University of São Paulo
"Luiz de Queiroz" College of Agriculture

VNIR, XRF, and LIBS spectroscopies for soil sensing on precision agriculture

Tiago Rodrigues Tavares

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Systems Engineering

.

Piracicaba
2021

Tiago Rodrigues Tavares
Agronomist

VNIR, XRF, and LIBS spectroscopies for soil sensing on precision agriculture

Advisor:
Prof. Dr. **JOSÉ PAULO MOLIN**
Coadvisor:
Prof. Dr. **ABDUL MOUNEM MOUAZEN**

Thesis presented to obtain the degree of Doctor in Science. Area: Agricultural Systems Engineering

Piracicaba
2021

Dados Internacionais de Catalogação na Publicação
DIVISÃO DE BIBLIOTECA – DIBD/ESALQ/USP

Tavares, Tiago Rodrigues

VNIR, XRF, and LIBS spectroscopies for soil sensing on precision agriculture / Tiago Rodrigues Tavares. - - Piracicaba, 2021.

164 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz de Queiroz”.

1 Química verde 2. Análise da fertilidade do solo 3. Laboratórios híbridos 4. Sistemas sensores inteligentes I. Título

ACKNOWLEDGMENTS

I would like to thank the following people and institutions who contributed to the well-development of this thesis:

- University of São Paulo - "Luiz de Queiroz" College of Agriculture (USP-ESALQ) and Agricultural Systems Engineering Graduate Program (PPGESA) for providing the infrastructure and the environment necessary to get my Doctor in Science degree.
- My adviser prof. Dr. José Paulo Molin for accepting me to be part of his research group; for the several opportunities presented to me during my doctorate; for believing, supporting and polishing my ideas; for giving me the autonomy necessary to consolidate the partnerships with the different research groups that were fundamental for this work; for your daily effort to conduct the Laboratory of Precision Agriculture (LAP), a center of excellence in research in its field, which enabled me to perceive new horizons during my doctorate.
- Prof. Dr. Hudson Wallace Pereira de Carvalho, head of the Laboratory of Analytical Chemistry (LIN), Center for Nuclear Energy in Agriculture (USP-CENA), for the knowledge shared during the course of Principles of Electronic and Nuclear Spectroscopy, essential for my growth in this area; for facilitating me to use the LIN's infrastructure; for the invaluable and unique productive provocations that somehow polished my logical reasoning; for the tips for scientific writing; for the friendly atmosphere developed during the many days I was at the LIN.
- My coadvisor prof. Dr. Abdul M. Mouazen for accepting me as a guest student in his research group; for his tireless reviews of my papers; for the numerous meetings that helped me to systematize a didactic sequence for the information obtained in this work. Thank you very much!
- Prof. Dr. Francisco José Krug, head of the Laboratory of Analytical Chemistry (LQA), Center for Nuclear Energy in Agriculture (USP-CENA), for sharing his enormous experience in a kind and motivating way, as well as for relying on me to use the LQA's equipment for sample preparation and LIBS analysis.
- Dra. Lidiane Cristina Nunes and Dr. Elton Eduardo Novais Alves for the numerous and important conversations and suggestions, always conducted in a kind and friendly way.
- Ghent University - Faculty of Bioscience Engineering, as well as the Precision Soil and Crop Engineering Group (Precision SCoRing), to accept me as a guest student, sharing valuable information for conducting the data analysis of this work.
- Prof. Dr. Fábio L. Melquiades and Dr. Felipe dos Santos Rodrigues, from Laboratory of Applied Nuclear Physics - Londrina State University (UEL), for their kind contribution regarding the methods used for modeling the spectral data.

- Prof. Dr. Paulo S. Pavinato for the knowledge shared about soil fertility and his helpful suggestions regarding total content analyses.
- Technicians Fátima Patreze and Liz Mary Bueno de Moraes, from Analytical Chemistry Laboratory (USP-CENA), for the support with the sample preparation; the technician Eduardo de Almeida, from Laboratory of Nuclear Instrumentation (USP-CENA), for the support with X-ray fluorescence equipment; and, also the technician Marina Colzato, from the Laboratory of Environmental Analysis (USP-ESALQ), for the support with soil analyses.
- My friends Dr. Mateus T. Eitelwein and prof. Dr. Maurício Cherubin for the tips shared at the beginning of the doctorate, fundamental for writing properly the research project.
- São Paulo Research Foundation (FAPESP), grant number 2017/21969-0; and also financial support promoted, at the beginning of the doctorate, by the Brazilian Federal Agencies: Coordination for the Improvement of Higher Education Personnel (CAPES) – Finance Code 001, and the National Council for Scientific and Technological Development (CNPq).
- My friend Victor Monseff de A. Campos, a member of the Nuffield Brazil Association (NuffieldBR) and the Ribersolo team, for the valuable discussions on the challenges and perspectives of proximal soil sensing in the Brazilian scenario.
- My professors, from the Federal University of Goiás - College of Agronomy (UFG-EA), prof. Dr. João Gaspar Farias and prof. M.Sc. Juliano Santana, for supporting my first steps in scientific research.
- All friends from LIN and LAP, including the guys from the mechanization and precision agriculture group (gMAP), for the pleasant working environment of the last years.
- To all employees of the Department of Biosystems Engineering, for cordiality and support during the daily activities, especially to Agnaldo F. Degaspari, Áureo S. de Oliveira, Davilmar A. D. Colevatti, Francisco de Oliveira (Chicão), José G. Gomes, and Juarez R. do Amaral.
- To Ana Paula Rotelli, Magali, and Bento Rodrigues for helping to handle the pandemic times that marked the end of the writing of this thesis.
- All my friends who somehow have been with me during those years, allowing life to go lighter. You guys were all very important too!
- My mom Kênia, an inspiration of strength, commitment and coherence, and my grandma Terezinha, for her unconditional love. This thesis is dedicated to you two!

Thank you all!!

“We have to do the best we can. This is our sacred human responsibility”

Albert Einstein (1879 – 1955)

"Vanity can turn into neurosis when, for example, we do not realize that we have too singular reasons to defend general principles"

Contardo Calligaris (1948 – 2021)

CONTENTS

RESUMO

Espectroscopias VNIR, XRF e LIBS para o sensoriamento do solo na agricultura de precisão

A espectroscopia de reflectância difusa no visível e infravermelho-próximo (VNIR), a espectroscopia de fluorescência de raios X (XRF) e a espectroscopia de emissão óptica com plasma induzido por laser (LIBS) são técnicas promissoras para uma caracterização rápida e ambientalmente limpa da fertilidade do solo. A integração destas técnicas com os métodos analíticos já estabelecidos de análise da fertilidade do solo pode permitir avanços na modernização de seus procedimentos analíticos, permitindo, *e.g.*, análises *in-situ* e em laboratoriais móveis, o que, por sua vez, conduziria a um novo paradigma em manejo do solo e em abordagens de agricultura de precisão. O principal objetivo desta tese é desenvolver estratégias para utilizar espectros gerados por sensores VNIR, XRF e LIBS na calibração de algoritmos agronômicos para prever os principais atributos de fertilidade do solo [argila, matéria orgânica (OM), capacidade de troca catiônica (CEC), pH, saturação de bases (V) e nutrientes extraíveis (ex-P, ex-K, ex-Ca e ex-Mg)]. Foi utilizado um conjunto de 102 amostras de solo coletadas em dois talhões agrícolas brasileiros. Nas primeiras etapas desta tese, foram conduzidos estudos para (i) simplificar o preparo de amostras de solo para análises com XRF, (ii) desenvolver um método simples e transparente para aquisição e processamento de dados de XRF, e (iii) encontrar um método preciso e eficiente para modelagem de dados de LIBS. Em seguida, foram comparados os desempenhos de predição individuais e combinados dos sensores VNIR, XRF e LIBS. Os resultados mostraram que, de modo geral, os melhores desempenhos preditivos se dão na seguinte sequencia: LIBS > XRF > VNIR. VNIR confirmou ser a melhor técnica para a predição do conteúdo de argila e OM [2,61 ≤ desvio residual da predição (RPD) ≤ 3,37], enquanto os atributos químicos CEC, V, ex-P, ex-K, ex-Ca e ex-Mg foram melhor preditos (1,82 ≤ RPD ≤ 4,82) pelas técnicas de análise elementar (*e.g.*, XRF e LIBS). Em relação às abordagens multi-sensor, a qualidade da predição diminuiu na seguinte ordem: VNIR + XRF + LIBS > XRF + LIBS > VNIR + LIBS > VNIR + XRF. Nossos resultados indicam que não há uma combinação de sensores ideal única para prever todos os principais atributos de fertilidade do solo e que esta combinação ótima deve ser específica de cada atributo. Contudo, ainda é cedo para esta afirmação, sendo necessário avaliar conjuntos de dados maiores que incluam outros tipos de solos, com variabilidade na mineralogia, classes texturais, faixa de concentração de atributos, e submetidos a diferentes práticas agrícolas. No último capítulo, avaliamos o potencial de aplicação do XRF para análises *in-situ*, concentrando-nos na avaliação do efeito tanto do conteúdo de umidade do solo quanto do tempo de análise sobre o desempenho dos sensores. Nossos resultados provaram que é possível fazer reduções drásticas no tempo de análise (de 90 para 2s) mantendo desempenhos satisfatórios (RPD ≥ 1,92) para a predição de atributos de fertilidade. Além disso, mostramos que embora o XRF seja menos sensível ao aumento de umidade do solo quando comparado ao sensor VNIR, a presença de água afeta seu desempenho preditivo de modo que métodos para

mitigar o efeito da umidade devem ser considerados para análises *in-situ* mais precisas. Esta tese trouxe alguns avanços para a aplicação dos sensores VNIR, XRF e LIBS em solos tropicais brasileiros como métodos analíticos rápidos e limpos para caracterizar atributos de fertilidade. As vantagens e desvantagens destas técnicas foram destacadas e apontamentos foram deixados para as próximas pesquisas que continuarão o amadurecimento tecnológico em direção a modernização dos diagnósticos de fertilidade do solo.

Palavras-chave: Química verde, Análise da fertilidade do solo, Laboratórios híbridos, Sistemas sensores inteligentes

ABSTRACT

VNIR, XRF, and LIBS spectroscopies for soil sensing on precision agriculture

Visible and near infrared diffuse reflectance spectroscopy (VNIR), X-ray fluorescence spectroscopy (XRF), and laser-induced breakdown spectroscopy (LIBS), are promising techniques for rapid and environmentally-friendly soil fertility characterisation. Integrating these sensing tools with the already established analytical methods can enable advances towards the modernisation of the traditional analytical procedures by allowing, *e.g.*, *in-situ* and mobile laboratory analysis, which, in turn, will enable new paradigm in soil management using precision agriculture approaches. The main objective of this thesis is to develop strategies for using the output generated by VNIR, XRF, and LIBS sensors in the calibration of agronomic algorithms for predicting key soil fertility attributes [clay, organic matter (OM), cation exchange capacity (CEC), pH, base saturation (V), and extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg)]. A set of 102 soil samples collected from two Brazilian agricultural fields was used. In the first stages, studies were conducted to (i) simplify the soil sample preparation for XRF sensor analysis, (ii) develop a simple and transparent method for XRF data acquisition and processing, and (iii) find an accurate and efficient method for LIBS data modeling. Afterwards, the individual and combined prediction performances of VNIR, XRF, and LIBS were compared. The results showed that overall the predictive performance decreased as follows: LIBS > XRF > VNIR. VNIR confirmed to be the best technique for the prediction of clay and OM content [2.61 ≤ residual prediction deviation (RPD) ≤ 3.37], while the chemical attributes CEC, V, ex-P, ex-K, ex-Ca, and ex-Mg were better predicted (1.82 ≤ RPD ≤ 4.82) by elemental analysis techniques. Regarding multi-sensor approaches, the prediction quality decreased in the order of: VNIR + XRF + LIBS > XRF + LIBS > VNIR + LIBS > VNIR + XRF. Our results indicate that there is no unique optimal sensor combination for predicting all the key soil fertility attributes and that this is attribute specific. However, we also consider that it is too soon to establish a clear trend, which requires further research evaluating larger data sets including other soil types, with greater variability in soil mineralogy, textural classes, attributes concentration range, and submitted to different agricultural practices. Lastly, we evaluated the potential of applying XRF for *in-situ* analysis, focusing on the effect of both soil moisture content and scanning time on the sensors' performance. Our results proved that it is possible to make drastic reductions in scanning time (from 90 to 2s) maintaining satisfactory performances (RPD ≥ 1.92) for predicting fertility attributes. Moreover, we showed that although XRF is less sensitive to soil moisture increment when compared to VNIR sensor, the presence of water impacts on the XRF's predictive performance and methods for mitigating soil moisture effect should be applied aiming at more accurate *in-situ* analysis. This thesis brought some advances for the application of VNIR, XRF, and LIBS sensors in Brazilian tropical soils as fast and clean analytical methods for characterizing fertility attributes. The

advantages and drawbacks of these techniques were highlighted and directions were left for future research that will continue the technological maturation toward the modernization of soil fertility diagnostics.

Keywords: Green chemistry, Soil fertility testing, Hybrid laboratories, Smart sensor systems

CHAPTER 1. GENERAL INTRODUCTION

Smart sensors are essential for the introduction of Information and Communication Technology in agriculture. They are the main providers of information that allows to intensify the monitoring of key variables in agricultural production systems (Bacco et al., 2019). It is important to mention that applying sensors directly in the field for soil and plant assessment is a classic tool of precision agriculture (PA), present in its approaches since the early years of PA development (Hummel et al., 1996). Nevertheless, recent technological advances have expanded the application of smart sensors in agriculture (Abbasi et al., 2014). The intensification of monitoring capabilities and the big amounts of data produced by smart sensors provide unprecedented decision-making capabilities (Wolfert et al., 2017), which, in turn, enables further advancement of PA approaches (Gebbers and Adamchuk, 2010).

PA approaches that seek to optimize the use of soil inputs [*e.g.*, variable rate (VR) applications of lime and fertilizers] have great potential to boost agronomic and environmental benefits in agricultural production systems (Mouazen et al., 2020). For a successful implementation of VR fertilization is fundamental a detailed characterisation of key soil fertility attributes in the field. A proper soil fertility management in Brazil is particularly important because we are the fourth largest consumer of fertilizers in the world (FAO, 2017) due to the predominance of acidic and low fertility tropical soils. Per year, it is estimated that about 4 million soil tests are carried out by Brazilian fertility analysis laboratories (Demattê et al., 2019). In addition, traditional soil analyses face other challenges related to the time required for performing the laboratory measurements (about 3 to 15 days), and also the hazardous reagents still used in some tests (*e.g.*, dichromate and sulfuric acid).

The establishment of a robust sensor-based method for soil fertility analysis—allowing farmers and laboratories to increase the number of analyses in a practical and clean way, without relying exclusively on traditional fertility soil tests—is a current need in Brazil and all over the world [*e.g.*, South America (Sanches et al., 2018), North America (Suduth et al., 2013), Europe (Riebe et al., 2019), Africa (Johnson et al., 2019), Asia (Xu et al., 2019), Oceania (O'Rourke et al., 2016)]. Integrating sensing techniques with laboratory analytical methods would modernise the traditional methods for soil analysis that will enable new paradigm in soil site-specfic management (Viscarra Rossel and Bouma, 2016). In order to create simple and environmentally-friendly alternatives for soil fertility analysis, efforts have been made worldwide to adapt and re-engineer sensor systems developed in other areas into soil sensing (Ji et al., 2016; Archbold et al., 2019), as well as assessing different combinations of sensors and data fusion methods (Weindorf et al., 2016; Xu et al., 2019; Zhang et al., 2019). Figure 1 presents a conceptual framework regarding how smart soil sensors working alone and/or integrated with multi-causal decision-making systems can drive localized input management in smart farms.

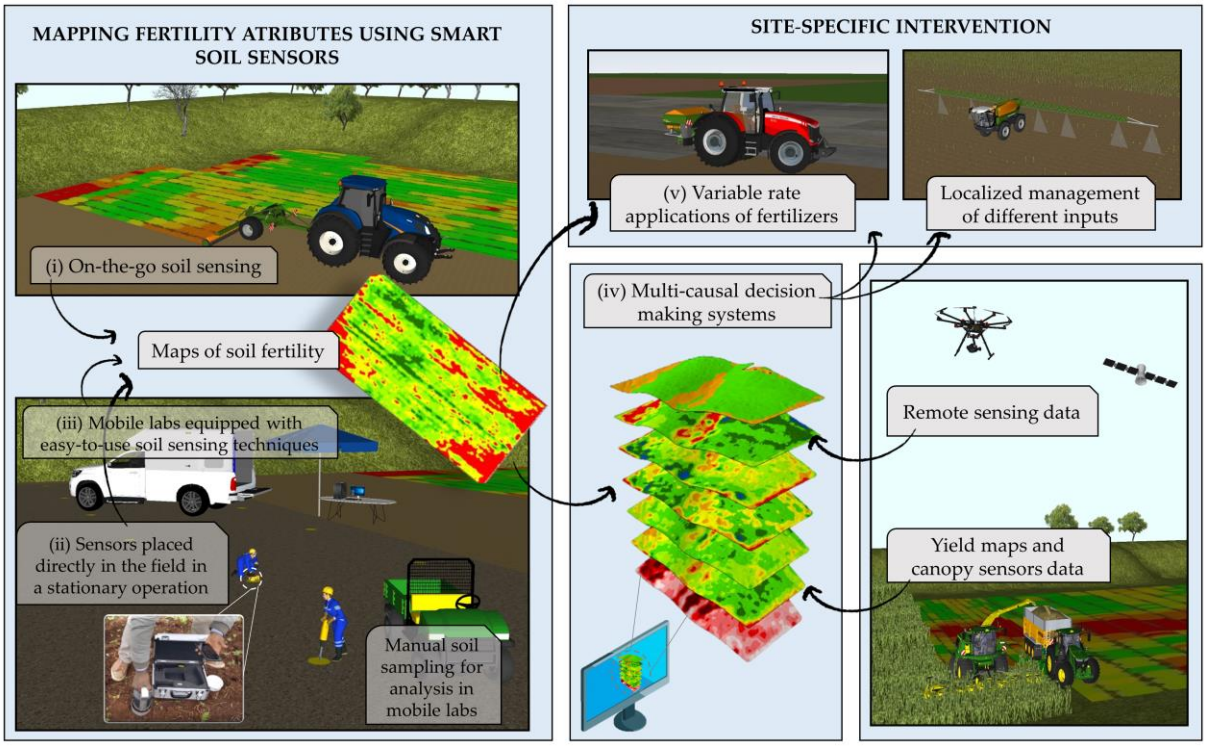


Figure 1. Conceptual framework for using smart soil sensors in smart farms. Smart soil sensors being used in different data acquisition fronts [(i) embedded in a mobile platform; (ii) equipament operated in a stationary way via manual or robot operations; and (iii) on benchtop inside mobile or hybrid laboratories], (iv) providing information for variable rate applications of fertilizers and lime, and (v) providing soil maps for multi-causal decision-making systems.

In the PA context, the sensors most addressed by the scientific literature for soil fertility prediction are visible, near-infrared (VNIR), and mid-infrared (Mid-IR) reflectance sensors, apparent electrical conductivity (ECa), and ion-selective electrodes (ISE). The first Brazilian studies evaluating VNIR and ECa sensors were published in the mid-2000s (Molin et al., 2005; Nanni and Demattê, 2006), and about 10 years later the first study evaluating ISE sensors was published (Silva and Molin, 2018). This shows that we are in the first stages of development of soil sensing applications. In addition, evaluating this literature it can be concluded that overall using VNIR, Mid-IR, and ECa sensors it is possible to predict texture and OM, and, in specific cases, CEC, pH, and extractable nutrients, while when using ISE it is possible infering about specific ions such as K^+, and H^+ (for pH estimations). Nevertheless, consistent predictions of soil chemical attributes, mainly extractable nutrients, are still a challenge.

Hence, it is now necessary to evaluate "new" sensing techniques that have been little explored for fertility attributes analysis. Recently, great attention from scientists commited with PA have been given for laser-induced breakdown spectroscopy (LIBS) (*e.g.*, Riebe et al., 2019; Erler et al., 2020) and X-ray fluorescence (XRF) (*e.g.*, Lima et al., 2019; Nawar et al., 2019), both sensors of elemental analysis that

are compatible with the direct analysis of soils, *e.g.* requiring minimal or no sample preparation. Both techniques have been employed for decades as laboratory analytical methods (Krug and Rocha, 2016), and now have evolved to equipment with reduced weight and size, greater robustness, and more efficient components (*e.g.*, detectors with better sensitivity), making them compatible with on-field operations (Galuszka et al., 2015). The XRF and LIBS are spectroanalytical techniques that measure the total content of a wide range of elements (*e.g.*, Si, Fe, Ca, K, Ti, Al, among others), which can be used as a proxy for infering fertility attributes. Another interesting feature of XRF and LIBS sensors is their complementary character to the VNIR sensor (O'Rourke et al., 2016), wich is a widespread technique in soil science (Stenberg et al., 2010) with extensive research reporting its potential to predict mineralogical and organic attributes successfully (Demattê et al., 2004; Kuang et al., 2012). In Brazilian soils, although scientific research using XRF and LIBS sensors for soil fertility analysis has expanded in recent years (Ferreira et al., 2015; Villas-Boas et al., 2016; Lima et al., 2019; Silva et al., 2019; Andrade et al., 2020; Santos et al., 2020), undoubtedly, we are still in the early stages of development of this application. In addition, some important aspects for applicating these sensors in PA have been little explored so far, *e.g.*, simplification of sample preparation, strategies for mitigating the matrix effect, synergy when combining different sensing technologies, effect of moisture content on sensor's performance, are few to mention among others.

In this scenario, the main task of the present thesis is to take advantage of the current knowledge regarding the use of VNIR, XRF, and LIBS sensors and adapt it by creating agronomic algorithms for predicting key soil fertility attributes in order to use these sensors as tools for the management of soil fertility. For that, the specific objectives were to (i) establish and validate prediction models of key soil fertility attributes using data of VNIR, XRF and LIBS and laboratory conventional analytical results, observing the relations between sensors' output and soil attributes (*e.g.*, relation between total and extractable contents of nutrients); (ii) evaluate the potential and applicability of data-fusion techniques to explore the synergism between elemental analysis sensors (*e.g.*, XRF and LIBS) and VNIR sensors to improve the prediction accuracy of soil fertility attributes; and (iii) evaluate the applicability of XRF sensors for *in-situ* analysis, investigating the effect of soil moisture content and scanning time on the sensors' performance. The hypothesis tested were: (i) elemental analysis sensors present greater potential for predicting soil chemical attributes compared to VNIR sensor; (ii) the combination of data provided by VNIR sensors with those from elemental analysis sensors will provide more comprehensive and accurate information about key fertility attributes compared to the single-sensor approaches; and (iii) although there are some critical factor that influences XRF's prediction accuracy in applications directly in the field (*e.g.*, sample moisture content and XRF scanning time), it is possible to establish strategies to maintain a satisfactory sensor performance.

To meet our objectives, this thesis is organized into eight chapters. Briefly, the **first** one presents a short introduction about the research topic studied. The **second** one addresses the theoretical framework for applying soil sensors in precision agriculture to map fertility attributes, focusing mainly on the spectroanalytical techniques explored in this thesis. The **third** one proposes a simplification of sample preparation for soil fertility analysis via XRF sensor, wich is compatible with the one already established for soil analysis using VNIR sensors. In the **fourth** one, it was established an optimized procedure for XRF data acquisition and processing to assess soil fertility attributes. The **fifth** proposes an optimised data analysis for modeling LIBS data to predict soil fertility attributes, also addressing the potential and limitation of this technique in the context of PA. The **sixth** one compares the individual and combined prediction performance of VNIR, XRF, and LIBS sensors for assessing fertility atttributes in Brazilian tropical soils. The **seventh** evaluates the effects of soil moisture content and the scanning time reduciton on the XRF performance for predicting fertility attributes, providing the first guidelines for *in-situ* applications. Finally, the **eigth** one provides the final considerations.

The main findings obtained during the execution of this thesis have resulted in the following scientific papers.

Molin, J.P.; Tavares, T.R. Sensor systems for mapping soil fertility attributes: challenges, advances, and perspectives in Brazilian tropical soils. *Engenharia Agrícola* **2019**, 39(SPE), pp.126-147, doi: 10.1590/1809-4430-eng.agric.v39nep126-147/2019.

Tavares, T.R.; Nunes, L.C.; Alves, E.E.N.; Almeida, E.D.; Maldaner, L.F.; Krug, F.J.; Carvalho, H.W.P.D.; Molin, J.P. Simplifying Sample Preparation for Soil Fertility Analysis by X-ray Fluorescence Spectrometry. *Sensors* **2019**, 19(23), p.5066, doi: 10.3390/s19235066.

Tavares, T.R.; Molin, J.P.; Nunes, L.C.; Alves, E.E.N.; Melquiades, F.L.; Carvalho, H.W.P.; Mouazen, A.M. Effect of X-ray tube configuration on measurement of key soil fertility attributes with XRF. *Remote Sensing* **2020**, 12, 963, doi: 10.3390/rs12060963.

Tavares, T.R.; Mouazen, A.M.; Alves, E.E.N.; dos Santos, F.R.; Melquiades, F.L.; Carvalho, H.W.P.; Molin, J.P. Assessing Soil Key Fertility Attributes Using a Portable X-Ray Fluorescence: A Simple Method to Overcome Matrix Effect. *Agronomy* **2020**, 10, 787, doi: 10.3390/agronomy10060787.

Tavares, T. R.; Molin, J. P.; Javadi, S. H.; Carvalho, H. W. P. D.; Mouazen, A. M. Combined Use of Vis-NIR and XRF Sensors for Tropical Soil Fertility Analysis: Assessing Different Data Fusion Approaches. *Sensors* **2021**, 21(1), 148, doi: 10.3390/s21010148.

Tavares, T. R.; Molin, J. P.; Nunes, L. C.; Wei, M. C. F.; Krug, F. J.; Carvalho, H. W. P. D.; Mouazen, A. M. Multi-sensor approach for tropical soil fertility analysis: comparison of individual and combined performance of VNIR, XRF, and LIBS spectroscopies. *Agronomy* **2021** (accepted).

Tavares, T.R.; Mouazen, A.M.; Nunes, L.C.; Santos, F.R.; Melquiades, F.L.; Silva, T. R.; Krug, F. J.; Molin, J.P. Laser-Induced Breakdown Spectroscopy (LIBS) for tropical soil fertility analysis. *Soil & Tillage* (submitted).

References

Abbasi, A.Z.; Islam, N.; Shaikh, Z.A. A review of wireless sensors and networks' applications in agriculture. *Comput. Stand. Interfaces* **2014**, 36(2), 263-270, doi: 10.1016/j.csi.2011.03.004.

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

Archbold, G.; Torres, H.B.; Ruiz, F.; Marín, M.N.; Chaves, D.M.; Arboleda, L.T.; Parra, C.; Carrillo, H.; Mouazen, A.M. pH Measurement IoT System for Precision Agriculture Applications. *IEEE Lat. Am. Trans.* **2019**, 17, 823–832, doi:10.1109/TLA.2019.8891951.

Bacco, M.; Barsocchi, P.; Ferro, E.; Gotta, A.; Ruggeri, M. The digitisation of agriculture: a survey of research activities on smart farming. *Array* **2019**, 3, 100009, doi: 10.1016/j.array.2019.100009.

Demattê, J.A.; Campos, R.C.; Alves, M.C.; Fiorio, P.R.; Nanni, M.R. Visible–NIR reflectance: a new approach on soil evaluation. *Geoderma* **2004**, 121, 95-112.

Demattê, JAM.; Dotto, AC.; Bedin, LG.; Sayão, VM.; Souza, AB. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. *Geoderma* **2019**, 337, 111-121. doi: 10.1016/j.geoderma.2018.09.010.

Erler, A.; Riebe, D.; Beitz, T.; Löhmannsröben, H.G.; Gebbers, R. Soil Nutrient Detection for Precision Agriculture Using Handheld Laser-Induced Breakdown Spectroscopy (LIBS) and Multivariate Regression Methods (PLSR, Lasso and GPR). *Sensors* **2020**, 20(2), 418, doi: 10.3390/s20020418.

FAO - Food and Agriculture Organization of the United Nations. *World Fertilizer Trends and Outlook to 2020*. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, 2017.

Ferreira, EC.; Neto, J.A.G.; Milori, D.M.; Ferreira, E.J.; Anzano, J.M. Laser-induced breakdown spectroscopy: Extending its application to soil pH measurements. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2015**, 110, 96-99. doi: 10.1016/j.sab.2015.06.002

Galuszka, A.; Migaszewski, ZM.; Namieśnik, J. Moving your laboratories to the field–Advantages and limitations of the use of field portable instruments in environmental sample analysis. *Environmental research* **2015**, 140, 593-603. doi 10.1016/j.envres.2015.05.017.

Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, 327, 828-831. doi: 10.1126/science.1183899.

Hummel, J. W.; Gaultney, L. D.; Sudduth, K. A. (1996). Soil property sensing for site-specific crop management. *Comput. Electron. Agric.* **1996**, 14(2-3), 121-136, doi: 10.1016/0168-1699(95)00043-7.

Ji, W.; Adamchuk, V.I.; Biswas, A.; Dhawale, N.M.; Sudarsan, B.; Zhang, Y.; Rossel, R.A.; Shi, Z. Assessment of soil properties *in-situ* using a prototype portable MIR spectrometer in two agricultural fields. *Biosyst. Eng.* **2016**, *152*, 14–27, doi 10.1016/j.biosystemseng.2016.06.005.

Johnson, J.M., Vandamme, E., Senthilkumar, K., Sila, A., Shepherd, K.D.; Saito, K. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. *Geoderma* **2019**, 354, 113840, doi: 10.1016/j.geoderma.2019.06.043.

Krug, F.J.; Rocha, F.R.P. *Métodos de preparo de amostras para análise elementar*. EditSBQ, Sociedade Brasileira de Química, São Paulo, 2016. p.572.

Kuang, B.; Mahmood, HS.; Quraishi, MZ.; Hoogmoed, WB.; Mouazen, AM.; Henten, EJV. Sensing soil properties in the laboratory, *in-situ*, and on-line: a review. *Advances in Agronomy* **2012**, 114, 155-223p.

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

Molin, J.P.; Gimenez, L.M.; Pauletti, V.; Schmidhalter, U.; Hammer, J. Mensuração da condutividade elétrica do solo por indução e sua correlação com fatores de produção. *Engenharia Agrícola* **2005**, 25(2), 420-426.

Mouazen, A. M.; Alexandridis, T.; Buddenbaum, H.; Cohen, Y.; Moshou, D.; Mulla, D.; ... & Sudduth, K. A. Monitoring. In *Agricultural Internet of Things and Decision Support for Precision Smart Farming* ; Castrignanò, A.; Buttafuoco, G.; Khosla, R.; Mouazen, A.; Moshou, D.; Naud, O., Ed.; Academic Press: London, United Kigdom, 2020; pp. 35-138.

Nanni, M.R.; Demattê, J.A.M. Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Science Society of America Journal* **2006**, 70, 393-407, doi: 10.2136/sssaj2003.0285.

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

O'Rourke, SM.; Stockmann, U.; Holden, NM.; McBratney, A.B.; Minasny B. An assessment of model averaging to improve predictive power of portable VNIR and XRF for the determination of agronomic soil properties. *Geoderma* **2016**, 279, 31-44. doi: 10.1016/j.geoderma.2016.05.005.

Riebe, D.; Erler, A.; Brinkmann, P.; Beitz, T.; Löhmannsröben, H.G.; Gebbers, R. Comparison of Calibration Approaches in Laser-Induced Breakdown Spectroscopy for Proximal Soil Sensing in Precision Agriculture. *Sensors* **2019**, 19(23), 5244, doi: 10.3390/s19235244.

Sanches, G.M.; Magalhães, P.S.; Remacre, A.Z.; Franco, H.C. Potential of apparent soil electrical conductivity to describe the soil pH and improve lime application in a clayey soil. *Soil and Tillage Research* **2018**, 175, 217-225, doi: 10.1016/j.still.2017.09.010.

Santos, F.R.; Oliveira, J.F.; Bona, E.; Santos, J.V.F.; Barboza, G.M.; Melquiades, F.L. EDXRF spectral data combined with PLSR to determine some soil fertility indicators. *Microchem. J.* **2020**, 152, 104275, doi:10.1016/j.microc.2019.104275.

Silva, F.C.D.S.; Molin, J.P. On-the-go tropical soil sensing for pH determination using ion-selective electrodes. *Pesquisa Agropecuária Brasileira* **2018**, 53(11), 1189-1202, doi: 10.1590/s0100-204x2018001100001.

Silva, E.A.; Weindorf, D.C.; Silva, S.H.; Ribeiro, B.T.; Poggere, G.C.; Carvalho, T.S.; Goncalves, M.G.; Guilherme, L.R.; Curi, N. Advances in Tropical Soil Characterisation via Portable X-Ray Fluorescence Spectrometry. *Pedosphere* **2019**, 29, 468–482, doi:10.1016/S1002-0160(19)60815-5.

Stenberg, B.; Viscarra Rossel, RA.; Mouazen, A.M.; Wetterlind, J. Visible and near-infrared spectroscopy in soil science. *Advances in Agronomy* **2010**, 107, 163-215. doi: 10.1016/S0065-2113(10)07005-7.

Sudduth, K.A.; Myers, D.B.; Kitchen, N.R.; Drummond, S.T. Modeling soil electrical conductivity–depth relationships with data from proximal and penetrating ECa sensors. *Geoderma* **2013**, 199, 12-21, doi: 10.1016/j.geoderma.2012.10.006.

Villas-Boas, P.R.; Romano, R.A.; Franco, M.A.M.; Ferreira, E.C.; Ferreira, E.J.; Crestana, S.; Milori, D.M.B.P. Laser-induced breakdown spectroscopy to determine soil texture: A fast analytical technique. *Geoderma* **2016**, 263, 195-202, 2016.

Viscarra Rossel, R.A.; Bouma, J. Soil sensing: A new paradigm for agriculture. *Agric. Syst.* **2016**, 148: 71-74. doi: 10.1016/j.agsy.2016.07.001

Weindorf, D.C.; Chakraborty, S.; Herrero, J.; Li, B.; Castañeda, C.; Choudhury, A. Simultaneous assessment of key properties of arid soil by combined PXRF and Vis–NIR data. *Eur. J. Soil Sci.* **2016**, 67, 173–183, doi:10.1111/ejss.12320.

Wolfert, S.; Ge, L.; Verdouw, C.; Bogaardt, M. J. Big data in smart farming–a review. *Agric. Syst.* **2017**, 153, 69-80, doi: 10.1016/j.agsy.2017.01.023.

Xu, D.; Zhao, R.; Li, S.; Chen, S.; Jiang, Q.; Zhou, L.; Shi, Z. Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. *Eur. J. Soil Sci.* **2019**, 70, 162–173.

Zhang, Y.; Hartemink, A.E. Data fusion of vis–NIR and PXRF spectra to predict soil physical and chemical properties. *Eur. J. Soil Sci.* **2019**, doi:10.1111/ejss.12875.

Connecting text for chapter 2

Chapter 2 brings up the historical context of soil fertility mapping in precision agriculture, rescuing the problematic that culminated in the introduction of sensor systems as a more efficient technological alternative for this purpose. This chapter makes a bibliographic review of using sensor systems in the context of precision agriculture, pointing out the possibilities of applying sensors to solve the soil sampling dilema and going deeper into the spectroanalytical techniques addressed in this thesis (VNIR, XRF, and LIBS spectroscopies).

An expanded version of this chapter has been published as a review paper in the *Engenharia Agrícola* Journal, as can be seen bellow.

Molin, J.P.; Tavares, T.R. Sensor systems for mapping soil fertility attributes: challenges, advances, and perspectives in Brazilian tropical soils. *Engenharia Agrícola* **2019**, 39, pp.126-147, doi: 10.1590/1809-4430-eng.agric.v39nep126-147/2019.

CHAPTER 2. SENSOR SYSTEMS FOR MAPPING SOIL FERTILITY ATTRIBUTES IN PRECISION AGRICULTURE: A REVIEW OF CHALLENGES, ADVANCES, AND PERSPECTIVES IN BRAZILIAN SOILS

Abstract

Soil fertility attributes have different scales and forms of spatial and temporal variations in agricultural fields. Adequate spatiotemporal characterisation of these attributes is fundamental for a successful development of strategies for variable rate application of fertilizers, enabling the classic benefits of precision agriculture (PA). Studies on Brazilian soils have shown that at least one sample ha^{-1} is required for a reliable mapping of key fertility attributes. However, the operational difficulty of sampling at high spatial density, as well as the cost of laboratory analysis, create a dilemma for this traditionally used soil characterisation sampling in PA. Alternatively, soil sensors have emerged as a practical and complementary tool for obtaining information on soil attributes, allowing analysis at high spatial density, without the production of chemical residues and at a reduced cost. Worldwide, scientists have devoted their attention to the development and application of sensor systems for this purpose. Applications of soil sensing techniques in PA involve different disciplines, such as instrumentation, data science, geostatistics, and soil science. The integration of these disciplines has allowed successful application of proximal sensors for the spatial diagnosis of soil fertility attributes. The current chapter aimed to present a bibliographic review of the concepts related to soil sensing and PA, focusing on the spectroanalytic techniques explored in this thesis (VNIR, XRF and LIBS sensors). We sought to present a broad view of the challenges, advances, and perspectives related to the application of soil sensing in Brazilian tropical soils in the context of PA.

Keywords: 1. Spatial variability 2. Proximal soil sensing 3. On-the-go sensing 4. Hybrid laboratory.

2.1. Introduction

Considered by some authors as one of the top 10 revolutions in agriculture (Crookston, 2006; Mulla, 2013), precision agriculture (PA) uses the technological advances of the last decades (*e.g.*, global positioning satellite systems and variable rate technologies) to create approaches for a proper treatment of spatial and temporal variability of crops (Molin et al., 2015). The objective of PA is to promote an optimized management of agricultural fields regarding its spatiotempoal variations, which increases yields and promotes a rational use of production resources. PA is one of the promising tools to overcome the global challenges of agriculture in the 21st century (Gebbers and Adamchuk, 2010). Challenges that include achieving food and energy security for more than 9 billion people in the coming decades (Godfray et al., 2010) and aligning agricultural production and environmental conservation practices (Foley et al., 2011).

Since PA conception in the mid-1980s, it has been closely associated with the management of soil spatial variability. Prior to the development of variable rate technologies and satellite positioning systems, soil scientists had already undertaken spatial sampling and got intrigued by the high spatial

variability of soil fertility in agricultural fields (Mulla and Khosla, 2016). After evaluating more than 70 agricultural fields in the state of Washington (USA) using intensive soil sampling, Dow et al. (1973a, b) concluded that recommendations for fertiliser application based on the average fertility of agricultural fields may be erroneous. These authors suggested the sampling intensification for a more accurate fertilizer prescription. During the following decade, the development of variable rate systems helped to evaluate the localized management of soil fertility. Regarding these evaluations, results of the pioneering works were positive and encouraging, showing increased yields (Mulla and Hammond, 1988; Hammond, 1993) and improved nutrient use efficiency (Khosla and Alley, 1999).

The traditional management of fertilizers and limestone, without considering the spatial variability, provides inefficient applications of these inputs with an excess or deficit at specific sites. In excessive applications, the nutrients that are not used by plants and retained by the soil are lost to the environment through leaching, runoff, and emissions. Regarding deficit applications, these impose negative influences on the development of crops and, consequently, on their production. Thus, adjusted and spatially accurate nutrient applications can reduce agronomic, economic, and environmental losses (Nawar et al., 2017).

Proper soil management in Brazil is particularly important because we are the fourth-largest fertilizer consumer in the world (FAO, 2017). This has been attributed to the predominance of soils with acid and low fertility character. Thus, not coincidentally, the main PA practice in Brazil is the site-specific management of soils (*e.g.*, correction of its fertility and acidity) based on the traditional mapping of its attributes at the field level. This PA approach is associated with significant amount of service providers and an adoption of approximately 15.3% of Brazilian grain producers, which corresponds to about 9 million ha of mapped soils using the traditional approach, *e.g.*, geo-referenced grid sampling (Molin, 2017).

For the successful implementation of the variable rate application of soil inputs, reliable procedures and technologies are required for the diagnosis of spatial variability of the key soil fertility attributes. The mapping appraoch predominantly used today in Brazil implicates the collection of samples in a regular grid (or occasionally using other arrangements), with geo-referenced sampling points. Then, these samples are sent to laboratories for tradicional fertility analysis and, subsequently, maps are created using interpolation methods. However, as mentioned above, the sampling density is limited by the costs of the laboratory tests and the operational difficulties regarding sampling logistics (McBratney et al., 2003; Demattê et al., 2015a). Previous studies on Brazilian soils have demonstrated that the density traditionally used today [*e.g.*, usually lower than 0.5 samples ha^{-1} (Molin, 2017)] is not adequate for a reliable characterisation of the spatial distribution of most fertility attributes (Nanni et al., 2011; Cherubin et al., 2015). This corroborates with different international studies conducted over

the past few decades (Webster and McBratney, 1987; McBratney and Pringle, 1999; Viscarra Rossel et al., 2011).

Despite the evolution of PA approaches and technologies, the spatial diagnosis of soil fertility attributes has developed slowly and is still a worldwide challenge (Viscarra Rossel and Bouma, 2016). The major challenge is how to increase the density of data on soil fertility attributes without sending more samples for laboratory testing. Soil sensing is a key solution to this problem (Viscarra Rossel and Bouma, 2016); this subject was carefully discussed by Viscarra Rossel et al. (2011). Recently, different applications using sensing techniques has received great attention from scientists commited with PA (Nawar et al., 2017), specialy using new technologies that were still little explored [*e.g.*, laser-induced breakdown spectroscopy (LIBS) (Riebe et al., 2019) and X-ray fluorescence spectroscopy (XRF) (Nawar et al., 2019)] and data fusion approaches between techniques with complementary capabilities [*e.g.*, XRF and visible and near infrared diffuse reflectance spectroscopy (VNIR) (Xu et al., 2019)].

Regarding the above-mentioned context of using sensor systems in the PA, the present chapter aimed at presenting a systematic review of concepts and technologies related to the application of soil sensors for fertility analysis in Brazilian tropical soils. We first characterized the dilemma of soil sampling taking into account research conducted on Brazilian soils. Afterwards, we discussed our vision on different possibilities of applying soil sensors to fill the restrictive information gap for an accurate soil fertility mapping. And lastly, we discussed challenges and advances of some sensing techniques for predicting key soil fertility attributes, giving an special attention to VNIR, XRF, and LIBS spectroscopies; techniques that are further explored in the following chapters of this thesis.

2.2. Diagnosis of spatiotemporal variability of soil fertility: a soil sampling dilema

The spatiotemporal variability of physical and chemical properties, such as the clay, sand, and organic matter (OM) content, cation exchange capacity (CEC), base saturation (V), pH, and extractable (ex-) nutrients, is dynamic, occurring with different amplitudes of variation and spatial patterns. These variations occur according to the classical factors of soil formation (McBratney et al., 2003), and owing to minor alterations (occurring on a small-scale) caused by a combination of local factors, such as relief, management, occurrence of pests and diseases, etc (Viscarra Rossel and Lobsey, 2016). Figure 1 presents images of several factors that are commonly associated with the spatiotemporal variability of soil fertility attributes, explaining the different ranges and patterns of variation in soil attributes.

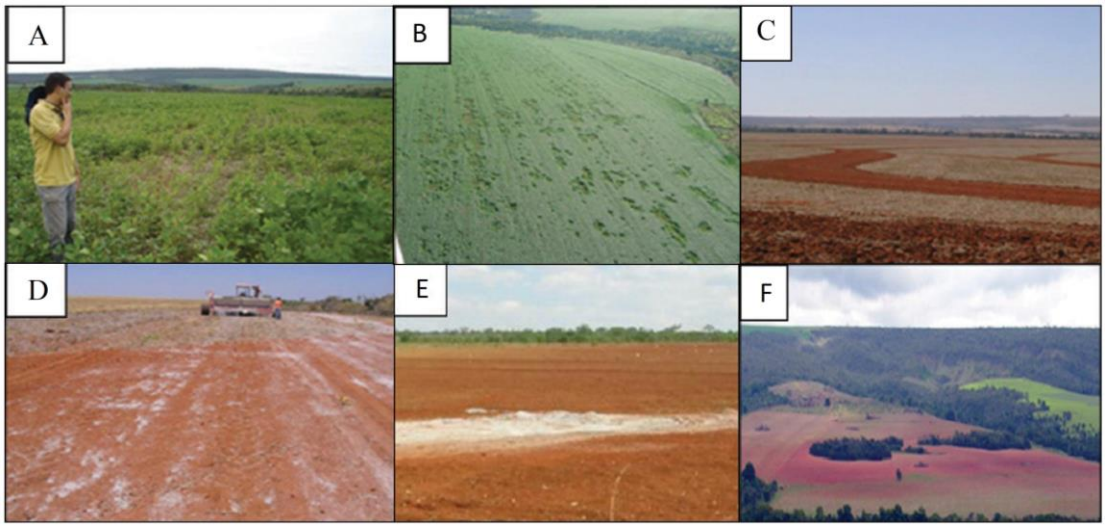


Figure 1. Factors associated with the spatiotemporal variability of soil fertility attributes (adapted from Resende and Coelho, 2017). Damage to growing crops by pests, diseases, or weather promotes heterogeneous nutrient export patterns in fields (A and B); Exposure of the subsoil in contour banks (C); Non-uniform application of fertilizers and lime (D); Limestone deposits (E); and abrupt natural changes in soil formation factors (F).

Understanding the possible ranges and patterns of spatial dependence of the variables to be mapped is fundamental for establishing the sampling density. Sampling must represent a whole statistical population. In the case of the spatial variation of agricultural fields, each "spot" and the transitions between them must be characterized by sampling points to discriminate not only the variation of the soil attribute content but also its spatial pattern (Molin et al., 2015). Soil fertility characterisation was the objective of the first studies related to PA. In the mid-1980s, research conducted by David Mulla at Washington State University evaluated the diagnosis of fertility attributes using different dimensions of sampling grids. The author suggested that the sampling points should be spaced between 30 and 60 m for an accurate representation of soil fertility attributes (reported by Veseth, 1986). In the 1990s, Wollenhaupt et al. (1994) reported that 32 m is the distance that best represented the soil fertility on a regular sample grid. The authors also stated that maps produced with a regular grid of 70 m had significantly reduced accuracy.

Several local studies have characterized the spatial dependence of physical and chemical attributes in Brazilian soils via geostatistical analyses to calculate the best sampling density (Nanni et al., 2011; Montanari et al., 2012; Cherubin et al., 2014a,b; Cherubin et al., 2015). Based on these studies, it is possible to conclude that the geostatistical results vary according to the local characteristics. However, sampling grids greater than 100 × 100 m (one sample ha^{-1}) are in general not efficient for characterizing the variability of most soil fertility attributes. An example of ex-P mapping using different sampling densities is shown in Figure 2. It can also be concluded that attributes related to the soil class and its formation (*e.g.*, texture, mineralogy) require a lower sampling density. On the other hand, pH, nutrients, and other

chemical attributes, that have variability closely related to soil and plant management, require a higher sampling density for reliable mapping (Wetterlind et al., 2010).

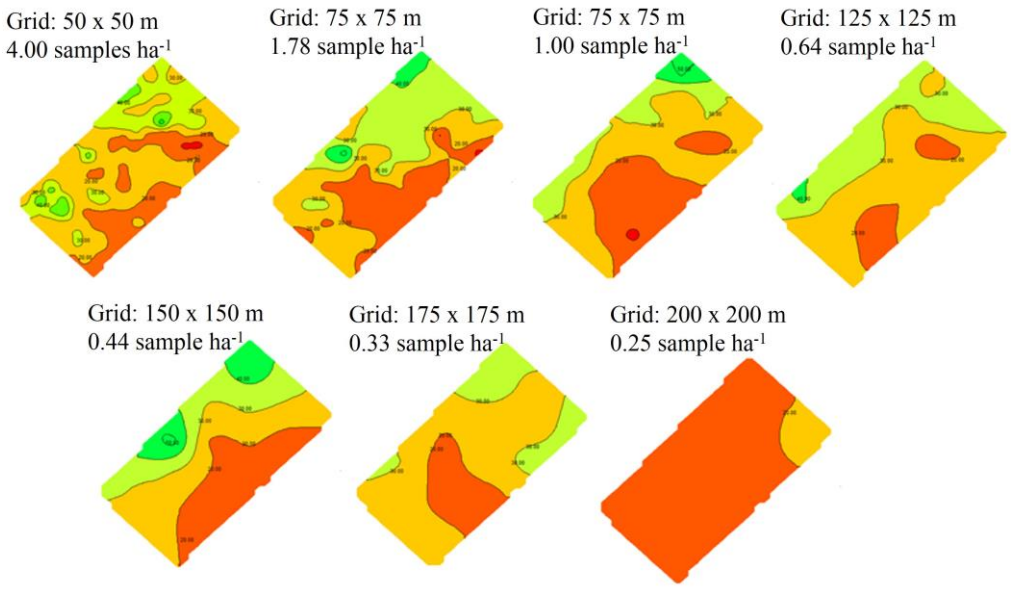


Figure 2. Spatial distribution of ex-P using different sampling densities in a 42-ha field (adapted from Cherubin et al., 2015).

Schirrmann and Domsch (2011) studying the spatial characterisation of fertility attributes (pH, ex-P, ex-K, and ex-Mg) compared a sampling using a 50 m grid-length (four samples ha^{-1}) with another one that simulates the sampling performed by a commercial on-the-go system (Adamchuk et al., 2006). In this study, the evaluated on-the-go system performed *in-situ* sample readings every 25 m (16 samples ha^{-1}). The authors found that the sampling with the higher resolution, which corresponds to the sampling done by the on-the-go system, improved the semivariograms results. Hence, allowing a clear identification of spatial structures of pH, ex-P, and ex-Mg, with higher predictive accuracy of interpolations. Interesting that despite better results with a high-resolution sampling for most of the evaluated attributes, Schirrmann and Domsch (2011) did not achieve good spatial models for the ex-K. According to the authors, the microscale variation of ex-K, with a spatial dependence smaller than 25 m, was limiting for the characterisation of this nutrient using the evaluated grids.

A short spatial dependence, marked by a short-range semivariogram, requires high spatial density sampling to generate reliable spatial diagnostics. A review conducted by Viscarra Rossel and Lobsey (2016) showed that the spatial variability of physical and chemical attributes in topsoils (0 to 20 cm of depth) varied between < 10 and 100 m (Table 1). Furthermore, it is not unusual for agricultural fields to have variations of nutrients in microscale, as reported by Schirrmann and Domsch (2011).

For a proper diagnosis of a specific soil attribute using grid sampling, researchers recommend that the minimum distance between samples should be equal to or less than half of the spatial dependence (Molin et al., 2015). Based on studies of spatial dependence of tropical soil attributes, this would mean a sampling density always higher than one sample ha^{-1}.

Table 1. Spatial and temporal variability of soil attributes (adapted from Viscarra Rossel and Lobsey, 2016).

Soil attribute	Spatial variability (amplitude)	Temporal variability
Texture	Moderate–high (between 20 and 98 m)	Low
Organic carbon	High (between 22 and 78 m)	Moderate (between 5 and 10 years)
CEC[1]	Moderate–high	Moderate
pH	Moderate–high (between 30 and 100 m)	Moderate–high (between growing seasons)
ex-P[2]	High (between < 10 and 31 m)	High (within and between growing seasons)
ex-K[2]	High (between 18 and 68 m)	Moderate–high (within and between growing seasons)

[1] Cation exchange capacity; [2] extractable (ex-) P and K.

2.3. Soil sensors as a solution for soil characterisation in precision agriculture

Soil sensing technologies are promise alternatives to face the soil sampling dilema, allowing rapid, cost-effectively and environmentally friendly analysis of key soil fertility attributes. The development of (i) sensor systems compatible with *in-situ* measurements and/or (ii) portable sensor systems well-matched with easy-to-use analytical procedures are two alternatives that can increase the quantity of soil information in the PA context (Figure 3). The first alternative is related to the proximal soil sensing (PSS) concept (Viscarra Rossel et al., 2011) and would be a complete solution, eliminating the effort of soil sampling, reducing costs with traditional laboratory analyses, and avoiding the use of reagents. On the other hand, the second alternative is related to practical analytical methods that would be executable by users, *e.g.*, user-friendly approaches; an alternative that would be compatible with mobile laboratories (Pandey et al., 2017). This second alternative would provide a partial solution, reducing laboratory costs and the use of reagents during soil analysis, but still requiring the execution of a geo-referenced sampling. An advantage of mobile laboratories is the possibility to bring a simplified laboratory structure close to the sampling site (Figure 3C).

Figure 3. Possibilities of soil sensing applications. (A) Sensors embedded in a mobile platform; (B) Sensors placed directly in the field in a stationary operation (Agrocares, Wageningen, Holland; https://www.agrocares.com/en); (C) Mobile laboratories for conducting simple sample preparation procedures and practical analysis (Agrocares, Wageningen, Holland; https://www.agrocares.com); and some sensor systems compatible with mobile laboratories: (D) laser-induced breakdown spectroscopy system developed by Agrorobótica company (São Carlos, SP, Brazil; https://agrorobotica.com.br) in partnership with Embrapa Agricultural Instrumentation (São Carlos, SP, Brazil), (E) Capillary electrophoresis system developed by Pessl (Weiz, Austria; https://metos.at/imetos-mobilab); and (F) Electrochemical sensor system developed by Ohaus™ (Parsippany, NJ, USA; https://br.ohaus.com).

Another application of sensor systems would be in hybrid laboratories, where sensors would work in an integrated way with traditional methods of soil analysis. According to Demattê et al. (2019), in a hybrid laboratory, part of the samples are analyzed by traditional methods and used to calibrate the predictive models, while the majority of them are analyzed using sensing technologies, using the previously calibrated models to predict the soil attributes. Hybrid laboratories are an interesting alternative to increase the efficiency of laboratory tests with low environmental impact. This should boost Brazilian research in the coming years to seek the best set of sensors compatible with direct analysis to predict soil attributes, as well as the best strategy for the calibration of predictive models at local and regional scales.

In PA, on-field sensor applications are a classic sensing alternative for increasing the density of soil information that has been proposed since the early 1990s for the localized management of agricultural fields (Sudduth and Hummel, 1991; Hummel et al., 1996). Applications of soil sensors that operate in contact or close to the soil surface (around 2 m) and are compatible with *in-situ* analysis define the concept of PSS (Viscarra Rossel et al., 2011). Although this concept excludes remote sensing

applications and measurements with sensors performed inside laboratories using benchtop equipment, such approaches are complementary to PSS because the development of many sensors starts with laboratory tests and calibrations to be used *in-situ* usually derived from laboratory measurements.

All the classic on-the-go sensors that acquire data in a kinematic way (Adamchuk et al. 2004) fit the PSS concept. These sensors can be embedded in agricultural machines, allowing data acquisition during field operations (Steinberger et al., 2009). Recently, companies have launched in the market sensor systems that are adaptable in seeders and tillage equipment, *e.g.*, iScan (Veris Technologies, Salina, USA) and SmartFirmer (Precision Planting, Tremont, USA). This strategy permits the acquisition of soil data during routine operations, as well as allowing real-time adjustments of the plant population and depth for intelligent seeders. These applications, although promising, require local assessments of sensor performance and agronomic algorithms to adjust the recommendations.

In PA, there is a continuous search for sensor systems that are compatible with on-the-go measurement, allowing interventions in real time (Molin et al., 2015). In these applications, data are processed at the same time they are collected, being transformed into decision using agronomic prescriptions. Although these systems tend to be practical and easy to implement in the field, their development requires specific analysis protocols and agronomic models, forming a complex system of intelligence for transforming data into agronomic decision (Weltzien, 2016). In the context of soil sensing, the interpretation of the relationship between sensor output and soil attributes is the main challenge for the development of smart sensor systems.

The assessment of sensing techniques' accuracy is traditionally performed by comparing the prediction given by sensors with that provided by traditional laboratory methods (Kuang et al., 2012). To build robust predictive models, the calibration strategy is fundamental, in which the responses of the sensors will be related to the reference values obtained by the laboratory (Kuang and Mouazen, 2011). Although this is the globally used approach for evaluating new sensing technologies, calibrate predictive models using laboratory determinations as reference has some drawbacks, as discussed by Viscarra Rossel and Bouma (2016). The main disadvantages include (i) propensity of different sources of error while conducting the different stages of analysis, *e.g.*, sampling, sample preparations, and analysis; and (ii) the fact that none of the different extracting solutions used for nutrient analysis produces a value that indeed represents the content available for plants. This second disadvantage happens because the availability of nutrients in the soil is dynamic and depends on multiple factors (*e.g.*, environmental conditions and soil-plant interactions). Regarding the first disadvantage, it is relatively common to observe discrepant results from different laboratories for the same soil sample analyzed using the same procedure. Such observations have been reported by studies conducted in Brazil (Demattê et al., 2019) and overseas (Viscarra Rossel and Bouma, 2016). The calibration of sensor

systems with doubtful laboratory results interferes with the sensor's performance evaluation. This should be considered by researchers, who should be aware of the methodological procedures and quality control of the laboratory they are utilizing. Sending samples to different laboratories could also be a strategy to verify the reliability of determinations.

Another interesting application of soil sensors is the association of their data with high spatial resolution data of yield (yield maps) and in-season development of crops (*e.g.*, vegetation indices obtained by canopy sensors or remote sensing). This approach has an enormous potential for improving on-farm trials methodologies (Viscarra Rossel and Bouma, 2016) and, consequently, improve site-specific recommendation of fertilizers (Trevisan et al., 2021). Such approaches would allow the creation of local databases, based on empirical estimations, and the development of fertilizer strategies specific for each field condition; hence, avoiding the use of highly generalized data, such as regional fertilizer recommendation guides.

2.4. Sensor systems for soil sensing

Some soil sensing technologies are not compatible with the PSS concept because their aplication still require some sample preparation to ensure satisfactory analytical performance (*e.g.*, LIBS and capillary electrophoresis system). However, these sensing technologies can be promising for soil analysis in the context of PA because they are compatible with the above-mentioned mobile and hybrid laboratories. Moreover, the development of faster and more efficient methods of sample preparation—a research area of Analytical Chemistry that has been greatly intensified in recent years (Krug and Rocha, 2016)—should allow the automation of this process in embedded equipment. This would enable applications of such technologies in the field. The study conducted by Sethuramasamyraja et al. (2008) is a classic example of an automated procedure for the collection and preparation of soil samples. The authors adapted a commercial soil sensing plataform, which consists of an on-the-go system using electrochemical sensors (pH Manager, Veris Technologies, USA), for the preparation of an aqueous soil solution before using ion-selective electrodes (ISE). In this approach, the entire procedure is performed cinematically, allowing on-the-go data acquisitions.

Soil sensors can be classified based on their design concept as follows: (i) optical/radiometric, (ii) electrical/electromagnetic, (iii) electrochemical, and (iv) mechanical (Adamchuk et al., 2004; Kuang et al., 2012). Respectively, the above-mentioned design concepts allow the measurement of the soil capacity to (i) absorb, reflect, and/or emit electromagnetic energy; (ii) accumulate or conduct electrical charge; (iii) release ions; and (iv) resist mechanical distortions (Viscarra Rossel and Lobsey, 2016). Figure 4 shows some promising soil sesing technologies.

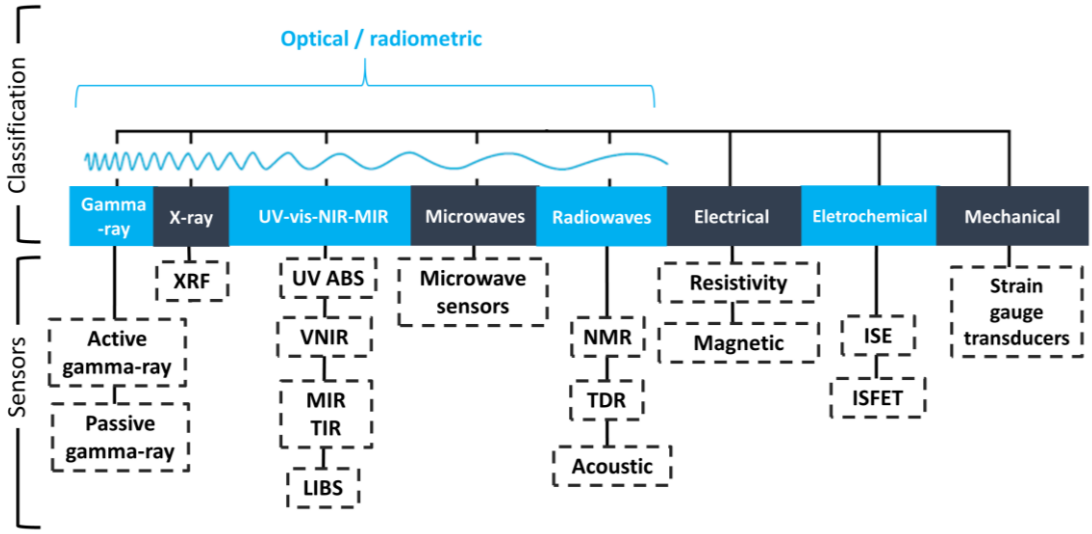


Figure 4. Some available technologies for the direct analysis of soils (adapted from Viscarra Rossel and Lobsey, 2016). The following abreviations were used: XRF for X-ray fluorescence; UV ABS for ultraviolet absorption; VNIR for visible and near-infrared spectroscopy; MIR for mid-infrared spectroscopy; TIR for thermal infrared spectroscopy; LIBS for laser-induced breakdown spectroscopy; NMR for nuclear magnetic resonance; TDR for time domain reflectometry; ISE for ion-selective electrodes; and ISFET for ion-selective field-effect transistors.

Studies have been conducted worldwide to evaluate and adapt the above-mentioned techniques to map soil attributes (Adamchuk et al., 2004; Kodaira and Shibusawa, 2013; Mouazen and Kuang, 2016; Sanches et al., 2018; Nawar et al., 2019). Some of them, bringing innovative applications for the localized management of P (Mouazen and Kuang, 2016) and limestone (Sanches et al., 2018), as well as combining sensor systems and multivariate statistical techniques for synthesize the main fertility attributes in fertility indices (Viscarra Rossel et al., 2010; Whetton et al., 2018). In general, consistent predictions of soil chemical attributes, mainly extractable nutrients, are still a challenge using the commonly applied sensing techniques. Hence, it is now necessary to evaluate "new" sensing techniques that have been little explored for soil fertility analysis. Among the possibilities, may be highlighted the XRF and the LIBS, both techniques count on portable sensors compatible with the direct analysis of soil samples (Galuszka et al., 2015), wich provide measurements of the total content of a wide range of elements (*e.g.*, Si, Fe, Ca, K, Ti, Al, among others).

In following sections, it is discussed challenges and advances of VNIR, XRF, and LIBS spectroscopies for predicting key fertility attributes. Thereafter, the complementarity between sensors and the importance of data fusion approaches are discussed.

2.5. Diffuse reflectance spectroscopy

Diffuse reflectance spectroscopy (DRS) involves soil sensing techniques characterized by their practicality, low operating cost, non-destructive nature, multi-informational spectra, and compatibility with little or no sample preparation (Stenberg et al., 2010). Another important feature of DRS is the possibility of registering spectral data on points or images using different platforms, *e.g.*, sensors directly on the field, benchtop sensors in the laboratory with sampled material, or remote sensing platforms with multi- or hyperspectral cameras. So, DRS comprises remote, proximal, or laboratory measurements and is a promising approach for applications in pedometrics (McBratney et al., 2003) and PA context (Adamchuk et al., 2004).

DRS has been used in Soil Science since the beginning of 1950s (Brooks, 1952). However, only in the last three decades it has gained importance, increasing the number of practical applications. This development is mainly associated with the establishment of chemometrics and multivariate statistical techniques in Analytical Chemistry (Viscarra Rossel et al., 2011).

Several scientific studies have successfully estimated soil physical and chemical properties using DRS in the spectral regions of visible (400–700 nm), near-infrared (700–2500 nm), and mid-infrared (2500–25000 nm) ranges (Ben-Dor and Banin, 1995; Viscarra Rossel et al., 2006). Moreover, DRS has been successfully applied directly in the field using sensors embedded in mobile platforms (Shibusawa et al., 1999; Mouazen et al., 2007; Christy, 2008) and portable sensors (Dhawale et al., 2015).

Diffuse reflectance consists of the percentage of incident radiation that is diffusely reflected by the soil at different wavelengths. The way the soil responds in reflectance to the different incident wavelengths constitutes its spectral behavior, which is represented by a spectrum. The spectra produced are the result of interactions between atoms and molecules of the soil and the incident radiation, which penetrates the first 10–50 μm of the sample surface (Demattê et al., 2016b). VNIR spectra provide information relate primarily to soil mineral constituents, organic compounds, and water content (Ben-Dor, 2002). In the spectrum, this information is represented by its intensity, shape, and absorptions (spectral features) at specific wavelengths (Demattê, 2002).

Absorptions in the visible region occur due to the excitation of valence electrons present in some atoms and functional groups (Clark and Roush, 1984). In organic molecules, these absorptions are restricted to certain functional groups (chromophores), which generate a rectilineal and concave shape across all spectra of the visible region. In inorganic species, such as iron oxides (*e.g.*, hematite and goethite), absorption occurs due to charge transfer, usually between 500 and 650 nm. In infrared region, especially MIR, the wavelengths that present energy equivalent to the natural vibrational frequencies of some molecules are absorbed, increasing the intensity of these vibrations (Pavia et al., 2010). In the NIR region, absorption occurs due to non-fundamental vibrations, which are overtones and

combination tones of the fundamental vibrations. Non-fundamental vibrations are considered secondary vibrations; they have lower intensity and can be understood as propagations of the fundamental vibrations that occur in the MIR (Demattê et al., 2016b). Important absorption features in soil spectra occur between 840 and 940 nm, related to the presence of Fe oxides (*e.g.*, hematite and goethite); between 1400 and 1900 nm due to the presence of water and hydroxyl molecules; between 2205 and 2225 nm related to kaolinite; and at 2265 nm associated to gibbsite (Demattê, 2002). The energy scattering, observed throughout the spectral intensity (the so-called albedo), is an important piece of information in the VNIR data. Under stable conditions of data acquisition geometry, surface roughness, and energy source intensity, the albedo should be related to soil granulometry (Ben-Dor, 2002).

Worldwide, many attempts have been made to predict physical and chemical attributes of soil using VNIR spectra. In general, successful calibrations of organic and total C, total N, and clay content are more likely to succeed because clay minerals and OM are spectrally active soil constituents, with well-known features in the VNIR region (Ben-Dor, 2002). Extractable nutrients and other soil attributes (*e.g.*, CEC, pH, and V) do not present absorption features in this spectral region and, hence, their correlations with VNIR spectra are generally weak (Stenberg et al., 2010). However, there are exceptions, as observed by Demattê et al. (2017) for ex-Mg and ex-K in Brazilian tropical soils and by Mouazen and Kuang (2016) for ex-P in temperate soils. These occasionally successful calibrations can be attributed to the covariance of those attributes with some spectrally active constituent (Kuang et al., 2012).

In Brazilian tropical soils, most studies have evaluated VNIR sensors in laboratories (Demattê et al., 2002, 2004, 2015a) and using remote sensing platforms (Nanni and Demattê, 2006; Demattê et al., 2016a) for digital soil survey. In both scenarios, satisfactory performances are most frequently reported for soil texture prediction (Demattê et al., 2015a; Demattê et al., 2016a; Cezar et al., 2019). In laboratory environment, satisfactory predictions can also be extended to OM (Demattê et al., 2004; Cezar et al., 2019), CEC (Nanni and Demattê, 2006), and extractable nutrients (Demattê et al., 2017).

An interesting approach using a VNIR sensor in a sugarcane field (185 ha) to characterize sub regions with different fertility classes was presented by Viscarra Rossel et al. (2010). The authors explored the multi-informational character of VNIR spectra, associating it with spatial information of the relief, to define a soil fertility index for different zones. This approach allowed to group three zones with contrasting fertility levels that should be managed using distinct fertilization strategies. Creating fertility indices are an interesting strategy to exploit soil sensors data, allowing to simplify soil fertility information in PA approaches (Iznaga et al., 2014; Askari et al., 2015). In addition, this approach is compatible with on-the-go surveys (Whetton et al., 2018), which would give more detailed spatial resolution.

Due to satisfactory performances in the laboratory environment, efforts have been made to develop VNIR equipment for on-the-go operations in agricultural fields (Shibusawa et al., 1999; Christy, 2008). One of the main challenges of *in-situ* applications using spectroanalytical techniques is to minimize the external effects (*e.g.*, moisture, surface roughness, granulometry, among others) that compromise part of the sensors' performance (Krug and Rocha, 2016). The sample preparation procedure predominantly used for data acquisition with VNIR sensors involves soil drying and sieving (< 2 mm), which ensures a better condition of the sample for spectral analysis (*e.g.*, homogeneity, reduction of the effects promoted by roughness and moisture) (Ben-Dor et al., 2015). Although on field sensing does not allow the execution of sample preparation, alternative methods can be adopted to minimize external effects, such as external parameter orthogonalization (EPO) (Roger et al., 2003) and direct standardization method (Wang et al., 1995). Some studies using VNIR spectroscopy directly in the field have achieved insensitive predictions of external effects using such methods (Wijewardane et al., 2016; Roudier et al., 2017). In Brazilian soils, Franceschini et al. (2018) pioneered the on-the-go data acquisition using a VNIR sensor. After using different methods to control the external effects, the authors obtained prediction models with potential for semi-quantitative determinations. They also highlighted that is necessary to ensure stable geometry conditions between the sensor and the soil surface during on-the-go surveys for optimal performance.

Globally, not many studies adressed the application of on-the-go VNIR data for delineating variable rate applications of fertilizers (*e.g.*, Maleki et al., 2008; Mouazen and Kuang, 2016). In this regard, an interesting research was presented by Mouazen and Kuang (2016). The authors presented results of the localized management of phosphatic fertilizers in a 21-ha field in the United Kingdom, where they used a VNIR on-the-go spectrometer to build the prescription maps. In this study, the diagnosis of P was performed over three years (2011, 2012 and 2013) using VNIR data after the harvest. As result of the site-specific management, an improvement in the spatial distribution of ex-P was observed, reducing its coefficients of variation from 26% to 25% and, subsequently, to 16% in 2011, 2012, and 2013, respectively. Mouazen and Kuang (2016) demonstrated the potential of this technique for a successful spatial diagnosis of fertility attributes in agricultural fields, which enabled a more assertive site-specific management.

2.6. Elemental analysis techniques

In recent years, there has been an increased interest in elemental analysis sensors (EAS), such as XRF and LIBS. These techniques are multi-elemental, allowing the measurement of the total content of a wide range of elements present in soil samples, such as Si, Fe, Al, K, P, Ca, among others. The

techniques are compatible with the direct analysis of soils, and can obtain satisfactory analytical performances in samples with little or no sample preparation. Moreover, their application do not generate chemical residues (Gredilla et al., 2016). Recent technological advances in optical and electronic components have allowed the development and miniaturization of these sensor systems, making both techniques more attractive for on-field analysis.

Although EAS allow determinations of the total nutrient contents, it is a consensus that their information can be important to infer about the soil fertility attributes. The correlation between extractable and total contents is the main reason that justifies using the elemental concentration of a given nutrient for the prediction of its respective extractable content (Viscarra Rossel and Lobsey, 2016). In addition, due to the weathered nature of tropical soils, their composition often shows an irrelevant presence of primary minerals (natural source of nutrients) (Fontes, 2012). Consequently, the relationships between the total and extractable contents of some nutrients (*e.g.*, ex-Ca, ex-K, and ex-Mg) are mainly influenced by soil agricultural management rather than natural processes (Lima et al., 2019, Andrade et al., 2020). This characteristic of tropical soils is attractive for using EAS to predict extractable nutrients. On the other hand, in soils with the presence of phylosilicates integrated with other soil minerals—such as mica (natural source of K) or calcite (natural source of Ca)—the relationship between the total and extractable content is more unlikely to occur. Furthermore, the soil chemistry in tropical soils regarding P is complex and characterized by high levels of P associated with strong bonds to Fe oxides (hematite and goethite) and Al (gibbsite). This complexity unbalances the relationship between its total content present in the soil and the amount of P accessible to plants (Hartemink, 2002; Schäefer et al., 2008); hence, develop calibrations using EAS that result in successful predictions of ex-P are more challenging than for the other nutrients.

Successful prediction of textural attributes in tropical soils using EAS are also expected, being explained by the relationship between the different particle size fractions and the total contents of Si, Al, and Fe (Lima et al., 2019). Si is commonly found in the sand and silt fractions (in the form of SiO_2) and, while Al, Fe, and Si are constituents of clay minerals, *e.g.*, kaolinite [$Al_2Si_2O_5(OH)_4$], gibbsite [$Al(OH)_3$], hematite (α-Fe_2O_3) and goethite (FeOOH) (Schäefer et al., 2008; Lima et al., 2019). Eventually, other cations, like microelements (Ni, Cr, Co, Cu, Zn, Ti, etc), can also be present in the iron oxides structure as isomorphic substitutes for Fe (Singh and Gilkes et al., 1992). Therefore, emission lines of these elements can also bring useful information for predicting texture attributes using EAS.

The satisfactory prediction of OM using EAS can occur using different information present in the spectrum, being able to occur using C or N emission lines, in the case of LIBS (Senesi and Senesi, 2016), or indirectly, exploring the relationship of this attribute with scattering peaks, in the case of XRF (Morona et al., 2017). Satisfactory predictive models can also be obtained for OM if it is correlated with

texture attributes (Lima et al., 2019), which is a common correlation found for tropical soils (Van Raij, 2011). For the other fertility attributes, such as CEC, pH, V, satisfactory predictive models can be obtained when such attributes present covariations with attributes more closely related to EAS spectra, such as textural attributes and extractable nutrients.

2.6.1. X-ray fluorescence spectroscopy (XRF)

The X-ray fluorescence technique is based on inducing fluorescence of a sample by means of its excitation with an incident X-ray source, followed by the measurement of specific photons that are emitted after this process (Kalnicky and Singhvi, 2001). The emitted photons also present energy in the X-ray region of the electromagnetic spectrum, this emitted energy is characteristic of each atom and, consequently, allows qualitative, and quantitative analyses of most elements present in the samples (Jenkins, 1995). Heavy elements [*e.g.*, the ones with higher atomic numbers (*e.g.*, Fe, Co, and Ni)] are more accurately measured than elements of low atomic number (*e.g.*, Si, Al, and P). So, higher limits of detection (LOD) are observed for light elements. For example, P presents LOD of approximately 5000 mg kg^{-1}, whereas heavier elements, such as U, present a lower LOD of ~ 5 mg kg^{-1} (Weindorf and Chakraborty, 2016). In addition, elements with atomic numbers lower than 12 (*e.g.*, N, B, C, Na, etc) cannot be detected by XRF technique (Jenkins, 1995); unless their concentrations are massively high and the XRF equipment has been configured to optimal conditions. The XRF spectra acquisition time is a parameter that must be configured on these sensors. Generally, it is recommended that the readings last for approximately 30–90 s, with the longer the dwell time, the greater the accuracy of the data (Weindorf and Chakraborty, 2016).

Regarding sample preparation, procedures such as drying and comminution (*e.g.*, reduction of particle size) allow the removal of moisture and promote homogenization of the sample, thus guaranteeing a gain in precision for XRF measurements (Krug and Rocha, 2016). However, simplification or omission of sample preparation may still produce promising results. O'Rourke et al. (2016) used dry samples with a particle size smaller than 2 mm (preparation procedure similar to that adopted for VNIR sensor) and obtained an R^2 higher than 0.70 for predicting the total content of K, Cr, Cu, Mn, and Zn. Weindorf et al. (2012) used a portable XRF to evaluate the chemical components in soil trenches, without any sample preparation. The authors reported that the technique allowed the observation of subtle chemical differences among the different horizons of the soil profile. Moreover, they also obtained good prediction models ($R^2 > 0.90$) for organic C.

An interesting advantage of the XRF technique for proximal sensing is the lower influence of soil moisture on its spectra (Horta et al., 2015), specially when compared to VNIR sensors. Ge et al.

(2005) presented a simple method for correcting the effect of moisture, stating that its application is necessary only in samples with a gravimetric moisture content above 20%. Despite these researches, to the best of our knowledge, the XRF performance on samples with different moisture contents has never been evaluated for the prediction of soil fertility attributes.

Regarding the performance of XRF for predicting fertility attributes, in soils of temperate regions, good prediction performances have been reported for pH prediction (Sharma et al. 2014), CEC (Sharma et al. 2015), V (Rawal et al., 2019), and soil texture (Zhu et al. 2011). In Brazilian tropical soils, satisfactory performances have been obtained for soil texture (Lima et al., 2019), chemical attributes (pH, V, and CEC) (Teixeira et al., 2018), organic matter (OM) (Morona et al., 2017), and extractable nutrients (ex-K, ex-Ca, and ex-Mg) (Silva et al., 2017; Tavares et al., 2019). Despite these studies, the application of XRF for tropical soil fertility analysis is in its early stages of development.

Finally, we comment that a comprehensive overview of the fundamentals and applications of the XRF technique for soil analysis, as well as discussions on quality control protocols, spectral interferences, equipment, and safety procedures, are shown in the United States Environmental Protection Agency's 6200 method (USEPA, 2007).

2.6.2. Laser-induced breakdown spectroscopy (LIBS)

This technique uses a laser as an energy source to vaporize part of the surface of the analyzed material, causing a microsampling by ablation and an immediate excitation of atoms and ions present at the ablation site. This excitation generates a microplasma that has its emission spectrum detected by a spectrometer, which is usually sensitive to the range between 200 and 900 nm (Harmon et al., 2013). The LIBS technique allows direct analysis of solids, having as important features its practicality, spectrum with multi-elemental capability, and micro-destructive character (Yu et al., 2016). Moreover, it can be considered a complementary technique to XRF since it allows the quantification of elements with atomic numbers lower than 12 (*e.g.*, N, C, and Na). LIBS spectra acquisition is extremely fast, with the collection of 1–20 spectra per second, depending on the laser frequency (Krug and Rocha, 2016).

For efficient predictive models using LIBS data, it is necessary to ensure that the laser promotes a congruent ablation in the analyzed samples, as well as has energy enough for an adequate plasma formation for the expulsion of the elements that compose the material (Senesi et al., 2009). In other words, the interaction between the laser and the sample—which involves ablation, vaporization, atomization, and excitation processes of the species in the plasma—must be reproducible (Gomes et al. 2011). Variations in the amount of the ablationated matter make it difficult to model the elements and replicate the generated models. Thus, the sample preparation before the application of the LIBS

technique is of paramount importance and is one of the main challenges for applying the LIBS technique to soil samples under field conditions (Jantzi et al., 2016). Thus, although technological advances have allowed the development of portable LIBS sensors (already avalaible in the market by some companies, *e.g.*, Sciaps, Inc., Woburn, USA), the application of such equipment for quantitative assessments in soil samples, directly in the field, still have the above-mentioned challenge to be overcome.

In benchtop analysis, the pelletizing of samples is recommended for soil analysis. Pressed pellets are prepared by applying a high pressure (~784.53 MPa or 8 t cm^{-2}) on dried and comminuted (with particle size usually < 150 μm) soil samples. The comminution of the material is usually performed with a cryogenic grinding mill or with a planetary ball mill. Sandy soils, with high presence of silicates, do not form pellets only by pressing the comminuted material and require the addition of a binder material (*e.g.*, cellulose or paraffin) for this purpose. When necessary, the binder material used should not contain the elements of interest in the analysis, *e.g.*, it must be inert. The pellet preparation, besides promoting material homogeneity, allows standardization of the density, porosity, and roughness of the sample surface (Krug and Rocha, 2016). The smaller the particle size to be pressurized, the more resistant and cohesive will be the pellet, and therefore will present more uniform craters after laser ablation and better measurement accuracy with LIBS (Carvalho et al., 2015).

The LIBS technique has been proposed in recent years for the characterisation of soil fertility attributes (Nicolodelli et al., 2019). In Brazilian soils, some researches have shown the potential of this technique to predict clay, sand, and silt content (Villas-Boas et al., 2016), C (Nicolodelli et al., 2014), and pH values (Ferreira et al., 2015). In this latter study, the pH was predicted using emission lines of Al, Ca, H, and O, in association with multivariate statistical techniques. Knadel et al. (2017) compared the performance between a LIBS and VNIR sensors to predict organic C and textural attributes in agricultural soils. The authors obtained good results using both techniques; even though, the determinations performed with LIBS generally achieved lower prediction errors. The authors also emphasized that the sample preparation performed for the LIBS analyses (pressed pellet preparation) is a key disadvantage of this technique compared to VNIR spectroscopy.

In order to improve the LIBS performance in soil analysis, some studies have suggested methodologies to reduce the physical matrix effect (Marangoni et al., 2016) and to correct spectral interferences (Nicolodelli et al., 2014). Nicolodelli et al. (2019) reviewed advances over the last decade for LIBS applications in soil analysis. They concluded that even though significant advances have been achieved in the detection limit of several elements, the reduction of matrix effects, the optimization of the signal-to-noise ratio, and the application of modern chemometric methods, other aspects, such as utilizing the LIBS application directly in the field remain challenging and unexplored. The authors also commented that the great progress achieved in LIBS instrumentation, as well as the advances for its

application in other areas (*e.g.*, environmental and geological science), should encourage the agro-scientific community to develop more studies on LIBS sensors. This would accelerate the development of new approaches that can overcome the many disadvantages and limitations associated with the LIBS technique in the agricultural sector.

2.7. Multi-sensor data integration (data fusion)

Soils offer numerous scales of spatial and temporal variation that can be monitored using different sensor systems. Although studies using single-sensor approaches have demonstrated potential for different applications, no sensor working alone can completely characterize the complexity of the soil (Nawar et al., 2017). Each soil sensor presents an exclusive perspective on the possibilities of predicting physical and chemical attributes along the soil profile, which is a function of its design concept and operation type. Multisensor systems and data fusion techniques allow integrating information collected at different scales (both vertically and horizontally in agricultural fields) and related to different soil properties in order to deal with the challenge of soil plurality (Grunwald et al., 2015).

Data fusion is a multidisciplinary field based on different areas, including information technology, signal processing, statistical assessment, and artificial intelligence (Khaleghi et al., 2013). Different forms of soil sensing integration can be applied, *e.g.*, multisensor systems acting on a benchtop, joint operation of different sensor systems directly in the field, integration between proximal and remote sensing techniques. Recently, the number of published studies on data fusion for the prediction of soil attributes has increased (Nawar et al., 2017). Some key advantages of sensor integration that have been reported by scientific publications, are: (i) increments in the prediction accuracy compared to the single-sensor case (O'Rourke et al., 2016); (ii) extend the number of soil attributes determined with satisfactory prediction performance (Mahmood et al., 2012); and (iii) the generation of new indices that allow subsidized localized management strategies (Benedetto et al., 2013; Mouazen et al., 2014).

An example of information complementarity between different sensors is observed between EAS (*e.g.*, LIBS and XRF) and DRS techniques (*e.g.*, VNIR). The elemental analysis performed using EAS enables the characterisation of the inorganic constituents of soil samples, while VNIR spectroscopy determines the mineralogical and organic components. The synergy between these techniques and the integrated use of these sensors has been attempted by recent studies (Horta et al., 2015; Wang et al., 2015; O'Rourke et al., 2016), which have shown better performance for predicting fertility attributes when using joint predictions rather than individual ones. Despite those recent advances, the ideal combination of sensors, as well as a robust and comprehensive strategy for data fusion for predicting

key fertility attributes are still unknown. Therefore, further studies in these areas should be encouraged within the soil sensing community, especially in Brazilian tropical soils.

2.8. Final considerations

The site-specific management of soil, based on the spatial diagnosis of fertility attributes, provides agronomic and environmental benefits to agricultural production systems. This localized management of soil fertility, wich is currently the main PA approach in Brazil, depends on an accurate mapping of key soil fertility attributes. Physical and chemical attributes related to soil fertility present different scales of spatial and temporal variations and their variability in the field is usually characterized by short-range semivariograms. Hence, for reliable mapping of most of these attributes, a high spatial density sampling (*e.g.*, ≥ 1 sample ha^{-1}) is required. This creates a soil sampling dilemma since increasing the density of soil sampling points means additional costs associated with laboratory analysis and more effort for sampling. Alternatively, soil sensor systems have been proposed to fill this information gap, allowing more practical, cheaper, and environmentally-friendly approaches.

Technological advances in a wide variety of sensors made them smaller, more accurate, more efficient, and less expensive; then, better suited for field applications. Some sensors are compatible with the proximal soil sensing (PSS) concept, allowing the direct soil analysis (*e.g.*, without sample preparation) *in-situ*. Other technologies still need some sample preparation to maintain an acceptable analytical performance [*e.g.*, laser-induced breakdown spectroscopy (LIBS)] and should be considered for mobile laboratory approaches.

The effective use of soil sensors for fertility attributes prediction involves overcoming several multidisciplinary challenges, which are listed below: (i) understanding the fundamentals of each technique, its potential and limitations, not to overestimate its potential; (ii) minimizing the loss of data quality when using PSS approaches; (iii) associating soil sensors spatial data with multivariate geostatistical approaches for a more accurate characterisation of the unsampled sites; (iv) adapting data fusion techniques with multi-sensor systems for a broader and more accurate characterisation of fertility attributes; and (v) developing general protocols to deal with field-by-field sensors' performance variations. Finally, it is importante to mention that in view of the enormous demand for fertilizers and lime in Brazilian soils, even as the worldwide calls for food security and a sustainable production environment, the development of successful sensor-based approaches for soil management should be seen as a necessity and further research should be encouraged in Brazilian tropical soils.

References

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

Adamchuk, VI.; Hummel, JW.; Morgan, MT.; Upadhyaya, SK. On-the-go soil sensors for precision agriculture. *Computers and electronics in agriculture* **2004**, 44, 71-91. doi: 10.1016/j.compag.2004.03.002

Adamchuk, V.I.; Morgan, M.T.; Brouder, S.M. Development of an on-the-go soil pH mapping method: analysis of measurement variability. *Applied Engineering Agriculture* **2006**, 22, 335-344, doi: 10.13031/2013.20450.

Askari, M.; O'Rourke, S.M.; Holden, N.M. Evaluation of soil quality for agricultural production using visible–near-infrared spectroscopy. *Geoderma* **2015**, 243, 80-91. doi: 10.1016/j.geoderma.2014.12.012.

Ben-Dor, E. Quantitative remote sensing of soil properties. *Advances in Agronomy* **2002**, 75, 173-244, doi: 10.1016/S0065-2113(02)75005-0.

Ben-Dor, E.; Banin, A. Near-infrared analysis as a rapid method to simultaneously evaluate several soil properties. *Soil Science Society of America Journal* **1995**, 59, 364-372. doi: 10.2136/sssaj1995.03615995005900020014x.

Ben-Dor, E.; Ong, C.; Lau, I.C. Reflectance measurements of soils in the laboratory: Standards and protocols. *Geoderma* **2015**, 245, 112-124, doi: 10.1016/j.geoderma.2015.01.002.

Benedetto, D.; Castrignanò, A.; Rinaldi, M.; Ruggieri, S.; Santoro, F.; Figorito, B.; Diacono, M.; Tamborrino, R. An approach for delineating homogeneous zones by using multi-sensor data. *Geoderma* **2013**, 199, 117-127, doi: 10.1016/j.geoderma.2012.08.028.

Brooks, F.A. Atmospheric radiation and its reflection from the ground. *Journal of Meteorology* **1952**, 9, 41–52, doi: 10.1175/1520-0469(1952)009<0041:ARAIRF>2.0.CO;2.

Carvalho, G.G.A.; Santos, Jr. D.; Gomes, M.S.; Nunes, L.C.; Guerra, M.B.B.; Krug, F.J. Influence of particle size distribution on the analysis of pellets of plant materials by laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2015**, 105, 130-135. doi: 10.1016/j.sab.2014.09.001.

Cezar, E.; Nanni, M.R.; Guerrero, C.; Silva Junior, C.A.; Cruciol, L.G.T.; Chicati, M.L.; Silva, G.F.C. Organic matter and sand estimates by spectroradiometry: Strategies for the development of models with applicability at a local scale. *Geoderma* **2019**, 340, 224-233. doi: 10.1016/j.geoderma.2019.01.021.

Cherubin, M.R.; Santi, A.L.; Eitelwein, M.T.; Ros, C.O.; Bisognin, M.B. Sampling grids used to characterise the spatial variability of pH, Ca, Mg and V% in Oxisols. *Revista Ciência Agronômica* **2014**a, 45, 659-672. doi: 10.1590/S1806-66902014000400004

Cherubin, M.R.; Santi, A.L.; Eitelwein, M.T.; Menegol, D.R.; Da Ros, C.O.; Castro Pias, O.H.; Berghetti, J. Eficiência de malhas amostrais utilizadas na caracterização da variabilidade espacial de fósforo e potássio. *Ciência Rural* **2014**b, 44, 425-432.

Cherubin, MR.; Santi, AL.; Eitelwein, MT.; Amado, TJC.; Simon DH.; Damian JM. Dimensão da malha amostral para caracterização da variabilidade espacial de fósforo e potássio em Latossolo Vermelho. *Pesquisa Agropecuária Brasileira* **2015**, 50, 168-177.

Christy, C.D. Real-time measurement of soil attributes using on-the-go near infrared reflectance spectroscopy. *Computers and electronics in agriculture* **2008**, 61, 10-19, doi: 10.1016/j.compag.2007.02.010.

Clark, R.N.; Roush, T.L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth* **1984**, 89, 6329-6340, doi: 10.1029/JB089iB07p06329.

Crookston RK (2006) A top 10 list of developments and issues impacting crop management and ecology during the past 50 years. Crop science 46(5): 2253-2262. DOI: 0.2135/cropsci2005.11.0416gas.

Demattê, J.A.M. Characterisation and discrimination of soils by their reflected electromagnetic energy. *Pesquisa Agropecuária Brasileira* **2002**, 37, 1445-1458. doi: 10.1590/S0100-204X2002001000013

Demattê, J.A;. Campos, R.C.; Alves, M.C.; Fiorio, P.R.; Nanni, M.R. Visible–NIR reflectance: a new approach on soil evaluation. *Geoderma* **2004**, 121, 95-112, doi: 10.1016/j.geoderma.2003.09.012.

Demattê, JAM.; Alves, MR.; Gallo, BC.; Fongaro, CT.; Souza, AB.; Romero, DJ.; Sato, MV. Hyperspectral remote sensing as an alternative to estimate soil attributes. *Revista Ciencia Agronomica* **2015**a, 46(2): 223–232, doi: 10.5935/1806-6690.20150001.

Demattê, J.A.M.; Alves, M.R.; Terra, F.D.S.; Bosquilia, R.W.D.; Fongaro, C.T.; Barros, P.P.D.S. Is it possible to classify topsoil texture using a sensor located 800 km away from the surface?. *Revista Brasileira de Ciência do Solo* **2016**a, 40, doi: http: 10.1590/18069657rbcs20150335.

Demattê, J.A.M.; Morgan, C.L.S.; Chabrillat, S.; Rizzo, R.; Franceschini, M.H.D.; Terra, F.; Wetterlind, J. Spectral sensing from ground to space in soil science: State of the art, applications, potential and perspectives. *Land Resources Monitoring, Modeling, and Mapping with Remote Sensing*. Thenkabail, PS, Ed, 661-732, 2016b.

Demattê, J.A.; Ramirez-Lopez, L.; Marques, K.P.P.; Rodella, A.A. Chemometric soil analysis on the determination of specific bands for the detection of magnesium and potassium by spectroscopy. *Geoderma* **2017**, 288, 8-22. doi: 10.1016/j.geoderma.2016.11.013.

Demattê, JAM.; Dotto, AC.; Bedin, LG.; Sayão, VM.; Souza, AB. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. *Geoderma* **2019**, 337, 111-121. doi: 10.1016/j.geoderma.2018.09.010.

Dhawale, N.M.; Adamchuk, VI.; Prasher, SO.; Viscarra Rossel, RA.; Ismail, AA.; Kaur, J. Proximal soil sensing of soil texture and organic matter with a prototype portable mid- infrared spectrometer. *European Journal of Soil Science* **2015**, 66, 661-669, doi: 10.1111/ejss.12265.

Dow, A.I.; James, D.W. Intensive Soil Sampling: A Principle of Soil Fertility Management in Intensive Irrigation Agriculture. In: *Washington Agricultural Experiment Station Bulletin 781*. Washington State University, Pullman, WA, 1973a.

Dow, A.I.; James, D.W.; Russell, T.S. Soil Variability in Central Washington and Sampling for Soil Fertility Tests. In: *Washington Agricultural Experiment Station Bulletin 788*, Washington State University, Pullman, WA, 1973b.

FAO - Food and Agriculture Organization of the United Nations. *World Fertilizer Trends and Outlook to 2020*. Food and Agriculture Organization of the United Nations (FAO), Rome, Italy, 2017.

Ferreira, EC.; Neto, J.A.G.; Milori, D.M.; Ferreira, E.J.; Anzano, J.M. Laser-induced breakdown spectroscopy: Extending its application to soil pH measurements. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2015**, 110, 96-99. doi: 10.1016/j.sab.2015.06.002

Fontes, M.P.F. Intemperismo de rochas e minerais. In: *Pedologia: fundamentos*. Ker, J.C.; Curi, N.; Schaefer, C.*E.G*.R.; Vidal-Torrado, P., Eds.; Sociedade Brasileira de Ciência do Solo: Viçosa, Brazil, 2012. pp. 181-205. (In Portuguese)

Foley, J.A.; Ramankutty, N.; Brauman, K. A..; Cassidy, E.S.; Gerber, J.S.; Johnston, M.; Mueller, N.D.; O'Connell, C.; Ray, D.K.; West, P.C.; Balzer, C.; Bennett, E.M.; Carpenter. S,R.; Hill, J.; Monfreda, C.; Polasky, S.; Rockstrom, J.; Sheehan, J.; Siebert, S.; Tilman, D.; Zaks, D.P.M. Solutions for a cultivated planet. *Nature* **2011**, 478(7369), 337–342.

Franceschini, M.H.D., Demattê, J.A.M., Kooistra, L., Bartholomeus, H., Rizzo, R., Fongaro, C.T.; Molin, J.P. Effects of external factors on soil reflectance measured on-the-go and assessment of potential spectral correction through orthogonalisation and standardisation procedures. *Soil and Tillage Research* **2018**, 177, pp.19-36, doi: 10.1016/j.still.2017.10.004.

Galuszka, A.; Migaszewski, ZM.; Namieśnik, J. Moving your laboratories to the field–Advantages and limitations of the use of field portable instruments in environmental sample analysis. *Environmental research* **2015**, 140, 593-603. doi 10.1016/j.envres.2015.05.017.

Ge, L.; Lai, W.; Lin, Y. Influence of and correction for moisture in rocks, soils and sediments on *in-situ* XRF analysis. *X- Ray Spectrometry Journal* **2005**, 34, 28-34. doi: 10.1002/xrs.782

Gebbers, R.; Adamchuk, V.I. Precision agriculture and food security. *Science* **2010**, 327, 828-831. doi: 10.1126/science.1183899.

Gomes, M.S.; Junior, D.S.; Nunes, L.C.; Carvalho, G.G.A.; Leme, F.O.; Krug, F.J. Evaluation of grinding methods for pellets preparation aiming at the analysis of plant materials by laser induced breakdown spectrometry. *Talanta* **2011**, 85, 1744-1750. doi: 10.1016/j.talanta.2011.06.069.

Godfray, H.C.J.; Beddington, J.R.; Crute, I.R.; Haddad, L.; Lawrence, D.; Muir, JF.; Pretty, J.; Robinson, S,.; Thomas, SM.; Toulmin, C. Food Security: The Challenge of Feeding 9 Billion People. *Science* **2010**, 327, 812-818, doi: 10.1126/science.1185383.

Gredilla, A.; Vallejuelo, S.F.O.; Elejoste, N.; Diego, A.; Madariaga, J.M. Non-destructive Spectroscopy combined with chemometrics as a tool for Green Chemical Analysis of environmental samples: A review. *TrAC Trends in Analytical Chemistry* **2016**, 76, 30-39, doi: 10.1016/j.trac.2015.11.011.

Grunwald, S.; Vasques, G.M.; Rivero, R.G. Fusion of soil and remote sensing data to model soil properties. *Advances in Agronomy* **2015**, 131, 1-109, doi: 10.1016/bs.agron.2014.12.004.

Hammond, M.W. Cost analysis of variable fertility management of phosphorus and potassium for potato production in central Washington. In: *Proceedings of soil specific crop management*. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America, 1993.

Harmon, R.S.; Russo, R.E.; Hark, R.R.; Applications of laser-induced breakdown spectroscopy for geochemical and environmental analysis: A comprehensive review. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2013**, 87: 11-26. doi: 10.1016/j.sab.2013.05.017.

Hartemink, A.E. Soil science in tropical and temperate regions—some differences and similarities. *Adv. Agron.* **2002**, 77, 269-292, doi: 10.1016/S0065-2113(02)77016-8.

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

Hummel, JW.; Gaultney, LD.; Sudduth, KA. Soil property sensing for site-specific crop management. *Computers and Electronics in Agriculture* **1996**, 14(2-3): 121-136, doi: 10.1016/0168-1699(95)00043-7.

Iznaga, A.C.; Orozco, M.R.; Alcantara, E.A.; Pairol, M.C.; Sicilia, Y.E.D.; Baerdemaeker, J.; Saeys, W. Vis/NIR spectroscopic measurement of selected soil fertility parameters of Cuban agricultural Cambisols. *Biosystems engineering* **2014**, 125, 105-121, doi: 10.1016/j.biosystemseng.2014.06.018.

Jantzi, S.C.; Motto-Ros, V.; Trichard, F.; Markushin, Y.; Melikechi, N.; Giacomo, A. Sample treatment and preparation for laser-induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy* **2016**, 115, 52-63. doi: 10.1016/j.sab.2015.11.002.

Jenkins, R. *Quantitative X-ray spectrometry*. CRC Press, 1995. Doi: 10.1201/9781482273380

Kalnicky, D.J.; Singhvi, R. Field portable XRF analysis of environmental samples. *Journal of hazardous materials* **2001**, 83, 93-122. doi: 10.1016/S0304-3894(00)00330-7.

Khaleghi, B.; Khamis, A.; Karray, F.O.; Razavi, S.N. Multisensor data fusion: A review of the state-of-the-art. *Information fusion* **2013**, 14(1), 28-44. doi: 10.1016/j.inffus.2011.08.001.

Khosla, R.; Alley, MM. Soil-specific nitrogen management on Mid-atlantic coastal plain soils. *Better Crop* **1999**, 83(3), 6-7.

Knadel, M.; Gislum, R.; Hermansen, C.; Peng, Y.; Moldrup, P.; Jonge, L.W.; Greve, M.H. Comparing predictive ability of laser-induced breakdown spectroscopy to visible near-infrared spectroscopy for soil property determination. *Biosystems engineering* **2017**, 156, 157-172, doi: 10.1016/j.biosystemseng.2017.01.007.

Kodaira, M.; Shibusawa, S. Using a mobile real-time soil visible-near infrared sensor for high resolution soil property mapping. *Geoderma* **2013**, 199, 64-79. doi: 10.1016/j.geoderma.2012.09.007.

Krug, FJ.; Rocha, FRP. *Métodos de preparo de amostras para análise elementar*. EditSBQ, Sociedade Brasileira de Química, São Paulo, 2016. 572,.

Kuang, B.; Mahmood, HS.; Quraishi, MZ.; Hoogmoed, WB.; Mouazen, AM.; Henten, EJV. Sensing soil properties in the laboratory, *in-situ*, and on-line: a review. *Advances in Agronomy* **2012**, 114, 155-223p.

Kuang B; Mouazen AM. Calibration of visible and near infrared spectroscopy for soil analysis at the field scale on three European farms. *European Journal of Soil Science* **2011**, 62, 629-636. doi: https://doi.org/10.1111/j.1365-2389.2011.01358.x.

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

Mahmood, H.S.; Hoogmoed, W.B.; van Henten, E.J. Sensor data fusion to predict multiple soil properties. *Precis. Agric.* **2012**, 13, 628-645, doi: 10.1007/s11119-012-9280-7.

Maleki, MR.; Mouazen, A.M.; Ketelaere, B.; Ramon, H.; Baerdemaeker, J. On-the-go variable-rate phosphorus fertilisation based on a visible and near-infrared soil sensor. *Biosystems Engineering* **2008**, 99, 35–46. doi: 10.1016/j.biosystemseng.2007.09.007.

Marangoni, B. S.; Silva, K. S.; Nicolodelli, G.; Senesi, G. S.; Cabral, J. S.; Villas-Boas, P. R.; ... & Milori, D. M. Phosphorus quantification in fertilizers using laser induced breakdown spectroscopy (LIBS): a methodology of analysis to correct physical matrix effects. *Analytical Methods* **2016**, 8(1), 78-82. doi: 10.1039/C5AY01615K

Mc Bratney, A. B.; Santos, M. M.; Minasny, B. On digital soil mapping. *Geoderma* **2003**, 117, 3-52, doi: 10.1016/S0016-7061(03)00223-4.

Mc Bratney, A.B.; Pringle, M.J. Estimating average and proportional variograms of soil properties and their potential use in precision agriculture. *Precision Agriculture* **1999**, 1, 125-152. doi: https://doi.org/10.1023/A:1009995404447.

Molin, J.P.; Amaral, L.R.; Colaço, A. *Agricultura de precisão*. São Paulo, Oficina de Textos, 2015. 238p.

Molin, J.P. Agricultura de precisão: números do mercado brasileiro. Boletim Técnico 03, 2017. Disponível em: http://www.agriculturadeprecisao.org.br/upimg/publicacoes/pub_-boletim-tecnico-03---agricultura-de-precisao-numeros-do-mercado-brasileiro-11-04-2017.pdf. Acessado 10 março, 2018.

Montanari, R.; Souza, G.S.A.; Pereira, G.T.; Marques, J.J.; Siqueira, D.S.; Siqueira, G.M. The use of scaled semivariograms to plan soil sampling in sugarcane fields. *Precision Agriculture* **2012**, 13, 542-552. doi: 10.1007/s11119-012-9265-6.

Morona, F.; Santos, F.R.; Brinatti, A.M.; Melquiades, F.L. Quick analysis of organic matter in soil by energy-dispersive X-ray fluorescence and multivariate analysis. *Applied Radiation and Isotopes* **2017**, 130, 13-20, doi: 10.1016/j.apradiso.2017.09.008

Mouazen, A.M.; Maleki, M.R.; Baerdemaeker, J.; Ramon, H. On-line measurement of some selected soil properties using a VNIR sensor. *Soil Tillage Research* **2007**, 93, 13-27. doi: 10.1016/j.still.2006.03.009.

Mouazen, A.M.; Alhwaimel, S.A.; Kuang, B.; Waine, T. Multiple on-line soil sensors and data fusion approach for delineation of water holding capacity zones for site specific irrigation. *Soil and Tillage Research* **2014**, 143, 95-105, doi: 10.1016/j.still.2014.06.003

Mouazen, A.M; Kuang, B. On-line visible and near infrared spectroscopy for in-field phosphorous management. *Soil and Tillage Research* **2016**, 155: 471-477. doi: 10.1016/j.still.2015.04.003.

Mulla, D.; Khosla, R. Historical Evolution and Recent Advances in Precision Farming. In: *Soil-Specific Farming Precision Agriculture*: 1-35, 2016.

Mulla, D.J.; Hammond, M.W. Mapping of soil test results from large irrigation circles. In: Proc. 39th Annual Far West Regional Fertilizer Conference, Bozeman, MT, 1988.

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

Nanni, M.R.; Demattê, J.A.M. Spectral reflectance methodology in comparison to traditional soil analysis. *Soil Science Society of America Journal* **2006**, 70, 393-407, doi: 10.2136/sssaj2003.0285.

Nanni, MR.; Povh, FP.; Demattê, JAM.; Oliveira, RBD.; Chicati, ML.; Cezar E. Optimum size in grid soil sampling for variable-rate application in site-specific management. *Scientia Agricola* **2011**, 68(3), 386-392, doi: 10.1590/S0103-90162011000300017.

Nawar, S.; Corstanje, R.; Mulla D.; Mouazen, AM. Delineation of soil management zones for variable-rate fertilization: A review. *Advances in agronomy* **2017** 143: 175-245, doi: 10.1016/bs.agron.2017.01.003.

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

Nicolodelli, G.; Marangoni, BS.; Cabral, JS.; Villas-Boas, PR.; Senesi, GS.; dos Santos CH.; ... & Milori, D. M. Quantification of total carbon in soil using laser-induced breakdown spectroscopy: a method to correct interference lines. *Applied optics* **2014**, 53, 2170-2176, doi: 10.1364/AO.53.002170

Nicolodelli, G.; Cabral, J.; Menegatti, C.R.; Marangoni, B.; Senesi, G.S. Recent advances and future trends in LIBS applications to agricultural materials and their food derivatives: an overview of developments in the last decade (2010–2019). Part I. Soils and fertilizers. *TrAC Trends in Analytical Chemistry* **2019**. doi: 10.1016/j.trac.2019.03.032

O'Rourke, SM.; Stockmann, U.; Holden, NM.; McBratney, A.B.; Minasny B. An assessment of model averaging to improve predictive power of portable VNIR and XRF for the determination of agronomic soil properties. *Geoderma* **2016**, 279, 31-44. doi: 10.1016/j.geoderma.2016.05.005

Pandey, S.; Bhatta, NP.; Paudel, P.; Pariyar, R.; Maskey, K.H.; Khadka, J.; Thapa, TB.; Rijal, B.; Panday, D. Improving fertilizer recommendations for Nepalese farmers with the help of soil-testing mobile van. *Journal of Crop Improvement* **2017**, 32, 19-32. doi: 10.1080/15427528.2017.1387837.

Pavia, D.L.; Lampman, G.M.; Kriz, G.S.; Vyvyan, J.R. *Introdução à espectroscopia*. Cengage Learning, 2010.

Riebe, D.; Erler, A.; Brinkmann, P.; Beitz, T.; Löhmannsröben, H.G.; Gebbers, R. Comparison of Calibration Approaches in Laser-Induced Breakdown Spectroscopy for Proximal Soil Sensing in Precision Agriculture. *Sensors* **2019**, 19(23), p.5244, doi: 10.3390/s19235244.

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

Resende, A.V.; Coelho, A. Amostragem para mapeamento e manejo da fertilidade do solo na abordagem de agricultura de precisão. International Plant Nutrition Institute (IPNI): Informações Agronômicas, 156: 1-8, 2017. Disponível em: http://www.ipni.net/publication/ia-brasil.nsf

Roger, J.M.; Chauchard, F.; Bellon-Maurel, V. EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometrics and Intelligent Laboratory Systems* **2003**, 66, 191–204, doi: 10.1016/S0169-7439(03)00051-0.

Roudier, P.; Hedley, C.B.; Lobsey, C.R.; Viscarra Rossel, R.; Leroux, C. Evaluation of two methods to eliminate the effect of water from soil vis–NIR spectra for predictions of organic carbon. *Geoderma* **2017**, 296, 98-107, doi: https: 10.1016/j.geoderma.2017.02.014.

Sanches, GM.; Magalhães, PS.; Remacre, AZ.; Franco, H.C. Potential of apparent soil electrical conductivity to describe the soil pH and improve lime application in a clayey soil. *Soil and Tillage Research* **2018**, 175: 217-225, doi: 10.1016/j.still.2017.09.010.

Schäefer, C.*E.G.*R.; Fabris, J.D., Ker, J.C. Minerals in the clay fraction of Brazilian latosols (oxisols): a review. *Clay Miner.* **2008**, 43, 137–154. doi: https://doi.org/10.1180/claymin.2008.043.1.11.

Schirrmann, M.; Domsch, H. Sampling procedure simulating on-the- go sensing for soil nutrients. *Journal of Plant Nutrition and Soil Science* **2011**, 174, 333-343. doi: 10.1002/jpln.200900367.

Silva, S.H.G.; Teixeira, A.F.D.S.; Menezes, M.D.D.; Guilherme, L.R.G.; Moreira, F.M.D.S.; Curi, N. Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). *Ciênc. Agrotec.* **2017**, 41, 648-664, doi: 10.1590/1413-70542017416010317.

Singh, B.; Gilkes, R.J. Properties and distribution of iron oxides and their association with minor elements in the soils of south-western Australia. *J. Soil Sci.* **1992**, 43, 77-98.

Senesi, GS.; Dell'Aglio, M.; Gaudiuso, R.; Giacomo, A.; Zaccone, C.; Pascale, O.; Miano, T.M.; Capitelli, M. Heavy metal concentrations in soils as determined by laser-induced breakdown spectroscopy (LIBS), with special emphasis on chromium. *Environmental research* **2009**, 109, 413-420. doi: 10.1016/j.envres.2009.02.005.

Senesi, G.S.; Senesi, N. Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review. *Analytica Chimica Acta* **2016**, 938, 7-17, doi: 10.1016/j.aca.2016.07.039.

Sethuramasamyraja, B.; Adamchuk, VI.; Dobermann, A.; Marx, DB.; Jones, DD.; Meyer, G.E. Agitated soil measurement method for integrated on-the-go mapping of soil pH, potassium and nitrate contents. *Computers and electronics in agriculture* **2008**, 60, 212-225. doi: 10.1016/j.compag.2007.08.003.

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

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

Shibusawa, S.; Li, M.Z.; Sakai, K.; Sasao, A.; Sato, H.; Hirako, S.; Otomo, A. Spectrophotometer for real-time underground soil sensing. American Society of Agricultural Engineers: St. Joseph, MI, USA, 1999.

Steinberger, G.; Rothmund, M.;. Auernhammer, H. Mobile farm equipment as a data source in an agricultural service architecture. *Computers and electronics in agriculture* **2009**, 65, 238-246, doi: 10.1016/j.compag.2008.10.005.

Stenberg, B.; Viscarra Rossel, RA.; Mouazen, A.M.; Wetterlind, J. Visible and near-infrared spectroscopy in soil science. *Advances in Agronomy* **2010**, 107, 163-215. doi: 10.1016/S0065-2113(10)07005-7.

Sudduth, K.A.; Hummel, J.W. Evaluation of reflectance methods for soil organic matter sensing. *Transactions of the ASAE* **1991**, 34, 1900-1909, doi: 10.13031/2013.31816.

Tavares, T.R.; Nunes, L.C.; Alves, E.E.N.; Almeida, E.; Maldaner, L.F.; Krug, F.J.; Carvalho, H.W.P.; Molin, J.P. Simplifying sample preparation for soil fertility analysis by X-ray fluorescence spectrometry. *Sensors* **2019**, 19, 5066, doi: 10.3390/s19235066.

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

Trevisan, R. G., Bullock, D. S., Martin, N. F. Spatial variability of crop responses to agronomic inputs in on-farm precision experimentation. *Precision Agriculture* **2021**, 22, 342-363, doi: 10.1007/s11119-020-09720-8.

USEPA – United States Environmental Protection Agency. Method 6200: Field portable X-ray fluorescence spectrometry for the determination of elemental concentrations in soil and sediment, 2007. Disponível em: http://www.epa.gov/osw/hazard/ testmethods/sw846/pdfs/6200.pdf. Acessado em: 11 de julho de 2019.

Van Raij, B. *Fertilidade do solo e manejo de nutrientes*. International Plant Nutrition Institute (IPNI): Piracicaba, São Paulo, Brazil, 2011; 420p. (In Portuguese)

Veseth, R. Managing variable soils. STEEP Extension Farming Update, Fall issue, 1986. 29-33p.

Villas-Boas, P.R.; Romano, R.A.; Franco, M.A.M.; Ferreira, E.C.; Ferreira, E.J.; Crestana, S.; Milori, D.M.B.P. Laser-induced breakdown spectroscopy to determine soil texture: A fast analytical technique. *Geoderma* **2016**, 263, 195-202. doi: 10.1016/j.geoderma.2015.09.018.

Viscarra Rossel, R.; Walvoort, D.J.J.; McBratney, AB.; Janik, LJ.; Skjemstad, J.O.;Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. *Geoderma* **2006**, 131, 59-75, doi: 10.1016/j.geoderma.2005.03.007.

Viscarra Rossel, RA.; Rizzo, R.; Demattê, JAM.; Behrens, T. Spatial modeling of a soil fertility index using visible–near-infrared spectra and terrain attributes. *Soil Science Society of America Journal* **2010**, 74, 1293-1300. doi: 10.2136/sssaj2009.0130.

Viscarra Rossel, RA.; Adamchuk, VI.; Sudduth, KA.; McKenzie, NJ.; Lobsey, C. Proximal soil sensing: an effective approach for soil measurements in space and time. *Advances in agronomy* **2011**, 113, 243-291.

Viscarra Rossel, R.A.; Bouma, J. Soil sensing: A new paradigm for agriculture. *Agricultural Systems* **2016**, 148: 71-74. doi: 10.1016/j.agsy.2016.07.001

Viscarra Rossel, R.A.; Lobsey, C. Scoping review of proximal soil sensors for grain growing. CSIRO, Australia, 2016.

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

Wang, Z.; Dean, T.; Kowalski, B.R. Additive background correction in multivariate instrument standardization. *Analytical Chemistry* **1995**, 67, 2379-2385. doi: 10.1021/ac00110a009.

Webster, R.; Mc Bratney, A.B Mapping soil fertility at Broom's Barn by simple kriging. Journal of the *Science of Food and Agriculture* **1987**, 38, 97-115, doi: 10.1002/jsfa.2740380203.

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

Weindorf, DC.; Chakraborty, S. Portable X-ray Fluorescence Spectrometry Analysis of Soils. In: *Methods of Soil Analysis*, SSSA Book Ser. 5. SSSA, Madison, WI, 2016. doi:10.2136/methods-soil.2015.0033.

Weltzien, C. Digital Agriculture or Why Agriculture 4.0 Still Offers Only Modest Returns. *Landtechnik* **2016**, 71, 66-68. doi: 10.15150/lt.2015.3123

Wetterlind, J.; Stenberg, B.; Söderström, M. Increased sample point density in farm soil mapping by local calibration of visible and near infrared prediction models. *Geoderma* **2010**, 156, 152-160. doi: https://doi.org/10.1016/j.geoderma.2010.02.012.

Whetton, R; Zhao, Y.; Mouazen, A.M. Quantifying individual and collective influences of soil properties on crop yield. *Soil Research* **2018**, 56, 19-27. doi: 10.1071/SR16264.

Wijewardane, N.K.; Ge, Y.; Morgan, C.L. Moisture insensitive prediction of soil properties from VNIR reflectance spectra based on external parameter orthogonalization. *Geoderma* **2016**, 267, 92-101. doi: 10.1016/j.geoderma.2015.12.014.

Wollenhaupt, N.C.; Wolkowski, R.P.; Clayton, M.K. Mapping soil test phosphorus and potassium for variable-rate fertilizer application. *Journal of Production Agriculture* **1994**, 7, 441-448, doi: 10.2134/jpa1994.0441.

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

Yu, K.Q.; Zhao, Y.R.; Liu, F.; He, Y. Laser-induced breakdown spectroscopy coupled with multivariate chemometrics for variety discrimination of soil. *Scientific reports* **2016**, 6, 27574. doi: 10.1038/srep27574.

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

Connecting text for chapter 3

The ideal sample preparation for applying elemental analysis sensors consists of preparing pellets. This procedure involves grinding the sample—until a particle size generally less than 50 μm (using, *e.g.,* ball or cryogenic mill)—followed by pressing the particles. This procedure is time-consuming and requires specific equipment (*e.g.,* a mill and a hydraulic press); therefore, it is not suitable with the practicality required for PSS applications. Although this procedure is, so far, mandatory for LIBS analyses, its simplification can be considered for XRF sensors.

In chapter 3 we evaluated the possibility of simplifying this sample preparation for the XRF sensor application. Samples only dried and sieved (< 2 mm) were evaluated and compared with the XRF performance in pellets. We emphasize that this simplified sample preparation is similar to the one already established for analysis of soil samples using VNIR sensors in a controlled environment. The satisfactory performance of this method makes it possible to evaluate soil samples simultaneously with both sensors.

The study conducted in this third chapter fulfills part of the third specific objective of this thesis, which proposes to evaluate the applicability of XRF for *in-situ* soil fertility analysis. Part of this chapter was presented at the 5th Global Workshop on Proximal Soil Sensing and and the full version was published as an original article in the special edition "Smart Sensing Technologies for Agriculture" of the Sensors Journal. The reference of these publications are listed bellow.

Tavares, T. R.; Nunes, L. C.; Alves, E.E.N.; Krug, J. F.; Carvalho, H.W.P.D.; and Molin, J. P. Sample preparation for assessing the potential use of XRF on soil fertility analysis. 5th Global Workshop on Proximal Soil Sensing, Columbia – EUA, May 28th to 31st, 2019.

Tavares, T.R.; Nunes, L.C.; Alves, E.E.N.; Almeida, E.D.; Maldaner, L.F.; Krug, F.J.; Carvalho, H.W.P.D.; and Molin, J.P., 2019. Simplifying Sample Preparation for Soil Fertility Analysis by X-ray Fluorescence Spectrometry. *Sensors* **2019**, 19(23), p.5066, doi: 10.3390/s19235066.

CHAPTER 3. SIMPLIFYING SAMPLE PREPARATION FOR SOIL FERTILITY ANALYSIS BY XRF SPECTROMETRY

Abstract

X-ray fluorescence (XRF) sensors allow one to collect digital data in a practical and environmentally friendly way, as a complementary method to traditional laboratory analyses. This work aimed to assess the performance of a XRF sensor to predict exchangeable nutrients in soil samples by using two contrasting strategies of sample preparation: pressed pellets and loose powder (<2 mm). Pellets were prepared using soil and a cellulose binder at 10% w w^{-1} followed by grinding for 20 min. Sample homogeneity was probed by X-ray fluorescence microanalysis. Extractable (ex-) nutrients were assessed by XRF furnished with a Rh X-ray tube and silicon drift detector. The calibration models were obtained using 58 soil samples and leave-one-out cross-validation. The predictive capabilities of the models were appropriate for both ex-K and ex-Ca determinations with coefficient of determination (R^2) ≥ 0.76 and ratio of performance to interquartile range (RPIQ) > 2.5. Although XRF analysis of pressed pellets allowed a slight gain in performance over loose powder samples for the prediction of ex-K and ex-Ca, satisfactory performances were also obtained with loose powders, which require minimal sample preparation. The prediction models with local samples showed promising results and encourage more detailed investigations for the application of XRF in tropical soils.

Keywords: 1. Physical matrix effect 2. Pelletizing 3. Proximal soil sensing 4. Soil testing.

3.1. Introduction

X-ray fluorescence (XRF) is a spectroanalytical technique compatible with direct soil analysis, which can be applied with minimal or no sample preparation (Gredilla et al., 2016). The recent technological advance of the optical and electronic components allowed the development and miniaturization of this technology, and it has become attractive for use in hybrid laboratories and *in-situ* analyses. Some studies have already pointed out the potential of using XRF sensors in proximal soil sensing (PSS) approaches (Weindorf and Chakraborty, 2016; Nawar et al., 2019). Despite that, XRF has been poorly explored for assessments of physical and chemical attributes of tropical soils, mainly under the context of PSS and precision agriculture.

To use XRF sensors as a practical analytical method in hybrid laboratories—in order to ensure a massive increase in the number of samples analyzed—it should be compatible with a simple soil sample preparation (*e.g.*, just air-dried and sieved rapidly). Recent studies involving XRF sensors for practical analysis of soil attributes have used dried samples with particle sizes smaller than 2 mm (Sharma et al., 2014, 2015; O'Rourke et al., 2016; Nawar et al., 2019). It is a consensus that pellet preparation after grinding the soil allows one to explore the potential of the XRF technique in soil analysis (Jantzi et al., 2016). The preparation of a pellet is recommended for analyses with the XRF

technique because it improves the homogeneity of the material and also allows one to control the density, porosity and surface roughness characteristics, reducing the physical matrix effects (Takahashi et al., 2015). Although it is known that the preparation of pellets guarantees better precision in the measurements performed with the XRF (Krug and Rocha, 2016; Shibata et al., 2009), recent studies have assumed that, when analyzing soil samples with particle sizes smaller than 2 mm, its heterogeneity and physical matrix effects can be neglected.

XRF analyses are more flexible with regards to sample preparation because—unlike other elemental analysis techniques, such as laser induced breakdown spectroscopy (LIBS)—they also allow one to evaluate loose powder (Shibata et al., 2009). However, we did not find any study comparing XRF performance for the prediction of fertility attributes on soil samples that were just dried and sieved (<2 mm) with samples prepared with the optimal sample preparation method. Therefore, the level of performance loss when neglecting the physical matrix effect and heterogeneity is unknown. For a robust development of the XRF technique as a practical tool for soil fertility analysis, one of the key points is to understand the tradeoff between analytical performance and sample preparation, in order to reduce or eliminate these procedures based on the analytical potential of the sensor for each sample condition. Such knowledge is important for the development of PSS applications using this tool.

To evaluate the possibility of simplifying the sample preparation procedure for XRF analyses, this work aimed to assess the performance of a portable XRF (pXRF) to predict extractable nutrients in soil samples prepared using two contrasting types of sample preparation: pellets and loose powder (≤2 mm). The effect of sample preparation in the spatial distribution of nutrients on the sample surface was evaluated using a benchtop microprobe X-ray fluorescence spectrometry (μ-XRF). Moreover, a procedure for preparation of soil pellets, involving planetary ball milling and the use of binding agents was also assessed.

3.2. Material and Methods

3.2.1. Soil samples

A set of 58 soil samples were selected for the comparison of their exchangeable nutrients content with the X-ray fluorescence produced by the pellet and loose powder samples. These samples were collected from 0 to 0.2 m soil depth in an agricultural field located at the southeast region of Brazil, in the municipality of Piracicaba, state of São Paulo (at coordinates 22°41'57.24" S and 47°38'33.33" W, WGS84 datum). The soil is classified as Lixisol (IUSS, 2014) with a clayey texture and high nutrient variability.

The soil samples were air-dried and sieved (<2 mm) and after that, three subsamples were separated: (i) 0.8 g was used for pelletizing, (ii) 10 g was analysed as loose powder, and (iii) about 30 g was used for the reference measurements.

3.2.2. Sample preparation

For pelletizing, the samples (particles < 2 mm) were initially dried at 105 °C for 24 h and thereafter ground in a planetary ball mill (Retsch model PM 200 mill, Germany) (Figure 1A) by using two grinding tungsten carbide jars (50 mL; Retsch, Germany) with 10 tungsten carbide balls (10 mm diameter) (Figure 1B). Grinding was performed at 400 rpm for 5 min clockwise/5 min counter clockwise with a 10-s stop before changing the rotation direction.

Preliminary experiments were carried out by just pressing the soil samples without a binder. It was observed that for the sandier sample (clay content of 175 g dm^{-3}) the resulting pellets were friable and easily crumbled (Figure 2A). Therefore, binder addition was decisive for improving the quality of the pellets. The binders tested were chosen based on their similarity to the analytical blank, *e.g.*, lower analyte mass fractions of elements evaluated in the soil fertility (*e.g.*, P, K, Ca, and Mg). In this case, binding agents, such as a microcrystalline cellulose powder (Sigma-Aldrich, Merck, Darmstadt, Germany), and cellulose (SPEX 3642, Metuchen, NJ, USA) were evaluated in the proportion of 10 and 15% w w^{-1}, with grinding/homogenization times of 10, 15, and 20 min.

Figure 1. Planetary ball mill (**A**), loose soil inside the tungsten carbide jars with the tungsten carbide balls (**B**) and hydraulic press (**C**), which were used at work.

The grinding and homogenized samples were pelletized in a hydraulic press (SPEX 3624B X-Press) (Figure 1C) by transferring 0.8 g of the powdered material to a stainless steel set and applying 8.0 t cm^{-2} for 3 min. Cylindrical pellets were approximately 15 mm diameter and 2 mm thick, with mass per

unit area of 0.45 g cm^{-2}. The pellets were visually inspected, evaluating their homogeneity aspect and integrity. Furthermore, an XRF spectra (obtained with pXRF, as described below in Section 3.2.5.) of a pellet and a loose soil sample were also compared in order to assess possible contamination during the milling process. Further experiments were performed with 10% w w^{-1} cellulose binder (Sigma-Aldrich, Merck, Darmstadt, Germany) and 20 min of grinding in a planetary ball mill.

Sample presentation in the form of loose powder was also considered for the analysis. The air-dried samples were sieved in a sieve with apertures of 2 mm. Ten grams of test sample was transferred to an XRF polyethylene cup of 31 mm (*n*. 1530, Chemplex Industries Inc., Palm City, FL, USA) assembled with a 4-μm thick polypropylene film (*n.* 3520, SPEX, USA).

3.2.3. Soil laboratory analysis

Soil testing conducted by a commercial laboratory determined the extractable (ex-) contents of P, K, Ca and Mg via ion exchange resin extraction. Clay content was quantified by the Bouyoucos hydrometer method in dispersing solution. The pseudo total content (ptc) of P, K, Ca and Mg were also analyzed following the USEPA Method 3051A (Element, 2007). The latter methods involve the extraction of ions using HNO_3 and HCl. The multielement quantification was made by inductively-coupled plasma optical emission spectrometry (ICP OES). The term ptc is used, because it is not a total digestion method. Despite this, this method presents proportional recoveries to the most aggressive methods for the determination of elements in tropical soils (Nogueirol, 2013), allowing one to understand the relationship between extractable and total content of the elements evaluated. In addition, it is a method that requires less time for digestion, less consumption of acids and lower risks of environmental contamination (Silva et al., 2014).

3.2.4. μ-XRF chemical images

μ-XRF is a type of energy dispersive X-ray fluorescence that employs a micrometric beam with a shape and size defined by a primary optic element; this can be done by a simple collimator, an optical capillary or a focusing mirror (Tian et al., 2014; Rodrigues et al., 2018). In this work, the μ-XRF technique was employed to characterize, on the sample surface, the influence of the sample preparation method on the spatial distribution of elements of interest.

The net intensities for K and Ca Kα emission lines were characterized with high spatial resolution on the surface of loose soil and pellet samples. A benchtop μ-XRF system (Orbis PC EDAX, United States) furnished with a Rh anode X-ray tube was used. The detection was carried out by a 30

mm² silicon drift detector (SDD). The μ-XRF tube current and voltage was operated at 15 kV and 200 μA, respectively; the beam size used was 30 μm and no primary filter was used; the live time was set to 2 s per spot; and the analysis was carried out under vacuum. In each sample, 800 points (matrix of 32 × 25 points) were evaluated in an area of about 2.32 mm² (1.60 × 1.45 mm).

Chemical images showing the variability of K and Ca, produced by Orbis Vision software, were linearly interpolated using Origin Lab 2016. The mean, maximum and minimum values, as well as the coefficient of variation (CV)—the ratio between the standard deviation and the mean expressed in percentage—were also calculated. P and Mg Kα emission lines were not identified in the samples, which did not allow the evaluation of these element lines. Similar μ-XRF analysis procedures are described by Rodrigues et al. (2018).

3.2.5. pXRF measurements and its performance evaluation

The measurements were carried out using a portable energy dispersive X-ray fluorescence spectrometer, Tracer III–SD model (Bruker AXS, Madison, USA), equipped with a 4 W Rh X-ray tube and 12 mm² of active area, and a X-Flash® Peltier-cooled SDD, with 2048 channels (Bruker AXS, Madison, USA). The tube operated at 23 μA and 15 kV, and emission intensities were measured for 90 s without vacuum. The voltage configuration was chosen based on the interest in low atomic number elements and the current, in order to keep deadtime below 15%, and avoiding spectral distortions and artifacts. Soil samples were measured in triplicate at different portions of its surface. To ensure the same attenuation conditions of the loose soil samples, the pellets were placed on a 4-μm-thick polypropylene thin-film.

All data were acquired using the software Bruker S1PXRF® (Bruker AXS, Madison, USA). The data were obtained through the deconvolution process using the Artax® (Bruker AXS, Madison, USA). The Kα emission characteristic lines of the elements P, K, Ca and Mg were evaluated. However, only K and Ca presented detectable emission lines, which were evaluated by their signal-to-noise ratio (SNR) and intensity, through the counts of photons per second (cps). The SNR was determined by dividing the characteristic X-ray net intensities by the background square root (Ernst et al., 2014), and the cps were obtained by the ratio of total X-ray intensity to detector live time. The standard deviation (SD) behavior of the intensity of the K and Ca Kα emission lines within the replicates was also evaluated for both sample preparations.

The intensity of K and Ca Kα emission lines, obtained from pellets and loose soil samples, were compared with the exchangeable contents of these nutrients. Calibration models were built using simple linear regressions (a univariate model), with the independent variable being pXRF data and the dependent variable being the soil property measured via commercial laboratory procedures. The prediction models

were validated using "leave-one-out" full cross-validation. The quality of the developed calibration was assessed with the coefficient of determination (R^2), the root mean square error (RMSE) and the ratio of performance to interquartile range (RPIQ), as recommended by Bellon-Maurel et al. (2010). The following arbitrary groups were used for simplifying the interpretation, as proposed by Nawar and Mouazen (2017): (1) excellent models (RPIQ > 2.5), (2) very good models (2.5 > RPIQ > 2.0), (3) good model (2.0 > RPIQ > 1.7), fair (1.7 > RPIQ > 1.4), and very poor model (RPIQ < 1.4). The descriptive statistics of soil fertility and the correlation between pseudo total and exchangeable contents were also determined.

3.3. Results

3.3.1. Soil pelletizing procedure

The pelletizing of sandy soil samples (*e.g.*, about 175 g dm^{-3} of clay content) was only possible with the addition of binder. In the initial tests, which evaluated different grinding times, they did not form pellets without the addition of binder (Figure 2A). In general, pellets produced after 10 and 15 min of grinding were brittle, except for pellets containing 15% w w^{-1} of binder (Figure 2B), which, in turn, were less homogeneous with white spots on their surface. The best cohesion between particles was obtained with 20 min of grinding. For this milling time, the pellet with 15% w w^{-1} of binder was slightly less heterogeneous than the pellet with 10% w w^{-1}. In relation to the brand, microcrystalline cellulose powder (Sigma-Aldrich, Merck, Darmstadt, Germany) presented more cohesive pellets. The best results were observed for pressed pellets prepared from soil mixed with cellulose binder at 10% w w^{-1} and ground for 20 min.

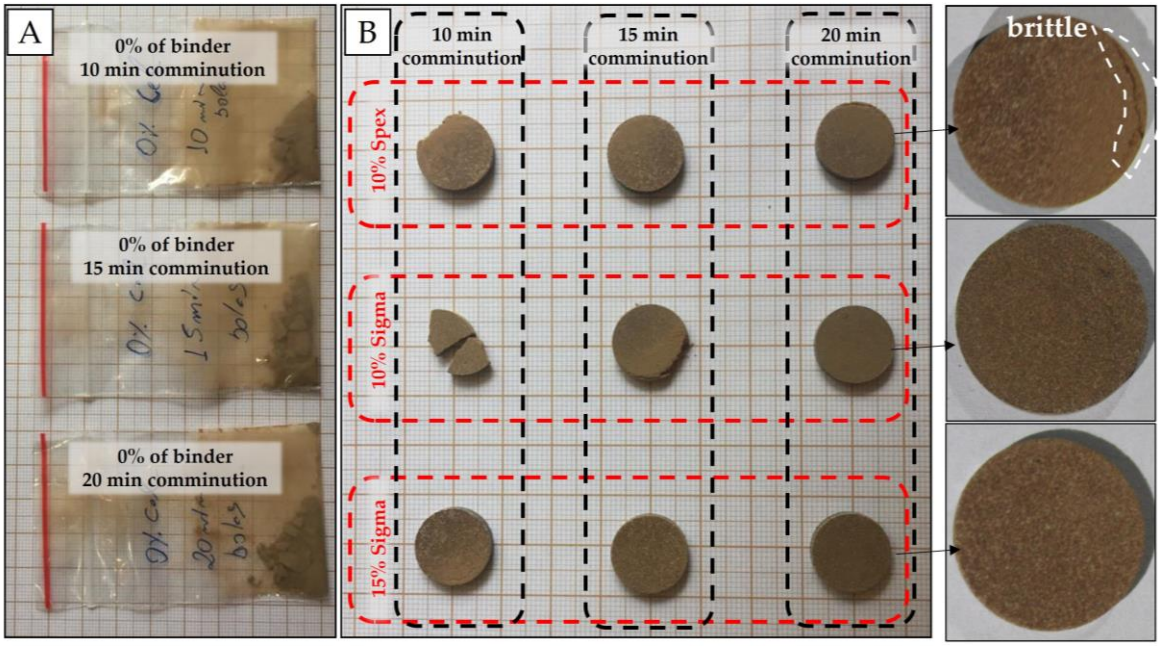


Figure 2. Soil samples without the addition of binder and with different grinding times (A); pellets resulting from tests with different grinding times, cellulose concentrations and brands (B).

Tungsten (W) contamination was observed in these ground soil samples. This contamination was caused by the tungsten carbide ball mill and it was evidenced by the W L-emission lines presented in the pellet spectrum, which were not observed in the loose soil spectrum (Figure 3). In the XRF spectra, W presents L and M-emission lines with energy lower than 15 keV, as highlighted in the red lines in the spectrum of Figure 3. In this range, the main W emission lines present energy of 1.77 (Mα), 1.83 (Mβ), 8.39 (Lα), 9.67 (Lβ_1) and 9.95 keV (Lβ_2). The effect of this contamination is best seen on L-emission lines, as they do not overlap with any other emission lines present. The W M-α line overlaps with Si (1.74 keV) and can also promote the enhancement of the Al Kα line (Al K_{edge} = 1.55 keV) due to secondary radiation excitation (chemical matrix effect). Thus, this W interference must be considered and corrected in the case of Al quantification. For the Ca and K determinations performed in this work, contamination with W was not a limiting factor.

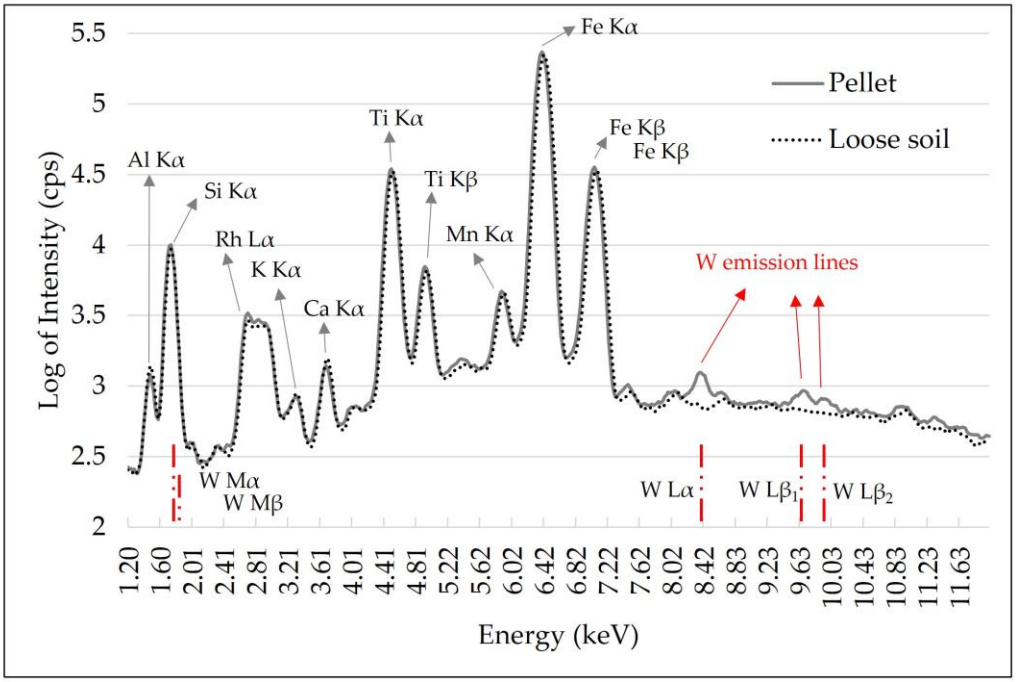


Figura 1. X-ray fluorescence spectra obtained with pXRF equipment using the pellet and loose soil sample. Tungsten (W) emission lines were identified with red lines. The emission spectra intensity is shown in the logarithm to reduce the differences in scales between the emission lines, allowing a better qualitative assessment.

3.3.2. µ-XRF chemical images and sample homogeneity

The spatial distribution patterns of Ca and K at the pellet and loose soil surface are shown in Figure 4. For both elements, the preparation of pellets resulted in more homogeneous surfaces than those observed for loose soil. The particle size reduction—required for pellet production—allows one to homogenize the distribution of the different elements in the sample, which occurs because it fragments

the regions where these elements are agglomerated (nuggets). In the loose soil sample, the presence of nuggets can be observed for the Ca and K (Figure 4B,C, respectively), which do not appear in the pelletized samples. The homogenization promoted by pelletizing drastically reduced the CV of both Kα emission line intensities, oscillating from 100.17% to 13.03% for Ca and from 46.09% to 18.01% for K, respectively.

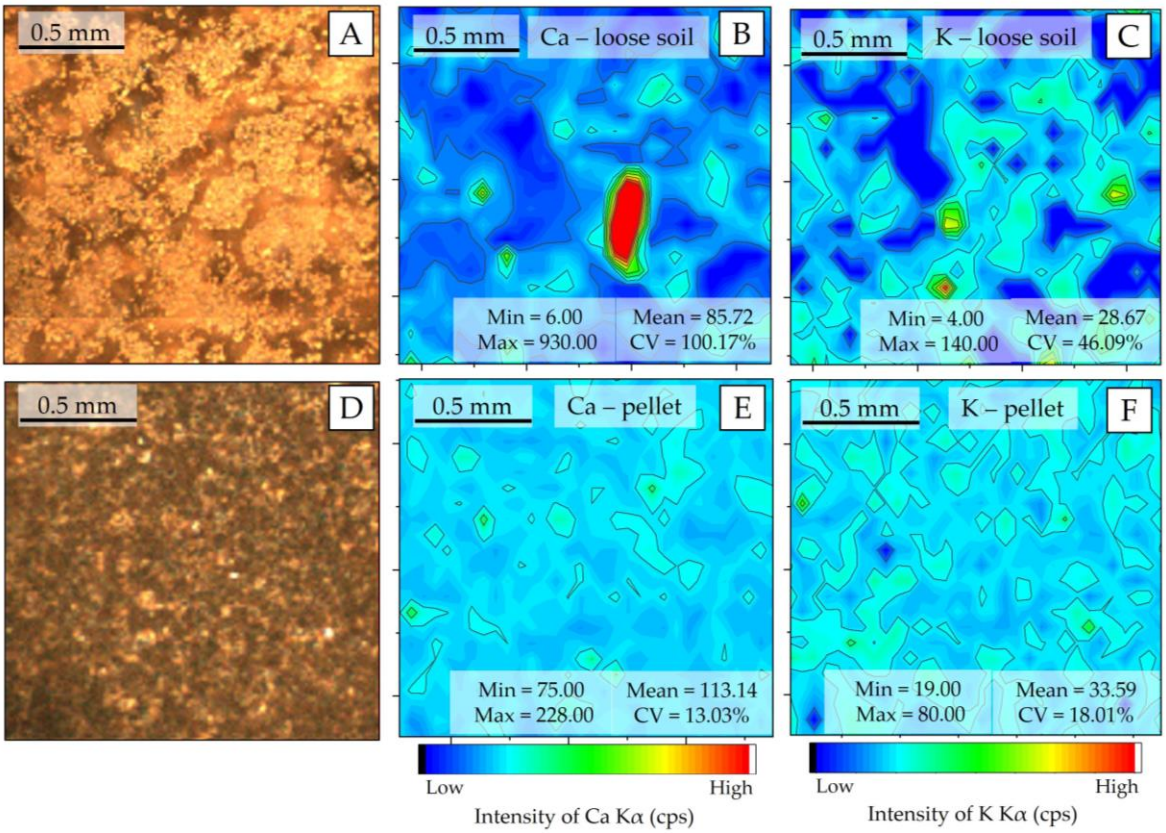


Figure 4. μ-XRF chemical images showing the spatial distribution of Ca and K in the loose soil sample (B and C, respectively); Ca and K in the pellet sample (E and F, respectively). Photo of the analyzed area of (A); the loose soil (A) and pellet (D).

3.3.3. Soil extractable nutrient prediction using a pXRF spectrometer

Soil samples were characterized by clayey texture, low variability of clay content (between 345 and 511 g dm^{-3}) and high variability of extractable nutrients (Table 1). According to the local fertility interpretation (Van Raij et al., 2001), the level of ex-P content oscillates between low to medium; and the level of ex-K, ex-Ca and ex-Mg content varies between medium to very high. These samples are also characterized by a significant correlation between extractable and pseudo total contents for all nutrients. A descriptive summary of these analyses is presented in Table 1.

Table 1. Descriptive statistics of extractable and pseudo total nutrients and the correlation between the respective nutrients.

	Exchangeable Nutrients				Pseudo Total Contents			
	ex-P	ex-K	ex-Ca	ex-Mg	ptc P	ptc K	ptc Ca	ptc Mg
	mg dm^{-3}	$mmol_c$ dm^{-3}			mg kg^{-1}			
Min	7.00	1.70	27.00	11.00	405.31	154.04	492.42	411.59
Mean	20.20	5.14	49.12	26.28	489.93	318.10	750.51	607.38
Max	46.00	10.30	78.00	54.00	669.18	477.18	1225.91	789.65
SD	8.42	1.78	12.89	10.66	55.74	79.00	159.46	112.31
CV (%)	41.69	34.60	26.25	40.57	11.38	24.84	21.25	18.49
Correlation with pseudo total	0.79 *	0.67 *	0.83 *	0.52 *				

* Significant correlation at the probability level of 0.01.

The qualitative evaluation of the XRF spectra (Figure 3) allowed us to identify the K and Ca emission lines, but no fluorescence emission was detected for Mg and P. The XRF intensity and the SNR of K and Ca were slightly higher for loose soil than pellet samples, and both had a highly significant correlation ($r > 0.9$) between pellet and loose soil, indicating that the changes promoted by sample preparation were well standardized for all samples (Figure 5). Despite the small gain in fluorescence intensity (an average of 4.21 and 15.31 cps for K and Ca, respectively) and in SNR (an average of 0.63 and 2.49 for K and Ca, respectively), when evaluating the behavior of the emission line intensity SD for the replicates, we can observe that the loose soil samples presented greater variation in relation to the pellets, both for K and Ca (Figure 5E,F, respectively). The replicate SD is an indicator of the reading precision. In this work, the lower precision of the loose soil samples might be related to their lower homogeneity in relation to the pelletized ones. Despite this loss of precision among the different replicates obtained in loose soil samples, the triplicate scans smoothed this effect. After averaging the replicates, the distribution of the XRF intensity showed a similar distribution between both sample preparations (Figure 5A,B).

Regarding the regression analysis, there was a slight reduction of precision in the calibration of ex-K and ex-Ca in loose powder soil samples, marked by a slight increase in error and reduction in R^2 (Figure 6). Comparing the loose soil samples in relation to the pellet samples, the prediction error of ex-K increased from 0.65 to 0.78 $mmol_c$ dm^{-3}. Similarly, for the ex-Ca prediction, the error increased from 5.89 to 6.12 $mmol_c$ dm^{-3}. Moreover, concerning the R^2 values, it oscillated from 0.87 to 0.81 for ex-K, and from 0.78 to 0.76 for ex-Ca. Besides that, all prediction models, obtained from both pellet and loose soil, showed excellent performance in their validation with RPIQ values above 2.5.

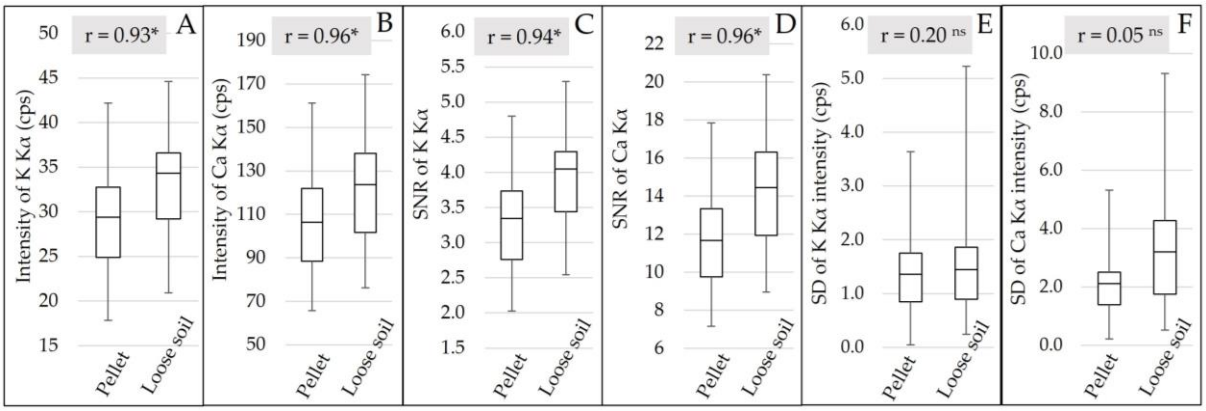


Figure 5. Box plot of the Kα emission line intensity of K and Ca (A and B, respectively), after averaging the replicates. Box plot of the signal-to-noise ratio (SNR) of the Kα emission lines of K and Ca (C and D, respectively), after averaging the replicates. Box plot of the standard deviation (SD) of the Kα emission line intensity of K and Ca (E and F, respectively) for the replicates. The Pearson correlation between the pellet and loose soil is also presented (correlations followed by * were significant at the probability level of 0.01; correlations followed by ^{ns} were not significant).

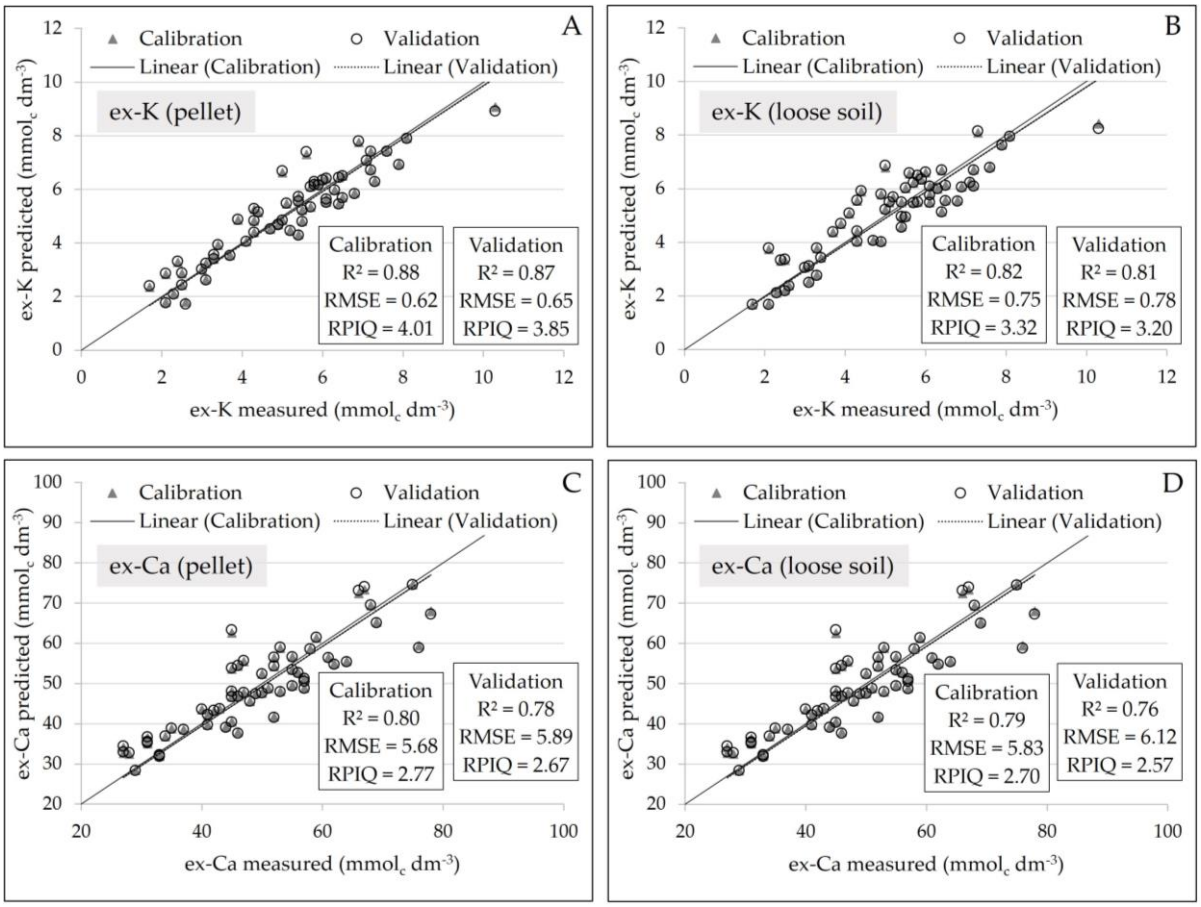


Figure 6. Scatter plots of measured versus predicted ex-K, for the pellets and loose soil (A and B, respectively), and of measured versus predicted ex-Ca, for the pellets and loose soil (C and D, respectively). Models were obtained using simple linear regressions with the Kα emission lines of each element (n = 58) and the validation was performed by "leave-one-out" full cross-validation.

3.4. Discussion

Even if the XRF technique is flexible concerning sample preparation, there is a consensus that pellet preparation increases the data precision (Krug and Rocha, 2016). Pellet preparation consists of conforming and binding the samples into a specific shape. For pellet formation, the soil samples must be grinded to an extremely fine powder by using a grinder; furthermore, a proper binding agent can also be necessary (Gondal et al., 2007). In this work, a procedure for soil pellet preparation was evaluated, with and without binding agents. Sandy soils are more likely to not form pellets without the use of a binder (Jantzi et al., 2016). In our work, it was perceived for samples with clay content around 175 g dm^{-3}. During testing for optimizing binder concentration and ball milling time, it was observed that increasing the grinding time, as well as the binder concentration, improves the cohesion between the particles of the pellet. In contrast, higher concentrations of binder (*e.g.*, 15% w w^{-1}, in this work) make it difficult to homogenize it. It is known that the smaller the particle size to be pressurized is, the more resistant and cohesive the pellet will be (Carvalho et al., 2015). However, pellet preparation in different soil sample sets can be optimized with different binder concentrations and milling time. To optimize these conditions, we suggest conducting preliminary tests, as described in this Chapter.

The reduction of the particle size, before pressing the material, promotes sample homogenization (Gondal et al., 2007). Both the element spatial patterns and the SD behavior on the replicates, clearly showed gain on homogeneity after pelletizing. The lower homogeneity of loose soil samples should be considered for determining the number of replicates during data acquisition with a pXRF device. Due to the high variability of element distribution in loose soil samples, a greater number of scans has to be acquired for more accurate representation of the sample surface, and then, different spectra can be averaged (Jantzi et al., 2016). In this work, pXRF spectra were obtained in triplicate (at different positions) and were sufficient for a good characterisation of both sample preparation samples.

One point to consider while milling samples is the possibility of contamination. Two types of contamination can occur, cross-contamination, due to inefficient cleaning when changing samples, and/or contamination with milling surfaces (*e.g.*, agate and tungsten carbide), due to the abrasion between the sample and mill components. In addition to the careful execution of cleaning procedures, the milling surfaces should be considered in terms of hardness and elemental composition to avoid the risk of sample contamination (Jantzi et al., 2016). In this work, the grinding in a ball mill made of tungsten carbide contaminated the samples with W. Soil samples usually have hard minerals, like silicates, and W should not be measured when tungsten carbide milling jars and balls are used for milling soil (Jantzi and Almirall, 2014). Specifically, in XRF analysis, W contamination can also be a problem for the determination of Si and Al, as described in the previous section. Iwansson and Landström (2000) showed that this kind of contamination is higher in quartz-rich samples and increases

with grinding time. When the element contamination (*e.g.*, W) is the same one that is to be quantified, the values must be adjusted, discarded, or over-looked to avoid misinterpretation of the results (Iwansson and Landström, 2000).

Pelletizing also reduces surface roughness effects and increases the density of the material (Gondal et al., 2007). Theoretically, reducing sample roughness means decreasing the physical matrix effect, which would attenuate part of the fluorescence produced by the analytes, and increasing the density of samples and fluorescence intensity (Shibata et al., 2009). Thus, a higher fluorescence yield and an upsurge in the SNR were expected for the emission lines of the pelletized samples, due to the increase of their density and reduction of the physical matrix effect (Takahashi, 2015). Nevertheless, this behavior was not observed in this work. In turn, the addition of binder (10% w w^{-1}), as well as the contamination by W, and the differences in homogeneity found (Figure 4B,E), appear to be the factors that influenced this behavior, slightly altering the chemical composition of the pellet samples and, consequently, their fluorescence production.

Soil samples are naturally heterogeneous and therefore comminution procedures are generally recommended for improving matrix homogenization, and should yield homogeneous pellets (Gondal et al., 2007). Ultimately, this can avoid heterogeneity effects, such as grain size effect, mineralogical effect, and segregation, factors that cause errors in the XRF analysis (Morikawa, 2014). However, in this work, the prediction models of ex-Ca and ex-K using pellets showed just a small performance gain over those obtained in loose powder soil samples. Using loose powder soil, prediction models for ex-Ca and ex-K calibrated with the 58 local samples obtained excellent performances, with RPIQ values over 2.5. These results are promising and encourage more detailed investigations on the application of the XRF technique.

The possibility of XRF application on soil samples that have just been dried and sieved (<2 mm), with satisfactory predictions of ex-K and ex-Ca using local models, is a promising alternative to increasing the efficiency of analytical procedures. In addition, this level of sample preparation is compatible with evaluations using VNIR diffuse reflectance, opening the potential to exploit joint XRF and VNIR sensors on these types of samples.

A simple calibration method was applied in this work, using only the emission line of the elements of interest for predictive modeling. XRF spectra are multi-informational, allowing the measurement of several soil properties from a single scan. This is possible because each spectrum stores a large amount of information along with its emission lines and different types of scattering (*e.g.*, Compton and Thomson scattering), not strictly related to the elementary constituents of the samples (Morona et al., 2017). In this sense, predictive modeling based on multivariate statistics and machine learning methods are an alternative to better exploit the hidden information present in XRF spectra

(Panchuk et al., 2018), and it can enable robust determinations of fertility attributes that have an indirect relationship with inorganic soil constituents such as pH, cation exchange capacity, base saturation and texture.

Finally, it is worth noting that this study was conducted under a clayey lixisol, which is a representative and common type of soil in Brazilian tropical areas (EMBRAPA, 2018). Therefore, this pioneering evaluation provided useful information to help XRF users—who aim to use this technique as a tool for practical soil analysis—to understand that the expected effects of sample preparation related to heterogeneity and physical matrix effects can be neglected. However, it is fundamental to also validate other types of soils with contrasting textural classes, which can proportionate distinct levels of physical matrix effects.

3.5. Conclusion

The addition of a binder was decisive for improving the quality of the pellets. The best results for soil preparation in the form of pellets were obtained with samples prepared with cellulose binder at 10% w w^{-1} and ground for 20 min.

Pressed pellets allowed a slight gain in performance over loose powder samples for the prediction of ex-K and ex-Ca. In spite of that, predictions in loose powder soil for ex-Ca and ex-K, calibrated with 58 local samples, obtained excellent performances in their validation, showing that it is possible to reduce the optimal sample preparation of XRF analyses for predicting soil nutrients. However, loose samples are less homogenous than pellets, and scanning loose soil samples in replicates is important for smoothing this effect.

The prediction models of ex-K and ex-Ca calibrated with local samples presented promising results. More detailed investigations are necessary to foster the application of the XRF technique in agricultural soil samples for determination of soil fertility attributes.

References

Bellon-Maurel, V.; Fernandez-Ahumada, E.; Palagos, B.; Roger, J.M.; McBratney, A. Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. *TrAC Trend. Anal. Chem.* **2010**, 29, 1073–1081, doi:10.1016/j.trac.2010.05.006.

Element, C.A.S. Method 3051A microwave assisted acid digestion of sediments, sludges, soils, and oils. *Z. Für Anal. Chem.* **2007**, 111, 362–366.

EMBRAPA Solos. *Brazilian Soil Classification System*, 5th ed.; EMBRAPA: Brasília, Brazil, 2018.

Ernst, T.; Berman, T.; Buscaglia, J.; Eckert-Lumsdon, T.; Hanlon, C.; Olsson, K.; Palenik, C.; Ryland, S.; Trejos, T.; Valadez, M.; et al. Signal-to-noise ratios in forensic glass analysis by micro X-ray fluorescence spectrometry. *X-Ray Spectrom.* **2014**, 43, 13–21, doi:10.1002/xrs.2437.

Gredilla, A.; de Vallejuelo, S.F.O.; Elejoste, N.; de Diego, A.; Madariaga, J.M. Non-destructive Spectroscopy combined with chemometrics as a tool for Green Chemical Analysis of environmental samples: A review. *TrAC Trend Anal. Chem.* **2016**, 76, 30–39, doi:10.1016/j.trac.2015.11.011.

IUSS Working Group WRB. World reference base for soil resources 2014. World Soil Resources Reports No. 106; Edited by Schad, P., van Huyssteen, C., Micheli, E.; FAO: Rome, Italy, 2014, 189p, ISBN 978-92-5-108369-7.

Iwansson, K.; Landström, O. Contamination of rock samples by laboratory grinding mills. *J. Radioanal. Nucl. Chem.* **2000**, 244, 609, doi:10.1023/A:1006769401251.

Jantzi, S.C.; Motto-Ros, V.; Trichard, F.; Markushin, Y.; Melikechi, N.; Giacomo, A. Sample treatment and preparation for laser-induced breakdown spectroscopy. *Spectrochim. Acta Part B* **2016**, 115, 52–63, doi:10.1016/j.sab.2015.11.002.

Krug, F.J.; Rocha, F.R.P. *Métodos de Preparo de Amostras Para Análise Elementar*; EditSBQ: São Paulo, Brazil, 2016; 572p. (In Portuguese)

Morikawa, A. Sample preparation for X-ray fluorescence analysis II. Pulverizing methods of powder samples. *Rigaku J.* **2014**, 30, 23–27.

Morona, F.; dos Santos, F.R.; Brinatti, A.M.; Melquiades, F.L. Quick analysis of organic matter in soil by energy-dispersive X-ray fluorescence and multivariate analysis. *Appl. Radiat. Isotopes* **2017**, 130, 13–20, doi:10.1016/j.apradiso.2017.09.008.

Nawar, S.; Mouazen, A.M. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *Catena* **2017**, 151, 118–129, doi:10.1016/j.catena.2016.12.014.

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

Nogueirol, R.C.; De Melo, W.J.; Bertoncini, E.I.; Alleoni, L.R.F. Concentrations of Cu, Fe, Mn, and Zn in tropical soils amended with sewage sludge and composted sewage sludge. *Environ. Monit. Assess.* **2013**, 185, 2929–2938, doi:10.1007/s10661-012-2761-3.

Rodrigues, E.S.; Gomes, M.H.F.; Duran, N.M.; Cassanji, J.G.B.; da Cruz, T.N.M.; Sant'Anna Neto, A.; Savassa, S.M.; de Almeida, E.; Carvalho, H.W.P. Laboratory Microprobe X-Ray Fluorescence in Plant Science: Emerging Applications and Case Studies. *Front. Plant Sci.* **2018**, 9, 1588, doi:10.3389/fpls.2018.01588.

Gondal, M.A.; Hussain, T.; Yamani, Z.H.; Baig, M.A. The role of various binding materials for trace elemental analysis of powder samples using laser-induced breakdown spectroscopy. *Talanta* **2007**, 72, 642–649, doi:10.1016/j.talanta.2006.11.039.

Carvalho, G.G.A.; Santos, D., Jr.; Gomes, M.S.; Nunes, L.C.; Guerra, M.B.B.; Krug, F.J. Influence of particle size distribution on the analysis of pellets of plant materials by laser-induced breakdown spectroscopy. *Spectrochim. Acta Part B* **2015**, 105, 130–135, doi:10.1016/j.sab.2014.09.001.

Jantzi, S.C.; Almirall, J.R. Elemental analysis of soils using laser ablation inductively coupled plasma mass spectrometry (LA-ICP-MS) and laser-induced breakdown spectroscopy (LIBS) with multivariate discrimination: Tape mounting as an alternative to pellets for small forensic transfer specimens. *Appl. Spectrosc.* **2014**, 68, 963–974, doi:10.1366/13-07351.

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

Panchuk, V.; Yaroshenko, I.; Legin, A.; Semenov, V.; Kirsanov, D. Application of chemometric methods to XRF-data—A tutorial review. *Anal. Chim. Acta* **2018**, 1040, 19–32, doi:10.1016/j.aca.2018.05.023.

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

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

Shibata, Y.; Suyama, J.; Kitano, M.; Nakamura, T. X-ray fluorescence analysis of Cr, As, Se, Cd, Hg, and Pb in soil using pressed powder pellet and loose powder methods. *X-Ray Spectrom.* **2009**, 38, 410–416, doi:10.1002/xrs.1195.

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

Takahashi, G. Sample preparation for X-ray fluorescence analysis III. Pellets and loose powder methods. *Rigaku J.* **2015**, 31, 26–30.

Tian, S.; Lu, L.; Xie, R.; Zhang, M.; Jernstedt, J.A.; Hou, D.; Ramsier, C.; Brown, P.H. Supplemental macronutrients and microbial fermentation products improve the uptake and transport of foliar applied zinc in sunflower (Helianthus Annuus L.) plants. Studies utilizing micro X-ray florescence. *Front. Plant. Sci.* **2015**, 5, 808, doi:10.3389/fpls.2014.00808.

Van Raij, B.; Andrade, J.C.; Cantarela, H.; Quaggio, J.A. *Análise química para avaliação de solos tropicais*; IAC: Campinas, Brazil, 2001; 285p. (In Portuguese)

Weindorf, D.C.; Chakraborty, S. *Portable X-ray Fluorescence Spectrometry Analysis of Soils*. In Methods of Soil Analysis; Hirmas, D., Madison, W.I., Eds.; Soil Science Society of America: Wisconsin, WI, USA, 2016; pp. 1–8, doi:10.2136/methods-soil.2015.0033.

Connecting text for chapter 4

Recent studies have evaluated the performance of XRF sensors for predicting soil fertility attributes. However, most of those studies have used pre-programmed packages [*e.g.*, Soil Mode (Innov-X Systems, Inc., MA, USA) and Geo Exploration (Bruker AXS, WI, USA)] for XRF data acquisition and processing, which are not fully transparent when implementing these actions. In addition, those packages have specific and non-flexible routines that limit the optimization of strategies for using XRF sensors in the precision agriculture context. A critical example of this lack of flexibility is the impossibility to evaluate the XRF performance with a reduced scanning time (*e.g.*, between 3 and 15 s),

which is mandatory for *in-situ* applications using this technique. Hence, it is necessary to establish transparent and optimized instrumental conditions (*e.g.*, X-ray tube configuration) and data processing (*e.g.,* pre-processing procedures and modeling approach) for assessing fertility attributes using XRF sensors. To the best of our knowledge, no previous study attempted to optimize those parameters for assessing soil fertility attributes.

Chapter 4 evaluates and suggests a novel methodology for XRF data acquisition and processing for soil fertility attributes assessment. The proposed method is simple and transparent, making it possible for further optimization related to XRF parameters, such as reducing its scanning time.

The study conducted in chapter 4 fulfills part of the first objective of this thesis and allows to proceed with the evaluation of the determining factors for *in-situ* applications with the XRF sensor, the third objective of this thesis (addressed in chapter 7). The main findings of chapter 4 were reported in two original articles, which were published in the journals Remote Sensing and Agronomy. The reference of these studies are listed bellow.

Tavares, T.R.; Molin, J.P.; Nunes, L.C.; Alves, E.E.N.; Melquiades, F.L.; Carvalho, H.W.P.; Mouazen, A.M. Effect of X-ray tube configuration on measurement of key soil fertility attributes with XRF. *Remote Sensing* **2020**, 12(6), p.963.

Tavares, T.R.; Mouazen, A.M.; Alves, E.E.N.; Santos, F.R.; Melquiades, F.L.; Carvalho, H.W.P.; Molin, J.P. Assessing Soil Key Fertility Attributes Using a Portable X-Ray Fluorescence: A Simple Method to Overcome Matrix Effect. *Agronomy* **2020**, 10(6), p.787.

CHAPTER 4. ASSESSING KEY SOIL FERTILITY ATTRIBUTES USING XRF: AN OPTIMIZED PROCEDURE FOR DATA ACQUISITION AND PROCESSING

Abstract

The successful use of X-ray fluorescence (XRF) sensors for soil analysis requires the selection of an optimal procedure of data acquisition and processing (*e.g.*, pre-processing and modeling). This Chapter has three objectives: (i) evaluate the performance of an XRF configured with different X-ray tube voltages (15 and 35kV) to predict fertility attributes [clay, organic matter (OM), cation exchange capacity (CEC), pH, base saturation (V), extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg)] in agricultural fields; (ii) evaluate the combined effect of Compton normalization (CN) and multivariate regressions [multiple linear regression (MLR) or partial least squares regression (PLSR)] on the attenuation of the soil matrix effects in order to improve the prediction accuracy of key soil fertility attributes; and (iii) compare the prediction performance achieved with the proposed solutions to that obtained from the Geo Exploration package (Bruker AXS, Madison, WI, EUA), a pre-developed measurement package commonly used by the scientific community. For that, 102 soil samples collected from two agricultural fields in Brazil with contrasting soil texture were used. The data analysis was divided in two parts. In part 1, two different X-ray tube configuration scenarios used to build the predictive models with MLR: (i) 10 emission lines of 15 keV spectra (EL-15), and (ii) 12 emission lines of 35 keV spectra (EL-35). In part 2, two pre-processing scenarios were applied to the emission lines of the EL-35 scenario, including with (ELC) and without (EL) the CN. Univariate regression models for the prediction of clay, CEC, ex-K and ex-Ca were compared with the corresponding MLR and PLSR models to assess matrix effect mitigation. The results of part 1 showed that the different X-ray tube configurations did not influence the prediction of the studied key fertility attributes. In addition, satisfactory predictions with residual prediction deviation (RPD) ≥ 1.54 and coefficient of determination (R^2) ≥ 0.70 were obtained for the eight out of nine studied soil fertility attributes (clay, OM, CEC, V, ex-K, ex-Ca, and ex-Mg). The results of part 2 showed the MLR and PLSR models improved the prediction results of the univariate models for both pre-processing methods (EL and ELC), proving to be promising strategies for mitigating the matrix effect. In turn, the CN also mitigated part of the matrix effect for ex-K, ex-Ca, and CEC predictions (not for clay), by improving the predictive performance of these elements when used in univariate and multivariate models. The prediction performances obtained using MLR and PLSR were comparable for all evaluated attributes. The combined use of CN with multivariate regressions (MLR or PLSR) achieved excellent prediction results for CEC (R^2 = 0.87), ex-K (R^2 ≥ 0.94), and ex-Ca (R^2 ≥ 0.96), whereas clay predictions were comparable with and without CN ($0.89 \leq R^2 \leq 0.92$). We suggest using multivariate regressions (MLR or PLSR) combined with the CN to remove the soil matrix effects and consequently result in optimal prediction results of the studied key soil fertility attributes. The prediction performance observed for this solution showed comparable results to the approach based on the preprogrammed measurement package tested (Geo Exploration package, Bruker AXS, Madison, WI, USA). The simple and transparent methodology suggested by this Chapter also enables future researches that seek to optimize the XRF scanning time in order to speed up the XRF analysis in soil samples.

Keywords: 1. Proximal Soil Sensing 2. X-ray tube configuration 3. Matrix effect attenuation 4. Soil testing.

4.1. Introduction

X-ray fluorescence (XRF) spectrometry is one of the promising proximal soil sensing (PSS) techniques, as it allows analyses with satisfactory performance using minimum sample preparation (Tavares et al., 2019). XRF sensors enable measuring the total content of specific elements in the soil (*e.g.*, Fe, Al, Si, Ca, and K) and can be used as a proxy for indirect predictions of other soil attributes (*e.g.*, pH, cation exchange capacity (CEC) and texture) (Weindorf and Chakraborty, 2016). Some researchers have applied XRF sensors to assess fertility attributes in soil samples and satisfactory results were reported for soil texture (Zhu et al., 2011), pH (Sharma et al., 2014), CEC (Sharma et al., 2015), base saturation (V) (Rawal et al., 2019), total content of nutrients (Nawar et al., 2019), and extractable nutrients (Towett et al., 2015). In Brazilian tropical soils, recent research has shown promising performance for organic matter (OM) (Morona et al., 2017), extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg) (Silva et al., 2017; Andrade et al., 2020), chemical attributes (*e.g.*, CEC, V, and pH) (Silva et al., 2017; Teixeira et al., 2018; Santos et al., 2020), and texture (Silva et al., 2016). Studies applying XRF in agricultural soils for predicting key fertility attributes are still preliminary and further works are needed for a comprehensive understanding of its applications particularly in tropical soils (Andrade et al., 2020; Silva et al., 2019).

Most of the studies that applied XRF in the PSS context used pre-developed (pre-programmed) measurement packages (*e.g.*, Soil Mode, Innov-X Systems, Inc., MA, USA), which have specific routines for spectra acquisition and processing (Zhu et al., 2011; Weindorf et al., 2012; Sharma et al., 2014, 2015; O'Rourke et al., 2016; Silva et al., 2017; Teixeira et al., 2018; Silva et al., 2019; Andrade et al., 2020). These factory calibrations involve the standardization of some parameters, such as (i) X-ray tube configurations (*e.g.*, voltage and current), (ii) number consecutive irradiations, (iii) scanning time (dwell time), (iv) filters and (v) types of spectral pre-processing (*e.g.*, Compton normalization, etc). Although these packages allow for a user-friendly approach that contributes to the popularization and expansion of the XRF application, they are not fully transparent and do not allow for flexibility regarding the time of analysis. Faster XRF analyses are fundamental to increase the agility of benchtop analysis, as well as creating new perspective for using this sensor embedded on mobile sensor platforms for on-line measurements. Therefore, the development of clear and objective methods of operating XRF should be prioritized (Bowers et al., 2019), since they allow for an unbiased scientific evaluation and also the optimization of data acquisition and processing routines necessary to reduce the time of analysis with XRF sensors.

The X-ray tube voltage and current are important configurations of the energy-dispersive X-ray fluorescence (ED-XRF) equipment. ED-XRF equipment uses a polychromatic beam (also called continuous radiation or *bremsstrahlung* radiation) to excite the sample and, by changing the voltage of the tube, the energy intensity and the amplitude of this continuous spectrum are changed (Jenkins,

2012). In turn, as the current is increased, the X-ray intensity (number of photons emitted from the tube filament, *e.g.*, the photon flux) of the continuous spectrum also increases (Lindgren, 2000). In other words, the voltage impacts the energy range and intensity of the continuous spectrum used for excitation while current impacts mostly its intensity. Hence, the X-ray tube configuration alters the fluorescence yield of a given analyte (Potts et al., 1984) and can, therefore, impact the performance of prediction models. Usually, lower voltages of the X-ray tube (≤15 keV) increase the fluorescence yield of light elements (*e.g.*, Si, Al, K, Ca, etc.), which can be a promising setup in XRF analyses of agricultural soils. On the other hand, a spectrum acquired at such low voltage does not contain the well-defined Kα scattering peaks, which are usually present when using higher energies. For example, in equipment with Rh-anode tube, the elastic (Thomson) and inelastic (Compton) scattering peaks appear between 18 and 22 keV. However, these scattered X-rays can also be important because they can be explored as a source of soil information (Panchuk et al., 2018).

Another challenge of applying XRF for soil analysis is the matrix effect. Due to the heterogeneous and complex nature, XRF analysis of soils is prone to matrix effects (Bowers, 2019), which refer to the influences of the different soil constituents in promoting absorptions or enhancements of emission lines of a given analyte under consideration (USEPA, 2007). Pre-developed measurement packages generally use strategies to mitigate the matrix effect (Bowers, 2019). A simple and widespread method for matrix effect correction is the Compton normalization (CN), which consists of using the Compton peak to normalize the analyte emission line (Yılmaz and Boydas, 2018). As the Compton peak intensity is inversely proportional to the average atomic mass of a sample, it may play an important role in discriminating between different sample sets each having similar matrices. Another alternative solution for mitigating the matrix effect is the use of multivariate statistics approaches (*e.g.*, multiple linear regression (MLR) or partial least squares regression (PLSR)) with emission lines intensities and scattering peaks used as input X-variables (Panchuk et al., 2018). However, models should be established with soil samples having similar matrices to ensure the best prediction performance. Combining CN with multivariate statistics could be the ideal transparent solution to overcome the matrix effects and achieve the best prediction accuracy.

Since most of the works that have evaluated XRF sensors in the PSS context used pre-developed measurement packages, the influence of X-ray tube configuration, as well as alternatives for the matrix effect migigation on fertility attribute predictions has not yet been addressed in the literature. Therefore, this work has three objectives: (i) evaluate the performance of a portable XRF configured with different X-ray tube voltages (15 and 35kV) to predict fertility attributes in agricultural fields; (ii) evaluate the combined effects of CN and multivariate regressions (MLR and PLSR) on the attenuation of the soil matrix effects (using data of two fields with considerable matrix effect) in order to improve

the prediction accuracy of key soil fertility attributes; and (iii) compare the prediction performance achieved with the proposed solutions to that obtained from the Geo Exploration package (Bruker AXS, Madison, WI, EUA), pre-developed measurement package commonly used by the scientific community.

4.2. Material and methods

The data analysis of this Chapter, as well as the presentation of its results and discussion, were performed in two parts. Data analysis part 1 addressed the first objective and data analysis part 2 addressed the second and third objectives. More specifically, in the part 1, the effect of the different X-ray tube configurations for predicting clay, OM, CEC, pH, V, ex-P, ex-K, ex-Ca, and ex-Mg was evaluated. Afterward, in the part 2, it was assessed the effect of Compton normalization and multivariate regressions for the prediction of clay, CEC, ex-K, and ex-Ca. In addition, also in the second part, the obtained predictions were also compared with predictions using the Geo Exploration package (Bruker AXS, Madison, WI, EUA).

4.2.1. Study sites and soil samples

The samples used in this study belong to the soil sample bank of the Precision Agriculture Laboratory (LAP) from Luiz de Queiroz College of Agriculture, University of São Paulo. They were collected from 0–20 cm depth and stored after being air-dried, ground and sieved (≤2 mm). The samples used in this work are from two different fields (Figure 1) that have been under active agricultural production over the years. Field 1 is located in the southeast region of Brazil, in the municipality of Piracicaba, State of São Paulo. Its soil is classified as Lixisol (IUSS, 2014) with a clayey texture and high nutrient variability. Field 2 is situated in Brazil's midwest region, in the municipality of Campo Novo do Parecis, State of Mato Grosso. Its soil is classified as Ferralsol (IUSS, 2014), with a texture varying between sandy loam and sandy clay loam. A set of 102 soil samples were selected for this work—58 soil samples from Field 1 and 44 from Field 2. The chemical analysis results of the LAP's soil sample bank were used to choose samples with wide ranges of variability of key fertility attributes in both study fields. After dataset selection, the samples were again subjected to laboratory chemical analyses, as described in Section 4.2.2, which provided the results of the reference analyses used in this work.

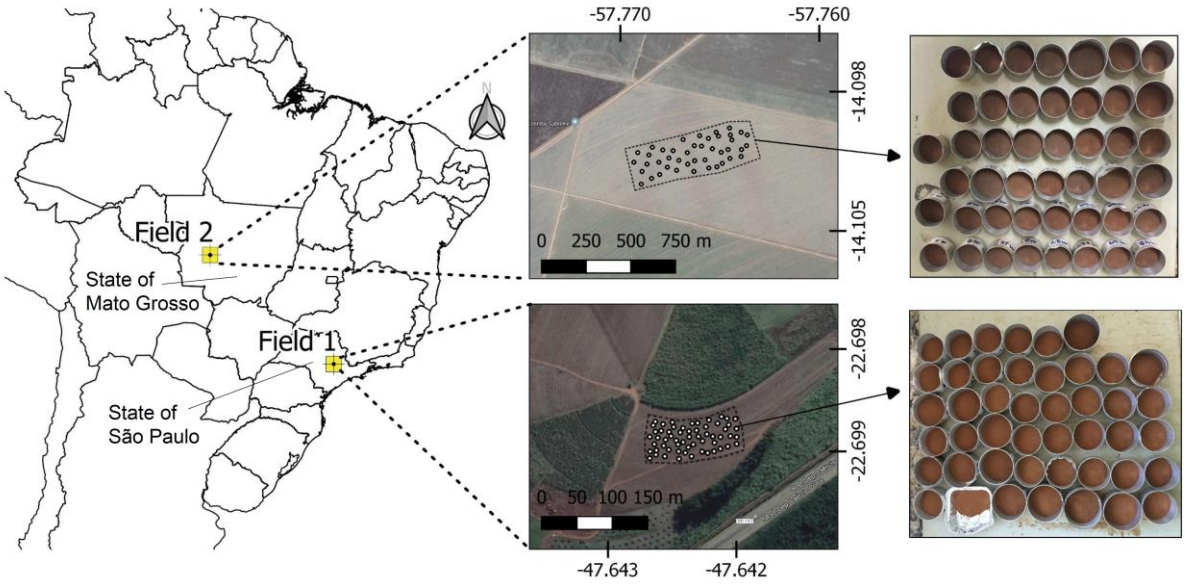


Figure 1. Location of Field 1 and Field 2 in Brazil, where the soil samples were collected.

4.2.2. Reference analyses

From each sample, an aliquot of 90 g was sent to the laboratory for regular soil fertility analysis. These analyses determined clay, OM, CEC, pH, V, ex-P, ex-K, ex-Ca, and ex-Mg, following the methods described by Van Raij et al. (2001). Clay content was quantified by the Bouyoucos hydrometer method in a dispersing solution. OM concentration was determined via oxidation with potassium dichromate solution. pH was determined via calcium chloride solution. Extractable nutrients were determined via ion exchange resin extraction. The soil potential acidity (H + Al) was quantified via pH in buffer solution method (SMP) and used to calculate the CEC, which corresponds to the sum of soil potential acidity and sum of bases (ex-Ca + ex-Mg + ex-K). Percent base saturation (V) was calculated by the ratio between the sum of bases and CEC.

The pseudo total content (ptc) of P, K, Ca, and Mg was also analysed following the United States Environmental Protection Agency (USEPA) Method 3051A (Element, 2007). This method comprises the chemical dissolution of pulverized soil samples using HNO_3 and HCl, with microwave-assisted digestion. Then, the multi-element quantification is performed using inductively coupled plasma optical emission spectrometry (ICP - OES). For the determination of elements in tropical soils, this method presents proportional recoveries compared to more aggressive methods (*e.g.*, EPA 3052, which uses HF) (Silva et al., 2014), thus allowing us to understand the relationship between the total and extractable contents of the elements of interest. Linear correlations were performed between the extractable and pseudo total content of P, K, Ca, and Mg. A correlation matrix was calculated among fertility attributes to better understand their interrelations.

4.2.3. Data analysis part 1: effect of X-ray tube configurations

4.2.3.1. XRF measurements with different X-ray tube configurations

Soil samples were scanned after being air-dried and sieved at 2 mm, as suggested by Tavares et al. (2019). Ten grams of each sample was placed in an XRF polyethylene cup of 31 mm diameter (n. 1530, Chemplex Industries Inc., USA) sealed at the bottom with a 4-μm thick polypropylene film (n. 3520, SPEX, USA). Each sample was scanned in triplicate and the sample cup was repositioned between scans. The replicates were subsequently averaged for analysis.

XRF spectra were acquired using a Tracer III-SD model (Bruker AXS, Madison, EUA). This hand-held instrument is equipped with a 4 W Rh X-ray tube and an X-Flash® Peltier-cooled Silicon Drift Detector (Bruker AXS, Madison, USA). This detector has 2048 channels and an active area of 10 mm^2. All samples were scanned using the XRF sensor, configured with two tube voltage conditions: 15 and 35 kV. For each voltage, the current was set so as not to exceed 15% of the detector deadtime, avoiding spectral distortions and artefacts; thus, the tube current was set at 23 and 7 μA for the voltages of 15 and 35kV, respectively. These conditions correspond to a tube working power of 0.245 and 0.345 W for higher and lower voltage, respectively. The two different XRF tube configurations provided the two XRF data scenarios evaluated in this study.

For both voltage conditions, the fluorescence emission was measured during 90 s at atmospheric pressure. No filter was used in order to enhance the signal of light elements. A cellulose pellet was used as a blank sample, being scanned at the beginning of the measurements, after 50 samples, and at the end, to ensure the contaminant-free operation of the equipment. The spectra of the blank samples are shown in Figure A1, in the Appendix Section.

The spectra were acquired using the Bruker S1PXRF® software (Bruker AXS, Madison, EUA). The Bayesian deconvolution process was applied using Artax® (Bruker AXS, Madison, EUA) to correct peaks overlaps (inter-elemental effects). The spectra were normalized by the detector live time; then, they were presented in counts of photons per second (cps). The means of XRF spectra of Field 1 and Field 2, acquired using both tube voltage conditions, are shown in Figure 2.

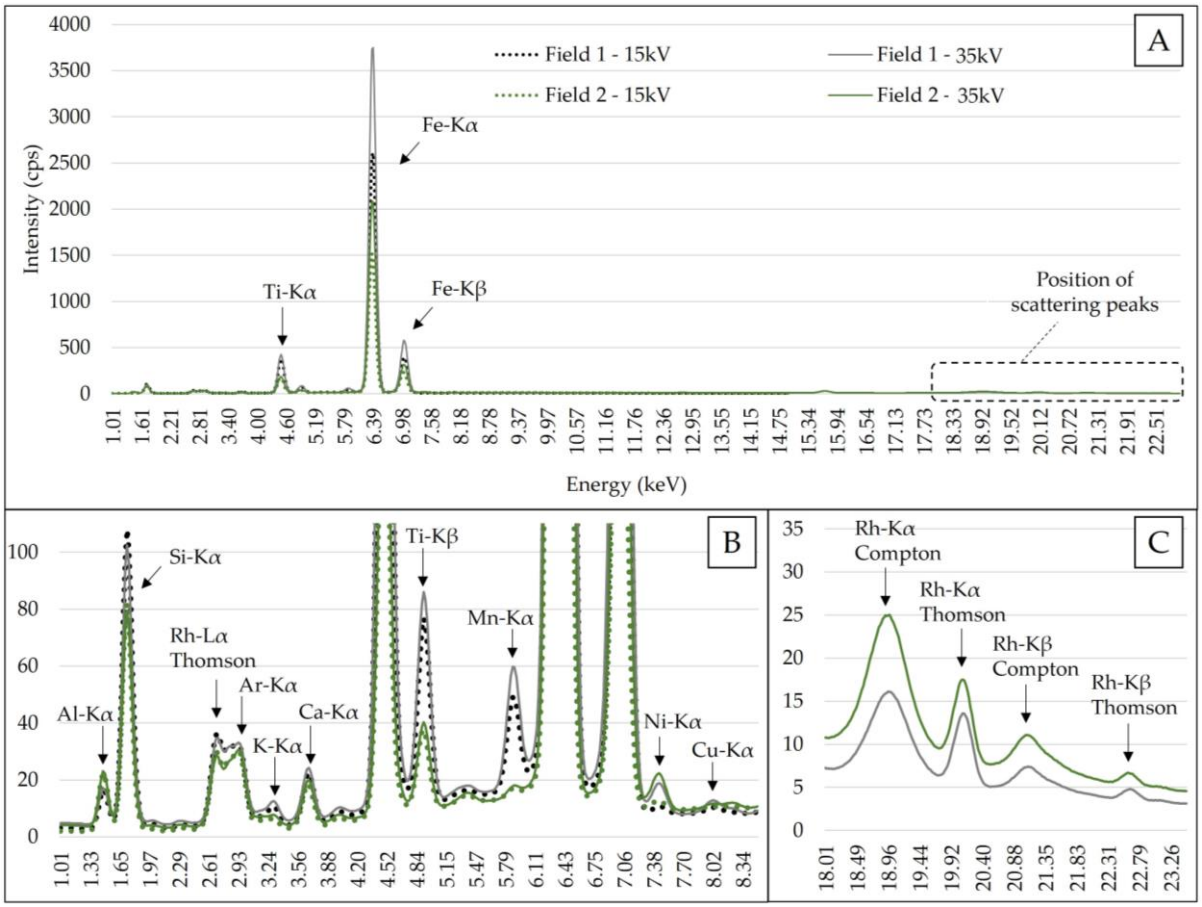


Figure 2. Mean of X-ray fluorescence (XRF) spectra of Field 1 and Field 2 obtained with an X-ray tube set at 15 and 35kV (A). Snapshot of the emission lines from 1.01 to 8.34 keV, which presented lower fluorescence emission in relation to the Kα emission lines of iron (Fe) and titanium (Ti) (B). Snapshot of scattering peaks (C). Counts of photons per second was abbreviated as cps.

The mean spectra were evaluated in order to select emission lines to be used as independent variables. This selection was based on the following criteria: (i) the element should be commonly found in agricultural soils; (ii) the signal-to-noise ratio (SNR) should be higher than 10 (Currie, 1968); and (iii) for elements with K and L emission lines, just K-lines was chosen due to their greater fluorescence intensity. Following those criteria, nine fluorescence lines (Kα emission lines of Al, Si, K, Ca, Ti, Mn, Fe, Ni, and Cu) and one scattering peak (Rh-Lα Thomson) were selected in the 15 keV spectra (totalling 10 independent variables). For the 35 keV spectra, the same 10 lines were selected, including the two higher scattering peaks present in the region between 18 and 23 keV (totalling 12 independent variables): Rh-Kα Compton and Rh-Kα Thomson. The 10 lines present in both spectra (Al-Kα, Si-Kα, K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, Cu-Kα, and Rh-Lα) were compared through box-plot graphs.

Net intensities of the emission lines selected in the two different XRF data scenarios were used to build the predictive models. The XRF data scenarios using the 15 and 35 keV spectra were designated in this work as EL-15 and EL-35, respectively.

4.2.3.2. Data modeling of part 1

Each one of the two XRF data scenarios (EL-15 and EL-35) was pulled in one matrix with the key fertility attributes (clay, OM, CEC, pH, V, ex-P, ex-K, ex-Ca, and ex-Mg), measured via the above detailed commercial laboratory procedures. Calibration models were built using MLR, as suggested by Weindorf and Chakraborty (2016), using the Unscrambler® version 10.5.1 (Camo AS, Oslo, Norway). Calibration models were developed for each of the studied soil fertility attributes. The two-field dataset (n = 102) was divided into two subsets by means of the Kennard Stone algorithm (Kennard and Stone, 1969), to 70% of data used for calibration and the remaining 30% used for validation. Descriptive statistical analyses were applied to characterize the variability of fertility attributes of both calibration and validation datasets. The methodology applied in this work was diagrammatically presented in Figure 3. The importance of XRF variables for the prediction of each fertility attribute was evaluated by the standardized regression coefficients of the regression analysis.

The quality of the developed calibration was assessed by means of the coefficient of determination (R^2), the residual prediction deviation (RPD), and the root-mean-square error (RMSE). Furthermore, we also used the normalized RMSE calculated by dividing the RMSE values by the range of each corresponding soil attribute. The RMSE was expressed in the units of the attribute of interest and in percentage, in the case on the normalized RMSE. The RPD was calculated as the ratio between the standard deviation of the laboratory measured soil property of interest and the RMSE in the prediction. The RMSE (%) and the RPD were used to compare the prediction performance of attributes that had different scales and units of measurement. Four RPD classes adapted from Chang et al. (2001) were used to evaluate the quality of the model prediction performance as poor models ($RPD < 1.40$), reasonable models ($1.40 \leq RPD < 2.00$), good models ($2.00 \leq RPD < 3.00$), and excellent models ($RPD \geq 3.00$).

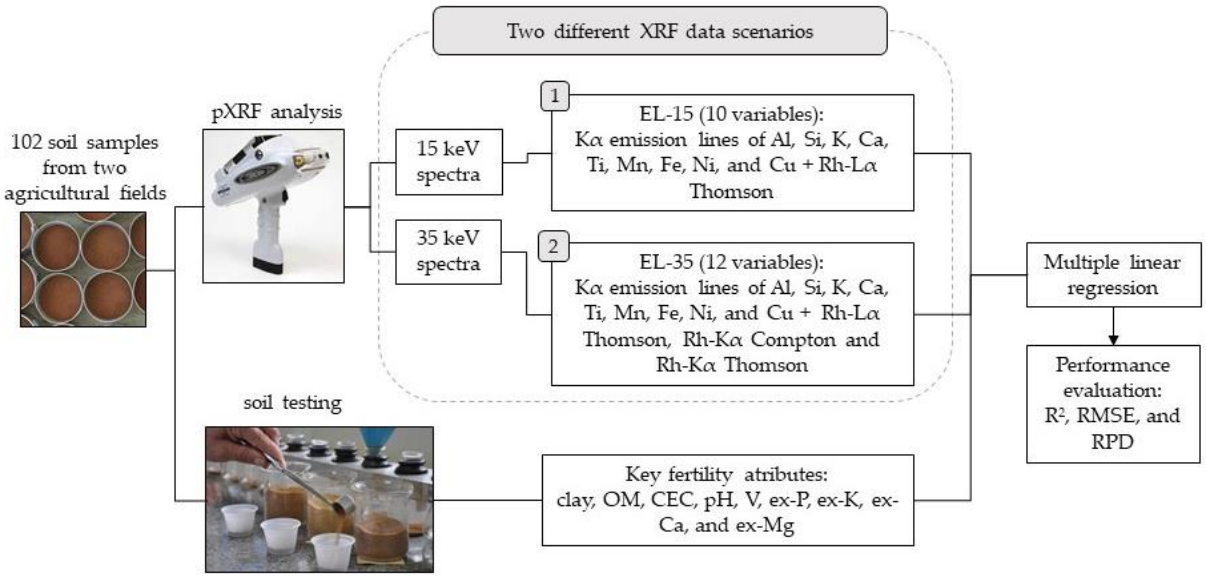


Figure 3. Framework of the methodology applied in part 1 for the development of calibration models for clay, organic matter (OM), cation exchange capacity (CEC), pH, base saturation (V), and extractable (ex-) P, K, Ca, and Mg using an X-ray fluorescence (XRF) sensor configured with two different X-ray tube voltages (15 and 35 kV). For the X-ray tube set at 15 kV, 10 XRF lines were selected, comprising the Kα emission line of aluminium (Al), silicon (Si), potassium (K), calcium (Ca), titanium (Ti), manganese (Mn), iron (Fe), nickel (Ni), copper (Cu), and the scattering peak Rh-Lα Thomson; this XRF scenario was called EL-15. For the X-ray tube set at 35 kV, in addition to these 10 XRF lines, Rh-Kα Compton and Thomson scattering peaks were included, this second scenario was called EL-35.

4.2.4. Data analysis part 2: matrix effect mitigation

For the development of data analysis part 2, the emission lines of the above-mentioned EL-35 scenario, which presents the scattering peaks Rh-Kα Compton, were used. Furthermore, all samples were also scanned using the Geo Exploration package (Bruker AXS, Madison, WI, EUA). This is an empirical package calibrated with standard reference materials. The measurements were performed in 3-phase oxide calibration which means that three different instrumental conditions are applied, *e.g.*, with different combinations of X-ray tube voltage and current and X-ray filters. This package also uses a standardization tool to modify the output of a new calibration in order to match with its type of sample matrix (Mantler et al., 2006). The block diagram of Figure 4 shows the methodology steps followed in the part 2.

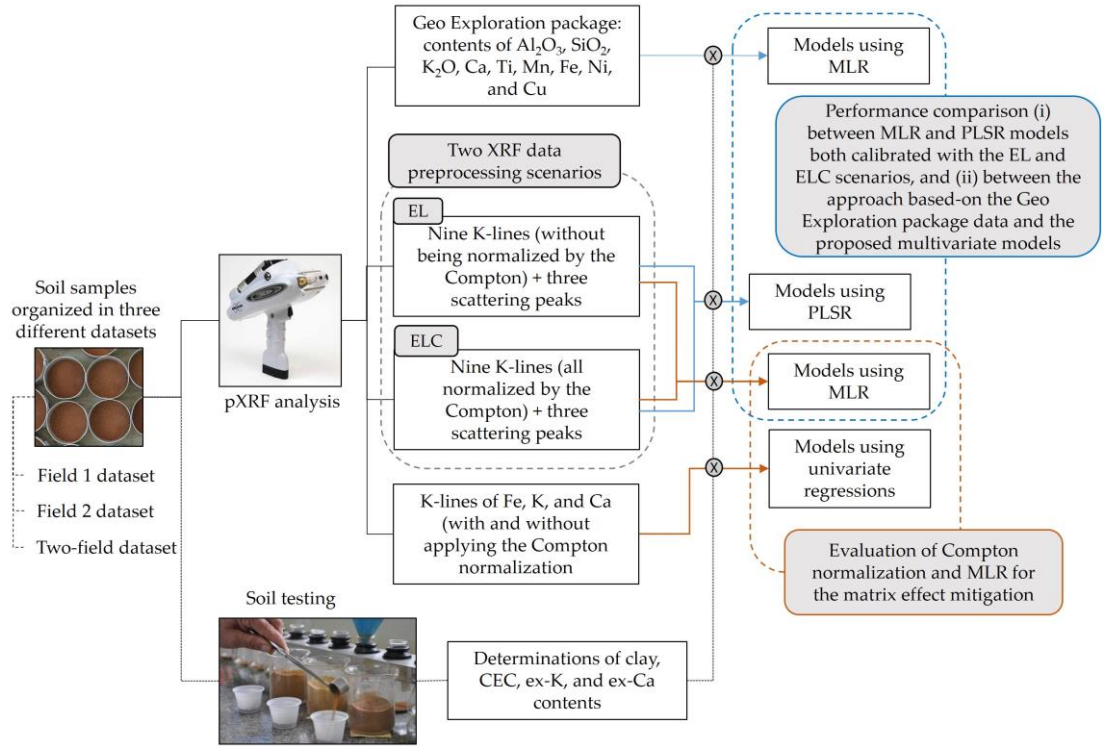


Figure 4. Framework of the methodology applied in part 2 for the development of calibration models for clay, cation exchange capacity (CEC), and exchangeable (ex-) potassium (K) calcium (Ca), and magnesium (Mg), using X-ray fluorescence device (XRF) and methods to attenuate the matrix effect. Two scenarios of XRF data pre-processing (emission lines with (ELC) and without (EL) Compton normalization), both using the 12 emission lines selected (Al-Kα, Si-Kα, K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, Cu-Kα, Rh-Lα Thomson, Rh-Kα Thomson, and Rh-Kα Compton), were subjected to multiple linear regression (MLR) and partial least squares regression (PLSR) analyses. For the ELC scenario, the fluorescence emission lines of Al-Kα, Si-Kα, K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, and Cu-Kα were normalized by the Compton peak, which was not applied for the EL scenario. Scattering peaks were not normalized in any of the scenarios. Total contents obtained with the Geo Exploration package (Bruker AXS, Madison, WI, USA) were also used to calibrate models for the studied soil attributes using MLR, having their performance compared with the other calibrated models.

4.2.4.1. XRF data pre-processing and related analyses applied in part 2

Two different XRF data pre-processing methods were applied to the emission lines of the EL-35 scenario. In the first method, designated as EL, the net intensities of the 12 emission lines were used as X-variables, whereas in the second method, designated as ELC, the normalized intensities of the same 12 emission lines with Compton peak were used as predictors. Scattering peaks were not normalized in any of these two pre-processing scenarios. Both pre-processing scenarios were subjected to multiple linear regressions (MLR and PLSR), in order to assess the effectiveness of this proposed combined solution to remove the matrix effect that exists in the current dataset and evaluate the improvement in the prediction accuracy of the studied soil fertility attributes.

Scatter plots and regression analysis of measured ptc of K, Ca, Ti, Mn, and Fe versus their respective Kα emission lines were presented. These five elements (namely, K, Ca, Ti, Mn, and Fe) were

analyzed since their ptc obtained via the USEPA method 3051A represent proportional recoveries to their total contents in tropical soils (Nogueirol et al., 2013; Silva et al., 2014). Boxplots graphs with the intensity of XRF lines (Al-Kα, Si-Kα, K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, and Cu-Kα) with and without CN were also shown. They were compared after being normalized by their range (*e.g.*, the difference between the maximum and minimum values), due to the different scales of both EL and ELC datasets.

4.2.4.2. Data modeling of part 2

To evaluate the matrix effect, calibration models for the merge two-field dataset were compared with the corresponding individual field dataset models (Field 1 and Field 2 datasets). Each of the three datasets was divided into two subsets of 70% for calibration and 30% for validation using the Kennard Stone algorithm (Kennard and Stone, 1969). MLR was used for the development of calibration models for both EL and ELC scenarios of the three above-mentioned datasets. Univariate regressions using emission lines with and without the CN were also carried out and the results were compared with those of the MLR. Univariate calibrations were established using the emission line that showed the greatest importance for the predictions built via MLR (presented in Table A1, in the Appendix Section). For example, for clay the Fe-Kα emission line was used; for ex-K, the K-Kα emission line; and for both ex-Ca and CEC, the Ca-Kα emission line.

For both EL and ELC scenarios, PLSR calibration models were also developed for each of the studied soil fertility attributes using the two-field merged dataset. The same calibration and validation sets mentioned above were used, and the number of latent variables (nLV) was selected based on the smallest root-mean-square error of cross-validation (RMSECV) applied by the leave-one-out technique.

The prediction performance of PLSR and MLR models calibrated with both XRF pre-processing scenarios (EL and ELC) were compared with models calibrated using the Geo Exploration package. The latter was calibrated for all the studied soil fertility attributes using MLR and for the contents of Al_2O_3, SiO_2, K_2O, Ca, Ti, Mn, Fe, Ni, and Cu as input X-variables, as suggested by Weindorf and Chakraborty (2016). The same above-mentioned calibration and validation datasets were used. All calibrations were performed using Unscrambler® version 10.5.1 (Camo AS, Oslo, Norway). The performance evaluation of the predictive models was carried out by means of R^2, RMSE, and RPD, similar to that performed in part 1.

4.3. Results

4.3.1. Laboratory measured soil properties

4.3.1.1. Two-field dataset (data analysis part 1)

The descriptive statistics of the fertility attributes for the calibration and validation datasets used for the data analysis part 1 are shown in Table 1. The range and SD of soil attributes in the calibration set are close to those in the validation set, which is expected since we applied Kennard Stone method in sample split to avoid undesirable influences on the prediction accuracy that are not sensor-related (Stenberg et al., 2010). In this study, the range and SD for both calibration and validation datasets are comparable. This was only not the case for ex-P, whose maximum value in the validation set (59.00 mg dm^{-3}) is explicitly lower than that in the calibration set (104.00 mg dm^{-3}), with smaller SD (15.98 mg dm^{-3}) for the calibration set compared to that of the validation set (10.65 mg dm^{-3}).

In general, the soil samples were characterized by a high variability of all fertility attributes, with a coefficient of variation (CV) larger than 23%, with the only exception being for pH (CV < 8%). According to the local fertility interpretation (Van Raij, 2011), the contents of ex-P, ex-K, ex-Ca, and ex-Mg range from low to very high. We also noticed that by joining data from both fields, the resulting dataset has a continuum range of values for all fertility attributes (Figure A2, in the Appendix Section), which allows its modeling as a single dataset.

A significant correlation exists between the extractable and pseudo total contents of K, Ca, and Mg ($0.75 \leq r \leq 0.90$), whereas weak correlations were observed for P ($r = 0.11$). These significant correlations suggest that the exchangeable fractions of K, Ca and Mg can be assessed by XRF using the total content values, once their emission line is detected.

Table 1. Descriptive statistics of fertility attributes for the calibration and validation dataset studied in the data analysis part 1.

	Clay	**OM[3]**	**CEC[4]**	**pH**	**V[5]**	**ex-P[6]**	**ex-K[6]**	**ex-Ca[6]**	**ex-Mg[6]**
	----- g dm^{-3} -----		$mmol_c$ dm^{-3}		%	mg dm^{-3}	------- $mmol_c$ dm^{-3} -------		
---------- Calibration set (n = 68) ----------									
Min	175.00	14.00	37.50	4.60	19.00	4.00	0.90	8.00	3.00
1st Qu.	251.00	19.00	60.03	5.00	43.50	12.00	1.00	16.00	7.00
Mean	352.00	24.78	81.75	5.38	64.99	22.65	3.41	35.69	18.06
3rd Qu.	431.00	30.75	98.58	5.78	84.00	31.5.	5.68	49.75	28.00
Max	511.00	37.00	148.90	6.30	92.00	104.00	10.30	78.00	54.00
SD[1]	95.21	6.13	25.86	0.47	21.96	15.98	2.48	19.08	12.58
CV[2] (%)	27.05	24.72	31.63	8.67	33.80	70.55	72.73	53.44	69.65
Skewness	−0.22	0.14	0.46	0.50	−0.57	2.26	0.59	0.25	0.81
Kurtosis	−1.22	−1.10	−0.41	−1.02	−1.16	8.75	−0.79	−1.01	−0.14
---------- Validation set (n = 34) ----------									
Min	175.00	18.00	42.50	4.70	28.00	7.00	0.90	8.00	3.00
1st Qu.	226.00	22.00	53.08	4.90	39.00	12.00	1.18	13.00	5.00
Mean	332.12	26.62	76.50	5.21	64.15	19.09	3.36	33.32	17.50
3rd Qu.	406.25	32.00	98.90	5.50	86.25	21.00	5.50	51.25	30.00
Max	463.00	35.00	138.40	6.20	91.00	59.00	7.90	75.00	47.00
SD[1]	92.03	5.48	26.14	0.37	23.42	10.65	2.26	19.71	12.81
CV[2] (%)	27.71	20.58	34.17	7.07	36.51	55.79	67.39	59.16	73.22
Skewness	−0.45	−0.11	0.53	0.83	−0.35	2.16	0.35	0.34	0.63
Kurtosis	−1.39	−1.45	−0.63	0.26	−1.62	5.78	−1.35	−1.11	−0.73

[1] Standard deviation, [2] coefficient of variation, [3] organic matter, [4] cation exchange capacity, [5] base saturation, and [6] extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg).

Interpreting the interrelationships between different fertility attributes (shown in Table 2) aid to understand why indirect (having no emission lines) predictions are still possible with XRF sensor data. Strong relationships exist between ex-Ca and ex-Mg ($r = 0.93$), as well as between each of them with V and CEC ($r \geq 0.84$). Positive correlations of CEC, V, ex-K, ex-Ca, and ex-Mg with clay range from 0.64 to 0.81. It should also be highlighted that ex-P and pH showed weak to reasonable correlations with all the other attributes, with r ranging from −0.22 to 0.10 for ex-P, and from −0.41 to 0.53 for pH. Finally, OM content has reasonable correlations with almost all attributes (except ex-P and pH), with r ranging from −0.65 to 0.62.

Table 2. Correlation matrix of the soil fertility attributes studied in the data analysis part 1. Correlations between extractable and pseudo total content (ptc) for the different macronutrients are also shown.

	Clay	**OM[1]**	**CEC[2]**	**pH**	**V[3]**	**ex-P[4]**	**ex-K[4]**	**ex-Ca[4]**	**ex-Mg[4]**
Clay	1.00	0.59**	0.67**	0.38**	0.81**	−0.06	0.73**	0.78**	0.64**
OM		1.00	0.43**	−0.01	0.62**	0.07	0.48**	0.55**	0.44**
CEC			1.00	0.42**	0.76**	0.02	0.64**	0.93**	0.93**
pH				1.00	0.53**	−0.22*	0.40**	0.46**	0.44**
V					1.00	−0.12	0.71**	0.92**	0.84**
ex-P						1.00	−0.1	0.01	−0.17
ex-K							1.00	0.68**	0.58**
ex-Ca								1.00	0.93**
ex-Mg									1.00
Correlation with ptc						0.11	0.90**	0.75**	0.83**

[1] Organic matter, [2] cation exchange capacity, [3] base saturation, and [4] extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg). * Significant correlation at the probability level of 0.05; ** Significant correlation at the probability level of 0.01; the correlations were presented on grayscale highlighting the highest values, which were evaluated ignoring the sign (module).

4.3.1.2. Field 1 and Field 2 datasets (data analysis part 2)

Table 3 shows a summary of the descriptive statistics of fertility attributes for the calibration and validation datasets used for data analysis part 2, all having comparable ranges and SDs. This was only not the case for the clay content of Field 1, whose clay's SD in the validation set (24.52 g dm^{-3}) is clearly lower than that in the calibration set (44.10 g dm^{-3}). It is also noteworthy that the ex-K concentration in Field 2 is of a considerably lower range and SD than the values in Field 1, *e.g.*, the ex-K's SD for the Field 2 (0.43 $mmol_c$ dm^{-3}) was about 4.5 times lower than that for Field 1 (1.94 $mmol_c$ dm^{-3}). According to the Brazilian soil fertility classes proposed by Van Raij (2011), ex-K in Field 1 ranges from medium to very high content, whereas it ranges from low to medium in Field 2.

Pearson's correlations between fertility attributes, as well as between ptc and extractable contents of K and Ca are shown in Table 4. For all datasets, a significant correlation exists between the ptc and exchangeable contents of K and Ca ($0.49 \leq r \leq 0.95$). These significant correlations suggest that all extractable nutrients evaluated can be inferred by their total content values assessed by XRF. The lowest correlation was observed in Field 2 between ptc and extractable content of K ($r = 0.49$), probably because of the low range and SD of K in this field. The highest correlation observed for all the three datasets evaluated was between ex-Ca and CEC, with r values of 0.93, 0.70, and 0.93 for the Field 1, Field 2, and two-field datasets, respectively.

Table 3. Descriptive statistics of the studied soil fertility attributes according to their calibration and validation sets used for data analysis part 2.

	Clay [1]		CEC [1]		ex-K [1]		ex-Ca [1]	
	Cal set	Val set	Cal set	Val set	Cal set	Val set	Cal set	Val set
				Field 1 [2]				
Min	345.00	378.00	58.90	62.50	1.70	2.30	27.00	28.00
Mean	424.03	405.47	97.52	93.23	5.11	5.07	49.21	47.53
Max	511.00	463.00	148.90	138.40	10.30	7.90	78.00	75.00
SD	44.10	24.52	20.93	21.19	1.94	1.50	12.53	13.52
CV (%)	10.40	6.05	21.46	22.73	37.92	29.48	25.47	28.44
				Field 2 [2]				
Min	175.00	175.00	37.50	42.50	0.90	0.90	8.00	8.00
Mean	255.14	239.20	60.54	55.31	1.12	1.19	17.52	15.33
Max	328.00	328.00	88.90	88.90	2.90	2.30	43.00	38.00
SD	45.90	51.34	13.89	13.13	0.43	0.39	7.81	7.78
CV (%)	17.99	21.46	22.94	23.74	38.38	33.23	44.58	50.74
				Two-field [2]				
Min	175.00	175.00	37.50	42.50	0.90	0.90	8.00	8.00
Mean	352.00	332.12	81.75	76.50	3.41	3.36	35.69	33.32
Max	511.00	463.00	148.90	138.40	10.30	7.90	78.00	75.00
SD	95.21	92.03	25.86	26.14	2.48	2.26	19.08	19.71
CV (%)	27.05	27.71	31.63	34.17	72.73	67.39	53.44	59.16

[1] The minimum (min), maximum (max), mean, and standard deviation (SD) values of clay content were given in g dm^{-3} and, in $mmol_c$ dm^{-3}, for cation exchange capacity (CEC) and the extractable (ex-) contents of K and Ca. [2] The number of samples (*n*) used for the calibration (cal) and validation (val) sets were, respectively, 39 and 19 for the Field 1 dataset, 29 and 15 for the Field 2 dataset, and 68 and 34 for the two-field dataset.

Table 4. Pearson's correlation between the studied soil fertility attributes for each dataset evaluated. Correlations between exchangeable (ex-) and pseudo total content (ptc) of K and Ca are also shown.

	Clay	CEC[1]	ex-K [2]	ex-Ca[2]	Clay	CEC	ex-K	ex-Ca	Clay	CEC	ex-K	ex-Ca
			Field 1				Field 2				Two-field dataset	
Clay	1.00	−0.02	−0.05	0.14	1.00	0.31 *	0.34 *	0.21	1.00	0.67 **	0.73 **	0.78 **
CEC		1.00	0.17	0.93 **		1.00	0.01	0.70 **		1.00	0.64**	0.93 **
ex-K			1.00	0.03			1.00	−0.01			1.00	0.68 **
ex-Ca				1.00				1.00				1.00
Correl [3] with ptc			0.67 **	0.83 **			0.49 **	0.95 **			0.90 **	0.75 **

* Significant correlation at the probability level of 0.05; ** Significant correlation at the probability level of 0.01; the correlations were presented on grayscale highlighting the highest values. [1] Cation exchange capacity. [2] extractable (ex-) K and Ca [3] Pearson's coefficient of correlation (r).

4.3.2. Data analysis part 1: effect of X-ray tube configurations

4.3.2.1. Effect of X-ray tube configuration on the XRF emission lines

The intensity of the emission lines present in both EL-15 and EL-35 scenarios [namely, Al-Kα, Si-Kα, Rh-Lα (Thomson), K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, and Cu-Kα] are plotted in box-plot graphs (Figure 5), allowing comparison between them, as a function of the X-ray tube configuration.

All the emission lines presented a high correlation coefficient ($r > 0.90$) between EL-15 and EL-35. The largest deviation was observed for Ni-Kα, with $r = 0.73$. This lower correlation was influenced by the low dispersion and intensity of the Ni-Kα XRF emission in EL-15.

The X-ray tube configuration affected the intensity of the emission lines of the different elements detected in both XRF scenarios. The Kα emission lines of the light elements Si and Al, with atomic number (Z) of 13 and 14, respectively, presented higher intensity when using the tube configured at 15 kV (EL-15) than using that configured at 35 kV (EL-35) (Figure 5a,b, respectively). The emission lines of K and Ca (with atomic numbers of 19 and 20, respectively) showed, in both scenarios (EL-15 and EL-35), a very similar intensity and dispersion of XRF emission (Figure 5d,e, respectively). However, Ti, Mn, Fe, Ni, and Cu (with atomic numbers 22, 25, 26, 28, and 29, respectively) showed higher XRF emission in the EL-35 scenario than that of the EL-15 (Figure 5f–j, respectively).

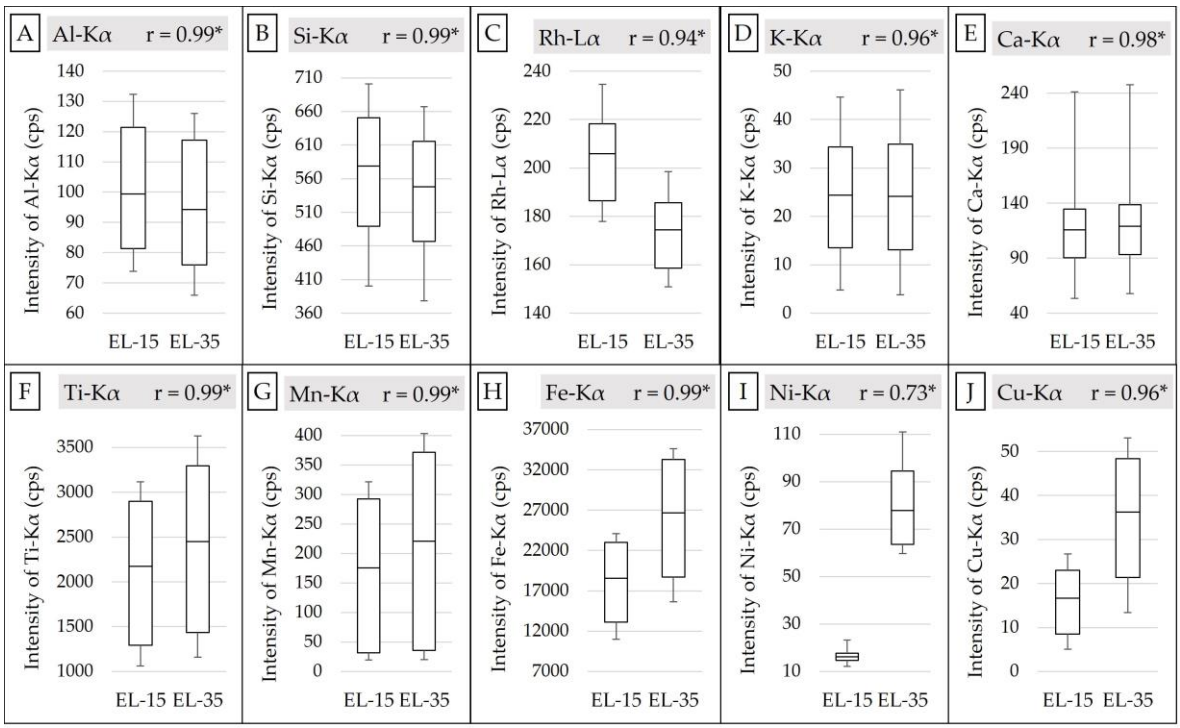


Figure 5. Box plot of the intensity of the lines Al-Kα, Si-Kα, Rh-Lα, K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, and Cu-Kα (**A**, **B**, **C**, **D**, **E**, **F**, **G**, **H**, **I,** and **J**, respectively) obtained using a portable X-ray fluorescence (XRF) sensor configured with two different X-ray tube voltages (15 and 35 kV). The XRF scenario using the X-ray tube set at 15 kV is named EL-15 and the XRF scenario using the X-ray tube set at 35 kV is named EL-35. The Pearson correlation between the EL-15 and EL-35 is also presented (correlations followed by * were significant at the probability level of 0.01).

4.3.2.2. Effect of X-ray tube configuration on XRF performance

The MLR prediction accuracy for the studied fertility attributes is similar between the EL-15 and EL-35 models (Table 5). The biggest differences were obtained for OM and pH, with R^2 values of 0.61 and 0.72, and 0.27 and 0.38, for EL-15 and EL-35, respectively. The prediction results for the remaining attributes are comparable with differences in R^2 values less than 0.06 between EL-15 and EL-35.

Table 5. Prediction results of the validation set (n = 34) obtained from multiple linear regressions (MLR) models calibrated using a portable X-ray fluorescence (XRF) sensor configured with two different X-ray tube voltages of 15 and 35 kV, designated as EL-15 and EL-35, respectively.

	Clay	**OM[1]**	**CEC[2]**	**pH**	**V[3]**	**ex-P[4]**	**ex-K[4]**	**ex-Ca[4]**	**ex-Mg[4]**
----------					R^2				
EL-15	0.88	0.61	0.80	0.27	0.95	0.00	0.90	0.91	0.85
EL-35	0.91	0.72	0.75	0.38	0.96	0.03	0.90	0.91	0.83
----------					RMSE				
EL-15	34.7	3.55	10.42	0.36	4.83	13.86	0.71	5.74	4.88
EL-35	31.72	2.87	11.08	0.32	4.82	13.41	0.65	5.63	4.97
----------					RMSE (%)				
EL-15	12.05	20.87	10.86	24.15	7.66	26.65	10.16	8.56	11.09
EL-35	11.01	16.87	11.55	21.17	7.65	25.79	9.31	8.40	11.29
----------					RPD				
EL-15	2.65	1.54	2.51	1.02	4.85	0.77	3.18	3.44	2.63
EL-35	2.90	1.91	2.36	1.16	4.86	0.79	3.47	3.50	2.58

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] exchangeable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg). The coefficient of determination (R^2) and residual prediction deviation (RPD) values are presented on grayscale, highlighting the highest values. The root-mean-square error (RMSE) was given in g dm^{-3} for clay and OM; in $mmol_c$ dm^{-3} for CEC, ex-K, ex-Ca, and ex-Mg; in % for V; and, for ex-P, the RMSE was given in mg dm^{-3}. The scatter plots of measured versus predicted fertility attributes are included as an annex (Figure A2).

The similarity between both XRF tube voltage data scenarios is also noted when analyzing the importance of the independent variables (X-variables) in both XRF scenarios, since the most important variables were always coincident in the prediction models calibrated with EL-15 and EL-35 (Table 6). More specifically, in both scenarios, Fe-Kα was the main variable for clay prediction, Mn-Kα for OM, K-Kα for ex-K, and Ca-Kα for CEC, V, ex-Ca, and ex-Mg. It is also observed that neither the Rh-Kα Compton and Thomson scattering peaks—present only in EL-35—nor the Ni-Kα were the main X-variable in any of the calibrated models (Table 6), which also contribute to the similarity between the predictive models calibrated with both X-ray tube configurations.

Table 6. Importance of X-ray fluorescence (XRF) variables for the prediction of studied soil fertility attributes, using the two XRF tube voltage data scenarios of EL-15 and EL-35, designated as EL-15 and EL-35, respectively. The values presented correspond to the t-value for each standardized coefficient obtained in the regressions calibration.

	Clay	OM[1]	CEC[2]	pH	V[3]	ex-P[4]	ex-K[4]	ex-Ca[4]	ex-Mg[4]
---------- EL-15 ----------									
Al-Kα	−1.35	1.48	0.11	−1.48	−0.21	−1.73	**−2.83**	0.21	1.29
Si-Kα	**−2.07**	−0.71	−0.37	−0.83	**2.04**	−0.94	**−2.40**	**2.26**	**2.07**
K-Kα	−1.13	−1.36	1.50	**2.20**	−0.25	0.07	**14.75**	0.86	0.27
Ca-Kα	−0.20	1.24	**5.03**	**2.90**	**7.49**	1.23	**−5.95**	**9.35**	**7.10**
Ti-Kα	**−2.31**	**−2.54**	**3.58**	**2.05**	−0.46	**−2.55**	0.01	1.55	**5.41**
Mn-Kα	−0.71	**3.54**	−1.79	−0.73	0.34	0.49	−0.83	**−2.16**	**−2.48**
Fe-Kα	**9.22**	−0.63	−0.57	−0.20	**2.49**	−0.40	1.98	**2.60**	−1.12
Ni-Kα	−0.09	−0.26	−0.05	0.03	0.59	**2.81**	0.37	0.41	−0.80
Cu-Kα	−1.77	0.44	0.26	−1.51	−1.84	**3.32**	−1.53	−0.09	−1.16
Rh-Lα Thomson	0.52	1.97	0.23	**−3.22**	1.97	0.20	**−4.06**	**2.28**	1.73
---------- EL-35 ----------									
Al-Kα	−1.14	0.93	−0.48	−0.24	−0.90	−0.55	−1.01	−1.11	−0.91
Si-Kα	**−2.08**	0.53	−0.73	−1.17	0.23	0.29	−0.81	0.52	−0.17
K-Kα	−1.20	−0.43	1.46	**2.04**	0.22	−0.65	**18.89**	0.92	0.41
Ca-Kα	−0.27	1.11	**4.45**	**3.33**	**7.89**	1.09	**−5.68**	**9.08**	**5.78**
Ti-Kα	**−2.99**	−1.91	**2.73**	1.30	−1.61	−1.61	−0.93	0.09	**3.54**
Mn-Kα	−0.40	**2.59**	−1.77	−0.14	0.74	0.64	−1.97	−1.92	−1.96
Fe-Kα	**5.09**	0.50	−0.27	−1.55	0.19	0.38	−1.42	0.64	−0.34
Ni-Kα	−1.12	−0.73	1.26	−0.07	0.09	1.44	−0.49	1.13	0.75
Cu-Kα	−1.30	0.42	0.90	−2.00	−1.85	**3.54**	−0.64	0.65	−0.94
Rh-Lα Thomson	−1.07	**2.32**	−0.88	**−2.40**	1.02	0.27	**−4.27**	0.65	−0.50
Rh-Kα Thomson	0.38	0.23	−0.25	0.90	1.78	−1.02	−1.02	1.27	0.22
Rh-Kα Compton	−0.61	1.07	−0.43	−1.83	−1.79	1.13	**−2.38**	−1.46	−0.63

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] exchangeable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg). The emboldened values indicate a significant correlation at the probability level of 0.05; significant values were presented on grayscale, with the most important variables having the darkest colour and vice versa

4.3.3. Data analysis part 2: matrix effect mitigation

4.3.3.1. Effect of Compton Normalization on the XRF data

The scatter plots of Figure 6 compare the ptc of the elements K, Ca, Ti, Mn, and Fe with their respective emission lines with and without being normalized by the Compton peak. Only K and Ca data showed continuous values when joining both fields (Figure 6A–D). The other elements (Ti, Mn, and Fe) presented a clear separation of data into contrasting groups, which highlights its different soil matrices.

For both K and Ca of the two-field dataset, CN reduced dispersion, and improved the correlations between ptc and the K-lines of K and Ca, as R^2 values increased from 0.84 to 0.89 for K, and from 0.69 to 0.92, for Ca (Figure 6A–D). The contrasting ranges leading to two clear groups of data for Ti, Mn, and Fe of the two fields resulted in an overestimation of R^2. However, for these elements, it can be observed that CN did not lead to clear increases in the R^2 values, as observed for Ca and K. The R^2

values with and without CN were similar for Fe (R^2 = 0.96, shown in Figure 6I and J), and slightly changed from 0.96 to 0.97 (Figure 6E,F) and from 0.98 to 0.97 (Figure 6G,H), for Ti and Mn, respectively.

Assessing the relationship between ptc and XRF emission lines for the Field 1 and Field 2 datasets separately, no clear performance improvement can be observed after the CN application on K-lines of K and Ca. Subtle performance reduction by CN for K and Ca is observed in Field 1, as R^2 values were reduced from 0.46 to 0.43, for K, and from 0.83 to 0.81, for Ca. For Field 2, the R^2 values for Ca-Kα slightly increased from 0.92 to 0.93. These results suggest a lower influence of CN on XRF data obtained from soil samples from the same field, confirming the lower influence of the matrix effect for individual field dataset, compared to the two-field dataset.

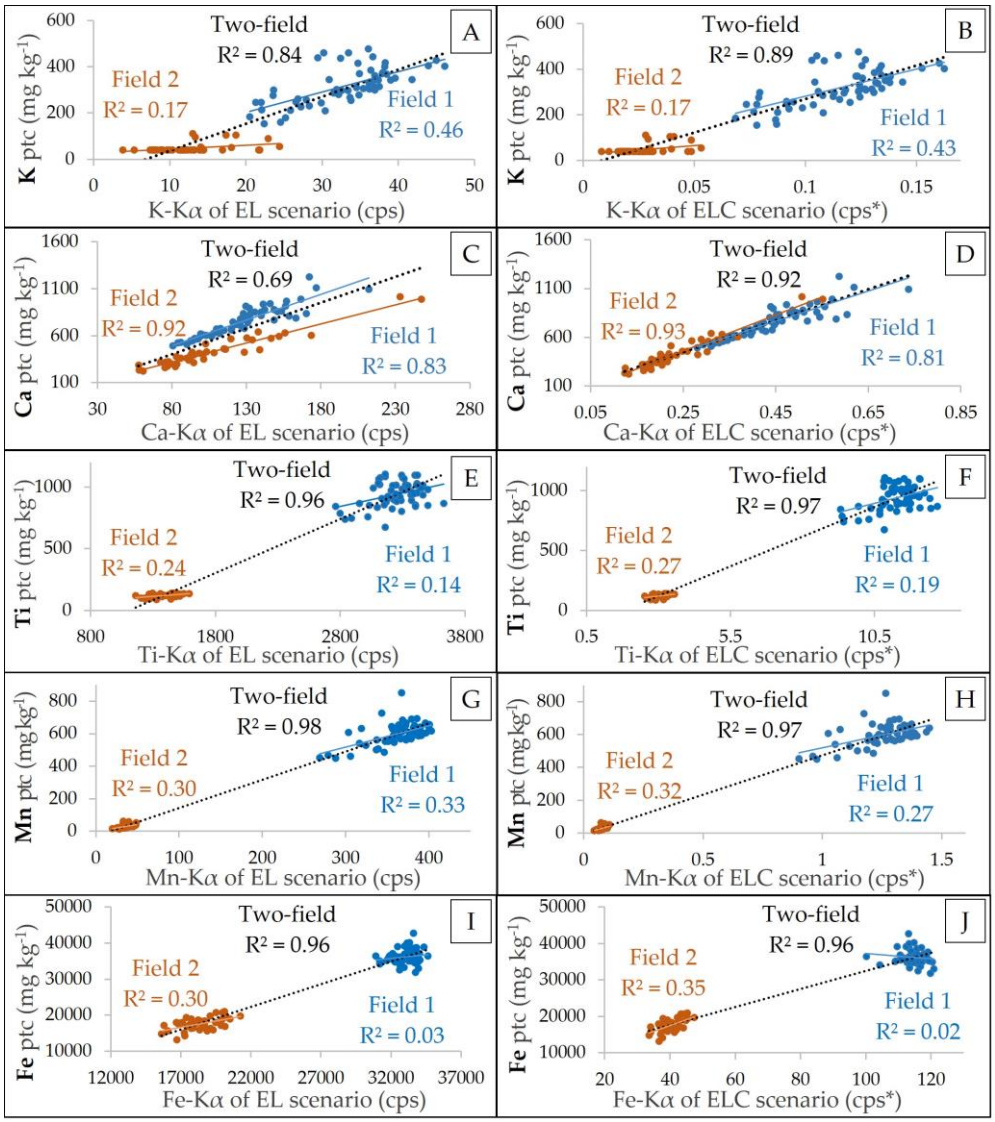


Figure 6. Scatter plots of measured pseudo total content (ptc) of potassium (K), calcium (Ca), titanium (Ti), manganese (Mn), and iron (Fe) versus their respective K-line intensity. The scatter plots were presented for the emission lines of both EL (on the left side, **A,C,E,G,I**) and ELC scenarios (on the right side, **B,D,F,H,J**). The intensity of the emission lines was presented in counts per second (cps), and the cps were normalized by Compton peak (cps*). The coefficient of determination (R^2) of each scatter plot was also presented. The points and R^2 presented in blue, orange, and black correspond, respectively, to the data of the Field 1, Field 2, and two-field datasets.

4.3.3.2. Attenuation promoted by Compton Normalization and MLR

The scatter plots of measured versus predicted ex-K and ex-Ca are presented in Figure 7A–H for calibrations with and without CN. For the two-field dataset, models for ex-Ca based on univariate regressions using the Ca-Kα without CN (Figure 7E) presented poor prediction performance (RPD = 1.39 and R^2 = 0.41). This is clearly attributed to the different sample locations, confirming the matrix effect. An explicitly better prediction performance (RPD = 3.02 and R^2 = 0.89) was obtained after CN, with points closely distributed around the 1:1 line, indicating the mitigation of part of the matrix effect present in the two-field dataset (Figure 7F). Even better prediction performance was observed for ex-Ca when using MLR (Figure 7G,H), with R^2 and RPD values increasing from 0.91 and 3.41 for EL (Figure 7G) to 0.94 and 4.05 for ELC (Figure 7H). A similar trend with large matrix effect can be observed for CEC (Figure 7M-P), for which the prediction results of the univariate model improved from being of poor performance for EL (R^2 = 0.35 and RPD = 1.25, Figure 7M) to reasonable performance with ELC (R^2 of 0.64 and RPD of 1.95, Figure 7N). Again MLR for CEC prediction showed further improvement in the prediction performance of CEC ($R^2 \geq 0.76$ and RPD ≥ 2.07, Figure 7O,P), with the best performance obtained after the ELC scenario (R^2 = 0.82 and RPD = 2.34, Figure 7P).

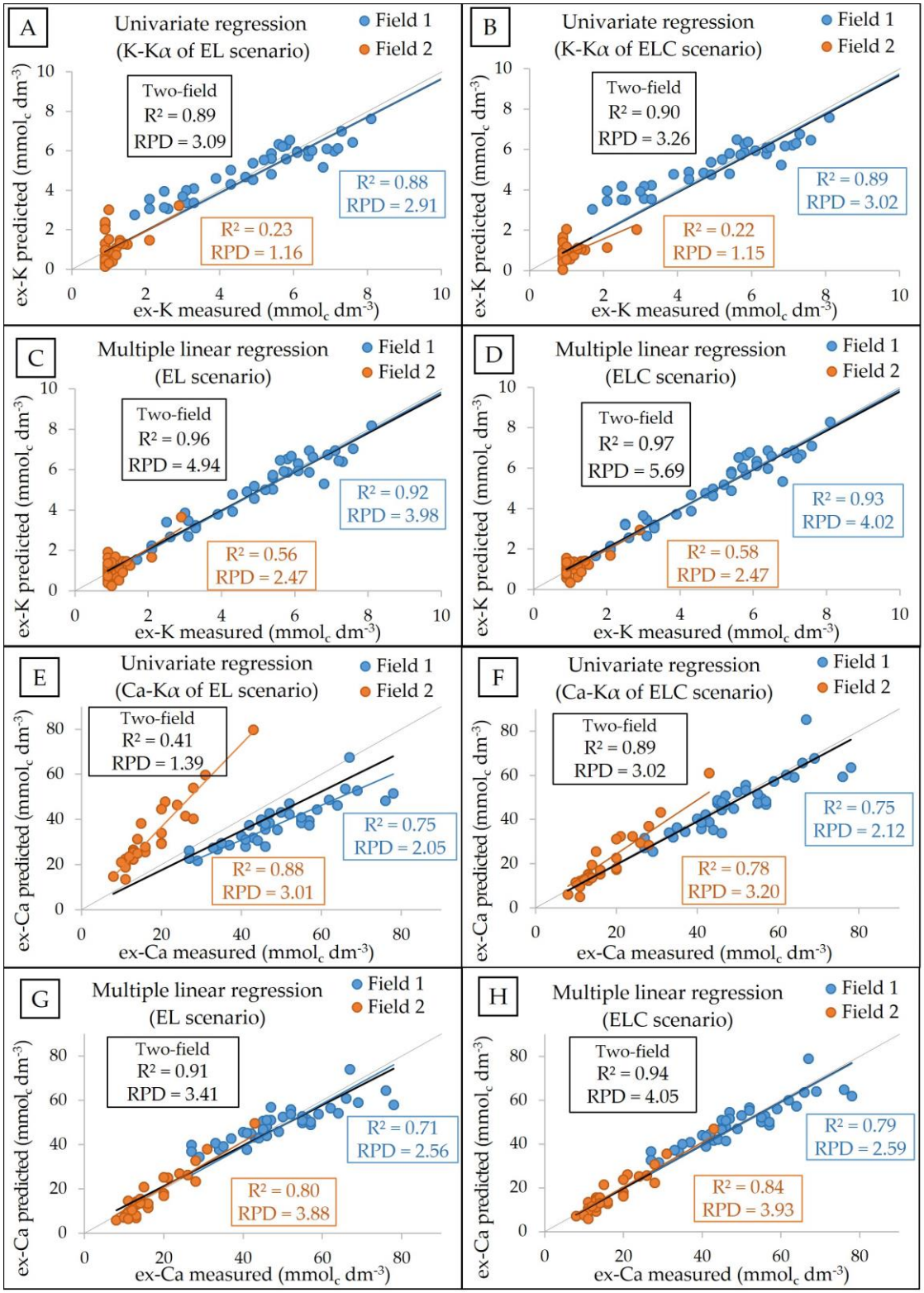
The ex-K prediction results (R^2 and RPD values of 0.89 and 3.09, respectively) with univariate models using the K-Kα without CN did not indicate a distinguished matrix effect in the two-field dataset (Figure 7A), which may be related to the low dispersion of ex-K in Field 2, masking the visualization of matrix effect. However, the ex-K prediction after CN has slightly improved, with R^2 and RPD values of 0.90 and 3.26, respectively (Figure 7B). Extra performance improvement was observed with MLR for ex-K prediction (Figures 7C and 7D).

Although the prediction results of clay using the univariate models were visually of good performances ($0.80 \leq R^2 \leq 0.85$ and $2.28 \leq$ RPD ≤ 2.60), point dispersions far from the 1:1 line indicating nonlinear behaviors, and consequently the clear matrix effect present in the two-field dataset (Figure 7I,J). Even CN did not lead to overcoming the matrix effect on clay prediction, behaving differently from the predictions for CEC, ex-K, and ex-Ca. With MLR, the clay prediction presented a linear dispersion of points around the 1:1 line with better performance indicators (R^2 = 0.92 and RPD ≥ 3.64, Figure 7K,L). However, unlike observed for CEC, ex-K, and ex-Ca predictions, the MLR model after ELC did not show improvement in the prediction performance of clay, compared to the EL scenario. The observed R^2 and RPD values for clay prediction using MLR were, respectively, 0.92 and 3.70, for the EL scenario (Figure 7K), and 0.92 and 3.64, for the ELC scenario (Figure 7L). Finally, the scatter plots of measured versus predicted ex-K, ex-Ca, clay, and CEC for calibrations using PLSR with and without CN are presented in Figure A4 (Appendix Section). The PLSR models showed similar behavior to MLR, with all models showing a linear dispersion of points around the 1:1 line, as well as showing a slight increase

in the prediction performance of ex-K, ex-Ca, and CEC for the ELC scenario in comparison with the EL scenario.

In summary, multivariate models (MLR and PLSR) for the prediction of ex-K, ex-Ca, clay, and CEC using both EL and ELC scenarios improved the linearity in the scatter plots, compared to the univariate models. In turn, CN has proved to be a good solution for mitigating the matrix effect in ex-K, ex-Ca, and CEC predictions, evidenced by increasing the predictive performance of these elements when used in both the univariate and multivariate models. However, CN was not useful for clay prediction, since no improvement in performance was observed in both the univariate and multivariate models.

It is also interesting to note that for predictions using individual field datasets the matrix effect was not as perceptible as seen for the two-field dataset. For the individual field models, the predictions with univariate regressions for EL showed a linear dispersion of points close the 1:1 line, particularly for ex-K in Field 1 (RPD = 2.91, Figure 7A), ex-Ca prediction in Field 1 (RPD = 2.05, Figure 7E), and clay in Field 2 (RPD = 1.43, Figure 7I).


A
Univariate regression
(K-Kα of EL scenario)
Field 1
Field 2
Two-field
R² = 0.89
RPD = 3.09
R² = 0.88
RPD = 2.91
R² = 0.23
RPD = 1.16
ex-K predicted (mmol$_c$ dm^{-3})
ex-K measured (mmol$_c$ dm^{-3})
B
Univariate regression
(K-Kα of ELC scenario)
Field 1
Field 2
Two-field
R² = 0.90
RPD = 3.26
R² = 0.89
RPD = 3.02
R² = 0.22
RPD = 1.15
ex-K predicted (mmol$_c$ dm^{-3})
ex-K measured (mmol$_c$ dm^{-3})
C
Multiple linear regression
(EL scenario)
Field 1
Field 2
Two-field
R² = 0.96
RPD = 4.94
R² = 0.92
RPD = 3.98
R² = 0.56
RPD = 2.47
ex-K predicted (mmol$_c$ dm^{-3})
ex-K measured (mmol$_c$ dm^{-3})
D
Multiple linear regression
(ELC scenario)
Field 1
Field 2
Two-field
R² = 0.97
RPD = 5.69
R² = 0.93
RPD = 4.02
R² = 0.58
RPD = 2.47
ex-K predicted (mmol$_c$ dm^{-3})
ex-K measured (mmol$_c$ dm^{-3})
E
Univariate regression
(Ca-Kα of EL scenario)
Field 1
Field 2
Two-field
R² = 0.41
RPD = 1.39
R² = 0.75
RPD = 2.05
R² = 0.88
RPD = 3.01
ex-Ca predicted (mmol$_c$ dm^{-3})
ex-Ca measured (mmol$_c$ dm^{-3})
F
Univariate regression
(Ca-Kα of ELC scenario)
Field 1
Field 2
Two-field
R² = 0.89
RPD = 3.02
R² = 0.75
RPD = 2.12
R² = 0.78
RPD = 3.20
ex-Ca predicted (mmol$_c$ dm^{-3})
ex-Ca measured (mmol$_c$ dm^{-3})
G
Multiple linear regression
(EL scenario)
Field 1
Field 2
Two-field
R² = 0.91
RPD = 3.41
R² = 0.71
RPD = 2.56
R² = 0.80
RPD = 3.88
ex-Ca predicted (mmol$_c$ dm^{-3})
ex-Ca measured (mmol$_c$ dm^{-3})
H
Multiple linear regression
(ELC scenario)
Field 1
Field 2
Two-field
R² = 0.94
RPD = 4.05
R² = 0.79
RPD = 2.59
R² = 0.84
RPD = 3.93
ex-Ca predicted (mmol$_c$ dm^{-3})
ex-Ca measured (mmol$_c$ dm^{-3})

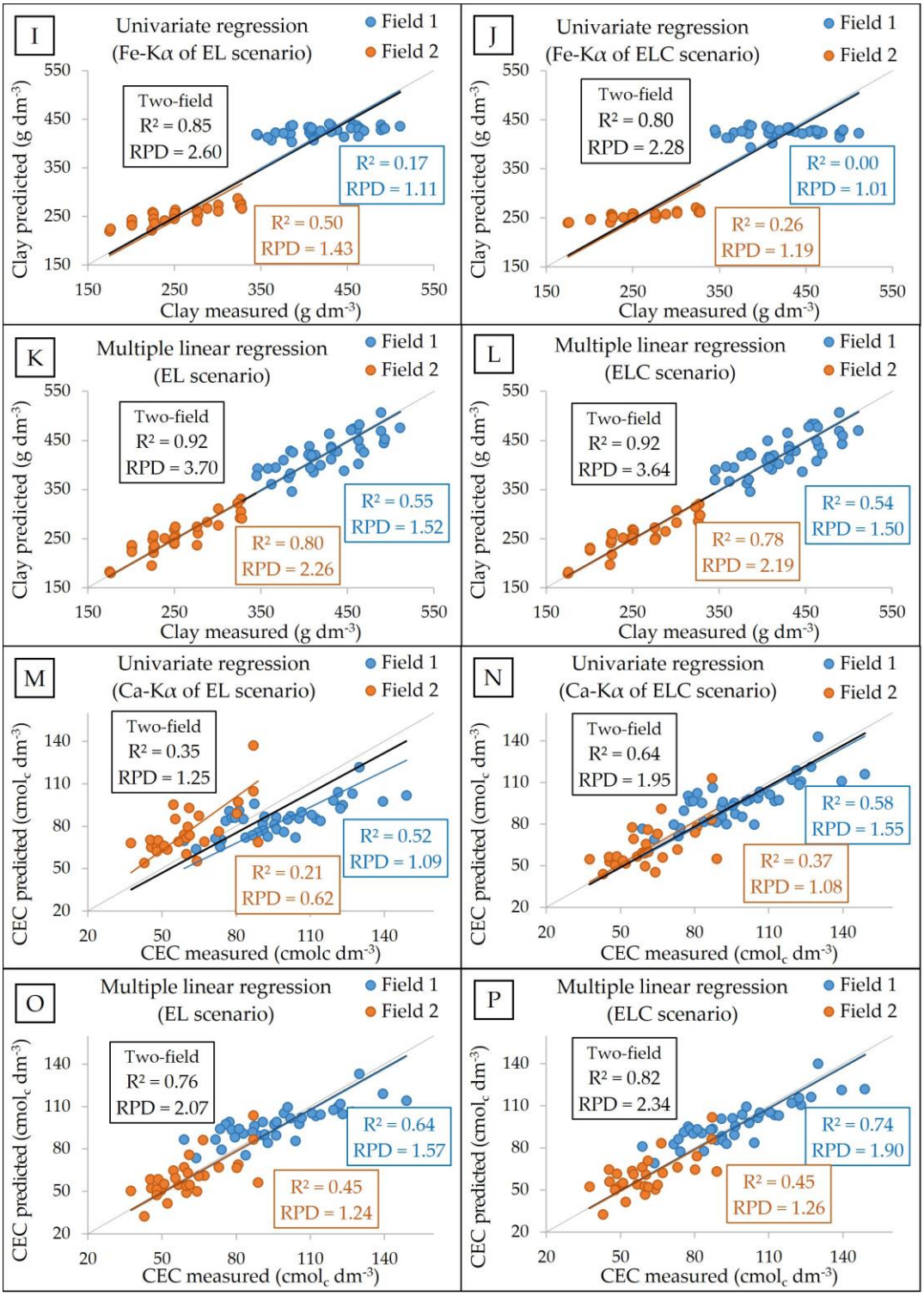


Figure 7. Scatter plots of measured versus predicted exchangeable (ex-) potassium (K; **A–D**) and calcium (Ca; **E–H**), clay (**I–L**), and cation exchange capacity (CEC; **M–P**) obtained from univariate regression using the K-lines without the Compton normalization (EL scenario; **A,E,I,M**, respectively); univariate regression using the K-lines after the Compton normalization (ELC scenario; **B,F,J,N**, respectively); multiple linear regression using all selected emission lines from the EL scenario (**C, G, K**, and **O**, respectively); and multiple linear regression using all selected emission lines from the ELC scenario (**D,H,L,P**, respectively). Models were obtained using the calibration set (*n* of 39, 29, and 68 for the Field 1, Field 2, and two-field datasets) and the validation was performed by "leave-one-out" full cross-validation. Figures of merit in blue, orange, and black correspond, respectively, to the calibrations performed using the Field 1, Field 2, and two-field datasets.

4.3.3.3. Comparison between MLR, PLSR, and the Geo Exploration Package

The results presented in Table 7 show comparable prediction performances for PLSR and MLR for both EL and ELC scenarios, with the biggest difference observed for clay prediction with the ELC scenario (R^2 values of 0.92 and 0.89, for MLR and PLSR, respectively). Smaller differences than 0.03 of R^2 values calculated for all properties indicate that both multivariate models tested allowed correcting the matrix effect present in the two-field dataset with comparable accuracies.

Table 7. Prediction results of the validation set (n = 34) obtained from multiple linear regressions (MLR) and partial least squares regression (PLSR) using portable X-ray fluorescence (XRF) measured emission lines with (ELC) and without (EL) Compton normalization. The results obtained with MLR using the Geo Exploration package data as input X-variables were also shown.

	Geo Exp. Package [3]	MLR		PLSR		
		EL	ELC	EL	ELC	nLV [4]
				R^2		
Clay	0.92	0.91	0.92	0.91	0.89	4
CEC [1]	0.83	0.83	0.87	0.81	0.87	3
ex-K [2]	0.92	0.92	0.95	0.91	0.94	3
ex-Ca [2]	0.90	0.92	0.96	0.91	0.96	4
				RMSE		
Clay	27.44	31.72	29.40	32.65	35.56	4
CEC	11.90	11.08	10.19	11.56	9.42	3
ex-K	0.72	0.65	0.53	0.75	0.60	3
ex-Ca	6.09	5.63	4.09	5.99	4.11	4
				RPD		
Clay	3.35	2.90	3.13	2.82	2.59	4
CEC	2.20	2.36	2.57	2.26	2.78	3
ex-K	3.16	3.47	4.26	3.01	3.77	3
ex-Ca	3.24	3.50	4.82	3.29	4.80	4

[1] Cation exchange capacity; [2] extractable (ex-) nutrients (ex-K and ex-Ca); [3] Geo Exploration package (Bruker AXS, Madison, WI, EUA); [4] number of latent variables used for the PLSR models calibrated for both EL and ELC scenarios. The coefficient of determination (R^2) and residual prediction deviation (RPD) values are presented on grayscale, highlighting the highest values. The root-mean-square error (RMSE) was given in g dm^{-3} for clay, and in $mmol_c$ dm^{-3} for CEC, ex-K, and ex-Ca.

Again models developed after the CN (ELC scenario) provided subtly better prediction performances for all soil attributes evaluated. The only exception was the PLSR model for clay, which provided better results for the EL scenario (R^2 of 0.91), compared to the ELC scenario (R^2 of 0.89). The largest improvements after ELC were observed for CEC and ex-Ca models, either with MLR or PLSR, with an increase in R^2 values for CEC of 0.04 and 0.06 and for ex-Ca of 0.04 and 0.05 for MLR and PLSR models, respectively. All other R^2 increases, when using the ELC scenario, were less than 0.04.

Finally, results achieved using the Geo Exploration package data showed R^2 values comparable to the MLR and PLSR models, with R^2 values of 0.83 for CEC, 0.90 for ex-Ca, and 0.92 for

clay and ex-K. Despite that, the prediction performances achieved using the Geo Exploration package data showed, in general, higher RMSE values. The only exception was the clay prediction, in which the Geo Exploration package presented the best result. For the other parameters (CEC, ex-K, and ex-Ca), the ELC models improved expressively the RMSE values. For CEC, the RMSE of PLSR-ELC model presented 21% lower than the one achieved by the Geo Exploration package; for ex-K, the MLR-ELC model presented a reduction of 26%; and for ex-Ca, the RMSE values of both PLSR-ELC and MLR-ELC reduced by 33% compared to the Geo Exploration package.

4.4. Discussion

4.4.1. Effect of X-ray tube configuration (data analysis part 1)

The X-ray tube anode emits a continuous spectrum when it is struck by accelerated electrons (Albertini et al., 2018). Essentially, when the voltage of the tube is increased, it also increases: (i) the energy range of the continuous spectrum; (ii) the overall emission intensity of the continuous spectrum; and (iii) the energy at maximum intensity in the continuous spectrum (Jenkins, 2012). In parallel, when increasing the current applied to the X-ray tube, increments only in the intensity of the continuous spectrum are observed, since the photon flow is increased (Lindgren, 2000). A schematic figure illustrating the behavior of the continuous spectrum emitted by the X-ray tube when modifying its voltage and current is presented in the Appendix Section (Figure A3).

Different configurations of X-ray tube—with different combinations of voltage and current—change the continuous spectrum emitted by the X-ray tube and, therefore, the fluorescence behavior emitted by the elements present in the sample (Jenkins, 2012). The results obtained by the data analysis part 1 show that the intensity of XRF emission of the different elements was indeed influenced by the configuration of the X-ray tube (Figure 5). Nevertheless, the emission lines detected in both the EL-15 and EL-35 scenarios [namely, Al-Kα, Si-Kα, Rh-Lα (Thomson), K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, and Cu-Kα] showed significant correlation with each other (Figure 5), while Ni-Kα showed the smallest correlations between the EL-15 and EL-35 scenarios (Figure 5I). The reduced intensity and dispersion of Ni XRF emission in EL-15 (Figure 5I) can be explained by the short energy range that is able to excite Ni (K_{edge} = 8.33 keV) when the tube was set at 15 kV. In turn, for the EL-35, the Ni XRF emission was favored, similar to that observed for elements with an absorption edge close to Ni (*e.g.*, Fe and Cu). In addition to that, it is important to highlight that part of the Ni signal is coming from the equipment itself (as can be seen in Figure A1); hence, it is also highly affected by the tube configuration.

X-ray tube configurations that produce continuous radiation with peak intensity closer to the absorption edge of light elements (*e.g.*, 1.30, 1.56, 1.84, and 2.15 keV, representing the K_{edge} of Mg, Al, Si, and P, respectively) benefit the XRF emission of these elements (Potts et al., 1984; Albertini et al., 2018). Therefore, it is expected that the EL-15 configuration would promote better predictions of light elements, such as ex-K and ex-Ca. However, the results obtained in this study do not support this assumption. The performance of the two X-ray tube configurations was very similar, even for the prediction of ex-K and ex-Ca, whose prediction models had a great weight of the emission lines of light elements (K-Kα and Ca-Kα, respectively). Working on tropical soils, Silva et al. (2016) compared different operational modes (Trace Mode and General Mode) of a XRF sensor (using a Bruker, S1 Titan LE, Bruker, Madison, USA) for the prediction of ex-Ca, observing slight variations in performance ($0.77 \leq R^2 \leq 0.85$) for ex-Ca prediction when comparing both operational modes. Although the details of the X-ray tube configuration of the operational modes evaluated are not disclosed, it is common that the instrumental conditions for "trace" analysis use low voltage and high current, similar to the EL-15 scenario applied in this work. Thus, this subtle oscillation of prediction performance due to changing the XRF operational mode observed by Silva et al. (2019) can be compared to the results shown in the present study.

The minimal effect of both X-ray tube configurations on the prediction of fertility attributes may well be explained by the low importance of the X-variables Rh-Kα Compton, Rh-Kα Thomson (scattering peaks present in the region between 18 and 23 keV), and the Ni-Kα. If these variables had a higher weight on the regression of some attribute, more relevant differences might be possible. The scattering peaks can be useful independent variables for OM prediction, as observed by Morona et al. (2017).

The spectra observed in this research did not present emission lines for Mg and P due to their low XRF intrinsic sensitivity (*e.g.*, lower fluorescence yield) and low Si-detector efficiency at this photon energy range, which is translated in high limits of detection. In this work, the greatest ptc of Mg and P were, respectively, 789.65 and 669.18 mg kg^{-1}, corresponding to an extractable content of 54 $mmol_c$ dm^{-3} and 104 mg dm^{-3} for ex-Mg and ex-P, respectively. X-ray tube configurations with energy levels below 15 kV should be further investigated for the detection of Mg and P, as they can promote more efficient excitation of these elements in soil samples (Potts et al., 1984). By further reducing the X-ray tube voltage setting, the intensity of the continuous spectrum in low energies is increased, as well as the limit of detection of low Z elements being reduced. Potts et al. (1984), working with an ED-XRF in rock powder samples, observed that Mg, Si, and Ca are most efficiently excited when using the X-ray tube configured at low energies (*e.g.*, 6–8 kV).

In view of our results, both scenarios of X-ray tube configuration (EL-15 and EL-35) can be used for predicting the studied fertility attributes, since their prediction results were comparable. Nevertheless, it is important to point out that some studies suggested that the information at the scattering region between 18 and 23 keV that is present only in spectra obtained with the Rh-tube configured at voltages above 23 kV may be promising for the prediction of organic attributes (Morona et al., 2017) and for the matrix effect attenuation using the Compton normalization (Jenkins, 2012).

4.4.2. Matrix effect attenuation (data analysis part 2)

Analyses of soil samples with heterogeneous composition using XRF sensors are subject to the called matrix effects. The matrix effect comprises the physical and chemical properties of the sample that influence the XRF spectra. Chemical matrix effects originate from differences in the concentrations of the elements that compose the soil matrix (*e.g.*, concentrations of Fe, Si, Al, Ti, etc.), which generate interferences from absorption and/or enhancement of the fluorescence produced by the sample (USEPA, 2007). The total contents of Ti, Mn, and Fe and their emission lines, are different between both fields, confirming the matrix difference between them (Figure 6). Prediction models of the univariate calibration using the emission lines as input without CN showed a clear nonlinear dispersion of points around the 1:1 scatter plot line, with the samples divided into two groups according to their field of origin, especially for clay (Figure 7I), CEC (Figure 7M), and ex-Ca (Figure 7E) models. For ex-K, the matrix effect was less evident than the others (Figure 7A), most probably due to the small variation range of this element in Field 2, as above commented (Section 4.3.1.2.).

The univariate calibration of clay, ex-Ca, and ex-K, used, respectively, the emission lines of Fe Kα, Ca-Kα, and K-Kα; the Ca-Kα was also used for the prediction of CEC. These lines can suffer from an absorption effect depending on the presence of elements that have an absorption edge with slightly lower energy than their fluorescent peaks (Jenkins, 2012). At the same time, enhancement effects may also occur due to the presence of elements that emit XRF with slightly higher energy than the absorption edge energy of these elements (USEPA, 2007). For example, Fe tends to absorb the Cu K-line, since Fe K_{edge} (7.11 keV) is close in energy to the Cu-Kα (8.05 keV), enhancing the intensity of Fe-Kα (6.41 keV). At the same time, Fe-Kα can be absorbed by the Cr, which presents K_{edge} at 5.99 keV, with energy slightly below that of Fe-Kα. Regarding the K-lines of K and Ca, the same interferences may take place between K (K-Kα = 3.31 keV and K_{edge} = 3.61 keV), Ar (Ar-Kα = 2.96 keV and K_{edge} = 3.20 keV), and Ca (Ca-Kα = 3.69 keV and K_{edge} = 4.04 keV), as well as between Ca, K, and Ti (Ti-Kα = 4.51 keV and K_{edge} = 4.97 keV).

The emission lines observed in this work presented well-defined peaks (Figure 2), indicating a low interference due to peak overlaps. However, chemical matrix effects due to spectral interferences

can occur in soil samples, especially when heavier elements (*e.g.*, Ba, Pb, W, etc.) are present in the soil matrix (Taggart et al., 1987). The energy of the L-lines of heavier elements occurs mainly within 0-10 keV, the region where the K-lines of important elements for soil assessment (*e.g.*, Ti, Cr, Fe, Co, and Ni) exist and lead to increase the intensity of the K-lines measured by the detector (Gallhofer and Lottermoser, 2018). Possible interference with K-Kα (3.31 keV) is the emissions of Cd-Lβ (3.33 keV) and Ti-Kα (4.51 keV) due to fluorescence of Ba-Lα (4.47 keV).

Solutions for matrix effect mitigation should be used to optimize the performance of prediction models in datasets with contrasting soil matrices (Markowicz, 2008). In this work, the use of CN combined with multivariate regressions was tested as a potential solution to improve the prediction accuracy of selected fertility attributes. Our results showed that CN allowed correcting part of the matrix effect present in the two-field dataset for CEC, ex-K, and ex-Ca predictions. This is evident by comparing the results of the univariate models using the K-lines with and without CN. Better linearity of points scattered around the 1:1 line and greater prediction accuracy obtained after CN with RPD value increases of 0.70 for CEC (Figure 7M,N), 0.17 for ex-K (Figure 7A,B), and 1.63 for ex-Ca (Figure 7E, F). This was not true only for clay prediction, which may be explained by the large difference in intensity observed for the Fe-Kα emission line of the Field 1 and Field 2 datasets (Figure 6I,J). The clay prediction using the two-field dataset required employing MLR, which resulted in reducing the matrix effect (Figure 7I–L) by using a multilayer of information available in XRF data and improving the prediction accuracy of this variable. The principal variables for clay prediction were the K-lines of Fe, Ti, and Si (Table A1, Appendix Section).

It should be noted that MLR without CN was sufficient to mitigate the error caused by the matrix effect and provide satisfactory prediction performances ($R^2 \geq 0.76$ and RPD ≥ 2.07) for clay, CEC, ex-K, and ex-Ca (Figure 7). These MLR models without CN had even slightly higher prediction performances than the corresponding univariate models after CN, with R^2 value increases of 0.12 for clay (Figure 7J,K) and CEC (Figure 7N,O), 0.06 for ex-K (Figure 7B,C), and 0.02 for ex-Ca (Figure 7F,G). This may be explained by the fact that the multiple regressions account for emission lines of different elements that compose the soil matrix (*e.g.*, Si, Al, Fe, Ti, etc.), which contain more information about the target element to be predicted than the single emission line accounted for by the univariate regression analysis (Kaniu et al., 2011; Morona et al., 2017). The MLR-ELC scenario provided the best performance for the prediction of CEC, ex-K, and ex-Ca in the two-field dataset, reaching $R^2 \geq 0.82$ and RPD ≥ 2.34. The MLR-ELC outperformed MLR-EL by increasing the R^2 values by 0.06, for CEC (Figure 7O,P), 0.01, for ex-K (Figure 7C,D), and 0.03, for ex-Ca (Figure 7G,H), while the prediction of clay remained similar in both scenarios, achieving R^2 of 0.92 (Figure 7K,L).

Multivariate models including MLR and PLSR are better solutions to exploit the hidden information present in the spectra (Panchuk et al., 2018), mitigating the matrix effect while improving the prediction accuracy of fertility attributes. Comparing the predictive performances of PLSR and MLR models, it was observed that both performed similarly (Table 7). Panchuk et al. (2018), reviewing the application of chemometric methods to XRF data, explained that PLSR models can perform similarly to MLR models when there are no overlaps in the emission lines selected for the models, otherwise PLSR models generally perform better than MLR. The spectra obtained in this work presented well-resolved peaks, which explain the similar prediction performance of MLR and PLSR models. In addition, results obtained with MLR and PLSR models showed prediction performances comparable to the Geo Exploration package (Table 7). These results indicate that the prediction performance of the multivariate models proposed in this paper may achieve a similar quality to those obtained by preprogrammed measurement packages (Table 7), and both approaches allowed the mitigation of the soil matrix effect. It is important to emphasize that preprogrammed measurement packages generally use calibration methods capable of correcting the matrix effect, as long as it is well-calibrated with accurate sensitivity coefficients.

In view of our results, a simple method for matrix effect mitigation by combining CN and multivariate models allowed a dataset of largely contrasting soil matrices, to be compatible for general calibrations of fertility attributes using XRF sensors. In general, the different strategies evaluated in this work can be presented in the following descending order of goodness of performance: multiple regressions using the selected emission lines after CN (ELC scenario) ≥ multiple regressions using the selected emission lines without CN (EC scenario) > univariate regression using emission lines after CN > univariate regression using emission lines without CN. Therefore, for the prediction of soil fertility attributes in datasets with the presence of matrix effect, we suggest the combined use of CN with any of the tested multivariate models (MLR or PLSR).

Most of the works that assessed XRF for predicting soil fertility attributes have used preprogrammed measurement packages (Zhu et al., 2011; Weindorf et al., 2012; Sharma et al., 2014, 2015; O'Rourke et al., 2016; Silva et al., 2017; Teixeira et al., 2018; Silva et al., 2019; Andrade et al., 2020), which use not disclosed general algorithms to preprocess the XRF data and mitigate the matrix effect (Bowers, 2019). Studies use these packages with a robust number ($n \geq 120$) of soil samples reported R^2 for clay prediction ranging between 0.85 and 0.97 (Zhu et al., 2011; Lima et al., 2019), from 0.86 to 0.92, for CEC (Sharma et al., 2015; Silva et al., 2017) , from 0.67 to 0.81, for ex-K (Silva et al., 2017; Andrade et al., 2020), and from 0.71 to 0.89, for ex-Ca (Silva et al., 2017; Andrade et al., 2020). The prediction accuracy for clay, CEC, ex-K, and ex-Ca observed in this research was similar or higher (Table 7) than those reported by the above-mentioned studies. Although preprogrammed measurement packages allow for

a user-friendly approach that contributes to the popularization and expansion of the XRF application, they are not fully transparent and do not allow flexibility regarding the XRF scanning time (Tavares et al., 2020). Furthermore, the preprogrammed measurement packages offer total content values to be used as input for calibration models, which demand an extra validation step to confirm that the total contents are accurately measured. In general, accuracy problems with these routines can occur, which demands a "correction factor" that can be determined based on the ratio of the certified concentration of standard reference material and the XRF measured concentration (Mantler et al., 2006; Koch et al.,2017; Chakraborty et al., 2019). In this sense, using intensity values of XRF emission lines is totally enough for establishing the prediction models avoiding this issue, and also the nonflexible procedures related to commercial packages (*e.g.*, pre-established X-ray tube settings and scanning time).

Although the matrix effect presented in the two-field dataset has been successfully mitigated in this work by combining CN and multivariate regressions, more investigations are needed to confirm that this approach can be applied for different datasets having different variation degrees of soil matrices. Therefore, it is essential that future research also evaluate other types of soils with contrasting soil characteristics that can proportionate distinct levels of chemical matrix effects, as well as evaluate more sophisticated methods for modeling XRF data (*e.g.*, machine learning, and computational models), especially for calibrations involving a larger number of samples.

4.5. Conclusion

The X-ray tube configuration of an X-ray fluorescence (XRF) spectrometer affected the intensity of the emission lines of the different elements detected. By reducing the Rh-anode X-ray tube voltage from 35 to 15 kV and incrementing the tube current from 7 to 23 μA (designated as EL-35 and EL-15 scenarios, respectively), increases in X-ray fluorescence emission intensity were observed for the Si and Al elements. The behavior of the Kα emission lines of K and Ca elements showed no differences in both X-ray tube configurations tested. All other emission lines detected (namely, Ti-Kα, Mn-Kα, Fe-Kα, Ni-Kα, and Cu-Kα) were higher at 35 kV. Emission lines of P and Mg were not detected using any of the tested configurations.

Both XRF sensor configurations performed satisfactorily ($0.61 \leq R^2 \leq 0.96$) to predict clay, OM, CEC, V, and extractable (ex-) K, Ca, and Mg, whereas the prediction results for pH and ex-P were unsatisfactory ($R^2 \leq 0.38$). The different XRF tube configurations (EL-15 and EL-35), applied in a Rh-anode X-ray tube, provided similar performances to predict all key fertility attributes evaluated. Therefore, data from this work support that both X-ray tube configurations can be used for predicting

fertility attributes. However, it is important to note that only tubes configured at 35kV have the Rh-scattering region between 18 and 23 keV, which may bring useful information in some datasets.

Regarding the matrix effect mitigation, solutions using Compton normalization (CN) combined with multivariate regressions (*e.g.*, multiple linear regression (MLR) and partial least squares regression (PLSR)) were evaluated and compared to a modeling approach using both univariate and multivariate regression analyses without CN. In addition, the proposed solutions were also compared to modeling based on a preprogrammed measurement package approach (Geo Exploration package, Bruker AXS, Madison, WI, EUA). Results showed that models developed with emission lines without CN were clearly affected by the matrix effect. For CEC, ex-K, and ex-Ca predictions, the matrix effect was attenuated by both CN and multivariate regressions (PLSR or MLR). For clay prediction the matrix effect was mitigated by the multivariate regressions only, for which CN had no effect. The prediction performances of MLR and PLSR models were comparable for all evaluated attributes. For the two-field dataset, the combined use of CN + multivariate regression achieved the best performances for the prediction of CEC (R^2 = 0.82), ex-K (R^2 = 0.97), and ex-Ca (R^2 = 0.94), whereas comparable performance was recorded for clay with and without CN (R^2 = 0.92). Therefore, it was suggested to use multivariate regressions (MLR or PLSR) combined with CN to predict key fertility attributes in sample sets with contrasting soil matrices.

The prediction performance observed for the solutions proposed in this work showed comparable results to the approach based on the Geo Exploration package, as well as to previously published works that used preprogrammed measurement packages. Moreover, the methodology applied in this study is simple and transparent, which allows the optimization of the procedures of XRF data acquisition, pre-processing, and modeling. Such knowledge is essential to enable more accurate and faster XRF analysis, as well as developing new applications of XRF sensors in the context of precision agriculture and soil science.

Appendix

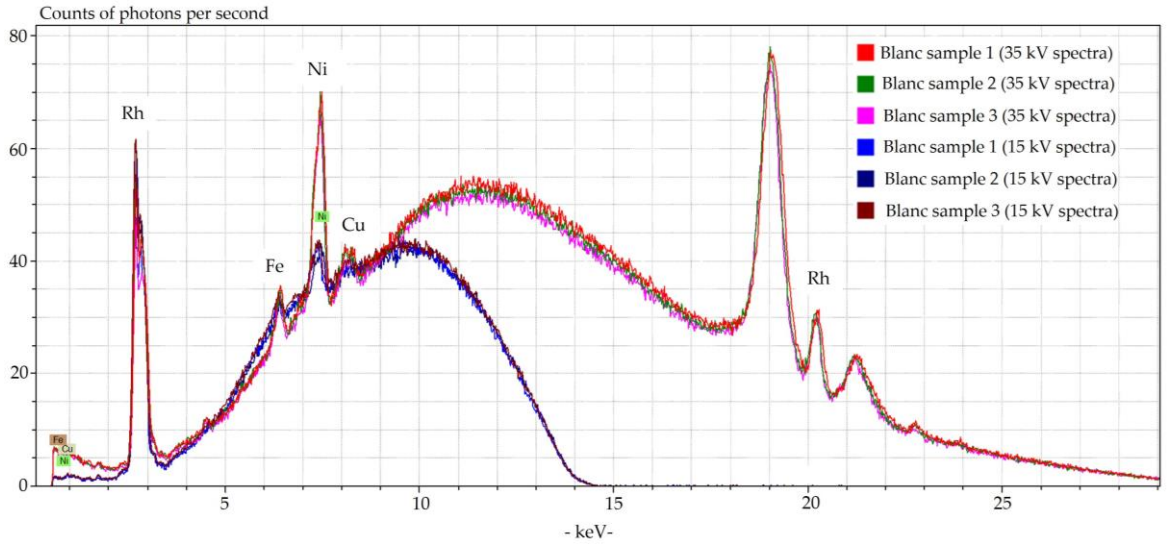


Figure A1. Spectra of the blanc sample obtained during the XRF data acquisition using the X-ray tube configured at 15 and 35 kV. The spectra acquired at the beginning, after 50 samples, and at the end of the measurements are designated in the legend as Blanc sample 1, 2, and 3, respectively.

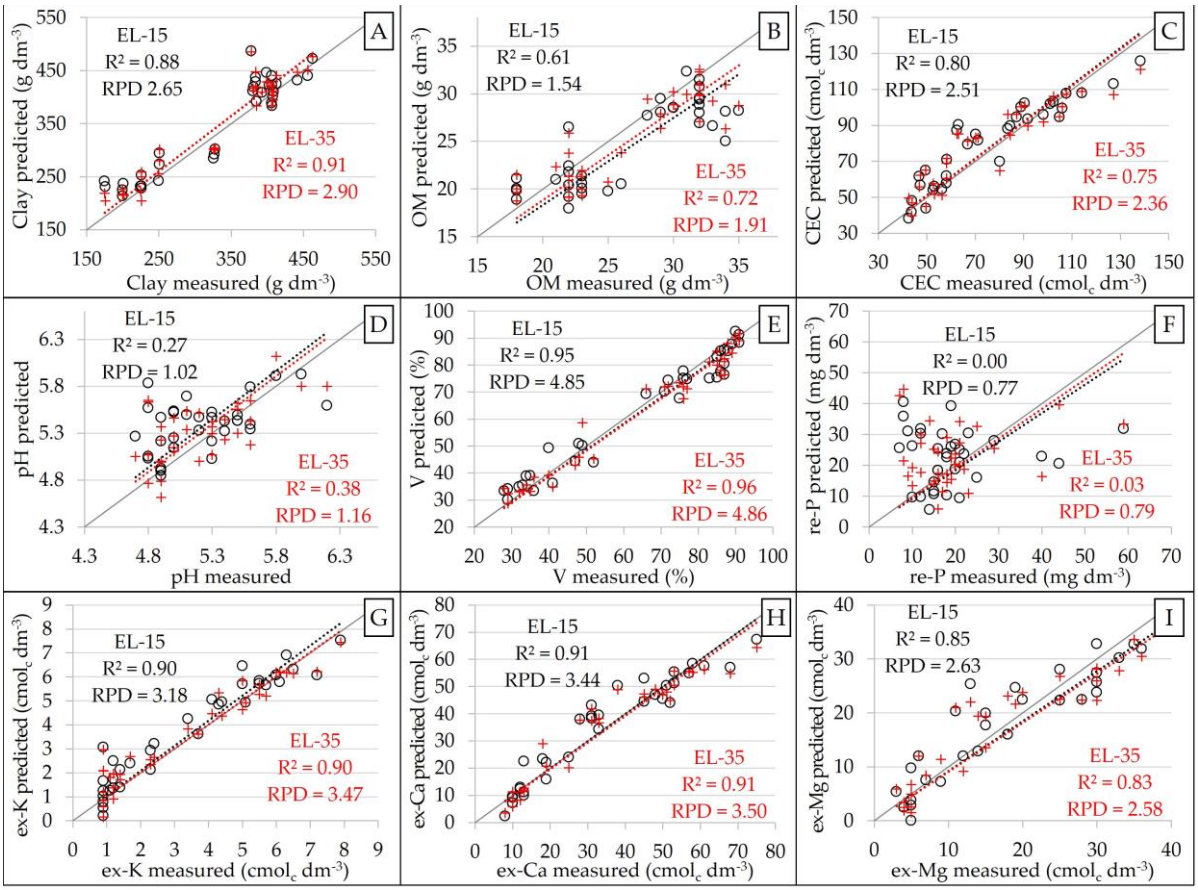


Figure A2. Scatter plots (for the calibration set, n = 34) of measured versus predicted clay, organic matter (OM), cation exchange capacity (CEC), pH, base saturation (V), and exchangeable (ex-) P, K, Ca, and Mg (**A-I**, respectively), which are predicted using multiple linear regressions (MLR). The data of EL-15 and EL-35 scenarios were represented by circles and crosses, respectively. The coefficient of determination (R^2) and residual prediction deviation (RPD) values are presented in black and red for EL-15 and EL-35 scenarios, respectively.

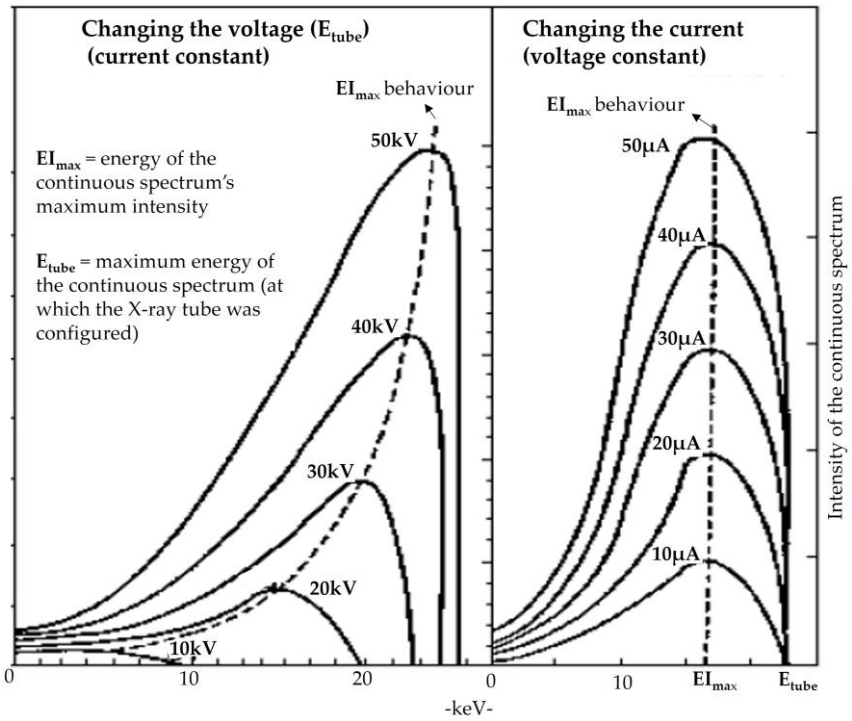


Figure A3. Behavior of the continuous spectrum emitted by the X-ray tube when modifying its voltage (keeping the current constant) and current (keeping the voltage constant) [adapted from Bertin (1978)].

Based on Figure A3, and using Equation (A1), as suggested by Taggart Jr et al. (1987), the energy of the continuous spectrum's maximum intensity (EI_{max}) was estimated, in keV, for each X-ray tube configuration.

$$EI_{max} = E_{tube} / 1.5, \quad \text{(A1)}$$

where E_{tube} is the maximum energy, in keV, of the continuous spectrum; in other words, that is the energy corresponding to the voltage that the X-ray tube was configured for. The calculated EI_{max} was 10.00 and 23.33 keV for the X-ray tube configured at 15 and 35 kV, respectively.

Table A1. Importance of X-ray fluorescence (XRF) variables for the prediction of studied soil fertility attributes on the two-field dataset, using both tested XRF scenarios, with (ELC scenario) and without (EL scenario) applying the Compton normalization to the selected emission lines. The values presented correspond to the t-value for each standardized coefficient obtained in the regressions calibration.

		Al-Kα	**Si-Kα**	**K-Kα**	**Ca-Kα**	**Ti-Kα**	**Mn-Kα**	**Fe-Kα**	**Ni-Kα**	**Cu-Kα**	**Rh-Lα[3]**	**Rh-Kα[3]**	**Rh-Kα[4]**
Clay	EL	−1.1	**−2.4**	−1.3	−0.4	**−2.8**	−0.5	**5.2**	−1.1	−1.2	−1.2	0.1	−0.8
	ELC	−0.6	**−4.3**	−1.1	−0.4	**−2.5**	−0.1	**3.1**	−1.2	−1.7	−1.7	1.1	**−2.5**
CEC[1]	EL	−0.5	−0.4	1.5	**4.6**	**2.5**	−1.6	−0.4	1.3	0.8	−0.8	0.0	−0.2
	ELC	−0.3	−0.5	1.4	**5.9**	**3.3**	−1.2	0.8	0.9	0.9	−0.2	−0.7	1.8
ex-K[2]	EL	−1.0	−1.0	**19.0**	**−5.8**	−0.8	**−2.1**	−1.4	−0.5	−0.6	**−4.4**	−1.2	**−2.5**
	ELC	−0.8	−0.6	**22.0**	**−5.4**	−0.1	**−2.3**	−0.8	0.8	−0.5	**−4.0**	−0.9	−1.9
ex-Ca[2]	EL	−1.1	0.8	1.0	**9.2**	−0.2	−1.8	0.5	1.1	0.6	0.8	1.5	−1.3
	ELC	−1.3	0.6	1.0	**11.9**	−0.4	−1.0	0.9	0.7	0.5	1.3	1.5	0.1

[1] Cation exchange capacity; [2] extractable (ex-) nutrients (ex-K and ex-Ca); [3] Thomson scattering peaks (Rh-Lα and Rh-Kα); [4] Compton scattering peak (Rh-Kα). The emboldened values indicate a significant correlation at the probability level of 0.05; significant values were presented on grayscale, with the most important variables having the darkest color and vice versa.

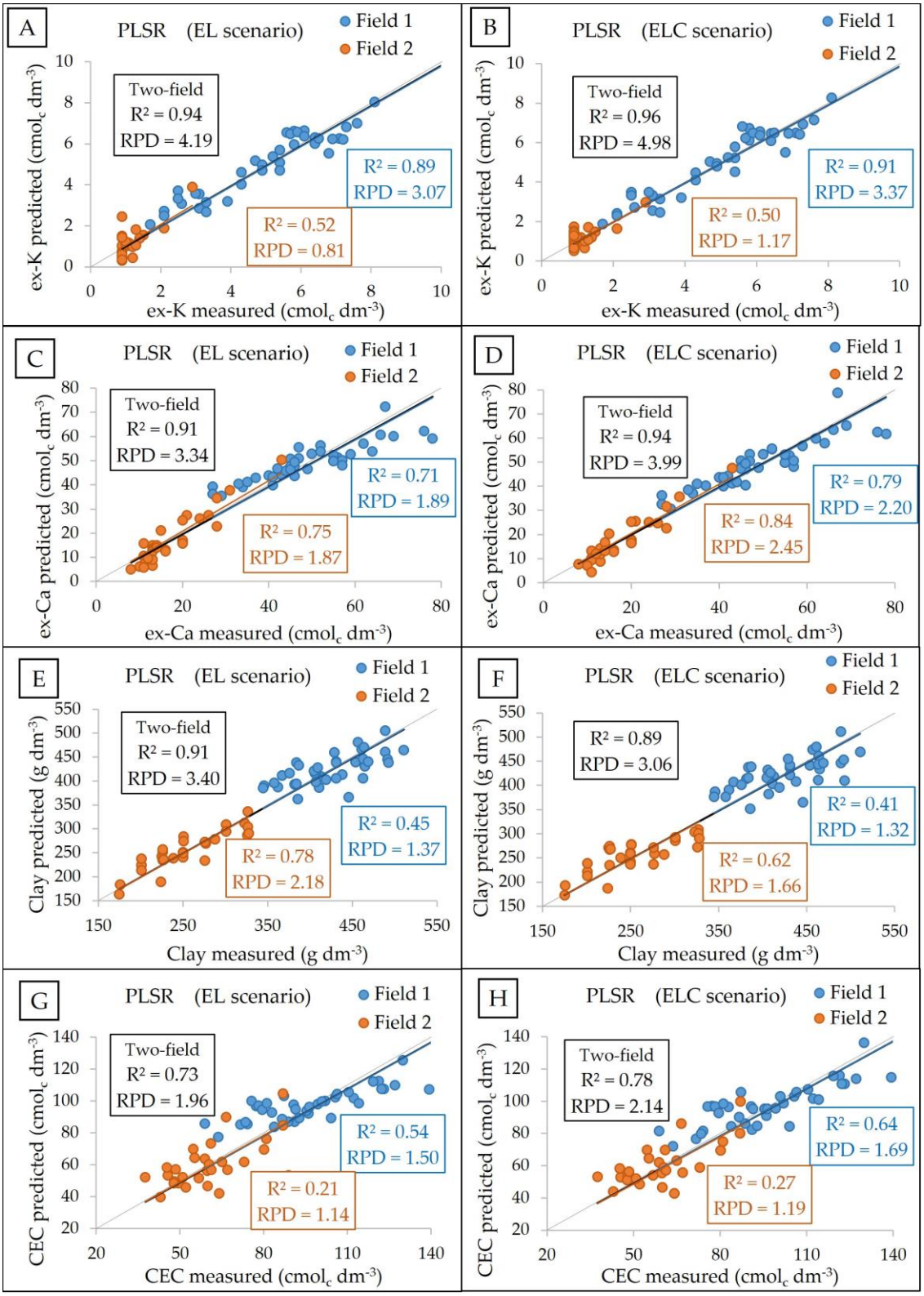


Figure A4. Scatter plots of measured versus predicted extractable (ex-) K (A,B), and Ca (C,D), clay (E,F), and cation exchange capacity (CEC) (G,H) obtained with partial least squares regression (PLSR) using all selected emission lines from the EL (B,D,F,H) and ELC scenario (A,C,E,G). Models were obtained using the calibration set (n of 39, 29, and 68 for the Field 1, Field 2, and two-field dataset) and the validation was performed by "leave-one-out" full cross-validation. Figures of merit in blue, orange, and black correspond, respectively, to the calibrations performed using the Field 1, Field 2, and two-field datasets.

References

Abreu, C.H., Jr.; Muraoka, T.; Lavorante, A.F. Relationship between acidity and chemical properties of Brazilian soils. *Sci. Agric.* **2003**, 60, 337–343, doi:10.1590/S0103-90162003000200019.

Albertini, V.R.; Paci, B.; Generosi, A. Energy dispersive, X-ray fluorescence analysis. In *Encyclopedia of Analytical Chemistry: Applications, Theory and Instrumentation*; John Wiley & Sons: New York, NY, USA, 2018; pp. 1–17, doi:10.1002/9780470027318.a6806.pub3.

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

Bertin, E.P. *Introduction to X-ray spectrometric analysis*; Plenum Press: New York, NY, USA, 1978; p.485.

Bowers, C. Matrix Effect Corrections in X-ray Fluorescence Spectrometry. *J. Chem. Educ.* **2019**, 96, 2597–2599, doi:10.1021/acs.jchemed.9b00630.

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

Chang, C.W.; Laird, D.A.; Mausbach, M.J.; Hurburgh, C.R. Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. *Soil Sci. Soc. Am. J.* **2001**, 65, 480–490, doi:10.2136/sssaj2001.652480x.

Currie, L.A. Limits for qualitative detection and quantitative determination. Application to radiochemistry. *Anal. Chem.* **1968**, 40, 586–593, doi:10.1021/ac60259a007.

Element, C.A.S. Method 3051A microwave assisted acid digestion of sediments, sludges, soils, and oils. Z. *Für Anal. Chem.* **2007**, 111, 362–366.

Fontes, M.P.F. Intemperismo de rochas e minerais. In *Pedologia: Fundamentos*; Ker, J.C., Curi, N., Schaefer, C.E.G.R., Vidal-Torrado, P., Eds.; Sociedade Brasileira de Ciência do Solo: Viçosa, Brazil, 2012; pp. 181–205. (In Portuguese)

Gallhofer, D.; Lottermoser, B.G. The influence of spectral interferences on critical element determination with portable X-ray fluorescence (pXRF). *Minerals* **2018**, *8*, 320, doi:10.3390/min8080320.

Hartemink, A.E. Soil science in tropical and temperate regions—Some differences and similarities. *Adv. Agron.* **2002**, 77, 269–292, doi:10.1016/S0065-2113(02)77016-8.

IUSS Working Group WRB. World reference base for soil resources 2014. In World Soil Resources Reports No. 106; Schad, P., van Huyssteen, C., Micheli, E., Eds.; FAO: Rome, Italy, 2014; p. 189, ISBN 978-92-5-108369-7.

Jenkins, R. X-Ray Fluorescence Spectrometry; John Wiley & Sons: Hoboken, NJ, USA, 2012; p. 232.

Kaniu, M.I.; Angeyo, K.H.; Mangala, M.J.; Mwala, A.K.; Bartilol, S.K. Feasibility for chemometric energy dispersive X- ray fluorescence and scattering (EDXRFS) spectroscopy method for rapid soil quality assessment. *X- Ray Spectrom.* **2011**, *40*, 432–440, doi:10.1002/xrs.1363.

Kennard, R.W.; Stone, L.A. Computer aided design of experiments. *Technometrics* **1969**, 11, 137–148, doi:10.1080/00401706.1969.10490666.

Koch, J.; Chakraborty, S.; Li, B.; Kucera, J.M.; Van Deventer, P.; Daniell, A.; Faul, C.; Man, T.; Pearson, D.; Duda, B.; et al. Proximal sensor analysis of mine tailings in South Africa: An exploratory study. *J. Geochem. Explor.* **2017**, *181*, 45–57, doi:10.1016/j.gexplo.2017.06.020.

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

Lindgren, E.S. X-ray fluorescence analysis-energy dispersive. In *Encyclopedia of Analytical Chemistry*; Meyers, R.A., Ed.; John Wiley & Sons: Chichester, UK, 2000; ISBN 0471 97670 9.

Markowicz, A.A. Chapter 2 Quantification and Correction Procedures. In *Portable X-ray Fluorescence Spectrometry: Capabilities for In-situ Analysis*; The Royal Society of Chemistry: London, UK, 2008; pp. 13–38.

Mantler, M.; Willis, J.P.; Lachance, G.R.; Vrebos, B.A.R.; Mauser, K.E.; Kawahara, N.; Rousseau, R.M.; Brouwer, P.N. Quantitative Analysis. In *Handbook of Practical X- Ray Fluorescence Analysis*; Burkhard, B., Kanngiesser, B., Langhoff, N., Wedell, R., Wolff, H., Eds.; Springer: Berlin/Heidelberg, Germany, 2006; pp. 309–410.

Molin, J.P.; Tavares, T.R. Sensor systems for mapping soil fertility attributes: Challenges, advances and perspectives in Brazilian tropical soils. *Eng. Agric.* **2019**, 39, 126–147, doi:10.1590/1809-4430-Eng.Agric.v39nep126-147/2019.

Morona, F.; Santos, F.R.; Brinatti, A.M.; Melquiades, F.L. Quick analysis of organic matter in soil by energy-dispersive X-ray fluorescence and multivariate analysis. *Appl. Radiat. Isotopes* **2017**, 130, 13–20, doi:10.1016/j.apradiso.2017.09.008.

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

Nogueirol, R.C.; De Melo, W.J.; Bertoncini, E.I.; Alleoni, L.R.F. Concentrations of Cu, Fe, Mn, and Zn in tropical soils amended with sewage sludge and composted sewage sludge. *Environ. Monit. Assess.* **2013**, *185*, 2929–2938, doi:10.1007/s10661-012-2761-3.

Oates, K.M.; Kamprath, E.J. Soil acidity and liming: I. Effect of the extracting solution cation and pH on the removal of aluminum from acid soils. *Soil Sci. Soc. Am. J.* **1983**, 47, 686–689, doi:10.2136/sssaj1983.03615995004700040016x.

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

Pavan, M.A.; Bingham, F.T.; Pratt, P.F. Redistribution of Exchangeable Calcium, Magnesium, and Aluminum Following Lime or Gypsum Applications to a Brazilian Oxisol. *Soil Sci. Soc. Am. J.* **1984**, 48, 33–38, doi:10.2136/sssaj1984.03615995004800010006x.

Panchuk, V.; Yaroshenko, I.; Legin, A.; Semenov, V.; Kirsanov, D. Application of chemometric methods to XRF-data—A tutorial review. *Anal. Chim. Acta* **2018**, 1040, 19–32, doi:10.1016/j.aca.2018.05.023.

Potts, P.J.; Webb, P.C.; Watson, J.S. Energy-dispersive x-ray fluorescence analysis of silicate rocks for major and trace elements. *X-Ray Spectrom.* **1984**, 13, 2–15, doi:10.1002/xrs.1300130103.

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

Rodrigues, E.S.; Gomes, M.H.F.; Duran, N.M.; Cassanji, J.G.B.; da Cruz, T.N.M.; Sant'Anna Neto, A.; Savassa, S.M.; de Almeida, E.; Carvalho, H.W.P. Laboratory Microprobe X-Ray Fluorescence in Plant Science: Emerging Applications and Case Studies. *Front. Plant Sci.* **2018**, 9, 1588, doi:10.3389/fpls.2018.01588.

Santos, F.R.; Oliveira, J.F.; Bona, E.; Santos, J.V.F.; Barboza, G.M.; Melquiades, F.L. EDXRF spectral data combined with PLSR to determine some soil fertility indicators. *Microchem. J.* **2020**, 152, 104275, doi:10.1016/j.microc.2019.104275.

Schäefer, C.*E.G.*R.; Fabris, J.D.; Ker, J.C. Minerals in the clay fraction of Brazilian latosols (oxisols): A review. *Clay Miner.* **2008**, 43, 137–154, doi:10.1180/claymin.2008.043.1.11.

Sharma, A.; Weindorf, D.C.; Man, T.; Aldabaa, A.A.A.; Chakraborty, S. Characterizing soils via portable X-ray fluorescence spectrometer: 3, Soil reaction (pH). *Geoderma* **2014**, 232, 141–147, doi:10.1016/j.geoderma.2014.10.001.

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

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

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

Silva, S.H.G.; Teixeira, A.F.D.S.; Menezes, M.D.D.; Guilherme, L.R.G.; Moreira, F.M.D.S.; Curi, N. Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). *Ciênc. Agrotec.* **2017**, 41, 648–664, doi:10.1590/1413-70542017416010317.

Silva, E.A.; Weindorf, D.C.; Silva, S.H.; Ribeiro, B.T.; Poggere, G.C.; Carvalho, T.S.; Goncalves, M.G.; Guilherme, L.R.; Curi, N. Advances in Tropical Soil Characterisation via Portable X-Ray Fluorescence Spectrometry. *Pedosphere* **2019**, 29, 468–482, doi:10.1016/S1002-0160(19)60815-5.

Singh, B.; Gilkes, R.J. Properties and distribution of iron oxides and their association with minor elements in the soils of south-western Australia. *J. Soil Sci.* **1992**, 43, 77–98.

Stenberg, B.; Viscarra Rossel, R.A.; Mouazen, A.M.; Wetterlind, J. Visible and near infrared spectroscopy in soil science. *Adv. Agron.* **2010**, 107, 163–215, doi:10.1016/S0065-2113(10)07005-7.

Taggart, J.E., Jr.; Lindsay, J.R.; Scott, B.A.; Vivit, D.V.; Bartel, A.J.; Stewart, K.C. Analysis of geologic materials by wavelength-dispersive X-ray fluorescence spectrometry. In *Methods for Geochemical Analysis*; US Geological Survey Bulletin: Denver, CO, USA, 1987; Volume 1770, pp. E1–E19.

Tavares, T.R.; Nunes, L.C.; Alves, E.E.N.; Almeida, E.; Maldaner, L.F.; Krug, F.J.; Carvalho, H.W.P.; Molin, J.P. Simplifying sample preparation for soil fertility analysis by X-ray fluorescence spectrometry. *Sensors* **2019**, 19, 5066, doi:10.3390/s19235066.

Tavares, T.R.; Molin, J.P.; Nunes, L.C.; Alves, E.E.N.; Melquiades, F.L.; Carvalho, H.W.P.; Mouazen, A.M. Effect of X-ray tube configuration on measurement of key soil fertility attributes with XRF. *Remote Sens.* **2020**, 12, 963, doi:10.3390/rs12060963.

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

Towett, E.K.; Shepherd, K.D.; Sila, A.; Aynekulu, E.; Cadisch, G. Mid-infrared and total x-ray fluorescence spectroscopy complementarity for assessment of soil properties. *Soil Sci. Soc. Am. J.* **2015**, 79, 1375–1385, doi:10.2136/sssaj2014.11.0458.

USEPA – United States Environmental Protection Agency (2007) M*ethod 6200: Field portable X-ray Fluorescence Spectrometry for the Determination of Elemental Concentrations in Soil and Sediment*. Available online: https://www.epa.gov/sites/production/files/2015-12/documents/6200.pdf (accessed on 25 March 2020).

Van Raij, B.; Andrade, J.C.; Cantarela, H.; Quaggio, J.A. *Análise Química Para Avaliação de Solos Tropicais*; IAC: Campinas, Brazil, 2001; p. 285. (In Portuguese)

Van Raij, B. *Fertilidade do Solo e Manejo de Nutrientes*; International Plant Nutrition Institute (IPNI): Piracicaba, São Paulo, Brazil, 2011; p. 420. (In Portuguese)

Viscarra Rossel, R.A.; Lobsey, C. *Scoping Review of Proximal Soil Sensors for Grain Growing*; CSIRO: Canberra, Australia, 2016.

Weindorf, D.C.; Zhu, Y.; Mc Daniel, P.; Valerio, M.; Lynn, L.; Michaelson, G.; Clark, M.; Ping, C.L. Characterizing soils via portable x-ray fluorescence spectrometer: 2. Spodic and Albic horizons. *Geoderma* **2012**, 189, 268–277, doi:10.1016/j.geoderma.2012.06.034.

Weindorf, D.C.; Chakraborty, S.; Herrero, J.; Li, B.; Castañeda, C.; Choudhury, A. Simultaneous assessment of key properties of arid soil by combined PXRF and V is–NIR data. *Eur. J. Soil Sci.* **2016**, 67, 173–183, doi:10.1111/ejss.12320.

Weindorf, D.C.; Chakraborty, S. Portable X-ray Fluorescence Spectrometry Analysis of Soils. In *Methods of Soil Analysis*; Hirmas, D., Madison, W.I., Eds.; Soil Science Society of America: Wisconsin, WI, USA, 2016; pp. 1–8, doi:10.2136/methods-soil.2015.0033.

Xu, D.; Zhao, R.; Li, S.; Chen, S.; Jiang, Q.; Zhou, L.; Shi, Z. Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. *Eur. J. Soil Sci.* **2019**, 70, 162–173.

Yılmaz, D.; Boydaş, E. The use of scattering peaks for matrix effect correction in WDXRF analysis. *Radiat. Phys. Chem.* **2018**, 153, 17–20, doi:10.1016/j.radphyschem.2018.08.035.

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

Connecting text for chapter 5

Applications of LIBS sensors for soil fertility assessment are even more recent than those of XRF; we are in the early stages of development of this approach in a worldwide context. In chapter 5, we evaluate the performance of LIBS sensor for predicting key fertility attributes in tropical soils. More specifically, we focus on assessing the performance of different spectral pre-processing and modeling techniques in dealing with the large number of spectral variables present in LIBS data (usually around tens of thousands of data points). The results show promising performance of LIBS for evaluating a wide range of fertility attributes in tropical soils. Furthermore, we suggest an accurate and efficient approach to model LIBS data that uses a reduced number of variables.

The study conducted in chapter 5 fulfills part of the first objective of this thesis. The study conducted in this chapter was submitted to Soil & Tillage Journal, as shown below:

Tavares, T.R.; Mouazen, A.M.; Nunes, L.C.; Santos, F.R.; Melquiades, F.L.; Silva, T. R.; Krug, F. J.; Molin, J.P. Laser-Induced Breakdown Spectroscopy (LIBS) for tropical soil fertility analysis. *Soil & tillage* (submitted).

CHAPTER 5. LASER-INDUCED BREAKDOWN SPECTROSCOPY (LIBS) FOR TROPICAL SOIL FERTILITY ANALYSIS

Abstract

The Laser Induced Breakdown Spectroscopy (LIBS) is a promising technique for soil fertility analysis in a rapid and environmentally friendly way. This application requires the selection of an optimal modeling procedure capable of handling the high spectral resolution of LIBS. This work aimed at comparing different modeling methods of LIBS data for the determination of key fertility attributes in Brazilian tropical soils. A benchtop LIBS system was used for the analysis of 102 soil samples, prepared in the form of pressed pellets. Models for the prediction of clay, organic matter, pH, cation exchange capacity, base saturation, and the extractable nutrients P, K, Ca, and Mg were developed using univariate linear regression (ULR), multiple linear regression (MLR) and partial least squares regression (PLS). The following input data for PLS were used: (i) the full spectra from 200 to 540 nm (38,880 variables), and (ii) variables selected by the interval successive projections algorithm (iSPA). The multivariate models achieved satisfactory predictions [residual prediction deviation (RPD) > 1.40] for eight out of nine fertility attributes. However, the best performances were obtained for the PLS with the variable ranges selected by the iSPA, which achieved satisfactory predictions (RPD ≥ 1.44) for seven out of the nine soil attributes studied. The MLR method obtained lower prediction performance than the iSPA-PLS using only 21 variables. The iSPA-PLS approach allowed a reduction from 3 to 160-fold in the total of variables compared to the full LIBS spectra, making it efficient and accurate modeling method that uses reduced number of variables. Although LIBS technique proved to be efficient for predicting fertility attributes in tropical soils, further research is encouraged in order to reduce the amount of sample preparation conducted in this study.

Keywords: 1. Soil fertility testing 2. Proximal soil sensing 3. Matrix effect mitigation 4. Extractable P determination.

5.1. Introduction

Research regarding LIBS applications for soil fertility analysis are still incipient. In temperate soils, the results of the few studies carried out are not always satisfactory. For example, in German agricultural soils, Erler et al. (Erler et al., 2020) have achieved good prediction performances for pH (R^2 between 0.91 and 0.95), reasonable predictions for humus content (R^2 between 0.54 and 0.66), and poor predictions for extractable (ex-) P (R^2 between 0.22 and 0.35). Similarly, in Chinese agricultural soils, Xu et al. (2019a) reported satisfactory predictions for pH, OM, and total N (R^2 between 0.60 and 0.75), and poor performances ($R^2 < 0.35$) for ex-K and ex-P. Also in Chinese soils, studies have shown reasonable predictions for organic matter (OM) ($R^2 = 0.58$) (Xu et al., 2019b), poor prediction performance for ex-K and ex-P, and good prediction for cation exchange capacity (CEC), OM, and pH ($R^2 \geq 0.81$) (Xu et al., 2019c). In Brazilian tropical soils, although good prediction performance ($R^2 \geq 0.85$) have been achieved for textural attributes (Villas-Boas et al., 2016) and pH (Ferreira et al., 2015), evaluation of the prediction

accuracy for other key fertility attributes (*e.g.*, CEC, OM, and extractable nutrients) has not been addressed in the literature yet.

The selection of an optimal modeling procedure for LIBS spectral data, capable of extracting useful but hidden information is crucial for accurate prediction of soil fertility attributes. Although there is no consensus regarding the best modeling strategy to adopt, recent studies suggested using multivariate techniques for the determination of soils elementary constitution when using LIBS spectra, in particular the partial least squares regression (PLS) (Riebe et al., 2019; Takahashi and Thornton, 2017). In addition, multivariate models are a simple and useful strategy to mitigate the matrix effect commonly present in soil samples (Takahashi and Thornton, 2017; Tavares et al., 2020a). Despite of the benefits, multivariate models using the entire spectrum require high computational capacity, especially due to the high spectral resolution of LIBS data that result in a large number of spectral variables (*e.g.*, tens of thousands data points) (Erler et al., 2020). In this regard, algorithms for the selection of most significant variables, such as deep learning or interval successive projection algorithm in partial least squares (iSPA-PLS) (Gomes et al., 2013; Niu et al., 2021; Riebe et al., 2019), can be a useful technique to reduce the amount of LIBS input variables and simplify data acquisition and processing procedures. To our best knowledge, no previous work reported the performance of iSPA-PLS for the prediction of key soil attributes in tropical soils, in comparison with full-spectra based PLS analyses.

This work aimed at evaluating different modeling approaches, namely, multiple linear regression (MLR) using selected emission lines of LIBS data, PLS using the full LIBS spectra from 200 to 540 nm (with 38,880 variables), and PLS using specific spectral regions selected by the iSPA algorithm for the prediction of key soil fertility attributes in Brazilian tropical soils. The best performing model was compared with univariate linear regression (ULR) in order to assess the influence of the matrix effect on the prediction of extractable nutrients (ex-P, ex-Ca, and ex-Mg).

5.2. Material and Methods

5.2.1. Study sites and soil samples

The study sites and soil samples used were described in the 4.2.1. Section (Chapter 4).

5.2.2. Reference analysis

The contents of the key soil fertility attributes clay, OM, CEC, pH, base saturation (V), ex-P, ex-K, ex-Ca, and ex-Mg were determined following the methods described by Van Raij et al. (2001). The

pseudo total content (ptc) of macronutrients (P, K, Ca, and Mg) were determined for comparison with its extractable contents. The determination of macronutrients ptc was performed following the United States Environmental Protection Agency (USEPA) Method 3051A (Element, 2007). Both ptc and soil fertility analyses were detailed in the 4.2.2. Section (Chapter 4).

5.2.3. Soil sample preparation (Pelletezing)

The conformation of soil samples in the form of pellets is of fundamental importance for soil analysis using LIBS (Jantzi et al., 2016). The samples with 10% w w^{-1} of microcrystalline cellulose powder (Sigma-Aldrich, Merck, Darmstadt, Germany) were ground and homogenized using a planetary ball mill (PM 200 mill, Retsch, Haan, Germany), for 20 min (10 cycles of 2 min grinding), and pelletized (15 mm diameter) in a hydraulic press (SPEX 3624B X-Press, Metuchen, EUA), as described elsewhere by Tavares et al. (2019).

5.2.4. XRF measurements and selection of emission lines

Experiments were carried out with a Q-switched Nd:YAG laser (Brilliant, Quantel, France) at 1064 nm, generating 5 ns pulses of up to 365 mJ, in a 6 mm diameter beam, at 10 Hz repetition rate. The laser was focused on the sample pellet by a convergent lens of 2.54 cm in diameter and 20 cm focal length (Newport, USA). Pellets were placed into a plastic sample holder positioned in a two axes manually-controlled translation stage movable in the plane orthogonal to the laser direction. In order to displace the atmospheric air from the sample surface, a laminar stream of argon (5.0 L min^{-1}) was continuously fed from the bottom of the sample holder as described elsewhere by Nunes et al. (2019).

The radiation emitted by the plasma was collected with a telescope composed by 50 mm and 80 mm focal length fused silica lenses (LLA Instruments GmbH, Germany) and coupled to the entrance slit of the spectrometer (model ESA 3000, LLA Instruments GmbH, Germany) with Echelle optics and ICCD detector using an optical fiber (1.5 m, 600 μm core). The collection angle with respect to the laser optical axis was 25°. The monitored spectral range was 200 – 780 nm with a resolution ranging from 5 pm at 200 nm to 19 pm at 780 nm (Figure 1). ESAWIN software (LLA Instruments GmbH, Germany) and NIST Atomic Spectra Database (Ralchenko and Kramida, 2020) were used for identification of the analytes emission lines.

The instrumental conditions including (i) laser pulse energy, (ii) lens-to-sample distance (LTSD), (iii) number of accumulated pulses, (iv) delay time, and (v) integration time gate, were optimized in initial tests in order to obtain the maximum signal-to-noise ratio (SNR) of the emission

lines of interest, in which the ICCD detector did not saturate (Nunes et al., 2009). The best results were obtained with 15 laser pulses, 2.0 μs delay time with a 7.0 μs integration time gate. The lens-to-sample distance and the pulse energy were adjusted at 19.5 cm and 65 mJ, respectively, leading to 250 J cm^{-2} at the sample surface.

In order to minimize drawbacks related to analytes microheterogeneity, the pressed pellets were sampled at 21 different sites, thus yielding 21 optical emission spectra. Approximately 1 mm distance was kept between sites to avoid re-ablation at the edges of the neighboring craters. Data processing was based on triplicate measurements, where seven spectra (corresponding to 7 sampling sites) of each sample were averaged in one representative spectrum.

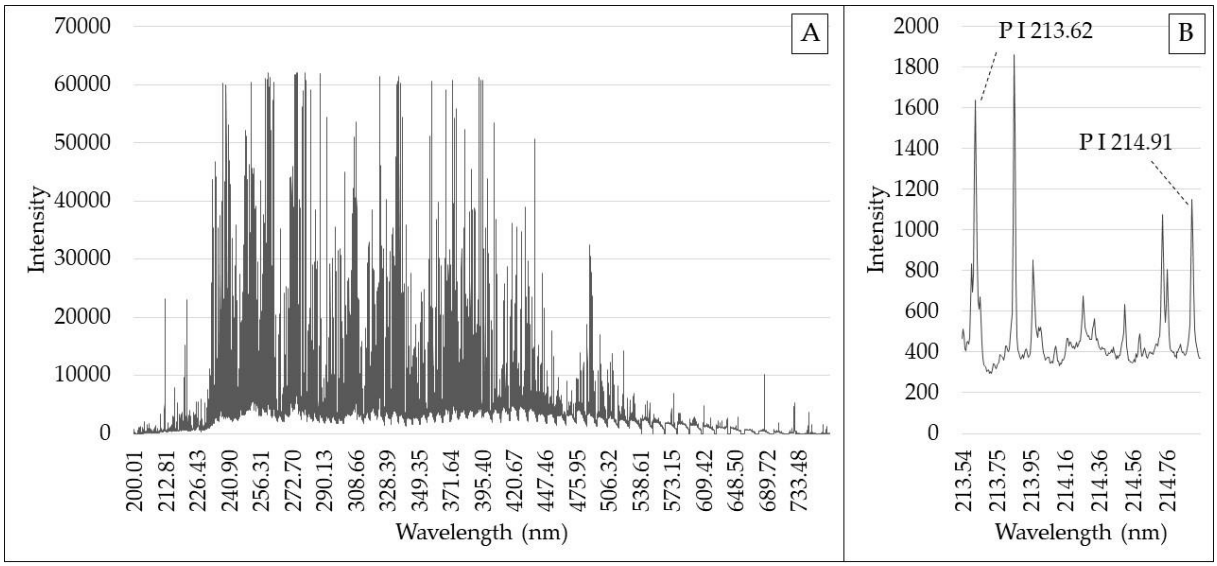


Figure 1. Full Laser Induced Breakdown Spectroscopy (LIBS) spectra obtained from the pellet soil sample (A), and P emission intensities lines in the range between 213.50 and 215.00 nm (B), under the defined experimental conditions.

5.2.5. Data modeling

For the development of the calibration models, 102 soil samples were separated using Kennard Stone algorithm (Kennard and Stone, 1969) into two sets of 68 and 34 samples, which were used for calibration and validation, respectively. Calibration models involving univariate linear regressions (ULR), multiple linear regressions (MLR), and partial least squares regressions (PLS) were built and validated.

Univariate models (ULR) were obtained from the regression of the peak area of the corresponding emission line (after background subtraction) against the reference soil extractable nutrient, by using the classical least squares regression model (Danzer and Currie, 1998). On the basis

of spectral selectivity and sensitivity, P I 214.914 nm, Ca II 315.887 nm, and Mg I 277.983 nm emission lines were selected for the prediction of ex-P, ex-Ca, and ex-Mg, respectively (Nunes et al., 2010; Riebe et al., 2019; Yu et al., 2016). K was not determined since it is below the detection limit of the LIBS technique, under the instrumental conditions used.

For the MLR models, 22 emission lines of elements commonly found in tropical soils (namely, C, Na, Mg, Al, Si, P, Ca, Ti, Mn, Fe, Cu, and Zn) were used as X-variables. Two well-defined emission lines for each of the above-mentioned element were selected. However, for C and Na only one clear emission line in the LIBS spectra was used. Only well-defined emission lines were chosen in order to avoid spectral interference from other elements. Following the mentioned criteria, the emission lines C I 247.856, Na II 330.135, Mg I 277.983, Mg I 285.213, Al I 308.2152, Al I 309.2709, Si I 212.412, Si I 390.552, P I 213.6182, P I 214.914, Ca II 315.887, Ca II 317.933, Ti II 353.541, Ti I 398.176, Mn II 294.920, Mn I 293.306, Fe I 305.908, Fe II 234.349, Cu I 324.754, Cu I 327.390, Zn II 206.200, and Zn I 213.857 nm were selected. No N emission lines were identified in our LIBS spectra.

For the PLS models, two calibration strategies have been performed: (i) using the full spectra within the region from 200 to 540 nm (designated in this study as FS-PLS), which covers most of the valuable and high-intensity emission lines of soil LIBS data (Yu et al., 2016), and (ii) using regions selected through the iSPA (Gomes et al., 2013). For the iSPA, the LIBS spectra from 200 to 540 nm were divided into 160 intervals of 243 variables. The iSPA uses projection operations iteratively to select intervals of variables with minimum collinearity to achieve good prediction ability (Gomes et al., 2013). The number of latent variables (LVs) used in PLS models was selected based on the lowest root mean square error value obtained using 5-fold cross-validation (RMSECV).

The models quality was assessed on the validation set by means of the coefficient of determination (R^2), the root mean square error (RMSE), the relative error (RMSE%), and the residual prediction deviation (RPD). The relative error was calculated by dividing the RMSE of each prediction by the range of the laboratory measured soil property, while the RPD was calculated as the ratio between the standard deviation (SD) of the laboratory measured soil property and the RMSE of its prediction. The RPD interpretations were carried out following the classes proposed by Chang et al. (Chang et al., 2001), in which RPD values below 1.40 indicate bad models, values between 1.40 and 2.00 determine reasonable models, and values above 2.00 indicate good models; being considered as satisfactory, predictions with RPD ≥ 1.40.

5.3. Results

5.3.1. Laboratory measured soil properties

The descriptive statistics and correlation matrix of the laboratory measured soil properties were presented in the 4.3.1.1. Section (Chapter 4).

5.3.2. Relationship between extractable and pseudo total content of nutrients

The scatter plots of Figure 2 show the relationships between ptc and extractable contents of the macronutrients for the individual field and merged Two-field dataset (Field 1 + Field 2) sets. The relationship between ptc and extractable contents are characterized by the R^2 of the regression line, as well as by the ratio between ptc and extractable contents (ptc-ex ratio) observed. In the Two-field dataset, the scatter plots between ptc and extractable contents of K, Ca, and Mg showed a linear dispersion of points with a significantly high R^2 values of 0.81, 0.84, and 0.69, for K, Ca, and Mg, respectively. In contrast, a non-linear behavior with R^2 of 0.01 was observed for P. For the individual Field 1 and Field 2 sets it is possible to observe that all macronutrients show a linear behavior with significant R^2 values. The non-linearity of P observed for the Two-field set is explained by the contrasting ptc ratio presented by each individual field. In other words, P was the only nutrient that showed a contrasting ptc-ex ratio between both fields (oscillating from 16.0 to 72.4 for Field 1, and from 3.8 to 30.4 for Field 2; Figura 2A), whereas for all other nutrients the ptc-ex ratio showed values below 5.6 (Figure 2B, C, D). These linear relations observed for K, Ca and Mg suggest that their extractable fractions can be successfully assessed by LIBS using its elementary information (once their emission line is detected). On the other hand, if satisfactory predictions for the ex-P is obtained, this should be explained by the ability of the prediction model to comprehend the different relationships between ptc and extractable content of P observed for each individual field. Satisfactory predictions for ex-P could also be attributed to the high correlations P has with K Ca, and Mg; however, this is not the case in this study due to the weak correlations of ex-P with the other attributes (r between -0.22 and 0.07).

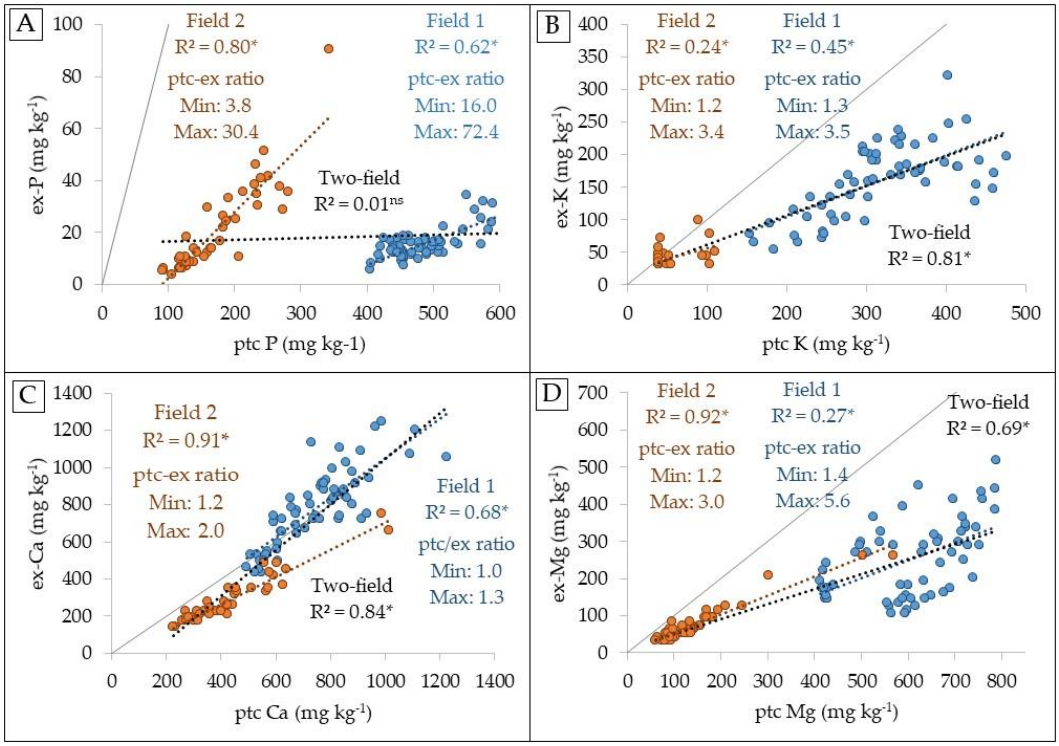


Figure 2. Scatter plots of extractable and pseudo total content (ptc) of P (A), K (B), Ca (C), and Mg (D). The minimum and maximum ratio between ptc and extractable contents (ptc-ex ratio), as well as the coefficient of determination (R^2) of the regression analysis, were presented. Information in blue, orange, and black correspond, respectively, to Field 1, Field 2, and Two-field data. Significant and non-significant regression analysis, at the probability level of 0.05, were marked with "*" and "NS", respectively.

5.3.3. Prediction performance of the different multivariate approaches

The results of the fertility attributes prediction using MLR and PLS are presented in Table 1. Complementing these results, Table 2 shows the performance gain (in percentage of RMSE reduction) of the best performing method compared to the others. In general, the LIBS' prediction performance was satisfactory (RPD ≥ 1.40) for clay, OM, CEC, V, ex-P, ex-K, ex-Ca, and ex-Mg using at least one of the multivariate approaches. Only pH did not show satisfactory predictions (RPD ≤ 1.12). The best prediction performances were obtained for CEC, ex-Ca, and ex-Mg, which had RPD always greater than 2.0, regardless of the multivariate approach applied. Good prediction performances (RPD > 2.0) were also obtained for clay, OM and ex-K prediction, using at least one of the evaluated multivariate methods.

Table 1. Prediction results of the validation (n = 34) obtained from partial least squares regression (PLS) and multiple linear regression (MLR) using LIBS data. The PLS models were built using the LIBS full spectra (FS) and spectral regions selected by the interval successive projection algorithm (iSPA). The MLR models were calibrated using the 21 selected emission lines (EL). The number of latent variables (LV) used in each model was also presented.

Regression approach	Variable selection	Clay	OM[1]	CEC[2]	pH	V[3]	ex-P[4]	ex-K[4]	ex-Ca[4]	ex-Mg[4]
		---------- R^2 ----------								
MLR	21 EL	0.80	0.64	0.88	0.06	0.86	0.59	0.58	0.91	0.89
PLS	FS	0.75	0.76	0.84	0.30	0.94	0.16	0.74	0.91	0.92
	iSPA	0.89	0.81	0.84	0.31	0.94	0.72	0.76	0.94	0.93
		---------- RMSE ----------								
MLR	21 EL	43.88	4.30	10.55	0.43	8.56	7.40	1.51	6.47	5.06
PLS	FS	47.16	2.89	10.75	0.34	5.74	10.69	1.14	6.19	4.16
	iSPA	27.68	2.56	10.69	0.33	5.49	5.84	1.10	4.85	3.51
		---------- RMSE% ----------								
MLR	21 EL	15.24	25.29	11.00	28.60	13.59	14.24	21.52	9.66	11.49
PLS	FS	16.38	18.52	14.95	23.02	10.62	21.47	16.42	9.48	9.45
	iSPA	12.38	15.05	11.14	21.95	8.71	11.23	15.75	7.23	7.97
		---------- RPD ----------								
MLR	21 EL	2.10	1.27	2.48	0.86	2.74	1.44	1.50	3.04	2.53
PLS	FS	1.95	1.89	2.43	1.08	4.07	1.00	1.98	3.19	3.08
	iSPA	3.06	2.14	2.45	1.12	4.27	1.82	2.05	4.07	3.65
		---------- LV ----------								
PLS	FS	4	6	4	3	5	6	2	5	6
	iSPA	5	6	6	4	3	5	4	5	4

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients. The coefficient of determination (R^2) and residual prediction deviation (RPD) values are presented on grayscale, highlighting the highest values. The root-mean-square error (RMSE) was given in g dm^{-3} for clay and OM; in $mmol_c$ dm^{-3} for CEC, ex-K, ex-Ca, and ex-Mg; in % for V; and, for ex-P, the RMSE was given in mg dm^{-3}.

Table 2. Improvement of prediction performance, represented by the reduction of root-mean-square error (RMSE), achieved with the best predictive model compared to the performance reached with the 2nd and 3rd predictive approaches. The partial least squares regression (PLS) models were built using the LIBS full spectra (FS) and spectral regions selected by the interval successive projection algorithm (iSPA). The multiple linear regression (MLR) models were calibrated using the 21 selected emission lines (EL). The presented results were obtained using the validation set (n = 34).

Regression approach	Variable selection	Clay	OM[1]	CEC[2]	pH	V[3]	ex-P[4]	ex-K[4]	ex-Ca[4]	ex-Mg[4]
		---------- RMSE ----------								
MLR	21 EL	43.88	4.30	**10.55**	0.43	8.56	7.40	1.51	6.47	5.06
PLS	FS	47.16	2.89	10.75	0.34	5.74	10.69	1.14	6.19	4.16
	iSPA	**30.03**	**2.56**	10.69	**0.33**	**5.49**	**5.84**	**1.10**	**4.85**	**3.51**
		---------- RMSE improvement of the best approach ----------								
compared to 2nd		-32%	-11%	-1%	-3%	-4%	-21%	-4%	-22%	-16%
compared to 3rd		-36%	-40%	-2%	-23%	-36%	-45%	-27%	-25%	-31%

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients. The RMSE values are presented on grayscale, highlighting the lowest ones. The lowest RMSE was also presented in bold. The RMSE was given in g dm^{-3} for clay and OM; in $mmol_c$ dm^{-3} for CEC, ex-K, ex-Ca, and ex-Mg; in % for V; and, for ex-P, the RMSE was given in mg dm^{-3}.

The comparison of the different multivariate modeling approaches (Table 2) revealed that in general the best performing is the iSPA-PLS models, obtaining the lowest RMSE for all attributes. The only exception was for CEC, whose the lowest prediction error was obtained by using MLR. In any case, MLR reduced the error of CEC prediction by only 1%, compared to the error obtained with iSPA-PLS. For the prediction of OM, pH, V, ex-K, ex-Ca and ex-Mg the sequence of methods giving the best to worst performance was iSPA-PLS > FS-PLS > MLR. For predicting clay and ex-P the sequence of best performing approaches was iSPA-PLS > MLR > FS-PLS, whereas for CEC, the sequence was MLR > iSPA-PLS > FS-PLS. Finally, it is worth noting that the gain in the prediction performance obtained for ex-P ranged from poor performance (RPD = 1.00 and RMSE = 10.69) with FS-PLS, to satisfactory performance with MLR and iSPA-PLS approaches (RPD of 1.44 and 1.82, respectively). This performance gain represented an error reduction of 21 % (RMSE = 7.40) and 45 % (RMSE = 5.84) for MLR and iSPA-PLS, respectively.

The spectral regions as well as the total number of variables used in the iSPA-PLS approach for predicting each soil attribute are presented in Table 3. Compared to the FS-PLS approach, iSPA-PLS allowed a 160-fold reduction in the total of variables for OM, ex-P, ex-K, ex-K and ex-Mg prediction; a 53-fold reduction for clay and pH prediction; and about 3-fold reduction for CEC and V prediction. It is also noteworthy that the variables used as input lie between 201.31 and 416.63 nm, while no variables after 416.63 nm were used.

Table 3. Information about the intervals of variables selected by the interval successive projection algorithm (iSPA), and used as input variables for partial least squares regression (PLS) analyses.

Attribute	Number of intervals selected	Total of variables used	Reduction*	Spectral ranges of the selected intervals (nm)
Clay	3	729	53 times	220.84 – 222.21, 246.77 – 248.29 and 251.37 – 252.94
OM[1]	1	243	160 times	295.08 – 296.92
CEC[2]	52	12,636	3 times	201.31 – 272.37, 287.89 – 298.73 and 414.15 – 416.63
pH	3	729	53 times	220.84 – 222.21, 237.82 – 239.27 and 391.78 – 394.22
V[3]	57	13,851	3 times	201.31 – 272.37, 279.18 – 300.60 and 396.59 – 399.10
ex-P[4]	1	243	160 times	214.14 – 215.44
ex-K[4]	1	243	160 times	375.28 – 377.53
ex-Ca[4]	1	243	160 times	389.37 – 391.77
ex-Mg[4]	1	243	160 times	284.36 – 286.14

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients. The variable importance of projection (VIP) scores of each interval of variables are shown in Figure A1 (Appendix Section). * Reduction in the number of variables used in the iSPA-PLS models when compared to the number of variables present in the range between 200 and 540 nm (n = 38,880 variables), which was used in the SF-PLS approach.

5.3.4. Matrix effect: comparison of univariate and multivariate approaches for extractable nutrients prediction

Figure 3 shows the scatter plots of measured and predicted values of ex-P, ex-Ca, and ex-Mg using ULR and the multivariate iSPA-PLS models. Univariate models for ex-P and ex-Ca (Figure 3A,C) showed poor performance (RPD ≤ 1.16), which is attributed to matrix effect, originated from sample physicochemical properties. The multivariate predictive models (Figure 3B,D,F) demonstrate less matrix-matching dependent, which showed markedly higher performances (RPD of 1.82, 4.07, and 6.65, for ex-P, ex-Ca, and ex-Mg, respectively) than the URL. For the multivariate models, points are closely distributed around the 1:1 line. The same trend is also seen for the other multivariate approaches (MLR and FS-PLS) in Figure A2 (Appendix Section). It is also noteworthy that the matrix effect for the prediction of ex-Mg was less significant than that observed for ex-P and ex-Ca, since ex-Mg provided very good prediction result (RPD = 2.75) with the ULR model, showing linear behavior with points closely spread around the 1:1 line (Figure 3E). Nevertheless, the the iSPA-PLS model showed clearly superior performance (RPD = 3.65) and much closer points dispersion to the 1:1 line, when compared to the ULR.

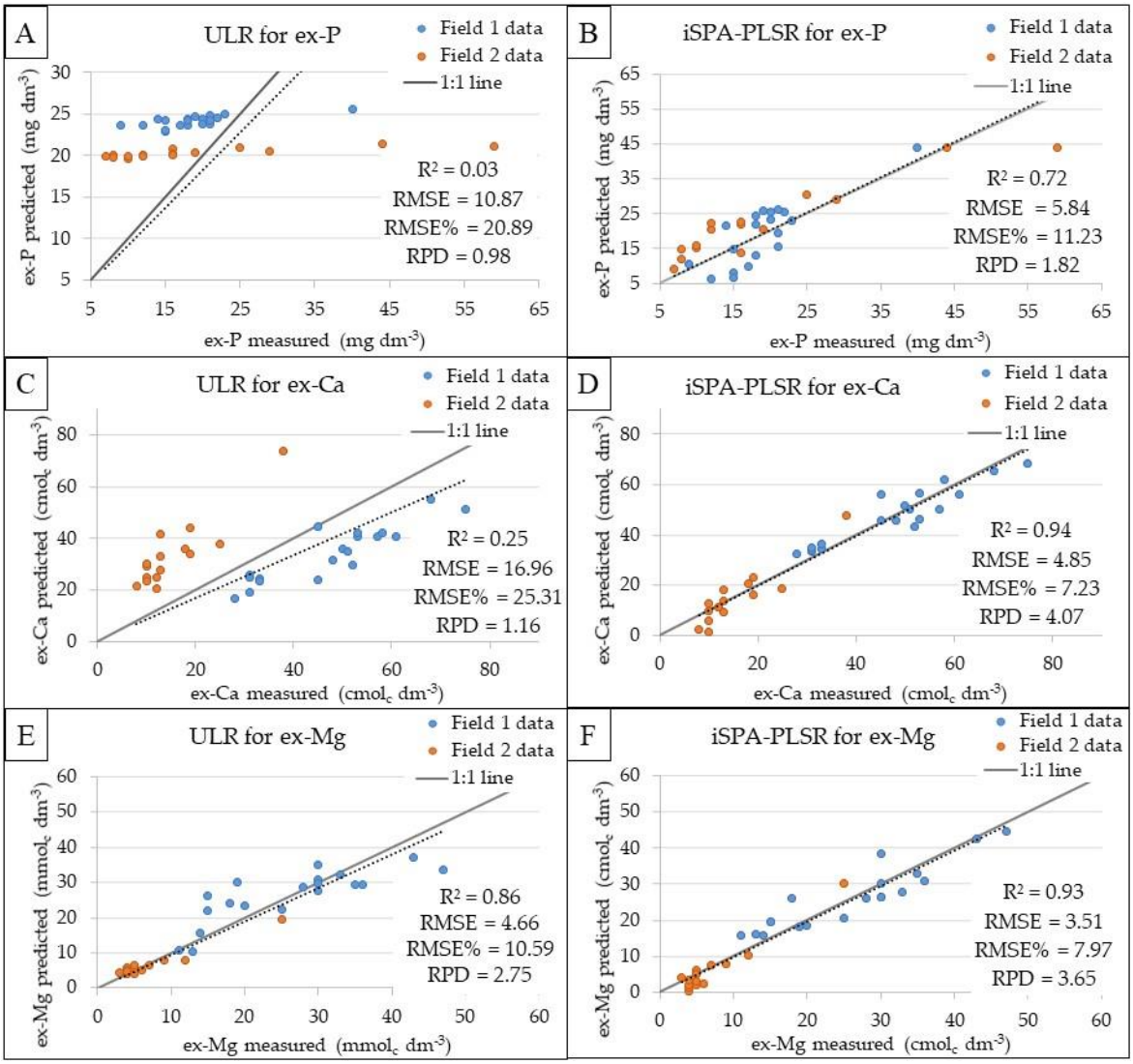


Figure 3. Scatter plots of measured versus predicted extractable (ex-) P (A, B), Ca (C, D), and Mg (E, F). The A,C, and E plots were obtained using univariate linear regression (ULR), whereas the B,D, and F plots were obtained with partial least squares regression (PLS) after interval variables selection by the interval successive projection algorithm (iSPA). The figures concerns the prediction performance of the models applied on the validation dataset (n = 34).

5.4. Discussion

5.4.1. Predicting key soil fertility attributes using LIBS

The LIBS spectra obtained from the pelletized soil samples allowed satisfactory prediction results ($0.59 \leq R^2 \leq 0.94$) using multivariate models developed for eight out of nine fertility attributes, although pH was the only attribute that showed poor prediction performance ($R^2 \leq 0.31$). Soil fertility attributes are indirectly predicted via elemental analysis sensors due to relationship existing between such attributes and the elemental constitution of the soil samples. Textural attributes are often related

to the content of Fe, Si, Al, Ti, among others, which are structural components of the different particle size fractions (Lima et al., 2019; Villas-Boas et al., 2016). The results obtained for clay prediction (R^2 = 0.89) were close to those obtained by Villas-Boas et al. (Villas-Boas et al., 2016) (R^2 = 0.83), working with a LIBS sensor for the analysis of tropical soils. Ferreira et al. (Ferreira et al., 2015) obtained successful models for pH (R^2 = 0.87) using Al, Ca, H, and O emission lines collected for tropical soils. However, the prediction results for pH were poor in our study. Satisfactory predictions of OM using LIBS were reported using C or N emission lines (Senesi and Senesi, 2016), or indirectly, through relationships with textural attributes (Lima et al., 2019). The performance of OM predictions obtained in this work oscillated according to the modeling approach used ($0.64 \leq R^2 \leq 0.81$), although this range of accuracy was similar that reported in studies conducted on Chinese soils that showed R^2 oscillating between 0.58 and 0.83 (Xu et al., 2019a; Xu et al., 2019b, 2019c). Although satisfactory predictions of CEC and V using LIBS have not yet been reported in the literature, good predictive models ($R^2 \geq 0.75$) can also be achieved (as achieved in this work), once they present covariations with well-predicted fertility attributes (*e.g.*, textural attributes, extractable nutrients, etc), as reported by Tavares et al. (2020b) and Silva et al. (2017), both using XRF sensors.

Regarding the extractable nutrients, the correlation between extractable and total contents is the main reason to explain successful predictions using the elemental concentration of a given nutrient. In this work, ptc and extractable contents of K, Ca and Mg showed significant correlations (Figure 2) in both the individual and Two-field datasets. Thus, the ex-Ca and ex-Mg predictions were very successful, obtaining R^2 value higher than 0.88. On the other hand, ex-K predictions, although satisfactory, presented inferior prediction performance ($0.58 \leq R^2 \leq 0.74$), which is attributed to the absence of identifiable K emission lines in the LIBS spectrum, and the level of accuracy obtained may well be due to its covariation with other fertility attributes, *e.g.*, clay (r = 0.73), ex-Ca (r = 0.68), and ex-Mg (r = 0.58). It is interesting to mention that due to the weathered nature of tropical soils, their composition are generally poor in primary minerals that are natural source of nutrients (Fontes, 2012). Consequently, the relationships between the total and extractable contents of some nutrients (*e.g.*, ex-Ca, and ex-Mg) are largely influenced by agricultural management (*e.g.*, fertiliser inputs) rather than natural processes (Andrade et al., 2020).

The soil chemistry regarding P in tropical soils is complex and characterized by strong association with Fe oxides (hematite and goethite) and Al (gibbsite). Due to this complexity, it is common to observe unbalanced relationships between its total content present in the soil and the amount of ex-P (Schäefer et al., 2008). This behavior was observed for P when merging the data from both fields (Two-field dataset). For this reason, we were not expecting satisfactory predictions for this nutrient via LIBS spectra using the Two-field dataset. On the other hand, in the individual field datasets,

ex-P showed a significant correlation with its ptc (Figure 2A). This is what would explain our satisfactory predictions for ex-P when using both multivariate iSPA-PLS and MLR models, and not when using the univariate model. In other words, the multivariate models allowed to correct the different ratios between total and extractable contents present in both individual fields, resulting in satisfactory models for ex-P (R^2 of 0.59 and 0.72 using MLR and iSPA-PLS, respectively).

Due to the indirect nature of predictions using LIBS sensors, it is important that further investigations address the temporal stability of the models' performance. Understanding the LIBS' performance for soil fertility prediction as a function of time/soil management will be fundamental to advance our knowledge regarding: (i) understanding and mitigating factors that influence sensor performance and (ii) developing spectral libraries of LIBS data for soil analysis.

5.4.2. Multivariate approaches, number of variables, and matrix effect mitigation

Using the iSPA algorithm, for the variables selection, combined with PLS showed to be the best method for predicting fertility attributes via LIBS sensor. iSPA-PLS provided accurate predictions (RPD ≥ 1.82) for eight out of the nine attributes (namely, clay, OM, CEC, V, ex-P, ex-K, ex-Ca, and ex-Mg), by employing a lower number of input variables than those used in the FS-PLS approach (Table 3). Although MLR generally underperformed the iSPA-PLS, its performance was satisfactory (RPD between 1.44 and 3.04) for the prediction of clay, CEC, V, ex-P, ex-K, ex-Ca, and ex-Mg by using 21 selected emission lines. Thus, MLR can also be considered a simple alternative method for predicting fertility attributes when drastic reductions in the number of variables are required. Reducing the number of input variables of LIBS data is necessary because their spectra contain a large number of variables, most of them are irrelevant for predicting fertility attributes. High volumes of data can negatively affect the performance of model calibration, as well as creates an unnecessary computational burden (Erler et al., 2020).

Although iSPA-PLS proved to be the best approach for predicting fertility attributes and mitigating the matrix effect for P, Ca, and Mg, all the other multivariate models also mitigated the matrix effect in some degree and achieved more accurate predictions than ULR models (Figure A2). The matrix effect in soil samples are associated with differences in material matrices, such as hardness, chemical composition, and density, which affects the intensities of the elemental emission lines (Senesi, 2020). The mitigation of the matrix effect with multivariate models can be explained by the fact that chemical and physical matrix effects are shown as elements' peaks. Therefore, multivariate regressions can extract this information in the spectra without requiring for prior assumptions of the plasma condition, as it is commonly used *e.g.*, in calibration-free LIBS (Takahashi and Thornton, 2017). The correction of the

matrix effect via multivariate approaches has also been reported by Tavares et al. (2020b) in analysis of fertility attributes using XRF sensors, which is also a matrix dependent technique.

5.4.3. The challenge of sample preparation

It is important to highlight operational aspects of sample preparation that are little discussed in scientific works aiming at using LIBS sensors as a practical method of soil analysis. Although some studies have evaluated the performance of LIBS sensors using non-pelletized soil samples (intact soil core) (Bricklemyer et al., 2013, 2018), most published studies used soil samples conformed in the form of pellets (Ferreira et al., 2015; Villas-Boas et al., 2016; Yu et al., 2016; Riebe et al., 2019; Xu et al., 2019a; Xu et al., 2019b, 2019c; Erler et al., 2020). This is because the pelletizing is of paramount importance before the data acquisition, as it improves homogeneity, as well as allow congruent ablations by laser, maintaining the stoichiometric proportion of the craters formed at each pulse, which is crucial for the replicability of the calibrated models (Jantzi et al., 2016). Quantitative analyses on loose powder soil samples would have drastically reduced performance compared to analyses on pelletized samples. In addition, although soil pelletizing does not require a reagent and is apparently simple to perform, soil comminution and pressing requires specific equipment (*e.g.*, mill and press), which adds costs and operational complexity to the method. Moreover, particle size reduction in soils containing different portions of quartz (common mineral in the sand fraction) require the use of mills with compatible hardness (*e.g.*, tungsten carbide), which are relatively more expensive and present extra challenges such as: (i) possibility of contamination of the sample with the milled material (Tavares et al., 2019), (ii) higher wear of the equipment. Thus, the afore-mentioned aspects bring into question the practicality of LIBS analysis when using pelletized soil samples. In this sense, it is important that further research needs to develop new solutions that avoid LIBS deteriorated performance when applied in non-pelletized soil samples. This is critical issue to be overcome for the expansion of LIBS as a practical soil analysis within the areas of precision agriculture and soil science.

5.5. Conclusion

LIBS technique proved to be efficient for predicting fertility attributes in tropical soils. The LIBS spectra obtained from the pelletized soil samples associated with the multivariate models achieved satisfactory predictions (RPD > 1.40) for eight out of the nine key soil fertility attributes, with pH being the only exception that showed poor predictive performance (RPD ≤ 1.12). The best prediction performances were obtained for CEC, ex-Ca, and ex-Mg, which had RPD always greater than 2.0,

regardless of the multivariate approach applied. Good performances (RPD > 2.0) were also obtained for clay, OM and ex-K prediction, using at least one of the evaluated multivariate methods.

Comparing the different multivariate approaches, multiple linear regressions (MLR) in general showed slightly lower performance than iSPA-PLS, it allowed satisfactory predictions (RPD between 1.44 and 3.04) for clay, CEC, V, ex-P, ex-K, ex-Ca, and ex-Mg by employing only 21 variables. Anyhow, the best performing models were those obtained after the interval successive projection algorithm in combination with partial least squares regression (iSPA-PLS). The iSPA-PLS allowed reducing the amount of variables by 160 times for OM, ex-P, ex-K, ex-K and ex-Mg, by 53 times, for clay and pH and by approximately 3 times for CEC and V, compared with the amount of variables in the full LIBS spectra from 200 to 540 nm (38,880 variables). The iSPA-PLS is an efficient and accurate modeling method that uses reduced number of variables to predict fertility attributes via LIBS sensor.

Due to the highlighted challenges related to the pelletizing of soil samples further research is encouraged to find solutions to allow LIBS applications in soil samples with a simpler sample preparation procedure. Overcoming this limitation would expand the possibilities of LIBS sensors application in precision agriculture and soil science.

Appendix

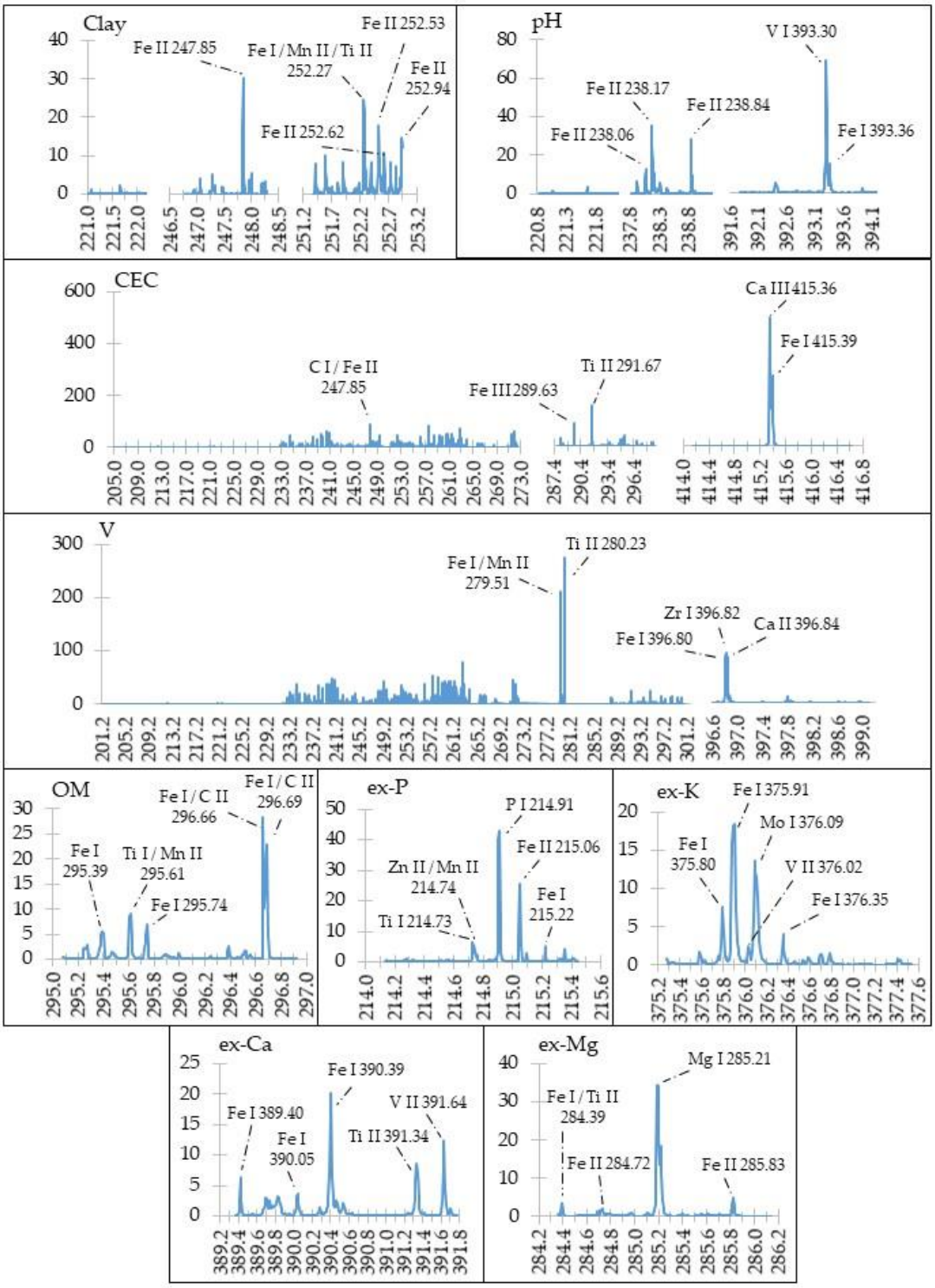


Figure A1. Variable importance of projection (VIP) scores of the interval of variables selected by the interval successive projection algorithm (iSPA) for the prediction of the key soil fertility attributes. The X-axis of the graphs represent the wavelengths (expressed in nm) of the LIBS spectrum and the Y-axis represents the VIP scores. The top five highest contributing emission lines to the prediction of each soil attribute were also identified.

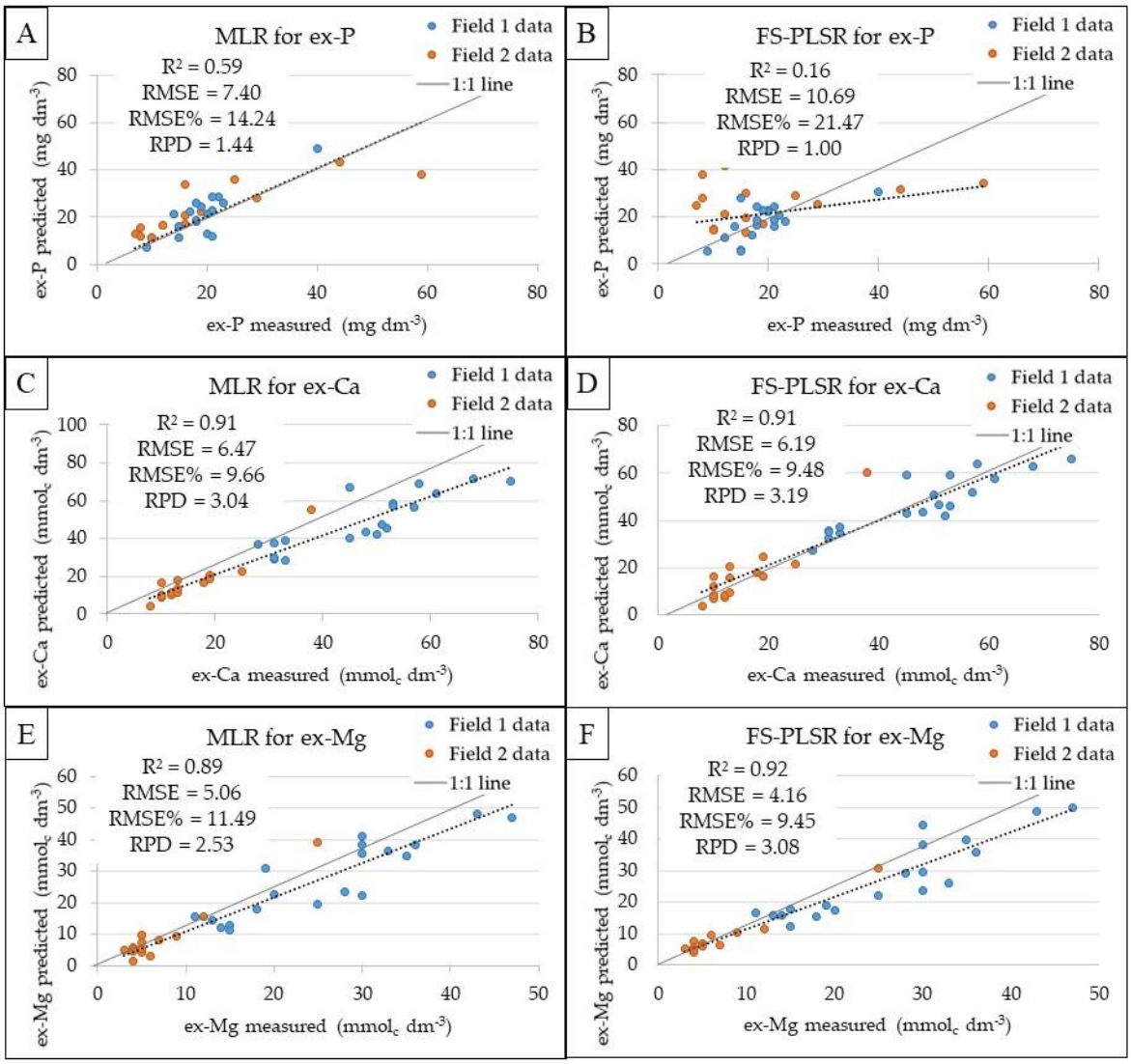


Figure A2. Scatter plots of measured versus predicted extractable (ex-) P (A, B), Ca (C, D), and Mg (E, F). In the left side (A,C,E), the models were calibrated using multiple linear regression (MLR) calibrated using the 21 selected emission lines. In the right side (B,D,F), the models were calibrated with partial least square regression using the spectra within the region from 200 to 540 nm (FS-PLS). The figures of merit presented correspond to the prediction performance of the models applied on the validation dataset (n = 34).

References

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

Bricklemyer, R. S.; Brown, D. J.; Turk, P. J.; Clegg, S. Comparing VNIRS, LIBS, and Combined VNIRS-LIBS for Intact Soil Core Soil Carbon Measurement. *Soil Science Society of America Journal* **2018**, 82(6), 1482–1496. https://doi.org/10.2136/sssaj2017.09.0332

Bricklemyer, R. S.; Brown, D. J.; Turk, P. J.; Clegg, S. M. Improved intact soil-core carbon determination applying regression shrinkage and variable selection techniques to complete spectrum laser-induced breakdown spectroscopy (LIBS). *Applied Spectroscopy* **2013**, 67(10), 1185–1199. https://doi.org/10.1366/12-06983

Chang, C.W.; Laird, D. A.; Mausbach, M. J.; Hurburgh, C. R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal* **2001**, 65(2), 480–490. https://doi.org/10.2136/sssaj2001.652480x

Danzer, K.; Currie, L. A. Guidelines for calibration in analytical chemistry. Part I. Fundamentals and single component calibration (IUPAC Recommendations 1998). *Pure and Applied Chemistry* **1998**, 70(4), 993–1014. https://doi.org/10.1351/pac199870040993

Element, C. A. S. (2007). Method 3051A microwave assisted acid digestion of sediments, sludges, soils, and oils. Z.Für Anal.Chem, 111, 362–366.

Erler, A.; Riebe, D.; Beitz, T.; Löhmannsröben, H.-G.; Gebbers, R. Soil Nutrient Detection for Precision Agriculture Using Handheld Laser-Induced Breakdown Spectroscopy (LIBS) and Multivariate Regression Methods (PLSR, Lasso and GPR). *Sensors* **2020**, 20(2), 418. https://doi.org/10.3390/s20020418

Ferreira, E. C.; Gomes Neto, J. A.; Milori, D. M. B. P.; Ferreira, E. J.; Anzano, J. M. Laser-induced breakdown spectroscopy: Extending its application to soil pH measurements. *Spectrochimica Acta - Part B Atomic Spectroscopy* **2015**, 110, 96–99. https://doi.org/10.1016/j.sab.2015.06.002

Fontes, M. P. F. (2012). Intemperismo de rochas e minerais. In J. C. Ker, N.; Curi, C. E. G. R.; Schaefer, P.; Vidal-Torrado (Eds.), Pedologia: Fundamentos (pp. 171–205). Sociedade Brasileira de Ciência do Solo.

Gomes, A. D. A.; Galvão, R. K. H.; De Araújo, M. C. U.; Véras, G.; Da Silva, E. C. The successive projections algorithm for interval selection in PLS. *Microchemical Journal* **2013**, 110, 202–208. https://doi.org/10.1016/j.microc.2013.03.015

Jantzi, S. C.; Motto-Ros, V.; Trichard, F.; Markushin, Y.; Melikechi, N.; De Giacomo, A. Sample treatment and preparation for laser-induced breakdown spectroscopy. *Spectrochimica Acta - Part B Atomic Spectroscopy* **2016**, 115, 52–63. https://doi.org/10.1016/j.sab.2015.11.002

Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, 11(1), 137–148. https://doi.org/10.1080/00401706.1969.10490666

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

Niu, C.; Cheng, X.; Zhang, T.; Wang, X.; He, B.; Zhang, W.; Feng, Y.; Bai, J.; Li, H. Novel Method Based on Hollow Laser Trapping-LIBS-Machine Learning for Simultaneous Quantitative Analysis of Multiple Metal Elements in a Single Microsized Particle in Air. *Analytical Chemistry* **2021**, 93(4), 2281–2290. https://doi.org/10.1021/acs.analchem.0c04155

Nunes, L. C.; Rocha, F. R. P.; Krug, F. J. Slope ratio calibration for analysis of plant leaves by laser-induced breakdown spectroscopy. *Journal of Analytical Atomic Spectrometry* **2019**, 34(11), 2314–2324. https://doi.org/10.1039/C9JA00270G

Nunes, L. C.; Braga, J. W.; Trevizan, L. C.; Souza, P. F.; Carvalho, G. G. A.; Júnior, D. S.; Poppi, R. J.; Krug, F. J. Optimization and validation of a LIBS method for the determination of macro and micronutrients in sugar cane leaves. *Journal of Analytical Atomic Spectrometry* **2010**, 25(9), 1453–1460. https://doi.org/10.1039/c003620j

Nunes, L. C.; Silva, G. A.; Trevizan, L. C.; Santos Júnior, D., Poppi, R. J.; Krug, F. J. Simultaneous optimization by neuro-genetic approach for analysis of plant materials by laser induced breakdown spectroscopy. *Spectrochimica Acta - Part B Atomic Spectroscopy* **2009**, 64(6), 565–572. https://doi.org/10.1016/j.sab.2009.05.002

Ralchenko, Y.; Kramida, A. Development of NIST Atomic Databases and Online Tools. *Atoms* **2020**, 8. https://doi.org/10.3390/atoms8030056

Riebe, D.; Erler, A.; Brinkmann, P.; Beitz, T.; Löhmannsröben, H. G.; Gebbers, R. Comparison of calibration approaches in laser-induced breakdown spectroscopy for proximal soil sensing in precision agriculture. *Sensors* (Switzerland) **2019**, 19(23), 1–15. https://doi.org/10.3390/s19235244

Schäefer, C. E. G. R.; Fabris, J. D.; Ker, J. C. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. *Clay Minerals* **2008**, 43(1), 137–154. https://doi.org/10.1180/claymin.2008.043.1.11

Senesi, G. S. A Survey of Recent Applications of Laser-induced Breakdown Spectroscopy (LIBS) to Soil Analysis. *Int. J. Earth Environ. Sci.* **2020**, 5, 172. https://doi.org/10.15344/2456-351X/2020/172

Senesi, G. S.; Senesi, N. Laser-induced breakdown spectroscopy (LIBS) to measure quantitatively soil carbon with emphasis on soil organic carbon. A review. *Analytica Chimica Acta* **2016**, 938, 7–17. https://doi.org/https://doi.org/10.1016/j.aca.2016.07.039

Silva, S. H. G.; Teixeira, A. F. dos S.; Menezes, M. D. de, Guilherme, L. R. G.; Moreira, F. M. de S.; Curi, N.; Silva, S. H. G.; Teixeira, A. F. dos S.; Menezes, M. D.; Guilherme, L. R. G.; Moreira, F. M. S.; Curi, N. Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). *Ciência e Agrotecnologia* **2017**, 41(6), 648–664. https://doi.org/10.1590/1413-70542017416010317

Takahashi, T.; Thornton, B. Quantitative methods for compensation of matrix effects and self-absorption in LIBS signals of solids. *Spectrochimica Acta - Part B Atomic Spectroscopy* **2017**, 138, 31–42. https://doi.org/10.1016/j.sab.2017.09.010

Tavares, T.R.; Nunes, L. C.; Alves, E. E. N.; Almeida, E.; Maldaner, L. F.; Krug, F. J.; Carvalho, H. W. P.; Molin, J. P. Simplifying sample preparation for soil fertility analysis by x-ray fluorescence spectrometry. *Sensors* (Switzerland) **2019**, 19(23). https://doi.org/10.3390/s19235066

Tavares, T. R.; Mouazen, A. M.; Alves, E. E. N.; Santos, F. R.; Melquiades, F. L.; Carvalho, H. W. P.; Molin, J. P. Assessing soil key fertility attributes using a portable X-ray fluorescence: A simple method to overcome matrix effect. *Agronomy* **2020a**, 10(6). https://doi.org/10.3390/agronomy10060787

Tavares, T. R.; Molin, J. P.; Nunes, L. C.; Alves, E. E. N.; Melquiades, F. L.; Carvalho, H. W. P.; Mouazen, A. M. Effect of x-ray tube configuration on measurement of key soil fertility attributes with XRF. *Remote Sensing* **2020b**, 12(6). https://doi.org/10.3390/rs12060963

Van Raij, B.; Andrade, J. C.; Cantarela, H.; Quaggio, J. A. *Análise química para avaliação de solos tropicais*. IAC: Campinas, Brazil, 2001; 285p.

Villas-Boas, P. R.; Romano, R. A.; Franco, M. A. M.; Ferreira, E. C.; Ferreira, E. J.; Crestana, S.; Milori, D. M. B. P. Laser-induced breakdown spectroscopy to determine soil texture: A fast analytical technique. *Geoderma* **2016**, 263, 195–202. https://doi.org/10.1016/j.geoderma.2015.09.018

Xu, D.; Zhao, R.; Li, S.; Chen, S.; Jiang, Q.; Zhou, L.; Shi, Z. Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. *European Journal of Soil Science* **2019a**, 70(1), 162–173. https://doi.org/10.1111/ejss.12729

Xu, X.; Du, C.; Ma, F.; Shen, Y.; Wu, K.; Liang, D.; Zhou, J. Detection of soil organic matter from laser-induced breakdown spectroscopy (LIBS) and mid-infrared spectroscopy (FTIR-ATR) coupled with multivariate techniques. *Geoderma* **2019b**, 355(January). https://doi.org/10.1016/j.geoderma.2019.113905

Xu, X.; Du, C.; Ma, F.; Shen, Y.; Zhou, J. Fast and simultaneous determination of soil properties using laser-induced breakdown spectroscopy (Libs): A case study of typical farmland soils in china. *Soil Systems* **2019c**, 3(4), 1–18. https://doi.org/10.3390/soilsystems3040066

Yu, K. Q.; Zhao, Y. R.; Liu, F.; He, Y. Laser-Induced Breakdown Spectroscopy Coupled with Multivariate Chemometrics for Variety Discrimination of Soil. *Scientific Reports* **2016**, 6(June), 1–10. https://doi.org/10.1038/srep27574

Connecting text for chapter 6

The chapter 6 is the first scientific work in a worldwide context to assesss the combined use of VNIR, XRF, and LIBS sensors for a complete characterisation of key fertility attributes in agricultural soils. These sensors are complementary for the characterisation of soil constituents and might provide more comprehensive information about the key soil fertility attributes. In addition, there is still no consensus on an optimal set of sensors to assess soil fertility, nor on an effective data fusion approach to explore the synergy among soil sensors. The present chapter highlights benefits and limitations of the in tandem application of VNIR, XRF, and LIBS spectroscopies for fertility analysis in tropical soils.

Chapter 6 fulfills the second objective of this thesis, which proposed to evaluate the combined use of the above mentioned sensors through data fusion strategies for predicting soil fertility attributes. The main findings of chapter 6 were published in two original articles, one in the Sensors Journal and the second in the Agronomy Journal, as shown below:

Tavares, T. R., Molin, J. P., Javadi, S. H., Carvalho, H. W. P. D., Mouazen, A. M. Combined Use of VNIR and XRF Sensors for Tropical Soil Fertility Analysis: Assessing Different Data Fusion Approaches. *Sensors* **2021**, 21(1), 148, doi: 10.3390/s21010148.

Tavares, T. R., Molin, J. P., Nunes, L. C., Wei, M. C. F., Krug, F. J., Carvalho, H. W. P. D., Mouazen, A. M. Multi-sensor approach for tropical soil fertility analysis: comparison of individual and combined performance of VNIR, XRF, and LIBS spectroscopies. *Agronomy* **2021** (accepted)

CHAPTER 6. MULTI-SENSOR APPROACH FOR TROPICAL SOIL FERTILITY ANALYSIS: COMPARISON OF INDIVIDUAL AND COMBINED PERFORMANCE OF VNIR, XRF, AND LIBS SPECTROSCOPIES

Abstract

Rapid, cost-effectively and environment-friendly analysis of key soil fertility attributes requires an ideal combination of sensors. The individual and combined performance of visible and near infrared diffuse reflectance spectroscopy (VNIR), X-ray fluorescence spectroscopy (XRF), and laser induced breakdown spectroscopy (LIBS) was assessed for predicting clay, organic matter (OM), cation exchange capacity (CEC), pH, base saturation (V), and extractable (ex-) nutrients in tropical soils. A set of 102 samples, collected from two agricultural fields, with broad ranges of fertility attributes were selected. Two contrasting data fusion approaches have been applied for modeling: (i) merging spectral data of different sensors followed by partial least squares regression (PLS), known as fusion before prediction; and (ii) applying the Granger and Ramanathan (GR) averaging approach, known as fusion after prediction. Results showed VNIR as individual technique to be the best for the prediction of clay and OM content [2.61 ≤ residual prediction deviation (RPD) ≤ 3.37], while the chemical attributes CEC, V, ex-P, ex-K, ex-Ca, and ex-Mg were better predicted (1.82 ≤ RPD ≤ 4.82) by elemental analysis techniques (*e.g.*, XRF and LIBS). Only pH cannot be predicted regardless the technique. The attributes OM, V, and ex-P were best predicted using single-sensor approaches, while the attributes clay, CEC, pH, ex-K, ex-Ca, and ex-Mg were overall best predicted using multi-sensor approaches. Regarding the performance of the multi-sensor approaches, ex-K, ex-Ca and ex-Mg, were best predicted (RPD of 4.98, 5.30, and 4.11 for ex-K, ex-Ca and ex-Mg, respectively) using two-sensor fusion approach (VNIR + XRF for ex-K and XRF + LIBS for ex-Ca and ex-Mg), while clay, CEC and pH were best predicted (RPD of 4.02, 2.63, and 1.32 for clay, CEC and pH, respectively) with the three-sensor fusion approach (VNIR + XRF + LIBS). Therefore, the best combination of sensors for predicting key fertility attributes proved to be attribute-specific, which is a drawback of the data fusion approach. The present work is pioneering in highlighting benefits and limitations of the in tandem application of VNIR, XRF and LIBS spectroscopies for fertility analysis in tropical soils.

Keywords: 1. Data fusion 2. Proximal soil sensing 3. Hybrid laboratory 4. Soil testing

6.1. Introduction

Visible and near infrared diffuse reflectance spectroscopy (VNIR), X-ray fluorescence spectroscopy (XRF), and laser induced breakdown spectroscopy (LIBS) are multi-source techniques compatible with direct analysis of soil samples. The visible part of VNIR spectra is sensitive to valence electron excitations of some atoms and functional groups, while the near infrared part responds to non-fundamental vibrations of specific molecules (Stenberg et al., 2010). In soil analysis, the spectral features present in the VNIR region are strictly related to mineralogical and organic components (Clark and Roush, 1984). In turn, elemental analysis techniques, as XRF and LIBS, provide the characterisation of the elemental constitution of a wide variety of chemical elements present in soil samples (Krug and Rocha, 2016). XRF and LIBS are also considered complementary (Krug and Rocha, 2016). For example,

elements with atomic number (Z) lower than 12 are hardly detected by XRF technique; however, elements such C (Z=6) and Na (Z=11) can be detected by LIBS. Hence, the study conducted in this chapter hypothesizes that the combination of data provided by VNIR sensors with those from elemental analysis sensors, *e.g.* XRF and LIBS, might provide complete characterisation of key soil fertility attributes (Castanedo, 2013; Mahmood et al., 2012).

Although VNIR, XRF, and LIBS techniques have been evaluated individually for predicting fertility attributes (Munnaf et al., 2019; Andrade et al., 2020; Erler et al., 2020; Riebe et al., 2019; Romero et al., 2018; Tavares et al., 2020a), the combination of these techniques is still at its early stages of development. Recent studies have evaluated different data fusion approaches for combining VNIR and XRF data for predicting fertility attributes (Javadi et al., 2021; Wan et al., 2020; Zhang and Hartemink, 2020). To the best of our knowledge, only Xu et al. (2019a) combined VNIR, XRF, and LIBS sensors for predicting soil fertility attributes [organic matter (OM), pH, extractable (ex-) P, ex-K, ex-N, and total N]. Although these studies show the synergy among these techniques, they do not make a comprehensive evaluation of all the key soil fertility attributes. Moreover, the combination of these three sensors have never been evaluated for the analysis of key fertility attributed in tropical soils.

Therefore, the present study aimed at assessing the performance of the individual and combined use of VNIR, XRF and LIBS sensors for predicting key soil fertility attributes in tropical soils. Two contrasting approaches of data fusion modeling has been applied: (i) a front-end approach that combines the raw data of each sensor, followed by partial least squares (PLS); and (ii) a back-end approach, that applies an adapted approach from the Granger and Ramanathan (GR) averaging method. While the former method can be called fusion before prediction, the latter can be designated as fusion after prediction.

6.2. Material and Methods

6.2.1. Study sites and soil samples

The study sites and soil samples used were described in the 4.2.1. Section (Chapter 4).

6.2.2. Reference analyses

The contents of the key soil fertility attributes clay, OM, cation exchange capacity (CEC), pH, base saturation (V), ex-P, ex-K, ex-Ca, and ex-Mg were determined following the methods described by Van Raij et al. (2001), which were detailed in the 4.2.2. Section (Chapter 4).

6.2.3. Sample preparation

The data acquisition with XRF and VNIR sensors was performed on dry and sieved samples with 2 mm particle size. For the XRF measurements, about ten grams of each loose soil sample was placed in a polyethylene cup of 31-mm in diameter (Chemplex Industries Inc., Palm City, USA), whose bottom was sealed with a 4-μm thick polypropylene film (SPEX CertiPrep Inc., Metuchen, USA). For the LIBS analysis, approximately eight grams of each soil sample was mixed (10% w w^{-1}) with a binder material (microcrystalline cellulose, Merck, Darmstadt, Germany), grounded, and homogenized using a planetary ball mill (PM 200 mill, Retsch, Haan, Germany) during 20 min before they were finally pelletized for scanning.

6.2.4. VNIR measurements and spectra pre-processing

Veris spectrometer (Veris Technologies, Salina, USA) was used in benchtop mode to acquire the VNIR spectra. This system uses a tungsten halogen lamp as the energy source and two detection systems, a CCD array (USB4000, Ocean optics, Largo, USA) and an InGaAs photodiode-array (C9914GB, Hamamatsu Photonics, Hamamatsu, Japan), to register spectra from 343 to 2222 nm, with ±5 nm of spectral resolution. As a quality control, the instrument checks its spectrum intensity using four factory reference materials during its initialization. The sensor was also self-calibrated, by collecting a dark and a white reference measurement, before each spectra acquisition. For scanning, the samples were placed in a sample holder that isolates the entrance of external light; so, there is no need for a dark room. Each sample was scanned in triplicate, by repositioning the sample after each reading to account for the micro heterogeneity of the sample (Tavares et al., 2019). Before the spectra pre-processing steps, the replicates were averaged.

The raw spectra were first reduced from 343–2222 nm to 437–2149 nm range due to noise at the spectrum edges. Thereafter, the link between the two detectors, an artifact at 1040 nm, was corrected following the method proposed by Mouazen et al. (Mouazen et al., 2009). These spectra were then pre-processed following the successive order: standard normal variate (SNV) > maximum normalization > first derivative with Savitzky–Golay > smoothing with Savitzky–Golay. These four pre-processing algorithms are frequently adopted in the literature. The SNV is a scattering correction method that is applied to soil VNIR spectra to remove the multiplicative interferences of particle size (Barnes et al., 1989). The maximum normalization was carried out to get all spectral data to approximately the same scale (Mouazen et al., 2006), while the first derivative was applied to improve the signal-to-noise ratio by enhancing weak spectral features (Ben-Dor et al., 1997). The smoothing approach was used to remove noise and improve the signal-to-noise ratio (Nawar and Mouazen, 2017; Rinnan et al., 2009) further. All

data pre-processing steps were performed using the Unscrambler® version 10.5.1 (Camo AS, Oslo, Norway).

6.2.5. XRF measurements and spectra pre-processing

A portable device Tracer III-SD model (Bruker AXS, Madison, EUA) was used for XRF data acquisition. This instrument has a 4 W Rh X-ray tube and a Peltier-cooled Silicon Drift Detector with 2048 channels. The X-ray tube was configured for voltage and current of 35 kV and 7 μA, respectively (Tavares et al., 2020a). No filter was used and the scanning time was 90 s, which was performed under atmospheric pressure (Tavares et al., 2020a). Three readings were taken from each soil specimen at three different spots, and these were then averaged to be considered for spectra pre-processing and modeling.

The acquired spectra were normalized by the detector live time and evaluated in counts of photons per second (cps). Considering the area under each peak, 12 spectral lines (K-lines of Al, Si, K, Ca, Ti, Mn, Fe, Ni, and Cu, and the scattering peaks Rh-Lα Thomson, Rh-Kα Compton, and Rh-Kα Thomson) were selected to be used as explanatory variables, following the criteria recommended by Tavares et al. (2020b). Finally, the K-line of Al, Si, K, Ca, Ti, Mn, Fe, Ni, and Cu were normalized by the Compton peak.

6.2.6. LIBS measurements and spectra pre-processing

The same equipment and instrumental conditions described in the 5.2.4. Section (Chapter 5) were used in the present study. Regarding the LIBS spectra pre-processing, they were initially downscaled to 200 to 540 nm, which is a region containing the most valuable and high-intensity emission lines (Yu et al., 2016). Afterwards, the interval successive projection algorithm (iSPA) was applied to select significant variables (Gomes et al., 2013). To employ this algorithm, the LIBS spectra from 200 to 540 nm were divided into 160 intervals of 243 variables. The regions selected for each fertility attribute are shown in Table A1 (Appendix Section). The mean raw spectra acquired with the VNIR, XRF, and LIBS technique are shown in Figure 1.

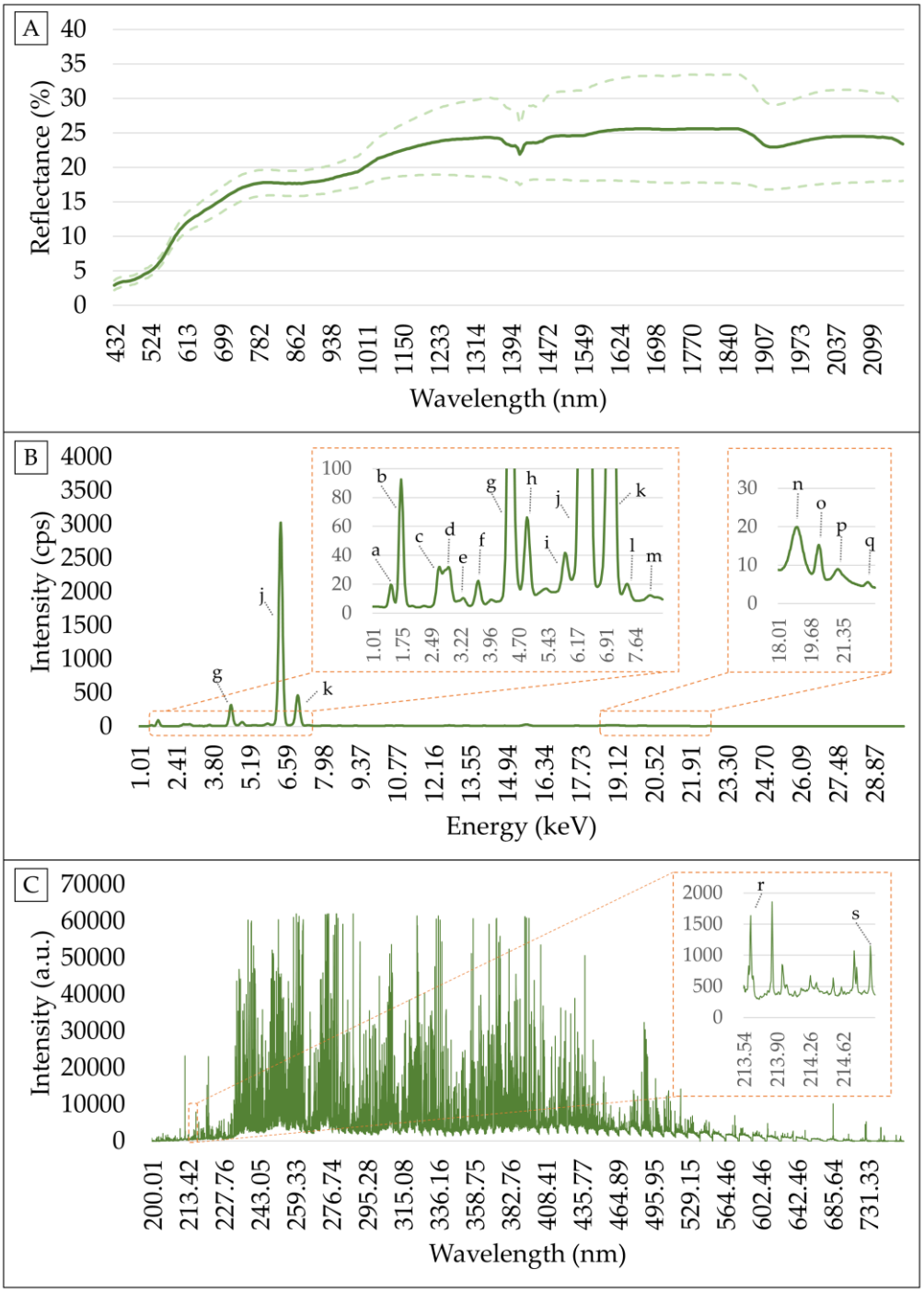


Figure 1. Mean spectra measured with visible and near infrared (VNIR) (A), X-ray fluorescence (XRF) (B), and (LIBS) (C) sensors. Snapshot of some XRF and LIBS emission lines are also shown. VNIR spectra are presented with their standard deviation (dash line) above and below the mean curve. The emission lines identified with the following letters correspond to: Al-Kα (a), Si-Kα (b), Rh-Lα Thomson (c), Ar-Kα (d), K-Kα (e), Ca-Kα (f), Ti-Kα (g), Ti-Kβ (h), Mn-Kα (i), Fe-Kα (j), -Fe-Kβ (k), Ni-Kα (l), Cu-Kα (m), Rh-Kα Compton (n), Rh-Kα Thomson (o), Rh-Kβ Compton (p), Rh-Kβ Thomson (q), P I 213.62 (r), and P I 214.91 (s). Atomic units and counts of photons per second have been abbreviated as a.u. and cps.

6.2.7. Modeling

After pre-processing, spectral variables were associated with the laboratory measured soil attributes by establishing predictive models. This was done using the spectral data of each single-sensor and for all possible combinations of sensors, through the following associations: VNIR + XRF, VNIR + LIBS, XRF + LIBS, and VNIR + XRF+ LIBS. The calibration and validation subsets comprised, respectively, of 70% and 30% of the total number of samples that were divided using the Kennard–Stone algorithm (Kennard and Stone, 1969) executed on the measured soil attributes. The prediction performance was evaluated through the determination coefficient (R^2), root mean square error (RMSE), and the residual prediction deviation (RPD). Based on the RPD values, the prediction quality of developed models were classified into four classes adapted from Chang et al. (2001): poor models (RPD < 1.40), reasonable models (1.40 ≤ RPD < 2.00), good models (2.00 ≤ RPD < 3.00), and excellent models (RPD ≥ 3.00). The relative improvement (RI) of the predictions accomplished by means of the multi-sensor approaches were also calculated, in terms of percentage of RMSE (Cardelli et al., 2017; O'Rourke et al., 2016), and compared to the best prediction obtained using the single-sensor (VNIR, XRF, and LIBS, individually) and two-sensor approach (VNIR + XRF, VNIR + LIBS, and XRF + LIBS).

6.2.7.1. Single-sensor predictive models

The predictive models using VNIR and LIBS data individually were established by means of PLS regression (Munnaf et al., 2019; Erler et al., 2020). The number of latent variables adopted for each PLS model was determined for the model in cross-validation that resulted in the lowest RMSE. For the XRF, models were calibrated with multiple linear regression (MLR), as suggested by Tavares et al. (2020b). All the calibrations and validations were performed using the Unscrambler software, version 10.5.1 (Camo AS, Oslo, Norway).

6.2.7.2. Data fusion approaches

Two data fusion approaches were used to build the predictive models using multi-sensor data. The first was designated as spectral fusion (SF) method. It is a front-end approach, consisting of merging the pre-processed spectral data of each combination of sensor (VNIR + XRF, VNIR + LIBS, XRF + LIBS, and VNIR + XRF + LIBS) in one matrix. The merged data were scaled by the standard deviation to create an even distribution of variances, before they were subjected to PLS regression to develop a model for each soil attribute. The number of latent variables for each PLS model was determined according to the leave-one-out cross-validation that resulted in the lowest RMSE.

The second data fusion method was adapted from the Granger and Ramanathan (GR) averaging method (Granger and Ramanathan, 1984), as performed by Tavares et al. (2021). In the present study, the GR approach consisted of calibrating a multiple linear regression (MLR) with independent variables being the predictions made with the sensors, plus the prediction made with the multi-sensor merged data (*e.g.*, SF approach). For example, for the combination between XRF and VNIR sensors, the GR method was applied according to Equation (1):

$$Y = W_0 + (W_{VNIR} \cdot X_{VNIR}) + (W_{XRF} \cdot X_{XRF}) + (W_{SF} \cdot X_{SF}), \quad (1)$$

where: Y is the soil property to be estimated; X_{VNIR} and X_{XRF} are the predictions made by VNIR and XRF single-models, respectively; XSF is the prediction made by the SF approach using the VNIR and XRF merged data; W_0, W_{VNIR}, W_{XRF}, WSF are the parameters of the MLR determined by least squares method, where the first parameter is the value of the line intercept and the others are the weights of the predictions.

Equation (1) was also applied for combinations of LIBS + VNIR, LIBS + XRF, and VNIR + XRF + LIBS. All data fusion models were calibrated and validated using the Unscrambler software, version 10.5.1 (Camo AS, Oslo, Norway).

6.3. Results

6.3.1. Laboratory measured soil properties

The descriptive statistics and correlation matrix of the laboratory measured soil properties were presented in the 4.3.1.1. Section (Chapter 4).

6.3.2. Prediction performances of single-sensor and multi-sensor approaches

The prediction results of individual and combined application of VNIR, XRF, and LIBS sensors are presented in Table 1 (see details in Table A2, in the Appendix Section). Considering the single-sensor prediction performance, VNIR presented excellent and good prediction performance for clay (RPD = 3.37), OM (RPD = 2.61), and V (RPD = 2.26). Among the three sensors, VNIR showed the lowest RMSE for clay and OM prediction (Table A2). The XRF sensor showed good prediction performance for clay (RPD = 3.13), CEC (RPD = 2.57), and ex-Mg (RPD = 2.99), and excellent predictions for V (RPD = 4.18), ex-K (RPD = 4.26), and ex-Ca (RPD = 4.82). Meanwhile, the LIBS sensor achieved excellent prediction performance for clay (RPD = 3.06), V (RPD = 4.27), ex-Ca (RPD = 4.07), and ex-Mg (RPD = 3.65), and good prediction performance for OM (RPD = 2.14). Among the three sensors when used individually,

XRF exhibited the lowest RMSE for CEC, ex-K and ex-Ca prediction, while LIBS showed the lowest RMSE for the prediction of pH, V, ex-P and ex-Mg (Table A2). The pH was the only attribute that could not be satisfactorily predicted (RPD < 1.12) by any of the single-sensor approaches, while the ex-P only was only satisfactorily predicted by LIBS (RPD = 1.82).

Table 1. Residual prediction deviation (RPD) obtained for the validation set (n = 34) of clay, organic matter (OM), cation exchange capacity (CEC), pH, base saturation (V), and extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg) using the visible and near infrared (VNIR), X-ray florescence (XRF), and laser induced breakdown spectroscopy (LIBS) data individually and different combination of them through spectra fusion (SF) and Granger and Ramanathan (GR) approaches.

	Clay	OM	CEC	pH	V	ex-P	ex-K	ex-Ca	ex-Mg
VNIR	3.37	2.61	1.40	1.10	2.26	0.88	1.89	1.79	1.45
XRF	3.13	1.82	2.57	1.11	4.18	0.80	4.26	4.82	2.99
LIBS	3.06	2.14	2.45	1.12	4.27	1.82	2.05	4.07	3.65
VNIR + XRF - SF	3.51	2.39	1.96	1.22	3.67	0.90	3.09	2.83	2.33
VNIR + XRF - GR	3.53	2.35	1.94	1.26	4.24	0.89	4.98	3.90	2.66
VNIR + LIBS - SF	3.90	2.38	2.55	1.31	3.69	1.07	1.96	3.71	2.86
VNIR + LIBS - GR	3.97	2.43	2.49	1.31	3.65	1.48	2.08	3.89	3.28
XRF + LIBS - SF	3.03	1.84	2.57	1.05	3.69	1.54	3.78	5.14	3.94
XRF + LIBS - GR	3.26	2.38	2.40	1.14	3.57	1.44	4.52	5.30	4.11
VNIR + XRF + LIBS - SF	4.02	2.42	2.60	1.31	3.88	1.05	2.42	3.97	3.47
VNIR + XRF + LIBS - GR	3.73	2.44	2.63	1.32	3.77	1.43	4.83	4.96	3.83

The RPD values for the same soil attribute were compared and presented on gray scale, highlighting the highest values within each soil attribute. The root-mean-square error (RMSE) are given in g dm^{-3} for clay and OM; in $mmol_c$ dm^{-3} for CEC, ex-K, ex-Ca, and ex-Mg; in % for V; and, and in mg dm^{-3} for ex-P. Results of coefficient of determination (R^2) and RMSE are detailed in the Appendix Section (Table A2).

Table 2 shows the RI for the predictions performed using multi-sensor data through the two applied data fusion approaches (SF and GR), compared with the best single-sensor approach. This indicator shows the improvement (when positive) or the deterioration (when negative) obtained in the sensor fusion relative improvement (Cardelli et al., 2017; O'Rourke et al., 2016). The attributes OM, V and ex-P were better predicted by the individual than combined sensors. All the other attributes (clay, CEC, pH, ex-K, ex-Ca, and ex-Mg) were better predicted by multi-sensor approaches.

The RI values for clay prediction ranged from -11.1 to 16.3%, for the SF approach, and from -3.2 to 15.2%, for the GR approach. For CEC prediction, the RI ranged from -31.1 to 1.4%, using SF, and from -32.0 to 2.3%, using GR. Although the pH models did not perform satisfactorily, their RI values ranged from -6.5 to 14.8% and 1.9 to 15.3% for SF and GR, respectively. For the ex-K prediction, the RI ranged from -117.4 to -12.6% for SF, and -104.5 to 14.4% for GR. For ex-Ca, the RI varied between -70.2 and 6.2% for SF, and between -23.9 and 9.2% for GR. For ex-Mg prediction, the RI values ranged from -28.5 to 24.2%, using SF, and from -12.4 to 27.2%, using GR. VNIR + XRF led to positive improvement in the prediction of clay, pH and ex-K (with GR only), although improvements were larger in VNIR + LIBS for clay and pH only. As expected, the fusion of XRF and LIBS improved the prediction of the extractable nutrients mainly,

ex-K, ex-Ca and ex-Mg, with minor improvement in CEC, with SF, and pH, with GR (Table 2). The fusion of the three sensors have scored the largest number of improvements with positive RI values for all studied attributes except for OM, V and ex-P. In general, the performance of the back-end data fusion strategy using the GR approach outperformed the SF approach in 72.2% of cases. In summary, the improvement in predictive performance varied according to the combination of sensors as well as the data fusion strategy used; there was no one optimal approach for the prediction of all the attributes analysed.

Table 2. Relative improvement (RI) and root-mean-square error (RMSE) achieved for each multi-sensor approach using spectra fusion (SF) and Granger and Ramanathan (GR) data fusion methods, in comparison with the best prediction obtained using the single-sensor approaches.

			Clay	OM[1]	CEC[2]	pH	V[3]	ex-P[4]	ex-K[4]	ex-Ca[4]	ex-Mg[4]
Best single-sensor approach		RMSE	27.32	**2.10**	10.19	0.33	**5.49**	**5.84**	0.53	4.09	4.28
		Tech[5]	VNIR	VNIR	XRF	LIBS	LIBS	LIBS	XRF	XRF	LIBS
VNIR +XRF	SF	RMSE	26.19	2.29	13.35	0.30	6.38	11.81	0.73	6.97	5.51
		RI (%)	4.1	-9.2	-31.1	8.6	-16.2	-102.3	-37.6	-70.2	-28.5
	GR	RMSE	26.05	2.33	13.45	0.29	5.52	11.91	**0.45**	5.06	4.82
		RI (%)	4.7	-10.7	-32.0	11.2	-0.7	-104.1	14.4	-23.5	-12.4
VNIR +LIBS	SF	RMSE	23.57	2.30	10.26	0.28	6.34	9.97	1.16	5.32	4.48
		RI (%)	13.7	-9.5	-0.7	14.5	-15.6	-70.7	-117.4	-30.0	-4.5
	GR	RMSE	23.18	2.25	10.49	0.28	6.42	7.20	1.09	5.07	3.90
		RI (%)	15.2	-7.1	-2.9	14.5	-16.9	-23.3	-104.5	-23.9	8.9
XRF +LIBS	SF	RMSE	30.35	2.98	10.17	0.35	6.34	6.90	0.60	3.84	3.25
		RI (%)	-11.1	-41.8	0.2	-6.5	-15.6	-18.2	-12.6	6.2	24.2
	GR	RMSE	28.19	2.30	10.89	0.32	6.56	7.39	0.50	**3.72**	**3.12**
		RI (%)	-3.2	-9.3	-6.8	1.9	-19.5	-26.6	5.8	9.2	27.2
VNIR + XRF +LIBS	SF	RMSE	**22.88**	2.27	10.04	0.28	6.04	10.17	0.93	4.96	3.70
		RI (%)	16.3	-7.8	1.4	14.8	-10.0	-74.2	-75.7	-21.2	13.7
	GR	RMSE	24.68	2.25	**9.95**	**0.28**	6.22	7.47	0.47	3.98	3.34
		RI (%)	9.7	-6.9	2.3	15.3	-13.3	-28.0	11.9	2.9	22.0

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg); [5] related technique. The positive RI values for the same soil attribute were compared and presented on gray scale, with higher RI values having darker color and vice versa. For each soil attribute, bolded RMSE values indicate the approach with the lowest prediction error.

Table 3 compares the prediction performances of the three-sensor approach with the best performances obtained with the two-sensor and the single-sensor approaches. Results revealed that the best predictions of OM, V and ex-P were obtained using the single-sensor approach (with VNIR for OM, and with LIBS for V and ex-P). Predictions of these attributes using two sensors showed degradation in predictive performance, with the best two-sensor approach showing an RMSE increment of 7.1, 0.7 and 18.2% for OM, V, and ex-P, respectively. A similar deterioration in the prediction performance of these attributes (OM, V, and ex-P) was also observed by the three-sensor approach, with an increase in the

prediction error oscillating between 6.9 and 74.2% in comparison to the best performing single-sensor approach.

Concerning ex-K, ex-Ca and ex-Mg, better prediction results were obtained when using the two-sensor approach. The VNIR + XRF combinations were the most promising for ex-K, while XRF + LIBS stood out for ex-Ca and ex-Mg. The performance increment obtained by the above-mentioned two-sensor approaches, relatively to the best single-sensor strategy, was 14.4, 9.2 and 27.2% for ex-K, ex-Ca and ex-Mg, respectively. When combining a third sensor, the predictive performance was degraded, with a reduction of 2.9, 6.5 and 6.8% for ex-K, ex-Ca and ex-Mg, respectively, relative to the best two-sensor approach.

Concerning the clay, CEC and pH the best prediction performances were recorded by the three-sensor approach, which reduced the error to a range between 2.3 and 16.3%, in comparison to the best single-sensor approach, and between 0.9 and 2.2%, relative to the best approach using two sensors.

Due to the presence of synergistic information in the data, improved predictions were obtained for a large number of attributes (six out of nine) when using the three-sensor approach. The combination between XRF and LIBS showed synergy for the prediction of five out of nine attributes, while both the combinations VNIR + LIBS and VNIR + XRF showed synergy for three out of nine attributes.

Table 3. Relative improvement (RI) achieved for the predictions using the best three-sensor approach in comparison with the best prediction obtained using the two-sensor, and the single-sensor approaches. The root mean square error (RMSE) of each approach was also presented.

	Best single-sensor approach		Best two-sensor approach				Three-sensor approach			
	RMSE	Tech[5]	RMSE	Tech[5]	DF[6]	RI comp. single-sensor[7]	RMSE	DF[6]	RI comp. single-sensor[7]	RI comp. two-sensor[8]
Clay	27.32	VNIR[9]	23.18	VNIR + LIBS	GR[12]	15.2	**22.88**	SF[13]	16.3	1.3
OM[1]	**2.10**	VNIR	2.33	VNIR + XRF	GR	-10.7	**2.25**	GR	-6.9	3.5
CEC[2]	10.19	XRF[10]	10.17	XRF + LIBS	SF	0.2	**9.95**	GR	2.3	2.2
pH	0.33	LIBS[11]	0.28	VNIR + LIBS	GR	14.5	**0.28**	GR	15.3	0.9
V[3]	**5.49**	LIBS	5.52	VNIR + XRF	GR	-0.7	6.04	SF	-10.0	-8.5
ex-P[4]	**5.84**	LIBS	7.20	VNIR + LIBS	GR	-23.3	7.47	GR	-28.0	-3.6
ex-K[4]	0.53	XRF	**0.45**	VNIR + XRF	GR	14.4	0.47	GR	11.9	-2.9
ex-Ca[4]	4.09	XRF	**3.72**	XRF + LIBS	GR	9.2	3.98	GR	2.9	-6.5
ex-Mg[4]	4.28	LIBS	**3.12**	XRF + LIBS	GR	27.2	3.34	GR	22.0	-6.8

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients; [5] related technique(s); [6] related data fusion approach; [7] RI compared with the best single-sensor approach; [8] RI compared with the best two-sensor approach; [9] visible and near infrared diffuse reflectance spectroscopy; [10] X-ray fluorescence spectroscopy; [11] laser induced breakdown spectroscopy; [12] Granger and Ramanathan approach; and [13] spectral fusion approach. For each soil attribute, bolded RMSE values indicate the approach with the lowest prediction error.

6.4. Discussion

Soil spectral analyses via the tandem use of VNIR, XRF and LIBS sensors have two potential benefits: (i) extend the set of predicted soil attributes in comparison with single-sensor approaches; and (ii) improve the models' predictive performance by exploring synergy in the data through data fusion approaches, both are discussed in Sections 6.4.1 and 6.4.2, respectively.

6.4.1. Individual performance of VNIR, XRF and LIBS sensors: coverage of key soil fertility attributes

The results observed in this work show that the tandem application of VNIR, XRF and LIBS techniques, allows extending the number of predicted fertility attributes with optimal performances (Table A3, in the Appendix Section). For individual sensor limited number of attributes could be detected and the attributes are technology specific. For example, excellent and good predictions could be achieved for clay and OM, respectively, with both the VNIR and LIBS sensors

Although the VNIR sensor allowed more accurate predictions of these attributes, since both attributes have direct spectral responses in the NIR range (Ben-Dor, 2002). A similar trend was observed for the chemical attributes CEC, V, ex-K, ex-Ca, and ex-Mg, whose optimal prediction performances were obtained with elemental analysis sensors (with RPD oscillating between 2.57 and 4.82) due to the fact these attributes are closely related to the emission lines present in these sensors.

The LIBS spectra presented emission lines for C (in 247.856 nm), P (in 213.6182 and 214.914 nm) and Mg (in 277.983 and 285.213 nm), which were not observable in the XRF spectra. The presence of these emission lines explains the better performance of LIBS for OM, ex-P and ex-Mg prediction than XRF. The XRF spectra showed K emission line (in 3.31 keV), which was not observable in the LIBS data, explaining its superior performance for ex-K prediction. The excellent ex-Ca prediction of both elemental techniques is well supported by the Ca emission lines (in 3.69 keV for XRF, and in 315.887 and 317.933 for LIBS). The presence of Ca emission lines also explains the similar performance for V and CEC prediction obtained by both XRF and LIBS (good predictions for CEC and excellent for V), since both attributes showed strong correlation with the ex-Ca ($r \geq 0.92$).

The attributes with the worst predictive performance were the ex-P and the pH. Regarding ex-P, LIBS was the only sensor to present a satisfactory prediction for this attribute. Although LIBS shows well defined emission lines for P, total contents of P may not always correlate with ex-P contents due to the complex chemistry of this element in tropical soils (Fontes, 2012). P in tropical soils is commonly associated, in a non-extractable manner, with Fe and Al oxides (Schäefer et al., 2008), which

makes the modeling of ex-P via elemental analysis techniques challenging (Andrade et al., 2020; Tavares et al., 2020a). A possible alternative to improve the predictive performance of ex-P and pH models is through the integration of electro-chemical sensors [*e.g.*, ion-selective electrodes (ISE) and the ion-sensitive field-effect transistor (ISFET)] to the multi-sensor array used, since with the former technology ensure direct soil analysis of the extractable contents of some ions (*e.g.*, H^+ and PO_4^{3-}) (Kuang et al., 2012; Molin and Tavares, 2019).

In general, the predictive capability of key soil fertility attributes decreased as fol-lows: LIBS > XRF > VNIR. These results differed from those presented by Xu et al. (2019a), which is the only research that explored VNIR, XRF and LIBS sensors for predicting soil fertility attributes. The authors obtained better prediction performances with the VNIR sensor, in comparison to the XRF and LIBS sensors, for OM, pH, ex-P, ex-K, and ex-N. Our results, however, are in line with those of other research carried out in tropical soils that also evaluated the standalone use of VNIR (Coutinho et al., 2019; Demattê et al., 2015, 2017; Demattê et al., 2019; Lacerda et al., 2016) and XRF (Andrade et al., 2020; Lima et al., 2019; Santos et al., 2020; Silva et al., 2019; Silva et al., 2017) techniques. Studies using LIBS sensors for the assessment of fertility attributes are incipient in tropical soils. However, the satisfactory performance obtained for clay is supported by that obtained by Villas-Boas et al. (2016). In turn, our poor pH predictions contrasted with the satisfactory predictions obtained by Ferreira et al. (2015).

6.4.2. Combined performance of VNIR, XRF and LIBS sensors: synergy for predicting key soil fertility attributes

The prediction of OM, V and ex-P did not show any synergy by the multi-sensor approaches evaluated in this study, as they were better predicted by each sensor individually. However, improved prediction performance was obtained for all the other attributes (clay, CEC, pH, ex-K, ex-Ca, and ex-Mg), with RI ranging between 0.2 and 27.2% when using at least one of the multi-sensor approaches. The absence of synergy in the information provided by VNIR and elemental analysis sensors has already been reported in the literature for pH, ex-K and OM (Tavares et al., 2021; Xu et al., 2019a). Conversely, O'Rourke et al. (2016), assessing the combination XRF + VNIR in Australian soils, reported RI ranging from 15 to 44% for clay, CEC, pH, ex-K, ex-Ca, and ex-Mg. Also evaluating data fusion between VNIR and XRF sensors, Zhang and Hartemink (2020) obtained RI values oscillating from 3 to 20% for the prediction of clay, pH, and total carbon in North American soils. A similar study achieved RI of 26% for CEC in Chinese soils (Wan et al., 2020). Also in Chinese soils, Xu et al. (2019b) reached an RI varying between 4 and 35% for OM prediction by the combination of LIBS and attenuated total reflectance Fourier-transform mid-infrared spectroscopy (FTIR-ATR).

In general, the prediction quality of the studied attributes by the sensor fusion decreased in the order of: VNIR + XRF + LIBS > XRF + LIBS > VNIR + LIBS > VNIR + XRF. Our results showed that there was no unique optimal sensor combination for predicting all the key soil fertility attributes, and that this is attribute specific. A similar behaviour was also observed by Xu et al. (2019a). Thus, the performance of different combinations of sensors should be evaluated by trial-and-error and that it is too soon to establish a clear trend, which requires further research. This means that data fusion is in its infancy stage, and more works will be necessary to recommend the optimal set of sensors for one or a set of soil attributes.

This study was conducted using 102 soil samples with broad variation of fertility attributes collected in Lixisols and Ferralsols, which are representative and common types of soil in Brazilian agricultural fields. In addition, this pioneering evaluation provided useful information to help PSS users to understand the advantages and drawbacks of the combined use of VNIR, XRF, and LIBS sensors for soil fertility analysis. Nevertheless, it is also fundamental to evaluate larger datasets including other types of soils, which present high variability in soil mineralogy, textural classes, concentration range, and submitted to different agricultural practices.

6.4.3. Operational aspects related to sample preparation

It is important to consider practical aspects of sample preparation necessary to ensure the application of each technique. The VNIR and XRF techniques are less demanding in terms of sample preparation, and can be applied on loose soil samples (*e.g.*, dried and sieved with 2 mm of particle size) without major impacts on the sensors performance (Coutinho et al., 2019; Tavares et al., 2019). However, the LIBS demands further sample preparation steps including particle size reduction with grinding, followed by pressing to create pellets that are necessary for optimal LIBS performance. Pelletizing promotes homogenization of the sample, and allows congruent ablations (necessary to maintain the stoichiometric proportion of the craters formed by the laser) (Jantzi et al., 2016). These extra sample preparation steps to ensure the collection of reliable data make it challenging to apply LIBS to run analyses directly in the field (Knadel et al., 2017). However, despite the extra time consigned to sample preparation, it would not make it unfeasible to use LIBS in controlled environments (*e.g.* hybrid laboratories and mobile laboratories) that enable the execution of such sample conditioning (Molin and Tavares, 2019).

6.5. Conclusions

This study shows that the fusion of data from visible and near infrared diffuse reflectance spectroscopy (VNIR), X-ray fluorescence spectroscopy (XRF), and laser induced breakdown spectroscopy (LIBS) can extend the number of fertility attributes predicted with optimal performance, as well as improve the prediction accuracy by exploiting the synergy through data fusion techniques.

Comparing among the individual sensor performance, the VNIR technique performed best for the prediction of clay and organic matter (OM) ($2.61 \leq RPD \leq 3.37$), while the elemental analysis techniques showed better performance for the prediction of cation exchange capacity (CEC), base saturation (V), and the extractable (ex-) K, Ca, and Mg ($2.61 \leq RPD \leq 3.37$). The LIBS stood out for V, ex-P and ex-Mg prediction, while the XRF resulted in optimal predictions for CEC, ex-K and ex-Ca. The attributes with the worst predictive performance were the ex-P ($0.80 \leq RPD \leq 1.82$) and pH ($RPD \leq 1.12$), with the latter performing poorly in all individual and combined sensor options tested in this study.

Regarding the performance of the multi-sensor approaches, ex-K, ex-Ca and ex-Mg, were better predicted when using two sensors (VNIR + XRF for ex-K, and XRF + LIBS for ex-Ca and ex-Mg), while clay, CEC and pH were best predicted with the three-sensor (VNIR + XRF + LIBS). However, the combined use of sensors did not always lead to improvement in the prediction results. For instance, the best sensing method for OM, V and ex-P prediction was the single-sensor approach, showing that there was no synergy between the evaluated sensors for the prediction of these attributes. Furthermore, the best combination of sensors for predicting key fertility attributes proved to be attribute-specific, which is a drawback of the in tandem application of these analytical techniques. Another drawback to integrate the addressed techniques is the need for different sample preparations, where LIBS needs extra steps to conform the loose soil samples into pellets.

Finally, the present work evidenced benefits and limitations of the in tandem application of VNIR, XRF and LIBS spectroscopies for fertility analysis in tropical soils. Further research should be encouraged to improve analytical protocols using multi-sensor approaches, to enable practical, environmentally friendly and accurate analysis, allowing its incorporation into future hybrid laboratories.

Appendix

Table A1. Intervals of variables selected by the interval successive projection algorithm (iSPA) in the LIBS spectra for the predictive modeling of each soil fertility attribute.

Attribute	Number of intervals selected	Spectral ranges of the selected intervals (nm)
Clay	3	220.84 – 222.21, 246.77 – 248.29, and 251.37 – 252.94
OM[1]	1	295.08 – 296.92
CEC[2]	52	201.31 – 272.37, 287.89 – 298.73, and 414.15 – 416.63
pH	3	220.84 – 222.21, 237.82 – 239.27, and 391.78 – 394.22
V[3]	57	201.31 – 272.37, 279.18 – 300.60, and 396.59 – 399.10
ex-P[4]	1	214.14 – 215.44
ex-K[4]	1	375.28 – 377.53
ex-Ca[4]	1	389.37 – 391.77
ex-Mg[4]	1	284.36 – 286.14

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients.

Table A2. Prediction results of the validation set (n = 34) obtained using single VNIR, XRF, and LIBS data alone and using multi-sensor data, combined through spectra fusion (SF) and Granger and Ramanathan (GR) approaches.

	Clay	OM[1]	CEC[2]	pH	V[3]	ex-P[4]	ex-K[4]	ex-Ca[4]	ex-Mg[4]
					R^2				
VNIR	0.93	0.86	0.51	0.19	0.80	0.07	0.74	0.68	0.52
XRF	0.92	0.74	0.88	0.34	0.95	0.01	0.95	0.96	0.89
LIBS	0.89	0.81	0.84	0.31	0.94	0.72	0.76	0.94	0.93
VNIR + XRF – SF	0.94	0.83	0.77	0.36	0.93	0.14	0.91	0.89	0.82
VNIR + XRF - GR	0.94	0.83	0.77	0.39	0.95	0.13	0.96	0.94	0.86
VNIR + LIBS - SF	0.95	0.84	0.85	0.45	0.93	0.32	0.75	0.93	0.89
VNIR + LIBS - GR	0.95	0.84	0.84	0.45	0.92	0.59	0.77	0.94	0.92
XRF + LIBS - SF	0.90	0.75	0.85	0.24	0.93	0.61	0.94	0.96	0.94
XRF + LIBS - GR	0.92	0.86	0.82	0.35	0.93	0.55	0.96	0.96	0.94
VNIR + XRF + LIBS - SF	0.95	0.84	0.85	0.46	0.94	0.31	0.83	0.94	0.92
VNIR + XRF + LIBS - GR	0.95	0.85	0.86	0.47	0.93	0.57	0.96	0.96	0.94
					RMSE				
vis-NIR	27.32	2.10	18.66	0.34	10.38	12.05	1.20	10.98	8.85
XRF	29.40	3.01	10.19	0.33	5.60	13.27	0.53	4.09	4.28
LIBS	30.03	2.56	10.69	0.33	5.49	5.84	1.10	4.85	3.51
VNIR + XRF - SF	26.19	2.29	13.35	0.30	6.38	11.81	0.73	6.97	5.51
VNIR + XRF - GR	26.05	2.33	13.45	0.29	5.52	11.91	0.45	5.06	4.82
VNIR + LIBS - SF	23.57	2.30	10.26	0.28	6.34	9.97	1.16	5.32	4.48
VNIR + LIBS - GR	23.18	2.25	10.49	0.28	6.42	7.20	1.09	5.07	3.90
XRF + LIBS - SF	30.35	2.98	10.17	0.35	6.34	6.90	0.60	3.84	3.25
XRF + LIBS - GR	28.19	2.30	10.89	0.32	6.56	7.39	0.50	3.72	3.12
VNIR + XRF + LIBS - SF	22.88	2.27	10.04	0.28	6.04	10.17	0.93	4.96	3.70
VNIR + XRF + LIBS - GR	24.68	2.25	9.95	0.28	6.22	7.47	0.47	3.98	3.34

[1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients (ex-P, ex-K, ex-Ca, and ex-Mg). The coefficient of determination (R^2) values for the same soil attribute were compared and presented on gray scale, highlighting the highest values within each soil attribute. The root-mean-square error (RMSE) are given in g dm^{-3} for clay and OM; in $mmol_c$ dm^{-3} for CEC, ex-K, ex-Ca, and ex-Mg; in % for V; and, for ex-P, the RMSE was given in mg dm^{-3}.

Table A3. Qualitative interpretation* of the residual prediction deviation (RPD) performance obtained for each of the evaluated approaches using single-sensor (VNIR, XRF, and LIBS) and multi-sensor (VNIR + XRF, VNIR + LIBS, XRF + LIBS, and VNIR + XRF + LIBS).

	VNIR	XRF	LIBS	VNIR + XRF	VNIR + LIBS	XRF + LIBS	VNIR + XRF + LIBS
Clay	Excel.	Excel.	Excel.	Excel.	Excel.	Excel.	Excel.
OM[1]	Good	Reason.	Good	Good	Good	Good	Good
CEC[2]	Reason.	Good	Good	Reason.	Good	Good	Good
pH	Poor	Poor	Poor	Poor	Poor	Poor	Poor
V[3]	Good	Excel.	Excel.	Excel.	Excel.	Excel.	Excel.
ex-P[4]	Poor	Poor	Reason.	Poor	Reason.	Reason.	Reason.
ex-K[4]	Reason.	Excel.	Good	Excel.	Good	Excel.	Excel.
ex-Ca[4]	Reason.	Excel.	Excel.	Excel.	Excel.	Excel.	Excel.
ex-Mg[4]	Reason.	Good	Excel.	Good	Excel.	Excel.	Excel.

* The qualitative interpretation was adapted from Chang et al. (2001) by using the following classes: poor models for RPD < 1.40 (highlighted in orange), reasonable models for 1.40 ≤ RPD < 2.00 (highlighted in light yellow), good models for 2.00 ≤ RPD < 3.00 (highlighted in light green), and excellent models for RPD ≥ 3.00 (highlighted in dark green); [1] Organic matter; [2] cation exchange capacity; [3] base saturation; [4] extractable (ex-) nutrients.

References

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

Barnes, R. J.; Dhanoa, M. S.; Lister, S. J. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. *Applied Spectroscopy* **1989**, 43(5), 772–777. https://doi.org/10.1366/0003702894202201

Ben-Dor, E. Quantitative remote sensing of soil properties. *Advances in Agronomy* **2002**, 75, 173–243. https://doi.org/https://doi.org/10.1016/S0065-2113(02)75005-0

Ben-Dor, E.; Inbar, Y.; Chen, Y. The reflectance spectra of organic matter in the visible near-infrared and short wave infrared region (400–2500 nm) during a controlled decomposition process. *Remote Sensing of Environment* **1997**, 61(1), 1–15. https://doi.org/https://doi.org/10.1016/S0034-4257(96)00120-4

Cardelli, V.; Weindorf, D. C.; Chakraborty, S.; Li, B.; De Feudis, M.; Cocco, S.; Agnelli, A.; Choudhury, A.; Ray, D. P.; Corti, G. Non-saturated soil organic horizon characterisation via advanced proximal sensors. *Geoderma* **2017**, 288, 130–142. https://doi.org/https://doi.org/10.1016/j.geoderma.2016.10.036

Castanedo, F. A Review of Data Fusion Techniques. *The Scientific World Journal* **2013**, 1–19. https://doi.org/10.1155/2013/704504

Chang, C.-W.; Laird, D. A.; Mausbach, M. J.; Hurburgh, C. R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal* **2001**, 65(2), 480–490. https://doi.org/10.2136/sssaj2001.652480x

Clark, R. N.; Roush, T. L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth* **1984**, 89(B7), 6329–6340. https://doi.org/10.1029/JB089iB07p06329

Coutinho, M.A.N.; Alari, F. de O.; Ferreira, M. M. C.; Amaral, L. R. do. Influence of soil sample preparation on the quantification of NPK content via spectroscopy. *Geoderma* **2019**, 338, 401–409. https://doi.org/10.1016/j.geoderma.2018.12.021

Demattê, J.A.M.; Alves, M. R.; Gallo, B. C.; Fongaro, C. T.; Romero, D. J.; Sato, M. V. Hyperspectral remote sensing as an alternative to estimate soil attributes. *Revista Ciência Agronômica* **2015**, 46(2), 223–232. https://doi.org/10.5935/1806-6690.20150001

Demattê, J.A.M.; Ramirez-Lopez, L.; Marques, K.P.P.; Rodella, A.A. Chemometric soil analysis on the determination of specific bands for the detection of magnesium and potassium by spectroscopy. *Geoderma* **2017**, 288, 8–22. https://doi.org/https://doi.org/10.1016/j.geoderma.2016.11.013

Demattê, J. A. M.; Dotto, A. C.; Bedin, L. G.; Sayão, V. M.; Souza, A. B. e. Soil analytical quality control by traditional and spectroscopy techniques: Constructing the future of a hybrid laboratory for low environmental impact. *Geoderma* **2019**, 337(December 2017), 111–121. https://doi.org/10.1016/j.geoderma.2018.09.010

Erler, A.; Riebe, D.; Beitz, T.; Löhmannsröben, H.-G.; Gebbers, R. Soil Nutrient Detection for Precision Agriculture Using Handheld Laser-Induced Breakdown Spectroscopy (LIBS) and Multivariate Regression Methods (PLSR, Lasso and GPR). *Sensors* **2020**, 20(2), 418. https://doi.org/10.3390/s20020418

Ferreira, E. C.; Gomes Neto, J. A.; Milori, D. M. B. P.; Ferreira, E. J.; Anzano, J. M. Laser-induced breakdown spectroscopy: Extending its application to soil pH measurements. *Spectrochimica Acta - Part B Atomic Spectroscopy* **2015**, 110, 96–99. https://doi.org/10.1016/j.sab.2015.06.002

Fontes, M.P.F. Intemperismo de rochas e minerais. In *Pedologia: Fundamentos*; Ker, J.C., Curi, N., Schaefer, C.*E.G.*R., Vidal-Torrado, P., Eds.; Sociedade Brasileira de Ciência do Solo: Viçosa, Brazil, 2012; pp. 171–205.

Gomes, A. D. A.; Galvão, R. K. H.; De Araújo, M. C. U.; Véras, G.; Da Silva, E. C. The successive projections algorithm for interval selection in PLS. *Microchemical Journal* **2013**, 110, 202–208. https://doi.org/10.1016/j.microc.2013.03.015

Granger, C. W. J.; Ramanathan, R. Improved methods of combining forecasts. *Journal of Forecasting* **1984**, 3(2), 197–204. https://doi.org/https://doi.org/10.1002/for.3980030207

Jantzi, S. C.; Motto-Ros, V.; Trichard, F.; Markushin, Y.; Melikechi, N.; De Giacomo, A. Sample treatment and preparation for laser-induced breakdown spectroscopy. *Spectrochimica Acta - Part B Atomic Spectroscopy* **2016**, 115, 52–63. https://doi.org/10.1016/j.sab.2015.11.002

Javadi, S. H.; Munnaf, M. A.; Mouazen, A. M. Fusion of Vis-NIR and XRF spectra for estimation of key soil attributes. *Geoderma* **2021**, 385, 114851. https://doi.org/10.1016/j.geoderma.2020.114851

Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, 11(1), 137–148. https://doi.org/10.1080/00401706.1969.10490666

Knadel, M.; Gislum, R.; Hermansen, C.; Peng, Y.; Moldrup, P.; de Jonge, L. W.; Greve, M. H. Comparing predictive ability of laser-induced breakdown spectroscopy to visible near-infrared spectroscopy for soil property determination. *Biosystems Engineering* **2017**, 156, 157–172. https://doi.org/10.1016/j.biosystemseng.2017.01.007

Krug, F. J.; Rocha, F. R. P. *Métodos de preparo de amostras para análise elementar*. EditSBQ: São Paulo, Brazil, 2016; 572p.

Kuang, B.; Mahmood, H. S.; Quraishi, M. Z.; Hoogmoed, W. B.; Mouazen, A. M.; van Henten, E. J. Sensing Soil Properties in the Laboratory, *In-situ*, and On-Line. *Advances in Agronomy* **2012**, 114, 155–223. https://doi.org/10.1016/B978-0-12-394275-3.00003-1

Lacerda, M. P. C.; Demattê, J. A. M.; Sato, M. V.; Fongaro, C. T.; Gallo, B. C.; Souza, A. B. Tropical Texture Determination by Proximal Sensing Using a Regional Spectral Library and Its Relationship with Soil Classification. *Remote Sensing* **2016**, 8 (9). https://doi.org/10.3390/rs8090701

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

Mahmood, H. S.; Hoogmoed, W. B.; van Henten, E. J. Sensor data fusion to predict multiple soil properties. *Precision Agriculture* **2012**, 13(6), 628–645. https://doi.org/10.1007/s11119-012-9280-7

Molin, J. P.; Tavares, T. R. Sensor systems for mapping soil fertility attributes: Challenges, advances, and perspectives in brazilian tropical soils. *Engenharia Agricola* **2019**, 39 (special issue), 126–147. https://doi.org/10.1590/1809-4430-ENG.AGRIC.V39NEP126-147/20190126

Mouazen, A. M.; Karoui, R.; De Baerdemaeker, J.; Ramon, H. Characterisation of Soil Water Content Using Measured Visible and Near Infrared Spectra. *Soil Science Society of America Journal* **2006**, 70(4), 1295–1302. https://doi.org/https://doi.org/10.2136/sssaj2005.0297

Mouazen, A. M.; Maleki, M. R.; Cockx, L.; Van Meirvenne, M.; Van Holm, L. H. J.; Merckx, R.; De Baerdemaeker, J.; Ramon, H. Optimum three-point linkage set up for improving the quality of soil spectra and the accuracy of soil phosphorus measured using an on-line visible and near infrared sensor. *Soil and Tillage Research* **2009**, 103(1), 144–152. https://doi.org/https://doi.org/10.1016/j.still.2008.10.006

Munnaf, M.; Nawar, S.; Mouazen, A. M. Estimation of Secondary Soil Properties by Fusion of Laboratory and On-Line Measured Vis–NIR Spectra. *Remote Sensing* **2019**, 11(23), 2819. https://doi.org/10.3390/rs11232819

Nawar, S.; Mouazen, A. M. Predictive performance of mobile vis-near infrared spectroscopy for key soil properties at different geographical scales by using spiking and data mining techniques. *CATENA* **2017**, 151, 118–129. https://doi.org/https://doi.org/10.1016/j.catena.2016.12.014

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

Riebe, D.; Erler, A.; Brinkmann, P.; Beitz, T.; Löhmannsröben, H. G.; Gebbers, R. Comparison of calibration approaches in laser-induced breakdown spectroscopy for proximal soil sensing in precision agriculture. *Sensors (Switzerland)* **2019**, 19(23), 1–15. https://doi.org/10.3390/s19235244

Rinnan, Å.; Berg, F. V. D.; Engelsen, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *TrAC Trends in Analytical Chemistry* **2009**, 28(10), 1201–1222. https://doi.org/https://doi.org/10.1016/j.trac.2009.07.007

Romero, D. J.; Ben-Dor, E.; Demattê, J. A. M.; Souza, A. B. E.; Vicente, L. E.; Tavares, T. R.; Martello, M.; Strabeli, T. F.; Barros, P. P. S., Fiorio, P. R.; Gallo, B. C.; Sato, M. V.; Eitelwein, M. T. Internal soil standard method for the Brazilian soil spectral library: Performance and proximate analysis. *Geoderma* **2018**, 312. https://doi.org/10.1016/j.geoderma.2017.09.014

Rossiter, D. G. Past, present & future of information technology in pedometrics. *Geoderma* **2018**, 324, 131–137. https://doi.org/https://doi.org/10.1016/j.geoderma.2018.03.009

Santos, F. R.; de Oliveira, J. F.; Bona, E.; Santos, J. V. F.; Barboza, G. M. C.; Melquiades, F. L. EDXRF spectral data combined with PLSR to determine some soil fertility indicators. *Microchemical Journal* **2020**, 152, 104275. https://doi.org/https://doi.org/10.1016/j.microc.2019.104275

Schäefer, C. E. G. R.; Fabris, J. D.; Ker, J. C. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. *Clay Minerals* **2008**, 43(1), 137–154. https://doi.org/10.1180/claymin.2008.043.1.11

Silva, E. A.; Weindorf, D. C.; Silva, S. H. G.; Ribeiro, B. T.; Poggere, G. C.; Carvalho, T. S.; Gonçalves, M. G. M.; Guilherme, L. R. G.; Curi, N. Advances in Tropical Soil Characterisation via Portable X-Ray Fluorescence Spectrometry. *Pedosphere* **2019**, 29(4), 468–482. https://doi.org/https://doi.org/10.1016/S1002-0160(19)60815-5

Silva, S. H. G.; Teixeira, A. F. dos S.; Menezes, M. D.; Guilherme, L. R. G.; Moreira, F. M. de S.; Curi, N.; Silva, S. H. G.; Teixeira, A. F. S.; Menezes, M. D. de, Guilherme, L. R. G.; Moreira, F. M. de S.; Curi, N. Multiple linear regression and random forest to predict and map soil properties using data from portable X-ray fluorescence spectrometer (pXRF). *Ciência e Agrotecnologia* **2017**, 41(6), 648–664. https://doi.org/10.1590/1413-70542017416010317

Stenberg, B.; Rossel, R. A. V.; Mouazen, A. M.; Wetterlind, J. Visible and Near Infrared Spectroscopy in Soil Science. In D. L. B. T.-A. Advances in Agronomy (Vol. 107, pp. 163–215). *Academic Press* **2010**. https://doi.org/10.1016/S0065-2113(10)07005-7

Tavares, T.R.; Nunes, L. C.; Alves, E. E. N.; de Almeida, E.; Maldaner, L. F.; Krug, F. J.; Carvalho, H. W. P.; Molin, J. P. Simplifying sample preparation for soil fertility analysis by x-ray fluorescence spectrometry. *Sensors* **2019**, 19(23). https://doi.org/10.3390/s19235066

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

Tavares, T.R.; Molin, J. P.; Nunes, L. C.; Alves, E. E. N.; Melquiades, F. L.; Carvalho, H. W. P.; Mouazen, A. M. Effect of x-ray tube configuration on measurement of key soil fertility attributes with XRF. *Remote Sensing* **2020**, 12(6). https://doi.org/10.3390/rs12060963

Tavares, T.R.; Mouazen, A. M.; Alves, E. E. N.; Dos Santos, F. R.; Melquiades, F. L.; Carvalho, H. W. P.; Molin, J. P. Assessing soil key fertility attributes using a portable X-ray fluorescence: A simple method to overcome matrix effect. *Agronomy* **2020**, 10(6). https://doi.org/10.3390/agronomy10060787

Van Raij, B.; Andrade, J. C.; Cantarela, H.; Quaggio, J. A. Análise química para avaliação de solos tropicais. IAC: Campinas, Brazil, 2001; 285p.

Villas-Boas, P. R.; Romano, R. A.; de Menezes Franco, M. A.; Ferreira, E. C.; Ferreira, E. J.; Crestana, S.; Milori, D. M. B. P. Laser-induced breakdown spectroscopy to determine soil texture: A fast analytical technique. *Geoderma* **2016**, 263, 195–202. https://doi.org/10.1016/j.geoderma.2015.09.018

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

Xu, D.; Zhao, R.; Li, S.; Chen, S.; Jiang, Q.; Zhou, L.; Shi, Z. Multi-sensor fusion for the determination of several soil properties in the Yangtze River Delta, China. *European Journal of Soil Science* **2019a**, 70(1), 162–173. https://doi.org/10.1111/ejss.12729

Xu, X.; Du, C.; Ma, F.; Shen, Y.; Wu, K.; Liang, D.; Zhou, J. Detection of soil organic matter from laser-induced breakdown spectroscopy (LIBS) and mid-infrared spectroscopy (FTIR-ATR) coupled with multivariate techniques. *Geoderma* **2019b**, 355(January). https://doi.org/10.1016/j.geoderma.2019.113905

Yu, K. Q.; Zhao, Y. R.; Liu, F.; He, Y. Laser-Induced Breakdown Spectroscopy Coupled with Multivariate Chemometrics for Variety Discrimination of Soil. *Scientific Reports* **2016**, 6(June), 1–10. https://doi.org/10.1038/srep27574

Zhang, Y.; Hartemink, A. E. Data fusion of vis–NIR and PXRF spectra to predict soil physical and chemical properties. *European Journal of Soil Science* **2020**, 71(3), 316–333. https://doi.org/https://doi.org/10.1111/ejss.12875

Connecting text for chapter 7

The influence of moisture content on sensor performance is a classic challenge for *in-situ* applications using soil sensors. Although this topic has been extensively studied for VNIR sensors, it has been little explored for XRF sensors, especially for the prediction of fertility attributes. Another challenge related to the *in-situ* application of XRF sensors is the scanning time (dwell time) typically used for soil analysis with XRF, which typically oscillates between 60 and 90 s. This analysis time is restrictive for *in-situ* applications of this sensor (*e.g.*, embedded in agricultural machines). In chapter 7, we evaluate the effect of reducing the scanning time, as well as the water increment, on XRF sensor`s performance.The present study is pioneering in showing the performance of an XRF sensor configured with a reduced scanning time (4 s) applied on wet samples for predicting fertility attributes.

Chapter 7 fulfills the third objective of this thesis, which aimed at investigating the effect of soil moisture content and scanning time on the XRF performance to predict soil fertility attributes.

CHAPTER 7. TOWARDS *IN-SITU* ANALYSIS USING XRF SENSOR FOR SOIL FERTILITY ASSESSMENT: EFFECTS OF SCANNING TIME REDUCTION AND SOIL MOISTURE

Abstract

To support future *in-situ*/on-the-go applications using X-ray fluorescence (XRF) sensors, the present chapter aimed at evaluating the trade-off among the XRF's performance for the prediction of fertility attributes and (i) the scanning time reduction and (ii) the increment of soil moisture content. 102 soil samples acquired in two Brazilian agricultural fields were considered for these analyses. To assess the effect of scanning time reduction, readings using a portable XRF were performed with its dwell time set at 90, 60, 30, 15, 10, 7, 4, and 2 s. For assessing the effect of moisture content, soil samples with moisture content of 0, 5, 10, 15, 20, and 25% g g^{-1} were scanned with XRF set with 4 s scanning time; in order to compare results a VNIR sensor was also included in this analysis. Our results showed that it is possible to make drastic reductions in scanning time (from 90 to 2s) maintaining satisfactory prediction performance [residual prediction deviation (RPD) > 1.40] of soil fertility attributes. XRF proved to be less affected by soil moisture compared to the VNIR technique. Regarding the moisture effect on XRF sensor's performance for fertility analysis, our results showed satisfactory prediction performances (RPD ≥ 1.45) for all fertility attributes on samples with up to 15% moisture content. However, the XRF gradually reduced its predictive performance by increasing the water level in the soil. Thus, methods to mitigate external effects on XRF data for more accurate *in-situ* applications might be useful, being necessary future research to find the best method for mitigating the effect of soil moisture content in XRF spectra. The study conducted in this chapter emphasize the potential of XRF for *in-situ* analysis of soil fertility attributes, being pioneering in showing that with a reduced sacanning time it is possible to obtain satisfactory prediction performances in wet samples.

Keywords: 1. Precision agriculture 2. On-the-go applications 3. Rapid soil analysis 4. Variable-rate applications.

7.1. Introduction

The application of X-ray fluorescence (XRF) sensors for soil fertility analysis has evolved rapidly in recent years (Lima et al., 2019; Nawar et al., 2019; Tavares et al., 2020a). The technique characterizes a wide range of soil elemental composition (*e.g.*, Ti, Fe, Cu, Ca, K, Si, among others), providing complementary information to other more common proximal soil sensing (PSS) techniques, such as mineralogical and organic constituents offered by visible and near infrared diffuse reflectance spectroscopy (VNIR) (O'Rourke et al., 2016). Due to the popularization of portable equipment associated with factory calibrations (*e.g.*, pre-programmed measurement packages), XRF sensors has been applied for in-field analysis with satisfactory results for heavy metal characterization in soils (Weindorf et al., 2013; Paulette et al., 2015) and geochemical evaluations in soil trenches (Weindorf et al., 2012; Stockmann et al., 2016).

Although XRF sensors are already applied *in-situ*, the scanning time used is typically around 30 to 90 s, which is quite contrasting when compared to the analysis time of other PSS techniques. For example, both apparent electrical conductivity (ECa) and VNIR techniques have an almost instantaneous scanning time (one s per scanning) that allows on-the-go data acquisitions with high spatial density (*e.g.*, 250 data points ha^{-1} at operating speeds around 4 m s^{-1}) (Molin and Tavares et al., 2019). On the other hand, on-the-go application using ion-selective electrodes (ISE) are an example closer to what would be realized for XRF, since the ISE needs a longer dwell time in contact with the sample to stabilize its reading (*e.g.*, approximately 10 to 15 s) (Adamchuk et al., 2007). This feature reduces the sensor's capability to acquire data at high spatial density in on-the-go operations. Kinematic data acquisitions using ISE systems usually measure about 15 to 30 data points ha^{-1} (Schirrmann et al., 2011). Hence, the normally used XRF scanning time limits analysis of rapid nature, such as on-the-go soil measurements, making it necessary the optimization of this instrumental condition to expand embedded applications of this sensor. To the best of our knowledge, the optimization of XRF sensor scanning time has not yet been addressed in the literature for soil fertility analysis.

Another challenge of *in-situ* analysis using soils sensors is to minimize the soil moisture effect that compromise part of the sensors' performance (Ge et al., 2005; Kuang and Mouazen, 2013). Strategies for calibrating of predictive models that are insensitive to the effect of soil moisture has received a great attention for *in-situ* analysis performed with VNIR sensors (Minasny et al., 2011; Kuang and Mouazen et al., 2013; Nawar et al., 2020). However, there is still a lack of information on the soil moisture effect in XRF data, especially for fertility attributes prediction. Horta et al. (2015) have mentioned that the XRF technique is less affected by soil moisture, compared to the VNIR technique. Ge et al. (2005) and Padilla et al. (2019) have complemented this information by indicating that correction of the moisture effect in XRF spectra is necessary only in samples with gravimetric moisture above 20%. However, this issue still needs further research to understand the effect of soil moisture on the prediction of fertility attributes using XRF data, especially for data collected with reduced scanning time.

In this backdrop, to support future *in-situ*/on-the-go applications using XRF sensors, the present chapter aimed at evaluating the trade-off among the XRF's performance for the prediction of fertility attributes and (i) the scanning time reduction and (ii) the increment of soil moisture. In addition, a VNIR sensor was also considered in the soil moisture effect analysis.

7.2. Material and Methods

The data analysis of this chapter was performed in two parts. In part 1, the effect of scanning time reduction on the XRF sensor's analytical performance was assessed, and in part 2, the effect of soil moisture was evaluated. The methodology applied in this chapter was diagrammatically presented in Figure 1.

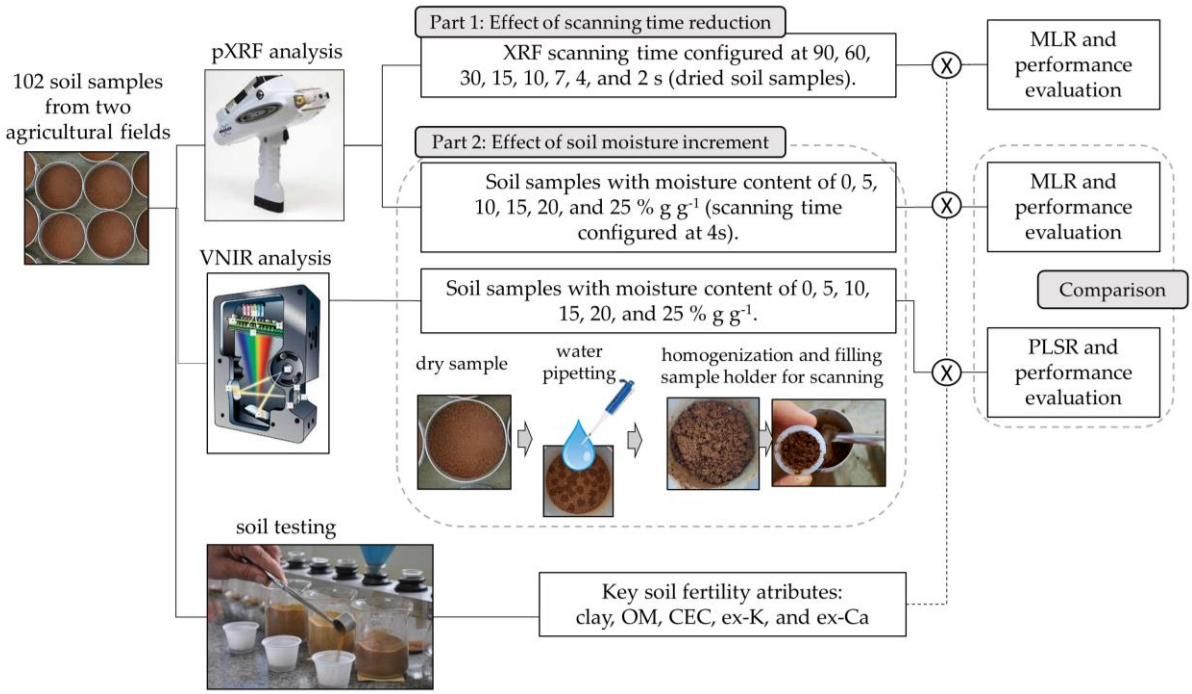


Figure 1. Framework of the methodology applied for evaluating, in part 1, the effect of scanning time reduction (part 1) and the influence of soil moisture increment (part 2) on the X-ray fluorescence (XRF) sensor's performance for the prection of clay, organic matter (OM), cation exchange capacity (CEC), exchangeable (ex-) K, and ex-Ca. In part 2, the influence of soil moisture increment was also evaluated for visible and near infrared diffuse reflectance spectroscopy (VNIR) sensor.

7.2.1. Study sites and soil samples

The study sites and soil samples used were described in the 4.2.1. Section (chapter 4).

7.2.2. Reference analyses

The contents of clay, cation exchange capacity (CEC), and the extractable (ex-) nutrients ex-K and ex-Ca were determined and used as Y-variables for the predictive modeling. The methods for these determinatons were described in the 4.2.2. Section (chapter 4).

7.2.3. Part 1: effect of scanning time reduction

7.2.3.1. XRF measurements with different scanning times

The equipment, instrumental conditions, and spectra pre-processing used are the same as those described in the 6.2.5. Section (chapter 6). The only exception was that, in the present study, the readings were performed with dwell time set at 90, 60, 30, 15, 10, 7, 4, and 2 s.

The characterization of XRF data as a function of scanning time reduction was performed by observing dispersion measures of the signal-to-noise ratio (SNR) of Al, Si, K, Ca, Ti, and Fe K-lines. These emission lines were chosen because they emit fluorescence at different energies (1.5, 1.7, 3.3, 3.7, 4.5, and 6.4 keV for Si, K, Ca, and Fe, respectivelly), representing different levels of influence of the matrix effect and the continuum spectrum of Rh X-ray tube (Tavares et al., 2020a,b).

7.2.3.2. Modeling

The set comprising 102 soil samples was divided into two groups, one for calibration (with 68 samples) and the other for validation (with 34 samples) of the predictive models, procedure performed using the Kennard–Stone algorithm (Kennard and Stone, 1969) executed on the soil fertility attributes (Y-variables). For each scanning time, calibration models were built using multiple linear regression (MLR). Each calibrated model was validated on the respective validation set of its scanning time, *e.g.*, models using XRF data acquired at 15 s were validated on XRF data also acquired at that scanning time. All the calibrations and validations were performed using the Unscrambler software, version 10.5.1 (Camo AS, Oslo, Norway).

The prediction performance was evaluated through the root mean square error (RMSE) and the residual prediction deviation (RPD), the latter was calculated as the ratio between the standard deviation (SD) of the laboratory measured soil property of interest and the RMSE in the prediction. Based on the RPD values, the prediction quality of developed models were classified into four classes adapted from Chang et al. (2001): poor models ($RPD < 1.40$), reasonable models ($1.40 \leq RPD < 2.00$), good models ($2.00 \leq RPD < 3.00$), and excellent models ($RPD \geq 3.00$).

7.2.4. Part 2: effect of soil moisture content

Soil samples with moisture content of 0, 5, 10, 15, 20, and 25% g g^{-1} were scanned with both VNIR and XRF sensors. Oven-dried samples (at 45 °C for 48 h) were considered at 0% moisture content

and used for moisture increment with deionized water. 15 g of soil were used for each round of water increment and the readings with both sensors were performed immediately after each water increment.

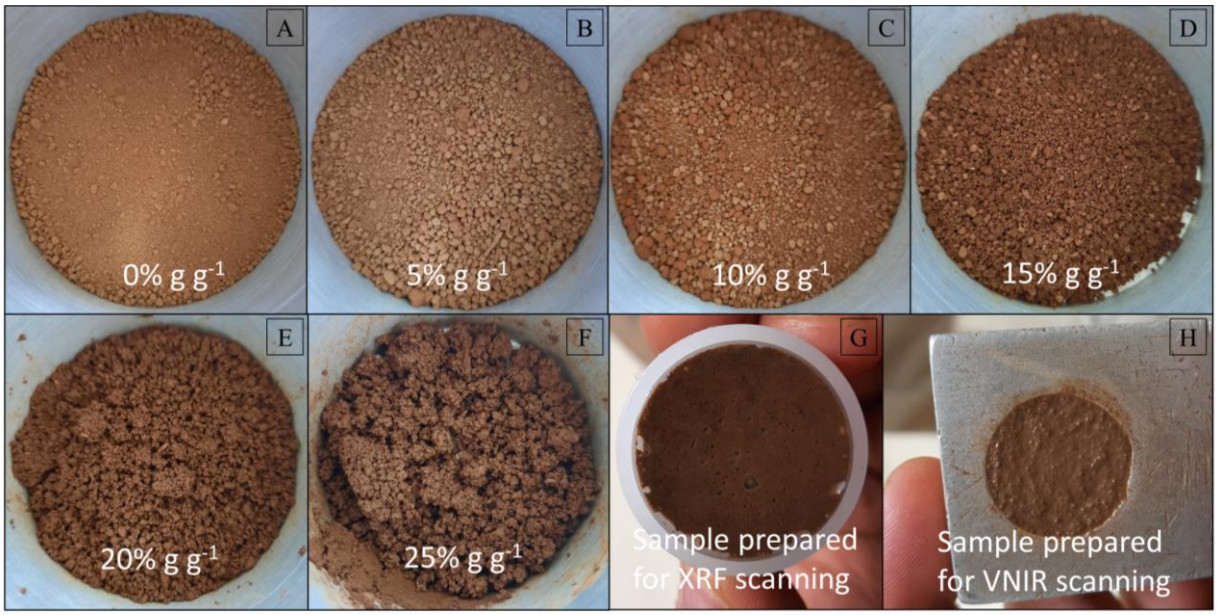


Figure 2. Soil sample (clay and sand content of 345 and 334 g dm^{-3}, respectively) with moisture content of 0, 5, 10, 15, 20, and 25 % g g^{-1} (A-F). The same soil sample with 25% moisture content in the sample holder for scanning with XRF (G) and VNIR (H) sensors.

7.2.4.1. XRF measurements and spectra pre-processing

The sample preparation, equipment, instrumental conditions, and XRF data pre-processing used in the present chapter are the same as those described in the 6.2.5. Section (chapter 6). The only exception is that the dwell time was set at 4 s in the present study.

7.2.4.2. VNIR measurements and spectra pre-processing

The sample preparation, equipment, instrumental characteristics, and spectra pre-processing used in the present study are the same as those described in the 6.2.4. Section (chapter 6).

7.2.4.3. Modeling

The same calibration and validation sets described above (in the 7.2.3.2. Section) were adopted in the present analysis. For evaluating the effect of moisture content o the XRF sensor's performance, prediction models were calibrated (using the calibration set) on samples with 0% moisture content and

validated (using the validation set) on samples with moisture content of 0, 5, 10, 15, 20, and 25%. The same figures of merit and RPD interpretation classes described in the 7.2.3.2. Section (present chapter) were also adopted in this evaluation.

7.3. Results

7.3.1. Laboratory Measured Soil Properties

The descriptive statistics and correlation matrix of the laboratory measured soil properties were presented in the 4.3.1.2. Section (chapter 4).

7.3.2. Effect of scanning time reduction on XRF sensor's performance

Figure 3 shows that reducing the sensor's dwell time leads to an increase in the SNR dispersion of the XRF emission lines, which is related to the increase of noise in spectra collected using reduced scanning time (Figure 3A). It can be seen that the increase in SNR dispersion occurs distinctly for different elements, in which Al, K, and Ca showed the highest dispersion [with coefficient of variation (CV) ranging from 1.2 to 25.3%] while Si, Ti, and Fe showed the lowest dispersion (with CV ranging from 0.3 to 4.7%). The greater dispersion must be related to elements with lower SNR values, since Al, K, and Ca had SNR values ranging from 2 to 22, whereas Si, Ti, and Fe had SNR values greater than 70. Among the elements with the highest SNR dispersion, K stood out with the highest CV values in all evaluated scanning times, with the only exception being the 2 s scanning time at which Ca had the highest CV. K was also the element with the lowest SNR, which ranged between 3 and 4.

Figure 4 shows the behavior of the prediction performance of clay, CEC, ex-K, and ex-Ca when reducing the XRF scanning time. Although the prediction error tends to increase as the dwell time decreases, the prediction performances of all fertility attributes remained satisfactory (RPD ≥ 1.92) at all evaluated scanning timesPredictions of clay, CEC, and ex-Ca maintained their good performance (RPD ≥ 2.00) over the entire scanning time reduction (from 90 to 2 s), showing an increase in RMSE ranging from 1.1 to 29.7 %. On the other hand, ex-K was the attribute that showed the greatest reduction in its prediction performance (with RMSE increasing between 24.3 and 133.1 %), which must be related to the reduced SNR of the K emission line.

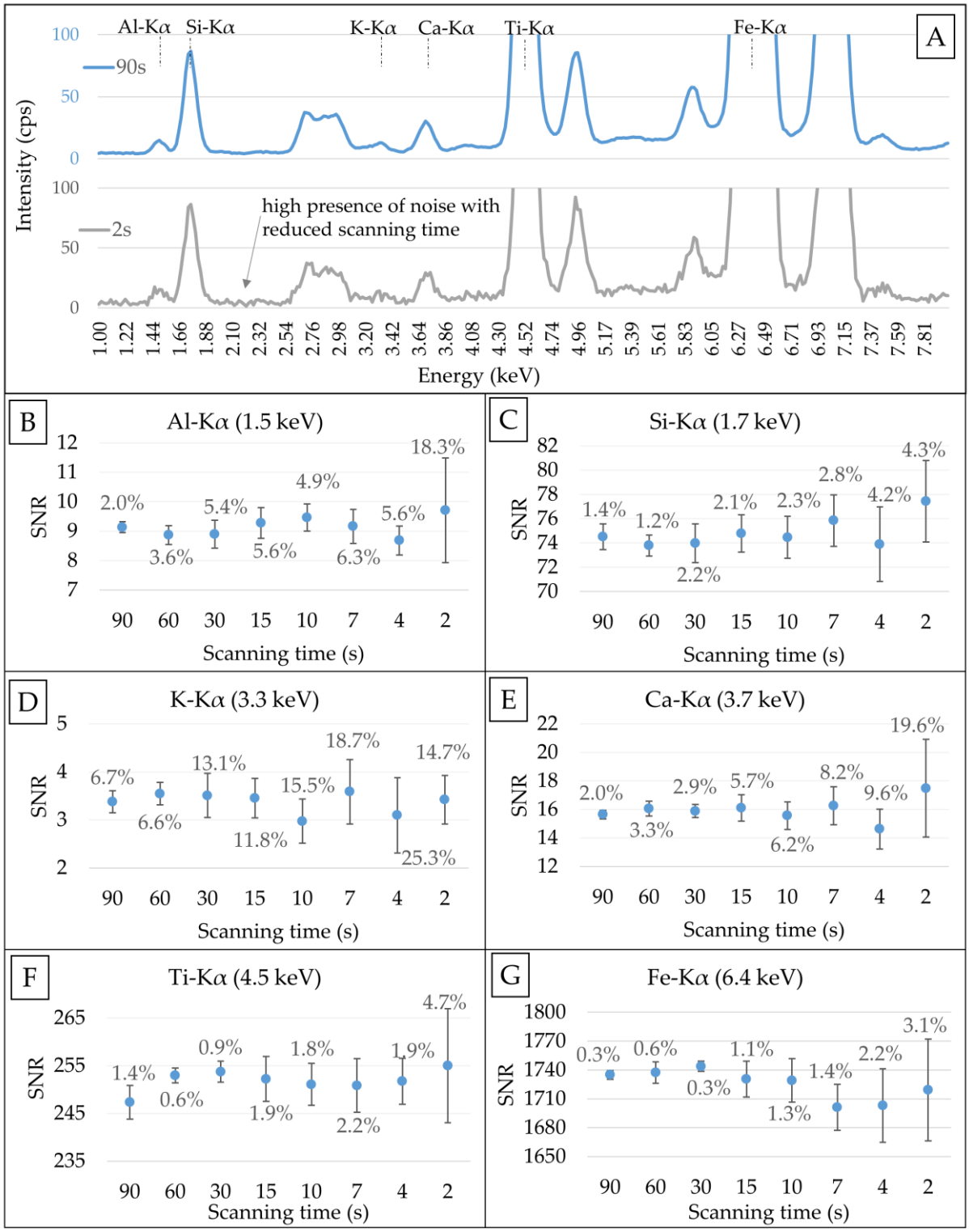


Figure 3. Effect of scanning time reduction on XRF spectra (A). Signal-to-noise ratio for the K-lines of Al (B), Si (C), K (D), Ca (E), Ti (F), and Fe (G) obtained at different scanning times (2, 4, 7, 10, 15, 30, 60, and 90 s). The bars represent the standard deviation and the values in percentage represent the coefficient of variation of five XRF measurements performed on the same soil sample after moving the sample cup position.

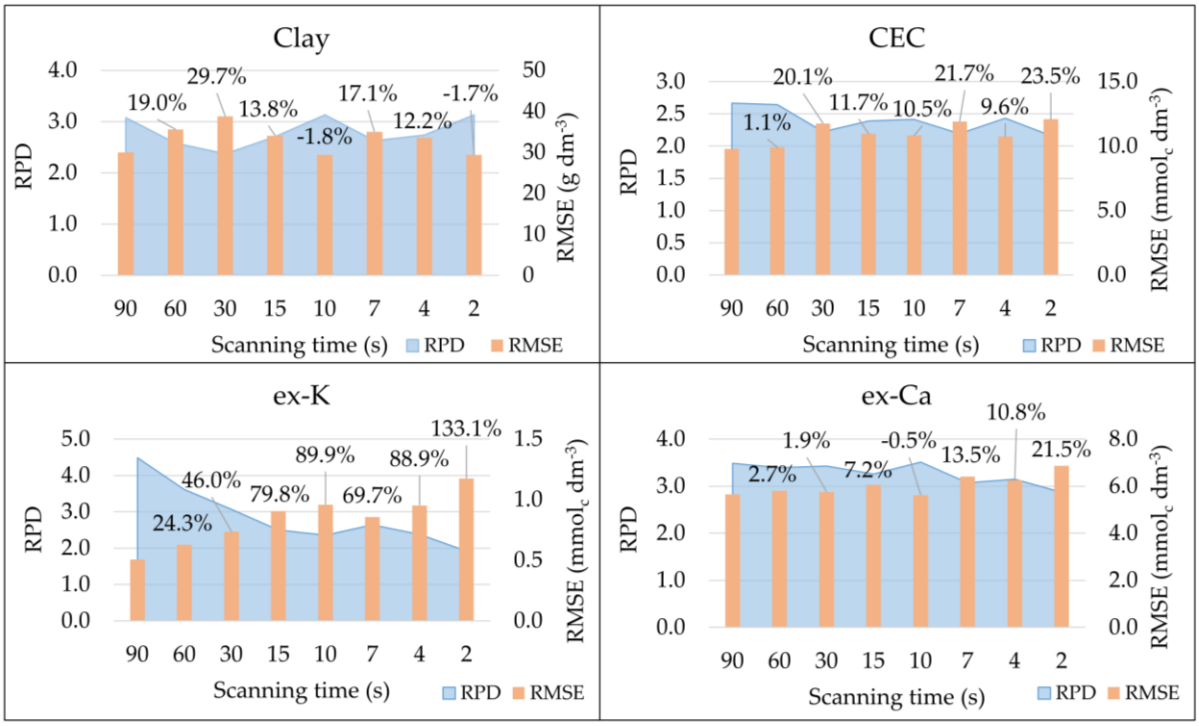


Figure 4. Effect of scanning time reduction on X-ray fluorescence (XRF) sensor performance for predicting (using the validation set, $n = 34$) clay, cation exchange capacity (CEC), and exchangeable (ex-) nutrients (ex-K and ex-Ca). The performance was evaluated via residual prediction deviation (RPD) and root-mean-square error (RMSE). The percentage values represent the variation of RMSE in relation to the performance obtained with 90s scanning time.

7.3.3. Effect of moisture content on VNIR and XRF sensors' performance

Figure 5 shows that the higher the moisture content in the soil sample, the lower is the intensity of its XRF lines. However, this behavior is inverted for the scattering peaks (Compton and Thomson scattering of Rh X-ray tube), which increase its intensity as the amount of water in the sample increases.

Regarding the prediction of fertility attributes using the XRF sensor, it is evident the reduction of its predictive performance as the moisture content increases. The prediction error of the XRF sensor increased between 7 and 50% in the first increment of water (samples with 5% moisture content), reaching an increase between 55 and 226% in samples with 25% moisture content. The VNIR sensor, on the other hand, proved to be more sensitive to the increment of moisture content, showing a more significant increase in error (between 1314 and 2796 % and 3738 and 10757 % for samples with 5 and 25 % moisture content, respectively). The lower water sensitivity of XRF sensor allowed satisfactory prediction performances (RPD ≥ 1.40) for all the fertility attributes evaluated in samples with up to 15% moisture content (Figure 6), which was not possible with the VNIR sensor that drastically reduced its predictive performance at the first addition of water (with prediction errors increasing by more than 1000% in samples with 5% moisture content).

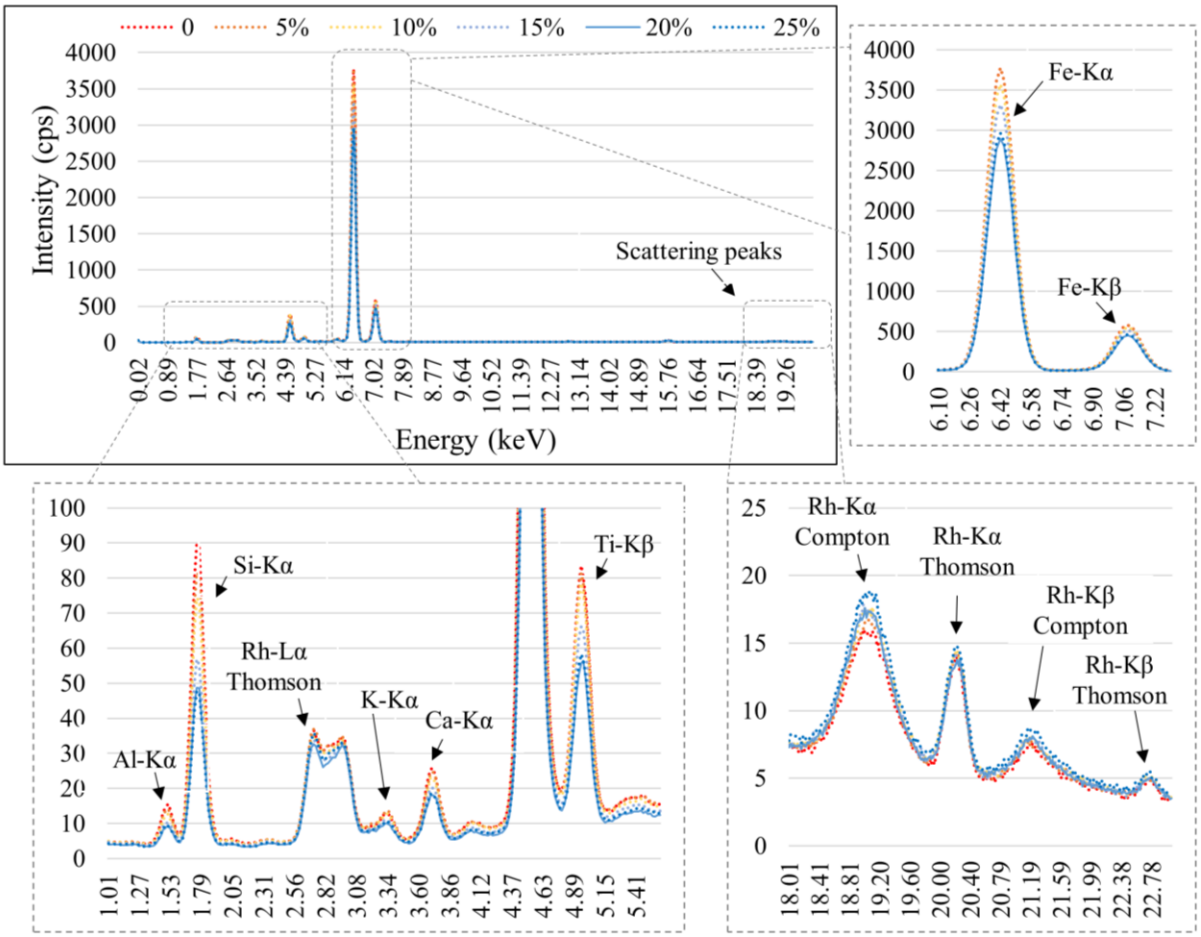


Figure 5. X-ray fluorescence emission of a dried soil sample (0%) and after increasing its gravimetric moisture content by 5, 10, 15, 20, and 25 %.

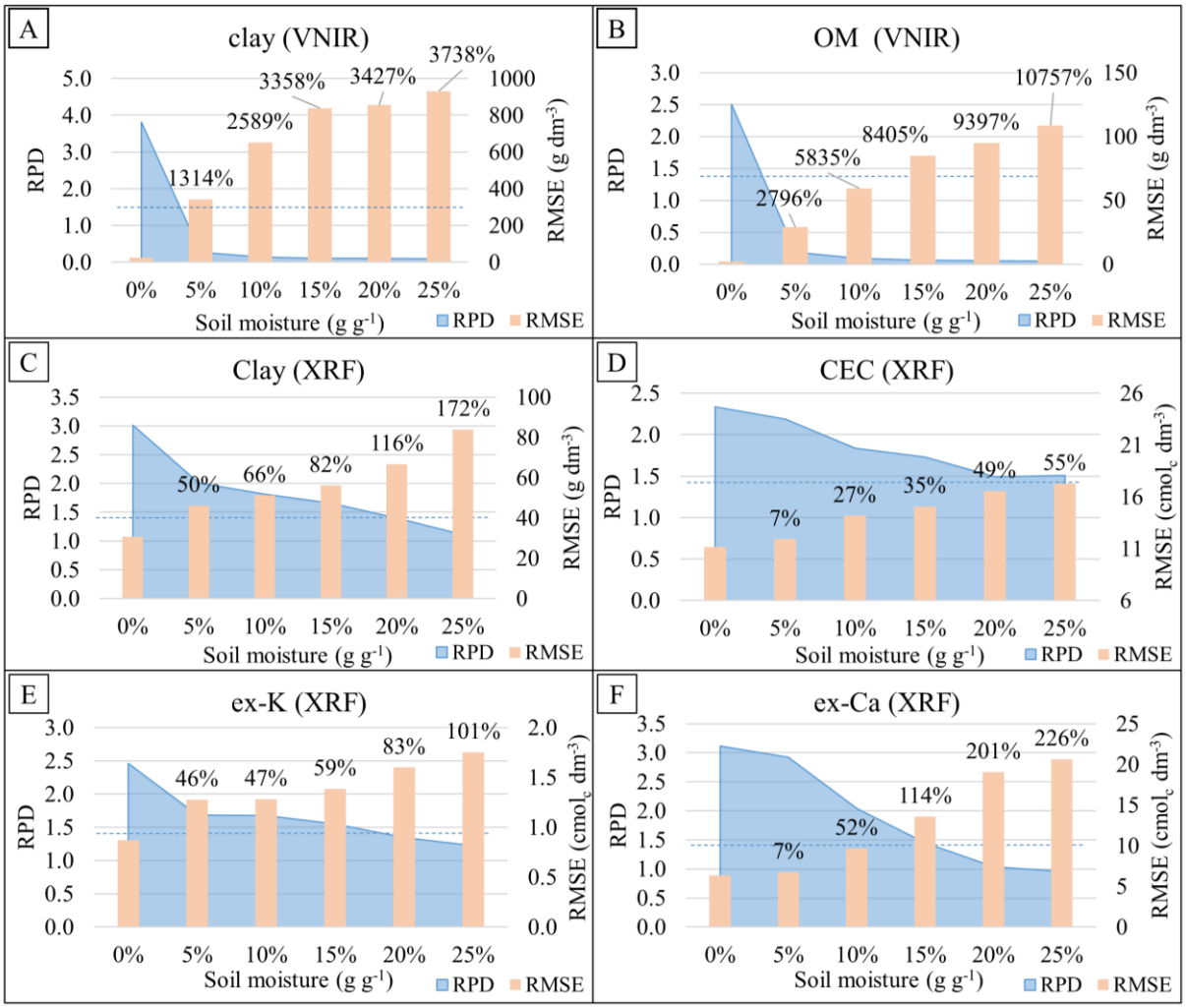


Figure 6. Effect of soil moisture increment on the performance of VNIR (A and B) and XRF (C, D, E, and F) sensors for predicting fertility attributes. Predictive models were calibrated on dry samples and validated on samples with different levels of gravimetric moisture (from 0 to 25%). The performance was evaluated via residual prediction deviation (RPD) and root-mean-square error (RMSE). The labels correspond to the increase in RMSE (in percentage) compared to the RMSE obtained on dry samples (0% of moisture content). The blue dashed line indicates a RPD of 1.4, values below this threshold indicate poor performance. XRF scanning time was configured at 4 s.

7.4. Discussion

Our results show that the reduction of XRF dwell time reduces its reading accuracy, which is an expected behavior as it is a typical effect of XRF analysis with reduced scanning time (Weindorf and Chakraborty, 2016). Nevertheless, the accuracy of the sensor for predicting fertility attributes does not degrade expressively and drastic reductions of the XRF scanning time (*e.g.*, with scanning time of up to 2 s) can be applied while maintaining satisfactory predictive performances (RPD ≥ 1.92). Alternatively, the reduced precision of individual measurements can be mitigated by replicating the data acquisition, especially if the attribute of interest depends on emission lines with reduced SNR (as observed for ex-K

in the present study). It is worth mentioning that SNR lower than 10 might be considered critical, as indicated by Currie (1968).

It is a consensus that the practicality promoted by rapid readings enables a massive increase in the spatial density of *in-situ* analyses, which in turn allows for a better characterization of soil spatial variability as a whole (Viscarra Rossel et al., 2011; Molin e Tavares, 2019). The results provided in this chapter show the robustness of XRF performance for rapid analysis of fertility attributes. This knowledge is not widespread among XRF users as it is common to use pre-programmed measurement packages (Horta et al., 2015; O'Rourke et al., 2016; Lima et al., 2019; Nawar et al., 2019), which are factory calibrations associated with a pre-established scanning time, generally between 60 to 90 s (Weindorf and Chakraborty, 2016).

Unlike laboratory measurements that are mainly conducted on dry and sieved samples, sensors applied in field conditions need to deal with external factors (*e.g.*, soil moisture, soil roughness, etc) that influence sensors' output (Horta et al., 2015, Mouazen and Al-Asadi, 2018). Soil moisture is one of the most studied external factor due to its marked influence on sensor's performance (Angelopoulou et al., 2020). Comparing XRF and VNIR sensors on wet soil samples, Horta et al. (2015) mentioned that the XRF is less affected by soil moisture compared to the VNIR technique, which was corroborated by the results obtained in the present study. Regarding the moisture effect on XRF sensor's performance for fertility analysis, our results showed satisfactory prediction performances (RPD ≥ 1.45) for all fertility attributes on samples with up to 15% moisture content. Thus, predictive models calibrated on dry samples can be satisfactorily replicated on wet samples with moisture content between 0 and 15 %. Nevertheless, it was also observed that the prediction performance of the XRF sensor already decreases (incresing the error between 7 and 50 %) at the first water increment (from 0 to 5 % moisture). So, in cases where it is necessary to maintain the prediction accuracy, methods adopted to minimize external effects such as external parameter orthogonalization (EPO) (Roger et al., 2003) and direct standardization method (Wang et al., 1995) should be considered. Although the abovementioned methods have shown promising results for VNIR sensors providing moisture-insensitive predictions of soil attributes (Minasny et al., 2011; Kuang and Mouazen et al., 2013; Nawar et al., 2020), they have not yet been explored for XRF, which should be encouraged in future research. In addition, studies applying XRF sensors directly in the field are also necessary in order to consider the influence of other factors, such as soil roughness, ambient temperature, residues (*e.g.*, stones and straw), and movement during spectral acquisition, which can also interfere *in-situ* measurements (Angelopoulou et al., 2020).

Finally, it is worth mentioning that typically moisture content may vary from 5 to 25 % g g^{-1} for sandy and from 5 to 30 % g g^{-1} for clay soils when the soil moisture is appropriate for conducting fieldwork (Adamchuk et al., 1999). In Brazilian tropical soils, soil testing is commonly performed in

the intercrop season, a period with drastically reduced rainfall, when soil moisture is usually not high (*e.g.*, generally smaller than 15 % g g^{-1}). This particularity would benefit the application of XRF in Brazilian tropical soils, since all the fertility attributes evaluated showed satisfactory prediction performance at 15 % g g^{-1} of soil moisture content. Lastly, our results allow raising the new hypothesis: "can models for predicting fertility attributes generated in dried samples be satisfactorily replicated in wet samples with up to 15 % g g^{-1} of water content?".

7.5. Conclusion

The reduction of X-ray fluorescence (XRF) sensor scanning time increases the standard deviation of XRF emission lines that, in turn, reduces the precision of sensor readings. In spite of that, our results showed that it is possible to make drastic reductions in scanning time (from 90 to 2s) maintaining satisfactory prediction performance [residual prediction deviation (RPD) ≥ 1.92] of soil fertility attributes.

The XRF proved to be less sensitive to soil moisture increment when compared to the VNIR sensor. For example, in a sample with 10% of moisture content, the XRF still achieved a satisfactory performance (RPD ≥ 1.45) for all the attributes evaluated, which was not reached with the VNIR sensor nor even using samples at 5% of moisture content. Despite this, the XRF gradually reduced its predictive performance by increasing the water level in the soil, *e.g.*, at the first level of moisture increment performed in this study (from 0 to 5% of gravimetric moisture) the prediciton error increased among 7 and 52%, depending on the attribute assessed. Thus, despite the lower sensitivity of the XRF sensor to soil moisture compared to VNIR, it is necessary to use methods to mitigate external effects on XRF data for more accurate *in-situ* applications.

The observed results emphasize the potential of XRF for *in-situ* analysis of soil fertility attributes, since with a reduced sacanning time it is possible to obtain satisfactory prediction performances in wet samples. Furthermore, our results encourage future research to find the best method for mitigating the effect of soil moisture content in XRF spectra.

References

Adamchuk, V.I.; Morgan, M.T.; Ess, D.R. An automated sampling system for measuring soil pH. *Transactions of the ASAE* **1999**, 42(4), p.885.

Adamchuk, V. I.; Lund, E. D.; Reed, T. M.; Ferguson, R. B. Evaluation of an on-the-go technology for soil pH mapping. *Precision Agriculture* **2007**, 8(3), 139-149, doi: 10.1007/s11119-007-9034-0.

Chang, C.-W.; Laird, D. A.; Mausbach, M. J.; Hurburgh, C. R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal* **2001**, 65(2), 480–490. https://doi.org/10.2136/sssaj2001.652480x

Currie, L.A. Limits for qualitative detection and quantitative determination. Application to radiochemistry. Anal. Chem. 1968, 40, 586–593.

Ge, L.; Lai, W.; Lin, Y. Influence of and correction for moisture in rocks, soils and sediments on *in-situ* XRF analysis. *X-Ray Spectrometry: An International Journal* **2005**, 34(1), 28-34, doi: 10.1002/xrs.782.

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

Kennard, R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, 11(1), 137–148. https://doi.org/10.1080/00401706.1969.10490666

Kuang, B.; Mouazen, A. M. Non-biased prediction of soil organic carbon and total nitrogen with vis–NIR spectroscopy, as affected by soil moisture content and texture. *Biosystems Engineering* **2013**, 114(3), 249-258, doi: 10.1016/j.biosystemseng.2013.01.005.

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

Minasny, B.; McBratney, A. B.; Bellon-Maurel, V.; Roger, J. M.; Gobrecht, A.; Ferrand, L.; Joalland, S. Removing the effect of soil moisture from NIR diffuse reflectance spectra for the prediction of soil organic carbon. *Geoderma* **2011**, 167, 118-124, doi: 10.1016/j.geoderma.2011.09.008.

Mouazen, A. M.; Al-Asadi, R. A. Influence of soil moisture content on assessment of bulk density with combined frequency domain reflectometry and visible and near infrared spectroscopy under semi field conditions. *Soil Till. Res.* **2018**, 176, 95-103, doi: 10.1016/j.still.2017.11.002.

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

Nawar, S.; Abdul Munnaf, M.; Mouazen, A.M. Machine Learning Based On-Line Prediction of Soil Organic Carbon after Removal of Soil Moisture Effect. *Remote Sens.* **2020**, 12, p.1308, doi: 10.3390/rs12081308.

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

Roger, J.M.; Chauchard, F.; Bellon-Maurel, V. EPO–PLS external parameter orthogonalisation of PLS application to temperature-independent measurement of sugar content of intact fruits. *Chemometrics and Intelligent Laboratory Systems* **2003**, 66(2): 191–204, doi: https://doi.org/10.1016/S0169-7439(03)00051-0.

Padilla, J.T.; Hormes, J.; Selim, H.M. Use of portable XRF: Effect of thickness and antecedent moisture of soils on measured concentration of trace elements. *Geoderma* **2019**, 337, 143-149, doi: 10.1016/j.geoderma.2018.09.022.

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

Schirrmann, M.; Gebbers, R.; Kramer, E.; Seidel, J. Soil pH mapping with an on-the-go sensor. *Sensors* **2011**, 11(1), 573-598, doi: https://doi.org/10.3390/s110100573.

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

Tavares, T.R.; Mouazen, A. M.; Alves, E. E. N.; Dos Santos, F. R.; Melquiades, F. L.; Carvalho, H. W. P.; Molin, J. P. Assessing soil key fertility attributes using a portable X-ray fluorescence: A simple method to overcome matrix effect. *Agronomy* **2020a**, 10(6). https://doi.org/10.3390/agronomy10060787

Tavares, T.R.; Molin, J. P.; Nunes, L. C.; Alves, E. E. N.; Melquiades, F. L.; Carvalho, H. W. P.; Mouazen, A. M. Effect of x-ray tube configuration on measurement of key soil fertility attributes with XRF. *Remote Sensing* **2020b**, 12(6). https://doi.org/10.3390/rs12060963

Viscarra Rossel, R.; Adamchuk, V.I.; Sudduth, K.A.; McKenzie, N.J.; Lobsey, C. Proximal soil sensing: An effective approach for soil measurements in space and time. *Adv. Agron.* **2011**, 113, 243–291, doi: 10.1016/B978-0-12-386473-4.00005-1

Wang, Z.; Dean, T.; Kowalski, B.R. Additive background correction in multivariate instrument standardization. *Analytical Chemistry* **1995**, 67, 2379–2385, doi: https://doi.org/10.1021/ac00110a009.

Weindorf, D.C.; Zhu, Y.; Mc Daniel, P.; Valerio, M.; Lynn, L.; Michaelson, G.; Clark, M.; Ping, C.L. Characterizing soils via portable x-ray fluorescence spectrometer: 2. Spodic and Albic horizons. *Geoderma* **2012**, 189, 268–277, doi:10.1016/j.geoderma.2012.06.034.

Weindorf, D. C.; Paulette, L.; Man, T. *In-situ* assessment of metal contamination via portable X-ray fluorescence spectroscopy: Zlatna, Romania. *Environmental pollution* **2013**, 182, 92–100, doi: https://doi.org/10.1016/j.envpol.2013.07.008

CHAPTER 8. FINAL CONSIDERATIONS

The necessity to intensify the density of soil information without relying exclusively on georeferenced sampling and costly laboratory testing represents the soil sampling dilemma of precision agriculture (PA). Spectro-analytical techniques that allow direct analysis of samples, *e.g.* those requiring minimal or no sample preparation, are promising for a rapid, cost-effective, and environmentally-friendly soil fertility characterisation, being alternatives for overcoming the soil sampling dilemma. The maturation of direct analysis techniques, such as visible and near infrared diffuse reflectance spectroscopy (VNIR), X-ray fluorescence spectroscopy (XRF), and laser induced breakdown spectroscopy (LIBS), would enable more accurate mapping of agricultural fields by means of *in-situ* or mobile laboratory analysis, which would represent new paradigm in soil management. A systematic review of this problematic in Brazilian tropical soils was presented in the *second chapter*, giving special attention to VNIR, XRF, and LIBS spectroscopies that were further explored in the subsequent chapters. For this thesis, the following three hypotheses were raised: (i) elemental analysis sensors (*e.g.*, XRF and LIBS) present greater potential for predicting soil chemical attributes compared to VNIR sensor; (ii) the combination of data provided by VNIR sensors with those from elemental analysis sensors will provide more comprehensive and accurate information about key fertility attributes compared to the single-sensor approaches; and (iii) although there are some critical factor that influences XRF's prediction accuracy in applications directly in the field (*e.g.*, sample moisture content and XRF scanning time), it is possible to establish strategies to maintain a satisfactory sensor performance. To test these hypotheses, we selected a set of 102 soil samples, with wide ranges of key fertility attributes, that was acquired in two Brazilian agricultural fields (Oxisol and Lixisol) with considerable textural variation. The overall findings associated with this project are described below according to the chapter in which they were presented.

In the *third chapter*, our results showed that it is possible to evaluate loose soil samples that were only dried and sieved (≤ 2 mm), obtaining satisfactory prediction results for extractable (ex-) K and ex-Ca [coefficient of determination (R^2) ≥ 0.76]. This simplified sample preparation performed only slightly worse than the analyses executed on pelletized soil samples ($R^2 \geq 0.78$), which is the ideal sample preparation for XRF analyses. Our results emphasize the flexibility of XRF technique regarding sample preparation and encourages its integrated use with VNIR technique, which is also compatible with loose soil analysis. In the *fourth chapter*, we suggested a novel methodology for XRF data acquisition and processing in order to assess soil fertility attributes. The proposed method consists in setting a Rh-anode X-ray tube at 35 kV of voltage and 7 μA of current. Then, for data processing, K-lines of elements commonly observed in agricultural soils (*e.g.*, Al-Kα, Si-Kα, K-Kα, Ca-Kα, Ti-Kα, Mn-Kα, Fe-Kα, Ni-

Kα, and Cu-Kα) are selected according to their signal-to-noise ratio (> 10) and normalized by the Compton peak to mitigate the matrix effect. Finally, prediction models are calibrated using a reduced number of variables (the normalized K-lines and the scattering peaks Rh-Lα Thomson, Rh-Kα Compton, and Rh-Kα Thomson) by means of multiple linear regression (MLR). This simple and transparent method allows for further optimizations related to XRF parameters necessary for *in-situ* applications, such as reducing its scanning time.

In the *fifth chapter*, we evaluate the performance of a benchtop LIBS sensor for predicting key soil fertility attributes. In general, predictions with satisfactory performance [residual prediction deviation (RPD) ≥ 1.82] were obtained for eight out of nine soil attributes evaluated [clay, organic matter (OM), base saturation (V), cation exchange capacity (CEC), ex-P, ex-K, ex-Ca, and ex-Mg], being pH the only exception that showed poor predictive performance (RPD ≤ 1.12). In addition, it was suggested a methodology to pre-process and model LIBS data that consists of selecting variables with the interval successive projections algorithm (iSPA) followed by model calibration via partial least squares regression (PLS). LIBS data is characterized by a very high spectral resolution (*e.g.*, usually around tens of thousands data points) and the suggested method proved to be an efficient and accurate approach to deal with this issue using reduced number of variables. Furthermore, we highlighted in the discussion that although LIBS proved to be efficient for predicting fertility attributes in tropical soils, further research is encouraged to find solutions that allows LIBS applications in soil samples with simpler sample preparation than soil pelletizing. The need for sample pelletizing makes *in-situ* applications with LIBS unfeasible, overcoming this limitation would expand its application possibilities in PA.

The individual and combined performance of VNIR, XRF, and LIBS for predicting key soil fertility attributes was assessed in the *sixth chapter*. The individual applications of XRF and LIBS sensors were performed using the optimized approaches provided by the chapters 3, 4, and 5. Results showed VNIR as individual technique to be the best for the prediction of clay and OM content (2.61 ≤ RPD ≤ 3.37), while the chemical attributes CEC, V, ex-P, ex-K, ex-Ca, and ex-Mg were better predicted (1.82 ≤ RPD ≤ 4.82) by elemental analysis techniques (confirming the *first hypothesis* of this thesis). Only pH cannot be predicted, regardless the technique. The combined use of sensors, with individual prediction models (*e.g.*, without using data fusion), allowed us to increase the range of soil fertility attributes determined with good and excellent predictions, emphasizing the complementary relationship of their data. At the same time, our results also showed that data fusion will not necessarily result in optimal prediction accuracy, compared to individual sensors. For example, the attributes OM, V, and ex-P were best predicted using single-sensor approaches; on the other hand, the attributes clay, CEC, pH, ex-K, ex-Ca, and ex-Mg were overall best predicted using multi-sensor approaches. Furthermore, the best combination of sensors for predicting key fertility attributes proved to be attribute-specific, *e.g.*, ex-K, ex-Ca and ex-Mg, were best

predicted using two-sensor fusion approach (VNIR + XRF for ex-K and XRF + LIBS for ex-Ca and ex-Mg), while clay and CEC were best predicted with the three-sensor fusion approach (VNIR + XRF + LIBS). Hence, the findings obtained in the chapter 6 confirm that combining the data provided by VNIR, XRF, and LIBS sensors provide more comprehensive information about key fertility attributes compared to the single-sensor approaches. However, the findings also evidence that data fusion is not always more accurate than approaches using individual sensors, which leads us to refute the *second hypothesis* of this thesis. Lastly, we emphasize that may be too soon to establish a clear trend to recommend the optimal set of sensors for predicting one or a set of soil attribute, which requires testing large number of data sets, with different soil types, soil origin, concentration range, and agricultural practices.

In the *seventh chapter*, we evaluate the applicability of XRF sensors for *in-situ* analysis, focusing on the effect of soil moisture content and scanning time on the XRF sensor's performance. Our results proved that it is possible to drastically reduce the scanning time (from 90 to 2s) while the performance for predicting fertility attributes is kept satisfactory (RPD > 1.40). Regarding the moisture effect, XRF proved to be less sensitive to soil moisture increment when compared to the VNIR sensor. Using XRF configured with scanning time of 4 s, satisfactory performances for clay, ex-K, and ex-Ca prediction were maintained when analysing samples with gravimetric moisture up to 15%, which was not possible for VNIR sensor. Despite this, the XRF gradually reduced its predictive performance by increasing the water level in the soil, which shows the importance of evaluating methods to mitigate this external effect for more accurate *in-situ* applications. Our study shows the potential of XRF for rapid *in-situ* analysis of soil fertility attributes and confirms the *third hypothesis* raised in this thesis.

This project brought some advances for using VNIR, XRF, and LIBS sensors in Brazilian tropical soils as rapid and environmentally-friendly analytical methods for characterizing fertility attributes. Benefits and drawbacks of sensor-based approaches to assess fertility attributes have been evidenced in the chapters of this thesis. In general, one of the main limitations is that sensors data are usually related to some external factors (*e.g.*, chemical and physical matrix effect, soil moisture in some cases, among others), which may compromise part of the sensors' predictive performance and can be a problem for replicating the calibrated models in subsequent analyses. Therefore, we would like to suggest that future research address the temporal stability of predictive models generated with sigle- and multi-sensor approaches that, to the best of our knowledge, has not been explored so far in the literature. Understanding temporal aspects of predictive models performance will be the basis for creating soil spectral libraries using XRF/LIBS data for soil fertility analysis, necessary to expand the development of *in-situ* and mobile laboratory analysis. Finally, it is worth mentioning that in view of the massive demand for fertilizers and lime in Brazilian soils, even as the worldwide calls for food security and sustainable farming systems, the development of successful sensor-based approaches for

soil management should be seen as a necessity and further research should be encouraged in Brazilian tropical soils.