

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA FUNDAMENTAL
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

VANESSA DA SILVA OLIVEIRA

**ANÁLISE DA AUTENTICIDADE DE CÉDULAS DE REAL UTILIZANDO
ESPECTROSCOPIA NIR PORTÁTIL E QUIMIOMETRIA**

Recife
2018

VANESSA DA SILVA OLIVEIRA

**ANÁLISE DA AUTENTICIDADE DE CÉDULAS DE REAL UTILIZANDO
ESPECTROSCOPIA NIR PORTÁTIL E QUIMIOMETRIA**

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Química do Departamento de Química Fundamental da Universidade Federal de Pernambuco como parte dos requisitos para obtenção do título de mestre em química.

Área de concentração: Química Analítica

Orientadora: Profa. Dra. Claudete Fernandes Pereira
Coorientadora: Profa. Dra. Fernanda A. Honorato
Coorientador: Dr. Ricardo Saldanha Honorato

Recife
2018

Catálogo na fonte
Bibliotecário Elaine Freitas CRB4-1790

S586a Oliveira, Vanessa da Silva
 Análise da autenticidade de cédulas de real utilizando espectroscopia nir
 portátil e quimiometria / Vanessa da Silva Oliveira . – 2018.
 44 f. : fig.

 Orientadora: Claudete Fernandes Pereira
 Mestrado (Dissertação) – Universidade Federal de Pernambuco. CCEN.
 Química Fundamental. Recife, 2018.
 Inclui referências e apêndice.

 1. Química analítica. 2. Análise espectral. 3. Espectroscopia de
 infravermelho. 4. Quimiometria. I. Pereira, Claudete Fernandes
 (Orientadora). II. Título.

 543.5 CDD (22. ed.) UFPE-FQ 2018-54

VANESSA DA SILVA OLIVEIRA

***ANÁLISE DA AUTENTICIDADE DE CÉDULAS DE REAL
UTILIZANDO ESPECTROSCOPIA NIR PORTÁTIL E
QUIMIOMETRIA***

Dissertação apresentada ao Programa de Pós-Graduação no Departamento de Química Fundamental da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Química.

Aprovada em: 27/07/2018

BANCA EXAMINADORA

Profa. Claudete Fernandes Pereira
Universidade Federal de Pernambuco
Departamento de Química Fundamental

Profa. Carolina Santos Silva
Universidade Federal de Pernambuco
Departamento de Engenharia Química

Profa. Simone da Silva Simões
Universidade Estadual da Paraíba
Departamento de Química

AGRADECIMENTOS

A minha orientadora, professora Claudete Pereira, por ter atendido meu pedido de trabalhar com química forense, pelos ensinamentos e pelos puxões de orelha que sem dúvida me tornaram uma profissional melhor.

A minha coorientadora, professora Fernanda Honorato, pelas importantíssimas contribuições e também por sua disponibilidade e paciência.

Ao meu coorientador Ricardo Honorato, por todo o engajamento neste projeto, tornando possível a sua realização.

Aos colegas do grupo de pesquisa: Eduardo, Vitor e Jailson por toda a ajuda no início deste projeto.

Aos membros do LAC, Francielson e João, pelo auxílio no uso dos equipamentos.

Ao CNPq pela minha bolsa de mestrado, ao INCTAA e NUQAAPE pelos recursos concedidos.

RESUMO

A falsificação de cédulas, mesmo com os investimentos governamentais em recursos gráficos avançados e itens de segurança, tem se tornado tão sofisticada que algumas são capazes de escapar aos testes de autenticidade. Portanto, o desenvolvimento de métodos mais eficientes, especialmente aqueles baseados na informação química, capazes de detectar a falsificação de maneira precisa e não destrutiva, torna-se relevante. Neste trabalho, a espectroscopia no infravermelho próximo (NIR) combinada com Modelagem Independente e Flexível por Analogia de Classes (SIMCA) e Análise Discriminante Linear (LDA) associada ao Algoritmo de Projeções Sucessivas (SPA) foi aplicada para identificar cédulas de Real (R\$20, R\$50 e R\$100) falsas e verdadeiras, usando um equipamento NIR portátil para a aquisição dos espectros, permitindo a realização de ensaios em campo, através de procedimento simplificado. Trezentas cédulas verdadeiras e duzentas e vinte sete falsas, analisadas na sede da Polícia Federal de Pernambuco, foram empregadas neste trabalho. Os espectros no infravermelho próximo (950-1650 nm) de sete áreas diferentes das cédulas (duas com tinta fluorescente, uma na marca d'água, três nas impressões calcográficas e uma no número de série impresso em tipografia) foram avaliados, usando um espectrômetro portátil, MicroNIRTM 1700 JSDU. Modelos SIMCA e SPA-LDA foram construídos usando os espectros pré-processados com a primeira derivada de uma das áreas calcográficas, a qual foi mais informativa para identificação da autenticidade. Gráficos de escores e resíduos versus influência foram utilizados para detectar amostras anômalas. Os modelos SPA-LDA para as cédulas de R\$20, R\$50 e R\$100 foram desenvolvidos utilizando-se doze, duas e dezenove variáveis, respectivamente. Para os modelos SIMCA, todas (300) as cédulas verdadeiras foram classificadas corretamente e nenhuma das falsas (227) foi classificada como verdadeira. Para os modelos SPA-LDA de duas classes (verdadeiras e falsas), todas as amostras de teste foram corretamente classificadas em suas respectivas classes. Para ambas as técnicas de classificação foram alcançadas sensibilidade, especificidade e eficiência igual a 100% e coeficiente de correlação de Matthew de +1. Assim, a combinação da espectroscopia NIR com instrumento portátil e modelos SIMCA e SPA-LDA é um método analítico efetivo, simples, rápido e não destrutivo para avaliar a autenticidade de cédulas de real.

Palavras-chave: Cédulas. Autenticidade. Infravermelho próximo portátil. Quimiometria. Classificação multivariada.

ABSTRACT

Banknote counterfeiting, even with the government investments in advanced graphic resources and security features, have become so sophisticated that some can escape to authenticity exams. Therefore, the development of more efficient methods, especially those based on chemical information, able to detect the counterfeit in an unambiguous and non destructive mode, becomes relevant. In this work, the near infrared spectroscopy (NIR) combined with Soft Independent Modeling of Class Analogy (SIMCA) and Linear Discriminant Analysis (LDA) associated to Successive Projections Algorithm (SPA) was applied to identify counterfeit and authentic Brazilian Real (R\$20, R\$50 e R\$100) banknotes, using portable NIR equipment for spectra acquisition, thus allowing in field analysis to be carried out, through a simplified procedure. Three hundred authentic and two hundred and twenty seven counterfeit Brazilian banknotes, analyzed at the headquarters of the Federal Police in Pernambuco, were employed in this study. Near infrared spectra (950-1650 nm) from seven different areas of the banknotes (two with fluorescent ink, one over watermark, three with intaglio printing process and one over the serial numbers with typography printing) were evaluated using a portable MicroNIRTM Spectrometer 1700 JSDU. SIMCA and SPA-LDA models were built using first derivative preprocessed spectral data from one of the intaglio areas, which was more informative to authenticity identification. Score and residual versus leverage plots were used to detect outliers. The SPA- LDA models to the R\$20, R\$50 and R\$100 banknotes were developed using twelve, two and nineteen variables, respectively. For the SIMCA models, all authentic (300) banknotes were correctly classified and none of the counterfeits (227) was classified as authentic. For the two classes SPA-LDA models (authentic and counterfeit currencies), all the test samples were correctly classified into their respective class. For both classification techniques were achieved sensitivity, specificity and efficiency equal to 100% and Matthews's correlation coefficient of +1. Thus, the combination of the NIR spectroscopy with portable equipment and SIMCA and SPA-LDA models is an effective, simple, fast and non-destructive analytical method to assessment Real banknotes authenticity.

Keywords: Banknotes. Authenticity. Portable near-infrared. Chemometrics. Multivariate classification.

LISTA DE FIGURAS

Figura 1 –	Distribuição de cédulas falsas retidas no Brasil em 2017.....	11
Figura 2 –	Modos vibracionais moleculares. + movimento para fora da página; - movimento para dentro de página.....	15
Figura 3 –	Diagramas da energia potencial: (a) oscilador harmônico, (b) curva 1, oscilador harmônico e curva 2, movimento anarmônico.....	17
Figura 4 –	Áreas analisadas nas cédulas: 1,7- fluorescentes; 2- marca d'água; 3,5,6- calcográficas, 4- tipográfica.....	29
Figura 5 –	Procedimento de aquisição espectral com o NIR portátil.....	30
Figura 6 –	Espectros médios sem pré-processamento das áreas analisadas nas cédulas. (A) R\$20, (B) R\$50 e (C) R\$100.	32
Figura 7 –	Espectros médios sem pré-processamento do papel sulfite e da marca d'água das cédulas verdadeiras e falsas de R\$20, R\$50 e R\$100.....	33
Figura 8 –	Gráfico de escores para R\$20(A e B), R\$50(C e D), R\$100(E e F) empregando os espectros pré-processados com a primeira derivada.	34
Figura 9 –	Gráficos dos pesos para todas as áreas analisadas das cédulas de R\$20, R\$50 e R\$100.....	35
Figura 10 –	Espectro médio pré-processado de todas as áreas analisadas nas cédulas. (A) R\$20, (B) R\$50 e (C) R\$100.	36
Figura 11 –	Número ótimo de variáveis no SPA-LDA para (A) área 6 de R\$20, (B) área 5 de R\$50 e (C) área 6 de R\$100.....	37
Figura 12 –	Variáveis selecionadas dos espectros médios para (A) área 6 R\$20, (B) área 5 R\$50 e (C) área 6 R\$100.	38

LISTA DE ABREVIATURAS

ATR	Reflectância total atenuada
AOFT	Filtro óptico- acústico sintonizável (Acoustoopic Tunable Filter)
E	Energia potencial
Fcal	Valor calculado para o teste F
Fcrit	Valor crítico adotado para o teste F
FIR	Infravermelho Distante (<i>Far Infrared</i>)
FN	Falso negativo
FP	Falso positivo
G	Risco médio de uma classificação incorreta pelo LDA
GA	Algoritmo genético (<i>Genetic Algorithm</i>)
HCA	Análise de agrupamento hierárquico (<i>Hierarchical Cluster Analysis</i>)
iPLS	Regressão dos mínimos quadrados parciais por intervalos (<i>Internal Partial Least Squares Regression</i>)
IR	Infravermelho (<i>Infrared</i>)
JK	<i>Jack-Knife</i>
KNN	k-ésimo vizinho mais próximo (<i>K-Nearest Neighbors</i>)
KS	<i>Kennard-Stone</i>
LDA	Análise discriminante linear (<i>Linear Discriminant Analysis</i>)
LVF	Filtro linear variável (<i>Linear Variable Filter</i>)
MIR	Infravermelho Médio (<i>Medium Infrared</i>)
MLR	Regressão Linear Múltipla (<i>Multiple Linear Regression</i>)
MSC	Correção Multiplicativa de Sinal (<i>Multiplicative Signal Correction</i>)
NIPALS	Algoritmo dos mínimos quadrados parciais iterativos não-lineares (<i>Nonlinear Iterative Partial Least Squares</i>)
NIR	Infravermelho Próximo (<i>Near Infrared</i>)
PCA	Análise de Componentes Principais (<i>Principal Components Analysis</i>)
PCs	Componentes principais (<i>Principal Components</i>)
SIMCA	Modelagem independente e flexível por analogia de classe (<i>Soft Independent Modeling of Class Analogy</i>)
SNV	Variação normal padrão (<i>Standard Normal Variation</i>)
SPA	Algoritmo das projeções sucessivas (<i>Successive Projections Algorithm</i>)
SVD	Algoritmo de decomposição por valores singulares (<i>Singular Value Decomposition</i>)
SW	<i>Stepwise</i>
UV	Radiação ultravioleta
VN	Verdadeiro negativo
VP	Verdadeiro positivo

SUMÁRIO

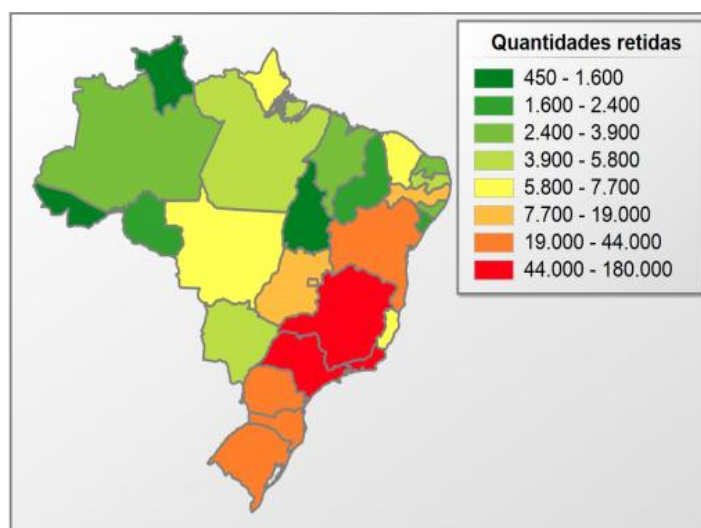
1	INTRODUÇÃO	11
2	FUNDAMENTAÇÃO TEÓRICA	13
2.1	Espectroscopia vibracional aplicada à análise da autenticidade de cédulas	13
2.2	Espectroscopia no infravermelho	14
2.3	Espectroscopia no infravermelho próximo	17
2.4	Quimiometria	19
2.5	Técnicas de reconhecimento de padrões	20
2.5.1	<i>PCA</i>	20
2.5.2	<i>SIMCA</i>	22
2.5.3	<i>LDA</i>	23
2.6	Validação dos modelos de classificação	23
2.7	Pré- processamento de dados	24
2.7.1	<i>Derivadas</i>	24
2.7.2	<i>Correção Multiplicativa de Sinal (MSC: Multiplicative Signal Correction)</i>	24
2.7.3	<i>Variação normal padrão (SNV: Standard Normal Variation)</i>	25
2.8	Seleção de amostras	25
2.9	Seleção de variáveis	26
2.10	Figuras de mérito	27
3	MATERIAIS E MÉTODOS	29
3.1	Amostras	29
3.2	Equipamento e aquisição dos espectros	30
3.3	Tratamento quimiométrico dos dados	30
4	RESULTADOS E DISCUSSÃO	32
4.1	Espectros NIR	32
4.2	Análise exploratória dos dados espectrais	34
4.3	Modelos de classificação	36

4.3.1	<i>SIMCA</i>	36
4.3.2	<i>SPA- LDA</i>	37
4.4	Figuras de mérito	38
5	CONCLUSÃO	39
	REFERÊNCIAS	40
	APÊNDICE A - PUBLICAÇÃO DECORRENTE DO TRABALHO	44

1 INTRODUÇÃO

Inflação, mercado clandestino, prejuízos para o comércio e para prestação de serviços, são algumas das consequências desastrosas da circulação de cédulas falsas, além do impacto severo na economia. A falsificação de cédulas é um crime recorrente em muitos países, inclusive no Brasil, que em 2017 registrou a apreensão de 500.726 cédulas falsas, em todos os estados, abrangendo todos os valores da primeira e segunda família do Real, vigentes desde 1994 e 2010, respectivamente. O estado de Pernambuco está entre os 10 primeiros na quantidade de cédulas falsas retidas, com 14.546 apreensões nesse mesmo ano. As cédulas com maior número de apreensões são as de R\$20, R\$50 e R\$100. A Figura 1 mostra a distribuição da quantidade de cédulas retidas no país, durante 2017. (BANCO CENTRAL DO BRASIL, 2017).

Figura 1 – Distribuição de cédulas falsas retidas no Brasil em 2017.



Fonte: Banco Central do Brasil (2017)

Como forma de inibir a falsificação, muitos governos têm investido esforços e custos para implantação de recursos gráficos avançados e itens de segurança em suas cédulas. A segunda família do Real é um exemplo do avanço tecnológico nessa área. Baseada na análise sensorial essa moeda traz itens que podem ser vistos (marca d'água, fios de segurança e quebra-cabeça), sentidos (alto relevo - calcografia e textura do papel) ou descobertos (faixa holográfica, números que mudam de cor, número escondido, elementos fluorescentes e microimpressões).

Na cartilha de treinamento disponibilizada pelo Banco Central do Brasil são

apresentados os testes para avaliação da autenticidade das cédulas com base nesses itens de segurança. (BANCO CENTRAL DO BRASIL, 2012). Esses exames, que são tácteis ou visuais, requerem treinamento prévio, alguns, inclusive exigem o emprego de equipamento, por exemplo, para detecção de tipo de impressão, microletras e de elementos fluorescentes. Além do mais, podem ser considerados subjetivos, uma vez que dependem da experiência do perito analista.

A fim de facilitar e agilizar os exames de autenticidade, estão disponíveis comercialmente muitos equipamentos portáteis baseados principalmente na radiação ultravioleta (UV). Porém, a característica fluorescente analisada pela maioria desses equipamentos é frágil, já que tintas fluorescentes comerciais (“UV Stuff”, [s.d.]) podem ser facilmente utilizadas pelos falsificadores, podendo causar um falso-positivo na detecção por radiação UV.

Neste sentido, os dispositivos existentes podem ser ineficientes em alguns casos, haja vista a sofisticação das falsificações e a habilidade dos fraudadores em burlar as atuais técnicas de verificação de autenticidade. Assim, o desenvolvimento de exames mais confiáveis, principalmente aqueles baseados na informação química, torna-se relevante.

Diferentes técnicas analíticas já vêm sendo utilizadas para identificar cédulas falsas com base na informação química, tais como, radiometria fototérmica infravermelha por laser induzido (OTHONOS et al., 1997), fluorescência de raios-X (APPOLONI; MELQUIADES, 2014; RUSANOV et al., 2009), espectrometria de massas (EBERLIN et al., 2010; KAO et al., 2013, 2016; SCHMIDT et al., 2014), espectroscopias Raman (BOŽIČEVIĆ; GAJOVIĆ; ZJAKIĆ, 2012; DE ALMEIDA et al., 2013), infravermelho médio (IMPERIO et al., 2015; SONNEX et al., 2014; VILA et al., 2006) e infravermelho próximo (DALE; KLATT, 1999). Contudo, durante o desenvolvimento deste trabalho não foi reportado na literatura o uso da espectroscopia NIR portátil na análise da autenticidade de cédulas.

Portanto, o objetivo deste trabalho foi o de desenvolver um método analítico eficiente, rápido, de baixo custo e não destrutivo baseado em medidas espectrais usando um espectrômetro NIR portátil e dois métodos de classificação, SIMCA e o SPA-LDA, para distinguir cédulas de Real falsas de autênticas (R\$20, R\$50 e R\$100) em análises de campo.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 Espectroscopia vibracional aplicada à análise da autenticidade de cédulas

As técnicas analíticas baseadas na espectroscopia vibracional vêm sendo reportadas na literatura para diferenciação de cédulas falsas e verdadeiras, bem como em estudos com papel moeda.

Vila e colaboradores (2006) distinguiram cédulas de euro (50€ e 100€) genuínas de cédulas falsas, de diferentes países (Espanha, Alemanha e França), utilizando espectroscopia no infravermelho médio (MIR) e análise de componentes principais (PCA). O procedimento foi baseado na aquisição dos espectros ($4000\text{-}720\text{ cm}^{-1}$) com um microscópio no modo de reflectância total atenuada (ATR) em quatro áreas diferentes, cada uma com $200\text{ }\mu\text{m}^2$, na frente e no verso da cédula. Análise dos espectros das notas falsas e verdadeiras revelaram diferenças nas bandas de absorção. As cédulas genuínas tiveram espectros reproduzíveis para cada uma das áreas analisadas, variando minimamente de acordo com a origem da nota. Os resultados da PCA mostraram uma distinção clara entre as cédulas originais e as falsificadas.

MIR também foi empregada por Sonnex (2014) para avaliar a autenticidade de cédulas de £20 de bancos ingleses, utilizando dois equipamentos portáteis de baixa resolução e um terceiro equipamento mais sofisticado com um sistema de imagem (microscópio). As cédulas verdadeiras e falsas foram analisadas a partir dos espectros adquiridos nos três equipamentos em três áreas diferentes da cédula (área em branco, filme polimérico que recobre a faixa holográfica e no logo do Banco da Inglaterra). Mesmo com baixa resolução, os espectros obtidos nos equipamentos portáteis mostraram diferenças nos perfis de absorção, permitindo a identificação das cédulas falsificadas. Nos casos ambíguos, os autores sugeriram o uso do equipamento mais sofisticado, o qual tem melhor resolução e imagens das áreas em branco do papel e do filme polimérico que cobre a faixa holográfica.

Recentemente, Imperio e colaboradores (2015) aplicaram as espectroscopias Raman e MIR no modo ATR para identificar cédulas falsas e avaliar o estado de conservação de cédulas de 1000£ impressas na Itália entre 1947 e 2002. As análises foram feitas em diversas áreas das cédulas como marcas d'água, marcas de segurança e números de série, além de fibras, colas e revestimentos da superfície; em cinco tipos diferentes de impressões durante esses 55 anos. Os espectros MIR e Raman mostraram diferenças no papel moeda e na composição dos pigmentos usados em cada época. Nas cédulas falsas, os espectros Raman revelaram a presença de pigmentos diferentes daqueles encontrados nas cédulas verdadeiras. Os espectros MIR demonstraram que o papel utilizado na confecção das notas falsas tinha

composição diferente daquela do papel moeda das cédulas autênticas.

Dale e Klatt (1999) reportaram o uso da reflectância difusa no infravermelho próximo (NIR) e do reconhecimento de padrões para verificar a possibilidade de separar papel moeda de papel estoque de boa qualidade. As amostras das cédulas foram analisadas em áreas sem pigmento e comparadas com os espectros do papel estoque de boa qualidade. O método do k-ésimo vizinho mais próximo (KNN) foi aplicado para determinar o grau de classificação. Os resultados mostraram que, mesmo as amostras tendo espectros similares e sobrepostos, foi possível separá-las.

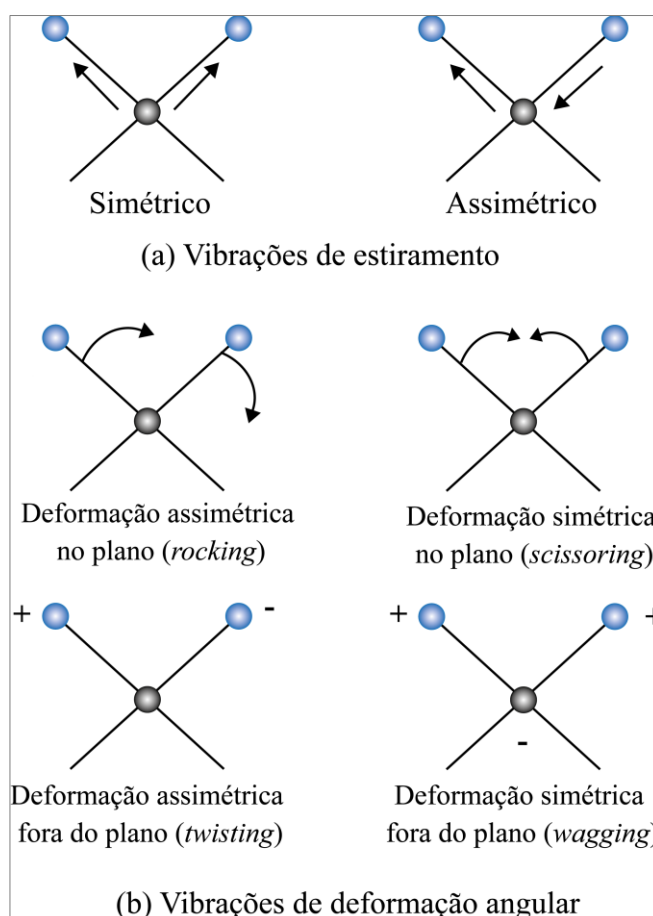
Os trabalhos anteriormente citados descrevem a habilidade das espectroscopias NIR, MIR e Raman em identificar cédulas falsas. Durante o desenvolvimento desta pesquisa não foi reportado na literatura o uso de equipamentos portáteis para essa finalidade. Devido à livre circulação das cédulas em bancos e no comércio, é relevante o desenvolvimento de um método analítico simples e prático para examinar a autenticidade das cédulas. A espectroscopia NIR apresenta esses requisitos e pode ser implementada usando um equipamento portátil, que permite uma análise em campo, com baixo custo se comparado aos equipamentos de bancada, empregados nos artigos citados. Além disso, análises mais confiáveis, quando comparadas àquelas obtidas com os equipamentos UV comerciais, podem ser obtidas, já que a espectroscopia NIR considera a informação da composição química tanto da tinta quanto do papel. Métodos que utilizam a informação química podem facilmente detectar cédulas falsas, pois distinguem a composição das tintas utilizadas, mesmo se as cédulas exibirem perfis visuais ou fluorescentes similares. A capacidade de análise em campo de forma eficiente e com equipamento de baixo custo é de extrema importância para uma posterior transferência da tecnologia proposta para setores da sociedade, como de serviço financeiro e de segurança. Ademais, é importante que as forças policiais tenham métodos que lhes permitam estar à frente de novas ações dos falsificadores.

2.2 Espectroscopia no infravermelho

A espectroscopia no infravermelho (IR) é uma técnica vibracional na qual a radiação eletromagnética interage com a amostra através do processo de absorção molecular (SKOOG; HOLLER; CROUCH, 2009). A radiação infravermelha não tem energia suficiente para realizar transições eletrônicas, como acontece na absorção da radiação visível e ultravioleta, ficando restrita a pequenas variações energéticas entre os diferentes estados vibracionais e rotacionais. Apenas algumas moléculas podem absorver a radiação IR, pois, para que a radiação incidente seja absorvida e o modo de vibração ou rotação seja detectado é necessário

que haja alteração do momento de dipolo durante esses movimentos, tal como acontece em moléculas heteronucleares, como HCl (SKOOG; HOLLER; CROUCH, 2009). Os diferentes modos de vibração e rotação das moléculas dão origem aos espectros na região do infravermelho, que correspondem à faixa de número de onda de $12.800 - 10 \text{ cm}^{-1}$. Esses diferentes tipos de vibrações que ocorrem nas ligações moleculares são mostrados na Figura 2.

Figura 2 – Modos vibracionais moleculares. + movimento para fora da página; - movimento para dentro de página.



Fonte: Adaptado de Skoog; Holler; Crouch (2009)

A faixa espectral do infravermelho pode ser dividida em três regiões: o IR-próximo (NIR) que abrange a faixa de 12800 a 4000 cm^{-1} e revela as bandas de sobretons e combinações, o IR-médio (MIR) que compreende os números de onda entre 4000 e 200 cm^{-1} e observa transições rotacionais- vibracionais, sendo a região mais utilizada nas análises qualitativas, e o IR- distante (FIR) que inclui a faixa com números de onda de 200 a 10 cm^{-1} e exibe bandas relacionadas às transições rotacionais, que, por serem quantizadas, aparecem como linhas discretas bem definidas.

As vibrações de estiramento podem ser explicadas de forma aproximada pelo modelo mecânico do oscilador harmônico, ilustrado na Figura 3 (a) (SKOOG; HOLLER; CROUCH, 2009). Sendo sua energia potencial máxima atingida quando a mola é comprimida ou esticada e zero quando as massas estão em repouso (Equação 1).

$$E = \frac{1}{2} kx^2 \quad (1)$$

onde k é a constante de força e x a distância entre as massas.

Contudo, o modelo descrito pela mecânica clássica não reflete o comportamento real das partículas atômicas, que têm energias de vibração quantizada (Equação 2), sendo necessário, portanto, a inclusão dessa característica pelo uso das equações de onda da mecânica quântica.

$$E_v = \left(v + \frac{1}{2}\right) h\nu \quad (2)$$

onde: E_v é a energia vibracional do v -ésimo nível quântico, v é o número quântico vibracional, h é a constante de Planck e ν é a frequência vibracional fundamental, que no modelo clássico é definida como:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{\mu}} \quad (3)$$

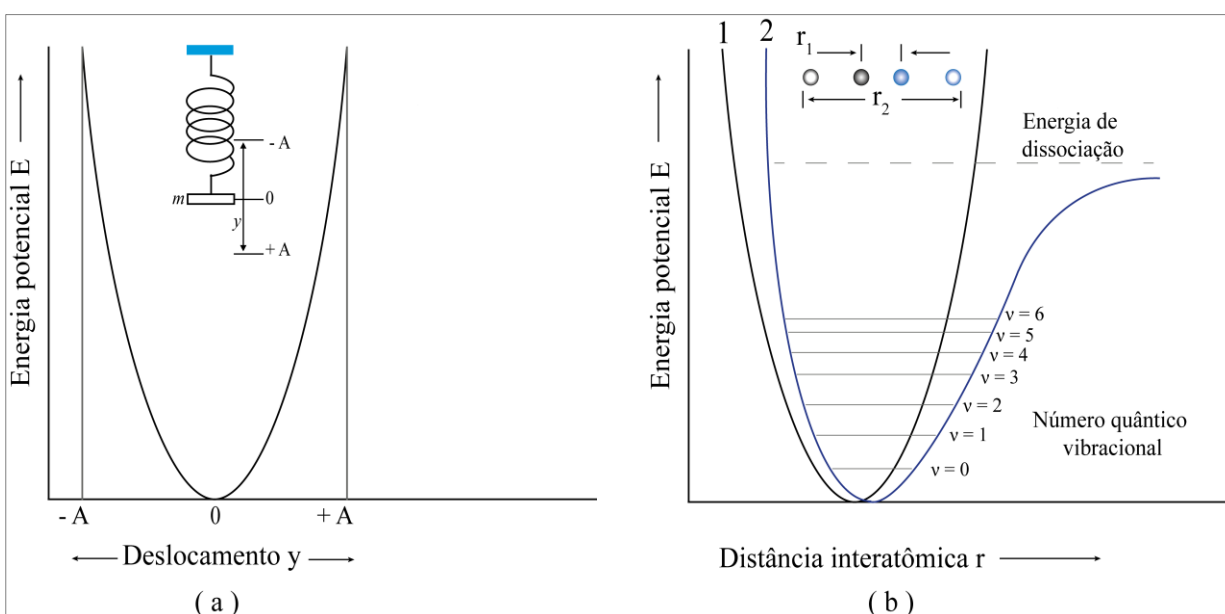
em que μ é a massa reduzida:

$$\mu = \frac{m_1 m_2}{m_1 + m_2} \quad (4)$$

Outra imposição do modelo harmônico é a de que as transições ocorram apenas entre os níveis adjacentes, ou seja, $\Delta v = \pm 1$, o que faz com a diferença de energia entre dois níveis adjacentes seja sempre a mesma (PASQUINI, 2003). Todavia, esse não é o comportamento observado a nível molecular, onde as contribuições das forças coulômbicas precisam ser consideradas. Assim, o modelo que melhor explica as vibrações moleculares é o do oscilador anarmônico, onde níveis de energia $\Delta v = 2$ ou mais elevados (sobretons) também podem ser alcançados e a existência de bandas de combinação é prevista. Esses dois tipos de bandas de absorção são mais comumente observadas na região do infravermelho próximo (NIR). No modelo anarmônico, a separação entre dois níveis de energia adjacentes diminui com o número quântico vibracional (PASQUINI, 2003). Esse modelo também considera a repulsão da nuvem eletrônica dos átomos envolvidos na ligação quando esses estão muito próximos, bem como a ruptura da ligação quando os átomos se afastam.

De forma mais detalhada, esse comportamento pode ser descrito da seguinte maneira (Figura 3): à medida que os núcleos se aproximam (r_1), a força de repulsão coulômbica entre eles estará no mesmo sentido da força restauradora da ligação fazendo com que a energia potencial aumente mais rapidamente do que aquela descrita pelo modelo do oscilador harmônico. Por outro lado, à proporção que os átomos se distanciam a força restauradora diminui assim como a energia potencial, que chega até um valor mínimo e volta a aumentar na medida em se aproxima da distância de dissociação (r_2) atômica. (SKOOG; HOLLER; CROUCH, 2009).

Figura 3 – Diagramas da energia potencial: (a) oscilador harmônico, (b) curva 1, oscilador harmônico e curva 2, movimento anarmônico.



Fonte: Adaptado de Skoog; Holler; Crouch (2009)

Os modos de vibração identificados na espectroscopia de infravermelho são importantes ferramentas tanto para análises qualitativas quanto quantitativas, como por exemplo, na identificação de compostos puros, elucidação de estruturas de compostos orgânicos, inorgânicos e espécies organometálicas, detecção de impurezas e ainda na estimativa da quantidade de um analito em uma mistura complexa, para áreas agrícolas (DONG et al., 2011) e industriais (KAUPPINEN et al., 2013).

2.3 Espectroscopia no infravermelho próximo

Descoberta em 1800, pelo alemão Frederick William-Herschel, a espectroscopia no infravermelho próximo (NIR) só passou a ser utilizada como técnica analítica quantitativa em

1960, por Karl Norris, na determinação de umidade em produtos agrícolas. Contudo, foi somente após 1980 que a espectroscopia NIR se consolidou como técnica analítica alternativa, devido aos avanços tecnológicos que permitiram o desenvolvimento dos equipamentos de aquisição e tratamento dos espectros e em especial, ao advento da quimiometria, que forneceu as ferramentas necessárias para a coleta de informações úteis desses espectros (PASQUINI, 2003).

A região do espectro eletromagnético com número de onda entre 12.800 e 400 cm^{-1} (780 a 2.500 nm) corresponde à região do NIR. As bandas de absorção observadas nessa região são essencialmente sobretons (frequências com amplitude aumentada devido à anarmonicidade das vibrações) e combinações (soma ou diferença entre duas ou mais frequências) de vibrações fundamentais que, em geral, envolvem as ligações C-H, O-H e N-H. Essas bandas têm intensidades 10 a 100 vezes mais fracas que as bandas fundamentais, logo não se confundem com essas (SKOOG; HOLLER; CROUCH, 2009).

Diferentemente das linhas discretas e bem definidas apresentadas pelos espectros no FIR, as bandas presentes nos espectros NIR são normalmente largas e sobrepostas dificultando a correlação simples e direta com a concentração do analito. Para isso, são utilizadas técnicas quimiométricas que auxiliam na interpretação desses dados.

A presença constante dos modos de vibração do hidrogênio torna a região do NIR essencial nas análises qualitativas, quantitativas e estudos de moléculas que contenham os grupos CH, OH e NH, como as proteínas, hidrocarbonetos de baixa massa molar e gorduras, sendo especialmente adotada nos processos industriais dos ramos da agricultura (FARDIM; FERREIRA; DURÁN, 2002), alimentos (RÍOS-REINA et al., 2018), e farmacêutico (DA SILVA; DA SILVA; PEREIRA, 2017). Além das aplicações forenses como quantificação de cocaína (CORREIA et al., 2018), identificação de resíduos biológicos, sangue e sêmen; (PEREIRA et al., 2017a; SILVA et al., 2017) e detecção de adulteração em documentos. (PEREIRA et al., 2017b).

Os instrumentos utilizados na região NIR podem ser de quatro tipos: os instrumentos dispersivos baseados em redes de difração, como aqueles utilizados na região do UV-VIS; os que utilizam filtros discretos como seletor de comprimento de ondas; os instrumentos com filtros óptico- acústicos sintonizáveis (AOFT: *acoustoopic tunable filter*) capazes de difratar a radiação em comprimentos de onda determinados; e ainda os espectrômetros com transformada de Fourier (FT) que têm como principais vantagens a reprodutibilidade e a relação sinal-ruído. As fontes de radiação, geralmente, empregadas nesses instrumentos são lâmpadas do tipo tungstênio-halogênio com janelas de quartzo e os detectores podem ser

fotodetectores de PbS e PbSe ou fotodiodos de InSb e InAs, ou ainda arranjos de detectores como InGaAs (SKOOG, D. A.; HOLLER, F.J.; CROUCH, 2009). Em conjunto com as fontes de radiação esses detectores produzem altas relações sinal-ruído, o que compensa, parcialmente, as baixas intensidades das bandas de absorção NIR (PASQUINI, 2003).

Os instrumentos baseados na transformada de Fourier agregam as melhores características em termos de precisão e exatidão na discriminação do comprimento de onda, alta relação sinal-ruído e velocidade de varredura. Contudo, a mesma tecnologia que possibilita todas essas vantagens, o uso do interferômetro, torna esse tipo de equipamento mais caro (PASQUINI, 2003), o que pode ser uma desvantagem para o uso em análises de rotina.

Os recentes avanços na microeletromecânica têm permitido a miniaturização dos equipamentos que operam na região NIR, possibilitando o desenvolvimento de espectrômetros portáteis mais práticos e econômicos que os atuais instrumentos de bancada. Um exemplo é o espectrômetro MicroNIR comercializado pela JSDU, que utiliza duas lâmpadas de tungstênio associadas a um filtro linear variável (LVF: *Linear Variable Filter*) que funciona como elemento dispersivo. O LVF é ligado diretamente a um conjunto de detectores, sem partes móveis, fazendo com que cada pixel detectado corresponda a um comprimento de onda diferente (O'BRIEN; HULSE; FRIEDRICH, 2012). Como resultado obtém-se um equipamento compacto, leve, resistente, de baixo custo, que, em geral, requer o mínimo de treinamento e pode ser facilmente empregado em análises de rotina, além da possibilidade de transporte do dispositivo.

A simplicidade da arquitetura dos instrumentos portáteis parece não ser um fator limitante para sua eficiência analítica, que em muitos casos é comparável à eficiência dos instrumentos de bancada como demonstrado em trabalhos de diferentes áreas, como química forense (RISOLUTI et al., 2018), análise de combustíveis (DA SILVA et al., 2017) e de alimentos (MALEGORI et al., 2017).

2.4 Quimiometria

É a área da química que utiliza ferramentas estatísticas, matemáticas e computacionais a fim de extrair de forma eficiente informações úteis de um conjunto de dados de interesse químico. As principais subáreas da Quimiometria são: planejamento e otimização de experimentos, reconhecimento de padrões e calibração multivariada. (DE SOUZA; POPPI, 2012).

No planejamento e otimização de experimentos o uso de ferramentas estatísticas

permite identificar as variáveis mais influentes em um processo, bem como a influência de suas interações, possibilitando que seja extraído o máximo de informação do sistema com um número mínimo de ensaios. Já no reconhecimento de padrões busca-se verificar a existência de similaridade ou diferenças entre as respostas obtidas sobre uma série de amostras. Quanto à calibração multivariada, medidas experimentais, como as espectrais, são relacionadas, por meio de modelos matemáticos, a alguma propriedade da amostra, como a concentração, por exemplo. A partir desses modelos é possível prever a concentração ou outras propriedades de amostras desconhecidas. Neste trabalho, foram empregadas as técnicas de reconhecimento de padrões.

2.5 Técnicas de reconhecimento de padrões

Nessa área da quimiometria as técnicas são divididas em: supervisionadas e não supervisionadas. Sendo as técnicas supervisionadas aquelas que usam uma série de dados (conjunto de treinamento), de classe conhecida, para "treinar" o algoritmo a distinguir as classes e, portanto, são usadas para a classificação. Enquanto que, nas técnicas não supervisionadas, as amostras se agrupam naturalmente sem o uso de informações prévias sobre as classes, sendo aplicadas na análise exploratória. Dentre as técnicas não supervisionadas estão a análise de agrupamento hierárquico (HCA: *Hierarchical Cluster Analysis*) e a análise de componentes principais (PCA: *Principal Components Analysis*). Já entre as técnicas supervisionadas podemos destacar o método do k-ésimo vizinho mais próximo (KNN: *K-Nearest Neighbors*), modelagem independente e flexível por analogia de classes (SIMCA: *Soft Independent Modeling of Class Analogy*) e análise discriminante linear (LDA: *Linear Discriminant Analysis*). As técnicas utilizadas neste trabalho são detalhadas a seguir.

2.5.1 PCA

É uma das técnicas não supervisionadas mais utilizadas para o reconhecimento de padrões, pois possibilita a detecção de amostras anômalas (*outliers*), a visualização da dispersão das amostras e a avaliação da relação entre as variáveis. A PCA é uma manipulação matemática feita na matriz de dados que projeta o máximo de informação (variância), utilizando o mínimo de variáveis não correlacionadas, chamadas de componentes principais (PCs) ou fatores (BEEBE; PELL; SEASHOLTZ, 1998). Os eixos das variáveis originais são combinados linearmente para construção de novos eixos ortogonais entre si, as PCs, o que permite a redução das dimensões e uma melhor visualização da matriz de dados,

possibilitando uma análise real da sua natureza multivariada. As similaridades e diferenças das amostras são evidenciadas nas coordenadas (escores) do novo sistema de eixos. A contribuição das variáveis originais para o novo sistema é denominada pesos (*loadings*), sendo calculada como o cosseno do ângulo entre o eixo da variável original e o eixo da PC. Quanto maior o módulo desse valor, maior será a importância da variável na PC. (AGELET; HURBURGH, 2010; BEEBE; PELL; SEASHOLTZ, 1998).

Para o cálculo da PCA são, frequentemente, empregados os algoritmos de decomposição por valores singulares (SVD: *Singular Value Decomposition*) e o dos mínimos quadrados parciais iterativos não-lineares (NIPALS: *Nonlinear Iterative Partial Least Squares*) (WOLD; ESBENSEN; GELADI, 1987), utilizado neste trabalho através do programa *The Unscrambler*. O algoritmo NIPALS decompõe a matriz de dados \mathbf{X} ($n \times m$) nos conjuntos de escores (\mathbf{t}) e pesos (\mathbf{p}). Segundo, Wold e colaboradores (1987) esse processo tem início com a estimativa do vetor de escores, que é a coluna de \mathbf{X} com maior variância. A segunda etapa é o cálculo do vetor peso, \mathbf{p} , obtido pela projeção de \mathbf{X} em \mathbf{t} (Equação 5):

$$\mathbf{p}^T = \frac{\mathbf{t}^T \cdot \mathbf{X}}{\mathbf{t}^T \cdot \mathbf{t}} \quad (5)$$

em seguida o vetor \mathbf{p} é normalizado para comprimento 1:

$$\mathbf{p}_{\text{norm}} = \frac{\mathbf{p}}{\sqrt{(\mathbf{p}^T \cdot \mathbf{p})}} \quad (5.1)$$

e a nova estimativa de escores é obtida:

$$\mathbf{t} = \frac{\mathbf{X} \cdot \mathbf{p}_{\text{norm}}}{\mathbf{p}_{\text{norm}}^T \mathbf{p}_{\text{norm}}} \quad (5.2)$$

Na sequência, verifica-se a convergência utilizando-se, por exemplo, a soma dos quadrados da diferença dos elementos de dois vetores de escores consecutivos. Se a convergência foi atingida, passa-se para etapa de estimativa dos resíduos. Se não houve convergência, o algoritmo retorna à segunda etapa, cálculo do vetor peso, até que o valor predefinido para convergência seja atingido. Por fim calculam-se os resíduos:

$$\mathbf{E} = \mathbf{X} - \mathbf{t} \cdot \mathbf{p}^T \quad (5.3)$$

Assim, a matriz \mathbf{X} é representada pelo modelo de componentes principais como:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^T \quad (5.4)$$

onde \mathbf{T} e \mathbf{P} são as matrizes de escores e pesos, respectivamente.

O gráfico de escores, obtido pela PCA, em conjunto com o gráfico de resíduo versus influência (*leverage*) é uma ferramenta muito útil para a detecção de amostras anômalas. As

amostras com alto resíduo não são descritas adequadamente pelo modelo e aquelas com alto *leverage* influenciam o próprio modelo, fazendo com que algumas PCs sejam utilizadas apenas para descrever o comportamento de variáveis associadas a determinadas amostras. Portanto, amostras que apresentam alto resíduo e alta influência são consideradas como potenciais amostras anômalas (*outliers*).

2.5.2 SIMCA

É uma técnica supervisionada baseada em PCA, que modela o espaço multidimensional formado pelas amostras para definição da classe (BEEBE; PELL; SEASHOLTZ, 1998). A classificação de amostras pelo SIMCA contrasta com o método KNN, que utiliza apenas a distância das amostras. No SIMCA, uma amostra é classificada se estiver dentro da caixa multidimensional construída pelo modelo. Assim, cada classe é modelada independentemente e a classificação pode ser feita pela comparação da variância residual (Equação 6) da amostra a ser classificada com a variância residual total da classe (Equação 7) (PONTES, 2009).

$$S_i = \sqrt{\frac{\sum_{j=1}^p (\text{res}_j)^2}{p-A}} \quad (6)$$

$$S_t = \sqrt{\frac{\sum_{i=1}^N \sum_{j=1}^p (\text{res}_{ij})^2}{(N-A-1)(p-A)}} \quad (7)$$

Onde N é o número de amostras do conjunto de treinamento da classe; A é o número de componentes principais utilizadas pela classe; p é o número de variáveis; i e j representam os índices das amostras e variáveis, respectivamente.

A localização da amostra em relação ao modelo pode ser feita utilizando um teste F, onde o valor calculado na Equação 8 (F_{cal}) é comparado com um valor crítico (F_{crit}), determinado de modo empírico ou tabelado dentro de certos níveis de confiança e graus de liberdade (PONTES, 2009). Se o valor de F_{cal} da amostra investigada for menor que o F_{crit} , ela pertencerá à classe em consideração.

$$F_{\text{cal}} = \frac{(S_i)^2}{(S_t)^2} \cdot \frac{N}{N-A-1} \quad (8)$$

Como o modelo SIMCA pode classificar a amostra em mais de uma classe ou mesmo em nenhuma classe, é importante considerar a possibilidade de ocorrer os seguintes erros: erro

tipo I, onde a amostra não é classificada na classe correta ou erro tipo II, quando a amostra é classificada na classe errada.

2.5.3 LDA

Nesse método o espaço amostral é dividido em tantas regiões quanto o número de classes analisadas, através de uma fronteira linear que separa as classes. Para isso, são utilizadas funções lineares discriminantes que maximizam a variância entre as classes e minimizam a variância dentro da classe (FISHER, 1936). A classificação é feita de acordo com a região em que a amostra se encontra. Sendo considerada para tal o quadrado da distância Mahalanobis, conforme a Equação 9, sendo a amostra atribuída à classe de menor r^2 :

$$r^2(\mathbf{x}, \mathbf{m}_j) = (\mathbf{x} - \mathbf{m}_j)^T \cdot \mathbf{S}^{-1} \cdot (\mathbf{x} - \mathbf{m}_j) \quad (9)$$

onde r^2 é o quadrado da distância de Mahalanobis entre a amostra \mathbf{x} e o centro da j -ésima classe; \mathbf{m}_j e \mathbf{S} são as estimativas do vetor média e da matriz de covariância para a classe j (DUDA; HART; STORK, 2001), obtidas a partir do conjunto de treinamento de classificação conhecida (BEEBE; PELL; SEASHOLTZ, 1998).

Para o cálculo da matriz inversa da Equação 9 é necessário que o número de amostras seja maior que o número de variáveis incluídas no modelo LDA, para evitar que a covariância estimada, \mathbf{S} , seja singular (PONTES, 2009). Assim, para dados espectrais o uso de métodos de seleção de variáveis ou de compressão espectral é um passo indispensável na construção de modelos LDA.

2.6 Validação dos modelos de classificação

A etapa de validação dos modelos de classificação é indispensável para analisar a confiabilidade dos modelos e pode ser feita através da validação cruzada (*cross validation*), normalmente implementada no algoritmo do programa de modelagem, ou pelo uso de um conjunto de amostras validação externas (amostras de classe conhecida não pertencentes ao conjunto de treinamento). Um dos procedimentos que podem ser escolhidos é o da validação cruzada total (*full cross validation*), onde uma amostra do conjunto de treinamento é retirada e o modelo reconstruído sem essa amostra. Em seguida, a classificação da amostra retirada é feita por esse novo modelo. Essa amostra é reinserida no conjunto de treinamento e o processo se repete para cada uma das demais amostras.

2.7 Pré-processamento de dados

Uma etapa fundamental em qualquer análise quimiométrica é o pré-processamento de dados, já que, é nessa etapa que as fontes de variação irrelevantes e aleatórias, são eliminadas ou pelo menos reduzidas (BEEBE; PELL; SEASHOLTZ, 1998). As principais ferramentas para o pré-processamento das amostras são: normalização, suavização e correções da linha de base. Para as variáveis podem ser usadas as técnicas de centrar na média, autoescalonamento, entre outras. Neste trabalho, foram testadas as técnicas para correção da linha de base: derivadas, correção multiplicativa de sinal (MSC: *Multiplicative Signal Correction*) e variação normal padrão (SNV: *Standard Normal Variation*).

2.7.1 Derivadas

Além dos ruídos, as medidas espectrais podem conter fontes de variação de baixa frequência que não estão relacionadas com as informações químicas de interesse. Essas variações são chamadas de variações da linha de base. Dentre as diversas formas de realizar a correção na linha de base estão as derivadas, que são utilizadas especialmente quando é difícil identificar os pontos da linha de base. As derivadas evidenciam os picos presentes nos dados originais, facilitando sua visualização. Um dos métodos mais utilizados para o cálculo das derivadas é o método de suavização Savitzky-Golay e Gorry, onde um polinômio de baixa ordem é estimado a partir de uma região do vetor da amostra, um intervalo (janela) é selecionado e o ponto central desse intervalo é substituído pela derivada do polinômio estimado nesse ponto (BEEBE; PELL; SEASHOLTZ, 1998). Enquanto a primeira derivada remove o efeito aditivo, a segunda derivada elimina efeitos multiplicativos (RINNAN; BERG; ENGELSEN, 2009; SAVITZKY; GOLAY, 1964).

2.7.2 Correção Multiplicativa de Sinal (MSC: *Multiplicative Signal Correction*)

Desenvolvido para corrigir o problema de espalhamento da luz na espectroscopia de reflectância, o MSC calcula os coeficientes da regressão linear e o erro de cada espectro em relação a um espectro de referência (Equação 10), geralmente o espectro médio (RINNAN; BERG; ENGELSEN, 2009). Para obter o espectro corrigido é feita a subtração do espectro original pela divisão desses coeficientes, como mostra a Equação 11:

$$\mathbf{X}_{\text{original}} = b_0 + b_1 \cdot \mathbf{X}_{\text{referência}} + \mathbf{E} \quad (10)$$

$$\mathbf{X}_{\text{corrigido}} = \frac{\mathbf{X}_{\text{original}} - b_0}{b_1} \quad (11)$$

onde $\mathbf{X}_{\text{original}}$ é a matriz dos espectros medidos, $\mathbf{X}_{\text{referência}}$ é o espectro de referência utilizado para processar todo o conjunto de dados, e \mathbf{E} é a matriz de resíduos, $\mathbf{X}_{\text{corrigido}}$ é a matriz com os espectros corrigidos, b_0 e b_1 são os parâmetros escalares que variam de acordo com a amostra (RINNAN et al., 2009) e estimam os efeitos aditivos e multiplicativos, respectivamente, enquanto o erro (ou matriz de resíduos) pode ser relacionado com as informações químicas.

2.7.3 Variação normal padrão (SNV: Standard Normal Variation)

Assim como o MSC, o SNV também elimina efeitos aditivos e multiplicativos causados pelo espalhamento da luz, embora seja mais eficiente na correção dos efeitos aditivos. Para isso, o SNV emprega o valor médio e o desvio padrão de cada espectro como mostra a Equação 12 (BARNES; DHANOA; LISTER, 1989; GELADI; MACDOUGALL; MARTENS, 1985).

$$\text{SNV} = \frac{\sum_{i=1}^p (x_i - \bar{x})}{\sqrt{\frac{\sum_{i=1}^p (x_i - \bar{x})^2}{p-1}}} \quad (12)$$

Onde x é o sinal analítico no comprimento de onda i e \bar{x} é a média dos p comprimentos de onda da amostra, que é calculada de acordo com a Equação 13:

$$\bar{x} = \frac{\sum_{i=1}^p x_i}{p} \quad (13)$$

2.8 Seleção de amostras

As amostras de treinamento utilizadas na construção dos modelos de classificação devem ser escolhidas de tal modo que sejam as mais representativas para formar a classe. Para isso, existe o algoritmo *Kennard Stone* (KS) que seleciona essas amostras a partir da distância euclidiana entre os vetores das respostas experimentais (\mathbf{x}) (BOUVERESSE et al., 1996; KENNARD; STONE, 1969) selecionando o par de amostras com maior distância. Dentre as amostras restantes, busca-se aquelas com menor distância em relação as amostras já

selecionadas, sendo retida aquela com a maior distância mínima. Esse procedimento se repete até que o número de amostras determinado pelo usuário seja atingido. A distância euclidiana, $d_x(p,q)$, entre x vetores das amostras p e q é determinada pela Equação 14:

$$d_x(p,q) = \sqrt{\sum_{j=1}^J [x_p(j) - x_q(j)]^2} \quad ; p, q \in [1, N] \quad (14)$$

onde $x_p(j)$ e $x_q(j)$ são os vetores das respostas instrumentais no j -ésimo comprimento de onda do par de amostras p e q . (GALVÃO et al., 2005).

O algoritmo ainda permite que após a seleção das amostras de treinamento, as amostras remanescentes sejam divididas nos conjuntos de validação e teste, conforme determinação prévia pelo analista. A divisão das amostras em subconjuntos possibilita a construção de modelos com amostras representativas (treinamento) e com as melhores condições de modelagem (validação) para cada classe. Além de viabilizar a avaliação final dos modelos a partir de um conjunto real (previsão) de amostras independentes que não estiveram envolvidas no desenvolvimento dos modelos, garantindo a confiabilidade do método.

2.9 Seleção de variáveis

O tamanho da matriz de dados pode ser um problema na análise multivariada, seja pelo pequeno número de amostras, seja pela grande quantidade de variáveis. Essas condições restringem a escolha e a atuação do método de reconhecimento de padrões escolhido. A fim de minimizar esses efeitos, muitos algoritmos de seleção de variáveis têm sido desenvolvidos para evidenciar as variáveis importantes e suprimir aquelas irrelevantes ou redundantes. Além disso, a seleção de variáveis pode ser útil na eliminação de ruído e na viabilização do uso de técnicas como a regressão linear múltipla (MLR: *Multiple Linear Regression*) e Análise discriminante linear (LDA: *Linear Discriminant Analysis*).

A seleção de variáveis é definida classicamente como a “seleção de um subconjunto de M variáveis provenientes de um conjunto de N variáveis ($M < N$). Neste caso, uma função de custo é empregada para otimização.” (DASH; LIU, 1997).

Os algoritmos mais conhecidos e empregados com essa finalidade são o algoritmo genético (GA: *Genetic Algorithm*), o *stepwise* (SW), a regressão dos mínimos quadrados parciais por intervalos (iPLS: *Internal Partial Least Squares Regression*), o algoritmo *Jack-Knife* (JK) e o algoritmo de projeções sucessivas (SPA: *Succesive Projections Algorithm*).

Neste trabalho foi utilizado como método de seleção de variáveis o SPA-LDA, que é

uma adaptação do método SPA, originalmente desenvolvido para reduzir problemas de colinearidade em modelos de regressão linear múltipla (MLR: *Multiple Linear Regression*). O SPA, essencialmente, determina um grupo de variáveis representativas com o mínimo de multicolinearidade e o máximo de informação. (ARAÚJO et al., 2001). No SPA-LDA, uma função de custo é aplicada a fim de avaliar o risco médio de o modelo LDA realizar uma classificação incorreta. Para o cálculo dessa função de custo as amostras utilizadas para validar o modelo LDA são consideradas, como mostra a Equação 15 (PONTES, 2009):

$$G = \frac{1}{K_V} \sum_{k=1}^{K_V} g_k \quad (15)$$

onde K_V é o número de amostras do conjunto de validação e g_k é o risco de uma classificação incorreta do objeto X_k da k -ésima amostra de validação e é definido como:

$$g_k = \frac{r^2(X_k, m_k)}{\min_{j \neq k} r^2(X_k, m_j)} \quad (15.1)$$

onde o numerador da equação é o quadrado da distância Mahalanobis entre o objeto X_k e a média de sua classe (m_k). E o denominador é o quadrado da distância Mahalanobis entre o objeto e o centro da classe errada mais próxima (m_j). Assim, g_k deve ser o menor possível garantindo que o objeto esteja o mais próximo de sua verdadeira classe e o mais distante das classes erradas. Para minimizar a colinearidade o algoritmo utiliza uma sequência de projeções vetoriais aplicadas às colunas da matriz das amostras de treinamento. Como um de seus resultados, o algoritmo fornece o gráfico da função de custo versus o número de variáveis, sendo o número ótimo de variáveis o valor mínimo desta curva.

2.10 Figuras de mérito

O desempenho dos modelos de classificação pode ser avaliado por diferentes parâmetros, sendo os mais utilizados: a sensibilidade, a especificidade e a eficiência. Esses parâmetros são baseados em quatro respostas binárias possíveis: verdadeiro positivo (VP), quando o modelo de classificação dá uma resposta positiva a uma amostra que de fato é positiva; falso positivo (FP), quando o modelo classifica como positivo uma amostra negativa; verdadeiro negativo (VN), quando o modelo dá uma resposta negativa a uma amostra que é negativa e o falso negativo (FN), quando uma amostra positiva é classificada como negativa. Neste trabalho, o resultado foi considerado positivo quando a amostra era uma

cédula autêntica e negativo quando a cédula era falsificada.

A sensibilidade, que é a fração das amostras corretamente classificadas na classe modelada, é medida pelo nível de confiança do espaço das classes (OLIVERI; DOWNEY, 2012):

$$\text{sensibilidade} = \frac{VP}{VP + FN} \quad (16)$$

Enquanto a especificidade corresponde à habilidade do modelo de identificar corretamente amostras que não pertencem à classe modelada (OLIVERI; DOWNEY, 2012):

$$\text{especificidade} = \frac{VN}{VN + FP} \quad (17)$$

Eficiência representa a habilidade de classificação total do modelo (OLIVERI; DOWNEY, 2012):

$$\text{eficiência} = \sqrt{\frac{VP \cdot VN}{(VP+FN) \cdot (VN+FP)}} \quad (18)$$

3 MATERIAIS E MÉTODOS

3.1 Amostras

Considerando o caráter forense das amostras e, portanto, a necessidade de preservar sua integridade, nenhum tratamento prévio foi aplicado. Para cada valor estudado (R\$20, R\$50 e R\$100) foram analisadas 100 cédulas verdadeiras. As cédulas falsas, impressas em jato de tinta e de diferentes apreensões, foram disponibilizadas pela Polícia Federal (PF) de Pernambuco para a aquisição espectral, por questões legais e de segurança, nas dependências da PF. Dessas cédulas, foram analisadas, 56 de R\$20, 18 de R\$50 e 153 de R\$100. Todas as cédulas falsas foram analisadas por peritos da Polícia Federal que certificaram a fraude nas cédulas através de exames físicos e visuais baseados nos itens de segurança definidos pelo Banco Central do Brasil. Sete áreas (Figura 4), na frente e no verso, das notas de Real foram selecionadas para este estudo: duas fluorescentes (1 e 7), a marca d'água (2), três calcográficas (3,5 e 6) e uma tipográfica (4).

Figura 4 – Áreas analisadas nas cédulas: 1,7- fluorescentes; 2- marca d'água; 3,5,6- calcográficas, 4- tipográfica.



Fonte: Autoria própria

Além das cédulas, também foi analisada uma amostra de papel sulfite branco com gramatura de 75g/m^2 , para fins de comparação espectral com o papel moeda.

3.2 Equipamento e aquisição dos espectros

Os espectros no infravermelho próximo foram obtidos com o espectrômetro portátil, MicroNIR™1700 JSDU, de dimensões 45 mm x 42 mm (diâmetro e altura), em uma faixa espectral de 950-1650nm com resolução $< 1,25\%$ do centro de comprimento de onda e média de 50 varreduras, com tempo de integração de 5 ms. Uma base de teflon redonda com o mesmo diâmetro do micronir foi colocada embaixo da cédula e utilizada como branco durante a aquisição dos espectros. Em cima da cédula foi fixado um molde de plástico transparente (transparência para retroprojeção), com circunferências vazadas, no mesmo diâmetro do micronir, em cada uma das áreas analisadas a fim de garantir que cada aquisição espectral fosse realizada exatamente no mesmo local e, portanto, assegurar a reprodutibilidade das medidas. (Figura 5).

Figura 5 – Procedimento de aquisição espectral com o NIR portátil.



Fonte: Autoria própria

Cada espectro foi obtido em triplicata, sendo utilizada a média dessas aquisições para a construção dos modelos de classificação.

3.3 Tratamento quimiométrico dos dados

Os espectros foram processados utilizando o programa The Unscrambler® X, versão 10.4 (CAMO S.A.). Diferentes técnicas de pré-processamento de dados foram utilizadas para a correção de fontes de variação irrelevantes e aleatórias, tais como variação normal padrão (SNV), correção de sinal multiplicativo (MSC), primeira e segunda derivadas (filtro Savitzky-Golay, polinômio de segunda ordem e janelas com diferentes tamanhos).

Inicialmente, PCA foi aplicada para análise exploratória. Em seguida, foram

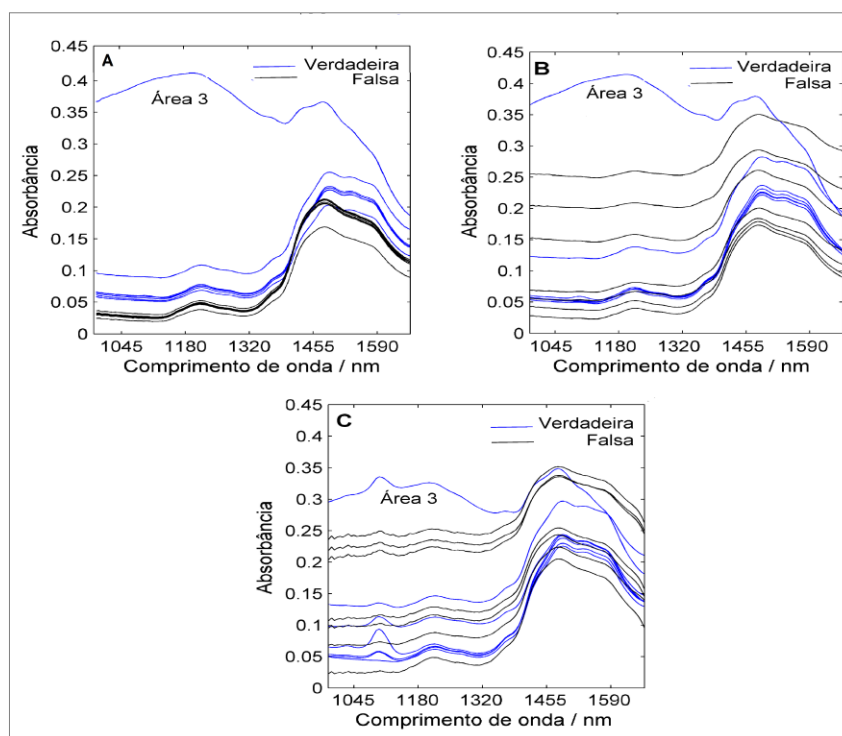
construídos modelos SIMCA apenas para as cédulas verdadeiras. O algoritmo *Kennard- Stone* (KS) foi utilizado para dividir essas amostras em grupos de treinamento (70% das amostras) e de previsão (30% das amostras). As cédulas falsas foram utilizadas como grupo de teste desses modelos. Para detecção de amostras anômalas (*outliers*) foram utilizados os gráficos de escores e o de resíduos versus influência (*leverage*). Como técnica de validação interna dos modelos foi utilizada a validação cruzada total *leave-one-out*. Também foram construídos modelos para cédulas falsas e verdadeiras utilizando SPA-LDA. Tanto no modelo das cédulas falsas, quanto no das cédulas verdadeiras os grupos de treinamento (50% das amostras), validação (25% das amostras) e teste (25% das amostras) foram divididos usando o KS. Os algoritmos KS e SPA-LDA foram aplicados utilizando o programa Matlab[®] R2010a 7.10.0.499.

4 RESULTADOS E DISCUSSÃO

4.1 Espectros NIR

Os espectros médios sem pré-processamento de todas as amostras em cada uma das sete áreas analisadas são mostrados na Figura 6. A diferença entre os espectros das cédulas autênticas e das falsas pode ser nitidamente observada.

Figura 6 – Espectros médios sem pré-processamento das áreas analisadas nas cédulas. (A) R\$20, (B) R\$50 e (C) R\$100.



Fonte: Autoria própria

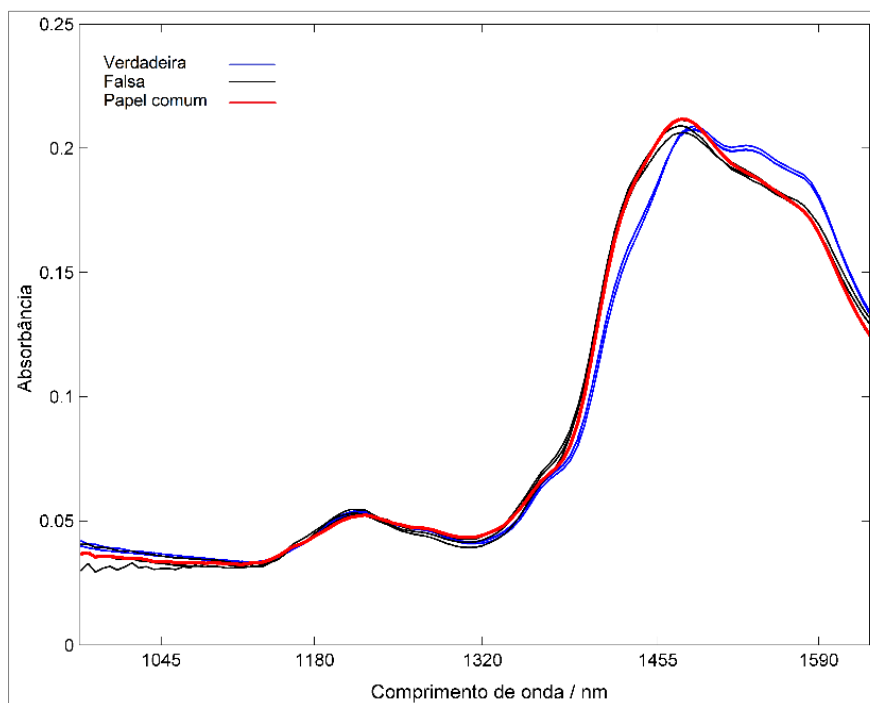
Considerando que o componente majoritário do papel moeda é a celulose, a atribuição das bandas foi feita com base no espectro NIR da celulose. As regiões espectrais entre 1212 e 1225 nm correspondem ao segundo sobretom de estiramento C-H. A banda com baixa intensidade em torno de 1366 nm está associada à combinação do primeiro sobretom do estiramento e deformação C-H. As bandas entre 1450 e 1650 nm estão associadas ao primeiro sobretom do estiramento O-H.

Também é possível visualizar na Figura 6 que existe uma variação na linha de base dos espectros devido ao efeito de espalhamento da radiação, o qual não está relacionado com a informação química e precisa ser removida usando técnicas de pré-processamento antes da construção dos modelos de classificação.

A comparação entre o espectro de uma amostra de papel sulfite e os espectros da

marca d'água, a área menos pigmentada, das cédulas autênticas e falsas é mostrada na Figura 7.

Figura 7 – Espectros médios sem pré-processamento do papel sulfite e da marca d'água das cédulas verdadeiras e falsas de R\$20, R\$50 e R\$100.



Fonte: Autoria própria

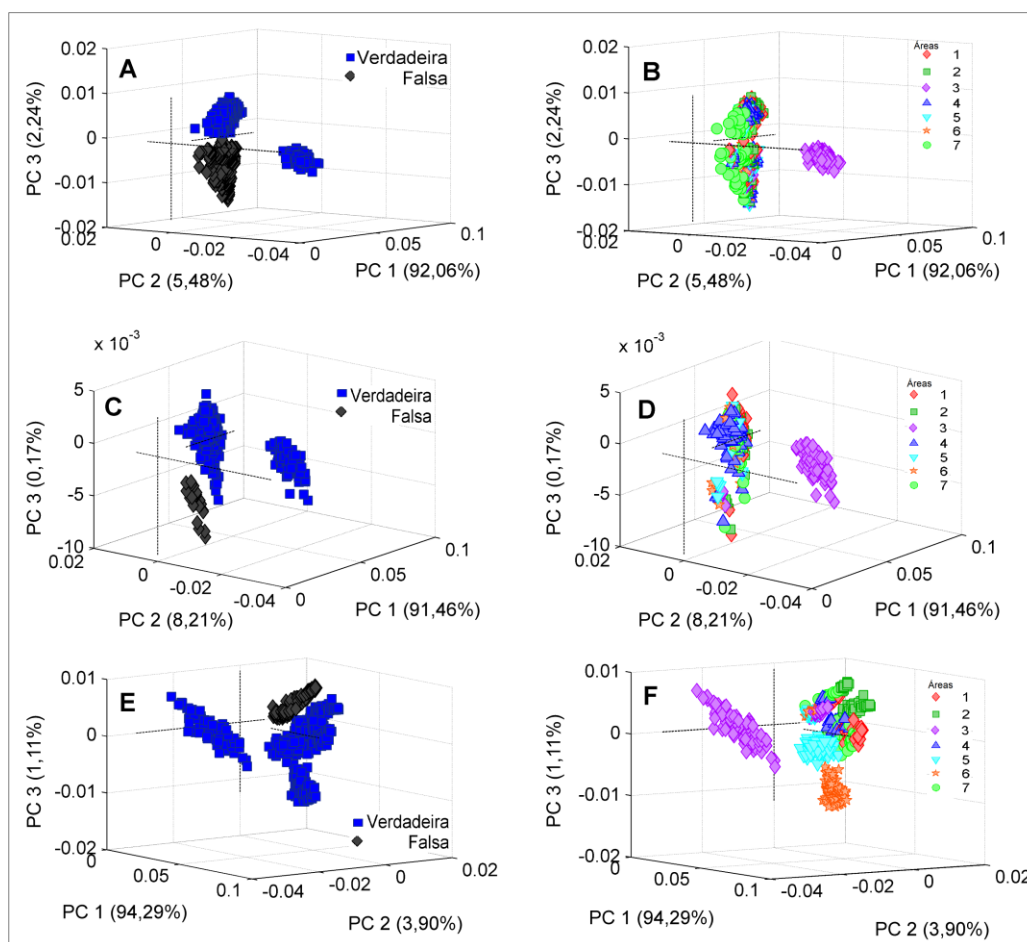
Apesar de a comparação ter sido feita entre um papel pigmentado, o papel moeda, e um papel branco, a Figura 7 mostra que ainda sim o espectro do papel sulfite é mais similar aos espectros das cédulas falsas, do que aos espectros das cédulas verdadeiras. Segundo Dale e Klatt (1999) é possível distinguir papel moeda de papel comum utilizando reflectância NIR difusa, o que comprova que eles têm propriedades químicas diferentes. Esse resultado sugere que as cédulas falsas são provavelmente feitas utilizando papel comum, enquanto as cédulas verdadeiras são impressas em papel especial, papel moeda, que tem uma composição diferente do papel comum.

Os espectros das áreas calcográficas 3 tiveram baixa reprodutibilidade porque a tinta nessa área não é homogênea (espectro mais diferente na Figura 6). As ranhuras presentes nessa área afetam a refletância, aumentando o espalhamento de radiação e, conseqüentemente, causando variações na linha de base dos espectros.

4.2 Análise exploratória dos dados espectrais

PCA foi, preliminarmente, aplicada usando todas as áreas a fim de identificar a área mais informativa para distinção entre cédulas falsas e autênticas. O gráfico de escores para cada tipo de cédula, R\$20, R\$50 e R\$100 é mostrado na Figura 8.

Figura 8 – Gráfico de escores para R\$20(A e B), R\$50(C e D), R\$100(E e F) empregando os espectros pré-processados com a primeira derivada.



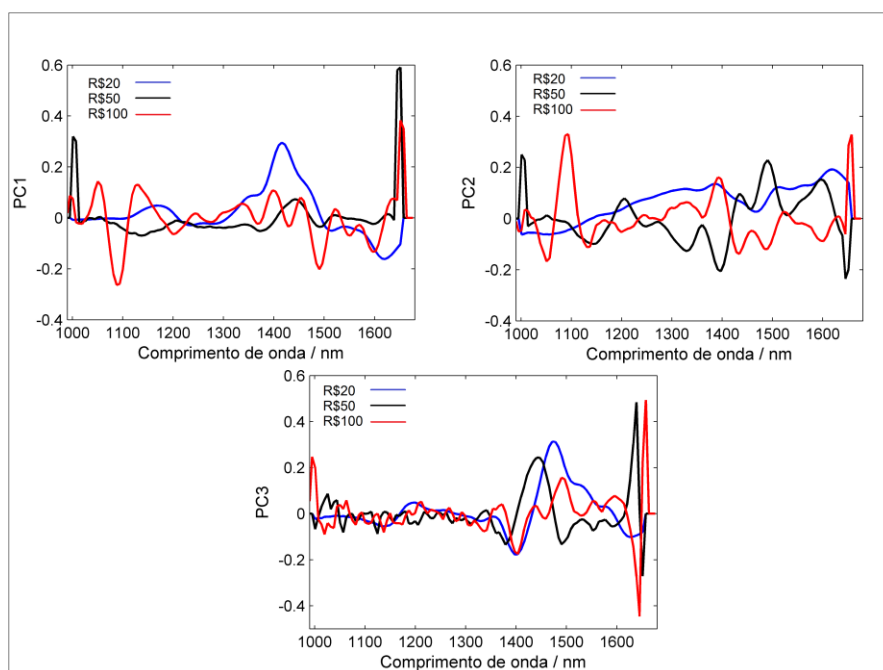
Fonte: Autoria própria

Como pode ser observado, a distinção entre as cédulas verdadeiras e falsas é evidente em todos os tipos de notas analisadas. Na Figura 8B, D e F é possível identificar um agrupamento da área 3, mostrando que essa área é diferente das outras e poderia ser suficiente para distinguir as amostras. Porém, como discutido anteriormente, essa área teve baixa reprodutibilidade e, portanto, não foi considerada para construção dos modelos SIMCA e SPA-LDA.

O gráfico dos pesos (*loadings*) é mostrado na Figura 9. Como se pode observar no gráfico de pesos da PC1, para as cédulas de R\$100 as principais variáveis responsáveis pela

separação das amostras estão em torno de 1100 nm, que correspondem às bandas características da tinta calcográfica. (MAEDER et al., 2013). As cédulas de R\$20 têm a maior variabilidade espectral na faixa 1400-1450 nm, relativa à banda atribuída ao estiramento O-H da celulose. E para as cédulas de R\$50 essa variabilidade ocorre em 1650 nm, ainda na região relativa ao estiramento O-H da celulose. Esse resultado é consistente, considerando que a composição do papel moeda é essencialmente celulose. Nos pesos da PC2 percebe-se que a variabilidade espectral ocorre novamente na região da tinta calcográfica das cédulas de R\$100. E no gráfico de pesos da PC3 a variabilidade espectral acontece na região atribuída ao estiramento O-H da celulose. Percebe-se, portanto, que as variáveis responsáveis pela distinção das cédulas falsas e verdadeiras estão tanto na região associada à tinta calcográfica quanto na região de estiramento O-H da celulose. Logo, as amostras podem ser diferenciadas por dois critérios, a presença da tinta calcográfica e o tipo de papel.

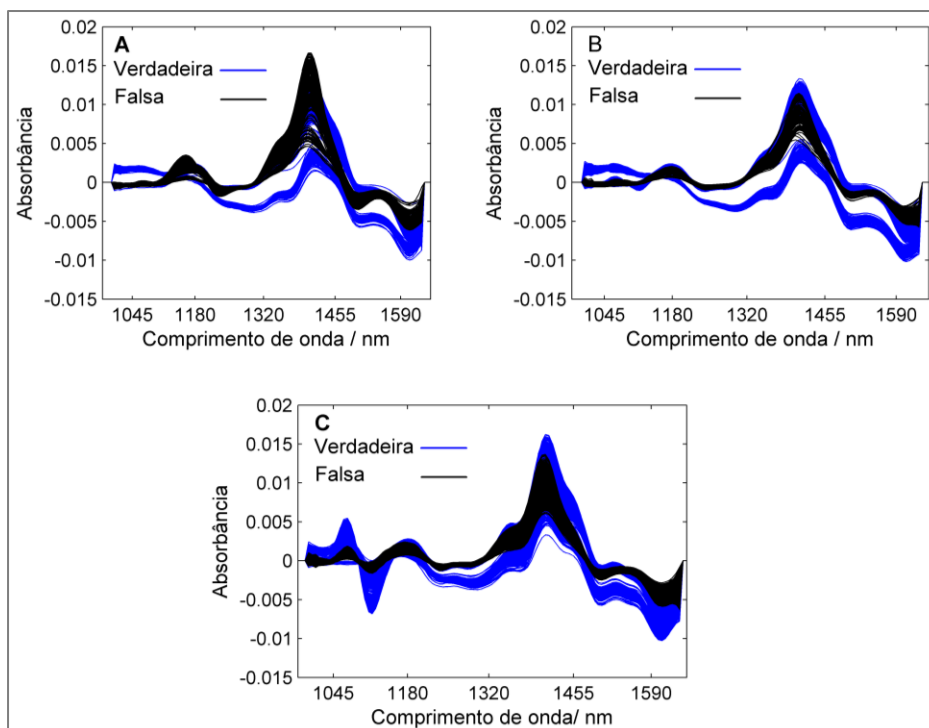
Figura 9 – Gráficos dos pesos para todas as áreas analisadas das cédulas de R\$20, R\$50 e R\$100.



Fonte: Autoria própria

Com esses resultados, pode-se afirmar que a área calcográfica (6) para as notas de R\$20 e R\$100, e a área calcográfica (5) para a de R\$50, foram suficientes para identificar as cédulas autênticas e diferenciá-las das falsas. Dentre todas as técnicas de pré-processamento testadas, a primeira derivada, com filtro Savitzky-Golay, polinômio de segunda ordem e janelas de 9 (R\$100) e 11 (R\$20 e R\$50) pontos, foi a mais efetiva na correção dos efeitos de espalhamento (Figura 10), de acordo com a visualização dos espectros.

Figura 10 – Espectro médio pré-processado de todas as áreas analisadas nas cédulas. (A) R\$20, (B) R\$50 e (C) R\$100.



Fonte: Autoria própria

4.3 Modelos de classificação

4.3.1 SIMCA

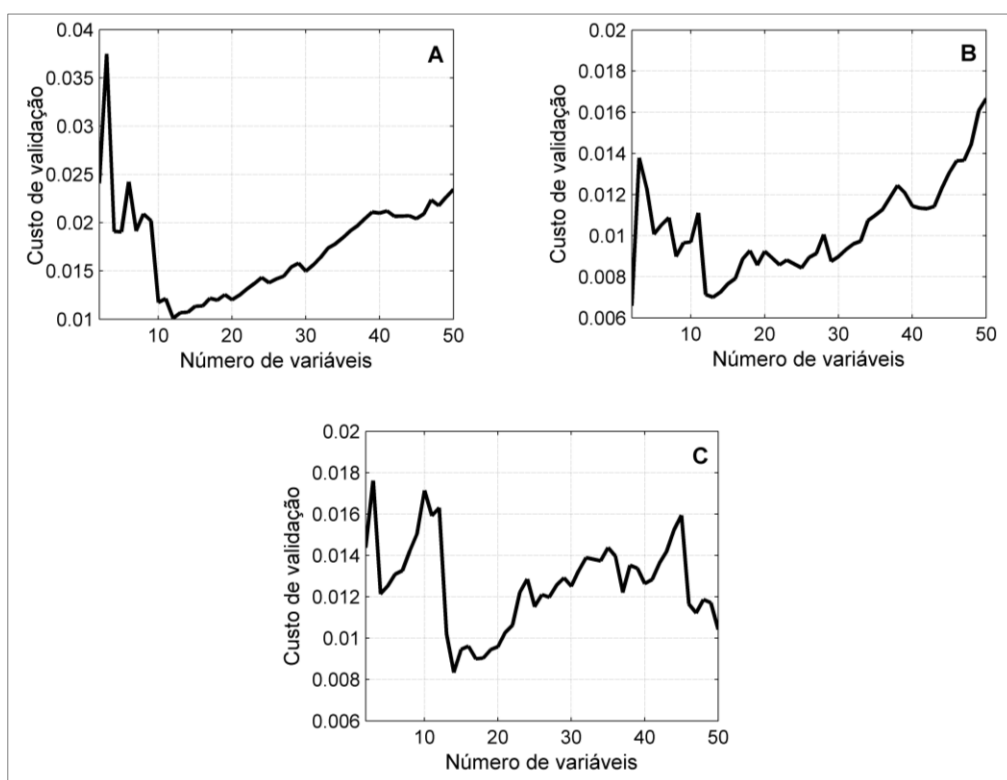
Os modelos SIMCA foram desenvolvidos empregando os espectros NIR das áreas selecionadas das cédulas autênticas. Para as amostras de R\$ 20 (área 6), o modelo requereu 5 PCs. Nesse modelo, as duas primeiras PCs explicaram 91% da variância total. O modelo para as cédulas de R\$50 (área 5) utilizou 4 PCs, sendo que as duas primeiras PCs explicaram 86% da variância dos dados. Para o modelo de R\$100 (área 6), foram utilizadas 5 PCs e as duas primeiras representaram 87% da variância total. Não foram encontradas amostras anômalas na construção dos modelos.

Usando os modelos SIMCA construídos para classificar as notas de R\$20, R\$50 e R\$100 como autênticas, no nível de confiança de 95%, todas as cédulas verdadeiras foram corretamente classificadas, ao passo que nenhuma cédula falsa foi classificada como autêntica. Portanto, não foram detectados erros do tipo I (amostra não incluída em sua classe) ou erros do tipo II (amostra incluída na classe errada).

4.3.2 SPA- LDA

Foram construídos modelos LDA para as classes verdadeira e falsa usando as variáveis selecionadas pelo SPA. O valor mínimo da função de custo G (PONTES et al., 2005) foi usado para determinar o número ótimo de variáveis nesses modelos. A Figura 11 mostra os valores mínimos das variáveis, 12, 2 e 14, bem localizados para as notas de R\$ 20, R\$ 50 e R\$ 100, respectivamente.

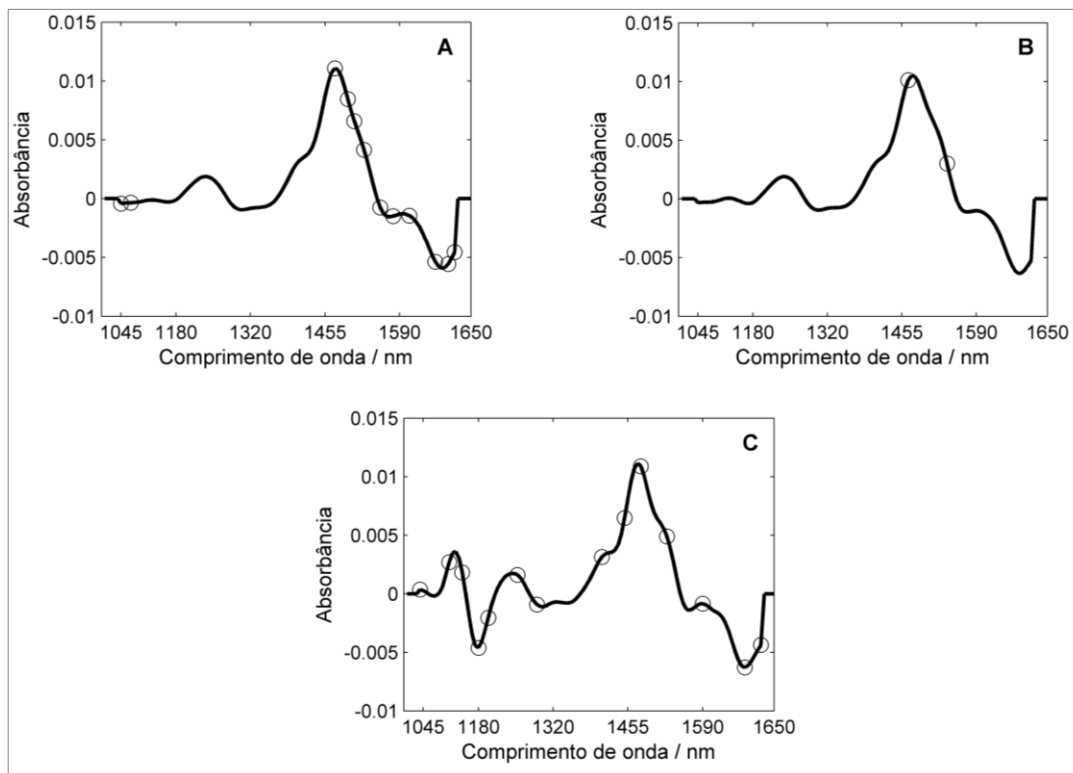
Figura 11 – Número ótimo de variáveis no SPA-LDA para (A) área 6 de R\$20, (B) área 5 de R\$50 e (C) área 6 de R\$100.



Fonte: Autoria própria

A Figura 12 mostra as variáveis selecionadas pelo SPA para cada tipo das cédulas analisadas. Os principais comprimentos de onda selecionados foram atribuídos ao estiramento O-H (primeiro sobretom) na região de 1450-1650nm. Especificamente, para as amostras de R\$ 100, a região entre 1000-1100 nm também foi selecionada, a qual corresponde à banda específica da tinta calcográfica.

Figura 12 – Variáveis selecionadas dos espectros médios para (A) área 6 R\$20, (B) área 5 R\$50 e (C) área 6 R\$100.



Fonte: Autoria própria

Embora os modelos SPA-LDA apresentem a vantagem de reduzir a não linearidade (GHASEMI-VARNAMKHAISTI et al., 2012) devido à seleção de variáveis, os modelos SIMCA são os mais adequados no estudo da autenticidade de cédulas, já que trata-se de um problema de modelagem de classes, uma vez que uma das classes de interesse, a classe das cédulas autênticas, pode ser eficientemente modelada de forma independente de qualquer outra classe. O que não é possível nos modelos LDA, em que devem existir no mínimo duas classes para modelagem.

4.4 Figuras de mérito

Os modelos SIMCA e SPA-LDA foram avaliados conforme os parâmetros expressos pelas Equações 16, 17 e 18 sendo obtidos 100% de sensibilidade, especificidade e eficiência. O que significa que os modelos classificaram corretamente as amostras nas classes modeladas identificaram corretamente as amostras que não pertenciam a uma determinada classe e que, portanto, foram eficientes na classificação total do modelo. Esses resultados demonstram a robustez do método analítico desenvolvido neste trabalho.

5 CONCLUSÃO

Neste trabalho foi apresentado um método analítico para o exame da autenticidade de cédulas de Real, R\$20, R\$50 e R\$100, empregando espectroscopia NIR com instrumento portátil e classificação das amostras através dos modelos quimiométricos SIMCA e SPALDA.

Com o método desenvolvido foi possível a classificação com 100% de sensibilidade, especificidade, eficiência, refletindo assim a robustez do método proposto para distinguir as cédulas verdadeiras das falsas.

O procedimento para avaliação da autenticidade de cada cédula durou cerca de 30 segundos, incluindo a aquisição dos espectros, pré-processamento dos dados e a classificação. A rotina dos modelos pode ser facilmente incorporada a um software amigável escrito em qualquer linguagem de programação, como Assembly, C++, Visual Basic, LabVIEW, entre outros. Não sendo necessário conhecimento técnico para interpretação da resposta dos modelos. Assim, a metodologia desenvolvida neste trabalho pode ser facilmente utilizada no comércio ou em bancos. Todavia, é necessário que um exame confirmatório da autenticidade seja realizado por peritos oficiais.

Além da efetividade, o método desenvolvido neste trabalho analisou uma característica ainda não explorada pelos equipamentos comerciais de verificação de autenticidade de cédulas, a tinta calcográfica, estando à frente de novas ações dos falsificadores. Assim, o método descrito é uma maneira simples, rápida, confiável e não destrutiva de avaliar a autenticidade de notas de Real e ainda pode ser utilizado na análise em campo, reduzindo o tempo de análise.

Como a calcografia está presente na maioria das moedas mundiais, o método analítico implementado neste trabalho tem potencial para ser aplicado de forma bem sucedida em Euros, Dólar e cédulas australianas, por exemplo. Contudo, é necessário o desenvolvimento de novos modelos, usando tanto cédulas autênticas quanto falsas da moeda em questão.

REFERÊNCIAS

- AGELET, L. E.; HURBURGH, C. R. A tutorial on near infrared spectroscopy and its calibration. **Critical Reviews in Analytical Chemistry**, v. 40, n. 4, p. 246–260, 2010.
- APPOLONI, C. R.; MELQUIADES, F. L. Portable XRF and principal component analysis for bill characterization in forensic science. **Applied Radiation and Isotopes**, v. 85, p. 92–95, fev. 2014.
- ARAÚJO, M. C. U. et al. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, v. 57, n. 2, p. 65–73, 2001.
- BANCO CENTRAL DO BRASIL. **Segunda Família do Real – Cartilha de Treinamento (2012)**. Disponível em: <<https://www.bcb.gov.br/novasnotas/assets/downloads/material-apoio/2e5/Cartilha.pdf>>. Acesso em: 30 mar. 2018.
- BANCO CENTRAL DO BRASIL. **Estatística de Falsificação - 2017**. Disponível em: <http://www.bcb.gov.br/htms/mecir/seguranca/EstatisticaFalsificacao_UFXDenominacao_internet_2017.pdf>. Acesso em: 30 mar. 2018.
- BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard normal variate transformation and de-trending of near-infrared diffuse reflectance spectra. **Applied Spectroscopy**, v. 43, n. 5, p. 772–777, 1989.
- BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, M. B. **Chemometrics a practical guide**. New York: John Wiley Y Sons, 1998.
- BOUVERESSE, E. et al. Standardization of near-infrared spectrometric instruments. **Analytical Chemistry**, v. 68, n. March, p. 982–990, 1996.
- BOŽIČEVIĆ, M. S.; GAJOVIĆ, A.; ZJAKIĆ, I. Identifying a common origin of toner printed counterfeit banknotes by micro-Raman spectroscopy. **Forensic Science International**, v. 223, n. 1–3, p. 314–320, 2012.
- CORREIA, R. M. et al. Portable near infrared spectroscopy applied to abuse drugs and medicine analyses. **Analytical Methods**, v. 10, p. 593–603, 2018.
- DA SILVA, N. C. et al. Standardization from a benchtop to a handheld NIR spectrometer using mathematically mixed NIR spectra to determine fuel quality parameters. **Analytica Chimica Acta**, v. 954, p. 32–42, 2017.
- DA SILVA, V. H.; DA SILVA, J. J.; PEREIRA, C. F. Portable near-infrared instruments: Application for quality control of polymorphs in pharmaceutical raw materials and calibration transfer. **Journal of Pharmaceutical and Biomedical Analysis**, v. 134, p. 287–294, 2017.
- DALE, J. M.; KLATT, L. N. Principal component analysis of diffuse near-infrared reflectance from paper currency. **Analysis**, v. 5, p. 123–129, 1999.
- DASH, M.; LIU, H. Feature selection for classification. **Intelligent data analysis**, 1997.

DE ALMEIDA, M. R. et al. Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation. **Microchemical Journal**, v. 109, p. 170–177, 2013.

DE SOUZA, A. M.; POPPI, R. J. Experimento didático de quimiometria para análise exploratória de óleos vegetais comestíveis por espectroscopia no infravermelho médio e análise de componentes principais: Um tutorial, parte I. **Química Nova**, v. 35, n. 1, p. 223–229, 2012.

DONG, Y.-W. et al. Determination of soil parameters in apple-growing regions by near- and mid- infrared spectroscopy. **Pedosphere**, v. 21, n. 5, p. 591–602, 2011.

DUDA, R.O.;HART, P.E.;STORK, D. G. **Pattern classification**. New York: John Wiley, 2001.

EBERLIN, L. S. et al. Instantaneous chemical profiles of banknotes by ambient mass spectrometry. **The Analyst**, v. 135, n. 10, p. 2533, out. 2010.

FARDIM, P.; FERREIRA, M. M. C.; DURÁN, N. Multivariate calibration for quantitative analysis of eucalypt kraft pulp by nir spectrometry. **Journal of Wood Chemistry and Technology**, v. 22, n. 1, p. 67–81, 2002.

FISHER, R. A. The use of multiple measurements in taxonomic problems. **Annals of Eugenics**, v. 7, n. 2, p. 179–188, 1936.

GALVÃO, R. K. H. et al. A method for calibration and validation subset partitioning. **Talanta**, v. 67, n. 4, p. 736–740, 2005.

GELADI, P.; MACDOUGALL, D.; MARTENS, H. Linearization and scatter-correction for NIR reflectance spectra of meat. **Applied Spectroscopy**, v. 39, n. 3, p. 491–500, 1985.

GHASEMI-VARNAMKHAHI, M. et al. Screening analysis of beer ageing using near infrared spectroscopy and the Successive Projections Algorithm for variable selection. **Talanta**, v. 89, p. 286–291, 2012.

IMPERIO, E. et al. Spectral investigations on 1000 £ banknotes throughout Italian Republic. **Vibrational Spectroscopy**, v. 79, p. 52–58, 2015.

KAO, Y.-Y. et al. Distinguishing authentic and counterfeit banknotes by surface chemical composition determined using electrospray laser desorption ionization mass spectrometry. **Journal of mass spectrometry : JMS**, v. 48, n. 11, p. 1129–35, 2013.

KAO, Y. Y. et al. Depth profiling of inks in authentic and counterfeit banknotes by electrospray laser desorption ionization/mass spectrometry. **Journal of Mass Spectrometry**, v. 51, n. 1, p. 20–27, 2016.

KAUPPINEN, A. et al. In-line multipoint near-infrared spectroscopy for moisture content quantification during freeze-drying. **Analytical Chemistry**, v. 85, n. 4, p. 2377–2384, 2013.

KENNARD, R. W.; STONE, L. A. Computer aided design of experiments. **Technometrics**, v. 11, n. 1, p. 137–148, 1969.

MAEDER, D. et al. IR-absorbing intaglio ink. **US pat. 8,362,130B2**, 2013.

MALEGORI, C. et al. Comparing the analytical performances of Micro-NIR and FT-NIR spectrometers in the evaluation of acerola fruit quality, using PLS and SVM regression algorithms. **Talanta**, v. 165, p. 112–116, 2017.

O'BRIEN, N. A.; HULSE, C. A.; FRIEDRICH, D. M. Miniature near-infrared (NIR) spectrometer engine for handheld applications. **Proceedings of SPIE**, v. 8374, p. 1–9, 2012.

OLIVERI, P.; DOWNEY, G. Multivariate class modeling for the verification of food-authenticity claims. **TrAC Trends in Analytical Chemistry**, v. 35, p. 74–86, maio 2012.

OTHONOS, A. et al. Laser photothermal diagnostics of genuine and counterfeit British and United States banknotes. **Optical Engineering**, v. 36, n. 2, p. 400–407, 1997.

PASQUINI, C. Near infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **Journal of the Brazilian Chemical Society**, v. 14, n. 2, p. 198–219, 2003.

PEREIRA, J. F. Q. et al. Evaluation and identification of blood stains with handheld NIR spectrometer. **Microchemical Journal**, v. 133, p. 561–566, 2017a.

PEREIRA, J. F. Q. et al. Projection pursuit and PCA associated with near and middle infrared hyperspectral images to investigate forensic cases of fraudulent documents. **Microchemical Journal**, v. 130, p. 412–419, 2017b.

PONTES, M. J. C. et al. The successive projections algorithm for spectral variable selection in classification problems. **Chemometrics and Intelligent Laboratory Systems**, v. 78, n. 1, p. 11–18, 2005.

PONTES, M. J. C. **Algoritmo das projeções sucessivas para seleção de variáveis espectrais em problemas de classificação**. [s.l.] Tese de doutorado, Paraíba: Universidade Federal da Paraíba, 2009.

RINNAN, Å.; BERG, F. VAN DEN; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **TrAC - Trends in Analytical Chemistry**, v. 28, n. 10, p. 1201–1222, 2009.

RÍOS-REINA, R. et al. NIR spectroscopy and chemometrics for the typification of Spanish wine vinegars with a protected designation of origin. **Food Control**, v. 89, p. 108–116, 2018.

RISOLUTI, R. et al. “Click and Screen” technology for the detection of explosives on human hands by a portable microNIR-chemometrics platform. **Analytical Chemistry**, v. 90, n. 7, p. 4288–4292, 2018.

RUSANOV, V. et al. Mössbauer and X-ray fluorescence measurements of authentic and counterfeited banknote pigments. **Dyes and Pigments**, v. 81, n. 3, p. 254–258, jun. 2009.

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

SCHMIDT, E. M. et al. Direct and non-destructive proof of authenticity for the 2nd generation of Brazilian real banknotes via easy ambient sonic spray ionization mass

spectrometry. **Science & Justice**, v. 54, n. 6, p. 459–464, dez. 2014.

SILVA, C. S. et al. Detecting semen stains on fabrics using near infrared hyperspectral images and multivariate models. **TrAC - Trends in Analytical Chemistry**, v. 95, p. 23–35, 2017.

SKOOG, D. A.; HOLLER, F.J.; CROUCH, S. R. **Princípios de análise instrumental**. 6. ed. Porto Alegre: Bookman, 2009.

SONNEX, E. et al. Identification of forged Bank of England £20 banknotes using IR spectroscopy. **Spectrochimica Acta - Part A: Molecular and Biomolecular Spectroscopy**, v. 118, p. 1158–1163, 2014.

UV Stuff. Disponível em: <<https://www.uvstuff.com/en/8-uv-ink>>. Acesso em: 30 mar. 2018.

VILA, A. et al. Development of a fast and non-destructive procedure for characterizing and distinguishing original and fake euro notes. **Analytica Chimica Acta**, v. 559, n. 2, p. 257–263, 2006.

WOLD, S.; ESBENSEN, K.; GELADI, P. Principal component analysis. **Chemometrics and Intelligent Laboratory**, v. 2, p. 37052, 1987.

APÊNDICE A - PUBLICAÇÃO DECORRENTE DO TRABALHO

OLIVEIRA,V. da S., HONORATO, R. S., HONORATO, F. A., PEREIRA, C. F. Authenticity assessment of banknotes using portable near infrared spectrometer and chemometrics. **Forensic Science International**, v.286, p. 121–127, 2018.