MARCELO HENRIQUE PROCÓPIO PELEGRINO

PREDICTION OF SOIL PROPERTIES VIA PORTABLE X-RAY FLUORESCENCE (pXRF) SPECTROMETRY IN BRAZIL

LAVRAS – MG

2019

MARCELO HENRIQUE PROCÓPIO PELEGRINO

PREDICTION OF SOIL PROPERTIES VIA PORTABLE X-RAY FLUORESCENCE (pXRF) SPECTROMETRY IN BRAZIL

Dissertação 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 a obtenção do título de mestre.

Prof. Dr. Sérgio Henrique Godinho Silva

Orientador

LAVRAS–MG

2019

Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).

Pelegrino, Marcelo Henrique Procópio.
Prediction of soil properties via portable X-ray fluorescence (pXRF) spectrometry in Brazil / Marcelo Henrique Procópio Pelegrino. - 2019.
88 p. : il.

Orientador(a): Sérgio Henrique Godinho Silva.

Dissertação (mestrado acadêmico) - Universidade Federal de Lavras, 2019.
Bibliografia.

1. Predição de propriedades do solo. 2. Sensores próximos. 3. Aprendizagem de máquina. I. Silva, Sérgio Henrique Godinho. II. Título.

MARCELO HENRIQUE PROCOPIO PELEGRINO

PREDICTION OF SOIL PROPERTIES VIA PORTABLE X-RAY FLUORESCENCE (pXRF) SPECTROMETRY IN BRAZIL

PREDIÇÃO DE PROPRIEDADES DO SOLO VIA ESPECTROMETRIA PORTÁTIL DE FLUORESCÊNCIA DE RAIOS-X (pXRF) NO BRASIL

Dissertação 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 a obtenção do título de mestre.

APROVADA em 15 de Fevereiro de 2019.

Dr. Teotonio Soares de Carvalho UFLA

Dr. Walbert Júnior Reis dos Santos IFSULDEMINAS

Prof. Dr. Sérgio Henrique Godinho Silva

Orientador

LAVRAS–MG

2019

A Deus,

À minha mãe, Maria Dolores, ao meu pai, Laércio Mário

Aos meus irmãos, Ricardo, Lucas e José

Aos amigos e colegas que fiz ao longo do caminho.

Dedico.

AGRADECIMENTOS

Agradeço primeiramente aos meus pais Laércio Mário e Maria Dolores pelo apoio incondicional nas minhas decisões, pelos sacrifícios, pelas preces e por nunca duvidarem do seu filho.

Agradeço aos meus irmãos Ricardo, José e Lucas, que me inspiram a ser melhor e continuar melhorando.

Ao professor PhD Sérgio Henrique Godinho Silva o qual admiro, meu orientador, conselheiro e amigo, sem o qual não estaria aqui hoje.

Aos professores do Departamento de Ciência do Solo, por sempre serem solícitos, em especial ao professor Nilton Curi, pelos conselhos, coorientação e amizade.

Ao grupo de pedologia, alunos da pós-graduação e da graduação, Mariana, Fernanda, Luíza, Renata, Marcelo, Wilson e Thaís do qual tenho orgulho de ser integrado.

A todos os funcionários do Departamento de Ciência do Solo, pela excelência e compromisso.

Aos amigos da República Taj Mahal "minha casa"

Aos amigos do UFLA Rugby Team "minha casa"

Ao professor Fernando Oliveira e ao CRIA Lavras "minha casa"

A todos que direta ou indiretamente participaram da concretização deste momento.

A CAPES e ao CNPq pelo apoio à pesquisa.

O presente trabalho foi realizado com apoio da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG)

RESUMO GERAL

Os solos são o principal substrato para a produção de alimentos. A crescente demanda e a pressão ambiental impõem maiores produtividades, rentabilidade e mitigação dos impactos ambientais nas etapas e técnicas de produção. Dessa ótica, é nítida a importância de se conhecer as propriedades químicas, físicas e biológicas dos solos. Para a agricultura, conhecer melhor a fertilidade dos solos possibilita um uso mais racional dos recursos e insumos no planejamento das culturas. Entretanto, a aquisição desse conhecimento requer amostragens e análises de solo, que aumentam em número e volume na medida em que se refina o conhecimento, o que torna o processo caro. Além disso, a usual abordagem pontual de amostragem limita a compreensão da variabilidade espacial e temporal das propriedades dos solos. Neste sentido, a aquisição de dados de sensores remotos (e.g., imagens de satélite) e próximos (e.g., espectrômetros portáteis), tem refinado e complementado o conhecimento das propriedades dos solos através da modelagem computacional. Como fonte amplamente difundida em modelagens ambientais, têm-se os atributos de terreno (p.ex., índice topográfico de umidade) facilmente obtidos a partir de modelos digitais de elevação (DEM) em ambientes SIG. Em relação aos sensores próximos a espectrometria de fluorescência de raios-X portátil (pXRF) tem como vantagens a facilidade, rapidez e a não geração de resíduos em sua operação, além das vantagens de um equipamento portátil. A presente dissertação é dividida em dois capítulos, cujos objetivos são: modelagem e predição espacial dos teores disponíveis dos micronutrientes Fe, Mn, Cu e Zn para as plantas, através de dados de atributos de terreno (TA), dados do pXRF e informação do material de origem (PM), para os horizontes superficiais, subsuperficiais e ambos somados (n = 153), em diferentes combinações e resoluções espaciais, utilizando-se o algoritmo random forest (RF); e modelagem e predição espacial dos teores trocáveis/disponíveis dos macronutrientes P, Ca e K, através dos dados do sensor pXRF somente dos horizontes superficiais (n= 90), utilizando regressão linear simples (LR), regressão polinomial (PR), regressão de potência (PwR), regressão linear múltipla (SMLR) e random forest (RF). A área de estudos é localizada entre as longitudes 501031 e 504192 mE e as latitudes 7651139 e 7653537 mN, fuso 23K, no *campus* da Universidade Federal de Lavras, com aproximadamente 315 ha. Seus solos têm como materiais de origem gnaisse, gabro e sedimentos aluviais. O clima é Cwa segundo classificação climática de Koppen, com temperaturas médias anuais de 20,4°C e precipitação média anual de 1460 mm. A amostragem foi realizada em um grid regular de 200 m de distância entre locais de coleta. As amostras forma submetidas as análises laboratoriais para determinação dos respectivos nutrientes. Posteriormente, uma porção de cada amostra foi analisada com pXRF modelo S1 Titan LE (Bruker Nano Analytics, Kennewick, WA, USA) no modo *Trace* durante 60 s em triplicata. Os AT foram gerados com o software SAGA GIS a partir de DEM de 5 e 10 m de resolução. Os dados foram separados em conjuntos de treinamento (70%) e validação (30%), e os modelos foram gerados no software R (RF) e SigmaPlot (LR, PR, PwR e SMLR). Para efeito de análise e comparação entre os modelos foram utilizadas as métricas de coeficiente de determinação (R^2), R^2 ajustado (R^2_{adj}), raiz do erro quadrático médio (RMSE), raiz do erro quadrático médio normalizado (nRMSE) e erro médio (ME) para o primeiro capítulo, e R^2, RMSE, erro médio absoluto (MAE) e o desvio residual das predições (RPD) para o segundo. Após a determinação dos melhores modelos, a predição espacial para geração dos mapas de nutrientes disponíveis foi realizada. As variáveis do pXRF quando presentes no modelo, foram espacializadas para toda a área através de interpolação pelo inverso da distância

(IDW). Os TA de 10 m foram superiores aos de 5 m para predições. Foi possível obter bons resultados na predição espacial de Fe disponível utilizando somente TA de 10 m (R^2 = 0,88; RMSE = 59.97 mg kg^{-1} e ME = 24.00 mg kg^{-1}) e para os demais com dados pXRF + AT de 10 m + PM (0,85; 29,65 mg kg^{-1}; 9,70 mg kg^{-1} para Mn, 0,64; 3,11 mg kg^{-1}; 0,71 mg kg^{-1} para Zn e 0,82; 1,17 mg kg^{-1}; 0.43 mg kg^{-1} para Cu, respectivamente). Nas predições dos macronutrientes, a abordagem PwR obteve os melhores resultados (R^2 = 0,80 e RMSE = 1,63 $cmol_c$ dm^{-3} para o Ca^{2+} trocável, e 0,53 e 6,92 mg dm^{-3} para o P disponível). Não foi possível estabelecer correlação entre os teores de K^+ disponível e o conteúdo total de K_2O fornecido pelo pXRF. Dados de sensores próximos associados a dados de atributos de terreno podem realizar predições de teores trocáveis/disponíveis de nutrientes com elevada acurácia em solos tropicais.

Palavras-chave: Predição de propriedades do solo. Sensores próximos. Aprendizagem de máquina.

GENERAL ABSTRACT

Soils are the main substrate for food production. Increasing environmental demand and pressure imposes greater productivity, profitability, and mitigation of environmental impacts in the stages and production techniques. From this perspective, the importance of knowing the chemical, physical, and biological soil properties is evident. For agriculture, a better understanding of soil fertility enables a more rational use of resources and inputs in crop planning. However, the acquisition of this knowledge requires soil sampling and analyses, which increases in number and volume as knowledge is refined. This makes the process expensive.. Moreover, the usual point-sampling approach limits understanding of the spatial and temporal variability of soil properties. In this sense, the acquisition of data through remote (e.g., satellite images) and proximal (e.g., portable spectrometers) sensors has refined and complemented the knowledge about soil properties with the aid of computational modeling. As a widely diffused source in environmental modeling, one can easily obtain terrain attributes (e.g., topographic wetness index) from digital elevation models (DEM) in GIS environments. Regarding proximal sensors, the portable X-ray fluorescence (pXRF) spectrometry has the advantages of ease, speed, and non-generation of waste in its operation, as well as the advantages of being portable. The present dissertation is divided in two chapters, whose objectives are: modeling and spatial prediction of the available micronutrients contents Fe, Mn, Cu, and Zn, through data obtained from terrain attributes (TA), pXRF, and parent material information (PM), for surface and subsurface horizons separately and combined (n = 153), in different combinations of datasets and spatial resolution, using the random forest (RF) algorithm; and modeling and spatial prediction of the available levels of the macronutrients P, Ca and K, through the pXRF sensor data for the surface horizon (n= 90), using simple linear regression (LR), polynomial regression (PR), power regression (PwR), multiple linear regression (SMLR) and random forest (RF). The study area is located between longitudes 501031 and 504192 mE and latitudes 7651139 and 7653537 mN, zone 23 K, located on the *campus* of the Federal University of Lavras, with approximately 315 ha. Its soils are developed from gneiss, gabbro and alluvial sediments. The climate is Cwa according to Köppen classification system, with average annual temperature of 20.4 °C and average annual rainfall of 1.460 mm. Samples were collected on a regular grid design of 200 m between sampling places., Samples were submitted to laboratory analysis to determine the respective nutrients. Subsequently, a portion of each sample was analyzed on a pXRF model S1 Titan LE (Bruker Nano Analytics, Kennewick, WA, USA) in Trace mode for 60 s in triplicate. The TA were generated with the SAGA GIS software from 5 and 10 m resolution DEM. The data were separated into training (70%) and validation sets (30%), and the models were generated in R software (RF) and SigmaPlot (LR, PR, PwR and SMLR). For the purposes of analysis and comparison between the models, we used the coefficient of determination (R^2), adjusted R^2 (R^2_{adj}), root mean squared error (RMSE), normalized root mean squared error (nRMSE) and mean error (ME) for the first chapter, and R^2, RMSE, mean absolute error (MAE) and the residual deviation of predictions (RPD) for the second chapter. After determination of the best models, the spatial prediction was followed to generate the available nutrient map. The variables of the pXRF when present in the model were spatialized for the entire area through inverse distance weighting (IDW) interpolation. The 10 m TA were better than the 5 m resolution for predictions. It was possible to obtain good results in the spatial prediction of available Fe using only 10 m TA (R^2 = 0.88; RMSE = 59.97 mg kg^{-1} and ME = 24.00 mg kg^{-1}) and for

the others with pXRF + 10 m TA + PM (0.85; 29.65 mg kg^{-1}; 9.70 mg kg^{-1} for Mn, 0.64; 3.11 mg kg^{-1}; 0.71 mg kg^{-1} for Zn e 0.82; 1.17 mg kg^{-1}; 0.43 mg kg^{-1} for Cu, respectively) In the predictions of the macronutrients, the PwR approach obtained the best results (R^2 = 0.80 and RMSE = 1.63 $cmol_c$ dm^{-3} for exchangeable Ca^{2+}, and 0.53 and 6.92 mg dm^{-3} for available P). It was not possible to establish a correlation between the available K^+ contents and the total K_2O content provided by pXRF. Proximal sensor data associated with TA data can accurately predict exchangeable/available nutrient contents in tropical soils.

Key-words: Soil property prediction. Proximal sensors. Machine learning.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

PRIMEIRA PARTE

1 INTRODUÇÃO

O Brasil é um país de enorme extensão territorial, compreendendo diversos climas e substratos litológicos, que resultam em grande variabilidade de classes de solos, compondo a paisagem nas mais diversas combinações. Isso resulta em consideráveis variações de atributos do solo, tanto horizontal e verticalmente, mesmo dentro de ambientes considerados homogêneos. Diante disso, conhecer essa variabilidade dos solos de forma detalhada pode contribuir para aumentos da produtividade e evitar a degradação dos solos.

Para atender as futuras demandas por alimento e garantir a segurança alimentar de maneira sustentável é necessário que haja esforços na introdução de novas tecnologias e metodologias que visem a aprimorar o uso da terra e dos insumos para produção. Neste sentido, o uso de sensores na agricultura vem crescendo e fornecendo importantes dados para a gestão, através da modelagem e mapeamento digital.

Portanto, os objetivos deste trabalho foram: gerar modelos preditivos e mapas digitais da disponibilidade de micronutrientes através de dados de atributos de terreno, espectrometria de fluorescência de raios-X portátil (pXRF) e informação do material de origem dos horizontes superficiais e subsuperficiais dos solos, utilizando um algoritmo de aprendizagem de máquina; e gerar modelos preditivos e mapas da disponibilidade de macronutrientes através dos dados do pXRF para o horizonte superficial dos solos, utilizando regressões lineares e múltiplas, além de algoritmo de aprendizagem de máquina em uma área com diferentes usos, solos, relevos e materiais de origem.

2. REFERÊNCIAL TEÓRICO

2.1 Modelagem e mapeamento digital de solos

Os esforços na compreensão da variabilidade espacial dos solos e suas propriedades utilizando o poder computacional provavelmente se iniciaram nos anos 60 nos primeiros Sistemas de Informação Geográfica (SIG) com abordagem inteiramente geoestatística (MCBRATNEY; SANTOS; MINASNY, 2003), ou seja, espacializando uma informação de um local em que ela seja conhecida (e.g. classe de solo) para locais

não amostrados através de métodos baseados na distância (LAGACHERIE et al., 2001). No decorrer do tempo, com maior poder de processamento e novos produtos disponíveis, como dados de satélites, o uso de modelos digitais de elevação (MDEs) e atributos de terreno (ATs) derivados do MDE como variáveis auxiliares tornaram-se comuns na tentativa de explicar a variabilidade dos solos (LAGACHERIE & McBRATNEY, 2006). MDEs e ATs são representações do terreno no ambiente computacional no formato raster, como uma matriz de células (ou pixels em uma fotografia). O tamanho da célula corresponde à resolução do raster, por exemplo, células de 5x5 m implicam em um raster com resolução de 5m, que contém uma informação do terreno a cada célula de 25 m² (HENGL, 2006).

No atual cenário, são inúmeras as ferramentas e fontes, o que gera uma gama de possíveis combinações e abordagens na tentativa de estabelecer bons modelos preditivos (ARROUAYS et al., 2014). Desde abordagens mais comuns, como regressões matemáticas (SILVA et al., 2015; LACARCE et al., 2012; JUHOS; SZABO; LADANYI, 2015; JUNIOR et al., 2016; MENEZES et al., 2016) até complexos algoritimos de aprendizagem de máquina (FORKUOR et al., 2017; HENGL et al., 2017; HEUNG et al., 2016; TAGHIZADEH-MEHRJARDI et al., 2015). Dentre estes, o Random Forest (BREIMAN et al., 2001) vem sendo cada vez mais utilizado (BELGIU & DRĂGU, 2016) e proporcionando bons resultados nas predições de propriedades do solo (GRIMM et al., 2008; HENGL et al., 2015; JUNIOR et al., 2016; BHERING et al. 2016; CHAGAS et al. 2016).

O Random Forest cria “n” cópias do banco de dados original reamostrando o mesmo. Estas diferentes cópias geram “n” modelos ou árvores que compõem uma floresta, cuja performace é avaliada através dos dados que ficaram fora da reamostragem. Desta forma, calcula-se um índice de erro que permite que a melhor composição de um modelo final e explicação da importância das variáveis preditoras no processo (LIAW; WIENER, 2002).

2.2 Espectrometria portátil de fluorescência de raios-X

Capaz de quantificar uma gama de elementos em segundos (RIBEIRO et al., 2017), o princípio operacional dos espectrômetros de fluorescência de raios-X portáteis (pXRF) é similar aos grandes aparatos usados em laboratórios (WEINDORF; BAKR;

ZHU, 2014). Fótons de raios-X de alta energia incidem sobre o material analisado a fim de excitar os elétrons das camadas internas dos átomos, fazendo-os sair de orbitais mais internos para orbitais mais externos. Estes, por sua vez, ao retornarem aos seus orbitais originais, emitem o excedente energético na forma de fluorescência. Esta é específica de cada elemento, o que permite a identificação e quantificação de diferentes elementos (WEINDORF et al., 2012b).

Excelentes resultados utilizando dados do pXRF já foram reportados na ciência do solo, em predições de pH (SHARMA et al., 2014; SILVA et al., 2017), capacidade de troca de cátions (SHARMA et al., 2015), textura (ZHU; WEINDORF; ZHANG, 2011), contaminação do solo (CHAKRABORTY et al., 2017), estudos de gênese e caracterização de perfil do solo (SWANHART et al., 2014; SILVA et al., 2018; STOCKMANN et al., 2016; WEINDORF et al., 2012a), além de variáveis para o mapeamento de classes de solo (SILVA et al., 2016) e outros.

REFERÊNCIAS

ARROUAYS, D. et al. GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. **Adv. Agron**. v. 125, p. 93–134, 2014.

BHERING, S. B. et al. Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. **Pesquisa Agropecuaria Brasileira**, v. 51, n. 9, p. 1359–1370, 2016.

BELGIU, M.; DRĂGU, L. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. **Remote Sens**. v. 114, p. 24–31, 2016.

BREIMAN, L. Random Forests. **Mach. Learn**. v. 45, p. 5–32, 2001.

CHAKRABORTY, S. et al. Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator kriging. **Geoderma**. v. 306, p. 108–119, 2017.

CHAGAS, C. da S. et al. Spatial prediction of soil surface texture in a semiarid region using random forest and multiple linear regressions. **Catena**. v. 139, p. 232–240. 2016.

FORKUOR, G.et al. 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**, v. 12, 2017.

GRIMM, R. et al. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random Forests analysis. **Geoderma**, 146, 102–113, 2008.

HENGL, T. Finding the right pixel size. **Computers & Geosciences**, v. 32, n. 9, p. 1283–1298, 2006.

HENGL, T.et al. Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. **PLoS One**, v. 10, p. 1–26, 2015.

HENGL, T. et al. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. **Nutrient Cycling in Agroecosystems**, v. 109, n. 1, p. 77–102, 2017.

HEUNG, B. et al. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. **Geoderma**, v. *265*, p. 62–77, 2016.

LAGACHERIE, P. et al. Mapping of reference area representativity using a mathematical soilscape distance. **Geoderma**, v. 101, p. 105– 118, 2001.

LACARCE, E. et al. Mapping soil Pb stocks and availability in mainland France combining regression trees with robust geostatistics. **Geoderma**, v. 170, p. 359–368, 2012.

LAGACHERIE, P.; MCBRATNEY, A. B. Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In: LAGACHERIE, P; MCBRATNEY, A. B.; VOLTZ, M. **Digital Soil Mapping: An Introductory Perspective**. 1. Ed., Amsterdam: Elsevier Science, 2007. p. 3-22.

LIAW, A.; WIENER, M. Classification and Regression by randomForest. **R News**, v. 2, p. 18–22. 2002.

MCBRATNEY, A. B.; SANTOS, M. L. M.; MINASNY, B. On digital soil mapping. **Geoderma**, v. 117, 2003.

MENEZES, M. D. de. et al. 2016. Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. **Sci. Agric**. v, 73, p. 274–285, 2016

JUNIOR, W. de C. et al. Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. **Pesqui. Agropecu. Bras**. v. 51, p. 1428–1437, 2016.

JUHOS, K.; SZABO, S.; LADANYI, M. Influence of soil properties on crop yield: A multivariate statistical approach**. Int. Agrophysics**, v. 29, p. 433–440, 2015.

RIBEIRO, B.T. et al. Portable X-ray fluorescence (pXRF) applications in tropical Soil Science. **Ciência e Agrotecnologia**, v. 41, p. 245–254, 2017.

SWANHART, S. et al. Soil salinity measurement via portable X-ray fluorescence spectrometry. **Soil Sci**, v. 179, p. 417–423, 2014.

SHARMA, A. et al. Characterizing soils via portable X-ray fluorescence spectrometer: 3. Soil reaction (pH). **Geoderma**, v. 232–234, p. 141–147, 2014.

SHARMA, A. et al. Characterizing soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). **Geoderma**, v. 239, p. 130–134, 2015.

SILVA, S.H.G. et al. Proximal sensing and digital terrain models applied to digital soil mapping and modeling of Brazilian Latosols (Oxisols). **Remote Sens**. v. 8, p. 614-635, 2016

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. Soil weathering analysis using a portable X-ray fluorescence (PXRF) spectrometer in an Inceptisol from the Brazilian Cerrado. **Appl. Clay Sci**, v. 162, p. 27–37, 2018.

STOCKMANN, U. et al. Utilizing portable X-ray fl uorescence spectrometry for in- fi eld investigation of pedogenesis, **Catena**, v. 139, p. 220–231, 2016.

TAGHIZADEH-MEHRJARDI, R. et al. Comparing data mining classifiers to predict spatial distribution of USDA-family soil groups in Baneh region, Iran. **Geoderma**, v. 253–254, p. 67–77, 2015.

WEINDORF, D.C.; BAKR, N.; ZHU, Y. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic applications. **Adv. Agron**, v. 128, p. 1- 45, 2014.

WEINDORF, D.C. et al. Enhanced pedon horizonation using portable X-ray fluorescence spectrometry. **Soil Sci. Soc. Am. J**, v. 76, p. 522-531, 2012a.

WEINDORF, D.C. et al. Characterizing soils via portable X-ray fluorescence spectrometer: 2. Spodic and Albic horizons. **Geoderma**, v. 189–190, p. 268–277, 2012b.

ZHU, Y., WEINDORF, D.C., ZHANG, W., 2011. Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. **Geoderma**, v. 167–168, p. 167–177, 2011.

SEGUNDA PARTE - ARTIGOS

ARTIGO 1

Synthesis of Proximal Sensing, Terrain Analysis, and Parent Material Information for Available Micronutrient Prediction in Tropical Soils

ABSTRACT

In developing countries, the use of proximal and remotely sensed data is of critical importance as a less expensive means of obtaining soils information. While proximal sensor approaches such as portable X-ray fluorescence (pXRF) spectrometry are becoming increasingly used to predict soil properties worldwide, remotely sensed data has also been used for terrain analysis in recent decades with the aid of powerful interpretive algorithms. The aims of this work were to apply a random forest algorithm to model and predict the available contents of Fe, Cu, Mn, and Zn from pXRF data in addition to terrain attributes (TAs) with 5 and 10 m spatial resolution and parent material information. The data were used separately and together in an area with high variability of parent materials. Soil samples (n=153) were collected, analyzed by pXRF, and subjected to laboratory analyses to determine the available contents of Fe, Cu, Mn, and Zn. Twelve TAs were generated from digital elevation models (DEM). These data were divided into five datasets (or random forest inputs): pXRF data; TA 5 m data; TA 10 m; pXRF + TA 5 m; and pXRF + TA 10 m. Predictions were performed to assess the importance of such variables. Models were validated with an independent set of samples. Finally, the best models were spatially rendered to cover the entire study area and maps were also validated. The combination of pXRF data and TA covariates in addition to parent material information allowed accurate predictions of available Fe, Mn, Cu, and Zn through the random forest algorithm. Parent material information improved the predictions. Pixel size of 10 m resolution promoted better results than 5 m

resolution. Available Fe contents were better predicted using only TA data. For the spatial prediction of available micronutrients, validation of maps resulted in R^2 of 0.88, RMSE of 59.97 mg kg^{-1} and ME of 24.00 mg kg^{-1} for Fe; 0.85, 29.65 mg kg^{-1}, 9.70 mg kg^{-1} for Mn, 0.64, 3.11 mg kg^{-1}, 0.71 mg kg^{-1} for Zn and 0.82, 1.17 mg kg^{-1}, 0.43 mg kg^{-1} for Cu, respectively. Available micronutrient contents can be accurately predicted using pXRF data in association with terrain and parent material information.

Keywords: portable X-ray fluorescence spectrometer, random forest, soil property prediction, tropical soils, Brazil.

INTRODUCTION

Precision agriculture is a rapidly expanding approach for optimizing agronomic productivity of agricultural lands. The use of contemporary equipment in support of precision agriculture offers a better understanding of spatial variability in soil properties which could decrease soil survey costs and improve mapping accuracy for agronomic development. Constraints concerning precision agriculture in tropical countries include difficulty regarding the use of software and equipment due to the lack of technical knowledge and economic limitations in acquiring such (Borghi et al. 2016).

Soil micronutrient deficiencies are an emerging limitation factor, especially for annual, and to a lesser degree perennial, crop production in Brazil (Fageria and Stone 2008). Since most areas do not have a long history of soil and crop management as compared with developed countries, soil forming factors commonly drive the spatial variability of soil nutrients in addition to land management. Thus, micronutrient occurrence in soils is largely related to parent materials, degree of weathering-leaching,

pedogenesis processes, and soil land use (Zhu et al. 2016). The use of such information is of strategic importance for development in tropical countries, since it can improve spatial prediction accuracy as an environmental covariate (Silva et al. 2016a; Menezes et al. 2016; Pelegrino et al. 2016), giving proper support to fertilizer application at variable rates or establishment of management zones.

To assess the nutrient content in soil, samples have traditionally been submitted to a laboratory for analysis. Conventional laboratory analyses are costly, time-consuming, and generate chemical residues. To overcome this issue, the application of new instruments that provide elemental information for soil samples in a rapid, economic, and environmentally friendly way is advantageous. In this sense, portable X-ray fluorescence (pXRF) spectrometry has shown suitable accuracy in several soil studies (Sharma et al. 2014; Stockmann et al. 2016; Weindorf et al. 2012). Further, the technique is non-invasive, non-destructive, and can be applied *in-situ*. Recently pXRF has been increasingly utilized to complement laboratory analyses, in both field and laboratory conditions, with applications for several branches of soil science (Ribeiro et al. 2017; Weindorf et al. 2014), geology (Ryan et al. 2017), archeology (Hunt and Speakman 2015), water (Pearson et al. 2017) and vegetal (McGladdery et al. 2018; Reidinger et al. 2012) analysis, ore grade control, and mineral exploration (Jakob et al. 2016).

Although pXRF provides total elemental contents, reported pXRF elemental data (or raw fluorescence energies) have been used as a proxy for the determination of soil texture (Zhu et al. 2011), pH (Sharma et al. 2014), cation exchange capacity (CEC) (Sharma et al. 2015), soil horizon differentiation (Weindorf et al. 2012), and assessment of soil contamination (Chakraborty et al. 2017; Horta et al. 2015; Paulette et al. 2015).

However, correlations between pXRF data and available micronutrient contents in tropical soils is lacking and warrants further evaluation.

In digital soil mapping (DSM), digital elevation models (DEMs) and terrain attribute (TA) derivatives from DEMs are commonly used as auxiliary information to improve spatial predictions (McBratney et al. 2003). With these topography-related data, pedologists search for parameters that explain soil variability (Lagacherie and McBratney 2006). TAs as well as DEMs are represented in raster format, which deliver information cell by cell (or pixel by pixel) in a continuous way (Arrouays et al. 2017). As a consequence, the spatial resolution may influence the results (Lecours et al. 2017; Mokarram and Hojati 2017; Penížek et al. 2016). Although fine resolution provides greater topographic detail, it can also introduce local artifacts and slow down data processing, whereas coarse resolution furnishes less detail and often provides missing channels and peaks (Hengl 2006).

However, DEMs and their derivatives do not always guarantee accurate spatial soil predictions in isolation. The integration of other input data such as climatic variables, remotely sensed data (e.g., hyperspectral imagery), and/or lithology information are recommended (Hengl et al. 2017). In areas featuring considerable variability of soil parent materials or areas where existing geology maps have small scales, pXRF and other proximal sensors can generate data related to parent materials. Such data has the potential to be used as covariates to support modeling and spatial predictions (Gray et al. 2016; Silva et al. 2016b).

A range of algorithms have been used by the digital soil mapping community to model and predict soil data. Among them, one increasingly used is an ensemble machine learning method called random forest (RF) (Breiman 2001). It is considered an ensemble method since it combines the results of many classifiers (Liaw and Wiener

2002), in this case, many classification or regression trees, to generate a final prediction (Breiman et al. 1984). Such trees, which are combined to form a forest, are created by resampling the data set (observations and predictors) with substitution (bootstrap) (Collard et al. 2014). By fitting a number of trees, the chance of model overfitting is minimum. RF is capable of fitting non-linear relationships and has proven to be a powerful algorithm, although its interpretability is difficult (Thompson et al. 2012). RF has been successfully applied to soil property prediction both spatially (Bhering et al. 2016; Chagas et al. 2016; Hengl et al. 2017; Silva et al. 2017) and non-spatially (Carvalho Junior et al. 2016).

The aims of this work were to: 1) characterize soils developed from variable parent materials with pXRF, evaluate random forest modeling for the prediction of plant available Fe, Cu, Mn, and Zn from pXRF data separately and in association with TA of 5 and 10 m spatial resolution and soil parent material as auxiliary information, and 2) assess the importance of such variables in micronutrient prediction in an area with different land uses and high spatial variability of parent materials, thus reducing the cost and time needed to characterize micronutrients in tropical soils. The hypothesis of this work is that combined high resolution TA, pXRF data, and parent material information will promote prediction of plant available micronutrients with higher accuracies than using TA in isolation. For that, models will be created using random forest algorithms for the different datasets and spatial predictions of the available micronutrient contents will be made and validated through an independent dataset in an area with variable topography and parent materials.

MATERIAL AND METHODS

Study area

The study area (Figure 1) is located between UTM longitudes 501031 and 504192 m E and latitudes 7651139 and 7653537 m N, zone 23K. The study area is ~315 ha, featuring native forest, eucalyptus and pinus plantations, agricultural crops, and pasture. The climate is Cwa per the Köppen classification, with hot and humid summers and cold and dry winters; average annual temperature is 20.4°C and average annual rainfall is 1,460 mm (Dantas et al. 2007).

In the study area, the soil parent materials are gneiss (metamorphic rock mainly composed by quartz, feldspars, and micas) (75%), gabbro (igneous rock basically

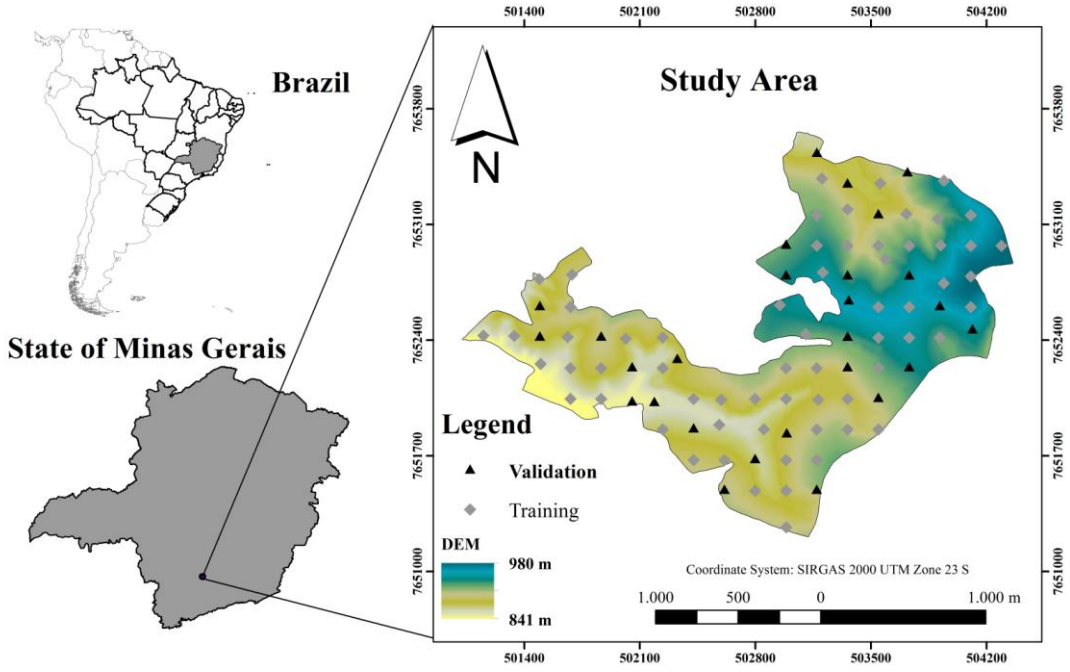

Fig. 1 Location map of the study area showing samples used for modeling and validation

composed by Ca-rich plagioclase and pyroxene minerals) (20%), and alluvial sediments (5%). Such parent material (PM) information was used in some of the models for predicting available micronutrient contents. The main soils developed from these parent materials are Oxisols, Ultisols, Inceptisols, Alfisols, Entisols and Histosols (Curi et al. 2017), per Keys to Soil Taxonomy (Soil Survey Staff 2014).

Soil sampling and laboratory analyses

Composite soil samples were collected in 90 places forming a regular grid design, with 200 m of distance between samples (Figure 1), comprising a total of 153 samples (90 - A horizon; 63 - B horizon). Such sampling density adequately encompasses all variability in soil classes and parent materials across the study area. The collected samples were air-dried, disaggregated to pass through a 2-mm sieve, and subjected to laboratory analyses for the determination of available Fe, Cu, Zn, and Mn per Mehlich-1 extraction (Mehlich 1953). Quantification was made via atomic absorption spectrophotometry on an AAS 800 (Perkin Elmer, Waltham, MA, USA) per Jackson (1958).

Samples were scanned with an S1 Titan LE (Bruker Nano Analytics, Kennewick, WA, USA) pXRF, in Trace (dual soil) mode per Weindorf and Chakraborty (2016). Each sample was scanned for 60 s in triplicate, with detected fluorescence analyzed via the integrated GeoChem software. Prior to each analysis with pXRF, the accuracy of the instrument was checked using two samples (2710a and 2711a) certified by the National Institute of Standards and Technology (NIST) as well as a sample provided by the pXRF manufacturer (check samples - CS). Since the element contents are known for those samples, the recovery percentage of each element was calculated (Table 1) to ensure the quality and accuracy of the generated data. In sequence, the elements (and

Table 1 Recovery percentage (%) of elements and compounds obtained by portable X-ray fluorescence spectrometer (pXRF) scanning of certified reference materials

Certified material	Al_2O_3	SiO_2	K_2O	CaO	Ti	Mn	Fe	Cu	Zn	Zr
Check sample	92	95	88	--[a]	--[a]	92	93	23	--[a]	--[a]
2710a	126	107	61	42	--[a]	67	70	83	96	106
2711a	129	107	60	71	75	65	71	74	88	--[a]

[a]No certified value.

compounds) that were detected in all samples and used for purposes of this study included: Al_2O_3, SiO_2, Cl, K_2O, CaO, Ti, Mn, Fe, Cu, Zn, and Zr.

Terrain attributes data set

Contour lines of 1 m vertical equidistance were derived from a planialtimetric survey, while DEMs of 5 and 10 m resolution were obtained using the *Topo to Raster* interpolation function (ANUDEM algorithm) in ArcGIS 10.3 (ESRI, The Redlands, CA, USA). Then, the *Fill* function was applied to make the DEM hydrologically consistent. The TA derivatives from the DEMs were created in SAGA GIS software (Conrad et al. 2015), and included: Aspect, Channel Network Base Level, Flow Accumulation, LS-Factor, Multiresolution Ridge Top Flatness, Multiresolution Valley Bottom Flatness, Plan Curvature, Profile Curvature, Slope, Saga Wetness Index, Topographic Wetness Index, and Vertical Distance to Channel Network. Such TAs were chosen based on several works published worldwide that successfully predicted soil properties from such terrain information (Adhikari et al. 2013; Akpa et al. 2014; Bishop et al. 2015; Florinsky et al. 2002; Forkuor et al. 2017; Gessler et al. 1995; Giasson et al. 2006; Menezes et al. 2018; Moore et al. 1993; Silva et al. 2016). Through georeferenced sampling sites, the values from each terrain attribute and altitude from the DEM were extracted. Using parent material information, elemental data from pXRF, and TA data for each sampling site, random forest models were adjusted to predict available Fe, Cu, Zn, and Mn.

Random forest

The random forest algorithm is an improved version of decision tree algorithms and operates similarly for classification and regression. The algorithm reduces the common overfitting of decision trees algorithms that frequently present problems of over-adjustment due to limited, sparse, or unrepresentative data using two meta-algorithms: bootstrap aggregation (bagging) and random sample with substitution (Breiman 2001). First, the algorithm creates a dataset of the same size as the original one adapted to the number of trees in the forest, created from random resampling with substitution (bootstrap). These are now the bootstrap datasets, which may contain duplicates of the original data and absence of some, which is termed 'bagging.' RF computes the mean of trained trees in different parts of the same training set, which reduces variance and generally increases model performance. While a single tree model is highly sensitive to training set noise, the average of a forest is not, since the trees are not correlated because they are trained with bootstraps sets (different sets of predictors originated from their resampling with substitution). This process estimates performance improvement as different predictors are used in trees and evaluates forecasting observations through an internal accuracy calculation from data that were not used in the model construction, termed "out-of-bag" (OOB) data. Such OOB predictions are, then, aggregated and used to calculate the OOB error. From this measure the mean square error (MSE_{OOB}) is calculated. Finally, percent variance explained (Var_{ex}) by the model is computed as Eq. 1:

$$Var_{ex} = 1 - \frac{MSE_{OOB}}{\sigma_y^2} \quad (1)$$

where σ_y^2 is the total variance of the variable calculated as the number of samples, instead of the number of samples minus 1 (Liaw and Wiener 2002). In addition, the models also provide the ranking of the most important variables for the model by

evaluating the increase in OOB error caused by the exclusion of one predictor from the dataset by bootstrapping while other predictors are maintained (Peters et al. 2007).

Predictions and accuracy assessment

Random forest models were adjusted for the samples per horizon (90 samples for A and 63 samples for B horizons) and also by combining samples of both horizons (153 samples), either considering or not considering PM information as follows: i) pXRF data, solely; ii) pXRF data + TA (5 m); iii) pXRF data + TA (10 m); iv) TA (5 m), solely; v) TA (10 m), solely. Data were separated using 70% for calibration (model development) and 30% for validation. For validation of the models, predictions were made for an independent set of samples (30% of total data). Then, the available Fe, Cu, Zn, and Mn predicted via modeling were compared to the observed contents of the validation data and the following statistical parameters were calculated between predicted and observed data: coefficient of determination (R^2), R^2_{adj}, root mean squared error (RMSE), and mean error (ME). Calculations for RMSE and ME follow as Eqs. 2 and 3, respectively:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}\left(X_{obs,i} - X_{model,i}\right)^2}{n}} \quad (2)$$

$$\text{ME} = \frac{1}{n}\sum_{i=1}^{n}\left(X_{obs,i} - X_{model,i}\right) \quad (3)$$

where X_{obs} represents observed values in laboratory analysis and X_{model} represents estimated values from pXRF and TA at time/place i in n observation points. The RMSE was normalized (RMSEn) to compare variables at different scales (Eq. 4):

$$\text{RMSEn} = \frac{RMSE}{\sigma} \quad (4)$$

where: σ is standard deviation.

Finally, optimal models for available Fe, Cu, Zn, and Mn were used to create maps of available micronutrient contents for the entire study area. pXRF results were modeled across the entire area through inverse distance weighting (IDW) interpolation per Eq. 5:

$$\hat{E} = \frac{\sum_{i=1}^{n} \frac{E}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \quad (5)$$

where $\hat{E}$ is a value to be estimated, E is a value obtained from pXRF, d_i^p is the distance from the n data points to the power of p of the point estimated. Thus, maps of available Fe, Cu, Zn, and Mn were made for the entire area based on the optimal models. These maps were validated with R^2, RMSE, and ME as aforementioned.

RESULTS AND DISCUSSION

Characterization of soils by pXRF according to parent material

Tables 2 and 3 present the elemental contents obtained by pXRF for A and B horizons of soils developed from different PM. Gabbro promoted the formation of soils with higher mean contents of Fe, whereas SiO_2, K_2O, and CaO contents were slightly lower compared to gneiss-derived soils. This is likely because most gabbro-derived soils are Oxisols, formed under intense weathering and leaching which remove bases from the soil (Curi et al. 2017; Resende et al. 2014; Shaetzl and Anderson 2005).

Results showed approximately five times greater CaO contents in the A horizon (Table 2) of soils derived from gabbro and gneiss when compared to the respective subsurface horizon values (Table 3). These differences may be due to liming residues, a common practice in the tropical soils (Lopes and Guilherme 2016) for superficial acidity correction.

Table 2 Summary statistics of portable X-ray fluorescence spectrometer (pXRF) data (mg kg^{-1}) analyzing A horizons of soils derived from gabbro, gneiss and alluvial sediments in Brazil

Parameter	Al_2O_3	SiO_2	Cl	K_2O	CaO	Ti	Mn	Fe	Cu	Zn	Zr
	----------Gabbro----------										
Min (mg kg^{-1})	139,5	119,53	353	847	219	5178	161	39,605	18	20	134
Max (mg kg^{-1})	245,7	338,56	820	3989	6968	16,23	1526	174,67	140	94	288
Mean (mg kg^{-1})	170,8	222,85	541	2138	2829	11,31	683	114,25	52	42	201
SD (mg kg^{-1})	31,81	77,712	127	1200	2046	3958	457	49,107	35	22	55
CV (%)	18.6	34.9	23.4	56.1	72.3	35.0	66.9	43.0	68.0	51.5	27.2
	----------Gneiss----------										
Min (mg kg^{-1})	103,1	124,46	285	912	492	2428	139	15,902	10	15	140
Max (mg kg^{-1})	240,2	414,75	935	34,10	20,45	18,30	2153	173,77	106	123	332
Mean (mg kg^{-1})	179,4	282,24	560	3863	4935	9356	640	74,483	33	41	220
SD (mg kg^{-1})	33,01	69,264	127	5599	4089	3737	416	40,792	19	22	46
CV (%)	18.4	24.5	22.6	144.9	82.9	39.9	65.0	54.8	57.7	53.8	21.1
	----------Alluvial Sediments----------										
Min (mg kg^{-1})	155,7	145,40	450	1202	873	8660	129	38,687	31	17	130
Max (mg kg^{-1})	216,6	348,87	586	2935	2120	12,75	353	83,454	114	57	209
Mean (mg kg^{-1})	184,7	269,42	528	1895	1561	11,03	252	66,880	62	40	171
SD (mg kg^{-1})	24,93	88,845	57	749	517	1735	93	20,038	37	17	32
CV (%)	13.5	33.0	10.9	39.5	33.1	15.7	36.8	30.0	59.1	42.2	18.9

SD= Standard Deviation; CV= coefficient of variation

Mean SiO_2 and K_2O contents were higher in soils derived from gneiss. Similar average values of Al_2O_3 were also observed for soils developed from gneiss and sediments, and slightly lower for gabbro-derived soils (Araujo et al. 2014; Lacerda et al. 2002; Marques Jr. et al. 1992). As expected, gabbro-derived soils presented lower contents of SiO_2 than gneiss-derived soils. Similar trends were observed for Ti, Mn, Cu, and Zn, whose values were greater for gabbro-derived soils, reflecting the effect of greater contents of such elements in the chemical composition of gabbro in relation to gneiss.

Table 3 Summary statistics of portable X-ray fluorescence spectrometer (pXRF) data (mg kg^{-1}) analyzing B horizons of soils derived from gabbro, gneiss, and alluvial sediments in Brazil

	Al_2O_3	SiO_2	Cl	K_2O	CaO	Ti	Mn	Fe	Cu	Zn	Zr
	----------Gabbro----------										
Min (mg kg^{-1})	141,24	101,1	394	705	156	44583	81.6	48663	26.6	25	129
Max (mg kg^{-1})	239,78	307,8	722	3742	778	15,66	894	180,54	141	86	273
Mean (mg kg^{-1})	183,85	200,3	548	2031	390	11,36	596	134,68	63	46	199
SD (mg kg^{-1})	29,521	78,06	122	1225	188	3848	268	48,317	38	21	54
CV (%)	16.1	39.0	22.4	60.3	48.2	33.9	44.9	35.9	59.7	45	27.
	----------Gneiss----------										
Min (mg kg^{-1})	113,47	138,3	348	869	154	2821	70	21,711	8	14	130
Max (mg kg^{-1})	226,27	262,6	635	3682	395	11,23	109	100,69	70	81	260
Mean (mg kg^{-1})	202,33	254,3	604	2851	802	8912	432	82,361	36	36	217
SD (mg kg^{-1})	37,279	31,90	52	1383	114	631	413	14,170	22	31	25
CV (%)	18.4	12.6	8.6	48.5	142.	7.1	95.6	17.2	62.0	85	11.
	----------Alluvial Sediments----------										
Min (mg kg^{-1})	167,56	284,6	506	1049	283	5912	160	29,904	17	26	169
Max (mg kg^{-1})	208,83	338,4	703	3036	215	8994	269	45,706	18	39	305
Mean (mg kg^{-1})	207,44	322,6	656	3026	139	7453	268	37,805	18	32	242
SD (mg kg^{-1})	1388	15,83	47	10	757	1541	1	7901	0	7	62
CV (%)	0.7	4.9	7.1	0.3	54.3	20.7	0.5	20.9	2.1	21	25.

SD= Standard Deviation; CV= coefficient of variation

The results for alluvial soils resemble those derived from gneiss. This might be a consequence of these sediments being surrounded by upper soils derived from gneiss and, therefore, a large portion of these sediments may be inherited from such soils.

Random forest modeling

Figure 2 shows the percentage of variance explained by random forest models as a measurement for predicting available contents of Fe, Cu, Zn, and Mn. The greater the percentage of variance explained, the better the model performed. In Fe modeling, all

models had positive variance. For Mn, all the models that contained pXRF data had positive variance. Modeling using only TA of 5 m resolution resulted in positive parameters for A horizon data.

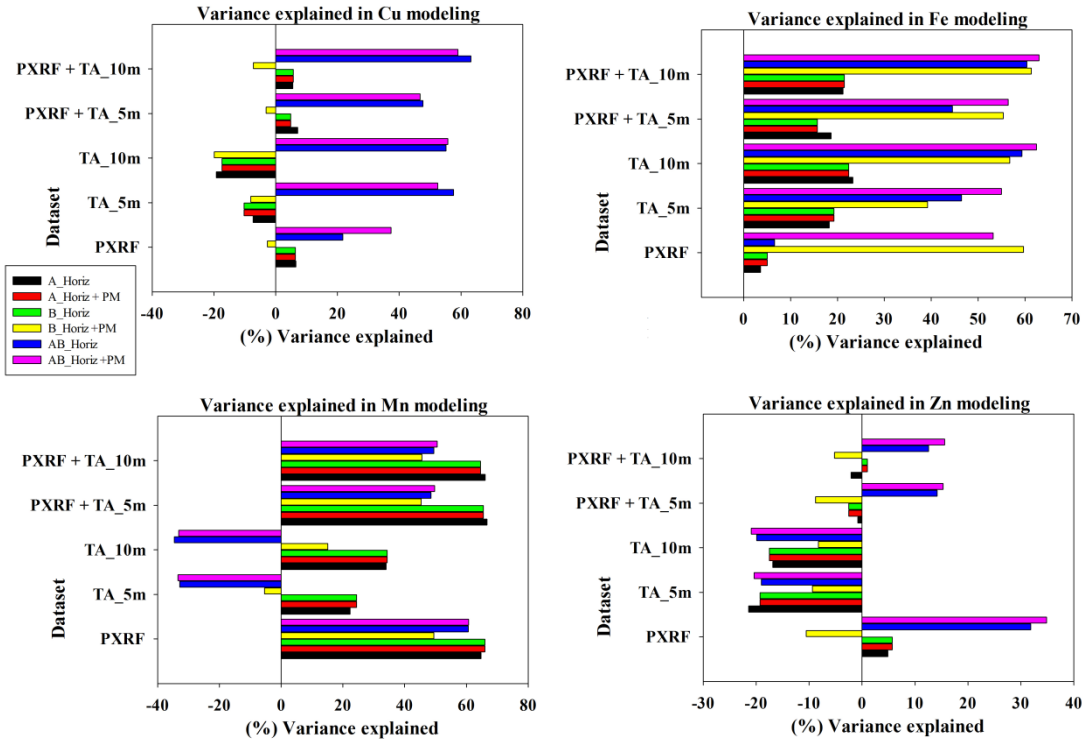

Fig. 2 Variance explained from random forest models in all dataset arrangements for available contents of Fe, Mn, Zn and Cu in Brazilian soils

When using 10 m TA information, positive parameters were obtained for A and B horizons. The data for A and B horizons combined using only TA information did not present any positive values.

For Zn modeling, the majority of models, except for A and B horizons together, had a negative variance. Only when pXRF data were present, were positive values observed. For Cu models using only TA information for A horizon data, neither 5 nor 10 m resolution achieved positive results. When pXRF data was used in isolation or in

tandem with PM, small positive variances were found. When utilizing data for A and B horizons together with or without PM, the models yielded positive variance.

Validation of the random forest models

Table 4 presents the R^2_{adj} obtained through comparison of estimated and real values of available Fe, Mn, Zn, and Cu, considering the different inputs on random forest models. A suitable adjustment was found for at least one model for all elements. For available Fe, only TA at 10 m resolution and PM were enough to provide the best predictions. For available/exchangeable Cu, pXRF data was needed in addition to those two types of data for optimal predictions.

Table 4 Coefficient R^2 adjusted for all dataset arrangements in the predictions of available Fe, Mn, Zn, and Cu in Brazilian soils

Dataset	Model	A Hor	A Hor + PM	B Hor	B Hor + PM	A+B Hor	A+B Hor + PM
pXRF	Fe	-0.04	0.37	0.01	0.44	0.10	0.10
	Mn	**0.55**	**0.56**	0.23	0.23	**0.63**	**0.63**
	Zn	0.06	0.08	0.06	0.09	0.06	0.06
	Cu	0.02	0.03	0.07	**0.47**	**0.52**	**0.52**
TA 5m	Fe	0.13	0.40	**0.58**	**0.77**	0.30	0.41
	Mn	0.04	0.05	-0.06	-0.06	0.07	0.07
	Zn	-0.03	-0.03	0.13	0.18	0.00	0.00
	Cu	-0.04	-0.04	0.32	0.45	0.09	0.08
TA 10m	Fe	**0.80**	**0.80**	**0.86**	**0.87**	**0.79**	**0.78**
	Mn	0.08	0.07	0.09	0.11	0.18	0.18
	Zn	-0.04	-0.04	0.28	0.28	-0.02	-0.01
	Cu	-0.04	-0.04	**0.49**	**0.57**	0.08	0.05
TA 5m + pXRF	Fe	**0.52**	0.21	0.29	**0.56**	0.30	0.42
	Mn	**0.58**	**0.59**	0.18	0.20	**0.68**	**0.68**
	Zn	0.07	0.05	0.17	0.16	0.25	0.26
	Cu	0.02	0.02	**0.65**	**0.70**	0.33	0.33
TA 10m + pXRF	Fe	**0.79**	**0.81**	**0.81**	**0.84**	**0.82**	**0.81**
	Mn	**0.58**	**0.57**	0.59	0.20	**0.70**	**0.71**

Zn	0.05	0.05	0.05	0.24	0.26	0.28
Cu	0.03	0.03	0.05	**0.75**	0.49	0.39

Hor = soil horizon; PM = parent material; pXRF = dataset from portable X-ray fluorescence spectrometer; TA 5 m = dataset from terrain attributes of 5 m resolution; TA 10 m = dataset from terrain attributes of 10 m resolution.

Utilizing only pXRF data, Mn prediction obtained a moderate R^2_{adj} for the models representing A horizon data (0.55); a slight increase was observed with addition of PM (0.56). Data from the combination of A and B horizons produced an R^2_{adj} of 0.63, which was the best result for available Mn since results for B horizon only reached a maximum of 0.23. For Cu, B horizon with PM and combined A and B horizons reached an R^2_{adj} of 0.47 and 0.52, respectively. The latter value was also obtained and optimal for A + B horizons in addition to PM. For Fe and Zn, the greatest R^2_{adj} values obtained were 0.44 and 0.09, respectively.

Using TA information with 5 m resolution, only predictions for Fe using data from B horizons plus PM obtained a high R^2_{adj} value (0.77), in contrast to 0.58 when PM was excluded. Notably, Cu showed the same tendency of optimal results when B horizons were jointly considered with PM (R^2_{adj} = 0.45). Validation of other nutrients was non-significant.

In the modeling data extracted from TA of 10 m resolution, available Fe showed high values of R^2_{adj} ranging from 0.78 to 0.87 for all horizons and in combinations with and without PM. Prediction models for available Cu with data for the B horizon followed the same trends as the TA with 5 m resolution data, but with greater R^2_{adj} value using 10 m resolution, reaching 0.57 when the PM was included. For available Mn and Zn, results of TA with 10 m outperformed results of TA with 5 m resolution,

but low R^2_{adj} values were obtained. Thus, available Mn and Zn were poorly predicted by terrain information alone.

When using data from TA with 5 m resolution together with pXRF data, the prediction for available Cu for B horizons plus PM showed an improvement (R^2_{adj} of 0.70) over the model featuring TA with 5 m resolution that reached 0.45. This illustrates the power of pXRF data as a covariate. Once again, the inclusion of PM promoted increased R^2_{adj} values, from 0.65 to 0.70. The prediction of available Fe through the same dataset for B horizons with PM worsened in relation to TA with 10 m resolution when pXRF data were added (R^2_{adj} of 0.56). However, this result was better than that with no PM included (0.29). For available Mn, the best results were obtained using data of A and B horizons together, with and without PM; both models reached an R^2_{adj} value of 0.68. This result is slightly better than that obtained with only pXRF data (0.63), showing that TA in combination with pXRF data promotes model improvement. For Zn, the best result (R^2_{adj} of 0.26) was obtained by combining A and B horizons with PM.

Using 10 m resolution TA information and pXRF data, predictions for available Fe remained high, with R^2_{adj} values ranging from 0.79 (A horizon data) to 0.84 (B horizon plus PM); this was slightly lower than using TA information with 10 m resolution in isolation. For available Mn, the best predictions were obtained with these combined covariates, reaching R^2_{adj} values of 0.70 and 0.71, respectively, with and without PM included. For available Cu, results were slightly lower (0.75 for B horizon plus PM) than those using only pXRF data (0.77). For Zn, low R^2_{adj} values were obtained, although this was the dataset that promoted the greatest R^2_{adj} among all the datasets, with 0.26 and 0.28 for A and B horizons without and with PM, respectively.

In Table 5, the validation errors assessed by RMSE and ME are shown. When PMs are inserted, lower RMSE and ME in Fe models for B horizon data were obtained, whereas for A horizon data, the error of predictions increased. For Mn models, PMs slightly affected the errors, mainly in B horizons with PMs. For Zn predictions, B horizon data featured lower RMSE and ME than those for A horizons plus PMs. For Cu predictions, the errors were similar between horizons and there was a considerable improvement using PM in B horizon models.

Figure 3 depicts the variation of normalized RMSE considering different inputs of random forest models. Different combinations of soil horizons heavily influenced model accuracy. For Fe and Zn, the lowest values of RMSEn were found for B horizons. The most accurate prediction was reached for Zn of B horizons (with and without PM). Whereas for Mn, the highest values were found for B horizons. Cu presented a stable variation of RMSEn values, independent of type of horizon/PM or the input models.

Optimal models were selected by observing the greatest R^2_{adj} and the lowest RMSE and ME values (Table 5). Models promoting the best predictions for each available micronutrient follow:

Fe: (TA of 10 m + PM) from B horizon R^2_{adj}= 0.86; RMSE = 55.06 mg kg^{-1}; ME = -25.08 mg kg^{-1}

Mn: (pXRF + TA of 10 m + PM) from A+B horizons R^2_{adj}= 0.71; RMSE = 37.27 mg kg^{-1}; ME = -8.27 mg kg^{-1}

Zn: (pXRF + TA of 10 m + PM) from A+B horizons R^2_{adj}= 0.28; RMSE = 4.17 mg kg^{-1}; ME = -1.80 mg kg^{-1}

Cu: (pXRF + TA of 10 m + PM) from B horizon R^2_{adj}= 0.74; RMSE = 1.32 mg kg^{-1}; ME = -0.39 mg kg^{-1}

Notably, three of the four datasets provided optimal models by utilizing pXRF data in combination with PM and TA information. Only Fe failed to use pXRF data, although the models that employed such variables also reached adequate accuracy. All of the models used TA data and PM. In contrast to TA data, which are easily obtained worldwide through DEM at diverse spatial resolutions (Bhering et al. 2016; Gray et al. 2016; Hengl et al. 2017; Heung et al. 2016; Lecours et al. 2017; Shangguan et al. 2014; Silva et al. 2016a), detailed PM information at large scales is scarce in many parts of the world, especially in developing countries such as Brazil. For this study, a PM map of

Table 5 Root mean square error (RMSE) and mean error (ME) of random forest models to predict available Fe, Mn, Zn, and Cu using different datasets from Brazil

	Error	Dataset	A Hor	A Hor + PM	B Hor	B Hor+PM	A+B Hor	A+ B Hor + PM
Fe	RMSE (mg kg^{-1})	pXRF	218.47	187.69	150.15	99.11	186.60	186.60
		TA_5m	205.45	195.24	**88.77**	**69.37**	170.26	152.94
		TA_10m	109.80	113.37	**58.21**	**55.06**	110.54	111.13
		PXRF + TA_5m	211.04	204.04	113.11	87.59	170.05	149.08
		PXRF + TA_10m	115.75	113.85	115.65	58.14	105.55	107.50
	ME (mg kg^{-1})	pXRF	19.49	16.84	-20.76	-19.29	22.01	22.01
		TA_5m	26.48	25.35	-11.78	-14.43	3.46	10.27
		TA_10m	-13.67	-11.71	-27.40	-25.08	29.94	31.04
		PXRF + TA_5m	34.41	31.98	-17.24	-16.51	-1.86	30.96
		PXRF + TA_10m	-4.61	-4.42	-4.65	-23.92	22.55	23.20
Mn	RMSE (mg kg^{-1})	pXRF	**47.27**	**47.19**	83.07	82.99	**41.96**	**41.96**
		TA_5m	70.40	69.68	93.18	93.43	64.48	64.32
		TA_10m	71.18	71.45	87.33	87.15	59.52	59.32
		PXRF + TA_5m	**44.65**	**44.32**	84.90	84.52	**39.47**	**39.56**
		PXRF + TA_10m	**45.03**	**45.63**	44.59	83.63	**37.86**	**37.27**
	ME (mg kg^{-1})	pXRF	-16.81	-16.96	20.51	20.76	-7.67	-7.67
		TA_5m	-16.46	-17.65	25.22	25.40	2.23	2.51
		TA_10m	-20.59	-20.20	24.07	24.43	0.12	0.73
		PXRF + TA_5m	-16.28	-16.32	22.38	22.22	-6.82	-7.39
		PXRF + TA_10m	-19.28	-19.26	-18.44	21.11	-7.86	-8.27

Zn	RMSE (mg kg^{-1})	pXRF	5.68	5.62	**1.43**	**1.41**	5.96	5.96
		TA_5m	5.19	5.18	**0.73**	**0.71**	3.77	3.72
		TA_10m	5.48	5.56	**0.79**	**0.82**	3.95	3.84
		PXRF + TA_5m	5.31	5.43	**0.83**	**0.81**	4.25	3.99
		PXRF + TA_10m	5.49	5.54	5.53	0.92	4.17	4.05
	ME (mg kg^{-1})	pXRF	-1.68	-1.70	-0.83	-0.82	-2.32	-2.32
		TA_5m	-1.20	-1.23	-0.09	-0.08	-0.72	-0.69
		TA_10m	-1.59	-1.68	-0.28	-0.29	-0.64	-0.57
		PXRF + TA_5m	-1.29	-1.35	-0.31	-0.26	-1.73	-1.60
		PXRF + TA_10m	-1.51	-1.63	-1.52	-0.44	-1.80	-1.77
Cu	RMSE (mg kg^{-1})	pXRF	2.03	2.13	2.53	1.84	2.05	2.05
		TA_5m	2.39	2.38	2.30	1.99	2.76	2.67
		TA_10m	2.68	2.82	**1.74**	**1.60**	2.11	2.29
		PXRF + TA_5m	2.11	1.97	**1.64**	**1.61**	2.49	2.34
		PXRF + TA_10m	2.02	2.11	**1.85**	**1.32**	1.53	1.78
	ME (mg kg^{-1})	pXRF	-0.51	-0.64	-0.54	-0.53	-1.09	-1.09
		TA_5m	-1.04	-0.96	-0.56	-0.43	-1.08	-0.97
		TA_10m	-0.77	-0.77	-0.15	-0.01	-0.58	-0.62
		PXRF + TA_5m	-0.67	-0.67	-0.73	-0.77	-1.03	-0.97
		PXRF + TA_10m	-0.70	-0.75	-0.63	-0.39	-0.63	-0.73

Hor = soil horizon; PM = parent material; pXRF = dataset from portable X-ray fluorescence spectrometer; TA 5 m = dataset from terrain attributes of 5 m resolution; TA 10 m = dataset from terrain attributes of 10 m resolution.

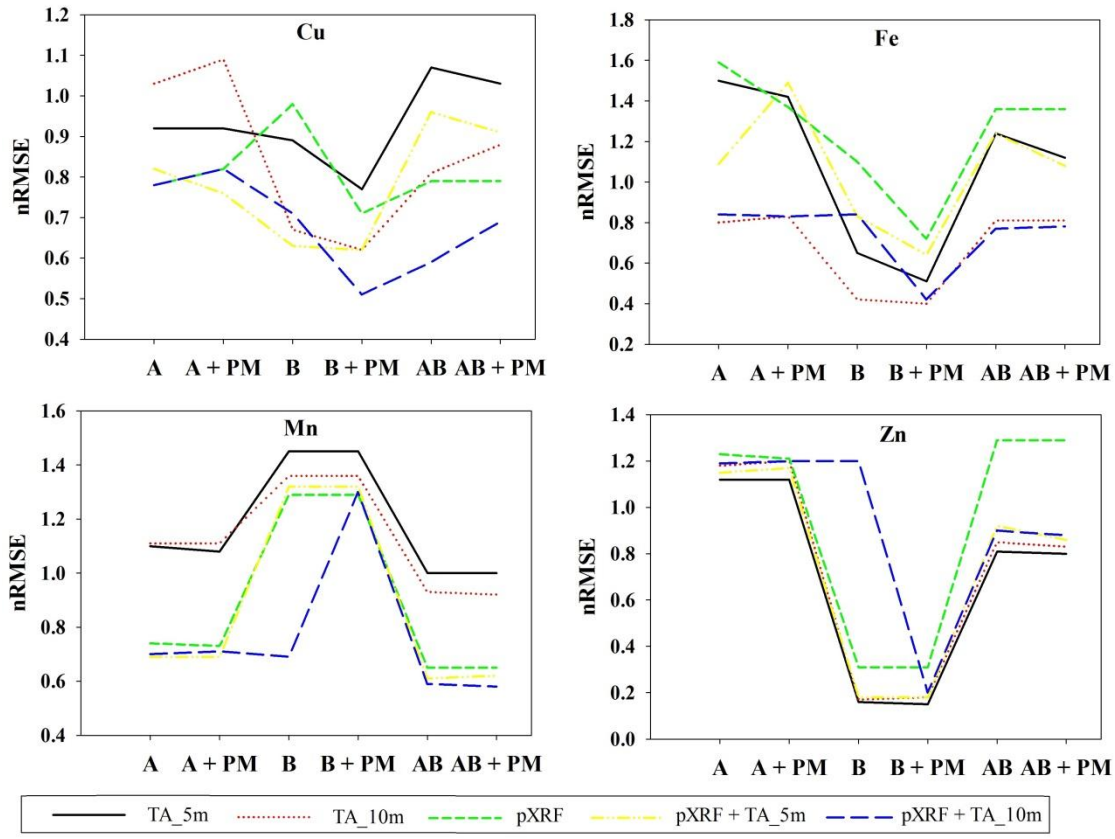

Fig. 3 Normalized root mean square error (nRMSE) of predictions for available Fe, Mn, Zn and Cu for different random forest inputs of Brazilian soils

the study area was obtained from Curi et al. (2017) who employed magnetic susceptibility data of soils using fuzzy logic. This approach provides an alternative to overcome a lack of detailed PM data which can subsequently improve soil property prediction.

The available Fe content tends to be higher in low landscape positions associated with sites where reduction processes occur (Resende et al. 2014). In such conditions, the crystalline structure of Fe-bearing minerals commonly found in tropical weathered-leached soils, such as hematite and goethite, are thermodynamically unstable, releasing Fe to the soil solution (Kämpf et al. 2012). Contrariwise at higher elevations, most Fe is structurally bound in minerals; as such, it is not available (Resende et al. 2014). Since pXRF detects both available and structural Fe, this might be the reason that pXRF data

did not improve the prediction models for this micronutrient; only models with TA information promoted the best predictions. This fact was confirmed by the use of the covariates Vertical Distance to Channel Network, Channel Network Base Level, and DEM in the best prediction models (see next section). The combination of slope, elevation, and baseline information on drainage networks and flow accumulation of water reflect the environmental conditions of oxidation/reduction of elements such as Fe in soil.

Mn undergoes faster reduction than Fe in the absence of oxygen, and is also common in tropical weathered-leached soils (Camargo et al. 1999). As such, accuracy increased by adding TA information. Cu and Zn occur in varying minerals in trace quantities, sometimes as isomorphous substitutes. Thus, the combination of pXRF data and TA information promoted greater R^2_{adj} values.

Analyzing the influence of pixel sizes, all the models including 10 m resolution presented better accuracy than 5 m. This conclusion supports other work that found that smaller pixel sizes do not always promote better predictions (Silva et al. 2016a; Maynard and Johnson 2014; Smith et al. 2006). Furthermore, higher resolution may generate too much noise and artefacts in the process (Cavazzi et al. 2013; Hengl 2006). Resolution has an inverse relationship with the size of the area represented by the basic mapping unit, the pixel. Also, the increase or decrease in the resolution directly implies the values extracted from the different terrain attributes derived from the DEM (Mashimbye et al. 2014) and can still vary with the degree of surface complexity (Lecours et al. 2017). Terrain attributes whose values are computed by accumulating adjacent pixel values (e.g., flow accumulation) have their range of values increased in smaller resolutions, whereas morphometric attributes (e.g., slopes, curvatures) decrease the range of values (Penížek et al. 2016).

Table 6 The five most important covariates for each dataset of each predicted micronutrient for Brazilian soils

Dataset	Order	Avail. Fe	Avail. Mn	Avail. Zn	Avail. Cu
pXRF	1	PM**	CaO***	Zn***	Cu**
	2	Fe	Mn	CaO	PM
	3	SiO_2	Al_2O_3	Fe	Zn
	4	K_2O	SiO_2	Ti	Al_2O_3
	5	Mn	Zn	PM	Mn
TA 5m	1	PM**	Slope*	Plan curvature**	Profile curvature**
	2	DEM 5m	MRVBF	PM	Slope
	3	VDCN	MRRTF	CNBL	MRVBF
	4	CNBL	SWI	TWI	MRRTF
	5	LS-Factor	LS-Factor	MRRTF	LS-Factor
pXRF + TA 5m	1	PM**	CaO***	CaO***	Cu**
	2	Fe	Mn	Zn	Profile curvature
	3	SiO_2	Al_2O_3	Ti	Zn
	4	VDCN	K_2O	PM	VDCN
	5	CNBL	SiO_2	Slope	MRRTF
TA 10 m	1	PM**	Plan curvature**	CNBL**	Profile curvature**
	2	VDCN	MRVBF	Slope	TWI
	3	DEM	LS-Factor	Flow accum	CNBL
	4	MRVBF	MRRTF	MRVBF	VDCN
	5	CNBL	SWI	Plan curvature	Aspect
pXRF + TA 10 m	1	VDCN**	CaO***	CaO***	Cu**
	2	Fe	Mn	Zn	Zn
	3	PM	Al_2O_3	Ti	Profile curvature
	4	SiO_2	Zn	PM	PM
	5	Flow accum	Plan curvature	Plan curvature	Ti

*A horizon data; ** B horizon data; *** A+B horizons data; P.M. = parent material; DEM = Digital elevation model; VDCN = Vertical distance to channel network; SWI = Saga wetness

index; MRVBF = Multiresolution index of valley bottom flatness; TWI = Topographic wetness index; MRRTF = Multiresolution index of ridge top flatness; Flow accum = Flow accumulation; CNBL= Channel network base level; R^2_{adj} = adjusted coefficient of determination.

Importance of covariates for prediction models

The combination of the five datasets (i - pXRF data, solely; ii - pXRF data + TA (5 m); iii - pXRF data + TA (10 m); iv – TA (5m), solely; v – TA (10 m), solely) evaluated for the prediction of the available contents of the four studied micronutrients for A, B, or A+B horizons with or without PM resulted in a total of 120 models. Using the same criteria of high R^2_{adj} and small RMSE and ME values, the best models obtained per dataset were selected. The combination of the predictions of the four micronutrient contents with the five datasets resulted in 20 models of which the five most important covariates for each model were defined by the random forest (Table 6).

In a general overview, the percentage of occurrence among the five most important covariates in models containing pXRF data appeared as follows: Zn (67%), CaO (50%), SiO_2 and Mn (42%), Fe and Al_2O_3 (33%), and K_2O (16%). Regarding the TA covariates, the percentages of occurrence among the models containing TA data were: VDCN (38%), CNBL (38%), MRVBF (31%), MRRTF (31%), plan curvature (31%), LS-Factor (25%), slope (25%), profile curvature (25%), TWI (13%), SWI (13%), and flow accumulation (13%). PM information occurred in 55% of the models. However, all the 20 best models presented in Table 6 used PMs as a covariate, regardless of whether this information was among the five most important covariates.

Fe, Mn, Cu, and Zn obtained by pXRF were always present within the five most important variables in models that used pXRF. For the prediction of available Fe, PM was the most important covariate, followed by pXRF Fe. Whereas, for available Mn, the

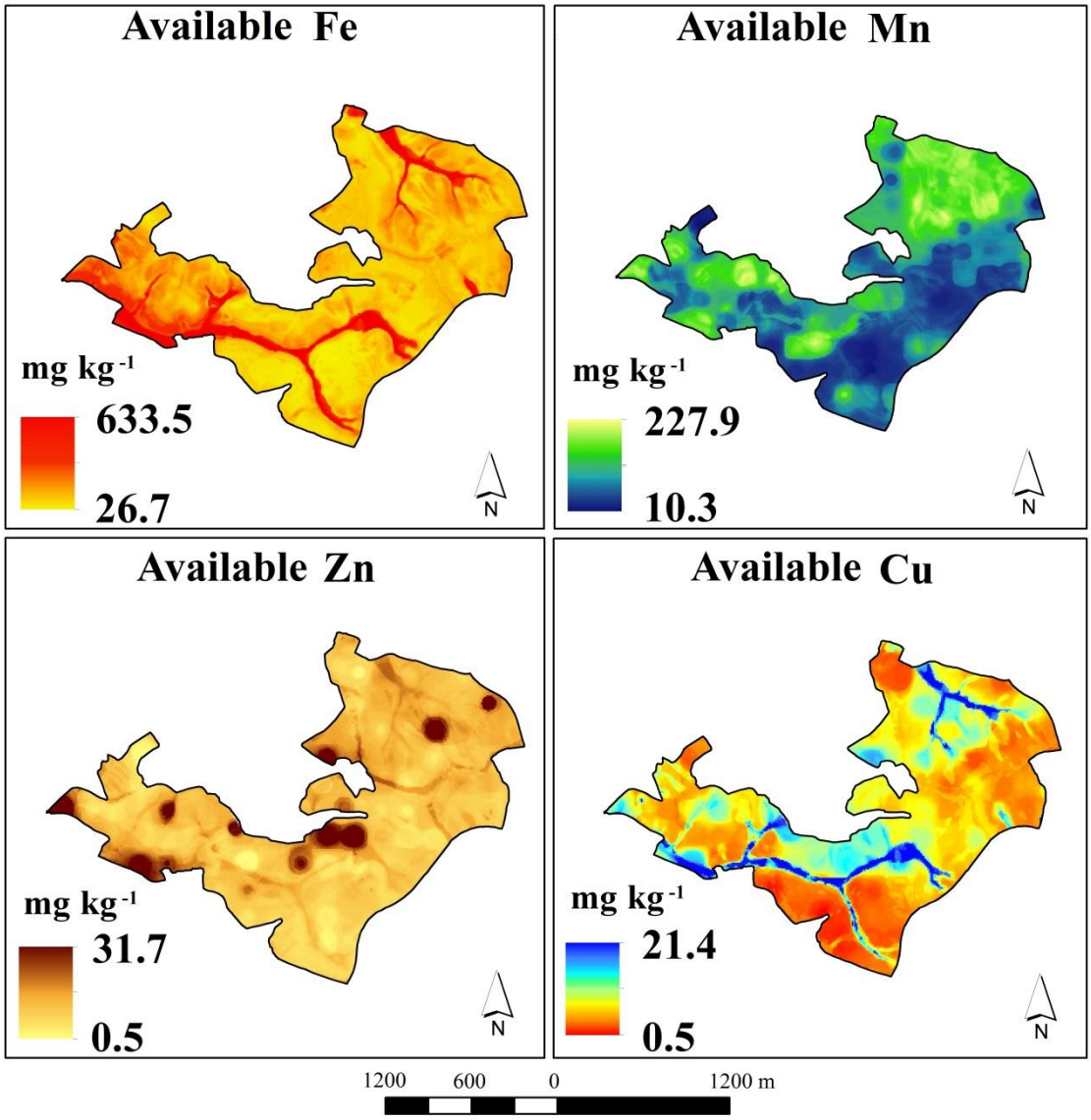

Fig. 4 Maps of available Fe, Mn, Zn and Cu contents obtained from the models of random forest algorithm for the study area in Brazilian soils

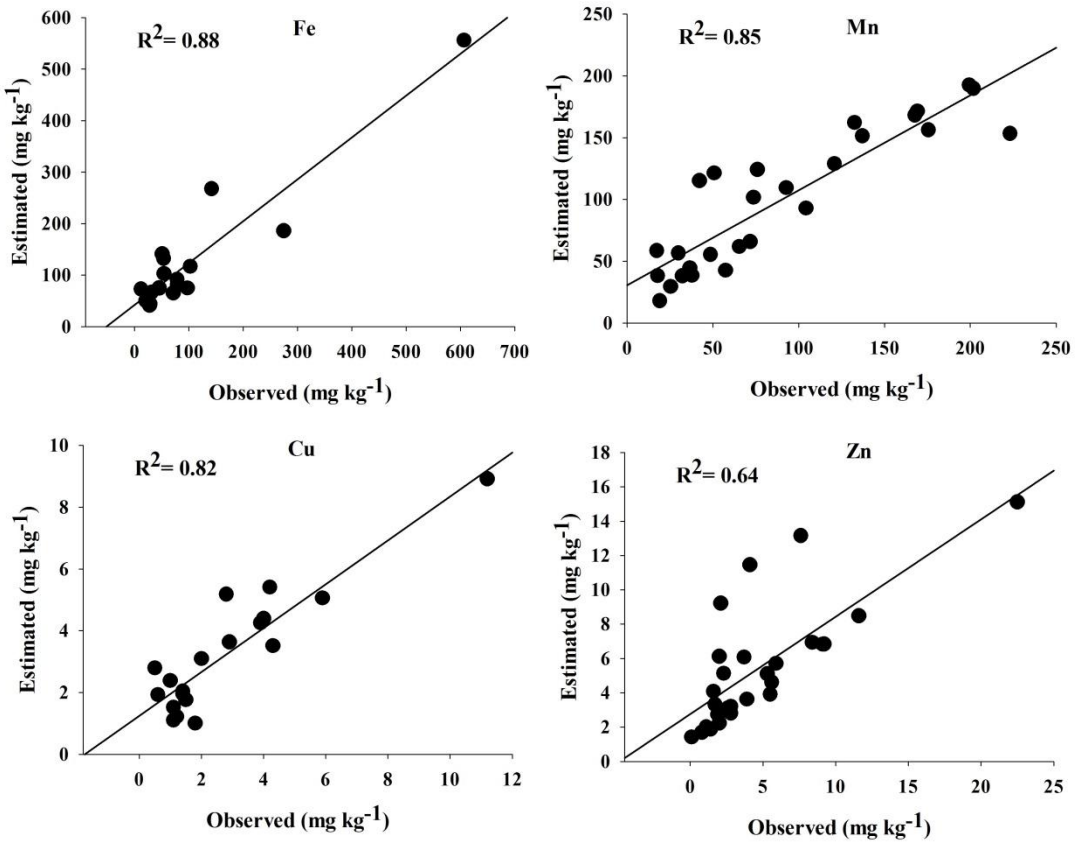

Fig. 5 A 1:1 relation between estimated and observed available Fe, Mn, Cu and Zn contents and the coefficient of determination R^2 for the validation dataset of spatial prediction for soil of the study area in Brazil

most important covariate was CaO, followed by pXRF Mn as the second most important covariate. Zn obtained by pXRF was the most important when pXRF data were used and the second most important in two models, after CaO. Cu by pXRF was the most important in all the three models that used pXRF data for predicting available Cu contents.

Spatial predictions of available Fe, Mn, Zn, and Cu

From the best models determined in the previous section, spatial predictions of available micronutrients were generated for the whole study area (Figure 4) based on

TA, PM, and spatialized elemental contents obtained by pXRF. The generated maps were validated with R^2, RMSE, and ME values using an independent set of samples (27 samples for A horizon and 19 samples for B horizon), making up 30% of total number of samples spread throughout the study area (Figure 1). Figure 5 shows the plots of estimated and observed available micronutrient contents. The validation of the spatial prediction resulted in an R^2 of 0.88, RMSE of 59.97 mg kg^{-1} and ME of 24.00 mg kg^{-1} for Fe; 0.85, 29.65 mg kg^{-1}, 9.70 mg kg^{-1} for Mn; 0.64, 3.11 mg kg^{-1}, 0.71 mg kg^{-1} for Zn; and 0.82, 1.17 mg kg^{-1}, 0.43 mg kg^{-1} for Cu, respectively.

These results indicate that the generated models can also be applied for spatial analysis, useful where intensive soil sampling and laboratory analyses is often unfeasible (Duda et al. 2017). Thus, it can contribute to differential soil management regarding fertilizer applications at varying rates, according to crop requirements.

CONCLUSIONS

The combination of pXRF data and TA covariates in addition to PM information allowed for accurate predictions of available Fe, Mn, Cu, and Zn through random forest algorithm in Brazilian soils. Many elements/compounds obtained by pXRF were key factors in the random forest algorithm for predictions of available Fe, Mn, Cu, and Zn; models were further improved when such data was associated with PM and TA data. Available Fe contents were better predicted using only TA data. Include of PM information improved the predictions. A pixel size of 10 m resolution yielded better results than 5 m resolution. High accuracy spatial prediction for Fe, Cu, Zn, and Mn can be useful for optimal land management via the combined use of proximal and remote sensors.

AKCNOWLEDGEMENTS

The authors would like to thank CNPq, CAPES and FAPEMIG for providing the financial support necessary for carrying out this work. The authors gratefully acknowledge the contributions of the BL Allen Endowment in Pedology at Texas Tech University in conducting this research.

REFERENCES

Adhikari, K., Kheir, R. B., Greve, M. B., Bøcher, P. K., Malone, B. P., Minasny, B., et al. (2013). High-Resolution 3-D Mapping of Soil Texture in Denmark. *Soil Science Society of America Journal*, *77*(3), 860. doi:10.2136/sssaj2012.0275

Akpa, S. I. C., Odeh, I. O. A., Bishop, T. F. A., & Hartemink, A. E. (2014). Digital mapping of soil particle-size fractions for Nigeria. *Soil Science Society of America Journal*, *78*(6), 1953–1966. doi:10.2136/sssaj2014.05.0202

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*(1), 11–25. doi:10.1590/S0100-06832014000100002

Arrouays, D., Lagacherie, P., & Hartemink, A. E. (2017). Digital soil mapping across the globe. *Geoderma Regional*, *9*, 1–4. doi:10.1016/j.geodrs.2017.03.002

Bhering, S. B., Chagas, C. da S., Junior, W. de C., Pereira, N. R., Filho, B. C., & Pinheiro, H. S. K. (2016). Mapeamento digital de areia, argila e carbono orgânico por modelos Random Forest sob diferentes resoluções espaciais. *Pesquisa Agropecuaria Brasileira*, *51*(9), 1359–1370. doi:10.1590/S0100-204X2016000900035

Bishop, T. F. A., Horta, A., & Karunaratne, S. B. (2015). Validation of digital soil maps at different spatial supports, *Geoderma*, *242*, 238–249.

Borghi, E., Avanzi, J. C., Bortolon, L., Luchiari Junior, A., & Bortolon, E. S. O. (2016). Adoption and Use of Precision Agriculture in Brazil: Perception of Growers and

Service Dealership. *Journal of Agricultural Science*, *8*(11), 89. doi:10.5539/jas.v8n11p89

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. doi:10.1023/A:1010933404324

Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and Regression Trees*. New York, NY: Chapman and Hall/CRC.

Camargo, F. A. D. O., Santos, G. D. A., & Zonta, E. (1999). Alterações eletroquímicas em solos inundados. *Ciência Rural*, *29*(1), 171–180. doi:10.1590/S0103-84781999000100032

Carvalho Junior, W., Calderano Filho, B., Chagas, C. D. S., Bhering, S. B., Pereira, N. R., & Pinheiro, H. S. K. (2016). Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. *Pesquisa Agropecuaria Brasileira*, *51*(9), 1428–1437. doi:10.1590/s0100-204x2016000900041

Cavazzi, S., Corstanje, R., Mayr, T., Hannam, J., & Fealy, R. (2013). Are fine resolution digital elevation models always the best choice in digital soil mapping? *Geoderma*, *195–196*, 111–121. doi:10.1016/j.geoderma.2012.11.020

Chagas, C. da S., de 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. doi:10.1016/j.catena.2016.01.001

Chakraborty, S., Man, T., Paulette, L., Deb, S., Li, B., Weindorf, D. C., & Frazier, M. (2017). Rapid assessment of smelter/mining soil contamination via portable X-ray fluorescence spectrometry and indicator kriging. *Geoderma*, *306*(June), 108–119. doi:10.1016/j.geoderma.2017.07.003

Collard, F., Kempen, B., Heuvelink, G. B. M., Saby, N. P. a., Richer de Forges, A. C., Lehmann, S., et al. (2014). Refining a reconnaissance soil map by calibrating regression models with data from the same map (Normandy, France). *Geoderma Regional*, *1*, 21–30. doi:10.1016/j.geodrs.2014.07.001

Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., et al. (2015).

System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. doi:10.5194/gmd-8-1991-2015

Curi, N., Silva, S. H. G., Poggere, G. C., & Menezes, M. D. de. (2017). *Mapeamento de solos e magnetismo no campus da UFLA como traçadores ambientais*. Lavras, MG: Editora UFLA.

Dantas, A. A. A., Carvalho, L. G. de, & Ferreira, E. (2007). Classificação e tendências climáticas em Lavras, MG. *Ciência e Agrotecnologia*, *31*(6), 1862–1866. doi:10.1590/S1413-70542007000600039

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. doi:10.1016/j.geoderma.2017.03.017

Fageria, N. k., & Stone, L. F. (2008). Micronutrient Deficiency Problems in South America. In B. J. Alloway (Ed.), *Micronutrient Deficiencies in Global Crop Production.* (pp. 245–266). Dordrecht: Springer. doi:https://doi.org/10.1007/978-1-4020-6860-7

Florinsky, I. ., Eilers, R. ., Manning, G. ., & Fuller, L. . (2002). Prediction of soil properties by digital terrain modelling. *Environmental Modelling & Software*, *17*(3), 295–311. doi:10.1016/S1364-8152(01)00067-6

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*(1), e0170478. doi:10.1371/journal.pone.0170478

Gessler, P. E., Moore, I. D., McKenzie, N. J., & Ryan, P. J. (1995). Soil-landscape modelling and spatial prediction of soil attributes. *International journal of geographical information systems*, *9*(4), 421–432. doi:10.1080/02693799508902047

Giasson, E., Clarke, R. T., Inda Junior, A. V., Merten, G. H., & Tornquist, C. G. (2006). Digital soil mapping using multiple logistic regression on terrain parameters in southern Brazil. *Scientia Agricola*. doi:10.1590/S0103-90162006000300008

Gray, J. M., Bishop, T. F. A., & Wilford, J. R. (2016). Lithology and soil relationships

for soil modelling and mapping. *Catena*, *147*, 429–440. doi:10.1016/j.catena.2016.07.045

Hengl, T. (2006). Finding the right pixel size. *Computers & Geosciences*, *32*(9), 1283–1298. doi:10.1016/j.cageo.2005.11.008

Hengl, T., Leenaars, J. G. B., Shepherd, K. D., Walsh, M. G., Heuvelink, G. B. M., Mamo, T., et al. (2017). Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutrient Cycling in Agroecosystems*, *109*(1), 77–102. doi:10.1007/s10705-017-9870-x

Heung, B., Ho, H. C., Zhang, J., Knudby, A., Bulmer, C. E., & Schmidt, M. G. (2016). An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma*, *265*, 62–77. doi:10.1016/j.geoderma.2015.11.014

Horta, A., Malone, B., Stockmann, U., Minasny, B., Bishop, T. F. A., McBratney, A. B., et al. (2015). Potential of integrated field spectroscopy and spatial analysis for enhanced assessment of soil contamination: A prospective review. *Geoderma*, *241–242*, 180–209. doi:10.1016/j.geoderma.2014.11.024

Hunt, A. M. W., & Speakman, R. J. (2015). Portable XRF analysis of archaeological sediments and ceramics. *Journal of Archaeological Science*, *53*, 628–638. doi:10.1016/j.jas.2014.11.031

Jackson, M. L. (1958). *Soil Chemical Analysis* (1st ed.). Englewood Cliffs: Prentice-Hall Inc.

Jakob, S., Gloaguen, R., & Laukamp, C. (2016). Remote sensing-based exploration of structurally-related mineralizations around Mount Isa, Queensland, Australia. *Remote Sensing*, *8*(5). doi:10.3390/rs8050358

Kämpf, N., Marques, J. J., & Curi, N. (2012). Mineralogia de Solos Brasileiros. In J. C. Ker, N. Curi, C. E. G. R. Schaefer, & P. Vidal-Torrado (Eds.), *Pedologia Fundamentos* (pp. 81-146). Viçosa, MG: SBCS.

Lacerda, M. P. C., Andrade, H., & Quemeneur, J. J. G. (2002). Pedogeoquimica em perfis de alteração na região de Lavras (MG): II - Elementos menores e elementos das terras raras. *Revista Brasileira de Ciência do Solo*, *26*(3), 87–102.

Lagacherie, P., & McBratney, A. B. (2006). Spatial soil information systems and spatial soil inference systems: perspectives for digital soil mapping. In P. Lagacherie, A. B. McBratney, & M. Voltz (Eds.), *Digital Soil Mapping: An Introductory Perspective* (pp. 3-22). Amsterdam: Elsevier Science.

Lecours, V., Devillers, R., Simms, A. E., Lucieer, V. L., & Brown, C. J. (2017). Towards a framework for terrain attribute selection in environmental studies. *Environmental Modelling and Software*, *89*, 19–30. doi:10.1016/j.envsoft.2016.11.027

Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, *2*, 18–22. doi:10.1023/A:1010933404324

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). doi:10.1016/bs.agron.2015.12.004

Marques Jr., J., Curi, N., & Lima, J. . (1992). Evolução diferenciada de Latossolo Vermelho-Amarelo e Latossolo Vermelho em função da litologia gnáissica na região de Lavras (MG). *R. Bras. Ci. Solo*, *16*, 235–240.

Mashimbye, Z. E., De Clercq, W. P., & Van Niekerk, A. (2014). An evaluation of digital elevation models (DEMs) for delineating land components. *Geoderma*, *213*, 312–319. doi:10.1016/j.geoderma.2013.08.023

Maynard, J. J., & Johnson, M. G. (2014). Scale-dependency of LiDAR derived terrain attributes in quantitative soil-landscape modeling: Effects of grid resolution vs. neighborhood extent. *Geoderma*, *230–231*, 29–40. doi:10.1016/j.geoderma.2014.03.021

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). *On digital soil mapping*. *Geoderma* (Vol. 117). doi:10.1016/S0016-7061(03)00223-4

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

Mehlich, A. (1953). *Determination of P, Ca, Mg, K, Na and NH_4*. Raleigh: North Carolina Soil Testing Division.

Menezes, M. D. de, Silva, S. H. G., Mello, C. R. de, Owens, P. R., & Curi, N. (2018). Knowledge-based digital soil mapping for predicting soil properties in two representative watersheds. *Scientia Agricola*, *75*(2), 144–153. doi:10.1590/1678-992x-2016-0097

Menezes, M. D., Silva, S. H. G., Mello, C. R., Owens, P. R., & Curi, N. (2016). Spatial prediction of soil properties in two contrasting physiographic regions in Brazil. *Scientia Agricola*, *73*(3), 274–285. doi:10.1590/0103-9016-2015-0071

Mokarram, M., & Hojati, M. (2017). Morphometric analysis of stream as one of resources for agricultural lands irrigation using high spatial resolution of digital elevation model (DEM). *Computers and Electronics in Agriculture*, *142*, 190–200. doi:10.1016/j.compag.2017.09.001

Moore, I. D., Gessler, P. E., Nielsen, G. A., & Peterson, G. A. (1993). Soil Attribute Prediction Using Terrain Analysis. *Soil Science Society of America Journal*, *57*(2), 443–452. doi:10.2136/sssaj1993.572NPb

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. doi:10.1016/j.geoderma.2014.12.025

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

Pelegrino, M. H. P., Silva, S. H. G., Menezes, M. D. de, Silva, E. da, Owens, P. R., & Curi, N. (2016). Mapping soils in two watersheds using legacy data and extrapolation for similar surrounding areas. *Ciência e Agrotecnologia*, *40*(5), 534–546. doi:10.1590/1413-70542016405011416

Penížek, V., Zádorová, T., Kodešová, R., & Vaněk, A. (2016). Influence of elevation data resolution on spatial prediction of colluvial soils in a luvisol region. *PLoS ONE*, *11*(11), 1–18. doi:10.1371/journal.pone.0165699

Peters, J., De Baets, B., Verhoest, N. E. C., Samson, R., Degroeve, S., Becker, P. De, & Huybrechts, W. (2007). Random forests as a tool for ecohydrological distribution modelling. *Ecological Modelling*, *207*(2–4), 304–318.

doi:10.1016/j.ecolmodel.2007.05.011

Reidinger, S., Ramsey, M. H., & Hartley, S. E. (2012). Rapid and accurate analyses of silicon and phosphorus in plants using a portable X-ray fluorescence spectrometer. *New Phytologist*, *195*(3), 699–706. doi:10.1111/j.1469-8137.2012.04179.x

Resende, M., Curi, N., Rezende, S. B., Corrêa, G. F., & Ker, J. C. (2014). *Pedologia: Base para distinção de ambientes* (6th ed.). Lavras, MG: Editora UFLA.

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*(3), 245–254. doi:10.1590/1413-70542017413000117

Ryan, J. G., Shervais, J. W., Li, Y., Reagan, M. K., Li, H. Y., Heaton, D., et al. (2017). Application of a handheld X-ray fluorescence spectrometer for real-time, high-density quantitative analysis of drilled igneous rocks and sediments during IODP Expedition 352. *Chemical Geology*, *451*, 55–66. doi:10.1016/j.chemgeo.2017.01.007

Shaetzl, R., & Anderson, S. (2005). *Soils: Genesis and geomorphology* (1st ed.). New York: Cambridge University Press.

Shangguan, W., Hengl, T., Jesus, J. M. de, Yuan, H., & Dai, Y. (2014). Mapping the global depth to bedrock for land suface modeling. *Journal of Advances in Modeling Earth Systems*, *6*, 513–526. doi:10.1002/2013MS000282.Received

Sharma, A., Weindorf, D. C., Man, T., Aldabaa, A. A. A., & 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, D. C., Wang, D., & 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, S. H. G., de Menezes, M. D., de Mello, C. R., de Góes, H. T. P., Owens, P. R., & Curi, N. (2016). Geomorphometric tool associated with soil types and properties spatial variability at watersheds under tropical conditions. *Scientia Agricola*, *73*(4). doi:10.1590/0103-9016-2015-0293

Silva, S. H. G., Poggere, G. C., de Menezes, M. D., Carvalho, G. S., Guilherme, L. R.

G., & 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. doi:10.3390/rs8080614

Silva, S. H. G., Poggere, G. C., Menezes, M. D., Carvalho, G. S., Guilherme, L. R. G., & Curi, N. (2016). Proximal sensing and digital terrain models applied to digital soil mapping and modeling of Brazilian Latosols (Oxisols). *Remote Sensing*, *8*(8), 614–35. doi:10.3390/rs8080614

Silva, S. H. G., Teixeira, A. F. S., Menezes, M. D., Guilherme, L. R. G., Moreira, F. M. de S., & Curi, N. (2017). Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence analyzer (pXRF). *Ciência e Agrotecnologia*, *41*(6), 648–664.

Smith, M. P., Zhu, a. X., Burt, J. E., & Stiles, C. (2006). The effects of DEM resolution and neighborhood size on digital soil survey. *Geoderma*, *137*(1–2), 58–69. doi:10.1016/j.geoderma.2006.07.002

Soil Survey Staff. (2014). *Keys to soil taxonomy* (12th ed.). USDA-NRCS. http://www.nrcs.usda.gov/Internet/FSE_DOCUMENTS/nrcs142p2_051546.pdf. Accessed 17 May 2016.

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. doi:10.1016/j.catena.2016.01.007

Thompson, J. A., Roecker, S., Grunwald, S., & Owens, P. R. (2012). *Digital Soil Mapping: Interactions with and Applications for Hydropedology*. *Hydropedology*. doi:10.1016/B978-0-12-386941-8.00021-6

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. doi:https://doi.org/10.1016/B978-0-12-802139-2.00001-9

Weindorf, D. C., Zhu, Y., Haggard, B., Lofton, J., Chakraborty, S., Bakr, N., et al. (2012). Enhanced pedon horizonation using portable X-ray fluorescence spectrometry. *Soil Science Society of America Journal*, *76*(2), 522–531. doi:10.2136/sssaj2011.0174

Zhu, H., Zhao, Y., Nan, F., Duan, Y., & Bi, R. (2016). Relative influence of soil chemistry and topography on soil available micronutrients by structural equation modeling, *Journal of Soil Science and Plant Nutrition, 16*(4), 1038–1051. doi:10.4067/S0718-95162016005000076

Zhu, Y., Weindorf, D. C., & 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

ARTIGO 2

Tropical Soil Nutrient Prediction Via Portable X-ray Fluorescence (pXRF) Spectrometry

Abbreviations: pXRF – portable X-ray fluorescence (pXRF); LR – linear regression; PR – polynomial regression; PwR – power regression; SMLR – stepwise multiple linear regression; RF – random forest; RMSE – root mean square error; MAE – mean absolute error; RPD – residual prediction deviation; DSM – digital soil mapping; NIST – National Institute of Standards and Technology; CS – check sample; OOB – out-of-bag; Var_{ex} – percentage of variance explained; Var_z – total variance of the variable; MSE_{oob} – mean of squared errors; IDW – inverse distance weighting; CV – coefficient of variation.

Abstract

Information on nutrient contents is fundamental for soil fertility management, contributing to optimal use of agricultural inputs, land suitability, and digital soil mapping techniques to enhance precision agriculture. In this sense, a representative sampling scheme is required to contemplate the intrinsic spatial variability of soils. In an attempt to address this variability and reduce costs, portable X-ray fluorescence (pXRF) spectrometry data was utilized for modeling and spatial prediction of exchangeable Ca^{2+} and available K^+ and P contents in tropical soils. Five different types of models were tested: linear regression (LR), polynomial regression (PR), power regression (PwR), stepwise multiple linear regression (SMLR), and random forest (RF) algorithm. Samples from soil A horizons were collected at 90 points in a regular grid design, and analyzed for exchangeable Ca^{2+} and available K^+ and P; pXRF scans analyses were also conducted on each sample. Regression models were calibrated and

afterwards validated with an independent set of samples through R^2, root mean square error (RMSE), mean absolute error (MAE), and residual prediction deviation (RPD). The best models were used to spatially predict nutrient contents. Validation data showed that exchangeable Ca^{2+} and available P were best estimated using PwR, and for available K^+ prediction, the SMLR outperformed the other models (R^2 = 0.28). For exchangeable Ca^{2+} and available P, respectively, the PwR model revealed R^2 of 0.89 and 0.52, and validation parameters: R^2 of 0.79 and 0.53; RMSE of 1.64 $cmol_c$ dm^{-3} and 7.07 mg dm^{-3}; MAE of 0.16 $cmol_c$ dm^{-3} and 0.79 mg dm^{-3}; and RPD of 1.61 and 1.45. When spatial prediction was performed, it achieved R^2 values of 0.80 for Ca^{2+} and 0.53 for P. For K^+ calibration, although low, only the SMLR model was able to produce some correlation, most likely by the erratic correlation between available and total contents since pXRF includes K^+ existing in the crystalline structure of some soil minerals. pXRF elemental data provided reliable local and spatial predictions of exchangeable Ca^{2+} and available P contents through PwR models, regardless of the high heterogeneity of the study area in terms of soil classes, parent materials, and land use.

Keywords: proximal sensor, regression analysis, soil fertility spatial prediction, random forest, digital soil mapping.

1. INTRODUCTION

Brazil is a major producers of grains, reaching records in soybean crop yields (Conab, 2018), and being the largest coffee producer in the world (Alvarenga and Arraes, 2017) producing several other agricultural and forestry products. In general, Brazilian soils are acidic and have low natural fertility, but they are highly suitable for mechanization (Trabaquini et al., 2015). Knowledge of the availability of plant nutrients

contents in the soil, in addition to other soil properties, is essential for proper soil management (Vasu et al., 2017).

Among the plant essential macronutrients, Ca, K, and P are of critical importance in tropical production systems. In Brazilian soils, Ca^{2+} is mainly provided by lime applications, help to correct soil pH (Hideo et al., 2016), improve root growth, and neutralize Al^{3+} (Lopes and Guilherme, 2016). Also, Ca^{2+} participates as a structural cellular wall component in plants (Malinovsky et al., 2014), among other roles. Conversely, K^{+} is essential in activating enzymatic functions in plants, improving vigor, and increasing resistance to pests and diseases (Wang et al., 2013). The natural source of K in soil is the weathering of primary and secondary K-bearing minerals, such as K-feldspars, biotite, muscovite, and vermiculite (Melo et al., 2004) or recycling of plant organic matter (Torres and Pereira, 2008). Commonly applied fertilizers are K-sulfates, K-nitrates, and K-chlorites (Bassoi, 2016). Although required in small quantities by plants, P is often the most limiting nutrient in tropical soil conditions (Lopes et al., 1982) due to its strong interaction with the solid fraction (Peluco et al., 2015), mainly Fe- and Al-oxides in acidic soils (Fink et al., 2014). P is vital to proper plant growth, respiration, and processes of energy storage (Laskar and Mukherjee, 2016).

The great demand for increased food productivity without exploring new areas under native vegetation has become a strong focus of agriculture in recent years (Qi et al., 2018). In this context, the assessment of spatial variability of soil nutrient contents is a viable way to identify and delineate critical nutrient deficiency zones (Vasu et al., 2017), driving the application of appropriate amounts of fertilizers and ameliorants where needed. However, soil nutrient information is generally scarce in tropical developing countries or exists merely at small scales.

To combat this limitation, digital soil mapping (DSM) has been increasingly applied worldwide (Arrouays et al., 2017). Through prediction models, DSM tools can provide inferences on the spatial and temporal variation of soil properties (Arrouays et al., 2014; Menezes et al., 2014; Pelletier et al., 2016) with considerable accuracy. For that, several techniques have been implemented for predicting variables of interest, including regression (Chagas et al., 2016; Lacarce et al., 2012; Souza et al., 2016), geostatistics (Gao et al., 2011; Shukla et al., 2016), fuzzy logic (Menezes et al., 2013; Shi et al., 2009; Silva et al., 2016a; Zhu et al., 2010), and machine learning techniques (Heung et al., 2016; Penížek et al., 2016), using data from remote sensing (Mulla, 2013; Silva et al., 2016b; Taylor et al., 2013; Zhu et al., 2016) and proximal sensors (Duda et al., 2017; Silva et al., 2017).

As a proximal sensor, the portable X-ray fluorescence (pXRF) spectrometry is able to quantify several chemical in seconds (Ribeiro et al., 2017; Weindorf et al., 2014). Thus, it has increasingly been used both in field and laboratory for a wide variety of studies (Stockmann et al., 2016; Ribeiro et al., 2018; Santana et al., 2018; Silva et al., 2018; Weindorf et al., 2012a) such as predicting soil texture (Silva et al., 2016b; Zhu et al., 2011), pH (Sharma et al., 2014), cation exchange capacity (Sharma et al., 2015) and other properties (Duda et al., 2017; Pelegrino et al., 2018; Silva et al., 2017; Suh et al., 2016). The elemental contents obtained by this methodology are considered to be total, inclusive of soil solution, soil mineral structure, fertilizers and conditioners, as well as elements strongly adsorbed to clay-sized particles.

The use of pXRF determined elements in soil has been extended to predict other soil properties owing to the fundamental chemistry reflected therein. Specifically, multiple linear regression, simple linear regression, random forest algorithm, and fuzzy logic (McLaren et al., 2012; O'Rourke et al., 2016; Sharma et al., 2014, 2015; Silva et al.,

2016b, 2017; Weindorf et al., 2012b; Zhu et al., 2016) have been used to predict the soil parameters of interest. Thus, pXRF analysis and development of robust prediction models can reduce the cost and time required for traditional laboratory analysis, as well as eliminate the chemical waste generated by them.

Adequate supply of Ca^{2+}, K^{+}, and P, as well as other nutrients in soil are essential to optimize plant growth. However, collecting and analyzing multiple samples requires substantive time and expense. Thus, this study sought to evaluate the efficacy of using pXRF for soil nutrient analysis in high leached, tropical soils. The objectives of the study were to: 1) establish predictive models for exchangeable Ca^{2+} and available K^{+} and P contents from pXRF elemental data by identifying the best regression model, and 2) generate spatial analysis of nutrient contents in tropical soils as a means of directing optimized corrective fertilizer application. We hypothesize that pXRF will be able to accurately predict nutrient contents in Brazilian soils with reasonable accuracy, reducing cost and time needed to adequately characterize soil fertility.

2. MATERIALS AND METHODS

2.1. Study area

The study area was located at the Federal University of Lavras, State of Minas Gerais, Brazil, between latitudes 7,651,207 and 7,653,478 mN and longitudes 501,962 and 503,957 mE, Zone 23K (Fig. 1). The area covers ~315 ha, encompassing native and planted forests, pasture, and agricultural crops. As per the Köppen classification, the climate of the region is Cwa with hot and humid summers and cold and dry winters, having an average annual temperature of 20.4°C and an average annual rainfall of 1,460 mm (Dantas et al., 2007).

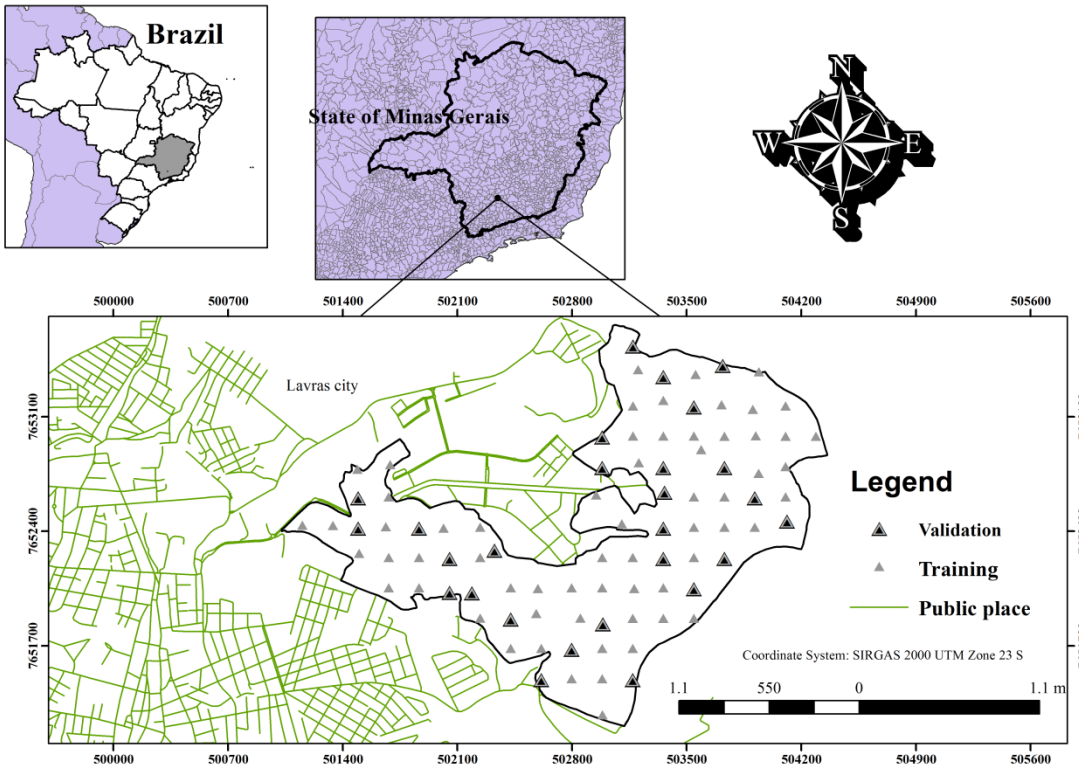

Fig. 1. Study area location at Lavras town, State of Minas Gerais, Brazil, and the points for modeling and validation of the predictions for available/exchangeable nutrients contents.

Soils of the area are derived from leucocratic and mesocratic granitic-gneiss (75%), gabbro intrusions (20%), and alluvial and coluvial-alluvial sediments (5%). The main soils are Oxisols, Ultisols, Inceptisols, Entisols, and Histosols per US Soil Taxonomy (Soil Survey Staff, 2014) (Curi et al., 2017).

2.2 Field sampling and laboratory analysis

Samples were collected from 90 points on the 0-20 cm layer using an auger (Schlindwein and Anghinoni, 2002). In areas featuring vegetative cover or organic material at the soil surface, such materials were gently scraped aside so as to only sample the mineral soil. The samples were air-dried, disaggregated to pass through a 2-mm sieve, and analyzed for exchangeable Ca^{2+} using KCl mol L^{-1} extractant (Raij and Bataglia, 1991) and available K^{+} and P using Mehlich-1 extractant (Mehlich, 1953).

Quantification of available K^+ was made on a Digimed atomic emission spectrometer model DM-63 (São Paulo, SP, Brazil); quantification of available P was made through molecular absorption spectrometry on a UV-visible spectrophotometer Metash model V-5000 (Shangai, China); and quantification of exchangeable Ca^{2+} was made on an atomic absorption spectrometer Perkin Elmer model AAnalyst 400 (Waltham, MA, USA).

One part of each soil sample was separated and scanned for 60s with a Bruker model S1 Titan LE pXRF (Billerica, MA, USA) using *Trace* mode (dual soil) and integrated GeoChem software. Prior to the analyses, the accuracy of the equipment was checked using certified reference materials NIST 2710a and NIST 2711a and a reference soil sample (CS) certified by the pXRF manufacturer. The elemental contents of those samples were compared to the contents obtained by pXRF through calculation of the recovery percentage (elemental content obtained by pXRF/certified elemental content) are (CS/2710a/2711a) (0 value indicates no reference value in the certified materials or no elemental detection by pXRF): Al_2O_3 (1.09/0.79/0.78); SiO_2 (1.05/0.93/0.93); Cl (0/0/0); K_2O (1.14/1.64/1.67); CaO (0/2.38/1.41); Ti (0/0/1.33); Mn (1.09/1.49/1.54); Fe (1.08/1.43/1.41); Cu (4.35/1.20/1.35); Zn (0/1.04/1.14); Zr (0/1.02/0); P_2O_5 (0/1.89/0.81); V (0/1.89/4.00); Ni (1.10/0/1.04); Cr (0/0/0.88).

2.3 Modeling

Prior to modeling, an exploratory data analysis was performed in order to avoid noise that could over- or underestimate predictions. Outliers were identified by the Cook's distance plot (Cook, 1977) and were removed *a priori*. Subsequently, the soil data were randomly separated into modeling (m) and validation (v) datasets as follows:

exchangeable Ca^{2+} (m = 62, v = 27); available P (m = 61; v = 26); and available K^{+} (m = 63; v = 27). The different number of samples per dataset was due to the removal of the outliers.

Linear regression (LR), 2nd degree polynomial regression (PR), power regression (PwR), stepwise multiple linear regression (SMLR), and random forest algorithm (RF) were executed for the predictions of soil nutrient contents. LR, PR, and PwR were generated considering the element/compound contents obtained by pXRF (CaO, K_2O, and P_2O_5) as independent variables and the exchangeable/available contents as dependent variables. For SMLR and RF, all 15 pXRF reported elements were used as independent variables to predict the exchangeable/available contents.

The SMLR was generated in SigmaPlot software (Systat Software, San Jose, CA), using the backward method. Initially, it incorporates all variables into the model followed by removal of the least important ones from the model according to a F-test at 5% of significance. The remaining variables compose the final SMLR model.

RF algorithm is an increasingly used ensemble machine learning technique (Belgiu and Drăgu, 2016). Using subsamples generated by bootstrap aggregation, several trees (a forest) are built, generating a final model. For each tree, predictors are randomly permuted in a out-of-bag (OOB) procedure sample, allowing for the calculations of the mean of squared errors (MSE_{oob}) (Eq. 1), the percentage of variance explained by the model (Var_{ex}) (Liaw and Wiener, 2002), and the importance of each predictor to the model: the greater the increases of errors when a predictor is removed from the trees, the more importan that variable is for the model. Var_{ex} is calculated as: $Var_{ex} = 1 - (MSE_{oob}/Var_z)$, in which Var_z is the total variance of the variable. The algorithm measures the predictions accuracy OOB procedure, reducing the variance of the trees by averaging them (Breiman, 2001). RF analysis was performed in RStudio software

version 1.1.419 (RStudio Team, 2016) applying the *randomForest* R package (Liaw and Wiener, 2015). The RF model was generated using the number of trees (*ntree* = 1000) and number of variables at each split (*mtry*) equals to one third of the number of variables (*mtry* = 5), as suggested by Liaw and Wiener (2002) where x_i is the observed value and x^{oob} is the mean of OOB prediction.

$$MSEoob = \frac{1}{n}\sum_{1}^{n}[x_i - x_i^{OOB}]^2 \quad (1)$$

All the generated models were validated with an independent data set. For the assessment of the best model, the following indexes were used: coefficient of determination (R^2), mean absolute error (MAE) (Eq. 2), root mean square error (RMSE) (Eq.3), and the residual prediction deviation (RPD), which is defined as the standard deviation of observed values divided by the RMSE, providing a metric of model validity that is more easily comparable across models. According to Chang et al. (2001), RPD values greater than 2.0 indicate good predictive models, values between 1.4 and 2.0 indicate fair models, and RPD values smaller than 1.4 indicate poor predictive models. The formulae for MAE and RMSE are given as:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|X_{obs,i} - X_{model,i}\right| \quad (2)$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(X_{obs,i} - X_{model,i})^2}{n}} \quad (3)$$

where X_{obs} represents observed values in laboratory analysis and X_{model} represents estimated values from regressions using pXRF elemental data at time/place *i* in *n* observations points.

2.4 Spatial analysis

After the best predictive model was determined for each exchangeable/available nutrient, the spatial prediction in a 10 m grid cell resolution was performed through the software ArcGIS 10.3 (ESRI, The Redlands, CA, USA) in order to provide maps of nutrient contents. Since pXRF generates discreet information (at places where samples were collected), these pXRF results were first spatialized across the entire area in order to generate a continuous dataset of all the elemental contents obtained by pXRF. The interpolation method used was the inverse distance weighting (IDW) (Eq. 4), after testing ordinary kriging, whose results were less accurate than IDW. Thus, the best model for each soil nutrient content was applied to such spatial information, generating a final map of the predicted nutrient contents. Finally, such maps were also validated according to the aforementioned statistical parameters. The formula applied for IDW is given as:

$$\hat{E} = \frac{\sum_{i=1}^{n} \frac{E}{d_i^p}}{\sum_{i=1}^{n} \frac{1}{d_i^p}} \quad (4)$$

where $\hat{E}$ is the value to be estimated, E is the value obtained from pXRF, d_i^p is the distance between two sampling points, n is the number of sampling points and p is a power parameter.

3. RESULTS AND DISCUSSION

3.1. Descriptive statistics

Table 1 presents the laboratory results obtained for the model calibration and validation datasets. The mean contents of the evaluated nutrients were interpreted according to (Ribeiro et al., 1999), including five categories: *Very low; Low; Medium; High; and Very high.* On average, the exchangeable Ca^{2+} and available K^{+} contents in the soils of the study area were *very high*. However, coefficient of variation (CV) values demonstrate the variability of such data across the area.

Available P in tropical soils are strongly influenced by soil texture and mineralogy (Fink et al., 2014; Motta et al., 2002). According to Curi et al. (2017), ~10% of the area contains very clayey soils (> 60% clay) and 67% clayey soils (35 – 60% clay); most are predominantly Oxisols and Ultisols. By contrast, 20% are medium-textured soils (15 – 35% clay), expressed mostly as Inceptisols and lowland soils. The average available P was *low* in clayey and medium-textured soils, and *medium* in very clayey soils. For the average of the validation set, the contents were considered *high* in very clayey soils, *medium* in clayey soils, and *low* in medium-textured soils.

Table 1. Descriptive statistics of the samples used in the modeling and prediction of soil properties in Brazil.

Soil nutrient	Calibration set					Validation set				
Content	Max	Min	Mean	SD	CV %	Max	Min	Mean	SD	CV %
K^{+} (mg dm^{-3})	392.0	48.0	157.1	74.5	47.4	352.0	40.0	132.7	76.1	57.4
P (mg dm^{-3})	40.0	0.8	7.2	24.1	109.4	40.9	1.4	8.6	10.3	119.0
Ca^{2+} ($cmol_c$ dm^{-3})	9.4	0.8	4.2	1.9	43.6	12.2	1.1	4.1	2.6	64.5

Max = maximum value; Min = minimum value; SD = standard deviation; CV = coefficient of variation.

The range of the elemental contents obtained through pXRF (Fig. 2) expresses the soil chemical variability caused by their parent materials, soil use and management, soil weathering degree and landscape position.

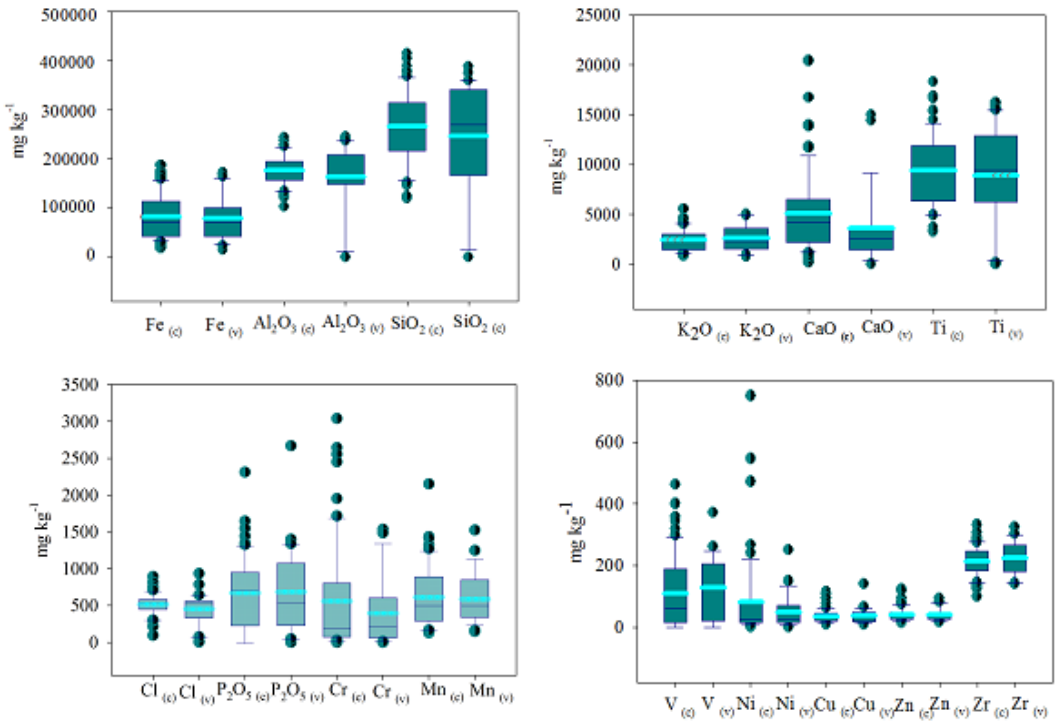

Fig 2. Variation of calibration ($_c$) and validation ($_v$) sets for each element induced by portable x-ray fluorescence (pXRF).

Gabbro-derived soils contributed to the greater contents of elements such as Fe and Ti, while granite-gneiss-derived soils presented greater contents of SiO_2 and K_2O. Silva et al. (2016b) compared the pXRF results obtained for Oxisols derived from gneiss and gabbro in Brazil and found a similar trend regarding these elemental contents. The high variability of some elements, such as CaO, K_2O, and P_2O_5 is a reflection not only of the parent materials and the varying weathering degrees of the soils, but also of the land use and management of the area, encompassing places both under agricultural cultivation, where fertilizers and other amendments have been frequently applied, and under native vegetation, with soils containing low contents of such elements due to their high weathering degree.

3.2 Regression models

Relative to laboratory derived data, Table 2 presents the regression models of pXRF predicted nutrient contents via LR, PR, PwR, and SMLR and their respective R^2.

Table 2. Adjusted models using portable X-ray fluorescence (pXRF) spectrometry data to predict soil nutrient contents in Brazil.

Nutrient	Method	Model	R^2
Ca^{2+}	LR	Ca^{2+}= 1.92*5E-4CaO	0.81
	PR	Ca^{2+}= 1.14 + 8E-4CaO -2.26E-8CaO^2	0.86
	PwR	Ca^{2+} = 0.035$CaO^{0.57}$	0.89
	SMLR	Ca^{2+}= 1 + 4.7E-4*CaO – 1.4E-4*Ti + 5.2E-3*V + 6.2E-4*Mn + 6.5E-3*Zr	0.85
K^+	LR	K^+= 155.25 + 3.6E-4K_2O	0.00
	PR	K^+= 154.4 + 0.000K_2O + 1E-8K_2O^2	0.00
	PwR	K^+= 155.64 + 2.51E-5$K_2O^{1.26}$	0.00
	SMLR	K^+= 81.70 + 0.04*P_2O_5 + 5.7E-4*Fe	0.22
P	LR	P = 0.39 + 1.1E-5P_2O_5	0.42
	PR	P = 4.05 – 0.01 P_2O_5 + 1.66E-5 $P_2O_5^2$	0.50
	PwR	P = 3.66 + 2.66E-12 $P_2O_5^{4.1}$	0.52
	SMLR	P= 18.78 + 0.01*P_2O_5 – 7.5E-5*SiO_2+ 0.03*V -1.5E-4*Fe + 0.04*Zr	0.58

LR = Linear regression; PR = 2nd degree polynomial regression; PwR = Power regression; SMLR = Stepwise multiple linear regression.

Comparing these models, the explanatory power of pXRF-sensed CaO and P_2O_5 for predicting exchangeable/available Ca^{2+} and P contents in all the regression models was evident. Adequate results for Ca^{2+} was obtained since the total contents do not differ much from the exchangeable contents obtained in tropical soils (Teixeira et al., 2018; Towett et al., 2015), where the main source of Ca in such soils tends to be liming and gypsum application (Lopes and Guilherme, 2016). Conversely, none of the regression

models was able to establish adequate correlations between the total K_2O contents and available K^+. Total K_2O contents may be in different forms in the soil, including solution-K, exchangeable-K, fixed-K, and structural-K (Meena et al., 2014).

RF alorithm provides MSE_{oob}, Var_{ex}, and the importance of each predictor variable as a way to show the performance of the models. The greater the Var_{ex} and the lower the MSE_{oob}, the better the model. The MSE_{oob} and Var_{exp} values computed by RF algorithm were: Ca^{2+} (1.01; 70.33%); K^+ (5571.22; -6.27%); and P (37.70; 37.96%). The high Var_{ex} and low MSR for Ca^{2+} and P indicate that models for predicting exchangeable Ca^{2+} and available P had adequate fit, while adequate available K^+ modeling through RF was not possible, similar to the aforementioned regression models (Table 3). Pelegrino et al. (2018) used RF to predict the available contents of micronutrients from pXRF data in Brazilian soils and found that models with Var_{ex} grater than c.a. 40% provided accurate predictions.

For exchangeable Ca^{2+}, the most important predictor variables in prediction were CaO, Al_2O_3, Mn, and P_2O_5 (Fig. 3). For available K^+ modeling, V, Fe, Ti, and K_2O were

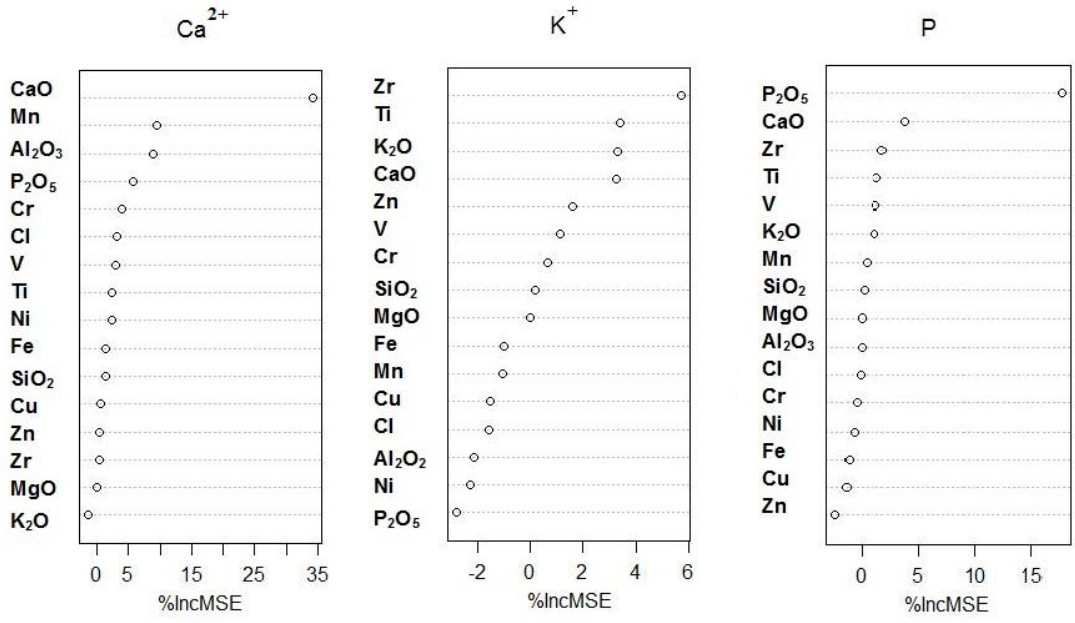

Fig. 3. Random forest variables importance plots based on the mean decreace accuracy

(%IncMSE). Notably, %IncMSE denotes the increase in mean squarred error of predictions (estimated with OOB-CV) as a result of permutation of variable j.

the major contributors to the model, while in the available P model, P_2O_5, Ni, Mn, and K_2O contributed more. For the predictions of exchangeable Ca^{2+} and available P, the most important variables, by far, were CaO and P_2O_5. Since the mineralogy of Brazilian soils lacks Ca-bearing minerals (Brinatti et al., 2010), which could be source of Ca^{2+} to plants through weathering, the contents obtained by pXRF tend to have a strong correlation with its exchangeable form. For P, although there is a similar lack of P-bearing minerals in Brazilian soils, this element is commonly adsorbed by clay minerals in acidic soils, which are common in Brazil (the mean pH in water of the soil in the study area was 5.7), rendering it unavailable to plants. Thus, P contents obtained by pXRF are not strongly correlated to available P contents, worsening such predictions.

3.3 Accuracy of the regressions models

In Table 3, the evaluation of the accuracy and errors of the models are presented. The RMSE, MAE, RPD, and R^2 measurements were used as criteria for selecting the best models. All adjustments obtained for exchangeable Ca^{2+} were adequate, with the best results found for PwR. Available P predictions showed intermediate performance, with better results also using PwR, whereas no model was able to efficiently predict available K^+.

Table 3 . Evaluation of the regression models predicting soil nutrient contents in Brazil.

Soil nutrient	LR	PR	PwR	SMLR	RF
			RMSE		
			$cmol_c\ dm^{-3}$		
Exchangeable Ca^{2+}	10.02	9.78	1.64	4.23	1.55

			mg dm^{-3}		
Available K^+	78.28	78.25	78.28	68.57	73.63
Available P	8.17	7.23	7.07	9.64	8.83
			MAE		
Exchangeable Ca^{2+}	1.99	2.62	0.16	0.98	0.22
Available K^+	23.54	23.47	23.51	24.75	25.4
Available P	1.30	0.77	0.73	2.44	0.1
			RPD		
Exchangeable Ca^{2+}	0.26	0.27	1.61	0.63	1.71
Available K^+	0.97	0.97	0.97	1.11	1.03
Available P	1.25	1.42	1.45	1.06	1.16
			R^2		
Exchangeable Ca^{2+}	0.63	0.20	0.79	0.63	0.77
Available K^+	0.00	0.00	0.00	0.28	0.15
Available P	0.42	0.50	0.53	0.17	0.20

RMSE = Root mean square error; MAE = mean absolute error; RPD = Residual prediction deviation, LR = Linear regression; PR = Polynomial regression; PwR = Power regression; SMLR = Stepwise multiple linear regression; RF = Random Forest; R^2 = Coefficient of determination.

Sources of soil K_2O content may involve K-feldspars and more commonly K-micas (Araujo et al., 2014) in addition to a range of secondary minerals, where K^+ can be strongly fixed in the interstitial clay mineral layers, rendering its plant availability marginal (O'Rourke et al. 2016). K is also present in soil in kaolinite with mica residual layers preserved in the mineral structure (Melo et al., 2004, 2005). In order to establish reliable correlations with available K contents, further evaluation with other auxiliary variables (e.g., clay content, organic matter content) and different approaches are necessary. Similar methodologies do not necessarily work well for the same predictions under different tropical environmental conditions. Silva et al. (2017) used pXRF data in an area with variable land uses, but more homogeneous area in terms of soil classes, and

obtained better models and validation by RF than SMLR for predicting soil properties. Their results for available soil K^+ predictions yielded $R^2 = 0.67$, RMSE = 52.67 mg dm^{-3}, and ME = - 23.00.

3.4 Spatial analysis

Optimal models (PwR for both exchangeable Ca^{2+} and available P) (Table 3) were used for spatial prediction by applying the prediction models to the raster calculator tool in ArcGIS 10.3 software (ESRI, The Redlands, CA, USA) (Fig. 4). Due to the low predictive power of the available K^+ model, it was not spatially rendered.

According to the generated maps, the current quantities of exchangeable Ca^{2+} are either high or very high, while the available P was shown to be very low following the soil fertility classification suggested by the 5^{th} approach of recommendations for the use of fertilizers and amendments in State of Minas Gerais, Brazil (Ribeiro et al., 1999) (Figs 4a and 4b). Some areas that showed greater contents of these nutrients are related to crop or experimental areas, where liming and fertilizers applications were performed. In the spatial predictions, the exchangeable Ca^{2+} contents in soils were, on average, *high* and *very high* (Fig. 4c). For available P, most soils of the area were classified as *very low* or *low* (Fig. 4d). The areas indicated on the map with *medium* and *high* contents correspond to experimental areas of coffee, corn, and orchard, reflecting the P-amendment applications.

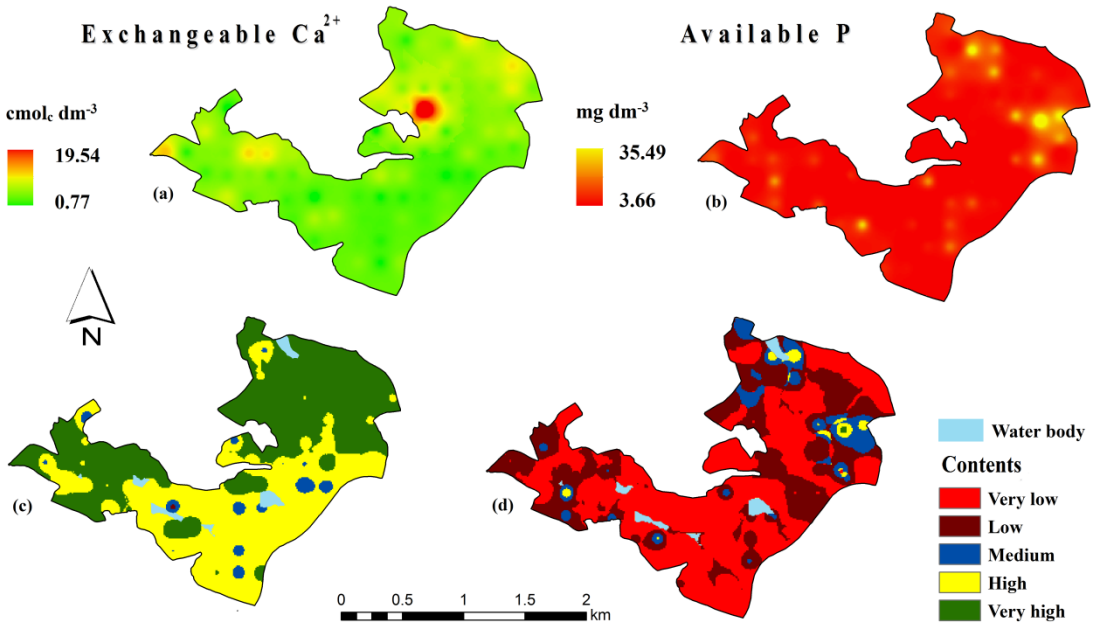

Fig. 4. Prediction of soil nutrient contents (a and b) and their classes of contents (c and d) based on the 5th approach of recommendations for the use of correctives and fertilizers in State of Minas Gerais, Brazil (Ribeiro et al., 1999).

A separate dataset was used in the validation of the generated maps of nutrient availability (Fig. 5). The spatial validation yielded R^2 = 0.80 and RMSE = 1.63 $cmol_c$ dm^{-3} for exchangeable Ca^{2+}, and R^2 = 0.53 and RMSE = 6.92 mg dm^{-3} for available P in soil

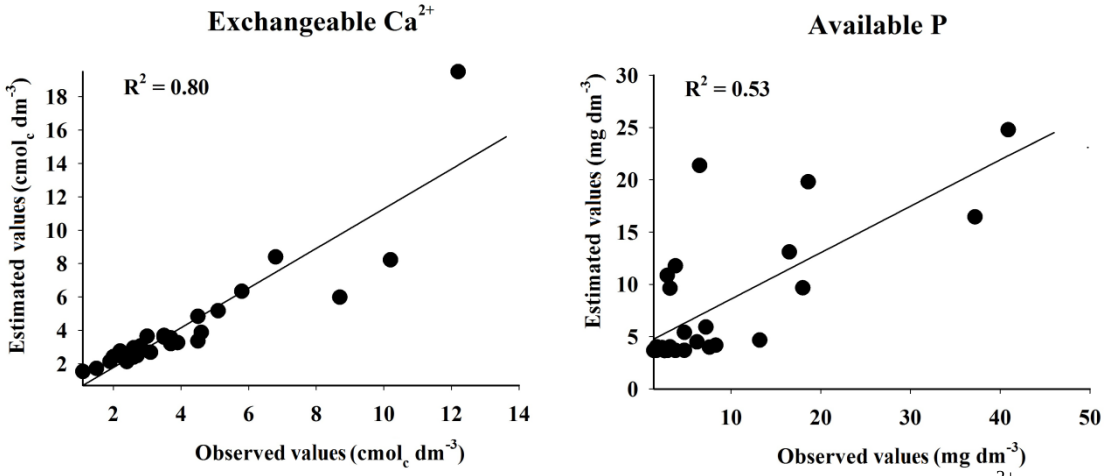

Fig. 5. Validation of spatialized models using power regression for exchangeable Ca^{2+} and available P contents in Brazilian soils

The best performance of the models for exchangeable Ca^{2+} was likely due to the lack of Ca-bearing minerals in most tropical soils. Contrariwise, in the case of P, Matos-Moreira et al. (2017) found large contributions of agricultural practices for predicting available P, identifying the main environmental, soil and land-use factors that affect its spatial distribution.

The elemental data generated via pXRF and modeled with simple regressions successfully predicted exchangeable Ca^{2+} and available P with reasonable accuracy, especially up to 8 $cmol_c$ dm^{-3}. Spatial analysis built upon such data was able to quickly identify the most nutrient deficient areas of the tested field, offering insight into the optimal locations for fertilizer application. The premise of using pXRF for rapid soil nutrient analysis in tropical soils has been soundly established by this study. Future work should evaluate the possibility of improving predictive models further via the inclusion of simple auxiliary input data such as soil texture (e.g., sand, silt, clay content), organic matter, etc.

4. CONCLUSIONS

PwR models provided the best fit and accurate estimates of exchangeable Ca^{2+} and available P from pXRF data for soil A horizon samples both punctually and spatially. Accurate modeling and predictions of available K^{+} from pXRF data was not possible. The pXRF data can be used to generate models to predict results of traditional laboratory analysis, allowing for reducing costs, time, and the production of chemical residues from such analyses. Also, it contributes to easily gather more soil data, allowing for spatially represent such nutrients availability to the whole area of interest with more details, especially for exchangeable Ca^{2+}, regardless of the variability of the area in terms of land uses, management, soil classes, and parent materials. Future efforts are needed to improve the results for predicting available K^{+}.

5. ACKNOWLEDGEMENTS

The authors would like to thank CNPq, CAPES and FAPEMIG for providing the financial support necessary for carrying out this study. The authors gratefully acknowledge the BL Allen Endowment in Pedology at Texas Tech University in conducting this research.

6. REFERENCES

Alvarenga, R.P., Arraes, N.A.M., 2017. Certificação fairtrade na cafeicultura brasileira: análises e perspectivas. Coffee Sci. 12, 124–147. https://doi.org/10.25186/cs.v12i1.1222

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. Rev.

Bras. Cienc. do Solo 38, 11–25. https://doi.org/10.1590/S0100-06832014000100002

Arrouays, D., Grundy, M.G., Hartemink, A.E., Hempel, J.W., Heuvelink, G.B.M., Hong, S.Y., Lagacherie, P., Lelyk, G., McBratney, A.B., McKenzie, N.J., Mendonca-Santos, M. d. L., Minasny, B., Montanarella, L., Odeh, I.O.A., Sanchez, P. A., Thompson, J. A., Zhang, G.-L., 2014. GlobalSoilMap: Toward a Fine-Resolution Global Grid of Soil Properties. Adv. Agron. 125, 93–134. https://doi.org/10.1016/B978-0-12-800137-0.00003-0

Arrouays, D., Lagacherie, P., Hartemink, A.E., 2017. Digital soil mapping across the globe. Geoderma Reg. 9, 1–4. https://doi.org/10.1016/j.geodrs.2017.03.002

Bassoi, L.H., 2016. Nitrate and potassium concentration in fertigated soil cultivated with wine vines. Ciência e Agrotecnologia. 40, 305–316. http://dx.doi.org/10.1590/1413-70542016403042215

Belgiu, M., Drăgu, L., 2016. Random forest in remote sensing: A review of applications and future directions. ISPRS J. Photogramm. Remote Sens. 114, 24–31. https://doi.org/10.1016/j.isprsjprs.2016.01.011

Breiman, L., 2001. Random Forests. Mach. Learn. 45, 5–32.

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

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

Chang, C., Laird, D., Mausbach, M.J., Hurburgh Jr, C.R., 2001. Near-Infrared

Reflectance Spectroscopy – Principal Components Regression Analyses of Soil Properties. Soil Sci. Soc. Am. J. 65, 480–490. https://doi.org/10.2136/sssaj2001.652480x.

Conab – Companhia Nacional de Abastecimento. Acompanhamento da safra brasileiras de grãos. v. 12 Safra 2017/18 - Décimo segundo levantamento, Brasília, p. 1-148, setembro 2018.

Cook, R.D., 1977. Detection of Influential Observation in Linear Regression. Technometrics 19, 15–18.

Curi, N., Silva, S. H. G., Poggere, G. C., Menezes, M. D. de., 2017. Mapeamento de solos e magnetismo no campus da UFLA como traçadores ambientais, first ed. Editora UFLA, Lavras.

Dantas, A.A.A., Carvalho, L.G. de, Ferreira, E., 2007. Classificação e tendências climáticas em Lavras, MG. Ciência e Agrotecnologia 31, 1862–1866. https://doi.org/10.1590/S1413-70542007000600039

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. https://doi.org/10.1016/j.geoderma.2017.03.017

Fink, J.R., Inda, A.V., Bayer, C., Torrent, J., Barrón, V., 2014. Mineralogy and phosphorus adsorption in soils of south and central-west Brazil under conventional and no-tillage systems. Acta Sci. Agron. 36, 379. https://doi.org/10.4025/actasciagron.v36i3.17937

Gao, Y., Gao, J., Chen, J., 2011. Spatial variation of surface soil available phosphorous and its relation with environmental factors in the Chaohu Lake watershed. Int. J. Environ. Res. Public Health 8, 3299–3317. https://doi.org/10.3390/ijerph8083299

Heung, B., Ho, H.C., Zhang, J., Knudby, A., Bulmer, C.E., Schmidt, M.G., 2016. An

overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265, 62–77. https://doi.org/10.1016/j.geoderma.2015.11.014

Hideo, C., Alexandre, C., Crusciol, C., Ferrari, J., Spadotti, G., Castro, A., 2016. Residual effects of superficial liming on tropical soil under no-tillage system. Pesqui. Agropecu. Bras. 51, 1633–1642. https://doi.org/10.1590/S0100-204X2016000900063

Lacarce, E., Saby, N.P.A., Martin, M.P., Marchant, B.P., Boulonne, L., Meersmans, J., Jolivet, C., Bispo, A., Arrouays, D., 2012. Mapping soil Pb stocks and availability in mainland France combining regression trees with robust geostatistics. Geoderma 170, 359–368. https://doi.org/10.1016/j.geoderma.2011.11.014

Laskar, S., Mukherjee, S., 2016. Optical sensing methods for assessment of soil macro-nutrients and other properties for application in precision agriculture: A review. ADBU J. Eng. Technol, 4, 206-209.

Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R news 2, 18–22. https://doi.org/10.1177/154405910408300516

Liaw, A., Wiener, M., 2015. Package “randomForest”. R Dev. Core Team. Available online at https://cran.r-project.org/web/packages/randomforest/randomforest.pdf. verified 20 dec. 2018.

Lopes, A.S., Vasconcellos, C.A., Novais, R.F. Adubação fosfatada em algumas culturas de Minas Gerais, Espírito Santo e Rio de Janeiro. In: Oliveira, A.J. de; Lourenço, S.; Goedert, W.J. Adubação fosfatada no Brasil. Brasília, EMBRAPA-DID, 1982. p.137-200. (EMBRAPA-DID, Documentos, 21).

Lopes, A.S., Guimarães, L.R.G., 2016. A career perspective on soil management in the Cerrado region of Brazil. Adv. Agron. 137, 1–72. https://doi.org/10.1016/bs.agron.2015.12.004

Malinovsky, F.G., Fangel, J.U., Willats, W.G.T., 2014. The role of the cell wall in plant immunity. Front. Plant Sci. 5, 1–12. https://doi.org/10.3389/fpls.2014.00178

Maluf, H.J.G.M., Silva, C.A., Curi, N., Norton, L.D., Rosa, S.D., 2017. Adsorption and availability of phosphorus in response to humic acid rates in soils limed with $CaCO_3$ or $MgCO_3$. Ciência e Agrotecnologia 42, 7–20. https://doi.org/10.1590/1413-70542018421014518

Matos-Moreira, M., Lemercier, B., Dupas, R., Michot, D., Viaud, V., Akkal-Corfini, N., Louis, B., Gascuel-Odoux, C., 2017. High-resolution mapping of soil phosphorus concentration in agricultural landscapes with readily available or detailed survey data. Eur. J. Soil Sci. 68, 281-294. doi: 10.1111/ejss.12420

McLaren, T.I., Guppy, C.N., Tighe, M.K., Forster, N., Grave, P., Lisle, L.M., Bennett, J.W., 2012. Rapid, Nondestructive Total Elemental Analysis of Vertisol Soils using Portable X-ray Fluorescence. Soil Sci. Soc. Am. J. 76, 1436. https://doi.org/10.2136/sssaj2011.0354

Meena, V.S., Maurya, B.R., Verma, J.P., 2014. Does a rhizospheric microorganism enhance K+ availability in agricultural soils? Microbiol. Res. 169, 337–347. https://doi.org/10.1016/j.micres.2013.09.003

Mehlich, A., 1953. Determination of P, Ca, Mg, K, Na and NH_4, North Carolina Soil Testing Division, Raleigh.

Melo, V.F., Ribeiro, A.N., Maschio, P.A., Corrêa, G.F., Lima, V.C., 2004. Mineralogia e formas de potássio e magnésio em diferentes classes de pesos e tamanhos da fração areia de solos do Triângulo Mineiro Rev. Bras. Ciênc. Solo 28, 219-231. http://dx.doi.org/10.1590/S0100-06832004000200001.

Melo, V.F., Corrêa, G.F., Ribeiro, A.N., Maschio, P.A., 2005. Cinética de liberação de potássio e magnésio pelos minerais da fração argila de solos do Triângulo Mineiro. Rev. Bras. Ciênc. Solo 29, 533–545. http://dx.doi.org/10.1590/S0100-06832005000400006

Menezes, M.D.de., Silva, S.H.G., Mello, C.R. de., Owens, P.R., Curi, N., 2014. Solum depth spatial prediction comparing conventional with knowledge-based digital soil mapping approaches. Sci. Agric. 71, 316-323. https://doi.org/10.1590/0103-9016-2013-0416

Menezes, M.D. de, Silva, S.H.G., Owens, P.R., Curi, N., 2013. Digital soil mapping approach based on fuzzy logic and field expert knowledge. Ciência e Agrotecnologia 37, 287–298. https://doi.org/10.1590/S1413-70542013000400001

Motta, P.E.F., Curi, N., Siqueira, J.O., Van Raij, B., Furtini Neto, A.E., Lima, J.M., 2002. Adsorção e formas de fósforo em Latossolos: influência da mineralogia e histórico de uso. Rev. Bras. Ciência do Solo 26, 349–359. https://doi.org/10.1590/S0100-06832002000200008

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosyst. Eng. 114, 358–371. https://doi.org/10.1016/j.biosystemseng.2012.08.009

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 Sci. Soc. Am. J. 80, 888. https://doi.org/10.2136/sssaj2015.10.0361

Pelegrino, M.H.P., Weindorf, D.C., Silva, S.H.G., de Menezes, M.D., Poggere, G.C., Guilherme, L.R.G., Curi, N., 2018. Synthesis of proximal sensing, terrain analysis, and parent material information for available micronutrient prediction in tropical

soils. Precis. Agric. 19, 1–21. https://doi.org/10.1007/s11119-018-9608-z

Pelletier, C., Valero, S., Inglada, J., Champion, N., Dedieu, G., 2016. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sens. Environ. 187, 156–168. https://doi.org/10.1016/j.rse.2016.10.010

Peluco, R.G., Júnior, J.M., Siqueira, D.S., Pereira, G.T., Barbosa, R.S., Teixeira, D. de B., 2015. Mapeamento do fósforo adsorvido por meio da cor e da suscetibilidade magnética do solo. Pesqui. Agropecu. Bras. 50, 259–266. https://doi.org/10.1590/S0100-204X2015000300010

Penížek, V., Zádorová, T., Kodešová, R., Vaněk, A., 2016. Influence of elevation data resolution on spatial prediction of colluvial soils in a luvisol region. PLoS One 11, 1–18. https://doi.org/10.1371/journal.pone.0165699

Qi, X., Fu, Y., Wang, R.Y., Ng, C.N., Dang, H., He, Y., 2018. Improving the sustainability of agricultural land use: An integrated framework for the conflict between food security and environmental deterioration. Appl. Geogr. 90, 214–223. https://doi.org/10.1016/j.apgeog.2017.12.009

Raij, B.van., Bataglia, O.C., 1991. Análise de laboratório, in: Oliveira, A.J., Garrido, W.E., Araujo, J.D., Lourenço, S.(Eds), Métodos de pesquisa em fertilidade do solo. Embrapa-SEA, Brasília, pp.81-101.

Ribeiro, A.C., Guimarães, P.T.G., Alvarez, V.H., 1999. Recomendações para o uso de corretivos e fertilizantes em Minas Gerais - 5ª aproximação. UFV, Viçosa.

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

Ribeiro, B.T., Weindorf, D.C., Silva, B.M., Tassinari, D., Amarante, L.C., Curi, N.,

Guilherme, L.R.G., 2018. The influence of soil moisture on oxide determination in tropical soils via portable X-ray fluorescence. Soil Sci. Soc. Am. J. 82, 632-644. https://doi.org/10.2136/sssaj2017.11.0380

Rourke, S.M.O., 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

RStudio Team, 2015. RStudio: Integrated development for R. Boston, MA, USA. Available online at http://www.rstudio.com/ (verified 03 Jan. 2019).

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, 648-655 https://doi.org/10.1071/SR18099

Schlindwein, J.A., Anghinoni, I., 2002. Tamanho da subamostra e representatividade da fertilidade do solo no sistema plantio direto. Cienc. Rural 32, 963–968. https://doi.org/ISSN 0103-8478

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

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

Shi, X., Long, R., Dekett, R., Philippe, J., 2009. Integrating different types of knowledge for digital soil mapping. Soil Sci. Soc. Am. J. 73, 1682–1692.

https://doi.org/10.2136/sssaj2007.0158

Shukla, A.K., Behera, S.K., Lenka, N.K., Tiwari, P.K., Prakash, C., Malik, R.S., Sinha, N.K., Singh, V.K., Patra, A.K., Chaudhary, S.K., 2016. Spatial variability of soil micronutrients in the intensively cultivated Trans-Gangetic Plains of India. Soil Tillage Res. 163, 282–289. https://doi.org/10.1016/j.still.2016.07.004

Silva, S.H.G., Menezes, M.D. de, Owens, P.R., Curi, N., 2016a. Retrieving pedologist's mental model from existing soil map and comparing data mining tools for refining a larger area map under similar environmental conditions in Southeastern Brazil. Geoderma 267, 65–77. https://doi.org/10.1016/j.geoderma.2015.12.025

Silva, S.H.G., Poggere, G.C., Menezes, M.D., Carvalho, G.S., Guilherme, L.R.G., Curi, N., 2016b. Proximal sensing and digital terrain models applied to digital soil mapping and modeling of Brazilian Latosols (Oxisols). Remote Sens. 8, 614-635. https://doi.org/10.3390/rs8080614

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. dos S., Menezes, M.D. de, Guilherme, L.R.G., Moreira, F.M. de 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

Soil Survey Staff, 2014. Keys to soil taxonomy. 12th ed. (USDA-NRCS)

Souza, E. de., Inácio, E., Filho, F., Ernesto, C., Reynaud, G., Batjes, N.H., 2016. Pedotransfer functions to estimate bulk density from soil properties and environmental covariates: Rio Doce basin. Sci. Agric. 73, 525–534. http://dx.doi.org/10.1590/0103-9016-2015-0485

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

Suh, J., Lee, H., Choi, Y., 2016. A rapid, accurate, and efficient method to map heavy metal-contaminated soils of abandoned mine sites using converted portable XRF data and GIS. Int. J. Environ. Res. Public Health 13, 1191-1208. https://doi.org/10.3390/ijerph13121191

Taylor, J.A., Jacob, F., Galleguillos, M., Prévot, L., Guix, N., Lagacherie, P., 2013. The utility of remotely-sensed vegetative and terrain covariates at different spatial resolutions in modelling soil and watertable depth (for digital soil mapping). Geoderma 193–194, 83–93. https://doi.org/10.1016/j.geoderma.2012.09.009

Teixeira, A.F.S, Weindorf, D.C., Silva, S.H.G., Guilherme, L.R.G., Curi, N. 2018. Portable X-ray fluorescence (pXRF) spectrometry aaplied to the prediction of chemical attributes in Inceptisols under different land uses. Ciência e Agrotecnologia 42, 501-512. https://doi.org/10.1590/1413-70542018425017518

Torres, J.L.R., Pereira, M.G., 2008. Dinâmica do potássio nos resíduos vegetais de plantas de cobertura no Cerrado. Rev. Bras. Cienc. do Solo 32, 1609–1618. https://doi.org/10.1590/S0100-06832008000400025

Towett, E.K., Shepherd, K.D., Tondoh, J.E., Winowiecki, L.A., Lulseged, T., Nyambura, M., Sila, A., Vågen, T.G., Cadisch, G., 2015. Total elemental composition of soils in Sub-Saharan Africa and relationship with soil forming

factors. Geoderma Reg. 5, 157–168. https://doi.org/10.1016/j.geodrs.2015.06.002

Trabaquini, K., Formaggio, A.R., Galvão, L.S., 2015. Changes in physical properties of soils with land use time in the Brazilian savanna environment. L. Degrad. Dev. 26, 397–408. https://doi.org/10.1002/ldr.2222

Vasu, D., Singh, S.K., Sahu, N., Tiwary, P., Chandran, P., Duraisami, V.P., Ramamurthy, V., Lalitha, M., Kalaiselvi, B., 2017. Assessment of spatial variability of soil properties using geospatial techniques for farm level nutrient management. Soil Tillage Res. 169, 25–34. https://doi.org/10.1016/j.still.2017.01.006

Wang, M., Zheng, Q., Shen, Q., Guo, S., 2013. The critical role of potassium in plant stress response. Int. J. Mol. Sci. 14, 7370–7390. https://doi.org/10.3390/ijms14047370

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

Weindorf, D.C., Zhu, Y., Haggard, B., Lofton, J., Chakraborty, S., Bakr, N., Zhang, W., Weindorf, W.C., Legoria, M., 2012a. Enhanced pedon horizonation using portable X-ray fluorescence spectrometry. Soil Sci. Soc. Am. J. 76, 522-531. https://doi.org/10.2136/sssaj2011.0174

Weindorf, D.C., Zhu, Y., McDaniel, P., Valerio, M., Lynn, L., Michaelson, G., Clark, M., Ping, C.L., 2012b. Characterizing soils via portable X-ray fluorescence spectrometer: 2. Spodic and Albic horizons. Geoderma 189–190, 268–277. https://doi.org/10.1016/j.geoderma.2012.06.034

Zhu, A. X., Yang, L., Li, B., Qin, C., Pei, T., Liu, B., 2010. Construction of membership functions for predictive soil mapping under fuzzy logic. Geoderma

155, 164–174. https://doi.org/10.1016/j.geoderma.2009.05.024

Zhu, H., Hu, W., Bi, R., Peak, D., Si, B., 2016. Scale- and location-specific relationships between soil available micronutrients and environmental factors in the Fen River basin on the Chinese Loess Plateau. Catena 147, 764–772. https://doi.org/10.1016/j.catena.2016.08.038

Zhu, Y., Weindorf, D.C., Zhang, W., 2011. Characterizing soils using a portable X-ray fluorescence spectrometer: 1. Soil texture. Geoderma 167–168, 167–177. https://doi.org/10.1016/j.geoderma.2011.08.010