



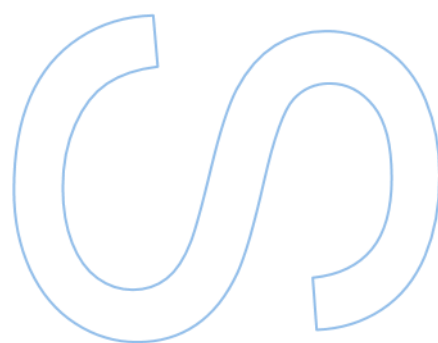
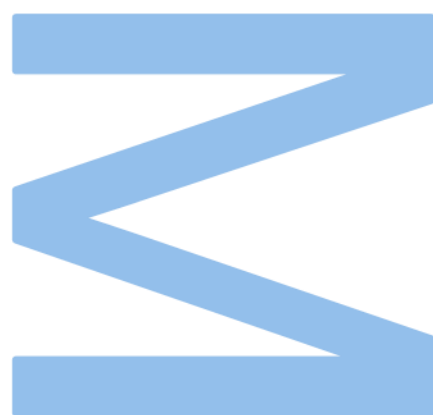
Desenvolvimento de um modelo de calibração MPLS-NIR para um alimento para animal

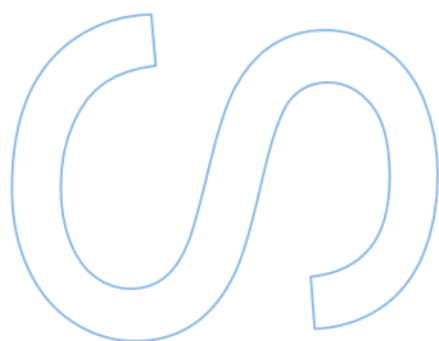
Ana Catarina Alves

Mestrado em Métodos Avançados e Acreditação em Análise
Química
Departamento Química e Bioquímica
2023

Orientador Académico

Carlos Alberto Rocha Gomes, Professor Auxiliar, Faculdade de
Ciências da Universidade do Porto.





Agradecimentos

A realização deste trabalho, assim como a conclusão do presente ciclo de estudos, não teria sido possível sem a colaboração de tantas pessoas que, de uma forma ou de outra, impactaram este percurso que agora chega ao fim. Assim sendo, expresso os meus sinceros agradecimentos a todos:

À Dr.^a Lúcia Medeiros pela oportunidade concebida para desenvolver este trabalho na Associação Interprofissional do Leite e Laticínios, bem como a oportunidade de começar um novo e importante ciclo da minha vida e pelo voto de confiança e apoio.

Ao Prof. Carlos Alberto Gomes, enquanto meu orientador académico, pela disponibilidade e paciência, por ser sempre capaz de apresentar uma solução quando as coisas não fluíam tão facilmente, pela confiança reiterada e por toda a compreensão e mentoria.

À Eng.^a Ana Lage, agradeço não apenas pela disponibilidade e ajuda a um nível mais profissional, mas também pelo apoio pessoal que sempre demonstrou que elevou este trabalho muito para além de um simples estágio curricular.

Aos meus colegas da equipa da ALIP, agradeço pelo apoio e ajuda em todas as etapas desta jornada por facilitarem a minha integração no grupo de trabalho e pela disponibilidade que sempre demonstraram, em especial às minhas colegas Liliana e Marina.

Aos meus colegas de mestrado agradeço pelo apoio, pela partilha e pelo convívio destes dois anos.

À minha família, especialmente aos meus pais, ao meu irmão e ao João, agradeço pelo apoio incondicional, por acreditarem sempre nas minhas capacidades e sacrificarem tanto para que nunca me faltasse nada, pelo apoio e carinho acompanhados sempre de uma palavra amiga. A todos os demais que estiveram presentes e a quem sempre agradeço todo o apoio incondicional. Que tenham tanto orgulho em mim como eu tenho em vocês.

Resumo

A Dieta Totalmente Mistura (TMR) desempenha um papel vital na dieta dos bovinos leiteiros, porém no que diz respeito à análise qualitativa e quantitativa desta mistura utilizando a espectroscopia de infravermelho próximo há uma lacuna significativa na literatura científica.

O propósito subjacente a este estudo consistiu em conceber um modelo de calibração fundamentado na regressão MPLS com o intuito de prever os parâmetros químicos e nutricionais essenciais da mistura, nomeadamente, matéria seca (MS), cinzas totais (CT), proteína bruta (PB), gordura bruta (GB) e fibra bruta (FB). Para alcançar este objetivo, foram utilizadas 255 amostras de mistura TMR em fresco, com as quais se desenvolveram equações de calibração. Estas equações foram elaboradas mediante a aplicação de tratamentos de SNV-DT em conjunto com a utilização de derivadas.

Após a elaboração das equações com as primeiras e segundas derivadas, foram selecionadas aquelas que apresentaram os melhores resultados na previsão dos parâmetros. Os coeficientes de correlação múltipla (RSQ) no conjunto de validação externa foram de 0,995 para a MS, 0,745 para as CT, 0,865 para a PB, 0,846 para a GB e 0,799 para a FB. Os desvios de previsões relativos indicaram que as equações de CT (1,97), PB (2,68) e GB (2,40) foram consideradas modelos eficientes para prever quantitativamente os parâmetros das misturas Unifeed em fresco. No entanto, as equações de MS (12,95) e FB (2,21) necessitam de ajustes, sendo crucial a aplicação de novas estratégias de calibração para melhorar a precisão dos modelos.

Palavras-chave: Espectroscopia de Infravermelho Próximo (NIR), quimiometria, calibração multivariada, regressão de quadrados mínimos parciais modificada (MPLS), validação, mistura Unifeed, TMR.

Abstract

The Total Mixed Diet (TMR) plays a vital role in the diet of dairy cattle, but there is a significant gap in the scientific literature when it comes to the qualitative and quantitative analysis of this mixture using near-infrared spectroscopy.

The purpose of this study was to design a calibration model based on MPLS regression in order to predict the essential chemical and nutritional parameters of the mixture, namely dry matter (DM), total ash (TC), crude protein (CP), crude fat (CrF) and crude fiber (CF). To achieve this, 255 samples of fresh Unifeed mix were used and calibration equations were developed. These equations were developed by applying SNV-DT treatments together with the use of derivatives.

After developing the equations with the first and second derivatives, those with the best results in predicting the parameters were selected. The multiple correlation coefficients (MCR) in the external validation set were 0.995 for MS, 0.745 for CT, 0.865 for PB, 0.846 for GB and 0.799 for FB. The relative prediction deviations indicated that the TC (1.97), PB (2.68) and GB (2.40) equations were considered efficient models for quantitatively predicting the parameters of fresh Unifeed mixtures. However, the MS (12.95) and FB (2.21) equations require adjustments, and it is crucial to apply new calibration strategies to improve the accuracy of the models

Keywords: Near Infrared Spectroscopy (NIR), chemometrics, multivariate calibration, modified partial least squares (MPLS) regression, validation, Unifeed mixture, TMR.

Índice

Lista de Tabelas	vi
Lista de Figuras	vii
Lista de Abreviaturas	ix
1. Objetivos e Enquadramento	1
2. Entidade de Acolhimento.....	2
3. Alimentos Para Animais.....	3
3.1 Dieta Totalmente Misturada (TMR).....	6
3.2 Composição Química dos Alimentos para Animais	9
3.2.1 Matéria Seca	10
3.2.2 Cinza Total	11
3.2.3 Proteína Bruta	12
3.2.4 Gordura Bruta	13
3.2.5 Fibra Bruta	14
4. Princípios da Espectroscopia	16
4.1 Radiação Infravermelho (IR)	20
5. Radiação Infravermelho Próximo (NIR)	23
5.1 Instrumentalização	26
5.1.1 Método de Amostragem.....	28
5.1.2 Métodos de medição	28
5.2 Zonas Características de Absorção.....	31
5.3 Aplicabilidade.....	33
5.4 Vantagens vs Desvantagens uso NIR	34
6. Quimiometria	36
6.1 Pré-tratamentos dos dados.....	37
6.1.1 Transformação dos dados.....	37
6.1.2 Pré-Processamento	45
6.2 Análise Exploratória dos Dados	47

6.3	Calibração Multivariada	51
6.3.1	Regressão por Mínimos Quadrados Parciais (PLS).....	52
6.3.2	Regressão por Mínimos Quadrados Parciais Modificado(MPLS)	54
6.4	Validação do Modelo	55
6.4.1	Validação Cruzada	55
6.4.2	Validação Externa	57
6.5	Outliers	57
7.	Materiais e Métodos	60
7.1	Preparação das amostras	60
7.2	Espectrofotómetro de refletância NIR	62
7.3	Análise química de referência	64
7.4	Desenvolvimento e validação da calibração	65
8.	Resultados e Discussão	68
8.1	Análise química de referência	68
8.2	Espectros NIRS.....	73
8.3	Estatística do desenvolvimento e validação da calibração	77
9.	Conclusão.....	90
	Referências Bibliográficas	92

Lista de Tabelas

Tabela 1: Métodos Oficiais das Frações.....	10
Tabela 2: Algumas zonas de absorção características de compostos presentes em alimentos para animais na região NIR. Adaptado ^{37,39}	32
Tabela 3: Resumo estatístico dos parâmetros analisados nas amostras de mistura Unifeed, expondo o mínimo, o máximo, a média e o desvio padrão de cada componente analisado.	68
Tabela 4: Análise de Componente Principal condições com diferentes tratamentos matemáticos.	78
Tabela 5: Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e primeira derivada (1,4,4,1) para cada parâmetro a analisado nas amostras de TMR em fresco.	83
Tabela 6: Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e primeira derivada (2,4,4,1) para cada parâmetro a analisado nas amostras TMR em fresco.	83
Tabela 7: Estatística de monitorização do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (1,4,4,1) e (2,4,4,1) para cada parâmetro de amostras TMR em fresco.....	85

Lista de Figuras

Figura 1: Representação do modelo Ondulatório.....	16
Figura 2: Espectroscopia de Absorção vs Espectroscopia de Emissão.....	17
Figura 3: Tipos de vibrações moleculares. Alongamento (A) simétrico e (B) assimétrico; Deformação angular (C) de tesoura, (D) de baloiço, (E) de oscilação e (F) de torção. Os sinais + e – denotam movimentos para a frente e para trás do plano da folha, respetivamente ⁵⁷	20
Figura 4: Descrição esquemática das regiões de radiação eletromagnética, mostrando os processos moleculares representativos que ocorrem em cada região cada região, com destaque para a região IR. Adaptado ²⁴	22
Figura 5: Análise de diversos espectros NIR e regiões de comprimento de onda das principais absorções moleculares. Retirado de ³⁰	25
Figura 6: Ilustração de uma configuração típica de um NIR equipado com um interferómetro. Adaptado ⁵⁸	28
Figura 7: Métodos de medição para a aquisição de espectros em espectroscopia NIR. A) Transmitância; B) Refletância; C) transletância; ; A. Amostra; I_R : Intensidade de radiação incidente; I_S : Intensidade de radiação emergente.	29
Figura 8: Principais características espectrais e as posições relativas dos picos para as absorções proeminentes na região do infravermelho próximo. Adaptado ³⁸	32
Figura 9: Representação esquemática da transformação das respostas experimentais em informação química através do modelo matemático.....	52
Figura 10: Esquema ilustrativo da decomposição dos dados pelo método dos mínimos quadrados. Adaptado ⁵⁹	53
Figura 11: Ensaio com diferentes número de rotações testados para 5 segundos.	61
Figura 12: Esquema de Preparação da amostra em Fresco.....	62
Figura 13: Equipamento FOSS NIRS TM DS2500.....	63
Figura 14: Blox Plot de MS.....	69
Figura 15: Blox Plot CT.....	70
Figura 16: Blox Plot PB.....	70
Figura 17: Blox Plot GB.....	71
Figura 18: Blox Plot FB.....	71
Figura 19: Representação gráfica da matriz das correlações existentes entre os parâmetros analisados nas amostras TMR, apresentado histogramas em relação ao parâmetro e a linha de tendência a vermelho para as relações entre dois parâmetros diferente.	72

Figura 20: <i>Scatter plot</i> Constituinte MS vs Contituinte PB	73
Figura 21: Espectro Bruto NIR de todas as amostras.	74
Figura 22: Espectro NIR com SNV+DT	74
Figura 23: Espectro NIR com apenas 1º derivada (1,4,4,1).....	75
Figura 24: Espectro Ruidoso 3º derivada (3,1,1,1).	76
Figura 25: Espectro com 1,4,4,1 SNV+DT	76
Figura 26: Espectro com 2,4,4,1 SNV+DT	77
Figura 27: Histograma de distância H, para 15 fatores.....	79
Figura 28: <i>Scatter Plot</i> 3D, PCA com 15 fatores.	79
Figura 29: Carta de Controlo T de Hotelling, com 15 fatores.	80
Figura 30: Gráfico construído com valores de Q.	81
Figura 31: Gráfico distribuição T ² de Hotelling vs Q	82
Figura 32: Comparação dos valores de CT de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (com outlier 222V incluído).....	87
Figura 33: Comparação dos valores de CT de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (sem outlier 222V).	88
Figura 34: Comparação dos valores de FB de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (com outlier 222V incluído).....	88
Figura 35: Comparação dos valores de FB de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (sem outlier 222V).	89

Lista de Abreviaturas

1-VR	Associação Interprofissional do Leite e Lacticínios
ADF	Acid Detergent Fibre (Fibra solúvel em Detergente Ácido)
ADL	Acid Detergent Lignin (Lenhina insolúvel em Detergente Ácido)
ALIP	Associação Interprofissional do Leite e Lacticínios
AOAC	Association of Official Analytical Chemists (Associação Oficial de Químicos Analíticos)
ANN	Artificial Neural Networks (Redes Neurais Artificiais)
CT	Cinza Total
DT	De-Trending (Eliminação da Tendência)
EMSC	Extended Multiplicative Scatter Correction (Correção de dispersão multiplicativa estendida)
FB	Fibra Bruta
FIR	Far-Infrared (Infravermelho distante)
FTIR	Fourier Transform Infrared Spectroscopy (Espectroscopia de Infravermelho com Transformada de Fourier)
GB	Gordura Bruta
GH	Global H value (GH distance of WinISI related to squared Mahalanobis distance D2) Distância de Mahalanobis
HCA	Hierarchical Cluster Analysis (Análise Hierárquica de Aglomerados)
IUPAC	International Union of Pure and Applied Chemistry (União Internacional de Química Pura e Aplicada)
IR	Infrared (Infravermelho)
LDA	Linear Discriminant Analysis (Análise Linear Discriminante)
LOO	Leave One Out (Deixa-um-fora)
MIR	Middle Infrared (Infravermelho Médio)
MPLS	Modified Partial Least Square (Mínimos Quadrados Parciais Modificados)
MSC	Multiplicative Scatter Correction (Correção de dispersão multiplicativa)
MRL	Multiple Linear Regression (Regressão Linear Múltipla)
NIR	Near-Infrared (Infravermelho próximo)
NIRS	Near-Infrared Reflectance Spectroscopy (Espectroscopia de Infravermelho Próximo)

NIPALS	Nonlinear Iterative Partial Least Squares (Mínimos Quadrados Parciais Iterativos Não Lineares)
NH	Neighbourhood H statistic (Estatística H da vizinhança)
NDF	Neutral Detergent Fibre (Fibra em Detergente Neutro)
OSC	Orthogonal Signal Correction (Correção Ortogonal de Sinal)
PCA	Principal Component Analysis (Análise de Componentes Principais)
PCR	Principal Component Regression (Regressão por Componentes Principais)
PLS	Partial Least Squares (Mínimos Quadrados Parciais)
PB	Proteína Bruta
PC	Principal Component (Componentes Principais)
PRESS	Predictive Residual Sum of Squares (Soma dos quadrados dos erros de previsão)
RER	Relative Error Ratio (Razão de Erro Relativo)
RPD	Ratio of Prediction to Deviation (Desvio de Previsão Relativo)
RMSECV	Root Mean Square Error of Cross Validation (Raiz Quadrada do Erro Médio Quadrático de Validação Cruzada)
RMSEP	Root Mean Squares Error of Prediction (Raiz Quadrada do Erro Médio Quadrático de Previsão)
RSQ	Squared Coefficient of Correlation (Coeficiente de Correlação Quadrado)
SEC	Standard Error of Calibration (Erro Padrão de Calibração)
SECV	Standard Error of Cross-Validation (Erro Padrão de Validação Cruzada)
SEP	Standard Error of Prediction (Erro Padrão de Previsão)
SD	Standard Deviation (Desvio Padrão)
SNV	Standard Normal Variate (Variação de Padrão Normal)
SNV-DT	Combination of SNV and DT (Combinação da SNV com DT)
TMR	Total Mixed Ration/Unifeed ration (Dieta Totalmente Misturada ou Dieta Completa)

1. Objetivos e Enquadramento

Uma Dieta Totalmente Misturada representa a combinação de todos os alimentos essenciais para garantir que uma vaca leiteira produza uma quantidade específica de leite diariamente. Estes alimentos são adicionados em proporções adequadas, assegurando que a vaca consuma todos os nutrientes necessários nas quantidades corretas. Comparada com os métodos tradicionais de alimentação, esta abordagem mostra-se mais eficaz, eficiente e economicamente vantajosa. No entanto, nem sempre os criadores seguem as proporções recomendadas pelos nutricionistas, tornando essencial verificar laboratorialmente os parâmetros nutricionais da mistura criada pelo produtor.

Com o aumento da produtividade das vacas, a rapidez na análise e na entrega dos resultados ao criador e ao nutricionista torna-se crucial para realizar ajustes necessários em tempo útil. Neste contexto, a análise das amostras frescas de TMR revela-se fundamental, pois elimina a necessidade de secagem da amostra, permitindo reduzir o tempo de resposta em aproximadamente 50%. Como resultado, espera-se uma melhoria substancial no serviço prestado, aumentando a satisfação do cliente que busca os serviços laboratoriais.

Desta forma, o objetivo primordial deste estudo é desenvolver um modelo de calibração MPLS capaz de prever parâmetros essenciais, tais como matéria seca, cinza total, proteína bruta, amido, fibra bruta e fibra solúvel em detergente neutro (NDF), em amostras TMR frescas. Este modelo não apenas contribuirá para a eficiência dos processos laboratoriais, mas também fortalecerá a confiabilidade das análises, oferecendo uma ferramenta valiosa para os profissionais envolvidos na criação e nutrição de vacas leiteiras.

2. Entidade de Acolhimento

A 19 de julho de 2002, com o objetivo de representar a fileira láctea nacional surge formalmente, em Lousada, a Associação para o Laboratório Interprofissional do Sector do Leite e Lacticínios (ALIP), a qual após um período de implementação e instalação do Laboratório, iniciou a atividade laboratorial apenas a 2 de outubro de 2006, durante esta época eram apenas efetuadas análises relacionadas com a classificação e valorização do leite ao produtor e contraste leiteiro.¹

No ano 2008 a ALIP foi reconhecida pelo Ministério da Agricultura como Organização interprofissional de Utilidade Pública, o que implicou a alteração da designação da sigla ALIP, passando a ser designada como “Associação Interprofissional do Leite e Lacticínios”. Posteriormente a 3 de junho de 2011, é reconhecida a qualidade dos serviços prestados, tendo sido alcançada a acreditação do laboratório pela IPAC- Instituto Português de Acreditação, a qual tem sido, ano após ano, renovada acompanhando a extensão a novos ensaios e setores. Do mesmo modo, em 2013 foi reconhecida competência técnica para a extensão dos ensaios microbiológicos e para águas de consumo, adicionalmente no final do ano de 2012 o laboratório iniciou a sua atividade no setor de Alimentos para Animais.¹

A ALIP, entidade que se encontra acreditada como Laboratório de Ensaaios, segundo a norma NP EN ISO/IEC 17025:2018, possui atualmente 25 métodos acreditados, dos quais 7 são do setor de Alimentos para Animais e todos os ensaios são realizados nas instalações permanentes do laboratório.² Esta Associação tem como visão e passando a citar: “ser reconhecida, a nível nacional e Internacional, como uma organização de referência para o desenvolvimento de toda a fileira do leite”, à qual acresce a missão “de promover o leite como produto de excelência e contribuir para a melhoria da qualidade do setor leiteiro nacional”³

3. Alimentos Para Animais

Os alimentos são substâncias que, após serem ingeridas pelos animais, são passíveis de serem submetidas a processos de digestão, absorção e utilização no organismo.

Num contexto mais abrangente, o termo "alimento" é empregue para descrever materiais comestíveis. Por exemplo, a erva e o feno são classificados como alimentos, embora nem todos os seus componentes sejam completamente passíveis de digestão pelo organismo⁴

Assim, os alimentos são caracterizados como substâncias comestíveis que englobam nutrientes, os quais representam compostos vitais e imprescindíveis que exercem uma função primordial no apoio às diversas funcionalidades relacionadas com a manutenção, crescimento, desenvolvimento, lactação, reprodução e saúde dos animais.

Os alimentos destinados a animais ruminantes podem ser classificados com base na sua composição e utilização. Diversas categorias são observadas, nomeadamente:

- **Silagens:** Exemplos incluem silagens de milho, erva e luzerna.
- **Forragens Secas:** Esta categoria abrange alimentos como feno, pastagens secas, folhas e caules.
- **Forragens Volumosas:** Engloba feno, palha e cascas com um teor de fibra bruta superior a 18%.
- **Pastagens Verdes:** Compreendem plantas de pastagem e forragens frescas.
- **Alimentos Proteicos:** Referem-se a fontes de proteína, como bagaço de soja.
- **Alimentos Energéticos:** Exemplos incluem milho, trigo e cevada.

Além destas categorias, é possível incorporar suplementos vitamínicos e minerais, bem como aditivos, para assegurar uma dieta equilibrada para os animais ruminantes.⁵

Uma dieta consiste numa combinação de alimentos que, quando fornecida em quantidade adequada e equilibrada a um animal, assegura a satisfação de todas as suas necessidades nutricionais. Já a refeição é tida como a porção de alimento consumida por um animal em momentos regulares, como de manhã e à noite.⁵

A alimentação representa a componente preponderante nos custos associados à operação de qualquer unidade de produção pecuária. Neste contexto, a rentabilidade económica da exploração está intrinsecamente ligada à eficácia com que os animais convertem os nutrientes consumidos em produtos finais. Adicionalmente, para alcançar elevados níveis de produção, torna-se imperativo assegurar o fornecimento de dietas equilibradas. Assim sendo, é crucial adquirir um conhecimento profundo acerca do valor nutricional dos alimentos utilizados.

Embora existam várias tabelas de composição química dos alimentos, a formulação de dietas com base apenas em valores tabelados deve ser encarada como uma abordagem de último recurso. Esta precaução decorre da reconhecida variação na composição química dos alimentos, que pode ser substancialmente afetada por diversos fatores, tais como o ano de produção, a origem e os processos de tratamento aplicados.⁶

As forragens desempenham um papel importante na alimentação de animais ruminantes, podendo ser empregues em estado verde ou sujeitas a processos de conservação. A colheita das forragens ocorre tipicamente antes da maturação plena das culturas, e frequentemente os locais de produção das culturas forrageiras diferem dos locais onde as mesmas serão utilizadas para fins alimentares. Nos contextos de explorações pecuárias de animais ruminantes, nomeadamente as direcionadas para a produção de leite, destaca-se a relevância das silagens de milho e de erva como componentes alimentares essenciais.⁷

A silagem é o resultado de um processo de fermentação láctica natural aplicado a culturas com elevado teor de humidade, com o intuito de conservar durante vários meses. Este processo, denominado ensilagem, visa principalmente a criação de um ambiente anaeróbico, impedindo a exposição das forragens ao oxigénio, colocando a silagem dentro de um silo. Esta abordagem possibilita a preservação eficaz deste alimento, permitindo a sua utilização como fonte de alimentação para animais ao longo de um período alargado. Sem este processo de ensilagem, as forragens tornar-se-iam não comestíveis e poderiam conter elementos potencialmente tóxicos.

Praticamente todas as culturas podem ser conservadas sob a forma de silagem, mas as mais comuns incluem as gramíneas, as leguminosas e os cereais inteiros, com destaque para o trigo e o milho. Para além de um ambiente anaeróbico, é crucial inibir a atividade de microrganismos indesejáveis, como as bactérias do género *Clostridium* ou da família *Enterobacteriaceae*, que podem produzir subprodutos de fermentação indesejáveis e desagradáveis. Estes microrganismos podem ser inibidos de duas maneiras: promovendo o crescimento de bactérias produtoras de ácido láctico ou através da utilização de aditivos químicos. As bactérias produtoras de ácido láctico fermentam os açúcares naturais presentes na forragem ensilada, principalmente glucose e frutose, resultando numa mistura de ácidos, sendo o ácido láctico o predominante. Os ácidos gerados aumentam a concentração de iões de hidrogénio até um nível em que as bactérias indesejáveis são inibidas. O pH crítico em que ocorre essa inibição varia de acordo com o teor de matéria seca da forragem ensilada. Logo

após a colheita da forragem e durante as fases iniciais da ensilagem, ocorrem alterações químicas devido à atividade das enzimas presentes nos tecidos vegetais. Os processos de respiração e proteólise desempenham um papel crucial na influência sobre o valor nutricional do produto final.

A respiração é um processo metabólico que ocorre nas plantas, onde os hidratos de carbono são degradados para produzir energia. Esta degradação envolve a oxidação de açúcares hexoses, que são os principais substratos utilizados, resultando na produção de dióxido de carbono e água, com libertação de energia.

Quando a forragem é colhida, a sua capacidade de realizar reações biossintéticas é limitada, levando a que a maior parte da energia gerada durante a respiração seja dissipada sob a forma de calor. Em condições normais, este calor seria libertado na atmosfera. No entanto, quando as forragens são armazenadas num silo, o calor é retido na massa de matéria vegetal, aumentando a temperatura interna.

Esta produção de calor durante a respiração é um processo ineficiente que pode levar à perda de hidratos de carbono solúveis, essenciais para a fermentação subsequente. Para evitar esta perda, é importante criar condições anaeróbias (sem oxigénio) no silo o mais rapidamente possível, limitando assim a respiração e preservando a qualidade da matéria vegetal armazenada

Após a colheita da erva fresca, a maior parte do azoto total presente encontra-se sob a forma de proteínas, representando cerca de 75-90% desse azoto. No entanto, logo após a colheita, inicia-se um processo de proteólise, que envolve a hidrólise das ligações peptídicas nas proteínas. Em poucos dias de exposição ao ambiente exterior, no campo, o teor de proteínas pode reduzir para cerca de 50%. A extensão deste processo de degradação de proteínas depende da espécie da planta, do teor de matéria seca e da temperatura ambiental.

Quando o material vegetal é ensilado, a proteólise continua, mas a sua atividade diminui à medida que o pH do meio diminui. Durante esse processo, são produzidos aminoácidos e péptidos com diferentes comprimentos de cadeia como produtos da degradação das proteínas. É importante notar que, embora as enzimas vegetais desempenhem um papel neste processo, a maioria da destruição dos aminoácidos da forragem é causada pela atividade microbiana e não pelas enzimas das próprias plantas.⁴

A silagem de milho representa um componente essencial nas dietas de animais ruminantes, especialmente em explorações de produção leiteira, ao longo de grande

parte do ano. Este alimento é facilmente sujeito a um processo de fermentação, apresentando um pH igual ou inferior a 4, o que o torna altamente valioso em termos nutricionais para os animais ruminantes, devido à sua riqueza em fibras digestíveis. No entanto, o seu valor nutricional está sujeito a influências variadas, que incluem o estado de maturação das plantas de milho, a altura do corte, as condições climáticas, a variedade de milho utilizada e a técnica de conservação implementada.⁷

Sendo que, o estado de maturação no momento de corte da cultura é o que mais terá impacto no valor nutritivo da silagem de milho, por essa mesma razão é recomendado aos agricultores a medição de matéria seca para a determinação do momento ideal de corte. É importante que os agricultores que se dedicam à produção de silagem recorram aos seus conselheiros técnicos de modo a identificarem as práticas culturais mais adequadas para as suas propriedades, de forma a encontrar em conjunto soluções e estratégias técnicas e tecnológicas que visem maximizar a produtividade e qualidade das culturas.^{8,9}

A silagem de erva é outra forma de forragem amplamente empregada em explorações leiteiras, desempenhando um papel crucial, especialmente durante o período em que a cultura de milho não está disponível. Este alimento é altamente valorizado devido à sua considerável quantidade de fibras facilmente digeríveis e também por representar uma fonte importante de proteína para os animais. No entanto, a qualidade da silagem de erva é notoriamente instável, sujeita a uma variedade de fatores que podem influenciar substancialmente o seu valor nutricional. Em alguns casos, a qualidade pode ser tão prejudicada que a silagem de erva é excluída das dietas dos animais. Consequentemente, a qualidade geral das silagens de erva é ainda considerada insatisfatória, mas existe uma tendência para melhorias no seu valor nutricional.^{10,11}

3.1 Dieta Totalmente Misturada (TMR)

Uma mistura, define-se como sendo a junção de duas ou mais substâncias, onde as partículas ou elementos constituintes de cada substância estão dispersos e intercalados entre si. No contexto da dieta animal, a mistura tem como objetivo a homogeneidade da dieta, independentemente do estado físico das substâncias.¹²

Uma mistura tida como "ideal" refere-se a um estado em que cada amostra representativa da mesma possui precisamente a mesma composição. No entanto, na realidade, alcançar uma mistura verdadeiramente "ideal" pode ser desafiante devido à segregação que ocorre entre partículas com diferentes tamanhos, densidades e formas. Estas discrepâncias nas características das partículas podem resultar na

separação e na falta de uniformidade na composição da mistura, mesmo quando são feitos esforços para misturar os componentes de forma homogênea.¹²

As TMR são misturas equilibradas de alimentos que fornecem todos os nutrientes necessários aos animais, pelo que são compostas por diversos alimentos, tais como erva, silagens, subprodutos vegetais, cereais, bagaços de oleaginosas, alimentos compostos complementares e suplementos minerais e vitamínicos. A combinação adequada destes alimentos permite atender às necessidades nutricionais dos animais, proporcionando uma dieta equilibrada. O termo dieta totalmente misturado (Total Mixed Ration - TMR) foi introduzido por Owen em 1971, com o propósito de designar uma dieta composta por alimentos grosseiros e concentrados que são misturados de forma homogênea.¹³

O uso da mistura homogênea de diversos alimentos impede que os ruminantes selecionem ingredientes específicos da sua dieta. Para além disso, o uso da dieta total possibilita que a vaca consuma pequenas quantidades de uma refeição equilibrada com maior frequência ao longo do dia, fornecendo um suprimento mais constante de nutrientes para os microrganismos presentes no rúmen, uma maior estabilidade do pH ruminal e uma maior eficiência de utilização da energia e proteína disponível no ecossistema retículo-rúmen.¹⁴

Deste modo, ao ser disponibilizado uma TMR, é assegurada uma distribuição constante de nutrientes aos microrganismos ao longo do dia. Esta prática auxilia na redução dos riscos de distúrbios digestivos, como a acidose provocada um teor de energia fermentável excessivo, contribuindo para a saúde dos animais. Isso facilita a utilização de outros alimentos ricos em fibras, especialmente quando os ingredientes são previamente triturados ou moídos, dificultando a seleção individual pelo animal. Além disso, a adoção de dietas completas permite uma maior utilização de matérias-primas que, quando fornecidas isoladamente, podem apresentar restrições nutricionais devido à sua baixa palatabilidade. Essas características individuais são atenuadas na TMR, enganando a avaliação sensorial do animal.

A elaboração do TMR requer a utilização de equipamentos especializados de pesagem e mistura de alimentos, frequentemente denominados de "reboques misturadores" ou "Unifeed". Por esta razão, na comunidade agrícola europeia, o TMR é frequentemente denominada de "Mistura Unifeed". O mercado de reboques oferece uma ampla variedade de opções, que vão desde reboques que apenas realizam a mistura de rações previamente preparadas até aqueles que permitem a pesagem dos alimentos a serem incorporados na mistura, bem como o seu corte ou trituração e

distribuição automática. Alguns desses reboques necessitam de ser acoplados a um trator, enquanto outros são autónomos. Cada tipo de equipamento apresenta características específicas, vantagens e desvantagens, assim como requisitos particulares em termos de mão de obra e instalações onde são utilizados.

A inclusão de matérias-primas provenientes da exploração agrícola ou de fontes próximas, devido ao seu menor custo, pode resultar numa significativa redução dos custos associados à alimentação animal. Contudo, é importante salientar que nem sempre é viável manter uma composição dietética constante ao longo do tempo devido à sazonalidade na disponibilidade dessas matérias-primas. Por essa razão, é aconselhável procurar a uniformização da dieta, especialmente no que concerne aos alimentos grosseiros, com o propósito de evitar variações abruptas na flora ruminal. Estas variações podem afetar adversamente a digestibilidade e a eficiência na utilização da dieta. Portanto, é fundamental utilizar matérias-primas de elevada qualidade e possuir um conhecimento profundo sobre o seu valor nutricional. Além disso, é crucial operar adequadamente os equipamentos de pesagem, mistura e distribuição, com vista a obter uma formulação precisa e garantir uma mistura homogênea.¹³

A utilização da TMR pode ser adaptada de acordo com o ciclo de lactação das vacas, permitindo uma flexibilidade acrescida na composição da dieta. Isto contribui para que as vacas mantenham um maior consumo de matéria seca (MS) após o parto, reduzindo assim os riscos de distúrbios digestivos no período pós-parto. Além disso, esta abordagem apresenta várias vantagens, como a redução de poeira, defecação e desperdício de alimentos na sala de ordenha, resultando em vacas mais serenas durante a ordenha, o que em alguns casos, pode mesmo aumentar a produção de leite. No entanto, existem desvantagens a considerar, sendo que vacas de baixa produção, com baixo potencial genético, podem ganhar peso excessivo e precisam de ser selecionadas criteriosamente. Os custos associados aos equipamentos necessários, como misturadores, vagões forrageiros e carregadores, podem tornar o sistema inviável para explorações pecuárias de menor dimensão. Além disso, é necessário agrupar com base no nível de produção de leite, fase de lactação, peso corporal e condição corporal para melhorar a eficiência alimentar. O movimento constante das vacas entre os grupos de produção pode resultar em confrontos e lesões, afetando a produção de leite. Por último, importa referir que a dieta total é mais adequada para silagens embora fenos picados ou com partículas mais longas possam ser utilizados em circunstâncias específicas.¹⁴

3.2 Composição Química dos Alimentos para Animais

A nutrição animal constitui uma área de estudo e intervenção científica que se dedica à preparação e formulação de rações e dietas, com vista a suprir as necessidades específicas dos animais nas diferentes fases de crescimento ou estágios de vida. Desempenha um papel fulcral na promoção do bem-estar e saúde dos animais, assegurando o fornecimento adequado dos nutrientes necessários para um desenvolvimento saudável e produtivo. A nutrição animal implica uma compreensão profunda das exigências nutricionais das espécies em causa, bem como a formulação precisa de dietas equilibradas, através da seleção criteriosa de ingredientes e da determinação das proporções ideais de nutrientes. Este processo leva em consideração variáveis como a idade, peso, espécie e nível de atividade dos animais.¹⁵

Para além do seu impacto na saúde e desempenho dos animais, a nutrição animal tem implicações importantes na saúde pública e segurança alimentar, uma vez que uma dieta adequada pode contribuir para a prevenção de doenças e para a melhoria da qualidade dos produtos de origem animal, tais como carne, leite e ovos. De facto, a obtenção de dados acerca dos perfis nutricionais dos alimentos geralmente implica a realização de uma série de análises laboratoriais. Contudo, é importante salientar que, para além da complexidade associada a esses procedimentos, que muitas vezes envolvem o uso de diversos solventes, essas análises podem não refletir com precisão a capacidade dos animais em aproveitar eficazmente os nutrientes dos alimentos. O primeiro método utilizado para determinar a composição nutricional dos alimentos para animais é conhecido como análise proximal ou de Weende. O sistema analítico de Weende, nomeado em homenagem à Estação Experimental de Weende na Alemanha, onde foi desenvolvido, foi concebido por Henneberg e Stohmann em 1865. Este sistema divide a matéria orgânica dos alimentos em várias frações, incluindo proteína bruta (PB), gordura bruta (GB), fibra bruta (FB) e extrativos não azotados. Todas estas frações são determinadas através de métodos analíticos, exceto a fração de extrativos não azotados, que é calculada por diferença. A análise dos alimentos por meio deste sistema não tem como objetivo isolar compostos químicos individuais, mas sim grupos analíticos que inicialmente se considerava serem uniformes em termos de utilização digestiva por todas as espécies de animais. Todavia, vários investigadores têm procurado estabelecer sistemas mais coerentes do ponto de vista químico e nutricional para fracionar a matéria orgânica dos alimentos de origem vegetal, em particular a sua

fração de hidratos de carbono. Na década de 1960, ocorreu uma abordagem revolucionária com o trabalho de investigação liderado por Peter van Soest, que propôs o sistema de soluções detergentes para a análise de alimento. Atualmente este sistema é utilizado na generalidade dos centros de investigação, a nível mundial, sobretudo na análise de forragens e de outros alimentos de natureza grosseira. Com o tempo, o sistema de Weende foi gradualmente substituído pelo sistema de Van Soest, sendo que introduzidas importantes modificações, nomeadamente a substituição da fração de fibra bruta (FB) pela fibra solúvel em detergente neutro (NDF), fibra solúvel em detergente ácido (ADF) e lenhina insolúvel em detergente ácido (ADL). Esta evolução permitiu uma análise mais precisa e abrangente da composição dos alimentos, especialmente no que diz respeito à fração de fibra e aos seus componentes específicos.^{4,5,15,16}

O sistema de análise proximal está organizado de forma a permitir a identificação das frações a serem determinadas, os principais componentes a serem analisados, um resumo dos procedimentos utilizados e os métodos oficiais aprovados pela AOAC (*Association of Official Analytical Chemists*) que são empregues para a análise de alimentos. (tabela 1)

Tabela 1: Métodos Oficiais das Frações.

Frações	Constituição	Método Oficial
Humidade (MS por diferença)	Água e compostos voláteis	Método AOAC 934,01 (gravimétrico)
Cinza Total	Elementos minerais	Método AOAC 942,05 (incineração)
Proteína Bruta	Proteínas, aminácidos, aminos e ácidos nucleicos	Método AOAC 954,01 (digestão)
Gordura Bruta	Lípidos, vitaminas lipossolúveis, pigmentos	Método AOAC 2003,06 (extração, gravimétrico)
Fibra Bruta	Celulose, hemicelulose, lenhina e pectina	Método AOAC 978,10 (incineração, gravimétrico)

3.2.1 Matéria Seca

A perda de massa de uma amostra durante o processo de secagem é determinada em condições específicas, as quais variam de acordo com a natureza da amostra. A avaliação da Matéria Seca (MS) é uma prática comum em análises nutricionais, sendo

particularmente relevante para alimentos de origem vegetal, que frequentemente exibem variações significativas no teor de água. Ter conhecimento da quantidade de água presente é fundamental para permitir comparações apropriadas entre diferentes tipos de alimentos. A determinação da MS pode ser realizada de duas maneiras distintas: pelo cálculo da diferença entre o teor de água (Equação 1)

$$MS(\%) = 100\% - H_2O(\%) \quad (1)$$

ou pela diferença de massa entre o alimento no estado fresco e no estado seco (Equação 2)

$$MS(\%) = \left(\frac{x_{\text{massa em fresco}(g)} - x_{\text{massa em seco}(g)}}{1000g} \right) \times 100\% \quad (2)$$

Desta forma, quanto mais substancial se afigurar a percentagem de MS, menor será o conteúdo de humidade presente nos alimentos. Eventuais erros nas análises de Matéria Seca podem estar associados à evaporação de ácidos gordos voláteis (AGV), óleos essenciais, ácido láctico em silagem ou quaisquer outros produtos resultantes de processos de fermentação.¹⁵

3.2.2 Cinza Total

A cinza total, composta principalmente por sílica, é o resíduo resultante da incineração completa de todos os nutrientes orgânicos presentes nos alimentos, é determinada gravimetricamente a partir da pesagem da decomposição de matéria orgânica por incineração (equação3). No entanto, é importante salientar que esta cinza pode conter minerais originados de compostos orgânicos, como enxofre e fósforo provenientes das proteínas. Durante o processo de incineração em temperaturas elevadas, pode ocorrer a perda de compostos voláteis. Portanto, a análise do teor de cinza não proporciona uma representação precisa dos constituintes inorgânicos dos alimentos, tanto em termos quantitativos como qualitativos.¹⁵

$$\text{Cinzas}(\%) = \frac{m_{\text{cinzas}}}{m_{\text{amostra}}} \times 100\% \quad (3)$$

(sendo que $m_{\text{cinzas}} = m_{\text{cadinho+amostra}} - m_{\text{cadinho vazio}}$)

Do ponto de vista nutricional, é importante destacar que os animais ruminantes não necessitam da cinza em si, mas sim dos elementos minerais individuais que ela contém. Por esse motivo, o parâmetro de determinação de cinza nos alimentos não é amplamente considerado de grande relevância na avaliação da qualidade nutricional dos mesmos.¹⁵

Consequentemente, os valores de cinzas podem subestimar ligeiramente os teores de minerais, no entanto, este erro é mínimo. A identificação de minerais individuais pode ser mais significativa e útil. A principal vantagem desse método é que permite calcular o extrato isento de azoto em comparação com a matéria seca (MS). Além disso, a determinação das cinzas totais indica a percentagem de compostos inorgânicos presentes nos alimentos para animais, sendo que valores elevados de cinzas podem sugerir a possibilidade de contaminação dos alimentos.¹⁵

3.2.3 Proteína Bruta

As proteínas são moléculas orgânicas complexas de elevado peso molecular. À semelhança dos hidratos de carbono e das gorduras, as proteínas contêm carbono, hidrogénio e oxigénio, mas distinguem-se por incluir também azoto e, geralmente, enxofre.^{4,15}

Estas macromoléculas desempenham um papel vital na vida dos animais, constituindo os principais elementos estruturais dos tecidos e estando presentes em todas as células. Para além da sua função estrutural, as proteínas desempenham várias outras funções cruciais para o funcionamento dos animais, tais como atividades bioquímicas, imunológicas, de transporte e reguladoras. Quando necessário, também podem ser utilizadas como fonte de energia. As necessidades proteicas dos animais variam ao longo das diferentes fases da vida, sendo mais elevadas durante os períodos de crescimento, gestação e lactação. Portanto, garantir uma ingestão adequada de proteínas é crucial para a saúde e o desenvolvimento adequado dos animais.⁴

Também é importante destacar que a maior parte do azoto presente nos alimentos está sob a forma de proteínas. Por essa razão, é prática comum e quase universal expressar as necessidades de azoto dos animais e o teor de azoto nos alimentos em termos de proteínas.⁴

O método utilizado para estimar a proteína bruta foi desenvolvido pelo químico dinamarquês Johan Kjeldahl e é amplamente reconhecido como o "método Kjeldahl".

A análise de Kjeldahl baseia-se na determinação do azoto (N) presente na amostra em análise. Para converter o teor de azoto medido na amostra em proteína bruta, é necessário aplicar um fator de cálculo de 6,25 (Equação 4). Esta abordagem fundamenta-se no facto de que todas as proteínas contêm aproximadamente 16% de azoto ($100/16 = 6,25$), ou seja, 16 gramas de azoto por cada 100 gramas de proteína. Portanto, 1 grama de azoto está associado a 6,25 gramas de proteína.^{4,15}

$$\text{Proteína Bruta (PB)} \left(\frac{g}{Kg} \right) = N(\text{azoto}) \left(\frac{g}{Kg} \right) \times 6,25 \quad (4)$$

No entanto, é fundamental notar que a suposição de que todo o azoto nos alimentos está exclusivamente na forma de proteínas é incorreta. De facto, nos alimentos podem estar presentes diversos compostos azotados simples, tais como amidas, aminoácidos, glicosídeos, alcalóides, sais de amónio e lípidos compostos. Portanto, a composição do azoto nos alimentos é mais diversificada do que apenas a sua associação com as proteínas.⁴

3.2.4 Gordura Bruta

Os lípidos, também designados como gorduras, são componentes presentes nos tecidos vegetais, a exemplo dos óleos vegetais, e nos tecidos animais, como na carne, ovos e leite. Sob o ponto de vista físico, os lípidos revelam-se relativamente insolúveis em água e mais solúveis em solventes orgânicos, como o hexano, o éter e o clorofórmio. Quimicamente, os lípidos consistem em compostos orgânicos na forma de ésteres de ácidos gordos e glicerol ou outro álcool similar. As gorduras desempenham um papel de extrema importância como fonte primordial de energia e são absolutamente essenciais para a sobrevivência dos animais. Além dessa função energética, têm uma contribuição fundamental ao fornecer ácidos gordos essenciais, uma vez que representam a única fonte destes compostos para os animais. Além das vantagens nutricionais, a incorporação de gorduras nas dietas dos animais conduz a uma elevação da qualidade dos alimentos, à diminuição da formação de poeira e à redução da segregação de partículas durante o processamento, o que na sua globalidade resulta numa maior atratividade dos alimentos e numa digestão mais eficiente por parte dos animais.^{4,15}

A extração da gordura pode ser realizada utilizando éter de petróleo, com ou sem prévia hidrólise, dependendo da amostra em questão. Sem a etapa de hidrólise, o éter

de petróleo é misturado com a amostra contendo gordura, e depois essa mistura é destilada para separar o éter de petróleo e o extrato seco, que é posteriormente pesado. Esta técnica é comumente aplicada a amostras de origem vegetal, com algumas exceções.¹⁷

Quando é necessária a hidrólise, a amostra é submetida a um processo em que é tratada com ácido clorídrico aquecido, depois arrefecida, filtrada e o resíduo resultante é lavado e seco. Em seguida, a gordura é extraída do resíduo hidrolisado utilizando éter de petróleo, o solvente é destilado e, por fim, o extrato seco é pesado. A hidrólise ácida tende a produzir resultados mais precisos em termos de quantificação da gordura, especialmente em amostras de alimentos sujeitas a tratamento térmico ou de origem animal. Esta técnica de extração de gordura com hidrólise é aplicável a uma ampla gama de amostras, incluindo aquelas de origem animal, bem como amostras vegetais como glúten, leveduras, proteínas de batata e soja, além de produtos alimentares que tenham sido submetidos a processos como extrusão, descamação e aquecimento e alimentos compostos em que pelo menos 20% da gordura presente advenha dos produtos anteriormente mencionados¹⁷

3.2.5 Fibra Bruta

A fibra é definida como a fração predominante das paredes celulares das plantas, sendo composta essencialmente por celulose, hemicelulose, lenhina e pectinas. Estes componentes não são passíveis de digestão nos mamíferos, mas nos ruminantes são submetidos a fermentação pelos microrganismos presentes no rúmen. Estes microrganismos desenvolveram estratégias que permitem uma digestão eficiente destes compostos, tornando-os valiosos fornecedores de energia para os animais.¹⁸

Adicionalmente, é de salientar que a sua textura física desempenha um papel fundamental no correto funcionamento das ações de mastigação, ruminação e secreção salivar. A análise proximal é uma técnica usada para determinar os valores dos macronutrientes nos alimentos. No entanto, a forma como a fibra bruta (FB) é analisada através deste sistema é alvo de críticas, já que não estima de forma precisa a parte insolúvel do alimento. Isso acontece porque a FB é solubilizada por ácido e base, o que resulta numa perda significativa de celulose e na solubilização de hemicelulose, lenhina e pectinas. A extração sequencial ácida e alcalina remove parcialmente a hemicelulose (80%), lenhina (50-90%) e recupera a celulose (50-80%). As pectinas são completamente solubilizadas, resultando na sua remoção total.¹⁵

Na década de 60, Peter Van Soest propôs melhorias nos métodos de análise da fibra, dividindo a fração de fibra em fibra solúvel em detergente neutro (NDF) e fibra solúvel

em detergente ácido (ADF). A NDF contém componentes insolúveis em detergente neutro, como hemicelulose, lenhina e celulose, além de cutina e proteínas ligadas à fibra. A NDF pode ser ainda mais fracionada por meio de extração com detergente ácido, que solubiliza a hemicelulose e as proteínas ligadas à fibra, resultando no ADF. Portanto, a NDF é a soma do ADF e da hemicelulose. A determinação da fração em detergente ácido é particularmente útil para forragens, já que está correlacionada com a digestibilidade do alimento.^{4,15}

Além disso, existe um terceiro fracionamento que dissolve a celulose em ácido sulfúrico, deixando apenas a lenhina como constituinte. Isso resulta na lenhina insolúvel em detergente ácido (ADL). Para animais ruminantes, a lenhina é indigestível, mas as hemiceluloses e celulosas são fermentadas pelos microrganismos no rúmen.¹⁵.

A amostra, possivelmente previamente desengordurada, é digerida com ácido sulfúrico a alta temperatura, depois é lavada e filtrada. Em seguida, é imediatamente digerida com hidróxido de potássio a alta temperatura, lavada novamente e filtrada. O resíduo seco resultante é então incinerado. A perda de peso resultante da incineração do resíduo seco corresponde à massa da fibra bruta.¹⁹

4. Princípios da Espectroscopia

A espectroscopia baseia-se na análise da interação entre uma porção da radiação eletromagnética e a matéria, através da qual é possível retirar informação estrutural. Uma vez que o espectro eletromagnético possui diferentes energias, tal permite que haja inúmeras formas de interação entre a radiação e a matéria, as quais induzem diferentes efeitos nos compostos/moléculas. Desta forma, a espectroscopia divide-se em subgrupos, de acordo com a energia a ser abordada e, consequentemente, com os efeitos que pode produzir ao interagir com a matéria.²⁰

A radiação eletromagnética pode ser descrita como uma forma de energia caracterizada por comportamento dual, quer isto dizer que possui propriedades quer ondulatórias quer propriedades de partícula. A propagação da radiação eletromagnética no espaço é explicada pelo carácter ondulatório da radiação, enquanto a quantificação de energia, que tem inerente a si, é descrita pelo modelo de partículas.

Em 1860 James C. Maxwell referiu-se à energia eletromagnética, como a energia que se desloca no espaço à velocidade da luz, pelo que o comportamento ondulatório consiste em dois campos simultâneos : Campo elétrico e o campo magnético, os quais se propagam num determinado sentido (direção de propagação), todavia com direções perpendiculares (Figura 1).

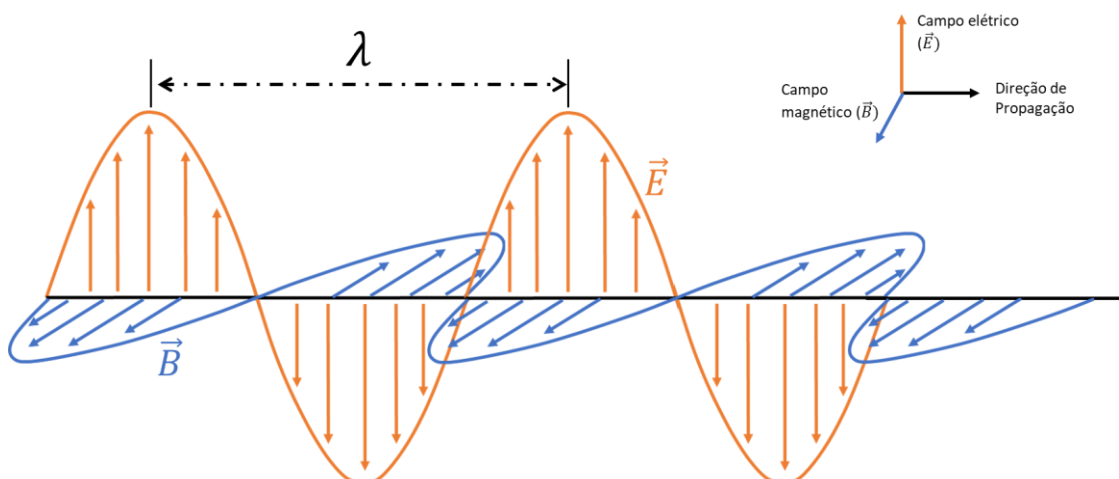


Figura 1: Representação do modelo Ondulatório.

O comportamento de partículas é determinado pelos fótons (feixe quântico), os quais possuem uma quantidade discreta de energia, designada por quantum, a qual é quantificada e associada à radiação em uso, através da constante de Planck (equação 5)

$$E = h\nu = \frac{hc}{\lambda} \quad (5)$$

(em que $E = \text{Energia (Joules)}$; $h = \text{constante de Planck} = 6.626 \times 10^{-34} \text{Js}$; $\nu = \text{frequência} = \frac{c}{\lambda} \text{ Hz}$; $c = \text{Velocidade da luz}$
 $= 2,99792458 \times 10^8 \text{ m/s}$; $\lambda = \text{comprimento de onda} = \frac{c}{\nu} \text{ nm}$);

As medidas espectroscópicas são classificadas de acordo com o tipo de interação entre a energia e a matéria, tendo em conta se ocorre absorção ou emissão de energia e se é aplicada a átomos ou a moléculas. ²⁰

No caso dos espectros de absorção, como por exemplo espectroscopia de absorção atômica, espectroscopia de infravermelho e espectroscopia de ressonância nuclear magnética (RMN), a espectroscopia de absorção ocorre quando ao incidir radiação a amostra absorve energia e passa para um estado energético superior, contrariamente ao que se sucede na espectroscopia de emissão, a qual ocorre quando a partícula que se encontra num estado superior de energia emite uma percentagem passando para um estado energético inferior, exemplo disso é a espectroscopia de fluorescência (figura 2).

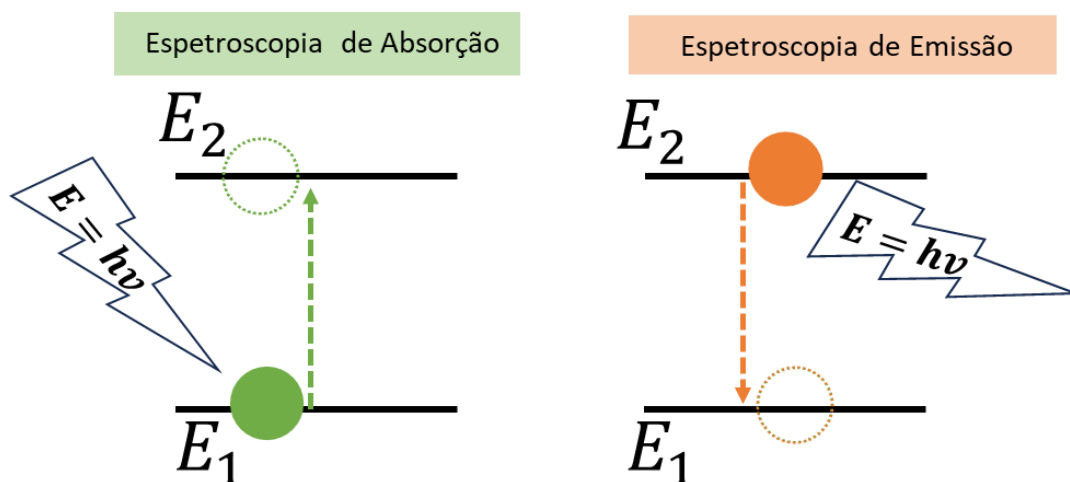


Figura 2: Espectroscopia de Absorção vs Espectroscopia de Emissão.

Quando se faz incidir a radiação eletromagnética numa molécula, apenas parte desta radiação é absorvida, sendo que a absorção só ocorre quando a energia incidente de um fóton é iguala à diferença entre os dois níveis energéticos (equação 6), é apenas absorvida a energia suficiente para passar do nível energético mais baixo (E_1) para o nível energético superior (E_2).²⁰

$$\Delta E = h\nu = E_2 - E_1 \quad (6)$$

Quanto maior for a diferença entre os dois estados energéticos, mais energética tem de ser a radiação incidente de forma que seja possível ocorrer absorção/emissão.

A vibração molecular pode ser explicada por dois modelos distintos, sendo um deles o modelo de oscilador harmónico, o modelo mais simples, em que os níveis vibracionais são equidistantes. Quer isto dizer, que a diferença de energia entre os níveis vibracionais é constante, uma vez que não contempla uma característica particular dos átomos, que é a sua carga elétrica, baseando-se apenas em duas massas sem carga e, portanto, as moléculas seriam capazes de armazenar infinitas quantidades de energia sem resultar na quebra das ligações químicas.²⁰

Para retratar este modelo, é necessário considerar duas massas esféricas m_1 e m_2 , as quais se encontram interligadas por uma mola com uma constante de força, k , e para a qual é descrita de acordo com a lei de Hooke a frequência da vibração (equação 7).

$$\nu = \frac{1}{2\pi i} \sqrt{\frac{k}{u}}, \text{ considerando } u = \frac{m_1 \times m_2}{m_1 + m_2} \quad (7)$$

Este modelo explica o comportamento de uma molécula diatómica baseado na mecânica quântica, que impõe que os átomos não podem vibrar em qualquer amplitude e que somente alguns estados energéticos vibracionais é que são permitidos à molécula, permitindo chegar à equação 8,²⁰

$$E_i = \left(v_i + \frac{1}{2}\right) h\nu; v_i = \{0, 1, 2, \dots\} \quad (8)$$

para a qual, E_i corresponde à energia vibracional de cada nível, v_i corresponde ao número quântico vibracional (0, 1, 2, ...) e h à constante de *Planck*.

Contudo, existe um modelo mais realista para descrever as vibrações moleculares, denominado por modelo do Oscilador Anarmónico, o qual provém de uma alteração do campo elétrico provocada pela incidência da radiação eletromagnética, que altera o momento dipolar²⁰

Desta forma, este modelo não apresenta a mesma diferença de energia entre os níveis vibracionais, mas a diferença de energia diminui com o aumento do número quântico. Sendo que, quanto mais elevado for o nível vibracional menor é a diferença entre os níveis, assim a separação entre os níveis adjacentes fica menor ao longo da molécula até que ocorre uma dissociação completa da mesma.²⁰

Em função do grau de anarmonicidade do modelo vibracional, que é determinado através das diferenças de energia entre os níveis vibracionais adjacentes para o modelo harmónico com as previstas entre dois níveis adjacentes pelo modelo anarmónico, podem ocorrer transições denominadas por transições de overtone (equação 9).

Quando ocorre a transição do estado fundamental de uma molécula para o seu segundo estado excitado, surgem bandas fracas que são designadas como o 1º Overtone. Da mesma forma, quando há a transição do estado fundamental para o terceiro estado excitado, observam-se bandas fracas conhecidas como o 2º Overtone. Por fim, a transição do estado fundamental para o quarto ou quinto estado excitado dá origem a bandas correspondentes ao 3º e 4º Overtone respetivamente.²¹

Esta transição resulta da absorção da radiação eletromagnética, e ocorre quando a energia vibracional de uma molécula promove a transição direta do estado fundamental (energia inicial do estado não excitado) para níveis superiores que não são adjacentes.^{20,21}

$$E_v = hv_e \left(v + \frac{1}{2} \right) - h_{x_e} v_e \left(v + \frac{1}{2} \right)^2 \quad (9)$$

onde h_{x_e} corresponde à constante de anarmonicidade.

Os modos vibracionais estão dependentes de uma forma geral da geometria e constituição da molécula. Os átomos presentes numa molécula nunca permanecem estáticos, encontrando-se em constante movimento. Para moléculas que possuem N átomos e não são lineares, há um total de 3N-6 graus de liberdade associados às vibrações moleculares. Porém, no caso de moléculas lineares, esse número é ligeiramente menor, com 3N-5 graus de liberdade, devido à característica das moléculas com esta geometria, as quais não possuem rotação em torno do eixo que liga os átomos. Estes graus de liberdade são representativos dos diferentes modos normais de vibração que podem ocorrer.²²

Existem dois tipos de modo vibracional molecular: *Stretching* e *Bending*, sendo que o *Stretching* diz respeito a alteração na distância da ligação, e pode ser subdividido em simétrico, os átomos movimentam-se na mesma direção e simultaneamente, ou assimétrico, enquanto um átomo se afasta o outro aproxima-se e vice-versa. *Bending* é relativo à alteração do ângulo de ligação química, subdividindo-se em tesoura (*scissors*), que ocorre de forma simétrica e no plano, de baloiço (*rocking*), que ocorre de forma assimétrica e no plano, de oscilação (*wagging*), em que a vibração é assimétrica mas fora do plano, e de torção (*twisting*), onde a vibração é igualmente fora do plano, mas simétrica, como representado na figura 3.

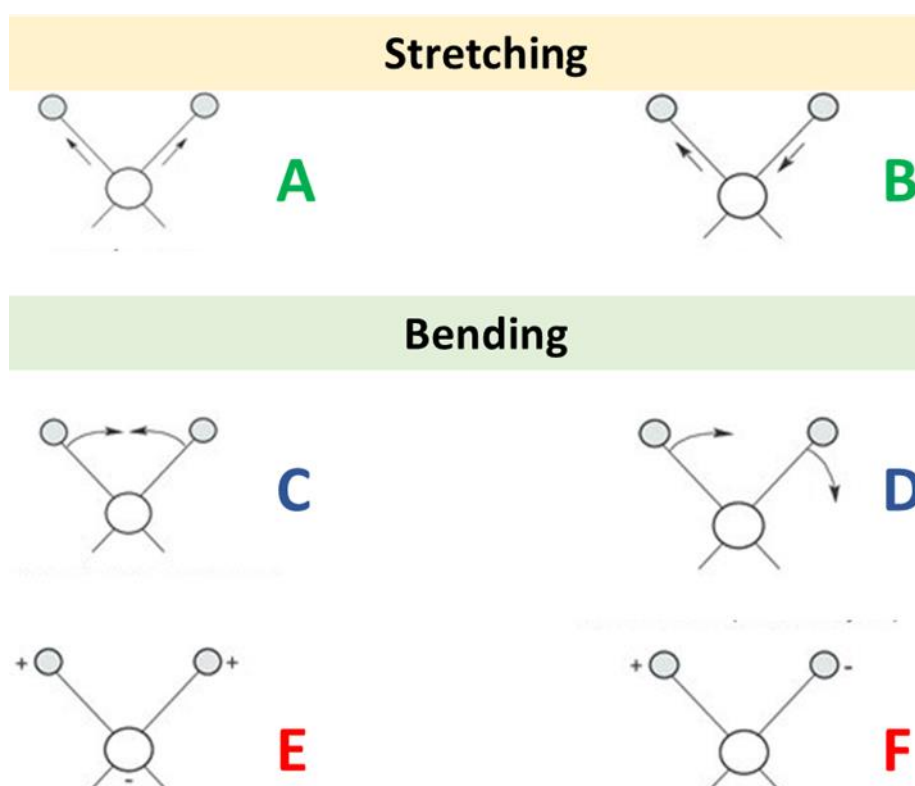


Figura 3: Tipos de vibrações moleculares. Alongamento (A) simétrico e (B) assimétrico; Deformação angular (C) de tesoura, (D) de baloiço, (E) de oscilação e (F) de torção. Os sinais + e - denotam movimentos para a frente e para trás do plano da folha, respetivamente ⁵⁷

4.1 Radiação Infravermelho (IR)

A espectroscopia no infravermelho remonta ao fim do século XVIII, aquando da descoberta de William Herschel da primeira região não visível do espectro de absorção, esta descoberta sucedeu no seguimento de um estudo acerca da diferença entre a temperatura e as diferentes cores obtidas da luz solar dividida por um prisma, em que Herschel descobriu um aumento significativo da temperatura mesmo fora da zona

vermelha da luz solar, pensou tratar-se de um tipo de radiação invisível diferente da luz visível para além da extremidade vermelha e por isso denominou como “heat ray”.

23

Em 1835, confirmou-se que o “heat ray” tinha um comprimento de onda superior ao da luz visível, e passou a ser designado como luz Infravermelha (IV).²³

Porém a sua aplicação foi deixada para segundo plano pois havia outros métodos analíticos que permitiam obter resultados mais inequívocos, principalmente relativamente à estrutura molecular.²³

Enquanto a transição de estado eletrónico ocorre na zona do visível, na zona do IR ocorre a transição de estado vibracional, uma vez que energia dos fótons da radiação Infravermelho corresponde à diferença entre a energia dos níveis vibracionais da molécula, os quais são dependentes do tipo e força da ligação. De modo que ocorra absorção de energia, para além da condição anteriormente referida, também tem de acontecer alteração no momento dipolar para que o sinal seja detetado. É relevante salientar que nem todos os modos vibracionais apresentarão estados excitados na espectroscopia de absorção no infravermelho. O princípio fundamental da espectroscopia de infravermelho baseia-se na deteção da radiação infravermelha que é absorvida por uma molécula, convertendo-se num aumento da energia vibracional, pelo que este aumento de energia tem o potencial de provocar uma mudança no momento dipolar da molécula, resultando em bandas de absorção nas regiões do espectro correspondentes ao comprimento de onda da energia absorvida. A intensidade dessas bandas está relacionada com a magnitude da mudança que ocorreu no momento dipolar. Portanto, apenas os modos de vibração que demonstram uma alteração no momento dipolar diferente de zero serão observáveis na espectroscopia de absorção no infravermelho. Assim, de uma forma geral, as moléculas diatómicas homonucleares, tais como H_2 ; N_2 e O_2 , não possuem momento dipolar e por essa mesma razão são inativas na absorção do infravermelho.²⁰ Por outro lado, as moléculas diatómicas heteronucleares, como por exemplo HCl , NO ; CO , por possuírem momento dipolar são ativas no IR.

As bandas do espectro IR correspondem a vibrações de fragmentos específicos das moléculas, são úteis para identificação de compostos e permitem uma análise qualitativa e quantitativa.²⁴

A Região do Infravermelho tem uma grande amplitude energética, variando de 150 KJmol^{-1} a $0,12 \text{ KJmol}^{-1}$, vai desde os $12\,500 \text{ cm}^{-1}$ até aos 10 cm^{-1} (figura 4), sendo que

apresenta 3 sub-regiões: Infravermelho próximo, médio (mid-IR; 4000-400 cm^{-1} ; 2500-25.000 nm) e distante (FIR; 400-10 cm^{-1} ; 0,025-1 mm), sendo a zona de Infravermelho Próximo (NIR) a com maior número de onda (12 500 – 4000 cm^{-1}) e consequentemente menor comprimento de onda (800-2500 nm), o que a faz a mais energética (150-50 KJ/mol) das três sub-regiões, sendo que quanto menor o λ maior a Energia da Radiação.²³⁻²⁶

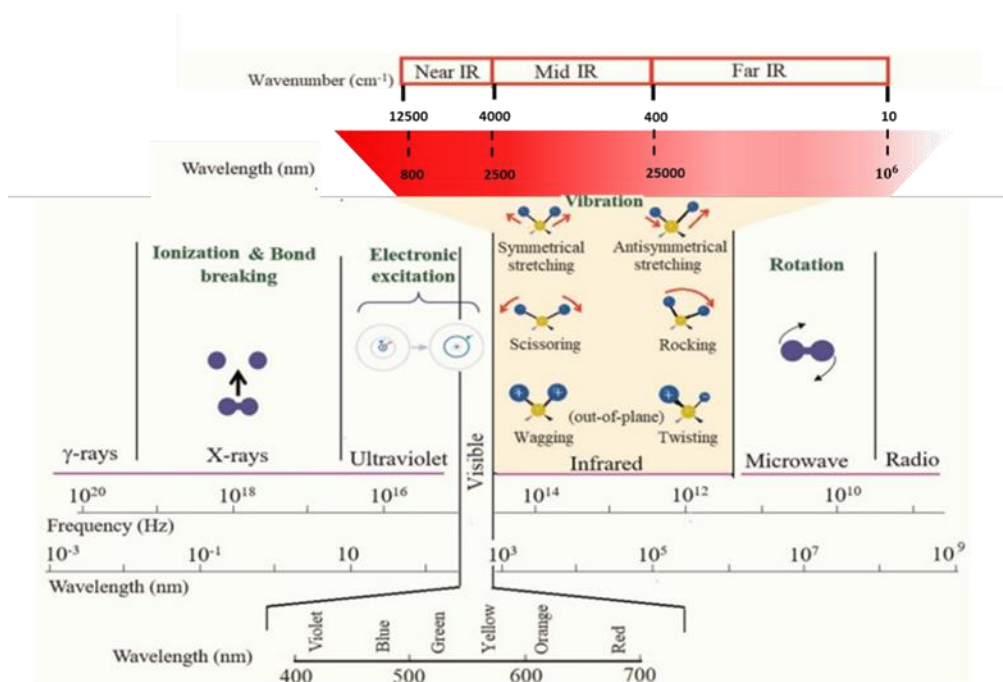


Figura 4: Descrição esquemática das regiões de radiação eletromagnética, mostrando os processos moleculares representativos que ocorrem em cada região cada região, com destaque para a região IR. Adaptado²⁴

Em geral, um espectro de IR pode ser dividido em quatro regiões para mais fácil interpretação, nomeadamente, **1)** 2500-4000 cm^{-1} : absorção envolvendo ligações simples formada por H e outros elementos, como por exemplo C-H, O-H e N-H; **2)** 2000-2500 cm^{-1} : absorção envolvendo ligações triplas, como por exemplo $\text{C}\equiv\text{C}$ e $\text{C}\equiv\text{N}$; **3)** 1500-2000 cm^{-1} : absorção envolve ligações duplas, tais como $\text{C}=\text{C}$ e $\text{C}=\text{N}$; e **4)** 400-1500 cm^{-1} , que corresponde à região que consiste numa maior diversidade de bandas, tornando-se mais complexa a sua interpretação, é única para cada composto, sendo muitas vezes designada por região de impressão digital, raramente é utilizada para identificação de algum grupo funcional em particular.²⁴

5. Radiação Infravermelho Próximo (NIR)

Um princípio fundamental da tecnologia NIR (Espectroscopia no Infravermelho Próximo) reside na absorção de energia por uma amostra quando esta é irradiada. Quando a radiação infravermelho atravessa a amostra, ocorre absorção somente se a frequência coincidir com a frequência de vibração da molécula, o que resulta na absorção seletiva de determinadas frequências em comprimentos de onda específicos. O espectro de absorção da amostra é construído a partir da relação entre a intensidade de absorção e o comprimento de onda da radiação incidente. No entanto, para que a vibração molecular ocorra, é necessário que a radiação infravermelha possa interagir com a molécula. Isso só acontece quando o deslocamento dos átomos num modo vibratório é capaz de causar uma mudança no momento dipolar da molécula. Portanto, a intensidade de um determinado espectro de absorção está diretamente relacionada com a magnitude da mudança no momento dipolar.^{21,27}

A energia que não foi absorvida é conseqüentemente, refletida, fornecendo informações acerca da composição da amostra. Para obter informações precisas sobre a concentração do analito de interesse, é necessário primeiro estabelecer uma relação entre as absorções, ou dados espectrais, e a composição química conhecida das amostras, obtida por meio de métodos laboratoriais, designados por métodos de referência. Essa relação é alcançada utilizando um método de regressão linear múltipla.⁴

A espectroscopia no infravermelho próximo (NIR) remete à região do espectro compreendida entre os 800 e 2500 nm, pelo que diz respeito tanto às transições eletrônicas quer as transições vibracionais.^{23,26,28}

A região NIR pode ser dividida em três outras regiões, designadas por Região I (800–1200 nm; 12,500–8500 cm⁻¹), Região II (1200–1800 nm; 8500–5500 cm⁻¹), e Região III (1800–2500 nm; 5500–4000 cm⁻¹), contudo os seus limites não são rigorosos.²³

O aparecimento da espectroscopia NIR carece de uma explicação inequívoca. No fim do século XIX, quer a radiação infravermelha quer a radiação visível já eram usadas para a medição de compostos simples e possivelmente a região NIR também devia ser usada para medição, porém não existe confirmação.²⁸

Todavia, julga-se que a falta de seletividade dos espectros de absorção na região NIR, constituiu a principal razão para a demora do desenvolvimento das primeiras

aplicações da espectroscopia de infravermelho próximo. Uma vez que, como os espectros eram constituídos por inúmeras bandas de absorção fortemente sobrepostas, a sua interpretação era bastante difícil, com a falta de interesse o desenvolvimento de equipamentos específicos para esta região também foi adiado.^{20,28}

Existiram alguns trabalhos que fizeram uso da espectroscopia NIR, como por exemplo na determinação de água em gelatina, mas foi através do trabalho de Karl Norris, engenheiro agrónomo, e os seus colaboradores, que se iniciou a disseminação de informação desta técnica no ambiente académico, sendo desta forma reconhecido o potencial da técnica de espectroscopia NIR.^{20,29}

O responsável pela inserção da espectroscopia NIR na indústria é Phil Williams, o qual substituiu o método de Kjeldahl, até então usado para determinação de proteína, pela determinação através da espectroscopia, sendo que atualmente tornou-se o método de controlo de qualidade eleito na indústria alimentar, pelas inúmeras vantagens que apresenta em relação a outras técnicas analíticas.²⁰

Contrariamente ao que se sucede nas regiões vizinhas do IR, a região do visível e ultravioleta, as quais são mais energéticas e promovem o aumento de energia eletrónica, a radiação NIR como transporta energia relativamente baixa, raramente possibilita que a sua interação com a matéria promova o aumento da energia eletrónica dos átomos ou moléculas.²⁰ A espectroscopia NIR é classificada como uma técnica de espectroscopia vibracional, dado que a energia transportada pela radiação NIR normalmente não é suficiente para induzir alterações significativas nos níveis eletrónicos de moléculas ou átomos. Contudo, essa energia é adequada para excitar os níveis de rotação, uma vez que ultrapassa as restrições impostas pelas leis da mecânica quântica.²¹

No espectro do infravermelho próximo (NIR), é possível identificar frequências relacionadas às vibrações normais, bem como a *overtones* e combinações dos modos vibracionais.²²

Na espectroscopia NIR, os espectros são frequentemente complexos devido à natureza menos intensa das absorções e à presença de *overtones*, os quais são anarmónicos e como tal não se comportam de forma simples, e combinações das ligações CH, OH e NH em vários grupos funcionais (Figura 5). Como o *overtone* ocorre entre o estado fundamental e o segundo ou terceiro estado de excitação de vibração e tendo em conta que estas transições de energia mais elevada são sucessivamente menos

propensas de acontecerem, consequentemente cada *overtone* sucessivamente possui menor intensidade.^{20,23,29}

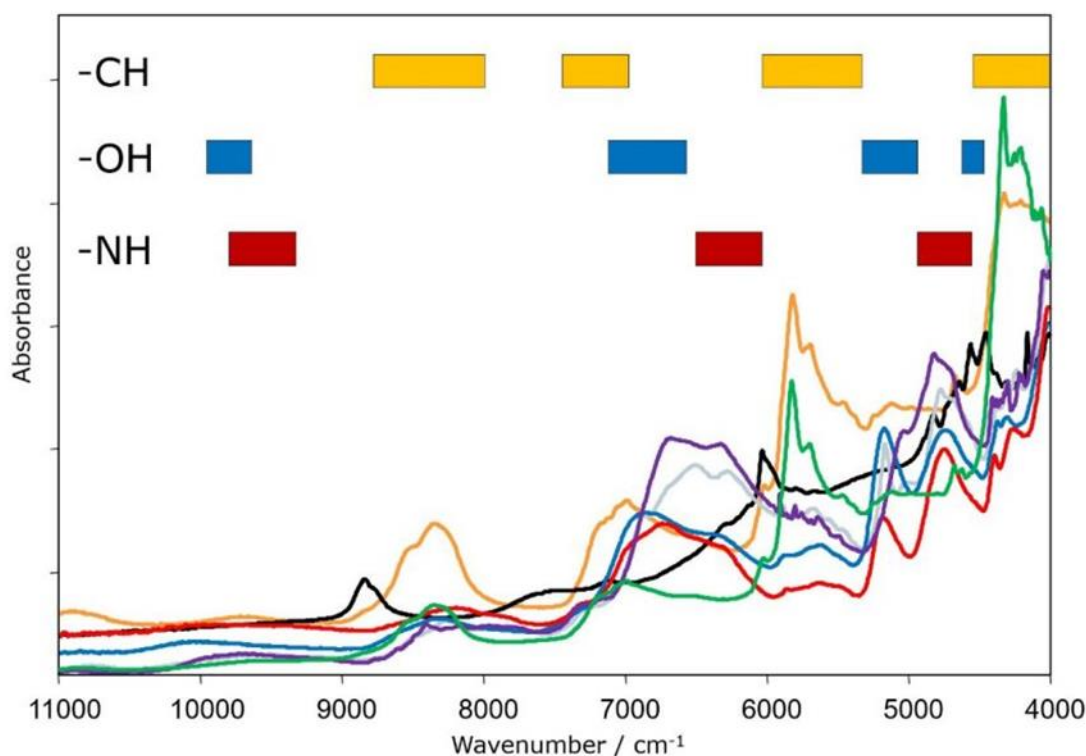


Figura 5: Análise de diversos espectros NIR e regiões de comprimento de onda das principais absorções moleculares. Retirado de ³⁰

Por conseguinte, torna-se necessário utilizar métodos estatísticos que envolvam uma calibração multivariada. Esta abordagem utiliza vários comprimentos de onda na determinação de uma propriedade.³¹

A integração dos princípios da Quimiometria, conforme abordado na secção Quimiometria, desempenhou um papel significativo no aprimoramento da atratividade da espectroscopia no infravermelho próximo (NIR). Isso possibilitou a capacidade de discriminação qualitativa e a determinação quantitativa de várias propriedades das amostras. Este avanço foi alcançado por meio do processamento avançado dos dados disponíveis, empregando ferramentas de análise de dados multivariados (para múltiplos comprimentos de onda) e técnicas de calibração, como a regressão por mínimos quadrados parciais (PLS).³²

Uma vez que as ligações OH, CH e NH têm uma influência significativa nos espectros NIR, esta característica facilita a determinação de parâmetros como humidade, teor de gordura, proteínas e até mesmo hidratos de carbono.³¹

A técnica NIR tem sido predominantemente utilizada para a determinação de compostos orgânicos. Todavia, vários estudos indicam a viabilidade da determinação de compostos inorgânicos. Ao contrário das ligações orgânicas, que apresentam bandas de absorção na região do NIR, os compostos inorgânicos só podem ser detetados quando estão presentes em complexos orgânicos e quelatos. Qualquer composto inorgânico que forme ligações covalentes dentro de um fragmento de um ião molecular, seja catião ou anião originará um espectro de absorção característico com grupos de frequência associados. A estrutura e a orientação do ião ou do complexo podem influenciar os espectros de absorção dos compostos inorgânicos.³³

5.1 Instrumentalização

A obtenção do espectro NIR requer um espectrofotómetro que inclui uma fonte luminosa, um sistema ótico, um suporte para as amostras e um detetor de sinal. Estes componentes são fundamentais para a análise de amostras no espectro próximo ao infravermelho, possibilitando a obtenção de informações valiosas sobre a composição das substâncias.^{20,23,34}

Uma das fontes luminosas de eleição na espectroscopia no infravermelho próximo (NIR) consiste na lâmpada de halogéneo com filamento de tungsténio. Esta lâmpada possibilita a emissão de radiação numa vasta gama de comprimentos de onda, abrangendo um intervalo que se estende desde os 320 nm até aos 2500 nm. Adicionalmente, é viável a utilização de diodos emissores de luz (LEDs), que, em virtude do seu baixo consumo energético, custo acessível e dimensões reduzidas, constituem uma alternativa atraente. Importa, no entanto, ressaltar que os LEDs possuem uma faixa de emissão espectral mais limitada, tornando-os particularmente adequados para aplicação em dispositivos de dimensões reduzidas.^{34,35}

O sistema ótico fragmenta a luz policromática em diferentes comprimentos de onda, procurando alcançar uma largura de banda estreita e uma intensidade elevada na região de absorção. A seleção dos comprimentos de onda tem como objetivo obter valores discretos e pode envolver a utilização de filtros de interferência, filtros ótico-acústicos sintonizáveis (AOTF) e redes de difração. Além disso, alguns dispositivos utilizam interferómetros para garantir alta precisão, exatidão e uma razão sinal-ruído significativa durante a aquisição rápida de dados.^{34,35}

O suporte para amostras depende do tipo de amostra, que pode ser sólida, líquida ou gasosa. A ausência de necessidade de pré-tratamento das amostras para a obtenção

dos espectros possibilita a utilização de uma ampla variedade de suportes adaptáveis a cada situação. O detetor é responsável pelo processo de converter a energia da radiação que nele incide num sinal elétrico, que é, em seguida, traduzido num sinal digital. O silício é utilizado para comprimentos de onda mais curtos, enquanto o sulfureto de chumbo (PbS), índio, gálio e arsénio (InGaAs) são preferenciais para comprimentos de onda mais longos. O detetor de PbS possui um tempo médio de vida útil, é económico e sensível, embora apresente limitações em termos de linearidade e velocidade de resposta, que podem ser atenuadas com a utilização dos detetores InGaAs. Detetores de semicondutor são ideais para a faixa visível até 1100 nm, sendo rápidos e sensíveis. Os detetores de PbS, apesar de mais lentos, são populares devido à durabilidade e sensibilidade na gama de 1100 nm a 2500 nm. Por sua vez, os detetores de InGaAs, embora mais dispendiosos, unem as características de velocidade e dimensões do detetor de silício com a amplitude de cobertura de comprimentos de onda do detetor de PbS.³⁴

Existem dois tipos distintos de espectrofotómetros NIR amplamente utilizados: os dispersivos e os de transformada de Fourier (FT-NIR). Os equipamentos originais de infravermelhos eram do tipo dispersivo, recorrendo a um prisma ou a um monocromador de rede para decompor a luz visível em diferentes frequências, sendo que o monocromador de rede oferecia uma separação mais precisa das frequências da radiação infravermelha, resultando numa maior resolução. Contudo, os sistemas dispersivos têm a limitação de medir apenas uma pequena parte do espectro de cada vez. Para obter o espectro completo, é necessário varrer a rede, os espelhos ou o detetor, ou mesmo uma combinação dos três. Nos espectrofotómetros, é comum empregar o interferómetro de Michelson, um componente essencial composto por um divisor de feixes (*beamsplitter*), um laser (habitualmente HeNe) e dois espelhos (um fixo e um móvel), como ilustrado na figura 6. Primeiramente, um colimador torna a radiação proveniente da fonte de radiação em feixes de luz paralelos e direciona-os para o interferómetro. No interferómetro, o divisor de feixes reflete uma porção dos feixes em direção ao espelho fixo e permite que os feixes restantes passem em direção ao espelho móvel. Estes feixes refletidos retornam ao *beamsplitter*, onde são reagrupados e encaminhados para o detetor. O laser controla o movimento preciso do espelho móvel, garantindo um alinhamento preciso do interferómetro. A interferência dos feixes é determinada pela diferença no percurso ótico percorrido. Esta variação na intensidade da radiação, registada ao longo do tempo, resulta num interferograma. Através da aplicação da transformada de Fourier, o interferograma é convertido num espectro de absorção no infravermelho. Esta abordagem permite realizar varreduras

rápidas numa vasta gama de comprimentos de onda, eliminando a necessidade de componentes como fendas ou filtros, que podem reduzir a sensibilidade do equipamento.²¹

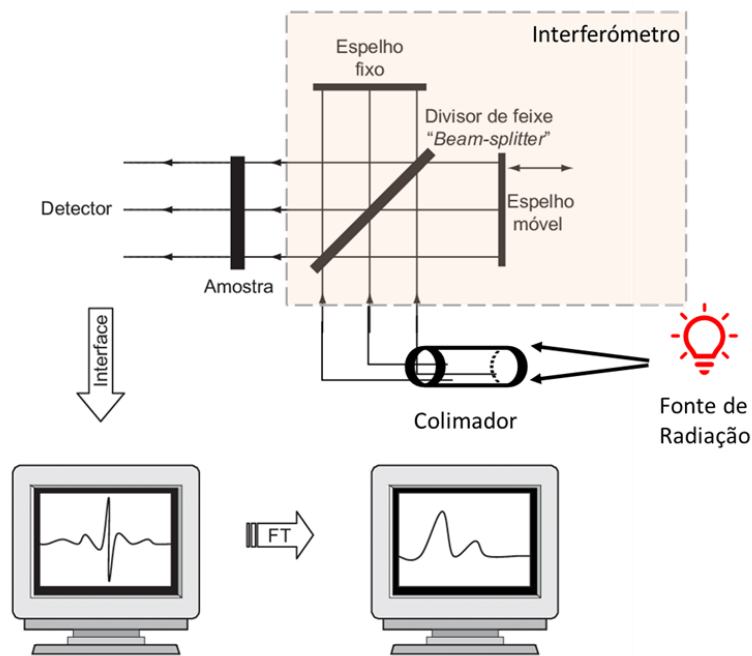


Figura 6: Ilustração de uma configuração típica de um NIR equipado com um interferómetro. Adaptado ⁵⁸

5.1.1 Método de Amostragem

Quando se deseja adquirir os espectros NIR de uma amostra, é imperativo recolher uma porção da área onde o processo está a decorrer. Para este propósito, existem diversos métodos de amostragem. O método "Off-line", o qual implica amostragem manual e o seu transporte para o laboratório. No método "At-line", a amostra é adquirida manualmente e analisada nas imediações do local do processo. No método "In-situ", uma sonda é inserida diretamente no recipiente em contacto direto com o meio para efetuar a medição. No método "On-line", a amostra é desviada do processo de fabrico, sendo automaticamente analisada antes de ser reintegrada no mesmo. Por fim, no método "In-line", a medição é realizada automaticamente sem retirar a amostra da corrente do processo. Cada um destes métodos oferece diferentes abordagens à amostragem de espectros NIR, adaptando-se às necessidades específicas de cada situação.^{34,36}

5.1.2 Métodos de medição

A escolha do método de aquisição mais adequado para uma determinada amostra depende das suas propriedades óticas e físicas. Em geral, materiais transparentes são

frequentemente analisados utilizando a técnica de transmitância. Por outro lado, amostras líquidas que sejam turvas, muito espessas ou sólidas podem ser medidas através das técnicas de refletância difusa ou transfletância.^{34,36}

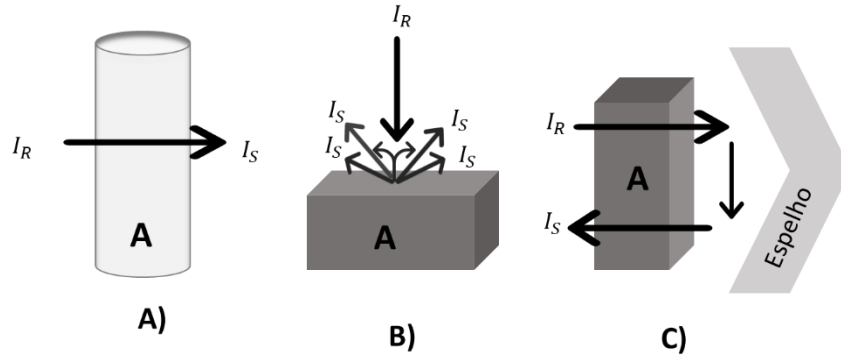


Figura 7: Métodos de medição para a aquisição de espectros em espectroscopia NIR. A) Transmitância; B) Refletância; C) transfletância; ; A. Amostra; I_R : Intensidade de radiação incidente; I_S : Intensidade de radiação emergente.

A transmitância é a denominação atribuída à medida de um feixe de luz capaz de atravessar uma amostra específica sem ser absorvida ou refletida. Este método encontra aplicação sobretudo na análise de amostras líquidas transparentes que não induzem dispersão da luz. O resultado pode ser apresentado diretamente em termos de transmitância (T) e/ou de absorvância (A).

$$A = \log_{10} \frac{I_0}{I} = \log_{10} \frac{1}{T} = -\log_{10} T \quad (10)$$

Considerando I_0 a intensidade da radiação incidente, I a intensidade da radiação transmitida, e pela lei de Lambert-Beer é possível determinar a equação 11 através equação 10.

$$T = \frac{I}{I_0} = 10^{-A} \quad (11)$$

A aplicação da lei de Lambert-Beer é justificada pela presença de uma relação proporcional entre a transmitância, a concentração do analito e a extensão do percurso ótico.

A espectroscopia de refletância compreende a medição da luz refletida a partir da superfície de uma amostra, apresentando duas vertentes distintas: a refletância especular e a refletância difusa. Contudo, a refletância especular, oferece informações

limitadas ou mesmo inexistentes sobre a composição da amostra, devido à sua interação superficial mínima com a mesma.³⁶

Este método de refletância difusa viabiliza a medição da reflectância (R), que corresponde à relação entre a intensidade da luz refletida pela amostra (I) e a refletida por uma superfície de fundo ou de referência (I_R). No espectro NIR, a radiação é capaz de penetrar consideravelmente no interior da amostra, onde pode ser absorvida devido a combinações de vibrações e ressonâncias parciais das substâncias em análise presentes na amostra. A radiação não absorvida é então refletida pela amostra em direção ao detetor.³⁶

Os espectros, utilizando este método, são frequentemente obtidos através de cálculos e transposições gráficas do logaritmo de ($1/R$) em relação ao comprimento de onda ou ao número de onda.

A refletância pode ser determinada da seguinte forma (equação 12):

$$R = \frac{I}{I_R} \quad (12)$$

Este método é geralmente aplicado a materiais sólidos. É fundamental que as condições de medição sejam altamente reprodutíveis, de amostra para amostra. Somente os espectros medidos em relação a um fundo que apresente as mesmas propriedades óticas podem ser comparados de forma significativa entre si.

A transfletância é um método especialmente utilizado para determinar a transmitância, envolvendo a combinação dos conceitos de transmitância e refletância. Este método é aplicado na análise de líquidos, independentemente da sua transparência, recorrendo a uma célula de quartzo e a uma placa refletora difusa. A radiação incidente atravessa a amostra, é refletida pela placa e, posteriormente, atravessa novamente a amostra. A diferença entre a transfletância e a transmitância simples reside na duplicação do percurso ótico, uma vez que o feixe de radiação passa duas vezes pela amostra. Esta configuração pode ser aplicada a um dispositivo que integra os sistemas de refletância e fibra ótica, nos quais a fonte luminosa e o detetor se encontram posicionados no mesmo lado da amostra. A amostra é examinada numa célula equipada com um espelho ou refletor de dispersão apropriado, que pode ser feito de metal ou de uma substância inerte que não absorva na região do infravermelho próximo.²¹

É determinado tendo em consideração que, I_T corresponde à Intensidade da radiação transfectida, sem amostra e I à Intensidade da radiação transmitida, medida na amostra

e como $A^* = \log_{10} (1/T^*)$, obtém-se que (equação 13):

$$T^* = \frac{I}{I_T} \quad (13)$$

Para efetuar medições precisas dos espectros NIR, é imperativo que a preparação das amostras seja adequada, especialmente quando se trata de amostras sólidas. As variações na densidade de empacotamento das amostras podem desencadear efeitos de dispersão da luz, os quais representam fontes substanciais de erro nos espectros.³⁶

5.2 Zonas Características de Absorção

Ao realizar a análise espectral de uma amostra desconhecida por meio da espectroscopia NIR, surgem várias questões pertinentes, como a identificação do tipo de esqueleto de carbono que constitui a amostra e a determinação dos compostos presentes. Esta técnica é notavelmente eficaz na previsão da existência de grupos funcionais e na identificação de compostos, uma vez que a maioria dos grupos característicos apresenta bandas de absorção distintas na região do infravermelho.³⁷

Na região NIR, as bandas de absorção mais proeminentes estão relacionadas com *overtones* e combinação de bandas das transições fundamentais das ligações C–H, O–H e N–H, como ilustrado na figura 8. No entanto, é importante salientar que, quando comparadas às suas contrapartes na região do infravermelho médio, estas bandas na região NIR tendem a ser menos intensas e apresentam maior largura.³⁸

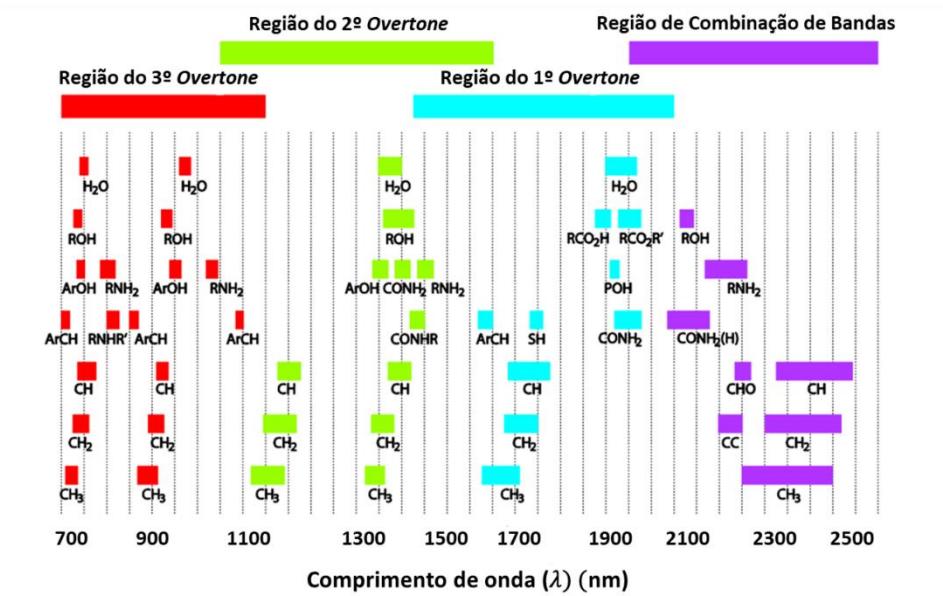


Figura 8: Principais características espectrais e as posições relativas dos picos para as absorções proeminentes na região do infravermelho próximo. Adaptado ³⁸

Assim, na Tabela 2, são fornecidos exemplos de regiões típicas de absorção de grupos funcionais encontrados em alimentos para animais, englobando humidade, proteína, gordura e fibra.^{37,39}

Tabela 2: Algumas zonas de absorção características de compostos presentes em alimentos para animais na região NIR. Adaptado ^{37,39}

Localização Característica das Bandas de Absorção			
λ (nm)			
Humidade	Proteína	Gordura	Fibra
950 – 980	975-1015	896-915	985 - 995
1440-1470	1480-1550	911-935	1170
1920-1940.	1620-1700	1010-1025	1410 – 1420
	1980	1410	1417 – 1440
	2030-2080	1720	1490
	2148-2200	1760	1500
		2070	1685

		2140	1780
		2310	1820
		2346	2200
		2380	2335
			2340
			2347
			2348
			2352
			2488

Em geral, ao compararmos os espectros obtidos através de transições vibracionais e transições eletrônicas, podemos constatar que os espectros no infravermelho próximo contêm informações mais abrangentes sobre os compostos presentes na amostra. As bandas obtidas pela espectroscopia vibracional proporcionam informações que podem ser utilizadas posteriormente para fins quantitativos e/ou qualitativos³⁷

5.3 Aplicabilidade

A espectroscopia no NIR destaca-se pela sua notável versatilidade em diversos aspetos. Em primeiro lugar, a sua amplitude de aplicações é notável, abrangendo desde laboratórios, fábricas, hospitais, ambientes exteriores, museus, locais de construção, estradas até a análise da atmosfera. Além disso, esta técnica pode ser aplicada a uma variedade de estados da matéria, incluindo sólidos, cristais, fibras, pós, líquidos, soluções e gases. Praticamente todos os tipos de materiais, desde amostras altamente purificadas até materiais a granel, podem ser submetidos a medições.²³

Esta técnica é mais amplamente utilizada na indústria de alimentos para animais, com o objetivo de monitorizar e garantir a qualidade de matérias-primas, produtos intermediários e produtos finais, uma vez que possibilita a estimativa da composição dos alimentos.⁴

5.4 Vantagens vs Desvantagens uso NIR

O interesse na espectroscopia NIR advém das suas vantagens em relação a outras técnicas instrumentais. De uma forma geral, esta técnica permite a análise não destrutiva de amostras, quer no estado líquido quer no estado sólido, sem a necessidade de preparação prévia, eliminando assim a necessidade de reagentes e a geração de resíduos. Adicionalmente, a aquisição dos espectros é rápida, o que a torna especialmente atrativa quando comparada com técnicas que requerem mais tempo de análise. No entanto, é importante salientar que estas vantagens dependem da construção prévia de modelos estatísticos, conhecidos como modelos quimiométricos.³⁸

No entanto, uma desvantagem notável é a baixa sensibilidade do sinal, o que pode limitar a deteção de compostos em baixas concentrações e simplificar os espectros. A dependência de outros métodos químicos de análise também é uma desvantagem, assim como o custo inicial elevado da instrumentação NIR. No entanto, avanços tecnológicos levaram ao desenvolvimento de espectrofotómetros NIR portáteis, que são mais acessíveis e fáceis de transportar.^{21,27}

Os procedimentos quimiométricos são normalmente aplicados aos espectros NIR para extrair informações relevantes, uma vez que os espectros são influenciados por características como a temperatura, o teor de humidade, o tamanho das partículas, a textura, a dispersão da luz, a manipulação da amostra e outras fontes de erro.⁴⁰

Os alimentos destinados a animais, devido à sua composição que engloba uma vasta gama de ingredientes, como soja, aveia, trigo, colza, entre outros, apresentam igualmente uma diversidade significativa de partículas com diferentes tamanhos e formas. Esta heterogeneidade pode levar a alterações nas características de reflexão e dispersão durante a análise NIR, devido às variações na densidade da amostra. No caso dos sólidos, é fundamental considerar a influência da superfície e do espalhamento da amostra, uma vez que a morfologia da superfície e o índice de refração afetam as propriedades de dispersão dos materiais sólidos. O grau de compactação, as dimensões variáveis das partículas e o acabamento da superfície podem resultar em diferenças significativas entre os espectros.^{41,42}

É de extrema importância o controlo da temperatura, uma vez que esta exerce uma influência significativa sobre a intensidade das vibrações moleculares, especialmente nas ligações de hidrogénio, amplamente reconhecidas como ligações fracas. Esta influência resulta em modificações nas bandas de absorção. Como exemplo concreto,

destacam-se as bandas de absorção da água, que, sob a influência da temperatura, sofrem um deslocamento para frequências mais elevadas.⁴²

6. Quimiometria

A designação de quimiometria apareceu pela primeira vez no ano 1971 num trabalho publicado pelo químico Svante Wold, numa revista sueca *Kemisk Tidskrift*. Todavia, foi apenas três anos mais tarde que apareceu pela primeira vez o termo *Chemometrics*, na literatura inglesa, com a publicação de um artigo baseado numa palestra feita pela palestrante Kowalski, no qual se referia a quimiometria como “um conjunto de métodos apropriados para a obtenção de informações significativas, sob o ponto de vista químico, a partir de um conjunto de dados”.⁴³

Pela IUPAC, quimiometria pode ser definida como a aplicação de estatística à análise de dados químicos e o planeamento de experiências químicas e simulações, porém devido à interdisciplinaridade que envolve esta temática, Chrétien propõe uma nova definição “Quimiometria é a administração e processamento de informações de natureza química”, atualmente a comunidade científica chegou a um consenso geral, em que assume a quimiometria como a interseção de três grandes áreas: Química, Matemática e Estatística.⁴³

O avanço das técnicas analíticas resultou na capacidade de realizar análises que geram uma grande quantidade de dados. Quando se mede a resposta de um sistema em relação a mais de duas variáveis, os dados resultantes tornam-se multivariados, e quanto maior o número de variáveis, ou seja, quanto maior dimensionalidade, mais complexa se torna a tarefa de compreender como essas variáveis afetam o sistema de forma simultânea. Portanto, através da aplicação de abordagens matemáticas e estatísticas específicas, é possível analisar uma grande quantidade de dados multivariados com o objetivo de extrair a informação que está correlacionada com uma determinada propriedade de interesse, este processo define-se como análise multivariada. Existem inúmeros métodos de análise multivariada, e a escolha do método a aplicar depende de fatores como o objetivo do modelo quimiométrico desejado e as características da matriz dos dados em questão. Na análise multivariada de dados espectrais, os métodos podem ser divididos em duas categorias principais: análise qualitativa/classificação e análise quantitativa/regressão. No contexto da análise qualitativa, o objetivo é classificar amostras com base nas características espectrais. Por outro lado, na análise quantitativa, o objetivo é determinar uma propriedade específica da amostra. Na análise de classificação, alguns métodos comuns incluem: Análise Discriminante Linear (LDA), Análise Discriminante por PLS (PLS-DA) e Análise de Componentes Principais (PCA). Em termos de análise quantitativa, existem vários métodos frequentemente utilizados:

Mínimos Quadrados Parciais (PLS), Regressão por Componentes Principais (PCR) e Redes Neurais Artificiais (ANN).

A análise multivariada de dados espectrais é uma abordagem fundamental na interpretação de informações complexas contidas em espectros NIR. Esta técnica é frequentemente empregue para lidar com a grande quantidade de dados de absorvância adquiridos em diferentes comprimentos de onda durante a análise de amostras. Cada espectro NIR pode apresentar correlações entre variáveis, tornando as absorvâncias em diferentes comprimentos de onda interdependentes, comumente denominado por multicolinearidade. A multicolinearidade pode representar um desafio ao desenvolver modelos quimiométricos robustos. Portanto, é essencial reduzir a complexidade dos dados espectrais brutos e abordar eficazmente a multicolinearidade para melhorar a precisão das análises e previsões.⁴⁴

6.1 Pré-tratamentos dos dados

O objetivo do pré-tratamento dos dados consiste, após a organização dos dados sob a forma de uma matriz, na redução das variáveis indesejáveis que não foram eliminadas durante a aquisição dos dados e que não serão removidas durante a análise subsequente, mas que ainda têm o potencial de influenciar os resultados.⁴³

No contexto da espectroscopia no infravermelho próximo (NIR), estas variações geralmente correspondem a perturbações físicas incontrolláveis nas amostras, como diferentes tamanhos de partículas ou uma distribuição não homogénea, que originam efeitos dispersivos responsáveis por alterações na linha de base e flutuações na intensidade das bandas de absorção. Adicionalmente, o pré-processamento dos dados possibilita torná-los mais comparáveis entre si, sobretudo quando se encontram em escalas distintas.⁴⁵

Contudo, é imprescindível efetuar qualquer tipo de tratamento com ponderação e critério, de forma a não distorcer, comprometer ou até mesmo inviabilizar os resultados da análise. Existem dois tipos de pré-processamento, um aplicado às amostras e outro às variáveis, designadamente transformação e pré-processamento, respetivamente.⁴³

6.1.1 Transformação dos dados

Cada amostra é submetida a um processo de pré-processamento individual, aplicando-se a mesma abordagem a todas elas. Existem várias técnicas matemáticas comuns que podem ser usadas com o objetivo de reduzir tanto as variações aleatórias

como as informações sistemáticas indesejáveis que não foram eliminadas. A componente aleatória, ou seja, o ruído experimental, pode ser reduzida através de técnicas de suavização/alisamento, enquanto as variações sistemáticas podem ser eliminadas ou reduzidas por meio de correções na linha de base e de outras transformações. Com a integração das interfaces dos instrumentos analíticos aos computadores, muitos desses procedimentos já estão incorporados nos equipamentos modernos de análise. Isso significa que, durante a aquisição dos dados, várias dessas transformações matemáticas podem ser aplicadas automaticamente.⁴³

1. Alisamento pelo método da média

Este tipo de técnica de suavização é recomendado para situações em que se pretende aumentar a relação sinal/ruído (S/R) e, simultaneamente, realizar uma compressão dos dados, o que implica uma redução no número de variáveis (J) de cada amostra. O processo envolve a seleção e o deslocamento de uma janela móvel ao longo de todo o espectro. Por exemplo, uma janela com tamanho $(2m+1)$ contém $(2m+1)$ variáveis, denominadas de $-m, -m+1, \dots, -1, 0, 1, \dots, m-1, m$, com o centro localizado em zero. A resposta obtida após a suavização (x_{alis}) é atribuída ao centro da janela, sendo uma simples média aritmética das respostas das $(2m+1)$ variáveis da respetiva janela. (equação 14)

$$x_{alis} = \sum_{k=-m}^{+m} c_k x_k = \frac{1}{(2m+1)} \sum_{k=-m}^{+m} x_k \quad (14)$$

O processo é repetido para a janela seguinte, que contém as variáveis no intervalo de $2m+1$ até $4m+2$, e assim por diante, até que todo o espectro tenha passado pelo processo de suavização. Como resultado, além da suavização, ocorre uma diminuição no número de variáveis, que inicialmente era J, e agora é aproximado para o número inteiro mais próximo de $J/(2m+1)$.⁴³

2. Alisamento pelo método da média móvel

Este método funciona de maneira semelhante ao alisamento pela média, com uma janela móvel percorrendo todo o espectro e o valor central da janela sendo substituído pela resposta média. A diferença reside no fato de que, neste caso, a janela desloca-se uma variável de cada vez, percorrendo todo o espectro. A cada deslocamento, a intensidade à esquerda é excluída e uma nova é incluída à direita. Neste método, o espectro suavizado basicamente mantém o número de variáveis que o original, não

ocorrendo a compressão dos dados. Contudo, é importante destacar que esta metodologia não permite a suavização completa do início e do final do espectro, uma vez que as respostas suavizadas são alocadas no centro das janelas.⁴³

3. Alisamento pelo método de Savitzky-Golay

A principal distinção entre o alisamento por média móvel e o método de Savitzky-Golay reside no fato de que, em vez de calcular uma média simples entre os pontos da janela, utiliza-se uma média ponderada, na qual os pesos são determinados através de um ajuste polinomial. O conceito é relativamente simples. Começa-se por ajustar um polinómio de grau n às respostas da primeira janela, que contém $(2m+1)$ pontos, indicados aqui como $k = -m, -m+1, \dots, -1, 0, 1, \dots, m-1, m$, onde $k = 0$ representa o centro. Em seguida, a resposta no ponto central da janela é substituída pelo valor do polinómio ajustado (equação 15). A janela move-se tal como no método da média móvel, e o valor do polinómio é recalculado no novo centro. Este processo repete-se até que todas as janelas tenham sido suavizadas.⁴³

$$x_k = a_0 + a_1k + a_2k^2 + \dots + a_nk^n ; k = -m, \dots, 0, \dots, +m. \quad (15)$$

A escolha do número de variáveis para uma janela é um aspeto crucial, uma vez que optar por um número reduzido pode não efetuar a filtragem adequada do ruído, enquanto a seleção de um número elevado pode resultar na eliminação de informação relevante e na distorção do espectro. Adicionalmente, o parâmetro "Gap", que determina o tamanho do intervalo utilizado no método, exerce também um efeito de suavização nas derivadas calculadas. Portanto, é essencial encontrar um equilíbrio apropriado na definição destes parâmetros de modo a garantir uma análise precisa e eficiente dos dados espectrais.

4. Alisamento com filtros de Fourier

Esta transformação também pode ser utilizada para aprimorar a relação sinal-ruído (S/R). Um espectrofotómetro com Transformada de Fourier, FTIR, captura um interferograma que incorpora informações relativas a todas as frequências emitidas pela fonte na região do infravermelho. No processo de suavização com filtros de Fourier, os espectros são transformados do domínio físico para o domínio da frequência através da Transformada de Fourier. Em seguida, o interferograma é multiplicado por uma função de apodização $A(k)$ que remove as componentes de alta

frequência no início e no final do interferograma, enquanto preserva as baixas frequências.⁴³

5. Derivadas

No contexto da espectroscopia, é possível observar um deslocamento global em todo o espectro, seja devido a questões instrumentais ou relacionadas com a amostragem. Este deslocamento resulta num desvio constante, tanto positivo quanto negativo, em relação à linha de base da absorvância, ou seja, em relação à referência zero. Para corrigir este desvio, pode-se utilizar a obtenção da primeira derivada do espectro. Dado que a derivada de uma constante é nula, após a devida correção, o espectro resultante é deslocado para o valor zero de absorvância. Contudo, se o espectro apresentar uma inclinação na linha de base à medida que o número de onda diminui, essa variação sistemática pode igualmente ser retificada através da obtenção da segunda derivada. A utilização de pré-tratamento através de derivadas é uma prática comum na análise de espectros de refletância difusa, uma vez que problemas relacionados com deslocamentos e inclinações da linha de base surgem frequentemente neste contexto.^{43,46}

A primeira derivada atenua desvios provocados por efeitos aditivos na linha de base, efetuando a correção do respetivo deslocamento. Simultaneamente, realça características de menor intensidade relativamente às de maior intensidade. Este processo torna as estruturas de menor amplitude mais proeminentes, simplificando a sua avaliação.⁴⁶

A derivada de uma função num ponto qualquer da curva representa a inclinação da reta tangente à curva nesse ponto. Quando aplicada a um pico, a sua derivada apresenta dois pontos extremos: um positivo, referente ao lado esquerdo do pico, e outro negativo, relacionado ao lado direito do pico. Os máximos positivo e negativo na derivada correspondem aos pontos de inflexão na curva original. No ponto de máxima intensidade do pico, a inclinação da reta tangente é nula, o que implica que a derivada é igual a zero. Uma maneira simples de calcular derivadas é utilizando a diferença entre dois valores sucessivos de absorvância, conforme indicado na equação 16, sendo que os intervalos de comprimento de onda ($\Delta\lambda = \delta$) são constantes.

$$2\delta \frac{dA}{d\lambda}(\lambda_j) = \Delta A(\lambda_j) = A_{\lambda_{j+\delta}} + A_{\lambda_{j-\delta}} \quad (16)$$

A segunda derivada, que corresponde à derivada da primeira derivada, é igualmente valiosa, pois fornece informações sobre a curvatura ou concavidade de uma curva. No ponto de inflexão, a concavidade é nula, enquanto no pico, onde a absorvância é máxima, a segunda derivada apresenta o seu valor negativo máximo. Como resultado, o gráfico da segunda derivada assemelha-se mais ao pico original do que o da primeira derivada, tornando a interpretação dos resultados mais simplificada. A equação 17 refere-se ao cálculo da segunda derivada, realizado através do método de diferenças, considerando δ como um pequeno intervalo de comprimentos de onda.⁴³

$$(2\delta)^2 \frac{d^2 A}{d^2 \lambda}(\lambda_j) = \Delta^2 A(\lambda_j) = A_{\lambda_{j+\delta}} + A_{\lambda_{j-\delta}} - 2A_{\lambda_j} \quad (17)$$

No que diz respeito à segunda derivada, esta também elimina inclinações na linha de base, mas amplifica o ruído espectral, tornando essencial a sua aplicação em regiões espectrais mais limitadas.⁴⁶

Apesar da simplicidade inerente ao método de cálculo da derivada por diferenças, este apresenta uma significativa desvantagem. Em primeiro lugar, o número de variáveis no espectro derivado é inferior ao do espectro original. Como resultado, o valor máximo de um pico no espectro original não se traduzirá em zero no espectro da primeira derivada, nem se alinhará com um mínimo no espectro da segunda derivada. Este método só funciona de acordo com o desejado para espectros de alta resolução. É de extrema relevância considerar que a relação entre a área sob um pico no espectro e a concentração é preservada ao aplicar a derivada. O que significa que os detalhes espectrais cruciais para a correlação entre as absorvância e as concentrações não sofrem alterações, e, conseqüentemente, a correspondência quantitativa entre o espectro e a concentração mantém-se válida.⁴³

O método mais amplamente utilizado para calcular derivadas de um sinal é o método de Savitzky-Golay. Este método é uma técnica de filtragem de média móvel na qual é efetuado um ajuste por mínimos quadrados de um polinómio de grau n aos pontos $(2m+1)$ da janela móvel ou segmento. Dado que cada ponto da curva é representado por uma expressão polinomial (equação 15), a derivada é obtida diretamente a partir da expressão analítica do polinómio.

A primeira derivada de x em relação a k é dada pela seguinte expressão:

$$\frac{d_x}{d_k} = a_1 + 2a_2 k + \dots + na_n k^{n-1} \quad (18)$$

Tendo em conta, o ponto central da janela como foco de interesse, tem-se que:

$$\left(\frac{d_x}{d_k}\right)_{k=0} = a_1 \quad (19)$$

O que significa que os pesos usados para determinar a primeira derivada, no ponto central da janela dependem somente de a_1 .

Em suma, o uso de derivadas em geral é o método mais apropriado para a correção de linhas de base em espectros. O seu cálculo não envolve subjetividade associada a funções paralelas e não introduz uma grande quantidade de variância nos dados. No entanto, a única desvantagem significativa é a redução indesejável da relação sinal-ruído. Isto não representa um problema quando a relação sinal-ruído é alta, mas pode ser inconveniente quando a relação sinal-ruído é baixa.^{20,43}

6. Logaritmo

A utilização do logaritmo pode ser aplicada com o objetivo de linearizar os dados, e a escolha da base logarítmica, seja ela decimal ($\log_{10}$) ou qualquer outra, não terá impacto na interpretação qualitativa dos resultados.

Por exemplo, na área da espectroscopia, a resposta medida corresponde à razão entre a intensidade que atravessa a amostra e a intensidade incidente na amostra, denominada por transmitância (T) e é calculada como indicado anteriormente na equação 11. É importante destacar que o decréscimo na intensidade luminosa à medida que atravessa uma amostra homogênea e transparente não é linear. No entanto, ao aplicar uma transformação logarítmica à transmitância, é possível estabelecer uma relação linear entre a intensidade luminosa e a concentração da substância em análise.⁴³

7. Normalização

Esta transformação envolve a divisão dos valores de cada variável de uma amostra por um fator de normalização, que pode ser a norma dessa amostra. Dessa forma, todas as amostras passam a ter uma escala pré-determinada.⁴³

8. Correção Multiplicativa de Sinal (MSC)

A transformação MSC (*Multiplicative Scatter Correction*) é utilizada para corrigir os efeitos de espalhamento aditivos e multiplicativos na absorvância. Os efeitos de espalhamento são causados por fenómenos físicos, como alterações no percurso ótico, na sensibilidade do detetor e do amplificador, flutuações na temperatura e na pressão, bem como variações no tamanho e na forma de partículas sólidas (pós e grãos), além de emulsões e dispersões. É relevante salientar que estes efeitos não estão relacionados com a composição das amostras. Os efeitos de espalhamento aditivos e multiplicativos são dependentes do comprimento de onda e devem ser eliminados, visto que representam fontes de variabilidade que não são pertinentes para o problema em análise. A correção desses efeitos de espalhamento implica o deslocamento e a adaptação de cada espectro de modo a ajustá-lo ao espectro "ideal". Os parâmetros de deslocamento retificam os efeitos aditivos, ao passo que os parâmetros de escala corrigem os efeitos multiplicativos. Ambos os efeitos podem ser estimados simultaneamente através de uma simples regressão linear.⁴³

Inicialmente, o MSC constrói um modelo de regressão linear do espectro médio em relação ao espectro de cada medição.⁴⁷

$$X_{ij} = k_{ij} \times \bar{X} + b_{ij} \quad (20)$$

Seguidamente, o espectro medido das amostras i e variáveis j ($X_{i,j}$) é corrigido tendo em conta o coeficiente do declive, $k_{i,j}$, e o coeficiente da interceção, $b_{i,j}$ ⁴⁷:

$$\widetilde{X}_{ij} = \frac{X_i - b_i}{k_i} \quad (21)$$

Aprimorando ainda mais as técnicas de pré-tratamento, a transformação MSC foi estendida, dando origem à correção de dispersão multiplicativa estendida (EMSC). O EMSC demonstra uma notável utilidade ao minimizar a variação da dispersão da luz, uma característica que apresenta dependência em relação ao comprimento de onda.⁴⁸

A vantagem do uso do MSC é a preservação da forma original do espectro, o que facilita a interpretação dos resultados, embora o MSC não aborde a correção da inclinação na linha de base.⁴³

9. Varição de Padrão Normal (SNV)

A normalização pelo desvio padrão normal (SNV), é um dos algoritmos mais prevalentes para mitigar interferências ocasionadas por disparidades na densidade das amostras e no tamanho das partículas, tal como a MSC. A aplicação deste método permite que os espectros revelem exclusivamente informações estruturais, eliminando a informação relacionada com a intensidade.⁴³

O método SNV, do inglês *Normal Variate Method*, normaliza cada espectro calculando e removendo o valor médio, bem como ajustando à variância, ou seja, baseia-se no princípio de centralizar o espectro em torno de zero, através da subtração do valor médio das intensidades do i -ésimo espectro ($\bar{X}$), e escalar os valores dividindo pelo desvio padrão das intensidades desse espectro (s_i) em todos os pontos dos espectros. (equação 22)

$$\tilde{X}_i = \frac{X_i - \bar{X}}{s_i} \quad (22)$$

Onde :

$$\bar{X} = \frac{1}{J} \sum_{j=1}^J X_{ij} \text{ e } s_i = \sqrt{\sum_{j=1}^J (X_{ij} - \bar{X})^2} \quad (23)$$

Devido à ausência de um espectro médio de referência, o SNV é considerado mais robusto em relação à presença de amostras atípicas (*outliers*) em comparação com o MSC.²⁰

O modelo SNV não simplifica o modelo matemático, nem reduz as interferências sistemáticas, mas contribui para a melhoria da precisão dos resultados.^{43,47}

10. Correção ortogonal de sinal (OSC)

A transformação OSC (*Orthogonal Signal Correction*) foi recentemente introduzida como uma alternativa à correção MSC, com o propósito de aprimorar a capacidade preditiva de modelos de calibração, embora já tenha sido previamente empregue noutras técnicas espectroscópicas. Este método fundamenta-se no facto de que a maior parte da variação sistemática presente na matriz de dados, X , apresenta uma correlação reduzida com a propriedade de interesse, tendo, consequentemente, um

valor preditivo limitado. Portanto, é desejável eliminar a variância ortogonal à propriedade de interesse do conjunto de dados. O aspecto notável desta correção espectral é que, ao eliminar as variações ortogonais, a complexidade do modelo, ou seja, o número de fatores no modelo, tende a diminuir, contudo, a sua capacidade preditiva mantém-se inalterada.⁴³

6.1.2 Pré-Processamento

Após a aplicação de alguns dos principais pré-tratamentos às amostras, é fundamental agora considerar os procedimentos de pré-processamento aplicados às variáveis, ou seja, a cada uma das colunas da matriz de dados. Este é um dos procedimentos de rotina mais significativos em qualquer análise quimiométrica.

1. Centragem dos dados na média

A centralização na média, também conhecida como *Mean Centering*, consiste na simples subtração da absorvância em cada comprimento de onda pela absorvância média correspondente a esse comprimento de onda em todas as amostras do conjunto (equação 24). Desta forma, o novo valor médio de absorvância para todas as amostras é igual a zero em cada comprimento de onda. Do ponto de vista estatístico, a centralização tem como objetivo evitar que os pontos mais distantes do centro dos dados tenham maior influência do que os pontos mais próximos, priorizando não apenas a intensidade observada, mas também a sua distância em relação ao valor médio. Geralmente, resulta na redução do número de fatores do modelo.⁴³

$$x_{ij(cm)} = x_{ij} - \bar{x}_j \quad (24)$$

Onde $\bar{x}_j$ é a média j -ésima coluna dos dados.

2. Escalamento pela Variância

Os dados são sujeitos a uma operação de escalonamento pelo fator de escala, que é uma medida de dispersão dos dados. Neste contexto, os dados são normalizados utilizando a variância como fator de escala. Isso implica que os elementos de cada uma das variáveis são divididos pelo fator de escala, que equivale ao respetivo desvio padrão.

$$x_{ij(v)} = \frac{x_{ij}}{s_j} \quad (25)$$

Onde s_j^2 é a variância da j -ésima variável.

3. De-Trending

Na etapa de pré-processamento, um dos procedimentos amplamente reconhecidos como sendo de elevada relevância é a remoção de tendências (De-Trending). O método de DT, ou redução de tendência espectral, é aplicado quando existe uma dependência em relação ao comprimento de onda, normalmente encontrada em espectros de refletância NIR de amostras em pó. Este método permite ajustar a subtração de polinômios, geralmente de segunda ordem, em cada espectro através de uma análise de regressão. Desta forma, o DT corrige as variações no deslocamento e na curvatura da linha de base, ou seja, procede à correção da linha de base. O uso de polinômios de ordens superiores pode ser considerado, no entanto, deve ser realizado com cautela, uma vez que existe um risco associado à remoção de informações relevantes. Normalmente, a aplicação da técnica DT é combinada com a técnica SNV, o que possibilita melhorar a dependência em relação aos comprimentos de onda ao remover a curvatura e ajustar o desvio da linha de base nos espectros.⁴⁹

O procedimento envolve a subtração de um polinómio ajustado em cada espectro i , que pode ser definido por uma equação polinomial de segunda ordem, como a seguinte:

$$x_{ij} = a_i + b_i \lambda_j + c_i \lambda_j^2 + e_{ij} \quad (26)$$

Onde a interceção a_i e os parâmetros de coeficiente de regressão b_i e c_i , foram determinados pelo método LS. O espectro sem tendência, é calculado pela seguinte expressão:

$$x_{i(detrend)} = x_{ij} - (a_i + b_i \lambda_i + c_i \lambda_i^2) \quad (27)$$

Em ambas equações λ_j corresponde à banda do comprimento de onda j .

4. Autoescalamento

O autoescalamento, também conhecido como *standardization*, é um método de normalização amplamente utilizado em dados provenientes de variáveis discretas, como medidas físico-químicas e concentrações de elementos, especialmente em contextos de modelação exploratória.⁴³

Este procedimento assemelha-se bastante ao SNV, uma vez que começa por centralizar os dados pela média e, subsequentemente, divide cada elemento pelo desvio padrão correspondente da respetiva coluna. Como resultado, todas as variáveis apresentam uma média igual a zero e uma variância igual a um, colocando-as numa mesma escala relativa e eliminando qualquer influência das escalas originais das variáveis. Este processo assegura que todas as variáveis tenham o mesmo peso, independentemente das suas escalas iniciais. Contudo, é crucial destacar que o autoescalamamento não é aconselhável para a aplicação a espectros em geral e, sobretudo, a espectros NIR. Nos contextos dos espectros, a aplicação do autoescalamamento pode ser desaconselhável, visto que os picos de maior intensidade frequentemente contêm informações mais informativas para o modelo. Adicionalmente, o autoescalamamento dos espectros pode amplificar a contribuição de variáveis ruidosas, especialmente em regiões espectrais com baixa absorbância, o que pode comprometer a qualidade do modelo resultante. Portanto, a decisão de aplicar ou não o autoescalamamento deve ser cuidadosamente ponderada, tendo em consideração as características específicas dos dados e os objetivos da análise em questão.²⁰

6.2 Análise Exploratória dos Dados

Após a etapa de pré-processamento, é ainda imperativo proceder à redução das variáveis, dada a vasta quantidade de informação contida nos espectros, eliminando aquelas que não demonstram correlação. Os métodos quimiométricos utilizados para identificar as semelhanças e diferenças em várias amostras de natureza diversa, com vista à sua agrupação e classificação, dividem-se em dois grupos distintos: os métodos "supervisionados" e os métodos "não supervisionados". Ambos os conjuntos de métodos se fundamentam na validade das seguintes pressuposições: as amostras do mesmo tipo apresentam semelhanças, existem diferenças significativas entre dois tipos distintos de amostras e o conjunto de medidas disponíveis é adequado para analisar essas semelhanças e diferenças. Nos métodos supervisionados, cada amostra analisada pertence a uma classe predefinida e esta informação é empregue durante a análise dos dados e na construção dos modelos de classificação. Os métodos não supervisionados prescindem dessa informação e, por conseguinte, não necessitam de qualquer conhecimento prévio acerca da classificação das amostras. Estas serão agrupadas de forma natural com base na informação contida nos dados experimentais em análise. Dois métodos matemáticos de análise exploratória, com conceções totalmente distintas e que podem ser aplicados a conjuntos de dados multivariados para possibilitar a visualização desejada, são a Análise de Componentes

Principais, frequentemente abreviada como PCA, e a Análise de Agrupamentos através de Métodos Hierárquicos, conhecida como HCA.⁴³

A Análise de Componentes Principais (PCA), ou *Principal Component Analysis* em inglês, foi introduzida por Karl Pearson em 1901. No entanto, o desenvolvimento formal deste método é atribuído ao trabalho de Hotelling, divulgado na década de 1930, o qual teve um impacto revolucionário na aplicação de métodos multivariados no domínio da psicologia. A PCA enquadra-se num conjunto de métodos pertencentes à categoria de classificação não supervisionada. Nestes métodos, não são disponibilizadas informações prévias sobre a categoria das amostras. Em vez disso, estabelecem-se critérios matemáticos que permitem expressar quantitativamente a semelhança entre duas amostras ou entre uma amostra e uma classe. Essa semelhança pode ser quantificada através de uma medida de distância, como a distância de Mahalanobis (H), que avalia a distância entre uma amostra e o centro do conjunto de amostras. Desta forma, é possível reduzir a dimensionalidade do espaço original do conjunto de dados, preservando as relações entre as amostras. Assim, as informações pertinentes são isoladas e realçadas, tornando-se mais visíveis para a inspeção visual. Através desta abordagem, é possível descobrir, visualizar e interpretar as discrepâncias presentes entre as variáveis, bem como examinar as relações que podem existir entre as amostras (objetos). Esta análise também possibilita a deteção de amostras que manifestem um comportamento atípico, uma vez que, com a projeção dos dados, estas amostras tendem a tornar-se evidentes, sendo comumente referidas como *outliers*. Para ilustrar o funcionamento do método de projeção, considera-se um exemplo simplificado, composto por um conjunto de pontos no espaço tridimensional R^3 . Coloca-se uma superfície plana atrás desses pontos e ilumina-se com luz incidente. Observa-se então as projeções destes pontos na superfície plana. Essencialmente, o que acontecesse é a projeção dos dados do espaço tridimensional R^3 para um espaço bidimensional R^2 na superfície plana, seguindo a direção da iluminação, resultando na redução da dimensão do espaço original.⁴³

A PCA é uma ferramenta quimiométrica com uma vasta gama de aplicações em várias áreas científicas. É frequentemente utilizada na identificação de grupos distintos e na seleção de amostras para a construção de modelos de calibração. A análise envolve a redução da dimensionalidade da matriz de dados espectrais (X), resultando na formação de componentes principais (PCs) que condensam as informações dos dados num menor número de variáveis. Com esta abordagem, torna-se possível estudar a

representação de uma amostra de natureza multivariada em dimensões mais reduzidas, o que facilita significativamente a sua visualização.^{20,43}

O ponto de partida para a análise exploratória de dados é a matriz pré-tratada dos dados, denominada por matriz X, descrita da seguinte maneira:

$$X = \begin{bmatrix} X_1^T \\ \vdots \\ X_I^T \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} \\ x_{21} & x_{22} & \cdots & x_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & x_{Ij} \end{bmatrix} = [X_1 \quad X_2 \quad \cdots \quad X_j] \quad (28)$$

Cada amostra é representada por um vetor-linha e cada variável por um vetor coluna:

$$X_i^T = [x_{i1} \quad x_{i2} \quad \cdots \quad x_{ij}] \text{ e } X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \vdots \\ x_{Ij} \end{bmatrix} \quad (29)$$

Do ponto de vista matemático, a PCA pode ser descrita como a decomposição da matriz X (I×J) em duas matrizes, uma de *scores* T e uma matriz *loadings* P, de tal maneira que:

$$X = TP^T \quad (30)$$

Onde os scores expressam as relações entre as amostras, enquanto os loadings indicam as reações entre as variáveis.

Havendo correlação entre as variáveis originais, nem todas as J componentes principais são necessárias para representar adequadamente os dados, mas apenas um subconjunto delas, as primeiras A, as restantes devem conter informação irrelevante ou aleatória.

Uma vez encontrado o valor de A, tendo em conta os produtos da matriz *scores* T (I, A) e uma matriz de *loadings* P (J,A), considerando o subconjunto, a matriz original pode ser reescrita da seguinte forma:

$$X = t_1 p_1^T + t_2 p_2^T + \cdots + t_A p_A^T + E = TP^T + E \quad (31)$$

Sendo A o número de Componentes Principais (PCs) no modelo PCA, t_1 e p_1 os vetores *scores* e *loadings* do primeiro PC, respetivamente, e E a matriz ($I \times J$) de resíduos.³¹

Os Componentes Principais, do inglês *Principal Components* (PCs) são definidos de forma a explicar as variâncias dos dados espectrais. Estes PCs são modelados em ordem decrescente das informações estatísticas, ou seja, o PC1 explica uma variância superior em relação ao PC2, o PC2 explica uma variância superior ao PC3 e assim sucessivamente. Cada PC é uma combinação linear das variáveis originais, no entanto, existe uma restrição ortogonal entre eles, indicando que os PCs são independentes e não estão correlacionados. Geralmente, apenas os primeiros PCs são responsáveis pela maior quantidade de variância dos dados e podem ser usados para representar o conjunto de dados de maneira mais simplificada.^{20,43}

No que diz respeito aos resíduos do modelo (E), este diminui à medida que aumenta o número de PCs, utilizados no modelo. Tal sucede porque, com o aumento do número de PCs, mais informação é retida, o que resulta na redução da quantidade de resíduos.

Do ponto de vista algébrico, os vetores *loadings* representa o cosseno dos ângulos entre o respetivo PC e o eixo das variáveis originais, ou seja, o vetor p fornece a informação do quanto a variável original contribuiu para a formação dos PCs. Em contrapartida, os vetores *scores* são as projeções ortogonais das amostras nos eixos de cada PC. Para a interpretação de modelos PCA, é fundamental entender que os *scores* fornecem a composição de cada PC relacionada aos objetos/amostras, enquanto os *loadings* relacionam essa mesma composição com as variáveis mais importantes. Uma das etapas críticas na PCA é a determinação do número apropriado de Componentes Principais. Idealmente, o número de PCs significativos deve corresponder ao número de fontes de variação linearmente independentes presentes nos dados. Por exemplo, na análise de espectros de misturas de compostos, esse número deve ser igual ao número de componentes químicos que contribuem de forma não correlacionada para o sinal espectral. O número de PCs no modelo de PCA depende da natureza específica dos dados. Uma das abordagens mais intuitivas para essa estimativa é baseada na quantidade de variância explicada pelo modelo PCA (V_{exp}).²⁰

Após a determinação do número de PCs, os dados da matriz X são projetados para um espaço de menor dimensão. A independência dos PCs torna possível investigar as relações entre as amostras nos primeiros PCs e as relações entre as variáveis nos

respetivos *loadings*. Os gráficos resultantes possibilitam a retenção de informações relevantes sobre tendências e características estruturais dos dados, bem como a identificação de grupos de amostras com características semelhantes e a deteção *outliers*.

Em suma, a Análise de Componentes Principais é um método com a finalidade de reduzir dados através da criação de combinações lineares das variáveis originais. Esta abordagem é amplamente aplicada na espectroscopia NIR, onde espectros com mais de mil variáveis podem ser visualizados de forma bidimensional. Adicionalmente, este modelo permite que apenas a informação relevante seja mantida, eliminando, por exemplo, variações associadas ao ruído.

6.3 Calibração Multivariada

Num processo de calibração, a quantificação da concentração de um componente ou de outra propriedade de interesse é realizada de forma indireta, através de medidas efetuadas num sistema químico. Quando uma variável seletiva que possa ser medida está disponível para esse sistema, recorre-se à calibração univariada. No entanto, em muitos casos, nomeadamente nos espectros NIR, é raro encontrar variáveis seletivas, mesmo em matrizes menos complexas. Assim, a alternativa é a utilização de métodos de regressão multivariada, como por exemplo Regressão Linear Múltipla (MRL), Regressão por Componentes Principais (PCR), Regressão por quadrados mínimos parciais (PLS), entre outros.

Para o desenvolvimento do modelo, são necessários pelo menos dois conjuntos de amostras. O primeiro é o conjunto de calibração, no qual os valores das concentrações ou propriedades a serem determinados são conhecidos, seja através da preparação de amostras com composição planeada ou pelo uso de métodos de referência, o qual se encontra representado na figura 9 pelo bloco y no caso de se tratar de apenas uma única propriedade de interesse, caso seja mais do que uma então designa-se de bloco Y e são designados de variáveis dependentes. Estes valores, juntamente com os espectros correspondentes, representados pelo bloco X na figura 9 e designado por variáveis independentes, são utilizados para estimar os coeficientes de regressão. O segundo conjunto de dados é o de validação, no qual o modelo desenvolvido é testado para prever a concentração de amostras independentes, que não foram incluídas na calibração.

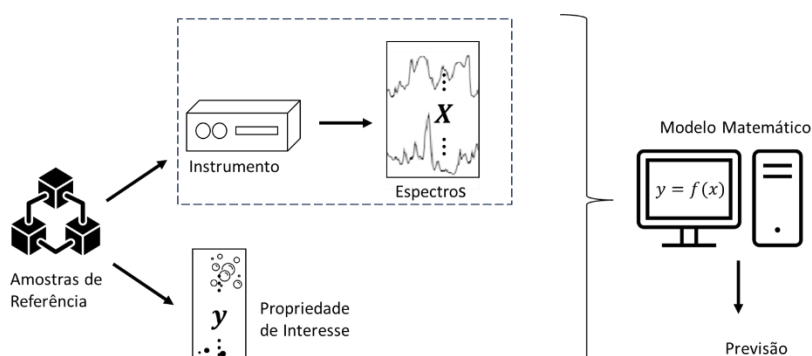


Figura 9: Representação esquemática da transformação das respostas experimentais em informação química através do modelo matemático.

6.3.1 Regressão por Mínimos Quadrados Parciais (PLS)

A Regressão por Mínimo Quadrados é uma ferramenta útil na construção de modelos preditivos perante um elevado número de variáveis (superior ao número de observações) que apresentam correlações significativas entre si. O PLS é um método de regressão multivariada amplamente utilizado para estabelecer a relação entre variáveis dependentes (Y), que dizem respeito à propriedade de interesse e variáveis independentes (X), que correspondem por exemplo aos espectros NIR. Com a decomposição simultânea de X e Y , ocorre uma ligeira perda da ortogonalidade das componentes principais, estas componentes recebem a designação mais ampla de Variáveis Latentes (VLs). As VLs são combinações lineares das variáveis em X que modelam de forma otimizada as variáveis dependentes em Y de maneira simultânea, são obtidas maximizando a covariância entre os scores t da matriz X e Y . O termo “parcial” do PLS deve-se ao facto de que a solução dos quadrados mínimos não se aplica a qualquer conjunto de pesos, mas somente aqueles que satisfazem a restrição de maximizar a covariância.⁴³

Em suma, na construção de modelos PLS, como descrito na figura 10, procede-se à decomposição simultânea das variáveis X e Y num conjunto de h Variáveis Latentes, que descrevem a direção de máxima variância. A matriz X possui dimensão $n \times p$, em que n corresponde ao número de amostras e p representa o número de variáveis independentes medidas. Por outro lado, a matriz Y pode ser um vetor (y) de dimensão $n \times 1$, quando o modelo PLS visa a previsão de apenas um analito, ou uma matriz de dimensão $n \times k$, quando o PLS prevê simultaneamente k analitos.

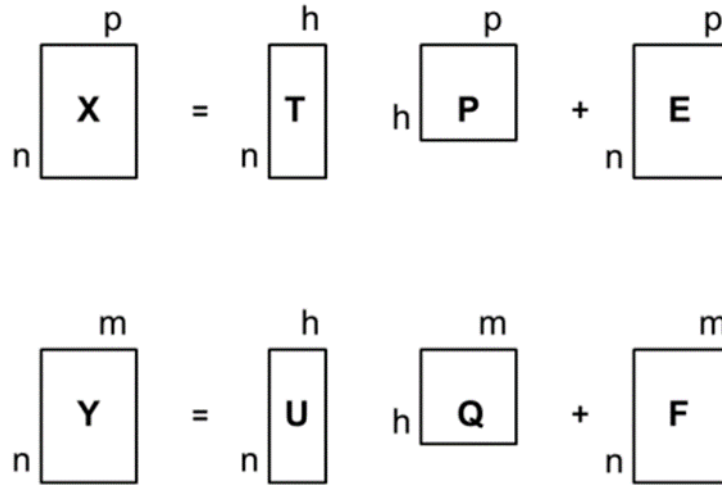


Figura 10: Esquema ilustrativo da decomposição dos dados pelo método dos mínimos quadrados. Adaptado ⁵⁹

As equações gerais do modelo regressão PLS, para decomposição da matriz X (Equação 32) e Y (Equação 33) são expressas da seguinte forma

$$X = TP' + E \quad (32)$$

$$Y = UQ' + F \quad (33)$$

Na equação apresentada, a matriz T representa os scores, a matriz P representa os loadings e a matriz E representa o erro residual de X , em contrapartida as matrizes U , Q e F englobam os scores, loadings e o erro residual de Y , respetivamente. Os scores representam as projeções das observações no espaço dos fatores obtidos através da regressão, enquanto os loadings correspondem às contribuições de cada variável na variação dos fatores. As matrizes residuais contêm a variância restante que não é explicada pelo modelo de regressão.

Uma característica fundamental do PLS é a consideração dos erros em ambas as matrizes X e Y , assumindo que esses erros estão igualmente distribuídos. Existem vários algoritmos para descobrir os fatores latentes e criar regressões por PLS, sendo o algoritmo de Mínimos Quadrados Parciais Iterativos Não Lineares (NIPALS) o mais comumente empregue.⁴⁶

De uma forma sucinta o algoritmo NIPALS na regressão por PLS inicia-se selecionando uma coluna de Y como uma estimativa inicial para os scores de Y (u_a),

posteriormente cada coluna de X é relacionada com os scores estimados u_a e assim obtém-se K regressões. Os coeficientes das regressões obtidas são tidos como os pesos de X (w_a). Os pesos obtidos descrevem a covariância entre as variáveis preditoras e as respostas e asseguraram que os fatores obtidos são ortogonais, ou que não se encontram relacionados. De seguida, cada linha da matriz X é relacionada com os w_a , surgindo novos coeficientes os quais são guardados como os scores de X (t_a), os quais são depois usados para calcular os *loadings* de X . Posteriormente, todas as colunas de Y são relacionadas com t_a , surgindo os pesos de Y (c_a) através dos coeficientes originários desta relação. Cada linha de Y é relacionada com c_a , caso estejam altamente relacionadas consequentemente terão coeficientes eles em u_a , devendo-se repetir novamente o processo até este ponto, terminando quando não houver variação significativa de u_a .⁵⁰

Os scores de X para além de caracterizem a matriz X , também funcionam como bons preditores de Y , podendo a propriedade de interesse ($\hat{y}$) ser estimada pela multiplicação da matriz X com um vetor regressão apropriado (b) obtido através dos parâmetros determinados pelo algoritmo da regressão.⁵⁰

O vetor de coeficientes de regressão linear (b) correlaciona os blocos X e Y de forma linear, para h VLs, da seguinte forma:

$$u_h = b_h t_h \quad (34)$$

Existem duas variantes do método PLS, conhecidas como PLS1 e PLS2. No método PLS1, cada coluna de Y é modelada individualmente, ou seja, para cada componente de interesse, é calculado um conjunto separado de scores e variáveis latentes. Por outro lado, no Método PLS2, é calculado um único conjunto de scores e variáveis latentes que abrange todas as colunas da matriz Y . Este conjunto é comum a todas as propriedades de interesse que estão a ser modeladas.⁴³

6.3.2 Regressão por Mínimos Quadrados Parciais Modificado(MPLS)

A Regressão por Mínimos Quadrados Parciais Modificados (MPLS) é uma ferramenta eficiente e amplamente utilizada na monitorização estatística multivariada de processos, é especialmente útil quando se necessita desconsiderar as principais fontes de variação espectral e concentrar-se em regiões com menor variação espectral, mas que apresentam fortes correlações.^{51,52}

O modelo MPLS pode ser estabelecido por X (equação 35) e Y (equação 36) da seguinte forma

$$X = TP^T + E \quad (35)$$

$$Y = XM + E \quad (36)$$

Em que M corresponde à matriz de coeficientes.⁵¹

O MPLS revela-se frequentemente mais estável e preciso do que o algoritmo PLS convencional. Nesta abordagem, os resíduos obtidos em cada comprimento de onda do NIR, após o cálculo de cada fator, são normalizados dividindo-os pelos desvios-padrão correspondentes. Esta normalização precede o cálculo do fator subsequente. Durante o desenvolvimento das equações MPLS, é altamente recomendável a aplicação de técnicas de validação cruzada para selecionar o número ótimo de fatores, prevenindo assim o sobre ajuste do modelo aos dados.^{51,53}

A única desvantagem observada ao trabalhar com MPLS reside na interpretação dos coeficientes de regressão. Estes não apresentam a mesma suavidade que os coeficientes PLS, tornando-se mais difíceis de relacionar com picos constituintes conhecidos. Contudo, é importante notar que os coeficientes da maioria dos modelos PLS, especialmente aqueles com pré-tratamento diversos e muitos fatores, também são difíceis de interpretar.⁵²

6.4 Validação do Modelo

Após a construção do modelo, é crucial validá-lo, testando-o em amostras não usadas durante a sua criação. A validação pode ser feita por validação cruzada (testando o modelo em subconjuntos dos dados) e validação empírica (avaliando-o em dados independentes). Ambos os métodos são essenciais para garantir a precisão das previsões. A validação cruzada determina o modelo e o número ideal de fatores, enquanto a validação empírica assegura a robustez do modelo.⁴³

6.4.1 Validação Cruzada

A seleção do número ótimo de fatores é crucial para a qualidade do modelo de regressão. Optar por um número de fatores inferior ao necessário pode resultar numa calibração insuficiente, já que o modelo terá menos informação do que necessita (*Underfitting*). Por outro lado, ao utilizar um número excessivo de fatores, o modelo pode ser prejudicado devido à inclusão de ruído e variações sistemáticas (*Overfitting*). Portanto, o objetivo é encontrar o número ideal de fatores que garanta uma previsão precisa para novas amostras, minimizando o erro associado.^{31,43}

A validação cruzada é um processo de reamostragem que visa determinar o número ideal de variáveis latentes que resulta no menor erro possível. Nos diferentes tipos de validação cruzada, as amostras são retiradas repetidamente do conjunto de dados para serem testadas pelo modelo construído. Um exemplo é o método LOO (*leave-one-out*), onde cada amostra é removida individualmente e testada. Este método é adequado para pequenos conjuntos de dados, geralmente com cerca de 20 amostras. Para conjuntos maiores, recomenda-se dividir as amostras em grupos e prever os grupos de forma aleatória, sistemática ou usando métodos venezianos (*venetian blinds*), onde as amostras são retiradas em intervalos constantes. Este processo gera parâmetros úteis para avaliar e aprimorar o modelo, incluindo a soma dos quadrados dos erros residuais previstos (PRESS), o erro quadrático médio da validação cruzada (RMSECV) e o erro quadrático médio de previsão (RMSEP).³¹

Os valores de PRESS são uma estimativa útil para selecionar o número ideal de variáveis latentes, minimizando o risco de *overfitting* e *underfitting*. Este parâmetro é calculado subtraindo o valor de referência (y_i) do valor previsto pelo modelo, excluindo uma medição ($\hat{y}_{i-1}$) (equação 37).

$$PRESS = \sum_{i=1}^n (y_i - \hat{y}_{i-1})^2 \quad (37)$$

O erro obtido por esta validação é conhecido como RMSECV e é calculado pela equação 38, onde n é o número de amostras de calibração, y_i é o valor de referência da amostra i , $\hat{y}_i$ é o valor previsto para a amostra i pelo modelo determinado. Portanto, um baixo valor de RMSECV indica um bom modelo, indicando que as previsões do modelo são próximas aos valores de referência das amostras.³¹

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n - 1}} \quad (38)$$

Para concluir o modelo de calibração, é essencial testá-lo utilizando amostras com concentrações conhecidas, que fazem parte do conjunto de validação e não são incluídas no conjunto de calibração. Nesta técnica, é calculado um erro denominado RMSEP (equação 39), que representa a diferença entre os valores previstos e os valores de referência das amostras i no conjunto de validação, composto por n amostras. O RMSEP é crucial para avaliar o quão bem o modelo escolhido pode prever

valores de amostras não utilizadas durante o processo de calibração. Quanto menor o valor do RMSEP, mais preciso é o modelo nas previsões.³¹

$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (39)$$

A calibração cruzada é preferível em situações onde o número de amostras não é suficiente para dividir entre grupos de calibração e validação. Em cenários onde há um número adequado de amostras, a validação externa é mais vantajosa, pois torna o sistema mais robusto. A robustez refere-se à capacidade do método em lidar com variações causadas por diferentes fatores, como operador, condições ambientais, compactação do material, alterações nos instrumentos, temperatura da amostra e sua idade. Um modelo robusto é menos sensível a variações irrelevantes e exibe maior estabilidade e aplicabilidade a longo prazo. Idealmente, estas variações devem ser incluídas tanto no conjunto de calibração quanto no conjunto de validação. Dessa forma, o sistema pode responder de maneira eficaz a novas amostras que apresentem essas variações, garantindo resultados precisos e consistentes.^{31,43}

6.4.2 Validação Externa

A qualidade da calibração é geralmente avaliada pela precisão na previsão de novas amostras desconhecidas, obtida durante o desenvolvimento do método. Durante este processo, calcula-se a média global do erro de previsão. O erro de previsão dos componentes de uma amostra pelo método NIR corresponde ao RMSCEV no caso da validação cruzada ou ao RMSEP no caso da utilização de validação externa. Na validação externa, os valores previstos pelos modelos desenvolvidos são comparados com os resultados obtidos por análises laboratoriais de referência. Esta comparação é essencial para avaliar a precisão e a confiabilidade do modelo, garantindo que ele seja capaz de prever de forma acurada os dados de amostras não utilizadas durante a construção do modelo. Para classificar um modelo de calibração multivariada como bom, é crucial assegurar que o modelo explique a maior parte da variância nos blocos X e Y. Tipicamente, um modelo de qualidade engloba aproximadamente 90% de toda a variância. Em situações onde a calibração não é satisfatória, observa-se uma grande discrepância entre os valores do erro padrão de previsão e do erro padrão de calibração.^{31,43}

6.5 Outliers

Outliers é a designação dada às amostras anómalas, as quais destacam-se no conjunto de dados devido ao seu comportamento distinto das restantes observações

do conjunto de calibração e/ou previsão. No contexto da calibração e previsão estatística, a presença de outliers pode comprometer a qualidade dos modelos desenvolvidos. A presença destes outliers no conjunto de calibração pode resultar em modelos com baixa capacidade de previsão e quando estão presentes no conjunto de validação, podem influenciar os resultados, frequentemente levando a conclusões que indicam que o modelo não é adequado, quando na realidade a capacidade de previsão poderia ser alcançada na sua ausência. Estas discrepâncias frequentemente resultam de medições ou leituras incorretas, podendo ser atribuídas a erros na técnica de calibração primária, transcrição de dados ou variações nas condições experimentais.^{43,54}

Identificar e tratar amostras anómalas é uma etapa essencial para otimizar os conjuntos de dados utilizados em calibração e validação. A exclusão cuidadosa de outliers do processo de modelagem permite a construção de modelos mais robustos e precisos, garantindo assim uma previsão confiável em situações futuras. Assim, é imperativo garantir que o conjunto de amostras utilizado para a calibração seja homogêneo, eliminando aquelas que se destacam significativamente das restantes amostras. Contudo, nem sempre é obrigatório excluir outliers para garantir uma boa capacidade preditiva na calibração. É possível minimizar a influência dos outliers, tornando o modelo menos sensível a eles, atribuindo-lhes um peso menor em relação às outras amostras. Esta abordagem possibilita uma calibração robusta e fiável, mantendo a integridade dos dados e melhorando a precisão das previsões.⁴³

A detecção de outliers pode ser conduzida através de várias abordagens. Três métodos comuns incluem a identificação de amostras com resíduos espectrais elevados (bloco X), elevada influência no modelo (*leverage*) e resíduos consideráveis nos valores das propriedades ou concentrações de compostos (bloco Y). No primeiro método, a detecção de valores anómalos é feita com base em resíduos não modelados de dados espectrais, comparando o desvio padrão dos resíduos totais ($s(e)$) (equação 40) com o desvio padrão de uma amostra específica ($s(e_i)$) (equação 41). Se $s(e_i)$ for significativamente maior do que $s(e)$, essa amostra é considerada um *outlier*.

$$s(e)^2 = \frac{1}{I_c J - J - A_{max}(I_c, J)} \sum_{i=1}^{I_c} \left(\sum_{j=1}^J (x_{ij} - \hat{x}_{ij})^2 \right) \quad (40)$$

$$s(e_i)^2 = \frac{I_c}{I_c J - J - A_{max}(I_c, J)} \sum_{j=1}^J (x_{ij} - \hat{x}_{ij})^2 \quad (41)$$

Em ambas as equações, J representa o número de variáveis espectrais, x_{ij} é o valor da absorvância da amostra i no comprimento de onda j e $\hat{x}_{ij}$ é o valor estimado com A variáveis latentes. O método de leverage analisa o afastamento das amostras em relação ao centro dos dados; se o valor GH de uma amostra ultrapassar um limiar específico, essa amostra deve ser excluída do conjunto de calibração. Por fim, os resíduos não modelados nas variáveis dependentes são fundamentais para avaliar a precisão de modelos estatísticos. A identificação de valores anómalos ocorre ao comparar a raiz do erro quadrático médio de calibração (RMSEC) com o erro absoluto da amostra. Se a discrepância entre o valor de referência (y_i) e a sua estimativa ($\hat{y}_i$) ultrapassar uma constante que varia de duas a três vezes o RMSEC, a amostra é identificada como um outlier.⁵⁴

7. Materiais e Métodos

7.1 Preparação das amostras

As amostras usadas neste trabalho, foram rececionadas no laboratório das instalações da ALIP, onde foram devidamente identificadas através de um ID interno garantindo as condições de confidencialidade. A recolha das amostras não é da responsabilidade do laboratório, ficando apenas ao seu encargo a amostragem das amostras rececionadas e não da sua recolha na fonte. Para a realização deste trabalho experimental utilizaram-se amostras de TMR, comumente designadas por mistura Unifeed ($n = 255$) que foram rececionadas à temperatura ambiente. Após a receção e identificação destas amostras, foram homogeneizadas manualmente (Figura 12 ponto 1), procedendo-se à amostragem seguindo o procedimento da NP ISO 6498 (2018), conforme as diretrizes internas estabelecidas. Do remanescente de cada amostra foram retiradas cerca de 400g que sofreram um processo de trituração em fresco, para o qual recorreu-se ao equipamento POWTEQ Knife Mill (Type HM100), durante 5 segundos a 5000 rotações por minutos (rpm) (Figura 12 ponto 2 e 3). Os parâmetros de trituração foram definidos tendo por base uma série de ensaios previamente realizados, os quais envolveram variações no tempo de trituração (5, 10, 15 e 20 segundos) e no número de rotações (de 3500 rpm a 10000 rpm). Na figura 11, são comparados os diferentes números de rotações para 5 segundos, tendo sido seleccionada a condição ótima 50 x 100 rpm. Esta seleção teve em consideração a necessidade de minimizar a perda de humidade causada pelo aquecimento associado ao atrito do processo, bem como a obtenção do tamanho de partícula desejado, desta forma foi estabelecido um equilíbrio entre a redução da perda de humidade e a obtenção da granulometria ótima para facilitar uma boa compactação do material.

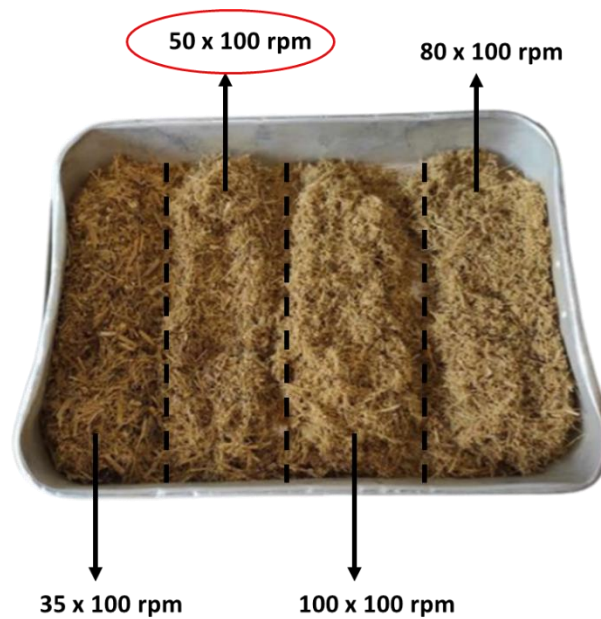


Figura 11: Ensaio com diferentes número de rotações testados para 5 segundos.

Após o processo de trituração, as amostras foram acondicionadas em sacos transparentes de plástico com fecho de correr, para minimizar as possíveis perdas de humidade devido às trocas com o ambiente atmosférico. Este procedimento foi temporariamente adotado para preservar as características das amostras durante um curto período, para que a temperatura estabilizasse de acordo com a temperatura do local onde se encontra o espectrofotómetro de refletância NIR. Posteriormente, as amostras trituradas foram submetidas a dois procedimentos distintos: uma leitura espectrofotométrica (Figura 12, ponto 4) , para obter o espectro em NIR em fresco e análises químicas de referência, para a qual foi necessário passar por um processo de secagem e moenda.



Figura 12: Esquema de Preparação da amostra em Fresco.

7.2 Espectrofotômetro de refletância NIR

Os espectros NIR de refletância difusas recolhidos das amostras recém-trituradas, foram obtidos num espectrofotômetro FOSS NIRS TM DS2500, conforme representado na Figura 13. Este equipamento é constituído por um monocromador que incide luz sobre as amostras numa gama espectral que se estende dos 400 nm aos 2500 nm. o seu sistema está equipado com dois detetores: um de silício sensível na faixa de 400 nm a 1100 nm e outro de sulfureto de chumbo sensível na faixa de 1100 nm a 2500 nm. É relevante salientar que este equipamento apresenta uma notável resolução espectral de 0,5 nm e integra circuitos de última geração concebidos para melhorar a relação sinal-ruído e minimizar interferências indesejadas que possam prejudicar o seu desempenho.



Figura 13: Equipamento FOSS NIRS TM DS2500.

Após o processo de preparação das amostras, estas foram inseridas e compactadas no recipiente de grande volume do equipamento, e a medição da energia refletida das amostras foi realizada à temperatura ambiente (Figura 12 ponto 4). O software ISIScan Nova, incorporado no FOSS NIRS TM DS2500, viabilizou a visualização e o armazenamento dos espectros NIR correspondentes a cada amostra. Os valores de refletância (R) são convertidos em valores de absorvância (A) espectral, sendo registados como $A = \log \frac{1}{R}$ com réplica.

Em paralelo foi criado o ficheiro de repetibilidade, em que foram efetuadas medições espectrais de amostras frescas, as quais foram submetidas a diferentes condições de preparação, com o intuito de simular possíveis condições/alterações que podem ocorrer durante a preparação das amostras e minimizar o seu efeito na calibração. Assim, as vertentes utilizadas foram, nomeadamente: a trituração, na qual promoveu-se leituras espectrais de amostras sem trituração e com 5 s e 10 s de trituração a 5000 rpm, testando diferentes granulometrias possíveis de ocorrer, a temperatura, em que algumas amostras foram submetidas a temperaturas mais baixas que variaram entre os 8 °C e os 11 °C, a temperaturas ambiente de 19°C a 21 °C, e a temperaturas mais elevadas entre os 25 °C e os 30 °C; a compactação, utilizando diferentes forças de

compactação ao introduzir as amostras no copo para realização da leitura, desde a mais leve, normal, forte e dois operadores diferentes. Todas as variáveis mencionadas exercem influência nos espectros obtidos, sendo desta forma possível mitigar os impactos decorrentes das variações nas condições de preparação das amostras.

7.3 Análise química de referência

Para determinar o teor de matéria seca (MS), as amostras foram sujeitas a um processo de secagem, segundo as diretrizes do procedimento interno LIPP071, numa estufa com circulação de ar forçado Memmert UF750, a uma temperatura de 80 °C, por cerca de 18 horas, sendo determinada gravimetricamente. Posteriormente, para as análises mencionadas a seguir, as amostras secas foram reduzidas a pó num moinho Christy Turner (modelo 8" Laboratory Mill) equipado com um crivo de 1 mm.

Os restantes parâmetros foram determinados tendo em conta a percentagem de MS a 103°C ± 2°C, seguindo as diretrizes internas que constam na LIPP085.

O teor total de cinzas totais foi determinado de acordo com os procedimentos internos, seguindo a LIPP072 baseada na norma NP ISO 5984 (2014). Este método envolve a incineração da amostra numa mufla a uma temperatura de 550°C ± 20°C, ao longo de aproximadamente 3 horas, a cinza total é determinada gravimetricamente como o resíduo remanescente após a incineração.

A determinação do teor de proteína bruta (PB) é calculada com base no azoto Kjeldahl multiplicado por 6,25, seguindo um procedimento interno LIPP073 fundamentado na norma NP ISO 5983 (2007). Este método compreende três fases distintas: a digestão, a destilação e a titulação. Inicialmente, a amostra é digerida num digestor FOSS Tecator Digestor com ácido sulfúrico concentrado e catalisadores, convertendo o azoto proteico em sulfato de amónio. Posteriormente, ocorre a adição de hidróxido de sódio para a libertação de amónia, a fase de destilação ocorre num sistema de arrastamento de vapor, utilizando um destilador Kjeltex TM 8200, com a amónia a ser recolhida numa solução recetora de ácido bórico. Por fim, realiza-se a titulação colorimétrica com ácido sulfúrico e desta forma determina-se o teor de proteína bruta utilizando o fator de conversão.

A gordura bruta é extraída através do uso de éter de petróleo com hidrólise que ocorre no equipamento hydrotec FOSS Tecator; a amostra é submetida a hidrólise com ácido clorídrico sob condições de aquecimento, sendo posteriormente arrefecida, filtrada e o resíduo resultante é lavado e seco. A gordura é, então, extraída do resíduo hidrolisado utilizando éter de petróleo no equipamento FOSS Extraction Unit, o solvente é

destilado e o extrato seco é pesado para a determinação quantitativa da gordura presente na amostra, seguindo as indicações que constam na LIPP074.

Para determinação do teor de fibra bruta segundo o procedimento interno LIPP127, a amostra é primeiramente digerida com ácido sulfúrico sob aquecimento, posteriormente lavada e filtrada. Em seguida, é submetida a uma segunda digestão com hidróxido de potássio sob calor, seguida de nova lavagem e filtração, ambas as etapas ocorrem no equipamento Gerhardt da Fribretherm. O resíduo seco remanescente é então seco, posteriormente pesado e incinerado. A perda de peso resultante da incineração do resíduo seco corresponde à massa de fibra bruta presente na amostra.

7.4 Desenvolvimento e validação da calibração

Após a obtenção dos espectros das amostras recentemente trituradas e das análises químicas de referência correspondentes, foram aplicados tratamentos matemáticos quimiométricos. Utilizou-se o software WinISI (versão 4.6.0; FOSS) para desenvolver e validar modelos de calibração que relacionam os dados espectrais do NIR com os valores de referência dos parâmetros analisados. Inicialmente, procedeu-se à caracterização dos dados de calibração através da análise dos parâmetros estatísticos e da visualização gráfica do ficheiro de calibração, que contém os espectros e os valores de referência das amostras. Procedeu-se a PCA com o propósito de avaliar a homogeneidade das amostras de TMR. Este cálculo foi realizado com base nos *scores* para determinar a qualidade das amostras, pois os resultados indicam o quão distintos são os espectros em relação à média do conjunto. Com base na distribuição estatística dos *scores*, as amostras foram classificadas como mais ou menos semelhantes ao espectro médio. Além disso, esta análise estatística permitiu a identificação de eventuais *outliers* presentes nos conjuntos de amostras utilizados. Subsequentemente, as amostras foram divididas em dois grupos: um para o desenvolvimento da calibração e validação cruzada ($n = 218$) e outro para a validação externa ($n = 37$). Esta divisão foi realizada pelo software WinISI, em que foram selecionadas as amostras de 5 em 5, a começar na terceira. No processo de calibração para cada parâmetro, aplicou-se a quantificação por MPLS, em conjunto com o arquivo de repetibilidade, de forma a obter as equações correspondentes da análise multivariada. Durante a construção dos modelos de calibração, procedeu-se ao pré-processamento dos espectros, incluindo SNV-DT, além da aplicação de tratamentos matemáticos específicos (1,4,4,1; 2,4,4,1; 3,5,5,1). Os tratamentos aplicados tiveram como objetivo corrigir a dispersão, eliminando os efeitos aditivos e

multiplicativos da linha de base, reduzindo as interferências de características físicas, como o tamanho das partículas e o comprimento do caminho das amostras para os espectros. No que concerne à derivatização, adotou-se uma nomenclatura com quatro dígitos (por exemplo, 1,4,4,1), onde o primeiro dígito representa o número da derivada, o segundo dígito refere-se ao intervalo no qual a derivada é calculada (gap), o terceiro dígito indica o número de pontos de dados numa média de execução ou suavização, e o quarto dígito representa a segunda suavização. Durante o processo de calibração, os intervalos de comprimento de onda utilizados foram os seguintes: 402 nm - 1098,2 nm; e 1102 nm - 2498,2 nm ou 408 nm a 1092,8 nm 1108 nm – 2492,8 nm. Os dois últimos intervalos foram escolhidos para excluir a luz visível, concentrando-se assim nas regiões específicas do espectro de interesse para a análise em questão

A distância H globais (GH) e de vizinhança (NH) podem ser utilizadas para criar conjuntos de dados, detetar valores atípicos e identificar grupos de amostras não abrangidos por uma calibração⁵² No software WinISI , o NH mede a proximidade entre uma amostra e a amostra mais próxima na base de dados, ou seja, um valor elevado de NH indica que a amostra está distante da amostra mais próxima na base de dados de calibração, portanto é diferente das amostras usadas para estabelecer o modelo. Por sua vez, o GH é uma medida que avalia a distância entre o espectro de uma amostra e o centro da base de dados, considera-se que um valor elevado de GH indica que a amostra está distante do grupo de amostras usado para calibração, ou seja, é diferente das amostras usadas para estabelecer o modelo. Este valor de GH é determinado por⁵⁵:

$$GH = \frac{\left(\frac{n}{n-1}\right) D^2}{k} \quad (42)$$

Em que k corresponde aos números de PCs do modelo e D^2 é o quadrado da distância de Mahalanobis. Assim, caso o valor de GH ultrapasse um limite predefinido, normalmente definido como 3, suspeita-se que a amostra está fora da gama de trabalho da calibração, sinalizando uma possível discrepância ou singularidade no espectro da amostra.⁵⁵

Os possíveis valores discrepantes foram identificados e removidos, e a calibração foi refeita.

Foi utilizada uma regressão MPLS para correlacionar os valores de referência com os dados espectrais do conjunto de calibração. Para avaliar os modelos de calibração desenvolvidos, foi adotada a técnica de validação cruzada, a fim de evitar o *overfitting*

dos dados. Neste tipo de validação, o conjunto de amostras da calibração foi dividido de forma a incluir apenas uma amostra (*Leave-One-Out*, LOO) ou um grupo de amostras (por exemplo, Group12345) para a verificação dos resultados, enquanto as demais compuseram o modelo de calibração. A estatística resultante da validação cruzada proporciona uma estimativa mais precisa do desempenho das calibrações em comparação com a estatística obtida apenas da própria calibração.

O critério utilizado para selecionar os melhores modelos de calibração para MS, CT, PB, GB e FB baseou-se nos menores valores de erro padrão da calibração (SEC) e do erro padrão da validação cruzada (SECV), juntamente com o maior coeficiente de determinação (RSQ) e o maior valor de $1-VR$ (uma estimativa da variância explicada). O coeficiente de determinação (RSQ) é crucial para explicar a relação entre os valores previstos por NIRS e os valores provenientes das análises químicas.

Finalmente, os modelos de calibração desenvolvidos foram validados ao utilizar um conjunto independente de amostras para a validação externa. Neste processo, os valores previstos pelos modelos foram comparados com os resultados obtidos através de análises laboratoriais de referência. Para avaliar a capacidade preditiva dos modelos de calibração escolhidos, foram aplicados critérios como o menor SEP, o menor BIAS (diferença média entre os valores NIR previstos e os valores de referência), a inclinação próxima de 1 e o Valor de Previsão Relativa (RPD).

O parâmetro RPD é obtido pela divisão do desvio padrão dos valores de referência pelo SEP do modelo desenvolvido, sendo usado para analisar a capacidade do modelo em prever dados futuros com precisão. Porém, quanto menor for o valor de SEP mais confiável é o modelo de calibração desenvolvido. No decorrer da fase de validação, foram detetados e removidos eventuais valores atípicos presentes no conjunto de amostras destinado à validação externa, através de métodos semelhantes aos utilizados na identificação e eliminação de outliers durante o desenvolvimento da calibração.

8. Resultados e Discussão

8.1 Análise química de referência

De forma a assegurar a precisão dos resultados, é fundamental que o conjunto de amostras utilizado no desenvolvimento da calibração seja altamente representativo das amostras que serão analisadas no futuro. Nesse contexto, foram escolhidas amostras que englobam uma vasta diversidade e uma ampla gama de constituintes para este estudo. A variação entre as amostras reflete as diferenças nas localidades de origem e nos períodos de colheita, bem como as diferenças na sua constituição, uma vez que a TMR não tem uma constituição definitiva nem restrita.

Na tabela 3 encontra-se detalhada a composição química das 255 amostras de TMR utilizadas no desenvolvimento da calibração NIRS em fresco.

Tabela 3: Resumo estatístico dos parâmetros analisados nas amostras de TMR, expondo o mínimo, o máximo, a média e o desvio padrão de cada componente analisado.

Constituintes	MS*	CT**	PB**	GB**	FB*
Mínimo	27,38	4,33	6,76	2,39	10,81
Máximo	63,19	17,00	20,96	5,29	44,10
Média	45,54	6,36	15,13	3,58	16,43
Desvio Padrão	5,46	1,24	1,96	0,54	3,62
Nº de Amostras	255	255	255	255	255

*Resultado apresentado em %. **Resultados apresentados em % de MS.

Na tabela em questão, encontram-se discriminados os intervalos de valores relativos aos parâmetros de Matéria Seca (MS), Cinza Total (CT), Proteína Bruta (PB), Gordura Bruta (GB) e Fibra Bruta (FB). Adicionalmente, são apresentadas as médias da composição para cada parâmetro, bem como os respetivos desvios padrão. A partir dos dados de referência obtidos, é possível verificar que foram incluídas diversas variedades de amostras de TMR na calibração, uma vez que ocorre uma variação notória nas concentrações nutricionais encontradas no conjunto de amostras.

Os parâmetros apresentam um desvio padrão baixo tendo em conta a sua respetiva média, o que indica que os valores deste conjunto de dados estão próximos da média aritmética. Por esta mesma razão, pode-se concluir que há uma maior consistência e previsibilidade nos dados, já que os valores estão menos propensos a afastar-se significativamente da média.

Comparando alguns valores obtidos da literatura para TMR tais como PB 17.36 % , MS 51.65%,GB 4.99% e FB 16.8%, com os valores descritos na tabela é possível confirmar que os mesmos encontram-se abrangidos pela gama de trabalho usada para a calibração em causa.⁵⁶

A partir software TIBCO Statistica™ (Version Trial). determinou-se um conjunto de diagramas *box plot* correspondente a cada parâmetro analisado nas amostras TMR. Nos diagramas é possível observar as variações existentes em cada componente numa perspetiva diferente, identificando assim o máximo, o mínimo, o 1ºquartil, a mediana, a média e o 3º quartil, bem como a sinalização de alguns outliers.

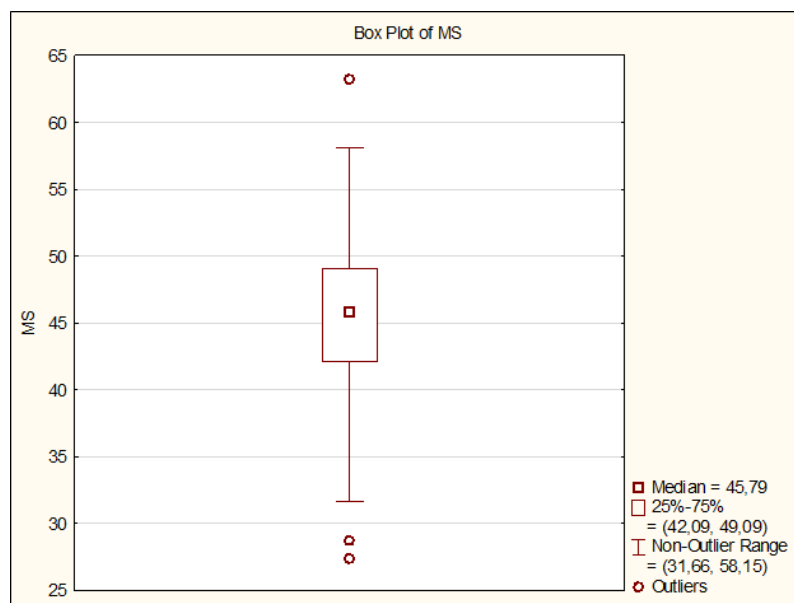


Figura 14: Blox Plot de MS.

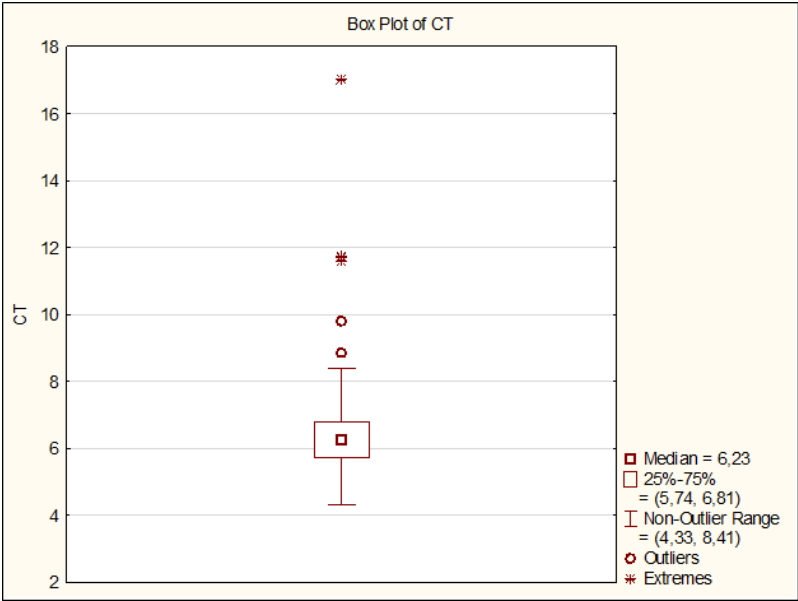


Figura 15: Blox Plot CT.

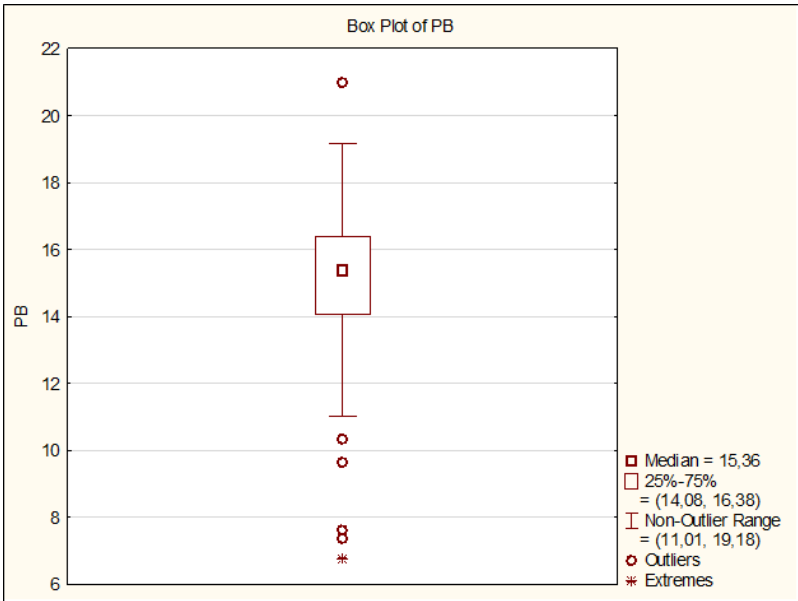


Figura 16: Blox Plot PB.

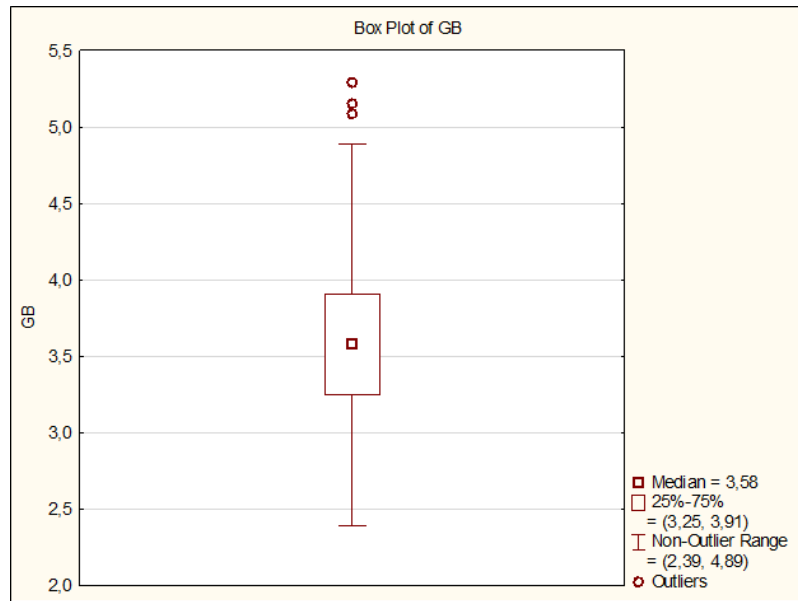


Figura 17: Blox Plot GB.

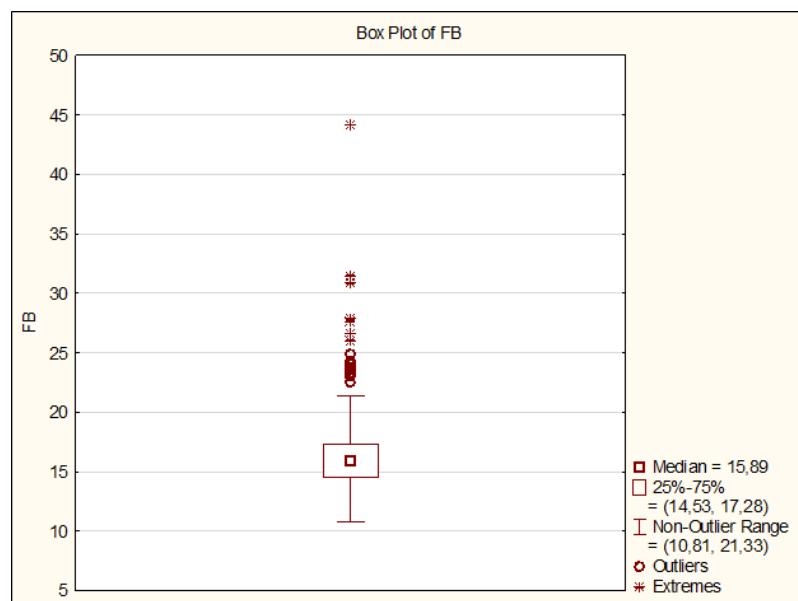


Figura 18: Blox Plot FB.

O *box plot* é uma representação gráfica estatística que oferece informações cruciais sobre a distribuição de um conjunto de dados. A caixa no meio do gráfico representa o intervalo interquartil, que é a diferença entre o primeiro quartil (Q1) e o terceiro quartil (Q3), é onde a maior parte dos dados se encontra. O retângulo menor dentro da caixa

representa a mediana (Q2), os bigodes do *box plot* estendem-se até 1,5 vezes o intervalo interquartil a partir das extremidades da caixa. Qualquer valor fora dessa faixa é considerado um valor atípico (outlier) e é representado como um ponto individual. Estes pontos são indicadores de dados que estão significativamente distantes dos quartis, sugerindo variações ou discrepâncias notáveis no conjunto de dados. O *box Plot* do parâmetro MS é o que apresenta maior dispersão dos resultados sendo que possui o intervalo interquartil maior, e em contrapartida a GB apresenta menor dispersão dos resultados uma vez que apresenta menor intervalo interquartil.

Nos diagramas, é possível identificar pontos que se alinham com a caixa e os bigodes, sugerindo que esses pontos podem ser considerados outliers, no diagrama da MS (figura 14) e no de GB (figura 17) são identificados 3 outliers, no da CT (figura 15), apenas 2, a PB (figura 16) apresenta cerca de 5 e a FB (figura 18) apresenta mais de 5 outliers, sendo o parâmetro que apresenta maior número de outliers.

Com o auxílio do software TIBCO Statistica™ analisou-se as correlações entre os parâmetros analisados por análise química, por meio de um gráfico de matriz, representado na figura 19.

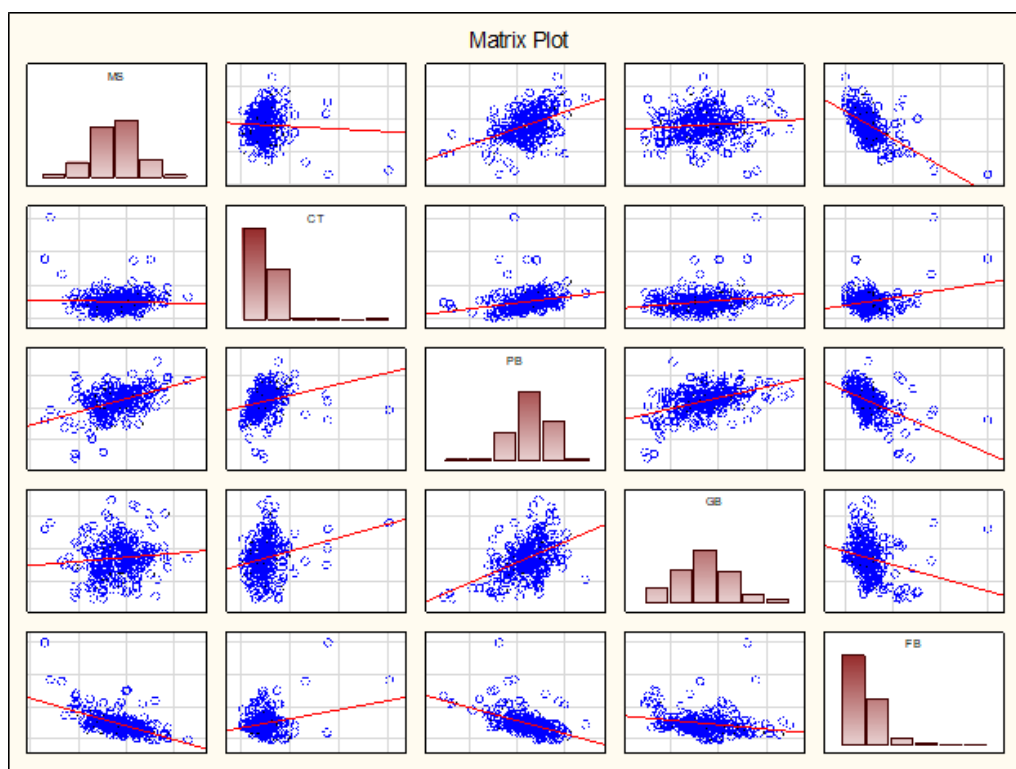


Figura 19: Representação gráfica da matriz das correlações existentes entre os parâmetros analisados nas amostras TMR, apresentado histogramas em relação ao parâmetro e a linha de tendência a vermelho para as relações entre dois parâmetros diferente.

Devido à composição variada característica da TMR, os parâmetros não apresentam uma boa correlação, pelo que apenas é possível observar alguma correlação entre o parâmetro de matéria Seca e a proteína Bruta, porém não é o suficiente para retirar conclusões plausíveis.

No gráfico apresentada na figura 20, construído a partir do *software* Winisi é possível observar que não há um bom ajuste dos dados (MS vs PB) ao modelo de regressão, pelo que o coeficiente de determinação é significativamente menor que 1.

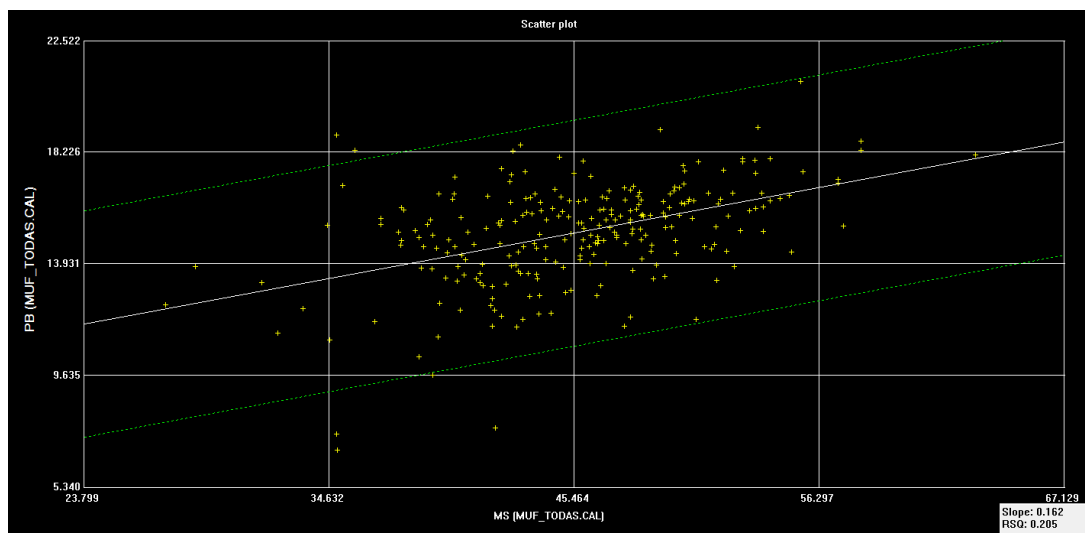


Figura 20: *Scatter plot* Constituinte MS vs Constituinte PB

No gráfico de matriz (figura 19), também é possível a identificação de possíveis valores atípicos devido às distâncias anómalas observadas em certos pontos ao correlacionar diferentes parâmetros, tal pode dever-se a possíveis erros durante as análises químicas ou à existência de amostras que se destacam das restantes utilizadas.

8.2 Espectros NIRS

Durante o processo de desenvolvimento da calibração MPLS, foram implementados tratamentos quimiométricos para eliminar informações irrelevantes, resultando numa redução das variáveis, deixando apenas a informação crucial para a calibração.

A Figura 21 apresenta os espectros brutos das amostras TMR sem pré-tratamento matemático e como tal contendo tanto informações relevantes quanto irrelevantes.

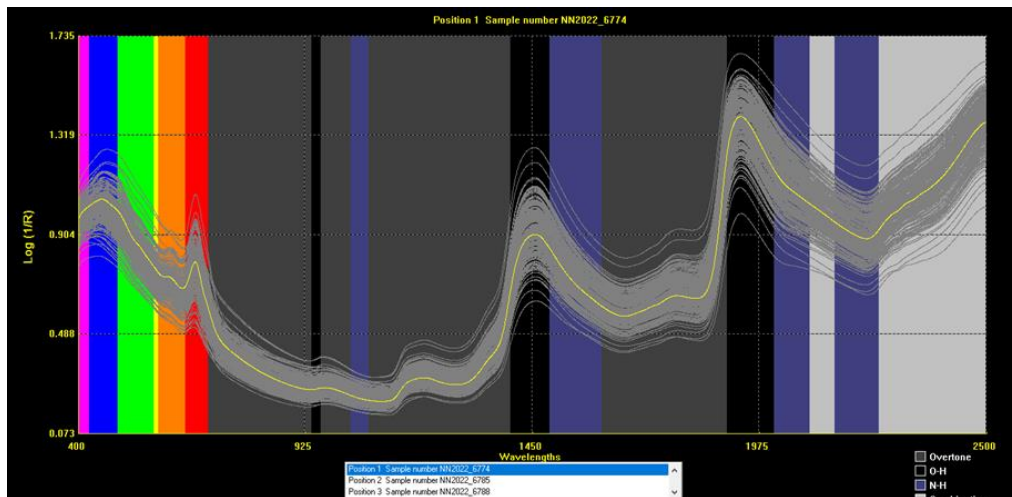


Figura 21: Espectro Bruto NIR de todas as amostras.

Na Figura 22 , é evidente o efeito da aplicação dos tratamentos SNV-DT, ocorrendo uma redução de variações indesejadas nos espectros e um realce das características específicas relacionadas com os constituintes das amostras.

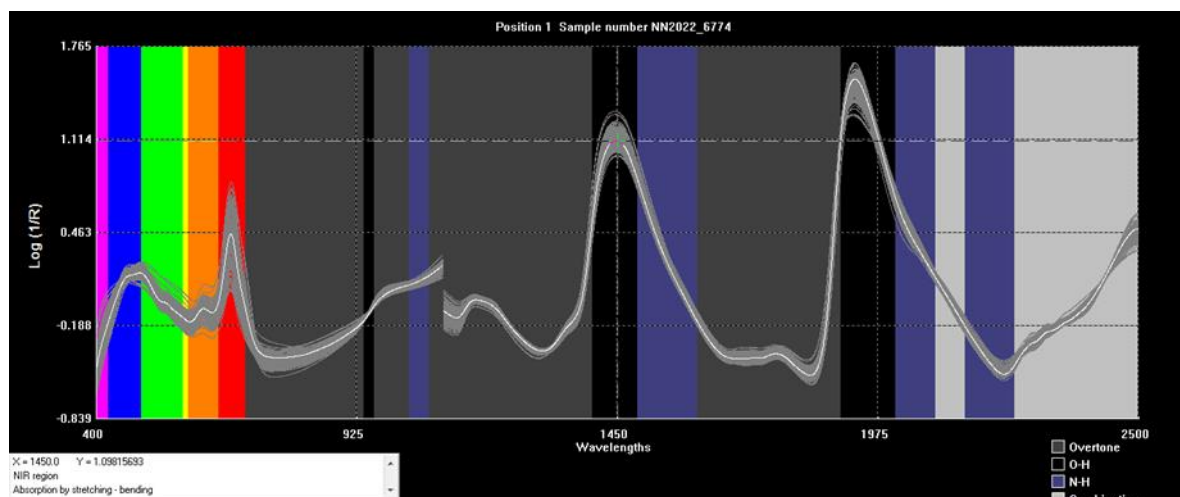


Figura 22: Espectro NIR com SNV+DT

Quando se aplica apenas o tratamento matemático da primeira derivada na condição (1,4,4,1), observam-se picos adicionais que são essenciais para distinguir bandas absorventes que podem estar sobrepostas no espectro bruto (figura 23). Esta distinção é crucial devido à presença de várias substâncias nas amostras, cada uma contribuindo para diferentes bandas de absorção. A separação nítida e identificação dessas contribuições individuais facilita significativamente a análise química e a interpretação dos resultados espectroscópicos, permitindo uma compreensão mais precisa e detalhada das amostras em estudo.

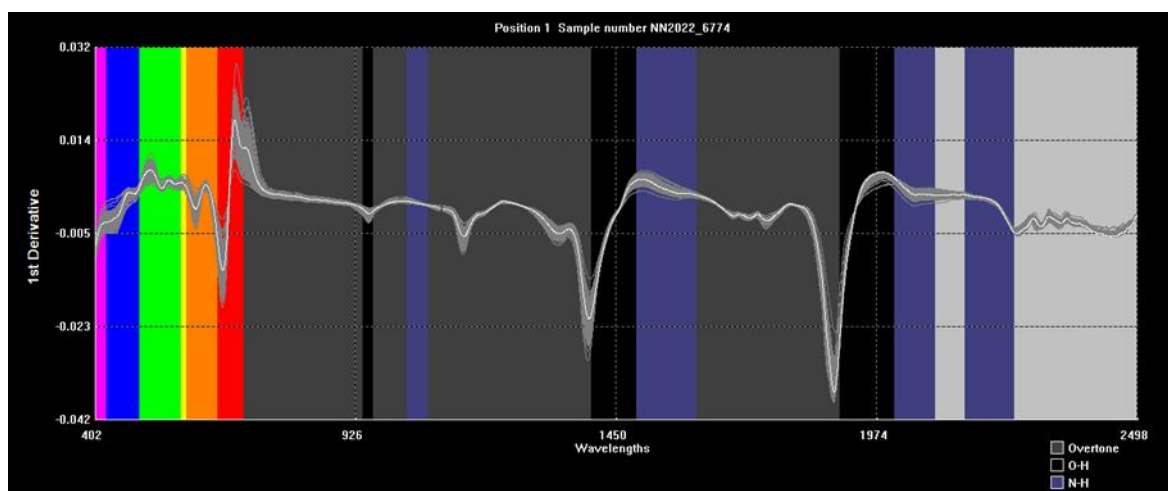


Figura 23: Espectro NIR com apenas 1º derivada (1,4,4,1).

Se ao espectro bruto sem tratamento matemático, for aplicada apenas a 3ª derivada, a ampliação das variações resultantes da 3ª derivada resulta numa maior ampliação do ruído, obtendo-se o espectro ruidoso (figura 24), permitindo perceber que para a calibração em questão a terceira derivada não apresenta vantagens.

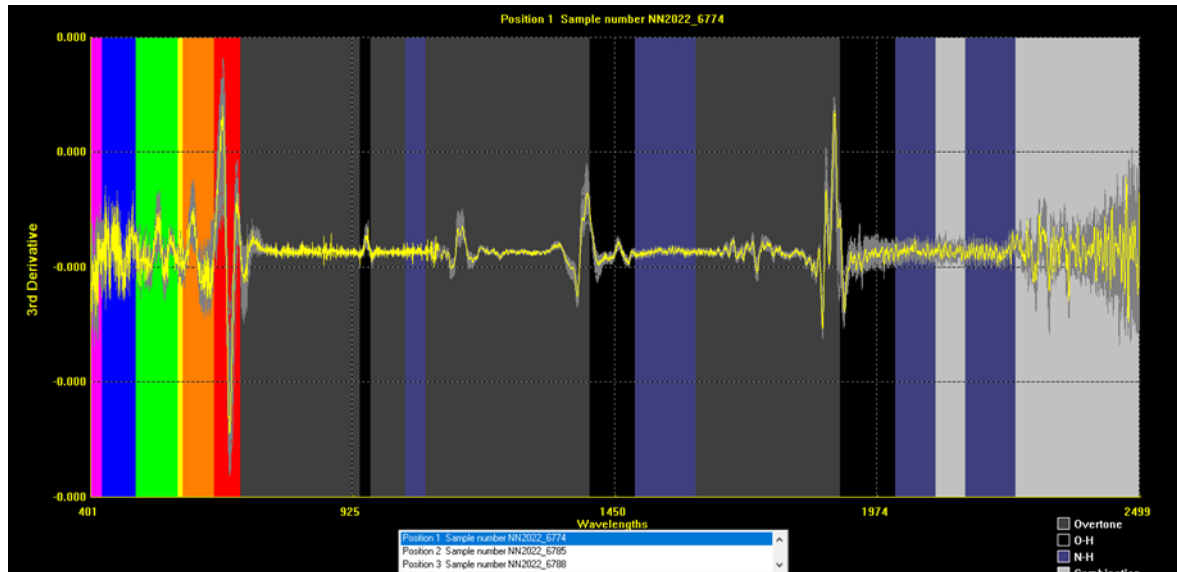


Figura 24: Espectro Ruidoso 3º derivada (3,1,1,1).

Nas figuras subsequentes, são exibidos os espectros após a aplicação dos tratamentos utilizados no desenvolvimento da calibração, nomeadamente, a aplicação do SNV-DT juntamente com a primeira derivada (Figura 25) e a segunda derivada (Figura 26), a qual também elimina inclinações na linha de base, mas amplifica o ruído espectral.

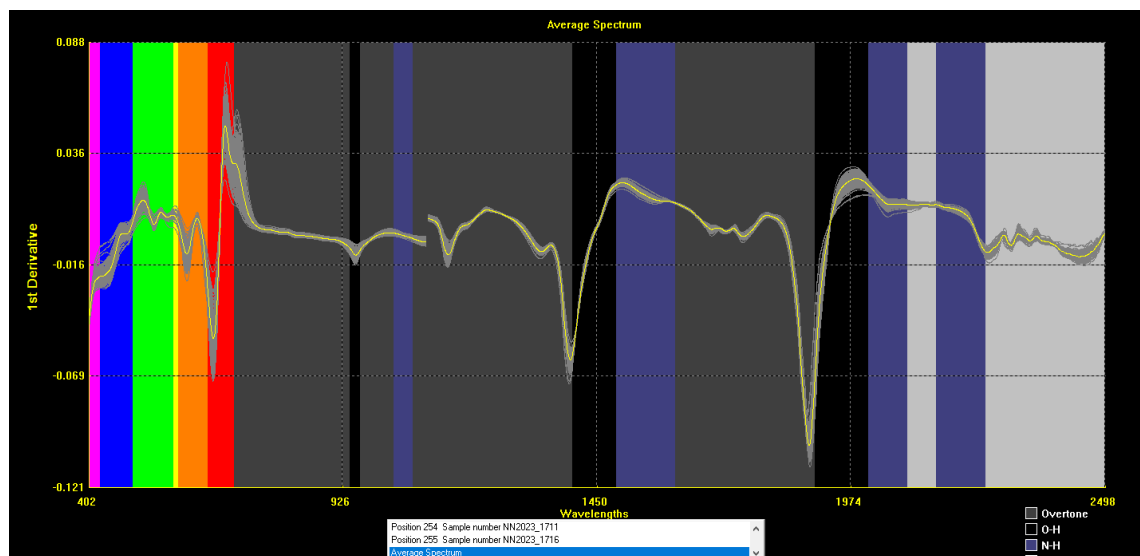


Figura 25: Espectro com 1,4,4,1 SNV+DT

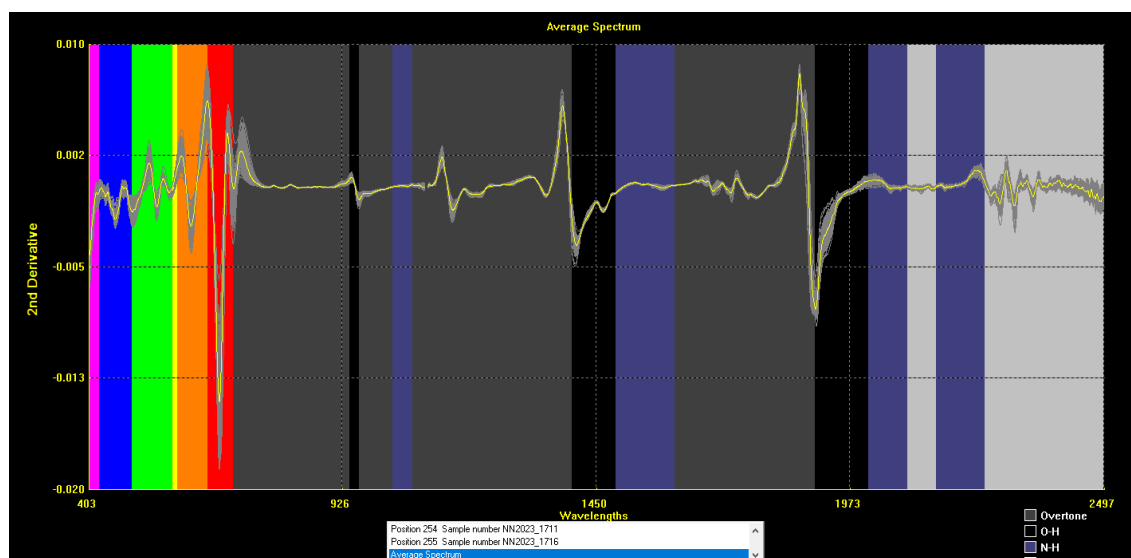


Figura 26: Espectro com 2,4,4,1 SNV+DT

Na análise dos espectros brutos de NIR, foram detetados picos com menor absorção nas regiões de comprimento de onda de 1450 nm e 1930 nm. Estes picos indicam uma menor absorção devido às bandas O-H, o que sugere a perda de humidade nas amostras analisadas. Adicionalmente, o software utilizado possibilita a identificação de zonas características nos espectros, associadas a *overtones*, O-H, N-H e combinações. Estas análises minuciosas são essenciais para compreender as propriedades das amostras e interpretar os resultados obtidos através da espectroscopia NIR.

8.3 Estatística do desenvolvimento e validação da calibração

Na análise de Componentes Principais (PCA) realizada, foram investigados diferentes tratamentos em segmentos específicos de comprimento de onda, nomeadamente na região visível entre 400 e 1100 nm e na região do infravermelho próximo entre 1100 e 2500 nm.

A aplicação de diferentes tratamentos espectrais foi meticulosamente considerada para identificar a combinação que proporcionasse os melhores indicadores. Através deste processo, foram gerados os loadings e os scores, que são essenciais para a

interpretação dos dados e a identificação de padrões relevantes nos espectros analisados. Este estudo permitiu uma compreensão mais profunda das características espectrais das amostras, destacando informações cruciais para as análises subsequentes.

Na tabela 4 consta as diferentes condições criadas, e para cada PCA o número de fatores e a variância explicada, bem como a distância H média e o número de outliers.

Tabela 4: Análise de Componente Principal condições com diferentes tratamentos matemáticos.

	Condição 1	Condição 2	Condição 3	Condição 4	Condição 5	Condição 6	Condição 7	Condição 8	Condição 9
Tratamento Matemático	1,4,4,1 SNV+DT	1,4,4,1 SNV+DT	1,4,4,1 SNV+DT	1,4,4,1 SNV+DT	1,4,4,1 SNV+DT	2,4,4,1 SNV+DT	2,4,4,1 SNV+DT	2,4,4,1 SNV+DT	2,4,4,1 SNV+DT
Comprimento Onda	402- 1098,2 1102- 2498,2	1102- 2498,2	1102- 2498,2	408- 1092,8 1108- 2492,8	1108- 2492,8	402- 1098,2 1102- 2498,2	1102- 2498,2	408- 1092,8 1108- 2492,8	1108- 2492,8
Nº Fatores	23	16	15	23	14	32	32	32	32
Variância Explicada	99.71	99.64	99.61	99.72	99.62	98.00	97.04	98.70	98.43
Média Distância H	0.8143	0.8097	0.7948	0.8088	0.7947	0.8702	0.9218	0.8884	0.9277
Outliers/GH	4	4	4)	4	6	2	2	2	2

Considerando a maior variância explicada com o menor número de fatores, juntamente com a menor distância média H e o menor número de outliers identificados, a condição que melhor se adequa é a condição 3, estas características indicam que esta condição é mais eficaz em capturar informações importantes dos dados, reduzir a distorção dos padrões e minimizar a influência de valores atípicos, proporcionando uma análise mais robusta e confiável.

Na figura 27 e 28 , histograma da distância H e do *scatter plot* 3D, respetivamente da PCA com 15 fatores e uma variância explicada de 99,61, onde é visualmente possível identificar a existência de *Outliers*.

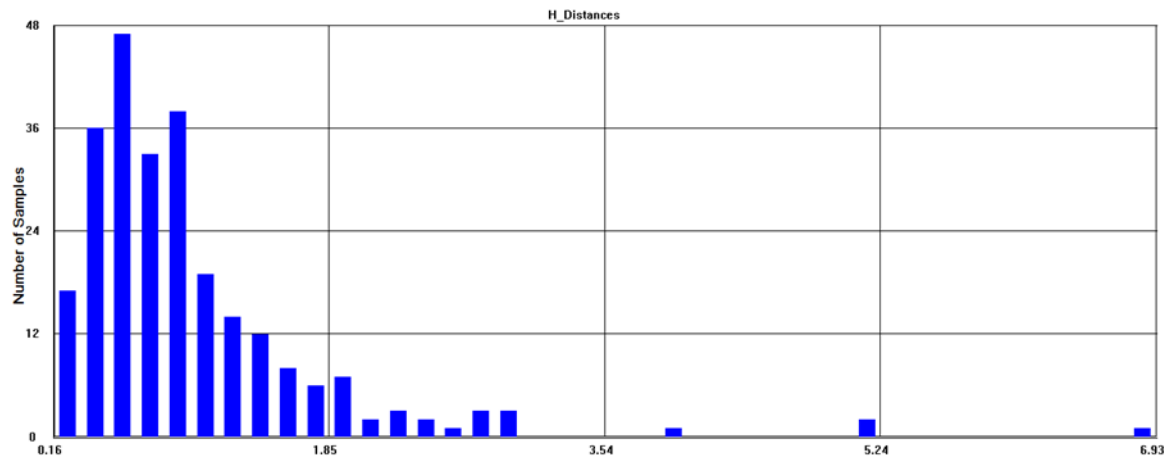


Figura 27: Histograma de distância H, para 15 fatores.

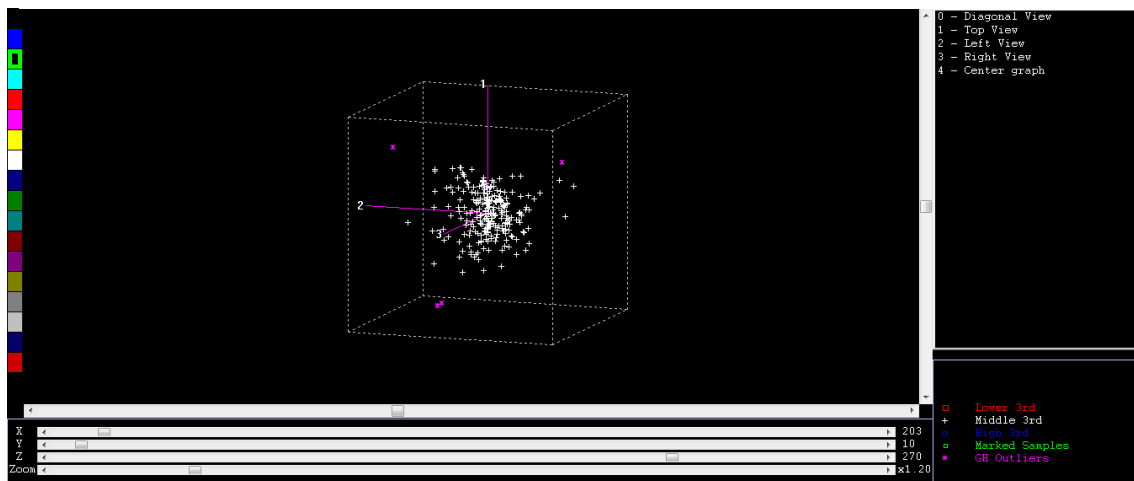


Figura 28: Scatter Plot 3D, PCA com 15 fatores.

Os quatro outliers identificados pelo PCA serão designados neste estudo pelas seguintes identificações: 48C, 63C, 211C e 222V. Através de avaliações visuais realizadas durante a amostragem e preparação, foi possível correlacionar essas identificações com algumas observações específicas. Por exemplo, a amostra 48C apresentava uma elevada quantidade de silagem de erva húmida, além de cascas de tomate, areia e bolores visíveis. Após a trituração, sua aparência tornou-se heterogénea, indicando que a granulometria ideal não foi alcançada. Já a amostra 222V, parte do conjunto de validação, também continha uma quantidade considerável de silagem de erva húmida, assim como cascas de tomate e bolores visíveis. Estas

características únicas tornaram essas amostras diferentes do comportamento observado no restante conjunto de amostras

Ao analisar os dados no software TIBCO StatisticaTM utilizando as mesmas condições de PCA mencionadas anteriormente, para um total de 15 PCs, observou-se a presença de mais do que os 4 outliers (figura 29) identificados pelo WinISI. No entanto, em ambos os casos, esses outliers destacaram-se, confirmando o comportamento distinto destas amostras em relação ao restante do conjunto de dados. Este padrão de comportamento reforça a validade da identificação destes pontos como outliers, indicando a sua relevância na análise e justificando exclusão dos mesmos em avaliações posteriores.

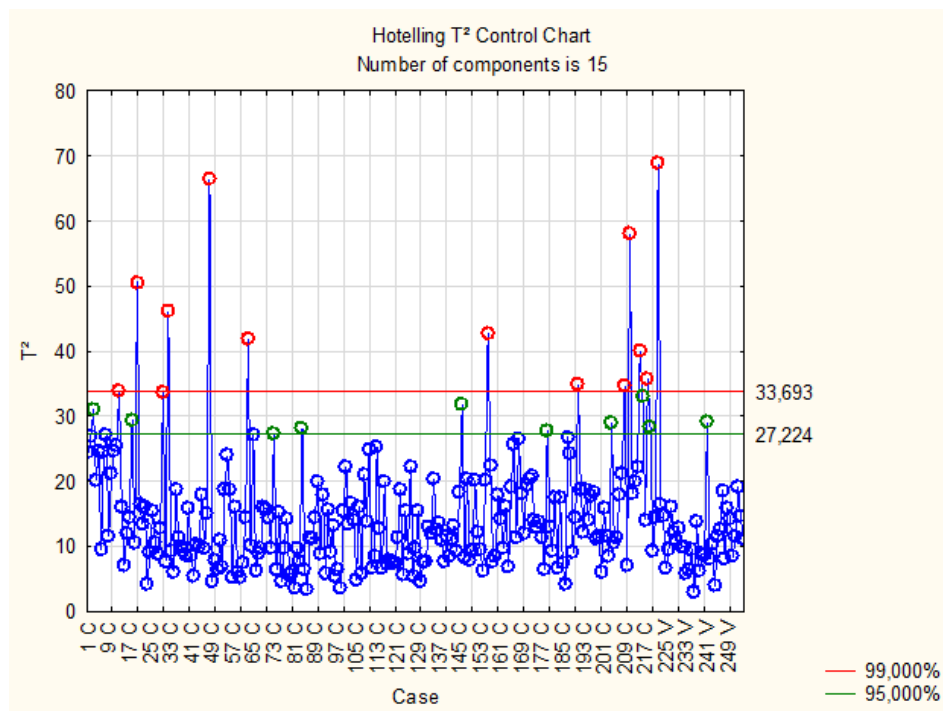


Figura 29: Carta de Controle T de Hotelling, com 15 fatores.

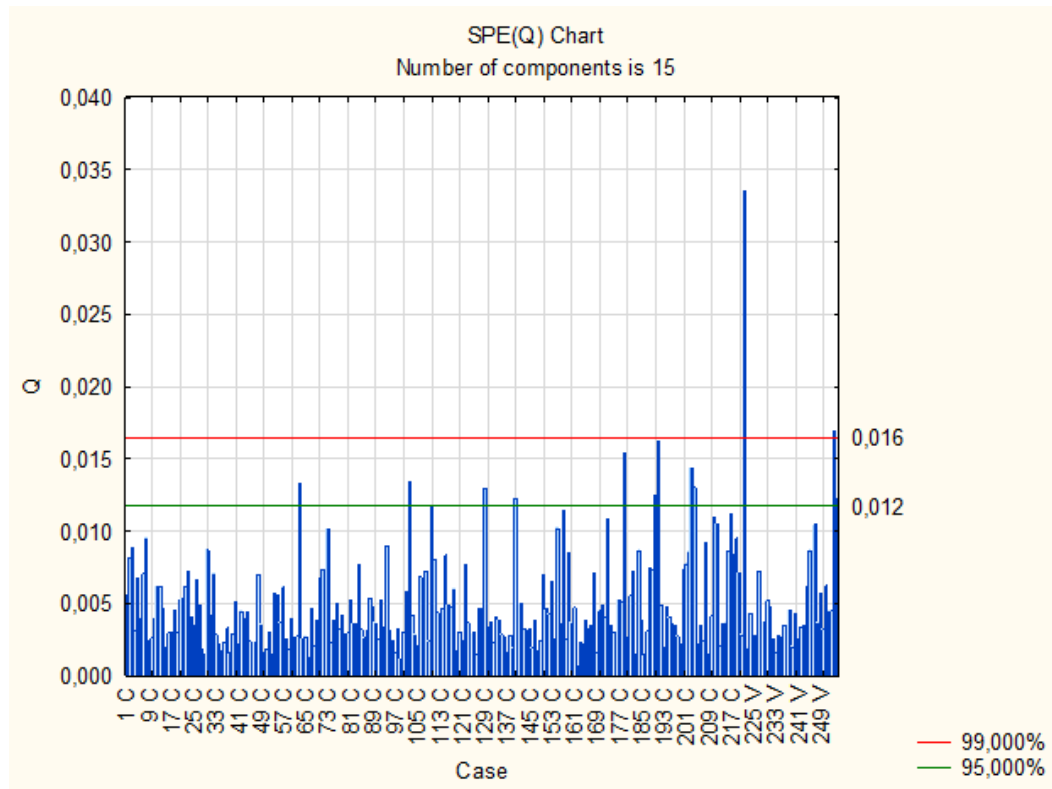


Figura 30: Gráfico construído com valores de Q.

Com os valores obtidos da distribuição T^2 de Hotelling (figura 29) e Q (figura 30), ao correlacionar os dois que correspondem ao GH vs NH, respetivamente no WinISI, é notório um outlier na zona crítica, correspondente à amostra 222V (figura 31). Isso reforça a importância dessa amostra como um ponto distinto e atípico em relação ao restante conjunto de dados. A sua presença na zona crítica destaca a necessidade de tratamento especial durante a análise, e confirma sua exclusão como um *outlier* significativo.

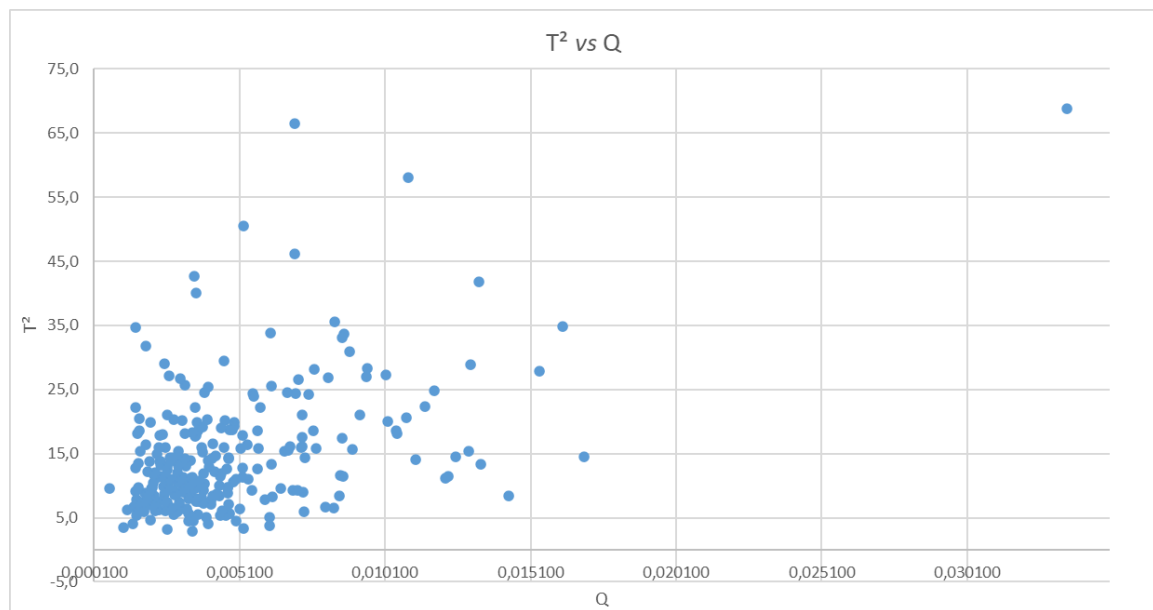


Figura 31: Gráfico distribuição T^2 de Hotelling vs Q

As tabelas 5 e 6 apresentam estatísticas descritivas, incluindo o número de amostras, média, desvio padrão, pré-processamento dos espectros, intervalos de comprimento de onda, tipo de validação cruzada utilizada, SEC, RSQ, 1-VR, SECV e variáveis latentes para os melhores modelos de calibração obtidos. Cada parâmetro analisado exibiu estatísticas de calibração e validação cruzada únicas, conforme esperado. É evidente que os resultados estatísticos foram impactados pelos tratamentos matemáticos aplicados, como a primeira derivada (1,4,4,1; Tabela 5) e a segunda derivada (2,4,4,1; Tabela 6). A terceira derivada não trouxe informações vantajosas para a calibração. De forma geral, as equações de calibração derivadas da primeira ou segunda derivada apresentaram desempenho superior, demonstrando maiores valores de RSQ e 1-VR, bem como menores valores de SEC e SECV para a maioria dos componentes nutricionais analisados.

Tabela 5: Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e primeira derivada (1,4,4,1) para cada parâmetro a analisado nas amostras de TMR em fresco.

Constituinte	λ	Cross Validation	N	Mean	SD	Est. Min	Est. Max	Calibração		Validação Cruzada		VL
								SEC	RSQ	SECV	1-VR	
MS	1108-2492,8	One out	218	45.5534	5.4085	29.3278	61.7790	0.3989	0.9946	0.5114	0.9910	12
CT	1102-2498,2	Grupos 12345	218	6.3322	1.1868	2.7719	9.8926	0.7086	0.6435	0.7797	0.5663	9
PB	408-1092,8 1108-2492,8	Grupos 12345	218	15.1340	1.9260	9.3561	20.9119	0.6865	0.8729	0.7606	0.8433	7
GB	408-1092,8 1108-2492,8	Grupos 12345	217	3.5894	0.5500	1.9394	5.2393	0.1948	0.8746	0.2305	0.8229	8
FB	1102-2498,2	Grupos 12345	216	16.4144	3.1358	7.0069	25.8219	0.9793	0.9025	1.2635	0.8501	10

Tabela 6: Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e primeira derivada (2,4,4,1) para cada parâmetro a analisado nas amostras TMR em fresco.

Constituinte	λ	Cross Validation	N	Mean	SD	Est. Min	Est. Max	Calibração		Validação Cruzada		VL
								SEC	RSQ	SECV	1-VR	
MS	1102-2498,2	One out	217	45.5560	5.4209	29.2933	61.8187	0.3237	0.9964	0.4993	0.9914	8
CT	402-1098,2 1102-2498,2	Grupos 12345	217	6.3076	1.1320	2.9115	9.7036	0.5365	0.7754	0.8081	0.5342	8
PB	1102-2498,2	Grupos 12345	217	15.1560	1.9027	9.4481	20.8640	0.4021	0.9553	0.7471	0.8488	11
GB	1102-2498,2	Grupos 12345	218	3.5880	0.5491	1.9407	5.2353	0.2031	0.8632	0.2280	0.8267	5
FB	1102-2498,2	One out	217	16.4135	3.1286	7.0277	25.7993	0.9198	0.9136	1.2489	0.8536	7

No que diz respeito ao teor de MS, observou-se que os valores de RSQ e 1-VR permaneceram consistentes ao aplicar equações de calibração de primeira e segunda derivada, mantendo-se em 0,99 tanto para RSQ como para 1-VR. No entanto, os resultados estatísticos SEC e do Erro SECV foram menores quando se utilizou o tratamento 2,4,4,1, apresentando valores de 0,32 e 0,49, respetivamente. Assim, conclui-se que a equação da segunda derivada resultou no melhor modelo de calibração para o parâmetro MS nesta fase inicial de avaliação. Contudo, é importante salientar que uma avaliação mais abrangente será realizada por meio da validação externa para validar a eficácia deste modelo. No que concerne ao teor de cinza total, o tratamento com a segunda derivada demonstrou ser o mais adequado em termos estatísticos de calibração. Este tratamento resultou em valores superiores de RSQ e 1-VR, 0,78 e 0,53 respetivamente, bem como valores inferiores de SEC (0,53), porém os valores de SECV são superiores aos da primeira derivada (0,80 > 0,77), a diferença não é considerada significativa. Assim, conclui-se que a equação de segunda derivada proporcionou o melhor modelo de calibração para o parâmetro de cinza total, com base nas métricas avaliadas. Para o parâmetro PB, a segunda derivada é a que se demonstra mais vantajosa, pela mesmo critério de análise, apresentando maior valores de RSQ e 1-VR, mas menores valores para SEC e SECV.

Na análise das estatísticas das equações de calibração desenvolvidas para o teor GB, observa-se uma falta de linearidade. Ou seja, os modelos com maiores valores de 1-VR e menores SEC não demonstram necessariamente maiores valores de RSQ, nem com menores SECV, e vice-versa. No entanto, após uma análise cuidadosa dos resultados obtidos, decidiu-se selecionar o modelo que apresenta os melhores resultados na validação cruzada. Especificamente, a equação de primeira derivada foi escolhida, exibindo valores de 0,19 e 0,23 para SEC e SECV, respetivamente, bem como valores de 0,87 e 0,82 para RSQ e 1-VR, respetivamente. No parâmetro da FB, tendo por base os mesmos critérios das seleções anteriores, optou-se pelo tratamento com a segunda derivada.

O procedimento de validação externa tem como propósito avaliar a capacidade preditiva dos modelos de calibração NIRS desenvolvidos. Os dados estatísticos provenientes da validação externa estão apresentados em detalhe na tabela 7, incluindo o número de amostras, a média e o desvio padrão dos valores laboratoriais, o Erro Padrão da Previsão (SEP), o coeficiente de determinação (RSQ) da validação externa, o bias, a inclinação e o Valor de Previsão Relativa (RPD). Os dados

estatísticos obtidos através da validação externa foram examinados com base nas equações de calibração mais eficazes, desenvolvidas através da regressão MPLS em conjunto com SNV-DT e as derivadas de primeira e segunda ordem de cada parâmetro.

Tabela 7: Estatística de monitorização do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (1,4,4,1) e (2,4,4,1) para cada parâmetro de amostras TMR em fresco.

Const.	Trat. Mat.	N	SEP	BIAS	BIAS LIMITE	SEP (C)	SEP (C) LIMITE	Mín LAB	Máx LAB	SD LAB	SD NIR	Med LAB	Méd NIR	Slope	RSQ	RDP	RER
MS	(1,4,4,1)	37	0,452	0,011	0,307	0,458	0,664	29,328	61,779	5,856	6,038	45,47	45,46	0,9670	0,9950	12,9558	71,7947
MS	(2,4,4,1)	37	0,481	0,034	0,299	0,487	0,649	29,293	61,819	5,856	5,981	45,47	45,44	0,9760	0,9940	12,1746	67,6204
CT	(1,4,4,1)	36	0,697	0,002	0,468	0,707	1,014	2,772	9,893	1,248	1,020	6,36	6,36	1,0080	0,6790	1,7905	10,2162
CT	(1,4,4,1)	37	1,01	0,124	0,468	1,016	1,014	2,772	9,893	1,501	1,013	6,50	6,38	1,0950	0,5460	1,4861	7,0502
CT	(2,4,4,1)	36	0,632	-0,036	0,485	0,64	1,050	2,912	9,704	1,248	1,183	6,36	6,40	0,9100	0,7450	1,9747	10,7470
CT	(2,4,4,1)	37	0,953	0,083	0,485	0,962	1,050	2,912	9,704	1,501	1,174	6,50	6,42	0,9810	0,5890	1,5750	7,1271
PB	(1,4,4,1)	37	0,805	-0,009	0,457	0,816	0,989	9,356	20,912	2,156	1,806	15,08	15,09	1,1110	0,8650	2,6783	14,3550
PB	(2,4,4,1)	36	0,89	0,093	0,448	0,898	0,971	9,448	20,864	2,136	1,852	15,16	15,06	1,0480	0,8250	2,4000	12,8269
GB	(1,4,4,1)	37	0,19	-0,049	0,139	0,186	0,300	1,939	5,239	0,462	0,441	3,54	3,58	0,9600	0,8380	2,6783	14,3550
GB	(2,4,4,1)	37	0,186	-0,051	0,137	0,181	0,296	1,941	5,235	0,462	0,415	3,54	3,59	1,0230	0,8460	2,4000	12,8269
FB	(1,4,4,1)	37	2,869	0,318	0,758	2,891	1,643	7,007	25,822	5,258	3,022	16,10	15,78	1,5560	0,8000	1,8327	6,5580
FB	(1,4,4,1)	36	1,052	-0,125	0,758	1,06	1,643	7,007	25,822	2,327	2,266	15,33	15,45	0,9180	0,7990	2,2120	17,8850
FB	(2,4,4,1)	37	2,933	0,125	0,749	2,971	1,624	7,028	25,799	5,258	2,816	16,10	15,98	1,6860	0,8160	1,7927	6,4001
FB	(2,4,4,1)	36	1,179	-0,326	0,749	1,149	1,624	7,028	25,799	2,327	2,027	15,33	15,65	0,9980	0,7560	1,9737	15,9216

Para interpretar os dados apresentados na tabela 7, é essencial considerar a interpretação dos vários parâmetros significativos. O SEP é um indicador crucial da precisão do modelo, sendo que representa a variabilidade percentual das previsões em relação aos valores reais: quanto menor o seu valor, mais precisa é a previsão. O BIAS, quando próximo de zero, sinaliza que o modelo não tem uma tendência sistemática de prever valores constantemente acima ou abaixo dos valores reais, o que implica que as previsões do modelo estão, em média, muito próximas dos valores reais de referência, sem uma inclinação constante para superestimar ou subestimar. Um BIAS próximo de zero reflete a imparcialidade do modelo, indicando previsões precisas e confiáveis. O RSQ mede a adequação das previsões aos valores de referência: um valor próximo de 1 indica um ajustamento excelente. A RDP, que é a razão entre o desvio padrão dos valores de referência e o SEP, reflete a qualidade do modelo: um RDP maior representa um modelo superior. O RER compara o erro do modelo com o erro dos dados de referência. Quanto menor o valor, melhor o modelo

em relação aos dados de referência. Por fim, o slope, quando igual a 1, indica que não há diferença significativa entre as previsões do modelo e os valores de referência. Valores menores que 1 indicam uma subestimação, enquanto valores maiores que 1 indicam uma superestimação das previsões do modelo em relação aos valores de referência. Estes parâmetros são essenciais para avaliar a qualidade e precisão das previsões do modelo em relação aos dados reais.

Analisando o parâmetro MS na condição (1,4,4,1), observa-se um SEP inferior (0,452) em comparação com a condição (2,4,4,1) (0,481). O BIAS é baixo (0,011), próximo de zero indica previsões imparciais e consistentes. A proximidade do RSQ a 1 indica um bom ajuste aos dados reais. O elevado valor RDP de 12,9558 sugere uma precisão robusta nas previsões. Nesta análise, a condição (1,4,4,1) destaca-se, apresentando um menor erro padrão e melhor ajuste aos dados reais. Porém é preciso avaliar novamente o modelo usado, pois apesar de a robustez ser habitualmente considerada uma qualidade desejável em métodos analíticos, um potencial problema de um método excessivamente robusto reside na sua falta de sensibilidade para detetar variações ou nuances subtis nas amostras. Portanto, é crucial encontrar um equilíbrio entre a robustez e sensibilidade ao desenvolver métodos analíticos.

Na condição (2,4,4,1) para o CT, na qual foi eliminado o *outlier* 222V, observa-se SEP inferior (0,632) em comparação com a condição (1,4,4,1) (0,697). O BIAS é próximo de zero o que indica previsões imparciais. O RSQ de 0,7450 sugere um ajuste razoável aos dados reais. O Valor de RDP de 1,9747 é aceitável, indicando uma precisão adequada nas previsões. Neste contexto, a condição (2,4,4,1) destaca-se ligeiramente, apresentando um SEP inferior e um RDP aceitável.

Já na condição (1,4,4,1) para a PB, observa-se um SEP inferior (0,805) em comparação com a condição (2,4,4,1) (0,89). O RSQ de 0,8650 sugere um ajuste robusto aos dados reais. O Valor de RDP de 2,6783 indica uma boa precisão nas previsões. Desta forma, a condição (1,4,4,1) é a que reúne mais condições, apresentando um SEP inferior e um RDP favorável, indicando maior precisão e confiabilidade nas previsões para a Proteína Bruta.

No contexto da análise da GB, a Condição (2,4,4,1) apresenta um SEP relativamente inferior (0,186) em comparação com a Condição (1,4,4,1) (0,19). O RDP de 2,4000 é considerado aceitável, indicando uma precisão razoável nas previsões. O coeficiente de determinação de 0,8460 sugere um bom ajuste do modelo aos dados reais, enquanto o BIAS próximo de zero confirma a imparcialidade nas previsões. Ambas as condições são parecidas em termos de desempenho, mas a (2,4,4,1) tem uma ligeira

vantagem devido ao SEP um pouco menor e um RSQ ligeiramente superior. Estes indicadores sugerem que, a Condição (2,4,4,1) pode ser a escolha mais favorável devido ao seu desempenho ligeiramente superior

No contexto da análise de FB, a Condição (1,4,4,1) na qual foi eliminado o *outlier* 222V, apresenta um SEP inferior (1,052) em comparação com a Condição (2,4,4,1) (1,179). Ambas as condições exibem um RDP aceitável, com um RDP de 1,9737 para a condição da segunda derivada e com um RDP de 2,2120 para a primeira derivada. O RSQ de 0,7560 sugere um ajuste razoável do modelo aos dados reais. Em geral, ambas as condições podem ser consideradas aceitáveis, mas a (1,4,4,1) tem uma ligeira vantagem devido ao seu SEP inferior, o que indica uma precisão um pouco melhor nas previsões da Fibra Bruta. No entanto, é importante notar que ambas as condições ainda podem beneficiar de melhorias para atingir um melhor desempenho.

As relações entre os resultados obtidos pela análise de referência e as previsões do NIRS para os parâmetros Cinza Total e Fibra Bruta, utilizando o tratamento matemático (2,4,4,1) com comprimentos de onda de 402-1098,2 e 1102-2498,2 nm e que foram avaliadas por meio de validação cruzada (Grupos 12345) e comprimento de onda de apenas 1102-2498 nm validação cruzada com exclusão de um único ponto (One out), respetivamente, são apresentadas nas Figuras 32 a 35.

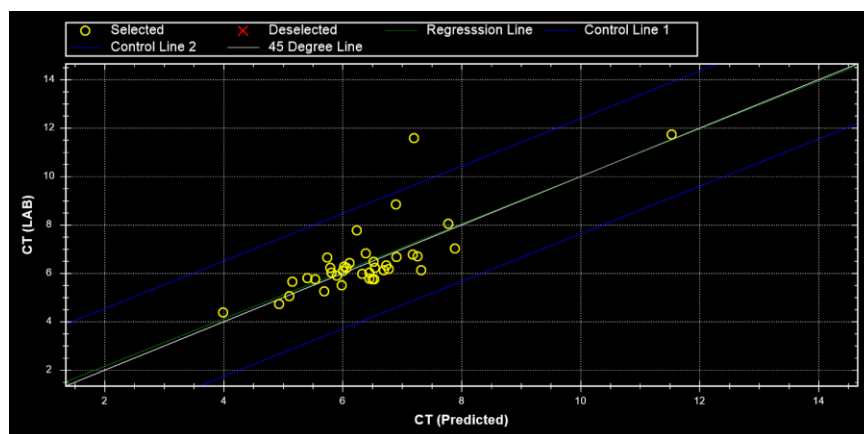


Figura 32: Comparação dos valores de CT de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (com outlier 222V incluído).

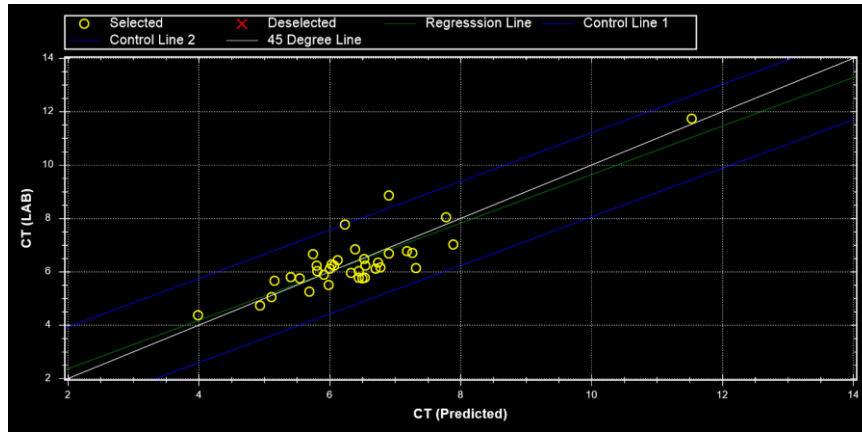


Figura 33: Comparação dos valores de CT de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (sem outlier 222V).

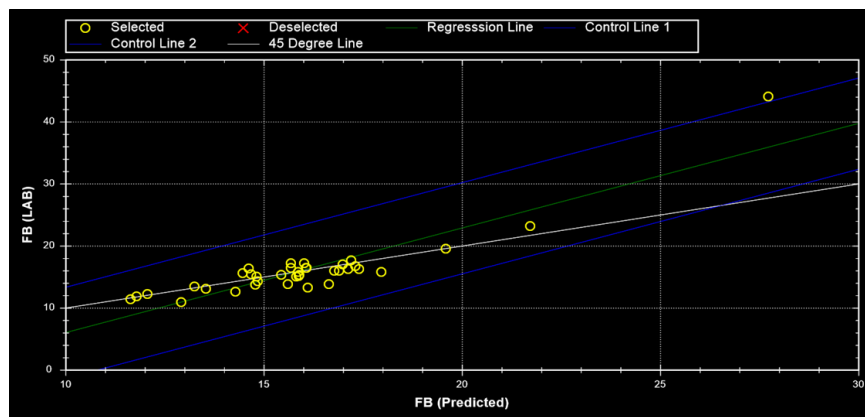


Figura 34: Comparação dos valores de FB de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (com outlier 222V incluído).

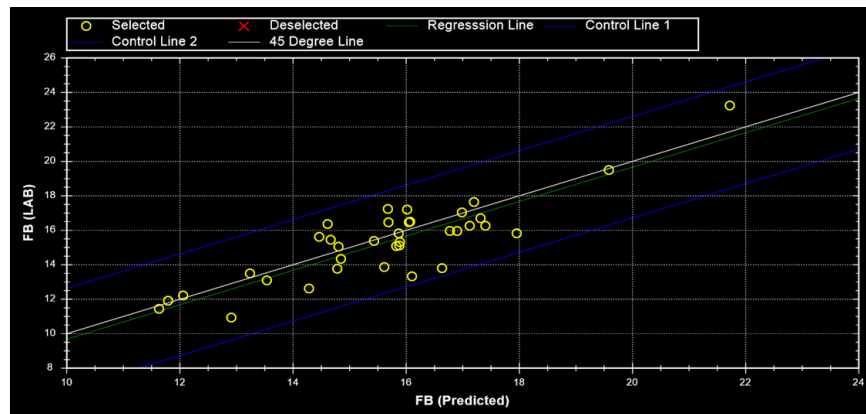


Figura 35: Comparação dos valores de FB de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa (sem outlier 222V).

Cada uma destas figuras representa um gráfico que ilustra a correlação nos tratamentos considerados, permitindo observar o impacto da presença de um *outlier* no modelo de regressão. Assim, é notável a influência significativa da presença de um *outlier*, pelo que a sua identificação e remoção cuidadosa é crucial para a precisão e robustez dos modelos de calibração NIRS. A presença destes valores discrepantes pode distorcer significativamente as previsões, enfatizando a importância da análise crítica dos dados e do processo de validação para garantir resultados confiáveis em análises NIRS de misturas TMR.

9. Conclusão

A aplicação da espectroscopia de infravermelho próximo em conjunto com técnicas de quimiometria revelou-se uma estratégia altamente eficaz para a previsão das propriedades da TMR. Ao substituir as análises químicas e nutricionais tradicionais, o NIRS aliado à quimiometria proporcionam uma alternativa inovadora e confiável para a avaliação precisa das composições das amostras TMR. Esta metodologia não só otimiza o tempo e os recursos, mas também contribui para a eficiência e qualidade dos processos na indústria agropecuária.

No desenvolvimento das calibrações para os parâmetros MS, CT, PB, GB e FB, foi utilizada a regressão MPLS em conjunto com o tratamento SNV-DT, combinado com tratamento de primeira e segunda derivada e diferentes intervalos de comprimento de onda. Após o desenvolvimento e validação das equações, a escolha da condição ideal dependeu do parâmetro específico em consideração. Para MS, PB e FB, a condição (1,4,4,1) revelou-se a melhor escolha, enquanto para CT e GB, a condição (2,4,4,1) foi preferível. É crucial notar que ambas as condições para FB ainda requerem ajustes para alcançar um melhor desempenho.

É imprescindível explorar estratégias alternativas de calibração com o intuito de obter resultados mais precisos. No âmbito deste estudo, considerando a constituição não específica da amostra, suscetível a variações, e a ausência de correlação linear entre os parâmetros analisados, ressalta-se a necessidade crucial de identificar e tratar valores aberrantes (*outliers*). Esta abordagem torna-se fundamental para assegurar a precisão e a validade dos resultados obtidos, bem como para a interpretação adequada dos dados, contribuindo assim para a robustez da investigação científica em questão.

Em suma, esta metodologia demonstrou um grande potencial para o melhoramento das calibrações, visando a previsão dos parâmetros analisados. Antevê-se a sua aplicação na otimização de modelos existentes e no desenvolvimento de novos componentes para a mistura Unifeed. Num futuro próximo, seria benéfico agrupar as amostras de acordo com características específicas, como a presença de casca de tomate, por exemplo. Dessa forma, estas amostras não seriam excluídas ou prejudicadas na calibração, mas sim integradas de maneira adequada, ampliando assim a abrangência e a precisão do estudo. Este trabalho desenvolvido não só revelou os desafios intrínsecos associados à complexa metodologia de calibração NIR

em amostras TMR, mas também apresentou soluções inovadoras para superar esses mesmo obstáculos.

Referências Bibliográficas

1. Associação Interprofissional do Leite e Lacticínios (ALIP). (2023). *História*. Recuperado de <https://www.alip.pt/historia.html>.
2. *Anexo Técnico de Acreditação L0512-1 ALIP-Associação Interprofissional do Leite e Lacticínios Laboratório No Produto Ensaio Método de Ensaio* (s.d). Recuperado de www.ipac.pt.
3. (ALIP-Associação Interprofissional do Leite e Lacticínios). (2023). *Missão*. Recuperado de <https://www.alip.pt/missao.html>.
4. McDonald, P., et al. (s.d). *Animal Nutrition* (7ª ed). Pearson. Recuperado de www.pearson-books.com.
5. Wu,G. (2018). *Principles of Animal Nutrition*.
6. Schingoethe, D. J. (1991). *Byproduct feeds. Feed analysis and interpretation. The Veterinary clinics of North America. Food animal practice*, (pp. 577–584). [https://doi.org/10.1016/S0749-0720\(15\)30787-8](https://doi.org/10.1016/S0749-0720(15)30787-8)
7. Lage, A. (2017). *Silagens de milho produzidas em Portugal: Avaliação da qualidade nutricional* (26ª ed), Revista da Agropecuária Ruminantes
8. Lage, A. (2019). *A importância da maturidade na qualidade nutricional das silagens de milho*. (44ª ed), Revista da Agropecuária Ruminantes.
9. Lage, A., & Salgueiro, M. (2019). *A Variação do Teor Proteico das Silagens de Milho em Portugal*.
10. Lage, A., & Salgueiro, M. (2020) *Silagem de erva: avaliação da qualidade nutricional*.
11. Lage,A. (2021). *As Forragens Vencedoras do CNF 2020*. Revista Ruminantes.
12. Kammel, D. W. (2019). *Design, Selection and Use of TMR Mixers*. Dairy-Cattle. Recuperado de <https://dairy-cattle.extension.org/design-selection-and-use-of-tmr-mixers/>.
13. Borges de Freitas, A. (2008). *Sistema de Alimentação UNIFEED: Rações Completas*. Notícias Limousine Nº17, (pp. 33-36).
14. Gonçalves, L. C., Borges, I., & Ferreira, P. D. S. (2009). *Alimentação de Gado de Leite*. Belo Horizonte: FEPMVZ-Editora.

15. Cherian, G. (s.d). *A Guide to the Principles of Animal Nutrition*. Oregon State University, Corvallis, OR.
16. Lloyd, L. E., McDonald, B. E., & Crampton, E. W. (1978). *Fundamentals of Nutrition* (2^a ed.). A Series of Books in Animal Science. G. W. Salisbury.
17. ISO. (1999). *Animal feeding stuffs – Determination of fat content* (ISO 6492). Geneva, Switzerland.
18. Bianchini, W., Rodrigues, É., Mendes Jorge, A., & Andrigheto, C. (2007). *Importância da fibra na nutrição de bovinos*. Revista Electrónica de Veterinaria, VIII(2)
19. ISO. (2000). *Animal feeding stuffs – Determination of crude fibre content – Method with intermediate filtration* (ISO 6865). Geneva, Switzerland: International Organization for Standardization.
20. Tibola, C. S., Medeiros, E. P. de, Simeone, M. L. F., & Oliveira, M. A. de. (2018). *Espectroscopia no infravermelho próximo para avaliar indicadores de qualidade tecnológica e contaminantes em grãos*. Embrapa Trigo.
21. Pasquini, C. (2003). *Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications*. Instituto de Química, Universidade Estadual de Campinas. doi:10.1590/S0103-50532003000200006
22. Scotter, C. N. G. (2005). *Near-Infrared, Infrared Spectroscopy*, 415-426. Elsevier.
23. Ozaki, Y., Huck, C., Tsuchikawa, S., & Engelsen, S. B. (2020). *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications*. In *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications* (pp. 1-593). Springer Singapore. doi:10.1007/978-981-15-8648-4.
24. El-Azazy, M. S., El-Shafie, A., & Al-Saad, K. (2023). *Introductory Chapter: Infrared Spectroscopy - Principles and Applications*. In *Infrared Spectroscopy - Perspectives and Applications*. IntechOpen. doi:10.5772/intechopen.109139.
25. Czarnecki, M. A., Morisawa, Y., Futami, Y., & Ozaki, Y. (2015). *Advances in Molecular Structure and Interaction Studies Using Near-Infrared Spectroscopy*. Chemical Reviews. doi:10.1021/cr500013u.

26. Mohamed, E. S., Saleh, A. M., Belal, A. B., & Gad, A. A. (2018). *Application of near-infrared reflectance for quantitative assessment of soil properties Egyptian Journal of Remote Sensing and Space Science*, doi:10.1016/j.ejrs.2017.02.001.
27. Pasquini, C. (2018). *Near infrared spectroscopy: mature analytical technique with new perspectives – A review*. *Analytica Chimica Acta* doi:10.1016/j.aca.2018.04.004
28. . Ozaki, Y., Huck, C., Tsuchikawa, S., & Engelsen, S. B. (2021). *Near-Infrared Spectroscopy: Theory, Spectral Analysis, Instrumentation, and Applications* (1^aed.). Springer.
29. Sun, D.-W. (2008). *Modern Techniques for Food Authentication* (1^aed.). Academic Press.
30. Bruker.. *O que é Espectroscopia FT-NIR*. Disponível em: <https://www.bruker.com/en/products-and-solutions/infrared-and-raman/ft-nir-spectrometers/what-is-ft-nir-spectroscopy.html>. Acedido a 9 de abril de 2023.
31. Næs, T. (2002). *A user-friendly guide to multivariate calibration and classification*. NIR Publications.
32. Sileoni, V., Marconi, O., Perretti, G., & Fantozzi, P. (2013). *Evaluation of different validation strategies and long term effects in NIR calibration models*. *Food Chem*.
33. Coates, J. (2000). *Interpretation of Infrared Spectra, A Practical Approach*. R.A. Meyers, Encyclopedia of Analytical Chemistry.
34. Reich, G. (2005). *Near-infrared spectroscopy and imaging: Basic principles and pharmaceutical applications*. *Advanced Drug Delivery Reviews*, doi:10.1016/j.addr.2005.01.020.
35. Agelet, L. E., & Hurburgh, C. (2010). *A Tutorial on Near Infrared Spectroscopy and Its Calibration*. *Critical Reviews in Analytical*.
36. Scarff, M., Arnold, S. A., Harvey, L. M., & McNeil, B. (2006). *Near infrared spectroscopy for bioprocess monitoring and control: Current status and future trends*. *Critical Reviews in Biotechnology*.
37. Workman, J. J. (1996). *Interpretive Spectroscopy for Near Infrared*. *Applied Spectroscopy Reviews*.

38. Steinbach, A.. *NIR Spectroscopy: A guide to near-infrared spectroscopic analysis of industrial manufacturing processes*. Monograph NIR Spectroscopy.
39. Alomar, D., et al. (2006). *Nutritional evaluation of commercial dry dog foods by near infrared reflectance spectroscopy*. Journal of Animal Physiology and Animal Nutrition.
40. Fernandez, V. M et.al (2006). *Evaluation of Pretreatment Strategies for Near-Infrared Spectroscopy Calibration Development of Unground and Ground Compound Feedingstuffs*. doi:10.1366/000370206775382839
41. Williams, P. (2015). *How it all began*. In: Near Infrared Spectroscopy: Proceedings of the International Conference. doi:10.17648/nir-2015-36736.
42. Williams, P. C., & Starkey, P. M. (1980). *Influence of Feed Ingredients Upon the Prediction of Protein in Animal Feed-mixes by Near-infrared Reflectance Spectroscopy*. Journal of Science and Food Agriculture.
43. Ferreira, M., et al. (2015). *Quimiometria: Conceitos, Métodos e Aplicações*. Editora da Unicamp.
44. Blanco, M., & Villarroya, I. (2002). *NIR spectroscopy: a rapid-response analytical tool*. TrAC Trends in Analytical Chemistry.
45. Huang, J., Romero-Torres, S., & Moshgbar, M. (2010). *Considerations in Data Pre-treatment for NIR and Raman Spectroscopy*. American Pharmaceutical Review.
46. Nobari Moghaddam, H., Tamiji, Z., Akbari Lakeh, M., Khoshayand, M. R., & Haji Mahmoodi, M. (2022). *Multivariate Analysis of Food Fraud: A Review of NIR Based Instruments in Tandem with Chemometrics*. Journal of Food Composition and Analysis, doi:10.1016/j.jfca.2021.104343.
47. Jiao, Y., Li, Z., Chen, X., & Fei, S. (2020). *Preprocessing Methods for Near-infrared Spectrum Calibration*. J Chemom.
48. Mancini, M., Toscano, G., & Rinnan, Å. (2019). *Study of the Scattering Effects on NIR Data for the Prediction of Ash Content Using EMSC Correction Factors*. J Chemom.
49. Silalahi, D. D., Midi, H., Arasan, J., Mustafa, M. S., & Caliman, J. P. (2018). *Robust Generalized Multiplicative Scatter Correction Algorithm on Pretreatment of Near Infrared Spectral Data*. Vib Spectrosc.

50. Dunn, K. (2023). *Process Improvement*. LearnChe.
51. Kong, X., Luo, J., Xu, Z., & Li, H. (2019). *Quality-Relevant Data-Driven Process Monitoring Based on Orthogonal Signal Correction and Recursive Modified PLS*. IEEE Access.
52. Westerhaus, M. (2014). *Eastern Analytical Symposium Award for Outstanding Achievements in Near Infrared Spectroscopy: My Contributions to Near Infrared Spectroscopy*. NIR News, 25, 16–20.
53. Hernández-Hierro, J. M., García-Villanova, R. J., & González-Martín, I. (2008). *Potential of Near Infrared Spectroscopy for the Analysis of Mycotoxins Applied to Naturally Contaminated Red Paprika Found in the Spanish Market*. Anal Chim Acta, 622, 189–194.
54. Valderrama, P., Braga, J. W. B., & Poppi, R. J. (2007). *Variable Selection, Outlier Detection, and Figures of Merit Estimation in a Partial Least-squares Regression Multivariate Calibration Model: A Case Study for the Determination of Quality Parameters in the Alcohol Industry by Near-infrared Spectroscopy*. J Agric Food Chem, 55, 8331–8338.
55. Garrido-Varo, A., Garcia-Olmo, J., & Fearn, T. (2019). *A Note on Mahalanobis and Related Distance Measures in WinISI and The Unscrambler*. J Near Infrared Spectrosc, 27, 253–258.
56. Doležal, P.. (2011). *Vliv Zkrmování Kvasinkové Kultury na Bachorovou Fermentaci a Krevní Ukazatele Krav Holštýnského Plemene*. Acta Veterinaria Brno, 80, 139–145.
57. Madeira, C. R. de Souza, Piantavini, M. S., & Trindade, A. C. L. B. (2015). *Application of infrared spectroscopy coupled with multivariate calibration for drug quality control - revision*. Visão Acadêmica, 16.
58. Helfer, G. A., Ferrão, M. F., Ferreira, C. D. V., & Hermes, N. (2006). *Using multivariate analysis methods in quality control of food essence by mid-infrared spectroscopy*. Ciencia e Tecnologia de Alimentos, 26, 779–786.
59. Geladi, P., & Kowalski, B. R. (1986). *Partial least-squares regression: A tutorial*. Analytica Chimica Acta, 186.