



**UNIVERSIDADE FEDERAL DO CEARÁ**  
**CENTRO DE CIÊNCIAS AGRÁRIAS**  
**DEPARTAMENTO DE CIÊNCIA DO SOLO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DO SOLO**

**FRANCISCO MATEUS DA CUNHA SILVA**

**AVALIAÇÃO DE FÓSFORO EM SOLOS UTILIZANDO ESPECTROSCÓPIA DE  
REFLECTÂNCIA E ALGORITMOS QUIMIOMÉTRICOS**

**FORTALEZA**

**2025**

FRANCISCO MATEUS DA CUNHA SILVA

AVALIAÇÃO DE FÓSFORO EM SOLOS UTILIZANDO ESPECTROSCÓPIA DE  
REFLECTÂNCIA E ALGORITMOS QUIMIOMÉTRICOS

Dissertação apresentado ao Programa de Pós-Graduação em Ciência do Solo, da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência do solo. Área de concentração: Pedologia.

Orientador: Prof. PhD. Adunias dos Santos Teixeira.

FORTALEZA

2025

Dados Internacionais de Catalogação na Publicação  
Universidade Federal do Ceará  
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

---

- S58a Silva, Francisco Mateus da Cunha.  
Avaliação de fósforo em solos utilizando espectroscopia de reflectância e algoritmos quimiométricos /  
Francisco Mateus da Cunha Silva. – 2025.  
84 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Ciências Agrárias, Programa de  
Pós-Graduação em Ciência do Solo, Fortaleza, 2025.  
Orientação: Prof. Dr. Adunias dos Santos Teixeira.
1. Análise multivariada. 2. Sensoriamento proximal. 3. Fertilidade do solo. 4. Sensores. I. Título.  
CDD 631.4
-

FRANCISCO MATEUS DA CUNHA SILVA

AVALIAÇÃO DE FÓSFORO EM SOLOS UTILIZANDO ESPECTROSCÓPIA DE  
REFLECTÂNCIA E ALGORITMOS QUIMIOMÉTRICOS

Dissertação apresentado ao Programa de Pós-Graduação em Ciência do Solo, da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Ciência do solo. Área de concentração: Pedologia.

Aprovada em: 19 / 05 / 2025

BANCA EXAMINADORA

---

Prof. Dr. Adunias do Santos Teixeira (Orientador)  
Universidade Federal do Ceará (UFC)

---

Prof. Dr. Raul Shiso Toma  
Universidade Federal do Ceará (UFC)

---

Profa. Dra. Rafaella da Silva Nogueira  
Universidade da Integração Internacional da Lusofonia Afro-Brasileira (UNILAB)

## AGRADECIMENTOS

Aos meus pais, Ivanilda e Antônio, minha irmã Ivana e avó Antônia pelo apoio e acolhimento durante todas as etapas da minha vida, que, mesmo distantes, se faziam presentes e me fortaleciam para continuar essa trajetória, por serem solo seguro nas horas mais difíceis, mas, acima de tudo pela confiança e entusiasmo a cada etapa cumprida.

Ao meu orientador, Prof. Dr. Adunias dos Santos Teixeira, pelo incentivo, confiança e acompanhamento durante este trabalho, pelas oportunidades de muitos aprendizados e por toda a contribuição nesta pesquisa.

Aos membros titulares da banca examinadora, Prof. Dr. Raul Shiso Toma e Prof. Dra. Rafaella da Silva Nogueira, pelas valiosas sugestões e apoio ao desenvolvimento do trabalho realizadas tanto na qualificação quanto na defesa da dissertação.

Aos colegas do Grupo de Automação em Manejo Agrícola Desenvolvimento Tecnológico (GAMA-DT), cujos ensinamentos me ajudaram a nortear etapas importantes da constituição desse trabalho.

Aos amigos que encontrei nessa jornada, ao companheirismo nas horas dedicadas aos estudos e pelo apoio mútuo.

À Fundação Cearense de Apoio ao Desenvolvimento Científico e Tecnológico (FUNCAP), pelo auxílio financeiro concedido, essencial a condução desta pesquisa.

Aos docentes e servidores do Programa de Pós-Graduação em Ciência do Solo e do Departamento de Engenharia Agrícola da UFC, pelo suporte institucional, pela infraestrutura laboratorial e pela cordialidade no atendimento às demandas deste trabalho.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho.

Muito obrigado!

## RESUMO

A manutenção de níveis adequados de fósforo no solo é fator determinante para o desenvolvimento vegetal, influenciando processos metabólicos vitais. Sendo sua determinação por métodos laboratoriais baseados em extração com reagentes químicos, que além de gerarem resíduos, essas técnicas podem ser custosas e demoradas, o que torna sua aplicação contínua e regular inviável em muitos cenários agrícolas, dificultando a caracterização de áreas agrícolas com precisão. No presente estudo, objetiva-se avaliar o potencial preditivo da espectroscopia de reflectância na quantificação do teor de fósforo disponível em solos, por meio da aplicação de diferentes algoritmos quimiométricos e de transformações nos dados de reflectância espectral adquiridos por espectrorradiômetro óptico na faixa de 350–2500 nm e seleção de variáveis. Para a condução da pesquisa, foram coletadas 75 amostras de um NEOSSOLO QUARTZARENICO e 75 amostras de um ARGISSOLO ACINZENTADO no distrito de irrigação do Baixo Acaraú, sendo posteriormente tratadas com 15 diferentes concentrações do fertilizante comercial MAP em 5 repetições. Para a coleta dos dados espectrais, o sensor óptico FieldSpec Pro FR 3 foi utilizado. Nas análises qualitativas e quimiométricas, foram considerados os espectros de reflectância bruta, bem como os dados pré-processados por meio de normalização, suavização e derivação de Savitzky-Golay e remoção do contínuo. Para a determinação dos teores disponíveis de fósforo nas amostras, foram realizadas leituras com espectrofotômetro UV/VIS com base na solução extratora de Mehlich 1. Posteriormente, para a calibração dos modelos de predição do fósforo disponível foram empregados a regressão linear múltipla (MLR), a regressão em componentes principais (PCR) e a regressão por mínimos quadrados parciais (PLSR), considerando todos os comprimentos de onda como variáveis preditoras e ainda a seleção de variáveis por Stepwise-Forward. Os resultados mostram que a aplicação de normalização espectral aliada à regressão PLSR sobre o espectro completo (450–2450 nm) foi a abordagem mais eficiente para prever o fósforo disponível. No conjunto de dados RQO + PAcD, obteve-se  $R^2 = 0,95$ ,  $RMSE = 12,75 \text{ mg kg}^{-1}$  e  $RPD = 4,31$ , para o conjunto RQO,  $R^2 = 0,95$ ,  $RMSE = 8,30 \text{ mg kg}^{-1}$  e  $RPD = 4,66$ , e no PAcD,  $R^2 = 0,92$ ,  $RMSE = 17,76 \text{ mg kg}^{-1}$  e  $RPD = 3,30$ , superando consistentemente os modelos PCR em todos os casos. A seleção de variáveis por filtro de correlação e stepwise-forward reforçou a relevância das bandas associadas à matéria orgânica (650–850 nm), óxidos de ferro (850–900 nm) e argilominerais 2:1 (1400, 1900 e 2200 nm). Após a seleção de variáveis, o desempenho dos modelos mostrou-se heterogêneo e, em geral, inferior aos calibrados com todos os comprimentos de onda, com o PCR superando os demais métodos de regressão e apenas o

modelo normalizado do conjunto combinado (RQO+ PAcD) atingindo categoria excelente com  $R^2$  de 0,84, RMSE de 23,65 mg kg<sup>-1</sup> e RPD de 2,49.

**Palavras-chaves:** análise multivariada; sensoriamento proximal; fertilidade do solo; sensores.

## ABSTRACT

Maintaining adequate phosphorus levels in the soil is a determining factor for plant development, influencing vital metabolic processes. Its determination through laboratory methods based on chemical extraction reagents not only generates waste but also makes these techniques costly and time-consuming, rendering their continuous and regular application unfeasible in many agricultural scenarios. This hinders the precise characterization of agricultural areas. The present study aims to evaluate the predictive potential of reflectance spectroscopy for quantifying available phosphorus content in soils, through the application of different chemometric algorithms and transformations of spectral reflectance data acquired by an optical spectroradiometer within the 350–2500 nm range, along with variable selection. For this research, 75 samples of a Neossolo quartzarenico and 75 samples of an Argissolo acinentado were collected in the Baixo Acaraú irrigation district, and subsequently treated with 15 different concentrations of the commercial fertilizer MAP in 5 replications. For spectral data acquisition, the FieldSpec Pro FR 3 optical sensor was used. In addition to raw reflectance, spectral normalization, Savitzky-Golay filtering, and continuum removal were applied for qualitative and chemometric analyses. Available phosphorus content in the samples was determined using a UV/VIS spectrophotometer based on the Mehlich 1 extraction solution. For the calibration of the phosphorus prediction models, multiple linear regression (MLR), principal component regression (PCR), and partial least squares regression (PLSR) were employed, considering all wavelengths as predictor variables, along with variable selection via Stepwise-Forward. The results showed that spectral normalization combined with PLSR regression on the full spectrum (450–2450 nm) was the most efficient approach for predicting available phosphorus. In the combined RQO + PAcd dataset,  $R^2 = 0.95$ ,  $RMSE = 12.75 \text{ mg kg}^{-1}$ , and  $RPD = 4.31$  were obtained; for RQO alone,  $R^2 = 0.95$ ,  $RMSE = 8.30 \text{ mg kg}^{-1}$ , and  $RPD = 4.66$ ; and for PAcd,  $R^2 = 0.92$ ,  $RMSE = 17.76 \text{ mg kg}^{-1}$ , and  $RPD = 3.30$  consistently outperforming the PCR models in all cases. Variable selection through correlation filtering and stepwise-forward reinforced the relevance of bands associated with organic matter (650–850 nm), iron oxides (850–900 nm), and 2:1 clay minerals (1400, 1900, and 2200 nm). After variable selection, model performance was heterogeneous and generally inferior to models calibrated with all wavelengths, with PCR outperforming the other regression methods and only the normalized model of the combined dataset achieving an excellent category with  $R^2 = 0.84$ ,  $RMSE = 23.65 \text{ mg kg}^{-1}$ , and  $RPD = 2.49$ .

**Keywords:** multivariate analysis; proximal sensing; sensors.

## LISTA DE FIGURAS

Figura 1	– Espectro eletromagnético: principais regiões. ....	18
Figura 2	– Mapa de localização e solos do Distrito de Irrigação do Baixo Acaraú. ....	29
Figura 3	– Mapa de localização dos lotes de coleta dos solos. ....	31
Figura 4	– Coleta das amostras de solo. ....	32
Figura 5	– Etapas de tratamento das amostras de solo: Peneiramento (A), pesagem e divisão de subamostras (B), adubação e incubação (C) e secagem (D). ....	32
Figura 6	– Esquema de aquisição de dados hiperespectrais (350–2500 nm) nas amostras de solo. ....	35
Figura 7	– Histogramas da distribuição dos teores de fósforo (mg/kg) nas classes de solo. ....	36
Figura 8	– Fluxograma metodológico para obtenção dos modelos de predição e validação dos atributos estudados. ....	38
Figura 9	– Amostras de Solos e Espectros de Reflectância: (A) Neossolo Quartzarênico (RQo), (B) Argissolo Acinzentado Distrófico (PAcd) e (C) Curvas Espectrais Representativas dos Solos. ....	41
Figura 10	– Boxplot da distribuição da Reflectância para os solos RQo e PAcd. ....	42
Figura 11	– Tratamentos dos dados espectrais médios das amostras dos solos RQo e PAcd Reflectância bruta(A), Normalização (B), Suavização da reflectância por Savitzky-Golay (C) e Continuum remove (D). ....	44
Figura 12	– Coeficientes de correlação de Pearson (r) e mapa de calor entre bandas espectrais VisNIR-SWIR (450 – 2.450 nm) como reflectância bruta e o teor de fósforo nas amostras dos solos RQo (A), no PAcd (B) e RQo + PAcd (C). ....	47
Figura 13	– Variância explicada acumulada e redução do RMSE em função do número de componentes. ....	50
Figura 14	– Loadings espectrais nos componentes principais do modelo PCR – Conjunto de dados AB (Refletância Bruta). ....	52
Figura 15	– Loadings espectrais nos componentes principais do modelo PLSR – Conjunto de dados AB (Refletância Bruta). ....	54
Figura 16	– Impacto do pré-processamento e do método de regressão no ranking de modelos para o conjunto de dados AB. ....	55

Figura 17 – Impacto do pré-processamento e do método de regressão no ranking de modelos para o conjunto de dados A. ....	56
Figura 18 – Impacto do pré-processamento e do método de regressão no ranking de modelos para o conjunto de dados B. ....	57
Figura 19 – Previsões do modelo PLSR com pré-tratamento espectral normalizado nos conjuntos AB de calibração e validação. ....	57
Figura 20 – Previsões do modelo PLSR com pré-tratamento espectral normalizado nos conjuntos A de calibração e validação. ....	58
Figura 21 – Previsões do modelo PLSR com pré-tratamento espectral normalizado nos conjuntos B de calibração e validação. ....	58
Figura 22 – Distribuição e quantidade de variáveis selecionadas por conjunto de dados e pré-tratamentos espectrais. ....	59
Figura 23 – Impacto do pré-processamento e do método de regressão no ranking de modelos com seleção de variáveis do conjunto de dados AB. ....	67
Figura 24 – Impacto do pré-processamento e do método de regressão no ranking de modelos com seleção de variáveis do conjunto de dados A. ....	68
Figura 25 – Impacto do pré-processamento e do método de regressão no ranking de modelos com seleção de variáveis do conjunto de dados B. ....	68
Figura 26 – Previsões de calibração e validação dos três modelos melhor ranqueados: (A) MLR com pré-tratamento espectral normalizado no conjunto RQO + PAcD; (B) PCR com pré-tratamento SG no conjunto RQO; (C) PCR com pré-tratamento CR no conjunto PAcD. ....	69

## LISTA DE TABELAS

Tabela 1	– Características químicas e físicas dos solos Neossolo Quartzarênico Órtico (RQo) e Argissolo Acinzentado Distrófico (PACd) (0–10 cm). .....	30
Tabela 2	– Descrição dos fertilizantes e concentrações de aplicação nas amostras de solo. ....	33
Tabela 3	– Distribuição quantitativa dos conjuntos amostrais utilizados para calibração e validação. ....	35
Tabela 4	– Teor de fósforo ( $\text{mg kg}^{-1}$ ) em Neossolo Quartzarênico (RQo) e Argissolo Acinzentado sob diferentes doses de aplicação de MAP.....	39
Tabela 5	– Tipos e Identificação dos Tratamentos Aplicados nos Modelos.....	49
Tabela 6	– Desempenho dos modelos de regressão PCR para estimativa de fósforo disponível durante a calibração e validação com o espectro completo.....	50
Tabela 7	– Desempenho dos modelos de regressão PLSR para estimativa de fósforo disponível durante a calibração e validação com o espectro completo. ....	53
Tabela 8	– Desempenho dos modelos de regressão MLR para estimativa de fósforo disponível para o conjunto de calibração e validação após a seleção de variáveis.....	61
Tabela 9	– Descrição dos modelos de melhor desempenho calibrados por MLR para os Conjuntos AB, A e B.....	62
Tabela 10	– Desempenho dos modelos de regressão PCR para estimativa de fósforo disponível para o conjunto de calibração e validação após a seleção de variáveis.....	63
Tabela 11	– Descrição dos modelos de melhor desempenho calibrados por PCR para os Conjuntos AB, A e B. ....	64
Tabela 12	– Desempenho dos modelos de regressão PLSR para estimativa de fósforo disponível para o conjunto de calibração e validação após a seleção de variáveis.....	65
Tabela 13	– Descrição dos modelos de melhor desempenho calibrados por PLSR para os Conjuntos AB, A e B. ....	66

## LISTA DE ABREVIATURAS E SIGLAS

CRR	Continuum Removed Reflectance
HCA	Análise de Agrupamento Hierárquico
IDE	Ambiente de Desenvolvimento Integrado
MLR	Regressão Linear Múltipla
MIR	Infravermelho médio
ND	Número Digital
NIR	Região espectral do Infravermelho Próximo
NT	Nitrogênio total
PCA	Análise de Componentes Principais
PCR	Regressão em Componentes Principais
PLS	Regressão por Mínimos Quadrados Parciais
REM	Radiação Eletromagnética
RMSE	Raiz do Erro Quadrático Médio
SMO	Smoothing
RPD	Razão da Performance do Desvio
SVD	Decomposição de Valores Singulares
SVG	Suavização Savitzky-Golay
SWIR	Região espectral do Infravermelho de Ondas Curtas
TFSA	Terra Fina Seca ao Ar
UV	Ultravioleta
VANT	Veículo Aéreo Não Tripulado
VIS	Região espectral do visível

## LISTA DE SÍMBOLOS

$R^2$	Coeficiente de determinação
$R^2_{\text{ajust}}$	Coeficiente de determinação ajustado
B	Coeficiente de regressão
E	Erro
$\sigma$	Desvio padrão

## SUMÁRIO

1	<b>INTRODUÇÃO</b> .....	15
2	<b>REFERENCIAL TEÓRICO</b> .....	16
2.1	<b>Sensoriamento remoto proximal</b> .....	16
2.2	<b>Comportamento espectral de solos</b> .....	19
2.3	<b>Quimiometria</b> .....	21
2.3.1	<i>Regressão linear múltipla</i> .....	22
2.3.2	<i>Regressão por componentes principais</i> .....	23
2.3.3	<i>Regressão por mínimos quadrados parciais</i> .....	24
2.4	<b>Avaliação espectral de solos</b> .....	25
2.5	<b>Dinâmica do fósforo em solos</b> .....	26
2.5.1	<i>Fatores de intensidade, qualidade e capacidade tampão</i> .....	26
2.5.2	<i>Adsorção e precipitação do fósforo</i> .....	27
3	<b>HIPOTESE</b> .....	28
4	<b>OBJETIVOS</b> .....	28
4.1	<b>Objetivos gerais</b> .....	28
4.2	<b>Objetivos específicos</b> .....	28
5	<b>METODOLOGIA</b> .....	29
5.1	<b>Descrição da área de coleta</b> .....	29
5.2	<b>Coleta e preparação das amostras de solo</b> .....	31
5.3	<b>Determinação de fósforo disponível</b> .....	33
5.4	<b>Avaliação espectral das amostras de solo</b> .....	34
5.5	<b>Análise quimiométrica e resposta espectral do solo</b> .....	35
5.5.1	<i>Pré-tratamento dos dados espectrais</i> .....	36
5.5.2	<i>Calibração e validação dos modelos preditivos</i> .....	36
5.6	<b>Implementação de Rotinas Estatísticas e Quimiométricas no R</b> .....	38
6	<b>RESULTADOS E DISCUSSÃO</b> .....	39
6.1	<b>Concentrações de fósforo obtidas com diferentes doses de fertilizantes em solos</b> .....	39
6.2	<b>Análise qualitativa dos dados hiperespectrais</b> .....	40
6.3	<b>Pré-tratamento espectral</b> .....	43

6.4	Correlação de Pearson .....	45
6.5	Estimativa do teor de fósforo disponível o espectro completo (450-2450 nm) .....	48
6.5.1	<i>Regressão por componentes principais utilizando o espectro completo .....</i>	48
6.5.2	<i>Regressão por mínimos quadrados parciais utilizando o espectro completo ...</i>	53
6.5.3	<i>Ranking de seleção de modelos utilizando todo o espectro (450 - 2450nm) ....</i>	55
6.6	Estimativa do teor de fósforo disponível com seleção de variáveis .....	59
6.6.1	<i>Estimação de fósforo no solo usando regressão linear múltipla (MLR) com variáveis selecionadas .....</i>	60
6.6.2	<i>Estimação de fósforo no solo usando regressão por componentes principais (PCR) com variáveis selecionadas .....</i>	62
6.6.3	<i>Estimação de fósforo no solo usando regressão por mínimos quadrados parciais (PLSR) com variáveis selecionadas .....</i>	64
6.6.4	<i>Ranking de seleção de modelos calibrados a partir da seleção de variáveis ....</i>	66
7	CONCLUSÃO .....	69
	REFERÊNCIAS .....	71
	APÊNDICE A – PARÂMETROS DE REGRESSÃO UTILIZADAS NOS MODELOS PREDITIVOS DE FÓSFORO PARA O NEOSSOLO QUARTZARÊNICO .....	79
	APÊNDICE B – PARÂMETROS DE REGRESSÃO UTILIZADAS NOS MODELOS PREDITIVOS DE FÓSFORO PARA O ARGISSOLO ACINZENTADO .....	81
	APÊNDICE C – PARÂMETROS DE REGRESSÃO UTILIZADAS NOS MODELOS PREDITIVOS DE FÓSFORO PARA O NEOSSOLO QUARTZARÊNICO E ARGISSOLO QUARTZARÊNICO.....	83

## 1 INTRODUÇÃO

A manutenção de níveis adequados de fósforo no solo é fator determinante para o desenvolvimento vegetal, influenciando processos metabólicos vitais como fotossíntese, transferência de energia (ATP) e síntese de ácidos nucleicos. Em solos tropicais, onde frequentemente se observam baixos teores de matéria orgânica e alta adsorção do nutriente em frações argilosas, a deficiência de fósforo pode limitar drasticamente o crescimento radicular, a floração e o rendimento final das culturas. Dessa forma, a compreensão do teor de fósforo disponível, em conjunto com outras características físicas, químicas e biológicas do solo torna-se fundamental para embasar práticas de manejo que garantam produtividade e sustentabilidade (CONKLIN, 2013).

Convencionalmente, a quantificação do fósforo disponível recorre a métodos laboratoriais baseados em extração com reagentes químicos, sendo o mais utilizado o Mehlich 1, seguidos de análise colorimétrica. Apesar de precisos, esses procedimentos são destrutivos, demandando a coleta e preparação de amostras representativas e por fim sua determinação (WALVOORT; MCBRATNEY, 2001). Dessa forma, além de gerarem resíduos, essas técnicas podem ser custosas e demoradas (VISCARRA ROSSEL et al., 2016).

A busca por métodos mais rápidos, sustentáveis e econômicos tem motivado o desenvolvimento de abordagens alternativas para análise de solos. Nesse contexto, a espectroscopia de refletância emerge como uma solução promissora. Essa técnica permite a avaliação simultânea de diversos parâmetros físicos, químicos e biológicos do solo, com uma preparação mínima de amostras e sem gerar resíduos perigosos. Além disso, proporciona a possibilidade de análises rápidas e não destrutivas, facilitando o monitoramento contínuo do solo no campo.

Diversos estudos têm empregado espectroscopia no infravermelho próximo e visível (350–2500 nm) para caracterizar e quantificar diretamente constituintes opticamente ativos do solo como carbono total e orgânico, nitrogênio total e fração argila, explorando suas bandas de absorção específicas às ligações C–H, N–H e O–H. Paralelamente, outras pesquisas têm buscado estimar de forma indireta constituintes não opticamente ativos (por exemplo, nutrientes, capacidade de troca catiônica e pH) por meio de modelos estatísticos que aproveitam a covariação dessas constituintes com os componentes opticamente ativos, embora tais análises apresentem maior variabilidade.

Embora o fósforo apresente distribuição complexa no solo manifestando-se em formas orgânicas e inorgânicas, ele mantém correlações significativas com constituintes

ópticamente ativos, tais como matéria orgânica, fração argila e óxidos de ferro (COUTINHO et al., 2019). Dessa forma, existe um potencial uso da espectroscopia de reflectância para estimar o P disponível em solos.

Para garantir a confiabilidade das previsões de fósforo disponível, é fundamental avaliar e otimizar os métodos de pré-processamento espectral. Estudos demonstram que escolhas inadequadas nessa etapa podem atenuar sinais importantes ou amplificar ruídos indesejados, comprometendo a calibração de modelos quimiométricos e conseqüentemente, a precisão das estimativas. Dessa forma, a avaliação de diferentes abordagens, envolvendo tanto a aplicação de transformações nos dados espectrais quanto a seleção das faixas espectrais mais informativas, é fundamental para o desenvolvimento de modelos robustos e com bom desempenho em distintos tipos de solos.

No entanto, apesar do potencial dessas técnicas espectrais, a quantidade massiva de dados gerados exige o uso de métodos estatísticos robustos para interpretação adequada. Técnicas quimiométricas, como regressão por mínimos quadrados parciais (PLSR) (ARAÚJO et al., 2015; JOHNSON et al., 2019; REYES; LIESS, 2024) e regressão em componentes principais (PCR) (MUNNAF; NAWAR; MOUAZEN, 2019), têm se mostrado eficazes na extração de informações significativas, permitindo a construção de modelos preditivos para diferentes atributos do solo. Dessa forma, a espectroscopia de reflectância, aliada a um processamento de dados adequado, configura-se como uma estratégia potencial para o manejo sustentável e eficiente da fertilidade do solo, em especial o fósforo.

## **2 REFERENCIAL TEÓRICO**

### **2.1 Sensoriamento remoto e proximal**

Nos últimos anos, o sensoriamento remoto e proximal tem sido utilizados como uma alternativa a caracterização de diferentes atributos do solo em detrimento de análises físicas e químicas laboratoriais padrões, se destacando pela rapidez e a aquisição de uma grande quantidade de informações a um custo relativamente baixo (VISCARRA-ROSSEL, 2006).

De forma mais abrangente, o sensoriamento remoto pode ser entendido como processo de obtenção de informações sem que haja contato na superfície física de objetos alvos sejam eles aéreas (uso de VANTS e aeronaves), orbitais (satélites) ou qualquer outra plataforma que permita a aquisição de dados sem o pleno contato com o alvo (NOVO, 2010; LORENZZETTI, 2015). Em contrapartida, o sensoriamento proximal as análises ocorrem a

partir da utilização de equipamentos para determinação de atributos em contato direto com a superfície do alvo, podendo estes serem a nível de laboratório (em condições controladas) ou de campo podendo sofrer interferência de diferentes fatores.

Embora possuam metodologias e equipamentos distintos, os princípios físicos por trás desses métodos de aquisição de dados são semelhantes e muitas vezes utilizados de forma combinados ou complementar para a aquisição de dados. Nesse princípio, as informações podem ser obtidas a partir de sensores capazes de detectar e mensurar mudanças que um determinado objeto impõe a diferentes campos de força que o circundam, sendo o principal deles a radiação eletromagnética (REM).

A radiação eletromagnética são ondas que se propagam através do espaço, transportando energia sem transportar matéria, podendo ser propagada até mesmo no vácuo. Elas são formadas pela combinação de campos elétricos e magnéticos oscilantes, que estão sempre em fase entre si (NOVO, 2010; LORENZZETTI, 2015). As ondas eletromagnéticas são classificadas de acordo com sua frequência e comprimento de onda. A frequência representa o número de oscilações completas que ocorrem em um segundo em relação a um ponto de referência, geralmente expressa em ciclos por segundo ou Hertz. O comprimento de onda representa a distância entre duas cristas sucessivas da onda, sendo expressa em metros.

A frequência e o comprimento de onda de uma onda eletromagnética estão inversamente relacionados, ou seja, quanto maior a frequência, menor o comprimento de onda. Dessa forma, ondas eletromagnéticas com frequências mais baixas têm comprimentos de onda mais longos, enquanto ondas eletromagnéticas com frequências mais altas têm comprimentos de onda mais curtos.

As ondas eletromagnéticas podem ser ainda divididas em diferentes faixas de frequência, chamadas de espectro eletromagnético. Cada faixa do espectro eletromagnético é caracterizada por um comprimento de onda e uma frequência específicos, e que devido a essas características promovem diferentes efeitos da REM com a matéria, sendo as principais a radiação gama, raios x, ultravioleta e radiação visível e radiação infravermelha.

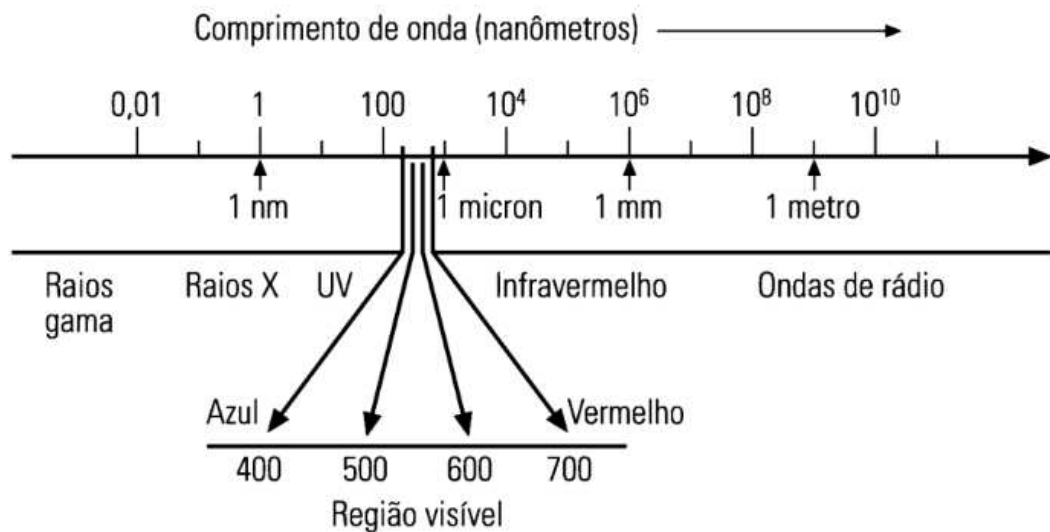
A radiação gama é emitida por materiais radioativos e pelo sol, se caracterizando por possuir alta frequência, devido a isso possui alta energia atribuída, estando na faixa espectral abaixo de 0,003 nm. As ondas do raio X, possui alta frequência e menor comprimento de onda estando localizado na faixa espectral entre 0,003 a 10nm. A radiação ultravioleta esta compreendida na faixa entre 10 a 400nm, essa faixa pode ser dividida em três banda a UV próximo (320 a 400 nm), UV distante (280 a 320nm) e UV máximo (10 a 280nm). O sol é o principal emissor natural dessa faixa, entretanto parte dessa radiação sofre interferência

atmosfera e é absorvida pela camada de Ozônio.

Diversos autores destacam a importância das faixas espectrais do visível, infravermelho próximo e infravermelho médio para estudos para caracterização de diferentes propriedades dos solos, incluindo física (HAGHI; PÉREZ-FERNÁNDEZ; ROBERTSON 2021), química DEMATTÊ et al., 2017; SHI et al., 2020; FARIDA et al., 2024 e mineralógica (COBLINSKI et al., 2021).

A radiação do visível (VIS) compreendida na faixa espectral entre 400 e 700 nm. Sendo essa faixa de comprimento de onda visível aos seres humanos e que passam a sensação de cor (Figura 1). Nessa faixa espectral ocorre a interação da energia a nível de átomo associados a transições eletrônicas, sendo possível derivar informações indiretamente, por efeito da aquisição específica de energia pelos íons (NOVO, 2010; Soriano-Disla et al., 2014).

**Figura 1.** Principais regiões do espectro eletromagnético.



Fonte: NOVO (2010).

A radiação infravermelha próximo (NIR), possui comprimento de onda entre 700 a 2500nm, envolve sobreposições e combinações de processos vibrações da molécula em seu estado fundamental. Por outro lado, o infravermelho médio (MIR), que abrange comprimentos de onda entre 2500 e 25000 nm, captura principalmente as vibrações fundamentais das moléculas.

Devido a sua natureza, cada alvo possui diferentes comportamentos à medida que interagem com a REM, tais como capacidade de absorção, emissão, transmissão e reflexão, sendo essas características influenciadas pela estrutura atômica e molecular de cada alvo,

podendo assim ser utilizados para fins analíticos (STENBERG et al., 2010; RIZZO, 2011).

A energia incidente (irradiância) interage com todo o constituído da matéria e de forma simultânea no alvo, entretanto, cada comprimento de onda possui uma afinidade e energia suficiente para atingir uma dada estrutura ou elemento específico (NOVO, 2010; LORENZZETTI, 2015). Dessa forma, ao atingir a superfície de um alvo parte da energia proveniente da REM é absorvida ou ainda transmitida para outros alvos e a energia excedente é refletida ou emitida.

Nesse sentido, a energia refletida por um alvo retrata o comportamento espectral do alvo, podendo este ser captada por sensores que possuem a capacidade de distinguir a variação no nível de energia refletida, emitida ou espalhada de um alvo ou superfície, varrendo uma ampla gama do espectro eletromagnético e registram quantitativamente a intensidade da energia refletida de um material, gerando um gráfico denominada de assinatura espectral de alvos, uma representação das amplitudes ou intensidades das frequências ao longo da região da radiação eletromagnética.

## **2.2 Comportamento espectral de solos**

A singularidade e variabilidade de cada tipo de solo se manifesta por meio de características distintas quanto sua composição mineral orgânica e inorgânica (BO STENBERG et al., 2010; WETTERLIND; STENBERG; ROSSEL, 2013). Devido a essas características únicas, sua interação com a radiação eletromagnética possibilita a caracterização do comportamento espectral em diversas faixas de comprimentos de onda e em diferentes intensidades, seja elas nas regiões espectrais do visível (VIS), infravermelho próximo (NIR) e infravermelho de ondas curtas (SWIR).

Ao se analisar cada faixa ou região espectral permite, a obtenção de informações quanto a natureza de uma amostra suas propriedades físicas, químicas, mineralógicas ou ainda biológicas dos solos.

A água exerce uma influência significativa na sua reflectância espectral dos solos em diferentes regiões do espectro eletromagnético, entretanto se destacam as regiões de absorções em 1400 nm e 1900 nm relacionadas à água absorvida e estruturalmente ligada a diferentes minerais (BO STENBERG et al., 2010; WETTERLIND; STENBERG; ROSSEL, 2013), na região do visível a presença da água no solo promove a redução da reflectância influenciando diretamente a cor observada destacando tons mais escuros (BAUMGARDNER et al., 1985).

Os minerais presentes no solo também influenciam a reflectância espectral. Minerais contendo ferro, como hematita e goethita, além de oxihidróxidos, apresentam bandas de absorção características na região visível do espectro, sendo sua presença e quantidade desses minerais podem ser estimadas usando a análise espectral (WETTERLIND; STENBERG; ROSSEL, 2013; CHAUHAN et al., 2021; DI RAIMO et al., 2022).

A matéria orgânica do solo, por sua vez, revela bandas de absorção distintas, mas amplas, na região visível, sendo dominadas por cromóforos e pela tonalidade mais escura (BO STENBERG et al., 2010).

Atualmente, os protocolos para análise físico-químicas dos solos baseiam-se na coleta de amostras em campo se caracterizando por serem destrutivas, necessidade de um tratamento prévio como é o caso da secagem, peneiramento, destorroamento e maceração, além da utilização de produtos químicos para diferentes fins seja eles extração, indicadores e titulação (WALVOORT; MCBRATNEY, 2001). Se caracterizando por serem processos dispendiosos, onerosos e potencialmente perigosos para quem manuseia, além da aquisição de resíduos que necessitam do descarte correto, caso contrário, se tornam potencialmente contaminadores do meio ambiente.

Em contraste com métodos tradicionais de análise de solo, a análise espectroscópica é um método não destrutivo se realizada diretamente no campo, ou ainda utilizando pequenas quantidades de solo quando realizado em laboratório, minimizando os impactos relacionados a coleta. Estas informações podem ainda serem utilizados tanto de forma superficial (para determinação de teores) ou ainda em diferentes profundidades do perfil do solo (para a classificação do tipo de solo). A análise espectroscópica é um método rápido, relativamente barato, possibilitando que com uma única leitura seja possível avaliar diferentes parâmetros simultaneamente, além disso, requerem uma baixa preparação de amostras (NANNI; DEMATTÊ, 2006).

Coutinho et al. (2019) realizaram um estudo comparativo para avaliar a influência do preparo de amostras de solo na análise de diferentes características físicas e químicas via espectroscopia Vis-NIR e Mir. Os autores concluíram que as amostras sem preparo apresentaram resultados equivalentes às amostras peneiradas em diferentes aberturas de malha, destacando a viabilidade de utilizar amostras de solo sem preparo para análises espectrais. A dispensabilidade do peneiramento pode simplificar e agilizar o processo analítico, além de reduzir custos, sendo uma temática que ainda deve ser estudado em diferentes condições e solos.

A mensuração da refletância dos solos pode ser realizada de várias maneiras, e duas abordagens proeminentes são relatadas com mais ênfase na literatura sendo eles o

sensoriamento remoto, por meio de sensores orbitais em satélites, e o sensoriamento proximal, utilizando espectrorradiômetros ou outros sensores ópticos em diferentes configurações. Esses dispositivos podem estar fixos em bancadas de laboratório, proporcionando condições controladas para medições precisas, ou serem portáteis, permitindo a coleta de dados in situ, diretamente no campo, a escolha entre sensoriamento remoto e proximal depende da aplicação específica e necessidades do trabalho.

### 2.3 Quimiometria

As modernizações instrumentais e computacionais permitiram o desenvolvimento de métodos de avaliação de dados cada vez mais complexos. Assim, a quimiometria surgiu da crescente demanda por métodos e técnicas que pudessem lidar com a complexidade e a quantidade cada vez maior de dados gerados nas diferentes áreas química. A quimiometria possui como base abordagens multivariadas para a interpretação de dados químicos, dessa forma, permitindo a extração de padrões, identificação de correlações, modelagem de relações entre variáveis (BARRA et al., 2021).

As técnicas quimiométricas foram ainda impulsionadas pela criação de softwares com interface intuitiva facilitando a sua aplicação, que até então eram necessários um conhecimento matemático, estatístico e computacional prévio e específico para o processamento de dados químicos (HAIR et al, 2009).

Sendo assim, embora as técnicas espectroscópicas permitam analisar diferentes atributos dos solos em diferentes faixas espectrais, geram uma grande quantidade de dados que muitas das vezes promovem mudanças imperceptíveis de serem observadas visualmente, necessitando assim, de técnicas estatísticas e processamento de dados adequadas que promovam uma melhor compreensão das informações.

Dentre as principais técnicas de modelagem estatística a quimiometria é priorizada em estudos espectrais em solos como por ARAÚJO et al., 2015; JOHNSON et al., 2019; MUNNAF; NAWAR; MOUAZEN, 2019; REYES; LIESS, 2024. Pois permitem extrair informações relevantes da base de dados espectrais e associar a diferentes parâmetros do solo, permitindo a construção de modelos estatísticos utilizados para encontrar respostas quantitativas e/ou qualitativas.

Os principais modelos quimiométricos possuem como base as análises multivariadas de dados, podendo ser considerada uma subárea voltada a análises de dados químicos. Os modelos quimiométricos podem ser divididos em técnicas ou algoritmos de

análise supervisionados e não supervisionados. As técnicas não supervisionadas ou ainda análise exploratórias de dados, não utiliza um conjunto de dados de treinamento que permite distinguir as classes ou parâmetros existentes dentro de um conjunto de dados, tendo como principal objetivo extrair informações do conjunto de dados a partir de relações entre as observações e variáveis estudadas.

Um dos principais exemplos da análise exploratória de dados incluem a análise de componentes principais (PCA), análise de agrupamento hierárquico (HCA) e análise de fatores, modelos capazes de identificar padrões e destacar as características mais relevantes em um conjunto de dados, reduzindo sua dimensionalidade ou ainda facilitando a visualização e interpretação ao criar grupos com base em suas similaridades ou dissimilaridades.

As técnicas supervisionadas em quimiometria, também conhecidas como calibração multivariada, se destaca na construção de modelos preditivos a partir da correlação entre informações instrumentais e parâmetros de interesse, dessa forma, dois conjuntos de dados distintos são necessários (FARIA; SJÖSTRÖM; ERIKSSON, 2001). O primeiro conjunto representa a matriz de dados com a resposta de aplicação da análise instrumental (no tocante a análises espectroscópicas são os dados de intensidade de energia refletida nos diferentes comprimentos de onda), enquanto o segundo conjunto de dados representa os valores respostas determinados por métodos de referências padronizados (FARIA; SJÖSTRÖM; ERIKSSON, 2001).

Dentre os principais modelos supervisionados possuem como principal finalidade a calibração de modelos matemáticos para gerar respostas quantitativas de parâmetros de interesse como a regressão linear múltipla (MLR), a regressão em componentes principais (PCR) e regressão por mínimos quadrados parciais (PLSR).

Assim, a capacidade de destacar informações das análises instrumentais e relacionar com parâmetros químicos de interesse, permite a utilização de técnicas quimiométricas supervisionadas seja ideal para análise de dados espectrais e utilizada na predição de diferentes atributos do solo.

### **2.3.1 Regressão linear múltipla**

A regressão linear múltipla possui como objetivo encontrar coeficientes  $\beta_0, \beta_1, \beta_2 \dots \beta_m$  que minimizem a diferença entre os valores reais das variáveis dependentes e os valores preditos pelo modelo, conforme a equação 1.

$$Y = \beta_0 + \beta_1 * x_1 + \beta_2 * x_2 \dots \beta_m * x_m + \varepsilon \quad \text{Eq. 1}$$

Onde, os coeficientes angulares ( $\beta$ ) ajusta a inclinação da reta em relação a uma variável independente específica. O intercepto ( $\beta_0$ ) representa o ponto onde a reta cruza o eixo Y, ou seja, o valor previsto de Y quando todas as variáveis X são iguais a zero.

### 2.3.2 Regressão por componentes principais

A regressão por componentes principais (PCR) tem como objetivo substituir a matriz de dados originais por uma nova matriz de pontuações (scores) durante o processo de predição e regressão. Essas pontuações são obtidas através da análise de componentes principais (PCA) e representam uma versão compacta dos dados originais, capturando a maior parte da variação presente neles.

Conforme Wehrens (2020), a regressão por componentes principais (PCR) oferece duas vantagens significativas. Primeiramente, as pontuações geradas por esse método são ortogonais entre si. Dessa forma, problemas de multicolinearidade entre as variáveis independentes, frequentemente observados em conjuntos de dados multivariados, são atenuados. Na medida em que as pontuações representam diferentes aspectos da variabilidade dos dados, sem apresentar correlação entre si. Além disso, outra vantagem importante da PCR é a redução na dimensionalidade dos dados, pois o número de pontuações selecionadas geralmente é inferior ao número de variáveis originais.

Para a definição dos componentes principais as matrizes de dados X e Y são normalizadas para garantir que tenham média igual a zero. A partir desses dados normalizados, as matrizes de covariância são calculadas. Em seguida, são calculados os autovalores e autovetores das matrizes de covariância, que são ordenados de forma crescente. Os componentes principais são calculados multiplicando a diferença entre a matriz de dados original (X) e sua média ( $\bar{X}$ ) pelo conjunto de autovetores (V), conforme a equação 2.

$$T = (X - \bar{X})V \quad \text{Eq. 2}$$

Após a seleção dos componentes principais, os coeficientes de regressão são estimados utilizando uma regressão linear entre os componentes principais (T) e a matriz de variável dependente (Y), conforme a equação 3.

$$B = (T^T - T)^{-1}T^TY \quad \text{Eq. 3}$$

### 2.3.3 Regressão por mínimos quadrados parciais

A regressão por mínimos quadrados parciais (PLSR) é uma técnica quimiométrica multivariada que combina os princípios da PCR com a MLR. Na PLSR, assim como no PCR são definidas variáveis latentes ortogonais a partir das matrizes de dados originais X e Y, possuindo como objetivo, maximizar a correlação entre essas matrizes de forma que as novas variáveis possuam a maior parte da variância dos dados originais (WEHRENS, 2020).

Uma diferença importante entre PLSR e PCR é o método de extração dos componentes. Enquanto os componentes do PCR são obtidos em uma única etapa usando decomposição de valores singulares (SVD), os componentes do PLSR são extraídos sequencialmente em cada iteração do algoritmo (WEHRENS, 2020). Durante cada iteração do algoritmo PLSR, a variação associada ao componente estimado é removida dos dados em um processo chamado de deflação. Durante o processo de iteração, os pesos (w) são recalculados de forma a maximizar a correlação entre a matriz de dados X e Y, atualizando as pontuações latentes (t), como também a matriz de resíduos (E), conforme a equação 4.

$$t = X * w = E * w \quad \text{Eq. 4}$$

Ao final do processo de iteração os valores t são utilizados para gerar a matriz T. Os valores de cargas para a matriz X (p) e matriz Y (q) são calculados pela regressão dos dados originais a partir das pontuações latentes, conforme as equações 5 e 6.

$$p = E^t t / (t^T t) \quad \text{Eq. 5}$$

$$q = F^t t / (t^T t) \quad \text{Eq. 6}$$

Posteriormente, os coeficientes de regressão são calculados conforme a equação 7.

$$B = (T^T - T)^{-1}T^TY \quad \text{Eq. 7}$$

Onde, T é a matriz de escores latentes e Y é a matriz de variável dependente.

## 2.4 Avaliação espectral de solos

Embora os íons e cátions disponíveis para as plantas sejam vitais para seu desenvolvimento, sua detecção direta no solo na região espectral do Vis-NIR (Visível e Infravermelho Próximo) é dificultada devido às suas características de absorção e respostas espectrais baixas nessas faixas espectrais (WETTERLIND; STENBERG; ROSSEL, 2013; Soriano-Disla et al., 2014). No entanto, quando se ligam a outros elementos presentes no solo mensuráveis nessas faixas espectrais, como óxidos e oxihidróxidos de alumínio (Al), ferro (Fe) ou manganês (Mn), argilominerais e matéria orgânica, umidade e tamanho de partículas ocorre uma modificação na resposta espectral padrão dos elementos (SORIANO-DISLA et al., 2014; MUKHERJEE; LASKAR, 2019).

Essa modificação permite a estimativa indireta da concentração desses íons e cátions através de modelos computacionais e técnicas quimiométricas. Dessa forma, a utilização dos elementos primários mencionados como informações complementares, promovem uma melhoria na predição de elementos secundários, incluindo os de interesse para fertilidade do solo e plantas (FARIDA et al., 2024).

Diversos autores como Demattê et al., (2017); Coutinho et al., (2019); Barra et al., (2021); Haghi; Pérez-Fernández; Robertson (2021) relatam que a aplicação de técnicas quimiométricas em conjunto com técnicas de espectroscopia em faixas individuais VIS, NIR, MIR, ou combinações VIS-NIR, SWIR se provaram capazes de fornecer informações confiáveis para diagnósticos de elementos primários e secundários do solo.

Coutinho et al., (2019) ao avaliarem o desempenho das técnicas espectroscopias reflectância difusa na região do visível e infravermelho próximo (Vis-NIR) e reflectância na região do infravermelho médio (MIR) em dois tipos de solos (arenoso e argiloso) e com diferentes tipos de preparo de amostras, para a quantificação do teor de N, P e K utilizando modelos de previsão a partir de técnicas quimiométricas. Os autores destacam correlações satisfatórias para o teor de P e K para a região Vis-NIR para solos argilosos com  $R^2$  variando entre 0.58 a 0.90 e 0.64 a 0.85 respectivamente, enquanto para solos arenosos foram encontrados  $R^2$  insatisfatórios para o teor de P variando 0.03 e 0.25 e satisfatório para o teor de K variando de 0.75 a 0.94. O teor de N obtiveram  $R^2$  variando entre moderadamente satisfatório a insatisfatório, destacando a dificuldade de avaliação de ambos os métodos espectroscópicos na detecção do nutriente em condições de baixas quantidades no solo.

Os autores destacam ainda que, para a variável P a faixa espectral entre 1900 e 2500 nm obteve uma resposta espectral devido a ligação com os oxihidróxidos de ferro e alumínio,

sendo estas utilizadas para prever indiretamente os valores e que devido a isso os melhores valores foram observados para o solo argiloso. Para a variável K as regiões espectrais mais importantes para sua predição são as correlacionadas com atributos primários do solo como a matéria orgânica e os argilominerais. Enquanto Demattê et al, (2017) observaram respostas espectrais nas regiões de 1400, 1900 e 2200nm para o P.

Haghi; Pérez-Fernández; Robertson (2021) ao avaliarem diferentes atributos do solo entre eles o Nitrogênio total, Mg e K nas faixas espectrais do NIR e MIR, utilizando diferentes modelos de regressão. O nitrogênio total e o Mg apresentaram ótima correlação com valores de  $R^2$  entre 0,90 e 0,95 e 0,72 e 0,83 respectivamente para os modelos de regressão estudados, enquanto o K apresentou uma correlação moderadamente boa com  $R^2$  variando entre 0,53 e 0,75 para os modelos estudados.

Zhou et al. (2019), desenvolveram um equipamento portátil para determinação do teor de nitrogênio total e umidade do solo em campo utilizando uma unidade óptica no infravermelho próximo, uma unidade de controle e uma unidade mecânica. Os autores obtiveram resposta espectral para a determinação teor de nitrogênio total nos comprimentos de onda de 1260, 1330, 1360, 1430, 1530, 1580 e 1660 nm, sendo estes utilização para a construção do equipamento, obtendo  $R^2$  de 0,891, 0,913, 0,905, 0,944, 0,887, 0,938, 0,884 e 0,951, para os respectivos comprimentos de onda sensíveis, enquanto o processo de calibração e validação do modelo de predição apresentaram um  $R^2$  0,832 e 0,783 respectivamente, alcançando resultados capazes de detectar em tempo real o teor de nitrogênio total.

## **2.5 Dinâmica do fósforo em solos**

O fósforo (P) no solo ocorre tanto em formas inorgânicas (minerais primários, precipitados, adsorvidos e em solução) como orgânicas (resíduos vegetais, microrganismos, ácidos nucleicos, etc.). Plantas absorvem apenas ortofosfato inorgânico, de modo que o P orgânico deve ser previamente mineralizado por enzimas fosfatases liberadas por raízes e microrganismos (WEIL; BRADY, 2017). Em solos tropicais altamente intemperizados, grande fração do P inorgânico apresenta-se adsorvido aos colóides ou retido em formas precipitadas de baixa solubilidade, em formas indisponíveis às plantas. (TIESSSEN et al., 1984).

### **2.5.1 Fatores de intensidade, quantidade e capacidade tampão**

A dinâmica do P no solo é governada por processos complexos que determinam sua

disponibilidade para as plantas. O modelo clássico de disponibilidade de P define dois compartimentos interligados: fator intensidade (I) e fator quantidade (Q) (HOLFORD; MATTINGLY, 1976).

O fator intensidade é a concentração de P na solução do solo, que em solos tropicais é muito reduzida por adsorção específica com alta força de ligação e reações de precipitação. O ácido ortofosfórico dissocia-se formando  $\text{H}_2\text{PO}_4^-$  (favorecido em pH ácido) e  $\text{HPO}_4^{2-}$  (em pH neutro a alcalino) (RUSSEL, 2014). As plantas preferencialmente absorvem o íon  $\text{H}_2\text{PO}_4^-$ , já que em solos brasileiros (pH geralmente ácido) ele predomina e sua absorção diminui em pH elevado.

O fator quantidade corresponde ao P-lábil do solo, ou seja, fosfatos adsorvidos aos coloides por ligações de energia relativamente baixa, em equilíbrio dinâmico com o P na solução. Dessa forma, quando o P da solução é absorvido por plantas, ou removidos, o fosforo adsorvido é desorvido (fator Q). Define-se também o fator capacidade (C) ou poder tampão do solo, dado por  $C=Q/I$ , que indica a rapidez com que Q repõe I. Em solos fortemente intemperizados e argilosos, I é muito pequeno e Q pode ser elevado, refletindo alta energia de ligação e baixa labilidade do P.

### 2.5.2 Adsorção e precipitação do fósforo

A adsorção de fosfatos nos coloides do solo ocorre inicialmente por atração eletrostática às superfícies carregadas positivamente (áreas de borda de argilas, hidróxidos de Fe/Al) e na sequência por ligações químicas covalentes específicas (NOVAIS, 2007). A adsorção é fortemente influenciada pelo teor e pelo tipo de argila, solos com alta quantidade de argila e de óxidos pouco cristalinos de ferro/alumínio apresentam capacidade de adsorção maior que solos de textura mais arenosa ou com argilas silicatadas.

A matéria orgânica do solo também afeta o processo, moléculas orgânicas de baixo peso, como ácido oxálico e fulvatos, podem bloquear sítios de adsorção inorgânica e reduzir a fixação de P pelas argilas (MELO; ALLEOINE, 2019). Além disso, compostos orgânicos podem ligar-se aos coloides por pontes metálicas ( $\text{Al}^{3+}$ ,  $\text{Fe}^{3+}$ ,  $\text{Ca}^{2+}$ ), conferindo caráter aniônico e adsorvendo diretamente fosfato.

Paralelamente, ocorrem reações de precipitação química, íons fosfato em solução combinam-se com cátions ( $\text{Ca}^{2+}$ ,  $\text{Al}^{3+}$ ,  $\text{Fe}^{3+}$ ) para formar novos minerais fosfatados pouco solúveis (MELO; ALLEOINE, 2019). Em solos ácidos dominam precipitados de fosfato com Fe e Al, enquanto solos com pH elevado e alto Ca formam fosfatos de cálcio. A labilidade dessas

fases precipitados segue a ordem  $P\text{-Al} > P\text{-Fe} > P\text{-Ca}$  (os fosfatos de alumínio são ligeiramente mais solúveis que os de ferro, ambos muito menos que os de cálcio). Essas reações de adsorção química e precipitação resultam em esgotamento do P em solução e estocagem majoritária do P como formas de baixa disponibilidade.

### **3 HIPOTESE**

O teor disponível de fósforo presente no solo altera os padrões de absorção e reflectância de energia em componentes opticamente ativos nos espectros. Essas variações espectrais podem ser empregadas como indicadores para estimar o teor de fósforo no solo por meio de análises quimiométricas, oferecendo uma precisão comparável aos métodos convencionais de laboratório.

### **4 OBJETIVOS**

#### **4.1 OBJETIVOS GERAIS**

Avaliar o potencial preditivo da espectroscopia de reflectância na quantificação do teor de fósforo disponível em solos, por meio da aplicação de diferentes algoritmos quimiométricos e de transformações nos dados de reflectância espectral adquiridos por espectrorradiômetro óptico na faixa de 350–2500 nm e seleção de variáveis.

#### **4.2 OBJETIVOS ESPECÍFICOS**

1. Selecionar e interpretar as regiões espectrais correlacionadas com o teor de fósforo disponível em amostras de solo.
2. Implementar diferentes técnicas de pré-processamento dos espectros e algoritmos quimiométricos, para a calibração de modelos preditivos de fósforo disponível em solos.
3. Avaliar o desempenho e efeito da combinação dos diferentes métodos de pré-processamento espectral e algoritmos quimiométricos na calibração de modelos na estimativa do fósforo disponível em solos.

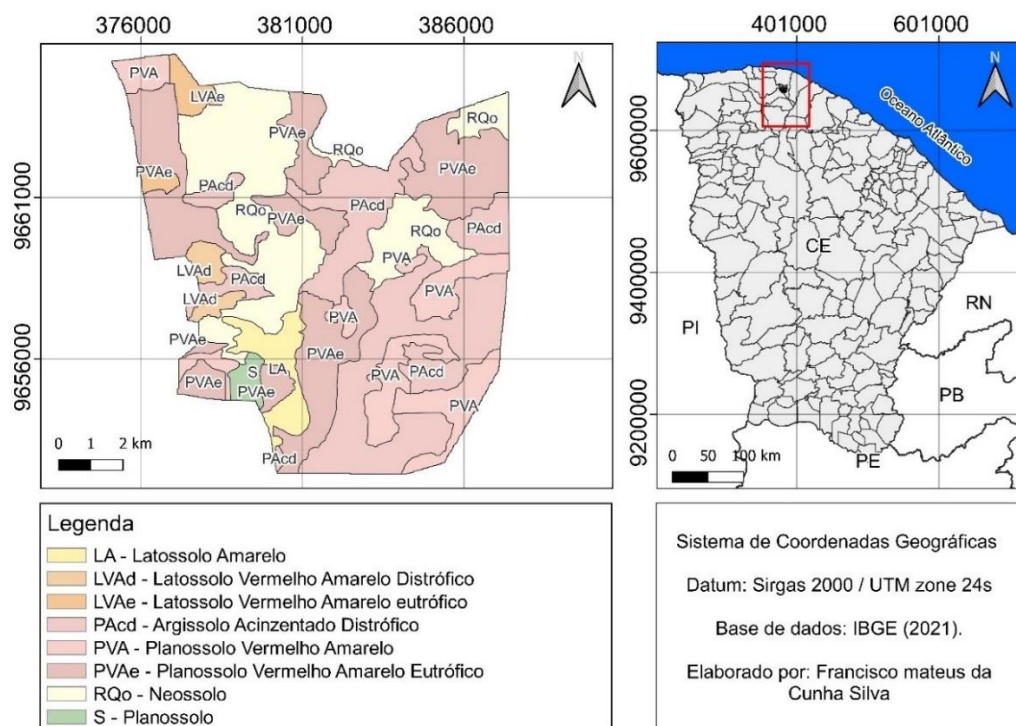
## 5 METODOLOGIA

### 5.1 Descrição da área de coleta

As amostras de solo foram coletadas no Distrito de Irrigação do Baixo Acaraú (DIBAU) que possui área total de 12.535 ha estendendo-se nos municípios de Acaraú, Bela Cruz e Marco, estando localizado entre as latitudes 3° 01' 00"S a 3° 09' 00"S e longitude 40° 01' 00."W a 40° 09' 00."W (Figura 2). Conforme a classificação de Koppen, o clima é do tipo AW caracterizando-se como tropical chuvoso com curto período seco no verão, e período chuvoso entre os meses de janeiro a maio e média anual de precipitação de 1.139 mm e temperatura 26 a 28°C. A topografia da região é caracterizada predominantemente como planícies, tabuleiros pré-litoraneos e depressões sertanejas com altitude média de 17m.

Os solos predominantes no município são o Argiloso Vermelho-amarelo (PVA), Argiloso Vermelho-amarelo Distrófico (PVAd), Argiloso Vermelho-Amarelo Eutrófico (PVAe), Neossolos quartzarênico Órtico (RQo), Latossolo Vermelho Eutrófico (LVAe), Latossolo Vermelho Amarelo Distrófico (LVAd), Latossolo Amarelo (LA) e Planossolos (S) (IPECE, 2017).

**Figura 2.** Mapa de localização e solos do Distrito de Irrigação do Baixo Acaraú.



**Fonte:** Elaborado pelo autor.

Para a coleta das amostras de solo, foram selecionados dois lotes agrícolas localizados no perímetro do DIBAU, caracterizados por condições edáficas distintas e baixa fertilidade em fósforo (Tabela 1).

**Tabela 1.** Características químicas e físicas dos solos Neossolo Quartzarênico Órtico (RQo) e Argissolo Acinzentado Distrófico (PACd) (0–10 cm).

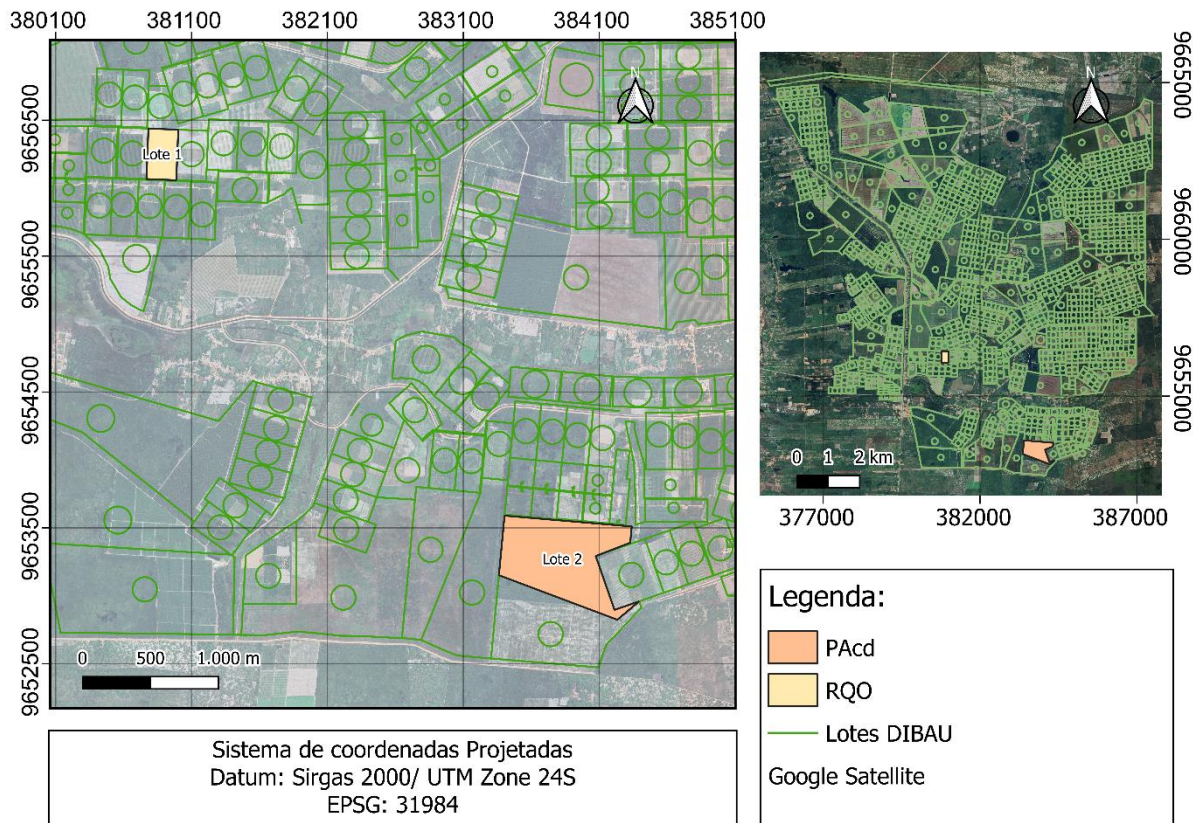
Análise química											
Solo	P mg.kg <sup>-1</sup>	pH	K <sup>+</sup>	Ca <sup>2+</sup>	Mg <sup>2+</sup>	Na <sup>+</sup>	H+Al	Al <sup>3+</sup>	SB	CTC	V
			----- mmolc/dm <sup>3</sup> -----								
RQo	15,54	5,5	0,71	0,56	0,49	0	0,44	0,01	1,76	2,20	80
PACd	13,31	6,1	0,91	0,8	1,13	0	1,26	0,01	2,84	4,1	69
Análise física											
Solo	Areia		Silte		Argila		Classe textural				
	----- g/kg -----		----- g/kg -----		----- g/kg -----						
RQo	790		50		145		Areia franca				
PACd	670		40		200		Franco argiloso arenoso				

**Fonte:** Elaborado pelo autor.

SB = Soma de Bases Trocáveis; CTC = Capacidade de Troca Catiônica; V = Saturação por bases.

O primeiro lote apresentava solo previamente manejado, com ausência de cobertura vegetal, sendo o cultivo anterior encerrado há aproximadamente três meses, sendo utilizado para a produção de melancia (*Citrullus lanatus*), o solo foi classificado como NEOSSOLO QUARTZARÊNICO (RQo) totalizando uma área de 8 ha. Por outro lado, o segundo lote encontrava-se em plena atividade agrícola, sendo destinado ao cultivo de coqueiros e classificado como ARGISSOLO ACINZENTADO (PACd), com área total de 40 ha (Figura 3).

**Figura 3.** Mapa de localização dos lotes de coleta dos solos.



**Fonte:** Elaborado pelo autor.

## 5.2 Coleta e preparação das amostras de solo

Neste estudo, foram selecionados dois lotes agrícolas caracterizados em duas diferentes classes. O primeiro lote representado por um NEOSSOLO QUARTZARÊNICO (RQ), enquanto o segundo lote correspondeu a um ARGISSOLO ACINZENTADO Distrófico (PAcd). As amostras foram coletadas na camada superficial de 0 a 10 cm, por ser a profundidade de maior interesse para análises de fertilidade do solo e por abranger a zona radicular das culturas de maior relevância agrícola.

Com o objetivo de minimizar variações intrínsecas ao solo, especialmente aquelas relacionadas às suas características químicas e mineralógicas e visando garantir maior homogeneidade nas amostras, reduzindo a influência de fatores não controlados e permitindo uma avaliação mais precisa dos efeitos dos fertilizantes aplicados para a coleta, foi estabelecido um ponto central em cada área de estudo (Figura 4).

**Figura 4.** Coleta das amostras de solo.



**Fonte:** Elaborado pelo autor.

Ao redor de cada ponto central, delimitou-se uma circunferência de 10 metros, dentro da qual as amostras foram coletadas de forma composta e deformadas. De cada lote agrícola, foram coletados aproximadamente 60 kg de solo, os quais foram devidamente identificados e armazenados em sacos plásticos para transporte.

Após a coleta, as amostras foram transportadas para a secagem no Laboratório de eletrônica e Mecânica agrícola (LEMA) da Universidade Federal do Ceará (UFC) para passarem pelo processo de secagem natural (ao ar). Uma vez secas, as amostras foram submetidas à destorroamento, peneiramento e homogeneização utilizando uma peneira com abertura de malha de 2 mm (Figura 5A).

**Figura 5.** Etapas de tratamentos das amostras de solo: Peneiramento (A), pesagem e divisão de subamostras (B), adubação e incubação (C) e secagem (D).



**Fonte:** Elaborado pelo autor.

Posteriormente, cada amostra de solo foi novamente homogeneizada e dividida em 75 subamostras de igual massa de 0.5kg cada e acondicionadas em sacos plásticos previamente identificados, para então modificar o teor de fósforo no solo a partir de fertilizante comercial MAP (fosfato monoamônico) purificado (Figura 5B). Para a preparação das subamostras foram utilizadas 15 diferentes concentrações em 5 repetições. Dessa forma, contabilizando um total de 150 amostras de solo para as duas classes. A Tabela 2, fornece informação sobre o fertilizante selecionado para o experimento.

**Tabela 2.** Descrição do fertilizante aplicado nas amostras de solo.

Nome comercial	Abreviação química	Componente P <sub>2</sub> O <sub>5</sub> (%)	Outros Componentes (%)	Solubilidade (g/L)	Índice salino / unidade
MAP purificado	NH <sub>4</sub> H <sub>2</sub>	61% de P <sub>2</sub> O <sub>5</sub>	12% de N	370	2,73

**Fonte:** Elaborado pelo autor.

De acordo com as classes de disponibilidade dos nutrientes descrita por Raij et al., (1996) e utilizadas por Coutinho et al., (2019) foram definidas concentrações de fósforo de 0; 5; 15; 25; 45; 60; 90; 120; 180; 210; 250; 290; 320; 350; 390 mg/kg de solo, definidas como D0 a D14. Para a aplicação dos fertilizantes, os mesmos foram pesados e dissolvidos individualmente em 100 ml de água, sendo então aplicada de forma homogênea e de maneira gradual e posteriormente misturadas para garantir a uniformidade da distribuição do fertilizante nas subamostras.

Após a aplicação dos fertilizantes, as subamostras de solo foram submetidas a um período de incubação de 14 dias, seguido por secagem, destorroamento e peneiramento (Figura 5C). As amostras peneiradas foram divididas em duas partes: a primeira destinada à análise química, e a segunda à análise espectral, na qual as amostras foram previamente secas em estufa a 45°C para garantir a homogeneidade e reduzir a interferência de umidade nos resultados (Figura 5D).

### 5.3 Determinação de fósforo disponível

A determinação do teor de fósforo disponível nos solos foi realizada no Laboratório de Química do Solo do Departamento de Ciência do Solo da Universidade Federal do Ceará (UFC), utilizando o método Mehlich 1, recomendado pela EMBRAPA (2017), sendo realizado em duplicatas. As amostras de solo foram submetidas ao processo de extração utilizando a solução extratora de Mehlich 1, composta por HCl 0,05 mol L<sup>-1</sup> e H<sub>2</sub>SO<sub>4</sub> 0,0125 mol L<sup>-1</sup>. Para

a extração, uma alíquota de solo (5 g) foi adicionada à 50 mL da solução extratora e agitada por 5 minutos para facilitar a solubilização do fósforo disponível.

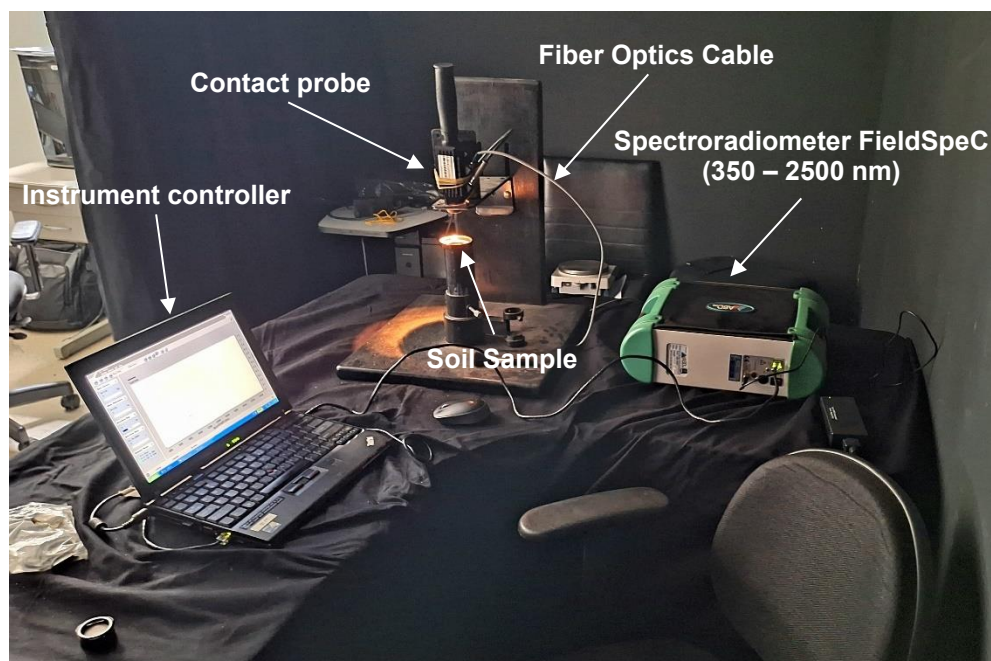
Após a agitação, a solução ficou em decantação por uma noite. Após, foram coletados 10 mL do extrato resultante que foi submetido à reação com uma solução colorimétrica composta por molibdato de amônio e ácido ascórbico. A leitura foi realizada em um espectrofotômetro AJX-1600, ajustado para comprimento de onda de 660 nm, com a concentração de fósforo determinada a partir de uma curva padrão preparada com soluções de concentrações conhecidas.

#### **5.4 Avaliação espectral das amostras de solo**

Para a avaliação espectral, foi realizada em ambiente controlado em relação a iluminação no laboratório de geoprocessamento pertencente ao Departamento de engenharia agrícola da Universidade Federal do Ceará (UFC). Para as medições foi utilizado o espectrorradiômetro FieldSpec (Analytical Spectral Devices, Inc., Boulder, Co) que possui resolução espectral de 1 nm, realizando as medições nas faixas de 350 - 2500nm, compreendendo as regiões espectrais do Visível (VIS), Infravermelho Próximo (NIR) e Infravermelho de Ondas Curtas (SWIR).

Uma placa branca espectral que permite a completa reflectância em todas as faixas espectrais foi utilizada como referência padrão, para a calibração do equipamento a cada 15 min. Foram conduzidas três repetições de medição utilizando o equipamento em diferentes posições na superfície da amostra. Em cada uma dessas posições, foram realizadas 10 leituras consecutivas. O valor médio do espectro foi então calculado com base nessas leituras. Posteriormente os dados foram convertidos de número digital (ND) para fator de refletância por meio do software ViewSpecPro. A metodologia de aquisição dos dados hiperespectrais em amostras de solo, incluindo o arranjo do espectrorradiômetro está ilustrada na Figura 6.

**Figura 6.** Esquema de aquisição de dados hiperespectrais (350–2500 nm) nas amostras de solo.



Fonte: Elaborado pelo autor.

### 5.6 Análise quimiométrica e resposta espectral do solo

Para a análise dos dados, foram considerados três conjuntos distintos: os dois primeiros representam as classes de solo individualizadas, RQo e PAcd, e o terceiro conjunto agrupa ambas as classes de solo (RQo + PAcd). Dentro de cada conjunto, foi adotada uma divisão de 70% dos dados para a calibração dos modelos e 30% para validação. A Tabela 3 detalha a quantidade de amostras utilizadas para calibração e validação em cada conjunto.

**Tabela 3.** Distribuição quantitativa dos conjuntos amostrais utilizados para calibração e validação.

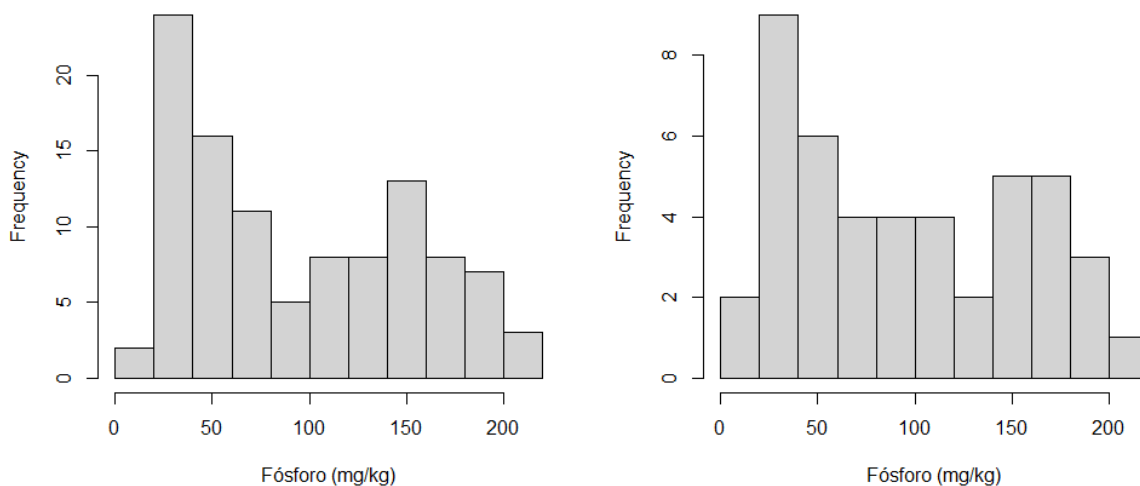
Conjuntos	Calibração	Validação	Total	Mínimo mg kg <sup>-1</sup>	Máximo mg kg <sup>-1</sup>
RQo	52	23	75	23.85	196.34
PAcd	52	23	75	17.68	188.37
RQo + PAcd	105	45	150	17.68	196.34

Fonte: Elaborado pelo autor.

Para a seleção dos dados de calibração e validação, foi adotada uma abordagem de amostragem estratificada. Inicialmente, as amostras foram ordenadas de acordo com o teor de fósforo. A partir dessa ordenação, as amostras foram agrupadas em subconjuntos, e posteriormente as amostras foram selecionadas aleatoriamente de cada subconjunto para garantir uma distribuição equilibrada dos teores de fósforo entre os conjuntos de calibração e validação. A Figura 7 ilustra a distribuição dos teores de fósforo (mg kg<sup>-1</sup>), sendo possível

visualizar a representatividade dos dados em ambos os conjuntos.

**Figura 7.** Histogramas da distribuição dos teores de fósforo (mg/kg) nas classes de solo.



**Fonte:** Elaborado pelo autor.

### 5.6.1 Pré-tratamento dos dados espectrais

Para aprimorar a qualidade dos dados e realçar características relevantes, é crucial aplicar filtros nos dados brutos de refletância. Nesse sentido, os dados obtidos pelo espectrorradiômetro passaram por três métodos matemáticos de pré-processamento: a normalização, a suavização da primeira derivada com Savitzky-Golay (SG) e o continuum removal (CR).

Esses métodos são empregados para reduzir o ruído nos dados de refletância, tornando-os mais legíveis e facilitando a identificação de padrões significativos. O SG remove ruídos de alta frequência, preservando ao máximo a forma dos picos espectrais. Enquanto o CR elimina características contínuas dos espectros, permitindo uma melhor visualização e destaque das características de absorção específicas dos parâmetros avaliados (MOURA-BUENO et al., 2020). Esses métodos foram implementados utilizando a biblioteca prospector no software R versão 3.6.3 (R CORE TEAM, 2024), conforme disponibilizado por Stevens e Ramirez-Lopez (2014).

### 5.6.2 Calibração e validação dos modelos preditivos

Para a calibração dos modelos de predição do fósforo disponível no solo, foram

empregados três métodos quimiométricos: a regressão linear múltipla (MLR), a regressão em componentes principais (PCR) e a regressão por mínimos quadrados parciais (PLSR). Para a avaliação dos métodos quimiométricos foram aplicadas duas abordagens. A primeira, todos os comprimentos de onda do espectro (350–2500 nm) foram usados diretamente para a calibração dos modelos, enquanto na segunda, realizou-se uma seleção de variáveis, eliminando inicialmente as bandas com baixa correlação estatística com o teor de fósforo por meio da correlação e, em seguida, aplicando-se o método Stepwise-Forward para escolher incrementalmente as bandas mais informativas.

A MLR permite estabelecer relações lineares entre as variáveis independentes e dependentes (WEHRENS, 2020). A PCR reduz a dimensionalidade dos dados através da transformação em componentes ortogonais, permitindo a análise das principais fontes de variação nos atributos químicos do solo (WEHRENS, 2020). Enquanto, a regressão por mínimos quadrados parciais (PLSR) é uma abordagem que combina as vantagens da regressão múltipla e da análise de componentes principais, sendo especialmente útil quando há multicolinearidade entre as variáveis independentes (WEHRENS, 2020).

Quatro métricas foram utilizados para avaliação do desempenho dos modelos de predição: O coeficiente de determinação ( $R^2$ ) (Equação 8), coeficiente de determinação ajustado ( $R^2_{ajust}$ ) (Equação 9), Raiz do Erro Quadrático Médio (RMSE) (Equação 10) e a Razão da Performance do Desvio (RPD) (Equação 11).

$$R^2 = 1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} \quad Eq. 8$$

$$R^2_{ajust} = 1 - \frac{(N - 1)(1 - R^2)}{N - (K + 1)} \quad Eq. 9$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{Y}_i - Y_i)^2}{N}} \quad Eq. 10$$

$$RPD = \frac{\sigma_{Y_o}}{RMSE} \quad Eq. 11$$

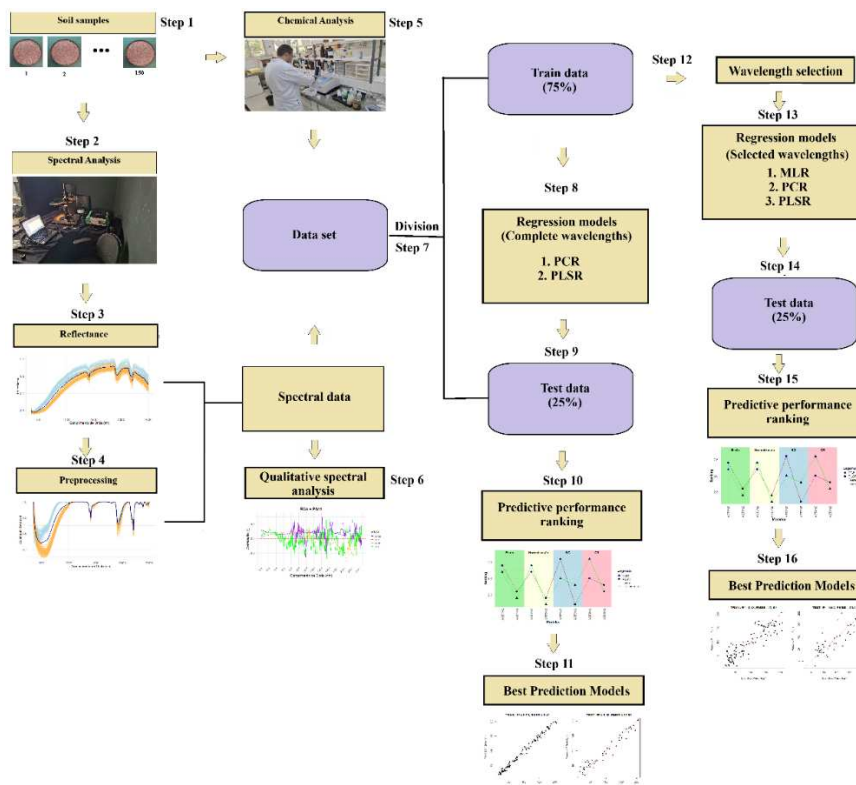
Onde,  $\hat{Y}_i$  = Valores preditos pelos modelos nas observações;  $Y_i$  = Valores determinados por método padrão;  $\bar{Y}_i$  = Média dos valores determinados por método padrão; N

= Número de observações;  $k$  = Total de variáveis independentes;  $\sigma$  = Desvio padrão observado ou predito.

Para seleccionar o modelo de melhor desempenho, adotou-se uma análise de decisão multicritério baseada no método de soma ponderada de Fishburn (1967). Dessa forma, foram utilizados como indicadores de desempenho as métricas  $r^2$ ,  $R^2$ , RMSE e RPD, avaliados tanto na calibração quanto na validação dos modelos. Para cada métrica, aplicou-se a normalização min–max. Em seguida, para cada modelo foi definido um *score* calculado pela soma ponderada dos valores normalizados, atribuindo-se peso igual a cada um dos critérios. Posteriormente, os *scores* foram ordenados e ranqueados, elegendo-se o de maior *score* como o de melhor desempenho geral dentro de cada conjunto de dados.

Na Figura 8 é ilustrado o fluxograma metodológico para obtenção dos modelos de predição e validação dos atributos estudados.

**Figura 8.** Fluxograma metodológico para obtenção dos modelos de predição e validação dos atributos estudados.



Fonte: Elaborado pelo autor.

## 5.7 Implementação de Rotinas Estatísticas e Quimiométricas no R

Todo o fluxo estatístico deste estudo foi programado integralmente do zero no R (v. 3.6.3), abrangendo desde o pré-processamento dos espectros até a construção, validação e comparação de modelos quimiométricos por meio de validação cruzada leave-one-out e análise multicritério baseada em Fishburn, sem recorrer a funções automatizadas de pacotes prontos. Para apoio em tarefas auxiliares de leitura e escrita de planilhas, manipulação de conjuntos de dados e representação gráfica, foram empregadas as bibliotecas readxl (WICKHAM; BRYAN, 2025), writexl (OOMS, 2025), dplyr (WICKHAM et al., 2023), ggplot2 (WICKHAM et al., 2025), MASS (RIPLEY; VENABLES, 2025), glmnet (FRIEDMAN et al., 2025) e pls (LILAND et al., 2025).

## 6 RESULTADOS E DISCUSSÃO

### 6.1 Concentrações de fósforo obtidas com diferentes doses de fertilizantes em solos

As concentrações de fósforo no solo, variaram conforme as doses de fertilizantes aplicadas nos dois tipos de solos distintos: o Neossolo Quartzarênico Órtico (RQo) e o Argissolo Acinzentado Distrófico (PAcd) (Tabela 4). A análise dos dados evidencia diferenças entre os dois solos em termos de capacidade de retenção e disponibilidade de fósforo, características associadas às suas propriedades físicas e químicas.

**Tabela 4.** Teor de fósforo ( $\text{mg kg}^{-1}$ ) em NEOSSOLO QUARTZARÊNICO (RQo) e ARGISSOLO ACINZENTADO sob diferentes doses de aplicação de MAP.

Tratamento	Tipo de solo	
	RQo	PAcd
D0	23,85	17,68
D1	31,83	26,85
D2	38,90	27,95
D3	44,31	36,83
D4	49,05	34,83
D5	54,78	45,67
D6	73,17	72,20
D7	90,56	73,19
D8	103,40	98,63
D9	139,26	114,77
D10	151,50	124,38
D11	154,14	141,01
D12	168,23	152,50
D13	179,27	181,50
D14	196,34	188,37

**Fonte:** Elaborado pelo autor.

No RQo, caracterizado por uma textura predominantemente arenosa e baixa capacidade de retenção de nutrientes. Os valores passaram de 23,85 mg kg<sup>-1</sup> no tratamento controle (D0) para 196,34 mg kg<sup>-1</sup> no tratamento D14, demonstrando uma maior facilidade na mobilização do fósforo para a solução do solo. Essa rápida resposta às aplicações fertilizantes reflete a pouca interação do fósforo com as partículas de areia, reduzindo a formação de compostos adsorvidos ou precipitados.

Por outro lado, no PAcd, um solo mais argiloso e rico em óxidos de ferro, as concentrações de fósforo alcançaram valores de 17,68 mg kg<sup>-1</sup> no D0 e atingindo 188,37 mg kg<sup>-1</sup> no D14. No entanto, os valores intermediários foram frequentemente menores em comparação ao RQo. Essas diferenças ilustram como a dinâmica do fósforo é fortemente influenciada pelas propriedades intrínsecas dos solos. Enquanto no RQo a mobilidade do nutriente é amplificada pela textura arenosa e baixa capacidade de retenção, no PAcd, a adsorção mais intensa, causada pelos óxidos de ferro e pela maior superfície específica das partículas de argila, resulta em uma retenção mais pronunciada. Esses comportamentos estão alinhados com estudos prévios, como os de Bastos et al. (2008), que destacam o papel da textura e da mineralogia na adsorção e difusão do fósforo no solo.

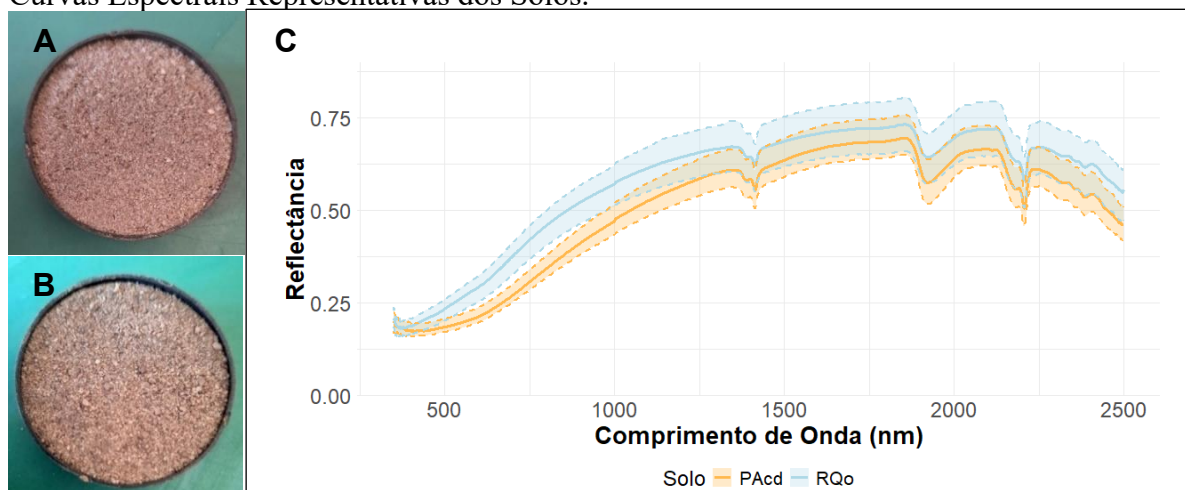
Em estudo realizado por Jalali e Jalali (2016) para avaliar a disponibilidade e a sorção de fósforo (P) em solos com diferentes texturas, os autores destacaram que a textura do solo tem uma influência significativa na dinâmica do fósforo em solos calcários. Solos com texturas arenosas e suas variações, como argiloso-arenoso e franco arenoso, apresentaram maiores teores de P extraído pelos diferentes métodos analisados. Em contrapartida, solos com texturas mais argilosas e franco-argilosas demonstraram maior capacidade-tampão de fósforo (PBC), indicando uma retenção mais eficiente e uma menor disponibilidade imediata desse nutriente. Além disso, o teor de silte foi identificado como um componente fundamental, exibindo uma correlação positiva com o coeficiente de sorção de Freundlich (Kf) e a capacidade tampão de fósforo (PBC), evidenciando sua relevância no controle da sorção e disponibilidade de fósforo no solo.

## **6.2 Análise qualitativa dos dados hiperespectrais**

Os espectros médios podem ser utilizados para avaliar qualitativamente as características dos solos, observando comprimentos de onda específicos na faixa visível, infravermelho próximo e infravermelho de ondas curtas, que revelam feições de absorção e reflectância associadas a diferentes constituintes básicos de todo solo, como matéria orgânica e

componentes orgânicos, óxidos de ferro, quartzo, argilominerais e teor de umidade (RIBEIRO et al., 2021; CHAUHAN et al., 2021; PAVLŮ et al., 2023) (Figura 9).

**Figura 9.** Amostras de Solos e Espectros de Reflectância: (A) NEOSSOLO QUARTZARÊNICO (RQo), (B) ARGISSOLO ACINZENTADO Distrófico (PAcd) e (C) Curvas Espectrais Representativas dos Solos.

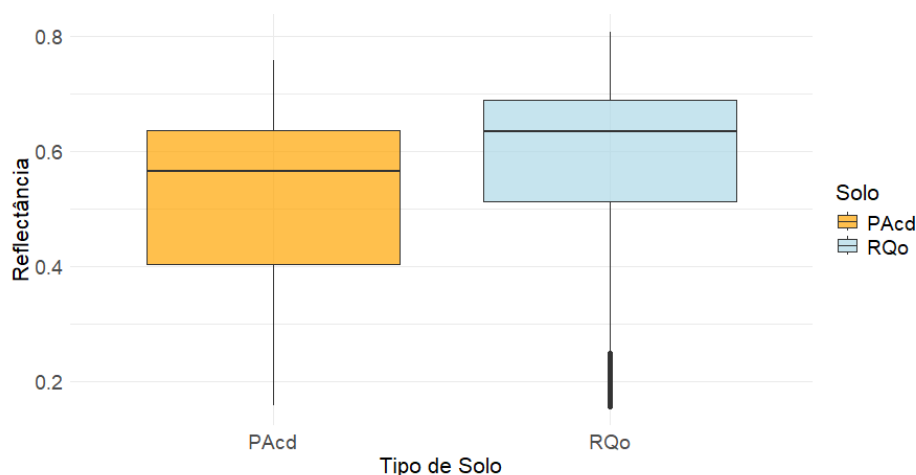


Fonte: Elaborado pelo autor.

Na Figura 9C, as curvas médias e espectros mínimo e máximo, que delimitam a variação de reflectância dentro das classes de solo estudadas, são apresentados. Observa-se que o espectro médio do RQo apresenta maior reflectância global, característica típica de solos arenosos, devido à predominância do quartzo. Solos com predominância de minerais de quartzo podem ser identificados por seu comportamento espectral característico, marcado por alta reflectância em determinadas faixas do espectro eletromagnético (COSTA et al., 2022). Por outro lado, o PAcd exibe menores valores de reflectância, o que pode ser atribuído à maior presença de argilominerais, óxidos de ferro ou ainda MO, que aumentam a absorção da REM em diferentes faixas.

A Figura 10 complementa essa análise ao ilustrar a distribuição da reflectância para cada tipo de solo. Observa-se que o RQo tende a apresentar valores medianos de reflectância mais elevados em comparação ao PAcd, o que pode estar relacionado à maior predominância de quartzo e à menor presença de argilominerais e óxidos de ferro, que geralmente estão associados à maior absorção da radiação eletromagnética. Além disso, a amplitude dos dados sugere uma variabilidade maior no RQo, o que pode refletir diferenças na textura e na composição mineralógica ao longo das amostras analisadas.

**Figura 10.** Boxplot da distribuição da Reflectância para os solos RQo e PAcD.



**Fonte:** Elaborado pelo autor.

Atualmente, o fósforo no solo não possui bandas de absorção bem definidas no espectro VIS-NIR-SWIR e, portanto, não pode ser avaliado de forma qualitativa diretamente por espectroscopia de reflectância. No entanto, ele pode ser inferido indiretamente a partir de outros componentes que influenciam sua disponibilidade e retenção no solo. Segundo Coutinho et al., (2019) e Potdar et al. (2021), a distribuição do fósforo no solo está fortemente associada à presença de óxidos de ferro, como hematita e goethita, e aluminossilicatos, como caulinita e gibbsita.

Na Figura 9C, são evidentes feições espectrais características na região do NIR-SWIR, particularmente em torno de 1400 nm, 1900 nm e 2200 nm, que refletem a interação da radiação com diferentes componentes do solo, incluindo os que têm relação com o fósforo no solo, como a umidade, os argilominerais 1:1 e a composição mineralógica dos solos analisados.

As bandas de absorção em torno de 1400 nm e 1900 nm indicam, respectivamente, a presença de sobretons de alongamento de grupos hidroxila (OH), seja na forma livre ou adsorvida por minerais da fração argilosa, e a combinação de estiramento de OH e da flexão de HOH, evidenciando moléculas de água estruturais (TERNIKAR; GOMEZ; KUMAR, 2024; ZHOU et al., 2024). Dessa forma, no PAcD, essas absorções são mais intensas em relação ao RQo, refletindo a maior quantidade de argila nesse solo.

Já a absorção em 2200 nm está relacionada a argilominerais 1:1 como caulinita e gibbsita (PAVÃO et al., 2024). No PAcD, essa feição espectral se destaca devido ao maior grau de desenvolvimento pedogenético e à maior presença de argilominerais. Em contrapartida, o RQo, composto predominantemente por quartzo, exibe uma resposta espectral mais fraca nessa região.

Conforme Carvalho et al., (2022), as curvas de reflectância espectral com menor reflectância geralmente indicam amostras com maior teor de óxidos de ferro e MOS, com bandas de absorção em 1400 e 1900 nm associadas às ligações OH e metal-OH nos minerais do solo, e bandas em torno de 2200 nm relacionadas à interação da radiação NIR com minerais de rede. Além disso, as bandas entre 2250 e 2450 nm estão associadas a ligações CH e são relacionadas à presença de ácidos húmicos e lignina. Variações no teor de argila podem afetar as características do espectro, alterando a absorção de energia devida às ligações hidroxila em minerais de argila.

Quanto aos valores de fósforo, os resultados mostram concentrações mais elevadas no RQo em relação ao PAcD, o que pode estar relacionado à menor presença de argilominerais no RQo. Esses minerais, como a caulinita e a gibbsita, têm alta capacidade de adsorção de fósforo, reduzindo sua disponibilidade no solo. No espectro, isso pode se manifestar em diferenças na intensidade das feições associadas à argila, que influenciam a retenção de nutrientes como o fósforo.

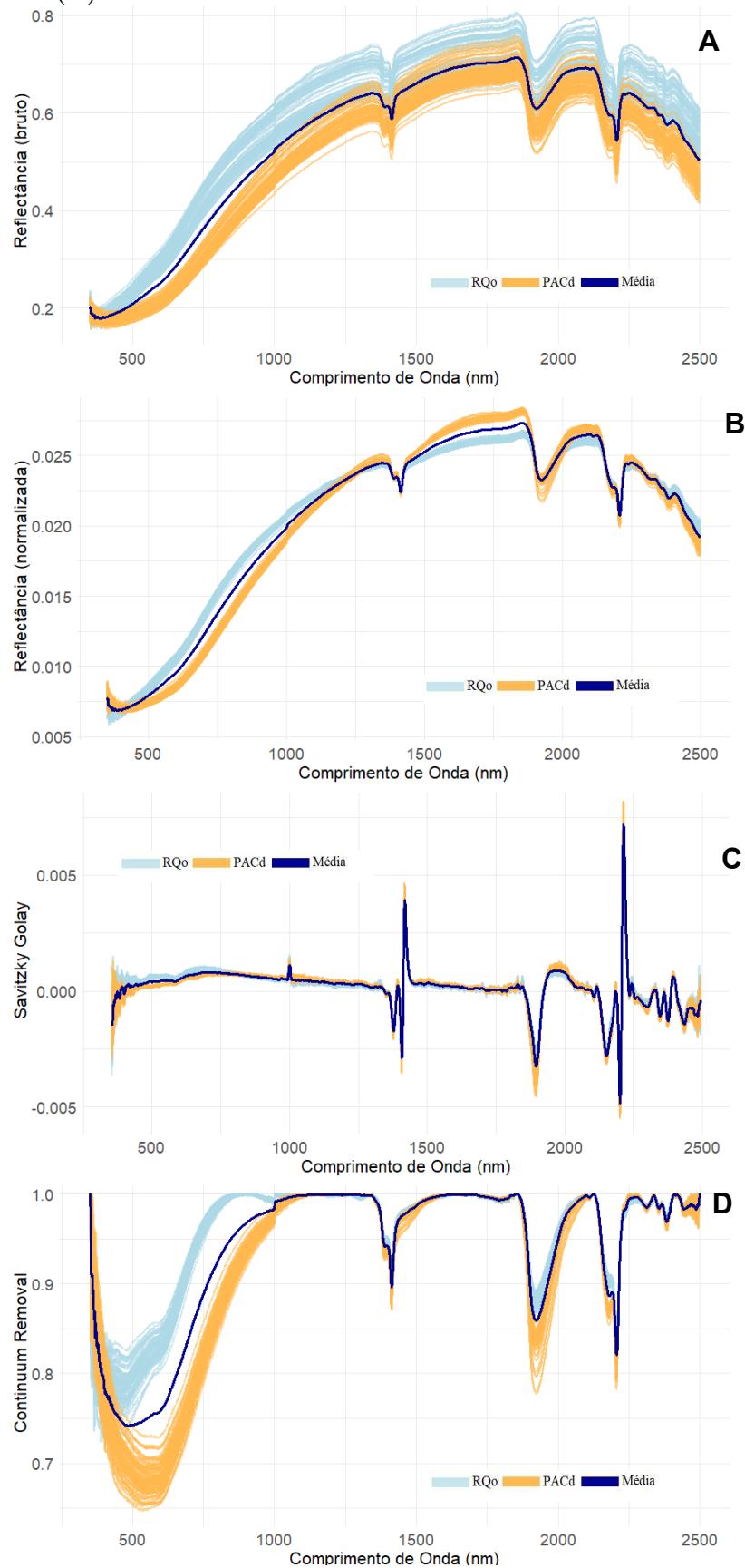
### **6.3 Pré-tratamento espectral**

A análise espectral é sensível aos efeitos de dispersão, que podem ser causados por variáveis intrínsecas à amostra (CARVALHO et al., 2022). Esse espalhamento pode ocorrer por diferentes fatores, como o tamanho, forma e estrutura das partículas, além da organização física da amostra. Quando a luz interage com essas partículas, ela não segue uma trajetória linear, o que pode alterar o comprimento do caminho percorrido e afetar as medições espectrais.

Esse efeito pode distorcer as características espectrais da amostra e introduzir variações nos resultados, dificultando a precisão das análises. Portanto, a dispersão pode gerar ruídos ou interferências no espectro, podendo ser necessário o uso de técnicas de pré-processamento para minimizar sua influência e realçar os sinais que são relevantes para constituintes presentes no solo.

O espectro bruto, abrangendo a faixa de 350 a 2500 nm, foi submetido a diferentes métodos de pré-processamento espectral, incluindo a normalização, a primeira derivada de Savitzky-Golay e a remoção do contínuo. A Figura 11 apresenta os resultados obtidos para cada técnica, destacando o espectro bruto utilizado como referência (Figura 11A), a normalização aplicada aos dados (Figura 11B), a transformação pela primeira derivada do filtro Savitzky-Golay (Figura 11C) e o método de remoção do contínuo (Figura 11D).

**Figura 11.** Tratamentos dos dados espectrais médios das amostras dos solos RQo e PACd Reflectância bruta(A), Normalização (B), Suavização da reflectância por Savitzky-Golay (C) e Continuum remove (D).



Fonte: Elaborado pelo autor.

A normalização (Figura 11B) foi eficiente em reduzir as diferenças na intensidade absoluta, destacando padrões relativos ao longo do espectro sem alterar significativamente a forma ou as feições características da curva. Com base nesse método, foi possível observar a diferenciação entre as amostras de solo RQo e PAcD, especialmente nas faixas de 500-1200 nm, 1500-1800 nm, e nas feições de absorção localizadas em 1400, 1900 e 2200 nm.

Observa-se que, com a aplicação da suavização por meio do filtro Savitzky-Golay (Figura 11C), tanto o desvio quanto a inclinação da linha de base foram significativamente reduzidos em comparação com os espectros brutos. Essa melhoria é especialmente relevante na extremidade dos espectros de 350 a 2500 nm, onde se evidencia uma grande variação dos dados, possivelmente indicando a presença de ruído. Com essa aproximação a uma linha de base horizontal, o filtro não apenas minimiza o ruído, mas também aprimora a resolução das bandas de absorção em torno de 1400, 1900 e 2200 nm, permitindo a identificação de picos entre 2300 e 2450 nm. Ao aproximar o espectro a uma linha de base horizontal, o filtro diminui o ruído, facilitando a interpretação das bandas de absorção, o que é fundamental para análises quantitativas e para a obtenção de melhores modelos preditivos (SHI; YAO; PAN, 2021).

Para o método do Continuum Removal (Figura 11D), observa-se uma variação acentuada dos dados na faixa de 500 a 1000 nm, sendo possível distinguir os solos estudados, evidenciando a presença de picos de absorção importantes que podem estar associados, por exemplo, a óxidos de ferro e matéria orgânica. Ao eliminar o efeito do contínuo de fundo, essa técnica ressalta as feições espectrais, permitindo uma distinção mais clara entre as assinaturas dos constituintes do solo, sem que a variabilidade global interfira na análise (CLARK; ROUSH, 1984). Assim, as bandas de absorção, já destacadas nos intervalos de 1400, 1900 e 2200 nm, ganham ainda mais evidência.

## 6.4 Correlação de Pearson

O coeficiente de correlação de Pearson ( $r$ ) constitui uma medida estatística essencial para quantificar a intensidade e a direção da relação linear entre duas variáveis quantitativas, podendo variar cujos valores podem variar de -1 a 1 (KUTNER et al., 2005; KHODADADI et al., 2024). Em geral, valores de  $r$  próximos a 1 indicam uma associação linear positiva, sugerindo que o incremento em uma variável está associado a um aumento proporcional na outra. Por outro lado, valores próximos a -1 evidenciam uma associação linear negativa forte, implicando que o aumento em uma variável está concomitantemente relacionado à diminuição na outra. Valores próximos a 0, entretanto, apontam para uma fraca ou inexistente relação linear,

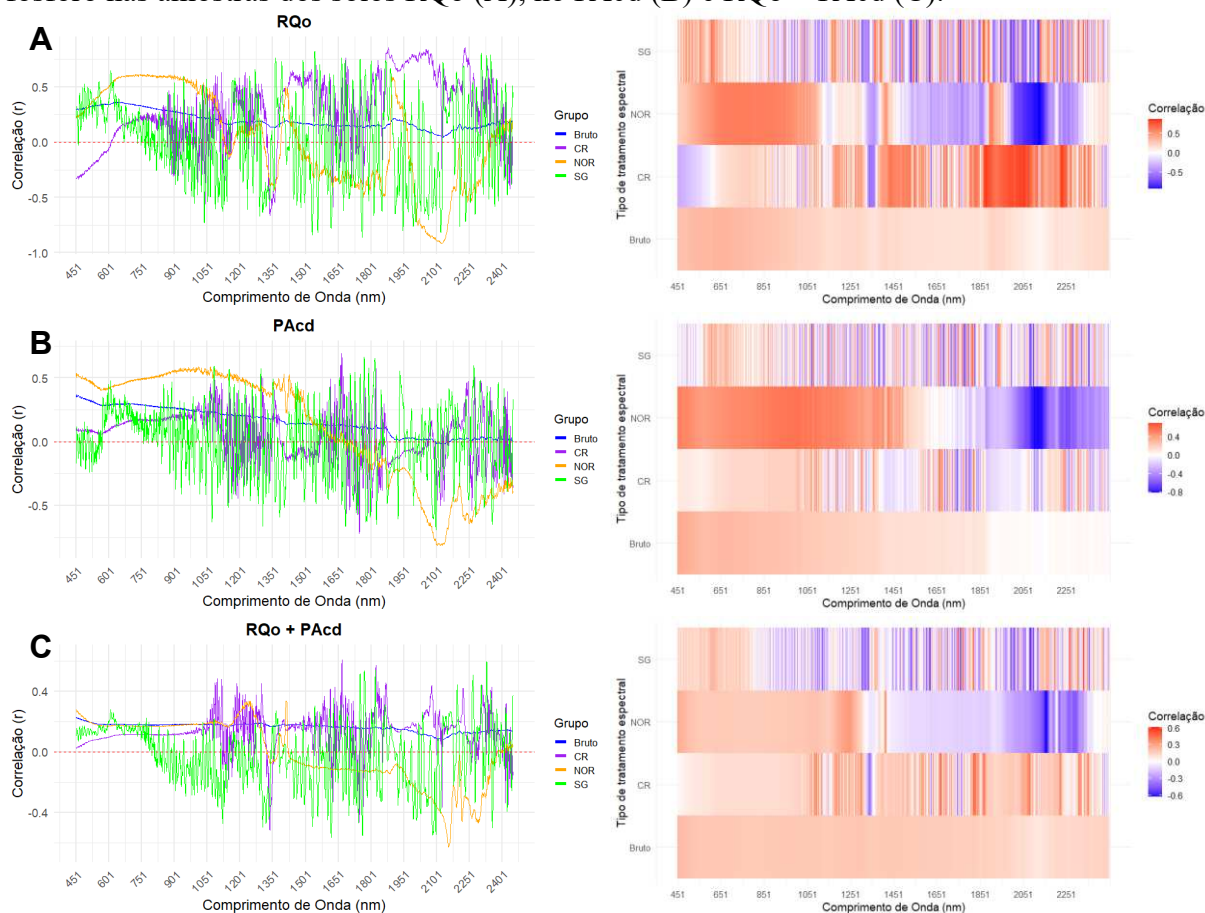
embora outras formas de associação possam estar presentes.

Durante a análise qualitativa e pré-tratamento dos espectros dos dados brutos, foram identificadas regiões ruidosas nas extremidades do espectro. Para mitigar esse efeito e melhorar a qualidade das medições de correlação e etapas subsequentes, essas regiões foram removidas, resultando em um intervalo útil de 451 a 2450 nm.

É importante ressaltar que, a aplicação do coeficiente de Pearson pressupõe que os dados se aproximem de uma distribuição normal e que a relação entre as variáveis seja fundamentalmente linear, sendo também sensível à presença de outliers, os quais podem distorcer significativamente os resultados da análise (KUTNER et al., 2005).

A Figura 10 apresenta os coeficientes de correlação de Pearson ( $r$ ) entre os valores de reflectância espectral no intervalo de 450 a 2450 nm e os teores de fósforo nas amostras de solo. Os resultados são mostrados para dois solos distintos, RQo (Figura 12 A) e PAcd (Figura 12 B), além da combinação dos dados de ambos os solos (Figura 12 C). Os gráficos à esquerda representam a correlação em função do comprimento de onda para quatro diferentes tratamentos espectrais. Já o mapa de calor à direita ilustra a distribuição dos coeficientes de correlação para os mesmos tratamentos espectrais ao longo do espectro eletromagnético, destacando regiões onde há agrupamentos de alta ou baixa correlação.

**Figura 12.** Coeficientes de correlação de Pearson ( $r$ ) e mapa de calor entre bandas espectrais VisNIR-SWIR (450 – 2.450 nm) nos diferentes níveis de processamento espectral e o teor de fósforo nas amostras dos solos RQo (A), no PAcd (B) e RQo + PAcd (C).



**Fonte:** Elaborado pelo autor.

De forma geral, é possível observar que a correlação para os dois tipos de solo e sua combinação, apresentaram uma grande variação de correlação ao longo do espectro em resposta aos pré-tratamentos espectrais utilizados. No solo RQo, a correlação variou entre -0.92 e 0.86, respectivamente aos pré-tratamentos espectrais normalização e CR (Figura 12 A). Para o solo PAcd foram encontrados uma variação dos valores de  $r$  entorno de -0.82 e 0.69, respectivamente para os métodos normalização e CR (Figura 12 B). Enquanto que para RQo + PAcd observou-se uma menor amplitude de variação, com correlação entre -0.62 a 0.61 também respectivamente para a normalização e CR. Evidenciando assim, uma grande variação de correlação ao longo do espectro (Figura 12 C).

Nota-se que a correlação para os dados brutos nos solos RQo, PAcd e RQo + PAcd foram predominantemente positivas e homogêneas ao longo das faixas do espectro, embora os comprimentos de onda na região do visível apresentaram correlações entre 0.3 a 0.4, sendo levemente superiores as regiões do NIR e SWIR, onde os valores correlação foram próximas

de 0.2.

Para os dados normalizados dos solos RQo, PAcD e RQo + PAcD, os valores de  $r$  demonstraram uma dispersão mais acentuada ao longo do espectro. No solo RQo, as faixas do visível e NIR apresentaram correlações moderadamente altas e positivas com valores de  $r$  situados em torno de 0.4 a 0.6 entre 600 a 1000 nm e próximo as feições de 1400nm e 1900nm, a região do SWIR evidenciaram uma tendência de inversão, atingindo picos negativos de -0.92 nos comprimentos de onda próximos a 2200 nm no solo (Figura 12 A). Embora apresentaram maiores valores de correlação na faixa do NIR os solos PAcD e RQo + PAcD seguiram o mesmo princípio do RQo.

Dentre os pré-tratamentos espectrais o SG obteve uma maior variação abruptas de  $r$  nas faixas do NIR e SWIR e mais estáveis e predominantemente positivas na faixa do visível para os três solos estudados. Sendo observado para o solo RQo a maior variação para o método SG com valores de  $r$  entre -0.86 a 0.82 na faixa dentro do NIR e SWIR. O mesmo ocorre para os solos PAcD e RQo + PAcD entretanto com -0.69 a 0.66 e -0.56 a 0.59 respectivamente.

Enquanto para o método CR, os três solos apresentaram correlações bastante abruptas. No solo RQo, os coeficientes de correlação na faixa do visível permaneceram próximos de zero, enquanto no NIR e SWIR apresentaram variações significativas, com correlações moderadamente altas tanto positivas quanto negativas. No solo PAcD, o método CR revelou um padrão diferente, com correlações mais homogêneas ao longo do espectro, embora ainda exibindo oscilações entre valores positivos e negativos. Já na combinação RQo + PAcD, a correlação mostrou-se mais estável, com menor amplitude de variação ao longo das bandas espectrais, indicando uma possível suavização dos efeitos individuais de cada solo quando analisados em conjunto.

## **6.5 Estimativa do teor de fósforo disponível o espectro completo (450-2450 nm)**

### **6.5.1 Regressão por componentes principais utilizando o espectro completo**

A Tabela 5 apresenta a organização dos tratamentos aplicados aos modelos, reunindo diferentes combinações de conjuntos de dados, métodos de pré-processamento e arquiteturas de modelo. Como resultado dessas combinações, foram formalizados 60 modelos distintos, divididos entre o uso de espectro completo (24 modelos) e a seleção de variáveis (36 modelos).

**Tabela 5.** Tipos e Identificação dos Tratamentos Aplicados nos Modelos.

Tipos de tratamento		Tratamento		
Conjuntos de dados	RQO (A)	PACd (B)	RQO + PACd (AB)	Sel. Var. (S)
Pré-processamento	BRUTO (T1)	NOR (T2)	SG (T3)	CR (T4)
Arq. Modelo	MLR (M1)	PCR (M2)	PLSR (M3)	

**Fonte:** Elaborado pelo autor.

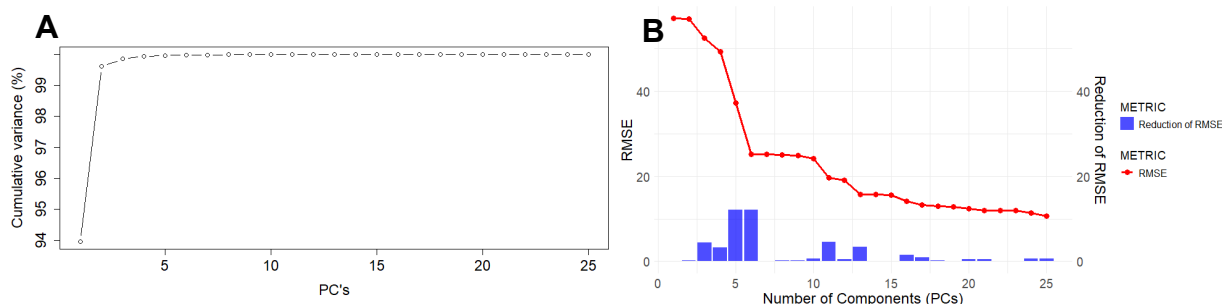
RQO = Neossolo Quartizarênico órtico; PACd = Argissolo acinzentado distrófico; Sel. Var = Seleção de variáveis; BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo; MLR = Regressão linear múltipla; PCR = Regressão por componentes principais; PLSR = Regressão por mínimos quadrados parciais.

Para a calibração dos modelos de regressão PCR e PLSR, utilizou-se o método de validação cruzada leave-one-out (LOO). Nesse procedimento, cada amostra do conjunto de calibração é removida sequencialmente para ser utilizada como validação, enquanto o modelo é ajustado com as demais amostras. Esse processo é repetido até que todas as amostras tenham sido validadas.

Diferentes modelos foram testados com o acréscimo progressivo de componentes principais. A escolha do número ideal de componentes para a calibração dos modelos foi baseada em dois critérios, a minimização do erro quadrático médio (RMSE) e a maximização da variância explicada.

A Figura 13 ilustra essa análise utilizada para definição da quantidade ideal de componentes para o modelo utilizando o conjunto de dados RQO + PACd em refletância bruta e método de regressão PCR (ABT1M2) onde o primeiro gráfico apresenta a variância cumulativa explicada à medida que mais componentes são adicionados, enquanto o segundo gráfico demonstra a redução do RMSE com o aumento do número de componentes. Observa-se que, a partir de duas componentes principais, a variância acumulada dos dados atinge aproximadamente 99%. No entanto, para garantir a minimização do erro entre os valores preditos e observados, a avaliação do RMSE foi utilizada como critério principal. Assim, para este modelo específico, foram selecionadas 11 componentes principais, visto que a partir da 12 componente principal a redução dos valores de RMSE se estabilizam, ficando próximo a zero.

**Figura 13.** Variância explicada acumulada (A) e redução do RMSE em função do número de componentes (B).



**Fonte:** Elaborado pelo autor.

Após o desenvolvimento e calibração dos modelos, a validação foi realizada utilizando um conjunto de dados inéditos, conforme demonstrado na Tabela 3, são demonstrado o quantitativo de amostras para cada conjunto de calibração e validação.

O desempenho das etapas de calibração e validação dos modelos de regressão PCR para a determinação do fósforo disponível, utilizando diferentes pré-processamentos dos espectros de reflectância na região VIS-NIR-SWIR (450-2500 nm) dos diversos conjuntos de dados, podem ser observados na Tabela 6.

**Tabela 6.** Desempenho dos modelos de regressão PCR para estimativa de fósforo disponível durante a calibração e validação com o espectro completo.

		Calibração				Validação					
Modelo	Dados	Proc.	Nº PCs	R <sup>2</sup>	RMSE	r <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup> ajust	RMSE	RPD	RPIQ
ABT1M2	RQO + PACd	Bruto	11	0.88	19.72	0.90	0.81	0.81	25.32	2.33	4.12
ABT2M2		NOR	10	0.87	20.95	0.89	0.80	0.79	26.36	2.23	3.96
ABT3M2		SG	12	0.90	18.50	0.94	0.87	0.86	21.32	2.76	4.89
ABT4M2		CR	13	0.81	25.13	0.83	0.68	0.68	32.70	1.80	3.19
AT1M2	RQO	Bruto	4	0.92	16.75	0.93	0.85	0.84	22.69	2.62	4.83
AT2M2		NOR	4	0.91	17.21	0.93	0.85	0.85	22.28	2.66	4.91
AT3M2		SG	3	0.89	19.34	0.95	0.90	0.90	18.15	3.27	6.03
AT4M2		CR	5	0.84	22.94	0.92	0.84	0.83	23.71	2.54	4.73
BT1M2	PACd	Bruto	11	0.93	14.83	0.96	0.90	0.90	18.17	3.22	5.91
BT2M2		NOR	8	0.86	21.35	0.89	0.78	0.77	26.95	2.17	3.98
BT3M2		SG	5	0.88	19.43	0.90	0.80	0.79	25.98	2.25	4.13
BT4M2		CR	10	0.91	17.09	0.80	0.57	0.56	37.71	1.55	2.85

**Fonte:** Elaborado pelo autor.

Nº PCs = Número de componentes principais; R<sup>2</sup> = Coeficiente de determinação; R<sup>2</sup> ajust = Coeficiente de determinação ajustado; RMSE = Raiz do erro quadrático médio; r<sup>2</sup> = coeficiente de correlação; RPD = Razão de performance de desvio; RPIQ = Razão entre desempenho e distância interquartil; RQO = Neossolo Quartizarênico órtico; PACd = Argissolo acinzentado distrófico; BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo.

De modo geral, os resultados para os três conjuntos de dados (AB, A e B) demonstraram que o uso da regressão PCR foi promissor na estimativa do fósforo disponível,

apresentando valores satisfatórios de  $R^2$  e RMSE na etapa de validação. Apresentando desempenho excelente, segundo a classificação de Chang et al. (2001), para todas as combinações de pré-processamento espectral e conjuntos de dados, exceto aqueles desenvolvidos com CR nos conjuntos AB e B, que foram classificados como de desempenho moderado na estimativa do teor de fósforo disponível em amostras de solo.

No conjunto AB, o desempenho dos métodos de pré-processamento variou entre os modelos, com valores de  $R^2$  entre 0.68 e 0.87, e RMSE e RPD variando de 21.32 a 32.70  $\text{mg kg}^{-1}$  e de 1.80 a 2.76, respectivamente (Tabela 6). Entre as abordagens avaliadas, apenas o método Savitzky-Golay (ABT3M2) apresentou resultados mais consistentes na calibração e validação, melhorando o ajuste em relação à reflectância bruta (ABT1M2). Isso se refletiu em  $R^2$  mais alto (0.87) e RMSE menor (21.32  $\text{mg kg}^{-1}$ ). Já o modelo com reflectância normalizada (ABT2M2), apesar de ter sido calibrado com um número menor de componentes principais, obteve desempenho comparável ao modelo de reflectância bruta, com valores de  $R^2$  e RMSE de 0.80 e 26.36  $\text{mg kg}^{-1}$ , respectivamente.

No conjunto de dados A, o comportamento foi similar ao observado nos dados AB. O modelo com pré-tratamento Savitzky-Golay (AT3M2) apresentou ligeiramente melhor desempenho, com  $R^2$  de 0.90 e RMSE de 18.15  $\text{mg kg}^{-1}$ , em comparação ao modelo com reflectância bruta (AT1M2), que obteve  $R^2$  de 0.85 e RMSE de 22.69  $\text{mg kg}^{-1}$ . Já os modelos que utilizaram reflectância normalizada (AT2M2) e continuum remove (AT4M2) demonstraram desempenho comparável, com  $R^2$  em torno de 0.85 e 0.84 e RMSE aproximado de 22.28 e 23.71  $\text{mg kg}^{-1}$ . No conjunto de dados B, o uso da reflectância bruta (BT1M2) apresentou desempenho superior aos demais pré-tratamentos espectrais na validação.

Cabe destacar que, a aplicação do pré-tratamento espectral continuum remove, embora tenha exigido em sua grande maioria o uso de um maior número de componentes em comparação aos demais métodos, demonstrou ser a abordagem menos precisa para os três conjuntos de dados analisados, tanto na calibração quanto na validação. Essa técnica, que visa eliminar a componente contínua dos espectros, não traduziu em ganhos no desempenho preditivo, evidenciado por valores de  $R^2$  mais baixos e RMSE mais altos, sugerindo que a remoção do contínuo pode ter eliminado informações relevantes ou intensificado a presença de ruídos no modelo.

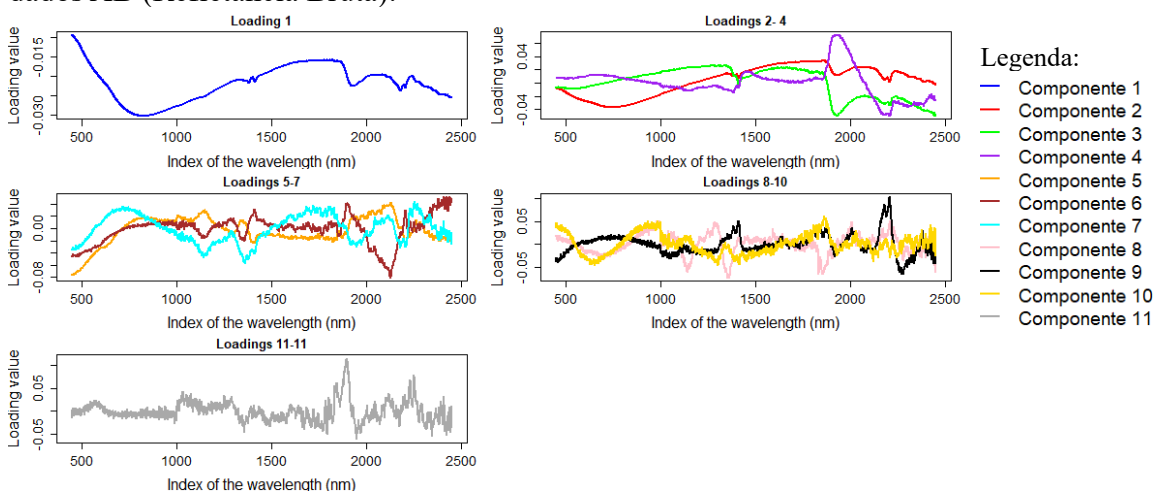
Em estudo sobre o desempenho dos modelos preditivos ao utilizar diferentes abordagens de pré-processamento em solos de Marrocos, Barra, Briak e Kebede (2022), observaram que, apesar de os pré-tratamentos na espectroscopia de solo terem o objetivo de reduzir o espalhamento e realçar sinais de interesse, a complexidade da matriz do solo exige a

preservação tanto dos dados de absorção pura quanto das informações do espalhamento. Assim, a remoção excessiva dos sinais e espalhamento pode prejudicar a performance dos modelos.

Na análise de componentes principais, os coeficientes denominados loadings refletem a influência individual de cada variável original na construção dos componentes latentes. Em particular, a utilização de gráficos de loadings permite identificar e quantificar a importância dos diferentes comprimentos de onda, evidenciando as contribuições específicas de cada um para os componentes principais e, conseqüentemente, aprimorando a compreensão da estrutura espectral dos dados.

A Figura 14 ilustra os loadings dos componentes principais selecionados para o conjunto de dados AB, obtidos a partir da reflectância bruta, os quais foram empregados na construção do modelo PCR. Nota-se que os loadings referentes ao componente 1 apresentam cargas negativas e com valores próximos a zero em todas as variáveis espectrais analisadas, o que indica uma contribuição uniforme, porém inversa, de cada faixa para esse componente. Contudo, determinadas faixas espectrais exerceram significativa influência na composição das componentes principais selecionadas.

**Figura 14.** Loadings espectrais nos componentes principais do modelo PCR – Conjunto de dados AB (Refletância Bruta).



**Fonte:** Elaborado pelo autor.

A região compreendida entre 850 e 900 nm evidenciou uma acentuada depressão para o primeiro loading, sendo essa faixa reconhecida por suas feições relacionadas aos óxidos de ferro. Ademais, as assinaturas espectrais atribuídas aos argilominerais do tipo 1:1 em torno de 1400, 1900 e 2200 nm destacaram-se na composição dos loadings. Demattê et al. (2017) e Coutinho et al. (2019) corroboram que os sinais em 1400, 1900 e 2200 nm respondem ao fósforo em virtude de sua ligação a oxi-hidróxidos de ferro e alumínio.

Os loadings sugerem a importância dos processos de sorção do fósforo em superfícies coloidais com carga variável, evidenciando que tais mecanismos são fundamentais para a regulação do nutriente entre as fases do solo. Durante a adubação fosfatada, a fonte mineral de fósforo se solubiliza, elevando sua concentração na solução, entretanto, para manter o equilíbrio entre as fases do solo, o excesso de fósforo é adsorvido na fase sólida na forma lábil, possibilitando reações de dessorção subsequente para a solução e consequentemente a disponibilidade do fósforo no solo, a depender do pH do solo (MELO; ALLEONI, 2019; NADEEM et al., 2024).

### 6.5.2 Regressão por mínimos quadrados parciais utilizando o espectro completo

Os resultados das etapas de calibração e validação dos modelos de regressão PLSR, aplicados na determinação do fósforo disponível, considerando diferentes arquiteturas de calibração que englobam os níveis de pré-processamento dos dados espectrais e os três conjuntos de dados utilizados estão apresentados na Tabela 7. A calibração dos modelos PLSR foi realizada de acordo com o mesmo protocolo empregado nos modelos PCR.

**Tabela 7.** Desempenho dos modelos de regressão PLSR para estimativa de fósforo disponível durante a calibração e validação com o espectro completo.

Calibração						Validação					
Modelo	Dados	Proc.	Nº PCs	R <sup>2</sup>	RMSE	r <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup> ajust	RMSE	RPD	RPIQ
ABT1M3	RQO + PACd	Bruto	13	0,99	5,83	0,97	0,94	0,94	14,04	4,19	7,43
ABT2M3		NOR	13	0,99	5,49	0,97	0,95	0,94	13,65	4,31	7,64
ABT3M3		SG	10	0,99	5,39	0,95	0,90	0,89	18,99	3,10	5,49
ABT4M3		CR	12	0,99	6,36	0,96	0,91	0,90	17,94	3,28	5,81
AT1M3	RQO	Bruto	10	0,99	5,31	0,98	0,95	0,94	12,96	4,58	8,45
AT2M3		NOR	10	0,99	4,80	0,98	0,95	0,94	12,75	4,66	8,59
AT3M3		SG	6	0,99	5,13	0,95	0,91	0,90	17,64	3,36	6,21
AT4M3		CR	10	0,99	3,29	0,94	0,89	0,85	19,93	3,02	5,63
BT1M3	PACd	Bruto	10	0,99	5,06	0,96	0,91	0,89	18,39	3,18	5,84
BT2M3		NOR	10	0,99	4,35	0,96	0,92	0,89	17,76	3,30	6,04
BT3M3		SG	7	0,99	3,66	0,91	0,82	0,79	25,19	2,32	4,26
BT4M3		CR	8	0,96	10,78	0,82	0,67	0,60	35,30	1,66	3,04

**Fonte:** Elaborado pelo autor.

Nº PCs = Número de componentes principais; R<sup>2</sup> = Coeficiente de determinação; R<sup>2</sup> ajust = Coeficiente de determinação ajustado; RMSE = Raiz do erro quadrático médio; r<sup>2</sup> = coeficiente de correlação; RPD = Razão de performance de desvio; RPIQ = Razão entre desempenho e distância interquartil; RQO = Neossolo Quartizarênico órtico; PACd = Argissolo acinzentado distrófico; BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo.

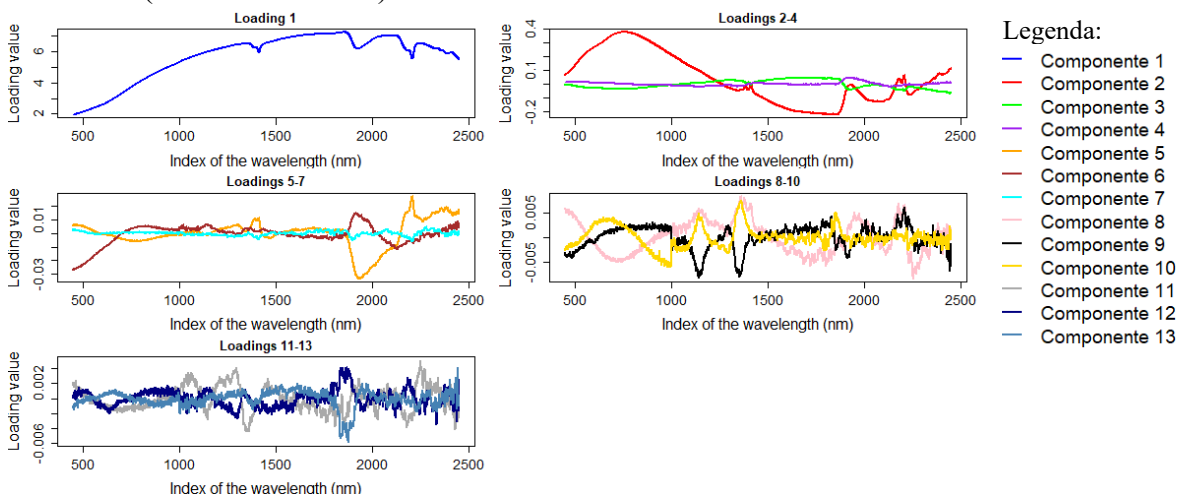
De forma geral, o desempenho dos modelos gerados a partir do método PLSR foram superiores aos observados pelo método PCR. O método PLSR se mostrou eficiente para a

calibração dos modelos preditivos de fósforo disponível considerando os níveis de processamento obtendo valores de  $R^2$  variando de 0,96 a 0,99, RMSE variando de 3,29 a 10,78  $\text{mg kg}^{-1}$  para calibração e valores de  $R^2$  variando entre 0,67 a 0,95, o RMSE situou-se entre 12,75 a 35,30  $\text{mg kg}^{-1}$  e RPD alcançou de 1,66 a 4,66. Obtendo desempenho segundo Chang et al., (2001) como excelente para todas as combinações de pré-processamento espectral e conjunto de dados ( $\text{RPD} > 2$  e  $R^2 > 0,8$ ), com exceção para o modelo desenvolvido com CR dentro do conjunto de dados do solo PAcd, que obteve desempenho classificados como moderado.

Dentro os níveis de pré-processamento, os dados com reflectância normalizados se mostrou superior aos demais métodos dentro dos três conjuntos de dados durante a etapa de validação do modelo embora os dados de reflectância bruta obteve desempenho comparável.

Os loadings da reflectância bruta do conjunto de dados AB, empregados na construção do modelo preditivo PLSR, são apresentados na Figura 15. O Loading 1 apresentou valores positivos crescentes ao longo do espectro, com destaque para inflexões acentuadas nas regiões de aproximadamente 1400, 1900 e 2200 nm, bandas amplamente reconhecidas por sua associação com a presença de água estrutural, grupos hidroxila e argilominerais do tipo 2:1.

**Figura 15.** Loadings espectrais nos componentes principais do modelo PLSR – Conjunto de dados AB (Refletância Bruta).



**Fonte:** Elaborado pelo autor.

O loading 2 mostra uma transição de positiva para negativa entre 1350 a 1400 nm, o que pode indicar sensibilidade a compostos orgânicos e óxidos entre 450 a 1100nm, além disso, faixas espectrais entre 1400 a 2200 nm demonstram oscilações marcadas, também ligadas a grupos funcionais de argilominerais e absorção de água.

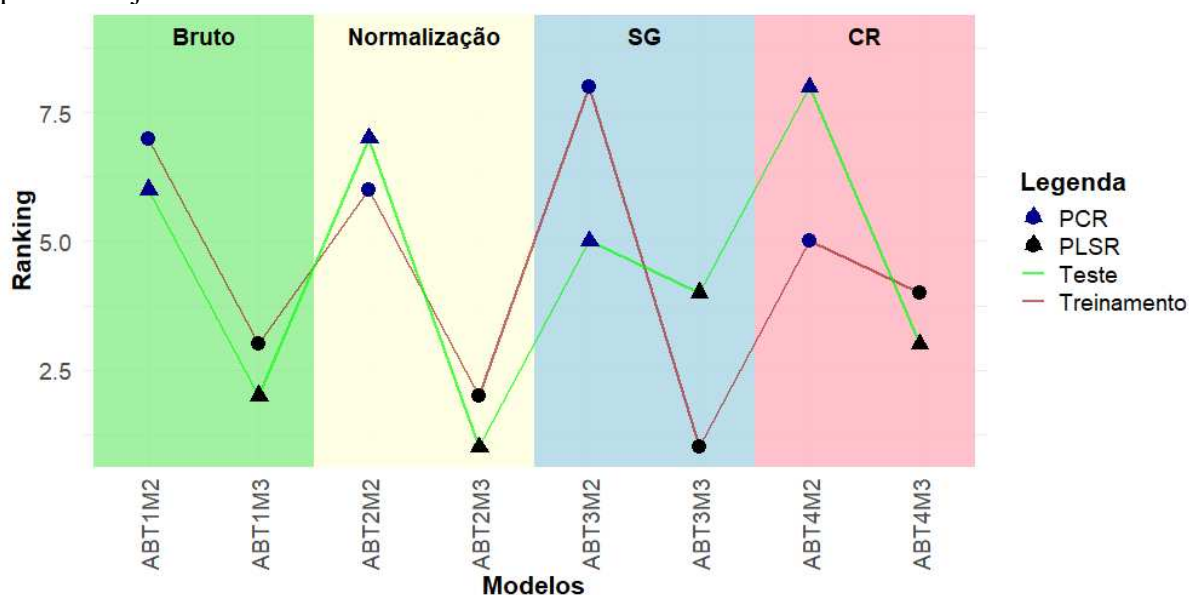
Os loadings 3 a 7 apresentaram valores uniformes e próximos de zero ao longo de todo o espectro, com pequenas variações localizadas. Em contraste, os loadings 8 a 13, apesar de exibirem feições associadas à água residual e a argilominerais, parecem refletir principalmente ruído ou componentes de pouca relevância na variação espectral total.

### 6.5.3 Ranking de seleção de modelos utilizando todo o espectro (450 - 2450nm)

Para selecionar o modelo com melhor desempenho, foi aplicada uma análise de decisão multicritério utilizando o método de soma ponderada proposto por Fishburn (1967). Foram considerados os valores de  $r^2$ ,  $R^2$ , RMSE e RPD durante as etapas de calibração e validação, atribuindo-se peso igual a cada parâmetro. A partir desses resultados, procedeu-se ao ranqueamento dos modelos para cada conjunto de dados.

Na Figura 16 apresenta-se o ranking de desempenho dos modelos preditivos construídos sobre o conjunto de dados AB, evidenciando o impacto das diferentes rotinas de pré-processamento e dos métodos de regressão.

**Figura 16.** Impacto do pré-processamento e do método de regressão no ranking de modelos para o conjunto de dados AB.



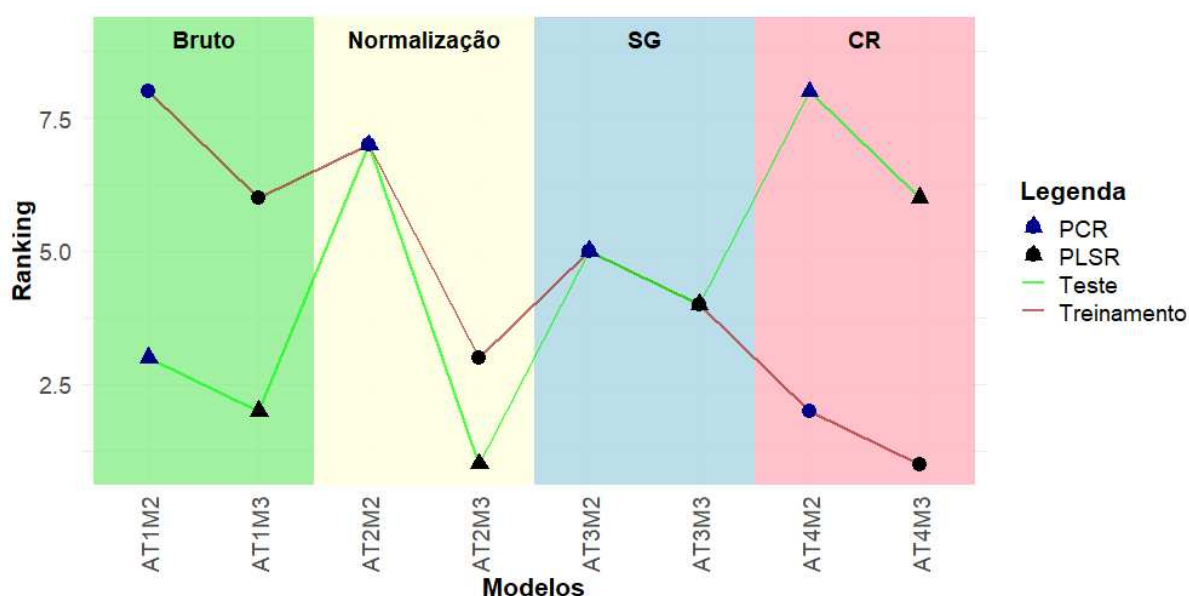
**Fonte:** Elaborado pelo autor.

BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo; PCR = Regressão por componentes principais; PLSR = Regressão por mínimos quadrados parciais.

Observou-se que a combinação da normalização da reflectância com o método PLSR (ABT2M3) apresentou o melhor desempenho, obtendo o menor ranking na fase de validação e, portanto, a maior capacidade preditiva para o teor de fósforo disponível no conjunto de dados RQO + PAcd.

De forma análoga, nos conjuntos isolados RQO (AT2M3) (Figura 17) e PAcd (BT2M3) (Figura 18), o PLSR aplicado à reflectância normalizada também superou as demais estratégias de pré-processamento, em especial aquelas baseadas no filtro de Savitzky-Golay e na remoção de contínuo.

**Figura 17.** Impacto do pré-processamento e do método de regressão no ranking de modelos para o conjunto de dados A.

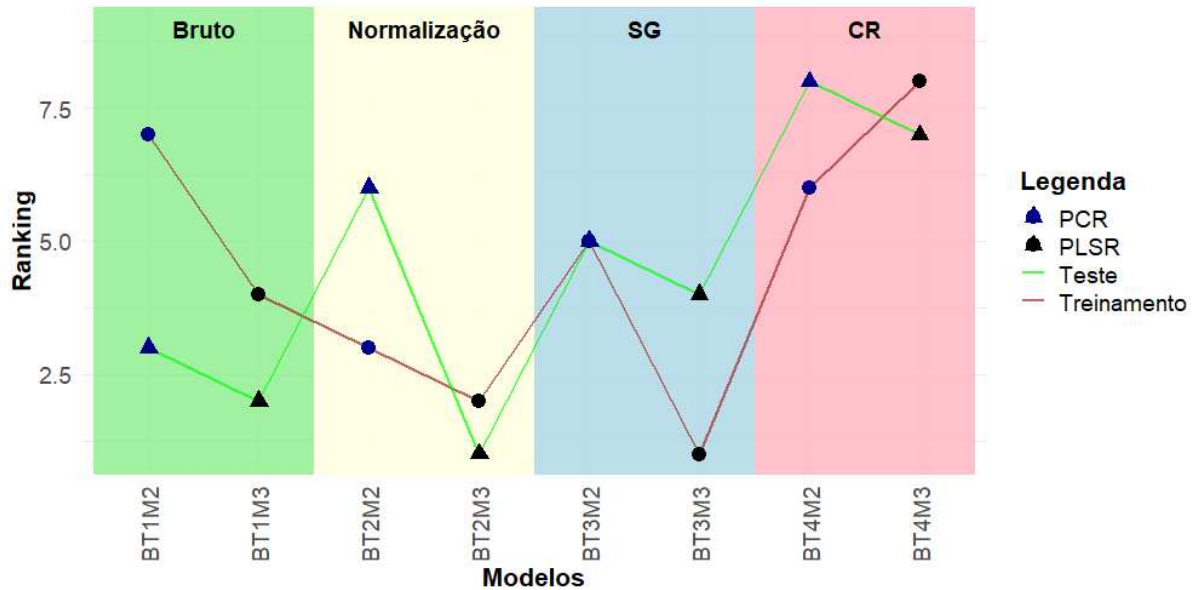


**Fonte:** Elaborado pelo autor.

BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo; PCR = Regressão por componentes principais; PLSR = Regressão por mínimos quadrados parciais.

Além disso, a comparação entre os métodos PCR e PLSR evidenciou a superioridade do PLSR em todos os níveis de pré-processamento durante a fase de validação, ressaltando sua maior sensibilidade às variações químicas nos dados espectrais relacionadas ao fósforo disponível.

**Figura 18.** Impacto do pré-processamento e do método de regressão no ranking de modelos para o conjunto de dados B.

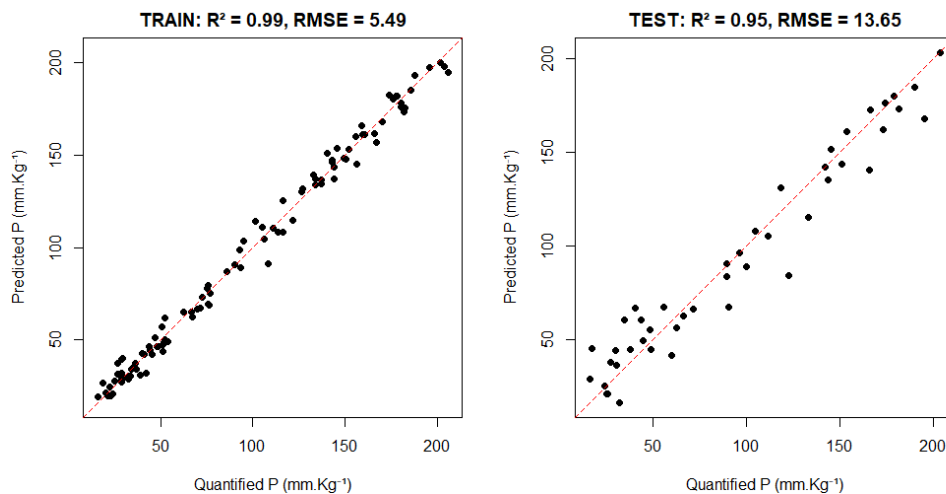


**Fonte:** Elaborado pelo autor.

BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo; PCR = Regressão por componentes principais; PLSR = Regressão por mínimos quadrados parciais.

Nas Figuras 19, 20 e 21 são apresentadas as correlações entre o fósforo quantificado e o fósforo predito pelo modelo PLSR com pré-processamento espectral por normalização, nos conjuntos AB, A e B, respectivamente.

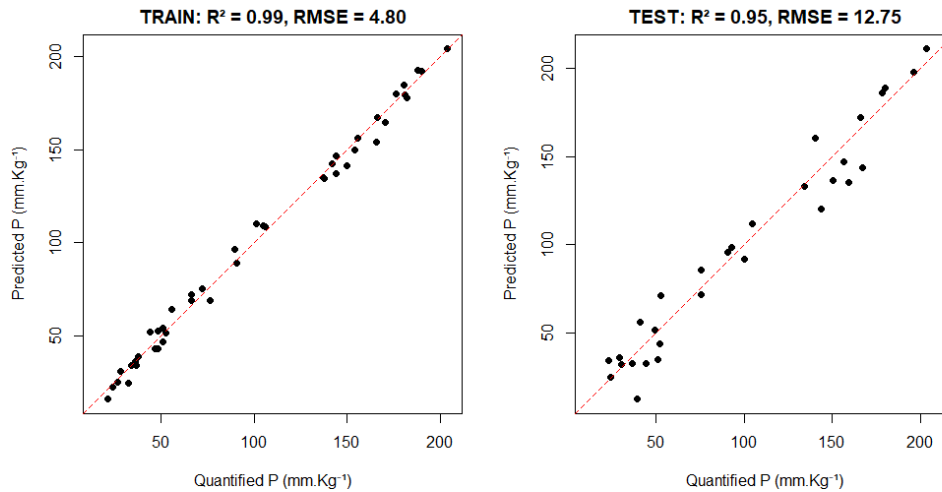
**Figura 19.** Previsões do modelo PLSR com pré-tratamento espectral normalizado nos conjuntos AB de calibração e validação.



**Fonte:** Elaborado pelo autor.

No conjunto AB, o modelo alcançou  $R^2 = 0,95$  e  $RMSE = 13,65 \text{ mg kg}^{-1}$ , no conjunto A,  $R^2 = 0,95$  e  $RMSE = 12,75 \text{ mg kg}^{-1}$ , e no conjunto B,  $R^2 = 0,95$  e  $RMSE = 17,76 \text{ mg kg}^{-1}$ , conforme já evidenciado na Tabela 7.

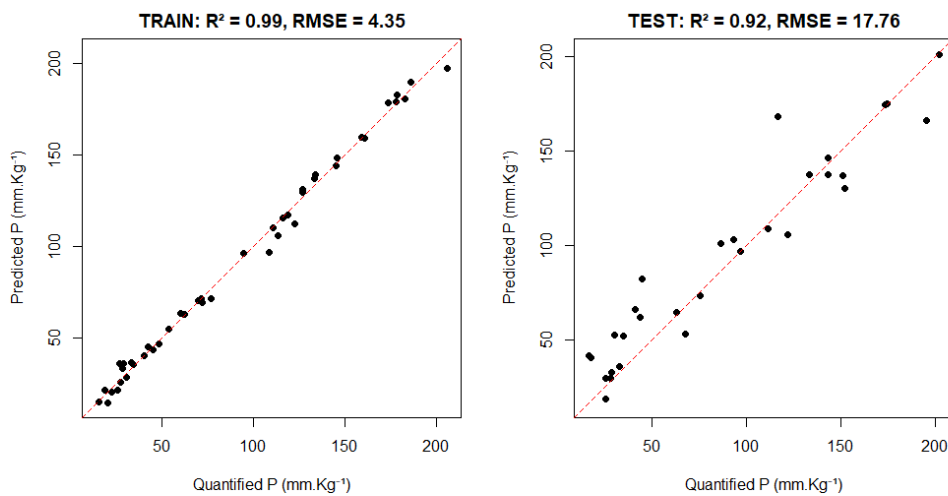
**Figura 20.** Previsões do modelo PLSR com pré-tratamento espectral normalizado nos conjuntos A de calibração e validação.



**Fonte:** Elaborado pelo autor.

Em todos os casos, a maioria dos pontos se alinha bem à reta 1:1, contudo, nota-se uma leve dispersão na faixa de concentração com teores mais baixos, em especial no ajuste do solo PACd (Figura 21).

**Figura 21.** Previsões do modelo PLSR com pré-tratamento espectral normalizado nos conjuntos B de calibração e validação.



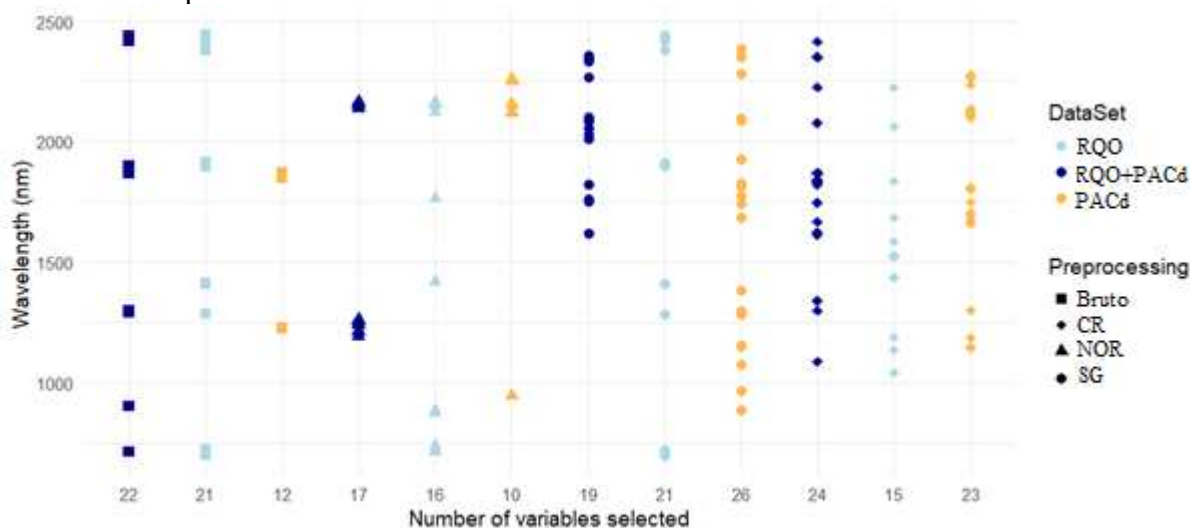
**Fonte:** Elaborado pelo autor.

## 6.6 Estimativa do teor de fósforo disponível com seleção de variáveis

Para reduzir a dimensionalidade do conjunto de dados e manter apenas os comprimentos de onda mais relevantes para a predição do teor de fósforo disponível, aplicou-se primeiramente um filtro de correlação, eliminando comprimentos de onda com baixa relação estatística com a concentração de fósforo. Em seguida, sobre as variáveis espectrais resultantes, utilizou-se o método Stepwise-Forward para selecionar de forma incremental as bandas mais significativas, contribuindo para que as variáveis escolhidas representem adequadamente a variação do teor de fósforo disponível em solo.

A Figura 22 apresenta os comprimentos de ondas selecionadas para o desenvolvimento dos modelos para predição de fósforo disponível, com base em diferentes pré-processamentos e conjuntos de dados (A, B e AB). As características de absorção selecionadas derivam da interação entre a radiação eletromagnética e os grupos funcionais presentes nas superfícies coloidais do solo, sendo observada uma ampla diversidade de grupos funcionais, tanto orgânicos quanto inorgânicos.

**Figura 22.** Distribuição e quantidade de variáveis selecionadas por conjunto de dados e pré-tratamentos espectrais.



**Fonte:** Elaborado pelo autor.

RQO = Neossolo Quartzarênico órtico; PACd = Argissolo acinzentado distrófico; BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo.

Os espectros brutos e normalizados apresentaram uma seleção de variáveis mais concentrada em regiões espectrais específicas, frequentemente associadas a componentes do

solo que influenciam direta ou indiretamente a disponibilidade de fósforo. Em contraste, os pré-processamentos por CR e SG resultaram em seleções mais distribuídas ao longo do espectro, com comprimentos de onda menos concentrados, porém ainda com feições relevantes.

Entre as variáveis escolhidas, observam-se comprimentos de onda nas faixas de 850 a 950 nm, que são comumente atribuídos a transições eletrônicas associadas ao  $\text{Fe}^{3+}$ , presente em minerais como goethita e hematita (RICHTER et al., 2009). Alguns comprimentos nessa faixa espectral (850 a 950 nm) e próximas de 2230 nm, por sua vez, também são frequentemente associadas a vibrações Al–Fe–OH, geralmente ligadas a argilominerais como esmectitas (STENBERG et al., 2010). Em torno de 2200 nm, ocorrem bandas amplamente descritas como características de combinações O–H e Al–OH, frequentemente relacionadas à presença de minerais como caulinita, mica (illita) e minerais do tipo 2:1 (vermiculita, esmectita).

Também foram selecionadas bandas em faixas entre 700 a 850 nm e 1000 a 1100 nm, regiões atribuídas a estiramentos de N–H e C–H que indicam a presença de compostos amínicos e de carbono aromático na matéria orgânica do solo (BEN-DOR; BANIN, 1995; DEMATTE et al., 2016).

De acordo com Fink et al. (2016), os grupos funcionais das substâncias húmicas ligam-se preferencialmente aos sítios de superfície de óxidos de alumínio ( $\text{Al}^{3+}$ ) e ferro ( $\text{Fe}^{3+}$ ), aumentando a competição com ânions fosfatados por sítios de adsorção ou ainda alterar as cargas superficiais desses sítios, dessa forma reduzindo a sorção de fósforo. Por outro lado, compostos orgânicos podem aumentar a disponibilidade de P por complexarem metais como o alumínio, inibindo a precipitação de fosfatos em formas de baixa solubilidade (INSKEEP; SILVERTOOTH, 1988; BORGGAARD et al., 2005).

Indiretamente, a matéria orgânica também eleva a capacidade de retenção hídrica e estimula a atividade microbiana, acelerando a mineralização de nutrientes e a mobilidade do fósforo, o que favorece sua disponibilidade para as plantas (MELO; ALLEONI, 2019).

Outros comprimentos de onda relevantes selecionados pelo algoritmo como 1200, 1400 e 1900 nm são descritos na literatura como associados às vibrações H–O–H e O–H da água, tanto na forma livre quanto adsorvida nos coloides do solo (STENBERG et al., 2010).

### **6.6.1 Estimação de fósforo no solo usando regressão linear múltipla (MLR) com variáveis selecionadas**

Após a seleção de variáveis, procedeu-se à calibração dos modelos utilizando os algoritmos quimiométricos PCR e PLSR, incluindo a regressão linear múltipla (MLR) como

parâmetro de comparação. Os resultados das etapas de calibração e validação dos modelos de regressão MLR após a seleção de variáveis, aplicados na determinação do fósforo disponível, considerando diferentes arquiteturas de calibração que englobam os níveis de pré-processamento dos dados espectrais e os três conjuntos de dados utilizados estão apresentados na Tabela 8.

**Tabela 8.** Desempenho dos modelos de regressão MLR para estimativa de fósforo disponível para o conjunto de calibração e validação após a seleção de variáveis.

Calibração						Validação					
Modelo	Dados	Proc.	Nº PCs	R <sup>2</sup>	RMSE	r <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup> ajust	RMSE	RPD	RPIQ
SABT1M1	RQO + PACd	Bruto	22	0,94	14,63	0,62	0,13	0,06	54,41	1,08	1,92
SABT2M1		NOR	17	0,92	16,41	0,92	0,84	0,73	23,65	2,49	4,41
SABT3M1		SG	19	0,88	19,96	0,85	0,69	0,46	32,26	1,83	3,23
SABT4M1		CR	24	0,90	18,03	0,81	0,62	0,16	35,96	1,64	2,90
SAT1M1	RQO	Bruto	21	0,86	21,75	0,59	-0,11	-3,01	61,41	0,97	1,78
SAT2M1		NOR	16	0,98	8,03	0,89	0,75	0,44	29,18	2,03	3,75
SAT3M1		SG	27	0,99	6,37	0,85	0,53	-5,75	39,82	1,49	2,75
SAT4M1		CR	15	0,97	9,25	0,89	0,76	0,49	28,76	2,09	3,90
SBT1M1	PACd	Bruto	12	0,77	27,21	0,74	0,28	-0,23	48,88	1,20	2,20
SBT2M1		NOR	10	0,93	15,34	0,75	0,45	0,17	42,56	1,38	2,52
SBT3M1		SG	26	0,98	8,26	0,51	-0,12	-9,85	60,96	0,96	1,76
SBT4M1		CR	23	0,98	7,48	0,73	0,42	-1,79	43,74	1,34	2,45

**Fonte:** Elaborado pelo autor.

Nº PCs = Número de componentes principais; R<sup>2</sup> = Coeficiente de determinação; R<sup>2</sup> ajust = Coeficiente de determinação ajustado; RMSE = Raiz do erro quadrático médio; r<sup>2</sup> = coeficiente de correlação; RPD = Razão de performance de desvio; RPIQ = Razão entre desempenho e distância interquartil; RQO = Neossolo Quartizarênico órtico; PACd = Argissolo acinzentado distrófico; BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo.

Nas calibrações para o fósforo com MLR, o desempenho mostrou-se bastante heterogêneo e, em sua maioria, inferior ao dos modelos PCR e PLSR, apresentando valores de R<sup>2</sup> variando de -0,12 a 0,84, RMSE entre 23,65 e 61,41 mg.kg<sup>-1</sup>, e RPD de 0,96 a 2,49. Em função dos baixos valores de R<sup>2</sup> e RPD, a maior parte dos modelos foi considerada insatisfatória para predição do fósforo. Apenas o modelo calibrado com a normalização da reflectância AB enquadrou-se na categoria excelente, conforme Chang et al. (2001), apresentando R<sup>2</sup> = 0,84, RMSE = 23,65 mg kg<sup>-1</sup> e RPD = 2,49.

Na Tabela 9 são apresentados os modelos de melhor desempenho para os conjuntos de dados AB, A e B, indicando o nível de pré-processamento espectral, os coeficientes de regressão e os valores de R<sup>2</sup> obtidos nos modelos calibrados por MLR. Para o MLR utilizando as variáveis selecionadas, o pré-processamento dos dados de reflectância normalizados se mostrou superior aos demais obtendo para os três conjuntos de dados um desempenho

ligeiramente melhor, com  $R^2$  de 0,84, 0,75 e 0,45 respectivamente para os conjuntos de dados AB, A e B.

**Tabela 9.** Descrição dos modelos de melhor desempenho calibrados por MLR para os Conjuntos AB, A e B.

Conjunto de dados	Tratamento espectral	Modelo	$R^2$
AB	NOR	$Y = 10468,45 - 488266 (1192 \text{ nm}) - 353477 (1211 \text{ nm}) + 385283,5 (1212 \text{ nm}) - 390095 (1214 \text{ nm}) + 327244,1 (1242 \text{ nm}) + 512235,5 (1243 \text{ nm}) - 575663 (1252 \text{ nm}) - 923038 (1253 \text{ nm}) + 272407 (1256 \text{ nm}) + 569758,4 (1259 \text{ nm}) + 559811,2 (1261 \text{ nm}) - 605870 (2136 \text{ nm}) + 523759,6 (2146 \text{ nm}) - 443661 (2154 \text{ nm}) + 290504,2 (2159 \text{ nm}) - 585123 (2162 \text{ nm}) + 506020,8 (2165 \text{ nm})$	0,84
A	NOR	$Y = 23326,93 - 483816 (714 \text{ nm}) + 1163605 (715 \text{ nm}) - 476444 (721 \text{ nm}) - 216250 (737 \text{ nm}) - 629736 (875 \text{ nm}) + 429788 (884 \text{ nm}) - 245193 (1417 \text{ nm}) + 276947,4 (1763 \text{ nm}) - 336309 (1766 \text{ nm}) - 567896 (2124 \text{ nm}) + 454820 (2151 \text{ nm}) - 642559 (2156 \text{ nm}) - 401301 (2161 \text{ nm}) + 845336,5 (2162 \text{ nm}) - 712489 (2163 \text{ nm}) + 567090,3 (2165 \text{ nm})$	0,75
B	NOR	$Y = 17620,46 + 1169165 (945 \text{ nm}) - 581393 (947 \text{ nm}) - 816635 (948 \text{ nm}) - 880769 (2119 \text{ nm}) + 313215,8 (2121 \text{ nm}) + 832743 (2153 \text{ nm}) - 1189242 (2155 \text{ nm}) + 383986,7 (2156 \text{ nm}) + 301555,7 (2251 \text{ nm}) - 261841 (2259 \text{ nm})$	0,45

**Fonte:** Elaborado pelo autor.

AB = Conjunto de dados do neossolo quartzarênico órtico + Argissolo acinzentado distrófico; A = Conjunto de dados do neossolo quartzarênico órtico; B = Conjunto de dados do Argissolo acinzentado distrófico;  $R^2$  = Coeficiente de determinação.

### 6.6.2 Estimação de fósforo no solo usando regressão por componentes principais (PCR) com variáveis selecionadas

Os resultados de calibração e validação dos modelos de regressão PCR, após a seleção de variáveis para determinação do fósforo disponível considerando diferentes níveis de pré-processamento espectral e os três conjuntos de dados estão apresentados na Tabela 10.

**Tabela 10.** Desempenho dos modelos de regressão PCR para estimativa de fósforo disponível para o conjunto de calibração e validação após a seleção de variáveis.

Modelo	Calibração					Validação					
	Dados	Proc.	Nº PCs	R <sup>2</sup>	RMSE	r <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup> ajust	RMSE	RPD	RPIQ
SABT1M2	RQO + PACd	Bruto	22	0,71	31,00	0,73	0,43	0,42	43,93	1,34	2,37
SABT2M2		NOR	17	0,92	16,41	0,92	0,84	0,73	23,65	2,49	4,41
SABT3M2		SG	19	0,88	19,96	0,85	0,69	0,69	32,26	1,83	3,23
SABT4M2		CR	24	0,90	18,03	0,81	0,62	0,16	35,96	1,64	2,90
SAT1M2	RQO	Bruto	21	0,61	36,12	0,83	0,69	0,67	32,75	1,81	3,34
SAT2M2		NOR	16	0,96	12,01	0,95	0,89	0,88	19,58	3,03	5,59
SAT3M2		SG	27	0,99	6,37	0,85	0,53	0,52	39,82	1,49	2,75
SAT4M2		CR	15	0,96	11,98	0,90	0,81	0,80	25,84	2,33	4,34
SBT1M2	PACd	Bruto	12	0,74	29,09	0,70	0,17	0,14	52,38	1,12	2,05
SBT2M2		NOR	10	0,89	18,61	0,80	0,60	0,59	36,25	1,61	2,96
SBT3M2		SG	20	0,92	16,22	0,85	0,71	0,70	31,06	1,88	3,46
SBT4M2		CR	23	0,92	16,57	0,85	0,70	0,69	31,47	1,86	3,41

**Fonte:** Elaborado pelo autor.

Nº PCs = Número de componentes principais; R<sup>2</sup> = Coeficiente de determinação; R<sup>2</sup> ajust = Coeficiente de determinação ajustado; RMSE = Raiz do erro quadrático médio; r<sup>2</sup> = coeficiente de correlação; RPD = Razão de performance de desvio; RPIQ = Razão entre desempenho e distância interquartil; RQO = Neossolo Quartizarênico órtico; PACd = Argissolo acinzentado distrófico; BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo.

Em comparação, os modelos PCR calibrados com variáveis selecionadas superaram aqueles obtidos por MLR e PLSR, apresentando, em sua maioria, maiores valores de R<sup>2</sup> e RPD entretanto maiores valores de RMSE quando comparados ao PLSR (Tabela 12). Os indicadores variaram de R<sup>2</sup> = 0,17 a 0,89, RPD = 1,12 a 3,03 e RMSE = 19,58 a 52,38 mg kg<sup>-1</sup>. Esse desempenho superior é corroborado pelo maior número de modelos classificados como excelentes e moderados para a predição do fósforo segundo Chang et al. (2001).

Na Tabela 11 são apresentados os modelos de melhor desempenho para os conjuntos de dados AB, A e B, indicando o nível de pré-processamento espectral, os coeficientes de regressão e os valores de R<sup>2</sup> obtidos nos modelos calibrados por PCR.

**Tabela 11.** Descrição dos modelos de melhor desempenho calibrados por PCR para os Conjuntos AB, A e B.

Amostra	Tratamento espectral	Modelo	R <sup>2</sup>
AB	NOR	Y = 10468,45 – 488266 (1192 nm) - 353477 (1211nm) + 385283,5 (1212 nm) - 390095 (1214nm) + 327244,1 (1242 nm) + 512235,5 (1243 nm) - 575663 (1252 nm) - 923038 (1253 nm) + 272407 (1256nm) + 569758,4 (1259 nm) + 559811,2 (1261 nm) - 605870 (2136 nm) + 523759,6 (2146 nm) - 443661 (2154 nm) + 290504,2 (2159 nm) - 585123 (2162 nm) + 506020,8 (2165nm)	0,84
A	NOR	Y = 30077,1 - 11637,9 (714 nm) - 50602,1 (715 nm) + 4070,537 (721 nm) + 40851,42 (737 nm) - 80618,1 (875 nm) - 184015 (884nm) - 412734 (1417nm) + 80689,03 (1763 nm) - 165384 (1766 nm) - 598968 (2124 nm) + 54747,76 (2151 nm) - 72612,4 (2156 nm) - 13832,2 (2161 nm) + 1678,506 (2162 nm) - 60316,7 (2163 nm) + 188379,8 (2165 nm)	0,89
B	SG	Y = 79,42317 + 37949,96 (886 nm) + 63720,19 (889 nm) - 55169,3 (968 nm) - 120044 (1074 nm) + 54112,37 (1150 nm) + 43801,44 (1154 nm) - 100153 (1285 nm) + 78857,74 (1296 nm) - 108651 (1382 nm) - 82833,9 (1685 nm) - 37534,2 (1743 nm) - 122229 (1777 nm) - 3000,85 (1815 nm) - 30058,1 (1819) + 10404,31 (1820 nm) - 166597 (1827 nm) - 183757 (1926 nm) + 5307,053 (2085 nm) + 13189,17 (2086 nm) - 75429,9 (2097 nm) - 117753 (2098 nm) - 11516,3 (2280 nm) + 104325,4 (2281 nm) - 68053,6 (2353 nm) + 47246,22 (2357 nm) + 92094,29 (2383 nm)	0,71

**Fonte:** Elaborado pelo autor.

AB = Conjunto de dados do neossolo quartzarênico órtico + Argissolo acinzentado distrófico; A = Conjunto de dados do neossolo quartzarênico órtico; B = Conjunto de dados do Argissolo acinzentado distrófico; R<sup>2</sup> = Coeficiente de determinação.

Para a calibração dos modelos com seleção de variáveis e algoritmo PCR, o pré-processamento dos dados de reflectância normalizados se mostrou superior para os conjuntos de dados AB e A com R<sup>2</sup> de 0,84 e 0,89, respectivamente. Enquanto para o conjunto de dados B os dados de reflectância corrigidos utilizando o filtro SG se mostrou superior com R<sup>2</sup> = 0,71.

### 6.6.3 Estimação de fósforo no solo usando regressão por mínimos quadrados parciais (PLSR) com variáveis selecionadas

A Tabela 12 mostra os resultados de calibração e validação dos modelos PLSR após seleção de variáveis para estimar o fósforo disponível em três conjuntos de dados, considerando diferentes níveis de pré-processamento espectral.

**Tabela 12.** Desempenho dos modelos de regressão PLSR para estimativa de fósforo disponível para o conjunto de calibração e validação após a seleção de variáveis.

Calibração					Validação						
Modelo	Dados	Proc.	Nº PCs	R <sup>2</sup>	RMSE	r <sup>2</sup>	R <sup>2</sup>	R <sup>2</sup> ajust	RMSE	RPD	RPIQ
SABT1M3	RQO +	Bruto	11	0,70	31,77	0,74	0,54	0,49	41,17	1,43	2,53
SABT2M3		NOR	7	0,80	25,90	0,82	0,67	0,65	34,28	1,72	3,04
SABT3M3		SG	10	0,86	21,32	0,86	0,73	0,71	32,02	1,84	3,26
SABT4M3		CR	9	0,90	18,55	0,81	0,65	0,62	35,40	1,66	2,95
SAT1M3	RQO	Bruto	7	0,60	36,71	0,77	0,60	0,52	37,25	1,59	2,94
SAT2M3		NOR	9	0,95	12,37	0,88	0,78	0,72	32,27	1,84	3,39
SAT3M3		SG	9	0,77	27,46	0,81	0,66	0,57	36,04	1,65	3,04
SAT4M3		CR	10	0,92	16,14	0,87	0,75	0,68	30,77	1,96	3,65
SBT1M3	PACd	Bruto	8	0,77	27,46	0,73	0,54	0,44	48,45	1,21	2,22
SBT2M3		NOR	7	0,68	32,41	0,60	0,36	0,24	52,53	1,11	2,04
SBT3M3		SG	8	0,93	14,86	0,86	0,75	0,69	29,11	2,01	3,69
SBT4M3		CR	7	0,97	9,03	0,74	0,55	0,46	41,28	1,42	2,60

**Fonte:** Elaborado pelo autor.

Nº PCs = Número de componentes principais; R<sup>2</sup> = Coeficiente de determinação; R<sup>2</sup> ajust = Coeficiente de determinação ajustado; RMSE = Raiz do erro quadrático médio; r<sup>2</sup> = coeficiente de correlação; RPD = Razão de performance de desvio; RPIQ = Razão entre desempenho e distância interquartil; RQO = Neossolo Quartizarênico órtico; PACd = Argissolo acinzentado distrófico; BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo.

Segundo a classificação de Chang et al. (2001), todos os modelos alcançaram desempenho moderado nos conjuntos AB e A. Já no conjunto B, apenas o modelo calibrado com espectros corrigidos pelo filtro SG manteve desempenho moderado, os demais foram classificados como de baixa capacidade preditiva para o fósforo disponível. Dessa forma, embora os modelos PLSR calibrados com o espectro completo tenham se mostrado superiores ao PCR, essa vantagem não se manteve após a seleção de variáveis.

No estudo de Sousa, Franco e Amaral (2019), avaliou-se a capacidade de estimar a disponibilidade de fósforo no solo por espectroscopia. Foram utilizados dois tipos de solo, mas os resultados obtidos não permitiram a estimativa efetiva da concentração de fósforo. As métricas de erro mostraram um R<sup>2</sup> de 0,48 e um RPD de 1,36, indicando desempenho para estimar o fosforo muito baixos. Embora o estudo tenha empregado diferentes tipos de solo e sem uso de métodos de pré-processamento espectral.

Rabie et al. (2020) obtiveram um modelo de predição de fósforo em quatro áreas agrícolas utilizando regressão por máquina de vetores de suporte (RSVM) associada a um algoritmo genético para otimização de PLS (GA-PLS). Ao estimar o fósforo extraível pelo método de Olson, o modelo RSVM-GA-PLS atingiu R<sup>2</sup> = 0,77, RMSE = 20,09 mg kg<sup>-1</sup> e RPD = 1,90.

Na Tabela 13 são apresentados os modelos de melhor desempenho para os

conjuntos de dados AB, A e B, indicando o nível de pré-processamento espectral, os coeficientes de regressão e os valores de  $R^2$  obtidos nos modelos calibrados por PLSR.

**Tabela 13.** Descrição dos modelos de melhor desempenho calibrados por PLSR para os Conjuntos AB, A e B.

Amostra	Tratamento espectral	Modelo	$R^2$
AB	SG	$Y = -352254 (1619 \text{ nm}) + 179234,8 (1752 \text{ nm}) + 183813,8 (1761 \text{ nm}) - 226083 (1762 \text{ nm}) + 224389,3 (1820 \text{ nm}) - 49963,8 (1821 \text{ nm}) - 324920 (2011 \text{ nm}) + 303693,3 (2013 \text{ nm}) + 316289,7 (2028 \text{ nm}) + 190255 (2055 \text{ nm}) - 95804,6 (2080 \text{ nm}) - 76063,1 (2087 \text{ nm}) - 427374 (2099 \text{ nm}) + 73978,47 (2100 \text{ nm}) - 308216 (2266 \text{ nm}) + 274926,9 (2267 \text{ nm}) + 138400,1 (2333 \text{ nm}) + 41800,21 (2348 \text{ nm}) - 90309,3 (2355 \text{ nm})$	0,73
A	NOR	$Y = 134407,3 (714 \text{ nm}) + 663879 (715 \text{ nm}) - 506853 (721 \text{ nm}) - 243541 (737 \text{ nm}) - 22798,3 (875 \text{ nm}) + 3351,274 (884 \text{ nm}) + 181165,7 (1417 \text{ nm}) + 619519,2 (1763 \text{ nm}) - 470851 (1766 \text{ nm}) - 591046 (2124 \text{ nm}) + 738718,1 (2151 \text{ nm}) - 501049 (2156 \text{ nm}) - 64779,6 (2161 \text{ nm}) + 338654,4 (2162 \text{ nm}) - 479245 (2163 \text{ nm}) + 262414,7 (2165 \text{ nm})$	0,78
B	SG	$Y = 140771,5 (886 \text{ nm}) + 157949,9 (889 \text{ nm}) - 288286 (968 \text{ nm}) - 127372 (1074 \text{ nm}) + 133121,6 (1150 \text{ nm}) + 154070,2 (1154 \text{ nm}) - 249050 (1285 \text{ nm}) + 134326,2 (1296 \text{ nm}) - 27358,7 (1382 \text{ nm}) - 121341 (1685 \text{ nm}) - 123712 (1743 \text{ nm}) - 162441 (1777 \text{ nm}) - 32907,8 (1815 \text{ nm}) - 246171 (1819) + 390031,9 (1820 \text{ nm}) - 43100,4 (1827 \text{ nm}) + 78423,89 (1926 \text{ nm}) + 237751 (2085 \text{ nm}) - 137826 (2086 \text{ nm}) + 87801,03 (2097 \text{ nm}) - 257699 (2098 \text{ nm}) - 79769,7 (2280 \text{ nm}) + 169605,7 (2281 \text{ nm}) - 122061 (2353 \text{ nm}) - 68558,5 (2357 \text{ nm}) + 85418,91 (2383 \text{ nm})$	0,75

**Fonte:** Elaborado pelo autor.

AB = Conjunto de dados do neossolo quartzarênico órtico + Argissolo acinzentado distrófico; A = Conjunto de dados do neossolo quartzarênico órtico; B = Conjunto de dados do Argissolo acinzentado distrófico;  $R^2$  = Coeficiente de determinação.

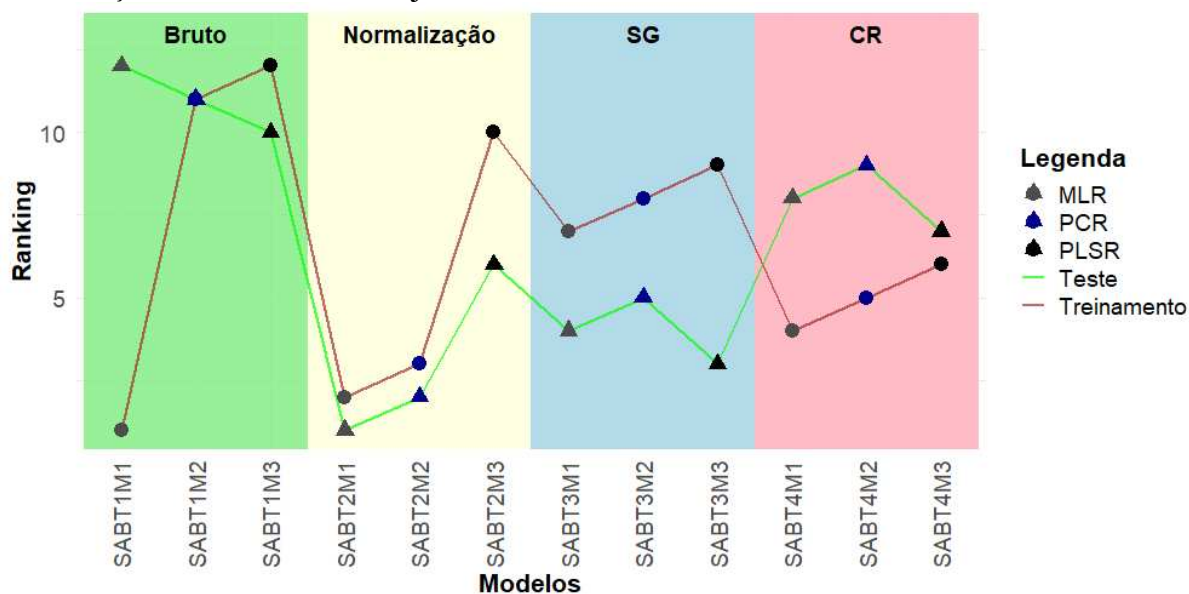
Para a calibração dos modelos com seleção de variáveis e algoritmo PLSR, o pré-processamento dos dados de reflectância corrigidos utilizando o filtro SG se mostrou superior para os conjuntos de dados AB e B com  $R^2$  de 0,73 e 0,75, respectivamente. Enquanto para o conjunto de dados A, os dados de reflectância normalizados se mostrou superior com  $R^2 = 0,78$ .

#### 6.6.4 Ranking de seleção de modelos calibrados a partir da seleção de variáveis

Assim como para o espectro completo, após a seleção de variáveis foi realizado o ranqueamento dos modelos em cada conjunto de dados AB, A e B. Para isso, consideraram-se os valores de  $r^2$ ,  $R^2$ , RMSE e RPD nas etapas de calibração e validação, atribuindo-se peso igual a cada métrica. Na Figura 23 apresenta-se o ranking de desempenho dos modelos preditivos

construídos sobre o conjunto de dados AB, evidenciando o impacto das diferentes rotinas de pré-processamento e dos métodos de regressão.

**Figura 23.** Impacto do pré-processamento e do método de regressão no ranking de modelos com seleção de variáveis do conjunto de dados AB.



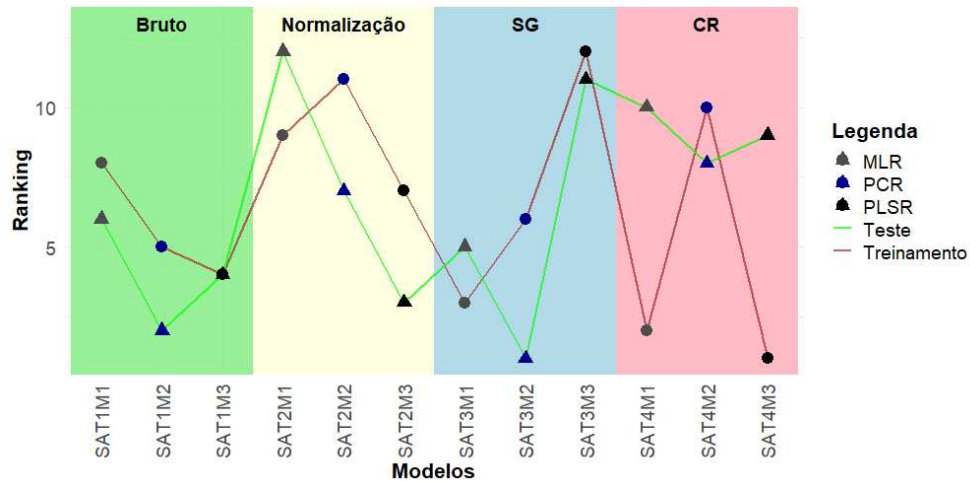
**Fonte:** Elaborado pelo autor.

BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo; MLR = Regressão linear múltipla; PCR = Regressão por componentes principais; PLSR = Regressão por mínimos quadrados parciais.

Nota-se que para o conjunto de dados AB, para a calibração dos modelos o MLR se mostrou superior em desempenho em todos os níveis de pré-processamento dos dados espectrais e obteve o melhor desempenho de validação geral quando combinado ao pré-processamento da reflectância normalizada (SABT2M1), entretanto, o mesmo não se mostrou eficiente durante os demais níveis de pré-processamento, sendo superado nos demais cenários de pré-processamento na etapa de validação pelos modelos desenvolvidos por PLSR (Figura 23).

Conforme a Figura 24, observa-se que a combinação do filtro SG com o método PCR (SAT3M2) apresentou o melhor desempenho, obtendo o menor ranking na fase de validação e, portanto, a maior capacidade preditiva para o teor de fósforo disponível no conjunto de dados RQO.

**Figura 24.** Impacto do pré-processamento e do método de regressão no ranking de modelos com seleção de variáveis do conjunto de dados A.

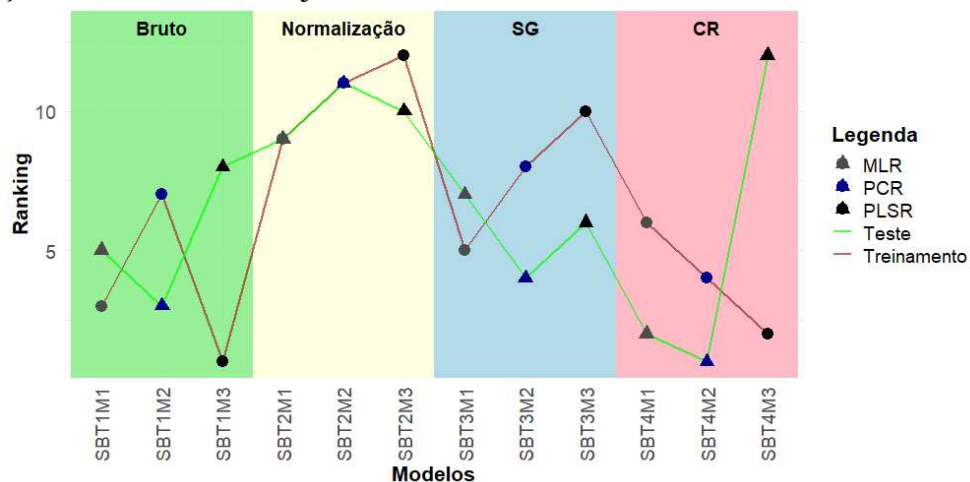


**Fonte:** Elaborado pelo autor.

BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo; MLR = Regressão linear múltipla; PCR = Regressão por componentes principais; PLSR = Regressão por mínimos quadrados parciais.

De forma análoga, no conjunto PAcd (SBT4M2) (Figura 25), o PCR foi o algoritmo quimiométrico de melhor desempenho quando aplicado à reflectância com remoção do contínuo.

**Figura 25.** Impacto do pré-processamento e do método de regressão no ranking de modelos com seleção de variáveis do conjunto de dados B.

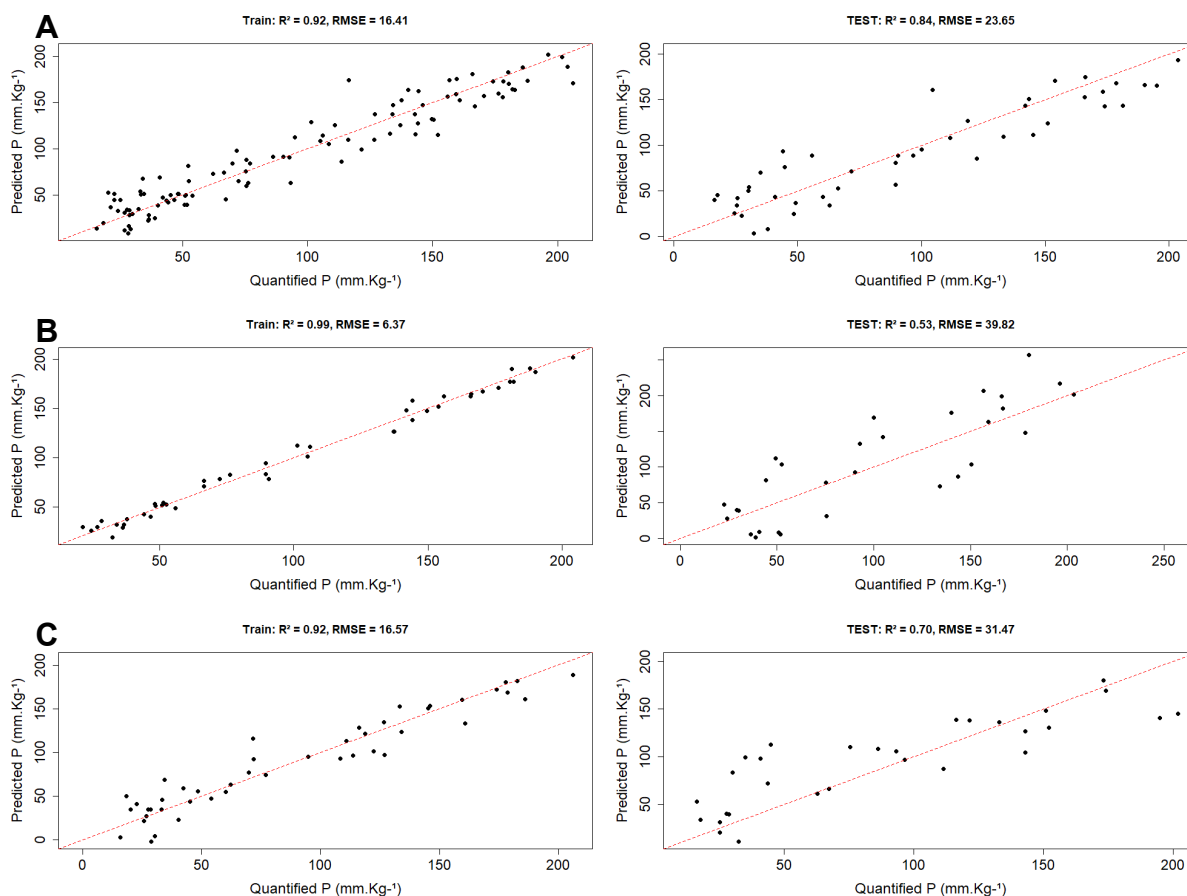


**Fonte:** Elaborado pelo autor.

BRUTO = Reflectância sem processamento; NOR = Reflectância normalizada por comprimento de vetor; SG = Primeira derivada com Savitzky-Golay; CR = Remoção do contínuo; MLR = Regressão linear múltipla; PCR = Regressão por componentes principais; PLSR = Regressão por mínimos quadrados parciais.

Nas Figuras 26A, 26B e 26C são apresentadas as correlações entre o fósforo quantificado e o fósforo predito pelos modelos com o maior ranking dentro dos conjuntos AB, A e B respectivamente.

**Figura 26.** Previsões de calibração e validação dos três modelos melhor ranqueados: (A) MLR com pré-tratamento espectral normalizado no conjunto RQO + PACd; (B) PCR com pré-tratamento SG no conjunto RQO; (C) PCR com pré-tratamento CR no conjunto PACd.



**Fonte:** Elaborado pelo autor.

No conjunto AB, o modelo alcançou  $R^2 = 0,84$  e  $RMSE = 23,65 \text{ mg kg}^{-1}$ , no conjunto A,  $R^2 = 0,53$  e  $RMSE = 39,82 \text{ mg kg}^{-1}$ , e no conjunto B,  $R^2 = 0,70$  e  $RMSE = 31,47 \text{ mg kg}^{-1}$ , conforme já evidenciado nas Tabelas 8 e 10. Nota-se uma leve dispersão dos dispersão no ajuste principalmente nos solos RQO e PACd (Figura 26 B e C).

## 7. CONCLUSÃO

Este estudo demonstra o potencial da espectroscopia de reflectância na faixa de 350 a 2.500 nm como método eficiente para estimar o fósforo disponível em solos, permitindo sua quantificação com desempenho comparável ao de métodos laboratoriais convencionais. Os

modelos calibrados apresentaram elevada precisão, com destaque para o uso da normalização espectral como pré-processamento, que realçou feições associadas ao fósforo e favoreceu o ajuste das calibrações. Entre os algoritmos testados, o PLSR aplicado aos dados normalizados obteve os melhores resultados nos diferentes tipos de solo avaliados.

O desempenho preditivo eficiente dos modelos em diferentes abordagens, aliado à identificação de bandas espectrais fortemente correlacionadas a propriedades físico-químicas que influenciam a dinâmica do fósforo, evidencia a aplicabilidade da técnica como alternativa complementar às análises laboratoriais, possibilitando maior agilidade, menor custo e sustentabilidade na caracterização do fósforo em solos.

## REFERÊNCIAS

- ARAÚJO, S. R.; SÖDERSTRÖM, M.; ERIKSSON, J.; ISENDAHL, C.; STENBORG, P.; DEMATTÊ, J.A.M. Determining soil properties in Amazonian Dark Earths by reflectance spectroscopy. **Geoderma**, v. 237-238, p. 308–317, 2015. DOI: <https://doi.org/10.1016/j.geoderma.2014.09.014>.
- BAH, A.; BALASUNDRAM, S.K; HUSNI, M.H. Sensor Technologies for Precision Soil Nutrient Management and Monitoring. **American Journal of Agricultural and Biological Sciences**, v. 7, n. 1, p. 43–49, 2012. DOI: <https://doi.org/10.3844/ajabssp.2012.43.49>.
- BARRA, I.; HAEFELE, S. M.; SAKRABANI, R.; KEBEDE, F. Soil spectroscopy with the use of chemometrics, machine learning and pre-processing techniques in soil diagnosis: Recent advances–A review. **TrAC Trends in Analytical Chemistry**, v. 135, p. 116166, 2021. DOI: <https://doi.org/10.1016/j.trac.2020.116166>.
- BARRA, I.; BRIAK, H.; KEBEDE, F. The application of statistical preprocessing on spectral data does not always guarantee the improvement of the predictive quality of multivariate models: Case of soil spectroscopy applied to Moroccan soils. **Vibrational spectroscopy**, v. 121, p. 103409–103409, 2022. DOI: <https://doi.org/10.1016/j.vibspec.2022.103409>.
- BASTOS, A. L.; COSTA, J. P. V. DA.; SILVA, I. DE F. DA.; RAPOSO, R. W. C.; SOUTO, J. S. Influência de doses de fósforo no fluxo difusivo em solos de Alagoas. **Revista Brasileira De Engenharia Agrícola E Ambiental**, v. 12, n.2, p.136–142, 2008. DOI: <https://doi.org/10.1590/S1415-43662008000200005>.
- BEC, K. B.; GRABSKA, J.; HUCK, C. W. Principles and applications of miniaturized near-infrared (NIR) spectrometers. **Chem. Eur. J.** v. 27, p.1514–1532, 2021. DOI: [doi:10.1002/chem.202002838](https://doi.org/10.1002/chem.202002838).
- BEN-DOR, E.; & BANIN, A. Near-Infrared Analysis as a Rapid Method to Simultaneously Evaluate Several Soil Properties. **Soil Science Society of America Journal**, v.59. n.2, p 364–372, 1995. <https://doi.org/10.2136/sssaj1995.03615995005900020014x>
- BO STENBERG; VISCARRA, R. R. A.; MOUAZEN, A. M.; WETTERLIND, J. Visible and Near Infrared Spectroscopy in Soil Science. **Advances in Agronomy**, p. 163–215, 2010. DOI: [https://doi.org/10.1016/s0065-2113\(10\)07005-7](https://doi.org/10.1016/s0065-2113(10)07005-7).
- BORGGGAARD, O. K.; RABEN-LANGE, B.; GIMSING, A. L.; STROBEL, B. W. Influence of humic substances on phosphate adsorption by aluminium and iron oxides. **Geoderma**, v. 127, p. 270–279, 2005. DOI: <https://doi.org/10.1016/j.geoderma.2004.12.011>.
- BREMNER, J. M.; MULVANEY, C. S. Nitrogen - Total. In: PAGE, A. L.; MILLER, R. H.; KEENEY, D. R. (Ed.). **Methods of soil analysis: part 2: chemical and microbiological properties**. 2nd ed. Madison: American Society of Agronomy, 1982. p. 595-624.
- CHANG, C., LAIRD, D.A., MAUSBACH, M.J. AND HURBURGH, C.R. Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. **Soil Sci. Soc. Am. J.**, n. 65, p. 480-490, 2001. DOI: <https://doi.org/10.2136/sssaj2001.652480x>.

CHAUHAN, R.; KUMAR, R.; KUMAR, V.; SHARMA, K.; SHARMA, V. On the discrimination of soil samples by derivative diffuse reflectance UV–vis-NIR spectroscopy and chemometric methods. **Forensic Science International**, v. 319, p. 110655, 2021. DOI: <https://doi.org/10.1016/j.forsciint.2020.110655>.

CLARK, R. N.; ROUSH, T. L. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. **J. Geophys. Res.**, v. 89, p. 6329–6340, 1984. DOI: <https://doi.org/10.1029/JB089iB07p06329>.

COBLINSKI, J. A.; INDA, A. V.; DEMATTÊ, J.A.M.; DOTTO, A.C.; GHOLIZADEH, A.; GIASSON, E. Identification of minerals in subtropical soils with different textural classes by VIS–NIR–SWIR reflectance spectroscopy. **CATENA**, v. 203, p. 105334, 2021. DOI: <https://doi.org/10.1016/j.catena.2021.105334>.

CONKLIN, A. R. **Introduction to Soil Chemistry: Analysis and Instrumentation**. Somerset: Wiley, 2013.

COSTA, J. F.; GIASSON, É.; DA SILVA, E. B.; TIECHER, T.; DE SENA, A.F.S.; SILVA, R.G.P.O. Soil texture prediction through stratification of a regional soil spectral library. **Pedosphere**, v. 32, n. 2, p. 294–306, 2021. DOI: [https://doi.org/10.1016/S1002-0160\(21\)60074-7](https://doi.org/10.1016/S1002-0160(21)60074-7).

COUTINHO, M. A.N.; ALARI, F. O.; FERREIRA, M. M.C.; AMARAL, L. R. Influence of soil sample preparation on the quantification of NPK content via spectroscopy. **Geoderma**, v. 338, p. 401–409, 2019. DOI: <https://doi.org/10.1016/j.geoderma.2018.12.021>.

DEMATTÊ, J. A. M.; MORGAN, C. L.S.; CHABRILLAT, S.; RIZZO, R.; FRANCESCHINI, M. H. D.; TERRA, F. DA S.; VASQUES, G. M.; WETTERLIND, J. Spectral Sensing from Ground to Space in Soil Science: State of the Art, Applications, Potential, and Perspectives. In: THENKABAIL, P. S. **Remote Sensing handbook**, vol II: Land resources monitoring, modelin d mapping with remote Sensing. Ed. Boca Raton, FL: CRC press (Taylor& Francis Group), 2016, p. 661-732.

DEMATTÊ, J. A.M.; RAMIREZ-LOPEZ, L.; KARINA P.P. 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. DOI: <https://doi.org/10.1016/j.geoderma.2016.11.013>.

DI RAIMO, L. A. D. L.; COUTO, E. G.; DE MELLO, D. C.; DEMATTÊ, J.A.M.; AMORIM, R.S.S.; TORRES, G.N.; BOCUTI, E.D.; VELOSO, G.V.; POPPIEL, R.R.; FRANCELINO, M.R.; FERNANDES-FILHO, E.I. Characterizing and Modeling Tropical Sandy Soils through VisNIR-SWIR, MIR Spectroscopy, and X-ray Fluorescence. **Remote Sensing**, v. 14, n. 19, p. 4823, 2022. DOI: <https://doi.org/10.3390/rs14194823>.

DOTTO, A. C.; SIMÃO, R.; CATEN, A. T; Gris, D. J.; FERNANDO, L. AlradSpectra: a Quantification Tool for Soil Properties Using Spectroscopic Data in R. **Revista Brasileira de Ciência do Solo (Impresso)**, v. 43, 2019. DOI: <https://doi.org/10.1590/18069657rbcscs20180263>.

EMBRAPA. **Manual de métodos de análise de solo**. - Portal Embrapa. 2017. Disponível

em: <https://www.embrapa.br/busca-de-publicacoes/-/publicacao/1085209/manual-de-metodos-de-analise-de-solo>. Acesso em: 10 nov. 2024.

FARIA, Svante; SJÖSTRÖM, Michael; ERIKSSON, Lennart. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, 2001.

FINK, J. R.; INDA, A. V.; TIECHER, T.; & BARRÓN, V. Iron oxides and organic matter on soil phosphorus availability. **Ciência e Agrotecnologia**, v. 40, n. 4, p. 369–379, jul. 2016. DOI: <https://doi.org/10.1590/1413-70542016404023016>.

FISHBURN, P.C. Additive Utilities with Incomplete Product Sets: Application to Priorities and Assignments. **Operations Research**, v. 15, n.3, p. 537-542, 1967. DOI: <https://doi.org/10.1287/opre.15.3.537>.

FRIEDMAN, J.; HASTIE, T.; TIBSHIRANI, R.; NARASIMHAN, B.; TAY, K.; SIMON, N.; QIAN, J.; YANG, J. **glmnet: Lasso and Elastic-Net Regularized Generalized Linear Models**. R package version 4.1-9, 2025. Disponível em: <https://cran.r-project.org/web/packages/glmnet/>.

HAGHI, R. K.; PÉREZ-FERNÁNDEZ, E.; ROBERTSON, A. H. J. Prediction of various soil properties for a national spatial dataset of Scottish soils based on four different chemometric approaches: A comparison of near infrared and mid-infrared spectroscopy. **Geoderma**, v. 396, p. 115071, 2021. DOI: <https://doi.org/10.1016/j.geoderma.2021.115071>.

HAIR, J.F; BLACK, W.C.; BABIN, B.J.; ANDERSON, R.E.; TATHAM, R.L. **Análise multivariada de dados**. 6ed. [S.L.]: Bookman, 2009.

HOLFORD, I.C.R.; MATTINGLY, G.E.G. Phosphate adsorption and availability plant of phosphate. **Plant Soil**. v. 44, p. 377–389, 1976. DOI: <https://doi.org/10.1007/BF00015889>.

INSKEEP, W. P.; SILVERTOOTH, J. C. Inhibition of hydroxyapatite precipitation in the presence of fulvic, humic, and tannic acids. **Soil Science Society of America Journal**, v. 52, p. 941–946, 1988. DOI: <https://doi.org/10.2136/sssaj1988.03615995005200040007x>.

INSTITUTO DE PESQUISA E ESTRATÉGIA ECONÔMICA DO CEARÁ. **Perfil básico municipal de Marco 2017**. Fortaleza, CE: Governo do Estado do Ceará, 2017.

JALALI, M.; JALALI, M. Relation between various soil phosphorus extraction methods and sorption parameters in calcareous soils with different texture. **Science of The Total Environment**, v. 566-567, p. 1080–1093, 2016. DOI: <https://doi.org/10.1016/j.scitotenv.2016.05.133>.

JOHNSON, J.; VANDAMME, E.; SENTHILKUMAR, K.; SILA, A.; SHEPHERD, K.D.; SAITO, K. Near-infrared, mid-infrared or combined diffuse reflectance spectroscopy for assessing soil fertility in rice fields in sub-Saharan Africa. **Geoderma**, v. 354, p. 113840–113840, 2019. DOI: <https://doi.org/10.1016/j.geoderma.2019.06.043>.

KATUWAL, S.; KNADEL, M.; NORGAARD, T.; MOLDRUP, P.; GREVE, M.H.; JONGE, L.W. Predicting the dry bulk density of soils across Denmark: Comparison of single-

parameter, multi-parameter, and vis–NIR based models. **Geoderma**, v. 361, p. 114080–114080, 2020. DOI: <https://doi.org/10.1016/j.geoderma.2019.114080>.

KHODADADI, M.; MEUSBURGER, K.; MIRZAEI, M.; STRAUSS, P. BLAKE, W. H.; MOGHASEH, E.; ALEWELL, C. Spatial cross-correlation of surface soil physicochemical properties with soil erosion estimated by fallout radionuclides in croplands in a semi-humid region of Iran. **Catena**, v. 237, p. 107836, 2024. DOI: <https://doi.org/10.1016/j.catena.2024.107836>.

KUTNER, M.H.; NACHTSHEIM, C.J.; NETER, J. LI, W. *Applied Linear Statistical Models*. McGraw-Hill, Boston, 5th Ed., p. 1369, 2005. ISBN: 0-07-238688-6.

LILAND, K.H.; MEVIK, B.-H.; WEHRENS, R.; et al. **pls**: Partial Least Squares and Principal Component Regression. R package version 2.8-5, 2025. Disponível em: <https://cran.r-project.org/web/packages/pls/>.

LORENZZETTI, João A. **Princípios Físicos de Sensoriamento Remoto**. [S.L.]: Editora Blucher, 2015.

MELLO, D.C.; SOUZA, A.B.; MELLO, F. A.O.; MARQUES, K.P.P.; POPPIEL, R.R.; BELINASSO, H.; DI RAIMO, L.D.L.; FRANCELINO, M.R.; FERNANDES-FILHO, E.I.; VELOSO, G.V.; SCHAEFER, C.E.G.R.; DEMATTÊ, J.A.M. Sensor-based field methods for pedology and soil surveys: Protocol suggestions for Brazilian tropical soils. **Geoderma Regional**, v. 33, p. 651, 2023. DOI: <https://doi.org/10.1016/j.geodrs.2023.e00651>.

MELO, V. F.; ALLEONI, L. R. F. **Química e mineralogia do solo**: conceitos básicos e aplicações. Sociedade Brasileira de Ciência do Solo, Viçosa, MG: ed. 1, p. 1364, 2019. ISBN: 978-85-86504-26-6.

MOURA-BUENO, J. M.; DALMOLIN, R. S. D.; HORST-HEINEN, T. Z.; CATEN, A.T.; VASQUES, G.M.; DOTTO, A.C.; GRUNWALD, S. When does stratification of a subtropical soil spectral library improve predictions of soil organic carbon content?. **Science of The Total Environment**, v. 737, p. 139895, 2020. DOI: <https://doi.org/10.1016/j.scitotenv.2020.139895>.

MUKHERJEE, S.; LASKAR, S. Vis–NIR-based optical sensor system for estimation of primary nutrients in soil. **Journal of Optics**, v. 48, n. 1, p. 87–103, 2019. DOI: <https://doi.org/10.1007/s12596-019-00517-1>.

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, n. 23, p. 2819, 2019. DOI: <https://doi.org/10.3390/rs11232819>.

NADEEM, F.; ABBAS, S.; WASEEM, F.; ALI, N.; MAHMOOD, R.; BIBI, S.; DENG, L.; WANG, R.; ZHONG, Y.; LI, X. Phosphorus (P) and Zinc (Zn) nutrition constraints: A perspective of linking soil application with plant regulations. **Environmental and Experimental Botany**, v. 226, p. 105875–105875, 2024. DOI: <https://doi.org/10.1016/j.envexpbot.2024.105875>.

NANNI, M. R.; DEMATTÊ, J. A. M. Spectral Reflectance Methodology in Comparison to

Traditional Soil Analysis. **Soil Science Society of America Journal**, v. 70, n. 2, p. 393–407, 2006. DOI: <https://doi.org/10.2136/sssaj2003.0285>.

NOCITA, M.; STEVENS, A.; WESEMAEL, B. V.; AITKENHEAD, M.; BACHMANN, M.; BARTHÈS, B.; BEN-DOR, E.; BROWN, D.; CLAIROTTE, M.; CSORBA, A.; DARDENNE, P.; DEMATTÊ, J.; GENOT, V.; GUERRERO, C.; KNADEL, M.; MONTANARELLA, L.; NOON, C.; RAMIREZ-LOPEZ, L.; ROBERTSON, J.; SAKAI, H.; WETTERLIND, J. Soil Spectroscopy: An Alternative to Wet Chemistry for Soil Monitoring. **Advances in Agronomy**, p. 139–159, 2015.

NOVAIS, R. F.; ALVAREZ, V.H.; BARROS, N.F.; FONTES, R.L.; CANTARUTTI, R.B.; NEVES, J.C.L (ed.). Fertilidade do solo. Viçosa: Sociedade brasileira de ciência do solo, 2007. p.472-537.

NOVO, E.M.L. **Sensoriamento Remoto: principios e aplicações**. [S.L.]: Editora Blucher, 2010.

OOMS, J. **writexl**: Export Data Frames to Excel 'xlsx' Format. R package version 1.5.4. Disponível em: <https://cran.r-project.org/web/packages/writexl/>.

PAVÃO, Q.S; RIBEIRO, P. G.; MACIEL, G. P.; SILVA.S.H.G.; ARAÚJO, S.R.; FERNANDES, A.R.; DEMATTÊ, J.A.M.; FILHO, P.W.M.S; RAMOS, S.J. Texture prediction of natural soils in the Brazilian Amazon through proximal sensors. **Geoderma Regional**, v. 37, p. e00813–e00813, 2024. DOI: <https://doi.org/10.1016/j.geodrs.2024.e00813>.

PAVLŮ, L.; ZÁDOROVÁ, T.; PAVLŮ, J.; TEJNECKÝ, V. DRÁBEK, O.; ROJAS, J.R.; THAI, S.; PENÍZEK, V. Prediction of the distribution of soil properties in deep Colluvisols in different pedogeographic regions (Czech Republic) using diffuse reflectance infrared spectroscopy. **Soil and Tillage Research**, v. 234, p. 105844–105844, 2023. DOI: <https://doi.org/10.1016/j.still.2023.105844>.

POTDAR, R. P.; SHIROLKAR, M. M.; VERMA, A. J.; MORE, P. S.; & KULKARNI, A. Determination of soil nutrients (NPK) using optical methods: a mini review. **Journal of Plant Nutrition**, v. 44, n. 12, p. 1826–1839, 2021. DOI: <https://doi.org/10.1080/01904167.2021.1884702>.

R Core Team. **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. 2024. Acesso em: <https://www.R-project.org>.

RABIE, R.; SAFFAJ, T.; ITQIQ, S. E.; BOUZIDA, I.; SAIDI, O.; YAAKOUBI, K.; LAKSSIR, B.; MERNISSI, N. E.; EL, M. Predicting soil phosphorus and studying the effect of texture on the prediction accuracy using machine learning combined with near-infrared spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 242, p. 118736–118736, 2020. DOI: <https://doi.org/10.1016/j.saa.2020.118736>.

REYES, J.; LIESS, M. Spectral Data Processing for Field-Scale Soil Organic Carbon Monitoring. **Sensors (Basel)**, v. 24, n. 3, p. 849–849, 2024. DOI: <https://doi.org/10.3390/s24030849>.

RIBEIRO, S. G.; TEIXEIRA, A. S.; DE OLIVEIRA, M. R. R.; COSTA, M.C.G.; ARAÚJO,

I.C.S.; MOREIRA, L.C.J.; LOPES, F.B. Soil Organic Carbon Content Prediction Using Soil-Reflected Spectra: A Comparison of Two Regression Methods. **Remote Sensing**, v. 13, n. 23, p. 4752, 2021. DOI: <https://doi.org/10.3390/rs13234752>.

RICHTER, N.; JARMER, T.; CHABRILLAT, S.; OYONART, C.; HOSTERT, P.; KAUFMANN, H. Free iron oxide determination in Mediterranean soils using diffuse reflectance spectroscopy. **Soil Science Society of American Journal**. v. 73, p. 72-81, 2009. DOI: <http://dx.doi.org/10.2136/sssaj2008.0025>.

RIPLEY, B.; VENABLES, W. **MASS**: Support Functions and Datasets for Venables and Ripley's MASS. R package version 7.3-65, 2025. Disponível em: <https://cran.r-project.org/web/packages/MASS/>.

RIZZO, R. **Técnicas espectroscópicas e atributos do terreno no mapeamento digital de solos**. 2011. Dissertação (Mestrado em Solos e Nutrição de Plantas) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2011. DOI:10.11606/D.11.2011.tde-02052011-091133.

RODRÍGUEZ-PÉREZ, José Ramón; MARCELO, Víctor; OBAYA, Dimas Pereira; GARCIA-FERNÁNDEZ, Marta; ABLANEDO, Enoc Sanz. Estimating Soil Properties and Nutrients by Visible and Infrared Diffuse Reflectance Spectroscopy to Characterize Vineyards. **Agronomy**, v. 11, n. 10, p. 1895, 2021. DOI: <https://doi.org/10.3390/agronomy11101895>.

RUMA, F. Y.; MUNNAF, M. A.; DE NEVE, S.; MOUAZEN, A.M. Visible and near infrared spectroscopy for predicting soil nitrogen mineralization rate: Effect of incubation period and ancillary soil properties. **Catena**, v. 235, p. 107649, 2024. DOI: <https://doi.org/10.1016/j.catena.2023.107649>.

RUSSEL, Alfred. **Introduction to soil chemistry: analysis and instrumentation**. Hoboken: J. Wiley, 2014.

SANTOS, D. R. DOS.; GATIBONI, L. C.; & KAMINSKI, J. Fatores que afetam a disponibilidade do fósforo e o manejo da adubação fosfatada em solos sob sistema plantio direto. **Ciência Rural**, v.38, n.2, p. 576–586, 2008. DOI: <https://doi.org/10.1590/S0103-84782008000200049>.

SHI, Pu; CASTALDI, Fabio; WESEMAEL, B. V.; OOST, K.V. Vis-NIR spectroscopic assessment of soil aggregate stability and aggregate size distribution in the Belgian Loam Belt. **Geoderma**, v. 357, p. 113958, 2020. DOI: <https://doi.org/10.1016/j.geoderma.2019.113958>.

SHI, X. W.; YAO, L. J.; PAN, T. Visible and Near-Infrared Spectroscopy with Multi-Parameters Optimization of Savitzky-Golay Smoothing Applied to Rapid Analysis of Soil Cr Content of Pearl River Delta. **Journal of Geoscience and Environment Protection**, v. 9, p.75-83, 2021. DOI: <https://doi.org/10.4236/gep.2021.93006>.

SOUZA, M. F. DE.; FRANCO, H. C. J.; AMARAL, L. R. DO. Estimation of soil phosphorus availability via visible and near-infrared spectroscopy. **Scientia Agricola**, v. 77, n. 5, p. e20180295, 2020. DOI: <https://doi.org/10.1590/1678-992X-2018-0295>.

SPOSITO, G. **The Chemistry of Soils**. New York, Ny: Oxford University Press, 2016.

STEVENS, A.; RAMIREZ-LOPEZ, L. **An introduction to the prospectr package**. R Package Vignette, Report No.: R Package Version 0.1, v. 3, 2014.

TAN, K. H., **Principles of Soil Chemistry, Fourth Edition**, [S.L.]: CRC Press, 2011.

TERNIKAR, C. R.; GOMEZ, C.; KUMAR, D. N. Visible and infrared lab spectroscopy for soil texture classification: Analysis of entire spectra v/s reduced spectra. **Remote Sensing Applications Society and Environment**, v. 35, p. 101242–101242, 2024. DOI: <https://doi.org/10.1016/j.rsase.2024.101242>.

TIESSEN, H.; STEWART, J.W.B.; COLE, C.V. Pathways of phosphorus transformations in soils of differing pedogenesis. **Soil Science Society of America. Journal**, Madison, v.48, n.2, p.853-858, 1984. DOI: <https://doi.org/10.2136/sssaj1984.03615995004800040031x>.

VAN RAIJ, B.; CANTARELLA, H.; QUAGGIO, J.A.; FURLANI, A.M.C. **Recomendações de Adubação e Calagem para o Estado de São Paulo**. 2nd ed. Instituto Agronômico & Fundação IAC, Campinas, 1996.

VISCARRA, R.R.A.; WALVOORT, D.J.J.; MCBRATNEY, A.B.; JANIK, L.J.; SKJEMSTAD, J.O. Visible, near infrared, mid infrared or combined diffuse reflectance spectroscopy for simultaneous assessment of various soil properties. **Geoderma**, v. 131, n. 1-2, p. 59–75, 2006. DOI: <https://doi.org/10.1016/j.geoderma.2005.03.007>.

WALVOORT, D.J.J.; MCBRATNEY, A.B. **Diffuse reflectance spectrometry as a proximal sensing tool for precision agriculture**. Montpellier: Agro Montpellier, 2001.

WEHRENS, Ron. **Chemometrics with R: Multivariate Data Analysis in the Natural and Life Sciences**. [s.l.]: Springer Science & Business Media, 2020.

WEIL, R.R.; N.C. BRADY. **Phosphorous and Potassium**, Chapter 14, p. 6433-695. The Nature and Properties of Soils, 15th ed. Pearson, Columbus, OH, USA, 2017.

WETTERLIND, J.; STENBERG, B.; VISCARRA, R. R. A. Soil Analysis Using Visible and Near Infrared Spectroscopy. **Plant Mineral Nutrients**, p. 95–107, 2012. DOI: [https://doi.org/10.1007/978-1-62703-152-3\\_6](https://doi.org/10.1007/978-1-62703-152-3_6).

WICKHAM, H.; FRANÇOIS, R.; HENRY, L.; Müller, K.; Vaughan, D. **dplyr: A Grammar of Data Manipulation**. R package version 1.1.4, 2023. Disponível em: <https://cran.r-project.org/web/packages/dplyr/>.

WICKHAM, H.; BRYAN, J. **readxl: Read Excel Files**. R package version 1.4.5, 2025. Disponível em: <https://cran.r-project.org/web/packages/readxl/>.

WICKHAM, H.; CHANG, W.; HENRY, L.; et al. **ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics**. R package version 3.5.2, 2025. Disponível em: <https://cran.r-project.org/web/packages/ggplot2/>.

ZHOU, P.; ZHANG, Y.; YANG, W.; LI, M.; LIU, Z.; LIU, X. Development and performance test of an in-situ soil total nitrogen-soil moisture detector based on near-infrared spectroscopy. **Computers and Electronics in Agriculture**, v. 160, p. 51–58, 2019. DOI: <https://doi.org/10.1016/j.compag.2019.03.016>.

ZHOU, Nan; HONG, Jin; SONG, Bo; WU, Shichao; WEI, Yichen; WANG, Tao. Feature Variable Selection Based on VIS-NIR Spectra and Soil Moisture Content Prediction Model Construction. **Journal of Spectroscopy**, v. 2024, p. 1–16, 2024. DOI: <https://doi.org/10.1155/2024/8180765>.

**APÊNDICE A – PARÂMETROS DE REGRESSÃO UTILIZADAS NOS MODELOS PREDITIVOS DE FÓSFORO PARA O NEOSSOLO QUARTZARÊNICO.**

I. Modelos construídos a partir do conjunto de dados do solo RQo e reflectância bruta.

Variável	MLR	PCR	PLSR
	Coeficientes		
<b>(Intercept)</b>	264,27	447,79	-
<b>X699</b>	-47655,97	-35839,22	-9652,62
<b>X706</b>	-70934,64	6651,38	-6475,68
<b>X707</b>	151922,60	10617,75	4744,72
<b>X708</b>	-119657,56	14962,08	-2701,94
<b>X722</b>	92932,20	9566,48	21058,28
<b>X1286</b>	-23949,14	-20922,66	-16871,25
<b>X1409</b>	34446,21	-211,74	-1708,54
<b>X1410</b>	-54810,50	5960,50	-2287,25
<b>X1411</b>	30941,96	4652,31	13972,77
<b>X1896</b>	33725,25	18968,99	24002,92
<b>X1909</b>	32598,43	39284,78	15494,30
<b>X1913</b>	-58792,92	-47076,46	-33798,48
<b>X2378</b>	-25805,70	-11631,26	-13439,04
<b>X2419</b>	-23820,63	-22857,10	-2275,77
<b>X2420</b>	51766,02	32119,57	27999,69
<b>X2421</b>	-80012,80	7864,89	-7519,30
<b>X2422</b>	71478,80	-9604,86	-1706,36
<b>X2433</b>	11090,57	13434,88	9628,54
<b>X2438</b>	-16109,25	-21291,21	-17882,95
<b>X2439</b>	27703,97	9125,76	6120,55
<b>X2440</b>	-17893,74	-3313,93	-5317,07
<b>R2</b>	-0,11	0,69	0,60
<b>RMSE</b>	61,41	32,74	37,25

II. Modelos construídos a partir do conjunto de dados do solo RQo e reflectância normalizado por comprimento de vetor.

Variável	MLR	PCR	PLSR
	Coeficientes		
<b>(Intercept)</b>	23326,93	30077,10	-
<b>X714</b>	-483815,87	-11637,85	134407,34
<b>X715</b>	1163604,87	-50602,08	663878,96
<b>X721</b>	-476444,22	4070,54	-506853,23
<b>X737</b>	-216250,25	40851,42	-243541,31
<b>X875</b>	-629736,30	-80618,10	-22798,29
<b>X884</b>	429788,05	-184015,44	3351,27
<b>X1417</b>	-245192,69	-412733,95	181165,74
<b>X1763</b>	276947,39	80689,03	619519,19
<b>X1766</b>	-336309,45	-165383,52	-470851,49
<b>X2124</b>	-567896,23	-598967,68	-591046,45
<b>X2151</b>	454819,95	54747,76	738718,11
<b>X2156</b>	-642559,38	-72612,36	-501049,22
<b>X2161</b>	-401301,48	-13832,24	-64779,56
<b>X2162</b>	845336,52	1678,51	338654,37
<b>X2163</b>	-712488,66	-60316,66	-479245,35
<b>X2165</b>	567090,32	188379,85	262414,74
<b>R2</b>	0,75	0,89	0,78
<b>RMSE</b>	29,18	19,58	32,27

III. Modelos construídos a partir do conjunto de dados do solo RQo e reflectância aplicada a primeira derivada de Savitzky-Golay.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	16447070	-64,93	-
X699	-127512800000	-6896,14	-110469,12
X706	-325145800000	45603,49	54306,71
X707	918780200000	33944,44	-12462,68
X708	-723066400000	36957,93	-141225,47
X722	204419800000	12574,24	115489,68
X1286	26316040000	46524,25	-24473,13
X1409	-1085470000000	36943,31	-165387,54
X1410	1640590000000	-1397,04	184068,97
X1411	-510533200000	-14524,58	6377,46
X1896	325120000000	178829,00	194528,77
X1909	-115557000000	-242728,65	20705,74
X1913	-509416900000	-245729,76	-593689,11
X2378	-170014600000	22314,81	-69874,15
X2419	339981700000	140350,12	92374,70
X2420	-638754800000	47504,38	-35231,18
X2421	515307800000	-12553,92	125122,93
X2422	-136455000000	-71915,79	-114351,96
X2433	-64322840000	-4549,99	-40541,77
X2438	-147829400000	-117358,14	-174885,37
X2439	-46411210000	-25177,06	76711,42
X2440	-8259202000	62328,63	-14302,80
<b>R2</b>	0,53	0,71	0,66
<b>RMSE</b>	39,82	31,29	36,04

IV. Modelos construídos a partir do conjunto de dados do solo RQo e reflectância aplicada a Remoção do contínuo.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	-56591,61	-51402,39	-
<b>X1042</b>	-8031,32	-9984,20	-17254,41
<b>X1136</b>	20065,11	25675,46	785,57
<b>X1188</b>	10719,42	5851,14	3251,34
<b>X1433</b>	24169,42	14894,04	28379,35
<b>X1437</b>	-18471,57	-7419,19	-11515,52
<b>X1438</b>	-8200,38	-12042,66	-14964,77
<b>X1521</b>	-7729,63	603,27	9654,13
<b>X1526</b>	-9312,64	6602,66	-24418,04
<b>X1527</b>	10605,26	-5308,69	-4845,03
<b>X1587</b>	12153,24	7168,28	12068,21
<b>X1685</b>	-27947,02	-4846,43	-29353,69
<b>X1686</b>	25527,99	-1139,37	13813,93
<b>X1835</b>	7306,70	6852,44	3082,14
<b>X2062</b>	20723,42	18445,29	26358,94
<b>X2225</b>	5477,25	6501,74	5458,53
<b>R2</b>	0,76	0,81	0,75
<b>RMSE</b>	28,76	25,84	30,77

**APÊNDICE B – PARÂMETROS DE REGRESSÃO UTILIZADAS NOS MODELOS PREDITIVOS DE FÓSFORO PARA O ARGISSOLO ACINZENTADO.**

I. Modelos construídos a partir do conjunto de dados do solo PACd e reflectância bruta

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	-197,098	-61,7518	-
X1227	35253,34	16136,71	38477,73
X1230	-27967,3	-8368,83	-30833
X1848	-63502,2	-63345,9	-61586,6
X1850	56276,71	67649,54	56948,65
X1851	-68566,9	-86034,3	-69235,1
X1853	-72230,4	-23602,4	-63104,3
X1854	54485,45	14100,56	46989,09
X1859	39482,58	29610,4	34488,77
X1861	28319,51	34042,29	23191,71
X1864	46454,69	42613,98	48916,54
X1870	22803,01	28462,3	30261,57
X1872	-49319,9	-49871,6	-53226
R2	0,28	0,17	0,54
RMSE	48,88	52,37	48,45

II. Modelos construídos a partir do conjunto de dados do solo PACd e reflectância aplicada a primeira derivada de Savitzky-Golay.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	-172,399	79,42317	-
X886	-277993	37949,96	140771,5
X889	248169,5	63720,19	157949,9
X968	-336025	-55169,3	-288286
X1074	-125394	-120044	-127372
X1150	387361,4	54112,37	133121,6
X1154	-675312	43801,44	154070,2
X1285	-559673	-100153	-249050
X1296	463034,9	78857,74	134326,2
X1382	-422546	-108651	-27358,7
X1685	-364752	-82833,9	-121341
X1743	-197238	-37534,2	-123712
X1777	-83823,5	-122229	-162441
X1815	280867	-3000,85	-32907,8
X1819	-1333030	-30058,1	-246171
X1820	1770391	10404,31	390031,9
X1827	302214,3	-166597	-43100,4
X1926	130499,3	-183757	78423,89
X2085	305470,6	5307,053	237751
X2086	-262260	13189,17	-137826
X2097	305487,2	-75429,9	87801,03
X2098	-241625	-117753	-257699
X2280	525086,3	-11516,3	-79769,7
X2281	-375370	104325,4	169605,7
X2353	-179682	-68053,6	-122061
X2357	120752,7	47246,22	-68558,5
X2383	125280,3	92094,29	85418,91
R2	-0,12	0,71	0,74
RMSE	60,95	31,05	29,11

III. Modelos construídos a partir do conjunto de dados do solo PACd e reflectância normalizado por comprimento de vetor.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	17620,46	18070,83	-
X945	1169165	228038,5	1549781
X947	-581393	-567295	-434968
X948	-816635	112408,5	-970995
X2119	-880769	-144859	-663058
X2121	313215,8	-471653	487703,8
X2153	832743	956703,7	365476,1
X2155	-1189242	-949753	-551612
X2156	383986,7	22734,14	67086,87
X2251	301555,7	102137,6	800663,8
X2259	-261841	-35111,7	-595370
R2	0,453139	0,603	0,357862
RMSE	42,55573	26,24	52,52797

IV. Modelos construídos a partir do conjunto de dados do solo PACd e reflectância aplicada a Remoção do contínuo.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	25451,37	36101,53	-
X1147	15966,81	-3217,1	12203,2
X1148	-9167,96	-5056,8	-7134,9
X1149	-15976,6	-7892,27	-18793,2
X1150	-16834,7	6907,433	-12550,5
X1187	5615,138	5551,263	6479,769
X1301	19724,87	11530,53	16817,8
X1661	8742,976	-370,661	8069,375
X1667	23593,12	18797,3	31441,27
X1687	8374,082	-16150,9	9179,35
X1703	-15362,5	-2908,43	-9508,11
X1748	-14299,8	-15376,3	-15693,4
X1801	-25153,7	-9997,05	-25220,3
X1807	24591,12	2572,079	23283,45
X2102	-16696,9	5190,413	-4982,17
X2103	20764,11	2232,823	11331,2
X2115	32154,21	8652,95	26160,52
X2119	-15749,2	-22986,3	-16155,2
X2120	-29903,8	-18635,3	-25910,5
X2126	-12895	-5365,4	-7344,1
X2135	-8363,05	-223,348	-761,394
X2235	12086,23	9404,211	12430,92
X2263	-9943,12	-7389,93	-10903,3
X2275	-6494,69	8745,76	-2174,16
R2	0,422313	0,70	0,549444
RMSE	43,73868	31,47	41,28045

**APÊNDICE C – PARÂMETROS DE REGRESSÃO UTILIZADAS NOS MODELOS PREDITIVOS DE FÓSFORO PARA O NEOSSOLO QUARTZARÊNICO E ARGISSOLO QUARTZARÊNICO.**

I. Modelos construídos a partir do conjunto de dados do solo RQo + PACd e reflectância bruta.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	-106,77	-106,77	-
X714	12932,23	12932,23	12388,24
X904	-16478,5	-16478,5	-15266,1
X1290	-38262,2	-38262,2	-39785,4
X1293	-18491,6	-18491,6	-24356,4
X1297	-25807,3	-25807,3	-27690,8
X1298	40691,74	40691,74	36578,19
X1300	53997,04	53997,04	65920,42
X1870	29066,66	29066,66	28497,27
X1873	-55730,4	-55730,4	-36534,8
X1874	63304,05	63304,05	39394,53
X1876	-54302,9	-54302,9	-47860,3
X1894	29209,1	29209,1	28504,59
X1899	-16914,6	-16914,6	-16142,7
X2417	-12916,6	-12916,6	-17074,5
X2419	26888,77	26888,77	26750,93
X2422	-13752,9	-13752,9	-12408,5
X2423	10491,76	10491,76	10725,94
X2425	45908,47	45908,47	38682,53
X2426	-42058,6	-42058,6	-38578,5
X2433	8157,697	8157,697	7762,012
X2434	-17369,4	-17369,4	-14256,3
X2438	-6473,11	-6473,11	-3273,73
R2	0,1265	0,4305	0,544819
RMSE	54,4	43,93	41,17139

II. Modelos construídos a partir do conjunto de dados do solo RQo + PACd e reflectância normalizado por comprimento de vetor.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	10468,45	10468,45	-
X1192	-488266	-488266	-931333
X1211	-353477	-353477	-240074
X1212	385283,5	385283,5	642627,2
X1214	-390095	-390095	435400,8
X1242	327244,1	327244,1	367571,2
X1243	512235,5	512235,5	17758,17
X1252	-575663	-575663	-952905
X1253	-923038	-923038	-987624
X1256	272407	272407	366018,5
X1259	569758,4	569758,4	825580,3
X1261	559811,2	559811,2	597317,9
X2136	-605870	-605870	-291438
X2146	523759,6	523759,6	375011,8
X2154	-443661	-443661	-721307
X2159	290504,2	290504,2	423473,8
X2162	-585123	-585123	-716416
X2165	506020,8	506020,8	808279,1
R2	0,83503	0,83503	0,669207
RMSE	23,64606	23,64606	34,27784

III. Modelos construídos a partir do conjunto de dados do solo RQo + PACd e reflectância aplicada a primeira derivada de Savitzky-Golay.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	84,98154	84,98154	-
X1619	-258047	-258047	-352254
X1752	145871,8	145871,8	179234,8
X1761	266988,7	266988,7	183813,8
X1762	-321924	-321924	-226083
X1820	242956,1	242956,1	224389,3
X1821	-111091	-111091	-49963,8
X2011	-333212	-333212	-324920
X2013	289419	289419	303693,3
X2028	247247,4	247247,4	316289,7
X2055	194783,4	194783,4	190255
X2080	-147699	-147699	-95804,6
X2087	-107342	-107342	-76063,1
X2099	-519164	-519164	-427374
X2100	223915,8	223915,8	73978,47
X2266	-380996	-380996	-308216
X2267	367021,4	367021,4	274926,9
X2333	60749,1	60749,1	138400,1
X2348	81042,69	81042,69	41800,21
X2355	-101619	-101619	-90309,3
R2	0,692945	0,692945	0,73452
RMSE	32,26004	32,26004	32,018

IV. Modelos construídos a partir do conjunto de dados do solo RQo + PACd e reflectância aplicada a Remoção do contínuo.

Variável	MLR	PCR	PLSR
	Coeficientes		
(Intercept)	-13606,4	-13606,4	-
X1088	7721,007	7721,007	8050,81
X1300	-8457,84	-8457,84	-6683,21
X1338	-11953,2	-11953,2	-11805,7
X1342	7713,715	7713,715	6909,212
X1615	19269,94	19269,94	16628,18
X1626	-15671	-15671	-18848,2
X1668	23644,89	23644,89	22003,83
X1747	-23272,4	-23272,4	-27231,9
X1822	11097,39	11097,39	8998,865
X1823	7878,677	7878,677	6089,94
X1824	-6738,44	-6738,44	-5492
X1831	12683,54	12683,54	3991,045
X1832	-18856,9	-18856,9	-10735,5
X1838	16505,94	16505,94	15069,15
X1841	-7389,23	-7389,23	-5969,1
X1866	-12759,2	-12759,2	-9552,04
X1869	16960,11	16960,11	14798,87
X1872	-9860,99	-9860,99	-9397,45
X2078	5092,936	5092,936	4439,136
X2226	3198,466	3198,466	3733,519
X2348	6377,851	6377,851	6587,403
X2349	-7904,65	-7904,65	-8612,36
X2353	4370,956	4370,956	4322,285
X2415	-5810,21	-5810,21	-7087,4
R2	0,618516	0,618516	0,652393
RMSE	35,95796	35,95796	35,4001