

UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Exatas

Departamento de Química

Ana Carolina da Costa Fulgêncio

**COMBINANDO MODELOS QUIMIOMÉTRICOS E PORTABILIDADE
PARA O CONTROLE DE QUALIDADE DE CERVEJA E OUTRAS
BEBIDAS FERMENTADAS**

Belo Horizonte

2023

UFMG/ICEX/DQ. 1.534
T. 698

Ana Carolina da Costa Fulgêncio

**COMBINANDO MODELOS QUIMIOMÉTRICOS E PORTABILIDADE PARA O
CONTROLE DE QUALIDADE DE CERVEJA E OUTRAS BEBIDAS
FERMENTADAS**

Tese apresentada ao Departamento de
Química do Instituto de Ciências Exatas
da Universidade Federal de Minas Gerais
como requisito parcial para a obtenção do
grau de Doutor em Ciências – Química.

Orientador: Marcelo Martins de Sena

Coorientador: Bruno Gonçalves Botelho

Belo Horizonte

2023

Ficha Catalográfica

F963c
2023
T

Fulgêncio, Ana Carolina da Costa.

Combinando modelos quimiométricos e portabilidade para o controle de qualidade de cerveja e outras bebidas fermentadas [manuscrito] / Ana Carolina da Costa Fulgêncio. 2023.

173 f. : il., gráfs., tabs.

Orientador: Marcelo Martins de Sena.

Coorientador: Bruno Gonçalves Botelho.

Tese (doutorado) – Universidade Federal de Minas Gerais – Departamento de Química.

Bibliografia: f. 120-136.

Apêndices: f. 137-173.

Encontram-se nos apêndices 4 (quatro) artigos da autora intitulados: Development of a simple and rapid method for color determination in beers using digital images -- Determination of alcohol content in beers of different styles based on portable near-infrared spectroscopy and multivariate calibration - - Screening method for the rapid detection of diethylene glycol in beer based on chemometrics and portable near-infrared spectroscopy -- Combining portable NIR spectroscopy and multivariate calibration for the determination of ethanol in fermented alcoholic beverages by a multi-product model.

1. Química analítica – Teses. 2. Cerveja – Teses. 3. Quimiometria – Teses. 4. Cerveja – Indústria – Controle de qualidade – Teses. 5. Mínimos quadrados – Teses. 6. Calibração – Teses. 7. Espectroscopia de infravermelho – Teses. 8. Bebidas fermentadas – Indústria – Teses. 9. Espectrofotômetro – Teses. 10. Análise de componentes principais – Teses. 11. Processamento de imagens – Técnicas digitais – Teses. 12. Análise multivariada – Teses. 13. Preparação de amostra (Química) – Teses. I. Sena, Marcelo Martins de, Orientador. II. Sena, Marcelo Martins de, Coorientador. III. Título.

CDU 043



UNIVERSIDADE FEDERAL DE MINAS GERAIS



"Combinando Modelos Quimiométricos e Portabilidade para o Controle de Qualidade de Cerveja e Outras Bebidas Fermentadas"

Ana Carolina da Costa Fulgêncio

Tese aprovada pela banca examinadora constituída pelos Professores:

Prof. Marcelo Martins de Sena - Orientador
UFMG

Prof. Bruno Gonçalves Botelho - Coorientador
UFMG

Profa. Cristina Malegori
University of Genoa

Profa. Patrícia Valderrama
UTFPR

Profa. Scheilla Vitorino Carvalho Souza Ferreira
UFMG

Prof. Rodinei Augusti
UFMG

Belo Horizonte, 16 de março de 2023.



Documento assinado eletronicamente por **Bruno Goncalves Botelho, Professor do Magistério Superior**, em 16/03/2023, às 17:06, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Rodinei Augusti, Professor do Magistério Superior**, em 16/03/2023, às 17:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Patrícia Valderrama, Usuária Externa**, em 16/03/2023, às 17:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Marcelo Martins de Sena, Professor do Magistério Superior**, em 16/03/2023, às 17:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Cristina Malegori, Usuária Externa**, em 16/03/2023, às 17:07, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Scheilla Vitorino Carvalho de Souza Ferreira, Professora do Magistério Superior**, em 16/03/2023, às 17:08, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2147411** e o código CRC **5F1F0E3C**.

AGRADECIMENTOS

Eu não consigo acreditar que finalmente estou escrevendo os agradecimentos da minha tese de doutorado! Eu finalmente posso dormir sem pensar em metodologias, análises de dados e revisões bibliográficas. Agradeço sinceramente a todas as pessoas que contribuíram para a conclusão bem-sucedida da minha tese de doutorado. Eu nunca teria conseguido sem a ajuda e apoio de todos vocês.

Em primeiro lugar, minha família merece um agradecimento especial. Rosa, Luiz, Pedro, Malu, Nina e Davidson, vocês sempre me apoiaram e me encorajaram a perseguir meus sonhos. Sem o amor e o apoio de vocês, eu não estaria aqui hoje.

Também gostaria de agradecer aos meus amigos de longa data da UFV - Paula, Naty, Carol e Junim - pela ajuda e incentivo ao longo de minha jornada acadêmica. Suas palavras de encorajamento e sua amizade significaram muito para mim.

Aos meus amigos da UFMG - Glaucimar (ou Jimmy, como ele é mais conhecido), Keila, Denise, Mariana de PL, Julia, Matheus, Victoria e Marina -, obrigado pelo apoio, amizade e pelas risadas compartilhadas durante os momentos difíceis. Vocês tornaram a minha jornada acadêmica muito mais agradável e me ajudaram a superar os obstáculos que surgiram no caminho.

Não posso esquecer de mencionar o meu co-orientador Bruno, que sempre me apoiou e me ajudou a manter o foco na minha pesquisa. Sua contribuição foi fundamental para a conclusão da minha tese.

Por fim, gostaria de agradecer ao meu orientador Marcelo por sua orientação, paciência e pelo seu comprometimento com a minha pesquisa. Seus insights foram inestimáveis para a minha pesquisa e eu serei eternamente grato por sua orientação.

E, claro, não posso deixar de mencionar meu amigo Jimmy, que merece um agradecimento especial. Jimmy, sua ajuda e apoio incondicionais durante todo o processo foram essenciais para o sucesso deste trabalho. Sua paciência, conhecimento e senso de humor foram inestimáveis e fizeram com que este trabalho fosse muito mais do que uma simples tese de doutorado. Muito obrigado!

Por fim, gostaria de agradecer aos amigos da vida - Paula e Thiago - por sempre me lembrarem de que a vida não é apenas estudo e trabalho. Obrigada :)

A todos vocês, obrigado por fazerem parte da minha jornada acadêmica. Estou profundamente grato por tudo o que fizeram por mim e espero que possamos manter a amizade por muitos anos.

*"Dê ao meu povo suficiente cervejas,
boas cervejas a preço acessível e não
haverá revolução partindo deles"*

Rainha Victória (1819-1901)

Resumo

Essa tese propôs o uso de portabilidade e quimiometria no desenvolvimento de novos métodos de análise para o controle de qualidade da cerveja. O mercado cervejeiro movimenta uma enorme rede que agrega desde a pesquisa, o cultivo, o processamento e a comercialização de insumos e matérias-primas até a entrega do produto finalizado. Sendo assim, é necessário considerar o controle de qualidade como uma etapa fundamental na indústria cervejeira. Neste estudo, novos métodos analíticos simples, rápidos, robustos e confiáveis foram desenvolvidos. Por ser a primeira característica sensorial das cervejas que chama a atenção do consumidor, a cor foi escolhida como o parâmetro inicial a ser determinado. O método desenvolvido combinou imagens digitais adquiridas através de um telefone celular e calibração multivariada (PLS, mínimos quadrados parciais). As próximas duas aplicações tiveram como objetivo a determinação do teor de álcool, porém usando estratégias e matrizes distintas. Para isso foi utilizado um espectrofotômetro portátil na região do infravermelho próximo em combinação com calibração multivariada (PLS). A segunda aplicação focou na indústria cervejeira, desenvolvendo assim um método para quantificar o teor alcóolico em cervejas de diferentes estilos. Já a terceira aplicação objetivou desenvolver um método a ser usado por órgãos de fiscalização, envolvendo um modelo quantitativo multiproduto para 4 bebidas fermentadas (cerveja, vinho, cidra e hidromel). A última aplicação combinou espectroscopia NIR portátil e PLS discriminante (PLS-DA) para desenvolver um método de triagem rápido e de interesse forense para a detecção direta de dietilenoglicol em cerveja. Este método foi inspirado em evento trágico recente de envenenamento de pessoas, incluindo mortes, por consumo de cerveja contaminada com esta substância. A validação analítica multivariada foi feita para atestar a eficiência de todos os métodos desenvolvidos. A disseminação da portabilidade abre novas perspectivas, modificando a forma como os usuários se comunicam com a ciência, abrindo novos mercados, com a possibilidade de incorporação em dispositivos cotidianos e aparelhos inteligentes, tornando-se nos próximos anos parte da Internet das Coisas, e contribuindo para a democratização da química analítica.

Palavras chaves: Cerveja; Quimiometria; Controle de qualidade cervejeiro; Portabilidade; Miniaturização

Abstract

This thesis combined portability and chemometrics in the development of new analytical methods for the quality control of beer. The beer market moves an extensive network that aggregates from the research, cultivation, processing, and marketing of inputs and raw materials to the delivery of the finished product. Given this fact, it is necessary to consider quality control as a fundamental step in the brewing industry. In this study, new simple, rapid, robust, and reliable analytical were developed. Color is the first sensory characteristic of beer that draws the consumer's attention, and thus was chosen as the initial parameter to be determined. The developed method combined digital images recorded by a smartphone and multivariate calibration (PLS, partial least squares) for the color determination of beers. The following two applications aimed at determining the alcohol content using different strategies and matrices. For this aim, a portable near-infrared (NIR) spectrophotometer was applied in combination with multivariate calibration (PLS). The second application focused on the brewing industry, thus developing a method for determining the alcohol content in beers of different styles. On the other hand, the third application focused on regulatory issues, thus developing a multi-product quantitative model for four fermented beverages (beer, wine, cider and mead). The last application combined portable NIR spectroscopy and PLS discriminant analysis (PLS-DA) to develop a fast screening method for the direct detection of diethylene glycol in beer. This is a forensic application based on a recent tragic occurrence of poisoning of dozens of people, including some deaths, by consuming beer contaminated with this substance. Multivariate analytical validation was carried out for all developed methods. The dissemination of portability opens up new perspectives. It can change the way users communicate with science, opening new markets and being able to be incorporated into consumer devices and smart appliances, becoming part of the Internet of Things in the coming years, and contributing to the democratization of analytical chemistry.

Keywords: Beer; Chemometrics; Brewery; Portability; Miniaturization.

LISTA DE FIGURAS

Figura 1. Modelo que descreve a preparação de pão e cerveja. Peça exposta no Museu da Antiguidade do Cairo { http://www.touregypt.net/featurestories/picture10082003.htm }.	25
Figura 2. (a) Gráfico de pizza contendo a participação de cada continente e (b) Mapa da produção mundial de cevada entre os anos de 2010 e 2017 segundo a Organização das Nações Unidas para Alimentação e Agricultura (FAO) (Adaptado de [17]).	29
Figura 3. Diagrama dos componentes do grão de cevada (Adaptado de [18])	31
Figura 4. Mapa Mundial da produção mundial de lúpulo entre os anos de 2010 e 2017 segundo a Organização das Nações Unidas para Alimentação e Agricultura (FAO) (Adaptado de [17]).	33
Figura 5. Esquema do espectrômetro portátil na região do infravermelho próximo MicroNIR® (Adaptado de [42]).	37
Figura 6. (a) Sistema usado para adquirir as imagens digitais. A câmera de madeira fabricada juntamente com sistema de iluminação; nela está uma pequena placa de Petri contendo 25 mL de amostra de cerveja (b) Uma imagem adquirida de uma placa de Petri como exemplo.....	52
Figura 7. Gráfico dos resíduos das amostras de calibração e validação para o modelo PLS global. Amostras de calibração (círculos pretos) e validação (triângulos vermelhos).....	54
Figura 8. Valores previstos de Y para o modelo PLS-DA. A linha tracejada horizontal indica o limite de classificação e a linha tracejada vertical indica a separação entre as amostras de treinamento e as amostras de teste. Amostras da faixa de coloração superior (triângulos vermelhos) e mostras da faixa de coloração inferior (asteriscos verdes)	56
Figura 9. Vetores informativos para o modelo PLS-DA. (a) vetores de regressão e (b) VIP escores.....	57
Figura 10. Valores de referência versus valores previstos para os modelos PLS (a) Modelo da faixa superior de SRM. (b) Modelo da faixa inferior de SRM. Amostras de calibração (círculos pretos) e amostras de validação (triângulos vermelhos)	60
Figura 11. Vetores de regressão para o modelo PLS. (a) Modelo para a escala superior de SRM e (b) modelo para a escalar inferior de SRM.....	62
Figura 12. Gráfico da região de confiança elíptica para a inclinação e intercepto da curva dos valores previstos <i>versus</i> de referência para os modelos PLS (a) faixa superior de SRM (b) faixa inferior de SRM. Símbolos: ponto preto para o ponto ideal e elipse vermelha da EJCR.	64

Figura 13. Intervalos de confiança para a previsão das amostras de validação, estimados baseando-se no SPE (<i>Standard Prediction Errors</i>). (a) modelo faixa superior de SRM (b) modelo faixa inferior de SRM. Valores previstos (triângulos) e valores de referência (círculos).	66
Figura 14. Curvas de calibração, sendo a curva contendo a matriz está representada por círculos vermelhos, já a curva de calibração no solvente está representada por triângulos azuis.	78
Figura 15. Espectros de NIR das 92 amostras de cervejas.	79
Figura 16. Gráfico de valores de referência versus valores previstos para o modelo PLS. Símbolos: círculos pretos para amostras de calibração, triângulos vermelhos para amostras de validação.	81
Figura 17. (a) Vetores de Regressão (b) VIP Escores para o modelo PLS.	82
Figura 18. Gráfico da região de confiança elíptica para a inclinação e intercepto da curva dos valores previstos <i>versus</i> de referência para o modelo PLS. Símbolos: círculo preto para o ponto ideal e elipse vermelha para EJCR.	84
Figura 19. Espectros NIR pré-processados por alisamento Savitzky-Golay e variação normal padrão (SNV) de todas as 153 amostras de bebidas analisadas.	93
Figura 20. Gráfico de teor alcoólico de referência <i>versus</i> valores previstos para o modelo PLS desenvolvido. (a) Representação em função de amostras de calibração e validação. Círculos vazios indicam amostras de calibração e triângulos cheios indicam amostras de validação. (b) Representação em função do tipo de bebida. Cruzes cinzas representam amostras de cerveja, amostras de sidra representadas por asteriscos verdes, amostras de hidromel por quadrados azuis e amostras de vinho por triângulos vermelhos.	95
Figura 21. Vetores informativos para o modelo PLS. (a) Escores VIP. A linha tracejada horizontal indica o limite de 1,0, acima do qual as variáveis foram consideradas como contribuindo significativamente para o modelo PLS. (b) Vetor de regressão.	97
Figura 22. (a) Uma região elíptica de confiança conjunta (EJCR) estimada no nível de confiança de 95% para a linearidade (inclinação <i>versus</i> intercepto) do modelo PLS desenvolvido. (b) Intervalos de confiança estimados para a previsão das amostras de validação. Os erros de predição padrão específicos da amostra (SPE) foram calculados com base na equação de erros em variáveis (EIV). Triângulos cinzas significam valores previstos e quadrados vermelhos significam os respectivos valores de referência.	100
Figura 23. Sistema usado para adquirir os espectros. Um suporte de acrilonitrila-butadieno-estireno especialmente projetado e construído para esta aplicação com uma impressora 3D, ao lado um Eppendorf de 2 mL contendo 1 mL de amostra de cerveja.	108

Figura 24. Espectros NIR pré-processados por alisamento Savitzky-Golay e variação normal padrão (SNV) de todas as 200 amostras de cerveja analisadas.	110
Figura 25. Valores de Y previstos para o modelo PLS-DA construído para discriminar amostras contaminadas e não contaminadas de DEG. A linha tracejada horizontal vermelha indica o limite Bayesiano e a linha tracejada vertical indica a separação entre as amostras de treinamento e de teste. Triângulos cheios indicam amostras não contaminadas (autênticas) e círculos vazios amostras contaminadas com DEG.....	112
Figura 26. Vetores informativos do modelo PLS-DA. (a) Escores (VIP). A linha pontilhada horizontal preta indica o limite de 1,0, acima do qual as variáveis são consideradas significativas para o modelo. (b) Coeficientes de regressão.....	114
Figura 27. (a) Espectro NIR de uma amostra DEG pura. (b) <i>Loadings</i> de VL2 do modelo PLS-DA, que foi atribuído a DEG.	115
Figura 28. Curva de probabilidade de detecção (POD) obtida por regressão logística, com resultados positivos em porcentagem e concentração em mg L ⁻¹	117

LISTA DE TABELAS

Tabela 1. Descrição dos parâmetros e suas fórmulas	47
Tabela 2. Tabela contendo as seguintes FOM: FPR (taxa de falso positivo), FNR (taxa de falso negativo), STR (taxa de sensibilidade), SPR (taxa de especificidade) e EFR (taxa de eficiência).....	58
Tabela 3. Resultados da otimização do modelo PLS da faixa superior de escala SRM desenvolvido através da detecção e remoção de <i>outliers</i>	59
Tabela 4. Resultados da otimização do modelo PLS da faixa inferior de escala SRM desenvolvido através da detecção e remoção de <i>outliers</i>	59
Tabela 5. Parâmetros para avaliação das figuras de mérito dos modelos PLS	63
Tabela 6. Amostras de cerveja usadas na construção e validação do modelo quimiométrico.	73
Tabela 7. Figuras de mérito para a determinação cromatográfica de álcool em um GC-FID.	77
Tabela 8. Resultados para a otimização do modelo PLS desenvolvido através da detecção e remoção de <i>outliers</i>	80
Tabela 9. Resultados para FOM (precisão e linearidade) do modelo PLS.	83
Tabela 10. Resultados para a otimização do modelo PLS desenvolvido através da detecção e remoção de <i>outliers</i>	94
Tabela 11. Outros FOM quantitativos estimados para o modelo PLS desenvolvido.	99
Tabela 12. Resultados para a otimização do modelo PLS-DA desenvolvido através da detecção e remoção de <i>outliers</i>	111
Tabela 13. FOM qualitativas estimadas para o modelo PLS-DA. Unidades em %.	116

LISTA DE ABREVIATURAS

ANOVA - Análise de variância

ASBC - Sociedade americana dos químicos cervejeiros (*American Society of Brewing Chemists*)

ATR – Reflectância Total Atenuada (*Attenuated Total Reflectance*)

CMYK – Sistema de cor ciano, magenta, amarelo e preto (*Cyan, Magenta, Yellow, Key/Black*)

CV - Validação Cruzada (*Cross Validation*)

CVAC - Procedimentos Analíticos baseados em Visão Computacional (*Computer Vision Algorithm Collection*)

CVCE - erro de classificação cruzada (*cross-validated classification error*)

DEG - Dietilenoglicol

EBC - Convenção de Cervejeiros da Europa (*European Brewing Convention*)

EFR - Taxa de eficiência (*Efficiency Rate*)

EIV - Erros em variáveis (*Errors In Variables*)

EJCR - Regiões Elípticas de Confiança Conjunta (*Elliptical Joint Confidence Regions*)

EUA - Estados Unidos da América

FAO - Organização das Nações Unidas para Alimentação e Agricultura (*Food and Agriculture Organization of the United Nations*)

FFT - Transformada Rápida de Fourier (*Fast Fourier Transform*)

FNR - Taxa de falso negativo (*False Negative Rate*)

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

FPR - Taxa de falso positivo (*False Positive Rate*)

FTIR - Espectroscopia no Infravermelho com Transformada de Fourier (*Fourier-transform infrared spectroscopy*)

GC - Cromatografia em Fase Gasosa (*Gas Chromatography*)

HSV – Sistema de cor matiz, saturação e valor/brilho (*Hue, Saturation and Value*)

IBU - Unidade internacional de amargor (*International Bitterness Unit*)

INMETRO - Instituto Nacional de Metrologia, Qualidade e Tecnologia

IPA – *India Pale Ale*

IS - Padrão Interno (*Internal Standard*)

LD - Limite teórico de detecção

LQ - Limite teórico de quantificação

VL - Variável Latente (*Latent Variable*)

MAPA – Ministério da Agricultura, Pecuária e Abastecimento

MIA - Análise Multivariada de Imagens (*Multivariate Image Analysis*)

MIR - Regressão Multivariada de Imagens (*Multivariate Images Regression*)

MSC – Correção de espalhamento multiplicativo (*Multiplicative Scatter Correction*)

NIPALS - Regressão Parcial de Mínimos Quadrados (*Nonlinear Iterative Partial Least Squares*)

NIR – Infravermelho próximo (*Near Infrared*)

NIRS – Espectroscopia no Infravermelho próximo (*Near Infrared Spectroscopy*)

OLS - Mínimos Quadrados Ordinários (*Ordinary Least Squares*)

PCA - Análise de Componentes Principais (*Principal Component Analysis*)

PLS - Regressão Parcial de Mínimos Quadrados (*Partial Least Squares*)

PLS-DA - Análise Discriminante por Mínimos Quadrados Parciais (*Partial Least Squares - Discriminant Analysis*)

RGB – Sistema de cor vermelho, verde, azul (*Red, Green, Blue*)

RMSE - Raiz quadrada do erro-médio (*Root Mean Square Error*)

RMSEC - Raiz quadrada do Erro Médio Quadrático da Calibração (*Root Mean Square Errors of Calibration*)

RMSECV - Raiz quadrada do Erro Médio Quadrático da Validação Cruzada (*Root Mean Square Errors of Cross Validation*)

RMSEP - Raiz quadrada do Erro Médio Quadrático da Previsão (*Root Mean Square Errors of Validation*)

RPD – Relação de Desempenho do Desvio (*Residual Prediction Deviation*)

RSD - Desvio Padrão Relativo (*Relative Standard Deviation*)

RSDR - Desvio Padrão Relativo nas condições de Precisão Intermediária

RSDr - Desvio Padrão Relativo nas condições de Repetibilidade

SEL – Seletividade

SG - Alisamento Savitzky-Golay (*Savitzky Golay smoothing*)

SPE - Erros de Previsão Padrão (*Standard Prediction Errors*)

SPR - Taxa de especificidade (*Specificity Rate*)

SRM - Método de Referência Padrão americano (*Standard Reference Method*)

STR - Taxa de seletividade (*Selectivity Rate*)

SVM - Máquinas de Vetores de Suporte (*Support Vector Machines*)

SVN – Variação Normal Padrão (*Standard Normal Variate*)

UNICAMP - Universidade Estadual de Campinas

UV/Vis - Espectroscopia no ultravioleta visível

VIP escores – Importância dos Vetores na Projeção dos escores (*Variable Importance on Projection*)

SUMÁRIO

1	INTRODUÇÃO	21
2	REVISÃO BIBLIOGRÁFICA	24
2.1	CERVEJA.....	24
2.1.1	Histórico	25
2.1.2	Matéria-Prima.....	27
2.1.3	Mercado	35
2.2	PORTABILIDADE	36
2.3	QUIMIOMETRIA	38
2.3.1	PCA.....	39
2.3.2	PLS	40
2.3.3	PLS-DA	40
2.3.4	Quimiometria no Controle de Qualidade de Cerveja.....	41
2.4	IMAGENS DIGITAIS	42
2.5	VALIDAÇÃO ANALÍTICA MULTIVARIADA.....	44
3	DESENVOLVIMENTO DE UM MÉTODO SIMPLES E RÁPIDO PARA DETERMINAÇÃO DE COR EM CERVEJAS USANDO IMAGENS DIGITAIS	49
3.1	OBJETIVOS	50
3.2	MATERIAIS E MÉTODOS	50
3.2.1	Equipamentos e <i>softwares</i>	50
3.2.2	Amostras	51
3.2.3	Procedimentos	51
3.2.4	Análise espectrofotométrica (método de referência).....	52
3.2.5	Modelos multivariados e validação analítica	52
3.3	RESULTADOS E DISCUSSÃO	54
3.3.1	Modelo global de PLS.....	54
3.3.2	Modelo PLS-DA	55
3.3.3	Modelos PLS locais.....	58
3.4	CONCLUSÕES	66
4	DETERMINAÇÃO DO TEOR DE ÁLCOOL EM CERVEJAS DE DIFERENTES ESTILOS COM BASE EM ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO PORTÁTIL E CALIBRAÇÃO MULTIVARIADA	69

4.1	OBJETIVOS	71
4.2	MATERIAIS E MÉTODOS	72
4.2.1	Equipamentos e <i>softwares</i>	72
4.2.2	Amostras	72
4.2.3	Procedimentos de Preparo de amostra.....	73
4.2.4	Análise cromatográfica (método de referência).....	74
4.2.5	Validação analítica do método multivariado proposto	75
4.3	RESULTADOS E DISCUSSÃO	76
4.3.1	Análise cromatográfica.....	76
4.3.2	Desenvolvimento do modelo PLS	79
4.4	CONCLUSÃO	84
5	MODELO DE CALIBRAÇÃO MULTIVARIADA MULTIPRODUTO PARA A DETERMINAÇÃO DO TEOR DE ETANOL EM BEBIDAS ALCOÓLICAS FERMENTADAS USANDO UM ESPECTROFOTÔMETRO NIR PORTÁTIL.....	86
5.1	INTRODUÇÃO	86
5.2	OBJETIVOS	89
5.3	MATERIAIS E MÉTODOS	89
5.3.1	Amostras	89
5.3.2	Instrumentos e <i>software</i>	90
5.3.3	Procedimentos	90
5.3.4	Método de Referência (GC-FID)	91
5.3.5	Modelo de calibração multivariada e validação analítica	91
5.4	RESULTADOS E DISCUSSÃO	92
5.5	CONCLUSÃO	100
6	MÉTODO DE TRIAGEM PARA DETECÇÃO RÁPIDA DE DIETILENOGLICOL EM CERVEJA COM BASE EM QUIMIOMETRIA E ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO PORTÁTIL	103
6.1	INTRODUÇÃO	103
6.2	OBJETIVOS	106
6.3	MATERIAIS E MÉTODOS	106
6.3.1	Equipamentos e <i>softwares</i>	106
6.3.2	Amostras de Cerveja.....	107
6.3.3	Método GC-MS para garantir a ausência de DEG em cervejas não contaminadas.....	107

6.3.4	Preparação das amostras e aquisição de espectros.....	107
6.3.5	Validação analítica qualitativa.....	108
6.4	RESULTADOS E DISCUSSÃO	109
6.5	CONCLUSÃO	118
7	CONCLUSÕES GERAIS	119
	REFERÊNCIAS.....	120
	APÊNDICE A.....	137
	APÊNDICE B.....	147
	APÊNDICE C.....	157
	APÊNDICE D.....	164

CAPÍTULO 1: INTRODUÇÃO

1 INTRODUÇÃO

A determinação de parâmetros de qualidade de cerveja e suas matérias-primas tradicionalmente envolve processos analíticos que, em sua maioria, consomem muito tempo, e utilizam solventes e reagentes tóxicos que podem oferecer risco ao analista. Além destas características, ela pode necessitar de equipamentos caros e que possuam manutenção dispendiosa. Esses fatores corroboram para a necessidade do desenvolvimento de novos métodos analíticos que sejam simples, rápidos e de baixo custo.

Devido ao grande interesse na fabricação de cervejas, o desenvolvimento de novos métodos analíticos e a rápida avaliação dos resultados para o controle de qualidade destes produtos, novos métodos analíticos são de grande importância como ferramentas de melhoria das técnicas de produção de cerveja. Assim como todo produto industrializado, o controle de qualidade do item concluído também é de fundamental importância na indústria cervejeira. Diante dos desafios enfrentados pelos processos analíticos tradicionais na determinação dos parâmetros de qualidade da cerveja, é crucial explorar novos métodos mais simples, rápidos e que ainda forneçam soluções mais econômicas para a indústria cervejeira. Isso exige a simplificação da tecnologia de análise de cerveja para torná-la mais acessível e econômica para cervejarias em todo o país. Como resultado, há uma demanda crescente por laboratórios especializados que oferecem serviços de análises físico-químicas para cervejas no Brasil.

Com a crescente sofisticação das técnicas instrumentais de análise química, técnicas de tratamentos de dados mais complexas do ponto de vista matemático e estatístico tornaram-se necessárias. Embora muitas metodologias de estatística multivariada tenham seu desenvolvimento originado na área de ciências sociais, elas atualmente são aplicadas quase como um procedimento de rotina na análise de dados oriundos de sistemas químicos. Dentre as áreas tradicionais da química, a química analítica é que mais se beneficia do uso de ferramentas quimiométricas. O modelo estatístico dos métodos multivariados considera a correlação entre muitas variáveis analisadas simultaneamente, permitindo a extração de uma quantidade maior de informação.

Tendo em vista o que foi mencionado acima, esta tese teve como principal objetivo a aplicação de métodos quimiométricos em conjunto com equipamentos de maior simplicidade, praticidade e viabilidade econômica (sistemas simples e portáteis de aquisição de imagens digitais, espectrofotômetros portáteis de infravermelho próximo), na análise de cerveja e suas matérias-primas. Após a introdução, o Capítulo 2 apresenta uma revisão bibliográfica focada na cerveja e suas matérias-primas, e equipamentos portáteis, além de discorrer brevemente sobre os métodos quimiométricos utilizados e aspectos da validação analítica multivariada. O Capítulo 3 descreve o desenvolvimento e validação de um novo método de análise multivariada de imagens digitais para determinação da cor de cervejas. O método desenvolvido é direto, simples e rápido, não exigindo o uso de técnicas de preparo de amostra e para garantir a robustez foi utilizada uma grande variedade de cervejas de diferentes estilos, marcas e empresas cervejeiras. O Capítulo 4 descreve o desenvolvimento e validação de um método de análise do teor de etanol em cerveja utilizando um equipamento portátil de espectroscopia no infravermelho próximo e PLS. O Capítulo 5 descreve o desenvolvimento e validação de um método multiproduto para análise de teor alcoólico em bebidas fermentadas, incluindo cerveja. Finalmente, o Capítulo 6 apresenta um método de triagem rápido e simples para a detecção direta de dietilenoglicol em cerveja, uma aplicação de grande importância forense.

CAPÍTULO 2: REVISÃO BIBLIOGRÁFICA

2 REVISÃO BIBLIOGRÁFICA

2.1 Cerveja

A cerveja é definida como uma bebida alcoólica produzida através da fermentação de ingredientes que contêm amido [1-2]. Essa bebida pode ser encontrada em muitos lugares, além de ser consumida e produzida em diversos países que têm o histórico de produção desde sua origem, especialmente na Ásia e na Europa. Em 2013, o Brasil já era o terceiro maior produtor de cerveja do mundo, tendo produzido quase 13,5 milhões de quilolitros desta bebida naquele ano. As cervejas Pilsner (*American standard lagers*) são as mais consumidas no Brasil [3].

A legislação brasileira define a cerveja como a bebida obtida pela fermentação alcoólica do mosto cervejeiro oriundo do malte de cevada e água potável, por ação da levedura, com adição de lúpulo [4]. A maior parte do mercado brasileiro de cerveja é dominada por uma das maiores empresas cervejeiras do mundo, dona das marcas mais consumidas pela população [3]. De acordo com uma pesquisa realizada pelo Ministério da Agricultura, Pecuária e Abastecimento (MAPA), o mercado cervejeiro no Brasil cresceu exponencialmente a partir de 2010. Somente em 2018, 210 novas cervejarias começaram a funcionar, representando um crescimento de 30% em relação a 2017, ou seja, o mais alto número da história brasileira [5], sendo que a maioria desses novos empreendimentos são cervejarias artesanais.

Como citado anteriormente, mesmo com toda a tecnologia atual, a fabricação de cerveja não é um processo simples, considerando as sequências críticas de etapas exigidas. É necessário predeterminar o tipo de cerveja a ser produzida para definir suas características sensoriais, como cor, aroma e amargor. Essas características devem ser obtidas na seleção de maltes, lúpulos e leveduras [6]. Os ingredientes da cerveja incluem água, que constitui uma parte importante do processo, além de lúpulo, levedura e outros ingredientes, como frutas, ervas e outras plantas.

2.1.1 Histórico

A cerveja é uma bebida fermentada apreciada há 8.000 anos. Um dos primeiros relatos de sua existência é descrito no antigo Egito, junto aos adoradores de Osíris, o deus egípcio da agricultura, sendo a descoberta da sua fermentação atribuída à intervenção divina [7]. A cerveja era muito mais do que apenas um alimento. Em uma sociedade que não possuía um sistema monetário, essa bebida era usada como unidade de troca, além de possuir um papel central na crença religiosa. De acordo com estudos oriundos de escavações, a fabricação de pão e cerveja costumava ocorrer em conjunto e as técnicas aplicadas na fabricação de cerveja pelos egípcios estão possivelmente ligadas à fabricação do pão, como é possível observar na [Figura 1](#), que apresenta uma foto de uma arte que é datada da Sexta Dinastia, correspondendo ao Reinado de Pepi (2289-2255 A.C.) e descoberta em 1894 [8,9]. Ainda não se sabe por completo como era fabricada a cerveja, mas uma variedade de plantas teria sido usada para exalar sabores, já que o lúpulo, principal ingrediente responsável pelo amargor, seria usado somente a partir do ano de 736 na Alemanha [10].



Figura 1. Modelo que descreve a preparação de pão e cerveja. Peça exposta no Museu da Antiguidade do Cairo {<http://www.touregypt.net/featurestories/picture10082003.htm>}.

Os egípcios foram os responsáveis por difundir as técnicas cervejeiras aos gregos e romanos. No entanto, na Grécia antiga e na Roma, o vinho era a bebida das classes privilegiadas, enquanto a cerveja era consumida pelas outras classes. A rota

destas técnicas seguiu para o norte, influenciada pelo conhecimento de manuseio da madeira da população do norte e centro da Europa, devido à sua capacidade de transformá-las em recipientes e barris de cerveja. Enquanto os gregos e outros do sul da Europa bebiam vinho estocado em cerâmica, as tribos alemãs consumiam bebidas baseadas em cevada ou trigo e estocadas em barris de madeira [7].

A indústria cervejeira ocidental foi estabelecida na região da Bohemia, que corresponde atualmente ao território da República Tcheca, e também Bavaria. Havia fabricantes de cerveja no Tribunal de Charlemagne que insistiam em tecnologia saudável na produção de cerveja, sendo assim a importância de um malte de qualidade para a fabricação da cerveja foi reconhecida, levando ao desenvolvimento de malterias especializadas. O cultivo de lúpulo na Alemanha foi registrado pela primeira vez em 736, e durante esse período uma mistura de ervas era utilizada para conceder amargor e aroma à cerveja [11]. Na Idade Média, século XIII, os cervejeiros alemães foram os primeiros a empregar lúpulo na cerveja, concedendo as características básicas da bebida atual. Em 1516, os duques Wilhelm IV e Ludwig X declararam a Lei da Pureza, que tinha como objetivo controlar a qualidade da cerveja produzida na Alemanha. Essa lei, também conhecida como *Reinheitsgebot*, restringe as matérias-primas da cerveja ao malte, lúpulo e água, sendo atualizada posteriormente para incluir leveduras. Com a revolução industrial, o modo de produção e distribuição de cerveja sofreu mudanças decisivas, estabelecendo-se fábricas cada vez maiores na Inglaterra e Alemanha.

A história da cerveja brasileira inicia-se em 1634, quando colonizadores holandeses chegaram pela primeira vez ao país, carregando ingredientes para sua produção. Em 1807, a Inglaterra enviou seus exércitos para defender Portugal contra Napoleão, forjando laços estreitos entre os dois países. Como resultado, os comerciantes ingleses estabeleceram negócios no Brasil e disponibilizaram cervejas nas principais cidades do país. Demorou mais de 50 anos para as primeiras cervejas artesanais locais, agora produzidas por imigrantes alemães, aparecerem. Os alemães colonizaram partes do sul do país e as primeiras cervejarias foram localizadas nos estados do Rio de Janeiro, São Paulo, Santa Catarina e Rio Grande do Sul [12]. Como país tropical, o Brasil não possuía um clima ideal para cultivar cevada e lúpulo. Ambos os ingredientes tiveram que ser importados de países produtores, o que era

complicado e caro na época. Além disso, a refrigeração ainda não estava disponível, tornando difícil evitar a contaminação das leveduras.

2.1.2 Matéria-Prima

2.1.2.1 Água

Essa substância é responsável por aproximadamente 94% do peso da cerveja, sendo esse líquido a matéria-prima presente em maior quantidade. Assim, a água e sua origem são importantes e têm o mesmo peso que a escolha das variedades de lúpulo, maltes e matérias-primas restantes que são utilizadas na fabricação de cerveja. A natureza da água é um fator essencial para determinar a qualidade da cerveja, pois fontes diferentes possuem perfis químicos distintos. Proveniente da rede pública ou de poços artesianos, ela é um fator essencial na escolha do local de instalação da cervejaria [13].

Uma fonte de água pode ser potável, mas não necessariamente adequada para a fabricação de cerveja. Na natureza, toda água contém sais dissolvidos e compostos orgânicos em quantidades diferentes de acordo com a região. Essa diferença influencia diretamente os processos químicos e enzimáticos que ocorrem durante a fermentação da cerveja, além da possibilidade de conter cloro ou cloraminas, gases dissolvidos ou compostos orgânicos que podem afetar adversamente o sabor da bebida [14].

Geralmente, a água para fabricação de cerveja deve possuir 50 mg L^{-1} de cálcio, no mínimo, para melhorar o desempenho da fermentação e clarificar a cerveja. A alcalinidade na água de fermentação tem sido tradicionalmente vista apenas como um fator a ser corrigido, porém, o nível recomendado de alcalinidade na água de produção varia com base na acidez da composição do malte e na característica desejada da cerveja. Em geral, a baixa alcalinidade é desejável para cervejas de cor mais clara e a necessidade de maior alcalinidade aumenta para maltes mais escuros e ácidos [15]. No entanto, se a água não for de boa qualidade ou não apresentar composição química adequada, poderá ser tratada por diferentes processos que visam purificá-la e, se necessário, efetuar algumas modificações nos níveis de íons inorgânicos presentes.

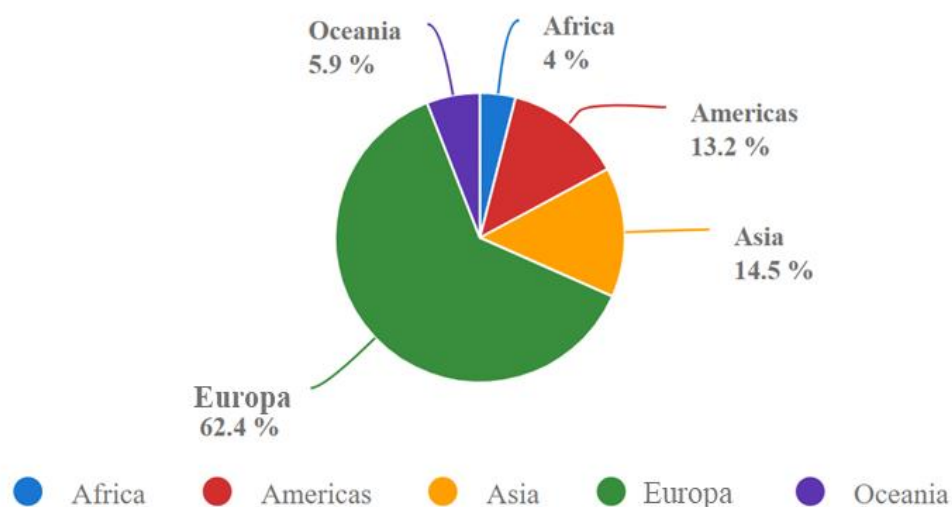
Além de ser uma matéria-prima, a água é consumida principalmente para limpeza e esterilização de instalações e equipamentos, geração de vapor, sistemas de resfriamento e lavagem. O gasto de água nas cervejarias varia de 4 a 10 vezes o volume de cerveja produzida [16].

2.1.2.2 Malte

Os historiadores acreditam amplamente que a relação entre os seres humanos e os grãos é anterior à história registrada, e uma das principais causas do distanciamento da caça e da coleta, bem como o desenvolvimento dos assentamentos, veio do desejo de cultivar grãos de cereais. Os grãos de cereais forneciam uma fonte estável e confiável de alimentos, permitindo que os primeiros humanos tivessem uma maior estabilidade alimentícia. Com o tempo, começou a haver um excesso de grãos. Pela falta de conhecimento sobre como realizar o armazenamento destes grãos, o excesso foi utilizado para produzir bebidas fermentadas, que também seriam fonte de açúcares, proteínas e vitaminas. Porém, este processo de fabricação dependia diretamente do tipo de grão utilizado [7].

Para entender melhor a distribuição da produção de cevada a Figura 2 ilustra um gráfico de pizza **(a)** e um mapa do cultivo de cevada mundial **(b)** entre os anos de 2010 e 2017. É possível observar que a Europa é a região de maior produção mundial, sendo responsável por 62,4% da produção total [17].

(a)



b)

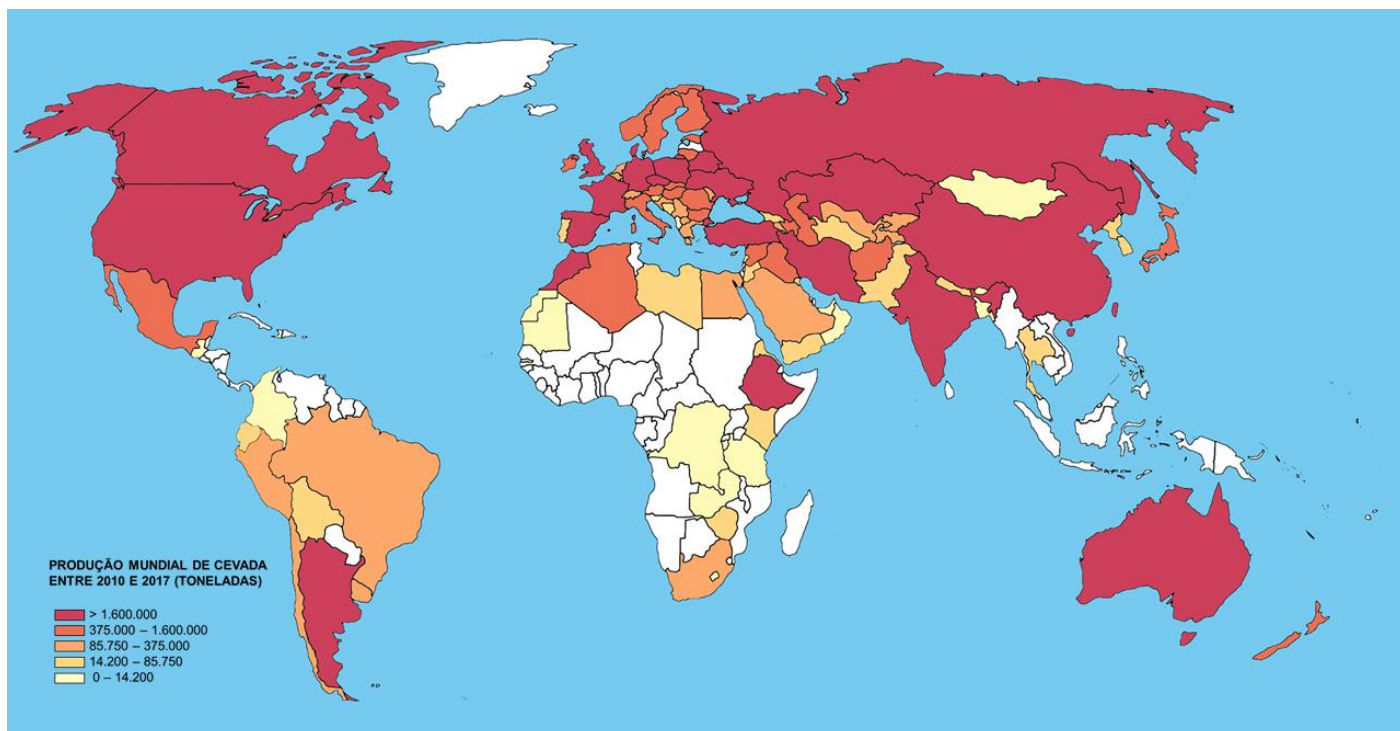


Figura 2. (a) Gráfico de pizza contendo a participação de cada continente e **(b)** Mapa da produção mundial de cevada entre os anos de 2010 e 2017 segundo a Organização das Nações Unidas para Alimentação e Agricultura (FAO) (Adaptado de [17]).

A cevada pertence à espécie *Hordeum Vulgare*, e pode ser classificada em duas categorias, dependendo de quando são semeados e colhidos os grãos: as variedades de inverno são semeadas no outono, enquanto as variedades de primavera são semeadas na primavera e colhidas após as variedades de inverno [18]. De um modo geral, considerando o hemisfério norte, quanto mais cedo no ano a cevada é semeada, menor será o teor de proteína no grão colhido e conseqüentemente maior será o seu rendimento, pois o amido é acumulado durante todo o período de crescimento [19].

Embora possa ser cultivada em locais quentes e úmidos, a cevada prefere climas frios e secos. A umidade no primeiro mês do cultivo garante que a cevada cresça até a maturidade, já a ausência de chuva durante o amadurecimento e a colheita contribui para uma colheita satisfatória. Como será feita a germinação controlada do grão na maltaria, é indesejável que este germine antecipadamente, seja na espiga ou na estocagem, em decorrência dos altos teores de umidade. O teor de umidade e a temperatura são as variáveis que determinam quanto tempo a cevada pode ser armazenada. Sendo assim, as cevadas com menor teor de água são mais valorizadas [19]. A aparência, o teor de umidade e a dureza relativa dos grãos são usados como

marcadores de sua maturidade. A acumulação de amido e outros sólidos no endosperma cessa quando o teor de umidade diminui para aproximadamente 35%. Posteriormente, o teor de umidade cai para menos de 15% e, nesse ponto, os grãos estão prontos para a colheita.

As variedades de cevada são classificadas com base na disposição das fileiras de sementes, podendo ter duas ou seis fileiras. Atualmente, todas as variedades de cevada são classificadas como *H. vulgare*, embora já tenham sido chamadas de *Hordeum distichon*. Todas as cevadas selvagens possuem duas fileiras e ambas as cevadas são usadas para fabricação de cerveja. Tradicionalmente, as cevadas de duas fileiras são predominantes usadas na Europa e no Brasil, enquanto as variedades de seis fileiras são preferidas pelos norte-americanos. Esta diferença não é apenas morfológica. Os grãos de cevada da variedade de duas fileiras são, em média, maiores e mais uniformes em comparação com os de seis fileiras [11].

Além disso, existem outras diferenças bioquímicas que têm importância em termos de qualidade do malte, como por exemplo, as variedades de duas fileiras possuem um teor de nitrogênio total relativamente baixo, além de conter mais extrato e α -amilase e menos β -glucana e potência diastática [11,20]. A cevada de seis fileiras apresenta dificuldade na produção de malte e na moagem dos grãos na cervejaria, resultando em um menor rendimento na mosturação, mas por outro lado, ela facilita a filtração do mosto e aceita maior proporção de adjunto na formulação da cerveja.

Do ponto de vista morfológico, o grão de cevada é constituído por um embrião, um endosperma amiláceo que funciona como fonte de alimento, e uma casca protetora. Cada um desses três elementos possui papéis importantes no processo de malteação e na fabricação da cerveja. A Figura 3 apresenta um diagrama dos componentes do grão de cevada. Nele, é possível identificar a extremidade basal, onde está localizado o embrião.

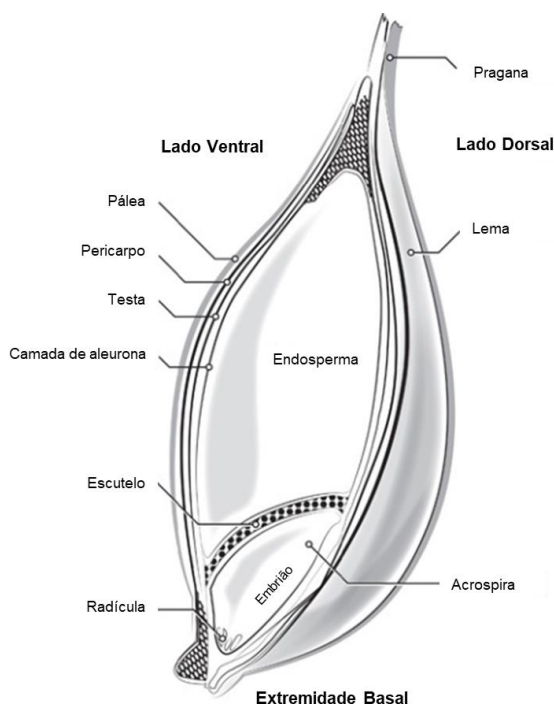


Figura 3. Diagrama dos componentes do grão de cevada (Adaptado de [18])

A casca é composta pelo lema e a pálea, já o pericarpo está localizado diretamente abaixo desta. A próxima camada é chamada de testa, que é um revestimento que protege o embrião e as reservas de energia [19]. A testa, juntamente com o pericarpo agem como um escudo, impedindo que a umidade e outros fatores ambientais atinjam o embrião. Sob a testa fica a camada de aleurona, que é uma fina camada de células vivas abrangendo a maior parte do endosperma. Essa camada não cresce, mas produz enzimas que acessam as reservas de energia armazenadas no endosperma necessárias ao embrião para uma germinação bem-sucedida. O endosperma é responsável por 80% do peso total do grão da cevada, sendo assim a maior estrutura no grão. A estrutura interna do endosperma consiste em uma mistura de grânulos grandes e pequenos de amido incorporados em uma matriz de proteínas.

Apesar do endosperma possuir a reserva energética, é o embrião que tem o controle final sobre essa reserva. Através da produção de hormônios, que migram para a aleurona, o embrião controla a ativação ou desativação da síntese enzimática. Os dois principais hormônios produzidos pelo embrião são as giberelinas e ácido abscísico. As giberelinas são responsáveis por regular o crescimento e influenciar vários processos de desenvolvimento, já o ácido abscísico inibe o crescimento e

desenvolvimento do grão [21]. O equilíbrio desses hormônios regula a extensão em que as várias enzimas são produzidas.

2.1.2.3 Lúpulo

O lúpulo é considerado uma importante matéria-prima da cerveja, sendo responsável por seu amargor e parte do aroma característico. A planta do lúpulo, *Humulus lupulus*, pertence à família das *Cannabaceae*, composta pelos gêneros *Humulus* e *Cannabis*. Ele é uma planta dióica, ou seja, formada por estruturas reprodutivas masculinas e femininas em indivíduos diferentes. O lúpulo para fabricação de cerveja é obtido da flor da planta feminina, e é rico em resinas e óleos, sendo a primeira fonte de substâncias amargas e o segundo fonte de sabor e aroma. Historicamente, o lúpulo era valorizado não somente por sua capacidade de fornecer essas características à bebida, mas também por melhorar sua preservação [7].

O *Humulus lupulus* é nativo do hemisfério norte e foi exportado para regiões do sul, sendo atualmente cultivado na Austrália, Nova Zelândia, Argentina e África do Sul. Existem cerca de 25 países onde o lúpulo é cultivado, sendo necessário solo fértil e condições climáticas específicas [11]. A Figura 4 apresenta o mapa do cultivo de lúpulo mundial entre os anos de 2010 e 2017. Segundo a Organização das Nações Unidas para Alimentação e Agricultura (FAO), os Estados Unidos foi o país líder em produção de lúpulo, com aproximadamente 28%, seguido pela Alemanha e Etiópia ambas com 27% e China com 7% [17].

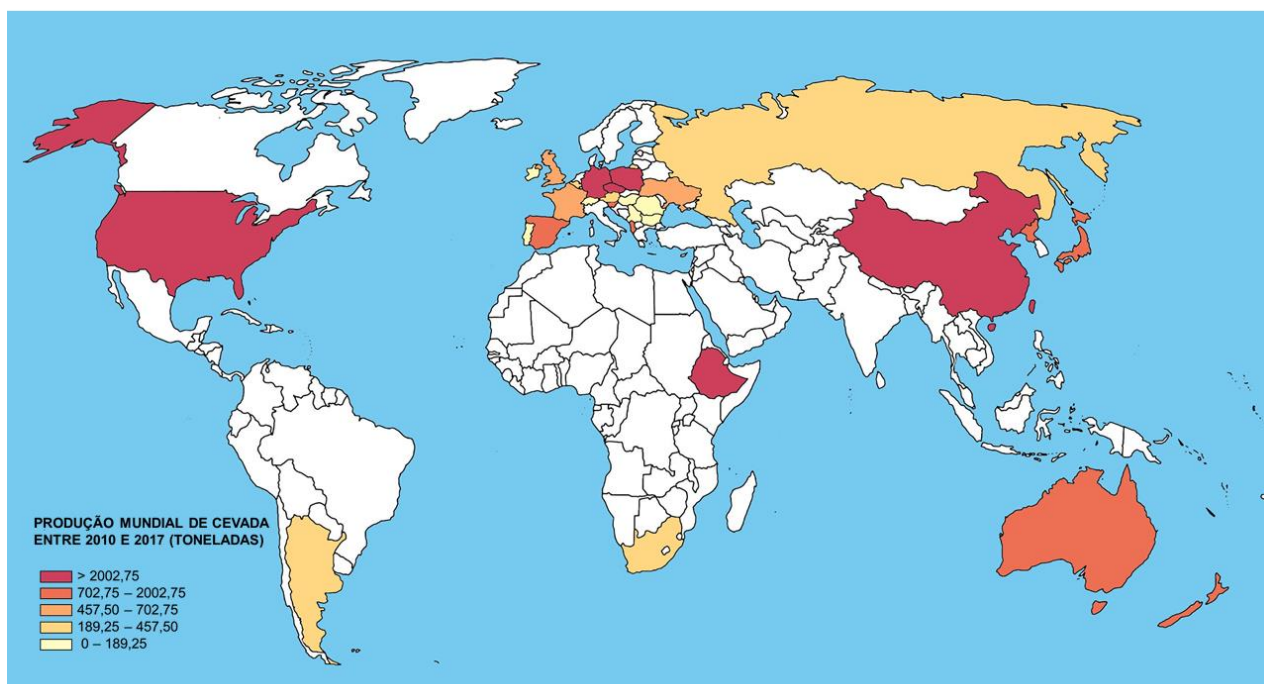


Figura 4. Mapa Mundial da produção mundial de lúpulo entre os anos de 2010 e 2017 segundo a Organização das Nações Unidas para Alimentação e Agricultura (FAO) (Adaptado de [17]).

Existem muitas variedades de lúpulo e a busca por tipos mais novos com características aprimoradas é contínua. Estas podem ser nomeadas de acordo com o nome da região onde são cultivadas ou com o nome da pessoa que as apresentou pela primeira vez [22]. Os lúpulos podem ser classificados em duas categorias, os lúpulos aromáticos e os lúpulos amargos. O que os diferencia é a proporção presente de resinas e óleos essenciais, sendo que se o lúpulo possui uma razão maior de óleos essenciais, ele é considerado aromático [11]. Atualmente, entre 50 e 100 variedades de lúpulo são cultivadas comercialmente e cada uma tem seu próprio conjunto de características. Uma outra diferenciação presente está na forma em que o lúpulo é comercializado, podendo ser convertido em pellets ou extratos, sendo esses dois tipos mais estáveis à oxidação, e de formato mais compacto, facilmente transportados e armazenados [18].

Uma grande variedade de compostos pode ser encontrada nesta planta, porém os seus principais constituintes são as resinas e óleos essenciais, ambos encontrados nas glândulas de lupulina, localizadas nos cones das plantas femininas. Sabe-se que os óleos essenciais variam de 0,03% a 3% do peso do lúpulo e são uma mistura complexa de aproximadamente 300 compostos, mas ainda não foi correlacionada a

composição química dos óleos essenciais com as características de aroma que eles proporcionam. Existem também vários componentes na porção de resinas, mas normalmente são considerados os α -ácidos e β -ácidos como componentes significativos para o amargor [23].

O lúpulo pode ser adicionado na fabricação de cerveja em dois momentos, antes e após a fermentação. O objetivo é o de tornar a cerveja amarga a um nível exato e reproduzível. Os α -ácidos presentes no lúpulo são insolúveis na cerveja e, portanto, para que possam acrescentar suas características à bebida é necessário que uma isomerização seja promovida. Para isso, ele é adicionado antes do processo de fervura [21]. Após esse processo, os α -ácidos presentes são isomerizados e formam iso- α -ácidos, que são mais solúveis no mosto.

Existe uma correlação entre o impacto geral no sabor de uma cerveja derivada de malte e levedura e sua amargura analítica. No entanto, o pH da cerveja também afeta a percepção de amargor, pois um pH mais alto do mosto, e a presença de íons como o Ca^{2+} , aceleram as reações de isomerização e promovem um maior aproveitamento do α -ácido presente no lúpulo [19,23].

2.1.2.4 Levedura

Dentre as matérias-primas usadas para a produção da cerveja, a levedura é o principal composto responsável pela etapa de fermentação. A levedura é um microrganismo classificado como fungo e que tem o objetivo de realizar a fermentação alcoólica, ou seja, converter os açúcares fermentescíveis em etanol e outros produtos [24]. A fermentação alcoólica possui diversas etapas, sendo seus produtos principais o etanol e dióxido de carbono além da formação dos subprodutos que têm elevada importância sensorial, gerando assim uma diversidade de sabores com diferentes tipos de matérias-primas [2].

O tipo de levedura usada em cervejas é a *Saccharomyces cerevisiae*, e este tipo conta com uma elevada diversidade de cepas. Dentro dessa espécie, são encontradas as antigas nomeações *S. uvarum* e *S. carlsbenguensis*, que a partir de 1984 foram unificadas dentro da espécie *S. cerevisiae* [25]. De modo mais prático, há uma classificação usual de grande importância em cervejarias, baseada no comportamento

da levedura durante o processo fermentativo. Esta classificação se divide em cervejas de alta fermentação (*Ale*) para fermentações com as leveduras localizadas na superfície do mosto, e cervejas de baixa fermentação (*Lager*), nas quais a levedura decanta no mosto [26].

O sabor da cerveja é influenciado diretamente pela temperatura imposta para a levedura. Leveduras *Ale* têm temperatura de trabalho entre 15 e 22°C, com algumas delas não funcionando em temperaturas abaixo de 12°C. Já as leveduras do tipo *Lager* são utilizadas em temperaturas entre 7 e 15°C, sendo que nessa faixa o crescimento ocorre de forma mais lenta e com menor volume de espumas o que vem a caracterizar a sedimentação dessa levedura. Sendo assim, as leveduras de fermentação de topo são usadas em cervejas tipo Porter, Stout, cervejas de trigo, Kolsch e Altbier, e as leveduras do tipo fermentação de fundo são empregadas em cervejas Marzen, American Malt Liquors, Pilsen, Bocks e Dortmund [27].

2.1.3 Mercado

A tradição da fabricação de cerveja no Brasil retorna à imigração alemã no início do século XIX. As primeiras cervejarias datam da década de 1830, a marca Bohemia é considerada a primeira cerveja brasileira, iniciando sua produção em 1853 na cidade de Petrópolis (Rio de Janeiro) e é a cerveja brasileira mais antiga que ainda está em produção. Antártica e Brahma, duas marcas importantes, começaram a serem produzidas na década de 1880 [12]. Depois de muitos anos disputando participação de mercado, a Brahma adquiriu a Antartica em 1999, tornando-se um dos principais atuantes no mercado mundial de cerveja [28]. A nova empresa foi nomeada Ambev, que mais tarde se fundiu com a gigante belga Interbrew para se tornar a segunda maior cervejaria do mundo, chamada InBev. Em 2009, a cervejaria belgo-brasileira conseguiu adquirir a maior empresa cervejeira do mundo, a Anheuser-Busch, criando uma nova empresa chamada Anheuser-Busch Inbev. Em 2015, a empresa já respondia por mais de 30% do mercado mundial de cerveja, produzindo no Brasil principalmente cervejas leves como Skol, Brahma, Antartica e outras [29].

Desde 2013, o Brasil já era o terceiro maior produtor de cerveja do mundo. Além disso, pode ser notado um grande crescimento do mercado de cerveja artesanal no Brasil. De acordo com o MAPA, 210 novas cervejarias entraram no mercado apenas

em 2018, o que representa um crescimento exponencial do mercado cervejeiro [5]. Minas Gerais possui um papel especial no mercado cervejeiro, sendo conhecida como a 'Bélgica brasileira' por possuir uma diversidade de estilos de cervejas e marcas, além de várias premiações internacionais [30]. Em volume de produção, o estado se encontra atrás somente de Santa Catarina, Rio Grande do Sul e São Paulo. Em diversidade, entretanto, Minas é destaque no ramo. É possível citar várias marcas reconhecidas nacionalmente e internacionalmente, como a Krug Bier, Wälls, Falke Bier e Antuérpia.

O mercado cervejeiro movimenta uma enorme rede que agrega desde a pesquisa, o cultivo, o processamento e a comercialização de insumos e matérias-primas até a entrega do produto. Tendo em vista esse fato, é necessário considerar o controle de qualidade como uma etapa fundamental na indústria cervejeira. Esse paradoxo entre crescimento versus controle de qualidade foi colocado em xeque quando, no início de 2020, a Backer foi apontada como responsável pela intoxicação de dezenove pessoas por dietilenoglicol, o que resultou em 10 fatalidades [31,32]. Nesse sentido, o desenvolvimento de métodos analíticos simples e rápidos tem se tornado um objetivo do mercado para facilitar e aprimorar o controle de qualidade do produto final.

Desse modo, métodos quimiométricos podem se destacar, e algumas alternativas de fácil acesso e de baixo custo, como aparelhos portáteis para obtenção de imagens digitais, pode ser utilizadas.

2.2 Portabilidade

Muitos instrumentos analíticos estão agora disponíveis em uma versão compacta. A recente disponibilidade dessas versões portáteis traz muitas vantagens para a ciência analítica, como economia de energia, preços mais baixos do que os instrumentos convencionais de bancada e facilidade de transporte para análise *in situ*. É possível mencionar diversos tipos de instrumentos portáteis, como os espectrofotômetros de absorção molecular no UV/visível e espectrômetros Raman, que possuem uma grande diversidade de aplicações, tais como caracterização de corantes e determinação de gordura em leite [33-34].

Entre os equipamentos portáteis, podemos destacar o espectrofotômetro na região do infravermelho próximo (NIR), que por muito tempo foi considerada como uma técnica “adormecida” entre as espectroscópicas. Embora a espectroscopia no infravermelho médio já fosse amplamente utilizada, a espectroscopia NIR foi ignorada pelos espectroscopistas na maior parte do século XX [35]. Os espectros NIR são frequentemente complexos e são compostos por bandas de absorção sobrepostas e fracas, por isso, essas bandas são pouco seletivas e sofrem influências de variáveis químicas e físicas.

A espectroscopia NIR e a quimiometria apresentam grande simbiose. A quimiometria fornece a ferramenta necessária para a interpretação dos dados analíticos oriundos da espectroscopia NIR. O uso conjunto de espectroscopia no infravermelho próximo (NIR) e calibração multivariada apresenta múltiplas aplicações, como análises sobre a composição, adulteração e controle de qualidade no setor de alimentos [36-39].

Embora os espectrômetros em miniatura estejam disponíveis comercialmente desde a década de 1990, foi somente uma década depois que um equipamento NIR portátil estaria disponível [40]. Pérez-Marín *et al.* publicaram a primeira aplicação de um espectrômetro portátil na região do infravermelho próximo (MicroNIR[®]), na qual estudaram a viabilidade do seu uso para medir conteúdo de sólidos solúveis, firmeza, variedade e duração do armazenamento pós-colheita sob refrigeração de ameixas intactas [41].

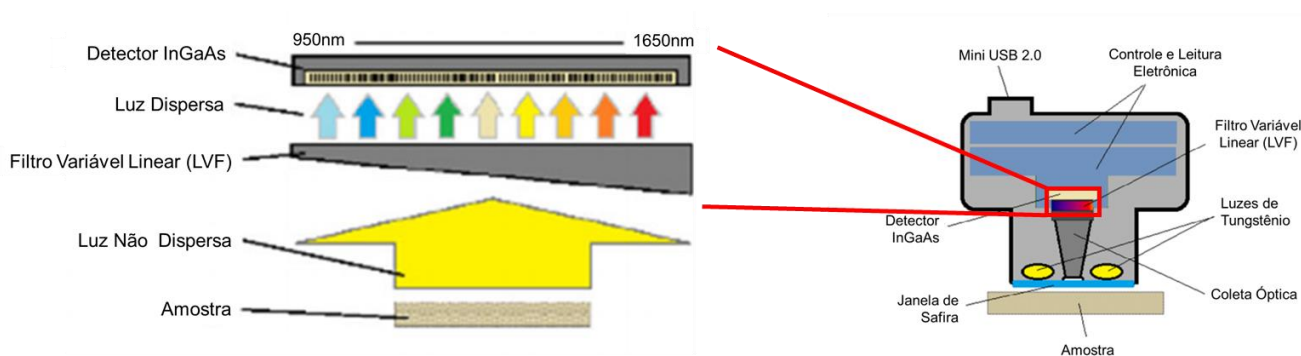


Figura 5. Esquema do espectrômetro portátil na região do infravermelho próximo MicroNIR[®] (Adaptado de [42]).

A Figura 5 apresenta um esquema do MicroNIR[®]. Nele, é possível observar seus principais componentes, como o Filtro Variável Linear (LVF). O LVF é um dispositivo

que passa frequências dentro de uma certa faixa e rejeita, ou atenua, frequências fora dela. Como o comprimento de onda central do *bandpass* (CWL) é uma função da espessura do revestimento, o comprimento de onda de pico transmitido através do filtro varia de forma linear na direção da cunha. Sendo assim, para o MicroNIR[®], o comprimento de onda específico foi selecionado ajustando o filtro para a posição linear apropriada. Portanto, o LVF se encontra posicionado acima de um detector de matriz InGaAs, que é um semicondutor de banda variável com excelentes propriedades ópticas, além de ser incomparável na faixa de comprimento de onda NIR para sensibilidade. A principal compensação na abordagem desse *design* é a eficiência da luz versus resolução, ou seja, a eficiência da luz é reduzida em comparação com um instrumento baseado na transformada de Fourier.

Um espectrofotômetro miniatura de infravermelho próximo, como o MicroNIR[®], possui vantagens analíticas quando comparado aos equipamentos NIR convencionais de bancada. Ele permite análises *in situ*, o que elimina fontes de erros relacionadas ao transporte de amostras e é mais fácil de operar. Além das vantagens do NIR convencional, tempos menores de análise podem ser alcançados, além de ser ambientalmente mais amigável [43]. As aplicações do MicroNIR[®] são ascendentes e foram relatadas em vários trabalhos com diferentes tipos de matrizes, como análise da qualidade do fruto da acerola [44], parâmetros de controle de qualidade na indústria cafeeira [45], adulteração de combustível [46] e identificação de matéria-prima farmacêutica [47].

2.3 Quimiometria

A quimiometria pode ser definida como sendo a aplicação de métodos estatísticos e matemáticos à análise de dados químicos e ao planejamento de experimentos químicos e simulações [48]. Segundo Kowalski, Brown e Vandeginste (1987), a palavra “quimiometria” foi usada pela primeira vez em 1971 pelo químico Svante Wold [49]. Kowalski teve um papel essencial para a introdução de quimiometria no Brasil ao lecionar um curso sobre o assunto no Instituto de Química da Universidade Estadual de Campinas (UNICAMP) em meados dos anos 1980 [50].

O crescimento e desenvolvimento da quimiometria estão conectados com o desenvolvimento e a popularização de microprocessadores, o que aumentou a geração de grandes quantidades de dados. Segundo Kowalski, houve uma ruptura do papel que um químico analítico deveria desempenhar, que passou a não ser mais um provedor de resultados analíticos e sim de informações. A química analítica foi transformada em ciência da informação, um processo no qual a quimiometria teve um papel importante, desenvolvendo e estudando ferramentas matemáticas e estatísticas, e incorporando essas ferramentas em *softwares* [49].

O problema gerado por uma quantidade de dados elevada é que seu volume pode dificultar a visualização de padrões e correlações. Assim, o objetivo de muitos dos métodos quimiométricos é a redução de dados. Frequentemente, há alguma correlação entre as variáveis, existindo assim informações redundantes. Os métodos quimiométricos utilizados para agrupar e classificar variados tipos de amostras podem ser divididos em dois grupos, os métodos supervisionados, e os métodos não-supervisionados de reconhecimento de padrões [51]. Nos métodos supervisionados, cada amostra analisada provém de uma classe já estabelecida, e essa informação é usada durante a análise dos dados. Os métodos não supervisionados não fazem uso dessa informação, sendo assim não requerem nenhum conhecimento prévio a respeito da classificação das amostras. As amostras em questão serão agrupadas com base na informação contida nos dados experimentais.

2.3.1 PCA

A Análise de Componentes Principais (PCA - *Principal Component Analysis*) é uma técnica usada para reduzir a quantidade de dados quando há correlação presente. Ela é um método de classificação não supervisionado utilizado para a identificação de padrões latentes existentes entre amostras com dados medidos para cada uma delas e organizados em grandes conjuntos de informações. A PCA realiza essa redução identificando componentes principais, ao longo das quais a variação nos dados é máxima [52].

O objetivo do método é a decomposição da matriz que contém as informações de modo a gerar novas matrizes formadas pela combinação linear das variáveis originais, e assim obter novas matrizes contendo as informações específicas das

amostras, chamadas de matrizes de escores, e das variáveis, chamadas de matrizes de pesos (*loadings*). Similaridades e diferenças entre objetos ou variáveis podem ser vistas através de gráficos em que os escores ou pesos são plotados uns contra os outros [53].

2.3.2 PLS

O método regressão por mínimos quadrados parciais (PLS - *Partial Least Squares*) é o principal método de calibração multivariada. Herman Wold na década de 1960 desenvolveu a primeira versão do método, até então conhecido como NIPALS (*Nonlinear Iterative Partial Least Squares*). Este já incluía elementos matemáticos fundamentais presentes no método PLS atual, mas seu uso era destinado à área de economia (econometria). Na década de 1980, seu filho Svante juntamente com Harald Martens deram prosseguimento ao seu trabalho e desenvolveram uma versão quimiométrica do método, que tinha como objetivo modelar dados com variáveis altamente correlacionadas em substituição ao NIPALS [54-55].

O PLS possui um princípio similar ao PCA, efetuando a decomposição simultânea do conjunto de dados instrumentais, representado no bloco \mathbf{X} , e do conjunto de dados contendo os valores de referência da propriedade a ser quantificada, no bloco \mathbf{Y} , de modo a maximizar a correlação entre essas matrizes. Com isso, é possível a obtenção de uma equação matemática, cujas entradas seriam as informações instrumentais multivariadas das amostras e a saída seria um vetor (ou matriz, no caso de mais de uma variável prevista) com a previsão da propriedade ou teor de um analito [56]. A informação da propriedade de interesse é incluída no cálculo das variáveis latentes, pois elas são definidas de modo a manter um compromisso entre a explicação da variância em \mathbf{X} e a previsão da variável dependente, em \mathbf{Y} [51].

2.3.3 PLS-DA

O PLS não foi originalmente projetado como uma ferramenta para discriminação estatística por Wold, mas apesar disso já eram encontradas na literatura aplicações para classificação na última década do século passado [57-60]. A análise discriminante por mínimos quadrados parciais (PLS-DA - *Partial Least Squares* -

Discriminant Analysis) foi formalmente apresentada por Matthew Barker e William Rayens em 2003. Eles defenderam o uso do PLS-DA em preferência ao PCA quando a discriminação é o objetivo e a redução da dimensão é necessária [61]. O PLS-DA pode ser considerado como um classificador linear de duas classes, ou seja, o método visa encontrar uma linha que divida o espaço entre duas regiões. Para esse fim, um limite de decisão, ou *threshold*, precisa ser estabelecido. Este limite pode ser arbitrário (0,5 por exemplo) ou calculado através da teoria Bayesiana.

O PLS-DA é uma adaptação do método PLS para modelos qualitativos. A principal diferença entre os dois está nas variáveis dependentes, uma vez que no PLS-DA elas representam valores qualitativos, ou seja, classes. Para isso, é necessário definir o bloco das variáveis dependentes do conjunto de treinamento, Y , que consiste em um conjunto de vetores y contendo números inteiros que codificam a classe da amostra, sendo um vetor para cada variável. Uma vez que as variáveis dependentes foram definidas, a etapa seguinte é a construção do modelo de discriminação utilizando o método PLS. Como o método PLS maximiza a relação entre a variável dependente e os escores, é possível afirmar que as variáveis latentes representam as direções que melhor discriminam as classes, e definem a máxima separação entre elas. O PLS-DA se torna assim uma solução eficaz para a classificação supervisionada, com apresentações gráficas, além de fornecer informações sobre as variáveis por meio de pesos e escores [62].

2.3.4 Quimiometria no Controle de Qualidade de Cerveja

Assim como todo produto industrializado, o controle de qualidade possui uma importância fundamental na indústria cervejeira. O MAPA exige que para se registrar uma cerveja seja informado, dentre outras características, a cor, o teor alcoólico e o extrato primitivo [63]. Recentemente, este tema recebeu atenção especial, com a repercussão da contaminação por dietilenoglicol encontrado na cerveja Belorizontina, pertencente à Cervejaria Backer, potencializando a discussão sobre o controle de qualidade em cervejarias artesanais, o que tornou a detecção de dietilenoglicol uma importante análise [64]. Além da situação especial deste parâmetro, é possível citar as seguintes análises como sendo as mais comuns: a determinação da cor da cerveja e do malte segundo os sistemas SRM (*Standard Reference Method* - Método de

Referência Padrão americano) ou EBC (*European Brewing Convention* - Convenção de Cervejeiros da Europa); qualificação da cerveja pelo nível de amargor utilizando a escala IBU (*International Bitterness Unit* - Unidade internacional de amargor); determinação da viabilidade da levedura e quantificação de óleos essenciais presentes no lúpulo [65].

Métodos quimiométricos no controle de qualidade cervejeiro já foram utilizados, por exemplo, para a determinação de amargor em cerveja usando espectrofluorimetria [66], caracterização de cerveja baseada em compostos fenólicos [67], análise de parâmetros como cor e amargor utilizando espectroscopia no infravermelho por transformada de Fourier (FTIR) [68] e análise de parâmetros como álcool e gravidade específica utilizando espectroscopia de ressonância magnética nuclear de alta resolução [69]. No entanto, os métodos citados não utilizaram equipamentos portáteis. Uma alternativa vantajosa para realizar o controle de qualidade é a combinação de técnicas como espectroscopia NIR e análise de imagens digitais com a quimiometria. Esse uso conjunto apresenta vantagens, como a exigência de volume menor de amostras e reagentes, preços mais baixos do que os instrumentos convencionais de bancada, além de poderem ser facilmente transportados para análise *in situ*.

2.4 Imagens Digitais

As imagens digitais são formadas por pixels, estes normalmente são organizados em uma matriz retangular ordenada. O tamanho de uma imagem é determinado pelas dimensões dessa matriz de pixels. A largura da imagem é o número de colunas e a altura da imagem é o número de linhas na matriz, formando assim uma matriz de M colunas e N linhas. Além do tamanho da imagem, existem dois outros parâmetros importantes a serem citados, que são a resolução e a intensidade. A resolução é a escala espacial dos pixels da imagem. Por exemplo, uma imagem de 3000 x 2500 pixels com uma resolução de 300 pontos por polegada (dpi) resultaria em um tamanho de imagem de 15 x 21 cm. Uma outra unidade que pode ser utilizada na resolução é ppi (pixels por polegada). Finalmente, a intensidade é necessária para definir uma imagem, pois é a mistura de intensidades de cores que produzem uma imagem colorida. Cada pixel tem seu próprio valor de intensidade, também conhecido

como brilho. Os dois tipos mais comuns de imagens digitais, preto e branco e colorida, são conhecidos como imagens em tons de cinza e RGB, respectivamente. Além do tipo de intensidade de cada pixel, a faixa de valores de intensidade também varia. Os valores de intensidade em imagens digitais são definidos por bits. Um bit é binário e tem apenas dois valores possíveis, 0 ou 1. Um intervalo de intensidade de 8 bits tem 256 valores possíveis, de 0 a 255. Isso pode ser visto matematicamente através da expressão $2^{n^{\circ} \text{ de bits}}$. A foto digital padrão usa um intervalo de valores de 8 bits. As imagens RGB usam intervalos de intensidade de 8 bits para cada cor e as imagens em preto e branco têm um único intervalo de intensidade de 8 bits [70].

Dentre as variáveis analíticas, a cor, percebida pelo observador, tem sido utilizada na análise química como uma das principais propriedades organolépticas que codifica informações qualitativas e semiquantitativas em reações como as de identificação e de separação. O surgimento de detectores de cores e sensores de imagem que duplicam e até aprimoram a visão humana por meio da aplicação de *software* e *hardware* próprios para essa finalidade, juntamente com o desenvolvimento de novas metodologias seletivas, abriu as portas para novas estratégias e oportunidades em química analítica. O uso de imagens digitais como método analítico tem se destacado no meio científico, uma vez que os métodos de captura de imagem se tornaram uma alternativa viável frente a outras técnicas analíticas já estabelecidas no cenário científico [71].

A aplicação da quimiometria ao processamento de imagens digitais deu origem à análise multivariada de imagens (MIA - *Multivariate Image Analysis*) e a regressão multivariada de imagens (MIR - *Multivariate Images Regression*) [72-73]. Nos últimos 15 anos, a análise química com base em medidas usando cores obtidas com os dispositivos de imagem aumentou muito, e essa tendência trouxe novos métodos analíticos simples, diretos, rápidos, *in situ* e de baixo custo, embasados em procedimentos analíticos baseados em visão computacional (CVAC - *Computer Vision Algorithm Collection*) [71].

A grande difusão de equipamentos portáteis para a aquisição de imagens digitais impulsionou o uso da cor como parâmetro para extrair informações químicas qualitativas e quantitativas e contribuiu para a democratização da ciência analítica [74]. Câmeras digitais, escâneres e *smartphones* equipados com câmeras de alta

resolução estão hoje disponíveis a preços muito baixos em comparação com a instrumentação analítica tradicional.

As estratégias de processamento para análise de imagens digitais podem variar bastante, desde regressões simples e univariadas, a modelos multivariados e até multidimensionais. Quando a detecção de imagem fornece um sinal seletivo, não há necessidade de quimiometria, podendo se utilizar nesses casos modelos univariados. Alguns exemplos disto incluem um método baseado em imagens digitais para determinar a acidez total em vinhos tintos usando a titulação ácido-base sem indicador [75] e uma plataforma de detecção por telefone celular para determinar biomarcadores no leite com o objetivo de checar contaminação [76]. Quando a matriz é mais complexa e/ou existem faixas analíticas mais amplas de analitos e interferências, a quimiometria apresenta potencial para ser usada. Uma das estratégias analíticas mais populares para MIA é a de modelar histogramas RGB (vermelho, verde e azul) por calibração multivariada, principalmente usando PLS. Esses histogramas podem ser modelados de maneira semelhante aos dados espectroscópicos tradicionais. Essa estratégia quimiométrica já foi aplicada na análise de alimentos e bebidas, como na quantificação de corantes artificiais em bebidas [77-78], na predição de biomassa de levedura imobilizada durante uma fermentação alimentar [79], na rápida determinação de etanol em aguardentes de cana [80], na avaliação da estabilidade térmica do leite [81] e na detecção de adulterações em hambúrgueres de frango [82]. Matrizes químicas mais complexas podem exigir estratégias mais sofisticadas, como a construção de modelos não lineares com máquinas de vetores de suporte (SVM - *Support Vector Machines*) [83].

2.5 Validação Analítica Multivariada

A validação de métodos analíticos multivariados ainda não é um assunto totalmente estabelecido. Uma concordância entre os aspectos de validação de métodos univariados e multivariados é essencial, pois o surgimento de novas metodologias multivariadas tem se tornado uma tendência recente. Para o reconhecimento de qualquer método como referência para a análise, a validação analítica multivariada é um procedimento fundamental [84].

Conceitos como seletividade, sensibilidade e curvas de calibração devem ser repensados e adaptados para validação multivariada. Os guias clássicos de validação comumente exigem que o método analítico seja seletivo em relação ao analito, o que não é coerente para calibração multivariada, pois seu uso só é vantajoso na ausência de um sinal seletivo. A construção de curvas de calibração convencionais para avaliar a linearidade também não é possível em métodos multivariados. Uma forma qualitativa de estimar a linearidade nesses casos seria analisar a distribuição dos resíduos do modelo, que devem ter comportamento aleatório para amostras de calibração e validação [36].

Entre os esforços feitos nos últimos anos para harmonizar aspectos univariados e multivariados de validação, surgiu um guia crucial que indica as práticas padrão para análise quantitativa multivariada utilizando espectroscopia no infravermelho [84] (mas, cujas recomendações podem se aplicar também a dados oriundos de outras técnicas analíticas). O guia ASTM abrange os critérios para validar o desempenho da calibração multivariada. As figuras de mérito (FOM) necessárias para a validação analítica são especificidade/seletividade (SEL), linearidade, exatidão, precisão e faixa de trabalho. Essas figuras de mérito são as mesmas indicadas por órgãos como o Instituto Nacional de Metrologia, Qualidade e Tecnologia (INMETRO) [84-86] Essas figuras de mérito são as mesmas citadas nos guias publicados pela AOAC/ISO/IUPAC [102], EURACHEM (2014) [186] e União Europeia (EC, 2021) [187].

A equação estatística usada para a raiz quadrada do erro médio (RMSE) é (Equação 1):

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

em que \hat{y}_i e y_i são respectivamente as concentrações do analito calculada/predita e de referência, e N o número de amostras. A equação do RMSE pode ser aplicada à calibração (RMSEC), à validação cruzada (RMSECV), e à validação externa (RMSEP). A exatidão em métodos multivariados é normalmente expressa através dos valores RMSEC e RMSEP. Este último é um parâmetro mais robusto, pois é estimado a partir de um conjunto externo independente de amostras de validação, não utilizado para construção do modelo [85].

A precisão costuma ser avaliada em dois níveis, repetibilidade e precisão intermediária. A repetibilidade expressa a variabilidade analítica sob as mesmas condições de trabalho dentro de um curto intervalo de tempo, e é determinada analisando repetidamente a mesma amostra de forma consecutiva de 6 a 10 vezes, de acordo com diretrizes nacionais e internacionais. Já a precisão intermediária inclui a influência de efeitos aleatórios dentro dos laboratórios, de acordo com o uso pretendido do procedimento, como diferentes dias e analistas. Ao confirmar que o método fornece um grau aceitável de exatidão, precisão e linearidade quando aplicado a amostras em toda a faixa do intervalo especificado do procedimento analítico, a faixa analítica é estabelecida [85].

Assim como os métodos quantitativos, os métodos qualitativos também devem passar por procedimentos de validação analítica. Em 2002, a União Europeia publicou um guia que cita parâmetros aplicáveis para validação de métodos qualitativos, mas, somente em 2006 que guias específicos começaram a ser publicados na literatura [188-189, 89]. Com o visível crescimento do número de métodos multivariados de classificação supervisionada, é necessário discutir como proceder à sua validação, tal como tem sido feito com os métodos de calibração multivariada. No entanto, este assunto ainda está em debate na literatura [87].

O primeiro passo para desenvolver um método qualitativo é definir um limiar que caracterize uma amostra como 0 (negativo/ausência) ou 1 (positivo/presença). Os métodos qualitativos modernos geralmente se baseiam em leituras instrumentais para classificar as amostras, e esse valor é chamado de *threshold* bayesiano, estimado usando a teoria de Bayes [88]. Depois que o limite é definido, o próximo passo é determinar o número de amostras de falso positivo (FP) e falso negativo (FN). FP é o número de amostras que não pertencem à classe alvo e são classificadas como pertencentes, por exemplo, uma amostra não adulterada sendo classificada como tal, e FN seria o parâmetro oposto. Em geral, a classe 1 é definida como relativa a amostras autênticas, embora isto não seja estritamente obrigatório e haja variação na literatura, o que às vezes pode causar confusão.

As figuras de mérito (FOM) qualitativas são estimadas utilizando os valores de FP e FN: Taxa de Falso Negativo (TFN), Taxa de Falso Positivo (TFP), Taxa de Seletividade, Taxa de Especificidade e Taxa de Eficiência [36,88]. Embora alguns

FOM utilizadas em validações qualitativas e quantitativas tenham o mesmo nome, os conceitos e formas de avaliá-las são diferentes. Figuras de mérito como Taxa de Falso Positivo, Taxa de Falso Negativo, Taxa de Eficiência, taxas de sensibilidade e especificidade expressam exatidão nas análises qualitativas, já outras FOM como Acordância (ACO) e Concordância (CON) expressam a precisão do método. Os parâmetros ACO e COM são considerados aceitáveis quando seus valores estimados se aproximam de 1 em uma faixa de concentração relevante para a aplicação do método. A ACO é a probabilidade de que duas amostras idênticas forneçam a mesma resposta (positiva ou negativa) quando realizadas pelo mesmo laboratório ou lote sob condições de repetibilidade. CON é a FOM qualitativa equivalente à estimativa de precisão intermediária ou reprodutibilidade em métodos quantitativos. Ele estima a probabilidade de duas amostras da mesma classe fornecerem o mesmo resultado em laboratórios ou lotes diferentes [89].

Tabela 1. Descrição dos parâmetros e suas fórmulas

Parâmetros	Fórmulas
Taxa de Falso Positivo (TFP)	$\frac{FP}{TN + FP}$
Taxa de Falso Negativo (TFN)	$\frac{FN}{FN + TP}$
Taxa de Sensibilidade (TSL)	$\frac{TP}{TP + FN} \cdot 100$
Taxa de Especificidade (TES)	$\frac{TN}{TN + FP} \cdot 100$
Taxa de Eficiência (TCF)	$100 - (TFP + TFN)$
Acordância	$ACO = \frac{\{k(k - 1) + (n - k)(n - k - 1)\}}{n(n - 1)}$
Concordância	$CON = \frac{2[k(k - nb)] + nb(nb - 1) - ACC[nb(n - 1)]}{n^2b(b - 1)}$

n = os resultados totais obtidos para um lote analítico; k = os resultados positivos obtidos para cada lote analítico; b = número de lotes analíticos; TN = verdadeiro negativo; TP = verdadeiro positivo

**CAPÍTULO 3: DESENVOLVIMENTO DE UM MÉTODO SIMPLES E RÁPIDO
PARA DETERMINAÇÃO DE COR EM CERVEJAS USANDO IMAGENS DIGITAIS**

3 DESENVOLVIMENTO DE UM MÉTODO SIMPLES E RÁPIDO PARA DETERMINAÇÃO DE COR EM CERVEJAS USANDO IMAGENS DIGITAIS

Como mencionado anteriormente, a cerveja é definida como uma bebida alcoólica produzida através da fermentação de malte, lúpulo, água e levedura [1,2]. Como para qualquer produto industrializado, o controle de qualidade da bebida finalizada também é fundamental na indústria cervejeira [65]. O desenvolvimento de novos métodos analíticos juntamente com a rápida avaliação dos resultados para o controle de qualidade são ferramentas de grande importância para aprimorar as técnicas de produção de cerveja. Além disso, um dos principais objetivos tem sido o de reduzir o tempo de resposta, simplificando o procedimento analítico sem comprometer os resultados. Para esta tarefa, os métodos quimiométricos ganharam relevância nos últimos anos. Esses métodos multivariados geralmente dependem de dados obtidos principalmente por espectrofotômetros de bancada (e relacionados a técnicas moleculares). No entanto, algumas alternativas de baixo custo surgiram recentemente, baseadas em equipamentos portáteis, como os usados para obter imagens digitais.

Entre as análises mais importantes utilizadas para avaliar a qualidade das cervejas, destaca-se a determinação da cor. A cor é a primeira característica sensorial das cervejas que chama a atenção do consumidor [90]. A determinação da cor é realizada de acordo com os dois sistemas atualmente aceitos, EBC (*European brewing convention* ou Convenção de Cervejeiros Europeus) e SRM (*standard reference method* ou método de referência padrão), ambos baseados em medições espectrofotométricas de absorção no visível, a 430 nm [91,92].

É possível citar trabalhos que empregaram ferramentas quimiométricas para o controle de qualidade das cervejas. Dentre eles, em 2007, Lachenmeier utilizou a espectroscopia no infravermelho por transformada de Fourier (FTIR) combinada com mínimos quadrados parciais (PLS) para prever nove parâmetros de cervejas, incluindo densidade relativa, teor de álcool, pH, amargor e cor na escala EBC [68]. No entanto, o modelo de calibração para cores apresentou baixas exatidão e linearidade,

conforme reconhecido pelo próprio autor. Blanco et al., em 2015, desenvolveram um método discriminante de estilos de cerveja e quantificaram a cor e o teor de álcool através de uma língua eletrônica portátil baseada em eletrodos impressos [93]. Neste artigo, o PLS forneceu uma boa linearidade para a determinação de cor, porém foi utilizado um pequeno número de amostras (25) e o modelo não foi validado com um conjunto de amostras independentes. Fengxia *et al.* em 2004 empregaram um escâner de mesa e uma câmera CCD para determinar a cor (escala EBC) de cervejas através de um modelo univariado com base no valor da saturação, um parâmetro estimado a partir das informações do histograma de cores RGB [94]. Uma desvantagem deste artigo foi sua faixa de trabalho muito estreita de valores de EBC (variando de 4 a 11, o que equivale a 2-6 unidades SRM), uma vez que apenas cervejas de cores claras foram usadas para construir o modelo.

3.1 Objetivos

O objetivo desta aplicação foi desenvolver um método de calibração multivariada, utilizando imagens digitais adquiridas através de um telefone celular e um modelo PLS para a determinação da cor de cervejas. Pretendeu-se construir um modelo robusto, utilizando uma grande variedade de cervejas, de diferentes estilos, marcas e cervejarias, e validá-lo de acordo com as diretrizes brasileiras e internacionais, com o objetivo de garantir sua confiabilidade e adequação ao uso pretendido.

3.2 Materiais e métodos

3.2.1 Equipamentos e softwares

Todas as imagens foram obtidas com um telefone celular iPhone® 7 (Apple, Cupertino, EUA), com capacidade de 128 GB, equipado com câmera de 12 MP e sensor Sony Exmor RS. Os dados foram tratados usando o *software* MATLAB®, versão 7.13 (MathWorks, Natick, EUA), juntamente com o PLS Toolbox, versão 6.5 (Eigenvector Technologies, Manson, EUA), e o Image Processing Toolbox, versão 8.0 (MathWorks). Um espectrofotômetro Agilent Cary 60 UV/vis (Agilent Technologies,

Santa Clara, EUA) foi utilizado para obter as medidas para o método de referência [92].

3.2.2 Amostras

Cento e vinte e oito amostras diferentes de cervejas foram adquiridas majoritariamente no comércio local de Belo Horizonte, MG, Brasil. As amostras foram armazenadas em temperatura ambiente até a análise. Foram obtidas amostras de cerveja de uma grande variedade de marcas e estilos, incluindo lager, weissbier, wit, IPA, bock, red ale, brown ale, dunkel, stout, porter, etc.

3.2.3 Procedimentos

Aproximadamente 300 mL de cada amostra foram coletados e desgaseificados usando banho ultrassônico por 5 min. Algumas gotas de 1-octanol (Neon, Suzano, Brasil) foram previamente adicionadas a cada amostra como um agente antiespumante. Na sequência, 25 mL de cada amostra foram transferidas para uma pequena placa de Petri (raio de 3,0 cm e 1,5 cm de altura) embutida em uma câmara escura de madeira, desenvolvida especificamente para a aquisição da imagem (Figura 6). Todas as imagens foram obtidas em 4032 x 3024 pixels (12 MB) no formato “.jpg”, com 16,8 milhões de cores e resolução de 300 dpi. A conversão das imagens nos histogramas RGB foi realizada no MATLAB. Primeiramente, foi selecionada uma área de 300x300 pixels a partir da parte central da placa, na parte homogênea da imagem. Essa área selecionada foi decomposta em um histograma RGB contendo 768 canais (256 para cada cor RGB) para cada amostra. Cada amostra foi digitalizada em triplicata e seu histograma médio foi utilizado para a construção de modelos PLS.

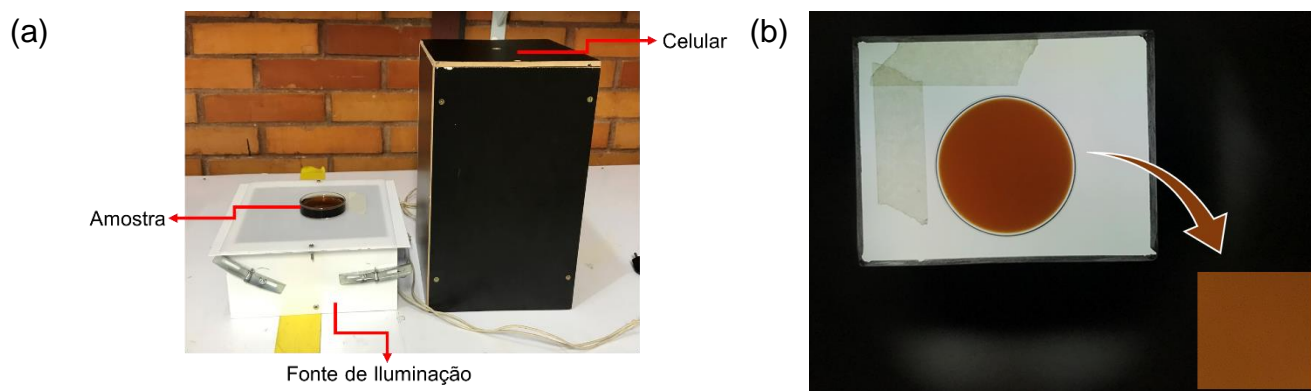


Figura 6. (a) Sistema usado para adquirir as imagens digitais. A câmera de madeira fabricada juntamente com sistema de iluminação; nela está uma pequena placa de Petri contendo 25 mL de amostra de cerveja (b) Uma imagem adquirida de uma placa de Petri como exemplo.

3.2.4 Análise espectrofotométrica (método de referência)

A análise espectrofotométrica foi realizada de acordo com o método oficial ASBC (*American Society of Brewing Chemists*) Beer-10 [92]. Este método se baseia em medições de absorvância a 430 nm das amostras contidas em uma cubeta de quartzo com percurso óptico de 10 mm. As amostras de cerveja não isentas de turbidez foram centrifugadas em uma centrífuga Awel MF 20-R (Awel SAS, Blain, França) por 5 min, a 3000 rpm. As cervejas escuras necessitaram uma etapa de diluição, pois suas absorvâncias passariam do valor de 1,0. Sendo assim, elas foram diluídas na proporção 1:10 (v/v) com água deionizada.

3.2.5 Modelos multivariados e validação analítica

Algumas tentativas de construir modelos univariados com base na frequência de canais unitários de RGB foram realizadas. Vários canais (2, 210, 222, 273, 456, 493, 530, 625 e 697) foram escolhidos de acordo com sua importância para o modelo multivariado, ou seja, canais com os maiores coeficientes de regressão em módulo. No entanto, nenhum desses modelos univariados forneceu exatidão e/ou linearidade aceitáveis, todos apresentando coeficientes de correlação abaixo de 0,80. Modelos construídos com as médias das cores dos canais vermelho, verde e azul também foram testados, mas igualmente não apresentaram bons resultados. Portanto, o uso da calibração multivariada é justificado. Um modelo PLS multivariado inicial foi

construído com todas as amostras de cerveja. Este modelo global apresentou erros de previsão muito grandes para algumas amostras na parte inferior da escala SRM (abaixo de 12 unidades). Desse modo, considerando as diluições utilizadas no método de referência, dois novos submodelos de calibração foram construídos. Um modelo PLS construído para a parte superior da escala SRM (amostras diluídas, com valores de referência acima de 12) e outro modelo PLS para amostras com valores de referência abaixo de 12 unidades. Um modelo discriminante (PLS-DA) foi construído para indicar em qual submodelo as amostras seriam inseridas.

Essa seleção resultou em 78 e 50 amostras nos submodelos das faixas superior e inferior da escala SRM, respectivamente. Esse modelo discriminante é essencial para os usuários dessa metodologia, uma vez que o analista de uma cervejaria não sabe de antemão qual modelo local de PLS deve ser usado para determinar amostras de cerveja cuja cor está próxima do limite entre as duas classes. Em resumo, a metodologia proposta consistiu em duas etapas. Um modelo PLS-DA é usado para atribuir o histograma da amostra de cerveja original a um desses dois modelos PLS locais. Na segunda etapa, a previsão da cor é fornecida pelo respectivo modelo de calibração. É importante ressaltar que, para o método quimiométrico proposto, nenhuma das amostras é diluída, pois a etapa de diluição é necessária apenas para o método de referência.

Em relação à regulamentação na área de análise de alimentos não há menção ao uso de estatística multivariada em grande parte das diretrizes de validação nacionais e internacionais. No entanto, para o método proposto foram estimados: exatidão, precisão, linearidade e faixa de trabalho. Além disso, as incertezas dos valores previstos foram calculadas a fim de estimar o intervalo de confiança. Adicionalmente, erros de previsão padrão específico das amostras (SPE – *Standard Prediction Errors*) foram calculados baseados na equação dos erros nas variáveis (EIV – *Errors in Variables*). Vale destacar que, considerando a natureza discreta da variável em análise (cor) os erros de calibração (RMSEC) e previsão (RMSEP) foram arredondados e representados sem nenhuma casa decimal.

3.3 Resultados e discussão

3.3.1 Modelo global de PLS

Inicialmente, foi construído um modelo contendo todas as amostras de cerveja, correspondendo à faixa analítica completa, de 3 a 210 unidades SRM. Apesar de ter um bom ajuste, esse modelo global de PLS apresentou erros relativos muito grandes para algumas amostras na parte inferior da escala SRM. Isso pode ser observado no gráfico de resíduos em função dos valores previstos (Figura 7). Para superar esse problema, as amostras foram divididas em duas classes para a construção de modelos PLS locais, de acordo com a necessidade ou não de diluição no método espectrofotométrico de referência. Verificou-se que as amostras que apresentam absorbâncias a 430 nm acima do limite de 1,0 correspondem a unidades SRM acima de 12.

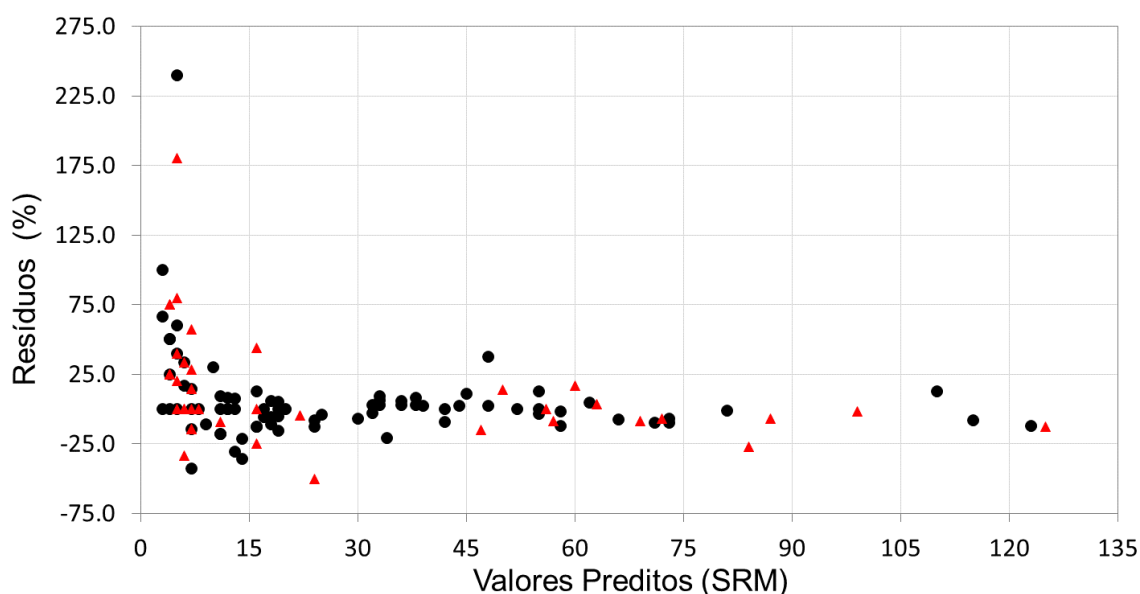


Figura 7. Gráfico dos resíduos das amostras de calibração e validação para o modelo PLS global. Amostras de calibração (círculos pretos) e validação (triângulos vermelhos)

Pode-se argumentar que, em situações reais, a atribuição de uma amostra de cerveja a um desses dois modelos locais não é previamente conhecida. Portanto, um modelo PLS-DA foi construído como uma etapa anterior para tomar essa decisão. Ele foi utilizado como modelo de triagem para a atribuição de amostras de cerveja a um dos dois modelos PLS locais, os de faixas superior e inferior na escala SRM. Como

uma absorvância de 1,0 é equivalente a 12 unidades SRM, esse valor foi escolhido como limite de decisão entre as duas classes modeladas.

Modelos PLS baseados em outros sistemas de cores, como HSV e CMYK, também foram testados, mas não foram observadas melhorias significativas em relação ao modelo usando as frequências dos histogramas RGB como variáveis independentes.

3.3.2 Modelo PLS-DA

Na prática, sem o modelo discriminante, não há como indicar a qual submodelo (faixa superior ou inferior) as amostras de cerveja seriam direcionadas. Nesse sentido, o modelo PLS-DA foi construído previamente para essa decisão ser tomada. Nesse modelo, os histogramas RGB das amostras foram organizados em uma matriz \mathbf{X} (128x768), na qual as amostras foram colocadas nas linhas e as frequências dos canais de cor nas colunas. O modelo foi construído correlacionando a matriz \mathbf{X} de variáveis independentes com um vetor \mathbf{y} de variáveis dependentes. Esse vetor contém valores de referência 1 para as amostras da faixa superior e 0 para as amostras da faixa inferior da escala SRM.

As amostras foram divididas em dois conjuntos, de treinamento e de teste. Para seleção das amostras de treinamento (cerca de 2/3 do total) foi usado o algoritmo de Kennard-Stone separadamente para cada classe [95]. Esse algoritmo assegurou a seleção das amostras de acordo com uma distribuição homogênea e representativa em toda a faixa analítica do método.

O melhor número de variáveis latentes (VL) foi escolhido por meio da validação cruzada em subconjuntos aleatórios (9 divisões e 20 iterações), de acordo com o menor erro de classificação cruzada (CVCE). O melhor modelo PLS-DA foi obtido com 2 VL, representando 61,3% da variância no bloco \mathbf{X} e 87,5% no bloco \mathbf{Y} . O limite Bayesiano foi estimado em 0,59. A Figura 8 apresenta o gráfico de classificação, mostrando os valores de \mathbf{Y} previstos de todas as amostras. Como pode ser observado, quase todas as amostras foram corretamente classificadas, com apenas um resultado falso negativo e um falso positivo observados no conjunto de treinamento, enquanto nenhuma amostra foi classificada erroneamente no conjunto teste.

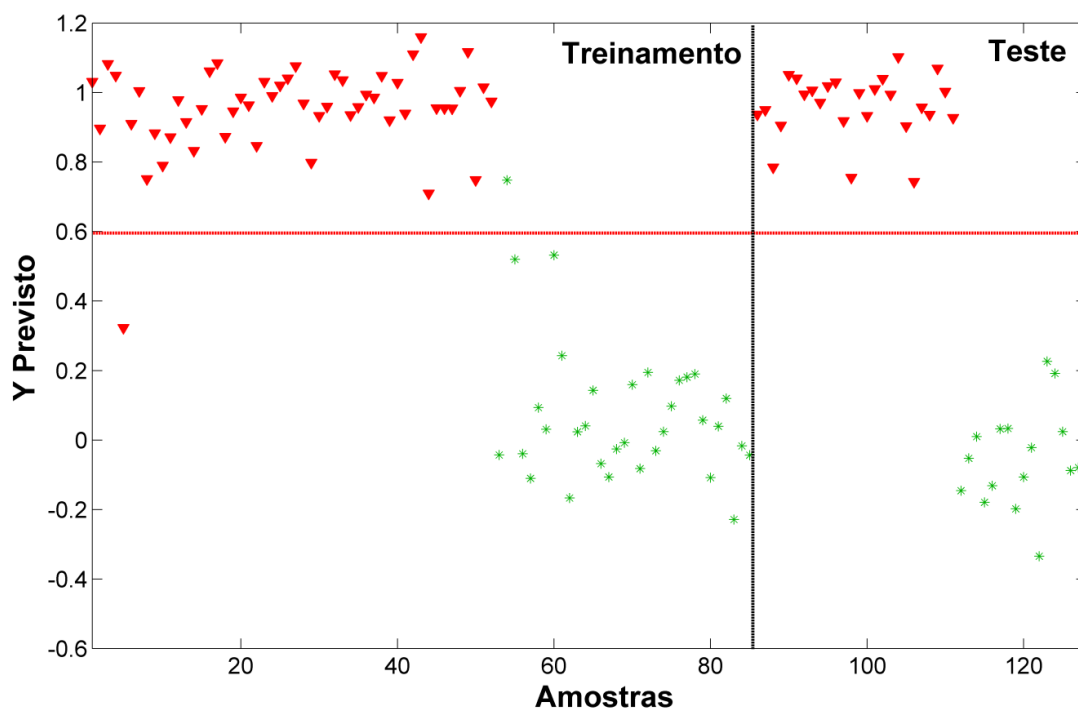


Figura 8. Valores previstos de Y para o modelo PLS-DA. A linha tracejada horizontal indica o limite de classificação e a linha tracejada vertical indica a separação entre as amostras de treinamento e as amostras de teste. Amostras da faixa de coloração superior (triângulos vermelhos) e amostras da faixa de coloração inferior (asteriscos verdes)

A interpretação de modelos PLS-DA e PLS pode ser feita através dos vetores informativos, tais como os coeficientes de regressão e os VIP escores. O último vetor mede a importância de cada variável na projeção usada por um modelo particular, em valores absolutos. Portanto, os VIP escores não permitem observar a contribuição da direção de cada variável. Essa informação pode ser estimada por meio dos vetores de regressão [54]. Para o modelo PLS-DA desenvolvido, os coeficientes de regressão e os VIP escores são mostrados nas [Figura 9\(a\)](#) e [Figura 9\(b\)](#), respectivamente. Em termos de interpretação, há uma estreita concordância entre esses dois vetores, ambos indicando a relação das variáveis que mais contribuem para o modelo em tons mais escuros de vermelho e azul, associados à faixa superior de coloração, e tons mais claros de verde, associados com a faixa inferior de coloração.

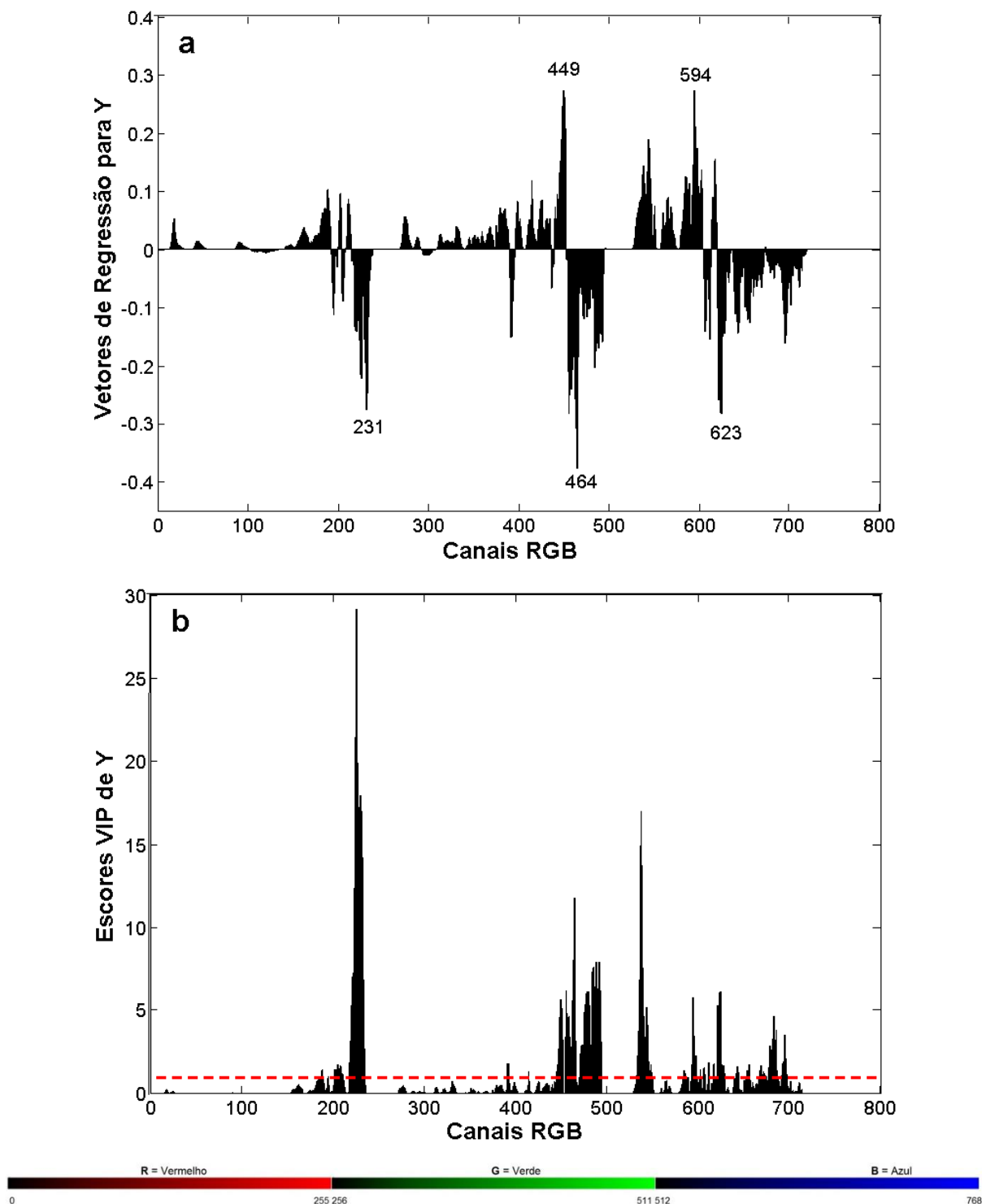


Figura 9. Vetores informativos para o modelo PLS-DA. **(a)** vetores de regressão e **(b)** VIP escores

Neste trabalho, algumas figuras de mérito (FOM) foram estimadas. Os cálculos utilizados para essas FOM foram baseados no artigo de Botelho *et al.* de 2015 [96]. Como pode ser visto na Tabela 2, cinco diferentes FOM foram estimadas para os conjuntos de treinamento e teste.

Tabela 2. Tabela contendo as seguintes FOM: FPR (taxa de falso positivo), FNR (taxa de falso negativo), STR (taxa de sensibilidade), SPR (taxa de especificidade) e EFR (taxa de eficiência)

	FPR (%)	FNR (%)	STR (%)	SPR (%)	EFR (%)
Treinamento	3,03%	1,92%	98,08%	96,97%	95,05%
Teste	0,00%	0,00%	100,00%	100,00%	100,00%

Para o conjunto de treinamento, um STR de 98,08% indicou uma alta taxa de verdadeiros positivos, enquanto um SPR de 96,97% significou que um falso positivo foi observado. Para o conjunto de testes, valores de SPR e STR de 100% implicam na ausência de falsos positivos e falsos negativos, respectivamente. Como consequência, também foram obtidos altos valores de EFR, uma vez que essa FOM expressa a diferença entre o total de resultados (100%) e a soma de FNR e FPR.

3.3.3 Modelos PLS locais

Como já mencionado anteriormente, dois modelos locais de calibração foram construídos. As 128 amostras foram divididas em 78 (faixa superior de coloração) e 50 (faixa inferior de coloração). Para cada um desses dois modelos PLS as amostras foram divididas em dois terços para calibração e um terço para validação, por meio do algoritmo de Kennard-Stone [95]. Para o modelo da parte superior, isso correspondeu a 52 e 26 amostras para a calibração e validação, respectivamente. Para o modelo da parte inferior, isso correspondeu a 33 e 17 amostras para a calibração e validação, respectivamente. O número de variáveis latentes foi selecionado de acordo com o menor valor de RMSECV, estimado usando validação cruzada por subconjuntos aleatórios (7 divisões e 20 iterações). Os melhores modelos foram construídos com 3 VL para ambos os modelos, faixa superior e faixa inferior.

Ambos os modelos foram otimizados por meio da detecção de *outliers*. Foram consideradas amostras *outliers* as que apresentaram altos resíduos em X , altos resíduos em Y ou altos valores de *leverage*, com nível de confiança de 95% [36,77]. O número de *outliers* detectados não excedeu o limite recomendado de $2/9$ (22,2%) [97, 100, 190]. Para o modelo da parte superior 8 amostras (10,3%) foram deletadas (3 amostras ou 5,8% da calibração e 5 amostras ou 19,2% da validação).

Tabela 3. Resultados da otimização do modelo PLS da faixa superior de escala SRM desenvolvido através da detecção e remoção de *outliers*. RMSEC (Raiz quadrada do Erro Médio Quadrático da Calibração), RMSEP (Raiz quadrada do Erro Médio Quadrático da Previsão) e RPD (Relação de Desempenho do Desvio).

	Nº de amostras de calibração	Nº de amostras de validação	Nº de variáveis latentes (VL)	RMSEC ^a	RMSEP ^a	RPD validação
Antes	52	26	3	10	27	1,5
Depois	49	21	3	5	6	5,6

^aUnidade SRM

Para o modelo da parte inferior, 3 das 50 amostras originais (17,6% da validação) foram deletadas como *outliers*. Após os *outliers* terem sido removidos, ambos os modelos foram reconstruídos. Os resultados da otimização dos dois modelos (RMSEC, RMSEP e RPD) estão apresentados nas Tabela 3 (modelo superior) e Tabela 4 (modelo inferior). A sigla RPD indica a relação de desempenho do desvio. O conceito de relação de RPD foi introduzido em 1986 por Phil Williams [98]. O RPD é a razão entre o desvio padrão da previsão e o erro quadrático médio da previsão (RMSEP) ou o erro quadrático médio da validação cruzada (RMSECV). O RPD pode ser interpretado como a razão da variação natural nas amostras pelos erros de previsão. Se o RPD for aproximadamente 1, ou seja, se a variação da amostra e os erros quadráticos tiverem valores aproximados, a previsão do método não será considerada boa para discriminação das amostras. Se este valor estiver acima de 2,4, é possível considerar que o método possui uma boa capacidade preditiva [36,99].

Tabela 4. Resultados da otimização do modelo PLS da faixa inferior de escala SRM desenvolvido através da detecção e remoção de *outliers*. RMSEC (Raiz quadrada do Erro Médio Quadrático da Calibração), RMSEP (Raiz quadrada do Erro Médio Quadrático da Previsão) e RPD (Relação de Desempenho do Desvio).

	Nº de amostras de calibração	Nº de amostras de validação	Nº de variáveis latentes (VL)	RMSEC ^a	RMSEP ^a	RPD validação
Antes	33	17	3	1	1	1,5
Depois	33	14	3	1	1	3,1

^aUnidade SRM

Como já foi dito anteriormente, RMSEC e RMSEP são parâmetros que descrevem a veracidade/exatidão do método. Para o modelo da parte superior, após a retirada dos *outliers*, os valores de RMSEC e RMSEP foram 5 unidades e 6 unidades SRM, respectivamente. Para o modelo da parte inferior de coloração, esses valores

foram ambos iguais a 1 unidade SRM. Os valores do RPD de validação foram 5,6 e 3,1 para os modelos da parte superior e inferior, respectivamente. Esses resultados são satisfatórios, uma vez que modelos PLS com valores de RPD acima de 2,4 são considerados com uma boa capacidade de previsão [36,99]. Os gráficos dos valores de referência *versus* previstos para os dois modelos podem ser vistos na Figura 10.

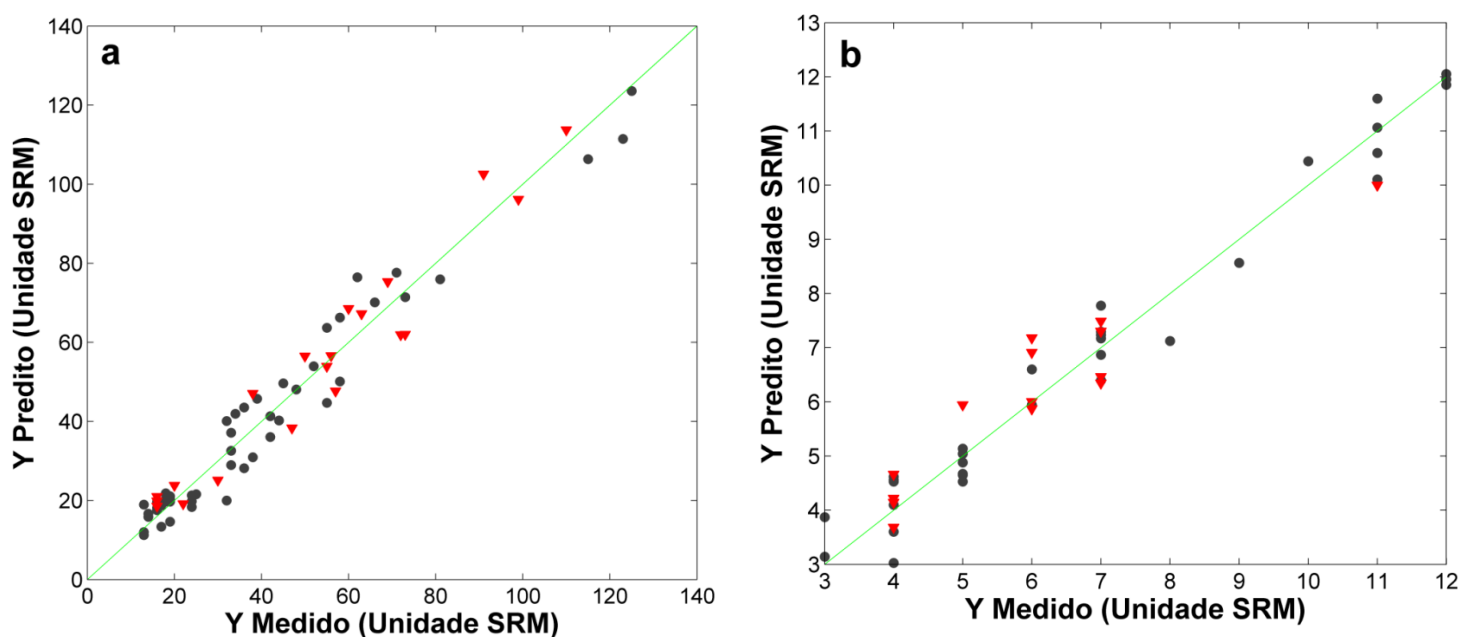


Figura 10. Valores de referência versus valores previstos para os modelos PLS (a) Modelo da faixa superior de SRM. (b) Modelo da faixa inferior de SRM. Amostras de calibração (círculos pretos) e amostras de validação (triângulos vermelhos)

Como apenas uma amostra presente no modelo da faixa superior apresentou SRM acima de 150 (210 unidades), ela foi excluída e um novo modelo foi desenvolvido na faixa de trabalho de 13 a 130 (Figura 10a).

Os coeficientes de regressão para ambos os modelos PLS são mostrados na Figura 11. Observando esses gráficos, é possível identificar os canais que mais contribuíram para a previsão da cor da cerveja. Canais do início da escala de cada cor primária no histograma (0 para vermelho, 256 para verde e 512 para azul) contribuíram para tons próximo do preto, enquanto canais que se encontram no final de cada parte do histograma RGB (255 para vermelho, 511 para verde e 768 para azul) representaram as cores primárias perto de suas intensidades máximas. Como observado na Figura 11a, os maiores coeficientes de regressão para o modelo da parte superior foram observados nos canais 96, 276 e 533, os quais apresentaram

forte influência na quantificação de cor da cerveja. Para o modelo da faixa inferior (Figura 11b), uma situação mais complexa é observada, com os canais 210, 456 e 625 tendo os mais positivos coeficientes de regressão, e os canais 222, 493 e 697 tendo os mais negativos.

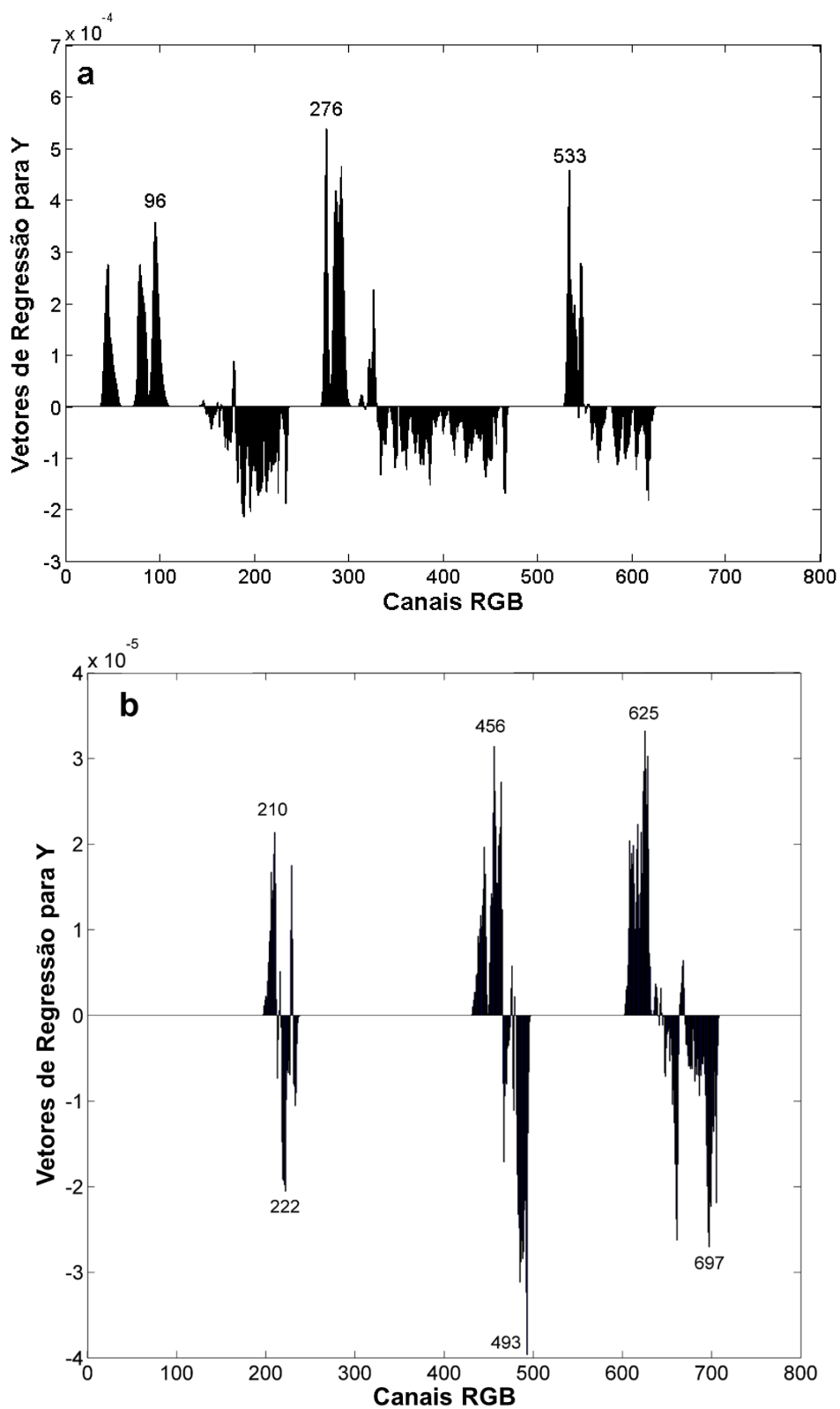


Figura 11. Vetores de regressão para o modelo PLS. **(a)** Modelo para a escala superior de SRM e **(b)** modelo para a escala inferior de SRM

Outras figuras de mérito estimadas para a validação dos métodos são mostradas na Tabela 5. A linearidade pode ser avaliada por meio da observação dos gráficos da Figura 10 e seus respectivos parâmetros de ajuste na Tabela 5 [89, 100-101].

Tabela 5. Parâmetros para avaliação das figuras de mérito dos modelos PLS locais

Figuras de Mérito	Parâmetros	Valores	
		Modelo Faixa Superior	Modelo Faixa Inferior
Precisão	Repetibilidade RSD (%)	0,7	0,8
		0,2	0,7
		0,5	0,4
	Precisão Intermediária RSD (%)	1,4	4,5
		2,2	3,4
Linearidade	Teste de Durbin-Watson	2,9	2,2
		2,24	2,06
	Inclinação	0,98±0,02	0,97±0,03
	Intercepto	1,05±0,95	0,19±0,22
Viés ^a	Coeficiente de Correlação	0,9743	0,9864
		1±5	0±1
Faixa de Trabalho ^a		13-130	3-12

^a Unidade SRM.

Coeficientes de correlação de 0,9743 e 0,9864 foram obtidos para os modelos da parte superior e inferior da escala SRM, respectivamente. O teste de Durbin-Watson foi aplicado a esses resíduos de ajuste para verificar sua independência. Valores estimados de 2,24 e 2,06 estão dentro da faixa de aceitação (1,50 - 2,50), o que confirma a ausência de autocorrelação no modelo. Um baixo valor de viés estimado para os modelos, considerando apenas as amostras de validação, constatou a ausência de erros sistemáticos. Os testes t de Student confirmaram que o viés zero está contido em ambos os intervalos de confiança a 95%.

As regiões elípticas de confiança conjunta (EJCR) também foram estimadas como uma ferramenta adicional para verificar a linearidade e a ausência de erros sistemáticos nos modelos (Figura 12). A estimativa de EJCR é baseada no intercepto e na inclinação da curva de ajuste. Para ambos os modelos, as estimativas da EJCR

(intervalo de confiança de 95%) continham o ponto ideal (inclinação 1,0 e intercepto 0,0).

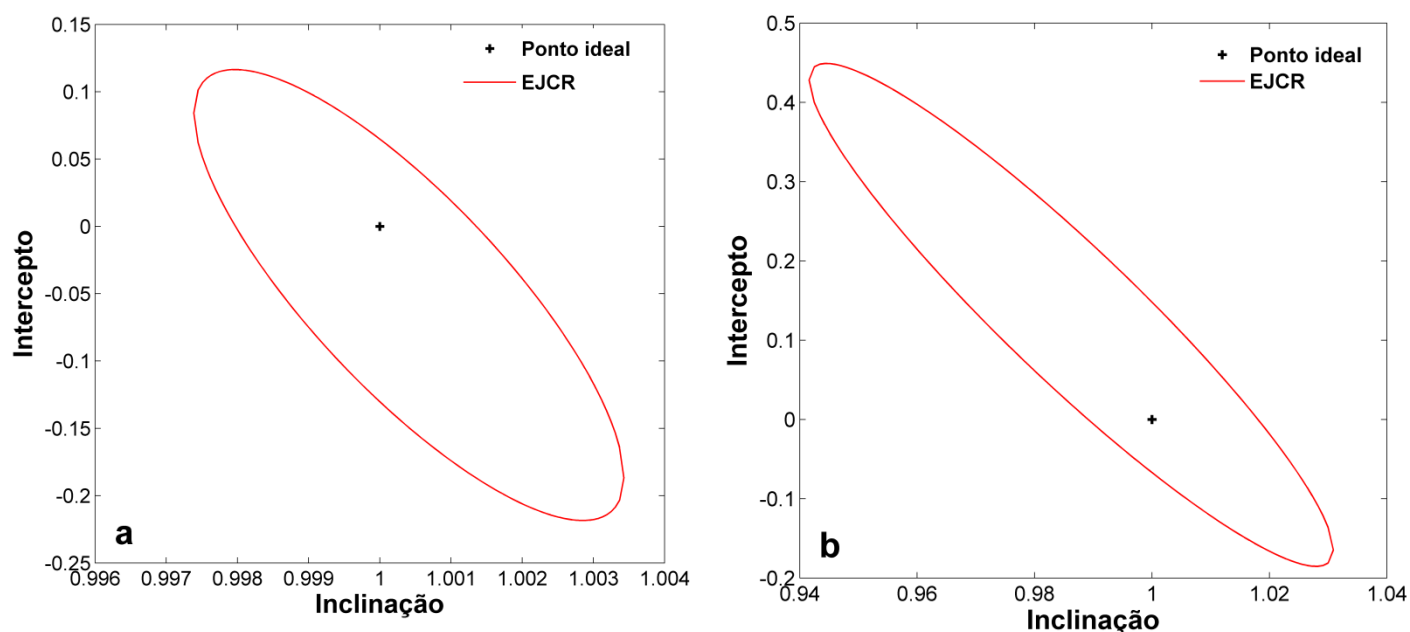


Figura 12. Gráfico da região de confiança elíptica para a inclinação e intercepto da curva dos valores previstos *versus* de referência para os modelos PLS **(a)** faixa superior de SRM **(b)** faixa inferior de SRM. Símbolos: ponto preto para o ponto ideal e elipse vermelha da EJCR.

A precisão do método foi avaliada por meio da estimativa dos desvios padrão relativos (RSD) para triplicatas de três amostras de cada modelo em dois níveis diferentes. Para cada modelo, essas amostras foram escolhidas de modo a representar valores de coloração baixo, médio e alto. Dois níveis de precisão, repetibilidade e precisão intermediária, foram avaliados. Para a repetibilidade, todas as triplicatas foram analisadas pelo mesmo analista no mesmo dia, enquanto para a precisão intermediária os resultados vieram de dois analistas diferentes em dias distintos. Os valores para a repetibilidade foram estimados em uma faixa de 0,2% a 4,5% (RSD) para o modelo de coloração superior, e entre 0,4% e 0,8% para o modelo de coloração inferior. A precisão intermediária (RSD) variou de 1,4% a 2,9% para o modelo da parte superior e de 2,2% a 4,5% para o modelo da parte inferior. Levando em consideração os bons resultados de exatidão e precisão, ambos os modelos podem ser considerados exatos e precisos. Portanto, as faixas de 3-12 e de 13-130 unidades SRM podem ser estabelecidas para os modelos da parte inferior e superior de coloração, respectivamente.

Por fim, intervalos de confiança foram estimados para a previsão das amostras de validação. O SPE calculado utilizando a equação EIV foi igual a 2 unidades de SRM para todas as amostras do modelo da faixa inferior, e variou de 4 a 5 unidades SRM para o modelo da faixa superior. A Figura 13 apresenta os valores de referência e previstos com os respectivos intervalos de confiança estimados para cada amostra de validação, em ambos os modelos. Para o modelo de coloração inferior, todos os valores de referência estavam dentro do intervalo de confiança estimado, representando uma concordância de 100%. Para o modelo de coloração superior, apenas uma em vinte e uma amostras estava fora do intervalo de confiança, representando 95,2% de concordância.

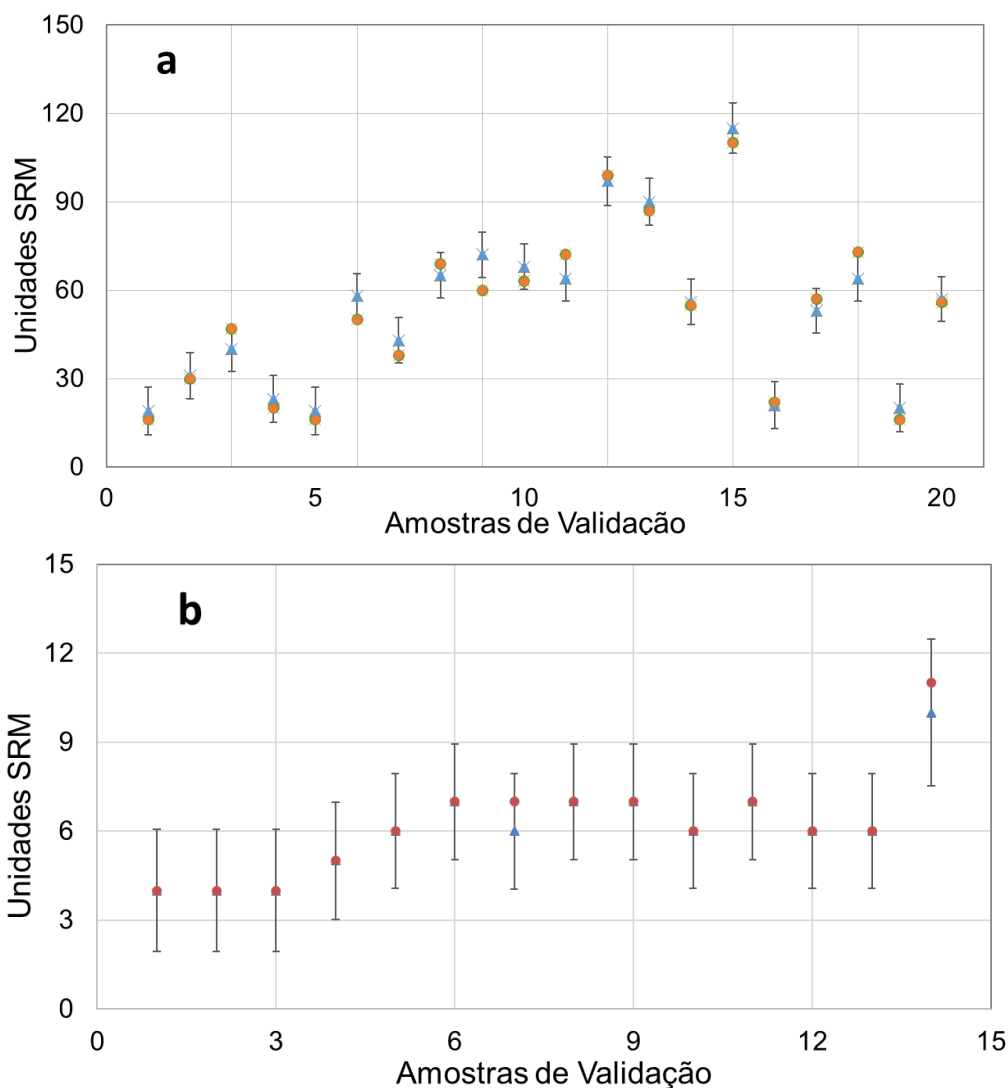


Figura 13. Intervalos de confiança para a previsão das amostras de validação, estimados baseando-se no SPE (*Standard Prediction Errors*). **(a)** modelo faixa superior de SRM **(b)** modelo faixa inferior de SRM. Valores previstos (triângulos) e valores de referência (círculos).

3.4 Conclusões

Um método multivariado baseado em imagens digitais obtidas em um aparelho celular, associadas à quimiometria, foi desenvolvido e validado para a determinação de cor de amostras de cerveja, de acordo com a escala de referência SRM. Devido à ampla faixa analítica (3-130 unidades), dois modelos PLS locais foram construídos, um que abrange a parte inferior de coloração e outro a parte superior. Esse novo método apresenta vantagens em relação ao método espectrofotométrico de referência, como um procedimento mais simples, mais rápido, de baixo custo e mínimo

pré-tratamento da amostra (retirada do gás). Esse método foi validado de acordo com as diretrizes nacionais e internacionais, sendo considerado exato, linear, preciso e adequado para o uso como uma metodologia oficial para a determinação de cor em cervejas [97,102]. No Brasil, assim como em outros países, o número de novas cervejarias artesanais tem apresentado um notável crescimento. Essas cervejarias muitas vezes são pequenas empresas, que produzem diversos estilos de cervejas, cujas cores variam das mais claras às mais escuras. Desse modo, o método proposto é adequado para esse nicho de mercado, especialmente se adaptado a uma plataforma analítica portátil, representando assim um grande potencial para o aumento da democratização da ciência analítica. Um artigo contemplando os resultados deste trabalho foi publicado na revista *Food Analytical Methods* (Apêndice A)

**CAPÍTULO 4: DETERMINAÇÃO DO TEOR DE ÁLCOOL EM CERVEJAS DE
DIFERENTES ESTILOS COM BASE EM ESPECTROSCOPIA NO
INFRAVERMELHO PRÓXIMO PORTÁTIL E CALIBRAÇÃO MULTIVARIADA**

4 DETERMINAÇÃO DO TEOR DE ÁLCOOL EM CERVEJAS DE DIFERENTES ESTILOS COM BASE EM ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO PORTÁTIL E CALIBRAÇÃO MULTIVARIADA

Neste capítulo, um espectrômetro portátil na região do infravermelho próximo foi usado para determinar o teor de álcool em amostras de cerveja. A cerveja é a bebida alcoólica mais consumida e comercializada no mundo, tendo a China seguida pelos EUA e Brasil como os principais produtores e a Alemanha como o principal exportador [103]. Em todo o mundo, o mercado cervejeiro é extremamente competitivo e cresceu significativamente, com exceção dos últimos dois anos. Apenas o mercado de cerveja dos EUA movimentou US\$ 114,2 bilhões de dólares em 2018 [104] e caiu para US\$ 94,1 bilhões de dólares em 2020 [105]. Cerca de 24% desse mercado é representado por cervejas artesanais. A produção e o consumo de cervejas artesanais tiveram um crescimento surpreendente nas últimas duas décadas, juntamente com a diversificação do produto em relação às suas características e estilos. Isto abriu novas oportunidades de negócios que vão desde a criação de um novo nicho de mercado no turismo, o turismo cervejeiro [106], até o mercado emergente de cervejas especiais de alto preço [107].

Uma tendência semelhante foi observada no Brasil, com um aumento de 30% no número de novas cervejarias entre 2017 e 2018 [108]. Muitas delas são pequenas cervejarias artesanais responsáveis por produzir diferentes estilos de cerveja, como lager, pale ale, weiss, stout e porter. Como citado anteriormente, a fabricação de cerveja é um processo complexo, e o controle de qualidade é um aspecto fundamental para que uma empresa se estabeleça nesse mercado exigente e competitivo. Frequentemente, as cervejarias artesanais terceirizam a análise química de suas cervejas. No entanto, a avaliação rápida dos parâmetros de controle de qualidade mais relevantes é importante para a melhoria da eficiência da produção. Para esta tarefa, equipamentos portáteis, como espectrômetros portáteis ou telefones celulares, podem ser benéficos quando combinados com quimiometria para desenvolver modelos de classificação supervisionada ou calibração multivariada [108].

O teor alcoólico é um dos parâmetros mais importantes no controle de qualidade da cerveja. Diferentes métodos que se baseiam na medição de alguma propriedade física ou química do etanol têm sido usados para determinar esse parâmetro em cervejas. De acordo com a *American Society of Brewing Chemists* (ASBC), existem diferentes maneiras de medir o teor de álcool nas cervejas: usando cromatografia gasosa (GC), refratometria, ensaio enzimático, instrumentos dedicados e densitometria [92], sendo o último método o mais utilizado por não demandar mão de obra especializada. Na literatura, há um crescimento no interesse de desenvolvimento de novos sensores para determinação direta de etanol em cerveja. Dentre eles, podem ser citados uma língua eletrônica potenciométrica [109], uma plataforma eletroquímica para o aprisionamento de enzimas [110], um biossensor eletroquímico nanomodificado [111], um nariz eletrônico de óxido metálico semicondutor [112] e hidrogéis de poliacrilamida sensíveis ao etanol [113]. Outra alternativa tem sido o uso de equipamentos dedicados baseados em espectroscopia no infravermelho próximo (NIR) [114], que podem ser combinados com uma etapa de destilação [115]. Além disso, o interesse no desenvolvimento de novos métodos cromatográficos, como os que utilizam GC com extração por *headspace* para quantificar álcool em cervejas, continua a aumentar na literatura de química de alimentos [104]. Enquanto alguns desses métodos são caros, laboriosos e demorados como os cromatográficos, o método densitométrico é simples, porém não seletivo e pode sofrer interferência de qualquer analito presente na matriz cervejeira que diminua a densidade do destilado. Assim, é de grande interesse para os produtores de cerveja desenvolver métodos rápidos, diretos e de baixo custo para controle de qualidade.

Vários instrumentos analíticos estão agora disponíveis em versões compactas. Como já mencionado, a recente disponibilidade desses equipamentos portáteis é particularmente útil para o controle de qualidade de alimentos e bebidas [116]. Eles permitem o desenvolvimento de métodos analíticos simples, rápidos, diretos e *in situ*. Como as amostras não precisam ser enviadas ao laboratório, uma vantagem importante vem da economia de tempo ao eliminar o atraso entre a amostragem e a disponibilização dos resultados, o que limita a frequência das análises e a otimização da produção [40]. A disseminação da portabilidade abre novas perspectivas, modificando a forma como se conduzem os negócios, abrindo novos mercados e

sendo incorporada em dispositivos de consumo e aparelhos inteligentes, com perspectiva de tornar-se nos próximos anos parte da Internet das Coisas [117], e contribuindo para a democratização da ciência analítica [74]. Especificamente, espectrômetros NIR portáteis podem ser considerados sensores químicos que apresentam sensibilidade próxima a instrumentos de bancada. Eles apresentam outras vantagens em relação aos espectrômetros de bancada, tais como tamanho reduzido, baixo custo, baixo consumo de energia e ausência de partes móveis em sua estrutura óptica [118].

Nos últimos anos, a espectroscopia NIR portátil tem sido combinada com a calibração multivariada, principalmente com modelos PLS, para o controle de qualidade de alimentos e bebidas. Alguns exemplos de aplicações incluem a quantificação de ácidos graxos em amendoim [119], determinação de parâmetros de qualidade e maturação em frutas cítricas [120], quantificação de parâmetros de qualidade de grãos de cacau [121], análise em tempo real de amostras de leite de vaca [122], determinação de parâmetros físicos e químicos na indústria de café torrado [45], determinação de acidez titulável e teor de ácido ascórbico em acerola durante o amadurecimento [44], avaliação *on site* da qualidade pós-colheita e comportamento de armazenamento refrigerado de ameixas [41] e avaliação da qualidade de damasco [123]. Apenas uma aplicação de espectroscopia NIR portátil foi encontrada discriminando cervejas de acordo com suas origens [124], mas nenhuma aplicação quantitativa foi publicada determinando qualquer parâmetro de qualidade da cerveja.

4.1 Objetivos

Utilizar um espectrofotômetro portátil na região do infravermelho próximo em combinação com calibração multivariada no desenvolvimento de um método de maior simplicidade, robustez e rapidez para determinação direta do teor de álcool em cervejas de diversos estilos. Além disso, validar o método analítico de acordo com as diretrizes brasileiras e internacionais.

4.2 Materiais e métodos

4.2.1 Equipamentos e softwares

Os espectros de infravermelho foram obtidos em um espectrofotômetro portátil MicroNIR[®] 1700, com a faixa de trabalho de 950-1650 nm, um equipamento fabricado pela Viavi Solutions (Milpitas, CA, EUA). Os dados foram processados usando o software MATLAB[®], versão 7.13 (MathWorks, EUA), juntamente com o PLS_Toolbox, versão 6.5 (Eigenvector Technologies, EUA).

4.2.2 Amostras

Noventa e duas amostras de cerveja foram adquiridas em sua maioria nas lojas locais de Belo Horizonte/MG e armazenadas em temperatura ambiente até a análise. Essas amostras de cerveja foram obtidas de uma grande variedade de cervejarias, marcas e estilos (lager, Vienna lager, pale ale, red ale, pilsen, IPA, weissbier, strong pale ale, stout, porter, etc.) com seu teor alcoólico variando de 3 a 11% (v/v). As informações (número de amostras para cada estilo de cerveja e a respectiva faixa de teor alcoólico) sobre todas as 92 amostras de cerveja analisadas estão resumidas na Tabela 6.

Tabela 6. Amostras de cerveja usadas na construção e validação do modelo quimiométrico.

Número de Amostras	Estilo	Teor de Álcool (%)
15	American Lager	3,2-6,7
1	Bock	8,1
6	Brown Ale	4,8-6,4
2	Dark Ale	6,1-10,1
8	Dark Lager	3,1-8,3
4	Dunkel	4,4-6,5
1	Japanese Rice Lager	5,5
1	Kölsch	5,6
3	Malzbier	4,2-4,6
10	Pale Ale	3,9-8,6
10	Pilsen	4,3-7,4
4	Porter	5,5-10,4
1	Radler	3,2
3	Red Ale	5,5-9,5
2	Red Lager	4,3-4,6
2	Specialty Beer	4,6-4,8
8	Stout	3,7-10,9
2	Strong Pale Ale	7,1-8,5
3	Vienna Lager	4,4-5,5
5	Weissbier	3,2-8,0
1	Wood Aged	10,9

4.2.3 Procedimentos de Preparo de amostra

Aproximadamente 50 ml de cada amostra foram coletadas e desgaseificadas em um equipamento ultrassom Equilab ULTRASONIK 28H (Madri, Espanha) por 5 minutos. Algumas gotas de 1-octanol (Neon, Suzano, Brasil) foram utilizadas para evitar espuma excessiva. Após a retirada do gás, 100 µL de cada amostra foram misturadas com 900 µL da solução de padrão interno IS (*Internal Standard*) (1-butanol 0,1% v/v em água deionizada) em um microtubo de 2 mL. Depois disso, cada tubo foi homogeneizado usando um agitador de vórtex (VELP Scientifica, Itália) durante 10 segundos a 3000 rpm. As amostras estavam prontas para serem injetadas no sistema cromatográfico, afim de se obter os valores de referência.

Para o método espectroscópico portátil proposto, 3 mL de cada amostra foram transferidos para uma pequena placa de Petri (3,5 cm de diâmetro x 1,2 cm de altura) que foi posicionada no espectrômetro portátil. Os espectros foram coletados como

uma média de 20 varreduras, de 950 a 1650 nm, com resolução de 6,25 nm. Dez espectros replicados da placa vazia também foram registrados nas mesmas condições, a fim de estimar o ruído instrumental. Durante todas essas medições, a temperatura da sala foi controlada a $25 \pm 1^\circ\text{C}$. É importante notar que o modo de refletância difusa foi usado. Este modo requer que o coeficiente de espalhamento seja mantido constante durante os experimentos, portanto, o volume da amostra e o recipiente utilizado permaneceram constantes [125]. Nos experimentos, teve-se o cuidado de garantir essas condições usando um artefato refletivo, controlando o volume da amostra, assim mantendo o comprimento do caminho ótico constante.

4.2.4 Análise cromatográfica (método de referência)

Neste trabalho, foi utilizado um sistema de cromatografia em fase gasosa Shimadzu GC-17A (Kyoto, Japão) equipado com um detector de ionizador de chama (FID) e uma coluna PoraPlot Q (10 m x 0,32 mm x 10 μm). O hidrogênio foi utilizado como gás de arraste em um fluxo de $2,2 \text{ cm}^3 \cdot \text{min}^{-1}$. A temperatura do injetor e detector foram mantidas a 250°C . A temperatura do forno foi mantida a 200°C por quatro minutos. 0,2 μL da amostra diluída foram injetados no modo *split* 1:2.

4.2.4.1 Validação do método de referência

Para avaliar o desempenho do método de referência desenvolvido para este trabalho, foi realizado um procedimento de validação analítica. Algumas figuras de mérito foram estimadas, tais como linearidade e efeito matriz, precisão, exatidão e limites teóricos de detecção e quantificação [126].

4.2.4.1.1 Linearidade e Efeito de Matriz

A linearidade do método foi avaliada de acordo com o procedimento descrito por Souza e Junqueira (2005). Uma curva de calibração de seis pontos foi construída usando pontos igualmente espaçados (1, 3, 6, 9, 12 e 15% v/v de álcool) usando água deionizada como solvente (curva de solvente) e cerveja não alcoólica (curva correspondente à matriz). Três réplicas foram analisadas para cada nível de

calibração. A presença de *outliers* em ambas as curvas foi avaliada pelo teste *jackknife* e os parâmetros de regressão foram estimados via OLS (*Ordinary Least Squares*). Os resíduos da regressão OLS proposta foram avaliados quanto à sua normalidade, homocedasticidade e independência, utilizando os testes Ryan-Joiner, Brown-Forsythe e Durbin-Watson, respectivamente. A falta de ajuste e a significância da regressão também foram testados para a regressão proposta [100-101].

Para a avaliação da presença do efeito da matriz, as inclinações da curva no solvente e da curva na matriz foram comparadas usando um teste t.

4.2.4.1.2 Precisão e recuperação

Para o estudo de precisão e recuperação, seis réplicas de três concentrações diferentes (1, 8 e 15 % v/v) foram preparadas e analisadas em três dias consecutivos por três analistas diferentes. A média da recuperação relativa foi estimada a partir das 18 réplicas de cada ponto (seis réplicas de três analistas) e os resultados de precisão, apresentados como desvio padrão relativo nas condições de repetibilidade (RSDr) e precisão intermediária (RSDR) foram estimados por ANOVA. Todas as premissas necessárias para a aplicação da ANOVA (normalidade, homoscedasticidade e independência) foram previamente testadas e confirmadas.

4.2.4.1.3 Limites de detecção e quantificação

Os limites teóricos de detecção (LD) e quantificação (LQ) foram estimados usando dez repetições de uma amostra em branco (usando água deionizada) e a inclinação da curva de calibração da matriz. O LD foi estimado em 3,3 vezes o desvio padrão das dez amostras de branco divididos pela inclinação, e o LQ foi estimado em 10 vezes o desvio padrão das dez amostras de branco divididos pela inclinação.

4.2.5 Validação analítica do método multivariado proposto

A validação analítica dos métodos de calibração multivariada ainda não é um assunto completamente estabelecido. É necessário buscar uma adaptação dos requisitos das diretrizes brasileiras [86] e internacionais [102] para estimar figuras de

mérito na calibração multivariada e assim harmonizar conceitos univariados e multivariados. Desta forma, foram estimadas as seguintes figuras de mérito: exatidão, precisão, linearidade, faixa de trabalho, sensibilidade, sensibilidade analítica, viés e relação de desempenho do desvio (RPD).

Além disso, uma região elíptica de confiança conjunta (EJCR) foi calculada para verificar a linearidade, e intervalos de confiança baseados nas incertezas dos valores previstos foram estimados para as amostras de validação. A equação de erros nas variáveis (EIV) foi utilizada para estimar os erros de predição padrão (SPE) específicos da amostra, que devem ser usados na ausência de viés [127]. A SPE para cada amostra foi estimada através da [Equação 2](#), na qual RMSEC é a raiz quadrada do erro médio quadrático da calibração, h é o *leverage* para cada amostra e $s_{(y_{ref})}$ é a estimativa do erro padrão para o método de referência:

$$\text{SPE} = \text{RMSEC} \sqrt{(1 + h) - s_{(y_{ref})}^2} \quad (2)$$

4.3 Resultados e discussão

4.3.1 Análise cromatográfica

A linearidade do método cromatográfico foi confirmada para a determinação do teor de álcool na faixa proposta (1-15 % v/v), para curvas combinadas com solvente (usando água deionizada) e com matriz (usando cerveja não alcoólica), com todas as premissas de regressão linear confirmadas para ambas as curvas, como pode ser visto na Tabela 7.

Tabela 7. Figuras de mérito para a determinação cromatográfica de álcool utilizando GC-FID. R² (coeficiente de determinação), R (parâmetro para avaliar a distribuição normal dos resíduos no teste Ryan-Joiner), t_L (parâmetro para avaliar a homoscedasticidade no teste Brown-Forsythe), d (parâmetro para avaliar a ausência de autocorrelação dos resíduos no teste de Durbin-Watson), F (valor do teste F para ANOVA), p (valor de p).

		Curva Calibração de solvente	Curva de Calibração contendo matriz
n		14	15
Equação		y=0,0832x+0,0153	y=0,0773x+0,0118
R²		0,999	0,998
Normalidade	R	0,989	0,987
	p	p>0,10	p>0,10
Homocedasticidade	t _i	0,342	0,042
	p	0,7383	0,967
Independência	d	2,182	1,539
	p	p>0,05	p>0,05
Regressão	F	5067	17135
	p	3,93x10 ⁻¹⁷	1,13x10 ⁻²¹
Falta de ajuste	F	0,306	1,249
	p	0,866	0,357

A presença de *outliers* foi testada usando Jackknife. Quatro outliers foram detectados e removidos na curva de solvente e três foram removidos da curva contendo a matriz. Ambas as remoções estavam de acordo com o limite recomendado para remoção de dados (22,2 % dos dados originais) [86]. Após a remoção, os resíduos de ambas as curvas foram avaliados como se distribuindo normalmente ($p > 0,10$), independentes ($p < 0,05$) e homoscedásticos ($p < 0,05$). A regressão também foi significativa ($p < 0,001$) e a falta de ajuste não foi significativa ($p > 0,05$) para ambas as curvas, o que confirma a linearidade da curva de solvente e da curva correspondente à matriz. Quando a inclinação da curva no solvente foi comparada com a curva contendo a matriz, é possível observar uma diferença significativa ($p < 0,05$). Como o efeito da matriz foi observado e confirmado, as etapas de validação adicionais foram realizadas usando curvas contendo a matriz (Figura 14).

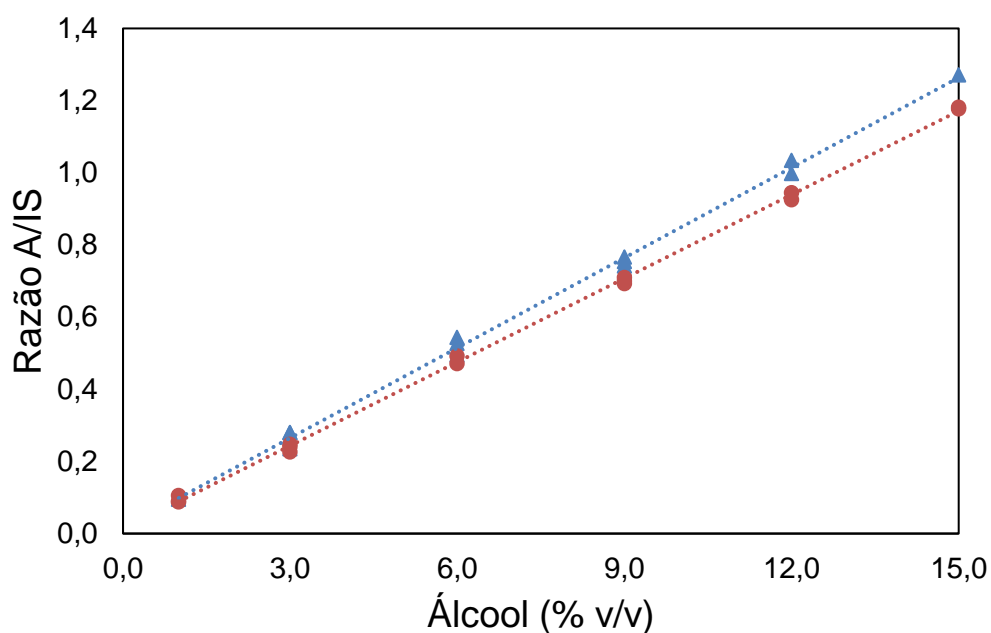


Figura 14. Curvas de calibração, sendo a curva contendo a matriz está representada por círculos vermelhos, já a curva de calibração no solvente está representada por triângulos azuis.

A recuperação e a precisão foram estimadas através da repetibilidade e da precisão intermediária. A recuperação foi estimada em três níveis diferentes (1, 8 e 15 % v/v) e a recuperação obtida variou de 98,4 a 101,4 %, estando de acordo com os limites definidos pelos guias da AOAC Internacional [126]. Os níveis de precisão também estavam de acordo com o guia de validação mencionado, variando de 1,9 a 2,2%. O LD teórico e LQ também foram estimados (0,06 e 0,19 % v/v, respectivamente) e ambos foram inferiores à primeira concentração da curva de calibração proposta.

O método proposto para determinação de álcool pode ser considerado elegível para determinar os valores de referência do teor de álcool nas cervejas que serão usadas para construir o modelo de calibração multivariada, pois todas as figuras de mérito avaliadas estão de acordo com o guia AOAC Internacional [126]. A facilidade do preparo de amostras (uma simples diluição com a solução IS) e o curto tempo de uma execução cromatográfica isotérmica (4 minutos) também foram compatíveis com o grande número de amostras que precisou ser analisado.

4.3.2 Desenvolvimento do modelo PLS

Os espectros obtidos com o MicroNIR[®] das 92 amostras de cervejas são mostrados na Figura 15. Ao observar esses espectros, foi possível atribuir as bandas mais intensas e largas, centradas em 1150 e 1300 nm, às vibrações das moléculas de álcool, correspondendo ao segundo sobretom de estiramento das ligações de CH₃ e CH₂ e ao primeiro sobretom de estiramento da ligação de ROH da água e etanol, respectivamente [128].

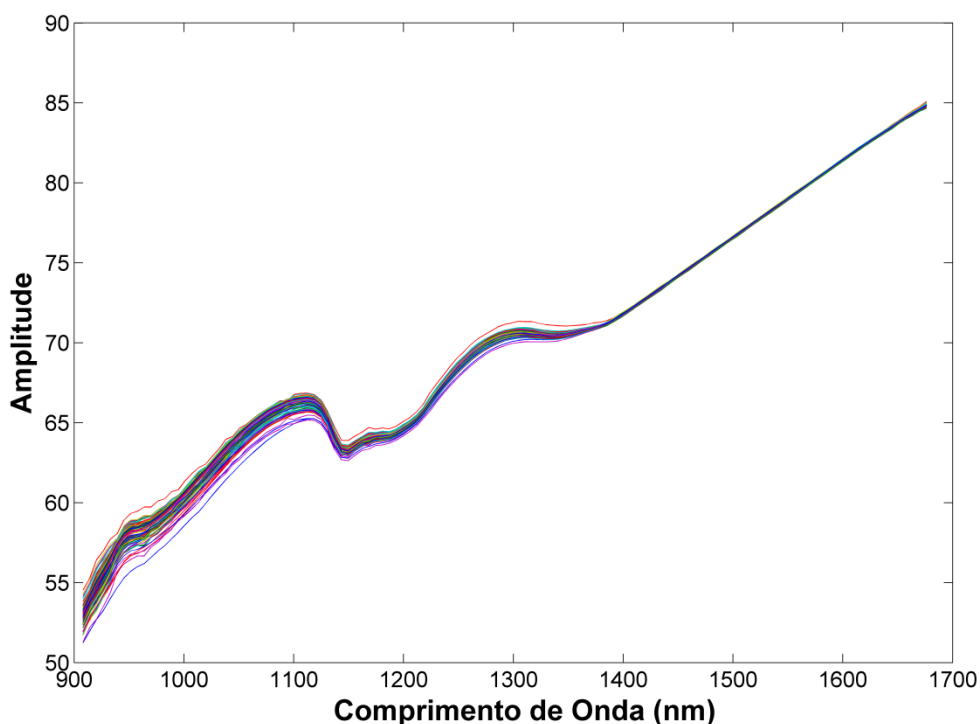


Figura 15. Espectros de NIR das 92 amostras de cervejas.

Alguns métodos estão disponíveis para a seleção eficiente de amostras de calibração, dentre eles o algoritmo Kennard-Stone, que foi usado para selecionar 61 (dois terços das amostras) para o conjunto de calibração, sendo as 31 amostras restantes utilizadas no conjunto de validação. Através de uma varredura uniforme dos dados espectrais, o algoritmo garantiu a presença das amostras mais representativas na calibração [95]. Na Figura 15, é possível observar a presença de desvios não lineares de linha-base, típicos em medições de refletância difusa de amostras sólidas, suspensões e amostras líquidas turvas, e causados pelo espalhamento de luz multiplicativo. Assim, a variação normal padrão (SNV) e o alisamento Savitzky-Golay (polinômio de primeira ordem e janela de 5 pontos), que visa aumentar a razão

sinal/ruído, foram utilizados como o pré-processamento dos dados, os quais também foram centrados na média [129].

O número de variáveis latentes (VL) foi selecionado de acordo com o menor erro quadrático médio da validação cruzada (RMSECV), estimado usando subconjuntos aleatórios (7 divisões de dados e 20 iterações). O melhor modelo foi selecionado inicialmente com seis VLS, e foi otimizado pela detecção de *outliers*. Foram consideradas amostras outliers as que apresentaram altos resíduos em X , altos resíduos em Y ou altos valores de *leverage*, com nível de confiança de 95% [36,77]. O número de *outliers* detectados não excedeu o limite recomendado de dois nonos, o que representa 22,2% do conjunto de calibração e do conjunto de validação separadamente [86, 100, 190]. Do modelo, 13 das 92 amostras originais foram removidas como *outliers*, sendo 8 no conjunto de calibração (13%) e 5 no conjunto de validação (16%). Após a remoção de *outliers*, o modelo foi reconstruído. Os resultados para a otimização do modelo, incluindo RMSEC, RMSEP e RPD, são mostrados na Tabela 8.

Tabela 8. Resultados para a otimização do modelo PLS desenvolvido através da detecção e remoção de *outliers*. RMSEC (Raiz quadrada do Erro Médio Quadrático da Calibração), RMSEP (Raiz quadrada do Erro Médio Quadrático da Previsão) e RPD (Relação de Desempenho do Desvio).

	Nº de amostras de calibração	Nº de amostras de validação	Nº de variáveis latentes (VL)	RMSEC	RMSEP	RMSECV	RPD validação	RPD calibração
Antes	61	31	6	0,8	0,9	1,0	2,1	1,8
Depois	53	26	6	0,5	0,6	0,7	2,8	2,7

RMSEC e RMSEP são parâmetros que descrevem a veracidade do método. Esses valores foram melhorados após a detecção de *outlier*, assim como também os valores de RPD. O RPD é um parâmetro simples para avaliação estatística de um modelo de calibração multivariada usando dados obtidos por espectroscopia no infravermelho próximo (NIR) [98]. Se este valor estiver acima de 2,4, é possível considerar que o método possui uma boa capacidade preditiva [36,99]. Como pode ser visto no gráfico dos valores previstos *versus* medidos, o modelo apresentou boa correlação e nenhuma irregularidade na estrutura de dados (Figura 16).

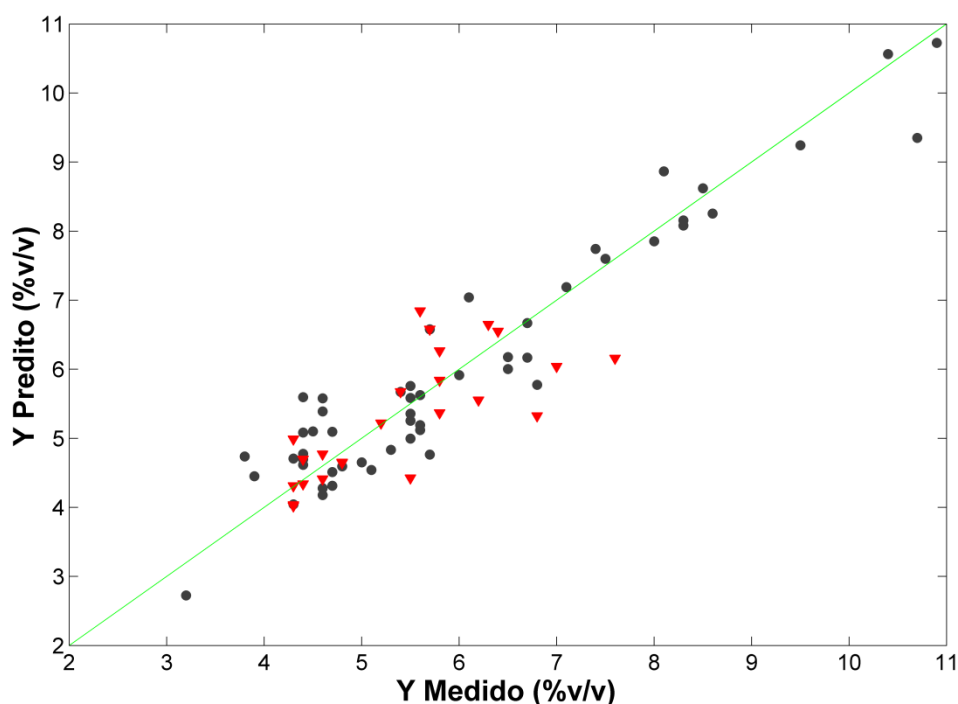


Figura 16. Gráfico de valores de referência versus valores previstos para o modelo PLS. Símbolos: círculos pretos para amostras de calibração, triângulos vermelhos para amostras de validação.

Os coeficientes de regressão estimados e os VIP escores do modelo PLS estão apresentadas na [Figura 17](#). Observando o gráfico da [Figura 17\(b\)](#), foi possível identificar as regiões espectrais que mais contribuem para prever o teor de álcool. Como era esperado, de acordo com a [Figura 17\(a\)](#), o coeficiente mais positivo está em cerca de 1317 nm, correspondendo ao primeiro sobretom do estiramento da ligação O-H do etanol e da molécula da água [130]. Por outro lado, os coeficientes mais negativos apresentam um pico centralizado em torno de 970 nm, que pode ser identificado como uma banda de combinação dos modos de estiramento simétrico e assimétrico da água e o terceiro sobretom de R-OH. O pico em 1187 nm, que apresenta o coeficiente negativo mais intenso, corresponde à banda de combinação do modo de vibração da água e do primeiro sobretom do estiramento da ligação O-H da molécula da água, além do segundo sobretom do estiramento da ligação de C-H do etanol.

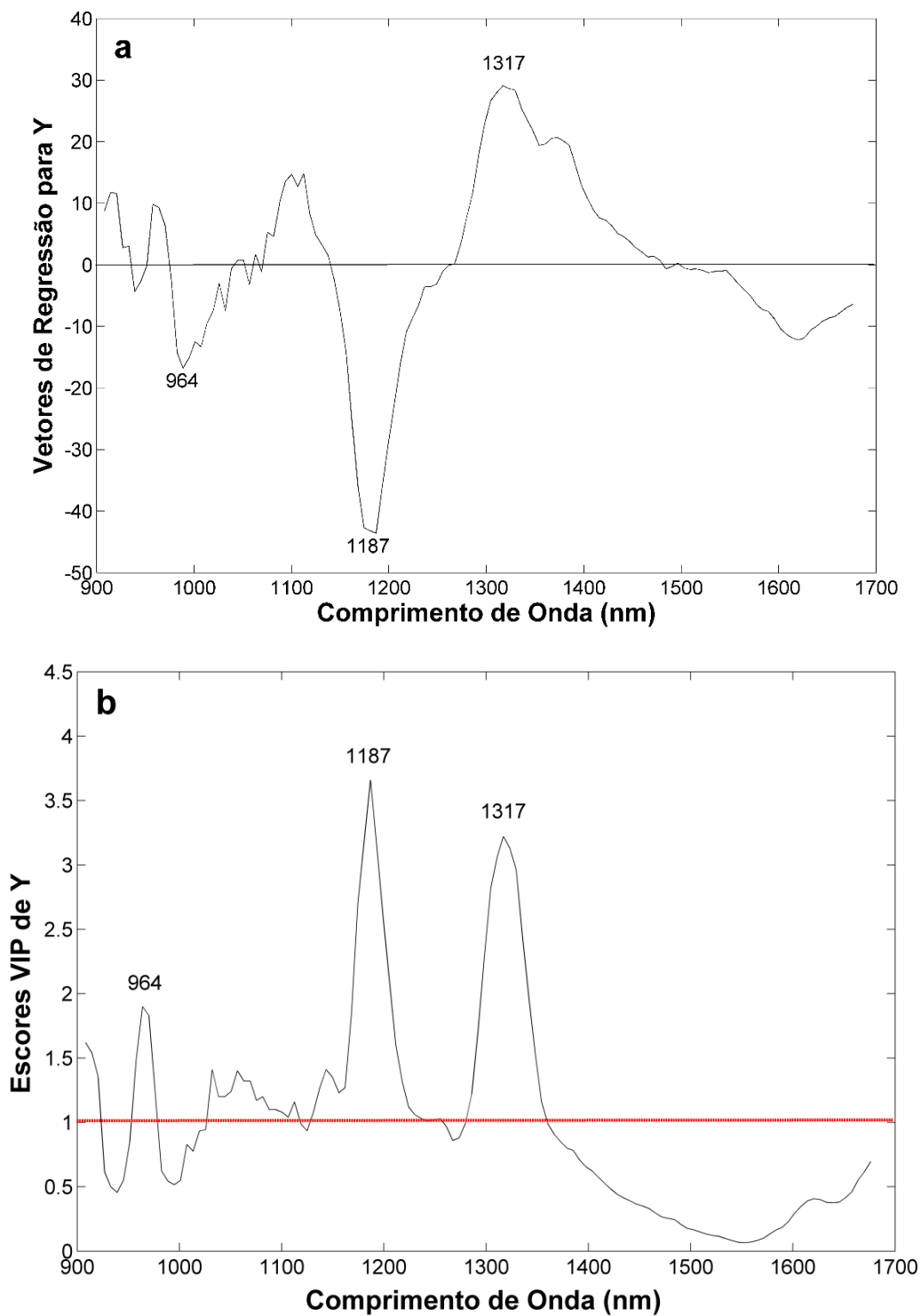


Figura 17. (a) Vetores de Regressão (b) VIP Escores para o modelo PLS.

Outras figuras de mérito (FOM) que foram estimadas para validar o método estão apresentadas na Tabela 9. O ajuste pode ser avaliado com o auxílio do gráfico de referência *versus* valores previstos, mostrados na Figura 16 e seus parâmetros presentes na Tabela 9. O coeficiente de correlação de 0,9355 foi obtido para o modelo.

Além disso, a independência dos resíduos foi certificada pelo valor do teste de Durbin-Watson (Tabela 9), que estava dentro dos limites da faixa de aceitação, entre 1,50 e 2,50 [101]. Considerando apenas as amostras de validação, os pequenos valores de viés estimados para o modelo corroboram a ausência de erros sistemáticos no modelo. Os testes t de Student corroboraram que não há viés no modelo no nível de confiança de 95%.

Tabela 9. Resultados para FOM (precisão e linearidade) do modelo PLS.

Figuras de Mérito	Parâmetros	Valores
Precisão	Repetibilidade RSD (%)	1,45
		1,14
		1,88
	Precisão Intermediária RSD (%)	2,47
		3,48
		3,92
Linearidade	Teste Durbin-Watson	2,21
	Inclinação	0,90±0,04
	Intercepto	0,56±0,23
	Coefficiente de Correlação	0,9355

Para melhor avaliar o erro do método, foi estimada uma região de confiança elíptica. Calculando uma EJCR, que é uma região elíptica de confiança conjunta usada para estimar a linearidade e a ausência de erros sistemáticos nos modelos, o intercepto e a inclinação estimadas foram comparadas com seus valores ideais de 0 e 1. Como pode ser visto na Figura 18, o ponto ($a = 0$, $b = 1$) está dentro do EJCR. Portanto, o EJCR para o modelo PLS mostrou que não houve diferença significativa entre os valores de referência $\%(v/v)$ e os valores previstos $\%(v/v)$, e que não houve evidência de viés no nível de confiança de 95%.

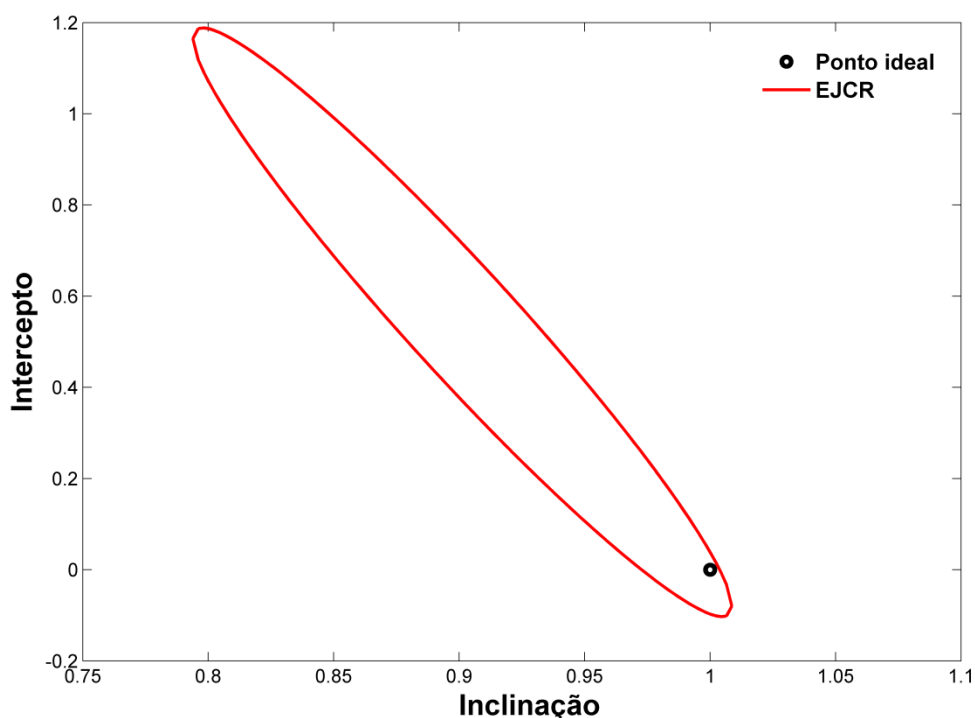


Figura 18. Gráfico da região de confiança elíptica para a inclinação e intercepto da curva dos valores previstos versus de referência para o modelo PLS. Símbolos: círculo preto para o ponto ideal e elipse vermelha para EJCR

4.4 Conclusão

Um equipamento portátil MicroNIR[®] foi usado para desenvolver e validar um método simples e rápido para determinar diretamente o teor de álcool em cervejas de diversos estilos, marcas e fabricantes. Esse novo método apresentou várias vantagens em relação ao método cromatógrafo de referência, como a facilidade de operação, mínima necessidade de pré-tratamento da amostra (desgaseificação), maior rapidez e menor custo. Este método foi validado de acordo com as diretrizes internacionais, sendo considerado linear, preciso e adequado ao uso como metodologia para a determinação de álcool em cervejas. Um artigo contemplando os resultados deste trabalho foi publicado na revista *Food Analytical Methods* (Apêndice B).

**CAPÍTULO 5: MODELO DE CALIBRAÇÃO MULTIVARIADA MULTIPRODUTO
PARA A DETERMINAÇÃO DO TEOR DE ETANOL EM BEBIDAS ALCOÓLICAS
FERMENTADAS USANDO UM ESPECTROFOTÔMETRO NIR PORTÁTIL**

5 MODELO DE CALIBRAÇÃO MULTIVARIADA MULTIPRODUTO PARA A DETERMINAÇÃO DO TEOR DE ETANOL EM BEBIDAS ALCOÓLICAS FERMENTADAS USANDO UM ESPECTROFOTÔMETRO NIR PORTÁTIL

5.1 Introdução

A expansão do mercado global de bebidas alcoólicas tem aumentado a demanda por seu controle de qualidade. Um dos principais parâmetros determinados no seu controle de qualidade físico-químico é o teor alcoólico da bebida, cuja medida de concentração representa a identidade e caracteriza o produto. Essa demanda analítica está ligada à garantia de qualidade na cadeia de abastecimento de alimentos, tributação e declaração do teor de etanol no rótulo [131]. Do ponto de vista forense, também é importante verificar se uma bebida foi adulterada ou falsificada [132]. Alguns países classificam as bebidas alcoólicas de acordo com seu teor de etanol. A categoria legal de uma bebida pode afetar o tipo de licença necessária para fabricar, distribuir e vender o produto; também pode influenciar na sua tributação [133-134]. Muitos países têm altos impostos especiais de consumo para bebidas alcoólicas fermentadas, e a declaração incorreta do teor alcoólico é uma das categorias de fraude encontradas nos regulamentos oficiais [135-136]. As bebidas fermentadas costumam ter até 16% de álcool por volume, enquanto as bebidas destiladas têm teores muito maiores, entre 30 e 60% [137]. A cerveja e o vinho estão entre as bebidas mais consumidas no mundo, sendo as duas bebidas alcoólicas mais consumidas na Europa [138]. No presente estudo, foram analisadas outras duas bebidas fermentadas menos consumidas, sidra e hidromel, cujas matérias-primas primárias são suco de maçã e mel, respectivamente.

O teor de etanol foi um dos primeiros parâmetros para os quais métodos analíticos quantitativos foram desenvolvidos no controle de qualidade de bebidas [139]. O método clássico de quantificação de etanol em bebidas alcoólicas baseia-se na determinação da densidade da bebida após uma etapa de destilação [126]. Esta é uma metodologia não seletiva, trabalhosa e demorada. Considerando a necessidade

de inspeção de controle de qualidade na cadeia de suprimentos de alimentos, diversos métodos analíticos têm sido desenvolvidos para a determinação de etanol em diferentes bebidas alcoólicas, tanto fermentadas quanto destiladas. Os artigos que desenvolvem métodos para quantificar o etanol em mais de um tipo de bebida têm sido baseados principalmente em técnicas cromatográficas, como cromatografia gasosa acoplada à espectrometria de massa (GC-MS) [139] ou detecção de ionização de chama (GC-FID) [104], e cromatografia líquida de alta eficiência (HPLC) [140]. Além de normalmente serem trabalhosos e demorados, os métodos cromatográficos vão contra os princípios da química verde, que busca evitar o uso de solventes tóxicos e minimizar a geração de resíduos químicos. Assim, muitos métodos alternativos têm sido desenvolvidos, principalmente nos últimos anos, buscando superar essas limitações. Esses métodos têm sido baseados em detecção amperométrica [141], biossensor microbiano empregando um sistema de duplo mediador eucarioto [142], análise direta em espectrometria de massa em tempo real (DART-MS) [132], uma reação colorimétrica de etanol com dicromato em um sistema de impressão 3D [143], espectros UV-visíveis adquiridos em um sistema de fluxo com difusão de gás em linha [131], ondas químicas visíveis registradas com um *smartphone* durante uma reação de Belousov-Zhabotinsky (BZ) [144], um quimio-ensaio cronométrico utilizando a reação oscilante BZ [145] e imagens digitais RGB obtidas com um escâner de mesa [146]. A maioria desses métodos consome reagentes [131,143-146], em alguns casos reagentes tóxicos [143]. Enquanto cerveja e vinho estão entre as bebidas mais analisadas em todos os artigos citados anteriormente, estudos envolvendo sidra e hidromel são escassos na literatura. Apenas um artigo determinou etanol especificamente em sidra empregando espectroscopia de ^1H RMN [147].

Nos últimos anos, as técnicas espectroscópicas vibracionais têm sido cada vez mais alternativas para o desenvolvimento de métodos analíticos mais simples, diretos, rápidos e verdes. Como essas técnicas propiciam a obtenção de sinais analíticos sobrepostos, sua combinação com ferramentas quimiométricas é quase obrigatória. Recentemente, as vantagens dessas técnicas foram aprimoradas devido à miniaturização dos espectrofotômetros. De fato, a portabilidade é uma tendência importante na química analítica em geral, não apenas na espectroscopia, proporcionando vantagens como economia de tempo, baixo custo de operação,

pequeno tamanho e peso dos equipamentos, menor consumo de energia e análise ecologicamente correta [117,148-151]. Um pré-tratamento mínimo ou nenhum da amostra é necessário, pequenos volumes de amostra são precisos, não há necessidade de consumir reagentes ou solventes e nenhum resíduo químico é gerado. Há vantagens evidentes em eliminar a necessidade de envio de amostras para o laboratório, pois o equipamento pode ser deslocado para o campo [148]. Particularmente, os dispositivos portáteis podem fornecer uma triagem rápida da autenticidade da bebida diretamente no ponto de necessidade [149]. Dentre as técnicas vibracionais, os dispositivos portáteis de infravermelho próximo (NIR) apresentam as vantagens de serem em sua maioria mais baratos, mais leves e de dimensões menores do que os equipamentos portáteis Raman e no infravermelho médio [150-151]. Devido a essas características, a espectroscopia NIR portátil está se tornando uma das plataformas analíticas mais desenvolvidas e empregadas comercialmente para determinar a autenticidade de alimentos, permitindo a integração em sistemas de rastreabilidade digital [150]. Em artigo anterior e recente, nosso grupo de pesquisa empregou espectroscopia NIR portátil e calibração multivariada para determinar o teor alcoólico de apenas uma matriz de bebida, a cerveja [152]. No presente estudo, desenvolvemos um modelo quantitativo multiproduto analisando quatro matrizes de bebidas diferentes.

Os modelos multiprodutos são obtidos a partir de dados espectrais para diferentes tipos de amostras ou matrizes, que são combinados em um único modelo de calibração multivariado [153-154]. Eles também são chamados de modelos de calibração globais ou robustos, embora esses dois termos também sejam usados para outras situações, e podem ser considerados muito convenientes e econômicos. A manutenção de vários modelos de calibração de um produto é trabalhosa e demorada. Passar da calibração de um produto para a calibração de vários produtos pode resultar em uma menor exatidão do modelo, embora limitada a um nível aceitável, quando um grande número de amostras para cada matriz estiver disponível. No entanto, quando os produtos são relativamente semelhantes e, principalmente, se o número de amostras de cada produto for pequeno, dificultando a modelagem local, os modelos multiprodutos tendem a apresentar melhor desempenho e menores erros de previsão [153]. Se a faixa analítica da variável dependente não for muito grande, o modelo

tradicional de calibração multivariada linear de mínimos quadrados (PLS) tende a fornecer boas previsões. O uso de dispositivos portáteis também facilita o desenvolvimento desse tipo de modelo, pois o mesmo espectrofotômetro pode ser transportado para diferentes locais de análise. Modelos de calibração multivariada multiprodutos foram desenvolvidos em combinação com dados NIR para vários tipos de matrizes, como diferentes alimentos [153], vários tipos de materiais para nutrição de plantas [154], várias raízes e tubérculos pulverizados [155], biomassa de diferentes origens [156] néctares de frutas industrializados e sucos de soja obtidos de diferentes frutas [157], uvas para vinho (*Vitis vinifera* L.) durante a maturação na videira e na colheita de 25 variedades de uvas brancas e tintas [158] e cinco diferentes produtos vacinais [159].

5.2 Objetivos

O objetivo deste trabalho foi desenvolver e validar um método multiproduto de calibração multivariada baseado na combinação de espectroscopia NIR portátil e PLS para determinação direta do teor de etanol em quatro bebidas alcoólicas fermentadas distintas. Os valores de referência foram obtidos com um método baseado em GC-FID. Quanto à robustez do método, uma variedade relativa de cervejas, sidras, hidroméis e vinhos foi incorporada ao modelo, correspondendo a uma faixa analítica de 4,3 a 15,3% (v/v). O método proposto também foi submetido a uma validação analítica multivariada de acordo com as diretrizes brasileiras e internacionais [36,77].

5.3 Materiais e Métodos

5.3.1 Amostras

Um total de 153 amostras de bebidas alcoólicas foram adquiridas em diferentes mercados locais (100% das amostras de cerveja e vinho e cerca de 10% das amostras de sidra e hidromel) ou preparadas artesanalmente (cerca de 90% de amostras de sidra e hidromel) para este estudo. As amostras utilizadas para a construção do modelo multiproduto foram cerveja (trinta e nove amostras), sidra (quarenta amostras), hidromel (trinta e seis amostras) e vinho (trinta e oito amostras). Essas amostras foram

obtidas ou preparadas para fornecer uma ampla variância, com seu teor de etanol variando entre 4 e 16% (v/v).

5.3.2 Instrumentos e software

Os espectros NIR foram adquiridos usando um espectrofotômetro portátil MicroNIR® 1700 da Viavi Solutions (Milpitas, CA, EUA), com um detector de matriz de fotodiodos InGaAs e tecnologia de filtro variável linear (LVF) como elemento dispersante. Este é um dos menores dispositivos NIR, pesando apenas 64 g, e com um dos melhores desempenhos do mercado [150]. Cada espectro foi obtido no modo de refletância difusa, com vinte varreduras na faixa de comprimento de onda de 908-1676 nm, com resolução de 6,25 nm. Os dados foram processados por meio do software MATLAB®, versão 7.13 (MathWorks, Natick, EUA), juntamente com o PLS Toolbox, versão 6.5 (Eigenvector Technologies, Manson, EUA). Um sistema de cromatografia gasosa Shimadzu GC-17A (Kyoto, Japão) equipado com um detector de ionização de chama (FID) e uma coluna PoraPlot Q (10 m × 0,32 mm × 10 µm) foram usados para obter os valores de referência.

5.3.3 Procedimentos

Aproximadamente 5 mL de cada amostra de cerveja e sidra foram degaseificadas usando um banho ultrassônico (Equilab ULTRASONIK 28H, Madrid, Espanha) por 5 minutos. As amostras de hidromel e vinho não precisaram ser degaseificadas. Após a degaseificação, aproximadamente 1,0 mL de cada amostra de bebida foi utilizado para o método de referência GC-FID e os 4,0 mL restantes para o método proposto. Na sequência, 3,0 mL de cada amostra foram transferidos para uma placa de Petri de poliestireno (1,7 cm de raio x 0,9 cm de altura). A placa foi colocada acima do espectrofotômetro MicroNIR® para registro dos espectros. Durante todas as medições, a temperatura do laboratório foi mantida em $25 \pm 1^\circ\text{C}$. Precauções para manter o coeficiente de espalhamento constante durante os experimentos foram tomadas, como o controle do volume da amostra e o uso de artefato refletivo.

5.3.4 Método de Referência (GC-FID)

Foi desenvolvido um método GC-FID para obtenção de valores de referência para as amostras, que foram utilizados como variáveis dependentes para a construção do modelo PLS. Para este método cromatográfico, 100 μL de cada amostra foram misturados com 900 μL da solução de padrão interno (PI) (1-butanol 1,0% v/v) em um tubo de microcentrífuga de 2 mL. O gás de arraste utilizado foi hidrogênio com um fluxo de $2,2 \text{ cm}^3 \text{ min}^{-1}$. A porta de injeção e o detector foram mantidos a 250°C e a temperatura do forno foi mantida a 200°C por 4 minutos. Uma alíquota de 0,2 μL da amostra diluída foi injetada com *split* de 1:2.

5.3.5 Modelo de calibração multivariada e validação analítica

Um modelo de calibração multivariada de multiproduto foi construído utilizando PLS. As amostras foram divididas em dois terços e um terço para os conjuntos de calibração e validação, respectivamente, aplicando o algoritmo Kennard-Stone [95]. Este algoritmo foi aplicado separadamente às amostras de cada matriz/produto, garantindo a representatividade de cada tipo de bebida nos conjuntos de calibração e validação. O algoritmo Kennard-Stone foi escolhido ao invés da amostragem aleatória pois garante a seleção de amostras representativas e homoganeamente distribuídas em toda a faixa analítica do espaço multivariado, de forma reprodutível com base em um critério sistemático [160].

A validação analítica realizada buscou adequar as exigências das diretrizes brasileiras e internacionais [86,102] ao estado da arte da estimativa das figuras de mérito em calibração multivariada [36,77,84]. Figuras de mérito (FOM) apropriadas foram estimadas para o método multiproduto proposto de acordo com os requisitos regulatórios, como veracidade, precisão, linearidade, faixa de trabalho, sensibilidade analítica, viés e RPD. Amostras com altos *leverages*, grandes resíduos no bloco X (dados espectrais) ou grandes resíduos no bloco Y (valores de concentração) com nível de confiança de 95% foram detectadas como *outliers* [36,77]. A incerteza dos valores previstos foi determinada para estimar os intervalos de confiança calculando os erros de previsão padrão específicos da amostra (SPE). A SPE foi calculada com base na equação de erro nas variáveis (EIV) [127]. Esta equação considera a

incerteza do método de referência e só deve ser utilizada na ausência de viés. Além disso, uma região de confiança conjunta elíptica (EJCR) foi calculada para verificar a linearidade e a ausência de erros sistemáticos no modelo.

5.4 Resultados e Discussão

Diretrizes específicas recomendam a utilização de no mínimo $6(nVL+1)$ e $4(nVL)$ amostras nos conjuntos de calibração e validação, respectivamente, para a construção de modelos quantitativos multivariados infravermelhos, onde nVL significa o número de variáveis latentes [84]. Levando em consideração que o número de amostras disponíveis para a construção de modelos de um produto era muito pequeno, decidiu-se construir um modelo de calibração multivariada multiproduto.

Conforme já mencionado na Seção 5.3.5, o algoritmo Kennard-Stone foi empregado para selecionar amostras para conjuntos de dados de calibração e validação, que foram compostos por 102 e 51 amostras, respectivamente. A proporcionalidade entre cada tipo de bebida foi mantida nesses dois conjuntos de dados. Os dados NIR foram pré-processados para melhorar a relação sinal-ruído e para evitar desvios causados por espalhamento multiplicativo de luz, que são típicos de espectros de refletância difusa [129]. Os espectros foram sequencialmente pré-processados por alisamento de Savitzky-Golay (largura de filtro de 5 pontos e ajuste polinomial de segunda ordem), por variação normal padrão (SNV) e centrados na média. A Figura 19 mostra todos os espectros pré-processados usados no desenvolvimento do modelo (antes de centrar na média). Ao observar esses espectros, os sinais mais proeminentes são as bandas de absorção centradas em cerca de 1200 nm e 1400 nm. Essas duas bandas espectrais podem ser atribuídas ao segundo sobretom do estiramento C-H do etanol e ao primeiro sobretom das vibrações de estiramento O-H do etanol e da água, respectivamente [161].

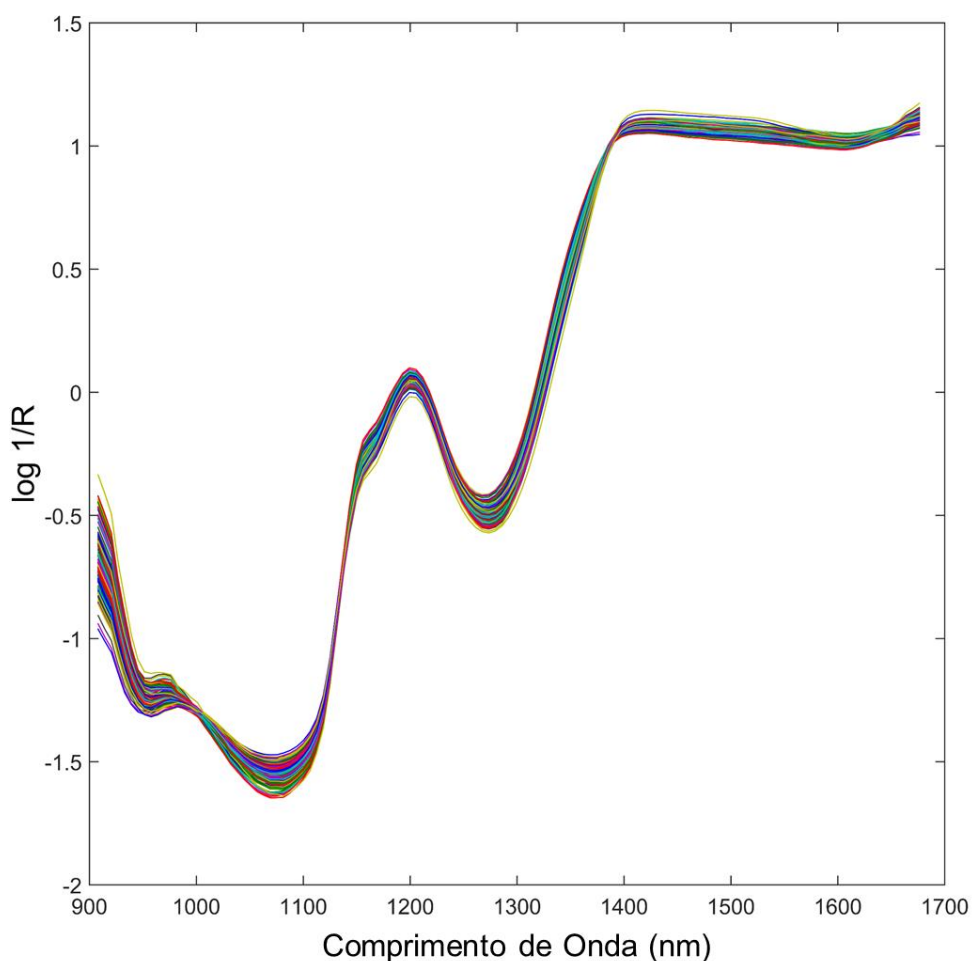


Figura 19. Espectros NIR pré-processados por alisamento Savitzky-Golay e variação normal padrão (SNV) de todas as 153 amostras de bebidas analisadas.

A validação cruzada por subconjuntos aleatórios (9 divisões e 20 iterações) foi aplicada para selecionar o melhor número de VL. Foi escolhido o número de VL que apresentou o menor erro quadrático médio da validação cruzada (RMSECV). O melhor modelo foi selecionado com 8 VL após otimização por detecção de outliers. Este procedimento pode melhorar a qualidade do modelo, uma vez que os outliers são definidos como observações que mostram algum tipo de desvio da maioria dos dados. O número de outliers não deve ultrapassar o limite de 22,2% dos conjuntos de treinamento e teste para cada bebida [97,100,190]. Dos dados originais, 17 das 153 amostras foram removidas como discrepantes (11,1%), dez do conjunto de calibração (9,8%) e sete do conjunto de validação (13,7%). A Tabela 10 especifica quantas amostras de cada bebida foram retiradas do modelo. Após a exclusão dos *outliers*, o modelo foi reconstruído e respondeu por 99,82% da variância no bloco **X** e 93,09% no bloco **Y**. Os resultados para a otimização do modelo, incluindo FOM como RMSEC,

RMSEP, e RPD para conjuntos de calibração e validação, são mostrados na Tabela 10.

Tabela 10. Resultados para a otimização do modelo PLS desenvolvido através da detecção e remoção de *outliers*. RMSEC (Raiz quadrada do Erro Médio Quadrático da Calibração), RMSEP (Raiz quadrada do Erro Médio Quadrático da Previsão) e RPD (Relação de Desempenho do Desvio).

		N° de amostras de calibração	N° de amostras de validação	N° de variáveis latentes (VL)	RMSEC (%)	RMSEP (%)	RMSECV (%)	RPD calibração	RPD validação
Antes	Cerveja	26	13	8	1,4	1,8	1,7	2,1	2,0
	Sidra	27	13						
	Hidromel	24	12						
	Vinho	25	13						
Depois	Cerveja	25	13	8	0,8	0,9	1,0	3,5	4,1
	Sidra	25	11						
	Hidromel	21	10						
	Vinho	21	10						

RMSEC e RMSEP são FOM relacionados à exatidão do método de calibração multivariada. Uma melhora desses parâmetros foi obtida após a detecção dos *outliers*, resultando em um RMSEC de 0,8% e um RMSEP de 0,9% v/v. Uma FOM apropriada para avaliar a capacidade preditiva dos modelos NIR em termos absolutos é o RPD. O conceito de RPD foi introduzido pela primeira vez em 1993 e corresponde à razão da variação natural nas amostras para o tamanho dos prováveis erros de previsão [98]. O modelo PLS desenvolvido apresentou RPD de 3,5 e 4,1 para os conjuntos de calibração e validação, respectivamente. Esses resultados são satisfatórios, pois os modelos PLS com valores de RPD superiores a 2,4 são considerados de boa capacidade preditiva [36, 99]. O ajuste pode ser avaliado através dos gráficos de regressão dos valores de referência *versus* valores previstos, que são mostrados na Figura 20. Ambas as figuras possuem o mesmo ajuste de retas e representam o mesmo modelo quimiométrico, o que mudou foi a indicação da origem das amostras. Na Figura 20a, as amostras estão representadas como de calibração ou validação. Já na Figura 20b, as amostras estão representadas em função do tipo de bebida. É possível observar uma diferença nas faixas de teor alcoólico entre as matrizes estudadas. Vinho e hidromel têm as maiores faixas de teor alcoólico, variando entre

8% e 16%. A cerveja tem a menor faixa de teor, variando entre 4% e 9%. Por fim, a sidra tem uma faixa de teor alcoólico intermediário, entre 7% a 14%.

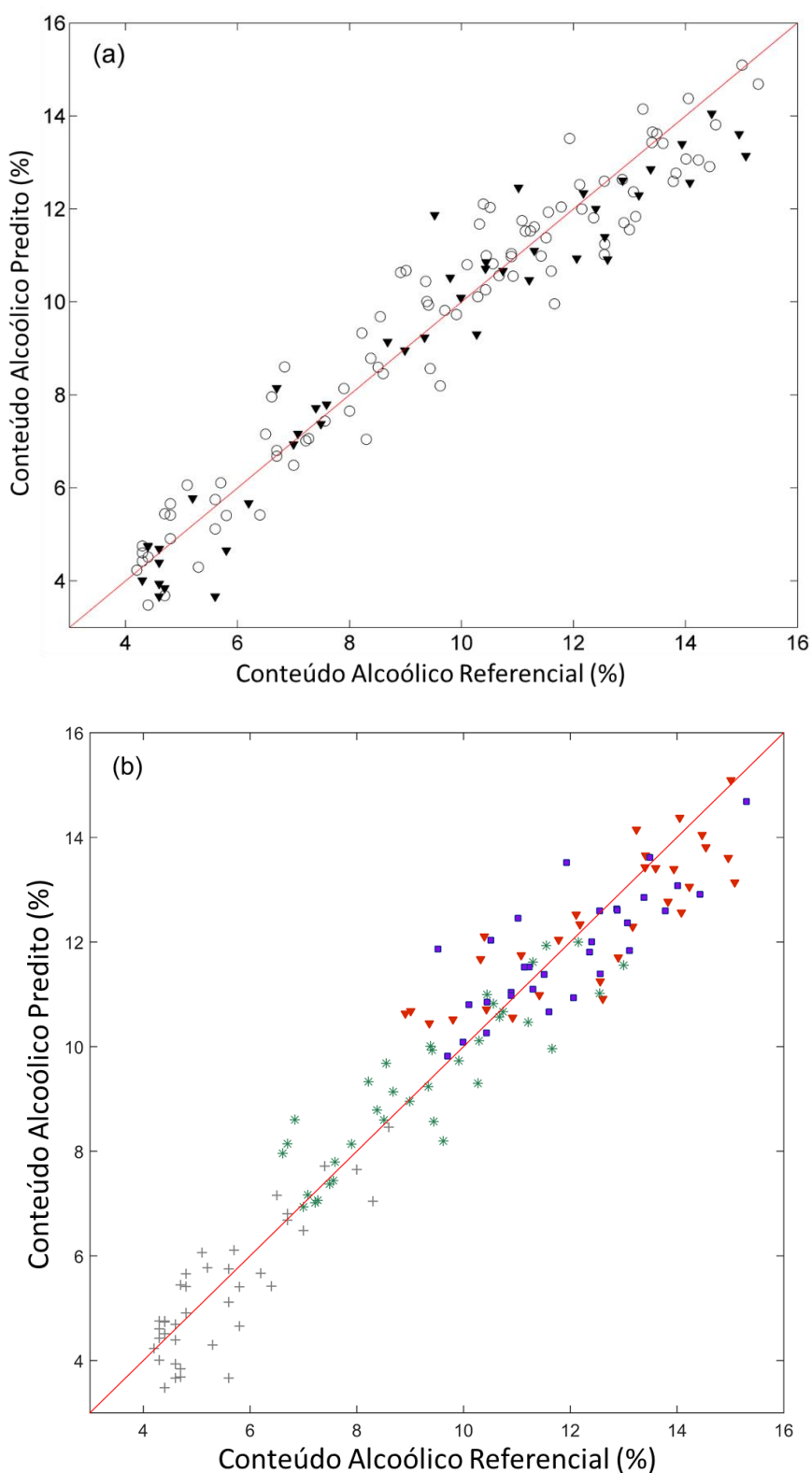


Figura 20. Gráfico de teor alcoólico de referência versus valores previstos para o modelo PLS desenvolvido. (a) Representação em função de amostras de calibração e validação. Círculos vazios

indicam amostras de calibração e triângulos cheios indicam amostras de validação. **(b)** Representação em função do tipo de bebida. Cruzes cinzas representam amostras de cerveja, amostras de sidra representadas por asteriscos verdes, amostras de hidromel por quadrados azuis e amostras de vinho por triângulos vermelhos.

Os vetores informativos mais importantes extraídos do modelo PLS desenvolvido foram usados para interpretação espectral. Os coeficientes de regressão e a importância dos vetores na projeção dos escores (VIP) são mostrados na Figura 21. Ao avaliar conjuntamente esses gráficos, é possível identificar as variáveis ou bandas que mais contribuem para as previsões do modelo. Os escores VIP (Figura 21a) representam uma medida do grau de importância de cada variável na projeção utilizada por um determinado modelo em valores absolutos. Complementarmente, os vetores de regressão (Figura 21b) discriminam a direção da contribuição de cada variável. Os coeficientes mais positivos contribuem para a predição do analito alvo, enquanto os mais negativos podem estar associados aos interferentes. Variáveis com escores VIP acima de um limite de 1,0 foram consideradas significativas para as previsões do modelo [54]. Também é importante notar que a espectroscopia NIR envolve interação entre matéria e luz e, embora as bandas espectrais possam ser atribuídas a modos vibracionais de grupos funcionais específicos, elas não são totalmente seletivas.

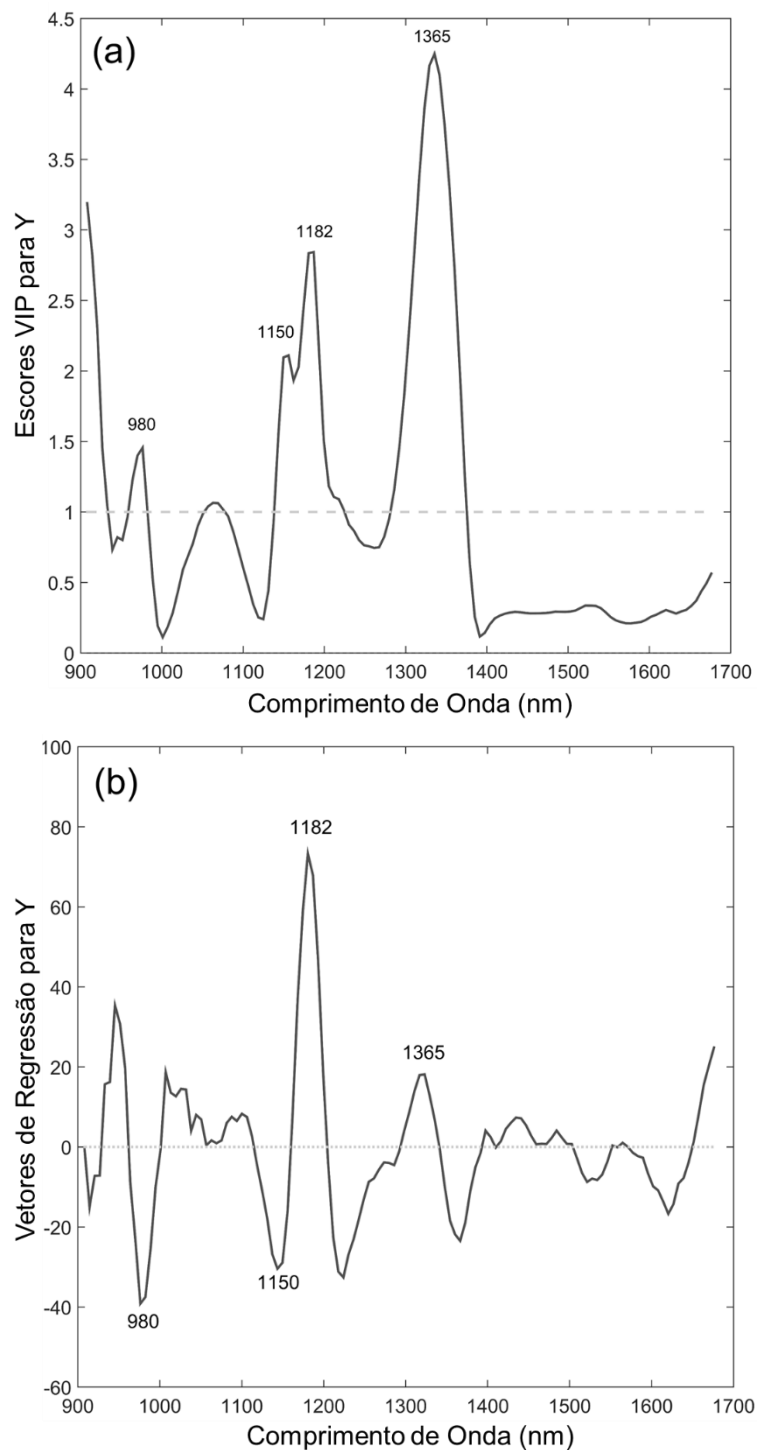


Figura 21. Vetores informativos para o modelo PLS. **(a)** EscORES VIP. A linha tracejada horizontal indica o limite de 1,0, acima do qual as variáveis foram consideradas como contribuindo significativamente para o modelo PLS. **(b)** Vetor de regressão.

De acordo com a Figura 21, há quatro bandas espectrais discriminantes a serem destacadas, duas regiões positivas e duas negativas de coeficientes de regressão, que correspondem aos quatro picos mais intensos de VIP escores. A banda centrada em torno de 980 nm pode ser atribuída ao segundo sobretom da banda de combinação

de estiramentos assimétricos e simétricos da ligação O-H da água, que é uma interferência na quantificação do etanol. Bandas espectrais centradas em torno de 1150 nm e 1182 nm podem ser atribuídas aos segundos sobretoms dos estiramentos de ligações CH₃ e CH₂ de carboidratos, que são componentes comuns de todas as bebidas analisadas. Por fim, o VIP escore mais intenso, em 1365 nm, está associado à banda espectral mais preditiva relacionada ao analito alvo. Esta banda pode ser atribuída ao primeiro sobretom do estiramento de C-H e ao primeiro sobretom do estiramento de O-H do etanol [161]. É interessante notar que a banda de absorção NIR mais intensa nas amostras, acima de 1400 nm (Figura 19) não contribuiu significativamente para o modelo, uma vez que os respectivos escores VIP apresentaram pequenos valores, abaixo do limiar de 1,0.

A Tabela 11 apresenta FOM complementares que foram avaliadas para o modelo PLS otimizado. Para avaliação da linearidade (Figura 20), foi estimado um coeficiente de correlação de 0,9312, juntamente com uma inclinação de $0,97 \pm 0,03$ e um intercepto de $0,85 \pm 0,22$. Além desses parâmetros, a linearidade do modelo foi verificada por meio da aplicação de testes estatísticos adequados, garantindo a ausência de resíduos sistemáticos. Assim, três testes foram aplicados na sequência: Ryan-Joiner para verificar a normalidade, Brown Forsythe para verificar a homocedasticidade, e Durbin-Watson para confirmar a ausência de autocorrelação dos resíduos, todos no nível de confiança de 95% [100]. Os resultados desses testes garantiram o comportamento aleatório dos resíduos, com valor estimado de 1,78 para o teste de Durbin-Watson, dentro da faixa de aceitação (1,50 - 2,50). O pequeno valor de viés estimado para o conjunto de validação, $0,3 \pm 0,9\%$, garantiu a ausência de erros sistemáticos no modelo. Um teste t de Student com 44 graus de liberdade corroborou a ausência de um viés diferente de zero. A precisão foi avaliada através da repetibilidade e precisão intermediária em três diferentes níveis de concentração para cada uma das quatro bebidas, por meio da estimativa do desvio padrão relativo (RSD). Os valores de RSD variaram entre 0,1 % e 1,7 % no nível de repetibilidade e entre 0,5 % e 1,9 % no nível de precisão intermediária. A sensibilidade analítica foi calculada a partir da estimativa do ruído instrumental, que foi obtida através do desvio padrão agrupado de dez réplicas de espectros de uma placa de Petri vazia. Seu valor inverso, $0,9 \%^{-1}$, fornece uma estimativa da diferença mínima de concentração que o método

pode distinguir e permitiu estabelecer o uso de apenas uma casa decimal para representar os resultados.

Tabela 11. Outros FOM quantitativos estimados para o modelo PLS desenvolvido.

Figuras de Mérito	Parâmetros	Valores	
Precisão	Repetibilidade RSD (%)	0,3	
		Cerveja	0,2
			0,5
			1,7
		Sidra	0,3
			0,4
			0,4
		Hidromel	0,1
			0,2
			0,4
		Vinho	0,3
			0,3
			1,6
			Cerveja
Precisão Intermediária RSD (%)		1,8	
		1,9	
	Sidra	1,1	
		1,1	
		1,8	
		0,8	
	Hidromel	0,5	
		1,0	
		1,1	
		Vinho	1,0
		0,6	
Linearidade	Teste Durbin-Watson	1,78	
	Inclinação	0,97±0,03	
	Intercepto	0,85±0,22	
	Coefficiente de correlação	0,9312	
Sensibilidade Analítica (%⁻¹)		0,9	
Viés (%)		0,3±0,9	
Faixa de Trabalho (%)		4,3-15,3	

Uma região elíptica de confiança conjunta (EJCR) também foi estimada para confirmar a linearidade do método desenvolvido (Figura 22a). Os limites da EJCR estimada no nível de confiança de 95% incluem o ponto ideal correspondente a um coeficiente angular (inclinação) de 1 e um coeficiente linear (intercepto) de 0 para o gráfico de valores preditos *versus* valores de referência. Portanto, os resultados de linearidade, precisão e veracidade permitiram garantir que o método desenvolvido pode ser considerado preciso na faixa de trabalho de 4,3% a 15,3%. Por fim, foram

estimados os intervalos de confiança para a previsão das amostras de validação. A equação EIV foi empregada para estimar SPE variando entre 1,5-1,6% (Figura 22b). Apenas duas das 44 amostras de validação não foram previstas dentro dos respectivos intervalos de confiança estimados (95%), representando uma concordância de 95,5%. Portanto, o método foi capaz de quantificar com precisão o teor alcoólico em cerveja, sidra, hidromel e vinho com apenas um modelo de calibração multivariado.

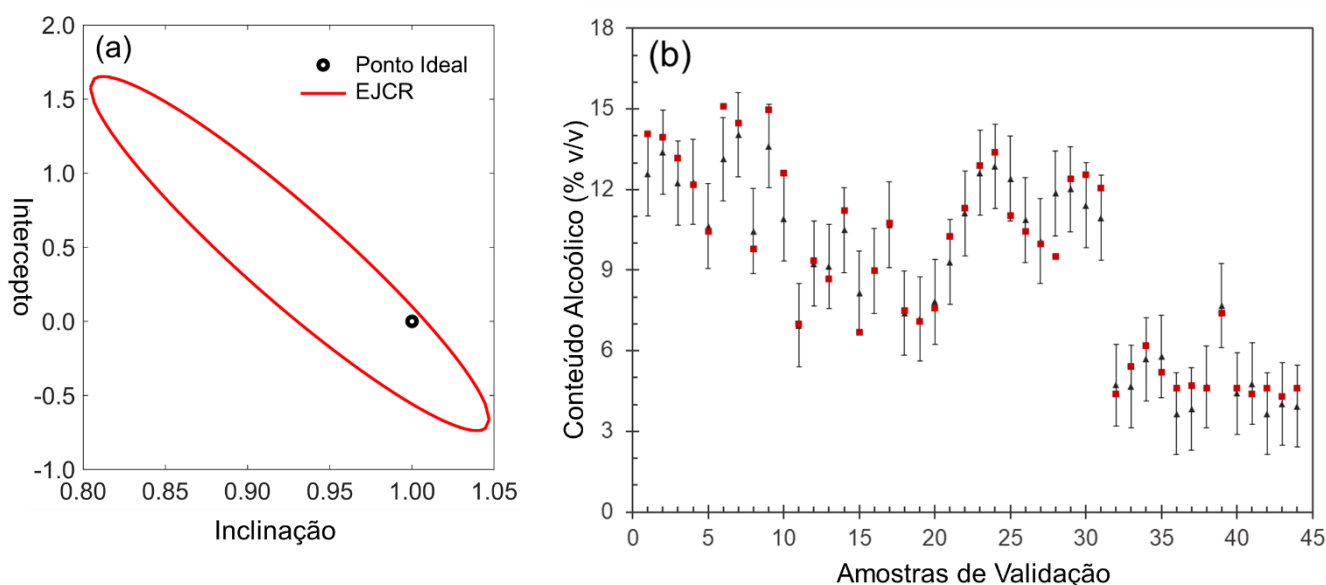


Figura 22. (a) Uma região elíptica de confiança conjunta (EJCR) estimada no nível de confiança de 95% para a linearidade (inclinação versus intercepto) do modelo PLS desenvolvido. **(b)** Intervalos de confiança estimados para a previsão das amostras de validação. Os erros de predição padrão específicos da amostra (SPE) foram calculados com base na equação de erros em variáveis (EIV). Triângulos cinzas significam valores previstos e quadrados vermelhos significam os respectivos valores de referência.

5.5 Conclusão

Neste capítulo, portabilidade, espectroscopia vibracional e quimiometria foram combinados para o desenvolvimento de um método analítico direto, simples, rápido e ambientalmente amigável. Um modelo robusto de calibração multivariada multiproduto foi construído com espectros NIR para a quantificação do teor de etanol em quatro diferentes bebidas alcoólicas fermentadas, demonstrando sua aplicabilidade em amostras reais. O método foi submetido a uma completa validação analítica multivariada por meio da estimativa de figuras de mérito apropriadas, garantindo

assim sua adequação à finalidade. A consistência do modelo também foi corroborada por uma interpretação espectral coerente por meio da análise de seus vetores informativos. O método desenvolvido apresentou várias vantagens sobre as alternativas analíticas tradicionais, mais trabalhosas e demoradas, baseadas em densitometria ou cromatografia. Essas vantagens permitem sugerir a incorporação do novo método em plataformas analíticas portáteis para controle e inspeção de qualidade de bebidas, contribuindo assim para proporcionar maior transparência na cadeia de abastecimento alimentar. Este método pode ser expandido para incorporar bebidas destiladas, ampliando assim sua faixa analítica. Um artigo contemplando os resultados deste trabalho foi publicado no *Talanta Open* (Apêndice D).

**CAPÍTULO 6: MÉTODO DE TRIAGEM PARA DETECÇÃO RÁPIDA DE
DIETILENOGLICOL EM CERVEJA COM BASE EM QUIMIOMETRIA E
ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO PORTÁTIL**

6 MÉTODO DE TRIAGEM PARA DETECÇÃO RÁPIDA DE DIETILENOGLICOL EM CERVEJA COM BASE EM QUIMIOMETRIA E ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO PORTÁTIL

6.1 Introdução

Os princípios de segurança alimentar garantem o acesso a alimentos e bebidas de qualidade e quantidade adequadas para uma alimentação saudável, além de representar um direito humano fundamental e universal. Esses princípios se tornaram um tópico imprescindível atualmente [163]. Quanto a outros produtos alimentícios industrializados, o controle de qualidade de uma bebida complexa como a cerveja também é imperativo [108,164,165]. O desenvolvimento de novos métodos analíticos em conjunto com a avaliação rápida de seus resultados para o controle de qualidade do produto final é uma ferramenta essencial para o aprimoramento das técnicas cervejeiras. Apesar de todos os protocolos de segurança adotados na produção cervejeira, problemas causados por contaminação podem ocorrer por diversos motivos, como falha acidental, falta de condições aceitáveis durante a produção ou distribuição, negligência, imperícia ou até mesmo ações criminosas.

No final de 2019, alguns casos incomuns de insuficiência renal aguda com alteração neurológica foram notificados no Estado de Minas Gerais, Brasil. Em janeiro de 2020, uma cervejaria local foi implicada na intoxicação alimentar de dezenas de pessoas relacionadas a essas ocorrências anteriores, um caso de grande repercussão na mídia brasileira. Posteriormente, análise pericial realizada pela polícia técnica detectou dietilenoglicol (DEG) por meio de cromatografia gasosa acoplada à espectrometria de massa (GC-MS) em pelo menos 41 lotes de cerveja desta cervejaria. Foram notificados mais de 30 casos suspeitos de intoxicação exógena por DEG e pelo menos sete mortes, e a cervejaria teve sua produção interrompida. Os resultados da investigação elucidaram as causas desta tragédia e respaldaram juridicamente a proibição do consumo continuado de cervejas desta cervejaria. O DEG era empregado como agente anticongelante (junto com monoetilenoglicol) na produção cervejeira,

porém foi detectado um furo na tubulação do anticongelante no tanque fermentador, por onde houve vazamento dessa substância, contaminando a cerveja [166-168].

O dietilenoglicol é um líquido límpido, praticamente inodoro, incolor e viscoso, de sabor adocicado e solúvel em água e etanol. Sua fórmula química é $(\text{HOCH}_2\text{CH}_2)_2\text{O}$. O DEG possui vários usos industriais, é utilizado como intermediário químico e solvente, como fluido de transferência de calor, fluido de freios, no processamento de cimento e como lubrificante, mas seu principal uso é como agente anticongelante. Intrinsecamente o DEG possui uma baixa toxicidade, sendo o principal problema sua biotransformação no fígado levando a metabólitos de alta toxicidade, como 2-hidroxiatoxiacético (2-HEAA) e ácido diglicólico (DGA) (e possíveis outros metabólitos ainda desconhecidos). Esses metabólitos podem persistir no organismo por muito tempo e causar síndrome tóxica aguda, resultando em insuficiência renal por degeneração tubular cortical e necrose tubular proximal. Os efeitos do envenenamento por DEG podem ser divididos em duas fases. A primeira fase consiste em sintomas gastrointestinais. O segundo estágio causa dano renal emergente e, na ausência de tratamento adequado, pode levar à morte [169,170]. Doze vítimas da intoxicação notificada em Minas Gerais apresentaram concentrações de DGA no sangue total variando entre 2 e 108 mg mL⁻¹. A dose mínima para toxicidade do DEG em humanos não foi claramente estabelecida, mas a ingestão de 1 mL kg⁻¹ é considerada letal [166]. Na verdade, a faixa de doses consideradas letais para o DEG varia amplamente, em mais de duas ordens de grandeza. Doses de 0,5 a 5 g kg⁻¹ têm sido sugeridas [168,169]. Ao contrário de produtos farmacêuticos e excipientes, não há limite máximo permitido para DEG em alimentos e bebidas [171].

De acordo com a literatura, nas últimas nove décadas, o DEG foi implicado como agente etiológico de pelo menos 15 incidentes envolvendo diversas matrizes, como medicamentos, produtos de saúde e bebidas alcoólicas [171-175]. Vários métodos analíticos têm sido relatados na literatura para a determinação de DEG nestes tipos de matrizes, principalmente aqueles baseados em técnicas cromatográficas. A técnica mais utilizada tem sido a cromatografia gasosa com detecção por ionização de chama (GC-FID). O GC-FID tem sido empregado para determinar o DEG em produtos farmacêuticos [173], cremes dentais [176], alimentos como refrigerantes, sucos, fórmulas infantis, cereais, farinha e salgadinhos [177] e vinho [172]. Este último

método foi desenvolvido devido a um caso real em que o DEG foi detectado em níveis superiores a 100 mg L^{-1} em vinhos, na Áustria; a adição de DEG ao vinho teria sido feita com o objetivo de melhorar seu sabor adocicado. A cromatografia líquida de ultra alto desempenho acoplada a espectrômetro de massas (UHPLC-MS) tem sido usada para determinar DEG em creme dental [178] e GC-MS para determinar o mesmo analito em plasma humano [179]. A GC-MS também foi utilizada recentemente para detectar e quantificar DEG em cerveja por um grupo de pesquisa brasileiro, reverberando a ocorrência de contaminação em Minas Gerais [171].

Apesar de suas vantagens, os métodos cromatográficos apresentam algumas desvantagens, como a exigência de análises onerosas, trabalhosas e demoradas, consumindo solventes orgânicos, gerando resíduos químicos e exigindo etapas prévias de preparo da amostra, como derivatização, limpeza e pré-concentração. Por outro lado, o surgimento da química verde trouxe preocupações sobre o desenvolvimento de novos métodos analíticos ambientalmente mais amigáveis. As técnicas espectroscópicas vibracionais apresentam menor sensibilidade em comparação com a cromatografia, mas representam uma alternativa verde, simples e mais rápida para desenvolver métodos de triagem. O uso concomitante de quimiometria é quase obrigatório. Assim, técnicas vibracionais têm sido combinadas com modelos quimiométricos, principalmente com o PLS, com a finalidade de detectar e quantificar DEG em diversas matrizes. O DEG foi determinado em glicerina através da espectroscopia Raman [175], em xarope para tosse à base de glicerina através das espectroscopias de infravermelho médio e próximo [174], e em creme dental e gel dentifrício por espectroscopia no infravermelho por transformada de Fourier com reflectância total atenuada (ATR-FTIR) [180].

A portabilidade é uma tendência importante na espectroscopia, permitindo em tempo real análises simples e robustas [117]. Os espectrômetros portáteis de infravermelho próximo (NIR) possuem sensibilidade e desempenho semelhantes aos de equipamentos de bancada, apresentando outras vantagens como tamanho reduzido, menor consumo de energia e ausência de partes móveis em sua estrutura óptica. Essas características melhoram a utilidade da espectroscopia NIR como alternativa às técnicas cromatográficas para o controle de qualidade de alimentos e bebidas. Particularmente para a fabricação de cervejas, os espectrômetros NIR

portáteis podem ser vistos como sensores úteis para estabelecer mais controle em todo o processo, reduzindo os custos de produção e melhorando a confiança do produto final. Não há mais a necessidade de monitoramento em escala laboratorial além das instalações de produção usadas para processos comerciais de maltagem e fabricação de cerveja. O monitoramento da fabricação de cerveja pode ser constante e automatizado. Além disso, a espectroscopia NIR portátil combinada com quimiometria pode ser utilizada para monitorar outros parâmetros no âmbito da tecnologia analítica de processo (PAT) [164,165]. Este equipamento pode ser usado, por exemplo, para determinar teor alcoólico [152] e cor [108] de cervejas.

6.2 Objetivos

O objetivo deste estudo foi combinar espectroscopia NIR portátil e análise discriminante PLS (PLS-DA) para desenvolver um método de triagem rápido e simples para a detecção direta de dietilenoglicol em cerveja. Com o objetivo de obter um modelo robusto, cervejas de diferentes estilos, teores alcoólicos, marcas e fabricantes foram utilizadas para construir o modelo multivariado. O método proposto foi validado incorporando aspectos multivariados por meio da estimativa de FOM qualitativas apropriadas.

6.3 Materiais e Métodos

6.3.1 Equipamentos e *softwares*

Os espectros NIR foram coletados usando um espectrofotômetro portátil MicroNIR® 1700 da Viavi Solutions (Milpitas, CA, EUA), com uma faixa de trabalho de comprimento de onda entre 900-1700 nm. Os dados foram processados por meio do software MATLAB®, versão 7.13 (MathWorks, Natick, EUA), juntamente com o PLS Toolbox, versão 6.5 (Eigenvector Technologies, Manson, EUA).

6.3.2 Amostras de Cerveja

Cem amostras de cerveja foram adquiridas principalmente no comércio local da Região Metropolitana de Belo Horizonte, Minas Gerais, Brasil, e armazenadas em temperatura ambiente até a análise. Essas amostras de cerveja foram compradas de uma grande variedade de marcas, cervejarias, teores de álcool e tipos, incluindo lager, pale ale, red ale, Vienna lager, Pilsen, IPA, weissbier, strong pale ale, porter e stout. Todas as amostras foram previamente verificadas quanto à ausência de DEG usando uma metodologia cromatográfica descrita na próxima seção.

6.3.3 Método GC-MS para garantir a ausência de DEG em cervejas não contaminadas

Um método cromatográfico foi desenvolvido para verificar a ausência de DEG em cervejas não contaminadas. As amostras de cerveja foram sonicadas por 5 minutos, diluídas dez vezes com metanol (grau HPLC), sendo uma alíquota de 100 µL transferida para um Eppendorf de 2 mL. A amostra foi centrifugada por 5 minutos a 3000 rpm, e transferida para frascos de 2 mL, antes da injeção no cromatógrafo. Um sistema Shimadzu GCMS-QP2010 Plus (Kyoto, Japão) foi utilizado em conjunto com uma coluna Innowax (60 m x 0,25 mm x 0,25 µm). O volume de injeção foi de 1,0 µL no modo *splitless*, as temperaturas da porta de injeção e do detector foram ajustadas em 250 °C, e o hélio foi usado como gás de arraste em uma vazão de 1,0 mL min⁻¹. Os íons de m/z 43 e 45, características de DEG, foram monitorados. A temperatura do forno foi mantida a 100 °C por 1 minuto, aumentada para 200 °C a uma taxa de 10 °C/min, mantida a 200 °C por 10 minutos, aumentada para 240 °C a uma taxa de 10 °C/min e mantida a 240 °C por 5 minutos.

6.3.4 Preparação das amostras e aquisição de espectros

Todos os reagentes usados são de grau analítico e foram utilizados como recebidos. Um volume de 10 mL de cada amostra de cerveja foi desgaseificado usando um banho ultrassônico (Equilab ULTRASONIK 28H, Madrid, Espanha) por 5 minutos. Uma gota de 1-octanol foi adicionada a cada amostra de cerveja para evitar a formação de espuma excessiva. Cem amostras de cervejas originais foram

selecionadas em ordem aleatória e enriquecidas com DEG em uma faixa de concentração de 10 a 1000 mg L⁻¹, com um incremento de 20 mg L⁻¹ entre as amostras. Um volume de 2,0 mL de cada amostra contaminada com DEG foi transferido para um tubo Eppendorf de 2 mL. Para amostras não contaminadas, 2,0 mL de cada amostra de cerveja original foram colocados em um tubo Eppendorf de 2 mL sem qualquer preparação adicional. Todos os tubos Eppendorf foram colocados em um suporte de acrilonitrila-butadieno-estireno especialmente projetado e construído para esta aplicação com uma impressora 3D (Creality Ender 3, Shenzhen, China), que ajudou a coletar os espectros das duzentas amostras com o espectrofotômetro MicroNIR[®] portátil (Figura 23). Os espectros foram registrados no modo de refletância difusa, com vinte varreduras cada. A faixa de comprimento de onda foi de 908 a 1676 nm com resolução de 6,25 nm. Durante todas essas medições, a temperatura do laboratório foi controlada em 25±1°C.

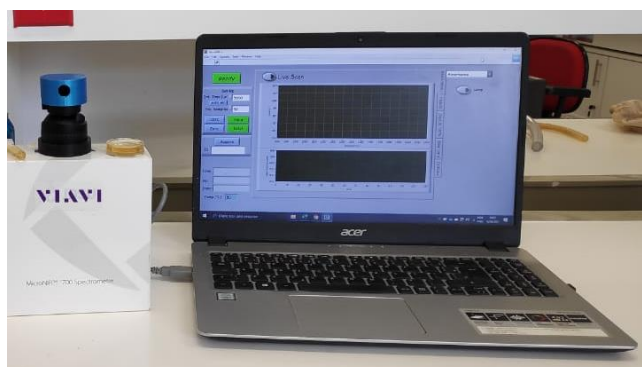


Figura 23. Sistema usado para adquirir os espectros. Um suporte de acrilonitrila-butadieno-estireno especialmente projetado e construído para esta aplicação com uma impressora 3D, ao lado um Eppendorf de 2 mL contendo 1 mL de amostra de cerveja.

6.3.5 Validação analítica qualitativa

A validação analítica do modelo desenvolvido foi avaliada pela estimativa de FOM qualitativas apropriadas, como taxa de falso-negativo (TFN), taxa de falso-positivo (TFP), taxa de sensibilidade (TSL), taxa de especificidade (TES), precisão, acordância (ACO), concordância (CON), capacidade de detecção (CC β) e limite de decisão (CC α). Precisão é uma FOM que expressa a taxa global de previsões incorretas, independentemente de serem positivas ou negativas. ACO e CON expressam a precisão de métodos qualitativos, sendo equivalentes à repetibilidade quantitativa e à precisão intermediária, respectivamente [89]. Essas FOM foram

inicialmente propostas para métodos univariados [181] e podem ser estendidas diretamente para métodos multivariados [96]. A ACO foi estimada usando dez réplicas em cinco níveis diferentes (50, 80, 100, 400 e 1000 mg L⁻¹) analisadas no mesmo lote em condições de repetibilidade (no mesmo dia e com o mesmo analista). A CON foi estimada usando o mesmo protocolo empregado para ACO, repetido em dois dias diferentes por dois analistas diferentes. CC α e CC β foram estimados a partir de curvas de probabilidade de detecção (PoD), que harmonizam os conceitos e parâmetros estatísticos entre a validação de métodos quantitativos e qualitativos. Este tipo de curva apresenta graficamente a relação entre a probabilidade de resultados positivos obtidos de métodos qualitativos, ou seja, os resultados dos modelos de classificação, e a concentração do analito. Os valores de probabilidade são calculados a partir de medições independentes em cada nível de adulteração [184,185]. Para estimar CC α e CC β , dez réplicas de amostras fortificadas foram preparadas em dez níveis de concentração diferentes (em branco, 10 mg L⁻¹, 20 mg L⁻¹, 40 mg L⁻¹, 60 mg L⁻¹, 80 mg L⁻¹, 100 mg L⁻¹, 400 mg L⁻¹, 700 mg L⁻¹ e 1000 mg L⁻¹).

6.4 Resultados e Discussão

A cerveja é uma matriz complexa. Seus principais ingredientes incluem água, que é um componente crítico no processo de fabricação de cerveja, além de lúpulo, levedura, e outros aditivos como frutas, ervas e outras plantas. A [Figura 24](#) mostra os espectros NIR pré-processados, com a faixa de comprimento de onda entre 908 e 1676 nm, das 200 amostras de cerveja analisadas, 100 contaminadas com DEG e 100 não contaminadas. A ausência de DEG nas amostras de cerveja não contaminadas foi verificada por GC-MS (seção 6.3.3). Observando os espectros, é possível destacar duas bandas de absorção mais intensas, uma centrada em torno de 1200 nm e outra centrada em torno de 1400 nm. A primeira banda espectral, na região entre 1100 e 1250 nm, pode ser atribuída ao segundo sobretom de estiramento das ligações de CH, CH₂ e CH₃; a segunda banda pode estar relacionada ao primeiro sobretom de estiramento de OH [128]. Mais especificamente, a primeira banda mencionada pode ser atribuída ao sobretom de estiramento C-H do etanol, enquanto a banda espectral em comprimentos de onda mais altos pode ser relacionada às ligações O-H do etanol e da água [108]. Os perfis espectrais de todas as amostras são semelhantes, evitando

assim a possibilidade de discriminação entre cervejas contaminadas e não contaminadas por inspeção visual. Considerando a seletividade limitada da espectroscopia NIR, o uso concomitante com a quimiometria para extrair informações é imperativo. Assim, foi construído um modelo PLS-DA de classificação multivariada supervisionada.

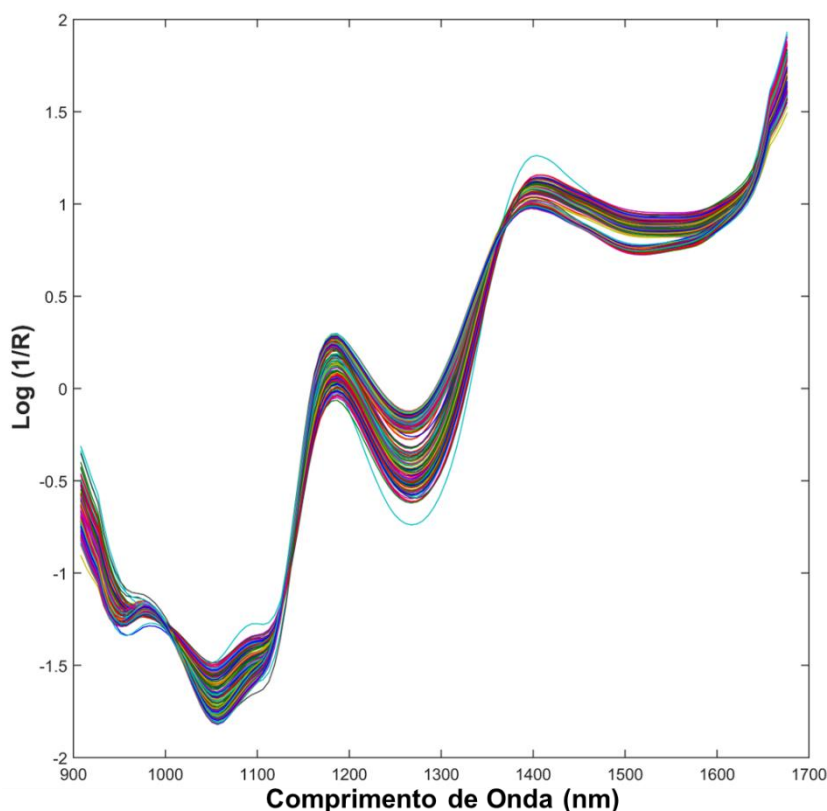


Figura 24. Espectros NIR pré-processados por alisamento Savitzky-Golay e variação normal padrão (SNV) de todas as 200 amostras de cerveja analisadas.

Para a construção do modelo PLS-DA, as amostras foram divididas em conjuntos de treinamento e teste, na proporção de aproximadamente dois terços e um terço, respectivamente. O algoritmo Kennard-Stone foi usado para selecionar amostras de treinamento de cada uma das duas classes [95]. Assim, 134 amostras de treinamento foram selecionadas, enquanto as 66 amostras restantes foram usadas para construir o conjunto de teste. Com o objetivo de eliminar desvios típicos de espectros infravermelho de refletância difusa e causados por espalhamento multiplicativo de luz, os dados foram sequencialmente pré-processados por alisamento de Savitzky-Golay (7 pontos e ajuste polinomial de primeira ordem), variação normal padrão (SNV) e centrar na média [129].

Subconjuntos aleatórios (2 divisões de dados e 20 iterações) foram aplicados para selecionar o melhor número de variáveis latentes (VL). Foi escolhido o número de VL que apresentou o menor erro de classificação de validação cruzada (CVCE). O melhor modelo foi selecionado com 6 VL após otimização por detecção e remoção de *outliers*. O número de *outliers* não pode ultrapassar o limite recomendado de 22,2% para os conjuntos de treinamento e teste [97]. Doze das duzentas amostras originais foram removidas como *outliers*, sendo 7 do conjunto de treinamento (5%) e 5 do conjunto de teste (8%). Após a remoção dos *outliers*, o modelo foi reconstruído. Os resultados para a otimização do modelo, incluindo sensibilidade, especificidade e erro de classificação de validação cruzada (CVCE), são mostrados na [Tabela 12](#).

Tabela 12. Resultados para a otimização do modelo PLS-DA desenvolvido através da detecção e remoção de *outliers*.

		Nº de Amostras de Treinamento	Nº de Amostras de Teste	Nº de variáveis latentes (VL)	Sensibilidade (CV) ^a	Especificidade (CV) ^a	Erro de classificação de validação cruzada (CVCE)
Antes	Livre de DEG	67	33	6	0,851	0,834	0,158
	DEG	67	33				
Depois	Livre de DEG	62	29	6	0,956	0,927	0,059
	DEG	65	32				

^aValidação Cruzada (CV)

Sensibilidade (TSL) e especificidade (TES) da validação cruzada são figuras específicas de mérito (FOM) usadas na validação analítica qualitativa. Como citado na seção 2.5, TSL e TES são considerados FOM relacionados à seletividade do método, e o parâmetro CVCE garante que o menor número de amostras de treinamento seja classificado erroneamente [96]. O modelo PLS-DA foi construído com 6 VL, o que representa 99,63% da variância espectral (bloco **X**) e 71,41% da variância da classe (bloco **Y**). O limite Bayesiano foi estimado em 0,51. Os valores previstos de **Y** de cada amostra para cada classe são mostrados na [Figura 25](#).

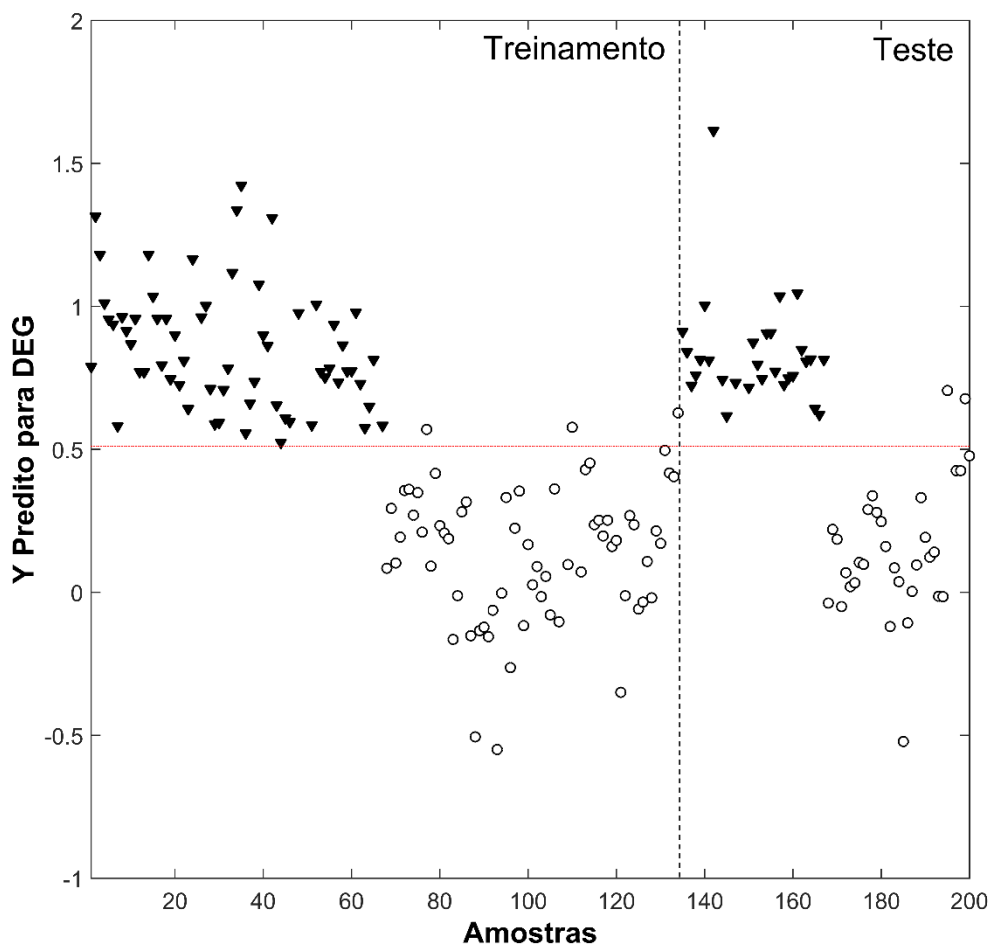
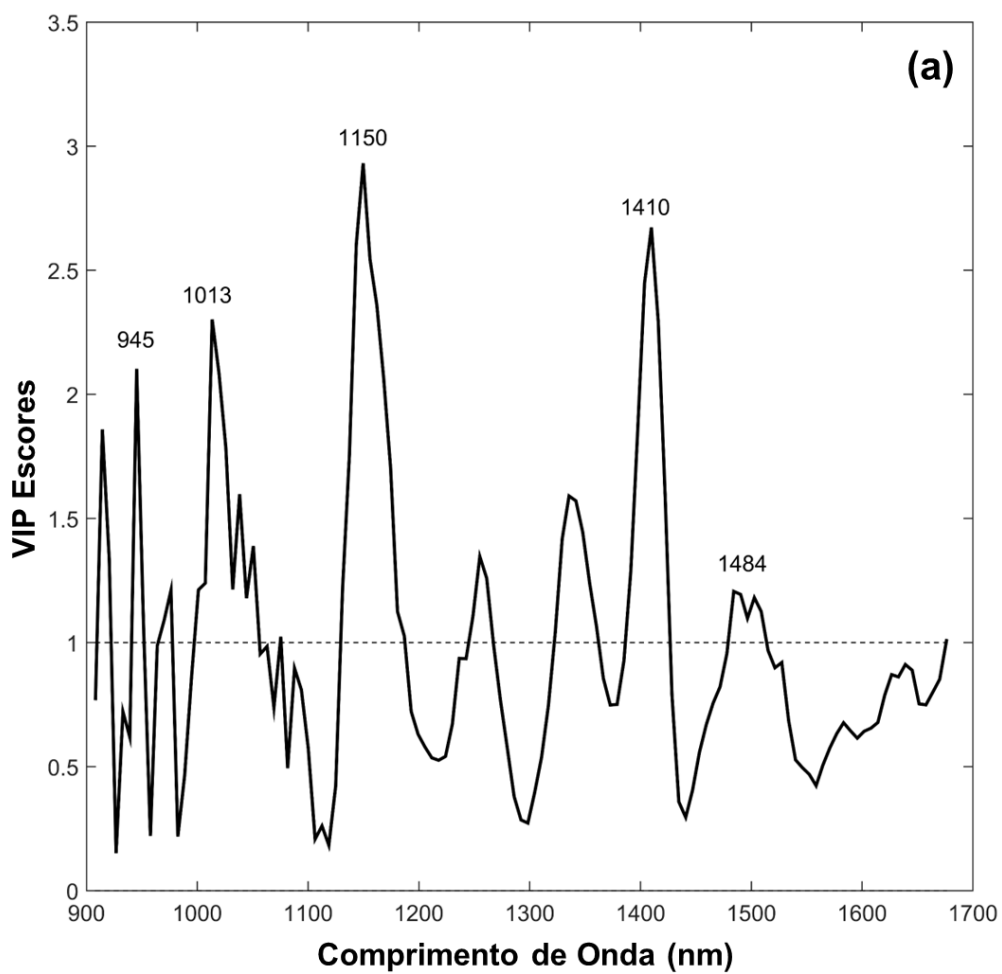


Figura 25. Valores de Y previstos para o modelo PLS-DA construído para discriminar amostras contaminadas e não contaminadas de DEG. A linha tracejada horizontal vermelha indica o limite Bayesiano e a linha tracejada vertical indica a separação entre as amostras de treinamento e de teste. Triângulos cheios indicam amostras não contaminadas (autênticas) e círculos vazios amostras contaminadas com DEG.

Os vetores estimados de coeficientes de regressão e VIP escores para o modelo PLS-DA desenvolvido são mostrados na [Figura 26](#). Os VIP escores ([Figura 26a](#)) indicam as variáveis que mais contribuem para cada previsão de classe em valores absolutos. Já os vetores de regressão ([Figura 26b](#)) podem mostrar contrastes entre variáveis com coeficientes mais positivos, característicos de cada classe alvo modelada específica (autêntica), e diretamente relacionados à classificação correta. Os coeficientes mais negativos podem ser associados às amostras contaminadas neste caso, ou seja, principalmente à molécula de DEG. Os escores VIP fornecem valores absolutos e não permitem estabelecer relações entre as variáveis mais importantes e o analito ou as interferências. Assim, a interpretação espectral do

modelo pode ser melhorada inspecionando conjuntamente os escores VIP e os coeficientes de regressão.



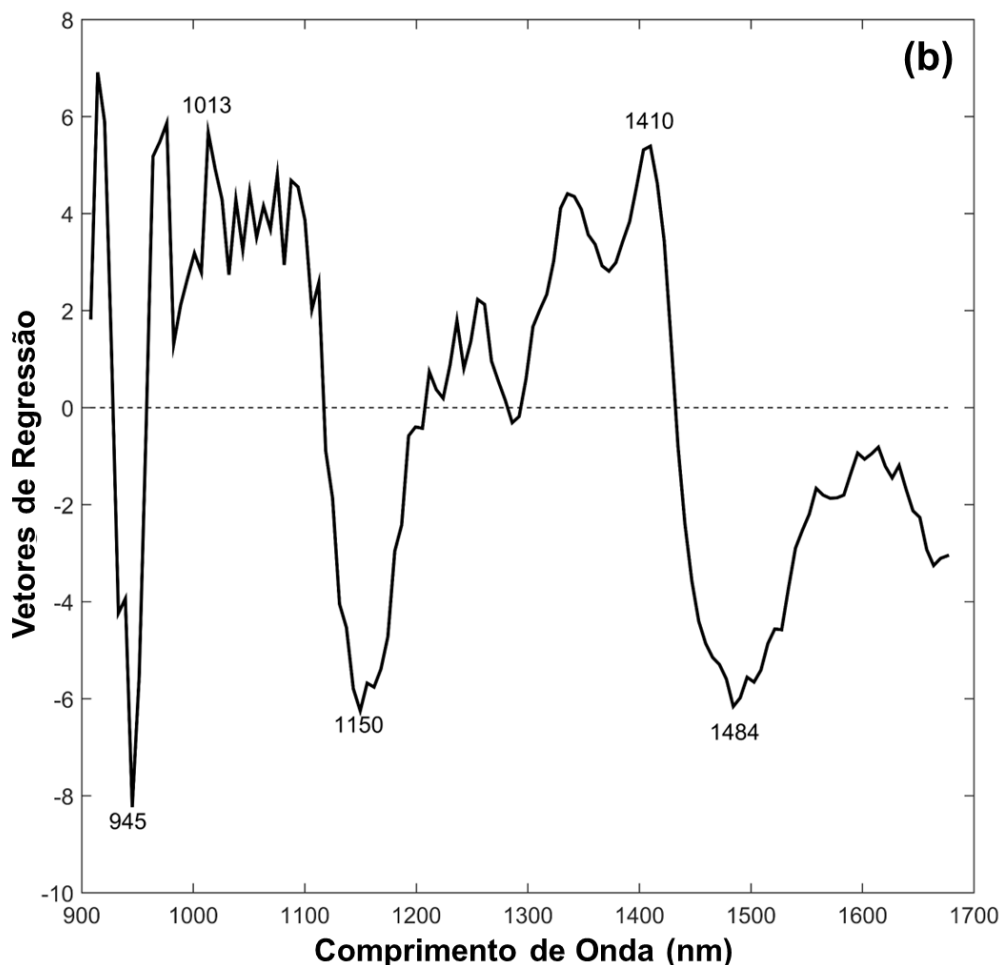


Figura 26. Vetores informativos do modelo PLS-DA. **(a)** Escores (VIP). A linha pontilhada horizontal preta indica o limite de 1,0, acima do qual as variáveis são consideradas significativas para o modelo. **(b)** Coeficientes de regressão.

As variáveis com coeficientes de regressão mais positivos contribuem para a caracterização da classe 1 (neste estudo, amostras de cerveja não contaminadas), enquanto as variáveis com coeficientes de regressão mais negativos caracterizam a classe 0 (amostras de cerveja contaminadas com DEG). As variáveis mais discriminantes não necessariamente coincidem com as absorções mais intensas. Visando complementar esta interpretação espectral, o espectro NIR de uma amostra de DEG puro é mostrado na Figura 27a, juntamente com o gráfico de *loadings* de VL2 na Figura 27b. Ao observar características espectrais, VL2 foi atribuído à contribuição DEG para o modelo PLS-DA.

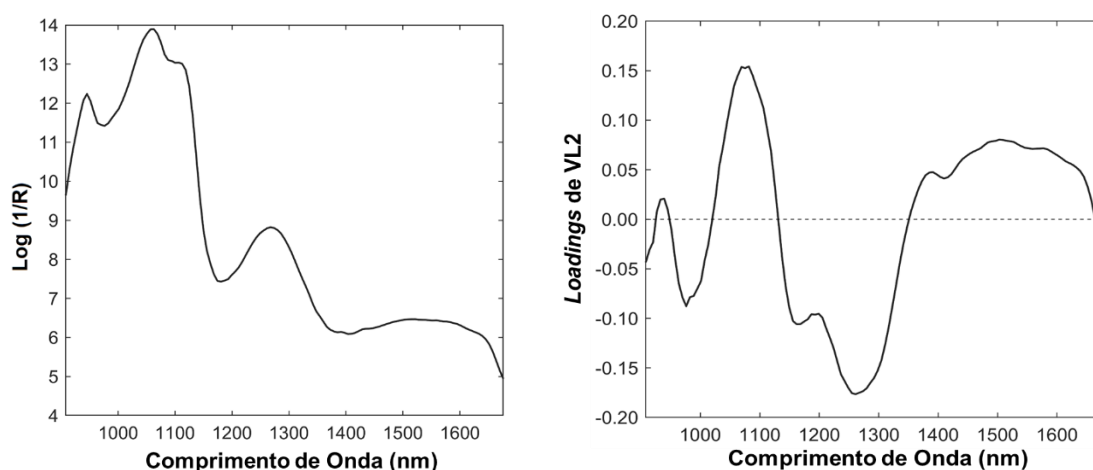


Figura 27. (a) Espectro NIR de uma amostra DEG pura. **(b)** *Loadings* de VL2 do modelo PLS-DA, que foi atribuído a DEG.

Observando as Figura 26 e 27, foi possível identificar as regiões espectrais no NIR que mais contribuiriam para a detecção do dietilenoglicol na cerveja. Existem três regiões negativas de coeficientes de regressão, as quais estão relacionadas com a presença de DEG. O pico negativo mais intenso e agudo dos coeficientes de regressão, em 945 nm, está relacionado ao terceiro sobretom das vibrações de CH₂ [128] e pode ser associado à estrutura química do DEG. A grande banda de coeficientes de regressão centrada em 1484 nm corresponde a uma intensa e ampla banda de absorção do espectro NIR de DEG e pode ser atribuído ao primeiro sobretom de suas ligações OH [174, 175]. A banda larga de coeficientes de regressão centrada em 1484 nm corresponde a uma intensa absorção no espectro NIR de DEG [181,182] e pode ser atribuída ao primeiro sobretom de suas ligações OH [174]. Por fim, a banda centrada em 1150 nm, que também apresenta os maiores escores VIP, pode ser atribuída ao segundo sobretom das ligações CH₂ e corresponde a uma banda de menor absorção no espectro NIR de DEG [181,182]. Todas essas atribuições espectrais também podem ser corroboradas pela observação das Figura 27a e Figura 27b. Em contraste, as regiões mais positivas dos coeficientes de regressão, que também correspondem a alguns dos maiores escores VIP, estão centradas em 1013 e 1410 nm. Elas podem ser atribuídas ao segundo e primeiro sobretoms do estiramento de ligações OH, que podem ser associadas a constituintes da cerveja, como etanol e água [128]. Esses sinais são interferências diretas na detecção de DEG.

Tabela 13. FOM qualitativas estimadas para o modelo PLS-DA. Unidades em %.

	TFP	TFN	TSL	TES	Precisão
Treinamento	0	4,5	95,5	100	95,5
Teste	0	6,1	93,9	100	93,9

	CON ^a					ACO ^a				
	50	80	100	400	1000	50	80	100	400	1000
Teste	25	81	100	100	100	44	80	100	100	100

^a concentração em mg L⁻¹

Neste trabalho, algumas FOM apropriadas foram estimadas. Na [Tabela 13](#), sete FOMs diferentes são mostradas para os conjuntos de treinamento e teste. O modelo PLS-DA não apresentou nenhum falso positivo, mas teve três falsos negativos no conjunto de treinamento e 2 no conjunto de teste. Isso significa que o DEG corretamente não foi detectado em todas as amostras não contaminadas. Os falsos positivos corresponderam a amostras contaminadas com os menores níveis de concentração de DEG, apontando assim para uma limitação na sensibilidade do método, que pode estar relacionada aos limites da técnica analítica, espectroscopia NIR. Ao realizar análises em amostras que foram classificadas erroneamente, é possível discutir um assunto interessante sobre equipamentos portáteis: Quais são os limites da técnica? Apesar de ser uma técnica que promete análises rápidas e precisas, é sempre necessário saber a melhor forma de aplicá-la. Considerando a limitação observada do método para detectar as menores concentrações de DEG, as FOM relacionados à precisão foram estimados em cinco níveis acima de 50 mg L⁻¹ (50, 80, 100, 400 e 1000 mg L⁻¹). A ACO foi avaliada em cinco níveis (50, 80, 100, 400 e 1000 mg L⁻¹) analisados em condições de repetibilidade. A CON foi estimada em dois dias diferentes por dois analistas distintos. O objetivo seria testar o modelo em diferentes faixas de concentração. Como esperado, a concentração de 50 mg L⁻¹ apresentou os menores valores de CON e ACO (25% e 44%, respectivamente). Para níveis iguais ou superiores a 100 mg L⁻¹, não foram observados erros de classificação e ambas FOM de precisão foram estimadas em 100%. [183]. Os resultados descritos acima indicam a falta de precisão do método em concentrações de DEG abaixo de 100 mg L⁻¹. Isso reforça a necessidade de estabelecer limites quantitativos para este

método qualitativo por meio da estimativa das medidas apropriadas de FOM, ou seja, capacidade de detecção ($CC\beta$) e limite de decisão ($CC\alpha$). [169].

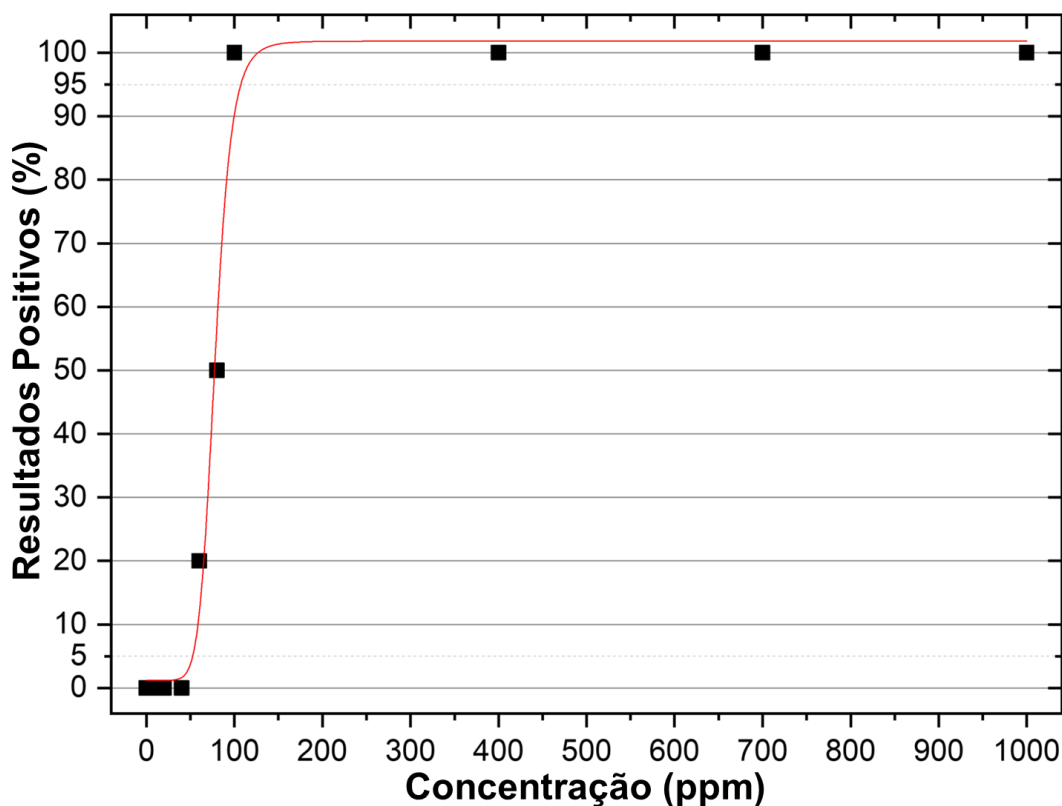


Figura 28. Curva de probabilidade de detecção (POD) obtida por regressão logística, com resultados positivos em porcentagem e concentração em mg L^{-1} .

Para calcular $CC\alpha$ e $CC\beta$, dez amostras replicadas fortificadas com DEG foram preparadas em dez níveis de concentração diferentes (branco, 10 mg L^{-1} , 20 mg L^{-1} , 40 mg L^{-1} , 60 mg L^{-1} , 80 mg L^{-1} , 100 mg L^{-1} , 400 mg L^{-1} , 700 mg L^{-1} e 1000 mg L^{-1}). Usando um ajuste logístico (Figura 28) e seguindo o procedimento descrito na ISO 11843, $CC\alpha$ e $CC\beta$ foram calculados. Segundo a literatura, os critérios de identificação devem ser satisfeitos em 95% ($1-\beta$) dos casos no $CC\beta$ e 5% no $CC\alpha$. Portanto, $CC\alpha$ e $CC\beta$ foram calculados de acordo com a equação de regressão logística, considerando X como 5% para $CC\beta$ e 95% para $CC\alpha$. A capacidade de detecção ($CC\beta$) e o limite de decisão ($CC\alpha$) foram de 52 mg L^{-1} e 106 mg L^{-1} , respectivamente [87,185]. Portanto, o método desenvolvido não pode ser considerado confiável para detectar contaminação por DEG em cerveja abaixo de 52 mg L^{-1} e sua faixa analítica deve ser restrita. Naturalmente, a presente metodologia baseada em espectroscopia NIR e quimiometria tem menor sensibilidade do que os métodos cromatográficos. Por

exemplo, um artigo publicado recentemente desenvolveu um método GC-MS para quantificar DEG em cerveja, que apresentou um limite de detecção de 5,0 mg L⁻¹ e um limite de quantificação de 10,0 mg L⁻¹ [171]. No entanto, nossa metodologia proposta tem várias vantagens e uma relação custo-benefício muito boa como método de triagem, pois pode ser implementada em uma plataforma analítica portátil para análise *in loco* com custos reduzidos.

6.5 Conclusão

Foi desenvolvido um método de triagem simples, rápido e direto para detectar o dietilenoglicol em cerveja baseado em um espectrômetro portátil na região do infravermelho próximo. Essa metodologia é de especial interesse, considerando a ocorrência recente no Brasil de envenenamento por dietilenoglicol causado pela contaminação da cerveja, que resultou em algumas mortes e dezenas de pessoas intoxicadas. Para o modelo desenvolvido de PLS-DA, o limite de decisão (CC α) foi estimado em 52 mg L⁻¹. O método foi validado de acordo com diretrizes brasileiras e internacionais [87,97], sendo considerado preciso. Também foi considerado robusto, incorporando uma grande variedade de cervejas de diferentes estilos e origens que foram usadas para construir o modelo. O método proposto exigiu uma pequena quantidade de amostra (2 mL), gastou apenas 30 s por aquisição espectral, não utilizou reagentes ou solventes e nem gerou resíduos químicos, além de exigir um mínimo preparo de amostra (desgaseificação). Por fim, esta metodologia pode ser adaptada a plataformas analíticas portáteis e implementada a baixo custo para o controle de qualidade na indústria cervejeira ou ser utilizada como opção de triagem rápida para análise forense *in loco* em casos de suspeita de intoxicação/envenenamento por cerveja. Um artigo contemplando os resultados deste trabalho foi publicado no *Food Chemistry* (Apêndice C).

7 CONCLUSÕES GERAIS

A análise de imagens digitais e a aplicação de técnicas espectroscópicas se mostraram como poderosas ferramentas em conjunto com a quimiometria na análise de cerveja. A associações entre estas técnicas e ferramentas quimiométricas como o PLS e PLS-DA permitiu o desenvolvimento de modelos capazes de determinar parâmetros como a cor e o conteúdo alcóolico de maneira simples, rápida e eficiente. As quatro aplicações apresentadas aqui mostraram apenas um pouco do potencial desses métodos na área cervejeira. Embora os métodos quimiométricos, como o PLS e PLS-DA, já estejam razoavelmente popularizados entre a comunidade de química analítica, as suas aplicações envolvendo equipamentos portáteis ainda são relativamente pouco conhecidos e muito do seu potencial ainda está por ser explorado. Nessa tese, as aplicações desenvolvidas focaram no uso de equipamentos portáteis, e na sua possível futura inclusão na internet das coisas (IoT), além da valorização da realização de uma validação analítica incluindo uma harmonização entre métodos univariados e multivariados.

REFERÊNCIAS

- [1] MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa nº 65, de 10 de dezembro de 2019. Brasília: Secretaria de Defesa Agropecuária, 2019. Disponível em:< <https://www.gov.br/agricultura/pt-br/assuntos/inspecao/produtos-vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/instrucao-normativa-no-65-de-10-de-dezembro-de-2019.pdf/view>>. Acesso em: março de 2023.
- [2] EBLINGER, H. M. **Handbook of brewing: processes, technology, markets**. Darmstadt: Willey-VCH, 2009
- [3] LAPERE, Y. **The beer sector in Brazil**. São Paulo: Flanders Investment & Trade Market Survey, 2015. Disponível em:< https://www.flandersinvestmentandtrade.com/export/sites/trade/files/market_studies/800151125155419/800151125155419_10.pdf >. Acesso em: outubro de 2022
- [4] MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Decreto Nº 9.902, de 8 de Julho de 2019. Brasília: Secretaria de Defesa Agropecuária, 2019. Disponível em:<[http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D9902.htm# art2](http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D9902.htm#art2)>. Acesso em: 19 fev 2022.
- [5] CILO, N. Pequenas e médias cervejarias pressionam grandes marcas. Estado de Minas, Belo Horizonte. Belo Horizonte: Estado de Minas, 2019. Disponível em:< https://www.em.com.br/app/noticia/economia/2019/02/08/internas_economia,1028869/pequena-e-media-cervejaria-pressionam-grandes-marcas.shtml >. Acesso em: outubro de 2022.
- [6] HUIGE, N. **Brewery by-products and effluents**. Em: Graham G. Stewart e Fergus G. Priest (Org.). Handbook of Brewing. Toronto: CRC Press, 2006.
- [7] HORNSEY, I. S. **A history of beer and brewing**. Cambridge: Royal Society of Chemistry, 2003.
- [8] MEUSSDOERFFER, F. G. **A comprehensive history of beer brewing**. Em: Hans Michael Eßlinge (Org.). Handbook of brewing: Processes, technology, markets. Weinheim: Wiley-VCH Verlag GmbH & Co, 2009.
- [9] SAMUEL, D. Archaeology of ancient Egyptian beer. **Journal of the American Society of Brewing Chemists**, 54, 3-12, 1996.
- [10] SAMUEL, D. Fermentation Technology 3,000 Years Ago - The Archaeology of Ancient Egyptian Beer. **SGM QUARTERLY**, 24, 3-5, 1997.
- [11] BAMFORTH, C. **Beer: tap into the art and science of brewing**. New York: Oxford University Press, 2009.

- [12] MARQUES, T. C. N.; OLIVEIRA, M. T. R. Inovação de produto ou saída para a crise? O lançamento da cerveja Brahma Chopp no verão de 1934. **História Econômica & História de Empresas**, 1, 87-120, 2003.
- [13] PALMER, J.; KAMINSKI, C. **Water: A comprehensive guide for brewers**. Boulder: Brewers Publications, 2013.
- [14] WOLFGANG, K. **Technology brewing and malting**. Dortmund: VLB Berlin, 1996.
- [15] BAMFORTH, C. **Brewing materials and processes: A practical approach to beer excellence**. Davis: Academic Press, 2016.
- [16] BARTH, R. **The Chemistry of Beer**. Hoboken: Owl's Nest Publishing, 2011
- [17] FAOSTAT, Food and Agriculture Organization Corporate Statistical Database. Countries - Select All; Regions - World + (Total); Elements - Production Quantity; Items - Barley; Years – 2010-2017. Roma: FAOSTAT, 2020. Disponível em: <<https://www.fao.org/faostat/en/#data/QCL/visualize>>. Acesso em: outubro de 2022.
- [18] MALLET, J. **Malt: a practical guide from field to brewhouse**. Boulder: Brewers Publications, 2014.
- [19] LEWIS, M.; BAMFORTH, C. **Essays in brewing science**. Davis: Springer Science & Business Media, 2006.
- [20] CAIÃO, E. **Cevada: história e evolução**. Brasília: Embrapa, 2008. Disponível em: <<https://www.alice.cnptia.embrapa.br/bitstream/doc/821797/1/ID100272008PLp289cevada.pdf>>. Acesso em: outubro de 2022.
- [21] BOULTON, C. **Encyclopaedia of brewing**. Oxford: John Wiley & Sons, 2013.
- [22] VERZELE, M.; DE KEUKELEIRE, D. **Chemistry and analysis of hop and beer bitter acids**. Gent: Elsevier, 2013.
- [23] HIERONYMUS, S. **For the love of hops: The practical guide to aroma, bitterness and the culture of hops**. Boulder: Brewers publications, 2012.
- [24] WHITE, C.; ZAINASHEFF, J. **Yeast: the practical guide to beer fermentation**. Boulder: Brewers publications, 2010.
- [25] DRAGONE, G.; SILVA, T. A. O. e SILVA, J. B. A. **Cerveja**. Em.: Venturini Filho, W. G. (Org.). **Bebidas Alcoólicas: Ciência e tecnologia**, vol.1. São Paulo: Blucher, 2016.
- [26] VENTURINI FILHO, W. G.; CEREDA, M. P. **Cerveja**. Em: Aquarone, E.; Borzani, W.; Scmidell, W. e Lima, U. A. (Org.). **Biotecnologia Industrial**, vol. 4 **Biotecnologia na produção de alimentos**. São Paulo: Edgard Blucher, 2001.
- [27] GOLDAMMER, T. **The Brewers' Handbook**. KVP Publishers: Virginia, 1999. 456p.

- [28] BRUNER, R. F.; CHAN, J. **Brazilian Beer Merger Negotiations: Companhia Cervejaria Brahma, SA**. Darden Business Publishing Cases, 2017.
- [29] KUMAR, R. B. **Mergers and Acquisitions by Anheuser-Busch InBev**. Em: KUMAR, R. B. (Org.). *Wealth Creation in the World's Largest Mergers and Acquisitions, Management for Professionals*. Dubai: Springer, 2019.
- [30] CASTRO, C. M. Minas é a 'Bélgica brasileira'. Belo Horizonte: O Tempo, 2013. Disponível em:< <https://www.otempo.com.br/economia/minas-e-a-belgica-brasileira-1.656508>>. Acesso em: 30 jan. de 2022.
- [31] AYER, F. Dietilenoglicol relacionado ao caso Backer já matou 750 em 10 países. Belo Horizonte: Estado de Minas, 2020. Disponível em:<https://www.em.com.br/app/noticia/gerais/2020/01/19/interna_gerais,1115300/dietilenoglicol-relacionado-ao-caso-backer-ja-matou-750-em-10-paises.shtml>. Acesso em: 30 jan. de 2022.
- [32] MANSUR, R. Caso Backer: mais de 2 anos após mortes por intoxicação, Justiça começa a ouvir testemunhas. G1 Minas, Belo Horizonte. Belo Horizonte: G1 Minas, 2022. Disponível em:< <https://g1.globo.com/mg/minas-gerais/noticia/2022/05/23/caso-backer-mais-de-2-anos-apos-mortes-por-intoxicacao-justica-comeca-a-ouvir-testemunhas.ghtml>>. Acesso em: 12 jul. de 2022.
- [33] ACETO, M.; AGOSTINO, A.; FENOGLIO, G., IDONE, A., GULMINI, M., PICOLLO, M., RICCIARDI, P.; DELANEY, J. K. Characterisation of colourants on illuminated manuscripts by portable fibre optic UV-visible-NIR reflectance spectrophotometry. **Analytical Methods**, 6, 1488-1500, 2014.
- [34] EL-ABASSY, R. M.; ERAVUCHIRA, P. J.; DONFACK, P.; VON DER KAMMER, B.; MATERNY, A.. Fast determination of milk fat content using Raman spectroscopy. **Vibrational Spectroscopy**, 56, 3-8, 2011.
- [35] WETZEL, D. L. Near-infrared reflectance analysis. **Analytical chemistry**, 55, 1165-1176, 1983.
- [36] BOTELHO, B. G.; MENDES, B. A. P.; SENA, M. M. Development and analytical validation of robust near-infrared multivariate calibration models for the quality inspection control of mozzarella cheese. **Food Analytical Methods**, 6, 881-891, 2013.
- [37] RODRIGUEZ-OTERO, J. L.; HERMIDA, M.; CENTENO, J. Analysis of dairy products by near-infrared spectroscopy: A review. **Journal of Agricultural and Food Chemistry**, 45, 2815-2819, 1997.
- [38] DING, H. B.; XU, R. J.. Near-infrared spectroscopic technique for detection of beef hamburger adulteration. **Journal of Agricultural and Food Chemistry**, 48, 2193-2198, 2000.

- [39] RIDGWAY, C.; CHAMBERS, J.; COWE, I. A.. Detection of grain weevils inside single wheat kernels by a very near infrared two-wavelength model. **Journal of Near Infrared Spectroscopy**, 7, 213-221, 1999.
- [40] ALCALÀ, M.; BLANCO, M.; MOYANO, D.I; BROAD, N. W.; O'BRIEN, N.; FRIEDRICH, D.; PFEIFER, F.; SIESLER, H. W. Qualitative and quantitative pharmaceutical analysis with a novel hand-held miniature near infrared spectrometer. **Journal of Near Infrared Spectroscopy**, 21, 445-457, 2013.
- [41] PÉREZ-MARÍN, D.; PAZ, P.; GUERRERO, J. E.; GARRIDO-VARO, A.; SÁNCHEZ, M. T. Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums. **Journal of Food Engineering**, 99, 294-302, 2010.
- [42] FRIEDRICH, D. M.; HULSE, C. A.; VON GUNTEN, . WILLIAMSON, E. P.; PEDERSON, C. G.; O'BRIEN, N. A. Miniature near-infrared spectrometer for point-of-use chemical analysis. **Photonic Instrumentation Engineering**, 8992, 7-17, 2014.
- [43] CORREIA, R. M.; DOMINGOS, E.; CÁO, V. M.; ARAUJO, B. R.; SENA, S.; PINHEIRO, L. U.; FONTES, A. M.; AQUINO, L. F. M.; FERREIRA, E. C.; FILGUEIRAS, P. R.; ROMÃO, W. Portable near infrared spectroscopy applied to fuel quality control. **Talanta**, 176, 26-33, 2018.
- [44] MALEGORI, C.; MARQUES, E. J. N.; FREITAS, S. T.; PIMENTEL, M. F.; PASQUINI, C.; CASIRAGHI, E. Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. **Talanta**, 165, 112-116, 2017.
- [45] BAQUETA, M. R.; COQUEIRO, A.; MARÇO, P. H.; VALDERRAMA, P. Quality control parameters in the roasted coffee industry: a proposal by using MicroNIR spectroscopy and multivariate calibration. **Food Analytical Methods**, 13, 1-11, 2019.
- [46] PAIVA, E. M.; ROHWEDDER, J. J. R.; PASQUINI, C.; PIMENTEL, M. F.; PEREIRA, C. F. Quantification of biodiesel and adulteration with vegetable oils in diesel/biodiesel blends using portable near-infrared spectrometer. **Fuel**, 160, 57-63, 2015.
- [47] SUN, L.; HSIUNG, C.; PEDERSON, C. G.; ZOU, P.; SMITH, V.; VON GUNTEN, M.; O'BRIEN, N. A. Pharmaceutical raw material identification using miniature near-infrared (MicroNIR) spectroscopy and supervised pattern recognition using support vector machine. **Applied Spectroscopy**, 70, 816-825, 2016.
- [48] KOWALSKI, B. Chemometrics. **Analytical Letters**, 11, 1-3, 1978.
- [49] KOWALSKI, B.; BROWN, S.; VANDEGINSTE, B. Editorial. **Journal of Chemometrics**, 1, 1-2, 1987.

- [50] BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. 25 anos de quimiometria no Brasil. **Química Nova**, 29, 1401-1406, 2006.
- [51] MILLER, J.; MILLER, J. **Statistics and chemometrics for analytical chemistry**. Harlow: Pearson Education Limited, 2018.
- [52] RINGNÉR, M. What is principal component analysis? **Nature biotechnology**, 26, 303-304, 2008.
- [53] BRO, R.; SMILDE, A. Principal component analysis. **Analytical Methods**, 6, 2812-2831, 2014.
- [54] WOLD, H. **Partial least squares**. Em: KOTZ, S.; READ, C. B.; BALAKRISHNAN, N.; VIDAKOVIC, B.; JOHNSON, N. L. (Org.). Encyclopedia of statistical sciences. New York: Wiley, 1985.
- [55] SANCHEZ, G. The saga of PLS. Território Britânico do Oceano Índico, 2015. Disponível em: < <https://sagaofpls.github.io> >. Acesso em: outubro de 2022.
- [56] BRERETON, R. G. **Applied chemometrics for scientists**. Chichester: John Wiley & Sons, 2007.
- [57] GOTTFRIES, J.; BLENNOW, K.; WALLIN, A.; GOTTFRIES, C.G. Diagnosis of dementias using partial least squares discriminant analysis. **Dementia and Geriatric Cognitive Disorders**, 6, 83-88, 1995.
- [58] WANG, C.-Y.; CHEN, C.-T.; CHIANG, C.-P.; YOUNG, S.-T.; CHOW, S.-N.; CHIANG, H. K. Partial least-squares discriminant analysis on autofluorescence spectra of oral carcinogenesis. **Applied Spectroscopy**, 52, 1190-1196, 1998.
- [59] ALSBERG, B. K.; KELL, D. B.; GOODACRE, R. Variable selection in discriminant partial least-squares analysis. **Analytical Chemistry**, 70, 4126-4133, 1998.
- [60] GONZÁLEZ-ARJONA, D.; LÓPEZ-PÉREZ, G.; GONZÁLEZ, A. G. Performing procrustes discriminant analysis with HOLMES. **Talanta**, 49, 189-197, 1999.
- [61] BARKER, M.; RAYENS, W. Partial least squares for discrimination. **Journal of Chemometrics**, 17, 166-173, 2003.
- [62] BRERETON, R. G.; LLOYD, G. R. Partial least squares discriminant analysis: taking the magic away. **Journal of Chemometrics**, 28, 213-225, 2014.
- [63] MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Instrução Normativa Nº 54, de 5 de Novembro de 2001. Brasília: Secretaria de Defesa Agropecuária, 2019. Disponível em: < <http://www.agricultura.gov.br/assuntos/vigilancia-agropecuaria/ivegetal/bebidas-arquivos/in-no-54-de-5-de-novembro-de-2001.doc/view> >. Acesso em: outubro de 2022.
- [64] MALHEIRO, F. Mapa identifica contaminação em lotes da Backer feitos antes de outubro de 2019. Belo Horizonte: O Tempo, 2020. Disponível em: < <https://www.otempo.com.br/cidades/mapa-identifica-contaminacao-em-lotes-da> >

[backer-feitos-antes-de-outubro-de-2019-1.2299374](#)>. Acesso em: outubro de 2022.

- [65] CURI, R. A.; VENTURINI FILHO, W. G., DUCATTI, C.; NOJIMOTO, T. Produção de cerveja utilizando cevada e maltose de milho como adjunto de malte: análises físico-química, sensorial e isotópica. **Brazilian Journal of Food Technology**, 11, 279-287, 2008.
- [66] CHRISTENSEN, J.; LADEFOGED, A. M.; NØRGAARD, L. Rapid determination of bitterness in beer using fluorescence spectroscopy and chemometrics. **Journal of the Institute of Brewing**, 111, 3-10, 2005.
- [67] GRANATO, D.; BRANCO, G. F.; FARIA, J. A. F.; CRUZ, A. G. Characterization of Brazilian lager and brown ale beers based on color, phenolic compounds, and antioxidant activity using chemometrics. **Journal of the Science of Food and Agriculture**, 91, 563-571, 2011.
- [68] LACHENMEIER, D. W. Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra. **Food Chemistry**, 101, 825-832, 2007.
- [69] LACHENMEIER, D. W.; FRANK, W.; HUMPFER, E.; SCHÄFER, H.; KELLER, S.; MÖRTTER, M.; SPRAUL, M. Quality control of beer using high-resolution nuclear magnetic resonance spectroscopy and multivariate analysis. **European Food Research and Technology**, 220, 215-221, 2005.
- [70] PRATT, W. K. Image Quantization. Em: PRATT, W. K. **Digital Image Processing**. 3rd. ed. Los Altos, US: John Wiley & Sons, 2001. cap. 6, p. 141-158.
- [71] CAPITÁN-VALLVEY, L. F.; LOPEZ-RUIZ, N.; MARTINEZ-OLMOS, A.; ERENAS, M. M.; PALMA, A. J.. Recent developments in computer vision-based analytical chemistry: A tutorial review. **Analytica Chimica Acta**, 899, 23-56, 2015.
- [72] GELADI, P.; GRAHN H. F. **Multivariate image analysis**. Chichester: John Wiley, 1996.
- [73] LIED, T. T.; GELADI P.; ESBENSEN K. H. Multivariate image regression (MIR): implementation of image PLSR-first forays. **Journal of Chemometrics**, 14, 585-598, 2000.
- [74] OZCAN, A. Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. **Lab on a Chip**, 14, 3187-3194, 2014.
- [75] TÔRRES, A. R.; LYRA, W. S.; ANDRADE, S. I. E.; ANDRADE, R. A. N.; SILVA, E. C.; ARAÚJO, M. C. U.; GAIÃO, E. N.. A digital image-based method for determining of total acidity in red wines using acid–base titration without indicator. **Talanta**, 84, 601-606, 2011.
- [76] LUDWIG, S. K.J.; ZHU, H.; PHILLIPS, S.; SHILEDAR, A.; FENG, S.; TSENG, D.; VAN GINKEL, L. A.; NIELEN, M. W.F.; OZCAN, A. Cellphone-based detection

platform for rbST biomarker analysis in milk extracts using a microsphere fluorescence immunoassay. **Analytical and Bioanalytical Chemistry**, 406, 6857-6866, 2014.

- [77] BOTELHO, B. G.; ASSIS, L. P.; SENA, M. M. Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages. **Food Chemistry**, 159, 175-180, 2014.
- [78] VIDAL, M.; GARCIA-ARRONA, R.; BORDAGARAY, A.; OSTRÁ, M.; ALBIZU, G. Simultaneous determination of color additives tartrazine and allura red in food products by digital image analysis. **Talanta**, 184, 58-64, 2018.
- [79] ACEVEDO, C. A.; SKURTYS, O.; YOUNG, M. E.; ENRIONE, J.; PEDRESCHI, F.; OSORIO, F. A non-destructive digital imaging method to predict immobilized yeast-biomass. **LWT- Food Science and Technology**, 42, 1444-1449, 2009.
- [80] BÖCK, F. C.; HELFER, G. A.; COSTA, A. B.; DESSUY, M. B.; FERRÃO, M. F. Rapid determination of ethanol in sugarcane spirit using partial least squares regression embedded in smartphone. **Food Analytical Methods**, 11, 1951-1957, 2018.
- [81] HELFER, G. A.; TISCHER, B.; FILODA, P. F.; PARCKERT, A. B.; SANTOS, R. B.; VINCIGUERRA, L. L.; FERRÃO, M. F.; BARIN, J. S.; COSTA, A. B. A new tool for interpretation of thermal stability of raw milk by means of the alizarol test using a PLS model on a mobile device. **Food Analytical Methods**, 11, 2022-2028, 2018.
- [82] FERNANDES, D. D. S.; ROMEO, F.; KREPPER, G.; DI NEZIO, M. S.; PISTONESI, M. F.; CENTURIÓN, M. E.; ARAÚJO, M. C. U.; DINIZ, P. H. G. D. Quantification and identification of adulteration in the fat content of chicken hamburgers using digital images and chemometric tools. **LWT - Food Science and Technology**, 100, 20-27, 2019.
- [83] BORIN, A.; FERRÃO, M. F.; MELLO, C.; CORDI, L.; PATACA, L. C.M.; DURÁN, N.; POPPI, R. J. Quantification of *Lactobacillus* in fermented milk by multivariate image analysis with least-squares support-vector machines. **Analytical and Bioanalytical Chemistry**, 387, 1105-1112, 2007.
- [84] ASTM. **Standard Practices for Infrared Multivariate Quantitative Analysis**. West Conshohocken: ASTM International, 2005.
- [85] INMETRO, Instituto Nacional de Metrologia. Orientações sobre validação de métodos de ensaios químicos. Rio de Janeiro: INMETRO, 2020.
- [86] MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Manual da Garantia da Qualidade. Brasília: Secretaria de Defesa Agropecuária, 2015. Disponível em:< <http://www.agricultura.gov.br/assuntos/inspecao/produtos->

vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/manual-de-garantia-da-qualidade-analitica_2015.pdf>. Acesso em: outubro de 2022.

- [87] EUROPEAN COMMISSION. Commission Decision 2002/657/EC of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. **Official Journal of the European Communities**, 50, 8-36, 2002.
- [88] LÓPEZ, M. I.; CALLAO, M. P.; RUISÁNCHEZ, I. A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. **Analytica Chimica Acta**, 891, 62-72, 2015.
- [89] GONDIM, C. S.; COELHO, O. A. M.; ALVARENGA, R. L.; JUNQUEIRA, R. G.; SOUZA, S. V. C. An appropriate and systematized procedure for validating qualitative methods: Its application in the detection of sulfonamide residues in raw milk. **Analytica Chimica Acta**, 830, 11-22, 2014.
- [90] CARVALHO, F. R.; MOORS, P.; WAGEMANS, J.; SPENCE, C. The Influence of Color on the Consumer's Experience of Beer. **Frontiers Psychology**, 19, 2205-2213, 2017.
- [91] SHELLHAMMER, T. H. Beer color. Em: BAMFORTH, Charles W.. (Org.). Beer: A quality perspective. **Handbook of Alcoholic Beverages**. Cambridge: Academic Press, 2009.
- [92] ASBC, American Society of Brewing Chemists. **Tristimulus analysis colorimetric or spectrophotometric**. Em: ASBC (Org.). Methods of Analysis. St. Paul: 2015.
- [93] BLANCO, C. A.; DE LA FUENTE, R.; CABALLERO, I.; RODRÍGUEZ-MÉNDEZ, M. L. Beer discrimination using a portable electronic tongue based on screen-printed electrodes. **Journal of Food Engineering**, 157, 57-62, 2015.
- [94] FENGXIA, S.; YUWEN, C.; ZHANMING, Z.; YIFENG, Y. Determination of Beer Color Using Image Analysis. **Journal of the American Society of Brewing Chemists**, 62, 163-167, 2004.
- [95] KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, 11, 137-148, 1969.
- [96] BOTELHO, B. G.; REIS, N.; OLIVEIRA, L. S.; SENA, M. M. Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. **Food Chemistry**, 181, 31-37, 2015.
- [97] MAPA, Ministério da Agricultura, Pecuária e Abastecimento. Manual da Garantia da Qualidade. Brasília: Secretaria de Defesa Agropecuária, 2015. Disponível em:< <http://www.agricultura.gov.br/assuntos/inspecao/produtos->

[vegetal/legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/manual-de-garantia-da-qualidade-analitica_2015.pdf](https://legislacao-1/biblioteca-de-normas-vinhos-e-bebidas/manual-de-garantia-da-qualidade-analitica_2015.pdf)>. Acesso em: novembro de 2022.

- [98] WILLIAMS, P. C.; SOBERING, D. C. N. Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. **Journal of Near Infrared Spectroscopy**, 1, 25-32, 1993.
- [99] WILLIAMS, P. **Implementation of Near-infrared Technology**. Em: WILLIAMS, P.; NORRIS, K. H. (Org.). Near-infrared Technology in the Agricultural and Food Industries. Madison: American Association of Cereal Chemists, 2001.
- [100] SOUZA, S. V. C.; JUNQUEIRA, R. G. A procedure to assess linearity by ordinary least squares method. **Analytica Chimica Acta**, 552, 25-35, 2005.
- [101] DURBIN, J.; WATSON, G. S. Testing for serial correlation in least squares regression: I. **Biometrika**, 38, 159-178, 1951.
- [102] THOMPSON, M.; ELLISON, S. L. R.; WOOD, R. Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). **Pure and Applied Chemistry**, 74, 835-855, 2002.
- [103] TÖRÖK, A.; SZERLETICS, A.; JANTYIK, L. Factors influencing competitiveness in the global beer trade. **Sustainability**, 12, 5957, 2020.
- [104] WACHEŁKO, O., SZPOT, P.; ZAWADZKI, M. The application of headspace gas chromatographic method for the determination of ethyl alcohol in craft beers, wines and soft drinks. **Food Chemistry**, 346, 128924, 2021.
- [105] BREWERS ASSOCIATION (USA). National Beer Sales & Production Data. The New Brewer, Boulder, USA, maio 2022. Disponível em: <https://www.brewersassociation.org/statistics-anddata/national-beer-stats/>. Acesso em: novembro de 2022.
- [106] KRAFTCHICK, J. F.; BYRD, E. T.; CANZIANI, B. & GLADWELL, N. J. Understanding beer tourist motivation. **Tourism Management Perspectives**, 12, 41-47, 2014.
- [107] DONADINI, G.; FUMI, M. D.; KORDIALIK-BOGACKA, E.; MAGGI, L.; LAMBRI, M.; SCKOKAI, P. Consumer interest in specialty beers in three European markets. **Food Research International**, 85, 301-314, 2016.
- [108] FULGÊNCIO, A. C. C.; ARAÚJO, V. P. T.; PEREIRA, H. V.; BOTELHO, B. G.; SENA, M. M. Development of a simple and rapid method for color determination in beers using digital images. **Food Analytical Methods**, 13, 303-312, 2019.
- [109] LVOVA, L.; PAOLESSE, R.; DI NATALE, C.; D'AMICO, A. Detection of alcohols in beverages: An application of porphyrin-based electronic tongue. **Sensors Actuators B**, 118, 439-447, 2006.

- [110] ADHIKARI, B. R.; SCHRAFT, H.; CHEN, A. A high-performance enzyme entrapment platform facilitated by a cationic polymer for the efficient electrochemical sensing of ethanol. **Analyst**, 142, 2595–2602, 2017.
- [111] CINTI, S.; BASSO, M.; MOSCONE, D.; ARDUINI, F. A paper-based nanomodified electrochemical biosensor for ethanol detection in beers. **Analytica Chimica Acta**, 960, 123-130, 2017.
- [112] VOSS, H. G. J.; MENDES Jr., J. J. A.; FARINELLI, M. E.; STEVAN Jr., S. L. A prototype to detect the alcohol content of beers based on an electronic nose. **Sensors**, 19, 2646, 2019.
- [113] ERFKAMP, J.; GUENTHER, M.; GERLACH, G. Hydrogel-based sensors for ethanol detection in alcoholic beverages. **Sensors**, 19, 1199, 2019.
- [114] KANEKO, T.; FURUSHO, S.; GANAHA, R.; INUI, T.; MATSUYAMA, A.; MIZUNO, A.; MORIMOTO, M.; TAKEMOTO, K. Comparison of the Anton Paar Alcozyzer method and the official GC-FID method of the National Tax Administration Agency Japan for the evaluation of alcohol content in beer, happo-shu, and nonalcoholic beer. **Journal of the American Society of Brewing Chemists**, 65, 246-247, 2007.
- [115] CLONINGER, L. Alcohol determination of malt-based beverages by rapid distillation. **Journal of the American Society of Brewing Chemists**, 76, 21-23, 2018.
- [116] ABASI, S.; MINAEI, S.; JAMSHIDI, B.; FATHI, D. Dedicated non-destructive devices for food quality measurement: A review. **Trends in Food Science and Technology**, 78, 197-205, 2018.
- [117] CROCOMBE, R. A. Portable spectroscopy. **Applied Spectroscopy**, 72, 1701-1751, 2018.
- [118] SANTOS, C. A. T.; LOPO, M.; PÁSCOA, R. N. M. J.; LOPES, J. A. A review on the applications of portable near-infrared spectrometers in the agro-food industry. **Applied Spectroscopy**, 67, 1215-1233, 2013.
- [119] YU, H.; LIU, H.; WANG, Q.; VAN RUTH, S. Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. **LWT**, 128, 109398, 2020.
- [120] SANTOS, C. S. P.; CRUZ, R.; GONCALVES, D. B.; QUEIROS, R.; BLOORE, M.; KOVACS, Z.; HOFMANN, I.; CASAL, S. Non-destructive measurement of the internal quality of citrus fruits using a portable NIR device. **AOAC International**, 104, 61-67, 2021.
- [121] ANYIDOHO, E. K.; TEYE, E.; AGBEMAFE, R.; AMUAH, C. L. Y.; BOADU, V. G. Application of portable near infrared spectroscopy for classifying and quantifying

cocoa bean quality parameters. **Journal of Food Processing and Preservation**, 1, e15445, 2021.

- [122] MUÑIZ, R.; CUEVAS-VALDÉS, M.; ROZA-DELGADO, B. Milk quality control requirement evaluation using a handheld near infrared reflectance spectrophotometer and a bespoke mobile application. **Journal of Food Composition and Analysis**, 86, 103388, 2020.
- [123] CAMPS, C.; CHRISTEN, D. Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy. **LWT**, 42, 1125-1131, 2009.
- [124] LI, H.; TAKAHASHI, Y.; KUMAGAI, M.; FUJIWARA, K.; KIKUCHI, R.; YOSHIMURA, N.; AMANO, T.; LIN, J.; OGAWA, N. A chemometrics approach for distinguishing between beers using near infrared spectroscopy. **Journal of Near Infrared Spectroscopy**, 17, 69-76, 2009.
- [125] BLITZ, J. P. Diffuse reflectance spectroscopy. Em: MIRABELLA, F. M. (ed.). **Modern techniques in applied molecular spectroscopy**. New York: Wiley, 1998. cap. 4, p.185–219.
- [126] AOAC, Association of Official Agricultural Chemists. Appendix F: Guidelines for standard method performance requirements. Em: AOAC (Org.). **Official Methods of Analysis**, 2012.
- [127] FABER, N. M.; SONG, X.; HOPKE, P. K.. Sample-specific standard error of prediction for partial least squares regression. **TrAC - Trends in Analytical Chemistry**, 22, 330-334, 2003.
- [128] NIRSYSTEMS, Metrohm. **A guide to near-infrared spectroscopic analysis of industrial manufacturing processes**. United States: Silver Spring, 2002.
- [129] RINNAN, Å.; van den BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **Trends in Analytical Chemistry**, 28, 1201-1222, 2009.
- [130] NORDON, A.; MILLS, A.; BURN, R. T.; CUSUCK, F. M.; LITTLEJOHN, D. Comparison of non-invasive NIR and Raman spectrometries for determination of alcohol content of spirits. **Analytica Chimica Acta**, 548, 148-158, 2005.
- [131] RATANAWIMARNWONG, N.; SINPUN, M.; CHANKAW, P.; CHOENGCHAN, N. Simple flow system with in-line gas-diffusion unit for determination of ethanol employing hypsochromic shift of visible absorbance band of methyl orange. **Talanta**, 206, 120234, 2020.
- [132] SISCO, E.; ROBINSON, E. L. Determination of ethanol concentration in alcoholic beverages by direct analysis in real time mass spectrometry (DART-MS). **Forensic Chemistry**, 18, 100219, 2020.
- [133] EUROPEAN COMMISSION. Excise duty on alcohol - Taxation and customs union. Bruxelas, 2020. Disponível em: < <https://taxation->

customs.ec.europa.eu/taxation-1/excise-duties/excise-duty-alcohol_en>.

Acesso em: novembro de 2022.

- [134] DEPARTMENT OF TAXATION AND FINANCE OF NEW YORK STATE. Alcoholic beverages tax. New York: Department of Taxation and Finance, 2022. Disponível em:< <https://www.tax.ny.gov/bus/bev/abt.htm#beer>>. Acesso em: novembro de 2022.
- [135] OFFICIAL JOURNAL OF THE EUROPEAN UNION. Regulation (EU) No 1168/2011 of the European Parliament and of the Council of 25 October 2011, 1-17. Disponível em:< <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2011:304:0001:0017:EN:PDF>>. Acesso em: novembro de 2022.
- [136] EUROPEAN COMMISSION. Labelling of alcoholic beverages in the EU: some facts. Bruxelas, 2020. Disponível em:< https://food.ec.europa.eu/safety/labelling-and-nutrition/food-information-consumers-legislation/alcohol-labelling_en>.
- Acesso em: novembro de 2022.
- [137] EUROPEAN COMMISSION. Food information to consumers - legislation. Bruxelas, 2020. Disponível em:< https://food.ec.europa.eu/safety/labelling-and-nutrition/food-information-consumers-legislation_en>. Acesso em: novembro de 2022.
- [138] WORLD HEALTH ORGANIZATION. Global status report on alcohol and health 2018. Disponível em:< <https://www.who.int/publications/i/item/9789241565639>>.
- Acesso em: novembro de 2022.
- [139] STUPAK, M.; KOCOUREK, V.; KOLOUCHOVA, I.; HAJŠLOVA, J. Rapid approach for the determination of alcoholic strength and overall quality check of various spirit drinks and wines using GC–MS. **Food Control**, 80, 307-313, 2017.
- [140] YARITA, T.; NAKAJIMA, R.; OTSUKA, S.; IHARA, T.; TAKATSU, A.; SHIBUKAWA, M. Determination of ethanol in alcoholic beverages by high-performance liquid chromatography–flame ionization detection using pure water as mobile phase. **Journal of Chromatography A**, 976, 387-391, 2002.
- [141] PAIXÃO, T. R. L. C.; CORBO, D.; BERTOTTI, M. Amperometric determination of ethanol in beverages at copper electrodes in alkaline medium. **Analytica Chimica Acta**, 472, 123-131, 2002.
- [142] NAKAMURA, H.; TANAKA, R.; SUZUKI, K.; YATAKA, M.; MOGI, Y. A direct determination method for ethanol concentrations in alcoholic beverages employing a eukaryote double-mediator system. **Food Chemistry**, 117, 509-513, 2009.
- [143] PINHEIRO, A. C. N.; FERREIRA, V. S.; LUCCA, B. G. Stamping method based on 3D printing and disposable napkin: Cheap production of paper analytical

devices for alcohol determination in beverages aiming forensics and food control. **Microchemical Journal**, 180, 107604, 2022.

- [144] SOMBOON, T.; SANSUK, S. An instrument-free method based on visible chemical waves for quantifying the ethanol content in alcoholic beverages. **Food Chemistry**, 253, 300-304, 2018.
- [145] SANSUK, S.; JUNTARAKOD, P.; TONGPHOOTHORN, W.; SIRIMUNGKALA, A.; SOMBOON, T. Visual chemo-chronometric assay for quantifying ethanol in alcoholic drinks by the colorimetric Belousov-Zhabotinsky oscillator. **Food Control**, 110, 107042, 2020.
- [146] FILGUEIRAS, M. F.; LIMA, B. d. O.; BORGES, E. M. A high-throughput, cheap, and green method for determination of ethanol in cachaça and vodka using 96-well-plate images. **Talanta**, 241, 123229, 2022.
- [147] ZURIARRAIN, A.; ZURIARRAIN, J.; VILLAR, M.; BERREGI, I. Quantitative determination of ethanol in cider by ¹H NMR spectrometry. **Food Control**, 50, 758-762, 2015.
- [148] GAŁUSZKA, A.; MIGASZEWSKI, Z. M.; NAMIEŚNIK, J. Moving your laboratories to the field – Advantages and limitations of the use of field portable instruments in environmental sample analysis. **Environmental Research**, 140, 593-603, 2015.
- [149] REZENDE, K. C. A.; DUARTE, L. M.; PINHEIRO, K. M. P.; CARDOSO, T. M. G.; NOGUEIRA, S. A.; COLTRO, W. K. T. Portable Analytical Platforms Associated with Chemometrics for Rapid Screening of Whisky Adulteration. **Food Analytical Methods**, 15, 2451–2461, 2022.
- [150] MCVEY, C.; ELLIOTT, C. T.; CANNAVAN, A.; KELLY, S. D.; PETCHKONGKAEW, A.; HAUGHEY, S. A. Portable spectroscopy for high throughput food authenticity screening: Advancements in technology and integration into digital traceability systems. **Trends in Food Science & Technology**, 118, 777-790, 2021.
- [151] ZHU, C.; FU, X.; ZHANG, J.; QIN, K.; WU, C. Review of portable near infrared spectrometers: Current status and new techniques. **Journal of Near Infrared Spectroscopy**, 30, 51-66, 2022.
- [152] FULGÊNCIO, A. C. C.; RESENDE, G. A. P.; TEIXEIRA, M. C. F.; BOTELHO, B. G.; SENA, M. M. Determination of Alcohol Content in Beers of Different Styles Based on Portable Near-Infrared Spectroscopy and Multivariate Calibration. **Food Analytical Methods**, 15, 307–316, 2022.
- [153] MICKLANDER, E.; KJELDAHL, K.; EGEBO, M.; NØRGAARD, L. Multi-product Calibration Models of near Infrared Spectra of Foods. **Journal of Near Infrared Spectroscopy**, 14, 395-402, 2006.

- [154] FAN, X.; TANG, S.; LI, G.; ZHOU, X. Non-Invasive Detection of Protein Content in Several Types of Plant Feed Materials Using a Hybrid Near Infrared Spectroscopy Model. **Plos One**, 11, e0163145, 2016.
- [155] MASITHOH, R. E.; LOHUMI, S.; YOON, W.; AMANAH, H. Z.; CHO, B. Development of multi-product calibration models of various root and tuber powders by Fourier transform near infra-red (FT-NIR) spectroscopy for the quantification of polysaccharide contents. **Heliyon**, 6, e05099, 2020.
- [156] RAMBO, M. K. D.; FERREIRA, M. M. C.; AMORIM, E. P. Multi-product calibration models using NIR spectroscopy. **Chemometrics and Intelligent Laboratory Systems**, 151, 108-114, 2016.
- [157] SANTOS, D. A.; LIMA, K.; CAVALCANTE, V.; COQUEIRO, A.; CONSOLIN, M.; CONSOLIN FILHO, N.; MARÇO, P. H.; VALDERRAMA, P. Multiproduct, Multicomponent and Multivariate Calibration: A Case Study by Using Vis-NIR Spectroscopy. **Food Analytical Methods**, 11, 1915-1919, 2018.
- [158] GONZÁLEZ-CABALLER, V.; PÉREZ-MARÍN, D.; LÓPEZ, M.; SÁNCHEZ, M. Optimization of NIR spectral data management for quality control of grape bunches during on-vine ripening. **Sensors**, 11, 6109-6124, 2011.
- [159] ZHENG, Y.; LAI, X.; BRUUN, S. W.; IPSEN, H.; LARSEN, J. N.; LØWENSTEIN, H.; SØNDERGAARD, I.; JACOBSEN, S. Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy. **Journal of Pharmaceutical and Biomedical Analysis**, 46, 592-596, 2008.
- [160] FERREIRA, R. A.; TEIXEIRA, G.; PETERNELLI, L. A. Kennard-Stone method outperforms the Random Sampling in the selection of calibration samples in SNPs and NIR data. **Ciência Rural**, 52, e20201072, 2021.
- [161] FULGÊNCIO, A. C. C.; RESENDE, G. A. P.; TEIXEIRA, M. C. F.; BOTELHO, B. G.; SENA, M. M.. Screening method for the rapid detection of diethylene glycol in beer based on chemometrics and portable near-infrared spectroscopy. **Food Chemistry**, 391, 133258, 2022.
- [162] STUART, B. H. **Infrared Spectroscopy: Fundamentals and Applications**. John Wiley & Sons, 2004.
- [163] PROSEKOV, A. Y.; IVANOVA, S. A. Food security: The challenge of the present. **Geoforum**, 91, 73-77, 2018.
- [164] SILEONI, V.; MARCONI, O.; PERRETTI, G. Near-infrared spectroscopy in the brewing industry. **Critical reviews in food science and nutrition**, 55, 1771-1791, 2015.
- [165] FOX, G. The brewing industry and the opportunities for real-time quality analysis using infrared spectroscopy. **Applied Sciences**, 10, 616, 2020.

- [166] GOULART, C. O. L.; BORDONI, L. S.; NASCENTES, C. C.; COSTA, L. M. Analysis of Diglycolic Acid after Mass Poisoning by Diethylene Glycol. **Journal of analytical toxicology**, 46, 64-68, 2020.
- [167] SANCHEZ, Z. M.; OLIVEIRA, C. L.; LARANJEIRA, R.; CAETANO, R. Intoxication and Deaths from Diethylene Glycol Present in Brazilian Craft Beers. **Journal of Studies on Alcohol and Drugs**, 81, 522-523, 2020.
- [168] LIMA, C. M. D.; BRAGA, L.; VENTURA, C. P.; GOULART, C. O. L. Diethylene glycol poisoning: A 1937 world retrospective study – From 1937 to the 2020 brewery case in Minas Gerais, Brazil. **Revista de Criminalística e Medicina Legal**, 6, 27-33, 2020.
- [169] SCHEP, L. J.; SLAUGHTER, R. J.; TEMPLE, W. A.; BEASLEY, M. G. Diethylene glycol poisoning. **Clinical toxicology**, 47, 525-535, 2009.
- [170] LANDRY, G. M.; MARTIN, S.; MCMARTIN, K. E. Diglycolic acid is the nephrotoxic metabolite in diethylene glycol poisoning inducing necrosis in human proximal tubule cells in vitro. **Toxicological Sciences**, 124, 35-44, 2011.
- [171] CALDEIRA, L. R.; MADUREIRA, F. D.; MAIA, T. D. F.; MULLER, C. V.; FERNANDES, C. Simultaneous quantification of ethylene glycol and diethylene glycol in beer by gas chromatography coupled to mass spectrometry. **Food Chemistry**, 346, 128871, 2021.
- [172] LAWRENCE, J. F.; CHADHA, R. K.; LAU, B. P. Y.; WEBER, D. F. Simplified routine method for the determination of diethylene glycol in wines by capillary gas chromatography with flame ionization detection. **Journal of Chromatography**, 367, 213-216, 1986.
- [173] BAFFI, P.; ELNESER, S.; BAFFI, M.; MELIN, M. D. Quantitative determination of diethylene glycol contamination in pharmaceutical products. **Journal of AOAC International**, 83, 793-801, 2000.
- [174] AHMED, M. K.; MCLEOD, M. P.; NÉZIVAR, J.; GIULIANI, A. W. Fourier transform infrared and near-infrared spectroscopic methods for the detection of toxic diethylene glycol (DEG) contaminant in glycerin-based cough syrup. **Journal of Spectroscopy**, 24, 601-608, 2010.
- [175] GRYNIEWICZ-RUZICKA, C. M.; ARZHANTSEV, S.; PELSTER, L. N.; WESTENBERGER, B. J.; BUHSE, L. F.; KAUFFMAN, J. F. Multivariate calibration and instrument standardization for the rapid detection of diethylene glycol in glycerin by Raman spectroscopy. **Applied Spectroscopy**, 63, 334-341, 2011.
- [176] HOLLOWAY, G.; MAHESWARAN, R.; LEEKS, A.; BRADBY, S.; WAHAB, S. Screening method for ethylene glycol and diethylene glycol in glycerin-containing

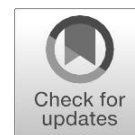
- products. **Journal of Pharmaceutical and Biomedical Analysis**, 51, 507-511, 2010.
- [177] RAHIM, A. A.; SAAD, B.; OSMAN, H.; HASHIM, N.; YAHYA, S.; TALIB, K. M. Simultaneous determination of diethylene glycol, diethylene glycol monoethyl ether, coumarin and caffeine in food items by gas chromatography. **Food Chemistry**, 126, 1412-1416, 2011.
- [178] HERNÁNDEZ, F.; IBÁÑEZ, M.; SANCHO, J. V. Fast determination of toxic diethylene glycol in toothpaste by ultra-performance liquid chromatography–time of flight mass spectrometry. **Analytical and Bioanalytical Chemistry**, 391, 1021-1027, 2008.
- [179] MAURER, H. H.; PETERS, F. T.; PAUL, L. D.; KRAEMER, T. Validated gas chromatographic–mass spectrometric assay for determination of the antifreezes ethylene glycol and diethylene glycol in human plasma after microwave-assisted pivalylation. **Journal of Chromatography B**, 754, 401-409, 2001.
- [180] LÓPEZ-SÁNCHEZ, M.; DOMÍNGUEZ-VIDAL, A.; AYORA-CAÑADA, M. J.; MOLINA-DÍAZ, A. Assessment of dentifrice adulteration with diethylene glycol by means of ATR-FTIR spectroscopy and chemometrics. **Analytica Chimica Acta**, 620, 113-119, 2008.
- [181] LI, X.; ARZHANTSEV, S.; KAUFFMAN, J. F.; SPENCER, J. A. Detection of diethylene glycol adulteration in propylene glycol - Method validation through a multiinstrument collaborative study. **Journal of Pharmaceutical and Biomedical Analysis**, 54, 1001-1006, 2011.
- [182] YIN, L.; ZHANG, X.; LI, X.; JIN, S. Development and implementation of incident response near-infrared models for analyzing contaminated medicines containing diethylene glycol solvent. **Journal of Innovative Optical Health Sciences**, 7, 1450035, 2014.
- [183] LANGTON, S.; CHEVENNEMENT, R.; NAGELKERKE, N.; LOMBARD, B. Analyzing collaborative trials for qualitative microbiological methods: Accordance and concordance. **International Journal of Food Microbiology**, 79, 175-181, 2002.
- [184] ISABEL LÓPEZ, M.; PILAR CALLAO, M.; RUISÁNCHEZ, I. A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. **Analytica Chimica Acta**, 891, 62-72, 2015.
- [185] GONDIM, C. S.; JUNQUEIRA, R. G.; SOUZA, S. V. C.; PILAR CALLAO, M.; RUISÁNCHEZ, I. Determining performance parameters in qualitative multivariate methods using probability of detection (POD) curves. Case study: Two common milk adulterants. **Talanta**, 168, 23-30, 2017.
- [186] EURACHEM. The Fitness for Purpose of Analytical Methods Food information to consumers. UK, 2014. Disponível em:<

https://www.eurachem.org/images/stories/Guides/pdf/MV_guide_2nd_ed_EN.pdf>. Acesso em: março de 2023.

- [187] EUROPEAN COMMISSION. Commission Implementing Regulation (EU) 2021/808 of 22 March 2021 on the performance of analytical methods for residues of pharmacologically active substances used in food-producing animals and on the interpretation of results as well as on the methods to be used for sampling and repealing Decisions 2002/657/EC and 98/179/EC. Bruxelas, 2021. Disponível em:<<https://op.europa.eu/en/publication-detail/-/publication/3dc2b06b-b9cf-11eb-8aca-01aa75ed71a1/language-en>>. Acesso em: março de 2023.
- [188] IUPAC. Validation of Qualitative and Semi-Quantitative (Screening) Methods by Collaborative Trial, **Chemistry International - Newsmagazine for IUPAC**, 28, 29-29, 2006.
- [189] MARCATHUR, R.; VON HOLSTB, C. A protocol for the validation of qualitative methods of detection. **Analytical Methods**, 4, 2744-2754, 2012.
- [190] HORWITZ, W. Protocol for the design, conduct and interpretation of method performance studies. **Pure and Applied Chemistry**, 67, 331-343, 1995.

APÊNDICE A

Food Analytical Methods (2020) 13:303–312
<https://doi.org/10.1007/s12161-019-01634-0>



Development of a Simple and Rapid Method for Color Determination in Beers Using Digital Images

Ana Carolina C. Fulgêncio¹ · Vinícius P. T. Araújo¹ · Hebert V. Pereira¹ · Bruno G. Botelho¹ · Marcelo M. Sena^{1,2}

Received: 2 April 2019 / Accepted: 14 August 2019 / Published online: 17 October 2019
 © Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Color is an important sensory parameter required for the quality control of beers. A new multivariate image analysis method for the color determination of beers was proposed and validated. Reference color values were determined using the SRM (standard reference method) system, which is based on absorbance measurements at 430 nm. Digital images were obtained with an iPhone 7 smartphone. The obtained RGB histograms were used for building partial least squares (PLS) models. The developed method is direct, simple, and rapid, not requiring sample pretreatment steps as the reference method. Beer samples of different styles, brands, and brewery companies were obtained in a large variety, totalizing 128 samples and comprising a range from 3 to 130 SRM units. A global PLS model built with all the beer samples presented too large prediction errors for some samples in the lower part of the SRM scale (below 12 units). Thus, considering the requirement of dilution prescribed by the reference method for samples with absorbances higher than 1.0, two local calibration models were built: for high SRM range (above 12 units) and low SRM range (equal or below 12 units) samples. A previous PLS discriminant analysis (PLS-DA) model was used to assign the beer samples to these two classes, resulting in 78 and 50 samples in the high- and low-range models, respectively. These two models were validated according to the Brazilian and international guidelines, being considered linear, accurate, precise, and unbiased. Uncertainties were also calculated for estimating confidence intervals for the predictions of the validation samples. The developed method could be easily adapted in a mobile platform, spreading its use and opening the possibility for the commercial production of a dedicated equipment.

Keywords RGB histograms · Image analysis · PLS · Multivariate calibration · Beer quality control · Chemometrics

Introduction

Beer is defined as an alcoholic beverage produced through the fermentation of starch-containing ingredients (Siqueira et al. 2008; EBlinger 2009). It can be found almost everywhere and has been consumed and produced in many countries since antiquity, especially in Asia and Europe. In 2013, Brazil was already the third largest beer producer in the world, having produced nearly 13.5 million kiloliters in that year. Pilsner

beers (American standard lagers) are the most consumed in Brazil. This style has light yellow color and the majority of the Brazilian beer market is dominated by one of the world's largest brewery companies, owner of the most consumed brands (Lapere 2015). According to a survey carried out by the Ministry of Agriculture, Livestock and Food Supply (MAPA for its initials in Portuguese), there has been an exponential growth of the brewing market in Brazil as of 2010. Only in 2018, 210 new breweries began to operate, representing a growth of 30% over 2017, the highest in Brazilian history (Cilo 2019). Most of these new producers are craft breweries.

Even with all the current technology, brewing beer is not a simple process considering the critical sequence of steps demanded to produce the final product. It is necessary to pre-determine the type of beer to be produced in order to define its sensorial characteristics, such as color, aroma, and bitterness. These characteristics should be obtained from the selection of malts, hops, and yeast (Huige 2006). As with any industrialized product, the quality control of the finished beverage is

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12161-019-01634-0>) contains supplementary material, which is available to authorized users.

* Marcelo M. Sena
 marcsen@ufmg.br

¹ Departamento de Química, ICEx, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil

² Instituto Nacional de Ciência e Tecnologia em Bioanálítica (INCT Bio), Campinas, SP 13083-970, Brazil

also fundamental in the brewing industry (Curi et al. 2008). The development of new analytical methods and the rapid assessment of the results for the quality control are of great importance as tools to improve beer production techniques. In addition, one of the current main goals has been to reduce the response time by simplifying the analytical procedure without compromising the results. For this task, chemometric methods have gained relevance in the last years. These multivariate methods usually rely on data obtained mainly by spectrophotometers. However, some low-cost alternatives have recently emerged, based on portable equipment used to obtain digital images.

Among the most important analyses used for assessing the quality of beers, the determination of color should be highlighted. Color is the first sensory characteristic of beers that draws the attention of the consumer (Carvalho et al. 2017). Color determination is performed according to the two currently accepted systems, EBC (European brewing convention) and SRM (standard reference method), which are both based on spectrophotometric absorption measurements at 430 nm (Shellhammer 2009; ASBC 2015).

Chemometric tools have been employed for the quality control of beers, and some articles have tried to quantify color. Fourier transform infrared spectroscopy (FTIR) has been combined with partial least squares (PLS) for predicting nine parameters of beers, including relative density, alcohol content, pH, bitterness, and EBC color (Lachenmeier 2007). Nevertheless, the calibration model for color presented low accuracy and low linearity as recognized by the own author. A portable electronic tongue based on screen-printed electrodes has been used to discriminate beer styles and to quantify color and alcohol strength (Blanco et al. 2015). In this paper, PLS provided a good linearity, but at the expense of using a small number of samples (25) and not validating the model with an independent test set. RoboBEER, a robotic pourer coupled with computer vision algorithm, has been combined with near infrared spectroscopy (NIRS) for the quality control of beers through the estimated fifteen parameters, though not including color (Viejo et al. 2018). Colorimetric sensor arrays have been used to build principal component analysis (PCA) and hierarchical cluster analysis (HCA) models to discriminate beer styles (Zhang et al. 2006). More specifically, a table scanner and a CCD camera have been employed to determine EBC color of beers through a univariate model based on the saturation value, a parameter estimated from the RGB color information (Fengxia et al. 2004). A drawback of this article was its very narrow working range of EBC values (ranging from 4 to 11, equivalent to 2–6 SRM units) since only lager beers of light colors were used to build the model.

The application of chemometrics to the processing of digital images originated multivariate image analysis (MIA) and multivariate image regression (MIR) (Geladi and Grahn 1996;

Lied et al. 2000). During the last 15 years, chemical analysis based on color changes obtained with imaging devices has grown increasingly. This notable trend has brought about simple, direct, rapid, in situ, and low-cost new analytical methods based on computer vision-based analytical (CVAC) procedures (Capitan-Vallvey et al. 2015). The great diffusion of portable equipments for the acquisition of digital images has promoted the use of color to extract qualitative and quantitative chemical information and improved the democratization of chemical science (Ozcam 2014). Digital cameras, scanners, and smartphones equipped with high-resolution cameras are available at very low prices compared to the traditional analytical instrumentation. As a consequence, analytical methods have been developed based on the concepts of lab-on-a-chip or point-of-care testing and applied to food analysis or as biosensors for clinical analysis (Preechaburana et al. 2014). Even cheap portable homemade devices containing built-in smartphones have been developed, such as a light-tight box for glucose biosensing (Chun et al. 2014) or an artisanal dark chamber for the quality control of hard candies (Botelho et al. 2017).

Processing strategies for digital image analysis can vary widely, from simple univariate regressions to multivariate and even the more complex multi-way models. When the image detection provides a selective signal, there is no need of chemometrics and univariate models can be used. Some examples include a digital image-based method for determining total acidity in red wines using acid–base titration without indicator (Torres et al. 2011) and a cell phone-based detection platform to determine biomarkers in milk aiming to detect contamination (Ludwig et al. 2014). When the matrix is more complex and/or there are wider analytical ranges of analytes and interferences, chemometrics is necessary. One of the most popular analytical strategies for MIA has been to model RGB (red, green, and blue) histograms by multivariate calibration, mainly using PLS. These histograms can be modeled in a similar way as the traditional spectroscopic data. This chemometric strategy has been applied in the analysis of food and beverages, such as in the quantification of artificial dyes in beverages (Botelho et al. 2014; Vidal et al. 2018), in the prediction of immobilized yeast-biomass during a food fermentation process (Acevedo et al. 2009), in the rapid determination of ethanol in sugarcane spirits (Böck et al. 2018), in the evaluation of thermal stability of raw milk (Helfer et al. 2018), and in the detection of adulterations in the fat content of chicken hamburgers (Fernandes et al. 2019). More complex chemical matrices can demand more sophisticated strategies, such as the building of non-linear models with support vector machines (SVM) (Borin et al. 2007) or the combination of a fast Fourier transform and multi-way PLS (N-PLS) for processing non-homogeneous images containing textural variations

(Huang et al. 2003; Godinho et al. 2010; Botelho et al. 2017).

The primary objective of this paper was to develop and validate a multivariate calibration method based on digital images recorded by a mobile phone coupled with PLS for the color determination of beers. Reference values were based on the SRM system and obtained by visible spectrophotometry. In truth, SRM and EBC systems differ only on the scale, being both based on the same experimental procedure. SRM was chosen for this work because it is more used in Brazil and America in general, while EBC predominates in Europe. In order to build a robust model, a great variety of beers, from different styles, brands, and breweries, was incorporated in the model, corresponding to a wide range of SRM units (3 to 130). The proposed multivariate method was also validated in accordance with the Brazilian and international guidelines (Botelho et al. 2013), aiming to assure its reliability and fit for the intended use.

Materials and Methods

Instruments and Software

All the images were recorded using a cellphone iPhone 7 (Apple, Cupertino, USA) with 128 GB of capacity, equipped with a camera of 12 MP, and a sensor Sony Exmor RS model. Data were processed using MATLAB software, version 7.13 (MathWorks, Natick, USA), along with the PLS Toolbox, version 6.5 (Eigenvector Technologies, Manson, USA), and the Image Processing Toolbox, version 8.0 (MathWorks). An Agilent Cary 60 UV/vis spectrophotometer (Agilent Technologies, Santa Clara, USA) was used for obtaining the measurements for the reference method.

Samples

One hundred twenty-eight different samples of beers were acquired mainly in the local commerce of Belo Horizonte, MG, Brazil. Samples were stored at room temperature until the analysis. Beer samples were obtained from a large variety of brands, breweries, and styles, including lager, weissbier, wit, IPA, bock, red ale, brown ale, dunkel, stout, porter, etc.

Procedure

Approximately 300 ml of each sample were collected and degassed using an ultrasonic bath for 5 min. A few drops of 1-octanol (Neon, Suzano, Brazil) were previously added to each sample as an antifoaming agent. In the sequence, 25 ml of each sample was transferred to a small Petri dish (3.0 cm radius \times 1.5 cm height) embedded in a wooden dark chamber specifically developed by the authors for image acquisition

(Fig. 1S, Supplemental material). All the images were obtained with 4032 \times 3024 pixels (12 MB) in “.jpg” format, with 16.8 million colors and 300 dpi resolution. The conversion of the images in RBG histograms was carried out in MATLAB. Firstly, a 300 \times 300 pixels frozen area was selected from the central part of the dish, in the homogeneous portion of the image. This selected area was decomposed in an RGB histogram containing 768 channels (256 for each RGB color) for each sample. Each sample was scanned in triplicate and its average histogram was used for building PLS models.

Spectrophotometric Analysis (Reference Method)

Spectrophotometric analysis was performed according to the official ASBC (American Society of Brewing Chemists) method Beer-10 (ASBC 2015). This method is based on absorbance measurements at 430 nm of the samples contained in a quartz cuvette of 10-mm optical path. Beer samples not free of turbidity were clarified by centrifugation for 5 min, at 3000 rpm, performed in an Awel MF 20-R centrifuge (Awel SAS, Blain, France). Dark beers needed to be diluted, and samples with absorbances higher than 1.0 were diluted in the ratio 1:10 (v/v) using deionized water.

Multivariate Models and Analytical Validation

Some attempts to build univariate models based on the frequency of single RGB channels were performed. Several channels (2, 210, 222, 273, 456, 493, 530, 625, and 697) were chosen according to their importance to the multivariate model, i.e., channels with the largest absolute regression coefficients. Nevertheless, none of these univariate models provided good accuracies or linearities, with correlation coefficients always below 0.80. Models built with the means of red, green, and blue channel colors were also tested but they also provided poor results. Therefore, the use of multivariate calibration is justified. A starting multivariate PLS model was built with all the beer samples. This global model presented too large prediction errors for some samples in the lower part of the SRM scale (below 12 units). Thus, considering the requirement of dilution prescribed by the reference method for samples with absorbances higher than 1.0, two new calibration submodels were constructed. One local PLS model was built for high SRM range (diluted) samples, which presented reference values above 12 units. Other local model was built for low SRM range (non-diluted) samples, which presented reference values equal or below 12 units. In order to check this criterion for sample selection between these two models from the measured histograms, a previous PLS discriminant analysis (PLS-DA) model was built to assign the beer samples to these two classes. This discriminant model is essential for the users of this methodology since the brewer's analyst does not know beforehand which local PLS model must be used to

determine beer samples whose color is close to the limit between the two classes.

This selection resulted in 78 and 50 samples belonging to the high- and low-range models, respectively. In resume, the proposed methodology consisted of two steps. A PLS-DA model is firstly used to assign the histogram from the original beer sample to one of these two local PLS models. In the second step, the prediction of the color is provided by the respective calibration model. It is important to stress that for the proposed chemometric method, none of the samples is diluted since the dilution step is only required for the reference method.

Analytical validation of multivariate calibration methods is currently not a completely well-established matter. In relation to food analysis, there is no mention of multivariate statistics in most of the Brazilian and international validation guidelines (MAPA 2011; EC 2002; Thompson et al. 2002). Harmonization between univariate and multivariate aspects in analytical validation guidelines is essential for the recognition and dissemination of multivariate methodologies. In order to cope with regulatory requirements, proper figures of merit (FOM) were estimated for the proposed method, such as trueness, precision, linearity, working range, bias, and residual prediction deviation (RPD) (Botelho et al. 2013). In addition, the uncertainty of the predicted values was calculated in order to estimate confidence intervals. Sample-specific standard prediction errors (SPE) were calculated based on the errors-in-variables (EIV) equation (Faber et al. 2003). This equation takes into account the uncertainty of the reference method and should be used in the absence of bias. A last remark should highlight that considering the discrete nature of the predicted variable (color), estimated uncertainties, as well as root mean square errors of calibration (RMSEC) and validation (RMSEP), were rounded and represented without decimal places.

Results and Discussion

Global PLS Model

Initially, a model containing all the beer samples, corresponding to the whole analytical range from 3 to 210 SRM units, was constructed. In spite of having a good fit, this global PLS model presented too large relative errors for some samples in the lower part of the SRM scale. This can be observed in the plot of the residuals as a function of the predicted values (Fig. 1). To overcome this problem, samples were divided into two classes for building local PLS models, according to the need or not of dilution in the reference spectrophotometric method. It was verified that samples presenting absorbances at 430 nm above the threshold limit of 1.0 correspond to SRM units above 12.

It can be argued that in practical real situations, the assignment of a beer sample to one of these two local models is not known beforehand. So, a PLS-DA model was built as a previous step to take this decision. It was used as a screening model for the assignment of beer samples to one of the two local PLS models, the high and the low SRM range ones. Since an absorbance of 1.0 is exactly equivalent to 12 SRM units, this value was chosen as a reference limit between the two modeled classes.

PLS models based on other color systems, namely HSV and CMYK, were also tested, but no significant improvements were observed in relation to the model using RGB histograms as independent variables.

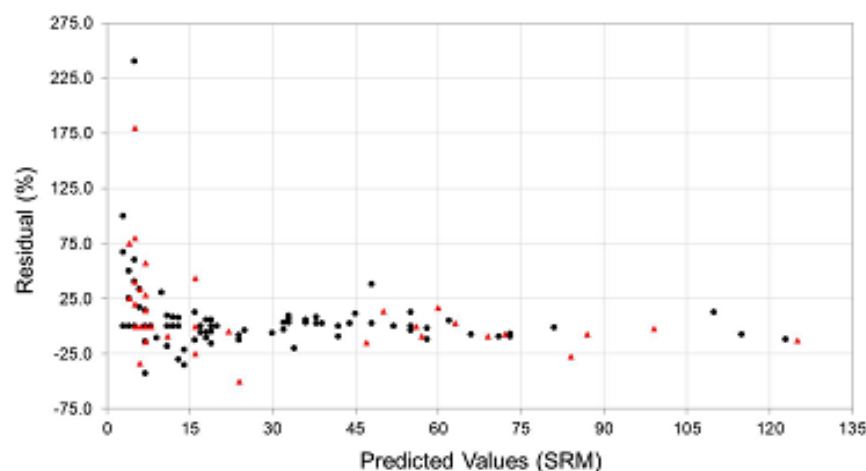
PLS-DA Model for Assigning Samples to Local PLS Models

As mentioned in the previous subsection, PLS-DA was used as a previous step of the proposed methodology to forward beer RGB histograms to the high-range or to the low-range PLS models. Sample RGB histograms were organized in a data matrix \mathbf{X} (128×768), in which the samples were placed in the lines and the color channels in the columns. This dataset was preprocessed by mean centering. A PLS-DA model was built by correlating the \mathbf{X} matrix of independent variables (frequencies at each channel) with a dummy \mathbf{y} vector of dependent variables. This vector assigns the reference values 1.0 to high-range samples and 0.0 to low-range samples. Then, the samples were split in 85 and 43 for the training and test sets, respectively, by applying the Kennard-Stone algorithm separately for each class (Kennard and Stone 1969). This procedure assures the selection of representative and homogeneously distributed samples in the whole analytical range of the multivariate space.

The best number of latent variables (LV) was chosen by random subset (9 splits and 20 iterations) cross-validation, according to the smallest cross-validation classification error (CVCE). The best PLS-DA model was obtained with 2 LV, accounting for 61.3% of the variance in the \mathbf{X} block and 87.5% in the \mathbf{Y} block. The Bayesian threshold was estimated as 0.59. The classification plot showing the \mathbf{Y} predicted values of all samples is depicted in Fig. 2. As can be seen, almost all the samples were correctly classified. Only one false-negative and one false-positive results were observed in the training set, while no sample was misclassified in the test set.

The interpretation of PLS-DA and PLS models can be performed through the observation of the informative vectors, such as regression coefficients and variable importance in projection (VIP) scores. The latter vector measures the importance of each variable in the projection used by a particular model in absolute values. Thus, VIP scores do not

Fig. 1 Plot of the residuals of the calibration and validation samples for the global PLS model. Calibration (black circles) and validation (red triangles) samples

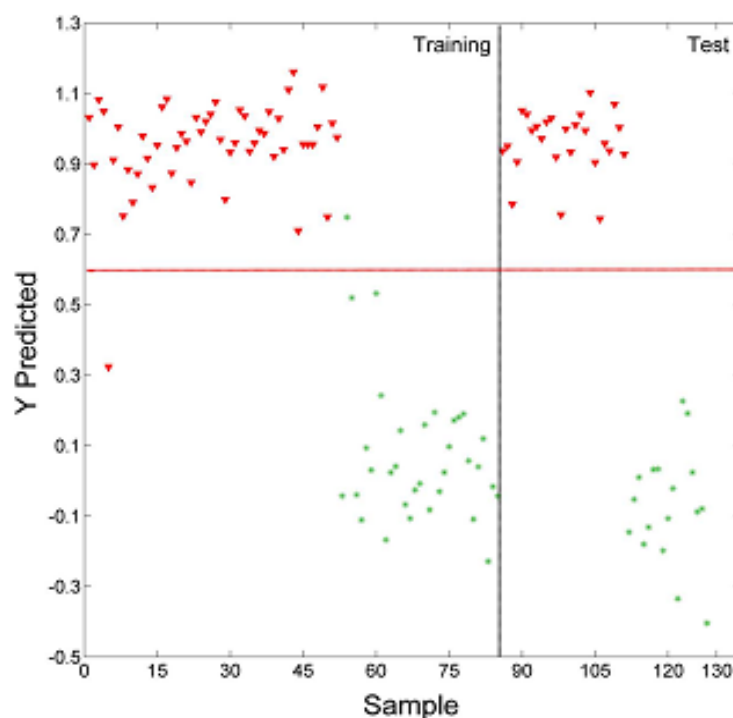


allow to observe the direction of each variable contribution. This information can only be estimated from the regression vectors. For the developed PLS-DA model, regression coefficients and VIP scores are shown in Figs. 3 a and b, respectively. There is a close agreement between these two vectors in terms of interpretation. Both of them indicate the relation of the variables that most contribute to the model to darker shades of red and blue, associated with high range/darker color beers, and lighter shades of green, associated with low range/lighter color beers.

Local PLS Models

As previously mentioned, the original 128 beer samples were divided into 78 to build the high-range model and 50 to build the low-range model. For each of these two local PLS models, samples were split into two-thirds for the calibration set and one-third for the validation set by using the Kennard-Stone algorithm. For the high-range model, this corresponds to 52 and 26 samples in the calibration and validation sets, respectively. For the low-range model, this corresponds to 33 and 17

Fig. 2 Predicted Y values for the PLS-DA model built to discriminate high- and low-range samples. Horizontal dashed line indicates the threshold and vertical dashed line indicates the separation between training and test samples. High- (down red triangles) and low-range (green asterisks) samples



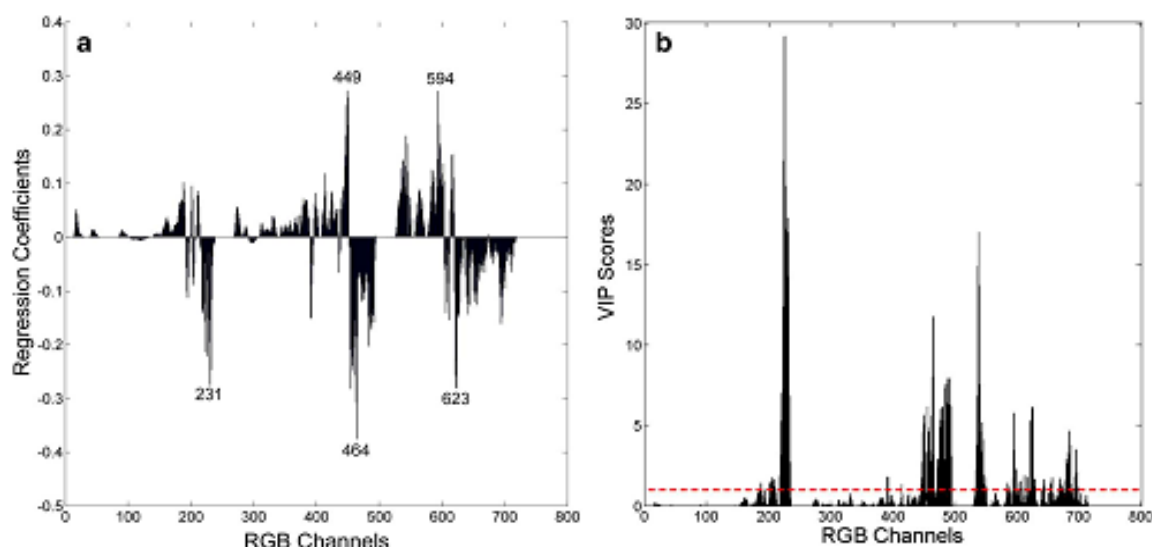


Fig. 3 Informative vectors for the PLS-DA model. **a** Regression vector and **b** variable importance in projection (VIP) scores. Horizontal dotted line in Fig. 3b indicates the threshold of 1.0, above which variable are considered to contribute significantly to the model

samples in the calibration and validation sets, respectively. All the samples were preprocessed by mean centering. The number of LV was selected according to the smallest root mean square error of cross-validation (RMSECV), estimated using random subsets CV (7 data splits and 20 iterations). The best models were built with 3 LV for both high and low ranges.

Both the models were optimized by outlier detection. Samples with large residuals in X (histograms), large residuals in Y (SRM units), or high leverages, at 95% confidence level, were considered outliers (Botelho et al. 2013, 2014, 2017) and deleted from the model. The number of detected outliers should not exceed the recommended limit of 2/9 (22.2%) (MAPA 2011). From the high-range model, 8 out of the original 78 samples (10.3%) were removed as outliers, being 3 in the calibration set and 5 in the validation set. From the low-range model, 3 out of the original 50 samples (6.0%) were detected as outliers, all in the validation set. After the removal of the outliers, the models were rebuilt. The results for the optimization of both the models, including RMSEC, RMSEP, and RPD, are shown in Table 1 (high-range model) and Table 2 (low-range model).

RMSEC and RMSEP are parameters describing the trueness of the methods. For the high-range model, these values

were very improved by the outlier detection, resulting in an RMSEC of 5 SRM units and an RMSEP of 6 SRM units. For the low-range models, these values were not significantly reduced after optimization, being both RMSEC and RMSEP equal to 1 SRM unit. RPD is the ratio of natural variation in the validation samples to the size of probable errors occurring during the prediction. It is dimensionless and evaluates the trueness of the model in absolute terms, independent of the analytical range. Validation RPD values were 5.6 and 3.1 for high- and low-range models, respectively. These results are satisfactory since PLS models with RPD values larger than 2.4 are considered to have good predictive capacity (Botelho et al. 2013; Williams 2001). The plot of the reference versus predicted values for these two models can be seen in Fig. 4.

Since only one sample in the high-range model presented SRM higher than 150 (210 units), it was excluded and a new model was developed in the working range from 13 to 130 (Fig. 4a).

The estimated regression coefficients for both PLS models are shown in Fig. 5. By observing these plots, it was possible to identify the channels that contribute more to predict the beer color. Channels at the beginning of the range of each primary color in the histogram (0 for red, 256 for green, and 512 for

Table 1 Results for the optimization of the developed high range PLS model through the detection of outliers

	Number of calibration samples	Number of validation samples	Number of LV	RMSEC ^a	RMSEP ^a	RPD validation
Before	52	26	3	10	27	1.5
After	49	21	3	5	6	5.6

^a SRM units

Table 2 Results for the optimization of the developed low-range PLS model through the detection of outliers

	Number of calibration samples	Number of validation samples	Number of LV	RMSEC ^a	RMSEP ^a	RPD validation
Before	33	17	3	1	1	1.5
After	33	14	3	1	1	3.1

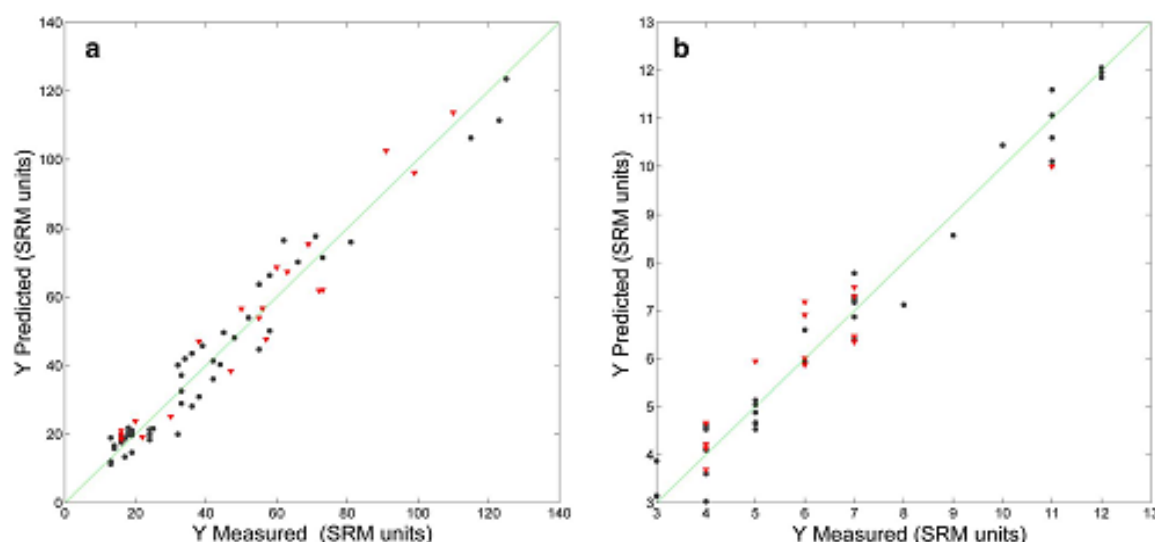
^aSRM units

blue) contribute to tones near to black. Oppositely, channels at the end of each part of the RGB histogram (255 for red, 511 for green, and 768 for blue) represent the primary colors close to their maximum intensities. As observed in Fig. 5a, the largest regression coefficients for the high-range model were observed at the channels 96, 276, and 533, which showed strong positive influences on the beer color quantification. This indicates the positive contribution of dark tones for the sample colors in this specific model. For the low-range model (Fig. 5b), a more complex situation is observed, with channels 210, 456, and 625 as the most positive regression coefficients and channels 222, 493, and 697 as the most negative ones.

Other estimated FOM for validating the methods are shown in Table 3. Linearity can be evaluated by observing the plots shown in Fig. 4 and their fit parameters in Table 3. Correlation coefficients of 0.9743 and 0.9864 were obtained for high- and low-range models, respectively. A Durbin-Watson test was applied to these fit residuals in order to check their independence. Estimated values of 2.24 and 2.06 are within the acceptance range (1.50–2.50), confirming the absence of autocorrelation in the model. The small values of bias estimated for both the models, considering only the validation samples, assured the absence of systematic errors. Student's *t* tests corroborated that zero bias is contained in both the estimated

confidence intervals at 95% level. Elliptical joint confidence regions (EJCR) were also estimated as an additional tool to verify the linearity and the absence of systematic errors in the models. The EJCR estimate is based on the intercept and slope of the fit curve. For both the models, EJCR estimates (95% confidence interval) contained the ideal point (slope 1.0 and intercept 0.0). These plots are shown in Fig. 2S (Supplemental material).

Method precision was also evaluated through the estimate of the relative standard deviation (RSD) for triplicates of three samples from each model at two different levels. For each model, these three samples were chosen to represent low, medium, and high color values, respectively. Precision levels at repeatability and intermediary precision were evaluated. For repeatability, all the triplicates were measured by the same analyst on the same day, while for intermediary precision, results came from two different analysts working in different days. Estimated repeatability RSD ranged between 0.2 and 0.7% for the high-range model, and between 0.4 and 0.8% for the low-range model. Intermediary precision RSD varied from 1.4 to 2.9% for the high-range model and from 2.2 to 4.5% for the low-range model. Taking into account the good results of trueness and precision, both these models can be considered accurate. Thus, the working ranges of 3–12 and

**Fig. 4** Reference versus predicted values for local PLS models. **a** High- and **b** low-range models. Calibration (black circles) and validation (red triangles) samples

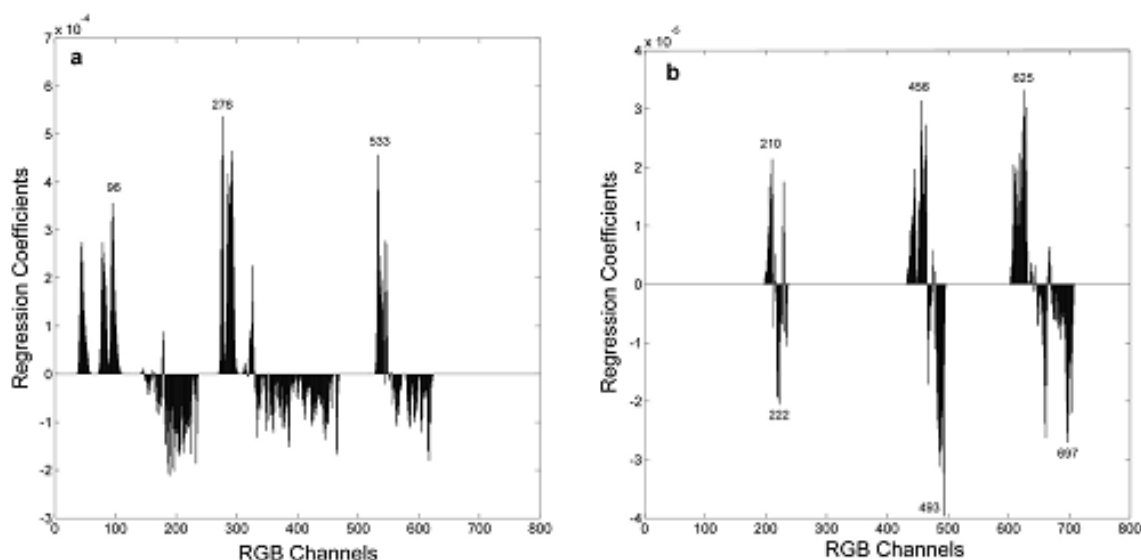


Fig. 5 Regression vector for local PLS models. **a** High and **b** low range models

13–130 can be established for the low and the high range models, respectively.

Finally, confidence intervals were estimated for the prediction of the validation samples. SPE calculated using the EIV equation were equal to 2 SRM units for all the samples in the low-range model and varied from 4 to 5 SRM units in the high-range model. Figure 6 presents reference and predicted values with the respective confidence intervals estimated for each validation sample, for both the models. For the low-range

model, all the reference values were within the estimated confidence interval, representing an agreement of 100%. For the

Table 3 Parameters for evaluating complementary FOM for the local PLS models

Figures of merit	Parameter	Values	
		High-range model	Low-range model
Precision	RSD repeatability (%)	0.7	0.8
		0.2	0.7
	RSD intermediary precision (%)	0.5	0.4
		1.4	4.5
Linearity	Durbin-Watson test	2.24	2.06
	Slope	0.98 ± 0.02	0.97 ± 0.03
	Intercept	1.05 ± 0.95	0.19 ± 0.22
	Correlation coefficient	0.9743	0.9864
Bias ^a		1 ± 5	0 ± 1
Working range ^b		13–130	3–12

^a SRM units

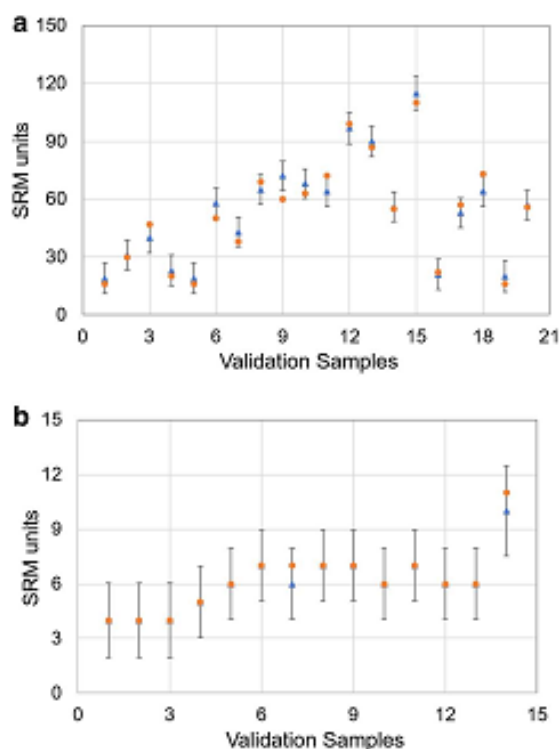


Fig. 6 Confidence intervals for the prediction of the validation samples, estimated based on the SPE. **a** High and **b** low range models. Predicted (triangles) and reference values (circles)

high-range model, only one out 21 samples was out of the confidence interval, representing 95.2% of agreement.

Conclusion

A multivariate image analysis method based on digital images obtained by a mobile phone coupled with chemometrics was developed and validated for the color determination in a great variety of beer styles, according to their SRM reference values. Due to the very wide analytical range (3–130 SRM units), two local PLS models were built to cover two different ranges of the values of the dependent variable. This new method presented several advantages over the current reference spectrophotometric method, such as a simpler procedure, without the need of any sample pretreatment in addition to degasification, more rapid, and of low cost. This method was validated in accordance with Brazilian and international guidelines, being considered linear, accurate, unbiased, and suitable for the use as an official methodology for the determination of color in beers. In Brazil, as well as in many other countries around the world, the number of new craft breweries is increasingly growing. Many of these breweries are small companies that artisanally produce many different styles of beers, whose colors may vary considerably from light to dark tones. The proposed method is very suitable for this type of business, especially when adapted in a portable analytical platform, with good potential to contribute to the democratization of the analytical science (Ozcam 2014).

Acknowledgments The authors want to thank the Brazilian agencies CNPq, CAPES (Finance Code 001), and FAPEMIG for financial support. A. C. C. F., H. V. P., and V. P. T. A. particularly acknowledge fellowships from CAPES and CNPq.

Compliance with Ethical Standards

Conflict of Interest Ana Carolina da Costa Fulgêncio declares that she has no conflict of interest. Vinícius Pires Teixeira Araújo declares that he has no conflict of interest. Hebert Vinícius Pereira declares that he has no conflict of interest. Bruno Gonçalves Botelho declares that he has no conflict of interest. Marcelo Martins Sena declares that he has no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent Not applicable.

References

Acevedo CA, Skurtys O, Young ME, Enrión J, Pedreschi F, Osorio F (2009) A non-destructive digital imaging method to predict immobilized yeast biomass. *LWT Food Sci Technol* 42:1444–1449

- ASBC (2015) *Methods of analysis: tristimulus analysis (colorimetric or spectrophotometric)*. American Society of Brewing Chemists, St. Paul
- Blanco CA, de La Fuente R, Caballero I, Rodríguez-Méndez ML (2015) Beer discrimination using a portable electronic tongue based on screen-printed electrodes. *J Food Eng* 157:57–62
- Böck FC, Helfer GA, Costa AB, Dessury MB, Ferrão MF (2018) Rapid determination of ethanol in sugarcane spirit using partial least squares regression embedded in smartphone. *Food Anal Methods* 11:1951–1957
- Borin A, Ferrão MF, Mello C, Cordi L, Pataca LCM, Durán N, Poppi RJ (2007) Quantification of *Lactobacillus* in fermented milk by multivariate image analysis with least-squares support-vector machines. *Anal Bioanal Chem* 387:1105–1112
- Botelho BG, Mendes BAP, Sena MS (2013) Development and analytical validation of robust near-infrared multivariate calibration models for the quality control of mozzarella cheese. *Food Anal Methods* 6:881–891
- Botelho BG, Assis LP, Sena MM (2014) Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages. *Food Chem* 159:175–180
- Botelho BG, Dantas KCF, Sena MM (2017) Determination of allura red dye in hard candies by using digital images obtained with a mobile phone and N-PLS. *Chemom Intell Lab Syst* 167:44–49
- Capitan-Vallvey LF, Lopez-Ruiz N, Martínez-Olmos A, Erenas MM, Palma AJ (2015) Recent developments in computer vision-based analytical chemistry: a tutorial review. *Anal Chim Acta* 899:23–56
- Carvalho FR, Moors P, Wagemans J, Spence C (2017) The influence of color on the consumer's experience of beer. *Front Psychol* 8:2205
- Chun HJ, Park YM, Han YD, Jang YH, Yoon HC (2014) Paper-based glucose biosensing system utilizing a smartphone as a signal reader. *BioChip J* 8:218–226
- Cilo N (2019) Pequenas e médias cervejarias pressionam grandes marcas. Estado de Minas, Belo Horizonte. https://www.em.com.br/app/noticia/economia/2019/02/08/intermas_economia,1028869/pequena-e-media-cervejaria-pressionam-grandes-marcas.shtml. Accessed March 2019
- Curi RA, Venturini Filho WG, Ducatti C, Nojimoto T (2008) Beer production using barley and corn maltose as malt adjuncts: physico-chemical, sensory and isotopic analyses. *Braz J Food Technol* 11: 279–287
- European Community (EC) (2002) Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. *Official Journal* 221:8–36 Brussels
- Eßlinger HM (2009) *Handbook of brewing: processes, technology, markets*. Wiley-VCH, Darmstadt
- Faber NM, Song XH, Hopke PK (2003) Sample-specific standard error of prediction for partial least squares regression. *TrAC – Trends Anal Chem* 22:330–334
- Fengxia S, Chang Y, Zhou Z, Yu Y (2004) Determination of beer color using image analysis. *J Am Soc Brew Chem* 62:163–167
- Fernandes DDS, Romco F, Krepper G, Di Nezio MS, Pistonesi MF, Centurión ME, Araújo MCU, Diniz PHGD (2019) Quantification and identification of adulteration in the fat content of chicken hamburgers using digital images and chemometric tools. *LWT Food Sci Technol* 100:20–27
- Geladi P, Grahn H (1996) *Multivariate image analysis*. John Wiley, Chichester
- Godinho MS, Oliveira AE, Sena MM (2010) Determination of interfacial tension of insulating oils by using image analysis and multi-way calibration. *Microchem J* 96:42–45
- Helfer GA, Tischer B, Filoda PF, Parckert AB, Santos RB, Vinciguerra LL, Ferrão MF, Barin JS, Costa AB (2018) A new tool for interpretation of thermal stability of raw milk by means of the alizarol test

- using a PLS model on a mobile device. *Food Anal Methods* 11: 2022–2028
- Huang J, Wium H, Qvist KB, Esbensen KH (2003) Multi-way methods in image analysis - relationships and applications. *Chemom Intell Lab Syst* 66:141–158
- Huige NJ (2006) Brewery by-products and effluents. In: Stewart GG, Priest FG (eds) *Handbook of Brewing*, 2nd edn. CRC Press, Toronto, pp 655–713
- Kenward RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11:137–148
- Lachenmeier DW (2007) Rapid quality control of spirit drinks and beer using multivariate data analysis of Fourier transform infrared spectra. *Food Chem* 101:825–832
- Lapere Y (2015) The beer sector in Brazil. Flanders Investment & Trade Market Survey, São Paulo. https://www.flandersinvestmentandtrade.com/export/sites/trade/files/market_studies/800151125155419/800151125155419_10.pdf. Accessed March 2019
- Lied TT, Geladi P, Esbensen KH (2000) Multivariate image regression (MIR): implementation of image PLSR-first forays. *J Chemom* 14: 585–598
- Ludwig SKJ, Zhu H, Phillips S, Shiledar A, Feng S, Tseng D, van Ginkel LA, Nielsen MWF, Ozcan A (2014) Cellphone-based detection platform for rbST biomarker analysis in milk extracts using a microsphere fluorescence immunoassay. *Anal Bioanal Chem* 406:6857–6866
- Ministério da Agricultura, Pecuária e Abastecimento (MAPA) (2011) *Manual da Garantia da Qualidade*. Secretaria de Defesa Agropecuária, MAPA, Brasília, p 227
- Ozcam A (2014) Mobile phones democratize and cultivate next-generation imaging, diagnostics and measurement tools. *Lab Chip* 14:3187–3194
- Preechaburana P, Suska A, Filippini D (2014) Biosensing with cell phones. *Trends Biotechnol* 32:351–355
- Shellhammer TH (2009) Beer color. In: Bamforth C (ed) *Beer. A quality perspective*. *Handbook of Alcoholic Beverages*. Academic Press, Cambridge, pp 213–227
- Siqueira PB, Bolini HMA, Macedo GA (2008) Beer production and its effects on the presence of polyphenols. *Braz J Food Nutr* 19:491–499
- Thompson M, Ellison SLR, Wold R (2002) Harmonized guidelines for single laboratory validation methods of analysis. *Pure Appl Chem* 74:835–855
- Torres AR, Lyra WS, de Andrade SJE, Andrade RAN, Silva EC, Araujo MCU, Gaião EN (2011) A digital image-based method for determining of total acidity in red wines using acid-base titration without indicator. *Talanta* 84:601–606
- Vidal M, Garcia-Arroña R, Bordagaray A, Ostra M, Albizu G (2018) Simultaneous determination of color additives tartrazine and allura red in food products by digital image analysis. *Talanta* 184:58–64
- Viejo CG, Fuentes S, Torrico D, Howell K, Dunshea FR (2018) Assessment of beer quality based on foamability and chemical composition using computer vision algorithms, near infrared spectroscopy and machine learning algorithms. *J Sci Food Agric* 98:618–627
- Williams P (2001) Implementation of near-infrared technology. In: Williams P, Norris K (eds) *Near-Infrared Technology in the Agricultural and Food Industries*, 2nd edn. American Association of Cereal Chemists Inc., St. Paul
- Zhang C, Bailey DP, Suslick KS (2006) Colorimetric sensor arrays for the analysis of beers: a feasibility study. *J Agric Food Chem* 54:4925–4931

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

APÊNDICE B

Food Analytical Methods (2022) 15:307–316
<https://doi.org/10.1007/s12161-021-02126-w>



Determination of Alcohol Content in Beers of Different Styles Based on Portable Near-Infrared Spectroscopy and Multivariate Calibration

Ana Carolina da Costa Fulgêncio¹ · Glaucimar Alex Passos Resende¹ · Marden Claret Fontoura Teixeira¹ · Bruno Gonçalves Botelho¹ · Marcelo Martins Sena^{1,2}

Received: 8 June 2021 / Accepted: 3 September 2021 / Published online: 15 September 2021
 © The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

The determination of alcohol content in beers is essential for the quality control of this beverage. This paper proposed and validated a new rapid and direct multivariate method for this aim using a portable near-infrared (NIR) spectrometer and partial least squares (PLS) regression. Reference values were obtained by a gas chromatography with flame ionization detection (GC-FID) method developed and validated for this purpose. Aiming at building a robust model, a great variety of beers, from different styles, brands, and breweries, was incorporated into the model. NIR spectra were recorded between 908 and 1676 nm for 92 beer samples, corresponding to a range from 3.2 to 10.9% (v/v) of alcohol content. PLS model provided accurate results with root-mean-square error of calibration (RMSEC) and prediction (RMSEP) of 0.5% and 0.6%, respectively. The developed method was validated through the estimate of figures of merit, such as linearity, trueness, precision, analytical sensitivity, bias, and residual prediction deviation (RPD). In addition, an elliptical joint confidence region was calculated to verify the linearity, and confidence intervals based on the standard prediction errors were estimated for the validation samples.

Keywords Handheld NIRS sensor · Chemometrics · Brewing · Beer quality control · Ethanol · Alcoholic strength

Introduction

Beer is the most consumed and traded alcoholic beverage in the world, with China followed by the USA and Brazil as the main producers and Germany as the main exporter (Török et al. 2020). Worldwide, the beer market is extremely competitive and has grown dramatically except for the last year. Only the USA beer market moved US \$114.2 billion in 2018 (Wachelko et al. 2021) and decreased to US \$94.1 billion in 2020 (Brewers Association 2020). Around 24% of this market represented craft beers. The production and consumption of craft beers have experienced an astonishing boom in the last two decades, along with the diversification of the product in relation to its characteristics and styles. This has opened new business opportunities ranging from the creation of a new market niche in tourism, beer tourism

(Kraftchick et al. 2014), to the emerging market of high-priced specialty beers (Donadini et al. 2016).

In Brazil, a similar trend has been observed, with an increase of 30% in the number of new breweries from 2017 to 2018 (Fulgêncio et al. 2020). Many of them are craft breweries and particularly small business, often producing different beers styles, such as lager, pale ale, weiss, stout, and porter. Brewing beer is a complex process, and the quality control is a key aspect for a company to establish itself in this demanding and competitive market. Frequently, craft breweries have outsourced chemical analysis. Nevertheless, the rapid assessment of the most relevant quality control parameters is important for the cost-effectiveness of the production. For this task, portable equipment, such as handheld spectrometers or mobile phones, can be especially useful when combined with chemometrics aiming at developing supervised classification or multivariate calibration models (Fulgêncio et al. 2020).

Alcohol content is one of the most important parameters in the beer quality control. Different methods have been used to determine this crucial parameter in beers, relying on measuring some physical or chemical property of ethanol. According to the American Society of Brewing Chemists

✉ Marcelo Martins Sena
 marcsen@ufmg.br

¹ Chemistry Department, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil

² Instituto Nacional de Ciência E Tecnologia Em Bioanalítica (INCT Bio), Campinas, SP 13083-970, Brazil

(ASBC), there are several official methods for this analysis, based on techniques such as gas chromatography (GC), refractometry, enzymatic assay, dedicated instruments, and densitometry (ASBC 2015), being the last method the most used due to not demanding skilled labor and utilizing low-cost glassware. In the literature, there has been a special interest in the development of new sensors for directly determining ethanol in beer. They have been based on a potentiometric porphyrin-based electronic tongue (Lvova et al. 2006), an electrochemical high-performance enzyme entrapment platform (Adhikari et al. 2017), a paper-based nanomodified electrochemical biosensor (Cinti et al. 2017), a metal oxide semiconductor electronic nose (Voss et al. 2019), and ethanol-sensitive polyacrylamide hydrogels (Erfkamp et al. 2019). Another alternative has been the use of dedicated equipment based on near-infrared (NIR) spectroscopy (Kaneko et al. 2007), which can be combined with a distillation step (Cloninger 2018). Still, the development of new methods using headspace GC for quantifying alcohol in beers has continued to raise interest in food chemistry literature (Wachefko et al. 2021). While some of these methods are expensive, laborious, and time-consuming such as the chromatographic ones, the simple densitometric method is non-selective and may suffer interference from any analyte that decreases the density of the distillate. Thus, it is of great interest for beer producers to develop rapid, direct, and low-cost methods for quality control.

Several analytical instruments are now available in compact versions. As already mentioned, the recent availability of these portable equipment is particularly useful for the quality control of food and beverages (Abasi et al. 2018). They allow the development of simple, rapid, direct, and in situ analytical methods. As samples do not need to be sent to the laboratory, an important advantage comes from the time saved by eliminating the delay between sampling and availability of the results, which limit the frequency of analysis and the optimization of the production (Alcalà et al. 2013). The dissemination of portability opens new perspectives by modifying the way the customers conduct business, opening new markets, and being incorporated in consumer devices and smart appliances, becoming in the coming years part of the Internet of Things (Croccombe 2018), and contributing to the democratization of analytical chemistry (Ozcam 2014). Specifically, portable NIR spectrometers can be considered chemical sensors that present sensitivity close to benchtop instruments. They show other advantages over benchtop spectrometers, such as small size, low cost, low energy consumption, and lack of moving parts in their optical structure (Santos et al. 2013).

In the last years, portable NIR spectroscopy has been combined with multivariate calibration, mainly with partial least squares (PLS), for the quality control of foods and beverages. Applications include fatty acid quantitation in

peanuts (Yu et al. 2020), determination of quality and ripeness parameters in citrus fruits (Santos et al. 2021), quantification of cocoa bean quality parameters (Anyidoho et al. 2021), real-time analysis of cow milk samples (Muñiz et al. 2020), determination of physical and chemical parameters in the roasted coffee industry (Baqueta et al. 2019), determination of titratable acidity and ascorbic acid content in acerola fruit during ripening (Malegori et al. 2017), on-site assessment of post-harvest quality and refrigerated storage behavior of plums (Pérez-Marín et al. 2010), and assessment of apricot fruit quality (Camps and Christen 2009). Only one application of portable NIR spectroscopy has been found discriminating beers according to their origins (Li et al. 2009), but no quantitative application has been published determining any beer quality parameter.

The aim of this paper was to combine portable NIR spectroscopy and PLS to develop and validate a simple, robust, and rapid method for directly determining alcohol content in beer. In order to obtain reference values for multivariate calibration, a method based on GC with flame ionization detection (GC-FID) was also developed and validated. Aiming at building a robust model, a great variety of beers, from different styles, brands, and breweries, was incorporated into the model. The proposed multivariate method was also validated in accordance with Brazilian and international guidelines, incorporating multivariate aspects (Botelho et al. 2013, 2014) with the aim to assure its reliability and fit for the intended purpose.

Materials and Methods

Instruments and Software

Spectra were recorded on a Viavi Solution MicroNIR® 1700 spectrometer (Milpitas, CA, USA). MicroNIR® 1700 had a LVF (linear variable filter) component setup directly connected to a linear detector array. The detector was a 128-pixel InGaAs array, which monitors the wavelength range from 900 to 1700 nm, achieving a geometric resolution of < 1%. Two tungsten lamps were used in the diffuse reflection mode providing an optimal broadband light output in 1000–2000 nm range. The diffuse reflected radiation was collected and sent to a computer. Data were processed using MATLAB® software, version 7.13 (MathWorks, Natick, MA, USA), along with the PLS toolbox, version 6.7.1 (Eigenvector Technologies, Wenatchee, WA, USA).

Samples

Ninety-two beer samples were purchased mainly in the local commerce of Belo Horizonte, MG, Brazil, and stored at room temperature until the analysis. These beer samples

were obtained from a large variety of breweries, brands, and styles (lager, Vienna lager, pale ale, red ale, Pilsen, IPA, weissbier, strong pale ale, stout, porter, etc.) with their alcohol content ranging from 3 to 11% (v/v). Information (number of samples for each beer style and the respective range of alcohol content) about all the 92 analyzed beer samples is summarized in Table 1.

Sample Preparation and Procedure

Approximately 50 mL of each beer sample were degassed using an ultrasonic bath (Equilab ULTRASONIK 28H, Madrid, Spain) for 5 min. Three drops of 1-octanol were added to each beer sample in order to avoid excessive foaming. After degassing, part of each sample was used for the reference method (GC-FID) and the remaining part for the proposed NIR method. For the reference method, 100 μ L of each sample was mixed with 900 μ L of the internal standard (IS) solution (1-butanol 0.1% v/v) in a 2-mL microcentrifuge tube. In the sequence, each tube was homogenized using a vortex mixer (VELP Scientifica, Usamate Velate, Italy) during 10 s at 3000 rpm. Samples were then ready to be injected in the chromatographic system. For the proposed portable spectroscopic method, 3 mL of each sample was poured (using an automatic burette) in a polystyrene Petri

dish (3.5 cm in diameter \times 1.2 cm in height), without no additional preparation. Then, the plate was placed above the MicroNIR® for spectra recording with 20 scans, resolution of 6.25 nm, and in the wavelength range of 908–1676 nm. Ten replicate spectra of the empty plate were also recorded in the same conditions in order to estimate the instrumental noise. During all these measurements, the temperature of the room was controlled at 25 ± 1 °C. It is important to note that the diffuse reflectance mode was used. This mode requires that the scattering coefficient should be kept constant during the experiments, so the sample volume and the container remain constant (Blitz 1998). In our experiments, care was taken to ensure these conditions by using a reflective artifact and controlling the sample volume, and the path length corresponded to 1.0 cm.

GC-FID Method

In this study, a Shimadzu GC-17A gas chromatography system (Kyoto, Japan) equipped with a flame ionizing detector (FID) and a Poraplot Q column (10 m \times 0.32 mm \times 10 μ m) was used. Hydrogen was used as carrier gas in a flow of 2.2 cm³.min⁻¹. Both the injection port and the detector were held at 250 °C. Oven temperature was kept at 20 °C for 4 min. A 0.2 μ L of the diluted sample was injected with split ratio of 1:2.

In order to assess the performance of the reference method developed for this work, a full analytical validation procedure was performed. Some figures of merit (FOM) were estimated, such as linearity and matrix effect, precision, trueness and limits of detection, and quantification (AOAC 2016).

Linearity and Matrix Effect

The linearity of the proposed method was evaluated according to the procedure described by Souza and Junqueira (2005). A six-point calibration curve was built with equally spaced points (1.0, 3.0, 6.0, 9.0, 12.0, and 15.0% v/v of alcohol), using ultrapure water as solvent (solvent curve) and non-alcoholic beer (matrix-matched curve). Three replicates were analyzed for each calibration level. The presence of outliers in both the curves was evaluated, and they were excluded using jackknife. Regression parameters were estimated for ordinary least squares (OLS) regression. The residuals from this regression were evaluated regarding its normality, homoscedasticity, and independence, using Ryan–Joiner, Brown–Forsythe, and Durbin–Watson tests, respectively. Lack of fit and significance of the regression were also tested for the proposed regression (Souza and Junqueira 2005). For testing the presence of matrix effect, the slope from the solvent curve and the matrix-matched curve were compared using a *t* test (95% confidence level).

Table 1 Information about all the 92 beer samples used in the developed PLS model

Number of samples	Beer style	Alcohol content (%)
15	American lager	3.2–6.7
1	Bock	8.1
6	Brown ale	4.8–6.4
2	Dark ale	6.1–10.1
8	Dark lager	3.1–8.3
4	Dunkel	4.4–6.5
1	Japanese rice lager	5.5
1	Kölsch	5.6
3	Malzbier	4.2–4.6
10	Pale ale	3.9–8.6
10	Pilsen	4.3–7.4
4	Porter	5.5–10.4
1	Radler	3.2
3	Red ale	5.5–9.5
2	Red lager	4.3–4.6
2	Specialty beer	4.6–4.8
8	Stout	3.7–10.9
2	Strong pale ale	7.1–8.5
3	Vienna lager	4.4–5.5
5	Weissbier	3.2–8.0
1	Wood aged	10.9

Precision and Recovery

For the precision and recovery evaluation, six replicates at three different concentrations (1.0, 8.0, and 15.0% v/v) were prepared with non-alcoholic beer and analyzed in 3 consecutive days by three different analysts. The mean relative recovery was estimated from 18 replicates (six replicates from three analysts). Precision results were presented as relative standard deviations under repeatability (DPR_r) and intermediary precision (DPR_p) and were evaluated through ANOVA. All assumptions needed for ANOVA (normality, homoscedasticity, and independency of the residuals) were previously tested and confirmed.

Limits of Detection and Quantification

The limits of detection (LOD) and quantification (LOQ) were estimated using ten replicates of a blank sample (ultrapure water) and the slope from the matrix-matched calibration. LOD was estimated as 3.3 times the standard deviation of ten blank samples divided by the slope, and LOQ was estimated as 10 times the standard deviation of these same blank samples divided by the slope.

Multivariate Analytical Validation

Analytical validation of multivariate calibration methods is still not a completely well-established matter. This study searched to harmonize univariate and multivariate concepts, aiming to adapt the requirements of the Brazilian (Brasil 2009) and international guidelines (Thompson et al. 2002; AOAC 2016) to the state of the art of how to estimate FOM in multivariate calibration (Botelho et al. 2013, 2014). The following FOM were estimated: trueness, precision, linearity, working range, analytical sensitivity, bias, and residual prediction deviation (RPD). In addition, an elliptical joint confidence region (EJCR) was calculated to verify the linearity, and confidence intervals based on the uncertainties of the predicted values were estimated for the validation samples. Errors-in-variables (EIV) equation was used to estimate sample-specific standard prediction errors (SPE), which should be used in the absence of bias (Faber et al. 2003). SPE for each sample was estimated through the following equation, in which RMSEC is the root-mean-square errors of calibration, h is the leverage for each sample, and $s(y_{ref})$ is the estimate of the standard error for the reference method:

$$SPE = RMSEC \sqrt{(1 + h) - s(y_{ref})^2} \quad (1)$$

Results and Discussion

Validation of the GC-FID Method for Obtaining Reference Values

Considering the chromatographic system available in our laboratory, a GC-FID method was developed and validated for quantifying alcohol in beers in order to obtain reference values for building a NIR-PLS model. Linearity of the chromatographic method was confirmed for ethanol content determination in the proposed range (1.0–15.0% v/v) for both solvent (using ultrapure water) and matrix-matched curves (using non-alcoholic beer) with all the assumptions of linear regression being assured for both curves, as can be seen on Table 2.

For this chromatographic method, outlier detection was performed using the Jackknife test. Four outliers were detected and removed in the solvent curve, and three outliers were removed from the matrix-matched curve. This is in accordance with the recommended limit for removing outliers (22.2% of the original data) (Souza and Junqueira 2005). After this removal, the residuals from both curves were normally distributed ($p > 0.10$), independent ($p < 0.05$), and homoscedastic ($p < 0.05$), according to the specific parameters estimated for Ryan–Joiner, Durbin–Watson, and Brown–Forsythe tests, respectively (Table 2). The regression (evaluated by ANOVA) was also significant ($p < 0.001$), and the lack of fit was not significant ($p > 0.05$) for both curves, confirming the linearity of the solvent and the matrix-matched curves. When the slope from solvent curve is compared to the matrix-matched curve, significant difference can be observed ($p < 0.05$). Since matrix effect was verified, further validation steps were performed using matrix-matched curves (Fig. 1).

Recovery and precision under conditions of repeatability and intermediary precision were also evaluated. Recovery was estimated at three different levels (1.0, 8.0, and 15.0% v/v) and ranged from 98.4 to 101.4%, being in accordance with the limits defined by the AOAC guidelines (AOAC 2016). Precision was evaluated through relative standard deviation (RSD), which ranged from 1.9 to 2.2%. LOD and LOQ were also estimated as 0.06% and 0.19% v/v, respectively. Both these limits were lower than the first concentration of the proposed calibration curve.

Thus, the proposed chromatographic method for alcohol content determination was considered suitable for generating the reference values used to build the multivariate calibration model. The ease of sample preparation (a simple dilution with IS solution) and the short time for an isothermal chromatographic run (3 min) were also consistent with the great number of samples that had to be analyzed.

Table 2 Figures of merits used to evaluate the linearity of the chromatographic method (GC-FID) developed for obtaining reference values for the determination of alcohol in beers

	Statistical parameter	Solvent calibration curve	Matrix-matched calibration curve
n		14	15
Equation		$y=0.0832x+0.0153$	$y=0.0773x+0.0118$
	R^2	0.999	0.998
Normality	R_{eq}	0.989	0.987
	p	> 0.10	> 0.10
Homoscedasticity	t_L	0.342	0.042
	p	0.7383	0.967
Independency	d	2.182	1.539
	p	> 0.05	> 0.05
Regression	F	5067	17,135
	p	3.93×10^{-17}	1.13×10^{-21}
Lack of fit	F	0.306	1.249
	p	0.866	0.357

R^2 , coefficient of determination; R_{eq} , parameter for evaluating the normal distribution of the residuals in the Ryan–Joiner test; t_L , parameter for evaluating homoscedasticity in the Brown–Forsythe test; d , parameter for evaluating the absence of autocorrelation of the residuals in the Durbin–Watson test; F , F test value for ANOVA; p , p value

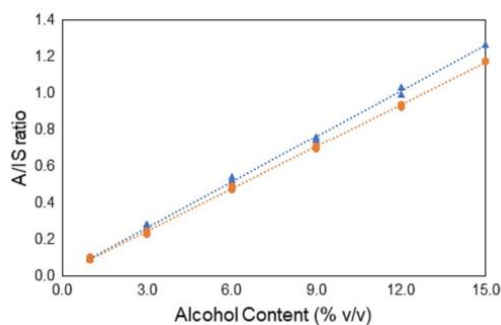


Fig. 1 Evaluation of the linearity of the GC-FID method. Blue triangles represent the solvent calibration curve, and orange circles represent the matrix-matched calibration curve. A/IS means analyte to internal standard

PLS Model Development

The NIR spectra of the 92 samples of beers are shown in Fig. 2. By observing these spectra, it was possible to assign the most intense and broad bands, centered around 1150 nm and 1300 nm, to ethanol vibrations, corresponding to the second overtone of CH_3 and CH_2 stretchings and to the second overtone of the R-OH stretching, respectively (Metrohm NIRSystems 2014).

The Kennard–Stone algorithm was applied to select 61 samples (two thirds) for the calibration set, while the remaining 31 samples were used for the validation set. Through a uniform scanning of the spectral data, this algorithm assured the presence of the most representative samples in

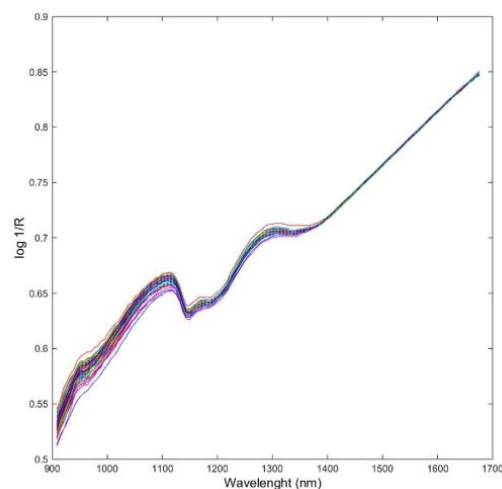


Fig. 2 NIR spectra of all the 92 analyzed beer samples

the calibration set (Kennard and Stone 1969). In Fig. 2, the presence of drifts can be noted. Drifts are nonlinear baseline deviations typical for diffuse reflectance spectra and caused by multiplicative light scattering. Thus, standard normal variate scaling (SNV) and Savitzky–Golay (SG) smoothing were used for data preprocessing in order to overcome this problem, previously to the mean centering (Rinnan et al. 2009).

The number of latent variables (LV) was selected according to the smallest root-mean-square error of cross-validation

(RMSECV), estimated using random subsets (7 data splits and 20 iterations). The best model was selected with 6 LV after optimization by outlier detection. Samples with large residuals in the *X* block (spectra), large residuals in the *Y* block (concentration values), or high leverages, at 95% confidence level, were detected as outliers (Botelho et al. 2014). The number of outliers cannot exceed the recommended limit of 22.2% of the calibration and validation sets (MAPA 2015). From the whole data, 13 out of the original 92 samples were removed as outliers, being 8 in the calibration set (13%) and 5 in the validation set (16%). After the removal of outliers, the model was rebuilt. The results for the optimization of the model, including root-mean-square errors of calibration (RMSEC) and prediction (RMSEP) and RPD, are showed in Table 3.

RMSEC and RMSEP are typical parameters for describing the trueness of multivariate calibration methods. These values were improved by the outlier detection, jointly with RPD. For the final PLS model, RMSEC and RMSEP of 0.5% and 0.6% were obtained, respectively. The concept of the RPD was introduced by Williams for NIR spectroscopy quantitative methods (Williams and Sobering 1993) aiming at evaluating their trueness in absolute terms, regardless of the analytical range. It corresponds to the ratio of natural variation in the samples to the size of likely prediction errors. Models with RPD above 2.4 present good predictive ability (Botelho et al. 2013; Williams 2001). Our final PLS model presented RPD of 2.8 and 2.7 for the calibration and validation sets, respectively. The plot of reference versus

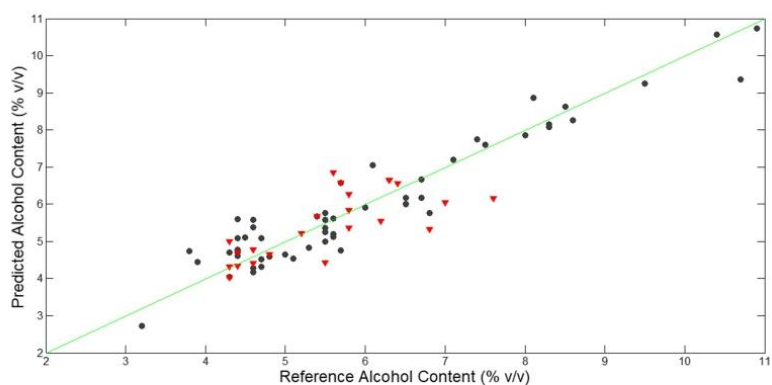
predicted values is shown in Fig. 3. As can be seen, the model presents good linearity, and no systematic trend was observed for its residuals.

An important aspect for any NIR multivariate calibration model is its spectral interpretation through the informative vectors. The estimated vectors of regression coefficients and variable importance in projection (VIP) scores for the developed PLS model are show in Fig. 4. VIP scores (Fig. 4b) show the importance of each variable for the PLS model in absolute terms (a threshold of 1.0 is suggested for the most significant variables), while the most positive regression coefficients are directly related to the ethanol prediction, and the most negative coefficients are related to the interferences (Fig. 4a). By observing Fig. 4, it was possible to identify the spectral regions that more contribute to predict the alcohol content. As it was expected, the most positive regression coefficients are centered around 1317 nm, which also correspond to the second most intense VIP scores peak. This spectral band can be assigned to the first overtone of C-H stretching and the first overtone of the O-H stretching of ethanol (Nordon et al. 2005). On the other hand, the second most negative regression coefficients (and the third most intense VIP scores) were centered around 970 nm and can be assigned to the third overtone of the combination band of symmetric and asymmetric stretching modes of the OH bond of water. The most negative regression coefficients and the most intense VIP scores are centered around 1187 nm. This band may be assigned to the second overtone of CH₂ stretching of carbohydrates (Metrohm NIRSystems 2014).

Table 3 Results for the optimization of the developed PLS model through the detection of outliers

	Number of calibration samples	Number of validation samples	Number of latent variables	RMSEC (%)	RMSEP (%)	RMSECV (%)	RPD calibration	RPD validation
Before	61	31	6	0.8	0.9	1.0	2.1	1.8
After	53	26	6	0.5	0.6	0.7	2.8	2.7

Fig. 3 Plot of reference versus predicted values for the PLS model. Black circles indicate calibration samples, and down red triangles indicate validation samples



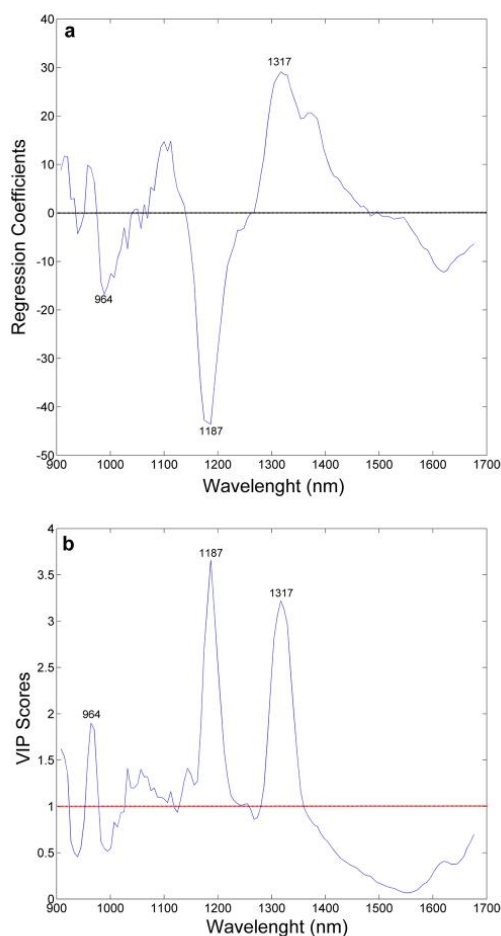


Fig. 4 Informative vectors for the PLS model. **a** Regression vector and **b** variable importance in projection (VIP) scores. Red horizontal dotted line in **b** indicates the threshold of 1.0, above which variables are considered to contribute significantly to the model

Other estimated FOM for validating the developed method are shown in Table 4. Linearity can be evaluated by observing the already discussed plot of reference versus predicted values (Fig. 3) and its fit parameters (Table 4). A correlation coefficient of 0.9355 and a slope of 0.92 ± 0.08 were estimated. In addition, the independency of the residuals was assured by the Durbin–Watson test, whose estimated parameter, 2.21, was within the acceptance limits (1.5–2.5). Bias was calculated as $0.1 \pm 0.6\%$, considering only the validation samples, and the obtained small value confirmed the absence of systematic errors in the model. This was checked by a Student's *t* test with 95% confidence level and 26 degrees of

Table 4 Parameters for evaluating complementary figures of merit for the developed portable NIR spectroscopic method

Figure of merit	Parameter	Value
Precision	RSD repeatability (%)	1.5
		1.1
	RSD intermediate precision (%)	1.9
		2.5
		3.5
Linearity	Durbin–Watson test parameter	3.9
	Slope	0.92 ± 0.08
	Intercept	0.58 ± 0.45
	Correlation coefficient	0.9355
Analytical sensitivity ($\%^{-1}$)		2.1
Bias (%)		0.1 ± 0.6
Working range (%)		3.2–10.9

freedom. An additional statistical tool, EJCR, was applied to check the linearity. As can be seen in Fig. 5, the ideal point (intercept = 0.0, slope = 1.0) was within the boundary of the estimated EJCR (95% confidence interval).

The method precision was also evaluated through the RSD for triplicates of three samples at two different levels, which were chosen to represent low, medium, and high alcohol contents, respectively. RSD varied between 1.1 and

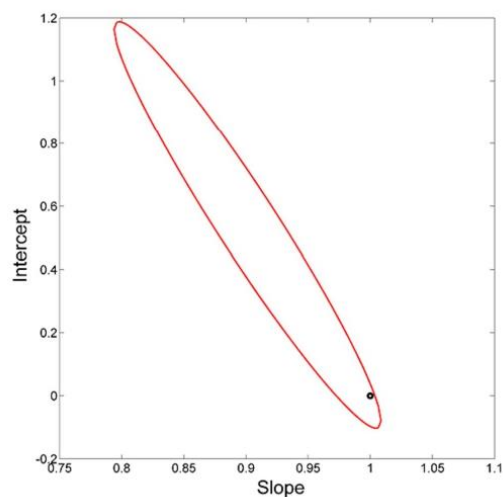


Fig. 5 Elliptical joint confidence region (EJCR) estimated (95% confidence interval) for the linearity (slope and intercept) of the PLS model. Black dot indicates the ideal point and red ellipse limits the EJCR

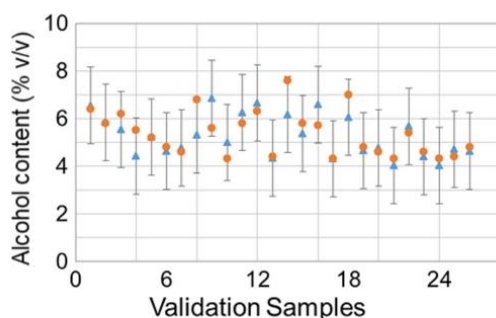


Fig. 6 Confidence intervals for the prediction of the validation samples, estimated based on the SPE. Predicted (blue triangles) and reference values (orange circles)

1.9% at the level of repeatability and between 2.5 and 3.9% at the level of intermediary precision (Table 4). All of these results were in accordance with the international guidelines (AOAC 2016) that prescribe a maximum RSD of 3.7% for repeatability and 6.0% for intermediate precision, considering the level of the concentration of the analytes (above 1 g kg^{-1}). Analytical sensitivity was calculated as $2.1\%^{-1}$, and its inverse, 0.5%, indicates the minimum content difference that can be discriminated by the method, considering the random instrumental noise as the only source of error. This last value also determines the number of decimal places, one, used to express the results. Finally, the results of trueness and precision allow to assure that the method can be considered accurate. Considering the linearity and accuracy studies, the working range of the method was established between 3.2 and 10.9% (v/v) of alcohol content.

Finally, confidence intervals were estimated for the prediction of the validation samples by employing Eq. (1) and the estimate of the standard error for the reference method (0.2%). SPE calculated using the EIV equation were equal to 1.6% (v/v) for all the samples in the model. Figure 6 shows reference and predicted values with the respective confidence intervals (95%) estimated for each validation sample. As can be observed, all the reference values were within the estimated confidence intervals, representing an agreement of 100%.

Conclusions

The combination of a portable NIR spectrometer and multivariate calibration was used to develop and validate a simple and rapid method for directly determining alcohol content in beers. It showed several advantages over the current reference chromatographic method, such as ease of operation, without the need for any complex sample treatment, being

environment friendly and of lower cost. It was validated following international guidelines and was considered linear, accurate, unbiased, and suitable for use as an official methodology for quantifying ethanol in beer. Since this methodology allowed the rapid assessment of one of the most relevant parameters in the beer quality control, thus increasing the frequency of analysis, it has a great potential to increase the cost-effectiveness of the production especially for small business craft breweries that often produce a wide range of styles.

Acknowledgements The authors would like to thank the Brazilian government agencies CNPq, FAPEMIG, and CAPES for financial support. ACCF acknowledges CAPES, and GAGR acknowledges CNPq for PhD scholarships. MCFE acknowledges CNPq for a scientific initiation scholarship. FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) is specially acknowledged for the acquisition of the portable NIR spectrophotometer (Viavi MicroNIR® 1700) through the project APQ03457-16.

Declarations

Ethics Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Consent to Participate Not applicable.

Conflict of Interest Ana Carolina da Costa Fulgêncio declares that she has no conflict of interest. Glaucimar Alex Passos Resende declares that he has no conflict of interest. Marden Claret Fontoura Teixeira declares that he has no conflict of interest. Bruno Gonçalves Botelho declares that he has no conflict of interest. Marcelo Martins Sena declares that he has no conflict of interest.

References

- Abasi S, Minaei S, Jamshidi B, Fathi D (2018) Dedicated non-destructive devices for food quality measurement: a review. *Trends Food Sci Technol* 78:197–205. <https://doi.org/10.1016/j.tifs.2018.05.009>
- Adhikari BR, Schraft H, Chen A (2017) A high-performance enzyme entrapment platform facilitated by a cationic polymer for the efficient electrochemical sensing of ethanol. *Analyst* 142:2595–2602. <https://doi.org/10.1039/c7an00594f>
- Alcalá M, Blanco M, Moyano D, Broad NW, O'Brien N, Friedrich D, Pfeifer F, Siesler HW (2013) Qualitative and quantitative pharmaceutical analysis with a novel hand-held miniature near infrared spectrometer. *J near Infrared Spectrosc* 21:445–457. <https://doi.org/10.1255/jnirs.1084>
- Anyidoho EK, Teye E, Agbemafe R, Amuah CLY, Boadu VG (2021) Application of portable near infrared spectroscopy for classifying and quantifying cocoa bean quality parameters. *J Food Process Pres* 1:e15445. <https://doi.org/10.1111/jfpp.15445>
- AOAC (2016). Appendix F: Guidelines for standard method performance requirements. Rockville AOAC International. http://www.eoma.aoc.org/app_f.pdf. Accessed 10 May 2021.
- ASBC (2015) Beer 4 - alcohol. ASBC methods of analysis. American Society of Brewing Chemists, Saint Paul, pp 1–7
- Baqueta MR, Coqueiro A, Março PH, Valderrama P (2019) Quality control parameters in the roasted coffee industry: a proposal

- by using MicroNIR spectroscopy and multivariate calibration. *Food Anal Methods* 13:1–11. <https://doi.org/10.1007/s12161-019-01503-w>
- Blitz JP (1998) Diffuse reflectance spectroscopy. In: Mirabella FM (ed) *Modern techniques in applied molecular spectroscopy*, 3rd edn. Wiley, New York, pp 185–219
- Botelho BG, Mendes BAP, Sena MM (2013) Development and analytical validation of robust near-infrared multivariate calibration models for the quality inspection control of mozzarella cheese. *Food Anal Methods* 6:881–891. <https://doi.org/10.1007/s12161-012-9498-z>
- Botelho BG, Assis LP, Sena MM (2014) Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages. *Food Chem* 159:175–180. <https://doi.org/10.1016/j.foodchem.2014.03.048>
- Brasil (2009) Dispõe sobre a padronização, a classificação, o registro, a inspeção, a produção e a fiscalização de bebidas; regulamenta a Lei nº 8918, de 14 de julho de 1994 (Decreto nº 6.871, de 04 de junho de 2009). *Diário Oficial da República Federativa do Brasil*, Brasília. http://www.planalto.gov.br/ccivil_03/_Ato2007-2010/2009/Decreto/D6871.htm. Accessed 10 May 2021
- Brewers Association (2020) National beer sales & production data. Boulder, USA. <https://www.brewersassociation.org/statistics-and-data/national-beer-stats/>. Accessed 11 May 2021
- Camps C, Christen D (2009) Non-destructive assessment of apricot fruit quality by portable visible-near infrared spectroscopy. *LWT - Food Sci Technol* 42:1125–1131. <https://doi.org/10.1016/j.lwt.2009.01.015>
- Cinti S, Basso M, Moscone D, Arduini F (2017) A paper-based nanomodified electrochemical biosensor for ethanol detection in beers. *Anal Chim Acta* 960:123–130. <https://doi.org/10.1016/j.aca.2017.01.010>
- Cloninger L (2018) Alcohol determination of malt-based beverages by rapid distillation. *J Am Soc Brew Chem* 76:21–23. <https://doi.org/10.1080/03610470.2017.1396839>
- Crocombe RA (2018) Portable spectroscopy. *Appl Spectrosc* 72:1701–1751. <https://doi.org/10.1177/0003702818809719>
- Donadini G, Fumi MD, Kordialik-Bogacka E, Maggi L, Lambri M, Skokai P (2016) Consumer interest in specialty beers in three European markets. *Food Res Int* 85:301–314. <https://doi.org/10.1016/j.foodres.2016.04.029>
- Erfkamp J, Guenther M, Gerlach G (2019) Hydrogel-based sensors for ethanol detection in alcoholic beverages. *Sensors* 19:1199. <https://doi.org/10.3390/s19051199>
- Faber NM, Song XH, Hopke PK (2003) Sample-specific standard error of prediction for partial least squares regression. *TrAC - Trends Anal Chem* 22:330–334. [https://doi.org/10.1016/S0165-9936\(03\)00503-X](https://doi.org/10.1016/S0165-9936(03)00503-X)
- Fulgêncio ACC, Araújo VPT, Pereira HV, Botelho BG, Sena MM (2020) Development of a simple and rapid method for color determination in beers using digital images. *Food Anal Methods* 13:303–312. <https://doi.org/10.1007/s12161-019-01634-0>
- Kaneko T, Furusho S, Ganaha R, Inui T, Matsuyama A, Mizuno A, Morimoto M, Takemoto K (2007) Comparison of the Anton Paar Alcolyzer method and the official GC-FID method of the National Tax Administration Agency Japan for the evaluation of alcohol content in beer, *happo-shu*, and nonalcoholic beer. *J Am Soc Brew Chem* 65:246–247. <https://doi.org/10.1094/ASBCJ-2007-1004-01>
- Kennard RW, Stone LA (1969) Computer aided design of experiments. *Technometrics* 11:137–148. <https://doi.org/10.2307/1266770>
- Kraftchick JF, Byrd ET, Canziani B, Gladwell NJ (2014) Understanding beer tourist motivation. *Tour Manag Perspect* 12:41–47. <https://doi.org/10.1016/j.tmp.2014.07.001>
- Li H, Takahashi Y, Kumagai M, Fujiwara K, Kikuchi R, Yoshimura N, Amano T, Lin J, Ogawa N (2009) A chemometrics approach for distinguishing between beers using near infrared spectroscopy. *J Near Infrared Spectrosc* 17:69–76. <https://doi.org/10.1255/jnirs.830>
- Lvova L, Paolesse R, Di Natale C, D'Amico A (2006) Detection of alcohols in beverages: an application of porphyrin-based electronic tongue. *Sens Actuators B* 118:439–447. <https://doi.org/10.1016/j.snb.2006.04.044>
- Malegori C, Marques EJM, Freitas ST, Pimentel MF, Pasquini C, Casiraghi E (2017) Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. *Talanta* 165:112–116. <https://doi.org/10.1016/j.talanta.2016.12.035>
- MAPA (2015) Manual de garantia da qualidade analítica. Ministério da Agricultura, Pecuária e Abastecimento, Brasília. <https://www.gov.br/agricultura/pt-br/assuntos/laboratorios/arquivos-publicacoes-laboratorio/manual-de-garantia-qualidade-analitica.pdf/@download/file/manual-de-garantia-qualidade-analitica.pdf>. Accessed 10 May 2021
- Metrohm NIRSystems (2014) A guide to near-infrared spectroscopic analysis of industrial manufacturing processes. Metrohm, Herisau. https://partners.metrohm.com/GetDocumentPublic?action=get_dms_document&docid=1275670. Accessed in 10 May 2021
- Muñiz R, Cuevas-Valdés M, Roza-Delgado B (2020) Milk quality control requirement evaluation using a handheld near infrared reflectance spectrophotometer and a bespoke mobile application. *J Food Compos Anal* 86:103388. <https://doi.org/10.1016/j.jfca.2019.103388>
- Nordon A, Mills A, Burn RT, Cusick FM, Littlejohn D (2005) Comparison of non-invasive NIR and Raman spectrometers for determination of alcohol content of spirits. *Anal Chim Acta* 548:148–158. <https://doi.org/10.1016/j.aca.2005.05.067>
- Ozcam A (2014) Mobile phones democratize and cultivate next generation imaging, diagnostics and measurement tools. *Lab Chip* 14:3187–3194. <https://doi.org/10.1039/c4lc00010b>
- Pérez-Marín D, Paz P, Guerrero JE, Garrido-Varo A, Sánchez MT (2010) Miniature handheld NIR sensor for the on-site non-destructive assessment of post-harvest quality and refrigerated storage behavior in plums. *J Food Eng* 99:294–302. <https://doi.org/10.1016/j.jfoodeng.2010.03.002>
- Rinnan A, van den Berg F, Engelsen SB (2009) Review of the most common pre-processing techniques for near-infrared spectra. *TrAC - Trends Anal Chem* 28:1201–1222. <https://doi.org/10.1016/j.trac.2009.07.007>
- Santos CAT, Lopo M, Páscoa RNJM, Lopes JA (2013) A review on the applications of portable near-infrared spectrometers in the agro-food industry. *Appl Spectrosc* 67:1215–1233. <https://doi.org/10.1366/13-07228>
- Santos CSP, Cruz R, Goncalves DB, Queiros R, Bloore M, Kovacs Z, Hoffmann I, Casal S (2021) Non-destructive measurement of the internal quality of citrus fruits using a portable NIR device. *J AOAC Int* 104:61–67. <https://doi.org/10.1093/jaoacint/qsaa115>
- Souza SVC, Junqueira RG (2005) A procedure to assess linearity by ordinary least squares method. *Anal Chim Acta* 552:25–35. <https://doi.org/10.1016/j.aca.2005.07.043>
- Thompson M, Ellison SLR, Wood R (2002) Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). *Pure Appl Chem* 74:835–855. <https://doi.org/10.1351/pac200274050835>
- Török A, Szerletics A, Jantyik L (2020) Factors influencing competitiveness in the global beer trade. *Sustainability* 12:5957. <https://doi.org/10.3390/su12155957>
- Voss HGJ, Mendes Júnior JJA, Farinelli ME, Stevan SL Jr (2019) A prototype to detect the alcohol content of beers based on an electronic nose. *Sensors* 19:2646. <https://doi.org/10.3390/s19112646>

- Wachelko O, Szpot P, Zawadzki M (2021) The application of head-space gas chromatographic method for the determination of ethyl alcohol in craft beers, wines and soft drinks. *Food Chem* 346:128924. <https://doi.org/10.1016/j.foodchem.2020.128924>
- Williams PC (2001) Implementation of near-infrared technology. In: Williams PC, Norris K (eds) *Near-infrared technology in the agricultural and food industries*. American Association of Cereal Chemists, St. Paul, 145–169
- Williams PC, Sobering DC (1993) Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds. *J near Infrared Spectrosc* 1:25–32. <https://doi.org/10.1255/jnirs.3>
- Yu H, Liu H, Wang Q, van Ruth S (2020) Evaluation of portable and benchtop NIR for classification of high oleic acid peanuts and fatty acid quantitation. *LWT - Food Sci Technol* 128:109398. <https://doi.org/10.1016/j.lwt.2020.109398>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

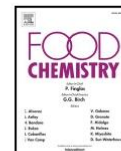
APÊNDICE C

Food Chemistry 391 (2022) 133258



Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem

Screening method for the rapid detection of diethylene glycol in beer based on chemometrics and portable near-infrared spectroscopy

A.C.C. Fulgêncio^a, Glaucimar Alex Passos Resende^a, Marden Claret Fontoura Teixeira^a, Bruno Gonçalves Botelho^a, Marcelo Martins Sena^{a,b,*}

^a Chemistry Department, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil

^b Instituto Nacional de Ciência e Tecnologia em Bioanálítica (INCT Bio), 13083-970 Campinas, SP, Brazil

ARTICLE INFO

Keywords:
Handheld NIRS sensor
Portability
Multivariate analysis
Beer quality control
Beer contamination
Forensic analysis

ABSTRACT

A recent case of contamination of some batches of a Brazilian beer brand with diethylene glycol (DEG) had great repercussion, resulting in at least seven deaths. In this article, a direct method was developed for the rapid detection of DEG in beer samples based on portable near-infrared spectroscopy combined with partial least squares discriminant analysis (PLS-DA). The discriminant model was built with 100 uncontaminated beer samples and 100 samples containing DEG in a concentration range between 10 and 1000 mg L⁻¹, totaling 200 samples of different brands and styles. The method was validated by estimating figures of merit, such as false positive and false negative rates, sensitivity, specificity, accuracy, concordance, and concordance. The decision limit (CC_α) of the method was 52 mg L⁻¹ and the detection capability (CC_β) was 106 mg L⁻¹. This method does not consume reagents/solvents and can be suitable for the beer industry quality control or forensic investigations.

1. Introduction

The principles of food safety ensure access to food and beverage of adequate quality and quantity for healthy eating, besides being a fundamental and universal human right. These principles have become an increasingly important topic nowadays (Prosekov & Ivanova, 2018). As for other industrialized food products, the quality control of a complex beverage such as beer is also imperative (Sileoni, Marconi, & Perreti, 2015; Fox, 2020; Fulgêncio, Araújo, Pereira, Botelho, & Sena, 2019). The development of new analytical methods jointly with the rapid evaluation of their results for the quality control of the final product is an essential tool for improving brewing techniques. Despite all safety protocols adopted in the brewing production, problems caused by contamination may occur due to several reasons, such as accidental failure, lack of acceptable conditions during production or distribution, negligence, malpractices, or even criminal actions.

At the end of 2019, some unusual cases of acute renal failure with neurological alteration were reported in the State of Minas Gerais, Brazil. In January 2020, a local brewery was implicated in the food poisoning of dozens of people related to these previous occurrences, resulting in at least seven deaths. This case had great repercussion in the

Brazilian media, and the brewery production was interrupted. The results of the investigation have elucidated the causes of this tragedy. DEG was employed as antifreeze agent, and a hole in the antifreeze piping of the fermenter tank was detected, through which this substance inadvertently leaked out and contaminated the beer (Goulart, Bordoni, Nascences, & Costa, 2020; Sanchez, Oliveira, Laranjeira, & Caetano, 2020; Lima, Braga, Ventura, & Goulart, 2021).

Diethylene glycol is a clear, practically odorless, colorless, and viscous liquid, with a sweetish taste and soluble in water and ethanol. Its chemical formula is (HOCH₂CH₂)₂O. DEG has numerous industrial uses, mainly as antifreeze agent, chemical intermediary and solvent, but also as heat transfer fluid, brake fluid, cement processing, and lubricant. DEG has a low toxicity for itself, the major problem being its biotransformation in the liver leading to highly toxic metabolites, such as 2-hydroxyethoxyacetic (2-HEAA) and diglycolic acid (DGA) (and possible other yet unknown metabolites). These metabolites may persist in the body for a long time and cause acute toxic syndrome, resulting in renal failure due to cortical tubular degeneration and proximal tubular necrosis. The minimum dose for DEG toxicity in humans has not been clearly established, but the range of doses reported as lethal for DEG has varied from 0.5 to 5 g kg⁻¹ (Schep, Slaughter, Temple, & Beasley, 2009; Landry,

* Corresponding author at: Departamento de Química, ICEx, Universidade Federal de Minas Gerais, 31270-901 Belo Horizonte, MG, Brazil.

E-mail address: marcsen@ufmg.br (M.M. Sena).

<https://doi.org/10.1016/j.foodchem.2022.133258>

Received 14 February 2022; Received in revised form 12 May 2022; Accepted 17 May 2022

Available online 20 May 2022

0308-8146/© 2022 Elsevier Ltd. All rights reserved.

Martin, & McMartin, 2011). Unlike pharmaceutical products and excipients, there is no maximum allowed limit for DEG in food and beverage (Caldeira, Madureira, Maia, Muller, & Fernandes, 2021).

DEG poisoning has been related to incidents involving matrices such as pharmaceuticals, healthcare products and alcoholic beverages. Several analytical methods have been reported in the literature for determining DEG in these types of matrices, mainly based on chromatographic techniques. The most used technique has been gas chromatography with flame ionization detection (GC-FID). GC-FID has been employed to determine DEG in pharmaceutical products (Baffi, Elneser, Baffi, & Melin, 2000), toothpaste (Holloway, Maheswaran, Leeks, Bradby, & Wahab, 2010), food items such as soft drinks, juice, infant formula, cereal, flour and snacks (Rahim et al., 2011), and wine (Lawrence, Chadha, Lau, & Weber, 1986). This last method has been developed because of a real case in which DEG has been detected at levels exceeding 100 mg L^{-1} in wines; the addition of DEG to wine would be made aiming to improve its sweet taste. Ultra-performance liquid chromatography–time of flight mass spectrometry (UHPLC-MS) has been used to determine DEG in toothpaste (Hernández, Ibáñez, & Sancho, 2008) and GC-MS to determine the same analyte in human plasma (Maurer, Peters, Paul, & Kraemer, 2001). GC-MS has also been very recently utilized to detect and quantify DEG in beer by a Brazilian research group, reverberating the contamination occurrence in Minas Gerais (Caldeira et al., 2021).

Despite their advantages, chromatographic methods present some drawbacks, requiring costly, laborious and time-consuming analysis, consuming organic solvents, generating chemical waste, and demanding previous sample preparation steps, such as derivatization, clean up and preconcentration. On the other hand, the emergence of green chemistry has brought up concerns about the development of new more environmentally friendly analytical methods. Vibrational spectroscopic techniques have lower sensitivity in comparison with chromatography, but represent a green, simple and more rapid alternative to develop screening methods. The concomitant use of chemometrics is almost mandatory. Thus, vibrational techniques have been combined with chemometric models, mainly partial least squares (PLS), to detect and quantify DEG in several matrices. DEG has been determined in glycerin with Raman spectroscopy (Gryniewicz-Ruzicka et al., 2011), in glycerin-based cough syrup with mid and near infrared spectroscopies (Ahmed, McLeod, Nézivar, & Giuliani, 2010), and in toothpaste and gel dentifrice with attenuated total reflectance Fourier transform infrared (ATR-FTIR) spectroscopy (López-Sánchez, Domínguez-Vidal, Ayora-Cañada, & Molina-Díaz, 2008).

Portability is an important trend in spectroscopy, allowing simple, robust and real-time analysis (Crocombe, 2018). Portable near infrared (NIR) spectrometers have sensitivity and performance similar to that of benchtop equipments, showing other advantages such as small size, lower energy consumption and absence of moving parts in their optical structure. These characteristics improve the usefulness of NIR spectroscopy as an alternative to chromatographic techniques for the quality control of food and beverage. Particularly for brewery production, handheld NIR spectrometers can be viewed as sensors useful to establishing more control of the whole process, reducing production costs, and improving the confidence of the final product. In addition, other parameters can be monitored in the scope of process analytical technology (PAT) (Sileoni et al., 2015; Fox, 2020). It can be used for instance to determine alcohol content (Fulgêncio, Resende, Teixeira, Botelho, & Sena, 2022) and color (Fulgêncio et al., 2019) of beers. Other liquid samples have also been checked for authenticity and quality control with portable devices, such as fresh juice (Chen et al., 2021), palm (Teye, Elliot, Sam-Amoah, & Mingle, 2019) and coriander oils (Kaufmann, Sampaio, García-Martín, & Barbin, 2022). Considering the points previously discussed, the scientific question approached in this article is to evaluate if a handheld NIR device can be used to detect DEG contamination in beer.

Thus, the aim of this study was to combine portable NIR spectroscopy

and PLS discriminant analysis (PLS-DA) to develop and validate a rapid and simple screening method for the direct detection of diethylene glycol in beer. Aiming to obtain a robust model, beers from different styles, alcohol contents, brands, and breweries were used to build the multivariate model. The proposed method was validated incorporating multivariate aspects through the estimate of proper qualitative figures of merit (FOM) (Botelho, Reis, Oliveira, & Sena, 2015; The Commission of the European Communities, 2012; Gondim, Junqueira, Souza, Pilar Callao, & Ruisánchez, 2017; Isabel López, Pilar Callao, & Ruisánchez, 2015). FOM related to concentration were also estimated by fitting a probability of detection (POD) curve with the PLS-DA outputs (Gondim et al., 2017; Isabel López et al., 2015).

2. Materials and methods

2.1. Instruments and software

NIR spectra were collected using a portable MicroNIR® 1700 spectrophotometer from Viavi Solution (Milpitas, CA, USA), with a wavelength working range between 900 and 1700 nm. One of the main components of this portable equipment is the linear variable filter (LVF), a part responsible for passing frequencies within the working range and rejecting or attenuating frequencies outside this range. Since the center wavelength (CWL) is chosen as a function of coating thickness, the peak wavelength transmitted through the filter will vary linearly in the wedge direction. Therefore, specifically for the MicroNIR®, wavelengths are selected by adjusting the filter to the appropriate linear position. Thus, the LVF is arranged above the InGaAs array detector. The detector is a variable-band semiconductor with excellent optical properties. The diffusely reflected radiation was collected and sent to a computer. Data were processed using MATLAB® software, version 7.13 (MathWorks, Natick, USA), coupled with the PLS Toolbox, version 6.5 (Eigenvector Technologies, Manson, USA).

2.2. Beer samples

One hundred beer samples were purchased mainly in the local commerce of the Metropolitan Region of Belo Horizonte, Minas Gerais, Brazil, and stored at room temperature until the analysis. These beer samples were purchased from a great variety of brands, breweries, alcohol contents and types, including lager, pale ale, red ale, Vienna lager, Pilsen, IPA, weissbier, strong pale ale, porter, and stout. All samples were previously checked for the absence of DEG using a chromatographic methodology described in the next section.

2.3. GC-MS method for assuring the absence of DEG in non-contaminated beers

A chromatographic method was developed for checking the absence of DEG in non-contaminated beers. Beer samples were sonicated for 5 min, diluted ten times with methanol (HPLC grade), and an aliquot of 100 μL was transferred to a 2 mL Eppendorf. Then, the sample was centrifuged for 5 min at 3000 rpm, and transferred to 2 mL vials, before the injection in the chromatograph. A Shimadzu GCMS-QP2010 Plus system (Kyoto, Japan) was utilized jointly with an Innovax column (60 $\text{m} \times 0.25 \text{ mm} \times 0.25 \mu\text{m}$). The injection volume was 1.0 μL in the splitless mode, the temperatures of the injection port and the detector were set at 250 °C, and helium was used as carrier gas in a flow of 1.0 mL min^{-1} . The ions of m/z 43 and 45, characteristics of DEG, were monitored. Oven temperature was kept at 100 °C for 1 min, increased to 200 °C at a rate of 10 °C/min, kept at 200 °C for 10 min, increased to 240 °C at a rate of 10 °C/min, and kept at 240 °C for 5 min.

2.4. Sample preparation and spectra acquisition

All the reagents were of analytical grade and used as received. A

volume of 10 mL of each beer sample was degassed using an ultrasonic bath (Equilab ULTRASONIK 28H, Madrid, Spain) for 5 min. One drop of 1-octanol was added to each beer sample to avoid excessive foaming. This step is not feasible for direct application in the industrial brewing process, but it can be adapted for an at-line monitoring. One hundred original beer samples were selected in random order and spiked with DEG in a concentration range from 10 to 1000 mg L⁻¹, with an increment of 20 mg L⁻¹ between samples. A volume of 2.0 mL of each sample contaminated with DEG was transferred to a 2 mL Eppendorf tube. For non-contaminated samples, 2.0 mL of each original beer sample was placed in a 2 mL Eppendorf tube without any additional preparation. All the Eppendorf tubes were placed in an acrylonitrile-butadienestyrene support especially designed and constructed for this application with a 3D printer (Creality Ender 3, Shenzhen, China), which helped to collect the spectra of the two hundred samples with the portable MicroNIR® spectrophotometer. Spectra were recorded in the diffuse reflectance mode, with twenty scans each. The wavelength range was from 908 to 1676 nm with a resolution of 6.25 nm. During all these measurements, the temperature of the laboratory was controlled at 25 ± 1 °C. The NIR spectrum of a sample of pure DEG was also recorded.

2.5. Qualitative analytical validation

The analytical validation of the developed model was assessed by estimating appropriate qualitative FOM, such as false-negative rate (FNR), false-positive rate (FPR), sensitivity rate (STR), specificity rate (SPR), accuracy, accordance (ACC), concordance (CON), detection capability (CC β), and decision limit (CC α). Accuracy is a FOM that expresses globally the rate of incorrect predictions regardless they are positive or negative. ACC and CON express the precision of qualitative methods, being equivalent to quantitative repeatability and intermediate precision, respectively. These FOM have been initially proposed for univariate methods (Langton, Chevenement, Nagelkerke, & Lombard, 2002) and can be directly extended to multivariate methods (Botelho, et al., 2015). ACC was estimated using ten replicates at five different levels (50, 80, 100, 400, and 1000 mg L⁻¹) analyzed in the same batch under repeatability conditions (the same day and the same analyst). CON was estimated using the same protocol employed for ACC, repeated at two different days by two different analysts. CC α and CC β were estimated from POD curves, which harmonize the statistical concepts and parameters between quantitative and qualitative method validation. This type of curve presents graphically the relation between the probability of positive results obtained from qualitative methods, *i.e.* the outcomes of classification models, and the analyte concentration. The probability values are calculated from independent measurements at each level of adulteration (Isabel López et al., 2015; Gondim et al., 2017). For estimating CC α and CC β , ten fortified sample replicates were prepared at ten different concentration levels (blank, 10 mg L⁻¹, 20 mg L⁻¹, 40 mg L⁻¹, 60 mg L⁻¹, 80 mg L⁻¹, 100 mg L⁻¹, 400 mg L⁻¹, 700 mg L⁻¹ and 1000 mg L⁻¹).

3. Results and discussion

3.1. PLS-DA model building

Beer is a complex matrix. Its main ingredients include water, which is a critical component in the brewing process, as well as hops and yeast, besides other additives such as fruits, herbs, and other plants. Fig. 1 shows the preprocessed NIR spectra, in the wavelength range between 908 and 1676 nm, of the 200 analyzed beer samples, 100 contaminated with DEG and 100 non-contaminated. The absence of DEG in the non-contaminated beer samples was verified by GC-MS (section 2.3). By observing these spectra, it is possible to highlight the two most intense absorption bands, one centered around 1200 nm, and other one centered around 1400 nm. The first spectral band, in the region between 1100 and 1250 nm, may be assigned to the second overtone stretching of CH₂

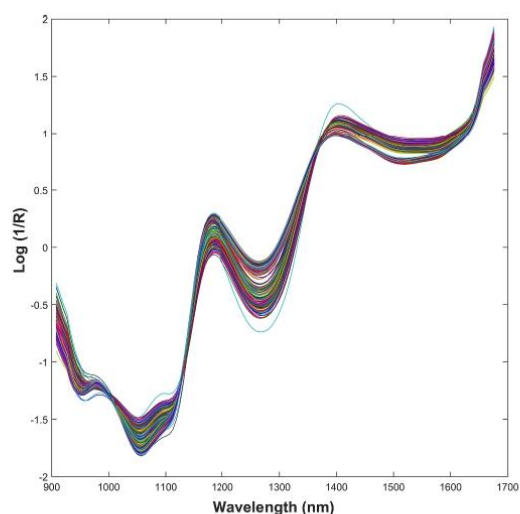


Fig. 1. NIR spectra preprocessed by Savitzky-Golay smoothing and standard normal variate (SNV) of all the 200 analyzed beer samples.

and CH₃ bonds; the second band might be related to the first overtone of the OH stretching (Metrohm NIRSystems, 2014). More specifically, the first mentioned band can be assigned to C–H stretching overtone of ethanol, while the spectral band at higher wavelengths can be related to O–H bonds of ethanol and water (Fulgêncio et al., 2019). Spectral profiles of all the samples are similar, thus preventing the possibility of discrimination between contaminated and non-contaminated beers by visual inspection. Considering the limited selectivity of NIR spectroscopy, the concomitant use of chemometrics to extract information is imperative. Thus, a multivariate supervised classification PLS-DA model was built.

For building the PLS-DA model, samples were split in training and test sets, in the ratio of approximately two thirds:one third. Kennard-Stone algorithm was used for selecting training samples from each of the two classes (Kennard & Stone, 1969). Thus, 134 training samples were selected, whereas the remaining 66 samples were used to construct the test set. Aiming to eliminate drifts typical of diffuse reflectance infrared spectra and caused by multiplicative light scattering, data was sequentially preprocessed by Savitzky-Golay smoothing (7 points and second order polynomial fit), standard normal variate (SNV), and mean centering (Rinnan, van den Berg, & Engelsen, 2009).

Non-contaminated samples were trained with *y* values equal to 1.0, while samples contaminated with DEG were associated to 0.0. Random subsets cross-validation (6 splits and 20 iterations) was applied to select the number of latent variables (LV), according to the smallest cross-validation classification error (CVCE). The best PLS-DA model was selected with 6 LV after optimization by outlier detection, accounting for 99.63% of the spectral variance (X block) and 71.41% of the dummy variables variance (Y block). In this study, outlier detection was performed based on leverage and Q (spectral) residues, both at 95% confidence level. The number of outliers cannot exceed the recommended limit of 2/9 (22.2 %) for both training and test sets (MAPA, 2014; Thompson, Ellison, & Wood, 2002). Twelve out of the original two hundred samples were removed as outliers, seven from the training set (5.2 %) and five from the validation set (7.6 %). After the removal of outliers, the model was rebuilt. The results for the optimization of this PLS-DA model, including STR and SPR for cross-validation, in addition to CVCE, are shown in Table 1. The Bayesian threshold was estimated at 0.51 and the plot of individual *y* predicted values is shown in Fig. 2.

Table 1

Results for the optimization by outlier detection of the PLS-DA model built for the detection of diethylene glycol in beer.

		Number of training samples	Number of test samples	Number of latent numbers	Sensitivity (CV) ^a	Specificity (CV) ^a	Cross-validation classification error (CVCE)
Before	DEG-Free	67	33	6	0.851	0.834	0.158
	DEG	67	33				
After	DEG-Free	62	29	6	0.956	0.927	0.059
	DEG	65	32				

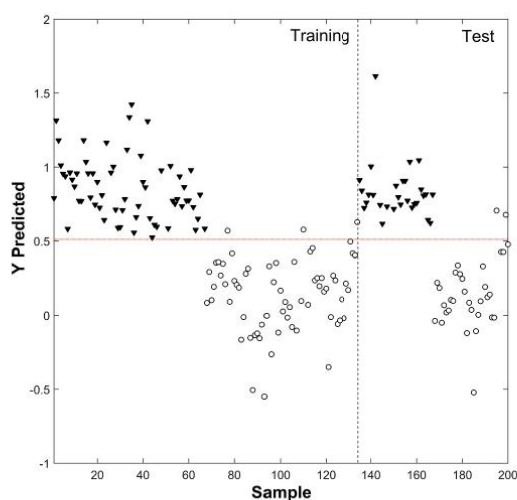
^a Cross-validation.

Fig. 2. Y predicted values for the PLS-DA model built to discriminate DEG contaminated and non-contaminated beer samples. Horizontal dashed line indicates the threshold and vertical dashed line indicates the separation between training and test samples. Full down triangles represent non-contaminated samples and empty circles represent samples contaminated with DEG.

3.2. Spectral interpretation of the PLS-DA model

Qualitative or quantitative chemometric models should be corroborated through the spectral interpretation of their informative vectors. The estimated vectors of regression coefficients and variable importance in projection (VIP) scores for the developed PLS-DA model are shown in Fig. 3. VIP scores (Fig. 3a) indicate the variables that most contribute to the model in absolute values (Chong & Jun 2005). Usually, variables with VIP scores higher than 1.0 are considered the most significant for the model. Since these parameters present absolute values, they should be evaluated as contributing to the model as whole and cannot be directly associated with one class. Thus, the spectral interpretation should be complemented by inspecting the regression vectors (Fig. 3b), because they can indicate to which class each variable is more related. Variables with the most positive regression coefficients contribute to the characterization of the class 1 (in this study, non-contaminated beer samples), while variables with the most negative regression coefficients characterize class 0 (beer samples contaminated with DEG). The most discriminant variables do not necessarily coincide with the most intense absorptions. Aiming to complement this spectral interpretation, the NIR spectrum of a pure DEG sample is shown in Fig. S1 (Supplementary Material), jointly with the loadings plot of LV2 in Fig. S2. By observing spectral features, LV2 was assigned to the DEG contribution to the PLS-DA model.

By observing Fig. 3, it was possible to identify the NIR spectral regions that most contribute to the detection of DEG in beer. There are three negative regions of regression coefficients, which are related to the presence of DEG. The most intense and sharp peak, at 945 nm, is related to the third overtone of CH₂ vibrations (Metrohm NIRSystems, 2014) and may be associated with the chemical structure of DEG. The large band of regression coefficients centered at 1484 nm corresponds to an intense and broad absorption band of the NIR spectrum of DEG (Yin, Zhang, Li, & Jin, 2014; Li, Arzhantsev, Kauffman, & Spencer, 2011) and can be assigned to the first overtone of its OH bonds (Ahmed et al., 2010). Finally, the band centered at 1150 nm, which also presents the highest VIP scores, can be assigned to the second overtone of CH₂ bonds and corresponds to a band of lower absorption in the NIR spectrum of DEG (Yin et al., 2014; Li et al., 2011). All these spectral assignments can also be corroborated by observing Figs. S1 and S2. By contrast, the most positive regions of regression coefficients, which also correspond to some of the highest VIP scores, are centered at 1013 and 1410 nm. They can be assigned to the second and first overtones of OH stretching (Metrohm NIRSystems, 2014), which can be attributed to constituents of beer, such as ethanol and water. These signals are direct interferences for the detection of DEG.

3.3. Analytical validation of the PLS-DA model

The analytical validation of the developed method was performed through the estimate of proper FOM. The values for seven different qualitative FOM are shown in Table 2. FNR, FPR, STR, SPR and accuracy were estimated for both training and test sets, while FOM related to precision, ACC and CON, were estimated only for the test set. PLS-DA model did not provide any false negative prediction, leading to STR of 100% for both training and test sets. This means that DEG was correctly not detected in all the non-contaminated samples. On the other hand, three false positive predictions for the training set and two false positives for the test set were observed. These results in SPR of 95.5% and 93.9% for training and test sets, respectively. All these false positives corresponded to samples contaminated with the lowest concentration levels of DEG, thus pointing to a limitation in the sensitivity of the method, which might be related to the limits of the analytical technique, NIR spectroscopy (Pasquini, 2018). Considering the observed limitation of the method to detect the lowest DEG concentrations, FOM related to precision were estimated at five levels above 50 mg L⁻¹ (50, 80, 100, 400, and 1000 mg L⁻¹). ACC corresponds to qualitative intra-run precision, while CON is related to inter-run precision. For DEG concentration at 50 mg L⁻¹, the smallest ACC and CON values were obtained, below 50%. For levels equal or above 100 mg L⁻¹, no classification errors were observed and both precision FOM were estimated as 100%.

The results described above indicate the lack of precision of the method at DEG concentration below 100 mg L⁻¹. This reinforces the need to establish quantitative limits for this qualitative method through the estimate of proper FOM, namely detection capability (CC_β) and decision limit (CC_α). These two FOM were calculated by a POD curve estimated by a logistic fit with the outputs of the PLS-DA model for DEG contaminated class (Isabel López et al., 2015; Gondim et al., 2017). This POD curve is shown in Fig. 4, and CC_α and CC_β were estimated as 52 mg

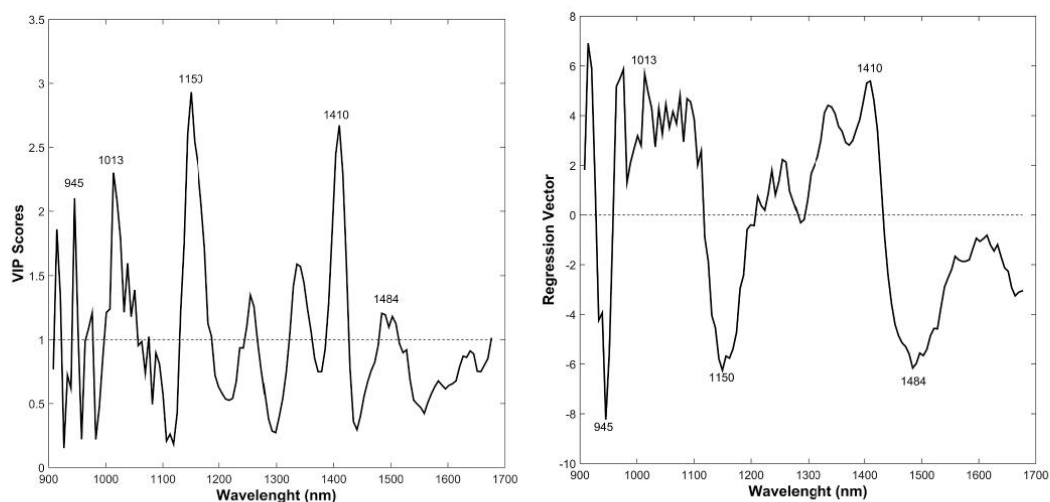


Fig. 3. Informative vectors for the PLS-DA model. (a) Variable importance in projection (VIP) scores. Black horizontal dotted line indicates the threshold of 1.0, above which variables are considered to significantly contribute to the model. (b) Regression vector.

Table 2

Estimated figures of merit for the qualitative analytical validation of the PLS-DA model. Units in %.

	FPR ^a	FNR ^b	STR ^c	SPR ^d	Accuracy	ACC ^e	CON ^f												
							50	80	100	400	1000	50	80	100	400	1000			
Training	4.5	0	100	95.5	95.5	–	–	–	–	–	–	–	–	–	–	–	–	–	–
Test	6.1	0	100	93.9	93.9	44	80	100	100	100	40	81	100	100	100	100	100	100	

^a False positive rate.

^b False negative rate.

^c Sensitivity rate.

^d Specificity rate.

^e Accordance for concentration levels in mg L⁻¹.

^f Concordance for concentration levels in mg L⁻¹.

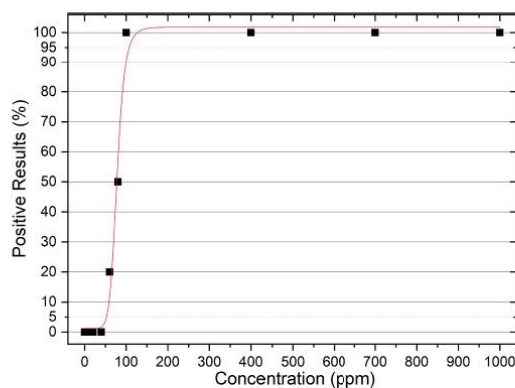


Fig. 4. Probability of detection (POD) curve obtained by logistic regression, with positive results in percentage and concentration in mg L⁻¹.

L⁻¹ and 106 mg L⁻¹, respectively. CC α particularly indicates the concentration limit at which the developed method detects the contaminant with a 5% of error chance of stating that the contaminant is present

when in fact it is not. Therefore, the developed method cannot be considered reliable for detecting DEG contamination in beer below 52 mg L⁻¹ and its analytical range should be restricted. Naturally, the present methodology based on NIR spectroscopy and chemometrics has lower sensitivity than chromatographic methods. For example, a recently published article developed a GC-MS method for quantifying DEG in beer, which presented a limit of detection of 5.0 mg L⁻¹ and a limit of quantification of 10.0 mg L⁻¹ (Caldeira et al., 2021). However, our proposed alternative has several advantages and a very good cost-benefit ratio as a screening method, since it can be implemented in a portable analytical platform for on-site analysis with very reduced costs.

4. Conclusions

A simple, rapid and direct screening method for detecting the toxic diethylene glycol in beer based on a portable near-infrared spectrometer was developed. This is of special interest considering the recent occurrence in Brazil of diethylene glycol poisoning caused by beer contamination, which resulted in some deaths and dozens of intoxicated people. For the developed PLS-DA model, the decision limit (CC α) was estimated as 52 mg L⁻¹. The method was validated according to Brazilian and international guidelines (MAPA, 2014; The Commission of the European Communities, 2012), being considered accurate and precise. It was also considered robust, incorporating the variance from beers of different styles and origins that were used to build the model. The proposed

method required a small amount of sample (2 mL), spent only 30 s per spectral acquisition, did not use reagents or solvents nor generate chemical waste, and did not demand any sample preparation in addition to degassing. Finally, this methodology can be adapted to portable analytical platforms and implemented at a low cost for the quality control in brewery industry or be used as a rapid screening option for on-site forensic analysis in cases of suspected intoxication/poisoning caused by beer.

Funding

This work was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), grant numbers APQ03457-16 and RED00042-16; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), grant number 1383/2020.

CRedit authorship contribution statement

A.C.C. Fulgêncio: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Glaucimar Alex Passos Resende:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation. **Marden Claret Fontoura Teixeira:** Data curation. **Bruno Gonçalves Botelho:** Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Project administration, Funding acquisition. **Marcelo Martins Sena:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank the Brazilian government agencies CNPq, FAPEMIG, and CAPES for financial support. ACCF acknowledges CAPES and GAPP acknowledges CNPq for Ph.D. scholarships. MCFT acknowledges CNPq for a scientific initiation scholarship. FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) is especially acknowledged for the acquisition of the portable NIR spectrophotometer (Viavi MicroNIR® 1700) through the project APQ03457-16. FAPEMIG is also acknowledge for the financial support of the project "Rede Mineira de Ciências Forenses", RED00042-16. The present work was performed with the financial support of "Programa de Cooperação Acadêmica em Segurança Pública e Ciências Forenses" (Edital CAPES 16/2020).

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.foodchem.2022.133258>.

References

- Ahmed, M. K., McLeod, M. P., Nézar, J., & Giuliani, A. W. (2010). Fourier transform infrared and near-infrared spectroscopic methods for the detection of toxic diethylene glycol (DEG) contaminant in glycerin based cough syrup. *Journal of Spectroscopy*, 24, 601–608.
- Baffi, P., Elneser, S., Baffi, M., & Melin, M. D. (2000). Quantitative determination of diethylene glycol contamination in pharmaceutical products. *Journal of AOAC International*, 83, 793–801.
- Botelho, B. G., Reis, N., Oliveira, L. S., & Sena, M. M. (2015). Development and analytical validation of a screening method for simultaneous detection of five adulterants in raw milk using mid-infrared spectroscopy and PLS-DA. *Food Chemistry*, 181, 31–37.
- Caldeira, L. R., Madureira, F. D., Maia, T. D. F., Muller, C. V., & Fernandes, C. (2021). Simultaneous quantification of ethylene glycol and diethylene glycol in beer by gas chromatography coupled to mass spectrometry. *Food Chemistry*, 346, Article 128871.
- Chen, W., Li, H., Zhang, F., Xiao, W., Zhang, R., Chen, Z., & Du, Y. (2021). Handheld short-wavelength NIR spectroscopy for rapid determination of sugars and carbohydrate in fresh juice with sampling error profile analysis. *Infrared Physics & Technology*, 115, Article 103732.
- Chong, I.-G., & Jun, C. H. (2005). Performance of some variable selection methods when multicollinearity is present. *Chemometrics and Intelligent Laboratory Systems*, 78, 103–112.
- Crocombe, R. A. (2018). Portable spectroscopy. *Applied Spectroscopy*, 72, 1701–1751.
- Fox, G. (2020). The brewing industry and the opportunities for real-time quality analysis using infrared spectroscopy. *Applied Sciences*, 10, 616.
- Fulgêncio, A. C. C., Araújo, V. P. T., Pereira, H. V., Botelho, B. G., & Sena, M. M. (2019). Development of a simple and rapid method for color determination in beers using digital images. *Food Analytical Methods*, 13, 303–312.
- Fulgêncio, A. C. C., Resende, G. A. P., Teixeira, M. C. F., Botelho, B. G., & Sena, M. M. (2022). Determination of alcohol content in beers of different styles based on portable near-infrared spectroscopy and multivariate calibration. *Food Analytical Methods*, 15, 307–316.
- Gondim, C. S., Junqueira, R. G., Souza, S. V. C., Pilar Calhao, M., & Ruisánchez, I. (2017). Determining performance parameters in qualitative multivariate methods using probability of detection (POD) curves. Case study: Two common milk adulterants. *Talanta*, 168, 23–30.
- Goulart, C. O. L., Bordoni, L. S., Nascentes, C. C., & Costa, L. M. (2020). Analysis of diglycolic acid after mass poisoning by diethylene glycol. *Journal of Analytical Toxicology*, 1–5. <https://doi.org/10.1093/jat/bkaa187>
- Gryniewicz-Ruzicka, C. M., Arzhantsev, S., Pelster, L. N., Westenberger, B. J., Buhse, L. F., & Kauffman, J. F. (2011). Multivariate calibration and instrument standardization for the rapid detection of diethylene glycol in glycerin by Raman spectroscopy. *Applied Spectroscopy*, 63, 334–341.
- Hernández, F., Ibáñez, M., & Sancho, J. V. (2008). Fast determination of toxic diethylene glycol in toothpaste by ultra-performance liquid chromatography–time of flight mass spectrometry. *Analytical and Bioanalytical Chemistry*, 391, 1021–1027.
- Holloway, G., Maheswaran, R., Leeks, A., Bradby, S., & Wahab, S. (2010). Screening method for ethylene glycol and diethylene glycol in glycerin-containing products. *Journal of Pharmaceutical and Biomedical Analysis*, 51, 507–511.
- Isabel López, M., Pilar Callao, M., & Ruisánchez, I. (2015). A tutorial on the validation of qualitative methods: From the univariate to the multivariate approach. *Analytica Chimica Acta*, 891, 62–72.
- Kaufmann, K. C., Sampaio, K. A., García-Martín, J. F., & Barbin, D. F. (2022). Identification of coriander oil adulteration using a portable NIR spectrometer. *Food Control*, 132, Article 108536.
- Kennard, R. W., & Stone, L. A. (1969). Computer aided design of experiments. *Technometrics*, 11, 137–148.
- Landry, G. M., Martin, S., & McMartin, K. E. (2011). Diglycolic acid is the nephrotoxic metabolite in diethylene glycol poisoning inducing necrosis in human proximal tubule cells in vitro. *Toxicological Sciences*, 124, 35–44.
- Langton, S., Chevennement, R., Nagelkerke, N., & Lombard, B. (2002). Analysing collaborative trials for qualitative microbiological methods: Accordance and concordance. *International Journal of Food Microbiology*, 79, 175–181.
- Lawrence, J. F., Chadha, R. K., Lau, B.-P.-Y., & Weber, D. F. (1986). Simplified routine method for the determination of diethylene glycol in wines by capillary gas chromatography with flame ionization detection. *Journal of Chromatography*, 367, 213–216.
- Li, X., Arzhantsev, S., Kauffman, J. F., & Spencer, J. A. (2011). Detection of diethylene glycol adulteration in propylene glycol - Method validation through a multi-instrument collaborative study. *Journal of Pharmaceutical and Biomedical Analysis*, 54, 1001–1006.
- Lima, C. M. D., Braga, L., Ventura, C. P., & Goulart, C. O. L. (2021). Diethylene glycol poisoning: A 1937 world retrospective study – From 1937 to the 2020 brewery case in Minas Gerais, Brazil. *Revista de Criminalística e Medicina Legal*, 6, 27–33.
- López-Sánchez, M., Domínguez-Vidal, A., Ayora-Cañada, & Molina-Díaz (2008). Assessment of dentifrice adulteration with diethylene glycol by means of ATR-FTIR spectroscopy and chemometrics. *Analytica Chimica Acta*, 620, 113–119.
- MAPA. (2014). Manual de garantia da qualidade analítica. Ministério da Agricultura, Pecuária e Abastecimento, Brasília, Brazil. Retrieved September 28, 2021, from <https://www.gov.br/agricultura/pt-br/assuntos/laboratorios/arquivos-publicacoes-laboratorio/manual-de-validacao-ia-e-iqi-nov2014.pdf>. Accessed in February 2022.
- Maurer, H. H., Peters, F. T., Paul, L. D., & Kraemer, T. (2001). Validated gas chromatographic-mass spectrometric assay for determination of the antifreezes ethylene glycol and diethylene glycol in human plasma after microwave-assisted pivalylation. *Journal of Chromatography B*, 754, 401–409.
- Metrohm NIRSystems. (2014). A guide to near-infrared spectroscopic analysis of industrial manufacturing processes. Metrohm, Herisau. Available in: <https://www.metrohm.com/en/documents/81085026>. Accessed in February 2022.
- Pasquini, C. (2018). Near infrared spectroscopy: A mature analytical technique with new perspectives – A review. *Analytica Chimica Acta*, 1026, 8–36.
- Prosekov, A. Y., & Ivanova, S. A. (2018). Food security: The challenge of the present. *Geoforum*, 91, 73–77.
- Rahim, A. A., Saad, B., Osman, H., Hashim, N., Yahya, S., & Talib, K. M. (2011). Simultaneous determination of diethylene glycol, diethylene glycol monoethyl ether, coumarin and caffeine in food items by gas chromatography. *Food Chemistry*, 126, 1412–1416.
- Rinnan, A., van den Berg, F., & Engelsen, S. B. (2009). Review of the most common pre-processing techniques for near-infrared spectra. *Trends in Analytical Chemistry*, 28, 1201–1222.

A.C.C. Fulgêncio et al.

Food Chemistry 391 (2022) 133258

- Sanchez, Z. M., Oliveira, C. L., Laranjeira, R., & Caetano, R. (2020). Intoxication and deaths from diethylene glycol present in Brazilian craft beers. *Journal of Studies on Alcohol and Drugs*, 81, 522–523.
- Schep, L. J., Slaughter, R. J., Temple, W. A., & Beasley, D. M. G. (2009). Diethylene glycol poisoning. *Clinical Toxicology*, 47, 525–535.
- Sileoni, V., Marconi, O., & Perreti, G. (2015). Near-infrared spectroscopy in the brewing industry. *Critical Reviews in Food Science and Nutrition*, 55, 1771–1791.
- Teye, E., Elliott, C., Sam-Amoah, L. K., & Mingle, C. (2019). Rapid and nondestructive fraud detection of palm oil adulteration with Sudan dyes using portable NIR spectroscopic techniques. *Food Additives & Contaminants: Part A*, 36, 1589–1596.
- The Commission of the European Communities. (2012). Commission Decision of 12 August 2002 implementing Council Directive 96/23/EC concerning the performance of analytical methods and the interpretation of results. *Official Journal of the European Communities*, 50, 1–22. <https://op.europa.eu/en/publication-detail/-/publication/ed928116-a955-4a84-b10a-cf7a82bad858/language-en>. Accessed in February 2022.
- Thompson, M., Ellison, S. L. R., & Wood, R. (2002). Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report). *Pure and Applied Chemistry*, 74, 835–855.
- Yin, L., Zhang, X., Li, X., & Jin, S. (2014). Development and implementation of incident response near-infrared models for analyzing contaminated medicines containing diethylene glycol solvent. *Journal of Innovative Optical Health Sciences*, 7, 1450035.

APÊNDICE D

Talanta Open 7 (2023) 100180



Contents lists available at ScienceDirect

Talanta Open

journal homepage: www.sciencedirect.com/journal/talanta-open



Combining portable NIR spectroscopy and multivariate calibration for the determination of ethanol in fermented alcoholic beverages by a multi-product model

Ana Carolina da Costa Fulgêncio^a, Glaucimar Alex Passos Resende^a,
Marden Claret Fontoura Teixeira^a, Bruno Gonçalves Botelho^a, Marcelo Martins Sena^{a,b,*}

^a Chemistry Department, Universidade Federal de Minas Gerais, Belo Horizonte, MG 31270-901, Brazil

^b Instituto Nacional de Ciência e Tecnologia em Bioanalítica (INCT Bio), Campinas, SP 13083-970, Brazil

ARTICLE INFO

Keywords:

Alcoholic strength
Handheld NIR sensor
Portability
Chemometrics
Miniaturization
Direct analysis

ABSTRACT

In this study, a multivariate calibration multi-product model was built by combining partial least square regression (PLS) and portable near infrared (NIR) spectroscopy for the determination of ethanol content in fermented alcoholic beverages. Reference values were obtained by gas chromatography with flame ionization detection (GC-FID). Aiming at building a robust model, a great variety of beers, ciders, meads, and wines were incorporated into the model. NIR spectra were recorded between 908 and 1676 nm for 153 alcoholic beverage samples, corresponding to a range from 4.3 to 15.3% (v/v) of alcohol content. PLS model provided accurate results with root mean square errors of calibration (RMSEC) and prediction (RMSEP) of 0.8% and 0.9%, respectively. The developed method was validated through the estimate of proper figures of merit, such as linearity, trueness, precision, analytical sensitivity, bias, and residual prediction deviation (RPD). This method was simple, direct, rapid, of low-cost and environmentally friendly, not consuming reagents or solvents nor generating chemical waste. It could be incorporated in analytical platforms for quality inspection, contributing to provide better transparency in the food supply chain.

1. Introduction

The expanding global market of alcoholic beverages has increased the demand for their quality control and inspection. One of the main parameters for this control is the ethanol content, whose concentration characterizes the product and may represent its identity. This analytical demand is linked to the quality assurance in the food supply chain, taxation and declaration of ethanol content on the label [1]. From the forensic point of view, it is also important to verify whether a beverage was adulterated or counterfeit [2]. Some countries classify alcoholic beverages according to their ethanol content. The legal category of a beverage can impact the type of license required to manufacture, distribute, and sell the product; it may also influence on taxation [3,4]. Many countries have high excise duties for fermented alcoholic drinks, and the incorrect declaration of alcoholic strength is one of the fraud categories found in official regulations [5,6]. Fermented drinks usually have up to 16% alcohol by volume, while distilled beverages have much higher contents, between 30 and 60% [7]. Beer and wine are among the

most consumed beverages in the world, being the two most consumed alcoholic drinks in Europe [8]. In the present study, two other less consumed fermented beverages were analyzed, cider and mead, whose primary raw materials are apple juice and honey, respectively.

Ethanol content was one of the earliest parameters for which quantitative analytical methods were developed in the quality control of beverages [9]. The classical method for quantifying ethanol in alcoholic beverages is based on determining the density of the beverage after a distillation step [10]. This is a non-selective, laborious and time-consuming approach. Considering the need for quality control inspection in the food supply chain, several analytical methods have been developed for determining ethanol in different alcoholic beverages, including both fermented and distilled. Articles developing methods to quantify ethanol in more than one type of beverage have been based mainly on chromatographic techniques, such as gas chromatography coupled to mass spectrometry (GC-MS) [9] or flame ionization detection (GC-FID) [11], and high-performance liquid chromatography (HPLC) [12]. Besides being laborious and time-consuming, chromatographic

* Corresponding author.

E-mail address: marcsen@ufmg.br (M.M. Sena).

<https://doi.org/10.1016/j.talo.2023.100180>

Received 10 September 2022; Received in revised form 13 November 2022; Accepted 16 November 2022

Available online 4 January 2023

2666-8319/© 2023 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

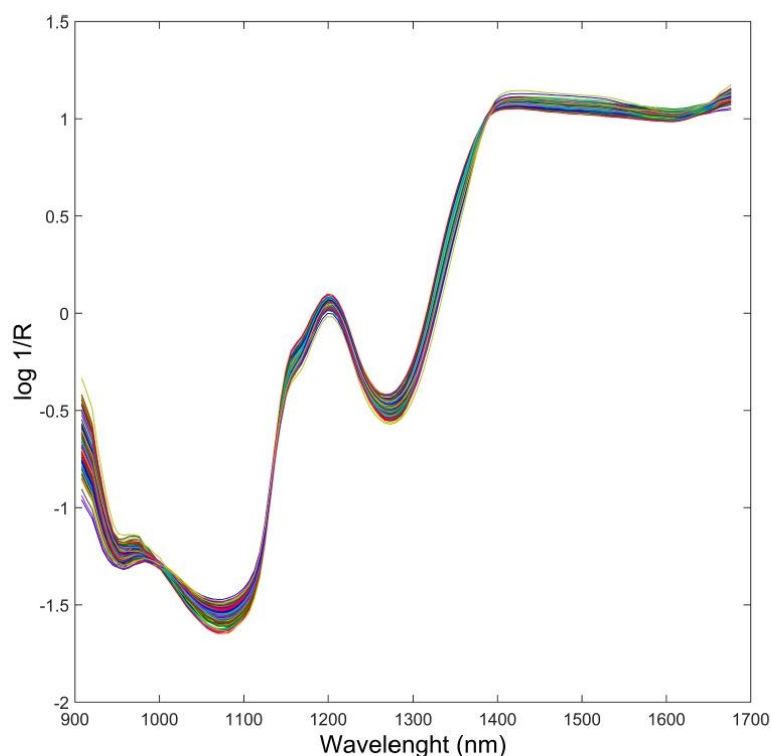


Fig. 1. NIR spectra preprocessed by Savitzky-Golay smoothing and standard normal variate (SNV) of all the 153 analyzed beverages samples.

Table 1
Results for the optimization of the developed PLS model through the detection of outliers.

		Number of calibration samples	Number of validation samples	Number of latent variables (LV)	RMSEC (%)	RMSEP (%)	RMSECV (%)	RPD calibration	RPD validation
Before	Beer	26	13	8	1.4	1.8	1.7	2.1	2.0
	Cider	27	13						
	Mead	24	12						
	Wine	25	13						
After	Beer	25	13	8	0.8	0.9	1.0	3.5	4.1
	Cider	25	11						
	Mead	21	10						
	Wine	21	10						

methods go against the principles of green chemistry, which search to avoid the use of toxic solvents and to minimize the generation of chemical waste. Thus, many alternative methods have been developed, mainly in the last years, seeking to overcome these limitations. These methods have been based on amperometric detection [13], microbial biosensing employing a eukaryote double-mediator system [14], direct analysis in real time mass spectrometry (DART-MS) [2], a colorimetric reaction of ethanol with dichromate in a 3D printing system [15], UV-visible spectra acquired in a flow system with in-line gas-diffusion [1], visible chemical waves registered with a smartphone during a Belousov-Zhabotinsky (BZ) reaction [16], a chemo-chronometric assay utilizing the BZ oscillating reaction [17] and RGB digital images obtained with a desktop scanner [18]. Most of these methods consume reagents [1,15-18], in some cases toxic reagents [15]. While beer and wine are among the most analyzed beverages in all previously mentioned articles, studies involving cider and mead are scarce in the

literature. Only one article has determined ethanol specifically in cider by employing ^1H NMR spectroscopy [19].

In the last years, vibrational spectroscopic techniques have increasingly represented alternatives for developing simpler, direct, faster, and greener analytical methods. Since these techniques provide overlapped analytical signals, their combination with chemometric tools is almost mandatory. Recently, the advantages of these techniques have been improved due to the miniaturization of the spectrophotometers. In fact, portability is an important trend in analytical chemistry in general, not only in spectroscopy, providing advantages such as time saving, low costs of operation, small size and light weight of the equipments, lower energy consumption and environmentally friendly analysis [20-24]. A minimum or no sample pretreatment is required, small sample volumes are necessary, there is no need to consume reagents or solvents, and no chemical waste is generated. There are evident advantages in eliminating the need of sending samples to the laboratory, as it can be moved

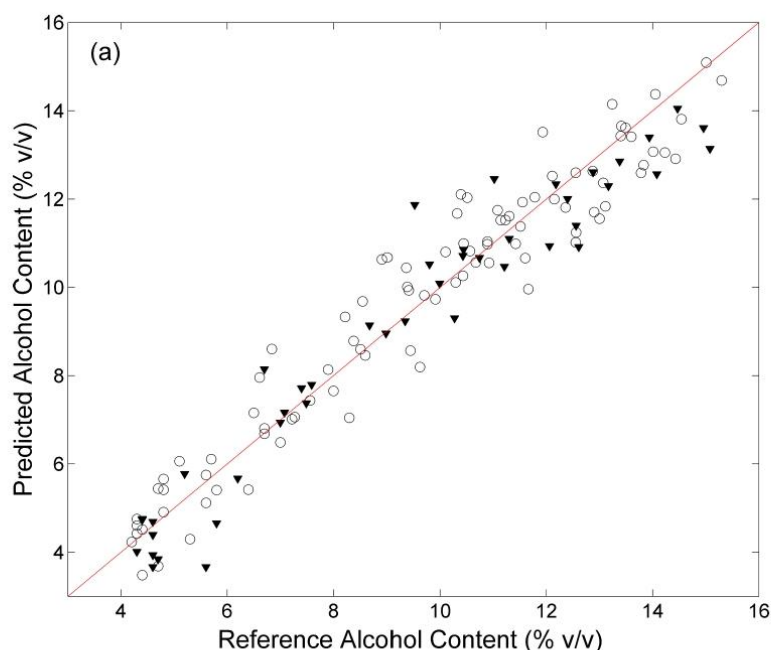


Fig. 2. Plot of reference alcohol content versus predicted values for the developed PLS model. (a) Representation as a function of calibration and validation samples. Empty circles indicate calibration samples and full down triangles indicate validation samples. (b) Representation as a function of the type of beverage. gray crosses represent beer samples, green asterisks cider samples, blue squares mead samples, and red down triangles wine samples.

to the field [20]. Particularly, portable devices can provide fast screening of beverage authenticity directly at the point of need [22]. Among vibrational techniques, handheld near infrared (NIR) devices present the advantages of being mostly cheaper, lighter and of smaller dimensions than Raman and mid infrared portable equipments [23,24]. Due to these characteristics, portable NIR spectroscopy is becoming one of the most commercially developed and employed analytical platforms for determining food authenticity, allowing integration into digital traceability systems [23]. In a previous and very recent article, our research group has employed portable NIR spectroscopy and multivariate calibration to determine the alcoholic strength of only one beverage matrix, beer [25]. In the present study, going beyond this previous reference, we developed a multi-product quantitative model by analyzing four different beverage matrices, which implies several advantages discussed in the next paragraph.

Multi-product models were obtained from spectral data for different types of samples or matrices, which are combined into a single multivariate calibration model [26,27]. They are also called global or robust calibration models (though these two terms have also been used for other situations) and can be very convenient and cost-effective. The maintenance of several one-product calibration models is laborious and time-consuming. Going from one-product to multi-product calibration may result in lower accuracy, though limited to an acceptable level, when a large number of samples for each matrix is available. However, when the products are relatively similar and especially if the number of samples of each product is small, making local modeling difficult, multi-product models tend to provide a better performance and lower prediction errors [26]. If the analytical range of the dependent variable is not too large, the traditional linear multivariate calibration model partial least squares (PLS) tends to provide good predictions. The use of portable devices also facilitates the development of this type of model

since the same spectrophotometer can be carried to different analysis sites. Multi-product multivariate calibration models have been developed in combination with NIR data for several types of matrices, such as different foods [26], several kinds of plant feed materials [27], various root and tuber powders [28], biomass from different product origins [29], industrialized fruit nectars and soy juices obtained from different fruits [30], wine grapes (*Vitis vinifera* L.) during on-vine ripening and at harvest from 25 white and red grape varieties [31], and five different vaccine products [32].

The objective of this paper was to develop and validate a multivariate calibration multi-product method based on the combination of portable NIR spectroscopy and PLS for directly determining ethanol content in four distinct fermented alcoholic beverages. Reference values were obtained with a method based on GC-FID. Concerning the robustness of the method, a relative great variety of beers, ciders, meads, and wines were incorporated into the model, corresponding to an analytical range from 4.3 to 15.3% (v/v). The proposed method was also submitted to a multivariate analytical validation according to the Brazilian and international guidelines [33,34].

2. Materials and methods

2.1. Samples

Sample beverages utilized for the construction of the multi-product model were beer (thirty-nine samples), cider (forty samples), mead (thirty-six samples), and wine (thirty-eight samples). A total of 153 samples of alcoholic beverages were purchased in different local marketplaces (100% of both beer and wine samples, 4 samples of cider and 4 samples of mead) or artisanally prepared (the remaining samples of both cider and mead) for this study. Since only a small variety of cider and

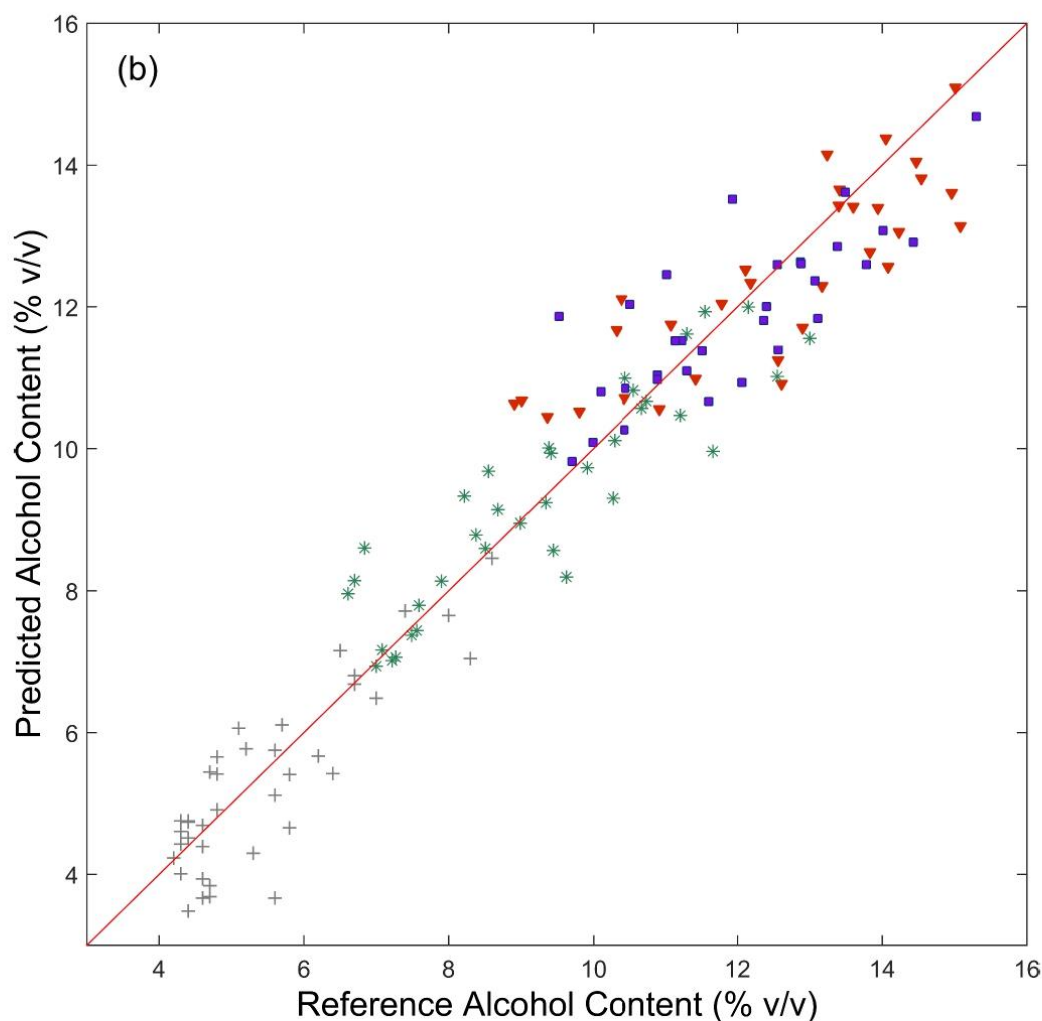


Fig. 2. (continued).

mead brands are available in the local market, about 90% of these beverage samples were artisanally produced in our laboratory with a wide variety of raw materials and alcohol content, aiming at obtaining a robust multi-product model. The whole set of samples was obtained or prepared to provide a wide variance with their ethanol content ranging from 4 to 16% (v/v).

2.2. Instruments and software

NIR spectra were acquired using a portable MicroNIR® 1700 spectrophotometer from Viavi Solutions (Milpitas, CA, USA), with an InGaAs photodiode array detector and linear variable filter (LVF) technology as dispersing element. This is one of the smallest NIR devices, weighing only 64 g, with one of the best performances in the market [23]. Each sample was scanned in the diffuse reflectance mode (and converted to $\log(1/R)$, which is equivalent to absorbance), with twenty scans at the wavelength range of 908–1676 nm, with resolution of 6.25 nm. Data

were processed using MATLAB® software, version 7.13 (MathWorks, Natick, USA), coupled with the PLS_Toolbox, version 6.5 (Eigenvector Technologies, Manson, USA). A Shimadzu GC-17A gas chromatography system (Kyoto, Japan) equipped with a flame ionization detector (FID) and a PoraPlot Q column (10 m × 0.32 mm × 10 μm) was used for obtaining reference values.

2.3. Procedure

Approximately 5 mL of each beer and cider sample were degassed using an ultrasonic bath (Equilab ULTRASONIK 28H, Madrid, Spain) for 5 min. Mead and wine samples did not need to be degassed. After degassing, approximately 1.0 mL of each beverage sample was used for the GC-FID reference method and the remaining 4.0 mL for the proposed multi-product NIR method. In the sequence, 3.0 mL of each sample was transferred to a polystyrene Petri dish (1.7 cm radius × 0.9 cm height). The plate was placed above the MicroNIR® spectrophotometer for

Table 2
Other estimated quantitative FOM for the developed PLS model.

Figures of merit	Parameter		Value		
Precision	RSD repeatability (%)	Beer	0.3		
			0.2		
			0.5		
		Cider	1.7		
			0.3		
			0.4		
		Mead	0.4		
			0.1		
			0.2		
		Wine	0.4		
			0.3		
			0.3		
		RSD intermediate precision (%)		Beer	1.6
					1.2
					1.8
				Cider	1.9
1.1					
1.8					
Mead	0.8				
	0.5				
	1.0				
Wine	1.1				
	1.0				
	0.6				
Linearity	Durbin-Watson test parameter				1.78
				Slope	0.97
					± 0.03
				Intercept	0.85
		± 0.22			
Correlation coefficient	0.9312				
	0.9				
Inverse of analytical sensitivity (%)			0.9		
Bias (%)			0.3 ± 0.9		
Working Range (%)			4.3–15.3		

spectra recording. During all these measurements, the temperature of the laboratory was set at 25 ± 1 °C. Precautions to maintain the scattering coefficient constant during the experiments were taken, such as the control of the sample volume and the use of a reflective artifact.

2.4. GC-FID analysis (reference method)

A GC-FID method was developed for obtaining reference values for the samples, which were used as dependent variables for the building of the PLS model. For this chromatographic method, 100 μL of each sample were mixed with 900 μL of the internal standard (IS) solution (1-butanol 1.0% v/v) in a 2 mL microcentrifuge tube. The carrier gas was hydrogen with a flow of $2.2 \text{ cm}^3 \text{ min}^{-1}$. The injection port and the detector were both held at 250 °C, and the oven temperature was maintained at 200 °C for 4 min. An aliquot of 0.2 μL of the diluted sample was injected with split ratio of 1:2.

2.5. Multivariate calibration model and analytical validation

A multi-product multivariate calibration model was built using PLS. Samples were split into two-thirds and one-third for the calibration and validation sets, respectively, by applying the Kennard-Stone algorithm [35]. This algorithm was applied separately to the samples of each matrix/product, guaranteeing the representativeness of each type of beverage in the calibration and validation sets. Kennard-Stone algorithm was chosen instead of random sampling because it ensures the selection of representative and homogeneously distributed samples in the whole analytical range of the multivariate space, in a reproducible manner based on a systematic criterion [36].

The performed analytical validation searched to adapt the requirements of the Brazilian and international guidelines [37,38] to the state of the art of estimate FOM in multivariate calibration [33,34,39].

Appropriate figures of merit (FOM) were estimated for the proposed multi-product method in accordance with regulatory requirements, such as trueness, precision, linearity, working range, analytical sensitivity, bias, and residual prediction deviation (RPD). Samples with extreme leverages, large residuals in the X block (spectral data) or large residuals in the y block (studentized residuals of the predicted concentration values) at 95% confidence level were detected as outliers [33,34]. The uncertainty of the predicted values was determined to estimate confidence intervals by calculating the sample-specific standard prediction errors (SPE). SPE were calculated based on the error-in-variables equation (EIV) [40]. This equation considers the uncertainty of the reference method and should only be used in the absence of bias. Thus, the absence of bias in the model should be previously verified by a t-test with the validation samples (the worst case in relation to the calibration samples) at 95% confidence level [39]. In addition, an elliptical joint confidence region (EJCR) was calculated to verify the linearity and the absence of systematic errors in the model [41,42].

3. Results and discussion

Specific guidelines recommend the utilization of a minimum of 6 (nLV+1) and 4(nLV) samples in the calibration and validation sets, respectively, for building infrared multivariate quantitative models, where nLV means the number of latent variables [39]. Taking into account that the number of samples available for building one-product models was too small, it was decided to build only one multi-product multivariate calibration model.

As already mentioned in Section 2.5, the Kennard-Stone algorithm was employed to select samples for calibration and validation datasets, which were composed of 102 and 51 samples, respectively. The proportionality among each type of beverage was maintained in these two datasets. NIR data were preprocessed to improve signal-to-noise ratio, and to avoid drifts caused by multiplicative light scattering, which are typical of diffuse reflectance spectra [43]. Spectra were sequentially preprocessed by Savitzky-Golay smoothing (5 points filter width and second order polynomial fit), standard normal variate scaling (SNV), and mean centering. Fig. 1 shows all preprocessed spectra used in the model development (before mean centering). By observing these spectra, the most prominent signals are the absorption bands centered at about 1200 nm and 1400 nm. These two spectral bands can be assigned to the second overtone of C–H stretching of ethanol, and to the first overtone of O–H stretching vibrations of ethanol and water, respectively [44].

Random subsets (9 splits and 20 iterations) cross-validation was applied to select the best number of latent variables (LV). The LV number presenting the smallest root mean square error of cross-validation (RMSECV) was chosen. The best model was selected with 8 LV after optimization by outlier detection. Outlier detection was performed based on the parameters mentioned in Section 2.5, all at 95% confidence level. This procedure can improve model quality since outliers are defined as observations showing some type of departure from most of the data. The number of outliers must not exceed the limit of 22.2% of the training and test sets for each beverage [38,45]. From the original data, 17 out of 153 samples were removed as outliers (11.1%), ten from the calibration set (9.8%), and seven from the validation set (13.7%). Table 1 specifies how many samples of each beverage were removed from the model. After deleting outliers, the model was reconstructed and accounted for 99.82% of the variance in X block and 93.09% in Y block. The results for the optimization of the model, including FOM such as root mean square errors of calibration (RMSEC) and prediction (RMSEP), and RPD for calibration and validation sets, are showed in Table 1.

RMSEC and RMSEP are FOM related to the trueness of the multivariate calibration method. An improvement of these parameters was obtained after the outliers' detection, resulting in an RMSEC of 0.8% and an RMSEP of 0.9% v/v. An appropriate FOM to assess the predictive

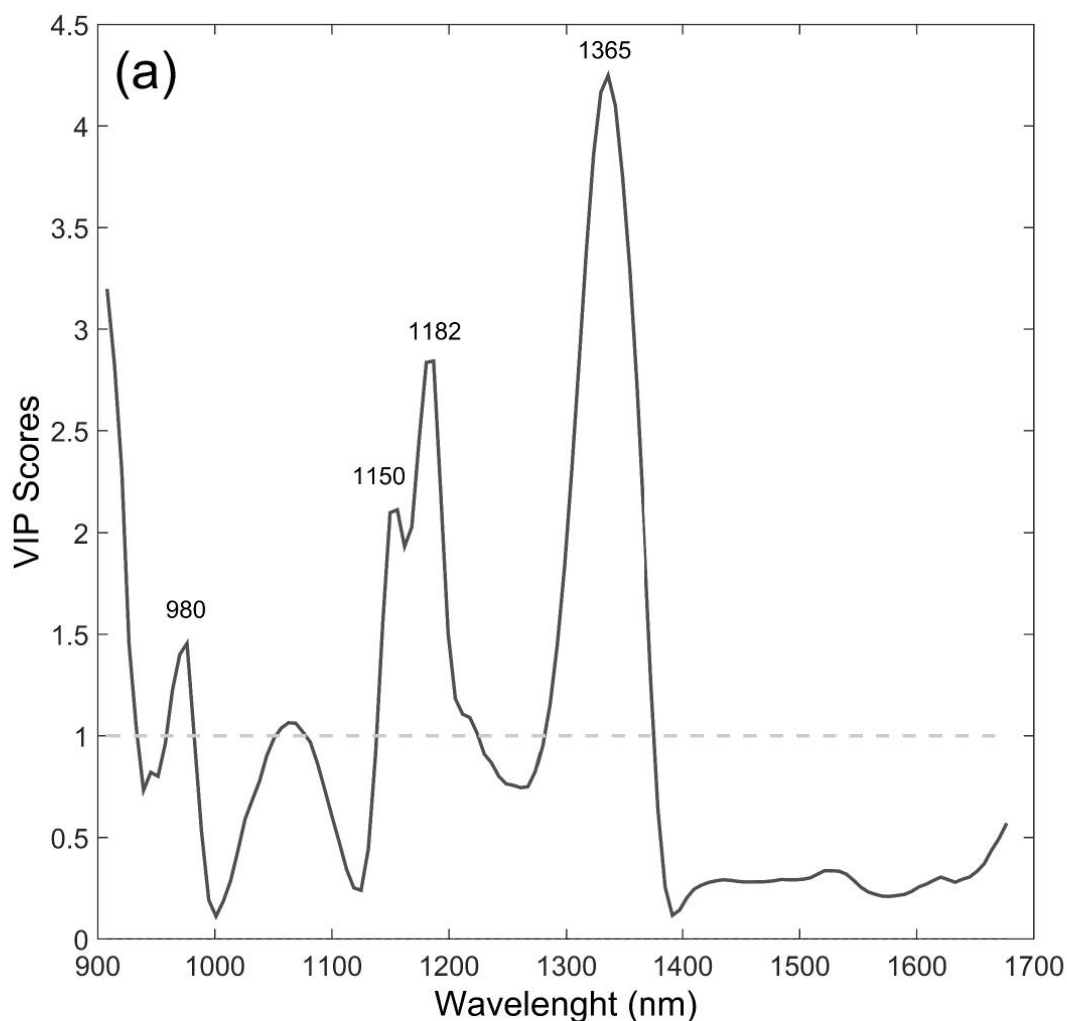


Fig. 3. Informative vectors for the PLS model. (a) Variable importance in projection (VIP) scores. Dashed horizontal line indicates the threshold of 1.0, above which variables are considered to contribute significantly to the PLS model. (b) Regression vector.

ability of NIR models in absolute terms is RPD, which is also a parameter for evaluating trueness. The concept of RPD was first introduced in 1993 and corresponds to the ratio of natural variation in samples to the size of probable prediction errors [46]. The developed PLS model presented RPD of 3.5 and 4.1 for the calibration and validation sets, respectively. These results are satisfactory, as PLS models with RPD values greater than 2.4 are considered to have a good predictive capacity [33,47]. The adjust can be evaluated through the regression plots of reference *versus* predicted values, which are shown in Fig. 2. The linearity will be evaluated by proper statistical tests presented in Table 2 and discussed later in this section, aiming to prove that residuals are randomly distributed. This figure represents the same fit as function of calibration and validation samples (Fig. 2a) and as a function of the type of beverage (Fig. 2b). It is possible to observe a difference in alcoholic strength amid the studied matrices. Wine and mead have the highest alcohol content,

ranging between 8% and 16%. Beer has the lowest content, ranging between 4% and 9%. Finally, cider has an intermediate alcohol content, between 7% to 14%.

The most important informative vectors extracted from the developed PLS model were used for spectral interpretation. Regression coefficients and variable importance in projection (VIP) scores are shown in Fig. 3. By jointly evaluating these plots, it is possible to identify the variables or bands most contributing to model predictions. The VIP scores (Fig. 3a) represent a measure of the degree of importance of each variable in the projection used by a particular model in absolute values. Complementarily, regression vectors (Fig. 3b) discriminate the direction of each variable contribution. The most positive coefficients contribute to the prediction of the target analyte, while the most negative ones can be associated with the interferences. Variables with VIP scores above a threshold of 1.0 are considered significant for the model predictions. It is

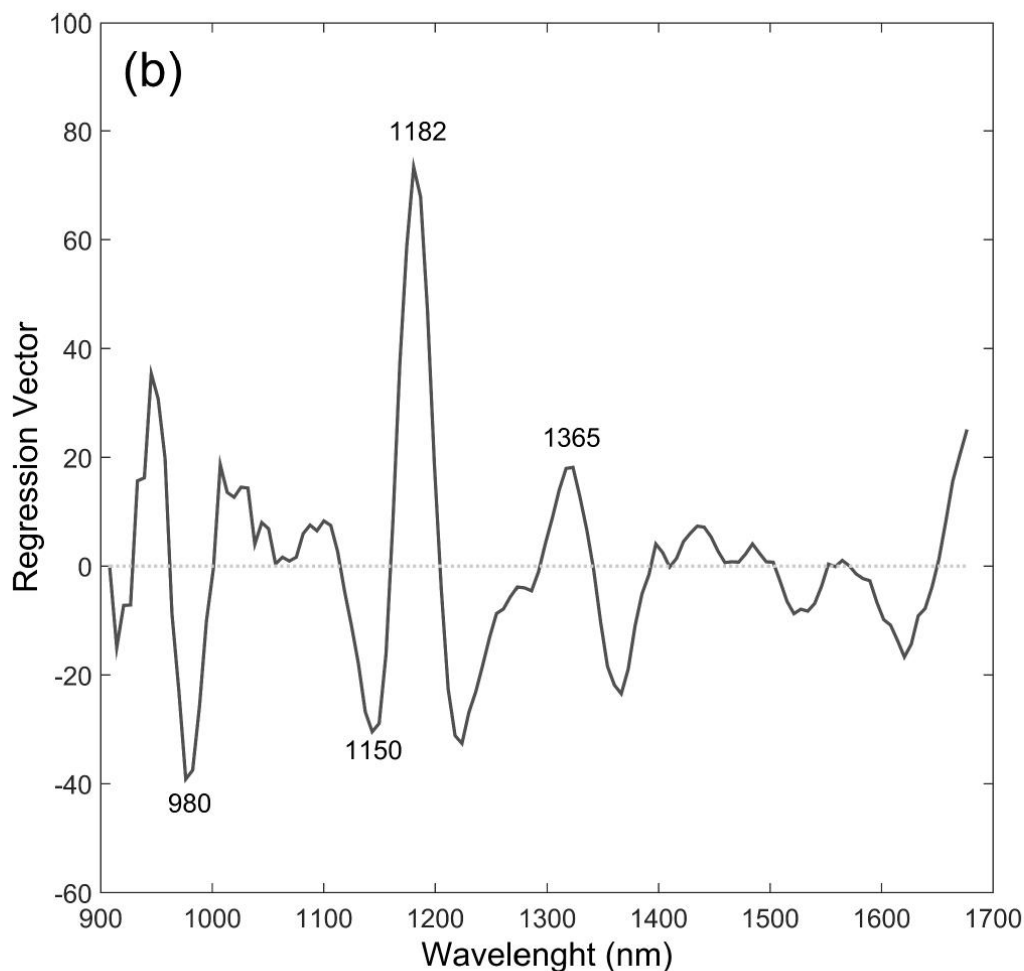


Fig. 3. (continued).

also important to note that NIR spectroscopy involves interaction between matter and light, and although the spectral bands may be assigned to vibrational modes of specific functional groups, they are not fully selective.

According to Fig. 3, there are four discriminant spectral bands to be highlighted, two positive and two negative regions of regression coefficients, which correspond to the four highest VIP scores peaks. The band centered around 980 nm can be assigned to the second overtone of the combination band of asymmetric and symmetric stretchings of the O—H bond of water, which is an interference in ethanol quantification. Spectral bands centered around 1150 nm and 1182 nm can be assigned to the second overtones of CH₃ and CH₂ stretchings of carbohydrates, which are common components of all analyzed beverages. Finally, the highest VIP score, at 1365 nm, is associated to the most predictive spectral band related to the target analyte. This band can be assigned to the first overtone of C—H stretching and the first overtone of the O—H stretching of ethanol [48]. It is interesting to note that the most intense NIR absorption band in the samples, above 1400 nm (Fig. 1) did not contribute significantly to the model, since the respective VIP scores

presented small values below the threshold of 1.0.

Table 2 presents complementary POM assessed for the optimized PLS model. For evaluating linearity (Fig. 2), a correlation coefficient of 0.9312 was estimated, jointly with a slope of 0.97 ± 0.03 and an intercept of 0.85 ± 0.22 . These results reinforced the accuracy results represented by RMSEC and RMSEP. In addition to these parameters, the linearity of the model should be verified by applying proper statistical tests to check the absence of systematic residuals. Thus, three tests were applied in the sequence: Ryan–Joiner for checking normality, Brown–Forsythe for checking homoscedasticity, and Durbin–Watson for confirming the absence of autocorrelation of the residuals, all at 95% confidence level [49]. The results of these tests assured the random behavior of the residuals, with an estimated value of 1.78 for Durbin–Watson test, within the acceptance range (1.50 - 2.50). The estimated small value of bias for the validation set, $0.3 \pm 0.9\%$, assured the absence of systematic errors in the model. A Student's *t*-test with 44 degrees of freedom corroborated the absence of a bias different from zero. Precision was evaluated as repeatability and intermediate precision at three different concentration levels for each of the four beverages, through estimating relative

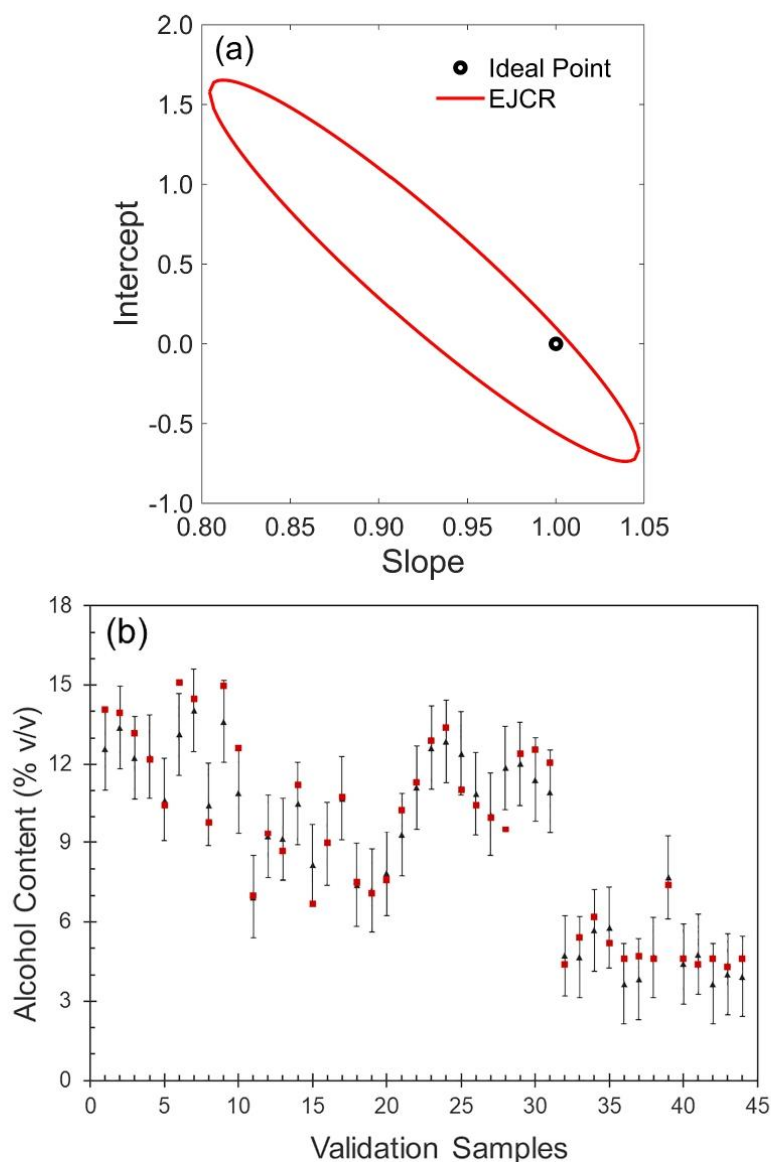


Fig. 4. (a) An elliptical joint confidence region (EJCR) estimated at 95% confidence level for the linearity (slope *versus* intercept) of the developed PLS model. (b) Confidence intervals estimated for the prediction of the validation samples. Sample-specific standard prediction errors (SPE) were calculated based on errors-in-variables (EIV) equation. gray up triangles mean predicted values and red squares mean the respective reference values.

standard deviation (RSD). RSD values varied between 0.1% and 1.7% at the level of repeatability and between 0.5% and 1.9% at the level of intermediate precision. These results were considered good, in accordance with the international guidelines that prescribe a maximum RSD of 3.7% for repeatability and 6.0% for intermediate precision, taking into account the level of the concentration of the analyte (above 1 g kg^{-1}) [10]. The analytical sensitivity was calculated from the estimate of the instrumental noise, which was obtained through the pooled standard

deviation of ten spectra replicates of an empty Petri dish [33]. Its inverse value, 0.9%, provides an estimate of the minimum concentration difference that the method can distinguish and established the use of only one decimal place to represent the results.

An elliptical joint confidence region (EJCR) was also estimated to confirm the linearity of the developed method (Fig. 4a). The limits of the estimated EJCR at 95% confidence level include the ideal point corresponding to an angular coefficient (slope) of 1 and a linear coefficient

(intercept) of 0 for the plot of predicted versus reference values. Moreover, EPCR can be considered an accuracy confirmation, showing no difference between the proposed method and the reference one. Therefore, the results of linearity, precision and trueness allowed to assure that the developed method can be considered accurate in the working range from 4.3% to 15.3%. Lastly, confidence intervals were estimated for the prediction of the validation samples. EIV equation was employed to estimated SPE varying between 1.5–1.6% (Fig. 4b). The reliability of the use of this equation was assured by the estimate of a non-significant bias, as previously mentioned in the last paragraph. Only two out of 44 validation samples were not predicted within the respective estimated confidence intervals (95%), representing an agreement of 95.5%. Therefore, the method was able to accurately quantify the alcohol content in beer, cider, mead, and wine with only one multivariate calibration model. The results shown in Fig. 4b also assured the trueness of the method jointly with other previously discussed FOM, such as RMSEC, RMSEP and RPD.

Conclusions

In this article, portability, vibrational spectroscopy and chemometrics were combined for the development of a direct, simple, rapid and environmentally friendly analytical method. Though the reference chromatographic method used solvents, once the model was established, the method will no longer consume reagents or solvents and can be considered green. A robust multi-product multivariate calibration model was built with NIR spectra for the quantification of alcohol content in four different fermented alcoholic beverages, demonstrating its applicability to real samples. The method was submitted to a full multivariate analytical validation through the estimate of proper figures of merit, thus ensuring its fit for purpose. The consistency of the model was also corroborated by a coherent spectral interpretation through analyzing its informative vectors. The developed method presented several advantages over traditional more laborious and time-consuming analytical alternatives based on densitometry or chromatography. These advantages allow to suggest the incorporation of the new method in portable analytical platforms for quality control of beverages, thus contributing to provide better transparency in the food supply chain. The synergy fostered by the combination of portability, NIR spectroscopy and a single multivariate calibration multi-product model can provide a green analytical tool for regulatory agencies performing quality inspection control of alcoholic beverages. This method may be expanded to incorporate other matrices, particularly distilled beverages, thus enlarging its analytical range.

Funding

This work was supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG), grant numbers APQ03457-16 and RED00042-16; and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), grant number 1383/2020.

CRedit authorship contribution statement

Ana Carolina da Costa Fulgencio: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Glaucimar Alex Passos Resende:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation. **Marden Claret Fontoura Teixeira:** Investigation, Data curation. **Bruno Gonçalves Botelho:** Conceptualization, Data curation, Investigation, Methodology, Resources, Supervision, Project administration, Funding acquisition. **Marcelo Martins Sena:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors would like to thank the Brazilian government agencies CNPq, FAPEMIG, and CAPES for financial support. ACGF acknowledges CAPES and GAPR acknowledges CNPq for Ph.D. scholarships. MCFT acknowledges CNPq for a scientific initiation scholarship. FAPEMIG (Fundação de Amparo à Pesquisa do Estado de Minas Gerais) is especially acknowledged for the acquisition of the portable NIR spectrophotometer (Viavi MicroNIR® 1700) through the project APQ03457-16.

References

- [1] N. Ratanawimarmwong, M. Sinpun, P. Chankaw, N. Choengchana, D. Nacapricha, Simple flow system with in-line gas-diffusion unit for determination of ethanol employing hypsochromic shift of visible absorbance band of methyl orange, *Talanta* 206 (2020), 120234, <https://doi.org/10.1016/j.talanta.2019.120234>.
- [2] E. Sisco, E.L. Robinson, Determination of ethanol concentration in alcoholic beverages by direct analysis in real time mass spectrometry (DART-MS), *Forensic Chem.* 18 (2020), 100219, <https://doi.org/10.1016/j.forc.2020.100219>.
- [3] European Commission, Taxation and customs union - excise duty on alcohol. Brussels, 2020. https://ec.europa.eu/taxation_customs/taxation-1/excise-duties/excise-duty-alcohol_en (accessed in August 2022).
- [4] Department of Taxation and Finance of New York State, Alcoholic beverages tax. Albany, USA, 2022. <https://www.tax.ny.gov/bus/bev/abt.htm#beer> (accessed in August 2022).
- [5] Official Journal of the European Union: Regulation (EU) No 1168/2011 of the European Parliament and of the Council of 25 October 2011, Luxembourg, 2011.
- [6] European Commission, Labelling of alcoholic beverages in the EU: some facts. Brussels, 2022. https://ec.europa.eu/food/safety/labelling-and-nutrition/food-information-consumers-legislation/alcohol-labelling_en (accessed in August 2022).
- [7] European Commission, Food information to consumers - legislation. Brussels, 2022. https://ec.europa.eu/food/safety/labelling-and-nutrition/food-information-consumers-legislation_en (accessed in August 2022).
- [8] WHO – World Health Organization, Global status report on alcohol and health 2018. Geneva, Switzerland, 2018. <https://www.who.int/publications/i/item/9789241565639>, (accessed in August 2022).
- [9] M. Stupak, V. Kocourek, I. Kolouchova, J. Hajslova, Rapid approach for the determination of alcoholic strength and overall quality check of various spirit drinks and wines using GC-MS, *Food Control* 80 (2017) 307–313, <https://doi.org/10.1016/j.foodcont.2017.05.008>.
- [10] G.W. Latimer, *Official Methods of Analysis of AOAC International*, 19th ed., AOAC International, Gaithersburg, USA, 2012.
- [11] O. Wachelko, P. Szpot, M. Zawadzki, The application of headspace gas chromatographic method for the determination of ethyl alcohol in craft beers, wines and soft drinks, *Food Chem.* 346 (2021), 128924, <https://doi.org/10.1016/j.foodchem.2020.128924>.
- [12] T. Yaritha, R. Nakajima, S. Otsuka, T. Ihara, A. Takatsu, M. Shibukawa, Determination of ethanol in alcoholic beverages by high-performance liquid chromatography-flame ionization detection using pure water as mobile phase, *J. Chromatogr. A* 976 (2002) 387–391, [https://doi.org/10.1016/S0021-9673\(02\)00942-1](https://doi.org/10.1016/S0021-9673(02)00942-1).
- [13] T.R.L.C. Paixão, D. Corbo, M. Bertotti, Amperometric determination of ethanol in beverages at copper electrodes in alkaline medium, *Anal. Chim. Acta* 472 (2002) 123–131, [https://doi.org/10.1016/S0003-2670\(02\)00942-X](https://doi.org/10.1016/S0003-2670(02)00942-X).
- [14] H. Nakamura, R. Tanaka, K. Suzuki, M. Yatska, Y. Mogi, A direct determination method for ethanol concentrations in alcoholic beverages employing a eukaryote double-mediator system, *Food Chem* 117 (2009) 509–513, <https://doi.org/10.1016/j.foodchem.2009.04.026>.
- [15] A.C.N. Pinheiro, V.S. Ferreira, B.G. Lucca, Stamping method based on 3D printing and disposable napkin: cheap production of paper analytical devices for alcohol determination in beverages aiming forensics and food control, *Microchem. J.* 180 (2022), 107604, <https://doi.org/10.1016/j.microc.2022.107604>.
- [16] T. Sombona, S. Sansuk, An instrument-free method based on visible chemical waves for quantifying the ethanol content in alcoholic beverages, *Food Chem.* 253 (2018) 300–304, <https://doi.org/10.1016/j.foodchem.2018.01.168>.
- [17] S. Sansuk, P. Juntarakod, W. Tongphothorn, A. Sirimungkala, T. Sombon, Visual chemo-chronometric assay for quantifying ethanol in alcoholic drinks by the

- colorimetric Belousov-Zhabotinsky oscillator, *Food Control* 110 (2020), 107042, <https://doi.org/10.1016/j.foodcont.2019.107042>.
- [18] M.F. Filgueiras, B.O. Lima, E.M. Borges, A high-throughput, cheap, and green method for determination of ethanol in cachaça and vodka using 96-well-plate images, *Talanta* 241 (2022), 123229, <https://doi.org/10.1016/j.talanta.2022.123229>.
- [19] A. Zuriarrain, J. Zuriarrain, M. Villar, I. Berregi, Quantitative determination of ethanol in cider by 1H NMR spectrometry, *Food Control* 50 (2015) 758–762, <https://doi.org/10.1016/j.foodcont.2014.10.024>.
- [20] A. Galuszka, Z.M. Migaszewski, J. Namiesnik, Moving your laboratories to the field—Advantages and limitations of the use of field portable instruments in environmental sample analysis, *Environ. Res.* 140 (2015) 593–603, <https://doi.org/10.1016/j.envres.2015.05.017>.
- [21] R.A. Crocombe, Portable spectroscopy, *Appl. Spectrosc.* 72 (2018) 1701–1751, <https://doi.org/10.1177/0003702818809719>.
- [22] K.C.A. Rezende, L.M. Duarte, K.M.P. Pinheiro, T.M.G. Cardoso, S.A. Nogueira, W.K. T. Coltro, Portable analytical platforms associated with chemometrics for rapid screening of whisky adulteration, *Food Anal. Methods* 15 (2022) 2451–2461, <https://doi.org/10.1007/s12161-022-02303-5>.
- [23] C. McVey, C.T. Elliott, A. Cannavan, S.D. Kelly, A. Petchkongkaew, S.A. Haughey, Portable spectroscopy for high throughput food authenticity screening: advancements in technology and integration into digital traceability systems, *Trends Food Sci. Technol.* 118 (2021) 777–790, <https://doi.org/10.1016/j.tifs.2021.11.003>.
- [24] C. Zhu, X. Fu, J. Zhang, K. Qin, C. Wu, Review of portable near infrared spectrometers: current status and new techniques, *J. Near Infrared Spectrosc.* 30 (2022) 51–66, <https://doi.org/10.1177/09670335211030617>.
- [25] A.C.C. Fulgencio, G.A.P. Resende, M.C.F. Teixeira, B.G. Botelho, M.M. Sena, Determination of alcohol content in beers of different styles based on portable near-infrared spectroscopy and multivariate calibration, *Food Anal. Methods* 15 (2022) 307–316, <https://doi.org/10.1007/s12161-021-02126-w>.
- [26] E. Micklander, K. Kjeldahl, M. Egebo, L. Norgaard, Multi-product calibration models of near infrared spectra of foods, *J. Near Infrared Spectrosc.* 14 (2006) 395–402, <https://doi.org/10.1255/jnirs.659>.
- [27] X. Fan, S. Tang, G. Li, X. Zhou, Non-invasive detection of protein content in several types of plant feed materials using a hybrid near infrared spectroscopy model, *PLoS ONE* 11 (2016), e0163145, <https://doi.org/10.1371/journal.pone.0163145>.
- [28] R.E. Masithoh, S. Lohumi, W.-S. Yoon, H.Z. Amanah, B.-K. Cho, Development of multi-product calibration models of various root and tuber powders by fourier transform near infra-red (FT-NIR) spectroscopy for the quantification of polysaccharide contents, *Heliyon* 6 (2020) e05099, <https://doi.org/10.1016/j.heliyon.2020.e05099>.
- [29] M.K.D. Rambo, M.M.C. Ferreira, E.P. Amorim, Multi-product calibration models using NIR spectroscopy, *Chemom. Intell. Lab. Syst.* 151 (2016) 108–114, <https://doi.org/10.1016/j.chemolab.2015.12.013>.
- [30] D.A. Santos, K.P. Lima, V. Cavalcante, A. Coqueiro, M.F.B. Consolin, N. Consolin Filho, P.H. Março, P. Valdeirama, Multiproduct, multicomponent and multivariate calibration: a case study by using vis-NIR spectroscopy, *Food Anal. Methods* 11 (2018) 1915–1919, <https://doi.org/10.1007/s12161-017-1099-4>.
- [31] V. González-Caballero, D. Pérez-Marín, M.-I. López, M.-T. Sánchez, Optimization of NIR spectral data management for quality control of grape bunches during on-vine ripening, *Sensors* 11 (2011) 6109–6124, <https://doi.org/10.3390/s111060109>.
- [32] Y. Zheng, X. Lai, S.W. Bruun, H. Ipsen, J.N. Larsen, H. Lowenstein, I. Søndergaard, S. Jacobsen, Determination of moisture content of lyophilized allergen vaccines by NIR spectroscopy, *J. Pharm. Biomed. Anal.* 46 (2008) 592–596, <https://doi.org/10.1016/j.jpba.2007.11.011>.
- [33] B.G. Botelho, B.A.P. Mendes, M.M. Sena, Development and analytical validation of robust near-infrared multivariate calibration models for the quality inspection control of mozzarella cheese, *Food Anal. Methods* 6 (2013) 881–891, <https://doi.org/10.1007/s12161-012-9498-z>.
- [34] B.G. Botelho, L.P. Assis, M.M. Sena, Development and analytical validation of a simple multivariate calibration method using digital scanner images for sunset yellow determination in soft beverages, *Food Chem.* 159 (2014) 175–180, <https://doi.org/10.1016/j.foodchem.2014.03.048>.
- [35] R.W. Kennard, L.A. Stone, Computer aided design of experiments, *Technometrics* 11 (1969) 137–148, <https://doi.org/10.2307/1266770>.
- [36] R.D.A. Ferreira, G. Teixeira, L.A. Peterlini, Kennard-Stone method outperforms the Random Sampling in the selection of calibration samples in SNPs and NIR data, *Cienc. Rural* 52 (2021), e20201072, <https://doi.org/10.1590/0103-8478r20201072>.
- [37] Brasil, *Dispõe Sobre a Padronização, a Classificação, o Registro, a Inspeção, a Produção e a Fiscalização de Bebidas; Regulamenta a Lei no 8918, De 14 De Julho De 1994 (Decreto no 6.871, De 04 De Junho De 2009)*, Diário Oficial da República Federativa do Brasil, Brasília, 2009.
- [38] M. Thompson, S.L.R. Ellison, R. Wood, Harmonized guidelines for single-laboratory validation of methods of analysis (IUPAC Technical Report), *Pure Appl. Chem.* 74 (2002) 835–855, <https://doi.org/10.1351/pac200274050835>.
- [39] ASTM, *Standard Practices for Infrared Multivariate Quantitative Analysis - E1655-05*, ASTM International, West Conshohocken, USA, 2012.
- [40] N.M. Faber, X.H. Song, P.K. Hopke, Sample-specific standard error of prediction for partial least squares regression, *TrAC - Trends Anal. Chem.* 22 (2003) 330–334, [https://doi.org/10.1016/S0165-9936\(03\)00503-X](https://doi.org/10.1016/S0165-9936(03)00503-X).
- [41] A.S. Luna, F.B. Gonzaga, W.F. Rocha, I.C. Lima, A comparison of different strategies in multivariate regression models for the direct determination of Mn, Cr, and Ni in steel samples using laser-induced breakdown spectroscopy, *Spectrochim. Acta B* 139 (2018) 20–26, <https://doi.org/10.1016/j.sab.2017.10.016>.
- [42] M. Ito, T. Suzuki, S. Yada, H. Nakagami, H. Teramoto, E. Yonemochi, K. Terada, Development of a method for nondestructive NIR transmittance spectroscopic analysis of acetaminophen and caffeine anhydride in intact bilayer tablets, *J. Pharm. Biomed. Anal.* 53 (2010) 396–402, <https://doi.org/10.1016/j.jpba.2010.04.029>.
- [43] A. Rinnan, F. van den Berg, S.B. Engelsen, Review of the most common pre-processing techniques for near-infrared spectra, *TrAC - Trends in Anal. Chem.* 28 (2009) 1201–1222, <https://doi.org/10.1016/j.trac.2009.07.007>.
- [44] A.C.C. Fulgencio, G.A.P. Resende, M.C.F. Teixeira, B.G. Botelho, M.M. Sena, Screening method for the rapid detection of diethylene glycol in beer based on chemometrics and portable near-infrared spectroscopy, *Food Chem* 391 (2022), 133258, <https://doi.org/10.1016/j.foodchem.2022.133258>.
- [45] MAPA, *Ministério da Agricultura, Pecuária e Abastecimento, Manual de Garantia da Qualidade Analítica, Áreas de Identidade e Qualidade de Alimentos e de Insumos*, Brasília, 2014.
- [46] P.C. Williams, D.C. Sobering, Comparison of commercial near infrared transmittance and reflectance instruments for analysis of whole grains and seeds, *J. Near Infrared Spectrosc.* 1 (1993) 25–32, <https://doi.org/10.1255/jnirs.3>.
- [47] P.C. Williams, Implementation of near-infrared technology, in: P.C. Williams, K. Norris (Eds.), *Near-infrared Technology in the Agricultural and Food Industries*, American Association of Cereal Chemists, St. Paul, 2001, pp. 145–169.
- [48] B.H. Stuart, *Infrared Spectroscopy: Fundamentals and Applications*, John Wiley & Sons, New York, 2004.
- [49] S.V.C. Souza, R.G. Junqueira, A procedure to assess linearity by ordinary least squares method, *Anal. Chim. Acta* 552 (2005) 25–35, <https://doi.org/10.1016/j.aca.2005.07.043>.



Marcelo Martins Sena is an Associate Professor at the Federal University of Minas Gerais, Belo Horizonte, Brazil, and Research Fellow from the Brazilian National Council for Scientific and Technological Development (CNPq). He holds a Ph. D. in Chemistry, and his current work focuses on chemometrics in analytical chemistry, including the application of molecular spectroscopic techniques (NIR, MIR, UV-vis, spectrofluorimetry, MS) to the analysis of food, pharmaceuticals, agricultural products and forensic samples. He has also published around 75 articles in international peer-reviewed journals.