

MSC

2.º
CICLO

FCUP
2022

U. PORTO

Desenvolvimento de um modelo de calibração
MPLS-NIR para amostras de silagem de milho em
fresco

Dulce Moraes

FC

U. PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO



Universidade do Minho
Escola de Engenharia

Desenvolvimento de um modelo de calibração MPLS-NIR para amostras de silagem de milho em fresco

Dulce Moraes

Dissertação de Mestrado apresentada à
Faculdade de Ciências da Universidade do Porto em
Quimiometria

2022



Universidade do Minho
Escola de Engenharia

U. PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO



Desenvolvimento de um modelo de calibração MPLS-NIR para amostras de silagem de milho em fresco

Dulce Moraes

Mestrado em Tecnologia e Ciência Alimentar
Departamento Química e Bioquímica
2022

Orientador

Victor Freitas, Professor Catedrático, FCUP

Coorientador

Lúcia Medeiros, Diretora Geral, ALIP





Universidade do Minho
Escola de Engenharia

Todas as correções determinadas
pelo júri, e só essas, foram efetuadas.

O Presidente do Júri,

Porto, ____/____/____





Universidade do Minho
Escola de Engenharia

MSc

O Presidente do Júri,

Porto,

____/____/____

**Desenvolvimento de um modelo de calibração MPLS-NIR
para amostras de silagem de milho em fresco**

Dulce Morais

Mestrado em Tecnologia e Ciência Alimentar

Departamento de Química e Bioquímica

2022

Desenvolvimento de um modelo de calibração MPLS-NIR para amostras de silagem de milho em fresco

Dulce Morais

Mestrado em Tecnologia e Ciência Alimentar

Departamento de Química e Bioquímica

MSCMSC

Agradecimentos

A elaboração deste trabalho experimental só se tornou possível com as colaborações, apoios e incentivos de diversas pessoas. Desde já, queria manifestar todo o meu apreço e gratidão a todas elas que, de uma forma ou de outra, contribuíram para que este projeto se realizasse.

Ao orientador Professor Doutor Victor Freitas, pela oportunidade de realização deste projeto, pela sua disponibilidade e incentivo.

À coorientadora Doutora Lúcia Medeiros, pela oportunidade de realização deste projeto na empresa ALIP, pela disponibilidade, pelo empenho, e incentivos.

As colegas de laboratório, nomeadamente Engenheira Ana Lage, pela sua disponibilidade, acompanhamento, empenho e dedicação; e Liliana Magalhães pelo incentivo e motivação.

À restante equipa Alip, pela forma como me receberam e acarinharam ao longo deste tempo.

Aos meus pais, irmã e namorado por estarem sempre presente nesta jornada, pelo incentivo, apoio e amor recebido, nunca deixando baixar os braços perante as dificuldades.

Ao professor Doutor Carlos Alberto Gomes que foi fulcral na realização deste projeto, essencialmente pelos ensinamentos, disponibilidade, empenho e dedicação.

A todas as outras pessoas, que diretamente ou indiretamente, ajudaram a tornar tudo isto possível.

Um Muito Obrigada!

Resumo

A silagem de milho é uma das bases alimentares mais usada nas explorações leiteiras. No entanto, poucos estudos foram realizados para avaliar qualitativa e quantitativamente as silagens de milho por espectroscopia de infravermelho próximo (NIRS). O objetivo deste estudo foi desenvolver um modelo de calibração por MPLS para prever parâmetros químicos e nutricionais da forragem, tais como: matéria seca (MS), cinza total, proteína bruta (PB), amido, fibra de detergente neutro (NDF) e pH. Para isso, 241 amostras de silagens de milho em fresco foram utilizadas para desenvolver equações de calibrações com tratamentos de SNV-DT juntamente com derivadas. As equações desenvolvidas com a primeira derivada foram as equações com melhores resultados nas previsões de todos os parâmetros. Os coeficientes de correlação múltipla (RSQ) do conjunto de validação externa foram 0,96 para MS, 0,63 para cinza total, 0,62 para PB, 0,76 para amido, 0,79 para NDF e 0,77 para pH. Os desvios de previsões relativos para os modelos de calibrações da primeira derivada indicam que as equações de MS (4,57), amido (2,07) e NDF (2,06) foram considerados bons modelos para prever quantitativamente os parâmetros das silagens de milho em fresco e que as equações de cinza total (1,56), PB (1,61) e pH (1,90) necessitam de ser ajustadas utilizando novas estratégias de calibração.

Palavras-chave: Espectroscopia de infravermelho próximo, quimiometria, calibração, MPLS, validação, silagem de milho.

Abstract

Corn silage is one of the most used food bases in dairy farms. However, few studies have been carried out to qualitatively and quantitatively evaluate corn silages by near infrared spectroscopy (NIRS). The objective of this study was to develop an MPLS calibration model to predict chemical and nutritional parameters of forage, such as: dry matter (MS), total ash, crude protein (PB), starch, neutral detergent fiber (NDF) and pH. For this, 241 fresh corn silage samples were used to develop calibration equations with SNV-DT treatments together with derivatives. The equations developed with the first derivative were the equations with the best results in the predictions of all parameters. The multiple correlation coefficients (RSQ) of the external validation set were 0,96 for MS, 0,63 for total ash, 0,62 for PB, 0,76 for starch, 0,79 for NDF and 0,77 for pH. The relative prediction deviations for the calibration models of the first derivative indicate that the equations of MS (4,57), starch (2,07) and NDF (2,06) were considered good models to quantitatively predict the parameters of silages of fresh corn and that the equations of total ash (1,56), PB (1,61) and pH (1,90) need to be adjusted using new calibration strategies.

Keywords: Near infrared spectroscopy, chemometrics, calibration, MPLS, validation, corn silage.

Índice

1. Introdução.....	1
2. Revisão Bibliográfica.....	3
2.1 Alimentos para animais ruminantes.....	3
2.2 Composição química dos alimentos	5
2.2.1 Matéria Seca	8
2.2.2 Cinza Total.....	8
2.2.3 Proteínas.....	9
2.2.4 Gordura.....	10
2.2.5 Fibras.....	11
2.2.6 Hidratos de Carbono	12
2.3 Espectroscopia NIR	12
2.3.1 Breve história	12
2.3.2 Importância na Indústria	13
2.3.3 Princípios Teóricos.....	14
2.3.4 Instrumentação	17
2.3.5 Métodos de medição	20
2.3.6 Zonas Características de Absorção	23
2.3.7 Fatores que afetam os espectros.....	24
2.3.8 Vantagens e desvantagens da espectroscopia NIR	25
2.4 Quimiometria	26
2.4.1 Análise Multivariada	27
2.4.2 Pré-processamentos de espectros NIR.....	28
2.4.3 Análise de Componentes Principais	35
2.4.4 Regressão por Mínimos Quadrados Parciais	37
2.4.5 Validação do modelo	40
2.4.6 Detecção de <i>Outliers</i>	41
3. Experimental	42

3.1 Preparação das amostras.....	42
3.2 Espectrofotômetro de refletância NIR.....	43
3.3 Análise química de referência.....	44
3.4 Desenvolvimento e validação da calibração	45
4. Resultados e Discussão	48
4.1 Análise química de referência.....	48
4.2 Espectros NIRS	52
4.3 Estatística do desenvolvimento e validação da calibração	55
5. Considerações finais e perspectivas futuras	65
6. Referências.....	66

Lista de Figuras

Figura 1. Gama de radiação eletromagnética.....	15
Figura 2. Diagrama de níveis de energia molecular. S_0 : Estado Fundamental; S_1 : 1º Estado Excitado; S_2 : 2º Estado Excitado	15
Figura 3. Modelo anarmônico - transições vibracionais.....	16
Figura 4. Métodos de medições para a aquisição de espectros em espectroscopia NIR. a) refletância; b) transmitância; c) transfletância; d) interatância. A: Amostra; I_0 : Intensidade de radiação incidente; I : Intensidade de radiação emergente	22
Figura 5. Espectro NIR de água sem tratamento com uma variação de temperatura.....	25
Figura 6. Representação da aplicação do método centralização na média a uma variável.....	30
Figura 7. Representação gráfica do resultado da derivação nos efeitos aditivos (verde) e aditivos mais multiplicativos (vermelho). O espectro azul apresenta um espectro sem quaisquer deslocamentos possíveis. A linha tracejada a preto é a linha zero.....	33
Figura 8. Representação gráfica de um conjunto de dados bidimensionais (λ_1 e λ_2) juntamente com eixos das componentes principais (PC1 e PC2). Linha cheia (—) representa os scores de uma amostra; linha tracejada (- - -) representa os loadings; as linhas tracejadas e ponteadas (— . . —) representam as coordenadas de uma amostra em relação aos eixos originais.....	37
Figura 9. Equipamento FOSS NIRS TM DS2500 utilizado para obtenção de espectros NIR juntamente com o copo de grande dimensão.....	44
Figura 10. Representação gráfica do conjunto de diagrama de bigodes de cada parâmetro analisado. a) Diagrama de bigodes referente à MS. b) Diagrama de bigodes referente à cinza total. c) Diagrama de bigodes referente à PB. d) Diagrama de bigodes referente ao amido. e) Diagrama de bigodes referente ao NDF. f) Diagrama de bigodes referente ao pH.....	50
Figura 11. Representação gráfica das correlações existentes entre os parâmetros analisados nas amostras de silagem de milho, expondo em histogramas apenas a relação de um componente e a linha de tendência a vermelho para as relações entre dois parâmetros diferentes.....	52
Figura 12. Espectros NIRS de amostras de silagem de milho sem e com tratamentos quimiométricos. a) Espectro NIRS em Log (1/R). b) Espectro NIRS com aplicação do tratamento SNV-DT. c) Espectro NIRS com aplicação dos tratamentos SNV-DT e a primeira derivada. d) Espectro NIRS com aplicação dos tratamentos SNV-DT e a segunda derivada. e) Espectro NIRS com aplicação dos tratamentos SNV-DT e a terceira derivada.....	53/54
Figura 13. Comparação dos valores de MS de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores utilizando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores utilizando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores utilizando MPLS juntamente com SNV-DT e 3,5,5,1.....	62

Figura 14. Comparação dos valores de cinza total de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.....62

Figura 15. Comparação dos valores de PB de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.....63

Figura 16. Comparação dos valores de amido de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.....63

Figura 17. Comparação dos valores de NDF de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.....64

Figura 18. Comparação dos valores de pH de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.....64

Lista de Tabelas

Tabela 1. Sistema de Weende: frações, componentes majoritários, procedimentos e métodos oficiais	7
Tabela 2. Resumo dos componentes essenciais de um espectrofotômetro NIR e seus respectivos tipos e materiais	20
Tabela 3. Exemplos de zonas típicas de absorção de compostos presentes em alimentos para animais na região NIR.....	23
Tabela 4. Resumo estatístico dos parâmetros analisados nas amostras de silagem de milho, expondo o mínimo, o máximo, a média e o desvio padrão de cada componente.....	48
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 silagem de milho em fresco.	56
Tabela 6. Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e segunda derivada (2,4,4,1) para cada parâmetro a analisado nas amostras de silagem de milho em fresco.....	57
Tabela 7. Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e terceira derivada (3,5,5,1) para cada parâmetro a analisado nas amostras de silagem de milho em fresco.....	57
Tabela 8. Estatística de monitoramento do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (1,4,4,1)) para cada parâmetro de silagem de milho em fresco.....	59
Tabela 9. Estatística de monitoramento do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (2,4,4,1)) para cada parâmetro de silagem de milho em fresco.....	60
Tabela 10. Estatística de monitoramento do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (3,5,5,1)) para cada parâmetro de silagem de milho em fresco.....	60

Lista de Abreviaturas

ADF – *Acid detergent fiber* (Fibra de detergente ácido)

ADL – *Acid detergent lignin* (Lenhina de detergente ácido)

ALIP – Associação Interprofissional do Leite e Laticínios

AOAC – *Association of official analytical chemists* (Associação dos químicos analíticos oficiais)

AOTF – *Acousto-optic tunable filters* (Filtro sintonizável acústico-ótico)

CCDs – *Charged coupled devices* (Dispositivos de carga acoplada)

DT – *De-Trending*

EMSC – *Extended multiplicative scatter correction* (Correção de dispersão multiplicativa estendida)

FB – Fibra bruta

FIR – *Far infrared* (Infravermelho distante)

FOM – *Figures of merit* (Figuras de mérito)

FT-NIR – *Fourier transform near-infrared* (Infravermelho próximo com transformada de Fourier)

GB – Gordura bruta

GH – Distância de Mahalanobis

InGaAs – *Indium-gallium-arsenide* (Índio-gálio-arsénio)

IR – *Infrared* (Infravermelho)

LED – *Light emitting diodes* (Díodos emissores de Luz)

LOO – *Leave-one-out*

MC – *Mean Centering* (Centralização na média)

MIR – *Middle infrared* (Infravermelho médio)

MPLS – *Modified partial least squares* (Mínimos quadrados parciais modificados)

MS – Matéria seca

MSC – *Multiplicative Scatter Correction* (Correção multiplicativa de sinal)

NDF – *Neutral detergent fibre* (Fibra de detergente neutro)

NIR – *Near-infrared reflectance* (Infravermelho próximo)

NIRS – *Near-infrared reflectance spectroscopy* (Espectroscopia de infravermelho próximo)

NW – Norris-Williams

PAT – *Process analytical techniques* (Técnicas de processos analíticos)

PB – Proteína bruta

PbS – *Lead sulfid* (Sulfureto de chumbo)

PCA – *Principal Component* (Análise de componentes principais)

PCs – *Principal components* (Componente principais)

PLS – Partial least squares (Mínimos quadrados parciais)

RPD – *Ratio of prediction to deviation* (Desvio de previsão relativo)

RSQ – *Squared coefficient of correlation* (Coeficiente de correlação quadrado)

SD – *Standard deviation* (Desvio padrão)

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)

SG – Savitzky-Golay

SLED- *Superluminescent light emitting diodes* (Díodos emissores de luz superluminescente)

SNR – Signal to noise ratio (Relação sinal-ruído)

SNV – *Standard Normal Variate* (Variação de padrão normal)

SNV-DT – *Combination of SNV and DT* (Combinação da variação de padrão normal com de-trending)

TMR – *Total mixed ration* (Mistura unifeed)

VL – Variáveis latentes

1-VR – *One minus the variance ratio* (Um menos a razão de variância)

1. Introdução

As forragens representam uma percentagem significativa na dieta dos animais ruminantes, podendo ser utilizadas em verde ou conservadas. As silagens de milho são uns dos alimentos base usados nas explorações leiteiras, pois apresentam-se como uma boa fonte de fibra digerível. Contudo, este tipo de alimento exhibe uma grande variação no valor nutritivo, devido a fatores como: clima, região, variedade da planta, práticas de processo de conservação, mas essencialmente pelo estado de maturação e a altura do corte (Lage, 2017, 2019). Dessa forma, a avaliação da composição química dos alimentos é um dos principais pontos a se considerar para obter sucesso na nutrição animal. A procura pelo conhecimento da composição nutricional de cada alimento, relacionado ao seu aproveitamento digestivo e metabólico, tem levado à procura de métodos analíticos mais eficientes, rápidos, baratos e de baixo impacto ambiental (Cherian, 2020). Isto deve-se ao facto de a quantificação de compostos, bem como a determinação de parâmetros de qualidade dos produtos alimentícios, envolverem etapas experimentais que necessitam de preparação da amostra, utilização de solventes que muitas vezes são de carácter tóxico e que requerem tempo para a sua execução.

A espectroscopia de infravermelho próximo (NIRS) é uma ciência que permite adquirir, medir e interpretar espectros resultantes da interação da radiação eletromagnética com a matéria. Dessa interação origina a absorção da radiação pelos constituintes de amostras a analisar que por consequente gera espectros de absorção que contém informações analíticas qualitativas e quantitativas (Pasquini, 2018b). No entanto, este tipo de espectroscopia tem de recorrer a tratamentos quimiométricos para promover uma simplificação e extração de informações relevantes, já que características como tamanho das partículas, temperatura, entre outros, influenciam a expressão dos espectros (Fernández-Cabanás et al., 2006). Basicamente, a NIRS surge como uma técnica de apoio a indústrias, visto que facilmente consegue determinar os parâmetros pretendidos de forma rápida e simultânea. Contudo existe a necessidade de desenvolver calibrações robustas para que seja possível quantificar e qualificar diferentes variedades dentro de cada alimento.

O laboratório da ALIP, situado em Lousada, distrito do Porto, iniciou a atividade a 02 de outubro de 2006 efetuando análises relacionadas com a classificação e valorização do leite ao produtor e contraste leiteiro de todo o território continental. Em janeiro de 2008, pelo despacho nº 4215/2008 emanado pelo Sr. Ministro da Agricultura, a ALIP foi reconhecida como OI (organização interprofissional), tendo a partir deste momento, o Estatuto de Utilidade Pública. Em junho de 2011 é concedida a acreditação pelo IPAC, de

acordo com a NP EN ISO/IEC 17025 (Anexo técnico nº L0512). Ainda no final de 2012, o Laboratório entrou noutra campo de ação, designadamente análises a alimentos para animais. A valorização da qualidade do leite constitui um fator de extrema importância no preço final pago ao produtor e tem como objetivo favorecer o desenvolvimento de programas de melhoria de eficiência nas explorações, aumentar os rendimentos no processo de transformação e garantir ao consumidor final a ausência de riscos sanitários. Cada vez mais, existe uma maior procura dos vários intervenientes do setor leiteiro, para a obtenção dos parâmetros nutricionais dos alimentos fornecidos aos seus animais, pois existe uma grande relação entre a qualidade dos alimentos fornecidos com a qualidade dos produtos obtidos dos mesmos, mais especificamente o leite.

Desta forma, o objetivo deste presente trabalho é o desenvolvimento de um modelo de calibração MPLS para prever diversos parâmetros, tais como: matéria seca, cinza total, proteína bruta, amido, fibra em detergente neutro e pH, em amostras de silagem de milho em fresco.

2. Revisão Bibliográfica

2.1 Alimentos para animais ruminantes

O alimento é toda a substância que é capaz de ser digerido, absorvido e utilizado pelos seres vivos, de forma a promover a sustentabilidade da vida, o crescimento, a saúde e a produtividade dos mesmos. Porém, nem todos os componentes de um alimento conseguem ser digeridos pelos animais. Do mesmo modo, os alimentos não apresentam a mesma capacidade de atender aos requisitos de manutenção, crescimento, reprodução e lactação (McDonald et al., 2011; Soest, 2018; Wu, 2018). Assim sendo, a nutrição começou de dia para dia a ganhar mais importância na alimentação dos animais, com finalidade de promover a manutenção e qualidade dos produtos (por exemplo, carne e leite), de minimizar o custo de produção e as perdas de nutrientes não digeridos. Dessa forma, a compreensão dos conceitos de nutrição é essencial para formular as dietas dos animais bem como o desenvolvimento de práticas alimentares, para aumentar a eficiência da produção dos alimentos, garantindo a proteção do meio ambiente e conservando o valor nutricional dos alimentos de origem animal (Cherian, 2020).

Os alimentos para animais de exploração, como bovinos ou ruminantes, aves, peixes e camarões, são frequentemente denominados de alimentos para animais. Os alimentos para os animais ruminantes apresentam classificações distintas conforme a sua composição e o uso, tais como: silagens (por exemplo, silagens de milho, erva e luzerna); forragens secas (por exemplo, feno, pastagens secas, folhas e caules); forragens volumosas (por exemplo, feno, palha e cascas com mais de 18% de fibra bruta); pastagens verdes; alimentos proteicos (por exemplo, farelo de soja, farinha de peixe e farinha de carne e ossos) e, por último, alimentos energéticos (por exemplo, milho, trigo, cevada e arroz). Contudo, alguns suplementos vitamínicos e minerais e aditivos podem ser incluídos em rações na preparação das dietas dos animais (Wu, 2018).

As forragens representam uma percentagem significativa na alimentação de animais ruminantes, e podem ser utilizadas em verde ou conservadas. O momento da colheita do mesmo é realizado antes da maturação completa da cultura forrageira, e os locais da produção das culturas diferem dos locais de utilização das forragens para a alimentação. As forragens mais utilizadas na alimentação de animais ruminantes em explorações, com o objetivo da produção de leite, são as silagens de milho e de erva (Lage, 2017), tornando-se essencial as informações sobre a sua composição química para equilibrar corretamente os nutrientes da dieta (Berzaghi et al., 2001). A silagem é o resultado de um processo de fermentação láctica natural de culturas com alto teor de

humidade, na qual pretendem ser conservadas por vários meses. Este processo, ensilagem, tem como primeiro objetivo a criação de um ambiente anaeróbio, evitando o contacto do oxigénio com as forragens, de forma a conservar este alimento, sendo possível alimentar os animais por um longo período de tempo, pois caso contrário torna-se não comestível e tóxico. Por outro lado, o processo de fermentação leva à formação de ácidos, essencialmente de ácido láctico, que provocam a diminuição do pH e que por sua vez inibe a atividade de microrganismos indesejáveis. Os fungos aeróbios e bactérias são microrganismos dominantes em forragens verdes, mas com o desenvolvimento de condições anaeróbias no silo estes microrganismos vão sendo substituídos por bactérias capazes de promover o seu crescimento na ausência de oxigénio. Após o corte da cultura e durante os primeiros estágios do processo de ensilagem ocorre mudanças de atividade enzimática presente nos tecidos vegetais, ou seja, sucede-se processos de respiração e de proteólise que influenciaram o valor nutritivo do produto final. Na respiração há a degradação oxidativa de compostos orgânicos, essencialmente hexoses, onde basicamente toda a energia é convertida em calor. Esse calor ao ficar aprisionado no silo, que por sua vez fica retido na massa da forragem, causando um aumento de temperatura. A respiração promove uma diminuição de carboidratos solúveis, diminuindo assim o valor nutritivo. Um método simples de mitigar este acontecimento é promover as condições anaeróbicas no silo o mais rapidamente possível. A proteólise é um processo que ocorre após a colheita e tem como efeito a diminuição do teor de proteína, uma vez que leva à hidrólise das ligações peptídicas. No entanto, a atividade proteolítica continua mesmo durante a ensilagem, apesar de diminuir a atividade com a diminuição do pH, fazendo com que os produtos resultantes sejam aminoácidos e péptidos de comprimento de cadeia variáveis. Verifica-se também uma diminuição de aminoácidos nas silagens, mas isso deve-se maioritariamente pela atividade microbiana (McDonald et al., 2011).

A silagem de milho é a forragem base dos regimes alimentares utilizadas em explorações leiteiras, durante grande parte do ano. Esta silagem, facilmente fermentada e de pH igual ou abaixo de 4, é considerada um bom alimento para os animais ruminantes, pois apresenta-se como uma excelente fonte de fibras digeríveis. No entanto, este tipo de alimento depara-se com vários fatores que influenciam o valor nutritivo, nomeadamente: o estado de maturação, altura do corte, o clima, variedade das plantas, e a forma de execução do processo de conservação (Lage, 2017). Destes fatores, o estado de maturação no momento de corte da cultura é o que mais terá impacto no valor nutritivo da silagem de milho, sendo recomendado aos agricultores a medição de matéria seca da planta para a determinação do momento ideal de corte (Lage, 2019).

A silagem de erva é outro tipo de forragem também bastante utilizado por explorações leiteiras, sendo considerado um dos recursos alimentares essenciais durante o período de cultura de milho. Este alimento para além de ser rico em fibras bastantes digeríveis, também é uma boa fonte de proteína para os animais. Contudo, a silagem de erva é um alimento muito instável a nível de qualidade da forragem, pois apresentam vários fatores que condicionam o seu valor nutritivo, podendo chegar ao ponto de serem excluídas na dieta. Dessa forma, a qualidade das silagens de erva é ainda considerada fraca, mas com tendência a sofrer melhorias a nível de valor nutritivo (Lage, 2018).

Antes da década de 1960, os animais ruminantes eram alimentados com base em forragens, mas devido ao aumento da produção de leite por animal tornou-se necessário a adição de mais nutrientes nas suas alimentações. Assim desenvolveu-se a invenção da mistura unifeed (TMR, *total mixed ration*), um sistema de ração completa que é constituída principalmente por forragens, grãos e concentrados de matérias-primas. Dessa maneira, pretende-se garantir que a alimentação das vacas seja uniforme e nutricionalmente completa. Dependendo do estado de fisiológico dos animais (por exemplo crescimento, produção de leite ou gestação) as rações podem ser reformuladas e/ou suplementadas conforme os objetivos pretendidos. As TMR com base em silagens permitem diluir e mascarar alguns sabores de ingredientes desagradáveis, apresentando assim menos restrições e mais flexibilidade de obter dietas de menores custos. Com uma alimentação TMR existe uma menor probabilidade de distúrbios digestivos e situações de ausência de alimentação aos animais relativamente aos concentrados e às forragens que são fornecidas separadamente. Isto reflete-se pelo facto de a alimentação ser considerada nutricionalmente mais equilibrada (Schingoethe, 2017).

Resumindo, a alimentação é um dos fatores mais importantes na criação de animais, pois para além de esta representar uma grande percentagem dos custos, posteriormente interfere na qualidade dos produtos gerados. Portanto, a avaliação nutricional dos alimentos disponibilizados aos animais é de extrema importância para garantir uma dieta equilibrada, de forma melhorar a longevidade, a produtividade, a saúde e o bem-estar animal (Cherian, 2020; Mcgrath et al., 2018).

2.2 Composição química dos alimentos

A nutrição animal é uma ciência de preparação e formulação de alimentos atendendo às necessidades dos animais em diferentes fases de crescimento ou fases de vida. Logo, a ação dos nutricionistas é conhecer quais os componentes nutricionais das

rações, das matérias-primas, e dos ingredientes utilizados nas formulações das rações. Dessa forma, torna-se possível os produtores otimizarem as necessidades nutricionais dos animais tendo em conta os custos das formulações alimentares (Cherian, 2020).

Normalmente, a obtenção de informações acerca de perfis nutritivos de alimentos leva a que sejam realizados diversos métodos de análises laboratoriais. No entanto, para além desses procedimentos requererem o uso de vários solventes, eles podem não representar verdadeiramente as capacidades de os animais utilizá-los com eficiência. O sistema de Weende foi concebido há mais de 100 anos por dois cientistas alemães, Wilhelm Henneberg e Friedrich Stohmann, de forma a obter informações extensas sobre a composição química dos alimentos. Primeiramente, o sistema recorre à eliminação de água nos alimentos a serem analisadas e depois procede-se à determinação de cinco frações presentes na matéria seca (MS) obtida: cinzas, proteína bruta (PB), gordura bruta (GB), fibra bruta (FB) e hidratos de carbono (Cherian, 2020; McDonald et al., 2011; Wu, 2018). Na tabela 1 está estruturado o sistema de análise proximal, onde é possível visualizar as frações a determinar, os componentes maioritários, os procedimentos utilizados de forma resumida e os métodos oficiais aprovados pela AOAC (*Association of official analytical chemists*) usados para análises de alimentos para animais.

Tabela 1. Sistema de Weende: frações, componentes majoritários, procedimentos e métodos oficiais (AOAC, 2005; Cherian, 2020; McDonald et al., 2011; Wu, 2018).

Frações	Componentes majoritários	Procedimento	Método oficial
Humidade (MS por diferença)	Água e compostos voláteis (100 % - H ₂ O = MS %)	A amostra é submetida a uma temperatura acima do ponto de ebulição da água até obtenção de um peso constante. A perda de peso corresponde à quantidade de água presente na amostra. No entanto, este procedimento em silagens pode levar a perdas significativas de compostos voláteis.	Método AOAC 934,01 (gravimétrico)
Cinza (matéria mineral)	Elementos minerais	A amostra é submetida a temperaturas entre 500 °C e os 600 °C durante 2 a 4 horas, de forma a queimar todos os nutrientes orgânicos.	Método AOAC 942,05 (incineração)
Proteína Bruta (As proteínas contêm em média 16% de azoto)	Proteínas, aminos, aminos e ácidos nucleicos	Determinação do azoto presente na amostra pelo método de Kjeldahl, posteriormente a quantidade de azoto é convertido em proteína bruta pela multiplicação do fator 6,25.	Método AOAC 954,01 (digestão)
Gordura Bruta	Lípidos, vitaminas lipossolúveis, pigmentos	Extração contínua com éter de petróleo por um tempo definido.	Método AOAC 2003,06 (extração, gravimétrico)
Fibra Bruta	Celulose, hemicelulose, lenhina e pectina	A amostra pode sofrer ou não um processo de desengorduramento conforme a sua composição e posteriormente a porção orgânica do resíduo é sujeita a uma digestão com recurso a uma solução ácido e uma solução alcalina.	Método AOAC 978,10 (incineração, gravimétrico)
Hidratos de Carbono	Amido, açúcares e porções variáveis de elementos presentes na fibra bruta	Determinação por diferença de um todo com a soma de todas as outras frações.	Determinação de amido: Método AOAC 2014,10 (enzimático colorimétrico)

Com o avançar dos anos, novas técnicas analíticas foram desenvolvidas e algumas técnicas foram modificadas. No entanto, o sistema de análise proximal continua ainda a referência para a declaração legal da composição dos alimentos na Europa (McDonald et al., 2011).

2.2.1 Matéria Seca

Nos laboratórios de análises nutricionais um dos procedimentos mais comumente realizado é a determinação da MS. Os alimentos vegetais podem variar muito no teor de água, por isso é essencial conhecer a quantidade de água para permitir comparações entre diferentes tipos de alimentos. A determinação de MS é efetuada pela diferença do teor de água ($MS\% = 100\% - H_2O$), ou então pela diferença de peso do alimento em fresco com o alimento seco ($((\text{Peso fresco} - \text{Peso seco}) \times 100\% = MS\%)$). Dessa forma, é possível deduzir que quanto maior a percentagem de MS menor será o teor de humidade presente nos alimentos (Cherian, 2020).

Normalmente, a maior parte das rações (alimentos compostos, matérias-primas e forragens desidratadas) apresentam cerca de 90% de MS, porém a maior parte das forragens frescas ou conservadas como as silagens apresentam uma MS inferior. Particularizando as silagens de milho, este apresenta uma MS cerca de 30% a 35% de MS, pois esse intervalo de percentagem corresponde ao momento ideal de corte da cultura de milho, tanto no ponto de vista produtivo como da qualidade da forragem. Um elevado teor em MS em plantações de milho ($> 35\%$ de MS) representa um aumento do peso de espiga e do grão, que consequentemente é contrabalançado pela senescência da planta, ou seja, progride numa diminuição da produtividade da planta. Quanto à qualidade, verifica-se que com a maturidade da planta existe uma diminuição da digestibilidade de MS bem como das paredes celulares da planta. No entanto, essa diminuição é compensada com um aumento de amido até uma certa percentagem de MS (FEDNA, 2016; Lage, 2019).

2.2.2 Cinza Total

A cinza total é o resíduo formado por constituintes inorgânicos dos alimentos após a incineração de todos os nutrientes orgânicos. O seu componente principal é a sílica, no entanto, ela pode conter minerais derivados de compostos orgânicos, como o enxofre e fósforo das proteínas. O uso de elevadas temperaturas pode causar a perda de alguns elementos voláteis, tais como cloretos, zinco, selênio, iodo, entre outros. Ou seja, o teor de

cinza analisado não representa verdadeiramente os constituintes inorgânicos no alimento, seja na forma quantitativa ou qualitativa. Nutricionalmente, os animais ruminantes não necessitam da cinza em si, mas precisam dos elementos minerais individuais que ela contém. Por esse motivo, o parâmetro da determinação de cinza nos alimentos não é considerado de grande interesse (Cherian, 2020; McDonald et al., 2011).

Embora, a cinza seja um parâmetro um pouco desprezado, a sua análise e quantificação podem dar indicação sobre eventuais contaminações com terra. Isto geralmente acontece quando os valores de cinza são superiores a 4% de MS em silagens de milho (Cherian, 2020; Lage, 2017).

2.2.3 Proteínas

As proteínas são compostos orgânicos complexos de alto peso molecular constituídas por carbono, hidrogênio e oxigênio, mas também contém azoto e geralmente enxofre. Estas são formadas através de um processo sequencial linear de resíduos de aminoácidos que se ligam entre si através de ligações covalentes (Cherian, 2020; McDonald et al., 2011).

As proteínas são vitais para a existência de vida e são os principais componentes estruturais dos tecidos animais, estando presentes em todas as células. Para além disso, elas ainda apresentam muitas outras funções que são essenciais para o funcionamento dos animais, nomeadamente: atividades bioquímicas, imunológicas, transporte, atividades reguladoras e quando necessário podem ser utilizadas para fornecer energia. Dessa forma, é extremamente importante a síntese de proteínas para a manutenção da vida. Embora o corpo animal consiga sintetizar alguns aminoácidos, existem outros que não consegue sintetizar em quantidades suficientes, ou então simplesmente não sintetiza. Nesses casos, são designados como aminoácidos essenciais, sendo fundamental o fornecimento de aminoácidos aos animais através de dietas alimentares. No entanto, as necessidades de aminoácidos essenciais variam de animais para animais, e no caso de animais ruminantes geralmente não acontece, pois são sintetizados pelos microrganismos do rúmen. As necessidades proteicas em animais variam conforme o estágio de vida, apresentando-se mais altas em alturas de crescimento, gestação e lactação (Cherian, 2020).

O objetivo da suplementação proteica em animais ruminantes é simplesmente fornecer uma fonte de azoto para a população microbiana do rúmen, para que os microrganismos possam produzir as suas próprias proteínas microbianas. As proteínas adquiridas na dieta são denominadas como proteínas degradáveis e não degradáveis no

rúmen, uma vez que a digestão de proteínas pode ser realizada em duas fases: (1) digestão no retículo rúmen e/ou (2) digestão no abomaso e intestino delgado. As proteínas não degradáveis são utilizadas logo na primeira fase para a produção de proteína microbiana. No entanto, no intestino delgado do ruminante, segunda fase, podem chegar proteínas não degradáveis e proteínas microbianas para serem digeridas e absorvidas. Ou seja, as necessidades de aminoácidos dos animais ruminantes são atendidas tanto por proteínas não degradáveis e proteínas microbianas sintetizadas a partir de proteínas degradáveis da dieta (Cherian, 2020; Wu, 2018).

As silagens de milho utilizadas para alimentação de animais ruminantes são normalmente classificadas de baixo valor proteico, uma vez que apresentam valores de PB entre os 6% e 8% de MS (FEDNA, 2016; Lage, 2017). No entanto, existem fatores como aumento da densidade das sementeiras e disponibilidade de água abundante que permitem uma maior captação de azoto do solo por parte das plantas. Estes fatores têm uma grande influência na qualidade nutricional da forragem, incluindo o teor de PB (Lage & Salgueiro, 2019).

2.2.4 Gordura

As gorduras são consideradas excelentes fontes de energia e necessárias para a sobrevivência dos animais. Por outro lado, apresentam um papel fundamental no fornecimento de ácidos gordos essenciais, uma vez que são a única fonte existente para os animais. Em animais ruminantes, os ácidos gordos que devem ser incluídos na dieta são o ácido linoleico e ácido α -linolénico, devido à incapacidade de formação de ligações duplas nas posições C3 e C6. As gorduras também podem fornecer vitaminas lipossolúveis, mas para os ruminantes esse papel torna-se indispensável, pois os alimentos utilizados normalmente são suplementados com vitaminas. A adição de gorduras aos alimentos leva a uma melhoria da qualidade, a uma redução da poeira e da separação de partículas durante o processamento, aumenta a palatabilidade e a digestibilidade do alimento (Cherian, 2020; McDonald et al., 2011).

A GB normalmente contribui com uma pequena percentagem em silagens de milho, por volta de 2% a 3% de MS (Judy et al., 2019). No entanto, a GB é bastante importante para análise de alimentos compostos, concentrados e TMR.

2.2.5 Fibras

A fibra tem um papel essencial na nutrição dos animais ruminantes, pois esta é uma fonte de energia e potencia os processos fermentativos (Júnior et al., 2007). A análise de FB pretende estimar a fração dos alimentos que não são digeríveis ou frações que são fermentadas posteriormente por microrganismos no intestino de animais ruminantes. A FB abrange um conjunto de carboidratos insolúveis associados à parede celular das plantas e que são resistentes à ação enzimática digestiva, ou seja, é constituída por celulose, hemicelulose, lenhina e pectinas (Cherian, 2020).

No entanto, a forma como a FB é analisada pelo sistema de análise proximal é bastante criticada, pois não estima de forma real a parte insolúvel do alimento (Wu, 2018). O motivo de tal contestação deve-se à utilização de ácido e de base que solubilizam grande parte da hemicelulose, lenhina e pectinas e a uma perda significativa de celulose (Cherian, 2020). Em média, a extração sequencial ácida e básica leva a uma remoção parcial de 80% de hemicelulose, de 50% a 90% de lenhina, e a uma recuperação de celulose entre os 50% a 80%. As pectinas são solubilizadas levando a que ocorra uma remoção total da mesma (Möller, 2014). No entanto, Peter Van Soest desenvolveu um trabalho na década de 60 onde melhorou os métodos de análise da fibra, onde propôs dividir a fração da fibra em fibra de detergente neutro (NDF), fibra de detergente ácido (ADF). A NDF contém na sua constituição componentes da parede celular insolúveis em detergente neutro, nomeadamente hemicelulose, a lenhina e a celulose, embora também possa conter cutina e proteínas ligadas à fibra. A NDF ainda pode ser mais fracionada utilizando uma extração com detergente ácido que solubiliza a hemicelulose e as proteínas ligadas à fibra, dando origem ao ADF. Assim, a NDF é o resultado da soma de ADF mais hemicelulose. A determinação da fração em detergente ácido é particularmente útil para em forragens, uma vez que existe uma boa correlação entre ela a digestibilidade do alimento (Cherian, 2020; McDonald et al., 2011; Wu, 2018). No entanto, ainda existe um terceiro fracionamento que promove a dissolução da celulose em ácido sulfúrico, restando apenas a lenhina como constituinte, dando origem à lenhina de detergente ácido (ADL) (Möller, 2014). Para os animais ruminante a lenhina é indigestível, mas as hemiceluloses e celulosas são fermentadas pelos microrganismos do rúmen (Cherian, 2020).

O fracionamento da fibra teve um grande impacto na formulação das dietas para animais ruminantes, uma vez que estes conseguem utilizar uma parte da fibra insolúvel do alimento, ou seja, a NDF (Júnior et al., 2007).

Os teores de fibras em silagens de milho são bastante variáveis, contudo estima-se um valor de FB entre 20% a 30% da MS para silagens de boa qualidade (FEDNA, 2016).

Os níveis de NDF são considerados bons na faixa dos 40% de MS, mas para o ADF os valores rodeiam os 30% de MS (Garcia et al., 2020; Lage, 2017). Relativamente ao ADL, uma vez que este apenas representa a fibra indigerida, lenhina, o valor em MS é relativamente baixo, cerca de 3% (FEDNA, 2016).

2.2.6 Hidratos de Carbono

Os hidratos de carbono são os principais constituintes dos tecidos vegetais, representando cerca de 60% a 90% de MS. Estes compostos podem estar presentes sobre a forma de amido, açúcares ou componentes da parede celular das plantas. Para além disso, os hidratos obtidos das dietas alimentares são uma importante fonte de energia para os animais, uma vez que estes correspondem à principal fonte de produção de energia para as células. A digestão dos hidratos de carbono em animais ruminantes ocorre através da fermentação no rúmen para produção de energia, que posteriormente serão utilizados por bactérias para o seu crescimento e síntese de proteínas. Os animais ruminantes obtêm energia através da absorção de ácidos gordos de cadeia curta que são produtos resultantes da fermentação no rúmen (Cherian, 2020). O teor em hidratos de carbono é calculado pela diferença da FB, uma vez que parte das fibras encontram-se nos hidratos de forma a serem contabilizados, uma vez que se dissolvem durante o processo de extração.

A determinação do amido é um parâmetro frequentemente analisado em silagens de milho, pois é o hidrato de carbono usado como fonte de reserva deste tipo de alimento. Os valores estimados para este parâmetro circundam os 30% de MS (FEDNA, 2016; Lage, 2017).

A cultura de milho apresenta uma ótima aptidão para o processo de ensilagem, uma vez que esta não carece de hidratos de carbono. Do ponto de vista nutricional, este tipo de silagem é um bom alimento, uma vez que apresenta um alto valor energético e com baixo teor em proteínas e minerais (FEDNA, 2016; Lage, 2017).

2.3 Espectroscopia NIR

2.3.1 Breve história

A radiação referente à região do infravermelho próximo (NIR) foi descoberta em 1800 pelo astrónomo William Herschel, enquanto estudava a contribuição das cores provenientes da luz solar no aumento da temperatura das substâncias expostas (Pasquini,

2003). Durante a sua experiência, Herschel verificou que a temperatura continuava a aumentar quando posicionava o termómetro para além da zona do vermelho, acabando por demonstrar a existência de energia posterior ao espectro visível (Agelet & Hurburgh, 2010). Esta descoberta permitiu construir a primeira parte do espectro eletromagnético não visível a ser identificada (Pasquini, 2003).

Contudo, a espectroscopia NIR apresentou um longo período de dormência, nomeadamente entre os anos 1800 a 1950, devido à sua falta de seletividade, uma vez que os espectros de absorção da região infravermelho próximo eram constituídos por inúmeras bandas de absorção fortemente sobrepostas, tornando-se de difícil interpretação. Desse modo, esta espectroscopia ficou para segundo plano relativamente a outros métodos analíticos que apresentavam uma maior capacidade de fornecer resultados mais inequívocos (Manley et al., 2008; Pasquini, 2003). No entanto, alguns trabalhos pioneiros foram desenvolvidos durante esse período, tais como: a utilização de pratos fotográficos com sensibilidade NIR, por Abney e Festing, para o registo dos espectros de compostos orgânicos, reconhecendo as ligações atômica e a importância da ligação do átomo de hidrogénio; a construção do primeiro espectrómetro que posteriormente permitiu com que Coblentz descobrisse que cada composto apresentava um espectro único; o uso da espectrofotometria NIR para a determinação de água em gelatina e de hidrocarbonetos em frações de petróleo; entre outras (Aenugu et al., 2011; Pasquini, 2003, 2018a). Porém, foram os trabalhos na área agroalimentar de Karl Norris, no Departamento de Agricultura nos Estados Unidos, que promoveram o reconhecimento do potencial desta técnica analítica, despertando interesse da comunidade científica e da indústria (Aenugu et al., 2011; Heinz W. Siesler, 2008; Manley et al., 2008; Pasquini, 2003). Nas décadas de 1970 e 1980, Thomas Hirschfeld desenvolveu trabalhos experimentais relacionados com aspetos fundamentais da espectroscopia NIR e sua instrumentação, permitindo assim aumentar o conhecimento relativo à região do infravermelho próximo (Heinz W. Siesler, 2008; Pasquini, 2018b, 2018a).

2.3.2 Importância na Indústria

Os trabalhos experimentais desenvolvidos por Karl Norris demonstraram que para ultrapassar a falta de seletividade relativa à informação expressa nos espectros NIR era necessário recorrer a ferramentas matemáticas, nomeadamente técnicas quimiométricas de tratamento multivariados de dados. Desse modo, a tecnologia NIR sofreu um rápido desenvolvimento tendo como base os fundamentos da espectroscopia vibracional, a instrumentação de espectrometria NIR e a quimiometria (Pasquini, 2003, 2018b).

A introdução da espectroscopia NIR numa envolvente industrial deve-se a Phil Williams, uma vez que promoveu a utilização de uma técnica analítica baseada em espectroscopia NIR, técnica limpa e rápida, em vez do método Kjeldahl para a determinação de proteínas (Williams, 2015). O potencial desta técnica foi evoluindo ao longo do tempo, tornando-se numa ferramenta bastante utilizada a nível industrial. Desde então, tem-se recorrido a esta técnica para obtenção de resultados rápidos de análise, para o controlo de qualidade e monitorização em diversos setores produtivos, assim como a nível das publicações de normas oficiais de forma a obter guias de desenvolvimento, validação e manutenção de métodos (Pasquini, 2018b; Porep et al., 2015). Devido ao avanço da tecnologia, a indústria agroalimentar utiliza a espectroscopia NIR no controlo da composição química das matérias-primas, produtos intermédios e produtos alimentares, permitindo o aumento da eficiência na produção e na deteção de problemas antecipadamente, de forma a manter a autenticidade do produto final (Manley et al., 2008).

As inovações mais recentes relativas a esta tecnologia são: as fibras óticas, a miniaturização dos espectrofotómetros, as imagens hiperespectrais NIR e o uso em técnicas analíticas de processo (PAT) (Pasquini, 2018b).

2.3.3 Princípios Teóricos

A espectroscopia é uma ciência que permite obter, medir e interpretar os espectros resultantes da interação da radiação eletromagnética com a matéria (Penner, 2017). Dos diversos fenómenos decorrentes dessa interação, a absorção da radiação pelos constituintes de uma amostra gera espectros de absorção com informações analíticas qualitativas e quantitativas (Pasquini, 2018b).

No espectro eletromagnético, a região do infravermelho (IR) encontra-se entre as gamas do visível e as micro-ondas, com a respetiva faixa de comprimento de onda de 780 nm a 100000 nm. Contudo, esta região pode ainda ser dividida em três sub-regiões distintas, nomeadamente: infravermelho próximo que se encontra compreendido entre os 780 nm a 2500 nm; infravermelho médio (MIR), compreendido entre os 2500 nm a 40000 nm e o infravermelho distante (FIR), compreendido entre os 40000-100000 nm (figura 1).



Figura 1. Gama de radiação eletromagnética. (Adaptado de Aenugu et al., 2011)

A espectroscopia NIR é designada como uma espectroscopia vibracional, pois a energia transportada pela radiação NIR é relativamente baixa para promover, na maioria dos casos, o aumento da energia eletrónica das moléculas ou átomos; porém excede a dos níveis rotacionais, não sendo concebida devido às restrições quânticas (figura 2) (Manley et al., 2008; Pasquini, 2003, 2018b; Porep et al., 2015). Ou seja, o espectro NIR é concebido pela transformação da energia da radiação fornecida a energia mecânica associada ao movimento dos átomos numa molécula (Pasquini, 2003).

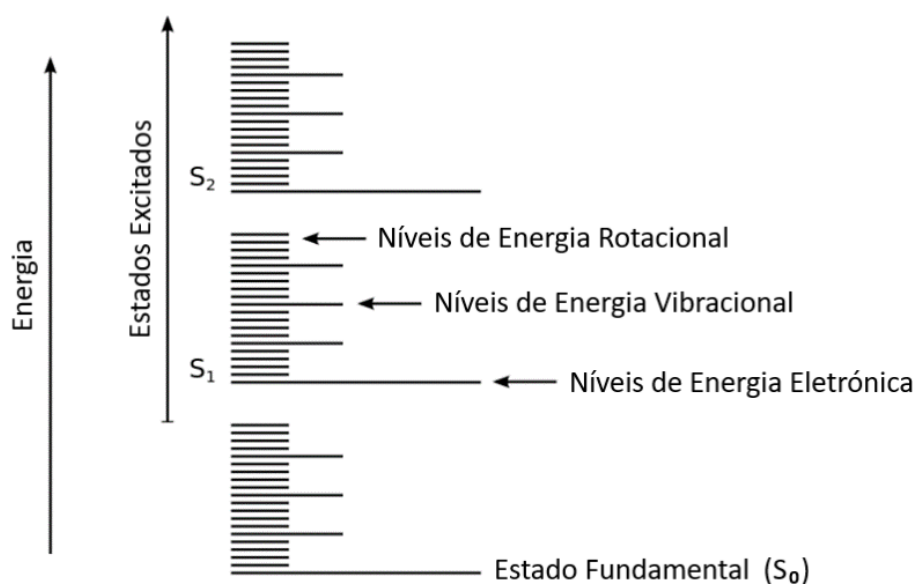


Figura 2. Diagrama de níveis de energia molecular. S_0 : Estado Fundamental; S_1 : 1º Estado Excitado; S_2 : 2º Estado Excitado. (Adaptado de Penner, 2017)

A espectroscopia NIR é a técnica analítica mais utilizada nas indústrias de alimentação animal, com o intuito de ter um controlo da qualidade de matérias-primas, produtos intermédios e produtos finais, estimando a composição dos alimentos (McDonald et al., 2011). Os espectros NIR dos constituintes dos alimentos apresentam uma alta

complexidade devido às largas bandas de absorção altamente sobrepostas e fracas, como já referido anteriormente, e às bandas de combinações vibracionais (Manley et al., 2008; Porep et al., 2015; Qu et al., 2015). A base desta metodologia reside na absorção da energia irradiada por grupos funcionais que contêm hidrogénio ligados a átomos mais pesados, sendo que nos espectros obtidos são principalmente as ligações C – H, O – H, N – H e S – H das moléculas presentes que dão origem as bandas de absorção. No entanto, as ligações químicas de alta energia, como C = C, C $\equiv$ C e C = O, também podem ser detetadas (McDonald et al., 2011; Pasquini, 2018a; Porep et al., 2015). O sistema molecular vibracional mais realista do sucedido é o modelo anarmónico quântico, onde os níveis energéticos vibracionais não são espaçados da mesma forma, mas diminuem a diferença entre si à medida que se consideram os níveis de energia vibracional mais altos. A energia vibracional de uma molécula pode promover a transição direta do estado fundamental para outros níveis superiores. A estas transições resultantes da absorção da radiação eletromagnética designa-se de *overtone*. Quando ocorre uma transição do estado fundamental da molécula para o segundo estado excitado originam-se bandas fracas denominadas de 1º *Overtone*, com a transição do estado fundamental para o terceiro estado excitado originam-se bandas fracas denominadas de 2º *Overtone* e da transição do estado fundamental para o quarto ou quinto estado excitado originam-se bandas correspondentes ao 3º e ao 4º *Overtone* (figura 3) (Pasquini, 2003, 2018b).

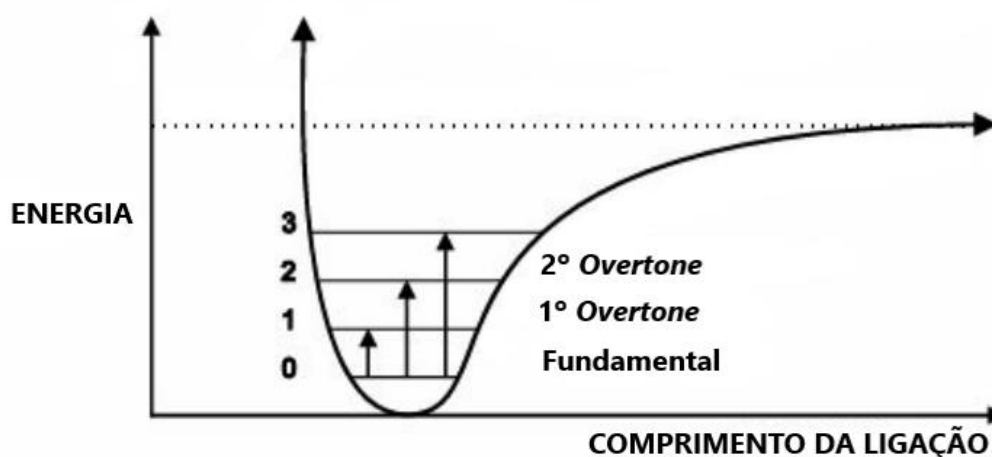


Figura 3. Modelo anarmónico - transições vibracionais. (Adaptado de Aenugu et al., 2011)

Um princípio fundamental da tecnologia NIR é a absorção da energia por uma amostra quando esta é irradiada. Uma amostra ao ser exposta a uma radiação com uma determinada frequência pode promover excitação das moléculas para níveis vibracionais superiores, uma vez que a energia envolvida é a energia exata entre dois níveis vibracionais,

de *overtones*, ou das combinações de duas ou mais vibrações. No entanto, a correspondência da energia irradiada com a dos níveis vibracionais não é suficiente para a absorção da radiação, é necessário que ocorra uma mudança no momento dipolo da molécula (Pasquini, 2003, 2018a). Dessa forma, apenas algumas frequências de uma determinada faixa de comprimento de ondas algumas frequências serão absorvidas. A intensidade de uma banda de absorção está relacionada com a magnitude da mudança do dipolo durante uma vibração com o seu grau de anarmonicidade (Pasquini, 2003).

Após a detecção da absorção da energia, é possível descrever a composição química do produto alimentício desconhecido, mesmo sem conhecer a estrutura molecular específica responsável pela absorção da energia vibracional. Portanto, apesar da dificuldade de interpretação dos espectros NIR, estes apresentam bastantes informações químicas e físicas acerca das moléculas orgânicas presentes na composição de um produto alimentar. Para a determinação do verdadeiro valor da espectroscopia NIR recorre-se a uma ferramenta analítica baseada em tratamentos estatísticos dos dados espectrais (Manley et al., 2008).

2.3.4 Instrumentação

Durante a última década do século XX, a espectroscopia NIR obteve grandes avanços tanto na instrumentação como na aquisição de dados. Esta evolução permitiu com que esta tecnologia estabeleça-se um lugar bem definido perante os outros métodos analíticos convencionais já estabelecidos, particularmente, métodos de referência (Manley et al., 2008).

A nível da instrumentação NIR existe uma vasta gama de variedades de configurações possíveis, porém, todos os espectrofotómetros NIR apresentam cinco componentes essenciais, tais como: o compartimento da amostra, a fonte de radiação, o sistema de seleção do comprimento de onda, o detetor de radiação e o processador de sinal (Agelet & Hurburgh, 2010; Manley et al., 2008). A variedade de apresentação de amostras é um dos pontos fortes da espectroscopia NIR, pois permite com que seja possível aplicar uma ampla gama de diferentes tipos de amostras (líquidos, pastas, em pó, amostras sólidas, etc.) a esta tecnologia, no entanto é importante assegurar um nível de espalhamento constante para cada amostra a fim de minimizar a dispersão da luz. (Manley et al., 2008).

Relativamente aos constituintes dos espectrofotómetros verifica-se que a nível dos compartimentos das amostras são usados, normalmente em laboratórios, copos de

amostras abertas constituídos por sílica ou por quartzo, material transparente à radiação infravermelho próximo. Alusivo às fontes de radiação NIR, estas podem ser classificadas como térmicas e não térmicas. A fonte mais utilizada é a lâmpada halógena de tungsténio, fonte térmica, que apresenta uma faixa de emissão de comprimento de onda de 320 nm a 2500 nm. As fontes não térmicas consistem em lâmpadas de descarga que emitem bandas de radiação mais estreitas, contudo tornam-se mais eficiente comparativamente às térmicas, porque a maior parte da energia consumida aparece como radiação emitida em uma faixa estreita de comprimentos de onda, reduzindo o consumo de energia. Das fontes de radiação não térmica, os díodos emissores de Luz (LED) exibem um baixo consumo de energia, pequeno porte e uma longa vida útil, sendo mais apropriados para instrumentos miniaturizados e para aplicações de triagens específicas fora de um laboratório; os lasers de díodo, também designados por díodos emissores de luz superluminiscente (SLED), são as fontes de luz não térmicas mais inovadoras com uma excelente resolução de comprimento de onda, menor frequência de ruído e economicamente mais baratos. Relativamente ao terceiro constituinte referido acima, pode-se distinguir os espectrofotómetros NIR com base na seleção do comprimento de onda, ou seja, comprimento de onda discreto ou espectro contínuo. Um espectrofotómetro de comprimento de onda discreto irradia uma amostra com apenas alguns comprimentos de onda selecionados utilizando filtros. Os instrumentos de filtro são projetados para aplicações particulares em regiões espectrais específicas, recorrendo ao uso de 6 a 20 filtros de interferência para selecionar comprimentos de onda que serão absorvidos por certas espécies moleculares em uma aplicação específica. Os instrumentos NIR de espectro contínuo podem incluir uma rede de difração, ser do tipo de matriz de díodo, filtro sintonizável acústico-ótico (AOTF) ou infravermelho próximo com transformada de Fourier (FT-NIR). Esses instrumentos são muito mais flexíveis do que os instrumentos de comprimento de onda discreto e podem ser usados numa ampla variedade de medições. Esses são geralmente os instrumentos mais fáceis de usar e mais baratos. No que diz respeito aos detetores, estes podem-se dividir em detetores de canal único ou multicanal. Relativo aos detetores de canal único podem-se enumerar os de silício, que são rápidos, silenciosos, pequenos e altamente sensíveis da região visível até 1100 nm; os de sulfureto de chumbo (PbS), que são mais lentos, mas muito utilizados pois apresentam uma alta durabilidade, baixo preço e são sensíveis desde 1100 nm até 2500 nm; e por fim o índio-gálio-arsénio (InGaAs), que são os mais caros, pois combinam as características de velocidade e tamanho do detetor de silicone com o alcance do comprimento de onda do detetor de PbS. Os detetores multicanais são compostos por matrizes de díodo, onde vários elementos de deteção estão dispostos em linhas ou então por dispositivos de carga acoplada (CCDs). Estes componentes transformam a energia da luz incidente em sinal

elétrico analógico. O sinal elétrico é então amplificado e transformado em digital, onde posteriormente é processado por um computador. O desempenho de um detetor de infravermelho depende de muitos fatores, incluindo velocidade de resposta, resposta espectral, limite de detecção e temperatura de operação. A seleção de detetores usados em diferentes instrumentos depende da faixa de comprimentos de onda a serem medidos. Os detetores e amplificadores são considerados as fontes mais comuns de ruído não sistemático em instrumentos, ruído aleatório. O ruído aleatório é reduzido na maioria dos instrumentos comerciais pela média de vários espectros à amostra, melhorando a relação sinal-ruído (SNR) (Agelet & Hurburgh, 2010; Lin Mengshi et al., 2009; Manley et al., 2008). Para os processadores de sinal é usual utilizar computadores ou então microprocessadores. Na tabela 2 está esquematizado, de forma sucinta, os componentes essenciais de um espectrofotómetro NIR juntamente com os respetivos tipos e materiais.

Tabela 2. Resumo dos componentes essenciais de um espectrofotômetro NIR e seus respectivos tipos e materiais.

Componentes	Tipos	Materiais
Compartimento da amostra	Copos de sílica	
	Copos de quartzos	
Fonte de radiação	Térmica	Lâmpada de tungsténio
	Não térmica	LED
		SLED
Seleção de comprimento de onda	Discreto	Filtros
		Rede de difração
	Espectro contínuo	Matrizes de díodo
		AOTF
		FT-NIR
Detetor	Canal único	Silício
		PbS
		InGaAs
	Multicanal	Matrizes de díodo
		CCDs
Processador de sinal/ computador	Computador	
	Microprocessador	

2.3.5 Métodos de medição

Em espectroscopia NIR, o modo apropriado de medição é ditado pelas características óticas das amostras. Porém, existem vários métodos para aquisição dos espectros em espectroscopia NIR, nomeadamente: refletância, transmitância, transfletância e interatância (Figura 4).

A espectroscopia de refletância, o método utilizado neste trabalho, mede a luz refletida pela superfície de uma amostra, no entanto esta é composta por duas vertentes: especular e difusa. A refletância especular, descrita pela Lei de Fresnel, contém pouca ou mesmo nenhuma informação sobre a composição da amostra devido ao seu contacto

mínimo com a amostra. O ângulo de reflexão da componente especular, que é o mesmo da luz incidente, é minimizado pelo ajuste da posição do detetor em relação à amostra. Por outro lado, a refletância difusa é a componente que permite a aquisição dos espectros por refletância, pois esta refere-se à parte do feixe incidente que atinge um certo grau de penetração no interior da amostra, que posteriormente é espalhada e absorvida em função da composição química da mesma. A radiação que não é absorvida é refletida pela amostra para o detetor (figura 4a). A medição espectral por refletância difusa é sobretudo apropriada para amostras sólidas, e foi a sua aplicação em espectroscopia NIR que permitiu o aumento da versatilidade e simplicidade da técnica em si. Na refletância difusa, a absorção da luz e o espalhamento das partículas sólidas contribuem para uma variação na intensidade do sinal, o que levou a que Kubelka e Munk fizessem um tratamento exigente do sinal neste tipo de medições. Desse tratamento resultou numa equação matemática, equação 1, onde substitui a lei de Lambert-Beer (válida apenas para materiais transparentes e homogêneos), estabelecendo uma relação linear entre a concentração (C) e a refletância difusa (R):

$$f(C) = \frac{(1 - R)^2}{2R} \quad (1)$$

onde R é a refletância obtida por:

$$R = \frac{I_R}{I_{R0}} \quad (2)$$

em que I_R é a intensidade da radiação refletida pela amostra e I_{R0} a mesma quantidade de radiação refletida por um material não absorvente em toda a faixa espectral de medição, equação 2. No entanto, a equação Kubelka-Munk (equação 1) não é muito aplicada, sendo substituída por uma função mais prática, equação 3, para desenvolver métodos analíticos, uma vez que esta não se afasta muito da previsão de Kubelka-Munk. No entanto, presume-se que, na maioria dos casos, pequenas variações na refletância obtêm-se um comportamento linear relativamente à concentração do analito.

$$f(C) = \text{Log} \frac{1}{R} \quad (3)$$

A transmitância é a designação que se dá à medição de um feixe de luz que consegue atravessar uma dada amostra, sem que esta seja absorvida ou refletida. Este método é utilizado, particularmente, para análises de amostras líquidas transparentes e que não provocam dispersão da luz. A lei de Lambert-Beer é aplicada neste contexto uma

vez que existe proporcionalidade entre a transmitância, a concentração do analito e o percurso ótico (figura 4b).

A transfletância é um método especial para a determinação da transmitância, porém esta resulta da combinação dos modos de transmitância e refletância. Este método é utilizado para análises de líquidos transparentes ou turvos, e recorre ao uso de uma célula de quartzo e uma placa refletora difusa. A radiação incidente é transmitida pela amostra, refletida pela placa e transmitida novamente pela amostra. A diferença da transfletância a uma simples transmitância está em duplicar o percurso ótico, visto que o feixe de radiação passa duas vezes pela amostra, como se pode observar na figura 4c.

A transmitância e a refletância também se podem combinar para dar origem ao método interatância. O modo de interatividade permite aumentar a probabilidade de interação entre um determinado feixe incidente e uma amostra, que, consequentemente, dará origem a um feixe emergente que contém informações mais realistas sobre a composição da amostra (figura 4d). Normalmente este método é aplicado para análises de amostras de grandes dimensões, como exemplo frutas inteiras (Manley et al., 2008; Pasquini, 2003).

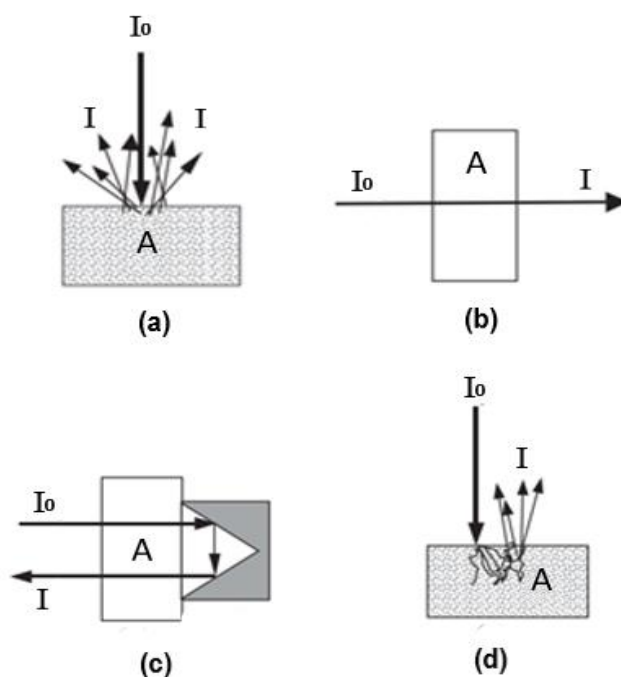


Figura 4. Métodos de medições para a aquisição de espectros em espectroscopia NIR. a) refletância; b) transmitância; c) transfletância; d) interatância. A: Amostra; I_0 : Intensidade de radiação incidente; I: Intensidade de radiação emergente (Adaptada de Pasquini, 2003).

2.3.6 Zonas Características de Absorção

Quando se promove o varrimento de uma amostra desconhecida por espectroscopia NIR, algumas questões são levantadas, tais como: que tipo de esqueleto carbónico constitui a amostra e quais os compostos que estão presentes. No entanto, esta técnica é muito útil para prever a presença de grupos funcionais e a identificação dos compostos, pois a maioria dos grupos característicos exigem bandas únicas de absorção na região infravermelho (Aenugu et al., 2011). Dessa forma, é possível observar na tabela 3 alguns exemplos de zonas típicas de absorção de grupos funcionais que se encontram presentes em alimentos para animais, nomeadamente, proteínas, gordura, fibra e amido.

Tabela 3. Exemplos de zonas típicas de absorção de compostos presentes em alimentos para animais na região NIR. (Adaptada de Shenk et al., 2008; Workman, 1996)

Zonas Típicas de absorção na região NIR			
Proteína	Gordura	Fibra	Amido
975 - 1015 nm	896-915 nm	875 - 900nm	850 - 900 nm
1480 - 1550 nm	911-935 nm	985 - 995 nm	1010 - 1030 nm
1620 – 1700 nm	1010-1025 nm	1170 nm	1450 nm
2030 - 2080 nm	1410 nm	1410 – 1420 nm	1540 nm
2148 - 2200nm	1720 nm	1417 – 1440 nm	1930 nm
1980 nm	1760 nm	1685 nm	1960 nm
2300 nm	2070 nm	1490 nm	2100 nm
2470 nm	2140 nm	1780 nm	2280 nm
2530 nm	2310 nm	1820 nm	2322 nm
---	2346 nm	2270 nm	2330 nm
---	2380 nm	2335 nm	2500 nm
---	---	2347 nm	---
---	---	2352 nm	---
---	---	2488 nm	---

A espectroscopia NIR, para além de prever e identificar os compostos mencionados anteriormente, também determina o teor de humidade, uma vez que a água apresenta bandas de absorção em comprimentos de ondas típicos, tais como: 950 – 980 nm, 1440 – 1470 nm e 1920 – 1940 nm (Workman, 1996).

Normalmente, quando se compara os espectros obtidos por transições vibracionais e transições eletrónica, chega-se à conclusão de que os espectros de infravermelho

próximo contem mais informação acerca dos compostos presentes na amostra. As bandas obtidas pela espectroscopia vibracional fornecem informações que, posteriormente, podem ser utilizadas para propósitos quantitativos e/ou qualitativos (Aenugu et al., 2011; Metrohm, 2014).

2.3.7 Fatores que afetam os espectros

A tecnologia NIR, apesar de ser um método de análise com muito prestígio a nível de avaliação de parâmetros químicos e físico-químicos, devido à sua rápida realização e preparação das amostras bem como aos instrumentos sofisticados que são utilizados, esta apresenta fatores que podem influenciar os espectros obtidos. Os artefactos espectrais podem surgir dos tamanhos das partículas variados, da distribuição desses mesmos tamanhos e da temperatura.

Os alimentos para animais apresentam na sua constituição uma ampla gama de partículas que diferem na forma, tamanho e densidade. Estas diferenças podem provocar alterações nas características de reflexão e dispersão de luz durante a análise NIR (Workman, 1996). A temperatura é um fator que influencia a intensidade das vibrações. O facto de as ligações de hidrogénio, consideradas fracas, serem umas das principais ligações que dão origem aos espectros NIR, variações na temperatura provoca consequentemente alterações nas bandas de absorção. As ligações intermoleculares, normalmente, são enfraquecidas com mudanças de temperatura, o que leva a que estas mudem a sua posição na banda de absorção de acordo com a temperatura da amostra (Cozzolino et al., 2007; Hageman et al., 2005; Ikoyi & Younge, 2020; Shah et al., 2016). Exemplo disso são as bandas de absorção da água que sofrem uma diminuição do comprimento de onda quando esta está sujeita a um aumento de temperatura (figura 5) (Segtnan et al., 2001).

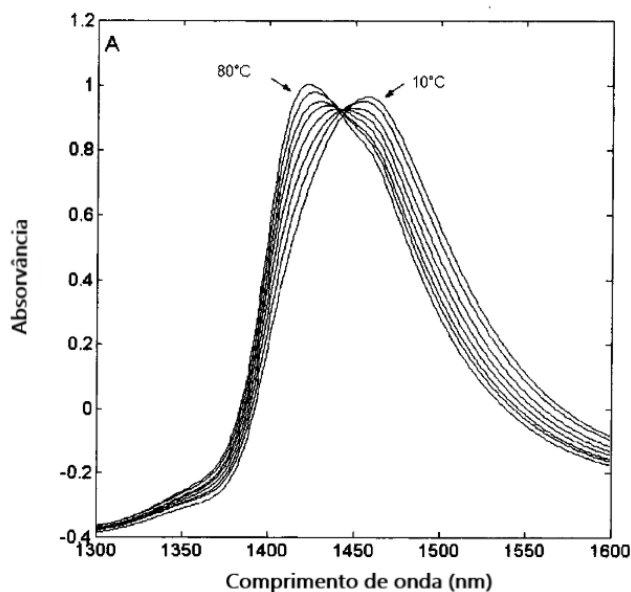


Figura 5. Espectro NIR de água sem tratamento com uma variação de temperatura. (Adaptado de Segtnan et al., 2001)

2.3.8 Vantagens e desvantagens da espectroscopia NIR

A evolução da tecnologia NIR, relativamente a outras técnicas analíticas, trouxe muitas vantagens na sua utilização. Primeiramente, pode-se enumerar o facto desta espectroscopia ser de manuseio fácil e rápido, e recorre ao uso de amostras minimamente preparadas ou amostras intactas sem qualquer tipo de pré-tratamento. O último aspeto mencionado promove a interação não convencional da radiação com a matéria, de forma a obter uma natureza não destrutiva e sem recorrer a reagentes químicos. Esta tecnologia pode ser utilizada para análises de amostras com alta humidade, líquidas, bem como sólidas, obtendo não só o registo de espectros completos, mas como também a análise de vários constituintes ou propriedades das mesmas de forma simultânea. O facto da penetração da luz NIR ser relativamente profunda permite também ter uma melhor representação das propriedades químicas da amostra (Aenugu et al., 2011; Manley et al., 2008; Nobari-Moghaddam et al., 2021; Pasquini, 2003; Wedelsbäck, 2015).

No que diz respeito às desvantagens da espectroscopia NIR, a baixa sensibilidade do sinal é considerada como uma das principais, pois esta pode limitar a determinação de compostos que estejam em baixas concentrações. Esta limitação leva à formação de espectros simplificados e restringe a informação extraída das ligações químicas fortes entre átomos leves, no entanto, o tratamento quimiométrico pode trazer informações reveladoras caso seja realizado de forma correta (Karoui & De Baerdemaeker, 2007). Outra das desvantagens desta tecnologia é o facto de ser uma técnica analítica que envolve, de forma dependente, outros métodos químicos de análise que muitas das vezes são menos

precisos e igualmente empíricos. Para além disso, a espectroscopia de infravermelho próximo requer um extenso e variado conjunto de dados para que possa obter uma calibração e validação de amostras a fim de produzir modelos de previsão robustos. A mencionar ainda que o uso e o desenvolvimento eficaz de novas calibrações exigem alguma experiência e treino (Manley et al., 2008; Wedelsbäck, 2015).

Relativamente à instrumentação NIR, pode-se encontrar tanto vantagens como desvantagens. O elevado custo inicial de aquisição de um instrumento NIR é um dos pontos negativos a apontar, uma vez que a maioria dos instrumentos são robustos, no entanto, estes podem ser utilizados tanto dentro como fora dos laboratórios. O avanço tecnológico promoveu o desenvolvimento de espectrofotómetros NIR portáteis, sendo estes instrumentos de pequeno porte e de fácil transporte, pois podem ser carregados em mochilas e montado com alguma facilidade. Este tipo de instrumento estão cada vez mais disponíveis nos dias de hoje (Aenugu et al., 2011; Manley et al., 2008; Pasquini, 2003).

2.4 Quimiometria

A quimiometria é designada como uma disciplina química que aplica métodos matemáticos e estatísticos de forma a delinear e/ou selecionar procedimentos e ensaios de medições ideais para obter o máximo de informação química, a partir de análises de dados químicos. O desenvolvimento e crescimento desta disciplina estão intimamente relacionados com os desenvolvimentos e popularização de microprocessadores, que se iniciou durante a década de 1970, resultando em equipamentos analíticos que permitiram a aquisição da informação química de uma amostra em tempo muito reduzido (Brereton, 2018; Otto, 2017). As análises quantitativas realizadas por métodos de referência, que por vezes são demoradas e pouco precisas, estão cada vez mais serem substituídas por técnicas instrumentais, nomeadamente espectroscopia de infravermelho. Com a sofisticação crescente dessas técnicas, que foi sendo acompanhada com a evolução dos microprocessadores e microcomputadores, tornou-se imprescindível o tratamento de dados mais complexos do ponto de vista matemáticos e estatísticos. Estes tratamentos permitiram relacionar os sinais obtidos dos espectros com os resultados desejados.

Particularizando o caso da espectroscopia NIR, os procedimentos quimiométricos são aplicados aos espectros para extrair informações relevantes, uma vez que estes são influenciados por características como o tamanho das partículas, a textura, a temperatura, o teor de humidade, a dispersão da luz, a manipulação das amostras e algumas fontes de erros (Fernández-Cabanás et al., 2006). Dessa forma, a espectroscopia NIR é uma técnica

que obrigatoriamente teve de recorrer a tratamentos quimiométricos por análises multivariadas, uma vez que as bandas presentes nos espectros surgem de sobreposições e combinações, o que gera um enorme número de sinais. Assim, a determinação de concentrações dos compostos presentes nas amostras analisadas tornou-se possível de se alcançar.

A quimiometria é uma ciência que teve um grande impacto na química analítica e, desse modo, é uma área com numerosas aplicações, essencialmente naquelas que a espectroscopia NIR está inserida: indústria química, farmacêutica e agroalimentar, saúde humana e estudos ambientais. (Fernández-Cabanás et al., 2006). Os métodos quimiométricos podem ser divididos em diferentes tipos conforme os objetivos pretendidos, tais como otimização ou planeamento experimental, calibração multivariada, classificação supervisionada e não supervisionada, entre outras. No entanto, como estamos dentro do âmbito espectroscopia NIR e pretende-se desenvolver uma calibração multivariada, serão abordados tópicos acerca da análise multivariada, tais como: pré-processamento de espectros, análise de componentes principais, regressão por mínimo quadrados parciais, validação do modelo e identificação de *outliers*.

Uma das vantagens mais pertinentes dos métodos quimiométricos é a possibilidade de quantificação ou classificação sem recorrer à resolução do sinal analítico, uma vez que os interferentes podem ser eliminados pelos tratamentos matemáticos que são utilizados. Por consequente, estes procedimentos quimiométricos têm contribuído para o desenvolvimentos de imensos métodos analíticos mais rápidos, simples e baratos (Sena & Almeida, 2018).

2.4.1 Análise Multivariada

Os estudos que se baseiam em dados com diversas variáveis não devem recorrer a uma análise univariada. Embora esta análise possa fornecer respostas ao estudo, não responde da forma pretendida, pois apenas utiliza uma variável de interesse. Assim sendo, para obtenção de bons resultados do estudo é necessário recorrer a uma análise multivariada, que emprega algoritmos e conceitos inerentes à quimiometria (Bau & Lamnek, 2016).

A análise multivariada adota uma sequência de etapas, com o objetivo de detetar o máximo de erros e de amostras anómalas, para além de extrair a máxima quantidade de informação dos dados relevantes (Sena & Almeida, 2018). Dessa sequência serão referidos os pré-processamentos e alguns algoritmos para a qualificação e quantificação

em aplicações NIR. A análise de componentes principais será apresentada como método de qualificação, enquanto a regressão por mínimos quadrados parciais será apresentada como método de quantificação.

2.4.2 Pré-processamentos de espectros NIR

Desse modo, o pré-processamento de dados é a primeira etapa a ser empregue durante a análise multivariada, na qual pretende identificar e remover informações espectrais (variações aleatórias e/ou sistemáticas indesejáveis) que interfiram na previsão ambicionada. Estas interferências provocam efeitos aditivos e/ou multiplicativos. Assim, a extração das informações espectrais poderá focar-se nos parâmetros que realmente são do interesse da análise (Sena & Almeida, 2018).

Como referido anteriormente, na análise dos dados espectrais NIR é expectável a presença de variações indesejáveis, que podem ser causadas pelo espectrofotómetro, na saturação do detetor e nas falhas do varrimento dos comprimentos de onda, ou por perturbações que levam a variações no percurso ótico e dispersões da luz, tais como as variações das densidades e tamanhos das partículas presentes em amostras biológicas, entre outros (Mark & Jr., 2007; Silalahi et al., 2018). Todos estes fatores, sobretudo as dispersões da luz, acabam por ter um grande impacto negativo nos espectros das amostras de produtos agroalimentares (Rinnan et al., 2009; Sena & Almeida, 2018). Os resultados mais comuns desses fatores traduzem-se, essencialmente, em espectros NIR com ruídos, espalhamentos e desvios da linha de base (Engel et al., 2013). Desse modo, os pré-processamentos são essenciais nos espectros NIR, pois os valores gerados levam a melhorias significativas no desenvolvimento de um modelo de calibração (Boulet & Roger, 2012; Rinnan et al., 2009).

Contudo, existem muitos métodos de pré-processamento que se foram desenvolvendo ao longo dos anos, mas a sua escolha deve ser realizada com cautela e baseada nos conhecimentos dos princípios de cada método, não descartando o conhecimento sobre as variações químicas e físicas presentes nas amostras. O uso inadequado de pré-processamentos pode levar à eliminação de informações relevantes que prejudicarão a qualidade do tratamento quimiométrico (Sena & Almeida, 2018). Porém, não existe um procedimento padrão para determinar qual pré-processamento proporciona melhores resultados para desenvolver calibrações NIRS (Fernández-Cabanás et al., 2006).

Para além disso, os métodos de pré-processamento podem ser aplicados em linhas ou em colunas da matriz de dados. Quando o pré-processamento é aplicado em linhas,

este é designado como pré-tratamento e é direcionado a uma amostra de cada vez, tendo em conta todas variáveis. Por sua vez, a utilização do método em colunas é aplicada a cada variável a analisar, considerando todas as amostras (Ferreira, 2015; Sena et al., 2017).

Como as variações indesejáveis e perturbações levam à formação de espectros NIR com não linearidades, de seguida são apresentados alguns métodos de pré-processamentos possíveis que têm como objetivo: remover variações sistemáticas; minimizar o ruído (smooth, alisamento de sinal); corrigir a linha de base; melhorar a definição dos picos que se encontram sobrepostos na mesma região, evidenciando os parâmetros de interesse e consequentemente aumentando a seletividade; diminuir os efeitos causados por diferenças no tamanho de partícula; e remover as informações irrelevantes. Este tratamento permite a continuação do desenvolvimento de uma calibração, sem que haja interferência de informações irrelevantes, permitindo assim uma maior linearidade.

Centralização na Média

A centralização na média ou *Mean Centering* (MC) é um dos pré-processamento mais requisitados, e baseia-se simplesmente na realização da média da absorvância de um determinado comprimento de onda, que posteriormente é subtraída a cada registo da absorvância desse mesmo comprimento de onda, para um conjunto de amostras. Esta operação pode ser aplicada não só à absorvância mas também a uma determinada variável, equação 4 (Ferreira et al., 1999; Hibbert, 2016; Hofer, 2017; Kramer, 1998).

$$a_{i,j}^* = a_{i,j} - \frac{1}{n} \sum_{i=1}^n a_{i,j} \quad (4)$$

em que i representa as amostras e o j as variáveis a analisar; n o número de amostras utilizadas; $a_{i,j}$ o valor inicial das variáveis j das amostras i ; e $a_{i,j}^*$ o valor das variáveis das amostras após aplicação da centralização na média.

Do ponto de vista estatístico, a centralização é um deslocamento da distribuição das absorvâncias a cada comprimento de onda, levando assim a que a média da variável resultante seja zero, como podemos observar na figura 6. Este método tem como objetivo prevenir que os pontos mais distantes do centro dos dados tenham mais influência no desenvolvimento de uma calibração do que os mais próximos, dando assim importância à distância do valor médio e não à intensidade verificada.

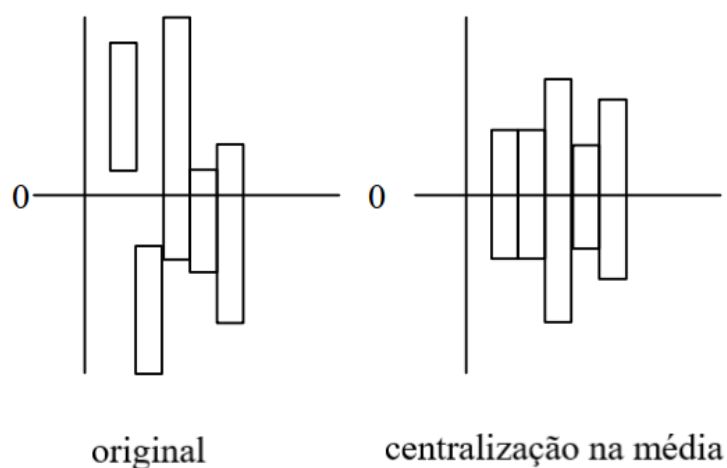


Figura 6. Representação da aplicação do método centralização na média a uma variável. (Adaptado de Nilsson, 1998)

Dessa forma, a centralização altera a escala de uma variável, mas não as suas unidades. A realização deste pré-processamento não afetar o desvio padrão dessa mesma variável (Hofer, 2017). Esta operação geralmente diminui o número de fatores do modelo de calibração.

Correção Multiplicativo de Sinal

A correção multiplicativa de sinal ou *Multiplicative Scatter Correction* (MSC) é um método amplamente aplicado em espectroscopia NIR devido às variações multiplicativas que ocorrem nos espectros obtidos. Essas variações são de carácter comum e geralmente são causadas por dispersões e fenómenos físicos (Windig et al., 2008). Dessa forma, a MSC é utilizada para eliminar o deslocamento da absorvância, bem como mudanças no declive que têm origem nos efeitos da dispersão da luz, que por sua vez advém da falta de homogeneidade das amostras (Hayati et al., 2020; Jiao et al., 2020; Mancini et al., 2019; Silalahi et al., 2018). No entanto, o procedimento deste método é simples, pois consiste na criação de um modelo de regressão linear de um espectro de referência, ou seja, espectro médio referente a todos os espectros presentes no conjunto da calibração, equação 5.

$$X_{i,j} = k_{i,j} * \bar{X} + b_{i,j} \quad (5)$$

Posteriormente, cada espectro das amostras i e variáveis j ($X_{i,j}$) é corrigido a $\tilde{X}_{i,j}$ tendo em conta o coeficiente do declive, $k_{i,j}$, e o coeficiente da interceção, $b_{i,j}$, como apresentado na equação 6 (Chu et al., 2018; Jiao et al., 2020).

$$\tilde{X}_{i,j} = \frac{X_{i,j} - b_{i,j}}{k_{i,j}} \quad (6)$$

O pré-processamento MSC foi estendida para aplicações mais elaborada, dando origem à correção de dispersão multiplicativa estendida (EMSC). O EMSC é utilizado para eliminar o comprimento do caminho incontrolável e/ou efeitos de dispersão. Contudo, este método é particularmente útil para minimizar a variação da dispersão da luz que é dependente do comprimento de onda (Mancini et al., 2019; Windig et al., 2008).

No desenvolvimento dos modelos de calibração, o método MCS reduz significativamente os fatores necessários, simplificando o modelo de calibração e aumentando a linearidade do mesmo.

Variação de Padrão Normal

A variação de padrão normal ou *Standard Normal Variate* (SNV) é um dos algoritmos mais utilizados para minimizar as interferências causadas pelas diferenças de densidade das amostras e do tamanho das partículas. A aplicação deste método faz com que os espectros apenas apresentem informação estrutural, descartando a informação relativa à intensidade (Conzen, 2014).

O método SNV usa como princípio a centralização na média do espectro em torno de zero, pela subtração da média ($\bar{X}$) das leituras espectrais da amostra i , e escalar pela divisão pelo desvio padrão das mesmas leituras ($SD(X_i)$) em todos os pontos dos espectros, equação 7, (Grisanti et al., 2018; Jiao et al., 2020).

$$\tilde{X}_i = \frac{X_i - \bar{X}}{SD(X_i)} \quad (7)$$

O modelo SNV não simplifica o modelo matemático, nem reduz as interferências sistemáticas, mas permite o melhoramento da precisão dos resultados.

Este método e o MSC, pré-processamentos da normalização, geralmente produzem resultados muito semelhantes sendo frequentemente considerados permutáveis. No entanto, as formas da determinação dos ajustes são realizadas de modo diferente, bem como o facto de o SNV utilizar apenas os dados referente a um só espectro e não ao conjunto de espectros empregues para a calibração (Bi et al., 2016; Grisanti et al., 2018; Q. Guo et al., 1999). O facto de o método SNV não utilizar um espectro de

referência torna-o mais vigoroso à detecção da presença de amostra anómalas (*outliers*) do que o MSC (Sena & Almeida, 2018).

Derivadas

A derivação espectral são ferramentas matemáticas bastantes utilizadas em espectroscopia NIR, uma vez que melhora a resolução dos sinais analíticos. Porém, este tipo de pré-processamento leva o aumento do ruído, efeito colateral. Desse modo, a utilização das derivadas nos tratamentos quimiométricos requer o uso em simultâneo de métodos de suavização (smoothing), a mencionar mais à frente (Sena & Almeida, 2018).

As derivadas podem ser classificadas como de primeira ou de segunda ordem, primeira derivada ou segunda derivada respetivamente, e as suas utilizações nos espectros NIR têm como objetivo dar mais ênfase aos picos, separando os que estão sobrepostos, permitindo aumentar a resolução dos mesmo e eliminar os efeitos aditivos. Estes métodos podem também eliminar as interferências originárias dos efeitos multiplicativos, nomeadamente os efeitos provocados pelos tamanhos das partículas (Nobari-Moghaddam et al., 2021; Sena & Almeida, 2018).

A primeira derivada elimina os desvios provocados pelos efeitos aditivos na linha de base, eliminando deslocamentos da mesma, e evidencia as características pequenas, face às características de maior intensidade. Isto leva a que estruturas mais pequenas obtenham uma forma mais acentuada, permitindo uma avaliação mais facilitada. Relativamente à segunda derivada, esta remove também os efeitos de inclinação presente na linha de base, no entanto leva a uma maior expressão do ruído espectral, tendo por isso de ser aplicada em regiões espectrais mais restritas (Conzen, 2014; Esquerre et al., 2012; Nobari-Moghaddam et al., 2021; Rinnan, 2014). Na figura 7 é possível observar a influência da aplicação das derivadas nos efeitos aditivos e efeitos aditivos mais multiplicativos em espectros sem tratamentos.

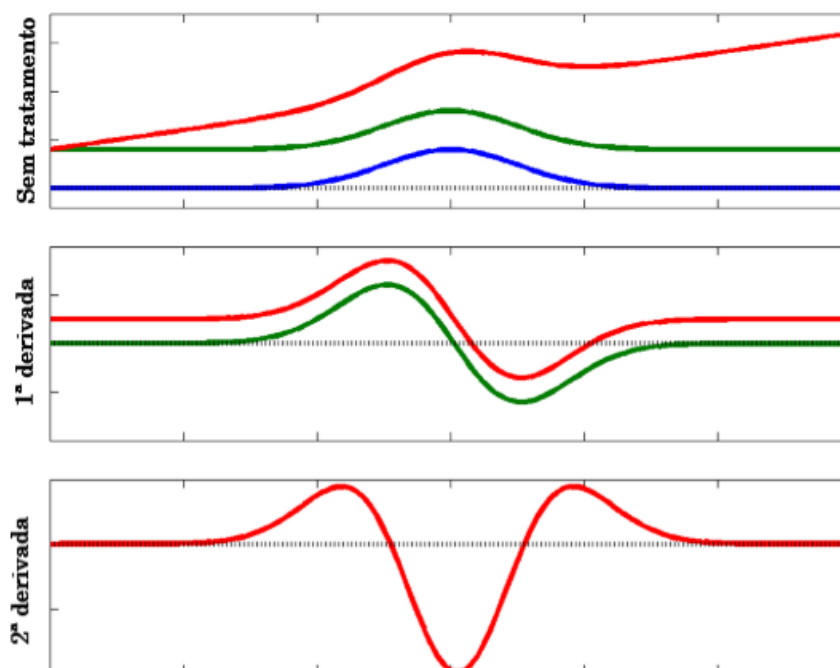


Figura 7. Representação gráfica do resultado da derivação nos efeitos aditivos (verde) e aditivos mais multiplicativos (vermelho). O espectro azul apresenta um espectro sem quaisquer deslocamentos possíveis. A linha tracejada a preto é a linha zero. (Adaptado de Rinnan et al., 2009)

Suavização

Os espectros NIR apresentam sempre um certo nível de ruído aleatório, na qual depende das amostras, do detetor do espectrofotómetro, das condições de medição e da utilização de derivadas como método de pré-processamento. Os métodos de suavização são requisitados para reduzir matematicamente o ruído, de forma a aumentar a razão sinal/ruído dos espectros, uma vez que os ruídos possuem uma alta frequência relativamente ao sinal de interesse (Sena & Almeida, 2018).

Os principais métodos mais utilizados para esse fim são: o método de Savitzky-Golay (SG) e o método de Norris-Williams (NW) (Rinnan, 2014). Estes métodos mencionados utilizam a suavização do espectro antes de calcular a derivadas, com o intuito de diminuir o efeito prejudicial do ruído. O método SG utiliza um filtro móvel onde ajusta qualquer ordem do polinómio, maioritariamente segunda e terceira ordem, em cada parte/janela do espectro. O método NW realiza a suavização dos espectros através do polinómio de ordem zero. A escolha do número de variáveis de uma janela é crucial, pois se recorrer a um número pequeno este pode não filtrar o ruído de forma correta, enquanto a utilização de um número grande pode levar à eliminação de informação relevante e distorcer o espectro. O parâmetro Gap, tamanho do intervalo em que é realizado o método, também tem um efeito de suavização nas derivadas calculadas (Rinnan, 2014; Rinnan et al., 2009; Sena & Almeida, 2018).

De-Trending

O método De-Trending (DT) ou redução da tendência espectral é usado quando existe uma dependência do comprimento de onda. Este método permite o ajuste da subtração dos polinômios, essencialmente, de 2ª ordem em cada espectro através de uma análise de regressão. Dessa maneira, o DT corrige as variações do deslocamento e da curvatura da linha de base, ou seja, corrige a linha de base (Fernández-Cabanás et al., 2006; Silalahi et al., 2018). O uso de polinômio de ordens superiores podem ser usados, contudo há um risco associado de remoção de informações relevantes (Esquerre et al., 2012). Estas variações descritas são normalmente observadas em espectros de amostras em pó ou compactadas (Nobari-Moghaddam et al., 2021).

Autoescalamento

O autoescalamento ou *standardization* é um método de normalização mais utilizado para dados que são obtidos de variáveis discretas em análises que não são fornecidas informações sobre as origens das amostras. Ou seja, é um método recomendado para a determinação de concentrações de compostos, medidas físico-químicas, entre outros.

O autoescalamento, um pré-processamento de normalização, é muito idêntico ao SNV, pois inicialmente utiliza a centralização na média, mas posteriormente, cada elemento sobre uma divisão do desvio padrão da respetiva coluna, equação 8 (Jiao et al., 2020):

$$\tilde{X}_i = \frac{X_i - \bar{X}}{SD(X_i)} \quad (8)$$

Por consequência, todas as variáveis apresentaram uma média de zero e uma variância de um, o que leva a que todas elas tenham o mesmo peso independentemente das escalas iniciais. Dessa forma, o autoescalamento em espectros NIR não são muito recomendados, pois os picos de maior intensidade, geralmente, contribuem com mais informação. Para além disso, este tipo de método em espectros aumenta a contribuição de variáveis de ruído em regiões de pouca absorvância (Sena & Almeida, 2018).

2.4.3 Análise de Componentes Principais

A análise de componentes principais ou *Principal Component Analysis* (PCA) pertence a um grupo de métodos designados de classificação não supervisionada, nas quais não são indicadas informações sobre a categoria das amostras. Estes métodos estabelecem critérios matemáticos que permitem que sejam expressos de forma quantitativa a similaridade entre duas amostras ou uma amostra e uma classe. Dessa forma, a similaridade pode ser expressa por uma medida de distância: distância de Mahalanobis (GH), mede a distância de uma amostra ao centro do conjunto de amostras (Blanco & Villarroya, 2002; NP, 2018b). O objetivo deste tipo de métodos é poder agrupar as amostras de forma natural, reduzindo as dimensões dos dados e extrair padrões ocultos (fatores latentes) das informações relevantes (Granato et al., 2018; Rodionova et al., 2021; Sena & Almeida, 2018). A PCA é uma ferramenta quimiométrica com uma ampla gama de aplicação em diversas áreas científicas, sendo normalmente utilizado na identificação de grupos distintos e na seleção de amostras para a construção de modelos de calibração.

A análise consiste na redução da dimensionalidade da matriz de dados espectrais (X), o que promove a formação de componentes principais (PCs) que condensam as informações dos dados em números menores de variáveis. Com esta aplicação, a representação de uma amostra de natureza multivariada pode ser estudada em menores dimensões, facilitando assim a sua visualização (Bro & Smilde, 2014; Granato et al., 2018).

Numa matriz de dados $X_{(n,m)}$, onde n linhas (amostras) e m colunas (comprimentos de onda), é realizada um modelo PCA onde a mesma é decomposta em produtos de uma matriz de *scores* $T_{(n,A)}$ e uma matriz de *loadings* $P_{(m,A)}$, conforme a equação 9:

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_A p_A^T + E = TP^T + E \quad (9)$$

sendo A o número de PCs no modelo PCA, t_1 e p_1 os vetores *scores* e *loadings* do primeiro PC, respetivamente, e E a matriz de resíduos (Bro & Smilde, 2014).

Os PCs são definidos de modo a explicar as variâncias dos dados espectrais. Estes componentes são modelados em ordem decrescentes das informações estatísticas, isto é: o PC1 explica uma maior variância dos dados relativamente ao PC2, o PC2 explica uma maior variância do que o PC3, e assim sucessivamente. Cada PC é uma combinação linear das variáveis originais, porém, quando relacionados, existe uma restrição ortogonal que indica que os PCs são independentes e que não se relacionam entre si. Normalmente, apenas os primeiros PCs são responsáveis pela maior quantidade de variância de dados e podem ser utilizados para representar todo o conjunto de dados de uma forma mais

simplista (Bro & Smilde, 2014; Granato et al., 2018; Sena et al., 2002; Sena & Almeida, 2018).

Os vetores *loadings* e *scores* são dos parâmetros mais importantes do modelo PCA. 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 as amostras nos eixos de cada PC, ou seja, relacionam a mesma composição dos PCs com as variáveis mais importantes.

Após a escolha do número de PCs, os dados da matriz X são projetados num espaço reduzido. O facto de os PCs serem independentes torna possível investigar as relações entre as amostras dos primeiros PCs e as relações entre as variáveis dos respetivos *loadings*. Os gráficos formados permitem reter informações úteis sobre as tendências e características estruturais acerca dos dados, bem como o agrupamento das amostras com características comuns e a deteção de amostras com *outliers* (Sena et al., 2002).

Com o intuito de melhor compreensão do método PCA, a figura 8 ilustra a uma representação de um gráfico bidimensional de um conjunto de 30 amostras ($n = 30$), em que são registadas as medidas referentes a duas variáveis ($m = 2$), que neste caso correspondem à intensidade de dois comprimentos de onda, λ_1 e λ_2 . Cada ponto do gráfico representa cada linha da matriz de dados. A função dos PCs é descrever a variação existente entre os pontos usando o menor número de eixos possíveis, ou seja, definisse novos eixos, PCs, que se alinham com os dados. A PC1 é o eixo que tem a direção que melhor descreve o máximo da variância das amostras, relativamente a qualquer uma das variáveis. A PC2, ortogonal à PC1, descreve a máxima variância restante. Posto isto, as coordenadas das amostras nos novos eixos, linha cheia, são designadas de *scores*, e como os PCs são formados pela combinação linear de variáveis originais, os coeficientes dessa combinação, vetor p , são denominados de *loadings*, representados pela linha a tracejado (Ferreira et al., 1999).

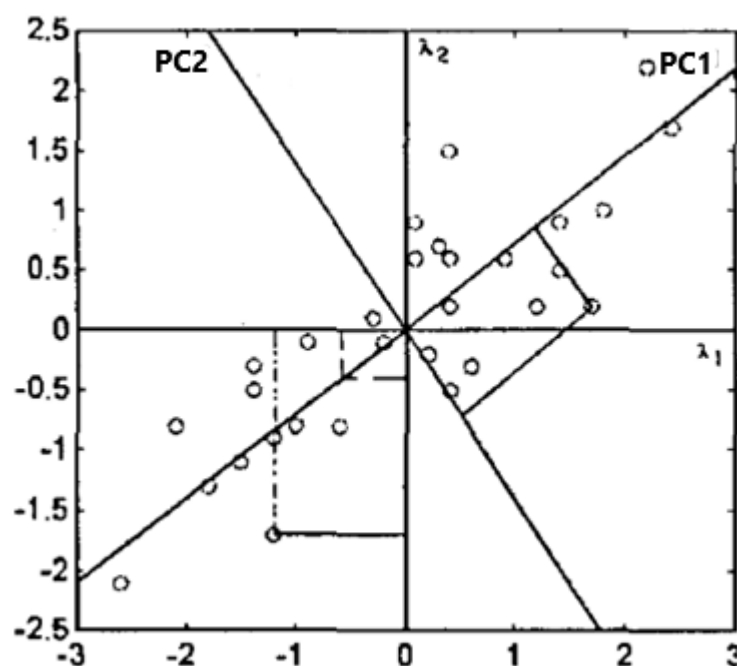


Figura 8. Representação gráfica de um conjunto de dados bidimensionais (λ_1 e λ_2) juntamente com eixos das componentes principais (PC1 e PC2). Linha cheia (—) representa os scores de uma amostra; linha tracejada (— —) representa os loadings; as linhas tracejadas e ponteadas (-.-.-) representam as coordenadas de uma amostra em relação aos eixos originais. (Adaptado de (Ferreira et al., 1999))

Em suma, a PCA é um método que tem como finalidade a redução da matriz de dados a partir de combinações lineares das variáveis originais, PCs. Por esse motivo, o PCA é um método muito utilizado em espectroscopia NIR, uma vez que os espectros apresentem inúmeras variáveis. Esta análise permite que apenas permaneçam as informações mais relevantes, excluindo as que estão associadas ao ruído (Sena & Almeida, 2018).

2.4.4 Regressão por Mínimos Quadrados Parciais

Uma calibração multivariada tem como principal objetivo estabelecer um modelo de regressão que relaciona os sinais analíticos medidos para uma amostra (por exemplo, um espectro NIR) com várias propriedades específicas. Para o desenvolvimento de uma calibração multivariada, onde muitas variáveis ou centenas ou milhares de espectros de infravermelho são utilizadas para construir um modelo, recorre-se a uma calibração inversa. Isto é, as variáveis independentes (por exemplo, os espectros) que estão contidas numa matriz X são definidos num bloco, e uma ou mais variáveis dependentes a ser previstas (y) estão inseridas numa matriz Y (por exemplo, as concentrações). Dessa forma, obtém-se um modelo matemático definido como, equação 10:

$$y = Xb + e \quad (10)$$

onde b representa o vetor dos coeficientes de regressão, de forma a minimizar os resíduos e , para cada variável de X . Os métodos de calibração multivariada são utilizados para determinarem o vetor de regressão (b), na qual os mínimos quadrados parciais é um dos métodos mais usados (Bro & Andersen, 2003; Ferreira et al., 1999; Sena & Almeida, 2018; Shao et al., 2010).

A regressão por mínimo quadrados parciais ou *Partial Least Squares* (PLS) é uma técnica que modela a relação de dois blocos de variáveis, a matriz de uma ou mais variáveis previstas (Y) com a matriz de diversa variáveis independentes (X), de forma a obterem uma correlação máxima entre si. O X é considerada uma matriz (n,p) em que n corresponde ao número de amostras e p ao número de variáveis medidas (absorvâncias de um determinado comprimento de onda, valores de diferentes tipos de propriedades, ou então uma intensidade relativa numa determinada razão massa/carga); enquanto que o Y pode ser considerado como um vetor y ($n,1$) quando o modelo PLS prevê apenas uma propriedade, sendo designado de PLS1, ou então um vetor y (n,k) quando k propriedades são previstas simultaneamente, modelo denominado de PLS2.

Este modelo de regressão é uma abordagem de fatores designados de variáveis latentes que modelam a estrutura de covariância entre dois blocos. O modelo tentará encontrar uma direção multidirecional na matriz X que explique a direção da variância máxima multidimensional na matriz Y (Abdi, 2010; Sharifi, 2018). Ou seja, este método determina o menor número de variáveis latentes que podem prever Y usando os dados em X da forma mais eficiente possível (Lourenço et al., 2008).

A matriz X é decomposta em vetores *scores* (t) e *loadings* (p) e na matriz residual E , enquanto a matriz Y é decomposta em vetores *scores* (u) e *loadings* (q) e na matriz residual F , tal como apresentado nas equações 11 e 12. O significado de h é o número de variáveis latentes utilizados pelo modelo.

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_h p_h^T + E = TP^T + E \quad (11)$$

$$Y = u_1 q_1^T + u_2 q_2^T + \dots + u_h q_h^T + F = UQ^T + F \quad (12)$$

As decomposições das matrizes X e Y são realizadas de forma a maximizar a covariações entre as novas matrizes T e U . Os blocos de matrizes X e Y são correlacionados de forma linear através do coeficiente de regressão linear b , conforme a equação 13:

$$u_h = b_h t_h \quad (13)$$

Dessa forma, os valores de b_h são agrupados numa matriz diagonal B, que contém os coeficientes de regressão entre as matrizes de scores T de X e U de Y, para que haja uma boa relação linear entre as matrizes T e U é necessário que ocorra pequenas rotações das variáveis latentes dos blocos X e Y.

Os valores estipulados para as novas amostras (Y') são calculados de acordo com a equação 14, tendo por base os seus scores T* (Chun & Keleş, 2010; Conzen, 2014; Ferreira et al., 1999; Sharifi, 2018).

$$Y' = T^* B Q \quad (14)$$

Um aspeto importante a destacar é o facto de os fatores decompostos no modelo PLS não serem necessariamente modelados por ordem decrescente da variância explicada. Este acontecimento ocorre, pois, as decomposições dos dados são realizadas ao mesmo tempo e estão sob a restrição de explicar simultaneamente a variância dos blocos, o que leva a uma pequena perda de ortogonalidade (Ferreira et al., 1999; Sena & Almeida, 2018). Na regressão PLS de uma análise química em que há simultaneidade e dependência mútua da factorização dos dados presentes na matriz X e Y, uma alteração dos dados espectrais provoca, por consequente, uma alteração na concentração do composto.

Este método de regressão é bastante requisitado devido à sua ampla gama de aplicações, bem como à sua facilidade de utilização. Além disso, é um método de quantificação capaz de suportar com grandes colinearidades e com ruídos dos dados, desde que estes interferentes estejam incluídos no conjunto de amostras da calibração (Ferreira et al., 1999; Shao et al., 2010).

No entanto, foram desenvolvidas modificações no algoritmo PLS (MPLS, mínimos quadrados parciais modificados), que permitem que os dados espectrais apresentem uma maior correlação com os dados de referência, sendo assim reduzidos a variáveis que respondam apenas pelas principais informações espectrais. Na regressão MPLS, os resíduos NIR obtidos em cada comprimento de onda, após cada fator, são padronizados antes do cálculo do próximo fator, ou seja, cada comprimento de onda é dividido pelos desvios padrões dos resíduos. Desta forma, o MPLS torna-se mais estável e preciso que o algoritmo PLS (Escuredo et al., 2018; T. Guo et al., 2021).

O erro referente à calibração, o SEC (*Standard Error of Calibration*), é apresentada na equação 15.

$$SEC = \sqrt{\frac{\sum_{i=1}^n (y_{i \text{ referência}} - y_{i \text{ previsto}})^2}{n}} \quad (15)$$

2.4.5 Validação do modelo

A validação do modelo pode ser realizada de duas maneiras, através da validação cruzada ou então pela validação externa.

Uma etapa crucial em PLS é a escolha de h , o número de variáveis latentes. Essa etapa é fundamental para qualificar o modelo, uma vez que, com diferentes seleções de número de h , adquirimos resultados distintos de calibrações. Por exemplo, se a escolha do número h for menor do que o necessário obtém-se uma calibração subajustada com resultados insatisfatórios, pois possuem informações a menos do que é indispensável; o mesmo acontece com a escolha de um número h superior ao ideal, já que é incorporado ruído e variações sistemáticas indesejáveis tornando a calibração sobreajustada. Por isso, neste processo é essencial encontrar o número ideal de variáveis latentes para fornecer uma melhor previsão de novas amostras. Para a determinação ideal do número de variáveis latentes recorre-se à validação cruzada (Faber & Rajkó, 2007).

A validação cruzada é basicamente um processo de reamostragem que define o número h apresentando menor erro possível. Ou seja, conforme o tipo de validação cruzada a utilizar são retiradas repetidamente uma amostra, ou um grupo delas, para serem conjeturadas pelo modelo construído. O LOO (*leave-one-out*) é um tipo de validação cruzada que recorre à remoção individual das amostras, onde, normalmente, recomenda-se apenas para pequenos conjuntos de calibrações, por volta de 20 amostras. Para conjuntos de dados maiores a serem utilizados em calibrações são recomendadas uma previsão das amostras por grupo: forma aleatória, sistemáticas ou venezianas (*venetian blinds*, retiram amostras em intervalos constantes). A seleção do número de variáveis latentes corresponde ao menor erro da validação cruzada, o SECV (*Standard Error of Cross Validation*), equação 16:

$$SECV = \sqrt{\frac{\sum_{i=1}^n (y_{i \text{ referência}} - y_{i \text{ previsto}})^2}{n - 1}} \quad (16)$$

onde n corresponde ao número de amostras utilizadas na calibração, $y_{i \text{ referência}}$ é o valor de referência da amostra i e $y_{i \text{ previsto}}$ é o valor previsto da amostra i pelo modelo construído.

Com o intuito de finalizar o modelo PLS é essencial recorrer a uma testagem e, para isso, utiliza-se um conjunto de amostras que apresentem concentrações conhecidas para validação. Estas amostras não podem ser usadas durante o processo de calibração. Para realização do mesmo determina-se um erro, o SEP (*Standard Error of Prediction*), equação 17.

$$SEP = \sqrt{\frac{\sum_{i=1}^n (y_{i \text{ referência}} - y_{i \text{ previsto}})^2}{n}} \quad (17)$$

onde n corresponde ao número de amostras utilizadas na validação, $y_{i \text{ referência}}$ é o valor de referência da amostra i de validação e $y_{i \text{ previsto}}$ é o valor previsto da amostra i de validação. A este tipo de testagem designa-se de validação externa, pois requer a avaliação da previsão do conjunto de amostras usadas só na validação e a estimativa do erro de previsão (SEP) (Faber & Rajkó, 2007; Sena & Almeida, 2018; Sun, 2009).

Para denotar que um modelo de calibração multivariada é considerado bom deve-se certificar que o mesmo explica a maior parte da variância nos dois blocos X e Y . Normalmente, um bom modelo abrange cerca de 90% de todas as variâncias. Quando deparamos com calibrações insatisfatórias observa-se uma grande discrepância entre os valores de SEP e de SEC.

Após a construção e validação do modelo de calibração, torna-se necessário recorrer à avaliação do desempenho do mesmo aquando à sua aplicação específica. Isto é, um processo de validação analítica que é realizado através de estimativas de figuras de mérito (FOM). Nas FOM existem vários parâmetros que estão abrangidos, nomeadamente: a robustez, a linearidade, a sensibilidade analítica, a exatidão, o bias e o desvio de previsão relativo (RPD) (Sena & Almeida, 2018).

2.4.6 Deteção de *Outliers*

Um *outlier* é a designação de amostras que são muito diferentes relativamente às restantes do conjunto de dados, ou seja, são amostras anómalas. As diferenças das mesmas podem ser provenientes de erros de medições, da sua origem, entre outros. A sua deteção é um processo crucial no desenvolvimento de calibrações multivariadas, uma vez que a sua presença prejudica a habilidade de previsão do modelo. Dessa forma, é necessário assegurar que o conjunto de amostras utilizadas para a calibração formem um conjunto homogéneo, removendo assim aquelas que se distanciam das restantes amostras (Bro & Andersen, 2003; Valderrama et al., 2007).

A detecção de *outliers* podem ser realizadas de várias maneiras, no entanto existem três formas simples de o fazer, nomeadamente a detecção de amostras com altos resíduos espectrais (bloco X), alta influência no modelo (*leverage*) e altos resíduos nos valores das propriedades ou concentração de compostos (bloco Y). A detecção de amostras com altos resíduos espectrais é realizada tendo por base a comparação do desvio padrão dos resíduos totais ($SD(e)$) com o desvio padrão de uma determinada amostras ($SD(e_i)$). Caso uma amostra apresente $SD(e_i) > 2,5 SD(e)$, a amostra é considerada um outlier. O *leverage* indica o quanto as amostras estão distantes do centro de dados, e se uma amostra apresentar um valor GH superior ao valor limite ela deve ser descartada do conjunto de amostras da calibração. Por último, a detecção de amostras anómalas com altos resíduos no bloco Y é realizada através da comparação do SEC com o erro absoluto dessa mesma amostra. Se o erro absoluto for maior duas ou três vezes maior que SEC, a amostras é um *outlier* (Valderrama et al., 2007).

3. Experimental

3.1 Preparação das amostras

Para a realização deste trabalho experimental utilizaram-se amostras de silagens de milho ($n = 241$) que foram rececionadas pela Alip, a temperatura ambiente. Para a realização dos ensaios experimentais utilizaram-se silagens de milho provenientes de várias regiões geográfica do país de forma a obter um leque abrangente de silagens a analisar.

As amostras, após serem rececionadas e devidamente identificadas, sofreram um processo de homogeneização manual, isto é, procedeu-se à realização da amostragem baseada na NP ISO 6498 (2018) conforme o procedimento interno. Posto isto, apenas cerca de 400 g de cada amostra foram usadas para um processo de trituração em fresco. Para a realização desse processo recorreu-se ao equipamento POWTEQ Knife Mill (Type HM100), durante 5 segundos a 3500 rotações por minutos (rpm). A definição dos parâmetros da trituração resultou de vários ensaios preliminares com variações tanto do tempo utilizados (5 s, 15 s, 20 s, 40 s) bem como do número de rotações (3500 rpm e 4000rpm). Dessa forma, para a seleção das condições de trituração teve-se em atenção a perda de humidade que poderia ocorrer por aquecimento e a diferença de granulometria obtida em diferentes ensaios, ou seja, estabeleceu-se um equilíbrio entre a menor perda de humidade e o melhor tamanho de partícula para favorecer uma boa compactação. Após a trituração, as amostras foram colocadas em sacos de plástico com fecho de forma a

minimizar mais perdas de humidades que poderiam ocorrer através das trocas com o ambiente atmosférico, durante um curto período de tempo.

Posteriormente, as amostras trituradas foram submetidas a dois procedimentos distintos: a uma leitura espectrofotométrica com intuito de obtenção de um espectro NIR e a uma realização de análises químicas de referência.

3.2 Espectrofotómetro de refletância NIR

Para a obtenção de espectros NIR por refletância das amostras trituradas em fresco utilizou-se o equipamento FOSS NIRSTM DS2500 (figura 9) constituído por um monocromador que ilumina as amostras numa faixa espectral de 400 nm a 2500 nm, um detetor de silício sensível de 400 nm a 1100nm e outro de sulfureto de chumbo sensível de 1100 nm a 2500nm. Este equipamento apresenta uma resolução espectral de 0.5 nm e circuitos de última geração para melhorar a relação sinal-ruído e minimizar ruídos estranhos que possam afetar o desempenho do mesmo.

Após a preparação das amostras, as mesmas foram colocadas e compactadas no copo de grande dimensão do equipamento e efetuou-se a medição da energia refletida das amostras a temperatura ambiente. O software ISIScan Nova incorporado no FOSS NIRSTM DS2500 permitiu visualizar e arrecadar os espectros NIR respetivos a cada amostra. Os valores de absorvância espectral foram registados como $\log 1/R$ sem réplicas, onde R corresponde à refletância da amostra.

Paralelamente, realizaram-se medições espectrais de amostras frescas que foram submetidas a diferentes condições de preparação. A implementação deste processo teve o intuito de simular possíveis condições/alterações que possam ocorrer durante a preparação das amostras, de forma a criar um ficheiro de repetibilidade. O ficheiro de repetibilidade serviu para minimizar a influência de efeitos indesejados nos resultados da calibração NIRS. 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 3500 rpm; a temperatura, em que algumas amostras foram submetidas a temperaturas mais baixas que rondavam entre os 8 °C e os 10 °C, temperaturas amenas de 19°C a 21 °C as quais as amostras se deparavam dentro dos sacos, e a temperaturas mais elevadas entre os 25 °C e os 28 °C; a compactação, utilizando diferentes forças de compactação ao introduzir as amostras no copo para realização da leitura; e, por último, a combinação de temperaturas mais amostragens diferentes. Todas estas vertentes têm impacto nos

espectros obtidos, e dessa forma tornou-se possível mitigar esses impactos causados por variações nas condições de preparação da amostragem.



Figura 9. Equipamento FOSS NIR5TM DS2500 utilizado para obtenção de espectros NIR juntamente com o copo de grande dimensão.

3.3 Análise química de referência

Para a determinação do pH utilizou-se amostras em fresco com recurso a um elétrico de contacto Mettler Toledom. Os restantes parâmetros eram determinados tendo em conta a percentagem de MS.

Para a determinação da MS, as amostras foram secas em estufa com circulação de ar forçada Memmert UF750, à temperatura de 80 °C, durante, sensivelmente, 18 horas.

Para as determinações analíticas a seguir referidas as amostras secas foram moídas em moinho Christy Turner (8" Laboratoy Mill) equipado com um crivo de 1 mm.

O teor em cinza total foi determinado conforme os procedimentos internos baseados na norma NP ISO 5984 (2014) que consiste num processo de inceneração da amostra numa mufla Lenton a temperatura a 550 °C, durante sensivelmente 3 horas. O resíduo incinerado representa a cinza total presentes nas amostras, uma vez que a matéria orgânica é decomposta neste processo.

A determinação do teor em PB é calculada tendo por base N Kjeldahl x 6,25, um procedimento interno baseado nas normas NP ISO 5983 (2007), em que consiste em três

fases distintas: a digestão, a destilação e a titulação. Primeiramente, a amostra é digerida num digestor FOSS Tecator Digestor 20 com ácido sulfúrico concentrado e catalisadores para converter o azoto proteico em sulfato de amónio, com posterior adição de hidróxido de sódio para a libertação de amónia. A destilação, fase seguinte, ocorre num sistema com arrastamento de vapor num destilador KjeltexTM 8200, através da solução recetora de ácido bórico. Por fim, a titulação colorimétrica com ácido sulfúrico. Dessa forma obtém-se o teor de azoto a partir da amónio produzida e o teor de PB multiplicando o resultado do teor de azoto com o fator 6,25.

O teor em amido foi determinado pelo método enzimático conforme o método interno, ou seja, baseado na hidrólise enzimática a temperaturas mais altas com recurso às enzimas α -amilase e amiloglucosidase da Megazine, sucessivamente. Posteriormente, efetuou-se as leituras de absorvância num espectrofotómetro Termo Fischer Scientific (Genesys 10uv) a 510 nm para a determinação do teor de amido nas amostras.

A determinação NDF foi determinada conforme descreve Van Soest, tendo sido adicionado α -amilase durante a extração em solução de detergente neutro, para remoção do amido presente nas amostras. A fração NDF foi expressa isenta de cinza residual, após inceneração em mufla Lenton AF 11/6B, durante 4 horas.

3.4 Desenvolvimento e validação da calibração

Após a realização das leituras espectrais das amostras trituradas em fresco e das análises químicas por referência das mesmas, procedeu-se à execução dos tratamentos matemáticos quimiométricos. O software utilizado WinISI (versão 4,6,0; FOSS) permitiu desenvolver e validar modelos de calibração que correlacionam os dados espectrais NIR com os valores de referência dos parâmetros em estudo.

Primeiramente, recorreu-se à análise dos dados espectrais por PCA para a determinação das GH, tendo sido calculada a partir dos scores para verificação da qualidade das amostras de silagem, uma vez que os resultados indicam o quão diferente é o espectro da amostra relativamente ao espectro médio do conjunto das mesmas. Com base na distribuição estatística de GH, as amostras foram classificadas como mais ou menos semelhante em relação ao espectro médio. Esta estatística permitiu também, numa fase mais avançada do desenvolvimento e validação dos modelos de calibrações, identificar possíveis *outliers* presentes nos conjuntos de amostras utilizados.

Posteriormente, as amostras foram divididas em dois grupos: para o desenvolvimento da calibração e validação cruzada ($n = 211$) e para a validação externa ($n = 30$). Esta divisão foi efetuada pelo software WinISI através de uma seleção automática. No desenvolvimento da calibração de cada parâmetro utilizou-se a quantificação através de MPLS, juntamente com o ficheiro de repetibilidade, de forma a obter as respetivas equações da análise multivariada. Durante o desenvolvimento dos modelos de calibração recorreu-se à utilização de pré-processamentos dos espectros, nomeadamente: SNV em combinação com DT (SNV-DT), juntamente com tratamentos matemáticos (1,4,4,1; 2,4,4,1; 3,5,5,1). Os pré-processamentos SNV-DT foram utilizados para correção da dispersão, de forma a remover os efeitos aditivos e multiplicativos da linha de base. A aplicação do SNV-DT levou a uma transformação dos dados espectrais brutos pela redução das interferências de características físicas, como o tamanho das partículas e o comprimento do caminho das amostras para os espectros. Com este tipo de tratamento obteve-se espectros com média zero e um desvio padrão de um (Rushing et al., 2016). Na derivatização foi utilizada uma nomenclatura com quatro dígitos (exemplo 1,4,4,1), em que o primeiro dígito representa o número da derivada, o segundo dígito diz respeito ao intervalo sobre o qual a derivada é calculada (gap), o terceiro dígito corresponde ao número de pontos de dados em uma média de execução ou suavização, e o quarto dígito representa a segunda suavização. Outro parâmetro que se fez variar durante o desenvolvimento da calibração foi os intervalos de comprimentos de onda, nomeadamente: 402 nm - 1098,2 nm; 408 nm a 1092,8 nm, 1102 nm - 2498,2 nm e 1108 nm – 2492,8 nm. Os dois últimos intervalos promoviam a exclusão a luz visível. A identificação de *outliers* foi realizada utilizando dois métodos distintos: o valor T ((valor de referência – valor previsto) / SEC) e o valor de GH, de modo a organizar os espectros de calibração. Após avaliação dos parâmetros T e GH, os possíveis outliers foram eliminados e a calibração foi novamente efetuada. Acredita-se que a discrepância de valores entre os previstos e os medidos em laboratório aconteceu devido à baixa qualidade dos valores medidos, não se correlacionando assim da melhor maneira com as características espectrais das amostras (Ikoyi & Younge, 2020).

Para avaliação dos modelos de calibração recorreu-se à validação cruzada, de forma a evitar o ajuste excessivo de dados, *overfitting*. Este tipo de validação promoveu uma divisão do conjunto de amostras da calibração. Essa divisão foi efetuada de modo a ter apenas uma amostra (LOO) ou um grupo de amostras (Group12345) para verificação de resultados, e os demais a constituírem o modelo de calibração. A estatística da validação cruzada apresenta a estimativa do desempenho das calibrações mais aprimorada do que à estatística resultante da própria calibração (Ikoyi & Younge, 2020).

O critério de escolha dos melhores modelos de calibração de MS, pH, cinza total, PB, amido e NDF utilizado foi o menor SEC e SECV, pelo maior coeficiente de correlação quadrado (RSQ) e o maior um menos a razão de variância (1-VR, estimativa da variância explicada) (Ikoyi & Younge, 2020; Rushing et al., 2016). O RSQ é responsável por explicar a relação entre os valores previstos por NIRS e os valores de análises químicas. Este parâmetro RSQ varia entre 0 e 1, em que a correlação aumenta quando a correlação entre as variáveis aumenta (Kumaravelu & Gopal, 2015).

Por fim, os modelos de calibração desenvolvidos foram validados aplicando o conjunto de amostras de validação independente. A validação externa compara os valores previstos pelos modelos desenvolvidos com os resultados obtidos de análises laboratoriais de referência. Para avaliar a capacidade de previsão dos modelos de calibração selecionados usaram-se o menor SEP, o menor bias (diferença média entre os valores NIR previsto e os valores de referência), a inclinação mais próxima de 1 e o RPD. O parâmetro RPD é determinado pela razão do desvio padrão dos valores de referência com o SEP do modelo desenvolvido, ou seja, é usado para discutir a capacidade de o modelo prever dados futuros (Yi et al., 2017). Para categorizar os valores de RPD obtidos recorreu-se a três categorias: categoria A para bons modelos de calibração, em que $SD/SEP > 2,0$; categoria B para modelos que possam ser melhorados recorrendo a diferentes estratégias de calibração, sempre que $1,4 \leq SD/SEP \leq 2,0$; e categoria C para modelo que não podem ser confiáveis para uso de previsão do método NIRS, assim que $SD/SEP < 1,4$. No entanto, quanto mais pequeno for o valor de SEP mais confiável é o modelo de calibração desenvolvido (Chang et al., 2001).

Nesta etapa de validação também se recorreu à eliminação de possíveis *outliers* que se encontravam no conjunto de amostras estipuladas para a validação externa pelos mesmos métodos utilizados na eliminação de *outliers* no desenvolvimento da calibração, reformulando assim a validação do modelo.

4. Resultados e Discussão

4.1 Análise química de referência

O conjunto de amostras utilizadas para o desenvolvimento da calibração deve de ser o mais representativo possível das amostras futuras a serem analisadas. Desse modo, conjuntos de amostras com ampla variação e grande faixa de constituintes foram utilizados para desenvolver este estudo. A variabilidade existente entre as amostras resulta da variação da localidade das mesmas e do momento de corte utilizado.

A composição química referente às 241 amostras de silagem de milho utilizadas no desenvolvimento da calibração NIRS em fresco está presente na tabela 4. Nessa tabela é possível observar o intervalo de valores relativamente aos parâmetros de MS, cinza total, PB, amido, NDF e pH, bem como os resultados da média da composição dos mesmos e o respetivo desvio padrão. Através dos dados de referência obtidos, é possível averiguar que foram utilizadas um leque abrangente de variedades de amostras de silagens de milho, uma vez que no conjunto das amostras houve uma ampla variação das concentrações nutricionais bem como no valor de pH. Isso tornou possível a construção de modelos de previsão entre os dados espectrais e análises laboratoriais para cada componente. O intervalo e os valores médios de MS, cinza total, PB, amido, NDF e pH no conjunto de amostras foram, respetivamente, 19,87% - 50,90% com uma média de 34,12%, 2,30% MS - 13,60% MS com média de 3,40% MS, 4,87% MS - 8,92% MS com média de 6,95% MS, 12,72% MS - 42,19% MS com média de 22,22% MS, 34,22% MS - 62,07% MS com média de 40,32% MS, e por último 3,33 - 5,33 com média de 3,65.

Tabela 4. Resumo estatístico dos parâmetros analisados nas amostras de silagem de milho, expondo o mínimo, o máximo, a média e o desvio padrão de cada componente.

	Mínimo	Máximo	Média	Desvio Padrão (SD)
MS*	19,87	50,90	34,12	3,53
Cinza Total**	2,30	13,60	3,40	0,91
PB**	4,87	8,92	6,95	0,57
Amido**	12,72	42,19	22,22	4,23
NDF*	34,22	62,07	40,32	3,62
pH	3,33	5,33	3,65	0,20

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

Segundo Lage (2017 e 2019), a qualidade das silagens de milho têm aumentado aos longos dos anos, no entanto existem sempre uma grande variabilidade no valor nutritivo das silagens. O efeito provocado pelo estado de maturação das plantas no momento do corte é o fator que mais influencia a variação do valor nutritivo. Essa variabilidade é bastante visível nos resultados obtidos na tabela 4. De acordo com os resultados nutricionais da campanha 2018-2019 verifica-se que existe uma grande similaridade, em termos médios, com os valores obtidos deste trabalho experimental. Os valores médios da campanha relativamente à MS, PB, Amido, NDF são, respetivamente: 34,1%, 6,97% MS, 32,2% MS e 42,6% MS, enquanto neste trabalho experimental obtiveram-se valores de: 34,12% para a MS, 6,95% MS para a PB, 32,22% MS para o amido e 40,32% de MS para o NDF.

Na figura 10 está presente um conjunto de diagramas barras e bigodes correspondente a cada parâmetro analisado nas amostras de silagem de milho. Por um lado, 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. Considerando as escalas utilizadas em cada gráfico, destacam-se os componentes da MS e do NDF como os parâmetros com maiores variações de valores e o componente do pH o parâmetro com menor variação. Por outro lado, nos diagramas observa-se uns pontos que estão na mesma linha que a respetiva caixa de bigodes, apresentando-se como possíveis anomalias, ou seja, *outliers* devido à discrepância de valores relativamente ao conjunto total de amostras. Desta forma, identifica-se seis valores atípicos no componente da MS, estando três pontos acima do limite máximo e três pontos abaixo do limite mínimo. No gráfico da cinza total estão presentes também seis valores discrepantes, no entanto todos eles encontram-se situados na parte superior do diagrama, ou seja, acima do limite máximo. Relativamente à PB, este apresenta cinco valores que não se enquadram com os restantes, expondo dois deles acima do valor máximo e três abaixo do valor mínimo. No que diz respeito ao parâmetro do amido, este é o único que exibe apenas possíveis anomalias na parte inferior do diagrama, sendo apenas quatro amostras com valores de amido inferior ao limite mínimo. O componente do NDF é aquele que apresenta o menor número de valores discrepantes relativamente ao conjunto total de amostras, estando todos eles, os três, na parte superior ao limite máximo do diagrama de bigodes. Por último, o parâmetro do pH apresenta, tal como a cinza total e o NDF, apenas valores atípicos na parte superior do gráfico, destacando-se quatro pontos do restante conjunto de amostras.

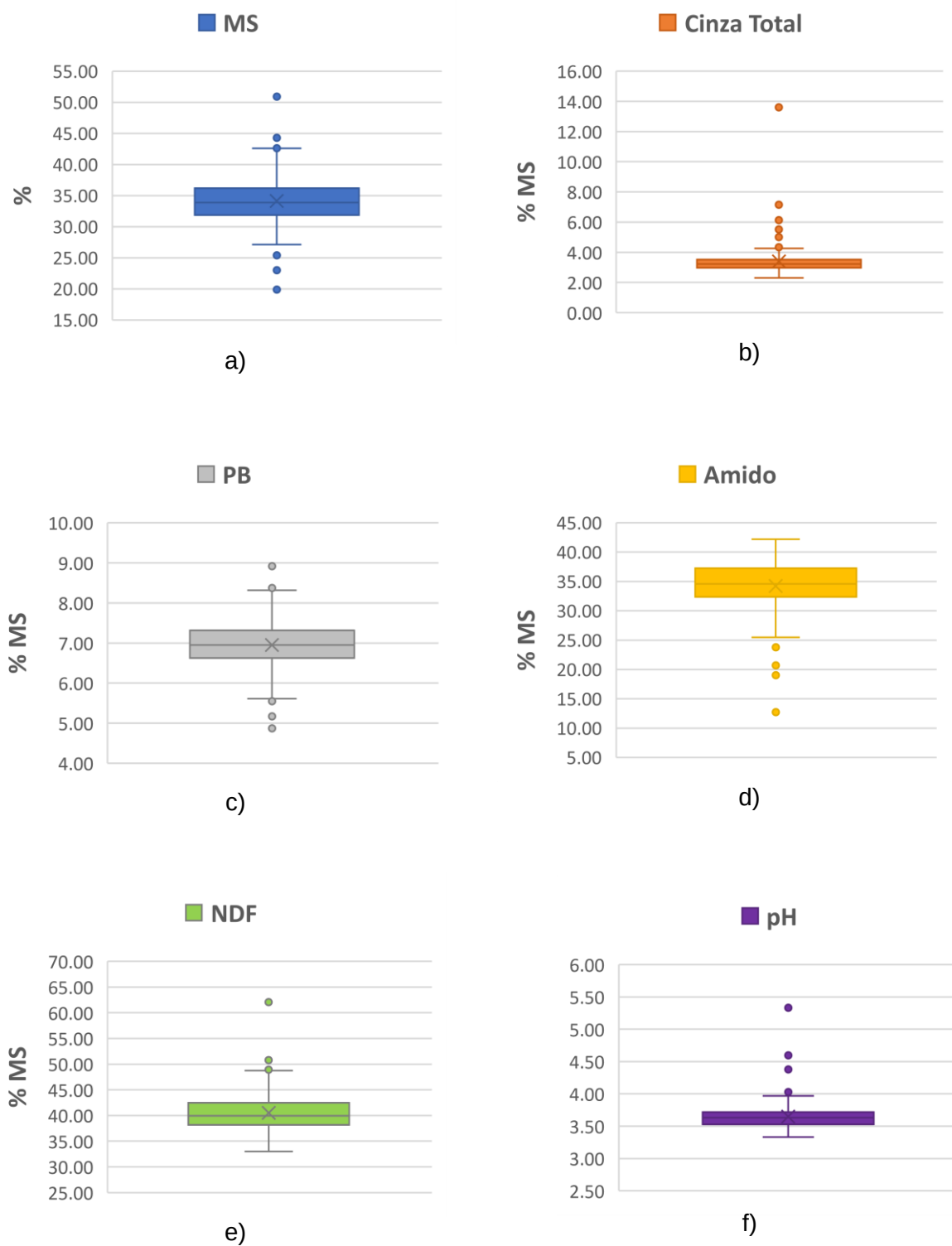


Figura 30. Representação gráfica do conjunto de diagrama de barras e bigodes de cada parâmetro analisado. a) Diagrama de bigodes referente à MS. b) Diagrama de bigodes referente à cinza total. c) Diagrama de bigodes referente à PB. d) Diagrama de bigodes referente ao amido. e) Diagrama de bigodes referente ao NDF. f) Diagrama de bigodes referente ao pH.

As correlações entre os parâmetros analisados por análise química estão apresentadas na figura 11 através de um conjunto de gráficos simplificados obtidos a partir do software TIBCO Statistica™ (Version Trial). Nesta imagem é possível reconhecer a existência de correlação já presumíveis entre alguns parâmetros, nomeadamente: a MS com o amido, a MS com o NDF e o amido com o NDF. Relativamente à relação da MS juntamente com o amido observa-se uma linha de tendência com declive positivo, ou seja, à medida que aumentam os valores de MS aumenta o teor em amido nas mesmas amostras. A correlação entre a MS e o NDF é negativa, uma vez que se observa uma diminuição dos valores de NDF aquando ao aumento do teor em MS. O mesmo acontece entre a correlação do amido com o NDF, apresentando assim uma linha de tendência com declive negativo. As correlações mencionadas já eram conjeturadas, porque durante a fase de crescimento da espiga na planta ocorre aumento dos teores de MS e de amido que estão relacionados com a deposição do amido nos grãos quando atinge a maturidade. O teor de fibra aumenta com a maturidade da planta, no entanto a sua percentagem vai diminuindo devido ao aumento das proporções da espiga relativamente ao caule e folhas. Tudo isto é traduzido numa correlação negativa e considerável quando se relaciona os teores de NDF com os de MS e de amido (Lage, 2019).

Nesta figura, tal como na anterior, são identificados possíveis *outliers*, devido à distância anormal existente em certos pontos quando se correlacionam diferentes parâmetros. Este tipo de resultados pode ocorrer de prováveis erros executados durante as análises químicas ou então de amostras que se diferenciam das restantes utilizadas.

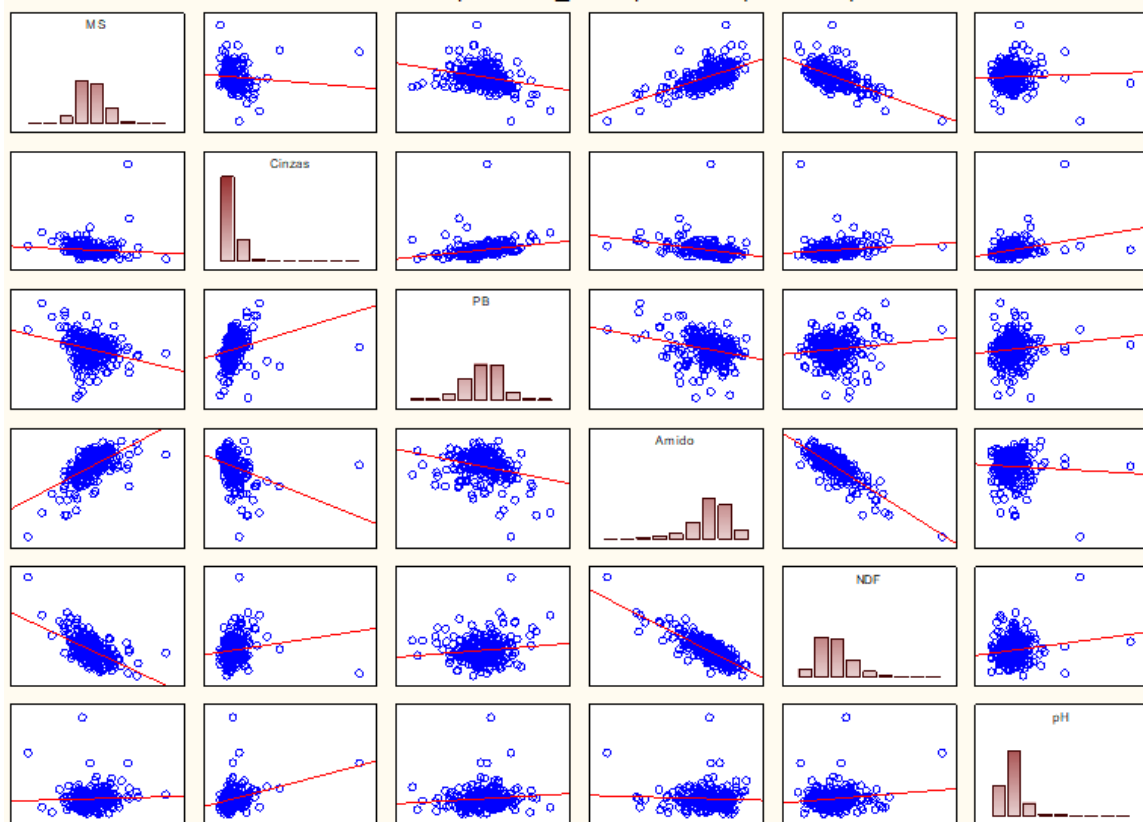


Figura 11. Representação gráfica das correlações existentes entre os parâmetros analisados nas amostras de silagem de milho, expondo em histogramas apenas a relação de um componente e a linha de tendência a vermelho para as relações entre dois parâmetros diferentes.

4.2 Espectros NIRS

No desenvolvimento da calibração MPLS recorreu-se a tratamentos quimiométricos para remoção de informação irrelevantes. Dessa forma obteve-se uma redução de variantes, apresentando apenas a informação importante para o desenvolvimento da calibração. Na figura 12a) é possível observar os espectros brutos obtidos das amostras utilizadas neste estudo sem recorrer a qualquer tipo de tratamento. Essa figura é constituída com informação relevante e irrelevante. Na figura 12b) é possível visualizar o efeito provocada pela aplicação dos tratamentos SNV-DT. Nas restantes figuras estão presentes os espectros obtidos com os tratamentos empregues no desenvolvimento da calibração, ou seja, aplicação do SNV-DT juntamente da primeira derivada (figura 12c), segunda derivada (figura 12d) e a terceira derivada (12e).

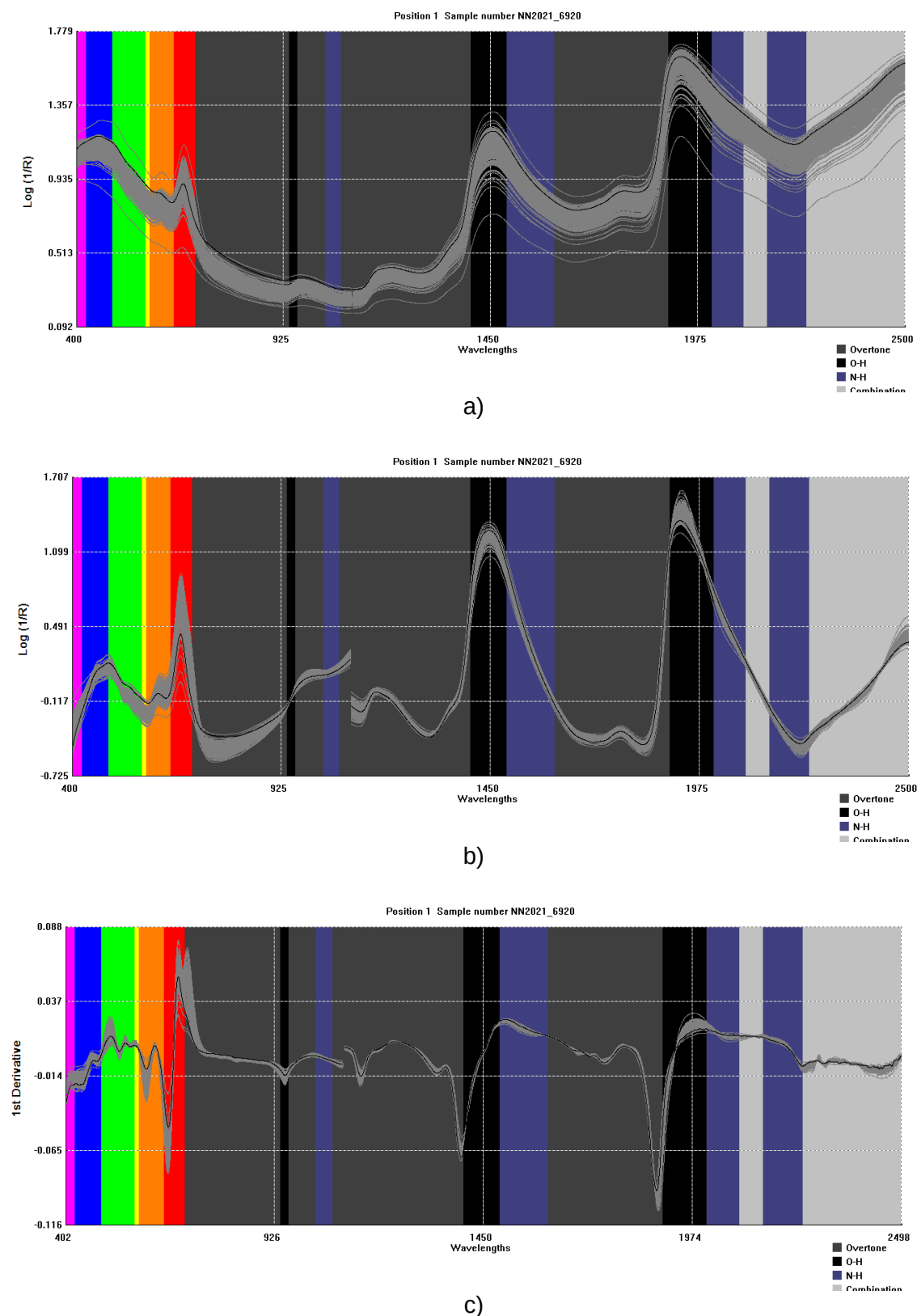
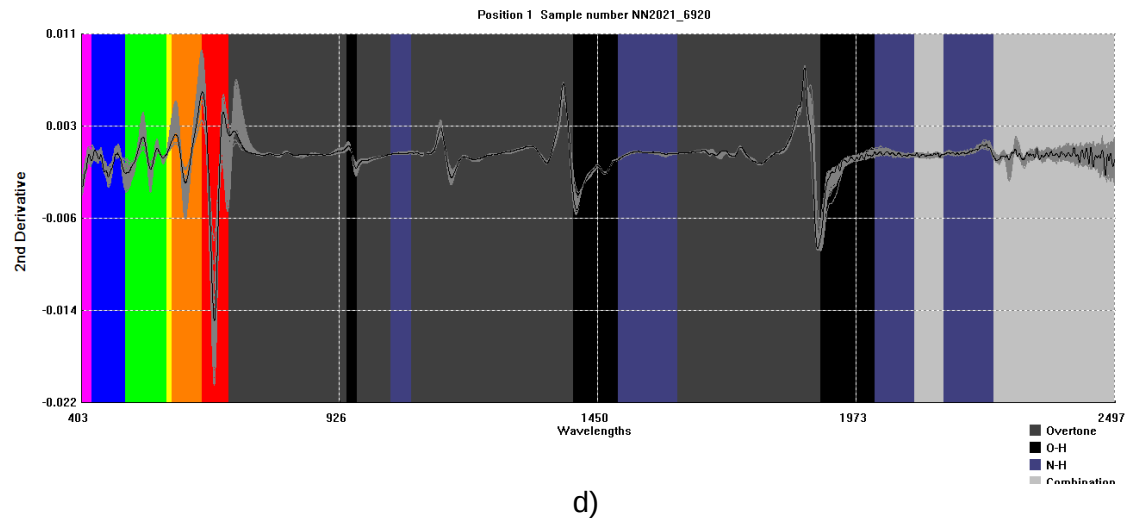
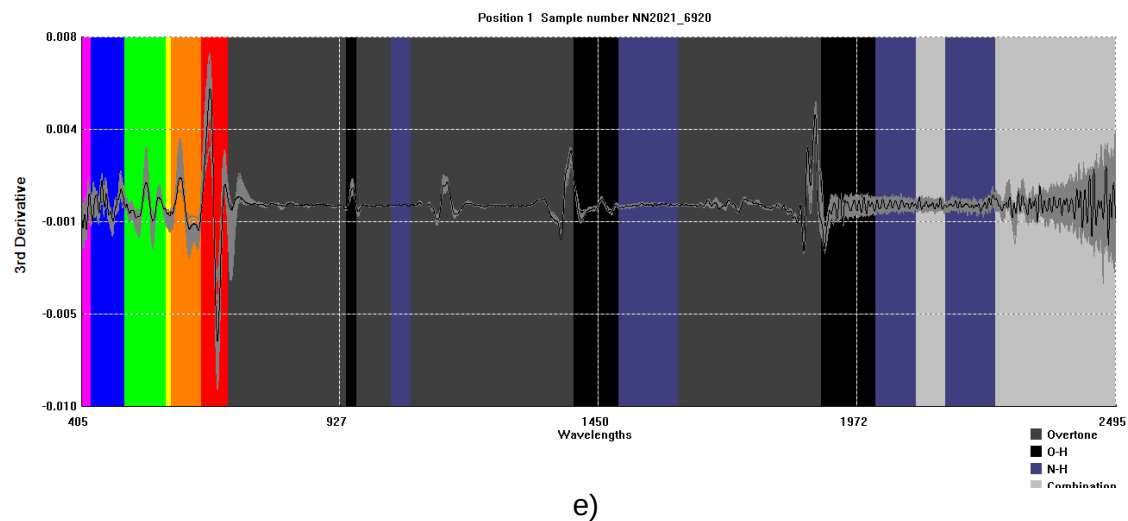


Figura 14. Espectros NIRS de amostras de silagem de milho em fresco sem e com tratamentos quimiométricos. a) Espectro NIRS em $\text{Log}(1/R)$. b) Espectro NIRS com aplicação do tratamento SNV-DT. c) Espectro NIRS com aplicação dos tratamentos SNV-DT e a primeira derivada.



d)



e)

Figura 15 continuação. Espectros NIRS de amostras de silagem de milho em fresco sem e com tratamentos quimiométricos. d) Espectro NIRS com aplicação dos tratamentos SNV-DT e a segunda derivada. e) Espectro NIRS com aplicação dos tratamentos SNV-DT e a terceira derivada.

O software utilizado permite ainda identificar zonas típicas ao longo dos espectros. Essas zonas correspondem a bandas de overtones, O-H, N-H e ainda combinações. No exame visual dos espectros brutos NIRS, estes mostraram picos com menor absorção em torno dos comprimentos de onda 1450 nm e 1930 nm. Estes picos de absorção reduzida nos espectros brutos correspondem a bandas O-H que são sinónimos de perda de humidade (Ikoyi & Younge, 2020).

4.3 Estatística do desenvolvimento e validação da calibração

As estatísticas descritivas relativas ao número de amostras, média, desvio padrão, pré-processamentos de espectros, intervalos de comprimentos de onda, tipo de validação cruzada utilizada, SEC, RSQ, 1-VR, SECV e variáveis latentes para os melhores modelos de calibração obtidos estão presentes nas tabelas 5 - 7. Cada componente analisado obteve estatísticas de calibração e de validação cruzadas diferentes entre si, como era de esperar. Obviamente, os resultados estatísticos de calibração e de validação cruzada foram afetados pelos tratamentos matemáticos, nomeadamente: a primeira derivada (1,4,4,1; tabela 5), a segunda derivada (2,4,4,1; tabela 6) e terceira derivada (3,5,5,1; tabela 7). Geralmente, as equações de calibrações da primeira ou da segunda derivada produziram melhores resultados, pois geraram maiores valores de RSQ e 1-VR e menores valores de SEC e SECV em quase todos os componentes nutricionais analisados.

Relativamente à MS, verifica-se que os valores de RSQ e 1-VR não variam quando se recorre a equações de calibração de primeira, segunda e terceira derivada, apresentando valores de 0,99 para RSQ e de 0,97 para 1-VR. No entanto, os resultados estatísticos relativos ao SEC e de SECV são menores quando se utiliza o tratamento 1,4,4,1, respetivamente 0,38 e 0,58. Portanto, considera-se que a equação de primeira derivada levou ao melhor modelo de calibração do parâmetro MS, numa primeira fase de avaliação, faltando ainda a avaliação através da validação externa.

A melhor estatística de calibração referente à cinza total corresponde à equação de primeira derivada, pois apresenta valores maiores de RSQ e 1-VR e menores de SEC e SECV, nomeadamente 0,76 para RSQ, 0,50 para 1-VR, 0,27 para SEC e 0,43 para SECV.

Na determinação das estatísticas das equações para o desenvolvimento de calibração correspondentes à PB verifica-se que não existe uma linearidade. Ou seja, os modelos com maiores valores de 1-VR e menores SEC não apresentam maiores valores de RSQ, nem os menores de SEC, e vice-versa. Contudo, após uma análise dos resultados optou-se por utilizar o modelo que apresenta melhores resultados relativos à validação cruzada, especificamente a equação de primeira derivada que apresenta valores de 0,18 e 0,29 para SEC e SECV, respetivamente, e valores de 0,90 e 0,73 para RSQ e 1-VR, respetivamente.

O mesmo acontece quando se analisa a estatística das equações referentes aos modelos de calibração do amido. Portanto, recorre-se à melhor estatística da validação cruzada para determinar qual o melhor modelo do amido. Dessa forma, averiguar que o

melhor modelo, nesta primeira fase de análise, corresponde também à equação da primeira derivada do desenvolvimento da calibração. Esta equação apresenta como resultados de SEC e SECV de 1,68 e 2,04, respetivamente, e resultados de RSQ e 1-VR de 0,86 e 0,76, respetivamente.

A estatística dos modelos de calibrações relativos ao parâmetro do NDF mostra que a equação obtida da primeira derivada apresenta os melhores valores relativos à validação cruzada, sendo 1,97 para 1-VR e 0,68 para SEC. No entanto, os resultados da calibração obtidos da equação da terceira derivada, relativamente ao SEC e RSQ, são melhores pois aparecem como 0,91 para SEC e 0,92 para RSQ.

O último parâmetro a ser analisado, o pH, demonstra também uma falta de linearidade entre a relação dos maiores RSQ e 1-VR e menores SEC e SECV. Deste modo, verifica-se a os melhores valores referentes à calibração correspondem à equação da primeira derivada com resultados de 0,06 para SEC e 0,83 para RSQ, e que os melhores valores referentes à validação cruzada estão presentes na equação da segunda derivada, com resultados de 0,50 para 1-VR e 0,14 para SECV. Nesta primeira fase de avaliação, seleciona-se a equação relativa à segunda derivada a melhor opção de modelo de calibração obtido.

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 silagem de milho em fresco. Comprimento de onda (λ); número de amostras utilizadas para o modelo de calibração (N); desvio padrão (SD), erro padrão de calibração (SEC); coeficiente de correlação múltipla da calibração (RSQ); um menos a razão da variância (1-VR); erro padrão de validação cruzada (SECV); variáveis latentes (VL).

	λ	Val. Cruzada	N	Média	SD	Calibração		Validação Cruzada		VL
						SEC	RSQ	1-VR	SECV	
MS	402-1098,2 1102-2498,2	G12345	210	34,24	3,49	0,38	0,99	0,97	0,58	11
Cinza Total	1102-2498,2	One Out	209	3,31	0,52	0,27	0,76	0,50	0,43	11
PB	402-1098,2 1102-2498,2	One Out	207	6,95	0,52	0,18	0,90	0,73	0,29	13
Amido	408-1092,8 1108-2492,8	G12345	208	34,35	4,15	1,68	0,86	0,76	2,04	6
NDF	402-1098,2 1102-2498,2	One Out	208	40,26	3,13	1,64	0,74	1,97	0,68	6
pH	1102-2498,2	One Out	209	3,63	0,15	0,06	0,83	0,48	0,14	12

Tabela 6. Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e segunda derivada (2,4,4,1) para cada parâmetro analisado nas amostras de silagem de milho em fresco. Comprimento de onda (λ); número de amostras utilizadas para o modelo de calibração (N); desvio padrão referente aos valores de laboratório (SD), erro padrão de calibração (SEC); coeficiente de correlação múltipla da calibração (RSQ); um menos a razão da variância (1-VR); erro padrão de validação cruzada (SECV); variáveis latentes (VL).

	λ	Val. Cruzada	N	Média	SD	Calibração		Validação Cruzada		VL
						SEC	RSQ	1-VR	SECV	
MS	402-1098,2 1102-2498,2	One Out	210	34,24	3,49	0,43	0,99	0,97	0,63	6
Cinza Total	408-1092,8 1108-2492,8	One Out	208	3,31	0,52	0,40	0,44	0,26	0,53	3
PB	402-1098,2 1102-2498,2	One Out	207	6,96	0,52	0,11	0,96	0,61	0,34	13
Amido	402-1098,2 1102-2498,2	One Out	209	34,33	4,15	1,57	0,87	0,74	2,16	5
NDF	1102-2498,2	One Out	206	40,28	3,14	1,46	0,79	0,66	2,03	4
pH	402-1098,2 1102-2498,2	One Out	209	3,63	0,15	0,08	0,72	0,50	0,14	4

Tabela 7. Resumo das melhores estatísticas de desenvolvimento de equações utilizando MPLS juntamente com SNV-DT e terceira derivada (3,5,5,1) para cada parâmetro analisado nas amostras de silagem de milho em fresco. Comprimento de onda (λ); número de amostras utilizadas para o modelo de calibração (N); desvio padrão referente aos valores de laboratório (SD), erro padrão de calibração (SEC); coeficiente de correlação múltipla da calibração (RSQ); um menos a razão da variância (1-VR); erro padrão de validação cruzada (SECV); variáveis latentes (VL).

	λ	Val. Cruzada	N	Média	SD	Calibração		Validação Cruzada		VL
						SEC	RSQ	1-VR	SECV	
MS	1102-2498,2	One Out	210	34,24	3,49	0,36	0,99	0,97	0,64	5
Cinza Total	402-1098,2 1102-2498,2	One Out	209	3,31	0,52	0,35	0,57	0,25	0,53	3
PB	402-1098,2 1102-2498,2	One Out	208	6,96	0,53	0,10	0,96	0,49	0,39	11
Amido	408-1092,8 1108-2492,8	One Out	206	34,33	4,11	1,21	0,92	0,67	2,43	7
NDF	402-1098,2 1102-2498,2	One Out	207	40,34	3,14	0,91	0,92	0,56	2,30	7
pH	402-1098,2 1102-2498,2	One Out	208	3,63	0,14	0,07	0,75	0,47	0,15	4

O procedimento da validação externa visa avaliar a previsibilidade dos modelos de calibração NIRS obtidos (Guo et al., 2021). As estatísticas da validação externa estão presentes nas tabelas 8 - 10, apresentando o número de amostras, a média, o desvio padrão referente aos valores de laboratório, o SEP, o RSQ da validação externa, o bias, a inclinação e o RPD. As estatísticas foram avaliadas relativamente às melhores equações de calibrações desenvolvidas pela regressão MPLS juntamente SNV-DT e primeira, segunda e terceira derivada de cada parâmetro (tabela 5 -7). As variações das amostras foram bem representadas pelas previsões obtidas do NIRS. Com base no SEP, bias, RSQ da validação e RPD, os modelos de equação previram com precisão os constituintes da MS, cinza total, PB, amido, NDF e pH. Na validação externa de todos os parâmetros

analisados, bem como nos diferentes tratamentos utilizados em cada um, verifica-se que existe uma linearidade relativamente aos valores de SEP, RSQ e RPD. Isto é, quanto menor for o valor de SEP maiores são os valores obtidos de RSQ da validação e o RPD, e vice-versa.

Os valores estatísticos relativos ao RSQ e RPD foram sempre maiores e o de SEP foram sempre menores nos modelos de calibração da primeira derivada do que nas equações de segunda e terceira derivada. Esta verificação foi observada em todos os parâmetros nutricionais analisados (tabela 8).

No que diz respeito ao componente da MS, o processo de validação externa confirmou que a equação da primeira derivada era o melhor modelo a utilizar para a calibração. Os resultados obtidos para o SEP, RSQ e RPD foram, respetivamente, 0,66, 0,96 e 4,57. Desta forma, podemos afirmar que neste modelo obteve-se uma boa correlação entre os dados laboratoriais e os dados previsto, e já que o valor de RPD >2, este é um bom modelo de calibração.

Relativamente ao parâmetro da cinza total, a equação da primeira derivada obteve resultados estatísticos melhores do que as restantes equações na validação externa. Este parâmetro apresentou resultados de 0,47 para SEP, 0,63 para RSQ e 1,56 para RPD. A correlação entre os valores de referência de laboratório com os previstos pelo NIRS não é muito boa, e o modelo de calibração pertence à categoria B quando relacionado com o valor de RPD, sendo aconselhável recorrer a diferentes estratégias calibração.

Os parâmetros estatísticos da validação externa alusivo à PB indicam que o modelo de calibração relativo à equação da primeira derivada pertence, tal como a cinza total, à categoria de melhoramento do modelo recorrendo a outras estratégias de calibração. O modelo está nesta categoria pois apresenta valores de RPD de 1,61. As restantes estatísticas da validação são de 0,38 para SEP e de 0,62 para RSQ, o que indica que também não existe uma ótima relação entre os resultados previstos pelo NIRS com os resultados obtidos em laboratório.

O amido apresenta como valores estatísticos de validação externa da equação de primeira derivada de 1,67 para o SEP, 0,76 para o RSQ e 2,07 para RPD. Este modelo confirma a previsão com precisão do método NIRS, apresentando como modelo de categoria A, sendo assim um bom modelo de calibração resultante dos tratamentos utilizados. No entanto, a correlação existente entre os valores de referência e os previstos não é considerada muito boa.

Alusivo ao parâmetro de NDF, este mostra que a equação de primeira derivada é um bom modelo de calibração quando avaliado pela validação externa, pois apresenta como RPD um valor de 2,06, ou seja, categoria A. Os restantes parâmetros da validação externa apresentam valores menor favoráveis, tal como 1,66 para SEP e 0,79 para RSQ. Desta forma, podemos averiguar que os resultados obtidos do modelo de calibração referentes à equação de primeira derivada do amido e do NDF são muito semelhantes.

Em relação ao componente do pH, o modelo de equação de primeira derivada apresenta um SEP ligeiramente baixo, igual a 0,11, no entanto os valores de RSQ e RPD não são muito altos, expondo-se como 0,77 e 1,90, respetivamente. Deste modo, pode-se concluir que este modelo de calibração não apresenta uma boa correlação entre os resultados obtidos por análise química em laboratório e os previstos por NIRS, e nem é considerado um bom modelo a ser utilizado, pois apresenta um valor entre $1,4 < SD/SEP < 2,0$, necessitando de ajustes através de utilização de diferentes estratégias de calibração.

Tabela 8. Estatística de monitoramento do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (1,4,4,1)) para cada parâmetro de silagem de milho em fresco. Número de amostras utilizadas para a validação externa (N); desvio padrão referente aos valores de laboratório (SD), erro padrão de previsão (SEP); coeficiente de correlação múltipla da validação (RSQ); desvio de previsão relativo (RPD).

	N	Média	SD	SEP	RSQ	Bias	inclinação	RPD
MS	29	33,12	3,00	0,66	0,96	-0,24	1,05	4,57
Cinza Total	29	3,42	0,73	0,47	0,63	0,04	1,42	1,56
PB	29	7,01	0,60	0,38	0,62	-0,01	0,85	1,61
Amido	29	34,68	3,45	1,67	0,76	-0,19	1,01	2,07
NDF	28	40,45	3,43	1,66	0,79	0,57	1,06	2,06
pH	30	3,65	0,21	0,11	0,77	0,01	1,37	1,90

Tabela 9. Estatística de monitoramento do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (2,4,4,1)) para cada parâmetro de silagem de milho em fresco. Número de amostras utilizadas para a validação externa (N); desvio padrão (SD); erro padrão de previsão (SEP); coeficiente de correlação múltipla da validação (RSQ); desvio de previsão relativo (RPD).

	N	Média	SD	SEP	RSQ	Bias	inclinação	RPD
MS	29	33,03	3,00	0,72	0,95	-0,15	1,10	4,18
Cinza Total	9	3,35	0,73	0,58	0,39	0,12	1,39	1,24
PB	29	7,00	0,60	0,45	0,42	0,01	0,89	1,33
Amido	29	34,82	3,45	1,86	0,71	-0,33	1,05	1,86
NDF	28	40,65	3,43	1,72	0,76	0,37	1,11	1,99
pH	30	3,66	0,21	0,14	0,61	0,01	1,51	1,50

Tabela 10. Estatística de monitoramento do conjunto de amostras de validação externa relativamente à calibração desenvolvida (MPLS juntamente com SNV-DT e (3,5,5,1)) para cada parâmetro de silagem de milho em fresco. Número de amostras utilizadas para a validação externa (N); desvio padrão (SD); erro padrão de previsão (SEP); coeficiente de correlação múltipla da validação (RSQ); desvio de previsão relativo (RPD).

	N	Média	SD	SEP	RSQ	Bias	inclinação	RPD
MS	29	33,07	3,00	0,76	0,94	0,18	1,00	3,92
Cinza Total	29	3,33	0,73	0,57	0,43	0,14	1,41	1,27
PB	29	6,98	0,60	0,57	0,15	0,02	0,57	1,05
Amido	29	35,07	3,45	2,33	0,57	-0,59	0,90	1,48
NDF	28	40,74	3,43	2,39	0,50	0,28	1,00	1,43
pH	30	3,66	0,21	0,14	0,59	0,01	1,49	1,47

A relação existente entre os resultados obtidos por análise de referência e os resultados previsto pelo NIRS de cada parâmetro nutricional analisado está presente nas figuras 13 – 18. Cada uma dessas figuras apresenta um conjunto de três gráficos que mostram essa correlação nos diferentes tipos de tratamentos utilizados. Isto é, a) alusivo ao tratamento com a primeira derivada (1,4,4,1); b) refere-se ao tratamento com a segunda derivada (2,4,4,1); e c) relativo ao tratamento utilizado com a terceira derivada (3,5,5,1).

Desta forma é possível observar as diferenças existentes nas previsões provocadas por cada modelo de calibração durante o processo de validação externa, em cada parâmetro.

Até agora poucos estudos foram realizados relativamente ao desenvolvimento de calibrações para silagens de milho e em fresco. No entanto, Cozzolino et al., (2006) desenvolveu um modelo de calibração de MPLS essencialmente com tratamentos de SNV-DT e derivadas. Nesse estudo apenas foi realizado o desenvolvimento da calibração e a validação cruzada, não executando a validação externa. Contudo, o modelo de calibração desenvolvido apresentou valores do coeficiente de calibração de 0,90 para a MS e PB, de 0,84 para o NDF e de 0,60 para o pH. O estudo apresenta também valores relativos ao RPD, porém utiliza o valor de SECV a dividir pelo SD de modo a avaliar previsão dos modelos executados. Outro estudo, mas mais antigo, foi o de Fontaneli et al., (2002) que desenvolveu um modelo de calibração PLS para os parâmetros de silagem de milho em condições de tratamentos diferentes, no entanto deste recorreu ao uso de validação externa. Apesar das diferentes condições, este apresentou valores referentes ao coeficiente de validação de 0,99 para MS, e NDF e 0,98 para a PB. Um outro estudo para prever os parâmetros fermentativos de silagem de milho por NIRS foi desenvolvido por Sørensen, (2004). Um dos parâmetros a ser analisado foi o pH, na qual obtiveram 0,62 como valor relativo ao coeficiente da validação e 1,8 como valor alusivo ao RPD.

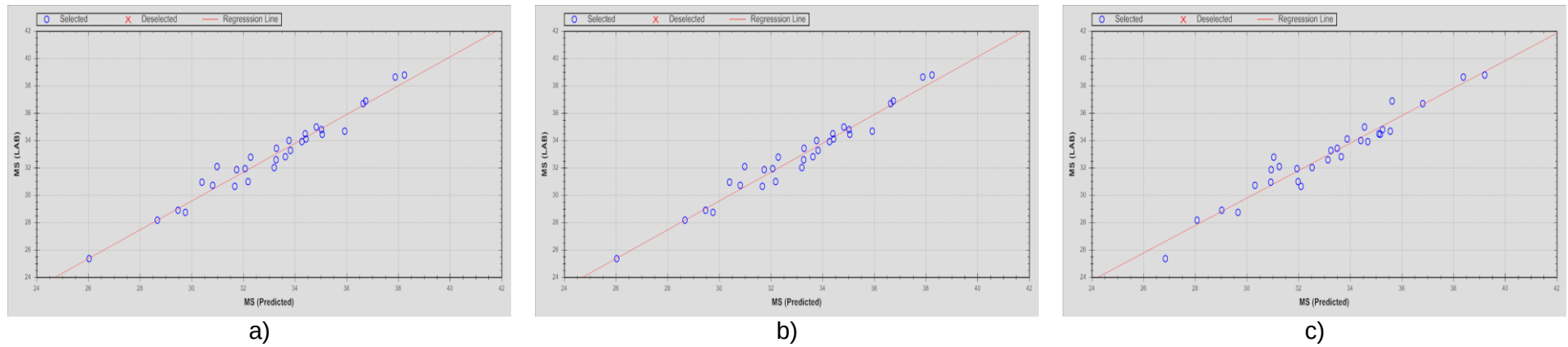


Figura 13. Comparação dos valores de MS de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores utilizando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores utilizando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores utilizando MPLS juntamente com SNV-DT e 3,5,5,1.

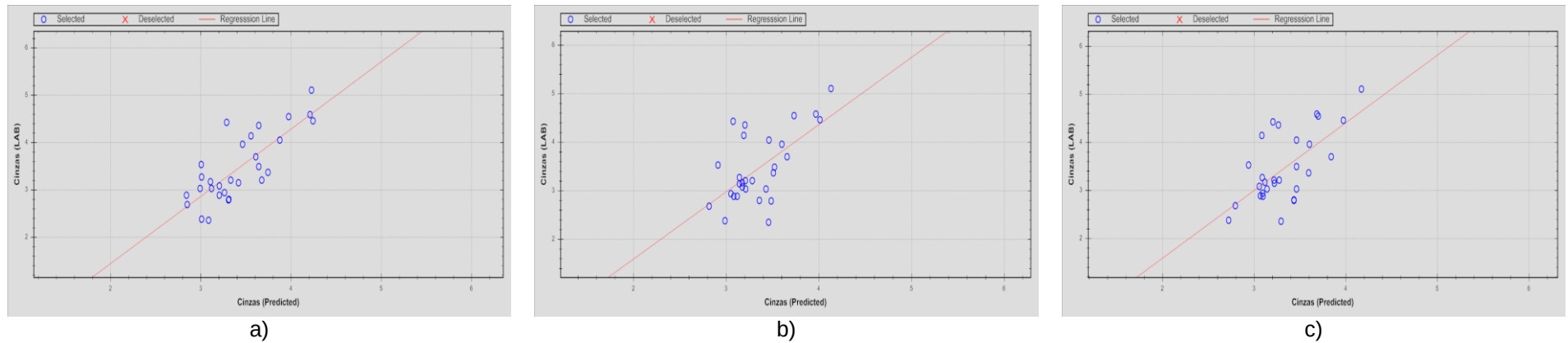
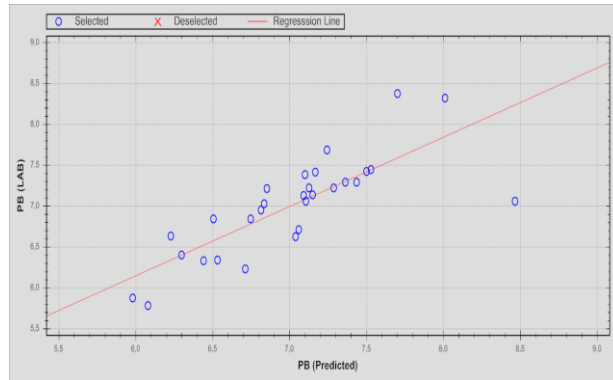
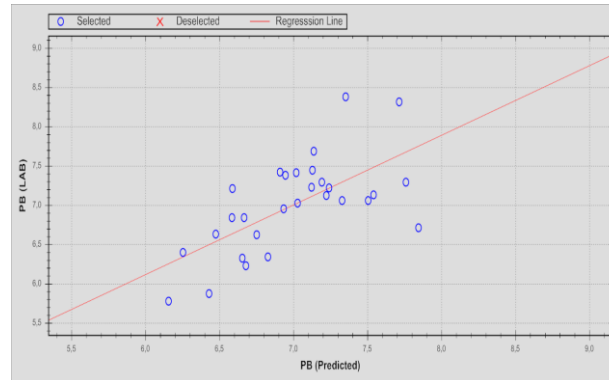


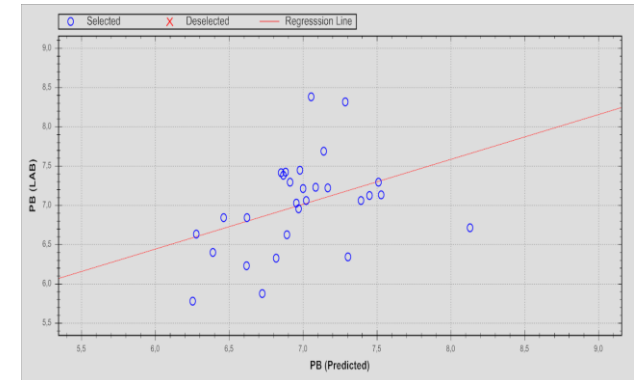
Figura 14. Comparação dos valores de cinza total de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.



a)

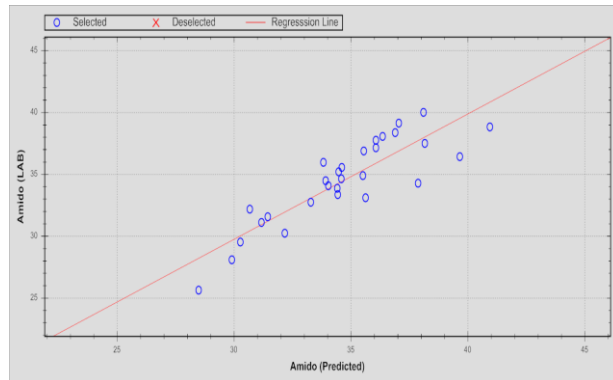


b)

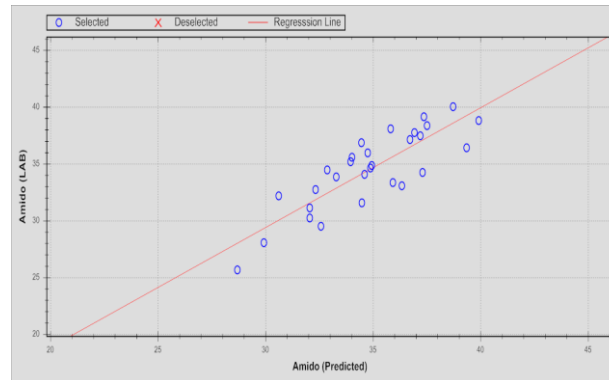


c)

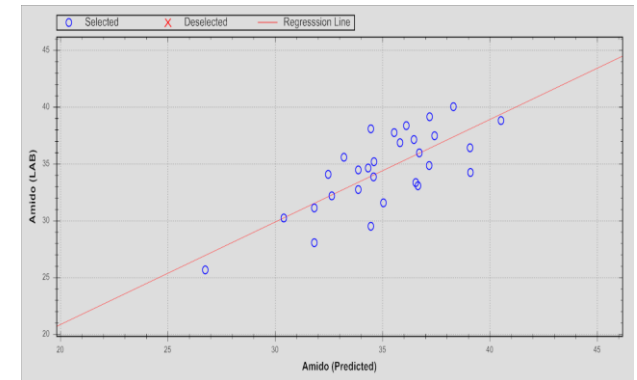
Figura 15. Comparação dos valores de PB de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.



a)

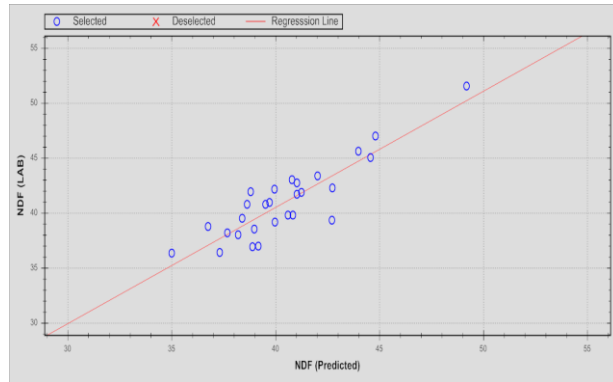


b)

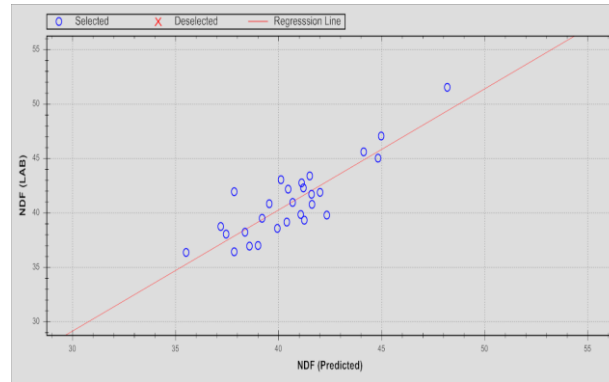


c)

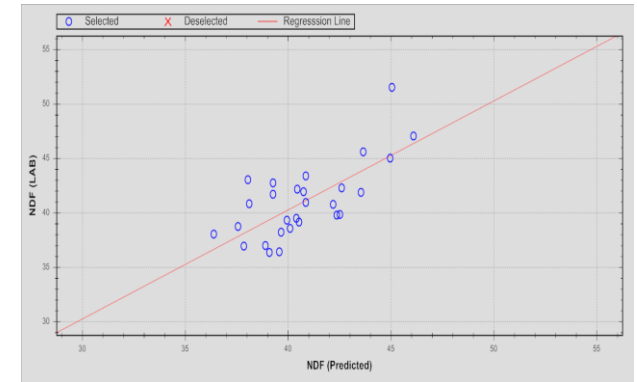
Figura 16. Comparação dos valores de amido de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.



a)

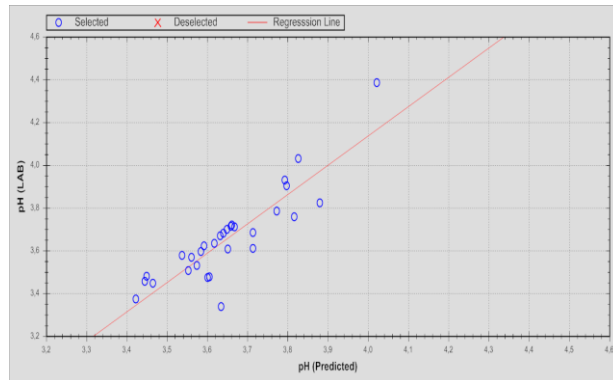


b)

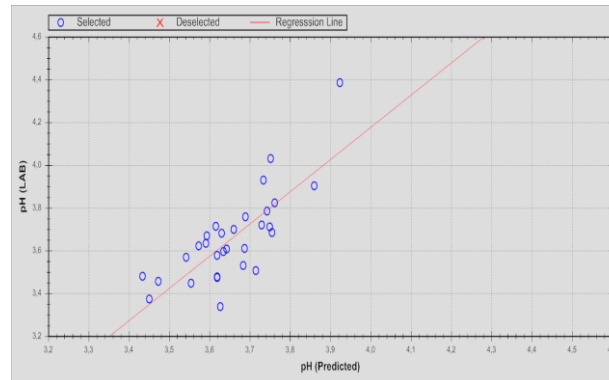


c)

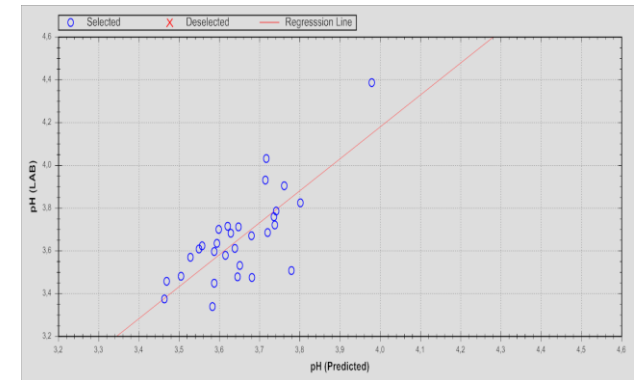
Figura 17. Comparação dos valores de NDF de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.



a)



b)



c)

Figura 18. Comparação dos valores de pH de análises de referência com os valores previstos pelo NIRS a partir da estatística de validação externa. a) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 1,4,4,1. b) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 2,4,4,1. c) Comparação dos valores aplicando MPLS juntamente com SNV-DT e 3,5,5,1.

5. Considerações finais e perspectivas futuras

A espectroscopia de infravermelho próximo associada às técnicas de quimiometria revelaram ser uma combinação eficaz na previsão das propriedades das silagens de milho. Dessa forma, esta metodologia analítica tornou-se um grande apoio às indústrias agropecuárias, pois é executada de uma forma rápida, e de baixo custo. Para além disso, esta técnica permite ainda determinar, em simultâneo, os vários parâmetros com elevada rapidez e precisão, fornecendo assim bons resultados que substituem os resultados de análises químicas e nutricionais das silagens.

Para o desenvolvimento das calibrações dos parâmetros MS, cinza total, PB, amido, NDF e pH recorreu-se à regressão MPLS juntamente com o tratamento SNV-DT. Adicionando ainda mais um tratamento utilizou-se um dos três tipos de derivadas: a primeira derivada (1,4,4,1), a segunda derivada (2,4,4,1) e a terceira derivada (3,5,5,1). Após desenvolvimento e validação das equações obtidas, verificou-se que, para cada parâmetro analisado, o melhor modelo de calibração para prever os respetivos parâmetros eram o que foram desenvolvidos por equações de primeira derivada. Isto porque todos esses apresentavam sempre um maior valor de RSQ e RPD e menor SEP na validação externa. No entanto, existe alguns desses modelos que apesar de serem classificados como os melhores do estudo, pertencem à categoria B, melhor dizendo, não representam um bom modelo, necessitando assim de recorrer outras estratégias de calibração para obter resultados mais fiáveis. Os parâmetros que se encontram nessa situação são a cinza total, a PB e o pH. Os restantes parâmetros, MS, amido, e o NDF, já possuem calibrações que possam prever os seus valores em futuras análises em silagens de milho em fresco.

Em suma, a metodologia apresentada mostrou-se ser bastante promissora para o desenvolvimento de calibrações com intuito de previsão dos parâmetros analisados, podendo-se perspetivar a sua aplicabilidade na melhoria dos modelos de categoria B e no desenvolvimento dos restantes componentes das silagens de milho em fresco, bem como desenvolvimento de calibrações noutros tipos de alimentos, nomeadamente silagens de erva e TMR.

6. Referências

- Abdi, H. (2010). Partial least squares regression and projection on latent structure regression (PLS Regression). *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(1), 97–106. <https://doi.org/10.1002/wics.51>
- Aenugu, H. P. R., Sathis Kumar, D., Srisudharson, Parthiban, N., Ghosh, S. S., & Banji, D. (2011). Near infra red spectroscopy- An overview. *International Journal of ChemTech Research*, 3(2), 825–836.
- Agelet, L. E., & Hurburgh, C. R. (2010). A tutorial on near infrared spectroscopy and its calibration. *Critical Reviews in Analytical Chemistry*, 40(4), 246–260. <https://doi.org/10.1080/10408347.2010.515468>
- AOAC. (2005). *Official Methods of Analysis*. <http://www.eoma.aoac.org/>
- Bau, N., & Lamnek, S. (2016). Multivariate analysis. *John Wiley & Sons, Ltd*, 1–3. <https://doi.org/10.1002/9781405165518.wbeosm133.pub2>
- Berzaghi, P., Flinn, P. C., Dardenne, P., Lagerholm, M., Shenk, J. S., Westerhaus, M. O., & Cowe, I. A. (2001). Comparison of linear and non-linear NIR calibration methods using large forage databases. In *Near Infrared Spectroscopy: Proceedings of the 10th International Conference* (pp. 107–114). NIR Publications.
- Bi, Y., Yuan, K., Xiao, W., Wu, J., Shi, C., Xia, J., Chu, G., Zhang, G., & Zhou, G. (2016). A local pre-processing method for near-infrared spectra, combined with spectral segmentation and standard normal variate transformation. *Analytica Chimica Acta*, 909, 30–40. <https://doi.org/10.1016/j.aca.2016.01.010>
- Blanco, M., & Villarroya, I. (2002). NIR spectroscopy: A rapid-response analytical tool. *TrAC - Trends in Analytical Chemistry*, 21(4), 240–250. [https://doi.org/10.1016/S0165-9936\(02\)00404-1](https://doi.org/10.1016/S0165-9936(02)00404-1)
- Boulet, J. C., & Roger, J. M. (2012). Pretreatments by means of orthogonal projections. *Chemometrics and Intelligent Laboratory Systems*, 117, 61–69. <https://doi.org/10.1016/j.chemolab.2012.02.002>
- Brereton, R. G. (2018). *Chemometrics Data Driven Extraction for Science* (2nd ed.). Wiley.
- Bro, R., & Andersen, C. M. (2003). Theory of net analyte signal vectors in inverse regression. *Journal of Chemometrics*, 17(12), 646–652. <https://doi.org/10.1002/cem.832>

- Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, 6, 2812–2831. <https://doi.org/10.1039/c3ay41907j>
- Chang, C.-W., Laird, D. A., Mausbach, M. J., & Hurburgh, C. R. (2001). Near-Infrared Reflectance Spectroscopy-Principal Components Regression Analyses of Soil Properties. *Soil Science Society of America Journal*, 65(2), 480–490. <https://doi.org/10.2136/sssaj2001.652480x>
- Cherian, G. (2020). *A Guide to the Principles of Animal Nutrition*. OREGON STATE UNIVERSITY CORVALLIS, OR.
- Chu, Y. W., Tang, S. S., Ma, S. X., Ma, Y. Y., Hao, Z. Q., Guo, Y. M., Guo, L. B., Lu, Y. F., & Zeng, X. Y. (2018). Accuracy and stability improvement for meat species identification using multiplicative scatter correction and laser-induced breakdown spectroscopy. 26(8), 10119–10127.
- Chun, H., & Keleş, S. (2010). Sparse partial least squares regression for simultaneous dimension reduction and variable selection. *Journal of the Royal Statistical Society. Series B: Statistical Methodology*, 72(1), 3–25. <https://doi.org/10.1111/j.1467-9868.2009.00723.x>
- Conzen, J.-P. (2014). *Multivariate Calibration. A practical guide for developing methods in the quantitative analytical chemistry* (3rd ed.). Bruker Optik GmbH.
- Cozzolino, D., Fassio, A., Fernández, E., Restaino, E., & La Manna, A. (2006). Measurement of chemical composition in wet whole maize silage by visible and near infrared reflectance spectroscopy. *Animal Feed Science and Technology*, 129(3–4), 329–336. <https://doi.org/10.1016/j.anifeedsci.2006.01.025>
- Cozzolino, D., Liu, L., Cynkar, W. U., Damberg, R. G., Janik, L., Colby, C. B., & Gishen, M. (2007). Effect of temperature variation on the visible and near infrared spectra of wine and the consequences on the partial least square calibrations developed to measure chemical composition. *Analytica Chimica Acta*, 588(2), 224–230. <https://doi.org/10.1016/j.aca.2007.01.079>
- Engel, J., Gerretzen, J., Szymanska, E., Jansen, J. J., Downey, G., Blanchet, L., & Buydens, L. M. C. (2013). Breaking with trends in pre-processing? *TrAC – Trends in Analytical Chemistry*, 50, 96–106.
- Escuredo, O., Seijo-Rodríguez, A., Inmaculada González-Martín, M., Shantal Rodríguez-Flores, M., & Carmen Seijo, M. (2018). Potential of near infrared spectroscopy for

- predicting the physicochemical properties on potato flesh. *Microchemical Journal*, 141, 451–457. <https://doi.org/10.1016/j.microc.2018.06.008>
- Esquerre, C., Gowen, A. A., Burger, J., Downey, G., & O'Donnell, C. P. (2012). Suppressing sample morphology effects in near infrared spectral imaging using chemometric data pre-treatments. *Chemometrics and Intelligent Laboratory Systems*, 117, 129–137. <https://doi.org/10.1016/j.chemolab.2012.02.006>
- Faber, N. M., & Rajkó, R. (2007). How to avoid over-fitting in multivariate calibration-The conventional validation approach and an alternative. *Analytica Chimica Acta*, 595(1-2 SPEC. ISS.), 98–106. <https://doi.org/10.1016/j.aca.2007.05.030>
- FEDNA. (2016). *Ensilado de maíz*. Fundación Española Para El Desarrollo de La Nutrición Animal. <http://www.fundacionfedna.org/forrajes/ensilado-de-maíz>
- Fernández-Cabanás, V. M., Garrido-Varo, A., Pérez-Marín, D., & Dardenne, P. (2006). Evaluation of pretreatment strategies for near-infrared spectroscopy calibration development of unground and ground compound feedingstuffs. *Applied Spectroscopy*, 60(1), 17–23. <https://doi.org/10.1366/000370206775382839>
- Ferreira, M. M. C. (2015). *Quimiometria: conceitos, métodos e aplicações*. Campinas: Ed. UNICAMP.
- Ferreira, M. M. C., Antunes, A. M., Melgo, M. S., & Volpe, P. L. O. (1999). Quimiometria I: Calibração Multivariada, um tutorial. *Química Nova*, 22(5), 724–731. <https://doi.org/10.1590/s0100-40421999000500016>
- Fontaneli, R. S., Durr, J. W., Scheffer-Basso, S. M., Haubert, F., & Bortolini, F. (2002). Validação do método da reflectância no infravermelho proximal para análise de silagem de milho. *Revista Brasileira de Zootecnia*, 31(2), 594–598. <https://doi.org/10.1590/s1516-35982002000300008>
- Garcia, P. H. de M., Calvante, J. T., Silva, D. F. da, Santos, G. C. de L., Lira, R. C., Ferreira, P. V., Beelen, P. M. G., Silva, M. T. da, Lima, G. S. de A., & Farias, I. M. S. C. de. (2020). Valor nutricional da silagem de genótipos de milho e sorgo cultivados em duas densidades de semeadura. *Zootecnia: Nutrição e Produção Animal*, 335–359. <https://doi.org/10.37885/200901491>
- Granato, D., Santos, J. S., Escher, G. B., Ferreira, B. L., & Maggio, R. M. (2018). Use of principal component analysis (PCA) and hierarchical cluster analysis (HCA) for multivariate association between bioactive compounds and functional properties in

- foods: A critical perspective. *Trends in Food Science and Technology*, 72(December 2017), 83–90. <https://doi.org/10.1016/j.tifs.2017.12.006>
- Grisanti, E., Totska, M., Huber, S., Calderon, C. K., Hohmann, M., Lingenfelter, D., & Otto, M. (2018). Dynamic localized SNV, Peak SNV, and partial peak SNV: Novel standardization methods for preprocessing of spectroscopic data used in predictive modeling. *Journal of Spectroscopy*, 2018, 1–14. <https://doi.org/10.1155/2018/5037572>
- Guo, Q., Wu, W., & Massart, D. L. (1999). The robust normal variate transform for pattern recognition with near-infrared data. *Analytica Chimica Acta*, 382(1–2), 87–103. [https://doi.org/10.1016/S0003-2670\(98\)00737-5](https://doi.org/10.1016/S0003-2670(98)00737-5)
- Guo, T., Dai, L., Yan, B., Lan, G., Li, F., Li, F., Pan, F., & Wang, F. (2021). Measurements of chemical compositions in corn stover and wheat straw by near-infrared reflectance spectroscopy. *Animals*, 11(3328), 1–12. <https://doi.org/10.3390/ani11113328>
- Hageman, J. A., Westerhuis, J. A., & Smilde, A. K. (2005). Temperature robust multivariate calibration: An overview of methods for dealing with temperature influences on near infrared spectra. *Journal of Near Infrared Spectroscopy*, 13(2), 53–62. <https://doi.org/10.1255/jnirs.457>
- Hayati, R., Munawar, A. A., & Fachruddin, F. (2020). Enhanced near infrared spectral data to improve prediction accuracy in determining quality parameters of intact mango. *Data in Brief*, 30(1–14). <https://doi.org/10.1016/j.dib.2020.105571>
- Heinz W. Siesler. (2008). Basic Principles of Near-Infrared Spectroscopy. In C. E. W. Burns Donald A. (Ed.), *Handbook of Near-Infrared Analysis* (pp. 1–19).
- Hibbert, D. B. (2016). Vocabulary of concepts and terms in chemometrics (IUPAC Recommendations 2016). *Pure and Applied Chemistry*, 88(4), 407–443. <https://doi.org/10.1515/pac-2015-0605>
- Hofer, M. (2017). Mean Centering. *John Wiley & Sons, Inc.*, 1–3. <https://doi.org/10.1002/9781118901731.iecrm0137>
- Ikoyi, A. Y., & Younge, B. A. (2020). Influence of forage particle size and residual moisture on near infrared reflectance spectroscopy (NIRS) calibration accuracy for macro-mineral determination. *Animal Feed Science and Technology*, 270, 1–13. <https://doi.org/10.1016/j.anifeedsci.2020.114674>
- Jiao, Y., Li, Z., Chen, X., & Fei, S. (2020). Preprocessing methods for near-infrared

spectrum calibration. *Journal of Chemometrics*, 34(11), 1–19. <https://doi.org/10.1002/cem.3306>

- Judy, J. V., Bachman, G. C., Brown-Brandl, T. M., Fernando, S. C., Hales, K. E., Harvatine, K. J., Miller, P. S., & Kononoff, P. J. (2019). Increasing the concentration of linolenic acid in diets fed to Jersey cows in late lactation does not affect methane production. *Journal of Dairy Science*, 102(3), 2085–2093. <https://doi.org/10.3168/jds.2018-14608>
- Júnior, gilberto de L. M., Zanine, A. de M., Borges, I., & Pérez, J. R. O. (2007). Qualidade da fibra para a dieta de ruminantes. *Ciência Animal*, 17(1), 7–17.
- Karoui, R., & De Baerdemaeker, J. (2007). A review of the analytical methods coupled with chemometric tools for the determination of the quality and identity of dairy products. *Food Chemistry*, 102(3), 621–640. <https://doi.org/10.1016/j.foodchem.2006.05.042>
- Kramer, R. (1998). Chemometric Techniques for Quantitative Analysis. In *Marcel Dekker, Inc.*
- Kumaravelu, C., & Gopal, A. (2015). A review on the applications of Near-Infrared spectrometer and Chemometrics for the agro-food processing industries. *IEEE International Conference on Technological Innovations in ICT for Agriculture and Rural Development*, 8–12. <https://doi.org/10.1109/TIAR.2015.7358523>
- Lage, A. (2017). Silagens de milho produzidas em portugal - Avaliação da qualidade nutricional. *Ruminantes*, 16–18.
- Lage, A. (2018). Silagens de erva : avaliação da qualidade nutricional. *Ruminantes*, 28–30.
- Lage, A. (2019). A importância da maturidade na qualidade nutricional das silagens de milho. *Ruminantes*, 34–37.
- Lage, A., & Salgueiro, M. D. (2019). A variação do teor em proteína das silagens de milho em portugal. *Ruminantes*, 52–54.
- Lin Mengshi, Rasco, B. A., Cavinato, A. G., & Al-Holy, M. (2009). Infrared (IR) Spectroscopy—Near- Infrared Spectroscopy and Mid-Infrared Spectroscopy. In *Infrared Spectroscopy for Food Quality Analysis and Control* (Vol. 43, Issue 6, pp. 119–143).
- Lourenço, N. D., Menezes, J. C., Pinheiro, H. M., & Diniz, D. (2008). Development of PLS calibration models from UV-Vis spectra for TOC estimation at the outlet of a fuel park wastewater treatment plant. *Environmental Technology*, 29(8), 891–898.

<https://doi.org/10.1080/09593330802015581>

- 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. *Journal of Chemometrics*, 33(4), 1–10. <https://doi.org/10.1002/cem.3111>
- Manley, M., Downey, G., & Baeten, V. (2008). Spectroscopic Technique: Near- infrared (NIR) Spectroscopy. In *Modern Techniques for Food Authentication* (pp. 65–116).
- Mark, H., & Jr., J. W. (2007). *Chemometrics in Spectroscopy*. Elsevier Inc.
- McDonald, P., Edwards, R. A., Greenhalgh, J. F. D., Morgan, C. A., Sinclair, L. A., & Wilkinson, R. G. (2011). Animal Nutrition. In *Pearson* (7th ed., Vol. 111, Issue 2793). <https://doi.org/10.1038/111651a0>
- Mcgrath, J., Duval, S. M., Tamassia, L. F. M., Kindermann, M., Stemmler, R. T., Gouvea, V. N. De, Acedo, T. S., Immig, I., Williams, S. N., & Celi, P. (2018). Nutritional strategies in ruminants: A lifetime approach. *Research in Veterinary Science*, 116, 28–39. <https://doi.org/10.1016/j.rvsc.2017.09.011>
- Metrohm. (2014). *A guide to near-infrared spectroscopic analysis of industrial manufacturing processes*. 1–43.
- Möller, J. (2014). Comparing methods for fibre determination in food and feed. *Foss*, 1, 1–6.
- Nilsson, J. (1998). Introduction to Chemometrics and Statistics. In *Multiway calibration in 3D QSAR: applications to dopamine receptor ligands* (pp. 35–52).
- Nobari-Moghaddam, H., Tamiji, Z., Akbari-Lakeh, M., Khoshayand, M. R., & Haji-Mahmoodi, M. (2021). Multivariate analysis of food fraud: A review of NIR based instruments in tandem with chemometrics. *Journal of Food Composition and Analysis*, 107(104343), 1–19. <https://doi.org/10.1016/j.jfca.2021.104343>
- NP. (2007). Alimentos para animais. Determinação do teor de azoto e cálculo do teor de proteína bruta (ISO 5983). *Instituto Português Da Qualidade*.
- NP. (2014). Alimentos para animais. Determinação da cinza bruta (ISO 5984). *Instituto Português Da Qualidade*.
- NP. (2018a). Alimentação para animais. Determinação da humidade e de outros teores de matérias votáveis (ISO 6496). *Instituto Português Da Qualidade*.

- NP. (2018b). Alimentos para animais, cereais e produtos à base de cereais moídos Diretrizes para a aplicação da espectrometria no infravermelho próximo (ISO 12099). Instituto Português Da Qualidade.
- Otto, M. (2017). *Chemometrics Statistics and Computer Application in Analytical Chemistry* (3rd ed.). Wiley-VCH.
- Pasquini, C. (2003). Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. *Journal of the Brazilian Chemical Society*, 14(2), 198–219. <https://doi.org/10.1590/S0103-50532003000200006>
- Pasquini, C. (2018a). Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36. <https://doi.org/10.1016/j.aca.2018.04.004>
- Pasquini, C. (2018b). Princípios da Espectroscopia no Infravermelho Próximo. In Casiane Salete Tibola, E. P. de Medeiros, M. L. F. Simeone, & M. A. de Oliveira (Eds.), *Espectroscopia no Infravermelho Próximo para Avaliar Indicadores de Qualidade Tecnológica e Contaminantes em Grãos* (pp. 13–30).
- Penner, M. H. (2017). Basic Principles of Spectroscopy. In S. S. Nielsen (Ed.), *Food Analysis* (pp. 79–88). https://doi.org/10.1007/978-3-319-45776-5_6
- Porep, J. U., Kammerer, D. R., & Carle, R. (2015). On-line application of near infrared (NIR) spectroscopy in food production. *Trends in Food Science and Technology*, 46(2), 211–230. <https://doi.org/10.1016/j.tifs.2015.10.002>
- Qu, J. H., Liu, D., Cheng, J. H., Sun, D. W., Ma, J., Pu, H., & Zeng, X. A. (2015). Applications of Near-infrared Spectroscopy in Food Safety Evaluation and Control: A Review of Recent Research Advances. *Critical Reviews in Food Science and Nutrition*, 55(13), 1939–1954. <https://doi.org/10.1080/10408398.2013.871693>
- Rinnan, Å. (2014). Pre-processing in vibrational spectroscopy-when, why and how. *Analytical Methods*, 6(18), 7124–7129. <https://doi.org/10.1039/c3ay42270d>
- Rinnan, Å., Berg, F. van den, & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28(10), 1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Rodionova, O., Kucheryavskiy, S., & Pomerantsev, A. (2021). Efficient tools for principal component analysis of complex data— a tutorial. *Chemometrics and Intelligent*

Laboratory Systems, 213(104304), 1–11.
<https://doi.org/10.1016/j.chemolab.2021.104304>

- Rushing, J. B., Saha, U. K., Lemus, R., Sonon, L., & Baldwin, B. S. (2016). Analysis of Some Important Forage Quality Attributes of Southeastern Wildrye (&i>Elymus glaberrimus) Using Near-Infrared Reflectance Spectroscopy. *American Journal of Analytical Chemistry*, 07(09), 642–662.
<https://doi.org/10.4236/ajac.2016.79060>
- Schingoethe, D. J. (2017). A 100-Year Review: Total mixed ration feeding of dairy cows. *Journal of Dairy Science*, 100(12), 10143–10150. <https://doi.org/10.3168/jds.2017-12967>
- Segtnan, V. H., Šašić, Š., Isaksson, T., & Ozaki, Y. (2001). Studies on the structure of water using two-dimensional near-infrared correlation spectroscopy and principal component analysis. *Analytical Chemistry*, 73(13), 3153–3161. <https://doi.org/10.1021/ac010102n>
- Sena, M. M., & Almeida, M. R. (2018). Quimiometria Aplicada aos Dados Espectrais no Infravermelho Próximo. In C. S. Tibola, E. P. de Medeiros, M. L. F. Simeone, & M. A. de Oliveira (Eds.), *Espectroscopia no Infravermelho Próximo para Avaliar Indicadores de Qualidade Tecnológica e Contaminantes em Grãos* (pp. 31–50).
- Sena, M. M., Almeida, M. R., Braga, J. W. B., & Poppi, R. J. (2017). Multivariate statistical analysis and chemometrics. In *Spectroscopic methods in food analysis* (pp. 273–314). Boca Raton: CRC Press.
- Sena, M. M., Tokeshi, H., Valarini, P., Frighetto, R. T., & Poppi, R. (2002). Discrimination of management effects on soil parameters by using principal component analysis: a multivariate analysis case study. *Soil & Tillage Research*, 67(2), 171–181.
- Shah, M., Bhatt, C., & Dave, J. (2016). Temperature dependent NIR spectroscopic analysis for food grain samples. *Indian Journal of Science and Technology*, 9(9), 1–4.
<https://doi.org/10.17485/ijst/2016/v9i9/85306>
- Shao, X., Bian, X., Liu, J., Zhang, M., & Cai, W. (2010). Multivariate calibration methods in near infrared spectroscopic analysis. *Analytical Methods*, 2(11), 1662–1666.
<https://doi.org/10.1039/c0ay00421a>
- Sharifi, A. (2018). Partial Least Squares-regression (PLS-regression) In Chemometrics. *1th National Conference on Achievements in Chemistry and Chemical Engineering, October*, 305–308.

- Shenk, J. S., Workman, J. J., Westerhaus, M. O., & Jr. (2008). Application of NIR Spectroscopy to Agricultural Products. In *Handbook of Near-Infrared Analysis* (pp. 347–386).
- 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. *Vibrational Spectroscopy*, 97(April), 55–65. <https://doi.org/10.1016/j.vibspec.2018.05.002>
- Soest, P. J. Van. (2018). *Nutritional Ecology of the Ruminant* (2nd ed.). Cornell University Press.
- Sørensen, L. K. (2004). Prediction of fermentation parameters in grass and corn silage by near infrared spectroscopy. *Journal of Dairy Science*, 87(11), 3826–3835. [https://doi.org/10.3168/jds.S0022-0302\(04\)73522-5](https://doi.org/10.3168/jds.S0022-0302(04)73522-5)
- Sun, D. W. (2009). *Infrared Spectroscopy for Food Quality Analysis and Control* (1st ed.). Elsevier Science.
- 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. *Journal of Agricultural and Food Chemistry*, 55, 8331–8338. <https://doi.org/10.1021/jf071538s>
- Wedelsbäck, H. (2015). in-Line Measurement With Nir-Advantages and Limitations. *Measurement of Moisture in Materials – Tribble M Conference*.
- Williams, P. (2015). How did it all start? *PROCEEDINGS OF THE 17TH INTERNATIONAL CONFERENCE ON NEAR INFRARED SPECTROSCOPY*, 128–132.
- Windig, W., Shaver, J., & Bro, R. (2008). Loopy MSC: A simple way to improve multiplicative scatter correction. *Applied Spectroscopy*, 62(10), 1153–1159. <https://doi.org/10.1366/000370208786049097>
- Workman, J. J. (1996). Interpretive spectroscopy for near infrared. In *Applied Spectroscopy Reviews* (Vol. 31, Issue 3, pp. 251–320). <https://doi.org/10.1080/05704929608000571>
- Wu, G. (2018). *Principles of Animal Nutrition*. Taylor & Francis Group, LLC.
- Yi, J., Sun, Y., Zhu, Z., Liu, N., & Lu, J. (2017). Near-infrared reflectance spectroscopy for the prediction of chemical composition in walnut kernel. *International Journal of Food*

Properties, 20(7), 1633–1642. <https://doi.org/10.1080/10942912.2016.1217006>