

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL  
FACULDADE DE AGRONOMIA - FAGRO  
PROGRAMA DE PÓS-GRADUAÇÃO EM FITOTECNIA – PPG-FITO

ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO PARA ESTIMAR A  
CONCENTRAÇÃO DE NITROGÊNIO EM PLANTAS DE ARROZ, MILHO, SOJA E  
TRIGO

Diego Lima Lucchesi  
Engenheiro Agrônomo/UFRGS

Dissertação apresentada como um dos requisitos  
à obtenção do Grau de Mestre em Fitotecnia  
Área de concentração Sistemas de Produção Vegetal

Porto Alegre (RS), Brasil  
Dezembro de 2023

### CIP - Catalogação na Publicação

Lucchesi, Diego Lima  
ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO PARA  
ESTIMATIVA DA CONCENTRAÇÃO DE NITROGÊNIO EM PLANTAS E  
GRÃOS NAS CULTURAS DO ARROZ, MILHO, SOJA E TRIGO /  
Diego Lima Lucchesi. -- 2024.  
52 f.  
Orientador: André Luis Vian.

Coorientador: Tales Tiecher.

Dissertação (Mestrado) -- Universidade Federal do  
Rio Grande do Sul, Faculdade de Agronomia, Programa de  
Pós-Graduação em Fitotecnia, Porto Alegre, BR-RS,  
2024.

1. Espectroscopia. 2. Nitrogênio. I. Vian, André  
Luis, orient. II. Tiecher, Tales, coorient. III.  
Título.

DIEGO LIMA LUCCHESI  
Engenheiro Agrônomo - UFRGS

## **DISSERTAÇÃO**

Submetida como parte dos requisitos  
para obtenção do Grau de

### **MESTRE EM FITOTECNIA**

Programa de Pós-Graduação em Fitotecnia  
Faculdade de Agronomia  
Universidade Federal do Rio Grande do Sul  
Porto Alegre (RS), Brasil

Aprovado em: (17/11/2023)  
Pela Banca Examinadora

ANDRÉ LUIS VIAN  
Orientador  
UFRGS

TALES TIECHER  
Coorientador  
UFRGS

SIMONE MUNDSTOCK JAHNKE  
Coordenadora do Programa de  
Pós-Graduação em Fitotecnia

CHRISTIAN BREDEMEIER  
UFRGS

LEANDRO SOUZA DA SILVA  
UFSM

CARLOS ALBERTO BISSANI  
Diretor da Faculdade de  
Agronomia

JEAN MICHEL MOURA BUENO  
UNICRUZ

# ESPECTROSCOPIA DE INFRAVERMELHO PRÓXIMO PARA ESTIMATIVA DA CONCENTRAÇÃO DE NITROGÊNIO EM PLANTAS E GRÃOS NAS CULTURAS DO ARROZ, MILHO, SOJA E TRIGO<sup>1</sup>

**Autor:** Diego Lima Lucchesi

**Orientador:** André Luis Vian

**Coorientador:** Tales Tiecher

## RESUMO

Os métodos químicos tradicionalmente empregados para a estimativa do teor de nitrogênio em materiais vegetais e grãos apresentam elevado custo, são operacionalmente ineficientes e necessitam o uso de reagentes químicos, elevando os riscos tanto para o laboratorista quanto para o meio ambiente. A espectroscopia de infravermelho próximo - NIR (*Near Infrared Spectroscopy*) é um método que possibilita a análise de forma mais eficiente, com menos custos e não geração de resíduos tóxicos. Para que esta técnica apresente boa acurácia e precisão, é necessário a criação de um modelo de calibração multivariada que associe os dados espectrais aos resultados obtidos nas análises químicas. O presente trabalho tem como objetivo a criação de modelos amplos de predição da concentração de nitrogênio em plantas e grãos de milho, arroz, trigo e soja com o uso de diferentes combinações de técnicas de pré-processamento e calibração multivariada. Para o presente estudo foram coletadas e analisadas um total de 899 amostras de folhas e 502 amostras de grãos. Foram utilizados os pré-processamentos *Detrend (DET)*, *Standard Normal Variate (SNV)* e *Savitzky-Golay Derivative (SGD)*. Os modelos de predição foram desenvolvidos a partir dos seguintes métodos de calibração multivariada: *Partial Least Squares Regression (PLSR)* e *Support Vector Machine (SVM)*. A calibração do modelo foi feita através da validação cruzada por meio de uma divisão aleatória em segmentos. Os valores preditos de nitrogênio através da espectroscopia foram comparados aos valores mensurados pelo método químico, sendo a acurácia da espectroscopia para este fim avaliada através dos seguintes parâmetros estatísticos: coeficiente de determinação ( $R^2$ ), relação entre o desempenho e a distância interquartil (RPIQ) e o erro médio quadrático de predição (RMSE). O uso da espectroscopia de infravermelho próximo foi capaz de prever com precisão as concentrações de nitrogênio em tecido vegetal, sendo que os melhores resultados foram encontrados com o uso do método de calibração SVM e pré-processamento SNV ( $R^2 = 0,99$ ). Adicionalmente, foi verificado que é possível a construção de modelos amplos de predição para os conjuntos de amostras com grãos de trigo e arroz, e folhas de trigo, arroz, milho e soja.

---

<sup>1</sup>Dissertação de Mestrado em Fitotecnia, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brasil. (52f.) Janeiro, 2024.

# NEAR INFRARED SPECTROSCOPY TO ESTIMATE NITROGEN CONCENTRATION IN PLANTS AND GRAINS IN RICE, CORN, SOYBEAN AND WHEAT CROPS<sup>1</sup>

**Author:** Diego Lima Lucchesi

**Adviser:** André Luis Vian

**Coadviser:** Tales Tiecher

## ABSTRACT

The chemical methods traditionally used to estimate the nitrogen content in plant materials and grains are expensive, operationally inefficient and require the use of chemical reagents, increasing risks for both the laboratory worker and the environment. Near Infrared Spectroscopy (NIR) is a method that enables analysis more efficiently, with less costs and does not generate toxic waste. For this technique to present good accuracy and precision, it is necessary to create a multivariate calibration model that associates the spectral data with the results obtained in chemical analyses. The present work aims to create broad models for predicting nitrogen concentration in leaves and grains of corn, rice, wheat and soybeans using different combinations of pre-processing and multivariate calibration techniques. For the present study, a total of 899 leaf samples and 502 grain samples were collected and analyzed. Detrend (DET), Standard Normal Variate (SNV) and Savitzky-Golay Derivative (SGD) pre-processing were used. The prediction models were developed based on the following multivariate calibration methods: Partial Least Squares Regression (PLSR) and Support Vector Machine (SVM). The model calibration was done through cross-validation through a random division into segments. The predicted nitrogen values through spectroscopy were compared to the values measured by the chemical method, with the accuracy of the spectroscopy for this purpose being evaluated through the following statistical parameters: coefficient of determination ( $R^2$ ), relationship between performance and interquartile distance (RPIQ) and the mean squared prediction error (RMSE). The use of near-infrared spectroscopy was able to accurately predict nitrogen concentrations in plant tissue, with the best results being found using the SVM calibration method and SNV pre-processing ( $R^2 = 0.99$ ). Additionally, it was verified that it is possible to build broad prediction models for sets of samples with wheat and rice grains, and wheat, rice, corn and soybean leaves.

---

<sup>2</sup>Master Dissertation in Plant Science, Faculdade de Agronomia, Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, Brazil. (52p.) January, 2024.

## SUMÁRIO

	Página
1 INTRODUÇÃO.....	1
2 REVISÃO BIBLIOGRÁFICA.....	4
2.1 Espectroscopia de reflectância na agricultura.....	4
2.2 Princípio da espectroscopia - NIR.....	5
2.3. Uso da espectroscopia de reflectância em plantas.....	6
2.4 Análise multivariada dos dados.....	8
2.5 Pré-processamentos espectrais.....	9
2.5.1 Detrend (DET).....	9
2.5.2 Savitski-Golay Derivative (SGD).....	9
2.5.3 Standard Normal Variate (SNV).....	10
2.6 Principal Component Analysis (PCA).....	10
2.7 Calibração Multivariada.....	11
2.7.1 Partial Least Squares Regression (PLSR).....	12
2.7.2 Support Vector Machine (SVM).....	12
2.8 Biblioteca Espectral.....	12
2.9 Referências.....	14
3 ARTIGO 1.....	17
3.1 Introdução.....	19
3.2 Material e métodos.....	21
3.2.1 Banco de amostras.....	21
3.2.2 Análises químicas.....	23
3.2.3 Análise de espectroscopia de infravermelho próximo (NIR).....	23
3.2.4 Pré-processamentos espectrais.....	24
3.2.5 Análise Discriminante.....	24
3.2.6 Análise dos componentes principais (PCA).....	24

	Página
3.2.7 Calibração multivariada.....	25
3.3 Resultados e discussão.....	27
3.3.1 Análise descritiva do conjunto de dados.....	27
3.3.2 Análise dos espectros de infravermelho próximo (NIR).....	29
3.3.3. Análise discriminante	30
3.3.4 Análise de componentes principais (PCA).....	30
3.3.5 Importância das bandas espectrais para predição de N	31
3.3.6 Construção dos modelos de calibração.....	33
3.3.6.1 Plantas.....	33
3.3.6.2. Grãos.....	35
3.4 Conclusão.....	37
3.5 Referências.....	38
4 CONSIDERAÇÕES FINAIS.....	43

## RELAÇÃO DE TABELAS

	Página
REVISÃO BIBLIOGRÁFICA	
1. Metabólitos das plantas, seus respectivos comprimentos de onda detectadas por espectroscopia na faixa de NIR e os grupos funcionais relacionados.....	7
2. Resumo dos estudos científicos que utilizaram a espectroscopia para estimar a concentração de N em plantas.....	8
ARTIGO 1	
1. Caracterização do banco de amostras de tecido vegetal em relação aos locais e as doses de nitrogênio.....	22
2. Estatística descritiva da concentração de N nas amostras de planta e grão.....	28
4. Parâmetros de acurácia dos modelos PLSR e SVM para estimativa de nitrogênio nos modelos construídos com amostras de planta para as culturas de arroz, milho, trigo e soja.....	33
5. Parâmetros de acurácia dos modelos PLSR e SVM para estimativa de nitrogênio nos modelos construídos com amostras de grãos de arroz, trigo e soja.....	35



## RELAÇÃO DE FIGURAS

REVISÃO BIBLIOGRÁFICA	Página
1. Espectro eletromagnético e suas regiões espectrais. A região do infravermelho é dividida em três partes: Infravermelho próximo (NIR), Infravermelho médio (MIR) e Infravermelho distante (FIR).....	5
2. Assinatura espectral típica de tecidos vegetais nas bandas espectrais do visível (VIS) e do infravermelho próximo (NIR).....	6
ARTIGO 1	
1. Amostras acondicionadas na Biblioteca Espectral de Tecido Vegetal (BETV). Sacos plásticos com laço de vedação e identificação das amostras (A e B); Amostras armazenadas em caixas vedadas com controle de umidade (C).....	23
2. Esquema das etapas realizadas para a calibração das análises multivariadas para a estimativa dos teores de N no tecido vegetal.....	26
3. Histograma dos teores de N das amostras de grão (trigo, soja e arroz) e planta (trigo, milho e arroz).....	28
4. Assinatura espectral média das amostras de planta e grão sem pré-processamento (RAW) (A), pré-processada com Detrend (DET) (B), Savitski-Golay Derivative (SGD) (C) e Standard Normal Variate (SNV) (D).....	29
5. Análise discriminante entre os grupos compostos por amostras de grão (trigo, soja e arroz) e planta (trigo, milho e arroz) .....	30

	Página
6. Análise de componentes principais (PCA) com os dados espectrais para os grupos compostos por amostras de grão (trigo, soja e arroz) e planta (trigo, milho e arroz) .....	31
7. Importância das bandas espectrais na a criação de modelos para estimativa de N em grãos (A) e plantas (B).....	32
8. Gráfico de dispersão do conteúdo de N medido e estimado nas amostras de arroz, trigo e milho, utilizando o algoritmo SVM com o pré-processamento espectral SNV.....	34
9. Gráfico de dispersão do conteúdo de N medido e estimado utilizando o modelo SVM com pré-processamento SNV em amostras de grãos de arroz, trigo e soja.....	36

## 1 INTRODUÇÃO

O nitrogênio (N) é um nutriente indispensável para o crescimento das plantas por ser um constituinte de proteínas, ácidos nucleicos, clorofila e demais metabólitos especializados (Liu *et al.*, 2022). A deficiência de N pode promover a redução do tamanho da folha e do teor de clorofila, afetando diretamente o rendimento e a qualidade das culturas agrícolas (Wang *et al.*, 2014). Já a fertilização excessiva pode levar a uma série de problemas ambientais, incluindo a poluição do solo e das águas subterrâneas, principalmente com nitrato (Huang *et al.*, 2020). Portanto, é desejável um manejo estratégico do N considerando as diferentes demandas pelas plantas durante seu ciclo de desenvolvimento fenológico (Liu *et al.*, 2022).

Os métodos mais usados para determinar a concentração de N são as análises químicas, como os métodos de Kjeldahl e Dumas (Tedesco *et al.*, 1995). No entanto, esses métodos tradicionais apresentam desvantagens como a destruição das amostras, longos tempos de análise e utilização de produtos químicos tóxicos (Liu *et al.*, 2022). Esses processos de detecção demorados deixam os agricultores com informações desatualizadas sobre o estado nutricional das plantas, o que é prejudicial à produção. Por estas razões, as análises químicas tradicionais não são apropriadas para monitorar a condição nutricional nas culturas, e um método alternativo deve ser usado (Miao *et al.*, 2023).

A espectroscopia de infravermelho próximo (NIR - Near Infrared Spectroscopy) é um método que surgiu como alternativa aos métodos analíticos tradicionais para a estimativa dos teores de nutrientes nos tecidos vegetais, de uma maneira mais rápida, menos onerosa e menos poluente (Van Maarschalkerweerd & Husted, 2015). Metodologias que empregam a espectroscopia NIR têm substituído parcialmente os métodos de análise laboratoriais em diversos setores como em análises de alimentos, produtos farmacêuticos, agrícolas, entre outros (Pasquini, 2018).

A espectroscopia NIR é um tipo de espectroscopia vibracional de alta energia realizada na faixa de comprimento de onda de 750 a 2500 nm (Pasquini, 2018). Essa técnica baseia-se na absorção de radiação eletromagnética na região do infravermelho próximo, que corresponde às frequências de vibração das ligações C-H, O-H e N-H encontradas nos grupos funcionais de moléculas orgânicas. O N faz parte dessas moléculas orgânicas e interage com os grupos funcionais, permitindo sua identificação e quantificação (Turker-Kaya & Huck, 2017). Entretanto, para que a técnica seja empregada com boa acurácia e precisão, é necessário a utilização de pré-processamentos espectrais e a construção de um modelo de calibração multivariada associando os dados espectroscópicos NIR aos resultados analíticos realizados pelos métodos de referência (Prananto *et al.*, 2020).

Recentemente diversos estudos utilizaram a espectroscopia NIR para estimativa de macro e micronutrientes em plantas cultivadas, entretanto os modelos são calibrados para uma espécie e local específico (Prananto *et al.*, 2020). Murguzur *et al.* (2019) calibrou um modelo para 97 espécies em diferentes regiões geográficas e obteve resultados satisfatórios para estimativa de nitrogênio, provando ser possível a construção de um modelo amplo que possa ser utilizado para mais de uma espécie e região. Modelos amplos podem aumentar a usabilidade do método, permitindo sua aplicação em diversos contextos, economizando tempo e recursos (Murguzur *et al.*, 2017). Bibliotecas espectrais podem ser construídas para dar suporte à técnica, permitindo que dados químicos e espectrais de diferentes espécies e regiões sejam reunidos em um único local, ficando disponíveis para novas calibrações e evitando a necessidade de grandes quantidades de novas amostras (Wijewardane *et al.*, 2023). As principais culturas agrícolas produzidas no RS são o arroz, milho, soja e trigo, portanto, é essencial modelos calibrados para a estimativa da concentração de nitrogênio nessas culturas (IBGE, 2021).

O objetivo deste estudo foi avaliar o uso da espectroscopia de infravermelho próximo (NIR) para estimar a concentração de N no tecido vegetal (folhas e grãos) das culturas de arroz, milho, soja e trigo por meio de técnicas de pré-processamento e calibração multivariada, a fim de desenvolver um modelo único de estimativa.

O presente trabalho tem como hipótese que a espectroscopia NIR pode substituir as análises químicas para estimar a concentração de N em plantas; a estimativa da concentração de N pode ser mais precisa utilizando técnicas adequadas de pré-

processamento e de calibração multivariada; é possível a calibração de modelos amplos para mais de uma espécie.

## 2 REVISÃO BIBLIOGRÁFICA

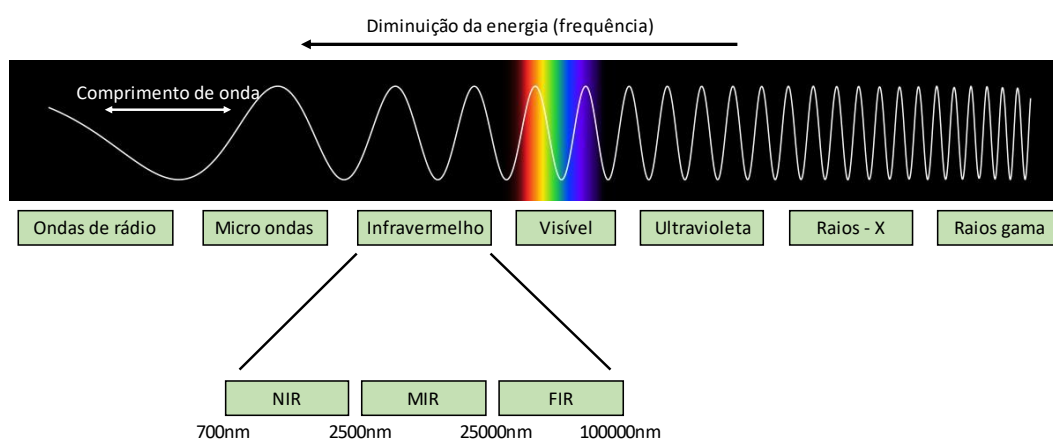
### 2.1 Espectroscopia de reflectância na agricultura

Nos anos 1700, Pierre Bouguer propôs a lei de Beer-Lambert, que forneceu aos pesquisadores uma base fisiológica para entender a relação entre compostos químicos e radiação eletromagnética. Em 1800, Frederick William Herschel, descobriu a "radiação infravermelha" e a relação entre os raios de luz visíveis e invisíveis e encontrou formas simples de examinar a absorção e a reflexão (Ring, 2000). Após essa descoberta, foram registrados estudos sobre a utilização desta técnica, porém, merece destaque, pela importância dos seus trabalhos, o grupo de pesquisadores de Karl Norris, que desempenharam um papel crucial no desenvolvimento de uma aplicação NIR para amostras orgânicas (Norris *et al.*, 1976). A partir dos anos 1980, a espectroscopia de reflectância começou a ser amplamente estudada, com a finalidade de analisar a umidade de grãos e, posteriormente, se expandiu para outras áreas do conhecimento, devido à sua eficiência na obtenção de informações (Zornoza *et al.*, 2008). As pesquisas sobre espectroscopia de NIR entre as décadas de 80 e 90 concentraram-se na estimativa de proteína em forragens e em grãos (Prananto *et al.*, 2020). Com base nas informações encontradas, buscou-se entender quais regiões do espectro eletromagnético tinham características de absorção conhecidas para a detecção de compostos químicos específicos, como nitrogênio (N), fósforo (P) e potássio (K) (Curran, 1989; Elvidge, 1990; Sasaki *et al.*, 1984). Em 1991, foi desenvolvido o primeiro modelo para análise de N em plantas de arroz visando a obtenção de melhores estratégias para aplicação de fertilizantes (Batten & Blakenev, 1991). A atual disseminação do NIRS a nível industrial deve ser creditada a Philip Williams e à sua crença no potencial da nova técnica enquanto trabalhava para a Comissão Canadense de Grãos. Este pesquisador substituiu a determinação de proteína no trigo, feita pelo método Kjeldahl convencional, pela técnica recentemente introduzida baseada no NIRS (Williams, 2015). O trabalho desenvolvido

por Williams teve também o mérito de motivar o interesse das empresas de instrumentos em produzir e comercializar instrumentos NIR, contribuindo decisivamente para a difusão da tecnologia NIR.

## 2.2 Princípio da espectroscopia - NIR

A radiação eletromagnética (REM) é uma forma de energia que se propaga por meio de ondas, as quais têm diferentes comprimentos de onda. As partículas que carregam energia através das ondas eletromagnéticas são chamadas de fótons, sendo esta energia diretamente relacionada à frequência da radiação eletromagnética (Niemz, 2007). A energia da radiação depende de sua frequência, sendo que quanto maior a frequência (comprimento de onda curto), maior a energia. O espectro eletromagnético compreende todos os comprimentos de onda, das ondas menos energéticas, como as ondas de rádio, até as mais energéticas, como os raios gama (Figura 1) (NIEMZ, 2007). A região do infravermelho é dividida em três regiões: infravermelho próximo – NIR (700 a 2500 nm), infravermelho médio – MIR (2500 a 25.000 nm) e infravermelho distante FIR (25.000 a 100.000 nm) (Turker-Kaya e Huck, 2017).



**FIGURA 1.** Espectro eletromagnético e suas regiões espectrais. A região do infravermelho é dividida em três partes: Infravermelho próximo (NIR), Infravermelho médio (MIR) e Infravermelho distante (FIR).

A espectroscopia NIR consiste basicamente na emissão de um feixe de radiação, com comprimento de onda conhecido, sobre uma amostra. Conforme a amostra é exposta ao feixe radiação, ela absorve parte da energia, resultando em vibrações moleculares (Zahir *et al.*, 2022). Quando a radiação incide na amostra, parte da luz recebida pode ser transmitida, absorvida ou refletida, e a luz resultante dessa interação será quantificada em

um detector, gerando a assinatura espectral da amostra (Pu *et al.*, 2020). A quantidade de luz absorvida pelas amostras é altamente influenciada pelas propriedades da amostra com as quais a luz interagiu, como densidade e composição química, e será usada para estimativa do conteúdo de nutrientes que quer ser quantificado (Carvalho *et al.*, 2015).

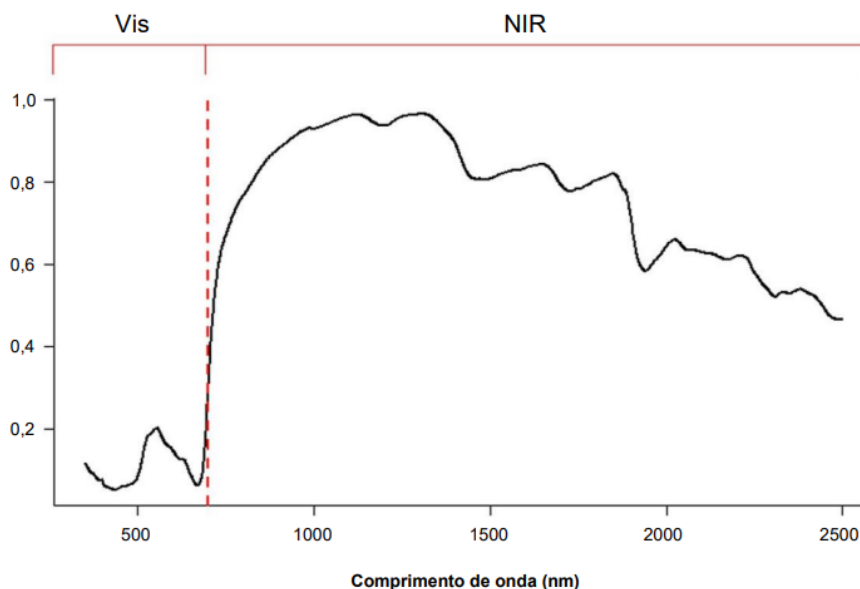
### 2.3. Uso da espectroscopia de reflectância em plantas

A mudança no estado de energia das moléculas é o princípio mais importante em espectroscopia e pode ser descrito pela equação de Planck, conforme Equação 1.

$$E = h\nu = hc / \lambda \quad (1)$$

Onde:  $E$  é a energia necessária para mudar o estado do elétron do estado fundamental para o estado de excitação,  $h$  é a constante de Planck,  $\nu$  é o número de onda,  $c$  é a velocidade da luz e  $\lambda$  é a comprimento de onda.

De acordo com a equação, quanto menor a energia necessária durante o estado de transição do elétron do estado fundamental para o estado excitado, maior o comprimento de onda da banda de absorção (Zahir *et al.*, 2022). Os tecidos vegetais têm valores diferentes de reflectância para cada comprimento de onda, desde o visível (VIS) até o infravermelho próximo (NIR). Para cada espécie de planta, o espectro de refletância é único, como pode ser observada na Figura 2 (Prananto *et al.*, 2020).



**FIGURA 2.** Assinatura espectral típica de tecidos vegetais nas bandas espectrais do visível (VIS) e do infravermelho próximo (NIR).



As amostras de tecido vegetal, sejam elas folhas ou grãos, apresentam uma resposta única ao espectro eletromagnético e, portanto, cada uma possui sua própria assinatura espectral (Richardson *et al.*, 2004). Este espectro é resultado da absorção da luz por compostos químicos e estruturais presentes nas plantas (Richardson *et al.*, 2004).

A espectroscopia de NIR pode fornecer uma grande quantidade de informações baseadas na absorção de energia por ligações como C–H, N–H e O–H, que são encontradas em moléculas orgânicas como carboidratos, lipídeos e proteínas. Assim, a composição mineral de uma planta pode ser estimada por NIR devido à associação dos nutrientes com estes grupos funcionais orgânicos, como podem ser observadas na Tabela 1 (Acosta *et al.*, 2023; Turker-Kaya & Huck, 2017).

**TABELA 1.** Metabólitos das plantas, seus respectivos comprimentos de onda detectadas por espectroscopia na faixa de NIR e os grupos funcionais relacionados.

<b>Metabólito</b>	<b>Comprimento de onda (nm)</b>	<b>Grupo funcional</b>
Carboidratos	1190	C–H
Carboidratos	1212	C–H
Carboidratos	1395	C–H
Proteínas	1432	N–H
Carboidratos	1482	O–H
Carboidratos	1501	N–H
Amido	1540	O–H
Proteínas	1564	N–H
Carboidratos	1614	C–H
C–H: carboidratos	1652	C–H
C–H: carboidratos	1696	C–H
Celulose, lignina e carboidratos	1777	C–H
Celulose, lignina e carboidratos	1816	O–H + C–O
Carboidratos	1953	C–O
Proteínas	2050	N–H
Álcoois	2073	O–H + O–H
Amido	2253	O–H + O–H
Proteínas	2292	N–H + C–O

## 2.4 Análise multivariada dos dados

Os dados dos espectros NIR contêm centenas a milhares de valores de reflectância em função dos comprimentos de onda (Prananto *et al.*, 2020). Além disso, os espectros são de difícil interpretação devido à sobreposição de sobretons e bandas de combinação. Assim, a análise multivariada de dados (MDA) é necessária para extrair as informações dos espectros NIR das amostras (Pu *et al.*, 2020). As abordagens analíticas de dados espectroscópicos incluem pré-processamento espectral, análise exploratória dos dados, calibração e validação (Pasquini, 2018). Nos últimos anos, diversos estudos foram publicados envolvendo a utilização de métodos de análise multivariada para estimativa de nitrogênio em plantas através da espectroscopia NIR (Tabela 2).

**Tabela 2.** Resumo dos estudos científicos que utilizaram a espectroscopia para estimar a concentração de N em plantas.

Referências	Cultura	Pré-processamento	Modelo
Abukmeil <i>et al.</i> (2022)	Batata	-	MLR
Acosta <i>et al.</i> (2023)	Caqui	MC + D1	PLSR
Acosta <i>et al.</i> (2023)	Caqui	MC + D2	PLSR
Alwi <i>et al.</i> (2021)	Eucalipto	SGD + SNV	PLSR
Azadnia <i>et al.</i> (2023)	Maçã	MSC + D2	RF
Chen <i>et al.</i> (2020)	Maça	SNV + D1	ELM
Cuq <i>et al.</i> (2019)	Uva	MSC	PLSR
Johnson <i>et al.</i> (2021)	Arroz	SGD + D2	PLSR
Murguzur <i>et al.</i> (2019)	97 espécies	-	PLSR
Ni <i>et al.</i> (2018)	Pinheiro	-	VWCNN
Prananto <i>et al.</i> (2021)	Algodão	SNV + DT	Cubist Regression
Rodrigues <i>et al.</i> (2020)	Soja	MSC	PLSR
Zhai <i>et al.</i> (2012)	Arroz	MSC + SNV + D1	SVM
Wang <i>et al.</i> (2021)	Pereira	-	PLSR

MC: Mean Center; D1: First Derivative; D2: Second Derivative; SGD: Savitzky–Golay Derivative; SNV: Standard Normal Variate; MSC: Multiplicative Scatter Correction; EPO: External Parameter Orthogonalization; PLSR: Partial Least Squares Regression; MLR: Multiple Linear Regression; RF: Random Forest; ELM: Extreme Learning Machine; VWCNN: Variable Weighted Convolutional Neural Network; ANN: Artificial Neural Networks; SVM: Suport Vector Machine.

## 2.5 Pré-processamentos espectrais

Os espectros brutos possuem ruídos que podem surgir da interferência de fatores físicos e/ou químicos, erros técnicos ou fatores aleatórios. A alta relação sinal-ruído (SNR) pode reduzir a precisão da calibração (Xu *et al.*, 2008). Além disso, o fundo e a dispersão de refletância (espalhamento) adiciona variáveis indesejadas aos dados espectrais que levam a modelos tendenciosos e de baixa precisão (Xu *et al.*, 2008).

Os pré-processamentos espectrais são aplicados para melhorar as características espectrais, suprimindo variações aleatórias, normalizando os espectros em relação às condições de medição e aprimorando as informações quimicamente relevantes. Essas técnicas incluem correções de linha de base, normalizações, derivadas e suavização, sendo o pré-processamento de espectros brutos antes da análise multivariada uma importante aliada (Salzer & Siesler, 2014).

A regra geral para a definição do método de pré-processamento depende da faixa espectral e padronizou-se o erro como o parâmetro a ser avaliado (Bec *et al.*, 2020). A seguir será detalhado os pré-processamentos utilizados neste estudo, sendo eles: *Detrend* (DET), Derivada de *Savitski-Golay* (SGD) e *Standard Normal Variate* (SNV).

### 2.5.1 *Detrend* (DET)

A técnica de pré-processamento DET visa remover a tendência sistemática presente nos espectros de refletância difusa no infravermelho próximo (NIR). Isso é alcançado ajustando uma função matemática à tendência observada nos espectros e subtraindo essa função dos espectros originais (Barnes *et al.*, 1989). Essa remoção da tendência permite destacar as variações espectrais relevantes, reduzindo o impacto de fatores não relacionados à composição química dos materiais. A técnica *Detrend* facilita a detecção de padrões e características específicas nos espectros, aprimorando a identificação e quantificação dos componentes químicos de interesse (Barnes *et al.*, 1989).

### 2.5.2 *Savitski-Golay Derivative* (SGD)

A técnica SGD tem como objetivo suavizar e diferenciar os dados espectrais, reduzindo o ruído e preservando características relevantes. A SGD baseia-se em um procedimento de mínimos quadrados ponderados, no qual polinômios locais são ajustados a pontos de dados vizinhos. Esses polinômios são utilizados para suavizar o sinal original,

minimizando a distorção introduzida. Ao ajustar os polinômios, a técnica leva em consideração as flutuações locais nos dados, permitindo uma suavização mais precisa (Savitzky & Golay, 1964). A SGD também é capaz de calcular a derivada dos dados, fornecendo informações sobre as taxas de variação. Isso é útil para identificar pontos de máximo ou mínimo nos dados e destacar características relevantes (Savitzky & Golay, 1964).

### **2.5.3 *Standard Normal Variate (SNV)***

A técnica consiste em transformar os espectros, de forma a remover as variações sistemáticas não relacionadas à composição química dos materiais. Essa técnica normaliza cada espectro individualmente, subtraindo a média do espectro e dividindo pelo desvio padrão. Essa transformação padroniza os espectros em termos de distribuição normal, tornando-os mais comparáveis e realçando as variações espectrais relevantes (Barnes *et al.*, 1989). O objetivo é reduzir o impacto de variações indesejadas, como espalhamento da luz e desvio instrumental, permitindo uma análise mais precisa das características espectrais relacionadas aos componentes químicos de interesse (Barnes *et al.*, 1989).

### **2.6 *Principal Component Analysis (PCA)***

É comum a utilização da técnica de análise de componentes principais (PCA) antes da etapa da calibração multivariada como forma de análise exploratória para investigar a presença de outliers e determinar se as amostras podem ser distinguidas (Miao *et al.*, 2023). A PCA é um método de mineração de dados, que reduz o número de variáveis a serem analisadas (como bandas espectrais), e promove, através de uma matriz de covariâncias composta por todas as bandas espectrais de todas as amostras espectrais, a transformação de essas variáveis em novos grupos de variáveis, chamados de componentes principais (CP) (Santos *et al.*, 2023). O PCA é uma técnica estatística que reduz a dimensionalidade dos espectros, revelando suas principais fontes de variação. Ao transformar os dados em componentes principais não correlacionados, o PCA simplifica a interpretação, destaca padrões relevantes e remove ruídos indesejados, possibilitando visualizar a estrutura dos espectros em um espaço de menor dimensão (Martens & Naes, 1992). Essa ferramenta desempenha um papel crucial na análise de dados espectrais, permitindo uma compreensão mais precisa e insights valiosos para diversas aplicações

(Turker-Kaya & Huck, 2017).

## 2.7 Calibração Multivariada

A calibração multivariada é uma técnica estatística usada para construir modelos matemáticos que relacionam variáveis independentes a uma variável dependente de interesse, permitindo a análise, previsão ou estimativa dessa variável dependente com base nas informações multivariadas disponíveis (Martens & Naes, 1992). Na espectroscopia, o processo de calibração visa converter a informação espectral em informação química da amostra foliar através de modelos estatísticos (Mark *et al.*, 2002). A qualidade da calibração determina a precisão da técnica na estimativa das propriedades químicas, ou seja, o conteúdo de nutrientes da folha (Nicolai *et al.*, 2007).

No processo de modelagem, geralmente são necessários dois conjuntos de dados: calibração e validação (Prananto *et al.*, 2020). A Calibração refere-se ao processo de estabelecer uma relação matemática entre as variáveis independentes (ou preditoras) e a variável dependente com base em um conjunto de amostras de calibração. Esse processo envolve a seleção adequada do modelo estatístico, ajuste dos parâmetros do modelo e estimativa das relações entre as variáveis. A calibração visa construir um modelo que seja capaz de descrever e capturar as relações presentes nos dados (Martens & Naes, 1992). A validação dos modelos, por outro lado, é o processo que avalia a capacidade e a qualidade do modelo construído durante a calibração. Ela envolve o uso de um conjunto de amostras independentes, denominado conjunto de validação, que não foi usado durante a calibração. O objetivo da validação é verificar se o modelo é capaz de fornecer resultados precisos e confiáveis para novas amostras (Martens & Naes, 1989).

A precisão e ajuste dos dados de calibração e validação são avaliados usando parâmetros, tais como o coeficiente de determinação ( $R^2$ ), raiz quadrada do erro médio (RMSE), o erro de previsão de *standard* (SEP) e o erro sistemático (viés) (Nicolai *et al.*, 2007). Outros parâmetros também são usados, como a razão entre RMSE e o desvio padrão dos dados (RPD) e a razão entre RMSE e intervalo interquartil dos dados (RPIQ) (Bellon-Maurel *et al.*, 2010). Algumas das técnicas de análise multivariada mais utilizadas na calibração multivariada são a Partial Least Squares Regression (PLSR) e Support Vector Machine (Chen *et al.*, 2007; Pasquini, 2018; Viscarra Rossel & Behrens, 2010).

### **2.7.1 Partial Least Squares Regression (PLSR)**

O *Partial Least Squares Regression* (PLSR) é um dos métodos de calibração multivariada mais utilizados na análise espectral de infravermelho e pode construir modelos de predição com múltiplos fatores e alta colinearidade (Wold *et al.*, 2001). A técnica estabelece uma relação entre um conjunto de variáveis independentes e uma variável dependente, construindo variáveis latentes, que são combinações lineares das variáveis independentes, buscando maximizar a variação das variáveis independentes e a covariância com a variável dependente. Esses componentes são obtidos de forma iterativa, ajustando pesos para as variáveis independentes e estimando os pesos para a variável dependente. Esses pesos são usados para prever a variável dependente para novas amostras, com base nas combinações lineares das variáveis independentes (Geladi *et al.*, 1986).

O PLSR projeta uma relação linear do parâmetro ou concentração modelado em função das variações espectrais. Não linearidades fracas podem ser superadas aumentando o número de variáveis latentes incluídas no modelo (Wold *et al.*, 2001). As principais vantagens da análise PLSR são a capacidade de lidar com mais variáveis descritoras com menor risco de correlação e a capacidade de analisar dados ruidosos e dados com variáveis incompletas (Zahir *et al.*, 2022).

### **2.7.2 Support Vector Machine (SVM)**

O *Support Vector Machine* (SVM) pode ser visto como um método de treinamento polinomial em que os pesos da rede são encontrados resolvendo um problema de programação quadrática com desigualdade linear e restrições de igualdade (Cortes & Vapnik, 1995). O SVM corrige a função de decisão de classificação com base no erro mínimo de risco estrutural em vez do erro mínimo da classificação incorreta no conjunto de treinamento para evitar problemas de sobre ajuste (Chen *et al.*, 2007). Ele executa o problema de classificação binária encontrando hiperplanos de margem máxima em termos de um subconjunto dos dados de entrada (vetores de suporte) entre diferentes classes. Utilizando esta técnica é possível derivar um plano espectral linear como uma função de decisão para problemas não lineares (Viscarra Rossel & Behrens, 2010).

## **2.8 Biblioteca Espectral**

O desenvolvimento de modelos de calibração requer um grande número de

amostras a serem analisadas em laboratório. Além disso, geralmente os modelos são calibrados em um conjunto limitado de amostras (uma única espécie, poucos genótipos e um ou dois ambientes), fazendo com que o desempenho desses modelos em outros conjuntos de dados independentes diminua substancialmente em relação ao seu desempenho inicial (Wijewardane *et al.*, 2023). Uma solução para esse problema é estabelecer uma biblioteca espectral em nível de folha que compreenda muitas amostras com dados de laboratório e dados espectrais (Wijewardane *et al.*, 2023). Com uma biblioteca espectral estabelecida, torna-se possível realizar uma abordagem chamada de 'spiking'. O conceito central de spiking é adicionar um pequeno número de amostras "locais" às amostras da biblioteca, fazendo com que aumente a precisão das próximas calibrações para as amostras com aquelas mesmas características das amostras locais incluídas (Wijewardane *et al.*, 2023).

As bibliotecas espectrais tiveram início na década de 80 com Stoner e Baumgardner nos Estados Unidos, e posteriormente, vários trabalhos científicos foram publicados sobre o tema. No Brasil, a primeira biblioteca foi elaborada pelo grupo de pesquisadores do Instituto Nacional de Pesquisas Espaciais (INPE), em 1992. Em nível de Universidade, a primeira Biblioteca foi iniciada no Departamento de Ciência do Solo da ESALQ em 1995.

A reflectância carrega consigo informações relacionadas à constituição físico-química do material, sendo que cada tipo de material (vegetação, solos, água, etc) apresenta um padrão de comportamento espectral distinto. Para um mesmo tipo material, como por exemplo, o tecido vegetal, existem diferentes comportamentos espectrais que podem variar de acordo com as características genéticas e do ambiente onde ele foi cultivado, como por exemplo o tipo de solo, nutrição, sanidade, condições hídricas, entre outros (Prananto *et al.*, 2020). Um grande banco de dados espectrais permite visualizar padrões de similaridade e variabilidade entre distintas amostras, e também serve como fonte de informação para estudos posteriores. Sendo assim, é de suma importância para a comunidade científica a criação de Bibliotecas Espectrais (Viscarra Rossel & Behrens, 2010).

## 2.9 Referências

- ACOSTA, M. *et al.* Non-destructive appraisal of macro-and micronutrients in persimmon leaves using Vis/NIR hyperspectral imaging. **Agriculture**, Basel, v. 13, n. 4, [art.] 916, 2023.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, Plainfield, v. 43, n. 5, p. 772-777, 1989.
- BATTEN, G. D.; BLAKENEY, A. B. Applying NIR to crop nutrient assessment. **NIR News**, New York, v. 2, n. 4, p. 10-11, 1991.
- BEĆ, K. B. *et al.* Principles and applications of vibrational spectroscopic imaging in plant science: a review. **Frontiers in Plant Science**, Lausanne, v. 11, [art.] 1226, 2020.
- BELLON-MAUREL, V. *et al.* Critical review of chemometric indicators commonly used for assessing the quality of the prediction of soil attributes by NIR spectroscopy. **Trends in Analytical Chemistry**, Amsterdam, v. 29, n. 9, p. 1073–1081, 2010.
- CARVALHO, G. G. A. *et al.* Direct determination of the nutrient profile in plant materials by femtosecond laser-induced breakdown spectroscopy. **Analytica Chimica Acta**, Amsterdam, v. 876, p. 26-38, 2015.
- CHEN, T.; MORRIS, J.; MARTIN, E. Gaussian process regression for multivariate spectroscopic calibration. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 87, n. 1, p. 59-71, 2007.
- CORTES, C.; VAPNIK, V. Support-vector networks. **Machine Learning**, Boston, v. 20, p. 273-297, 1995.
- CURRAN, P. J. Remote sensing of foliar chemistry. **Remote Sensing of Environment**, New York, v. 30, n. 3, p. 271-278, 1989.
- ELVIDGE, C. D. Visible and near infrared reflectance characteristics of dry plant materials. **Remote Sensing**, Basel, v. 11, n. 10, p. 1775-1795, 1990.
- GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, Amsterdam, v. 185, p. 1-17, 1986.
- HUANG, C. H. *et al.* Early diagnosis and management of nitrogen deficiency in plants utilizing Raman spectroscopy. **Frontiers in Plant Science**, Lausanne, v. 11, [art.] 663, 2020.
- IBGE - INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **Ranking – Agricultura** - Valor da produção (2022). Rio de Janeiro: IBGE, 2022. Disponível em: <https://www.ibge.gov.br/explica/producao-agropecuaria/rs>. Acesso em: 14 ago. 2023.



JUAN, A. D. *et al.* Chemometric tools for image analysis. *In*: SALZER, R.; SIESLER, H. W. (ed.). **Infrared and raman spectroscopic imaging**. Weinheim: John Wiley, 2014. p. 57-110.

LIU, W. *et al.* Non-destructive measurements of *Toona sinensis* chlorophyll and nitrogen content under drought stress using near infrared spectroscopy. **Frontiers in Plant Science**, Lausanne, v. 12, [art.] 809828, 2022.

MAARSCHALKERWEERD, M.; HUSTED, S. Recent developments in fast spectroscopy for plant mineral analysis. **Frontiers in Plant Science**, Lausanne, v. 6, [art.] 169, 2015.

MARK, H. *et al.* Validation of a near-infrared transmission spectroscopic procedure, part A: validation protocols. **Journal of Pharmaceutical and Biomedical Analysis**, Oxford, v. 28, n. 2, p. 251-260, 2002.

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

MIAO, X. *et al.* Measurement of nitrogen content in rice plant using near infrared spectroscopy combined with different PLS algorithms. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, Amsterdam, v. 284, [art.] 121733, 2023.

MURGUZUR, F. J. A. *et al.* Towards a global arctic-alpine model for Near-infrared reflectance spectroscopy (NIRS) predictions of foliar nitrogen, phosphorus and carbon content. **Scientific Reports**, London, v. 9, n. 1, [art.] 8259, 2019.

NICOLAI, B. M. *et al.* Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: a review. **Postharvest Biology and Technology**, Amsterdam, v. 46, n. 2, p. 99-118, 2007.

NIEMZ, M. H. Light and matter. *In*: HOUCK, M. M.; SIEGEL, J. A. **Fundamentals of forensic science**. Amsterdam: Academic Press, 2007. p. 94-119.

NORRIS, K. H. *et al.* Predicting forage quality by infrared reflectance spectroscopy. **Journal of Animal Science**, Champaign, v. 43, n. 4, p. 889-897, 1976.

PASQUINI, C. Near infrared spectroscopy: a mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, Amsterdam, v. 1026, p. 8-36, 2018.

PRANANTO, J. A.; MINASNY, B.; WEAVER, T. Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues. **Advances in Agronomy**, San Diego, v. 164, p. 1-49, 2020.

PU, Y. Y. *et al.* Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing. **International Dairy Journal**, Champaign, v. 103, [art.] 104623, 2020.

RICHARDSON, A. D.; REEVES III, J. B.; GREGOIRE, T. G. Multivariate analyses of

- visible/near infrared (VIS/NIR) absorbance spectra reveal underlying spectral differences among dried, ground conifer needle samples from different growth environments. **New Phytologist**, Oxford, v. 161, n. 1, p. 291-301, 2004.
- RING, E. F. J. The discovery of infrared radiation in 1800. **The Imaging Science Journal**, London, v. 48, n. 1, p. 1-8, 2000.
- SANTOS, D. G. L. A. A. *et al.* Spectral method for macro and micronutrient prediction in soybean leaves using interval partial least squares regression. **European Journal of Agronomy**, Amsterdam, v. 143, [art.] 126717, 2023.
- SASAKI, K.; KAWATA, S.; MINAMI, S. Estimation of component spectral curves from unknown mixture spectra. **Applied Optics**, Denver, v. 23, n. 12, p. 1955-1959, 1984.
- SAVITZKY, A.; GOLAY, M. J. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, Washington, DC, v. 36, n. 8, p. 1627-1639, 1964.
- TÜRKER-KAYA, S.; HUCK, C. W. A review of mid-infrared and near-infrared imaging: principles, concepts and applications in plant tissue analysis. **Molecules**, Basel, v. 22, n. 1, [art.] 168, 2017.
- VISCARRA ROSSEL, R. A.; BEHRENS, T. Using data mining to model and interpret soil diffuse reflectance spectra. **Geoderma**, Amsterdam, v. 158, n. 1/2, p. 46- 54, 2010.
- WANG, Y. *et al.* Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. **Plant**, New York, v. 10, n. 1, [art.] 36, [p. 1-11], 2014.
- WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, Amsterdam, v. 58, n. 2, p. 109-130, 2001.
- WIJEWARDANE, N. K. *et al.* A leaf-level spectral library to support high-throughput plant phenotyping: predictive accuracy and model transfer. **Journal of Experimental Botany**, Oxford, v. 74, n. 14, p. 4050-4062, 2023.
- XU, Lu *et al.* Ensemble preprocessing of near-infrared (NIR) spectra for multivariate calibration. **Analytica Chimica Acta**, Amsterdam, v. 616, n. 2, p. 138-143, 2008.
- ZAHIR, S. A. D. M. *et al.* A review of visible and near-infrared (Vis-NIR) spectroscopy application in plant stress detection. **Sensors and Actuators A: Physical**, Iowa, v. 338, [art.] 113468, 2022.
- ZORNOZA, R. *et al.* Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. **Soil Biology and Biochemistry**, Amsterdam, v. 40, n. 7, p. 1923-1930, 2008.

### **3 ARTIGO 1**

**Espectroscopia de infravermelho próximo para estimativa da concentração de nitrogênio em plantas e grãos de arroz, trigo, milho e soja**

## RESUMO

Os métodos químicos tradicionalmente empregados para a estimativa do teor de nitrogênio em plantas apresentam elevado custo, são operacionalmente ineficientes e necessitam o uso de reagentes químicos, elevando os riscos tanto para o laboratorista quanto para o meio ambiente. A espectroscopia de infravermelho próximo (*Near Infrared Spectroscopy* - NIR) é um método que possibilita a análise de forma mais eficiente, com menos custos e não geração de resíduos tóxicos. Para que esta técnica apresente boa acurácia e precisão, é necessário a criação de um modelo de calibração multivariada que associe os dados espectrais aos resultados obtidos nas análises químicas. O presente trabalho tem como objetivo a criação de um modelo único que seja capaz de fazer a predição da concentração de nitrogênio em plantas de milho, arroz, trigo e soja com o uso de diferentes combinações de técnicas de pré-processamento e calibração multivariada. Para o presente estudo foram coletadas e analisadas um total de 899 amostras foliares. Foram utilizados os pré-processamentos *Detrend (DET)*, *Standard Normal Variate (SNV)* e *Savitzky-Golay Derivative (SGD)*. Os modelos de predição foram desenvolvidos a partir dos seguintes métodos de calibração multivariada: *Partial Least Squares Regression (PLSR)* e *Support Vector Machine (SVM)*. A calibração do modelo foi feita através da validação cruzada por meio de uma divisão aleatória em segmentos. Os valores preditos de nitrogênio através da espectroscopia foram comparados aos valores mensurados pelo método químico, sendo a acurácia da espectroscopia para este fim avaliada através dos seguintes parâmetros estatísticos: coeficiente de determinação ( $R^2$ ), relação entre o desempenho e a distância interquartil (RPIQ) e o erro médio quadrático de predição (RMSE). O uso da espectroscopia de infravermelho próximo foi capaz de prever com precisão as concentrações de nitrogênio em tecido vegetal, sendo que os melhores resultados foram encontrados com o uso do método de calibração SVM e pré-processamento SNV ( $R^2 = 0,99$ ). Adicionalmente, foi verificado que é possível a construção de um modelo único que seja capaz de estimar a concentração de N para o conjunto de dados com as quatro culturas.

**Palavras-chave:** Espectroscopia de infravermelho próximo; estimativa de nitrogênio; pré-processamento espectral; tecido vegetal.

## ABSTRACT

The chemical methods traditionally used to estimate the nitrogen content in plants are expensive, operationally inefficient and require the use of chemical reagents, increasing risks for both the laboratory worker and the environment. Near infrared spectroscopy - NIR (Near Infrared Spectroscopy) is a method that enables analysis more efficiently, with less costs and does not generate toxic waste. For this technique to present good accuracy and precision, it is necessary to create a multivariate calibration model that associates the spectral data with the results obtained in chemical analyses. The present work aims to create a unique model that is capable of predicting nitrogen concentration in corn, rice, wheat and soybean leaves using different combinations of pre-processing and multivariate calibration techniques. For the present study, a total of 899 leaf samples were collected and analyzed. *Detrend (DET)*, *Standard Normal Variate (SNV)* and *Savitzky-Golay Derivative (SGD)* pre-processing were used. The prediction models were

developed based on the following multivariate calibration methods: Partial Least Squares Regression (PLSR) and Support Vector Machine (SVM). The model calibration was done through cross-validation through a random division into segments. The predicted nitrogen values through spectroscopy were compared to the values measured by the chemical method, with the accuracy of the spectroscopy for this purpose being evaluated through the following statistical parameters: coefficient of determination ( $R^2$ ), relationship between performance and interquartile distance (RPIQ) and the mean squared prediction error (RMSE). The use of near-infrared spectroscopy was able to accurately predict nitrogen concentrations in plant tissue, with the best results being found using the SVM calibration method and SNV pre-processing ( $R^2 = 0.99$ ). Additionally, it was verified that it is possible to build a single model that is capable of estimating the N concentration for the data set with the four crops.

**Key-words:** Near infrared spectroscopy; nitrogen estimation; spectral preprocessing; plant tissue.

### 3.1 Introdução

O nitrogênio (N) é um nutriente indispensável para o crescimento das plantas por ser um constituinte de proteínas, ácidos nucleicos, clorofila e metabólitos especializados (Li *et al.*, 2020). A deficiência de N pode levar à redução do tamanho da folha e do teor de clorofila, o que afeta diretamente o rendimento e a qualidade das culturas (Wang *et al.*, 2014), enquanto a fertilização excessiva pode levar a uma série de problemas ambientais, incluindo a poluição do solo e das águas subterrâneas (Huang *et al.*, 2020). Portanto, é desejável um manejo estratégico do N considerando as diferentes demandas pelas plantas durante cada estágio fenológico (Li *et al.*, 2020), o que pode ser feito através de análises foliares. A análise da concentração de N também é importante para os grãos, pois influencia diretamente a segurança alimentar, o valor comercial e a satisfação do consumidor (Pandiselvam *et al.*, 2021). A quantidade de nitrogênio (N) nos grãos se relaciona diretamente com seu conteúdo de proteína (Mariotti *et al.*, 2008), característica importante usada para avaliar sua qualidade (Kamboj *et al.*, 2022), e é uma informação importante em programas de manejo de fertilidade do solo.

Os métodos mais usados para avaliar o teor de N no tecido vegetal são as análises químicas, como os métodos de Kjeldahl e Dumas (Tedesco *et al.*, 1995). No entanto, esses métodos tradicionais possuem as desvantagens de serem técnicas onerosas, que exigem a destruição das amostras e que consomem reagentes químicos tóxicos que podem afetar a saúde dos laboratoristas e, posteriormente, exigem tratamento de resíduos específicos (Liu *et al.*, 2022). Esses métodos são também demorados, retardando, assim,

o diagnóstico sobre o estado nutricional das plantas e, conseqüentemente, atrasando a tomada de decisão. Por estas razões, as análises químicas tradicionais não são apropriadas para monitorar o status de N nas culturas e um método alternativo deve ser usado (Miao *et al.*, 2023).

A espectroscopia de infravermelho próximo (*Near Infrared* - NIR) é um método que surgiu como alternativa aos métodos analíticos tradicionais para a estimativa dos teores de nutrientes nos tecidos vegetais, de uma maneira mais rápida, menos onerosa e menos poluente (Vanmaarschalkerweerd & Husted, 2015). Metodologias que empregam a espectroscopia NIR têm se mostrado eficientes na substituição de métodos de análise laboratoriais em diversos setores como em análises de alimentos, produtos farmacêuticos e agrícolas entre outros (Pasquini, 2018).

A espectroscopia de infravermelho próximo (NIRS) é um tipo de espectroscopia vibracional de alta energia realizada na faixa de comprimento de onda de 750 a 2500 nm (Pasquini, 2018). Essa técnica baseia-se na absorção de radiação eletromagnética na região do infravermelho próximo, que corresponde às frequências de vibração das ligações C-H, O-H e N-H encontradas nos grupos funcionais de moléculas orgânicas. Alguns nutrientes, como, por exemplo, o N, fazem parte dessas moléculas orgânicas e interagem com os grupos funcionais, permitindo sua identificação e quantificação (Turker-Kaya & Huck, 2017). Para que a técnica seja empregada com boa acurácia e precisão, é necessário a utilização de pre-processamentos espectrais e a construção de um modelo de calibração multivariado associando os dados espectroscópicos NIR aos resultados analíticos realizados pelos métodos de referência (Prananto *et al.*, 2020).

Um dos desafios da metodologia NIRS é que sua aplicação geralmente é limitada a amostras populacionais fechadas, ou seja, mesmo ambiente e espécies (Chodack, 2008; Prananto *et al.*, 2020). Isso significa que as características espectrais dos tipos de amostra não incluídos no modelo de calibração podem interferir nas previsões do modelo. Estudos em folhas de árvores apoiam o potencial para um modelo NIRS amplo (Serbin *et al.*, 2014; Smis *et al.*, 2014). Recentemente Murguzur *et al.* (2017) construíram um modelo calibrado para diversas espécies em diferentes regiões geográficas e obtiveram resultados satisfatórios para estimativa de N, provando ser possível a construção de modelos globais que possam ser utilizados para mais de uma espécie e região. Portanto, levantamos a hipótese de que seja possível construir modelos de estimativa de N para múltiplas culturas agrícolas de importância econômica, com objetivo de aumentar a usabilidade do método, permitindo

sua aplicação em diversos contextos, economizando tempo e recursos.

Entre as principais culturas agrícolas produzidas no Rio Grande do Sul (RS), estão a soja, arroz, milho e trigo (IBGE, 2022). Portanto, é essencial modelos calibrados para a estimativa da concentração de N nessas culturas. Alguns trabalhos já fizeram calibrações de modelos estatísticos para predição de N nestas culturas, entretanto todos criam modelos para uma única cultura específica (Rodrigues *et al.*, 2020; Santos *et al.*, 2023; Miao *et al.*, 2023; Ge *et al.*, 2019; Tahmasbian *et al.*, 2021). O presente trabalho foi pioneiro na criação de um modelo abrangente que possa ser utilizado na estimativa de N para essas quatro culturas.

Os objetivos deste trabalho foram avaliar o uso da espectroscopia de infravermelho próximo para estimar a concentração de N em tecidos vegetais e grãos de trigo, soja, milho e arroz através de diferentes técnicas de pré-processamento e calibração multivariada, testar a construção de modelos globais para essas culturas e iniciar a construção de uma biblioteca espectral para tecidos vegetais.

## **3.2 Material e métodos**

### **3.2.1 Banco de amostras**

As amostras de material vegetal (planta e grãos) de trigo, milho e soja foram obtidas de experimentos conduzidos na Estação Experimental Agronômica da Universidade Federal do Rio Grande do Sul (EEA/UFRGS) e o material vegetal de arroz foi obtido na Estação Experimental Agronômica do Instituto Rio Grandense do Arroz (EEA/IRGA). A EEA/UFRGS localiza-se no município de Eldorado do Sul (RS), pertencente à região fisiográfica da Depressão Central do RS. O solo da área experimental é classificado como Argissolo Vermelho Distrófico típico (Bergamaschi *et al.*, 2003). A EEA/IRGA localiza-se no município de Cachoeirinha (RS), pertence à região da encosta inferior do nordeste do RS. Os solos encontrados na região são Planossolos caracterizado por horizonte superficial de cores claras, seguido de horizonte B plânico (Klamt *et al.*, 1985). Em associação aos Planossolos, há os Gleissolos, oriundos de minerais formados em condições de saturação por água, que ocorrem em áreas mais baixas e apresentam horizonte cinzento (Zanon *et al.*, 2018). As amostras utilizadas foram coletadas de experimentos conduzidos por estudantes dos Programas de Pós-graduação em Fitotecnia e Ciência do Solo da UFRGS, onde buscou-se ampliar a variabilidade dos teores de N em nossa biblioteca espectral (Tabela 1).

Para gerenciar as amostras, criou-se a Biblioteca Espectral de Tecidos Vegetais

(BETV), a qual possui um formato colaborativo e atualmente segue recebendo material vegetal. A BETV está sediada nas dependências do Grupo de Estudos em Agricultura Digital (GEAD – UFRGS) e tem como objetivo ser base para o desenvolvimento de novas ferramentas para determinação de teores nutricionais em tecidos vegetais, fomentando um novo polo de desenvolvimento e geração de conhecimento científico. Para o estudo, a BETV foi composta com amostras de tecido vegetal (material de estádios vegetativo e reprodutivo), separado em plantas, com 193 amostras de milho, 138 amostras de arroz e 135 amostras de trigo, totalizando 466 amostras. A BETV também foi composta por grãos de soja ( $n= 305$ ), grãos de trigo ( $n= 93$ ) e grãos de arroz ( $n= 73$ ), totalizando 471 amostras de grãos.

**TABELA 1.** Caracterização do banco de amostras de tecido vegetal em relação aos locais e as doses de nitrogênio.

<b>Cultura</b>	<b>Cultivares</b>	<b>Local</b>	<b>Safra</b>	<b>Doses de N (kg ha<sup>-1</sup>)</b>
Soja	BRS 5601 RR; BASF 15b70			0, 50, 100, 150, 200
Trigo	BRS Parrudo; BRS Reponde; TBIO Toruk; TBIO Sossego; BRS Bela Jóia	EEA/UFRGS	2019; 2020; 2021	0, 15, 30, 45, 60
Milho	B2418 VYHR P3016 VYHR			0, 25, 50, 75, 100, 120, 150, 200
Arroz	IRGA 424; IRGA 431	EEA/IRGA		0, 60, 120, 180, 240

As amostras foram secas em estufa com circulação de ar a 65-70 °C e moídas na peneira em frações de até 2,5 mm. Essa moagem foi realizada para aumentar a área superficial, bem como a densidade aparente, e porosidade, assim, removendo as barreiras da camada superficial para a radiação penetrar na matriz interior do material, dando uma maior chance de absorver a energia incidente durante a aquisição dos espectros NIR (Oyinloye & Yoon, 2020; Safarzadeh, 2021). Posteriormente, as amostras foram acondicionadas em sacos plásticos com lacre de vedação e armazenadas com sílica gel em caixas vedadas para a manutenção da baixa umidade relativa do ar (Figura 1).





**FIGURA 1.** Amostras acondicionadas na Biblioteca Espectral de Tecido Vegetal (BETV). Sacos plásticos com lacre de vedação e identificação das amostras (A e B); Amostras armazenadas em caixas vedadas com controle de umidade (C).

### 3.2.2 Análises químicas

As análises químicas para a determinar das concentrações de N total das amostras foram realizadas no Laboratório de Fisiologia Vegetal do Departamento de Plantas de Lavoura da Faculdade de Agronomia da UFRGS. A determinação dos teores de N foi realizada pelo método semi-micro de Kjeldahl descrito em Tedesco *et al.* (1995). O método consiste na solubilização úmida, seguida por destilação a vapor e titulação para a quantificação dos teores de  $\text{NH}_4^+$ . A solubilização sulfúrica ( $\text{H}_2\text{SO}_4$  + catalisadores) transforma proteína e aminoácidos do tecido vegetal em  $\text{N-NH}_4^+$ , que é destilado e complexado em ácido bórico indicador e, por fim, titulado com solução padronizada de  $\text{H}_2\text{SO}_4$  (Tedesco *et al.*, 1985).

### 3.2.3 Análise de espectroscopia de infravermelho próximo (NIR)

As análises espectroscópicas foram realizadas no Laboratório Federal de Defesa Agropecuária (LFDA – Ministério da Agricultura e Pecuária, MAPA), localizado em Porto Alegre (RS). Os espectros de infravermelho próximo (NIR) foram coletados na faixa de 780–2500 nm, utilizando o espectrômetro Bruker MPA, com a função de reflectância difusa com 60 leituras por espectro. Para realização das leituras espectrais, foi utilizado placas de petry com 10 g de cada amostra.

### 3.2.4 Pré-processamentos espectrais

Após a aquisição dos dados espectrais das amostras, os espectros brutos (RAW) foram submetidos a três técnicas de pré-processamento espectral, sendo a *Detrend*, derivada de Savitski-Golay e a *Standard Normal Variate*. A técnica *Detrend* (DET) permite destacar as variações espectrais mais relevantes, reduzindo o impacto de fatores não relacionados à composição química das amostras (Barnes et al., 1989). A técnica *Derivada de Savitski-Golay* (SGD) tem como objetivo suavizar e diferenciar os dados espectrais, reduzindo o ruído e preservando características relevantes (Savitzky & Golay, 1964). E a técnica *Standard Normal Variate* (SNV) que normaliza cada espectro de forma individual, subtraindo a média do espectro e dividindo pelo desvio padrão, com objetivo de reduzir o impacto de variações indesejadas, como espalhamento da luz e desvio instrumental, permitindo uma análise mais precisa das características espectrais relacionadas aos componentes químicos das amostras (Barnes et al., 1989).

A definição dessas três técnicas foi baseada nas diferenças que existem no processo matemático que cada uma utiliza para transformar os dados espectrais. Todos os pré-processamentos foram realizados utilizando os pacotes *prospectrs* e *clusterSim* no software R (Stevens & Ramirez-Lopez, 2020; Walesiak & Dudek, 2020; R Core Team, 2020).

### 3.2.5 Análise Discriminante

A análise discriminante foi realizada antes da construção dos modelos, visando aprofundar a compreensão das diferenças existentes entre os conjuntos de amostras considerados. A análise discriminante busca encontrar uma combinação linear das variáveis que maximize a separação entre grupos (Klecka, 1980). A análise discriminante possibilita encontrar padrões entre os conjuntos de amostras, servindo como fonte de informação para tomada de decisão sobre quais conjuntos de dados usar em cada modelo.

### 3.2.6 Análise dos componentes principais (PCA)

A análise de componentes principais (PCA), com objetivo de diferenciar os conjuntos de amostras e reduzir a dimensionalidade dos dados. Essa técnica é realizada buscando combinações lineares das variáveis originais, chamadas de componentes principais, que capturam a maior parte da variabilidade dos dados (Martens & Naes, 1992). A redução da dimensionalidade simplifica a interpretação dos resultados e melhora

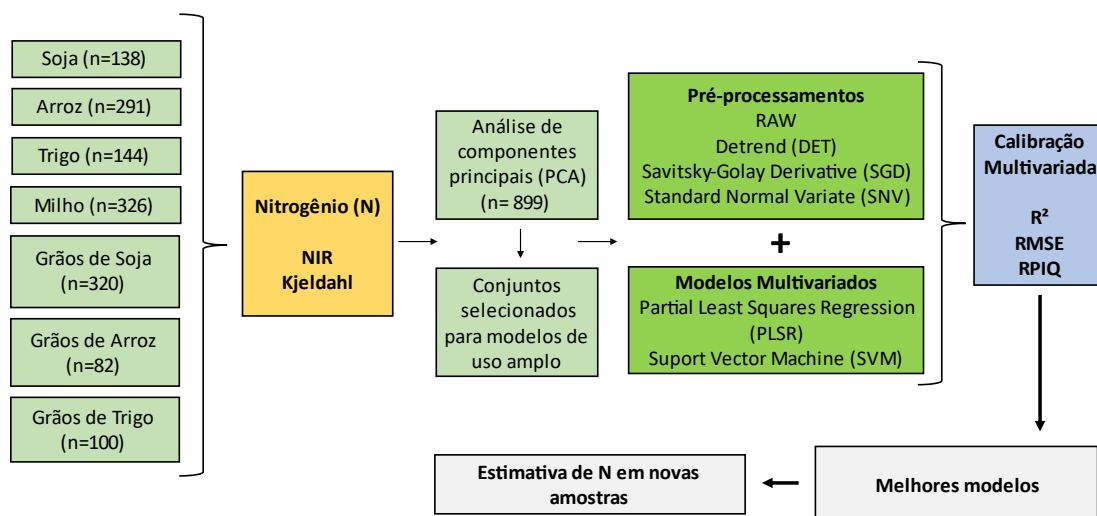
a eficiência dos modelos.

### 3.2.7 Calibração multivariada

A calibração multivariada foi realizada no banco de dados utilizados, composto por amostras de milho (n = 326), arroz (n = 291), trigo (n = 144) e soja (n = 138), totalizando 899 amostras. Os modelos multivariados foram calibrados e correlacionados com os teores de N das plantas, avaliada pelos dois métodos (químico e espectroscópico). Na calibração dos modelos, foram utilizados dois métodos multivariados, sendo um modelo paramétrico (Partial Least Squares Regression - PLSR) e outro não paramétrico (Support Vector Machine - SVM). Foram utilizados os pacotes do software *R pls e 1071* do software R, respectivamente (Mevik *et al.*, 2016; Meyer *et al.*, 2019).

Optou-se por utilizar os diferentes modelos devido à ocorrência de correlações lineares e não lineares entre o teor de N e as variáveis espectrais (Rossel & Behrens, 2010). O método PLSR reduz grandes conjuntos de variáveis preditoras a um pequeno número de fatores ortogonais não correlacionados para minimizar a soma dos quadrados dos erros das estimativas (Varmuza & Filzmoser, 2016). O método SVM foi utilizado com a função kernel, que separa os dados de calibração em hiperplanos e busca estabelecer correlações entre as variáveis dependentes e independentes quando estas apresentam comportamento não linear (Ivanciuc, 2007).

A calibração dos modelos com PLSR e SVM para as amostras de planta foi feita combinando os três pré-processamentos espectrais, sendo *Detrend* (DET), *Derivada de Savitski-Golay* (SGD), *Standard Normal Variate* (SNV) e os espectros brutos (RAW). Todos os modelos foram calibrados por validação cruzada usando uma divisão aleatória em 10-fold (Brus *et al.*, 2011). Após a modelagem e interpretação dos resultados dos modelos calibrados, selecionando os modelos mais acurados e que possibilitaram a predição de N (Figura 2).



**FIGURA 2.** Esquema das etapas realizadas para a calibração das análises multivariadas para a estimativa dos teores de N no tecido vegetal.

O desempenho dos modelos foi avaliado pelo coeficiente de determinação ( $R^2$ ) (Equação 1), relação entre o desempenho e a distância interquartil (RPIQ) (Equação 2), erro médio quadrático de predição (RMSE) (Equação 3) e Bias (Equação 4) (Johnson *et al.*, 2021; Miao *et al.*, 2023; Murguzur *et al.*, 2017; Prananto *et al.*, 2020; Santos *et al.*, 2023).

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

$$RPIQ = \frac{(Q3 - Q1)}{RMSE} \quad (2)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (3)$$

$$Bias = \frac{1}{N} \sum_{i=1}^n (\hat{y}_i - y_i)^2 \quad (4)$$

Onde: n = número total de observações;  $y_i$  = valor observado;  $\bar{y}_i$  = média das observações;  $\hat{y}_i$  = valor estimado; Q3 = terceiro quartil; Q1 = primeiro quartil.

Para a interpretação, utilizou-se a classificação de qualidade dos modelos sugerida por Johnson *et al.* (2019), onde:  $R^2 > 0,75$ , indicam modelos com nível de predição alto;

$0,50 \leq R^2 \leq 0,75$ , indicam modelos com nível de predição médio; e  $R^2 < 0,50$ , são modelos não recomendados. Os métodos de RMSE avalia as diferenças entre os valores previstos e os valores medidos e quantifica a precisão comparando os erros de previsão de diferentes modelos. Já o RPIQ considera o erro de previsão em relação a variação dos valores medidos (Johnson *et al.*, 2021). O bias "positivo" indicam predições superestimadas e o bias "negativo" indicam predições subestimadas, e quando próximo de zero indica melhor qualidade do modelo (Rodrigues *et al.*, 2020). Considerou-se um alto nível de confiabilidade  $RPIQ \geq 3,0$ ; médio nível de confiabilidade  $\geq 1,9$  e  $< 3,0$ , e baixo nível de confiabilidade  $\geq 1,5$  e  $< 1,9$  (Veum *et al.*, 2015; Chang *et al.*, 2001). Os parâmetros de acurácia ( $R^2$ , RMSE, RPIQ e Bias) obtidos com os diferentes métodos de calibração, pré-processamentos espectrais e conjunto de dados foram comparados pelo teste de média de Tukey ao nível de significância de  $p < 0,05$  utilizando o software R (R CORE TEAM, 2020). Todas as análises e modelagens foram implementadas no software R (R Core Team, 2020).

### **3.3 Resultados e discussão**

#### **3.3.1 Análise descritiva do conjunto de dados**

A variabilidade da concentração de N nas amostras analisadas pelo método químico padrão foram resumidas e apresentadas na forma de análise descritiva na Tabela 2. A amplitude na concentração de N nas espécies analisadas ocorreu devido as doses de adubação nitrogenada nos experimentos conduzidos nas EEA/UFRGS e EEA/IRGA. Esta variação de concentração de N é considerada importante para que o desenvolvimento de modelos possam ser robustos do ponto de vista de calibração e validação (Johnson *et al.*, 2021).

Os teores de N nas plantas de trigo variaram entre 1,04% e 5,86%, valores semelhantes aos encontrados por Ecarnot *et al.* (2013), cujas variações ficaram entre 0,40% e 4,94%. Além disso, é relevante notar que o teor médio de 2,55% de N nas amostras de planta de arroz é próximo ao valor obtido por Miao *et al.* (2023), que registraram uma média de 2,37% em um conjunto de 295 amostras. Para o milho, os teores de N variaram entre 0,95% a 4,53%, resultados semelhantes aos encontrados por Wen *et al.* (2019), que encontraram resultados variando entre 1,10% e 4,07%. O valor médio da concentração de N nas amostras de grão de arroz foi de 1,13%, sendo semelhante ao utilizado por Mishra *et al.* (2021), que validou modelos em um conjunto de amostras com

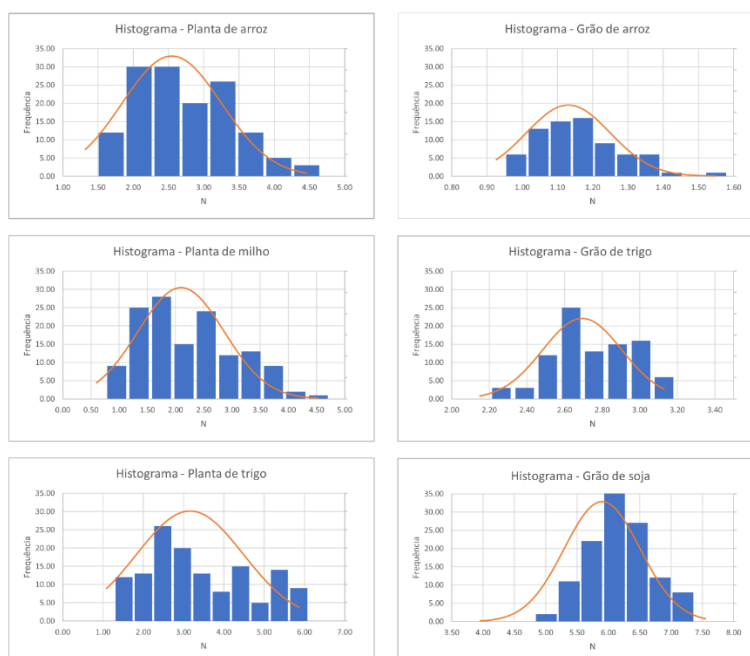
teores médios de 1,39%. A concentração média de N nas amostras de grão de soja foi de 5,90%, semelhante à média de 6,54% disponíveis na literatura (Shi *et al.*, 2022). Para as amostras de trigo a média encontrada foi de 2,69%, sendo uma média acima dos valores encontrados por Kamboj *et al.* (2022), que foram de 1,80%.

**TABELA 2.** Estatística descritiva da concentração de N nas amostras de planta inteira e grãos.

Estatística Descritiva								
Material	$n^1$	Mínimo	Máximo	Média	Mediana	Desvio Padrão	1° Q <sup>2</sup>	3° Q <sup>2</sup>
Arroz Grão	73	0,92	1,55	1,13	1,11	0,12	1,04	1,19
Soja Grão	305	3,92	7,55	5,90	5,92	0,60	5,42	6,31
Trigo Grão	93	2,14	3,13	2,69	2,67	0,21	2,57	2,86
Arroz planta	138	1,29	4,46	2,55	2,38	0,70	1,97	3,14
Milho planta	193	0,95	4,53	2,15	2,10	0,72	1,61	2,64
Trigo Planta	135	1,04	5,86	3,16	2,87	1,32	2,08	4,18

$n$  - Número de amostras. <sup>2</sup> Quartil.

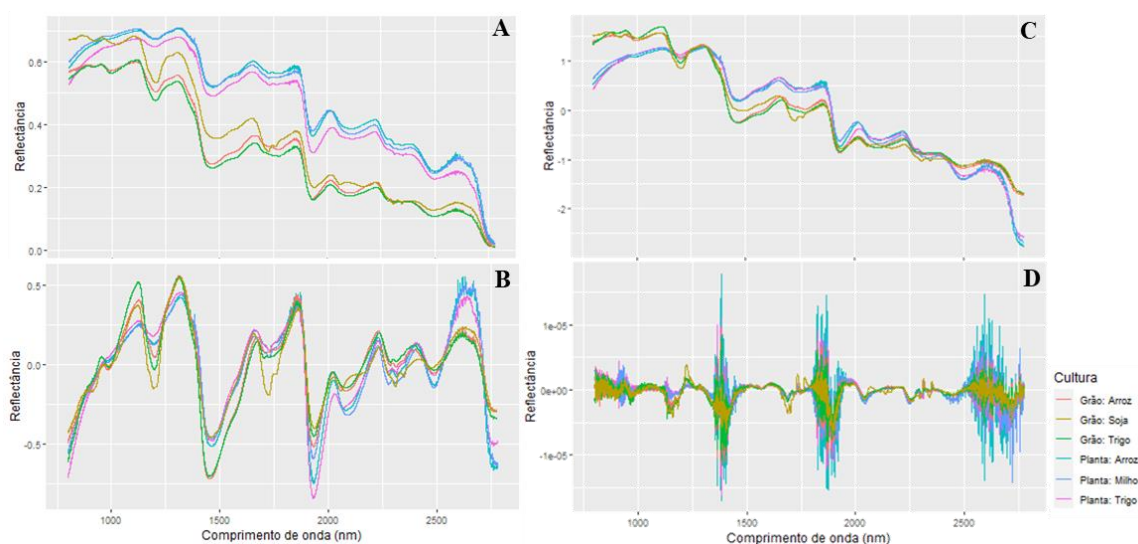
O histograma das concentrações de N nas amostras analisadas é apresentado na Figura 3, onde pode-se observar que a maior quantidade de amostras de planta tiveram concentração de N variando entre 1% e 4%, enquanto a maior parte das amostras de grãos tiveram a concentração de N variando entre 1% e 7%.



**FIGURA 3.** Histograma dos teores de N das amostras de grão (trigo, soja e arroz) e planta (trigo, milho e arroz).

### 3.3.2 Análise dos espectros de infravermelho próximo (NIR)

Na Figura 4 são apresentadas as médias das assinaturas espectrais (reflectâncias) das amostras de planta e de grão, com e sem a utilização dos pré-processamentos. Foi verificado que mesmo entre espécies diferentes, houve comportamento espectral semelhante. Cada espécie de planta tem um espectro de reflectância ou absorção único nas faixas do NIR (Prananto *et al.*, 2020), entretanto, existe uma tendência geral para as plantas, que é resultado de um complexo padrão de dispersão e absorção causados por componentes bioquímicos e estruturais presentes no tecido vegetal (Richardson *et al.*, 2004). Os espectros de reflectância no NIR são afetados por sobretons moleculares e combinações vibracionais relacionadas ao alongamento e dobramento das ligações moleculares como O–H, N–H e C–H encontradas em grupos funcionais de moléculas orgânicas como água, carboidratos, lipídeos e proteínas (Das *et al.*, 2020).



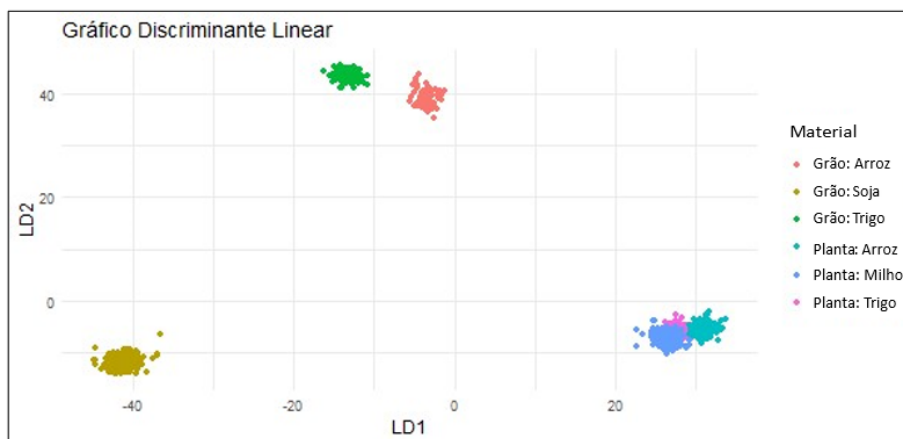
**FIGURA 4.** Assinatura espectral média das amostras de planta e grão sem pré-processamento (RAW) (A), pré-processada com *Detrend* (DET) (B), *Standard Normal Variate* (SNV) (C), *Savitski-Golay Derivative* (SGD) (D).

A utilização do pré-processamento com a técnica DET removeu a tendência sistemática presente nos espectros e destacou as bandas com maiores e menores reflectâncias, reduzindo o impacto de fatores não relacionados à composição química das plantas (Barnes *et al.*, 1989). A técnica SGD usou a primeira derivada para calcular a taxa de variação da reflectância em relação aos comprimentos de onda, resultando em

transformações nos espectros. Assim, observou-se que os espectros apresentaram feições realçadas e o sinal de linha de base das amostras foi removido (Rinnan *et al.*, 2009). A técnica SNV suavizou cada espectro individualmente, subtraindo a média do espectro e dividindo pelo desvio padrão. Essa transformação padroniza os espectros em termos de distribuição normal, tornando-os mais comparáveis e realçando as variações espectrais relevantes (Barnes *et al.*, 1989). Por isso, as assinaturas das diferentes espécies ficaram mais próximas no SNV em comparação ao RAW.

### 3.3.3. Análise discriminante

A análise discriminante auxiliou na identificação e no entendimento dos dados obtidos, permitindo identificar os padrões e diferenças entre os grupos de amostras (Figura 5). Com a análise foi possível verificar que os grupos compostos por plantas tiveram uma distribuição semelhante, enquanto os grupos compostos por grãos se diferenciaram, sendo as amostras de grãos de trigo e arroz semelhantes, porém diferentes das amostras de grão de soja. A análise contribuiu para a compreensão das complexidades presentes nos dados, e serviu como base para a criação de modelos adaptados a grupos específicos.



**FIGURA 5.** Análise discriminante entre os grupos compostos por amostras de grão (trigo, soja e arroz) e planta (trigo, milho e arroz).

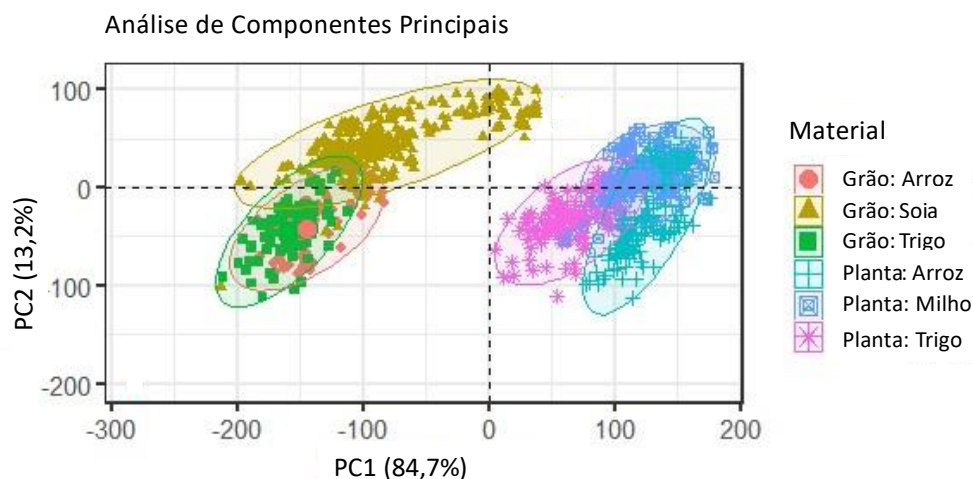
### 3.3.4 Análise de Componentes Principais (PCA)

A análise de componentes principais (PCA) foi realizada no conjunto de dados



composto por amostras de plantas de milho, arroz e trigo e de grãos de soja, arroz e trigo. A distribuição dos pontos no gráfico da PCA, mostrou as semelhanças entre os grupos analisados, onde sugere-se uma relação próxima em termos de composição química e características espectrais (Figura 6). O 1º componente explicou 84,7% da variação dos dados, enquanto o 2º componente explicou 13,2%, o que explicou 97,9% da relação. Da mesma forma, as amostras dos diferentes grãos também apresentaram uma proximidade entre si. No entanto, ao realizar uma comparação entre os grupos de plantas e grãos, observou-se um comportamento distante na distribuição dos dados, sugerindo características distintas entre esses dois tipos de material.

Os resultados indicam que a construção de um modelo único para plantas e outro modelo para os grãos seria a estratégia mais adequada, pensando em construir modelos multiespécie, com potencial de proporcionar uma solução mais eficiente na estimativa de N.



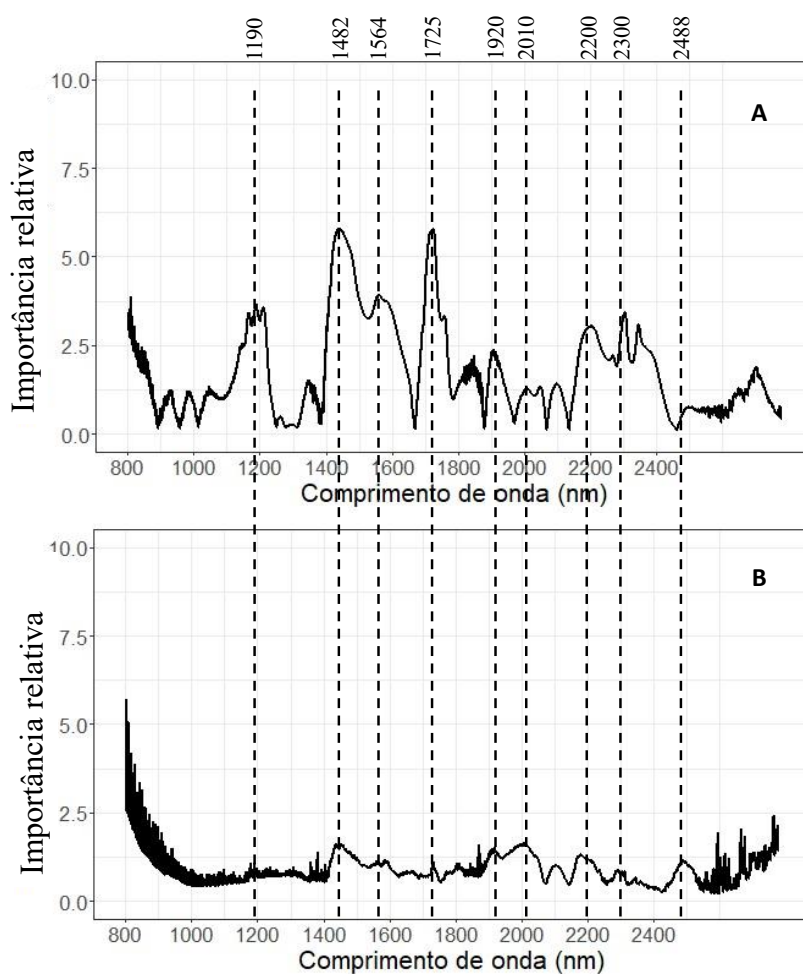
**FIGURA 6.** Análise de componentes principais (PCA) com os dados espectrais para os grupos compostos por amostras de grão (trigo, soja e arroz) e planta (trigo, milho e arroz).

### 3.3.5 Importância das bandas espectrais para predição de N

Na Figura 7, são apresentados os scores das bandas do NIR mais relevantes para o processo de modelagem. As bandas de absorção a 1482 nm e 1725 nm apresentaram os maiores scores ( $\sim 5,5$ ) durante o processo de criação do modelo para grãos. Esses comprimentos de onda correspondem às vibrações das ligações O–H e C–H presentes em grupos funcionais de carboidratos (Turker-Kaya & Huck, 2017). A banda em 1564 nm

apresentou uma importância relativa de  $\sim 3,75$  e está relacionada com ligações N–H presentes em proteínas (Turker-Kaya & Huck, 2017). Os comprimentos de onda de 2200 nm e 2300 nm correspondem às vibrações entre C–H de moléculas de carboidrato e proteínas, e tiveram uma importância relativa de 3,12 e 3,40 (Turker-Kaya & Huck, 2017; Aenugu *et al.*, 2011).

Para a modelagem feita com as amostras de planta a diferença entre os scores foi menor, sendo que os comprimentos de onda de 1482, 1820, 1920 e 2010 nm, tiveram as maiores importâncias, ficando próximo a 1,60. Esses comprimentos de onda se relacionam as ligações de O–H de carboidratos, O–H da celulose, C–H de carboidratos e O–H + C–O do amido respectivamente (Turker-Kaya & Huck, 2017; Aenugu *et al.*, 2011).



**FIGURA 7.** Importância das bandas espectrais na a criação de modelos para estimativa de N em grãos (A) e plantas (B).

### 3.3.6 Construção dos modelos de calibração

#### 3.3.6.1 Plantas

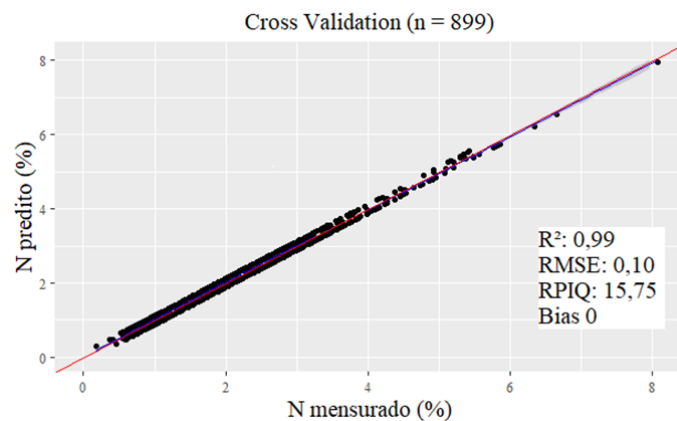
A calibração multivariada foi realizada com os modelos estatísticos Regressão de Mínimos Quadrados Parciais (PLSR) e Máquina de Vetores de Suporte (SVM). Além disso, testou-se as técnicas (SNV, DET e SGD) de pré-processamento nos dados espectrais (Tabela 4).

**TABELA 3.** Parâmetros de acurácia dos modelos PLSR e SVM para estimativa de nitrogênio nos modelos construídos com amostras de planta para as culturas de arroz, milho e trigo.

<b>Modelo</b>	<b>Pré-processamento</b>	<b>R<sup>2</sup></b>	<b>RMSE</b>	<b>RPIQ</b>	<b>Bias</b>
PLSR	RAW	0,94	0,25	5,55	0,00
PLSR	DET	0,94	0,25	5,49	0,00
PLSR	SNV	0,94	0,25	5,59	0,00
PLSR	SGD	0,94	0,25	5,53	0,00
SVM	RAW	0,98	0,15	11,31	0,00
SVM	DET	0,99	0,11	15,59	0,00
SVM	SNV	0,99	0,10	15,75	0,00
SVM	SGD	0,99	0,11	15,47	0,00

PLSR: Partial Least Squares Regression; SVM: Suport Vector Machine. SGD: Savitzky–Golay Derivative; SNV: Standard Normal Variate; DET: Detrend.

De acordo com o modelo de classificação sugerida por Johnson *et al.* (2019), os modelos que apresentaram altos valores de  $R^2$  ( $> 0,90$ ), indicam variações nas concentrações de N, que podem ser bem estimadas pelos modelos com base em dados espectrais. As calibrações feitas com os dados espectrais sem pré-processamentos (RAW) tiveram os menores valores de  $R^2$ , sendo 0,94 para PLSR e 0,98 para SVM. Foi observado que as calibrações feitas com SVM foram superiores as feitas com PLSR, sendo que o melhor modelo de todo o conjunto foi SVM com o pre-processamento SNV onde foi obtido o  $R^2$  de 0,99, RMSE de 0,10, RPIQ de 15,75 e Bias 0 (Figura 8).



**FIGURA 8.** Gráfico de dispersão do conteúdo de N medido e estimado nas amostras de arroz, trigo e milho, utilizando o algoritmo SVM com o pré-processamento espectral SNV.

Estudos recentes desenvolveram modelos de predição de N para as culturas do arroz, milho, trigo e soja, porém os modelos foram desenvolvidos de forma individualizada para cada espécie. Johnson *et al.* (2021), calibraram modelos PLSR para a cultura do arroz utilizando o pré-processamento SGD e foi alcançado parâmetros de  $R^2$  variando de 0,81 a 0,87, de RPIQ variando de 3,62 a 4,41 e RMSECV variando de 0,14 a 0,17. Calibrações feitas com PLSR alcançaram  $R^2 > 0,94$  quando combinadas com os pré-processamentos *Standard Normal Variate* (SNV) e *First Derivative* (FD) (Miao *et al.*, 2023), resultados semelhantes ao encontrados no presente estudo, comprovando a precisão de NIRS em estimar N em plantas de arroz.

Ecarnot *et al.* (2013), calibraram modelos PLSR para predição de N em amostras secas e frescas de trigo, e obtiveram  $R^2 > 0,93$  com  $RMSE < 0,32$ . Tahmasbian *et al.* (2021), calibraram modelos PLSR para o trigo com  $R^2 = 0,96$  e  $RMSE = 0,92$  sem a utilização de pré-processamentos espectrais. Já modelos de aprendizado de máquina foram mais precisos para predição de N na cultura do trigo ( $R^2 = 0,96$  e  $RMSE = 0,21$ ), com o método SVM sendo mais estável em comparação ao PLSR (Yao *et al.*, 2015). Pandey *et al.* (2017), calibraram modelos PLSR com espectroscopia NIR para a predição de N na cultura do milho e obtiveram  $R^2 = 0,88$  e  $RMSE = 0,47$ . Silva *et al.* (2023), obtiveram  $R^2 = 0,80$ ,  $RMSE = 0,33$ ,  $RPIQ = 1,92$  e  $Bias = 0,0$  utilizando o modelo PLSR. Ge *et al.* (2019), encontraram padrões semelhantes ao desenvolver modelos PLSR para o milho com  $R^2 = 0,87$  e  $RMSE = 0,25$ .

A espectroscopia NIR demonstrou ser eficiente na estimativa da concentração de

N em tecido vegetal, apresentando forte correlação entre o teor de N das plantas com os espectros da região NIR (Prananto *et al.*, 2020; Johnson *et al.*, 2021; Miao *et al.*, 2023; Tahmasbian *et al.*, 2021; Silva *et al.*, 2023). Os resultados encontrados neste estudo, indicam que é possível a construção de um modelo único para as três culturas (arroz, trigo e milho) e oferece uma abordagem mais eficiente e versátil para a predição do teor de N, permitindo uma aplicação mais ampla e simplificada em diferentes cenários agrícolas.

### 3.3.6.2. Grãos

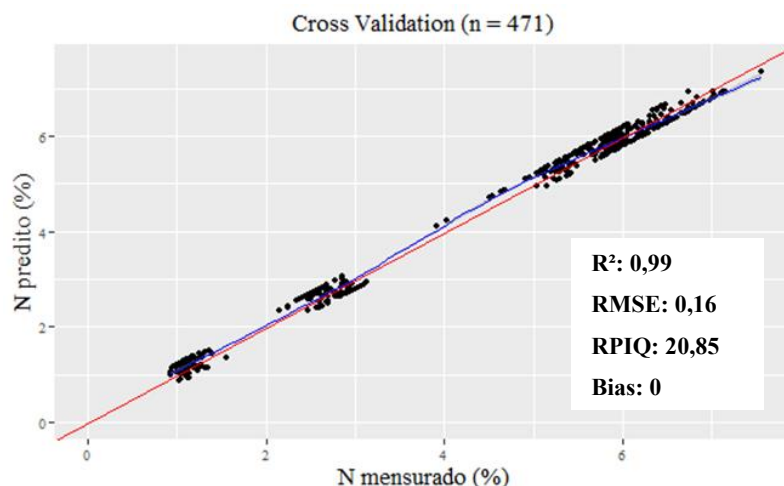
A calibração multivariada foi realizada com os modelos estatísticos Regressão de Mínimos Quadrados Parciais (PLSR) e Máquina de Vetores de Suporte (SVM). Além disso, testou-se as técnicas (SNV, DET e SGD) de pré-processamento nos dados espectrais (Tabela 5). Todas calibrações apresentaram  $R^2 = >0,97$ , indicando que os modelos podem ser utilizados para as técnicas de pré-processamento (Johnson *et al.*, 2019). O modelo de PLSR apresentou os melhores parâmetros estatísticos quando associado com o pré-processamento SNV. Já o modelo SVM obteve o melhor pré-processamento com o SNV.

**TABELA 4.** Parâmetros de acurácia dos modelos PLSR e SVM para estimativa de nitrogênio nos modelos construídos com amostras de grãos de arroz, trigo e soja.

<b>Modelo</b>	<b>Pré-processamento</b>	<b>R<sup>2</sup></b>	<b>RMSE</b>	<b>RPIQ</b>	<b>Bias</b>
PLSR	RAW	0,97	0,34	10,02	0,00
PLSR	DET	0,98	0,30	11,25	0,00
PLSR	SNV	0,98	0,28	11,88	0,00
PLSR	SGD	0,97	0,32	10,48	0,00
SVM	RAW	0,99	0,17	19,88	0,00
SVM	DET	0,99	0,16	20,70	0,00
SVM	SNV	0,99	0,16	20,85	0,00
SVM	SGD	0,99	0,16	20,78	0,00

PLSR: Partial Least Squares Regression; SVM: Suport Vector Machine. SGD: Savitzky–Golay Derivative; SNV: Standard Normal Variate; DET: Detrend.

Com os parâmetros encontrados nos modelos construídos, observou-se que as calibrações feitas com o SVM foram superiores às com PLSR, sendo que o melhor modelo de todo o conjunto foi SVM com o pré-processamento SNV (Figura 9).



**FIGURA 9.** Gráfico de dispersão do conteúdo de N medido e estimado utilizando o modelo SVM com pré-processamento SNV nas amostras de grãos.

Existe um número reduzido de trabalhos disponíveis na literatura que relacionam a espectroscopia NIR para estimativa de N em grãos, destacando o pioneirismo dos resultados. Porém, na literatura outros estudos utilizam o NIR em grãos para realizar a predição de proteína. Kawamura *et al.* (2003), utilizaram o NIR para detectar o teor de proteína em grãos de arroz com modelo de predição para o arroz integral de  $R^2 = 0,70$ , e para arroz beneficiado de  $R^2 = 0,76$ . A predição do teor de proteína e amilase em farinha de arroz utilizando o modelo *Modified Partial Least Squares* (MPLS) e pré-processamento *Multiple Scatter Correction* (MSC) foi destacado por Xie *et al.* (2014). O teor de proteína em amostras de farinha de arroz foi previsto com  $R^2 = 0,94$  e o teor de amilose foi previsto para farinha de arroz integral ( $R^2 = 0,75$ ). O potencial preditivo da espectroscopia NIR para três formas físicas de arroz: grão de arroz, arroz integral e arroz em pó, com a utilização do algoritmo PLSR sem pré-processamentos observou-se o menor RMSE para a farinha de arroz (0,36%) (Mishra *et al.*, 2021).

A composição de glúten em farinha de trigo foi estimada por espectroscopia NIR utilizando modelos PLS com pré-processamento First Derivative (FD) e os resultados foram considerados aceitáveis com base no RMSE de 6,09 mg/g (Schuster *et al.*, 2023). O teor de proteína em grãos de trigo foi estimado por NIR utilizando modelos PLS e os resultados encontrados indicam uma alta precisão de estimativa do teor de proteína ( $R^2 = 0,97$ , RMSE 0,093 e RPD = 10,15) (Huan *et al.*, 2021). O modelo PLS para predição do teor de proteína em grãos de trigo obteve-se  $R^2 = 0,90$ , RMSE = 0,47 e Bias 0,03 (Mishra & Lohme, 2021).

Shi *et al.* (2022), determinaram as concentrações de aminoácidos e proteína bruta na soja com técnicas NIR. Os modelos de calibração foram feitos com PLS utilizando os pré-processamentos SNV e DET e a maioria dos aminoácidos e proteína apresentaram  $R^2$  de 0,95 e 0,90, respectivamente.

Chadalavada *et al.* (2022), desenvolveu modelos para serem utilizados em múltiplas espécies de grãos ( $R^2 \geq 0,90$ ,  $RMSE \leq 0,91$ ,  $RPD \geq 3,08$ ), porém sem a inclusão de grãos de arroz e trigo. No geral, os estudos alcançam as métricas de precisão dos métodos de estimativa ( $R^2 \geq 0,85$ ;  $RPD > 3,0$ ) sendo compatíveis com os resultados encontrados no presente estudo.

Os resultados demonstraram que é possível desenvolver modelos abrangentes capazes de abranger múltiplas espécies, possibilitando alcançar e aplicar a técnica de modo amplo. Em resumo, os resultados deste estudo apontam para a viabilidade e a importância de desenvolver modelos de predição para múltiplas espécies. Um dos principais benefícios dessa técnica é de realizar a estimativa do teor de proteínas além do teor de nitrogênio, pelo fato do N ser componente fundamental das proteínas.

### 3.5 Conclusão

A estimativa da concentração de nitrogênio em plantas de milho, arroz e trigo, e grãos de soja, trigo e arroz foi mais precisa utilizando o modelo de calibração multivariado *Support Vector Machine* (SVM) combinado com o pré-processamento *Standard Normal Variate* (SNV). Adicionalmente, foi verificado que é possível a calibração de um modelo único para estimar a concentração de nitrogênio nas plantas (milho, arroz e trigo) e grãos (soja, arroz e trigo), não havendo necessidade de criar modelos específicos para cada espécie.

Em suma, este estudo abre novos horizontes ao demonstrar a viabilidade e as vantagens de um modelo amplo de predição para múltiplas culturas agrícolas. Com resultados promissores, o modelo proposto pode ser um impulsionador para pesquisas futuras na busca por uma abordagem mais integrada e abrangente na estimativa de parâmetros agrícolas.

### 3.6 Referências

AENUGU, H, P, R. *et al.* Near infrared spectroscopy - An overview. **International Journal of Chem Tech Research**, Oxford, v. 3, n. 2, p. 825-836, 2011.

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, Plainfield, v. 43, p. 772-777, 1989.

BRUS, D. J.; KEMPEN, B.; HEUVELINK, G. B. M. Sampling for validation of digital soil maps. **European Journal of Soil Science**, Oxford, v. 62, n. 3, p. 394-407, 2011.

CHADALAVADA, K. *et al.* NIR instruments and prediction methods for rapid access to grain protein content in multiple cereals. **Sensors**, Basel, v. 22, n. 10, p. 3710, 2022.

CHANG, C. W. *et al.* Near-infrared reflectance spectroscopy–principal components regression analyses of soil properties. **Soil Science Society of America Journal**, Madison, v. 65, n. 2, p. 480-490, 2001.

DAS, B. *et al.* Spectroscopy based novel spectral indices, PCA-and PLSR-coupled machine learning models for salinity stress phenotyping of rice. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, Amsterdam, v. 229, p. 117983, 2020.

SANTOS, D. G. L. A. A. *et al.* Spectral method for macro and micronutrient prediction in soybean leaves using interval partial least squares regression. **European Journal of Agronomy**, Amsterdam, v. 143, p. 126717, 2023.

ECARNOT, M; COMPAN, F; ROUMET, P. Assessing leaf nitrogen content and leaf mass per unit area of wheat in the field throughout plant cycle with a portable spectrometer. **Field Crops Research**, Amsterdam, v. 140, p. 44-50, 2013.

GE, Y. *et al.* High-throughput analysis of leaf physiological and chemical traits with VIS–NIR–SWIR spectroscopy: a case study with a maize diversity panel. **Plant Methods**, New York, v. 15, n. 1, p. 1-12, 2019.

HUANG, C. H. *et al.* Early diagnosis and management of nitrogen deficiency in plants utilizing Raman spectroscopy. **Frontiers in Plant Science**, Lausanne, *11*, 663, 2020.

HUAN, K. *et al.* Variable selection in near-infrared spectra: Application to quantitative non-destructive determination of protein content in wheat. **Infrared Physics & Technology**, Washington, v. 119, p. 103937, 2021.

IVANCIUC, O. *et al.* Applications of support vector machines in chemistry. **Reviews in Computational Chemistry**, Weinheim, v. 23, p. 291, 2007.

JIE, W. *et al.* Determination of Nitrogen Concentration in Fresh Pear Leaves by Visible/Near-Infrared Reflectance Spectroscopy. **Agronomy Journal**, v. 106, n. 5, p. 1867-1872, 2014.



JOHNSON, J.B. An overview of near-infrared spectroscopy (NIRS) for the detection of insect pests in stored grains. **Journal of Stored Products Research**, Oxford, v. 86, p. 101558, 2020.

JOHNSON, J. M. *et al.* Application of infrared spectroscopy for estimation of concentrations of macro-and micronutrients in rice in sub-Saharan Africa. **Field Crops Research**, Amsterdam, v. 270, p. 108222, 2021.

KAMBOJ, U; GUHA, P; A, S. Comparison of PLSR, MLR, SVM regression methods for determination of crude protein and carbohydrate content in stored wheat using near Infrared spectroscopy. **Materials Today: Proceedings**, London, v. 48, p. 576-582, 2022.

KAWAMURA, S. *et al.* Development of an automatic rice-quality inspection system. **Computers and Electronics in Agriculture**, Amsterdam, v. 40, p. 115-126, 2003.

KLECKA, W. R. **Discriminant analysis**. London: Sage, 1980.

LIU, W. *et al.* Non-destructive measurements of toona sinensis chlorophyll and nitrogen content under drought stress using near infrared spectroscopy. **Frontiers in Plant Science**, Lausanne, v. 12, p. 809828, 2022.

MARIOTTI, F; TOMÉ, D; MIRAND, P. P. Converting nitrogen into protein—beyond 6.25 and Jones' factors. **Critical Reviews in Food Science and Nutrition**, Philadelphia, v. 48, n. 2, p. 177-184, 2008.

MARTENS, H; NAES, T. **Multivariate calibration**, John Wiley & Sons, New York, 1992.

MEVIK, B. H.; WEHRENS, R.; LILAND, K. H. **Partial least squares and principal component regression**, Vienna, 2016.

MEYER, D. *et al.* e1071: misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien. **R package version**, Vienna, v. 1, n. 2, 2019.

MIAO, X.X. *et al.* Measurement of nitrogen content in rice plant using near infrared spectroscopy combined with different PLS algorithms. **Spectrochemical Acta Part a: Molecular and Biomolecular Spectroscopy**, Amsterdam, v. 284, p. 121733, 2023.

MISHRA, P; ANGILERI, M; WOLTERING, E. Identifying the best rice physical form for non-destructive prediction of protein content utilising near-infrared spectroscopy to support digital phenotyping. **Infrared Physics & Technology**, Southampton, v. 116, p. 103757, 2021.

MISHRA, P; LOHUMI, S. Improved prediction of protein content in wheat kernels with a fusion of scatter correction methods in NIR data modelling. **Biosystems Engineering**, London, v. 203, p. 93-97, 2021.

OYINLOYE, T. M; YOON, W.B. Effect of freeze-drying on quality and grinding process of food produce: A review. **Processes**, Basel, v. 8, n. 3, p. 354, 2020.

PANDISELVAM, R. *et al.* Recent applications of vibrational spectroscopic techniques in the grain industry. **Food Reviews International**, London, v. 39, n. 1, p. 209-239, 2023.

PANDEY, P. *et al.* High throughput in vivo analysis of plant leaf chemical properties using hyperspectral imaging. **Frontiers in Plant Science**, Lausanne, v. 8, p. 1348, 2017.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives—A review. **Analytica Chimica Acta**, Amsterdam, v. 1026, p. 8-36, 2018.

PRANANTO, J. A; MINASNY, B; WEAVER, T. Near infrared (NIR) spectroscopy as a rapid and cost-effective method for nutrient analysis of plant leaf tissues. **Advances in Agronomy**, San Diego, v. 164, p. 1-49, 2020.

Pu, Y.Y. *et al.* Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing. **International Dairy Journal**, Champaign, v. 103, p. 104623, 2020.

R CORE TEAM. **R: a language and environment for statistical computing**. Vienna: R Foundation for Statistical Computing, 2020. Disponível em: <http://www.R-project.org/>. Acesso em: 4 mar. 2021.

RICHARDSON, D.; REEVES III, J. B.; GREGOIRE, T.G. Multivariate analyses of visible/near infrared (VIS/NIR) absorbance spectra reveal underlying spectral differences among dried, ground conifer needle samples from different growth environments. **New Phytologist**, Oxford, v. 161, n. 1, p. 291-301, 2004.

RODRIGUES, M. *et al.* Vis–NIR spectroscopy: from leaf dry mass production estimate to the prediction of macro-and micronutrients in soybean crops. **Journal of Applied Remote Sensing**, Basel, v. 14, n. 4, p. 044505-044505, 2020.

SAFARZADEH M, F. Roles of drying, size reduction, and blanching in sustainable extraction of phenolics from olive leaves. **Processes**, Basel, v. 9, n. 9, p. 1662, 2021.

SAVITZKY, A.; GOLAY, M. J. Smoothing and differentiation of data by simplified least squares procedures. **Analytical Chemistry**, Washington, v. 36, n. 8, p. 1627-1639, 1964.

SHI, D. *et al.* Estimation of crude protein and amino acid contents in whole, ground and defatted ground soybeans by different types of near-infrared (NIR) reflectance spectroscopy. **Journal of Food Composition and Analysis**, Orlando, v. 111, p. 104601, 2022.

SCHUSTER, C; HUEN, J; SCHERF, K. A. Prediction of wheat gluten composition via near-infrared spectroscopy. **Current Research in Food Science**, Amsterdam, v. 6, p. 100471, 2023.

SILVA, F. H. C. A. *et al.* Comparison of UV, visible and near-infrared, and mid-infrared spectrometers to estimate maize and sorghum leaf nutrients using dry-intact and ground leaves. **Computers and Electronics in Agriculture**, Amsterdam, v. 211, p. 108001,

2023.

STEVENS, A.; RAMIREZ-LOPEZ, L. **An introduction to the prospectr package**. [S.l.: s.n.], 2020. Disponível em: <https://cran.rproject.org/web/packages/prospectr/index.html>. Acesso em: 21 fev. 2021.

TAHMASBIAN, I. *et al.* Comparison of hyperspectral imaging and near-infrared spectroscopy to determine nitrogen and carbon concentrations in wheat. **Remote Sensing**, New York, v. 13, n. 6, p. 1128, 2021.

TEDESCO, M. J. *et al.* **Análises de solo, plantas e outros materiais**. Porto Alegre: UFRGS, 1995.

TURKER-KAYA, S; HUCK, C. W. A review of mid-infrared and near infrared imaging: principles, concepts and applications in plant tissue analysis. **Molecules**, Basel, v. 22, n. 1, p. 168, 2017.

VAN MAARSCHALKERWEERD, M; HUSTED, S. Recent developments in fast spectroscopy for plant mineral analysis. **Frontiers in Plant Science**, Lausanne, v. 6, p. 169, 2015.

VARMUZA, K; FILZMOSER, P. **Introduction to multivariate statistical analysis in chemometrics**, New York, 2016.

VEUM, K. *et al.* Estimating a soil quality index with VNIR reflectance spectroscopy. **Soil Science Society of America Journal**, Madison, v. 2, p. 637-649, 2015.

WANG, Y. *et al.* Estimating rice chlorophyll content and leaf nitrogen concentration with a digital still color camera under natural light. **Plant Methods**, New York, v. 10, p. 1-11, 2014.

WALESIAK, M.; DUDEK, A. The choice of variable normalization method in cluster analysis. In: SOLIMAN, K. S. (ed.). **Education excellence and innovation management: a 2025 vision to sustain economic development during global challenges**. Sewilla: International Business Information Management Association, p. 325-340, 2020.

WEN, P. *et al.* Estimating leaf nitrogen concentration considering unsynchronized maize growth stages with canopy hyperspectral technique. **Ecological indicators**, New York, v. 107, p. 105590, 2019.

WIJEWARDANE, N. K. *et al.* A leaf-level spectral library to support high-throughput plant phenotyping: predictive accuracy and model transfer. **Journal of Experimental Botany**, Oxford, p. erad129, 2023.

WILLIAMS, P; MANLEY, M; ANTONISZYN, J. **Near infrared technology: getting the best out of light**, London, 2019.

XIE, L.H. *et al.* Optimisation of near-infrared reflectance model in measuring protein and amylose content of rice flour. **Food Chemistry**, London, 142, p.92-100, 2014.

ZANON, A. J. *et al.* **Ecofisiologia da soja**. Santa Maria: Pallotti, p. 135, 2018.

ZORNOZA, R.; GUERRERO, C. Near infrared spectroscopy for determination of various physical, chemical and biochemical properties in Mediterranean soils. **Soil Biology & Biochemistry**, Amsterdam, v. 40, n. 7, p. 1923–1930, 2008.

## 5. CONSIDERAÇÕES FINAIS

Os resultados deste estudo demonstram a eficácia da espectroscopia NIR na estimativa de teor de nitrogênio (N) em folhas de milho, soja, trigo e arroz, e grãos de soja, trigo e arroz. Uma das principais descobertas deste trabalho é a possibilidade de criar um único modelo para prever o teor de N nas folhas dessas quatro culturas, o que sugere uma maior aplicabilidade da técnica. Para os grãos, foi necessário desenvolver modelos separados, sendo um específico para trigo e arroz, e outro para soja. Isso ressalta a importância de levar em consideração a estrutura e composição de cada cultura na calibração dos modelos.

Além disso, este estudo também demonstrou que a espectroscopia NIR pode ser usada com sucesso para estimar o teor de proteína nos grãos, uma vez que a proteína é composta principalmente por N. Essas informações são de grande relevância, uma vez que a determinação do teor de proteína é fundamental na avaliação da qualidade nutricional dos grãos.

Em suma, os resultados deste trabalho abrem novas perspectivas para o uso da espectroscopia NIR na agricultura, facilitando a monitorização do teor de N nas folhas e grãos de várias culturas, bem como a estimativa do teor de proteína nos grãos. Essa abordagem não apenas economiza tempo e recursos, mas também contribui para o avanço da pesquisa agrícola e para o desenvolvimento de práticas mais eficientes na produção de alimentos. Portanto, é evidente que a espectroscopia NIR é uma ferramenta promissora com amplo potencial para aplicações futuras na agricultura.