Ba, Bi, Ca, Ce, Cl, Cr, Cu, Fe, K, Mn, Nb, Ni, P, Pb, Rb, Rh, S, Si, Sr, Ta, Ti, Tl, V, Y, Zn, and Zr.

To assess the accuracy of pXRF, the elemental contents of two National Institute of Standards and Technology (NIST) certified reference materials (2710a and 2711a) and a check sample (CS) provided by the pXRF manufacturer were analyzed. The results were used to calculate the percentage of recovery for each element: [% recovery = 100 × (content obtained by pXRF/certified content)]. Thus, the recovery values per element (%) are presented as follows (2710a / 2711a / CS): Ag (0/0/0); Al (58,8/103,2/97,4); As (0/0/0); Ba

(0/0/0); Bi (0/0/0); Ca (19,3/66,1/0); Ce (0/0/0); Cl (0/0/0); Cr (0/0/0); Cu (82,6/67,1/ 95,8); Fe (73,6/69,2/94,8); K (83,9/56,2/92,2); Mn (69/60/92,3); Nb (0/0/0); Ni (0/23/106); P (314,3/1145,5/0); Pb (0/0/0); Rb (0/86,4/ 0); Rh (0/0/0); S (0/0/0); Si (93,8/91,5/96,4); Sr (100/89,8/0); Ta (0/ 0/0); Ti (100,4/75,5/0); Tl (0/0/0); V (0/0/0); Y (0/0/0); Zn (97/85,3/ 0); and Zr (0/99/0). Zero values indicate that pXRF did not detect the element or that the element had no certified value. Raw elemental contents were used to calibrate the models; no correction on these values was made considering the recovery values.

2.3. Statistical analyses

Initially, the SOM content obtained via standard laboratory analysis and the data generated by the Nix Pro™ and pXRF sensors were submitted to descriptive statistics to obtain the minimum, maximum, mean, standard deviation (SD) and coefficient of variation (CV) of the dataset. Next, to help the understanding of the importance of the variables, Pearson's correlation coefficient (r) was calculated between the data obtained by the two sensors separately and the SOM content, using the entire data set (n = 705). All statistical analyses were performed in R software (R Core Team, 2018).

Before generating the SOM prediction models, all data obtained by both sensors were normalized and scaled using the standard scale function in the R software. This function is given by: standardized elemental content = [x - average (x)/standard deviation (x)], where x is the total elemental content obtained by pXRF or the dimensionless value of the variables obtained by Nix Pro™. At first, the entire dataset (n = 705) was used to generate the models, where they were randomly separated into modeling (n = 495, ~ 70%) and validation (n = 210, ~ 30%) subsets using the createDataPartition function provided by the caret package of R software (Kuhn et al., 2018). Special care was taken to keep samples from the A and B horizons of the same soil profile in the same dataset (either modeling or validation), to maintain true data independence during the validation procedure. It should be noted that the use of a dataset that contains different horizons (A and B) allows investigating the possibility of creating a general and robust model that is suitable for samples from both A and B horizons or when the objective is to determine the content of SOM with depth, which is important in the process of evaluating and improving soil profile fertility.

Thus, SOM prediction models were generated for five different conditions: i) only with data from the Nix Pro™ in a dry soil condition (n = 15 predictor variables); ii) only with data from the Nix Pro™ in a moist soil condition (n = 15 predictor variables); iii) only with

data from the pXRF in a dry soil condition (n = 29 predictor variables); iv) combining data from Nix Pro™ + pXRF, both in a dry soil condition (n = 44 predictor variables); and v) combining data from the Nix Pro™ in a moist soil condition + pXRF in a dry soil condition (n = 44 predictor variables). In all described conditions, the models were also generated with or without using the information of the horizons collected as auxiliary input data. For this, three different machine learning algorithms were used in the five conditions above mentioned: stepwise multiple linear regression (SMLR), support vector machine with linear kernel (SVM) and random forest (RF) through the R "Caret" package (Kuhn, 2008).

The SMLR is a semi-automatic model construction process that was executed after applying the variance inflation factor (VIF) to minimize the problem of multicollinearity that occurs when working with a dataset with many predictive variables. The least important variables were removed using the Akaike information criterion through the step function (Akaike, 1973). This procedure ensures that only variables that contribute to the model are incorporated into the final model. Variable selection was based on model root mean square error (RMSE).

SVM uses kernel functions, i.e., the input data is plotted into a new hyperspace, where separations are performed (Forkuor et al., 2017). The ultimate goal is to obtain an ideal hyperspace for adjusting and predicting data using the insensitive loss function that tolerates errors less than the constant defined as a limit (Hastie et al., 2009).

The RF algorithm is a machine learning method that operates by constructing a multitude of decision trees and by averaging the results of all the individual trees. Each regression tree is manufactured based on the bootstrap samples of the data. To fit each tree, a random subset of predictors is applied. The growth of the tree is continued until the specified minimum node size is reached (Breiman, 2001; Liaw and Wiener, 2002; Raeesi et al., 2019). For this algorithm, the following parameters were established: number of trees (ntrees) = 500, number of variables used in each tree (mtry) = 6, corresponding to one third of the number of predictors (Liaw and Wiener, 2002). The RF does not provide a final equation, but it does provide the importance of the predictor variables for the model. One of the metrics used for this purpose is the percentage increase in the mean square error (%IncMSE). The higher the value of %IncMSE, the greater the importance of the variable for the model (Forkuor et al., 2017; Gonzalez et al., 2015).

2.4. Validation of the models

The performance of the SOM prediction models in each situation was assessed by the validation dataset (remaining 30% of the samples) through the coefficient of determination (R^2), the root mean square error (RMSE) and the residual prediction deviation (RPD). The R^2 value describes the proportion of the total variance of the observed data that can be explained by the model, whose value ranges from 0 to 1, with higher values indicating better agreement (Wan et al., 2020). According to Viscarra Rossel et al. (2010), the R^2 values for soil attribute predictions are considered very good (>0.81), good (0.61–0.8), regular (0.41–0.6) and poor (<0.4). The RMSE is a reliable standard statistical metric for model evaluation, commonly used in several studies with this approach (Chai and Draxler, 2014). Models with RMSE values closer to zero present better performance. For the RPD, Chang et al. (2001) defined three categories for this parameter related to the quality and reliability of the models: accurate (RPD > 2), moderately accurate (1.4 ≤ RPD ≤ 2), and inaccurate models (RPD < 1.4). Thus, the models that provide the highest values of R^2 and RPD and the lowest values of RMSE were considered optimal for SOM content prediction.

3. Results and discussion

3.1. Descriptive statistics for SOM, Nix Pro™ and pXRF data

3.1.1. Descriptive statistics for SOM

The descriptive statistics of SOM obtained from laboratory analyses for A + B (n = 705), A (n = 298) and B (n = 407) horizons are presented in Fig. 2. SOM contents in the A horizon ranged from 0.80 to 18.21% with mean value of 3.94%, standard deviation of 2.45% and a CV value of 62.18%. For the B horizon data, SOM contents ranged from 0.04 to 9.09% with mean value of 1.62%, standard deviation of 0.95% and a CV value of 58.64%. Thus, the high CV (84.10%) value for the entire dataset used in this work (A + B horizons combined; n = 705) indicates high range and variability of SOM contents in the study areas in relation to the general mean (2.53%).
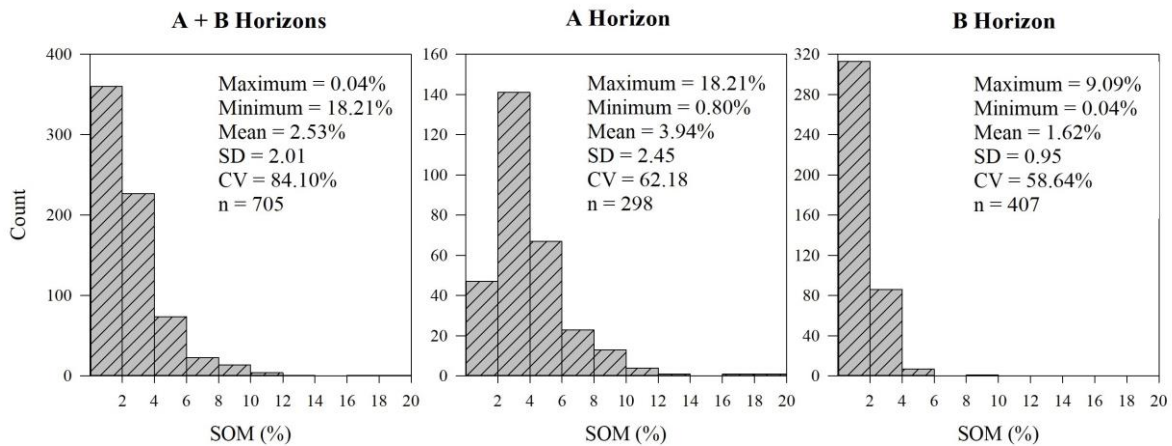
Fig. 2. Descriptive statistics for soil organic matter (SOM) content obtained from laboratory analyses for A + B (n = 705), A (n = 298) and B (n = 407) horizons of soils collected in different Brazilian states. SD – standard deviation; CV – coefficient of variation.

This trend was consistent with the findings of Gomes et al. (2019), which also reported high range and variability in SOM content across Brazil. Countrywide, these authors also confirmed that the highest content of SOM is concentrated in the topsoil, reducing drastically with depth, as observed in the present study (Fig. 2). SOM content is influenced by several factors such as vegetation (Gomes et al., 2019), topography (Cardinael et al., 2017), climate (Munoz-Rojas et al., 2017), soil class (Zhao et al., 2006), soil depth/soil horizons (Li et al., 2017), soil mineralogy (Dwivedi et al., 2017), soil biota (Komarov et al., 2017), soil management (Li et al., 2017) and land use (Silva et al., 2017). In the present work, considering the great geographical expression of the sampling places, the aforementioned factors explain the high SOM content variability in the present study.

3.1.2. Descriptive statistics for Nix Pro[TM] color sensor

In general, the descriptive statistics of soil color showed high variability according to the high CV value obtained in each parameter evaluated (Fig. 3). Results of dry soil condition showed, in most cases, the highest CV values ranging from 8.7% (Y-CMYK, yellow) to 51.4% (K-CMYK, black), as expected, reflecting the high variability of SOM content in this study. K represents black color, and the dataset includes both A and B soil horizons, hence resulting in a larger amplitude of SOM contents. This might indicate the usefulness of K as a predictor variable of SOM content.

Color systems - Nix Pro™
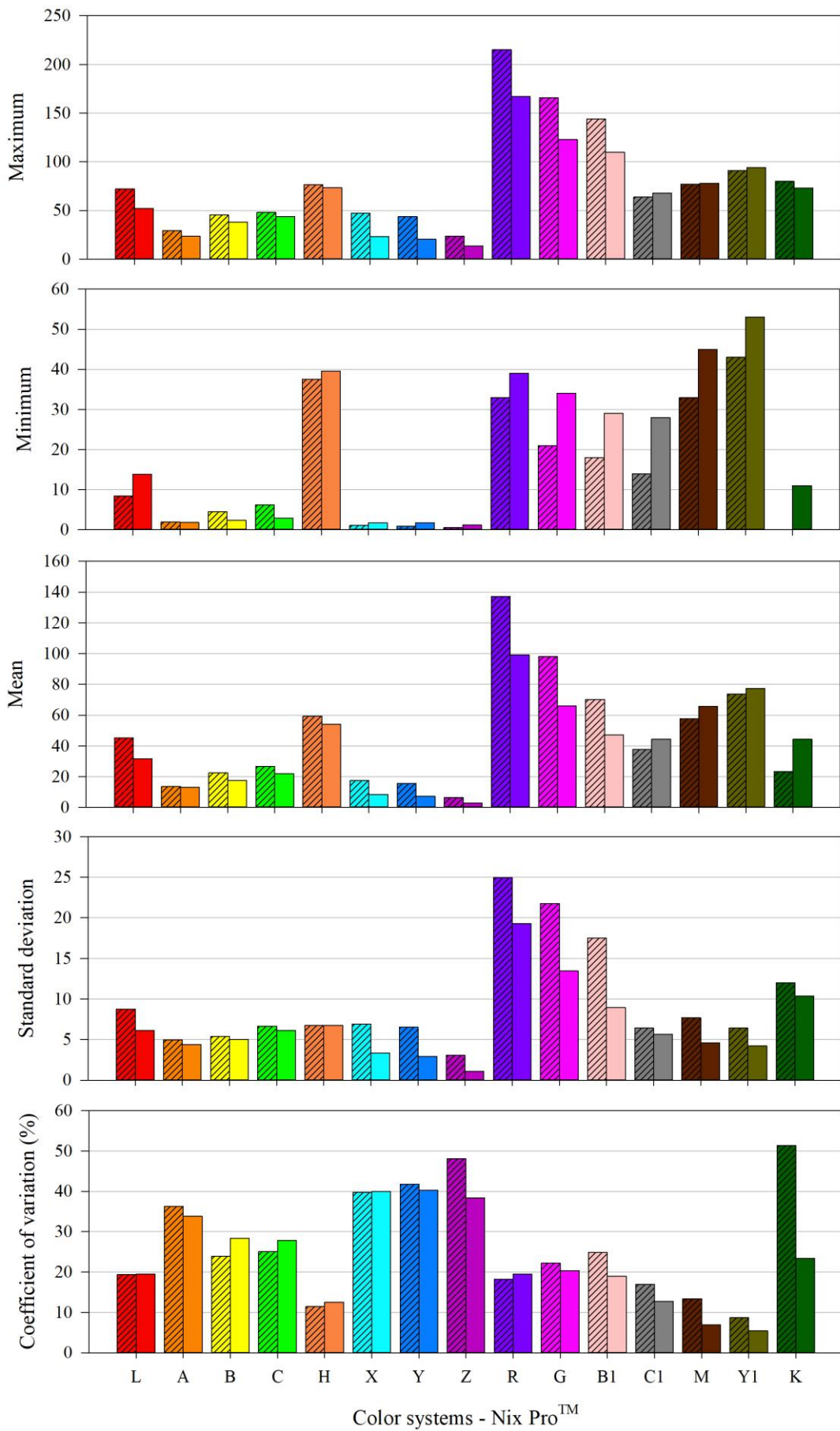
Fig. 3. Descriptive statistics for each parameter that compose the color systems: [1]CIELAB, [2]CIELCH, [3]XYZ, [4]RGB and [5]CMYK obtained by the Nix Pro™ color sensor in Brazilian soil samples under dry and moist conditions (n = 705). Hatched bars – dry soil samples; Unhatched bars – moist soil samples.

Baumgardner et al. (1969) reported that SOM contents >2.0% tend to mask the effects of iron oxides, mainly in surface horizons. Nonetheless, in tropical red soils with high hematite content the pigmenting power of this iron oxide mineral prevails (Resende et al., 2011). Accordingly, this might indicate that color variables that express the "colorfulness" of samples, i.e., chroma, could be relevant as predictor variables. In this study, about 50% (n = 353) of the samples have SOM content >2%, most of them in A horizon (Fig. 2).

The analyses performed on dry soil samples also provided the highest standard deviation and amplitudes of color average values, except for the CMYK system (Fig. 3). In this condition, the R (red) and G (green) parameters of the RGB system showed the highest average values. In the CMYK system, the Y (yellow) and M (magenta) parameters showed the highest average value regardless of the moisture condition of the samples. Considering that in CMYK the red color is expressed by Y and M combined, this result reflects the mineralogy of the clay fraction of tropical soils, which contains significant amounts of hematite and goethite iron oxides (Curi et al., 2017; Kampf et al., 2012; Poggere et al., 2018). Hence, tropical soils with low SOM content (<2%) have predominantly yellow colors at the surface where SOM favors goethite over hematite formation (Kampf and Curi, 2000) and red, red-yellow or yellow colors in the subsurface horizons (Baumgardner et al., 1970; Correa et al., 2008; Resende et al., 2011; Ramos et al., 2020).

The lower variation of results obtained in most parameters of the evaluated color systems in moist soil samples were expected since soil color changes when samples are dry and moist (Viscarra Rossel et al., 2006). Specifically, moisture darkens soil color, increasing its tonality and reducing its brightness (Demattê et al., 2006; Shields et al., 1968; Shonk et al., 1991). Thus, color of soil samples becomes more homogeneous when moist (Viscarra Rossel et al., 2008), as occurred in this current study. In order to illustrate the differences in the relative variation of soil color when samples are dry and moist, Fig. 4 shows the plots of some color variables reported by Nix Pro™ and the corresponding color of the soil samples when dry and moist. Moisture strongly influenced soil color regardless of the color systems reported by Nix Pro™. Andrade et al. (2020d) reported darker soil colors when samples were moist, decreasing soil sample's reflectance for tropical soils, similar to the present study.

Conversely, Stiglitz et al. (2016) observed that moisture did not cause strong changes in color of temperate soils analyzed via Nix Pro™.
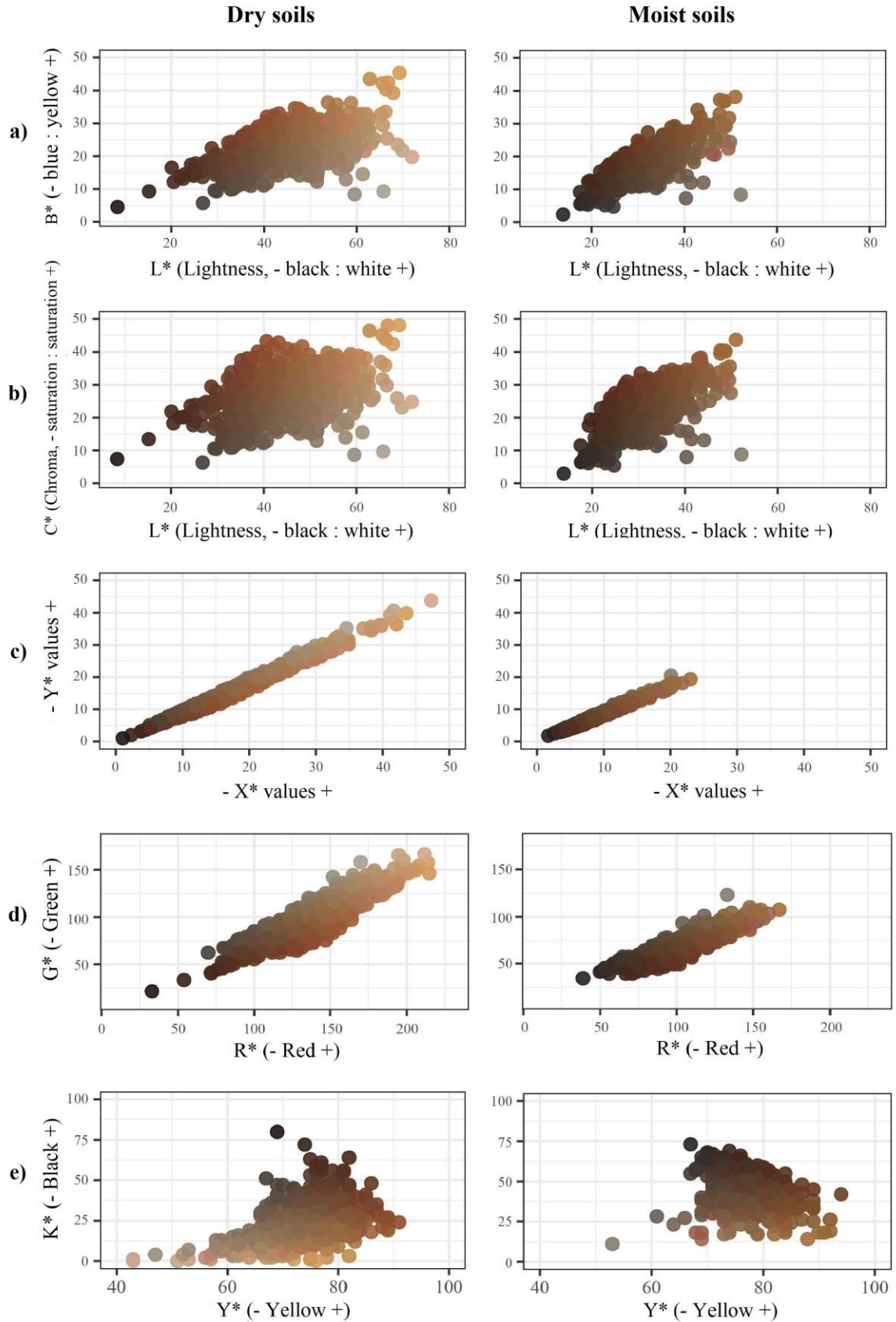
Fig. 4. Plots of variables from the CIELAB (a), CIELCH (b), XYZ (c), RGB (d) and CMYK (e) color systems obtained from Brazilian soil samples under dry and moist conditions (n = 705) using Nix Pro™ sensor. In X and Y axes, the variables of each color system that obtained the greatest ranges were chosen (see the descriptive analysis, Fig. 3).

3.1.3. Descriptive statistics for pXRF sensor

The total elemental contents acquired by pXRF also showed considerable variation with high CV value for each element, with values between 19.7% (Al) and 352.9% (Nb) (Fig. 5). This variation is likely due to mineralogy diversity caused by different parent materials, soil classes and degree of weathering as well as other factors that affect elemental contents such as land uses, SOM content and soil fertility management (Benedet et al., 2020b; Chakraborty et al., 2019; Dijair et al., 2020; Faria et al., 2020; Mancini et al., 2019; Paulette et al., 2015; Rawal et al., 2019; Teixeira et al., 2020).

The highest mean values were observed for Si followed by Al, Fe, Ti, K and Ca (Fig. 5), which are due to quartz (SiO_2) and muscovite ($KAl_2(Si_3Al)O_{10}(OH, F)_2$), the dominant minerals in the sand fraction and strongly resistant to weathering. Conversely, kaolinite ($Al_2Si_2O_5(OH)_4$), hematite (α-Fe_2O_3), goethite (α-FeOOH) and gibbsite ($Al(OH)_3$) are the main minerals in the clay fraction, whilst the silt fraction contains minerals present in both the previous fractions (sand and clay) (Araujo et al., 2014; Brinatti et al., 2010; Kampf et al., 2012; Pacheco et al., 2018; Silva et al., 2018). The abundance of these minerals in Brazilian soils explains the high values of Al and Fe observed in this study, as these elements tend to residually accumulate in weathered-leached soils where they make part of the crystalline structure of the aforementioned minerals. High contents of Si are mainly due to the presence of quartz and muscovite in the sand fraction and kaolinite in the clay fraction. These results corroborate other studies investigating soils in different regions of Brazil and also observed high contents of Si, Al and Fe via pXRF when compared with other detected and quantified elements (Benedet et al., 2020a; Borges et al., 2020; Dijair et al., 2020; Lima et al., 2019; Pelegrino et al., 2019; Silva et al., 2019a, 2019b; Teixeira et al., 2020).

The high contents of Ti are mainly due to the presence of ilmenite ($FeTiO_2$), rutile (TiO_2), both found in the sand fraction, and anatase (TiO_2 - polymorph of rutile) in the clay fraction of tropical soils (Costa et al., 2014; Kampf et al., 2012). Benedet et al. (2020a), Faria et al. (2020) and Teixeira et al. (2020) also observed that this element was regularly detected and quantified in tropical soils via pXRF.
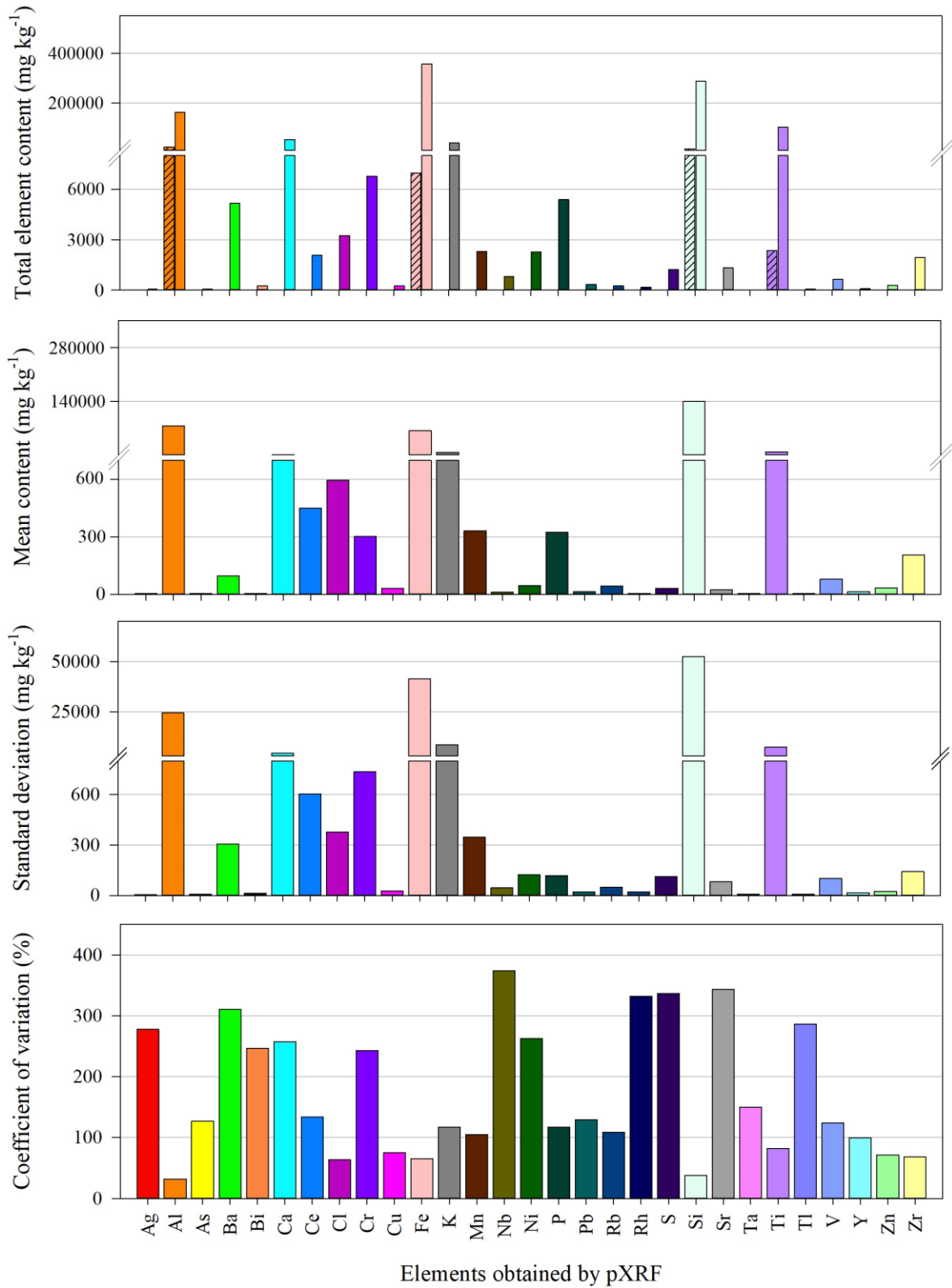
Fig. 5. Descriptive statistics for the total elemental content obtained by portable X-ray fluorescence (pXRF) spectrometry in Brazilian soil samples under dry conditions (n = 705). Hatched bars - minimum values; Unhatched bars - maximum values.

Calcium content is directly related to the agronomic management of soils under cultivated areas (e.g., application of liming and/or agricultural gypsum and mineral fertilizers), since in most Brazilian soils Ca is mainly found in exchangeable form; minerals commonly found in Brazilian soils are not a source of Ca (Araujo et al., 2014; Brinatti et al., 2010; Kampf et al., 2012; Pacheco et al., 2018; Silva et al., 2018), except for traces of Ca-feldspars in sand fractions (Kampf et al., 2003). The contents of K were also influenced by agronomic management, as well as by the presence of muscovite ($KAl_2(Si_3Al)O_{10}(OH,F)_2$) and traces of K-feldspars ($KAlSi_3O_8$) (Kampf et al., 2012).

In general, the variability of data obtained for SOM content determined by traditional laboratory methods and by the Nix Pro™ and pXRF sensors may contribute to the generation of robust and reliable SOM content prediction models, with possible application for soils of different tropical biomes.

3.2. Correlations between SOM content and soil color

The correlation between the SOM content and soil color in five color systems (CIELAB, CIELCH, XYZ, RGB and CMYK) is shown in Table 1. Results indicate that the correlation coefficient between SOM and color parameters of each color system were different under dry or moist soil condition. The analyses on moist soil samples provided relatively greater correlation values when compared with the dry soil samples. Similar results were observed by Demattê et al. (2011), Stiglitz et al. (2017a, 2017b), and Raeesi et al. (2019).

Table 1

Correlation coefficient and significance value (*p*) between soil organic matter (SOM) content and Nix Pro™ sensor parameters in dry and moist soil samples collected in Brazil.

Soil	Nix Pro™ sensor parameters														
	L	A	B	C	H	X	Y	Z	R	G	B	C	M	Y	K
Dry	-0.19 ***	-0.26 ***	-0.32 ***	-0.33 ***	0.09 *	-0.25 ***	-0.22 ***	-0.12 ***	-0.28 ***	-0.15 ***	-0.07 ns	0.35 ***	0.03 ns	-0.08 *	0.19 ***
Moist	-0.28 ***	-0.37 ***	-0.42 ***	-0.44 ***	0.03 ns	-0.32 ***	-0.28 ***	-0.13 ***	-0.36 ***	-0.21 ***	-0.06 ns	0.43 ***	-0.09 **	-0.37 ***	0.29 ***

*, **, ***Significant at $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively.

For moist soil samples, the highest negative correlation values were obtained for the parameters: C–CIELCH (represents chroma) (r = − 0.44 p < 0.001), B–CIELAB (represents

yellowness) (r = − 0.42 p < 0.001), A–CIELAB (represents redness) and Y–CMYK (represents the yellow color) (both r = − 0.37 p < 0.001), and R–RGB (represents the red color) (r = − 0.36 p < 0.001). Also, a positive correlation was obtained for parameters C–CMYK (represents the cyan color) (r = 0.43 p < 0.001) and K–CMYK (represents black) (r = 0.30 p < 0.001). These correlation values illustrate that as SOM content increases, there is an equivalent increase in dark and grayish color parameters relative to others (i.e. combination of K and C-CMYK, dark-green colors), whilst chroma values reduce. For comparison purposes, the highest correlations obtained under dry soil conditions were − 0.33 (p < 0.001), − 0.32 (p < 0.001) and − 0.26 (p < 0.001) for the parameters C (chroma), B (yellowness) and A (redness) of the CIELCH and CIELAB systems, besides a positive correlation of 0.35 (p < 0.001) for the parameter C of the CMYK system (Table 1). Thus, moist soil promotes a greater correlation between SOM and the color parameters of each color system.

Higher correlation values in moist soil samples indicate that the use of the Nix Pro™ color sensor can provide better results for the prediction of SOM content in this condition than under dry soil conditions. Chroma or saturation represents the intensity or purity of a color in relation to gray (Viscarra Rossel et al., 2006), and an increase in SOM content reduces C values and the shade of soil colors (Baumgardner et al., 1969; Demattˆe et al., 2011). A more detailed analysis shows that a reduction in the tonality of soil color parameters occurred in red (parameters A-Redness and R-Red) and yellow (parameters B-Yellowness and Y-Yellow), which are very common colors in tropical soils as previously discussed.

Stiglitz et al. (2017a, 2017b) correlated SOC and the CIELAB color system and observed that the highest correlations were − 0.63 and − 0.84 for parameters A (redness) and B (yellowness) in moist soil samples and − 0.71 and − 0.73 in dry soil samples. Mukhopadhyay et al. (2020), working with the same color system, also observed correlation values of − 0.83 and − 0.77 between SOC and L (lightness) and A (redness) parameters in dry soil samples. Correlation values of the above cited studies were higher than those obtained in this study for the CIELAB system, which were caused by the greater number of soil classes with more oxidic mineralogy, soil samples, and the use of both A and B horizons samples in the same dataset herein. In addition to the contrasting environmental conditions, these differences created larger data variability resulting in lower correlation values, corroborating the results obtained by Raeesi et al. (2019).

3.3. Correlation between SOM and elemental contents obtained by pXRF

Correlation values between SOM and elemental contents obtained by pXRF are shown in Table 2. The greatest positive correlations were obtained between SOM and calcium (Ca) (r = 0.34, p < 0.001), manganese (Mn) (r = 0.32, p < 0.001), sulfur (S) (r = 0.26, p < 0.001) and phosphorus (P) (r = 0.20, p < 0.001). The other results showed positive or negative correlation values close to zero. Most tropical soils have only traces of Ca-, S- and P-bearing minerals (Araujo et al., 2014; Brinatti et al., 2010; Kampf et al., 2012; Pacheco et al., 2018; Silva et al., 2018). Thus, most Ca, S and P detected by pXRF in soil samples is related to agronomic practices such as application of limestone and/or agricultural gypsum and mineral fertilizers (mainly phosphate), commonly used in Brazilian soils to correct superficial and/or subsuperficial acidity and improve fertility in agricultural areas (Lopes and Guilherme, 2016). In general, these agronomic practices combined with crop rotation/succession and soil conservation management/practices (e.g., no-tillage with cover crops or minimum cultivation), positively influence productivity, shoot and root biomass of cultivated plants (Li et al., 2018; Zandoná et al., 2015). Consequently, it raises the input of plant residues to the soil, which causes an increase in SOM content over time, mainly in the surface layer (0–30 cm) (Bertol et al., 2004; Castro Filho et al., 1998).

Table 2 Correlation coefficient and significance value (p) between soil organic matter (SOM) and elemental contents determined via portable X-ray fluorescence spectrometry (pXRF) in dry soil samples collected in Brazil.

Soil	pXRF-determined elemental contents														
	Ag	Al	As	Ba	Bi	Ca	Ce	Cl	Cr	Cu	Fe	K	Mn	Nb	Ni
Dry	-0.09	-0.09	-0.06	-0.06	-0.06	0.34	-0.18	-0.02	0.12	0.08	0.14	0.01	0.32	-0.04	0.05
	*	*	ns	ns	ns	***	***	ns	***	*	***	ns	***	ns	ns
	P	Pb	Rb	Rh	S	Si	Sr	Ta	Ti	Tl	V	Y	Zn	Zr	
	0.20	0.05	-0.01	-0.12	0.26	-0.06	-0.01	-0.15	0.02	-0.07	0.15	-0.01	0.13	-0.02	
	***	ns	ns	**	***	ns	ns	***	ns	ns	***	ns	***	ns	

*, **, ***Significant at p < 0.05, p < 0.01, and p < 0.001, respectively.

Manganese has been strongly correlated with pH (Faria et al., 2020) and Ca (Teixeira et al., 2020) in Brazilian agricultural soils. This might be explained by the application of liming and/or agricultural gypsum, which may contain Mn as impurity. Thus, frequent applications of limestone and gypsum may increase the contents of this element in soil, as well as pH (Valadares et al., 1974). Also, Mn-oxide minerals influence the development of tropical soil darker colors (lower values and chromas) (Carvalho Filho et al., 2011).

3.4. Model performance in SOM content prediction

Performance of models trained with the dataset acquired by the Nix Pro™ and pXRF separately and combined, in dry and moist soil samples, with and without horizons data is shown in Fig. 6. When horizons were not used as auxiliary input data, the best models for SOM content prediction were generated through the combination of data from Nix Pro™ and pXRF, with Nix Pro™ data obtained on moist soil samples. In this condition, RF outperformed the other algorithms, delivering values of $R^2 = 0.63$, RMSE = 1.40% and RPD of 1.73. Using RF only with pXRF data provided models with intermediate performance presenting values of $R^2 = 0.56$, RMSE = 1.34% and RPD = 1.60. In contrast, the worst SOM content prediction models were generated only with Nix Pro™ data, where the best model was obtained with moist soil samples through the RF algorithm with values of $R^2 = 0.33$, RMSE = 1.34% and RPD = 1.33.

The incorporation of soil horizons data into the models increased R^2 and RPD and reduced RMSE values, mainly in prediction models generated with data from each sensor (Fig. 6). However, even with improvements, the worst results were likewise produced by models generated only with data from Nix Pro™, with $R^2 = 0.50$, RMSE = 1.40% and RPD = 1.53 through the RF algorithm, using moist soil samples, while models generated only with pXRF data separately delivered superior results to those using only the Nix Pro™ via RF, delivering $R^2 = 0.64$, RMSE = 1.20% and RPD of 1.78. Lastly, combining Nix Pro™ and pXRF data produced the best models in both dry and moist conditions. The RF algorithm generated the model with the best performance using moist soil samples ($R^2 = 0.73$, RMSE = 1.09% and RPD = 2.00).

Figs. 6 and 7 suggest that the use of sensors separately tend to underestimate SOM content in soils with high contents of this attribute, regardless the algorithm, the use of soil horizon data as predictor variables or the moisture condition of the samples. However, this trend did not occur with the combination of sensors in moist soil samples via the best performing algorithm (RF), even with soil horizons as auxiliary input data. This approach accurately predicted high and low SOM contents, as the scatter points and the 1:1 line were closer than those based on single sensor datasets.
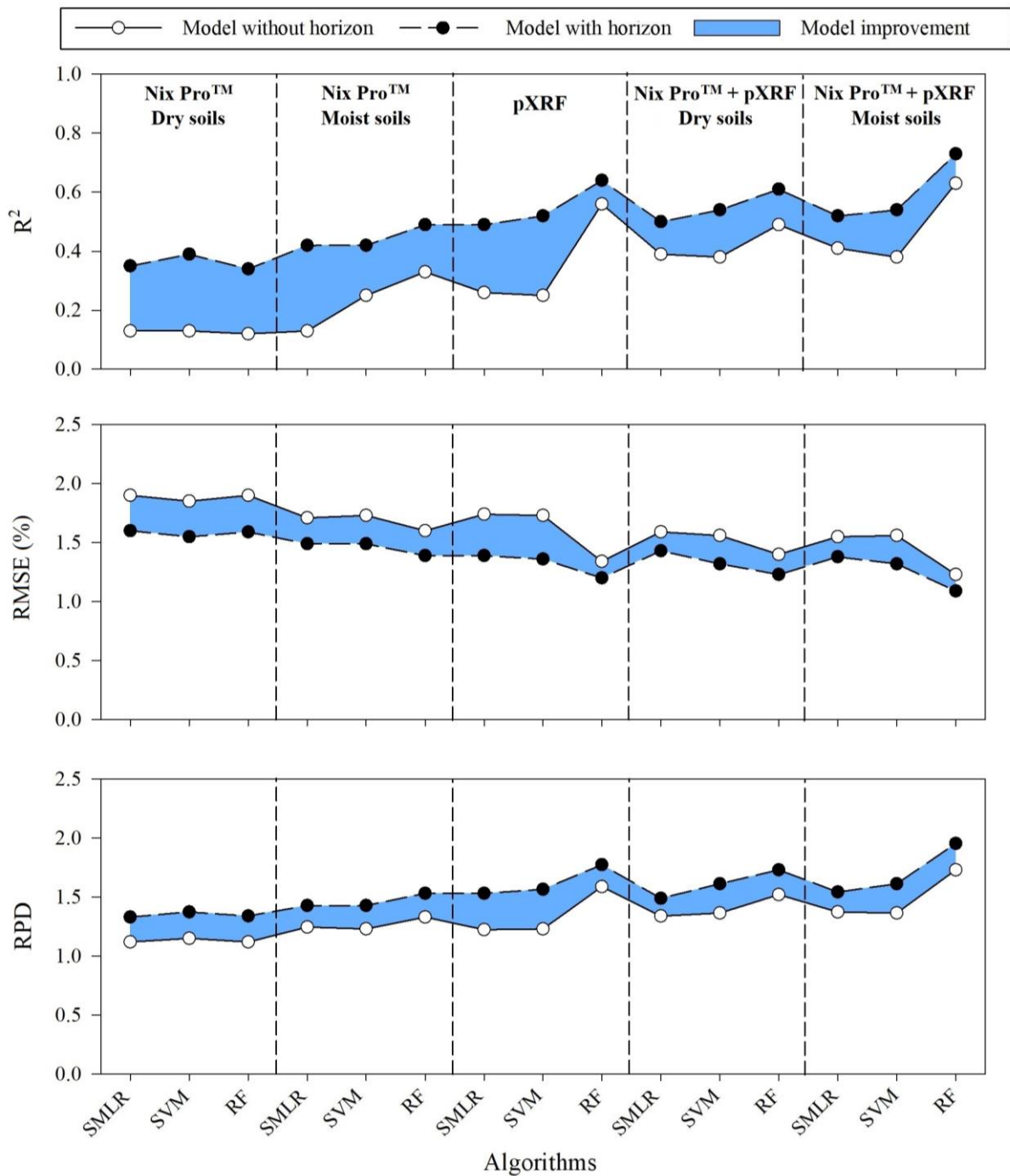
Fig. 6. Coefficient of determination (R^2), root mean square error (RMSE) and residual prediction deviation (RPD) for soil organic matter (SOM) content prediction models in Brazilian soil samples. SMLR - stepwise multiple linear regression; SVM - support vector machine with linear core; RF - random forest.
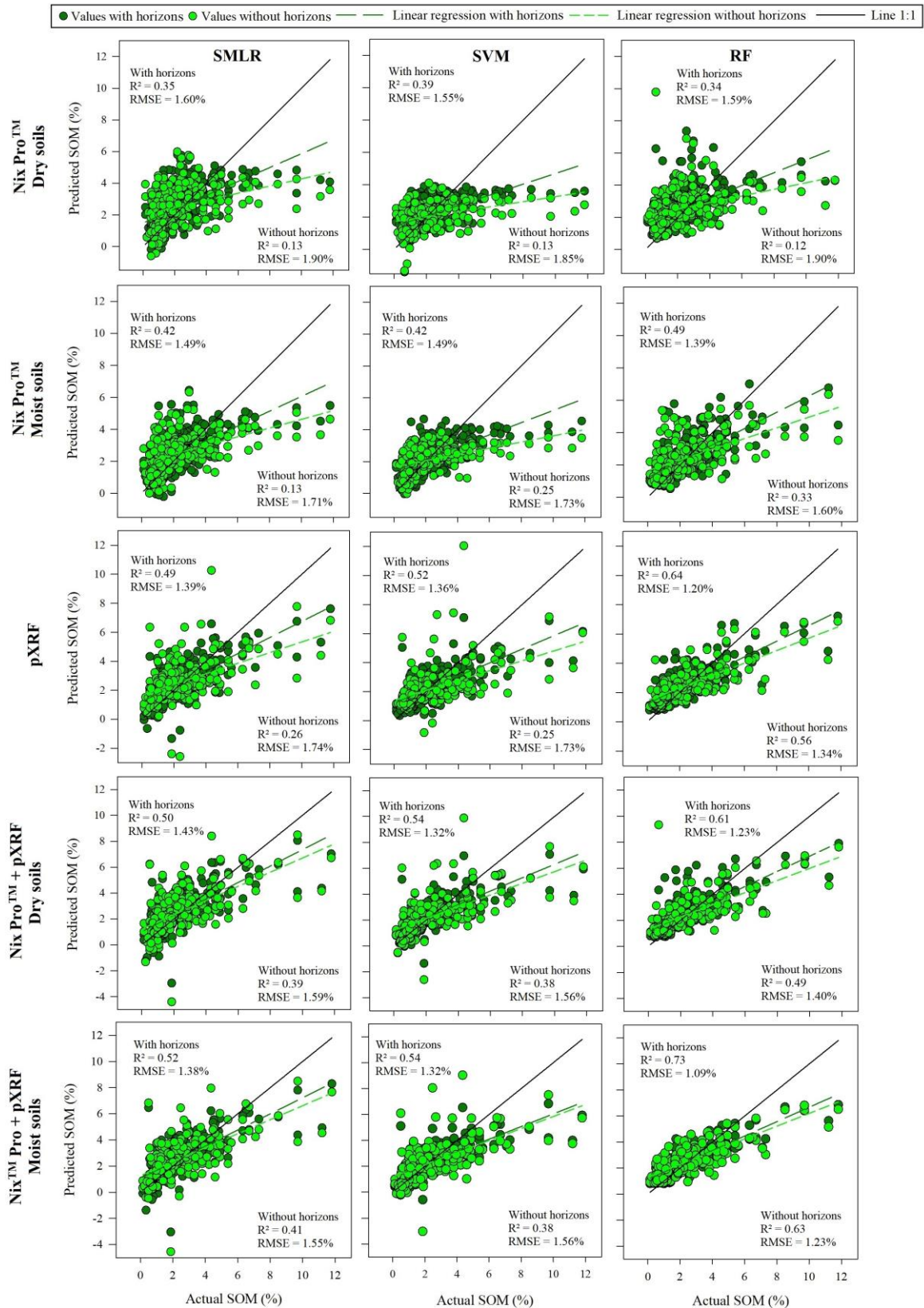
Fig. 7. Scatter plots between the soil organic matter (SOM) obtained from traditional laboratory analysis and content predicted by models trained with data obtained by Nix Pro™ and portable X-ray fluorescence (pXRF) spectrometry separately and combined, in dry and moist soil samples, with or without the introduction of soil horizons data as a predictor

variable in Brazilian soil samples. SMLR - stepwise multiple linear regression; SVM - support vector machine with linear core; RF - random forest.

The Nix Pro™ color sensor has been more commonly used in soil science since 2016 (Stiglitz et al., 2016, 2017a) by demonstrating that Nix Pro™ can successfully generate models capable of predicting SOM and SOC contents. The authors concluded that only the L (lightness) and A (redness) parameters of the CIELAB color system combined with soil sampling depth were sufficient to generate SOC content prediction models with values of R^2 = 0.80 for dry soil samples and R^2 = 0.73 for moist soil samples through multiple linear regression from soils of the USA. Similar results were obtained by Mukhopadhyay et al. (2020), who obtained R^2 values of 0.70 for SOC content prediction models in dry soil samples in India; however, the authors did not include data from soil horizons in the model. In contrast, Raeesi et al. (2019), used Nix Pro™ data to predict SOM content in soils of arid and semi-arid areas in Iran and observed that the best prediction model was generated using the RF algorithm for moist soil samples with an R^2 of 0.62. In the present study, using only Nix Pro™ data, the best R^2 value was 0.50 via the RF algorithm in moist soil samples and with the addition of soil horizons data as an auxiliary variable. This result was considerably lower than those reported in the literature, regardless of soil moisture condition and the algorithm used, probably due to the large variability of soil colors used herein, associated with the much greater content of Fe-oxide minerals and their strong pigmenting power, mainly hematite. Conceivably, models developed per soil color (e.g., red, red-yellow, yellow, gray, black, etc.) could generate better results using only Nix Pro™.

The use of only pXRF data for prediction of SOM or SOC content can also be found in the literature. Wang et al. (2015) concluded that pXRF data were able to predict SOC content with an R^2 of 0.77 via RF in soils from the USA. However, for soils in tropical regions, Silva et al. (2017), Andrade et al. (2020b) and Tavares et al. (2020b) obtained prediction models for SOM with R^2 of 0.60 (Random Forest), 0.59 (eXtreme Gradient Boosting) and 0.61 (multiple linear regression), respectively. Conversely, sensor data combination (Nix Pro™ + pXRF) tested in this study was sufficient to obtain good predictions with an R^2 = 0.73, RMSE = 1.09% and RPD = 2.00 using the RF for moist soil samples with soil horizons as auxiliary input data; a promising achievement for tropical soils research. These findings also highlight the importance of soil moisture conditions in the generation of reliable and robust models for the prediction of SOM content in tropical soils.

The lower accuracy of predictive models using data only from Nix Pro™ in soils from different regions of the world reinforces that additional tests are important to achieve an ideal methodology to provide the best results for each environmental condition. The present study, to the authors' knowledge, is the first attempt to use data from Nix Pro™ in the generation of models capable of predicting SOM content in Brazilian soils. However, the results obtained herein reinforce that the information from a single proximal sensor is sometimes not enough to obtain promising and reliable predicted results due to the complex nature and large variability of SOM contents (Wan et al., 2019, 2020), mainly for Brazilian soils, which present a peculiar simple mineralogy with a much higher proportion of Fe-oxide minerals, mainly hematite, than temperate soils. This unique condition requires the combination of proximal sensors and robust algorithms for predicting SOM content with greater accuracy. The number of studies exploring this approach is constantly growing in Brazil and worldwide (Andrade et al., 2020c, 2020d; Benedet et al., 2020a, 2020b; O'Rourke et al., 2016; Wan et al., 2019; Wang et al., 2015; Xu et al., 2019).

3.5. Importance of predictor variables for the best model

Prediction results showed that the combination of data obtained by both sensors, using soil horizons data as a predictor variable, generated the best prediction model via the RF algorithm, with an $R^2 = 0.73$ in a large and variable dataset (Figs. 6 and 7). The relative importance of the predictor variables calculated using the percent of increment in mean square error (%IncMSE) for the best SOM content prediction model is shown in Fig. 8. In general, the most important variables for predicting SOM content were those with the highest correlations with this attribute (Tables 1 and 2), as discussed in the previous sections. Thus, the most important variable for the model was soil horizons, followed by Ca, chroma (C – CIELCH), Si, yellowness (B–CIELAB), and Mn (Fig. 8). As SOM content decreases with soil depth (Jobbagy and Jackson, 2000; Li et al., 2020), it explains the importance of the soil horizons as an auxiliary variable for the models. The importance of soil horizons data in SOM content prediction models has already been reported in studies conducted by Mikhailova et al. (2017) and Stiglitz et al. (2017a), but only in models developed for temperate soils. The importance of the color variable chroma relates to the decrease in chroma values as SOM increases, and B – which represents color variation from blue to yellow – might be associated with changes in saturation caused by SOM content variation.
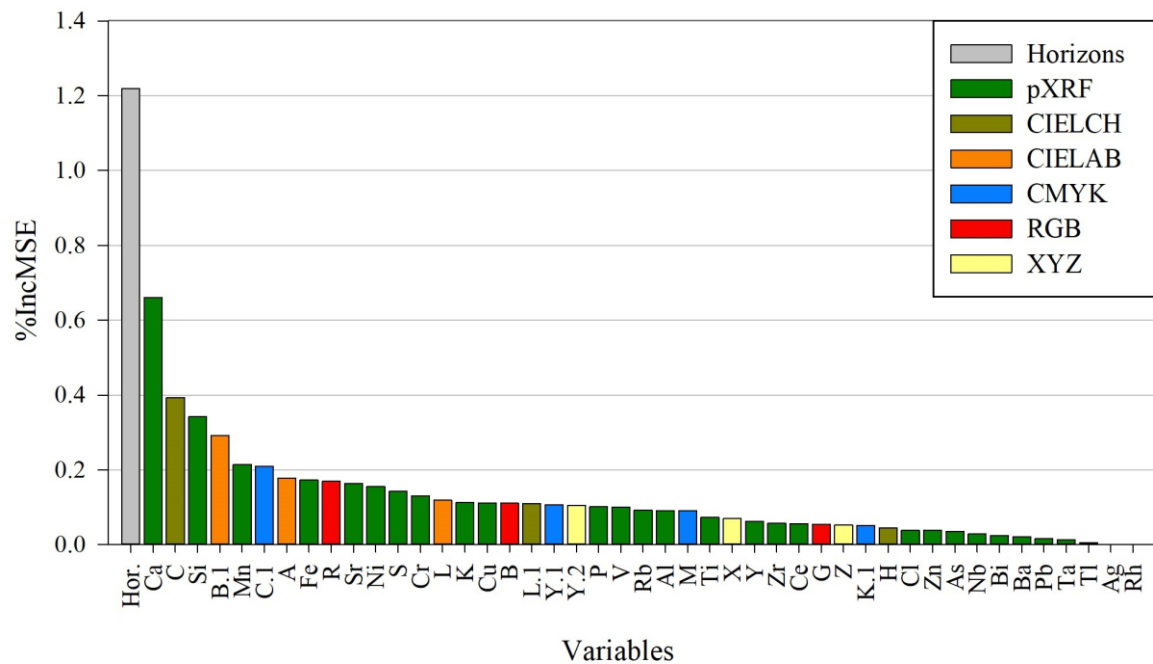
Fig. 8. Importance of variables in decreasing order for the model built using the random forest (RF) algorithm trained from combined data obtained by Nix Pro™ and portable X-ray fluorescence (pXRF) spectrometry, in moist Brazilian soil samples, combined with soil horizons as an auxiliary input data.

The importance of Ca and Mn is related to the application of limestone and/or agricultural gypsum to correct surface and/or subsurface soil acidity (Lopes and Guilherme, 2016). In addition, Mn-oxide minerals favor darker colors of tropical soils. The importance of Si is mainly associated with quartz, a mineral very resistant to weathering, commonly found in the coarse fractions of tropical soils, besides kaolinite in the clay fraction. Thus, sandy soils are largely dominated by quartz and showed the lowest contents of SOM, due to the dominance of sparse vegetation in these areas, causing little deposition of plant biomass to the soil, and high decomposition rates through microorganisms, associated with great porosity. Conversely, clayey soils, due to the higher specific surface area, tend to present a greater accumulation of SOM (Dijair et al., 2020; Kampf et al., 2012; Silva et al., 2019a, 2019b; Silva et al., 2020), taking into account more available microsites for strong interaction between these soil particles.

The importance of the color variable chroma relates to the decrease in chroma values as SOM increases, especially when soil is moist (Baumgardner et al., 1969; Demattˆe et al., 2006; Demattê et al., 2011; Shields et al., 1968; Shonk et al., 1991). Variables L (lightness of

color) and K (black) were not as relevant as variables associated with red/ yellow colors and saturation, showing that the discoloring caused by SOM was more significant to predictions than color brightness (Table 1). This might indicate a trend for very weathered-leached tropical soils, naturally presenting greater color saturation due to the much higher content of Fe-oxide minerals.

Finally, Nix Pro™ alone was not sufficient to generate reliable prediction models for SOM content in tropical soils regardless of the tested algorithms, soil moisture conditions and the introduction of sampled horizon as auxiliary data. These findings differ from those reported by Stiglitz et al. (2017a) using Nix Pro™ in temperate soils. However, it is advantageous to use the combined data of Nix Pro™ + pXRF, along with the addition of soil horizons information as an auxiliary input data, in the generation of robust models through the RF algorithm, capable of predicting SOM content in tropical soils, mainly in moist soils. Thus, despite the large variability of soil classes, sampling depths, land uses and sampling sites, the best prediction model for SOM (Random Forest) in this study with tropical soils was classified according to Chang et al. (2001) and Viscarra Rossel et al. (2010) as good and with satisfactory predictive capacity, since it presents values of $R^2 = 0.73$, RMSE = 1.09% and RPD = 2.00. This indicates the reliability of this method for general SOM content assessments and motivates further investigations on the use of this approach in new studies about soil carbon dynamics and predictions via proximal sensors. The prospects for more specific uses of SOM modeling require further research, but the results herein demonstrate the possibility of SOM prediction even in complex and highly variable soil samples. Advantages of this approach include reduced time to obtain SOM results in comparison to traditional wet-chemistry laboratory analyses, minimal sample preparation, elimination of chemical waste (environmentally-friendly method) and reduction in costs of analyses (Benedet et al., 2021).

4. Conclusions

SOM content in Brazilian soils was predicted with accuracy using pXRF and Nix Pro™ sensors ($R^2 = 0.73$, RMSE = 1.09% and RPD = 2.00) (analyses via Nix Pro™ were performed in moist soil samples). These results were attained via the RF algorithm with combined data from the Nix Pro™ and pXRF sensors and using soil horizons as auxiliary input data. However, the use of sensors data separately was not sufficient to generate reliable SOM content prediction models, regardless of soil moisture conditions, probably because the soils were rife with Fe-oxide minerals, mainly hematite. For tropical soils, adding horizons as

auxiliary input data and using moist soil samples in the analyses via Nix Pro™ yielded promising results in all evaluated models. The most important predictor variable for SOM content was soil horizons, followed by variables: Ca, chroma (C–CIELCH), Si, yellowness (B–CIELAB), and Mn. The correlations obtained between SOM and the total elemental contents obtained by pXRF showed higher values for the elements Ca, Mn, S and P. Results show the applicability of this method for general SOM prediction in tropical soils, but seemingly indicate the promising possibility of using such proximal sensors in more specific assessments of carbon in soils, especially considering more homogeneous datasets. The acquired results showed that the combination of data provided by Nix Pro™ and pXRF can accurately be used to model and predict SOM content for a wide variety of Brazilian soils, including different soil classes, parent materials, sampling depths, land uses and sampling sites.

5. Acknowledgments

References

Akaike, H., 1973. Information theory and an extension of maximum likelihood principle. In: Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory, Akademia Kiado, Budapest. pp. 267–281.

Alvares, C.A., Stape, J.L., Sentelhas, P.C., Moraes G.J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorologische Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Andrade, R, Faria, W., Silva, S., Chakraborty, S., Weindorf, D., Mesquita, L., Guilherme, L., Curi, N., 2020a. Prediction of soil fertility via portable X-ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. Geoderma 357, 113960. https://doi.org/10.1016/j.geoderma.2019.113960

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

https://10.1016/j.geoderma.2019.113957

Andrade, R., Silva, S.H.G., Faria, W.M., Poggere, G.C., Barbosa, J.Z., Guilherme, L.R.G., Curi, N., 2020c. Proximal sensing applied to soil texture prediction and mapping in Brazil. Geoderma Regional 23, e00321. https://doi.org/10.1016/j.geodrs.2020.e00321

Andrade, R., Silva, S.H.G., Weindorf, D.C., Chakraborty, S., Missina, W., Guilherme, L.R.G., Curi, N., 2020d. Tropical soil order and suborder prediction combining optical and X-ray approaches. Geoderma Regional 23, e00331. https://doi.org/10.1016/j.geodrs.2020.e00331

Anjos, L.H.C., Jacomine, P.T.K., Santos, H.G., Oliveira, V.A., Oliveira, J.B., 2012. Sistema brasileiro de classificação de solos. In: Ker, J.C., Curi, N., Schaefer, C.E.G.R., Vidal-Torrado, P. (Eds.), Pedologia - Fundamentos. SBCS, Viçosa. pp. 303–343.

Araujo, M.A., Pedroso, A.V., Amaral, D.C., Zinn, Y.L., 2014. Paragênese mineral de solos desenvolvidos de diferentes litologias na região sul de Minas Gerais. Revista Brasileira de Ciencia do Solo 38, 11–25.

Baumgardner, M.F., Kristof, S., Johannsen, C.J., Zachary, A., 1969. Effects of organic matter on the multispectral properties of soils. Agricultural Experiment Station Journal 3939, 413–422.

Baumgardner, M.F., Kristof, S.J., Johannsen, C.J., Zachary, A.L., 1970. The effects of organic matter on multispectral properties of soils. Proceedings of the Indiana Academy of Science 79, 413–422.

Benedet, L., Acuna-Guzman, S.F., Faria, W.M., Silva, S.H.G., Mancini, M., Teixeira, A.F.S., Pierangeli, L.M.P., Acerbi Júnior, F.W., Gomide, L.R., Padua 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. https://doi.org/10.1016/j.catena.2020.105003

Benedet, L., Faria, W.M., Silva, S.H.G., Mancini, M., Demattê, J.A.M., Guilherme, L.R.G., Curi, N., 2020a. Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. Geoderma 376. https://doi.org/10.1016/j.geoderma.2020.114553

Benedet, L., Faria, W.M., Silva, S.H.G., Mancini, M., Guilherme, L.R.G., Demattê, J.A.M., Curi, N., 2020b. Soil subgroup prediction via portable X-ray fluorescence and visible near-infrared spectroscopy. Geoderma 365, 114212. https://doi.org/10.1016/j.geoderma.2020.114212

Bertol, I., Albuquerque, J.A., Leite, D., Amaral, A.J., Zoldan Junior, W.A., 2004. Physical

soil properties of conventional tillage and no-tillage, in crop rotation and succession, compared with natural pasture. Revista Brasileira de Ciencia do Solo 28, 155–163.

Borges, C.S., Weindorf, D.C., Nascimento, D.C., Curi, N., Guilherme, L.R.G., Carvalho, G.S., Ribeiro, B.T., 2020. Comparison of portable X-ray fluorescence spectrometry and laboratory-based methods to assess the soil elemental composition: Applications for wetland soils. Environmental Technology & Innovation 19, 100826. https://doi.org/10.1016/j.eti.2020.100826

Broadbent, F. E. (1953). The soil organic fraction. Advances in Agronomy, 153–183. doi:10.1016/s0065-2113(08)60229-1

Brodský, L., Vašát, R., Klement, A., Zádorová, T., Jakšík, O., 2013. Uncertainty propagation in VNIR reflectance spectroscopy soil organic carbon mapping. Geoderma 199, 54–63. http://dx.doi.org/10.1016/j.geoderma.2012.11.006

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Brinatti, A.M., Mascarenhas, Y.P., Pereira, V.P., Partiti, C.S.M., Macedo, Á., 2010. Mineralogical characterization of a highly-weathered soil by the rietveld method. Scientia Agricola 67, 454–464. https://doi.org/10.1590/s0103-90162010000400013

Cardinael, R., Chevallier, T., Cambou, A., Beral, C., Barthes, B.G., Dupraz, C., Chenu, C., 2017. Increased soil organic carbon stocks under agroforestry: a survey of six different sites in France. Agriculture, Ecosystems & Environment 236, 243–255.

Carvalho Filho, A., Curi, N., Marques, J.J.G.S.M., Shinzato, E., Freitas, D.A.F., Jesus, E.A., Massahud, R.T.R., 2011. Óxidos de Manganês em Solos do Quadrlátero Ferrífero (MG). Revista Brasileira de Ciencia do Solo 35, 793–804.

Castro Filho, C., Muzilli, O., Padanoschi, A.L., 1998. Soil aggregate stability and its relation with organic carbon in a typic haplorthox, as a function of tillage systems, crop rotations and soil sample preparation. Revista Brasileira de Ciencia do Solo 22, 527–538.

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geoscientific Model Development 7, 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chakraborty, S., Li, B., Weindorf, D.C., Deb, S., Acree, A.P., Panda, P., 2019. Use of portable X-ray fluorescence spectrometry for classifying soils from different land use land cover systems in India. Geoderma 338, 5–13. https://doi.org/10.1016/j.geoderma.2018.11.043

Chang, C., Laird, D.A., Mausbach, M.J., Hurburgh, C.R.J., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Science

Society of America Journal 65, 480–490. https://doi.org/10.2136/sssaj2001.652480x

Chatterjee, A., Lal, R., Wielopolski, L., Martin, M.Z., Ebinger, M.H., 2009. Evaluation of different soil carbon determination methods. Critical Reviews in Plant Science 28, 164–178. https://doi.org/10.1080/07352680902776556

Correa, M.M., Ker, J.C., Barrón, V., Fontes, F.M.P., Torrent, J., Curi, N., 2008. Caracterização de óxidos de ferro de solos do ambiente Tabuleiros Costeiros. Revista Brasileira Ciência do Solo 32, 1017–1031. https://doi.org/10.1590/S0100-06832008000300011

Costa, S.Â.D., Ker, J.C., Simões, D.F.F., Fontes, M.P.F., Fabris, J.D., Andrade, F.V., 2014. Pedogênese e classificação de latossolos desenvolvidos de itabiritos no Quadrilátero Ferrífero, MG. Revista Brasileira de Ciencia do Solo 38, 359–371. https://doi.org/10.1590/S0100-06832014000200001

Curi, S., Silva, S.H.G., Poggere, G.C., Menezes, M.D., 2017. Mapeamento de Solos e Magnetismo no Campus da UFLA Como Traçadores Ambientais, First. ed. 147p.

Demattê, J.A.M., Bortoletto, M.A.M., Vasques, G.M., Rizzo, R., 2011. Quantificação de matéria orgânica do solo através de modelos matemáticos utilizando colorimetria no sistema munsell de cores. Bragantia 70, 590–597. https://doi.org/10.1590/S0006-87052011005000006

Demattê, J.A.M., Sousa, A.A., Alves, M.C., Nanni, M.R., Fiorio, P.R., Campos, R.C., 2006. Determining soil water status and other soil characteristics by spectral proximal sensing. Geoderma 135, 179–195.

Dijair, T.S.B., Silva, F.M., Teixeira, A.F.S., Silva, S.H.G., Guilherme, L.R.G., Curi, N., 2020. Correcting field determination of elemental contents in soils via portable X-ray fluorescence spectrometry. Ciência e Agrotecnologia 44. https://doi.org/10.1590/1413-7054202044002420

Dotto, A.C., Dalmolin, R.S.D., Caten, T.A., Grunwald, S., 2018. A systematic study on the application of scatter-corrective and spectral-derivative preprocessing for multivariate prediction of soil organic carbon by Vis-NIR spectra. Geoderma 314, 262–274. https://doi.org/10.1016/j.geoderma.2017.11.006

Duda, B.M., Weindorf, D.C., Chakraborty, S., Li, B., Man, T., Paulette, L., Deb, S., 2017. Soil characterization across catenas via advanced proximal sensors. Geoderma 298, 78–91.

Dwivedi, D., Riley, W.J., Torn, M.S., Spycher, N., Maggi, F., Tang, J.Y., 2017. Mineral properties, microbes, transport, and plant-input profiles control vertical distribution and

age of soil carbon stocks. Soil Biology and Biochemistry 107, 244–259. https://doi.org/10.1016/j.soilbio.2016.12.019

Faria, A.J.G., Silva, S.H.G., Melo, L.C.A., Andrade, R., Mancini, M., Mesquita, L.F., Teixeira, A.F.S., Guilherme, L.R.G., Curi, N., 2020. Soils of the Brazilian Coastal Plains biome : prediction of chemical attributes via portable X-ray fluorescence (pXRF) spectrometry and robust prediction models. Soil Research 59, 640–653. https://doi.org/10.1071/SR20136

Forkuor, G., Hounkpatin, O., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using Remote Sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. PLoS ONE 12, 1–21. https://doi.org/10.1371/journal.pone.0170478

Gomes, L.C., Faria, R.M., Souza, E., Veloso, G.V., Schaefer, C.E.G.R., Filho, E.I.F., 2019. Modelling and mapping soil organic carbon stocks in Brazil. Geoderma 340, 337–350. https://doi.org/10.1016/j.geoderma.2019.01.007

González, S., Herrera, F., García, S., 2015. Monotonic random forest with an ensemble pruning mechanism based on the degree of monotonicity. New Generation Computing 33, 367–388.

Hastie, T., Tibshirani, R.J., Friedman, J.H., 2009. The elements of statistical learning: data mining, inference, and prediction, second. ed. https://doi.org/10.1007/b94608

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T.F.A., McBratney, A.B., Pallasser, R., Pozza, L., 2015. Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. Geoderma 241–242, 180–209. https://doi.org/10.1016/j.geoderma.2014.11.024

Howard, P. (1965). The carbon-organic matter factor in various soil types. Oikos, 15(2), 229-236. doi:10.2307/3565121

Jia, X., Chen, S., Yang, Y., Zhou, L., Yu, W., Shi, Z., 2017. Organic carbon prediction in soil cores using VNIR and MIR techniques in an alpine landscape. Scientific Reports 7, 1–9. https://doi.org/10.1038/s41598-017-02061-z

Jobbagy, E.G., Jackson, R.B., 2000. The vertical distribution of soil organic carbon and its relation to climate and vegetation. Ecological Applications 10, 423–436.

Johnston, A.E., Poulton, P.R., Coleman, K., 2009. Soil Organic Matter. Its Importance in Sustainable Agriculture and Carbon Dioxide Fluxes, 1st ed, Advances in Agronomy. Elsevier Inc. https://doi.org/10.1016/S0065-2113(08)00801-8

Kagiliery, J., Chakraborty, S., Acree, A., Weindorf, D.C., Brevik, E.C., Jelinski, N.A., Li, B.,

Jordan, C., 2019. Rapid quantification of lignite sulfur content: Combining optical and X-ray approaches. International Journal of Coal Geology 216 https://doi.org/10.1016/j.coal.2019.103336.

Kämpf, N., Marques, J.J., Curi, N., 2012. Mineralogia de Solos Brasileiros, in: In: Pedologia Fundamentos. SBCS, Viçosa, MG. p. 343.

Kämpf, N., Woods, W.I., Sombroek, W., Kern, D.C., Cunha, T.J.F., 2003. Classification of Amazonian Dark Earths and other ancient anthropic soils. p. 77-102. In: Lehmann, J.; Kern, D.C.; Glaser, B.; Woods, W.I. (Eds.). Amazonian Dark Earths. Origin, properties and management. Kluwer Academic Publishers, Dordrecht.

Kämpf, N., Curi, N., 2000. Óxidos de ferro: Indicadores de atributos e ambientes pedogenêicos e geoquímicos. In: Novais, R.F., Alvarez V.H., Schaefer, C.E.G.R., eds. Tópicos em ciência do solo. Viçosa, MG, Sociedade Brasileira de Ciência do Solo, p.107-138.

Kheir, R.B., Greve, M.H., Bøcher, P.K., Greve, M.B., Larsen, R., McCloy, K., 2010. Predictive mapping of soil organic carbon in wet cultivated lands using classification-tree based models: The case study of Denmark. Journal of Environmental Management 91, 1150–1160. https://doi.org/10.1016/j.jenvman.2010.01.001

Komarov, A., Chertov, O., Bykhovets, S., Shaw, C., Nadporozhskaya, M. Frolov, P., Zubkova, E., 2017. RomulHum model ofsoil organic matter formation coupled with soil biota activity. I. Problem formulation, model description, and testing. Ecological Modelling 345, 113–124.

Kuhn, M., 2008. Building Predictive Models in R Using the caret Package. Journal of Statistical Software 28, 1–26. https://doi.org/10.18637/jss.v028.i05

Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., Team, R.C., Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2018. Package 'caret'. https://cran.r-project.org/web/packages/caret/caret.pdf. Accessed 30 Jul 2020.

Lal, R., Kimble, J.M., Follett, R.F., Stewart, B.S., 2001. Methods of Assessment of Soil Carbon. CRC Press, Boca Raton, FL. 696p.

Le Quéré, C., Andrew, R.M., Friedlingstein, P., Sitch, S., Pongratz, J., Manning, A.C., Et., A., 2018. Global Carbon Budget 2018. Earth System Science Data Discussions 10, 2141–2194. https://doi.org/10.5194/essd-10-2141-2018

Li, Q., Li, A., Dai, T., Fan, Z., Luo, Y., Li, S., Yuan, D., Zhao, B., Tao, Q., Wang, C., Li, B., Gao, X., Li, Y., Li, H., Wilson, J.P., 2020. Depth-dependent soil organic carbon

dynamics of croplands across the Chengdu Plain of China from the 1980s to the 2010s. Global Change Biology 26, 4134–4146. https://doi.org/10.1111/gcb.15110

Li, Y., Cui, S., Chang, S.X., Zhang, Q., 2018. Liming effects on soil pH and crop yield depend on lime material type , application method and rate , and crop species : a global meta-analysis. 128-137 19, 1393–1406. https://doi.org/10.1007/s11368-018-2120-2

Li, Z., Liu, C., Dong, Y., Chang, X., Nie, X., Liu, L., Zeng, G., 2017. Response of soil organic carbon and nitrogen stocks to soil erosion and land use types in the Loess hilly–gully region of China. Soil & Tillage Research 166, 1–9.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2, 18–22.

Lima, M.T., Weindorf, D.C., Curi, N., Guilherme, L.R.G., Lana, R.M.Q., Ribeiro, B.T., 2019. Geoderma Elemental analysis of Cerrado agricultural soils via portable X-ray fluorescence spectrometry : Inferences for soil fertility assessment. Geoderma 353, 264–272. https://doi.org/10.1016/j.geoderma.2019.06.045

Lopes, A.S., Guilherme, L.R.G., 2016. A career perspective on soil management in the Cerrado Region of Brazil. Advances in Agronomy 137, 1–72.

Lorenz, K., Lal, R., 2016. Environmental Impact of Organic Agriculture, Advances in Agronomy. Elsevier Inc. https://doi.org/10.1016/bs.agron.2016.05.003

Mancini, M., Weindorf, D.C., Chakraborty, S., Silva, S.H.G., Teixeira, A.F.S., Guilherme, L.R.G., Curi, N., 2019. Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. Geoderma 337, 718–728. https://doi.org/10.1016/j.geoderma.2018.10.026

Mancini, M., Weindorf, D.C., Monteiro, M.E.C., Faria, Á.J.G., Teixeira, A.F.S., Lima, W., Lima, F.R.D., Dijair, T.S.B., Marques, F.D.A., Ribeiro, D., Silva, S.H.G., Chakraborty, S., Curi, N., 2020. From sensor data to Munsell color system: Machine learning algorithm applied to tropical soil color classification via Nix[TM] Pro sensor. Geoderma 375, 114471. https://doi.org/10.1016/j.geoderma.2020.114471

McCarty, G.W., Reeves, J.B., 2006. Comparison of IR and MIR diffuse reflectance spectros-copy for field-scale measurement of soil fertility parameters. Soil Science 171, 94–102. https://doi.org/10.1097/01.ss.0000187377.84391.54

McDowell, M.L., Bruland, G.L., Deenik, J.L., Grunwald, S., Knox, N.M., 2012. Soil total carbon analysis in Hawaiian soils with visible, near-infrared and mid-infrared diffuse reflectance spectroscopy. Geoderma 189–190, 312–320. https://doi.org/10.1016/j.geoderma.2012.06.009

Mikhailova, E.A., Stiglitz, R.Y., Post, C.J., Schlautman, M.A., Sharp, J.L., Gerard, P.D., 2017. Predicting Soil Organic Carbon and Total Nitrogen in the Russian Chernozem from Depth and Wireless Color Sensor Measurements. Eurasian Soil Science 50, 1414–1419. https://doi.org/10.1134/S106422931713004X

Moritsuka, N., Kawamura, K., Tsujimoto, Y., Rabenarivo, M., Andriamananjara, A., Rakotoson, T., Razafimbelo, T., 2019. Comparison of visual and instrumental measurements of soil color with different low-cost colorimeters. Soil Science and Plant Nutrition 65, 605–615. https://doi.org/10.1080/00380768.2019.1676624

Mukhopadhyay, S., Chakraborty, S., Bhadoria, P.B.S., Li, B., Weindorf, D.C., 2020. Assessment of heavy metal and soil organic carbon by portable X-ray fluorescence spectrometry and NixPro^{TM} sensor in landfill soils of India. Geoderma Regional 20, e00249. https://doi.org/10.1016/j.geodrs.2019.e00249

Munõz-Rojas, M., Abd-Elmabod, S.K., Zavala, L.M., De la Rosa, D., Jordan, A., 2017. Climate change impacts on soil organic carbon stocks ofMediterranean agricultural areas: a case study in Northern Egypt. Agriculture, Ecosystems & Environment 238, 142–152. https://doi.org/10.1016/j.agee.2016.09.001

Muñoz, J.D., Kravchenko, A., 2011. Soil carbon mapping using on-the-go near infrared spectroscopy, topography and aerial photographs. Geoderma 166, 102–110. https://doi.org/10.1016/j.geoderma.2011.07.017

Nawar, S., Mouazen, A.M., 2019. On-line vis-NIR spectroscopy prediction of soil organic carbon using machine learning. Soil and Tillage Research 190, 120–127. https://doi.org/10.1016/j.still.2019.03.006

O'Rourke, S.M., Minasny, B., Holden, N.M., McBratney, A.B., 2016. Synergistic use of Vis-NIR, MIR, and XRF spectroscopy for the determination of soil geochemistry. Soil Science Society of America Journal 80, 888. https://doi.org/10.2136/sssaj2015.10.0361

Pacheco, A.A., Ker, J.C., Schaefer, C.E.G.R., Fontes, M.P.F., Andrade, F. V., Martins, E.D.S., Oliveira, F.S.D., 2018. Mineralogy, micromorphology, and genesis of soils with varying drainage along a hillslope on granitic rocks of the Atlantic Forest. Revista Brasileira de Ciencia do Solo 42, e0170291.

Paulette, L., Man, T., Weindorf, D.C., Person, T., 2015. Rapid assessment of soil and contaminant variability via portable X-ray fluorescence spectroscopy: Copşa Mică, Romania. Geoderma 243–244, 130–140. https://doi.org/10.1016/j.geoderma.2014.12.025

Pelegrino, M.H.P., Weindorf, D.C., Silva, S.H.G., de Menezes, M.D., Poggere, G.C., Guilherme, L.R.G., Curi, N., 2019. Synthesis of proximal sensing, terrain analysis, and

parent material information for available micronutrient prediction in tropical soils. Precision Agriculture 20, 746–766. https://doi.org/10.1007/s11119-018-9608-z

Poggere, G.C., Inda, A.V., Barrón, V., Kämpf, N., Brito, A.D.B., Barbosa, J.Z., Curi, N., 2018. Maghemite quantification and magnetic signature of Brazilian soils with contrasting parent materials. Applied Clay Science 161, 385–394. https://doi.org/10.1016/j.clay.2018.05.014

R Core Team, 2018. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.

Raeesi, M., Zolfaghari, A.A., Yazdani, M.R., Gorji, M., Sabetizade, M., 2019. Prediction of soil organic matter using an inexpensive colour sensor in arid and semiarid areas of Iran. Soil Research 57, 276–286. https://doi.org/10.1071/SR18323

Raij, B. V., 1969. A capacidade de troca de cátions das frações orgânica e mineral em solos. Bragantia 28, 85–112.

Ramos, P.V., Inda, A.V., Barrón, V., Siqueira, D.S., Marques Júnior, J., Teixeira, D.D.B., 2020. Color in subtropical Brazilian soils as determined with a Munsell chart and by diffuse reflectance spectroscopy. Catena 193, 104609. https://doi.org/10.1016/j.catena.2020.104609

Rawal, A., Chakraborty, S., Li, B., Lewis, K., Godoy, M., Paulette, L., Weindorf, D.C., 2019. Determination of base saturation percentage in agricultural soils via portable X-ray fluorescence spectrometer. Geoderma 338, 375–382. https://doi.org/10.1016/j.geoderma.2018.12.032

Resende, M., Curi, N., Ker, J.C., Rezende, S.B., 2011. Mineralogy of Brazilian soils: interpretation and applications. Editora UFLA: Lavras. 201p.

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. https://doi.org/10.1590/1413-70542017413000117

Santos, R.D., Lemos, R.C., Santos, H.G., Ker, J.C., Anjos, L.H.C., Shimizu, S.H., 2015. Manual de descrição e coleta de solo no campo. 7. edição revisada e ampliada. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 170 p.

Santos, H.G., Carvalho Júnior, W., Dart, R.O., Áglio, M.L.D., Sousa, J.S., Pares, J.G., Fontana, A., Martins, A.L.S., Oliveira, A.P.O., 2011. O novo mapa de solos do Brasil: legenda atualizada, Empresa Brasileira de Pesquisa Agropecuária. 67p.

Santos, H.G., Jacomine, P.K.T., Anjos, L.H.C., Oliveira, V.A., Lumbreras, J.F., Coelho, M.R., Almeida, J.A., Araujo Filho, J.C., Oliveira, J.B., Cunha, T.J.F., 2018. Sistema

Brasileiro de Classificação de Solos. 5th. Embrapa Solos, Brasília revista e ampliada ed. 353 p.

Sharma, A., Weindorf, D.C., Wang, D.D., Chakraborty, S., 2015. Characterizing soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). Geoderma 239, 130–134. https://doi.org/10.1016/j.geoderma.2014.10.001

Shields, J., Paul, E., Arnaud, R., Head, W., 1968. Spectrophotometric measurement of soil color and its relationship to moisture and organic matter. Canadian Journal of Soil Science 48, 271–280.

Shonk, G.A., Gaultney, L.D., Schulze, D.G., Van Scoyoc, G.E., 1991. Spectroscopic sensing of soil organic matter content. Transactions - American Society of Agricultural Engineers 34, 1978–1984.

Silva, E.A., Weindorf, D.C., Silva, S.H.G., Ribeiro, B.T., Poggere, G.C., Carvalho, T.S., Gonçalves, M.G.M., Guilherme, L.R.G., Curi, N., 2019. Advances in tropical soil characterization via portable X-ray fluorescence spectrometry. Pedosphere 29, 468–482. https://doi.org/10.1016/S1002-0160(19)60815-5

Silva, F.M., Weindorf, D.C., Silva, S.H.G., Silva, E.A., Ribeiro, B.T., Guilherme, L.R.G., Curi, N., 2019. Tropical soil toposequence characterization via pXRF spectrometry. Soil Science Society of America Journal 83, 1153–1166. https://doi.org/10.2136/sssaj2018.12.0498

Silva, J.E., Lemanski, J., Resck, D.V.S., 1994. Perdas de matéria orgânica e suas relações com a capacidade de troca de catiônica em solos da região de Cerrados do oeste baiano. Revista Brasileira de Ciência do Solo 18, 541–547.

Silva, S.H.G., Ribeiro, B.T., Guerra, M.B.B., Carvalho, H.W.P., Lopes, G., Carvalho, G.S., Guilherme, L.R.G., Resende, M., Mancini, M., Curi, N., Rafael, R.B.A., Cardelli, V., Cocco, S., Corti, G., Chakraborty, S., Weindorf, D.C. 2021. pXRF in tropical soils: Methodology, applications, achievements and challenges. Advances Agronomy. 167, p. 1-62. https://doi.org/10.1016/bs.agron.2020.12.001

Silva, S.H.G., Silva, E.A., Poggere, G.C., Guilherme, L.R.G., Curi, N., 2018. Tropical soils characterization at low cost and time using portable X-ray fluorescence spectrometer (pXRF): Effects of different sample preparation methods. Ciência e Agrotecnologia 42, 80–92. https://doi.org/10.1590/1413-70542018421009117

Silva, S.H.G., Teixeira, A.F.S., Menezes, M.D., Guilherme, L.R.G., Moreira, F.M. S., Curi, N., 2017. Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). Ciência e

Agrotecnologia 41, 648–664. https://doi.org/10.1590/1413-70542017416010317

Silva, S.H.G., Weindorf, D.C., Pinto, L.C., Faria, W.M., Acerbi Junior, F.W., Gomide, L.R., Mello, J.M., Pádua Junior, A.L., de Souza, I.A., Teixeira, A.F. dos S., Guilherme, L.R.G., Curi, N., 2020. Soil texture prediction in tropical soils: A portable X-ray fluorescence spectrometry approach. Geoderma 362. https://doi.org/10.1016/j.geoderma.2019.114136

Soil Survey Staff, 2014. Keys to Soil Taxonomy, 12th ed. ed. USDA, Washington, DC.

Stanley, P.L., Rowntree, J.E., Beede, D.K., DeLonge, M.S., Hamm, M.W., 2018. Impacts of soil carbon sequestration on life cycle greenhouse gas emissions in Midwestern USA beef finishing systems. Agricultural Systems 162, 249–258. https://doi.org/10.1016/j.agsy.2018.02.003

Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., Sharp, J., 2017a. Using an inexpensive color sensor for rapid assessment of soil organic carbon. Geoderma 286, 98–103. https://doi.org/10.1016/j.geoderma.2016.10.027

Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., Sharp, J., 2016. Evaluation of an inexpensive sensor to measure soil color. Computers and Electronics in Agriculture 121, 141–148. https://doi.org/10.1016/j.compag.2015.11.014

Stiglitz, R., Mikhailova, E., Post, C., Schlautman, M., Sharp, J., Pargas, R., Glover, B., Mooney, J., 2017b. Soil color sensor data collection using a GPS-enabled smartphone application. Geoderma 296, 108–114. https://doi.org/10.1016/j.geoderma.2017.02.018

Tavares, T.R., Molin, J.P., Nunes, L.C., Alves, E.E.N., Melquiades, F.L., Carvalho, H.W.P., Mouazen, A.M., 2020a. Effect of x-ray tube configuration on measurement of key soil fertility attributes with XRF. Remote Sensing 12. https://doi.org/10.3390/rs12060963

Tavares, T.R., Mouazen, A.M., Alves, E.E.N., Santos, F.R., Melquiades, F.L., Carvalho, H.W.P., Molin, J.P., 2020b. Assessing soil key fertility attributes using a portable X-ray fluorescence: A simple method to overcome matrix effect. Agronomy 10. https://doi.org/10.3390/agronomy10060787

Teixeira, A.F. S., Pelegrino, M.H.P., Faria, W.M., Silva, S.H.G., Gonçalves, M.G.M., Acerbi Júnior, F.W., Gomide, L.R., Pádua Júnior, A.L., Souza, I.A., Chakrabortyd, S., Weindorfe, D.C., Guilherme, L.R.G., Curi, N., 2020. Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry. Geoderma 361, 114132. https://doi.org/10.1016/j.geoderma.2019.114132

Valadares, J.M.A., Bataglia, O.C., Furlani, P.R., 1974. Estudo de materiais calcários usados como corretivo do solo no Estado de São Paulo. IQ - Determinação de Mo, Co, Cu, Zn,,

Mn e Fe. Bragantia 33, 147–152.

Verdade, F.C., 1956. Influência da matéria orgânica na capacidade de troca de cátions do solo. Bragantia 15, 35–42.

Viscarra Rossel, R.A., Adamchuk, V.I., Sudduth, K.A., McKenzie, N.J., Lobsey, C., 2011. Proximal soil sensing: an effective approach for soil measurements in space and time, Advances in Agronomy 113, 237-282. https://doi.org/10.1016/B978-0-12-386473-4.00005-1

Viscarra Rossel, R.A., Fouad, Y., Walter, C., 2008. Using a digital camera to measure soil organic carbon and iron contents. Biosystems Engineering 100, 149–159. https://doi.org/10.1016/j.biosystemseng.2008.02.007

Viscarra Rossel, R.A., McBratney, A.B., Minasny, B., 2010. Proximal soil sensing. 1st ed. Springer Science Business Media B.V.: Dordrecht. 446p. https://doi.org/10.1007/978-90-481-8859-8

Viscarra Rossel, R.A., Minasny, B., Roudier, P., McBratney, A.B., 2006. Colour space models for soil science. Geoderma 133, 320–337. https://doi.org/10.1016/j.geoderma.2005.07.017

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.

Wan, M., Hu, W., Qu, M., Li, W., Zhang, C., Kang, J., Hong, Y., Chen, Y., Huang, B., 2020. Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. Geoderma 363, 114163. https://doi.org/10.1016/j.geoderma.2019.114163

Wan, M., Qu, M., Hu, W., Li, W., Zhang, C., Cheng, H., Huang, B., 2019. Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals. Process Safety and Environmental Protection 132, 73–81. https://doi.org/10.1016/j.psep.2019.09.025

Wang, D., Chakraborty, S., Weindorf, D.C., Li, B., Sharma, A., Paul, S., Ali, M.N., 2015. Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. Geoderma 243–244, 157–167. https://doi.org/10.1016/j.geoderma.2014.12.011

Wang, X., Wang, J., Zhang, J. , 2012. Comparisons of Three methods for organic and inorganic carbon in calcareous soils of Northwestern China. PLoS One 7, e44334. https://doi.org/10.1371/journal.pone.0044334

Weindorf, D.C., Bakr, N., Zhu, Y., 2014. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications, Advances in Agronomy 128, 1-45. https://doi.org/10.1016/B978-0-12-802139-2.00001-9

Weindorf, D.C., Chakraborty, S., 2016. Portable X-ray fluorescence spectrometry analysis of soils. In: Hirmas, D. (Ed.), Methods of Soil Analysis. Soil Science Society of America, Madison, WI, pp. 1–8. https://doi.org/10.2136/methods-soil.2015.0033.

Xu, D., Zhao, R., Li, S., Chen, S., Jiang, Q., Zhou, L., Shi, Z., 2019. Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. European Journal of Soil Science 70, 162–173. https://doi.org/10.1111/ejss.12729

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, 1467–1476.

Zandoná, R.R., Beutler, A.N., Burg, G.M., Barreto, C.F., Schmidt, M.R., 2015. Gypsum and lime increase soybean and maize yield and decrease drought stress. Pesquisa Agropecuária Tropical 45, 128–137. https://doi.org/10.1590/1983-40632015v4530301

Zhao, L., Sun, Y., Zhang, X., Yang, X., Drury, C.F., 2006. Soil organic carbon in clay and silt sized particles in Chinese mollisols: relationship to the predicted capacity. Geoderma 132, 315–323. https://doi.org/10.1016/j.geoderma.2005.04.026

ARTICLE 3 - Relationship between elemental content determined via portable X-ray fluorescence and traditional acid-digestion-based methods in tropical soils

Álvaro José Gomes de Faria[a], Sérgio Henrique Godinho Silva[a], Leônidas Carrijo Azevedo Melo[a], Lívia Botelho[a], Luiz Roberto Guimarães Guilherme[a], Nilton Curi[a,]

[a]Department of Soil Science, Federal University of Lavras – UFLA, Doutor Sylvio Menicucci Avenue, Lavras, Minas Gerais State, 37200-900, Brazil. E-mail addresses: ajgomesdefaria@hotmail.com, sergio.silva@ufla.br, leonidas.melo@ufla.br, livia.botelho@ufla.br, guilherm@ufla.br, niltcuri@ufla.br.

Abstract

Context: Studies comparing the elemental contents obtained via portable X-ray fluorescence (pXRF) spectrometry under different scanning conditions and traditional time-consuming concentrated acid-digestion-based methods are rare in tropical soils. **Aims**: i) To compare the contents of Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr obtained by pXRF with the results of the USEPA 3051a method; and ii) to evaluate the impacts of different conditions on the pXRF results. **Methods**: Soil samples were analyzed with pXRF under five conditions: field (F), post-field (PF), air-dried fine earth (ADFE, < 2 mm), macerated (M), and macerated and sieved (MS, <150 µm). Linear regressions were adjusted between the USEPA 3051a and pXRF results for the five conditions. Coefficient of determination (R²), root-mean-square error (RMSE), and residual prediction deviation (RPD) were used as validation parameters for the models. **Key results**: The different scanning conditions provided contrasting results for the total elemental contents via pXRF. The validation generated good models, as indicated by the condition, R^2, and RPD values, and the results were, respectively: Ca (M; 0.88; 3.00), Cu (F; 0.91; 3.29), Fe (ADFE; 0.94; 4.14), Mn (F; 0.85; 2.65), Cr (ADFE; 0.86; 2.77), and Ni (ADFE and M; 0.74 for both; 2.10 and 2.08). **Conclusions**: PXRF can accurately determine the contents of Ca, Cu, Fe, Mn, Cr, and Ni in tropical soils compared with the 3051a method.

Implications: Results of the USEPA 3051a method can be accurately predicted via pXRF and regression models, reducing time, cost, and work required.

Keywords: Tropical soils, chemical attributes modeling, ICP atomic emission spectroscopy, pedology, geochemistry.

1. Introduction

Soil is a fundamental, slowly-renewable natural resource for many human activities, sustaining life on the planet and aiding the economic development of nations (Adhikari and Hartemink 2016). Soil characterization provides a solid support for decision-making related to engineering, geochemical mapping, environmental monitoring, and food production (Zhang and Hartemink 2019). Consequently, strategies such as *in situ* observations, sampling, and laboratory analyses with different levels of complexity for soil characterization are fundamental (Shepherd and Walsh 2002; Evanylo and McGuinn 2009; Zhang and Hartemink 2019).

Soil elemental composition is an important chemical attribute for both agronomic purposes and environmental monitoring, because it affects soil reactions, salinity, cation exchange capacity, plant nutrient availability, and heavy metal contamination (Towett *et al.* 2015; Weindorf and Chakraborty 2020), among others. Traditionally, total elemental content in soil has been determined by concentrated acid-digestion-based methods, such as USEPA 3050b, 3051a and 3052, followed by elemental content determination via atomic absorption spectrometry (AAS) or inductively coupled plasma optical emission spectrometry (ICP-OES) (Chen and Ma 1998; USEPA 2007a; Silva *et al.* 2014; Weindorf and Chakraborty 2020). Although these methods have been widely used, they require intensive, time-consuming, and laborious preparation of soil samples. Moreover, these methods involve the use of strong acids (e.g., HNO_3, H_2SO_4, HCl, and HF) that produce hazardous chemical wastes (Chen and Ma 1998; Qu *et al.* 2019; Silva *et al.* 2019), which need to be handled properly to avoid undesirable pollution. Thus, novel, environmentally friendly methods for total elemental content assessment that are chemical-free and require minimal soil sample preparation are urgently needed. Proximal sensors, such as the portable X-ray fluorescence (pXRF) spectrometer, are being increasingly adopted for multiple applications in soil science and other fields of study.

The pXRF equipment has been recognized worldwide as a reliable, easy to use, fast, no cost, and environmentally friendly method for total elemental content determination of soils both in the field (*in situ*) and laboratory (*ex situ*), with minimal or no sample preparation (Horta *et al.* 2015; Stockmann et al., 2016a; Ribeiro *et al.* 2017; Feng *et al.* 2021; Silva *et al.* 2021). Considering environmental sciences, several methods of analysis methods have been proposed for characterization of soils and sediments. For instance, method 6200 (USEPA 2007b) was originally established and further improved for temperate soils by the Soil Survey Staff (2014a) and Weindorf and Chakraborty (2020), and recently for tropical soils by Silva *et al.* (2021). The pXRF has been increasingly adopted for numerous applications in soils from temperate and tropical regions for agronomic, pedological, and environmental purposes (Rouillon and Taylor 2016; Stockmann *et al.* 2016a; Mancini *et al.* 2019; Qu *et al.* 2019; Acree *et al.* 2020; Borges *et al.* 2020; Faria *et al.* 2020; Teixeira *et al.* 2020; Benedet *et al.* 2021).

Comparison of pXRF with acid-digestion-based methods for total elemental content determination has been more explored in soils from temperate regions, and strong relations between these methods have been determined (Kilbride *et al.* 2006; Radu and Diamond 2009; Peinado *et al.* 2010; McLaren *et al.* 2012; Rouillon and Taylor 2016; Horta *et al.* 2021). Conversely, few studies have been conducted for tropical soils (Silva *et al.* 2019; Borges *et al.* 2020). Silva *et al.* (2019) found strong correspondences between pXRF and the 3051a method for the elements Ca, Cu, Fe, Mn, Cr, V, and Ni, with values of R^2 ranging from 0.85 to 0.95. However, a poor correlation was obtained for Al, K, Ti, and Zr, ranging from 0.1 to 0.42. Borges *et al.* (2020) also observed strong correlations between the results of pXRF and extraction by the 3051a method, followed by determination via graphite-furnace atomic absorption spectrometry (GFAAS) for Cu ($R^2 = 0.83$), Fe ($R^2 = 0.83$), and Pb ($R^2 = 0.82$), but found poor correlations for As ($R^2 = 0.47$), Zn ($R^2 = 0.42$), and Mn ($R^2 = 0.28$). These studies, however, have not investigated the effect of sample preparation method and conditions of sample analyses on the pXRF results and their correlation with the acid-digested-based method.

Soil sample preparation in the previously mentioned studies included air-dried, ground, and sieved (<2 mm) samples for further scanning with pXRF in a laboratory (*ex situ*). However, it is well-known that pXRF results obtained under field conditions differ from those obtained in a laboratory; even sieving samples at different particle sizes may cause variation in the pXRF results (Stockmann *et al.* 2016a; Silva *et al.* 2018; Dijair *et al.* 2020), but little

has been done aimed at evaluating the effect of different sample preparation conditions on the correspondence between pXRF and 3051a results, mainly in tropical soils.

In Brazil, these investigations have been rare, and strong correlations with pXRF may contribute to reduced cost, time, the amount of the chemical residues generated, and the financial resources needed for such analyses. However, particle size distribution, moisture, and soil structure may affect pXRF results, raising questions about the best condition for analyzing samples via pXRF for correlations with the 3051a method. In this context, the aims of this pilot study involving tropical soils were to: i) evaluate the impact of different pXRF scanning conditions of soil samples, as follows: in the field (F), post-field (PF), air-dried, disaggregated, and sieved at 2 mm (ADFE), ADFE followed by grinding (M), and M sieved at 150 µm (MS); the five scanning conditions were used to determine the total elemental contents of Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr; ii) determine the relationship between the 12 elements obtained with the five scanning conditions with pXRF and the traditional acid-digested-based method (USEPA 3051a/ICP-OES) via linear regression models and validation.

2. Materials and methods

2.1 Soil sampling and analyses

This study was conducted with composite soil samples collected in Lavras, Minas Gerais State, Brazil, representing the main soil classes of the Brazil. There was no application of amendments. According to the Köppen classification system, the region's climate is framed as Cwa, with dry winters and rainy summers (Alvares *et al.* 2013). The mean annual precipitation and temperature are 1450 mm and 20.4°C (Dantas *et al.* 2007; Alvares *et al.* 2013). The study area and a flowchart of the procedures performed to obtain data with pXRF and the reference method (USEPA 3051a/ICP-OES) are shown in Fig. 1.
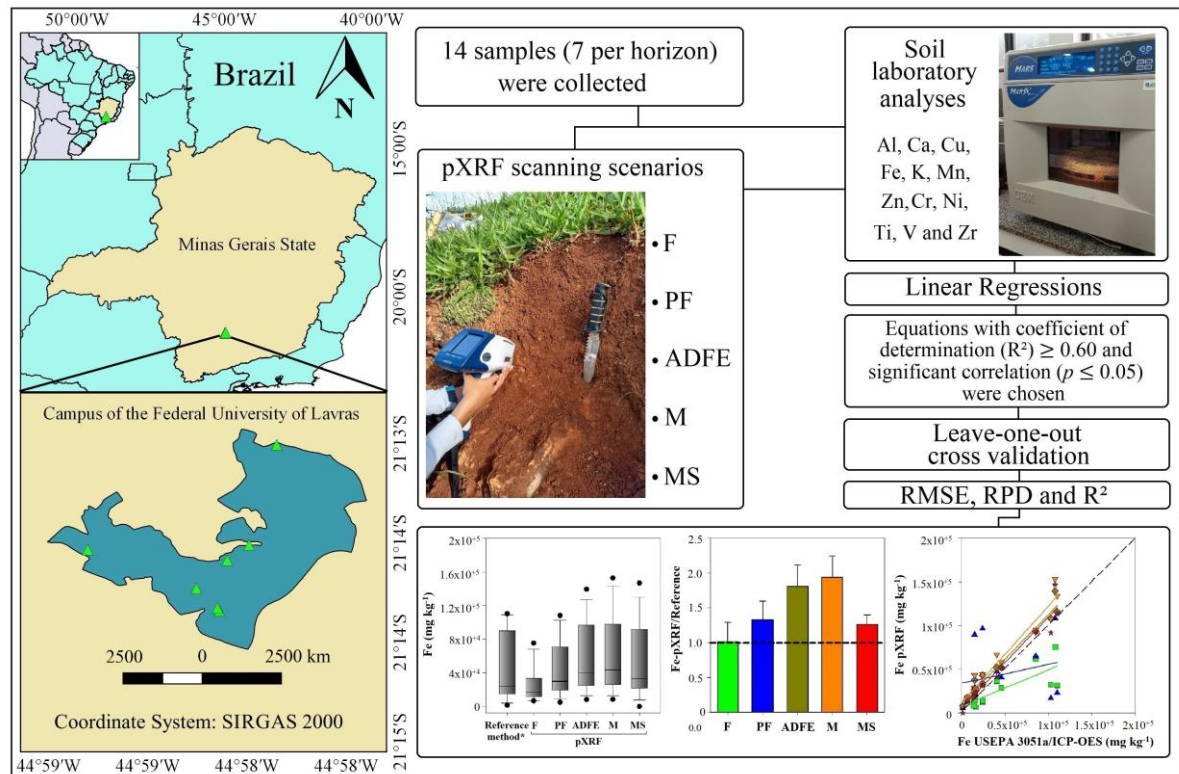
Fig. 1. Location of the sampling sites and a flowchart of the procedures conducted in this study. F, field; PF, post-field; ADFE, air-dried fine earth; M, macerated; MS, macerated and sieved. SIRGAS 2000 stands for Geocentric Reference System for the Americas, which has been the official reference system for datum of geographic coordinates since 2005 for the activities of Brazilian cartography.

Seven soil profiles were morphologically described (Santos *et al.* 2015) and classified according to the Brazilian Soil Classification System (Santos *et al.* 2018) and the *US Soil Taxonomy* (Soil Survey Staff 2014) as: Red Latosol (Anionic Acrudox), Red-Yellow Latosol (Typic Hapludox), Haplic Gleysol (Typic Endoaquent), Haplic Cambisol (Typic Dystrustept), Yellow Argisol (Inceptic Hapludult), Red Argisol (Typic Rhodudult), and Red Nitosol (Rhodic Kandiudult). The soil parent materials were gabbro (Anionic Acrudox and Rhodic Kandiudult), colluvio-alluvial sediments (Typic Endoaquent), and gneiss (the other soils) (Curi *et al.* 2017).

In each soil profile, A and B horizons (A and Cg horizons in the case of Typic Endoaquent) were analyzed via pXRF (better described below) directly on the soil profile wall (field - F – condition). Then, samples were collected from the soil profile, placed into plastic bags, and taken to the laboratory for analyses via pXRF and USEPA 3051a. In the laboratory, samples were analyzed while still moist (post-field – PF – condition). Then, they

were air-dried, disaggregated, and passed through a 2-mm sieve (air-dried fine earth - ADFE –condition) with sample moisture ranging from 1.24 to 3.84% (Fig. 2). Then, soil samples were analyzed again via pXRF, followed by grinding the ADFE samples in a mortar with agate pistil (macerated - M – condition) and another pXRF scanning was performed. In sequence, these soil samples were sieved at 150 µm (macerated and sieved - MS – condition) and analyzed again via pXRF. All pXRF analyses were performed in triplicate. After that, MS samples were analyzed following the USEPA 3051a method (USEPA 2007a).
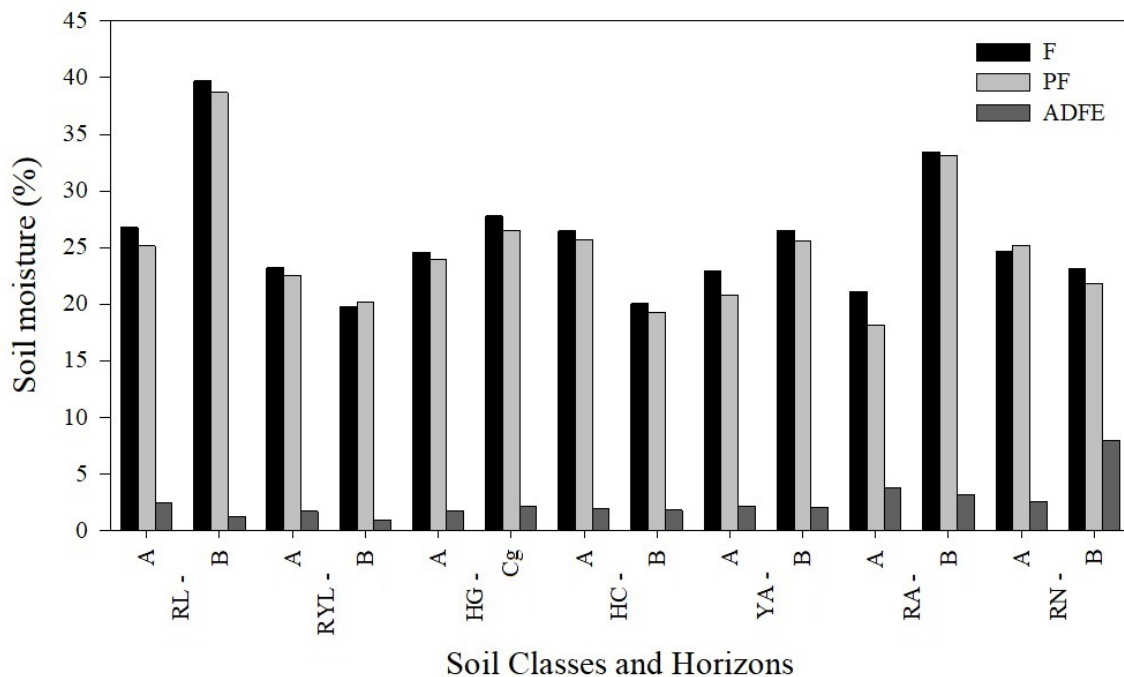
Fig. 2. Mean moisture content of soil samples under different scanning conditions with pXRF. A, A horizon; B, B horizon; Cg, Cg horizon; F, field; PF, post-field; ADFE, air-dried fine Earth; RL, Red Latosol (Anionic Acrudox); RYL, Red-Yellow Latosol (Typic Hapludox); HG, Haplic Gleysol (Typic Endoaquent); HC, Haplic Cambisol (Typic Dystrustept); YA, Yellow Argisol (Inceptic Hapludult); RA, Red Argisol (Typic Rhodudult); RN, Red Nitosol (Rhodic Kandiudult).

All soil samples were analyzed individually by a pXRF model S1 Titan LE (Bruker Nano Analytics, Kennewick, WA, USA), according to Silva *et al.* (2021) for tropical soils. Samples were scanned in triplicate in Trace (dual soil) mode integrated with the Geochem software with a scanning time of 60 s. The pXRF equipment has an X-ray tube with a thin

rhodium (Rh) window, 4 W, 15–50 keV X-ray generator with 5–100 μA, and a silicon drift detector (SDD) with typical resolution <145 eV.

Quality control of the pXRF analyses was tested by running two samples certified by the National Institute of Standards and Technology (NIST), 2710a and 2711a, and a sample certified by the pXRF manufacturer (Check Sample - CS) to determine the recovery rates for each element [% Recovery = (elementary content via pXRF/certified elementary content) x 100]. The recovery values (%) for the elements identified in all samples were (2710a/2711a/CS): Al (64/127/97), Ca (86/45/--), Cu (84/67/95), Fe (108/98/90), K (52/33/89), Mn (73/64/88), Zn (99/81/--), Cr (--/129/--), Ni (--/104/110), Ti (169/118/--), V (--/20/--), and Zr (105/--/--). Dashed lines (--) indicate either the absence of reference values for an element or its value was below the detection limit of the pXRF equipment.

For the USEPA 3051a method, 0.5 g of the MS samples were weighed, in triplicate, and transferred to 50 ml Teflon® vessels containing 5 mL of concentrated HNO_3. Samples were heated for 10 min in a microwave (CEM MARS-5, CEM Corp., Matthews, NC, USA) at 180 ºC under a pressure of 448 kPA (USEPA 2007a). After completing the digestion process, samples were filtered through a Whatman filter paper Nº. 40 and rinsed with ultrapure water until the 50 mL volume was completed. After that, the extracts were quantified by inductively coupled plasma optical emission spectrometry (ICP-OES), model Spectro Blue (Spectro Analytical Instruments, Kleve, Germany) (Soltanpour *et al.* 1996). The following elements, based on the following wavelengths (nm), were identified and quantified: Al (396.152), Ca (315.887), Cu (324.754), Fe (373.486), K (766.491), Mn (403.076), Zn (213.856), Cr (205.618), Ni (232.003), Ti (334.941), V (309.311), and Zr (343.823). A multi-elemental calibration curve was used for the quantification of elements ranging from 0.1 to 50 mg L^{-1}.

The USEPA 3051a method provides semi-total contents of soil elements and it has been a reference method for environmental monitoring analyses in Brazil, where most soils are highly weathered and leached and have very low reserves of plant nutrients, whose dominant minerals are quartz and muscovite in the sand fraction, quartz in the silt fraction, and hydroxy-interlayered vermiculite, kaolinite, gibbsite, hematite, and goethite, in different proportions, in the clay fraction (Brinatti *et al.* 2010; Kämpf *et al.* 2012). In these soils, the USEPA 3051a method presents recovery values, for most elements, comparable to more acid-aggressive methods (e.g., USEPA 3052), which employ HF, thus dissolving also silicate minerals (Silva *et al.* 2014). Thus, the USEPA 3051a method was considered as a reference method for comparisons with pXRF results obtained from the five different sampling conditions.

2.2 Dataset generated with pXRF and statistical analyses

Datasets obtained with pXRF analyses were compared with one another for elemental contents to check for variations among the soil sample conditions. These datasets included sample variations in particle size distribution, moisture degree, and soil structure, acquired both in the field (*in situ*) and in the laboratory (*ex situ*). Results from each pXRF condition were also compared with those obtained by the reference method (USEPA 3051a/ICP-OES) in order to assess which condition would be more appropriate in terms of cost and time required to process and obtain information comparable to a reference method.

Soil moisture degree (%) was also determined for F, PF, and ADFE conditions by the formula: [(WW-DW) / DW] x 100, where WW is the wet weight and DW is the dry weight obtained after 24 h at 105 °C, to verify the influence of soil moisture degree on pXRF performance under different conditions. Results of soil moisture for different soil samples are shown in Fig. 2 (Silva *et al.* 2018).

The contents of Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr obtained by the USEPA 3051a and by the pXRF methods, under the five scanning conditions, were submitted to descriptive statistics to obtain the maximum, minimum, average, and coefficient of variation (CV%) values. The elemental contents obtained from pXRF and the USEPA 3051a methods were also analyzed using boxplots to assist in the characterization and identification of intrinsic differences in elemental quantification due to soil moisture degree, particle size distribution, soil structure, and locality of analyses (field vs. laboratory).

A linear regression model (y = a + bx) was adjusted between the contents of Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr to assess the correspondence between the USEPA 3051a and pXRF methods under the different conditions. Pearson's correlation coefficients, as well as the ratios between the methods for different elemental contents, were also calculated.

For linear regression equations obtained for each element, the values of coefficient of determination (R^2, *Eq. 1*) and correlation (r) were considered as statistical parameters for evaluation. Properly adjusted equations ($R^2 \geq 0.60$) (Viscarra Rossel *et al.* 2010) with a significant correlation ($p \leq 0.05$) were validated by the leave-one-out cross validation (LOOCV) method using the caret package (Kuhn *et al.* 2018) of the R software (R Development Core Team 2019). Validation of the models was performed by R^2 (*Eq. 1*), root-mean-square error (RMSE, *Eq. 2*), and residual prediction deviation (RPD, *Eq. 3*) values,

where the highest values of R^2 and RPD and the lowest values of RMSE were considered the best ones.

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(yi - mi)^2}{\sum_{i=1}^{n}(\bar{Y} - mi)^2} \qquad\qquad Eq. 1$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi - mi)^2} \qquad\qquad Eq. 2$$

$$RPD = SD/RMSE \qquad\qquad Eq. 3$$

Where: n - number of observations, yi - value estimated by the model, mi - value obtained using the USEPA 3051a method, $\bar{Y}$ - average of values obtained by USEPA 3051a method, and SD is the standard deviation of the mean.

3. Results and discussion

3.1 Overview

Descriptive statistics for Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr obtained by the reference (USEPA 3051a) and pXRF methods under the different conditions for scanning soil samples are shown in Table 1. Regarding the 3051a method, the greatest concentrations were obtained for Fe (43,608.16 mg kg^{-1}) and Al (41,073.15 mg kg^{-1}), which reflect the predominance of highly weathered and leached soils used in this study that tend to accumulate hydroxy-interlayered vermiculite, kaolinite, and Fe and Al oxide minerals in the clay fraction (Brinatti *et al.* 2010; Resende *et al.* 2011; Kämpf *et al.* 2012; Barbosa *et al.* 2021). Moreover, the USEPA 3051a method has higher capacity to dissolve oxide minerals than silicate minerals in the soil clay fraction, because it does not include hydrofluoric acid in its methodology.

In relation to number of samples with results (SwR) for each element, six soil samples showed Ni values below the quantification limit for the USEPA 3051a method, including samples from Typic Hapludox, Inceptic Hapludult, and Typic Dystrustept, all developed from gneiss, as well as Typic Endoaquent formed from alluvial-colluvial sediments. Similar results were obtained by Silva *et al.* (2019) for soils from the same region. High Ni concentrations

are generally found in soils derived from mafic rocks (Kabata-Pendias 2010; Tsadilas *et al.* 2019), as observed in this study for the Anionic Acrudox derived from gabbro.

Table 1. Descriptive analysis for the elements Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr (mg kg^{-1}) obtained by the reference method (USEPA 3051a) and by pXRF under different soil sample conditions.

Method/ Conditions	Al					Ca				
	SwR[6]	Max.[7]	Min.[8]	Mean	CV[9]	SwR	Max.	Min.	Mean	CV
3051a	14	98,720.97	6,336.10	41,073.15	63.50	14	1,134.96	208.33	510.73	48.43
F[1]	14	59,520.28	24,616.64	38,739.41	27.59	00	<LOD	<LOD[10]	<LOD	<LOD
PF[2]	14	64,209.31	33,489.56	48,355.15	17.29	12	1,238.79	<LOD	421.01	88.46
ADFE[3]	14	160,737.71	67,016.21	109,318.87	26.46	12	1,818.08	<LOD	492.93	92.39
M[4]	14	148,011.91	68,422.45	113,524.53	22.55	13	1,771.20	<LOD	590.14	77.83
MS[5]	14	144,015.84	49,532.03	86,730.37	25.24	13	1,889.01	<LOD	468.59	97.27

Method/ Conditions	Cu					Fe				
	SwR	Max.	Min.	Mean	CV	SwR	Max.	Min.	Mean	CV
3051a	13	79.06	<LOQ[11]	19.85	123.21	14	110,178.55	1,532.47	43,608.16	92.06
F	14	81.50	13.00	27.75	72.32	14	75,442.02	6,651.45	25,819.17	80.44
PF	14	90.00	6.33	27.12	80.90	14	108,232.00	5,071.48	43,270.10	77.26
ADFE	14	147.00	11.67	40.62	88.82	14	139,337.50	8,242.38	55,163.43	75.04
M	14	165.33	16.33	45.21	88.69	14	152,762.26	8,413.63	60,007.92	75.36
MS	14	177.33	13.33	42.00	102.72	14	146,749.92	0.65	50,939.65	84.66

Method/ Conditions	K					Mn				
	SwR	Max.	Min.	Mean	CV	SwR	Max.	Min.	Mean	CV
3051a	14	703.17	334.37	516.72	20.85	14	998.10	4.93	367.80	93.63
F	14	2,759.47	1,760.98	2,280.02	15.17	13	581.66	<LOD	205.43	95.19
PF	14	4,023.84	620.56	1,722.33	58.97	13	680.35	<LOD	194.62	107.57
ADFE	14	6,848.61	660.40	2,577.58	65.47	14	1,158.68	21.16	336.87	105.43
M	14	6,880.70	622.78	2,683.23	65.97	14	909.71	18.83	333.08	88.28
MS	14	4,844.99	625.27	2,223.04	60.29	14	873.33	32.51	307.14	99.85

Method/ Conditions	Zn					Cr				
	SwR	Max.	Min.	Mean	CV	SwR	Max.	Min.	Mean	CV
3051a	14	41.20	0.36	16.71	80.36	14	1,943.80	58.56	515.87	123.44
F	14	68.50	24.00	42.94	36.87	14	1,143.00	22.67	344.69	107.48
PF	14	54.00	15.67	30.74	37.52	13	2,684.33	<LOD	714.67	138.61
ADFE	14	70.00	21.33	39.24	32.90	14	3,631.33	102.00	1,107.29	122.54
M	14	68.67	22.67	41.88	32.67	14	4,784.67	144.00	1,313.93	122.70
MS	14	82.00	20.33	40.69	38.79	14	4,146.00	95.00	1,139.97	125.39

Method/ Conditions	Ni					Ti				
	SwR	Max.	Min.	Mean	CV	SwR	Max.	Min.	Mean	CV
3051a	08	174.33	<LOQ	46.35	134.43	14	839.30	3.75	249.01	127.86
F	14	183.33	4.67	72.92	89.61	14	3,751.74	648.92	2,126.19	42.02
PF	14	238.33	4.33	85.17	96.84	14	5,745.41	2,232.87	3,608.90	36.65
ADFE	14	345.67	17.00	128.21	97.56	14	6,822.81	2,024.02	4,810.74	32.60
M	14	385.00	17.33	140.74	96.84	14	7,404.04	2,253.84	5,135.38	33.67
MS	14	409.00	13.33	127.79	102.82	14	7,493.69	2,007.65	4,314.98	39.66

Method/ Conditions	V					Zr				
	SwR	Max.	Min.	Mean	CV	SwR	Max.	Min.	Mean	CV
3051a	14	443.57	52.33	184.99	61.77	14	41.36	6.85	19.43	56.90
F	03	42.50	<LOD	6.49	208.96	14	244.67	94.33	144.82	29.50
PF	06	260.00	<LOD	56.29	163.27	14	233.67	89.00	150.57	28.52
ADFE	13	404.33	<LOD	136.86	94.52	14	341.67	130.33	198.81	28.36
M	10	393.00	<LOD	136.10	88.83	14	334.00	143.00	233.40	25.54
MS	13	344.67	<LOD	122.31	93.19	14	287.67	137.00	194.62	22.29

[1] F – Field; [2] PF – Post-Field; [3] ADFE – Air-dried fine earth, [4] M - Macerated; [5] MS – Macerated and sieved; [6] SwR – Samples with results; [7] Max. - Maximum values; [8] Min. - Minimum values; [9] CV – Coefficient of variation (%); [10] <LOD - < limit of detection; [11] <LOQ - < limit of quantification.

The pXRF data under the F condition showed the lowest average concentrations for Al (38,739.41 mg kg^{-1}), Ca (<LOD), Fe (25,819.17 mg kg^{-1}), Mn (205.43 mg kg^{-1}), Cr (344.69 mg kg^{-1}), and V (6.49 mg kg^{-1}), with values lower than the USEPA 3051a, but also lower than the other scanning conditions (Table 1). The same trend was observed for the PF condition, whose average concentrations obtained for Al (48,355.15 mg kg^{-1}), Ca (421.01 mg kg^{-1}), Fe (43,270.10 mg kg^{-1}), Mn (194.62 mg kg^{-1}), Cr (714.67 mg kg^{-1}), and V (56.29 mg kg^{-1}) were lower than the other pXRF scanning conditions and the reference method. These results show that both F and PF scanning conditions can underestimate the elemental content for these elements via pXRF, likely due to a greater moisture content (between 18.14 and 39.70%) (Fig. 2), as has been reported in other studies (Hu *et al.* 2014; Stockmann *et al.* 2016a; Dijair *et al.* 2020; Horta *et al.* 2021). The presence of water molecules between the mineral and organic particles in soil can cause the scattering of primary X-rays or absorbance of secondary X-rays, promoting an underestimation of the content of a particular element, especially those with low atomic number (Parsons *et al.* 2013; Weindorf *et al.* 2014b; Ribeiro *et al.* 2018; Santana *et al.* 2018, 2019; Padilla *et al.* 2019).

Conversely, both F and PF scanning conditions with pXRF showed greater average concentrations for Cu (27.75 and 27.12 mg kg^{-1}), K (2,280.02 and 1,722.33 mg kg^{-1}), Zn (42.94 and 30.74 mg kg^{-1}), Ni (72.92 and 85.17 mg kg^{-1}), Ti (2,126.19 and 3,608.90 mg kg^{-1}), and Zr (144.82 and 150.57 mg kg^{-1}) than the USEPA 3051a method. All samples under the F condition with pXRF showed Ca content below the limit of detection, while only two soil samples showed this limitation in the PF condition. Concerning V, 11 and 8 samples were below the limit of detection for F and PF conditions, respectively. Conversely, ADFE, M, and MS conditions revealed a much greater number of soil samples with contents above the limit of detection for such elements.

Both F and PF conditions had a similar moisture degree (Fig. 2), but contrasting results for elemental concentrations, except for Cu and Mn (Table 2). The results showed that both soil moisture content and soil structure affect the results obtained by pXRF in the PF scanning condition. Silva *et al.* (2018) also found similar results evaluating different sample preparation methods on pXRF results. The authors observed that scanning the same samples with pXRF directly in the soil profile and in the laboratory provided different concentrations

of Fe_2O_3 and TiO_2, whose diverse results were attributed to be a consequence of either the disturbance or the preservation of structure of the soil samples, because in both scanning conditions the samples had a similar moisture degree.

The M condition showed the highest average concentrations for most elements among the ADFE, M and MS, as follows: Al (113,524.53 mg kg^{-1}), Ca (590.14 mg kg^{-1}), Cu (45.21 mg kg^{-1}), Fe (60,007.92 mg kg^{-1}), K (2,683.23 mg kg^{-1}), Cr (1,313.93 mg kg^{-1}), Ni (140.74 mg kg^{-1}), Ti (5,135.38 mg kg^{-1}), and Zr (233.40 mg kg^{-1}) compared to the ADFE and MS conditions (Table 1). For Mn and V, the greatest concentrations obtained in the M condition were still lower than the USEPA 3051a method. Similarly, Silva *et al.* (2018) found the greatest concentrations for Fe_2O_3 and TiO_2 via pXRF by scanning the samples under the M condition followed by ADFE.

Soil samples with finer particle sizes tend to produce greater concentrations of elements measured via pXRF than soil samples with coarser particle sizes (Laiho and Peramaki 2005; Markowicz 2008). Thus, it was expected that the MS condition would provide the greatest elemental contents in this study, but such results were lower than those under M and ADFE conditions, which have larger particle sizes. These results indicate that grinding samples to a smaller particle size (<150 µm) does not always promote greater content reported by pXRF for all the elements evaluated. It is also worth mentioning that a good homogenization of soil samples is important before performing pXRF analyses, in order to guarantee that elements present in trace contents may be evenly distributed in the analyzed sample layer by pXRF (Markowicz 2008).

Soil samples under the ADFE, M and MS conditions had only 2, 1, and 1 samples with results below the detection limit for Ca, and 1, 4, and 1 samples with results below the limit of detection for V, respectively, via pXRF (Table 1). This suggests that for these elements a minimum soil sample preparation, e. g., drying and homogenization, is required to achieve the best performance by pXRF. Thus, reducing soil moisture content seems to be sufficient and effective for determination of such elements by this method.

As these three conditions, as cited above, have low moisture content (Fig. 2), the influence of moisture on Ca detection should be minimal (Parsons *et al.* 2013; Weindorf *et al.* 2014a). The low Ca detection under the ADFE, M, and MS conditions should be associated with the fact that most Brazilian soils tend to be deprived of Ca-bearing minerals; so, soils without application of fertilizers, lime, and agricultural gypsum, as is the case in this study, have naturally very low Ca contents (Kämpf *et al.* 2012; Resende *et al.* 2021).

The coefficient of variation (CV) ranged from 15.17% (for K) to 107.48% (for Cr) for the F condition, without considering the element V due to the low number of samples with results below the limit of detection. For the PF condition, CV ranged from 17.29% to 138.61% for Al and Cr, respectively, again without considering the element V. For the ADFE and M conditions, CV results varied between 26.46% (Al) and 122.54% (Cr), and between 22.55% (Al) and 122.70% (Cr), respectively. Finally, for the MS condition, the scans provided CV between 22.29% (Zr) and 125.39% (Cr). Thus, for most conditions evaluated with pXRF, the lowest and highest variations in results were obtained for Al and Cr, and the PF condition provided the largest range of CV values among the studied elements, reflecting the large variability of soil samples used in this study, which represent the most important soil classes in Brazil.

Boxplots show the distribution of samples per element according to the values generated by both the USEPA 3051a and pXRF methods (Fig. 3). In general, there was a high dispersion and variability of the elements obtained in each data set, mainly for the ADFE, M, and MS conditions via pXRF. This was due to the high variability of the studied soil classes (Latosol, Gleysol, Cambisol, Argisol, and Nitosol), different soil horizons, and different soil parent materials (gneiss, gabbro, and alluvial-colluvial sediments). The highest contents of Al, Ca, Cu, Fe, K, Mn, Cr, Ni, Ti, V, and Zr were found under the M scanning condition, followed by ADFE (Fig. 3). Under the F and PF scanning conditions, soil moisture had a great effect in causing underestimation and lower data dispersion for Al, Ca, Fe, Mn, Cr, and V in comparison with the USEPA 3051a method and the other pXRF scanning conditions.

Several regression models (e.g., simple linear, polynomial, power regression, and more robust algorithms) can be generated for assessing the correspondence of elements obtained with pXRF and the USEPA 3051a methods. Auxiliary variables (e.g., soil organic matter, moisture content, and texture) can be added to models in order to improve the results obtained in each condition and, consequently, approach the values generated by standard methods under laboratory conditions (Stockmann *et al.* 2016b; Ribeiro *et al.* 2018; Silva *et al.* 2018; Qu *et al.* 2019; Dijair *et al.* 2020). Simple linear models will be discussed later in the section "Elementary correspondence between pXRF and USEPA 3051a methods".
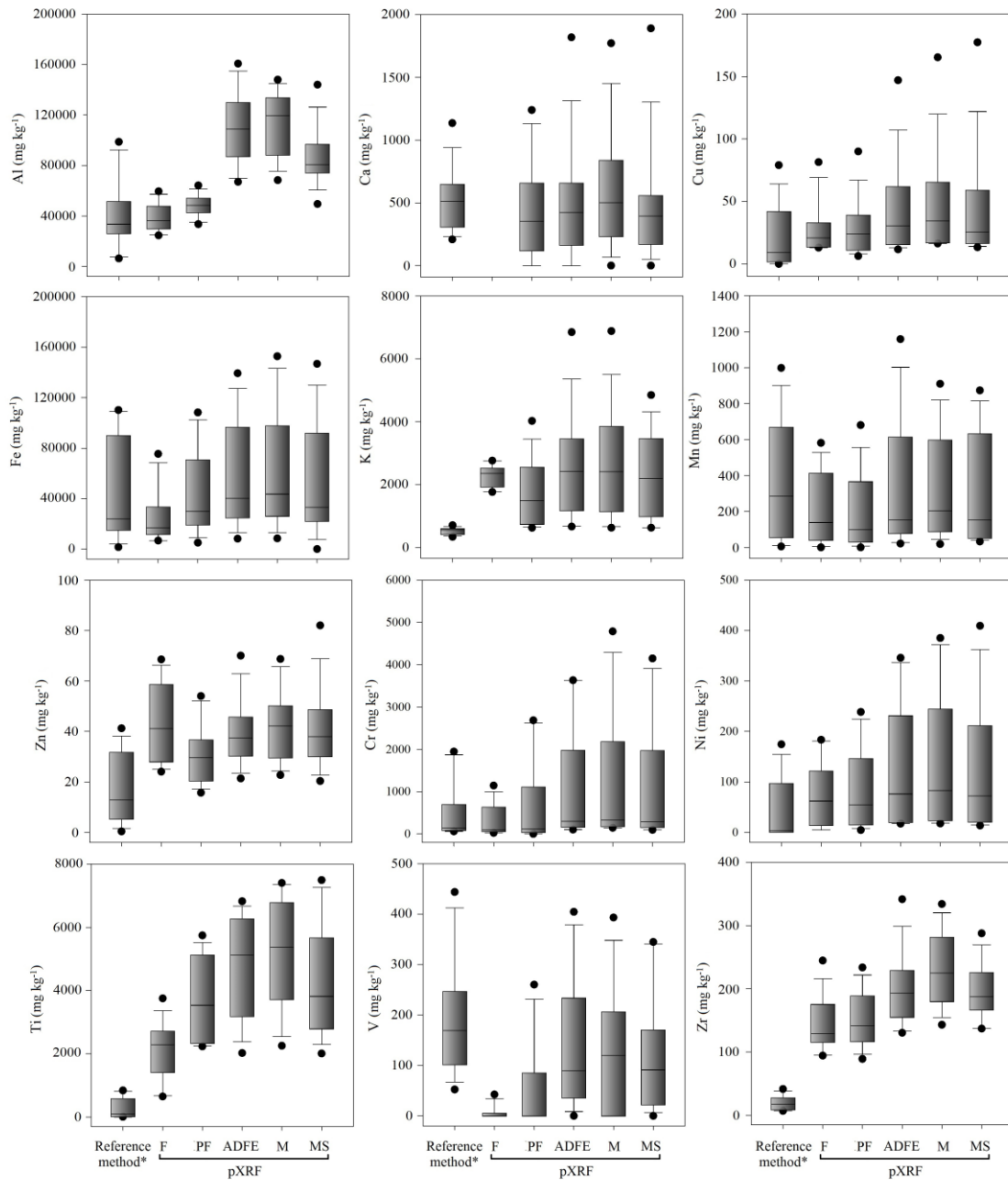
Fig. 3. Boxplots comparing the elemental content of Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V and Zr in soil samples obtained using the reference method (USEPA 3051a) and the pXRF in the field (F), post-field (PF), air-dried fine earth (ADFE), macerated (M) and macerated and sieved (MS) conditions.

3.2 Ratios between pXRF and USEPA 3051a methods

Mean ratios between contents of Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr obtained by pXRF and the USEPA 3051a methods in each scanning condition are shown in Fig. 4. In general, for Cu, K, Zn, Ti, and Zr contents, high ratios were found, regardless of the

scanning condition, which indicates that much greater concentrations were obtained for the pXRF than for the USEPA 3051a method. The highest ratios were found for Ti followed by Zr, whose values ranged from 15.8 (F) to 28.8 (M), and from 11.0 (F and PF) to 17.4 (M), respectively. In contrast, the average ratios for V were less than 0.8 in all scanning conditions with pXRF, and the lowest values were observed for the F (0.04) and PF (0.4) conditions, confirming the lower elemental contents via pXRF in such conditions.
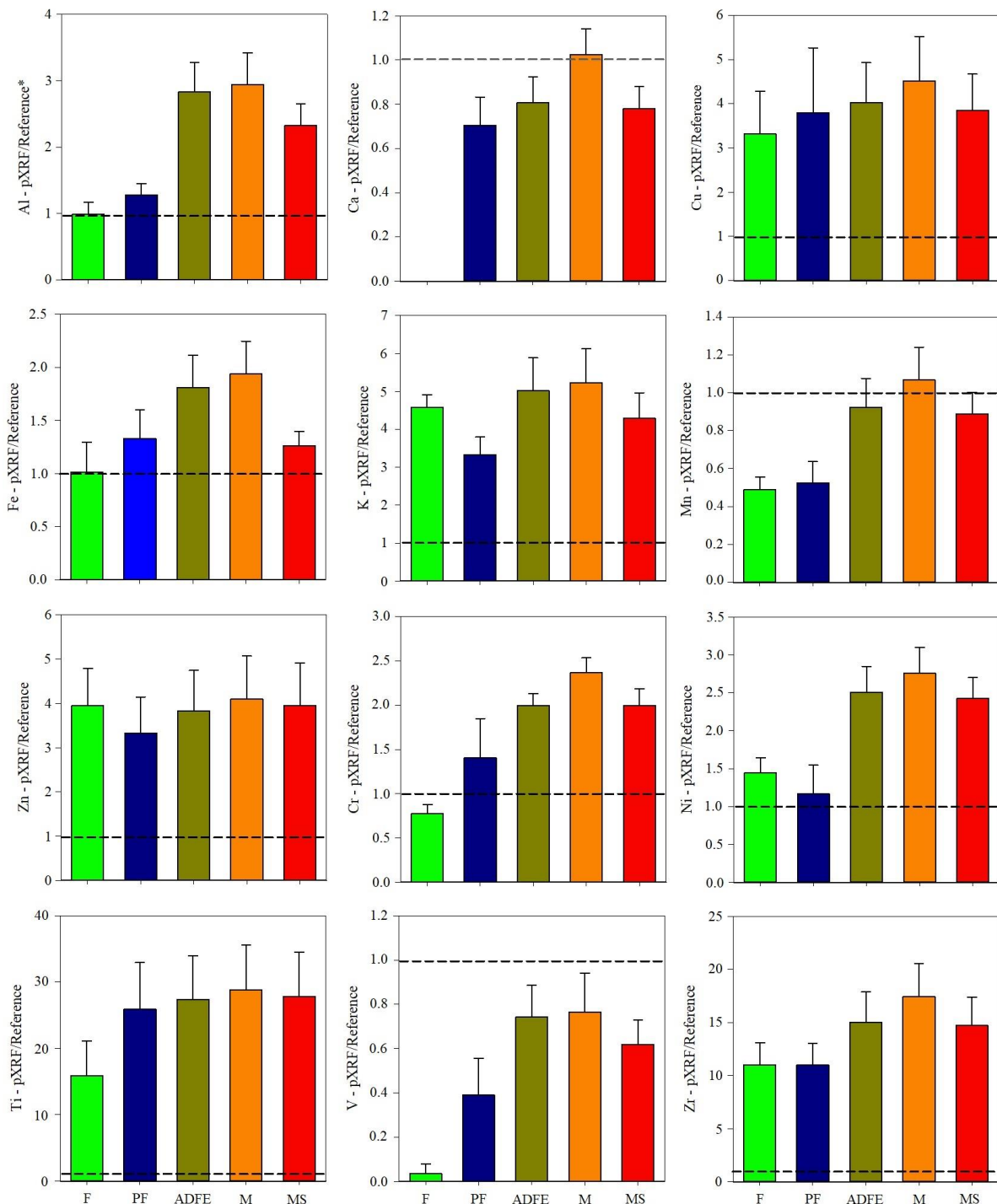
Fig. 4. Mean ratios between the reference (USEPA 3051a) and pXRF methods under different conditions for Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V and Zr in tropical soil samples. F, field; PF, post-field; ADFE, air-dried fine earth; M, macerated; MS, macerated and sieved. Error bars show the standard error (n = 14); dashed lines show the 1:1 ratio.

For Al, Fe, Cr, and Ni, ratios close to 1 were observed for pXRF only in F and PF scanning conditions, indicating that in these situations the pXRF is capable of delivering values close to those of the reference method. However, for ADFE, M, and MS conditions, high ratios were observed for these elements. For Ca, the M condition would be the most appropriate one because it provided ratios close to 1, while for Mn, the highest ratios were obtained under the ADFE and M conditions. The other conditions provided lower ratios for Ca and Mn.

Quantification of Al, Fe, Cr, and Ni contents, both under F and PF conditions, is promising, because they provide values close to the USEPA 3051a method in a fast, inexpensive, and reliable way (Fig. 4). However, for Ca and Mn, the M and ADFE conditions are recommended for scanning with pXRF. Borges *et al.* (2020) evaluated the ratios between the values obtained by the pXRF under the MP condition and the USEPA 3051a method assisted by GFAAS in hydromorphic tropical soils, similar to the Gleysol studied here, and found high ratios for Zn, Cu, Fe, As, and Mn, varying between 2.2 and 7.8, while for Pb the ratios were close to 1.

3.3 Elementary correspondence between pXRF and USEPA 3051a methods

The linear regression equations, the coefficients of determination (R^2), the correlation coefficients (r), and the *p*-values of the relationships between the elemental contents obtained by the reference (USEPA 3051a) and by the pXRF methods under various conditions are shown in Table 2, while Fig. 5 shows the graphical representation for each element, whose dispersion points (n = 14) are distributed around the 1:1 line. In general, an adequate linear regression for each element is totally dependent on the soil scanning condition via pXRF. Only 10 out of 60 linear regression models were interpreted as good (R^2 = 0.61 to 0.80), while 19 models were classified as very good ($R^2 = \geq 0.81$) (highlighted in bold in Table 3), according to Viscarra Rossel *et al.* (2010). The prediction accuracy of these models for each element through validation will be discussed later.

Table 2. Linear equations, coefficient of determination (R²), and correlation (r) for the relations between elemental contents obtained via the reference method (USEPA 3051a) and pXRF under different soil sampling conditions.

Elements	Conditions	Linear Equation	R²	R	p - value
Al	F[1]	Y = 52255.5546 - 0.3291x	**0.65**	-0.80	<0.01
	PF[2]	Y = 44791.0042 + 0.0868x	0.07	0.27	ns
	ADFE[3]	Y = 124097.9723 - 0.3598x	0.11	-0.32	ns
	M[4]	Y = 123673.5857 - 0.2471x	0.06	-0.25	ns
	MS[5]	Y = 93708.0767 - 0.1699x	0.04	-0.20	ns
Ca	F	--	--	--	--
	PF	Y = -186.1206 + 1.2269x	**0.70**	0.87	<0.01
	ADFE	Y = -414.0340 + 1.7753x	**0.91**	0.92	<0.01
	M	Y = -300.6874 + 1.7533x	**0.91**	0.96	<0.01
	MS	Y = -429.1067 + 1.7486x	**0.87**	0.94	<0.01
Cu	F	Y = 12.0655 + 0.7901x	**0.93**	0.96	<0.01
	PF	Y = 17.2756 + 0.4958x	0.31	0.55	<0.05
	ADFE	Y = 12.4335 + 1.4198x	**0.93**	0.96	<0.01
	M	Y = 14.2441 + 1.5600x	**0.91**	0.95	<0.01
	MS	Y = 9.4384 + 1.6402x	**0.87**	0.93	<0.01
Fe	F	Y = 7703.2874 + 0.4154x	**0.65**	0.80	<0.01
	PF	Y = 34037.1126 + 0.2117x	0.07	0.25	ns
	ADFE	Y = 10965.8920 + 1.0135x	**0.97**	0.98	<0.01
	M	Y = 12179.8055 + 1.0968x	**0.95**	0.97	<0.01
	MS	Y = 5598.2291 + 1.0397x	**0.94**	0.97	<0.01
K	F	Y = 1821.9826 + 0.8864x	0.08	0.28	ns
	PF	Y = -92.4834 + 3.5122x	0.14	0.37	ns
	ADFE	Y = 106.0205 + 4.7831x	0.09	0.31	ns
	M	Y = -37.0965 + 5.2646x	0.10	0.32	Ns
	MS	Y = -355.3052 + 4.9898x	0.16	0.40	Ns
Mn	F	Y = 16.2718 + 0.5233x	**0.87**	0.94	<0.01
	PF	Y = 31.3278 + 0.4538x	0.56	0.76	<0.01
	ADFE	Y = 23.2822 + 0.8526x	**0.68**	0.83	<0.01
	M	Y = 66.9078 + 0.7237x	**0.72**	0.85	<0.01
	MS	Y = 2.7429 + 0.8280x	**0.86**	0.93	<0.01
Zn	F	Y = 27.1763 + 0.9432x	**0.64**	0.80	<0.01
	PF	Y = 30.2531 + 0.0290x	0.00	0.03	ns
	ADFE	Y = 29.7183 + 0.5696x	0.35	0.59	<0.05
	M	Y = 31.7185 + 0.6081x	0.36	0.60	<0.05
	MS	Y = 29.4263 + 0.6740x	0.33	0.57	<0.05
Cr	F	Y = 154.2734 + 0.3698x	0.40	0.63	<0.05
	PF	Y = 945.0711 -0.3190x	0.04	-0.15	ns
	ADFE	Y = 50.5874 + 2.0524x	**0.92**	0.96	<0.01
	M	Y = 66.0106 + 2.4237x	**0.91**	0.96	<0.01
	MS	Y = 71.9094 + 2.0737x	**0.85**	0.93	<0.01
Ni	F	Y = 37.0918 + 0.8899x	**0.86**	0.92	<0.01
	PF	Y = 68.0372 - 0.0951x	0.01	-0.05	ns
	ADFE	Y = 59.1555 + 1.7463x	**0.84**	0.95	<0.01
	M	Y = 65.5612 + 1.8996x	**0.84**	0.94	<0.01

	MS	Y = 50.2901 + 1.8700x	**0.82**	0.94	<0.01
	F	Y = 2156.5606 - 0.1220x	0.00	-0.04	ns
	PF	Y = 3791.8750 - 0.7348x	0.03	-0.18	ns
Ti	ADFE	Y = 4069.8662 + 2.9753x	0.37	0.60	<0.05
	M	Y = 4182.6988 + 3.8259x	0.50	0.70	<0.01
	MS	Y = 3187.6807 + 4.5271x	**0.71**	0.84	<0.01
	F	--	--	--	--
	PF	Y = 87.3052 - 0.2666x	0.04	-0.06	ns
V	ADFE	Y = -9.6339 + 0.8426x	**0.61**	0.76	<0.01
	M	Y = 53.2672 + 0.6487x	**0.69**	0.79	<0.01
	MS	Y = -22.1618 + 0.8010x	**0.67**	0.83	<0.01
	F	Y = 184.6614 - 2.0507x	0.28	-0.53	ns
	PF	Y = 165.4130 - 0.7640x	0.04	-0.20	ns
Zr	ADFE	Y = 247.5426 - 2.5085x	0.24	-0.49	ns
	M	Y = 289.2736 - 2.8758x	0.28	-0.53	<0.05
	MS	Y = 241.3783 - 2.4069x	0.38	-0.61	<0.05

[1] F – Field; [2] PF – Post-Field; [3] ADFE – Air-dried fine earth; [4] M - Macerated; [5] MS – Macerated and sieved.
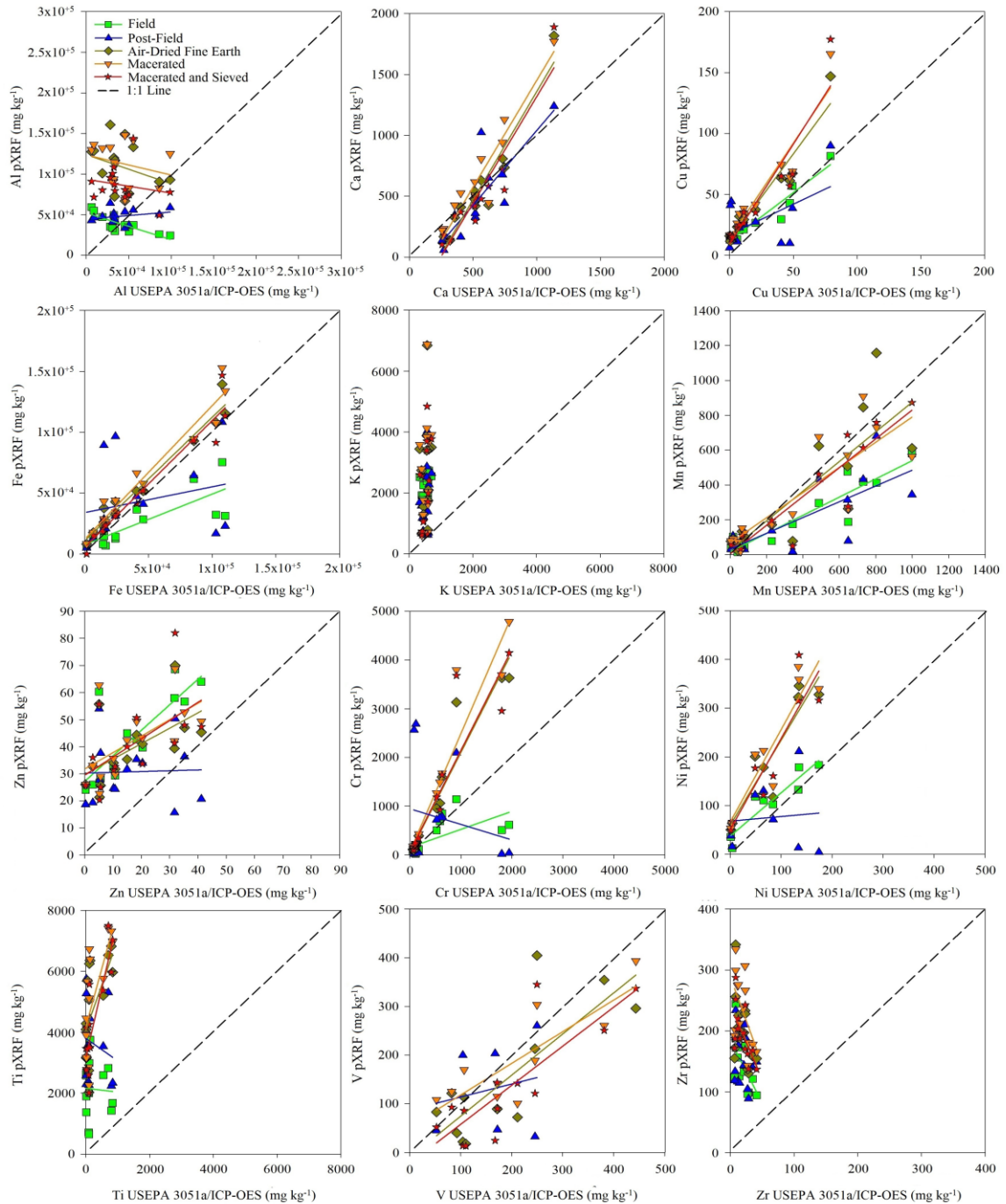
Fig. 5. Graphical representation of the relations between the elemental content obtained by the reference (USEPA 3051a) and pXRF methods under different scanning conditions of soil samples: field (F); post-field (PF); air-dried fine earth (ADFE); macerated (M); macerated and sieved (MS).

The vast majority of soil-sample-preparation conditions provided equations with high R^2 values (between 0.61 and 0.97) in addition to strong and significant correlations (between 0.76 and 0.98) for Ca, Cu, Fe, Mn, Cr, Ni, and V (Table 2). Thus, the results indicate that the

use of pXRF to predict the content of these elements obtained by the USEPA 3051a method is adequate, except for the F and PF conditions for Ca (R^2 = 0.00 and 0.70; r = 0.00 and 0.87), Fe (R^2 = 0.65 and 0.07; r = 0.80 and 0.25), Cr (R^2 = 0.40 and 0.04; r = 0.63 and -0.15), and V (R^2 = 0.00 and 0.04; r = 0.00 and -0.06), when compared to laboratory scanning conditions (ADFE, M, and MS). This is also true for Cu (R^2 = 0.31; r = 0.55), Mn (R^2 = 0.56; r = 0.76), and Ni (R^2 = 0.01; r = -0.05), but only under the PF condition. Therefore, pXRF scanning under F and PF conditions are not recommended for Ca, Fe, Cr, and V, while for Cu, Mn, and Ni scanning under the PF condition is not recommended. Low or no correspondence between field results through pXRF scanning and standard laboratory methods has also been reported by Hu *et al.* (2014), Santana *et al.* (2018), Silva *et al.* (2018), and Horta *et al.* (2021) for the elements As, Pb, Cu, Zn, Ti, and Si.

For Al (except for the F condition), K, Zn (except for the F condition), Ti (except for the MS condition), and Zr, linear models with low R^2 values (between 0.08 and 0.38) and non-significant correlations were obtained between results of pXRF and the USEPA 3051a method, regardless of the scanning condition (Table 2). However, for Al and Zn, the F condition provided linear models with R^2 values of 0.65 and 0.64, as well as correlation values of -0.80 and 0.80, respectively, while for Ti, R^2 values of 0.71 and correlation values of 0.84 were obtained under the MS condition.

The low correspondence between pXRF data and the reference method for Al and K is because in most Brazilian soils these elements are commonly found in the crystalline structure of silicate minerals, such as muscovite [$K(Al_2)Si_3Al)O_{10}(OH)_2$] in the sand fraction and kaolinite $Al_2Si_2O_5(OH)_4$ and hydroxy-interlayered vermiculite [$nH2O,Al_2(Si_4Al)O_{10}(OH)_2$] in the clay fraction (Brinatti *et al.* 2010; Inda *et al.* 2010; Resende *et al.* 2011; Kämpf *et al.* 2012), and they are not dissolved by the USEPA 3051a method, because it does not employ hydrofluoric acid. Likewise, Zr is commonly found in silicate minerals that are resistant to weathering, such as zirconite ($ZrSiO_4$) present in the sand fraction (Curi and Franzmeier 1987; Santos *et al.* 2016; Stockmann *et al.* 2016a). Also, Ti is widely found in the crystalline structure of non-silicate primary minerals (titanium oxides), such as ilmenite ($FeTiO_2$) and rutile (TiO_2), which are generally found in the clay fraction, and they are quite resistant to weathering and tend to accumulate in most tropical soils (Kämpf *et al.* 2012; Costa *et al.* 2014; Santos *et al.* 2016; Lyu *et al.* 2017). Therefore, these elements are greatly underestimated by the USEPA 3051a acid digestion method (USEPA 2007; Silva *et al.* 2014, 2019; Tavares *et al.* 2020). Conversely, pXRF determines the total elemental content of soils (Weindorf *et al.* 2014b; Ribeiro *et al.* 2017; Silva *et al.* 2021), which explains the high

contents for Al, K, Zr and Ti (Table 2) and, consequently, their low or no correspondence with the reference method (Table 3). The low performance of the USEPA 3051a method in the dissolution of Ti oxides minerals (ilmenite and rutile) was also observed by Silva *et al.* (2019), who found Ti values 20 times lower and a low correspondence between methods (R^2 = 0.31) for the reference in comparison with the pXRF method.

3.4 Validation of linear equations

Validation parameters for linear regression models with $R^2 \geq 0.60$ and significant ($p \leq$ 0.05) correlation values are shown in Table 3. As expected, due to the low correspondence with the reference method (Table 2), the validation for Al and Zn in the field (F), and for V in the air-dried fine-earth (ADFE), macerated (M), and macerated and sieved (MS) conditions, resulted in low values of R^2 (from 0.40 to 0.56) and RPD (from 1.23 to 1.48), as well as high values of RMSE (from 9.26 to 13757.48 mg kg^{-1}). Thus, models generated for these elements were classified as having poor prediction performance (Chang *et al.* 2001; Viscarra Rossel *et al.* 2010). For Ti, the model generated for the MS condition was considered good with a reasonable prediction performance, providing R^2 of 0.61 and RPD of 1.65, as well as RMSE of 192.57 mg kg^{-1}. In general, estimating Al, Zn, V, and Ti with pXRF is only recommended when moderate accuracy of the models is sufficient.

In contrast, Ca validation in the M condition produced superior results in comparison with other conditions, with R^2 values of 0.88, RPD of 3.00, and the lowest values of RMSE (80.28 mg kg^{-1}). For Cu, the R^2 values were above 0.70 in the F, ADFE, M, and MS conditions, but the best result was observed for the F condition (R^2 = 0.91, RMSE = 7.44 mg kg^{-1} and RPD = 3.29). Scanning with pXRF for the ADFE condition produced the best validation results for Fe, reaching R^2 of 0.94 and RPD of 4.15, as well as the lowest values of RMSE (9679.38 mg kg^{-1}) when compared with other scanning conditions.

For Mn, the greatest values of R^2 and RPD were obtained for the F condition (R^2 = 0.85 and RPD = 2.65), which made it possible to generate the lowest RMSE values (129.7 mg kg^{-1}). The ADFE condition produced the best validation parameters for Cr with values of R^2, RMSE and RPD of 0.86, 229.55 mg kg^{-1}, and 2.77, respectively. The ADFE and M conditions provided the best validation results for Ni, with values of R^2, RMSE, and RPD reaching 0.74, 30.09 mg kg^{-1}, and 2.10 for the ADFE condition, while for the M condition the values were 0.74, 30.37 mg kg^{-1}, and 2.08, respectively. Therefore, the linear regression models obtained were classified as very good for Ca, Cu, Fe, Mn, and Cr, and as good for Ni under the ADFE

and M conditions (Viscarra Rossel *et al.* 2010). All the models for these elements had excellent prediction performance (Chang *et al.* 2001).

Table 3. Results of equations to validate predictions concerning the contents of elements determined by the reference method (USEPA 3051a), which were compared to the data obtained via pXRF under different conditions of soil-sample scanning.

Conditions	Al			Ca		
	R²	RMSE[6]	RPD[7]	R²	RMSE	RPD
F[1]	**0.52**	**17660.41**	**1.48**	--	--	--
PF[2]	--	--	--	0.45	174.52	1.39
ADFE[3]	--	--	--	0.84	114.55	2.06
M[4]	--	--	--	**0.88**	**80.28**	**3.00**
MS[5]	--	--	--	0.75	183.41	1.31

Conditions	Cu			Fe		
	R²	RMSE	RPD	R²	RMSE	RPD
F	**0.91**	**7.44**	**3.29**	0.57	26006.91	1.54
PF	--	--	--	--	--	--
ADFE	0.84	12.31	1.99	**0.94**	**9679.38**	**4.15**
M	0.80	13.89	1.76	0.92	11280.77	3.56
MS	0.71	20.87	1.17	0.88	14068.12	2.85

Conditions	Mn			Zn		
	R²	RMSE	RPD	R²	RMSE	RPD
F	**0.85**	**129.77**	**2.65**	**0.50**	**9.26**	**1.45**
PF	--	--	--	--	--	--
ADFE	0.56	228.73	1.51	--	--	--
M	0.62	208.20	1.65	--	--	--
MS	0.83	138.51	2.49	--	--	--

Conditions	Cr			Ni		
	R²	RMSE	RPD	R²	RMSE	RPD
F	--	--	--	0.74	83.71	0.75
PF	--	--	--	--	--	--
ADFE	**0.86**	**229.55**	**2.77**	**0.74**	**30.09**	**2.10**
M	0.84	243.14	2.61	**0.74**	**30.37**	**2.08**
MS	0.72	326.41	1.95	0.65	37.13	1.70

Conditions	Ti			V		
	R²	RMSE	RPD	R²	RMSE	RPD
F	--	--	--	--	--	--
PF	--	--	--	--	--	--
ADFE	--	--	--	0.40	93.68	1.23
M	--	--	--	**0.56**	**80.82**	**1.41**
MS	**0.61**	**192.57**	**1.65**	0.46	84.77	1.35

[1] F – Field; [2] PF – Post-Field; [3] ADFE – Air-dried fine earth; [4] M - Macerated; [5] MS – Macerated and sieved; [6] RMSE – Root Mean Square Error; [7] RPD - Residual Prediction Deviation. "--" - The validation parameters are related to linear regression models with R² < 0.60 and with non-significant correlation values ($p > 0.05$). K and Zr do not appear in the table because the coefficients of determination (R²) of the linear regression models for all different scanning conditions with pXRF remained below 0.60.

The results demonstrate strong potential of applying data generated by pXRF to estimate results obtained by the USEPA 3051a method in tropical soils for Ca (M condition), Cu (F condition), Fe (ADFE condition), Mn (F condition), Cr (ADFE condition), and Ni (ADFE and M conditions) using regression equations. This may simplify sample preparation, reduce cost and time, besides being an environmentally friendly, chemical-free technique. We also emphasize that for Ca, Cu, Fe, Mn, Cr, and Ni, the ADFE condition (commonly used for scans with pXRF) provided good results. Silva *et al.* (2019) found lower R^2 and greater RMSE values for Ca, Cu, Fe, Mn, Cr, and Ni via pXRF scanning soil samples when compared with the best results found in the present study for these same elements, when both studies considered the USEPA 3051a method as reference (Table 3).

4. Conclusion

Variations in particle size distribution, degree of soil moisture, alteration or preservation of soil structure, and evaluation of locality (field vs. laboratory) provided contrasting results for the total elemental content obtained from pXRF for Al, Ca, Cu, Fe, K, Mn, Zn, Cr, Ni, Ti, V, and Zr. For the vast majority of elements evaluated, the M condition, followed by ADFE, provided the largest number of samples reaching quantification and the greatest elemental concentrations in relation to the other scanning conditions with pXRF, when the reference method was USEPA 3051a / ICP-OES. In contrast, the results obtained by pXRF under F and PF conditions were highly influenced by soil moisture.

Contents of Ca (M condition), Cu (F condition), Fe (ADFE condition), Mn (F condition), Cr (ADFE condition), and Ni (ADFE and M conditions) obtained by pXRF can be used to predict results obtained by the USEPA 3051a method. However, moderate relationships were observed for Al (F condition), Zn (F condition), Ti (MS condition), and V (M condition). Moreover, it was not possible to reasonably predict K and Zr contents, regardless the pXRF scanning conditions. These results show that pXRF is a promising tool for determining the elemental composition of tropical soils, and it can be considered a replacement alternative for the costly and time-consuming USEPA 3051a method, especially for Ca, Cu, Fe, Mn, Cr, and Ni. It is advisable to remove moisture from the soil sample to

improve the accuracy of the method. The ADFE condition is recommended, because of the ease of sample preparation in relation to the M and MS conditions. Similar results were provided by the ADFE, M, and MS conditions, and, in general for most of the elements evaluated, they give more reliable results than the F and PF conditions evaluated.

5. Acknowledgements

References

Acree A, Weindorf DC, Paulette L, Van Gestel N, Chakraborty S, Man T, Jordan C, Prieto JL (2020) Soil classification in Romanian catenas via advanced proximal sensors. *Geoderma* **377**, 114587. https://doi.org/10.1016/j.geoderma.2020.114587

Adhikari K, Hartemink AE (2016) Linking soils to ecosystem services — a global review. *Geoderma* **262**, 101–111. https://doi.org/10.1016/j.geoderma.2015.08.009

Alvares CA, Stape JL, Sentelhas PC, Moraes GJL, Sparovek G (2013) Köppen's climate classification map for Brazil. *Meteorologische Zeitschrift* **22**, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Barbosa JZ, Poggere G, Silva SHG, Mancini M, Motta ACV, Marques JJGSM, Curi N (2021) National-scale spatial variations of soil magnetic susceptibility in Brazil. *Journal of South American Earth Sciences* **108**, 103191. https://doi.org/10.1016/j.jsames.2021.103191

Benedet L, Acuña-Guzman SF, Faria WM, Silva SHG, Mancini M, Teixeira AFS, Pierangeli LMP, Acerbi Júnior FW, Gomide LR, Pádua Júnior AL, Souza IA, Menezes MD, Marques JJ, Guilherme LRG, Curi N (2021) Rapid soil fertility prediction using X-ray fluorescence data and machine learning algorithms. *Catena* **197**. https://doi.org/10.1016/j.catena.2020.105003

Borges CS, Weindorf DC, Nascimento DC, Curi N, Guilherme LRG, Carvalho GS, Ribeiro BT (2020) Comparison of portable X-ray fluorescence spectrometry and laboratory-

based methods to assess the soil elemental composition: Applications for wetland soils. *Environmental Technology & Innovation* **19**, 100826. https://doi.org/10.1016/j.eti.2020.100826

Brinatti AM, Mascarenhas YP, Pereira VP, Partiti CSM, Macedo Á (2010) Mineralogical characterization of a highly-weathered soil by the rietveld method. *Scientia Agricola* **67**, 454–464. https://doi.org/10.1590/s0103-90162010000400013

Chang C, Laird DA, Mausbach MJ, Hurburgh CRJ (2001) Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. *Soil Science Society of America Journal* **65**, 480–490. https://doi.org/10.2136/sssaj2001.652480x

Chen M, Ma LQ (1998) Comparison of Four USEPA Digestion Methods for Trace Metal Analysis Using Certified and Florida Soils. *Journal of Environmental Quality* **27**, 1294–1300. https://doi.org/10.2134/jeq1998.00472425002700060004x

Costa SÂD, Ker JC, Simões DFF, Fontes MPF, Fabris JD, Andrade FV (2014) Pedogênese e classificação de latossolos desenvolvidos de itabiritos no Quadrilátero Ferrífero, MG. *Revista Brasileira de Ciência do Solo* **38**, 359–371. https://doi.org/10.1590/S0100-06832014000200001

Curi N, Franzmeier DP (1987) Effect of parent rocks on chemical and mineralogical properties of some Oxisols in Brazil. *Soil Science Society of America Journal* **51**, 153–158.

Curi S, Silva SHG, Poggere GC, Menezes MD (2017) Mapeamento de Solos e Magnetismo no Campus da UFLA Como Traçadores Ambientais, First. ed. 147p.

Dantas AAA, Carvalho LG, Ferreira E (2007) Classificação e tendências climáticas em Lavras, MG. *Ciência e Agrotecnologia* **31**, 1862–1866. https://doi.org/10.1590/S1413-70542007000600039

Dijair TSB, Silva FM, Teixeira AFS, Silva SHG, Guilherme LRG, Curi N (2020) Correcting field determination of elemental contents in soils via portable X-ray fluorescence spectrometry. *Ciência e Agrotecnologia* **44**. https://doi.org/10.1590/1413-7054202044002420

Evanylo G, McGuinn R (2009) Agricultural management practices and soil quality: measuring, assessing, and comparing laboratory and field test kit indicators of soil quality attributes. Virginia: Polytechnic Institute and State University. Publication 452-400. p.12.

Faria AJG, Silva SHG, Melo LCA, Andrade R, Mancini M, Mesquita LF, Teixeira AFDS, Guilherme LRG, Curi N (2020) Soils of the Brazilian Coastal Plains biome: Prediction

of chemical attributes via portable X-ray fluorescence (pXRF) spectrometry and robust prediction models. *Soil Research* **58**, 683–695. https://doi.org/10.1071/SR20136

Feng X, Zhang H, Yu P (2020) X-ray fluorescence application in food, feed, and agricultural science: a critical review. *Critical Reviews in Food Science and Nutrition* **61**, 2340-2350. https://doi.org/10.1080/10408398.2020.1776677

Horta A, Azevedo L, Neves J, Soares A, Pozza L (2021) Integrating portable X-ray fluorescence (pXRF) measurement uncertainty for accurate soil contamination mapping. *Geoderma* **382**, 114712. https://doi.org/10.1016/j.geoderma.2020.114712

Horta A, Malone B, Stockmann U, Minasny B, Bishop TFA, McBratney AB, Pallasser R, Pozza L (2015) Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. *Geoderma* **241–242**, 180–209. https://doi.org/10.1016/j.geoderma.2014.11.024

Hu W, Huang B, Weindorf DC, Chen Y (2014). Metals analysis of agricultural soils via portable x-ray fluorescence spectrometry. *Bulletin of Environmental Contamination and Toxicology* **92**, 420–426. https://doi.org/10.1007/s00128-014-1236-3

Inda AV, Torrent J, Barron V, Bayer C (2010) Aluminum hydroxy-interlayered minerals and chemical properties of a subtropical Brazilian Oxisol under no-tillage and conventional tillage. *Revista Brasileira de Ciência do Solo* **34**, 33–41. https://doi.org/10.1590/S0100-06832010000100004

Kabata-Pendias A (2010) Trace elements in soils and plants. CRC Press, Fourth Ed. ed. p.548. https://doi.org/10.1201/b10158

Kämpf N, Marques JJ, Curi N (2012) Mineralogia de Solos Brasileiros, in: Ker JC, Curi N, Schaefer CEGR, Vidal-Torrado P. Pedologia Fundamentos. SBCS, Viçosa, MG. p.343.

Kilbride C, Poole J, Hutchings T (2006). A comparison of Cu, Pb, As, Cd, Zn, Fe, Ni and Mn determined by acid extraction/ICP–OES and ex situ field portable X-ray fluorescence analyses. *Environmental Pollution* **143**, 16–23. https://doi.org/10.1016/j.envpol.2005.11.013

Kuhn M, Wing J, Weston S, Williams A, Keefer C, Engelhardt A, Al E (2018) Classification and regression training. Available online at https://cran.r-project.org/web/packages/caret/caret.pdf (verified on Jan 23, 2021). URL https://cran.r-project.org/web/packages/caret/caret.pdf (accessed 7.18.18).

Laiho JVP, Peramaki P (2005) Evaluation of portable X-ray fluorescence (pXRF) sample preparation methods. *Special Paper - Geological Survey of Finland* **38**, 73–82.

Lyu S, Wei X, Chen J, Wang C, Wang X, Pan D (2017) Titanium as a beneficial element for crop production. *Frontiers in Plant Science* **8**, 1–19. https://doi.org/10.3389/fpls.2017.00597

Mancini M, Weindorf DC, Chakraborty S, Silva SHG, Santos TAF, Guilherme LRG, Curi N. (2019) Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. *Geoderma* **337**, 718–728. https://doi.org/10.1016/j.geoderma.2018.10.026

Markowicz AA (2008) Quantification and correction procedures., in: Potts PJ, West M. (Eds.), Portable X-Ray Fluorescence Spectrometry. Capabilities for In Situ Analysis. The Royal Society of Chemistry, Cambridge, U.K. pp. 13–38.

McLaren TI, Guppy CN, Tighe MK, Forster N, Grave P, Lisle LM, Bennett JW (2012) Rapid, Nondestructive Total Elemental Analysis of Vertisol Soils using Portable X-ray Fluorescence. *Soil Science Society of America Journal* **76**, 1436–1445. https://doi.org/10.2136/sssaj2011.0354

Padilla JT, Hormes J, Magdi SH (2019) Use of portable XRF: Effect of thickness and antecedent moisture of soils on measured concentration of trace elements. *Geoderma* **337**, 143–149. https://doi.org/10.1016/j.geoderma.2018.09.022

Parsons C, Margui GE, Pili E, Floor GH, Roman-Ross G, Charlet L (2013) Quantification of trace arsenic in soils by field-portable X-ray fluorescence spectrometry: Considerations for sample preparation and measurement conditions. *Journal of Hazardous Materials* **262**, 1213–1222. https://doi.org/10.1016/j.jhazmat.2012.07.001

Peinado FM, Ruano SM, González MGB, Molina CE (2010) A rapid field procedure for screening trace elements in polluted soil using portable X-ray fluorescence (pXRF). *Geoderma* **159**, 76–82. https://doi.org/10.1016/j.geoderma.2010.06.019

Qu M, Chen J, Li W, Zhang C, Wan M, Huang B, Zhao Y (2019) Correction of in-situ portable X-ray fluorescence (PXRF) data of soil heavy metal for enhancing spatial prediction. *Environmental Pollution* **254**, 112993. https://doi.org/10.1016/j.envpol.2019.112993

R Development Core Team (2019) R: A language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing, 2019. Available in: https://www.r-project.org/. Access in: Jul. 17, 2020.

Radu T, Diamond D (2009) Comparison of soil pollution concentrations determined using AAS and portable XRF techniques. *Journal of Hazardous Materials* **171**, 1168–1171. https://doi.org/10.1016/j.jhazmat.2009.06.062

Resende M, Curi N, Ker JC, Rezende SB (2011) Mineralogy of Brazilian soils: interpretation and applications, second edi. ed. Editora UFLA, Lavras.

Resende M, Curi N, Poggere GC, Barbosa JZ, Pozza AAA, Teixeira AFS (2021) Pedologia, fertilidade, água e planta: Inter-relações e aplicações, 2nd ed. Editora UFLA, Lavras. p.263.

Ribeiro BT, Silva SHG, Silva EA, Guilherme LRG (2017) Portable X-ray fluorescence (pXRF) applications in tropical Soil Science. *Ciência e Agrotecnologia* **41**, 245–254. https://doi.org/10.1590/1413-70542017413000117

Ribeiro BT, Weindorf DC, Silva BM, Tassinari D, Amarante LC, Curi N, Guilherme LRG (2018) The influence of soil moisture on oxide determination in tropical soils via portable x-ray fluorescence. *Soil Science Society of America Journal* **82**, 632. https://doi.org/10.2136/sssaj2017.11.0380

Rouillon M, Taylor MP (2016) Can field portable X-ray fluorescence (pXRF) produce high quality data for application in environmental contamination research? *Environmental Pollution* **214**, 255–264. https://doi.org/10.1016/j.envpol.2016.03.055

Santana MLT, Carvalho GS, Guilherme LRG, Curi N, Ribeiro BT (2019) Elemental concentration via portable x-ray fluorescence spectrometry: Assessing the impact of water content. *Ciência e Agrotecnologia* **43**. https://doi.org/10.1590/1413-7054201943029218

Santana MLT, Ribeiro BT, Silva SHG, Poggere GC, Guilherme LRG, Curi N (2018) Conditions affecting oxide quantification in unknown tropical soils via handheld X-ray fluorescence spectrometer. *Soil Research* **56**, 648. https://doi.org/10.1071/sr18099

Santos AC, Pereira MG, Anjos LHC, Bernini TA, Cooper M (2016) Genesis of soils formed from mafic igneous rock in the Atlantic Forest environment. *Revista Brasileira de Ciência do Solo* **40**, 1–21. https://doi.org/10.1590/18069657rbcs20150056

Santos RD, Lemos RC, Santos HG, Ker JC, Anjos LHC, Shimizu SH (2015) Manual de descrição e coleta de solo no campo. 7. edição revisada e ampliada. Viçosa, MG: Sociedade Brasileira de Ciência do Solo, 170 p.

Santos H, Jacomine PK, Anjos LH, Oliveira V, Lumbreras J, Coelho M, Almeida J, Araujo Filho J, Oliveira J, Cunha TJF (2018). Sistema Brasileiro de Classificação de Solos, 5th ed. ed. Embrapa, Brasília, p.355.

Shepherd KD, Walsh MG (2002) Development of Reflectance Spectral Libraries for Characterization of Soil Properties. *Soil Science Society of America Journal* **66**, 988. https://doi.org/10.2136/sssaj2002.9880

Silva EA, Weindorf DC, Silva SHG, Ribeiro BT, Poggere GC, Carvalho TS, Gonçalves MGM, Guilherme LRG, Curi N (2019) Advances in Tropical Soil Characterization via Portable X-Ray Fluorescence Spectrometry. *Pedosphere* **29**, 468–482. https://doi.org/10.1016/s1002-0160(19)60815-5

Silva SHG, Ribeiro BT, Guerra MBB, Carvalho HWP, Lopes G, Carvalho GS, Guilherme LRG, Resende M, Mancini M, Curi N, Rafael RBA, Cardelli V, Cocco S, Corti G, Chakraborty S, Li B, Weindorf DC (2021) pXRF in tropical soils: Methodology, applications, achievements and challenges. Advances Agronomy **167**, 1–62. https://doi.org/10.1016/bs.agron.2020.12.001

Silva SHG, Silva EA, Poggere GC, Guilherme LRG, Curi N (2018) Tropical soils characterization at low cost and time using portable X-ray fluorescence spectrometer (pXRF): Effects of different sample preparation methods. *Ciência e Agrotecnologia* **42**, 80–92. https://doi.org/10.1590/1413-70542018421009117

Silva SHG, Silva EA, Poggere GC, Junior ALP, Gonçalves MGM, Guilherme LRG, Curi N (2020) Modeling and prediction of sulfuric acid digestion analyses data from pxrf spectrometry. *Scientia Agricola* **77**. https://doi.org/10.1590/1678-992x-2018-0132

Silva YJAB, Nascimento CWA, Biondi CM (2014) Comparison of USEPA digestion methods to heavy metals in soil samples. *Environmental Monitoring and Assessment* **186**, 47–53. https://doi.org/10.1007/s10661-013-3354-5

Soil Survey Staff (2014a) Soil survey field and laboratory methods manual. Version 2. USDA-NRCS. Available online at: https://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/stelprdb1244466.pdf (verified 15 Feb. 2021).

Soil Survey Staff (2014b) Keys to soil taxonomy, 12th ed. USDA-NRCS. Retrieved from http://www.nrcs. usda.gov/Inter net/FSE_DOCUM ENTS/nrcs1 42p2_05154 6.pdf.

Soltanpour PN, Johnson GW, Workman SM, Jones BJ, Miller RO (1996) Inductively coupled plasma emission spectrometry and inductively coupled plasma-mass spectrometry, in: Methods of Soil Analysis. Part 3. Chemical Methods. pp. 91–139. https://doi.org/10.2136/sssabookser5.3.c5

Stockmann U, Cattle SR, Minasny B, McBratney AB (2016a) Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. *Catena* 139, 220–231. https://doi.org/10.1016/j.catena.2016.01.007

Stockmann U, Jang HJ, Minasny B, McBratney AB (2016b) The Effect of Soil Moisture and Texture on Fe Concentration Using Portable X-Ray Fluorescence Spectrometers, in: Hartemink A, Minasny B. *Digital Soil Morphometrics*. pp. 63–71. https://doi.org/10.1007/978-3-319-28295-4_5

Tavares TR, Molin JP, Nunes LC, Alves EEN, Melquiades FL, Carvalho HWP, Mouazen AM (2020) Effect of x-ray tube configuration on measurement of key soil fertility attributes with XRF. *Remote Sensing* **12**, 963. https://doi.org/10.3390/rs12060963

Teixeira AFS, Pelegrino MHP, Faria WM, Silva SHG, Gonçalves MGM, Acerbi Júnior FW, Gomide LR, Pádua Júnior AL, Souza IA, Chakrabortyd S, Weindorfe DC, Guilherme LRG, Curi N (2020) Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry. *Geoderma* **361**, 114132. https://doi.org/10.1016/j.geoderma.2019.114132

Towett EK, Shepherd KD, Tondoh JE, Winowiecki LA, Lulseged T, Nyambura M, Sila A, Vågen TG, Cadisch G (2015) Total elemental composition of soils in Sub-Saharan Africa and relationship with soil forming factors. *Geoderma Regional* **5**, 157–168. https://doi.org/10.1016/j.geodrs.2015.06.002

Tsadilas CD, Rinklebe J, Selim HM (2019) Nickel in Soils and Plants. CRC Press, 1st Editio. ed. New York. p.432. https://doi.org/10.1201/9781315154664-3

USEPA (2007a) Method 3051a (SW-846): Microwave assisted acid digestion of sediments, sludges, soils, and oils. US EPA. https://www.epa.gov/sites/ production/files/2015-12/documents/3051a.pdf. (Acessed 27 jan. 2021).

USEPA (2007b) Method 6200: Field portable X-ray fluorescence spectrometry for the determination of elemental concentrations in soil and sediment. US EPA. https://www.epa.gov/sites/production/files/2015-12/documents/6200.pdf. (Accessed 10 Dez. 2019).

Viscarra Rossel RA, McBratney AB, Minasny B (2010) Proximal soil sensing (Progress in Soil Science), Springer; 2010ª edição. p.671.

Weindorf DC, Bakr N, Zhu Y, Mcwhirt A, Ping CL, Michaelson G, Nelson C, Shook K, Nuss S (2014) Influence of Ice on Soil Elemental Characterization via Portable X-Ray Fluorescence Spectrometry. *Pedosphere* **24**, 1–12. https://doi.org/10.1016/S1002-0160(13)60076-4

Weindorf DC, Bakr N, Zhu Y (2014) Advances in portable X-ray fluorescence (pXRF) for environmental, pedological, and agronomic applications, *Advances in Agronomy* **128**, 1-45. https://doi.org/10.1016/B978-0-12-802139-2.00001-9

Weindorf DC, Chakraborty S (2020) Portable X-ray fluorescence spectrometry analysis of soils. *Soil Science Society of America Journal* **84**, 1384–1392. https://doi.org/10.1002/saj2.20151

Zhang Y, Hartemink AE (2019) Data fusion of vis–NIR and pXRF spectra to predict soil physical and chemical properties. *European Journal of Soil Science* **71**, 316–333. https://doi.org/10.1111/ejss.12875

ARTICLE 4 - Rapid elemental prediction of heterogeneous tropical soils from pXRF data: a comparison of models via linear regressions and machine learning algorithms

Manuscript to be submitted to the journal of Geochemical Exploration

Alvaro José Gomes de Faria[1], Sérgio Henrique Godinho Silva[1], Luiza Carvalho Alvarenga Lima[2], Renata Andrade[1], Lívia Botelho[1], Leônidas Carrijo Azevedo Melo[1], Nilton Curi[1]

[1]Department of Soil Science, Federal University of Lavras – UFLA, Doutor Sylvio Menicucci Avenue, Lavras, Minas Gerais State, 37200-900, Brazil.
[2]Department of Mechanical Engineering, Federal University of Lavras – UFLA, Doutor Sylvio Menicucci Avenue, Lavras, Minas Gerais State, 37200-900, Brazil.

Abstract

In Brazil, USEPA 3051a is considered a standard analytical methodology for the extraction of inorganic substances in soils commonly used for environmental monitoring, mainly geochemical balance. However, these analyzes are expensive, time-consuming and produce chemical residues. Conversely, proximal sensors such as portable X-ray fluorescence (pXRF) spectrometry reduce analysis time, costs and consequently offer a valuable alternative to laboratory analyses. This study aimed to investigate the development and validation of models to predict the results of the USEPA 3051a method for 28 chemical elements from pXRF data, through linear regressions and robust machine learning algorithms. A total of 179 samples (considering horizons A and B) representing a large area from Brazil were analyzed for elemental composition using the USEPA 3051a method and pXRF. Four prediction models - simple linear regression (SLR), stepwise multiple linear regressions (SMLR), support vector machine (SVM) and random forest (RF) - were tested and compared. Modeling was developed with 70% of the data, while the remaining 30% were used for validation by calculating R^2, residual prediction deviation (RPD), root mean square error (RMSE) and normalized root mean square error (NRMSE). The results demonstrated that most of the elements obtained via pXRF were well correlated with the USEPA 3051a method. Machine learning algorithms (SVM and RF) performed better than SLR and SMLR for the prediction of Al, Ba, Bi, Ca, Cd, Ce, Co, Cr, Cu, Fe, Mg, Mn, Mo, P, Pb, Sn, Sr, Ti, Tl, V, Zn and Zr in tropical soils, whose R^2 and RPD values ranged from 0.52 to 0.94 and 1.43 to 3.62, respectively, as well as the lowest values of RMSE and NRMSE values (0.28 to 0.70 mg kg⁻

[1]). Unsatisfactory results ($R^2 < 0.50$; RPD < 1.4) were generated for K, Ni, Be, Li, Na and S, regardless of the prediction models. Most USEPA 3051a results can be accurately predicted from pXRF data through SVM and RF algorithms, saving cost, time, and ensuring routine geochemical characterization of tropical soils on a large scale in an environmentally friendly way.

Keywords: portable X-ray fluorescence spectrometry, rapid soil analysis, total elemental content, support vector machine, random forest.

1. Introduction

Portable X-ray fluorescence (pXRF) spectrometry is one of the proximal sensing techniques that has received great attention in soil science in recent years (O'Rourke et al., 2016a; Ravansari et al., 2020). This growing use of pXRF occurs due to technological advances, which include equipment miniaturization, advances in computing power, battery life capability, and high-energy X-ray generating tubes (Declercq et al., 2019; Javadi and Mouazen, 2021; Tavares et al., 2021). However, in tropical developing countries, studies with this technique are still scarce and recent considering the wide range of purposes in which this tool can be used for (e.g., agronomic, pedological, environmental and geochemical) (O'Rourke et al., 2016b; Ravansari et al., 2020; Silva et al., 2021).

PXRF has changed the acquisition of multielemental soil data (from Mg to U on the Periodic Table) with a wide concentration range (ppm or mg kg^{-1}), non-destructive, fast (approximately 60 s), low-cost and environmentally correct both in the field and in the laboratory (Hseu et al., 2016; Vasques et al., 2020; Weindorf et al., 2014). In temperate soils, several methods have been proposed for soil and sediment characterization (Soil Survey Staff, 2014; USEPA, 2007a; Weindorf and Chakraborty, 2020), and a methodology for tropical soils via pXRF has also recently been made available (Silva et al., 2021). Studies have shown that, in addition to soil, different matrices can also be characterized using this tool, including parent material (Mancini et al., 2019b, 2019a), different types of sediments (Ferreira et al., 2021; Knight et al., 2021), leaf tissues (Borges et al., 2020a), water (Pearson et al., 2016), and organic compounds (e.g., manure and biochar) (Faria et al., 2021; Weindorf et al., 2018).

Commonly, the quantification of the total or semi-total elemental concentration in soils is performed by traditional wet methods based on concentrated acid digestion (e.g., USEPA 3050b, 3051a and 3052), which convert solid samples into liquid extracts with

quantification in inductively coupled plasma optical emission spectrometry (ICP-OES) (Chen and Ma, 1998; Silva et al., 2014; USEPA, 2007b). According to the Brazilian Environmental Legislation, the USEPA 3051a method is a standard semi-total wet analytical procedure for the extraction of inorganic substances in soils (CETESB, 2014; CONAMA, 2009; COPAM, 2011). This method is widely used in highly weathered and leached tropical soils because it has recovery values comparable to more aggressive methods (e.g., USEPA 3052), without requiring hydrofluoric acid manipulation (Bispo et al., 2021; Bocardi et al., 2020; Faria et al., 2022; Ribeiro et al., 2019). However, unlike pXRF, these traditional wet methods based on acid digestion, including USEPA 3051a or more aggressive ones (e.g., USEPA 3052) can be quite expensive, time consuming and hazardous as they may involve the use of large amounts of strong acids and the generation of chemical residues which can have impacts on the environment and human health (Benedet et al., 2021; Nascimento and Tenuta Filho, 2010; Wan et al., 2020).

To overcome the negative environmental impact of wet geochemistry, efforts are being made to establish an accurate elementary relationship between pXRF data and conventional laboratory analysis results, mainly in tropical developing countries, whose studies in this regard are still very limited. Silva et al. (2019) and Faria et al. (2022) used pXRF to collect elementary data from only 22 and 14 soil samples, respectively, in order to estimate the results obtained by the 3051a method by applying simple linear regressions (SLR). These authors concluded that the prediction results were satisfactory only for the elements Ca, Cu, Fe, Mn, Cr and Ni with R^2 values ranging from 0.56 (Mn) to 0.94 (Fe). Borges et al. (2020b) used 144 soil samples combined with simple linear regressions and also observed a strong elementary relationship between the pXRF results and the 3051a method for some elements, such as: Cu ($R^2 = 0.83$), Fe ($R^2 = 0.83$) and Pb ($R^2 = 0.82$). However, these works involve more homogeneous soils, which instigates the investigation of this new approach involving a more diverse set of soils. In addition, previous studies have shown that machine learning algorithms, such as: random forest (RF) and support vector machine (SVM) with linear kernel, have provided highly accurate results (Andrade et al., 2021; Faria et al., 2020; Liu et al., 2021; Meier et al., 2018; Silva et al., 2020; Wu et al., 2018), especially in large heterogeneous areas with great variability of soils, parent materials and land use (Araújo et al., 2014; Lucà et al., 2017; Wan et al., 2019).

The authors of the present study are unaware of any other work carried out in tropical soils using linear and non-linear models combined with a heterogeneous database for predicting the results of the USEPA 3051a method from data obtained by pXRF. Thus, the

objectives of this study were: i) to use pXRF to characterize the elemental content of highly heterogeneous tropical soils; and ii) develop and validate prediction models for 28 elements traditionally obtained by the USEPA 3051a method based on data generated by pXRF through machine learning algorithms (RF and SVM) and simple (SLR) and multiple (SMLR) linear regressions. We hypothesize that the results of soil chemical composition determined by pXRF will be well correlated with the USEPA 3051a method for tropical soils, despite their different quantification methodologies, which will allow the development of robust prediction models using only pXRF data as input variables associate with machine learning algorithms.

2. Material and Methods

2.1 *Soil sampling*

This study was carried out with 179 soil samples collected in the Brazilian states of Minas Gerais, Bahia, Espírito Santo, Rio de Janeiro, São Paulo and Santa Catarina, whose territorial extension reached is 1,587,169 km^2 (Fig. 1). According to the Köppen classification system, these regions have different tropical and subtropical climate conditions, as shown in Fig. 1 (Alvares et al., 2013).
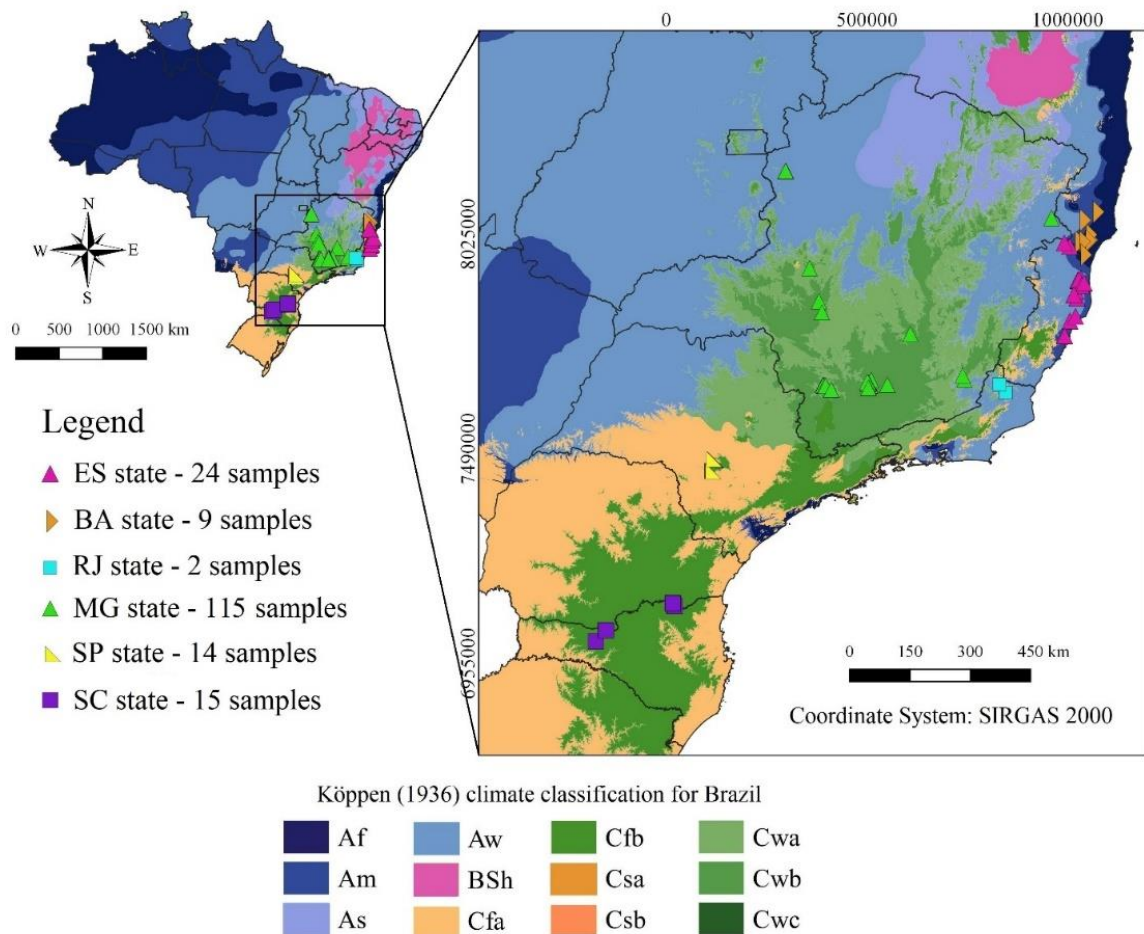
Fig. 1. Study area, distribution of the collected samples in six Brazilian states and Köppen (1936) climate classification for Brazil. A - tropical zone (Af - without dry season, Am - monsoon, Aw - with dry winter, As - with dry summer), B - dry zone [Bs - semiarid (BSh - low latitude and altitude)], C - Humid subtropical zone [Cf - oceanic climate, without dry season (Cfa - with hot summer, Cfb - with temperate summer), Cw - with dry winter (Cwa - and hot summer, Cwb - and temperate summer, Cwc - and short and cool summer), Cs - with dry summer (Csa - and hot, Csb - and temperate)].

The samples were collected in a wide range of soil classes described according to Santos et al. (2015) and classified according to the *Brazilian Soil Classification System* (Santos et al., 2018) as (n = number of samples): Argisols (n = 57), Latosols (n = 48), Cambisols (n = 32), Nitosols (n=16), Neosols (n=12), Espodosols (n=09), Luvisols (n=02), Plintosols (n=02) and Chernosols (n=01). The approximate correspondence by *US Soil Taxonomy* (Soil Survey Staff, 2014) is: Ultisols, Oxisols, Inceptisols, Oxisols and Ultisols, Entisols, Spodosols, Alfisols, Oxisols (Plinthaquoxes) and Mollisols, respectively. In total, 77 samples were collected on the A horizon (between 0 – 20 cm) and 102 samples on the B

horizon (between 60 and 120 cm), totaling the 179 soil samples used in this study. Notably, these mentioned soil classes correspond to 90% of the Brazilian territory (~ 7,665,400 km^2) (Santos et al., 2018).

The following parent materials were identified in the soil sample collection areas during fieldwork: granite, gneiss, gabbro, basalt, alluvial and colluvial sediments, shale, sandstone, quartzite, phyllite, itabirite, mica-schist, amphibolite, charnockite and tuff. Finally, the areas where the samples were collected were under different land uses, such as: native vegetation, planted or native pastures and cultivated with coffee (*Coffea arabica* and/or *canephora*) and eucalyptus (*Eucalyptus* spp). In cultivated areas, agronomic management included the application of limestone and agricultural gypsum prior to mineral fertilizers, according to the needs of each crop. No management practices were carried out in areas under native vegetation.

2.2 *Laboratory analyses by USEPA 3051a and pXRF*

All 179 soil samples were initially air-dried, passed through a 2 mm sieve (air-dried fine earth – ADFE) and individually analyzed by both wet (USEPA 3051a) and dry (pXRF) geochemical methods. To facilitate the USEPA 3051a analysis, a portion of each ADFE (~15 g) was ground using an agate mortar and passed through a 150 µm sieve. Subsequently, in triplicate, 0.5 g of the ground soil was transferred to 50-mL Teflon® vessels containing 5 mL of concentrated HNO$_3$. Digestion was performed by heating the samples for 10 min in a microwave (CEM MARS-5, CEM Corp., USA) at 180 °C under a pressure of 448 kPA (USEPA, 2007b). After completing the digestion process, the samples were passed through N° 40 Whatman filter paper, rinsing with ultrapure water until the volume of 50 mL was completed. The extracts obtained were analyzed by inductively coupled plasma optical emission spectrometry (ICP-OES), Spectro Blue model (Spectro Analytical Instruments, Germany) (Soltanpour et al., 1996). Thus, 28 elements, based on the following wavelengths (nm), were obtained: Al (196,152), Ba (233,527), Be (313,042), Bi (206,170), Ca (315,887), Cd (214,438), Ce (413,380), Co (230,786), Cr (267,716), Cu (327,396), Fe (238,204), K (766,491), Li (670,780), Mg (285,213), Mn (403,076), Mo (203,909), Na (588,995), Ni (231,604), P (178,287), Pb (220,353), S (182,034), Sr (460,733), Sn (189,991), Ti, (336,121), Tl (276,787), V (292,402), Zn (206,200) and Zr (339,198). A multi-elemental calibration curve between 0.1 to 50 mg L^{-1} was used to quantify these elements.

For comparison purposes, a new portion of ADFE (~15 g) was analyzed using a pXRF model S1 Titan LE (Bruker Nano Analytics, Kennewick, WA, USA), according to the methodologies recommended by Silva et al. (2021) for tropical soils. The samples were scanned in triplicate in Trace mode (dual soil), integrated into the GeoChem software, using two sequential operating beams. Each beam was set to scan for 30s, so the entire scan was completed in 60s. The pXRF features a Rh tube (4 W, 15–50 keV, and 5–100 µA) and a silicon shift detector (SDD) with typical resolution <145 eV, which allows the elemental identification and quantification between Mg and U in the Periodic Table. Despite the large amount of elements that pXRF can identify and quantify (~ 45 elements), only 15 elements that presented readings in more than 50% of the soil samples were considered for this study (e.g., Al, Ca, Cr, Cu, Fe, K, Mn, Ni, P, Pb, Sr, Ti, V, Zn and Zr – all in mg kg^{-1}). Thus, elements with no or just a few readings (< 50%) were disregarded.

The pXRF performance was assessed by scanning certified reference materials (CRM) by the National Institute of Standards and Technology (NIST) (2710a and 2711a) and a sample certified by the manufacturer (check sample - CS) of the equipment for the elements detected in most samples. The recovered values [% Recovery = (elementary content via pXRF/certified elementary content) x 100] for these elements compared to the information from CRM 2710a and 2711a and from the manufacturer's sample (CS), were respectively: Al (64/127/97), Ca (86/45/--), Cr (--/129/--), Cu (84/67/95), Fe (108/98/90), K (52/33/89), Mn (73/64/88), Ni (--/104/110), P (--/69/93), Pb (--/--/98), Sr (198/172/--), Ti, (169/118/--), V (--/20/--), Zn (99/81/--) and Zr (105/--/--). Dashed lines (--) indicate the absence of reference values for an element or it was below the pXRF detection limit.

2.3 *Statistical analysis and validation of models*

The 28 elements obtained by the USEPA 3051a wet method, as well as the 15 elements delivered by pXRF were initially analyzed using boxplots and histograms to assist in the characterization and identification of intrinsic differences in the elemental quantification obtained by both methods in highly heterogeneous tropical soils. All statistical analyses were performed using R software version 4.0.3 (R Development Core Team, 2018).

In order to evaluate the elementary relationship between the used methods (USEPA 3051a and pXRF) in a very simple way, simple linear regressions (SLR) were created between the elements obtained by both USEPA 3051a and pXRF methods (Al, Ca, Cr, Cu, Fe, K, Mn, Ni, P, Pb, Sr, Ti, V, Zn and Zr). Also, in order to generate more robust prediction

models, stepwise multiple linear regression (SMLR), random forest (RF) and support vector machine (SVM) with linear kernel were also created to predict the 28 elements obtained via USEPA 3051a based on the 15 elements detected and quantified by pXRF in all samples.

The prediction models were created using R software, with RF contained in the "randomForest" package (Liaw and Wiener, 2015) and SVM contained in the "e1071" package (Hornik et al., 2015). The RF analysis was performed with the following parameters: number of trees of the model (ntrees) = 1000, node size = 5, and number of variables used in each tree (mtry) = 4, equivalent to the number of variables divided by 3, as suggested by (Liaw and Wiener, 2002). The RF does not provide a final equation, but the variable importance of the generated model can be assessed. One of the metrics released by the algorithm is the percent increase in mean square error (% IncMSE). The higher the % IncMSE value, the greater the importance of the variable for the model (González et al., 2015).

Prior to the modeling, all pXRF results from scans of soil samples were normalized (normalized elemental content = (x - mean (x)) / std(x), where x is the original elementary content, mean (x) is the mean of the values for each element, and std(x) is the standard deviation of the values of each element) and scaled (scaled elemental content = (x – mean(x)), where x is the original elemental content).

For SLR, SMLR, RF and SVM prediction models, the total database (n = 179) was randomly separated into modeling (125 samples - 70%) and validation sub-datasets (54 samples - 30%). Samples from A and B horizons from the same soil profile were kept in the same dataset (either modeling or validation) in order to ensure true data independence during the validation procedure. The accuracy of the prediction models was assessed by comparing the predicted with the observed values through the following statistical indexes: determination coefficient (R^2, Eq. 1), root mean square error (RMSE, Eq. 2), normalized root mean square error (NRMSE, Eq. 3) and residual prediction deviation (RPD, Eq. 4).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (ei - mi)^2}{\sum_{i=1}^{n} (ei - \bar{Y}_i)^2} \qquad\qquad \text{Eq. 1}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (ei - mi)^2} \qquad\qquad \text{Eq. 2}$$

$$NRMSE = \frac{RMSE}{SD}$$
<div align="right">Eq. 3</div>

$$RPD = \frac{SD}{RMSE}$$
<div align="right">Eq. 4</div>

Where: n - number of observations; ei - values estimated by the model; mi - values obtained through USEPA 3051a; $\bar{Y}i$ - represents the mean of the values observed in the USEPA 3051a analyses; SD - is the standard deviation of the values observed in the USEPA 3051a analyses.

The R^2 values of the generated models and their validations were classified as: very good (> 0.81), good ($0.61 - 0.80$), regular ($0.41 - 0.60$) and low ($< 0. 4$), according to Viscarra Rossel et al. (2010). The normalized RMSE facilitates the comparison between datasets or models with different data ranges. Models with lower RMSE and NRMSE values indicate better performance and reliability of predictions (Chai and Draxler, 2014; Jalali et al., 2019; Taebi and Mansy, 2017). The generated RPD were classified into three classes: RPD > 2, models with accurate predictions, $1.4 \leq RPD \leq 2$, models with reasonable predictions and RPD < 1.4, models without prediction capability (Chang et al., 2001). The models with greater R^2 and RPD and smaller RMSE and NRMSE were considered the best ones for predicting USEPA 3051a laboratory analyses.

3. Results and Discussion

3.1 Descriptive analysis

The boxplots and histograms for the 28 elements determined by USEPA 3051a, as well as for the 15 elements delivered by pXRF, can be seen in Fig. 2. High amplitude and variability were observed in the data determined by the USEPA 3051a method. Of the 28 elements determined by the USEPA 3051a method, the averages values followed in the order of Fe > Al > Ca > K > Mn > P > Ti, and ranged from 141 (Ti) to 21,525.40 mg kg^{-1} (Fe). In contrast, the lowest values were found in Mo > Sr > Sn > Co > Li > Cd > Be, ranging from 0.30 (Be) to 8.30 mg kg^{-1} (Mo). Similar range values were observed by Coringa et al. (2014) and Souza et al. (2015) using USEPA 3051a method for geochemical evaluation of a wide variety of tropical soils.
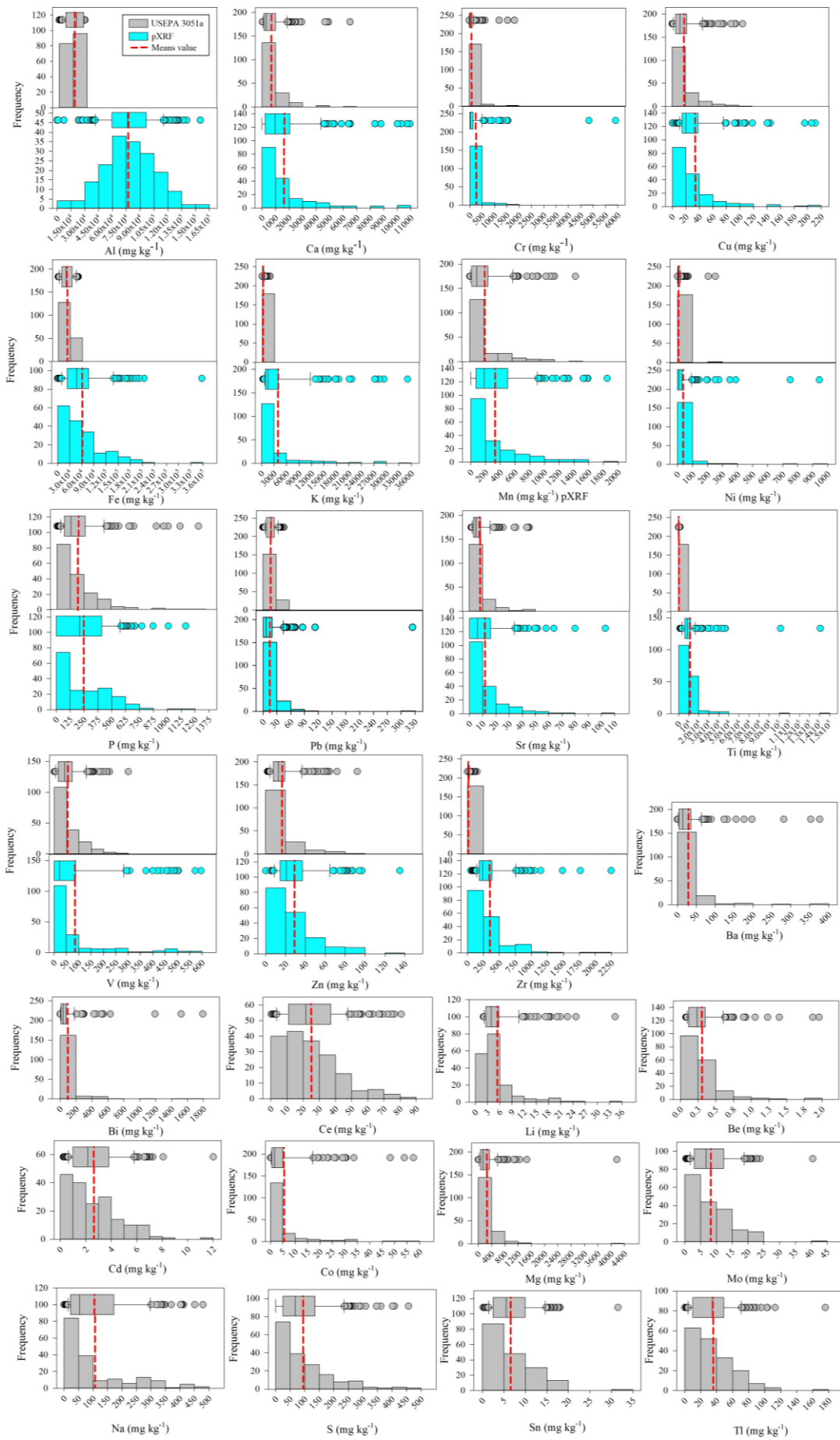
Fig. 2. Boxplots, histograms, and mean values for elemental content obtained by the USEPA 3051a (28 elements) and pXRF (15 elements) methods in Brazilian tropical soils.

The contents of Ba (29.60 mg kg^{-1}), Co (5.46 mg kg^{-1}), Cu (18.02 mg kg^{-1}), Mo (8.30 mg kg^{-1}), Ni (10.86 mg kg^{-1}), V (54.31 mg kg^{-1}) and Zn (16.88 mg kg^{-1}) are below the limits stablished by the Brazilian Environmental Legislation (CETESB, 2014; CONAMA, 2009; COPAM, 2011), that consider the USEPA 3051a method as standard analytical method for the extraction of inorganic substances in soils. Conversely, Cd (2.62 mg kg^{-1}) and Cr (92.84 mg kg^{-1}) are above the recommended levels, and can potentially poses risk to human health. High levels of Cr in tropical soils were also reported by Rosolen et al. (2015) and Souza et al. (2015), being attributed to agricultural practices and the chemical composition of parent materials.

The content of the 15 elements determined by both methods (USEPA 3051a and pXRF) presented completely different results (Fig. 2), which was expected due to the intrinsic characteristics of each method. Commonly, USEPA 3051a in tropical soils extract semi-total contents, mainly due to the non-dissolution of silicates. Thus, satisfactory recovery rates of elements are usually obtained from non-silicate minerals in different soil granulometric fractions, and from soil organic matter (SOM) fraction (Silva et al., 2019, 2014; USEPA, 2007b). Contrariwise, the elemental content delivered by pXRF is considered a total determination (Costa et al., 2019; Pelegrino et al., 2021; Ravansari et al., 2021; Silva et al., 2021), which explains the difference in results between the methods.

The results obtained by pXRF presented higher amplitudes, variability, and mean values for all 15 elements evaluated compared to the USEPA 3051a method. PXRF yielded the highest values for Al, Fe, Ti and K, whose values were: 77,407.56, 61,228.23, 11,361.60 and 4,012.00 mg kg^{-1}, respectively. This was likely to occur because Brazilian tropical soils have different proportions of minerals such as muscovite, ilmenite, and rutile in the sand fraction, as well as kaolinite, hematite, goethite, gibbsite and anatase in the silt and clay fractions (Brinatti et al., 2010; Kämpf et al., 2012; Melo et al., 2002b). The abundance of these minerals explains the high values of Al, Fe and Ti obtained in this study, which are present in the crystalline structure of these minerals. In addition to the presence in muscovite, K levels are also due to the presence of K-feldspar traces and agronomic amendments, such as potassium fertilization (Kämpf et al., 2012). Overall, the mean values for most of the 15 elements evaluated were consistent with Marques et al. (2004a) that evaluated the content of 20 elements via wavelength scattering X-ray fluorescence in Brazilian Cerrado soils.

The greater data variability determined by pXRF was mainly associated with the parent material diversity, which in turn directly influences the chemical composition of soils (Gozukara et al., 2021; Mancini et al., 2019b; Souza et al., 2021; Stockmann et al., 2016), in addition to different management practices, such as application of acidity correctives and fertilizers. The dataset in the present study presents soils developed from acidic rocks (e.g., granite, gneiss, sandstone, and quartzite) and soils developed from basic rocks (e.g., gabbro and basalt), explaining the greater variability of the chemical element contents.

3.2 *Correlations between USEPA 3051a and pXRF*

The correlation between the 28 elements obtained by the USEPA 3051a method with the 15 elements delivered by pXRF for all soil samples (n = 179) (Fig. 3) demonstrates the vast majority of positive correlations. Evaluating the content of the elements obtained by both methods (USEPA 3051a and pXRF), positive and strong correlations were observed for Cr (r = 0.98), Ni (r = 0.95), Ca (r = 0.87), Cu (r = 0.85), Fe (r = 0.82), Mn (r = 0.82), Zn (r = 0.78) and V (r = 0.71). The strong correlations are due to the fact that most of these elements can be found in tropical soils in the following situations: i) constituents of the crystalline structure of the Fe, Al, and Mn oxides (e.g., Cr and V) (Hseu et al., 2016; Marques et al., 2004a), ii) strongly adsorbed on clay-sized minerals, mainly iron oxides, as well as forming stable complexes with SOM (e.g., Cr, Ni, Cu, Zn and Mn) (Marques et al., 2004b; Meurer et al., 2010), and iii) present in soil solution, such as Ca and P, since most tropical soils have only traces of minerals containing this element, where their contents are basically derived from the application of correctives and soil conditioners (Brinatti et al., 2010; Kämpf et al., 2012; Lopes and Guilherme, 2016). Therefore, the contents of Cr, Ni, Ca, Cu, Fe, Mn, Zn and V are easily recoverable using the USEPA 3051a method, which in turn provided high correlation values with the data via pXRF. Similar results were reported by Silva et al. (2019) in tropical soils, that showed correlation indices between pXRF and USEPA 3051a for these same eight elements ranging from 0.57 (Zn) to 0.97 mg kg^{-1} (Cr).

USEPA 3051a

pXRF \ USEPA 3051a	Al	Ba	Be	Bi	Ca	Cd	Ce	Co	Cr	Cu	Fe	K	Li	Mg	Mn	Mo	Na	Ni	P	Pb	S	Sn	Sr	Ti	Tl	V	Zn	Zr
Al	0.52	-0.03	-0.02	0.18	-0.03	0.29	0.3	0	0.18	0.05	0.25	0.06	0.06	-0.03	-0.03	0.49	0.39	0.08	0.1	0.43	0.09	0.48	0.04	-0.05	0.19	0.15	0.19	0.01
Ca	-0.19	0.19	-0.04	-0.07	0.87	-0.21	-0.1	0.04	-0.07	0.01	-0.2	0.16	-0.13	0.34	0.2	-0.21	0	-0.01	0.31	-0.14	0.56	-0.19	0.77	-0.11	-0.14	-0.16	0.19	-0.11
Cr	0.26	0.01	0.12	0.98	0.03	0.44	0.02	0.59	0.98	0.27	0.37	-0.06	0	0	0.39	0.39	0.3	0.9	0.17	0.37	0.1	0.39	0.04	0.39	0.43	0.31	0.27	0.03
Cu	0.32	0.43	0.57	0.34	0.04	0.53	0.38	0.65	0.34	0.85	0.63	0.12	0.18	0.14	0.57	0.31	0.15	0.39	0.49	0.37	0.11	0.32	0.14	0.75	0.61	0.71	0.57	0.61
Fe	0.58	0.23	0.4	0.49	-0.05	0.87	0.4	0.56	0.49	0.59	0.82	0.03	0.09	0.04	0.43	0.66	0.18	0.45	0.41	0.62	0.06	0.68	0.05	0.77	0.89	0.71	0.43	0.52
K	-0.18	0.19	0.18	-0.13	-0.09	-0.18	0.01	-0.03	-0.13	-0.03	-0.13	0.52	0.27	0.22	0.04	-0.21	-0.04	-0.07	-0.15	-0.05	-0.17	-0.21	-0.06	-0.16	-0.14	-0.18	-0.01	-0.12
Mn	0.24	0.52	0.5	0.18	0.24	0.41	0.38	0.62	0.17	0.59	0.53	0.23	0.05	0.25	0.82	0.2	0.15	0.26	0.45	0.37	0.26	0.22	0.33	0.6	0.56	0.54	0.48	0.51
Ni	0.25	0.1	0.23	0.93	0.02	0.46	0.1	0.69	0.93	0.35	0.42	0	0.04	0.07	0.47	0.35	0.31	0.95	0.19	0.36	0.07	0.35	0.05	0.43	0.48	0.36	0.31	0.15
P	-0.13	0.28	0.12	-0.09	0.39	-0.12	0	0.11	-0.09	0.25	-0.07	0.29	0.07	0.21	0.2	-0.17	0.1	-0.04	0.5	-0.08	0.32	-0.16	-0.03	-0.06	-0.02	0.31	-0.02	
Pb	0.17	0.09	0.18	0.22	-0.08	0.5	0.06	0.28	0.22	0.21	0.34	0	0.02	-0.02	0.16	0.26	0	0.22	0.13	0.33	-0.01	0.28	-0.04	0.55	0.55	0.26	0.13	0.3
Sr	-0.09	0.25	0.18	-0.04	0.02	0.02	0.1	0.2	-0.04	0.11	0.07	0.21	-0.01	0.21	0.14	-0.07	-0.06	0.06	0.05	-0.05	-0.04	-0.06	0.1	0.09	0.08	0.02	0.02	0.47
Ti	0.08	0.06	0.1	0.01	-0.1	0.23	0.18	0.11	0	0.19	0.23	-0.05	-0.05	0	0.06	0.14	-0.11	0.04	0.1	0.2	0.05	0.16	-0.04	0.37	0.23	0.25	0.04	0.58
V	0.38	0.25	0.4	0.43	-0.01	0.63	0.35	0.54	0.43	0.63	0.63	-0.02	0.1	0.08	0.44	0.47	0.08	0.43	0.39	0.43	0.06	0.48	0.07	0.81	0.64	0.71	0.4	0.69
Zn	0.4	0.56	0.58	0.37	0.21	0.44	0.4	0.66	0.37	0.78	0.55	0.31	0.19	0.3	0.64	0.3	0.28	0.42	0.62	0.43	0.31	0.31	0.33	0.55	0.53	0.57	0.78	0.35
Zr	-0.31	-0.14	-0.23	-0.19	-0.1	-0.21	-0.31	-0.18	-0.19	-0.2	-0.28	-0.17	-0.17	-0.13	-0.25	-0.21	-0.24	-0.16	-0.16	-0.25	-0.01	-0.2	-0.07	-0.1	-0.27	-0.24	-0.21	0.12

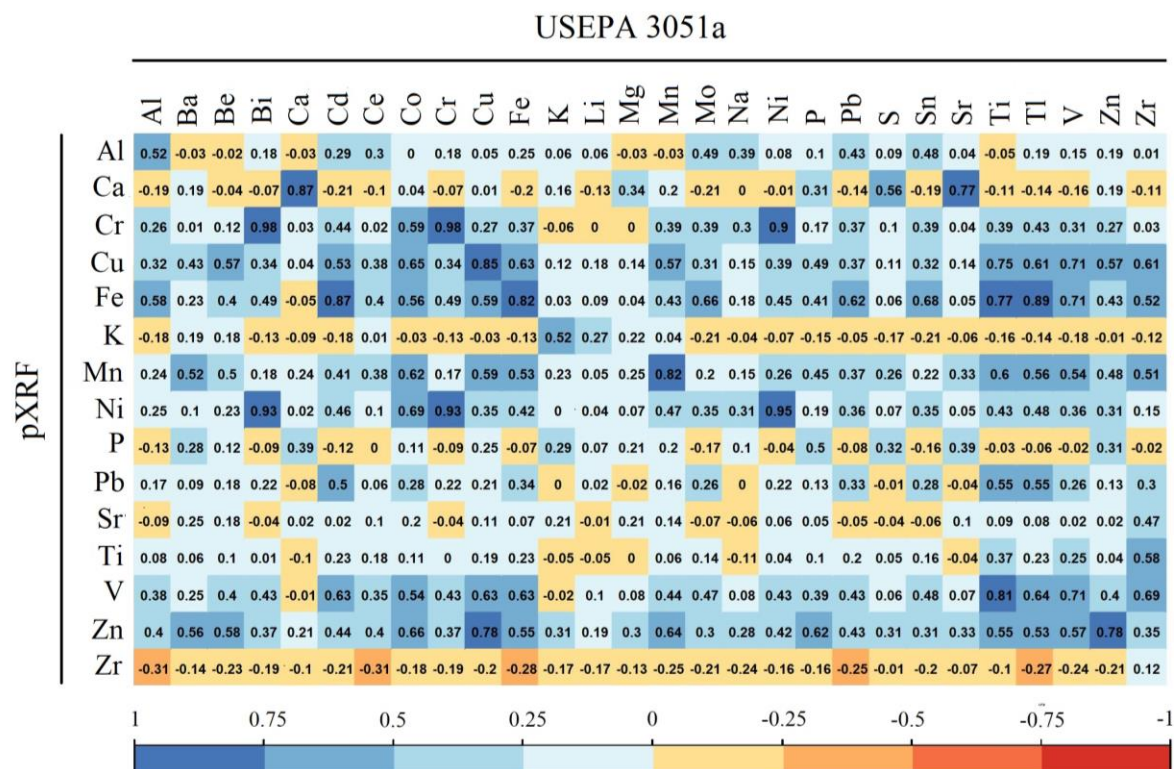1   0.75   0.5   0.25   0   -0.25   -0.5   -0.75   -1

Fig. 3. Color correlation between the elemental contents obtained by the standard method USEPA 3051a (28 elements) and by pXRF method (15 elements) in Brazilian tropical soil samples (n = 179).

The fact that SOM forms stable complexes with most divalent and trivalent cations in soil, making them easily recovered by both methods (USEPA 3051a and pXRF), may also explain other highly significant correlations observed, such as: $Bi_{USEPA\ 3051a}$ (Bi^{3+}) with Cr_{pXRF} (Cr^{3+}) (r = 0.98) and $Bi_{USEPA\ 3051a}$ (Bi^{3+}) with Ni_{pXRF} (Ni^{2+}) (r = 0.93), in addition to the results between $Cu_{USEPA\ 3051a}$ (Cu^{2+}) and Zn_{pXRF} (Zn^{2+}) (r = 0.78) and $Cr_{USEPA\ 3051a}$ and Ni_{pXRF} (r = 0.93) (Fig. 3). The close relationship between most of these elements and SOM can be seen in studies carried out by Lima et al. (2019), Faria et al. (2022), Murata (2010) and Hou et al. (2005).

The results also demonstrated that Cd content (USEPA 3051a) was strongly correlated (r = 0.87) with Fe (pXRF) (Fig. 3). Regarding the high correlation values between $Sr_{USEPA\ 3051a}$ and Ca_{pXRF} (r = 0.77), Drouet et al. (2007) and Kabata-Pendias (2010) explain that both elements have similar geochemical and biochemical behavior during weathering processes. The results also demonstrated that Ti via USEPA 3051a was well correlated with Cu (r = 0.75), Fe (r = 0.77) and V (r = 0.81) delivered by pXRF. A correlation value of 0.89 was obtained between $Tl_{USEPA\ 3051a}$ and Fe_{pXRF}, due to Tl being associated with Fe oxides, and to

SOM (Kabata-Pendias and Mukherjee, 2007) which, consequently, indirectly explains the high correlations between Tl and Cr, Cu and Zn.

Considering that the methods are semi-total (USEPA 3051a) and total (pXRF) determination, the results showed that both methods tested can provide similar recovery values for many elements present in tropical soils, indicating that they can be used as a basis for the development of promising and accurate prediction models in the geochemical assessment of soils from tropical regions. More details will be discussed in the following sections.

In Fig. 3 is also shown positive and moderate correlations for Al (r = 0.52), K (r = 0.52) and P (r = 0.50), as well as low or no positive correlations for Ti (r = 0.37), Pb (r = 0.33), Zr (r = 0.12), and Sr (r = 0.10) obtained by both the USEPA 3051a and pXRF methods in soils. Despite the correlation results delivered for Al, K and Zr were superior to those found by Silva et al. (2019) and Faria et al. (2022), due to the presence of these elements in the crystal structure of silicate minerals not digested by USEPA 3051a, stronger correlations between the methods were not found. Thus, the contents of Al, K, Ti and Zr was not fully recovered by the USEPA 3051a method, since this method has a greater capacity to dissolve non-silicate minerals present in the different granulometric fractions of tropical soils (Abbruzzini et al., 2014; Chen and Ma, 1998; Silva et al., 2014). The use of pXRF provides the ability to identify and quantify the total multielemental content of soil (Silva et al., 2021; Weindorf et al., 2014), which explains the higher content of these elements compared to the traditional method (Fig. 2) and, consequently, the low correlation results (Fig. 3).

Contrariwise, an unexpected result occurred since in tropical soils Ti is found mainly in non-silicate minerals (e.g., ilmenite, rutile and anatase) present in the sand and clay fraction (Kämpf et al., 2012; Marques et al., 2004a; Melo et al., 2002a) which, in turn, should be digestible by USEPA 3051a. However, the standard method used was not able to satisfactorily recover this element when compared to pXRF, resulting in a low correlation (r = 0.37) (Fig. 3). This result is probably due to the large amount of these minerals in the coarse fraction (e.g., ilmenite and rutile), resistant to weathering and with a tendency to accumulate in most tropical soils, interfering on USEPA 3051a digestion process.

Studies with low or no correlation between Al, K, Ti and Zr contents obtained by traditional methodologies and by pXRF are already reported in soils of the tropical regions (e.g., Faria et al., 2022; Silva et al., 2019). However, Declercq et al. (2019) evaluating 128 soil samples collected in 10 different countries, mostly distributed in temperate regions,

reported correlations above 0.80 for K and Al obtained between pXRF and acid digestion methods.

The low correlation results for Pb and Sr, commonly found in minerals such as galena and celestite (non-silicate) in the clay fraction, may be due to the low natural concentration of these elements in tropical soils (Kabata-Pendias, 2010; Marques et al., 2004a), and to the great variability of factors in this study (e.g., degree of soil weathering, parent materials, collection depth, soil use and management practices) which may have contributed to a non-linear relationship between USEPA 3051a and pXRF. This variability of factors may also have contributed to the large amount of low correlations (negative or positive) between elemental contents via USEPA 3051a and pXRF, especially for elements with low concentrations, such as: Ce, Li, Be, Na and Sn (Fig. 3).

Finally, P showed moderate and positive correlation (r = 0.50) between methods, which was already expected due to the main source of this element in Brazilian soils being of anthropic origin via successive applications of phosphate fertilizers (Benedet et al., 2021; Yang and Post, 2011), easily determined by both methods. In tropical soils, Silva et al. (2019) and Lima et al. (2019) found strong positive correlations (0.71 and 0.68) between P content via digestion methods and pXRF.

3.3 *Modeling and validation from pXRF data via*:

3.3.1 Simple linear regression and stepwise multiple linear regression

Models generated through simple linear regressions (SLR) and stepwise multiple linear regression (SMLR) with their respective R^2 values for prediction of Al, Ca, Cr, Cu, Fe, K, Mn, Ni, P, Pb, Sr, Ti, V, Zn and Zr are shown in Table 1. In general, the R² values of the models generated by the SLR ranged from 0.00 (Zr) to 0.96 (Cr and Ni), while for SMLR, the R² values ranged between 0.38 (Zr) and 0.84 (Cr).

Table 1. Prediction models generated by simple linear regression (SLR) and stepwise multiple linear regression (SMLR) with their respective R² to predict 15-element results via the USEPA 3051a method from portable X-ray fluorescence (pXRF) spectrometry data in Brazilian tropical soils.

Elements		Prediction Models	R^2
Al	SLR	Al $_{USEPA\ 3051a}$ = 1.6198 Al$_{pXRF}$ + 52075	0.26
	SMLR	Al $_{USEPA\ 3051a}$ = 15780.7 + 4390.5Al + 6442.7Fe - 1333.5Mn + 1908.3P - 1987.1Pb + 1348.7Ti	0.50
Ca	SLR	Ca $_{USEPA\ 3051a}$ = 1.941 Ca$_{pXRF}$ + 278.56	0.82
	SMLR	Ca $_{USEPA\ 3051a}$ = 745.25 + 126.71Al + 870.88Ca + 68.02Fe - 97.93K + 126.59P - 57.21Ti	0.83
Cr	SLR	Cr $_{USEPA\ 3051a}$ = 3.5078 Cr$_{pXRF}$ - 93.586	0.96
	SMLR	Cr $_{USEPA\ 3051a}$ = 95.880 + 6.785Al + 195.626Cr - 19.065Cu + 32.920Fe + 30.028Ni + 6.253P - 19.421Pb + 10.676Ti	0.84
Cu	SLR	Cu $_{USEPA\ 3051a}$ = 1.5395 Cu$_{pXRF}$ + 5.2731	0.75
	SMLR	Cu $_{USEPA\ 3051a}$ = 17.0199 - 6.3559Cr + 13.6047Cu + 3.9761Fe + 1.2706K - 2.3862Mn + 5.6033Ni + 2.3353P - 1.8938Pb + 2.2116Zn	0.76
Fe	SLR	Fe $_{USEPA\ 3051a}$ = 2.6233 Fe$_{pXRF}$ + 1095.8	0.74
	SMLR	Fe $_{USEPA\ 3051a}$ = 20477.1 - 1186.7Ca - 8992.0Cr - 1922.2Cu + 16909.3Fe + 1090.1K + 8337.2Ni + 1361.4P - 4665.2Pb + 1979.3Ti	0.74
K	SLR	K $_{USEPA\ 3051a}$ = 11.577 K$_{pXRF}$ + 1426.2	0.28
	SMLR	K $_{USEPA\ 3051a}$ = 288.59 + 96.69Fe + 102.58K - 84.04Pb + 128.68Sr - 82.86V + 79Zn	0.46
Mn	SLR	Mn $_{USEPA\ 3051a}$ = 1.0197 Mn$_{pXRF}$ + 92.515	0.72
	SMLR	Mn $_{USEPA\ 3051a}$ = 221.771 - 17.807Al + 39.183Cu - 56.842Fe - 27.411K + 268.869Mn + 134.391Ni + 43.331Sr - 25.153Ti - 56.313V	0.78
Ni	SLR	Ni $_{USEPA\ 3051a}$ = 3.6101 Ni$_{pXRF}$ - 3.4733	0.96
	SMLR	Ni $_{USEPA\ 3051a}$ = 11.6196 - 1.4216Al - 1.0173K + 1.5590Mn + 30.2747Ni - 0.9719P - 1.8194Pb + 2.9803Sr	0.73
P	SLR	P $_{USEPA\ 3051a}$ = 0.5647 P$_{pXRF}$ + 146.97	0.26
	SMLR	P $_{USEPA\ 3051a}$ = 198.37 + 23.01Al + 32.39Ca - 61.09Cu + 92.35Fe - 62.14K - 27.19Ni + 110.22P + 110.02Zn	0.53
Pb	SLR	Pb $_{USEPA\ 3051a}$ = 0.7495 Pb$_{pXRF}$ + 0.4526	0.17
	SMLR	Pb $_{USEPA\ 3051a}$ = 18.1187 + 3.0021Al + 1.3258Cr - 2.0862Cu + 7.1281Fe + 1.5390P + 3.7537Ti - 2.7515V + 1.7137Zn - 1.8988Zr	0.50
Sr	SLR	Sr $_{USEPA\ 3051a}$ = 0.6803 Sr$_{pXRF}$ + 7.4886	0.27
	SMLR	Sr $_{USEPA\ 3051a}$ = 8.3426 + 0.9898Al + 5.7868Ca + 1.5523Fe - 2.1433K + 1.2559P + 2.0973Sr - 0.7700Ti - 1.1799V + 1.2941Zn	0.57
Ti	SLR	Ti $_{USEPA\ 3051a}$ = 14.607 Ti$_{pXRF}$ + 8462.9	0.08
	SMLR	Ti $_{USEPA\ 3051a}$ = 125.93 - 29.7Al + 108.63Cr + 94.59Cu - 86.17Ni - 33.99P - 33.98Pb + 52.2Ti + 127.86V - 33.81Zr	0.70
V	SLR	V $_{USEPA\ 3051a}$ = 2.1084 V$_{pXRF}$ - 28.759	0.63

	SMLR	$V_{\text{USEPA 3051a}} = 49.733 - 3.973Ca - 27.706Cr + 8.798Cu + 22.511Fe + 26.032Ni - 12.843Pb + 5.935Ti + 18.676V$	0.64
Zn	SLR	$Zn_{\text{USEPA 3051a}} = 1.2828\ Zn_{\text{pXRF}} + 7.9433$	0.64
	SMLR	$Zn_{\text{USEPA 3051a}} = 16.4514 - 3.3954Cu + 2.7334Fe - 1.1970K - 1.4173Mn + 3.0195P - 2.8368Pb + 1.6409Ti + 13.5105Zn$	0.66
Zr	SLR	$Zr_{\text{USEPA 3051a}} = -1.1264\ Zr_{\text{pXRF}} + 346.43$	0.00
	SMLR	$Zr_{\text{USEPA 3051a}} = 15.4435 + 3.0327K - 4.2521Ni - 2.1370P - 6.6159Pb - 2.1958Sr + 5.3779Ti + 16.6906V$	0.38

The results showed that both SLR and SMLR provided promising prediction models for Ca (R^2 = 0.82 and 0.83), Cr (R^2 = 0.96 and 0.84), Cu (R^2 = 0.75 and 0.76), Fe (both with R^2 = 0.74), Mn (R^2 = 0.72 and 0.78), Ni (R^2 = 0.96 and 0.73), V (R^2 = 0.63 and 0.64), Ti (R^2 = 0.70 for SMLR only) and Zn (R^2 = 0.64 and 0.66) (Table 1). Similar results were observed by Borges et al. (2020b) and Faria et al. (2022) using SLR with R^2 values from 0.61 (V) to 0.97 (Fe).

Models with regular or low accuracy due to their low R^2 values (from 0.00 Zr – SLR to 0.57 Sr – SMLR), were generated for Al, K, P, Pb, Sr, Zr and Ti, the latter just by SLR. The results suggest that there are non-linear relationships between the content obtained by both USEPA 3051a and pXRF methods for some elements and the use of more robust non-linear algorithms may provide better results (Wan et al., 2020). The generation of unreliable models is due to the occurrence of elements in the crystalline structure of silicate (e.g., Al, K and Zr) and non-silicate (e.g., P and Ti) minerals present mainly in the sand fraction of tropical soils (Brinatti et al., 2010; Kämpf et al., 2012) and, consequently, not accessed by the traditional method USEPA 3051a (Chen and Ma, 1998; Silva et al., 2014). The low concentrations of Pb and Sr in tropical soils, together with the great variability of factors in this study, may have contributed to a non-linear response of these elements.

It was observed in the models generated by the SMLR that of the 15 elements provided by pXRF, between 6 and 10 elements were incorporated into the models (Table 1), where Fe and P appeared more frequently (both in 12 equations), followed by Ti (11 equations), Al (9 equations), Ni (9 equations) and Cu (9 equations). This result demonstrates that some elements that are easily accessed with pXRF can be well correlated with the occurrence of other elements in soil, facilitating the development of prediction models.

The validation indices (R^2, RPD, RMSE, NRMSE) between observed and predicted values generated for validation sub-dataset (n = 54) using SLR and SMLR, as well as their graphical representation (1:1 graphs), can be observed in Fig. 4. Overall, the models delivered

by SMLR showed better performance in predicting the 15 elements compared to models via SLR. According to Forkuor et al. (2017), SMLR is one of the widely used linear regression techniques and has a good ability to handle data multicollinearity in estimating soil attributes based on linear relationships between soil properties and pXRF elemental contents. However, only the model for predicting Cr via SMLR was considered reliable and with excellent prediction capacity (R^2 = 0.97, RPD = 3.37, RMSE = 54.24 mg kg^{-1} and NRMSE = 0.30). Conversely, the validation results for the prediction models for Ca, Cu, P, Pb, Ti and Zn were classified between good and regular with reasonable prediction capacity, with R^2 and RPD values ranged from 0.49 to 0.70 and 1.40 to 1.65 respectively, and consequently presented the lowest RMSE and NRMSE.
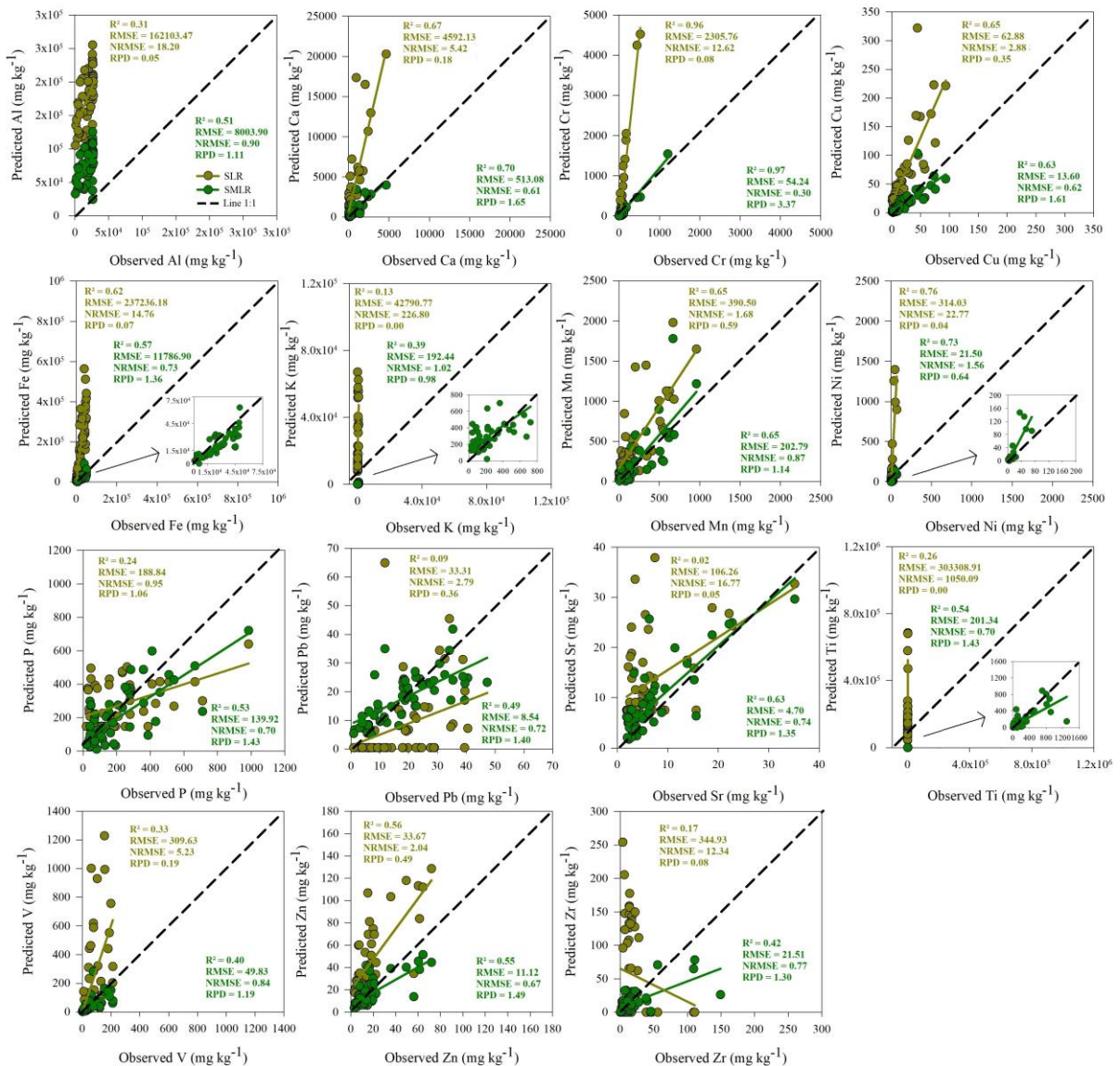
Fig. 4. Scatter plots and validation parameters between observed laboratory values (USEPA 3051a) *versus* values predicted by simple (SLR) and multiple linear regression (SMLR) for 15 elements using data obtained by pXRF in Brazilian tropical soils.

Despite the SLR having provided relatively high R^2 values for some elements (e.g., Ca, Cr, Cu, Fe, Mn and Ni), the models in general were classified as poor and unreliable for the prediction, according to the RPD, RMSE and NRMSE values (Chang et al., 2001). High RPD combined with low RMSE and NRMSE values is recommended to ensure the smallest errors, better performance and reliability for the model (Chai and Draxler, 2014; Jalali et al., 2019; Taebi and Mansy, 2017), which did not occur with the models generated by the SLR and for some models via SMLR. This can be the result of non-linear nature of some elements determined by USEPA 3051a and pXRF, mainly in the validation of Al, K, Ti and Zr. The validation for these elements were already expected to be low according to the correlation analysis (Fig. 3) and the R^2 of the models (Table 1). Similar results were observed by others authors in soils of tropical region (e.g., Andrade et al., 2020; Benedet et al., 2021; Faria et al., 2022; Santana et al., 2018; Silva et al., 2019). The use of machine learning algorithms for these elements as well as for others is encouraged and will be discussed below.

3.3.2 Validation via random forest and support vector machine

Based on the validation results presented in Fig. 5, the use of random forest (RF) and support vector machine (SVM) algorithms provided promising models, capable of predicting accurately and reliably for 13 out of the first 15 elements evaluated in the studied soils when compared to SLR and SMLR. Despite of the good performance of the SVM, the models generated by RF were superior to all other models generated (e.g., SVM, SLR and SMLR) and provided the highest values of R^2 (between 0.51 and 0.83) and RPD (between 0.51 and 0.83) (Fig. 5). Therefore, the values predicted by RF for each element had a smaller dispersion in relation to the 1:1 line, and were close to those measured in the laboratory (Fig. 5). However, a more detailed observation shows that RF had difficulty in predicting high values measured in the laboratory (e.g, P, Ti, V, Zn and Zr) underestimating the results and negatively influencing the validation parameters. The underestimations may be due to the relative scarcity of samples with high contents at the upper ends of each element, making it impossible to effectively train the models for these situations, as already described by Wang et al. (2015) and Benedet et al. (2021).
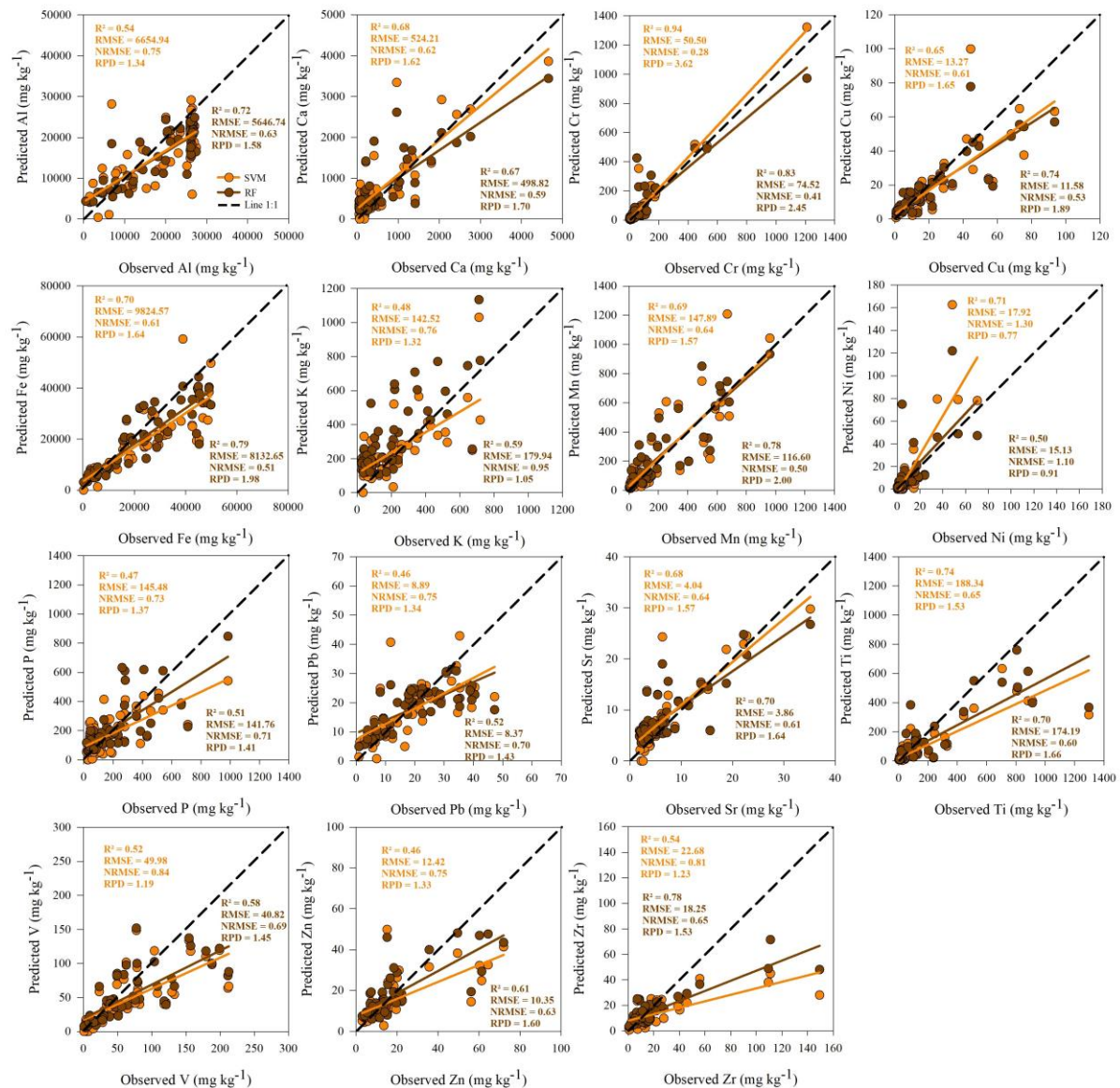
Fig. 5. Scatter plots and validation parameters between measured/observed laboratory values (USEPA 3051a) *versus* values predicted by Support Vector Machine with (SVM) linear kernel and Random Forest (RF) for 15 elements using datasets obtained by pXRF in Brazilian tropical soils.

The use of RF generated a greater number of models classified as reliable and with great predictive capacity, such as: Cr, Fe and Mn. The R^2 values for these elements were 0.83, 0.80 and 0.78, and values for RPD of 2.45, 2.00 and 2.00, respectively. These elements also stand out from the others for providing the lowest NRMSE values (Cr = 0.41, Fe = 0.51 and Mn = 0.50) among all the first 15 predicted elements, through the tested models. For Cr, despite the excellent performance of RF, it is important to show that the models delivered by

SVM and SMLR were superior (R² = 0.94 and 0.97; RPD = 3.62 and 3.37; NRMSE = 0.28 and 0.30, respectively), and considered the best among all elements (n = 28) predicted in this study.

The number of models considered good and regular with reasonable prediction capacity also increased with the use of RF algorithm and was observed for Al, Ca, Cu, P, Pb, Sr, Ti, V, Zn and Zr, with R² values between 0.51 (P) to 0.78 (Zr), RPD between 1.41 (P) to 1.89 (Cu), NRMSE between 0.53 (Cu) to 0.71 (P) and one of the lowest RMSE results. For P, the model generated by SMLR was still superior to machine learning algorithms (SVM and RF) for delivering the highest values of R² = 0.53, RPD = 1.43, RMSE = 139.92 mg kg^{-1} and NRMSE = 0.70, delivering a regular model with reasonable predictive ability.

According to the literature, RF algorithm is one of the most promising machine learning techniques currently available to perform predictions of various soil attributes in tropical and temperate regions (Chagas et al., 2016; Dharumarajan et al., 2017; Faria et al., 2022; Mancini et al., 2020; Nawar et al., 2019; Wang et al., 2015). A similar behavior was observed for the prediction of semi-total elemental content in heterogeneous tropical soils, as it was shown that among the 15 elements detailed in Fig. 4, where RF promoted promising results for 13 of them (Fig. 5). However, even with the use of robust algorithms, it was still not possible to obtain satisfactory models for K and Ni.

3.4 *Modeling and validation for 13 remaining elements from pXRF data*

3.4.1 Stepwise multiple linear regression, support vector machine and random forest

Models generated by stepwise multiple linear regression (SMLR) with their respective R^2 values for prediction of 13 elements (e.g., Ba, Be, Bi, Cd, Ce, Co, Li, Mg, Mo, Na, S, Sn and Tl) acquired via USEPA 3051a based on data obtained by pXRF are shown in Table 2. The R² values of the models generated by the SMLR ranged from 0.14 (Li) to 0.86 (Bi). However, only the model for predicting Bi was interpreted as very good (R² ≥ 0.81), while models for predicting Cd, Mo, Sn and Tl were classified as good (R² between 0.61 to 0.80). Similarly, to the results presented in Table 1, these results also indicate the potential of using pXRF to provide promising elementary variables that fit the prediction models of elements that are uncommon in works with this approach in heterogeneous tropical soils, therefore, it contributes for the most complete geochemical characterization.

Table 2. Models generated by stepwise multiple linear regression with their respective R² to predict results of 13 elements via acid digestion (USEPA 3051a) from portable X-ray fluorescence spectrometry (pXRF) data in tropical soils.

Elements	Prediction Model	R²
Ba	$Ba_{USEPA\ 3051a} = 34.515 - 6.319Al + 15.060Cu - 7.306K + 15.646Mn - 10.356Pb + 23.412Sr - 17.619V + 26.910Zn$	0.34
Be	$Be_{USEPA\ 3051a} = 0.31847 - 0.06797Al - 0.04745Ca - 0.22001Cr + 0.19155Cu + 0.06862Fe - 0.04348K + 0.22249Ni - 0.05450P - 0.04400Pb + 0.17737Sr - 0.11980V + 0.06088Zn - 0.06159Zr$	0.48
Bi	$Bi_{USEPA\ 3051a} = 96.668 + 7.578Al + 192.236Cr - 19.687Cu + 34.797Fe + 28.057Ni + 6.670P - 19.66Pb + 10.837Ti$	0.86
Cd	$Cd_{USEPA\ 3051a} = 2.39769 + 0.27973Al + 1.71069Fe - 0.28504Mn + 0.16583Ni + 0.19971P - 0.33858Pb + 0.18738Ti$	0.77
Ce	$Ce_{USEPA\ 3051a} = 25.100 + 4.453Al - 14.885Cr + 8.848Fe + 2.063K + 12.171Ni - 6.362Pb + 7.294Ti - 5.839Zr$	0.46
Co	$Co_{USEPA\ 3051a} = 5.6312 - 0.7841Al + 2.0686Cu - 1.0912K + 3.2285Mn + 5.5202Ni + 0.9930Pb + 1.3571Sr - 1.6918Ti$	0.56
Li	$Li_{USEPA\ 3051a} = 5.9269 - 0.8942Ca - 4.3790Cr + 2.3372Cu + 0.8718K - 0.8471Mn + 4.4136Ni - 0.7992Pb + 1.2483Sr$	0.14
Mg	$Mg_{USEPA\ 3051a} = 288.21 - 73.47Al + 61.31Ca - 61.89K - 100.68P - 119.29Pb + 288.25Sr + 117.36Zn - 73.12Zr$	0.34
Mo	$Mo_{USEPA\ 3051a} = 7.5068 + 2.1338Al + 2.9126Cr - 0.9806Cu + 5.0856Fe - 1.0392Mn - 2.2942Ni + 1.0919P - 1.1572Pb + 0.7677Ti$	0.70
Na	$Na_{USEPA\ 3051a} = 112.307 + 60.174Al + 132.061Cr - 97.033Ni + 33.255P$	0.31
S	$S_{USEPA\ 3051a} = 97.393 + 41.123Ca + 91.392Cr - 27.323Cu - 26.196K - 83.421Ni + 11.028P + 16.584Ti - 20.507V + 53.520Zn$	0.41
Sn	$Sn_{USEPA\ 3051a} = 5.9821 + 1.6313Al + 2.0528Cr - 0.8654Cu + 4.0085Fe - 0.7023Mn - 1.5654Ni + 0.8992P - 0.9117Pb + 0.6360Ti$	0.69
Tl	$Tl_{USEPA\ 3051a} = 32.722 - 9.47Cr - 2.817Cu + 24.846Fe + 11.641Ni + 2.330P - 3.861Pb$	0.79

Among the 15 elements determined by pXRF, from 4 to 13 elements were incorporated into the models (Table 2), as they present satisfactory correlation values (Fig. 3) with USEPA 3051a method. In the 13 models, it is observed that Al appeared the most (10 equations), followed by Cr and Cu (both in 9 equations). Cr and Cu measured by pXRF showed moderate to high correlations for most elements obtained by the USEPA 3051a method (Fig. 2), which favored the inclusion of both elements in most prediction models. Among them, the prediction model for Bi stands out, which obtained a high value of R² = 0.86. The probable explanation for Al and Fe inclusion in several models (Table 1 and 2) is due to the fact that these elements tend to accumulate residually in large amounts in tropical

soils, as they are part of the crystalline structure of several minerals present in the clay fraction, as well as their interaction or geochemical behavior are similar to other elements (Brinatti et al., 2010; Kabata-Pendias, 2010; Kämpf et al., 2012).

The performance of the models generated by SMLR, SVM and RF algorithms for the prediction of the 13 elements via the USEPA 3051a method are presented in Fig. 6. Overall, the validation for the 13 elements was promising both in the use of SMLR and in the use of SVM and RF. However, only for Bi that all models tested were classified as reliable and with excellent predictive capacity with $R^2 \geq 0.80$, RPD ≥ 2.00 and very low values of RMSE and NRMSE. Thus, the validation parameters for Bi delivered $R^2 = 0.81$ (RF), 0.94 (SVM) and 0.96 (SMLR), RPD = 2.30 (RF), 3.37 (SMLR) and 3.49 (SVM), as well as NRMSE = 0.29 (SMLR)), 0.30 (SVM) and 0.43 (RF) which, together with those observed for Cr (SVM and SMLR with NRMSE = 0.28 and 0.30, respectively), were the best and most accurate results among all elements (n = 28) tested in this study.
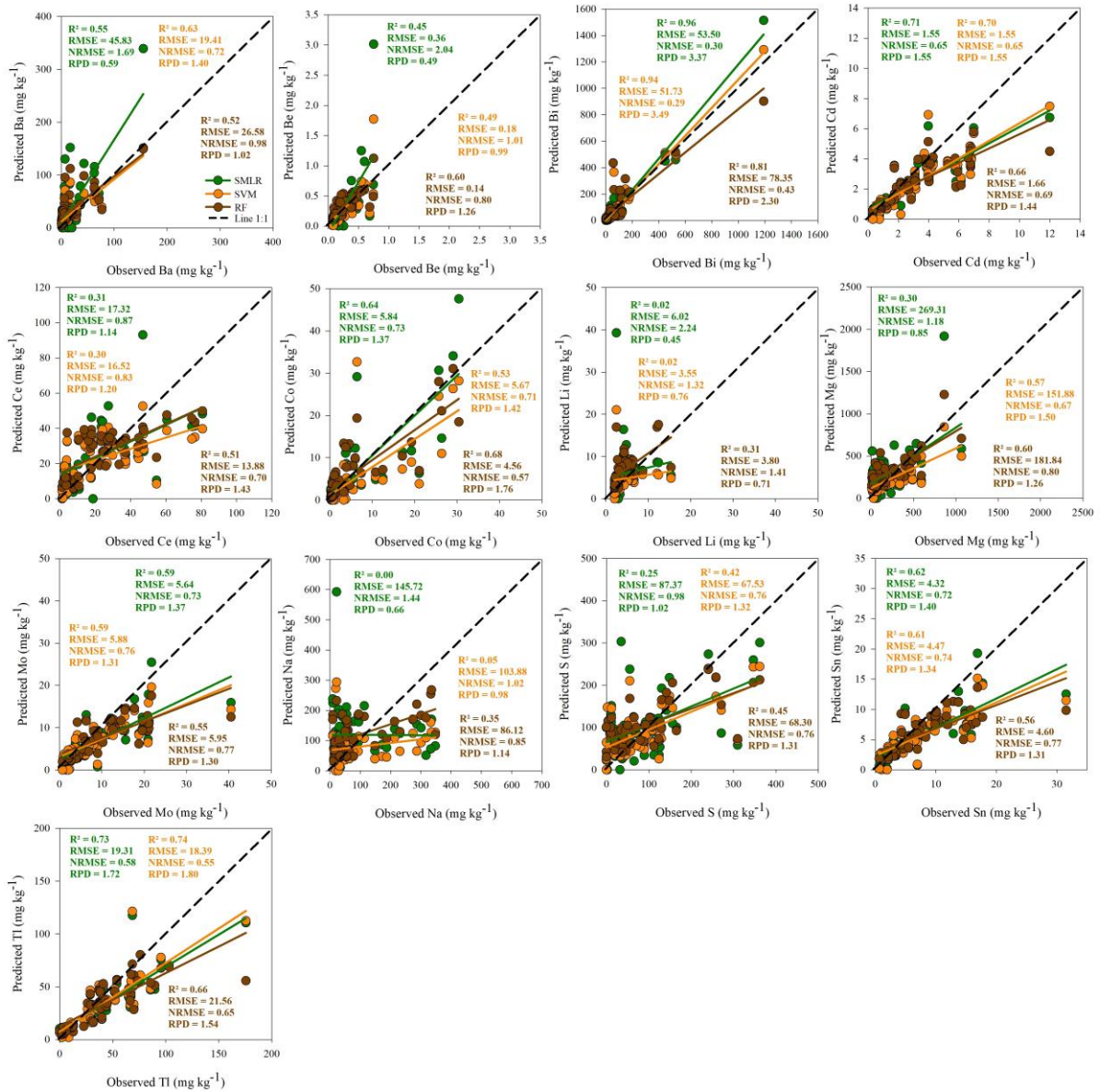
Fig. 6. Scatter plots and validation parameters between measured/observed values in the laboratory (USEPA 3051a) *versus* values predicted by Stepwise Multiple Linear Regression (SMLR), Support Vector Machine (SVM) with linear kernel and Random Forest (RF) for 13 elements using datasets obtained by pXRF in Brazilian tropical soils.

Unlike Cr, Bi is considered a rare metal in the earth's crust and is usually found associated with some metals (e.g., Ag, Co, Pb, Zn, Al and Fe). Thus, Cr contents in soils are much higher than Bi, as observed in Fig. 2, where both are directly inherited from the parent material, mainly in ultramafic and mafic igneous rocks. Even with the lowest levels of Bi in the soils, after the weathering of the minerals bismuthinite (Bi_2S_3) and bismite (Bi_2O_3), there is a direct contribution of SOM and contents of Fe and Al in the retention of this element in

tropical soils (Hou et al., 2005; Kabata-Pendias, 2010), which explains the high correlations of Bi with Ni, Cr and Zn, metals complexed by MOS, and moderate correlations (r = 0.49) mainly with Fe (Fig. 3).

The prediction of Cd and Tl also provided satisfactory and similar validation results, regardless of the model generated. For Cd, models generated by SMLR and SVM showed more accuracy by delivering values of R^2 = 0.71, RPD = 1.55, RMSE = 1.55 mg kg^{-1} and NRMSE = 0.65 for both. As for Tl, SMLR (R^2 = 0.73, RPD = 1.72, RMSE = 19.31 mg kg^{-1} and NRMSE = 0.58) and SVM (R^2 = 0.74, RPD = 1.80, RMSE = 18.39 mg kg^{-1} and NRMSE = 0.55) delivered models with superior validation. Unlike Bi, Cd and Tl, the best prediction performances for Ba, Co and Sn were achieved by only one prediction model (Fig. 6). Thus, the results delivered for Ba showed that the use of SVM was better than the others (R^2 = 0.63, RPD = 1.40, RMSE =19.41 mg kg^{-1} and NRMSE = 0.72). As for the prediction of Co, the model generated by the RF was superior to the results delivered by SMLR and SVM, with values of R^2, RPD, RMSE and NRMSE corresponded to 0.68, 1.76, 4.56 mg kg^{-1} and 0.57, respectively. In contrast, for Sn, the validation values generated by SMLR provided the best result (R^2 = 0.62, RPD = 1.40, RMSE = 4.32 mg kg^{-1} and NRMSE = 0.72) when compared with other algorithms tested. Therefore, the models generated for Cd, Tl, Ba, Co and Sn were classified as good and with reasonable prediction capacity (R^2 between 0.61 to 0.80, RPD between 1.4 to 2.00, as well as low values of RMSE and NRMSE).

For Ce and Mg, the results obtained via RF and SVM, respectively, were also relevant and promising, classified with reasonable prediction capacity (R^2 between 0.41 to 0.60 and RPD between 1.4 to 2.00). Due to the relative scarcity of samples with high contents, the best models also had difficulty in predicting mainly high values measured in the laboratory, underestimating the results, and negatively influencing the validation parameters. Similar to what was observed for K and Ni, for the elements Be, Li, Mo, Na and S, the use of pXRF data alone was not able to generate satisfactory predictive models, regardless of the use of linear regressions or algorithms of machine learning.

In summary, in terms of predictive performance presented in this study, the use of the SVM and RF algorithms produced excellent adjustments in relation to linear models by providing high values of R^2 and RPD, the lowest values of RMSE and especially the lowest values of NRMSE (Fig. 7). Despite the RF being considered the best prediction model for 18 elements, it is also important to highlight that the use of this algorithm provided accurate and reliable prediction models for 13 elements (e.g., Al, Ca, Cu, Fe, Mn, Pb, Sr, Ti, V, Zn, Zr, Ce and Co), since SVM and SMLR provided such an approach in only six (e.g., Cr, Ba, Bi, Cd,

Mg and Tl) and three (e.g, P, Cd and Sn) elements, respectively (Table 3). Thus, even using a large and heterogeneous dataset, the RF was able to deliver R^2 values between 0.52 and 0.79, RPD ranging from 1.43 to 2.00, NRMSE between 0.50 and 0.70 and the lowest values of RMSE.
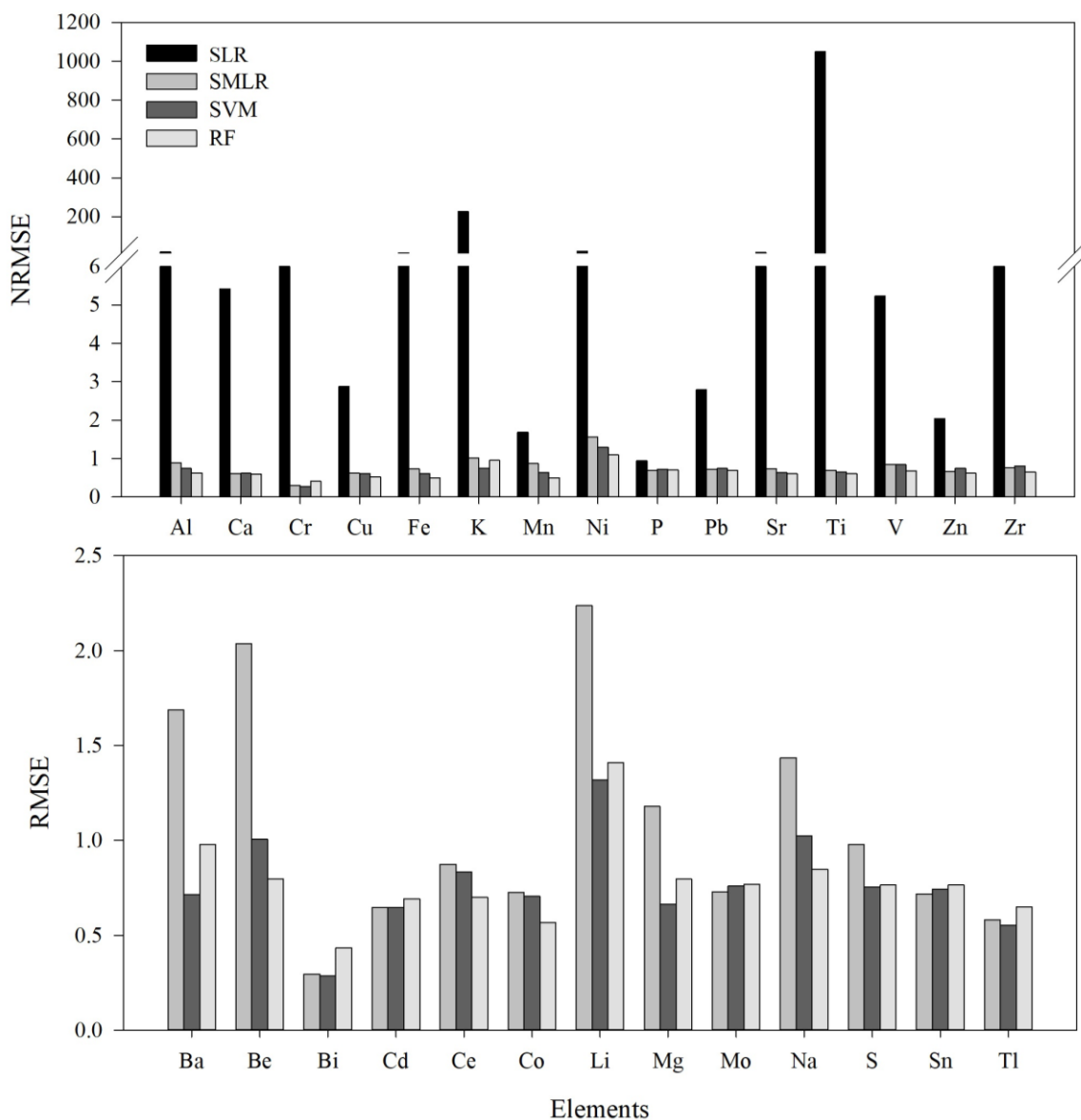

Fig. 7. Normalized root mean square error (NRMSE) corresponding to the validation between measured/observed values in the laboratory (USEPA 3051a) *versus* values predicted by simple linear regression (SLR), Stepwise Multiple Linear Regression (SMLR), Support Vector Machine (SVM) with linear kernel and Random Forest (RF) using datasets obtained by pXRF in Brazilian tropical soils.

Table 3. Summary of the final interpretation of the best models generated using pXRF data for the prediction of 28 elements commonly acquired via USEPA 3051a in Brazilian tropical soils.

Elements	Best Model	Parameters			
		R²	RPD	RMSE (mg kg⁻¹)	NRMSE
Al	RF	0.72	1.58	5646.74	0.63
Ca	RF	0.67	1.70	498.82	0.59
Cr	SVM	0.94	3.62	50.50	0.28
Cu	RF	0.74	1.89	11.58	0.53
Fe	RF	0.79	1.98	8132.65	0.51
K	RF	0.59	1.05	179.94	0.95
Mn	RF	0.78	2.00	116.60	0.50
Ni	RF	0.50	0.91	15.13	1.10
P	SMLR	0.53	1.43	139.92	0.70
Pb	RF	0.52	1.43	8.37	0.70
Sr	RF	0.70	1.64	3.86	0.61
Ti	RF	0.70	1.66	174.19	0.60
V	RF	0.58	1.45	40.82	0.69
Zn	RF	0.61	1.60	10.35	0.63
Zr	RF	0.78	1.53	18.25	0.65
Ba	SVM	0.63	1.40	19.41	0.72
Be	RF	0.60	1.26	0.14	0.80
Bi	SVM	0.94	3.49	51.73	0.29
Cd	SMLR/SVM	0.70	1.55	1.55	0.65
Ce	RF	0.51	1.43	13.88	0.70
Co	RF	0.68	1.76	4.56	0.57
Li	RF	0.31	0.71	3.80	1.41
Mg	SVM	0.57	1.50	151.88	0.67
Mo	SMLR	0.59	1.37	5.64	0.73
Na	RF	0.35	1.14	86.12	0.85
S	SVM	0.42	1.32	67.53	0.76
Sn	SMLR	0.62	1.40	4.32	0.72
Tl	SVM	0.74	1.80	18.39	0.55

RF - Random Forest; SMLR - Stepwise Multiple Linear Regression; SVM - Support Vector Machine with linear kernel; R² - coefficient of determination; RPD – residual prediction deviation; RMSE - root mean square error; NRMSE - normalized root mean square error.

Therefore, the results delivered by this study show that when dealing with a very heterogeneous set of samples, arising from large areas, in which the soil composition can vary considerably, the precision of linear regression techniques decreases due to the non-linear

nature of the relationship between the data obtained by pXRF and the dependent variable (USEPA 3051a results). Thus, to obtain robustness and reliability in the prediction, an appropriate nonlinear model, mainly RF, is an essential factor for a successful elementary prediction. Thus, the results of this study indicate an advance for soils of tropical region in relation to what has been reported in the literature for the indirect quantification of elemental contents in such soils, contributing to a fast and environmentally friendly geochemical analysis of these weathered and leached soils.

The present study is the first attempt to use pXRF-only data linked to the use of simple and multiple linear regression, as well as machine learning algorithms in the generation and validation of models capable of predicting the elementary content of a large number of elements in highly heterogeneous tropical soils based on the USEPA 3051a method. The few existing studies approach the prediction of a small number of elements through the use of simple linear regression only, use a relatively smaller dataset for this purpose and with more similar soils than in this study (e.g., Borges et al., 2020b; Faria et al., 2022; Silva et al., 2019).

3.5. *Important variables for Random Forest prediction models*

Assessing all models created via Random Forest (RF) using the dataset via pXRF (n = 179), it is possible to identify the variables that most contributed to the predictions of the 28 elements of the studied soils, only using the percentage of increment in the mean square error (% IncMSE) (Fig. 8). In general, when analyzing the importance of the variables up to the fifth position, we can observe that 23 out of the 28 predicted elements had Zn among the most important variables, followed by Fe, Cu and V, which were also highly relevant for the predictions of the contents of 19, 17 and 15 elements, respectively. Zinc, Fe, Cu and V delivered by pXRF were highly and positively correlated with the vast majority of the 28 elements determined by the USEPA 3051a standard method (Fig. 3), which also explains the strong influence of these elements on elementary prediction through the RF algorithm.
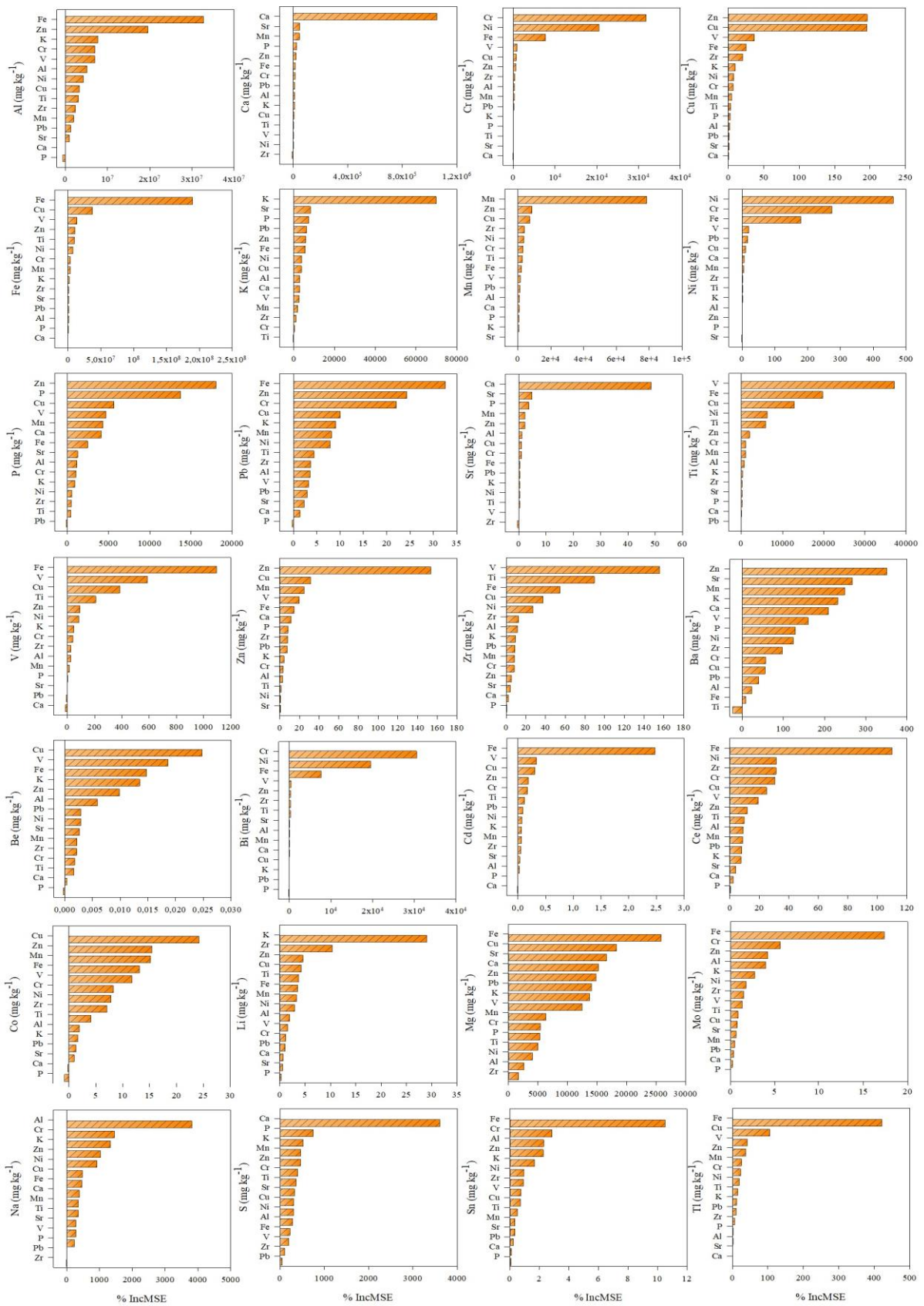
Fig. 8. Importance of variables in decreasing order expressed by percentage increase in mean square error (% IncMSE) for models built using the Random Forest (RF) algorithm trained from data generated by pXRF in Brazilian tropical soils.

Analyzing the first 15 elements obtained via USEPA 3051a, it can be observed that their correspondents delivered by pXRF are presented as the first or second most important variable (Fig. 8). The correlation results between the elements common to both methods already showed the trend of being satisfactorily for the vast majority of elements, as discussed previously. In contrast, the 13 elements shown in Fig. 8 do not have their counterparts via pXRF in this study, since they were mostly below the detection limits of pXRF, but the validation data of the predictions were satisfactory for seven of these elements (e.g., Ba, Bi, Cd, Ce, Co, Mg and Tl) using machine learning algorithms (SVM or RF). Therefore, the use of algorithms combined with available pXRF data becomes efficient in the indirect prediction of elements of interest. This result further demonstrates that elements easily accessed with pXRF can be well correlated with the occurrence of other elements in tropical, heterogeneous, and leached soils.

4. Conclusion

The pXRF data proved to be of great importance for the characterization and geochemical monitoring of tropical soils in a fast, accurate and environmentally friendly way. Strong positive correlations (ranging from 0.50 to 0.98) between the elementary content provided by pXRF and the USEPA 3051a method were obtained, allowing the development of accurate prediction models.

The use of the RF algorithm was superior to others models for providing accurate and reliable prediction models for 13 elements (e.g., Al, Ca, Cu, Fe, Mn , Pb, Sr, Ti, V, Zn, Zr, Ce and Co),. However, models for Cr and Bi via SVM delivered the best results in terms of precision and accuracy among all. Unsatisfactory results ($R^2 < 0.50$; RPD < 1.4) were generated for K, Ni, Be, Li, Mo, Na and S, regardless of the models used.

For the RF algorithm, the predictor variables Zn, Fe, Cu and V via pXRF were highly relevant for the prediction of the vast majority of the 28 elements for being positively correlated with the results of the USEPA 3051a method. Therefore, the results of this study clearly show that it is possible to develop prediction models for chemical element contents resulting from the USEPA 3051a method, using only the pXRF sensor and machine learning

algorithms, even using a large and heterogeneous dataset (wide range of soils, with different classes, parent materials, sampling depths, land uses, management practices and sampling sites).

Acknowledgments

References

Abbruzzini, T.F., Silva, C.A., de Andrade, D.A., Carneiro, W.J. de O., 2014. Influência de métodos de digestão na recuperação de teores de ferro, zinco, níquel, cromo, cádmio e chumbo em 11 resíduos orgânicos. Rev. Bras. Cienc. do Solo 38, 166–176. https://doi.org/10.1590/S0100-06832014000100016

Alvares, C.A., Stape, J.L., Sentelhas, P.C., De Moraes Gonçalves, J.L., Sparovek, G., 2013. Köppen's climate classification map for Brazil. Meteorol. Zeitschrift 22, 711–728. https://doi.org/10.1127/0941-2948/2013/0507

Andrade, R., Faria, W.M., Silva, S.H.G., Chakraborty, S., Weindorf, D.C., Mesquita, L.F., Guilherme, L.R.G., Curi, N., 2020. Prediction of soil fertility via portable X-ray fluorescence (pXRF) spectrometry and soil texture in the Brazilian Coastal Plains. Geoderma 357, 113960. https://doi.org/10.1016/j.geoderma.2019.113960

Andrade, R., Silva, S.H.G., Weindorf, D.C., Chakraborty, S., Faria, W.M., Guilherme, L.R.G., Curi, N., 2021. Micronutrients prediction via pXRF spectrometry in Brazil: Influence of weathering degree. Geoderma Reg. 27, e00431. https://doi.org/10.1016/j.geodrs.2021.e00431

Anjos, L.H.C., Jacomine, P.T.K., Santos, H.G., Oliveira, V.A., Oliveira, J.B., 2012. Sistema brasileiro de classificação de solos, in: In: Ker, J.C., Curi, N., Schaefer, C.E.G.R., Vidal-Torrado, P. (Eds.), Pedologia - Fundamentos. SBCS, Viçosa. pp. 303–343.

Araújo, S.R., Wetterlind, J., Demattê, J.A.M., Stenberg, B., 2014. Improving the prediction performance of a large tropical vis-NIR spectroscopic soil library from Brazil by clustering into smaller subsets or use of data mining calibration techniques. Eur. J. Soil

Sci. 65, 718–729.

Benedet, L., Acuña-Guzman, S.F., Faria, W.M., Silva, S.H.G., Mancini, M., Teixeira, A.F. dos S., Pierangeli, L.M.P., Acerbi Júnior, F.W., Gomide, L.R., Pádua Júnior, A.L., Souza, I.A. de, de 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. https://doi.org/10.1016/j.catena.2020.105003

Benedet, L., Faria, W.M., Silva, S.H.G., Mancini, M., Guilherme, L.R.G., Demattê, J.A.M., Curi, N., 2020. Soil subgroup prediction via portable X-ray fluorescence and visible near-infrared spectroscopy. Geoderma 365, 114212. https://doi.org/10.1016/j.geoderma.2020.114212

Bispo, F.H.A., de Menezes, M.D., Fontana, A., Sarkis, J.E. de S., Gonçalves, C.M., de Carvalho, T.S., Curi, N., Guilherme, L.R.G., 2021. Rare earth elements (REEs): geochemical patterns and contamination aspects in Brazilian benchmark soils. Environ. Pollut. 289, 117972. https://doi.org/10.1016/j.envpol.2021.117972

Bocardi, J.M.B., Pletsch, A.L., Melo, V.F., Quinaia, S.P., 2020. Quality reference values for heavy metals in soils developed from basic rocks under tropical conditions. J. Geochemical Explor. 217, 106591. https://doi.org/10.1016/j.gexplo.2020.106591

Borges, C.S., Weindorf, D.C., Carvalho, G.S., Guilherme, L.R.G., Takayama, T., Curi, N., Lima, G.J.E.O., Ribeiro, B.T., 2020a. Foliar Elemental Analysis of Brazilian Crops via Portable X-ray Fluorescence Spectrometry. Sensor 20, 2509.

Borges, C.S., Weindorf, D.C., Nascimento, D.C., Curi, N., Guilherme, L.R.G., Carvalho, G.S., Ribeiro, B.T., 2020b. Comparison of portable X-ray fluorescence spectrometry and laboratory-based methods to assess the soil elemental composition: Applications for wetland soils. Environ. Technol. Innov. 19, 100826. https://doi.org/10.1016/j.eti.2020.100826

Brinatti, A.M., Mascarenhas, Y.P., Pereira, V.P., Partiti, C.S. de M., Macedo, Á., 2010. Mineralogical characterization of a highly-weathered soil by the rietveld method. Sci. Agric. 67, 454–464. https://doi.org/10.1590/s0103-90162010000400013

Carvalho Filho, A. de, Curi, N., Marques, J.J.G. de S. e M., Shinzato, E., Freitas, D.A.F. de, Jesus, E.A. de, Massahud, R.T.L.R., 2011. Óxidos de manganês em solos do quadrilátero ferrífero (MG). Rev. Bras. Ciência do Solo 35, 793–804. https://doi.org/10.1590/s0100-06832011000300015

CETESB, 2014. Companhia de Tecnologia de Saneamento Ambiental. Decisão de Diretoria 045/2014/E/C/I, de 20-02-2014. Dispõe sobre a aprovação dos Valores Orientadores para

Solos e Águas Subterrâneas no Estado de São Paulo - 2014, em substituição aos Valores Orientadores .

Chagas, C. da S., Carvalho Junior, W., Bhering, S.B., Calderano Filho, B., 2016. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. Catena 139, 232–240. https://doi.org/10.1016/j.catena.2016.01.001

Chai, T., Draxler, R.R., 2014. Root mean square error (RMSE) or mean absolute error (MAE)? -Arguments against avoiding RMSE in the literature. Geosci. Model Dev. 7, 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Chang, C., Laird, D.A., Mausbach, M.J., Hurburgh, C.R.J., 2001. Near-infrared reflectance spectroscopy-principal components regression analyses of soil properties. Soil Sci. Soc. Am. J. 65, 480–490.

Chen, M., Ma, L.Q., 1998. Digestion Methods for Trace Metal Analysis. J. Environ. Qual. 27, 1294–1300.

CONAMA, 2009. Conselho Nacional do Meio Ambiente. Resolução Conama n∘ 420, de 28 de dezembro de 2009: Dispõe sobre critérios e valores orientadores de qualidade do solo quanto à presença de substâncias químicas e estabelece diretrizes para o gerenciamento ambiental de .

COPAM, 2011. Conselho Estadual de Política Ambiental, Deliberação Normativa Copam n∘ 166, de 29 de junho de 2011. Altera o Anexo I da Deliberação Normativa Conjunta Copam CERH no 2 de 6 de setembro de 2010, estabelecendo os Valores de Referência de Qualidade dos Solos.

Coringa, E. de A.O., Couto, E.G., Torrado, P.V., 2014. Geoquímica de solos do pantanal norte, Mato Grosso. Rev. Bras. Ciência do Solo 38, 1784–1793. https://doi.org/10.1590/s0100-06832014000600013

Costa, Y.T., Ribeiro, B.T., Curi, N., de Oliveira, G.C., Guilherme, L.R.G., 2019. Organic Matter Removal on Oxide Determination in Oxisols Via Portable X-ray Fluorescence. Commun. Soil Sci. Plant Anal. 50, 673–681. https://doi.org/10.1080/00103624.2019.1589479

Declercq, Y., Delbecque, N., De Grave, J., De Smedt, P., Finke, P., Mouazen, A.M., Nawar, S., Vandenberghe, D., Van Meirvenne, M., Verdoodt, A., 2019. A comprehensive study of three different portable XRF scanners to assess the soil geochemistry of an extensive sample dataset. Remote Sens. 11. https://doi.org/10.3390/rs11212490

Dharumarajan, S., Hegde, R., Singh, S.K., 2017. Spatial prediction of major soil properties using Random Forest techniques - A case study in semi-arid tropics of South India.

Geoderma Reg. 10, 154–162. https://doi.org/10.1016/j.geodrs.2017.07.005

Drouet, T., Herbauts, J., Gruber, W., Demaiffe, D., 2007. Natural strontium isotope composition as a tracer of weathering patterns and of exchangeable calcium sources in acid leached soils developed on loess of central Belgium. Eur. J. Soil Sci. 58, 302–319.

Faria, A.J.G., Silva, S.H.G., Andrade, R., Mancini, M., Melo, L.C.A., Weindorf, D.C., Guilherme, L.R.G., Curi, N., 2022. Prediction of soil organic matter content by combining data from Nix Pro TM color sensor and portable X-ray fluorescence spectrometry in tropical soils. Geoderma Reg. 28, e00461. https://doi.org/10.1016/j.geodrs.2021.e00461

Faria, Á.J.G., Silva, S.H.G., Melo, L.C.A., Botelho, L., Guilherme, L.R.G., Curi, N., 2022. Relationship between elemental content determined via portable X-ray fluorescence and traditional acid_digestion-based methods in tropical soils. Soil Res. Press. https://doi.org/https://doi.org/10.1071/SR21272

Faria, A.J.G. de, Rufini, M., Leite, A. do A., Ribeiro, B.T., Silva, S.H.G., Guilherme, L.R.G., Melo, L.C.A., 2021. Elemental analysis of biochar-based fertilizers via portable X-ray fluorescence spectrometry. Environ. Technol. Innov. 23, 101788. https://doi.org/10.1016/j.eti.2021.101788

Faria, Á.J.G. De, Silva, S.H.G., Melo, L.C.A., Andrade, R., Mancini, M., Mesquita, L.F., Teixeira, A.F.D.S., Guilherme, L.R.G., Curi, N., 2020. Soils of the Brazilian Coastal Plains biome: Prediction of chemical attributes via portable X-ray fluorescence (pXRF) spectrometry and robust prediction models. Soil Res. 58, 683–695. https://doi.org/10.1071/SR20136

Ferreira, G.W.D., Ribeiro, B.T., Weindorf, D.C., Teixeira, B.I., Chakraborty, S., Li, B., Guilherme, L.R.G., Scolforo, J.R.S., 2021. Assessment of iron-rich tailings via portable X-ray fluorescence spectrometry: the Mariana dam disaster, southeast Brazil. Environ. Monit. Assess. 193, 1–19. https://doi.org/10.1007/s10661-021-08982-7

Forkuor, G., Hounkpatin, O.K.L., Welp, G., Thiel, M., 2017. High resolution mapping of soil properties using Remote Sensing variables in south-western Burkina Faso: A comparison of machine learning and multiple linear regression models. PLoS One 12. https://doi.org/10.1371/journal.pone.0170478

González, S., Herrera, F., García, S., 2015. Monotonic random forest with an ensemble pruning mechanism based on the degree of monotonicity. New Gener. Comput. 33, 367–388.

Gozukara, G., Zhang, Y., Hartemink, A.E., 2021. Using vis-NIR and pXRF data to

distinguish soil parent materials – An example using 136 pedons from Wisconsin, USA. Geoderma 396, 115091. https://doi.org/10.1016/j.geoderma.2021.115091

Hornik, K., Weingessel, A., Leisch, F., Davidmeyerr-Projectorg, M.D.M., 2015. Package 'e1071''.'

Hou, H., Takamatsu, T., Koshikawa, M.K., Hosomi, M., 2005. Migration of silver, indium, tin, antimony, and bismuth and variations in their chemical fractions on addition to uncontaminated soils. Soil Sci. 170, 624–639. https://doi.org/10.1097/01.ss.0000178205.35923.66

Hseu, Z.-Y., Chen, Z.-S., Tsai, C.-C., Jien, S.-H., 2016. Portable X-Ray Fluorescence (pXRF) for Determining Cr and Ni Contents of Serpentine Soils in the Field, in: In: Hartemink A., Minasny B. (Eds) Digital Soil Morphometrics. Progress in Soil Science. Springer, Cham. pp. 37–50. https://doi.org/10.1007/978-3-319-28295-4_3

Jalali, S.A.S., Navidi, M.N., Mohammadi, J.S., Meymand, A.Z., Mohammad Esmail, Z., 2019. Prediction of Soil Cation Exchange Capacity Using Different Soil Parameters by Intelligent Models. Commun. Soil Sci. Plant Anal. 50, 2123–2139. https://doi.org/10.1080/00103624.2019.1654501

Javadi, S.H., Mouazen, A.M., 2021. Data fusion of xrf and vis-nir using outer product analysis, granger–ramanathan, and least squares for prediction of key soil attributes. Remote Sens. 13. https://doi.org/10.3390/rs13112023

Kabata-Pendias, A., 2010. Trace elements in soils and plants. CRC Press, Fourth Ed. ed. https://doi.org/10.1201/b10158

Kabata-Pendias, A., Mukherjee, A.B., 2007. Trace elements from soil to human. Springer-Verlag. Berlin.

Kämpf, N., Marques, J.J., Curi, N., 2012. Mineralogia de Solos Brasileiros, in: Ker, J.C. Curi, N. Schaefer, C.E.G.R. Vidal-Torrado, P. Pedologia Fundamentos. SBCS, Viçosa, MG. p. 343.

Knight, R.D., Kjarsgaard, B.A., Russell, H.A.J., 2021. An analytical protocol for determining the elemental chemistry of Quaternary sediments using a portable X-ray fluorescence spectrometer. Appl. Geochemistry 131, 105026. https://doi.org/10.1016/j.apgeochem.2021.105026

Liaw, A., Wiener, M., 2015. Package "randomForest". R Dev. Core Teamhttps://doi.org/10.5244/C.22.54.

Liaw, A., Wiener, M., 2002. Classification and regression by random forest. R News 2, 18–22.

Lima, T.M. de, Weindorf, D.C., Curi, N., Guilherme, L.R.G., Lana, R.M.Q., Ribeiro, B.T., 2019. Elemental analysis of Cerrado agricultural soils via portable X-ray fluorescence spectrometry: Inferences for soil fertility assessment. Geoderma 353, 264–272. https://doi.org/10.1016/j.geoderma.2019.06.045

Liu, Y., Wang, C., Xiao, C., Shang, K., Zhang, Y., Pan, X., 2021. Prediction of multiple soil fertility parameters using VisNIR spectroscopy and PXRF spectrometry. Soil Sci. Soc. Am. J. 85, 591–605. https://doi.org/10.1002/saj2.20223

Lopes, A.., Guilherme, L.R.., 2016. A career perspective on soil management in the Cerrado region of Brazil. Adv. Agron. 137, 1–72.

Lucà, F., Conforti, M., Castrignanò, A., Matteucci, G., Buttafuoco, G., 2017. Effect of calibration set size on prediction at local scale of soil carbon by Vis-NIR spectroscopy. Geoderma 288, 175–183.

Mancini, M., Weindorf, D.C., Chakraborty, S., Silva, S.H.G., dos Santos Teixeira, A.F., Guilherme, L.R.G., Curi, N., 2019a. Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. Geoderma 337, 718–728. https://doi.org/10.1016/j.geoderma.2018.10.026

Mancini, M., Weindorf, D.C., Monteiro, M.E.C., de Faria, Á.J.G., dos Santos Teixeira, A.F., de Lima, W., de Lima, F.R.D., Dijair, T.S.B., Marques, F.D.A., Ribeiro, D., Silva, S.H.G., Chakraborty, S., Curi, N., 2020. From sensor data to Munsell color system: Machine learning algorithm applied to tropical soil color classification via Nix^{TM} Pro sensor. Geoderma 375, 114471. https://doi.org/10.1016/j.geoderma.2020.114471

Mancini, M., Weindorf, D.C., Silva, S.H.G., Chakraborty, S., Teixeira, A.F. dos S., Guilherme, L.R.G., Curi, N., 2019b. Parent material distribution mapping from tropical soils data via machine learning and portable X-ray fluorescence (pXRF) spectrometry in Brazil. Geoderma 354, 113885. https://doi.org/10.1016/j.geoderma.2019.113885

Marques, J.J., Schulze, D.G., Curi, N., Mertzman, S.A., 2004a. Trace element geochemistry in Brazilian Cerrado soils. Geoderma 121, 31–43. https://doi.org/10.1016/j.geoderma.2003.10.003

Marques, J.J., Schulze, D.G., Curi, N., Mertzman, S.A., 2004b. Major element geochemistry and geomorphic relationships in Brazilian Cerrado soils. Geoderma 119, 179–195. https://doi.org/10.1016/S0016-7061(03)00260-X

Meier, M., Souza, E., Francelino, M.R., Fernandes Filho, E.I., Schaefer, C.E.G.R., 2018. Digital soil mapping using machine learning algorithms in a tropical mountainous area. Rev. Bras. Cienc. do Solo 42, e0170421. https://doi.org/10.1590/18069657rbcs20170421

Melo, V.F., Novais, R.F., Schaefer, C.E.G.R., Fontes, M.P.F., Singh, B., 2002a. Mineralogia das frações areia, silte e argila de sedimentos do grupo barreiras no município de Aracruz, estado do Espírito Santo. Rev. Bras. Ciência do Solo 26, 29–41. https://doi.org/10.1590/s0100-06832002000100004

Melo, V.F., Schaefer, C.E.G.R., Singh, B., Novais, R.F., Fontes, M.P.F., 2002b. Propriedades químicas e cristalográficas da caulinita e dos óxidos de ferro em sedimentos do grupo barreiras no município de Aracruz, estado do Espírito Santo. Rev. Bras. Ciência do Solo 26, 53–64.

Meurer, E.J., Rheinheimer, R.D., Bissani, C.A., 2010. Fenômeno de Sorção em Solos. In: Fundamentos de Química do Solo. 4ed. MEURER, E. J. (ed). Porto Alege: Evangraf.

Murata, T., 2010. Bismuth solubility through binding by various organic compounds and naturally occurring soil organic matter. J. Environ. Sci. Heal. - Part A Toxic/Hazardous Subst. Environ. Eng. 45, 746–753. https://doi.org/10.1080/10934521003651465

Nascimento, E.S., Tenuta Filho, A., 2010. Chemical waste risk reduction and environmental impact generated by laboratory activities in research and teaching institutions. Brazilian J. Pharm. Sci. 46, 187–198.

Nawar, S., Delbecque, N., Declercq, Y., De Smedt, P., Finke, P., Verdoodt, A., Van Meirvenne, M., Mouazen, A.M., 2019. Can spectral analyses improve measurement of key soil fertility parameters with X-ray fluorescence spectrometry? Geoderma 350, 29–39. https://doi.org/10.1016/j.geoderma.2019.05.002

O'Rourke, S.M., Stockmann, U., Holden, N.M., McBratney, A.B., Minasny, B., 2016. An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. Geoderma 279, 31–44. https://doi.org/10.1016/j.geoderma.2016.05.005

Pearson, D., Chakraborty, S., Duda, B., Li, B., Weindorf, D.C., Deb, S., Brevik, E., Ray, D.P., 2016. Water analysis via portable X-ray fluorescence spectrometry. J. Hydrol. 544, 172–179. https://doi.org/10.1016/j.jhydrol.2016.11.018

Pelegrino, M.H.P., Silva, S.H.G., de Faria, Á.J.G., Mancini, M., Teixeira, A.F. dos S., Chakraborty, S., Weindorf, D.C., Guilherme, L.R.G., Curi, N., 2021. Prediction of soil nutrient content via pXRF spectrometry and its spatial variation in a highly variable tropical area. Precis. Agric. https://doi.org/10.1007/s11119-021-09825-8

Qu, M., Chen, J., Li, W., Zhang, C., Wan, M., Huang, B., Zhao, Y., 2019. Correction of in-situ portable X-ray fluorescence (PXRF) data of soil heavy metal for enhancing spatial prediction. Environ. Pollut. 254, 112993. https://doi.org/10.1016/j.envpol.2019.112993

R Development Core Team, 2019. R: A language and environment for statistical computing. Vienna, Austria. R Foundation for Statistical Computing, 2019. Available in: https://www.r-project.org/. Access in: Jul. 17, 2020. [WWW Document].

Ravansari, R., Wilson, S.C., Tighe, M., 2020. Portable X-ray fluorescence for environmental assessment of soils: Not just a point and shoot method. Environ. Int. 134, 105250. https://doi.org/10.1016/j.envint.2019.105250

Ravansari, R., Wilson, S.C., Wilson, B.R., Tighe, M., 2021. Rapid PXRF soil organic carbon and organic matter assessment using novel modular radiation detector assembly. Geoderma 382, 114728. https://doi.org/10.1016/j.geoderma.2020.114728

Ribeiro, B.T., Nascimento, D.C., Curi, N., Guilherme, L.R.G., Costa, E.T. de S., Lopes, G., Carneiro, J.P., 2019. Assessment of trace element contents in soils and water from cerrado wetlands, triângulo mineiro region. Rev. Bras. Cienc. do Solo 43. https://doi.org/10.1590/18069657rbcs20180059

Rosolen, V., De-Campos, A.B., Govone, J.S., Rocha, C., 2015. Contamination of wetland soils and floodplain sediments from agricultural activities in the Cerrado Biome (State of Minas Gerais, Brazil). Catena 128, 203–210. https://doi.org/10.1016/j.catena.2015.02.007

Santana, M.L.T., Ribeiro, B.T., Silva, S.H.G., Poggere, G.C., Guilherme, L.R.G., Curi, N., 2018. Conditions affecting oxide quantification in unknown tropical soils via handheld X-ray fluorescence spectrometer. Soil Res. 56, 648. https://doi.org/10.1071/sr18099

Santos, H.., Jacomine, P.K.., Anjos, L.H.., Oliveira, V.., Lumbreras, J.., Coelho, M.., Almeida, J.., Araujo Filho, J.., Oliveira, J.., Cunha, T.J.F., 2018. Sistema Brasileiro de Classificação de Solos, 5th ed. ed. Embrapa, Brasília.

Santos, R.D., Santos, H.G., Ker, J.C., Anjos, L.H.C., Shimizu, S.H., 2015. Manual de descrição e coleta de solos no campo, 7ª revisad. ed. Viçosa, MG.

Silva, E.A., Weindorf, D.C., Silva, S.H.G., Ribeiro, B.T., Poggere, G.C., Carvalho, T.S., Gonçalves, M.G.M., Guilherme, L.R.G., Curi, N., 2019. Advances in Tropical Soil Characterization via Portable X-Ray Fluorescence Spectrometry. Pedosphere 29, 468–482. https://doi.org/10.1016/s1002-0160(19)60815-5

Silva, S.H.G., Ribeiro, B.T., Guerra, M.B.B., de Carvalho, H.W.P., Lopes, G., Carvalho, G.S., Guilherme, L.R.G., Resende, M., Mancini, M., Curi, N., Rafael, R.B.A., Cardelli, V., Cocco, S., Corti, G., Chakraborty, S., Li, B., Weindorf, D.C., 2021. pXRF in tropical soils: Methodology, applications, achievements and challenges. Adv. Agron. 167, 1–62. https://doi.org/10.1016/bs.agron.2020.12.001

Silva, S.H.G., Silva, E.A., Poggere, G.C., Junior, A.L.P., Gonçalves, M.G.M., Guilherme, L.R.G., Curi, N., 2020. Modeling and prediction of sulfuric acid digestion analyses data from pxrf spectrometry. Sci. Agric. 77. https://doi.org/10.1590/1678-992x-2018-0132

Silva, Y.J.A.B., Nascimento, C.W.A., Biondi, C.M., 2014. Comparison of USEPA digestion methods to heavy metals in soil samples. Environ. Monit. Assess. 186, 47–53. https://doi.org/10.1007/s10661-013-3354-5

Soil Survey Staff, 2014. Soil survey field and laboratory methods manual. Version 2. USDA-NRCS. Available online at: https://www.nrcs.usda.gov/Internet/FSE_ DOCUMENTS/stelprdb1244466.pdf (verified 15 Feb. 2021).

Soltanpour, P.N., Johnson, G.W., Workman, S.M., Jones, B.J., Miller, R.O., 1996. Inductively coupled plasma emission spec- trometry and inductively coupled plasma-mass spectrometry, in: Methods of Soil Analysis. Part 3. Chemical Methods. pp. 91–139. https://doi.org/10.2136/ sssabookser5.3.c5

Souza, A.B., Demattê, J.A.M., Bellinaso, H., Mello, D.C. de, da Silva Lisboa, C.J., de Oliveira Mello, F.A., Marques, K.P.P., de Resende, M.E.B., Reis, J.V., Mancini, M., Silva, S.H.G., Curi, N., 2021. A sensors-based profile heterogeneity index for soil characterization. Catena 207, 105670. https://doi.org/10.1016/j.catena.2021.105670

Souza, J.J.L.L., Abrahão, W.A.P., Mello, J.W.V., Silva, J., Costa, L.M., Oliveira, T.S., 2015. Geochemistry and spatial variability of metal(loid) concentrations in soils of the state of Minas Gerais, Brazil. Sci. Total Environ. 505, 338–349. https://doi.org/10.1016/j.scitotenv.2014.09.098

Stockmann, U., Cattle, S.R., Minasny, B., McBratney, A.B., 2016. Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. Catena 139, 220–231. https://doi.org/10.1016/j.catena.2016.01.007

Taebi, A., Mansy, H.A., 2017. Time-frequency distribution of seismocardiographic signals: A comparative study. Bioengineering 4, 1–21. https://doi.org/10.3390/bioengineering4020032

Tavares, T.R., Molin, J.P., Hamed Javadi, S., de Carvalho, H.W.P., Mouazen, A.M., 2021. Combined use of vis-nir and xrf sensors for tropical soil fertility analysis: Assessing different data fusion approaches. Sensors 21, 1–23. https://doi.org/10.3390/s21010148

Teixeira, A.F. dos S., Pelegrino, M.H.P., Faria, W.M., Silva, S.H.G., Gonçalves, M.G.M., Acerbi Júnior, F.W., Gomide, L.R., Pádua Júnior, A.L., Souza, I.A. de, Chakrabortyd, S., Weindorfe, D.C., Guilherme, L.R.G., Curi, N., 2020. Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry. Geoderma 361,

114132. https://doi.org/10.1016/j.geoderma.2019.114132

USEPA, 2007a. Method 6200: Field portable X-ray fluorescence spectrometry for the determination of elemental concentrations in soil and sediment. US EPA. https://www.epa.gov/sites/production/files/2015-12/documents/6200.pdf. (Accessed 10 Dez. 2019).

USEPA, 2007b. Method 3051a (SW-846): Microwave assisted acid digestion of sediments, sludges, soils, and oils. US EPA. https://www.epa.gov/sites/ production/files/2015-12/documents/3051a.pdf. (Acessed 27 jan. 2021).

Vasques, G.M., Rodrigues, H.M., Coelho, M.R., Baca, J.F.M., Dart, R.O., Oliveira, R.P., Teixeira, W.G., Ceddia, M.B., 2020. Field proximal soil sensor fusion for improving high-resolution soil property maps. Soil Syst. 4, 1–22. https://doi.org/10.3390/soilsystems4030052

Viscarra Rossel, R.A., Bui, E.N., De Caritat, P., McKenzie, N.J., 2010. Mapping iron oxides and the color of Australian soil using visible-near-infrared reflectance spectra. J. Geophys. Res. Earth Surf. 115, 1–13. https://doi.org/10.1029/2009JF001645

Wan, M., Hu, W., Qu, M., Li, W., Zhang, C., Kang, J., Hong, Y., Chen, Y., Huang, B., 2020. Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. Geoderma 363, 114163. https://doi.org/10.1016/j.geoderma.2019.114163

Wan, M., Qu, M., Hu, W., Li, W., Zhang, C., Cheng, H., Huang, B., 2019. Estimation of soil pH using PXRF spectrometry and Vis-NIR spectroscopy for rapid environmental risk assessment of soil heavy metals. Process Saf. Environ. Prot. https://doi.org/10.1016/j.psep.2019.09.025

Wang, D., Chakraborty, S., Weindorf, D.C., Li, B., Sharma, A., Paul, S., Ali, M.N., 2015. Synthesized use of VisNIR DRS and PXRF for soil characterization: Total carbon and total nitrogen. Geoderma 243–244, 157–167. https://doi.org/10.1016/j.geoderma.2014.12.011

Weindorf, D.C., Bakr, N., Zhu, Y., 2014. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications, Advances in Agronomy. Elsevier. https://doi.org/10.1016/B978-0-12-802139-2.00001-9

Weindorf, D.C., Chakraborty, S., 2020. Portable X-ray fluorescence spectrometry analysis of soils. Soil Sci. Soc. Am. J. 84, 1384–1392. https://doi.org/10.1002/saj2.20151

Weindorf, D.C., Chakraborty, S., Li, B., Deb, S., Singh, A., Kusi, N.Y., 2018. Compost salinity assessment via portable X-ray fluorescence (PXRF) spectrometry. Waste Manag.

78, 158–163. https://doi.org/10.1016/j.wasman.2018.05.044

Wu, W., Li, A. Di, He, X.H., Ma, R., Liu, H. Bin, Lv, J.K., 2018. A comparison of support vector machines, artificial neural network and classification tree for identifying soil texture classes in southwest China. Comput. Electron. Agric. 144, 86–93. https://doi.org/10.1016/j.compag.2017.11.037

Yang, X., Post, W.M., 2011. Phosphorus transformations as a function of pedogenesis: a synthesis of soil phosphorus data using Hedley fractionation method. Biogeosciences 8, 2907–2916