

JAMILLE CARVALHO DE SOUZA

**EMPREGO DA ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO PARA
CARACTERIZAÇÃO *IN SITU* DA COMPOSIÇÃO DE ALIMENTOS E DE
SUBSTÂNCIAS ILÍCITAS**

Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Agroquímica para obtenção do título de *Doctor Scientiae*.

Orientadora: Prof.^a Dr.^a Maria do Carmo
Hespanhol

**VIÇOSA - MINAS GERAIS
2024**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade
Federal de Viçosa - Campus Viçosa**

T

S729e
2024

Souza, Jamille Carvalho de, 1996-
Emprego da espectroscopia no infravermelho próximo para a caracterização *in situ* da composição de alimentos e de substâncias ilícitas / Jamille Carvalho de Souza. – Viçosa, MG, 2024.

1 tese eletrônica (117 f.): il. (algumas color.).

Inclui apêndices.

Orientador: Maria do Carmo Hespanhol.

Tese (doutorado) - Universidade Federal de Viçosa, Departamento de Química, 2024.

Referências bibliográficas: f. 105-110.

DOI: <https://doi.org/10.47328/ufvbbt.2024.608>

Modo de acesso: World Wide Web.

1. Análise espectral - Instrumentos. 2. Espectroscopia de infravermelho próximo. 3. Análise multivariada. 4. Café - Qualidade. 5. Cocaína. 6. Adoçantes artificiais. 7. Açúcares (Química orgânica). I. Hespanhol, Maria do Carmo, 1966-. II. Universidade Federal de Viçosa. Departamento de Química. Programa de Pós-Graduação em Agroquímica. III. Título.

CDD 22. ed. 543.5


JAMILLE CARVALHO DE SOUZA

**EMPREGO DA ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO PARA
CARACTERIZAÇÃO *IN SITU* DA COMPOSIÇÃO DE ALIMENTOS E DE
SUBSTÂNCIAS ILÍCITAS**


Tese apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Agroquímica para obtenção do título de *Doctor Scientiae*.

APROVADA: 10 de julho de 2024.

Assentimento:

Documento assinado digitalmente
 **JAMILLE CARVALHO DE SOUZA**
Data: 03/10/2024 20:31:40-0300
Verifique em <https://validar.iti.gov.br>

Jamille Carvalho de Souza
Autora

Documento assinado digitalmente
 **MARIA DO CARMO HESPANHOL**
Data: 03/10/2024 18:19:32-0300
Verifique em <https://validar.iti.gov.br>

Maria do Carmo Hespagnol
Orientadora

Dedico este trabalho à minha mãe, por todo amor e por todos os esforços para sempre fazer de meus objetivos, seus objetivos e de meus sonhos, seus sonhos.

AGRADECIMENTOS

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES)-Segurança Pública e Ciências Forenses (CAPES PROCAD – SPCF) [número do processo 88887.608483/2021-00] pelo fornecimento da bolsa de doutorado.

Ao Instituto Nacional de Ciências e Tecnologias Analíticas Avançadas (INCTAA)/Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) [número do processo INCTAA/CNPq 465768/2014-8; CNPq 305649/2021-3] e à CAPES PROCAD – SPCF [número do processo 88881.516445/2020-01] pelo auxílio financeiro.

À Universidade Federal de Viçosa (UFV), ao Departamento de Química (DEQ – UFV), ao Programa de Pós-Graduação em Agroquímica (PPAQ – UFV) e ao Instituto Nacional de Criminalística (INC – DPF), por permitirem a concretização deste trabalho.

À Professora Dra. Maria do Carmo Hespanhol, por ter aceitado me orientar, pelos ensinamentos, conselhos, paciência e parceria, durante o desenvolvimento deste trabalho e também ao longo de todos os anos de orientação.

Ao Professor Dr. Celio Pasquini, pelas aulas teóricas, ensinamentos, discussões e contribuições no desenvolvimento deste trabalho.

Ao Perito Criminal Federal Dr. Adriano Otávio Maldaner, pela parceria e contribuições no desenvolvimento deste trabalho e também pelos conselhos e acolhimento em Brasília.

Aos professores que aceitaram o convite para participar desta defesa.

Aos meus pais, Jorge e Miriam, pelo amor, apoio e confiança incondicionais.

Ao meu irmão Jefferson e à minha tia Renata, pelo amor incondicional e apoio.

À minha sobrinha Ágatha, por me dar forças nos dias difíceis com sua doçura, inocência e perspicácia.

Aos meus amigos, em especial, Ana Carolina, Ana Luiza e Gabriel, por torcerem pelo meu sucesso e por tornarem essa caminhada mais fácil.

Aos meus colegas de trabalho do GAES, pelo apoio e por me fornecerem toda a ajuda necessária para a realização deste trabalho. Em especial, ao Kaíque e à Thaís, pela presença, companheirismo e amizade.

Aos meus colegas de trabalho do LCP e do DEF – UFV, pela acolhida, apoio e compreensão.

Aos novos amigos e colegas que o desenvolvimento deste trabalho me trouxe. Em especial, à Jennifer, pela amizade e parceria.

Enfim, a todos que de alguma forma contribuíram para a realização deste trabalho, o meu sincero agradecimento!

*“Porque se chamavam homens
Também se chamavam sonhos
E sonhos não envelhecem.”*

Milton Nascimento – Clube da esquina nº 2

RESUMO

SOUZA, Jamille Carvalho, D.Sc., Universidade Federal de Viçosa, julho de 2024. **Emprego da espectroscopia no infravermelho próximo para caracterização *in situ* da composição de alimentos e substâncias ilícitas.** Orientadora: Maria do Carmo Hespanhol.

Dois espectrofotômetros operando na região do infravermelho próximo (NIR) miniaturizados e de baixo custo: NanoNIR (*Texas Instruments Inc., EUA*), que monitora a faixa espectral de 900 a 1700 nm, e NeoSpectra micro-FT-NIR (*Si-ware, Egito*), operando entre 1350 e 2550 nm; foram avaliados para determinar características de cafés comerciais e para caracterizar amostras de cocaína apreendidas pela Polícia Federal. Além disso, também foi proposta uma nova abordagem analítica baseada na espectroscopia NIR, usando o NeoSpectra, análise de correlação e análise multivariada de dados para selecionar produtos comerciais. Essa abordagem foi aplicada em amostras de edulcorantes vendidas à granel. Para tanto, foram coletados espectros NIR de conjuntos abrangentes de amostras de cafés de cápsulas, cocaína e edulcorantes. Esses espectros foram pré-processados e tratados com o software Unscrambler 11.0 (*Aspen Technologies, Noruega*), avaliados utilizando análise de componentes principais (PCA), modelagem independente e flexível por analogias de classes (SIMCA), resolução de curva multivariada de mínimos quadrados alternantes (MCR-ALS) e mínimos quadrados parciais (PLS). No caso dos cafés, diferenças entre lotes de um mesmo tipo de café puderam ser observadas e modelos quantitativos baseados em PLS foram construídos para determinar as características de intensidade, amargor, corpo, torra e acidez das amostras. Já para as amostras de cocaína, ambos os instrumentos mostraram a diferença entre cocaína base e cloridrato de cocaína empregando-se PCA. Além disso, foram construídos modelos de classificação SIMCA e de regressão PLS, a fim de identificar e quantificar, respectivamente, o teor de cocaína nas amostras. E, por fim, o estudo de caso dos edulcorantes à granel mostrou que a combinação de técnicas multivariadas com inspeção visual de espectros e correlação espectral produz resultados qualitativos e quantitativos confiáveis, reduzindo o custo de implementação ao evitar o uso de instrumentos caros e grandes conjuntos de dados de amostras analisadas por métodos analíticos de referência. Os resultados revelaram o potencial dos espectrofotômetros NIR miniaturizados e de baixo custo para caracterizar qualitativa e quantitativamente diversos tipos de amostras de interesse comercial e forense.

Palavras-chave: Espectroscopia no infravermelho próximo; Espectrofotômetros NIR portáteis; Análise *in situ*; Análise multivariada de dados; Análise de correlação espectral; Café torrado e moído; Cloridrato de cocaína; Cocaína base livre; Edulcorantes à granel.

ABSTRACT

SOUZA, Jamille Carvalho, D.Sc., Universidade Federal de Viçosa, July, 2024. **Use of near-infrared spectroscopy for in situ characterization of the composition of foods and illicit substances.** Adviser: Maria do Carmo Hespanhol.

Two miniaturized, low-cost near-infrared (NIR) spectrophotometers: NanoNIR (Texas Instruments Inc., USA), which monitors the spectral range from 900 to 1700 nm, and NeoSpectra micro-FT-NIR (Si-ware, Egypt), operating between 1350 and 2550 nm; were evaluated to determine characteristics of commercial coffees and to characterize cocaine samples seized by the Brazilian Federal Police. Furthermore, a new analytical approach based on NIR spectroscopy, using NeoSpectra, correlation analysis and multivariate data analysis to select commercial products was also proposed. This approach was applied to bulk sweeteners. Therefore, NIR spectra were collected from comprehensive sample sets of capsule coffee, cocaine and sweeteners. These spectra were pre-processed and treated with Unscrambler 11.0 software (Aspen Technologies, Norway), and were evaluated using principal component analysis (PCA), soft independent modelling by class analogy (SIMCA), multivariate curve resolution – alternating least squares (MCR-ALS), and partial least squares (PLS). In the case of coffees, differences between batches of the same type of coffee could be observed and quantitative models based on PLS were built to determine intensity, bitterness, body, roast and acidity of the samples. For the cocaine samples, both instruments showed the difference between free base cocaine and cocaine hydrochloride employing PCA. Furthermore, SIMCA classification and PLS regression models were constructed in order to identify and quantify, respectively, the cocaine content in the samples. And finally, the bulk sweeteners case study showed that combining multivariate techniques with visual inspection of spectra and spectral correlation produces reliable qualitative and quantitative results, reducing implementation cost by avoiding the use of expensive instruments and large sets of data from samples analyzed by reference analytical methods. The results revealed the potential of miniaturized, low-cost NIR spectrophotometers to qualitatively and quantitatively characterize different kinds of samples of commercial and forensic interest.

Keywords: Near infrared spectroscopy; Portable near infrared spectrophotometers; In situ analysis; Multivariate data analysis; Spectral correlation analysis; Ground roasted coffee; Cocaine hydrochloride; Free base cocaine; Bulk sweeteners.

SUMÁRIO

INTRODUÇÃO GERAL	11
Capítulo 1: Viabilidade do uso de espectrofotômetros compactos de infravermelho próximo e análise multivariada de dados para avaliar características de café torrado e moído	14
RESUMO.....	14
1.1 INTRODUÇÃO	16
1.2 OBJETIVO	18
1.3 MATERIAIS E MÉTODO	19
1.3.1 Amostras	19
1.3.2 Espectrômetros NIR compactos.....	21
1.3.3 Protocolo de medição das amostras.....	21
1.3.4 Tratamento dos dados.....	22
1.3.5 Análise quimiométrica	22
1.4 RESULTADOS E DISCUSSÃO	24
1.4.1 Resultados multivariados qualitativos	24
1.4.1.1 Avaliação preliminar das variáveis sensoriais	24
1.4.1.2 PCA dos conjuntos de dados para ambos os instrumentos	25
1.4.1.3 Agrupamento de acordo com o teor de cafeína	27
1.4.1.4 Agrupamento de acordo com os tipos de café.....	29
1.4.1.5 Agrupamento de acordo com a variedade.....	31
1.4.1.6 Agrupamento de acordo com a origem do café	31
1.4.1.7 Agrupamento de acordo com as características sensoriais.....	32
1.4.1.8 Avaliação conjunta da viabilidade de agrupamento	34
1.4.1.9 Análise detalhada de agrupamento.....	34
1.4.2 Resultados quantitativos multivariados para as características sensoriais	37
1.5 CONCLUSÕES.....	44
Capítulo 2: Emprego da espectroscopia no infravermelho próximo para a identificação, autenticidade e adulterações de amostras de cocaína apreendidas pela Polícia Federal.	46
RESUMO.....	46
2.1 INTRODUÇÃO	47
2.2 OBJETIVO	49
2.3 MATERIAIS E MÉTODO	50
2.3.1 Amostras	50
2.3.2 Espectrômetros infravermelho próximo	51

2.3.3	Frasco de medida e protocolo de aquisição dos espectros	51
2.3.4	Análise do conjunto de dados.....	52
2.4	RESULTADOS E DISCUSSÃO	54
2.4.1	Desempenho dos espectrômetros	54
2.4.2	Características dos espectros de infravermelho próximo de amostras de cocaína	56
2.4.3	Análise de Componentes Principais (PCA).....	59
2.4.4	Modelos de classificação SIMCA	62
2.4.5	Modelos de regressão PLS.....	65
2.5	CONCLUSÕES.....	71
Capítulo 3: Explorando a quimiometria para análise de triagem por espectroscopia no infravermelho próximo sem o apoio de métodos analíticos de referência: aplicação a edulcorantes à granel		
	RESUMO.....	72
3.1	INTRODUÇÃO	73
3.2	OBJETIVO	76
3.3	MATERIAIS E MÉTODO	77
3.3.1	Amostras	77
3.3.2	Espectrofotômetro NIR e dispositivo de sondagem de amostras.....	78
3.3.3	Protocolo de medição das amostras.....	78
3.3.4	Tratamento dos dados espectrais.....	78
3.4	RESULTADOS E DISCUSSÃO	80
3.4.1	Abordagens multivariadas qualitativas	80
3.4.1.1	PCA e correlação espectral.....	80
3.4.1.2	Projeção e SIMCA para classificação de edulcorantes comerciais à granel	87
3.4.2	Abordagens multivariadas quantitativas.....	90
3.4.2.1	Avaliação do MCR-ALS para estimar o teor dos adulterantes.....	90
3.4.2.2	Avaliação do PLS para a estimativa do teor dos adulterantes.....	98
3.4.2.3	Comparação entre MCR-ALS e PLS para fins quantitativos.....	100
3.5	CONCLUSÕES.....	101
CONCLUSÕES GERAIS		
	REFERÊNCIAS BIBLIOGRÁFICAS	105
Apêndice A. Capítulo 1		
	Apêndice B. Capítulo 3.....	116

INTRODUÇÃO GERAL

Investir no desenvolvimento de metodologias analíticas rápidas e que não destruam ou alterem o material analisado pode ser muito interessante para alguns tipos de amostras, especialmente no caso de alimentos, medicamentos e substâncias ilícitas. Nesse sentido, o uso da espectroscopia no infravermelho próximo (NIRS) se mostra como uma alternativa viável para grande parte dessas matrizes, uma vez que é uma técnica não-destrutiva, rápida e que apresenta resposta analítica para amostras que contenham ligações C-H, N-H, S-H, O-H, entre outras, em sua composição [1]. Além disso, com o advento da miniaturização e a produção de instrumentos portáteis de mais baixo custo, há a possibilidade de desenvolver diversas metodologias baseadas em NIRS, que permitam a análise direta, rápida e *in situ* de amostras, contribuindo assim para a triagem e o controle de qualidade de substâncias e produtos em diferentes circunstâncias.

Entretanto, é importante destacar os principais desafios do emprego da NIRS nessas condições, que são o de atestar a qualidade e a robustez dos instrumentos portáteis de baixo custo disponíveis, e o de extrair informações analíticas dos espectros NIR, dada a complexidade da composição das matrizes e a variação do teor dos analitos nas amostras analisadas, que se refletem na complexidade dos espectros NIR obtidos (visto que, na análise direta, não há nenhum tipo de preparo de amostra e todos os seus componentes poderão contribuir para o sinal analítico).

Os principais instrumentos portáteis utilizados no desenvolvimento de métodos empregando NIRS, descritos na literatura, são: MicroNIR – Viavi Solutions; NIRScan – Texas (Innospectra, Telspec-Telspec Inc.); SciO – Consumer Physics; NeoSpectra – Si-Ware; MicroPhazir – Thermo Fisher [2].

A fim de avaliar a capacidade desses instrumentos de produzir informações analíticas úteis, dois espectrofotômetros NIR miniaturizados de baixo custo foram empregados nos trabalhos que se seguem, sendo eles: NIRScan – Texas; e NeoSpectra – Si-Ware.

O espectrofotômetro NIRScan (Texas Instruments Inc.) monitora a região espectral de 900 a 1700 nm com uma resolução nominal de 10 nm. Duas pequenas fontes de radiação de filamento de tungstênio (0,7 W) estão incorporadas no instrumento, próximas à sua janela de medição. A medição é feita no modo de reflectância difusa e a radiação NIR, que retorna da amostra, passa por uma fenda de entrada e é direcionada para uma rede de difração fixa que projeta a radiação dispersa na superfície de um conjunto bidimensional de microespelhos. O arranjo espectrométrico do NanoNIR emprega a tecnologia MEMS (*Micro Electro Mechanical*

System), que controla o conjunto bidimensional de microespelhos, desviando um comprimento de onda selecionado em direção ao detector de ponto único de InGaAs, que registra as intensidades da radiação NIR [3].

O outro espectrofotômetro (NeoSpectra, Si-Ware, Egito) apresenta um interferômetro baseado em MEMS para monitorar a faixa espectral de 1350 a 2550 nm, e emprega a transformada de Fourier (FT) para recuperar os espectros de absorção. A resolução espectral reivindicada pelo fabricante está na faixa de 8 – 16 nm [4].

Juntamente com o uso de ferramentas quimiométricas, propôs-se desenvolver novos métodos para monitorar a qualidade e as características sensoriais de cafés de cápsula, a composição de amostras suspeitas de conterem cocaína, e a natureza e a autenticidade de edulcorantes comercializados à granel.

Inicialmente, esses dois instrumentos foram empregados para avaliar as características sensoriais de cafés torrados e moídos. O conjunto amostral foi constituído por amostras de cápsulas de café espresso da marca *Nespresso*. Essas amostras possuem um controle de qualidade rigoroso, o que minimiza a variabilidade entre amostras de mesmo tipo, e, além disso, o fornecedor disponibiliza diversas informações acerca das características dos cafés. Dessa forma, apesar de não terem sido feitas análises por outras técnicas analíticas, foi possível explorar amplamente os dados espectrais NIR obtidos e avaliar as potencialidades do uso desses instrumentos portáteis no desenvolvimento de métodos analíticos.

Posteriormente, esses instrumentos também foram empregados na análise de amostras de cocaína, apreendidas pela Polícia Federal do Brasil (PF). Neste caso, todas as amostras já haviam sido previamente analisadas por outros métodos de referência. Entretanto, trata-se de um conjunto de amostras mais heterogêneo. Sendo assim, o objetivo foi investigar a viabilidade do uso desses instrumentos portáteis e da técnica NIR para extrair informações acerca da composição das amostras e propor um novo método rápido e não-destrutivo de triagem de substâncias suspeitas.

E, por fim, no terceiro capítulo, é apresentada uma ideia inovadora de utilizar a espectroscopia NIR, empregando um instrumento portátil (neste caso, apenas o NeoSpectra foi usado), juntamente com a análise multivariada de dados, para rastrear a autenticidade de produtos comerciais, sem a necessidade de utilizar técnicas de referência. E, como estudo de caso, amostras comerciais de edulcorantes adquiridos à granel foram analisadas.

Além de visar o desenvolvimento de metodologias portáteis de análise direta de amostras sólidas, este trabalho também propôs reduzir os custos de implementação dos novos

métodos, através do emprego de instrumentos de baixo custo e da redução do emprego de métodos de referência.

Capítulo 1: Viabilidade do uso de espectrofotômetros compactos de infravermelho próximo e análise multivariada de dados para avaliar características de café torrado e moído

Este capítulo é baseado no manuscrito publicado:

Jamille Carvalho Souza, Celio Pasquini, Maria C. Hespanhol. Food Control, 138 (2022) 109041. <https://doi.org/10.1016/j.foodcont.2022.109041>

RESUMO

Dois espectrofotômetros no infravermelho próximo (NIR) miniaturizados e de baixo custo, operando em duas faixas espectrais diferentes (900–1700 nm e 1300–2500 nm) e empregando diferentes tecnologias espectrométricas, foram avaliados para determinar diversas características de cafés comerciais torrados e moídos, usando análise multivariada. O conjunto abrangente de amostras compreendeu cento e quarenta e cinco amostras de diferentes procedências, misturas de variedades (arábica/arábica e arábica/robusta) ou variedade única (arábica), descafeinados ou regulares, apresentando diversos graus de características sensoriais (amargor, intensidade, acidez, corpo e torra), conforme informado pelo fabricante. Os espectros representativos das amostras foram adquiridos utilizando um dispositivo de sondagem que proporciona deslocamento rotacional e linear do frasco da amostra. O conjunto de dados espectrais foi avaliado utilizando modelos de análise de componentes principais (PCA) com o objetivo de verificar a capacidade de discriminação de ambos os instrumentos. Quando avaliados até a quarta componente principal por ambos os instrumentos compactos, os cafés são agrupados e bem diferenciados de acordo com a origem, variedade, e presença de cafeína pelos *scores* da PCA. Diferenças entre lotes de um mesmo tipo de café podem ser observadas. Modelos quantitativos baseados na regressão de mínimos quadrados parciais (PLS) foram construídos para avaliar a viabilidade dos espectrofotômetros miniaturizados em determinar as características de intensidade, amargor, corpo, torra e acidez das amostras. Essas características sensoriais podem ser previstas por modelos de mínimos quadrados parciais com erros quadráticos médios de validação externa de 1,3, 0,7, 0,6, 0,5 e 0,8 e 1,2, 0,7, 0,5, 0,5 e 0,7 para intensidade, amargor, corpo, torra, e acidez, respectivamente, utilizando os conjuntos de dados obtidos pelos instrumentos NanoNIR e NeoSpectra e empregando variáveis selecionadas pelo algoritmo Jack-Knife. Os resultados quantitativos podem ser melhorados utilizando modelos de regressão local construídos com um conjunto restrito de amostras. Em geral, os resultados revelam o potencial dos espectrofotômetros NIR miniaturizados e de baixo custo para

caracterizar qualitativa e quantitativamente o café torrado e moído comercial, com uma ligeira vantagem de usar o instrumento que fornece a faixa de comprimento de onda mais abrangente e é baseado na tecnologia da transformada de Fourier.

Palavras-chave: Espectroscopia no infravermelho próximo; Espectrofotômetros compactos de infravermelho próximo; Café torrado e moído; Determinação de características de café; Classificação comercial do café; Análise multivariada.

1.1 INTRODUÇÃO

O café é uma das bebidas mais consumidas no mundo. Os grãos torrados, moídos e prontos para uso são a forma mais comum de comercialização do produto à disposição do cliente final. Recentemente, vários revendedores estão fornecendo cafés moídos de uma única origem, com misturas de variedades distintas (arábica/robusta ou arábica/arábica), processados para atingir uma diversidade de atributos sensoriais desejados. Cafés descafeinados também estão disponíveis para clientes intolerantes à cafeína. Intensidade, amargor, corpo, acidez e grau de torra são atributos sensoriais de cafés moídos normalmente reportados para produtos comerciais.

A variedade de produtos comercializados exige métodos analíticos precisos, rápidos, de baixo custo e ecologicamente corretos (verdes), capazes de avaliar objetivamente as características qualitativas e sensoriais do café moído comercial.

Métodos analíticos baseados em espectroscopia vibracional como espectroscopia infravermelho (IR) e infravermelho próximo (NIRS) e análise multivariada atendem à maioria dos requisitos para avaliar qualitativa e quantitativamente as principais características do café moído. Esse assunto já foi revisado [5], [6]. Diversos trabalhos relatam a eficácia dos métodos baseados na NIRS para determinar a origem geográfica [7], [8], [9], [10], variedade [11], [12], atributos sensoriais [13], [14], [15], qualidade [16], [17], adulteração [18], [19], composição [12], [20], e características específicas [21], [22] de grãos de café verdes ou torrados, inteiros ou moídos. Em comum, no entanto, esses métodos baseiam-se em conjuntos restritos de dados espectrais, considerando a variabilidade das amostras comercializadas, obtidos com instrumentos de bancada de alto custo.

Recentemente, as tecnologias de miniaturização permitiram a produção de instrumentos NIR compactos e de baixo custo (US\$ 1.000–5.000). Mais de vinte instrumentos miniaturizados estão atualmente disponíveis [1]. Poucos trabalhos utilizando esse tipo de instrumento são encontrados na literatura, descrevendo seu uso em métodos analíticos para avaliar um número restrito de características do café moído, como grau de torra, empregando amostras da mesma origem, blend ou monovarietal [23], [24], [25].

Devido à sua introdução relativamente recente no mercado, os instrumentos compactos e de baixo custo precisam ser avaliados de forma mais abrangente para determinar sua viabilidade na estimativa eficaz das características do café moído.

Este trabalho avalia o potencial de dois instrumentos NIR compactos em produzir informações precisas capazes de distinguir a origem, blends, marcas e quantificar as

propriedades sensoriais do café torrado moído. Foram avaliados dois instrumentos compactos e de baixo custo. Um é baseado na tecnologia dispersiva e o outro na tecnologia de transformada de Fourier (FT), monitorando diferentes faixas de comprimento de onda.

O objetivo deste trabalho não é validar definitivamente os resultados dos métodos de classificação e regressão, mas avaliar a viabilidade e as limitações do uso de instrumentos compactos e análise multivariada para discriminar entre diversos produtos comerciais de acordo com seus atributos de qualidade e características dos cafés torrados e moídos, conforme relatados pelo revendedor. Portanto, este trabalho utiliza amostras e emprega as informações fornecidas por uma empresa de café idônea para formar um conjunto abrangente de dados espectrais NIR, que permite avaliar e comparar o desempenho de instrumentos NIR compactos e análise multivariada a fim de avaliar diversas características de cafés torrados e moídos.

1.2 OBJETIVO

Desenvolver um novo método espectroscópico para avaliar características sensoriais de cafés torrados e moídos, empregando dois espectrofotômetros NIR portáteis e de baixo custo e análise multivariada de dados.

1.3 MATERIAIS E MÉTODO

1.3.1 Amostras

O conjunto amostral empregado neste trabalho compreende 145 cafés torrados e moídos retirados de cápsulas comerciais de café expresso (Nespresso, Lausanne, Suíça) de 28 tipos diferentes. A Tabela 1.1 apresenta as características do conjunto de dados de acordo com as informações encontradas na página do fabricante e nos rótulos das embalagens dos produtos.

Vários lotes (1 a 8) de diferentes tipos de café foram adquiridos diretamente do revendedor/fabricante durante um ano. Duas cápsulas de cada lote foram abertas e o conteúdo imediatamente transferido para frascos de vidro cilíndricos de 20 mL. Os tipos de café “Nicarágua” e “Volluto Descafeinado” foram selecionados para avaliar a repetibilidade (mesma amostra) e reprodutibilidade do mesmo lote (5 conjuntos de duas cápsulas da mesma embalagem) obtidas pelos espectros registrados pelos dois instrumentos NIR compactos. Aproximadamente, 10 g de café moído são suficientes para encher o frasco. O pó foi compactado dentro do frasco, vedado com filme de polipropileno e tampado. Depois de abertas, as amostras foram conservadas em geladeira a 10 °C.

Tabela 1.1. Descrição do conjunto amostral empregado neste trabalho

Tipo	N	R/D	Variedade	Origem	Intensidade	Amargor	Acidez	Corpo	Torra
Napoli	8	R	Blend Arábica/Robusta	América do Sul/Uganda	13	5	1	5	5
Kazaar	7	R	Blend Arábica/Robusta	Não especificado	12	5	1	5	5
India	6	R	Blend Arábica/Robusta	Índia	11	5	1	4	5
Ristretto	7	R	Blend Arábica/Robusta	América do Sul/África Oriental/Sul da Índia	10	4	3	4	4
Ristretto D	2	D	Blend Arábica/Robusta	América do Sul/África Oriental/Sul da Índia	10	4	3	4	4
Arpeggio	7	R	Blend Arábica	América Latina (Costa Rica)	9	4	2	4	4
Arpeggio D	2	D	Blend Arábica	América Latina (Costa Rica)	9	4	2	4	4
Cafezinho do Brasil	1	R	Blend Arábica	Brasil	9	3	1	2	3
Envivo Lungo	7	R	Blend Arábica/Robusta	Índia/México	9	4	1	4	4
Fortissimo Lungo	7	R	Blend Arábica	Índia	8	3	2	3	4
Indonesia	6	R	100% Arábica	Indonésia	8	3	2	4	4
Roma	6	R	Blend Arábica/Robusta	Brasil/México	8	4	4	3	3
Venezia	7	R	Blend Arábica/Robusta	Brasil/Colômbia/África	8	3	1	4	4
Galapagos	1	R	Não especificado	Galápagos	7	2	2	2	1
Sumatra Vieilli	1	R	Não especificado	Sumatra	7	2	0	2	3
Colombia	7	R	100% Arábica	Colômbia	6	3	4	3	3
Livanto	5	R	Blend Arábica	Brasil/Colômbia/Guatemala	6	3	3	3	3
Capriccio	5	R	Blend Arábica/Robusta	América do Sul (Brasil)	5	3	3	2	2
Nicaragua	9	R	100% Arábica	Nicarágua	5	2	2	2	2
Shanghai Lungo	2	R	Blend Arábica	Quênia/China/Indonésia	5	1	2	2	2
Buenos Aires Lungo	2	R	Blend Arábica/Robusta	Colômbia/Uganda	4	1	2	1	1
Cosi	6	R	Blend Arábica	Costa Rica/Quênia	4	2	3	2	2
Ethiopia	7	R	100% Arábica	Etiópia	4	2	4	2	2
Linizio Lungo	4	R	Blend Arábica	Brasil/Colômbia	4	1	1	2	2
Vivalto Lungo	6	R	Blend Arábica	Brasil/Colômbia	4	2	2	2	3
Vivalto Lungo D	1	D	Blend Arábica	Brasil/Colômbia	4	2	2	2	3
Volluto	8	R	Blend Arábica	Brasil/Colômbia	4	2	3	2	2
Volluto D	8	D	Blend Arábica	Brasil/Colômbia	4	2	3	2	2

N: número de lotes; R: café regular; D: café descafeinado.

1.3.2 Espectrômetros NIR compactos

Dois instrumentos NIR compactos foram empregados neste estudo. Um (instrumento 1) é baseado em tecnologia dispersiva e usa um conjunto de microespelhos baseados em MEMS (sistemas micro-eletromecânicos) para analisar os comprimentos de onda na faixa de 900 a 1700 nm (NanoNIR -Texas Instruments Inc. Austin, Texas, EUA). A resolução espectral nominal é de 10 nm. Os espectros foram registrados com 228 pontos de comprimento de onda. Detalhes sobre o desempenho deste espectrofotômetro podem ser encontrados em artigo anterior [26].

O outro espectrofotômetro (aqui denominado de instrumento 2 – NeoSpectra, Si-Ware, Egito) apresenta um interferômetro baseado em MEMS para monitorar a faixa espectral de 1350 a 2550 nm, empregando transformada de Fourier (FT) para recuperar os espectros de absorção. Foram registrados 257 pontos de comprimento de onda por espectro. A resolução espectral reivindicada pelo fabricante está na faixa de 8–16 nm.

1.3.3 Protocolo de medição das amostras

Os frascos de vidro contendo amostras de café foram medidos representativamente no modo de reflectância difusa, empregando o dispositivo de sondagem linear-rotacional descrito em artigo anterior [27]. As medidas foram feitas na temperatura do laboratório de 25 °C. Os instrumentos compactos foram inseridos no compartimento próprio do dispositivo de sondagem de forma intercambiável, e todas as amostras tiveram seus espectros de reflectância adquiridos e expressos em absorbância em função do comprimento de onda.

A referência utilizada foi um frasco preenchido com Spectralon®, garantindo 100% de reflectância sobre os comprimentos de onda de ambos os espectrofotômetros e compensando qualquer absorção da parede do frasco de vidro. Para o instrumento 1, cada espectro representativo das amostras foi o resultado da média de três varreduras com 90 espectros cada, adquiridos em um intervalo de tempo de 20 s. Para o instrumento 2, os espectros representativos resultaram de três varreduras durante um tempo de integração de 20 s cada. O fabricante relata que o interferômetro realiza milhares de subvarreduras durante o tempo de integração empregado. Para ambos os instrumentos, o tempo total de medição por amostra foi de cerca de 60 s.

1.3.4 Tratamento dos dados

Todos os espectros médios representativos foram transformados pela segunda derivada usando o algoritmo Savitzky-Golay, empregando uma janela simétrica de 15 pontos e um polinômio de 2º grau. Os espectros adquiridos pelo instrumento 1 tiveram as extremidades da faixa de comprimentos de onda (que possuem baixa razão sinal-ruído) removidas após a derivação, resultando em espectros constituídos por 179 pontos na faixa de 998–1625 nm. Toda a faixa espectral monitorada pelo instrumento 2 foi empregada.

1.3.5 Análise quimiométrica

Os conjuntos de dados foram processados utilizando o software Unscrambler 11.0 (Aspen Technologies, Noruega). Modelos baseados em análise de componentes principais (PCA) de todo o conjunto de amostras foram construídos usando os conjuntos de dados com segunda derivada, obtidos pelos dois espectrofotômetros compactos. Os dados foram centrados na média e o algoritmo NIPALS foi empregado para extrair as componentes principais.

Um modelo PCA adicional foi construído utilizando os valores das cinco características sensoriais reportadas pelas amostras de café, empregando o conjunto de dados centrado e padronizado.

Modelos de regressão baseados em mínimos quadrados parciais (PLS), avaliados para prever as características sensoriais de amostras de café utilizando informações espectrais, foram construídos centralizando os dados espectrais e os valores das características. O melhor número de variáveis latentes para cada modelo foi encontrado com base nos resultados da validação cruzada realizada pela divisão aleatória do conjunto amostral de calibração em 20 segmentos.

Neste trabalho, foram utilizados os valores das características sensoriais fornecidos pelo fabricante do café para construir e avaliar o desempenho dos modelos PLS. No entanto, eles foram selecionados usando modelos PLS preliminares e o recurso da relação X-Y de outlier fornecido pelo programa Unscrambler [28]. Este recurso analisa o comportamento dos *scores* das características versus os *scores* dos dados espectrais para cada fator PLS, ajudando a identificar amostras que são discrepantes na relação da propriedade modelada com as características espectrais. No presente caso, nenhuma amostra pôde ser excluída com base nos resultados da análise da relação X-Y de outliers, contribuindo para atestar a qualidade das informações supridas pelo fabricante.

O algoritmo Jack-Knife oferecido pelo Unscrambler 11.0 foi empregado para selecionar

as variáveis espectrais que fornecem os coeficientes de regressão mais estáveis [28]. Essas variáveis selecionadas foram utilizadas para construir e validar novos modelos PLS para as características sensoriais.

1.4 RESULTADOS E DISCUSSÃO

1.4.1 Resultados multivariados qualitativos

1.4.1.1 Avaliação preliminar das variáveis sensoriais

A primeira avaliação do conjunto de dados foi realizada utilizando apenas as variáveis de características sensoriais informadas pelo revendedor dos cafés. A correlação entre intensidade, acidez, amargor, corpo e grau de torra foi avaliada e está apresentada na Tabela 1.2. Foi observada forte correlação ($r > 0,9$) entre intensidade, amargor, corpo e grau de torra. A acidez apresentou correlação inversa e fraca ($r = -0,3$ a $-0,5$) com as demais características.

Tabela 1.2. Resultados da análise de correlação entre as características sensoriais dos cafés analisados

Características sensoriais	Intensidade	Amargor	Acidez	Corpo	Torra
Intensidade	1	0,937	-0,520	0,942	0,927
Amargor	0,937	1	-0,329	0,904	-0,545
Acidez	-0,520	-0,329	1	-0,478	-0,545
Corpo	0,942	0,904	-0,478	1	0,946
Torra	0,927	0,891	-0,545	0,946	1

Um modelo PCA construído com os dados das características (intensidade, acidez, amargor, corpo e grau de torra) padronizados (centrados na média e ponderados pelo desvio padrão) também evidencia a forte correlação entre o conjunto de características sensoriais, como pode ser observado pelo gráfico de *loadings*, representado na Figura 1.1.

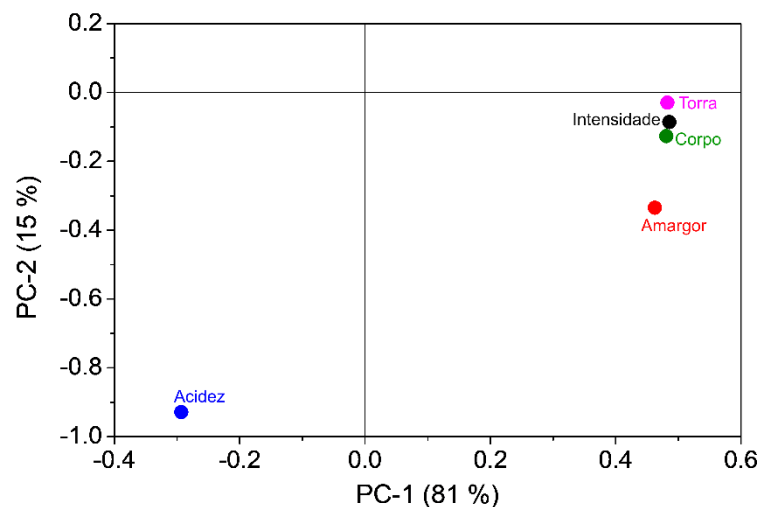


Figura 1.1. Gráfico de *loadings* resultante da PCA das características sensoriais das amostras de café.

Com base nesses resultados, é possível antecipar que, se presentes nos espectros NIR, as mesmas informações NIR do conjunto de amostras de cafés provavelmente serão utilizadas por modelos de regressão visando determinar essas características sensoriais correlacionadas. Este fato é relevante para interpretar os resultados obtidos por PCA e PLS. A única característica sensorial independente identificada no conjunto de dados atual é a acidez.

1.4.1.2 PCA dos conjuntos de dados para ambos os instrumentos

Todos os conjuntos de dados espectrais obtidos pelos dois instrumentos, mostrados nas Figuras 1.2A e B e 1.3A e B, foram submetidos à PCA. Esses resultados permitem avaliar diversas informações qualitativas sobre a viabilidade de ambos os instrumentos analisarem as semelhanças e diferenças entre as amostras de café e suas características.

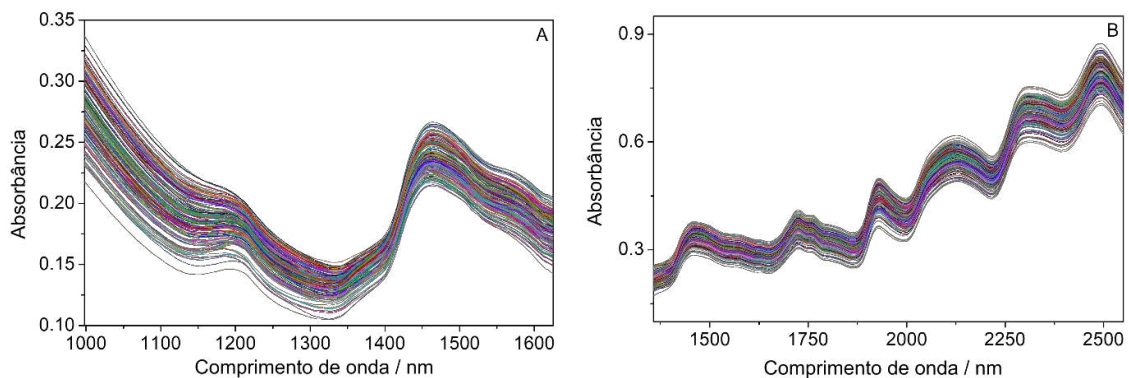


Figura 1.2. Conjuntos de espectros brutos das amostras de café torrado e moído obtidos por (A) instrumento 1, (B) instrumento 2.

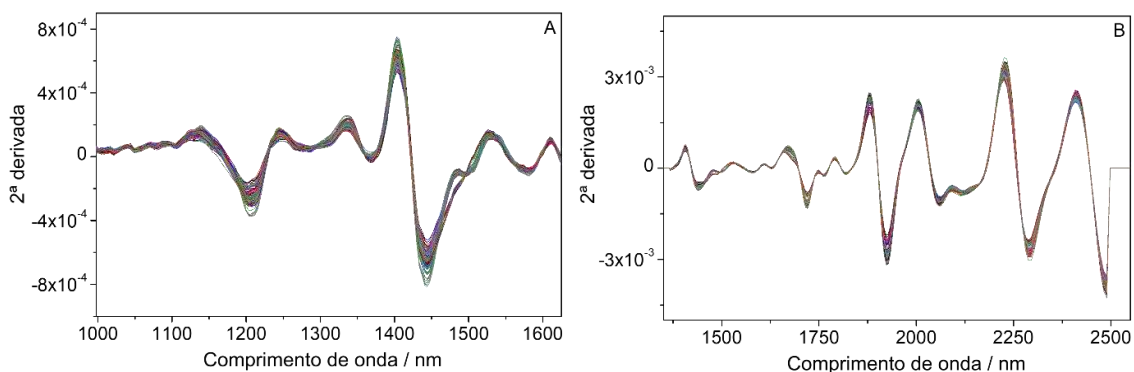


Figura 1.3. Segunda derivada (Savitzky-Golay, janela simétrica de 15 pontos, polinômio de 2º grau) dos conjuntos de espectros das amostras de café torrado e moído obtidos pelo (A) instrumento 1, (B) instrumento 2.

Para avaliar a real capacidade dos instrumentos em discriminar entre as diferentes amostras de café, é essencial estabelecer limites de referência de variabilidade para os valores de *scores*. Essas referências foram estabelecidas observando-se a dispersão dos valores de

scores para as três réplicas da mesma amostra (repetibilidade), medidas sequencialmente em um curto intervalo de tempo (~5 min), e para a mesma amostra do mesmo lote medida cinco vezes em dias diferentes durante oito meses (reprodutibilidade). Este estudo utilizou dois tipos de café (“Volluto descafeinado” e “Nicarágua”).

A Figura 1.4 mostra a distribuição dos *scores* de PC-1 e PC-2, a partir dos dados espectrais obtidos pelos instrumentos 1 (A) e 2 (B), respectivamente. Juntas, essas duas PCs capturam 80 e 87% da variância total do conjunto de dados, para os instrumentos 1 e 2, respectivamente.

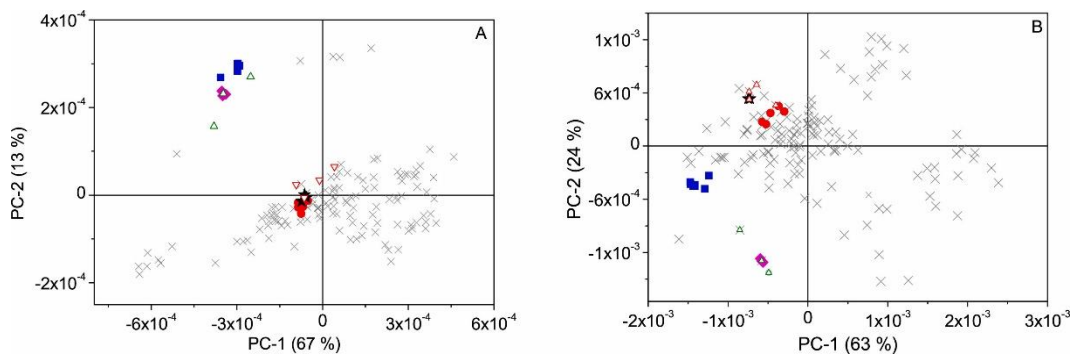


Figure 1.4. Dispersão dos valores de *scores* da PCA, obtidos para os testes de repetibilidade e reprodutibilidade, em comparação com a dispersão dos *scores* de todo o conjunto de dados. (A) Instrumento 1; (B) Instrumento 2. \times outros tipos de café; \bullet reprodutibilidade para o café *Nicarágua*; \blacksquare reprodutibilidade para o *Volluto descafeinado*; \star repetibilidade do café *Nicarágua*; \blacklozenge repetibilidade do café *Volluto descafeinado*; ∇ outros lotes de *Nicarágua*; \triangle outros lotes de *Volluto descafeinado*.

A repetibilidade e reprodutibilidade dos valores dos *scores* podem ser observadas na figura acima. O conhecimento dessas dispersões permite avaliar com segurança se um agrupamento observado ou distâncias detectadas considerando qualquer outra característica das amostras de café são, de fato, significativas. Por exemplo, pequenas diferenças entre lotes de mesmos tipos de café podem ser observadas por uma dispersão de *score* mais significativa do que a obtida nos testes de repetibilidade e reprodutibilidade. Este tipo de informação pode ser utilizado para comprovar o controle de qualidade dos produtos comercializados.

Os *loadings* das PCAs, até a quarta componente principal, para todos os conjuntos de dados espectrais podem ser observados a seguir, nas Figuras 1.5A–B e 1.6A–B, para os instrumentos 1 e 2, respectivamente.

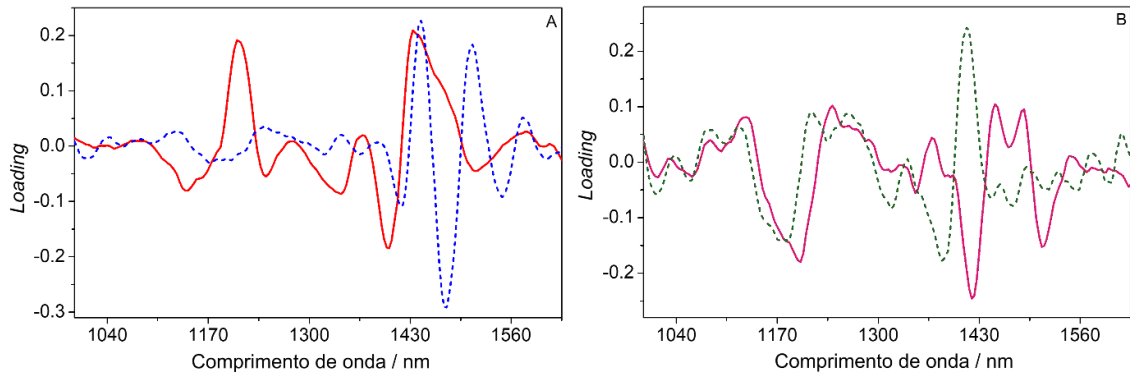


Figura 1.5. Gráficos de *loadings* da PCA obtida analisando-se todo o conjunto de dados espectrais, utilizando o instrumento 1. (A) — PC-1 e - - - PC-2; (B) — PC-3 e - - - PC-4.

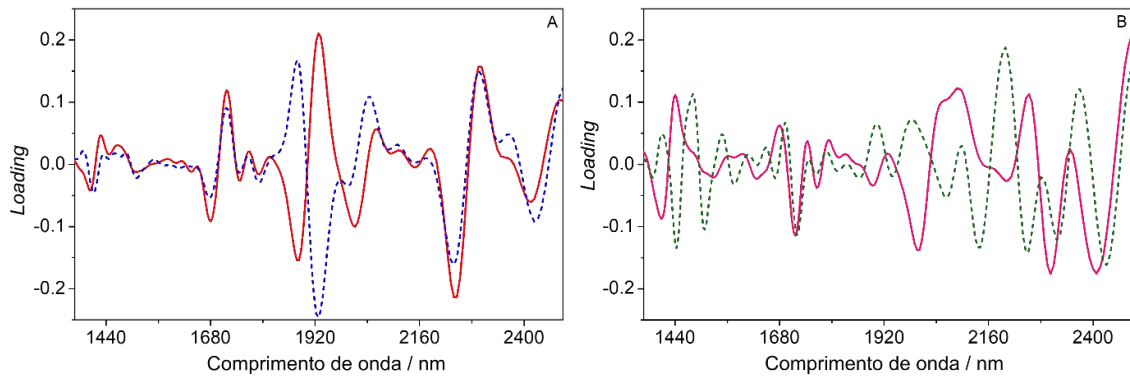


Figura 1.6. Gráficos de *loadings* da PCA obtida analisando-se todo o conjunto de dados espectrais, utilizando o instrumento 2. (A) — PC-1 e - - - PC-2; (B) — PC-3 e - - - PC-4.

1.4.1.3 Agrupamento de acordo com o teor de cafeína

A Figura 1.7A e B mostra que ambos os instrumentos podem fornecer informações espectrais para distinguir entre amostras de café normal e descafeinado. A classificação independe do atributo de intensidade e de outras características da amostra. O instrumento 1 manifesta a diferença logo nas duas primeiras PCs, enquanto o instrumento 2 captura as informações relevantes para distinguir as classes na quarta PC. Este fato mostra que os espectros obtidos pelo instrumento 1 são limitados no conteúdo de informação em comparação com o instrumento 2. Esta limitação resulta da menor faixa espectral monitorada pelo instrumento 1. Além disso, sua faixa espectral refere-se principalmente ao segundo e terceiro harmônicos dos modos vibracionais, enquanto o instrumento 2 pode avaliar a faixa espectral onde se encontram as combinações mais intensas e informativas dos modos vibracionais.

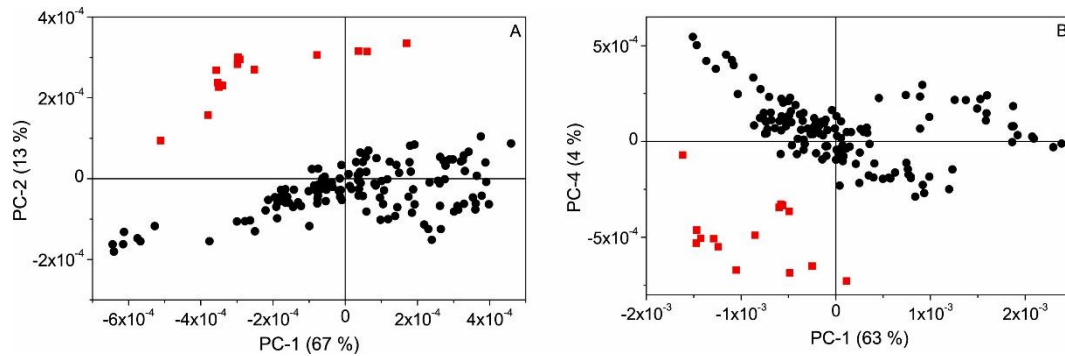


Figura 1.7. Gráficos de distribuição de *scores* da PCA de todo o conjunto de dados, mostrando a discriminação entre café descafeinado e café regular (com cafeína). (A) Instrumento 1; (B) Instrumento 2. ● Regular e ■ descafeinado.

A 2ª PC resultante da análise do conjunto de dados obtido pelo instrumento 1 é a mais relevante para distinguir entre amostras de cafés regulares e descafeinadas. O gráfico de *loadings* da PCA para esta componente (Figura 1.5A) mostra elevados valores de *loadings* coincidindo com os comprimentos de onda que evidenciam diferenças espectrais significativas entre os dois tipos de café. O mesmo comportamento é observado para a 4ª PC, utilizando o instrumento 2 (Figura 1.6B).

As diferenças espectrais que resultam na distinção entre os cafés normais e descafeinados coincidem parcialmente com as características espectrais presentes no espectro puro da cafeína [29]. Este é o caso das regiões “b” e “c” dos espectros obtidos pelo instrumento 2, mostradas na Figura 1.8 a seguir.

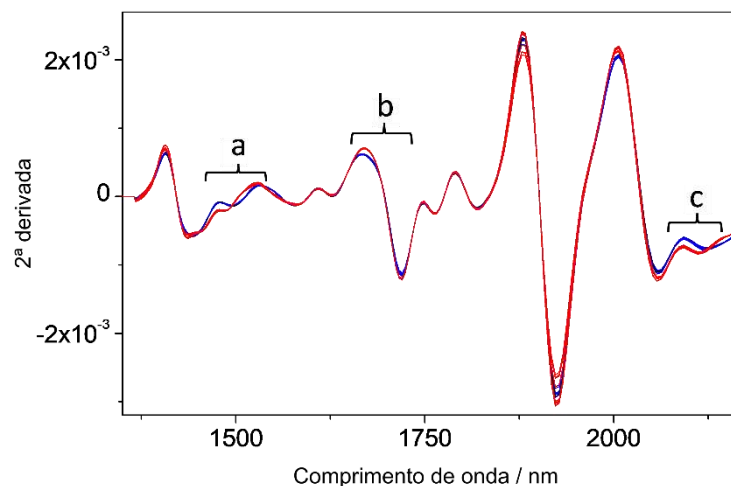


Figura 1.8. Segunda derivada do conjunto de espectros obtido pelo instrumento 2 do café regular (em vermelho) e descafeinado (em azul) do tipo *Volluto*. As regiões “a”, “b” e “c” evidenciam os comprimentos de onda que apresentam maiores diferenças espectrais entre o café regular e o descafeinado.

Essas regiões podem ser atribuídas a características espectrais associadas ao primeiro e segundo sobretons com uma pequena contribuição de combinações binárias e ternárias (b) e de combinações binárias e ternárias de modos vibracionais (c). Os espectros derivados são do mesmo tipo de café (*Volluto*), regular ou descafeinado, apresentando as mesmas características sensoriais, origem e variedade. No entanto, outras diferenças observadas não podem se correlacionar com o espectro puro da cafeína. Em particular, a diferença observada no espectro da segunda derivada na região entre 1464 e 1533 nm (claramente notada nos espectros da derivada do instrumento 2 da Figura 1.8, região “a”). Embora seja uma das regiões mais evidentes para distinguir as amostras de café regular das descafeinadas, não pode ser associada às características espectrais da cafeína[29]. Provavelmente, esta diferença é causada por alterações adicionais na composição química do café resultantes do processo de descafeinação.

1.4.1.4 Agrupamento de acordo com os tipos de café

Para avaliar o potencial de ambos os instrumentos em distinguir entre os tipos de café, foi realizada uma PCA utilizando apenas amostras de cafés regulares. O conjunto de dados contém 24 tipos diferentes de cafés regulares e 132 amostras.

A Figura 1.9A e B mostra os *scores* resultantes das PCAs obtidas para os instrumentos 1 e 2, respectivamente, categorizados de acordo com os tipos de cafés. Os *loadings* das PCAs, até a quarta componente principal, para os conjuntos de dados espectrais de amostras regulares, podem ser observados nas Figuras 1.10A e B e 1.11A e B, para os instrumentos 1 e 2, respectivamente.

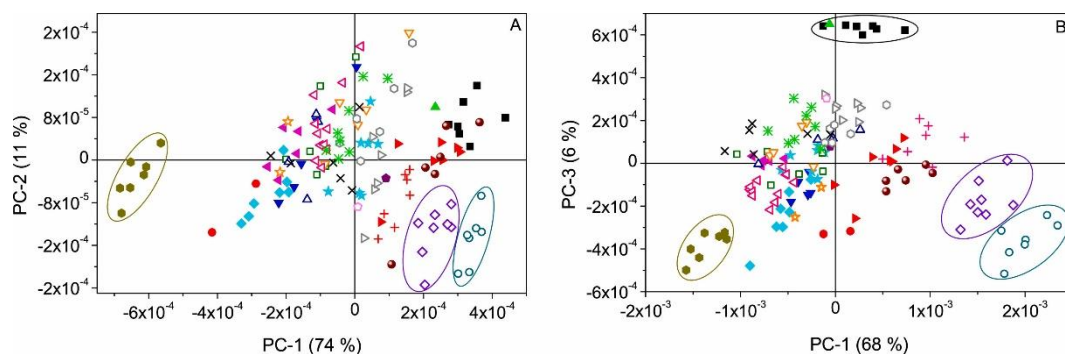


Figura 1.9. Distribuição dos *scores* das PCAs categorizados de acordo com os tipos de cafés regulares. (A) Instrumento 1, (B) Instrumento 2. ■ *Arpeggio*, ● *Buenos Aires Lungo*, ▲ *Cafezinho do Brasil*, ▼ *Capriccio*, ◆ *Colombia*, ◆ *Cosi*, ► *Envivo Lungo*, ◆ *Ethiopia*, ★ *Fortissimo Lungo*, ◆ *Galapagos*, ● *India*, □ *Indonesia*, ○ *Kazaar*, △ *Linizio Lungo*, ▽ *Livanto*, ◆ *Napoli*, ▽ *Nicaragua*, ▽ *Ristretto*, ○ *Roma*, ★ *Shanghai Lungo*, ◆ *Sumatra Vieilli*, + *Venezia*, × *Vivalto Lungo*, e * *Volluto*.

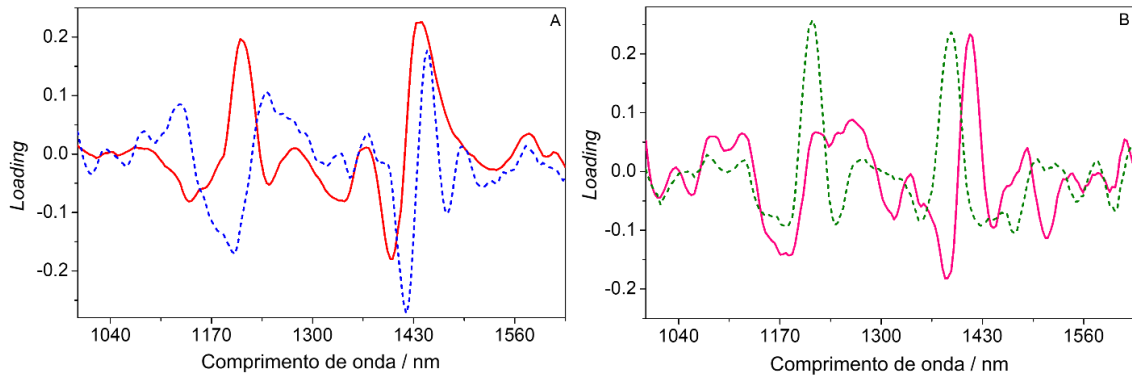


Figura 1.10. Loadings da PCA obtidos analisando o conjunto de dados de amostras de cafés regulares usando o instrumento 1. (A) — PC1, e - - - PC2; (B) — PC3, e - - - PC4.

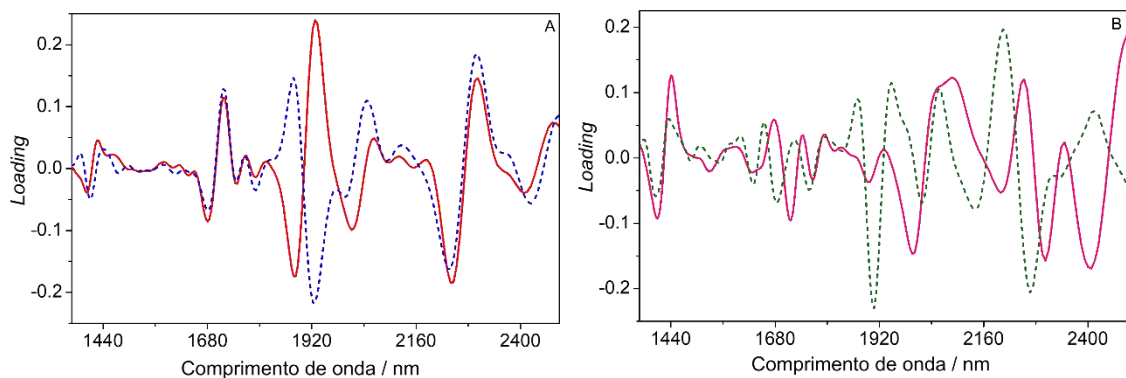


Figura 1.11. Loadings da PCA obtidos analisando o conjunto de dados de amostras de cafés regulares usando o instrumento 2. (A) — PC1, e - - - PC2; (B) — PC3, e - - - PC4.

Uma melhor diferenciação foi observada analisando-se os *scores* de PC1 e PC2 e os *scores* de PC1 e PC3 para os instrumentos 1 e 2, respectivamente. A distinção entre os tipos fica evidente em diversas amostras, enquanto muitas outras são apresentadas como um único grupo. Os dados do instrumento 1 permitem identificar claramente três grupos referentes aos cafés do tipo *Kazaar*, *Napoli* e *Ethiopia*.

O instrumento 2 apresenta desempenho um pouco melhor exibindo quatro grupos bem separados ao analisar os *scores* de PC1 e PC3. Além das três classes identificadas pelo instrumento 1, o instrumento 2 identifica os cafés do tipo *Arpeggio*. O potencial qualitativo também foi avaliado utilizando PCs de ordem inferior. Porém, não foi obtido ganho significativo de classificação para ambos os instrumentos ao observar os resultados até a quinta componente principal.

É possível observar uma baixa dispersão dos *scores* para todos os tipos de cafés, atestando o bom controle do processo de fabricação.

1.4.1.5 Agrupamento de acordo com a variedade

Existem três classes de variedades presentes no conjunto de dados dos cafés regulares: arábica puro, blends de arábica e blends de arábica e robusta. O melhor agrupamento foi observado analisando-se os *scores* das duas primeiras e da primeira e terceira PCs dos instrumentos 1 e 2, respectivamente, conforme mostrado na Figura 1.12A e B.

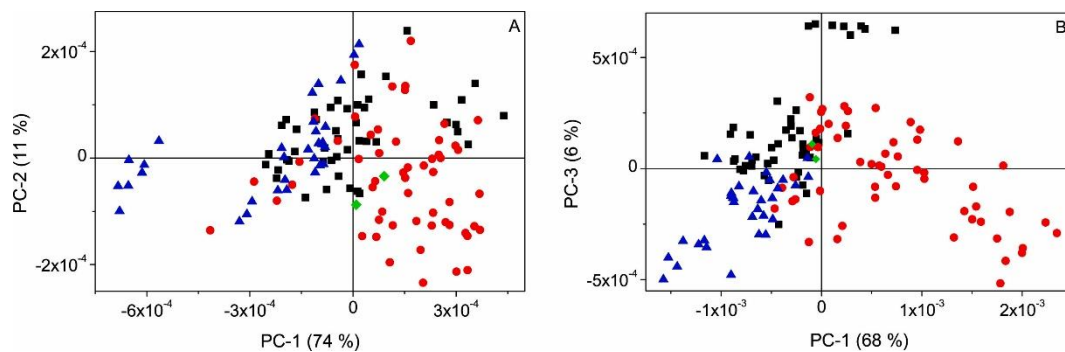


Figura 1.12. Distribuição dos *scores* da PCA para as amostras regulares de café regular, categorizadas pela variedade. (A) Instrumento 1, (B) Instrumento 2. ● Blends de arábica e robusta, ■ Blend de arábica, ▲ arábica puro, e ◆ não especificado.

A separação de classes entre os blends de arábica/robusta e os blends de arábica e arábica puro é mais evidente. O instrumento 2 apresenta uma separação mais aparente entre as classes quando comparado ao instrumento 1. A distinção entre arábica puro e blends de arábica não é possível usando as informações espectrais do instrumento 1. Usando o instrumento 2, pode-se visualizar uma tendência de separação entre essas duas classes com as amostras de arábicas puros assumindo predominantemente valores negativos de *scores* em PC-1 e PC-3. A análise da identificação da variedade é prejudicada pelo fato da proporção de cada variedade pura nos blends não ser conhecida, uma vez que esta informação não é disponibilizada pelo fornecedor. Portanto, alguns blends de arábica/robusta podem estar misturados com arábica puro e blends de arábica/arábica em função do teor de arábica.

1.4.1.6 Agrupamento de acordo com a origem do café

O revendedor comercial reporta produtos de café regulares de origem única e misturas de cafés de diferentes regiões, incluindo continentes e países. O total de classes é vinte e uma classe não é especificada.

A Figura 1.13A e B mostra a distribuição dos *scores* categorizados pela origem dos cafés para os instrumentos 1 e 2, respectivamente. O instrumento 1 permite identificar um grupo associado à origem etíope, detectado através da análise dos *scores* de PC1 e PC2. Os outros dois grupos mal podem ser distinguidos no quadrante inferior direito. Um grupo, que se refere

a uma amostra da América do Sul/Uganda, e o outro, cuja origem não é especificada pelo fabricante. Já o instrumento 2 permite identificar seis grupos associados à origem das amostras: Etiópia, Índia/México, Brasil/Colômbia/África, Índia, e América do Sul/Uganda, de acordo com a distribuição dos *scores* em PC1 e PC2. Também foi identificado mais um grupo, cuja origem não é informada pelo fornecedor, encontrado próximo ao grupo da América do Sul/Uganda (tipo *Kazaar*). O instrumento 2 apresenta melhor desempenho para distinguir a origem do café do que o instrumento 1.

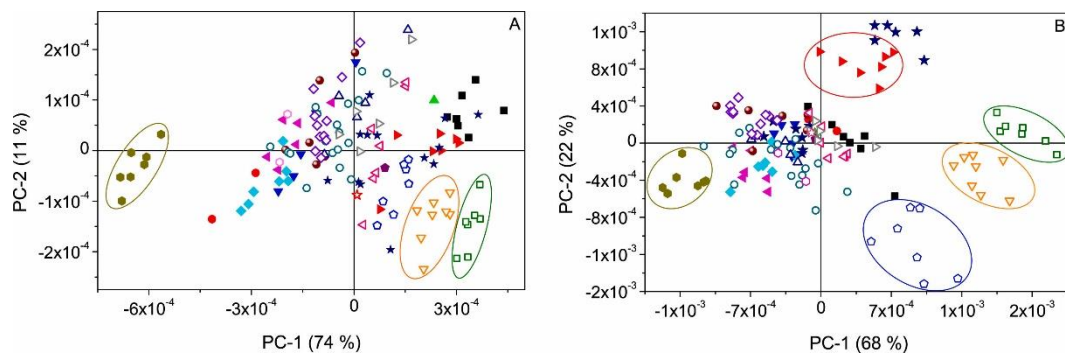


Figura 1.13. Distribuição dos *scores* da PCA categorizados de acordo com a origem do café. (A) Instrumento 1, (B) Instrumento 2. ■ América Latina (Costa Rica), ● Colômbia/Uganda, ▲ Brasil, ▼ América do Sul (Brasil), ◆ Colômbia, ◆ Costa Rica/Quênia, ► Índia/México, ● Etiópia, ★ Índia, ◆ Galápagos, ● Indonésia, □ Não especificado, ○ Brasil/Colômbia, △ Brasil/Colômbia/Guatemala, ▽ América do Sul/Uganda, ◇ Nicarágua, ◁ América do Sul/África Oriental/Sul da Índia, ▷ Brasil/México, ◊ Quênia/China/Indonésia, ☆ Sumatra, and ◊ Brasil/Colômbia/África.

Notavelmente, o café da Etiópia é sempre encontrado em grupos bem definidos em todas as análises de PCA realizadas neste estudo. A planta do café é originária da Etiópia. O país é reconhecido como aquele onde a bebida do café se originou. Portanto, os cultivares de arábica daquele país podem ter sofrido menores modificações do que em outros países, permanecendo mais próximos do original.

1.4.1.7 Agrupamento de acordo com as características sensoriais

Conforme mencionado, quatro das características sensoriais das amostras de café estão altamente correlacionadas. Portanto, embora a análise tenha sido realizada para todas as variáveis sensoriais, apenas são apresentados os resultados da PCA para intensidade e acidez, que apresentam menor correlação com as demais variáveis. As Figuras. S1.1A–C e S1.2A–C, encontradas no Apêndice A, ilustram os *scores* categorizados pelas demais variáveis sensoriais (torra, corpo e amargor).

A Figura 1.14A e B mostra os *scores* categorizados pelos valores de intensidade de PC1 e PC2 obtidos pelos instrumentos 1 e 2, respectivamente. Há uma clara tendência de aumento da intensidade da esquerda para a direita ao longo do PC1 para ambos os instrumentos. O instrumento 1 não distingue claramente os grupos de amostras, exceto as amostras da Etiópia (intensidade 4) agrupadas e bem separadas, apresentando valores de *scores* mais baixos para PC1. Grupos não bem definidos também são detectados para amostras que apresentam valores de intensidade iguais a 11, 12 e 13.

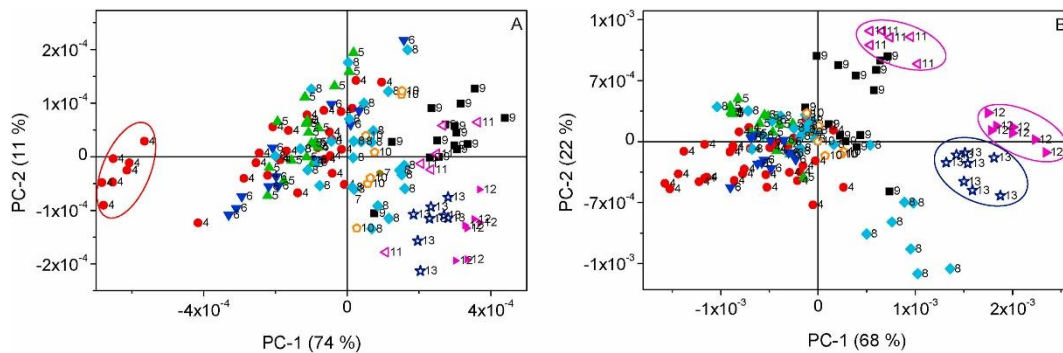


Figura 1.14. Distribuição dos *scores* da PCA categorizadas pela intensidade das amostras regulares de café. (A) instrumento 1, (B) instrumento 2. Os números próximos aos símbolos referem-se aos valores de intensidade das amostras.

O instrumento 2 mostra uma definição superior de grupos amostrais associados a características de intensidade com valores de 11, 12 e 13. Amostras com valores de intensidade 4, 8 e 9 são divididas em dois ou três grupos (intensidade 4), refletindo outros fatores que afetam os perfis espectrais. Este assunto é discutido abaixo.

A Figura 1.15A e B representa os valores de *scores* para PC1 e PC2 categorizados pelo valor de acidez.

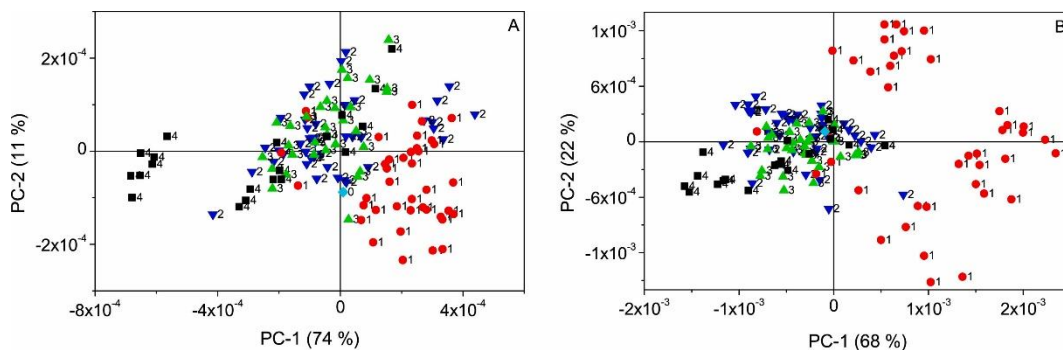


Figura 1.15. Distribuição dos *scores* de PCA categorizados pela acidez das amostras regulares de café. (A) Instrumento 1, (B) Instrumento 2. Os números próximos aos símbolos referem-se aos valores de acidez.

Como esperado, com base na análise prévia das variáveis sensoriais (Figura 1.1 e Tabela 1.2), a acidez varia no sentido oposto da intensidade. Para o instrumento 1, um grupo de amostras com acidez igual a 4, correspondente às amostras da Etiópia, encontra-se separado das demais amostras. Os *scores* obtidos através dos dados do instrumento 2 mostram vários grupos com o mesmo valor de acidez, como exemplificam as amostras com acidez = 1.

1.4.1.8 Avaliação conjunta da viabilidade de agrupamento

Uma análise global dos resultados da PCA aplicada aos conjuntos de dados produzidos pelos dois instrumentos, considerando todas as características dos cafés simultaneamente, aponta para as dificuldades de decisão sobre os fatores responsáveis pelo agrupamento das amostras de café. Por exemplo, os grupos reconhecidos com base nos tipos de café *India*, *Kazaar* e *Napoli*, utilizando o conjunto de dados do instrumento 2, são facilmente confundidos com os valores de intensidade de 11, 12 e 13, respectivamente, apresentados por estas amostras. Por outro lado, amostras com os mesmos valores de intensidade podem ser encontradas divididas em grupos distintos. Um exemplo claro são as amostras com intensidade igual a 4 ou 9.

Geralmente, a análise de conjuntos de dados espectrais NIR gerados por amostras de café processado (torrado e moído) sempre transmitirá algumas incertezas sobre o agrupamento (e possível uso desses resultados para posterior classificação do café), porque, como mostram os resultados deste trabalho, às vezes a característica espectral da amostra pode ser definida por fatores como origem e cultivar do café (ou blends de cultivares), enquanto em outras situações o processamento da amostra, como o grau de torra, pode dominar a assinatura espectral e, conseqüentemente, prevalecer sobre outros fatores.

É razoável considerar que qualquer conclusão sobre agrupamento ou classificação de amostras de café processado deve ser cuidadosamente tomada e validada considerando todas as possíveis fontes de variabilidade espectral, incluindo os efeitos do processamento do produto, como o grau de torra.

1.4.1.9 Análise detalhada de agrupamento

Para aumentar a capacidade dos instrumentos compactos de produzir informações úteis para a classificação de amostras de café, os conjuntos de dados foram explorados mais detalhadamente, limitando a variabilidade de variáveis sensoriais influentes, como a intensidade. Assim, apenas as 33 amostras de café regular apresentando intensidade igual a 4 foram analisadas separadamente. Como mostrado anteriormente, a intensidade está

significativamente correlacionada com as outras três variáveis sensoriais, cujos valores resultaram parcialmente do processamento do café e devem refletir alterações químicas nas amostras. Particularmente, a intensidade está correlacionada com o grau de torra, fator controlado pelo processo que confere a maior parte da alteração na composição química do café.

A Figura 1.16A, B e C mostra a distribuição dos *scores* do conjunto de dados obtidos pelo instrumento 2. Resultados equivalentes obtidos com o instrumento 1 podem ser observados na Figuras S1.3A–C no material suplementar (Apêndice A).

Quando a PCA é restrita a amostras com intensidade 4, aquelas da mesma variedade ficam agrupadas de forma mais evidente. Essas amostras definem três grupos de acordo com o tipo de café. Um grupo muito bem separado contendo as amostras da Etiópia, outro contendo o *Buenos Aires Lungo* e um terceiro grupo no qual os tipos *Volluto*, *Vivalto Lungo*, *Cosi* e *Linizio Lungo* se encontram misturados. É interessante notar que nesta análise detalhada, os dados espectrais NIR produzidos pelo instrumento 2 podem distinguir pequenas diferenças entre os quatro tipos de café presentes no terceiro grupo, evidenciando algumas diferenças (principalmente sensoriais) entre eles.

Considerando os resultados obtidos pela análise qualitativa dos conjuntos de dados produzidos por ambos os instrumentos, é possível afirmar que o instrumento 2 é mais eficaz em fornecer informações espectrais úteis para a caracterização de amostras de café torrado e moído. A faixa mais abrangente de comprimento de onda monitorada por este instrumento, que inclui a região de bandas de combinação, é o diferencial que fornece informações adicionais sobre as amostras de café torrado e moído.

No entanto, os resultados da análise de dados espectrais provenientes de amostras de café processado precisam ser avaliados com cuidado, pois os efeitos das fontes nativas de variabilidade, como variedade e origem, podem ser superados por mudanças na composição química das amostras, transmitidas pelo processamento industrial.

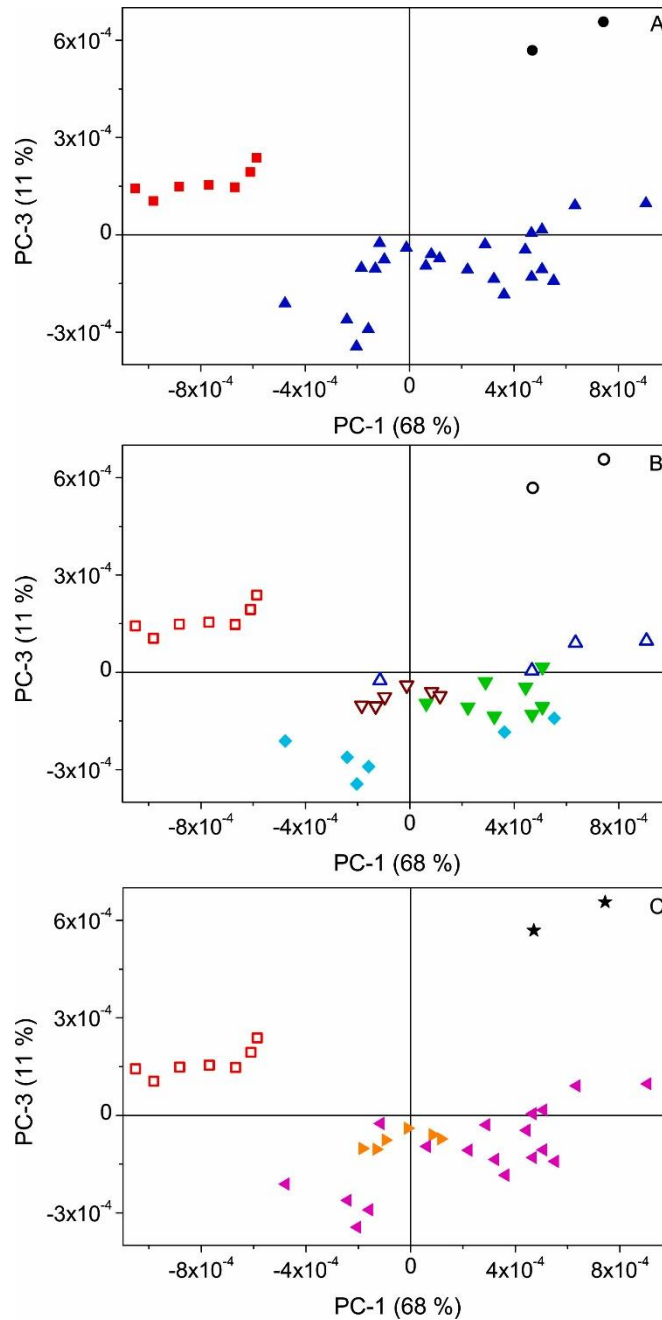


Figura 1.16. Distribuição dos *scores* de PCA dos cafés regulares com intensidade igual a 4, utilizando o conjunto de dados obtido pelo instrumento 2. (A) de acordo com a variedade, ■ Arábica puro, ▲ Blends de Arábica, ● Blend Arábica e Robusta; (B) tipo, ○ *Buenos Aires Lungo*, □ *Ethiopia*, ▽ *Cusi*, △ *Linizio Lungo*, ◆ *Vivalto Lungo*, ▼ *Volluto*; e (C) origem, ★ Colômbia/Uganda, ► Costa Rica/Quênia, ▲ Brasil/Colômbia.

De modo geral, a viabilidade de um instrumento compacto para acessar informações espectrais relevantes visando a caracterização do café é atestada pelos resultados deste trabalho. Além disso, esta informação demonstra que os diferentes tipos comerciais podem ser avaliados utilizando instrumentação NIR compacta. Grandes e pequenas diferenças podem ser observadas entre diferentes tipos de café e, às vezes, é possível caracterizar a origem dessas diferenças.

Mesmo quando a distinção entre os cafés não é prática, a consistência espectral entre vários lotes de um mesmo tipo de café também pode ser verificada para atestar boas práticas de fabricação e outras características intrínsecas do café, como origem e cultivar.

1.4.2 Resultados quantitativos multivariados para as características sensoriais

Os conjuntos de dados espectrais obtidos pelos dois instrumentos foram empregados para avaliar a viabilidade do uso de espectrofotômetros compactos para quantificar as características sensoriais de amostras de café torrado e moído quando um conjunto de dados abrangente é empregado.

Os valores das características sensoriais fornecidos pelo fabricante das cápsulas de café foram utilizados para a avaliação quantitativa do desempenho dos instrumentos. O algoritmo de mínimos quadrados parciais (PLS) foi empregado para construir modelos de regressão visando prever a intensidade, corpo, amargor, torra e acidez utilizando os conjuntos de espectros da segunda derivada.

O conjunto amostral foi dividido em dois subconjuntos utilizando o algoritmo Kennard-Stone aplicado nos *scores* das PCAs realizadas em ambos os conjuntos de dados. Noventa amostras foram empregadas para calibrar os modelos PLS e as 55 restantes foram empregadas para a validação do modelo. Todas as variáveis foram inicialmente empregadas para construir os modelos PLS.

A Tabela 1.3 resume os resultados de calibração e validação obtidos para as cinco características sensoriais e os dois instrumentos avaliados, utilizando todo o conjunto de comprimentos de onda disponíveis.

Não foram observadas diferenças significativas na predição dos traços sensoriais entre os modelos de regressão construídos a partir do conjunto de dados obtidos pelos dois instrumentos.

A Tabela 1.4 mostra os resultados dos modelos PLS construídos utilizando os conjuntos de variáveis selecionados pelo algoritmo Jack-Knife, baseado na estabilidade dos coeficientes de regressão dos modelos construídos utilizando os conjuntos inteiros de variáveis. A seleção é feita analisando os 20 conjuntos de coeficientes de regressão produzidos após o processamento da validação cruzada do modelo preliminar. Os resultados da validação desses modelos usando um conjunto externo de amostras de teste são mostrados nas Figuras 1.17A–E e 1.18A–E para os instrumentos 1 e 2, respectivamente.

Tabela 1.3. Resumo dos resultados obtidos pelos modelos PLS para a determinação das características sensoriais do café torrado e moído, empregando todo o conjunto de variáveis espectrais sondadas pelos dois instrumentos

Instrumento	Característica sensorial	Calibração				Validação		
		NC	LV	R ²	RMSECV	NV	R ²	RMSEP
1	Intensidade	90	6	0.743	1.5	55	0.777	1.3
	Amargor		3	0.658	0.7		0.747	0.7
	Corpo		7	0.756	0.5		0.731	0.7
	Torra		7	0.775	0.5		0.721	0.6
	Acidez		2	0.428	0.7		0.452	0.8
2	Intensidade	90	6	0.786	1.4	55	0.846	1.1
	Amargor		5	0.626	0.7		0.669	0.7
	Corpo		5	0.673	0.6		0.801	0.5
	Torra		5	0.739	0.6		0.848	0.4
	Acidez		5	0.587	0.7		0.586	0.7

NC: número de amostras de calibração, LV: variáveis latentes empregadas pelo modelo PLS, RMSECV: erro quadrático médio de validação cruzada, NV: número de amostras de validação, RMSEP: erro quadrático médio de predição.

Tabela 1.4. Resumo dos resultados obtidos pelos modelos globais PLS para a determinação de características sensoriais de café torrado e moído, empregando o conjunto de variáveis espectrais selecionadas pelo algoritmo Jack-Knife

Instrumento	Característica sensorial	Calibração				Validação		
		NC	LV	R ²	RMSECV	NV	R ²	RMSEP
1	Intensidade	90	5	0.798	1.3	55	0.782	1.3
	Amargor		1	0.650	0.7		0.646	0.7
	Corpo		4	0.767	0.5		0.714	0.6
	Torra		5	0.778	0.5		0.791	0.5
	Acidez		1	0.424	0.8		0.367	0.8
2	Intensidade	90	3	0.787	1.4	55	0.820	1.2
	Amargor		4	0.600	0.7		0.664	0.7
	Corpo		6	0.688	0.6		0.733	0.5
	Torra		5	0.747	0.6		0.803	0.5
	Acidez		4	0.620	0.7		0.566	0.7

NC: número de amostras de calibração, LV: variáveis latentes empregadas pelo modelo PLS, RMSECV: erro quadrático médio de validação cruzada, NV: número de amostras de validação, RMSEP: erro quadrático médio de predição.

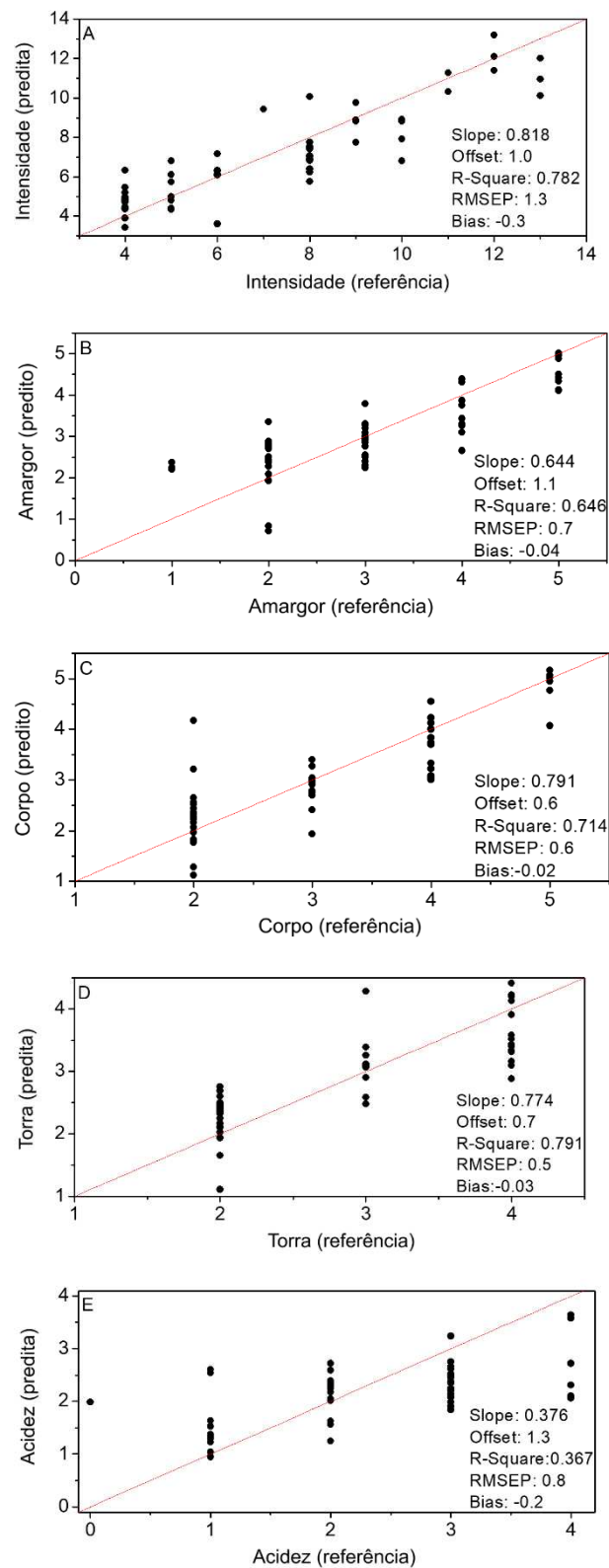


Figura 1.17. Comparação dos resultados de validação previstos pelos modelos PLS com os valores de referência para características do café, empregando o conjunto de dados espectrais obtidos pelo instrumento 1 e as variáveis selecionadas pelo algoritmo Jack-knife. (A) intensidade, (B) amargor, (C) corpo, (D) torra e (E) acidez.

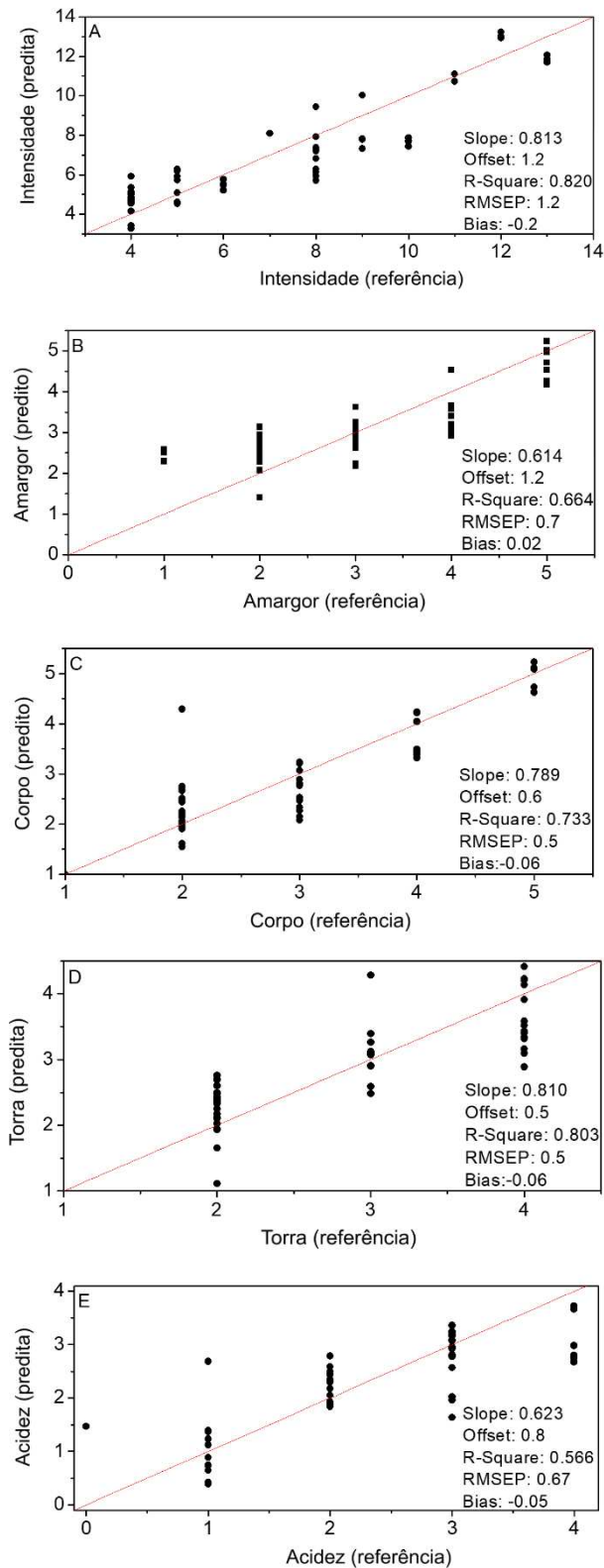


Figure 1.18. Comparação dos resultados de validação previstos pelos modelos PLS com os valores de referência para características do café, empregando o conjunto de dados espectrais obtidos pelo instrumento 2 e as variáveis selecionadas pelo algoritmo Jack-knife. (A) intensidade, (B) amargor, (C) corpo, (D) torra e (E) acidez.

Em geral, utilizando o algoritmo Jack-Knife, observa-se uma redução no número de variáveis latentes necessárias para otimizar os modelos. No entanto, apenas pequenas melhorias foram observadas na precisão do modelo. Com base no princípio da parcimônia, os modelos construídos com os conjuntos de variáveis selecionados devem ser recomendados para avaliar as características sensoriais de amostras de café torrado e moído [28].

Considerando que os valores das características sensoriais são sempre expressos como valores inteiros pelo fabricante, o fato de os valores se referirem aos lotes de café, e não à amostra específica medida pelo NIRS, a incerteza usual apresentada pelas variáveis sensoriais, e o conjunto abrangente de dados empregado neste estudo para construir modelos de regressão global, o RMSEP para intensidade, corpo e torra pode ser considerado satisfatório para determinar esses parâmetros, utilizando os dados espectrais NIR obtidos pelos dois instrumentos compactos avaliados neste estudo. O desempenho quantitativo global dos dois instrumentos pode ser considerado equivalente. Os outros dois parâmetros (amargor e acidez) apresentaram erros mais significativos e só poderiam ser utilizados para triagem dessas características se modelos globais fossem empregados.

Os resultados quantitativos podem ser melhorados através da construção de modelos de regressão locais. Por exemplo, a Tabela 1.5 mostra os resultados da validação cruzada e da validação externa obtidos usando apenas as amostras com intensidade na faixa de 9 a 13 para ambos os instrumentos. O subconjunto contém 46 amostras de 9 tipos de cafés diferentes. Trinta amostras, selecionadas pelo algoritmo Kennard-Stone, foram empregadas para construção dos modelos e o restante utilizado para validação do modelo. O algoritmo Jack-Knife foi empregado para selecionar as variáveis utilizadas pelos modelos PLS. Embora o subconjunto contenha amostras com todos os valores de intensidade nessa faixa, as demais propriedades estão restritas a apenas dois valores sucessivos. No entanto, uma melhoria significativa nos resultados é obtida utilizando modelos locais. O desempenho geral do instrumento 2 é ligeiramente melhor que o do instrumento 1.

Tabela 1.5. Resumo dos resultados obtidos pelos modelos locais PLS para a determinação de características sensoriais de café torrado e moído, empregando o conjunto de variáveis espectrais selecionadas pelo algoritmo Jack-Knife e o subconjunto amostral com valores de intensidade na faixa 9–13

Instrumento	Característica sensorial	Calibração				Validação		
		NC	LV	R ²	RMSECV	NV	R ²	RMSEP
1	Intensidade	30	5	0.772	0.7	16	0.784	0.7
	Amargor		2	0.747	0.3		0.795	0.2
	Corpo		1	0.810	0.2		0.886	0.2
	Torra		3	0.767	0.3		0.799	0.2
	Acidez		2	0.508	0.6		0.656	0.5
2	Intensidade	30	6	0.845	0.5	17	0.915	0.4
	Amargor		3	0.861	0.2		0.835	0.2
	Corpo		4	0.908	0.2		0.967	0.1
	Torra		5	0.854	0.2		0.842	0.2
	Acidez		7	0.787	0.4		0.830	0.3

NC: número de amostras de calibração, LV: variáveis latentes empregadas pelo modelo PLS, RMSECV: erro quadrático médio de validação cruzada, NV: número de amostras de validação, RMSEP: erro quadrático médio de predição.

1.5 CONCLUSÕES

Este trabalho avaliou dois instrumentos NIR compactos e de baixo custo e análise multivariada para determinar diversas características de amostras de café torrado e moído de um conjunto de amostras abrangente. Os desempenhos dos instrumentos nas avaliações qualitativa e quantitativa das características do café são semelhantes, sendo os resultados obtidos pelo instrumento 2 (NeoSpectra) melhores, principalmente, considerando os resultados quantitativos.

Os resultados da PCA mostraram que ambos os instrumentos podem identificar variabilidades entre diversos lotes de café do mesmo tipo fabricados pela mesma empresa, uma vez que a variabilidade espectral observada entre os lotes (indicada pela dispersão dos *scores* da PCA) excede a repetibilidade e reprodutibilidade espectral observada para ambos os instrumentos. A menor dispersão dos *scores* da PCA nas componentes principais relevantes pode ser usada para atestar boas práticas de fabricação.

Ambos os instrumentos podem distinguir cafés de uma única variedade de cafés de blends de variedades e cafés regulares de descafeinados, mesmo no complexo conjunto de amostras empregado neste estudo. A origem das amostras de café é mais difícil de atestar, porque o processamento industrial do café pode contrabalancear a impressão digital química e espectral original dos grãos verdes. Contudo, a origem etíope pode ser identificada de forma inequívoca entre outras amostras, mesmo considerando o efeito do processamento do café. É claro que a identificação da origem do café é melhorada quando são consideradas apenas as amostras processadas aproximadamente da mesma forma, como mostram os cafés regulares apresentando intensidade igual a 4.

O efeito do processamento industrial pode ser observado, por exemplo, pelo grau de torra, que transmite alterações severas na composição química original dos grãos de café verde. Essas alterações causam variações nos espectros NIR das amostras registradas por ambos os instrumentos e contribuem para confundir outras características. Isto é demonstrado, por exemplo, pela dificuldade de agrupar as amostras de diferentes origens e variedades (arábica e mistura de arábica) quando as intensidades das amostras são significativamente diferentes. No entanto, estender a análise qualitativa a amostras de café que apresentem a mesma intensidade melhora as informações sobre a origem e a variedade. Os resultados quantitativos também podem ser melhorados através da utilização de modelos de regressão locais em vez de modelos de regressão globais.

Este trabalho evidenciou a dificuldade de estender as conclusões obtidas por um conjunto de dados espectrais NIR constituído por amostras com um número restrito de características de café torrado e moído, como observado em diversos trabalhos previamente desenvolvidos, para um conjunto de dados mais abrangente.

Capítulo 2: Emprego da espectroscopia no infravermelho próximo para a identificação e detecção de adulterações de amostras de cocaína apreendidas pela Polícia Federal

RESUMO

A caracterização de substâncias suspeitas no momento da apreensão por agentes policiais é de grande importância para evitar falhas e permitir a aplicação da lei de forma correta e eficaz, fornecendo provas confiáveis para futuras decisões judiciais. Atualmente, os testes rápidos realizados para identificação *in situ* da cocaína baseiam-se em reações colorimétricas, que, geralmente, não são muito confiáveis, pois podem produzir resultados falso-positivos. Assim, este trabalho teve como objetivo desenvolver um método *in situ*, rápido e não destrutivo para caracterizar quimicamente amostras de cocaína, utilizando espectrômetros NIR portáteis de baixo custo e análise multivariada de dados. Para isso, foram coletados espectros NIR de 1034 amostras de cocaína apreendidas pela Polícia Federal em diferentes localidades do Brasil durante o período de 2020 a 2022. Essas amostras, previamente caracterizadas no Instituto Nacional de Criminalística (DF, Brasil), foram analisadas por dois espectrômetros portáteis: o NanoNIR (Texas Instruments Inc. Austin, EUA), que monitora a faixa espectral de 900 – 1700 nm; e o NeoSpectra micro-FT-NIR (Si-ware, Egito), operando entre 1350 e 2550 nm. Posteriormente, o conjunto de dados foi pré-processado e tratado com o software Unscrambler 11.0 (Aspen Technologies, Noruega). Os dados foram transformados usando a 2ª derivada (Savitzky-Golay, 15 pontos, polinômio de 2º grau). Em seguida, foi realizada a análise de componentes principais (PCA) neste conjunto de dados. Os espectros obtidos pelos instrumentos NanoNIR e NeoSpectra permitem diferenciar entre cocaína base e cloridrato de cocaína nas duas primeiras PCs. Uma vez que a forma química da cocaína tem grande peso na distribuição dos *scores* dessas amostras, as análises seguintes foram realizadas utilizando, separadamente, os subgrupos cloridrato de cocaína e cocaína base. A PCA também foi utilizada para diferenciar amostras puras de adulteradas e para identificar qualitativamente os principais adulterantes de cocaína. Além disso, foram construídos modelos de classificação SIMCA e de regressão PLS, a fim de identificar e quantificar, respectivamente, a cocaína.

Palavras-chave: Cloridrato de cocaína, Cocaína base livre, Amostras apreendidas, Espectroscopia NIR, Análise multivariada, Análise *in situ*.

2.1 INTRODUÇÃO

As amostras forenses são aquelas que possuem alguma relevância/interesse na área forense, de inteligência e criminalística. Podem ser de diversas naturezas, sendo que, para a toxicologia, as drogas de abuso são as mais abundantes e estudadas, em especial, a cocaína. Essa substância é estimulante do sistema nervoso central (SNC), possui propriedades anestésicas, e pode provocar sensação de poder e euforia, de modo que é muito utilizada recreativamente [30].

A cocaína, denominada quimicamente como metilbenzoilecgonina, é um alcalóide extraído das folhas de plantas do gênero *Erytroxylum*, que se encontram distribuídas geograficamente em regiões tropicais de altitudes elevadas, principalmente, da América do Sul. Sua forma natural é a base livre, entretanto, em meio ácido, o agrupamento amino da molécula é protonado e a cocaína é convertida em um sal de amônio quaternário, conhecido como cloridrato de cocaína (Figura 2.1) [31].

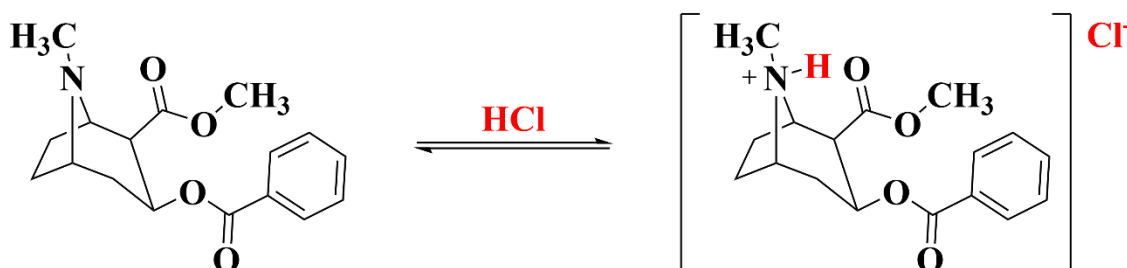


Figura 2.1. Estruturas da cocaína base (à esquerda) e do cloridrato de cocaína (à direita).

Os maiores produtores mundiais de cocaína são a Colômbia, o Peru e a Bolívia. E o Brasil, apesar de não ser um país produtor da droga, destaca-se no cenário mundial como um dos principais países de trânsito da cocaína. Principalmente, por causa da sua posição geográfica estratégica, que liga os países produtores aos países de destino da droga, como no caso da Europa e da África. Esse fato contribui para que o Brasil ocupe as primeiras posições em termos de volume de apreensão de produtos baseados em cocaína [32].

A principal técnica analítica utilizada para identificar e quantificar cocaína em amostras suspeitas é a cromatografia gasosa com detector de ionização em chama (GC-FID) [30]. Entretanto, essa técnica é laboriosa, demorada e requer um laboratório e instrumentação de bancada para ser executada. Dessa forma, atualmente, são utilizados em campo testes rápidos de triagem com o objetivo principal de identificar, no material suspeito, a presença ou ausência de cocaína.

Os principais testes rápidos utilizados são os testes colorimétricos (que se baseiam na mudança de cor da amostra) [33]. Entretanto, esses testes, apesar de sensíveis, geralmente, não são muito seletivos e podem gerar falsos positivos. A fim de tentar contornar esse problema, existe a necessidade da criação de novos métodos de triagem de substâncias suspeitas, que sejam simples, eficientes, portáteis e de baixo custo. Nesse sentido, o uso de técnicas espectroscópicas de análise direta de amostras se mostra como uma alternativa.

A espectroscopia no infravermelho próximo (NIRS, do inglês, *near infrared spectroscopy*) é um tipo de espectroscopia vibracional capaz de obter informações analíticas de amostras moleculares complexas. Essas informações advêm da interação da radiação infravermelha com a matéria e essa técnica se caracteriza por ser precisa, rápida e não-destrutiva [33]. Por isso, a técnica NIR tem sido muito empregada para diversos tipos de amostras e problemas analíticos [1].

Com os avanços na área de instrumentação analítica, nas últimas duas décadas, têm sido desenvolvidos e comercializados espectrômetros NIR portáteis, miniaturizados e de baixo custo, com diferentes configurações instrumentais [1]. Esses instrumentos permitem a análise de amostras em campo e o desenvolvimento de novas metodologias analíticas mais acessíveis e interessantes.

Dessa forma, este trabalho propõe investigar o potencial e o desempenho de dois espectrômetros NIR de baixo custo na análise de amostras de interesse forense, exemplificadas aqui pela cocaína.

2.2 OBJETIVO

Desenvolver um novo método de triagem, que seja rápido, não-destrutivo e portátil para caracterizar quimicamente cocaína em amostras suspeitas, no momento de apreensão, utilizando dois espectrômetros NIR miniaturizados e de baixo custo e análise multivariada de dados.

2.3 MATERIAIS E MÉTODO

2.3.1 Amostras

Para a realização deste trabalho, foram analisadas 1034 amostras de cocaína apreendidas pela Polícia Federal do Brasil nos anos de 2020 a 2022, sendo que, desse total, 275 amostras eram constituídas de cocaína base, 735 de cloridrato de cocaína e 24 foram denominadas “N.D.”, indicando que não puderam ser classificadas nem como cocaína base, nem como cloridrato de cocaína.

Essas amostras foram previamente analisadas por espectroscopia de infravermelho por transformada de Fourier com refletância total atenuada (FTIR/ATR) e por cromatografia gasosa acoplada a detector por ionização de chama (GC-FID) no Instituto Nacional de Criminalística da Polícia Federal (INC/DPF, Brasília – DF), a fim de se determinar a presença e forma de apresentação da cocaína, a natureza química e o teor de seus constituintes respectivamente, e assim, obter o perfil químico dessas amostras.

Em linhas gerais, o perfil químico dessas amostras mostrou que elas possuem um teor médio de 85,5 % de cocaína, sendo que a amostra mais concentrada possui um teor de 98,4 % e a menos concentrada, de 6,9 %. A mediana, que representa o valor central de um conjunto de dados numéricos ordenados [34] equivale a 91,8 % de cocaína para esse conjunto de amostras e a moda, que corresponde ao valor mais frequente no conjunto de dados, equivale a 94 %.

Além da cocaína, também foram identificados e quantificados por CG-FID os principais adulterantes encontrados em amostras de cocaína, a saber: aminopirina, benzocaína, cafeína, diltiazem, fenacetina, hidroxizina, levamisol, lidocaína e procaína. Dessa forma, constatou-se, no conjunto das 1034 amostras, que 903 amostras não estavam adulteradas e as outras 131 amostras continham pelo menos um desses adulterantes em sua composição.

Os espectros NIR dos padrões analíticos (Sigma-Aldrich, EUA) dessas substâncias utilizadas para adulteração foram coletados. Também foram adicionados ao conjunto de dados, os espectros NIR dos padrões analíticos das substâncias que também podem ser encontradas como adulterantes de cocaína (clobenzorex, paracetamol e tetracaína) e os espectros NIR de substâncias que podem ser utilizadas como diluentes de amostras de cocaína, tais como: ácido bórico, amido e irganox.

Para a etapa de construção dos modelos de classificação e de regressão e para a etapa de validação, o conjunto total de amostras foi dividido da seguinte forma: as amostras apreendidas nos anos de 2020 e 2021 (que representam 907 amostras do conjunto total, sendo 218 amostras de cocaína base, 673 de cloridrato e 16 “N.D.”) foram utilizadas para construir os

modelos; enquanto as amostras apreendidas no ano de 2022 (equivalente a 127 amostras do total, das quais 57 amostras eram cocaína base, 62 eram cloridrato de cocaína e 8 eram “N.D.”) foram utilizadas para validar os modelos.

Uma amostra de cocaína, utilizada como controle nas análises cromatográficas foi utilizada como controle nas medidas espectroscópicas a fim de atestar o desempenho dos espectrofotômetros portáteis NIR e o protocolo de medidas. Assim, espectros NIR dessa amostra foram coletados no início e no final de cada dia de experimentos durante o período de aquisição dos dados utilizados para a construção dos modelos.

2.3.2 Espectrômetros infravermelho próximo

Os espectros de absorbância no infravermelho próximo (NIR) do conjunto de dados deste trabalho foram coletados utilizando-se dois espectrômetros portáteis e de baixo custo: o NanoNIR (*Texas Instruments Inc.* Austin, Texas, EUA) [3] e o NeoSpectra micro-FT-NIR (*SiWare*, Egito) [4].

Como detalhado em trabalhos anteriores [1], [26],[35], [36], o NanoNIR opera na faixa espectral de 900 a 1700 nm e possui uma resolução nominal de 10 nm, enquanto o NeoSpectra varre a região entre 1350 e 2550 nm, com uma resolução de 16 nm.

Para a aquisição dos dados espectrais, empregando o NanoNIR, foi utilizado o modo de varredura simples, com 228 pontos amostrados ao longo da faixa espectral, e o espectro resultante foi uma média de 25 varreduras. No caso dos espectros obtidos com o auxílio do NeoSpectra, o número de pontos amostrados ao longo da faixa espectral foi de 257 pontos e o tempo de exposição dos espectros de cada amostra à radiação foi de 2 segundos.

Os desempenhos dos dois espectrômetros foram verificados ao longo do período de aquisição dos dados espectrais deste trabalho através do monitoramento do ruído, da variação da linha base e da dispersão dos espectros da amostra controle.

2.3.3 Frasco de medida e protocolo de aquisição dos espectros

Para a obtenção dos espectros NIR, cada amostra de cocaína foi transferida para um frasco de vidro de borossilicato tipo 1, modelo “vial shell clear”, 15 x 45 mm (4 mL), marca *Flowsupply* [37]. Tomou-se o cuidado para que as amostras ocupassem completamente o fundo do frasco, preenchendo-o até uma altura superior a 5 mm (altura mínima necessária para evitar que parte do feixe de radiação passasse através da amostra e escapasse).

Como medida de fundo (*background*), utilizou-se também um frasco de vidro (idêntico ao usado para armazenar as amostras) preenchido com Spectralon® (fluoropolímero cuja

reflectância difusa nominal na região do infravermelho próximo é de 100%) [38].

Antes da aquisição dos espectros, os espectrômetros foram conectados a notebooks por cabos USB e mantidos ligados por 20 min para estabilização. Após transcorrido esse tempo, os frascos de vidro foram posicionados sobre as janelas ópticas dos instrumentos, e os espectros foram obtidos através de medidas de reflectância difusa – Figura 2.2.

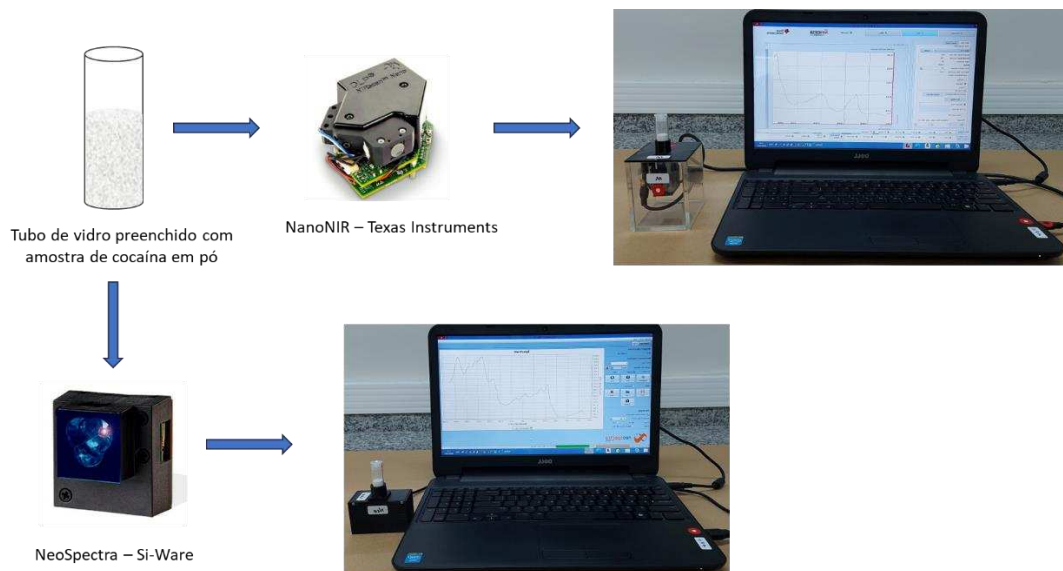


Figura 2.2. Esquema representando a montagem experimental da obtenção dos espectros NIR das amostras de cocaína.

O protocolo de aquisição dos dados se deu na seguinte ordem: primeiramente, foi realizada a medida do sinal de referência (100 % de reflectância); em seguida, coletou-se o espectro da amostra controle e, por fim, seguiu-se adquirindo-se os espectros das amostras de cocaína que compuseram o conjunto de dados. Novos sinais de referência e do controle foram obtidos no início e no final de cada dia de experimento e após 50 amostras analisadas consecutivamente. A variabilidade da linha base (absorbância zero) e a dispersão dos espectros da amostra controle foi estimada ao longo de todo o período de aquisição dos dados espectrais.

2.3.4 Análise do conjunto de dados

O conjunto de dados foi pré-processado, tratado e analisado utilizando-se o software quimiométrico *The Unscrambler* 11.0 (CAMO, Noruega).

Para os dados obtidos com o NanoNIR, as extremidades dos espectros foram removidas e a faixa útil empregada nas análises foi de 970 a 1670 nm. Com o NeoSpectra, utilizou-se toda sua faixa de trabalho (1350 a 2550 nm). Como pré-tratamento, em ambos os casos, foi aplicada a segunda derivada (algoritmo de Savitsky-Golay, janela de 15 pontos, polinômio de 2º grau)

sobre os dados espectrais.

Posteriormente, realizaram-se análises de componentes principais (PCAs) desses conjuntos de dados e construíram-se modelos de classificação SIMCA (*soft independent modeling by class analogy*) globais e para as classes “cloridrato de cocaína” e “cocaína base”. Além disso, também foram construídos modelos de regressão PLS (*partial least squares*) globais e para as classes “cloridrato de cocaína” e “cocaína base”, com o objetivo de quantificar o teor de cocaína, utilizando os espectros derivados e a informação de referência do teor de cocaína nas amostras (obtida por análise cromatográfica).

As figuras de mérito dos modelos de classificação SIMCA foram expressas como taxa de falsos positivos (FPR), taxa de falsos negativos (FNR), seletividade (STR), especificidade (SPR) e eficiência (EFR), calculadas usando, respectivamente, as equações de (1) a (5) que se seguem:

$$FPR = \frac{FP}{(FP+TN)} \times 100 \quad (1)$$

$$FNR = \frac{FN}{(FN+TP)} \times 100 \quad (2)$$

$$STR = \frac{TP}{(TP+FN)} \times 100 \quad (3)$$

$$SPR = \frac{TN}{(TN+FP)} \times 100 \quad (4)$$

$$EFR = 100 - (FPR + FNR) \quad (5)$$

Onde FP representa o número de falsos positivos; FN, o número de falsos negativos; TP equivale ao número de verdadeiros positivos; e TN, ao número de verdadeiros negativos.

Já as figuras de mérito dos modelos de regressão PLS foram expressas como RMSECV (raiz quadrada do erro quadrático médio de validação cruzada) para as amostras de calibração (no estágio de calibração) e RMSEP (raiz quadrada do erro quadrático médio de previsão) para as amostras de validação (no estágio de predição).

2.4 RESULTADOS E DISCUSSÃO

2.4.1 Desempenho dos espectrômetros

Durante o desenvolvimento deste trabalho, os desempenhos dos espectrofotômetros empregados na coleta dos dados foram avaliados. A Figura 2.3 mostra os desempenhos gerais do NanoNIR e do NeoSpectra.

Na Figura 2.3A e 2.3D, estão representados os desvios padrão das medidas de sinal de fundo (sinal do *Spectralon*®), expressos em absorbância por comprimento de onda, obtidas utilizando-se, respectivamente, o NanoNIR e o NeoSpectra. Para se calcular esses desvios, foram utilizadas 42 medidas do sinal de fundo, coletadas ao longo do período de aquisição dos dados utilizados neste trabalho.

Também estão representados, na Figura 2.3B e 2.3E, os desvios padrão das medidas da amostra controle, expressos em absorbância por comprimento de onda, para o NanoNIR e o NeoSpectra respectivamente. Foram utilizadas, neste caso, 32 medidas da cocaína controle, coletadas ao longo do período de aquisição dos dados.

Por fim, na Figura 2.3C e 2.3F, estão representados os gráficos da relação sinal-ruído (SNR) da amostra controle em função do comprimento de onda, para os espectrômetros NanoNIR e NeoSpectra respectivamente. A relação sinal-ruído foi calculada, dividindo-se a média do sinal (absorbância) em cada comprimento de onda por seu respectivo desvio padrão.

Os desempenhos do NanoNIR e do NeoSpectra se mostraram adequados, com desvios padrão médios da linha base de 0,003 e 0,0006, respectivamente. Entretanto, no caso do NanoNIR, como pode ser observado nos gráficos (A), (B) e (C) da Figura 2.3, seu desempenho piora nas extremidades, delimitadas pelas linhas vermelhas. Desse modo, a faixa de trabalho escolhida foi de 960 a 1660 nm. Já para o NeoSpectra, apesar do pior desempenho no início de sua faixa espectral (entre 1350 e 1550 nm, aproximadamente), os valores de desvio se mantiveram-se todos abaixo de 0,002, no caso da variação do sinal de fundo, e abaixo de 0,01 para as medidas da amostra controle (resultados melhores do que os obtidos utilizando-se o NanoNIR), como pode ser observado nos gráficos (D), (E) e (F) da Figura 2.3. Portanto, toda a faixa espectral do NeoSpectra (1350 a 2550 nm) foi utilizada na análise dos dados.

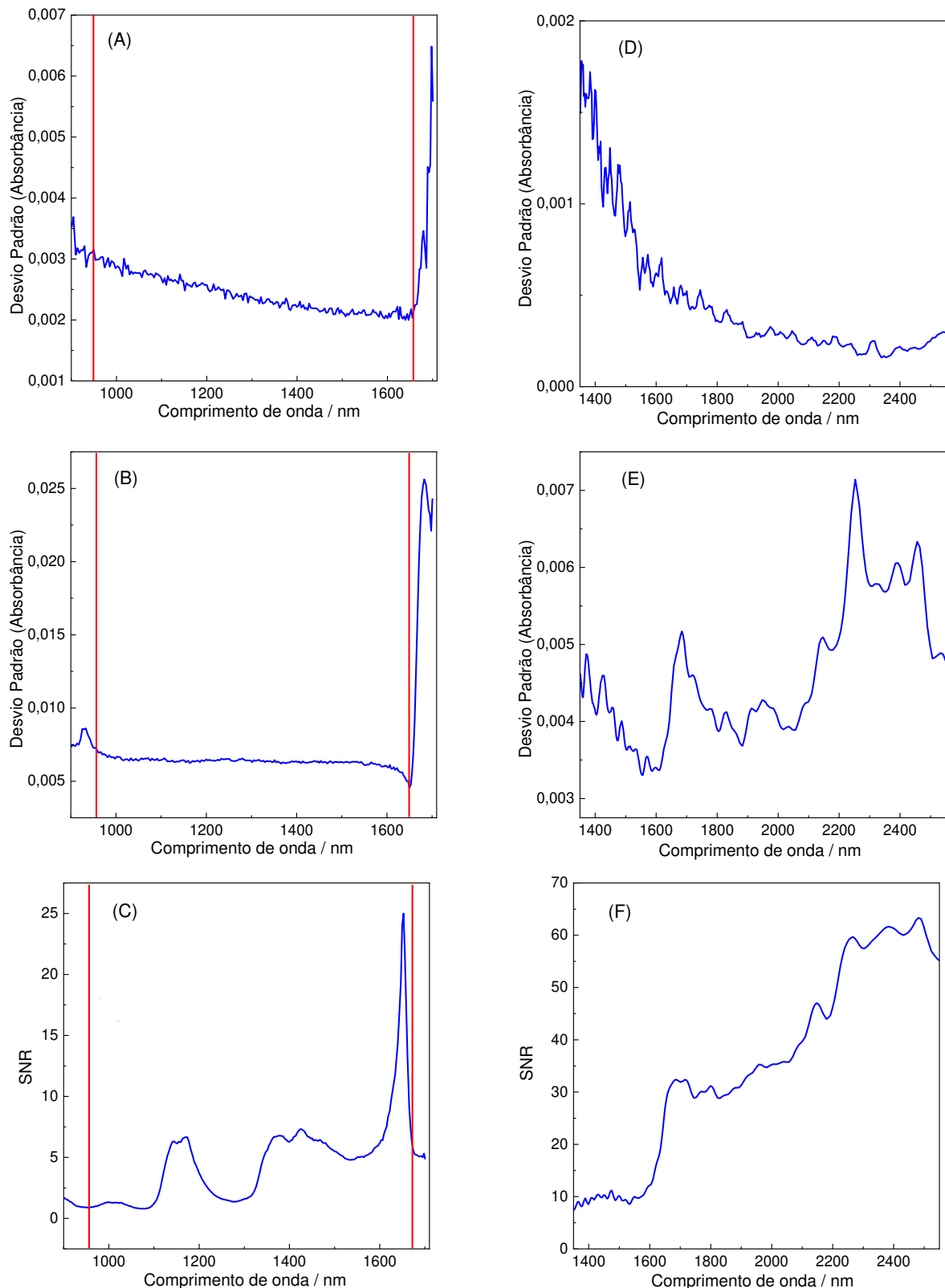


Figura 2.3. Desempenho geral dos espectrômetros portáteis. (A) e (D) desvio padrão da linha base; (B) e (E) desvio padrão da amostra controle; (C) e (F) relação sinal-ruído (SNR) do espectro da amostra controle, para os espectrômetros NanoNIR e NeoSpectra, respectivamente. As linhas vermelhas delimitam a região espectral considerada ideal para a faixa de trabalho do NanoNIR.

O conhecimento do desempenho a longo prazo dos espectrômetros, em especial, para os espectros das amostras de cocaína, é essencial para validar os resultados obtidos a partir da análise multivariada dos dados coletados.

2.4.2 Características dos espectros de infravermelho próximo de amostras de cocaína

Os espectros originais de infravermelho próximo de todas as amostras de cocaína, obtidos utilizando-se os espectrômetros NanoNIR e NeoSpectra, estão mostrados na Figura 2.4. Os espectros estão classificados de acordo com a forma de apresentação das amostras (cocaína base, cloridrato de cocaína ou forma não-definida), identificadas previamente por FTIR/ATR [39].

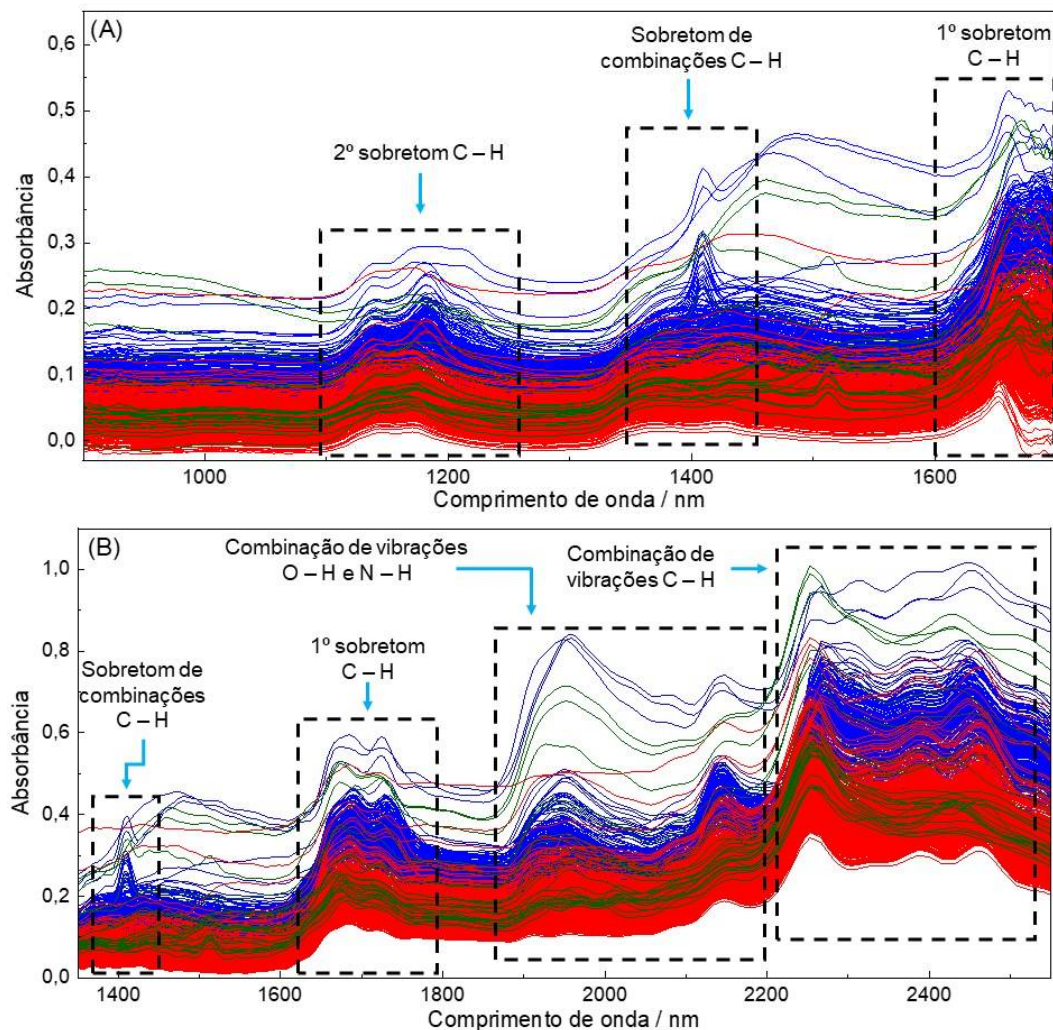


Figura 2.4. Espectros NIR originais do conjunto total de amostras de cocaína. (A) NanoNIR; (B) NeoSpectra. (—) espectros das amostras de cocaína base; (—) amostras de cloridrato de cocaína; (—) amostras não identificadas (N.D.)

As características espectrais mais relevantes ocorrem após 1100 nm. Observando-se os espectros das amostras de cocaína obtidos utilizando-se o NanoNIR (Figura 2.4A), é possível perceber três regiões com bandas de absorção proeminentes: a primeira região entre os comprimentos de onda de 1100 a 1250 nm (referente ao 2° sobretom do alongamento da ligação C – H); a segunda na faixa espectral de 1350 a 1450 nm (associada ao sobretom de combinações C – H); e, por fim, a terceira região, entre 1600 e 1700 nm (referente ao 1° sobretom do alongamento da ligação C – H) [40].

O final da faixa espectral monitorada pelo NanoNIR apresenta os maiores valores de absorvância para as amostras de cocaína, entretanto, como discutido na seção anterior (2.4.1), o desempenho do NanoNIR piora a partir de 1660 nm. Dessa forma, a qualidade das informações químicas contidas nos espectros NIR nessa região pode ser prejudicada, quando o esse instrumento é utilizado na coleta dos dados.

As características espectrais presentes nas segunda e terceira regiões destacadas para o NanoNIR, porém, também estão presentes na faixa de trabalho do NeoSpectra, visto que os dois instrumentos compartilham a região espectral entre 1350 e 1700 nm. Além dessas regiões, a faixa de trabalho do NeoSpectra também comporta outras duas regiões relevantes, com bandas de absorção intensas: uma entre 1800 e 2200 nm (relacionada a combinação de vibrações O – H e N – H) e a outra entre 2200 e 2500 nm (referente a combinação de vibrações C – H) [40], como pode ser observado na Figura 2.4B.

Há importantes diferenças espectrais entre as amostras de cocaína base (cujos espectros estão representados, na Figura 2.4, pelas linhas azuis) e de cloridrato de cocaína (linhas em vermelho), que podem auxiliar na discriminação das amostras de cocaína quanto à sua forma química.

Na faixa espectral abrangida pelo NanoNIR, as intensidades das bandas de absorção em 1142 nm e, principalmente, em 1194 nm mostram as maiores diferenças em relação às formas cloridrato de cocaína e cocaína base. O pico de absorção em 1142 nm está associado ao sobretom da ligação C – H de hidrocarbonetos aromáticos [40]. Ambas as formas básica e ácida da cocaína possuem, em sua estrutura, o anel aromático proveniente do éster do ácido benzoico [31], absorvendo nesse comprimento de onda. Já o pico em 1194 nm, referente ao sobretom da ligação C – H de grupos metil de hidrocarbonetos alifáticos [40], é mais proeminente nos espectros das amostras de cocaína base e pode estar associado à presença do grupo metil ligado à amina da base nitrogenada [31].

Já na faixa espectral varrida pelo NeoSpectra, as intensidades das bandas de absorção

em 1685 nm, 1730 - 1740 nm, 1955 nm e 2145 nm são elevadas, principalmente, para as amostras de cocaína base. O pico de absorção em 1685 nm é característico do 1º sobretom da ligação C – H de hidrocarbonetos aromáticos [40]. Entretanto, há absorção proveniente desse sobretom em toda a faixa de 1680 a 1770 nm [40]. Adicionalmente, em 1732 nm, há absorção advinda também do 1º sobretom da ligação C – H de hidrocarbonetos alifáticos [40].

Em 1735 nm, há absorção relacionada ao sobretom do estiramento da ligação C – H de amina [40] e a banda de absorção nessa região é mais pronunciada para amostras de cocaína base. Isso também pode estar associado à presença do grupo metil ligado ao nitrogênio da amina na cocaína base.

Por fim, a banda de absorção na região entre 1950 e 1960 nm está associada à vibração da ligação C = O proveniente de ésteres, enquanto o pico em 2145 nm relaciona-se à vibração da ligação C – H provenientes de ácidos carboxílicos [40].

Ainda analisando a Figura 2.4, pode-se observar que os espectros de absorção NIR das amostras de cocaína apresentam fortes efeitos aditivos e multiplicativos de espalhamento da radiação, devido, principalmente, ao fato de terem sido coletados diretamente nas amostras sólidas (com diferentes tamanhos de partículas).

A fim de tentar minimizar esses efeitos e realçar apenas as características espectrais informativas acerca da composição química das amostras, os espectros NIR foram pré-tratados, calculando-se a segunda derivada (algoritmo de Savitzky-Golay, com uma janela simétrica de 15 pontos e polinômio de segundo grau).

Os espectros derivados podem ser observados na Figura 2.5. Na Figura 2.5A, está representada a segunda derivada dos espectros obtidos a partir do NanoNIR. Optou-se por mostrar no gráfico apenas a faixa útil de trabalho (960 a 1660 nm), apresentada na seção 2.4.1, uma vez que fonte de variação fora dessa faixa poderia ser proveniente de ruído instrumental. E, na Figura 2.5B, está representada a segunda derivada dos espectros adquiridos utilizando-se o NeoSpectra.

Pode-se notar que a segunda derivada foi capaz de minimizar os efeitos de espalhamento da radiação e realçar as regiões espectrais importantes. As amplitudes, positivas e negativas, dos espectros derivados indicam as regiões mais informativas, com maior fonte de variabilidade, das amostras.

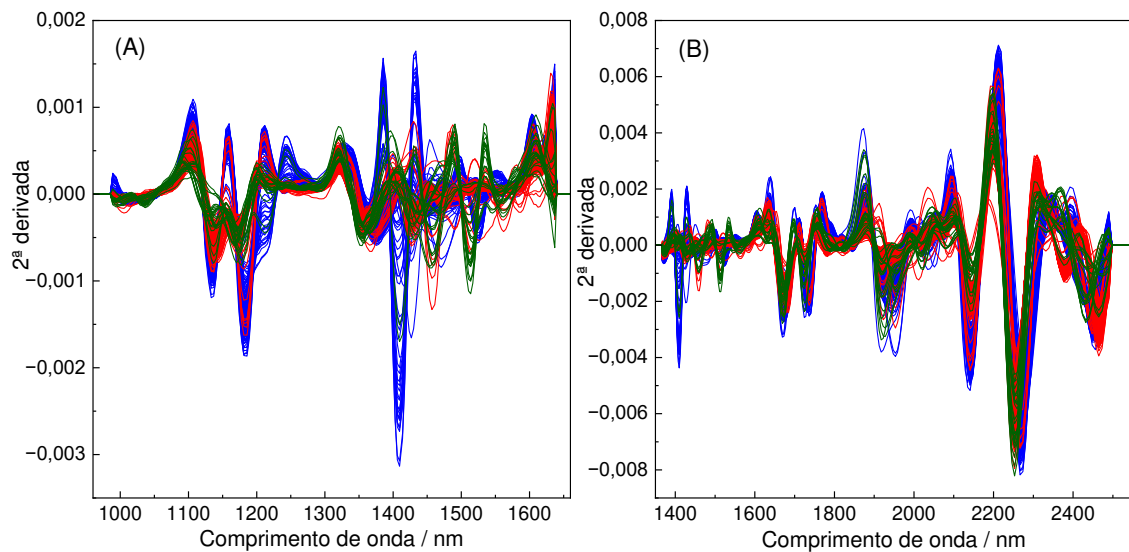


Figura 2.5. Espectros NIR das amostras de cocaína, após aplicar a 2ª derivada (Savitzky-Golay, janela simétrica de 15 pontos e polinômio de 2º grau). (A) NanoNIR; (B) NeoSpectra. (—) espectros derivados das amostras de cocaína base; (—) espectros das amostras de cloridrato de cocaína; (—) espectros das amostras não identificadas (N.D.).

Além disso, também é possível perceber que ambos os espectrômetros utilizados nas coletas dos dados, apesar de suas diferenças instrumentais, apresentam resposta analítica para as amostras de cocaína, visto que tanto na faixa de trabalho do NanoNIR quanto na do NeoSpectra, há a ocorrência de diversas regiões de máximos e mínimos.

2.4.3 Análise de Componentes Principais (PCA)

O conjunto de espectros pré-tratado das amostras de cocaína foi então submetido a análise de componentes principais (PCA), usando o algoritmo NIPALS (*Non-linear iterative partial least squares*) [41].

As Figuras 2.6A e 2.6B mostram os gráficos de *scores* para a primeira e a segunda componentes principais (PCs) das amostras de cocaína, utilizando-se os espectros obtidos com o NanoNIR e o NeoSpectra, respectivamente. Essas duas primeiras componentes, juntas, capturam 85 e 90 % da variância total do conjunto de dados espectrais, obtidos usando o NanoNIR e o NeoSpectra, respectivamente. Pode-se perceber que a PC-1, nos dois casos, tende a separar as amostras de acordo com a sua forma de apresentação (cocaína base ou cloridrato).

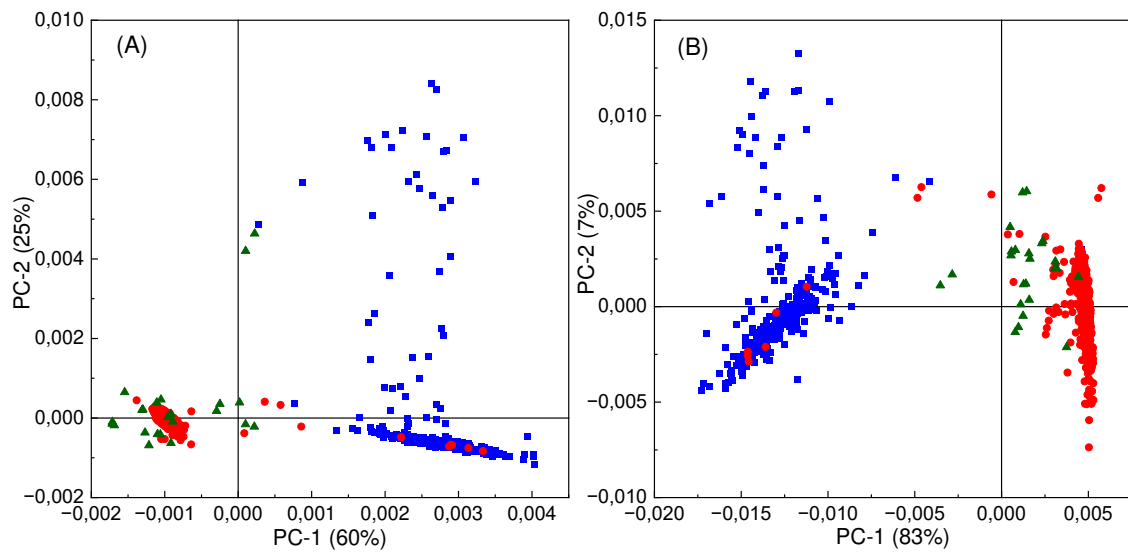


Figura 2.6. Distribuição de *scores* das PCAs das amostras de cocaína. (A) NanoNIR; (B) micro-FT-NIR NeoSpectra. ■ amostras de cocaína base; ● amostras de cloridrato de cocaína; ▲ amostras não-determinadas (N.D.).

Já nas Figuras 2.7A e 2.7B, estão representados os *scores*, para as 1ª e 2ª PCs, apenas das amostras de cocaína que contêm algum tipo de adulteração. A Figura 2.7A se refere aos dados obtidos utilizando-se o NanoNIR e a Figura 2.7B, aos dados obtidos com o NeoSpectra.

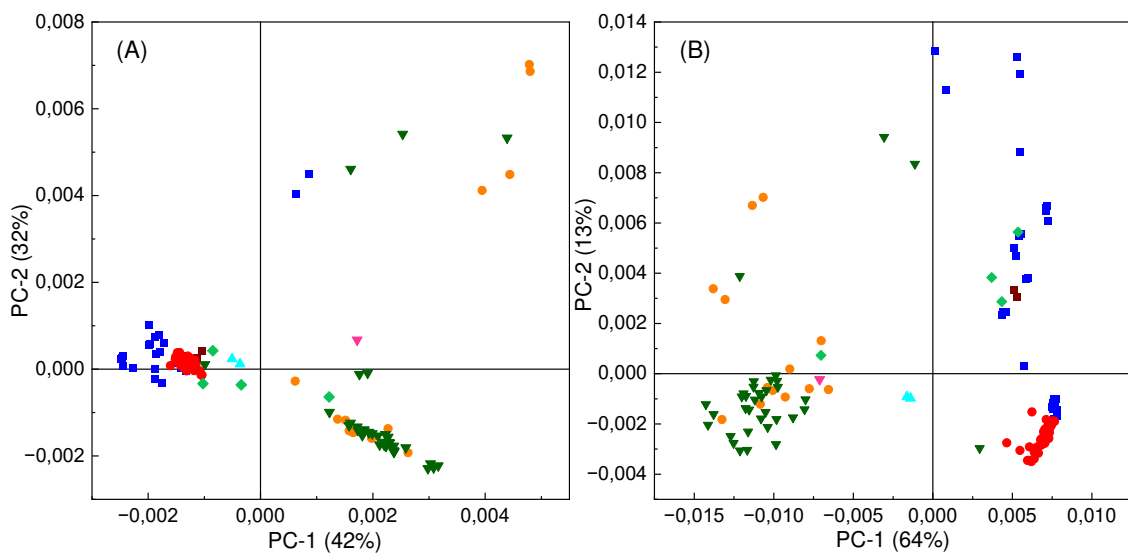


Figura 2.7. Distribuição de *scores* das PCAs das amostras de cocaína adulteradas. (A) NanoNIR; (B) micro-FT-NIR NeoSpectra. Amostras adulteradas majoritariamente com: ● aminopirina; ■ cafeína; ■ cafeína e lidocaína; ▲ fenacetina; ▲ fenacetina e lidocaína; ▼ hidroxizina; ● levamisol; ◆ lidocaína.

Em ambas as figuras, pode-se perceber que, para as amostras de cloridrato de cocaína (cujos *scores* estão projetados em valores negativos de PC-1 na Figura A e em valores positivos na Figura B), os principais adulterantes encontrados são a cafeína e o levamisol. Já para as

amostras de cocaína base (cujos *scores* estão projetados em valores positivos de PC-1 na Figura A e em valores negativos na Figura B), os principais adulterantes são a aminopirina e a fenacetina.

Ainda observando a Figura 2.7, pode-se perceber que há uma tendência no agrupamento das amostras adulteradas com levamisol em ambos os gráficos. Já as amostras adulteradas com cafeína tendem a se agrupar melhor no gráfico A (relacionado aos dados obtidos com o NanoNIR). No caso das amostras adulteradas com aminopirina e fenacetina, há uma sobreposição dos *scores* dessas amostras em ambos os gráficos, não sendo possível agrupá-las de acordo com o tipo de adulteração, empregando-se as duas primeiras componentes principais. Os gráficos de *scores* dessas amostras foram avaliados até a 6ª PC e a sobreposição dessas amostras permaneceu em todas as componentes principais.

Como a forma química da cocaína é uma característica com grande influência na distribuição dos *scores*, realizaram-se novas PCAs, considerando as amostras de cloridrato de cocaína e de cocaína base separadamente. A Figura 2.8 mostra os gráficos de distribuição de *scores* das PCAs, utilizando-se apenas as amostras de cloridrato de cocaína, enquanto a Figura 2.9 mostra as PCAs obtidas a partir das amostras de cocaína base.

No geral, tanto para as amostras de cloridrato de cocaína quanto para as amostras de cocaína base, as informações obtidas pelos dois instrumentos foram semelhantes.

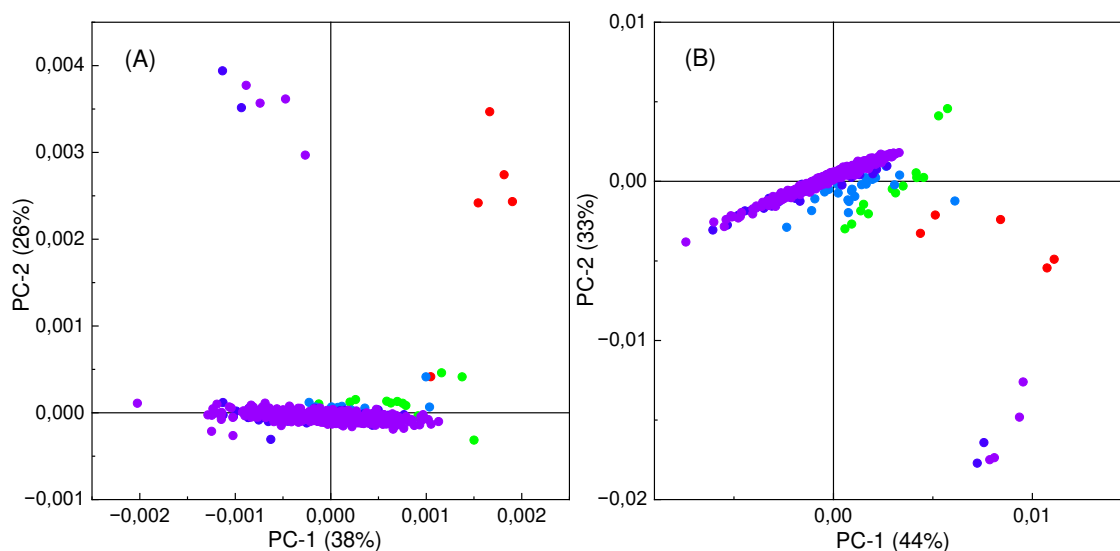


Figura 2.8. Distribuição de *scores* das PCAs das amostras de cloridrato de cocaína. (A) NanoNIR; (B) NeoSpectra. Os *scores* das amostras estão pintadas de acordo com suas respectivas faixas de teor de cocaína: ● amostras com teor de cocaína entre 0 e 20 %; ● 40 a 60 %; ● 60 a 80 %; ● 80 a 90 %; ● 90 a 100 %.

Observando-se a Figura 2.8, pode-se perceber que a maioria das amostras se concentrou

ao redor da origem dos eixos PC-1 e PC-2. Essas amostras possuem elevados teores de cocaína (acima de 80%), o que pode justificar o agrupamento.

É possível perceber também, na Figura 2.8B, que existe uma tendência de distribuição das amostras de acordo com seus respectivos teores de cocaína. Em relação à PC-2, há dispersão de algumas amostras ao longo desse eixo. Essas amostras possuem características peculiares, como baixo grau de oxidação ou presença de adulterantes, que podem ter sido responsáveis pelo deslocamento no eixo.

Um comportamento semelhante pode ser observado na Figura 2.9, que representa a distribuição de *scores* das PCAs das amostras de cocaína base para os dois espectrômetros.

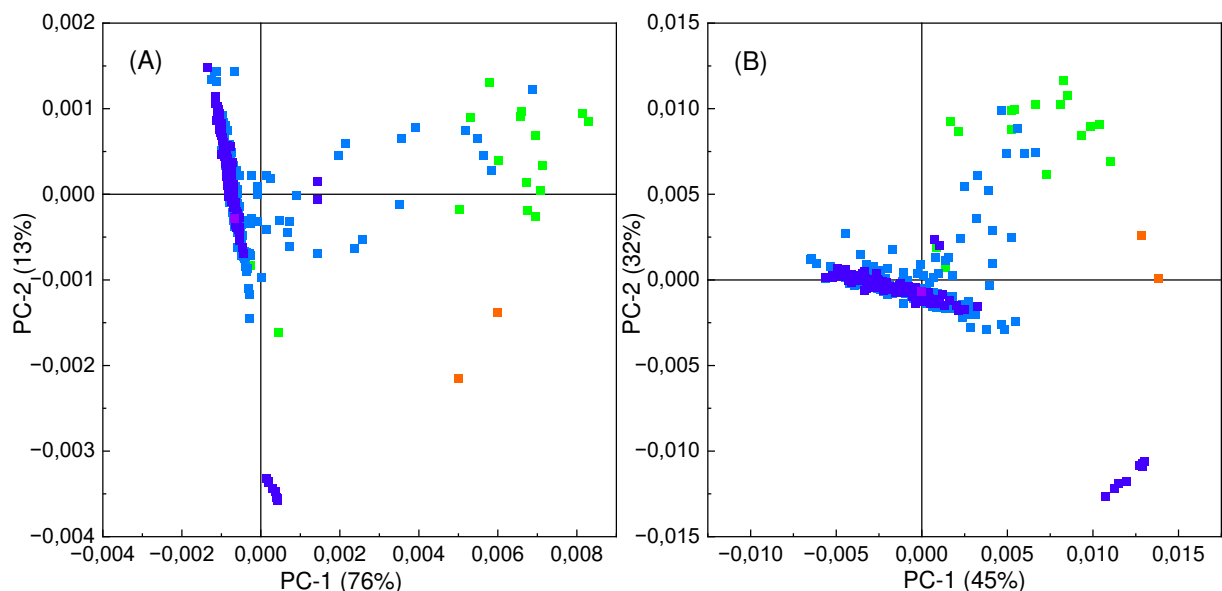


Figura 2.9. Distribuição de *scores* das PCAs das amostras de cocaína base (A) NanoNIR; (B) micro-FT-NIR NeoSpectra. Os *scores* das amostras estão pintadas de acordo com suas respectivas faixas de teor de cocaína: ■ amostras com teor de cocaína entre 20 e 40 %; ■ 40 a 60 %; ■ 60 a 80 %; ■ 80 a 90 %; ■ 90 a 100 %.

A análise dessas PCAs permite observar que a técnica NIR é capaz de extrair informações acerca da composição química das amostras de cocaína e, com isso, tem o potencial para ser usada na triagem dessas amostras.

2.4.4 Classificação SIMCA

Na triagem de amostras suspeitas, o principal objetivo é dizer se a amostra apreendida contém ou não cocaína. Nesse sentido, a classificação SIMCA foi utilizada empregando-se modelos PCA construídos com as informações espectrais obtidas nos dois instrumentos. Inicialmente, foram construídos modelos globais com o intuito de identificar a presença de

cocaína nas amostras independentemente da sua concentração, forma de apresentação ou presença concomitante de outras substâncias (adulterantes e/ou diluentes). Em seguida, foram construídos modelos para as formas de apresentação da cocaína mais recorrentes, base e cloridrato, a fim de tentar classificar a forma de apresentação da droga em amostras desconhecidas.

Posteriormente, um conjunto de teste independente, composto por 62 amostras de cloridrato de cocaína, 57 amostras de cocaína base, 8 amostras N.D. (que não puderam ser classificadas previamente nem como base e nem como cloridrato, mas que continham cocaína) e 14 amostras de adulterantes/diluentes, foi utilizado para testar o desempenho dos modelos.

A Tabela 2.1 apresenta as figuras de mérito de classificação SIMCA global e para as classes cloridrato e base, construídos a partir dos dados obtidos em cada instrumento. A indicação “*Texas*” entre parênteses se refere aos modelos construídos a partir dos espectros obtidos com o espectrômetro NanoNIR; enquanto a indicação “*NeoS*”, refere-se aos modelos construídos com os dados obtidos do micro-FT-NIR NeoSpectra.

Tabela 2.1. Resultados da validação externa dos modelos SIMCA para a classificação da cocaína

Modelo SIMCA	Faixa espectral / nm	Amostras modelagem	PCs	TP	TN	FP	FN	FPR (%)	FNR (%)	STR (%)	SPR (%)	EFR (%)
Global (<i>Texas</i>)	970 a 1670	907	7	113	14	0	14	0	11	89	100	89
Global (<i>NeoS</i>)	1350 a 2550	907	7	113	14	0	14	0	11	89	100	89
Cloridrato (<i>Texas</i>)	970 a 1670	673	7	56	74	5	6	6	10	90	94	84
Cloridrato (<i>NeoS</i>)	1350 a 2550	673	7	57	74	5	5	6	8	92	94	86
Base (<i>Texas</i>)	970 a 1670	218	7	51	82	2	6	2,4	11	89	98	87
Base (<i>NeoS</i>)	1350 a 2550	218	6	52	82	2	5	2,4	8,8	91	98	89

TP: Total de amostras positivas reais; TN: total de amostras negativas reais; FP: número de falsos positivos; FN: número de falsos negativos; FPR = $[FP / (FP + TN)] \times 100$, taxa de falsos positivos; FNR = $[FN / (FN + TP)] \times 100$, taxa de falsos negativos; STR = $[TP / (TP + FN)] \times 100$, seletividade; SPR = $(TN / (TN + FP)) \times 100$, especificidade; EFR = $100 - (FPR + FNR)$, eficiência.

No geral, os modelos construídos a partir dos dados dos dois instrumentos tiveram um desempenho similar. Em ambos os modelos globais, não houve nenhum falso positivo, ou seja, os modelos não classificaram como cocaína nenhuma amostra que não continha cocaína (no conjunto testado, as amostras sem cocaína foram compostas pelos principais adulterantes e diluentes analisados). Entretanto, também nos dois modelos globais, houve 14 falsos negativos,

ou seja, 14 amostras que continham cocaína não foram classificadas, pelos modelos, como pertencentes ao grupo de amostras com cocaína (o que representa 11 % das amostras testadas desse grupo). Nos dois casos, essas amostras foram praticamente as mesmas, sendo que, 7 amostras pertenciam à classe de amostras “N.D.”, 4 amostras eram de base e 3 de cloridrato de cocaína e, além disso, 12 dessas 14 amostras estavam adulteradas, principalmente, com altos teores de cafeína, fenacetina, levamisol e/ou lidocaína.

O mesmo conjunto de teste foi utilizado para avaliar o desempenho dos modelos locais (modelos para as classes “cloridrato” e “base”). Observando-se a Tabela 2.1, pode-se notar que os números de falsos positivos aumentaram e os números de falsos negativos diminuíram nos modelos locais em relação aos modelos globais.

No caso dos falsos positivos, todas as amostras continham cocaína e os modelos locais falharam apenas em classificar a forma de apresentação da droga. Os modelos locais de cloridrato, para ambos os instrumentos, apresentaram cinco falsos positivos, dos quais quatro eram amostras de cocaína base e um era uma amostra “N.D.”. Já os modelos locais de base, apresentaram dois falsos positivos, que consistiam em amostras de cloridrato de cocaína que foram erroneamente classificadas como cocaína base pelos dois instrumentos.

Já no caso dos falsos negativos, houve uma pequena melhora nas taxas de falsos negativos e na seletividade para os modelos locais de cloridrato de ambos os instrumentos e no modelo local de base do Neospectra, em relação aos modelos globais. O modelo local de base do Texas apresentou o mesmo desempenho que o modelo global do Texas para esses parâmetros. O fato da diminuição dos números de falsos negativos nos modelos locais em comparação aos modelos globais não ter influenciado tanto no resultado das taxas de falsos negativos e de seletividade se deve às 7 amostras “N.D.”, presentes nos grupos de falsos negativos dos modelos globais. Essas amostras, quando testadas pelos modelos locais, também resultaram em respostas negativas. Entretanto, como suas formas químicas não puderam ser determinadas pelo método de referência, ou seja, não se pôde determinar se se tratavam de cloridrato ou de base, elas foram classificadas como verdadeiros negativos pelos modelos locais de cloridrato e base.

Dessa forma, comparando-se os modelos globais e locais, há a variação de alguns parâmetros, entretanto, essas variações não refletem em mudanças concretas nos desempenhos dos modelos. Além disso, pode-se perceber, para os modelos globais, que ambos os instrumentos tiveram o mesmo desempenho nas classificações. Já para os modelos locais, os resultados de classificação obtidos com o NeoSpectra foram um pouco melhores que os obtidos

com o Texas.

2.4.5 Modelos de regressão PLS

Além de conseguir informar a presença ou ausência de cocaína em substâncias suspeitas e sua forma de apresentação, este trabalho também propôs estimar o teor de cocaína nas amostras apreendidas. Para isso, foram construídos modelos de regressão PLS globais (contendo amostras de cocaína base e cloridrato) e modelos de regressão PLS locais para as classes “cloridrato de cocaína” e “cocaína base”, utilizando os dados espectrais obtidos com o NanoNIR e o NeoSpectra.

A Figura 2.10 mostra os gráficos dos valores preditos *versus* de referência para o modelo global de regressão PLS do teor de cocaína das amostras analisadas pelo NanoNIR, para os conjuntos de calibração e de validação. O número de variáveis latentes (LV) utilizadas nesse modelo foi igual a sete.

Já a Figura 2.11 mostra os gráficos dos valores preditos *versus* de referência para o modelo global de regressão PLS do teor de cocaína das amostras analisadas pelo NeoSpectra, para os conjuntos de calibração e de validação. Nesse modelo, o número de variáveis latentes utilizadas foi seis.

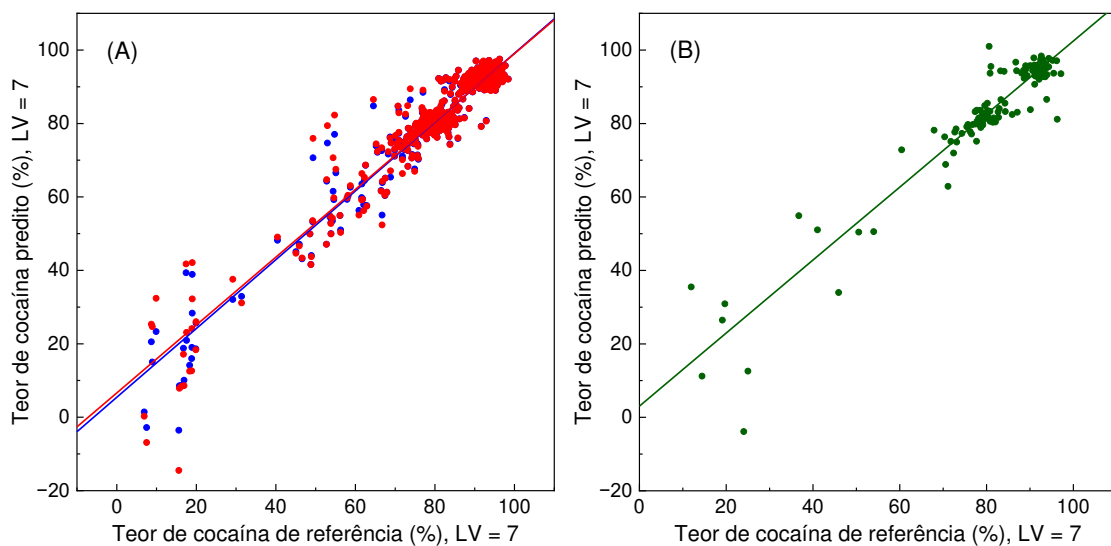


Figura 2.10. Valores preditos *versus* de referência para o modelo global de regressão PLS do teor de cocaína das amostras analisadas pelo NanoNIR. (A) Conjunto de calibração, onde ● representa os valores das amostras de calibração; ● representa os valores da validação cruzada das amostras do conjunto de calibração. (B) Conjunto de validação, onde ● representa os valores das amostras do conjunto de predição.

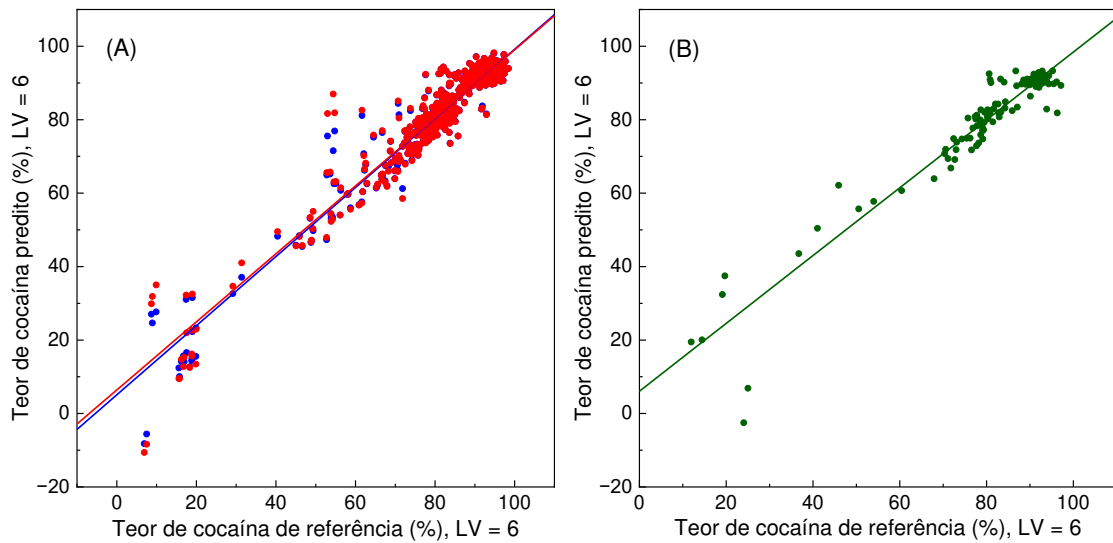


Figura 2.11. Valores preditos *versus* de referência para o modelo global de regressão PLS do teor de cocaína das amostras analisadas pelo NeoSpectra. (A) Conjunto de calibração, onde ● representa os valores das amostras de calibração; e ● representa os valores da validação cruzada das amostras do conjunto de calibração. (B) Conjunto de validação, onde ● representa os valores das amostras do conjunto de predição.

As Figuras 2.12 e 2.13 mostram os gráficos dos valores preditos *versus* de referência para o modelo local “cloridrato” de regressão PLS do teor de cocaína, para os conjuntos de calibração e validação, das amostras analisadas pelo NanoNIR e Neospectra, respectivamente. Os números de variáveis latentes utilizadas foram iguais a seis em ambos os casos.

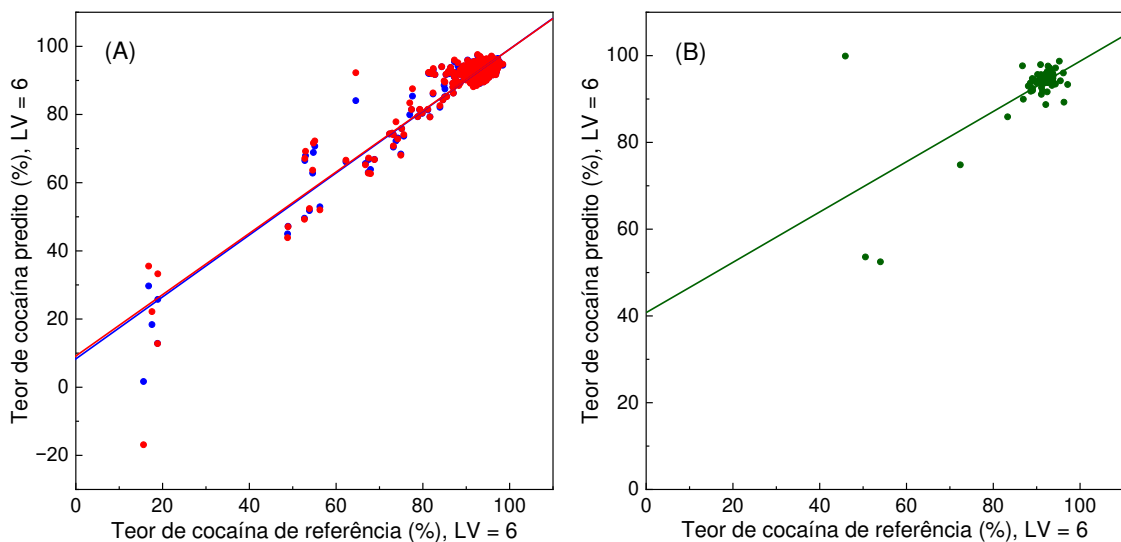


Figura 2.12. Valores preditos *versus* de referência para o modelo local “cloridrato” de regressão PLS do teor de cocaína das amostras analisadas pelo NanoNIR. (A) Conjunto de calibração, onde ● representa os valores das amostras de calibração; ● representa os valores da validação cruzada das amostras do conjunto de calibração. (B) Conjunto de validação, onde ● representa os valores das amostras do conjunto de predição.

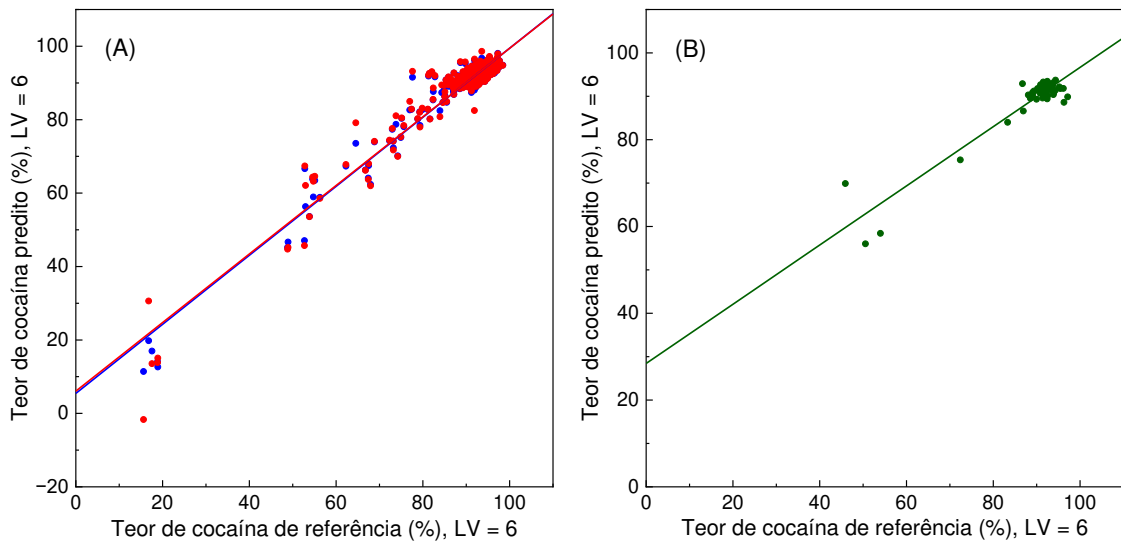


Figura 2.13. Valores preditos *versus* de referência para o modelo local “cloridrato” de regressão PLS do teor de cocaína das amostras analisadas pelo NeoSpectra. (A) Conjunto de calibração, onde ● representa os valores das amostras de calibração; e ● representa os valores da validação cruzada das amostras do conjunto de calibração. (B) Conjunto de validação, onde ● representa os valores das amostras do conjunto de predição.

Por fim, as Figuras 2.14 e 2.15 mostram os gráficos dos valores preditos *versus* de referência para o modelo local “base” de regressão PLS do teor de cocaína, para os conjuntos de calibração e validação, das amostras analisadas pelo NanoNIR e Neospectra, respectivamente.

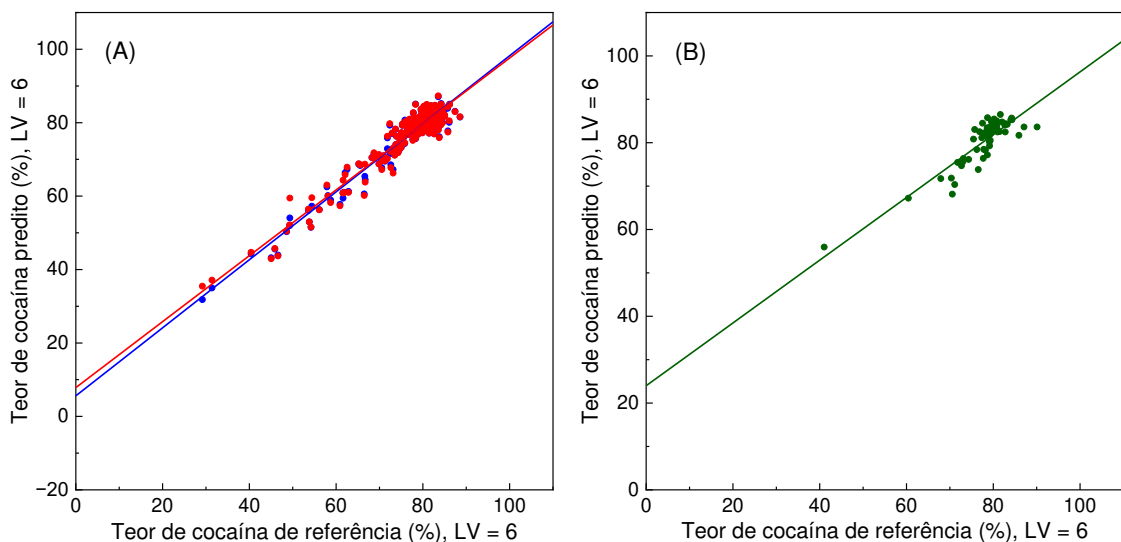


Figura 2.14. Valores preditos *versus* de referência para o modelo local “base” de regressão PLS do teor de cocaína das amostras analisadas pelo NanoNIR. (A) Conjunto de calibração, onde ● representa os valores das amostras de calibração; ● representa os valores da validação cruzada das amostras do conjunto de calibração. (B) Conjunto de validação, onde ● representa os valores das amostras do conjunto de predição.

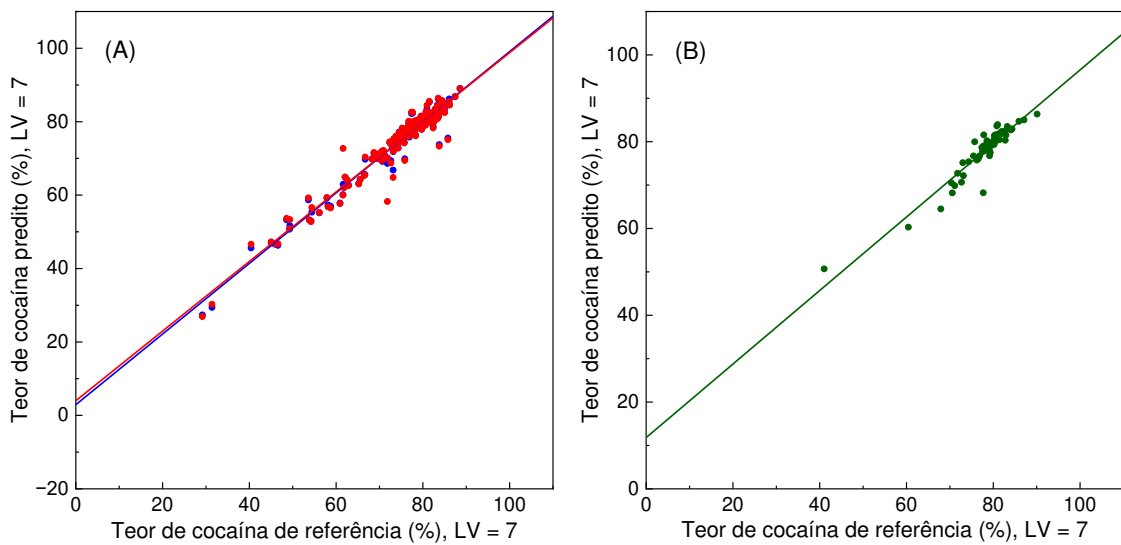


Figura 2.15. Valores preditos *versus* de referência para o modelo local “base” de regressão PLS do teor de cocaína das amostras analisadas pelo NeoSpectra. (A) Conjunto de calibração, onde ● representa os valores das amostras de calibração; e ● representa os valores da validação cruzada das amostras do conjunto de calibração. (B) Conjunto de validação, onde ● representa os valores das amostras do conjunto de predição.

As figuras de mérito para os modelos globais e para os modelos locais “cloridrato” e “base” estão sintetizadas na Tabela 2.2.

Apesar do desempenho dos dois instrumentos ter sido equiparável nos modelos de classificação SIMCA, o NeoSpectra apresentou melhores desempenhos nos modelos de regressão PLS, tanto globais quanto locais, quando comparado ao NanoNIR.

O modelo global (*Texas*), referente ao NanoNIR, apresentou valores de RMSECV e RMSEP (4,1 e 6,4 %, respectivamente) maiores do que os do modelo global (*NeoS*) (3,9 e 5,2 %, respectivamente), que se refere ao NeoSpectra. Além disso, os coeficientes de determinação (R^2), tanto na etapa de calibração quanto na de validação, foram melhores para o modelo global (*NeoS*) e os valores de *bias* (erro sistemático) foram iguais na etapa de calibração e menor na etapa de validação para o modelo global (*NeoS*), quando comparados ao modelo global (*Texas*).

Um comportamento semelhante pode ser observado nos modelos locais. O modelo local “cloridrato” (*Texas*) apresentou valores de RMSECV e RMSEP (3,3 e 7,7 %, respectivamente) maiores do que os do modelo local “cloridrato” (*NeoS*) (2,6 e 3,9 %, respectivamente). E o modelo local “base” (*Texas*) também apresentou valores de RMSECV e RMSEP (2,9 e 4,1 %, respectivamente) maiores do que os do modelo local “base” (*NeoS*) (2,4 e 2,4 %, respectivamente).

Tabela 2.2. Resultados para a calibração e validação de modelos PLS globais e locais construídos para a determinação do teor de cocaína, utilizando-se os espectrômetros NanoNIR e NeoSpectra

Calibração						
Modelo de regressão	Faixa espectral / nm	Nº de amostras	Nº de variáveis latentes (LV)	RMSECV / %	R ²	Bias / %
Global (Texas)	970 a 1670	907	7	4,1	0,919	0,06
Global (NeoS)	1350 a 2550	907	6	3,9	0,926	0,06
Cloridrato (Texas)	970 a 1670	673	6	3,3	0,870	0,04
Cloridrato (NeoS)	1350 a 2550	673	6	2,6	0,919	0,03
Base (Texas)	970 a 1670	218	6	2,9	0,912	0,08
Base (NeoS)	1350 a 2550	218	7	2,4	0,940	0,04
Validação						
Modelo de regressão	Faixa espectral / nm	Nº de amostras	Nº de variáveis latentes (LV)	RMSEP / %	R ²	Bias / %
Global (Texas)	970 a 1670	127	7	6,4	0,864	2,58
Global (NeoS)	1350 a 2550	127	6	5,2	0,909	-0,09
Cloridrato (Texas)	970 a 1670	62	6	7,7	0,348	3,15
Cloridrato (NeoS)	1350 a 2550	62	6	3,9	0,875	0,00
Base (Texas)	970 a 1670	57	6	4,1	0,782	2,44
Base (NeoS)	1350 a 2550	57	7	2,4	0,882	-0,08

RMSECV = erro quadrático médio de validação cruzada; RMSEP = erro quadrático médio de predição; R² = coeficiente de determinação.

Comparando-se os modelos locais “cloridrato” e “base” (Texas) com o modelo global (Texas), pode-se perceber que os erros de validação cruzada (RMSECV) foram menores e os erros de predição (RMSEP) também foram menores, exceto para o modelo local “cloridrato” (Texas), cujo RMSEP foi igual a 7,7 %, o que representou o maior erro de predição entre todos os modelos. Além disso, o modelo local “cloridrato” (Texas) apresentou o menor coeficiente de determinação (R² = 0,348) entre os modelos e um bias elevado (3,15 %).

Esse pior desempenho do modelo local “cloridrato” (Texas) na etapa de predição pode ser justificado pela presença de uma amostra no conjunto de validação que não foi bem explicada pelo modelo, alavancando o modelo de predição, como pode ser observado na

Figura 2.12B.

No caso dos modelos locais “cloridrato” e “base” (*NeoS*), quando comparados ao modelo global (*NeoS*), pode-se perceber que os parâmetros erro quadrático médio e *bias*, apresentados na Tabela 2.2, são melhores para os modelos locais, tanto na etapa de calibração quanto na etapa de validação.

Dessa forma, no geral, os modelos construídos utilizando os dados do NeoSpectra foram melhores que os modelos construídos com os dados do Texas e os modelos globais apresentaram desempenhos inferiores aos modelos locais “cloridrato” e “base”, com exceção do modelo local “cloridrato” (*Texas*).

2.5 CONCLUSÕES

Com base nos desempenhos dos dois instrumentos, tanto em relação aos modelos de classificação quanto nos modelos de predição do teor de cocaína, é possível propor um novo método de triagem, alternativo aos métodos já empregados, utilizando a espectroscopia NIR e a análise multivariada de dados, capaz de identificar a presença de cocaína, a sua forma de apresentação e de estimar o seu teor em amostras suspeitas.

Como esse método utiliza a técnica NIR e instrumentos portáteis e de baixo custo, ele permite a caracterização da amostra suspeita no momento de sua apreensão, através da análise direta e não-destrutiva do material, preservando, assim, possíveis evidências de crimes. Além disso, é possível extrair mais informações acerca das amostras no momento da apreensão quando comparado aos métodos colorimétricos, por exemplo.

Apesar dos dois instrumentos apresentarem bons desempenhos, o espectrômetro micro-FT-NIR NeoSpectra (*Si-ware*), no geral, produziu modelos melhores. Contudo, o NanoNIR (*Texas Instruments*) também foi capaz de fornecer modelos de classificação e de regressão satisfatórios e apresenta um valor de mercado cinco vezes menor que o NeoSpectra, tornando-se assim uma opção mais viável economicamente.

Capítulo 3: Explorando a quimiometria para análise de triagem por espectroscopia no infravermelho próximo sem o apoio de métodos analíticos de referência: aplicação a edulcorantes à granel

Este capítulo é baseado no manuscrito publicado:

Jamille Carvalho Souza, Maria C. Hespanhol, Fábio R. P. Rocha, Celio Pasquini
Microchemical Journal, 190 (2023) 108732. <https://doi.org/10.1016/j.microc.2023.108732>

RESUMO

Uma nova abordagem analítica baseada em espectroscopia de infravermelho próximo (NIR), usando um espectrofotômetro miniaturizado baseado em transformada de Fourier, análise de correlação e análise multivariada de dados por componentes principais (PCA), resolução de curva multivariada por mínimos quadrados alternados (MCR-ALS), e regressão por mínimos quadrados parciais (PLS) é proposta para selecionar produtos comerciais e matérias-primas. A combinação de técnicas multivariadas com inspeção visual de espectros e correlação espectral produz resultados qualitativos e quantitativos confiáveis, reduzindo o custo de implementação ao evitar o uso de instrumentos caros e grandes conjuntos de dados de amostras analisadas por métodos analíticos de referência caros e trabalhosos. Como estudo de caso, o procedimento é aplicado a edulcorantes comercializados à granel comumente empregados em substituição à sacarose em dietas pobre em carboidratos. O efeito do tamanho das partículas de produtos comerciais e suas misturas no espectro NIR é identificado e discutido. Estratégias para determinar a autenticidade e a adulteração são propostas e avaliadas utilizando apenas uma amostra de referência para cada edulcorante testado. A resposta espectral não linear observada para misturas binárias dos edulcorantes é discutida. Misturas de edulcorantes, simulando adulterações de produtos, são analisadas por meio de PCA, MCR-ALS não linear e regressão de mínimos quadrados parciais (PLS), ajudando a confirmar os adulterantes e estimar o nível de adulteração de amostras falsificadas. É fornecido um roteiro para desenvolver e empregar o procedimento analítico proposto.

Palavras-chave: Espectroscopia no infravermelho próximo; Análise de correlação espectral; Análise espectral multivariada; Método de análise de triagem; Análise de edulcorantes à granel.

3.1 INTRODUÇÃO

O uso de técnicas multivariadas e espectroscópicas vibracionais, como a espectroscopia no infravermelho próximo (NIR), para avaliar a autenticidade e a adulteração de alimentos e medicamentos é amplamente relatado na literatura [42], [43], [44], [45], [46], [47], [48].

Os procedimentos descritos para o desenvolvimento desses métodos analíticos baseiam-se em grandes conjuntos de dados amostrais, muitas vezes exigindo análise por métodos analíticos de referência caros e trabalhosos, como cromatografia e espectrometria de massas.

A precisão e a incerteza aceitáveis dos resultados analíticos podem ser reduzidas se aplicadas para fins de triagem. Assim, torna-se possível projetar procedimentos analíticos rápidos desenvolvidos sem técnicas analíticas de referência, empregando um mínimo de amostras certificadas em alguns casos. No entanto, a literatura carece de descrever estratégias eficazes para empregar a coleta de informações analíticas por técnicas instrumentais, como a espectroscopia NIR, para orientar o analista a produzir resultados de triagem confiáveis, empregando análise multivariada na ausência de grandes conjuntos de dados de amostras de referência. Em princípio, as informações fornecidas pela técnica instrumental devem ser autossuficientes para apoiar o desenvolvimento de métodos de análise de triagem, considerando que um método não pode ter desempenho melhor do que as informações oferecidas pela técnica associada. Portanto, estes métodos poderiam ser desenvolvidos explorando ferramentas já disponíveis para inspeção e análise de dados, incluindo análise multivariada, para extrair informações de conjuntos de dados, utilizando poucas amostras de referência para orientar a interpretação das informações.

A espectroscopia NIR está compreendida na faixa de comprimento de onda de 750 a 2500 nm da radiação eletromagnética. Nesta faixa de energia, a radiação eletromagnética promove a excitação de combinações e sobretons de modos vibracionais fundamentais de átomos ligados ou grupos de átomos que constituem as moléculas. O espectro de absorção resulta principalmente da interação da radiação para excitar harmônicos e combinações dos modos vibracionais de grupos funcionais constituídos principalmente por ligações C – H, N – H e O – H [1], [33], [49]. Os efeitos secundários causados pelas interações intra ou intermoleculares da estrutura molecular envolvendo esses grupos funcionais provocam alterações na energia e na anarmonicidade das ligações químicas, refletindo alterações nos espectros de absorção das substâncias [49]. Essas variações espectrais conferem seletividade à espectroscopia NIR, que é empregada para fins analíticos qualitativos e quantitativos [1].

O consumo de edulcorantes naturais tem aumentado em todo o mundo para atender às restrições alimentares impostas pelo diabetes ou para reduzir o número de calorias ingeridas em dietas mais saudáveis. Vários adoçantes são adquiridos à granel, muitas vezes vendidos sem identificação de origem e certificados de qualidade. É o caso de edulcorantes como eritritol, maltitol, xilitol, maltodextrina e sorbitol. Devido à sua aparência semelhante a olho nu como um pó branco com diferentes graus de tamanho de partícula, eles são propensos à adulteração e à falsificação, principalmente por razões econômicas.

A maioria desses edulcorantes são álcoois de carboidratos com cadeias de carbono de tamanhos diferentes, com alguns grupos substituintes. Os edulcorantes à granel apresentam características sensoriais semelhantes, mas não iguais. Por exemplo, os seus índices glicêmicos e valores calóricos diferem significativamente. A falsificação ou adulteração de qualquer um desses edulcorantes à granel pode prejudicar a saúde do consumidor ou, pelo menos, desrespeitar o seu direito de conhecer o real produto que está consumindo.

A técnica mais usual para analisar esses edulcorantes em diversos tipos de alimentos é a cromatografia líquida de alta eficiência (HPLC) [50], [51], [52]. A determinação de eritritol, maltitol, xilitol e sorbitol em chocolates, usando eletroforese capilar com detecção de condutividade sem contato acoplada capacitivamente também foi descrita [53]. A espectrometria de massas tem sido usada para detectar adulteração em galactooligosacarídeos comerciais [54]. Embora estas técnicas possam ser adaptadas para analisar edulcorantes à granel prontos para consumo, elas são demoradas e requerem instrumentos caros. No entanto, nenhum trabalho pode ser encontrado na literatura considerando este assunto.

A espectroscopia vibracional tem sido empregada para analisar diversos edulcorantes. A espectroscopia Raman foi utilizada para determinar as formas polimórficas de sorbitol e manitol em medicamentos [55]. A mesma técnica, empregando um espectrômetro portátil, também é proposta para detectar produtos falsificados de estévia e quantificar sacarina e ciclamato em formulações de mesa [56], [57].

Quanto à utilização da espectroscopia NIR para análise de edulcorantes, a literatura relata alguns trabalhos dedicados às substâncias de origem artificial e natural. A quantificação dos edulcorantes artificiais sacarina e ciclamato em formulações de mesa usando um espectrômetro NIR portátil e regressão de mínimos quadrados parciais (PLS) foi descrita [57]. A espectroscopia NIR foi empregada para determinar edulcorantes naturais como esteviosídeo e rebaudiosídeo A em folhas de *stevia rebaudiana Bertoni*, empregando análise multivariada [58]. Utilizando espectroscopia NIR, são descritos o monitoramento em tempo real da produção

de xilitol em um biorreator e a equivalência de dextrose de produtos de maltodextrina resultantes da hidrólise [59], [60]. Essa técnica também é empregada para determinar o teor de umidade em manitol liofilizado usando microespectrômetros [61] e avaliar o mecanismo de granulação úmida em misturador de alta velocidade com silicato de cálcio poroso e álcoois de açúcar [62]. Recentemente, a espectroscopia NIR foi comparada à espectroscopia de infravermelho ATR-FT e Raman para identificar vários adoçantes e açúcares artificiais rapidamente [63]. As substâncias avaliadas foram sacarose, galactose, frutose, glicose, rebaudiosídeo A; os adoçantes artificiais sacarina, sucralose, neotame, acessulfame K. No entanto, nenhum edulcorante à granel comum foi avaliado.

A espectroscopia NIR foi recentemente impulsionada pela introdução dos espectrofotômetros NIR compactos de baixo custo [1], [64]. Esses espectrofotômetros miniaturizados apresentam desempenho comparável aos instrumentos de bancada em diversas aplicações analíticas de espectroscopia NIR [64] e custam algo entre US\$ 1.000 e 5.000. Além disso, os dispositivos são portáteis e podem ser usados para análise em campo. Estas características dos instrumentos NIR compactos são relevantes para programas de inspeção que visam rastrear rapidamente a autenticidade e adulterações de produtos comerciais. Com base em um levantamento da literatura atual, não se encontraram trabalhos que visem avaliar a autenticidade e adulterações de edulcorantes à granel utilizando espectroscopia NIR.

Este trabalho propõe uma nova estratégia analítica baseada em espectroscopia NIR, com um número mínimo de amostras de referência, e informações fornecidas por análise de componentes principais (PCA), análise de correlação, modelagem independente e flexível por analogias de classes (SIMCA) [28], resolução de curva multivariada por mínimos quadrados alternados (MCR-ALS) [65], [66], e regressão de mínimos quadrados parciais (PLS) [28]. Quando combinadas, as informações fornecidas pelas técnicas multivariadas, a avaliação gráfica de espectro e a análise de correlação produzem resultados analíticos confiáveis com o objetivo de rastrear a autenticidade e adulteração de produtos comerciais. O protocolo proposto é aplicado a edulcorantes comerciais à granel como um estudo de caso.

3.2 OBJETIVO

Desenvolver uma nova abordagem analítica, utilizando espectroscopia NIR, um espectrofotômetro NIR portátil e de baixo custo e análise multivariada de dados, para rastrear a autenticidade e a qualidade de produtos comerciais, sem o suporte de análises de referência.

3.3 MATERIAIS E MÉTODO

3.3.1 Amostras

A Tabela 3.1 descreve o conjunto amostral de edulcorantes empregados neste trabalho e suas principais características dietéticas. Os edulcorantes à granel foram adquiridos de fornecedores brasileiros tanto na internet quanto em lojas físicas durante um período de 6 meses. O conjunto amostral contempla a maior parte das fontes de variabilidade, incluindo a granulometria, o lote e a aparência física dos edulcorantes mais consumidos (eritritol e xilitol). Amostras de referência dos edulcorantes foram adquiridas da Sigma-Aldrich com pureza de 98 a 99%. Vários outros edulcorantes artificiais e naturais foram adicionados ao conjunto de amostras para avaliar a capacidade da espectroscopia NIR de discriminar entre edulcorantes naturais e artificiais. Todas as amostras comerciais e de referência se mostram como pós brancos com diferentes graus de tamanho de partícula e aparência cristalina.

Tabela 3.1. Descrição do conjunto de amostras comerciais utilizadas e suas principais características dietéticas

Rótulo da amostra	Número de amostras	Índice glicêmico*	Valor calórico / kcal/g
xilitol	44	12	2,4
eritritol	28	1	0,2
sorbitol	5	4	2,6
maltitol	2	35	2,4
maltodextrina	5	105-110	3,8
frutose	2	23	4
sucrose	5	65	4
estévia	6	0	0
sucralose	6	0	0

* <http://www.sugar-and-sweetener-guide.com/glycemic-index-for-sweeteners.html>

Amostras contendo misturas de maltitol/eritritol, maltodextrina/eritritol, sorbitol/eritritol, xilitol/maltitol e sorbitol/xilitol, com o conteúdo das substâncias variando na faixa de 0 – 100% (m/m) (em incrementos de 10%) foram preparadas usando substâncias comerciais autênticas, para representar prováveis práticas de adulteração. As massas adequadas das substâncias, totalizando cerca de 30 g, foram pesadas em saco de polietileno, utilizando balança analítica, homogeneizadas e transferidas para frascos de vidro de medição. As misturas contendo maltodextrina e eritritol foram posteriormente trituradas manualmente em almofariz para homogeneizar a granulometria dos constituintes.

3.3.2 Espectrofotômetro NIR e dispositivo de leitura de amostras

Foi utilizado um espectrofotômetro compacto NeoSpectra micro-FT-NIR (Si-Ware Systems, Egito). O instrumento opera na faixa de comprimento de onda de 1350–2550 nm. A resolução espectral em 1550 nm é de 16 nm. Os espectros compreendem 257 pontos (comprimentos de onda) após a transformação de Fourier do interferograma médio. Um tempo de integração de 20 s foi utilizado por aquisição de espectro único. O instrumento foi conectado a uma porta USB de um notebook e ligado 20 minutos antes do início das medições. A fonte de radiação do espectrofotômetro foi ligada apenas durante a aquisição do espectro. As medições foram realizadas à temperatura ambiente do laboratório de 25 °C.

O espectrofotômetro NIR compacto foi acoplado a um dispositivo de leitura de amostras descrito anteriormente [27] para obter um espectro NIR representativo. O dispositivo gira e move simultaneamente o frasco de amostra linearmente sobre a janela de medição do espectrofotômetro, permitindo que uma quantidade mais significativa de amostra seja acessada para produzir um espectro representativo.

3.3.3 Protocolo de medição das amostras

As amostras foram inseridas em frascos de vidro de 25 mL, selados e tampados. Os frascos foram totalmente preenchidos e acoplados ao dispositivo de leitura contendo o espectrofotômetro compacto NIR para aquisição dos espectros representativos. Os frascos de amostra foram escaneados por 20 s enquanto se moviam pelo dispositivo de leitura. Um frasco de referência contendo Spectralon®, que reflete 100% da radiação NIR, foi utilizado para registrar o sinal de referência utilizado para cálculo de absorvância. As amostras foram medidas em triplicatas para produzir um espectro médio representativo com uma relação sinal-ruído melhorada. Uma nova referência (sinal de fundo) foi adquirida a cada quinze varreduras.

3.3.4 Tratamento dos dados espectrais

O conjunto de dados espectrais, a PCA, a análise de correlação, os modelos MCR-ALS e PLS foram processados utilizando o software Unscrambler 11.0 (Aspen Technologies). Os espectros originais foram suavizados usando o algoritmo Savitzky-Golay, utilizando 11 pontos na janela de suavização e polinômio de 2° grau, e transformados pela variação normal padrão (SNV) [67] para minimizar os efeitos de espalhamento. Os dados foram centralizados para regressão PLS. O MCR-ALS empregou apenas concentrações não negativas como restrição e

os espectros dos edulcorantes comerciais autênticos (puros) para a estimativa inicial do algoritmo.

3.4 RESULTADOS E DISCUSSÃO

A Figura 3.1 mostra o conjunto de espectros NIR dos edulcorantes de referência (98 – 99% de pureza) antes e depois da transformação por suavização e SNV. O efeito da dispersão da radiação pelas partículas sólidas, como esperado, é significativamente reduzido pela SNV. Podem ser observadas diferenças entre as assinaturas espectrais dos edulcorantes mais comuns, principalmente entre os poliálcoois. A Figura S3.1 no material suplementar (Apêndice B) mostra o conjunto completo de dados espectrais de 103 amostras antes e depois da suavização e da transformação SNV.

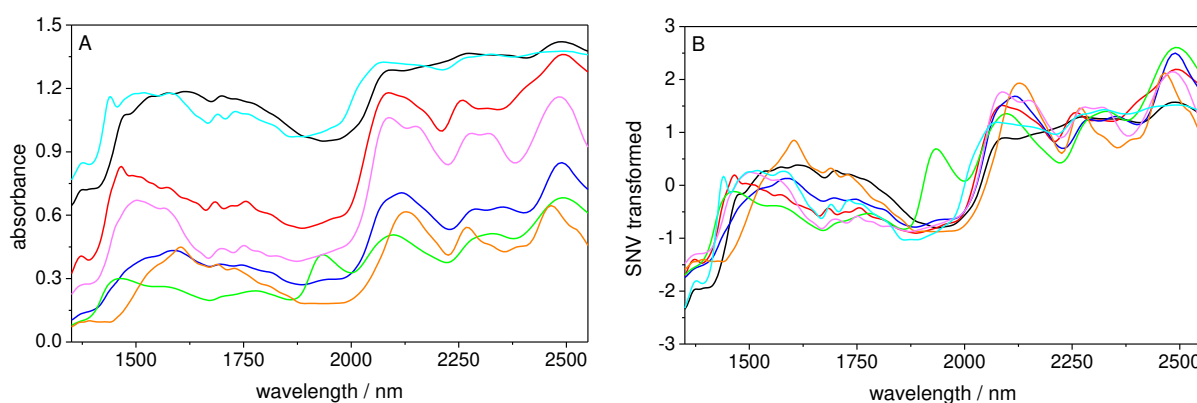


Figura 3.1. (A) espectros brutos dos edulcorantes de referência; (B) espectros dos edulcorantes de referência após suavização e transformação SNV. — xilitol, — frutose, — sorbitol, — maltodextrina, — eritritol, — maltitol, and — sacarose.

3.4.1 Abordagens multivariadas qualitativas

3.4.1.1 PCA e correlação espectral

A avaliação qualitativa do conjunto de dados foi iniciada pela observação das características espectrais das amostras, comparando visualmente os perfis espectrais com os das referências (padrões Sigma-Aldrich), realizando uma PCA preliminar e análise de correlação de todo o conjunto de espectros.

A Figura 3.2 mostra o gráfico dos *scores* da PCA para a primeira (PC1) e segunda (PC2) componentes principais e PC1 e PC3. Juntas, as três PCs capturam 94% da variância do conjunto de dados, validando interpretação a seguir.

Cinco clusters são observados no gráfico de PC1 x PC2. Eles foram identificados comparando a localização dos *scores* das substâncias de referência e utilizando os rótulos dos produtos comerciais. Aglomerados bem definidos são observados para os álcoois de carboidratos xilitol, eritritol, maltitol e sorbitol. A maltodextrina também está bem agrupada. Os *scores* da sacarose, frutose e maltitol estão próximos uns dos outros. Os *scores* do gráfico

PC1 x PC3 distinguem melhor entre amostras de sacarose e maltitol. No entanto, os *scores* da sacarose e frutose permanecem próximos.

Embora a identificação baseada na similaridade dos *scores* das amostras com os das substâncias de referência seja confiável, seus *scores* apresentam alguma diferença com os das amostras comerciais. Ao mesmo tempo, os *scores* de produtos comerciais autênticos apresentam dispersão variável. Por exemplo, a diferença é evidente para o eritritol, cujo *score* da referência se encontra deslocado do cluster dos produtos comerciais, e pode ser observada uma dispersão significativa dos *scores* dos produtos considerados autênticos causada principalmente em consequência de uma alteração no tamanho das partículas. Os valores de *scores* das amostras de xilitol mostram uma grande dispersão, provavelmente também em função do tamanho das partículas. No caso da maltodextrina, a variabilidade pode ser atribuída às diferentes quantidades de glicose, maltose e oligossacarídeos resultantes da hidrólise do amido de milho ou de mandioca empregada na fabricação. A variabilidade da composição da maltodextrina é avaliada pelo seu valor equivalente de dextrose (DE) que varia de 2 a 20 [60]. A maltodextrina de referência tem DE entre 16 - 17.

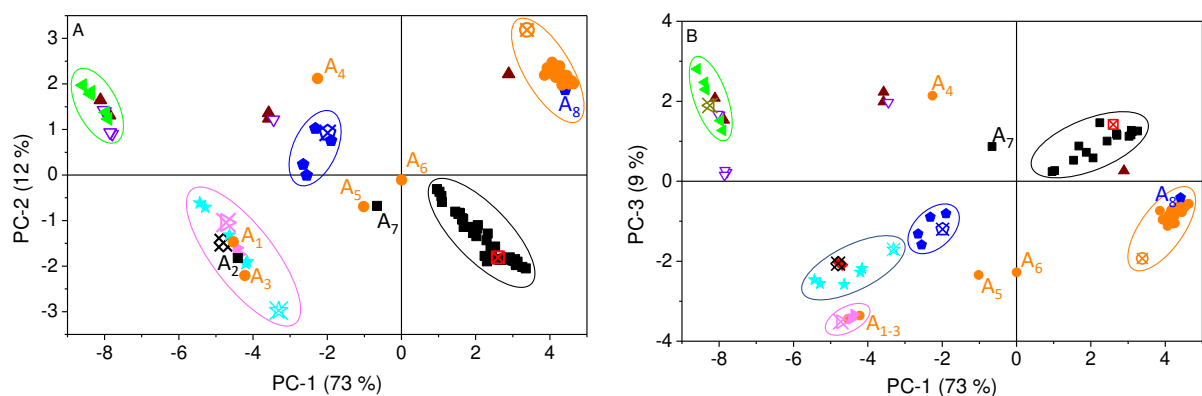


Figura 3.2. Distribuição dos *scores* da PCA obtida para todo o conjunto de espectros das amostras, incluindo os edulcorantes de referência. (A) PC1 x PC2; (B) PC1 x PC3. A₁ – A₈, possíveis amostras comerciais falsificadas. ◆ frutose, ⊠ frutose referência, ◆ sorbitol, ⊠ sorbitol referência, ★ sacarose, ⊠ sacarose referência, ▲ estévia, ▼ sucralose, ▲ maltodextrina, ⊠ maltodextrina referência, ▶ maltitol, ⊠ maltitol referência, ● eritritol, ⊠ eritritol referência, ■ xilitol, e ⊠ xilitol referência.

A identificação precisa de amostras autênticas de cada edulcorante presente no conjunto de dados é altamente relevante considerando o contexto do presente trabalho. A tentativa de classificar amostras sem o apoio de técnicas analíticas de referência requer a quantidade máxima de evidências baseadas exclusivamente em informações espectrais NIR. A informação fornecida pela PCA pode ser reforçada explorando a correlação entre os espectros pré-tratados das amostras com os espectros de referência de cada substância.

O primeiro passo no uso da correlação espectral para aumentar a evidência para a identificação de amostras autênticas é definir um limite para o coeficiente de correlação, acima do qual a amostra pode ser identificada como autêntica. Foram obtidos os coeficientes de correlação entre o espectro de referência testado e os espectros das demais referências. Os valores limites para os coeficientes de correlação (R_t) foram definidos por:

$$R_t = [(1 - R_h) / 2] + R_h \quad (\text{equação 1})$$

onde R_h é o coeficiente de correlação mais alto encontrado comparando o espectro de referência da substância testada quanto à autenticidade com os espectros de referência de todas as outras substâncias de interesse. Este limite tolera alguma variabilidade espectral nos espectros reais de amostras positivas e negativas, considerando fatores como a granulometria da amostra. Porém, o critério é conservador, evitando falsos positivos.

O espectro de uma amostra avaliada é classificado como substância autêntica se seu valor R_t for superior a 0,9899, 0,9815, 0,9779, 0,9616, 0,9899, 0,9620 e 0,9803, quando for verificada a autenticidade de uma amostra supostamente constituída por frutose, sorbitol, maltodextrina, eritritol, maltitol, xilitol ou sacarose, respectivamente. Os coeficientes de correlação fornecem um critério de classificação objetivo e mais direto que a PCA, empregam as variáveis reais e automatizam o procedimento de seleção, mantendo informações sobre a similaridade das características espectrais originais entre as amostras.

A Figura 3.3 mostra os coeficientes de correlação observados comparando o conjunto de espectros das substâncias de referência e a posição dos valores de R_t . Por exemplo, a Figura 3.4 mostra os coeficientes de correlação entre todos os espectros do conjunto de amostras e o espectro de referência do eritritol. As amostras identificadas como autênticas concordam com aquelas identificadas pela PCA. As amostras identificadas incorretamente em seus rótulos também são classificadas corretamente pelo critério do coeficiente de correlação.

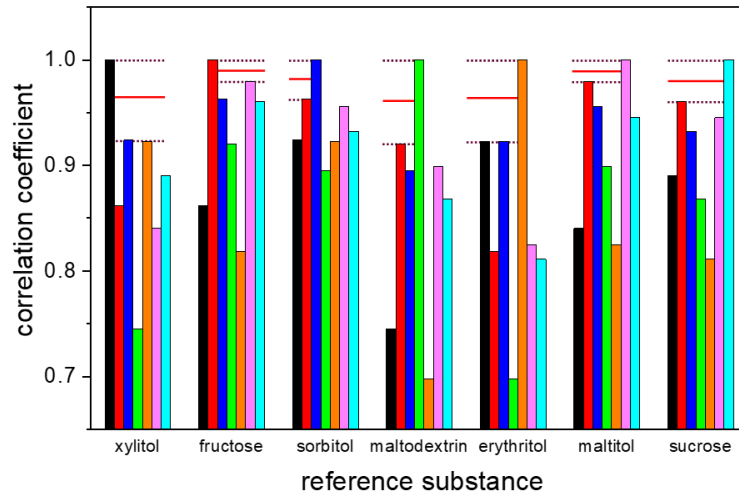


Figura 3.3. Coeficientes de correlação entre os espectros das substâncias de referência. | xilitol, | frutose, | sorbitol, | maltodextrina, | eritritol, | maltitol, | sacarose. As linhas sólidas vermelhas referem-se aos valores R_t utilizados para discriminar entre a substância de teste e os outros edulcorantes.

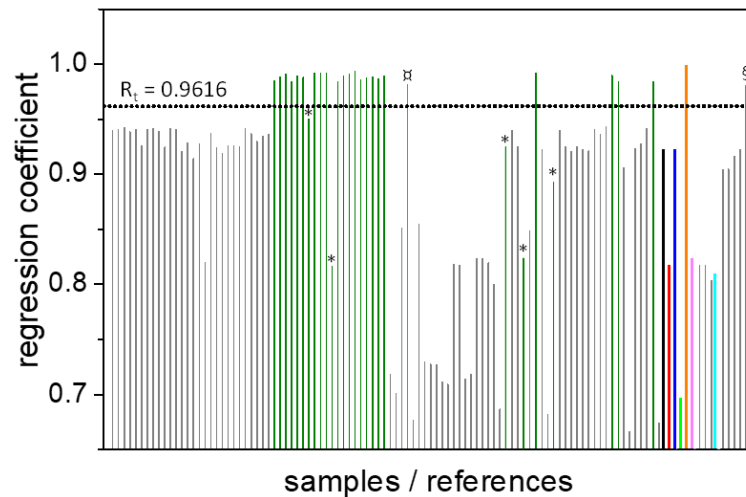


Figura 3.4. Coeficientes de correlação do conjunto de espectros das amostras comparados com o espectro de referência do eritritol. | outras amostras, | amostras rotuladas como eritritol, | xilitol, | frutose, | sorbitol, | maltodextrina, | eritritol, | maltitol, | sacarose. * Amostras rotuladas incorretamente como eritritol; x um adoçante de estévia, que utiliza eritritol como diluente; § uma amostra rotulada incorretamente como sorbitol, mas que é eritritol.

A Figura 3.5 mostra o conjunto de espectros de amostras comerciais de eritritol classificadas como autênticas com base na distância de seus respectivos *scores* para o da referência e do R_t . O espectro de referência também é exibido. Não são observadas diferenças entre as características espectrais das amostras e da referência. Diferenças nas intensidades relativas de absorção de amostras autênticas são observadas na faixa de 2.000 a 2.500 nm. Nesta região espectral, as características de absorção são atribuídas às combinações de modos vibracionais fundamentais. Em particular, o pico encontrado na faixa 2100 - 2250 nm, presente

em todos os álcoois de carboidratos e sacarose, é atribuído à combinação do estiramento de O – H com o estiramento da ligação C – O [68]. Após a transformação SNV dos espectros das amostras autênticas de eritritol, a variância restante também é evidente na faixa de combinação. Em contraste, a região espectral 1350 – 1950 nm mostra uma redução da intensidade das bandas de absorção.

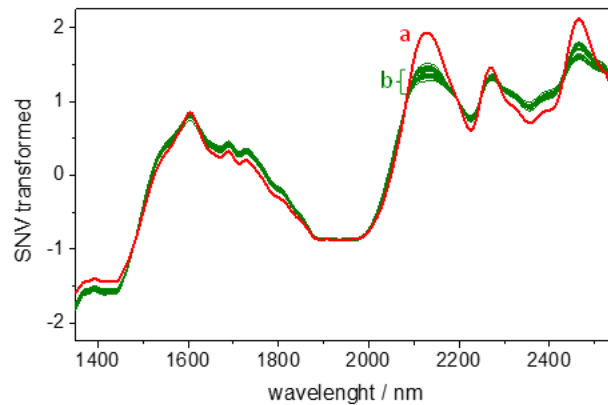


Figura 3.5. Conjunto de espectros de amostras comerciais de eritritol classificadas como autênticas usando PCA e análise de correlação, e o espectro de referência do eritritol. (a) eritritol referência; (b) amostras autênticas de eritritol.

A causa das diferenças observadas parece estar associada à granulometria e, talvez, à cristalinidade das amostras. As bandas de absorção em torno de 2.000 - 2.500 nm aumentam, enquanto as de todos os polióis e sacarose diminuem na faixa de 1.350 - 1.950 nm para amostras com tamanhos de partículas menores. A Figura 3.6 mostra o efeito da trituração de amostras comerciais autênticas de sacarose cristalina e eritritol em seus espectros NIR transformados por SNV, permitindo observar os efeitos descritos em função do tamanho das partículas.

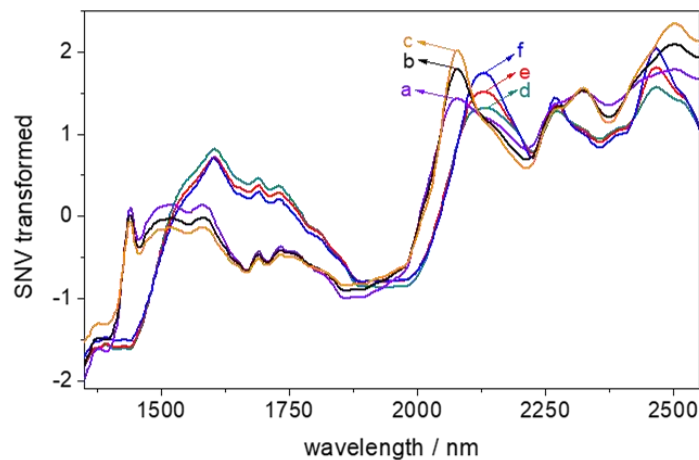


Figura 3.6. Comparação de espectros de amostras de sacarose e eritritol com vários tamanhos de partículas. — (a), — (b), e — (c), sacarose de granulometria grossa, média e fina, respectivamente; — (d), — (e), — (f), eritritol de granulometria grossa, média e fina, respectivamente.

Embora a transformação SNV minimize os efeitos de espalhamento, ela tem limitações. O aumento da absorção observado na região espectral de combinação em relação à região de sobretons pode ser parcialmente atribuído ao comportamento não linear do espalhamento da radiação em função do comprimento de onda. A radiação é espalhada mais intensamente em comprimentos de onda mais curtos para o mesmo tamanho médio de partícula. Portanto, a absorção é reduzida porque a radiação espalhada não consegue interagir para ser absorvida pelas partículas. Por outro lado, a radiação de comprimento de onda maior é menos espalhada e pode interagir com as partículas, aumentando a absorção na região de combinação NIR. Outros efeitos associados ao resultado experimental observado podem decorrer da redução da cristalinidade do material triturado, alterando a proporção entre as ligações químicas envolvidas na estrutura cristalina das partículas e aquelas em sua superfície.

A variabilidade inerente dos espectros NIR de edulcorantes comerciais autênticos requer um conjunto de dados abrangente, incluindo amostras com diferentes tamanhos de partículas. No entanto, a base de dados de substâncias autênticas pode ser estabelecida sem utilizar análises laboratoriais. Em vez disso, pode-se empregar PCA, exame visual de espectro e correlação espectral.

Os *scores* das amostras de sucralose e edulcorantes extraídos da planta estévia estão distribuídos na Figura 3.2A-B de acordo com as principais substâncias diluentes informadas em seus rótulos, devido ao maior índice de doçura desses edulcorantes (600 e 300 vezes maior que a sacarose para sucralose e estévia, respectivamente). O critério baseado no coeficiente de correlação também classifica estas amostras como autênticas quando o espectro de referência do diluente é testado em relação aos espectros da amostra. Os *scores* para três produtos de estévia e cinco amostras de sucralose podem ser encontrados próximos ao grupo da maltodextrina. Seus rótulos mencionam o uso dessa substância como diluente primário. O *score* de um produto de estévia está próximo do grupo de eritritol. Esta substância também é indicada como diluente no rótulo do produto. Os *scores* de duas amostras de estévia e uma sucralose são encontrados próximos uns dos outros, e distantes de qualquer cluster. A sucralose e um dos produtos de estévia especificam a lactose como diluente. O diluente da outra amostra de estévia não foi identificado, mas pode ser lactose devido à proximidade dos *scores* com as outras duas amostras.

A PCA e os coeficientes de correlação revelam oito amostras comerciais não conformes. Na Figura 3.2A-B, os *scores* de duas amostras rotuladas como eritritol (A_1 e A_3) e uma amostra rotulada como xilitol (A_2) são encontrados dentro do cluster atribuído ao maltitol; uma amostra

rotulada como sorbitol (A_8) é encontrada no grupo do eritritol autêntico. Os espectros dessas amostras são comparados aos dos produtos autênticos na Figura S3.2A-D (Apêndice B), ajudando a demonstrar a troca de rótulos.

As amostras rotuladas como eritritol ou xilitol (A_1 - A_3), cuja composição é maltitol, representam um sério risco à saúde do consumidor, pois o índice glicêmico do eritritol e do xilitol é 1 e 12, respectivamente, enquanto o do maltitol é 35. Além disso, o valor calórico do maltitol ($2,4 \text{ kcal g}^{-1}$) é superior ao do eritritol ($0,2 \text{ kcal g}^{-1}$) e igual ao do xilitol ($2,4 \text{ kcal g}^{-1}$).

Outras quatro amostras rotuladas como eritritol ou xilitol (A_4 - A_7) são encontradas distantes de quaisquer agrupamentos no gráfico PC1 x PC2, caracterizando potenciais adulterações. Essas amostras foram investigadas detalhadamente para identificar e quantificar os adulterantes.

O *score* do eritritol adulterado, rotulado como A_4 na Figura 3.2B, está próximo do agrupamento das amostras autênticas de sorbitol. Portanto, é razoável supor que o sorbitol foi empregado como adulterante. No entanto, o espectro desta amostra adulterada reflete o de uma mistura de maltodextrina e eritritol. Em particular, o pico em 1932 nm da maltodextrina, ausente no espectro do eritritol, está presente no espectro da amostra A_4 . Este fato aumenta a probabilidade da maltodextrina ser o adulterante suspeito. O custo médio da maltodextrina comercial é cerca de três vezes menor que o do eritritol, estimulando a adulteração com motivação econômica. Esse tipo de adulteração também apresenta risco ao consumidor devido ao alto índice glicêmico da maltodextrina. A Figura 3.7 compara os espectros de referência do eritritol e da maltodextrina com os da amostra A_4 .

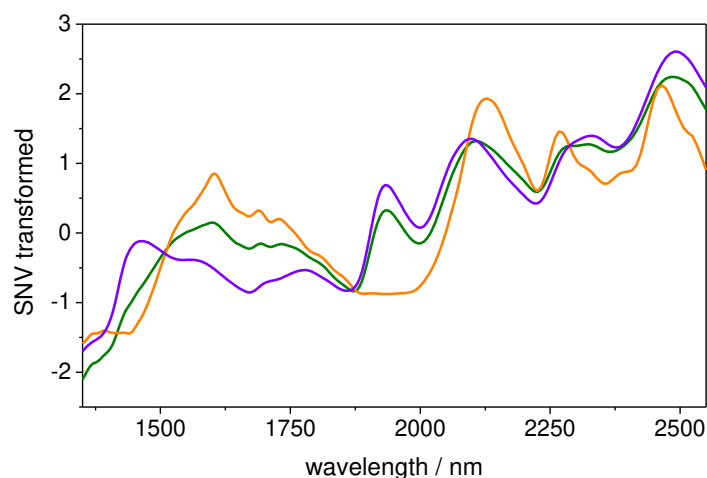


Figura 3.7. Espectros de referências e amostra falsificada A_4 . — eritritol referência, — maltodextrina referência, — amostra A_4 rotulada como eritritol.

Os *scores* das amostras adulteradas A₅ e A₆, rotuladas como eritritol, e A₇, rotulada como xilitol, mostram alguma semelhança com os clusters de sorbitol e maltitol, como pode ser observado nas Figuras 3.2A e 3.2B.

3.4.1.2 Projeção e SIMCA para classificação de edulcorantes comerciais à granel

Foi construído um modelo PCA empregando apenas as amostras previamente identificadas como autênticas e suas respectivas referências. As oito amostras não conformes foram projetadas neste modelo. A Figura 3.8A mostra a distribuição dos *scores* para PC1 e PC2. Podem ser observados os cinco clusters, quatro associados a edulcorantes autênticos. As localizações dos *scores* das amostras projetadas são semelhantes às aquelas mostradas na Figura 3.2A.

A Figura 3.8B mostra a localização dos *scores* projetados no gráfico PC2 x PC3. As mesmas conclusões sobre as características do conjunto de amostras projetado podem ser tiradas dos resultados da projeção, incluindo as indicações sobre falsificações e possíveis adulterantes.

A projeção é uma abordagem simples para rastrear não conformidades de amostras testadas. Requer apenas que as substâncias de referência verifiquem a autenticidade das amostras comerciais para definir o conjunto de amostras utilizado para construir o modelo PCA. As falsificações são prontamente detectadas e as informações sobre adulterantes podem ser antecipadas. No entanto, requer intervenção do analista para interpretar os resultados.

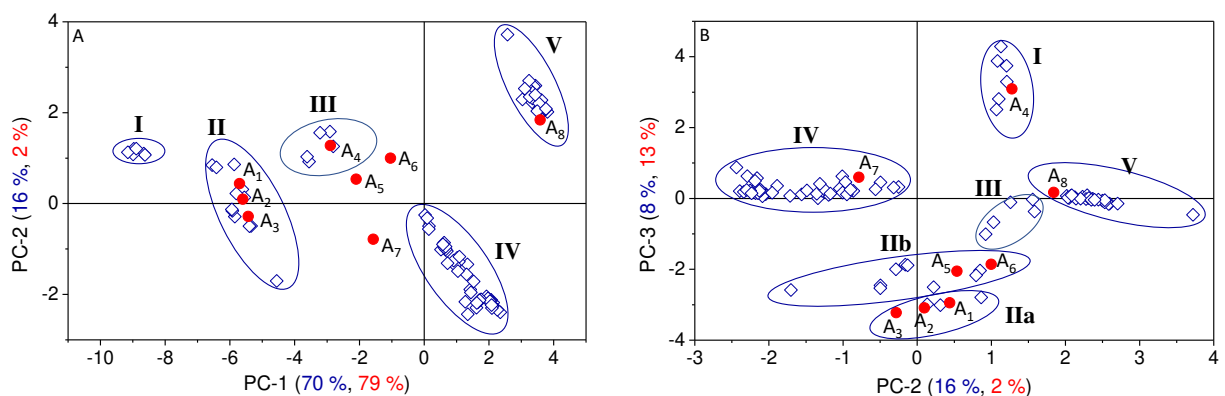


Figura 3.8. Projeção das amostras falsificadas no modelo PCA construído com edulcorantes autênticos. (A) PC1 x PC2; (B) PC2 x PC3. A₁ – A₈, amostras falsificadas. \diamond Amostras autênticas de edulcorantes. \bullet Amostras projetadas. I, maltodextrina; II, sacarose+frutose+maltitol; IIa, maltitol; IIb, sacarose+frutose; III, sorbitol; IV, xilitol; e V, eritritol.

Um modelo PCA adicional foi construído empregando os conjuntos de espectros das misturas binárias preparadas a partir de edulcorantes autênticos, melhorando a identificação de

possíveis adulterantes. A Figura 3.9A mostra o gráfico de *scores* para PC1 x PC2 para as misturas e amostras projetadas, e a Figura 3.9B mostra o mesmo para PC2 x PC3.

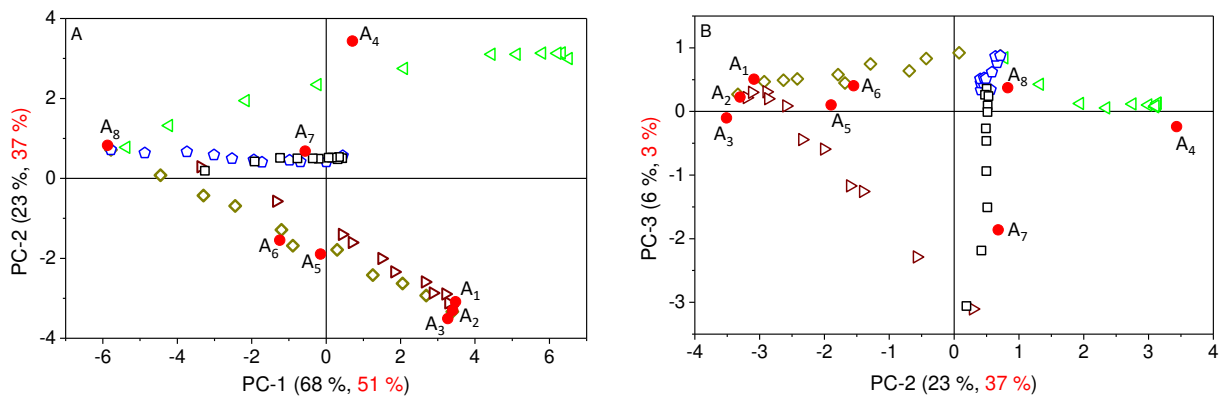


Figura 3.9. Projeção dos espectros de amostras falsificadas no modelo PCA construído com misturas autênticas de edulcorantes. A1 – A8, amostras falsificadas. (A) *scores* de PC1 x PC2; (B) *scores* de PC2 x PC3. \blacklozenge eritritol/maltitol; \blacktriangleleft eritritol/maltodextrina; \blacktriangleright eritritol/sorbitol; \blacktriangleright xilitol/maltitol; \blacksquare xilitol/sorbitol; \bullet amostras projetadas.

Este modelo PCA, que inclui a variabilidade espectral produzida por misturas binárias de edulcorantes com possíveis adulterantes, auxilia na identificação do adulterante em amostras falsificadas. Os *scores* das amostras A₁, A₂ e A₃ são encontrados próximos dos *scores* de 100% de maltitol. Os *scores* da amostra A₄ são encontrados próximos às misturas de eritritol/maltodextrina. Os *scores* das amostras A₅ e A₆ estão localizados próximos dos *scores* das misturas de maltitol/eritritol. Os *scores* da amostra A₇ situam-se próximos dos *scores* das misturas de sorbitol/xilitol. Por fim, os *scores* da amostra A₈ estão próximos de 100% (m/m) de eritritol.

A análise combinada das posições dos *scores* na dimensão reduzida das PCs melhora a identificação de adulterantes em edulcorantes falsificados. Embora somente as misturas mais prováveis tenham sido consideradas, o conjunto de dados utilizado para produzir o modelo PCA para as misturas preparadas pode ser expandido para incluir várias outras composições presumidas.

O conjunto de espectros de edulcorantes autênticos foi empregado para construir modelos de PCA individuais para cada classe de substância. Os conjuntos de amostras de eritritol e xilitol (os mais comuns e com maior número de amostras autênticas no conjunto de dados) foram divididos usando o algoritmo Kennard-Stone [69] em dois subconjuntos empregados na construção dos modelos PCA e para avaliar o desempenho preditivo para amostras externas positivas reais. As amostras dos outros edulcorantes estão em pequeno

número no conjunto de dados. Assim, todas foram utilizadas para construir os modelos PCA, e nenhum verdadeiro positivo foi utilizado para avaliar o desempenho da classificação SIMCA.

O algoritmo SIMCA foi empregado para classificar as amostras restantes e o conjunto externo de eritritol e xilitol. A Tabela 3.2 apresenta os valores da classificação das figuras de mérito [70].

Todos os falsos positivos (FP) detectados pelo SIMCA referem-se a adoçantes à base de sucralose ou estévia, identificados pela projeção PCA empregando a substância testada como diluente, ou a amostras rotuladas equivocadamente como uma substância quando são, na verdade, outro edulcorante (a substância testada pelo SIMCA). Conseqüentemente, a classificação incorreta dessas amostras pode ser facilmente identificada examinando seus rótulos e espectros. Se essas amostras forem removidas do conjunto de falsos positivos, todas as figuras de mérito tornam-se iguais a 100%.

Tabela 3.2. Figuras de mérito para a classificação de edulcorantes à granel usando SIMCA

Modelo PCA	Amostras modeladas	Outliers removidos	PCs	TP	TN	FP	FN	FPR (%)	FNR (%)	STR (%)	SPR (%)	EFR (%)
Eritritol	14	0	3	10	86	1	0	1.1	0	100	98.9	98.9
Xilitol	23	0	3	20	67	0	0	0	0	100	100	100
Maltodextrina	6	0	2	-	104	5	-	0	-	-	95.5	-
Sorbitol	5	0	3	-	105	0	-	0	-	-	100	-
Maltitol	3	0	1	-	107	2	-	0	-	-	98.2	-

TP: Total de amostras positivas reais; TN: total de amostras negativas reais; FN: número de falsos negativos; FP: número de falsos positivos; $FPR = [FP / (FP + TN)] \times 100$, taxa de falsos positivos; $FNR = [FN / (FN + TP)] \times 100$, taxa de falsos negativos; $STR = [TP / (TP + FN)] \times 100$, seletividade; $SPR = (TN / (TN + FP)) \times 100$, especificidade; $EFR = 100 - (FPR + FNR)$, eficiência.

Os conjuntos de dados espectrais de misturas binárias também foram classificados para avaliar a sensibilidade dos modelos de classificação quando são testados edulcorantes adulterados. Poucos erros de classificação ocorreram quando o conteúdo de uma substância é baixo na mistura binária, e seu espectro é confundido com o da substância pura devido ao comportamento espectral não linear, como será discutido a seguir. Ocorreram classificações erradas para três misturas de xilitol, contendo até 30% de maltitol, e uma mistura de eritritol, contendo 10% de maltitol, todas classificadas como maltitol; uma mistura de maltodextrina contendo 10% de eritritol, classificada como maltodextrina; duas misturas de sorbitol contendo 10 e 20% de xilitol, classificadas como sorbitol; e uma mistura de maltodextrina contendo 10% de eritritol, classificada como maltodextrina. Misturas de eritritol ou xilitol, dois dos edulcorantes mais consumidos, contendo 10% (m/m) ou mais de qualquer adulterante estudado, não foram classificadas como autênticas empregando SIMCA.

Deve-se tomar cuidado com possíveis erros causados por diluentes de adoçantes artificiais. Os edulcorantes à granel são utilizados como diluentes em altas concentrações e podem causar classificação incorreta desse tipo de amostra.

SIMCA é uma excelente alternativa para automatizar a classificação de edulcorantes, fornecendo figuras de mérito adequadas com poucos falsos positivos ou negativos reais e detectando a presença de adulterantes no nível mais baixo de 10% (m/m), investigado neste trabalho. Além disso, é capaz de gerar resultados rápidos, característica desejável para análises de triagem. Porém, reduz a quantidade de informações obtidas sobre as características das amostras. Por exemplo, se empregado sozinho, supostos adulterantes não podem ser antecipados.

3.4.2 Abordagens multivariadas quantitativas

3.4.2.1 Avaliação do MCR-ALS para estimar o teor dos adulterantes

Conforme mencionado, um dos objetivos deste trabalho é minimizar o uso de amostras de referência analisadas por métodos caros para triagem de amostras não autênticas ou adulteradas e estimar o conteúdo de adulterantes para fins de triagem. Uma abordagem potencial consiste em empregar MCR-ALS para avaliar o conteúdo relativo do adulterante, assumindo que o material falsificado consiste em uma mistura binária.

Várias misturas binárias descritas anteriormente foram estudadas. Suas composições foram escolhidas com base em resultados anteriores obtidos por análise qualitativa do conjunto de dados usando PCA, projeção e SIMCA e na probabilidade de utilização de substâncias de baixo custo em adulterações economicamente motivadas.

O procedimento analisa o conjunto de espectros de misturas binárias pelo MCR-ALS, usando os espectros das duas substâncias autênticas (“puras”) como estimativa inicial para o algoritmo. Os *scores* MCR-ALS, relacionados com a proporção de cada componente na mistura, são utilizados para ajustar uma curva analítica utilizando as concentrações reais das substâncias misturadas. Em seguida, os espectros das amostras identificadas como não conformes são incorporados um de cada vez no conjunto de espectros das misturas. O MCR-ALS é executado novamente e os *scores* MCR-ALS resultantes para as amostras não conformes são obtidos. Simultaneamente, o resíduo da amostra analisada é comparado com o resíduo médio das misturas preparadas. Quando o resíduo da amostra estiver mais próximo do resíduo médio das misturas, o adulterante é considerado presente e quantificado, convertendo o *score* MCR-ALS para concentração real por meio da curva analítica.

Todas as misturas binárias apresentam resposta espectral não linear em função da concentração do adulterante. As adições de até 40% do provável adulterante maltodextrina ao eritritol provocam ligeiras alterações nos espectros das misturas. As Figuras 3.10A e 3.10C-E mostram os conjuntos de espectros das misturas binárias.

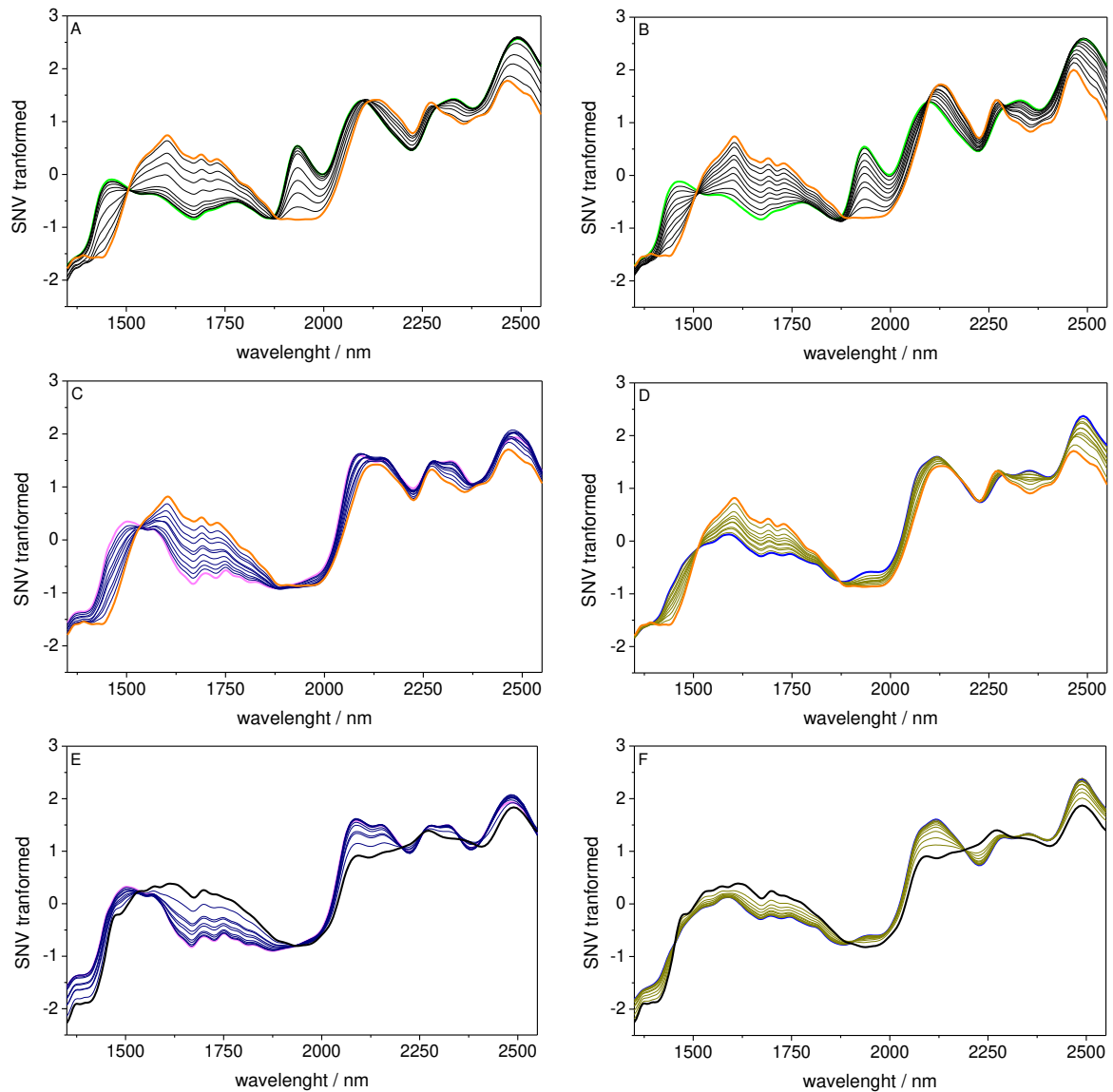


Figura 3.10. Conjunto de espectros para as misturas binárias e substâncias puras de: (A) — maltodextrina 100 %(m/m)/ — eritritol 100 %(m/m); (B) — maltodextrina 100 %(m/m) / — eritritol 100 %(m/m), amostras cominuídas; (C) — maltitol 100 %(m/m)/ — eritritol 100 %(m/m); (D) — sorbitol 100 %(m/m)/ — eritritol 100 %(m/m); (E) — maltitol 100 %(m/m) / — xilitol 100 %(m/m); (F) — sorbitol 100 %(m/m) / — xilitol 100 %(m/m).

O comportamento não linear observado é relevante porque a maioria dos algoritmos usuais de regressão quantitativa multivariada, como o PLS, assumem linearidade da resposta espectral com a concentração do analito avaliado.

A não linearidade da resposta espectral pode ser explicada pela interação física da mistura de pós de diferentes granulometrias e cristalinidade com a radiação NIR. Quando uma substância com pequeno tamanho de partícula está presente em alto teor em relação a outra de maior tamanho de partícula, as substâncias com pequeno tamanho de partícula podem interagir e cobrir as partículas do constituinte de menor concentração, impedindo sua interação com a radiação NIR. Além disso, partículas de tamanho pequeno contribuem mais para espalhar a radiação, que é refletida para o sistema de detecção do espectrofotômetro de forma especular, sem interagir com a amostra. Quando o conteúdo da substância com o tamanho de partícula menor é reduzido na mistura binária, ocorre o acesso da radiação às partículas do outro constituinte, e a resposta espectral tende à linearidade.

Da mesma forma, o comportamento não linear da resposta espectral de misturas binárias deve ser considerado quando estas são empregadas para avaliar o desempenho de métodos de classificação. A não linearidade observada nas misturas preparadas certamente ocorre também nas amostras comerciais possivelmente adulteradas. A semelhança dos espectros de substâncias não adulteradas com espectros de misturas contendo em alguns casos até 30% (m/m) do adulterante pode restringir o desempenho dos modelos de classificação, aumentando a probabilidade de falsos positivos.

Os resultados do SIMCA evidenciam a possibilidade de classificação incorreta devido à resposta espectral não linear das misturas. Porém, no presente caso, os edulcorantes mais consumidos não produziram nenhum falso positivo se o teor do adulterante for maior ou igual a 10% (m/m), o menor nível testado neste trabalho.

O comportamento dos *scores* MCR-ALS em função da concentração real assemelha-se ao comportamento não linear das respostas espectrais, como pode ser observado, por exemplo, na Figura 3.11A-F.

Os espectros NIR das misturas de maltodextrina/eritritol trituradas foram estudados para verificar o efeito da homogeneização do tamanho de partícula dos constituintes das misturas binárias. A Figura 3.10A-B permite comparar o efeito da cominuição nos espectros das misturas. Os *scores* do MCR-ALS alteram sua relação com as concentrações reais, apresentando comportamento quase linear em função do teor de maltodextrina nas misturas trituradas, contrastando com a correlação evidentemente não linear apresentada pelas misturas não trituradas. As Figuras 3.11A e 3.11B nos permitem observar a correlação entre os *scores* MCR-ALS com a concentração real para as misturas original e triturada de maltodextrina/eritritol.

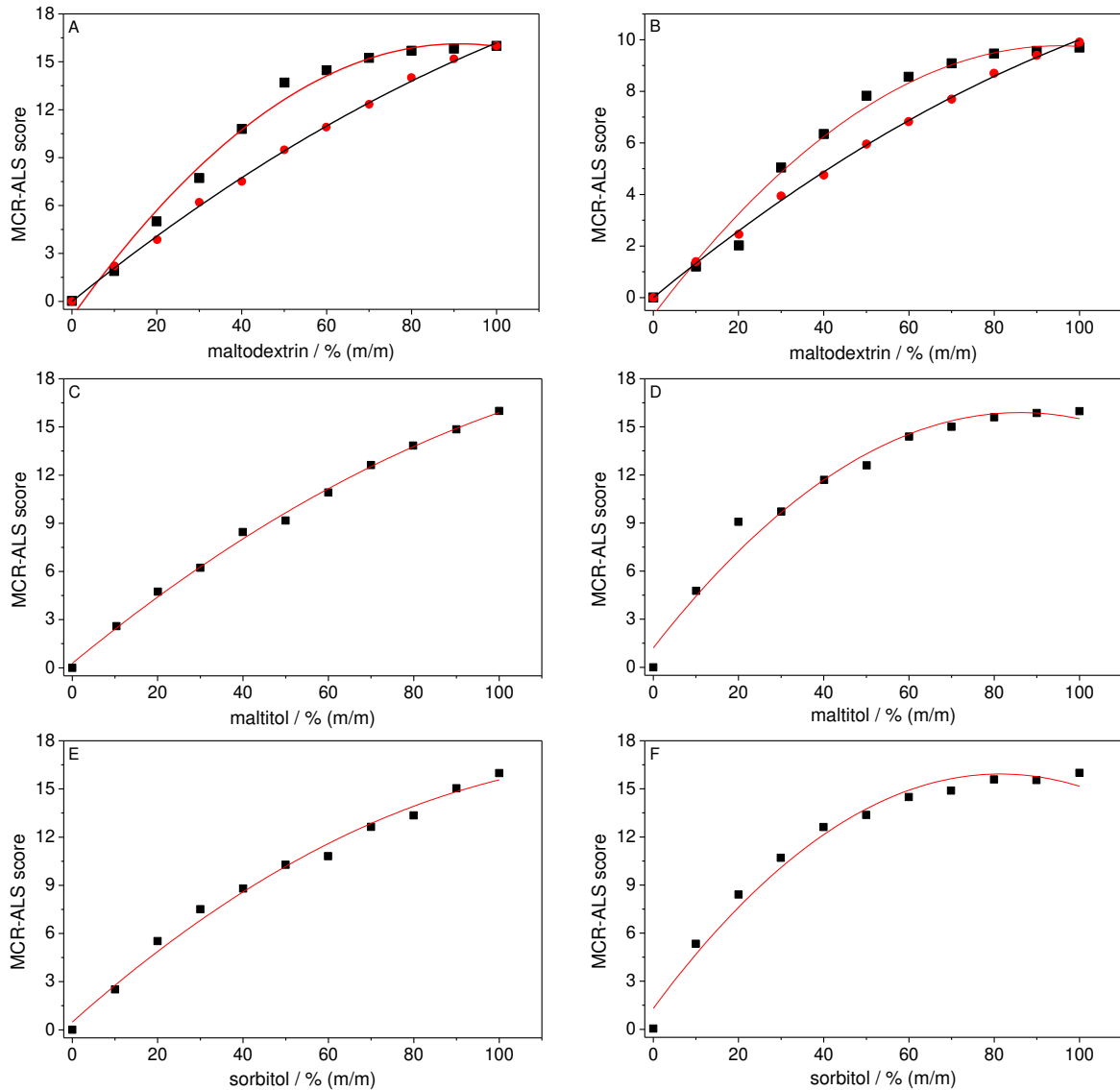


Figura 3.11. Relação entre os *scores* MCR-ALS e as concentrações reais das misturas binárias. (A) maltodextrina/eritritol (■) bruta e (●) triturada, misturas usando toda a faixa espectral; (B) maltodextrina/eritritol (■) bruta e (●) triturada, misturas usando a faixa espectral reduzida; (C) maltitol/eritritol misturas brutas; (D) maltitol/xilitol misturas brutas; (E) sorbitol/eritritol misturas brutas; (F) sorbitol/xilitol misturas brutas.

Este resultado é relevante para a aplicação da espectroscopia NIR de refletância para triar produtos constituídos por misturas de sólidos com diferentes granulometrias. Neste trabalho, optou-se por construir os modelos utilizando misturas e amostras originais (não trituradas) para detectar e avaliar amostras falsificadas na sua forma mais provável, próxima do que deve ser a prática mais comum dos adulteradores. Além disso, a cominuição retarda o método analítico, reduzindo a sua eficácia nas tarefas de inspeção. Por outro lado, a cominuição aumenta a sensibilidade espectral a concomitantes em menores quantidades. O desempenho da trituração para diferenciar edulcorantes à granel de edulcorantes à base de estévia ou sucralose será investigado no futuro.

O procedimento utilizado para rastrear a concentração de adulterantes requer a adoção de uma abordagem não linear para converter os *scores* MCR-ALS dos componentes da mistura em concentração real. Assim, os *scores* MCR-ALS são convertidos em concentração real ajustando um polinômio aos *scores* MCR e às concentrações reais. A curva analítica resultante é empregada para converter o *score* MCR da amostra de teste na concentração real do adulterante. Um polinômio de segundo grau ajustou-se bem a todos os casos estudados. A Tabela 3.3 mostra os resultados do ajuste.

Tabela 3.3. Resultados de ajuste para curvas analíticas não lineares relacionando os *scores* MCR-ALS e a concentração real dos adulterantes de edulcorantes comuns

Mistura adulterante/edulcorante	R ²	B ₀ (± SE)	B ₁ (± SE)	B ₂ (± SE)
sorbitol/xilitol	0.9759	1.31 (± 0.60)	0.359 (± 0.028)	-2.21E-03 (± 3.00E-04)
sorbitol/eritritol	0.9882	0.49 (± 0.42)	0.237 (± 0.020)	-8.62E-04 (± 1.89E-04)
maltodextrina/eritritol	0.9868	-0.97 (± 0.52)	0.375 (± 0.024)	-2.05E-03 (± 2.34E-04)
maltodextrina/eritritol-low	0.9778	-0.69 (± 0.41)	0.219 (± 0.019)	-1.15E-03 (± 1.85E-04)
maltitol/xilitol	0.9719	1.21 (± 0.66)	0.341 (± 0.031)	-1.98E-03 (± 2.94E-04)
maltitol/eritritol	0.9968	0.28 (± 0.23)	0.219 (± 0.011)	-6.27E-04 (± 1.01E-04)

SE = Erro padrão; equação da curva de calibração: $MCR-ALS\ score = B_0 + B_1x + B_2x^2$; x = concentração real do adulterante, % (m/m); R² = coeficiente de determinação.

Os espectros NIR das oito amostras falsificadas foram incorporados, um de cada vez, no conjunto de espectros das misturas binárias e substâncias puras, simulando vários níveis de adulteração, e o MCR-ALS foi executado para encontrar seus *scores*. Se o resíduo da amostra prevista for aceitável, os *scores* MCR-ALS são convertidos para concentração real % (m/m), usando a curva analítica pertinente. Os resíduos da amostra são empregados para avaliar a similaridade da amostra prevista com as amostras de calibração. Arbitrariamente, aceita-se a concentração estimada de amostras que apresentem resíduos iguais ou inferiores a dez vezes o resíduo médio calculado para as misturas e substâncias puras. A Figura 3.12 mostra dois exemplos de MCR-ALS para triagem do conteúdo adulterante, supondo a amostra falsificada constituída por eritritol adulterado com maltitol. Neste caso, foram previstas uma amostra rotulada como eritritol suspeita de conter maltitol (A₆) e outra que supunha conter eritritol e maltodextrina (A₄), conforme previsto pela PCA.

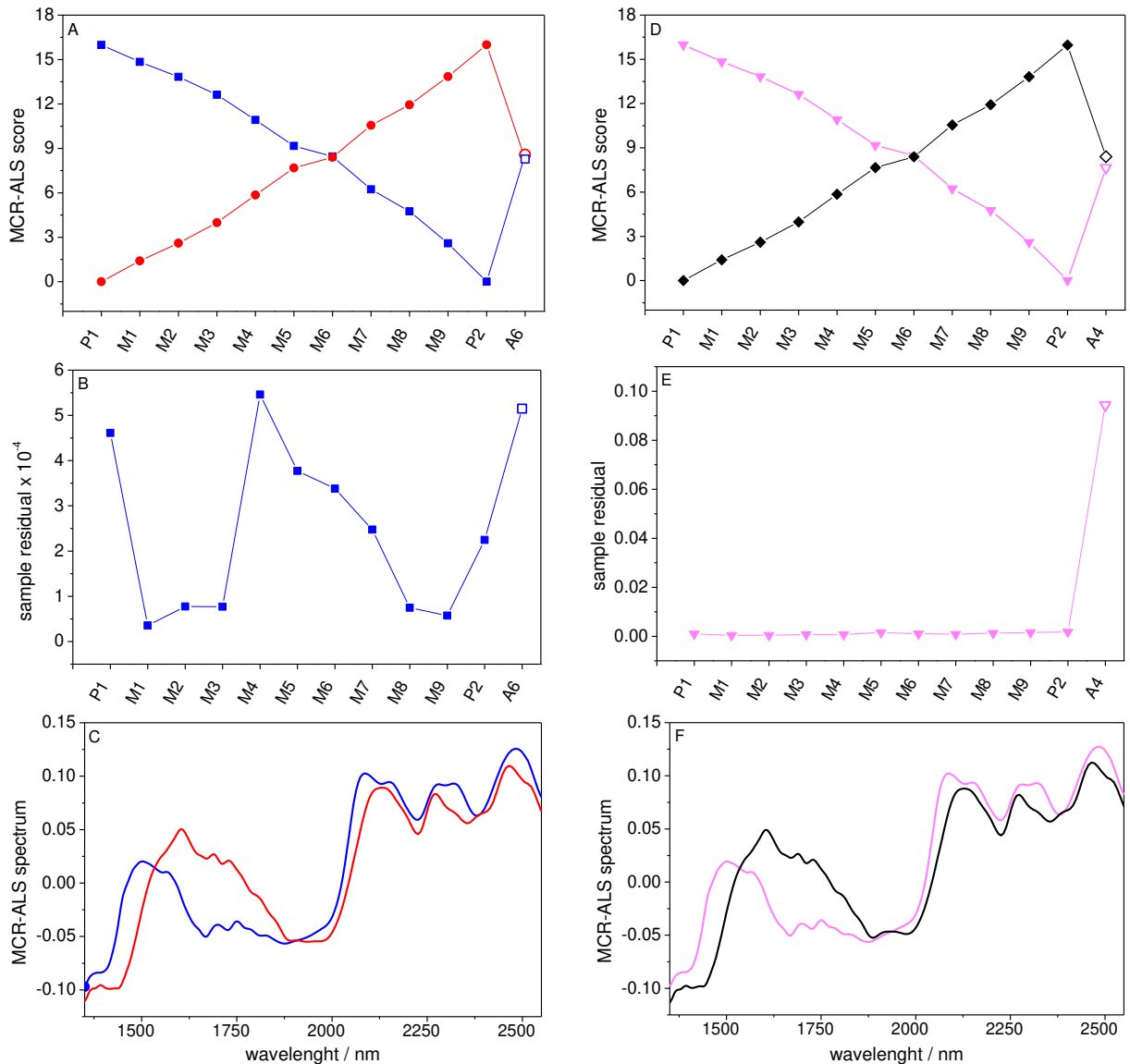


Figura 3.12. Resultados do MCR-ALS para análise das amostras falsificadas A₄ e A₆ utilizando o conjunto de espectros das misturas de maltitol/eritritol. (A) *scores* MCR-ALS para as misturas e amostra A₆; (B) resíduos para as misturas e amostra A₆. (D) *scores* MCR-ALS para as substâncias puras, misturas e amostra A₄; (E) resíduos para as substâncias puras, misturas e amostra A₄; (C) e (F), espectro de componentes MCR-ALS resultante após análise das amostras A₆ e A₄, respectivamente. —■—, —▽—, —, e — maltitol; —●—, —◆—, —, e — eritritol; □ amostra A₆; ▽ amostra A₄.

As concentrações estimadas de maltitol pelo MCR-ALS nas amostras falsificadas e os respectivos resíduos amostrais, utilizando o conjunto de dados espectrais das misturas maltitol/eritritol e maltodextrina/eritritol, são apresentados na Tabela 3.4.

Tabela 3.4. Resultados para quantificação de adulterantes em amostras falsificadas usando MCR-ALS

misturas	amostra	A1	A2	A3	A4	A5	A6	A7	A8
maltitol/eritritol	<i>score</i> MCR-ALS	15.96	16.00	15.98	7.636	10.40	8.578	8.076	0.000
	maltitol / % (m/m)	101	101	101	38	55	43	40	-1
	a = resíduo da amostra	1.43E-04	3.89E-04	2.40E-03	9.43E-02	1.28E-03	5.15E-04	5.25E-02	1.79E-03
	b = resíduo médio das misturas	2.28E-04	2.34E-04	2.85E-04	1.00E-03	2.36E-04	2.29E-04	6.66E-04	2.86E-04
	razão (a/b)	0.6	1.7	8.4	94	5.4	2.2	78	6.3
maltodextrina/eritritol	<i>score</i> MCR-ALS	10.86	10.68	10.41	9.410	6.386	5.030	6.660	0
	maltodextrina / % (m/m)	41	40	38	34	22	18	23	3
	a = resíduo da amostra	1.06E-01	1.14E-01	1.23E-01	1.14E-02	4.59E-02	3.20E-02	3.60E-02	8.69E-04
	b = resíduo médio das misturas	1.17E-03	1.22E-03	1.28E-03	4.23E-04	9.05E-04	8.29E-04	7.67E-04	3.45E-04
	razão (a/b)	91	93	96	27	51	38	47	2.5

Os resultados permitem estabelecer diversas conclusões sobre as amostras falsificadas. As amostras A₅ e A₆, que a PCA supostamente identificou como misturas de eritritol e maltitol, tiveram sua concentração de maltitol prevista em 55 e 43% (m/m), respectivamente, apresentando resíduos aceitáveis que corroboram a composição suspeita. As amostras rotuladas como eritritol ou xilitol (A₁, A₂ e A₃), mas que são, de fato, maltitol, conforme previamente identificado pela PCA, tiveram resíduos aceitáveis e suas concentrações estimadas de maltitol foram próximas a 100%. A amostra rotulada como sorbitol, que é eritritol (A₈), também foi prevista com resíduo aceitável, contendo cerca de 0% de maltitol. As amostras A₄ e A₇ não foram bem previstas, apresentando grandes resíduos. Portanto, a sua composição pode diferir das misturas binárias preparadas de maltitol ou maltodextrina com eritritol.

Alguns cuidados precisam ser considerados na utilização desta abordagem devido à variabilidade de tamanhos de partículas e composição de amostras comerciais e seu efeito na resposta espectral. Por exemplo, a amostra falsificada A₄ não pôde ser prevista com confiança aceitável usando MCR-ALS e os espectros das misturas de eritritol com maltodextrina, o adulterante inferido a partir da informação qualitativa da PCA.

O problema da maltodextrina é que os produtos comerciais são misturas com composições variáveis. Para este trabalho, foram preparadas misturas contendo apenas um tipo de maltodextrina, enquanto as adulterações podem ser feitas com maltodextrina com diferentes composições. A inspeção do conjunto de espectros de amostras autênticas de maltodextrina mostrou que a variabilidade dos espectros NIR de produtos comerciais é mais significativa na faixa de 1900 – 2500 nm, contendo características associadas a mudanças em sua composição [21], enquanto a região entre 1350 – 1880 nm mostra variabilidade reduzida. Assim, a amostra A₄ foi reanalisada, empregando apenas a faixa de comprimento de onda inferior. A concentração de maltodextrina foi estimada em 34% (m/m), comprovada por uma baixa razão entre os resíduos, igual a 3.

A composição da amostra falsificada A₇ rotulada como xilitol não pôde ser estabelecida de forma satisfatória utilizando todos os modelos construídos.

O MCR-ALS funciona muito bem na identificação de amostras com rótulos trocados. As substâncias autênticas reais foram previstas corretamente com teores próximos a 100% e baixos resíduos quando um dos constituintes das misturas binárias é a substância real presente na amostra testada.

3.4.2.2 Avaliação do PLS para a estimativa do teor dos adulterantes

Modelos de regressão baseados em PLS foram construídos para estimar a quantidade de adulterante em edulcorantes, empregando os conjuntos de dados das misturas. Os modelos PLS requerem 2 ou 3 fatores em vez de 1 fator esperado para modelar misturas binárias, para levar em conta a não linearidade observada entre a resposta espectral e a concentração. A Tabela 3.5 mostra os resultados da validação cruzada “*leave-one-out*”, obtidos para todos os modelos PLS.

O conteúdo de adulterantes das oito amostras falsificadas foi estimado por meio dos modelos. Os resultados da previsão podem ser observados na Tabela 3.6.

Intervalos de confiança (chamados “desvios”) dos valores previstos [28] estimados pelo programa Unscrambler podem ser usados da mesma forma que os resíduos do MCR-ALS para decidir se o modelo PLS adequado foi empregado e se os resultados previstos são confiáveis. Valores de desvio iguais ou inferiores a três vezes o RMSECV do respectivo modelo, expressos diretamente em unidades de % (m/m), podem ser associados a concentrações previstas de forma confiável. Todas as concentrações estimadas utilizando outros modelos PLS apresentaram resultados pouco confiáveis, fornecendo concentrações negativas, conteúdos muito superiores a 100% ou intervalos de confiança sempre maiores que três vezes o RMSECV dos modelos.

Tabela 3.5. Principais características dos modelos de regressão PLS utilizados para quantificar adulterantes em edulcorantes à granel em misturas binárias

Modelo PLS	R ²	RMSECV / %	Bias / %	Número de fatores PLS
maltitol+eritritol	0.998	1.5	-0.03	2
xilitol+maltitol	0.986	4.0	-0.97	2
sorbitol+xilitol	0.925	9.1	1.91	3
eritritol+sorbitol	0.994	2.7	0.08	2
eritritol+maltodextrina	0.965	6.5	-0.25	3

R² = coeficiente de determinação; RMSECV = erro quadrático médio de validação cruzada “*leave-one-out*”.

Os resultados mostram que a utilização do modelo PLS adequado baseado em misturas de eritritol e maltitol pode determinar com confiança aceitável a concentração de maltitol em seis amostras adulteradas. As amostras A₅ e A₆ foram previamente identificadas como provável eritritol adulterado com maltitol através de projeção nos modelos PCA de edulcorantes e misturas autênticas. Os conteúdos previstos de maltitol nestas amostras foram de 58 e 45% (m/m).

A₁, A₂ e A₃ são amostras rotuladas como eritritol ou xilitol que são, na verdade, maltitol autêntico, conforme identificado por PCA e SIMCA. Os teores previstos de maltitol destas amostras foram próximos de 100% (m/m), e os desvios são baixos para os dois modelos

construídos com misturas contendo este composto. A amostra A₈ é rotulada como sorbitol, mas é, na verdade, eritritol. Por conseguinte, o teor estimado de maltitol é próximo de zero, enquanto o teor de sorbitol é próximo de 100%.

Tabela 3.6. Resultados da concentração estimada de adulterantes em amostras falsificadas usando PLS

Modelo PLS										
mistura	maltitol/eritritol		maltitol/xilitol		sorbitol/xilitol		eritritol/sorbitol		eritritol/maltodextrina	
adulterante	maltitol	D*	maltitol	D*	sorbitol	D*	sorbitol	D*	maltodextrina	D*
% (m/m)										
A ₁	98	2 [†]	87	3 [†]	574	167	-7	49	83	126
A ₂	101	1 [†]	100	4 [†]	622	182	-73	52	99	130
A ₃	103	4	115	8 [†]	670	198	-74	55	117	136
A ₄	46	27	-29	51	73	109	-35	44	77	35
A ₅	58	3 [†]	42	16	233	82	11	33	60	91
A ₆	45	3 [†]	30	23	122	68	32	27	43	78
A ₇	49	20	12	29	207	88	-7	39	98	76
A ₈	1	4 [†]	4	45	199	139	104	10	22	17

*D = desvio; †, indica valores de desvio que atribuem confiabilidade às concentrações estimadas. A₁₋₈, amostras não conformes.

A amostra A₄ deveria ter sido bem prevista pelo modelo construído utilizando as misturas de maltodextrina/eritritol. No entanto, tal como observado quando o MCR-ALS foi utilizado, o teor de maltodextrina desta amostra não pôde ser estimado com confiança utilizando o modelo PLS baseado em toda a faixa espectral. Assim, um modelo PLS adicional foi construído utilizando a faixa espectral de 1350 a 1880 nm. O modelo empregou 3 fatores, apresentando RMSECV de 6,5 %, *bias* igual a 0,27 % (m/m) e coeficiente de determinação igual a 0,965. A concentração de maltodextrina prevista para a amostra A₄ foi de 30 % (m/m), e o desvio igual a 13% (m/m) foi considerado satisfatório.

O conteúdo da amostra A₇, rotulada como xilitol, não pôde ser previsto com confiança razoável por qualquer modelo PLS, indicando outro tipo de adulterante ou composição, conforme observado usando MCR-ALS.

O procedimento sugerido para estimar a concentração de adulterantes usando PLS em amostras falsificadas requer uma coleção de modelos de regressão construídos usando misturas binárias ou talvez terciárias de edulcorantes e possíveis adulterantes. No entanto, não são necessárias amostras analisadas por métodos analíticos de referência. Conforme exemplificado neste trabalho, o intervalo de confiança, que é calculado considerando os resíduos espectrais do espectro previsto [26], pode certificar os resultados estimados para fins de triagem.

3.4.2.3 Comparação entre MCR-ALS e PLS para fins quantitativos

Tanto o MCR-ALS quanto o PLS puderam estimar a concentração de adulterantes em amostras de edulcorantes não conformes e aumentar a confiança sobre a identidade do adulterante. Ambos fornecem meios para decidir a confiabilidade dos resultados, quando são utilizadas misturas de substâncias autênticas para construir os modelos, em vez de amostras reais previamente analisadas em quantidades significativas por métodos de referência dispendiosos, como é habitual.

Os conteúdos de maltitol nas amostras A₅ e A₆, rotuladas como eritritol, foram previstos como 55 e 43% (m/m) por MCR-ALS e 58 e 45% (m/m) usando PLS, respectivamente, indicando uma adulteração grave e concordância entre os dois métodos. A amostra A₄, também rotulada como eritritol, foi bem avaliada por PLS e MCR-ALS como contendo 30 e 34% de maltodextrina, respectivamente, utilizando a faixa de comprimento de onda mais baixa.

O MCR-ALS requer a inclusão dos espectros da amostra avaliada no conjunto de espectros de misturas antes da execução do algoritmo. O método é eficaz na previsão de constituintes puros em qualquer mistura binária. No entanto, o procedimento requer maior intervenção do analista, embora um software aplicativo possa ser desenvolvido para automatizar o processamento computacional.

O PLS estima as concentrações de forma automática e imediata, facilitando o trabalho do analista em comparação com o MCR-ALS. Entretanto, a utilização desta abordagem com misturas ternárias certamente trará dificuldades crescentes.

A principal desvantagem de ambas as abordagens é a falta de representatividade das amostras autênticas, utilizadas para preparar as misturas utilizadas para calibração do modelo. A incerteza da concentração prevista aumenta à medida que os espectros da amostra avaliada contêm substâncias com granulometria ou composição muito diferente daquela utilizada para preparar as misturas. No entanto, este erro pode ser aceito nas análises de triagem.

A caracterização qualitativa e quantitativa de amostras suspeitas é efetivamente alcançada usando PCA, análise de correlação, MCR-ALS ou PLS. Por outro lado, o método exige o envolvimento do analista para interpretar as informações extraídas pelos métodos multivariados.

O custo de implementação do método é substancialmente reduzido usando um espectrofotômetro NIR compacto e de baixo custo, poucas amostras de referências e evitando técnicas analíticas caras, comumente empregadas para construir modelos multivariados de classificação e regressão.

Várias não conformidades de edulcorantes comerciais à granel foram identificadas no presente estudo utilizando o método proposto. A troca de rótulos por motivos econômicos entre eritritol ou xilitol e maltitol, detectada em três amostras comerciais, e uma grave adulteração de uma amostra de eritritol com maltodextrina é preocupante devido à diferença significativa no índice glicêmico dessas substâncias.

A estratégia descrita neste trabalho pode ser aplicada a diversos outros produtos e matérias-primas comerciais, visando rastrear não conformidades e auxiliar nas tarefas de inspeção.

CONCLUSÕES GERAIS

O desenvolvimento destes três trabalhos evidenciou o grande potencial que a espectroscopia NIR e a análise multivariada de dados têm de extrair informações analíticas de amostras alimentícias e de drogas ilícitas.

O emprego de instrumentos miniaturizados e de baixo custo permite a implementação de métodos mais rápidos, portáteis e suficientemente bons para análises de triagem, apesar das limitações características desse tipo de instrumentação. Também foi possível comparar os desempenhos dos dois instrumentos empregados nos trabalhos. Mesmo possuindo características diferentes, esses dois instrumentos entregaram resultados satisfatórios na maioria dos casos, sendo o NeoSpectra um pouco melhor que o NanoNIR da Texas.

Na avaliação das características sensoriais de cafés torrados e moídos, os dois espectrofotômetros foram capazes de discriminar os cafés quanto à variedade do grão (uma única variedade arábica de blends de variedades) e à presença de cafeína. Também ficou evidente no desenvolvimento deste trabalho que o processamento industrial do café altera a composição química original dos grãos de café verde, de forma a modificar os espectros NIR, dificultando assim a obtenção de informações acerca da origem dos cafés através de dados espectrais. Entretanto, a origem etíope pode ser identificada de forma inequívoca, mesmo considerando o efeito do processamento do café. Modelos quantitativos baseados na regressão de mínimos quadrados parciais (PLS) foram construídos para determinar as características de intensidade, amargor, corpo, torra e acidez das amostras de cafés, utilizando os dois instrumentos.

No segundo capítulo, esses dois espectrofotômetros também foram utilizados para caracterizar quimicamente cocaína em amostras suspeitas. Ambos os instrumentos foram capazes de identificar a presença de cocaína, a sua forma de apresentação e de estimar o seu teor em amostras suspeitas, permitindo assim a proposta de um novo método de triagem de substâncias suspeitas no momento de apreensão.

E, no terceiro capítulo, uma ideia inovadora foi apresentada. Neste último trabalho, foi proposto combinar a espectroscopia NIR com a análise multivariada de dados, empregando apenas o NeoSpectra, para rastrear a autenticidade de produtos comerciais, sem a necessidade de utilizar técnicas de referência. Essa ideia foi aplicada a amostras de edulcorantes, vendidas à granel, e foram empregadas também algumas poucas amostras de referências. A caracterização qualitativa e quantitativa dessas amostras foi efetivamente alcançada usando PCA, análise de correlação, MCR-ALS ou PLS. Além disso, várias não conformidades de

edulcorantes comerciais à granel foram identificadas.

As principais dificuldades que puderam ser percebidas no desenvolvimento desses novos métodos são a obtenção de um conjunto amostral representativo, os desafios de se trabalhar diretamente com amostras sólidas (como a influência do tamanho de partícula na obtenção dos espectros NIR), a obtenção de amostras de referência ou a possibilidade de se analisar o conjunto amostral por outros métodos de referência e a grande variabilidade na composição das amostras de interesse. Todavia, foi possível propor métodos inovadores, não-destrutivos, portáteis, rápidos e de baixo custo para a triagem de amostras de café, cocaína e edulcorantes, utilizando instrumentos compactos e de baixo custo.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] C. Pasquini, “Near infrared spectroscopy: A mature analytical technique with new perspectives – A review,” *Anal Chim Acta*, vol. 1026, pp. 8–36, Oct. 2018, doi: 10.1016/j.aca.2018.04.004.
- [2] Letícia. P. Foli, M. C. Hespanhol, K. A. M. L. Cruz, and C. Pasquini, “Miniaturized Near-Infrared spectrophotometers in forensic analytical science – a critical review,” *Spectrochim Acta A Mol Biomol Spectrosc*, vol. 315, p. 124297, Jul. 2024, doi: 10.1016/j.saa.2024.124297.
- [3] T. Instruments and I. Dlpu, “DLP ® NIRscan™ Nano EVM User’s Guide User’s Guide,” 2015. [Online]. Available: www.ti.com
- [4] “NeoSpectra Micro-Development Kit Evaluation Brief Guide.”
- [5] D. F. Barbin, A. L. de S. M. Felicio, D.-W. Sun, S. L. Nixdorf, and E. Y. Hirooka, “Application of infrared spectral techniques on quality and compositional attributes of coffee: An overview,” *Food Research International*, vol. 61, pp. 23–32, Jul. 2014, doi: 10.1016/j.foodres.2014.01.005.
- [6] T. Ferreira, L. Galluzzi, T. de Paulis, and A. Farah, “Three centuries on the science of coffee authenticity control,” *Food Research International*, vol. 149, p. 110690, Nov. 2021, doi: 10.1016/j.foodres.2021.110690.
- [7] J. Medina, D. Caro Rodríguez, V. A. Arana, A. Bernal, P. Esseiva, and J. Wist, “Comparison of Attenuated Total Reflectance Mid-Infrared, Near Infrared, and ¹H-Nuclear Magnetic Resonance Spectroscopies for the Determination of Coffee’s Geographical Origin,” *Int J Anal Chem*, vol. 2017, pp. 1–8, Nov. 2017, doi: 10.1155/2017/7210463.
- [8] A. Giraud, S. Grassi, F. Savorani, G. Gavoci, E. Casiraghi, and F. Geobaldo, “Determination of the geographical origin of green coffee beans using NIR spectroscopy and multivariate data analysis,” *Food Control*, vol. 99, pp. 137–145, May 2019, doi: 10.1016/j.foodcont.2018.12.033.
- [9] F. Kurniawan, I. W. Budiastara, Sutrisno, and S. Widoyotomo, “Classification of Arabica Java Coffee Beans Based on Their Origin using NIR Spectroscopy,” *IOP Conf Ser Earth Environ Sci*, vol. 309, no. 1, p. 012006, Sep. 2019, doi: 10.1088/1755-1315/309/1/012006.
- [10] Yusmanizar and A. A. Munawar, “Environmental origin classification of coffee beans using infrared spectroscopy,” *IOP Conf Ser Earth Environ Sci*, vol. 922, no. 1, p. 012014, Nov. 2021, doi: 10.1088/1755-1315/922/1/012014.
- [11] I. Esteban-Díez, J. M. González-Sáiz, and C. Pizarro, “An evaluation of orthogonal signal correction methods for the characterisation of arabica and robusta coffee varieties by NIRS,” *Anal Chim Acta*, vol. 514, no. 1, pp. 57–67, Jun. 2004, doi: 10.1016/j.aca.2004.03.022.
- [12] I. Esteban-Díez, J. M. González-Sáiz, C. Sáenz-González, and C. Pizarro, “Coffee varietal differentiation based on near infrared spectroscopy,” *Talanta*, vol. 71, no. 1, pp. 221–229, Jan. 2007, doi: 10.1016/j.talanta.2006.03.052.

- [13] I. Esteban-Díez, J. M. González-Sáiz, and C. Pizarro, "Prediction of sensory properties of espresso from roasted coffee samples by near-infrared spectroscopy," *Anal Chim Acta*, vol. 525, no. 2, pp. 171–182, Nov. 2004, doi: 10.1016/j.aca.2004.08.057.
- [14] K. M. Santos, M. F. V. Moura, F. G. Azevedo, K. M. G. Lima, I. M. Raimundo, and C. Pasquini, "Classification of Brazilian Coffee Using Near-Infrared Spectroscopy and Multivariate Calibration," *Anal Lett*, vol. 45, no. 7, pp. 774–781, May 2012, doi: 10.1080/00032719.2011.653905.
- [15] M. R. Baqueta, A. Coqueiro, and P. Valderrama, "Brazilian Coffee Blends: A Simple and Fast Method by Near-Infrared Spectroscopy for the Determination of the Sensory Attributes Elicited in Professional Coffee Cupping," *J Food Sci*, vol. 84, no. 6, pp. 1247–1255, Jun. 2019, doi: 10.1111/1750-3841.14617.
- [16] I. Esteban-Díez, J. M. González-Sáiz, and C. Pizarro, "Prediction of Roasting Colour and other Quality Parameters of Roasted Coffee Samples by near Infrared Spectroscopy. A Feasibility Study," *J Near Infrared Spectrosc*, vol. 12, no. 5, pp. 287–297, Oct. 2004, doi: 10.1255/jnirs.437.
- [17] J. R. Santos, M. C. Sarraguça, A. O. S. S. Rangel, and J. A. Lopes, "Evaluation of green coffee beans quality using near infrared spectroscopy: A quantitative approach," *Food Chem*, vol. 135, no. 3, pp. 1828–1835, Dec. 2012, doi: 10.1016/j.foodchem.2012.06.059.
- [18] J. K. Winkler-Moser *et al.*, "Detection of Corn Adulteration in Brazilian Coffee (*Coffea arabica*) by Tocopherol Profiling and Near-Infrared (NIR) Spectroscopy," *J Agric Food Chem*, vol. 63, no. 49, pp. 10662–10668, Dec. 2015, doi: 10.1021/acs.jafc.5b04777.
- [19] D. A. P. Forchetti and R. J. Poppi, "Detection and Quantification of Adulterants in Roasted and Ground Coffee by NIR Hyperspectral Imaging and Multivariate Curve Resolution," *Food Anal Methods*, vol. 13, no. 1, pp. 44–49, Jan. 2020, doi: 10.1007/s12161-019-01502-x.
- [20] P. C. Ayu, I. W. Budiastara, and A. Rindang, "NIR spectroscopy application for determination caffeine content of Arabica green bean coffee," *IOP Conf Ser Earth Environ Sci*, vol. 454, no. 1, p. 012049, Feb. 2020, doi: 10.1088/1755-1315/454/1/012049.
- [21] L. Alessandrini, S. Romani, G. Pinnavaia, and M. D. Rosa, "Near infrared spectroscopy: An analytical tool to predict coffee roasting degree," *Anal Chim Acta*, vol. 625, no. 1, pp. 95–102, Sep. 2008, doi: 10.1016/j.aca.2008.07.013.
- [22] F. de C. Pires, R. G. F. A. Pereira, M. R. Baqueta, P. Valderrama, and R. Alves da Rocha, "Near-infrared spectroscopy and multivariate calibration as an alternative to the Agtron to predict roasting degrees in coffee beans and ground coffees," *Food Chem*, vol. 365, p. 130471, Dec. 2021, doi: 10.1016/j.foodchem.2021.130471.
- [23] R. M. Correia *et al.*, "Portable near infrared spectroscopy applied to quality control of Brazilian coffee," *Talanta*, vol. 176, pp. 59–68, Jan. 2018, doi: 10.1016/j.talanta.2017.08.009.
- [24] M. R. Baqueta, A. Coqueiro, P. H. Março, and P. Valderrama, "Quality Control Parameters in the Roasted Coffee Industry: a Proposal by Using MicroNIR Spectroscopy and Multivariate Calibration," *Food Anal Methods*, vol. 13, no. 1, pp. 50–60, Jan. 2020, doi: 10.1007/s12161-019-01503-w.

- [25] N. Yergenson and D. E. Aston, "Online determination of coffee roast degree toward controlling acidity," *J Near Infrared Spectrosc*, vol. 28, no. 4, pp. 175–185, Aug. 2020, doi: 10.1177/0967033520924493.
- [26] M. C. Hespanhol, J. C. Souza, and C. Pasquini, "Feasibility of a portable, low-cost near-infrared spectrophotometer for the quality screening of omega-3 dietary supplements," *J Pharm Biomed Anal*, vol. 189, p. 113436, Sep. 2020, doi: 10.1016/j.jpba.2020.113436.
- [27] C. Pasquini and M. C. Hespanhol, "A rotational-linear sample probing device to improve the performance of compact near-infrared spectrophotometers," *Microchemical Journal*, vol. 170, p. 106747, Nov. 2021, doi: 10.1016/j.microc.2021.106747.
- [28] K. H. Esbensen and B. Swarbrick, *Multivariate data analysis: an introduction to multivariate analysis, process analytical technology and quality by design*, 6^a. Camo, 2018.
- [29] J. Grabska, K. B. Beć, Y. Ozaki, and C. W. Huck, "Anharmonic DFT Study of Near-Infrared Spectra of Caffeine: Vibrational Analysis of the Second Overtones and Ternary Combinations," *Molecules*, vol. 26, no. 17, p. 5212, Aug. 2021, doi: 10.3390/molecules26175212.
- [30] "CADERNOS ANP A CRIMINALÍSTICA DO SÉCULO XXI E ANÁLISE DE DROGAS E INTELIGÊNCIA," 2008.
- [31] N. V. S. Rodrigues, "Caracterização química e análise quimiométrica de amostras de cocaína apreendidas em Minas Gerais pelo Departamento de Polícia Federal," 2010, *Universidade Federal de Minas Gerais, Belo Horizonte*.
- [32] UNITED NATIONS OFFICE ON DRUGS AND CRIME & DIVISION FOR TREATY AFFAIRS., *WORLD DRUG REPORT 2022*. UNITED NATIONS, 2022.
- [33] C. Pasquini, "Near Infrared Spectroscopy: fundamentals, practical aspects and analytical applications," *J Braz Chem Soc*, vol. 14, no. 2, pp. 198–219, Apr. 2003, doi: 10.1590/S0103-50532003000200006.
- [34] "Revisão de média, mediana e moda," <https://pt.khanacademy.org/math/statistics-probability/summarizing-quantitative-data/mean-median-basics/a/mean-median-and-mode-review>.
- [35] J. C. Souza, C. Pasquini, and M. C. Hespanhol, "Feasibility of compact near-infrared spectrophotometers and multivariate data analysis to assess roasted ground coffee traits," *Food Control*, vol. 138, p. 109041, Aug. 2022, doi: 10.1016/j.foodcont.2022.109041.
- [36] J. C. Souza, M. C. Hespanhol, F. R. P. Rocha, and C. Pasquini, "Exploiting chemometrics for screening analysis by near-infrared spectroscopy without the support of reference analytical methods: Application to bulk sweeteners," *Microchemical Journal*, vol. 190, p. 108732, Jul. 2023, doi: 10.1016/j.microc.2023.108732.
- [37] "Catálogo de produtos flowsupply." Accessed: Aug. 01, 2023. [Online]. Available: <http://www.flowsupply.com.br/brasil/>
- [38] G. T. Georgiev and J. J. Butler, "Long-term calibration monitoring of Spectralon diffusers BRDF in the air-ultraviolet," *Appl Opt*, vol. 46, no. 32, p. 7892, Nov. 2007, doi: 10.1364/AO.46.007892.

- [39] A. O. Maldaner *et al.*, “Chemical Profiling of Street Cocaine from Different Brazilian Regions,” *J Braz Chem Soc*, 2015, doi: 10.5935/0103-5053.20150321.
- [40] E. W. Ciurczak, B. Igne, J. Workman, and D. A. Burns, “Handbook of Near- Infrared Analysis.” [Online]. Available: <https://www.routledge.com/Practical->
- [41] P. Geladi and B. R. Kowalski, “Partial least-squares regression: a tutorial,” *Anal Chim Acta*, vol. 185, pp. 1–17, 1986, doi: 10.1016/0003-2670(86)80028-9.
- [42] S. H. F. Scafi and C. Pasquini, “Identification of counterfeit drugs using near-infrared spectroscopy,” *Analyst*, vol. 126, no. 12, pp. 2218–2224, Dec. 2001, doi: 10.1039/b106744n.
- [43] A. A. Bunaciu and H. Y. Aboul-Enein, “Adulterated drug analysis using FTIR spectroscopy,” *Appl Spectrosc Rev*, vol. 56, no. 5, pp. 423–437, May 2021, doi: 10.1080/05704928.2020.1811717.
- [44] K. Dégardin, A. Guillemain, N. V. Guerreiro, and Y. Roggo, “Near infrared spectroscopy for counterfeit detection using a large database of pharmaceutical tablets,” *J Pharm Biomed Anal*, vol. 128, pp. 89–97, Sep. 2016, doi: 10.1016/j.jpba.2016.05.004.
- [45] R. Deidda *et al.*, “Vibrational spectroscopy in analysis of pharmaceuticals: Critical review of innovative portable and handheld NIR and Raman spectrophotometers,” *TrAC Trends in Analytical Chemistry*, vol. 114, pp. 251–259, May 2019, doi: 10.1016/j.trac.2019.02.035.
- [46] J. Mendez, L. Mendoza, J. P. Cruz-Tirado, R. Quevedo, and R. Siche, “Trends in application of NIR and hyperspectral imaging for food authentication,” *Scientia Agropecuaria*, vol. 10, no. 1, pp. 143–161, Mar. 2019, doi: 10.17268/sci.agropecu.2019.01.16.
- [47] M. De Géa Neves, R. J. Poppi, and M. C. Breitzkreitz, “Authentication of plant-based protein powders and classification of adulterants as whey, soy protein, and wheat using FT-NIR in tandem with OC-PLS and PLS-DA models,” *Food Control*, vol. 132, p. 108489, Feb. 2022, doi: 10.1016/j.foodcont.2021.108489.
- [48] M. Manley and V. Baeten, “Spectroscopic Technique: Near Infrared (NIR) Spectroscopy,” in *Modern Techniques for Food Authentication*, Elsevier, 2018, pp. 51–102. doi: 10.1016/B978-0-12-814264-6.00003-7.
- [49] K. Norris and P. Williams, *Near-infrared technology: in the agricultural and food industries*, 2 nd. St. Paul: American Association of Cereal Chemists, 2001.
- [50] A. Saraiva, C. Carrascosa, D. Raheem, F. Ramos, and A. Raposo, “Maltitol: Analytical Determination Methods, Applications in the Food Industry, Metabolism and Health Impacts,” *Int J Environ Res Public Health*, vol. 17, no. 14, p. 5227, Jul. 2020, doi: 10.3390/ijerph17145227.
- [51] Y.-H. Wang, B. Avula, W. Tang, M. Wang, M. A. Elsohly, and I. A. Khan, “Ultra-HPLC method for quality and adulterant assessment of steviol glycosides sweeteners - Stevia rebaudiana and stevia products,” *Food additives & contaminants. Part A, Chemistry, analysis, control, exposure & risk assessment*, vol. 32, no. 5, p. 674–685, 2015, doi: 10.1080/19440049.2015.1021863.

- [52] W. Yan, N. Wang, P. Zhang, J. Zhang, S. Wu, and Y. Zhu, "Simultaneous determination of sucralose and related compounds by high-performance liquid chromatography with evaporative light scattering detection," *Food Chem*, vol. 204, pp. 358–364, Aug. 2016, doi: 10.1016/j.foodchem.2016.02.099.
- [53] A. G. Coelho and D. P. de Jesus, "A simple method for determination of erythritol, maltitol, xylitol, and sorbitol in sugar-free chocolates by capillary electrophoresis with capacitively coupled contactless conductivity detection," *Electrophoresis*, vol. 37, no. 22, pp. 2986–2991, Nov. 2016, doi: 10.1002/elps.201600263.
- [54] C. Shin, Y. Lim, E. Yoon, and S.-H. Kim, "Detection of adulteration in commercial galactooligosaccharide by flow-injection electrospray ionization tandem mass spectrometry," *Food Control*, vol. 36, no. 1, pp. 280–286, Feb. 2014, doi: 10.1016/j.foodcont.2013.08.039.
- [55] M. E. Auer, U. J. Griesser, and J. Sawatzki, "Qualitative and quantitative study of polymorphic forms in drug formulations by near infrared FT-Raman spectroscopy," *J Mol Struct*, vol. 661–662, pp. 307–317, Dec. 2003, doi: 10.1016/j.molstruc.2003.09.002.
- [56] P. Vargas Jentsch, S. Torrico-Vallejos, S. Mendieta-Brito, L. A. Ramos, and V. Ciobotă, "Detection of counterfeit stevia products using a handheld Raman spectrometer," *Vib Spectrosc*, vol. 83, pp. 126–131, Mar. 2016, doi: 10.1016/j.vibspec.2016.01.015.
- [57] S. R. Karunathilaka *et al.*, "Quantitation of Saccharin and Cyclamate in Tabletop Formulations by Portable Raman and NIR Spectrometers in Combination with Partial Least Squares Regression," *Food Anal Methods*, vol. 11, no. 4, pp. 969–979, Apr. 2018, doi: 10.1007/s12161-017-1057-1.
- [58] Y. Martono, S. Trihandaru, and F. S. Rondonuwu, "Determination of Stevioside and Rebaudioside A in <i>Stevia rebaudiana</i> Bertoni Leaves Using near Infrared Spectroscopy and Multivariate Data Analysis," *Indonesian Journal of Chemistry*, vol. 18, no. 4, p. 664, Nov. 2018, doi: 10.22146/ijc.25580.
- [59] A. Ortega-Platas, V. Corro-Herrera, M. G. Aguilar-Uscanga, and J. Gómez-Rodríguez, "Real-time monitoring of xylitol production in a bioreactor by *Candida tropicalis* IEC5-ITV using Near-Infrared Spectroscopy (NIRS)," *Rev Mex Ing Quim*, vol. 20, no. 3, pp. 1–10, Jul. 2021, doi: 10.24275/rmiq/Bio2420.
- [60] E. Storz and K. Steffens, "Feasibility Study for Determination of the Dextrose Equivalent (DE) of Starch Hydrolysis Products with Near-Infrared Spectroscopy (NIRS)," *Starch - Stärke*, vol. 56, no. 2, pp. 58–62, Feb. 2004, doi: 10.1002/star.200300220.
- [61] C. R. Muzzio, N. G. Dini, and L. D. Simionato, "Determination of moisture content in lyophilized mannitol through intact glass vials using NIR micro-spectrometers," *Brazilian Journal of Pharmaceutical Sciences*, vol. 47, no. 2, pp. 289–297, Jun. 2011, doi: 10.1590/S1984-82502011000200010.
- [62] N. Hirai, T. Takatani-Nakase, and K. Takahashi, "Application of Near-Infrared Spectrometry to Evaluate the Mechanism of Wet Granulation Using a High-Speed Mixer with Porous Calcium Silicate and Sugar Alcohols," *Chem Pharm Bull (Tokyo)*, vol. 66, no. 11, pp. 1027–1034, Nov. 2018, doi: 10.1248/cpb.c18-00474.

- [63] B. GUVEN, S. DURAKLI-VELİOĞLU, and İ. H. BOYACI, “RAPID IDENTIFICATION OF SOME SWEETENERS AND SUGARS BY ATTENUATED TOTAL REFLECTANCE-FOURIER TRANSFORM INFRARED (ATR-FTIR), NEAR-INFRARED (NIR) AND RAMAN SPECTROSCOPY,” *Gıda*, vol. 44, no. 2, pp. 274–290, Apr. 2019, doi: 10.15237/gida.GD18119.
- [64] K. B. Beć, J. Grabska, and C. W. Huck, “Principles and Applications of Miniaturized Near-Infrared (NIR) Spectrometers,” *Chemistry – A European Journal*, vol. 27, no. 5, pp. 1514–1532, Jan. 2021, doi: 10.1002/chem.202002838.
- [65] C. Ruckebusch and L. Blanchet, “Multivariate curve resolution: A review of advanced and tailored applications and challenges,” *Anal Chim Acta*, vol. 765, pp. 28–36, Feb. 2013, doi: 10.1016/j.aca.2012.12.028.
- [66] A. Malik, A. de Juan, and R. Tauler, “Multivariate Curve Resolution: A Different Way To Examine Chemical Data,” 2015, pp. 95–128. doi: 10.1021/bk-2015-1199.ch005.
- [67] R. J. Barnes, M. S. Dhanoa, and S. J. Lister, “Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra,” *Appl Spectrosc*, vol. 43, no. 5, pp. 772–777, Jul. 1989, doi: 10.1366/0003702894202201.
- [68] Jr. , J. Workman and L. Weyer, *Practical Guide to Interpretive Near-Infrared Spectroscopy*. CRC Press, 2007. doi: 10.1201/9781420018318.
- [69] R. W. Kennard and L. A. Stone, “Computer Aided Design of Experiments,” *Technometrics*, vol. 11, no. 1, p. 137, Feb. 1969, doi: 10.2307/1266770.
- [70] E. Trullols, I. Ruisánchez, and F. X. Rius, “Validation of qualitative analytical methods,” *TrAC Trends in Analytical Chemistry*, vol. 23, no. 2, pp. 137–145, Feb. 2004, doi: 10.1016/S0165-9936(04)00201-8.

Apêndice A. Capítulo 1

Material suplementar do manuscrito publicado: Viabilidade do uso de espectrofotômetros compactos de infravermelho próximo e análise multivariada de dados para avaliar características de café torrado e moído

Jamille Carvalho Souza, Celio Pasquini, Maria C. Hespanhol. Food Control, 138 (2022) 109041. <https://doi.org/10.1016/j.foodcont.2022.109041>

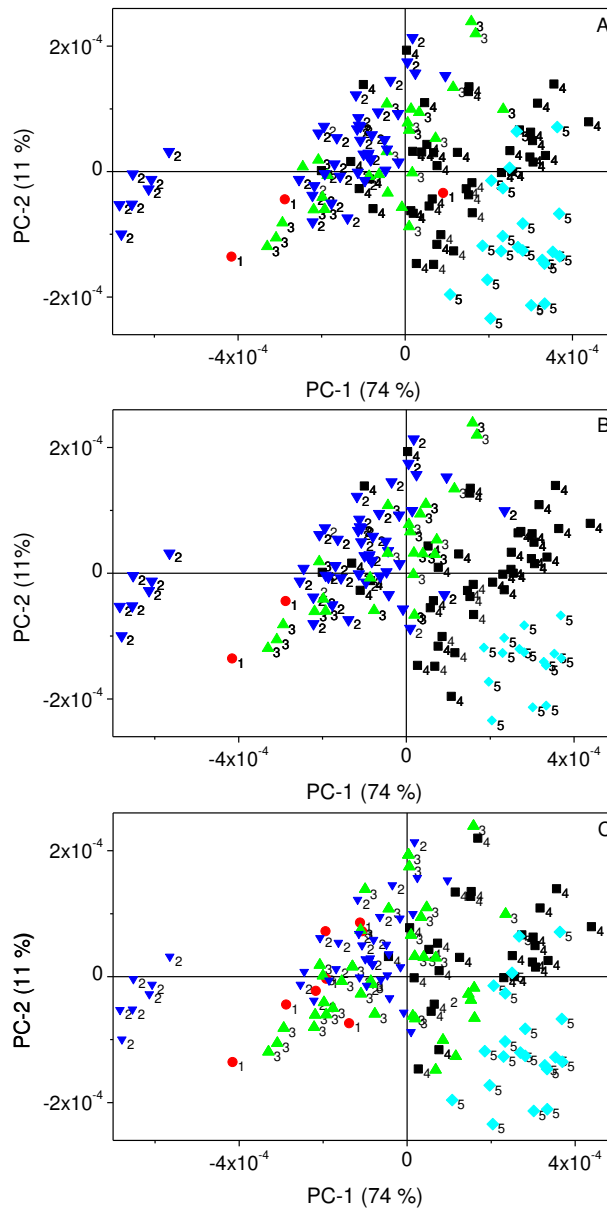


Figura S1.1. Distribuição dos *scores* da PCA do conjunto de dados obtido pelo instrumento 1, categorizados por: (A) grau de torra, (B) corpo, e (C) amargor. Os números próximos aos símbolos referem-se aos valores das características sensoriais.

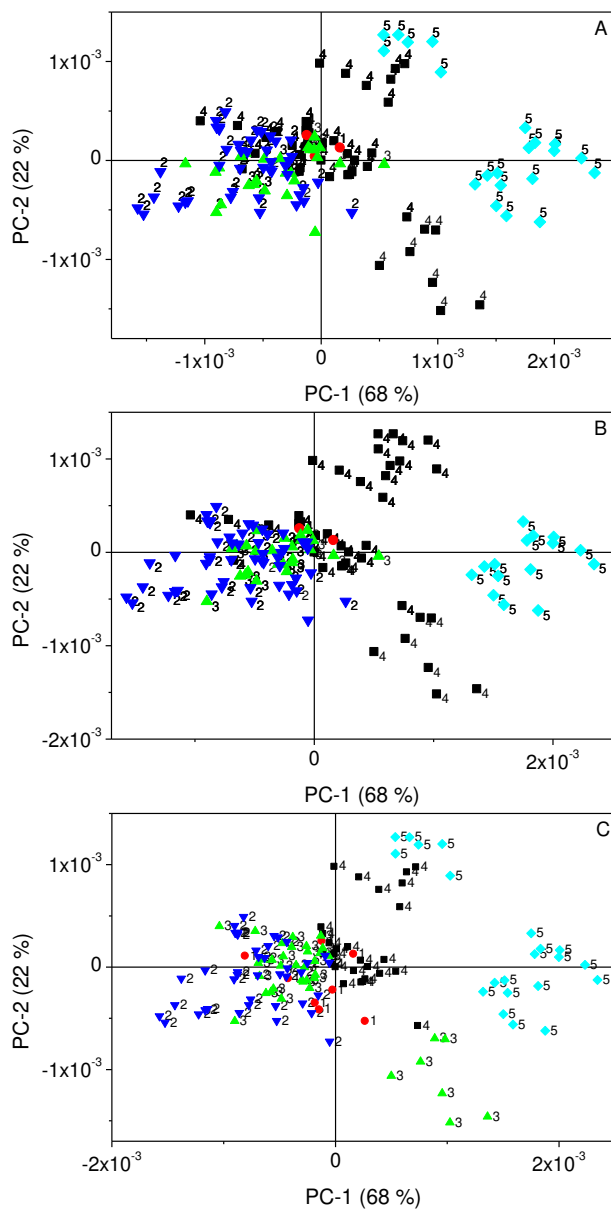


Figura S1.2. Distribuição dos *scores* da PCA do conjunto de dados obtido pelo instrumento 2, categorizados por: (A) grau de torra, (B) corpo, e (C) amargor. Os números próximos aos símbolos referem-se aos valores das características sensoriais.

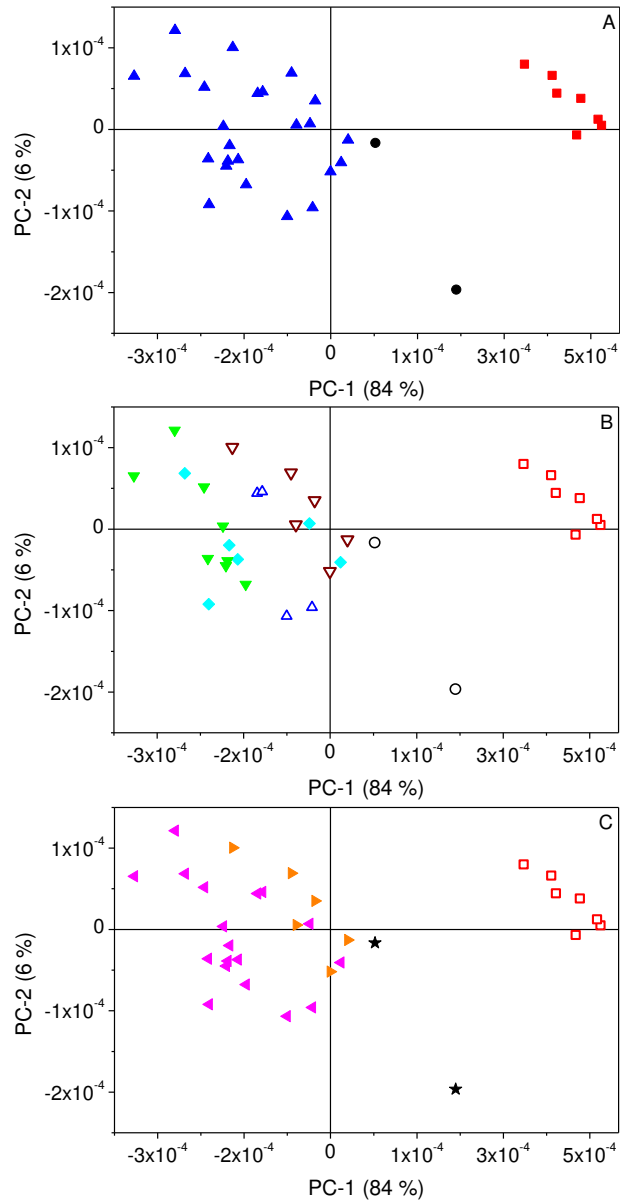


Figura S1.3. Distribuição dos scores de PCA dos cafés regulares com intensidade igual a 4, utilizando o conjunto de dados obtido pelo instrumento 1. (A) de acordo com a variedade, ■ Arábica puro, ▲ Blends de Arábica, ● Blend de Arábica e Robusta; (B) tipo, ○ *Buenos Aires Lungo*, □ *Ethiopia*, ▽ *Cosi*, △ *Linizio Lungo*, ◆ *Vivalto Lungo*, ▼ *Volluto*; e (C) origem, ★ Colômbia/Uganda, ► Costa Rica/Quênia, ◀ Brasil/Colômbia.

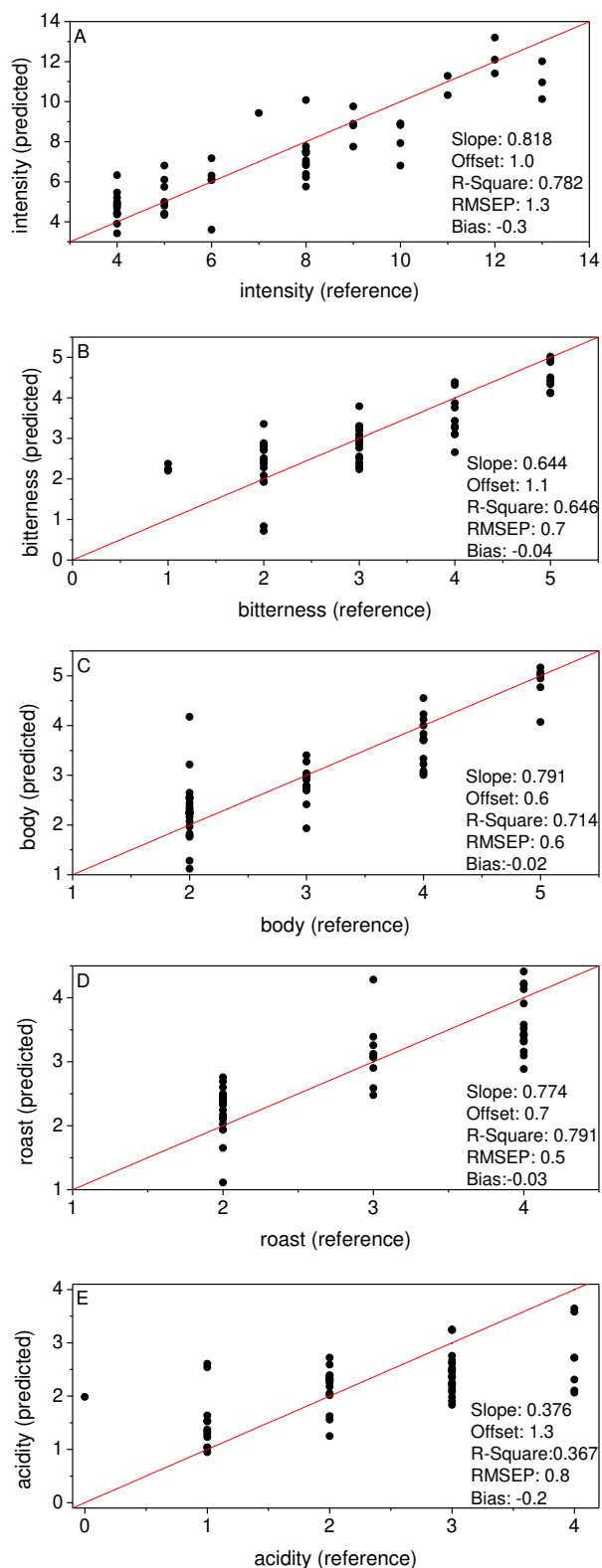


Figura S1.4. Comparação dos resultados de validação preditos pelos modelos PLS com valores de referência para características do café empregando o conjunto de dados espectrais obtidos pelo instrumento 1 e as variáveis selecionadas pelo algoritmo Jack-knife. (A) intensidade, (B) amargor, (C) corpo, (D) torra, e (E) acidez.

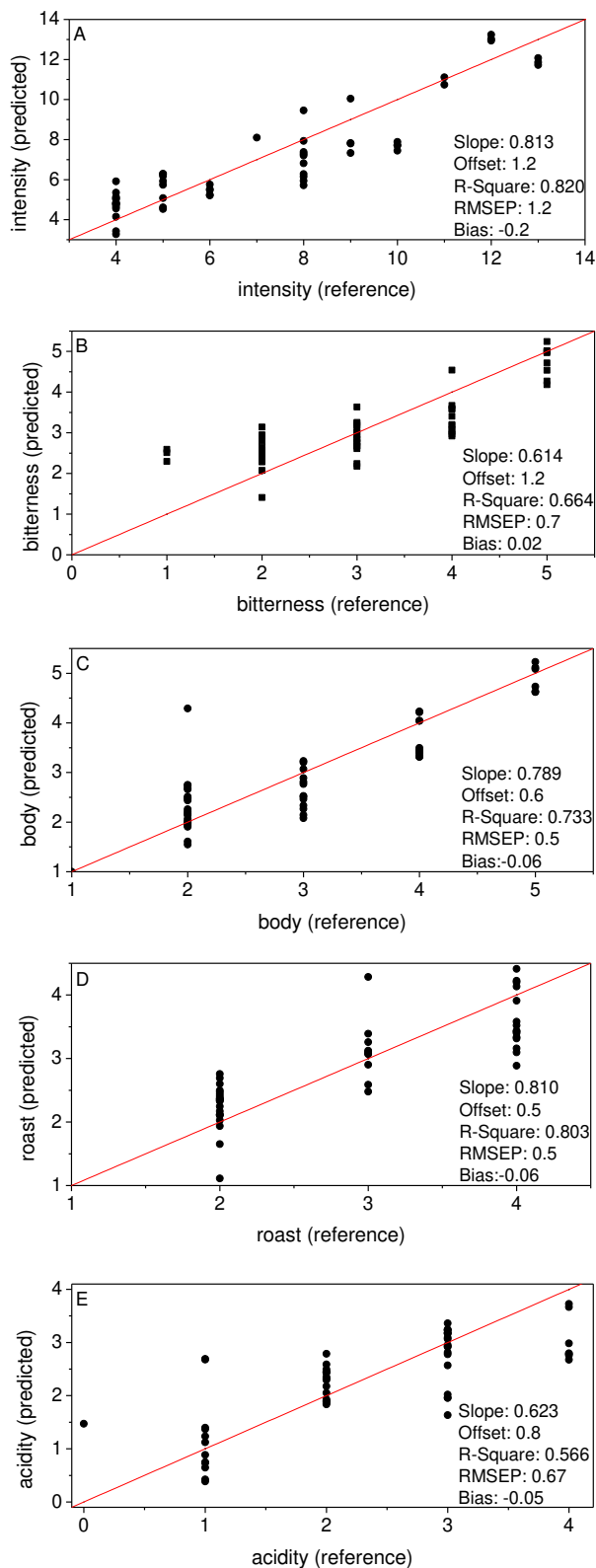


Figura S1.5. Comparação dos resultados de validação preditos pelos modelos PLS e valores de referência para características do café empregando o conjunto de dados espectrais obtidos pelo instrumento 2 e as variáveis selecionadas pelo algoritmo Jack-knife. (A) intensidade, (B) amargor, (C) corpo, (D) torra, e (E) acidez.

Apêndice B. Capítulo 3

Material suplementar do manuscrito publicado: Explorando a quimiometria para análise de triagem por espectroscopia no infravermelho próximo sem o apoio de métodos analíticos de referência: aplicação a edulcorantes à granel.

Jamille Carvalho Souza, Maria C. Hespanhol, Fábio R. P. Rocha, Celio Pasquini *Microchemical Journal*, 190 (2023) 108732. <https://doi.org/10.1016/j.microc.2023.108732>

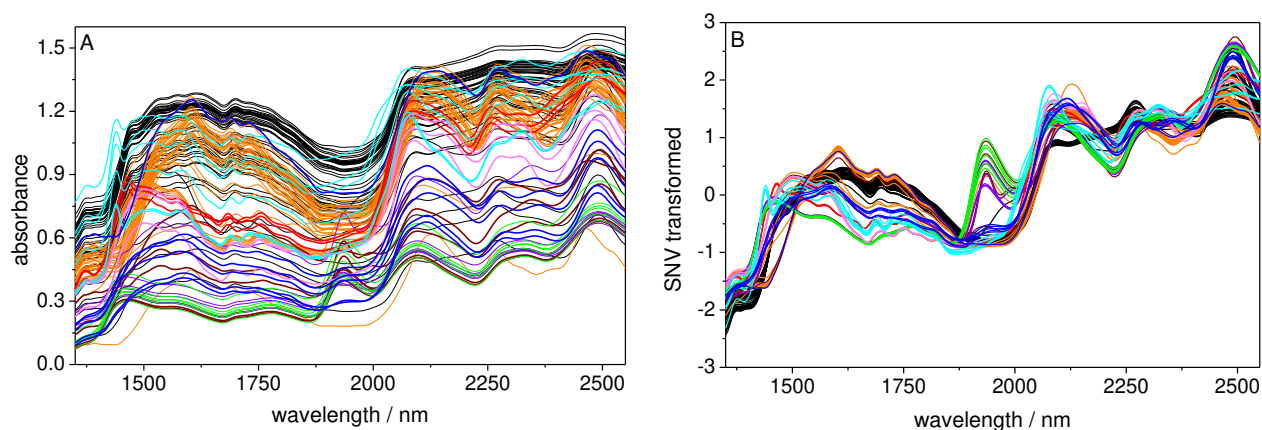


Figura S3.1. Conjunto de espectros NIR de todas as amostras e substâncias de referência empregadas. (A) conjunto de espectros brutos; (B) conjunto de espectros após transformação SNV. — xilitol; — frutose; — sorbitol; — maltodextrina; — eritritol; — maltitol; — sacarose; — estévia; e — sucralose.

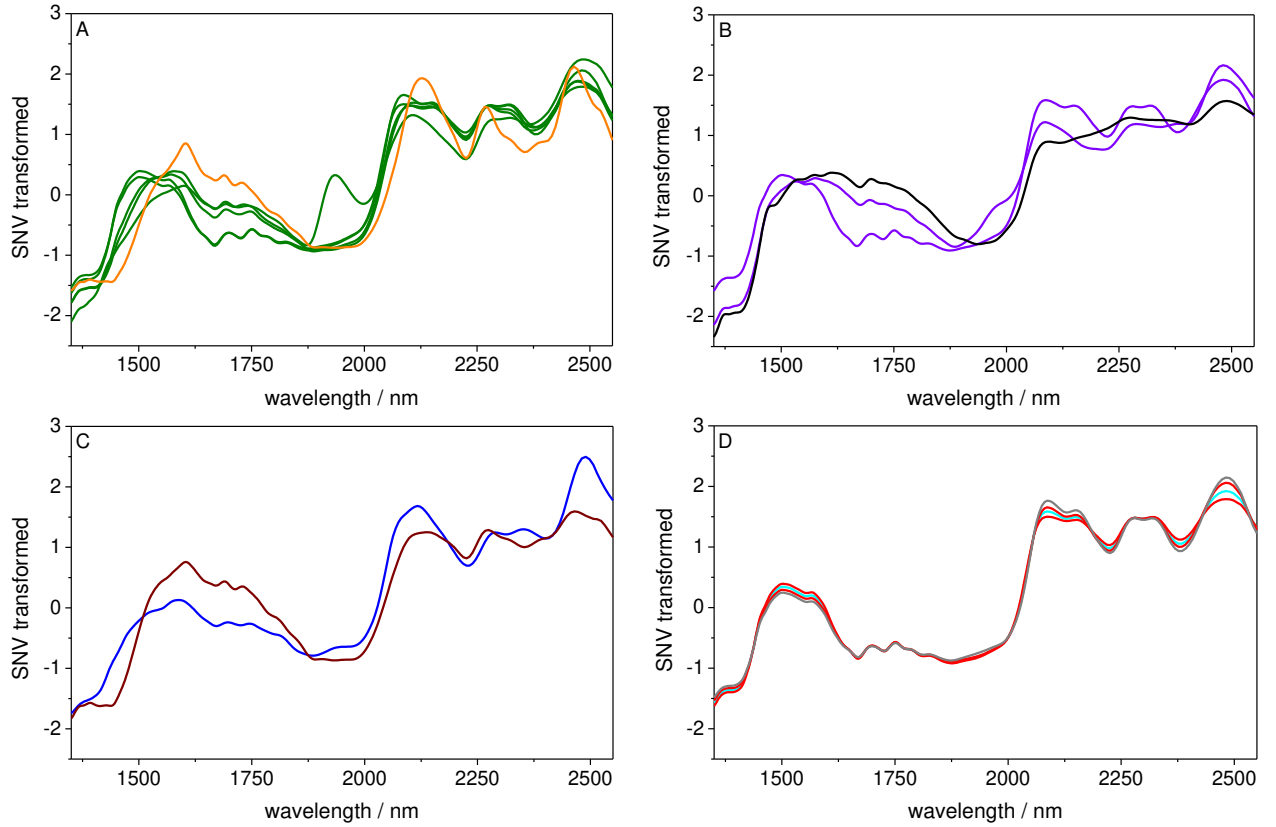


Figura S3.2. Comparação dos espectros das amostras falsificadas e suas respectivas substâncias de referência. (A) — amostras rotuladas como eritritol (A_1 , A_3 , A_4 , A_5 , A_6), — eritritol de referência, (B) — amostras rotuladas como xilitol (A_2 , A_7), — xilitol de referência, (C) — amostra rotulada como (A_8), — sorbitol de referência, (D) — amostras rotuladas como eritritol, — amostra rotulada como xilitol (A_2), — maltitol de referência.