



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

**DISSERTAÇÃO DE MESTRADO**

**METODOLOGIAS ANALÍTICAS PARA A IDENTIFICAÇÃO DE NÃO  
CONFORMIDADES EM AMOSTRAS DE ÁLCOOL COMBUSTÍVEL**

**ADENILTON CAMILO DA SILVA**



***João Pessoa – PB - Brasil***  
***Agosto/2013***



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

**DISSERTAÇÃO DE MESTRADO**

**METODOLOGIAS ANALÍTICAS PARA A IDENTIFICAÇÃO DE NÃO  
CONFORMIDADES EM AMOSTRAS DE ÁLCOOL COMBUSTÍVEL**

**ADENILTON CAMILO DA SILVA\***

**Dissertação apresentada como  
requisito para obtenção do título  
de Mestre em Química pela  
Universidade Federal da Paraíba.**



**Orientador (a): Prof. Dr. MARCIO JOSE COELHO DE PONTES**

**\*BOLSISTA CNPq**

***João Pessoa – PB - Brasil***  
***Agosto/2013***

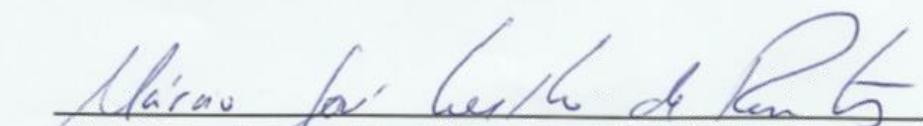
S586m Silva, Adenilton Camilo da.  
Metodologias analíticas para a identificação de não conformidades em amostras de álcool combustível / Adenilton Camilo da Silva.- João Pessoa, 2013.  
82f. : il.  
Orientador: Márcio José Coelho de Pontes  
Dissertação (Mestrado) – UFPB/CCEN  
1. Química. 2. AEHC. 3. Espectroscopia no infravermelho.  
4. Voltametria. 5. Classificação multivariada. 6. Seleções de variáveis.

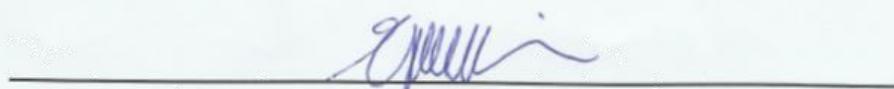
UFPB/BC

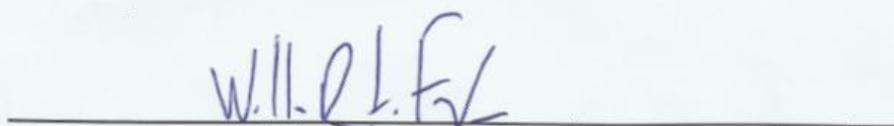
CDU: 54(043)

# **“Metodologias Analíticas para a Identificação de não Conformidades em Amostras de Álcool Combustível”.**

Dissertação de Mestrado de Adenilton Camilo da Silva aprovada pela banca examinadora em 27 de agosto de 2013:

  
Prof. Dr. Márcio José Coelho de Pontes  
Orientador/Presidente

  
Prof. Dr. Everaldo Paulo de Medeiros  
Examinador

  
Prof. Dr. Wallace Duarte Fragoso  
Examinador

A **Deus**, rei de amor e misericórdia. Que me permite o dom da vida a cada dia. Dando-me paz, alegrias e realizações.

A minha querida mãe, **Maria do Socorro**, que com todo amor e coragem, enfrentou os desafios da vida para dar o melhor aos seus filhos, a educação.

A minha eterna amada **Diva Glaucy** (*in memoriam*), que por graça de Deus, alcancei o privilégio de conhecer, amar, e guardar seu amor e seus ensinamentos.

COM MUITO CARINHO E GRATIDÃO,

**DEDICO.**

## **AGRADECIMENTOS**

Aos meus irmãos, **Adailton e Aline**, que são presença firme de Deus na minha vida, e que estão sempre ao meu lado, torcendo pela minha vitória.

Aos meus queridos amigos, **José Wanderley e Glaucia Maria e sua família**, por terem amado com tanta intensidade minha querida e amada **Diva Glaucy** (*in memoriam*), e por me amarem como filho.

Aos meus avós, **Anésio** (*in memoriam*) e **Francisca** (*in memoriam*), que cuidaram com tanto amor e carinho de seus filhos e netos, mostrando-os caminhos da retidão, amor e honestidade.

Ao professor orientador, **Márcio José Coelho de Pontes**, por ser um grande amigo, que teve paciência e dedicação ao me orientar. Sou extremamente grato pela confiança, à alegria de querer ver seus orientandos bem, por contribuições valiosas para minha vida e principalmente compreensão nos momentos em que precisei de tempo.

**Aos meus familiares e amigos**, que são presentes valiosos dados por Deus, e que torceram por mim, apoiaram, ampararam nos momentos difíceis, e tiveram paciência e compreensão com meus estudos.

Aos amigos do **LAQA**, pelo o apoio nos momentos difíceis, pelas alegrias, conversas e contribuições valiosas. Grandes guerreiros, coordenados pelo grande **Dr. Mário Cesar Ugulino**, o qual também sou verdadeiramente grato.

Ao amigo **Eduardo Matos**, que durante sua iniciação científica, contribuiu grandiosamente para a realização deste trabalho, de forma parceira, paciente e buscando sempre ajudar.

A prof<sup>a</sup>. Dr<sup>a</sup>. Maria Fernanda Pimentel (UFPE / LAC) por fornecer as amostras de AEHC empregados neste estudo.

Ao **CNPq** pelo apoio financeiro e financiamento da bolsa de pesquisa.

Agradeço de coração a todos que contribuíram e contribuem com a minha caminhada. Este trabalho é nosso! Muito obrigado!

“Foi pela sabedoria que o Senhor criou a terra, foi com inteligência que ele formou os céus. Foi pela ciência que se fenderam os abismos, por ela as nuvens destilam o orvalho. Meu filho, guarda a sabedoria e a reflexão, não as percas de vista. Elas serão a vida de tua alma e um adorno para teu pescoço. Então caminharás com segurança, sem que o teu pé tropece. Se te deitares, não terás medo. Uma vez deitado, teu sono será doce. Não terás a recear nem terrores repentinos, nem a tempestade que cai sobre os ímpios, porque o Senhor é tua segurança e preservará teu pé de toda cilada.” (Provérbios 3, 19:26)

## **SUMÁRIO**

<b>ÍNDICE DE FIGURAS .....</b>	<b>iii</b>
<b>ÍNDICE DE TABELAS .....</b>	<b>vii</b>
<b>LISTA DE ABREVIATURAS E SIGLAS .....</b>	<b>viii</b>
<b>RESUMO.....</b>	<b>x</b>
<b>ABSTRACT .....</b>	<b>xi</b>
<b>CAPÍTULO I – INTRODUÇÃO .....</b>	<b>01</b>
1.1. Etanol combustível .....	02
1.2. Adulteração de AEHC.....	04
1.3. Objetivos gerais.....	09
1.3.1. Objetivos específicos.....	09
<b>CAPÍTULO II – FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>10</b>
2.1. Espectroscopia no infravermelho .....	11
2.2. Voltametria cíclica e uso de eletrodo de cobre .....	15
2.3. Quimiometria .....	18
2.3.1. Análise de componentes principais (PCA) .....	20
2.3.2. Análise discriminante pelos mínimos quadrados parciais (PLS-DA) .....	22
2.3.3. Análise discriminante linear (LDA).....	24
2.3.3.1. Seleção de variáveis.....	26
<b>3. CAPÍTULO III – EXPERIMENTAL .....</b>	<b>29</b>
3.1. Amostras.....	30
3.2. Aquisição dos espectros no infravermelho (NIR e MIR) ....	30
3.3. Aquisição dos voltamogramas .....	32
3.4. Procedimentos quimiométricos e software .....	33
<b>4. CAPÍTULO IV – RESULTADOS E DISCUSSÃO .....</b>	<b>37</b>
4.1. Estudo com espectroscopia no infravermelho .....	38
4.1.1. Avaliação dos espectros NIR e MIR.....	38
4.1.2. Análise exploratória dos dados de infravermelho.....	40
4.1.3. Validação dos modelos de classificação.....	43
4.1.3.1. Seleção de variáveis e validação dos modelos LDA	

.....	43
4.1.4. Uso dos modelos de classificação para previsão .....	47
4.2. Estudo com voltametria cíclica .....	52
4.2.1. Avaliação dos voltamogramas .....	52
4.2.2. Análise exploratória dos dados voltamétricos .....	55
4.2.3. Validação dos modelos de classificação .....	57
4.2.3.1. Seleção de variáveis e validação dos modelos LDA .....	57
4.2.4. Uso dos modelos de classificação para Previsão .....	60
<b>CAPÍTULO V – CONCLUSÕES .....</b>	<b>63</b>
5.1. Conclusões .....	64
5.2. Propostas futuras .....	65
<b>CAPÍTULO VI – REFERÊNCIAS BIBLIOGRÁFICAS .....</b>	<b>68</b>
<b>Anexos .....</b>	<b>78</b>

## ÍNDICE DE FIGURAS

<b>Figura 1.1</b> - Índices de não conformidade de combustíveis líquidos brasileiros .....	4
<b>Figura 2.1</b> - Espectro de radiação eletromagnética e distribuição do infravermelho .....	11
<b>Figura 2.2</b> - Curvas de energia potencial de uma molécula diatômica: modelo do oscilador harmônico .....	12
<b>Figura 2.3</b> - Curvas de energia potencial de uma molécula diatômica: (1) modelo do oscilador harmônico e (2) modelo do oscilador anarmônico...	13
<b>Figura 2.4</b> - (a) Sinais de excitação de tensão <i>versus</i> tempos empregados na voltametria cíclica, (b) registro da corrente <i>versus</i> potencial (voltamograma).....	16
<b>Figura 2.5</b> - Voltamograma cíclico para eletrodo metálico de cobre em solução de 0.1 mol/L de NaOH, faixa de potencial -1,0 a 0,70 V vs Ag / AgCl (3,0 mol L <sup>-1</sup> de KCl).....	16
<b>Figura 2.6</b> - Plano bidimensional com duas variáveis e representação dos dois primeiros Componentes Principais .....	21
<b>Figura 2.7</b> – Representação da organização dos dados para realização de PLS-DA para duas classes.....	23
<b>Figura 2.8</b> – Representação da organização dos dados para realização de PLS-DA para três ou mais classes .....	23
<b>Figura 2.9</b> – Esquema de discriminação entre duas classes (A e B), baseando-se nas variáveis originais ( $y_1$ e $y_2$ ), ou na linha discriminatória. ....	24
<b>Figura 3.1</b> - Espectrofotômetro FT-IR da Bomem (FTLA 2000-160) utilizado para o registro dos espectros NIR .....	31

- Figura 3.2** - Espectrofotómetro FTIR Varian 640 utilizado para o registro dos espectros do MIR.....31
- Figura 3.3** - Potenciostato Autolab PGSTAT 101 (Eco Chemie) utilizado para aquisição dos voltamogramas. ....32
- Figura 4.1** - Espectros registrados na região (a) NIR e (b) MIR.....38
- Figura 4.2** - Espectros na região do NIR (7449 a 3769  $\text{cm}^{-1}$ ) para as classes de amostras: (a) não adulteradas, (b) adulteradas com água e (c) adulteradas com metanol .....39
- Figura 4.3** - Espectros MIR (3500 a 698  $\text{cm}^{-1}$ ) para as classes de amostras: (a) não adulteradas, (b) adulteradas com água e (c) adulteradas com metanol. ....40
- Figura 4.4** - Gráfico dos escores da PCA aplicada aos espectros NIR. (a) PC1xPC2 e (b) PC2xPC3. (■: amostras não adulteradas, ●: amostras adulteradas com água, ▲: amostras adulteradas com metanol) .....41
- Figura 4.5** - Gráficos dos pesos da PCA (PC1, PC2 e PC3) aplicada aos espectros NIR das amostras de AEHC.....41
- Figura 4.6** - Gráfico de escores da PCA (PC1xPC2) aplicada aos espectros MIR originais. (■: amostras não adulteradas, ●: amostras adulteradas com água, ▲: amostras adulteradas com metanol) .....42
- Figura 4.7** - Gráficos dos pesos da PCA aplicada aos espectros MIR das amostras de AEHC .....43
- Figura 4.8** - Gráficos de custo para seleção da cadeia de variáveis pelo SPA nos dados (a) NIR e (b) MIR. ....44
- Figura 4.9** - Espectros médios com as variáveis selecionadas pelo SPA para os dados (a)NIR e (b) MIR .....44

- Figura 4.10** - Espectros médios com as variáveis selecionadas pelo SW para os dados (a)NIR e (b) MIR .....46
- Figura 4.11** - Espectros médios com as variáveis selecionadas pelo GA para os dados (a)NIR e (b) MIR .....47
- Figura 4.12** - Gráfico dos escores (FD1×FD2) para o conjunto completo de espectros NIR utilizando variáveis selecionadas pelo (a) SW, (b) SPA e (c) GA. (■: amostras não adulteradas, ●: amostras adulteradas com água,▲: amostras adulteradas com metanol). .....50
- Figura 4.13** - Gráfico dos escores (FD1×FD2) para o conjunto completo de espectros MIR utilizando variáveis selecionadas pelo (a) SW, (b) SPA e (c) GA. (■:amostras não adulteradas, ●: amostras adulteradas com água,▲: amostras adulteradas com metanol) .....51
- Figura 4.14** – Voltamogramas cíclicos das amostras de AEHC, registrados na janela de potencial de - 0,9 V a 0,8 V ..... 52
- Figura 4.15** – Voltamogramas cíclicos das amostras de AEHC, no sentido anódico (0.4 a 0.8 V), para amostras (a) não adulteradas e (b) adulteradas com metanol ..... 53
- Figura 4.16** – Voltamogramas cíclicos das amostras de AEHC, com remoção de áreas não informativas, para amostras (a) não adulteradas e (b) adulteradas com metanol ..... 53
- Figura 4.17** – Voltamogramas das amostras de AEHC na região de oxidação de moléculas de álcool, para amostras (a) não adulteradas e (b) adulteradas com metanol .....54
- Figura 4.18** – Gráficos dos escores pela PCA aplicada aos Voltamogramas das amostras de AEHC. (a) PC1×PC2 e (b) PC1×PC3 (■: amostras não adulteradas, ▲: amostras adulteradas com metanol) .....55

- Figura 4.19** – Gráficos dos pesos da PCA aplicada aos voltamogramas cíclicos das amostras de AEHC. (PC1: —, PC2: —, PC3: —●—).....56
- Figura 4.20** – Gráficos de custo para seleção da cadeia de variáveis pelo SPA nos dados de voltametria .....57
- Figura 4.21** – Voltamograma médio com as variáveis selecionadas pelo SPA para os dados de voltametria.....58
- Figura 4.22** – Voltamograma médio com as variáveis selecionadas pelo SW para os dados de voltametria .....59
- Figura 4.23** – Voltamograma médio com as variáveis selecionadas pelo SPA para os dados de voltametria.....60
- Figura 4.24** - Gráfico dos escores (FD1) para o conjunto completo de voltamogramas utilizando variáveis selecionadas pelo (a) SW, (b) SPA e (c) GA. (■: amostras não adulteradas, ▲: amostras adulteradas com metanol) .....62

## **ÍNDICE DE TABELAS**

<b>Tabela 3.1</b> – Número de amostras de treinamento, validação e teste em cada classe para os dados de espectroscopia no Infravermelho .....	34
<b>Tabela 3.2</b> – Número de amostras de treinamento, validação e teste em cada classe para os dados de voltametria. ....	34
<b>Tabela 4.1</b> – Resultados de classificação apresentados pelo algoritmo Stepwise (SW) para o conjunto de validação .....	45
<b>Tabela 4.2</b> – Resultados de classificação dos modelos SW-LDA, SPA-LDA, GA-LDA e PLS-DA aplicados aos espectros NIR para o conjunto de teste. ....	48
<b>Tabela 4.3</b> – Resultados de classificação dos modelos SW-LDA, SPA-LDA, GA-LDA e PLS-DA aplicados aos espectros MIR para o conjunto de teste. ...	49
<b>Tabela 4.4</b> – Resultados de classificação apresentados pelo algoritmo Stepwise (SW) para o conjunto de validação usando os dados de voltametria. ....	59
<b>Tabela 4.5</b> – Resultados de classificação dos modelos SW-LDA, SPA-LDA, GA-LDA e PLS-DA, aplicados aos voltamogramas para o conjunto de teste. ....	61

**SIGLAS E ABREVIATURAS**

Proálcool	Programa Nacional do Álcool
AEAC	Álcool Etílico Anidro Combustível
AEHC	Álcool Etílico Hidratado Combustível
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
NBR	Normas Brasileiras
ABNT	Agência Brasileira de Normas Técnicas
ASTM	American Society for Testing and Materials
PMQC	Programa de Monitoramento e Qualidade dos Combustíveis
PMQL	Programa de Monitoramento e Qualidade dos Lubrificantes
NIR	Near-Infrared
MIR	Middle-Infrared
PCA	Principal Component Analysis
PLS	Partial Least Square
FTIR	Fourier Transform Infrared
FIR	Far-Infrared
LDA	Linear Discriminant Analysis
PLS-DA	Partial Least Squares Discriminant Analysis
VC	Voltametria Cíclica
LED	Light Emitting Diode
APS	Algoritmo das Projeções Sucessivas
SPA	Successive Projections Algorithm
GA	Genetic Algorithm
SW	Stepwise
LAC	Laboratório de Combustíveis
UFPE	Universidade Federal de Pernambuco
ATR	Attenuated Total Reflectance

KS	Kennard Stones
PC	Principal Component
E <sub>vib</sub>	Energia vibracional
v <sub>v</sub>	Velocidade de Varredura
t <sub>n</sub>	Vetor de Scores
p <sub>n</sub>	Vetor de Loadings

## RESUMO

No Brasil, uma das formas de comercialização do etanol combustível é na forma hidratada (AEHC - Álcool Etílico Hidratado Combustível). As adulterações encontradas nas amostras de AEHC são preocupantes, pois podem gerar prejuízos fiscais e à sociedade. Dentro dessa perspectiva, este trabalho propõe o desenvolvimento de novas metodologias analíticas baseadas no uso da espectroscopia no infravermelho (próximo - NIR e médio - MIR) e Voltametria Cíclica (com eletrodo cobre), em conjunto com técnicas quimiométricas de reconhecimento de padrões, visando à identificação das adulterações de AEHC com água ou metanol. Um total de 184 amostras de AEHC, coletadas de diferentes postos de combustíveis foram analisadas. Estas amostras foram divididas em três classes: (1) não adulteradas; (2) adulteradas com água (0,5% a 10,0%) e (3) adulteradas com metanol (2,0% a 13,0% m.m<sup>-1</sup>). A análise por componentes principais (PCA) foi aplicada aos dados, sendo possível verificar, principalmente, uma tendência à formação de agrupamentos das classes de amostras não adulteradas e adulteradas. Modelos de classificação foram baseados na análise discriminante linear (LDA) com prévia seleção de variáveis pelos algoritmos: SPA (Algoritmo das projeções sucessivas), GA (Algoritmo genético), SW (Stepwise). A técnica de PLS-DA (Análise discriminante pelos mínimos quadrados parciais) também foi aplicada nos dados. Avaliando os espectros MIR, 100% de acerto de classificação foram alcançados com todos os modelos. Para os dados NIR, utilizando SPA-LDA e SW-LDA houve uma taxa de classificação correta (TCC) de 84,4% e 97,8%, respectivamente, enquanto em PLS-DA e GA-LDA classificaram-se corretamente todas as amostras de teste. Na avaliação dos dados voltamétricos, tanto o SPA-LDA como o PLS-DA alcançaram uma TCC de 93%, mas os modelos GA-LDA e SW-LDA apresentaram melhores resultados, classificando corretamente todas as amostras de teste. Portanto, os métodos propostos são alternativas promissoras para a identificação, de forma rápida e segura, de adulteração em amostras de AEHC com água ou metanol.

**Palavras Chave:** AEHC; Espectroscopia no Infravermelho; Voltametria; Classificação Multivariada; Seleção de Variáveis.

## ABSTRACT

In Brazil, ethanol fuel is marketed in the hydrated form (HEAF– Hydrated Ethyl Alcohol Fuel). The adulterations found in HEAF can generate fines, and possible risks to society. With this perspective, this work proposes developing new analytical methods based on the use of infrared spectroscopy (NIR and MIR), and Cyclic Voltammetry (copper electrode), and chemometric pattern recognition techniques, to identify HEAF adulterations (with water or methanol). A total of 184 HEAF samples collected from different gasoline stations were analyzed. These samples were divided in three classes: (1) unadulterated, (2) adulterated with water (0.5% to 10%mm<sup>-1</sup>), and (3) adulterated with methanol (2% to 13% mm<sup>-1</sup>). Principal Components Analysis (PCA) was applied, permitting verification of a tendency to form clusters for unadulterated and adulterated samples. Classification models based on Linear Discriminant Analysis (LDA), with variable selection algorithms: SPA (Successive Projections Algorithm), GA (Genetic Algorithm), and SW (Stepwise) were employed. PLS-DA (Discriminant Analysis by Partial Least Squares) was applied to the data. Assessing the MIR spectra, 100% correct classification was achieved for all models. For NIR data, SPA-LDA and LDA-SW achieved a correct classification rate (RCC) of 84.4%, and 97.8%, respectively, while PLS-DA and GA-LDA correctly classified all test samples. In the evaluation of voltammetric data, as SPA-LDA as PLS-DA achieved a 93% RCC, but the GA-LDA and SW-LDA models showed better results, correctly classifying all test samples. The results suggest that the proposed methods are promising alternatives for identifying HEAF samples adulterated with water or methanol both quickly and securely.

**Keywords:** HEAF; Infrared Spectroscopy; Voltammetry; Multivariate Classification; Variables Selection.

Capítulo I

# **INTRODUÇÃO**

## 1. INTRODUÇÃO

### 1.1. Etanol combustível

O etanol é frequentemente usado como matéria-prima para bebidas, cosméticos, produtos químicos e farmacêuticos, e também como combustíveis. <sup>[1]</sup> A utilização do etanol como combustível para motores traz uma série de vantagens, tais como: redução da dependência de combustíveis derivados do petróleo, diversificação da matriz energética e, ainda, contribuição para diminuição dos impactos ambientais. <sup>[2]</sup>

O etanol combustível é produzido em diferentes países com diferentes matérias-primas. O Brasil ocupa um lugar de destaque no cenário mundial e tem a cana-de-açúcar como sua principal matéria prima. <sup>[3-4]</sup> O uso da cana-de-açúcar como matéria-prima para produção de etanol traz uma série de vantagens como: possuir um processo de produção mais econômico que outras matérias-primas <sup>[5]</sup>, apresenta um alto rendimento em termos de volume percentual de etanol obtido por área plantada, e ainda, vantagens em termos de balanço energético de produção quando comparado à produção de álcool combustível à base de milho. <sup>[6]</sup>

A inclusão do etanol combustível na matriz energética brasileira iniciou-se na década de 70, através do Proálcool (Programa Nacional do Álcool), como uma alternativa aos combustíveis derivados de petróleo, que apresentaram uma alta variação do preço, mediante a crise do petróleo ocorrida naquele período. <sup>[7-8]</sup>

Boa parte da frota de veículos de carga leve do Brasil tem seu funcionamento parcial ou totalmente à base de etanol combustível, o qual pode ser encontrado na forma anidra (álcool etílico anidro combustível - AEAC), misturado à gasolina com um percentual obrigatório, atualmente, de vinte e cinco por cento (25%), ou na forma hidratada (álcool etílico

hidratado combustível - AEHC), sendo utilizado em automóveis de motor dedicado ou ainda em carros *bicombustíveis*, que utilizam tanto a gasolina quanto álcool como combustível. [9-10] De acordo com a Portaria ANP nº 45/01 [11], para efeito de comercialização, os teores mínimos de etanol para AEHC e AEAC são, respectivamente, 92,5% e 99,3% (m/m).

No Brasil, a produção de etanol no mês de dezembro de 2012 foi de 320 milhões de litros de etanol, sendo 50% desse volume referente a etanol anidro e a outra metade referente ao etanol hidratado. Outro dado que vale a pena salientar é o fato de que, desde o ano de 2007 até o ano dezembro de 2012, os carros *bicombustíveis* têm representado mais de 80% de veículos leves licenciados no país. [12]

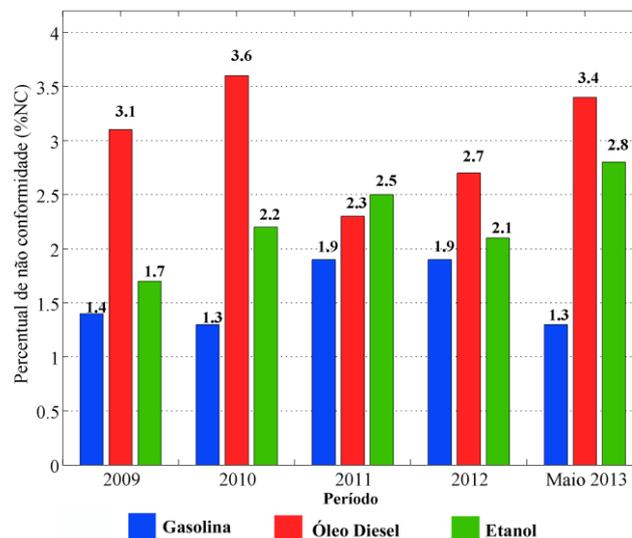
Mediante a grande demanda de utilização do etanol como combustível, os órgãos reguladores têm gerado diretrizes para fiscalizar a qualidade do mesmo. No Brasil, o controle de produção e comercialização do etanol combustível é realizado pela agência nacional do petróleo, gás natural e biocombustíveis (ANP), que tem a função de proteger os interesses dos consumidores quanto ao preço, qualidade e oferta de produtos, bem como especificar a qualidade dos derivados de petróleo, gás natural e seus derivados e dos biocombustíveis. [13]

Para comercialização do etanol combustível em suas diferentes aplicações é necessário que, em todo o território nacional, os diversos agentes econômicos que comercializam o produto sigam as instruções contidas na resolução ANP nº 7, de 9 de fevereiro de 2011. [14] Além disso, é necessário que o etanol combustível, seja ele anidro ou hidratado, esteja dentro das especificações contidas no regulamento técnico ANP nº 3/2011 [14], o qual se baseia em métodos de ensaio nacionais (NBR – Normas Brasileiras da ABNT – Agência Brasileira de Normas Técnicas) e internacionais (ASTM - *American Society for Testing and Materials*).

## 1.2. Adulteração de AEHC

A adulteração é o ato criminoso pelo qual se transforma um produto, modificando suas características originais, inviabilizando a garantia do produto. Em particular, a adulteração dos combustíveis se caracteriza principalmente pela adição irregular de qualquer substância, sem recolhimento de impostos, com vistas à obtenção de lucro. [15]

Dentro dessa perspectiva, a ANP mantém dois programas para monitorar constantemente a conformidade dos combustíveis líquidos (gasolina, etanol e óleo diesel) e dos óleos lubrificantes comercializados: o programa de monitoramento da qualidade dos combustíveis líquidos – PMQC e o programa de monitoramento da qualidade dos lubrificantes – PMQL. Os resultados obtidos no PMQC e no PMQL são importantes ferramentas que geram uma política de controle de qualidade de combustíveis junto à população, bem como orientam a fiscalização e ações de diferentes órgãos. [16] Segundo dados fornecidos pelo PMQC de maio de 2013 [17], o etanol e o óleo diesel apresentaram, nos últimos anos, as maiores taxas de não conformidades de todos os combustíveis monitorados no Brasil (**Figura 1.1**).



**Figura 1.1** - Índices de não conformidade de combustíveis líquidos brasileiros.

As não conformidades apresentadas pelo AEHC podem está relacionadas a adulterações, as quais podem trazer uma série de prejuízos, que em termos gerais estão relacionados a danos aos veículos automotores, danos ambientais e à saúde, sonegação de impostos e tributos. Tais condições violam o dever do fornecedor em garantir a qualidade do produto que é comercializado. As principais formas de falsificação do AEHC aparecem com a adição ilegal de água ao etanol ou, em casos mais graves, a substituição do etanol pelo metanol.

A adulteração de AEHC por adição de água consiste em misturar este solvente, que é mais barato e de fácil acesso, acima do valor máximo permitido, que equivale a 4,9% (v/v), segundo regulamento técnico nº 3/2011, presente na resolução nº 7 da ANP. <sup>[14]</sup> Este tipo de fraude ocorre, pois as propriedades da água e do AEHC permitem que os mesmos estejam misturados em diferentes proporções sem que a adulteração seja perceptível visualmente. Como danos gerados pela adição de água ao AEHC estão: a perda de potência do motor e aumento no consumo de combustível.

A adulteração por substituição do etanol por metanol consiste em mascarar a quantidade de etanol do combustível por meio da adição de metanol ao mesmo. O valor máximo permitido de metanol equivale a 1% (v/v), segundo a resolução nº 7 da ANP. <sup>[14]</sup> Este tipo de fraude ocorre, pois as propriedades físico-químicas do etanol e metanol são bem semelhantes, como: ambos são incolores, solúveis em água e com valores de densidades próximos. <sup>[18]</sup> Tais parâmetros permitem que os mesmos estejam misturados sem percepção visível, e não ocorra modificação expressiva no teor alcoólico do meio, permitindo a diminuição do volume de etanol sem que seja perceptível através da avaliação dos métodos convencionais. Ainda, não incide sobre o metanol impostos como combustível.

O metanol é extremamente tóxico, podendo causar sérios problemas de saúde, tais como: dor de cabeça, náuseas, vômitos, cegueira e até morte. [19] Por isso, é contra a lei utilizá-lo como combustível no Brasil.

Segundo o regulamento técnico ANP nº 3/2011 [14], os métodos de ensaio citados para determinação do teor de água em AEHC são: NBR 15531 [20], NBR 15888 [21] e ASTM E203 [22], os quais se baseiam no método de Karl Fisher, tanto volumétrico quanto coulométrico. No que diz respeito à determinação do teor de metanol em amostras de AEHC, as seguintes normatizações foram propostas recentemente para esse fim: ABNT NBR 16041 [23] e ASTM D5501 [24], as quais propõem o uso de cromatografia gasosa para determinação dos teores de metanol.

Apesar de bastante difundidos, os métodos de ensaio apresentados para determinação de água ou metanol em AEHC possuem alguns inconvenientes, como: complexidade dos reagentes para o método Karl Fischer, e ainda, aparecimento de reações de competição durante as determinações. [25] Já o método de cromatografia gasosa requer que a amostra esteja isenta de impurezas, o equipamento é de alto custo, com demanda de tempo, e especialistas para execução dos ensaios.

Em face do exposto, torna-se importante o desenvolvimento de métodos analíticos rápidos e precisos que identifiquem com segurança a autenticidade em amostras de AEHC, bem como a detecção de adulteração com metanol e/ou água.

A espectroscopia na região no infravermelho próximo (NIR: near infrared) associada aos métodos de análise multivariada vem sendo bastante utilizada no desenvolvimento de novas metodologias analíticas que possam substituir os métodos convencionais de análise, auxiliando no controle de qualidade de amostras de biocombustíveis / combustíveis. [26-29] Trata-se de uma técnica eficiente, rápida, não destrutiva, com baixo uso reagente, e que se diferencia dos métodos clássicos que normalmente

necessitam de padrões e reagentes caros, apresentam complexidade nas determinações e alto custo de operação e manutenção.

Mendes et. al. (2003) [26] utilizaram as espectroscopias NIR e RAMAN em combinação com a regressão pelos mínimos quadrados parciais (PLS: *Partial Least Square*) para determinar o teor de álcool etílico em amostras de AEHC. Segundo os autores, o modelo construído com espectros NIR apresentou um desempenho equivalente ao método de referência (ASTM D4052).

A espectroscopia NIR foi também utilizada para a determinação simultânea de etanol e metanol em amostras de gasolinas. [27] Para esse propósito, os autores utilizaram a faixa de 1105 a 1682 nm para a aplicação dos modelos PLS. Segundo os autores os resultados foram satisfatórios. Na literatura consultada existem poucos trabalhos envolvendo a espectroscopia no infravermelho e análise multivariada para propósito de investigação de adulteração de amostras de AEHC com metanol e/ou água. [28-29] Destaca-se o trabalho de Carneiro e colaboradores [28], o qual as espectroscopias MIR (*mid infrared*) e NIR foram utilizadas, em combinação com modelos PLS, para determinação desses solventes em amostras de AEHC. Os autores fizeram uso de uma técnica multivariada quantitativa (PLS), mas nenhum estudo envolvendo métodos de reconhecimento de padrões/classificação e/ou seleção de comprimentos de ondas representativos foi apresentado.

Outra técnica que merece destaque é a voltametria cíclica, sobretudo quando empregada como ferramenta para estudos fundamentais, fornecendo informações qualitativas sobre os processos eletroquímicos. Na voltametria cíclica, a espécie eletroativa em solução é oxidada ou reduzida na superfície do eletrodo de trabalho quando um potencial apropriado é aplicado para forçar a transferência de elétrons, resultando em uma corrente que é registrada. A partir de um potencial inicial, o potencial é variado com velocidade constante até um potencial final, e

então retorna com a mesma velocidade ao potencial inicial. A corrente é registrada em função do potencial elétrico aplicado entre os eletrodos de trabalho e referência, e dessa forma, sua ilustração gráfica corresponde ao voltamograma. [30]

À medida que os equipamentos foram se aperfeiçoando, as técnicas eletroanalíticas ganharam uma grande importância no contexto da Química Analítica [31]. Contudo, estas técnicas apresentam, em sua grande maioria, aplicações voltadas à análise univariada. Um perfil voltamétrico, por exemplo, que apresenta sobreposições de picos analíticos que, em geral, parece ser um problema na análise univariada, podem ser úteis quando métodos baseados em análise multivariada (classificação ou calibração multivariada, por exemplo) são utilizados para extrair o máximo de informações relevantes.

Na literatura, poucos trabalhos são encontrados envolvendo métodos eletroanalíticos com ou sem quimiometria com o intuito de identificação de adulteração de amostras de AEHC com metanol [32] e/ou água [33], sendo que, apenas uma referência para cada tipo de adulteração pode ser destaca. O trabalho apresentado por Pereira e colaboradores [32], visou determinar metanol em etanol por meio do uso de voltametria cíclica empregando eletrodo de ouro em meio alcalino. Os valores de correntes obtidas em dois potenciais distintos foram selecionados para construção de duas curvas de calibração univariadas. Já Bueno e colaboradores [33] buscaram discriminar adulteração de AEHC com água para o desenvolvimento de um eletrodo interdigitado de cobre e medidas de capacitância associados a ferramentas quimiométricas de reconhecimento de padrões não supervisionadas (PCA e HCA). Ainda, os scores da PCA foram utilizados para construção de uma curva de calibração univariada para quantificação das adulterações. Nenhum dos trabalhos apresentados na literatura propôs o uso de métodos eletroanalíticos e quimiometria para reconhecimento de padrões supervisionado e/ou seleção de potenciais representativos para classificar amostras adulteradas de AEHC.

### 1.3. Objetivos Gerais

Desenvolver novas metodologias analíticas baseadas em medidas no infravermelho (NIR e MIR) e por voltametria cíclica em conjunto com técnicas de reconhecimento de padrão, para a identificação de não conformidades em álcool etílico hidratado combustível (AEHC).

#### 1.3.1. Objetivos Específicos

- Realizar um estudo exploratório empregando a análise de componentes principais (PCA) aplicada aos dados espectrais NIR e MIR e de voltametria cíclica com intuito de identificar padrões de amostras de AEHC não adulteradas e adulteradas.
- Construir modelos de classificação a partir dos espectros e voltamogramas usando análise discriminante pelos mínimos quadrados parciais (PLS-DA), e em análise discriminante linear (LDA) com diferentes estratégias de seleção de variáveis (comprimentos de onda ou potenciais) utilizando: algoritmo das projeções sucessivas (SPA), algoritmo genético (GA) e stepwise (SW).
- Comparar o desempenho dos modelos construídos com seleção de variáveis (SPA-LDA, GA-LDA e SW-LDA) com aqueles obtidos através dos espectros NIR e MIR ou voltamograma completo (PLS-DA) em função da taxa de classificação correta (TCC) para o conjunto de Teste.

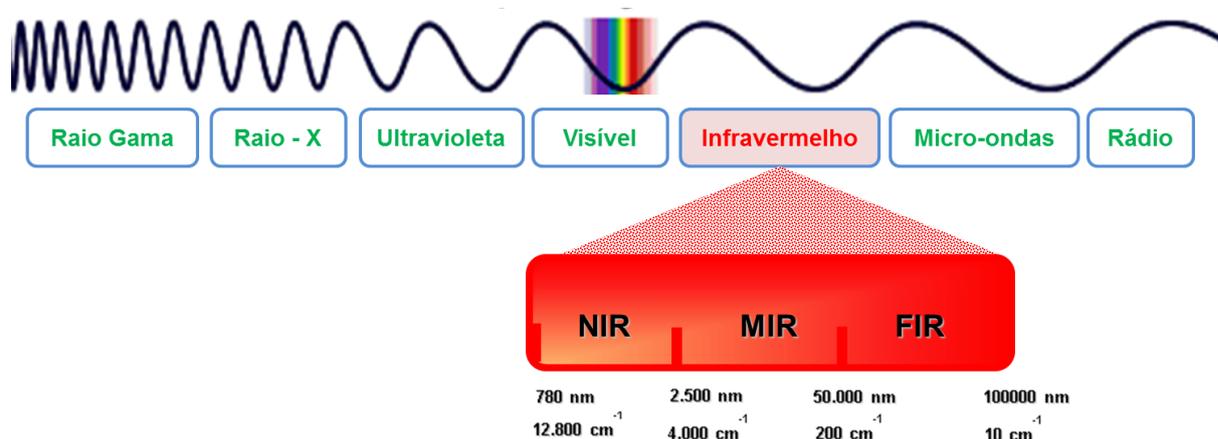
Capítulo II

# **FUNDAMENTAÇÃO TEÓRICA**

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. Espectroscopia no infravermelho

A espectroscopia na região do infravermelho é uma técnica analítica que as frequências das transições vibracionais das moléculas. [34] A região do espectro eletromagnético para transições vibracionais compreende a faixa de 12.800 a 10  $\text{cm}^{-1}$ , a qual compreende: próximo (NIR), médio (MIR) e distante (FIR), como apresentado na **Figura 2.1**. [34]

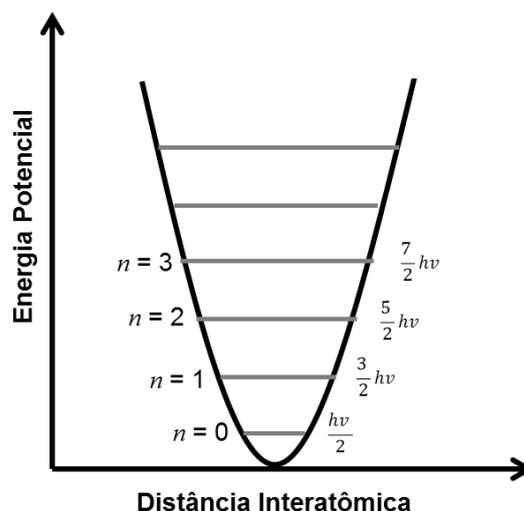


**Figura 2.1** – Espectro de radiação eletromagnética e distribuição do infravermelho.

O modelo mais simples para estudar vibrações de moléculas diatômicas é considerar que o comportamento vibracional é similar ao de um oscilador harmônico (**Figura 2.2**) que, pela mecânica quântica, apresenta os níveis de energia vibracional quantizados, conforme a **Equação 2.1**: [34-36]

$$E_{vib} = hv \left( n + \frac{1}{2} \right) \quad (2.1)$$

Onde: em que  $E_{vib}$  é a energia vibracional,  $h$  é a constante de Plank,  $v$  é a frequência vibracional clássica e  $n$  é o número quântico vibracional.



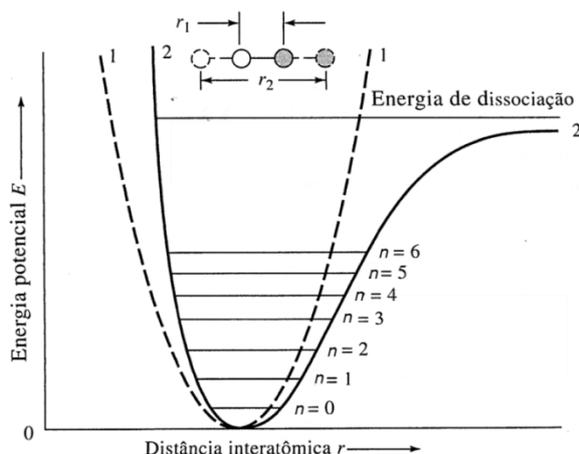
**Figura 2.2-** Curvas de energia potencial de uma molécula diatômica: modelo do oscilador harmônico.

Na **equação 2.1** observa-se um resultado importante da mecânica quântica, para  $n=0$  o valor da energia não é zero, mas tem o valor  $hv/2$ , conhecido como *energia do estado fundamental* ou energia do ponto zero. Existe ainda uma regra de seleção para o oscilador harmônico, que são os níveis de energia equiespaçados, permitindo apenas transições para  $\Delta n$  é  $\pm 1$ . [35]

O modelo do oscilador harmônico (**Figura 2.2**) é simplificado, apresentando alguns inconvenientes, como o fato de não prevê a dissociação da molécula, e não apresentar a existência de transições com  $\Delta n$  de  $\pm 2, \pm 3, \dots$  que é um fenômeno observado em alguns espectros. Dentro dessa perspectiva, a vibração molecular diatômica pode ser comparada com o oscilador anarmônico (**Figura 2.3**), que apresenta os níveis de energia vibracional, conforme a **Equação 2.2**:

$$E_{vib} = hv \left( n + \frac{1}{2} \right) - hvw \left( n + \frac{1}{2} \right)^2 \quad (2.2)$$

Onde: onde  $E_{vib}$  é a energia vibracional,  $h$  é a constante de Plank,  $v$  é a frequência vibracional clássica e  $n$  é o número quântico vibracional e  $w$  a constante de anarmonicidade. [35-38]



**Figura 2.3-** Curvas de energia potencial de uma molécula diatômica: (1) modelo do oscilador harmônico e (2) modelo do oscilador anarmônico. Adaptado de: SKOOG, D. A., et. al (2009) [36]

No modelo do oscilador anarmônico (**Figura 2.3**) a regra de seleção muda para  $\Delta n = \pm 1, \pm 2, \pm 3 \dots$ , ainda, é possível observar que ao aumentando no número quântico vibracional  $n$ , a energia dos níveis diminui. Além disso, nota-se neste modelo, uma energia de dissociação, não existente no oscilador harmônico.

Quando a molécula vibra, pode haver transição para outros estados energéticos, sendo a diferença de energia observada nestas transições que origina as bandas espectrais. A absorção no infravermelho é um fenômeno no qual só ocorre absorção de radiação que tenha exatamente a mesma energia da transição dos níveis energéticos. No caso das moléculas diatômicas, a condição para que haja absorção da radiação infravermelha é que uma molécula apresente uma variação no momento de dipolo, tal fenômeno não ocorre em moléculas homonucleares, como  $O_2$ ,  $N_2$  ou  $Cl_2$ . [36-37]

O número de possíveis modos de vibração normais para uma molécula linear é  $3N - 5$ , e para uma molécula não linear é  $3N - 6$ , em que  $N$  é o número de átomos na molécula. Não surpreendentemente, os espectros de infravermelho muitas vezes mostram um número considerável de bandas de absorção. [36-38] Quando o número de bandas

obtidas é menor que o número esperado, podem está ocorrendo vibrações sem variação do momento de dipolo ou a presença de energias de vibrações idênticas. Caso contrário, existam mais bandas que o esperado, pode está ocorrendo a presença de sobretons e bandas de combinação. [37-39]

A espectroscopia no infravermelho (IR) é uma técnica analítica não destrutiva que permite a determinação direta, rápida e com confiança de várias propriedades da amostra sem pré-tratamento. Cada região do IR apresentam características específicas, como é o caso do MIR e NIR.

No MIR (4000 a 200  $\text{cm}^{-1}$ ) ocorrem às vibrações fundamentais, que ocorrem quando as variações nos níveis vibracionais ( $\Delta n$ ) são iguais a  $\pm 1$ . Normalmente essas variações são intensas, com melhor coeficiente de absorção e definição de bandas, que permite a boa análise das estruturas das moléculas. Já o NIR (12.800 a 4000  $\text{cm}^{-1}$ ) está relacionado aos sobretons e combinações. Os sobretons ocorrem quando as variações nos níveis vibracionais ( $\Delta n$ ) são diferentes de  $\pm 1$ , e apresentam intensidades cerca de 10 a 1000 vezes menores que as observadas para as transições fundamentais. Já as combinações ocorrem quando um fóton excita simultaneamente dois estados vibracionais diferentes. A frequência da banda de combinação é aproximadamente igual à soma ou à diferença entre as duas frequências das bandas fundamentais que interagem para a sua formação. Nesse sentido o NIR apresenta menor definição de bandas, com um perfil mais voltado à análise quantitativa. [35-39]

Os sinais de baixa intensidade gerados por matrizes complexas, a dificuldade em se elaborar uma correlação satisfatória entre grupos de átomos envolvidos na molécula, bandas de interpretação difícil e de alta complexidade, dentre outros; [38] fazem com que o uso da espectroscopia no infravermelho, sobretudo na região do NIR, dependa da utilização de métodos quimiométricos multivariados na maioria das aplicações analíticas.

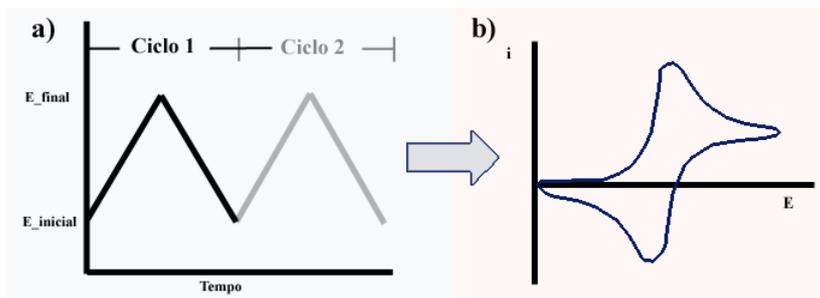
## 2.2. Voltametria cíclica e uso do eletrodo de cobre

A voltametria cíclica (VC) é uma técnica amplamente utilizada para identificação de sistemas redox, compreensão de intermediários de reação e para verificar estabilidade de produtos de reação. [40]

Um conjunto de vantagens é apresentado pela voltametria cíclica, como: determinações diretas e com baixo consumo de amostra, curto tempo na realização das análises, baixo custo da instrumentação e dos materiais utilizados, robustez diante da presença de partículas de sólidos dispersas. [41]

Os métodos voltamétricos são caracterizados, em geral, pela aplicação de um potencial (E) a um eletrodo de trabalho, sendo que o potencial varia sistematicamente com o tempo, enquanto a resposta de corrente é medida. Diferentes condições de variação do potencial no tempo podem ser utilizadas, as quais são conhecidas como sinais de excitação. [42]

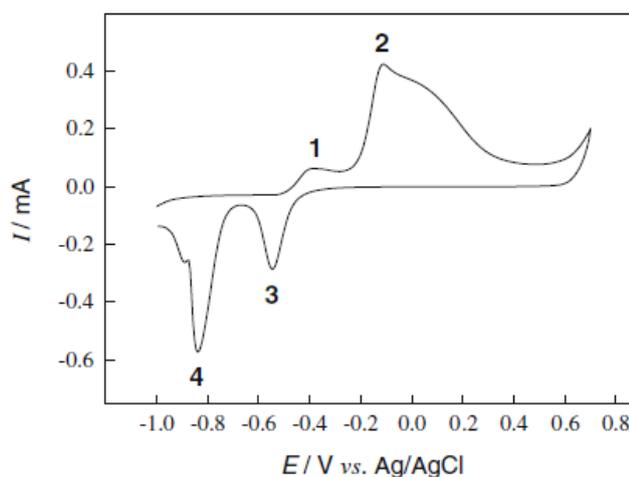
Na voltametria cíclica, a forma de controlar o potencial no tempo é a forma triangular. [42] Nesse tipo de sinal de excitação, o potencial aplicado em um eletrodo varia linearmente entre um valor inicial e um valor final, retornando para o valor inicial em certa velocidade de varredura, enquanto um voltamograma (perfil de corrente que evolui no sistema em função do potencial) é registrado (**Figura 2.4**). Através da avaliação do voltamograma é possível extrair características e informações qualitativas importantes sobre as espécies analisadas.



**Figura 2.4** – (a) Sinais de excitação de tensão *versus* tempos empregados na voltametria cíclica, (b) registro da corrente *versus* potencial (voltamograma).

Na literatura é possível encontrar trabalhos que utilizam a voltametria cíclica para identificar o mecanismo associado à oxidação de etanol e compostos orgânicos em superfície do eletrodo de cobre, para diferentes amostras, como: açúcares, polióis e ácidos carboxílicos [43-44], etanol em bebidas [45], álcool na respiração [46], etanol em amostras de gasolina [47], demanda química de oxigênio [48], glicerol em biodiesel [49].

No trabalho de Silva, et al. (2009) [48] é apresentado o voltamograma cíclico típico obtido com eletrodo de cobre em meio alcalino, o mesmo pode ser visualizado na figura **Figura 2.5**.

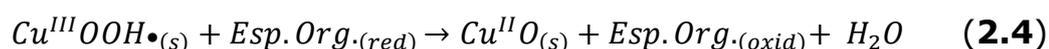
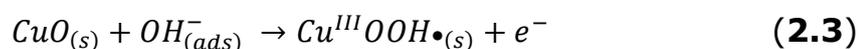


**Figura 2.5** – Voltamograma cíclico para eletrodo metálico de cobre em solução de 0.1 mol/L de NaOH, faixa de potencial -1,0 a 0,70 V vs Ag / AgCl (3,0 mol L<sup>-1</sup> de KCl)

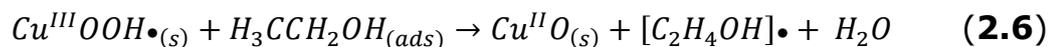
Fonte: Adaptado de Silva et al. (2009) [48]

Na **Figura 2.5** é possível observar picos que estão relacionados à diferentes estágios de oxidação do cobre. O pico anódico **1** que aparece em aproximadamente 0.3 V, corresponde a formação de óxido de cobre (I), na forma de Cu<sub>2</sub>O. Já o pico **2**, que surge com a maior intensidade de corrente, aproximadamente em -0.1 V, está relacionado à formação de cobre (II), principalmente nas formas de CuO e Cu(OH)<sub>2</sub>. Avaliando o sentido inverso da varredura, é possível observar o pico catódico **3** por volta de -0.55 V, e está ligado à redução do Cu(II) à Cu(I). Já, no potencial de aproximadamente -0.84 V, o pico **4** provavelmente esteja relacionado à mudança de Cu(I) a Cu(0). [48]

Dentre os diferentes estados de oxidação apresentados pelo cobre durante o processo de varredura utilizando a voltametria cíclica em meio alcalino, o pico apresentado em potenciais de aproximadamente 0.7V, que pode ser visto na **Figura 2.5**, está ligado à formação da espécie livre solúvel de Cu (III), que é a porção ativa que medeia à oxidação de compostos orgânicos. [45-48] Tal fato já postulado por voltametria com RRDE (Eletrodo rotatório de disco-anel) [50], e por isso, tem sido sugerido para explicar o bom desempenho de eletrodos de cobre aplicados na identificação de diferentes compostos orgânicos. [45-47] A seguir é apresentada a **Equação 2.3** e **2.4** com a representação do um mecanismo proposto Silva et al. (2009) para oxidação de compostos orgânicos. [48]



Mecanismo semelhante para oxidação de álcoois são encontrados na literatura, como é o caso da oxidação de metanol proposto por Heli et al. (2004) [51]. Ainda, Paixão et al. (2002) [45] propuseram a **Equação 2.5** e **2.6**, referentes à oxidação de etanol:



Mesmo com o conhecimento da capacidade de detecção de moléculas orgânicas, em principal de álcool, por meio da oxidação utilizando eletrodo de cobre em meio alcalino, a identificação de adulteração de AEHC, principalmente por metanol utilizando voltametria pode ser uma tarefa difícil, pois as moléculas de etanol e metanol possuem propriedades químicas bem semelhantes, podendo apresentar sinais sobrepostos. Desse modo, uso de técnicas baseadas em análise multivariada aplicadas aos dados voltamétricos torna-se importante para auxiliar no processo de detecção de adulteração de amostras de AEHC.

### 2.3. Quimiometria

Os conceitos de quimiometria começaram a ser aplicados a partir dos anos 70 com o intuito de descrever a aplicação de modelos matemáticos e de estatística à química. [52] A quimiometria possui uma série de aplicações específicas, entre estas: planejamento e otimização de experimentos, reconhecimento de padrões, métodos de resolução de curvas, calibração multivariada, transferência de calibração, entre outros. [53]

A evolução dos computadores é um dos motivos da facilitação e divulgação das técnicas quimiométricas. Os cálculos multivariados passaram a ser realizados de forma mais rápida, possibilitando a análise de resultados de técnicas analíticas que geram um grande conjunto de dados, como é o caso da quimiometria aplicada às técnicas de infravermelho [54-56] e eletroanalíticas [57-59].

Diante desta realidade, a quimiometria passou a auxiliar na aquisição respostas analítica rápidas, que podem contribuir a tomada de decisões

por meio do uso de métodos alternativos, o que se conhece por análise *screening*.

A necessidade de respostas analíticas rápidas e confiáveis é uma prática comum nos laboratórios de pesquisa e de análise de rotina. Métodos laboratoriais convencionais fornecem informações qualitativas e quantitativas detalhadas sobre as amostras. Na maioria das vezes, tais métodos são demorados, demandam de uma grande quantidade de amostra, uso de reagentes caros e específicos. Contudo, métodos baseados na análise *screening* são mais simples e rápidos, podendo proporcionar resolução de questões como: sim ou não? Continue ou pare? Pertence ou não pertence? Adulterada ou não adulterada? Dessa maneira, algumas vantagens importantes podem ser alcançadas, incluindo a redução dos custos, a rapidez e simplificação de metodologias. [60-62]

As técnicas de reconhecimento de padrão podem ser aplicadas a dados espectrais e voltamétricos com intuito de construir e validar metodologias baseadas na análise *screening*, para identificar e classificar a presença ou não de adulteração em amostras de AEHC. Essas técnicas podem ser realizadas de forma não supervisionada e supervisionada. No primeiro caso, uma avaliação exploratória é realizada com objetivo de identificar semelhanças e diferenças em conjuntos de amostras que foram analisadas.

A PCA [63] é uma poderosa ferramenta amplamente utilizada para esse propósito. Nos métodos de reconhecimento de padrão supervisionado, uma informação adicional sobre os membros das classes é utilizada para a elaboração de modelos que possam classificar amostras desconhecidas.

Nesse contexto, os métodos baseados na LDA e no PLS-DA vêm se destacando em diversos problemas analíticos de classificação. [64-67] As bases teóricas da PCA, LDA e PLS-DA são discutidas nas próximas seções.

### 2.3.1 Análise por componentes principais (PCA)

A análise de componentes principais (PCA) é um método de análise multivariada de dados muito usado para análise exploratória. Ela permite redução de dimensionalidade, verificação do comportamento das amostras baseando-se na variância, e também, serve como base para métodos classificatórios e de regressão. [63]

A PCA normalmente realizada sob uma matriz de dados  $\mathbf{X}$  ( $n \times m$ ), sendo que, nas linhas encontram-se os  $n$  objetos (amostras) e, nas colunas, as  $m$  variáveis. No caso de dados espectrais, as variáveis correspondem aos valores de absorbância nos respectivos números de onda, já para os dados voltamétricos, as variáveis correspondem aos valores de correntes referentes aos potenciais.

Para realização da PCA, a matriz  $\mathbf{X}$  é fatorada em dois subespaços, o de escores ( $\mathbf{t}$ ) e o de pesos ( $\mathbf{l}$ ). Nesse processo, a fatoração é realizada de forma sequencial, sendo que, cada par ( $\mathbf{t}$ ,  $\mathbf{l}$ ) recebe o nome de “componente principal (PC)”.

Considerando a soma dos quadrados da matriz  $\mathbf{X}$ , que representa a variância total dos dados, receberá o nome de primeira componente principal (PC1), a que explicar a maior fração possível da variância. A segunda componente principal (PC2) será ortogonal a primeira e deve possuir a segunda maior fração da variância explicada. E, assim, sucessivamente, até que um número ótimo de componentes principais (PC<sub>*n*</sub>) seja obtido, sendo a matriz  $E$  (resíduo) correspondente a variância não explicada que apresenta o ruído amostral. O sistema de fatoração em componentes principais pode ser representado pela **Equação 2.7** e **2.8** apresentadas a seguir:

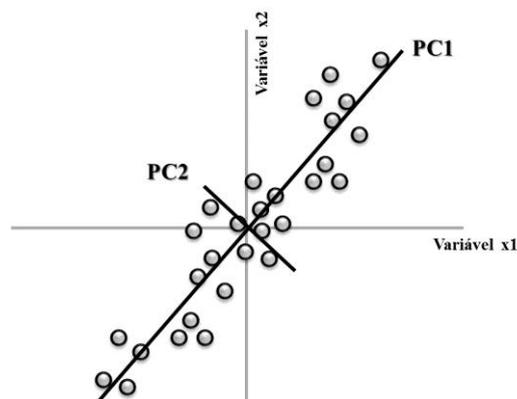
$$\mathbf{X} = \mathbf{t}_1 \mathbf{l}_1^T + \mathbf{t}_2 \mathbf{l}_2^T + \mathbf{t}_n \mathbf{l}_n^T + E \quad (2.7)$$

Em que:  $\mathbf{X}$  é uma matriz instrumental,  $t_n$  equivale ao vetor de escores da  $PC_n$ ,  $l_n^T$  corresponde ao vetor de pesos da  $PC_n$ ,  $E$  equivale ao resíduo (Variância não explicada).

$$\mathbf{X} = \mathbf{T}\mathbf{L}^T + E \quad (2.8)$$

Em que:  $\mathbf{X}$  é uma matriz instrumental,  $\mathbf{T}$  equivale a matriz de escores da  $PC_1$  a  $PC_n$ ,  $\mathbf{L}^T$  corresponde a matriz de pesos da  $PC_1$  a  $PC_n$ ,  $E$  equivale a variância total não explicada (resíduo total).

Na **Figura 2.6** é possível observar o plano bidimensional para duas variáveis  $x_1$  e  $x_2$  com dados centrados na média, com as duas primeiros componentes principais representados em negro. [63, 68 e 69]



**Figura 2.6** – Plano bidimensional com duas variáveis e representação dos dois primeiros Componentes Principais

Fonte: Adaptado de Beebe, et. al (1998) [68]

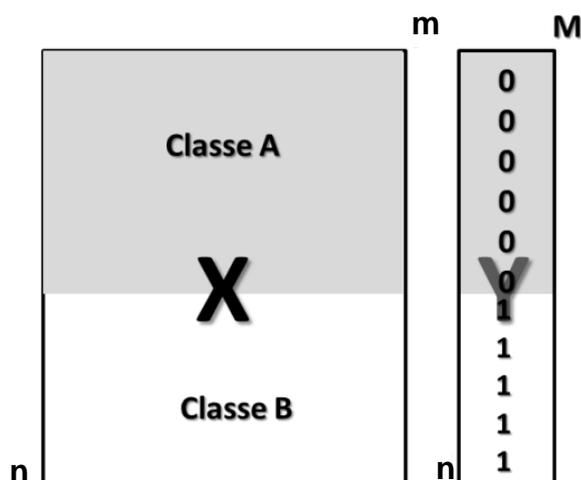
Na **Figura 2.6**, as coordenadas das amostras com relação às PC's correspondem aos escores ( $\mathbf{t}$ ) e são matematicamente definidos como as projeções das amostras no eixo das PC's. A contribuição de cada variável para as PC's são considerados os pesos ( $\mathbf{l}$ ) e é calculado como o cosseno do ângulo entre o eixo da variável e o eixo da PC, ou seja, as variáveis

originais com maior peso na combinação linear das primeiras componentes principais são as mais importantes. [68, 69 e 70]

### 2.3.2 Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)

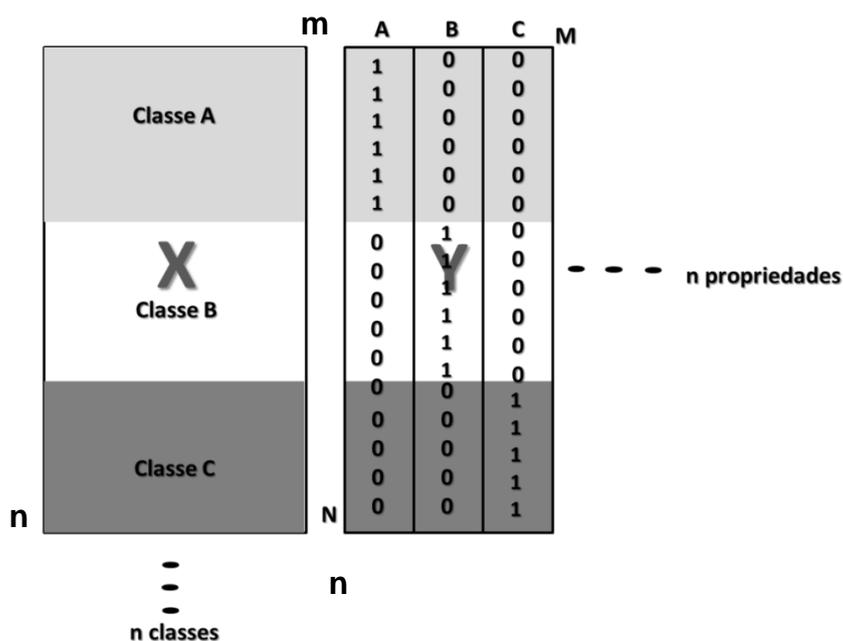
A análise discriminante pelos mínimos quadrados parciais (PLS-DA) é um método de reconhecimento de padrão supervisionado baseado no algoritmo PLS, técnica de calibração inversa que busca a relação entre a resposta instrumental e propriedades de interesse. No contexto da calibração multivariada, esta relação é estabelecida por meio de uma matriz contendo as repostas instrumentais,  $\mathbf{X}$  ( $n \times m$ ), com um vetor ou uma matriz contendo a (s) propriedade (s) de interesse,  $\mathbf{Y}$  ( $n \times M$ ). No caso do PLS-DA a matriz  $\mathbf{X}$  permanece com os dados instrumentais e a matriz  $\mathbf{Y}$  recebe a designação da classe a qual a amostra pertence. [71, 72]

Para o estudo com duas classes A e B, o vetor  $\mathbf{y}$  é uma variável indicadora que recebe os valores de 0 ou 1, em que, as amostras da classe A recebem o valor 0 e as da classe B o valor 1. Neste caso, em que apenas duas classes são modeladas, o cálculo é efetuado usando o método de calibração pelos Mínimos Quadrados Parciais (PLS1). [72] Na **Figura 2.7** apresenta-se um esquema de organização das matrizes para o cálculo do PLS-DA com duas classes.



**Figura 2.7** – Representação da organização dos dados para realização de PLS-DA para duas classes.

Casos em que a modelagem inclui a presença de três ou mais classes, utiliza-se o método PLS2. A variável indicadora é uma matriz **Y** com um número de colunas igual ao número de classes. Valores 0 e 1 são utilizados nas colunas que correspondem a cada uma das classes, sendo que, as amostras pertencentes à essa classe recebem o valor 1, enquanto as demais recebem o valor 0 (**Figura 2.8**).



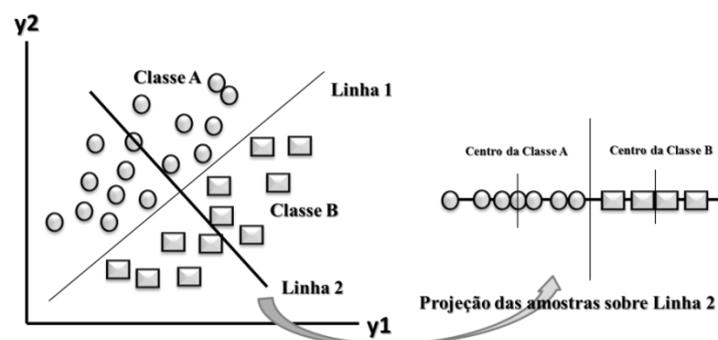
**Figura 2.8** – Representação da organização dos dados para realização de PLS-DA para três ou mais classes.

Os valores previstos pelo modelo PLS-DA serão idealmente entre zero e um. Um limiar entre 0 e 1 deve ser estabelecido para indicar qual classe a amostra pertencerá. Então, a amostra que apresentar um valor previsto, em módulo, maior que o valor limiar adotado, será considerado não membro; já um valor previsto, em módulo, menor que o limiar, pertencerá à classe.

### 2.3.3 Análise Discriminante Linear (LDA)

Análise discriminante é uma técnica comumente utilizada para classificar um grupo de observações em classes pré-definidas. [68] Dentro dessa perspectiva, Fisher (1936) [73] propôs a resolução do problema da discriminação entre dois ou mais grupos, visando posterior classificação. A proposta de Fisher consiste em obter funções matemáticas capazes de classificar uma amostra  $\mathbf{x}$  em um dos vários grupos (populações)  $\mu_i$ , ( $i=1, 2, \dots, g$ ), variando da classe 1 até a classe  $g$ , com base na combinação linear de características (variáveis) medidas, de forma a maximizar a razão da distância entre as classes pela distância no interior da classe. [69, 74]

Para um maior entendimento, a **Figura 2.9** apresenta um esquema de discriminação entre duas classes (A e B), considerando duas variáveis ( $y_1$  e  $y_2$ ) medidas.



**Figura 2.9** – Esquema de discriminação entre duas classes (A e B), baseando-se nas variáveis originais ( $y_1$  e  $y_2$ ), ou na linha discriminatória.

Fonte: Adaptado de Brereton (2003) [74]

Como pode ser visto na **Figura 2.9** as variáveis originais ( $y_1$  e  $y_2$ ), quando avaliadas individualmente, não conseguem discriminar as classes A e B. Porém, quando combinadas, a separação entre as duas classes é observada, estando demarcada pela Linha 1, em que as amostras da classe A ficam acima e da classe B, abaixo. Adicionalmente, é possível representar a discriminação das classes por meio da linha 2. As amostras podem ser projetadas sobre a mesma, e as classes poderão ser determinadas por distância. Os valores menores equivalem à classe A e valores maiores podem ser pertencentes a classe B. Caso sejam calculadas as médias das classes, as amostras que tiverem valores mais próximos a média da classe A, pertencerão à mesma, e o mesmo se dará para amostras com valores próximos à média da classe B.

A LDA baseia-se, para o propósito de classificação, no cálculo da distância de Mahalanobis [68, 69, 74]. Sendo uma amostra descrita por  $\mathbf{x} = [x_1, x_2 \dots x_m]^T$ , em que  $[x_1, x_2 \dots x_m]$  corresponde as variáveis referentes à técnica aplicada, como as medidas de absorvância em diferentes comprimentos de onda, ou valores de corrente para cada potencial aplicado. Teremos que, o quadrado da distância de Mahalanobis  $r^2(\mathbf{x}, \mu_j)$  entre  $\mathbf{x}$  e o centro da  $j$ -ésima classe ( $j=1,2,\dots, c$ ), será definido conforme a **equação 2.9**:

$$r^2(\mathbf{x}, \mu_j) = (\mathbf{x} - \mu_j)^T \cdot \Sigma_j^{-1} \cdot (\mathbf{x} - \mu_j) \quad (2.9)$$

Em que:  $\mathbf{x}$  é um vetor linha de uma observação,  $\mu_j$  e  $\Sigma_j^{-1}$  são respectivamente a medida da média (centro) e matriz de variância-covariância para a classe  $j$ -ésima classe.

Uma dificuldade encontrada com o uso dessa distância de Mahalanobis é que o número de objetos (amostras) deve ser maior que o número de variáveis. Se houver  $m$  variáveis deve haver pelo menos  $m +$

2 amostras, caso contrário, a matriz de variância-covariância será singular. [69, 74]

Dados provenientes da espectroscopia no infravermelho ou da voltametria cíclica possuem, normalmente, um número de variáveis maior que o número de amostras. Assim procedimentos de seleção de variáveis e/ou redução de dimensionalidade são requeridos.

### 2.3.3.1 Seleção de variáveis

As técnicas de seleção de variáveis têm sido utilizadas para diferentes propósitos de classificação. Entre elas, destacam-se: algoritmo das projeções sucessivas (SPA) [75-77], algoritmo genético (GA) [78-79] e o stepwise (SW). [80-81]

O SPA, foi proposto inicialmente por Araújo et al. (2001) [75] e tem como objetivo buscar subconjuntos de variáveis que contenham informação minimamente redundante, minimizando a colinearidade dos dados. Para tanto, o SPA baseia-se num método de *inclusão*, no qual, inicia-se com uma variável, e incorpora-se outra variável a cada iteração, até que a quantidade de N (número de variáveis necessárias) seja atingida. A essência do SPA consiste em operações de projeção efetuadas na matriz de dados, cujas linhas e colunas correspondem, respectivamente, às amostras de treinamento e variáveis (número de onda, potenciais, entre outros.). Cadeias de variáveis são geradas e avaliadas por meio do risco médio G (Equação 2.10) de uma classificação incorreta pela LDA, calculada com base nas amostras do conjunto de validação. O subconjunto de variáveis que expressar o menor valor de G será, então, selecionado. [75-77]

$$G = \frac{1}{k_V} \sum_{k=1}^{k_V} g_k \quad (2.10)$$

Em que:  $g_k$  (risco de uma classificação incorreta do objeto  $x_k$  da  $k$ -ésima amostra de validação) é definido como na **equação 2.11**:

$$g_k = \frac{r^2(x_k, \mu_{Ik})}{\min_{Ij \neq Ik} r^2(x_k, \mu_{Ij})} \quad (2.11)$$

Na equação anterior, o numerador  $r^2(x_k, \mu_{Ik})$  é o quadrado da distância de Mahalanobis entre o objeto  $x_k$  (com índice de classe  $Ik$ ) e a média de sua classe ( $\mu_{Ik}$ ). O denominador da **Equação 2.11** corresponde ao quadrado da distância de Mahalanobis entre o objeto  $x_k$  e o centro da classe errada mais próxima. Idealmente,  $g_k$  deverá ser tão pequeno quanto possível, ou seja, o objeto  $x_k$  deverá estar perto do centro da sua verdadeira classe e distante dos centros das demais classes.

O algoritmo genético (GA) foi proposto John H. Holland na década de 60. Trata-se de uma técnica de otimização e busca aleatória baseada no princípio da biologia de evolução e genética natural. [78] O GA assimila-se ao processo evolutivo natural, pois sua estrutura permite que as variáveis do sistema em estudo sejam codificadas de forma binária, ficando com uma estrutura análoga aos dos cromossomos biológicos. Por meio de um gerador aleatório uma população de cromossomos é gerada aleatoriamente, de forma a evitar influências tendenciosas externas. Em seguida, testa-se a aptidão do indivíduo, verificando, matematicamente, quais apresentam melhor resposta para o sistema. Nessa dissertação, a aptidão usada para o GA foi considerada como o inverso do risco  $G$  (**Equação 2.10**) calculado usando os números de onda ou potenciais codificados no cromossomo. Os indivíduos que gerarem as melhores respostas irão ter seus parâmetros cruzados para geração de novas combinações, descendentes. Alguns indivíduos descendentes podem ainda receber características diferentes dos pais e sofrerem mutações. Gerações e mutações ocorrem até que a condição de avaliação seja alcançada. [79]

O algoritmo stepwise (SW) baseia-se no método de inclusão ou exclusão de variáveis. Em outras palavras, o algoritmo verifica a

importância das variáveis, incluindo-as ou excluindo-as do modelo com base no potencial discriminatório das mesmas. A variável com maior grau de discriminabilidade  $x_i$  é selecionada, enquanto um procedimento de validação cruzada leave-one-out, é executado e o número de erros é verificado. Antes da próxima etapa, as variáveis que são altamente correlacionadas com a variável  $x_i$  são descartadas com base em um coeficiente de correlação indicado pelo analista. O procedimento é repetido até que o conjunto de variáveis que apresentar menor número de erros de validação cruzada seja escolhido. [80-81]

Capítulo III

# **EXPERIMENTAL**

### 3. EXPERIMENTAL

#### 3.1 Amostras

As amostras utilizadas neste estudo foram fornecidas pelo Laboratório de Combustíveis (LAC) da Universidade Federal de Pernambuco (UFPE), e tiveram suas coletas realizadas em diferentes postos de combustíveis localizados nos estados de Alagoas, Pernambuco e Sergipe. O LAC é responsável pelo monitoramento dos combustíveis/biocombustíveis, incluindo o AEHC, comercializados nesses estados. Todas as amostras utilizadas para esse trabalho encontravam-se dentro de especificação da ANP.

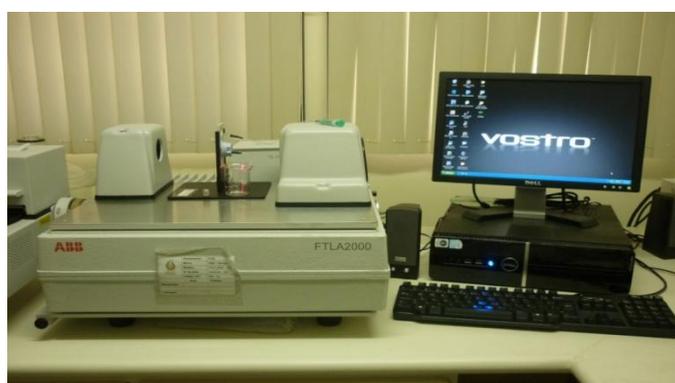
Um total de 184 amostras de AEHC foram utilizadas nesse trabalho, as quais foram divididas em três conjuntos, cada conjunto com aproximadamente 60 amostras. Cada conjunto de amostra compôs uma classe, da seguinte forma: **(1)**: amostras que não passaram por adulterações - *não adulteradas* (60); **(2)** amostras em que foi adicionado alíquotas de água - *adulteradas com água* (60) e **(3)** amostras que tiveram adição de metanol - *adulteradas com metanol* (60). As adulterações foram realizadas em laboratório por meio da adição de metanol e água nos níveis de 1,9-13,0% m.m<sup>-1</sup> e 0,5-10,0% m.m<sup>-1</sup>, respectivamente. Para incluir uma maior variabilidade na composição das amostras adulteradas, as adulterações foram realizadas com diferentes amostras e de forma aleatória.

#### 3.2 Aquisição dos espectros no infravermelho (NIR e MIR)

Um espectrofotômetro FTIR da Bomem (FTLA 2000-160) (**Figura 3.1**) equipado com uma célula de quartzo com percurso óptico de 1,0 mm foi utilizado para obter espectros NIR (15800 a 370 cm<sup>-1</sup>). Um

espectrofotômetro FTIR Varian 640 (**Figura 3.2**) equipado com um acessório de amostragem ATR foi utilizado para obter os espectros na faixa do MIR (4000 a 698  $\text{cm}^{-1}$ ).

Todos os espectros foram registrados com uma média de 32 varreduras, e uma resolução espectral de 8  $\text{cm}^{-1}$ . Os espectros do branco foram obtidos usando uma célula vazia limpa ou o acessório de ATR. A temperatura foi controlada a  $23 \pm 1$  °C durante todo o processo de aquisição espectral.



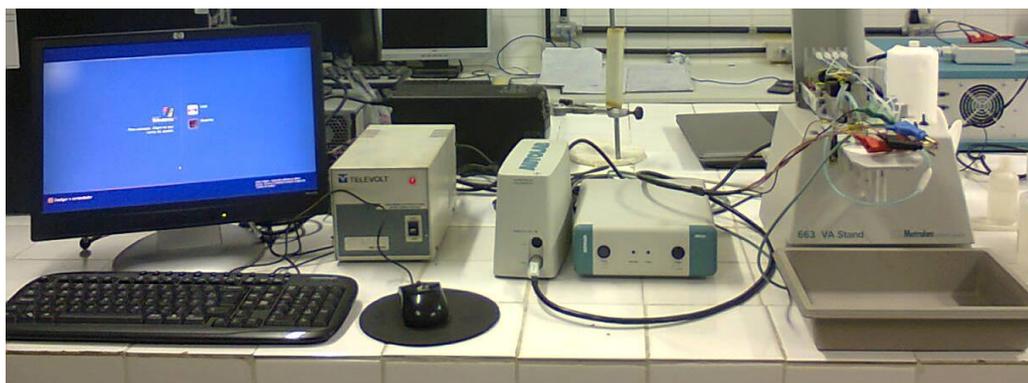
**Figura 3.1** – Espectrofotômetro FT-IR da Bomem (FTLA 2000-160) utilizado para o registro dos espectros NIR.



**Figura 3.2** – Espectrofotômetro FTIR Varian 640 utilizado para o registro dos espectros do MIR.

### 3.3 Aquisição dos voltamogramas

A aquisição dos voltamogramas foi realizada em um potenciostato Autolab PGSTAT 101 (Eco Chemie) (**Figura 3.3**) com o software de aquisição de dados (Nova 1.7). Uma configuração de três eletrodos foi utilizada, compreendendo um eletrodo de Ag/AgCl (KCl saturado) como de referência, um fio de platina como contra-eletrodo, e um fio de cobre de área planar e circular ( $d = 2,0$  mm) recoberto por uma vareta de vidro, como eletrodo de trabalho. Uma solução de  $1,0 \text{ mol L}^{-1}$  de hidróxido de sódio (Vetec) foi usada como eletrólito de suporte.



**Figura 3.3** – Potenciostato Autolab PGSTAT 101 (Eco Chemie) utilizado para aquisição dos voltamogramas.

Antes da realização das medidas voltamétricas, visando à remoção de sujeiras e contaminantes, o eletrodo de cobre passou por um processo de polimento mecânico em uma superfície de feltro utilizando pasta de alumina com  $3,0 \mu\text{m}$ . Depois o mesmo foi lavado abundantemente com água deionizada até que a superfície lisa e com aparência espelhada. Em seguida, visando obter uma maior formação de espécies de Cu(III) em solução, assim como proposto por Paixão et. al (2004) [46], o eletrodo foi polarizado no eletrólito de suporte pela aplicação de um potencial de  $-0,1 \text{ V vs Ag / AgCl}$  por 5 segundos. Todos os voltamogramas foram registrados após a adição de  $100 \mu\text{L}$  de etanol em  $10 \text{ mL}$  de solução de NaOH. A

janela de potencial avaliada foi de um potencial inicial ( $E_i$ ) de -0.89 V para o potencial final ( $E_f$ ) de 0,8 V vs Ag / AgCl, sentido anódico, com os seguintes parâmetros: taxa de incremento potencial e velocidade de varredura de 5 mV e 50 mV/s, respectivamente.

Os parâmetros voltamétricos utilizados nesse trabalho foram baseados em diferentes trabalhos encontrados na literatura que utilizaram o eletrodo de cobre em meio alcalino para determinação de compostos orgânicos, em especial, álcool em outras matrizes. [45, 48, 51] Ainda, o parâmetro velocidade de varredura foi avaliado em dois níveis: 10 mV/s e 50mV/s. Porém, a velocidade de varredura de 50mV/s foi utilizada, pois não apresentou diferenças significativas, e ao mesmo tempo, permite que a medida fosse realizada mais rapidamente.

### 3.4 Procedimentos Quimiométricos e Software

Antes da construção dos modelos quimiométricos, as amostras de cada uma das classes de AEHC foram divididas em três subconjuntos (treinamento, validação e teste) utilizando o algoritmo Kennard-Stone (KS) [82]. O algoritmo KS foi aplicado separadamente para cada classe, como reportado por Pontes et. al [76]. É importante destacar que, para o estudo envolvendo a espectroscopia IR, três classes foram estudadas: **(1)** Não adulteradas; **(2)** adulteradas com metanol e **(3)** adulteradas com água. Para os dados voltamétricos, foram consideradas apenas as classes de amostras não adulteradas e aquela composta por amostras adulteradas com metanol. O número de amostras em cada em cada conjunto para cada banco de dados, após a remoção de amostras anômalas, é apresentado nas **Tabelas 3.1 e 3.2**.

**Tabela 3.1** – Número de amostras de treinamento, validação e teste em cada classe para os dados de espectroscopia no Infravermelho.

Classe	Dados NIR			Dados MIR		
	Treinamento	Validação	Teste	Treinamento	Validação	Teste
<b>Não Adulteradas</b>	30	15	15	28	15	15
<b>Adulteradas com água</b>	31	15	15	28	16	16
<b>Adulteradas com metanol</b>	30	15	15	28	14	14
<b>Total</b>	91	45	45	84	45	45

**Tabela 3.2** – Número de amostras de treinamento, validação e teste em cada classe para os dados de voltametria.

Classe	Dados Voltametria		
	Treinamento	Validação	Teste
<b>Não Adulteradas</b>	30	15	15
<b>Adulteradas com metanol</b>	30	15	15
<b>Total</b>	60	30	30

É importante mencionar que, tanto as amostras de treinamento quanto as de validação foram utilizadas nos procedimentos de modelagem em todas as técnicas de classificação. As amostras de teste, no entanto, só foram utilizadas na comparação do desempenho de classificação dos diferentes modelos construídos. Todos os dados foram centrados na média antes de procedimentos de modelagem.

Os espectros e voltamogramas obtidos foram inicialmente avaliados com o intuito de determinar a faixa de trabalho a ser utilizada, bem como, a necessidade de pré-processamentos.

Os espectros IR foram avaliados sem pré-processamentos e pré-processados com a primeira derivada pelo método de Savitzky-Golay (variando o número de janelas em 7, 11, 15 e 21). O desempenho desses

dados (originais e pré-processados) foi avaliado em função da taxa de classificação correta para o conjunto de teste. Para os dados voltamétricos, apenas os voltamogramas originais (sem pré-processamento) foram empregados.

Uma análise exploratória baseada na PCA foi realizada com o intuito de verificar a formação natural de agrupamentos baseadas na variância dos dados, tanto voltamétricos, quanto no infravermelho (NIR e MIR).

Na elaboração dos modelos de classificação, quatro métodos foram avaliados: análise discriminante pela regressão por mínimos quadrados parciais (PLS-DA); análise discriminante Linear com prévia seleção de variáveis pelo algoritmo das projeções sucessivas (SPA-LDA), algoritmo genético (GA-LDA) e stepwise (SW-LDA). No processo de modelagem, as amostras de treinamento e validação foram utilizadas. Todos os dados foram centrados na média antes de procedimentos de modelagem.

Para avaliação final dos modelos, a taxa de classificação correta (TCC) das amostras de teste foi utilizada.

$$TCC(\%) = \left( \frac{\text{Número de acertos de classificação}}{\text{Número total de amostras}} \right) \times 100\% \quad (3.1)$$

A rotina de GA foi executada para 50 gerações com 100 cromossomos em cada geração. Cruzamento e mutação foram definidos com probabilidades de 60% e 10%, respectivamente. Devido à natureza estocástica do algoritmo GA-LDA, a rotina foi repetida dez vezes (a partir de diferentes populações iniciais aleatórias) e a melhor solução encontrada durante estas 10 execuções foi mantida.

A metodologia do stepwise (SW) utilizada no presente trabalho adotou o critério apresentado por Caneca et al. (2006) [80]. Para tanto, sete valores (0,10, 0,30, 0,50, 0,70, 0,90, 0,95 e 0,99) de coeficientes de correlação foram testados no modelo SW-LDA, a fim de decidir quais as

variáveis deveriam ser descartadas. O melhor valor foi selecionado com base nos erros de classificação no conjunto de validação.

Na execução do PLS-DA, as classes foram codificadas com os valores 0 e 1, e foi adotado como limiar o valor de 0.5.

O pré-processamento dos dados, a PCA e PLS-DA foram realizados utilizando Unscrambler ® X.1 (CAMO SA). Os algoritmos KS, SPA-LDA, GA-LDA e SW-LDA foram executados em Matlab 6.5 (Mathworks, EUA).

Capítulo IV

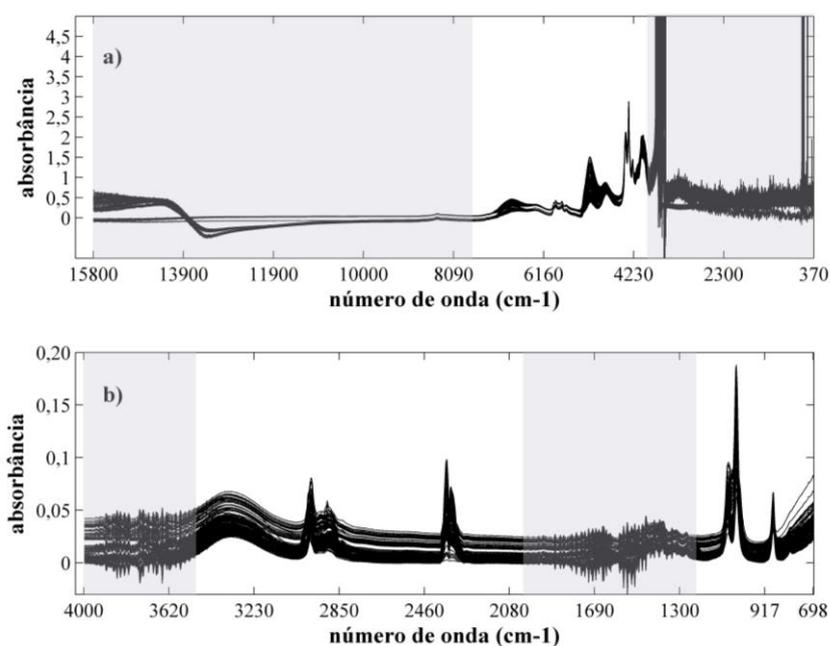
# **RESULTADOS E DISCUSSÃO**

## 4. RESULTADOS E DISCUSSÃO

### 4.1. Estudo com Espectroscopia no Infravermelho

#### 4.1.1. Avaliação dos espectros NIR e MIR

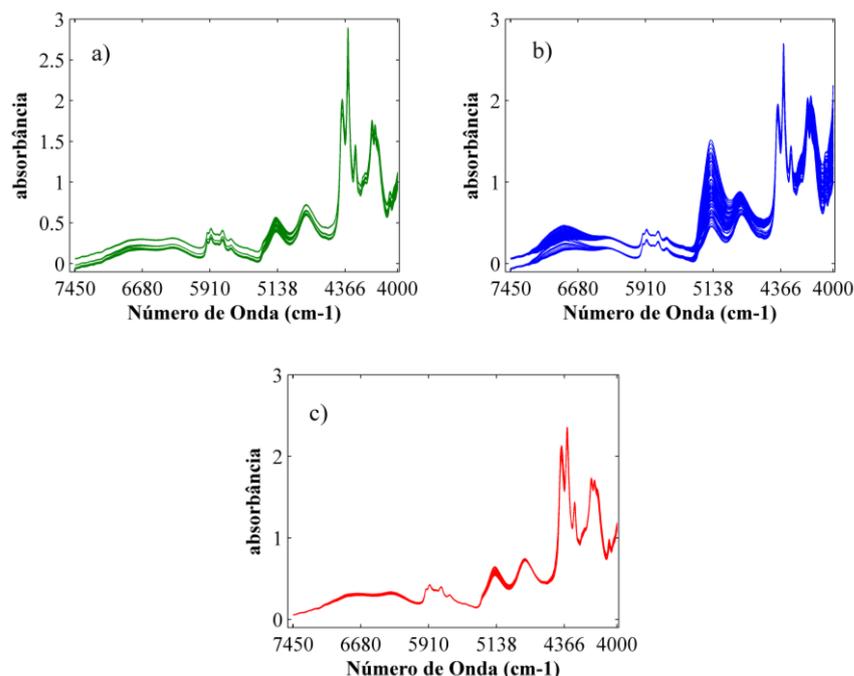
Os espectros das amostras de AEHC foram registrados nas regiões do NIR (15800 a 370  $\text{cm}^{-1}$ ) (**Figura 4.1a**) e do MIR (4000 a 698  $\text{cm}^{-1}$ ) (**Figura 4.1b**), conforme apresentado a seguir.



**Figura 4.1** – Espectros registrados na região (a) NIR e (b) MIR

As regiões sombreadas encontradas na **Figura 4.1**, não foram utilizadas no tratamento quimiométrico. De fato, tais regiões apresentam ou um sinal de absorvância muito baixo (15800-7500  $\text{cm}^{-1}$  na **Figura 4.1a**) ou saturação do detector com um sinal ruidoso (3800 a 370  $\text{cm}^{-1}$  na **Figura 4.1a**, e nas faixas de 4000 a 3500  $\text{cm}^{-1}$  e 2000  $\text{cm}^{-1}$  a 1200  $\text{cm}^{-1}$  na **Figura 4.1b**).

Os espectros NIR originais registrados na faixa de 7449–3769  $\text{cm}^{-1}$  para as três classes: não adulteradas (**Figura 4.2a**), adulteradas com água (**Figura 4.2b**) e adulteradas com metanol (**Figura 4.2c**).

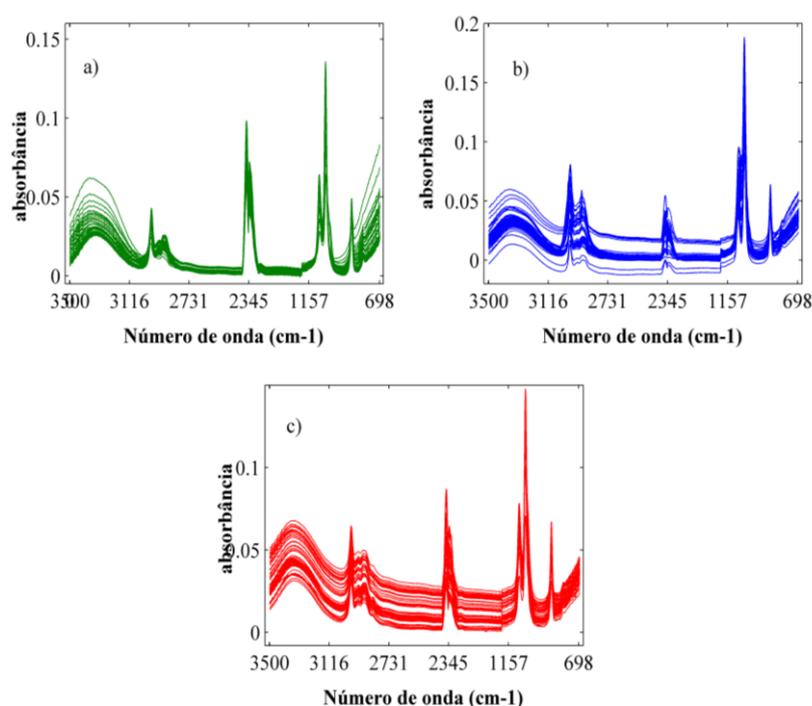


**Figura 4.2** – Espectros na região do NIR (7449 a 3769  $\text{cm}^{-1}$ ) para as classes de amostras: (a) não adulteradas, (b) adulteradas com água e (c) adulteradas com metanol.

Na **Figura 4.2** (a, b e c) é possível observar uma grande semelhança entre os espectros de absorção na região do NIR para as amostras não adulteradas e adulteradas com água ou com metanol. Os picos encontrados podem ser associados ao primeiro sobretom (5.000-6.000  $\text{cm}^{-1}$ ) e às regiões de combinação (4.600-4.000  $\text{cm}^{-1}$ ) de estiramento C-H. [84] Além disso, o primeiro sobretom do OH da água, etanol e do metanol ocorre em torno de 6.870  $\text{cm}^{-1}$ . Bandas de combinação de estiramento e flexão angular do OH na água ocorrem na região próxima de 5.128  $\text{cm}^{-1}$ . [85-86]

Na **Figura 4.3**, observa-se em sequencia, os espectros MIR na faixa de 3500  $\text{cm}^{-1}$  a 698  $\text{cm}^{-1}$  para as três classes estudadas, as quais apresentaram perfis espectrais semelhantes. Em particular, os valores de absorbância encontrados em torno de 1180-840  $\text{cm}^{-1}$  correspondem aos

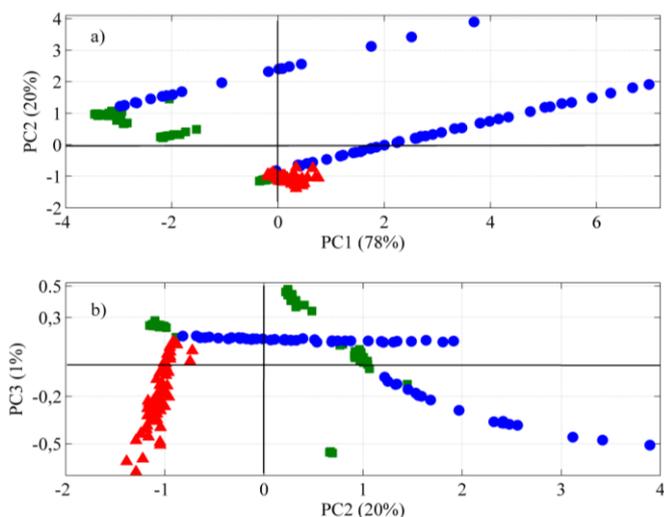
modos de deformação axial C-O e deformação simétrica C-C-O para álcoois primários. Deformação axial O-H pode ser observado em torno de  $3500\text{ cm}^{-1}$  e os picos também de estiramento C-H, C-H<sub>2</sub> e CH<sub>3</sub> também podem ser observados em torno de  $2.900\text{ cm}^{-1}$ . [87] Ainda, os picos apresentados em torno de  $2345\text{ cm}^{-1}$  podem, possivelmente, serem atribuídos às vibrações de grupos nitrilas (C≡N) [36], que, por sua vez, podem estar relacionados aos aditivos adicionados ao AEHC, previstos no artigo 9 clausula §5, da resolução nº 7 da ANP. [14]



**Figura 4.3** – Espectros MIR ( $3500$  a  $698\text{ cm}^{-1}$ ) para as classes de amostras: (a) não adulteradas, (b) adulteradas com água e (c) adulteradas com metanol.

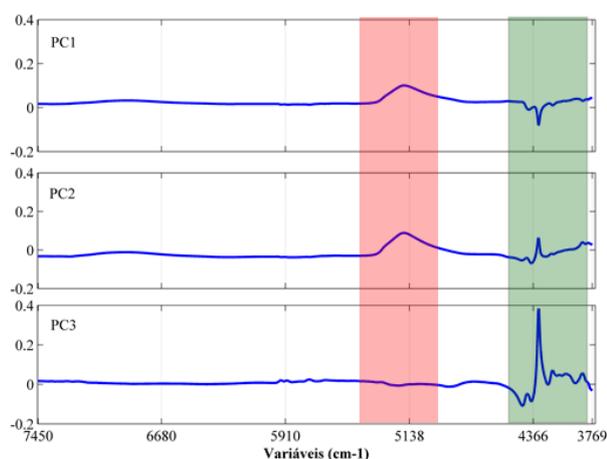
#### 4.1.2. Análise Exploratória dos dados de Infravermelho

O comportamento natural dos dados em termos de variância para as amostras não adulteradas e adulteradas, foi avaliado com a aplicação de uma PCA ao conjunto completo dos espectros, sendo avaliados os gráficos dos escores e pesos para NIR (Figura 4.4 e 4.5) e MIR (Figura 4.6 e 4.7).



**Figura 4.4** – Gráfico dos escores da PCA aplicada aos espectros NIR. (a) PC1xPC2 e (b) PC2xPC3. (■: amostras não adulteradas, ●: amostras adulteradas com água, ▲: amostras adulteradas com metanol)

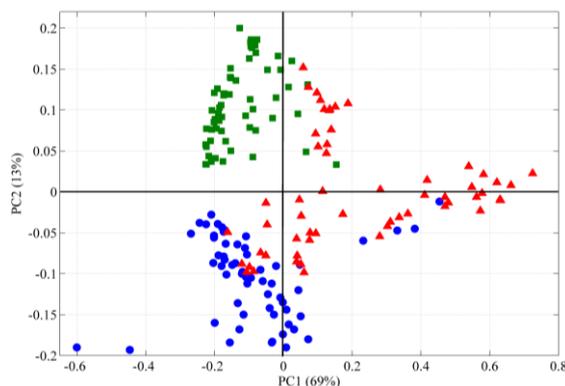
Para a distribuição de escores apresentada na **Figura 4.4**, não é observada a formação de agrupamentos das classes estudadas. A dispersão nos valores de escores pode estar relacionada à variabilidade inerente as diferentes origens das amostras não adulteradas. E também à utilização de diferentes níveis de adulteração, tanto com metanol, quanto com a água. Na figura a seguir são observados os valores de pesos da PCA, os quais podem auxiliar na compreensão do comportamento encontrado no gráfico de escores.



**Figura 4.5** – Gráficos dos pesos da PCA (PC1, PC2 e PC3) aplicada aos espectros NIR das amostras de AEHC.

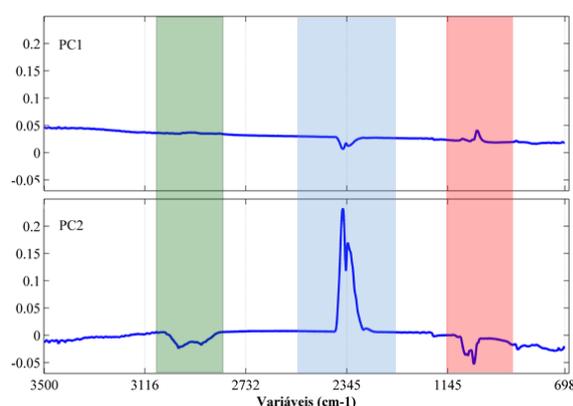
Os valores de pesos obtidos pela PCA (**Figura 4.5**) para a PC1 e PC2, possuem regiões relevantes que podem ser associadas à flexão angular do OH acoplada à deformação simétrica de C-H (região avermelhada do gráfico de pesos). Essa região ocorre por volta de  $5.128\text{ cm}^{-1}$  [85-86], e que justificaria a maior dispersão das amostras adulteradas com água, caracterizadas pelo aumento das forças intermoleculares das ligações de hidrogênio, que aumentam com o aumento do percentual de água. Ainda, valores de pesos significativos em regiões referentes a estiramento C-H nas moléculas de álcool (região esverdeada do gráfico de pesos), que ocorre por volta de  $4.600\text{-}4.000\text{ cm}^{-1}$  [84], foram encontrados para a PC1, PC2 e principalmente na PC3, que justificaria a dispersão das amostras não adulteradas e adulteradas com metanol.

A partir do gráfico de escores apresentado a seguir, na **Figura 4.6**, é verificado que ocorre uma tendência de formação de agrupamentos das classes por meio da combinação das componentes principais PC1 e PC2. Observa-se que a grande maioria dos escores das amostras adulteradas com metanol é encontrado na direção dos valores positivos da PC1, enquanto que as demais amostras apresentam valores de escores mais negativos. A PC2 é responsável por discriminar as amostras adulteradas com água (escores negativos) daquelas não adulteradas (escores positivos).



**Figura 4.6** – Gráfico de escores da PCA (PC1xPC2) aplicada aos espectros MIR originais. (■: amostras não adulteradas, ●: amostras adulteradas com água, ▲: amostras adulteradas com metanol)

Na **Figura 4.7** é apresentado o gráfico dos pesos obtidos pela PCA aplicada aos espectros MIR. Possivelmente, os valores de pesos positivos para PC1, que aparecem em torno de  $1030\text{ cm}^{-1}$  (região avermelhada do gráfico de pesos), podem estar relacionados aos estiramentos C-O [87] ( $1.030\text{ cm}^{-1}$ ). Para os pesos em PC2, duas regiões com valores negativos são encontradas: uma delas, citada anteriormente para a PC1, e outra próxima de  $2960\text{ cm}^{-1}$  (região esverdeada do gráfico de pesos), que pode ser atribuídos aos estiramentos C-H [87]. Os valores de peso em torno de  $2345\text{ cm}^{-1}$  estão relacionados indiretamente com a presença de água. As ligações de hidrogênio das moléculas de água interagem diminuindo a intensidade de vibrações dos grupos nitrilas.



**Figura 4.7** – Gráficos dos pesos da PCA aplicada aos espectros MIR das amostras de AEHC.

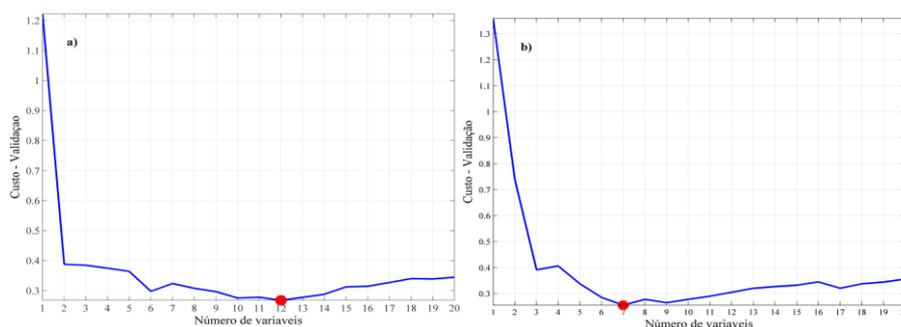
### 4.1.3. Validação dos modelos de classificação

#### 4.1.3.1. Seleção de variáveis e validação dos modelos LDA

Os modelos LDA foram criados e validados utilizando cadeias de variáveis selecionadas baseadas nos parâmetros utilizados por cada algoritmo (SPA, SW e GA). Os resultados são apresentados a seguir.

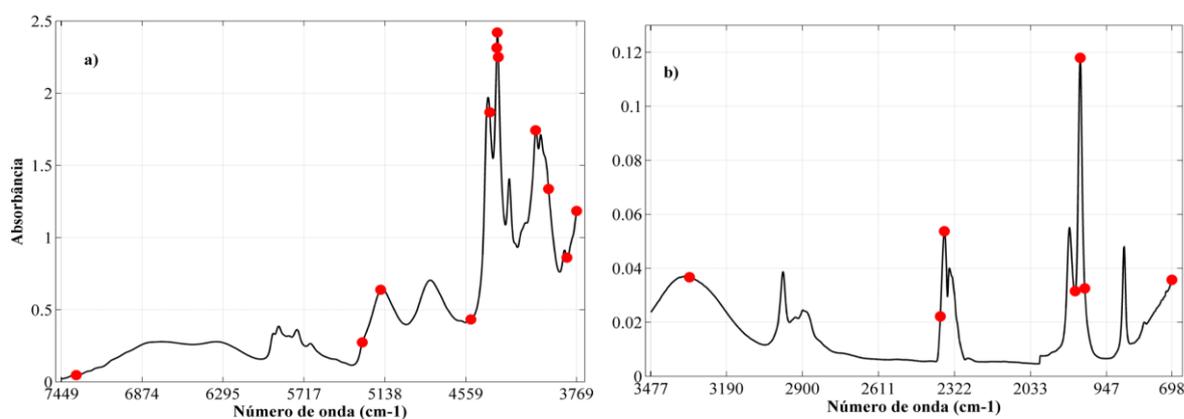
Na **Figura 4.8** é apresentado o gráfico de custo utilizado pelo o SPA como critério para seleção da cadeia de variáveis. Como pode ser visto, as

curvas exibem um mínimo em 12 e 7 variáveis para os dados (a) NIR e (b) MIR, respectivamente.



**Figura 4.8** – Gráficos de custo para seleção da cadeia de variáveis pelo SPA nos dados (a) NIR e (b) MIR.

A seguir, na **Figura 4.9**, são apresentadas as variáveis selecionadas pelo SPA.



**Figura 4.9** – Espectros médios com as variáveis selecionadas pelo SPA para os dados (a) NIR e (b) MIR.

As variáveis selecionadas pelo SPA para o modelo usando os dados NIR (**Figura 4.9a**) encontram-se razoavelmente bem distribuídas em diferentes regiões dos espectros. Em particular, variáveis selecionadas em torno de  $4.330\text{ cm}^{-1}$  e  $4.260\text{ cm}^{-1}$ , podem ser associadas ao segundo sobretom de flexão e à combinação de estiramento -CH-, respectivamente. [86] Variáveis selecionadas em torno de  $4.000\text{ cm}^{-1}$  e  $3.760\text{ cm}^{-1}$  estão relacionadas à região de combinação de vibração C-H.

Ainda, um número de onda isolado foi selecionado em torno de  $7.441\text{ cm}^{-1}$  e pode ser associado à região de combinação de estiramento C-H.

As variáveis selecionadas pelo SPA para o modelo usando os dados MIR (**Figura 4.9b**) também se encontram razoavelmente bem distribuídas em diferentes regiões dos espectros. Variáveis referentes aos picos entre  $1.180\text{ cm}^{-1}$  e  $840\text{ cm}^{-1}$  foram selecionadas, as quais podem ser relacionadas a estiramentos C-O e C-C-O [87]. A seleção de variáveis próximas de  $2300\text{ cm}^{-1}$  pode ser associada à presença de um grupo nitrila ( $\text{C}\equiv\text{N}$ ) [36]. Ainda, uma variável em torno de  $3400\text{ cm}^{-1}$  foi selecionada, que corresponde a estiramento OH.

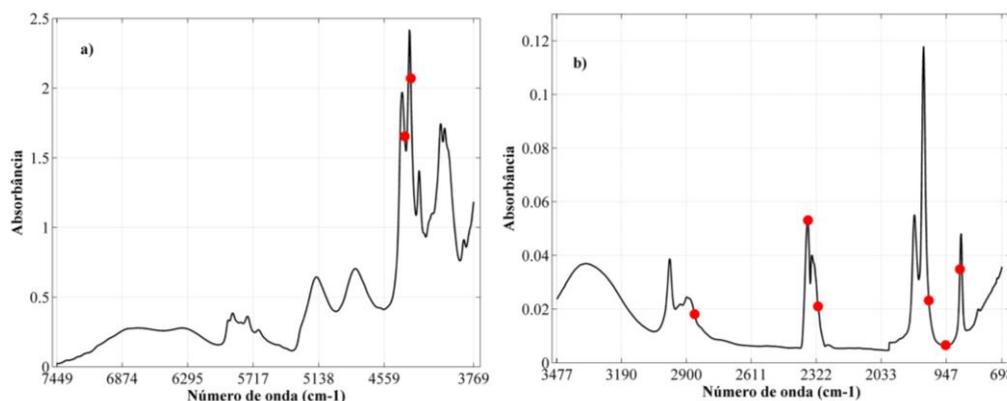
Os modelos SPA-LDA construídos resultaram em uma taxa de classificação correta para o conjunto de validação de 100%, e 97,7% (44 das 45 amostras classificadas corretamente), para os dados do NIR e MIR, respectivamente.

Na **Tabela 4.1** são observados os erros de classificação para o grupo de validação utilizando o algoritmo stepwise (SW) para diferentes valores de correlação. Como resultado, os menores valores de erros de classificação das amostras de validação foram encontrados com duas variáveis selecionadas para os dados NIR e seis para o MIR.

**Tabela 4.1** – Resultados de classificação apresentados pelo algoritmo Stepwise (SW) para o conjunto de validação.

NIR							
Coeficiente Correlação	0,10	0,30	0,50	0,70	0,90	0,95	0,99
Erros de validação	3	1	3	3	4	4	5
N. de variáveis selecionadas	2	2	3	3	4	4	4
MIR							
Coeficiente Correlação	0,10	0,30	0,50	0,70	0,90	0,95	0,99
Erros de validação	3	3	3	1	0	0	1
N. de variáveis selecionadas	2	4	4	5	6	7	9

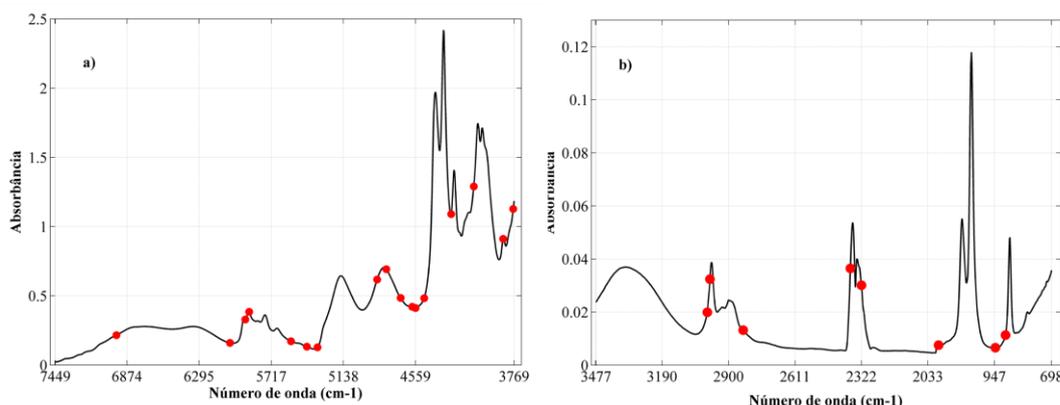
Na **Figura 4.10** são apresentados os espectros médios com as variáveis selecionadas pelo SW para os dados **(a)** NIR e **(b)** MIR.



**Figura 4.10** – Espectros médios com as variáveis selecionadas pelo SW para os dados **(a)** NIR e **(b)** MIR.

As variáveis selecionadas pelo SW para o modelo usando os dados NIR (**Figura 4.10a**) apresentaram-se localizadas em apenas uma região do espectro, especificamente, em torno de  $4.330\text{ cm}^{-1}$  e  $4.260\text{ cm}^{-1}$ , que pode ser relacionado ao segundo sobretom de flexão -CH- e à combinação de estiramentos -CH-, respectivamente [86]. Para os dados MIR (**Figura 4.10b**), as variáveis selecionadas encontraram-se melhor distribuídas pelo espectro, sendo que, seleções em torno dos picos entre  $1.180\text{ cm}^{-1}$  e  $840\text{ cm}^{-1}$  podem ser relacionada a estiramentos C-O e C-C-O [87] e em torno de  $2.900\text{ cm}^{-1}$ , correspondem ao estiramento C-H [84]. Variáveis selecionadas próximas ao número de onda de  $2300\text{ cm}^{-1}$  estão relacionadas com o grupo nitrila ( $\text{C}\equiv\text{N}$ ) [36]. Ainda, uma variável em torno de  $3400\text{ cm}^{-1}$  foi selecionada, que corresponde a estiramento OH.

Para o GA, os melhores resultados das 10 execuções testadas resultaram em um total de 17 variáveis para os dados do NIR, enquanto que 8 variáveis para os dados do MIR. As variáveis selecionadas podem ser observadas na **Figura 4.11**.



**Figura 4.11** – Espectros médios com as variáveis selecionadas pelo GA para os dados (a) NIR e (b) MIR.

As variáveis selecionadas pelo GA para o modelo usando os dados NIR (**Figura 4.11a**) apresentaram-se em maior número quando comparadas com as estratégias utilizadas anteriormente, sendo que as seleções em torno de  $4.000\text{ cm}^{-1}$  e  $3.760\text{ cm}^{-1}$  estão relacionadas à região de combinação de e vibração C-H. Variáveis próximas do número de onda  $5.867\text{ cm}^{-1}$  podem ser atribuídas ao grupo metil ( $-\text{CH}_2$ ) [84]. Para as variáveis selecionadas para os dados MIR (**Figura 4.11b**), destacam-se os picos em torno de  $1.180\text{ cm}^{-1}$  e  $840\text{ cm}^{-1}$  relacionados a estiramentos C-O e C-C-O [87]. Outra região que merece destaque encontra-se entre  $3.000\text{ cm}^{-1}$  e  $2.900\text{ cm}^{-1}$ , correspondendo ao estiramento C-H [84]. Regiões provavelmente relacionadas ao grupo nitrila ( $\text{C}\equiv\text{N}$ ) ( $2300\text{ cm}^{-1}$ ) [36], foram selecionadas.

Os modelos GA-LDA construídos com essas variáveis alcançaram uma TCC de 100% para o conjunto de validação tanto para os dados do NIR, quanto para MIR.

#### 4.1.4. Uso dos modelos de classificação para Previsão

Os modelos de LDA obtidos com as variáveis selecionadas pelos algoritmos SW, SPA e GA, assim como o modelo de PLS-DA, foram aplicados para a classificação do conjunto de teste. Os resultados são

apresentados de forma detalhada nas **tabelas 4.2** (dados do NIR) e **4.3** (dados do MIR).

**Tabela 4.2** – Resultados de classificação dos modelos SW-LDA, SPA-LDA, GA-LDA e PLS-DA aplicados aos espectros NIR para o conjunto de teste.

Dados NIR	SW-LDA (2)*			SPA-LDA (12)			GA-LDA (17)			PLS-DA (9)			
	Classe Predita			Classe Predita			Classe Predita			Classe Predita			
Índice Classe	N	1	2	3	1	2	3	1	2	3	1	2	3
<b>1</b>	15	14	1	-	15	-	-	15	-	-	15	-	-
<b>2</b>	15	5	10	-	1	14	-	-	15	-	-	15	-
<b>3</b>	15	1	-	14	-	-	15	-	-	15	-	-	15

Dados: Amostras de AEHC (1) não adulteradas, (2) adulteradas com água e (3) adulteradas com metanol. O número de variáveis ou fatores utilizado em cada modelo é indicado entre parênteses. N indica o número de amostras de teste utilizado neste estudo, \* valor de limiar: 0.3.

A partir dos resultados observados na **Tabela 4.2**, foi constatado que os piores resultados dos modelos LDA foram obtidos com o uso das variáveis selecionadas com o algoritmo SW, que apresentou uma taxa de classificação correta (**TCC**) de 84,4% para as amostras de teste. Tal resultado pode apresentar relações com o baixo número de variáveis selecionado por esse método (**Figura 4.10a**), e ainda, as mesmas se encontram em regiões muito próximas, o que diminui o nível de informações discriminantes se comparado às demais estratégias de seleção. Dentro dessa perspectiva, os modelos com maiores números de variáveis selecionadas, GA-LDA e PLS-DA (espectro completo), apresentaram os melhores resultados, com uma **TCC** de 100% para as amostras de teste. Ainda, o modelo de classificação SPA-LDA, teve uma taxa de classificação correta (**TCC**) de 97,8% para as amostras de teste.

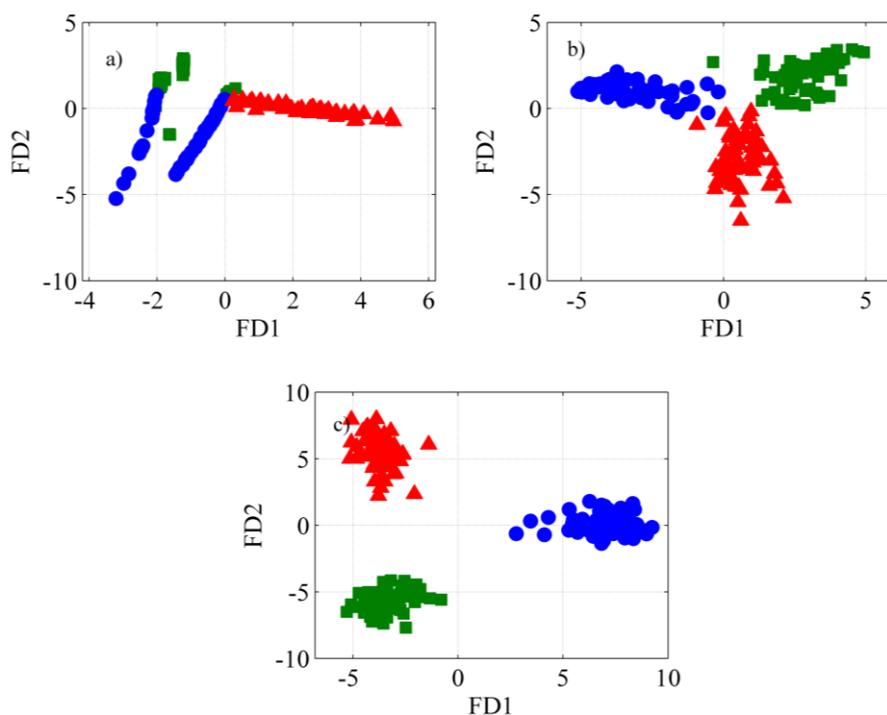
**Tabela 4.3** – Resultados de classificação dos modelos SW-LDA, SPA-LDA, GA-LDA e PLS-DA aplicados aos espectros MIR para o conjunto de teste.

Dados MIR	SW-LDA (6)*			SPA-LDA (7)			GA-LDA (8)			PLS-DA (11)			
	Classe Preditada			Classe Preditada			Classe Preditada			Classe Preditada			
Índice Classe	N	1	2	3	1	2	3	1	2	3	1	2	3
<b>1</b>	15	15	-	-	15	-	-	15	-	-	15	-	-
<b>2</b>	16	-	16	-	-	16	-	-	16	-	-	16	-
<b>3</b>	14	-	-	14	-	-	14	-	-	14	-	-	14

Dados: Amostras de AEHC (1) não adulteradas, (2) adulteradas com água e (3) adulteradas com metanol. O número de variáveis ou fatores utilizado em cada modelo é indicado entre parênteses. N indica o número de amostras de teste utilizado neste estudo, \* valor de limiar: 0.90.

Os resultados de classificação para os dados na região do MIR (**Tabela 4.3**), observa-se que todos os modelos de classificação, SW-LDA, SPA-LDA, GA-LDA e PLS-DA, apresentaram resultados correspondentes a uma taxa de classificação correta (**TCC**) de 100% para as amostras de teste. Tal resultado pode ser relacionado ao que foi comentado anteriormente, já que todas as estratégias de seleção de variáveis apresentaram seleções em diferentes pontos do espectro. Outro fator importante é que a região do MIR apresenta modificações significativas na distribuição dos picos de absorção mediante pequenas modificações na estrutura ou constituição das moléculas. É por essa característica que a região do MIR recebe o nome de região de impressão digital <sup>[35]</sup>.

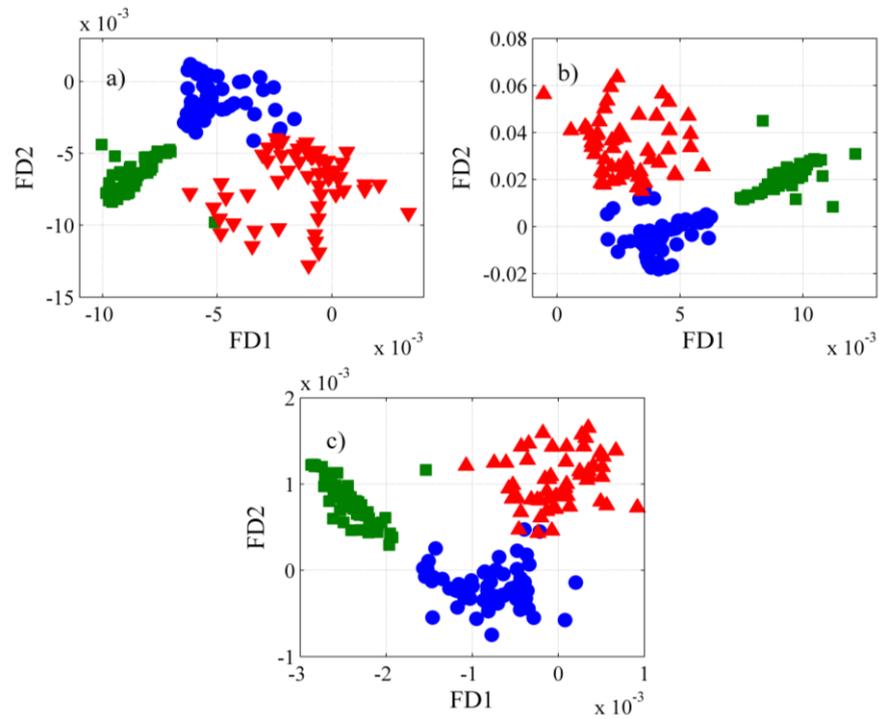
Por meio de uma inspeção visual realizada nos gráficos dos escores (**Figura 4.12 e 4.13**), é possível verificar a capacidade discriminante dos três métodos avaliados.



**Figura 4.12** – Gráfico dos escores (FD1×FD2) para o conjunto completo de espectros NIR utilizando variáveis selecionadas pelo (a) SW, (b) SPA e (c) GA. (■: amostras não adulteradas, ●: amostras adulteradas com água, ▲: amostras adulteradas com metanol)

É possível observar na **Figura 4.12** que a melhor discriminação das classes é encontrada com os dados na região do NIR quando o GA é utilizado. Neste caso, as amostras adulteradas com água são separadas dos demais grupos ao longo de FD1. Já a FD2 é responsável por discriminar as amostras não adulteradas daquelas contaminadas com metanol.

Observando a **Figura 4.13**, é possível verificar que os escores para os diferentes modelos LDA utilizando os dados MIR, apresentaram perfis bastante semelhantes, sendo que a discriminação entre as amostras não adulteradas e adulteradas ocorre ao longo da FD1. Além disso, a FD2 é responsável por discriminar o tipo de adulteração.

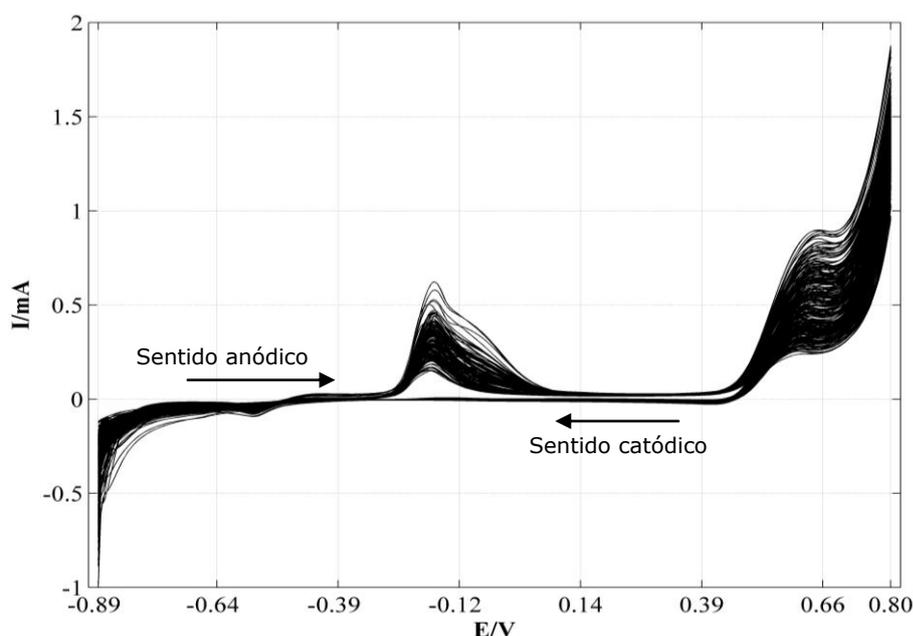


**Figura 4.13** – Gráfico dos escores (FD1×FD2) para o conjunto completo de espectros MIR utilizando variáveis selecionadas pelo (a) SW, (b) SPA e (c) GA. (■: amostras não adulteradas, ●: amostras adulteradas com água, ▲: amostras adulteradas com metanol)

## 4.2. Estudo com Voltametria cíclica

### 4.2.1. Avaliação dos Voltamogramas

Os voltamogramas cíclicos registrados entre os potenciais de -0,9 V a 0,8 V para todas as amostras analisadas são apresentados na **Figura 4.14**.

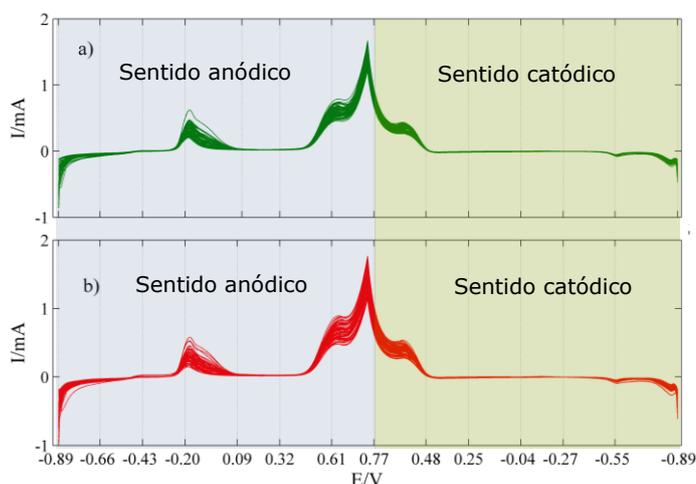


**Figura 4.14** – Voltamogramas cíclicos das amostras de AEHC, registrados na janela de potencial de - 0,9 V a 0,8 V.

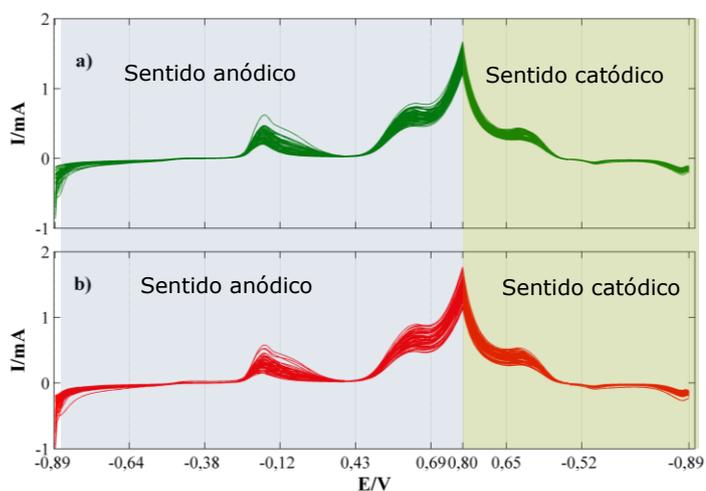
Os voltamogramas apresentados anteriormente estão de acordo com aqueles encontrados na literatura [48-50], em que, os principais picos encontrados correspondem provavelmente à formação da espécie Cu (II) (pico anódico em aproximadamente -0,20V), e à formação de Cu (III) e oxidação das moléculas de etanol e metanol no eletrodo (pico anódico potencial de aproximadamente 0,6V).

Para o estudo quimiométrico, três diferentes regiões de trabalho foram avaliadas. A primeira consistiu em analisar todo o perfil voltamétrico, sem remoção de variáveis, como apresentado na **Figura 4.15**. A segunda estratégia baseou-se na remoção das áreas do

voltamograma que apresentaram valores de corrente próximo de zero, ou seja, regiões de baixa informação e de alta correção. O resultado é apresentado na **Figura 4.16**.



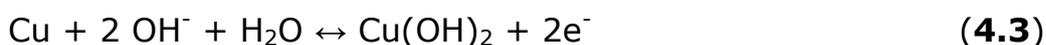
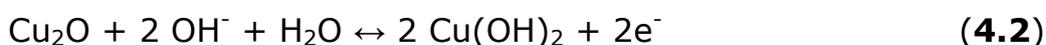
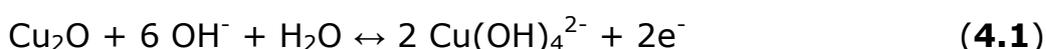
**Figura 4.15** – Voltamogramas cíclicos das amostras de AEHC, no sentido anódico (0,4 a 0,8 V), para amostras (a) não adulteradas e (b) adulteradas com metanol.



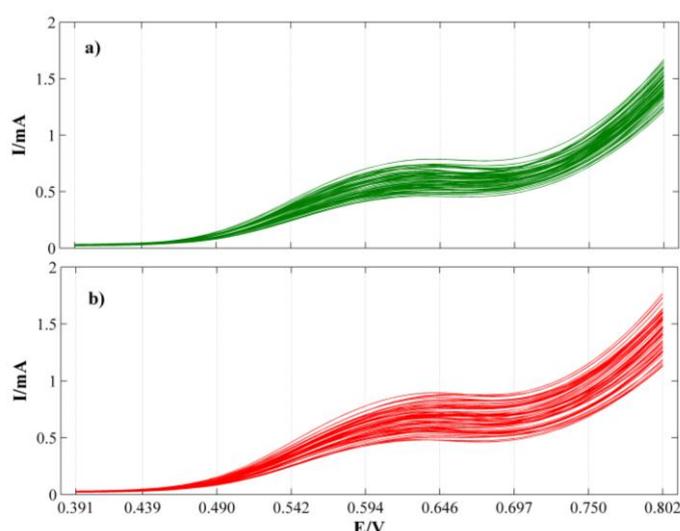
**Figura 4.16** – Voltamogramas cíclicos das amostras de AEHC, com remoção de áreas não informativas, para amostras (a) não adulteradas e (b) adulteradas com metanol.

Quando utilizadas as áreas apresentadas na **Figura 4.15** e **4.16** os piores resultados para os modelos de classificação foram encontrados. Dois fatores podem está relacionados à deficiência dos modelos. Um deles

está relacionado à presença de áreas com variáveis não informativas, as quais poderiam ser incluídas nos modelos, e não auxiliar na discriminação das amostras. Outro fator é a região próxima de -0,20 V, onde ocorre a formação de Cu (II). Nessa faixa de potencial não existem relações diretas com os componentes do AEHC, ainda, diferentes mecanismos podem acontecer paralelamente para formação do Cu (II) <sup>[89]</sup> (**Equação 4.1** a **4.4**). A alta variabilidade na formação da espécie Cu (II), pode influenciar negativamente nos modelos de classificação.



E por último, foi avaliada apenas a região onde ocorre a oxidação das moléculas de etanol e metanol (0,40 a 0,80V, sentido anódico), as quais estão presentes no AEHC. A região é apresentada na **Figura 4.17**.

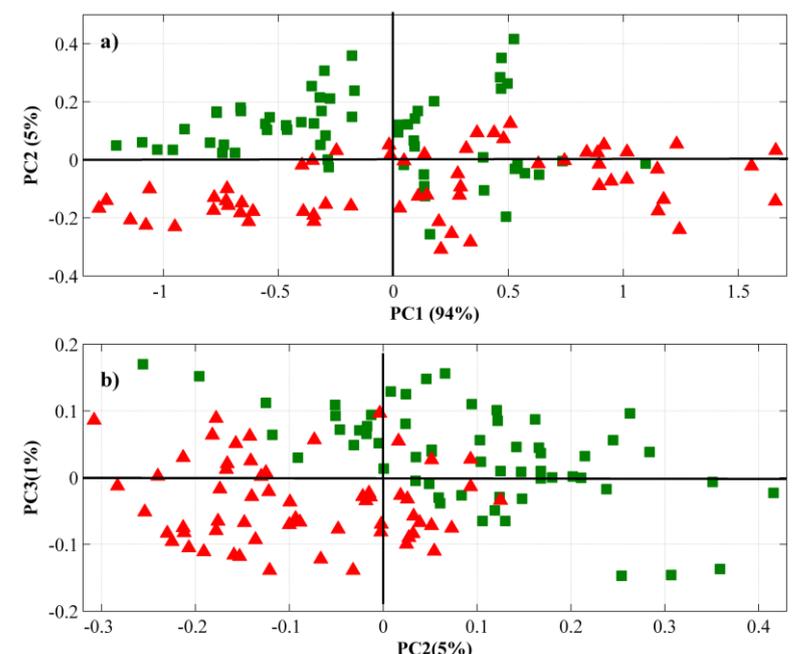


**Figura 4.17** – Voltamogramas das amostras de AEHC na região de oxidação de moléculas de álcool, para amostras (a) não adulteradas e (b) adulteradas com metanol.

Os melhores resultados de classificação foram alcançados utilizando apenas o sentido anódico na faixa potencial de 0,4 a 0,8 V (**Figura 4.17**). A seguir são apresentados os estudos quimiométricos para essa região.

#### 4.2.2. Análise Exploratória dos dados Voltamétricos

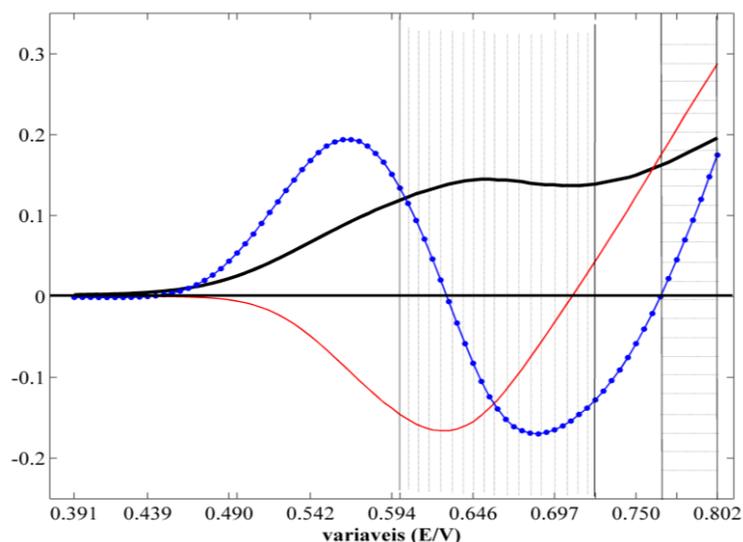
Com o intuito de verificar a capacidade discriminatória baseado na variabilidade dos voltamogramas, uma PCA foi realizada, e os escores (**Figura 4.18**) e pesos (**Figura 4.19**) foram avaliados para as 120 amostras.



**Figura 4.18** – Gráficos dos escores pela PCA aplicada aos Voltamogramas das amostras de AEHC. (a) PC1×PC2 e (b) PC1×PC3 (■: amostras não adulteradas, ▲: amostras adulteradas com metanol)

Avaliando o gráfico de escores apresentado na **Figura 4.20**, verifica-se que, ao longo de PC1, há uma grande dispersão das classes não adulteradas e adulteradas. Já para PC2, verifica-se que existe certa tendência de discriminação, sendo que, as amostras não adulteradas apresentam-se com valores de escores positivos e adulteradas, valores de

escores negativos. Ainda, no gráfico de PC2 x PC3 (**Figura 4.20b**), a tendência de discriminação observada pela PC2 torna-se mais visível. A seguir, na **Figura 4.21**, são apresentados os gráficos de pesos de PCA.



**Figura 4.19** – Gráficos dos pesos da PCA aplicada aos voltamogramas cíclicos das amostras de AEHC. (PC1: —, PC2: —, PC3: —●—).

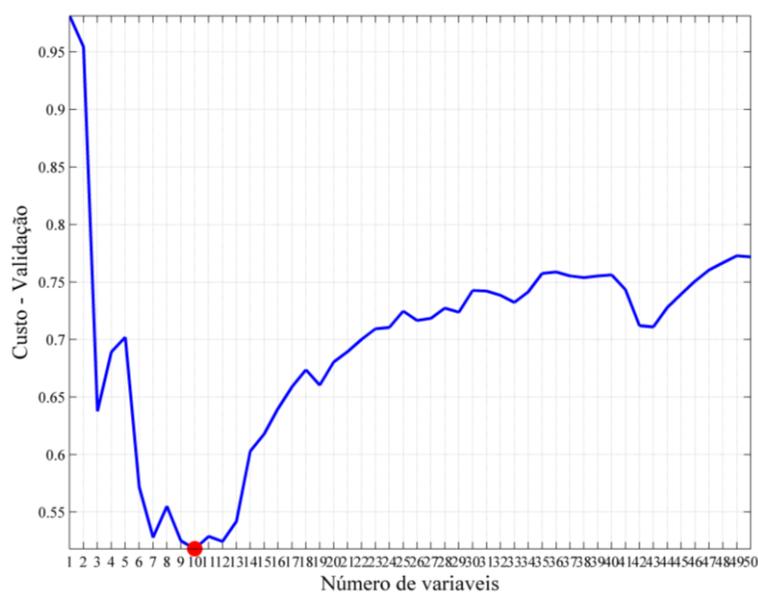
Ao avaliar os resultados obtidos com o gráfico de pesos (**Figura 4.19**), verifica-se que os mesmos corroboram com o comportamento dos escores (**Figura 4.18**), no qual a PC2 combinada com a PC3, forneceu maior tendência de separação entre as classes. Esse fato está relacionado a PC3 e principalmente a PC2 possuírem maiores valores de pesos na região referente aos potenciais onde ocorre a oxidação das moléculas de etanol (por volta de 0,7V - área de linhas horizontais). A partir desse potencial, a intensidade da corrente tende a variar principalmente de acordo com a quantidade de etanol [30, 88]. Ainda, pesos negativos estão relacionados a potenciais próximos a 0,5V (área de linhas verticais), que estão relacionados à formação de Cu (III) [45] e oxidação do metanol [51]. O perfil de pesos apresentado pela PC1 assemelha-se ao voltamograma em estudo (**Figura 4.17**), comportamento que pode estar relacionado à multicolinearidade apresentada pelas variáveis.

### 4.2.3. Validação dos Modelos de Classificação

#### 4.2.3.1. Seleção de variáveis e validação dos modelos LDA

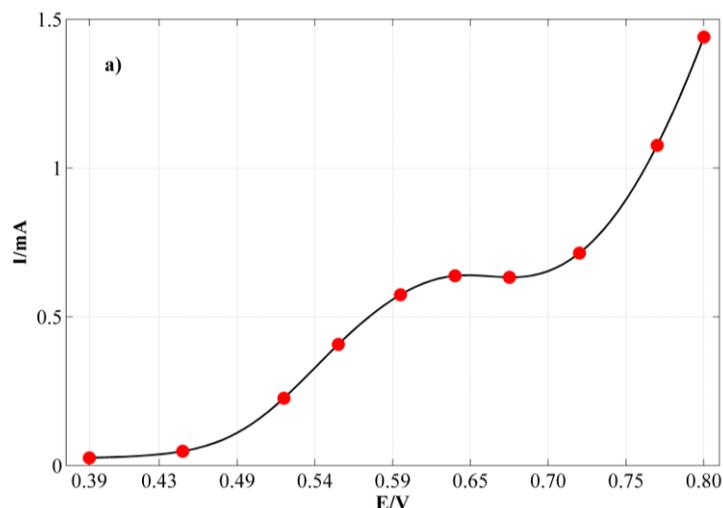
Os modelos LDA foram criados utilizando cadeias de variáveis selecionadas baseadas nos parâmetros utilizados por cada algoritmo (SPA, SW e GA). Em seguida os modelos foram validados. Os resultados são observados a seguir.

Para o SPA o critério utilizado para seleção da cadeia de variáveis é o gráfico de custo, o qual é apresentado na **Figura 4.20**. Como resultado, a curva exibe um mínimo em 10 variáveis.



**Figura 4.20** – Gráficos de custo para seleção da cadeia de variáveis pelo SPA nos dados de voltametria.

Na **Figura 4.21**, são apresentadas as variáveis selecionadas pelo SPA.



**Figura 4.21** – Voltamograma médio com as variáveis selecionadas pelo SPA para os dados de voltametria.

As variáveis selecionadas pelo SPA para o modelo usando os dados de voltametria (**Figura 4.21**) encontram-se bem distribuídas ao longo de todo o voltamograma. As variáveis selecionadas próximas à faixa de potencial de 0,55 a 0,70 V podem ser atribuídas à presença de metanol [51] nas amostras de AEHC, assim como à mudança da concentração de etanol [30,88]. Já os potenciais selecionados próximos a 0,8 V estão relacionados aos processos anódicos que ocorrem mais fortemente com o etanol.

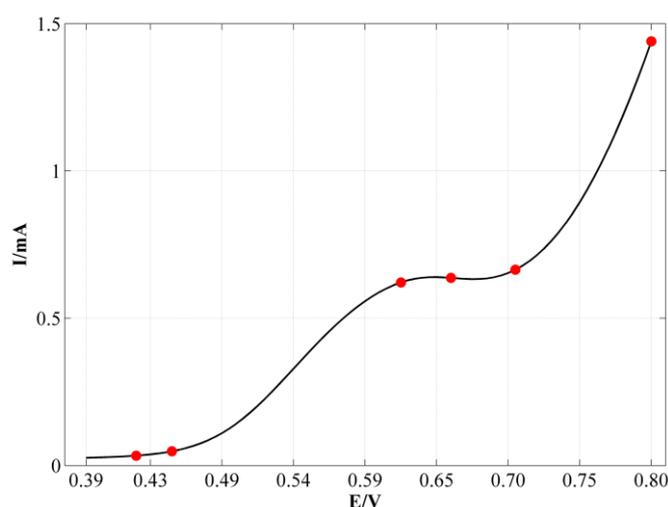
O modelo SPA-LDA construído resultou em uma **TCC** para o conjunto de validação de 90% (27 das 30 amostras classificadas corretamente), para os dados de voltametria.

Na **Tabela 4.4** são fornecidos os erros de classificação para o grupo de validação utilizando o algoritmo stepwise (SW) para diferentes valores de correlação utilizando os dados de voltametria. Como resultado, os menores valores de erros foram encontrados com 6 variáveis selecionadas.

**Tabela 4.4** – Resultados de classificação apresentados pelo algoritmo Stepwise (SW) para o conjunto de validação usando os dados de voltametria.

Voltametria							
Coeficiente Correlação	0,10	0,30	0,50	0,70	0,90	0,95	0,99
Erros de validação	11	11	10	9	6	6	5
N. de variáveis selecionadas	1	1	2	2	3	4	6

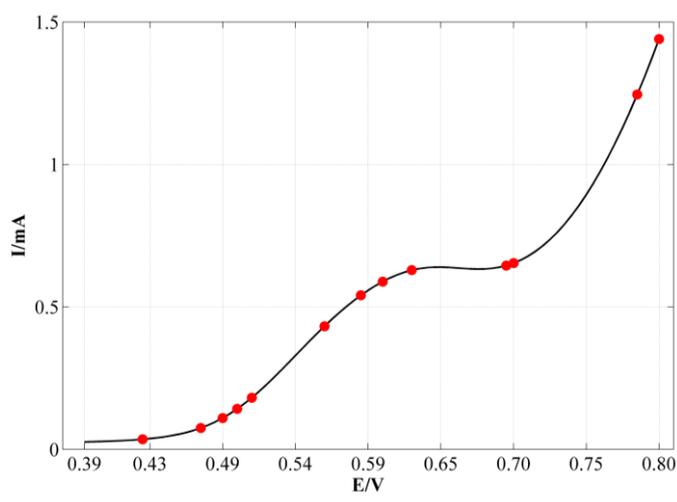
Na **Figura 4.22**, apresentada a seguir, é apresentado o voltamograma médio com as variáveis selecionadas pelo SW para os dados de voltametria.



**Figura 4.22** – Voltamograma médio com as variáveis selecionadas pelo SW para os dados de voltametria.

As variáveis selecionadas pelo SW para o modelo usando os dados de voltametria (**Figura 4.22**) apresentaram-se em três regiões principais. Especificamente, em torno de 0,43V, início da curva de formação de Cu (III) e oxidação do metanol. O número maior de variáveis foi selecionado na região referente à oxidação do metanol e etanol (entre 0,60 e 0,70V) [51, 30,88].

Para o GA, os melhores resultados das 10 execuções testadas resultaram em um total de 13 variáveis para os dados de voltametria. As variáveis selecionadas podem ser observadas na **Figura 4.23**.



**Figura 4.23** – Voltamograma médio com as variáveis selecionadas pelo SPA para os dados de voltametria.

Como pode ser visto, o GA seleciona um número de variáveis maior, quando comparado com os outros algoritmos testados. Tais variáveis encontram-se próximas à faixa de potencial de 0,55 a 0,70 V e podem ser atribuídas à presença de metanol [51] nas amostras, assim como à mudança da quantidade de etanol [30,88].

Os modelos GA-LDA construídos com essas variáveis alcançaram uma **TCC** de 99,6% (29 acertos de 30 amostras) para o conjunto de validação para os dados de voltametria.

#### 4.2.4. Uso dos modelos de classificação para Previsão

Os modelos LDA obtidos com as variáveis selecionadas pelos algoritmos SW, SPA e GA, assim como o modelo PLS-DA, foram aplicados para a classificação do conjunto de teste (**Tabela 4.5**).

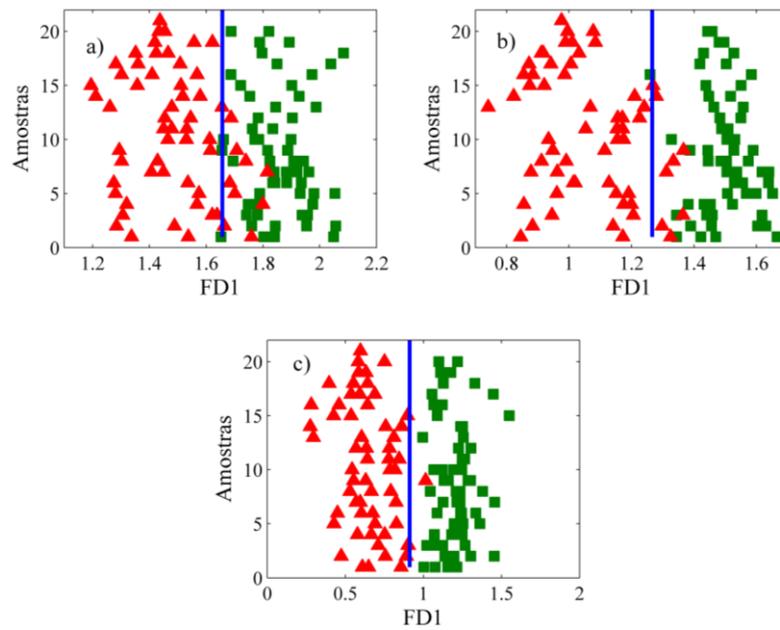
**Tabela 4.5** – Resultados de classificação dos modelos SW-LDA, SPA-LDA, GA-LDA e PLS-DA, aplicados aos voltamogramas para o conjunto de teste.

Dados Voltametria	SW-LDA (6)*		SPA-LDA (10)		GA-LDA (13)		PLS-DA (9)		
	Classe Preditada		Classe Preditada		Classe Preditada		Classe Preditada		
Índice Classe	N	1	2	1	2	1	2	1	2
<b>1</b>	15	15	-	15	-	15	-	15	-
<b>2</b>	15	-	15	2	13	-	15	2	13

Dados: Amostras de AEHC (1) não adulteradas e (2) adulteradas com metanol. O número de variáveis ou fatores utilizado em cada modelo é indicado entre parênteses. N indica o número de amostras de Teste utilizado neste estudo. Limiar: 0.99

A partir dos resultados apresentados na **Tabela 4.5**, observa-se que os modelos LDA obtidos com o uso das variáveis selecionadas com o algoritmo SPA e PLS-DA apresentaram uma menor **TCC** para as amostras de teste (93,3%). Os erros encontrados para esses modelos são provenientes de amostras adulteradas sendo classificadas como não adulteradas. Em contrapartida, os modelos SW-LDA e GA-LDA classificaram corretamente todas as amostras do conjunto de previsão.

Na **Figura 4.24** é exposto o gráfico dos escores obtidos pela LDA usando os diferentes métodos de seleção de variáveis para o conjunto completo de amostras (treinamento, validação e teste). Como pode ser observado, os modelos apresentaram perfis de discriminação bastante parecidos. O modelo GA-LDA (**Figura 4.24c**) apresentou uma menor quantidade de erros globais para todas as amostras quando comparados aos demais modelos.



**Figura 4.24** - Gráfico dos escores (FD1) para o conjunto completo de voltamogramas utilizando variáveis selecionadas pelo (a) SW, (b) SPA e (c) GA. (■: amostras não adulteradas, ▲: amostras adulteradas com metanol).

Capítulo V

# **CONCLUSÕES**

## 5.1 Conclusões

Como conclusões, temos que na aplicação da análise de componentes principais (PCA) para todas as técnicas, foi possível encontrar a formação de alguns agrupamentos referentes ao AEHC não adulterado e aqueles adulterados. Contudo, a sobreposição de algumas amostras também foi observada. Para análise classificatória, modelos SPA-LDA, GA-LDA, SW-LDA e PLS-DA, quando aplicados aos dados provenientes do infravermelho médio, foi observada uma TCC de 100 % para o conjunto de Teste para todos os modelos. Para os dados do NIR, o SPA-LDA e SW-LDA classificaram corretamente 84,4 % e 97,8 %, respectivamente, enquanto o PLS-DA e GA-LDA apresentaram melhores resultados, com 100% das amostras do conjunto de Teste classificadas corretamente. Os modelos também foram aplicados aos dados voltamétricos, sendo que, para esses dados, o SPA-LDA e PLS-DA revelaram desempenhos semelhantes, com 93% de acerto na classificação das amostras de teste. No entanto, melhores resultados são encontrados com o uso dos modelos SW-LDA e GA-LDA, que classificaram corretamente 100% das amostras do conjunto de Teste. Em termos gerais, foi possível realizar a detecção de adulteração em Amostras de Álcool Etílico Combustível (AEHC) por água ou por metanol com o uso de metodologias analíticas alternativas baseadas na espectroscopia no infravermelho (NIR e MIR) e na voltametria cíclica, associada a técnicas quimiométricas de reconhecimento de padrão.

## 5.2 Propostas Futuras

Propõe-se, como continuidade desse trabalho:

- Construir modelos de classificação utilizando potenciostato portátil, com o intuito e fazer a detecção de adulteração do AEHC, direto do posto de combustível.
- Ampliar o modelo desenvolvido utilizando técnicas eletroanalíticas e modelos de calibração multivariada para a determinação de parâmetros de qualidade em amostras de AEHC.
- Desenvolvimento de equipamentos portáteis, a partir da utilização dos números de onda ou potenciais selecionados com as técnicas de seleção de variáveis para os dados NIR e de voltametria.

Capítulo VI

# **REFERÊNCIAS BIBLIOGRÁFICAS**

## 6. REFERÊNCIAS BIBLIOGRÁFICAS

- [1] WYMAN, C. E. Ethanol Fuel. *Reference module in earth systems and environmental sciences, from Encyclopedia of Energy*, p. 541–555, 2004.
- [2] JAEGER, W. K.; EGELKRAUT T. M. Biofuel economics in a setting of multiple objectives and unintended consequences. *Renewable and Sustainable Energy Reviews*, v. 15, p. 4320–4333, 2011.
- [3] SÁNCHEZ, Ó. J.; CARDONA, C. A. Trends in biotechnological production of fuel ethanol from different feedstocks. *Bioresource Technology*, v. 99, p. 5270–5295, 2008.
- [4] RENEWABLE FUELS ASSOCIATION – RFA, World fuel ethanol production, Disponível em: < <http://ethanolrfa.org/pages/World-Fuel-Ethanol-Production>>, Acesso em: 13/06/2013.
- [5] MUSSATO, S. I.; DRAGONE, G.; GUIMARÃES, P.M.R.; SILVA, J. P. A.; CARNEIRO, L. M.; ROBERTO, I. C.; VICENTE, A.; DOMINGUES, L.; TEIXEIRA, J. A. Technological trends, Global market, and challenges of bio-ethanol production. *Biotechnology Advances*, v. 28, p. 817–830, 2010.
- [6] CRAGO, C. L.; KHANNA M., BARTON J. Competitiveness of brazilian sugarcane ethanol compared to US corn ethanol. *Energy Policy*, v. 38, p. 7404–7415, 2010.
- [7] FREITAS, L. C.; KANEKO, S. Ethanol demand in Brazil: Regional approach, *Energy Policy*, v. 39, p. 2289–2298, 2011.
- [8] ROSILLO-CALLE, F.; CORTEZ, L. A. B. Towards ProAlcool II — a review of the Brazilian bioethanol programme. *Biomass and Bioenergy*, v. 14, p. 115–124, 1998.

- [9] ANP. Agência nacional do petróleo, gás natural e biocombustíveis. Resolução ANP Nº 36, de 6.12.2005. Disponível em: < <http://anp.gov.br/>>. Acesso em: 12/06/2013.
- [10] FERREIRA, A. L.; PRADO, F. P. A.; SILVEIRA, J. J. Flex cars and the alcohol price. *Energy Economics*, V. 31, p. 382–394, 2009.
- [11] ANP. Agência nacional do petróleo, gás natural e biocombustíveis. Portaria ANP Nº 45, de 16.3.2001. Disponível em: < <http://anp.gov.br/>>. Acesso em: 12/06/2013.
- [12] MME. Ministério de Minas e Energia. Boletim Mensal dos Combustíveis Renováveis. Edição Nº 61, Fevereiro de 2013.
- [13] SBQ. Sociedade Brasileira de Química. Apresentação. Disponível em: < <http://www.anp.gov.br/>>. Acesso em: 12/06/2013.
- [14] ANP. Agência nacional do petróleo, gás natural e biocombustíveis. Resolução ANP Nº 7, de 9.2.2011. Disponível em: < <http://www.anp.gov.br/>>. Acesso em: 12/06/2013.
- [15] ANP. Agência nacional do petróleo, gás natural e biocombustíveis. Orientações ao consumidor: qualidade e adulteração de combustíveis. Disponível em: < <http://www.anp.gov.br/?pg=46746>>. Acesso em: 11/06/2013.
- [16] ANP. Agência nacional do petróleo, gás natural e biocombustíveis. Boletins da Qualidade. Disponível em: < <http://www.anp.gov.br/>>. Acesso em: 11/06/2013.
- [17] ANP. Agência nacional do petróleo, gás natural e biocombustíveis. Boletim Mensal Do Monitoramento Dos Combustíveis Líquidos Automotivos. Ano 12, Maio 2013, Revisão 1, 23 junho 2013.

- [18] CHEREMISINOFF, N. P., *Industrial Solvents Handbook*, 2nd Ed., Marcel Dekker, Inc: New York, 2003.
- [19] GOSSELIN, R. E.; HODGE, H. C.; SMITH, R. P.; GLEASON, M. N. *Clinical Toxicology of Comercial Products*. Williams & Wilkins, Baltimore, 1976.
- [20] NBR 15531, *Álcool etílico-Determinação do teor de água - Método volumétrico de Karl Fischer*, 2009.
- [21] NBR 15888, *Etanol - Determinação do teor de água - Método coulométrico de Karl Fischer*, 2010.
- [22] ASTM E203, *Standard Test Method for Water Using Voltametric Karl Fischer Titration*, 2008.
- [23] ABNT NBR 16041, *Etanol combustível — Determinação dos teores de metanol e etanol por cromatografia gasosa*, 2012.
- [24] ASTM D5501, *Standard test method for determination of ethanol and methanol content in fuels containing greater than 20% ethanol by gas chromatography*, 2013.
- [25] GRÜNKE, S. Main and side reactions in the Karl Fischer solution *Food Control*, v. 12, p. 419–426, 2001.
- [26] MENDES, L. S.; OLIVEIRA, F. C. C.; SUAREZ, P. A. Z. RUBIM, J. C. Determination of ethanol in fuel ethanol and beverages by Fourier transform (FT)-near infrared and FT-Raman spectrometries. *Analytica Chimica Acta* , v. 493, p. 219–231, 2003.
- [27] FERNANDES, H. L.; RAIMUNDO JR., I. M.; PASQUINI, C.; ROHWEDDER, J. J. R. Simultaneous determination of methanol and ethanol in gasoline using NIR spectroscopy: Effect of gasoline composition. *Talanta*, v. 75, p. 804–810, 2008.

- [28] CARNEIRO, H. S. P.; MEDEIROS, A. R. B.; OLIVEIRA, F. C. C. Determination of Ethanol Fuel Adulteration by Methanol Using Partial Least-Squares Models Based on Fourier Transform Techniques. *Energy & Fuels*, v. 22, p. 2767-2770, 2008.
- [29] SILVA, A. C.; PONTES, L. F. B. L.; PIMENTEL, M. F.; PONTES, M. J. C. Detection of adulteration in hydrated ethyl alcohol fuel using infrared spectroscopy and supervised pattern recognition methods. *Talanta*, v. 93, p. 129-134, 2012.
- [30] DAVIES, T. J.; COMPTON, R. G. The cyclic and linear sweep voltammetry of regular and random arrays of microdisc electrodes: Theory. *Journal of Electroanalytical Chemistry*, v. 585, p.63-82, 2005.
- [31] SOUZA, D.; MACHADO, S. A. S.; AVACA, L. A. Voltametria de onda quadrada. Primeira parte: Aspectos teóricos. *Química Nova*, v.26, p.81-89, 2003.
- [32] PEREIRA, P. F.; SOUSA, R. M.F.; MUNOZ, R. A.A.; RICHTER, E. M. Simultaneous determination of ethanol and methanol in fuel Ethanol using cyclic voltammetry. *Fuel*, v.103, p.725-729, 2013.
- [33] BUENO, L.; PAIXÃO, T. R.L.C. A copper interdigitated electrode and chemometrical tools used for the discrimination of the adulteration of ethanol fuel with water. *Talanta*, v.87, p.210-215, 2011.
- [34] RUBINSON, KENNETH A. *Análisis Instrumental*, Prentice Hall, 2001.
- [35] LIMA, K. M. G.; RAIMUNDO JR. I. M.; SILVA, A. M. S. S.; PIMENTEL, M. F. Sensores ópticos com detecção no infravermelho próximo e médio, *Química Nova*, v. 32, n. 6, p. 1635-1643, 2009.
- [36] SKOOG, D. A.; HOLLER, F. J.; CROUCH, S. R. *Princípios de Análise Instrumental*. 6ª ed. Porto Alegre: Editora Bookman, 2009.

- [37] SALA, O. I<sub>2</sub>- Uma molécula didática. *Química Nova*, v. 31, p. 914-920, 2008.
- [38] PASQUINI, C. Near Infrared Spectroscopy: Fundamentals, Practical Aspects and Analytical Applications. *Journal of the Brazilian chemical society*, v.14, p.198-219, 2003.
- [39] COATES, J. Interpretation of Infrared Spectra, A Practical Approach. *Encyclopedia of Analytical Chemistry*, p. 10815–10837, 2000.
- [40] Handbook of Instrumental Techniques for Analytical Chemistry, Frank A. Settle, 1997, Prentice Hall PTR (ECS Professional).
- [41] SOUZA, D.; MACHADO, S. A. S.; AVACA, L. A. Voltametria de onda quadrada. Primeira parte: Aspectos teóricos. *Química Nova*, v.26, p. 81-89, 2003.
- [42] SKOOG, WEST, HOLLER, CROUCH. Fundamentos de Química Analítica, Tradução da 8ª Edição norte-americana, Editora Thomson, São Paulo-SP, 2006.
- [43] CHEN, Z.-L.; HIBBERT, D.B. Simultaneous amperometric and potentiometric detection of sugars, polyols and carboxylic acids in flow systems using copper wire electrodes. *Journal of Chromatography A*, v.766, p.27–33, 1997.
- [44] FU, C.; SONG, L.; FANG, Y. Simultaneous determination of sugars and organic acids by co-electroosmotic capillary electrophoresis with amperometric detection at a disk-shaped copper electrode. *Analytica Chimica Acta*, v. 371, p.81–87, 1998.
- [45] PAIXÃO, T. R. L. C.; CORBO, D.; BERTOTTI, M. Amperometric determination of ethanol in beverages at copper electrodes in alkaline medium. *Analytica Chimica Acta*, v.472, p.123–131, 2002.

- [46] PAIXÃO, T. R. L. C.; BERTOTTI, M., Development of a breath alcohol sensor using a copper electrode in an alkaline medium. *Journal of Electroanalytical Chemistry*, v. 571, p.101–109, 2004.
- [47] PAIXÃO, T. R. L. C.; CARDOSO, J. L.; BERTOTTI, M. The use of a copper microelectrode to measure the ethanol content in gasohol samples. *Fuel*, v.86,p.1185–1191, 2007.
- [48] SILVA, C. R.; CONCEIÇÃO, C. D. C.; BONIFÁCIO, V. G.; FILHO, O. F.; TEIXEIRA, M. F. S. Determination of the chemical oxygen demand (COD) using a copper electrode: a clean alternative method. *Journal of Solid State Electrochemistry*, v.13,p.665-669, 2009.
- [49] MARUTA, A. H.; PAIXÃO, T. R. L. C. Flow injection analysis of free glycerol in biodiesel using a copper electrode as an amperometric detector. *Fuel*, v.91, p.187–191, 2012.
- [50] MILLER, B. Split-Ring Disk Study of the Anodic Processes at a Copper Electrode in Alkaline Solution. *Journal Electrochemical Society*, v.116, p.1675,1969.
- [51] HELI, H.; JAFARIAN, M.; MAHJANI, M.G.; GOBAL, F. Electro-oxidation of methanol on copper in alkaline solution. *Electrochimica Acta*, v.49, p. 4999–5006, 2004.
- [52] KENNEDY, J. F.; KACZMAREK, A. Practical Guide to Chemometrics, 2nd edition, P. Gemperline (Ed.). CRC Press, Boca Raton, FL, USA (2006).
- [53] HOPKE, P. K. The evolution of chemometrics. *Analytica Chimica Acta*, v. 500, p.365–377, 2003.
- [54] ROGGO, Y.; CHALUS, P.; MAURER, L. A review of near infrared spectroscopy and chemometrics in pharmaceutical technologies. *Journal of Pharmaceutical and Biomedical Analysis*, v. 44, p. 683–700, 2007.

- [55] MUEHLETHALER, C.; MASSONNET, G.; ESSEIVA, P. The application of chemometrics on Infrared and Raman spectra as a tool for the forensic analysis of paints. *Forensic Science International*, v.209, p.173–182, 2011.
- [56] VERAS, G.; GOMES, A. A.; SILVA, A. C.; BRITO, A. L. B.; ALMEIDA, P. B. A.; MEDEIROS, E. P. Classification of biodiesel using NIR spectrometry and multivariate techniques. *Talanta*, v.83, p.565–568, 2010.
- [57] ESTEBAN, M.; ARIÑO C.; DÍAZ-CRUZ, J.M. Chemometrics for the analysis of voltammetric data. *Trends in Analytical Chemistry*, v. 25, p.86–92, 2006.
- [58] ESTEBAN, M.; ARIÑO C.; DÍAZ-CRUZ, J.M., 4.13 – Chemometrics in electrochemistry comprehensive chemometrics, Editors-in-Chief: Stephen D. Brown, Romà Tauler, and Beata Walczak, Oxford, 2009, Pages 425–458.
- [59] NI, Y.; KOKOT, S., Does chemometrics enhance the performance of electroanalysis?. *Analytica Chimica Acta*, v626, p.130–146, 2008.
- [60] VALCÁRCEL, M.; CÁRDENAS, S. Qualitative Analysis, Encyclopedia of Analytical Science (Second Edition), p.405-411, 2005.
- [61] MUÑOZ-OLIVAS, R. Screening analysis: an overview of methods applied to environmental, clinical and food analyses, *TrAC - Trends in Analytical Chemistry*, v.23, p. 203–216, 2004.
- [62] VALCÁRCEL, M.; CÁRDENAS, S.; GALLEGO, M. Sample screening systems in analytical chemistry. *Trends in Analytical Chemistry*, v.18, p.685–694, 1999.
- [63] ESBENSEN, K.H.; GELADI, P. Principal Component Analysis: Concept, Geometrical Interpretation, Mathematical Background, Algorithms, History, Practice Comprehensive Chemometrics, Editors-in-Chief:

Stephen D. Brown, Romà Tauler, and Beata Walczak, Oxford, p. 211–226, 2009.

[64] NETO, F. F. G.; MARINO, G.; ARAÚJO, M. C. U.; GALVÃO, R. K. H.; PONTES, M. J. C.; MEDEIROS, E. P.; LIMA, R. S. Classification of edible vegetable oils using square wave voltammetry with multivariate data analysis. *Talanta*, v.77, p. 1660, 2009.

[65] PONTES, M. J. C.; PEREIRA, C. F.; PIMENTEL, M. F.; VASCONCELOS, F.V. C.; SILVA, A. G. B. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. *Talanta*, v.85, p.2159, 2011.

[66] SOUTO, U. T. C. P.; PONTES, M. J. C.; SILVA, E. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SANCHES, F. A. C.; CUNHA, F. A. S.; OLIVEIRA, M. S. R. UV–Vis spectrometric classification of coffees by SPA–LDA. *Food Chem.* v.119,p.368, 2010.

[67] SILVA, C. S.; BORBA, F. S. L.; PIMENTEL, M. F.; PONTES, M. J. C.; HONORATO, R. S.; PASQUINI, C. Classification of blue pen ink using infrared spectroscopy and linear discriminant analysis. *Microchem. J.*, v.109,p.122-127,2013.

[68] BEEBE, K. R.; PELL, R. J.; Seasholtz, M. B.; "Chemometrics: A Practical Guide"; Wiley, New York 1998.

[69] MASSART, D. L. et al., Handbook of Chemometrics and Qualimetrics: Part A. Amsterdam: Elsevier Science B. V., 1997.

[70] Practical guide to chemometrics / edited by Paul Gemperline.--2nd ed. 2006.

[71] BARKER, M.; RAYENS, W. Partial least squares for discrimination. *Journal of Chemometrics*, v.17, p. 166-173, 2003.

[72] ALMEIDA, M. R.; CORREA, D. N.; ROCHA, W. F.C.; SCAFI, F. J.O.; POPPI, R. J. Discrimination between authentic and counterfeit banknotes using Raman spectroscopy and PLS-DA with uncertainty estimation, *Microchemical Journal*, V. 109, p. 170–177, 2013.

[73] FISHER, R. A. The Use of Multiple Measurements in Taxonomic Problems, *F.R.S.*,v.7,p.179–188, 1936.

[74] BRERETON, R.G; Chemometrics: data analysis for the laboratory and chemical plant. John Wiley Professio, 2003, 504 p.

[75] ARAÚJO, M. C. U.; SALDANHA, T. C. B.; GALVÃO, R. K. H. The successive projections algorithm for variable selection in spectroscopic multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems*, v. 57, p.65–73, 2001.

[76] PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; et al., The successive projections algorithm for spectral variable selection in classification problems. *Chemometrics and Intelligent Laboratory Systems*, v.78,p.11–18, 2005.

[77] SOARES, S. F. C.; GOMES, A. A.; ARAUJO, M. C. U.; et al., The successive projections algorithm. *Trends in Analytical Chemistry*, v.42, p.84–98, 2013.

[78] BANDYOPADHYAY, S.; MURTHY, C.A.; PAL, S.K. Pattern classification with genetic algorithms, *Pattern Recognition Letters*, v. 16, p.801–808, 1995.

[79] FILHO, P. A. C.; POPPI, R. J. Algoritmo genético em química. *Química Nova*, v.22, 1999.

[80] CANECA, A. R.; PIMENTEL, M. F.; GALVÃO, R. K. H.; MATTA, C. E.; CARVALHO, F. R.; RAIMUNDO JR., I. M.; PASQUINI, C.; ROHWEDDER, J.

J.R. Assessment of infrared spectroscopy and multivariate techniques for monitoring the service condition of diesel-engine lubricating oils. *Talanta*, v.70, p.344–352, 2006.

[81] SENRA, L. F. A. C.; NANJI, L. C.; MELLO, J. C. C. B. S.; MEZA, L. A. Estudo sobre métodos de seleção de variáveis em DEA. *Pesquisa Operacional*. v.27, 2007.

[82] KENNARD, R. W.; STONE, L. A. Computer Aided Design of Experiments. *Technometrics*, v.11, p.137, 1969.

[83] SOARES, A.S.; GALVÃO, R. K. H. Fault detection using Linear Discriminant Analysis with selection of process variables and time lags. 2010 IEEE International Conference on Industrial Technology (ICIT), p. 217 – 222, 2010.

[84] LIRA, L. F. B.; VASCONCELOS, F. V. C.; PEREIRA, C. F.; PAIM, A. P. S., STRAGEVITCH, L.; PIMENTEL, M. F. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*, v. 89, pg.405–409, 2010.

[85] XIAOBO, Z.; JIEWEN, Z.; POVEY, M. J.W.; HOLMES, M.; HANPIN, M. Variables selection methods in near-infrared spectroscopy. *Analytica Chimica Acta*, v. 667, p.14–32, 2010.

[86] WORKMAN, J.; WEYER, L. Practical guide to interpretative near-infrared spectroscopy. CRC Press, New York, 2008.

[87] SANDLER, S. R.; KARO, W.; BONESTEEL, JO-ANNE; PEARCE, E. M. Experiment 14 – Infrared spectroscopy, polymer synthesis and characterization, San Diego, 1998, Pages 98–107.

[88] M. FLEISCHMANN, K. KORINEK, D. PLETCHER. The kinetics and mechanism of the oxidation of amines and alcohols at oxide-covered

nickel, silver, copper, and cobalt electrodes. *Journal of the Chemical Society*, p.1396, 1972.

[89] J. AMBROSE, R.G. BARRADAS, D.W. SHOESMITH. Investigations of copper in aqueous alkaline solutions by cyclic voltammetry. *Journal of Electroanalytical Chemistry and Interfacial Electrochemistry*, v.47, p.47, 1973.

**ANEXOS**



Contents lists available at ScienceDirect

Electrochimica Acta

journal homepage: [www.elsevier.com/locate/electacta](http://www.elsevier.com/locate/electacta)

## An electroanalytical method to detect adulteration of ethanol fuel by using multivariate analysis



Adenilton Camilo Silva, José Eduardo Matos Paz, Liliana F.B. Lira Pontes, Sherlan Guimarães Lemos, Márcio José Coelho Pontes\*

Universidade Federal da Paraíba, Departamento de Química, João Pessoa, PB, Brazil

### ARTICLE INFO

Article history:  
Received 1 January 2013  
Received in revised form 5 June 2013  
Accepted 28 July 2013  
Available online xxx

**Keywords:**  
Hydrated ethyl alcohol fuel  
Voltammetry  
Copper electrode  
Linear discriminant analysis  
Variable selection

### ABSTRACT

This work proposes a simple and inexpensive electroanalytical method to identify adulteration in hydrated ethyl alcohol fuel (HEAF) samples by using multivariate analysis. The methodology was validated in a case study with HEAF samples divided into two categories: (1) unadulterated and (2) adulterated with methanol in different concentration levels (1.0–13%  $\text{m} \cdot \text{m}^{-3}$ ). Staircase voltammetry was the electrochemical technique employed. Voltammograms were recorded with a copper electrode in an alkaline solution. Four different pattern recognition methods were evaluated for classification of the resulting voltammograms: partial least squares discriminant analysis (PLS-DA); linear discriminant analysis (LDA) with variable selection by the successive projections algorithm (LDA/SPA), genetic algorithm (LDA/GA) and a stepwise formulation (LDA/SW). The best results were achieved with LDA/GA and LDA/SW models, which all prediction samples were correctly classified.

© 2013 Elsevier Ltd. All rights reserved.

### 1. Introduction

The utilization of biomass for energy production is a subject of great importance all over the world. In fact, with the spread of initiatives to prevent climate change, corporate investments in renewable biofuels have increased substantially in recent years. Ethanol produced on a large scale from sugarcane industries is the biofuel most widely used in Brazil. As automotive fuel, it can be found in an anhydrous form, which is blended with gasoline, or as hydrated ethyl alcohol fuel (HEAF) [1]. In the last years, the illegal addition of methanol to HEAF has become a concern of Brazilian regulatory agencies. In fact, the lower price of methanol in relation to ethanol and similar physical–chemical properties (density values, solubility in water, organoleptic properties) of these alcohols [2] contribute to the adulteration of HEAF with methanol in Brazil. This illicit practice could cause serious problems for both the economy and the health of consumers [3]. In Brazil, the acceptable maximum level of methanol in HEAF is 1.0% (v/v) [4].

Recently, the Brazilian Association of Technical Standards (ABNT) and the American Society for Testing and Materials (ASTM

International) have published two standards [5,6] to determine the amounts of ethanol and methanol in ethanol fuel by using gas chromatography (GC). However, GC presents some inconveniences such as the time required to prepare a large variety of standard solutions that can generate dangerous residues, and the use of expensive and high-maintenance instruments.

There are a number of works that have been developed to monitor the quality of ethanol fuel with emphasis on determination of quality parameters using different analytical methods [7–10]. However, little specific research has been devoted to detect adulteration in ethanol fuel [10–13] particularly in terms of adulteration of ethanol fuel with methanol [10–12].

The works reported in Refs. [12,13] exemplify the potential that electroanalytical techniques could have for building classification models based on existing methanol content in HEAF samples. Therefore, it is possible to combine the low cost and fast response of these techniques with the accuracy required to create screening methods for HEAF adulteration, reducing the effective number of samples submitted to analysis by official chromatographic method, thereby reducing the total cost of a monitoring program. In this sense, the present work proposes a simple and low-cost electrochemical method to detect the adulteration of HEAF samples with methanol by using staircase voltammetry performed with a copper oxide electrode and pattern recognition techniques.

The use of copper-based electrodes is supported by its electrocatalytic activity toward the anodic oxidation of simple alcohols and small organic molecules in highly alkaline medium [14–21]. Copper has the highest rate constant among several oxide covered

\* Corresponding author at: Universidade Federal da Paraíba, Departamento de Química – Laboratório de Automação e Instrumentação em Química Analítica/Quimilometria LAQA, CEP 58051-970 João Pessoa, PB, Brazil.  
Tel.: +55 83 3216 7438; fax: +55 83 3216 7438.  
E-mail addresses: [marciocoelho@quimica.ufpb.br](mailto:marciocoelho@quimica.ufpb.br), [marcio.quimica@gmail.com](mailto:marcio.quimica@gmail.com) (M.J.C. Pontes).



Contents lists available at ScienceDirect

Talanta

journal homepage: [www.elsevier.com/locate/talanta](http://www.elsevier.com/locate/talanta)

## Detection of adulteration in hydrated ethyl alcohol fuel using infrared spectroscopy and supervised pattern recognition methods

Adenilton Camilo Silva<sup>a</sup>, Liliana Fátima Bezerra Lira Pontes<sup>a</sup>, Maria Fernanda Pimentel<sup>b</sup>, Márcio José Coelho Pontes<sup>a,\*</sup><sup>a</sup> Universidade Federal da Paraíba, Departamento de Química, João Pessoa, PB, Brazil<sup>b</sup> Universidade Federal de Pernambuco, Departamento de Engenharia Química, Recife, PE, Brazil

### ARTICLE INFO

#### Article history:

Received 23 November 2011

Received in revised form 25 January 2012

Accepted 30 January 2012

Available online 4 February 2012

#### Keywords:

Hydrated ethyl alcohol fuel

Infrared spectrometry

Supervised pattern recognition methods

Partial least squares – discriminant analysis

Linear discriminant analysis

Wavenumber selection

### ABSTRACT

This paper proposes an analytical method to detect adulteration of hydrated ethyl alcohol fuel based on near infrared (NIR) and middle infrared (MIR) spectroscopies associated with supervised pattern recognition methods. For this purpose, linear discriminant analysis (LDA) was employed to build a classification model on the basis of a reduced subset of wavenumbers. For variable selection, three techniques are considered, namely the successive projection algorithm (SPA), the genetic algorithm (GA) and a stepwise formulation (SW). For comparison, models based on partial least squares discriminant analysis (PLS-DA) were also employed using full-spectrum. The method was validated in a case study involving the classification of 181 hydrated ethyl alcohol fuel samples, which were divided into three different classes: (1) authentic samples; (2) samples adulterated with water and (3) samples contaminated with methanol. LDA/GA and PLS-DA models were found to be the best methods for classifying the spectral data obtained in NIR region, which achieved a correct prediction rate of 100% in the test set, while the LDA/SPA and LDA/SW were correctly classified at 84.4% and 97.8%, respectively. For MIR data, all models (PLS-DA and LDA coupled with the SW, SPA and GA) employed in this study correctly classified all samples in the test set.

© 2012 Elsevier B.V. All rights reserved.

### 1. Introduction

Ethanol is frequently used as raw material for beverages, cosmetics, pharmaceuticals and in the chemical industries. Due to the variety of renewable resources available in Brazil, ethanol is widely used as an automotive vehicle fuel [1,2]. This is usually found in an anhydrous form (blended with the gasoline) and as hydrated ethyl alcohol fuel (HEAF) [3].

According to the Brazilian National Agency for Petroleum, Natural Gas, and Biofuels (ANP) the HEAF has had, in recent years, one of the highest rates of nonconformity among all fuels monitored in Brazil [4]. The main form of adulteration in HEAF samples is the illegal addition of water to ethanol fuel or, in more serious cases, replacement of ethyl alcohol by methyl alcohol. According to Resolution ANP [5], the permitted maximum levels of water

and methanol in HEAF samples are 4.0% ( $\text{v v}^{-1}$ ) and 1.0% ( $\text{v v}^{-1}$ ), respectively.

Methanol and ethanol present similar physical-chemical properties, including: solubility in water, density values, look and smell [6]. The lower price of methanol in relation to ethanol and the similarity of these alcohols contribute to the ease of adulteration of HEAF with methanol in Brazil. Methanol is extremely toxic and can cause serious health problems such as headache, nausea, vomiting, blindness and even death. Therefore it is against the law for it to be used as fuel in Brazil [7].

The reference method to measure alcohol content in HEAF samples is based on the ABNT NBR 5992 standard [8]. This method uses a glass hydrometer to measure the alcohol content; but it is not possible to distinguish the type of alcohol (ethyl or methyl) in the HEAF samples. The Brazilian Standards Association (ABNT) has elaborated a draft standard (draft 34:007.01-006), now in the public enquiry stage, to determine the amount of ethanol and methanol concentration in ethanol fuel samples with a method based on gas chromatography.

Considering the possibility of HEAF adulteration with water or methanol, it is important to develop a fast, simple and low cost method able to detect whether the permitted maximum levels of

\* Corresponding author at: Universidade Federal da Paraíba, Departamento de Química – Laboratório de Automação e Instrumentação em Química Analítica/Quimiometria (LAQA), CEP 58051-970, João Pessoa, PB, Brazil.

Tel.: +55 83 3216 7438; fax: +55 83 3216 7438.

E-mail addresses: [marcio.quimica@gmail.com](mailto:marcio.quimica@gmail.com), [marciocoelho@quimica.ufpb.br](mailto:marciocoelho@quimica.ufpb.br) (M.J.C. Pontes).