DOI: http://dx.doi.org/10.1590/1678-992X-2018-0132

ISSN 1678-992X

SCIENTIA AGRICOLA

Research Article

Modeling and prediction of sulfuric acid digestion analyses data from PXRF spectrometry in tropical soils

Sérgio Henrique Godinho Silva[1], Elen Alvarenga Silva[1], Giovana Clarice Poggere[2], Alceu Linares Pádua Junior[3], Mariana Gabriele Marcolino Gonçalves[1], Luiz Roberto Guimarães Guilherme[1], Nilton Curi[1]*

[1]Universidade Federal de Lavras – Depto. de Ciência do Solo, C.P. 3037 – 37200-000 – Lavras, MG – Brasil.
[2]Universidade Tecnológica Federal do Paraná – Depto. de Ciências Biológicas e Ambientais – Campus Medianeira, Av. Brasil, 4232 – Independência – 84884-000 – Medianeira, PR – Brasil.
[3]Universidade Federal dos Vales do Jequitinhonha e Mucuri/Instituto de Ciências Agrárias – Campus Unaí, Av. Vereador João Narciso, 1380 – Cachoeira – 38610-000 – Unaí, MG – Brasil.
*Corresponding author <niltcuri@ufla.br>

Edited by: Tiago Osório Ferreira

Received April 26, 2018
Accepted January 10, 2019

ABSTRACT: Sulfuric acid digestion analyses (SAD) provide useful information to environmental studies, in terms of the geochemical balance of nutrients, parent material uniformity, nutrient reserves for perennial crops, and mineralogical composition of the soil clay fraction. Yet, these analyses are costly, time consuming, and generate chemical waste. This work aimed at predicting SAD results from portable X-ray fluorescence (pXRF) spectrometry, which is proposed as a "green chemistry" alternative to the current SAD method. Soil samples developed from different parent materials were analyzed for soil texture and SAD, and scanned with pXRF. The SAD results were predicted from pXRF elemental analyses through simple linear regressions, stepwise multiple linear regressions, and random forest algorithm, with and without incorporation of soil texture data. The modeling was developed with 70 % of the data, while the remaining 30 % was used for validation through calculation of R^2, adjusted R^2, root mean square error, and mean error. Simple linear regression can accurately predict SAD results of Fe_2O_3 (R^2 0.89), TiO_2 (R^2 0.96), and P_2O_5 (R^2 0.89). Stepwise regressions provided accurate predictions for Al_2O_3 (R^2 0.87) and Ki - molar weathering index (SiO_2/Al_2O_3) (R^2 0.74) by incorporating soil texture data, as well as for SiO_2 (R^2 0.61). Random forest also provided adequate predictions, especially for Fe_2O_3 (R^2 0.95), and improved results of Kr - molar weathering index ($SiO_2/(Al_2O_3 + Fe_2O_3)$) ($R^2$ 0.66), by incorporation of soil texture data. Our findings showed that the SAD results could be accurately predicted from pXRF data, decreasing costs, time and the production of laboratory waste.

Keywords: soil clay fraction, weathering indices, random forest, proximal sensors, green chemistry

Introduction

Applied research using portable X-ray fluorescence (pXRF) spectrometry has increased dramatically over the last few years in different fields of Soil Science (Duda et al., 2017; Chakraborty et al., 2017; Silva et al., 2017; Stockmann et al., 2016). Weindorf et al. (2014) suggested that, in the future, great efforts will be made to establish correlations between pXRF data and results of conventional laboratorial analyses.

Recent studies have used pXRF data to predict various soil chemical and physical properties resulted from conventional laboratory analyses (Aldabaa et al., 2015; Sharma et al., 2014; Sharma et al., 2015; Silva et al., 2017; Zhu et al., 2011). This means that, analyses that are costly, difficult to be performed, time-consuming, and that generate chemical residues could be replaced or at least reduced, if accurate predictions of their results are achieved from pXRF data (Rouillon and Taylor, 2016; McGladdery et al., 2018).

In Brazil, sulfuric acid digestion analyses (SAD) are important for studies concerning geochemical balance of nutrients, parent material uniformity, nutrient reserves for perennial crops, as well as mineralogical composition of the soil clay fraction, among others (Curi and Kämpf, 2012). These analyses provide contents of some elements expressed on the oxide basis (Al_2O_3, SiO_2, Fe_2O_3, TiO_2, and P_2O_5). Furthermore, this data allows calculating two indices used in the Brazilian Soil Classification System and in soil surveys to differentiate highly weathered soils, that is, Ki (SiO_2/Al_2O_3) and Kr ($SiO_2/(Al_2O_3 + Fe_2O_3)$). However, conventional SAD is expensive, time-consuming, and generates considerable amounts of chemical waste. In an attempt to overcome this issue, Silva et al. (2018) used pXRF to estimate SAD results applying simple linear regression, obtaining accurate predictions only for Fe_2O_3 and TiO_2. Nevertheless, several more robust statistical models have been used in other studies, generating suitable results, such as multiple linear regressions (Rourke et al., 2016; Forkuor et al., 2017) and random forest algorithm (Chagas et al., 2016; Souza et al., 2016; Silva et al., 2017).

Considering that SAD determines the chemical composition of fine fractions (clay fraction mainly) and that pXRF determines the soil bulk chemical composition, we hypothesize that the results of soil bulk chemical composition determined by pXRF could be well correlated with SAD results in tropical soils, allowing the replacement of SAD by prediction models using pXRF data as input variables. In this sense, this study aimed at predicting the results of SAD from pXRF data, applying simple and multiple linear regressions as well as random forest algorithm. We expect not only to reduce cost, time and laboratory waste, but also to facilitate the use of SAD-derived information in studies on tropical soils.

Materials and Methods

Soil sampling and laboratory analyses

This study was conducted using 52 soil samples collected from the southern, southeastern, and northeastern regions of Brazil, encompassing four states, 19 soil classes according to Soil Taxonomy (Soil Survey Staff, 2014) and 14 parent materials (Table 1).

Soils were morphologically described, and samples from A and B horizons were collected for laboratory analyses. The soil texture analyses were performed according to Baver et al. (1972) and Gee and Bauder (1986). For SAD (Embrapa, 1997), 1 g of soil was mixed with 500 mL of sulfuric acid and 500 mL of water. Then, the solution was boiled for 30 min followed by addition of 50 mL of water. This mixture was filtered and the Fe_2O_3 and Al_2O_3 contents were determined by titration, whereas TiO_2 and P_2O_5 contents were determined by colorimetry, and SiO_2, by gravimetry. From these results, weathering indices [Ki = SiO_2/Al_2O_3 and Kr = $SiO_2/(Al_2O_3 + Fe_2O_3)$] were calculated.

pXRF analyses

Soil samples from A and B horizons were air-dried, ground and sieved through a 2-mm mesh (air-dried fine earth, ADFE). Next, about 15 g of ADFE were analyzed by pXRF spectrometry, in the Trace (dual soil) mode, during 60 s, in triplicate, using the Geochem software, as recommended by Weindorf and Chakraborty (2016). This equipment has a Rh X-ray tube, 50 keV, 100 µA, and a silicon drift detector (SDD) with resolution of < 145 eV, which allows detection of several chemical elements, from Mg to U (Weindorf et al., 2014; Ribeiro et al., 2017). The pXRF performance was checked through scanning reference materials (CRM) certified by National Institute of Standards and Technology (NIST) (2710a and 2711a) and a sample certified by the equipment manufacturer for the elements detected in most samples. The values recovered (100 × pXRF results/certified results) for these elements in comparison to information from CRM 2710a and 2711a and the manufacturer's sample, were Al_2O_3 (84/65/91), Fe (81/66/85), SiO_2 (64/47/85), P_2O_5 (381/547/-), Ti (82/69/-), K_2O (60/47/89), Cr (-/112/-), Mn (74/59/85), Ni (-/96/101), Ca (40/46/-), Cu (-/77/92), Zn (-/85/-), V (51/27/-), Zr (-/105/-), Rb (104/102/-), and Pb (107/108/104), respectively. Lack of recovery values indicates that either a certified concentration of that element was not available in the reference sample or the pXRF was not able to detect that element. It is worth mentioning that the results of pXRF may be influenced by particle size, moisture content, scanning time, interelemental interference, and atomic weight (Peinado et al., 2010; Weindorf et al., 2014; Ribeiro et al., 2017), which may explain some low recovery values for some elements.

Statistical analyses

Simple linear regressions were generated between the pXRF elemental results and elemental results from SAD, such as pXRF Al_2O_3 to predict SAD Al_2O_3. Similarly, weathering indices were calculated from both SAD and pXRF results and simple linear regressions were adjusted between them.

Contrary to the procedure to adjust simple linear regression models, when only pXRF data equivalent to the elements provided by SAD (Al_2O_3, SiO_2, Fe_2O_3, TiO_2, and P_2O_5) were used, stepwise multiple linear regressions were created to predict SAD results based on all the elements detected by pXRF for all samples, that is, Al_2O_3, Fe, SiO_2, P_2O_5, Ti, K_2O, Cr, Mn, Ni, Cu, Zn, allowing the generation of more robust models. Furthermore, two regression models per element resulted from SAD were created varying the datasets, as follows: a) using only pXRF data; and, b) using pXRF data in addition to soil texture data of each sample (sand, silt, and clay contents). In this case, we adopted the backward stepwise method in which the least important variables for the model were removed. The addition of soil texture data may improve predictions, due to relationships found between the mineral composition of the soil and the particle size fraction in which each mineral is commonly found. Quartz (SiO_2) is the main mineral found in the sand fraction of tropical soils, while most clay particles correspond to kaolinite and Fe- and Al-oxides (Kämpf et al., 2012).

The random forest algorithm was applied to predict the SAD results as well as weathering indices through the same elements used in stepwise multiple linear regression. Random forest was performed in R software, through the random Forest package (Liaw and Wiener, 2015). The number of trees grown by the model was 1000, the number of variables used per node was 5, and the number of variables inserted per tree was four, which is 1/3 of the total number of predictors, as suggested by Liaw and Wiener (2002). The model provides the mean square of residuals OOB (out-of-bag), the percentage of variance explained by the model, and the importance of each variable for the model (Breiman, 2001; Liaw and Wiener, 2002). Similar to stepwise multiple linear regressions, random forest models were created with different datasets: a) only pXRF data; and, b) pXRF data in association with sand, silt, and clay contents of each sample.

For the generation of models through linear regressions, stepwise multiple linear regressions, and random forest algorithm, the total database was separated into two datasets: 37 samples (70 %) for the creation of models, and 15 samples (30 %) for validation of the generated models to ensure that they provide accurate predictions of SAD results and weathering indices through pXRF data.

Validation

The predicted values resulted from the linear regressions, stepwise multiple linear regressions, and random forest models were compared with the observed values through calculations of R^2, adjusted R^2 (R^2_{adj}), root

Table 1 – Soils, parent material, location, number of samples (n), soil horizon, texture and results of sulfuric acid digestion analysis.

Soil	Parent material	Location (State)	Coordinates			Horizon	Clay	Silt	Sand	Sulfuric acid digestion analysis						
			Latitude	Longitude	Atitude					SiO_2	Al_2O_3	Fe_2O_3	TiO_2	P_2O_5	Ki[1]	Kr[2]
							%									
1-Rhodic Kandiudult	amphibolite	Miraí (MG)	21°02'18" S	42°42'28" W	388 m	B	43	32	25	24.8	17.5	15.4	3.1	0.12	2.4	1.5
2-Rhodic Kandiudult	basalt	Água Doce (SC)	26°38'20" S	51°32'22" W	1223 m	B	67	25	8	26.3	22	22.8	2.9	0.15	2	0
3-Anionic Acrudox	gabbro	Lavras (MG)	21°13'47" S	44°58'42" W	896 m	B	68	17	15	12.3	28.2	23.2	2	0.06	0.7	1.2
4-Anionic Acrudox	gabbro	Lavras (MG)	21°13'49" S	44°58'11" W	923 m	A	70	11	19	13.5	26.5	20.2	2	0.07	0.9	1
5-Anionic Acrudox	gabbro	Lavras (MG)	21°13'49" S	44°58'11" W	923 m	B	69	16	15	10.8	28.9	23	2.5	0.03	0.6	3.7
6-Anionic Acrudox	itabirite	Itabirito (MG)	20°07'00" S	43°58'07" W	1313 m	B	54	15	31	1.9	15.1	64.8	3.1	0.09	0.2	0.8
7-Humic Xanthic Hapludox	granite-gneiss	Miraí (MG)	21°07'56" S	42°39'52" W	727 m	B	65	6	29	16.5	25.2	13.6	1.6	0.1	1.1	2.5
8-Inceptic Hapludult	gneiss	Lavras (MG)	21°14'04" S	44°58'30" W	873 m	A	43	7	50	13.2	16.6	5.7	1.5	0.07	1.3	0.5
9-Inceptic Hapludult	gneiss	Lavras (MG)	21°14'04" S	44°58'30" W	873 m	B	60	7	33	20.3	24.4	6.4	1.3	0.05	1.4	2.4
10-Lithic Calciustoll	limestone	Italva (RJ)	21°23'56" S	41°39'50" W	47 m	A	34	18	48	16.8	9	4.4	0.8	0.08	3.2	0.4
11-Lithic Udorthent	gabbro	Lavras (MG)	21°11'51" S	44°59'36" W	960 m	A	48	20	32	21.7	22.7	14.3	2	0.07	1.6	3.3
12-Rhodic Eutrustox	tuffite	Patos de Minas (MG)	18°38'05" S	46°21'31" W	1042 m	B	45	40	15	18.6	13.1	36	11.1	0.73	2.4	2.1
13-Rhodic Hapludox	gneiss	Lavras (MG)	21°14'05" S	44°58'13" W	989 m	A	59	15	26	19.3	24.2	10.8	1.5	0.1	1.4	1.9
14-Rhodic Hapludox	gneiss	Lavras (MG)	21°14'05" S	44°58'13" W	989 m	B	61	16	23	19.6	25.9	12.1	1.8	0.06	1.3	2.9
15-Rhodic Hapludox	phyllite-micaschist	Nazareno (MG)	21°15'45" S	44°30'47" W	1031 m	B	66	22	12	11.8	32.6	16.4	2.8	0.07	0.6	1
16-Rhodic Haplustox	pellitic rocks	Unaí (MG)	16°26'30" S	46°54'15" W	618 m	B	64	24	12	23	21	9.9	0.6	0.03	1.9	1.9
17-Rhodic Kandiudult	gabbro	Lavras (MG)	21°13'12" S	44°58'01" W	893 m	A	53	10	37	12.8	20.9	10.9	1.4	0.1	1	1.1
18-Rhodic Kandiudult	gabbro	Lavras (MG)	21°13'12" S	44°58'01" W	893 m	B	43	11	46	14.7	17.4	8.7	1.8	0.06	1.4	2.1
19-Rhodic Kandiudult	gabbro	Lavras (MG)	21°13'13" S	44°58'00" W	896 m	B	46	18	36	13.6	15.5	9.8	1.5	0.12	1.5	0.8
20-Rhodic Kandiudult	gabbro	Lavras (MG)	21°13'54" S	44°58'19" W	912 m	A	45	16	39	11.7	17.2	17.7	1.5	0.29	1.2	2.6
21-Rhodic Kandiudult	gabbro	Lavras (MG)	21°13'54" S	44°58'19" W	912 m	B	58	29	13	17.5	25.6	30.3	1.9	0.11	1.2	2.4
22-Typic Argiustoll	amphibolite	Italva (RJ)	21°24'10" S	41°40'12" W	48 m	B	48	23	29	24.5	18.6	11.5	2.6	0.09	2.2	0.9
23-Typic Argiustoll	tuffite	Patos de Minas (MG)	18°38'03" S	46°21'33" W	1040 m	B	46	28	26	15.2	10.7	32.8	11.1	1.37	2.4	0.1
24-Typic Dystrustept	pellitic rocks	Unaí (MG)	16°26'50" S	46°54'15" W	611 m	B	44	53	3	21.4	14.8	6.3	0.6	0.04	2.5	1.2
25-Typic Dystrustept	pellitic rocks	Unaí (MG)	16°26'49" S	46°53'55" W	609 m	B	40	57	3	18.5	12.6	6.8	0.4	0.03	2.5	1.3
26-Typic Dystrustept	pellitic rocks	Unaí (MG)	16°26'43" S	46°54'02" W	610 m	B	35	51	14	16.8	12.9	5.8	0.5	0.02	2.2	2.1
27-Typic Dystrustept	pellitic rocks	Unaí (MG)	16°26'36" S	46°54'15" W	616 m	B	53	37	10	20.2	15.8	8.2	0.5	0.04	2.2	0.7
28-Typic Dystrustept	gneiss	Lavras (MG)	21°13'58" S	44°58'41" W	874 m	B	34	29	37	20.1	25.7	4.9	0.7	0.05	1.3	1.7
29-Typic Dystrustept	gneiss	Lavras (MG)	21°13'50" S	44°59'10" W	867 m	A	33	16	51	14.9	20.4	4	0.6	0.05	1.2	0.2
30-Typic Dystrustept	gneiss	Lavras (MG)	21°13'50" S	44°59'10" W	867 m	B	17	39	44	21.5	24.4	3.5	0.6	0.04	1.5	0.8
31-Typic Dystrustept	tuffite	Patos de Minas (MG)	18°38'00" S	46°21'34" W	1043 m	B	48	30	22	23.2	11.1	31.4	13.1	1.68	3.5	0.2
32-Typic Endoaquent	alluvial sediments	Lavras (MG)	21°14'13" S	44°58'22" W	885 m	A	46	20	34	22.7	21.1	3	1.2	0.04	1.8	0.3
33-Typic Endoaquent	alluvial sediments	Lavras (MG)	21°14'13" S	44°58'22" W	885 m	B	34	12	54	16.2	15.5	14.2	0.8	0.02	1.8	1.4
34-Typic Hapludox	basalt	Vargem Bonita (SC)	26°52'30" S	51°47'46" W	1053 m	B	81	15	4	23.7	22.9	21.1	4.5	0.15	1.8	0.5
35-Typic Hapludox	phyllite-micaschist	Nazareno (MG)	21°15'47" S	44°30'45" W	1030 m	B	63	26	11	12.8	31	17	2	0.05	0.7	0.2
36-Typic Hapludox	gneiss	Lavras (MG)	21°13'39" S	44°57'42" W	874 m	B	52	22	26	16.3	22.4	17	3.5	0.12	1.2	0.5
37-Typic Hapludox	gneiss	Lavras (MG)	21°14'12" S	44°58'23" W	887 m	A	47	9	44	16.8	20.8	3.9	0.8	0.04	1.4	0.5
38-Typic Hapludox	gneiss	Lavras (MG)	21°14'12" S	44°58'23" W	887 m	B	61	6	33	19.4	24.1	5.5	0.7	0.04	1.4	2
39-Typic Hapludox	gneiss	Lavras (MG)	21°14'14" S	44°58'25" W	896 m	A	45	11	44	15.9	19.8	4.2	0.8	0.05	1.4	0.3
40-Typic Hapludox	gneiss	Lavras (MG)	21°14'14" S	44°58'25" W	896 m	B	58	12	30	17.5	28.8	5.8	0.9	0.03	1	2.5
41-Typic Hapludult	gneiss	Lavras (MG)	21°13'56" S	44°58'17" W	895 m	A	35	15	50	14.6	16	4.3	0.9	0.28	1.6	0.5
42-Typic Hapludult	gneiss	Lavras (MG)	21°13'56" S	44°58'17" W	895 m	B	61	12	27	21.2	27	7.1	0.8	0.29	1.3	3
43-Typic Hapludult	gneiss	Lavras (MG)	21°13'51" S	44°58'45" W	884 m	B	60	8	32	21.4	23.2	5.2	1.1	0.04	1.6	0.4
44-Typic Kandiustalf	charnockite	Itaperuna (RJ)	21°13'08" S	41°49'18" W	129 m	B	50	27	23	28.9	21.8	10.2	1.2	0.06	2.3	0.2
45-Typic Kandiustalf	micaschist-gneiss	Cabroró (PE)	08°30'09" S	39°19'39" W	332 m	B	35	19	46	19.7	13.1	8.4	2.3	0.02	2.6	0
46-Typic Plinthaquox	sediments	Campos Altos (MG)	19°37'44" S	46°04'26" W	1136 m	B	72	19	9	27.2	34.8	12.8	2.5	0.08	1.3	0
47-Typic Rhodudult	gneiss	Lavras (MG)	21°13'35" S	44°57'39" W	976 m	B	54	14	32	22	23.3	9	1.3	0.08	1.6	0.1
48-Typic Rhodudult	gneiss	Lavras (MG)	21°13'50" S	44°58'46" W	888 m	A	44	14	42	20.2	18.7	7.3	0.8	0.04	1.8	0.1
49-Typic Rhodudult	gneiss	Lavras (MG)	21°13'50" S	44°58'46" W	888 m	B	38	30	32	24.9	25	4.8	0.6	0.02	1.7	1.9
50-Typic Udorthent	basalt	Água doce (SC)	26°38'27" S	51°31'39" W	1218 m	B	33	43	24	30.8	18.8	20.9	3.5	0.17	2.8	1.8
51-Typic Udorthent	gabbro	Lavras (MG)	21°11'51" S	44°59'36" W	959 m	A	40	26	34	13.1	19.5	15.5	2.3	0.09	1.1	0.4
52-Xanthic Haplustox	cover sediments	São Gotardo (MG)	19°22'41" S	46°08'36" W	1200 m	B	74	21	5	6.1	40.5	14.9	4.9	0.12	0.3	5.4

[1]Ki = SiO_2/Al_2O_3; [2]Kr = $SiO_2/(Al_2O_3 + Fe_2O_3)$.

mean square error (RMSE) (Equation 1), and mean error (ME) (Equation 2). The greater the R^2 and R^2_{adj}, and the lower the RMSE and ME, the more accurate the prediction models, considering the defined parameter. Then, the best method was determined to calculate the content of each element, as well as to predict weathering indices resulted from SAD.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(ei - mi)^2} \quad (1)$$

$$ME = \frac{1}{n}\sum_{i=1}^{n}(ei - mi) \quad (2)$$

where *n*: number of observations; *ei*: values estimated by the model; and *mi*: values obtained through SAD.

Results and Discussion

Soil chemical attributes

The range of SAD values result from different factors of soil formation (Tables 1, 2, and 3), similar to values found for other soils from other regions of Brazil (Curi and Franzmeier, 1987; Vasconcelos et al., 2013; Santos et al., 2014; Carvalho Filho et al., 2015). Soils developed from itabirite, basalt, gabbro, and tuffite presented the highest Fe_2O_3 and TiO_2 contents and the lowest SiO_2 contents. Fe_2O_3 and TiO_2 decrease and SiO_2 increases as the parent material becomes felsic, with greater quartz amounts, such as soils derived from gneiss. The lowest Fe_2O_3 contents were found in the Typic Endoaquent, due to the reduction of Fe^{3+} to Fe^{2+} followed by solubilization of Fe-bearing minerals (Schaetzl and Anderson, 2005) and Fe^{2+} leaching.

Fe_2O_3, Al_2O_3, and P_2O_5 pXRF contents were lower or greater than those found for SAD according to soil mineralogy and, hence, soil texture, which presented a wide variation for the studied soils due to the diversity of parent materials and weathering degree of soils (Table 4). For Al_2O_3 and P_2O_5, 75 and 77 % of the samples, respectively, presented SAD contents greater than those obtained by pXRF. The opposite trend was observed for SiO_2, Fe_2O_3, and TiO_2, which presented 73, 63, and 67 % of the samples with pXRF contents greater than SAD contents. The SAD analyses are more likely to provide the digestion of clay-sized particles (Curi and Kämpf, 2012). Since Brazilian soils have large contents of SiO_2 in the sand fraction, mainly as quartz (Brinatti et al., 2010; Kämpf et al., 2012), the SiO_2 content is only accessed by pXRF, justifying its larger content than that found with SAD, similar to the results for Fe_2O_3 and TiO_2, which are components of minerals also occurring in the sand fraction, such as magnetite, rutile, and ilmenite (Kämpf et al., 2012). Since higher contents of Al and P are commonly found in the clay fraction (Brinatti et al., 2010), these contents obtained through SAD were higher than those obtained with pXRF.

Simple linear regression modeling and predictions

Table 5 presents the values of R^2, R_{adj} and the equations obtained by simple linear regressions between SAD and pXRF data. For the predictions of Fe_2O_3, TiO_2, and P_2O_5, R^2 and R_{adj} values were higher than 0.80, showing an adequate fit of these regressions for the prediction of these SAD results directly from pXRF data.

The prediction of Fe_2O_3 had the highest R^2 and R^2_{adj}, followed by P_2O_5. It is noteworthy that SAD quantifies mostly the elemental contents in the clay fraction (Resende et al., 1987; Curi and Kämpf, 2012). Consequently, since Fe_2O_3 of tropical soils is concentrated in this fraction due to the high weathering degree of these soils, the contents obtained by pXRF should be well correlated to those from SAD. However, Fe_2O_3 is also present in the form of magnetite in the sand fraction (Schaefer et al., 2008). Thus, the presence of this mineral in this fraction may have prevented an even better adjustment.

For Al_2O_3 and SiO_2, adequate adjustments were not possible. Conversely, the Ki and Kr indices had R^2 of 0.59 and 0.53, respectively. Possible reasons for SAD Al_2O_3 and SiO_2 predictions not to be viable include the frequent occurrence of quartz (SiO_2) and some presence of phyllosilicates (containing both Al and Si, among other elements) in the sand fraction of Brazilian soils. Again, as SAD quantifies mainly the elemental content of the clay fraction (Resende et al., 1987), portions of Si and Al detected by pXRF were not quantified by SAD, hindering an adequate fit of linear regressions between these values. Another factor that may have influenced the adjustment of these prediction models is the low recovery values obtained for Si and Al, probably due to factors that influence the pXRF analysis, such as particle size, moisture, sample weight, sample preparation, data collection, and instrument alignment (Weindorf et al., 2014; Silva et al., 2018; Santana et al., 2018; Ribeiro et al., 2018; Peinado et al., 2010), although the samples of this study were analyzed in similar conditions regarding these influencing factors. In fact, this may be a constraint to the pXRF analysis if such factors are not taken into account.

In contrast, the equations generated to predict SAD Fe_2O_3, TiO_2, and P_2O_5 from the contents obtained by pXRF were validated (Figure 1). TiO_2 obtained R^2 and R_{adj} of 0.96 with very low values of RMSE and ME. P_2O_5 and Fe_2O_3 presented an R^2 of 0.89. For P_2O_5, the ME and RMSE values were low, whereas for Fe_2O_3 the RMSE values were the highest among all elements (Figure 1) due to the greater range of Fe_2O_3 contents.

These results confirm the possibility of using pXRF to predict Fe_2O_3, TiO_2, and P_2O_5 contents of SAD. In a preliminary study, only Fe_2O_3 and TiO_2 yielded adequate fit values for linear regressions (Silva et al., 2018). Santana et al. (2018) found adequate results for predicting SAD Fe_2O_3 and TIO_2 for Brazilian soils. In this study, this trend was confirmed in addition to an adequate adjustment for P_2O_5 contents, which has not been previously reported.

Table 2 – Portable X-ray fluorescence (pXRF) spectrometry data for A- and B-horizon samples of the soils studied.

Soil	Hor.	SiO_2	Al_2O_3	Fe_2O_3	TiO_2	P_2O_5	K_2O	Cl	CaO	Cr	Cu	Mn	Ni	Pb	Rb	V	Zn	Zr
		%						mg kg^{-1}										
1-Rhodic Kandiudult	B	23.6	11.6	15	3	0.04	0.18	487	292	419	35	296	100	8	8	155	29	197
2-Rhodic Kandiudult	B	24.4	11.8	21.1	3.8	0.05	0.20	191	0	0	196	1254	0	50	23	473	79	262
3-Anionic Acrudox	B	11.7	17.4	27.8	2.2	0	0.09	269	0	4832	79	42	388	60	13	454	48	175
4-Anionic Acrudox	A	14.7	17.2	22.1	1.9	0.03	0.09	198	662	3631	61	341	328	37	4	354	45	158
5-Anionic Acrudox	B	12.6	17.6	23.6	1.7	0	0.08	482	0	3631	63	100	323	46	3	296	39	155
6-Anionic Acrudox	B	3.6	7.1	51.1	5.3	0	0.07	281	0	0	15	737	0	325	93	293	0	432
7-Humic Xanthic Hapludox	B	23.6	19.6	15.4	1.9	0.02	0.13	185	178	186	28	138	19	8	6	72	40	312
8-Inceptic Hapludult	A	34.3	19.1	7.7	1.6	0.07	0.19	668	883	111	32	807	18	0	0	113	34	225
9-Inceptic Hapludult	B	27.3	28.2	8.8	1.4	0.05	0.16	427	590	192	38	173	20	7	2	213	32	197
10-Lithic Calciustoll	A	29.5	9.1	5.6	0.9	0.1	0.30	313	55229	73	25	1114	36	18	20	21	40	166
11-Lithic Udorthent	A	27.6	15.7	15	1.8	0.06	0.35	354	3341	201	80	1532	52	10	17	221	68	158
12-Rhodic Eutrustox	B	16.8	8.3	30.7	16.9	0.14	0.53	0	603	0	206	1850	349	86	50	307	77	1764
13-Rhodic Hapludox	A	23.5	19	12.8	2	0.02	0.19	482	1237	272	38	336	31	8	5	90	25	217
14-Rhodic Hapludox	B	19.5	16.1	13.8	2.1	0	0.22	725	126	271	36	347	39	17	2	0	24	214
15-Rhodic Hapludox	B	12.8	20.8	19.8	3.1	0	0.09	520	0	382	25	222	0	19	3	223	13	279
16-Rhodic Haplustox	B	32.8	13.8	9.4	1.8	0.07	1.33	571	658	48	35	118	27	27	152	40	22	298
17-Rhodic Kandiudult	A	25.5	16.2	13.5	1.8	0.05	0.33	638	1174	1542	38	713	195	19	15	71	41	242
18-Rhodic Kandiudult	B	31.2	17.6	10.6	1.7	0.1	0.41	220	1133	956	34	1497	201	19	15	89	44	228
19-Rhodic Kandiudult	B	18.6	10.7	9	1.1	0.03	0.25	489	6906	814	28	546	144	9	15	99	32	200
20-Rhodic Kandiudult	A	21.7	12.7	19	1.4	0.08	0.42	354	2546	1597	67	656	117	35	22	72	47	130
21-Rhodic Kandiudult	B	14.7	14.4	28.5	1.8	0	0.08	254	599	3131	147	790	346	66	15	404	70	137
22-Typic Argiustoll	B	28.5	13.4	11.7	1.9	0.02	0.13	98	6117	107	36	585	124	20	9	171	38	208
23-Typic Argiustoll	B	14.6	5.2	24.5	17.2	0.45	0.67	0	4979	0	245	1972	371	47	104	0	124	1937
24-Typic Dystrustept	B	50.9	10.3	5.5	1.1	0.12	2.74	529	2048	57	27	198	21	9	140	71	47	179
25-Typic Dystrustept	B	48.2	9.5	6.6	1.1	0.11	2.55	538	3123	77	29	523	20	7	133	60	51	189
26-Typic Dystrustept	B	51.7	11.3	4.3	1	0.08	2.74	900	904	55	27	139	22	0	131	68	41	207
27-Typic Dystrustept	B	38.5	11.7	8	1.3	0.08	1.66	413	1772	63	31	123	28	8	146	22	26	276
28-Typic Dystrustept	B	26.7	21.6	6.8	0.9	0.02	0.48	806	805	236	16	49	96	23	8	0	53	190
29-Typic Dystrustept	A	38.7	22.1	5.2	0.8	0.03	0.47	314	0	215	16	221	50	6	10	18	35	172
30-Typic Dystrustept	B	35.9	30.4	4.8	0.6	0	0.83	560	457	102	15	43	89	6	7	40	56	155
31-Typic Dystrustept	B	17	5.1	25.3	13.4	0.58	0.84	82	3496	0	247	1685	488	47	82	245	111	1425
32-Typic Endoaquent	A	38.4	24.1	3.6	1.2	0.03	0.33	361	575	390	23	114	63	9	7	122	33	256
33-Typic Endoaquent	B	41.1	24.4	1.7	1.1	0	0.42	340	238	164	16	99	26	7	9	83	26	342
34-Typic Hapludox	B	21.5	12.2	19.7	4	0.08	0.19	85	0	0	145	620	21	48	19	572	77	335
35-Typic Hapludox	B	15.5	18.3	17.6	2.7	0	0.07	363	0	30	27	266	7	19	0	233	10	198
36-Typic Hapludox	B	6.2	4.6	7.2	1.5	0	0.07	481	0	0	0	222	11	18	0	161	41	308
37-Typic Hapludox	A	31.4	22.7	5.2	0.9	0.05	0.25	428	758	137	15	133	17	0	4	22	26	205
38-Typic Hapludox	B	26.7	25.2	6.7	0.9	0.06	0.21	438	195	184	12	27	18	0	4	0	21	189
39-Typic Hapludox	A	26.8	24.7	6.9	0.9	0.03	0.24	330	728	197	17	28	12	0	0	20	18	212
40-Typic Hapludox	B	23.1	22.9	7.1	1	0	0.23	607	237	159	14	76	16	0	2	0	19	220
41-Typic Hapludult	A	33.2	17.4	5.1	1.1	0.05	0.42	547	3902	54	18	690	37	30	12	0	31	292
42-Typic Hapludult	B	23.5	20	8.4	0.9	0.04	0.31	946	864	214	16	183	67	26	10	0	29	195
43-Typic Hapludult	B	28.2	22.1	7	1.4	0.02	0.17	400	702	348	29	174	73	0	8	68	28	179
44-Typic Kandiustalf	B	24	14.2	8.3	1	0	0.61	526	497	0	39	182	8	0	38	0	51	98
45-Typic Kandiustalf	B	33.3	15.8	10.5	1.6	0.04	2.24	498	5847	92	43	659	50	8	99	203	64	189
46-Typic Plinthaquox	B	19.8	22.1	11.4	4.8	0	0.08	328	151	0	31	300	27	29	0	249	33	540
47-Typic Rhodudult	B	23.3	18	9.5	1.2	0.04	0.19	452	145	605	31	228	93	0	6	0	26	183
48-Typic Rhodudult	A	29.3	17	7.8	0.6	0	0.39	699	205	654	21	160	83	7	11	23	33	172
49-Typic Rhodudult	B	27.8	21.5	5.8	0.6	0	0.46	586	0	364	19	0	71	6	8	0	28	140
50-Typic Udorthent	B	28.9	11.5	19.9	3.1	0.11	0.36	76	323	0	211	1590	14	27	41	434	135	213
51-Typic Udorthent	A	24.2	14.4	15.5	1.7	0.03	0.47	328	10293	229	99	1816	53	23	19	241	64	185
52-Xanthic Haplustox	B	5.1	23.7	16.1	6.5	0	0.07	271	0	0	15	283	0	8	0	488	10	683

Table 3 – Maximum, minimum, and mean values of data from sulfuric acid digestion analyses (SAD) and elemental analyses by portable X-ray fluorescence (pXRF) spectrometry of A- and B-horizon samples of the soils studied.

	Element/ Weathering indices	Modeling[1]				Validation[1]			
		Min	Max	Mean	STD	Min	Max	Mean	STD
		SAD							
Method	SiO_2	1.91	27.24	18.38	5.36	12.27	30.84	17.88	6.03
	Al_2O_3	8.99	40.52	21.50	6.56	10.74	31.04	20.61	6.09
	Fe_2O_3	2.97	64.79	13.66	11.92	4.25	32.84	13.18	8.35
	TiO_2	0.47	13.12	2.29	2.63	0.36	11.06	2.11	2.61
	P_2O_5	0.02	1.68	0.14	0.29	0.02	1.37	0.17	0.34
	Ki	0.22	3.59	1.60	0.73	0.71	2.82	1.57	0.7
	Kr	0.06	2.45	1.15	0.48	0.49	1.85	1.10	0.47
		pXRF							
pXRF	SiO_2	3.55	51.7	25.61	10.86	11.66	48.16	25.62	10.07
	Al_2O_3	4.60	30.39	17.09	6.44	5.24	21.59	15.29	4.76
	Fe_2O_3	1.69	51.1	13.08	9.66	5.06	27.81	13.33	7.54
	TiO_2	0.56	16.91	2.64	3.35	0.58	17.24	2.63	4.1
	P_2O_5	0.00	0.58	0.05	0.1	0.00	0.45	0.07	0.11
	Ki	0.21	4.93	1.68	0.99	0.67	5.05	1.9	1.14
	Kr	0.06	3.30	0.98	0.68	0.32	3.41	1.19	0.78

[1]Values calculated for A and B horizons.

Table 4 – Minimum, maximum, mean, and standard deviation values of soil texture for A- and B-horizon samples of the studied soils.

	Sand	Silt	Clay	Sand	Silt	Clay
	%					
	Modeling			Validation		
Min	3	6	17	3	6	33
Max	54	53	81	50	57	70
Mean	28	20	52	27	25	48
STD	14	11	13	14	14	13

Table 5 – Equations, R^2, and R_{adj} for prediction of sulfuric acid digestion results from elemental characterization by portable X-ray fluorescence (pXRF) spectrometry (in ppm).

	R^2	R_{adj}	Equations
Al_2O_3	0.29	0.27	$Al_2O_{3SAD} = 12.162 + 0.547*Al_2O_{3pXRF}$
SiO_2	0.15	0.12	$SiO_{2SAD} = 13.541 + 0.189*SiO_{2pXRF}$
P_2O_5	0.83	0.82	$P_2O_{5SAD} = -0.00456 + 2.697*P_2O_{5pXRF}$
Fe_2O_3	0.91	0.91	$Fe_2O_{3SAD} = -1.728 + 1.177*Fe_2O_{3pXRF}$
TiO_2	0.90	0.89	$TiO_{2SAD} = 0.328 + 0.745*TiO_{2pXRF}$
Ki	0.59	0.58	$Ki_{SAD} = 0.648 + 0.563*Ki_{pXRF}$
Kr	0.53	0.52	$Ki_{SAD} = 0.641 + 0.513*Kr_{pXRF}$

Stepwise modeling and predictions

The stepwise multiple linear regression method allowed the evaluation of these models to predict SAD chemical composition from pXRF values using additional predictors in relation to those used in the simple linear regression, that is, other pXRF elemental data and soil texture data. The generated models that used either pXRF data only or pXRF + texture data as predictor variables and their R^2 and R_{adj} values are shown in Table 6.

In the stepwise regressions, the R^2 values were all above 0.79 for models without incorporation of soil texture data and above 0.83, after incorporating texture data as predictor variables, including Al_2O_3 and SiO_2, which did not yield adequate models through simple linear regression. A better adjustment of the equations when including texture data was observed for predictions of SAD Al_2O_3 and SiO_2. For Fe_2O_3, TiO_2, P_2O_5, Ki, and Kr, models with or without soil texture data presented basically the same R^2 and R^2_{adj} values. This behavior indicates that not only texture, but also other elements provided by the pXRF analysis, in addition to Fe, Al, Si, P, and Ti, contributed to modeling SAD results.

Regarding the pXRF elemental data, the number (in parenthesis) of models in which each element was present is: Zr (12), Fe (12), Mn (9), Al (8), Ti (8), Rb (8), Ni (8), Cr (8), Si (7), V (6), Cu (6), P (6), Ca (6), Pb (5), K (4), Cl (3), Zn (3). This fact demonstrates that some elements that now are easily accessed with pXRF may be well correlated with soil properties, facilitating the development of prediction models. Rubidium (Rb), for instance, was found to be correlated with clay content in soils from the United States by Zhu et al. (2011). Regarding the texture variables in 7 models, sand, silt, and clay were present, respectively, in 3, 2, and 3 models. All the models incorporating texture data as predictor variables presented at least one size fraction in the equation, except for the Kr model, highlighting the expectations confirmed of sand presence in SAD SiO_2 model, while clay was one of the variables of the SAD Fe_2O_3 and Al_2O_3 models, which reflects Brazilian soils mineralogy, as previously discussed.

Although the Al_2O_3 and SiO_2 content did not present a good fit with simple linear regressions (Table 5), in stepwise multiple linear regressions these two elements achieved high R^2 and R_{adj}, that is, 0.79 and 0.76 for Al_2O_3, and 0.79 and 0.71 for SiO_2, with and without incorporating texture. When incorporating texture data as predictors, the model for SAD Al_2O_3 included clay and silt variables, which can be explained by the weathering process and formation of kaolinite and gibbsite in greater proportion in these size fractions.

In contrast, the SAD SiO_2 model used sand as a variable. Quartz predominates in the sand fraction of tropical soils, which reinforces the importance of this particle size fraction for the prediction of SAD SiO_2. The non-adjustment of the simple linear regression for SiO_2 may be because SAD does not digest quartz, as this method is intended for minerals in the clay fraction (Resende et al., 1987). The SiO_2 content could be underestimated in SAD, while in the pXRF analysis, all SiO_2 is detected from soil bulk composition. In the stepwise regression, the use of the sand fraction in the model as a predictor variable corrected this effect leading to better model adjustments.

For the prediction of SAD Fe_2O_3, the clay fraction was one of the independent variables inserted into the model, reinforcing the concentration of the Fe-bearing

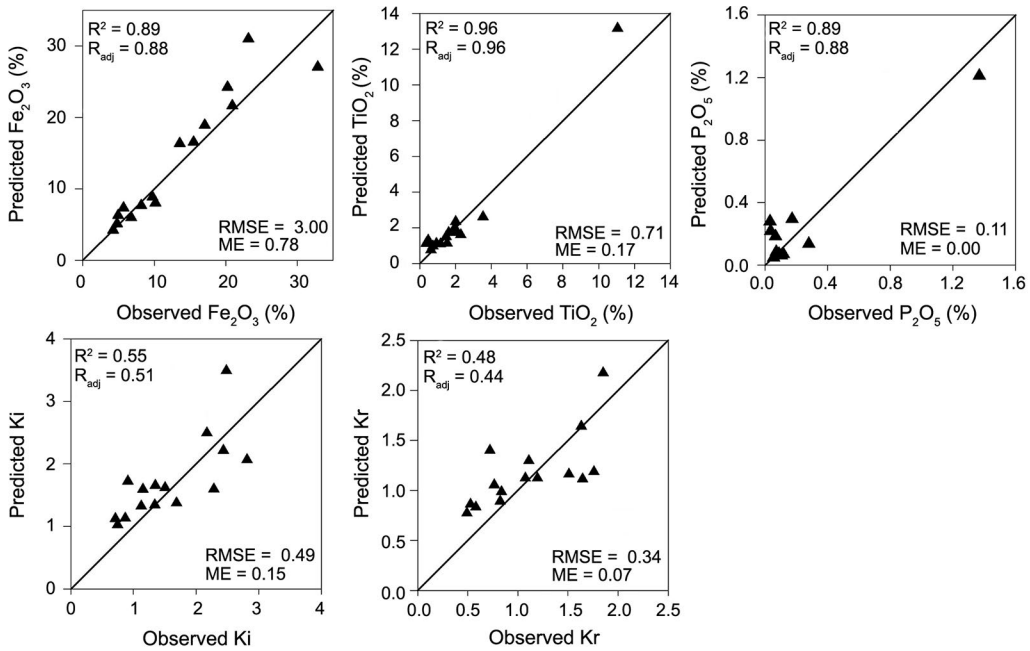

Figure 1 – Validation of simple linear regressions to predict results of sulfuric acid digestion analyses from data obtained with portable X-ray fluorescence (pXRF) spectrometry. RMSE = root mean square error; ME = mean error.

Table 6 – Equations, R^2, and R_{adj} for stepwise multiple linear regressions to predict sulfuric acid digestion (SAD) results from data of portable X-ray fluorescence (pXRF) spectrometry, with and without incorporating soil texture data (in %) into the models.

Elements/ Weathering indices	R^2	R_{adj}	Equations
			Without soil texture
Al_2O_3	0.79	0.76	$Al_2O_{3SAD} = 16.854 + 215.02*V - 0.403*SiO_2 - 0.235*Fe_2O_3 + 113.898*Cl + 0.587*Al_2O_3$
SiO_2	0.79	0.71	$SiO_{2SAD} = 29.042 - 438.925*Zr + 3.635*TiO_2 + 861.21*Rb + 550.93*Ni - 39.404*Mn - 4.844*K_2O - 0.658*Fe_2O_3 + 807.238*Cu - 69.24*Cr - 31.562P_2O_5$
P_2O_5	0.97	0.96	$P_2O_{5SAD} = -0.231 + 3.074*Zr + 15.679*Zn - 15.344*Rb - 0.708*Mn + 0.00616*Fe_2O_3 + 2.027*Cl + 2.296*P_2O_5$
Fe_2O_3	0.98	0.97	$Fe_2O_{3SAD} = 5.322 + 292.806*Zr + 94.698*V - 2.867*TiO_2 - 214.855*Rb + 760.132*Pb + 0.755*Fe_2O_3 - 1.005*CaO - 0.343*Al_2O_3$
TiO_2	0.99	0.99	$TiO_{2SAD} = 0.236 + 41.251*Zr + 185.155*Zn + 32.409*V - 104.257*Rb - 161.695*Pb + 54.744*Ni - 7.963*Mn + 0.147*Fe_2O_3 - 112.044*Cu - 8.487*Cr + 6.424*P_2O_5 - 0.0677*Al_2O_3$
Ki	0.90	0.86	$Ki_{SAD} = 1.854 - 24.874*Zr + 0.244*TiO_2 + 0.0252*SiO_2 + 40.464*Ni - 3.821*Mn - 0.0297*Fe_2O_3 + 44.141*Cu - 4.994*Cr + 0.283*CaO - 0.0358*Al_2O_3$
Kr	0.84	0.80	$Kr_{SAD} = 0.904 - 21.289*Zr + 0.195*TiO_2 + 0.0223*SiO_2 + 18.214*Ni - 0.0205*Fe_2O_3 - 2.58*Cr + 0.209*CaO$
			With soil texture
Al_2O_3	0.94	0.92	$Al_2O_{3SAD} = -0.441 - 0.427*TiO_2 - 0.366*SiO_2 - 222.791*Pb + 23.88*Mn - 255.975*Cu + 0.726*Al_2O_3 + 0.289*Clay + 0.291*Silt$
SiO_2	0.83	0.75	$SiO_{2SAD} = 28.549 - 428.159*Zr + 2307.631*Zn - 142.37*V + 4.149*TiO_2 + 1051.561*Rb + 361.983*Ni - 43.614*Mn - 7.926*K_2O - 0.566*Fe_2O_3 - 51.997*Cr - 29.803*P_2O_5 - 0.118*Sand$
P_2O_5	0.97	0.97	$P_2O_{5SAD} = - 0.123 + 2.466*Zr - 0.00633*SiO_2 - 6.898*Rb - 0.77*Mn + 11.87*Cu + 1.755*Cl + 2.15*P_2O_5 + 0.00291*Silt + 0.00346*Sand$
Fe_2O_3	0.98	0.97	$Fe_2O_{3SAD} = 9.247 + 283.057*Zr + 125.875*V - 2.832*TiO_2 + 661.863*Pb - 1.617*K_2O + 0.772*Fe_2O_3 - 1.245*CaO - 0.333*Al_2O_3 - 0.0774*Clay$
TiO_2	0.99	0.99	$TiO_{2SAD} = 1.184 + 35.331*Zr + 39.467*V - 0.0338*SiO_2 - 110.469*Rb - 131.694*Pb + 69.278*Ni - 8.447*Mn + 0.51*K_2O + 0.1*Fe_2O_3 - 9.786*Cr + 5.27*P_2O_5 - 0.0694*Al_2O_3 + 0.0181*Sand$
Ki	0.90	0.86	$Ki_{3SAD} = 3.036 - 31.479*Zr + 0.308*TiO_2 + 33.954*Rb + 32.739*Ni - 3.586*Mn - 0.0501*Fe_2O_3 + 62.353*Cu - 4.229*Cr + 0.26*CaO - 0.0241*Al_2O_3 - 0.0114*Clay$
Kr	0.84	0.80	$Kr_{SAD} = 0.901 - 20.877*Zr + 0.188*TiO_2 + 0.0223*SiO_2 + 18.648*Ni - 0.0202*Fe_2O_3 - 2.633*Cr + 0.21*CaO$

minerals in the smallest particle fractions in soils as the weathering processes advance, mainly as hematite and goethite (Kämpf et al., 2012). The sand fraction was included as predictor variable into the SAD TiO_2 model, whereas both sand and silt fractions were present in the model for the SAD P_2O_5 prediction. This result indicates the presence of minerals containing P and Ti in the sand and silt fractions in the soils studied. The sand and silt fractions of soils developed from basalt, gabbro, tuffite, and amphibolite may present considerable amounts of minerals such as anatase, rutile, and titanomagnetite, which present Ti in their composition (Fabris et al., 1997; Fabris et al., 1998).

The Ki and Kr indices were also submitted to stepwise multiple linear regressions, yielding R^2 values of 0.90 and 0.84, respectively. Models with and without soil texture data had similar R^2 values. Most predictor variables were the same in models with and without incorporation of soil texture data, except for the prediction of Ki values. In this last case, by incorporating soil texture data into the model, the clay content was added to the model while Si was replaced by Rb.

The validation indices for prediction of SAD values and weathering indices from the formerly presented equations are shown in Figure 2. Except for Kr, Fe_2O_3, and SiO_2, validation indices were always better when incorporating soil texture into the models. P_2O_5, TiO_2, Ki, and Kr presented RMSE and ME values close to zero for validation of models with and without incorporating soil texture data, which is in agreement with the low range of their values. R^2 and R^2_{adj} for Fe_2O_3 validation without incorporating texture data were the same as those

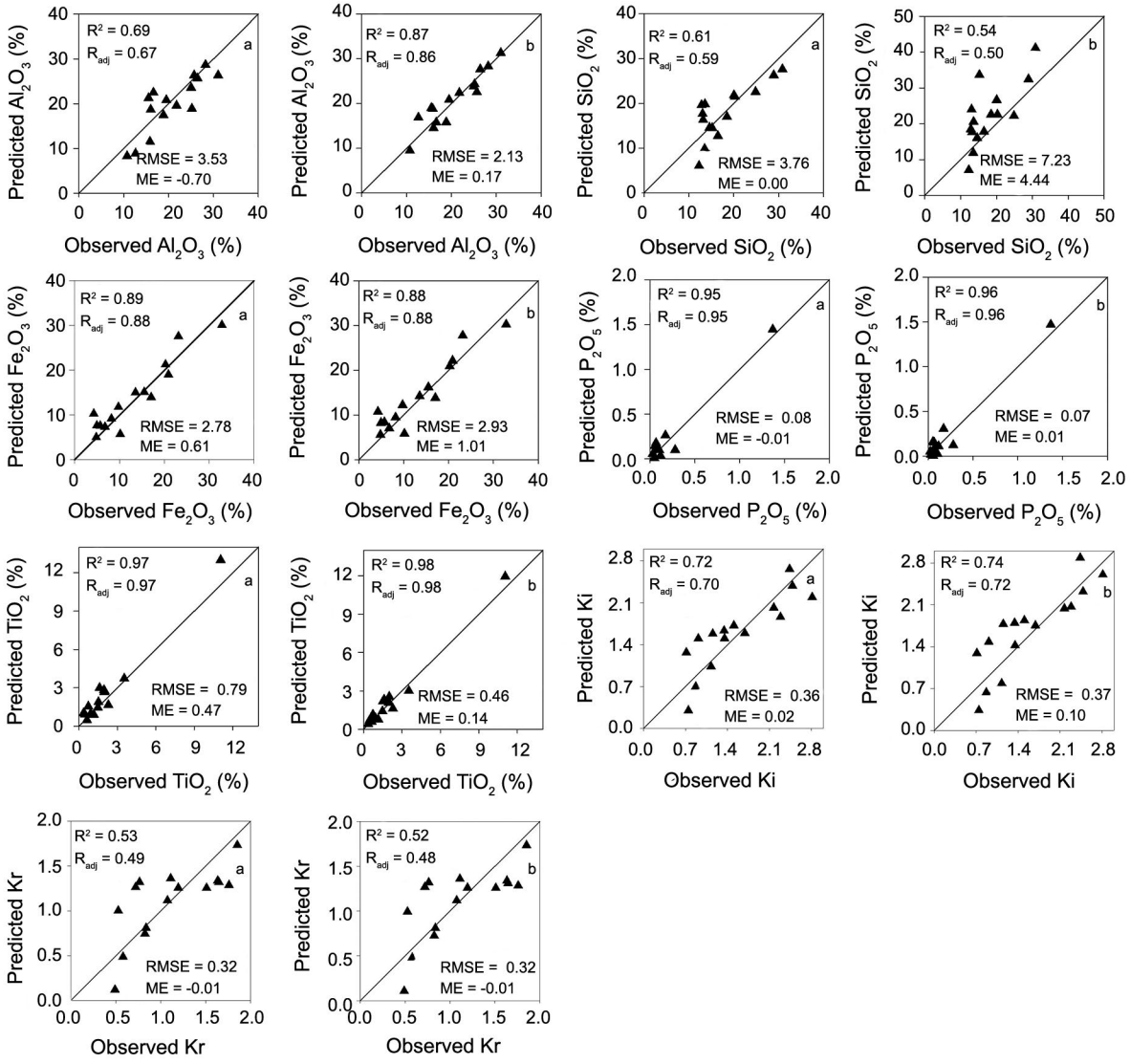

Figure 2 – Validation of stepwise multiple linear regressions to predict results of sulfuric acid digestion analyses from portable X-ray fluorescence (pXRF) spectrometry data without (a) and with (b) soil texture data. RMSE = root mean square error; ME = mean error.

found with simple linear regressions, but RMSE and ME values were lower with the stepwise regression. These values were slightly better than those that incorporated texture data as predictors. P_2O_5 and TiO_2 had a similar trend, with validation of models incorporating texture presenting better results. Predictions of these elements were also improved in comparison to simple regressions.

A considerable improvement in the modeling and validations was achieved for Al_2O_3 and SiO_2 values, whose models reached R^2 and R^2_{adj} that were much higher than those obtained with simple regressions. Al_2O_3 validation indicated that texture improved validation indices and the opposite trend was found for SiO_2. Ki and Kr also improved in modeling and prediction quality compared to simple regression, although their validation indices were lower than those obtained for the predictions of SAD elemental contents.

In general, when using multiple regressions, texture data and pXRF elemental data other than Fe, Ti, Al, Si, and P were fundamental to improve prediction of SAD results, although predictions of SAD Fe_2O_3, TiO_2, and P_2O_5 also provided adequate results using simple linear regressions. This shows that besides soil texture, the mineralogical composition of different particle size fractions also influences the results and can be accessed by pXRF. Attention should be drawn to the fact that some of the elements used in the equations presented low recovery values, which is important since they tend to be present in small concentrations in the soils.

Random forest modeling and predictions

Random forest models for most SAD elemental contents as well as for Ki and Kr presented high percentage of variance explained and small errors (Table 7), indicating good adjustments of the models. Little variation occurred by adding information on soil texture into the models, except for Al_2O_3, which showed an improvement of 13 % of variance explained in the presence of such data.

The most important variables for predicting SAD SiO_2, Fe_2O_3, Ki and Kr were these elemental contents/weathering indices obtained from pXRF even in the presence of sand, silt, and clay data (Table 8). For the SAD Al_2O_3 model, the clay content was the most important variable, followed by pXRF Rb and Al_2O_3 contents.

Table 7 – Modeling results to predict sulfuric acid digestion results from portable X-ray fluorescence (pXRF) spectrometry by random forest algorithm, with and without incorporating soil texture data into the models.

Paramenter[1]	SiO_2	Al_2O_3	Fe_2O_3	TiO_2	P_2O_5	Ki	Kr
				Without soil texture			
VarEx	25.72	38.63	47.59	67.88	40.69	65.08	64.81
MSE	24.63	30.46	86.33	2.73	0.06	0.20	0.08
				With soil texture			
VarEx	24.33	51.20	49.04	66.47	40.72	60.2	63.14
MSE	25.10	24.22	83.93	2.85	0.06	0.23	0.08

[1]VarEx = percentage of variance explained by the model; MSE = mean square of residuals out-of-bag method.

Table 8 – Importance of the variables (Imp) for random forest models using only pXRF data or pXRF together with texture data to predict results from sulfuric acid digestion analyses.

Imp	Predicted variable							Predicted variable						
	SiO_2	Al_2O_3	Fe_2O_3	TiO_2	P_2O_5	Ki	Kr	SiO_2	Al_2O_3	Fe_2O_3	TiO_2	P_2O_5	Ki	Kr
	Only pXRF data							pXRF + texture data						
1	SiO_2	Rb	Fe_2O_3	V	SiO_2	Ki _pXRF	Kr_pXRF	SiO_2	Clay	Fe_2O_3	V	Zr	Ki _pXRF	Kr_pXRF
2	Cr	Al_2O_3	Pb	TiO_2	TiO_2	SiO_2	SiO_2	Zn	Rb	Pb	TiO_2	TiO_2	SiO_2	SiO_2
3	Zn	K_2O	TiO_2	Zr	Mn	Al_2O_3	K_2O	Cr	Al_2O_3	TiO_2	Fe_2O_3	Pb	Silt	K_2O
4	Ni	SiO_2	Cu	Fe_2O_3	Cu	P_2O_5	Cr	Silt	K_2O	V	Zr	SiO_2	P_2O_5	Zn
5	Cu	P_2O_5	V	Cr	Zr	Zn	Fe_2O_3	Cu	P_2O_5	Cu	Cu	Mn	Zn	Pb
6	K_2O	V	SiO_2	Mn	Cl	K_2O	Zn	K_2O	SiO_2	SiO_2	Mn	P_2O_5	K_2O	CaO
7	Pb	CaO	Ni	Cu	P_2O_5	CaO	CaO	Fe_2O_3	Sand	Zr	Cl	Cu	Al_2O_3	Cr
8	Fe_2O_3	Pb	Zr	Ni	Ni	Cu	Pb	Sand	Cr	Mn	Cr	Cl	CaO	Silt
9	Cl	Mn	Al_2O_3	Cl	Fe_2O_3	Pb	Al_2O_3	Ni	V	Zn	P_2O_5	Ni	Clay	Fe_2O_3
10	CaO	Cr	Mn	P_2O_5	Pb	Fe_2O_3	Cu	Cl	CaO	Al_2O_3	Ni	Sand	Pb	Al_2O_3
11	Mn	Fe_2O_3	Zn	SiO_2	CaO	Ni	Rb	CaO	Fe_2O_3	Ni	CaO	Zn	Cr	Clay
12	V	Zr	CaO	K_2O	Al_2O_3	Cr	Ni	Al_2O_3	Silt	Sand	SiO_2	Fe_2O_3	Ni	Rb
13	Zr	Cl	Cr	Zn	Zn	Rb	P_2O_5	TiO_2	Pb	Cl	Zn	CaO	Sand	Cu
14	Al_2O_3	Zn	Rb	Al_2O_3	K_2O	Mn	TiO_2	Pb	Mn	CaO	Al_2O_3	Rb	Cu	TiO_2
15	TiO_2	Ni	Cl	CaO	Rb	Cl	V	Mn	Cu	K_2O	K_2O	Clay	Cl	Ni
16	P_2O_5	TiO_2	K_2O	Pb	Cr	TiO_2	Cl	P_2O_5	Cl	Silt	Pb	Al_2O_3	Mn	P_2O_5
17	Rb	Cu	P_2O_5	Rb	V	V	Zr	Rb	Ni	Rb	Clay	Cr	TiO_2	V
18	–	–	–	–	–	Zr	Mn	V	TiO_2	Cr	Silt	K_2O	Fe_2O_3	Cl
19	–	–	–	–	–	–	–	Zr	Zr	Clay	Sand	V	Zr	Sand
20	–	–	–	–	–	–	–	Clay	Zn	P_2O_5	Rb	Silt	Rb	Zr
21	–	–	–	–	–	–	–	–	–	–	–	–	V	Mn

In the model without soil texture variables, Rb was the most important, followed by Al_2O_3. This fact demonstrates the importance of variables provided by pXRF for these predictions. Zhu et al. (2011) used pXRF data for predicting sand and clay contents in soils from Lousiana and New Mexico (USA) and found a correlation of 0.91 between Rb and clay contents. These findings may indicate a similar trend also for Brazilian soils, although these authors are aware of differences between those soils and the ones in our study.

For SAD SiO_2 prediction, silt and sand were defined as the 4th and 8th most important variables, and clay was the least important. In both models for SAD TiO_2, pXRF V was the most important variable, followed by pXRF TiO_2. As expected, the indices Ki and Kr calculated using pXRF SiO_2, Al_2O_3, and Fe_2O_3 results were the most important variables to predict SAD Ki and Kr, followed by pXRF SiO_2, since the latter element is taken into account in the formulae used to calculate those indices.

The validation of the random forest models is presented in Figure 3. The validation results, in general, were very similar with and without incorporation of soil texture data into the models. Thus, for random forest algorithms, soil texture data did not improve the predictions of SAD results and Ki and Kr values. However, random forest provided better results for SAD Al_2O_3 with texture data, and SAD Fe_2O_3 and Kr with and without this attribute data in comparison to stepwise regression. It demonstrates that, without information on soil texture, random forest algorithms can be used to deliver better predictions of SAD Fe_2O_3, and Kr than stepwise multiple linear regressions.

Influence of soil texture on the prediction of weathering indexes

The prediction of elemental contents (expressed on the oxide basis) and weathering indices from the pXRF analysis was substantially improved when texture was used, especially in the multiple regression analysis for Al_2O_3 and SiO_2. The reasons for this occurrence are, first, due to the nature of the SAD and pXRF analyses: while the SAD is efficient in the dissolution of minerals of the clay fraction, the pXRF analysis provides results of the total chemical composition of the sample. Thus,

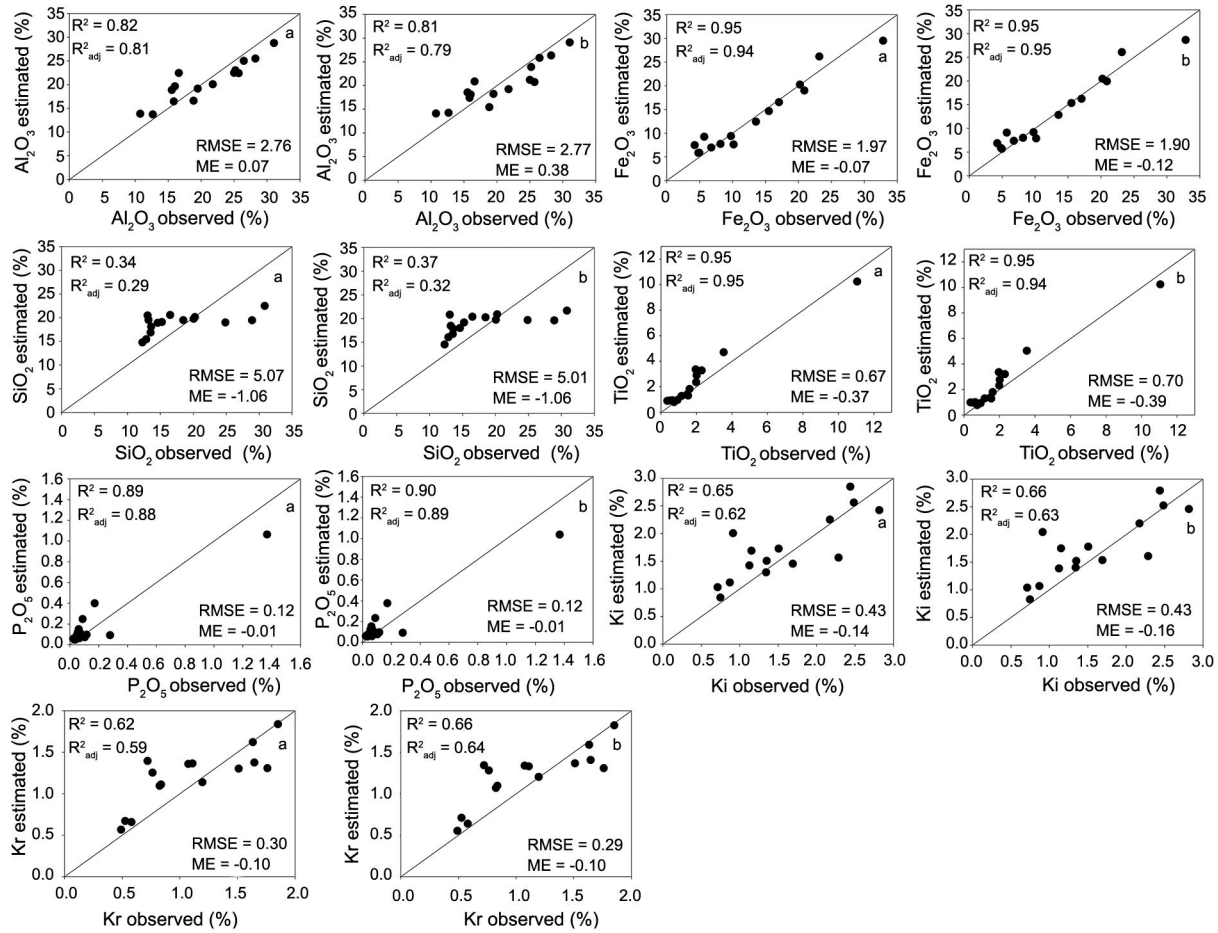

Figure 3 – Validation of random forest models to predict results of sulfuric acid digestion analyses from portable X-ray fluorescence data without (a) and with (b) soil texture data. RMSE = root mean square error; ME = mean error.

such pXRF results may differ from those obtained from other analyses, including SAD, which explains the reason why the incorporation of soil texture into the models in addition to the contents of elements in the minerals was decisive for the improvement of the prediction models. Second, mineralogy of Brazilian soils is mostly dominated by kaolinite, Fe and Al oxides (hematite, goethite, and gibbsite) in the clay fraction and quartz in association with smaller contents of muscovite in the sand fraction (Melo et al., 2001; Inda et al., 2010; Kämpf et al., 2012; Silva et al., 2012; Carvalho Filho et al., 2015). The dominance of either sand or clay fraction in most Brazilian soils reflects the soil parent material. The silt content is much lower due to the high weathering degree of these soils (Tables 3 and 4) and represents the maximum instability fraction. Thus, the sand and clay contents as well as the mineralogical composition of these particle size fractions influenced primarily the predictions of SAD Al_2O_3 and SiO_2 contents. Thus, considering the high validation indices for all the SAD results and weathering indices, pXRF could be adopted as an alternative method to provide such soil data. The creation of better models to predict SAD results is encouraged, mainly through incorporation of more soil data.

Conclusions

Accurate predictions of SAD Al_2O_3, Fe_2O_3, SiO_2, P_2O_5, and TiO_2 results as well as Ki and Kr weathering indices can be obtained using pXRF data with and without incorporation of soil texture data into the models, through simple linear regressions, stepwise multiple linear regressions, and random forest algorithm. The clay and sand contents were crucial to improve the models to predict SAD Al_2O_3 and SiO_2, respectively. These findings demonstrate that it is possible to use pXRF to reduce costs, time, and the amount of chemical waste produced by the SAD analyses. In addition, these results contribute to speeding up not only soil chemical characterization, but also the assessment of information on soil weathering degree, geochemical balance of nutrients, parent material homogeneity, reserve of nutrients for perennial crops, mineralogy of the clay fraction, among others. Finally, the association of new tools and robust algorithms enhance soil characterization for varying purposes, while providing a fast, cost-effective, and "green chemistry" alternative for the SAD analyses.

Authors' Contributions

Conceptualization: Silva, S.H.G.; Silva, E.A.; Poggere, G.C.; Guilherme, L.R.G.; Curi, N. Data acquisition: Silva, S.H.G.; Silva, E.A.; Poggere, G.C.; Pádua Junior, A.L.; Gonçalves, M.G.M. Data analysis: Silva, S.H.G.; Silva, E.A.; Poggere, G.C.; Pádua Junior, A.L.; Gonçalves, M.G.M. Design of Methodology: Silva, S.H.G.; Silva, E.A.; Poggere, G.C.; Curi, N. Writing and editing: Silva, S.H.G.; Silva, E.A.; Guilherme, L.R.G.; Curi, N.

References

Aldabaa, A.A.A.; Weindorf, D.C.; Chakraborty, S.; Sharma, A.; Li, B. 2015. Combination of proximal and remote sensing methods for rapid soil salinity quantification. Geoderma 239: 34-46.

Baver, L.D.; Gardner, W.H.; Gardner, W.R. 1972. Soil Physics. John Wiley, New York, NY, USA.

Breiman, L. 2001. Random forests. Machine Learning 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. Scientia Agricola 67: 454-464.

Carvalho Filho, A.; Inda, A.V.; Fink, J.R.; Curi, N. 2015. Iron oxides in soils of different lithological origins in Ferriferous quadrilateral (Minas Gerais, Brazil). Applied Clay Science 118: 1-7.

Chagas, C.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.

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: 108-119.

Curi, N.; Franzmeier, D.P. 1987. Effect of parent rocks on chemical and mineralogical properties of some Oxisols in Brazil. Soil Science Society of American Journal 51: 153-158.

Curi, N.; Kämpf, N. 2012. Soil characterization = Caracterização do solo. p. 147-170. In: Ker, J.C.; Curi, N.; Schaefer, C.E.G.R.; Vidal-Torrado, P., eds. Pedology: fundamentals = Pedologia: fundamentos. SBCS, Viçosa, MG, Brazil (in Portuguese).

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.

Empresa Brasileira de Pesquisa Agropecuária [Embrapa]. 1997. Manual of soil analysis methods = Manual de métodos de análises de solos. Embrapa Solos, Rio de Janeiro, RJ, Brazil (in Portuguese).

Fabris, J.D.; Jesus Filho, M.F.; Coey, J.M.D.; Mussel, W.N.; Goulart, A.T. 1997. Iron-rich spinels from Brazilian soils. Hyperfine Interactions 110: 23–32.

Fabris, J.D.; Coey, J.M.D.; Mussel, W.N. 1998. Magnetic soils from mafic lithodomains in Brazil. Hyperfine Interactions 114: 249–258.

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: e0170478.

Gee, G.W.; Bauder, J.W. 1986. Particle-size analysis. p. 383-412. In: Klute, A., ed. Methods of soil analysis. American Society of Agronomy, Madison, WI, USA.

Inda, A.V.; Torrent, J.; Barrón, V.; Bayer, C. 2010. Aluminum hydroxy-interlayered minerals and chemical properties of a subtropical brazilian Oxisol under no-tillage and conventional tillage. Revista Brasileira de Ciência do Solo 34: 33-41.

Kämpf, N.; Marques, J.J.; Curi, N. 2012. Mineralogy of brazilian soils = Mineralogia de Solos Brasileiros. p. 81-146. In: Pedology fundamentals = Pedologia fundamentos. SBCS, Viçosa, MG, Brazil (in Portuguese).

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

Liaw, A.; Wiener, M. 2015. Package "randomForest". R Development Core Team 2: 1-29.

McGladdery, C.; Weindorf, D.C.; Chakraborty, S.; Li, B.; Paulette, L.; Podar, D.; Pearson, D.; Kusi, N.Y.O.; Duda, B. 2018. Elemental assessment of vegetation via portable X-ray fluorescence (PXRF) spectrometry. Journal of Environmental Management 210: 210-225.

Melo, V.F.; Singh, B.; Schaefer, C.E.G.R.; Novais, R.F.; Fontes, M.P.F. 2001. Chemical and mineralogical properties of kaolinite-rich brazilian soils. Soil Science Society America Journal 65: 1324-1333.

Peinado, F.M.; Ruano, S.M.; González, M.G.B.; Molina, C.E. 2010. A rapid field procedure for screening trace elements in polluted soil using portable X-ray fluorescence (PXRF). Geoderma 159: 76-82.

Resende, M.; Bahia Filho, A.F.C.; Braga, J.M. 1987. Clay mineralogy of Latossols estimated by chemical allocation of total oxides contents by H_2SO_4 digestion. Revista Brasileira de Ciência do Solo 23: 17-23 (in Portuguese, with abstract in English).

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.

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 Science Society of America Journal 82: 632-644.

Rouillon, M.; Taylor, M.P. 2016. Can field portable X-ray fluorescence (pXRF) produce high quality data for application in environmental contamination research? Environmental Pollution 214: 255-264.

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.

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 Research 56: 648-655.

Santos, W.J.R.; Curi, N.; Silva, S.H.G.; Fonseca, S.; Silva, E.; Marques, J.J. 2014. Detailed soil survey of an experimental watershed representative of the brazilian Coastal Plains and its practical application. Ciência e Agrotecnologia 38: 50-60.

Schaefer, C.E.G.R.; Fabris, J.D.; Ker, J.C. 2008. Minerals in the clay fraction of brazilian Latosols (Oxisols): a review. Clay Minerals 43: 137-154.

Schaetzl, R.J.; Anderson, S. 2005. Soil: Genesis and Geomorphology. Cambridge University Press, New York, NY, USA.

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.

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.

Silva, E.A.; Gomes, J.B.V.; Filho, J.C.A.; Vidal-Torrado, P.; Cooper, M.; Curi, N. 2012. Morphology, mineralogy and micromorphology of soils associated to summit depressions of the northeastern brazilian Coastal Plains. Ciência e Agrotecnologia 36: 507-517.

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

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.

Soil Survey Staff. 2014. Keys to Soil Taxonomy, 12ed. USDA-Natural Resources Conservation Service, Washington, DC, USA.

Souza, E.; 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. Scientia Agricola 73: 525-534.

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.

Vasconcelos, F.C.W.; Carvalho, S.A.; Silva, S.H.G., Silva, E. A., Guerreiro, M.C.; Curi, N. 2013. MACRO simulator (Version 5.0) for predicting atrazine herbicide behavior in Brazilian Latosols. Ciência e Agrotecnologia 37: 211-220.

Weindorf, D.C.; Bakr, N.; Zhu, Y. 2014. Advances in portable X-ray fluorescence (PXRF) for environmental, pedological, and agronomic application. Advances in Agronomy 128: 1-45.

Weindorf, D.C.; Chakraborty, S. 2016. Portable X-ray fluorescence spectrometry analysis of soils. Methods of Soil Analysis. DOI:10.2136/methods-soil.2015.0033.

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.