



UNIVERSIDADE
ESTADUAL DE LONDRINA

FELIPE RODRIGUES DOS SANTOS

**USO DE FLUORESCÊNCIA DE RAIOS X E ANÁLISE
MULTIVARIADA NA DETERMINAÇÃO DE ATRIBUTOS DE
FERTILIDADE DE SOLOS EM UMA TOPOSSEQUÊNCIA
DERIVADA DO BASALTO**

Londrina
2020

FELIPE RODRIGUES DOS SANTOS

**USO DE FLUORESCÊNCIA DE RAIOS X E ANÁLISE
MULTIVARIADA NA DETERMINAÇÃO DE ATRIBUTOS DE
FERTILIDADE DE SOLOS EM UMA TOPOSSEQUÊNCIA
DERIVADA DO BASALTO**

Tese apresentada ao Programa de Pós-graduação em Física da Universidade Estadual de Londrina - UEL, como requisito parcial para a obtenção do título de Doutor.

Orientador: Prof. Dr. Fábio Luiz Melquiades
Coorientadora: Dra. Graziela Moraes de
Cesare Barbosa

Londrina
2020

Ficha de identificação da obra elaborada pelo autor, através do Programa de Geração Automática do Sistema de Bibliotecas da UEL

D722 dos Santos, Felipe Rodrigues.
USO DE FLUORESCÊNCIA DE RAIOS X E ANÁLISE MULTIVARIADA NA DETERMINAÇÃO DE ATRIBUTOS DE FERTILIDADE DE SOLOS EM UMA TOPOSSEQUÊNCIA DERIVADA DO BASALTO / Felipe Rodrigues dos Santos. - Londrina, 2020.
172 f.

Orientador: Fábio Luiz Melquiades.
Coorientador: Graziela Moraes de Cesare Barbosa.
Tese (Doutorado em Física) - Universidade Estadual de Londrina, Centro de Ciências Exatas, Programa de Pós-Graduação em Física, 2020.
Inclui bibliografia.

1. Fluorescência de raios X - Tese. 2. Quimiometria - Tese. 3. Atributos de fertilidade do solo - Tese. I. Melquiades, Fábio Luiz. II. de Cesare Barbosa, Graziela Moraes . III. Universidade Estadual de Londrina. Centro de Ciências Exatas. Programa de Pós-Graduação em Física. IV. Título.

CDU 53

FELIPE RODRIGUES DOS SANTOS

**USO DE FLUORESCÊNCIA DE RAIOS X E ANÁLISE MULTIVARIADA
NA DETERMINAÇÃO DE ATRIBUTOS DE FERTILIDADE DE SOLOS
EM UMA TOPOSSEQUÊNCIA DERIVADA DO BASALTO**

Tese apresentada ao Programa de Pós-graduação em Física da Universidade Estadual de Londrina - UEL, como requisito parcial para a obtenção do título de Doutor.

BANCA EXAMINADORA

Orientador: Prof. Dr. Fábio Luiz Melquiades
Universidade Estadual de Londrina - UEL

Prof. Dr. Marcelo Martins de Sena
Universidade Federal de Minas Gerais - UFMG

Prof. Dr. Hudson Wallace Pereira de Carvalho
Universidade de São Paulo - USP

Prof. Dr. Avacir Casanova Andrello
Universidade Estadual de Londrina - UEL

Prof. Dr. Paulo Sérgio Parreira
Universidade Estadual de Londrina – UEL

Londrina, 08 de Outubro de 2020.

AGRADECIMENTOS

Inicialmente eu gostaria de agradecer ao meu orientador Prof. Dr. Fábio Luiz Melquiades pela oportunidade, incentivo e pelos preciosos conhecimentos compartilhados durante esses 10 anos, que desde a iniciação científica tem sido fundamental para a minha formação.

A minha coorientadora Dra. Graziela Moraes de Cezare Barbosa, ao Dr. José Francirlei de Oliveria e ao José Victor Freitas que contribuíram muito para a realização desse trabalho, disponibilizando as amostras e análises convencionais. Agradeço por toda a contribuição e discussões que foram fundamentais para o desenvolvimento do trabalho.

Ao Prof. Dr. Evandro Bona por toda a colaboração e discussão com relação às análises multivariadas.

A todos os colegas do Laboratório de Física Nuclear Aplicada da UEL, em especial ao Guilherme, Aline, Leonardo, Matheus e Débora que me auxiliaram nas medidas e preparação das amostras e ao Sandro e Fábio Lopes pelos cafezinhos da tarde e todo o momento de descontração.

Aos colegas da pós-graduação que estiveram comigo nos momentos de estudo e resolução de listas.

A todos os professores, desde o ensino básico até a pós-graduação, que contribuíram para a minha formação.

Ao CNPq pela concessão da bolsa de doutorado durante esse período, a qual foi fundamental para a realização desse trabalho.

A todas as instituições que de alguma forma contribuíram para realização desse trabalho: ao Laboratório de Solos e Tecido Vegetal do IAPAR onde as análises convencionais foram realizadas, ao Laboratório de Análises de Raios X da UEL onde foram realizadas as medidas por EDXRF e a toda estrutura disponibilizada pelo Laboratório de Física Nuclear Aplicada da UEL.

A todos os meus amigos, em especial Halisson e William, pelo companheirismo, apoio e amizade, que mesmo de longe sempre estiveram presentes gerando bons momentos de descontração.

Não poderia deixar de agradecer a minha cachorrinha Agnes que esteve comigo durante todo esse período dando amor, carinho e alegria mesmo nos dias mais cansativos e difíceis.

A minha companheira Andressa, por estar ao meu lado durante todos esses anos, compartilhando momentos importantes e me apoiando em tudo. Muito obrigado pelo incentivo, amor, atenção, compreensão e encorajamento nos momentos difíceis.

E por fim agradeço a toda a minha família, minha irmã Bianca, meus padrinhos Carlos e Marli, minha avó Dona Aneide e em especial aos meus pais, Valdemir (*in memoriam*) e Cleuse, que são os principais responsáveis por todas as conquistas da minha vida, sendo grandes incentivadores e dando muita força para que nunca desistisse dos meus objetivos. Orgulho-me muito de vocês e serei eternamente grato a tudo o que foi feito por mim.

DOS SANTOS, Felipe Rodrigues. **Uso de fluorescência de raios x e análise multivariada na determinação de atributos de fertilidade de solos em uma topossequência derivada do basalto**. 2020. 172 f. Tese (Doutorado em Física) – Universidade Estadual de Londrina, Londrina, 2020.

RESUMO

A fluorescência de raios X (XRF) combinada com calibração multivariada é uma das metodologias que tem sido propostas para análise dos atributos de fertilidade do solo de forma rápida, barata e com baixo impacto ambiental. No entanto, a maioria dos estudos relatados na literatura utilizam dados de concentrações elementares determinados pela técnica, o que pode causar a perda de informações úteis contidas em seus espectros. Além disso, em alguns casos, essas concentrações elementares podem não ser acuradas. Nesse sentido, o uso de dados espectrais de XRF pode ser mais adequado no desenvolvimento de modelos de calibração multivariada para determinação de parâmetros do solo. No entanto, nem todas as variáveis do espectro contribuem para a modelagem do parâmetro de interesse. Além disso, ainda não existe um procedimento bem estabelecido para o pré-processamento desses dados. Portanto, o objetivo desse estudo foi desenvolver modelos de regressão por mínimos quadrados parciais (PLSR) com dados espectrais de EDXRF sob duas condições de medidas (com 15 kV e 50 kV no tubo) para determinação de carbono orgânico total (COT), fósforo disponível (P_{disp}), cálcio trocável (Ca^{2+}), magnésio trocável (Mg^{2+}), potássio trocável (K^+), soma de bases trocáveis (SB), capacidade de troca catiônica total (CTC), saturação por bases (V%), pH e acidez potencial ($H^+ + Al^{3+}$) em solos de uma topossequência derivada do basalto. Buscou-se avaliar o uso de diferentes técnicas de pré-processamento e métodos de seleção de variáveis com o objetivo de melhorar a capacidade preditiva dos modelos verificando as faixas espectrais que mais contribuíram para a determinação de cada parâmetro. Além disso, para verificar a eficiência dos modelos, foram avaliadas algumas figuras de mérito. Os resultados demonstraram que o uso de 15 kV no tubo foi a melhor condição de medida para todos os atributos avaliados, com exceção do pH. Além disso, o uso do escalamento de Pareto como pré-processamento gerou os menores erros de previsão para praticamente todos os parâmetros. Em geral, os métodos de seleção de variáveis resultaram em uma melhora na capacidade preditiva dos modelos de calibração, contribuindo para identificar as regiões espectrais que mais influenciaram na modelagem de cada atributo e a análise das figuras de mérito foi essencial para demonstrar a viabilidade da metodologia. Somente o COT, Ca^{2+} e SB apresentaram modelos preditivos considerados bons, Mg^{2+} , CTC e V% apresentaram modelos preditivos considerados razoáveis e os demais atributos apresentaram modelos fracos. Por fim, estes resultados contribuem para o estabelecimento de uma abordagem no tratamento de dados espectrais de EDXRF para calibração multivariada, bem como para uma melhor interpretação das variáveis que influenciam na modelagem dos atributos de fertilidade do solo.

Palavras-chave: EDXRF. Calibração multivariada. Pré-processamento. Seleção de variáveis. Solos.

DOS SANTOS, Felipe Rodrigues. **Use of X-ray fluorescence and multivariate analysis to determine soil fertility attributes in a toposequence derived from basalt.** 2020. 172 p. Thesis (PhD in Physics) – State University of Londrina, Londrina, 2020.

ABSTRACT

X-ray fluorescence (XRF) combined with multivariate calibration is one of the methodologies that has been proposed for soil fertility analysis in a fast, cost effective and environmental friendly manner. However, most of studies reported in the literature have been limited to the use of elemental data determined by XRF, which may cause loss of useful information contained in their spectra. Also, in some cases, these elemental data might not be accurate. In this sense, the use of XRF spectral data may be more appropriate in the development of multivariate calibration models for soil parameters determination. However, not all variables in the spectrum contribute to the parameter of interest modeling. In addition, there is no well-established approach for the pre-processing of soil XRF spectra data for multivariate calibration. Therefore, the objective of this study was to develop multivariate calibration models using partial least square regression (PLSR) with EDXRF spectral data under two measurement conditions (with 15 and 50 kV at tube) to determine total organica carbon (COT), available phosphorus (P_{disp}), exchangeable calcium (Ca^{2+}), exchangeable magnesium (Mg^{2+}), exchangeable potassium (K^+), sum of exchangeable bases (SB), total cation exchange capacity (CTC), base saturation percentage (V%), pH, and potential acidity ($\text{H}^+ + \text{Al}^{3+}$) in a toposequence derived from basalt. It was evaluated the use of different pre-processing techniques and variable selection methods, aiming to improve the predictive ability of the models and to interpret the spectral ranges that most contributed to the determination of each attribute. Also, to verify the models efficiency, some figures of merit were evaluated. The results showed that the use of 15 kV at tube was the best measurement condition for all evaluated attributes, except for pH. Moreover, the use of Pareto scaling as pre-processing generated the smallest predictive errors for almost all parameters. In general, the variable selection methods resulted in an improvement in the calibration models predictive ability, contributing to identify the spectral regions that most influenced the modeling of each attribute. The analysis of the figures of merit was essential to demonstrate the viability of the methodology. Only COT, Ca^{2+} , and SB presented predictive models considered good, Mg^{2+} , CTC, and V% presented predictive models considered reasonable, and the other attributes presented fair models. Finally, these results contribute toward the establishment of an approach for soil EDXRF spectra data treatment for multivariate calibration as well as for a better interpretation of the EDXRF variables that influence the soil fertility attributes modeling.

Keywords: EDXRF. Multivariate calibration. Pré-processing. Variable selection. Soils.

LISTA DE FIGURAS

Figura 1	Espectro contínuo de raios X para o alvo de tungstênio.	27
Figura 2	Representação esquemática do efeito fotoelétrico.	29
Figura 3	Representação esquemática do espalhamento Rayleigh.	31
Figura 4	Representação esquemática do espalhamento Compton.	32
Figura 5	Diagrama de transição de raios X e suas respectivas denominações.	35
Figura 6	Representação esquemática do efeito Auger.	36
Figura 7	Rendimento da fluorescência das camadas K e L em função do número atômico Z.	36
Figura 8	Etapas básicas de um sistema de EDXRF.	37
Figura 9	Comportamento do espectro de raios X emitido por um tubo de Rh.	38
Figura 10	Organização da matriz de dados espectrais X.	40
Figura 11	Localização dos pontos amostrais de acordo com a topografia do terreno.	46
Figura 12	Fluxograma com os passos utilizados no desenvolvimento dos modelos de calibração multivariada.	55
Figura 13	Espectros de EDXRF das 396 amostras de solo aplicando 15 kV (a) e 50 kV (b) no tubo de raios X.	64
Figura 14	Estimativa da contribuição relativa da absorção fotoelétrica, espalhamento Compton e Rayleigh para a energia de 2,7 (a) e 20,2 keV (b).	66
Figura 15	Box-plot com as recuperações de todas as medidas dos quatro MRC utilizados para validação (preto) com os limites de quantificação de cada elemento (vermelho).	69
Figura 16	Correlação de Pearson entre os atributos de fertilidade do solo determinados pelos métodos convencionais e as concentrações elementares determinadas por EDXRF.	71
Figura 17	Gráfico de scores (a e b) e loadings (c) da PCA com os atributos de fertilidade do solo determinados pelos métodos convencionais.	73
Figura 18	Gráfico de scores e loadings da PCA com as concentrações	

	elementares determinadas por EDXRF.....	74
Figura 19	Gráfico de scores e loadings da PCA nos dados espectrais de EDXRF na faixa de Ti-U centrados na média e na escala de Pareto.....	75
Figura 20	Métodos de pré-processamento aplicados sobre os dados espectrais de Na-Sc.....	80
Figura 21	Métodos de pré-processamento aplicados sobre os dados espectrais de Ti-U.....	81
Figura 22	Intervalos de variáveis selecionadas para os melhores modelos de COT com escalamento de Pareto e centragem na média.....	85
Figura 23	Vetor de regressão dos melhores modelos PLSR para o COT com escalamento de Pareto e centragem na média.....	86
Figura 24	Erro relativo para o conjunto de previsão dos melhores modelos para COT.....	89
Figura 25	Intervalos de variáveis selecionadas para o melhor modelo de Pdisp (a) com seu respectivo vetor de regressão (b).....	95
Figura 26	Erro relativo para o conjunto de previsão do melhor modelo para Pdisp.....	97
Figura 27	Intervalos de variáveis selecionadas para os melhores modelos de Ca ²⁺ com escalamento de Pareto e centragem na média.....	101
Figura 28	Vetor de regressão dos melhores modelos PLSR para o Ca ²⁺ com escalamento de Pareto e centragem na média.....	102
Figura 29	Erro relativo para o conjunto de previsão dos melhores modelos para Ca ²⁺	104
Figura 30	Intervalos de variáveis selecionadas para os melhores modelos de Mg ²⁺ com escalamento de Pareto e centragem na média.....	108
Figura 31	Vetor de regressão dos melhores modelos PLSR para o Mg ²⁺ com escalamento de Pareto e centragem na média.....	109
Figura 32	Erros relativos para o conjunto de previsão dos melhores modelos para Mg ²⁺	111
Figura 33	Variáveis selecionadas para os melhores modelos de K ⁺ (a) com escala de Pareto e (b) centrados na média com seus respectivos vetores de regressão (c) e (d).....	115
Figura 34	Erros relativos para o conjunto de previsão dos melhores	

	modelos para K+.....	117
Figura 35	Variáveis selecionadas para os melhores modelos de SB (a) com espectro de Na-Sc e (b) espectro de Ti-U com seus respectivos vetores de regressão (c) e (d).....	121
Figura 36	Erros relativos para o conjunto de previsão dos melhores modelos para SB.....	124
Figura 37	Intervalos de variáveis selecionadas para os melhores modelos de CTC com escalamento de Pareto e centragem na média.....	127
Figura 38	Vetor de regressão dos melhores modelos PLSR para CTC com escalamento de Pareto e centragem na média.	128
Figura 39	Erros relativos para o conjunto de previsão dos melhores modelos para CTC.	131
Figura 40	Variáveis selecionadas para os melhores modelos de V% (a) com espectro de Na-Sc e (b) espectro de Ti-U com seus respectivos vetores de regressão (c) e (d).....	135
Figura 41	Erros relativos para o conjunto de previsões dos melhores modelos para V%.....	137
Figura 42	Intervalos de variáveis selecionadas para os melhores modelos de pH com escalamento de Pareto e centragem na média.....	140
Figura 43	Vetor de regressão dos melhores modelos PLSR para o pH com escalamento de Pareto e centragem na média.	141
Figura 44	Erros relativos para o conjunto de previsão dos melhores modelos para pH.....	143
Figura 45	Variáveis utilizadas para os melhores modelos de acidez potencial com seus respectivos vetores de regressão.....	147
Figura 46	Erros relativos para o conjunto de previsões dos melhores modelos para acidez potencial.....	149
Figura 47	Gráfico com os valores de referência de cada atributo de fertilidade do solo versus valores previstos pelos melhores modelos.....	150

LISTA DE TABELAS

Tabela 1	Etapas do algoritmo NIPLS para o método PLSR.....	45
Tabela 2	Técnicas de pré-processamento aplicadas nos dados espectrais.....	51
Tabela 3	Estatística descritiva dos atributos do solo determinados pelos métodos convencionais.....	63
Tabela 4	Concentração certificada do MRC PTXRF_IAEA04, coeficiente elementar, limites de detecção (LD) e limites de quantificação (LQ) do sistema de EDXRF com 95% de nível de confiança.....	67
Tabela 5	Comparação entre concentrações certificadas e medidas do MRC IAEA09, com as respectivas recuperações.....	67
Tabela 6	Comparação entre as concentrações certificadas e medidas do MRC IAEA13, com as respectivas recuperações.....	68
Tabela 7	Comparação entre as concentrações certificadas e medidas do MRC IPT32, com as respectivas recuperações.....	68
Tabela 8	Comparação entre as concentrações certificadas e medidas do MRC IPT42, com as respectivas recuperações.....	68
Tabela 9	Resultados da estatística descritiva das concentrações elementares totais determinadas por EDXRF com seu respectivo desvio padrão (DP).....	70
Tabela 10	Estatística descritiva dos atributos do solo determinados pelos métodos convencionais.....	77
Tabela 11	Resultados dos modelos para COT com diferentes pré-processamentos aplicados aos dados espectrais de EDXRF.....	78
Tabela 12	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de COT com os espectros centrados na média e com escala de Pareto.....	82
Tabela 13	Teste de randomização para comparação dos melhores modelos para o COT em um intervalo de confiança de 95%.	84
Tabela 14	Figuras de mérito dos melhores modelos para COT.....	89
Tabela 15	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de Pdisp com os espectros centrados na média e com escala de Pareto.....	92

Tabela 16	Teste de randomização para comparação dos melhores modelos para o Pdisp em um intervalo de confiança de 95%......	93
Tabela 17	Figuras de mérito do melhor modelo para Pdisp.....	96
Tabela 18	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de Ca ²⁺ com os espectros centrados na média e com escala de Pareto.....	99
Tabela 19	Teste de randomização para comparação dos melhores modelos para o Ca ²⁺ em um intervalo de confiança de 95%.	100
Tabela 20	Figuras de mérito dos melhores modelos para Ca ²⁺	104
Tabela 21	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de Mg ²⁺ com os espectros centrados na média e com escala de Pareto.....	106
Tabela 22	Teste de randomização para comparação dos melhores modelos para o Mg ²⁺ em um intervalo de confiança de 95%......	107
Tabela 23	Figuras de mérito dos melhores modelos para Mg ²⁺	111
Tabela 24	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de K ⁺ com os espectros centrados na média e com escala de Pareto.....	113
Tabela 25	Teste de randomização para comparação dos melhores modelos para o K ⁺ em um intervalo de confiança de 95%.	114
Tabela 26	Figuras de mérito dos melhores modelos para K ⁺	117
Tabela 27	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de SB com os espectros centrados na média e com escala de Pareto.....	119
Tabela 28	Teste de randomização para comparação dos melhores modelos para o SB em um intervalo de confiança de 95%.	120
Tabela 29	Figuras de mérito dos melhores modelos para SB.	123
Tabela 30	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de CTC com os espectros centrados na média e com escala de Pareto.....	125
Tabela 31	Teste de randomização para comparação dos melhores modelos para o CTC em um intervalo de confiança de 95%.	126
Tabela 32	Figuras de mérito dos melhores modelos para CTC.	131
Tabela 33	Parâmetros estatísticos dos métodos de seleção de variáveis	

	para os modelos de V% com os espectros centrados na média e com escala de Pareto.....	133
Tabela 34	Teste de randomização para comparação dos melhores modelos para o V em um intervalo de confiança de 95%.	134
Tabela 35	Figuras de mérito dos melhores modelos para V%.....	136
Tabela 36	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de PH com os espectros centrados na média e com escala de Pareto.....	138
Tabela 37	Teste de randomização para comparação dos melhores modelos para o pH em um intervalo de confiança de 95%.	139
Tabela 38	Figuras de mérito dos melhores modelos para pH.....	142
Tabela 39	Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de acidez potencial com os espectros centrados na média e com escala de Pareto.	144
Tabela 40	Teste de randomização para comparação dos melhores modelos para a acidez potencial em um intervalo de confiança de 95%.	146
Tabela 41	Figuras de mérito dos melhores modelos para acidez potencial..	148
Tabela 42	Síntese dos resultados dos melhores modelos de cada parâmetro.....	149

LISTA DE ABREVIATURAS E SIGLAS

ANN	Redes Neurais Artificiais
BiPLS	Mínimos Quadrados Parciais por Intervalos “em sentido contrário”
COT	Carbono Orgânico Total
CTC	Capacidade de Troca Cationica
CV	Coeficiente de Variação
EDXRF	Fluorescência de Raios X por Energia Dispersiva
ER	Erro Relativo
FiPLS	Mínimos Quadrados Parciais por Intervalos Progressivos
GAM	Modelo Aditivo Generalizado
GLM	Modelo Linear Generalizado
iPLS	Mínimos Quadrados Parciais por Intervalos
iSPA-PLS	Algoritmo das Projeções Sucessivas por Intervalos para Mínimos Quadrados Parciais
LIBS	Espectroscopia de Emissão Óptica com Plasma Induzido por Laser
LV	Variáveis Latentes
MCR	Material de Referência Certificado
MLR	Regressão Linear Múltipla
MO	Matéria Orgânica
MSC	Correção Multiplicativa de Sinais
PC's	Componentes Principais
PCA	Análise de Componentes Principais
PLSR	Regressão por Mínimos Quadrados Parciais
RF	Floresta Aleatória
RMSEC	Raiz do Erro Quadrático Médio de Calibração
RMSECV	Raiz do Erro Quadrático Médio de Validação Cruzada
RMSEP	Raiz do Erro Quadrático Médio de Previsão
RPD	Razão de Desempenho de Desvio
RPIQ	Razão de Desempenho Interquartil
RT	Árvore de Regressão
SB	Soma de Bases Trocáveis
SiPLS	Mínimos Quadrados Parciais por Intervalos Sinérgicos

SNV	Padronização Normal de Sinal
SVM	Máquina de Vetor de Suporte
V%	Saturação por Bases
Vis-NIR	Espectroscopia do Infravermelho Próximo e Visível
XRF	Fluorescência de Raios X

SUMÁRIO

1	INTRODUÇÃO	16
2	OBJETIVOS	20
3	REVISÃO DA LITERATURA	21
4	FUNDAMENTAÇÃO TEÓRICA	27
4.1	FUNDAMENTOS DA FLUORESCÊNCIA DE RAIOS X	27
4.1.1	Produção De Raios X	27
4.2	ANÁLISE MULTIVARIADA	39
5	METODOLOGIA	46
5.1	AMOSTRAGEM	46
5.2	ANÁLISES CONVENCIONAIS	47
5.3	MEDIDAS POR EDXRF	48
5.4	ANÁLISE EXPLORATÓRIA	50
5.5	PRÉ-PROCESSAMENTOS	51
5.6	CALIBRAÇÃO MULTIVARIADA	52
5.7	SELEÇÃO DE VARIÁVEIS	55
5.8	FIGURAS DE MÉRITO	57
6	RESULTADOS E DISCUSSÕES	63
6.1	RESULTADOS DOS MÉTODOS CONVENCIONAIS	63
6.2	RESULTADOS DE EDXRF	64
6.3	ANÁLISE EXPLORATÓRIA	70
6.4	CALIBRAÇÃO MULTIVARIADA	76
7	CONCLUSÕES	152
	REFERÊNCIAS	154
	APÊNDICE	170
	APÊNDICE 1 – Publicações e participação em eventos durante o período do doutorado.	170

1. INTRODUÇÃO

Atualmente, estima-se que a população mundial é de 7,6 bilhões de pessoas, com expectativa de chegar a 9,8 bilhões em 2050 (UN-DESA POPULATION DIVISION, 2017). Com esse crescimento populacional, a demanda por alimentos, energia e demais produtos oriundos da agricultura tem aumentado. Diante desse contexto, a produção agrícola precisa ser otimizada, visando a sustentabilidade e a segurança alimentar.

A produção agrícola depende, entre outros fatores, da disponibilidade equilibrada dos nutrientes do solo. Portanto, a análise periódica dos atributos de fertilidade do solo é essencial para avaliação e monitoramento de suas condições químicas e físicas, indicando a necessidade de técnicas de fertilização, calagem e conservação (DEMATTÊ et al., 2019a).

Tradicionalmente, essas análises são realizadas em laboratórios através de extrações ácidas ou básicas, usando técnicas que, embora acuradas, exigem preparação das amostras, são destrutivas, morosas, utilizam reagentes e geram resíduos (WAISER et al., 2007).

Anualmente, milhões de análises de diversos atributos relacionados à química, fertilidade e classificação do solo são realizadas no Brasil (DE SOUZA et al., 2016). Com o desenvolvimento da agricultura de precisão, a demanda por essas análises tem se tornado ainda maior (DEMATTÊ et al., 2017). No entanto, com base nos métodos convencionais, realizar análises em um grande número de amostras requer mais recursos financeiros, tempo e reagentes químicos que geram resíduos.

Algumas análises, por exemplo, podem levar de 3 a 15 dias para ficarem prontas, o que não é adequado considerando a velocidade exigida na agricultura de precisão. Além disso, deve-se levar em conta o volume de substâncias químicas utilizadas nessas análises. De acordo com Demattê et al. (2019b), somente para análise de matéria orgânica (MO) estima-se um consumo global de 3 milhões de litros de ácido sulfúrico e 840 toneladas de dicromato e sulfato ferroso amoniacal. Além do custo dessas análises, existe a questão ambiental envolvida na destinação correta dos resíduos gerados.

Assim, o desenvolvimento de novos métodos para análise de solos baseados na química verde, nos quais agricultores e laboratórios aumentem o número de análises de forma rápida e eficiente, vem tornando-se uma prioridade nas

últimas décadas. Dessa forma, muitos pesquisadores vêm explorando o uso de métodos espectroscópicos associados à análise multivariada para determinar diversos atributos do solo (SORIANO-DISLA et al., 2014; VILLAS-BOAS et al., 2019; WEINDORF; BAKR; ZHU, 2014), com destaque para a espectroscopia no infravermelho próximo e visível (Vis-NIR, do inglês *Visible and Near Infrared Spectroscopy*), a qual tem se mostrado bastante eficiente para a determinação de diversas propriedades do solo, como teor de argila, matéria orgânica e composição mineralógica (SORIANO-DISLA et al., 2014). No entanto, o número de estudos que reportam a aplicabilidade dessa técnica na determinação de cátions trocáveis e composição química do solo é bem menor (DEMATTÊ et al., 2017). De acordo com Stenberg et al. (2010) os modelos utilizando dados espectrais de Vis-NIR ainda não atingiram níveis de acurácia ótimos para determinação dos cátions trocáveis devido à fraca relação com as bandas de absorção características.

Visando avançar neste tema, a fluorescência de raios X por dispersão em energia (EDXRF, do inglês *Energy Dispersive X-Ray Fluorescence*) combinada com análise multivariada tem surgido como uma possibilidade viável na determinação dos atributos do solo, a qual pode tornar-se uma técnica promissora para avaliação destes parâmetros de forma rápida, sem uso de reagentes e geração de resíduos com a vantagem de possibilitar a determinação simultânea da concentração elementar (DOS SANTOS et al., 2020; KANIU et al., 2012; MORONA et al., 2017; SHARMA et al., 2015; SILVA et al., 2017; TAVARES et al., 2020b; ZHU; WEINDORF; ZHANG, 2011). Além disso, alguns estudos têm mostrado que a fusão de dados de EDXRF com Vis-NIR pode melhorar de forma considerável a capacidade preditiva dos atributos de fertilidade do solo em modelos de calibração multivariada (BENEDET et al., 2020a; COMINO et al., 2018; O'ROURKE et al., 2016; WAN et al., 2020; WANG et al., 2015a).

A EDXRF é uma técnica analítica multielementar frequentemente empregada na análise qualitativa e quantitativa de uma grande variedade de amostras (BYERS; MCHENRY; GRUNDL, 2019; KHUDER et al., 2010; LEMIÈRE, 2018; MARGUÍ; QUERALT; HIDALGO, 2009; ROUILLON; TAYLOR, 2016; SCIUTTO et al., 2012). Devido à sua habilidade em identificar e quantificar espécies inorgânicas, de forma rápida, simples, com baixo custo operacional, necessitar de mínima e às vezes nenhuma preparação de amostra e, principalmente, por ser, em geral, uma técnica não destrutiva, a EDXRF tornou-se bastante disseminada

(JENKINS, 1988). Ela é aplicável a uma extensa faixa de concentração e possui exatidão analítica adequada. No entanto, a determinação de elementos com $Z < 18$ é afetada pelo baixo rendimento de fluorescência. Além disso, vários outros fatores ainda são um desafio para análise de solo por EDXRF, como efeitos físicos de matriz (teor de umidade, tamanho das partículas e condições da superfície) e efeitos químicos de matriz (causados pela absorção ou intensificação do sinal de emissão do elemento de interesse por elementos que constituem a matriz amostral) (MARGUÍ; QUERALT; HIDALGO, 2009).

Como discutido anteriormente, a determinação de atributos de fertilidade do solo por meio de métodos espectroscópicos combinados com calibração multivariada tem se mostrado promissora e inovadora. No entanto, a maioria dos estudos que relatam a combinação de dados de EDXRF com análise multivariada se limitam ao uso das concentrações elementares, as quais, devido à complexidade da correção do efeito matriz para amostras de solo, em alguns casos podem não ser acuradas, conduzindo a interpretações erradas e falta de robustez nos modelos. Além disso, o uso de dados elementares pode causar a perda de informações úteis contidas nos espectros, como efeitos de espalhamento, cuja contribuição para caracterização de amostras orgânicas complexas tem se mostrado importante, como demonstrado por diversos autores (AIDENE et al., 2020; BORTOLETO; BORGES; BUENO, 2007; DE ALMEIDA et al., 2020; MELQUIADES et al., 2012; BUENO et al., 2005; VERBI; PEREIRA-FILHO; BUENO, 2005). Nesse sentido, o uso de dados espectrais de EDXRF pode ser mais adequado no desenvolvimento de modelos de calibração multivariada para determinação de parâmetros do solo.

No entanto, ainda não existe um procedimento bem estabelecido para o pré-processamento dos dados espectrais de EDXRF para calibração multivariada. Além disso, nem todas as variáveis do espectro contribuem para a modelagem dos parâmetros de interesse. Nesse contexto, o uso de métodos de seleção de variáveis pode melhorar a previsão dos modelos, permitindo uma interpretação mais simples da contribuição das variáveis (ANDERSEN; BRO, 2010; DOS SANTOS et al., 2020; WANG et al., 2018).

Portanto, o objetivo geral desse estudo foi desenvolver modelos de calibração multivariada utilizando regressão por mínimos quadrados parciais (PLSR, do inglês *Partial Least Square Regression*) com dados espectrais de EDXRF sob

duas condições de medidas para determinação de diversos atributos de fertilidade de solos de uma topossequência localizada na região norte do estado do Paraná. Especificamente, buscou-se avaliar o uso de diferentes técnicas de pré-processamento e métodos de seleção de variáveis nos dados espectrais de EDXRF com o objetivo de melhorar a capacidade preditiva dos modelos verificando as faixas espectrais que mais contribuíram para a determinação de cada atributo. Além disso, para verificar a eficiência dos modelos, foram avaliadas algumas figuras de mérito como linearidade, precisão, robustez, sensibilidade e limites de detecção.

2. OBJETIVOS

2.1 OBJETIVO GERAL

Desenvolver modelos de calibração multivariada utilizando dados espectrais de EDXRF para determinação de carbono orgânico total (COT), fósforo disponível (P_{disp}), cálcio trocável (Ca^{2+}), magnésio trocável (Mg^{2+}), potássio trocável (K^+), soma de bases trocáveis (SB), capacidade de troca catiônica total (CTC), saturação por bases (V%), pH e acidez potencial ($\text{H}^+ + \text{Al}^{3+}$) em solos de uma toposequência derivada do basalto, localizado no município de Cambé-PR.

2.2 OBJETIVOS ESPECÍFICOS

- Determinar os atributos de fertilidade do solo pelos métodos convencionais;
- Calibrar e validar um sistema de bancada de EDXRF para determinar os teores totais de Al, Si, P, S, Ca, K, Mn, Fe, Cu e Zn em solos agrícolas;
- Realizar análise exploratória com os dados dos atributos de fertilidade do solo, dados de concentrações elementares e dados espectrais de EDXRF;
- Construir modelos de calibração multivariada utilizando PLSR para determinação dos atributos de fertilidade do solo;
- Verificar a influência das condições de medidas por EDXRF nos modelos de calibração multivariada;
- Avaliar o uso de diferentes técnicas de pré-processamento e métodos de seleção de variáveis nos dados espectrais de EDXRF;
- Verificar as regiões espectrais com maior contribuição na construção dos modelos;
- Realizar a validação analítica avaliando as figuras de mérito dos modelos de calibração multivariada.

3 REVISÃO DA LITERATURA

A demanda por informações do estado nutricional de solos agrícolas de forma rápida, barata e com baixo impacto ambiental vem crescendo nos últimos anos. Dessa forma, o desenvolvimento de métodos mais rápidos que atendam os princípios da química verde, tem se tornado uma prioridade.

Nas últimas décadas, novas técnicas e conceitos, envolvendo métodos espectroscópicos e análise multivariada, têm sido desenvolvidos. Estas novas técnicas são promissoras e muitas delas bem sucedidas na previsão das propriedades do solo. Por exemplo, a espectroscopia do visível e infravermelho próximo (Vis-NIR) juntamente com a calibração multivariada tem sido relatada como uma ferramenta eficaz para medir uma variedade de propriedades do solo, incluindo mineralogia (SERBIN et al., 2009; VISCARRA ROSSEL et al., 2009a), textura (BRICKLEMYER; BROWN, 2010; BROWN, 2007; DE SANTANA; DE SOUZA; POPPI, 2018; HOBLEY; PRATER, 2019; VISCARRA ROSSEL et al., 2009b) carbono orgânico total (BELTRAME et al., 2016; DEMATTÊ et al., 2019b; GE et al., 2011; MORGAN et al., 2009; SANKEY et al., 2008; SITHOLE; NCAMA; MAGWAZA, 2018; STEVENS et al., 2013; TERRA; VISCARRA ROSSEL; DEMATTÊ, 2019; VOHLAND et al., 2011), cor (RAMOS et al., 2020; ROSSEL; CHEN, 2011; VISCARRA ROSSEL et al., 2009b), macronutrientes (DEMATTÊ et al., 2019a, 2019b; MALMIR et al., 2019; MUNNAF; NAWAR; MOUAZEN, 2019; TERRA; DEMATTÊ; VISCARRA ROSSEL, 2015; WANG et al., 2015b) e demais indicadores de fertilidade do solo (DE SANTANA; DE SOUZA; POPPI, 2018; DEMATTÊ et al., 2019a, 2019b; MUNNAF; NAWAR; MOUAZEN, 2019; TERRA; DEMATTÊ; VISCARRA ROSSEL, 2015; VOHLAND et al., 2014; WANG et al., 2015b).

Recentemente, alguns estudos têm reportado o uso de fluorescência de raios X (XRF) combinado com análise multivariada como uma possibilidade viável na previsão de propriedades do solo.

Zhu; Weindorf & Zhang (2011) desenvolveram modelos de regressão linear múltipla (MLR, do inglês *Multiple Linear Regression*) com seleção de variáveis por *backward elimination* para determinação da textura (teores de argila, areia e silte) do solo utilizando dados de concentração de 15 elementos (K, Ca, Ti, Cr, Mn, Fe, Co, Cu, Zn, As, Rb, Sr, Zr, Ba, and Pb) determinados por um equipamento portátil de fluorescência de raios X (pXRF). Este estudo foi realizado

com 426 amostras de solo coletadas no estado da Louisiana, EUA e 157 amostras de solo coletadas na região do vulcão Capulin no Novo México, EUA. Os resultados de validação externa demonstraram bom desempenho dos modelos com coeficiente de determinação (R^2) de 0,975, 0,854 e 0,682 para argila, areia e silte da Louisiana e R^2 de 0,876, 0,891 e 0,875 para argila, areia e silte de Capulin.

Kaniu et al. (2012) realizaram um dos primeiros estudos demonstrando o uso de EDXRF para determinação de macronutrientes biodisponíveis em solo. Nesse estudo foram determinados 5 macronutrientes biodisponíveis em solos (C, N, Na, Mg e P) utilizando regressão por mínimos quadrados parciais (PLSR) e redes neurais artificiais (ANN) com dados espectrais do espalhamento Compton. Para tal, os autores utilizaram 66 amostras de Latossolo coletadas na região leste (Katumani) e oeste (Kitale) do Quênia. O método de ANN apresentou desempenho melhor que o PLSR somente para a modelagem de Mg. A validação externa apresentou altos valores de R^2 para todos os parâmetros. Além disso, o teste *one-way* ANOVA, em um nível de confiança de 95%, indicou que não houve diferença significativa entre os teores dos macronutrientes de referência e previstos pelos modelos PLSR e ANN.

Sharma et al. (2014) desenvolveram modelos de MLR para determinar o pH de solos agrícolas utilizando dados elementares determinados por pXRF. Para esse estudo foram coletados dois conjunto de dados: um denominado Dataset A, que consiste de 100 amostras coletadas em diversos estados dos EUA, e outro denominado Dataset B, contendo 639 amostras coletadas no estado da Louisiana, EUA. Para o Dataset A, o modelo MLR foi construído com os teores de Al, Si, Mn, Fe, K, Ca e Zn, gerando um R^2 de previsão de 0,57. Para o Dataset B, foram utilizados dados de concentração de 15 elementos (K, Ca, Ti, Cr, Mn, Fe, Co, Cu, Zn, As, Rb, Sr, Zr, Ba e Pb) gerando um R^2 de previsão de 0,77.

Em outro estudo, Sharma et al. (2015) construíram modelos de MLR usando dados de concentrações de pXRF para análise da CTC de solos agrícolas. Para isso foram empregadas 450 amostras de solo coletadas no Nebraska e na Califórnia, EUA. Foram utilizados os dados dos teores de Ca, Ti, V, Cr, Fe, Cu, Sr e Zr para construção dos modelos de MLR, o que gerou um R^2 de 0,908.

Morona et al. (2017) determinaram os teores de matéria orgânica (MO) e carbono orgânico total (COT) em solos agrícolas de 27 municípios do estado do Paraná e discriminaram a sua proveniência utilizando dados espectrais de

EDXRF combinados com regressão por mínimos quadrados parciais e análise de componentes principais. Para construção dos modelos de calibração multivariada foi utilizado a região de espalhamento do espectro, gerando um R^2 de 0,60 e razão de desempenho de desvio (RPD) de 1,5 para validação externa do COT e R^2 de 0,82 e RPD=2,5 para validação cruzada da MO. Além disso, nesse estudo os autores determinaram outras figuras de mérito como o inverso da sensibilidade analítica, limites de detecção e quantificação.

Silva et al. (2017) construíram modelos de MLR e floresta aleatória (RF, do inglês *Random Forest*) para determinar diversas propriedades do solo (pH, K^+ , Ca^{2+} , Mg^{2+} , Al^{3+} , P-remanescente, MO, CTC e V%) utilizando dados elementares de pXRF. Este estudo foi realizado em uma propriedade rural no município de Campos Altos – MG. Os resultados demonstraram que o melhor desempenho foi obtido na modelagem com RF, apresentando R^2 de 0,88, 0,39, 0,95, 0,71, 0,86, 0,76, 0,60, 0,38 e 0,95 para previsão do pH, K^+ , Ca^{2+} , Mg^{2+} , Al^{3+} , P-remanescente, MO, CTC e V%, respectivamente. O Ca foi a variável mais importante para modelagem de todos os parâmetros, exceto para a MO e CTC, que tiveram como variável mais importante o P e o Si, respectivamente.

Rawal et al. (2019) avaliaram o uso de quatro modelos de calibração multivariada (Modelo aditivo generalizado (GAM), regressão linear múltipla (MLR), floresta aleatória (RF) e árvore de regressão (RT)) para determinar a saturação por bases (V%) e a capacidade de troca catiônica (CTC) em solos agrícolas de seis estados dos EUA. Para isso foram utilizados dados elementares de Mg, K e Ca medidos por pXRF. Em geral, os modelos não apresentaram um desempenho satisfatório. O melhor modelo para V% foi com RT que gerou um R^2 de validação de 0,68 e o melhor modelo para CTC foi utilizando GAM que gerou R^2 de validação igual a 0,69. Nesse caso, a razão de desempenho de desvio (RPD) para ambos os modelos foi 1,8 e a razão de desempenho interquartil (RPIQ) foi 2,0 e 2,7 para o V% e CTC, respectivamente.

Silva et al. (2020) construíram modelos de regressão por meio de modelo linear generalizado (GLM), máquina de vetor de suporte (SVM) e floresta aleatória (RF) com dados elementares de pXRF para determinar a textura de uma ampla variedade de solos do Brasil (1565 amostras coletadas nos horizontes superficiais e sub-superficiais nos estados de MG, SP, ES, RJ, BA, PE e SC). Foram construídos modelos para os horizontes superficiais e sub-superficiais de forma

separada e combinada. Os modelos SVM e RF apresentaram resultados preditivos melhores do que os modelos GLM, nos quais, com o conjunto de dados dos horizontes combinado, o R^2 do modelo RF foi igual a 0,73 para a previsão do teor de areia e o R^2 dos modelos SVM foram iguais a 0,72 e 0,84 para previsão dos teores de silte e argila, respectivamente.

Teixeira et al. (2020) utilizaram dados elementares de pXRF com três algoritmos (*Cubist*, RF e MLR) pra prever o pH, SB, V%, CTC e saturação por Al de uma ampla variedade de solos do Brasil. Foram utilizadas 2017 amostras de solo coletadas nos horizontes superficiais (682 amostras) e sub-superficiais (1335 amostras) de sete estados brasileiros (MG, SP, ES, RJ, BA, PE e SC). No geral, os melhores resultados foram atingidos pelo método *Cubist* seguido por RF, onde o RPD variou de 1,5 a 1,6 para o pH; de 1,4 a 2,6 para SB; de 1,5 a 1,7 para CTC; de 1,4 a 2,3 para V% e de 1,6 a 2,0 para saturação por Al. Nesse estudo, os autores também demonstram que o Ca foi a variável mais importante na modelagem de todos os parâmetros.

Andrade et al. (2020) utilizaram dados elementares de pXRF para desenvolver modelos de GLM e RF para prever os macronutrientes (Al^{3+} , Ca^{2+} , Mg^{+2} , K^+ e P disponível) de solos da Planície costeira do ES e BA, bem como avaliar a influência da mineralogia e do manejo do solo na sua composição elementar. Os resultados demonstram que a mineralogia do solo teve influência principalmente na composição do Al. Por outro lado, o manejo do solo influenciou na composição do Ca e K. Os melhores resultados foram atingidos utilizando RF, sendo que o RPD ficou entre 1,4 e 2,0 para o Ca^{2+} , Mg^{+2} e K^+ . No entanto, para o Al^{3+} e P disponível os modelos não apresentaram desempenho satisfatório, com RPD entre 0,5 e 1,5. Os autores identificaram que as variáveis mais importantes para modelagem do Ca^{2+} , K^+ e P disponível foram seus respectivos teores determinados por pXRF. Além disso, eles argumentaram que o fraco desempenho dos modelos para Al^{3+} e P disponível foi devido à forte influência da mineralogia sobre o Al^{3+} e aos baixos teores de P disponível nos solos em estudo.

TAVARES et al. (2020a) avaliaram o desempenho de um equipamento de pXRF operado em duas condições de medidas (utilizando 15 kV no tubo e corrente de 23 μA e com 35 kV no tubo e corrente de 7 μA) para prever alguns atributos de fertilidade do solo (teor de argila, MO, CTC, pH, V% e macronutrientes trocáveis (P, K, Ca e Mg)). Foram construídos modelos MLR

utilizando as intensidades das linhas de emissão de diferentes elementos (Al, Si, K, Ca, Ti, Mn, Fe, Ni e Cu) bem como as intensidades dos picos de espalhamento. Para isso, utilizou-se 102 amostras coletadas em dois campos agrícolas (em Piracicaba – SP e Campo Novo do Parecis – MT). Os resultados demonstraram que apesar da configuração no tubo de raio X afetar a intensidade das linhas de emissão, não houve diferença significativa na acurácia dos modelos MLR. Além disso, os modelos apresentaram desempenho satisfatório para todos os parâmetros avaliados, com exceção do pH e P trocável.

Em outro estudo, Tavares et al. (2020b) avaliaram o uso da normalização Compton em dados de intensidade de pXRF para determinar os teores de argila, CTC, Ca e K trocável de dois campos agrícolas (Piracicaba – SP e Campo Novo do Parecis – MT). Nesse estudo, foram construídos modelos MLR e PLSR. Os resultados demonstram que o uso da normalização Compton melhorou a capacidade preditiva dos modelos, reduzindo o valor de RMSEP de 7 a 27%. Além disso, os autores demonstraram que a normalização Compton foi eficiente para mitigar o efeito matriz devido à heterogeneidade das amostras dos diferentes campos de coleta.

Além desses estudos relatados, o uso de dados de EDXRF tem sido útil para avaliação de diversas outras propriedades de interesse agrícola, como avaliação da salinidade (SWANHART et al., 2014; WEINDORF et al., 2018), intemperismo e pedogênese (MANCINI et al., 2019a, 2019b; SILVA et al., 2018; STOCKMANN et al., 2016), proveniência de sedimentos (DE CASTILHOS et al., 2015; MELQUIADES et al., 2019), classificação do uso e tipo do solo (BENEDET et al., 2020b; CHAKRABORTY et al., 2019; MELQUIADES; ANDREONI; THOMAZ, 2013), entre outros.

Vale a pena ressaltar, que além das técnicas de Vis-NIR e EDXRF, a técnica de espectroscopia de emissão óptica com plasma induzido por laser (LIBS, do inglês *Laser Induced Breakdown Spectroscopy*) também tem se mostrado eficiente na determinação de diversos atributos do solo (FERREIRA et al., 2015; GOUEGUEL et al., 2019; HE et al., 2018; SENESI; SENESI, 2016; VILLAS-BOAS et al., 2016; Xu et al., 2019).

De forma geral, esses estudos relatados na literatura mostram o potencial da técnica de EDXRF combinada com a análise multivariada na previsão das propriedades do solo. No entanto, a maioria dos estudos citados se limita ao uso

de dados de concentrações elementares determinadas por XRF usando rotinas quantitativas desenvolvidas pelos fabricantes dos equipamentos, as quais em alguns casos não são acuradas e não levam em conta os efeitos de espalhamento. Além disso, sempre que um procedimento analítico é proposto ou desenvolvido, é necessário investigar se o método apresenta um desempenho adequado para as condições em que será aplicado. Poucos estudos determinam figuras de mérito como precisão, sensibilidade, robustez, limite de detecção e linearidade para modelos multivariados usando dados de EDXRF. Do mesmo modo, é pouco explorada a influência dos pré-processamentos nos dados, da seleção de variáveis e das condições de medidas de EDXRF em modelos de calibração multivariada. Sendo assim, uma das propostas deste estudo é preencher esta lacuna e realizar um procedimento cauteloso de validação.

4 FUNDAMENTAÇÃO TEÓRICA

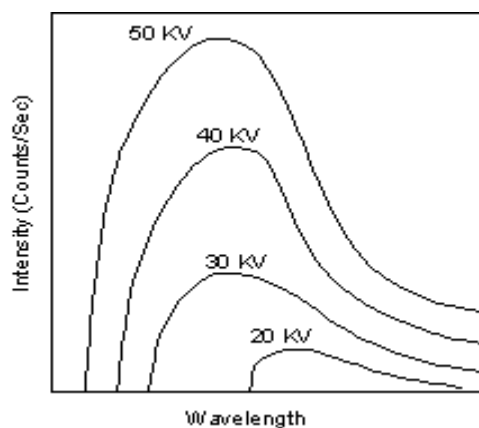
4.1 FUNDAMENTOS DA FLUORESCÊNCIA DE RAIOS X

4.1.1 Produção de Raios X

Os raios X compreendem uma forma de radiação eletromagnética de pequeno comprimento de onda. Sua descoberta ocorreu acidentalmente em 1895 pelo físico alemão Wilhelm Conrad Röntgen. Hoje, sabe-se que os raios X têm origem em dois processos físicos que ocorrem em nível atômico: *Bremsstrahlung* e raios X característicos (JENKINS, 1988).

O *Bremsstrahlung* (do alemão que significa desaceleração da radiação) ocorre na interação de elétrons com o campo elétrico dos átomos de um material alvo em tubos de raios X. Nesse processo de interação os elétrons perdem parte de sua energia cinética, a qual é convertida em fótons de raios X. Como consequência disso, é emitido um espectro contínuo de radiação eletromagnética, que está diretamente associado com a alta tensão aplicada ao tubo (EISBERG; RESNICK, 1979), conforme mostra a Figura 1.

Figura 1 - Espectro contínuo de raios X para o alvo de tungstênio.



Fonte: EISBERG; RESNICK, 1979.

Os raios X característicos podem ser originados na interação de partículas carregadas ou fótons com um alvo. Isso ocorre quando partículas carregadas ou fótons incidentes fornecem energia suficiente para arrancar os elétrons das camadas mais internas do elemento alvo, deixando o átomo num

estado altamente excitado. Ao retornar para o estado fundamental, é emitido um conjunto de fótons com energias bem definidas, que correspondem à diferença de energia entre os dois orbitais envolvidos e estão diretamente associados ao número atômico do elemento alvo (EISBERG; RESNICK, 1979; JENKINS, 1988).

Um aspecto interessante dos raios X característicos é que as frequências e comprimentos de onda das linhas variam lentamente de elemento a elemento. Essa regularidade foi inicialmente observada por Moseley, em 1913. Ele fez um levantamento de espectros de raios X característicos para vários elementos. O resultado que ele encontrou é que o comprimento de onda de qualquer linha do espectro discreto decresce à medida que o número atômico do alvo aumenta, ou seja, ele encontrou uma relação linear entre a raiz quadrada da frequência dos raios X emitidos e o número atômico Z , conhecida como a lei de Moseley (EISBERG; RESNICK, 1979). Este resultado foi muito importante, pois conseguiu reformular a tabela periódica desenvolvida por Mendeleev e foi um dos primeiros resultados experimentais que concordava com o modelo atômico de Bohr (EISBERG; RESNICK, 1979). Com isso, teve início a técnica de fluorescência de raios (XRF) embora, somente a partir de 1948, a XRF tenha se consolidado com um trabalho apresentado por Freidman & Birks (JENKINS, 1988).

4.1.2 Interação de Raios X com a matéria

Quando um feixe de raios X incide sobre um material, alguns de seus fótons podem interagir com os átomos deste material, podendo originar três tipos de interação: absorção fotoelétrica, espalhamento coerente (ou Rayleigh) e espalhamento incoerente (ou Compton). As probabilidades dessas interações podem ser expressas pela seção de choque de cada fenômeno. Portanto, a probabilidade de interação de um fóton com um átomo por centímetro quadrado da matéria é dada pela soma das seções de choque por átomo dos fenômenos predominantes, conforme Equação 1 (VAN GRIEKEN; MARKOWICZ, 2002).

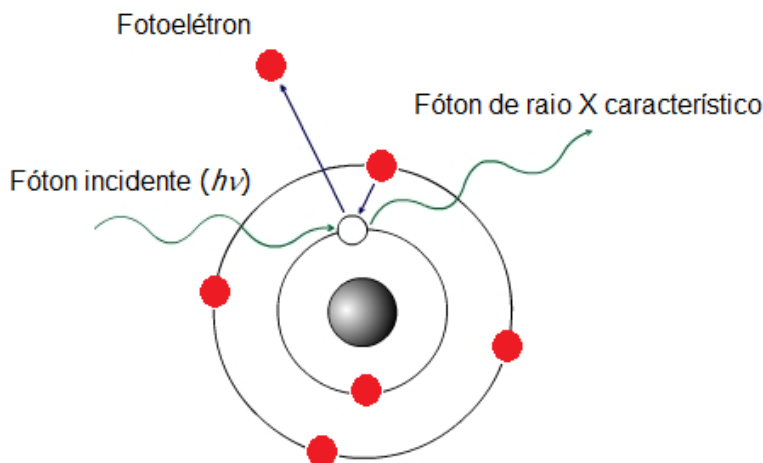
$$\sigma_{\text{tot}} = \sigma_{\text{F}} + \sigma_{\text{R}} + Z\sigma_{\text{C}} \quad (1)$$

onde σ_{F} é a seção de choque de absorção fotoelétrica total por átomo, σ_{R} é a seção de choque para o espalhamento Rayleigh, σ_{C} é a seção de choque para o espalhamento Compton e Z é o número atômico do átomo.

4.1.2.1 Efeito fotoelétrico

No efeito fotoelétrico, o fóton incidente interage com um elétron das camadas mais internas de um átomo. A energia do fóton é completamente absorvida pelo elétron, que é ejetado do material (EISBERG; RESNICK, 1979), conforme Figura 2. Esse fenômeno ocorre somente se o fóton incidente tiver energia maior do que a energia de ligação do elétron ao átomo, denominado como corte de absorção, a qual pode ser estimada de forma semi-empírica, aplicando-se a teoria atômica de Bohr para o átomo de hidrogênio com algumas considerações sobre as experiências de Moseley (NASCIMENTO FILHO, 1999). Cabe ressaltar, que após a liberação do fotoelétron, o átomo fica em um estado excitado e para desexcitação desse átomo um fóton de raio X característico ou, em alguns casos, um elétron Auger é liberado.

Figura 2 - Representação esquemática do efeito fotoelétrico.



Fonte: Adaptado de <<http://slideplayer.com.br/slide/11774338/>>

O efeito fotoelétrico ocorre com maior probabilidade para fótons incidentes de baixas energias (até 100 keV) e para elementos de alto número atômico. A seção de choque do efeito fotoelétrico para energias não relativísticas ($h\nu \ll m_0c^2$) pode ser estimada pela Equação 2 (KAPLAN, 1978).

$$\sigma_F = \frac{8\pi}{3} \left(\frac{e^2}{m_0c^2} \right)^2 Z^5 \alpha^4 4\sqrt{2} \left(\frac{m_0c^2}{h\nu} \right)^{7/2} \quad (2)$$

onde $h\nu$ é a energia do fóton incidente, m_0c^2 é a energia de repouso do elétron e $\alpha = 1/137$ é a constante de estrutura fina.

A Equação 2 descreve a probabilidade de ocorrer o efeito fotoelétrico no nível da camada K. Nesse nível, a probabilidade para ejeção dos fotoelétrons é maior devido à conservação de momento dos elétrons, visto que seus momentos de recuo são absorvidos pelo núcleo atômico.

4.1.2.2 Espalhamento rayleigh

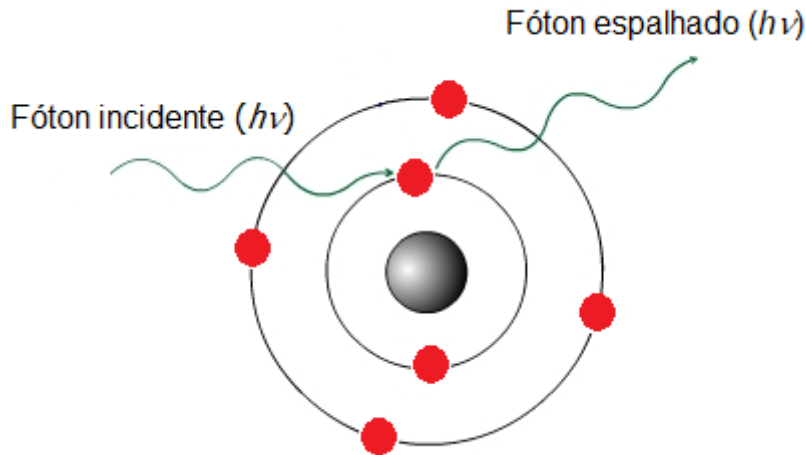
O espalhamento Rayleigh consiste num espalhamento elástico dos fótons incidentes, de baixa energia, em um átomo como um todo. Nesse processo, todos os elétrons do átomo participam de maneira coerente (LEO, 1994). Por esse motivo, ele também é chamado de espalhamento coerente. Por se tratar de uma colisão elástica, não há transferência de energia entre os fótons e os elétrons. Assim, os fótons são espalhados com a mesma energia com que interagiram com os átomos, mudando apenas o seu momento linear (VAN GRIEKEN; MARKOWICZ, 2002), conforme mostra a representação esquemática da Figura 3.

O espalhamento Rayleigh ocorre principalmente para baixas energias e para materiais com alto Z, cuja seção de choque diferencial para energias não relativísticas é expressa pela Equação 3 (VAN GRIEKEN; MARKOWICZ, 2002).

$$\frac{d\sigma_R}{d\Omega} = \frac{1}{2} r_0^2 (1 + \cos^2 \theta) |F(x, Z)|^2 \quad (3)$$

onde r_0 corresponde ao raio clássico do elétron, $F(x, Z)$ é o fator de espalhamento atômico que leva em consideração a distribuição de carga da nuvem eletrônica e $d\Omega$ é o ângulo sólido para o ângulo de difusão θ .

Figura 3 - Representação esquemática do espalhamento Rayleigh.



Fonte: Adaptado de <<http://slideplayer.com.br/slide/11774338/>>

4.1.2.3 Espalhamento Compton

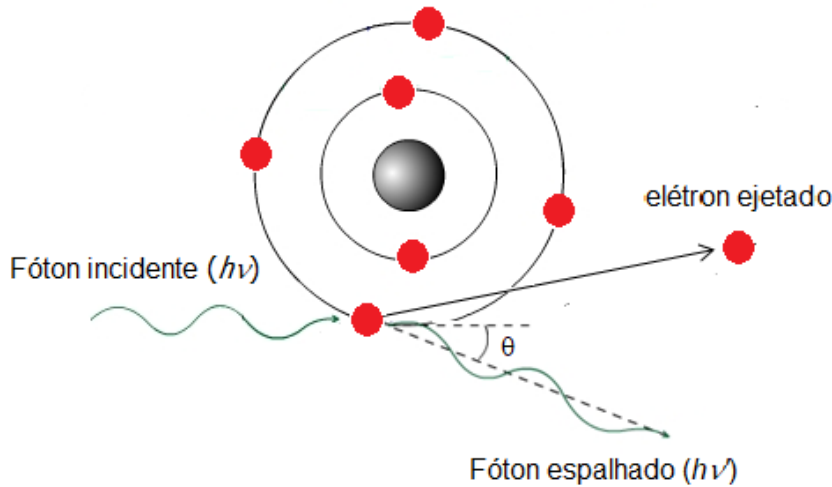
O espalhamento Compton surge da interação de um fóton com um elétron atômico fracamente ligado (VAN GRIEKEN; MARKOWICZ, 2002). Através dessa interação parte da energia do fóton incidente é transferida para o elétron, ejetando-o do átomo. Nesse caso, o fóton é espalhado num certo ângulo com relação à sua direção de incidência e com energia menor, sendo igual à energia de incidência menos a energia transferida ao elétron (EISBERG; RESNICK, 1979). Uma representação esquemática do espalhamento Compton é apresentada na Figura 4.

A energia do fóton espalhado pode ser determinada considerando a conservação do momento e energia do sistema, conforme Equação 4.

$$h\nu' = \frac{h\nu}{1 + \gamma(1 - \cos \theta)} \quad (4)$$

Onde $\gamma = h\nu/m_0c^2$ e θ é o ângulo com que o fóton incidente foi espalhado (EISBERG; RESNICK, 1979).

Figura 4 - Representação esquemática do espalhamento Compton.



Fonte: Adaptado de <<http://slideplayer.com.br/slide/11774338/>>

O efeito Compton é predominante para altas energias (de 100 keV a 1,02 MeV) e em elementos de baixo Z (CESAREO et al., 1992), sendo que a probabilidade de ocorrer esse fenômeno aumenta quando a energia de ligação dos elétrons orbitais se torna desprezível comparada com a energia do fóton incidente. Dessa forma, os elétrons podem ser considerados livres no processo de interação. A fórmula de Klein- Nishina (Equação 5) descreve a seção de choque diferencial da interação de um fóton com um elétron livre com relação ao ângulo sólido Ω .

$$\frac{d\sigma_c}{d\Omega} = \frac{1}{2} r_0^2 \frac{1}{[1 + \gamma \cos \theta]^2} \left[1 + \cos^2 \theta + \frac{\gamma^2 (1 - \cos \theta)^2}{1 + \gamma (1 - \cos \theta)} \right] \quad (5)$$

onde r_0 corresponde ao raio clássico do elétron e $\gamma = hv/m_0c^2$.

Se o efeito Compton ocorrer com elétrons ligados, alguns efeitos de ligação atômica devem ser considerados na Equação 5. Isso envolve aplicar um fator multiplicativo $S(x,Z)$ chamado de função de espalhamento incoerente (CESAREO et al., 1992).

4.1.2.4 Atenuação de raios X na matéria

A intensidade de fótons transmitidos quando um feixe de fótons, com intensidade I_0 e energia E , atravessa um material puro de espessura x é dada por (CESAREO et al., 1992):

$$I = I_0 e^{-\mu(E)x} \quad (6)$$

Onde $\mu(E)$ é o coeficiente de atenuação linear por centímetro que atravessa o absorvedor, que pode ser expresso como a seção de choque total multiplicada pelo número de átomos por centímetro cúbico do absorvedor, conforme Equação 7 (VAN GRIEKEN; MARKOWICZ, 2002). Em alguns casos, se utiliza $\mu' = \mu(E)/\rho$, que é definido como o coeficiente de atenuação de massa para um material de densidade ρ e número atômico Z . O coeficiente de atenuação de massa tem a vantagem de ser independente da densidade do material, sendo proporcional à seção de choque das interações dos fótons por átomo (CESAREO et al., 1992).

$$\mu = \sigma_{\text{tot}} \rho \frac{N_A}{A} \quad (7)$$

onde ρ é a densidade do material, A é o número de massa, N_A é o número de Avogrado.

4.1.3 Fluorescência De Raios X Por Dispersão Em Energia

A fluorescência de raios X por dispersão em energia (EDXRF) é classificada como uma técnica de emissão atômica, baseada na interação de raios X com a matéria. Quando um feixe de raios X interage com um átomo presente na amostra, elétrons das camadas mais internas do átomo podem ser ejetados devido ao efeito fotoelétrico. Isto cria vacâncias, deixando o átomo em um estado instável. Para estabilizar o átomo, elétrons das camadas orbitais mais externas realizam transições eletrônicas para preencher as vacâncias geradas, liberando a diferença de energia entre os orbitais envolvidos na forma de um fóton de raio X característico,

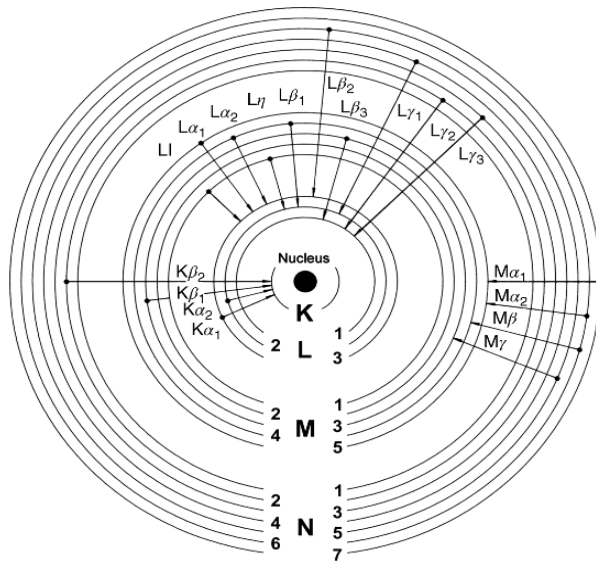
que tem energia bem definida para cada elemento (JENKINS, 1988; VAN GRIEKEN; MARKOWICZ, 2002).

Assim, os raios X característicos podem ser diretamente utilizados na identificação qualitativa dos elementos, e como a sua intensidade é diretamente proporcional à concentração do elemento, a técnica também fornece informações que podem ser utilizadas para fins quantitativos (JENKINS, 1988). Portanto, a EDXRF é comumente utilizada para identificar e quantificar teores totais de espécies inorgânicas do Mg ao U. No entanto, para a faixa de energia na qual a maioria dos equipamentos de EDXRF trabalha, a sensibilidade da técnica é maior para elementos de $19 < Z < 41$. Nessa faixa de energia, é difícil determinar elementos de baixo número atômico ($Z < 18$), principalmente devido ao baixo rendimento fluorescente e à baixa seção de choque de absorção dos raios X (JENKINS, 1988; VAN GRIEKEN; MARKOWICZ, 2002).

Os raios X característicos de cada elemento são identificados de acordo com as camadas envolvidas na transição. Segundo a nomenclatura de Siegbahn (VAN GRIEKEN; MARKOWICZ, 2002), a camada a ser preenchida pelo elétron é que determina a primeira parte do nome. A segunda parte é definida pela camada de origem do elétron, se a camada de origem for a mais próxima, será denominada de α , se for das camadas posteriores serão denominadas sucessivamente β , γ , δ , etc. Por exemplo, quando a vacância situada na camada K for preenchida por um elétron da camada L ou M, ocorrerá a emissão dos raios X $K\alpha$ ou $K\beta$, respectivamente. Quando a vacância na camada L for preenchida por um elétron da camada M ou N, ocorrerá a emissão dos raios X $L\alpha$ ou $L\beta$, respectivamente, conforme ilustrado na Figura 5.

Em cada camada eletrônica existem vários subníveis com valores de energias muito próximos. Cada subnível é caracterizado por uma sequência de números quânticos (n , l , m_l , s , m_s , j e m_j), que dão as informações quânticas do estado.

Figura 5 - Diagrama de transição de raios X e suas respectivas denominações.



Fonte: KLOCKENKÄMPER; VON BOHLEN, 2015.

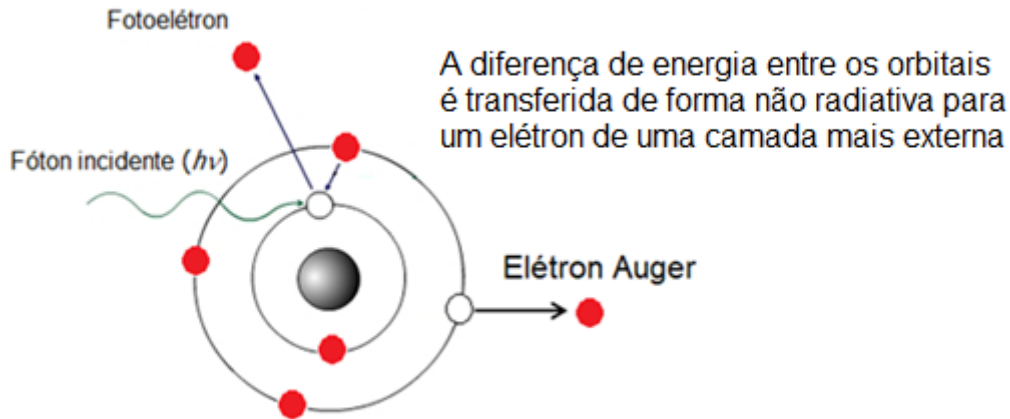
As transições eletrônicas entre os níveis de energia de um dado átomo não ocorrem aleatoriamente, elas são governadas pelas leis da mecânica quântica. Nesse caso, as transições eletrônicas devem obedecer às chamadas regras de seleção, dadas por:

$$\Delta n \neq 0 \quad \Delta l \geq \pm 1 \quad \Delta j = \pm 1 \text{ ou } 0 \quad (8)$$

n = número quântico principal; l = número quântico do momento angular orbital; j = número quântico do momento angular total.

Algumas vezes, ao invés de ocorrer a emissão de raios X característicos em um átomo excitado, o excesso de energia pode ser transferido de forma não radiativa para um elétron de uma camada mais externa, ejetando-o do átomo, conforme ilustra a Figura 6. Esse fenômeno é conhecido como efeito Auger e tem maior relevância para elementos com baixo número atômico (LEO, 1994; VAN GRIEKEN; MARKOWICZ, 2002).

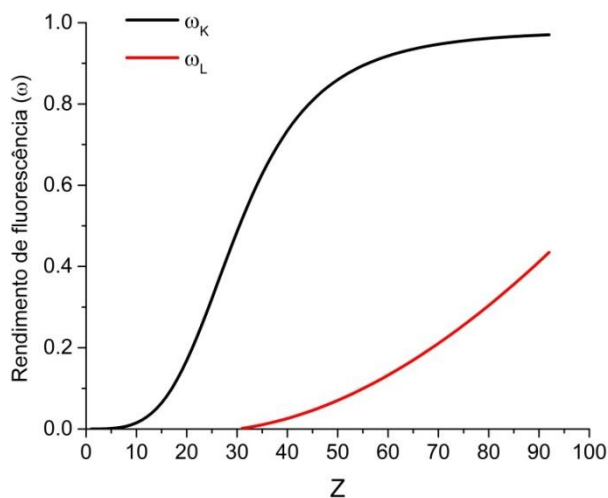
Figura 6 - Representação esquemática do efeito Auger.



Fonte: Adaptado de <<http://slideplayer.com.br/slide/11774338/>>

Com isso, um parâmetro importante a ser considerado na EDXRF é o rendimento de fluorescência, definido como o número de raios X característicos efetivamente emitidos em relação ao número de vacâncias produzidas em uma dada camada (KLOCKENKÄMPER; VON BOHLEN, 2015). A Figura 7 apresenta o rendimento de fluorescência (ω) das camadas K e L com relação ao número atômico Z. Pode-se notar o baixo rendimento de fluorescência da camada K para $Z < 20$ e da camada L para $Z < 60$.

Figura 7 - Rendimento da fluorescência das camadas K e L em função do número atômico Z.

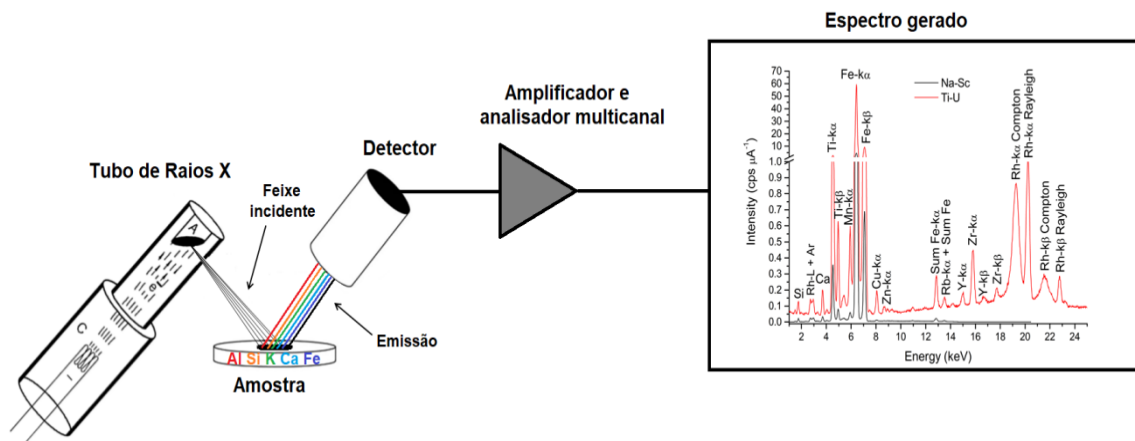


Fonte: do autor (adaptado de VAN GRIEKEN; MARKOWICZ, 2002).

4.1.4 Instrumentação

Um sistema de EDXRF é constituído de uma fonte para a excitação das amostras, um detector que identifica e separa os raios X característicos, uma placa multicanal que registra o espectro obtido e a eletrônica necessária para a alimentação do sistema e amplificação dos sinais provenientes do detector (PARREIRA, 2006), conforme esquema da Figura 8. A fonte de excitação pode ser através de um tubo de raios X, fontes radioativas ou radiação síncrotron. A maioria dos espectrômetros comerciais de EDXRF utiliza tubo de raios X como fonte de excitação, devido à sua versatilidade e por apresentarem menor risco (JANSSENS, 2003).

Figura 8 - Etapas básicas de um sistema de EDXRF

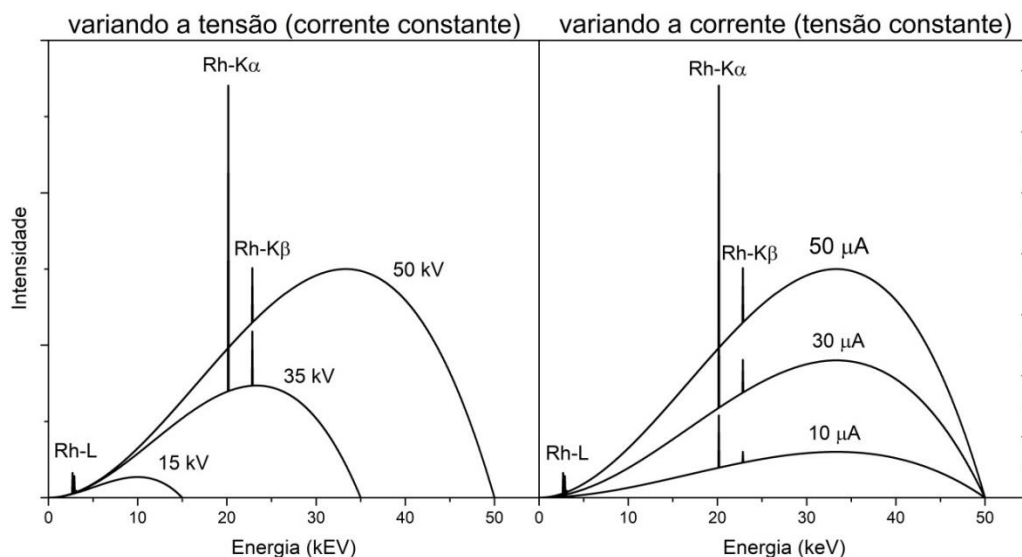


Fonte: do autor.

O tubo de raios X é constituído por um cátodo e um ânodo. No cátodo normalmente se utiliza um filamento de tungstênio que é aquecido para emissão termiônica de elétrons em uma câmara de vácuo. Quanto maior a corrente, mais aquecido será o filamento e mais elétrons serão emitidos do cátodo para o ânodo (JANSSENS, 2003; KLOCKENKÄMPER; VON BOHLEN, 2015). Estes elétrons são acelerados, por uma alta tensão, em direção a um alvo (metal de alta pureza como Cr, Ag, W, Mo, Rh, Pd, ...) que serve como ânodo (JANSSENS, 2003). Devido à interação dos elétrons acelerados com o ânodo, são produzidos raios X de energia contínua (*bremstrahlung*) com linhas características do material ânodo

sobrepostas ao contínuo. Ao aumentar a tensão do tubo, também se aumentam a faixa de energia do espectro contínuo (energia máxima), a intensidade de emissão do espectro contínuo e a energia correspondente à intensidade máxima no espectro contínuo (JENKINS, 1988). Paralelamente, quanto maior o número atômico do material do ânodo e a intensidade de corrente aplicada ao tubo, mais intenso é o feixe de radiação produzido no tubo. A Figura 9 ilustra o comportamento do espectro emitido por um tubo de raios X com ânodo de Rh ao modificar sua tensão e corrente.

Figura 9 - Comportamento do espectro de raios X emitido por um tubo de Rh.



Fonte: do autor (adaptado de KLOCKENKÄMPER; VON BOHLEN, 2015).

O sistema de detecção é composto por um detector semiconductor que pode ser de Si (Li), Si-PIN ou SDD (*Silicon Drift Detector*). Em geral os detectores mais empregados eram os de Si(Li) e algumas vezes o de Ge(Li), sendo dispositivos muito bons para a espectroscopia de raios X, com resolução em energia de aproximadamente 160 eV na energia de 5,9 keV. Porém, esses detectores exigem um sistema de resfriamento, necessitando de grandes bancadas para a sua instalação e levando alguns minutos para serem usados (BUTLER et al., 2010). Atualmente, os detectores Si-PIN e SDD são os mais utilizados em sistemas comerciais de XRF. Estes detectores possuem dimensões reduzidas e não necessitam de resfriamento com nitrogênio líquido, pois são resfriados por efeito *Peltier* (WEST et al., 2012). Dessa forma, é possível trabalhar à temperatura

ambiente, possibilitando diversas aplicações, inclusive análises *in situ* (MELQUIADES; APPOLONI, 2004).

Os raios X são selecionados através de pulsos eletrônicos no detector, sendo que as amplitudes desses pulsos são diretamente proporcionais às energias dos raios X. Assim, é gerado um espectro que mostra a quantidade de fótons de raios X característicos detectados (eixo Y) versus a energia dos mesmos (eixo X). Como a energia de cada elemento é bem definida e o número total de fótons sob um pico característico (área do pico) é proporcional à quantidade daquele elemento na amostra, é possível a identificação e quantificação destes elementos de forma não destrutiva, rápida, simples, com baixo custo operacional, necessitando de mínima e às vezes nenhuma preparação de amostra (JENKINS, 1988). No entanto, existem dois fatores de grande importância que se constituem como os principais limitantes para a definitiva consolidação da técnica de EDXRF como ferramenta analítica. Estes fatores são os efeitos físicos de matriz, como teor de umidade, tamanho das partículas e condições da superfície da amostra, e efeitos químicos de matriz causados pela absorção ou intensificação do sinal de emissão do elemento de interesse por elementos que constituem a matriz amostral (MARGUÍ; QUERALT; HIDALGO, 2009).

4.2 ANÁLISE MULTIVARIADA

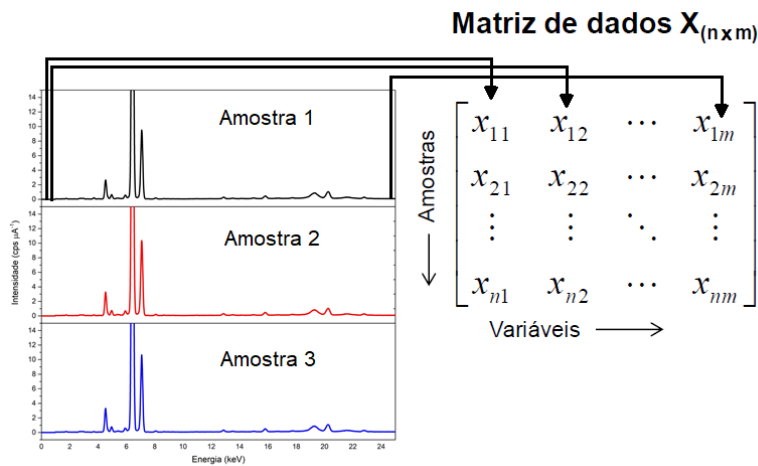
O uso de estatística multivariada permite que muitas variáveis sejam analisadas simultaneamente, possibilitando extrair muito mais informações de um conjunto de dados. Dessa forma, a estatística multivariada aplicada à análise espectroscópica possibilita diversas melhorias, como a maior precisão dos resultados, modelagem na presença de interferentes e a possibilidade de extrair informações que não estão explícitas (FERREIRA, 2015). Em química, esse tipo de análise é conhecida como quimiometria que é definida como o uso de ferramentas estatísticas e matemáticas em dados químicos multivariados (GELADI; KOWALSKI, 1986). Nesse contexto, os métodos quimiométricos têm sido utilizados para diversas finalidades, como processamento de sinais analíticos, planejamento e otimização de experimentos, análise exploratória baseada em reconhecimento de padrões, classificação de dados, calibração multivariada, entre outros (BEEBE et al., 1998;

OTTO, 2007). Neste trabalho foram utilizados apenas análise exploratória e calibração multivariada, mais especificamente análise de componentes principais (PCA, do inglês *Principal Component Analysis*) e regressão por mínimos quadrados parciais (PLSR, do inglês *Partial Least Square Regression*).

4.2.1 Organização E Pré-Processamento Dos Dados

Para análise multivariada, os dados devem ser inicialmente organizados e recomenda-se realizar um pré-tratamento e/ou pré-processamento nos dados para reduzir variações indesejáveis (FERREIRA, 2015). No caso de espectros, os dados são organizados na forma de uma matriz $\mathbf{X}_{(n \times m)}$, na qual as linhas (n) representam as amostras e as colunas (m) representam as variáveis (que são as intensidades em cada canal de energia para o caso da EDXRF), conforme Figura 10.

Figura 10 - Organização da matriz de dados espectrais \mathbf{X} .



Fonte: do autor.

Os pré-tratamentos, tais como alisamento e derivadas, são aplicados nas amostras (nas linhas). Por outro lado, os pré-processamentos são aplicados nas variáveis (nas colunas).

A centragem dos dados na média é o tipo de pré-processamento mais comum em dados espectrais, especialmente em EDXRF (AIDENE et al., 2020; BORTOLETO; BORGES; BUENO, 2007; DA-COL; BUENO; MELQUIADES, 2015;

DOS SANTOS et al., 2020; NOGUEIRA et al., 2019; MELQUIADES et al., 2012; BUENO et al., 2005). Esse pré-processamento previne que os pontos mais distantes do centro dos dados tenham influência maior que os pontos mais próximos. Isso é realizado por meio da subtração do valor médio de cada variável dos dados, resultando em uma translação dos eixos para o valor médio de cada uma das variáveis. Dessa maneira, a estrutura dos dados permanece preservada. Portanto, a centragem na média é usada para focar na variação relevante entre as amostras (BRO; SMILDE, 2003; FERREIRA, 2015; VAN DEN BERG et al., 2006).

Os métodos de escalamento são recomendados quando as variáveis são expressas em diferentes escalas ou a variação dos dados é grande. Esses métodos são usados para igualar o potencial de cada variável, minimizando o efeito de uma variável dominante (BRO; SMILDE, 2003; FERREIRA, 2015; VAN DEN BERG et al., 2006). Neste estudo, foram utilizados o autoescalamento e o escalamento de Pareto. No autoescalamento, todas as variáveis são divididas por seu desvio padrão após serem centradas na média. No entanto, variáveis sem importância que contêm apenas ruído (e, portanto, têm baixo desvio padrão) tornam-se muito importantes. Por outro lado, no escalamento de Pareto, em vez do desvio padrão, é utilizada a raiz quadrada do desvio padrão como fator de escala. Esse pré-processamento reduz a importância de valores altos, mas não tão drasticamente quanto no autoescalamento. Assim, a estrutura dos dados permanece parcialmente inalterada.

O algoritmo Savitzky-Golay é um dos métodos mais comuns para alisamento dos dados, sendo utilizado para o aumento da razão sinal/ruído. Este algoritmo é baseado no ajuste de polinômios para muitas pequenas janelas de dados (FERREIRA, 2015; ZIMMERMANN; KOHLER, 2013).

Em alguns métodos espectroscópicos, os efeitos da linha de base podem ocorrer deixando o sinal com um deslocamento vertical constante ou uma inclinação da linha de base. Isso pode ser corrigido usando uma derivada no sinal de entrada (ENGEL et al., 2013; FERREIRA, 2015). A primeira derivada é eficaz para remover o deslocamento constante da linha de base e a segunda derivada eliminará tanto o deslocamento constante quanto a inclinação da linha de base. Geralmente, o algoritmo de Savitzky-Golay é utilizado para calcular as derivadas dos polinômios ajustados, visto que a derivada pode reduzir a relação sinal/ruído (RINNAN; BERG; ENGELSEN, 2009).

A correção multiplicativa de sinais (MSC, do inglês *Multiplicative Signal Correction*) e a padronização normal de sinal (SNV, do inglês *Standard Normal Variate*) são pré-processamento amplamente utilizado para corrigir efeitos de espalhamento em dados espectrais, especialmente em aplicações com NIR (RINNAN; BERG; ENGELSEN, 2009). Esses efeitos de espalhamento são causados por fenômenos físicos (sensibilidade do sistema de detecção, variações de temperatura, diferenças no tamanho e forma das partículas sólidas das amostras, etc.) que não estão relacionados à composição das amostras.

4.2.2 Análise dos Componentes Principais

A análise de componentes principais (PCA) é classificada como uma análise exploratória que permite a visualização de informações importantes de um conjunto de dados. Dessa forma, a PCA é muito aplicada para reconhecimento de padrões. Esta técnica é baseada na redução da dimensionalidade original dos dados através de uma transformação linear ortogonal para um pequeno número de componentes principais (PC's, do inglês *Principal Components*) que explicam a maior parte da variabilidade dos dados.

Geometricamente, a PCA pode ser definida como uma técnica de projeção, na qual uma matriz \mathbf{X} é projetada num subespaço de dimensões reduzidas. Para isso, a PCA decompõe a matriz original de dados \mathbf{X} em um produto de outras duas matrizes denominadas \mathbf{T} (*scores*) e \mathbf{P} (*loadings*), podendo ser representado pela Equação 9. Os resíduos desta decomposição ficam contidos em uma matriz de erro \mathbf{E} (BEEBE et al., 1998; FERREIRA, 2015; WOLD; ESBENSEN; GELADI, 1987).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^A \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E} \quad (9)$$

onde \mathbf{T} é a matriz de *scores* com n linhas (amostras) e A colunas (componentes principais); \mathbf{P} é a matriz de *loadings* com m linhas (variáveis) e A colunas (componentes principais); \mathbf{t}_i é o vetor de *scores* que representa a i ésima coluna da matriz de *scores* \mathbf{T} e \mathbf{p}_i é o vetor de *loadings*.

Cada PC é construída pela combinação linear das variáveis originais. A primeira componente principal (PC1) é traçada na direção da maior variabilidade nos dados, a segunda (PC2) é traçada ortogonalmente à primeira, descrevendo a maior variação não contida na PC1 e assim sucessivamente. Os *scores* representam as novas coordenadas das amostras no sistema de eixos formado pelas PC's. Dessa forma, por meio da representação gráfica dos *scores* é possível verificar similaridades e diferenças entre as amostras. Os *loadings* são os coeficientes da combinação linear, representando o quanto cada variável original contribui para a formação de cada PC (BEEBE et al., 1998; FERREIRA, 2015; WOLD; ESBENSEN; GELADI, 1987). Ou seja, os *loadings* são os cossenos do ângulo entre o eixo da PC e o eixo das variáveis originais. Dessa forma, quanto mais próximo de ± 1 , maior a influência que determinada variável tem na descrição da PC.

Existem diversos algoritmos descritos na literatura para calcular as matrizes **T** e **P** a partir da matriz **X**. Dentre os principais estão a decomposição por valores singulares (SVD, do inglês *Singular Values Deconvolution*), diagonalização da matriz de correlação (ou covariância) e o algoritmo NIPALS (*Non Linear Iterative Partial Least Squares*), o qual foi utilizado neste estudo (FERREIRA, 2015; WOLD; ESBENSEN; GELADI, 1987).

4.2.3 Regressão por Mínimos Quadrados Parciais

A regressão por mínimos quadrados parciais (PLSR, do inglês *Partial Least Square Regression*) é um dos métodos mais utilizados na literatura para o desenvolvimento de modelos de calibração multivariada, oferecendo a possibilidade da análise de dados como espectros, que apresentam sinais sobrepostos e determinações simultâneas de vários constituintes (FERREIRA; BRAGA; SENA, 2013; GÓMEZ-NUBLA et al., 2018; GONZAGA; PASQUINI, 2010; SIMEONE et al., 2017). Esse método permite estabelecer uma relação entre uma matriz **X** com as variáveis independentes e uma matriz **Y** (ou um vetor de dados **y**) com as variáveis dependentes. Para isso, inicialmente as matrizes **X** e **Y** são decompostas por PCA, conforme segue:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum_{i=1}^A \mathbf{t}_i \mathbf{p}_i^T + \mathbf{E} \quad (10)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F} = \sum_{i=1}^A \mathbf{u}_i \mathbf{q}_i^T + \mathbf{F} \quad (11)$$

onde: \mathbf{X} é a matriz das variáveis independentes; \mathbf{Y} é a matriz das variáveis dependentes; \mathbf{T} e \mathbf{U} são as matrizes dos *scores*; \mathbf{P} e \mathbf{Q} são as matrizes dos *loadings*; \mathbf{E} e \mathbf{F} são as matrizes dos erros e o subíndice i representa o número de variáveis latentes que vai de 1 até A .

Como na PCA a decomposição das matrizes \mathbf{X} e \mathbf{Y} são realizadas de forma independente, pode-se obter uma relação não linear entre os *scores* dessas matrizes. Portanto, no método PLSR a informação de \mathbf{Y} é incorporada de forma que cada PC do modelo sofre uma leve rotação até que se atinja a máxima covariância entre \mathbf{X} e \mathbf{Y} . Devido a isso, há uma perda de ortogonalidade das componentes principais, que passam a receber a terminologia de variáveis latentes (LV, do inglês *Latent Variables*) (BEEBE et al., 1998; FERREIRA, 2015; GELADI; KOWALSKI, 1986).

Todo esse processo é realizado em várias etapas que são repetidas para cada PC a partir de um processo iterativo, no qual se otimiza ao mesmo tempo a projeção das amostras sobre os *loadings* para a determinação dos *scores* e o ajuste linear entre os *scores* de \mathbf{X} e \mathbf{Y} , de acordo com a Equação 12. Deste modo, o método PLSR busca um compromisso entre a explicação da variância em \mathbf{X} e a obtenção da maior correlação com \mathbf{Y} .

$$\mathbf{u}_i = b_i \mathbf{t}_i ; \quad b_i = \frac{\mathbf{u}_i^T \mathbf{t}_i}{\mathbf{t}_i^T \mathbf{t}_i} \quad (12)$$

Há duas variantes do método PLS, conhecidas como PLS1 e PLS2. No método PLS1 cada coluna \mathbf{y} é modelada individualmente. Ou seja, para cada parâmetro de interesse é calculado um conjunto separado de *scores* e variáveis latentes. No método PLS2 todas as variáveis (parâmetros de interesse) de \mathbf{Y} contribuem para modelagem. Isso pode melhorar o desempenho do modelo, principalmente quando os dados nas diferentes colunas de \mathbf{Y} estão correlacionadas

(FERREIRA, 2015). Neste trabalho foi utilizado o método PLS1. Assim, levando em conta a Equação 12, a decomposição do vetor \mathbf{y} fica:

$$\mathbf{y} = \mathbf{T}\mathbf{q}^T + \mathbf{f} = \sum_{i=1}^A \mathbf{t}_i q_i + \mathbf{f} \quad (13)$$

Atualmente, existem diversos algoritmos descritos na literatura para calcular os parâmetros de um modelo PLSR (ANDERSSON, 2009), dentre os quais destacam-se o algoritmo de scores não ortogonalizado proposto por Martens & Naes (1992) e o mais conhecido e utilizado, NIPALS, proposto por Wold (GELADI; KOWALSKI, 1986). A Tabela 1 explica as etapas do algoritmo NIPALS de modo a levar em consideração a informação de \mathbf{y} na decomposição da matriz \mathbf{X} . Nesse caso, considera-se um fator peso \mathbf{W} , que substitui a matriz de *loadings* \mathbf{P} para garantir a ortogonalidade de \mathbf{T} e para que se encontre a melhor covariância entre \mathbf{X} e \mathbf{Y} .

Tabela 1 - Etapas do algoritmo NIPLS para o método PLSR.

Etapa	Descrição	Equação
1	Pré-processar \mathbf{X} e \mathbf{y}	
2	Considere	$\mathbf{X} = \mathbf{X}_0$ e $\mathbf{y} = \mathbf{y}_0$
3	Calcule o fator-peso que é a projeção das variáveis de \mathbf{X} sobre \mathbf{y}_{i-1} . Isso assegura que os vetores \mathbf{t} e \mathbf{y} sejam altamente correlacionados.	$\mathbf{w}_i = \mathbf{X}_{i-1}^T \mathbf{y}_{i-1} (\mathbf{y}_{i-1}^T \mathbf{y}_{i-1})^{-1}$
4	Normalize o fator-peso:	$\mathbf{w}_{i,norm} = \mathbf{w}_i (\ \mathbf{w}_i\)^{-1}$
5	Calcule os <i>scores</i> de \mathbf{X} :	$\mathbf{t}_i = \mathbf{X}_{i-1}^T \mathbf{w}_{i,norm} (\mathbf{w}_i^T \mathbf{w}_i)^{-1}$
6	Estime os <i>loadings</i> de \mathbf{X} que são os coeficientes de regressão da projeção das variáveis de \mathbf{X}_i sobre \mathbf{t}_i .	$\mathbf{p}_i = \mathbf{X}_{i-1}^T \mathbf{t}_i (\mathbf{t}_i^T \mathbf{t}_i)^{-1}$
7	Estime os <i>loadings</i> de \mathbf{y} que são os coeficientes de regressão de \mathbf{y}_i sobre \mathbf{t}_i .	$q_i = (\mathbf{t}_i^T \mathbf{t}_i)^{-1} \mathbf{t}_i^T \mathbf{y}_{i-1}$
8	Remova os efeitos do <i>i</i> ésimo fator da matriz original. Para $i=1, 2, \dots, A$.	$\mathbf{X}_i = \mathbf{X}_{i-1} - \mathbf{t}_i \mathbf{p}_i^T$ $\mathbf{y}_i = \mathbf{y}_{i-1} - \mathbf{t}_i q_i$
9	Repita as etapas de 3 a 8 até A variáveis latentes	
10	Calcule o vetor de regressão PLSR	$\mathbf{b} = \mathbf{W}(\mathbf{P}^T \mathbf{W})^{-1} \mathbf{q}$

Fonte: Adaptado de (FERREIRA, 2015).

5 METODOLOGIA

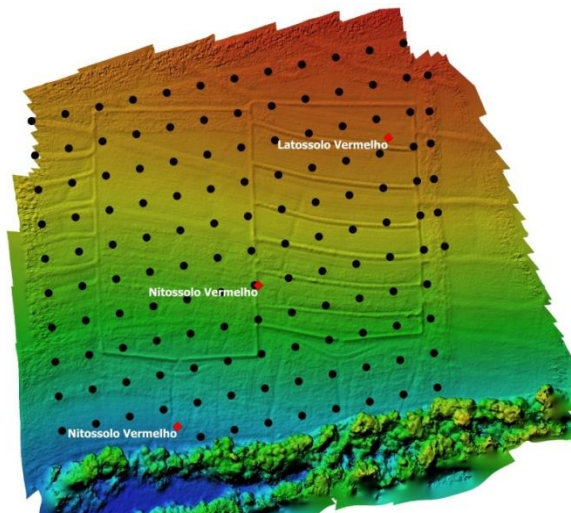
5.1 AMOSTRAGEM

Este estudo foi realizado na bacia hidrográfica do Ribeirão Vermelho, localizado no município de Cambé-PR, entre as latitudes de 23°09'59.70" a 23°09'50.14" S e longitudes de 51°14'42.27" a 51°14'56.87" W. A área encontra-se no Terceiro Planalto do Paraná, apresenta solos classificados como Latossolo Vermelho Eutroférico (Rhodic Ferralsol) e Nitossolo Vermelho Eutroférico (Rhodic Nitisol) (IUSS WORKING GROUP WRB, 2015; SANTOS et al., 2006), altamente intemperizados, oriundos de basalto, com textura muito argilosa. A declividade média da área é de 0,09 m m⁻¹, onde a altitude varia entre 544 m a 514 m (DOS SANTOS, 2019).

A área estudada é manejada com plantio direto há mais de 14 anos, mas a escarificação é realizada em média a cada cinco anos. As culturas utilizadas são predominantemente soja cultivada no verão e milho cultivado no inverno. A aveia é cultivada esporadicamente no inverno.

A amostragem foi realizada em malha regular de 28 X 28 m com 6 pontos adicionais (fora do grid), conforme Figura 11. As amostras foram coletadas nas profundidades de 0 a 5 cm, 5 a 10 cm e 10 a 20 cm em cada ponto amostral, totalizando 396 amostras.

Figura 11 - Localização dos pontos amostrais de acordo com a topografia do terreno.



Fonte: (DOS SANTOS, 2019).

5.2 ANÁLISES CONVENCIONAIS

Os parâmetros analisados foram: carbono orgânico total (COT), fósforo disponível (P_{disp}), cálcio trocável (Ca^{2+}), magnésio trocável (Mg^{2+}), potássio trocável (K^+), soma de bases trocáveis (SB), capacidade de troca catiônica total (CTC), saturação por bases (V%), pH, e acidez potencial ($\text{H}^+ + \text{Al}^{3+}$). Todas as análises foram realizadas no Laboratório de Solos e Tecido Vegetal do IAPAR, seguindo as recomendações do manual de Pavan, et al. (1992).

Os teores de COT foram determinados pelo método de oxidação de dicromato (Walkley-Black). Para isso, 0,5 g de amostra de solo foram transferidas para um balão Erlenmeyer de 250,0 mL, ao qual foram adicionados 10,0 mL de solução de dicromato de potássio $0,4 \text{ mol L}^{-1}$ ($\text{K}_2\text{Cr}_2\text{O}_7$). Em seguida, um tubo de ensaio (25,0 mm de diâmetro e 250,0 mm de altura), cheio de água, foi colocado no balão Erlenmeyer (isso funciona como um sistema condensador). Esse sistema foi aquecido e após o resfriamento adicionou-se 80,0 mL de água destilada, 2,0 mL de ácido ortofosfórico (H_3PO_4) e três gotas do indicador de difenilamina. Após isso, o sistema foi titulado lentamente com uma solução de sulfato ferroso de amônio $0,1 \text{ mol L}^{-1}$ até a cor mudar de azul para verde. Uma amostra branco também foi titulada. Dessa forma, o COT foi determinado pela Equação 14 (DONAGEMMA, et al., 2011), onde V_s é o volume (mL) usado na titulação do sistema, V_b é o volume (mL) usado na titulação da amostra branco e 0,6 é um fator utilizado para transformações unitárias.

$$\text{COT (g kg}^{-1}\text{)} = (40 - V_s) \left(\frac{40}{V_b} \right) 0.6 \quad (14)$$

O Ca^{2+} e Mg^{2+} foram determinados utilizando o extrator KCl 1 mol L^{-1} , sendo a solução medida por espectrometria de absorção atômica. Utilizando a solução extratora Mehlich-1, determinaram-se os teores de K^+ e P_{disp} , por fotometria de chama e espectrofotometria de absorção molecular UV-Vis, respectivamente.

SB, CTC e V% foram estimados pelas seguintes equações:

$$\text{SB} = \text{Ca}^{2+} + \text{Mg}^{2+} + \text{K}^+ \quad (15)$$

$$\text{CTC} = \text{Ca}^{2+} + \text{Mg}^{2+} + \text{K}^+ + \text{H}^+ + \text{Al}^{3+} \quad (16)$$

$$V\% = \frac{100 \times SB}{CTC} \quad (17)$$

As leituras de pH foram realizadas por meio de potenciômetro em solução de CaCl_2 na proporção de 1:2,5. Adicionando-se a solução tampão SMP, foram realizadas as leituras do pH-SMP para determinar os valores de $\text{H}^+ + \text{Al}^{3+}$.

5.3 MEDIDAS POR EDXRF

Para realizar as medidas por EDXRF, retirou-se uma alíquota de 5 g de cada amostra, acondicionando-as em suporte porta amostra de polietileno (n. 1530, Chemplex Industries Inc., EUA) coberto com filme Mylar de 2,5 μm de espessura. Dessa forma, as amostras ficaram com uma espessura de aproximadamente 10 mm dentro dos recipientes, o que garante a condição de amostra infinita (KALNICKY; SINGHVI, 2001). As medidas foram realizadas em um equipamento de bancada da Shimadzu, modelo EDX720, pertencente ao Laboratório de Análise por Raios X (LARX) da UEL.

A excitação foi realizada com um tubo de raio X de Rh de 50 W em atmosfera de ar. Com objetivo de ter uma adequada excitação para uma ampla faixa de elementos, o tubo de raio X foi operado em duas condições de medidas (15 kV por 200s, chamada neste estudo de faixa Na-Sc e 50 kV por 100 s, chamada de faixa Ti-U), em que a corrente foi ajustada automaticamente para um tempo morto máximo de 30% e nenhum filtro foi usado no feixe primário. A detecção foi realizada por um detector de Si (Li) resfriado com nitrogênio líquido e os espectros foram gravados em 2048 canais.

A condição de tensão no tubo de raio X afeta a intensidade das linhas de emissão de diferentes elementos. A voltagem de 15 kV (faixa Na-Sc) favorece a excitação dos elementos com $Z < 20$. Por outro lado, a intensidade de emissão dos elementos com $Z > 20$ é maior quando é aplicado 50 kV no tubo (faixa Ti-U) (TAVARES et al., 2020a; VAN GRIEKEN; MARKOWICZ, 2002).

Foram realizadas três medidas de cada amostra, agitando-se o suporte porta amostra antes de cada medição. Este procedimento foi realizado para irradiar diferentes áreas da amostra, garantindo um espectro representativo.

5.3.1 Quantificação Do Teor Total De Elementos Inorgânicos

A quantificação elementar das amostras de solo se deu pela técnica de EDXRF. Esta técnica relaciona a concentração dos elementos presentes na amostra com a intensidade dos picos de cada um desses elementos. Este procedimento geralmente é realizado pelo método de Parâmetros Fundamentais (PF), o qual baseia-se na modelagem teórica dos fenômenos físicos fundamentais do processo de fluorescência, utilizando fatores de probabilidade, coeficientes de atenuação para cada elemento e valores experimentais fundamentais (PAVLINSKY et al., 2006). Porém, para matrizes de solo o método de PF torna-se complexo, nesse caso, a análise quantitativa pode ser realizada pelo método da curva de calibração. No entanto, para algumas aplicações, é difícil obter um número suficiente de padrões certificados com matrizes semelhantes às amostras estudadas. Nesses casos é possível calibrar o sistema de EDXRF, de acordo com a Equação 18, utilizando um material de referência certificado (MRC) de matriz semelhante às amostras a serem analisadas e validar esta calibração por meio de outros MRCs.

$$I_i = K_i \cdot C_i \quad (18)$$

Onde: I_i é a intensidade do raio X característico do elemento i ; K_i é um coeficiente de calibração que corresponde ao produto da sensibilidade com o fator de absorção do elemento i e C_i é a concentração do elemento i presente na amostra.

Portanto, neste estudo, os coeficientes de calibração do Al, Si, P, S, K, Ca, Ti, Mn, Fe, Cu e Zn foram determinados utilizando um MRC de solo argiloso da Agência Internacional de Energia Atômica (PTXRF_IAEA04). Para validar esta calibração, foram utilizados outros quatro MRCs (de sedimento argiloso de rio (IAEA09), solo argiloso (IAEA13), argila plástica (IPT32) e argila São Simão (IPT42)). Todos esses MRCs foram medidos por seis vezes no equipamento de EDXRF.

Para determinar os coeficientes de calibração do Al, Si, P, S, K, Ca utilizou-se o espectro na faixa de Na-Sc. Por outro lado, os coeficientes do Ti, Mn, Fe, Cu e Zn foram determinados utilizando as medidas na faixa de Ti-U.

Além disso, também foram determinados os limites de detecção (LD) e quantificação (LQ) para cada elemento de interesse de acordo com as Equações (19) e (20).

$$LD = \frac{3}{K_i} \sqrt{\frac{I_{i(BG)}}{t \cdot A}} \quad (19)$$

$$LQ = \frac{10}{K_i} \sqrt{\frac{I_{i(BG)}}{t \cdot A}} \quad (20)$$

Onde: LD = Limite de detecção para o analito; $I_{i(BG)}$ = intensidade (cps μA^{-1}) do background sob o pico do analito; K_i = coeficiente de calibração do analito; t = tempo de excitação/detecção; A = intensidade de corrente no tubo; LQ = limite de quantificação.

5.4 ANÁLISE EXPLORATÓRIA

Para verificar o comportamento dos dados dos atributos de fertilidade do solo determinados pelos métodos convencionais e das concentrações elementares determinadas por EDXRF foi realizado, inicialmente, análise da estatística descritiva dos dados. Em seguida, com o objetivo de identificar correlações entre os atributos de fertilidade e concentrações elementares, bem como encontrar similaridades entre as amostras para relacioná-las com suas características, foi desenvolvida a análise de correlação de Pearson e análise de componentes principais (PCA). Além disso, também se realizou PCA com os dados espectrais de EDXRF.

Antes de desenvolver qualquer modelo multivariado, recomenda-se um pré-processamento dos dados para reduzir variações indesejáveis devido à amostras anômalas e ruídos experimentais (FERREIRA et al., 1999). Dessa forma, para realizar os modelos de PCA, os dados dos parâmetros de fertilidade do solo e de concentrações elementares foram autoescalados. Por outro lado, para os dados espectrais testaram-se dois métodos de pré-processamento, sendo a centragem na média e o escalamento de Pareto.

5.5 PRÉ-PROCESSAMENTOS

Os modelos de calibração multivariada foram construídos com dados espectrais de EDXRF, os quais foram pré-processados para construção dos modelos. Neste estudo, antes de qualquer pré-processamento, a parte final dos espectros, que corresponde a uma região de fundo sem picos característicos (de 15,2 a 20,0 keV para a faixa Na-Sc e de 26,4 a 40,0 keV para a faixa Ti-U), foi removida. Dessa forma, o espectro de Na-Sc ficou com 1472 variáveis e o espectro de Ti-U ficou com 1270 variáveis.

Foram testadas 23 técnicas de pré-processamento espectral, tais como: centragem na média (MC, do inglês *Mean Centering*), autoescalamamento (Auto), escalamento de Pareto (Pareto), alisamento pelo método de Savitzky-Golay (SG) com ajuste polinomial de terceira ordem e janela de 11 pontos, primeira derivada (1º der) usando o algoritmo de Savitzky-Golay com polinômio de 3ª ordem e 11 pontos de dados, segunda derivada (2º der) usando o algoritmo de Savitzky-Golay com polinômio de 3ª ordem e 11 pontos de dados, correção multiplicativa de sinais (MSC), padronização normal de sinal (SNV), bem como suas combinações.

A maioria desses métodos de pré-processamento são aplicados e explorados com mais detalhes em espectroscopia vibracional, sendo que faltam testes avaliando os efeitos desses pré-processamentos em dados espectrais de EDXRF, bem como na capacidade preditiva dos modelos de calibração multivariada. Portanto, nesse estudo foram avaliados os efeitos das técnicas de pré-processamento descritas na Tabela 2 para previsão do COT. Para a modelagem dos demais atributos de fertilidade foram utilizados somente os pré-processamentos que se mostraram mais adequados para os dados espectrais de EDXRF.

Tabela 2 - Técnicas de pré-processamento aplicadas nos dados espectrais.

Nenhum	SG	SNV	MSC+MC	2ºder+Auto	1ºder+Pareto
MC	1º der	SG+MC	SNV+MC	MSC+Auto	2ºder+Pareto
Auto	2º der	1ºder+MC	SG+Auto	SNV+Auto	MSC+Pareto
Pareto	MSC	2ºder+MC	1ºder+Auto	SG+Pareto	SNV+Pareto

5.6 CALIBRAÇÃO MULTIVARIADA

Antes de desenvolver os modelos de calibração, o conjunto de dados foi dividido em 2/3 para calibração (264 amostras) e 1/3 para validação (132 amostras). Esse procedimento foi realizado por meio do algoritmo de amostragem uniforme Kennard-Stone que levou em consideração a matriz com dados espectrais de EDXRF ($\mathbf{X}_{396 \times 2742}$), no qual as variáveis espectrais da faixa de Na-Sc e Ti-U foram concatenadas (KENNARD; STONE, 1969). Esse procedimento foi realizado com o objetivo de ter o mesmo conjunto de dados de calibração e previsão para todos os parâmetros de fertilidade do solo em ambas as condições de medidas de EDXRF.

Neste estudo, foi utilizada regressão por mínimos quadrados parciais (PLSR) para desenvolver os modelos de calibração multivariada para determinação dos parâmetros de fertilidade do solo com dados espectrais de EDXRF. O PLSR estabelece uma relação quantitativa entre uma matriz \mathbf{X} , com respostas analíticas (dados espectrais de EDXRF, neste caso), e um vetor \mathbf{y} , com a propriedade de interesse, através de suas decomposições sucessivas e simultâneas, conforme o algoritmo da Tabela 1 (GELADI; KOWALSKI, 1986).

O número de variáveis latentes (LV) nos modelos PLSR foi escolhido de acordo com o valor mínimo da raiz do erro quadrático médio de validação cruzada (RMSECV, do inglês *Root Mean Square Error of Cross-Validation*). Essa otimização foi realizada utilizando o procedimento *10-fold*, que consiste em dividir aleatoriamente o conjunto total de calibração em 10 subconjuntos de tamanho aproximadamente igual. Um desses subconjuntos é retirado para validar o modelo construído pelos demais. Esse processo é, então, repetido para os demais subconjuntos. Dessa forma, os erros da validação cruzada de cada subconjunto são utilizados para produzir uma única estimativa de RMSECV.

A detecção de *outliers* nos modelos PLSR baseou-se em dados com *leverage* extremos e resíduos não modelados nos dados espectrais, de acordo com a norma ASTM E1655-00 (FERREIRA; BRAGA; SENA, 2013; VALDERRAMA; BRAGA; POPPI, 2007).

O *leverage* representa o grau em que uma amostra está distante do conjunto de dados e em modelos PLSR pode ser representado por:

$$h_i = \hat{\mathbf{t}}_i^T (\hat{\mathbf{T}}^T \hat{\mathbf{T}})^{-1} \hat{\mathbf{t}}_i \quad (21)$$

Onde $\hat{\mathbf{T}}$ é a matriz de scores de todas as amostras e $\hat{\mathbf{t}}_i$ é o vetor de score da i ésima amostra.

De acordo com a ASTM E1655-00, amostras com $h_i > 3 \frac{(LV+1)}{n}$ são consideradas *outlier*, onde LV é número de variáveis latentes do modelo e n é o número de amostras de calibração (VALDERRAMA; BRAGA; POPPI, 2007).

A identificação de *outliers* baseados em resíduos não modelados nos dados espectrais é obtida por comparação do desvio padrão residual total ($s(\hat{\epsilon})$) com o desvio padrão residual da i ésima amostra ($s(\hat{\epsilon}_i)$). Neste estudo, amostras com $(s(\hat{\epsilon}_i)) > 3 (s(\hat{\epsilon}))$ foram consideradas *outliers*.

Os modelos PLSR foram construídos com os pré-processamentos descritos acima e seus desempenhos foram avaliados de acordo com o valor mínimo da raiz do erro quadrático médio de validação cruzada (RMSECV) e previsão (RMSEP). Após isso, diferentes métodos de seleção de variáveis foram testados com o objetivo de melhorar a interpretação dos modelos, verificando as variáveis de EDXRF mais importantes na modelagem dos parâmetros de interesse. O desempenho dos métodos de seleção de variáveis também foi avaliado pelo RMSE e coeficiente de correlação ao quadrado (R^2) de calibração, validação cruzada e previsão:

$$RMSECV = \sqrt{\frac{1}{n_c} \sum_{i=1}^{n_c} (\hat{y}_i - y_i)^2} \quad (22)$$

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_p} (y_i - \hat{y}_i)^2}{n_p}} \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (24)$$

onde \hat{y}_i e y_i são os valores previstos e medidos da amostra i nos conjuntos de calibração e previsão; \bar{y} é a média do método de referência para todas as amostras nos conjuntos de calibração e previsão e n é o número de amostras de calibração

(n_c) e previsão (n_p).

Além disso, a razão de desempenho de desvio (RPD, do inglês *Ratio Performance to Deviation*) também foi utilizada para avaliar o desempenho dos modelos. O RPD (Eq. 25) compara o RMSEP com o desvio padrão do conjunto de previsão (DP_{pred}).

$$RPD = \frac{DP_{pred}}{RMSEP} \quad (25)$$

Neste estudo, grupos arbitrários foram utilizados para simplificação da interpretação, conforme proposto por Viscarra Rossel; Mcglynn; Mcbratney (2006): (1) modelo excelente, que pode ser considerado estável e acurado ($RPD > 2,5$); (2) modelo bom, que pode ser considerado acurado ($2,5 > RPD > 2,0$); (3) modelo razoável, que pode ser utilizado para avaliação quantitativa ($2,0 > RPD > 1,8$); (4) modelo fraco, que pode ser utilizado apenas para uma avaliação qualitativa e correlação ($1,8 > RPD > 1,4$); (5) modelo muito ruim ($RPD < 1,4$).

Para comparar o desempenho dos modelos com relação ao pré-processamento e seleção de variáveis, foi realizado um teste de randomização com nível de significância de 0,05, conforme descrito por Van der Voet (1994). Esse teste consiste em avaliar os resíduos de previsão dos modelos a serem comparados, conforme as hipóteses: hipótese nula (H_0), onde $RMSEP_A = RMSEP_B$, e a hipótese alternativa (H_1), onde $RMSEP_A \neq RMSEP_B$. Para tal, o teste de randomização executa os seguintes passos:

- (i) Calcula a diferença entre os resíduos quadráticos de previsão dos modelos A e B;

$$d_i = e_{Ai}^2 - e_{Bi}^2$$

- (ii) Calcula a diferença média quadrática dos resíduos dos modelos A e B;

$$\bar{d}_i = \frac{1}{n} \sum_{i=1}^n d_i$$

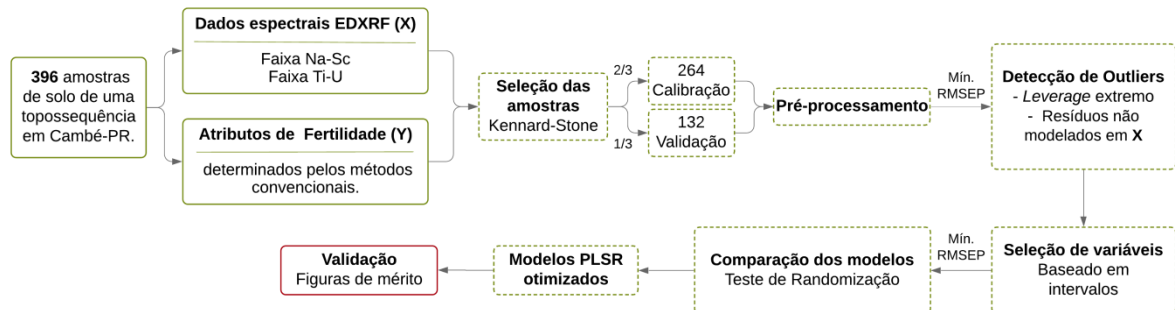
- (iii) Adiciona randomicamente sinais à diferença dos resíduos d_i ;
 (iv) Calcula a diferença média quadrática dos resíduos após o passo (iii), \bar{d}_i^* ;
 (v) Repete as etapas (iii) e (iv) k-vezes.

Neste trabalho, as etapas (iii) e (iv) foram repetidas 1999 vezes e o

p-valor do teste foi determinado pela proporção do número de vezes em que $\bar{d}_i^* > \bar{d}_i$. Se o p-valor for menor que o nível de significância de 0,05, não existem evidências significativas para aceitar H_0 , indicando que os modelos têm exatidões diferentes. Caso contrário, não há evidências para rejeitar H_0 , indicando que não há diferença entre a veracidade dos modelos.

A Figura 12 apresenta um fluxograma detalhando os passos utilizados no desenvolvimento dos modelos de calibração multivariada.

Figura 12 - Fluxograma com os passos utilizados no desenvolvimento dos modelos de calibração multivariada.



Fonte: do autor.

Todas as análises de pré-processamento, seleção de variáveis e calibração multivariada por PLSR foram realizadas no MATLAB R2015a (The MathWorks Inc., Natick, EUA).

5.7 SELEÇÃO DE VARIÁVEIS

A calibração multivariada é dedicada ao estabelecimento de modelos de calibração que relacionam as variáveis da técnica analítica às propriedades de interesse. Notavelmente, no uso de dados espectrais de EDXRF, cada canal de energia (pontos espectrais) é tratado como uma variável na modelagem. Contudo, nem todas as variáveis do espectro contribuem para a modelagem do parâmetro de interesse, visto que muitos conjuntos de dados possuem variáveis irrelevantes, ruidosas ou com pouca influência preditiva (ANDERSEN; BRO, 2010; WANG et al., 2018). Portanto, a remoção destas variáveis pode melhorar as previsões (ou em último caso mantê-las constantes em termos de

erro), reduzindo simultaneamente a complexidade do modelo. Nesse contexto, os métodos de seleção de variáveis têm desempenhado um papel fundamental em calibração multivariada, permitindo uma interpretação mais simples e com maior capacidade preditiva (WANG et al., 2018).

Neste trabalho foi testado o desempenho dos modelos utilizando cinco diferentes métodos de seleção de variáveis baseado em intervalos: mínimos quadrados parciais por intervalos (iPLS), mínimos quadrados parciais por intervalos progressivos (FiPLS), mínimos quadrados parciais por intervalos “em sentido contrário” (BiPLS), mínimos quadrados parciais por intervalos sinérgicos (SiPLS) e algoritmo das projeções sucessivas por intervalos para mínimos quadrados parciais (iSPA-PLS).

O algoritmo iPLS utilizado neste estudo foi desenvolvido por Nørgaard et al. (2000). Ele constrói modelos PLSR locais em subintervalos equidistantes do espectro completo e seleciona o intervalo com o menor RMSECV. Dessa forma, é possível avaliar e identificar os subconjuntos de variáveis que apresentam informações relevantes. Assim, o iPLS dá uma primeira impressão das informações de diferentes subintervalos e localiza o melhor intervalo individual. No entanto, ele não leva em consideração o sinergismo entre os intervalos, o que poderia levar a um modelo com melhor desempenho (WANG et al., 2018).

Os métodos FiPLS, BiPLS e SiPLS são três extensões básicas do iPLS que consideram a combinação de múltiplos intervalos para melhorar a capacidade preditiva dos modelos. Como no iPLS, o FiPLS, BiPLS e SiPLS dividem os dados espectrais em um determinado número de intervalos equidistantes. No FiPLS, depois de selecionar o primeiro intervalo com base no valor mínimo de RMSECV (como no iPLS), novos modelos de PLSR são calculados combinando o primeiro intervalo selecionado com cada um dos outros intervalos restantes, um de cada vez. Novamente, a melhor combinação de dois intervalos é determinada pelo menor RMSECV. Isso é repetido até atingir o mínimo RMSECV possível (XIAOBO et al., 2007). O algoritmo BiPLS usado neste estudo também foi desenvolvido por Nørgaard et al. (2000). Este é um método reverso do FiPLS, no qual os modelos PLSR são calculados com cada intervalo deixado de fora. O primeiro intervalo removido é o que fornece o maior RMSECV. Este procedimento continua até minimizar o RMSECV (LEARDI; NØRGAARD, 2004). O SiPLS é uma estratégia que permite a seleção da melhor combinação entre todas as combinações de dois, três e

quatro intervalos. No final, a combinação com o RMSECV mais baixo é considerada a faixa otimizada.

O algoritmo de projeções sucessivas (SPA) foi originalmente proposto com o objetivo de minimizar os problemas de colinearidade na regressão linear múltipla (MLR) (ARAÚJO et al., 2001). O SPA usa operações de projeção em um espaço vetorial para selecionar subconjuntos de variáveis com colinearidade mínima e, assim, alcançar boa capacidade de previsão. O algoritmo iSPA-PLS, proposto por Gomes et al. (2013), é uma extensão do SPA para selecionar intervalos de variáveis para uso em PLSR. Esse método também divide os dados em um número de intervalos equidistantes, combinando as propriedades de redução de ruído do PLS com a possibilidade de descartar variáveis não informativas do SPA (GOMES et al., 2013; MARIANI et al., 2015).

Neste estudo, o espectro completo de Na-Sc foi dividido em 16 subintervalos equidistantes com 92 variáveis em cada intervalo e o espectro completo de Ti-U foi dividido em 10 intervalos com 127 variáveis cada. Esses intervalos foram definidos tentanfo evitar o corte de parte dos picos.

5.8 FIGURAS DE MÉRITO

No presente trabalho foram avaliadas as seguintes figuras de mérito para os modelos de calibração multivariada: veracidade, precisão, linearidade, erros sistemáticos (*bias*), sensibilidade (SEN), inverso da sensibilidade analítica (γ^{-1}), limites de detecção (LOD), limites de quantificação (LOQ) e robustez. As definições para essas figuras de mérito seguem as recomendações presentes nos protocolos emitidos pela IUPAC (International Union of Pure and Applied Chemistry) (CURRIE, 1999) e pela norma E1655–0017 da ASTM (American Society for Testing and Materials) (ASTM E1655-05, 2012).

5.8.1 Veracidade

De acordo com de acordo com o Vocabulário Internacional de Metrologia (INMETRO, 2012) e a norma ABNT NBR ISO 5725-1:2018 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018), veracidade é o grau de concordância entre a média de um número de medidas repetidas e um valor de

referência. Vale a pena ressaltar que, de acordo com essas normas, precisão é o grau de concordância entre diferentes valores medidos, obtidos por medições repetidas no mesmo objeto ou em objetos similares, sob condições especificadas e exatidão (veracidade + precisão) é o grau de concordância entre o valor medido e um valor de referência (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2018; INMETRO, 2012).

A veracidade dos modelos de calibração multivariada foi avaliada através de parâmetros, como RMSEC, RMSECV e RMSEP. Entre estes, o RMSEP é o parâmetro mais robusto, porque é estimado a partir de amostras de validação externas (VALDERRAMA; BRAGA; POPPI, 2009). Além desses parâmetros, a veracidade dos resultados previstos pela aplicação dos modelos PLSR também foi avaliada pelo erro relativo nos valores previstos (ER), calculado conforme Equação 26, e erros relativos individuais.

$$ER(\%) = 100 \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n y_i^2}} \quad (26)$$

Onde y_i e \hat{y}_i são os valores dos parâmetros do solo determinados pelos métodos convencionais e previstos pelos modelos de calibração multivariada, respectivamente.

5.8.2 Precisão

A precisão expressa o grau de concordância de uma série de medidas realizadas de forma independente para uma mesma amostra homogênea em condições especificadas (VALDERRAMA; BRAGA; POPPI, 2009). Nesse estudo, avaliou-se a precisão no nível de repetitividade, no qual foi determinado o coeficiente de variação, conforme Equação 27, de medidas em triplicata (realizadas pelo mesmo analista e no mesmo dia) de três níveis de concentração cobrindo a faixa de trabalho do modelo de calibração.

$$CV(\%) = \frac{DP}{\bar{y}} 100 \quad (27)$$

Onde DP é o desvio padrão das medidas em triplicada e \bar{y} é a média dessas medidas.

5.8.3 Linearidade

A avaliação da linearidade em modelos de calibração univariada é realizada por meio da análise de variância (ANOVA) e falta de ajuste da regressão (BARROS NETO; SCARMINIO; BRUNS, 2001). No entanto, para calibração multivariada por PLSR essas análises são impossíveis, uma vez que a calibração é construída com uma matriz de variáveis decomposta em A variáveis latentes. Dessa forma, a linearidade em modelos de calibração multivariada geralmente é avaliada verificando a ausência de tendência nos resíduos. No entanto, isso muitas vezes é realizado de forma subjetiva, por meio da inspeção visual de um gráfico de dispersão dos resíduos (VALDERRAMA; BRAGA; POPPI, 2009). Portanto, para avaliar a linearidade dos modelos desse trabalho de forma mais objetiva, foi utilizado o teste de Durbin-Watson (DW) o qual verifica autocorrelações entre sucessivos resíduos do gráfico de resíduos parciais aumentados (APARP, do inglês Augmented Partial Residual Plots), conforme sugerido por Centner; De Noord; Massart, (1998). Este teste examina a hipótese nula de que não há correlação entre os resíduos sucessivos, versus a hipótese alternativa de que a correlação existe. Neste estudo foi estimada a probabilidade associada à estatística DW da Equação 28, em que valores maiores que 0,05 indicam ausência de não linearidades.

$$DW = \frac{\sum_{i=1}^{n-1} (e_{i+1} - e_i)^2}{\sum_{i=1}^{n-1} e_i^2} \quad (28)$$

onde e_i é o resíduo da i -ésima amostra e n é o número de amostras de calibração.

5.8.4 Erros Sistemáticos (*Bias*)

Segundo a IUPAC (CURRIE, 1999), o termo *bias* é atribuído a erros sistemáticos que são calculados pela diferença entre a média populacional e o valor verdadeiro e são todas as componentes de erro que não são aleatórias. A existência desse tipo de erro afeta a exatidão e a determinação dos intervalos de confiança. A

norma E1655-00 da ASTM recomenda a avaliação deste parâmetro através de um teste- t para as amostras de validação no nível de 95% de confiança, conforme Equações 29, 30 e 31 (ASTM E1655-05, 2012). Caso o valor de t -calculado seja maior do que seu valor crítico para n_p-1 graus de liberdade em um nível de significância α , em que n_p corresponde ao número de amostras de previsão, há um indicativo de que os erros sistemáticos presentes no modelo são significativos.

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n_p} \quad (29)$$

$$SDV = \sqrt{\frac{\sum_{i=1}^n [(y_i - \hat{y}_i) - bias]^2}{n_p - 1}} \quad (30)$$

$$t = \frac{|bias| \sqrt{n_p}}{SDV} \quad (31)$$

Onde y_i e \hat{y}_i são os valores dos parâmetros do solo determinados pelos métodos convencionais e previstos pelos modelos de calibração multivariada, respectivamente.

5.8.5 Sensibilidade

Corresponde à fração do sinal responsável pelo acréscimo de uma unidade de concentração da propriedade de interesse (VALDERRAMA; BRAGA; POPPI, 2009). Para modelos de calibração multivariada, como PLSR, pode ser calculado conforme Equação 32.

$$SEN = \frac{1}{\|\mathbf{b}\|} \quad (32)$$

Onde $\|\mathbf{b}\|$ é a norma Euclidianas do vetor de regressão do modelo PLSR.

5.8.6 Inverso Da Sensibilidade Analítica

A sensibilidade analítica apresenta a sensibilidade do método em termos da unidade de concentração que é utilizada, sendo definida como a razão entre a sensibilidade (SEN) e o desvio padrão do ruído instrumental (δ_x) (VALDERRAMA; BRAGA; POPPI, 2009).

Alguns estudos demonstram que a parte final do espectro de EDXRF, que corresponde a uma região de fundo sem picos característicos, geralmente oferece dados suficientes para determinação do ruído experimental (MELQUIADES et al., 2012; MORONA et al., 2017). Nesse caso, δ_x foi obtido calculando o desvio padrão da região de 15,2 a 20,0 keV para a faixa Na-Sc e de 26,4 a 40,0 keV para a faixa Ti-U do conjunto de calibração.

O inverso da sensibilidade analítica (γ^{-1}) fornece uma estimativa da diferença mínima que é discernível pelo método analítico considerando o ruído experimental aleatório como a única fonte de erro. Ou seja, essa figura de mérito indica a resolução do método e, conseqüentemente, o número de casas decimais a ser usado para expressar os valores de previsão (FERREIRA; BRAGA; SENA, 2013).

5.8.7 Limites de detecção e quantificação

O limite de detecção (LOD) equivale à menor concentração do parâmetro de interesse que pode ser detectada, mas não necessariamente quantificada. Por outro lado, o limite de quantificação (LOQ) representa a menor concentração da substância de interesse que pode ser medida com uma incerteza máxima de 10% (CURRIE, 1999). Para modelos PLSR o LOD e LOQ podem ser determinados por meio das equações 33 e 34, respectivamente.

$$LOD = 3,3\delta_x\|\mathbf{b}\| \quad (33)$$

$$LOQ = 10\delta_x\|\mathbf{b}\| \quad (34)$$

Onde δ_x é o desvio padrão do ruído instrumental e $\|\mathbf{b}\|$ é o vetor de regressão do modelo PLSR.

5.8.8 Robustez

De acordo com o INMETRO (INMETRO, 2010), a robustez de um método é a medida de sua capacidade de permanecer inalterado sob pequenas variações de determinados fatores nas quais o método pode estar sujeito como, por exemplo, temperatura, umidade, reagentes, etc. Neste estudo a robustez foi avaliada com relação à granulometria das amostras de solo. Para tal, foram selecionados, por Kennard-Stone (KENNARD; STONE, 1969), 39 amostras do conjunto de previsão, as quais foram maceradas e peneiradas em uma granulometria de 125 μm para posterior análise. Para verificar se houve variação significativa nos resultados de acordo com a granulometria, foi realizado um teste de randomização, em um nível de significância de 0,05 conforme proposto por Dejaegher et al. (2006), comparando os resultados preditivos dessas amostras (39 amostras cobrindo toda a faixa de concentração) com granulometria de 2 mm e 125 μm .

6 RESULTADOS E DISCUSSÕES

6.1 RESULTADOS DOS MÉTODOS CONVENCIONAIS

A Tabela 3 apresenta a estatística descritiva dos atributos de fertilidade do solo determinados pela análise convencional. De acordo com Warrick e Nielsen (1980), valores de CV inferiores a 12 % expressam baixa variabilidade, valores entre 12 e 60 % variabilidade média, e valores acima de 60 %, variabilidade alta. Portanto, o pH e CTC apresentaram baixa variabilidade. Por outro lado, o teores de P_{disp} e K^+ tiveram alta variabilidade na área de estudo. Os demais atributos de fertilidade apresentaram variabilidade média.

As faixas de valores para os atributos de fertilidade do solo reportados na Tabela 3 são semelhantes aos encontrados em Latossolos manejados com sistema de plantio direto (BRIEDIS et al., 2016), que é o caso da área de estudo. Os valores desses atributos também dependem do uso de fertilizantes e insumos. Portanto, a variação de P_{disp} e K^+ no solo pode estar relacionada à aplicação de fertilizantes em linha, bem como à adsorção e mobilidade desses elementos (SANTI et al., 2012).

Tabela 3 - Estatística descritiva dos atributos do solo determinados pelos métodos convencionais.

Atributo	Média ± DP	CV (%)	Med.	Mín.	Máx.	Curt.	Ass.
Todas as amostras (n=396)							
COT	19,8 ± 4,3	22	20,2	7,6	30,5	-0,26	-0,50
P_{disp}	13,9 ± 9,9	71	12,3	0,4	69,8	5,45	1,73
Ca^{2+}	6,0 ± 1,5	25	6,2	1,4	9,2	-0,48	-0,44
Mg^{2+}	2,0 ± 0,6	29	2,0	0,7	3,7	0,56	-0,21
K^+	0,24 ± 0,16	67	0,18	0,05	1,15	3,65	1,59
SB	8,2 ± 1,9	23	8,4	2,1	12,6	-0,53	-0,14
CTC	13,7 ± 1,5	11	13,8	9,3	17,9	-0,20	-0,33
V	60 ± 10	16	61	18	80	-0,92	1,61
pH	5,3 ± 0,4	8	5,2	4,2	6,5	0,32	0,15
$H^+ + Al^{3+}$	5,4 ± 1,1	20	5,3	2,7	9,7	0,31	0,91

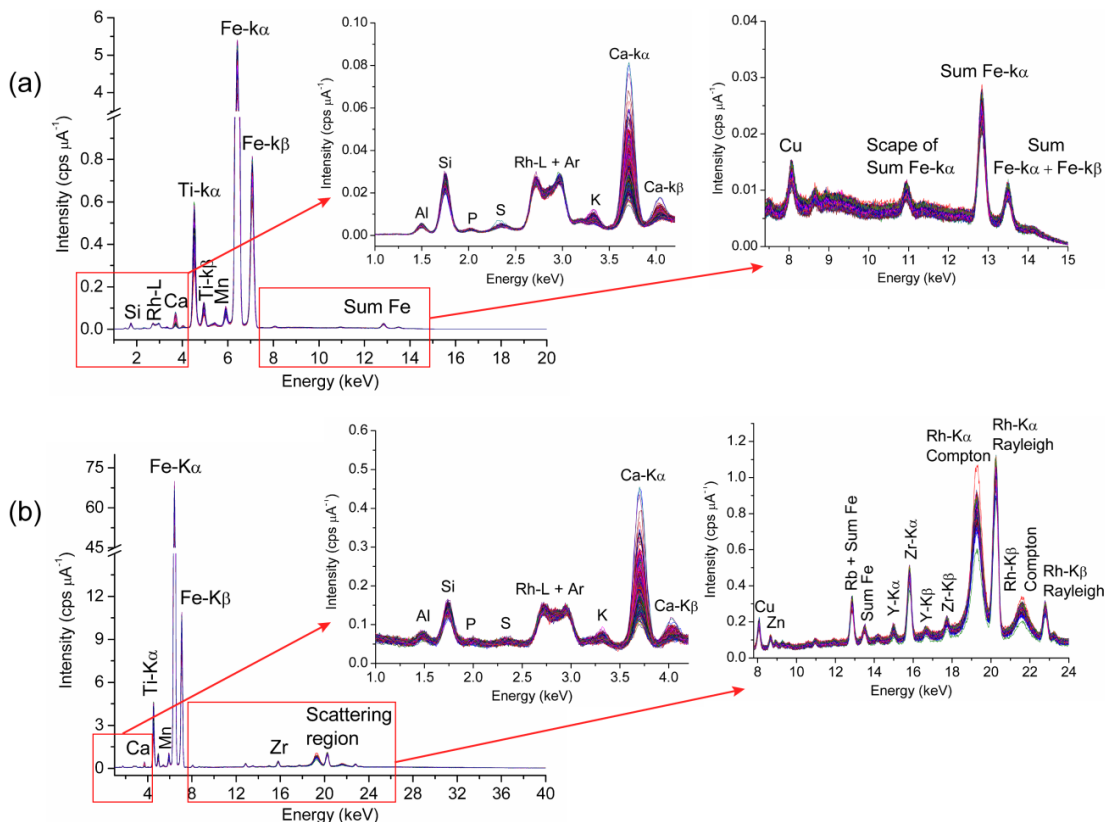
DP = desvio padrão; CV = coeficiente de variação; Med. = mediana; Mín. = mínimo; Máx = máximo; Curt. = curtose; Ass. = assimetria; Valores do COT em $g\ kg^{-1}$, do P_{disp} em $mg\ kg^{-1}$, do Ca^{2+} , Mg^{2+} , K^+ , SB, CTC e $H^+ + Al^{3+}$ em $cmol_c\ kg^{-1}$ e V em %.

6.2 RESULTADOS DE EDXRF

6.2.1 Avaliação Qualitativa

A Figura 13 mostra os espectros de EDXRF na faixa Na-Sc (a) e Ti-U (b) para as 396 amostras de solo medidas neste estudo. Quando é aplicado 15 kV no tubo (faixa Na-Sc), a energia dos raios X não é suficiente para excitar adequadamente os elementos com $Z > 26$. Nesse caso, a faixa de Na-Sc é mais eficiente para excitar os elementos com $Z < 20$ (TAVARES et al., 2020a; VAN GRIEKEN; MARKOWICZ, 2002). Por outro lado, no espectro Ti-U (50 kV) a fluorescência dos elementos com $Z > 20$ é favorecida e os picos de espalhamento (Compton e Rayleigh) do tubo de Rh tornam-se evidentes.

Figura 13 - Espectros de EDXRF das 396 amostras de solo aplicando 15 kV (a) e 50 kV (b) no tubo de raios X.



Fonte: dos Santos et al. (2020).

Além disso, o ganho para a faixa de Na-Sc é melhor do que para a faixa de Ti-U, pois os 2048 canais do espectro são divididos em um intervalo de

energia menor (de 0 a 20 keV). Para a faixa Ti-U, existem 40 keV divididos em 2048 canais de espectro. A partir da Figura 13, pode-se notar também que a resposta instrumental é diferente de acordo com a voltagem aplicada ao tubo de raios X. Dessa forma, avaliou-se a influência dessas condições nos modelos de calibração multivariada.

Foi possível identificar nos espectros (Figura 13) a presença de Al, Si, P, S, K, Ca, Ti, Mn, Fe, Cu, Zn, Rb, Y e Zr, sendo que o pico mais pronunciado é do Fe. Devido a isso, a região de 10 a 14 keV é muito afetada pelos picos soma do Fe. A alta intensidade do pico de Fe está relacionada ao fato de que este elemento é o principal componente formador do solo em estudo (SCHWERTMANN; TAYLOR, 1989).

Os picos de Rh (tubo de raios X com alvo de Rh), especialmente os picos do espalhamento Compton, são devidos principalmente a interações com elementos leves (H, C, N, O) que causam alto espalhamento e baixa absorção da radiação incidente (MELQUIADES; DOS SANTOS, 2015; VERBI; PEREIRA-FILHO; BUENO, 2005).

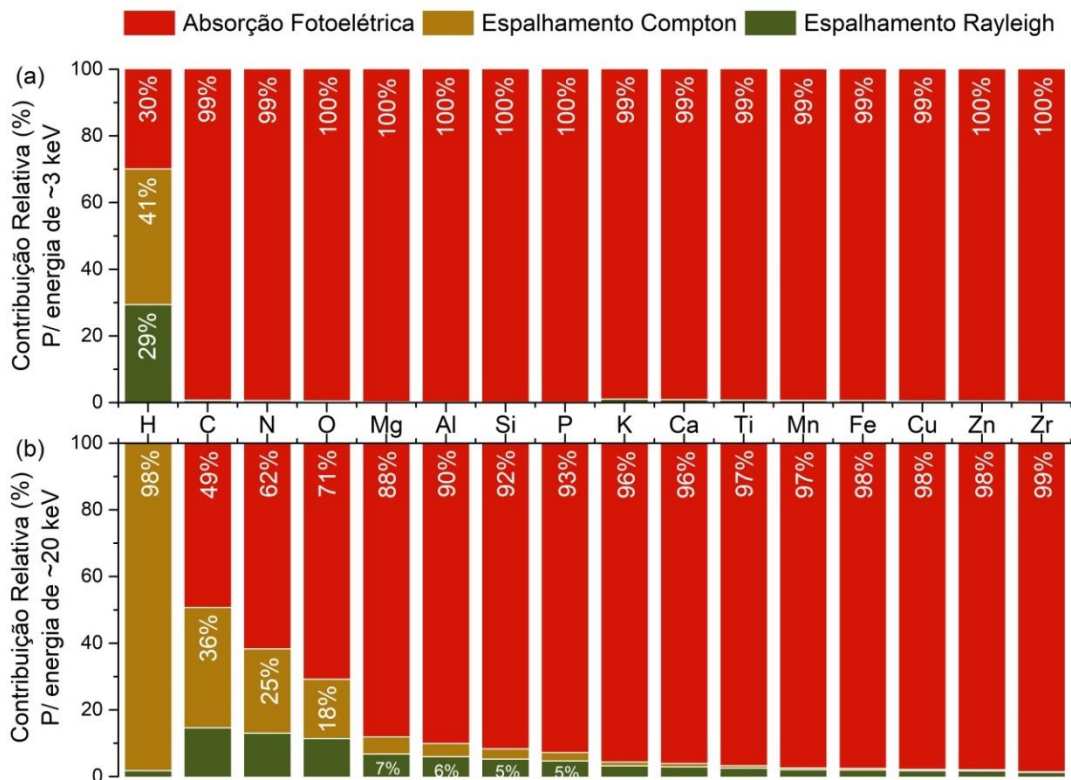
A Figura 14 apresenta uma estimativa da contribuição relativa da absorção fotoelétrica, espalhamento Compton e Rayleigh para a energia de 2,7 (a) e 20,2 keV (b), que correspondem às energias das linhas $L\alpha$ e $K\alpha$ do tubo de Rh, respectivamente. Esses dados foram extraídos do software NIST XCOM, disponível em: <https://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html>.

Pode-se observar na Figura 14, que a probabilidade de ocorrer efeitos de espalhamento é muito pequena para a energia de aproximadamente 3 keV. Nesse caso, o efeito dominante é a absorção fotoelétrica. É importante ressaltar que também existe a possibilidade da radiação incidente passar pela amostra sem nenhuma interação. Por outro lado, para a energia dos picos Rh- $K\alpha$ (20,2 keV) e Rh- $K\beta$ (22,7 keV), a contribuição relativa do espalhamento Rayleigh variou de 1% a 15% para a faixa de elementos identificados nas amostras, sendo que os elementos com maior contribuição relativa foram o C, N e O. Nessa faixa de energia, a contribuição para o espalhamento Compton foi principalmente devido ao H, C, N e O. No entanto, também se observa uma pequena contribuição relativa de aproximadamente 5% do Mg e 3% do Al, Si e P para o espalhamento Compton.

Apesar de haver espalhamento para todos os elementos, na faixa de

energia que a maioria dos equipamentos comerciais de EDXRF trabalham, a contribuição relativa dos efeitos de espalhamento, principalmente para o efeito Compton, é muito pequena para elementos mais pesados que K. Nesse caso, o efeito predominante é a absorção fotoelétrica, que contribui com mais de 95% para os elementos de Z=19 a Z=41 (CESAREO et al., 1992).

Figura 14 - Estimativa da contribuição relativa da absorção fotoelétrica, espalhamento Compton e Rayleigh para a energia de 2,7 (a) e 20,2 keV (b).



Fonte: do autor adaptado de < <https://physics.nist.gov/PhysRefData/Xcom/html/xcom1.html> >

6.2.2 Avaliação Quantitativa

Para calibrar o sistema de medidas de EDXRF, mediu-se por seis vezes um MRC de solo argiloso (PTXRF_IAEA04) e calculou-se a intensidade média para os elementos Al, Si, P, S, K, Ca, Ti, Mn, Fe, Cu e Zn. Por meio das Equações (18), (19) e (20), determinaram-se os coeficientes elementares, limites de detecção e limites de quantificação para cada elemento. A Tabela 4 apresenta as concentrações certificadas, coeficientes elementares, limites de detecção e limites de quantificação para os elementos de interesse.

Utilizou-se os MRCs IAEA09, IAEA13, IPT32 e IPT42 para validar estas calibrações. Estes MRCs também foram medidos por seis vezes. Posteriormente, as intensidades médias dos elementos de interesse foram calculadas para determinar suas concentrações por meio da Equação 18. Os resultados dessas concentrações elementares foram comparados com as concentrações certificadas de acordo com as Tabelas 5, 6, 7 e 8.

Tabela 4 - Concentração certificada do MRC PTXRF_IAEA04, coeficiente elementar, limites de detecção (LD) e limites de quantificação (LQ) do sistema de EDXRF com 95% de nível de confiança.

Elemento	Concentração certificada (%)	Coeficiente	LD (%)	LQ (%)
Al	11,1 ± 0,2	0,0057	0,3	1,0
Si	25,3 ± 0,2	0,0306	0,1	0,3
P	0,0448 ± 0,0014	0,1373	0,0162	0,0541
S	0,0613 ± 0,0019	0,6079	0,0059	0,0195
K	1,96 ± 0,04	0,9082	0,01	0,02
Ca	0,207 ± 0,005	0,9105	0,004	0,015
Ti	0,557 ± 0,001	8,871	0,007	0,024
Mn	0,0193 ± 0,0007	19,64	0,0038	0,0127
Fe	4,79 ± 0,08	34,33	0,002	0,008
Cu	0,0022 ± 0,0001	201,3	0,0004	0,001
Zn	0,0161 ± 0,0006	62,01	0,001	0,004

Tabela 5 - Comparação entre concentrações certificadas e medidas do MRC IAEA09, com as respectivas recuperações.

Elemento	Concentração certificada	Concentração medida	Recuperação
Al	5,94 ± 0,09	6,80 ± 0,19	114 ± 3
Si	32,2 ± 0,3	32,2 ± 0,3	100 ± 1
P	0,110 ± 0,003	0,143 ± 0,006	130 ± 5
S	0,0328 ± 0,0011	0,054 ± 0,004	164 ± 11
K	1,95 ± 0,04	1,896 ± 0,014	97 ± 1
Ca	1,38 ± 0,03	1,38 ± 0,07	100 ± 5
Ti	0,430 ± 0,010	0,404 ± 0,009	94 ± 2
Mn	0,100 ± 0,003	0,109 ± 0,005	109 ± 5
Fe	2,97 ± 0,05	2,837 ± 0,013	96 ± 1
Cu	0,00201 ± 0,00010	0,00200 ± 0,00001	99 ± 1
Zn	0,0096 ± 0,0004	0,0113 ± 0,0005	118 ± 5

Todos os valores estão expressos em % e o erro da concentração medida corresponde a um desvio padrão.

Tabela 6 - Comparação entre as concentrações certificadas e medidas do MRC IAEA13, com as respectivas recuperações.

Elemento	Concentração certificada	Concentração medida	Recuperação
Al	10,00 ± 0,14	7,9 ± 0,4	79 ± 4
Si	26,1 ± 0,2	24,5 ± 0,8	94 ± 3
P	< LD	-	-
S	< LD	-	-
K	0,117 ± 0,003	0,114 ± 0,002	97 ± 2
Ca	1,54 ± 0,03	1,871 ± 0,008	122 ± 1
Ti	0,442 ± 0,016	0,446 ± 0,009	101 ± 2
Mn	0,148 ± 0,004	0,157 ± 0,008	106 ± 5
Fe	5,38 ± 0,08	5,25 ± 0,02	98 ± 1
Cu	0,0055 ± 0,0002	0,00400 ± 0,00001	73 ± 1
Zn	0,0066 ± 0,0003	0,0065 ± 0,0005	98 ± 8

Todos os valores estão expressos em % e o erro da concentração medida corresponde a um desvio padrão.

Tabela 7 - Comparação entre as concentrações certificadas e medidas do MRC IPT32, com as respectivas recuperações.

Elemento	Concentração certificada	Concentração medida	Recuperação
Al	15,08 ± 0,11	15,3 ± 0,7	101 ± 4
Si	24,21 ± 0,14	22,5 ± 0,6	93 ± 2
P	0,057 ± 0,009	0,031 ± 0,003	55 ± 5
K	0,66 ± 0,03	0,570 ± 0,011	86 ± 2
Ca	0,121 ± 0,014	0,127 ± 0,005	104 ± 4
Ti	0,893 ± 0,012	0,912 ± 0,011	102 ± 1
Fe	2,42 ± 0,05	2,49 ± 0,02	103 ± 1

Todos os valores estão expressos em % e o erro da concentração medida corresponde a um desvio padrão.

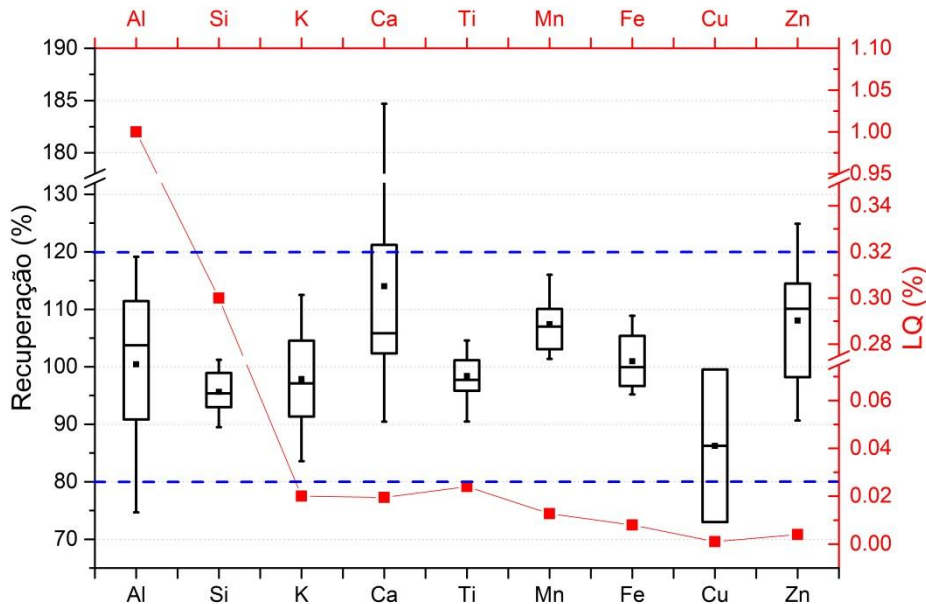
Tabela 8 - Comparação entre as concentrações certificadas e medidas do MRC IPT42, com as respectivas recuperações.

Elemento	Concentração certificada	Concentração medida	Recuperação
Al	17,04 ± 0,11	18,3 ± 0,8	107 ± 5
Si	24,26 ± 0,05	23,2 ± 0,8	96 ± 3
P	< LQ	-	-
K	0,39 ± 0,04	0,434 ± 0,003	111 ± 1
Ca	0,036 ± 0,007	0,047 ± 0,014	130 ± 39
Ti	0,58 ± 0,02	0,555 ± 0,005	96 ± 1
Fe	0,76 ± 0,03	0,823 ± 0,005	108 ± 1

Todos os valores estão expressos em % e o erro da concentração medida corresponde a um desvio padrão.

A Figura 15 mostra as recuperações de todas as medidas realizadas nos quatro MRC utilizados para validação (n=24 para o Al, Si, K, Ca, Ti e Fe e n=12 para o Mn, Cu e Zn) com os limites de quantificação de cada elemento. Observa-se que todos os elementos tiveram recuperações próximas a 100%, sendo que somente o Ca e Cu apresentaram recuperações médias fora do intervalo de 90 a 110%.

Figura 15 - Box-plot com as recuperações de todas as medidas dos quatro MRC utilizados para validação (preto) com os limites de quantificação de cada elemento (vermelho).



Fonte: do autor.

Utilizando os coeficientes elementares da Tabela 4 foi possível determinar as concentrações dos elementos de interesse nas amostras de solo. A Tabela 9 apresenta os resultados da estatística descritiva das concentrações de Al, Si, K, Ca, Ti, Mn, Fe, Cu e Zn nas amostras de solo em estudo.

Observa-se (Tabela 9) que somente as concentrações de K, Ca e Mn tiveram variabilidade média. As concentrações de todos os demais elementos tiveram baixa variabilidade, demonstrando a homogeneidade destas variáveis na área de estudo (WARRICK; NIELSEN, 1980).

Tabela 9 - Resultados da estatística descritiva das concentrações elementares totais determinadas por EDXRF com seu respectivo desvio padrão (DP).

Elemento	Média ± DP	CV	Mediana	Mínimo	Máximo
Al	8,0 ± 0,8	9	8,1	6,0 ± 0,6	10,0 ± 0,9
Si	9,1 ± 0,7	8	9,1	7,2 ± 0,3	11,3 ± 0,3
K	0,052 ± 0,017	32	0,050	0,027 ± 0,002	0,120 ± 0,004
Ca	0,28 ± 0,10	36	0,27	0,105 ± 0,004	0,83 ± 0,06
Ti	2,5 ± 0,3	12	2,5	1,66 ± 0,05	3,8 ± 0,3
Mn	0,200 ± 0,038	19	0,195	0,120 ± 0,002	0,335 ± 0,006
Fe	15,3 ± 0,7	5	15,4	12,5 ± 0,2	17,0 ± 0,3
Cu	0,007 ± 0,001	8	0,007	0,005 ± 0,001	0,008 ± 0,001
Zn	0,008 ± 0,001	9	0,008	0,007 ± 0,001	0,011 ± 0,001

Todos os valores estão em %.

6.3 ANÁLISE EXPLORATÓRIA

6.3.1 Análise de Correlação

Foi aplicado o teste de correlação de Pearson com o objetivo de verificar correlações entre os atributos de fertilidade do solo determinados pelos métodos convencionais e concentrações elementares determinadas por EDXRF. Esses resultados são mostrados na Figura 16.

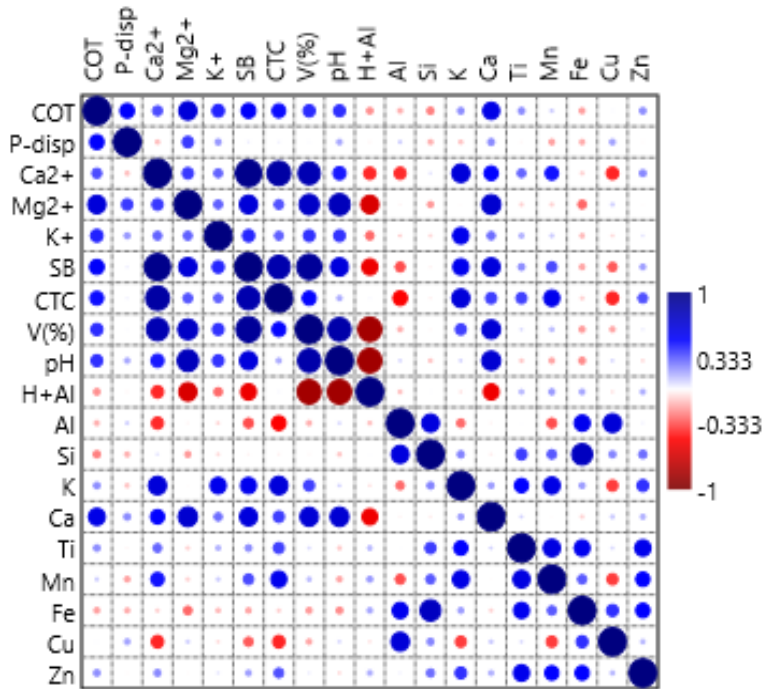
Em geral, observam-se correlações significativas entre todos os atributos de fertilidade do solo, exceto para o P_{disp} , que demonstrou correlação positiva somente com COT e Mg^{2+} . A acidez potencial ($H^+ + Al^{3+}$) foi o único parâmetro de fertilidade que apresentou correlação inversa com todos os demais, sendo essa correlação inversa mais significativa com o pH ($r = -0,86$), $V\%$ ($r = -0,83$) e Mg^{2+} ($r = -0,64$).

O COT e o Mg^{2+} apresentaram correlações com todos os atributos de fertilidade analisados. Além disso, esses parâmetros tiveram uma alta correlação com o Ca total determinado por EDXRF ($r = 0,60$ para o COT e $r = 0,69$ para o Mg^{2+}). Observa-se também uma pequena correlação inversa entre os teores de Fe e Si com o COT e Mg^{2+} .

O Ca^{2+} apresentou alta correlação com SB, CTC e $V(\%)$. Isso faz sentido, visto que o Ca^{2+} é a principal base trocável do solo em questão. Além disso, o Ca^{2+} apresentou correlações positivas entre e os teores totais de K ($r = 0,65$), Ca ($r = 0,50$) e Mn ($r = 0,47$) e correlações inversas com Cu ($r = -0,43$) e Al ($r = -0,42$). Um

fato que chama a atenção é o cálcio trocável (Ca^{2+}) ter uma correlação maior com o K total do que com o Ca total.

Figura 16 - Correlação de Pearson entre os atributos de fertilidade do solo determinados pelos métodos convencionais e as concentrações elementares determinadas por EDXRF.



Fonte: do autor.

O K^+ teve correlação positiva com todos os parâmetros de fertilidade analisados com exceção da acidez potencial. No entanto, essas correlações foram pequenas ($0,20 < r < 0,43$). A maior correlação do K^+ foi com o K total determinado por EDXRF, com r de 0,57.

A SB, CTC e V% apresentaram correlações semelhantes, tendo altas correlações com Ca^{2+} , Mg^{2+} , K^+ e COT. A SB e V% também apresentaram alta correlação com pH. Com relação aos teores inorgânicos, a SB apresentou correlação positiva com Ca ($r = 0,64$) e K ($r = 0,57$) seguida pelo Mn ($r = 0,36$) e correlação negativa com Al ($r = -0,35$) e Cu ($r = -0,32$). A CTC apresentou correlação positiva com K ($r = 0,64$) e Mn ($r = 0,57$), seguida pelo Ca ($r = 0,38$), Ti ($r = 0,37$) e Zn ($r = 0,34$) e correlação negativa com Al ($r = -0,51$) e Cu ($r = -0,43$). Por fim, V% apresentou correlação somente com o Ca ($r = 0,66$) e K ($r = 0,38$).

O pH teve correlação positiva com COT, Ca^{2+} , Mg^{2+} , K^+ , SB e V%, sendo esta maior com o V% ($r = 0,83$) e Mg^{2+} ($r = 0,76$). Também se observou uma alta correlação negativa entre o pH e acidez potencial, como descrito anteriormente. Entre os teores inorgânico, o pH teve correlação somente com o Ca ($r = 0,67$). Notavelmente, a ocorrência de Ca em solos brasileiros decorre principalmente da aplicação de calcário, sendo uma prática comum de manejo para aumentar o pH do solo. Como o Mg^{2+} também está presente no calcário, existe uma boa correlação entre Ca e Mg^{2+} com o pH do solo (LOPES; GUILHERME, 2016).

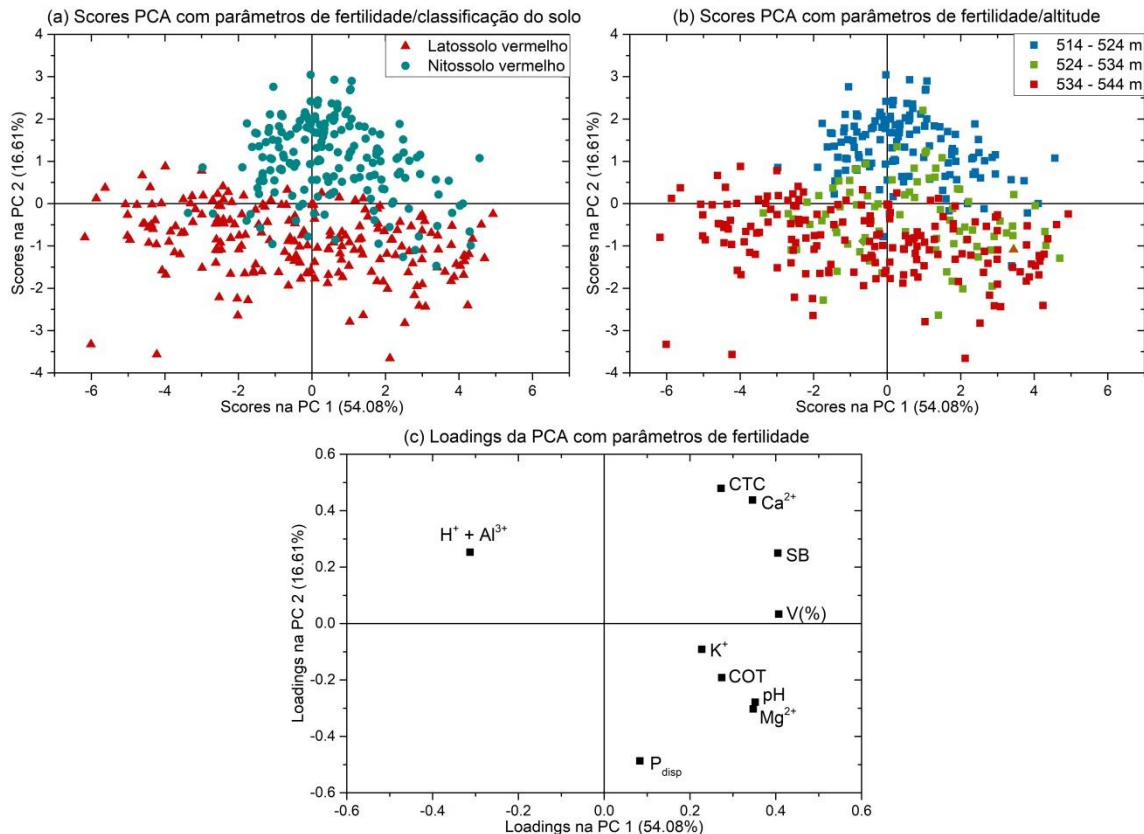
6.3.2 Análise de Componentes Principais (PCA)

A Figura 17 mostra o gráfico de *scores* e *loadings* da PCA construída com os dados dos atributos de fertilidade do solo ($\mathbf{X}_{396 \times 10}$) determinados pelos métodos convencionais. A PC1 explicou 54,1% da variância dos dados e a PC2 explicou 16,6%. Foi possível observar uma tendência de separação das amostras de acordo com a classificação do solo (Figura 17 (a)) e altitude (Figura 17 (b)), visto que as amostras classificadas como Latossolo Vermelho estão localizadas na região mais alta da área de estudo. Nesse caso, as amostras classificadas como Latossolo Vermelho agruparam-se no eixo negativo da PC2 e as amostras classificadas como Nitossolo Vermelho ficaram mais concentradas no eixo positivo da PC2. A transição de uma classe de solo para outra ocorre na região de altitude intermediária (entre 524 e 534 m) (Figura 11). De acordo com os *scores* da Figura 17 (b), as amostras localizadas entre as altitudes de 524 e 534 m estão distribuídas em ambos os eixos da PC2, sendo mais concentradas próximo da origem (zero). Portanto, isso demonstra que as amostras de Nitossolo Vermelho que ficaram no eixo negativo da PC2 (Figura 17 (a)) são da região de transição entre as diferentes classes de solo da área de estudo.

Por meio dos *loadings* da Figura 17 (c), observa-se que o parâmetro que mais contribuiu para essa separação foi o P_{disp} , o qual tem os menores teores na região mais baixa do terreno, onde estão as amostras de Nitossolo Vermelho. Nessa região estão localizadas as maiores declividades e também os maiores valores para fluxo acumulado de água (DOS SANTOS, 2019). De acordo com dos Santos (2019), as perdas de P_{disp} podem estar relacionadas ao fluxo de água, sendo maiores nas áreas com declividade mais acentuada, o que causa menor valor de

P_{disp} na região de Nitossolo Vermelho. Por outro lado, os teores de Ca trocável (Ca^{2+}) e CTC são maiores na região com Nitossolo Vermelho (região de menor altitude).

Figura 17 - Gráfico de *scores* (a e b) e *loadings* (c) da PCA com os atributos de fertilidade do solo determinados pelos métodos convencionais.



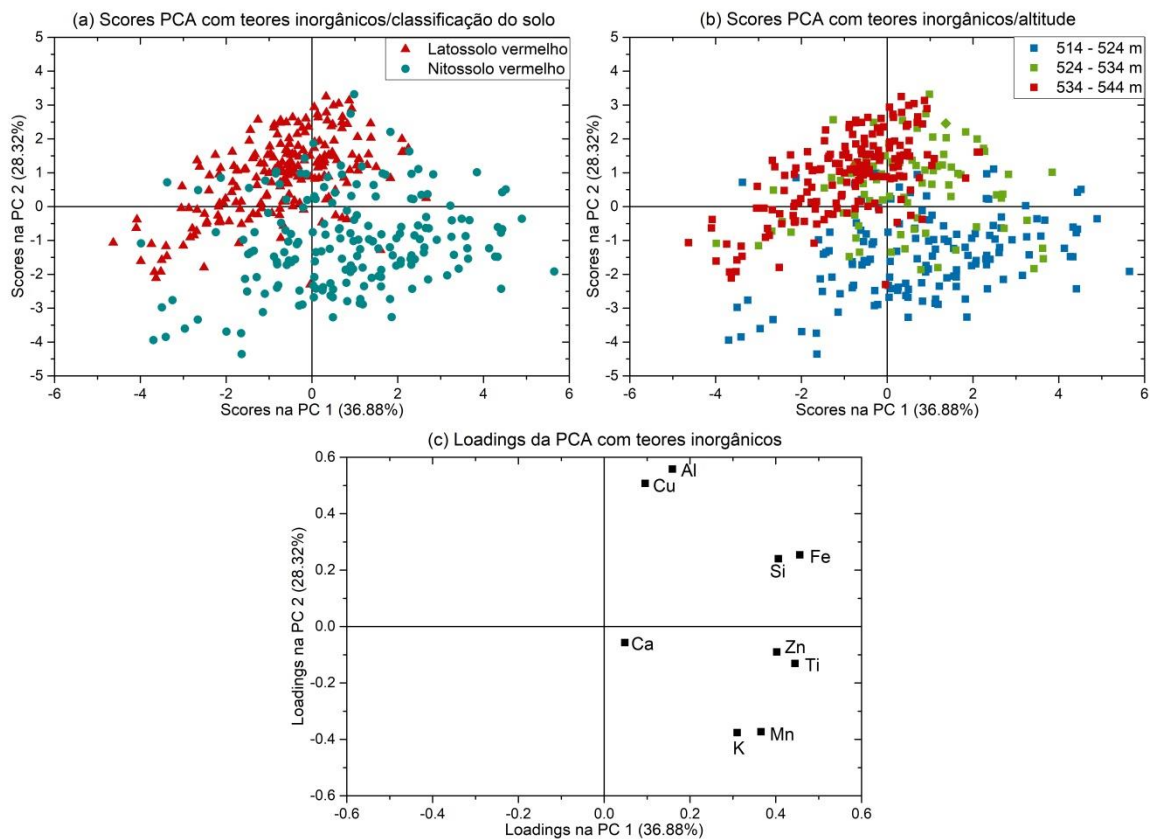
Fonte: do autor.

A Figura 18 mostra os gráficos de *scores* e *loadings* da PCA realizada com os dados de concentração elementar determinados por EDXRF (X_{396X9}). A PC1 explicou 36,9% da variância dos dados e a PC2 explicou 28,3%. Assim como na PCA com os atributos de fertilidade do solo determinados pelos métodos convencionais (Figura 17), foi observado uma tendência de separação das amostras de acordo com a classificação do solo e altitude da área de estudo. Pelo gráfico de *scores* (Figura 18 (a)), observa-se que as amostras classificadas como Latossolo Vermelho estão concentradas principalmente na região negativa da PC1 e positiva da PC2. Essa é a mesma região onde estão concentradas as amostras de maior altitude (Figura 18 (b)).

Pelo gráfico de *loadings* (Figura 18 (c)), nota-se que a PC1 contribui

para separação das amostras de acordo com os teores dos elementos determinados neste estudo, de forma que as amostras localizadas no eixo negativo da PC1 têm as menores concentrações destes elementos. Por outro lado, observa-se que as amostras localizadas no eixo positivo da PC2 têm as maiores concentrações de Cu e Al e as menores concentrações de K e Mn. Portanto, estes resultados indicam que as concentrações de K, Mn, Ti e Zn foram ligeiramente maiores nas amostras classificadas previamente como Nitossolo Vermelho.

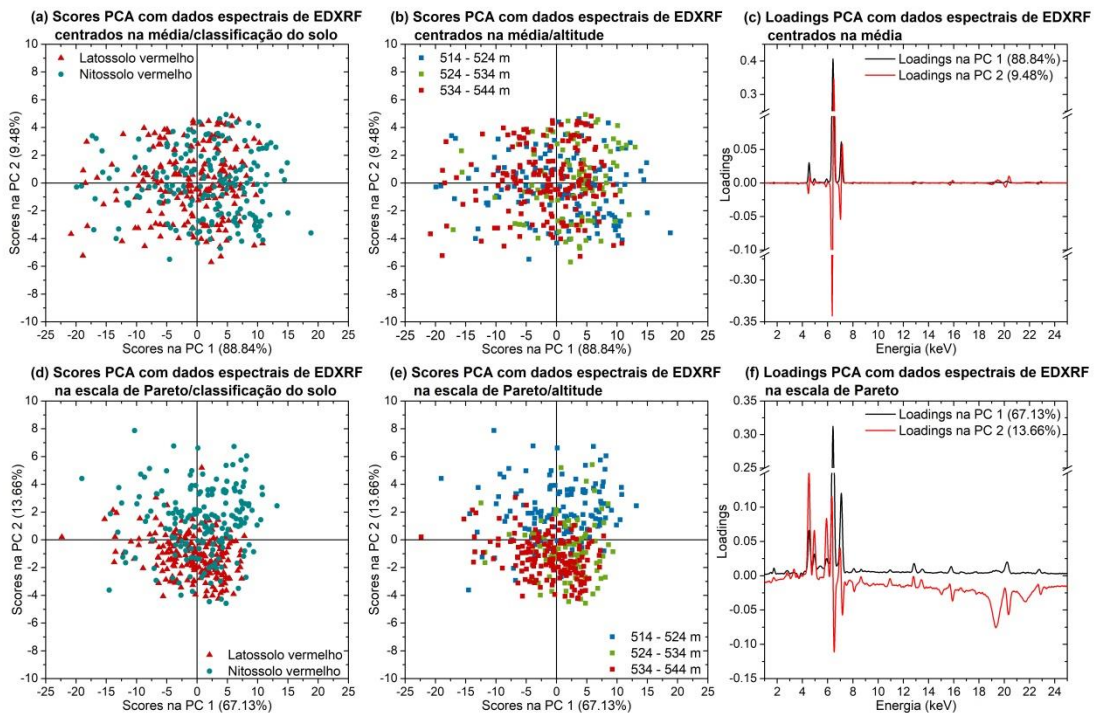
Figura 18 - Gráfico de *scores* e *loadings* da PCA com as concentrações elementares determinadas por EDXRF.



Fonte: do autor.

Também foi construído um modelo de PCA com dados espectrais de EDXRF, conforme Figura 19. É importante ressaltar que as PCAs foram contruídas com os dados da faixa de Na-Sc e Ti-U. No entanto, como os resultados foram semelhantes, optou-se por apresentar somente os resultados da PCA com os dados espectrais de EDXRF na faixa de Ti-U ($X_{396X1270}$) (Figura 19).

Figura 19 - Gráfico de *scores* e *loadings* da PCA com dados espectrais de EDXRF na faixa de Ti-U centrados na média e na escala de Pareto.



Fonte: do autor.

Neste caso, as PCAs foram construídas com dois diferentes pré-processamentos, um com os dados espectrais centrados na média e outro com os dados na escala de Pareto. De acordo com a Figura 19, observa-se que o pré-processamento foi fundamental para separação das amostras. Com os dados centrados na média não foi possível distinguir as amostras de acordo com nenhuma classe. No entanto, quando se aplicou o escalamento de Pareto, foi observada uma tendência de separação das amostras da mesma forma como nas PCAs das Figuras 17 e 18 (com os dados dos parâmetros de fertilidade do solo determinados pelos métodos convencionais e com os dados de concentrações elementares determinados por EDXRF, respectivamente). Nesse caso, as amostras classificadas como Latossolo Vermelho (que estão localizadas na região de maior altitude) ficaram agrupadas no eixo negativo da PC2, enquanto as amostras classificadas como Nitossolo Vermelho (que estão localizadas na região de menor altitude) estão na sua maioria no eixo positivo da PC2. De acordo com o gráfico de *loadings* (Figura 19 (f)), as regiões com os picos de Ti e Mn foram as que mais contribuíram para o agrupamento das amostras no eixo positivo da PC2 e a região do espectro com os

picos de espalhamento (entre 18 e 24 keV) foi a que mais contribuiu para o agrupamento das amostras no eixo negativo da PC2. Esses resultados demonstram que a aplicação do pré-processamento adequado é essencial para o uso de dados espectrais em análise multivariada.

É importante ressaltar que foram testadas outras classes nas PCAs, como profundidade de coleta da amostra (0-5, 5-10 e 10-20 cm) e inclinação do terreno. No entanto, não foi observado nenhum agrupamento com relação a essas outras classes.

6.4 CALIBRAÇÃO MULTIVARIADA

Antes de construir os modelos de calibração multivariada, o conjunto de dados foi dividido em 2/3 para calibração (264 amostras) e 1/3 para validação (132 amostras). A Tabela 10 apresenta a estatística descritiva dos atributos de fertilidade do solo determinados pela análise convencional para o conjunto de calibração e para o conjunto de previsão. É possível verificar que o algoritmo Kennard-Stone separou, de forma representativa, as amostras que formam os conjuntos de calibração e previsão, preservando os valores médio, máximo e mínimo.

O coeficiente de variação (CV) é um fator que tem influência nos modelos de calibração. Altos RMSE podem ser induzidos por altos CVs (STENBERG et al., 2010). Nesse sentido, observou-se que os CVs do conjunto de calibração e previsão foram semelhantes (Tabela 10), enfatizando a adequação do algoritmo de Kennard-Stone para separar os conjuntos de calibração e previsão externa. Além disso, observa-se que o intervalo de valores de cada parâmetro, para o conjunto de calibração e previsão estão próximos.

Valores de curtose e assimetria próximos a zero indicam que os dados possuem distribuição normal (RYU, 2011). Em geral, se observa que os valores de curtose e assimetria do conjunto de calibração e previsão são semelhantes, onde os maiores valores foram encontrados para o P_{disp} , K^+ e $V\%$, indicando não normalidade nesses dados. A modelagem de dados de parâmetros do solo que não seguem uma distribuição normal é bastante comum (TAVARES et al., 2020a). No entanto, isso pode causar efeitos com potencial de redução da qualidade preditiva, conforme relatado por Mouazen; de Baerdemaeker; ramon (2006).

Tabela 10 - Estatística descritiva dos atributos do solo determinados pelos métodos convencionais.

Atributo	Média ± DP	CV (%)	Med.	Mín.	Máx.	Curt.	Ass.
Amostras de calibração (n=264)							
COT	20,0 ± 4,4	22	20,5	7,6	30,5	-0,35	-0,43
P _{disp}	14,4 ± 10,6	73	13,0	0,4	69,8	5,49	1,76
Ca ²⁺	6,0 ± 1,5	25	6,3	1,4	9,2	-0,59	-0,23
Mg ²⁺	2,0 ± 0,6	29	2,0	0,7	3,7	0,61	-0,07
K ⁺	0,25 ± 0,17	68	0,20	0,05	1,15	3,36	1,53
SB	8,4 ± 1,9	23	8,7	2,1	12,6	-0,68	0,15
CTC	13,8 ± 1,6	11	13,9	9,3	17,9	-0,26	-0,35
V	60 ± 10	16	62	18	80	-1,10	2,46
pH	5,3 ± 0,4	8	5,2	4,2	6,5	0,28	0,21
H ⁺ + Al ³⁺	5,4 ± 1,1	19	5,3	2,7	9,7	0,36	1,39
Amostras de previsão externa (n=132)							
COT	19,4 ± 4,2	22	19,4	9,5	29,2	-0,10	-0,55
P _{disp}	12,8 ± 8,4	66	11,1	0,6	49,6	3,02	1,38
Ca ²⁺	5,8 ± 1,5	25	5,8	2,8	8,7	-0,16	-0,85
Mg ²⁺	2,0 ± 0,6	29	2,0	1,1	3,7	0,45	-0,49
K ⁺	0,21 ± 0,13	61	0,18	0,05	0,65	1,34	1,44
SB	8,0 ± 1,8	23	8,0	4,1	12,0	-0,25	-0,57
CTC	13,5 ± 1,5	11	13,5	9,4	17,2	-0,13	-0,16
V	59 ± 10	16	60	31	76	-0,57	0,10
pH	5,3 ± 0,4	8	5,2	4,4	6,5	0,40	0,09
H ⁺ + Al ³⁺	5,5 ± 1,1	20	5,3	3,1	9,0	0,22	0,17

DP = desvio padrão; CV = coeficiente de variação; Med. = mediana; Mín. = mínimo; Máx = máximo; Curt. = curtose; Ass. = assimetria; Valores do COT em g kg⁻¹, do P_{disp} em mg kg⁻¹, do Ca²⁺, Mg²⁺, K⁺, SB, CTC e H⁺ + Al³⁺ em cmol_c kg⁻¹ e V em %.

6.4.1 Modelos para COT

6.4.1.1 Avaliação do pré-processamento espectral para os modelos de COT

Foram desenvolvidos vários modelos PLSR para o COT utilizando diferentes técnicas de pré-processamento nos dados espectrais de EDXRF. O desempenho dos modelos foi avaliado com base nos menores valores de RMSECV e RMSEP. A Tabela 11 mostra um resumo com os resultados dos modelos para todos os pré-processamentos testados. Esses resultados (Tabela 11) fornecem o número de variáveis latentes (LV) dos modelos, RMSECV, RMSEP e a melhora relativa (I (%)) de cada pré-processamento. É importante ressaltar que nenhuma

amostra do conjunto de calibração e previsão foi identificada como outlier nesses modelos.

Para ambas as faixas espectrais (Na-Sc e Ti-U), o pré-processamento que apresentou o menor RMSEP foi o escalamento de Pareto. Observa-se (Tabela 11) que os modelos com escalamento de Pareto apresentaram uma melhora relativa de 18% no espectro de Na-Sc e de 12% no espectro de Ti-U. De forma geral, o espectro de Na-Sc foi mais suscetível aos pré-processamentos testados.

Somente em alguns casos, o pré-processamento piorou o desempenho dos modelos, como a aplicação de MSC e SNV no espectro de Na-Sc e o uso da primeira derivada combinada com centragem na média no espectro Ti-U.

Tabela 11 - Resultados dos modelos para COT com diferentes pré-processamentos aplicados aos dados espectrais de EDXRF.

Pré-processamento	Espectro de Na-Sc				Espectro de Ti-U			
	LV	RMSECV	RMSEP	I(%)	LV	RMSECV	RMSEP	I(%)
Nenhum	4	3,2	2,5	0%	6	3,2	2,5	0%
MC	5	3,2	2,4	1%	6	3,2	2,4	4%
Auto	6	2,7	2,1	13%	3	3,1	2,4	7%
Pareto	6	2,8	2,0	18%	7	3,0	2,2	12%
SG	4	3,2	2,5	0%	6	3,2	2,4	6%
1º der	4	3,4	2,4	1%	5	3,2	2,5	2%
2º der	4	3,2	2,5	0%	5	3,2	2,5	1%
MSC	4	3,1	2,5	-1%	5	3,2	2,4	4%
SNV	4	3,1	2,5	-1%	5	3,2	2,4	4%
SG+MC	6	3,0	2,1	15%	8	3,0	2,3	9%
1ºder+MC	4	3,2	2,4	1%	5	3,2	2,5	2%
2ºder+MC	4	3,2	2,4	1%	5	3,2	2,6	-1%
MSC+MC	3	3,1	2,5	-1%	4	3,2	2,4	4%
SNV+MC	3	3,1	2,5	-1%	4	3,2	2,4	4%
SG+Auto	5	2,7	2,3	7%	4	3,1	2,3	8%
1ºder+Auto	2	3,0	2,3	8%	2	3,2	2,5	2%
2ºder+Auto	2	3,1	2,3	6%	2	3,2	2,5	1%
MSC+Auto	6	2,7	2,1	16%	3	3,1	2,4	6%
SNV+Auto	6	2,7	2,1	13%	4	3,0	2,4	7%
SG+Pareto	5	3,0	2,0	17%	7	3,0	2,2	11%
1ºder+Pareto	6	3,1	2,3	6%	4	3,1	2,4	4%
2ºder+Pareto	4	3,1	2,5	0%	4	3,2	2,5	2%
MSC+Pareto	5	2,9	2,1	14%	6	3,0	2,3	9%
SNV+Pareto	5	3,0	2,1	14%	6	3,0	2,3	9%

Variáveis latentes (LV), RMSECV (g kg^{-1}), RMSEP (g kg^{-1}) e melhora relativa (I (%)) em relação aos dados espectrais sem nenhum pré-processamento.

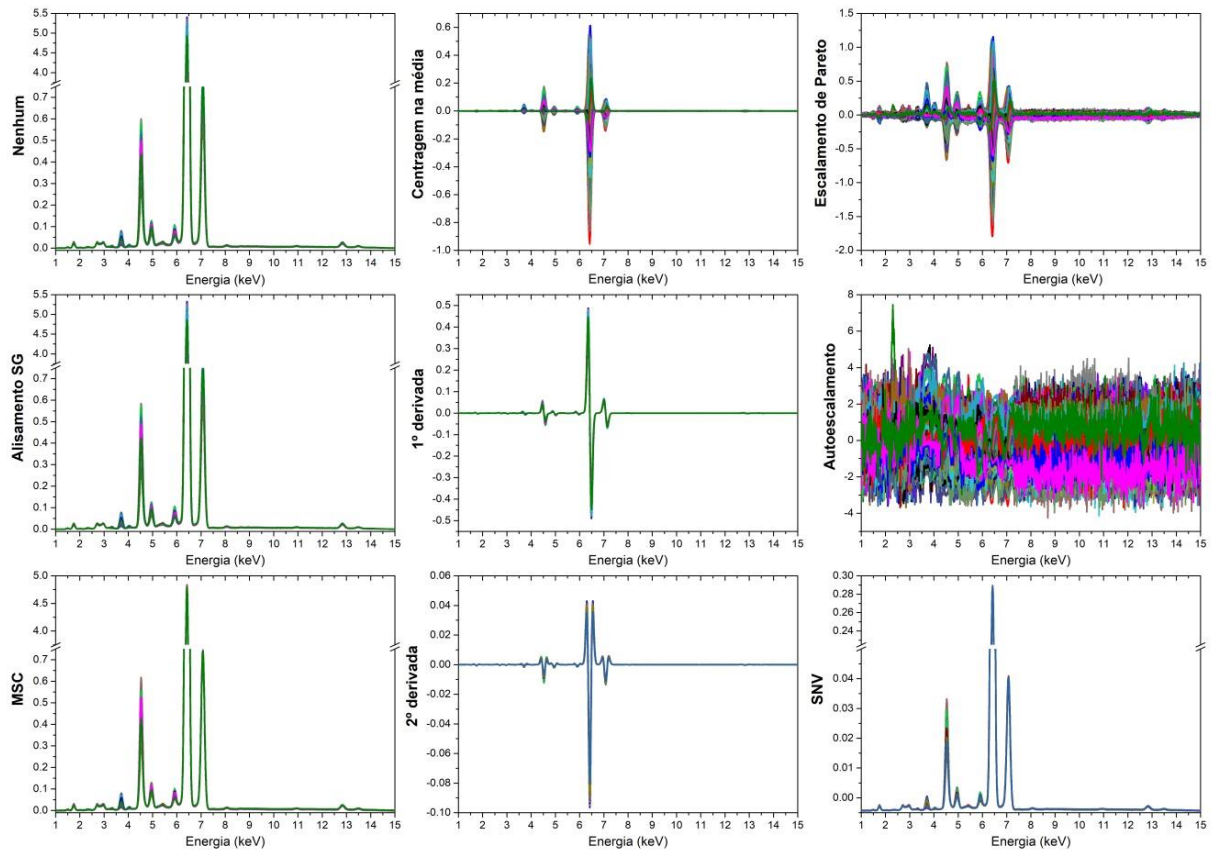
Não há uma abordagem bem estabelecida de pré-processamento para análise multivariada com dados espectrais de EDXRF. Normalmente, é aplicado apenas a centragem na média (AIDENE et al., 2020; BORTOLETO; BORGES; BUENO, 2007; DA-COL; BUENO; MELQUIADES, 2015; DOS SANTOS et al., 2020; NOGUEIRA et al., 2019; MELQUIADES et al., 2012; BUENO et al., 2005). Por esse motivo, os resultados dos modelos com escalamento de Pareto (o pré-processamento que apresentou o melhor desempenho) foram comparados com a centragem na média. Para isso, aplicou-se o teste de randomização, o qual mostrou que houve diferença significativa (p -valor = 0,01) entre a acurácia dos modelos construídos com escalamento de Pareto e os modelos com centragem na média sobre o espectro de Na-Sc. Os valores de RMSEP para os modelos com o espectro Na-Sc centrados na média e na escala de Pareto foram $2,4 \text{ g kg}^{-1}$ com cinco LVs e $2,0 \text{ g kg}^{-1}$ com seis LVs, respectivamente. Para o espectro Ti-U, não foi observado diferença significativa (p -valor = 0,13) entre a acurácia dos modelos usando a centragem na média e o escalamento de Pareto. Para esse espectro, os valores de RMSEP foram $2,4 \text{ g kg}^{-1}$ com seis LVs e $2,2 \text{ g kg}^{-1}$ com sete LVs, utilizando centragem na média e escalamento de Pareto, respectivamente.

As Figuras 20 e 21 mostram os dados espectrais de EDXRF com diferentes pré-processamentos para as faixas de Na-Sc e Ti-U, respectivamente. Observa-se que a escala de Pareto reduziu a importância relativa dos picos mais intensos como o de Fe, evidenciando picos menos intensos como Si, K, Ca, Mn e os picos de espalhamento do Rh. Além disso, a estrutura dos dados permaneceu parcialmente intacta e dados sem importância como ruídos não se tornaram tão evidentes como acontece no autoescalamento. Isso pode explicar porque o desempenho dos modelos com escalamento de Pareto foi melhor do que com os demais pré-processamentos.

Além de gerar o melhor desempenho, o uso do escalamento de Pareto é adequado para esse conjunto de dados, pois os picos de Fe são muito intensos em comparação aos outros elementos. Além disso, o espectro de EDXRF é afetado pelo rendimento fluorescente, ou seja, a probabilidade de emissão de um raio X característico é diferente para cada elemento. Assim, o escalamento de Pareto visa ponderar essas diferentes probabilidades de emissão sem ser altamente influenciado por variáveis sem importância, como ruídos. Dessa forma, todos os modelos para os demais atributos de fertilidade do solo foram construídos utilizando

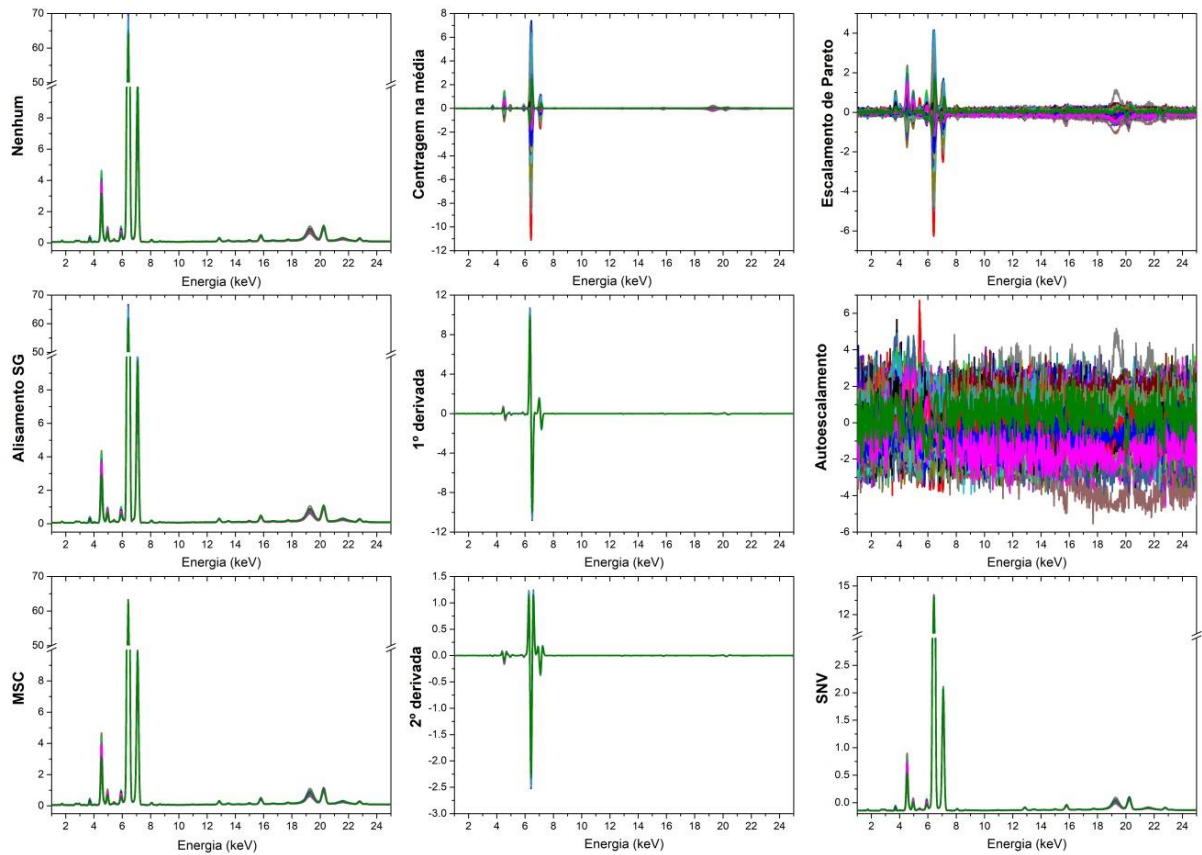
escalamento de Pareto e centragem na média, o qual é comumente utilizado em dados espectrais de EDXRF.

Figura 20 - Métodos de pré-processamento aplicados sobre os dados espectrais de Na-Sc.



Fonte: do autor.

Figura 21 - Métodos de pré-processamento aplicados sobre os dados espectrais de Ti-U.



Fonte: do autor.

6.4.1.2 Seleção de variáveis dos modelos de COT

Diferentes métodos de seleção de variáveis baseados em intervalos (iPLS, BiPLS, FiPLS, SiPLS e iSPA-PLS) foram testados para verificar as regiões espectrais que mais contribuíram para os modelos com os dados espectrais de EDXRF centrados na média e com escala de Pareto. Os resultados para a calibração e previsão dos métodos de seleção de variáveis nas faixas espectrais de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 12.

Tabela 12 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de COT com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	276	552	368	736
	LV	6	3	5	6	5	6
	R²_{cal}	0,77	0,71	0,74	0,85	0,71	0,72
	RMSEC	2,1	2,4	2,2	1,7	1,9	2,3
	R²_{cv}	0,59	0,66	0,67	0,69	0,69	0,66
	RMSECV	2,8	2,5	2,5	2,5	2,5	2,6
	R²_{pred}	0,77	0,69	0,74	0,71	0,74	0,79
	RMSEP	2,0	2,4	2,1	2,3	2,1	1,9
	RPD	2,1	1,8	2,0	1,8	2,0	2,2
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	184	460	368	460
	LV	5	3	4	5	4	5
	R²_{cal}	0,50	0,71	0,74	0,77	0,76	0,69
	RMSEC	3,1	2,4	2,2	2,0	2,1	2,4
	R²_{cv}	0,48	0,66	0,66	0,62	0,62	0,64
	RMSECV	3,2	2,6	2,5	2,6	2,7	2,6
	R²_{pred}	0,66	0,64	0,66	0,71	0,71	0,74
	RMSEP	2,4	2,5	2,4	2,3	2,3	2,1
	RPD	1,7	1,7	1,7	1,8	1,8	2,0
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	635	762	508	508
	LV	7	3	6	5	5	5
	R²_{cal}	0,86	0,48	0,71	0,83	0,69	0,71
	RMSEC	1,6	3,2	2,0	1,8	2,4	2,4
	R²_{cv}	0,52	0,44	0,55	0,53	0,53	0,52
	RMSECV	3,0	3,3	3,0	3,0	3,0	3,0
	R²_{pred}	0,72	0,62	0,72	0,76	0,76	0,76
	RMSEP	2,2	2,6	2,2	2,1	2,1	2,1
	RPD	1,9	1,6	1,9	2,0	2,0	2,0
Faixa Ti-U centrado na média	nº Variáveis	1270	127	635	762	508	635
	LV	6	3	6	6	6	6
	R²_{cal}	0,53	0,45	0,61	0,71	0,59	0,61
	RMSEC	3,0	3,2	2,5	2,4	2,8	2,7
	R²_{cv}	0,46	0,42	0,50	0,52	0,48	0,48
	RMSECV	3,2	3,3	3,1	3,0	3,1	3,1
	R²_{pred}	0,67	0,58	0,71	0,71	0,71	0,72
	RMSEP	2,4	2,7	2,3	2,3	2,3	2,3
	RPD	1,7	1,5	1,8	1,8	1,8	1,8

*EC=Espectro completo; Valores de RMSE em g kg⁻¹.

Em geral, nota-se que a seleção de variáveis melhorou a capacidade preditiva dos modelos, exceto o iPLS, o qual piorou o desempenho de todos os modelos. O iPLS é um método que seleciona apenas uma região espectral específica. Portanto, ele não leva em consideração outras regiões que poderiam contribuir para o modelo. Isso demonstra que mais de uma região do espectro de EDXRF contribui para a determinação do COT. No entanto, a região selecionada por iPLS foi repetida em todos os outros métodos de seleção de variáveis, evidenciando a região do espectro de EDXRF que mais contribuiu para a modelagem do COT.

O principal parâmetro para comparação entre os modelos foi o RMSEP, que demonstrou que a seleção de variáveis por iSPA-PLS e BiPLS foi mais eficiente para os espectros Na-Sc e Ti-U, respectivamente. O valor de RMSEP foi reduzido em 5% (de 2,0 para 1,9 g kg⁻¹) utilizando iSPA-PLS no espectro Na-Sc com escala de Pareto e em 13% (de 2,4 para 2,1 g kg⁻¹) no espectro de Na-Sc centrado na média. A seleção de variáveis por BiPLS reduziu os valores de RMSEP em 7% para ambos os pré-processamento no espectro de Ti-U (de 2,2 a 2,1 g kg⁻¹ para escala de Pareto e de 2,4 para 2,3 g kg⁻¹ para centragem na média). Para comparar a acurácia desses modelos foi realizado um teste de randomização em um intervalo de confiança de 95% (Tabela 13).

Os resultados do teste de randomização (Tabela 13) mostram que a seleção de variáveis por iSPA-PLS no espectro de Na-Sc foi significativa em termos de acurácia. Por outro lado, a seleção de variáveis no espectro Ti-U não apresentou melhora significativa. Entretanto, ao comparar a combinação do pré-processamento com a seleção de variáveis, observa-se que os espectros com escalamento de Pareto foram mais acurados do que com a centragem na média. Isso também pode ser atestado pelo valor de RPD, que demonstra que apenas os modelos com escalamento de Pareto tiveram um desempenho considerado bom ($2,5 > \text{RPD} > 2,0$).

Em relação à condição de medida, o teste de randomização (Tabela 13) não mostrou diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U. Portanto, esses resultados sugerem que o escalamento de Pareto foi mais adequado para pré-processar esses dados espectrais.

Tabela 13 - Teste de randomização para comparação dos melhores modelos para o COT em um intervalo de confiança de 95%.

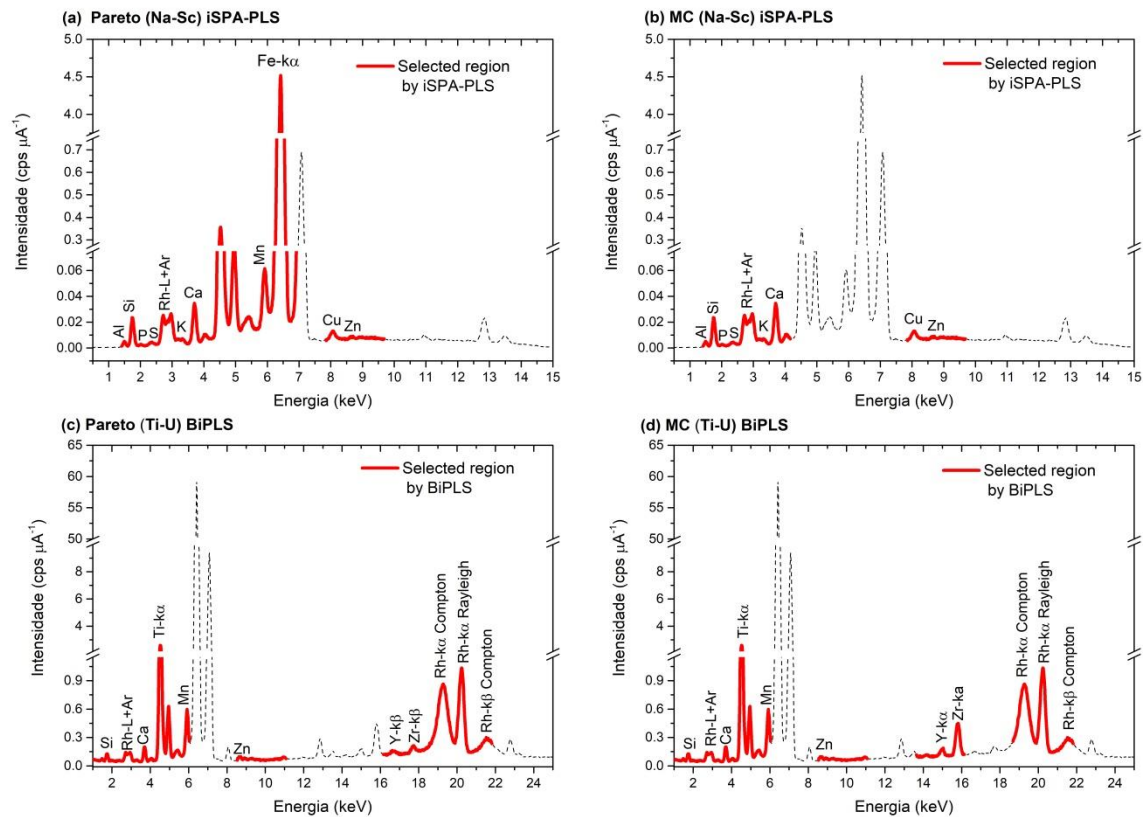
Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X	0,005	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com escalamento de Pareto.
Pareto (Na-Sc) iSPA		
MC (Na-Sc) EC X	0,03	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com centragem na média.
MC (Na-Sc) iSPA		
Pareto (Ti-U) EC X	0,11	Não houve diferença significativa na seleção de variáveis por BiPLS no espectro de Ti-U com escalamento de Pareto.
Pareto (Ti-U) BiPLS		
MC (Ti-U) EC X	0,13	Não houve diferença significativa na seleção de variáveis por BiPLS no espectro de Ti-U com centragem na média.
MC (Ti-U) BiPLS		
Pareto (Na-Sc) iSPA X	0,01	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
MC (Na-Sc) iSPA		
Pareto (Ti-U) BiPLS X	0,02	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
MC (Ti-U) BiPLS		
Pareto (Na-Sc) iSPA X	0,16	Não houve diferença significativa entre os melhores modelos de cada condição de medida
Pareto (Ti-U) BiPLS		

6.4.1.3 Interpretação espectral das variáveis selecionadas nos modelos de COT

As variáveis selecionadas para os melhores modelos de cada pré-processamento são mostradas na Figura 22 e os vetores de regressão para os respectivos modelos são mostrados na Figura 23.

Em geral, percebem-se algumas semelhanças entre as regiões selecionadas (Figura 22) para os diferentes pré-processamento em cada faixa espectral (espectro de Na-Sc e Ti-U). Essas semelhanças são mais evidentes no espectro de Ti-U. A principal diferença nas regiões selecionadas para cada pré-processamento foi observada no espectro de Na-Sc, para o qual o FiPLS e o iSPA-PLS selecionaram a região com o pico de Fe-K α nos dados com escalamento de Pareto, a qual não foi selecionada com os dados centrados na média. Provavelmente, isso foi possível devido à escala de Pareto reduzir a intensidade relativa dos picos de Fe com relação aos demais.

Figura 22 - Intervalos de variáveis selecionadas para os melhores modelos de COT com escalamento de Pareto e centragem na média.

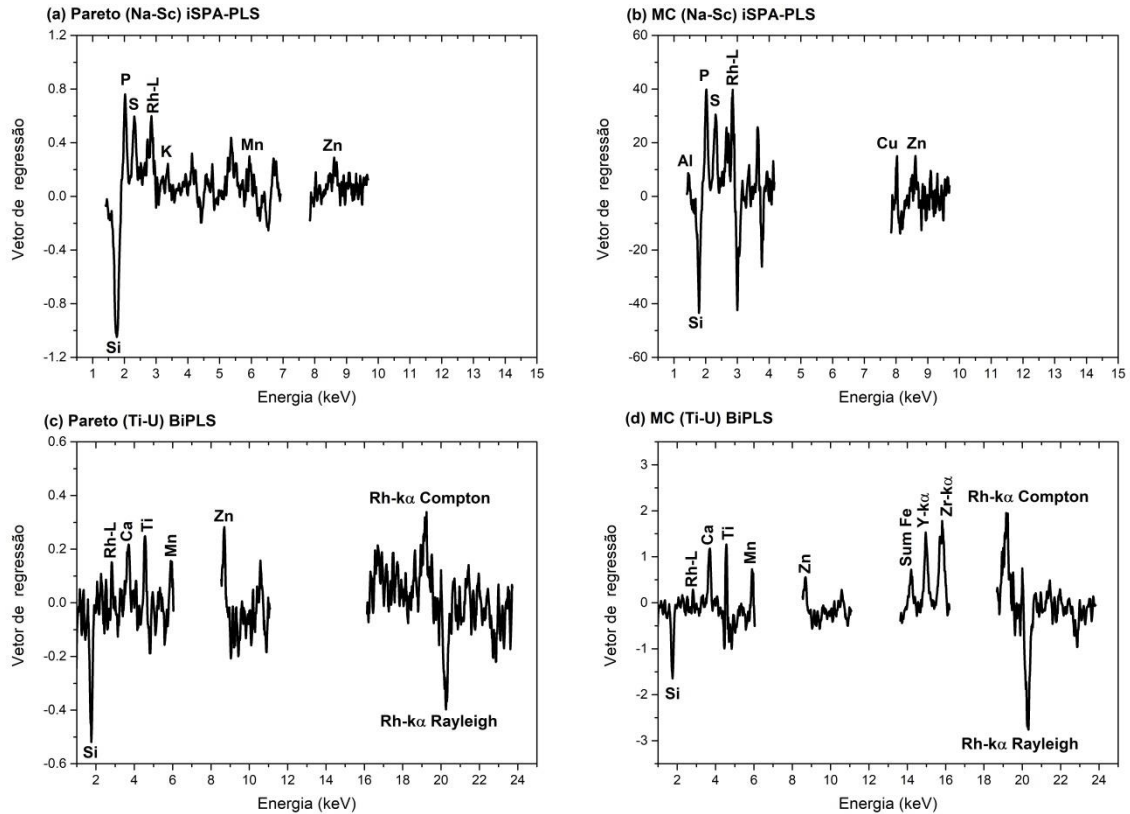


Fonte: do autor.

Para todos os modelos do espectro na faixa de Na-Sc (Tabela 12), foi selecionada a região de 1,4 a 2,3 keV, que abrange os picos de Al, Si, P e S. Para o espectro na faixa de Ti-U, a região de 1,0 a 6,0 keV foi selecionada por todos os métodos de seleção de variáveis. A exceção foi para o método iPLS, que selecionou a região entre 3,5 e 6,0 keV. A região de 1,0 a 6,0 keV contém os picos de Al, Si, P, S, K, Ca, Ti, Mn e Rh-L.

Avaliando os vetores de regressão (Figura 23), observa-se que o Si teve uma contribuição negativa para ambos os modelos de cada faixa espectral. Os picos P, S e Rh-L contribuíram positivamente para os modelos. Portanto, torna-se claro que essa região inicial do espectro, que contém os picos de elemento com $Z < 18$, foi a que mais contribuiu para a modelagem do COT. Isso poderia explicar porque os modelos com o espectro de Na-Sc tiveram um desempenho ligeiramente melhor do que com o espectro de Ti-U, já que a condição de medida de Na-Sc é mais indicada para excitar os elementos com $Z < 20$ (TAVARES et al., 2020a; VAN GRIEKEN; MARKOWICZ, 2002).

Figura 23 - Vetor de regressão dos melhores modelos PLSR para o COT com escalamento de Pareto e centragem na média.



Fonte: do autor.

Em geral, o vetor de regressão para cada pré-processamento teve um comportamento semelhante. A principal diferença foi no peso relativo do Si, que foi maior nos modelos com escalamento de Pareto. Isso reforça que a escala de Pareto equalizou a região inicial do espectro com os demais picos, principalmente com o pico de Fe-K α que é aproximadamente 200 vezes maior que os picos na região de $Z = 13$ a $Z = 20$. Além disso, o rendimento fluorescente para esse intervalo de número atômico é baixo. Portanto, esses resultados demonstram que o uso do escalamento de Pareto foi adequado para esse conjunto de dados, pois dimensionou as diferentes regiões do espectro reduzindo a influência do rendimento fluorescente no desempenho dos modelos.

Para os modelos com o espectro de Ti-U, nota-se que a região de espalhamento (Rh-K α Compton e Rh-K α Rayleigh) teve uma grande contribuição nos vetores de regressão, como já demonstrado por Morona et al. (2017). No

presente estudo, o pico Rh-K α Compton teve uma contribuição positiva para os modelos, enquanto o pico de Rh-K α Rayleigh contribuiu negativamente. Além desses picos de espalhamento, há também a contribuição das linhas L do Rh. Todos esses picos de espalhamento do Rh, especialmente o pico Compton-K α , são devidos à interação do feixe de excitação com elementos de baixo número atômico ($Z < 11$), bem como componentes orgânicos das amostras de solo que tem baixa seção de choque para absorção da radiação incidente e alta seção de choque para o espalhamento (BORTOLETO; BORGES; BUENO, 2007; MORONA et al., 2017). Portanto, os picos de espalhamento também carregam informações sobre o COT, uma vez que os fótons de raios X característicos do carbono não aparecem nos espectros de EDXRF devido ao seu baixo rendimento fluorescente.

No geral, os métodos de seleção de variáveis selecionaram regiões espectrais ao longo de todo o espectro, demonstrando a natureza complexa da matriz do solo analisada. Por exemplo, Fe e Al são elementos importantes para prever o COT em solos ácidos. Eles possuem alta área superficial específica para interações iônicas, sendo parte altamente reativa aos minerais do solo (KLEBER et al., 2015). O carbono orgânico (CO) em Latossolos e Nitossolos é protegido principalmente pela formação do complexo OC-Fe e OC-Al (KLEBER et al., 2005, 2015). Além disso, a fertilização para a produção das culturas ao longo do tempo melhora a estabilização do CO pela formação de frações não cristalinas de Fe e aumento da agregação nas camadas superficiais do solo (WANG et al., 2019).

Embora o Si não seja predominante na fração de argila, espera-se que as superfícies neutras de siloxano dos minerais do grupo caulino liguem-se às moléculas orgânicas não polares ou porções moleculares como carbono alquila ou aromático via interações hidrofóbicas e de van der Waals (KLEBER et al., 2015). Por outro lado, as espécies de hidrólise catiônica de Al e Fe (III) prevalecem nos locais de troca de filossilicatos nos solos ácidos e são muito mais efetivos na formação de complexos de esfera interna com ligantes orgânicos fortes. Assim, a interação restrita e fraca com o COT pode explicar o efeito negativo do Si nos modelos, indicando que a principal interação do COT é com a fração de argila do solo.

A CTC dos Latossolos e Nitossolos é predominantemente originada pelo CO (até 60%) (SOARES; ALLEONI, 2008). Assim, elementos como K, Ca e Mn podem ser encontrados predominantemente associados ao CO e, nos modelos,

podem representar a fração mais instável do COT. Além disso, a contribuição positiva das regiões com os picos de P e S indicam que esses elementos podem estar adsorvidos ou complexados com o CO.

Portanto, esses resultados contribuem para uma melhor interpretação das variáveis do espectro de EDXRF que influenciaram na modelagem do COT.

6.4.1.4 Figuras de mérito dos modelos para COT

A Tabela 14 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia dos modelos para COT. O teste DW em um nível de confiança de 95% indicou que todos os modelos apresentaram linearidade. A veracidade dos modelos foi avaliada com base nos valores de RMSEP e RPD (Tabela 12). Além desses parâmetros, o erro relativo (ER (%)) e os erros relativos individuais (Figura 24) do conjunto de previsão também foram utilizados para avaliar a veracidade dos modelos. O erro relativo foi de 10% para os modelos com escalamento de Pareto e 11% para os modelos com os dados centrados na média. Pode ser visto na Figura 24 que somente um pequeno número de amostras apresentou erros relativos individuais acima de $\pm 20\%$. Além disso, de acordo com a Tabela 14, nenhum modelo apresentou *bias*.

A precisão/repetitividade foi avaliada pelo coeficiente de variação (CV) de medidas em triplicata de três diferentes níveis de concentração, baixo ($10,8 \text{ g kg}^{-1}$), médio ($20,6 \text{ g kg}^{-1}$) e alto ($28,8 \text{ g kg}^{-1}$). O alto nível de concentração foi o que apresentou os piores CVs. Em geral, todos os CVs foram inferiores a 10%, exceto para o modelo BiPLS no espectro Ti-U com os dados centrados na média, que apresentou CV de 13% no nível alto de concentração.

A robustez dos modelos com relação à variação de granulometria foi avaliada por meio de um teste de randomização, no qual se comparou os resultados preditivos de um conjunto de amostras (39 amostras cobrindo toda a faixa de concentração) com granulometria de 2 mm e 125 μm . Esse teste demonstrou que somente os modelos com escalamento de Pareto foram robustos com relação à variação de granulometria.

Tabela 14 - Figuras de mérito dos melhores modelos para COT.

	Pareto (Na-Sc) iSPA-PLS	MC (Na-Sc) iSPA-PLS	Pareto (Ti-U) BiPLS	MC (Ti-U) BiPLS
Linearidade (p-valor DW)	0,65	0,76	0,77	0,93
SEN^a	0,14	0,003	0,25	0,06
γ^{-1} (g kg⁻¹)	0,1	0,1	0,2	0,1
 Bias (t)^b	0,21 (1,28)	0,05 (0,20)	0,16 (1,14)	0,24 (1,50)
ER(%)	10	11	10	11
CV(%)	4/5/8	7/8/9	6/2/6	5/6/13
Repetitividade^c				
Robustez (p-valor) ^d	0,32	0,0005	0,47	0,03
LOD (g kg⁻¹)	0,4	0,4	0,8	0,4
LOQ (g kg⁻¹)	1,1	1,1	2,5	1,1

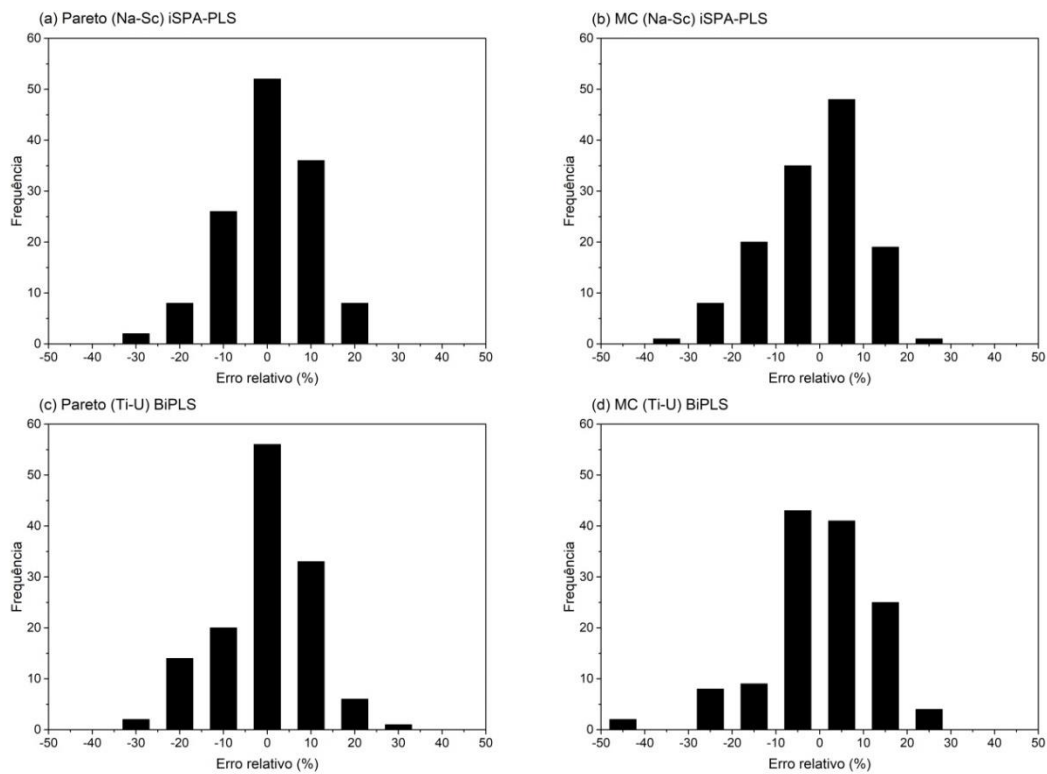
^a Valor expresso como a razão entre intensidade e g kg⁻¹

^b $t_{critico} = 1,98$

^c Valores em nível baixo (10,8 g kg⁻¹), médio (20,6 g kg⁻¹) e alto (28,8 g kg⁻¹).

^d p-valor do teste de randomização em 95% de confiança.

Figura 24 - Erro relativo para o conjunto de previsão dos melhores modelos para COT.



Fonte: do autor.

Também foram avaliadas outras figuras de mérito como sensibilidade pura (SEN), inverso da sensibilidade analítica (γ^{-1}), limite de detecção (LOD) e limite de quantificação (LOQ). O γ^{-1} indicou que a resolução do método foi de 0,2 g kg⁻¹ para o modelo Pareto (Ti-U) BiPLS e 0,1 g kg⁻¹ para os demais modelos. Este aumento em γ^{-1} do modelo Pareto (Ti-U) BiPLS pode ter sido causado pela intensificação do ruído experimental devido ao pré-processamento. No entanto, essa intensificação do ruído experimental não foi significativa para o modelo Pareto (Na-Sc) iSPA-PLS. Os LOD e LOQ mostram valores coerentes, já que outros estudos usando Vis-NIR e EDXRF demonstram LOD variando de 0,3 a 4,4 g kg⁻¹ (BELTRAME et al., 2016; DE SOUZA et al., 2016; MORONA et al., 2017).

De forma geral, esses resultados atestaram que todos os modelos mostraram veracidade e precisão adequadas. Os modelos com escalamento de Pareto (Pareto (Na-Sc) iSPA-PLS e Pareto (Ti-U) BiPLS) foram os mais exatos e precisos, demonstrando robustez com relação à variação de granulometria. Assim, esses modelos mostraram-se adequados para prever o COT neste conjunto de dados, com valores de RPD de 2,2 e 2,0, os quais foram considerados satisfatórios, visto que estudos anteriores relatam valores de RPD variando de 1,1 a 2,9 com a técnica de Vis-NIR (BRICKLEMYER; BROWN, 2010; GOMEZ; VISCARRA ROSSEL; MCBRATNEY, 2008; SITHOLE; NCAMA; MAGWAZA, 2018; STEVENS et al., 2013; VOHLAND et al., 2011). Em geral, os bons resultados de predição usando Vis-NIR para a determinação do COT devem-se principalmente às amplas bandas de absorção na parte visível devido à cor da matéria orgânica e a algumas bandas características na parte do infravermelho próximo causadas pelas ligações C-O e C-H (STENBERG et al., 2010).

Portanto, O presente estudo demonstrou que o uso de dados espectrais de EDXRF produziu resultados equivalentes aos relatados na literatura com Vis-NIR para determinação de COT em solos agrícolas, permitindo uma interpretação do ponto de vista inorgânico, a qual não pode ser feita por vis-NIR. Além disso, com EDXRF, foi possível determinar simultaneamente a concentração de elementos inorgânicos presentes nas amostras.

6.4.2 Modelos para P_{disp}

6.4.2.1 Pré-processamento e seleção de variáveis dos modelos de P_{disp}

Diferentes métodos de seleção de variáveis baseados em intervalos (iPLS, BiPLS, FiPLS, SiPLS e iSPA-PLS) foram testados para verificar as regiões espectrais que mais contribuiriam para modelagem do P_{disp} com os dados espectrais centrados na média e com escala de Pareto. Os resultados para a calibração e previsão dos métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 15. Para ambos os modelos foram excluídas 10 amostras de calibração e 2 de previsão, as quais foram identificadas como *outliers*.

De forma geral, é possível notar que todos os modelos geraram altos RMSEPs e RMSECVs para o P_{disp} , indicando falta de acurácia. Observa-se (Tabela 15) que os modelos utilizando escalamento de Pareto sobre o espectro de Na-Sc apresentaram valor de RMSEP menor do que com a centragem na média, sendo reduzido em 13% (de 6,7 para 5,8 mg kg⁻¹) para os modelos com o espectro completo. Por meio do teste de randomização (p-valor = 0,001), verificou-se que essa diferença foi significativa em um nível de confiança de 95%. Para o espectro de Ti-U não houve diferença significativa entre os modelos com escalamento de Pareto e centragem na média.

Em geral, a seleção de variáveis melhorou a capacidade preditiva dos modelos, principalmente para os dados centrados na média. Assim como para o COT, o desempenho dos modelos piorou somente com o uso do iPLS. No entanto, para a faixa espectral de Na-Sc, a região selecionada por esse algoritmo se repetiu para todos os outros métodos de seleção de variáveis. Por outro lado, esse comportamento não foi observado para a faixa espectral de Ti-U.

Tabela 15 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de P_{disp} com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	368	368	368	368
	LV	7	4	7	6	7	7
	R²_{cal}	0,56	0,53	0,66	0,58	0,59	0,59
	RMSEC	5,1	5,5	4,7	5,2	5,1	5,1
	R²_{cv}	0,42	0,49	0,56	0,50	0,52	0,52
	RMSECV	6,2	5,7	5,4	5,7	5,6	5,6
	R²_{pred}	0,41	0,38	0,44	0,44	0,46	0,46
	RMSEP	5,8	5,9	5,7	5,6	5,4	5,4
	RPD	1,3	1,3	1,3	1,3	1,4	1,4
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	184	1380	368	92
	LV	4	2	3	4	4	2
	R²_{cal}	0,38	0,45	0,61	0,38	0,59	0,45
	RMSEC	6,3	5,9	5,0	6,3	5,1	5,9
	R²_{cv}	0,35	0,42	0,48	0,36	0,45	0,42
	RMSECV	6,5	6,1	5,9	6,5	5,9	6,1
	R²_{pred}	0,20	0,34	0,38	0,20	0,35	0,34
	RMSEP	6,7	6,0	5,9	6,7	6,0	6,0
	RPD	1,1	1,2	1,2	1,1	1,2	1,2
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	762	762	508	381
	LV	4	3	4	4	5	5
	R²_{cal}	0,44	0,36	0,42	0,42	0,67	0,62
	RMSEC	6,1	6,4	6,1	6,1	4,6	4,9
	R²_{cv}	0,37	0,32	0,37	0,37	0,36	0,37
	RMSECV	6,4	6,4	6,4	6,4	6,5	6,5
	R²_{pred}	0,27	0,14	0,44	0,44	0,31	0,32
	RMSEP	6,5	7,0	6,5	6,5	6,3	6,3
	RPD	1,1	1,1	1,1	1,1	1,2	1,2
Faixa Ti-U centrado na média	nº Variáveis	1270	127	508	762	508	635
	LV	5	3	6	5	6	6
	R²_{cal}	0,46	0,37	0,56	0,45	0,58	0,58
	RMSEC	5,9	6,4	6,2	6,0	5,1	5,2
	R²_{cv}	0,32	0,32	0,38	0,32	0,40	0,37
	RMSECV	6,7	6,6	6,4	6,6	6,2	6,5
	R²_{pred}	0,25	0,12	0,30	0,25	0,28	0,29
	RMSEP	6,5	7,0	6,3	6,5	6,4	6,3
	RPD	1,1	1,1	1,2	1,1	1,2	1,2

*EC=Espectro completo; Valores de RMSE em mg kg⁻¹

Por meio dos valores de RMSEP e RPD observa-se que a seleção de variáveis por iSPA-PLS e SiPLS foram as que apresentaram o melhor desempenho para ambas as faixas espectrais com escala de Pareto. Sendo que o valor do RMSEP foi reduzido em 7% (de 5,8 para 5,4 mg kg⁻¹) utilizando estes métodos de seleção de variáveis na faixa espectral de Na-Sc com escala de Pareto e em 3% (de 6,5 para 6,3 mg kg⁻¹) na faixa espectral de Ti-U com escala de Pareto. Para os dados centrados na média, o método de seleção de variáveis que apresentou o melhor desempenho foi o FiPLS, reduzindo os valores de RMSEP em 12% (de 6,7 para 5,9 mg kg⁻¹) e 3% (de 6,5 para 6,3 mg kg⁻¹) com relação aos modelos com o espectro completo de Na-Sc e Ti-U, respectivamente. Para comparar a acurácia desses modelos foi realizado um teste de randomização em um intervalo de confiança de 95%, conforme Tabela 16.

Tabela 16 - Teste de randomização para comparação dos melhores modelos para o P_{disp} em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X	0,04	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com escalamento de Pareto.
Pareto (Na-Sc) iSPA		
MC (Na-Sc) EC X	0,01	Houve diferença significativa na seleção de variáveis por FiPLS no espectro de Na-Sc com centragem na média.
MC (Na-Sc) FiPLS		
Pareto (Ti-U) EC X	0,22	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com escalamento de Pareto.
Pareto (Ti-U) iSPA		
MC (Ti-U) EC X	0,20	Não houve diferença significativa na seleção de variáveis por FiPLS no espectro de Ti-U com centragem na média.
MC (Ti-U) FiPLS		
Pareto (Na-Sc) iSPA X	0,01	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
MC (Na-Sc) FiPLS		
Pareto (Ti-U) iSPA X	0,37	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
MC (Ti-U) FiPLS		
Pareto (Na-Sc) iSPA X	0,01	Houve diferença significativa entre os melhores modelos de cada condição de medida.
Pareto (Ti-U) iSPA		

Os resultados do teste de randomização mostram que a seleção de variáveis na faixa espectral de Na-Sc apresentou melhora significativa em termos de acurácia para ambos os pré-processamentos. Por outro lado, a seleção de variáveis

no espectro Ti-U não apresentou melhora significativa. Ao comparar os melhores modelos de cada pré-processamento, observa-se que somente para a faixa espectral de Na-Sc o modelo com a escala de Pareto foi mais acurado do que com a centragem na média. Além disso, somente os modelos com a seleção de variáveis por iSPA-PLS e SiPLS sobre a faixa espectral de Na-Sc com escala de Pareto apresentaram $RPD=1,4$, o qual, apesar de ser considerado um modelo fraco, ainda pode ser utilizado para avaliações qualitativas e de correlação de acordo com a classificação de Viscarra Rossel; McGlynn; Mcbratney (2006). Todos os demais modelos apresentaram $RPD < 1,4$, sendo considerados muito ruins.

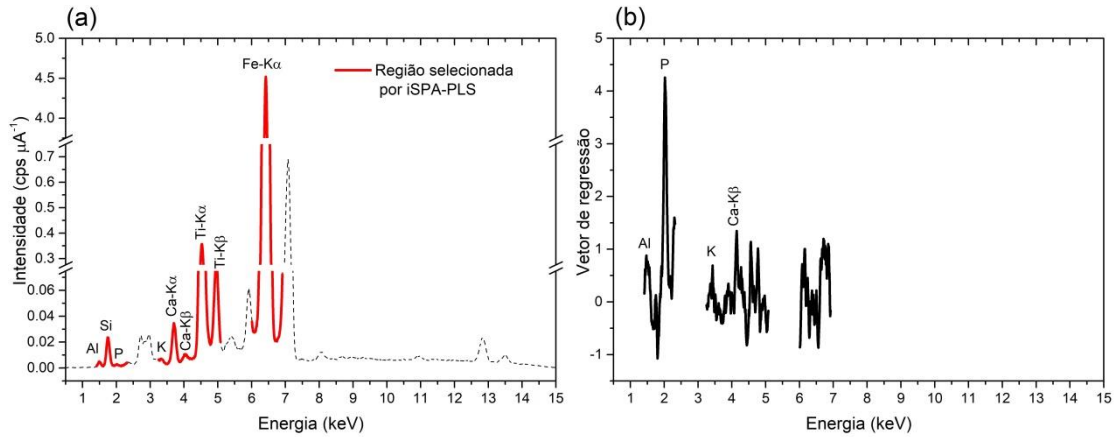
Dessa forma, observa-se que, para o P_{disp} , houve diferença entre os modelos com relação à condição de medida, sendo os modelos com a faixa espectral de Na-Sc foram mais acurados. Isso pôde ser verificado pelo teste de randomização (Tabela 16), que mostrou diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U em um nível de confiança de 95% (p -valor = 0,01).

6.4.2.2 Interpretação espectral das variáveis selecionadas nos modelos de P_{disp}

Conforme discutido no item anterior, o melhor modelo para o P_{disp} foi utilizando a seleção de variáveis por iSPA-PLS sobre o espectro de Na-Sc com escalamento de Pareto (Pareto (Na-Sc) iSPA-PLS). A Figura 25 apresenta as variáveis selecionadas para este modelo com seu respectivo vetor de regressão. Percebe-se que foram selecionadas as regiões do espectro que abrangem os picos de Al, Si, P, K, Ca, Ti e Fe- $K\alpha$. Avaliando o vetor de regressão (Figura 25 (b)), observa-se que o elemento que mais contribuiu para a modelagem foi o P, seguido do Ca, Al e K.

Vale ressaltar que em todos os modelos para a faixa espectral de Na-Sc foi selecionada a região de 1,4 a 2,3 keV, sendo que nessa região está presente o pico de P ($P-K\alpha = 2,0$ keV). Por outro lado, para o espectro de Ti-U essa região só foi selecionada nos modelos que apresentaram o melhor desempenho.

Figura 25 - Intervalos de variáveis selecionadas para o melhor modelo de P_{disp} (a) com seu respectivo vetor de regressão (b).



Fonte: do autor.

O P no solo apresenta-se em formas orgânicas e inorgânicas, dependendo da natureza do composto a que está ligado. O fósforo inorgânico compreende as formas de P dos minerais primários, adsorvidos aos colóides do solo com diferentes graus de estabilidade, como em ligações com Fe, Al, Ca, argilas silicatadas e óxidos de Fe e Al da fração argila (DOS SANTOS; GATIBONI; KAMINSKI, 2008). Para extração do P_{disp} utilizam-se soluções que promovem a dissolução parcial dos colóides, sendo extraídas, em ordem decrescente de efetividade, as formas de P ligadas a Ca, Al e Fe (KAMPRATH; WATSON, 1980). Portanto, as regiões selecionadas pelo método iSPA-PLS bem como suas contribuições no vetor de regressão mostraram-se coerentes com o que é relatado na literatura. Nesse caso, a contribuição do K pode estar associada à fertilização da área em estudo. Isso também poderia explicar porque o modelo que teve o melhor desempenho foi utilizando a faixa espectral de Na-Sc com escalamento de Pareto, uma vez que essa condição de medida é mais adequada para a excitação do P, bem como para os demais elementos (Al, K e Ca) (TAVARES et al., 2020a; VAN GRIEKEN; MARKOWICZ, 2002). Além disso, o rendimento fluorescente para esses elementos é baixo, o que demonstra que o uso do escalamento de Pareto foi adequado para dar um peso relativo maior a eles na modelagem do P_{disp} .

6.4.2.3 Figuras de mérito dos modelos para P_{disp}

A Tabela 17 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia do melhor modelo para o P_{disp} . O teste DW em um nível de confiança de 95% demonstra a ausência de atocorrelação entre os resíduos, indicando linearidade do modelo. O erro relativo geral do modelo foi de 38%, esse resultado demonstra que o modelo para P_{disp} não teve veracidade adequada, como atestado pelos valores de RPD. Além disso, a Figura 26 mostra que a maioria das amostras do conjunto de previsão tiveram erros relativos acima de $\pm 20\%$, sendo que algumas amostras ultrapassaram o erro relativo de $\pm 100\%$, chegando, em alguns casos, a 240%. No entanto, não há erros sistemáticos presentes neste modelo.

A precisão/repetitividade foi avaliada pelo coeficiente de variação (CV) de medidas em triplicata de três diferentes níveis de concentração, baixo ($2,4 \text{ mg kg}^{-1}$), médio ($12,0 \text{ mg kg}^{-1}$) e alto ($25,4 \text{ mg kg}^{-1}$). Todos os CVs foram superiores a 10%, sendo que o nível baixo de concentração foi o que apresentou o pior CV (61%). Isso indica que além de não haver veracidade, esses resultados não foram precisos. Além disso, o método não se mostrou robusto com relação à variação de granulometria.

Tabela 17 - Figuras de mérito do melhor modelo para P_{disp} .

	Pareto (Na-Sc) iSPA-PLS
Linearidade (p-valor DW)	0,09
SEN^a	0,66
γ^{-1} (mg kg^{-1})	0,1
 Bias (t)^b	0,33 (0,69)
ER(%)	38
CV(%) repetitividade^c	61/18/16
Robustez (p-valor)^d	0,01
LOD (mg kg^{-1})	0,4
LOQ (mg kg^{-1})	1,1

^a Valor expresso como a razão entre intensidade e mg kg^{-1}

^b $t_{critico} = 1,98$

^c Valores em nível baixo ($2,4 \text{ mg kg}^{-1}$), médio ($12,0 \text{ mg kg}^{-1}$) e alto ($25,4 \text{ mg kg}^{-1}$).

^d p-valor do teste de randomização em 95% de confiança.

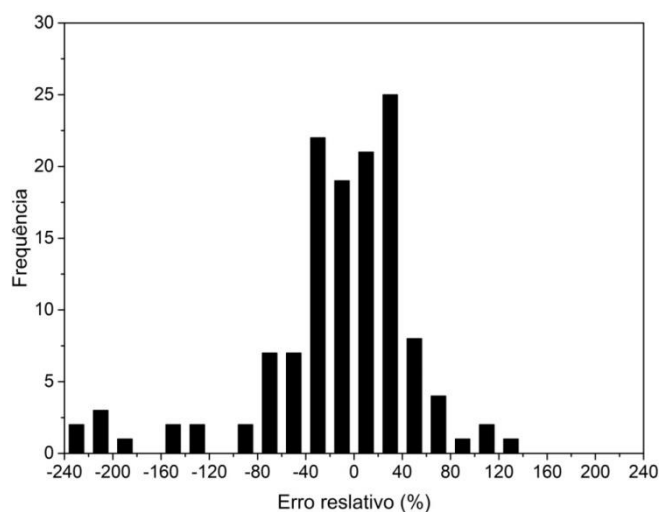
Avaliaram-se também outras figuras de mérito como sensibilidade pura (SEN), inverso da sensibilidade analítica (γ^{-1}), limite de detecção (LOD) e limite

de quantificação (LOQ). O γ^{-1} indicou que a resolução do método foi de $0,1 \text{ mg kg}^{-1}$, com limite mínimo de detecção de $0,4 \text{ mg kg}^{-1}$ e LOQ de $1,1 \text{ mg kg}^{-1}$. Neste caso, algumas amostras do conjunto de calibração e previsão têm valores inferiores ao LOQ do modelo.

Esses resultados atestaram que nem o melhor modelo para P_{disp} apresentou veracidade e precisão adequadas, onde o desempenho deste modelo foi considerado fraco ($\text{RPD}=1,4$) (VISCARRA ROSSEL; MCGLYNN; MCBRATNEY, 2006). Diversos estudos na literatura, utilizando dados de concentração de XRF bem como dados espectrais de Vis-NIR, relatam dificuldades na modelagem do P_{disp} , com R^2 variando de 0,03 a 0,60 onde na maioria dos casos o RPD é $<1,4$ (ANDRADE et al., 2020; DEMATTÊ et al., 2019a; TEIXEIRA et al., 2020; PINHEIRO et al., 2017; TAVARES et al., 2020a). Nesses estudos, os autores atribuem a dificuldade de determinar o P_{disp} devido a seus baixos teores, que dificultam a determinação por EDXRF e também à sua alta variabilidade e labilidade, que podem tornar as determinações convencionais imprecisas.

Dessa forma, mais estudos devem ser realizados para verificar a viabilidade do uso de dados espectrais de EDXRF para determinação de P_{disp} em solos agrícolas.

Figura 26 - Erro relativo para o conjunto de previsão do melhor modelo para P_{disp} .



Fonte: do autor.

6.4.3 Modelos para Ca^{2+}

6.4.3.1 Pré-processamento e seleção de variáveis dos modelos de Ca^{2+}

Os resultados para a calibração e previsão dos métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 18. Nenhuma amostra foi identificada como *outlier* para os modelos com os dados centrados na média. No entanto, duas amostras de calibração foram identificadas como *outliers* nos modelos com escalamento de Pareto.

De forma geral, é possível notar (Tabela 18) que os RMSEPs e RMSECVs de todos os modelos para o Ca^{2+} foram similares. A exceção foi para os modelos com espectro completo centrados na média, que apresentaram valores de RMSEP e RMSECV superiores aos demais. Os modelos utilizando escalamento de Pareto apresentaram valores de RMSEP menor do que os modelos com centragem na média, sendo essa diferença mais evidente para os modelos com o espectro completo.

O RMSEP para o modelo com espectro completo de Na-Sc foi reduzido em 26% (de 0,92 para 0,68 $\text{cmol}_c \text{kg}^{-1}$) utilizando escalamento de Pareto. Por meio do teste de randomização (p -valor = 0,0005), verificou-se que essa diferença foi significativa em um nível de confiança de 95%. Para o espectro de Ti-U, o teste de randomização demonstrou que não houve diferença significativa entre os modelos com escalamento de Pareto e centragem na média (Tabela 19). No entanto, o RMSEP para o modelo com espectro completo de Ti-U foi reduzido em 9% (de 0,80 para 0,73 $\text{cmol}_c \text{kg}^{-1}$) com escalamento de Pareto.

Em geral, a seleção de variáveis melhorou a capacidade preditiva dos modelos, sendo esta melhora muito mais evidente para os dados centrados na média. Neste caso, utilizando a seleção de variáveis por FiPLS e iSPA-PLS, os valores de RMSEP foram reduzidos em 25% (de 0,92 para 0,69 $\text{cmol}_c \text{kg}^{-1}$) e 6% (de 0,80 para 0,75 $\text{cmol}_c \text{kg}^{-1}$) com relação aos modelos com o espectro completo de Na-Sc e Ti-U, respectivamente. Novamente, a seleção de variáveis por iPLS foi o único método que piorou o desempenho dos modelos. Apesar disso, a região selecionada por esse algoritmo (de 3,2 a 4,2 keV) se repetiu para todos os outros métodos de seleção de variáveis, evidenciando que ela foi a que mais contribuiu para a modelagem do Ca^{2+} .

Tabela 18 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de Ca^{2+} com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	920	920	368	828
	LV	6	3	5	5	4	6
	R²_{cal}	0,77	0,62	0,77	0,77	0,77	0,77
	RMSEC	0,71	0,93	0,72	0,72	0,73	0,72
	R²_{cv}	0,71	0,61	0,72	0,72	0,72	0,72
	RMSECV	0,82	0,95	0,79	0,79	0,81	0,81
	R²_{pred}	0,77	0,72	0,79	0,79	0,77	0,79
	RMSEP	0,68	0,77	0,67	0,67	0,69	0,67
	RPD	2,2	1,9	2,2	2,2	2,1	2,2
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	552	736	368	276
	LV	5	3	4	5	4	5
	R²_{cal}	0,56	0,58	0,72	0,71	0,71	0,71
	RMSEC	1,01	0,93	0,81	0,82	0,82	0,84
	R²_{cv}	0,52	0,56	0,67	0,67	0,67	0,66
	RMSECV	1,05	1,02	0,88	0,88	0,89	0,89
	R²_{pred}	0,61	0,72	0,77	0,77	0,77	0,77
	RMSEP	0,92	0,78	0,69	0,69	0,69	0,69
	RPD	1,6	1,9	2,1	2,1	2,1	2,1
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	508	1143	508	635
	LV	6	3	6	6	6	6
	R²_{cal}	0,72	0,56	0,74	0,72	0,74	0,69
	RMSEC	0,81	1,01	0,77	0,81	0,77	0,85
	R²_{cv}	0,64	0,50	0,67	0,64	0,67	0,64
	RMSECV	0,93	1,07	0,88	0,93	0,88	0,92
	R²_{pred}	0,76	0,59	0,76	0,76	0,76	0,76
	RMSEP	0,73	0,92	0,71	0,73	0,71	0,72
	RPD	2,0	1,6	2,0	2,0	2,0	2,0
Faixa Ti-U centrado na média	nº Variáveis	1270	127	1016	1016	508	889
	LV	7	5	7	7	7	7
	R²_{cal}	0,62	0,53	0,90	0,90	0,83	0,88
	RMSEC	0,93	1,05	0,48	0,48	0,63	0,50
	R²_{cv}	0,50	0,49	0,61	0,61	0,59	0,61
	RMSECV	1,09	1,10	0,96	0,96	0,99	0,96
	R²_{pred}	0,71	0,58	0,74	0,74	0,69	0,74
	RMSEP	0,80	0,94	0,75	0,75	0,81	0,75
	RPD	1,8	1,5	1,9	1,9	1,8	2,0

*EC=Espectro completo; Valores de RMSE em $\text{cmol}_c \text{ kg}^{-1}$

Os resultados do teste de randomização (Tabela 19) demonstram que não houve diferença significativa entre os melhores modelos de cada pré-processamento. Além disso, o teste também mostra que não houve diferença significativa entre os melhores modelos de cada condição de medida (Na-Sc e Ti-U). Entretanto, apenas os modelos com o espectro de Na-Sc apresentaram $RPD > 2,0$.

Tabela 19 - Teste de randomização para comparação dos melhores modelos para o Ca^{2+} em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X Pareto (Na-Sc) FiPLS	0,42	Não houve diferença significativa na seleção de variáveis por FiPLS no espectro de Na-Sc com escalamento de Pareto.
MC (Na-Sc) EC X MC (Na-Sc) FiPLS	0,0005	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com centragem na média.
Pareto (Ti-U) EC X Pareto (Ti-U) FiPLS	0,17	Não houve diferença significativa na seleção de variáveis por FiPLS no espectro de Ti-U com escalamento de Pareto.
MC (Ti-U) EC X MC (Ti-U) iSPA	0,15	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com centragem na média.
Pareto (Na-Sc) FiPLS X MC (Na-Sc) FiPLS	0,11	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
Pareto (Ti-U) FiPLS X MC (Ti-U) iSPA	0,16	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
Pareto (Na-Sc) FiPLS X Pareto (Ti-U) FiPLS	0,10	Não houve diferença significativa entre os melhores modelos com o espectro de Na-Sc e Ti-U

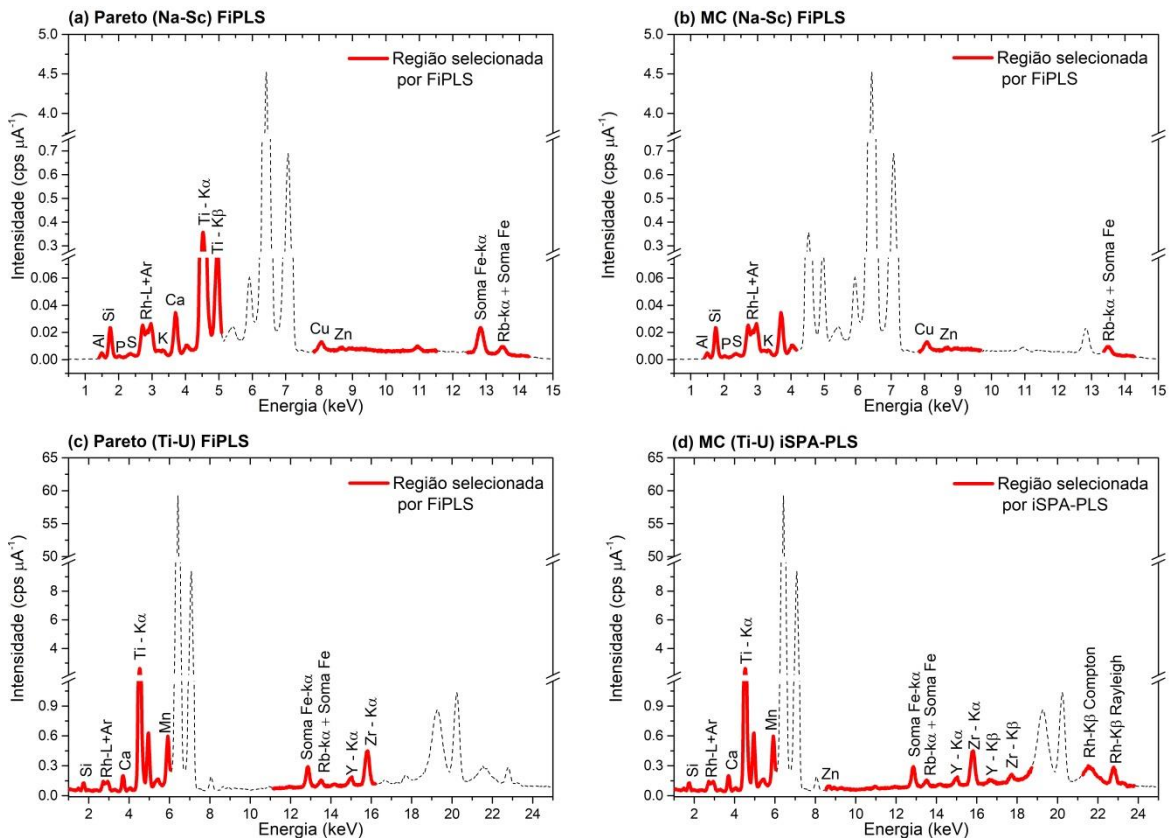
6.4.3.2 Interpretação espectral das variáveis selecionadas nos modelos de Ca^{2+}

As variáveis selecionadas para os melhores modelos de cada pré-processamento são mostradas na Figura 27 e os vetores de regressão para os respectivos modelos são mostrados na Figura 28.

Por meio da Figura 27 observa-se que a região inicial do espectro, que contém os picos de Al, Si, P, S, K, Ca e linhas L do Rh, foi utilizada em todos os modelos. Isso pode ser uma das causas para o desempenho ligeiramente melhor dos modelos com o espectro de Na-Sc, visto que essa condição é mais adequada

para essa faixa inicial do espectro (TAVARES et al., 2020a; VAN GRIEKEN; MARKOWICZ, 2002). Além dessa região, também foi utilizado a faixa espectral que compreende os picos de Cu, Zn e picos soma do Fe para os melhores modelos do espectro de Na-Sc, e os picos de Y, Zr e soma do Fe para os melhores modelos com o espectro de Ti-U.

Figura 27 - Intervalos de variáveis selecionadas para os melhores modelos de Ca^{2+} com escalamento de Pareto e centragem na média.



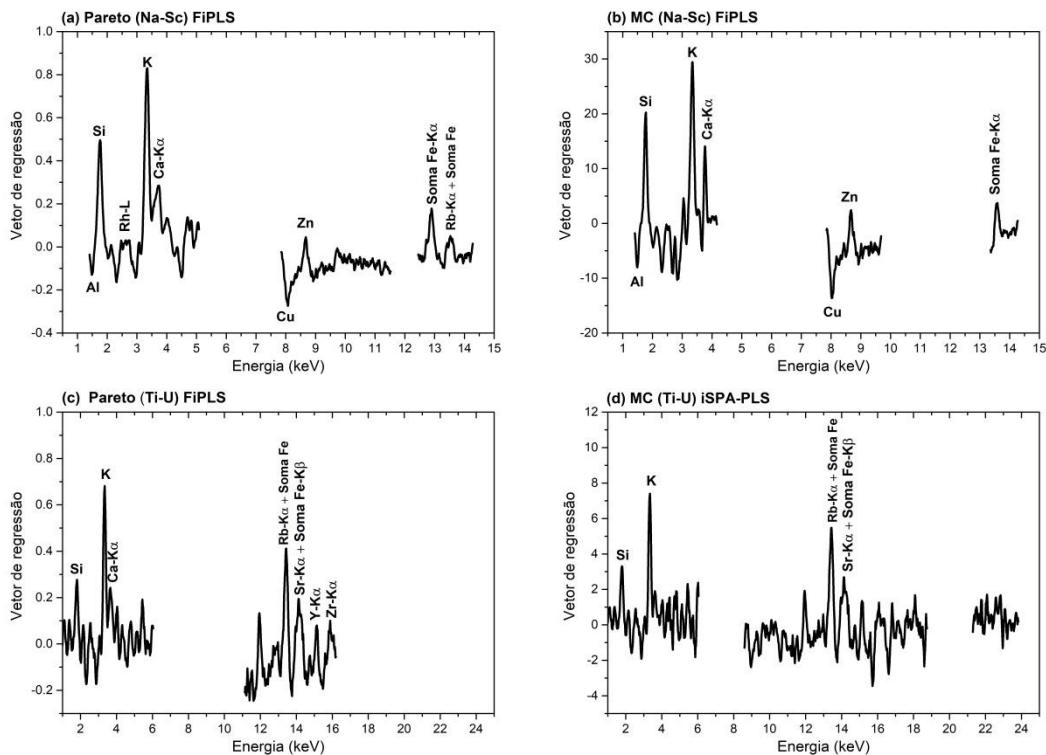
Fonte: do autor.

Em geral, os vetores de regressão (Figura 28) dos melhores modelos (de cada condição de medida e pré-processamento) tiveram um comportamento semelhante, sendo que o pico de K foi o que apresentou a maior contribuição para modelagem do Ca^{2+} . Esse mesmo comportamento foi observado na análise de correlação da Figura 16, na qual o Ca trocável (Ca^{2+}) apresentou um coeficiente de correlação (r) de 0,65 com o K total e r de 0,50 com o Ca total. De fato, esperava-se que os picos de Ca fossem os que apresentassem a maior contribuição, uma vez que os teores totais desse elemento em solos brasileiros estão associados principalmente aos seus teores trocáveis (Ca^{2+}) (TEIXEIRA et al.,

2020). Isso se deve ao fato que os principais minerais dos solos brasileiros são os quartzos, micas, caulinitas e óxidos de Fe e Al, sendo que o Ca não prevalece em sua estrutura cristalina (BRINATTI et al., 2010; SILVA et al., 2018).

Como já mencionado no item 6.3.1, a presença de Ca em solos brasileiros está associada principalmente à prática de calagem que usualmente é utilizada para aumentar o pH do solo. Essa prática, além de aumentar os teores de Ca, adiciona Mg no solo. Além disso, alguns estudos também têm encontrado correlação entre a aplicação de calcário com o Sr (MYRVANG et al., 2017; TEIXEIRA et al., 2018). Portanto, a contribuição dos picos de K para a previsão de Ca^{2+} pode estar associada ao manejo e fertilização do solo e isso também pode indicar que há uma relação sinérgica entre o Ca e K.

Figura 28 - Vetor de regressão dos melhores modelos PLSR para o Ca^{2+} com escalamento de Pareto e centragem na média.



Fonte: do autor.

Além da contribuição positiva dos picos de K e Ca, também se observa que os picos de Si, Zn, Rb-K α , Sr-K α e picos soma do Fe contribuíram positivamente e Al e Cu contribuíram negativamente para modelagem do Ca^{2+} . Essa

relação negativa do Al e Cu com o Ca^{2+} faz sentido, uma vez que existe uma relação antagônica entre esses elementos na troca de cargas do solo (SPOSITO, 2008).

6.4.3.3 Figuras de mérito dos modelos para Ca^{2+}

A Tabela 20 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia dos modelos para Ca^{2+} . De acordo com o teste DW, todos os modelos apresentaram ausência de autocorrelação entre os resíduos, indicando linearidade. Além disso, nenhum modelo apresentou erro sistemático.

A veracidade dos modelos foi avaliada com base nos valores de RMSEP e RPD (Tabela 18), que atestaram que a maioria dos modelos apresentou veracidade adequada ($\text{RPD} > 2,0$). Além desses parâmetros, o erro relativo geral (ER (%)) e os erros relativos individuais (Figura 29) do conjunto de previsão também foram utilizados para avaliar a veracidade dos modelos. O erro relativo geral foi de 11% para o modelo com seleção de variáveis por FiPLS no espectro de Na-Sc com escalamento de Pareto (Pareto (Na-Sc) FiPLS), 12% para os modelos com seleção de variáveis por FiPLS no espectro de Na-Sc centrado na média (MC (Na-Sc) FiPLS) e no espectro de Ti-U com escalamento de Pareto (Pareto (Ti-U) FiPLS) e 13% para o modelo com seleção de variáveis por iSPA-PLS no espectro de Ti-U centrado na média (MC (Ti-U) iSPA-PLS). A Figura 29 mostra que o modelo Pareto (Na-Sc) FiPLS foi o que apresentou os menores erros relativos individuais, sendo que a maioria das amostras apresentaram erros relativos abaixo de $\pm 20\%$.

Três diferentes níveis de concentração (baixo ($3,1 \text{ cmol}_c \text{ kg}^{-1}$), médio ($5,8 \text{ cmol}_c \text{ kg}^{-1}$) e alto ($8,2 \text{ cmol}_c \text{ kg}^{-1}$)) foram utilizados para avaliar a precisão dos modelos por repetitividade. Os CVs das medidas em triplicata ficaram abaixo de 7% para os modelos com escalamento de Pareto. O modelo MC (Ti-U) iSPA-PLS apresentou a pior precisão, com CV de 27% no nível baixo e 10 % no nível alto de concentração. Com exceção do modelo MC (Na-Sc) FiPLS, todos os outros mostraram-se robustos quanto a variação de granulometria de 2 mm para $125 \mu\text{m}$ do conjunto de validação.

Tabela 20 - Figuras de mérito dos melhores modelos para Ca^{2+} .

	Pareto (Na-Sc) FiPLS	MC (Na-Sc) FiPLS	Pareto (Ti-U) FiPLS	MC (Ti-U) iSPA-PLS
Linearidade (p-valor DW)	0,52	0,21	0,75	0,29
SEN^a	0,21	0,006	0,32	0,02
γ^{-1} (cmol_c kg⁻¹)	0,04	0,01	0,2	0,2
 Bias (t)^b	0,09 (1,49)	0,04 (0,68)	0,04 (0,64)	0,02 (0,35)
ER(%)	11	12	12	13
CV(%)				
repetitividade^c	4/2/6	1/3/11	7/6/6	27/4/10
Robustez (p-valor) ^d	0,35	0,002	0,48	0,11
LOD (cmol_c kg⁻¹)	0,12	0,03	0,6	0,6
LOQ (cmol_c kg⁻¹)	0,36	0,10	1,9	1,7

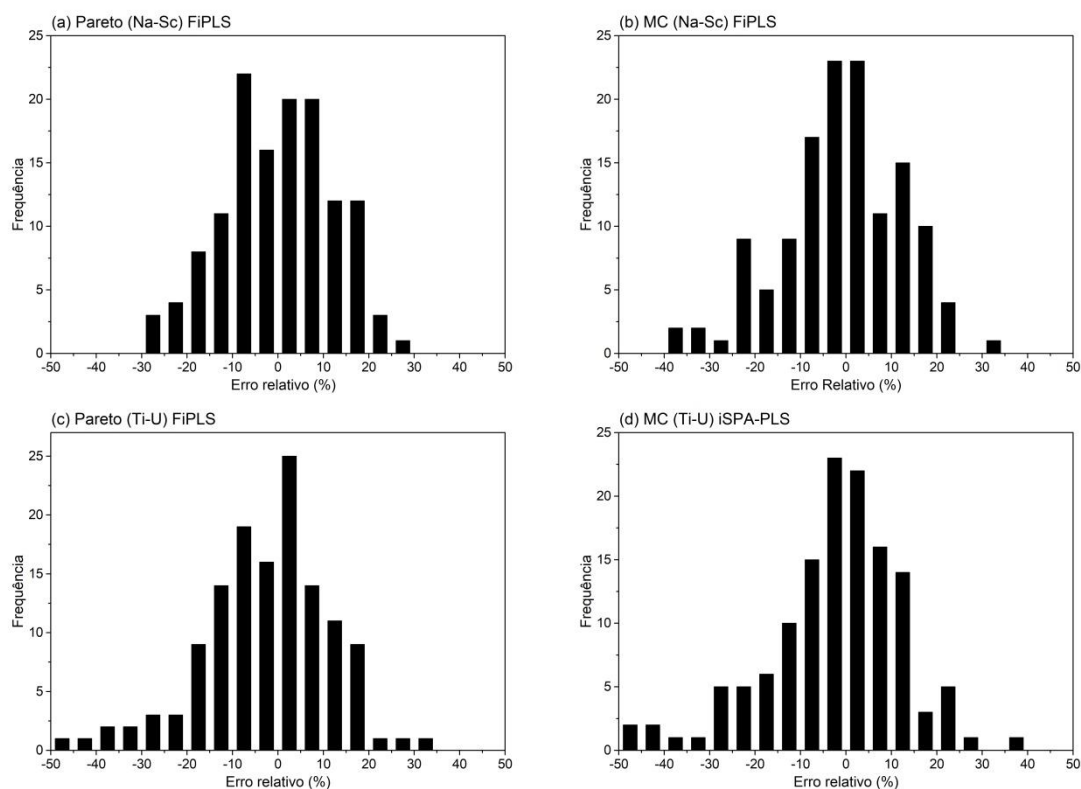
^a Valor expresso como a razão entre intensidade e cmol_c kg⁻¹

^b $t_{critico} = 1,98$

^c Valores em nível baixo (3,1 cmol_c kg⁻¹), médio (5,8 cmol_c kg⁻¹) e alto (8,2 cmol_c kg⁻¹).

^d p-valor do teste de randomização em 95% de confiança.

Figura 29 - Erro relativo para o conjunto de previsão dos melhores modelos para Ca^{2+} .



Fonte: do autor.

Além disso, também foram avaliados a SEN, γ^{-1} , LOD e LOQ. O γ^{-1} indicou que a resolução do método foi menor para os modelos com espectro de Na-Sc. Os LOD ficaram abaixo de 0,12 $\text{cmol}_c \text{kg}^{-1}$ para os modelos com espectro de Na-Sc e abaixo de 0,60 $\text{cmol}_c \text{kg}^{-1}$ para os modelos com o espectro de Ti-U.

Portanto, esses resultados atestaram que apenas os modelos com escalamento de Pareto mostraram robustez, veracidade e precisão adequadas. O valor de RPD foi de 2,2 para o modelo Pareto (Na-Sc) FiPLS e 2,0 para o modelo Pareto (Ti-U) FiPLS, indicando um bom desempenho para prever o Ca^{2+} neste conjunto de dados. Outros estudos utilizando dados de Vis-NIR, XRF, bem como a fusão de dados também demonstram bom desempenho dos modelos para previsão do Ca trocável, com RPD variando de 1,3 a 4,8 (ANDRADE et al., 2020; DEMATTÊ et al., 2019a; TEIXEIRA et al., 2020; LIMA et al., 2019; MUNNAF; NAWAR; MOUAZEN, 2019; O'ROURKE et al., 2016; PINHEIRO et al., 2017; SILVA et al., 2017; TAVARES et al., 2020a, 2020b), indicando a viabilidade dessas técnicas na previsão desse parâmetro.

6.4.4 Modelos para Mg^{2+}

6.4.4.1 Pré-processamento e seleção de variáveis dos modelos de Mg^{2+}

Os resultados para a calibração e previsão dos modelos de Mg^{2+} com os métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 21. É importante ressaltar que nenhuma amostra foi identificada com *outlier* nesses modelos.

De forma geral, é possível notar que os RMSEPs e RMSECVs de todos os modelos para o Mg^{2+} foram similares, sendo que o único modelo com desempenho considerado razoável ($2,0 > \text{RPD} > 1,8$) foi com a seleção de variáveis por BiPLS no espectro de Na-Sc com escalamento de Pareto (Pareto (Na-Sc) BiPLS). Esses resultados (Tabela 21) são similares com o que é relatado na literatura, em que modelos utilizando dados de Vis-NIR e de XRF apresentam R^2 de previsão variando de 0,33 a 0,85 e RPD variando de 1,2 a 2,6 (ANDRADE et al., 2020; DEMATTÊ et al., 2019a; MUNNAF; NAWAR; MOUAZEN, 2019; PINHEIRO et al., 2017; SILVA et al., 2017; TAVARES et al., 2020a).

Tabela 21 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de Mg^{2+} com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	368	644	368	276
	LV	3	3	4	3	5	5
	R²_{cal}	0,62	0,58	0,66	0,67	0,67	0,66
	RMSEC	0,35	0,38	0,34	0,33	0,33	0,34
	R²_{cv}	0,59	0,55	0,64	0,62	0,64	0,62
	RMSECV	0,37	0,39	0,35	0,35	0,34	0,35
	R²_{pred}	0,64	0,59	0,64	0,67	0,64	0,64
	RMSEP	0,34	0,37	0,34	0,33	0,35	0,34
	RPD	1,7	1,6	1,7	1,8	1,7	1,7
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	644	736	368	368
	LV	4	3	4	4	4	4
	R²_{cal}	0,61	0,59	0,64	0,66	0,62	0,62
	RMSEC	0,37	0,37	0,35	0,34	0,35	0,37
	R²_{cv}	0,58	0,55	0,59	0,59	0,59	0,59
	RMSECV	0,38	0,39	0,37	0,37	0,37	0,38
	R²_{pred}	0,61	0,61	0,66	0,64	0,64	0,64
	RMSEP	0,37	0,37	0,34	0,35	0,35	0,36
	RPD	1,6	1,6	1,7	1,6	1,6	1,6
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	381	635	508	762
	LV	6	2	3	6	6	3
	R²_{cal}	0,69	0,58	0,62	0,66	0,66	0,62
	RMSEC	0,33	0,38	0,36	0,34	0,34	0,36
	R²_{cv}	0,59	0,55	0,59	0,59	0,59	0,59
	RMSECV	0,38	0,39	0,37	0,37	0,37	0,37
	R²_{pred}	0,66	0,62	0,62	0,62	0,64	0,64
	RMSEP	0,33	0,36	0,36	0,35	0,35	0,35
	RPD	1,7	1,6	1,6	1,6	1,7	1,6
Faixa Ti-U centrado na média	nº Variáveis	1270	127	762	762	508	762
	LV	6	4	4	4	4	4
	R²_{cal}	0,67	0,5776	0,6084	0,6084	0,6241	0,6084
	RMSEC	0,34	0,38	0,36	0,36	0,35	0,36
	R²_{cv}	0,58	0,5476	0,5776	0,5776	0,6084	0,5776
	RMSECV	0,38	0,39	0,38	0,38	0,36	0,38
	R²_{pred}	0,62	0,6241	0,6241	0,6241	0,64	0,6241
	RMSEP	0,35	0,35	0,35	0,35	0,34	0,35
	RPD	1,7	1,7	1,7	1,7	1,7	1,7

*EC=Espectro completo; Valores de RMSE em $cmol_c kg^{-1}$

Os resultados do teste de randomização (Tabela 22) demonstram que a seleção de variáveis só apresentou melhora significativa para os modelos baseados nos espectros de Na-Sc centrados na média, para o qual a seleção de variáveis por FiPLS reduziu o valor de RMSEP em 8% (de 0,37 para 0,34 $\text{cmol}_c \text{ kg}^{-1}$) com relação ao espectro completo. No caso dos modelos na faixa espectral de Ti-U com escalamento de Pareto, observa-se que os métodos de seleção de variáveis aumentaram os valores de RMSEP com relação ao modelo com o espectro completo. Também, de acordo com o teste de randomização, não houve diferença significativa entre os melhores modelos de cada pré-processamento nem diferença significativa entre os melhores modelos de cada condição de medida (Na-Sc e Ti-U). Entretanto, apenas o modelo Pareto (Na-Sc) BiPLS apresentou desempenho considerado razoável ($2,0 > \text{RPD} > 1,8$), enquanto os demais modelos apresentaram desempenho considerado fraco ($1,8 > \text{RPD} > 1,4$) (VISCARRA ROSSEL; MCGLYNN; MCBRATNEY, 2006).

Tabela 22 - Teste de randomização para comparação dos melhores modelos para o Mg^{2+} em um intervalo de confiança de 95%.

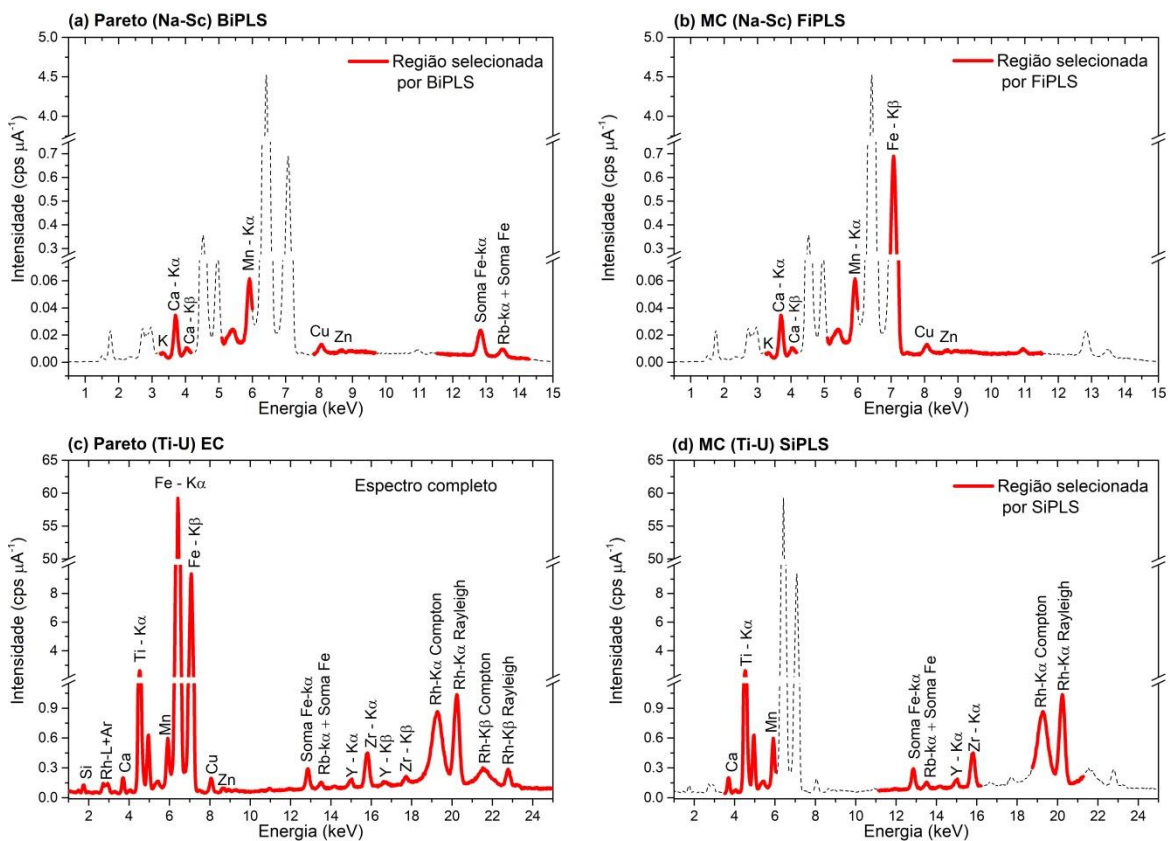
Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X	0,20	Não houve diferença significativa na seleção de variáveis por BiPLS no espectro de Na-Sc com escalamento de Pareto.
Pareto (Na-Sc) BiPLS		
MC (Na-Sc) EC X	0,01	Houve diferença significativa na seleção de variáveis por FiPLS no espectro de Na-Sc com centragem na média.
MC (Na-Sc) FiPLS		
MC (Ti-U) EC X	0,20	Não houve diferença significativa na seleção de variáveis por SiPLS no espectro de Ti-U com centragem na média.
MC (Ti-U) SiPLS		
Pareto (Na-Sc) BiPLS X	0,17	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
MC (Na-Sc) FiPLS		
Pareto (Ti-U) EC X	0,71	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
MC (Ti-U) SiPLS		
Pareto (Na-Sc) BiPLS X	0,06	Não houve diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U
Pareto (Ti-U) EC		

6.4.4.2 Interpretação espectral das variáveis selecionadas nos modelos de Mg^{2+}

As variáveis selecionadas para os melhores modelos de cada pré-processamento são mostradas na Figura 30 e os vetores de regressão para os respectivos modelos são mostrados na Figura 31.

De forma geral, observa-se que foram selecionadas regiões ao longo de todo o espectro para a modelagem de Mg^{2+} . Porém, devido ao baixo rendimento fluorescente para a região de baixas energias, não foi possível detectar o raio X característico do Mg ($Mg-K\alpha = 1,3 \text{ keV}$). Dessa forma, poucos modelos utilizaram a região em torno de 1,3 keV para modelagem do Mg^{2+} . Por outro lado, a região com os picos de Ca foi utilizada em todos os modelos.

Figura 30 - Intervalos de variáveis selecionadas para os melhores modelos de Mg^{2+} com escalamento de Pareto e centragem na média.

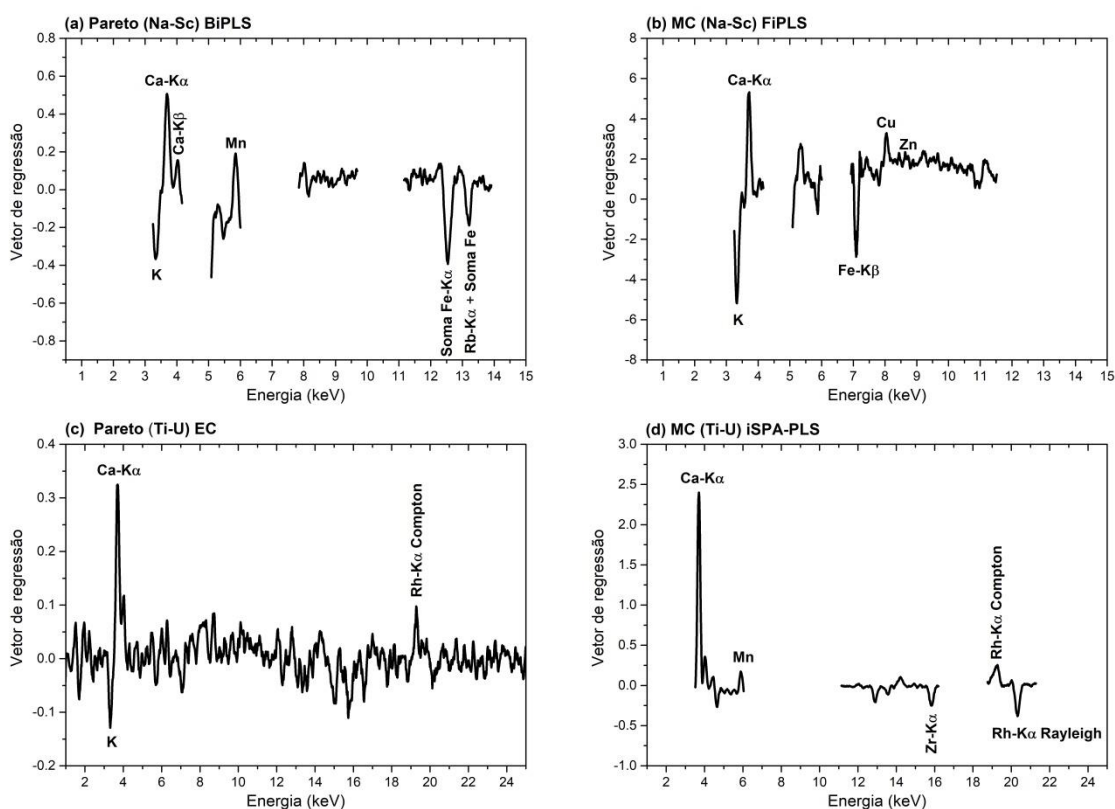


Fonte: do autor.

Os vetores de regressão (Figura 31) dos melhores modelos (de cada condição de medida e pré-processamento) para o Mg^{2+} também apresentaram um comportamento semelhante, sendo que os picos de Ca-K α e K-K α foram os que

apresentaram a maior contribuição para todos os modelos. O Ca teve uma contribuição positiva e o K teve uma contribuição negativa. Além disso, observa-se que os picos de Mn-K α e Cu-K α contribuíram positivamente e os picos Fe-K β e soma do Fe contribuíram negativamente para modelagem do Mg²⁺.

Figura 31 - Vetor de regressão dos melhores modelos PLSR para o Mg²⁺ com escalamento de Pareto e centragem na média.



Fonte: do autor.

Por meio da Figura 16, observa-se que o Mg²⁺ apresenta correlações em ordem decrescente com o pH, COT, Ca²⁺ e K⁺. Isso indica que os teores de Mg²⁺ são altamente influenciados pelo manejo do solo. Portanto, a contribuição positiva dos picos de Ca na modelagem deste parâmetro provavelmente está relacionada à aplicação de calcário no solo. Por outro lado, a contribuição negativa dos picos de K indica uma relação antagônica entre esses elementos. Nesse caso, essa relação antagônica sugere que o K total advém de fontes naturais, visto que, de acordo com a Figura 16, há uma correlação positiva entre o Mg²⁺ e K⁺. De forma geral, estes resultados estão de acordo com o que é

relatado por outros autores que utilizaram dados de XRF para determinação de Mg^{2+} em solos tropicais, os quais também observaram que o Ca foi a variável que mais contribuiu para a modelagem do Mg^{2+} (ANDRADE et al., 2020; SILVA et al., 2017; TAVARES et al., 2019).

Para os modelos com o espectro de Ti-U, nota-se que a região de espalhamento (Rh- $K\alpha$ Compton e Rh- $K\alpha$ Rayleigh) teve uma grande contribuição para a modelagem do Mg^{2+} , sendo que o pico Rh- $K\alpha$ Compton teve uma contribuição positiva e o pico de Rh- $K\alpha$ Rayleigh contribuiu negativamente. Como já mencionado, os picos de espalhamento são devidos principalmente à interação do feixe de excitação com elementos de baixo número atômico ($Z < 11$) e componentes orgânicos presentes na amostra (BORTOLETO; BORGES; BUENO, 2007; MORONA et al., 2017). Portanto, apesar da região de espalhamento também carregar informações a respeito do Mg, provavelmente a contribuição destes picos está associada a alta correlação entre o Mg^{2+} e o COT.

6.4.4.3 Figuras de mérito dos modelos para Mg^{2+}

A Tabela 23 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia dos modelos para Mg^{2+} . De acordo com essas figuras de mérito todos os modelos demonstram ser lineares e não apresentam *bias*, em um nível de confiança de 95%. O erro relativo de todos os modelos foi de 16%. No entanto, de acordo com a Tabela 21, somente o modelo Pareto (Na-Sc) BiPLS pode ser considerado razoável ($2,0 > RPD > 1,8$). Além disso, através da Figura 32, observa-se que esse modelo foi o que apresentou os menores erros relativos individuais. Contudo, um número significativo de amostras apresentaram erros relativos acima de $\pm 20\%$, chegando, em alguns casos, a um erro relativo de $|42\%|$.

A precisão dos modelos por repetitividade (medidas em triplicata) foi avaliada um nível de concentração baixo ($1,2 \text{ cmol}_c \text{ kg}^{-1}$), médio ($2,1 \text{ cmol}_c \text{ kg}^{-1}$) e alto ($3,3 \text{ cmol}_c \text{ kg}^{-1}$). Em geral, observa-se que os modelos utilizando a faixa espectral de Na-Sc mostraram-se mais precisos, com $CV < 8\%$. Além disso, o teste de randomização para robustez demonstrou que somente os modelos com escalamento de Pareto foram robustos quanto a variação de granulometria.

Tabela 23 - Figuras de mérito dos melhores modelos para Mg^{2+} .

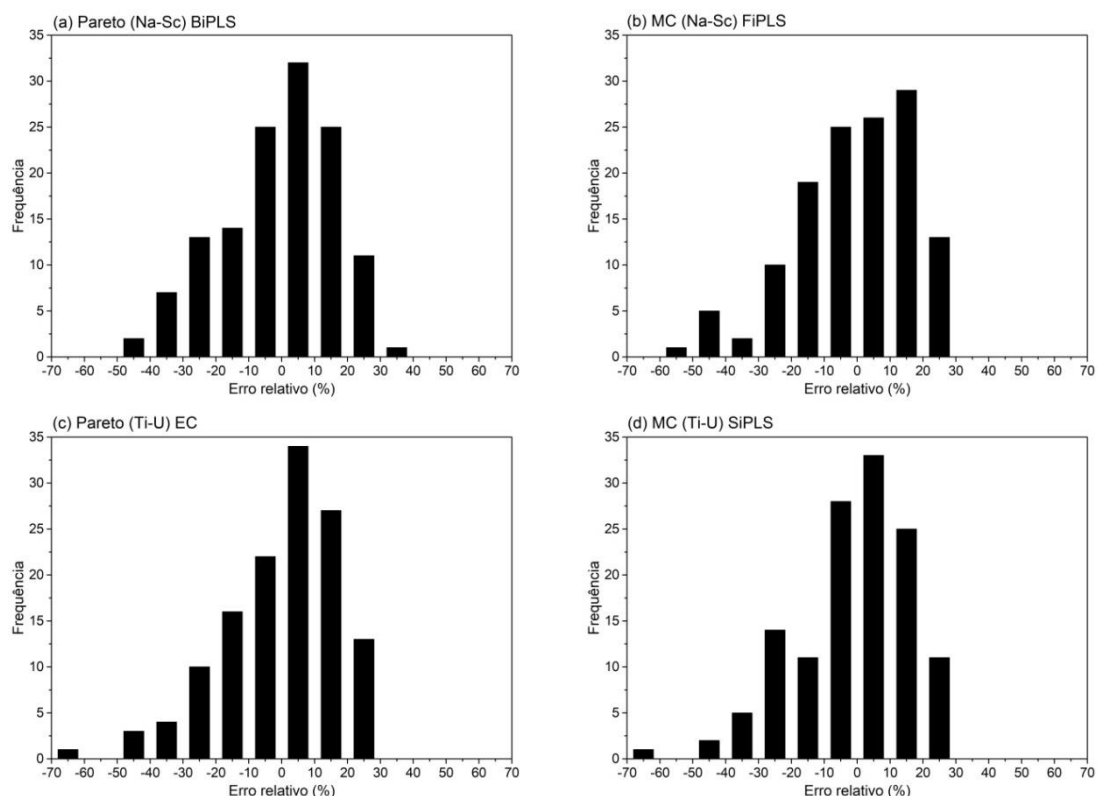
	Pareto (Na-Sc) BiPLS	MC (Na-Sc) FiPLS	Pareto (Ti-U) EC	MC (Ti-U) SiPLS
Linearidade (p-valor DW)	0,34	0,83	0,92	0,77
SEN^a	0,24	0,02	0,63	0,16
γ^{-1} (cmol_c kg⁻¹)	0,03	0,003	0,1	0,02
 Bias (t)^b	0,03 (0,91)	0,05 (1,63)	0,05 (1,78)	0,04 (0,80)
ER(%)	16	16	16	16
CV(%) repetitividade^c	5/2/8	2/3/7	3/5/13	5/4/10
Robustez (p-valor)^d	0,26	0,0005	0,27	0,0005
LOD (cmol_c kg⁻¹)	0,10	0,01	0,3	0,08
LOQ (cmol_c kg⁻¹)	0,31	0,03	1,0	0,25

^a Valor expresso como a razão entre intensidade e cmol_c kg⁻¹

^b $t_{critico} = 1,98$

^c Valores em nível baixo (1,2 cmol_c kg⁻¹), médio (2,1 cmol_c kg⁻¹) e alto (3,3 cmol_c kg⁻¹).

^d p-valor do teste de randomização em 95% de confiança.

Figura 32 - Erros relativos para o conjunto de previsão dos melhores modelos para Mg^{2+} .

Fonte: do autor.

Além de apresentar a pior veracidade, precisão e robustez, o modelo Pareto (Ti-U) Full também apresentou os piores LOD ($0,3 \text{ cmol}_c \text{ kg}^{-1}$) e LOQ ($1,0 \text{ cmol}_c \text{ kg}^{-1}$). Neste caso, LOQ foi superior à concentração de algumas amostras do conjunto de calibração. Para os demais modelos, os LOD ficaram abaixo de $0,1 \text{ cmol}_c \text{ kg}^{-1}$ e os LOQ abaixo de $0,3 \text{ cmol}_c \text{ kg}^{-1}$.

6.4.5 Modelos para K^+

6.4.5.1 Pré-processamento e seleção de variáveis dos modelos de K^+

Os resultados para a calibração e previsão dos modelos de K^+ com os métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 24. Duas amostras do conjunto de calibração foram identificadas como *outliers*. Dessa forma, todos os modelos foram construídos com 262 amostras de calibração e 162 amostras de previsão.

Os resultados da Tabela 24 mostram que todos os modelos geraram altos RMSEPs e RMSECVs para o K^+ , indicando falta de acurácia dos modelos. No entanto, observa-se que os modelos utilizando escalamento de Pareto apresentaram valores de RMSEP menor do que os modelos com a centragem na média, sendo reduzido em 12% (de 0,09 para $0,08 \text{ cmol}_c \text{ kg}^{-1}$) para o modelo com o espectro completo de Na-Sc e em 7% (de 0,12 para $0,11 \text{ cmol}_c \text{ kg}^{-1}$) para o modelo com o espectro completo de Ti-U. Verificou-se, por meio do teste de randomização, que essa diferença foi significativa em um nível de confiança de 95%, com p-valor = 0,01 para os modelos com o espectro completo de Na-Sc e p-valor = 0,03 para os modelos com o espectro completo de Ti-U.

Tabela 24 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de K^+ com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	828	828	368	828
	LV	9	4	8	8	8	9
	R²_{cal}	0,79	0,45	0,74	0,74	0,69	0,74
	RMSEC	0,08	0,12	0,08	0,08	0,09	0,08
	R²_{cv}	0,59	0,42	0,62	0,62	0,56	0,62
	RMSECV	0,10	0,12	0,10	0,10	0,10	0,10
	R²_{pred}	0,58	0,38	0,56	0,56	0,55	0,59
	RMSEP	0,08	0,10	0,08	0,08	0,09	0,08
	RPD	1,5	1,2	1,5	1,5	1,4	1,5
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	736	736	368	368
	LV	4	3	8	8	6	8
	R²_{cal}	0,59	0,45	0,72	0,72	0,64	0,67
	RMSEC	0,10	0,12	0,08	0,08	0,09	0,09
	R²_{cv}	0,45	0,42	0,61	0,61	0,58	0,58
	RMSECV	0,12	0,12	0,10	0,10	0,10	0,10
	R²_{pred}	0,45	0,40	0,56	0,56	0,49	0,53
	RMSEP	0,10	0,10	0,08	0,08	0,09	0,09
	RPD	1,3	1,3	1,5	1,5	1,4	1,4
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	254	508	254	254
	LV	5	2	3	6	3	4
	R²_{cal}	0,55	0,35	0,64	0,72	0,64	0,58
	RMSEC	0,11	0,13	0,09	0,08	0,09	0,10
	R²_{cv}	0,26	0,24	0,44	0,41	0,44	0,40
	RMSECV	0,14	0,14	0,12	0,12	0,12	0,12
	R²_{pred}	0,28	0,42	0,44	0,42	0,44	0,46
	RMSEP	0,11	0,10	0,10	0,10	0,10	0,09
	RPD	1,1	1,3	1,2	1,2	1,2	1,3
Faixa Ti-U centrado na média	nº Variáveis	1270	127	381	762	381	127
	LV	5	2	3	5	3	2
	R²_{cal}	0,35	0,35	0,58	0,35	0,58	0,35
	RMSEC	0,13	0,13	0,10	0,13	0,10	0,13
	R²_{cv}	0,26	0,25	0,42	0,26	0,42	0,25
	RMSECV	0,14	0,14	0,12	0,14	0,12	0,14
	R²_{pred}	0,16	0,42	0,38	0,17	0,38	0,42
	RMSEP	0,12	0,10	0,10	0,12	0,10	0,10
	RPD	1,0	1,3	1,2	1,0	1,2	1,3

*EC=Espectro completo; Valores de RMSE em $\text{cmol}_c \text{ kg}^{-1}$

Em geral, todos os métodos de seleção de variáveis melhoraram a

capacidade preditiva dos modelos. A única exceção foi para os modelos utilizando a faixa espectral de Na-Sc com escalamento de Pareto, cujo modelo com o espectro completo foi o que apresentou o menor valor de RMSEP. No caso do espectro de Na-Sc centrado na média, a seleção de variáveis por FiPLS e BiPLS (que selecionaram a mesma região) reduziu o valor de RMSEP em 14 % (de 0,10 para 0,08 $\text{cmol}_c \text{kg}^{-1}$). Para o espectro de Ti-U a seleção de variáveis por iSPA reduziu o valor de RMSEP em 17% (de 0,11 para 0,09 $\text{cmol}_c \text{kg}^{-1}$) com os dados na escala de Pareto e em 20 % (de 0,11 para 0,09 $\text{cmol}_c \text{kg}^{-1}$) com os dados centrados na média. De acordo com o teste de randomização (Tabela 25), esses métodos de seleção de variáveis produziram uma melhora significativa no desempenho dos modelos.

Os resultados do teste de randomização (Tabela 22) demonstram que não houve diferença significativa ao comparar os melhores modelos dentro de cada faixa espectral. No entanto, o teste de randomização indicou que houve diferença significativa entre o RMSEP dos melhores modelos de cada condição de medida. Isso também pôde ser atestado pelos valores de RPD, que demonstram que somente os modelos com espectro de Na-Sc apresentaram valor de RPD acima de 1,4 (VISCARRA ROSSEL; MCGLYNN; MCBRATNEY, 2006).

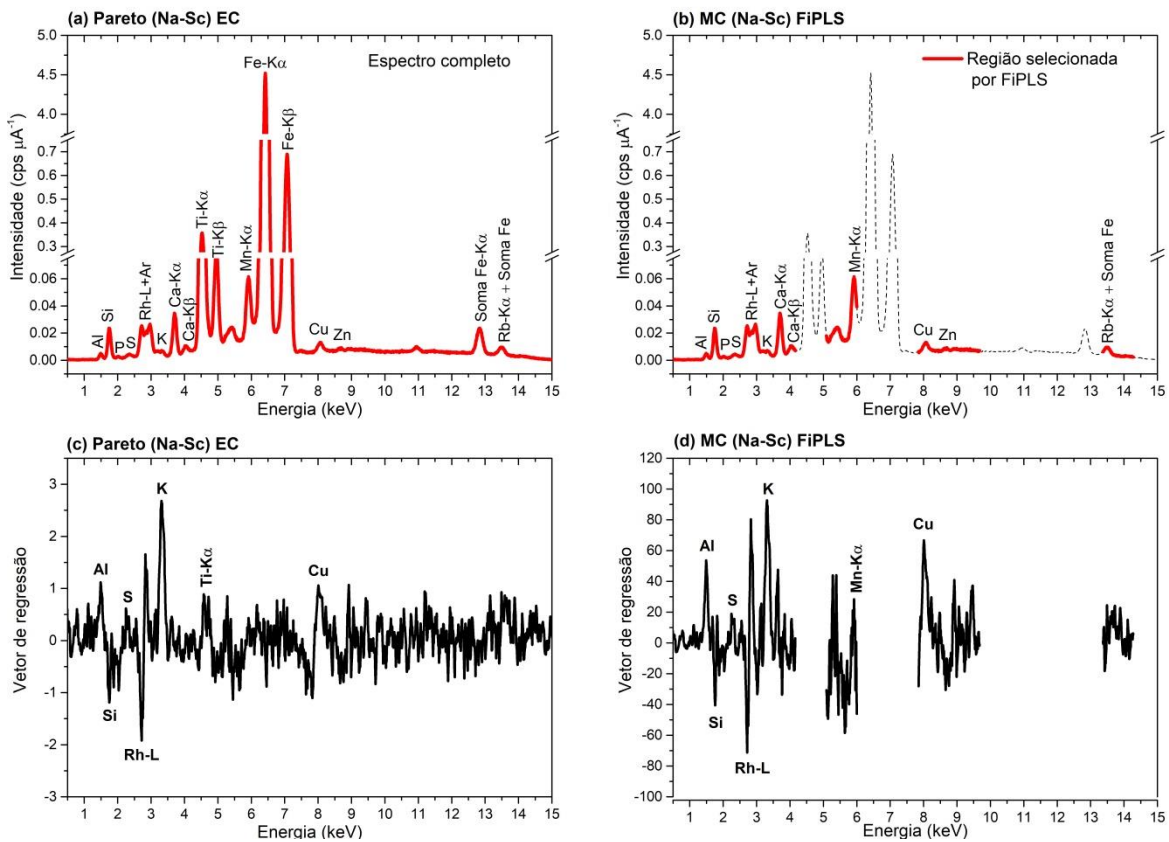
Tabela 25 - Teste de randomização para comparação dos melhores modelos para o K^+ em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
MC (Na-Sc) full X	0,03	Houve diferença significativa na seleção de variáveis por FiPLS no espectro de Na-Sc com centragem na média.
MC (Na-Sc) FiPLS		
Pareto (Ti-U) full X	0,0005	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com escalamento de Pareto.
Pareto (Ti-U) iSPA		
MC (Ti-U) full X	0,003	Houve diferença significativa na seleção de variáveis por iPLS no espectro de Ti-U com centragem na média.
MC (Ti-U) iSPA-PLS		
Pareto (Na-Sc) full X	0,42	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
MC (Na-Sc) FiPLS		
Pareto (Ti-U) iSPA X	0,35	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
MC (Ti-U) iSPA		
Pareto (Na-Sc) full X	0,03	Houve diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U
Pareto (Ti-U) iSPA		

6.4.5.2 Interpretação espectral das variáveis selecionadas nos modelos de K^+

De acordo com o item anterior, todos os modelos com o espectro de Ti-U apresentaram desempenho ruim, sendo que somente alguns modelos com o espectro de Na-Sc apresentaram desempenho considerado fraco, mas que ainda podem ser utilizados para correlações. Dessa forma, foram avaliadas somente as variáveis do melhor modelo de cada pré-processamento (escala de Pareto e centrado na média) sobre a faixa espectral de Na-Sc. A Figura 33 apresenta as variáveis selecionadas para estes modelos com seus respectivos vetores de regressão. Percebe-se que para o modelo com espectro de Na-Sc centrado na média foram selecionadas as regiões com os picos de Al, Si, P, S, Rh-L, K, Ca, Mn, Rb- $K\alpha$ e soma do Fe.

Figura 33 - Variáveis selecionadas para os melhores modelos de K^+ (a) com escala de Pareto e (b) centrados na média com seus respectivos vetores de regressão (c) e (d).



Fonte: do autor.

De forma geral, os vetores de regressão para ambos os modelos

tiveram um comportamento semelhante, sendo que a principal contribuição para modelagem foi o próprio pico de K, seguido pelos picos de Al e Cu, que tiveram contribuição positiva, e Si e Rh-L, que contribuíram negativamente. A alta correlação entre os teores totais e disponíveis de K também foi observada por Andrade et al. (2020) e Tavares et al. (2020b). No entanto, nesses trabalhos os autores demonstram desempenho satisfatório para a previsão de K^+ , atingindo R^2 de previsão de 0,68 e 0,95, respectivamente.

6.4.5.3 Figuras de mérito dos modelos para K^+

A Tabela 26 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia dos melhores modelos para K^+ . Os resultados do teste DW para linearidade e o teste t para *bias* demonstram que os modelos não apresentam autocorrelação entre os resíduos nem erros sistemáticos. O erro relativo do modelo Pareto (Na-Sc) EC foi de 34% e do modelo MC (Na-Sc) FiPLS foi de 35%. A Figura 34 mostra que os erros relativos individuais para esses modelos foram muito grandes, com a maioria das amostras de previsão apresentando erro relativo superior a $\pm 20\%$, ultrapassando, para algumas amostras, um erro relativo de 200%. Portanto, esses resultados indicam a falta de veracidade desses modelos para K^+ .

Além de não serem acurados, os resultados do CV para repetitividade indicaram que os modelos não foram precisos. Os piores CVs foram no nível baixo de concentração ($0,05 \text{ cmol}_c \text{ kg}^{-1}$), cujo modelo Pareto (Na-Sc) EC apresentou CV de 161% e o modelo MC (Na-Sc) apresentou CV de 312%. Para os demais níveis de concentração (médio ($0,25 \text{ cmol}_c \text{ kg}^{-1}$) e alto ($0,52 \text{ cmol}_c \text{ kg}^{-1}$)), os CVs variaram de 16 a 30%. Além disso, de acordo com o teste de randomização, esses modelos não foram robustos com relação à variação de granulometria.

Esses resultados ruins de veracidade, precisão e robustez podem ser devidos aos limites de quantificação dos modelos, que são superiores à faixa de trabalho de $0,05 \text{ cmol}_c \text{ kg}^{-1}$ a $0,75 \text{ cmol}_c \text{ kg}^{-1}$.

Tabela 26 - Figuras de mérito dos melhores modelos para K^+ .

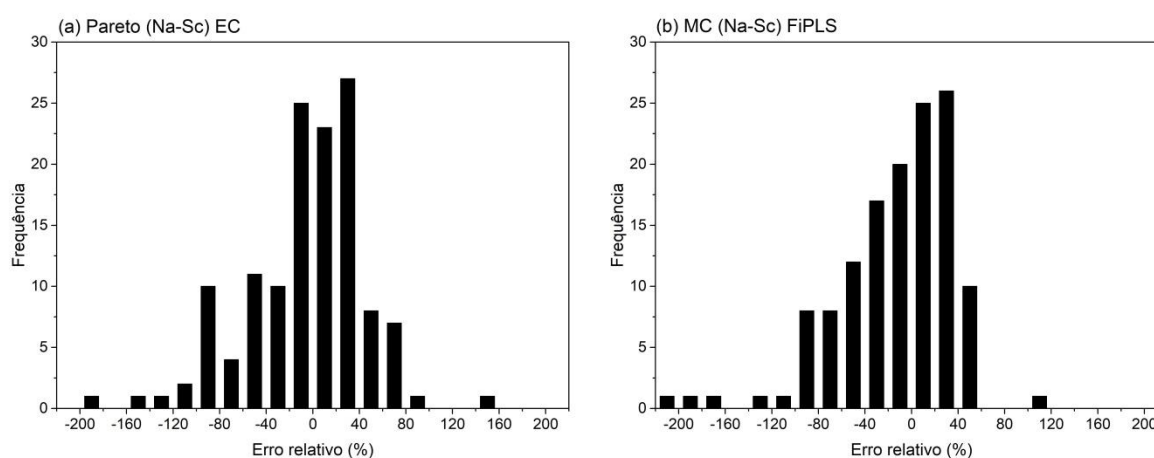
	Pareto (Na-Sc) EC	MC (Na-Sc) FiPLS
Linearidade (p-valor DW)	0,66	0,58
SEN^a	0,05	0,002
γ^{-1} (cmol_c kg⁻¹)	0,1	0,03
 Bias (t)^b	0,001 (0,16)	0,002 (0,43)
ER(%)	34	35
CV(%) repetitividade^c	161/22/16	312/28/30
Robustez (p-valor)^d	0,04	0,02
LOD (cmol_c kg⁻¹)	0,4	0,12
LOQ (cmol_c kg⁻¹)	1,4	0,35

^a Valor expresso como a razão entre intensidade e cmol_c kg⁻¹

^b $t_{critico} = 1,98$

^c Valores em nível baixo (0,05 cmol_c kg⁻¹), médio (0,25 cmol_c kg⁻¹) e alto (0,52 cmol_c kg⁻¹).

^d p-valor do teste de randomização em 95% de confiança.

Figura 34 - Erros relativos para o conjunto de previsão dos melhores modelos para K^+ .

Fonte: do autor.

Esses resultados mostram que o uso de dados espectrais de EDXRF não foi adequado para modelagem do K^+ nesse conjunto de dados, sendo inexato e impreciso. De forma geral, vários autores demonstram a viabilidade do uso de dados tanto de XRF quanto de Vis-NIR para a previsão de K^+ , com valores de RPD variando de 1,8 a 4,3 (ANDRADE et al., 2020; DEMATTÊ et al., 2019a; TAVARES et al., 2020a, 2020b). No entanto, nem todos os estudos demonstram isso, sendo que, em alguns casos, são reportados valores de RPD variando de 1,0 a

1,6 (MUNNAF; NAWAR; MOUAZEN, 2019; O'ROURKE et al., 2016; SILVA et al., 2017). Dessa forma, mais estudos devem ser realizados para verificar a viabilidade do uso de dados espectrais de EDXRF para determinação de K^+ em solos agrícolas.

6.4.6 Modelos para SB

6.4.6.1 Pré-processamento e seleção de variáveis dos modelos de SB

Os resultados para a calibração e previsão dos modelos de SB com os métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 27.

Nenhuma amostra foi identificada como *outlier* nos modelos utilizando a faixa espectral de Na-Sc. No entanto, para a faixa espectral de Ti-U quatro amostras do conjunto de calibração foram identificadas como *outliers*, as quais foram excluídas para a construção dos modelos.

Observa-se, na Tabela 27, que os modelos utilizando escalamento de Pareto apresentaram valores de RMSEP menor do que os modelos com os dados centrados na média, sendo essa diferença mais evidente para os modelos com o espectro completo. Utilizando o escalamento de Pareto, os valores de RMSEP para o modelo com espectro completo de Na-Sc e Ti-U foram reduzidos em 16% (de 0,99 para 0,83 $\text{cmol}_c \text{kg}^{-1}$) e 10% (de 0,99 para 0,89 $\text{cmol}_c \text{kg}^{-1}$), respectivamente. O teste de randomização demonstrou que essas diferenças foram significativas em um nível de confiança de 95%, com p-valor de 0,001 para o modelo com espectro completo de Na-Sc e p-valor = 0,02 para o modelo com espectro completo de Ti-U.

Tabela 27 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de SB com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	644	1012	368	276
	LV	6	3	4	4	5	5
	R²_{cal}	0,79	0,67	0,74	0,74	0,74	0,72
	RMSEC	0,87	1,08	0,96	0,95	0,96	0,98
	R²_{cv}	0,71	0,66	0,71	0,71	0,71	0,69
	RMSECV	1,04	1,12	1,04	1,03	1,04	1,04
	R²_{pred}	0,79	0,76	0,79	0,81	0,79	0,79
	RMSEP	0,83	0,89	0,83	0,78	0,85	0,83
	RPD	2,2	2,1	2,2	2,3	2,1	2,2
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	644	644	368	368
	LV	4	3	5	5	5	5
	R²_{cal}	0,62	0,66	0,74	0,74	0,72	0,72
	RMSEC	1,17	1,12	0,98	0,98	1,00	1,00
	R²_{cv}	0,59	0,62	0,69	0,69	0,69	0,69
	RMSECV	1,21	1,17	1,06	1,06	1,07	1,07
	R²_{pred}	0,71	0,76	0,79	0,79	0,79	0,79
	RMSEP	0,99	0,90	0,85	0,85	0,84	0,84
	RPD	1,8	2,0	2,1	2,1	2,2	2,2
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	635	1143	508	1016
	LV	6	5	6	6	6	6
	R²_{cal}	0,76	0,66	0,81	0,77	0,81	0,77
	RMSEC	0,95	1,09	0,81	0,80	0,83	0,91
	R²_{cv}	0,67	0,64	0,76	0,72	0,76	0,72
	RMSECV	1,10	1,13	0,95	1,00	0,94	1,00
	R²_{pred}	0,76	0,72	0,79	0,79	0,79	0,81
	RMSEP	0,89	0,97	0,85	0,83	0,85	0,81
	RPD	2,0	1,9	2,1	2,2	2,1	2,3
Faixa Ti-U centrado na média	nº Variáveis	1270	127	635	635	381	762
	LV	6	5	4	4	4	4
	R²_{cal}	0,66	0,66	0,69	0,69	0,69	0,19
	RMSEC	1,11	1,11	1,05	1,05	1,05	1,03
	R²_{cv}	0,59	0,62	0,67	0,67	0,67	0,67
	RMSECV	1,21	1,16	1,09	1,09	1,09	1,08
	R²_{pred}	0,71	0,71	0,76	0,76	0,74	0,76
	RMSEP	0,99	0,97	0,90	0,90	0,93	0,89
	RPD	1,8	1,9	2,0	2,0	2,0	2,0

*EC=Espectro completo; Valores de RMSE em $\text{cmol}_c \text{ kg}^{-1}$

Em geral, a seleção de variáveis melhorou a capacidade preditiva dos modelos, sendo esta melhora muito mais evidente para os dados centrados na média. Para o espectro de Na-Sc com escalamento de Pareto, a seleção de variáveis por BiPLS reduziu o RMSEP em 6% (de 0,83 para 0,78 $\text{cmol}_c \text{kg}^{-1}$). No caso do espectro de Na-Sc centrado na média, essa redução foi de 15% (de 0,99 para 0,84 $\text{cmol}_c \text{kg}^{-1}$) utilizando os métodos de SiPLS e iSPA-PLS, os quais selecionaram as mesmas variáveis. Para o espectro de Ti-U, a seleção de variáveis por iSPA-PLS reduziu o RMSEP em 9% (de 0,89 para 0,81 $\text{cmol}_c \text{kg}^{-1}$) e em 10% (de 0,99 para 0,89 $\text{cmol}_c \text{kg}^{-1}$) com relação ao espectro completo na escala de Pareto e centrado na média, respectivamente. De acordo com o teste de randomização (Tabela 28), esses métodos de seleção de variáveis apresentaram melhora significativa com relação aos modelos com espectro completo.

Tabela 28 - Teste de randomização para comparação dos melhores modelos para o SB em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X	0,01	Houve diferença significativa na seleção de variáveis por BiPLS no espectro de Na-Sc com escalamento de Pareto.
Pareto (Na-Sc) BiPLS		
MC (Na-Sc) EC X	0,005	Houve diferença significativa na seleção de variáveis por SiPLS no espectro de Na-Sc com centragem na média.
MC (Na-Sc) SiPLS		
Pareto (Ti-U) EC X	0,009	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com escalamento de Pareto.
Pareto (Ti-U) iSPA		
MC (Ti-U) EC X	0,005	Houve diferença significativa na seleção de variáveis por iSPA no espectro de Ti-U com centragem na média.
MC (Ti-U) iSPA		
Pareto (Na-Sc) BiPLS X	0,001	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
MC (Na-Sc) SiPLS		
Pareto (Ti-U) iSPA X	0,002	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
MC (Ti-U) iSPA		
Pareto (Na-Sc) BiPLS X	0,25	Não houve diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U
Pareto (Ti-U) iSPA		

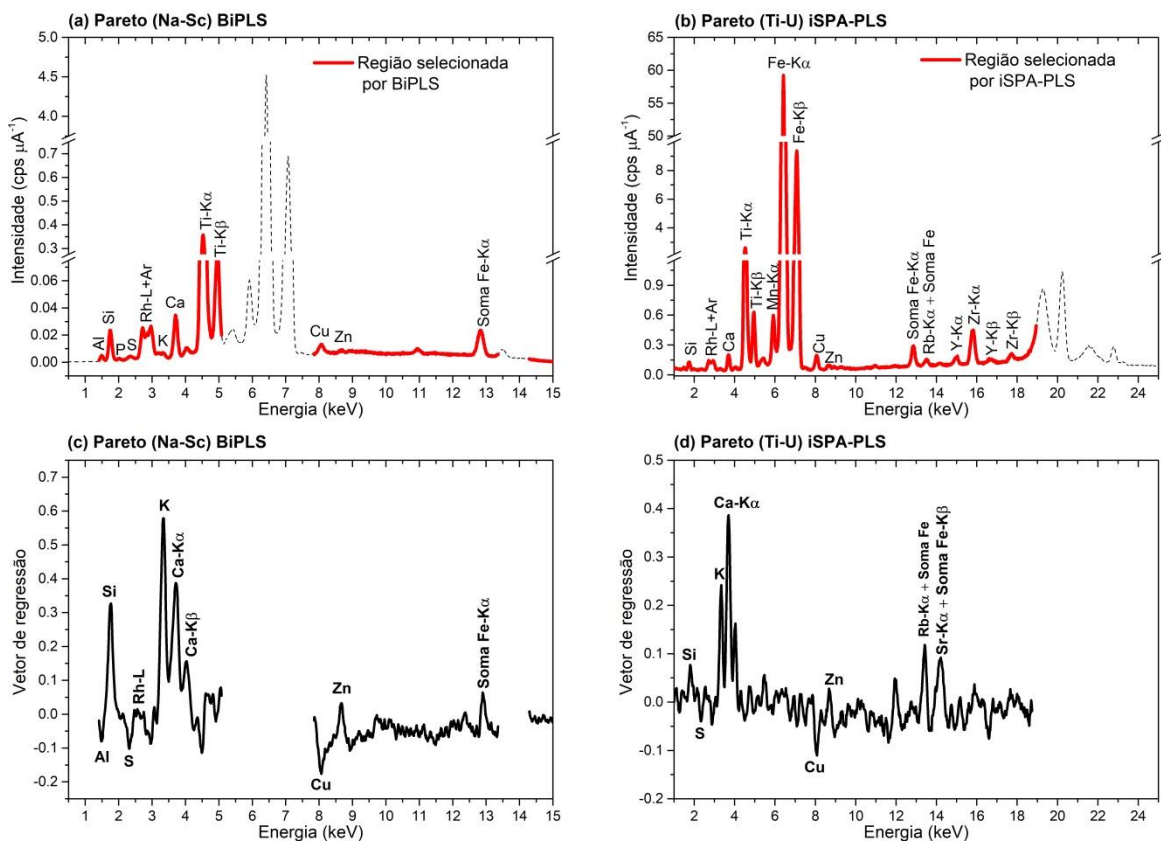
Ao comparar a combinação do pré-processamento com a seleção de variáveis, observa-se que os modelos com escalamento de Pareto (Pareto (Na-Sc) BiPLS e Pareto (Ti-U) iSPA) foram mais acurados do que os modelos com os dados

centrados na média (MC (Na-Sc) SiPLS e MC (Ti-U) iSPA). Isso também pode ser atestado pelo valor de RPD, que demonstra que apenas esses modelos apresentaram $RPD = 2,3$. Entretanto, o teste de randomização (Tabela 28) não mostrou diferença significativa entre os melhores modelos de cada condição de medida (Na-Sc e Ti-U).

6.4.6.2 Interpretação espectral das variáveis selecionadas nos modelos de SB

De acordo com o item anterior, os modelos que apresentaram o melhor desempenho para SB foram com seleção de variáveis por BiPLS sobre o espectro de Na-Sc com escalamento de Pareto (Pareto (Na-Sc) BiPLS) e com seleção de variáveis por iSPA-PLS sobre o espectro de Ti-U com escalamento de Pareto (Pareto (Ti-U) iSPA). A Figura 35 apresenta as variáveis selecionadas para estes modelos com seus respectivos vetores de regressão.

Figura 35 - Variáveis selecionadas para os melhores modelos de SB (a) com espectro de Na-Sc e (b) espectro de Ti-U com seus respectivos vetores de regressão (c) e (d).



Fonte: do autor.

Para ambos os modelos foram selecionadas regiões ao longo de todo o espectro. Novamente, os vetores de regressão de ambos os modelos tiveram um comportamento semelhante, a principal diferença foi no peso dos picos de K e Ca. Para os modelos com o espectro de Na-Sc o pico K-K α teve uma contribuição maior do que o pico Ca-K α . Por outro lado, para os modelos com o espectro de Ti-U o pico K-K α teve uma contribuição menor do que o pico Ca-K α . Além disso, para os modelos com o espectro de Na-Sc o peso dos picos de Si e Rh-L foram maiores do que para os modelos com o espectro de Ti-U. Provavelmente, essa diferença entre os pesos desses picos nos vetores de regressão se deve ao fato da condição de medida de Na-Sc (15 kV aplicado no tubo) ser mais adequada para excitação dessa região inicial do espectro.

A contribuição positiva do K e Ca nos modelos de SB faz sentido, já que as principais bases trocáveis do solo são Ca²⁺, K⁺ e Mg²⁺. Além disso, percebe-se que os vetores de regressão dos modelos para SB (Figura 35 (c) e (d)) foram muito semelhantes aos vetores de regressão dos modelos para Ca²⁺ (Figura 28), que é a principal base trocável do solo. Em um estudo realizado por Teixeira et al. (2020) em solos tropicais, também é apontado que o Ca e K foram as variáveis que mais contribuíram para a modelagem da SB. É importante ressaltar que o Al e o Cu possuem uma relação antagônica com o Ca na troca de cargas do solo. O Al é mais solúvel em solos com baixo pH. Nessa condição, a SB é baixa e o Al compõe uma proporção maior da CTC. No entanto, quando os cátions básicos, como Ca (CaCO₃), são adicionados no solo, o Al é precipitado na forma Al(OH)₃. Assim, o Al é deslocado da CTC aumentando a SB (SPOSITO, 2008). Este processo explica o efeito negativo do Al no modelo SB. O Cu é um outro micronutriente que desempenha um papel importante na dinâmica de troca de cargas do solo. O teor de Cu no solo é inversamente correlacionado com a SB, sendo que o teor de Cu trocável diminui com o aumento do SB e do pH (SPOSITO, 2008). Isso também pode ser observado nos modelos para SB, sendo que a região com o pico de Cu teve uma contribuição negativa na modelagem desse parâmetro.

6.4.6.3 Figuras de mérito dos modelos para SB

Para verificar a eficácia dos modelos para SB foram avaliadas

algumas figuras de mérito, que são mostradas na Tabela 29. Os resultados do teste DW para linearidade e o teste t para *bias* indicaram que os modelos são lineares e estão ausentes de erros sistemáticos. O erro relativo de ambos os modelos foi de 10%. Além disso, a Figura 36 mostra que poucas amostras apresentaram erros relativos individuais superiores a $\pm 20\%$. No entanto, observa-se que os erros relativos individuais do modelo Pareto (Na-Sc) BiPLS foram menores do que os do modelo Pareto (Ti-U) iSPA-PLS. A precisão por repetitividade dos modelos foi avaliada em três diferentes níveis de concentração, baixo ($4,6 \text{ cmol}_c \text{ kg}^{-1}$), médio ($8,2 \text{ cmol}_c \text{ kg}^{-1}$) e alto ($11,6 \text{ cmol}_c \text{ kg}^{-1}$). Os CVs das medidas em triplicata ficaram abaixo de 5% para o modelo Pareto (Na-Sc) BiPLS e abaixo de 9% para o modelo Pareto (Ti-U) iSPA-PLS. Além disso, o teste de randomização para robustez demonstrou que ambos os modelos foram robustos quanto à variação de granulometria de 2 mm para $125 \mu\text{m}$ do conjunto de previsão. Portanto, esses resultados demonstram que ambos os modelos apresentaram veracidade, precisão e robustez adequadas, sendo que o modelo Pareto (Na-Sc) BiPLS apresentou um desempenho um pouco melhor dessas figuras de mérito. Além disso, os LOD e LOQ do modelo Pareto (Na-Sc) BiPLS foram menores do que para o modelo Pareto (Ti-U) iSPA-PLS.

Tabela 29 - Figuras de mérito dos melhores modelos para SB.

	Pareto (Na-Sc) BiPLS	Pareto (Ti-U) iSPA-PLS
Linearidade (p-valor DW)	0,93	0,86
SEN^a	0,28	0,60
γ^{-1} ($\text{cmol}_c \text{ kg}^{-1}$)	0,03	0,1
 <i>Bias</i> (t)^b	0,07 (0,85)	0,07 (1,02)
ER(%)	10	10
CV(%) repetitividade^c	4/5/5	9/4/7
Robustez (p-valor)^d	0,45	0,35
LOD ($\text{cmol}_c \text{ kg}^{-1}$)	0,09	0,3
LOQ ($\text{cmol}_c \text{ kg}^{-1}$)	0,27	1,0

^a Valor expresso como a razão entre intensidade e $\text{cmol}_c \text{ kg}^{-1}$

^b $t_{\text{critico}} = 1,98$

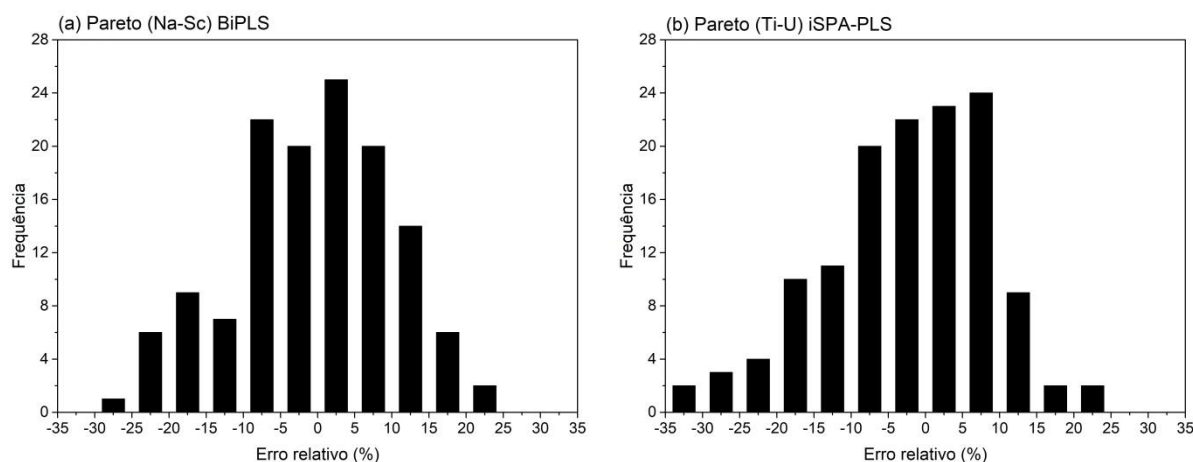
^c Valores em nível baixo ($4,6 \text{ cmol}_c \text{ kg}^{-1}$), médio ($8,2 \text{ cmol}_c \text{ kg}^{-1}$) e alto ($11,6 \text{ cmol}_c \text{ kg}^{-1}$).

^d p-valor do teste de randomização em 95% de confiança.

De forma geral, esses resultados demonstram a viabilidade do uso

de dados espectrais de EDXRF para previsão da SB nesse conjunto de dados. Esses resultados foram equivalentes aos relatados na literatura ($1,4 > RPD < 2,6$), com dados de concentração de XRF e dados espectrais de Vis-NIR (DEMATTÊ et al., 2019b, 2019a; TEIXEIRA et al., 2020; PINHEIRO et al., 2017).

Figura 36 - Erros relativos para o conjunto de previsão dos melhores modelos para SB.



Fonte: do autor.

6.4.7 Modelos para CTC

6.4.7.1 Pré-processamento e seleção de variáveis dos modelos de CTC

Para construção dos modelos na faixa espectral de Ti-U, utilizou-se 260 amostras para calibração e 132 para previsão, sendo que quatro amostras do conjunto de calibração foram excluídas, as quais foram identificadas como *outliers*. Por outro lado, nenhuma amostra foi identificada como *outlier* nos modelos utilizando faixa espectral de Na-Sc.

Os resultados para a calibração e previsão dos modelos de CTC com os métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 30. Observa-se que o uso do escalamento de Pareto gerou uma pequena melhora no desempenho dos modelos com espectro completo, reduzindo os valores de RMSEP em 7% (de 0,91 para 0,85 $\text{cmol}_c \text{kg}^{-1}$) para o espectro de Na-Sc e 3% (de 0,91 para 0,88 $\text{cmol}_c \text{kg}^{-1}$) para o espectro de Ti-U.

Tabela 30 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de CTC com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	368	644	368	736
	LV	6	2	4	5	4	6
	R²_{cal}	0,74	0,58	0,71	0,71	0,69	0,71
	RMSEC	0,80	1,01	0,85	0,83	0,87	0,84
	R²_{cv}	0,64	0,55	0,66	0,64	0,67	0,64
	RMSECV	0,94	1,05	0,91	0,93	0,90	0,93
	R²_{pred}	0,67	0,64	0,67	0,67	0,67	0,69
	RMSEP	0,85	0,88	0,84	0,84	0,85	0,83
	RPD	1,7	1,7	1,7	1,7	1,7	1,8
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	552	552	368	276
	LV	4	3	4	4	4	5
	R²_{cal}	0,58	0,61	0,71	0,71	0,71	0,67
	RMSEC	1,01	0,97	0,85	0,86	0,86	0,89
	R²_{cv}	0,55	0,53	0,66	0,66	0,66	0,62
	RMSECV	1,04	1,06	0,91	0,91	0,91	0,96
	R²_{pred}	0,61	0,61	0,67	0,69	0,69	0,66
	RMSEP	0,91	0,91	0,84	0,82	0,82	0,85
	RPD	1,6	1,6	1,8	1,8	1,8	1,7
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	762	762	508	635
	LV	4	4	4	4	5	5
	R²_{cal}	0,66	0,58	0,64	0,64	0,69	0,76
	RMSEC	0,93	1,01	0,93	0,93	0,87	0,78
	R²_{cv}	0,59	0,55	0,61	0,61	0,61	0,61
	RMSECV	1,00	1,06	0,98	0,98	0,98	0,99
	R²_{pred}	0,64	0,59	0,64	0,64	0,66	0,66
	RMSEP	0,88	0,93	0,88	0,88	0,87	0,86
	RPD	1,7	1,6	1,7	1,7	1,7	1,7
Faixa Ti-U centrado na média	nº Variáveis	1270	127	635	635	508	635
	LV	6	5	4	4	4	4
	R²_{cal}	0,64	0,56	0,62	0,62	0,61	0,61
	RMSEC	0,92	1,02	0,96	0,96	0,96	0,97
	R²_{cv}	0,58	0,53	0,59	0,59	0,58	0,58
	RMSECV	1,01	1,06	0,99	0,99	1,00	1,01
	R²_{pred}	0,61	0,59	0,62	0,62	0,61	0,61
	RMSEP	0,91	0,94	0,90	0,90	0,91	0,92
	RPD	1,6	1,6	1,6	1,6	1,6	1,6

*EC=Espectro completo; Valores de RMSE em $\text{cmol}_c \text{ kg}^{-1}$

Em geral, os métodos de seleção de variáveis também

apresentaram um desempenho sutilmente melhor com relação aos modelos com espectro completo. No entanto, a seleção de variáveis só apresentou melhora significativa para o espectro de Na-Sc centrado na média, sendo que o método de SiPLS reduziu em 10 % o valor de RMSEP (de 0,91 para 0,82 $\text{cmol}_c \text{kg}^{-1}$).

Tabela 31 - Teste de randomização para comparação dos melhores modelos para o CTC em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X	0,10	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com escalamento de Pareto.
Pareto (Na-Sc) iSPA		
MC (Na-Sc) EC X	0,005	Houve diferença significativa na seleção de variáveis por SiPLS no espectro de Na-Sc com centragem na média.
MC (Na-Sc) SiPLS		
Pareto (Ti-U) EC X	0,24	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com escalamento de Pareto.
Pareto (Ti-U) iSPA		
MC (Ti-U) EC X	0,25	Não houve diferença significativa na seleção de variáveis por FiPLS no espectro de Ti-U com centragem na média.
MC (Ti-U) FiPLS		
Pareto (Na-Sc) iSPA X	0,27	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
MC (Na-Sc) SiPLS		
Pareto (Ti-U) iSPA X	0,12	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
MC (Ti-U) FiPLS		
MC (Na-Sc) SiPLS X	0,07	Não houve diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U
Pareto (Ti-U) iSPA		

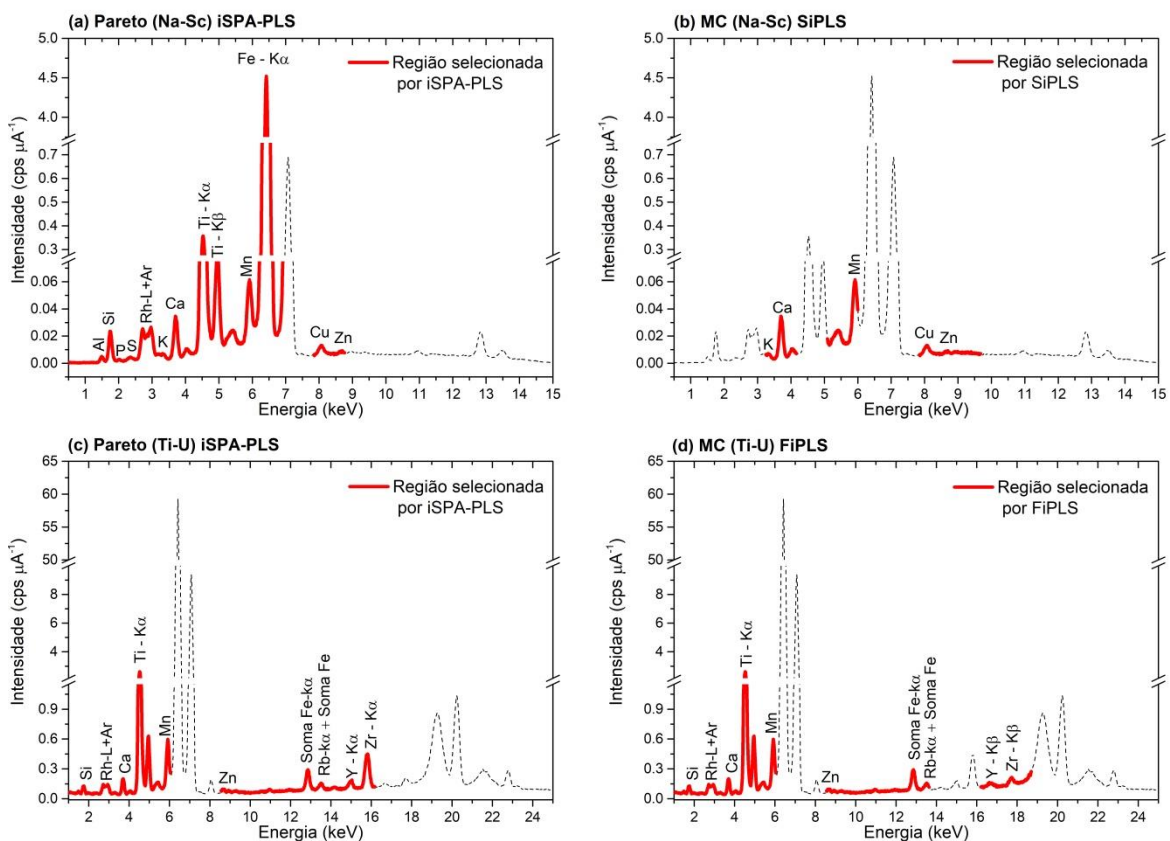
Os resultados do teste de randomização (Tabela 31) demonstram que não houve diferença significativa no desempenho dos modelos de acordo com o pré-processamento aplicado (escalamento de Pareto e centragem na média). Além disso, de acordo com o teste de randomização, não houve diferença significativa entre os melhores modelos de cada condição de medida (Na-Sc e Ti-U). Entretanto, somente alguns modelos com espectro de Na-Sc apresentaram desempenho considerado razoável ($2,0 > \text{RPD} > 1,8$), com $\text{RPD} = 1,8$. Estudos na literatura utilizando dados de XRF, Vis-NIR e fusão de dados Vis-NIR+XRF apresentam valores de R^2 variando de 0,38 a 0,94 e valores de RPD variando de 0,9 a 2,8 (DEMATTÊ et al., 2019a, 2019b; TEIXEIRA et al., 2020; O'ROURKE et al., 2016; PINHEIRO et al., 2017; RAWAL et al., 2019; SHARMA et al., 2015; SILVA et al.,

2017; TAVARES et al., 2020a, 2020b; WAN et al., 2020). De forma geral, a maior parte desses estudos demonstra desempenho de bom a excelente para a previsão da CTC.

6.4.7.2 Interpretação espectral das variáveis selecionadas nos modelos de CTC

As variáveis selecionadas para os melhores modelos de cada pré-processamento são mostradas na Figura 37 e os vetores de regressão para os respectivos modelos são mostrados na Figura 38.

Figura 37 - Intervalos de variáveis selecionadas para os melhores modelos de CTC com escalamento de Pareto e centragem na média.

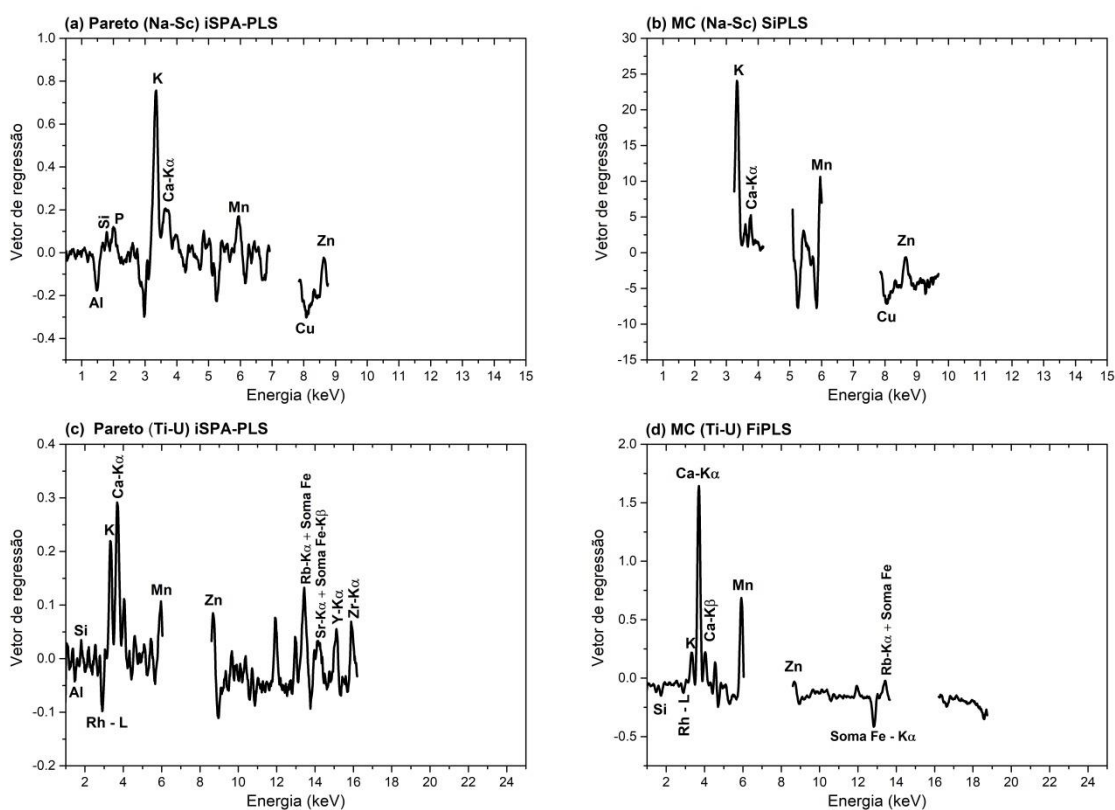


Fonte: do autor.

Para o espectro de Na-Sc na escala de Pareto a seleção de variáveis por iSPA-PLS selecionou as regiões de 0,5 a 6,9 keV e de 7,9 a 8,8 keV, que contém os picos de Al, Si, P, Rh-L, K, Ca, Ti, Mn, Fe-K α , Cu e Zn. Por outro lado, para os espectros de Na-Sc centrados na média, o melhor modelo foi com a

seleção de variáveis por SiPLS, o qual selecionou somente as regiões com os picos de K, Ca, Mn-K α , Cu e Zn. Para os modelos com o espectro de Ti-U, observa-se (Figura 37) que as regiões selecionadas para o modelo na escala de Pareto foram semelhantes às regiões selecionadas com os dados centrados na média. Nesse caso, foram selecionadas as regiões com os picos de Al, Si, P, Rh-L, K, Ca, Ti, Mn, Zn, Rb, Y, Zr e picos soma do Fe.

Figura 38 - Vetor de regressão dos melhores modelos PLSR para CTC com escalamento de Pareto e centragem na média.



Fonte: do autor.

Os vetores de regressão (Figura 38) dos melhores modelos (de cada condição de medida e pré-processamento) para a CTC também tiveram um comportamento semelhante, sendo que os picos de K, Ca, Mn e Zn contribuíram positivamente e os picos de Al e Cu contribuíram negativamente para a modelagem da CTC. Assim como para os modelos de SB, a principal diferença observada entre os vetores de regressão foi o peso dos picos de K e Ca, os quais para os modelos com o espectro de Na-Sc o pico K-K α teve uma contribuição maior do que o pico

Ca-K α . Por outro lado, para os modelos na faixa de Ti-U o pico K-K α teve uma contribuição menor do que o pico Ca-K α . Nota-se também que o peso dos picos com $Z < 18$ foi maior para os modelos com o espectro de Na-Sc. Como já discutido anteriormente, isso pode ser devido à energia de excitação da condição de medida de Na-Sc ser mais adequada para essa região inicial do espectro.

As duas principais fontes de carga para CTC em solos tropicais e subtropicais são decorrentes do conteúdo de carbono orgânico e coloides inorgânicos. Assim, a CTC é dependente do conteúdo e qualidade do carbono orgânico, mineralogia e pH do solo.

A CTC proveniente de coloides inorgânicos é predominante de minerais 2:1. Em geral, o hidróxido de Al neutraliza a CTC destes minerais, ocupando os locais de troca das suas camadas internas, o que desloca os cátions trocáveis básicos (Ca, Mg e K) para a solução do solo (COSTA et al., 2004). Esses processos explicam, em partes, as relações negativas entre o Al e os modelos de CTC. Por outro lado, o Si pode aumentar a CTC do solo, uma vez que a gibbsita (Al(OH)₃) geralmente pode ser redissolvida e combinada com o ácido silícico para formar uma nova caulinita, cuja capacidade de troca é maior que da gibbsita (COSTA et al., 2004). No entanto, no processo de intemperismo, durante a formação de Latossolos, quase todos os minerais são removidos da fração de areia e silte. Por outro lado, na fração de argila esse intemperismo causa a dominância de caulinita, gibbsita e minerais de Fe, resultando em baixos valores de CTC. Tipicamente, nesse processo, ocorre a remoção de sílica (dessilicatização) dos materiais primários, levando a um acúmulo de óxidos e hidróxidos de Fe, Mn e Al (OSMAN, 2013).

Além disso, o conteúdo de carbono orgânico também influencia na CTC do solo, principalmente em áreas manejadas com plantio direto (como no presente estudo). Ramos et al. (2018) observaram que a CTC aumentou 25% a cada 1,8 Kg m⁻² de carbono orgânico estocado. Por outro lado, a redução do carbono orgânico diminuiu a CTC de 8,40 cmol_c kg⁻¹ para 4,82 cmol_c kg⁻¹ (RAMOS et al., 2018). Os autores observaram também que o estoque de carbono apresentou uma correlação positiva com a SB e atribuiu essa relação aos grupos carboxílico (-COOH) e fenólico (-OH) (RAMOS et al., 2018). Em Latossolos, esses grupos orgânicos interagem com óxidos de ferro e alumínio, formando micro agregados (diâmetro menor que 0,25 mm) que estão fortemente associados aos teores de K, Ca e Mg no solo (SCHAEFER; FABRIS; KER, 2008).

Toda essa dinâmica da CTC e a sua correlação com elementos presentes no solo é complexa. No entanto, observam-se as semelhanças entre os vetores de regressão dos modelos de CTC e SB com os vetores de regressão dos modelos de Ca^{2+} . Essas semelhanças entre os vetores de regressão, bem como as altas correlações entre esses atributos (Figura 16), sugerem uma influência do manejo sobre as variáveis que contribuem para modelagem desses parâmetros.

6.4.7.3 Figuras de mérito dos modelos para CTC

A Tabela 32 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia dos modelos para CTC. De acordo com essas figuras de mérito, em um nível de confiança de 95%, nenhum modelo apresentou *bias* e autocorrelação entre os resíduos. Observa-se que o erro relativo dos modelos foi baixo, sendo 6% para os modelos Pareto (Na-Sc) BiPLS, MC (Na-Sc) SiPLS e Pareto (Ti-U) iSPA-PLS e 7% para o modelo MC (Ti-U) FiPLS. Além disso, observa-se que os erros relativos individuais de todos os modelos (Figura 39) também foram baixos, sendo que praticamente todas as amostras tiveram erros relativos individuais menores que $\pm 20\%$. No entanto, de acordo com a Tabela 30, somente os modelos Pareto (Na-Sc) iSPA e MC (Na-Sc) SiPLS tiveram um desempenho considerado razoável ($2,0 > \text{RPD} > 1,8$), com RPD de 1,8.

A precisão dos modelos por repetitividade (medidas em triplicata) foi avaliada em um nível de concentração baixo ($10,1 \text{ cmol}_c \text{ kg}^{-1}$), médio ($13,6 \text{ cmol}_c \text{ kg}^{-1}$) e alto ($16,1 \text{ cmol}_c \text{ kg}^{-1}$). Para todos os modelos os CVs foram baixos, indicando que esses modelos são precisos. Os maiores CVs foram observados no nível baixo de concentração, os quais variaram de 6 a 2%. No entanto, somente o modelo Pareto (Na-Sc) iSPA-PLS apresentou robustez quanto à variação de granulometria.

Os limites de quantificação dos modelos variaram de 1,3 a $0,07 \text{ cmol}_c \text{ kg}^{-1}$. Observa-se que os LOD e LOQ dos modelos com escalamento de Pareto foram maiores do que dos modelos com os dados centrados na média. Isso pode ter sido causado pela intensificação do ruído experimental devido ao pré-processamento.

Tabela 32 - Figuras de mérito dos melhores modelos para CTC.

	Pareto (Na-Sc) iSPA-PLS	MC (Na-Sc) SiPLS	Pareto (Ti-U) iSPA-PLS	MC (Ti-U) FiPLS
Linearidade (p-valor DW)	0,60	0,75	0,57	0,57
SEN^a	0,22	0,009	0,46	0,17
γ^{-1} (cmol_c kg⁻¹)	0,03	0,007	0,1	0,02
 Bias (t)^b	0,004 (0,06)	0,02 (0,23)	0,04 (0,55)	0,08 (1,02)
ER(%)	6	6	6	7
CV(%) repetitividade^c	2/0,1/1	4/2/4	6/3/3	2/1/2
Robustez (p-valor)^d	0,33	0,03	0,001	0,01
LOD (cmol_c kg⁻¹)	0,11	0,02	0,4	0,07
LOQ (cmol_c kg⁻¹)	0,34	0,07	1,3	0,23

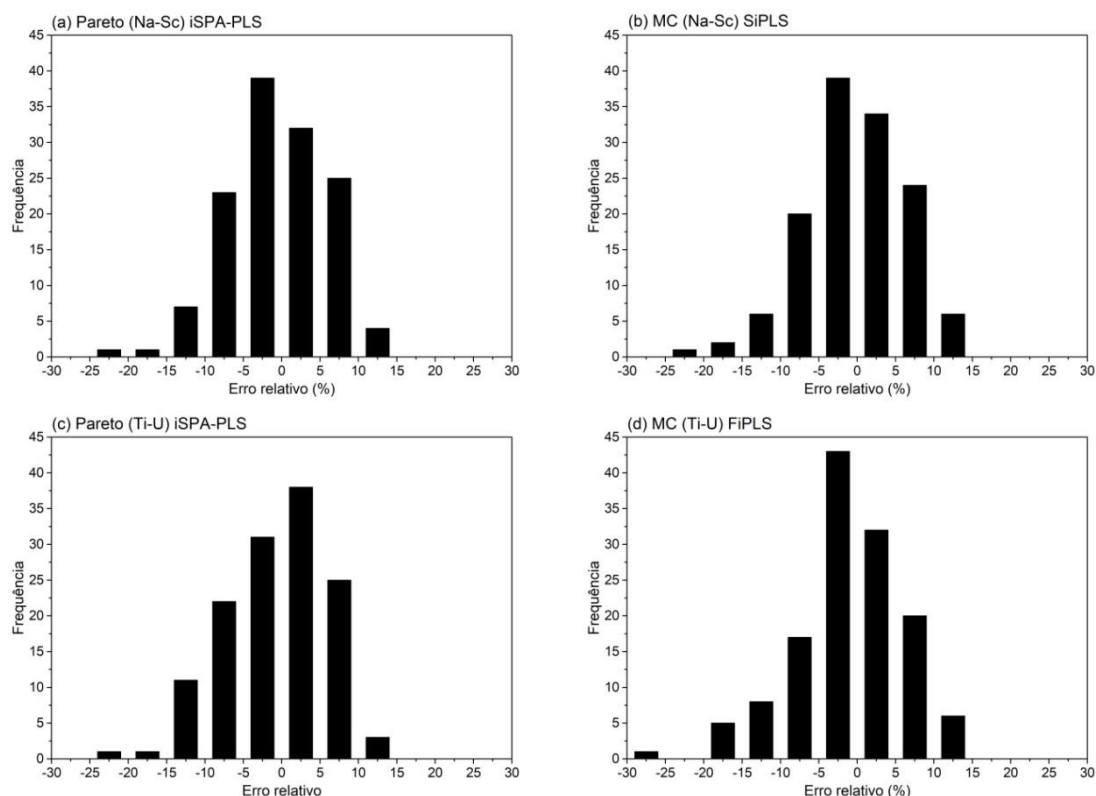
^a Valor expresso como a razão entre intensidade e cmol_c kg⁻¹

^b $t_{critico} = 1,98$

^c Valores em nível baixo (10,1 cmol_c kg⁻¹), médio (13,6 cmol_c kg⁻¹) e alto (16,1 cmol_c kg⁻¹).

^d p-valor do teste de randomização em 95% de confiança.

Figura 39 - Erros relativos para o conjunto de previsão dos melhores modelos para CTC.



Fonte: do autor.

6.4.8 Modelos para V%

6.4.8.1 Pré-processamento e seleção de variáveis dos modelos de V

Os resultados para a calibração e previsão dos modelos de V% com os métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 33. Para ambos os modelos foram excluídas 10 amostras de calibração e uma de previsão que foram identificadas como *outliers*.

Em geral, o uso do escalamento de Pareto e da seleção de variáveis melhorou o desempenho dos modelos. Para os modelos com espectro completo de Na-Sc, o uso do escalamento de Pareto reduziu em 13% o RMSEP (de 5,89 para 5,10 %) com relação aos dados centrados na média. De acordo com o teste de randomização (p-valor = 0,0003), essa melhora foi significativa em um nível de confiança de 95%. No entanto, para o espectro completo de Ti-U, o escalamento de Pareto aumentou o RMSEP em 2% (de 5,76 para 5,85 %), porém essa diferença não foi significativa (p-valor = 0,22). No caso da seleção de variáveis, o uso do iSPA-PLS melhorou o desempenho de todos os modelos (com diferentes pré-processamentos e condições de medida). Entretanto, essa melhora só foi significativa (Tabela 34) para o espectro de Na-Sc centrado na média e para o espectro de Ti-U com escalamento de Pareto.

Ao comparar a combinação do pré-processamento com a seleção de variáveis por iSPA-PLS (que se mostrou mais eficiente para esse conjunto de dados), observa-se que os espectros com escalamento de Pareto foram mais acurados do que com centragem na média, apresentando uma melhora significativa de acordo com o teste de randomização (Tabela 34). Isso também pode ser atestado pelo valor de RPD, que demonstra que apenas os modelos com escalamento de Pareto apresentaram $RPD > 1,8$. Em relação à condição de medida, o teste de randomização (Tabela 34) não mostrou diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U. Portanto, esses resultados sugerem que o escalamento de Pareto foi mais adequado para pré-processar esses dados espectrais para modelagem de V%.

Tabela 33 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de V% com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	184	920	368	1104
	LV	6	2	4	6	5	6
	R²_{cal}	0,74	0,64	0,74	0,74	0,74	0,74
	RMSEC	4,27	5,03	4,51	4,29	4,27	4,36
	R²_{cv}	0,64	0,62	0,64	0,66	0,66	0,64
	RMSECV	5,15	5,19	4,87	4,95	4,97	5,04
	R²_{pred}	0,71	0,62	0,71	0,71	0,67	0,71
	RMSEP	5,10	5,68	5,35	5,06	5,31	5,05
	RPD	1,9	1,7	1,8	1,9	1,8	1,9
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	460	460	368	368
	LV	4	2	5	5	5	5
	R²_{cal}	0,61	0,59	0,72	0,72	0,69	0,69
	RMSEC	5,34	5,39	4,54	4,54	4,70	4,73
	R²_{cv}	0,58	0,58	0,66	0,66	0,64	0,64
	RMSECV	5,52	5,49	5,05	5,05	5,10	5,10
	R²_{pred}	0,61	0,66	0,66	0,66	0,67	0,67
	RMSEP	5,89	5,55	5,45	5,45	5,31	5,29
	RPD	1,6	1,7	1,7	1,7	1,8	1,8
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	762	762	508	889
	LV	5	2	5	5	5	6
	R²_{cal}	0,67	0,56	0,77	0,77	0,72	0,77
	RMSEC	4,78	5,48	4,00	4,00	4,33	3,99
	R²_{cv}	0,58	0,55	0,61	0,61	0,59	0,59
	RMSECV	5,41	5,61	5,25	5,25	5,29	5,26
	R²_{pred}	0,61	0,61	0,64	0,64	0,66	0,69
	RMSEP	5,85	5,90	5,55	5,55	5,44	5,16
	RPD	1,6	1,6	1,7	1,7	1,8	1,9
Faixa Ti-U centrado na média	nº Variáveis	1270	127	635	889	508	508
	LV	7	3	3	3	3	3
	R²_{cal}	0,64	0,55	0,59	0,59	0,59	0,59
	RMSEC	5,22	5,93	5,59	5,55	5,55	5,58
	R²_{cv}	0,52	0,50	0,56	0,56	0,56	0,56
	RMSECV	6,11	6,14	5,79	5,77	5,75	5,77
	R²_{pred}	0,62	0,62	0,62	0,64	0,64	0,64
	RMSEP	5,76	5,93	5,69	5,64	5,64	5,63
	RPD	1,7	1,6	1,7	1,7	1,7	1,7

*EC=Espectro completo; Valores de RMSE em %

Tabela 34 - Teste de randomização para comparação dos melhores modelos para o V em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X	0,11	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com escalamento de Pareto.
Pareto (Na-Sc) iSPA		
MC (Na-Sc) EC X	0,001	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com centragem na média.
MC (Na-Sc) iSPA		
Pareto (Ti-U) EC X	0,001	Houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com escalamento de Pareto.
Pareto (Ti-U) iSPA		
MC (Ti-U) EC X	0,23	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com centragem na média.
MC (Ti-U) iSPA		
Pareto (Na-Sc) iSPA X	0,02	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
MC (Na-Sc) iSPA		
Pareto (Ti-U) iSPA X	0,009	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
MC (Ti-U) iSPA		
Pareto (Na-Sc) iSPA X	0,28	Não houve diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U
Pareto (Ti-U) iSPA		

6.4.8.2 Interpretação espectral das variáveis selecionadas nos modelos de V%

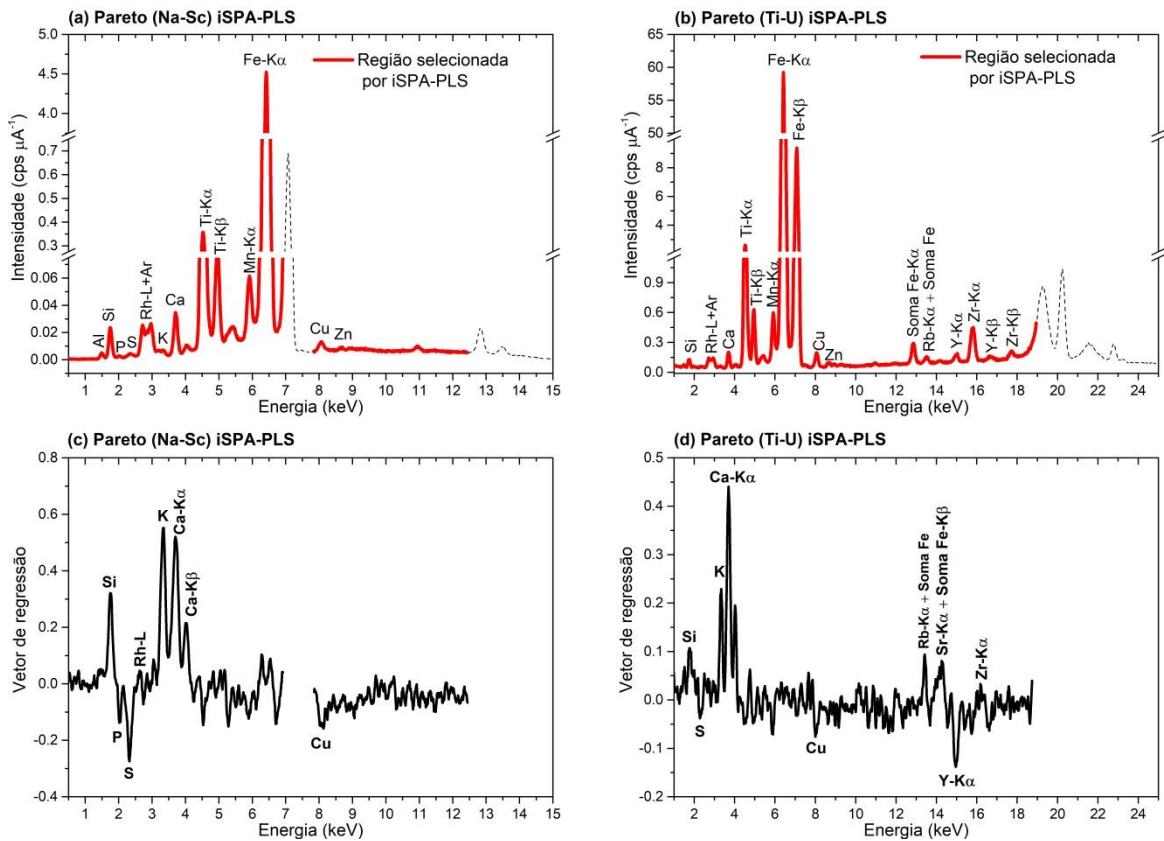
De acordo com o item anterior, os modelos que apresentaram o melhor desempenho para V% foram com seleção de variáveis por iSPA-PLS sobre os espectros de Na-Sc e Ti-U, ambos com escalamento de Pareto (Pareto (Na-Sc) iSPA-PLS) e (Pareto (Ti-U) iSPA-PLS). A Figura 40 apresenta as variáveis selecionadas para estes modelos com seus respectivos vetores de regressão.

Para ambos os modelos foram selecionadas regiões ao longo de todo o espectro, sendo essas regiões muito semelhantes às regiões selecionadas para os melhores modelos de SB. Isso faz sentido visto que V% é definido como a porcentagem da CTC do solo que é ocupado pelas bases trocáveis (SB).

Assim como para SB, os vetores de regressão dos melhores modelos para V% tiveram um comportamento semelhante, sendo que a principal diferença foi no peso dos picos de K e Ca. Para os modelos com os espectros de

Na-Sc, os picos $K\text{-K}\alpha$ e $\text{Ca-K}\alpha$ tiveram pesos semelhantes. Por outro lado, para os modelos com os espectros de Ti-U o pico $\text{Ca-K}\alpha$ teve uma contribuição maior do que o pico $K\text{-K}\alpha$. Além disso, para os modelos com os espectros de Na-Sc o peso dos picos de Si, P, S e Rh-L foram maiores do que para os modelos com os espectros de Ti-U. Como discutido anteriormente, essa diferença entre os pesos desses picos nos vetores de regressão pode ser devido ao fato da condição de medida de Na-Sc (15 kV aplicado no tubo) ser mais adequada para excitação dessa região inicial do espectro.

Figura 40 - Variáveis selecionadas para os melhores modelos de V% (a) com espectro de Na-Sc e (b) espectro de Ti-U com seus respectivos vetores de regressão (c) e (d).



Fonte: do autor.

Observa-se que as maiores contribuições para modelagem de V% foram dadas pelos picos de K e Ca, sendo estes as principais bases trocáveis do solo. Além disso, também se observa uma contribuição positiva dos picos de Si e Rh-L e negativa dos picos de P, S e Cu. Para o modelo com o espectro de Ti-U, além dessas contribuições descritas, nota-se que os picos soma do Fe, Rb- $K\alpha$ e Sr-

$K\alpha$ também contribuíram positivamente e o pico Y - $K\alpha$ teve uma influência negativa na modelagem de $V\%$.

6.4.8.3 Figuras de mérito dos modelos para V

Para verificar a eficácia dos modelos para $V\%$ foram avaliadas algumas figuras de mérito, que são mostradas na Tabela 35. Os resultados do teste DW indicaram linearidade e o teste t para *bias* demonstrou ausência de erros sistemáticos nos modelos. O erro relativo do modelo com a faixa espectral de Na-Sc foi de 8% e do modelo com a faixa espectral de Ti-U foi 9%. Além desses baixos valores de erro relativo, a Figura 41 mostra que poucas amostras, em ambos os modelos, apresentaram erros relativos individuais superiores a $\pm 20\%$. Nota-se que os erros relativos individuais do modelo Pareto (Na-Sc) iSPA-PLS foram um pouco menores do que os erros do modelo Pareto (Ti-U) iSPA-PLS, o qual apresentou, em alguns casos, erros relativos individuais superiores a 30%.

Tabela 35 - Figuras de mérito dos melhores modelos para $V\%$.

	Pareto (Na-Sc) iSPA-PLS	Pareto (Ti-U) iSPA-PLS
Linearidade (p-valor DW)	0,17	0,69
SEN^a	0,22	0,58
γ^{-1} (cmol_c kg⁻¹)	0,03	0,1
 Bias (t)^b	0,005 (0,01)	0,47 (1,05)
ER(%)	8	9
CV(%) repetitividade^c	2/2/2	3/3/2
Robustez (p-valor)^d	0,40	0,20
LOD (cmol_c kg⁻¹)	0,12	0,3
LOQ (cmol_c kg⁻¹)	0,35	1,1

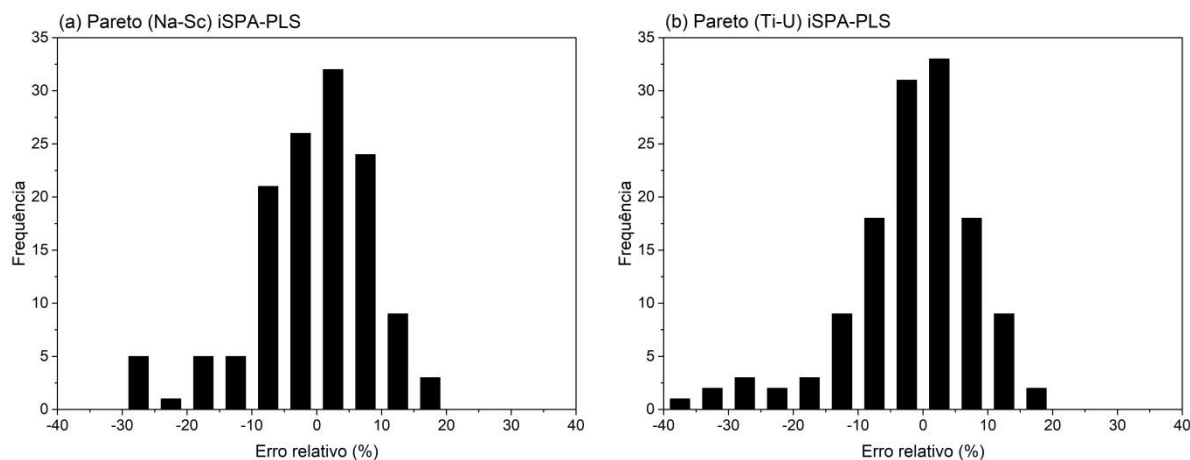
^a Valor expresso como a razão entre intensidade e % (cmol_c kg⁻¹/ cmol_c kg⁻¹)

^b $t_{critico} = 1,98$

^c Valores em nível baixo (38%), médio (60%) e alto (76%).

^d p-valor do teste de randomização em 95% de confiança.

Figura 41 - Erros relativos para o conjunto de previsões dos melhores modelos para V%.



Fonte: do autor.

A precisão dos modelos foi avaliada em um nível baixo (38%), médio (60%) e alto (76%). Ambos os modelos demonstraram boa precisão, com CV de 2% a 3% para todos os níveis. Além disso, não foi observada diferença significativa no conjunto previsão com relação à variação de granulometria, indicando robustez dos modelos. Portanto, esses resultados demonstram que ambos os modelos apresentaram veracidade e precisão adequadas, sendo que o modelo Pareto (Na-Sc) iSPA-PLS apresentou um desempenho ligeiramente melhor dessas figuras de mérito.

6.4.9 Modelos para pH

6.4.9.1 Pré-processamento e seleção de variáveis dos modelos de pH

Os resultados para a calibração e previsão dos modelos de pH com os métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 36. É importante ressaltar que nenhuma amostra foi identificada como *outlier* nessas amostras.

De forma geral, os valores de RMSEP e RMSECV de todos os modelos para o pH foram semelhantes, sendo que os modelos com escalamento de Pareto apresentaram um desempenho ligeiramente melhor do que os modelos com

os dados centrados na média.

Tabela 36 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de PH com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	368	1012	368	552
	LV	3	3	4	3	4	3
	R²_{cal}	0,55	0,52	0,59	0,59	0,69	0,56
	RMSEC	0,27	0,28	0,26	0,26	0,23	0,27
	R²_{cv}	0,52	0,49	0,55	0,55	0,56	0,53
	RMSECV	0,29	0,29	0,28	0,28	0,27	0,28
	R²_{pred}	0,55	0,52	0,55	0,55	0,55	0,56
	RMSEP	0,27	0,28	0,27	0,27	0,27	0,27
RPD	1,5	1,4	1,5	1,5	1,5	1,5	
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	920	1104	368	552
	LV	4	2	4	4	3	3
	R²_{cal}	0,58	0,52	0,58	0,58	0,58	0,56
	RMSEC	0,27	0,28	0,27	0,27	0,27	0,27
	R²_{cv}	0,55	0,50	0,55	0,55	0,52	0,53
	RMSECV	0,27	0,29	0,28	0,28	0,29	0,28
	R²_{pred}	0,55	0,52	0,55	0,53	0,53	0,56
	RMSEP	0,28	0,28	0,28	0,28	0,28	0,27
RPD	1,5	1,4	1,5	1,4	1,4	1,5	
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	254	508	254	381
	LV	6	2	2	4	2	2
	R²_{cal}	0,71	0,66	0,66	0,67	0,66	0,66
	RMSEC	0,22	0,23	0,23	0,23	0,23	0,24
	R²_{cv}	0,64	0,62	0,62	0,64	0,62	0,61
	RMSECV	0,24	0,24	0,24	0,24	0,24	0,25
	R²_{pred}	0,58	0,58	0,59	0,59	0,59	0,59
	RMSEP	0,27	0,26	0,26	0,27	0,26	0,26
RPD	1,5	1,6	1,6	1,5	1,6	1,6	
Faixa Ti-U centrado na média	nº Variáveis	1270	127	635	1016	254	381
	LV	6	4	5	5	5	4
	R²_{cal}	0,64	0,56	0,59	0,59	0,59	0,56
	RMSEC	0,25	0,27	0,26	0,26	0,26	0,27
	R²_{cv}	0,53	0,53	0,53	0,53	0,53	0,52
	RMSECV	0,28	0,28	0,28	0,28	0,28	0,26
	R²_{pred}	0,55	0,58	0,53	0,52	0,55	0,58
	RMSEP	0,27	0,27	0,28	0,28	0,28	0,27
RPD	1,5	1,5	1,5	1,4	1,5	1,5	

*EC=Espectro completo

A seleção de variáveis por iSPA-PLS apresentou uma pequena melhora para ambos os pré-processamentos no espectro de Na-Sc e para os dados centrados na média do espectro de Ti-U. No caso dos modelos utilizando a faixa espectral de Ti-U com escalamento de Pareto, os métodos FiPLS e SiPLS (que selecionaram as mesmas variáveis) foram os que apresentaram o melhor desempenho para modelagem do pH (menor RMSEP e maior R^2_{pred}).

De acordo com o teste de randomização em um nível de confiança de 95% (Tabela 37), não houve diferença significativa entre as comparações realizadas. No entanto, somente alguns modelos utilizando a faixa espectral de Ti-U com escalamento apresentaram RPD = 1,6.

Tabela 37 - Teste de randomização para comparação dos melhores modelos para o pH em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
Pareto (Na-Sc) EC X Pareto (Na-Sc) iSPA	0,06	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com escalamento de Pareto.
MC (Na-Sc) EC X MC (Na-Sc) iSPA	0,12	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Na-Sc com centragem na média.
Pareto (Ti-U) EC X Pareto (Ti-U) FiPLS	0,17	Não houve diferença significativa na seleção de variáveis por FiPLS no espectro de Ti-U com escalamento de Pareto.
MC (Ti-U) EC X MC (Ti-U) iSPA	0,14	Não houve diferença significativa na seleção de variáveis por iSPA-PLS no espectro de Ti-U com centragem na média.
Pareto (Na-Sc) iSPA X MC (Na-Sc) iSPA	0,10	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
Pareto (Ti-U) FiPLS X MC (Ti-U) iSPA	0,07	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
Pareto (Na-Sc) iSPA X Pareto (Ti-U) FiPLS	0,07	Não houve diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U

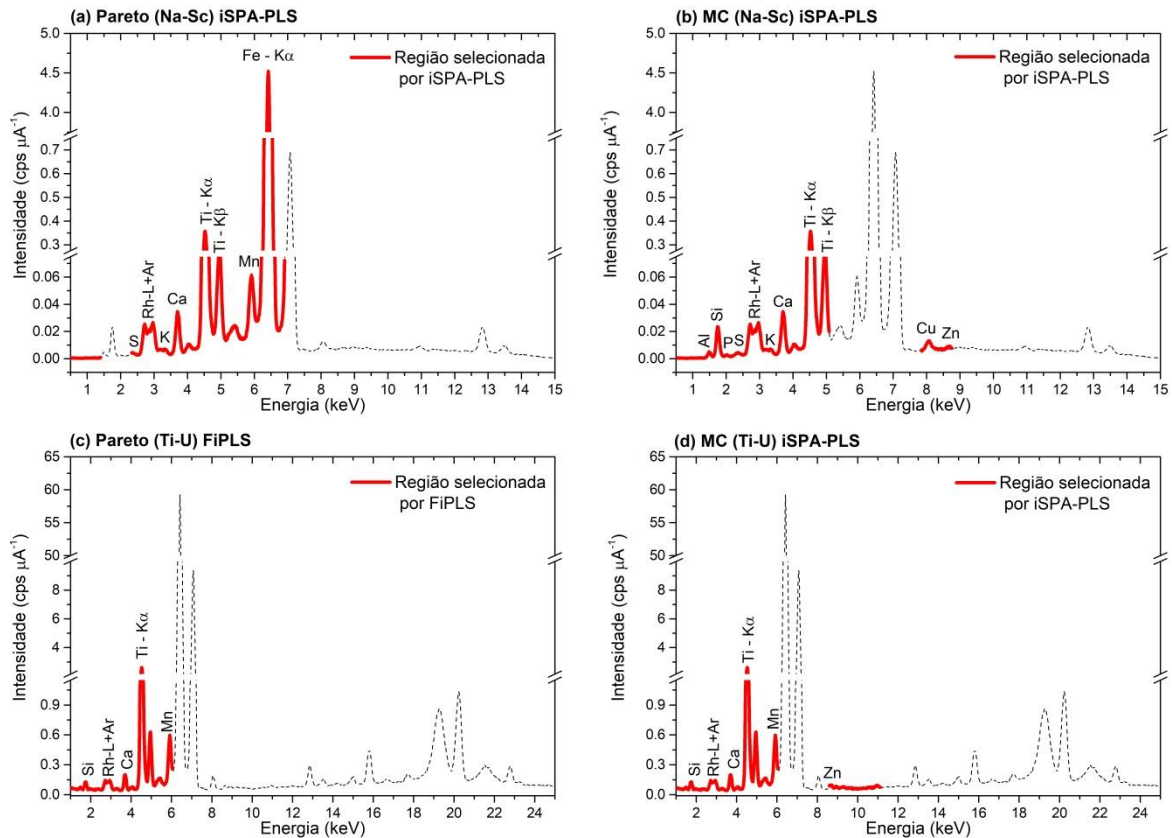
6.4.9.2 Interpretação espectral das variáveis selecionadas nos modelos de pH

As variáveis selecionadas para os melhores modelos de cada pré-processamento são mostradas na Figura 42 e os vetores de regressão para os

respectivos modelos são mostrados na Figura 43.

De forma geral, as variáveis selecionadas para os melhores modelos de pH foram semelhantes. Observam-se pequenas diferenças como a presença dos picos de Cu e Zn nos modelos com os dados centrados na média, os quais não estão presentes nos modelos com escalamento de Pareto, e a ausência dos picos de Al e Si no modelo Pareto (Na-Sc) iSPA-PLS, bem como a presença do pico Fe- $K\alpha$ nesse modelo.

Figura 42 - Intervalos de variáveis selecionadas para os melhores modelos de pH com escalamento de Pareto e centragem na média.



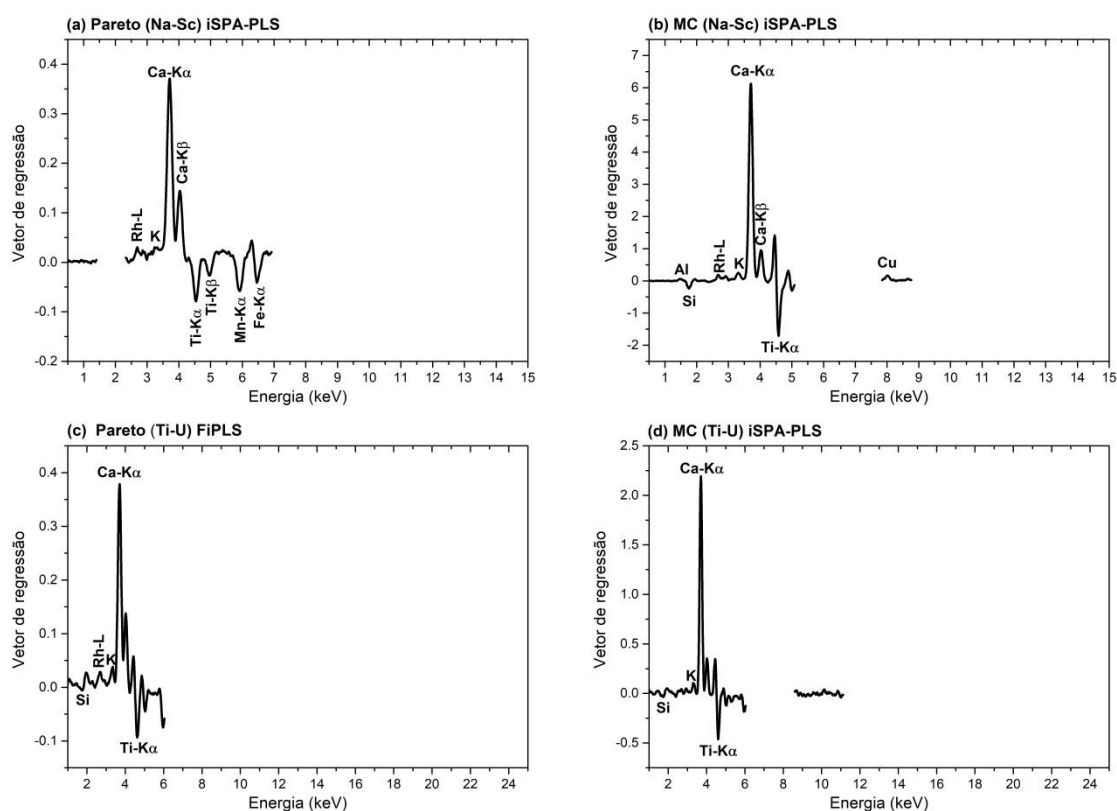
Fonte: do autor.

De acordo com os vetores de regressão (Figura 43), observa-se que as principais contribuições para modelagem do pH foram os picos de Ca, que tiveram uma correlação positiva com o pH, seguido pelos picos de Ti, que tiveram uma correlação negativa. Os demais picos, como Al, Rh-L, K e Cu, tiveram uma pequena contribuição positiva, enquanto o pico de Si teve uma pequena contribuição negativa. Para o modelo Pareto (Na-Sc) iSPA-PLS, além das contribuições

mencionadas acima, observa-se uma contribuição negativa relativamente grande dos picos Mn-K α e Fe-K α .

Provavelmente, a alta contribuição dos picos de Ca na modelagem do pH é em decorrência da aplicação de calcário no solo aumentar o seu pH.

Figura 43 - Vetor de regressão dos melhores modelos PLSR para o pH com escalamento de Pareto e centragem na média.



Fonte: do autor.

6.4.9.3 Figuras de mérito dos modelos para pH

A Tabela 38 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia dos modelos para pH. De acordo com essas figuras de mérito todos os modelos apresentaram linearidade e estão ausentes de *bias*, em um nível de confiança de 95%. Observa-se que o erro relativo de todos os modelos foi de 5%. Além desse baixo valor de erro relativo, a Figura 44 mostra que os erros relativos individuais de todos os modelos foram baixos, sendo que praticamente todas as amostras tiveram erros relativos individuais menores que \pm

10%. A precisão dos modelos foi avaliada em um nível de pH baixo (4,3), médio (5,6) e alto (6,2). Para todos os modelos, os CVs por repetitividade ficaram abaixo de 4%, indicando boa precisão. Os maiores CVs foram observados no nível baixo de pH, sendo 4% para os modelos com espectro de Na-Sc e 3% para os modelos com espectro de Ti-U. Os LOQ dos modelos variaram de 0,7 a 0,01, sendo maiores para os modelos com espectro de Ti-U. No entanto, somente os modelos utilizando escalamento de Pareto mostraram-se robustos quanto à variação de granulometria.

Essas figuras de mérito (Tabela 38) demonstram que ambos os modelos apresentaram boa precisão e baixos valores de desvio relativo. No entanto, somente os modelos utilizando escalamento de Pareto mostraram-se robustos quanto à variação de granulometria, sendo que o modelo Pareto (Ti-U) FiPLS foi o que apresentou melhor desempenho, com R^2 de previsão de 0,59 e RPD=1,6. Esses resultados estão de acordo com a maioria dos estudos utilizando dados de XRF para previsão do pH em solos tropicais. Teixeira et al. (2020) também reportaram desempenho satisfatório para previsão do pH em diversos solos de sete estados brasileiros, com RPD de 1,6 e R^2 de 0,63. Já Silva et al. (2017) demonstram excelentes resultados de previsão do pH, com R^2 de 0,88 e RPD de 2,5. Por outro lado, Tavares et al. (2020a) não obtiveram desempenho satisfatório para previsão do pH em solos do estado de São Paulo e Mato Grosso. Nesse caso o R^2 de previsão foi de 0,38 com RPD de 1,2.

Tabela 38 - Figuras de mérito dos melhores modelos para pH.

	Pareto (Na-Sc) iSPA-PLS	MC (Na-Sc) iSPA-PLS	Pareto (Ti-U) FiPLS	MC (Ti-U) iSPA-PLS
Linearidade (p-valor DW)	0,73	0,97	0,36	0,66
SEN^a	0,63	0,006	0,85	0,18
γ^{-1}	0,01	0,001	0,07	0,02
 Bias (t)^b	0,01 (0,51)	0,02 (0,77)	0,02 (0,74)	0,02 (0,81)
ER(%)	5	5	5	5
CV(%) repetitividade^c	4/1/2	4/1/3	3/1/1	3/2/2
Robustez (p-valor)^d	0,46	0,001	0,30	0,001
LOD	0,04	0,004	0,24	0,07
LOQ	0,12	0,012	0,73	0,21

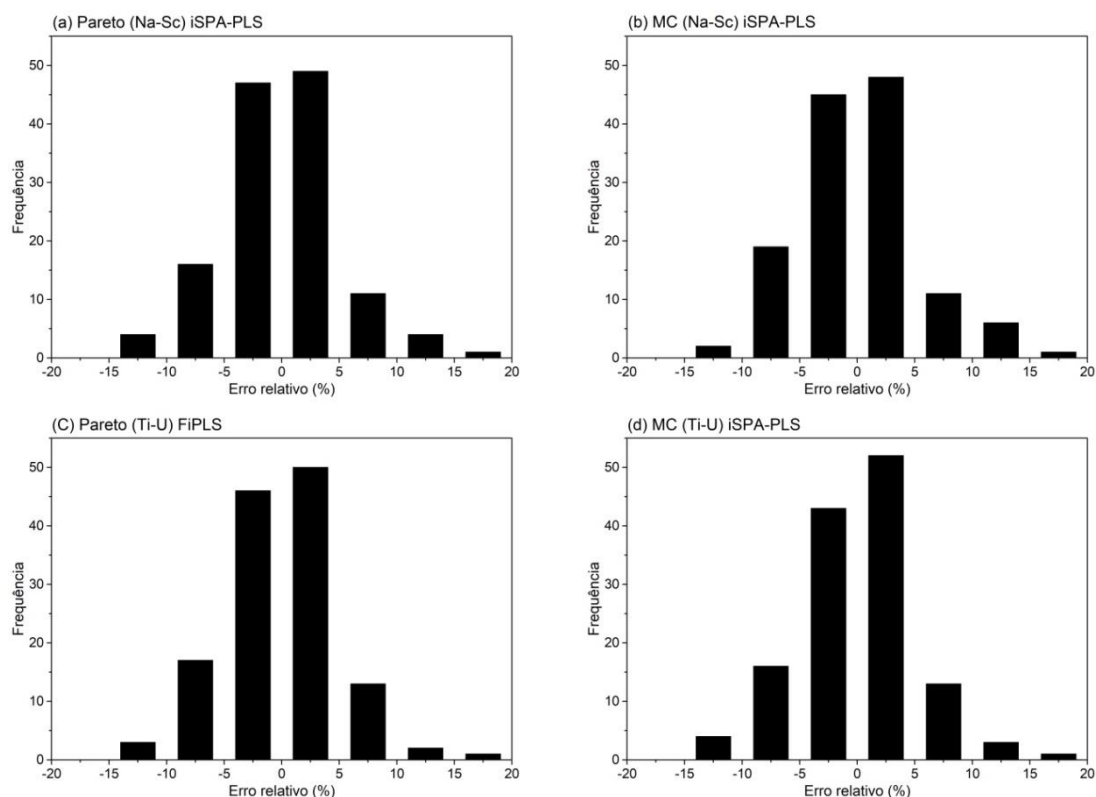
^a Valor expresso como a razão entre intensidade e valor de pH

^b $t_{critico} = 1,98$

^c Valores em nível baixo (4,3), médio (5,6) e alto (6,2).

^d p-valor do teste de randomização em 95% de confiança.

Figura 44 - Erros relativos para o conjunto de previsão dos melhores modelos para pH.



Fonte: do autor.

6.4.10 Modelos para acidez potencial

6.4.10.1 Pré-processamento e seleção de variáveis dos modelos de acidez potencial

Os resultados para a calibração e previsão dos modelos de acidez potencial com os métodos de seleção de variáveis nos espectros de Na-Sc e Ti-U usando os dados centrados na média e com escala de Pareto são mostrados na Tabela 39. Todos os modelos foram construídos com 258 amostras de calibração e 132 amostras de previsão, uma vez que 6 amostras do conjunto de calibração foram identificadas como *outliers*.

Tabela 39 - Parâmetros estatísticos dos métodos de seleção de variáveis para os modelos de acidez potencial com os espectros centrados na média e com escala de Pareto.

		EC*	iPLS	FiPLS	BiPLS	SiPLS	iSPA
Faixa Na-Sc com escalamento de Pareto	nº Variáveis	1472	92	460	1380	368	736
	LV	6	2	4	6	4	4
	R²_{cal}	0,66	0,38	0,62	0,52	0,50	0,52
	RMSEC	0,58	0,79	0,62	0,69	0,68	0,67
	R²_{cv}	0,41	0,35	0,45	0,42	0,45	0,42
	RMSECV	0,79	0,81	0,75	0,77	0,72	0,74
	R²_{pred}	0,55	0,38	0,50	0,55	0,49	0,55
	RMSEP	0,75	0,88	0,78	0,76	0,80	0,77
	RPD	1,5	1,3	1,4	1,5	1,4	1,5
Faixa Na-Sc centrado na média	nº Variáveis	1472	92	184	184	184	460
	LV	4	2	5	5	5	4
	R²_{cal}	0,45	0,38	0,55	0,55	0,55	0,46
	RMSEC	0,73	0,77	0,27	0,27	0,27	0,72
	R²_{cv}	0,41	0,36	0,50	0,50	0,50	0,41
	RMSECV	0,75	0,78	0,69	0,69	0,69	0,75
	R²_{pred}	0,45	0,38	0,50	0,50	0,50	0,48
	RMSEP	0,83	0,87	0,79	0,79	0,79	0,82
	RPD	1,3	1,3	1,4	1,4	1,4	1,3
Faixa Ti-U com escalamento de Pareto	nº Variáveis	1270	127	508	508	254	762
	LV	6	2	5	5	5	6
	R²_{cal}	0,53	0,44	0,56	0,56	0,55	0,53
	RMSEC	0,66	0,73	0,65	0,65	0,66	0,64
	R²_{cv}	0,42	0,41	0,38	0,38	0,46	0,44
	RMSECV	0,75	0,75	0,72	0,72	0,72	0,74
	R²_{pred}	0,53	0,44	0,52	0,52	0,50	0,53
	RMSEP	0,77	0,84	0,78	0,78	0,79	0,78
	RPD	1,4	1,3	1,4	1,4	1,4	1,4
Faixa Ti-U centrado na média	nº Variáveis	1270	127	508	1016	381	889
	LV	6	4	5	5	5	4
	R²_{cal}	0,53	0,45	0,48	0,55	0,49	0,55
	RMSEC	0,67	0,73	0,71	0,66	0,70	0,66
	R²_{cv}	0,40	0,40	0,41	0,37	0,37	0,37
	RMSECV	0,76	0,76	0,76	0,79	0,78	0,79
	R²_{pred}	0,42	0,45	0,41	0,46	0,46	0,45
	RMSEP	0,85	0,84	0,86	0,82	0,83	0,83
	RPD	1,3	1,3	1,3	1,4	1,3	1,3

*EC=Espectro completo; Valores de RMSE em cmolc kg⁻¹

Por meio da Tabela 39, observa-se que os modelos utilizando escalamento de Pareto apresentaram valores de RMSEP menores do que os modelos com os dados centrados na média. Para os modelos com espectro completo das faixas de Na-Sc e Ti-U, o escalamento de Pareto reduziu em 10% (de 0,83 para 0,75 $\text{cmol}_c \text{ kg}^{-1}$) e 9% (de 0,85 para 0,77 $\text{cmol}_c \text{ kg}^{-1}$) o RMSEP com relação à centragem na média, respectivamente. De acordo com o teste de randomização, essa diferença nos valores de RMSEP foram significativas em um nível de confiança de 95%, com p-valor de 0,002 para os modelos com espectro completo de Na-Sc e p-valor de 0,001 para os modelos com espectro completo de Ti-U.

Todos os métodos de seleção de variáveis pioraram o desempenho dos modelos com escalamento de Pareto para ambas as faixas espectrais. Por outro lado, a seleção de variáveis mostrou-se eficiente para os modelos com os dados centrados na média. Nesse caso, o uso de FiPLS, BiPLS e SiPLS (que selecionaram as mesmas regiões) reduziu em 5% (de 0,83 para 0,79 $\text{cmol}_c \text{ kg}^{-1}$) o RMSEP dos modelos com espectro de Na-Sc centrados na média. Para os modelos com o espectro de Ti-U centrados na média, essa redução foi de 4% (0,85 para 0,82 $\text{cmol}_c \text{ kg}^{-1}$) com o uso do método BiPLS. De acordo com o teste de randomização (Tabela 40), o uso destes métodos de seleção de variáveis apresentou uma melhora significativa para os modelos com os dados centrados na média.

Ao comparar os melhores modelos dentro de cada pré-processamento (escalamento de Pareto e centragem na média) utilizando a faixa espectral de Na-Sc, o teste de randomização (Tabela 40) demonstrou que não houve diferença significativa em um nível de confiança de 95%. Por outro lado, ao comparar os melhores modelos de cada pré-processamento no espectro de Ti-U, o teste de randomização mostrou que houve diferença significativa. Com relação à condição de medida, não foi observado diferença significativa entre os melhores modelos de cada faixa espectral (Na-Sc e Ti-U). Entretanto, somente os modelos para a faixa de Na-Sc apresentaram $\text{RPD} > 1,4$, sendo que o melhor modelo foi com o espectro completo na faixa de Na-Sc com escalamento de Pareto. Esse modelo apresentou R^2 de previsão de 0,55 e $\text{RPD}=1,5$. Existem poucos estudos avaliando o uso de dados de XRF para previsão da acidez potencial em solos. Até onde se sabe, um dos únicos trabalhos nesse sentido foi realizado por Towett et al. (2015). Nesse caso, os autores utilizaram dados de TXRF, MIR e a fusão de dados MIR com TXRF

(MIR+TXRF) para determinar diversos atributos de fertilidade do solo, entre eles a acidez potencial, a qual foi determinada com R^2 de previsão de 0,58, 0,47 e 0,55 para os dados de TXRF, MIR e MIR+TXRF, respectivamente. Em solos tropicais, Demattê et al. (2019a) e Pinheiro et al. (2017) determinaram a acidez potencial utilizando dados de Vis-NIR com R^2 de previsão variando de 0,70 a 0,89.

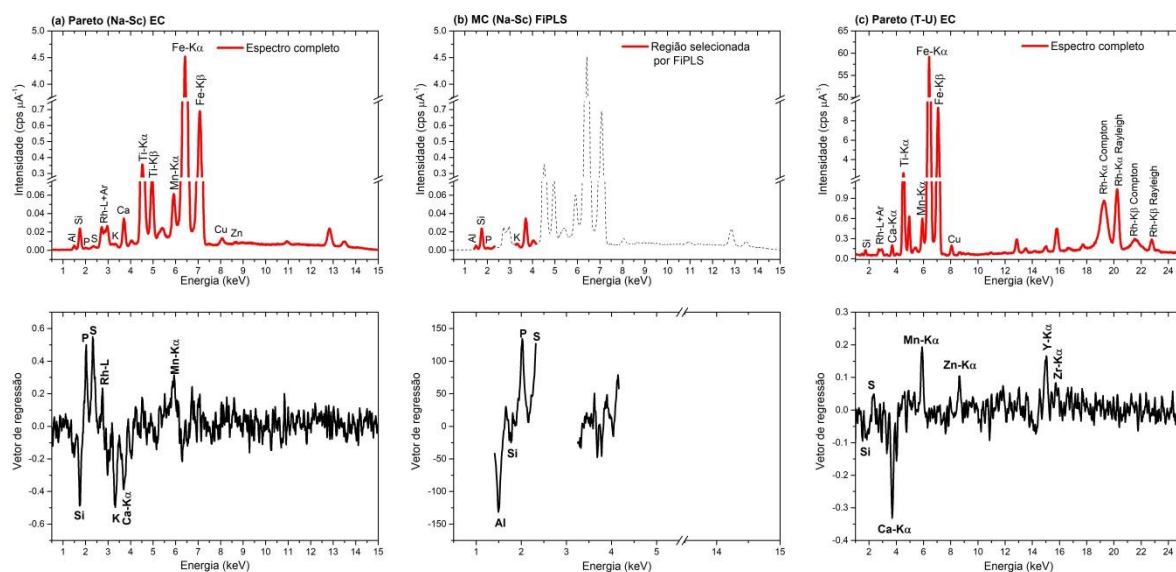
Tabela 40 - Teste de randomização para comparação dos melhores modelos para a acidez potencial em um intervalo de confiança de 95%.

Modelos	p-valor	Comentários
MC (Na-Sc) EC X MC (Na-Sc) FiPLS	0,04	Houve diferença significativa na seleção de variáveis por FiPLS no espectro de Na-Sc com centragem na média.
MC (Ti-U) EC X MC (Ti-U) BiPLS	0,02	Houve diferença significativa na seleção de variáveis por BiPLS no espectro de Ti-U com centragem na média.
Pareto (Na-Sc) EC X MC (Na-Sc) FiPLS	0,12	Não houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Na-Sc.
Pareto (Ti-U) EC X MC (Ti-U) BiPLS	0,001	Houve diferença significativa entre os melhores modelos de cada pré-processamento no espectro de Ti-U.
Pareto (Na-Sc) EC X Pareto (Ti-U) EC	0,22	Não houve diferença significativa entre os melhores modelos do espectro de Na-Sc e Ti-U

6.4.10.2 Interpretação espectral das variáveis utilizadas nos modelos de acidez potencial

De acordo com o item anterior, os modelos que apresentaram o melhor desempenho para acidez potencial foram construídos com o espectro completo nas faixas de Na-Sc e Ti-U, ambos com escalamento de Pareto (Pareto (Na-Sc) EC) e (Pareto (Ti-U) EC), e com a seleção de variáveis por FiPLS no espectro de Na-Sc centrado na média (MC (Na-Sc) FiPLS). A Figura 45 apresenta as variáveis selecionadas para estes modelos com seus respectivos vetores de regressão. Observa-se que o modelo MC (Na-Sc) FiPLS utilizou as regiões com os picos e Al, Si, P, S, K e Ca.

Figura 45 - Variáveis utilizadas para os melhores modelos de acidez potencial com seus respectivos vetores de regressão.



Fonte: do autor.

O vetor de regressão do modelo Pareto (Na-Sc) EC mostra que os picos de Si, P e S foram os que mais contribuíram para modelagem da acidez potencial, sendo que o pico de Si teve correlação negativa com a acidez potencial e os picos de P e S tiveram correlação positiva. Além desses picos, observa-se a contribuição positiva das linhas L do Rh e do Mn-K α , bem como as contribuições negativas dos picos de K e Ca. Para o modelo MC (Na-Sc) FIPLS, os picos de P e S também foram os que mais contribuíram positivamente. No entanto, a maior contribuição negativa foi do pico de Al seguido pelo Si. O vetor de regressão do modelo Pareto (Ti-U) EC teve um comportamento similar ao vetor de regressão do modelo Pareto (Na-Sc) EC, para o qual os picos de Si e Ca apresentaram uma contribuição negativa e os picos de S e Mn contribuíram positivamente. Nesse caso, nota-se que o peso dessas contribuições foi diferente. O pico de Ca-K α foi o que apresentou o maior peso (com correlação negativa), em seguida veio o pico Mn-K α o qual teve correlação positiva. Além desses, também se observa uma grande contribuição positiva dos picos Zn-K α e Y-K α .

6.4.10.3 Figuras de mérito dos modelos para acidez potencial

A Tabela 41 apresenta algumas figuras de mérito que foram avaliadas para verificar a eficácia dos modelos para acidez potencial. De acordo com essas figuras de mérito, em um nível de confiança de 95%, todos os modelos apresentaram linearidade e estão ausentes de *bias*. O erro relativo foi de 13% para o modelo Pareto (Na-Sc) EC e 14% para os demais. Através da Figura 46, observa-se que o modelo MC (Na-Sc) FiPLS apresentou erros relativos individuais menores com relação aos demais modelos. Contudo, um número significativo de amostras apresentaram erros relativos acima de $\pm 20\%$, chegando, em alguns casos, a erros relativos individuais superiores a $\pm 40\%$.

Tabela 41 - Figuras de mérito dos melhores modelos para acidez potencial.

	Pareto (Na-Sc) EC	MC (Na-Sc) FiPLS	Pareto (Ti-U) EC
Linearidade (p-valor DW)	0,08	0,38	0,29
SEN^a	0,22	0,002	0,61
γ^{-1} (cmol_c kg⁻¹)	0,03	0,04	0,1
 Bias (t)^b	0,05 (0,77)	0,05 (0,59)	0,02 (0,01)
ER(%)	13	14	14
CV(%) repetitividade^c	4/1/9	10/3/2	7/4/11
Robustez (p-valor)^d	0,35	0,0005	0,47
LOD (cmol_c kg⁻¹)	0,11	0,12	0,3
LOQ (cmol_c kg⁻¹)	0,34	0,36	1,0

^a Valor expresso como a razão entre intensidade e cmol_c kg⁻¹

^b $t_{critico} = 1,98$

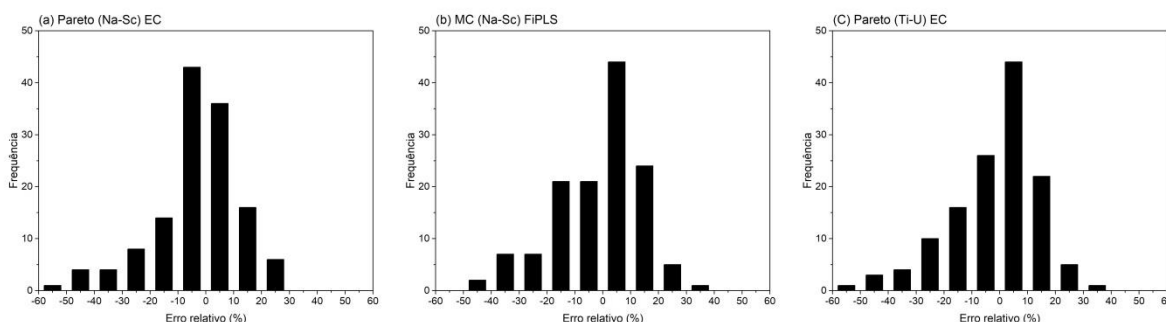
^c Valores em nível baixo (3,1 cmol_c kg⁻¹), médio (5,4 cmol_c kg⁻¹) e alto (7,5 cmol_c kg⁻¹).

^d p-valor do teste de randomização em 95% de confiança.

De acordo com o teste de randomização para robustez, somente os modelos utilizando o escalamento de Pareto apresentaram robustez quanto à variação de granulometria de 2 mm para 125 μm do conjunto de previsão. A precisão dos modelos foi avaliada em níveis de concentração baixo (3,1 cmol_c kg⁻¹), médio (5,4 cmol_c kg⁻¹) e alto (7,5 cmol_c kg⁻¹). Em geral, os CVs por repetitividade

ficaram abaixo de 10%, exceto para o alto nível de concentração do modelo Pareto (Ti-U) EC, que apresentou CV de 11%. Além de apresentar a pior precisão e veracidade, o modelo Pareto (Ti-U) EC também apresentou os piores LOD ($0,3 \text{ cmol}_c \text{ kg}^{-1}$) e LOQ ($1,0 \text{ cmol}_c \text{ kg}^{-1}$).

Figura 46 - Erros relativos para o conjunto de previsões dos melhores modelos para acidez potencial.



Fonte: do autor.

6.4.11 Síntese dos resultados

A Tabela 42 apresenta uma síntese dos resultados dos melhores modelos com base no valor máximo de R^2 de previsão e valor mínimo de RMSEP de cada atributo de fertilidade avaliado.

Tabela 42 - Síntese dos resultados dos melhores modelos de cada parâmetro.

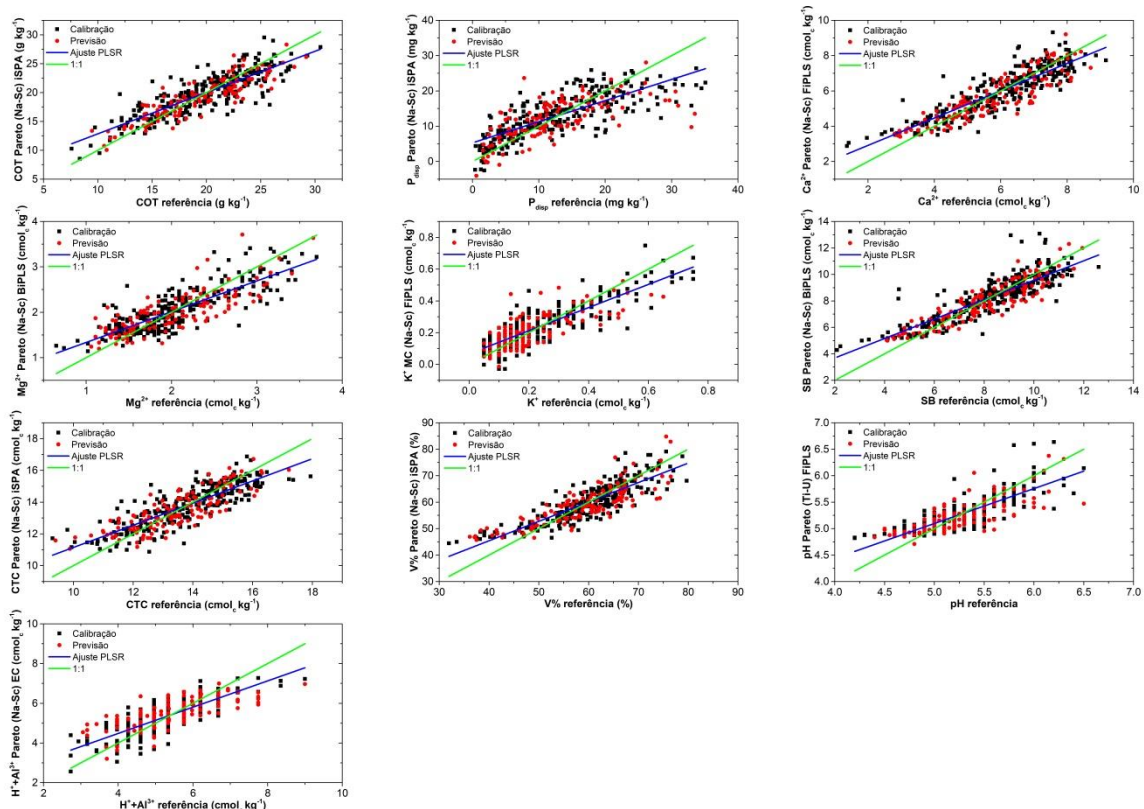
Parâmetro	Melhor condição de medida*	Melhor modelo*	RPD	Elementos que mais contribuíram	ER(%)	LOQ
COT	Na-Sc	Pareto iSPA	2,2	Si, P, S e Rh-L	10	1,1
P_{disp}	Na-Sc	Pareto iSPA	1,4	P, Ca, Al e K	38	1,1
Ca²⁺	Na-Sc	Pareto FiPLS	2,2	K, Si e Ca	11	0,4
Mg²⁺	Na-Sc	Pareto BiPLS	1,8	Ca, K e Mn	16	0,3
K⁺	Na-Sc	Pareto iSPA	1,5	K, Rh-L, Cu e Al	34	0,35
SB	Na-Sc	Pareto BiPLS	2,3	K, Ca, Si e Cu	10	0,3
CTC	Na-Sc	Pareto SiPLS	1,8	K, Ca, Cu e Al	6	0,3
V%	Na-Sc	Pareto iSPA	1,9	K, Ca, Si e S	8	0,4
pH	Ti-U	Pareto FiPLS	1,6	Ca e Ti	5	0,7
(H⁺ + Al³⁺)	Na-Sc	Pareto EC	1,5	Si, P, S, K e Ca	13	0,3

* De acordo com os valores máximos de R^2 de previsão e mínimos de RMSEP; ER – Erro relativo; LOQ – Limite mínimo de quantificação

De acordo com a classificação de Viscarra Rossel; Mcdglynn; Mcbratney (2006) para atributos do solo, nenhum desses modelos foi classificado como excelente. Somente o COT, Ca^{2+} e SB apresentaram modelos considerados bons ($2,5 > \text{RPD} > 2,0$), nos quais os resultados quantitativos foram obtidos com acurácia e precisão. Mg^{2+} , CTC e V% apresentaram modelos razoáveis e os demais atributos apresentaram modelos fracos ($1,8 > \text{RPD} > 1,4$), que podem ser utilizados apenas para avaliações qualitativas e de correlação.

A Figura 47 mostra os gráficos de dispersão com os valores de referência dos atributos de fertilidade do solo *versus* os previstos pelos melhores modelos de cada atributo.

Figura 47 - Gráfico com os valores de referência de cada atributo de fertilidade do solo *versus* valores previstos pelos melhores modelos.



Fonte: do autor.

Observa-se, pela Tabela 42, que a faixa de Na-Sc foi a melhor condição de medida para praticamente todos os atributos avaliados, exceto para o pH. Além disso, o uso do escalamento de Pareto como pré-processamento gerou os menores RMSEPs para todos os parâmetros e mostrou-se mais robusto para a

variação de granulometria de 2 mm para 125 μm do conjunto de previsão. Em geral, observou-se que os elementos com $Z < 20$ foram os que mais contribuíram para modelagem destes parâmetros, somente em alguns casos houve a contribuição da região com os picos de Mn, Cu e Zn. Isso poderia explicar porque os modelos com o espectro de Na-Sc apresentaram um desempenho ligeiramente melhor do que com o espectro de Ti-U, visto que a condição de medida de Na-Sc é mais indicada para excitar os elementos com $Z < 20$.

Devido ao alto intemperismo de solos tropicais, a relação entre os teores elementares e os atributos de fertilidade é altamente influenciada pelo manejo do solo. Além disso, essas relações também podem variar dependendo de condições climáticas, topografia, uso do solo, entre outros. Portanto, a qualidade dos modelos preditivos, bem como as variáveis que mais contribuem para a sua modelagem, podem oscilar entre diferentes áreas, culturas e até mesmo no tempo. Dessa forma, mais estudos devem ser realizados para entender a dinâmica desses atributos de fertilidade no solo e suas relações com dados de XRF.

7 CONCLUSÕES

Os resultados deste estudo mostraram que por meio de EDXRF foi possível determinar os teores totais de Al, Si, K, Ca, Ti, Mn, Fe, Cu e Zn em solos agrícolas e a análise exploratória destes dados, por PCA, mostrou uma separação das amostras de acordo com a altitude da topossequência e/ou classificação do tipo de solo.

Além da aplicação convencional de EDXRF para a determinação dos teores inorgânicos totais em solo, os resultados deste estudo demonstram a viabilidade do uso de dados espectrais de EDXRF na determinação de COT, Ca^{2+} , SB, CTC e V% nos solos em questão. No entanto, para $\text{H}^+\text{+Al}^{3+}$, pH, K^+ e P_{disp} os resultados preditivos não foram satisfatórios, apresentando $\text{RPD} < 1,8$. Neste sentido, mais estudos devem ser realizados com o objetivo de verificar a viabilidade do uso de dados espectrais de EDXRF para previsão desses atributos.

Avaliou-se a influência de duas condições de medidas (Na-Sc e Ti-U) na previsão dos atributos de fertilidade do solo. Em geral, o uso de espectros na faixa Na-Sc gerou os melhores resultados de previsão, indicando que os picos característicos dos elementos mais leves (de número atômico de $Z = 13$ a $Z = 20$), bem como o espalhamento das linhas L do tubo de Rh foram de grande importância para a modelagem desses dados. Por outro lado, o uso de espectros na faixa de Ti-U gerou os melhores resultados de previsão para o pH, principalmente devido à importância dos picos de Ti na modelagem deste parâmetro.

O uso do escalamento de Pareto como pré-processamento gerou os menores RMSEPs e mostrou-se robusto com relação à variação de granulometria para praticamente todos os parâmetros. Além de gerar o melhor desempenho dos modelos, o uso do escalamento de Pareto mostrou-se adequado para esse conjunto de dados, uma vez que o espectro de EDXRF é afetado pelo rendimento de fluorescência e os picos de Fe são muito intensos para o solo em estudo.

Em geral, os métodos de seleção de variáveis resultaram em uma melhora na capacidade preditiva dos modelos de calibração. Isso deixa claro que o uso do espectro completo de EDXRF introduz ruídos e interferentes nos modelos de calibração, prejudicando o seu desempenho.

Através da avaliação do vetor de regressão foi possível identificar as regiões dos espectros de EDXRF que mais contribuíram na determinação desses

atributos e a análise das figuras de mérito foi essencial para demonstrar a viabilidade da metodologia e a robustez dos modelos preditivos.

Por fim, estes resultados contribuem para o estabelecimento de uma abordagem no tratamento de dados espectrais de EDXRF para calibração multivariada, bem como para uma melhor interpretação das variáveis que influenciam na modelagem dos atributos de fertilidade do solo. Estudos adicionais utilizando diferentes tipos de solo, equipamentos portáteis, métodos de preparação de amostras e otimização do tempo de medida podem ser necessários para implementar a metodologia na rotina de laboratórios de análise de solo ou em estações de agricultura de precisão. Com isso, o presente estudo demonstra o potencial da metodologia proposta, como um procedimento eficiente, econômico e ecológico para a determinação dos atributos de fertilidade do solo.

REFERÊNCIAS

- AIDENE, S.; SEMENOV, V.; KIRSANOV, DENIS; KIRSANOV, DMITRY; PANCHUK, V. Assessment of the physical properties, and the hydrogen, carbon, and oxygen content in plastics using energy-dispersive X-ray fluorescence spectrometry. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 165, p. 105771, 2020.
- DE ALMEIDA, E.; MELQUIADES, F. L.; MARQUES, J. P. R.; MARGUÍ, E.; DE CARVALHO, H. W. P. Determination of the polymeric thin film thickness by energy dispersive X-ray fluorescence and multivariate analysis. **Spectrochimica Acta Part B: Atomic Spectroscopy**, v. 167, p. 105818, 2020.
- ANDERSEN, C. M.; BRO, R. Variable selection in regression-a tutorial. **Journal of Chemometrics**, v. 24, p. 728–737, 2010.
- ANDERSSON, M. A comparison of nine PLS1 algorithms. **Journal of Chemometrics**, v. 23, p. 518–529, 2009.
- 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**, v. 357, p. 113960, 2020.
- ARAÚJO, M. C. U.; SALDANHA, T. C. B.; GALVÃO, R. K. H.; YONEYAMA, T.; CHAME, H. C.; VISANI, V. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 57, p. 65–73, 2001.
- ABNT.NBR ISO 5725-1:2018. **Exatidão (veracidade e precisão) dos métodos e dos resultados de medição - Parte 1: Princípios gerais e definições**. Rio de Janeiro, 2018.
- ASTM E1655-05. **Standard Practices for Infrared Multivariate Quantitative Analysis**. Conshohocken, 2012.
- BARROS NETO, B. DE; SCARMINIO, I. S.; BRUNS, R. E. **Como fazer experimentos: pesquisa e desenvolvimento na ciência e na indústria**. Campinas: Editora da UNICAMP, 2001.
- BEEBE, K. R.; BEEBE, K. R.; PELL, R. J.; PELL, R. J.; SEASHOLTZ, M. B. **Chemometrics: A Practical Guide**. 1º Ed. ed. New York: Wiley, 1998.
- BELTRAME, K. K.; SOUZA, A. M.; COELHO, M. R.; WINKLER, T. C. B.; SOUZA, W.

- E.; VALDERRAMA, P. Soil organic carbon determination using NIRS: Evaluation of dichromate oxidation and dry combustion analysis as reference methods in multivariate calibration. **Journal of the Brazilian Chemical Society**, v. 27, p. 1527–1532, 2016.
- BENEDET, L.; FARIA, W. M.; SILVA, S. H. G.; MANCINI, M.; DEMATTÊ, J. A. M.; GUILHERME, L. R. G.; CURI, N. Soil texture prediction using portable X-ray fluorescence spectrometry and visible near-infrared diffuse reflectance spectroscopy. **Geoderma**, v. 376, p. 114553, 2020a.
- BENEDET, L.; FARIA, W. M.; SILVA, S. H. G.; MANCINI, M.; GUILHERME, L. R. G.; DEMATTÊ, J. A. M.; CURI, N. Soil subgroup prediction via portable X-ray fluorescence and visible near-infrared spectroscopy. **Geoderma**, v. 365, p. 114212, 2020b.
- VAN DEN BERG, R. A.; HOEFSLOOT, H. C. J.; WESTERHUIS, J. A.; SMILDE, A. K.; VAN DER WERF, M. J. Centering, scaling, and transformations: improving the biological information content of metabolomics data. **BMC Genomics**, v. 7, p. 142, 2006.
- BORTOLETO, G. G.; BORGES, S. S. O.; BUENO, M. I. M. S. X-ray scattering and multivariate analysis for classification of organic samples: A comparative study using Rh tube and synchrotron radiation. **Analytica Chimica Acta**, v. 595, p. 38–42, 2007.
- BRICKLEMYER, R. S.; BROWN, D. J. On-the-go VisNIR: Potential and limitations for mapping soil clay and organic carbon. **Computers and Electronics in Agriculture**, v. 70, p. 209–216, 2010.
- BRIEDIS, C.; DE MORAES SÁ, J. C.; LAL, R.; TIVET, F.; FERREIRA, A. O.; FRANCHINI, J. C.; SCHIMIGUEL, R.; HARTMAN, D. C.; SANTOS, J. Z. Can highly weathered soils under conservation agriculture be C saturated? **Catena**, v. 147, p. 638–649, 2016.
- BRINATTI, A. M.; MASCARENHAS, Y. P.; PEREIRA, V. P.; PARTITI, C. S. DE M.; MACEDO, Á. Mineralogical characterization of a highly-weathered soil by the Rietveld Method. **Scientia Agricola**, v. 67, p. 454-464, 2010.
- BRO, R.; SMILDE, A. K. Centering and scaling in component analysis. **Journal of Chemometrics**, v. 17, p. 16–33, 2003.
- BROWN, D. J. Using a global VNIR soil-spectral library for local soil characterization and landscape modeling in a 2nd-order Uganda watershed. **Geoderma**, v. 140, p. 444–453, 2007.

- BUENO, M. I. M. S.; CASTRO, M. T. P. O.; DE SOUZA, A. M.; DE OLIVEIRA, E. B. S.; TEIXEIRA, A. P. X-ray scattering processes and chemometrics for differentiating complex samples using conventional EDXRF equipment. **Chemometrics and Intelligent Laboratory Systems**, v. 78, p. 96–102, 2005.
- BUTLER, O.; EVANS, H.; FISHER, A.; HILL, S.; HARRINGTON, C.; TAYLOR, A.; WEST, M.; ELLIS, A. Atomic Spectrometry Updates: A 25-year retrospective. **Journal of Analytical Atomic Spectrometry**, v. 25, p. 1546–1566, 2010.
- BYERS, H. L.; MCHENRY, L. J.; GRUNDL, T. J. XRF techniques to quantify heavy metals in vegetables at low detection limits. **Food Chemistry: X**, v. 1, p. 100001, 2019.
- DE CASTILHOS, N. D. B.; MELQUIADES, F. L.; THOMAZ, E. L.; BASTOS, R. O. X-ray fluorescence and gamma-ray spectrometry combined with multivariate analysis for topographic studies in agricultural soil. **Applied Radiation and Isotopes**, v. 95, p. 63–71, 2015.
- CENTNER, V.; DE NOORD, O. E.; MASSART, D. L. Detection of nonlinearity in multivariate calibration. **Analytica Chimica Acta**, v. 376, p. 153–168, 1998.
- CESAREO, R.; HANSON, A. L.; GIGANTE, G. E.; PEDRAZA, L. J.; MATHABOALLY, S. Q. G. Interaction of keV photons with matter and new applications. **Physics Reports**, v. 213, p. 117–178, 1992.
- 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**, v. 338, p. 5–13, 2019.
- COMINO, F.; AYORA-CAÑADA, M. J.; ARANDA, V.; DÍAZ, A.; DOMÍNGUEZ-VIDAL, A. Near-infrared spectroscopy and X-ray fluorescence data fusion for olive leaf analysis and crop nutritional status determination. **Talanta**, v. 188, p. 676–684, 2018.
- COSTA, A. C. S.; BIGHAM, J. M.; TORMENA, C. A.; PINTRO, J. C. Clay mineralogy and cation exchange capacity of Brazilian soils from water contents determined by thermal analysis. **Thermochimica Acta**, v. 413, p. 73–79, 2004.
- CURRIE, L. A. Nomenclature in evaluation of analytical methods including detection and quantification capabilities (IUPAC Recommendations 1995). **Analytica Chimica Acta**, v. 391, p. 105–126, 1999.
- DA-COL, J. A.; BUENO, M. I. M. S.; MELQUIADES, F. L. Fast and direct Na and K determination in table, marine, and low-sodium salts by X-ray fluorescence and chemometrics. **Journal of Agricultural and Food Chemistry**, v. 63, p. 2406–2412,

2015.

DEJAEGHER, B.; CAPRON, X.; SMEYERS-VERBEKE, J.; HEYDEN, Y. VANDER. Randomization tests to identify significant effects in experimental designs for robustness testing. **Analytica Chimica Acta**, v. 564, n. 2, p. 184–200, 2006.

DEMATTE, 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**, v. 337, p. 111–121, 2019a.

DEMATTE, J. A. M.; DOTTO, A. C.; PAIVA, A. F. S.; et al. The Brazilian Soil Spectral Library (BSSL): A general view, application and challenges. **Geoderma**, v. 354, p. 113793, 2019b.

DEMATTE, 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**, v. 288, p. 8–22, 2017.

DONAGEMA, G. K.; CAMPOS, D. V. B. DE; CALDERANO, S. B.; TEIXEIRA, W. G. . V. **Manual de métodos de análise de solo**. 2. ed. Rio de Janeiro: Embrapa Solos, 2011.

EISBERG, R. M.; RESNICK, R. **Física quântica: átomos, moléculas, sólidos, núcleos e partículas**. Rio de Janeiro: ELSEVIER EDITORA, 1979.

ENGEL, J.; GERRETZEN, J.; SZYMAŃSKA, E.; JANSEN, J. J.; DOWNEY, G.; BLANCHET, L.; BUYDENS, L. M. C. Breaking with trends in pre-processing? **TrAC Trends in Analytical Chemistry**, v. 50, p. 96–106, 2013.

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**, v. 110, p. 96–99, 2015.

FERREIRA, M. H.; BRAGA, J. W. B.; SENA, M. M. Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. **Microchemical Journal**, v. 109, p. 158–164, 2013.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v. 22, p. 724–731, 1999.

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas:

Editora da Unicamp, 2015.

GE, Y.; MORGAN, C. L. S.; GRUNWALD, S.; BROWN, D. J.; SARKHOT, D. V. Comparison of soil reflectance spectra and calibration models obtained using multiple spectrometers. **Geoderma**, v. 161, p. 202–211, 2011.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 185, p. 1–17, 1986.

GOMES, 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**, v. 110, p. 202–208, 2013.

GÓMEZ-NUBLA, L.; ARAMENDIA, J.; FDEZ-ORTIZ DE VALLEJUELO, S.; MADARIAGA, J. M. Analytical methodology to elemental quantification of weathered terrestrial analogues to meteorites using a portable Laser-Induced Breakdown Spectroscopy (LIBS) instrument and Partial Least Squares (PLS) as multivariate calibration technique. **Microchemical Journal**, v. 137, p. 392–401, 2018.

GOMEZ, C.; VISCARRA ROSSEL, R. A.; MCBRATNEY, A. B. Soil organic carbon prediction by hyperspectral remote sensing and field vis-NIR spectroscopy: An Australian case study. **Geoderma**, v. 146, p. 403–411, 2008.

GONZAGA, F. B.; PASQUINI, C. A low cost short wave near infrared spectrophotometer: Application for determination of quality parameters of diesel fuel. **Analytica Chimica Acta**, v. 670, p. 92–97, 2010.

GOUEGUEL, C. L.; SOUMARE, A.; NAULT, C.; NAULT, J. Direct determination of soil texture using laser-induced breakdown spectroscopy and multivariate linear regressions. **Journal of Analytical Atomic Spectrometry**, v. 34, p. 1588–1596, 2019.

VAN GRIEKEN, R.; MARKOWICZ, A. **Handbook of X-ray spectrometry**. New York: Marcel Dekker, 2002.

HE, Y.; LIU, X.; LV, Y.; et al. Quantitative analysis of nutrient elements in soil using single and double-pulse laser-induced breakdown spectroscopy. **Sensors**, v. 18, p. 1526, 2018.

HOBLEY, E. U.; PRATER, I. Estimating soil texture from vis-NIR spectra. **European Journal of Soil Science**, v. 70, p. 83–95, 2019.

INSTITUTO NACIONAL DE METROLOGIA NORMALIZAÇÃO E QUALIDADE INDUSTRIAL – INMETRO. **Orientação sobre validação de métodos de ensaios químicos**. Rio de Janeiro, 2010.

INSTITUTO NACIONAL DE METROLOGIA NORMALIZAÇÃO E QUALIDADE INDUSTRIAL – INMETRO. **Vocabulário Internacional de Metrologia – Conceitos fundamentais e gerais e termos associados**. Rio de Janeiro, 2012.

IUSS WORKING GROUP WRB. **World Reference Base for Soil Resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World Soil Resources Reports No. 106**. Rome, 2015.

JANSSENS, K. X-Ray Fluorescence Analysis. In: G. Gauglitz; T. Vo-Dinh (Orgs.); **Handbook of Spectroscopy**. p.363–420, 2003. Weinheim: WILEY-VCH Verlag.

JENKINS, R. **X-Ray Fluorescence Spectrometry**. New York: Wiley, 1988.

KALNICKY, D. J.; SINGHVI, R. Field portable XRF analysis of environmental samples. **Journal of Hazardous Materials**, v. 83, p. 93–122, 2001.

KAMPRATH, E. J.; WATSON, M. E. Conventional Soil and Tissue Tests for Assessing the Phosphorus Status of Soils. In: F. E. Khasawneh; E. C. Sample; E. J. Kamprath (Orgs.); **The Role of Phosphorus in Agriculture**, A. p.433–469, 1980. Madison: American Society of Agronomy.

KANIU, M. I.; ANGEYO, K. H.; MWALA, A. K.; MANGALA, M. J. Direct rapid analysis of trace bioavailable soil macronutrients by chemometrics-assisted energy dispersive X-ray fluorescence and scattering spectrometry. **Analytica Chimica Acta**, v. 729, p. 21–25, 2012.

KAPLAN, I. **Física nuclear**. 2ª Ed. ed. Rio de Janeiro: Guanabara Dois, 1978.

KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, p. 137–148, 1969.

KHUDER, A.; AHMAD, M.; HASAN, R.; SAOUR, G. Improvement of X-ray fluorescence sensitivity by dry ashing method for elemental analysis of bee honey. **Microchemical Journal**, v. 95, p. 152–157, 2010.

KLEBER, M.; MIKUTTA, R.; TORN, M. S.; JAHN, R. Poorly crystalline mineral phases protect organic matter in acid subsoil horizons. **European Journal of Soil Science**, v. 56, p. 717–725, 2005.

KLEBER, M.; EUSTERHUES, K.; KEILUWEIT, M.; MIKUTTA, C.; MIKUTTA, R.; NICO, P. S. Mineral–Organic Associations: Formation, Properties, and Relevance in Soil Environments. **Advances in Agronomy**, v. 130, p. 1–140, 2015.

KLOCKENKÄMPER, R.; VON BOHLEN, A. **Total-Reflection X-Ray Fluorescence Analysis and Related Methods: Second Edition**. Second ed. Hoboken: John Wiley

& Sons, 2015.

LEARDI, R.; NØRGAARD, L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. **Journal of Chemometrics**, v. 18, p. 486–497, 2004.

LEMIÈRE, B. A review of pXRF (field portable X-ray fluorescence) applications for applied geochemistry. **Journal of Geochemical Exploration**, v. 188, p. 350–363, 2018.

LEO, W. R. **Techniques for Nuclear and Particle Physics Experiments**. 2^a ed. New York: pringer-Verlag, 1994.

LIMA, T. M. DE; WEINDORF, D. C.; CURI, N.; GUILHERME, L. R.G.; LANA, R. M.Q. RIBEIRO, B. T. Elemental analysis of Cerrado agricultural soils via portable X-ray fluorescence spectrometry: Inferences for soil fertility assessment. **Geoderma**, v. 353, p. 264–272, 2019.

LOPES, A. S.; GUILHERME, L. R. G. A Career Perspective on Soil Management in the Cerrado Region of Brazil. **Advances in Agronomy**, v. 137, p. 1–72, 2016.

MALMIR, M.; TAHMASBIAN, I.; XU, Z.; FARRAR, M. B.; BAI, S. H. Prediction of soil macro- and micro-elements in sieved and ground air-dried soils using laboratory-based hyperspectral imaging technique. **Geoderma**, v. 340, p. 70–80, 2019.

MANCINI, M.; WEINDORF, D. C.; CHAKRABORTY, S.; SILVA, S. H. G.; TEIXEIRA, A. F. S. GUILHERME, L. R. G.; CURI, N. Tracing tropical soil parent material analysis via portable X-ray fluorescence (pXRF) spectrometry in Brazilian Cerrado. **Geoderma**, v. 337, p. 718–728, 2019a.

MANCINI, M.; WEINDORF, D. C.; SILVA, S. H. G.; CHAKRABORTY, S.; TEIXEIRA, A. F. S. GUILHERME, L. R. G.; CURI, N. Parent material distribution mapping from tropical soils data via machine learning and portable X-ray fluorescence (pXRF) spectrometry in Brazil. **Geoderma**, v. 354, p. 113885, 2019b.

MARGUÍ, E.; QUERALT, I.; HIDALGO, M. Application of X-ray fluorescence spectrometry to determination and quantitation of metals in vegetal material. **TrAC - Trends in Analytical Chemistry**, v. 28, p. 362–372, 2009.

MARIANI, N. C. T.; TEIXEIRA, G. H. A.; DE LIMA, K. M. G.; MORGENSTERN, T. B.; NARDINI, V.; CUNHA JÚNIOR, L. C. NIRS and iSPA-PLS for predicting total anthocyanin content in jaboticaba fruit. **Food Chemistry**, v. 174, p. 643–648, 2015.

MARTENS, H.; NAES, T. **Multivariate Calibration**. New York: Wiley, 1992.

MELQUIADES, F. L.; ANDREONI, L. F. S.; THOMAZ, E. L. Discrimination of land-

- use types in a catchment by energy dispersive X-ray fluorescence and principal component analysis. **Applied Radiation and Isotopes**, v. 77, p. 27–31, 2013.
- MELQUIADES, F. L.; APPOLONI, C. R. Application of XRF and field portable XRF for environmental analysis. **Journal of Radioanalytical and Nuclear Chemistry**, v. 262, p. 533–541, 2004.
- MELQUIADES, F. L.; BASTOS, R. O.; ANDREONI, L. S.; THOMAZ, E. L.; ANTONELI, V. Coupling soil transfer from hillslope to riparian zone through natural fingerprint in a catchment with tobacco crop. **Journal of Soils and Sediments**, v. 19, p. 1928–1936, 2019.
- MELQUIADES, F. L.; BORTOLETO, G. G.; MARCHIORI, L. F. S.; BUENO, M. I. M. S. Direct determination of sugar cane quality parameters by X-ray spectrometry and multivariate analysis. **Journal of Agricultural and Food Chemistry**, v. 60, p. 10755–10761, 2012.
- MELQUIADES, F. L.; DOS SANTOS, F. R. Preliminary Results: Energy Dispersive X-Ray Fluorescence and Partial Least Squares Regression for Organic Matter Determination in Soil. **Spectroscopy Letters**, v. 48, p. 286–289, 2015.
- MORGAN, C. L. S.; WAISER, T. H.; BROWN, D. J.; HALLMARK, C. T. Simulated in situ characterization of soil organic and inorganic carbon with visible near-infrared diffuse reflectance spectroscopy. **Geoderma**, v. 151, p. 249–256, 2009.
- 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. **Applied Radiation and Isotopes**, v. 130, p. 13–20, 2017.
- MOUAZEN, A. M.; DE BAERDEMAEKER, J.; RAMON, H. Effect of wavelength range on the measurement accuracy of some selected soil constituents using visual-near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, v. 14, p. 189–199, 2006.
- MUNNAF, M. A.; NAWAR, S.; MOUAZEN, A. M. Estimation of secondary soil properties by fusion of laboratory and on-line measured Vis-NIR spectra. **Remote Sensing**, v. 11, p. 2819, 2019.
- MYRVANG, M. B.; HEIM, M.; KROGSTAD, T.; ALMÅS, Å. R.; GJENGEDAL, E. The use of carbonatite rock powder as a liming agent. **Journal of Plant Nutrition and Soil Science**, v. 180, p. 326–335, 2017.
- NASCIMENTO FILHO, V. F. **Técnicas Analíticas Nucleares De Fluorescência De Raios X Por Dispersão De Energia (Ed-Xrf) E Por Reflexão Total (Txrf)**. Depto.

- de Física e Meteorologia, CENA, Piracicaba, 1999.
- NOGUEIRA, G. A. L.; PARREIRA, P. S.; DOS SANTOS, F. R.; MELQUIADES, F. L. Determination of metal content in industrial powder ink and paint thickness over steel plates using X-Ray Fluorescence. **Applied Radiation and Isotopes**, v. 150, p. 168–174, 2019.
- NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**, v. 54, p. 413–419, 2000.
- 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**, v. 279, p. 31–44, 2016.
- OSMAN, K. T. **Plant Nutrients and Soil Fertility Management**. Dordrecht: Springer Netherlands, 2013.
- OTTO, M. **Chemometrics: Statistics and Computer Application in Analytical Chemistry**. 2º Ed. Weinheim: Wiley, 2007.
- PARREIRA, P. S. Metodologia de EDXRF e aplicações com um sistema portátil. 2006. Londrina: LFNATEC - Publicação Técnica do Laboratório de Física Nuclear Aplicada.
- PAVAN, M. A.; BLOCH, M. F. M.; ZEMPULSKI, H. C.; MIYAZAWA, M.; ZOCOLER, D. C. **Manual de análise química de solo e controle de qualidade**. Londrina: IAPAR Londrina, 1992.
- PAVLINSKY, G. V.; DUKHANIN, A. Y.; BARANOV, E. O.; PORTNOY, A. Y. Theory of the implementation of the fundamental parameter method for the X-ray fluorescence determination of low-atomic-number elements. **Journal of Analytical Chemistry**, v. 61, p. 654–661, 2006.
- PINHEIRO, É. F. M.; CEDDIA, M. B.; CLINGENSMITH, C. M.; GRUNWALD, S.; VASQUES, G. M. Prediction of soil physical and chemical properties by visible and near-infrared diffuse reflectance spectroscopy in the Central Amazon. **Remote Sensing**, v. 9, p. 293, 2017.
- RAMOS, F. T.; DORES, E. F. DE C.; WEBER, O. L. DO. S.; BEBER, D. C.; CAMPELO, J. H.; MAIA, J. C. S. Soil organic matter doubles the cation exchange capacity of tropical soil under no-till farming in Brazil. **Journal of the Science of**

- Food and Agriculture**, v. 98, p. 3595–3602, 2018.
- RAMOS, P. V.; INDA, A. V.; BARRÓN, V.; SIQUEIRA, D. S.; MARQUES JÚNIOR, J.; TEIXEIRA, D. B. Color in subtropical brazilian soils as determined with a Munsell chart and by diffuse reflectance spectroscopy. **CATENA**, v. 193, p. 104609, 2020.
- 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**, v. 338, p. 375–382, 2019.
- RINNAN, Å.; BERG, F. VAN DEN; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC Trends in Analytical Chemistry**, v. 28, p. 1201–1222, 2009.
- ROSSEL, R. A. V.; CHEN, C. Digitally mapping the information content of visible–near infrared spectra of surficial Australian soils. **Remote Sensing of Environment**, v. 115, p. 1443–1455, 2011.
- ROUILLON, M.; TAYLOR, M. P. Can field portable X-ray fluorescence (pXRF) produce high quality data for application in environmental contamination research? **Environmental Pollution**, v. 214, p. 255–264, 2016.
- RYU, E. Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. **Behavior Research Methods**, v. 43, p. 1066–1074, 2011.
- SANKEY, J. B.; BROWN, D. J.; BERNARD, M. L.; LAWRENCE, R. L. Comparing local vs. global visible and near-infrared (VisNIR) diffuse reflectance spectroscopy (DRS) calibrations for the prediction of soil clay, organic C and inorganic C. **Geoderma**, v. 148, p. 149–158, 2008.
- DE SANTANA, F. B.; DE SOUZA, A. M.; POPPI, R. J. Visible and near infrared spectroscopy coupled to random forest to quantify some soil quality parameters. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 191, p. 454–462, 2018.
- SANTI, L. A.; FIORIN, J. E.; COCCO, K. T.; CHERUBIN, M. R.; EITELWEIN, M. T.; AMADO, T. J. C. Distribuição horizontal e vertical de fósforo e potássio em área manejada com ferramentas de agricultura de precisão. **Revista Plantio Direto**, p. 18–25, 2012.
- DOS SANTOS, D. R.; GATIBONI, L. C.; KAMINSKI, J. Fatores que afetam a disponibilidade do fósforo e o manejo da adubação fosfatada em solos sob sistema plantio direto. **Ciencia Rural**, v. 38, n. 2, p. 576–586, 2008.

- DOS SANTOS, F. R.; DE OLIVEIRA, J. F.; BONA, E.; DOS 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**, v. 152, p. 104275, 2020.
- SANTOS, H. G.; JACOMINE, P. K. T.; ANJOS, L. H. C.; OLIVEIRA, V. Á.; DE COELHO; M. R.; LUMBRERAS, J. F.; CUNHA, T. J. F. **Sistema brasileiro de classificação de solos**. Rio de Janeiro: Embrapa Solos, 2006.
- DOS SANTOS, J. V. F. **FORMAS DE FÓSFORO EM UMA TOPOSSEQUÊNCIA DE SOLOS ORIUNDOS DO BASALTO**, 2019. INSTITUTO AGRONÔMICO DO PARANÁ (IAPAR).
- TEIXEIRA, A. F. S.; PELEGRINO, M. H. P.; FARIA, W. M.; SILVA, S. H. G.; GONÇALVES, M. G. M.; ACERBI JÚNIOR, F. W.; GOMIDE, L. R.; PÁDUA JÚNIOR, A. L.; DE SOUZA, I. A.; CHAKRABORTY, S.; WEINDORF, D. C.; GUILHERME, L. R. G.; CURI, N. Tropical soil pH and sorption complex prediction via portable X-ray fluorescence spectrometry. **Geoderma**, v. 361, p. 114132, 2020.
- SCHAEFER, C. E. G. R.; FABRIS, J. D.; KER, J. C. Minerals in the clay fraction of Brazilian Latosols (Oxisols): a review. **Clay Minerals**, v. 43, p. 137–154, 2008.
- SCHWERTMANN, U.; TAYLOR, R. M. Iron Oxides. **Minerals in Soil Environments**, SSSA Book Series SV - 1. p.379–438, 1989. Madison, WI: Soil Science Society of America.
- SCIUTTO, G.; OLIVERI, P.; PRATI, S.; QUARANTA, M.; BERSANI, S.; MAZZEO, R. An advanced multivariate approach for processing X-ray fluorescence spectral and hyperspectral data from non-invasive in situ analyses on painted surfaces. **Analytica Chimica Acta**, v. 752, p. 30–38, 2012.
- 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**, v. 938, p. 7–17, 2016.
- SERBIN, G.; DAUGHTRY, C. S. T.; HUNT, E. R.; REEVES, J. B.; BROWN, D. J. Effects of soil composition and mineralogy on remote sensing of crop residue cover. **Remote Sensing of Environment**, v. 113, p. 224–238, 2009.
- 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**, v. 232–234, p. 141–147, 2014.
- SHARMA, A.; WEINDORF, D. C.; WANG, D.; CHAKRABORTY, S. Characterizing

soils via portable X-ray fluorescence spectrometer: 4. Cation exchange capacity (CEC). **Geoderma**, v. 239–240, p. 130–134, 2015.

SILVA, S. H. G.; HARTEMINK, A. E.; TEIXEIRA, A. F. DOS S.; INDA, A. V.; GUILHERME, L. R. G.; CURI, N. Soil weathering analysis using a portable X-ray fluorescence (PXRF) spectrometer in an Inceptisol from the Brazilian Cerrado. **Applied Clay Science**, v. 162, p. 27–37, 2018.

SILVA, S. H. G.; TEIXEIRA, A. F. DOS S.; MENEZES, M. D. DE; 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**, v. 41, p. 648–664, 2017.

SILVA, S. H. G.; WEINDORF, D. C.; PINTO, L. C.; FARIA, W. M.; ACERBI JUNIOR, F. W.; GOMIDE, L. R.; DE MELLO, J. M.; DE PÁDUA JUNIOR, A. L.; DE SOUZA, I. A.; TEIXEIRA, A. F. DOS S.; GUILHERME, L. R. G.; CURI, N. Soil texture prediction in tropical soils: A portable X-ray fluorescence spectrometry approach. **Geoderma**, v. 262, p. 114136, 2020.

SIMEONE, M. L. F.; PARRELLA, R. A. C.; SCHAFFERT, R. E.; et al. Near infrared spectroscopy determination of sucrose, glucose and fructose in sweet sorghum juice. **Microchemical Journal**, v. 134, p. 125–130, 2017.

SITHOLE, N. J.; NCAMA, K.; MAGWAZA, L. S. Robust Vis-NIRS models for rapid assessment of soil organic carbon and nitrogen in Feralsols Haplic soils from different tillage management practices. **Computers and Electronics in Agriculture**, v. 153, p. 295–301, 2018.

SOARES, M. R.; ALLEONI, L. R. F. Contribution of Soil Organic Carbon to the Ion Exchange Capacity of Tropical Soils. **Journal of Sustainable Agriculture**, v. 32, p. 439–462, 2008.

SORIANO-DISLA, J. M.; JANIK, L. J.; VISCARRA ROSSEL, R. A.; MACDONALD, L. M.; MCLAUGHLIN, M. J. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. **Applied Spectroscopy Reviews**, v. 49, p. 139–186, 2014.

DE SOUZA, A. M.; FILGUEIRAS, P. R.; COELHO, M. R.; FONTANA, A.; WINKLER, T. C. B.; VALDERRAMA, P.; POPPI, R. J. Validation of the near infrared spectroscopy method for determining soil organic carbon by employing a proficiency assay for fertility laboratories. **Journal of Near Infrared Spectroscopy**, v. 24, p. 293–303, 2016.

- SPOSITO, G. **The Chemistry of Soils**. New York: Oxford University Press, 2008.
- STENBERG, B.; ROSSEL, R. A. V.; MOUAZEN, A. M.; WETTERLIND, J. Visible and Near Infrared Spectroscopy in Soil Science. In: D. L. Sparks (Org.); **Advances in Agronomy**, v. 107, p.163–215, 2010. Burlington: Academic Press.
- STEVENS, A.; NOCITA, M.; TÓTH, G.; MONTANARELLA, L.; VAN WESEMAEL, B. Prediction of Soil Organic Carbon at the European Scale by Visible and Near InfraRed Reflectance Spectroscopy. **PLOS ONE**, v. 8, p. e66409, 2013.
- STOCKMANN, U.; CATTLE, S. R.; MINASNY, B.; MCBRATNEY, A. B. Utilizing portable X-ray fluorescence spectrometry for in-field investigation of pedogenesis. **CATENA**, v. 139, p. 220–231, 2016.
- SWANHART, S.; WEINDORF, D. C.; CHAKRABORTY, S.; BAKR, N.; ZHU, Y.; NELSON, C.; SHOOK, K.; ACREE, A. Soil salinity measurement via portable X-ray fluorescence spectrometry. **Soil Science**, v. 179, p. 417–423, 2014.
- TAVARES, R. T.; MOLIN, P. J.; NUNES, C. L.; ALVES, E. E.; MELQUIADES, F. L.; DE CARVALHO, H. W.; MOUAZEN, A. M. Effect of X-Ray Tube Configuration on Measurement of Key Soil Fertility Attributes with XRF. **Remote Sensing**, v. 12, p. 963, 2020a.
- TAVARES, T. R.; MOUAZEN, A. M.; ALVES, E. E. N.; DOS SANTOS, F. R.; MELQUIADES, F. L.; DE 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**, v. 10, p. 787, 2020b.
- TAVARES, T. R.; NUNES, L. C.; ALVES, E. E. N.; DE ALMEIDA, E.; MALDANER, L. F.; KRUG, F. J.; DE CARVALHO, H. W. P.; MOLIN, J. P. Simplifying Sample Preparation for Soil Fertility Analysis by X-ray Fluorescence Spectrometry. **Sensors**, v. 19, p. 5066, 2019.
- TEIXEIRA, A. F. DOS 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 . **Ciência e Agrotecnologia**, 2018.
- TERRA, F. S.; DEMATTÊ, J. A. M.; VISCARRA ROSSEL, R. A. Spectral libraries for quantitative analyses of tropical Brazilian soils: Comparing vis-NIR and mid-IR reflectance data. **Geoderma**, v. 255, p. 81–93, 2015.
- TERRA, F. S.; VISCARRA ROSSEL, R. A.; DEMATTÊ, J. A. M. Spectral fusion by Outer Product Analysis (OPA) to improve predictions of soil organic C. **Geoderma**, v.

335, p. 35–46, 2019.

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 Science Society of America Journal**, v. 79, p. 1375–1385, 2015.

UN-DESA POPULATION DIVISION. **World Population Prospects: The 2017 Revision**. 2017.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 55, p. 8331–8338, 2007.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Quimica Nova**, v. 32, p. 1278–1287, 2009.

VERBI, F. M.; PEREIRA-FILHO, E. R.; BUENO, M. I. M. S. Use of X-ray scattering for studies with organic compounds: A case study using paints. **Microchimica Acta**, v. 150, p. 131–136, 2005.

VILLAS-BOAS, P. R.; FRANCO, M. A.; MARTIN-NETO, L.; GOLLANY, H. T.; MILORI, D. M. B. P. Applications of laser-induced breakdown spectroscopy for soil analysis, part I: Review of fundamentals and chemical and physical properties. **European Journal of Soil Science**, v. 71, p. 789–804, 2019.

VILLAS-BOAS, P. R.; ROMANO, R. A.; DE MENEZES FRANCO, M. A.; et al. Laser-induced breakdown spectroscopy to determine soil texture: A fast analytical technique. **Geoderma**, v. 263, p. 195–202, 2016.

VISCARRA ROSSEL, R.A.; CATTLE, S. R.; ORTEGA, A.; FOUAD, Y. In situ measurements of soil colour, mineral composition and clay content by vis–NIR spectroscopy. **Geoderma**, v. 150, n. 3–4, p. 253–266, 2009.

VISCARRA ROSSEL, R. A.; MCGLYNN, R. N.; MCBRATNEY, A. B. Determining the composition of mineral-organic mixes using UV-vis-NIR diffuse reflectance spectroscopy. **Geoderma**, v. 137, p. 70–82, 2006.

VAN DER VOET, H. Comparing the predictive accuracy of models using a simple randomization test. **Chemometrics and Intelligent Laboratory Systems**, v. 25, p. 313–323, 1994.

VOHLAND, M.; BESOLD, J.; HILL, J.; FRÜND, H.-C. Comparing different

- multivariate calibration methods for the determination of soil organic carbon pools with visible to near infrared spectroscopy. **Geoderma**, v. 166, p. 198–205, 2011.
- VOHLAND, M.; LUDWIG, M.; THIELE-BRUHN, S.; LUDWIG, B. Determination of soil properties with visible to near- and mid-infrared spectroscopy: Effects of spectral variable selection. **Geoderma**, v. 223–225, p. 88–96, 2014.
- WAISER, T. H.; MORGAN, C. L. S.; BROWN, D. J.; HALLMARK, C. T. In Situ Characterization of Soil Clay Content with Visible Near-Infrared Diffuse Reflectance Spectroscopy. **Soil Science Society of America Journal**, v. 71, p. 389–396, 2007.
- WAN, M.; HU, W.; QU, M.; et al. Rapid estimation of soil cation exchange capacity through sensor data fusion of portable XRF spectrometry and Vis-NIR spectroscopy. **Geoderma**, v. 363, p. 114163, 2020.
- 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 characterization: Total carbon and total nitrogen. **Geoderma**, v. 243–244, p. 157–167, 2015.
- WANG, L.-L.; LIN, Y.-W.; WANG, X.-F.; Xiao, N.; Xu, Y.-D.; Li, H.-D.; Xu, Q.-S. A selective review and comparison for interval variable selection in spectroscopic modeling. **Chemometrics and Intelligent Laboratory Systems**, v. 172, p. 229–240, 2018.
- WANG, P.; WANG, J.; ZHANG, H.; DONG, Y.; ZHANG, Y. The role of iron oxides in the preservation of soil organic matter under long-term fertilization. **Journal of Soils and Sediments**, v. 19, p. 588–598, 2019.
- WANG, Y.; HUANG, T.; LIU, J.; Lin, Z.; Li, S.; Wang, R.; Ge, Y. Soil pH value, organic matter and macronutrients contents prediction using optical diffuse reflectance spectroscopy. **Computers and Electronics in Agriculture**, v. 111, p. 69–77, 2015.
- WARRICK, A. W.; NIELSEN, D. R. Spatial Variability of Soil Physical Properties in the Field. In: D. HILLEL (Org.); **Applications of Soil Physics**. p.319–344, 1980. New York: Academic Press.
- WEINDORF, D. C.; BAKR, N.; ZHU, Y. Advances in Portable X-ray Fluorescence (PXRF) for Environmental, Pedological, and Agronomic Applications. In: D. L. Sparks (Org.); **Advances in Agronomy**, v. 128, p.1–45, 2014. Academic Press.
- WEINDORF, D. C.; CHAKRABORTY, S.; LI, B.; DEB, S.; SINGH, A.; KUSI, N. Y. Compost salinity assessment via portable X-ray fluorescence (PXRF) spectrometry. **Waste Management**, v. 78, p. 158–163, 2018.

WEST, M.; ELLIS, A. T.; POTTS, P. J.; et al. Atomic spectrometry update—X-ray fluorescence spectrometry. **Journal of Analytical Atomic Spectrometry**, v. 27, p. 1603–1644, 2012.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 2, p. 37–52, 1987.

XIAOBO, Z.; JIEWEN, Z.; XINGYI, H.; YANXIAO, L. Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of ‘Fuji’ apple based on different PLS models. **Chemometrics and Intelligent Laboratory Systems**, v. 87, p. 43–51, 2007.

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 Syst.**, v. 3, p. 66, 2019.

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

ZIMMERMANN, B.; KOHLER, A. Optimizing savitzky-golay parameters for improving spectral resolution and quantification in infrared spectroscopy. **Applied Spectroscopy**, v. 67, p. 892–902, 2013.

APÊNDICE

APÊNDICE 1 – PUBLICAÇÕES E PARTICIPAÇÃO EM EVENTOS DURANTE O PERÍODO DO DOUTORADO.

Artigos publicados em periódicos:

DOS SANTOS, F.R.; DE OLIVEIRA, J.F.; BONA, E.; DOS 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**, v. 152, p. 104275, 2020.

TAVARES, T.R.; MOUAZEN, A.M.; ALVES, E.E.N.; DOS SANTOS, F.R.; MELQUIADES, F.L.; PEREIRA DE CARVALHO, H.W.; MOLIN, J.P. Assessing Soil Key Fertility Attributes Using a Portable X-ray Fluorescence: A Simple Method to Overcome Matrix Effect. **Agronomy-Basel**, v. 10, p. 787, 2020.

SARTO, G.; LOPES, F.; DOS SANTOS, F.R.; PARREIRA, P.S., DE ALMEIDA, L.C. Characterization OF Cu₂O/TiO₂NTs Nanomaterials using EDXRF, XRD and DRS for photocatalytic applications. **Applied Radiation and Isotopes**, v. 151, p. 124-128, 2019.

LISBOA NOGUEIRA G.A.; PARREIRA P.S.; RODRIGUES DOS SANTOS F.; MELQUIADES F.L. Determination of metal content in industrial powder ink and paint thickness over steel plates using X-Ray Fluorescence. **Applied Radiation and Isotopes**, v. 150, p. 168-174, 2019.

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. **Applied Radiation and Isotopes**, v. 130, p. 13-20, 2017.

RODRIGUES DOS SANTOS, F.; DE ALMEIDA, E.; DA CUNHA KEMERICH, P.D.; MELQUIADES, F.L. Evaluation of metal release from battery and electronic components in soil using SR-TXRF and EDXRF. **X-Ray Spectrometry**, v. 46, p. 512, 2017.

Trabalhos apresentados em eventos científicos:

SANTOS, F. R.; BONA, E. ; Melquiades, F. L. . Pre-treatment and variable selection of EDXRF spectral data for determination of soil TOC using PLSR. In: IV Escola de Inverno de Quimiometria, 2019, Porto Alegre.

DOS SANTOS, F.R.; BONA, E.; OLIVEIRA, J.F.; SANTOS, J.V.F.; BARBOSA, G.M.C.; MELQUIADES, F.L. Avaliação do poder preditivo de modelos PLS e SVM na determinação da saturação por base em solos agrícolas medidos por EDXRF. In: X Workshop de quimiometria, 2019, Salvador.

DOS SANTOS, F.R.; SANTOS, J.V.F.; OLIVEIRA, J.F.; BARBOSA, G.M.C.; Melquiades, F.L. Determination of exchangeable calcium by EDXRF and partial least square regression. In: XVI Conferencia Latinoamericana de Análisis por Técnicas de Rayos X, 2018, Pucón.

NOGUEIRA, G.A.L.; PARREIRA, P.S.; DOS SANTOS, F.R.; MELQUIADES, F.L. Determination of metal content in industrial powder ink and paint thickness over iron plates using X-ray fluorescence. In: Conferencia Latinoamericana de Análisis por Técnicas de Rayos X, 2018, Pucón.

LOPEZ, D.S.; GUAITA, M.G.D.; URBANO, A.; LOPES, F.; DOS SANTOS, F.R.; LOURENCO, J. Estudos estruturais e químicos de películas de óxidos litiados ternários por EDXRF e DRX. In: XVI Conferencia Latinoamericana de Análisis por Técnicas de Rayos X, 2018, Pucón.

SARTO, G.; ALMEIDA, L.C.; LOPES, F.; DOS SANTOS, F.R.; PARREIRA, P.S.; LOURENCO, J. Characterization of TiO₂/CuO₂ nanomaterials by using EDXRF, DRX and DRS for photocatalytic applications. In: XVI Conferencia Latinoamericana de Análisis por Técnicas de Rayos X, 2018, Pucón.

DOS SANTOS, F.R.; MORONA, F.; CARVALHO, H. W. P.; ALMEIDA, E.; MELQUIADES, F. L. Determination of soil organic matter with PXRF spectral data. In: 4 th Brazilian Soil Physics Meeting, 2017, Piracicaba.

Artigo aceito para publicação:

DOS SANTOS, F.R.; DE OLIVEIRA, J.F.; BONA, E.; BARBOZA, G.M.C.; MELQUIADES, F.L. Evaluation of pre-processing and variable selection on EDXRF spectral data with partial least square regression: a case of study for soil organic carbon prediction. **Spectrochimica Acta part B**.

Artigo submetido:

SANTANA, E. J.; DOS SANTOS, F. R.; MASTELINI, S. M.; MELQUIADES, F. L.; BARBON JR., S. Improved prediction of soil properties with Multi-target Stacked Generalisation on EDXRF spectra. **Artigo submetido para Chemometrics and Intelligent Laboratory Systems**.