

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DO SOLO E
QUALIDADE DE ECOSISTEMAS**

**USO DA ESPECTROSCOPIA Vis-NIR NA AVALIAÇÃO
DA FERTILIDADE DOS SOLOS DO ESTADO DA BAHIA**

FÁBIO OLIVEIRA DA SILVA

CRUZ DAS ALMAS - BAHIA

DEZEMBRO - 2020

**UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA
CENTRO DE CIÊNCIAS AGRÁRIAS, AMBIENTAIS E BIOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DO SOLO E
QUALIDADE DE ECOSISTEMAS**

**USO DA ESPECTROSCOPIA Vis-NIR NA AVALIAÇÃO DA
FERTILIDADE DOS SOLOS DO ESTADO DA BAHIA**

FÁBIO OLIVEIRA DA SILVA

Dissertação apresentada ao programa de pós-graduação em Solos e Qualidade de Ecossistemas, da UNIVERSIDADE FEDERAL DO RECÔNCAVO DA BAHIA, como requisito parcial para obtenção do título de Mestre em Ciências do Solo.

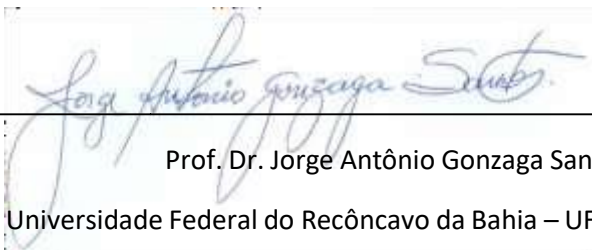
ORIENTADOR: PROF. Dr. JORGE ANTÔNIO GONZAGA SANTOS

CRUZ DAS ALMAS - BAHIA

DEZEMBRO - 2020

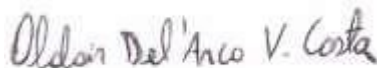
Universidade Federal Do Recôncavo Da Bahia Centro De Ciências Agrárias, Ambientais E Biológicas
Programa De Pós-Graduação Em Ciências Do Solo E Qualidade De Ecossistemas

Comissão examinadora da Defesa de Dissertação de Fábio Oliveira da Silva




Prof. Dr. Jorge Antônio Gonzaga Santos

Universidade Federal do Recôncavo da Bahia – UFRB



Prof. Dr. Oldair Del'Arco Vinhas Costa

Universidade Federal do Recôncavo da Bahia – UFRB



Prof^ª. Dr^ª. Deorgia Tayane Mendes Souza

Universidade Estadual de Feira de Santana - UEFS

Dissertação homologada pelo Colegiado do Curso de Mestrado em Solos e Qualidade de Ecossistemas em _____

Conferindo o grau de Mestre em Ciências do Solo _____.

SUMÁRIO

RESUMO

ABSTRACT

1. INTRODUÇÃO	7
2. REVISÃO BIBLIOGRÁFICA	11
2.1. Variabilidade espacial dos atributos do solo	11
2.2 Espectroscopia Vis-NIR	12
2.3 Breve histórico da espectroscopia	13
2.4 Fundamentos da espectroscopia	14
2.5 Quimiometria	16
2.6 Calibração multivariadas	18
2.7 Regressão por mínimos quadrados parciais	20
2.8 Validação do modelo	21
3. MATERIAIS E MÉTODOS	24
3.1 Análises Químicas	27
3.2 Aquisição Espectral	28
3.3 Análise de Dados	39
4. RESULTADOS E DISCUSSÃO	36
4.1 Características da Distribuição das amostras	36
4.2 Pré-processamentos	38
4.3 Capacidade preditiva dos modelos de regressão	41
CONCLUSÃO	49
REFERÊNCIAS BIBLIOGRÁFICAS.....	50

USO DA ESPECTROSCOPIA Vis-NIR NA AVALIAÇÃO DA FERTILIDADE DOS SOLOS DO ESTADO DA BAHIA

Autor: Fábio Oliveira da Silva

Orientador: Prof. Dr. Jorge Antônio Gonzaga Santos

RESUMO: Espectroscopia de infravermelho próximo e visível (Vis-NIR) é uma técnica econômica alternativa para análises químicas na previsão das propriedades do solo. Para consolidá-la, são necessários estudos que determinem o algoritmo de combinação de banda ideal para um método eficaz de extrair variáveis espectrais, considerando as informações de interação entre as concentrações determinadas a partir das análises convencionais e os dados espectrais vis-NIR. Neste estudo foi experimentado diferentes técnicas de pré-processamento como, SNV, Savitzky-Golay, filtro mediano e suas combinações, como o método de calibração multivariada PLS, com o propósito de estimar as propriedades químicas do solo relacionadas aos teores de cálcio, magnésio, carbono orgânico, potássio, fósforo, hidrogênio mais alumínio, alumínio e sódio. Para tanto, foram selecionadas 154 amostras de 32 perfis de solos, totalizando 9 classes de solo distintas. As amostras foram secas, destorroadas e divididas em dois grupos, o primeiro foi submetido às análises químicas laboratoriais convencionais, e o segundo grupo foi usado para obtenção dos espectros Vis-NIR. Os espectros foram pré-processados utilizando as diferentes técnicas, e os dados das diferentes análises foram correlacionados utilizando o PLS. Para validação do método foi utilizado o coeficiente de determinação R^2 e o desvio preditivo residual RPD. Os valores dos coeficientes utilizados para validação demonstraram bons resultados para os métodos de predição dos teores de C.O., Na e Mg, para os elementos Ca, K e H + Al o modelo foi aceitável e para Al e P não foi possível estabelecer um método de calibração adequado.

Palavras chave: Espectroscopia Vis-NIR, Calibração multivariada, Atributos do solo.

USE OF VIS-NIR SPECTROSCOPY IN ASSESSING SOIL FERTILITY IN THE STATE OF BAHIA

Author: Fábio Oliveira da Silva

Adviser: Prof. Dr. Jorge Antônio Gonzaga Santos

ABSTRACT: Near and visible infrared spectroscopy (Vis-NIR) is an alternative economical technique for chemical analysis to predict soil properties. To consolidate it, studies are needed to determine the ideal band combination algorithm for an effective method of extracting spectral variables, considering the interaction information between the wave bands of the vis-NIR spectral data and conventional analyzes. In this study, different pre-processing techniques were tried, such as SNV, Savitzky-Golay, median filter and their combinations, such as the multivariate calibration method PLS, with the purpose of estimating the chemical properties of the soil related to the levels of calcium, magnesium, organic carbon, potassium, phosphorus, hydrogen plus aluminum, aluminum and sodium. For this purpose, 154 samples from 32 profiles were selected, totaling 9 different soil classes, the samples were dried and distilled were divided into two groups, the first was subjected to conventional laboratory chemical analyzes, and the second group was used to obtain the spectra Vis-NIR. The spectra were pre-processed using the different techniques, and the data from the different analyzes were correlated using the PLS, to validate the method, the R² determination coefficient and the residual predictive deviation RPD were used. The values of the coefficients used for validation demonstrated the construction of good methods to predict the levels of CO, Na and Mg, for the elements Ca, K and H + Al the model was acceptable and for Al and P it was not possible to establish a method of proper calibration.

Key words: Vis-NIR spectroscopy, Multivariate calibration, Soil attributes.

1. INTRODUÇÃO

O avanço da produção agrícola do país tem requerido o monitoramento constante das propriedades químicas, físicas e biológicas para uso eficiente do solo em níveis sustentáveis. Segundo Santos (2011) o crescimento da produtividade das lavouras tem sido conseguido graças ao uso racional de insumos, o que torna necessário promover retornos adequados sobre os investimentos, demandando confiáveis análises em laboratórios, que permitam diagnósticos corretos, principalmente a problemas relacionados a fertilidade química e possíveis contaminantes do solo.

No Brasil muitas áreas agrícolas ainda são desconhecidas, apenas com estudos adequados dos solos destas áreas, será possível identificar conhecer suas características colaborando para o entendimento de seu real potencial de uso, permitindo o crescimento de forma sustentável pela utilização de práticas de manejo adequadas, que garantam o controle da produtividade das culturas e dos riscos de contaminação para o ambiente (López, 2009).

É comum encontrar dentro de uma mesma propriedade diferentes classes de solos, mas para que sejam implementados diferentes usos e manejos, é necessário avaliar, de forma específica para cada situação, os atributos químicos do solo. Uma vez que, variabilidade espacial dos níveis destes atributos podem ser causadas por: interações do manejo com os fatores e processos de formação do solo, resultando em diferentes respostas das culturas e impactando sobre as saídas de nutrientes.

Estudos da variabilidade espacial dos atributos do solo tornam possível gerenciar tais diferenças, uma vez que o conhecimento da variabilidade espacial dos atributos do solo permitem controlar a produtividade das culturas e os riscos de contaminação do ambiente. Para o estudo das características do solo relacionadas com o seu potencial de uso e para a avaliação da fertilidade, as análises químicas e físicas de rotina são os métodos utilizados convencionalmente. Estas análises permitem determinar as necessidades de aplicação de fertilizantes e calagem. Além de oferecem suporte para a implementação de irrigação, determinação de classes de solos, definição de zonas específicas de manejo, entre outros.

O nível de precisão destas análises vem aumentando de forma considerável, de tal forma, que hoje é possível avaliar quase todos os elementos presentes na tabela periódica em concentrações de até parte por trilhão. Por meio de técnicas já consolidadas como a espectroscopia de absorção atômica, espectrometria de absorção atômica com forno de grafite e espectrometria de emissão atômica com plasma acoplado indutivamente.

Entretanto, a utilização destes equipamentos, requer a abertura química das amostras utilizando reagentes com alto grau de pureza, por isso são caros, e produzem resíduos tóxicos os quais precisam ser tratados antes de serem descartados. Além disso as metodologias adotadas pela maioria dos laboratórios para análises químicas dos solos, envolvem muitas etapas e procedimentos demorados, demandando assim, tempo e mão de obra. Estes problemas são ainda mais potencializados quando há necessidade de avaliar um número grande de amostras, principalmente quando se emprega os princípios da agricultura de precisão.

Uma alternativa é o uso de aparelhos que não requerem a abertura química dessas amostras, que tenham a mesma eficiência analítica, que sejam menos onerosas e que não gere descarte de resíduos agressivos ao meio ambiente, ou seja, favorecendo o menor impacto ambiental. A utilização da espectrometria de infravermelho próximo, vem sendo empregada pelo mundo na predição de propriedades químicas e físicas ativas nessa região do espectro. Diversos autores como SINGH, *et. al* (2019), Eitelwein (2017), Demattê (2019), Santos (2011), entre outros, apontam que a técnica é de fato muito promissora, tendo sido inclusive já experimentada para análise de solos e contaminantes, entretanto sua confiabilidade estar diretamente relacionada a capacidade representativa da biblioteca espectral, a qual servirá como matriz para correlação dos resultados a partir de métodos estatísticos que possibilitem a predição das propriedades desejadas.

A utilização da espectrometria de infravermelho próximo tem despertado bastante o interesse dos estudiosos de solos, por possuir uma série de vantagens sobre as análises convencionais, devido principalmente, a sua alta eficiência na obtenção de informações. Segundo Shepherd e Walsh (2002), esta é uma das técnicas analíticas mais eficientes, econômicas, reproduzíveis e disponíveis no século 21. Uma vez que, podem ser obtidos uma grande quantidade de atributos físicos e químicos do solo com apenas uma leitura da resposta espectral, sempre que existirem modelos estatísticos calibrados para extrair estas informações (SINGH, *et. al* 2019).

A espectroscopia infravermelha apresenta grande potencial para avaliação de propriedades e atributos com conseqüente mapeamento de solo com alta resolução, uma vez que, se trata de uma técnica que permite obter resultados rápidos e precisos com emprego de métodos simples e de baixo custo, quando comparado com métodos convencionais (Soares, 2017).

Diversos pesquisadores apontam a existência de inúmeras evidências da interação entre a espectroscopia e os atributos do solo. Entretanto, a elevada sensibilidade às variações da composição do solo, cria o desafio de construção e

calibração de modelos capazes de representar tais variações, obtidos a partir de diferentes densidades de informações do solo geradas indiretamente por meio de sensores espectroscópicos.

Diante do exposto, objetivou-se com o presente trabalho calibrar e validar modelos de predição de atributos químicos do solo a partir da resposta espectral (Vis-NIR). Para isso, foram contemplado os seguintes objetivos específicos: a) Construir uma biblioteca de dados espectrais com solos representativos do Estado da Bahia, que possibilite estabelecer modelos preditivos de atributos relacionados com fertilidade do solo, utilizando espectrometria da região próxima ao infravermelho; b) Calibrar e validar modelos para estimativa de atributos químicos para os solos mais representativos do estado da Bahia.

2. REVISÃO BIBLIOGRÁFICA

2.1. Variabilidade espacial dos atributos do solo

O solo é um sistema dinâmico, que sofre a interação de diversos fatores, principalmente em ambientes tropicais. Sua composição química apresenta comportamento diferente quanto a sua distribuição no solo, exigindo muitas vezes uma maior amostragem na avaliação da variabilidade em uma determinada área (Zanão Junior et al., 2007).

Essa variabilidade é consequência de complexas interações dos fatores e processos de sua formação (Corá et al., 2004). Áreas com características pedológicas similares podem apresentar diferenças significativas em seus atributos quando submetidas às diferentes práticas de manejo (Silva et al, 2007).

No Brasil a pratica de implementação de insumos pode induzir ao erro uma vez que é baseada em teores médios da fertilidade do solo, o conhecimento detalhado da variabilidade espacial dos atributos químicos, pode gerar aplicações de corretivos e fertilizantes mais efetivas, melhorando o controle do sistema de produção das culturas e consequentemente das contaminações ambientais (Cavalcante et al., 2007). Algo muitas vezes impraticável devido aos custos demandados em análises convencionais nessa escala.

Sendo assim, faz-se necessário a implementação de novas tecnologias que possibilitem a avaliação dos atributos do solo de forma mais eficiente, com o objetivo de melhorar a capacidade produtiva e conservação dos recursos ambientais sobretudo para um manejo racional dos insumos agrícolas (Grego et al., 2005).

Inúmeros estudos apontam para a importância do uso de sensores eletromagnéticos na avaliação da distribuição dos componentes do solo que fornecem valiosas informações de suas características possibilitando a separação em campo de áreas homogêneas (Adamchuk et al., 2004). Grandes dificuldades são encontradas para caracterização e estruturação da variabilidade espacial dos atributos químicos do solo, entre elas se destaca a necessidade de um grande número de análises de solo, elevando o custo para obtenção da informação (Santos, 2011). Estudos como os de Montezano et al. (2006), apontam que a variabilidade espacial de componentes é alta para fósforo disponível, e média, para matéria orgânica, cálcio, magnésio, acidez potencial, e baixa, para pH e potássio em uma área cultivada com milho, reforçando a necessidade de uma maior amostragem para resultados mais efetivos.

As análises convencionais de solo envolve levantamento, amostragem e análise química, que muitas vezes é caro, destrutivo e requer o uso de vários reagentes químicos. Além disso, o processo típico de análise laboratorial pode levar dias e até meses para uma amostragem maior. Rápida determinação e quantificação de nutrientes do solo de forma eficiente e precisa são essenciais, para um bom manejo. A espectroscopia de infravermelho de refletância difusa (IR) é um método rápido e não destrutivo que correlaciona a radiação refletida de um superfície do material com propriedades desse material através do uso de estatísticas multivariadas. Também permite a previsão simultânea de propriedades múltiplas usando o mesmo espectro (Wartini, et. al, 2020).

2.2 Espectroscopia Vis-NIR

De maneira geral, a espectroscopia consiste na avaliação da radiação emitida ou absorvida por um determinado corpo, podendo ser na forma de luz visível, infravermelho, ultravioleta, entre outras. Já a espectrofotometria é o emprego da técnica analítica que permite medir a energia eletromagnética emitida ou absorvida por átomos ou moléculas relacionando-a com propriedades específicas dos mesmos (Cienfuegos; Vaitsmain, 2000).

Quando certa irradiação incide sobre uma amostra partes da mesma é seletivamente absorvida dando origem ao que conhecemos como espectro. Graças a construção química específica de cada material estas respostas para a absorção da luz serão diferentes o que pode possibilitar a identificação dos mesmos. Estes espectros são obtidos com o emprego de equipamentos conhecidos como espectrofotômetros, que por apresentarem características de construção distintas operam em faixas de comprimento de onda diferentes e que reflete também na forma de obtenção da medida espectrofotométrica podendo ser reflectância ou transmitância (Saliba, 1999).

Os espectros, sejam de absorção ou emissão, para átomos são apresentados geralmente por linhas bem distintas, já para as moléculas essas linhas de absorção são muito próximos e por isso são conhecidas como bandas (Saliba, 1999). As energias associadas as bandas de uma molécula qualquer estão relacionadas as energia de vibração, rotação e as energias provenientes das nuvens eletrônicas, esta última é observada nas regiões do espectro (tabela 2.1) conhecidas como ultravioleta e visível, as rotacionais na região do micro-ondas já as vibracionais na região do infravermelho (Skoog et al., 2002; Schrader, 1995).

Sendo muito extensa a região correspondente a faixa do infravermelho, esta é convenientemente dividida em infravermelho próximo – NIR, Infravermelho médio – MIR e infravermelho distante – FIR (Skoog et al., 2002).

Tabela 2.1 Regiões espectrais

Região		Comprimento de onda (nm)
Ultravioleta		200 a 400
Vísivel		400 a 780
Próximo - NIR		780 a 2500
Infravermelho	Médio - MIR	2500 a 5000
	Distante - FIR	5000 a 100.000
Micro-ondas		1.000.000 a 1.000.000.000

Fonte: Skoog et al. (2002) Princípios da análise instrumental.

Vale salientar que esses valores não possuem limites específicos bem definidos, sendo comum encontrar na literatura diferentes valores, pois a transição de uma faixa para outra é gradual (Pasquini, 2003).

Para estudos de avaliação das propriedades do solo tem sido comumente utilizadas duas faixas espectrais, a do visível e a do infravermelho próximo (Soriano-Disla et al., 2014). Comparado ao modelo tradicional de obtenção dos dados referentes a fertilidade do solo, o uso de informações espectrais possui muitas vantagens principalmente por permitir a realização de estudos qualitativos e quantitativos, possibilitando o trabalho com grupamentos orgânicos e inorgânicos além da extração de várias informações da mesma amostra com apenas um espectro (Rossel et al., 2006). Outro fator a ser considerado para adoção da espectroscopia Vis-NIR, diz respeito a exigir pouco ou nenhum preparo das amostras, constituindo ainda uma técnica de fácil manuseio e de aquisição espectral rápida (Rossel et al., 2006).

Muitos autores relatam bons resultados na geração de modelos de predição de atributos químicos do solo, utilizando a espectroscopia Vis-NIR (Rossel et al., 2006; Demattê; Terra; Viscarra Rossel, 2015; Zhang et al., 2016). Estes autores ressaltam a existência de obstáculos como variabilidade da matriz do solo, sensibilidade do equipamento utilizado além da construção de modelos que sejam mais eficientes em relacionar sinais analíticos com a propriedade desejada. Estes desafios precisam ser

transpostos para que a espectroscopia possa ser dominada como uma ferramenta para quantificação dos atributos do solo relacionados a fertilidade.

2.3 Breve histórico da espectroscopia

A radiação eletromagnética referente à faixa do infravermelho próximo (NIR) foi identificada pela primeira vez em 1800 por Sir William Herschel, tendo sido a primeira parte do espectro eletromagnético não visível a ser identificada. Passaram-se quase um século e meio após sua descoberta, até que fossem realizadas as primeiras aplicações analíticas da radiação NIR. Um dos fatores para este atraso, está relacionado à falta de seletividade dos espectros de absorção na região NIR, sempre referida pelos espectroscopistas de então como uma região espectral de difícil interpretação, na qual os espectros são constituídos por inúmeras bandas de absorção fortemente sobrepostas (Pasquini, 2003).

Esta dificuldade gerou desinteresse resultando no adiamento até a década de 1960 para construção de equipamentos dedicados a explorar analiticamente esta faixa do espectro NIR com qualidade. Além disso, métodos quantitativos clássicos para o tratamento de dados demandam alta seletividade, dificultando o aproveitamento analítico das informações geradas pela espectroscopia NIR (Sousa, 2016).

Embora exista registros na literatura de trabalhos anteriores que fizeram uso da espectroscopia NIR, foram os trabalhos desenvolvidos por Karl Norris, no Cereal Centre, em Beltsville, nos Estados Unidos, que de fato disseminou os princípios da técnica analítica no meio acadêmico. Estes estudos foram prosseguidos por Thomas Hirschfeld, um espectroscopista reconhecido, nas décadas de 1970 e 1980 por relevantes trabalhos, colaborando para o desenvolvimento da técnica no ambiente acadêmico e que permitiu aumentar o conhecimento sobre esta região espectral até então negligenciada (Pasquini, 2018).

Já a inserção da espectroscopia NIR no âmbito industrial pode ser atribuída a Phil Williams, que ao substituir o método de Kjeldahl, massivamente difundido na Canadian Grain Commission, pelo então recentemente introduzido método de determinação de proteína baseado em espectroscopia NIR trouxe as atenções comerciais da época para o uso prático de uma nova tecnologia analítica rápida e limpa na substituição de um método via úmida que além de demandar várias horas na realização de uma única análise, empregava reagentes tóxicos de difícil descarte, por uma técnica rápida e limpa, despertando o interesse de companhias fabricantes de instrumentos, que começaram a produzir equipamentos com especificações

necessárias para o desenvolvimento da técnica baseada na faixa espectral do NIR (Naes et al., 2002).

Os memoráveis trabalhos desenvolvidos por Karl Norris revelaram que para o desenvolvimento da espectroscopia NIR era necessário contornar o problema da falta de seletividade da informação analítica contida nos espectros, por isso, a busca por ferramentas matemáticas se fez necessário (Magalães et al., 2006). De fato, para a determinação da umidade presente na farinha de trigo, descrita no primeiro trabalho de Karl Norris, foi necessário estabelecer a correlação entre o conteúdo de água com a diferença de absorbância e comprimentos de ondas específicos, a fim de corrigir os efeitos de espalhamento da radiação, comumente observado em sólidos pulverizados. Esta necessidade para implementação da tecnologia NIR motivou a incorporação de técnicas quimiométricas no tratamento multivariado dos dados (Naes et al., 2002).

A utilização da técnica de regressão por quadrados mínimos parciais uma das mais empregadas até os dias de hoje foi aplicada pela primeira vez como método analítico para espectroscopia NIR em 1984 por Harold Martens, durante a 2ª conferência Internacional em Reflectância Difusa, realizada no Estados Unidos e desde então, passou a ser desenvolvida e difundida como suporte aos fundamentos da espectroscopia vibracional, a instrumentação espectrofotométrica NIR e a quimiometria (Williamms, 2015).

A técnica vêm passando por evoluções ao longo dos anos e hoje é possível observar a implementação de fibras ópticas, o uso de técnicas analíticas e até a miniaturização dos espectrofotômetros que possibilitam inclusive as imagens hiperespectrais, estas inovações tem permitido a incorporação rotineira da técnica em diversas áreas do setor produtivo graças a relevância dos métodos analíticos qualitativos e quantitativos que permitem a validação e manutenção de novos métodos (Pasquini, 2018).

2.4 Fundamentos da espectroscopia

O ramo da ciência que se dedica ao estudo da geração da radiação eletromagnética e sua interação com a matéria é denominado de espectroscopia (Skoog,2002). Uma vez que, o espectro eletromagnético se estende por uma ampla faixa de energia, são inúmeras as formas de interação entre a matéria e a radiação e por consequência são inúmeras também os fenômenos provocados por essas interações muitas vezes relacionados a transferência de energia para seus constituintes (Pasquini, 2018).

Dáí a importância de estudar o espectro eletromagnético dividido em faixas mais ou menos específicas, de acordo com suas energias e consequentemente com os

fenômenos que elas podem provocar ao interagir com a matéria. Entre as muitas formas de interação da radiação com a matéria, a absorção da radiação eletromagnética pelos constituintes da amostra é a mais interessante para o estudo da espectroscopia analítica, uma vez que, os espectros gerados a partir deste fenômeno possuem informações analíticas qualitativas e quantitativas a cerca de uma amostra (Sousa, 2016).

A energia contida nas ondas eletromagnéticas NIR determina a forma de interação com a matéria e causam transferência energéticas em processos de absorção da radiação. Um exemplo disso, está no fato de a energia transportada pela radiação NIR ser relativamente baixa tornando rara a sua interação com a matéria ao ponto de promover aumento da energia eletrônica das moléculas ou átomos, como é percebido nas interações entre a matérias com as suas regiões vizinhas do visível e ultravioleta, de maior energia, apenas poucos e raros elementos químicos possuem essa capacidade (Pasquini, 2018).

A absorção de luz no infravermelho próximo está relacionada a natureza das ligações moleculares e pelas relações entre os átomos que formam tais moléculas. O princípio que envolve a absorção no infravermelho consiste na irradiação da amostra com luz de comprimento de ondas específicos, a absorção da luz provoca vibração das moléculas da amostra, sendo medida pela diferença entre a radiação emitida pelo infravermelho e a refletida pelo material. Isso só é possível, graças ao fato, das ligações moleculares não serem estáticas no espaço, vibrando o tempo todo, em movimentos específicos de estiramento e compressão que variam de acordo com os elementos envolvidos e das ligações entre eles (Almeida, 2009).

Para que uma molécula absorva radiação infravermelha é necessário que seu momento dipolo (diferença de polaridade entre os átomos de uma molécula) seja alterado, sendo consequência do movimento vibracional ou rotacional da mesma (Skoog, 2002). A absorção da radiação eletromagnética ocorre quando a energia radiante corresponde a diferença entre dois níveis energéticos vibracionais. Para tanto é necessário entender que há duas formas básicas para que haja mudança dos níveis energéticos de uma molécula, quando ela recebe a radiação e a mudança de nível energético ocorre do estado 0 para 1 diz-se que ocorreu uma transição fundamental e são geradas por movimentos vibracionais harmônicos, mais quando o movimento vibracional é enarmônico, de principal interesse da espectroscopia NIR, fazendo com que a molécula ao mudar de nível energético o faça do 0 para o 2, ou seja, quando há o salto de dois ou mais níveis energéticos chama-se essas transições de sobretom (Pasquini, 2018).

Desta forma, as moléculas que principalmente interagem com a radiação NIR são as diatômicas heteronucleares, poliatômicas com vibrações assimétricas e grupos funcionais ligados a polímeros, englobando a maioria dos compostos orgânicos e muitos inorgânicos (Solomons, 2001). Já as moléculas homonucleares O_2 , N_2 ou Cl_2 , não absorvem radiação NIR durante a vibração de suas estruturas (Skoog, 2002).

O espectro de determinada amostra de solo obtido pela radiação NIR, é resultado da absorção de energia pelas moléculas orgânicas que possuem ligações do átomo de hidrogênio com outro elemento mais eletronegativo como carbono, nitrogênio ou oxigênio (Solomons, 2001).

Como manifestado, o grau de movimentos não harmônicos das ligações químicas envolvidas nos modos vibracionais é muito relevante para a espectroscopia NIR. Desta forma, ligações químicas com esta característica têm maior probabilidade de gerarem bandas de absorção na região NIR. Além disso, para que uma banda de absorção, por causa de um sobreton, seja observada na região NIR, a energia da transição fundamental, no caso dos sobretons, ou das transições fundamentais nas combinações, deve ser alta. Assim, a energia destes modos vibracionais situa-se na região espectral NIR. Estes requisitos são atendidos por ligações envolvendo o átomo de hidrogênio ligado a um átomo mais pesado, como carbono, nitrogênio oxigênio e enxofre. Desta forma, espectros NIR sempre apresentam bandas de absorção referentes aos sobretons e combinações de modos vibracionais associados às ligações C – H, N – H, O – H e S – H, denominadas de grupos funcionais. Outra possibilidade vem de ligações químicas de alta energia, como C = C, C \equiv C e C = O, cujas interações com a radiação eletromagnética podem ser observadas na região NIR (Pasquini, 2018).

Desta forma, pode-se dizer que a resposta espectral dos solos é devido principalmente aos argilo-minerais, óxido de ferro e a matéria orgânica. Como os solos possuem diferentes constituintes estes determinam o grau de radiação refletida pela amostra comparada ao grau de radiação que incidiu sobre a mesma, fornecendo a medida da reflectância que é captada pelos sensores, como a medida da reflectância é determinada pelos diferentes constituintes do solo, isso permite que os mesmos possam ser quantificados pela análise de suas respostas espectrais (Dalmolin et al., 2005).

2.5 Quimiometria

Nos estudos que envolvem centenas e milhares de variáveis, o estudo univariado, apesar de fornecer informações importantes não responde com o interesse pretendido, por isso, é adotado com alternativa a esse problema o estudo com análise

multivariada apoiando-se em algoritmos que permitam gerar as informações pretendidas. Modelos estatísticos que empregam métodos multivariados consideram a correlação entre diversas variáveis simultaneamente, possibilitando a extração de uma quantidade muito maior quando comparada a análise univariada (Sena et al., 2000).

Analises multivariadas possuem como princípio fundamental a utilização de muitas variáveis, como por exemplo valores de absorbância para muitos comprimentos de onda, possibilitando a quantificação de outra variável de interesse (Sena et al., 2000).

O maior problema que se observa ao trabalhar com os dados referentes a espectroscopia de infravermelho é o processamento e interpretação das informações geradas. Devido ao grande número dessas informações é necessário o emprego de uma ferramenta que permita a retirada das informações necessárias as análises qualitativas e quantitativas (Ferrarini, 2004).

Da necessidade de extrair informações é que surge a quimiometria, consistindo em uma técnica que aplica ferramentas estatísticas e matemáticas em dados de origem química. O uso desses tratamentos permitem a separação física de determinadas espécies pela separação quimiométrica de seus sinais analíticos (Dias, 2011).

A implementação de ferramentas matemáticas e estatísticas nos estudos químicos, outrora já empregadas em outras áreas da ciência, possibilitou o desenvolvimento da quimiometria, impulsionada pelo avanço tecnológico na área instrumental. Quando espectrômetros e cromatógrafos se destacaram no ambiente industrial, em 1960, trazendo consigo uma avalanche de dados, se fez necessário, buscar ferramentas matemáticas e estatísticas capazes de extrair dos dados gerados o máximo de informação útil, associado ao aparecimento de computadores capazes de processar essa grande quantidade de dados, este ambiente de evoluções favoreceu o surgimento e rápido desenvolvimento da quimiometria (Sousa, 2016).

Desde então a quimiometria tem se apresentado como solução para diversos desafios na ciência. Define-se a quimiometria como a aplicação de ferramentas matemáticas e estatísticas para o tratamento, a interpretação e a previsão de dados estatísticos (Ferreira, 2015). O que torna a quimiometria indispensável para muitas áreas da química, inclusive para espectroscopia de infravermelho, é a classificação das amostras, a partir de suas propriedades possibilitando a quantificação de suas variáveis além da previsão de resultados (Williamms, 2015).

Por possuir diferentes aplicações, pode-se dividir a quimiometria em 5 etapas, no processamento dos sinais analíticos, otimização para planejamento de

experimentos, na classificação de dados a partir do reconhecimento de padrões, na calibração multivariada e por fim em métodos de inteligência artificial (Mazur, 2012).

Sendo considerada uma das mais recentes incorporações da química analítica por tornar possível o processamento e interpretação de dados que antes não poderiam ser analisados, a quimiometria tem se encarregado de empregar ferramentas matemáticas e estatísticas associadas a técnicas instrumentais na produção de dados, que posteriormente serão utilizados na elaboração de modelos empíricos para estimar propriedades de ordem qualitativas e quantitativas (Sousa, 2016).

A espectroscopia NIR gera um volume imenso de dados que necessitam o emprego da quimiometria para melhor interpretação, nesse sentido, esta ciência tem sido empregada na construção de modelos eficazes na predição de propriedades químicas. Espectros NIR são caracterizados pela sobreposição de bandas tornando necessário o emprego de técnicas multivariadas para análise dos dados dos espectros, se destacando a Regressão em componentes principais (PCR – Principal Components Regression) e a Regressão por quadrados mínimos parciais (PLS – Partial Least Squares) (Mazur, 2012).

Devido às dificuldades em interpretar os espectros NIR caracterizados por bandas sobrepostas, como apresentado anteriormente, foram desenvolvidas técnicas multivariadas para análise dos dados dos espectros, como análise de componentes principais (PCA) e regressão por mínimos quadrados parciais (PLSR), as quais permitem a construção de modelos quantitativos e qualitativos, além de análise de reconhecimento de padrões (MAZUR, 2012). Neste tipo de calibração multivariada, a propriedade de interesse, é determinada por metodologias padrão e representada no método vetor, já a resposta instrumental é fornecida em forma de matriz (Valderrama, 2007).

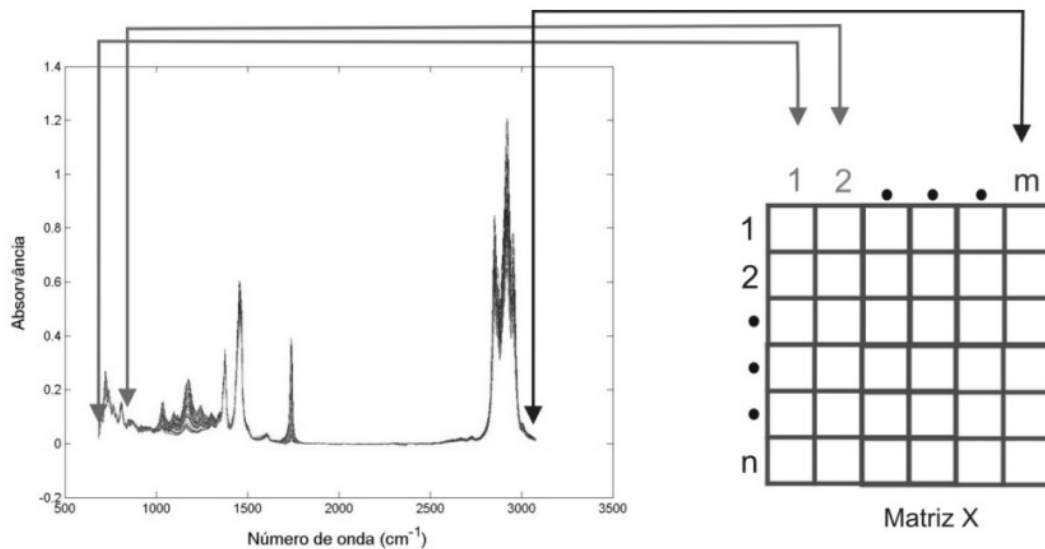
2.6 Calibração multivariadas

Na maioria das vezes as análises químicas tem por finalidade a determinação qualitativa e/ou quantitativa de determinado componente em amostras. Instrumentos laboratoriais não possuem a capacidade de fornecer essa resposta diretamente, para tanto é necessário a elaboração de um método empírico, que permita a extração dessa informação, dos dados gerados por estes equipamentos. Nesta perspectiva, surge a calibração como ferramenta para aplicação de um algoritmo matemático que colabore na interpretação dos dados gerados, estabelecendo uma relação entre dois grandes

blocos de dados, de um lado a informação química e do outro as medidas instrumentais (Valderrama, 2005).

A calibração multivariada consiste basicamente em duas etapas, o desenvolvimento ou calibração e a validação (Naes et al., 2002). Na calibração são estabelecidas relações entre os sinais analíticos, representados em uma matriz (figura 2.1), e a concentração dos analitos, que são estimadas pela verificação de padrões representativos para o conjunto de amostras analisadas (Borges Neto, 2005). De forma geral, o desenvolvimento do modelo de calibração depende estritamente da relação entre estas duas informações sempre que as propriedades que caracterizam as mesmas apresentam dependência entre si.

Figura 2.1 – Representação da construção de uma matriz para calibração multivariada.



Fonte: Souza (2015). Quantification of adulterations in extra virgin flexseed oil using MIR and PLS

Para aplicações de análises multivariadas de quantificação e qualificação na espectroscopia NIR existem muitos algoritmos, entre eles se destacam a análise de cluster, regressão múltipla linear, PCA, PCR e a PLS.

A PLS vem sendo muito empregada para análise de atributos do solo a partir do espectro Vis-NIR, por ser um método linear multivariado e possuir uma vantagem natural no gerenciamento de colinearidade forte com baixa complexidade computacional quando comparado a outros métodos (Zhang et al., 2021).

2.7 Regressão por mínimos quadrados parciais

Herman Wold desenvolveu este método por volta dos anos 1960, tendo sido aplicada inicialmente a econometria, sendo posteriormente difundida com a ajuda de Swante Wold (Geladi e Kowalski, 1986). A regressão por mínimos quadrados parciais (PLS), do inglês Partial Least Square Regression (PLSR), é o método de calibração mais empregado em análises multivariadas, principalmente em aplicações que envolvem a espectroscopia de infravermelho médio e próximo (Ferreira, 2015). O método é eficaz em situações em que a resposta instrumental não é específica para a espécie química de interesse conseguindo extrair a informação desejada, estabelecendo a relação entre a propriedade estudada e a resposta espectral (Dias, 2011).

O método se apoia na vantagem multicanal que aparece quando é utilizado duas ou mais variáveis correlacionadas, assim, mesmo as variáveis sofrendo interferências como ruídos, a informação poderá ser extraída graças a informação redundante presentes nas variáveis utilizadas o que torna o método robusto, uma vez que, as distorções causadas por interferentes são minimizadas pela ação da informação redundante, já que todas as variáveis contribuem com o cálculo do modelo (Naes et al., 2002).

Outra etapa presente no método é a compressão dos dados, que consiste em uma nova matriz com o mesmo número de variáveis da matriz original, mais o cálculo realizado condensa a grande quantidade de informações permitindo que a matriz original seja reconstruída em poucas variáveis latentes, funcionando como um filtro que elimina parte da interferência, geralmente ruído instrumental (Ferreira, 2015).

No processamento das matrizes as informações referentes as variações espectrais e os valores de referência são traduzidos em duas matrizes gerando vetores próprios, chamados de fatores ou componentes principais – CPs, estes por sua vez, são utilizados na previsão da concentração, já que os espectros NIR originais contêm todas as informações relevantes para o sistema em estudo. Esta decomposição em fatores possui como vantagem o fato de que toda a informação importante ficar comprimida podendo ser utilizada na calibração (Almeida, 2009).

O modelo de calibração PLS permite decompor as matrizes X (matriz dos espectros ou scores) e Y (Matriz das concentrações ou loadings) simultaneamente, buscando a máxima covariância entre as CPs e suas respectivas matrizes, gerando duas variantes do PLS, PLS1 e PLS2 necessárias para construção dos modelos de calibração. Enquanto o PLS1 possibilita apenas que as CPs sejam correlacionadas uma

por vez a cada propriedade de estudo, a PLS2 permite que as CPs possam ser correlacionadas simultaneamente com diferentes propriedades (Ferreira, 2015).

A utilização da metodologia PLS possibilita ainda, o emprego de diferentes algoritmos como o SVD (Singular Value Decomposition), e o NIPALS (Nonlinear Iterative Partial Squares), que geram modelos mais robustos e precisos, mas demandam algum tipo de pré-processamento para o melhoramento dos resultados.

No entanto, os espectros Vis-NIR do solo contêm vários sinais complexos com diferentes absorções sobrepostas e multicolinearidade, então eles são principalmente inespecíficos. Portanto, a estratégia de pré-processamento espectral é indicada para remover o ruído sistemático indesejado, para corrigir mudanças de linha de base e para restaurar a relação linear e, conseqüentemente, melhorar o desempenho de calibração (Zhang et al., 2021).

2.8 Validação do modelo

Após a construção do modelo de calibração multivariada, é necessário avaliar a sua capacidade preditiva através do processo de validação, que pode ser dividida em validação interna e externa. A validação interna utiliza as mesmas amostras outrora empregadas na construção do método, já a validação externa utiliza amostras que não participaram da construção do modelo e que possuam características semelhantes (Ferreira, 2015).

Um dos métodos mais comumente empregados na validação interna é a validação cruzada (cross validation), onde uma amostra é deixada de fora da elaboração do modelo para que seja realizada a previsão das propriedades em estudo dessa amostra, este procedimento é repetido até que todas as amostra tenha sido excluídas pelo menos uma vez. Posteriormente, os valores da previsão são comparados aos valores de referência para obtenção do erro, prevalecendo o modelo que gerar o menor erro (Naes et al., 2002).

Este procedimento resulta no erro denominado RMSECV (Root Mean Squared Error of Cross Validation) representado na equação 2.1, onde n representa o número de amostras usadas na calibração, y_i referência o valor de referência da amostra i e y_i previsto o valor previsto para a amostra i . Espera-se valores baixos de RMSECV para bons modelos.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i \text{ referência} - y_i \text{ previsto})^2}{n-1}} \quad (2.1)$$

Por meio dessa técnica é possível definir o número ótimo de CPs para o modelo e quais amostras possuem perfis atípicos às outras, a definição desse número é relevante pois, sendo esse número excessivo favorecerá um modelo superajustado levando a deterioração da análise por ter incorporado valores do ruído e de variações sistemáticas, entretanto valores insuficientes de fatores resultam em modelos pouco eficazes por possuírem menos informação do que se necessita, sendo o ideal encontrar o número de fatores que forneçam a melhor previsão com o menor erro possível (Ferreira, 2015).

A calibração cruzada pode ser empregada em situações onde não existam número de amostras suficientes para serem divididas em dois grupos de calibração e validação, se o número de amostras for suficiente será mais vantajoso a validação externa por tornar o sistema mais robusto. A robustez está relacionada a capacidade do método de adequação a variações causadas pelo operador, condições ambientais, compactação do material, alterações instrumentais, entre outras, ou seja é menos sensível a variações irrelevantes possuindo maior estabilidade. Deve-se considerar ainda que estas variações devem estar presentes em ambos os conjuntos de calibração e validação, para que sejam obtidas respostas mais eficazes para novas amostras a serem analisadas (Almeida,2009).

A qualidade do método de calibração é definido pela precisão que apresenta ao prever propriedades de novas amostras, a partir da média global do erro de previsão, este erro de previsão para amostras que empregam espectros NIR em seu método corresponde ao RMSECV para casos de validação cruzada e RMSEP para casos de validação externa (Naes et al., 2002).

Outros fatores também são utilizados para determinação no número ótimo de fatores, como é o caso da soma quadrática do erro residual de predição PRESS (Prediction Residual Error Sum Square) e da raiz da média quadrática RMSEP (Root Mean Squares Error of Prediction) representados nas equações a seguir:

$$PRESS = \sum_{i=1}^n (y_i \text{ referência} - y_i \text{ previsto})^2 \quad (2.2)$$

$$RMSEP = \sqrt{\frac{PRESS}{n}} \quad (2.3)$$

Outra forma de validar a precisão do método de calibração é com base no coeficientes de determinação R^2 , sendo a razão de desvio padrão (SD) do conjunto de dados para a raiz do erro quadrático médio da validação cruzada (RMSECV) ou da

previsão (RMSEP). Na verdade, R^2 indica a porcentagem da variação na variável Y que pode ser contabilizado pela variável X.

A eficiência dos modelos também pode ser medida pelo desvio preditivo residual (RPD), calculado a partir da proporção do desvio padrão dos dados medidos em laboratório (referência) pelo RMSE da validação. É o fator que mede se a precisão da previsão foi aumentada em comparação com a média dos dados originais. Os valores de RPD são classificados da seguinte forma: RPD 1.0 indica modelo com previsões muito pobres e seu uso não é recomendado; RPD entre 1,0 e 1,4 indicam modelos com previsões ruins, onde apenas os valores altos e baixos são distinguível; RPD entre 1,4 e 1,8 indicam modelos justos com previsões que podem ser usadas para avaliação e correlação; Valores de RPD entre 1,8 e 2,0 indicam bons modelos onde as previsões quantitativas são possíveis; RPD entre 2,0 e 2,5 indicam que os modelos podem ser utilizados para previsões quantitativas; RPD igual ou superior a 2.5 indicam excelentes modelos (Viscarra et al., 2006).

Todos conjuntos de dados podem apresentar amostras com comportamento discrepante, Modelos multivariado podem ser melhorados pela identificação e exclusão dessas anomalias no conjunto de calibração, pois tais informações discrepantes favorecem falsas conclusões dos dados gerados pelos modelos. As amostras anômalas (outliers) podem ser identificadas pelo emprego de dois parâmetros, o primeiro conhecido como leverage representa a influência de cada amostra para o modelo de regressão e o segundo resíduos de Student indica se a amostra possui distribuição normal com nível de confiança de 95% admitindo-se limites de variação de ± 2 (Ferreira, 2015).

3. MATERIAL E MÉTODOS

O estudo foi realizado em amostras de terra de 32 perfis de solos, totalizando 154 amostras. Cerca de 130 amostras utilizadas no trabalho são resultantes de dados secundários oriundos de outros trabalhos de dissertação do programa de pós-graduação em Solos e Qualidade de Ecossistemas da Universidade Federal do Recôncavo da Bahia. As amostras foram coletadas de horizontes de solos representativos do território do Estado da Bahia, distribuídos como mostra a figura 3.1, os mesmos foram selecionados de modo a incluir a diversidade de classes de solos com diferentes granulometria, mineralogia e os teores de carbono, fosforo, sódio, potássio, cálcio, magnésio, alumínio e hidrogênio.

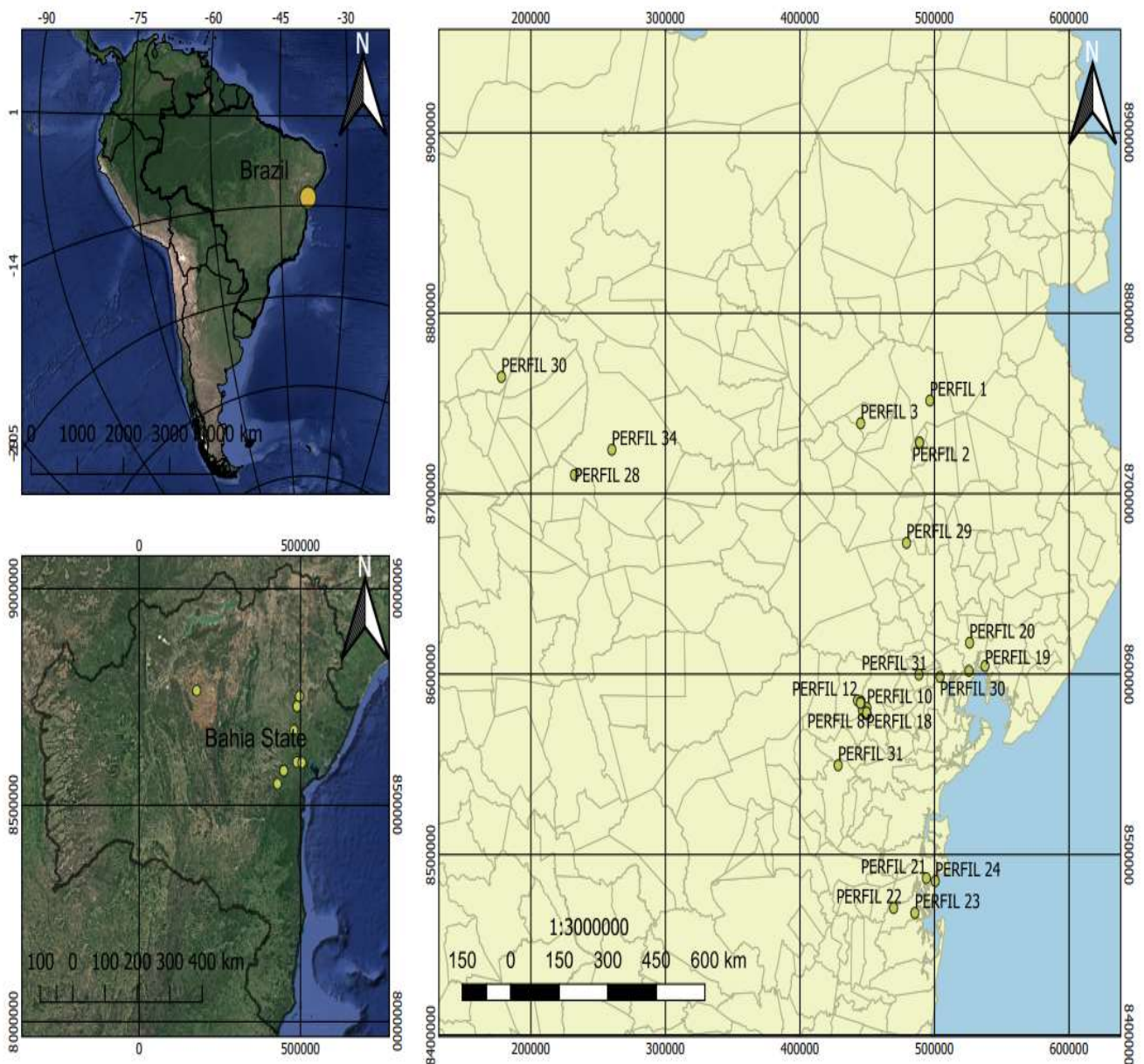


Figura 3.1 – Mapa com a localização dos perfis coletados.

As amostras selecionadas foram secas ao ar, destorroadas, peneiradas, armazenadas e divididas em dois lotes de aproximadamente 50g. Todas as amostras foram submetidas a avaliação por reflectância de infravermelho por meio do espectrorradiômetro FieldSpec 3 da panalytical e análises da fertilidade do solo utilizando os métodos convencionais (conforme Embrapa, 2017), como demonstrado pelo fluxograma apresentado na figura 3.2.

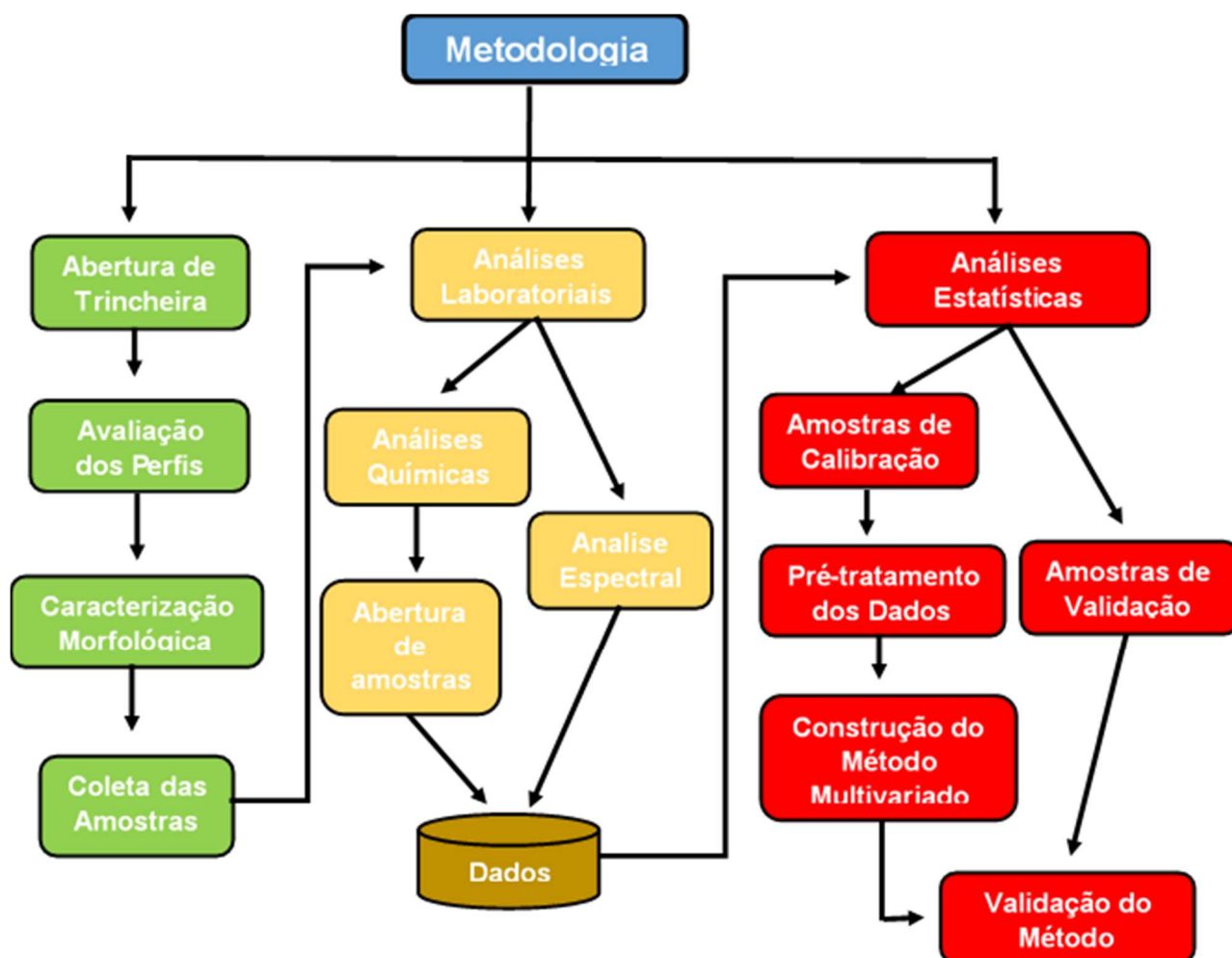


Figura 3.2 – Fluxograma das atividades desenvolvidas para as análises químicas de solos do estado da Bahia.

As amostras foram coletadas em dois dos três biomas identificados na Bahia, Caatinga e Mata Atlântica, foram selecionadas amostras de 32 perfis (Tabela 3.1), englobando 9 classes de solos, com esta diversidade amostral buscou-se tornar o estudo o mais representativo possível, uma vez que as diferentes condições em cada ambiente colaboraram para a variabilidade de suas propriedades.

Tabela 3.1 – Identificação dos perfis, classes e localização de solos avaliados no Estado da Bahia.

Identificação	Classe de Solos	Localização
Perfil 1	Planossolo Háplico Eutrófico típico	Araci
Perfil 2	Neossolo Regolítico Eutrófico Léptico	Araci
Perfil 3	Neossolo Regolítico Eutrófico léptico	Valente
Perfil 4	Cambissolo Háplico Ta Eutrófico vértico	São Domingos
Perfil 5	Neossolo Litólico Hístico típico	Santa Terezinha
Perfil 6	Organossolo Fólico	Santa Terezinha
Perfil 7	Latossolo Amarelo Distrófico típico	Santa Terezinha
Perfil 8	Neossolo Litólico Hístico típico	Santa Terezinha
Perfil 9	Neossolo Litólico Húmico típico	Santa Terezinha
Perfil 10	Latossolo Amarelo Distrófico típico	Santa Terezinha
Perfil 11	Latossolo Amarelo Distrófico argissólico	Santa Terezinha
Perfil 12	Latossolo Vermelho Distrófico argissólico	Santa Terezinha
Perfil 13	Organossolo Háplico	Santa Terezinha
Perfil 14	Neossolo Litólico Distrófico fragmentário	Santa Terezinha
Perfil 15	Cambissolo Háplico Tb Distrófico latossólico	Santa Terezinha
Perfil 16	Cambissolo Háplico Tb Distrófico léptico	Santa Terezinha
Perfil 17	Argissolo Vermelho-Amarelo Distrófico planossólico	Santa Terezinha
Perfil 18	Cambissolo Háplico Tb Distrófico latossólico	Santa Terezinha
Perfil 19	Argissolo Vermelho-Amarelo Distrófico típico	Santo Amaro
Perfil 20	Gleissolo Sáfico Sódico vertissólico	Santo Amaro
Perfil 21	Argissolo Amarelo	Ituberá
Perfil 22	Latossolo Vermelho-Amarelo	Ituberá
Perfil 23	Argissolo Vermelho	Igrapiúna
Perfil 24	Espodossolo Humilúvico	Nilo Peçanha
Perfil 25	Latossolo Amarelo Distrófico típico	Elísio Medrado
Perfil 26	Neossolo Litólico Distrófico típico	Elísio Medrado
Perfil 27	Cambissolo Háplico Tb Distrófico	Elísio Medrado
Perfil 28	Espodossolo Humilúvico	Morro do Chapéu
Perfil 29	Plintossolo Pétrico Concrecionário típico	Ubaíra
Perfil 30	Cambissolo Háplico Eutrófico Ta A moderado	Irecê
Perfil 31	Latossolo Amarelo	Cruz das Almas
Perfil 32	Latossolo Amarelo	Morro do Chapéu

A classificação dos perfis foi realizada seguindo os critérios estabelecidos pelo Sistema de Classificação de Solos (Embrapa, 2018). As classes bem como as quantidades de perfis utilizadas neste estudo podem ser conferidas na tabela 3.2.

Tabela 3.2 – Número de perfis de cada classe de solo avaliada no Estado da Bahia

Classe	Número de perfis
Argissolo	5
Cambissolo	5
Espodossolo	3
Gleissolo	1
Latossolo	10
Neossolo	4
Organossolo	2
Planossolo	1
Plintossolo	1
TOTAL	32

3.1 Análises Químicas

A determinação do carbono total do solo foi realizado via úmida pelo método adaptado de Yeomans e Brenner (1998). Em síntese, 0,1g a 0,5g de TFSA macerada e peneirada com malha 0,2mm ao qual foi acrescentado 5 mL de $K_2Cr_2O_7$ 0,167mol.L⁻¹ e 7,5mL de H_2SO_4 concentrado. A mistura foi aquecido em bloco digestor a temperatura de 170°C por 30 minutos. Após esfriamento das amostras a temperatura ambiente foi adicionado 0,3mL de sulfato ferroso amoniacal 0,2 mol.L⁻¹, a solução é levada a titulação com sulfato ferroso amoniacal. O percentual de carbono orgânico é determinado com base nos volumes gastos na titulação da amostra (V), do branco aquecido (Vba) e do branco não aquecido (Vbn) como descritos nas equações a seguir:

$$A = \frac{(Vba - V)(Vbn - Vba)}{Vbn} + (Vba - V) \quad (3.1)$$

$$CO(dag\ kg^{-1}) = \frac{(A)(molaridade\ Sulf.Ferroso)(3)(100)}{peso\ da\ amostra(mg)} \quad (3.2)$$

Os teores de P, K⁺, Na⁺, Ca²⁺, Mg²⁺, Al³⁺ foram realizadas de acordo com método adaptado da Embrapa (2017). Em resumo, os teores de P, K e Na foram determinados em um extrato da solução Mehlich 1 (HCl e 0,05mol L⁻¹ e H_2SO_4 a 0,0125mol L⁻¹)

adicionada as amostras na relação 10:1 (solução:solo) agitado por 16 horas seguida pela leitura, o P por fotocolorimetria, no comprimento de onda de 660nm, já os valores de Na e K foram determinados pela leitura direta em fotômetro de chama.

As concentrações de cálcio, magnésio e alumínio foram extraídos na relação 1:10 solo:solução de cloreto de potássio. Os teores de Ca+Mg, Ca e Al foram feitas por titulometria utilizando EDTA para Ca e Mg e hidróxido de sódio para determinação de alumínio. A acidez potencial ($H^+ + Al^{+3}$) foi determinada por titulometria após extração da amostra com acetato de cálcio (relação 1:10 solo:solução) seguido de titulação com NaOH 0,025 mol L⁻¹ usando como indicador fenolftaleína.

3.2 Aquisição Espectral

Amostras de TFSA foram moídas no moinho de amostras vibratórias C.M.T. Company LTD Modelo TI-100 (Figura 3.3) com frascos e barras em aço Inoxidável em intervalos de tempo com duração entre 5 a 15 segundos, de acordo com a resistência oferecida pelo material. A padronização das condições de leitura foi feita pela nivelção da superfície para diminuir a influência da rugosidade do solo, oriundo dos diferentes tamanhos de partículas (Eitelwein, 2017).

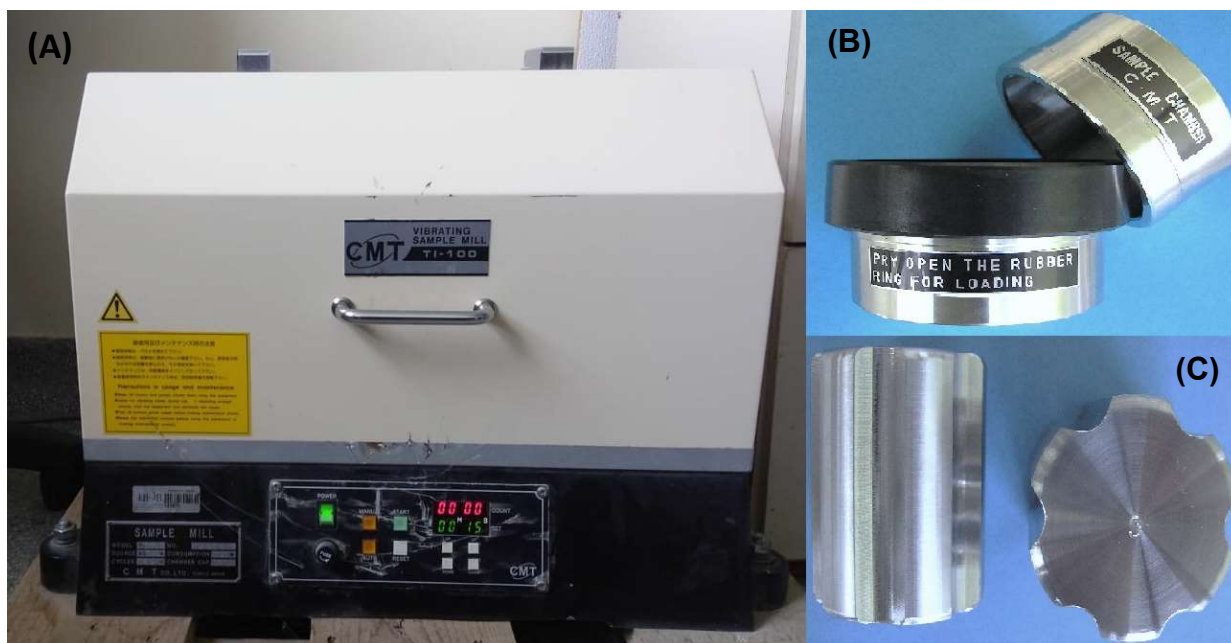
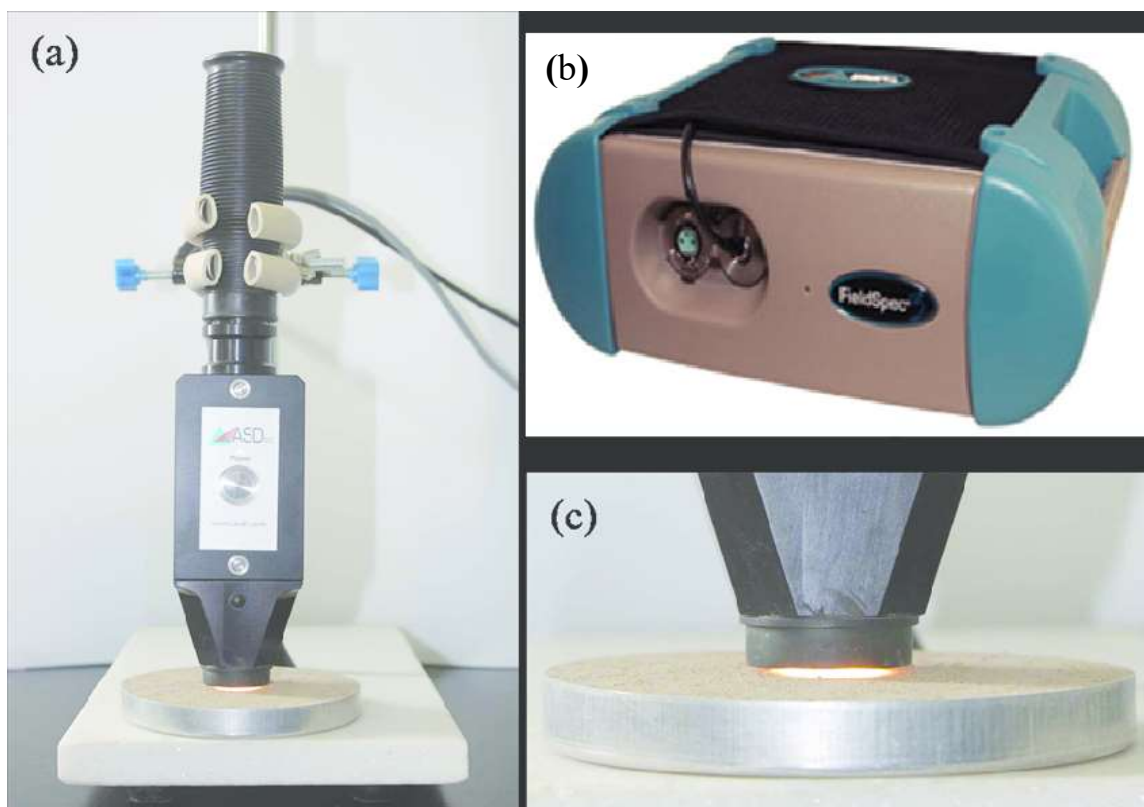


Figura 3.3. – Moinho de amostras vibratórias: (A) Moinho C.M.T. Modelo TI-100; (B) Frasco amostral; (C) Barra moedora.

Cada amostra foi transferida para uma placa de petri e o escaneamento espectral foi feito com o espectrorradiômetro FieldSpec 3 (Figura 3.4) calibrado para realizar 100

leituras por amostra nas faixas de 350nm a 2500nm, o que inclui as faixas do visível e do infravermelho próximo, obtendo-se 2150 bandas espectrais por leituras. Os espectros foram gerados no Laboratório de Sensoriamento Remoto e Espectroradiometria – LABESPECTRO, em parceria com o Programa de pós-graduação em Ciências da Terra e Modelagem – PPGM, da Universidade Estadual de Feira de Santana – UEFS.



Fonte: adaptado de <https://doi.org/10.1371/journal.pone.0151536>

Figura 3.4 – Espectrorradiometro: (a) Sonda de contato; (b) ASD FieldSpec 3; (c) Amostra de solos.

3.3 Análise de Dados

As amostras foram divididas em dois grupos, 32 amostras de validação (Tabela 3.3) e 122 amostras de calibração (Tabela 3.4). Os dados obtidos das amostras de calibração foram utilizados para avaliação multivariada PLS. Por causa da vasta quantidade de informações espectrais fornecidas pelos espectrofotômetros NIR, o grande número de amostras necessárias para construir modelos de classificação e calibração e a alta correlação em espectros, é necessário o emprego de métodos de

redução de variável que permita reduzir o tamanho dos dados originais para algumas variáveis, contendo apenas informações relevantes das amostras. Para isso são empregadas algumas técnicas de estatística multivariada, neste estudo o modelo matemático utilizado foi o PLS com validação cruzada (leave-one-out).

Tabela 3.3. Dados químicos das amostras de validação utilizados para análise dos solos do Estado da Bahia.

Amostra	C.O. g.kg ⁻¹	P mg.kg ⁻³	Na cmol.kg ⁻¹	K	Ca	Mg	Al	H + Al
P1E	1.16	2.00	0.05	0.13	0.70	0.40	0.10	0.20
P7 A	25.24	1.50	0.03	0.08	0.40	0.40	1.00	8.25
P7AB	18.07	0.90	0.03	0.04	0.37	0.17	0.77	6.16
P7BA	13.45	0.50	0.02	0.02	0.20	0.07	0.40	4.84
P7Bw1	8.96	0.57	0.01	0.05	0.27	0.13	0.23	2.92
P9 A1	42.31	3.05	0.08	12.00	0.93	0.27	1.57	10.56
P9A2	22.48	1.82	0.06	0.08	0.57	0.13	1.63	10.84
P9AC	20.38	1.42	0.04	0.05	0.33	0.07	1.77	8.42
P9Cr	14.54	1.11	0.04	0.04	0.23	0.13	1.67	7.81
P13H1	110.62	5.75	0.32	0.09	0.40	0.27	1.47	22.11
P13H2	47.08	2.05	0.11	0.02	0.17	0.03	0.70	12.43
P20 Agnz	10.87	31.40	68.80	0.14	0.70	8.40	0.20	0.20
P21A	12.18	1.00	0.00	0.04	1.20	0.18	0.00	1.10
P21AB	8.12	1.00	0.03	0.01	1.90	1.50	0.00	1.65
P21BA	4.64	1.00	0.03	0.01	0.49	0.04	0.70	3.19
P21Bt1	4.64	1.00	0.03	0.01	0.49	0.04	0.70	3.19
P21Bt2	3.48	1.00	0.01	0.01	0.12	0.03	1.80	4.18
P28A3	24.62	3.65	0.03	0.05	0.00	0.03	1.30	16.17
P28-3E	2.79	6.41	0.04	0.02	0.00	0.00	0.00	2.64
P28BH	5.30	5.92	0.11	0.02	0.30	0.40	1.00	8.75
P31Bfc	6.00	3.00	0.05	0.01	0.07	0.06	0.35	6.39
P32A	9.85	10.43	0.03	0.20	2.40	2.00	0.15	1.70
P32AB	8.91	5.44	0.03	0.14	2.40	2.45	0.10	4.10
P32BA	4.00	1.26	0.04	0.03	1.90	1.80	0.00	2.10
P32B1	5.18	1.72	0.04	0.02	1.90	0.95	0.10	1.20
P32B2	3.83	0.79	0.03	0.03	1.60	1.25	0.10	1.45
P34A	1.08	2.41	0.02	0.08	0.20	0.80	0.85	4.62
P34AB	0.72	1.98	0.02	0.08	0.20	0.30	0.85	4.79
P34BA	0.71	1.70	0.04	0.02	0.15	1.00	0.88	3.96
P34BW1	0.61	2.21	0.04	0.02	0.40	0.70	0.55	3.30
P34BW2	0.59	1.67	0.06	0.01	0.30	0.50	0.25	0.66
P34BW3	0.34	1.59	0.04	0.01	0.20	1.00	0.15	5.61

Tabela 3.4. Dados químicos das amostras de calibração utilizados para análise dos solos do Estado da Bahia – parte 1.

Amostra	C.O. g.kg ⁻¹	P mg.kg ⁻³	Na cmol.kg ⁻¹	K	Ca	Mg	Al	H + Al
P1A	5.80	4.00	0.03	0.29	1.00	0.30	0.10	0.20
P1AE	2.90	2.00	0.02	0.17	0.80	0.20	0.10	0.20
P1-2Bt	1.74	4.00	0.65	0.04	4.20	6.00	0.00	0.20
P2 A	10.40	5.00	0.01	0.39	1.30	0.40	0.00	3.40
P2C1	4.06	2.00	0.01	0.34	1.10	0.40	0.00	0.20
P2C2	1.74	2.00	0.01	0.27	1.20	0.30	0.00	0.20
P2C3	0.58	2.00	0.02	0.23	0.90	0.30	0.00	0.60
P2-2Cr	1.74	2.00	0.40	0.12	2.00	2.70	0.00	1.80
P3 A	3.48	17.00	0.06	0.25	3.00	0.70	0.10	0.60
P3Bi	2.90	15.00	0.10	0.15	3.30	1.10	0.10	1.20
P3C1	2.32	14.00	0.09	0.10	2.50	1.10	0.10	1.00
P3C2	1.16	9.00	0.10	0.09	2.40	1.30	0.20	0.30
P4A	12.18	103.00	0.39	0.35	12.60	10.00	0.20	2.20
P4Bi1	4.64	68.00	1.72	0.34	18.10	14.50	0.00	0.80
P4Bi2	2.32	88.00	2.45	0.33	14.80	14.50	0.00	0.10
P4BC	4.06	95.00	2.49	0.28	13.20	12.70	0.00	0.30
P5O	63.97	12.29	0.05	0.03	0.30	0.20	2.83	16.06
P6O1	84.50	3.05	0.03	0.03	0.27	0.10	0.90	14.52
P6O2	91.38	2.10	0.03	0.02	0.20	0.00	1.17	18.10
P6O3	91.60	1.46	0.03	0.01	0.10	0.23	1.13	18.26
P6O4	76.17	1.34	0.01	0.01	0.20	0.10	0.77	15.29
P6C1	70.79	0.78	0.02	0.01	0.10	0.17	0.73	19.25
P6C2	93.58	0.74	0.02	0.00	0.10	0.40	1.03	21.84
P7Bw2	7.58	0.62	0.01	0.01	0.27	0.00	0.53	3.58
P8 A	70.71	14.18	0.10	0.11	1.10	0.05	2.30	21.78
P8C1	53.15	7.94	0.10	0.05	0.60	0.12	2.70	17.93
P10 A1	14.30	3.95	0.03	0.09	0.73	0.30	0.50	3.52
P10A2	7.35	3.30	0.01	0.03	0.40	0.10	1.33	3.74
P10AB	7.89	1.52	0.02	0.02	0.50	0.20	1.17	3.63
P10BA	5.67	1.51	0.01	0.03	0.63	0.43	0.87	3.08
P10Bw1	4.33	0.54	0.02	0.05	0.60	0.20	0.06	2.42
P10Bw2	1.99	1.11	0.05	0.03	0.73	0.03	1.13	1.87
P11 A	20.30	2.34	0.06	0.13	0.95	0.15	0.66	6.27
P11AB	11.62	1.48	0.05	0.05	0.37	0.08	1.10	5.61
P11BA	6.65	0.60	0.04	0.02	0.22	0.08	1.16	2.81
P11Bw1	5.96	1.07	0.04	0.02	0.18	0.13	1.16	4.13
P11Bw2	5.61	0.40	0.04	0.01	0.18	0.15	1.18	4.40
P11Bt1	3.91	0.45	0.04	0.01	0.22	0.13	1.25	3.08
P11Bt2	17.30	0.71	0.06	0.01	0.23	1.22	1.30	3.96
P12 A	15.67	7.05	0.14	0.65	1.68	1.38	0.08	3.25
P12AB	9.05	2.59	0.11	0.39	0.35	0.95	0.35	4.51
P12BA	7.07	1.90	0.15	0.23	0.32	0.88	0.31	2.97
P12Bw1	5.19	1.10	0.17	0.12	0.40	0.78	0.53	3.19
P12Bw2	3.27	0.63	0.20	0.08	0.30	0.73	0.21	2.37

Tabela 3.4. Dados químicos das amostras de calibração utilizados para análise dos solos do Estado da Bahia – parte 2

Amostra	C.O. g.kg ⁻¹	P mg.kg ⁻³	Na cmol.kg ⁻¹	K	Ca	Mg	Al	H + Al
P12Bw3	2.10	0.58	0.17	0.07	0.30	0.90	0.25	2.81
P12Bw4	2.14	0.85	0.23	0.09	0.27	0.95	0.11	2.26
P14A	22.90	2.16	0.07	0.18	0.17	0.50	1.43	7.98
P15A	11.62	1.26	0.03	0.06	0.47	0.13	0.97	4.68
P15AB	8.21	0.93	0.02	0.03	0.33	0.27	0.87	3.47
P15Bi1	11.05	0.57	0.02	0.09	0.40	0.00	0.77	3.30
P15-2B	8.67	0.76	0.03	0.01	0.20	0.10	0.57	3.08
P15-2C	4.35	0.62	0.03	0.01	0.20	0.10	0.20	1.98
P16A	30.17	2.64	0.06	0.15	0.40	0.00	1.37	9.90
P16Bi	27.34	1.61	0.06	0.12	0.30	0.00	1.30	8.80
P16Cr	17.87	1.06	0.08	0.07	0.20	0.10	1.23	5.72
P17 A	13.66	2.93	0.05	0.27	1.40	0.67	0.00	4.46
P17AB	7.78	2.08	0.08	0.11	0.83	0.30	0.30	3.85
P17BA	7.11	1.14	0.08	0.10	0.67	0.13	0.80	3.52
P17Bt	4.19	0.00	0.10	0.05	0.53	0.17	0.97	4.90
P18A	19.79	4.56	0.14	0.50	2.33	1.30	0.00	2.75
P18AB	11.05	2.38	0.10	0.36	1.07	0.70	0.10	2.92
P18BA	6.45	1.97	0.08	0.37	0.93	0.37	0.10	1.73
P18Bi	2.50	1.67	0.08	0.36	0.90	0.50	0.00	2.20
P18BC	1.65	1.47	0.08	0.21	0.80	0.53	0.00	1.21
P19 A	8.48	0.60	0.08	0.15	0.70	0.50	1.40	3.20
P19AB	7.03	0.00	0.08	0.12	0.40	1.70	2.30	3.90
P19B/A(A)	4.96	0.00	0.08	0.12	0.30	0.60	2.65	3.80
P19B/A (B)	6.58	0.00	0.08	0.13	0.30	0.70	3.55	4.20
P19B1	6.58	0.00	0.08	0.13	0.60	0.30	3.10	5.20
P20ACgnz	8.91	29.20	43.60	0.10	0.90	6.60	0.20	0.00
P20Cgnz	7.67	22.80	42.96	0.11	1.00	5.20	0.35	0.30
P20-2Cgnz	8.89	23.50	54.28	0.10	1.00	8.10	0.00	1.30
P20-3Cgnz	15.31	30.60	57.91	0.11	1.50	7.90	0.60	0.00
P21Bt3	2.32	1.00	0.01	0.01	0.06	0.04	2.40	3.63
P22A	31.32	3.00	0.06	0.13	0.60	0.97	0.80	9.13
P22AB	16.24	1.00	0.03	0.04	0.00	0.08	0.70	5.83
P22BA	8.12	1.00	0.02	0.02	0.00	0.08	0.30	3.30
P22Bw1	6.38	1.00	0.03	0.02	0.00	0.07	0.20	2.97
P22Bw2	4.64	1.00	0.03	0.01	0.20	0.08	0.20	2.86
P22Bw3	4.06	1.00	0.03	0.00	0.30	0.09	0.10	2.09
23A	5.22	1.00	0.06	0.08	0.18	0.30	2.70	8.69
P23AB	4.64	1.00	0.04	0.02	0.19	0.29	3.00	5.61
P23Bt1	4.64	1.00	0.03	0.02	0.10	0.28	4.50	7.15
P23Bt2	4.06	1.00	0.04	0.03	0.08	0.28	5.60	7.92
P23Bt3	2.32	1.00	0.03	0.03	0.08	0.28	6.60	10.12
P23Bt4	2.32	1.00	0.04	0.07	0.09	0.18	8.70	11.66
P24A1	36.54	2.00	0.12	0.05	0.82	0.56	1.10	12.10
P24A2	26.68	2.00	0.09	0.04	0.56	0.44	1.20	10.23

Tabela 3.4. Dados químicos amostras calibração – parte 3

Amostra	C.O. g.kg ⁻¹	P mg.kg ⁻³	Na cmol.kg ⁻¹	K	Ca	Mg	Al	H + Al
P24AE	4.64	2.00	0.03	0.01	0.11	0.19	0.20	1.76
P24E	9.28	1.00	0.05	0.01	0.14	0.18	0.30	3.30
P24EB	29.00	1.23	0.05	0.01	0.13	0.31	2.80	17.05
P24Bhg	59.16	2.00	0.07	0.01	0.23	0.41	4.70	29.81
P24Bhgx	73.67	2.00	0.07	0.01	0.42	0.52	4.80	30.91
P25 A	23.00	4.85	0.03	0.13	0.70	1.25	0.45	6.44
P25AB	19.00	2.23	0.04	0.13	0.45	0.30	0.85	5.78
P25BA	31.00	1.47	0.04	0.07	0.40	0.45	0.80	3.14
P25B1	14.00	1.18	0.04	0.06	0.20	0.10	0.55	3.30
P25B2	10.00	0.54	0.06	0.14	0.40	0.80	0.25	2.48
P25B3	5.00	0.69	0.06	0.16	0.30	0.70	0.15	1.98
P26 A1	32.00	2.23	0.05	0.09	1.20	0.40	1.50	5.94
P26A2	23.00	2.83	0.02	0.07	1.45	0.15	0.15	3.47
P26C	30.00	2.87	0.02	0.07	1.20	0.00	0.20	3.14
P27 A	23.00	5.31	0.03	0.13	0.70	1.25	0.45	9.74
P27AB	19.00	2.29	0.04	0.13	0.45	0.30	0.85	5.78
P276BA	31.00	1.03	0.04	0.07	0.40	0.45	0.80	3.47
P27Bi	14.00	0.69	0.04	0.06	0.20	0.10	0.55	2.48
P28A1	21.00	4.33	0.10	0.09	0.50	1.10	0.20	8.25
P28A2	28.85	4.65	0.04	0.04	0.25	0.75	0.60	12.38
P28-2A1	26.74	3.87	0.03	0.03	0.10	0.40	1.50	20.96
P28-2A2	21.00	4.07	0.03	0.03	0.15	0.35	1.40	14.19
P28-3AE	10.13	9.97	0.07	0.04	0.20	0.95	0.70	7.43
P28-3EB	10.13	8.34	0.07	0.02	0.15	0.33	0.70	5.78
P28BSX	14.96	8.77	0.11	0.02	0.00	0.00	0.00	12.38
P31 Ac	25.00	3.00	0.12	0.08	1.02	0.71	1.47	16.92
P31Bwc	16.00	2.00	0.07	0.15	0.47	0.17	1.19	11.99
P31Bfc2	5.00	5.00	0.04	0.02	0.08	0.07	0.26	4.39
P31Bfc3	4.00	4.00	0.05	0.07	0.09	0.06	0.20	3.75
P32A1	1.82	7.44	0.05	0.11	0.40	0.25	0.65	6.11
P32A2	1.61	8.01	0.05	0.08	0.25	0.55	0.75	5.78
P32AB	0.66	4.60	0.06	0.09	0.20	1.10	0.95	5.12
P32BT1	0.34	2.87	0.09	0.14	0.10	0.80	0.85	3.47
P32BT2	0.33	2.17	0.69	0.18	0.15	1.05	0.75	2.15
	0.33	2.18	0.09	0.12	0.03	1.05	1.05	2.64

Para avaliação de desempenho do modelo de calibração multivariada foi calculado o coeficiente de determinação (R^2) (eq. 3.4), a raiz quadrado do erro médio (RMSE) (eq. 3.5) e o desvio preditivo residual (RPD) (eq. 3.6). Esses valores expressam a exatidão do modelo, ou seja, representa quão próximo o valor calculado pelo modelo está do valor obtido por outro método de análise (Geladi; Kowalski, 1986).

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3.4)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3.5)$$

$$RPD = \frac{sd}{RMSE} \quad (3.6)$$

Onde: \hat{y} é o valor predito, y é o valor observado e \bar{y} é a média dos valores observados, n é o número de amostras com i variando de 1 a n

Para utilização do R^2 na avaliação dos modelos de calibração foram utilizados os critérios estabelecidos por Sayes et al. (2005), onde ele define quatro classificações para esses valores, a) Variações de R^2 entre 0,5 a 0,65 indicam que o modelo permite a discriminação entre baixas e altas concentrações; b) Valores de R^2 entre 0,66 e 0,80 indicam modelos aceitáveis; c) Valores variando de 0,81 a 0,9 indicam bons modelos; e d) Para valores superiores a 0,90 indicam modelos com capacidade de predição excelente.

A classificação adotadas para utilização do RPD está de acordo com Viscarra et al. (2006b) onde: a) Valores inferiores a 1 indicam modelos não adequados a predição; b) RPD entre 1,0 e 1,4 dão indicação de que o modelo pode ser usado para predição, mais que é possível apenas para valores muito diferentes; c) Variações de RPD entre 1,4 e 2,0 indicam bom modelo, com previsões justas, que podem ser usadas para avaliação e correlação; d) RPD entre 2,0 e 2,5 indicam modelo com previsões quantitativas muito boas; e e) Para modelos com RPD acima de 2,5 a indicação de previsões excelentes.

As amostras escolhidas para o grupo de validação seguiram o critério de representatividade do universo amostral de acordo com suas classes de solo, garantindo que a representatividade de cada classe em ambos os grupos sejam as mesmas. As amostras de cada grupo, bem como suas respectivas características químicas estão descritas nas tabelas 3.3 e 3.4.

Os espectros gerados em refletância foram transformados em absorbância ($\log_{10} \frac{1}{R}$) para facilitar a identificação das feições de absorção nos espectros. A elaboração do modelo de calibração multivariada consiste em relacionar dois grupos de dados, onde X corresponde a matriz dos dados experimentais, sendo as amostras representadas em linhas. As variáveis, neste caso os comprimentos de onda são apresentados nas colunas. Enquanto que Y é a matriz composta pelas variáveis

estudadas, que neste caso são as análises químicas de fertilidade. O objetivo da calibração multivariada PLS é relacionar as propriedades de interesse com as muitas respostas espectrais, construindo um modelo $Y = f(X)$, que permita explicar a variação de Y em função dos valores observados para X.

Antes da construção do modelo de calibração multivariada empregando o algoritmo PLS, os espectros foram submetidos a pré-processamentos que envolveram a suavização de valores espectrais através da variação normal padronizada (Standard Normal Variate – SNV) e os filtros ou alisamentos (Smoothing) empregando as técnicas da derivada 1ª de Savitzky-Golay e do filtro Mediano (Median Filter). A escolha dessas técnicas de pré-processamento teve como base estudos anteriores que revelaram bons resultados para modelos de calibração multivariada associados a estes tratamentos como demonstram os estudos realizados por Maghimi et al. (2010); Liu et al. (2019); Zhao et al. (2021); Rodrigues (2015); entre outros. O software The Unscrambler X versão 10.4 foi utilizado para o pré-processamento e para os métodos de calibração.

Os pré-processamentos foram empregados em todos espectros do grupo de amostras de calibração. A variação normal padronizada (Standard Normal Variate – SNV) foi usada para corrigir o efeito da dispersão da luz. O alisamento através da derivada 1ª de Savitzky-Golay teve como parâmetros adotados polinômio de segunda ordem com janela de 7 pontos de suavização tendo sido aplicado para suavizar os espectros, já o filtro mediano (Median filter) foi empregado com o mesmo propósito de suavização, tendo sido adotado com parâmetro para sua utilização o tamanho de janela de 3 pontos.

O modelo de calibração empregado foi o PLS, com a utilização máxima de componentes igual a 7, identificando ou outliers e com o emprego da média dos dados centrais, a validação foi cruzada, o algoritmo empregado para correlacionar os dados de atributos químicos e valores espectrais foi Kernel PLS. O modelo de calibração foi empregado individualmente para cada um dos elementos em estudo.

4. RESULTADOS E DISCUSSÃO

4.1 Características da Distribuição das amostras

Os solos são sistemas dinâmicos e complexos, com características e composições distintas. A resposta do espectro eletromagnético para cada um dos seus constituintes depende da sua concentração na amostra. A obtenção de modelos de predição de atributos do solo por meio da espectroscopia Vis-NIR requer a utilização de banco de dados com uma boa representatividade dos diversos solos que se deseja avaliar (Sayes et al., 2008).

As características do universo amostral utilizado neste estudo estão representados na Tabela 4.1. Os valores da assimetria, medida da distorção da distribuição de probabilidade de uma variável aleatória sobre sua média; e curtose, medida do efeito combinado da distribuição dos dados na cauda, indicam que o banco de dados utilizado no estudo não apresenta a distribuição normal. Na distribuição normal, o valor da assimetria se aproxima de zero, devido a equivalência da distribuição dos dados para a direita e para esquerda do ponto central. As características dos atributos avaliados no estudo, Tabela 4.1, indicam que H+Al e K são os parâmetros com menor assimetria e Na, Ca e P são os atributos com maior assimetria.

O valor da curtose é normalmente comparado ao da distribuição normal, que é igual a 3. O atributo pode ser considerado com a cauda pesada ou leve se o valor da curtose for maior ou menor que 3, respectivamente. Existe uma boa relação entre os valores de assimetria e curtose dos atributos avaliados. Com a tendência geral dos atributos apresentarem maiores valores acima do ponto central (Ginzburg, 2018).

Tabela 4.1 – Estatística descritiva das amostras utilizadas no estudo para as análises químicas convencionais.

Atributos	Unidades	Média	Assimetria	curtose	Median a	Desvio padrão
C.O.	g kg ⁻¹	15,89	2,37	5,30	7.84	20.40
P	mg kg ⁻¹	6,20	4,86	24,57	2.00	14.28
Na	Cmol _c kg ⁻¹	1,66	5,64	30,90	0.06	9.37
K		0.11	2,01	4,84	0.07	0.94
Ca		1,04	5,14	27,34	0.40	2.30
Mg		1.16	3,75	14,58	0.32	2.38
Al		0.98	2,94	10,94	0.63	1.24
H+Al		5.92	1,99	4,22	3.61	5.69

A reflectância de 154 amostras foi adquirida por meio do Vis/NIR na região entre 350nm e 2500nm. Para a construção do modelo de calibração as amostras do estudo foram divididas em dois grupos: calibração e validação. O agrupamento das amostras foi feito de forma preservar a representatividade da distribuição das características dos atributos químicos avaliados, gráficos 4.1 e 4.2. A sobreposição das linhas de reflectância em um dado comprimento de onda indica que os materiais pertencem a uma mesma classe de materiais com características gerais (composição química, mineralógica e outras) semelhantes porem presente, por exemplo em diferentes concentrações. (Rodrigues, 2018)

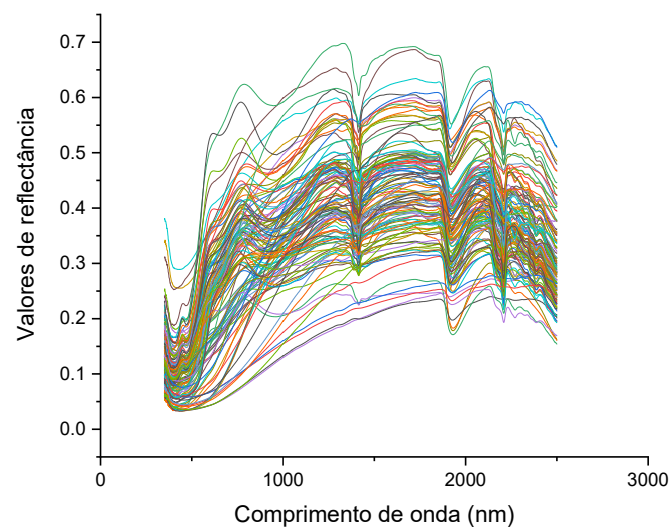


Figura 4.1 – Espectros de reflectância das amostras de calibração dos solos do Estado da Bahia utilizados no estudo.

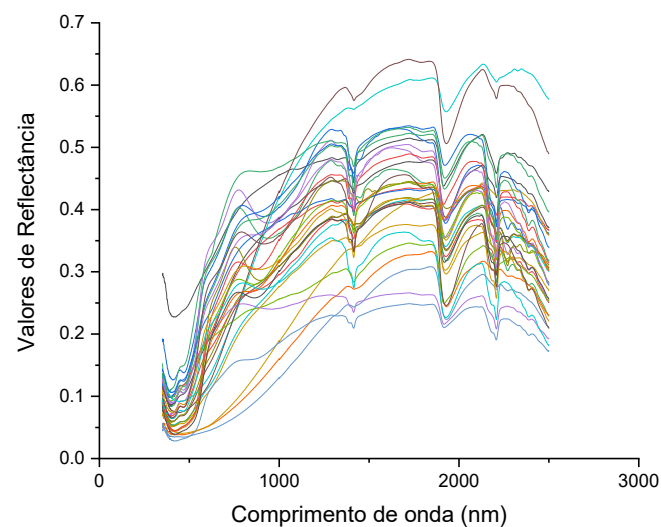


Figura 4.2 – Espectros de reflectância das amostras de validação dos solos do Estado da Bahia utilizados no estudo.

Para que os sinais analíticos dos espectros de cada constituinte do solo possa ser relacionado com uma propriedade química específica, como a concentração de determinado elemento, demanda a utilização de recursos matemáticos. O emprego direto desta técnica aos espectros Vis/NIR sem nenhum tratamento torna o modelo de calibração pouco efetivo (Sayes, 2005)

4.2 Pré-processamentos

Os dados obtidos em reflectância foram inicialmente transformados em absorbância que apresenta uma relação mais direta com a capacidade intrínseca dos materiais em absorver radiações eletromagnéticas em frequências específicas (Hollas, 2004). A absorbância tem sido usada frequentemente em trabalhos de geologia e na ciências do solo com o objetivo de determinar os constituintes de minerais de rochas e solos, uma vez que, os espectros de absorção possuem relação direta com a concentração da maioria dos elementos químicos (Zornoza, 2008).

Pré-processamentos são técnicas usualmente empregadas aos dados espectrais para remover informações irrelevantes, como ruído, incertezas, variabilidade e interações o que torna os modelos de calibração mais confiáveis, precisos e estáveis. A combinação de diferentes métodos de pré-processamento tende a tornar mais eficiente a remoção dos ruídos (Cen & He, 2007). Os novos valores espectrais obtidos a partir da aplicação das técnicas podem ser conferidos nas figuras 4.3; 4.4; 4.5; 4.6; e 4.7.

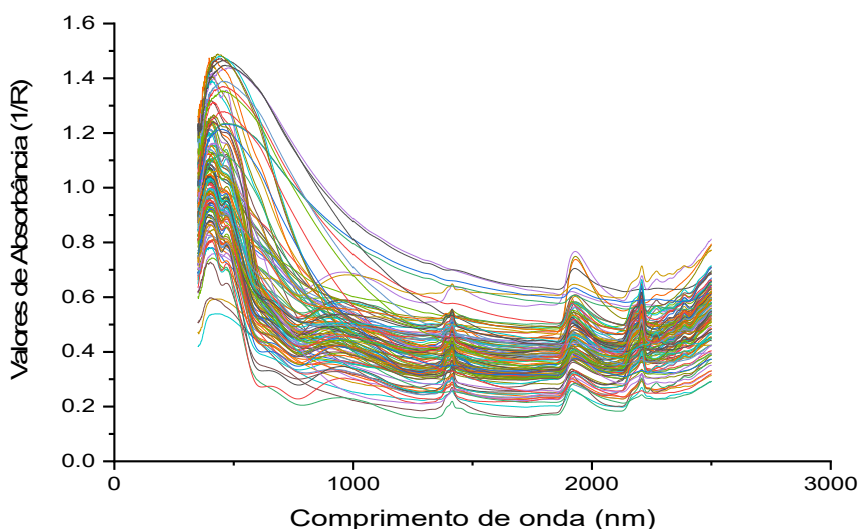


Figura 4.3 – Espectros das amostras em estudo convertidos em absorbância.

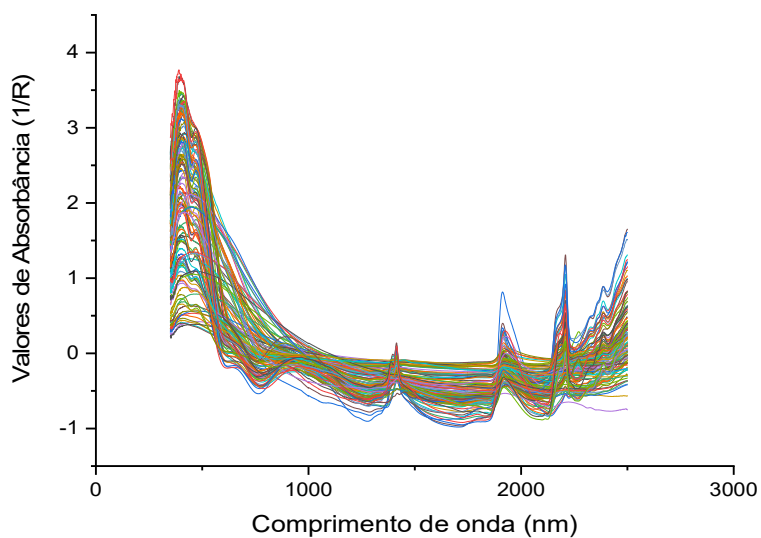


Figura 4.4 – Espectros das amostras de calibração pré-processadas com SNV.

A comparação dos espectros de absorvância sem tratamento (figura 4.3) com os espectros tratados com SNV (figura 4.4) permite observar a centralização dos dados próximos a absorvância zero, o que minimiza os efeitos de espalhamentos da luz e interferências causadas pelo tamanho de partículas.

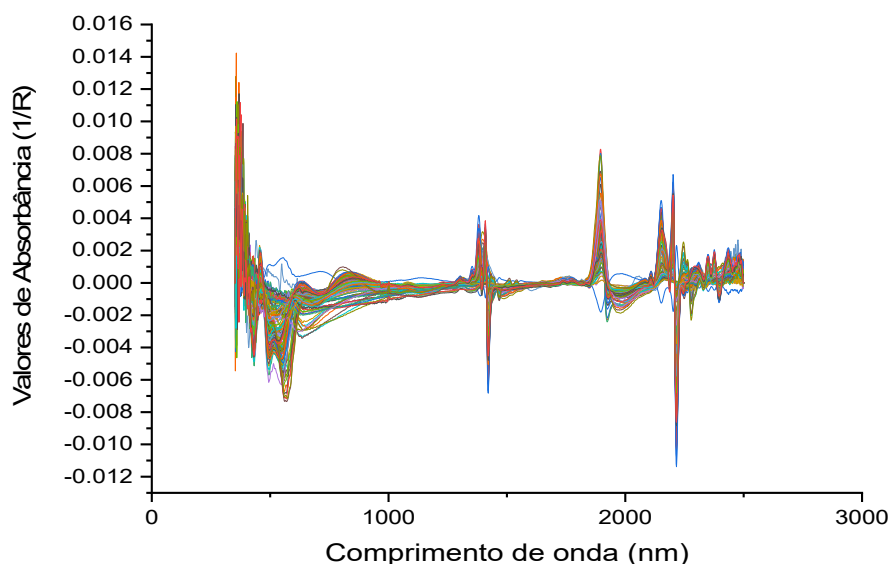


Figura 4.5 – Espectros das amostras de calibração pré-processadas com a derivada 1ª de Savitzky-Golay.

A aplicação da derivada 1ª de Savitzky-Golay modificou a forma dos espectros (figura 4.6). A aplicação da derivada gera ruídos, diminuindo a relação sinal/ruído, daí a

necessidade de aplicação de um filtro que permita diminuir esse efeito. A combinação das técnicas de derivação e alisamento foi proposto por Savitzky-Golay, baseando-se em alisamento com aproximação polinomial por mínimos quadrados (Luo et al., 2005).

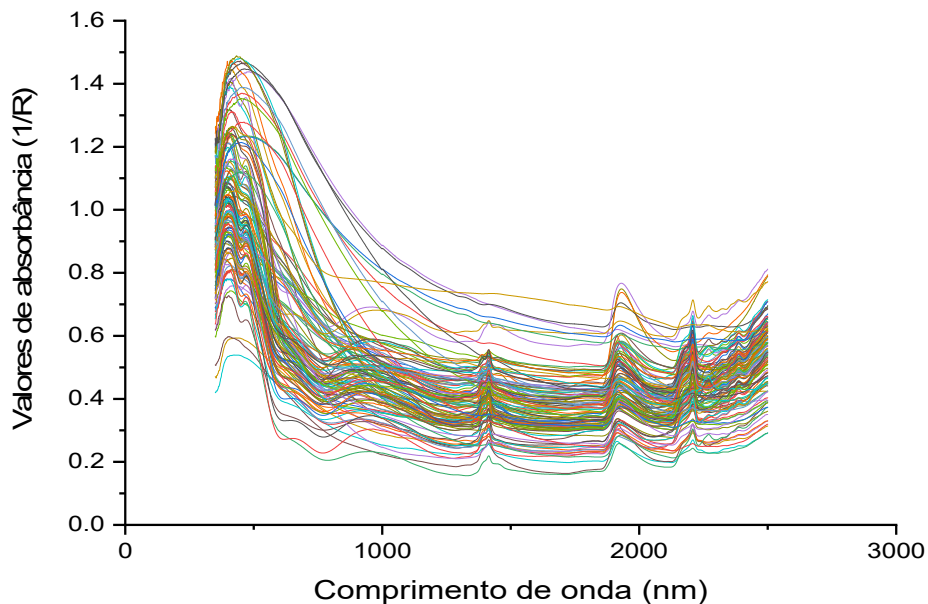


Figura 4.6 – Espectros das amostras de calibração pré-processadas com o filtro Mediano.

O emprego do filtro mediano aos espectros é muito similar ao proposto por Savitzky-Golay, mas sem a utilização da derivada, ele é aplicado para aumentar a relação sinal ruído e por isso as modificações nos espectros são mínimas (figura 4.6).

As combinações entre pré-tratamentos SNV mais a derivada 1ª de Savitzky-Golay (figura 4.7) e SNV mais o filtro mediano 4.8) foram empregadas com o propósito de corrigir eventuais falhas individuais das técnicas no processamento dos dados espectrais.

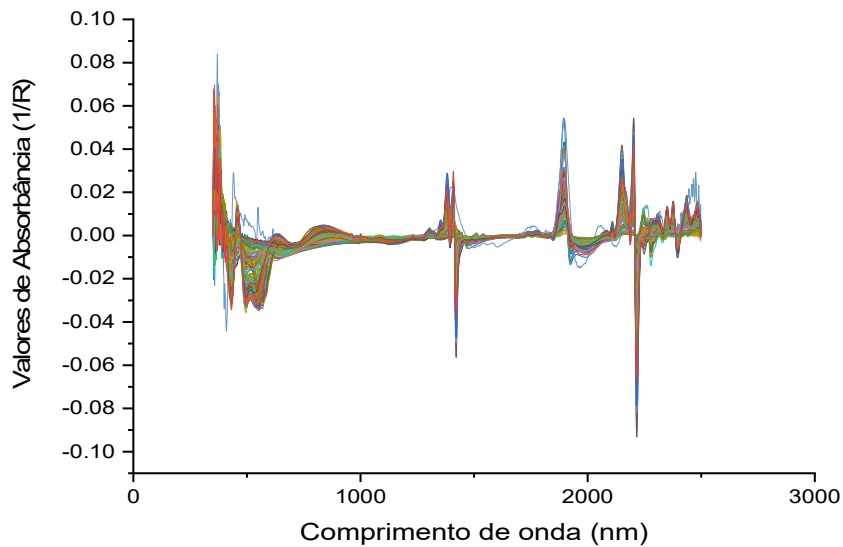


Figura 4.7 – Espectros das amostras de calibração pré-processadas com SNV mais a derivada 1ª de Savitzky-Golay.

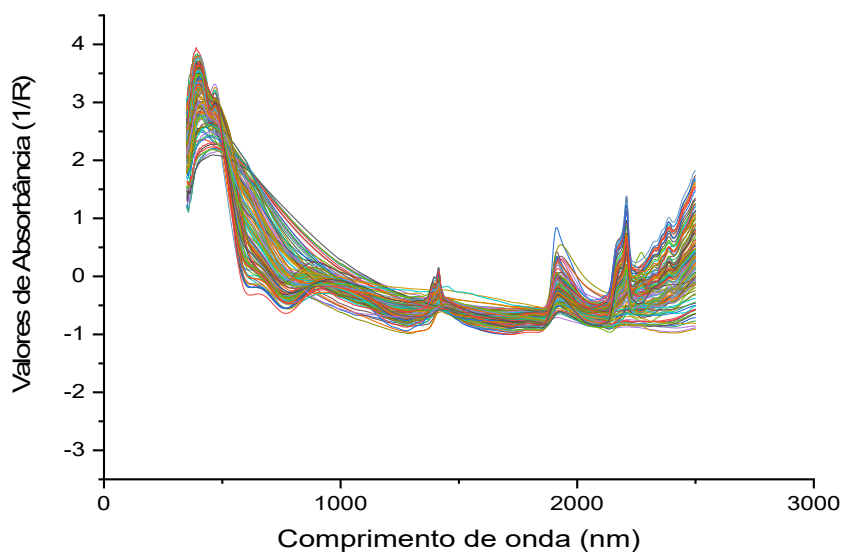


Figura 4.8 – Espectros das amostras de calibração pré-processadas com SNV mais o filtro Mediano.

4.3 Capacidade preditiva dos modelos de regressão

A capacidade preditiva do teor de alumínio pelo PLS não foi boa, tabela 4.1. Os baixos valores de correlação indicam que os pré-processamentos não foram efetivos relações efetivas entre os dados espectrais e os valores da concentração de Al, impossibilitando em alguns casos calcular o R^2 . Os valores de RPD indicam também que o modelo não pode ser usado na predição do Al. Zornosa et al., (2008) atribuiu o

fato a ausência de resposta primária do Al, P, Ca, Mg e K nas regiões do visível e do infravermelho próximo e por esta razão não podem ser preditos com confiança pela espectroscopia.

Tabela 4.1 – Valores dos erros de calibração (RMSEC) e de validação (RMSEP), coeficientes de determinação (R^2) para calibração e validação, e desvio preditivo residual (RPD), associados aos pré-tratamentos aplicados na regressão PLS para Al.

Atributo	Pré-tratamento	Calibração		Validação Interna		Validação Externa			RPD
		R^2	RMSEC	R^2	RMSEP	R^2	RMSEP	Correl.	
Al	SNV	0,06	1,31	0	1,4	0	1,0	-0,31	0,37
	Savitzky Golay	0,3	1,25	0,2	1,3	0,3	1,21	0,5	0,40
	Filtro Mediano	0,22	1,22	0,2	1,3	0	1,31	-0,14	0,55
	SNV + S. Golay	0,21	1,20	0,1	1,4	0	1,32	-0,49	0,9
	SNV + F. Mediano	0,1	1,3	0	1,4	0	1,0	-0,31	0,37
C.O.	SNV	0,6	13,00	0,5	15,1	0,2	20	0,38	0,53
	Savitzky Golay	0,85	8,00	0,69	11,60	0,85	8,03	0,92	1,02
	Filtro Mediano	0,68	11,60	0,61	13,00	0,40	16,02	0,68	0,73
	SNV + S. Golay	0,85	5,0	0,84	5,2	0,81	6,72	0,92	1,74
	SNV + F. Mediano	0,61	13,00	0,51	14,65	0,20	20,36	0,38	0,55
P	SNV	0,2	13,9	0,2	14,5	0,8	7,8	0,3	0,81
	Savitzky Golay	0,7	7,5	0,6	9,6	0,2	5,9	0,5	0,96
	Filtro Mediano	0,5	5,6	0,4	6,4	0,2	5,3	0,4	1,02
	SNV + S. Golay	0,8	7,6	0,6	10	0,2	7,5	0,4	0,86
	SNV + F. Mediano	0,3	13,0	0,2	14,0	0,1	7,8	0,3	0,91
K	SNV	0,02	0,11	0	0,11	0,02	2,10	0,13	0,01
	Savitzky Golay	0,20	0,10	0,13	0,11	0	2,11	-0,19	0,02
	Filtro Mediano	0,19	0,10	0,09	0,11	0,10	2,07	0,09	0,04
	SNV + S. Golay	0,14	0,11	0,07	0,11	0	2,11	-0,44	0,01
	SNV + F. Mediano	0,26	0,11	0,06	0,11	0,18	0,11	0,16	0,02
Na	SNV	0,54	6,02	0,35	7,42	0,52	8,23	0,76	0,78
	Savitzky Golay	0,63	5,37	0,28	7,64	0,52	8,32	0,72	0,90
	Filtro Mediano	0,38	6,98	0,21	7,97	0,42	9,57	0,65	0,51
	SNV + S. Golay	0,78	4,14	0,56	5,86	0,76	5,7	0,90	1,41
	SNV + F. Mediano	0,69	4,89	0,47	6,68	0,67	6,81	0,87	1,05
Ca	SNV	0,20	2,33	0,12	2,48	0	1,32	-0,35	0,61
	Savitzky Golay	0,83	1,06	0,78	1,23	0,48	0,5	0,81	1,73
	Filtro Mediano	0,54	1,76	0,41	2,02	0	1,58	-0,05	0,87
	SNV + S. Golay	0,84	1,05	0,68	1,50	0,97	0,33	0,57	1,2
	SNV + F. Mediano	0,67	0,75	0,13	1,24	0	0,67	-0,25	0,71
Mg	SNV	0,74	1,30	0,40	1,97	0,24	1,31	0,60	1,10
	Savitzky Golay	0,87	0,93	0,80	1,15	0,43	1,14	0,66	0,96
	Filtro Mediano	0,62	1,60	0,47	1,90	0,26	1,73	0,40	1,02
	SNV + S. Golay	0,89	0,82	0,84	1,03	0,71	0,980	0,85	1,63
	SNV + F. Mediano	0,74	1,31	0,49	1,86	0,29	1,31	0,60	0,23
H+Al	SNV	0,18	5,47	0,14	5,65	0,32	4,14	0,57	0,47
	Savitzky Golay	0,63	3,66	0,48	4,39	0,44	3,58	0,74	0,96
	Filtro Mediano	0,70	3,29	0,58	3,93	0,31	3,98	0,62	0,87
	SNV + S. Golay	0,67	3,46	0,49	4,37	0,28	4,14	0,57	0,47
	SNV + F. Mediano	0,18	5,47	0,14	5,67	0,26	4,14	0,57	0,47

Os coeficientes de determinação para carbono orgânico nas amostras de calibração para os diferentes pré-tratamento variou de 0,60 (SNV e SNV+ F. mediano) a 0,85 (Savitzky Golay e SNV + S. Golay). Os pré-tratamentos com maiores valores de R^2 de calibração também apresentaram maiores valores de R^2 para validação interna das amostras (0,69 e 0,84, respectivamente). Os Valores de RPD calculados para C.O. (tabela 4.1), pelos critérios adotados, indicam que o modelo construído associando os pré-tratamentos SNV e Savitzky-Golay apresentam capacidade para realizarem boas previsões, tendo sido estabelecido correlações de 0,92 (tabela 4.1) entre o sinal analítico e os valores de C.O., apresentando RPD de 2,3 e R^2 de 0,81 na validação externa, o que também indica ser um bom modelo. Para este caso a combinação de técnicas de pré-processamento melhoraram o desempenho do modelo de calibração uma vez que as técnicas utilizadas individualmente não alcançaram o mesmo valor de RPD. Estes valores estão de acordo com o descrito em outros trabalhos, como Rodrigues (2015) que obteve um RPD de 2,13 e um R^2 0,85, na validação externa. Já Santos et al. (2010) obteve valores para R^2 que variaram de 0,83 a 0,87. Os valores de RMSEP (erro de predição) maiores que os de RMSEC (erro de calibração) indicando que o modelo tende a superestimar os resultados (Parisotto, 2009).

Estas características estão de acordo com o esperado para o C.O. pois apresenta resposta primária, gerando influência ao longo de toda a curva espectral. Dalmolin (2002) relatou o aumento da reflectância dos solos ao longo de todo o espectro com o decréscimo da matéria orgânica. Para Demattê et al. (2004a) a predição da matéria orgânica do solo pode ser realizada em diferentes bandas do espectro óptico, sendo fixadas de acordo com o sensor óptico disponível.

A correlação estabelecida para o P não foi satisfatória na elaboração do modelo de calibração. Evidenciado pelos valores de RPD para todos os tratamentos empregados (tabela 4.1), sendo inferiores a 1, o que de acordo com a classificação de Viscarra (2006) indica que o modelo não é adequado a predição. Fato corroborado pelos valores de R^2 inferiores a 0,5. Estes valores foram observados os trabalhos de Dunn et al. (2002) que encontrou R^2 de 0,18 e RPD igual a 1,1, Silva (2018) também encontrou valores R^2 igual a 0,20 e RPD de 1,1. Estas observações podem estar associadas as respostas espectroscópicas, como o P não possui resposta primaria na região NIR, tais sinais são dependentes da relação do fósforo com a matéria orgânica e com minerais primários.

Para que sejam estabelecidas as relações entre os teores de P e os espectros Vis-NIR, é necessário também que haja correlação de P com outros atributos do solo, porque quando o método Vis-NIR é utilizado em combinação com técnicas multivariadas de análise dos dados espectrais, ela tem potencial para determinar qualitativa ou quantitativamente a composição química de uma amostra de solo (Xiaobo et al., 2010), Por isso a caracterização e predição do fósforo utilizando-se a espectroscopia Vis-NIR é possível, devido às correlações que esse elemento apresenta com o carbono, hidrogênio ou oxigênio, principalmente, nas formas de P=O, P=O ligado a hidrogênio, P-fenil, P-H, P-OH e fosfatos PO₃ e PO₄ ou C-H, C=OH, C-O, P-O e P=O presentes no fosfato monoéster ou diéster (Wu et al., 2010; Pasquini, 2003).

Estas correlações são específicas, o que por vezes, impede a extrapolação dos resultados para outros solos, as relações dos espectros com os atributos secundários, como o P, não é precisa e as interações mais específicas entre estes parâmetros precisam ser melhores entendidas (Wu et al., 2010).

Embora os valores de RPD para K, (tabela 4.1), calculados a partir do emprego do SNV, do filtro mediano e da combinação de SNV com filtro mediano, apresentem valores iguais ou superiores a 1 o que pela classificação adotada reflete em um modelo com capacidade de distinção apenas de valores muito diferentes, a sua utilização, mesmo para esse fim não é indicado uma vez que a avaliação do R², para todos os casos foi inferior a 0,3, demonstra que o modelo construído não atende aos padrões de referência. Além do mais os valores de RMSEP muito superiores aos de RMSEC indicam tendência de superestimação dos valores. Os modelos construídos com os pré-processamentos SNV e Filtro mediano tiveram RPD igual a 1, no entanto quando estas técnicas foram combinadas o RPD foi elevado a 1,15, evidenciando que a combinação de pré-processamentos podem melhorar a capacidade preditiva do PLS.

Os resultados obtidos são equiparáveis aos descritos por Silva (2018) com R² 0,30 e RPD de 0,99 utilizando os mesmos pré-tratamentos. Também observado por Dann et al. (2002) com valores máximo de 0,68 e 1,4 para R² e RPD respectivamente. Wartini et al. (2020) encontrou valores de R² variando entre 0,19 e 0,56 onde os autores associam esses valores a baixa interação do potássio com a matéria orgânica, fato essencial para predição de atributos que não possuem resposta direta na faixa do infravermelho.

O modelo construído para o Na apresentou excelentes correlações entre os sinais analíticos e as concentrações deste elemento, variando de 0,65 a 0,90 (tabela

4.1), com RPD variando de 1,25 até 2,1, para o modelo construído com a combinação dos algoritmos SNV e Savitzky-Golay. Estes resultados permitem a classificação deste método para previsões quantitativas muito boas, com R^2 igual a 0,76 indicando também que o modelo é aceitável, apresentando também valores RMSEC e RMSEP próximos mais com indicação da tendência para previsão de valores acima dos reais. Vale destacar ainda que o modelo de calibração utilizando os algoritmos SNV e filtro mediano combinados, também apresentaram bons resultados com RPD igual a 1,76 e R^2 de 0,67 indicando que o modelo é aceitável e possui a capacidade de fazer previsões justas.

Wartini et al. (2020) ao utilizar as espectroscopia Vis-NIR para previsão de atributos químicos do solo encontrou valores de R^2 variando entre 0,24 a 0,45, valores baixos quando comparados a este estudo. Já Silva (2018) encontrou valores de R^2 de 0,6 e RPD de 1,55, embora tenha utilizado como forma de entrada dos dados espectrais a reflectância para os mesmos tratamentos espectrais empregados neste estudo. Como já evidenciado anteriormente os valores de absorbância tendem a apresentar melhores resultados na predição de atributos químicos do solo, justificando a melhora encontrada nos resultados deste estudo.

Valores mais próximos dos encontrados neste trabalho foram evidenciados por Dunn et al. (2002) com R^2 variando de 0,69 a 0,85 e RPD de 1,7 a 2,3. No entanto, os autores relacionaram os resultado dos modelos de predição a seleção das amostras, Chang et al. (2001) ao utilizar um conjunto de 439 amostras com teores de Na variando entre 0,1 a 1,8 com média de 0,2 cmol/kg não obteve sucesso é predizer este atributo com o Vis-NIR, segundo o descrito pelos autores o conjunto de dados possuía uma significativa distorção na distribuição, apontando para baixas concentrações fazendo com que os sinais analíticos resultantes fossem facilmente confundidos com o ruído e acabavam sendo descartado durante o processo. Desta forma, os resultados observados neste estudo possuem uma maior variabilidade dos teores de Na somando-se ao fato de que os pré-processamentos podem ter sido mais efetivos em removerem apenas o ruído permitindo que o modelo de regressão conseguisse estabelecer uma correlação, com valores de até 0,90, entre os sinais analíticos e as concentrações de Na.

A derivada primeira de Savitzky-Golay com RPD igual a 1,4 (tabela 4.1) foi o melhor pré-tratamento utilizados para a construção do método PLS para avaliação dos teores de Ca. Segundo a classificação empregada neste estudo este valor de RPD está no limiar entre as classificações de modelo adequado para distinção de valores altos e

baixos, a modelo com previsões justas. Baseado na classificação para R^2 o modelo é bom para discriminação de altas e baixas concentrações, já os valores de RMSEC e RMSEP não revelam tendência de superestimação. Valores similares foram observados por Wartini et al. (2020) com valores de R^2 que variaram entre 0,71 e 0,81; Franceschini et al. (2013) com valores de R^2 entre 0,42 e 0,77 e RPD de 1,3 e 1,4;

Os valores de RPD para Mg variaram de 1,12 a 1,88 (tabela 4.1), apresentando o melhor valor, 1,88, para aplicação combinada dos algoritmos SNV e primeira derivada de Savitzky-Golay, este valor indica que o modelo é bom e pode ser empregado em previsões quantitativas com boa correlação, o valor de R^2 igual a 0,77 para validação externa reforça o RPD com classificação de modelo aceitável, os valores de RMSEC e RMSEP indicam leve tendência a previsão de valores superestimados. Dunn et al. (2002) obteve valores de R^2 igual a 0,85 e RPD de 2,7 para validação, já Saeys et al. (2005) obteve valores de R^2 entre 0,80 e 0,86 e RPD de 1,84 a 2,65.

Os modelos PLS empregados na avaliação dos valores de H + Al tiveram RPD variando de 1,16 a 1,33 para todos os pré-tratamentos utilizados (tabela 4.1), o que revela modelos com capacidade de distinguir valores altos e baixos, entre eles se destacou o modelo que empregou a derivada primeira de Savitzky-Golay com RPD de 1,33 e valores do RMSEC e RMSEP indicando uma leve tendência de elevação dos valores de predição, no entanto os modelos apresentaram valores de R^2 inferiores a 0,44, com valores mínimos de 0,26, demandando a associação com outros critérios para avaliação dos teores de H + Al. Os valores encontrados aqui estão condizentes com os relatados por Franceschini et al. (2013) com valores de R^2 e RPD 0,47 e 1,4 respectivamente e por Silva (2018) com R^2 igual a 0,49 e RPD 1,31.

Como descrito anteriormente, os valores dos erros mostram que os modelos testados não são capazes de realizar a predição para alumínio. Esses índices sofreram uma leve melhoria ao ser avaliados os teores de H + Al, isto pode estar associado ao fato de o hidrogênio apresentar maior interação com a matéria orgânica em diferentes faixas gerando respostas espectrais mais significativas. Isso faz sentido porque a espectroscopia NIR fornece informações sobre as proporções relativas de ligações, como C – H, N – H, S – H e O – H, presentes nos compostos orgânicos (O – H também incluído na molécula de água) (Zornoza et al., 2008).

A construção dos modelos de calibração multivariada para previsão de atributos químicos do solo neste estudo podem ser divididas em três grupos. O primeiro diz respeito aos modelos que não demonstraram efetividade em seu propósito, independente

das técnicas de pré-processamento empregadas, neste grupo estão os modelos aplicados aos elementos Al e P, Os valores de alumínio estão de acordo com outros estudos como Rodrigues (2018), já os valores de P estão abaixo dos apresentados em outros trabalhos como Felix et al. (2016) e Oliveira et al. (2015). No entanto estes valores baixos podem estar associados a forte influência da umidade do solo e dos minerais hidratados, além da interação com a matéria orgânica, por possuírem resposta espectral nas mesmas faixas (Vendrame et al., 2012).

Embora as análises tenha sido feitas em amostras de solo seco ao ar, ainda restam água adsorvida nas áreas superficiais dos minerais argilosos e da matéria orgânica, em equilíbrio com o vapor de água na atmosfera. Resultados semelhantes também foram observados em outros estudos (Zornoza et al., 2008; Chang et al., 2001; Dunn et al., 2002).

No segundo grupo pode-se inserir os modelos estabelecidos para K, Ca e H + Al, com valores de RPD para os modelos de calibração multivariada que remetem a sua utilização apenas para predição da existência destes elementos no solo pela distinção entre valores altos e baixos. Entretanto os valores de RPD encontrados neste estudo estão em desacordo com os valores experimentados em outros trabalhos. Enquanto que para K, Askari et al. (2016), encontram valores que variaram de 1,9 a 2,3, e Tekin et al. (2016) com valor de 2,03 o melhor valor encontrado neste estudo foi de 1,15. Já para o Ca Segundo Araújo et al. (2015), os valores de RPD foram estimados em 2,1, neste estudo os valores chegaram a 1,4.

A falta de eficiência dos modelos em predizer as concentrações dos elementos P, Ca, K e Al, pode estar associada a forma de interação destes materiais com os espectros NIR, não ocorrendo de forma direta, necessitando da interação destes elementos com minerais primário e com a matéria orgânica para que haja resposta no espectro. Assim, a capacidade do NIR-PLS em predizer P, K, Ca e Mg, está relacionada à forte dependência dessas propriedades com a matéria orgânica e argilas, que possuem uma resposta primária na região do NIR. Essas propriedades são controladas principalmente pelo tipo e conteúdo de argila e matéria orgânica, que apresentam grupos funcionais com cargas variáveis responsáveis pela adsorção dos diferentes cátions e água (Chang et al., 2001).

Ainda para Demattê et al. (2017) os baixos teores de determinados elementos no solo dificultam a sua estimativa através dos modelos de calibração. Uma vez que, o conjunto de amostras de calibração nos quais predominam amostras que contribuam fracamente com o sinal analítico, podem ser confundidas com sinais indesejados e

acabem não sendo utilizados pelo modelo de regressão. Fato esse que pode estar associado a baixa eficiência dos modelos estabelecidos no estudo para a estimativa dos teores dos elementos K, Ca e H + Al. A partir do banco de dados estatísticos do universo amostral utilizado no estudo (tabela 4.1), embora a variabilidade dos teores destes elementos sejam altas existem a predominância significativa dos valores de baixa concentração para estes elementos, o que possivelmente dificultou a construção de modelos de calibração eficientes.

No terceiro grupo estão os modelos com RPD superior a 1,8, sendo que o Mg apresentou RPD igual a 1,88 e R^2 de 0,71, valores que estão de acordo com outros estudos como os realizados por Demattê et al. (2017) com RDP variando entre 1,9 e 2,3 e de R^2 variando de 0,63 a 0,66, e os de Tekin et al. (2016) com valores de RDP entre 1,56 e 2,2 e valores de R^2 entre 0,60 e 0,79. Os valores de RDP e R^2 para Na são 2,1 e 0,76 respectivamente, a exemplo dos estudos de Wartini et al. 2020; Silva 2018; Dunn et al. 2002 e observando os critérios estabelecidos por Viscarra et al. (2006b) e por Sayes et al. (2005) o modelo pode ser classificado como aceitável por possuir a capacidade de previsões quantitativas muito boas. Para o C.O. o RPD de 2,3 e o R^2 de 0,81 indicam que o modelo é bom com capacidades preditivas quantitativas muito boas, como também observado por Rodrigues (2015) com valores de RDP variando de 1,4 a 2,3 e R^2 de 0,76 a 0,85, para Santos (2011) os valores de R^2 alcançaram 0,84.

5. CONCLUSÃO

No geral os modelos de calibração multivariada PLS empregados a espectroscopia Vis-NIR apresentaram bons resultados, mesmo sendo utilizado um banco de dados de muita variabilidade e tendo sido avaliado um número grande de elementos, principalmente para a combinação das técnicas SNV e primeira derivada de Savitzky-Golay, evidenciadas principalmente por esta última, demonstrando que a associação de técnicas de pré-processamento podem resultar em modelos mais eficientes.

As técnicas empregadas para avaliação dos solos do Estado da Bahia mostraram-se mais eficazes na avaliação dos elementos C.O., Na e Mg, inclusive para previsões quantitativas. Para avaliação K, Ca e H + Al os modelos permitiram apenas avaliações qualitativas. Já para Al e P não foi possível estabelecer boas correlações.

Outra alternativa que pode potencializar os resultados é o emprego de outros algoritmos de pré-processamento como MSC, segunda derivada, centralização na média, contínuo removido, entre outros, além de combinações entre diferentes técnicas, associados a outros algoritmos como o NIPALS e o PCR na construção dos modelos de calibração multivariada, incorporando ainda outras faixas espectrais pouco estudadas como o MIR e o FIR. Podendo inclusive ser associada a difração de raio X (DRX).

REFERÊNCIAS BIBLIOGRÁFICAS

ADAMCHUK, V.I.; HUMMEL, J.W.; MORGAN, M.T.; UPADHYAYA, S.K. On-the-go soil sensors for precision agriculture. *Computers and Electronics in Agriculture*, vol.44, p.71–91, 2004.

ALMEIDA, F. M. N. Wspwctrosopia de infravermelho próximo com transformada de Fourier (FT-NIR) na caracterização de farinhas para alimentação pueril. 2009. Dissertação em engenharia biológica – Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa.

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.

ASKARI, M.S.; CUI, J.; O'ROURKE, S.M.; HOLDEN, N.M. Evaluation of soil structural quality using VIS–NIR spectra. *Soil & Tillage Research*, v.146, n.1, p.108-117, 2015.

BORGES NETO, W.; MELLO, C.; POPPI, R. J. Adulteração de óleo de oliva por espectroscopia raman e calibração multivariada. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.

BROWN, D.J.; SHEPHERD, K.D.; WALSH, M.G.; MAYS, M.D.; REINSCH, T.G. Global soil characterization with VNIR diffuse reflectance spectroscopy. *Geoderma*, v.132, p. 273-290, 2006.

BURNS, D. A.; CIURCZAK, E. W. *Handbook of Near Infrared Analysis*. 3ª. ed. Parkway – USA: CRC Press, 2007.

CAVALCANTE, E.G.S.; ALVES, M.C.; SOUZA, Z.M.; PEREIRA, G.T. Variabilidade espacial de atributos químicos do solo sob diferentes usos e manejos. *Revista Brasileira de Ciência do Solo*, v.31, p.1329-1339, 2007.

CEN, H.; HE, Y. Theory and application of near infrared reflectance spectroscopy in determination of food quality. *Trends in Food Science e Techonology*. v.18, n. 2, p. 72-83, 2007.

CIENFUEGOS, Freddy; VAITSMAN, Delmo. *Análise Instrumental*. 1.ed. Rio de Janeiro: Editora Interciência, 2000.

CORÁ, J. E.; ARAUJO, A. V.; PEREIRA, G. T.; BERALDO, J. M. G. Variabilidade espacial de atributos do solo para adoção do sistema de agricultura de precisão na cultura de cana-de-açúcar. *Revista Brasileira de Ciência do Solo*, Viçosa, v.28, n.6, p.1013-1021, 2004.

DALMOLIN, R. S. D.; GONÇALVES, C. N.; KLAMT, E.; DICK, D. P. Relações entre os contaminantes do solo e seu comportamento espectral. *Ciência Rural*, Santa Maria, V. 35, n. 2, p. 481-489, 2005.

DALMOLIN, R.S.D. *Matéria orgânica e características físicas, químicas, mineralógicas e espectrais de Latossolos de diferentes ambientes*. 2002. 151f. Tese (Doutorado em

Ciência do Solo) - Programa de Pós-graduação em Ciência do Solo, Universidade Federal do Rio Grande do Sul.

DEMATTÊ, J.A.M. et al. Effect of fermentation residue on the spectral reflectance properties of soil. *Geoderma*, Amsterdam, v.120, p.187-200, 2004a.

DEMATTÊ, J.A.M.; L. RAMIREZ-LOPEZ, K.P.P. MARQUES, A.A. RODELLA. Chemometric soil analysis on the determination of specific bands for the detection of magnesium and potassium by spectroscopy, *Geoderma*, v.288, n.1, p. 8-22, 2017.

DIAS, Y. G. Espectroscopia NIR e calibração multivariada aplicadas ao controle de qualidade de gases combustíveis naturais e derivados do petróleo. 2011. Dissertação (Mestrado em química) – Instituto de química, Universidade Estadual de Campinas, Campinas.

DUNN, B. W.; et al. The potencial of near-infrared reflectance spectroscopy for soil analysis – a case study from the Riverine Plain of South-eastern Australia. *Australian Journal of Experimental Agriculture*, V. 42, p. 607-614, 2002.

EITELWEIN, M. T. Sensoramento proximal de solo para a quantificação de atributos químicos e físicos. 2017. Tese (Doutorado em Engenharia de Ciências Agrícolas) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

EMBRAPA (EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA). Manual de análises químicas de solos, plantas e fertilizantes. 2.ed. Rio de Janeiro. Centro Nacional de Pesquisa de Solos. 627p. 2009.

FELIX, J. C. Predição de fósforo, carbono e nitrogênio em solos de basalto, por meio da espectroscopia NIR. *Pesquisa agropecuária brasileira*, v.51, n.9, p.1405-1416, 2016.

FERRARINI, H. Determinação de teores nutricionais do milho por espectroscopia no infravermelho e calibração multivariada. 2004. Dissertação (Mestrado em Química) – Universidade do Paraná, Curitiba.

FERREIRA, M. M.C. QUMIOMETRIA – Conceitos, Métodos e Aplicações. 1 ed. Campinas. Unicamp, 2015.

FERREIRA, M. M.C. Qumiometria – Conceitos, Métodos e Aplicações. 1 ed. Campinas. Unicamp, 2015.

FERRUCIO, C. E. Espectroscopia Vis-NIR para predição de atributos químicos do solo. 2016, Dissertação (Mestrado em agronomia), Universidade Estadual de Londrina, Londrina.

FRANCESCHINI, M. H. D.; Sensoriamento remoto hiperespectral nos níveis laboratório, campo aéreo como ferramentas auxiliares no manejo do solo. 2013. Dissertação (Mestrado em Solos e Nutrição de Plantas) – Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba.

GELADI, P.; KOWALSKI, B. R. Partial Least-Squares Regression: A tutorial. *Analytical Chimica Acta*, v. 185, p. 1-17, 1986.

GINZBURG, I. Determination of the diffusivity, dispersion, skewness and kurtosis in heterogeneous porous flow. Part II: Lattice Boltzmann schemes with implicit interface. *Advances in Water Resources*, v. 118, p. 49-82, 2018.

GREGO, C.R.; VIEIRA, S.R. Variabilidade espacial de propriedades físicas do solo em uma parcela experimental. *Revista Brasileira de Ciência do Solo*, v.29, p.169-177, 2005.

HOLLAS, J. M. Eletromagnetic radiation and its interaction with atoms and molecules. In: *Modern spectroscopy*. Chichester: John Wiley, 2004. P. 27-39.

LIU, Y. et al. The influence of spectral pretreatment on the selection of representative calibration samples for soil organic matter estimation using Vis-NIR reflectance spectroscopy. *Remote Sensing*, v.11, n.450, 2019.

LÓPEZ, Leonardo Ramírez. *Pedologia quantitativa: espectrometria VIS-NIR-SWIR e mapeamento digital de solos*. 2009.

LUO, J.; YING, K.; HE, P.; BAI, J. Properties of Savitzky–Golay digital differentiators. *Digit. Signal Process.* 15, 122–136, 2005.

MAGALHÃES, W. L. E.; ZANGISKI, F.; KLEIN, C. H.; HIGA, A. R. Uso da espectroscopia no infravermelho próximo (NIR) para predição não-destrutiva de densidade básica da madeira de *Pinus taeda*. Colombo: Embrapa Florestas, 2006. 4 p. (Embrapa Florestas. Comunicado técnico, 159).

MAZUR, L. Aplicação de metodologia por NIR para determinação METILXANTINAS PRESENTES NA ERVA-MATE (*Ilex paraguariensis*). 2012. Dissertação de mestrado, Programa de Pós- Graduação em Engenharia de Alimentos – UFPR, Curitiba.

MCMURRY, John. *Química Orgânica*. São Paulo: Pioneira Thomson Learning, 2005. 2v

MOGHIMI, A.; AGHKHANI, M. H.; SAZGARNIA, A.; SARMED, M. Vis/NIR spectroscopy and chemometrics for the prediction of soluble solids content and acidity (pH) of kiwifruit. *Biosystems Engineering*, v.106, p. 295-302, 2010.

MONTEZANO, Z.F.; CORAZZA, E.J.; MURAOKA, T. Variabilidade espacial da fertilidade do solo em área cultivada e manejada homoganeamente. *Revista Brasileira de Ciência do Solo*, vol.30, p.839-847, 2006.

MURRAY, I.; WILLIAMS, P. C. In *Near-Infrared Technology in the Agricultural and Food Industries*; Williams, P. C.; Norris, K. H., eds.; American Association of Cereal Chemists: St. Paul, MN, 1987, chap. 2.

NAES, T., ISAKSSON, T., FEARN, T., DAVIES, T. *A User-Friendly Guide to Multivariate Calibration and Classification*. NIR Publications. 2002.

OLIVEIRA, J.F. de; BROSSARD, M.; CORAZZA, E.J.; MARCHÃO, R.L.; VENDRAME, P.R.S.; BRITO, O.R.; GUIMARÃES, M. de F. VIS-NIR spectrometry, soil phosphate

extraction methods and interactions of soil attributes. *Química Nova*, v.38, p.342-350, 2015.

PARISSOTO, G. et al. Desenvolvimento de metodologias no infravermelho aplicadas para a quantificação de cloridrato de metformina em formulações farmacêuticas. *Orbital*, v.1, n.1, p. 38-48, 2009.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the brazilian chemical society*, São Paulo, v.14, n.2, p. 407-409, mar/abr. 2003.

PASQUINI, C. Princípios da espectroscopia no infravermelho próximo. In: TIBOLA, C. S.; Medeiros, M. L. F.; OLIVEIRA, M. A. *Espectroscopia no infravermelho próximo para avaliar indicadores de qualidade tecnologia e contaminantes em grãos*. Brasília: Embrapa, 2018. p. 13-31.

RODRIGUES, G. Seleção de bandas espectrais apoiada pela metaheurística PSO para predição de alumínio trocável de amostras de solo. 2018. Dissertação (Mestrado em computação aplicada), Universidade Federal de Ponta Grossa, Ponta Grossa.

RODRIGUES, M. J. M. Avaliação do potencial de espectroscopia no infravermelho próximo na quantificação de carbono em solos do estado do Acre. 2015. Dissertação – (Mestrado em Ciências, Inovação e Tecnologia para a Amazônia), Univesidade Federal do Acre, Rio Branco.

SALIBA, P. R. Desenvolvimento de metodologias analíticas utilizando espectroscopia visível e infravermelho próximo: classificação de açúcares, determinação de brix e acidez total em suco de laranja. 1999. Tese (doutorado em química) Universidade Estadual de Campinas, São Paulo.

SANTOS, Ana Paula dos. *Espectroscopia de infravermelho próximo em análises de solos e plantas*. p. 1–38, 2011.

SANTOS, H.G. dos; CARVALHO JUNIOR, W. de; DART, R. de O.; AGLIO, M.L.D.; SOUSA, J.S. de; PARES, J.G.; FONTANA, A.; MARTINS, A.L. da S.; OLIVEIRA, A.P. de. *O novo mapa de solos do Brasil: legenda atualizada*. Rio de Janeiro: Embrapa Solos, 2011. 67p. (Embrapa Solos. Documentos, 130).

SAYES, W.; MOUAZEN, A.M.; RAMON, H. Potential for onsite and online analysis of pig manure using visible and near infrared reflectance spectroscopy. *Biosystems Engineering*, v.91, n.4, p.393–402, 2005.

SCHRADER, B. *Infrared and Ramam Spectroscopy: methods and applications*. Weinheim: VHC, 1995.

SENA, M. M.; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI< P. J> Avaliação do uso de métodos quimiométricos em análise de solos. *Revista Química Nova*, São Paulo, V.23, p. 547-556, 2000.

SHEPHERD, Keith D.; WALSH, Markus G. Development of reflectance spectral libraries for characterization of soil properties. *Soil Science Society of America Journal*, v. 66, n. 3, p. 988–998, 2002.

SILVA, H. G. et al. Digestibility of nutrients in diets containing ammoniated elephant grass and cocoa meal or palm kernel cake fed to sheep. *Rev. Bras. Zootec.*, 36 (2): 499-506, 2007.

SILVA, R. G. P. O. Espectroscopia de reflectância no mapeamento e modelagem de solos no semiárido. 2018, Dissertação (Mestrado em ciências do solo), - Departamento de Ciências do Solo, Universidade Federal do Ceará, Fortaleza.

SINGH, Kanika; MAJEED, Israr; PANIGRAHI, Niranjan; et al. Catena Near infrared diffuse reflectance spectroscopy for rapid and comprehensive soil condition assessment in smallholder cacao farming systems of Papua New Guinea. *Catena*, v. 183, n. May, p. 104185, 2019. Disponível em: <<https://doi.org/10.1016/j.catena.2019.104185>>.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. Princípios de análise instrumental. 5. Ed. São Paulo: Bookman, 2002.

SOARES, Valdeir Moraes. Uso de sensores vis-nir para predição de atributos químicos do solo. 2017. Tese, pós-graduação em Agricultura Tropical, da Universidade Federal de Mato Grosso, Cuiabá.

SOLOMONS, T. W. G. FRUHLE, C. B. Química Orgânica, 7ª edição, LTC – Livros Técnicos e Científicos Editora S. A. Rio de Janeiro, 2001.

SORIANO-DISLA, J. M.; JANIK, L. J.; VISCARRA ROSSEL, R. a; MACDONALD, L. M.; MCLAUGHLIN, M. J. The Performance of Visible, Near-, and Mid-Infrared Reflectance Spectroscopy for Prediction of Soil Physical, Chemical, and Biological Properties. *Applied Spectroscopy Reviews*, v. 49, n. 2, p. 139–186, 2014.

SOUSA, J. S. Desenvolvimento de Modelos de Calibração Multivariada em Espectroscopia de Infravermelho Próximo para Ácidos Graxos em Amostras de Carne Bovina. 2016. Dissertação (Química analítica) – Centro de ciências exatas e de tecnologia departamento de química, Universidade Federal de São Carlos, São Carlos.

SOUZA, L. M., SANTANA, F. B., GONTIJO, L. C., MAZIVILA, S. J., BORGES NETO, W. Quantification of adulterations in extra virgin flexseed oil using MIR and PLS. *Food Chemistry*, v. 182, p. 35-40, 2015.

TEKIN, Y.; TÜMSAVAŞ, Z.; ULUSOY, Y.; MOUAZEN A.M. On-line VisNIR sensor determination of soil variations of sodium, potassium and magnesium. *IOP Conference Series: Earth and Environmental Science*, v.41, n.1, p.1-11, 2016.

VALDERRAMA, P. Avaliação de figura de mérito em calibração multivariada na determinação de controle de qualidade em indústria alcooleira por espectroscopia de infravermelho próximo. 2005. Dissertação (Mestrado em Química Analítica) – Universidade Estadual de Campinas, Campinas.

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

VENDRAME, P.R.S.; MARCHÃO, R.L.; BRUNET, D.; BECQUER, T. The potential of NIR spectroscopy to predict soil texture and mineralogy in Cerrado Latosols. *European Journal of Soil Science*, v.63, p.743-753, 2012.

VISCARRA-ROSSEL, R.A.; McGLYNN, R.N.; McBRATNEY, A.B. Determining the composition of mineral-organic mixes using UV-VISNIR diffuse reflectance spectroscopy. *Geoderma*, v.137, n.1, p.70-82, 2006b.

WARTINI, Ng. Et al. Developing a soil spectral library using a low-cost NIR spectrometer for precision fertilization in Indonesia. Elsevier – *Geoderma Regional*, n.22, 2020.

WILLIAMS, P. How it all began. In: INTERNATIONAL CONFERENCE ON NEAR INFRARED SPECTROSCOPY, 17. 2015, Foz do Iguassu. Proceedings... Campinas: International Council for Near Infrared Spectroscopy, 2015. p. 128-132. Disponível em: Acesso em: 20 fev. 2020.

WU, C. Y.; JACOBSON, A. R.; LABA, M.; KIM, B.; BAVEYE, P. C.; *Water Air Soil Pollut.* 2010, 209, 377.

XIAOBO, Z.; JIEWEN, Z.; POVEY, M. J. W.; HOLMES, M.; HANPIN, M.; *Anal. Chim. Acta* 2010, 667, 14.

ZANÃO JÚNIOR, L.A.; LANA, R.M.Q.; GUIMARÃES, E.C. Variabilidade espacial do pH, teores de matéria orgânica e micronutrientes em profundidades de amostragem num Latossolo Vermelho sob semeadura direta. *Ciência Rural*, v.37, n.4, p.1000-1007, 2007.

ZHANG, Z. et al. Strategies for the eficiente estimation of soil organic matter in salt-affected soils through Vis-NIR spectroscopy: Optimal band combination algorithm and spectral degradation. Elsevier – *Geoderma*, n. 382, 2021.

ZHAO, D.; ARSHAD, M.; TRIANTAFILIS, J. Predicting soil physical and chemical properties using Vis-NIR in australian cotton áreas. *Catena*, v.196, 2021.

ZORNOZA, R. et al. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in mediterranean soils. *Soil Biology e Biochemistry*, v.40, p1923-1930, 2008.