



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

## **DISSERTAÇÃO DE MESTRADO**

**Uso da região espectral de sobretons para  
determinação do teor de biodiesel e classificação de  
misturas diesel/biodiesel adulteradas com óleo vegetal**

**Fernanda Vera Cruz de Vasconcelos**

João Pessoa – PB - Brasil  
Abril/2011



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

## DISSERTAÇÃO DE MESTRADO

**Uso da região espectral de sobretons para  
determinação do teor de biodiesel e classificação de  
misturas diesel/biodiesel adulteradas com óleo vegetal**

**Fernanda Vera Cruz de Vasconcelos\***

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal da Paraíba como parte dos requisitos para obtenção do título de Mestre em Química, área de concentração Química Analítica.

Orientadora: **Profa. Dra. Claudete Fernandes Pereira**

2ª Orientador: **Prof. Dr. Márcio José Coelho de Pontes**

\* Bolsista (Capes)

João Pessoa – PB - Brasil  
Abril/2011

V331u Vasconcelos, Fernanda Vera Cruz de.

Uso da região espectral de sobretons para determinação do teor de biodiesel e classificação de misturas diesel/biodiesel adulteradas com óleo vegetal / Fernanda Vera Cruz de Vasconcelos.- João Pessoa, 2011.

93f. : il.

Orientadora: Claudete Fernandes Pereira

Co-orientador: Márcio José Coelho de Pontes

Dissertação (Mestrado) – UFPB/CCEN

1. Química. 2. Misturas diesel/biodiesel. 3. Espectroscopia NIR. 4. Calibração multivariada.

UFPB/BC

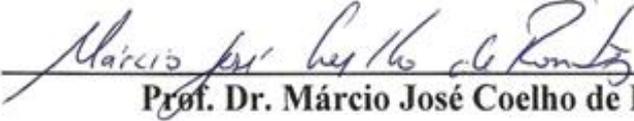
CDU: 54(043)

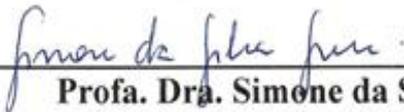
**Uso da Região Espectral de Sobretons para  
Determinação do Teor de Biodiesel e  
Classificação de Misturas Diesel/Biodiesel  
Adulteradas com Óleo Vegetal.**



Aprovada pela banca examinadora:

  
\_\_\_\_\_  
**Profa. Dra. Claudete Fernandes Pereira**  
Orientadora/Presidente

  
\_\_\_\_\_  
**Prof. Dr. Márcio José Coelho de Pontes**  
2º Orientador

  
\_\_\_\_\_  
**Profa. Dra. Simone da Silva Simões**  
Examinadora

  
\_\_\_\_\_  
**Profa. Dra. Teresa Cristina Bezerra Saldanha**  
Examinadora

***"Só existem dois dias no ano em que nada pode ser feito. Um se chama ontem e o outro se chama amanhã, portanto hoje é o dia certo para amar, acreditar, fazer e principalmente viver."***

Dalai Lama.

Aos meus pais e aos meus irmãos, pessoas  
as quais tenho maior carinho e admiração.

A todas as pessoas que mesmo de longe me  
deram força para seguir nessa jornada,

Dedico.

## AGRADECIMENTOS

- A **Deus**, por iluminar sempre meu caminho, sem Ele eu nada seria;
- Ao meu querido pai e amigo, **Adalberto**, pelo carinho, amor, compreensão e apoio nos momentos difíceis;
- A minha mãe, **Alice**, pelo apoio, compreensão, e pelos sábios e valiosos conselhos, aos meus irmãos (**Flávia**, **Laís** e **Danillo**) que sempre torceram por mim;
- A Profa. Dra. **Claudete Fernandes Pereira** pela confiança, amizade, orientação, ensinamentos e dedicação demonstrados durante a realização desse trabalho;
- A Profa. Dra. **Maria Fernanda Pimentel** pela oportunidade de aprendizagem, incentivo, sugestões e ensinamentos desde a época da Iniciação Científica;
- Ao Prof. Dr. **Márcio Coelho Pontes** pela colaboração, disponibilidade, ensinamentos e sugestões acadêmicas;
- Aos professores do Departamento de Química da UFPB, em especial aos profs. Drs. **Mário César Ugulino** e **Edvan Cirino** pela boa acolhida e ensinamentos;
- Aos coordenadores do LAC-UFPE, em especial Prof. Dr. **Luiz Stragevitch**, por disponibilizar sempre recursos para o bom andamento das pesquisas;
- Aos alunos de Iniciação Científica **Paulo Fernandes** e **Alinne Giliane** pela importante colaboração na fase experimental desse trabalho, sem vocês não teria conseguido;
- Aos colaboradores do LAC-UFPE, em especial **Ismália** e **Fabricia** pela ajuda nas análises cromatográficas;

- Aos amigos do Departamento de Engenharia Química da UFPE: **Viviane**, **Josemar** e **Jonathan** pela amizade, companheirismo e incentivo desde a época da graduação;
- As amigas **Carol** e **Alianda**, pela força, incentivo, paciência e participação nos momentos importantes;
- A **Rosimeri**, **Rosângela** e **Anabel** pela amizade, companheirismo e boa acolhida em João Pessoa;
- A todos os amigos que conquistei durante esse período no LAQA-UFPB, em especial a: **Paulo Henrique**, **Sófacles Figueredo**, **Marcelo Batista** e **Heberty Viana**, pela paciência, força, incentivo, sugestões acadêmicas e agradáveis gargalhadas;
- A **TRANSPETRO** (Petrobrás Transporte S.A), por fornecer às amostras de óleo diesel utilizadas na realização deste trabalho;
- Aos colegas de trabalho do LAFEPE: **Antônio Diógenes**, **Elcy Araújo**, **Tereza Raquel**, **Thamirys** e **Luciana** pelo apoio e incentivo;
- As empresas **BRASBIODIESEL** e **BRACOL HOLDING**, por fornecerem as amostras de biodiesel certificadas (B100);
- A Capes pela bolsa concedida;

Enfim, a todos aqueles que direta ou indiretamente contribuíram para realização desse trabalho.

# SUMÁRIO

<b>LISTA DE FIGURAS .....</b>	<b>iii</b>
<b>LISTA DE TABELAS .....</b>	<b>vi</b>
<b>LISTA DE ABREVIATURAS .....</b>	<b>viii</b>
<b>RESUMO.....</b>	<b>x</b>
<b>ABSTRACT.....</b>	<b>xi</b>
<b>1. INTRODUÇÃO .....</b>	<b>2</b>
<b>2. OBJETIVOS.....</b>	<b>5</b>
2.1 Objetivo Geral .....	5
2.2 Objetivos Específicos .....	5
<b>3. REVISÃO BIBLIOGRÁFICA .....</b>	<b>7</b>
3.1. O Biodiesel .....	7
3.1.1 Processo de produção do biodiesel .....	8
3.2 Misturas Diesel/Biodiesel .....	11
3.2.1 Controle de qualidade das Misturas diesel/biodiesel .....	11
3.3 Espectroscopia na região do infravermelho .....	14
3.4 Quimiometria .....	19
3.4.1 Pré-processamento.....	19
3.4.2 Reconhecimento de padrões .....	22
3.4.3 Calibração multivariada .....	24
3.4.4 Seleção de variáveis.....	28
3.4.5 Seleção de amostras .....	31
3.5 Métodos analíticos baseados na espectroscopia no infravermelho para análise de misturas diesel/biodiesel .....	32
<b>4. MATERIAIS E MÉTODOS.....</b>	<b>37</b>
4.1 Amostras.....	37
4.2 Síntese do biodiesel .....	37

4.3	Parâmetros de qualidade do biodiesel .....	37
4.4	Preparação das misturas .....	39
4.5	Aquisição dos espectros NIR .....	40
4.6	Tratamento quimiométrico dos dados .....	41
4.6.1	Pré-processamento.....	41
4.6.2	Reconhecimento de padrões .....	41
4.6.3	Calibração Multivariada .....	41
<b>5.</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>46</b>
5.1	Espectros NIR .....	46
5.2	Reconhecimento de padrões .....	47
5.2.1	Análise por Componente Principais .....	48
5.2.2	Modelos LDA/SPA .....	49
5.3	Calibração multivariada.....	55
5.3.1	Caminho óptico de 1,0 mm (9000-4000 cm <sup>-1</sup> ).....	55
5.3.2	Caminho óptico de 10 mm (9000-6300 cm <sup>-1</sup> ).....	60
5.3.3	Caminho óptico de 50 mm (7500-6300 cm <sup>-1</sup> ).....	66
5.3.4	Comparação dos resultados obtidos para os caminhos ópticos de 1,0, 10 e 50 mm.....	72
<b>6</b>	<b>CONCLUSÃO .....</b>	<b>75</b>
<b>7</b>	<b>PROPOSTAS FUTURAS .....</b>	<b>77</b>
<b>8</b>	<b>REFERÊNCIAS BIBLIOGRÁFICAS.....</b>	<b>79</b>
<b>ANEXO</b>	<b>.....</b>	<b>93</b>

## LISTA DE FIGURAS

<b>Figura 3.1</b> – Reação de transesterificação de triacilglicerídeos.....	09
<b>Figura 3.2</b> – Evolução dos índices de não conformidade dos principais combustíveis comercializados no território brasileiro.....	12
<b>Figura 3.3</b> – Distribuição das não conformidades por item especificado para o diesel.....	13
<b>Figura 3.4</b> – Tipos de vibrações moleculares. Legenda: + = vibração fora do plano; - = vibração para dentro do plano.....	16
<b>Figura 3.5</b> – Variação de energia potencial em função da distância internuclear para: (a) oscilador harmônico, (b) oscilador anarmônico. $d_e$ = distância de equilíbrio.....	17
<b>Figura 3.6</b> – Representação das matrizes de dados ( <b>X</b> e <b>y</b> ) na calibração multivariada.....	25
<b>Figura 4.1</b> – Planta piloto utilizada para produção de biodiesel de girassol, localizada no Laboratório de Combustíveis (LAC) da UFPE.....	38
<b>Figura 4.2</b> – Espectrofotômetro Infravermelho FT-IR (FTLA 2000-160) da Bomen® .....	41
<b>Figura 5.1</b> – Espectros NIR das 132 amostras registrados na faixa de $14000-4000\text{ cm}^{-1}$ para os caminhos ópticos: (a) 1,0 mm; (2) 10 mm; (c) 50 mm. As regiões espectrais utilizadas no tratamento quimiométrico para os diferentes caminhos ópticos estão destacadas em vermelho.....	46

- Figura 5.2** – Gráfico dos escores da PC2 × PC1 para as 150 amostras do conjunto de dados nos caminhos ópticos: (a) 1,0 mm, (b) 10 mm e (c) 50 mm. (\*: D, ●: DO, ▲: DBO e ■: DB). O percentual da variância explicada por cada PC está indicado entre parênteses..... 49
- Figura 5.3** – Variáveis espectrais selecionadas pelo SPA na construção dos modelos de classificação LDA utilizando-se o caminho óptico de 1,0 mm..... 50
- Figura 5.4** – Variáveis espectrais selecionadas pelo SPA na construção dos modelos de classificação LDA utilizando-se os caminhos ópticos de (a) 10 mm e (b) 50 mm..... 51
- Figura 5.5** – Gráfico dos escores da FD2×FD1 para as 150 amostras do conjunto de dados nos caminhos ópticos: (a) 1,0 mm, (b) 10 mm e (c) 50 mm. (\*: D, ●: DO, ▲: DBO e ■: DB)..... 52
- Figura 5.6** – Espectros NIR das 128 amostras: (a) originais e (b) pré-processados por correção de linha de base e polinômio de SG com janela de 11 pontos (LB/SSG)..... 57
- Figura 5.7** – Gráfico dos valores previstos para teor de biodiesel versus valores de referência para os modelos: (a) PLS; (b) PLS/JK e (c) MLR/SPA utilizando o caminho óptico de 1,0 mm (●: amostras de calibração e ●: validação externa)..... 59
- Figura 5.8** – Variáveis espectrais selecionadas pelas estratégias: (a) PLS/JK e (b) MLR/SPA para determinação do teor de biodiesel utilizando o caminho óptico de 1,0 mm..... 60
- Figura 5.9** – Espectros NIR das 129 amostras: (a) originais e (b) pré-processados por correção de linha de base com suavização por média móvel com janela de 5 pontos (LB/SMM)..... 63

- Figura 5.10** – Gráfico dos valores previstos para teor de biodiesel versus valores de referência para os modelos: (a) PLS; (b) PLS/JK e (c) MLR/SPA utilizando o caminho óptico de 10 mm. (●: amostras de calibração e ●: validação externa)..... 65
- Figura 5.11** – Variáveis espectrais selecionadas pelas estratégias: (a) PLS/JK e (b) MLR/SPA para determinação do teor de biodiesel utilizando o caminho óptico de 10 mm..... 66
- Figura 5.12** – Espectros NIR das 129 amostras: (a) originais; (b) e pré-processados por correção de linha de base com suavização por média móvel com janela de 5pts (LB / SMM)..... 69
- Figura 5.13** – Gráfico dos valores previstos para teor de biodiesel versus valores de referência para os modelos: (a) PLS; (b) PLS/JK e (c) MLR/SPA utilizando o caminho óptico de 50 mm. (●: amostras de calibração e ●: validação externa)..... 71
- Figura 5.14** – Variáveis espectrais selecionadas pelas estratégias: (a) PLS/JK e (b) MLR/SPA para determinação do teor de biodiesel utilizando o caminho óptico de 50 mm..... 72

## LISTA DE TABELAS

<b>Tabela 3.1</b> – Regiões espectrais do infravermelho.....	14
<b>Tabela 4.1</b> – Número de amostras dos conjuntos de treinamento, validação e teste utilizados para construção do modelo LDA/SPA.....	42
<b>Tabela 5.1</b> – Erros de classificação obtidos pelos modelos LDA/SPA para as amostras do conjunto de teste utilizando-se os caminhos ópticos de 1,0, 10 e 50 mm.....	53
<b>Tabela 5.2</b> – Resultados obtidos para os modelos de calibração multivariada PLS, PLS-JK e MLR/SPA (caminho óptico: 1,0 mm) utilizando-se diferentes estratégias de pré-processamento dos espectros.....	55
<b>Tabela 5.3</b> – Resultados para o teste $F$ (nível de 95% de confiança) obtidos da comparação das diferentes técnicas de pré-processamento empregadas na construção dos modelos (PLS, PLS/JK e MLR/SPA) para o caminho óptico de 1,0 mm.....	56
<b>Tabela 5.4</b> – Resultados do teste $F$ (nível de 95% de confiança) comparando-se os modelos PLS, PLS/JK e MLR/SPA construídos utilizando-se os espectros NIR (caminho óptico = 1,0 mm) pré-processados com correção de linha de base e polinômio móvel com filtro SG (janela de 11 pontos).....	58
<b>Tabela 5.5</b> – Resultados obtidos para os modelos de calibração multivariada PLS, PLS-JK e MLR/SPA (caminho óptico: 10 mm) utilizando-se diferentes estratégias de pré-processamento dos espectros.....	61

<b>Tabela 5.6</b> – Resultados para o teste $F$ (nível de 95% de confiança) obtidos da comparação das diferentes técnicas de pré-processamento empregadas na construção dos modelos (PLS, PLS/JK e MLR/SPA) para o caminho óptico de 10 mm.....	62
<b>Tabela 5.7</b> – Resultados do teste $F$ (nível de 95% de confiança) comparando-se os modelos PLS, PLS/JK e MLR/SPA construídos utilizando-se os espectros NIR (caminho óptico = 10 mm) pré-processados com correção de linha de base e suavização por filtro média móvel (janela de 5 pontos).....	64
<b>Tabela 5.8</b> – Resultados obtidos para os modelos de calibração multivariada PLS, PLS-JK e MLR/SPA (caminho óptico: 50 mm) utilizando-se diferentes estratégias de pré-processamento dos espectros.....	67
<b>Tabela 5.9</b> – Resultados para o teste $F$ (nível de 95% de confiança) obtidos da comparação das diferentes técnicas de pré-processamento empregadas na construção dos modelos (PLS, PLS/JK e MLR/SPA) para o caminho óptico de 50 mm.....	68
<b>Tabela 5.10</b> – Resultados do teste $F$ (nível de 95% de confiança) comparando-se os modelos PLS, PLS/JK e MLR/SPA construídos utilizando-se os espectros NIR (caminho óptico = 50 mm) pré-processados com correção de linha de base e suavização por filtro média móvel (janela de 5 pontos).....	70
<b>Tabela 5.11</b> – Resultados do teste $F$ (nível de 95% de confiança) para os modelos PLS, PLS/JK e MLR/SPA construídos utilizando-se os espectros NIR para os caminhos ópticos de 1,0; 10 e 50 mm .....	73

## LISTA DE ABREVIATURAS

ABNT	Associação Brasileira de Normas Técnicas
AG	Algoritmo Genético
ANN	Redes Neurais Artificiais
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
ATR	Reflectância Total Atenuada
BS	<i>British Standard</i>
CLS	Método Clássico dos Mínimos Quadrados
FTIR	Espectrômetro com Transformada de Fourier
HCA	Análise de Agrupamentos Hierárquicos
H-PLS	Método PLS-Hierárquico
Serial-PLS	Regressão por Mínimos Quadrados Parciais Serial
IR	Infravermelho
iPLS	Regressão pelos Mínimos Quadrados Parciais por Intervalo
JK	Algoritmo Jack Knife
KS	Algoritmo Kennard-Stone
KNN	Método do K-ésimo Vizinho mais Próximo
LAC	Laboratório de Combustíveis
LB/SMM	Correção de linha de base e suavização com filtro média móvel
LB/SSG	Correção de linha de base e suavização por polinômio móvel e filtro Savitzky-Golay
LDA	Análise Discriminante Linear
MIR	Infravermelho Médio
MLR	Regressão Linear Múltipla

MLR/SPA	Regressão Linear Múltipla com seleção de variáveis pelo Algoritmo das Projeções sucessivas
MSC	Correção Multiplicativa do Sinal
NIR	Infravermelho Próximo
PC	Componentes Principais
PCA	Análise por Componentes Principais
PCR	Regressão por Componentes Principais
PLS	Regressão por Mínimos Quadrados Parciais
PLS/JK	Regressão por Mínimos Quadrados Parciais com coeficientes de regressão selecionados pelo algoritmo Jack-Knife
PMQC	Programa Nacional de Monitoramento de Qualidade de Combustíveis
PNPB	Programa Nacional de Produção e Uso do Biodiesel
RMSECV	Raiz quadrada do erro médio quadrático de validação cruzada
RMSEP	Raiz quadrada do erro médio quadrático de previsão
SBQ	Superintendência de Biocombustíveis e de Qualidade do Produto
SIMCA	Modelagem Independente e Flexível por Analogia de Classe
SNV	Variação Normal Padrão
SG	<i>Savitzky-Golay</i>
SPA	Algoritmo das Projeções Sucessivas
SPXY	Partição de amostra baseado na distância de <b>X-y</b>

## RESUMO

**Título:** Uso da região espectral de sobretons para determinação do teor de biodiesel e classificação de misturas diesel/biodiesel adulteradas com óleo vegetal

**Autor:** Fernanda Vera Cruz de Vasconcelos

Neste trabalho foi avaliado o uso da região espectral de sobretons (NIR) para classificação e determinação do teor de biodiesel em misturas diesel/biodiesel adulteradas com óleo vegetal (*in natura*). Para tanto, foram utilizados espectros NIR registrados em diferentes caminhos ópticos: 1,0 (9000-4000  $\text{cm}^{-1}$ ), 10 (9000-6300  $\text{cm}^{-1}$ ) e 50 mm (7500-6300  $\text{cm}^{-1}$ ). A Análise de Componentes Principais (PCA) e a Análise Discriminante Linear com seleção de variáveis pelo Algoritmo das Projeções Sucessivas (LDA/SPA) foram utilizadas na identificação de adulteração de misturas diesel/biodiesel com óleo vegetal. Os modelos de calibração multivariada (Regressão por Mínimos Quadrados Parciais - PLS, Regressão por Mínimos Quadrados Parciais com coeficientes de regressão selecionados pelo algoritmo Jack-Knife - PLS/JK e Regressão Linear Múltipla com seleção de variáveis pelo Algoritmo das Projeções Sucessivas - MLR/SPA) foram desenvolvidos para determinação do teor de biodiesel. Os resultados obtidos demonstraram que a região espectral de sobretons aliada à estratégia de classificação LDA/SPA pode ser utilizada em estudos preliminares para detecção de adulteração de misturas diesel/biodiesel por óleos vegetais *in natura*, visto que apresentou bons resultados de classificação, com índice de acerto de 86% para os caminhos ópticos de 10 mm e 50 mm. Além disso, foi demonstrada a aplicabilidade da região de sobretons associada à estratégia de regressão MLR/SPA para determinação do teor de biodiesel em misturas diesel/biodiesel na presença de óleos vegetais *in natura*, na faixa de 0,0 a 10,0% v/v. Tal estratégia é simples, rápida e utiliza poucas variáveis espectrais. Neste contexto, a região espectral de sobretons pode ser útil na construção de instrumentos de baixo custo para controle de qualidade de misturas diesel/biodiesel, considerando-se o menor custo dos componentes ópticos para essa região espectral.

Palavras-chave: misturas diesel/biodiesel; espectroscopia NIR; calibração multivariada; classificação.

## ABSTRACT

**Title:** Using overtone regions to determine biodiesel content and classification of diesel/biodiesel blends adulterated with vegetable oils

**Author:** Fernanda Vera Cruz de Vasconcelos

This work evaluates the use of overtone regions (NIR) for the classification and determination of biodiesel content in diesel/biodiesel blends adulterated with vegetable oil. For this purpose, NIR spectra were obtained using three different optical paths: 1.0 mm (9000-4000  $\text{cm}^{-1}$ ); 10 mm (9000-6300  $\text{cm}^{-1}$ ) e 50 mm (7500-6300  $\text{cm}^{-1}$ ). Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) with the Successive Projections Algorithm (SPA) were employed to build screening models to identify adulteration of diesel/biodiesel blends with vegetable oils. The multivariate calibration models (Partial Least Square - PLS, Partial Least Square with significant regression coefficients selected by Jack-Knife algorithm - PLS/JK and Multiple Regression Linear with variable selection by Successive Projections Algorithm - MLR/SPA) were developed to determine the biodiesel content. The results showed that the overtone regions with the classification strategy LDA/SPA can be used in preliminary studies to detect adulteration of diesel/biodiesel blends with vegetable oil. This strategy showed positive results for classifying, with a correct classification rate of 86% for the optical paths of 10 mm and 50 mm. In addition, the work demonstrated the potential of overtone regions with MLR/SPA regression strategy to determine biodiesel content in diesel/biodiesel blends, in the range from 0.0% to 10.0%v/v, considering the possible presence of raw oil as a contaminant. This strategy is simple, fast and uses a fewer number of spectral variables. Considering this, the overtone regions can be usefull to develop low cost instruments for the quality control of diesel/biodiesel blends, considering the lower cost of optical components for this spectral region.

**Keywords:** diesel/biodiesel blends; near infrared spectrometry; multivariate calibration; classification.

# **INTRODUÇÃO**

## **Capítulo 1**

## 1 INTRODUÇÃO

No Brasil, o biodiesel foi inserido na matriz energética em 2005 (Lei nº 11.097/2005) de acordo com o Programa Nacional de Produção e Uso do Biodiesel (PNPB) do Governo Federal. Esse programa se apoiou na crescente demanda por combustíveis de fontes renováveis e no potencial brasileiro para atender parte expressiva dessas necessidades.

Segundo essa lei, a partir de janeiro de 2008 tornou-se obrigatória, em todo território nacional, a mistura B2, ou seja, 2% de biodiesel e 98% de diesel, em volume. Em janeiro de 2010, essa obrigatoriedade passou para 5% (B5). A lei prevê ainda o uso de misturas com teores mais elevados de biodiesel e até mesmo o biodiesel puro mediante autorização da ANP (Agência Nacional do Petróleo, Gás Natural e Bicompostíveis) (B100) (<http://www.anp.gov.br/> acesso 13/11/2010).

Devido à crescente utilização do biodiesel, diferentes métodos analíticos têm sido propostos para determinar a concentração desse combustível em misturas com diesel (MONTEIRO et al, 2009; PRATES, et al, 2010). Além disso, sob o aspecto do controle de qualidade, o desenvolvimento de tais métodos é importante, visto que, óleos vegetais *in natura* podem estar presentes nas misturas devido à: (1) conversão incompleta a éster durante a reação ou (2) adição ilegal desse material, o qual é mais barato que o biodiesel (KNOTHE et al, 2005).

Dentre os métodos analíticos propostos, as espectroscopias nas regiões do infravermelho médio (MIR) e próximo (NIR), utilizando métodos de calibração multivariada, surgem como boas alternativas para prever os principais parâmetros de qualidade de combustíveis.

Em geral, esses métodos apresentam vantagens intrínsecas à espectroscopia na região do infravermelho, que é uma técnica analítica rápida, não destrutiva, não invasiva e que permite a análise de vários parâmetros, com pouco ou nenhum preparo da amostra (PASQUINI et al, 2003).

---

De acordo com a literatura, os resultados obtidos utilizando as espectroscopias MIR/NIR associadas às ferramentas quimiométricas para determinação do teor de biodiesel (PIMENTEL et al, 2006) e óleo vegetal (OLIVEIRA et al, 2007) no diesel apresentaram-se bastante promissores. Entretanto, pouca atenção tem sido dada à região espectral de sobretons, a qual pode ser útil na construção de instrumentos de baixo custo para controle de qualidade de misturas diesel/biodiesel.

Nesse contexto, o presente trabalho propõe o desenvolvimento de um método analítico para classificação e determinação do teor de biodiesel em misturas diesel/biodiesel adulteradas com óleo vegetal (*in natura*) utilizando a região espectral de sobretons (NIR).

# **OBJETIVOS**

## **Capítulo 2**

---

## 2 OBJETIVOS

### 2.1 Objetivo Geral

- Avaliar o uso da região espectral de sobretons do NIR para determinação do teor de biodiesel e classificação de misturas diesel/biodiesel adulteradas com óleo vegetal *in natura*.

### 2.2 Objetivos específicos

- Avaliar o comportamento espectral na região do NIR de misturas diesel/biodiesel adulteradas com óleo vegetal utilizando diferentes caminhos ópticos (1,0, 10 e 50 mm);
- Avaliar o uso da espectroscopia NIR associada à Análise por Componentes Principais (PCA) e Análise Discriminante Linear com seleção de variáveis pelo Algoritmo das Projeções Sucessivas (LDA/SPA) na identificação de adulteração de misturas diesel/biodiesel;
- Construir modelos de calibração multivariada utilizando diferentes técnicas de regressão para determinação do teor de biodiesel em misturas diesel/biodiesel adulteradas com óleo vegetal, empregando diferentes caminhos ópticos (1,0, 10 e 50 mm);
- Avaliar o desempenho dos modelos construídos utilizando-se as técnicas de Regressão por Mínimos Quadrados Parciais (PLS) e Regressão Linear Múltipla e dois métodos de seleção de variáveis: (1) coeficientes de regressão mais significativos selecionados pelo algoritmo Jack-Knife (PLS/JK) e (2) Algoritmo das Projeções Sucessivas (MLR/SPA).

# **REVISÃO BIBLIOGRÁFICA**

## **Capítulo 3**

---

## 3 REVISÃO BIBLIOGRÁFICA

### 3.1 O Biodiesel

De acordo com a Resolução N° 7 de 2008 da ANP, o biodiesel é definido como um combustível para uso em motores a combustão interna com ignição por compressão, renovável e biodegradável, composto de alquil ésteres de ácidos graxos de cadeia longa derivados de óleos vegetais ou de gorduras animais (RESOLUÇÃO ANP N° 7, 2008).

Historicamente, o uso de óleos vegetais como combustível teve início com o desenvolvimento dos motores diesel no final do século XIX por Rudolph Diesel (PINTO et al, 2005). Todavia, no início do século XX, o diesel passou a ser o combustível utilizado para esse tipo de motor devido ao seu baixo custo e por razões técnicas.

Durante as duas crises do petróleo ocorridas nas décadas de 30 e 70, foi retomada a idéia do uso de óleos vegetais como combustível; entretanto, a mesma não obteve sucesso (SUAREZ et al, 2007). Além disso, houve uma maior preocupação em atender à crescente demanda energética mundial, o que, conseqüentemente, levou ao crescimento da consciência ambiental devido às mudanças climáticas associadas à liberação de gases da queima de combustíveis fósseis. Todos esses fatores impulsionaram a busca por combustíveis de fontes renováveis, os quais voltaram a ser fortemente discutidos e pesquisados. (PIMENTEL et al, 2006; LÔBO et al, 2009).

O uso direto de óleos vegetais em motores a combustão interna, apesar de ser favorável do ponto de vista energético, apresenta diversos problemas, tais como: combustão incompleta, formação de depósitos de carbono nos sistemas de injeção e contaminação do óleo lubrificante. As causas desses problemas podem ser atribuídas à polimerização dos triacilglicerídeos, à baixa volatilidade e, principalmente, à alta viscosidade dos óleos (KNOTHE et al, 2005).

Para viabilizar o uso dos óleos vegetais como combustíveis várias abordagens têm sido investigadas, dentre as quais destacam-se: craqueamento, transesterificação, microemulsificação e esterificação (SUAREZ et al, 2009). O processo mais comum é a transesterificação que leva à obtenção de alquil ésteres de óleo vegetal ou de gordura animal, hoje conhecido como biodiesel, com propriedades físico-químicas semelhantes às do diesel mineral podendo ser utilizado puro ou em misturas, exigindo pouca ou nenhuma alteração do motor (SCHUCHARDT et al, 2007; KNOTHE, 2006).

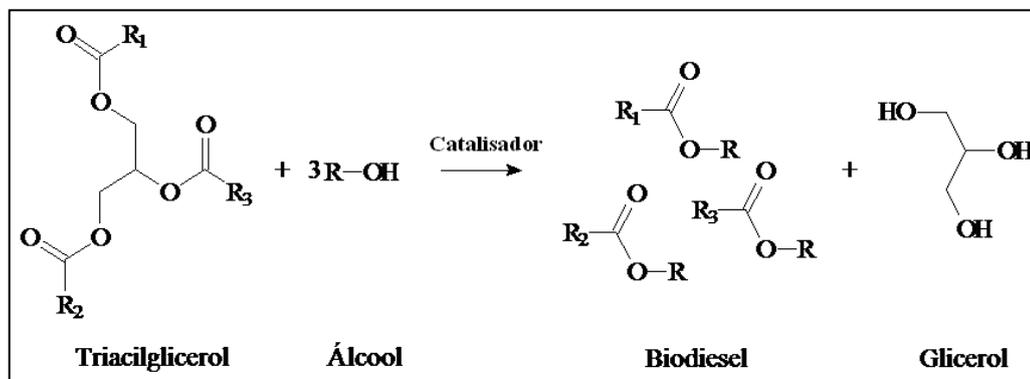
O biodiesel apresenta diversas vantagens em relação ao diesel é biodegradável, praticamente isento de enxofre, e sua combustão é mais eficiente que o diesel. Diversos estudos têm indicado que o uso desse combustível reduz a emissão de gases associados ao efeito estufa, tais como CO<sub>2</sub> e CO. Além da redução da emissão de materiais particulados e óxidos de enxofre responsável pela chuva ácida (FANGRUI et al, 1999; SUAREZ et al, 2007). Apesar dessas vantagens, foi demonstrado que há um aumento da emissão de NO<sub>x</sub> relacionado ao uso do biodiesel (BASHA et al, 2009).

### **3.2.1 Processo de produção do biodiesel**

A transesterificação ou alcoólise é o processo químico considerado mais viável para a obtenção do biodiesel. Nessa reação os triacilglicerídeos provenientes de óleos vegetais ou gorduras animais reagem com um álcool de cadeia curta, na presença de um catalisador, produzindo o biodiesel e o glicerol (**Figura 3.1**). O processo global é composto por três etapas reversíveis e consecutivas, nas quais diacilglicerídeos e monoacilglicerídeos são formados como intermediários (GERIS et al, 2007; TAPANES et al, 2008).

Diferentes fatores podem afetar a reação de transesterificação, tais como: o tipo de catalisador (básico, ácido, enzimático ou heterogêneo), a razão molar álcool/óleo, o tipo de álcool (metanol ou etanol), a

temperatura da reação e a velocidade de agitação (MEHER et al, 2006), além da qualidade da matéria-prima utilizada para produção do biodiesel.



**Figura 3.1** – Reação de transesterificação de triacilglicerídeos (LÔBO et al, 2009).

Nos últimos anos, várias pesquisas têm sido desenvolvidas visando a otimização desses parâmetros a fim de aumentar o rendimento da reação e facilitar o processo de purificação do biodiesel (MARCHETTI et al, 2007; DEMIRBAS, 2008; LEUNG et al, 2010). Dentre os processos estudados, o que vem sendo mais empregado em escala industrial é a transesterificação alcalina homogênea, utilizando metanol ou etanol como agente transesterificante.

A rota metílica vem sendo amplamente utilizada, pois o metanol apresenta maior reatividade que o etanol, o que implica em menor tempo e temperatura de reação. O etanol, porém, é consideravelmente menos tóxico e pode ser obtido a partir de fontes renováveis. Além disso, produz um biodiesel com maior índice de cetano. Uma desvantagem do etanol é que promove uma maior dispersão da glicerina no biodiesel dificultando a separação das fases, o que, conseqüentemente, requer uma maior razão molar álcool/óleo para aumentar o rendimento da reação (BOUAID et al, 2007; ENCINAR et al, 2007).

Com relação ao tipo de catalisador básico, os alcóxidos são mais eficientes e apresentam alto rendimento. Os hidróxidos de sódio e potássio, embora menos reativos que seus respectivos alcóxidos, promovem rendimentos satisfatórios e apresentam menor custo. Por esse

---

motivo, eles têm sido mais empregados na indústria (LÔBO et al, 2009; SUAREZ et al, 2009).

O biodiesel pode ser obtido a partir de óleos vegetais extraídos de diferentes oleaginosas (soja, algodão, girassol, pinhão manso, óleos residuais, dentre outras), além de gordura animal (sebo bovino ou óleo de peixe). Todavia, a transesterificação básica homogênea requer matérias-primas com baixo percentual de ácidos graxos livres e que sejam isentas de água para evitar a formação de sabão (RAMADHAS et al, 2005).

Após o processo de transesterificação, a massa reacional é constituída de duas fases separáveis por decantação ou centrifugação. A fase menos densa é composta de uma mistura de ésteres metílicos ou etílicos, dependendo da natureza do álcool adotado, e dos resíduos de catalisador e do álcool empregado em excesso. Já a fase mais densa é composta de glicerina bruta, álcool e impurezas provenientes da matéria-prima.

A etapa de purificação dos ésteres consiste basicamente na separação da glicerina, evaporação do excesso de álcool e lavagem ácida para a neutralização de resíduos de catalisador, seguida pela etapa de secagem, obtendo como produto final o biodiesel (GERIS et al, 2007). O coproduto da reação, a glicerina, após ser devidamente purificado, poderá ser reaproveitado, por exemplo, em indústrias farmacêuticas e de cosméticos (MOTA et al, 2009).

Finalmente, para ser comercializado, o biodiesel deverá atender as especificações estabelecidas pelas agências regulamentadoras. No Brasil, a ANP é o órgão responsável por estabelecer as especificações e por fiscalizar as atividades relativas à produção, controle de qualidade, distribuição, revenda e comercialização do biodiesel e das misturas com diesel (<http://www.anp.gov.br/?pg=33410>, acesso em 09/11/2010). A Resolução ANP N° 7 de 2008 apresenta as especificações para o biodiesel comercializado em todo o território nacional.

## **3.2 Misturas Diesel/Biodiesel**

Em decorrência da crescente demanda por combustíveis de fontes renováveis e do potencial brasileiro em atender parte expressiva dessas necessidades, o Governo Federal lançou em 2004 o Programa Nacional de Produção e Uso do Biodiesel. Esse programa interministerial teve como objetivo a implantação, de forma sustentável, da cadeia de produção e uso do biodiesel, visando à geração de emprego e renda na agricultura familiar (<http://www.biodiesel.gov.br> acesso 09/11/2010).

Em 2005, com a publicação da Lei Federal Nº 11.097 foi autorizada a adição de 2% (em volume) de biodiesel ao diesel mineral, mistura conhecida como B2, que passou a ser obrigatória em 2008. Segundo essa lei, o percentual de biodiesel adicionado ao diesel aumentaria para 5% apenas em 2013. Entretanto, essa obrigatoriedade foi antecipada e, desde janeiro de 2010 a mistura B5 vem sendo comercializada em todo o território brasileiro (<http://www.anp.gov.br/> acesso 13/11/2010). Com a expansão da capacidade produtiva do biodiesel no Brasil, há uma tendência ao longo dos próximos anos da comercialização de misturas com teores cada vez mais elevados de biodiesel, mediante a autorização da ANP.

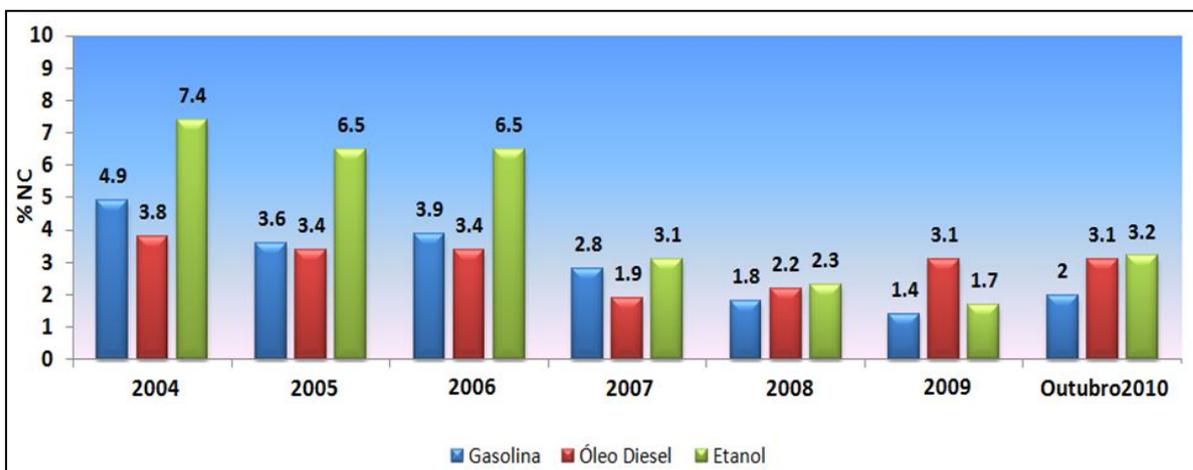
### **3.2.1 Controle de qualidade das Misturas diesel/biodiesel**

A ANP, dentro de suas atribuições, instituiu em 1999 o Programa Nacional de Monitoramento de Qualidade de Combustíveis (PMQC). Os principais objetivos do programa são: (1) realização de um levantamento periódico dos indicadores gerais da qualidade dos combustíveis (etanol, gasolina, diesel e óleos lubrificantes) comercializados nos postos revendedores do Brasil e (2) identificação dos focos de não conformidade, visando orientar e aperfeiçoar a atuação da ANP como órgão fiscalizador. Os resultados obtidos são uma ferramenta importante, pois fornecem subsídios para identificar e coibir fraudes e adulterações. Para a execução do PMQC, a ANP conta com uma rede composta por cerca de 20

instituições e centros de pesquisa em todo o país (<http://www.anp.gov.br/?pg=33410>, acesso em 09/11/2010).

No que se refere às misturas diesel/biodiesel comercializadas no Brasil, essas devem atender às especificações estabelecidas na Resolução ANP N° 42/2009 e qualquer desvio em relação a um dos itens especificados é considerado como uma não conformidade (NC). Por exemplo, a presença de óleo vegetal nas misturas diesel/biodiesel devido à conversão incompleta da reação de transesterificação, um maior ou menor teor de biodiesel contido nas misturas com óleo diesel, ou ainda devido à adição ilegal de alguma substância, como o óleo vegetal (in natura).

Através da Superintendência de Biocombustíveis e de Qualidade do Produto (SBQ) e do PMQC, a ANP lança mensalmente boletins avaliando a qualidade dos combustíveis líquidos automotivos brasileiros. O gráfico ilustrado na **Figura 3.2** mostra a evolução dos índices de não conformidade dos principais combustíveis comercializados no território brasileiro.

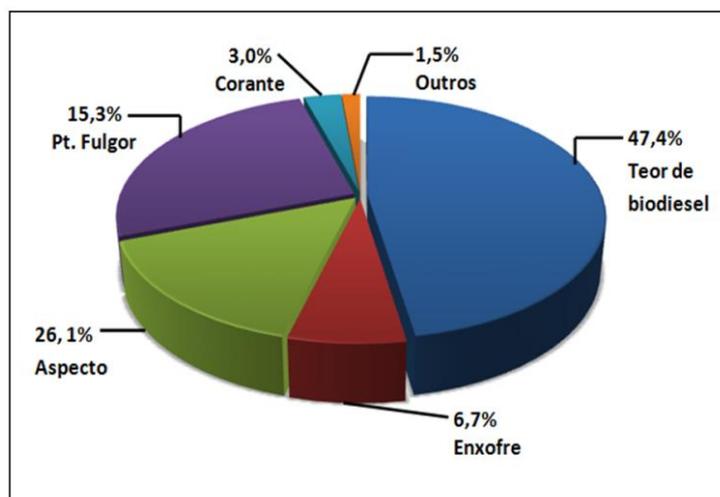


**Figura 3.2** – Evolução dos índices de não conformidade dos principais combustíveis comercializados no território brasileiro. (Fonte: ANP, dados 2004 a outubro de 2010).

Como pode ser observado na **Figura 3.2** o diesel apresentou os menores índices de não conformidade até o ano de 2007 em relação ao etanol e à gasolina. Entretanto, em 2009 chegou a atingir os maiores índices. E, em outubro de 2010 manteve o índice de 2009 (3,1% NC),

praticamente empatado com etanol que apresentou 3,2 % de não conformidades.

Analisando-se o total de não conformidades para o diesel no mês de outubro de 2010, em relação aos itens especificados pela Resolução ANP Nº 42 de 2009, o que apresentou o maior número de não conformidades foi o teor de biodiesel, correspondendo a 47,4% (**Figura 3.3**). De acordo com a legislação vigente, o percentual de biodiesel adicionado ao diesel é de 5% v/v, podendo ter uma faixa de tolerância de 4,5 a 5,5% v/v. Isto é, mais de 40% das misturas diesel/biodiesel estão fora dos limites especificados.



**Figura 3.3** – Distribuição das não conformidades por item especificado para o diesel (Fonte: ANP, dados de outubro de 2010).

O método analítico adotado pela ANP para determinação do teor de biodiesel em misturas com diesel está descrito na norma EN 14078:2009, o qual utiliza a espectroscopia na região do infravermelho médio (MIR). Esse método se baseia na construção de uma curva de calibração (univariada) e utiliza o pico de absorção associado à vibração de estiramento da carbonila (C=O) presente nos ésteres ( $\sim 1745 \text{ cm}^{-1}$ ) para determinar o teor de biodiesel presente nas misturas (BS EN 14078:2009).

Em 2008, a Associação Brasileira de Normas Técnicas (ABNT) publicou a norma NBR 15568 para determinação do teor de biodiesel utilizando a espectroscopia na região do MIR e modelos de calibração multivariada. O método se baseia na construção de modelos de regressão por mínimos quadrados parciais (PLS), utilizando as regiões do espectro características das ligações C-O ( $1100-1200\text{ cm}^{-1}$ ) e C=O ( $1735-1750\text{ cm}^{-1}$ ) presente nos ésteres (ABNT NBR 15568:2008).

Na literatura são encontrados diferentes procedimentos para análise de misturas diesel/biodiesel utilizando as espectroscopias nas regiões espectrais do MIR e do infravermelho próximo (NIR) aliadas às ferramentas quimiométricas.

A seguir, serão apresentados os princípios fundamentais da espectroscopia no infravermelho, com ênfase à região espectral NIR, além das ferramentas quimiométricas utilizadas no presente trabalho.

### 3.3 Espectroscopia na região do infravermelho

A região espectral no infravermelho compreende a faixa de radiação com números de onda no intervalo de aproximadamente  $12800$  a  $10\text{ cm}^{-1}$  (SKOOG, et al 2006). O espectro na região do infravermelho é usualmente dividido em infravermelho próximo (NIR), infravermelho médio (MIR) e infravermelho distante (FIR). Os limites aproximados para cada região espectral são mostrados na **Tabela 3.1**.

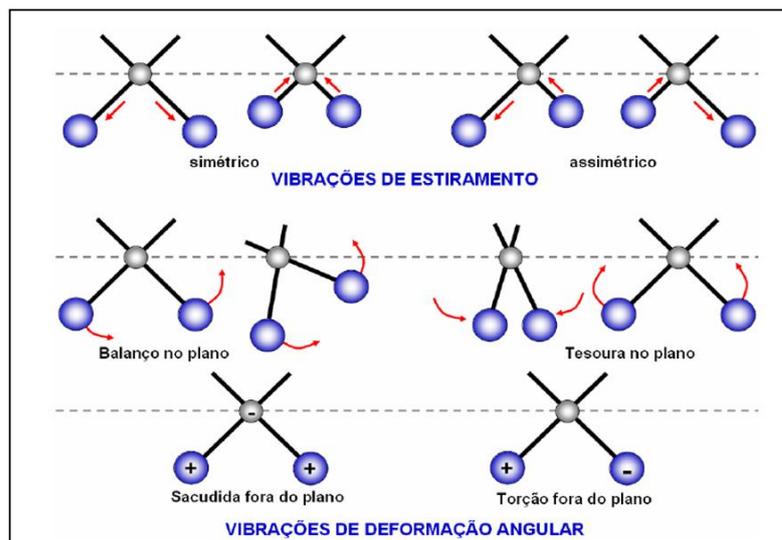
**Tabela 3.1** – Regiões espectrais do infravermelho (SKOOG, et al 2006).

Região	Número de onda ( $\text{cm}^{-1}$ )	Comprimento de onda (nm)
NIR	12.800 – 4000	780 – 2500
MIR	4000 – 400	2500 – 5000
FIR	200 – 10	5000 – 100.000

A radiação infravermelha não é energética o suficiente para causar transições eletrônicas como em outras regiões espectrais (ultravioleta, visível, raios X) e a absorção dessa radiação é muito restrita às espécies que possuem pequenas diferenças de energia entre os vários estados rotacionais e vibracionais. Para que haja absorção da radiação IV é necessário que a molécula apresente variações no momento de dipolo como consequência do movimento vibracional ou rotacional. Apenas nessas situações o campo elétrico alternado da radiação é capaz de interagir com a molécula e causar variações na amplitude de um de seus movimentos (SKOOG et al, 2006).

Moléculas homonucleares como  $O_2$ ,  $N_2$  ou  $Cl_2$  não sofrem variações efetivas no momento de dipolo durante a vibração e rotação e, conseqüentemente, não absorvem na região do infravermelho. Ao contrário, moléculas diatômicas heteronucleares (HCl, CO, entre outras) apresentam modos vibracionais de absorção ativos no infravermelho (SKOOG et al, 2006).

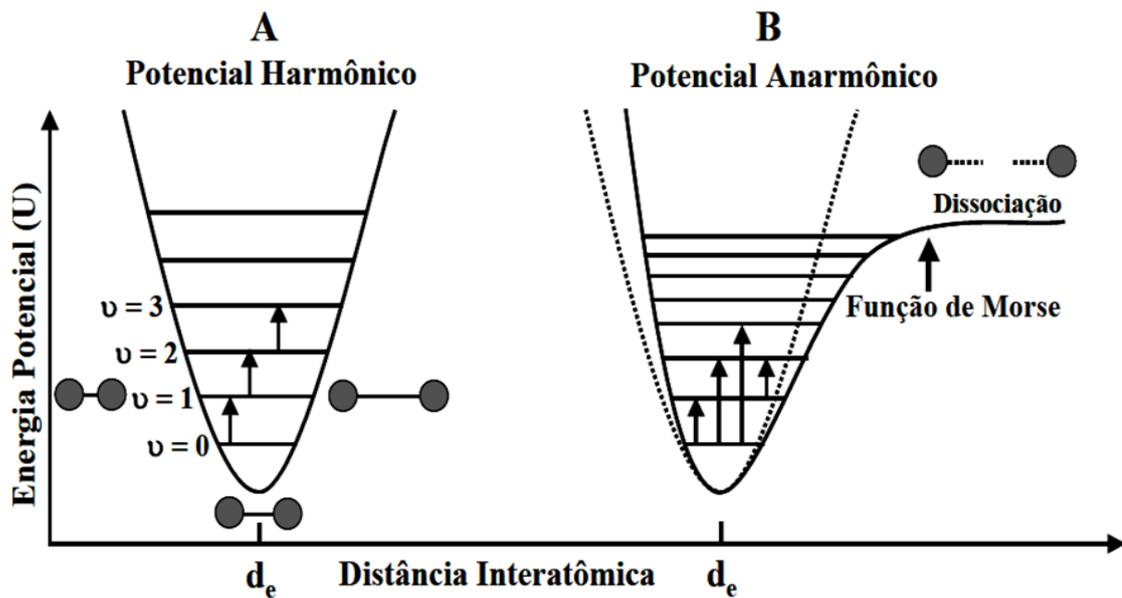
As posições relativas dos átomos em uma molécula não estão fixadas exatamente, pois as mesmas variam continuamente como consequência de inúmeros tipos de vibrações e rotações em torno das ligações da molécula (SKOOG et al, 2006). As vibrações podem ser classificadas nas categorias de estiramento e deformação. Uma vibração de estiramento envolve uma variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos. Já as vibrações de deformação são caracterizadas por variações no ângulo entre duas ligações e podem ser de quatro tipos: balanço no plano, tesoura no plano, sacudida fora do plano e torção fora do plano (**Figura 3.4**).



**Figura 3.4** – Tipos de vibrações moleculares. Legenda: + = vibração para fora do plano; - = vibração para dentro do plano (NASCIMENTO, 2010).

Um modelo simples para descrever a vibração em moléculas diatômicas é a do oscilador harmônico (PASQUINI, 2003). Segundo esse modelo, os níveis energéticos vibracionais seriam igualmente espaçados e apenas transições vibracionais ( $\Delta v = \pm 1$ ) seriam permitidas (**Figura 3.5a**). Apesar de ser útil para o entendimento da energia vibracional, esse modelo falha para sistemas microscópicos (moléculas) por não considerar as forças coulômbicas de atração e repulsão e as dissociações encontradas em moléculas reais (BLANCO et al, 2002).

Deste modo, o comportamento das moléculas pode ser melhor representado pelo modelo do oscilador anarmônico, no qual os níveis de energia não são igualmente espaçados (**Figura 3.5b**). Esse modelo prevê a ocorrência de transições com  $\Delta v > \pm 2$  (sobretons) e a existência de bandas de combinação. Esses dois tipos de bandas são predominantes na região do infravermelho próximo (PEREIRA, 2006).



**Figura 3.5** – Variação de energia potencial em função da distância internuclear para: (a) oscilador harmônico, (b) oscilador anarmônico.  $d_e$  = distância de equilíbrio (PEREIRA, 2006).

O espectro na região do infravermelho próximo compreende a faixa de  $12800$  a  $4000\text{ cm}^{-1}$ , na qual as bandas de absorção são associadas principalmente a sobretons e combinações de vibrações fundamentais de ligações N-H, C-H, O-H e S-H que ocorrem na região do infravermelho médio (SKOOG et al, 2006).

As bandas de absorção na região do NIR são bem menos intensas (cerca de 10 a 1000 vezes) e mais sobrepostas que no MIR. Entretanto, isso não é uma desvantagem do NIR, visto que já se encontram disponíveis, para uso nessa região, fontes de radiação intensas e detectores de alta eficiência que contribuem para o aumento da relação sinal/ruído (HONORATO, 2006). Além disso, bons resultados podem ser obtidos quando se usa refletância difusa em amostras sólidas devido à melhor penetração da radiação (DE LIRA, 2010).

Uma das vantagens do NIR em relação ao MIR é que os instrumentos operados nessa região podem ser construídos utilizando sistemas ópticos semelhantes aos que operam na região do UV-Vis (SKOOG et al, 2006). Em geral, esses instrumentos são mais robustos quando comparados aos que operam na região do MIR, pois sua óptica é menos sensível à

umidade do ar. Os instrumentos na região do MIR também são eficientes; entretanto, possuem um custo mais elevado (PASQUINI, 2003).

A região do MIR compreende a faixa de 4000 - 400  $\text{cm}^{-1}$  do espectro eletromagnético. Nessa região, os espectros apresentam picos bem definidos correspondentes às transições fundamentais (primeiro harmônico). As bandas são intensas e os picos possuem absorbâncias bastante altas o que facilita a interpretação dos espectros. A faixa espectral do MIR conhecida como região de impressão digital (1200-600  $\text{cm}^{-1}$ ) é bastante útil para identificação de compostos, visto que pequenas diferenças na estrutura e na constituição das moléculas resultam em mudanças significativas no perfil e na distribuição dos picos de absorção. Deste modo é possível identificar, por exemplo, um composto orgânico comparando-se seu espectro com bancos de dados pré-existentes (SKOOG et al, 2006).

Enquanto a região do MIR ganhou ampla aceitação por parte dos espectroscopistas, a região do NIR foi pouco utilizada até 1950 (PASQUINI, 2003). A principal razão para isso foi atribuída à complexidade dos espectros, que apresentam bandas largas resultantes da sobreposição de picos individuais, dificultando sua interpretação. Nesse sentido, as primeiras publicações envolvendo a espectroscopia NIR só surgiram em 1965 com o trabalho de Karl Norris (BLANCO et al, 2002; PASQUINI, 2003).

A partir da década de 80 o número de artigos publicados utilizando essa técnica aumentou consideravelmente (McCLURE, 2003; PASQUINI, 2003). Tal aumento decorreu principalmente do desenvolvimento da microeletrônica e do uso de microcomputadores, que proporcionaram um avanço nas análises instrumentais, possibilitando a aquisição rápida e fácil de um grande número de dados de uma mesma amostra. Como consequência, o tratamento dos dados obtidos passou a exigir modelos mais complexos que iam além da calibração univariada (COSTA FILHO et al, 2002). Sendo assim, o uso de ferramentas quimiométricas se tornou

---

essencial à análise de dados químicos de natureza multivariada (ESBENSEN, 2002).

### **3.4 Quimiometria**

A Quimiometria pode ser considerada como um dos principais fatores que contribuíram para a difusão da espectroscopia NIR como uma poderosa ferramenta de análise em diversas aplicações qualitativas e quantitativas na química analítica (COSTA FILHO et al, 2002).

Os métodos quimiométricos se baseiam em ferramentas matemáticas e estatísticas para extrair informações relevantes de dados multivariados, muitas vezes complexos e de difícil interpretação, obtidos por diferentes técnicas instrumentais, tais como a espectroscopia na região do infravermelho (BARROS NETO, et al 2006). Esses métodos auxiliam a espectroscopia IR em várias situações, notadamente: (1) métodos de planejamento e otimização de experimentos; (2) pré-processamento de dados espectrais; (3) reconhecimento de padrões/classificação de amostras; (4) seleção de amostra e variáveis espectrais; (5) calibração multivariada (BEEBE et al, 1998).

#### **3.4.1 Pré-processamento**

O pré-processamento pode ser definido como qualquer manipulação matemática dos dados realizada antes da construção dos modelos quimiométricos (BEEBE et al, 1998). Em geral, as técnicas de pré-processamento são empregadas a dados espectrais (p.ex.: espectros NIR) visando remover ou reduzir fontes de variação sistemática ou aleatória que não estejam relacionadas ao problema químico de interesse. Tais variações podem ocorrer devido à instabilidade do instrumento, ao espalhamento da radiação durante a realização das medidas, ou ainda devido à variabilidade das propriedades físicas das amostras (SIMÕES, 2008).

As técnicas de pré-processamento de dados podem ser aplicadas tanto às amostras quanto às variáveis. Dentre as técnicas de pré-processamento aplicadas às amostras tem-se: normalização, ponderação, suavização, correção de linha de base, correção multiplicativa de sinal (MSC) e variação normal padrão (SNV). Já as técnicas mais comuns aplicadas às variáveis são: centralização dos dados na média, escalonamento e autoescalonamento (BEEBE et al, 1998).

A normalização e a ponderação são técnicas de pré-processamento muito semelhantes que consistem na divisão ou multiplicação de cada variável de um vetor de amostra por uma constante, removendo assim variações sistemáticas relativas às amostras (BEEBE et al, 1998).

Já a suavização é um método utilizado para reduzir matematicamente ruídos aleatórios de alta frequência a fim de aumentar a relação sinal/ruído. Para esse procedimento existem diferentes filtros digitais, tais como: média móvel, Savitzky-Golay (BEEBE et al, 1998), transformada de Fourier (CERQUEIRA et al, 2000) e transformada Wavelet (GALVÃO et al, 2001). Neste trabalho foram abordadas apenas as técnicas suavização com filtro média móvel e polinômio móvel com filtro Savitzky-Golay.

A idéia básica da suavização com filtro média móvel é a divisão do sinal analítico em uma série de intervalos de tamanho  $n$  (janela), de modo sequencial e ponto a ponto. Posteriormente é feita uma substituição do valor do ponto central do intervalo pelo valor médio desse (CERQUEIRA et al, 2000). Apesar de simples, o uso dessa técnica pode distorcer picos estreitos levando a perda de informação útil, uma vez que a suavização se dá pela média entre os pontos do sinal (BEEBE et al, 1998). Sendo assim, sua aplicação fica reduzida a espectros compostos por bandas largas, nos quais a largura do intervalo necessário para a redução do ruído não é crítica (CERQUEIRA et al, 2000).

O filtro de Savitzky-Golay é menos sensível ao tamanho do intervalo utilizado, pois, nesse caso um polinômio de baixa ordem é ajustado aos pontos de uma janela (intervalo) pelo método dos mínimos quadrados (CERQUEIRA et al, 2000). Esta característica possibilita que o método seja

---

aplicado a sinais analíticos com picos estreitos, apresentando resultados superiores aos obtidos pela média móvel. Entretanto, se o número de pontos da janela for elevado, picos podem ser removidos e os remanescentes distorcidos (BEEBE et al, 1998).

A primeira derivada é uma das técnicas de pré-processamento mais utilizadas para corrigir variações de linha de base em espectros MIR/NIR (DE LIRA, 2010), pois remove deslocamentos constantes de linha de base, além de possibilitar uma melhor visualização de picos existentes nos sinais originais (PONTES, 2009). Entretanto, o uso dessa técnica provoca, no caso de espectros, uma diminuição na relação sinal-ruído, tornando-se necessário o uso de filtros de suavização, tais como Savitzky-Golay citado anteriormente (DE LIRA, 2010).

A correção de linha de base é outro tipo de pré-processamento que pode ser aplicada para minimizar variações sistemáticas de linha de base. Nesse procedimento o espectro pré-processado é obtido subtraindo-se os dados originais a partir de uma função de feição de linha de base, a qual pode ser uma função linear (BEEBE et al, 1998). Essa ferramenta também pode ser associada a técnicas de suavização para minimização do ruído, tais como: filtro média móvel e polinômio móvel com filtro de Savitzky-Golay citadas anteriormente.

É importante salientar que a escolha da técnica de pré-processamento é uma etapa crítica da modelagem, pois a mesma pode influenciar tanto positiva quanto negativamente no desempenho do modelo. Desse modo, para que se execute um bom pré-processamento é necessário que haja uma interação entre a técnica utilizada e o reexame dos dados originais, além do conhecimento da natureza dos dados, bem como do processo químico ou físico do sistema em estudo (BEEBE et al, 1998).

### 3.4.2 Reconhecimento de padrões

As técnicas quimiométricas de reconhecimento de padrões (RP) utilizam ferramentas matemáticas e estatísticas que permitem explorar os resultados obtidos por meio de análises químicas, como por exemplo, medidas espectrométricas. E assim, verificar a existência de similaridades e diferenças entre as amostras que estejam relacionadas diretamente a sua composição química (CORREIA et al, 2007).

Entre as técnicas de RP têm-se as não supervisionadas: (1) Análise de Componentes Principais (PCA) e (2) Análise de Agrupamentos Hierárquicos (HCA). Essas são utilizadas apenas para examinar similaridades ou diferenças entre amostras, bem como para identificar a formação de padrões no espaço multidimensional.

Além dessas, tem-se as técnicas supervisionadas: (1) Modelagem Independente e Flexível por Analogia de Classe (SIMCA), (2) Método do K-ésimo Vizinho mais Próximo (KNN) e (3) Análise Discriminante Linear (LDA), que são utilizadas para classificação de amostras ou agrupamento em classes de acordo com suas características físicas e químicas (BEEBE et al, 1998). Neste trabalho foram abordadas apenas as técnicas de análise por componentes principais e análise discriminante linear.

A PCA consiste em uma manipulação matemática da matriz de dados a fim de representar as variações presentes em muitas variáveis através de um número menor de fatores, preservando o máximo de informação (CORREIA et al, 2007). Isso é possível pela combinação linear das variáveis originais, constituindo um novo sistema de eixos ortogonais entre si, denominados componentes principais (PC), fatores ou variáveis latentes (VL's) (FERREIRA et al, 1999). As coordenadas das amostras no novo sistema de eixos das PC's são chamadas de escores. A contribuição que cada variável original exerce sobre uma determinada PC é denominada peso (*loading*), que pode ser definida como sendo o cosseno do ângulo entre o eixo da variável e o eixo da PC (FERREIRA et al, 1999). Assim, cada amostra passa a ser representada por um ponto no espaço

---

multidimensional, do qual é possível extrair informações sobre a mesma e/ou sobre agrupamentos ou classes que apresentem características semelhantes (NASCIMENTO, 2008).

Entretanto, vale salientar que a PCA permite apenas a realização de uma análise exploratória dos dados. Para classificar uma amostra futura como pertencente a um ou mais agrupamentos inicialmente caracterizados utilizam-se técnicas de RP supervisionados, tais como a LDA (PONTES, 2009).

A LDA é uma técnica de reconhecimento de padrões supervisionada, que consiste na determinação de funções discriminantes lineares (FD's), as quais maximizem a variância entre as classes e minimizem a variância dentro de uma mesma classe. Assim como a PCA, a LDA pode ser considerada como um método de redução da dimensionalidade dos dados. A diferença entre elas é que, enquanto a PCA seleciona uma direção que contém a máxima variância dos dados, a LDA seleciona uma direção que alcança a separação máxima entre as classes avaliadas (PONTES, 2009).

O processo de classificação da LDA encontra-se descrito na literatura. Entretanto, é importante citar que nesse processo as amostras do conjunto de treinamento são sempre classificadas como pertencentes a uma ou outra classe, ou seja, se uma amostra não for incluída em sua própria classe, a mesma será incluída em uma classe mais próxima, o que caracteriza erros do tipo I e tipo II (PONTES et al, 2005).

Uma das limitações da LDA em relação às demais técnicas de reconhecimento de padrões supervisionadas é que a mesma é apropriada apenas para conjunto de dados com pequenas dimensões. Além disso, sua eficiência é bastante comprometida por problemas de colinearidade (PONTES, 2009). Desse modo, torna-se necessário o uso de técnicas de seleção de variáveis para eliminar as variáveis não informativas e reter as que resultem na máxima exatidão, principalmente quando se trata de dados espectrais. Um dos métodos de seleção de variáveis que pode ser empregado é o algoritmo das projeções sucessivas (SPA), que será apresentado na **Seção 3.4.4**.

### 3.4.3 Calibração multivariada

A Calibração multivariada é provavelmente a área da Quimiometria que tem atraído um maior interesse em aplicações da espectrofotometria na região do infravermelho (ESBENSEN, 2002). O propósito da calibração é estabelecer uma relação matemática quantitativa entre os dados de saída de um instrumento e o parâmetro físico ou químico de interesse previamente determinado por uma técnica independente (FERREIRA et al, 1999).

O processo geral de calibração consiste basicamente em três etapas: calibração, validação e previsão. A etapa de calibração estabelece um modelo matemático, obtido relacionando-se a matriz de dados das variáveis medidas (matriz  $\mathbf{X}$ ), por exemplo, espectros no infravermelho, com a matriz de dados das propriedades de interesse determinados por um método de referência (matriz  $\mathbf{Y}$ ). Na etapa de validação, o modelo é avaliado utilizando-se ferramentas estatísticas de diagnóstico. Por fim, o modelo validado é aplicado para prever propriedades de amostras desconhecidas, que possuam a mesma natureza das amostras do conjunto de calibração, a partir do seu sinal espectral (PERALTA-ZAMORA et al, 1997; FERREIRA et al, 1999).

A organização dos dados espectrais na calibração multivariada pode ser representada na forma de uma matriz  $\mathbf{X}$ , de dimensão  $n \times m$ , na qual as linhas representam as amostras cujos espectros foram registrados e as colunas as variáveis (comprimentos ou números de onda). O vetor  $\mathbf{y}$  com dimensão  $n$  representa o parâmetro de interesse (**Figura 3.6**).

$$\mathbf{X} = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{pmatrix} \quad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

Variáveis independentes                      Variável dependente

**Figura 3.6** – Representação das matrizes de dados ( $\mathbf{X}$  e  $\mathbf{y}$ ) na calibração multivariada (DE LIRA, 2010).

Durante a construção dos modelos de calibração multivariada, geralmente utiliza-se o método de validação cruzada completa para estabelecer a dimensionalidade do modelo. Na validação cruzada, uma amostra do conjunto de calibração é removida e um novo modelo é construído com os dados restantes. A amostra que foi removida tem seu valor previsto pelo modelo e um erro de calibração é calculado. Esse processo é repetido para cada amostra até que cada uma delas tenha sido excluída em algum momento (FERREIRA et al, 1999). A raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) é um dos parâmetros utilizados para avaliar a eficiência dos modelos calibração, conforme descrito pela **Equação 3.1** (BEEBE et al, 1998).

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n_{\text{cal}}} (y_{i,\text{cal}} - \hat{y}_{i,\text{cal}})^2}{n_{\text{cal}}}}, \quad (3.1)$$

onde  $n_{\text{cal}}$  é o número de amostras do conjunto de calibração,  $y_{i,\text{cal}}$  é o valor do parâmetro de interesse da  $i$ -ésima amostra do conjunto de calibração e  $\hat{y}_{i,\text{cal}}$  é o valor do parâmetro estimado pelo modelo.

A capacidade preditiva do modelo de calibração pode ser avaliada através de amostras que não participaram da etapa de construção dos modelos (conjunto de validação externa). Uma figura de mérito bastante utilizada para esse fim é a raiz quadrada do erro médio quadrático de

previsão (RMSEP), conforme descrito na **Equação 3.2** (BEEBE et al, 1998).

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^{n_{\text{prev}}} (y_{i,\text{prev}} - \hat{y}_{i,\text{prev}})^2}{n_{\text{prev}}}}, \quad (3.2)$$

onde  $n_{\text{prev}}$  é o número de amostras do conjunto de previsão,  $y_{i,\text{prev}}$  é o valor do parâmetro de interesse da *i-ésima* amostra do conjunto de previsão e  $\hat{y}_{i,\text{prev}}$  é o valor do parâmetro estimado pelo modelo.

Existem vários métodos matemáticos para realização da calibração multivariada, tais como: a Regressão Linear Múltipla (MLR), a Regressão por Componentes Principais (PCR) e a Regressão por Mínimos Quadrados Parciais (PLS) (PERALTA-ZAMORA et al, 1997).

Entre os métodos citados, a MLR é considerada a mais simples e de fácil interpretação, pois utiliza as variáveis originais na obtenção dos modelos (BEEBE et al, 1998). Nesse método, cada variável dependente do vetor,  $\mathbf{y}$ , é expressa como uma combinação linear das variáveis independentes da matriz  $\mathbf{X}$  por intermédio da **Equação 3.3**:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}, \quad (3.3)$$

onde  $\mathbf{b}$  é o vetor dos coeficientes de regressão e  $\mathbf{e}$  o vetor dos resíduos. O vetor  $\mathbf{b}$  dos coeficientes lineares é estimado, na etapa de calibração, utilizando o critério dos mínimos quadrados, que minimiza o vetor dos resíduos ( $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$ ) e é calculado pela **Equação 3.4**:

$$\mathbf{b} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}, \quad (3.4)$$

As estimativas para previsão dos parâmetros de interesse de amostras desconhecidas de mesma natureza das amostras de calibração podem ser obtidas de acordo com a **Equação 3.5**:

$$\hat{\mathbf{y}} = \mathbf{X}^* \mathbf{b} + \mathbf{e}, \quad (3.5)$$

onde  $\mathbf{X}^*$  é a matriz de dados para as novas amostras, e  $\mathbf{e}$  representa o vetor dos resíduos.

Apesar da simplicidade do método, o mesmo apresenta alguns problemas que limitam sua aplicação. Um desses problemas é que o número de variáveis deve ser menor que o número de amostras. Caso contrário, a matriz  $(\mathbf{X}^T \mathbf{X})$  não poderá ser invertida e a solução não existirá. Outro problema é a colinearidade, pois o cálculo de  $(\mathbf{X}^T \mathbf{X})^{-1}$  só poderá ser realizado se o determinante da inversa for diferente de zero. Portanto, se há colunas em  $\mathbf{X}$  que são colineares, ou seja, se as variáveis forem redundantes, o  $\det[(\mathbf{X}^T \mathbf{X})^{-1}]$  é zero ou próximo desse valor, causando problemas de instabilidade numérica e propagação de erros na determinação de  $\mathbf{y}$  (SENA et al, 2000). Em função disso, torna-se necessário o uso de algoritmos matemáticos para selecionar as variáveis relevantes para o parâmetro de interesse (DE LIRA, 2010).

O PLS é o método mais usado em calibração multivariada e difere do PCR por usar tanto a informação da matriz de dados  $\mathbf{X}$ , como também do vetor  $\mathbf{y}$  (ou da matriz  $\mathbf{Y}$ , no caso do PLS2) no cálculo das variáveis latentes. O PLS2 é uma versão do método PLS, no qual várias variáveis (ou parâmetros)  $\mathbf{Y}$  são modeladas simultaneamente (BEEBE et al, 1998).

No caso do PLS1, as matrizes  $\mathbf{X}$  e  $\mathbf{y}$  são decompostas em uma soma de  $h$  variáveis latentes, conforme as equações a seguir:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E}, \quad (3.6)$$

e

$$\mathbf{y} = \mathbf{UQ}^T + \mathbf{F} = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{E}, \quad (3.7)$$

onde as matrizes dos escores, pesos e resíduos são, respectivamente,  $\mathbf{T}$ ,  $\mathbf{P}$  e  $\mathbf{E}$  para a matriz  $\mathbf{X}$ , e  $\mathbf{U}$ ,  $\mathbf{Q}$  e  $\mathbf{F}$  para a matriz  $\mathbf{y}$ , respectivamente.

A correlação entre os dois blocos  $\mathbf{X}$  e  $\mathbf{y}$  é simplesmente uma relação linear obtida pelo coeficiente de regressão linear, tal como descrito abaixo:

$$\mathbf{u}_h = b_h \mathbf{t}_h, \quad (3.8)$$

Para "h" variáveis latentes, os valores de  $b_h$  são agrupados na matriz diagonal  $\mathbf{B}$ , que contém os coeficientes de regressão entre a matriz de escores  $\mathbf{U}$  de  $\mathbf{y}$  e a matriz de escores  $\mathbf{T}$  de  $\mathbf{X}$ . A melhor relação linear possível entre os escores desses dois blocos é obtida através de pequenas rotações das variáveis latentes dos blocos de  $\mathbf{X}$  e  $\mathbf{y}$ .

A matriz  $\mathbf{y}$  pode ser calculada de  $\mathbf{u}_h$ ,

$$\mathbf{y} = \mathbf{T}\mathbf{B}\mathbf{Q}^T + \mathbf{F}, \quad (3.9)$$

Por fim, as estimativas das propriedades de interesse de amostras "desconhecidas" são realizadas a partir dos novos escores,  $\mathbf{T}^*$ , substituídos na **Equação 3.10**.

$$\hat{\mathbf{y}} = \mathbf{T}^* \mathbf{B} \mathbf{Q}^T, \quad (3.10)$$

Nesse processo o número ótimo de variáveis latentes geralmente é encontrado utilizando-se o método de validação cruzada. Apesar do modelo PLS apresentar bons resultados em várias aplicações envolvendo a espectroscopia IR, os modelos de regressão que utilizam VL's ao invés das variáveis originais são em geral de difícil interpretação, pois nem sempre essas novas variáveis possuem algum significado físico ou físico-químico apropriado (DE LIRA, 2010).

#### 3.4.4 Seleção de variáveis

Vários métodos quimiométricos de RP e calibração multivariada apresentam limitações quando aplicados a um conjunto de dados com grande número de variáveis e que apresentam alguma correlação, conforme foi discutido anteriormente.

Nesse contexto, vários algoritmos de seleção de variáveis têm sido desenvolvidos com o intuito de solucionar esses problemas, dentre os quais se destacam: algoritmo genético (GA) (LEARDI, 2001), *stepwise* (SW) (FORINA et al, 2007), redes neurais artificiais (ANN) (CERQUEIRA et al, 2001), regressão por mínimos quadrados parciais por intervalo (iPLS) (NØRGAARD et al, 2000), algoritmo Jack Knife (JK) (NAES et al, 2002) e algoritmo das projeções sucessivas (SPA) (ARAÚJO et al, 2001). Nesse trabalho foram utilizados os algoritmos JK e o SPA.

O algoritmo JK pode ser utilizado para selecionar as variáveis mais informativas na construção dos modelos de regressão, incluindo-se o PLS. Nesse algoritmo, os erros padrão das estimativas dos coeficientes de regressão obtidos pelo modelo são avaliados utilizando-se uma série de subconjuntos de dados. A partir daí, pode-se calcular os intervalos de confiança para os coeficientes de regressão e avaliar a significância de cada variável utilizada no modelo (HONORATO et al, 2007).

O algoritmo SPA foi proposto por Araújo e colaboradores (ARAÚJO et al, 2001) como um método de seleção de variáveis no âmbito da calibração multivariada, demonstrando ser eficiente quando aplicado a modelos MLR (PONTES, 2009). Esse algoritmo utiliza operações simples de projeção para reduzir a colinearidade do conjunto de dados espectrais, de modo a melhorar o condicionamento numérico e reduzir a propagação de ruídos na construção de modelos MLR (HONORATO et al, 2007).

Os fundamentos teóricos, bem como os detalhes das projeções do SPA, têm sido amplamente apresentados em outros trabalhos (GALVÃO et al, 2001; HONORATO et al, 2007; PONTES, 2009). O SPA utiliza como base os conjuntos de calibração e validação, ambos contendo dados de respostas instrumentais ( $\mathbf{X}$ ) e valores dos parâmetros de referência ( $\mathbf{y}$ ). Basicamente, o objetivo do algoritmo é selecionar um subconjunto representativo de variáveis espectrais com ênfase na redução da colinearidade. A escolha do subconjunto ótimo de variáveis foi realizada inicialmente com base no menor valor da raiz quadrada do erro médio

quadrático de validação (RMSEV) obtido para o modelo MLR, conforme mostrado na **Equação 3.11**:

$$\text{RMSEV} = \sqrt{\frac{\sum_{i=1}^{n_{\text{val}}} (y_{i,\text{val}} - \hat{y}_{i,\text{val}})^2}{n_{\text{val}}}} \quad (3.11)$$

Galvão e colaboradores (GALVÃO et al, 2007) adaptaram o SPA para a validação cruzada, no qual a escolha das variáveis é baseada no menor valor do RMSECV (**Equação 3.1**). Desse modo, o melhor subconjunto de variáveis é aquele que, após a construção do modelo MLR, apresentar o menor RMSECV. Posteriormente, o modelo validado é utilizado para estimar (prever) o valor do parâmetro de interesse de amostras desconhecidas de mesma natureza das amostras de calibração a partir do sinal espectral.

Em geral, os modelos MLR/SPA são mais simples e parcimoniosos que os modelos PLS por utilizar poucas variáveis espectrais na modelagem. Além disso, vários estudos têm demonstrado que a capacidade preditiva dos modelos MLR/SPA é comparável, ou até melhor do que os modelos PLS com todas as variáveis (DI NEZIO et al, 2007; HONORATO et al, 2008; PISTONESI et al, 2010; PONTES et al, 2011).

No âmbito da classificação, o algoritmo SPA foi adaptado por Pontes e colaboradores (PONTES et al, 2005) para o uso na LDA que, conforme discutido anteriormente, também é bastante afetada por problemas de colinearidade. Com essa adaptação, o risco médio  $G$  de uma classificação incorreta pela LDA (**Equação 3.12**), calculada com base no conjunto de validação, foi adotada como função de custo para a escolha das variáveis.

$$G = \frac{1}{K_v} \sum_{k=1}^{K_v} g_k, \quad (3.12)$$

onde  $g_k$  é o risco de uma classificação incorreta do objeto  $x_k$  da  $k$ -ésima amostra de validação e é definido como:

$$g_k = \frac{r^2(\mathbf{x}_k, \mu_{Ik})}{\min_{Ij \neq Ik} r^2(\mathbf{x}_k, \mu_{Ij})} \quad (3.13)$$

Na equação acima (**Equação 3.13**), o numerador  $r^2(\mathbf{x}_k, \mu_{Ik})$  representa o quadrado da distância de Mahalanobis entre o objeto  $\mathbf{x}_k$  (com índice de classe  $Ik$ ) e a média de sua classe  $\mu_{Ik}$ . E o denominador corresponde ao quadrado da distância de Mahalanobis entre o objeto  $\mathbf{x}_k$  e o centro da classe errada mais próxima. Idealmente, o valor de  $g_k$  deverá ser tão pequeno quanto possível, ou seja, o objeto  $\mathbf{x}_k$  deverá estar perto do centro da sua verdadeira classe e distante dos centros das demais classes (PONTES, 2009).

Vários trabalhos descritos na literatura têm demonstrado o bom desempenho da LDA/SPA em problemas de classificação (PONTES et al, 2009; MOREIRA et al, 2009; GAMBARRA-NETO et al, 2009; SOUTO et al, 2010).

### 3.4.5 Seleção de amostras

Na Quimiometria um algoritmo clássico utilizado para seleção de amostras é o Kennard-Stone (KS) (KENNARD e STONE, 1969). Nesse algoritmo, as distâncias euclidianas entre os vetores das respostas instrumentais ( $\mathbf{X}$ ) das amostras selecionadas são maximizadas (GALVÃO et al, 2005). Tal distância é descrita como  $d_x(p,q)$  entre os X-vetores de cada par  $(p,q)$  de amostras selecionadas. A **Equação 3.14** descreve o cálculo da distância Euclidiana  $d_x(p,q)$  utilizada para seleção de amostras pelo algoritmo KS.

$$d_{x(p,q)} = \sqrt{\sum_{j=1}^J [x_p(j) - x_q(j)]^2}; \quad p, q \in [1, N] \quad (3.14)$$

Para dados espectrais,  $x_p(j)$  e  $x_q(j)$  representam as respostas instrumentais no  $j$ -ésimo comprimento de onda para as amostras "p" e "q", respectivamente, e "j" corresponde ao número de comprimentos de onda no espectro.

Em problemas de classificação, geralmente utiliza-se o algoritmo KS para selecionar as amostras dos conjuntos de treinamento, validação e teste, como no caso da LDA empregada neste trabalho. Já na calibração multivariada, geralmente utiliza-se o algoritmo SPXY (*Sample set Partitioning based on joint  $\mathbf{X}$ - $\mathbf{y}$  distances*) (GALVÃO et al., 2005) na seleção de amostras representativas de calibração/validação.

O SPXY é uma adaptação do KS, e diferentemente deste, considera tanto as diferenças de  $\mathbf{X}$  quanto as de  $\mathbf{y}$  no cálculo das distâncias inter-amostras. Nesse caso, cada distância  $d_y(p,q)$  pode ser calculada para cada par de amostras "p" e "q", conforme descrito na **Equação 3.15**:

$$d_y(p, q) = \sqrt{(y_p - y_q)^2} = |y_p - y_q|; p, q \in [1, N] \quad (3.15)$$

Para assegurar a igualdade de importância para a distribuição das amostras no espaço de  $\mathbf{X}$  e  $\mathbf{y}$ , as distâncias  $d_x(p,q)$  e  $d_y(p,q)$  são divididas pelos seus valores máximos no conjuntos de dados, tendo-se assim uma distância  $Xy$  normalizada, conforme descrito **Equação 3.16**:

$$d_{xy}(p, q) = \frac{d_x(p,q)}{\max_{p,q \in [1, N]} d_x(p,q)} + \frac{d_y(p,q)}{\max_{p,q \in [1, N]} d_y(p,q)}; p, q \in [1, N] \quad (3.16)$$

### **3.5 Métodos analíticos baseados na espectroscopia no infravermelho para análise de misturas diesel/biodiesel**

Assim como acontece para outros combustíveis, tais como: gasolina (FELÍCIO et al, 2005; HONORATO et al, 2008), óleo diesel (BREITKREITZ et al, 2003) e etanol (BROWN et al, 2011) o uso das espectroscopias MIR/NIR para monitoramento da qualidade tanto do biodiesel (B100) quanto das misturas diesel/biodiesel já é uma realidade e os resultados obtidos até o momento são bastante promissores.

Recentes trabalhos têm demonstrado a viabilidade do uso da espectroscopia na região do infravermelho associada à calibração multivariada para: (1) o monitoramento da qualidade do biodiesel (ZAGONEL et al, 2004; FELIZARDO et al, 2007; BAPTISTA et al, 2008; DE

LIRA et al, 2010); (2) a previsão dos principais parâmetros de qualidade de misturas diesel/biodiesel (DE LIRA et al, 2010); (3) a determinação do teor de biodiesel; e (4) a identificação de adulteração por óleo vegetal. Alguns trabalhos citados na literatura referentes aos itens 3 e 4 são apresentados a seguir.

PIMENTEL et al (2006) descreveram a utilização de modelos de calibração multivariada (PLS) baseados em espectros MIR (4000-650  $\text{cm}^{-1}$ ; sonda ATR-refletância total atenuada) e NIR (12000-4000  $\text{cm}^{-1}$ ; caminho óptico = 1,0 mm) para determinar o teor de biodiesel em misturas com óleo diesel considerando a presença de óleos vegetais. Os valores do RMSEP obtidos foram 0,18% v/v (NIR) e 0,25% v/v (MIR). De acordo com o teste F (a um nível de confiança de 95%), não houve diferença estatisticamente significativa entre os modelos construídos empregando as duas regiões espectrais. Este estudo demonstrou ainda que é possível identificar contaminações com óleos vegetais crus usando PCA e espectros MIR.

Ainda em 2006, OLIVEIRA et al avaliaram o uso das espectroscopias NIR na faixa de 9000-4700  $\text{cm}^{-1}$  (caminho óptico = 2,0 mm) e MIR na faixa de 4000-650  $\text{cm}^{-1}$  (sonda ATR-refletância total atenuada) associadas a modelos de calibração PLS e Redes Neurais (ANN) para determinação de ésteres metílicos em misturas de biodiesel (0-100% m/m) com diesel. Os resultados demonstraram que o modelo PLS/FTIR-NIR (RMSEP = 0,06% m/m) apresentou um melhor desempenho com relação aos modelos PLS/FTIR-ATR (RMSEP = 0,20% m/m), ANN/FTIR-NIR (RMSEP = 0,21% m/m) e ANN/FTIR-ATR (RMSEP = 0,21% m/m).

OLIVEIRA et al (2007) investigaram a adulteração de misturas B2 e B5 com óleos vegetais por meio das espectroscopias vibracionais NIR (caminho óptico = 2,0 mm) e Raman (caminho óptico = 0,5 cm) com transformada de Fourier. Eles avaliaram o desempenho de modelos de calibração multivariada baseados em PLS, PCR e Redes Neurais. Os resultados obtidos demonstraram que entre os modelos de calibração

construídos, aqueles que empregaram Redes Neurais e FT-Raman apresentaram uma melhor exatidão para amostras externas de validação.

ALISKE et al (2007) e GUARIEIRO et al (2008) desenvolveram métodos semelhantes para determinação do teor de biodiesel em misturas diesel/biodiesel utilizando a espectroscopia MIR e a calibração univariada. As determinações foram realizadas utilizando a área do pico de absorção referente ao grupo carbonila (C = O), 1700 – 1800  $\text{cm}^{-1}$ . Os métodos propostos mostraram-se satisfatórios para medidas de biodiesel em misturas com diesel, mesmo em baixas concentrações de biodiesel (0,1% v/v).

Em 2008, SOARES et al avaliaram a adulteração de amostras de biodiesel de diferentes fontes (algodão, mamona e palma) com óleo de soja *in natura*, utilizando espectroscopia MIR e modelos PLS com e sem seleção de variáveis. Eles avaliaram dois métodos de seleção de variáveis: *forward* (FW) e *stepwise* (SW). Os modelos construídos utilizando-se os métodos de seleção de variáveis apresentaram uma melhor correlação entre os valores de referência e os valores previstos pelo modelo.

MAHAMUNI e ADEWUYI (2009) desenvolveram um método analítico utilizando a espectroscopia no infravermelho (FTIR-ATR) para monitoramento da reação de transesterificação na obtenção do biodiesel e para a determinação do teor de biodiesel (1-100% m/m) em misturas com diesel. Além disso, após pequenas modificações no método, o mesmo foi utilizado para quantificar o teor de óleo vegetal (1-40% m/m) adicionado às misturas diesel/biodiesel, simulando esse tipo de adulteração. Os modelos de calibração multivariada foram construídos utilizando o método clássico dos mínimos quadrados (CLS). Os valores de RMSEP obtidos com os modelos para quantificação do biodiesel durante a reação e para determinação do teor de biodiesel em misturas diesel/biodiesel foram 0,13 %m/m e 0,12 %m/m, respectivamente. Já para determinação do teor de óleo em misturas diesel/biodiesel o valor de RMSEP foi de 0,20 %v/v.

---

GAYDOU e DUPUY (2010) investigaram o uso simultâneo dos espectros NIR (10000 - 4430  $\text{cm}^{-1}$ ; caminho óptico = 2,0 mm) e MIR (4000 - 650  $\text{cm}^{-1}$ , FTIR-ATR) para quantificação do óleo vegetal (0-30% m/m) adicionado às misturas diesel/biodiesel (0-10% m/m). O método PLS-Hierárquico (H-PLS) foi utilizado inicialmente para combinar as duas regiões (matrizes), e em seguida os modelos de calibração foram construídos utilizando o método PLS-Serial (Serial-PLS). Também, foram construídos modelos PLS para as duas faixas espectrais separadamente. Os resultados obtidos demonstraram que, entre os modelos de calibração construídos, aquele baseado em FTIR-NIR e PLS apresentou melhor exatidão para amostras externas de validação (RMSEP = 0,36% m/m).

# **MATERIAIS E MÉTODOS**

## **Capítulo 4**

## 4 MATERIAIS E MÉTODOS

### 4.1 Amostras

Foram preparadas misturas contendo diesel/biodiesel, diesel/óleo e diesel/biodiesel/óleo. Para tanto, foram utilizadas diferentes amostras de diesel mineral fornecidas pela TRANSPETRO (Petrobrás Transporte S.A), localizada no pólo petroquímico de Suape/PE. As amostras de biodiesel (B100) consistiram de: uma amostra de biodiesel de soja, uma mistura de biodiesel de sebo (62% v/v) com biodiesel de soja (38% v/v) e uma amostra de biodiesel de girassol. As duas primeiras amostras de biodiesel foram fornecidas pelas empresas BRASBIODIESEL e BRACOL HOLDING, respectivamente. A amostra de biodiesel de girassol foi produzida em escala laboratorial, conforme procedimento descrito na **Seção 4.2**. No caso das amostras de óleos vegetais *in natura*, foram utilizados óleos comerciais de soja, girassol e óleo de fritura residual.

### 4.2 Síntese do biodiesel

O biodiesel de girassol foi produzido em uma planta piloto (**Figura 4.1**), localizada no Laboratório de Combustíveis (LAC) da UFPE (Universidade Federal de Pernambuco), a partir da reação de transesterificação do óleo vegetal com álcool metílico na razão molar 6:1 utilizando-se o hidróxido de potássio como catalisador (1,0% m/m).

Inicialmente, aqueceu-se o óleo de girassol a temperatura de 60°C em um reator (1,5L). Em seguida, adicionou-se o hidróxido de potássio completamente dissolvido em metanol. Deixou-se o sistema sob agitação constante (300 rpm) a temperatura de 60°C por 1h. Após esse período, a mistura reacional foi transferida para um funil de decantação deixando-se separar as fases. Recolheram-se separadamente as fases pesada (glicerina) e leve (biodiesel). O excesso de álcool presente no biodiesel foi recuperado e, em seguida, o biodiesel foi lavado.

A primeira lavagem foi feita utilizando-se uma solução de HCl (0,1 mol/L) para neutralizar o catalisador e, em seguida, foram feitas

sucessivas lavagens com água destilada até que um pH entre 6,0 e 7,0 fosse atingido. Concluída esta etapa, o biodiesel foi aquecido para retirada da água e posteriormente seco com sulfato de sódio anidro.



**Figura 4.1** – Planta piloto utilizada para produção do biodiesel de girassol, localizada no Laboratório de Combustíveis (LAC) da UFPE.

### 4.3 Parâmetros de qualidade do biodiesel

Os parâmetros de qualidade das amostras de biodiesel de soja, girassol e da mistura de biodiesel de sebo (62%v/v) e soja (38%v/v) atendem às especificações da ANP (Resolução N° 7 de 2008).

Apesar das amostras de biodiesel não apresentarem um teor de éster igual a 100% m/m, esta suposição foi aceita para fins deste estudo, visto que não foram observadas quantidades significativas de triacilglicerídeos, glicerina, metanol, dentre outros contaminantes, nos certificados apresentados pelas empresas, bem como nos cromatogramas obtidos para o biodiesel de girassol.

A determinação do teor de éster do biodiesel de girassol (EN 14103) foi realizada utilizando-se um cromatógrafo da marca Shimadzu GC 17-A

com detector de ionização de chama, equipado com uma coluna (DB-WAX) 30m, 0,25mm e espessura do filme de 0,25 $\mu$ m.

#### 4.4 Preparação das misturas

As misturas contendo diesel/biodiesel, diesel/óleo, diesel/biodiesel/óleo, além de diesel puro, foram preparadas segundo um planejamento de misturas de três componentes (concentração de diesel, biodiesel e óleo vegetal) com ponto central. O planejamento de misturas utilizado foi o *Simplex Lattice Design* (REIS et al, 1996), o qual foi calculado utilizando-se o software Unscrambler<sup>®</sup> 9.7.

As amostras foram preparadas à temperatura ambiente utilizando-se uma balança semi-analítica (Mettler<sup>®</sup>,  $\pm 0,01$ g). Para esse procedimento, foi adicionada uma alíquota de biodiesel e/ou óleo de massa conhecida em um frasco âmbar de 250 mL, seguida da adição do diesel e posterior homogeneização. As frações mássicas (% m/m) dos componentes das misturas foram convertidas em frações volumétricas (% v/v) através da densidade, a fim de expressar os resultados para teor de biodiesel na mesma unidade de medida referenciada nas normas EN14078 e ABNT15568.

As concentrações de biodiesel e óleo nas misturas com diesel variaram de 0,0 a 10% v/v, com incrementos de 1,0% v/v. A razão entre as concentrações de biodiesel/óleo contidas nas misturas ternárias (diesel/biodiesel/óleo) variou de 0,25 a 4,0, de acordo com o planejamento mostrado no **Anexo 1**. Esse planejamento foi aplicado a cada uma das amostras de biodiesel (soja, mistura do biodiesel de sebo (68%v/v) e soja (32%v/v) e girassol) e óleos correspondentes (soja, óleo de fritura e girassol), utilizando-se diferentes amostras de diesel, totalizando 99 misturas.

De modo a atender a atual legislação brasileira, que obriga a adição de 5,0% v/v de biodiesel ao óleo diesel (RESOLUÇÃO CNPE N° 6, 2009) foram adicionadas ao planejamento mais 13 amostras de diesel/biodiesel

com esse percentual, além das três já existentes no planejamento. Além dessas, foram adicionadas também 5 misturas diesel/biodiesel com percentual de biodiesel variando de 2,5% a 6,5% v/v (com incrementos de 1,0%) e mais 5 misturas diesel/óleo nas mesmas proporções. Além de 8 amostras de óleo diesel, e 2 misturas diesel/biodiesel/óleo com razão entre as concentrações de biodiesel/óleo igual a 6,0. No total foram preparadas 132 amostras (conjunto de dados).

#### 4.5 Aquisição dos espectros NIR

Os espectros NIR (14000 a 4000  $\text{cm}^{-1}$ ) foram registrados em um espectrofotômetro de infravermelho com transformada de Fourier (FT-IR) modelo FTLA2000-160 da marca Bomen<sup>®</sup> (**Figura 4.2**).

As medidas espectrais foram realizadas utilizando-se células de quartzo com diferentes percursos ópticos (1,0, 10 e 50 mm), a fim de avaliar a região de sobretons cujas absorções serão maiores em função do aumento do caminho óptico. Todos os espectros foram registrados a temperatura de  $23 \pm 1^\circ\text{C}$ , com resolução espectral de  $8,0 \text{ cm}^{-1}$  e média de 32 varreduras. Após cada medida foi realizada a limpeza da célula utilizando hexano.



**Figura 4.2**– Espectrofotômetro Infravermelho FT-IR (FTLA 2000-160) da Bomen<sup>®</sup>.

## 4.6 Tratamento quimiométrico dos dados

### 4.6.1 Pré-processamento

Em geral, os espectros NIR das amostras analisadas ao longo deste estudo apresentaram variações sistemáticas de linha de base e pouco ruído aleatório. Para minimizar esses problemas, diferentes estratégias de pré-processamento dos espectros foram avaliadas: (1) correção de linha de base seguida de suavização com filtro média móvel (LB/SMM); (2) correção de linha de base seguida de suavização por polinômio móvel com filtro Savitzky-Golay (LB/SSG) e (3) primeira derivada com filtro Savitzky-Golay e polinômio de segunda ordem, variando-se o tamanho das janelas de 5 a 11 pontos. A correção de linha de base linear também foi estudada. As variáveis (números de onda) foram centradas na média antes da modelagem.

### 4.6.2 Reconhecimento de padrões

Inicialmente foi realizada uma análise exploratória dos dados utilizando-se a PCA a fim de avaliar se é possível identificar similaridades e diferenças entre as amostras do conjunto de dados utilizadas na construção dos modelos de classificação e calibração de acordo com a sua composição química para cada um dos caminhos ópticos estudados (1,0, 10 e 50 mm).

Em seguida, foram construídos modelos de classificação utilizando-se a Análise Discriminante Linear com seleção de variáveis pelo Algoritmo das Projeções Sucessivas (LDA/SPA).

Para esse procedimento foram utilizados os espectros NIR pré-processados das 132 amostras do conjunto de dados, com mais 22 amostras de diesel puro, totalizando 154 amostras. Antes da obtenção dos modelos foram excluídas quatro amostras anômalas (*outliers*) do conjunto de dados, com base no gráfico dos resíduos versus o leverage, perfazendo um total de 150 amostras.

A escolha da estratégia de pré-processamento dos dados espectrais foi baseada na análise dos gráficos de escores e nos erros de classificação apresentados por cada uma das técnicas utilizadas.

Para os estudos da PCA, as 150 amostras foram divididas em quatro conjuntos (classes): diesel (35 amostras), diesel/biodiesel (42 amostras), diesel/óleo (35 amostras), e diesel/biodiesel/óleo (38 amostras).

Para os estudos de classificação LDA/SPA, dividiu-se cada uma das classes acima em três subconjuntos: treinamento (50%), validação (25%) e teste (25%) utilizando o algoritmo KS. A **Tabela 4.1** mostra o número de amostras utilizado para construção dos modelos de classificação LDA/SPA.

**Tabela 4.1** – Número de amostras dos conjuntos de treinamento, validação e teste utilizados para construção dos modelos LDA/SPA.

Classe	Conjunto		
	Treinamento (50%)	Validação (25%)	Teste (25%)
D (35)	19	8	8
DB (42)	22	10	10
DO (35)	19	8	8
DBO (38)	20	9	9
<b>Total</b>	<b>80</b>	<b>35</b>	<b>35</b>

As amostras dos conjuntos de treinamento e validação foram empregadas na etapa de seleção de variáveis pelo SPA. Em particular, as amostras do conjunto de validação foram utilizadas na seleção do número ótimo de variáveis (números de onda) com base na minimização da função de custo G (PONTES et al, 2005). As amostras do conjunto de teste foram utilizadas para avaliar a eficiência dos modelos de classificação.

#### 4.6.3 Calibração multivariada

A fim de determinar o teor de biodiesel em misturas com diesel adulteradas com óleo vegetal *in natura*, foram construídos modelos de

calibração multivariada utilizando-se os espectros NIR pré-processados para cada caminho óptico (1,0, 10 e 50 mm) e suas respectivas concentrações de biodiesel. Para tanto, foram empregadas as técnicas de Regressão por Mínimos Quadrados Parciais (PLS) e a Regressão linear múltipla (MLR) e dois métodos de seleção de variáveis: (1) coeficientes de regressão significativos calculados pelo algoritmo de Jack-Knife (PLS/JK) e (2) Algoritmo das Projeções Sucessivas (MLR/SPA).

Para verificar a existência de amostras anômalas foi construído um modelo PLS com as 132 amostras do conjunto de dados e o gráfico dos resíduos versus o *leverage*, tanto em  $\mathbf{X}$  quanto em  $\mathbf{y}$ , foi avaliado. As amostras com alto resíduo e alto *leverage* foram excluídas do conjunto de dados. Em seguida, um novo modelo foi construído e os parâmetros como coeficiente de correlação (R) e o RMSECV foram avaliados. Esse procedimento foi aplicado para cada caminho óptico (1,0, 10 e 50 mm) separadamente.

Após essa etapa, as amostras foram divididas nos conjuntos de calibração (2/3) e validação externa (1/3) utilizando-se o algoritmo SPXY (GALVÃO, et al 2005).

As amostras do conjunto de calibração foram empregadas nos procedimentos de modelagem utilizando o método da validação cruzada completa para determinação do número ideal de fatores (variáveis latentes) para construção dos modelos PLS. O número de variáveis latentes foi escolhido com base no menor valor de RMSECV e na análise das ferramentas de diagnóstico do modelo (gráfico da variância explicada versus RMSECV, gráfico dos *loadings* e do resíduo dos modelos). Os coeficientes de regressão mais significativos empregados na construção dos modelos PLS foram selecionados pelo algoritmo de Jack-Knife disponível no programa Unscrambler®.

O método da validação cruzada também foi empregado no algoritmo SPA para selecionar o melhor conjunto de números de onda para os modelos MLR. Para esse procedimento foram utilizadas todas as amostras de calibração.

A capacidade preditiva dos modelos foi avaliada através do RMSEP, obtido para o conjunto de validação externa. Um teste  $F$ , em um nível de confiança de 95%, foi utilizado para avaliar se existem diferenças estatisticamente significativas entre os valores de RMSEP dos modelos PLS, PLS/JK e MLR/SPA. Os valores de  $F_{cal}$  foram obtidos calculando-se a razão entre os valores quadráticos do maior e menor RMSEP obtidos pelos modelos. Esses valores de  $F_{cal}$  foram comparados com os valores de  $F_{crit}$  ( $n, n, 0,95$ ) tabelado.

Segundo a norma ABNT 15568:2008, o RMSEP não pode ser superior a 0,10% v/v para os modelos PLS construídos na faixa de concentração de biodiesel de 0-8,0% v/v, e nem superior a 1,0 % v/v para os modelos PLS construídos na faixa de concentração de biodiesel de 8,0-30% v/v, utilizando-se espectros medidos no Infravermelho Médio (MIR). Entretanto, vale salientar que os modelos de calibração desenvolvidos no presente estudo utilizaram concentrações de biodiesel e óleo variando de 0 a 10% v/v.

O programa Unscrambler<sup>®</sup> 9.7 (CAMO S.A) foi empregado no pré-processamento dos dados e na obtenção dos modelos PCA, PLS e PLS/JK. Os algoritmos de Kennard-Stone (KS), LDA/SPA, SPXY e MLR/SPA foram executados utilizando-se rotinas escritas em Matlab<sup>®</sup> R2008a.

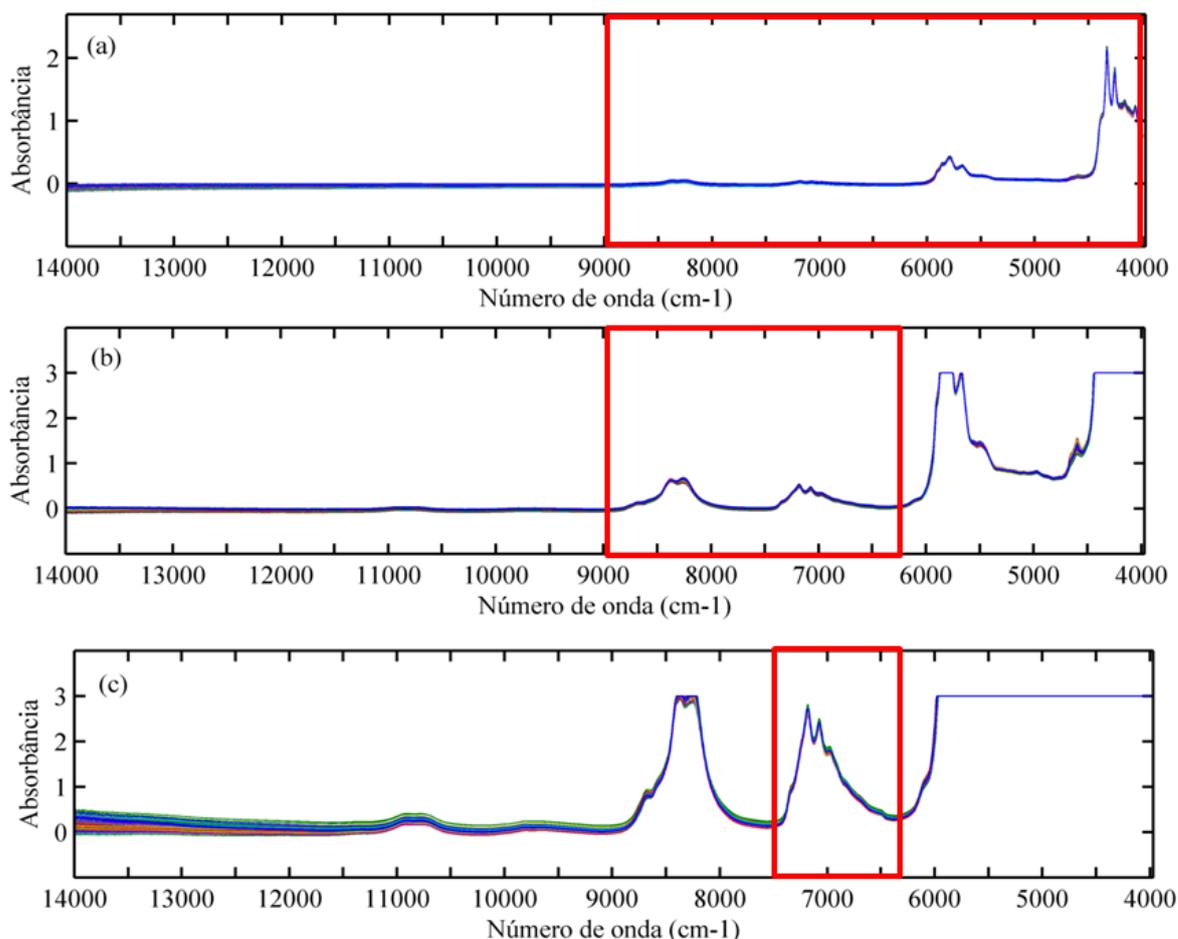
# **RESULTADOS E DISCUSSÃO**

## **Capítulo 5**

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Espectros NIR

As **Figuras 5.1a-c** mostram os espectros NIR das 132 amostras registrados na faixa de 14000-4000  $\text{cm}^{-1}$  para os diferentes caminhos ópticos utilizados no presente estudo.



**Figura 5.1** – Espectros NIR das 132 amostras registrados na faixa de 14000-4000  $\text{cm}^{-1}$  para os caminhos ópticos: (a) 1,0 mm; (b) 10 mm; (c) 50 mm. As regiões espectrais utilizadas no tratamento quimiométrico para os diferentes caminhos ópticos estão destacadas em vermelho.

Esses espectros apresentam bandas associadas à região das combinações (4800-4000  $\text{cm}^{-1}$ ), primeiro sobretudo das combinações (7400-6900  $\text{cm}^{-1}$ ), primeiro sobretudo (6300-5180  $\text{cm}^{-1}$ ) e segundo sobretudo (8910-8000  $\text{cm}^{-1}$ ) do estiramento de ligações C-H. Além dessas, são observadas bandas em torno de 5500  $\text{cm}^{-1}$  e na região de 4740-4370  $\text{cm}^{-1}$ , as quais são atribuídas ao segundo sobretudo do grupamento C=O e

à combinação das bandas de estiramento de ligações C–H e C=O respectivamente. Para o caminho óptico de 50 mm (**Figura 5.1c**), observa-se também a banda ( $11200\text{-}10500\text{ cm}^{-1}$ ) associada ao terceiro sobretom dos modos de estiramento das ligações C–H (WORKMAN, 2008).

Para avaliar o uso da região de sobretoms para determinação do teor de biodiesel em misturas com diesel adulteradas com óleo vegetal e classificação foram selecionadas as seguintes faixas espectrais:  $9000\text{-}6300\text{ cm}^{-1}$  (10 mm) e  $7500\text{-}6300\text{ cm}^{-1}$  (50 mm), destacadas nas **Figuras 5.1b** e **c**, respectivamente. Para fins de comparação com as faixas espectrais supracitadas foi selecionada a faixa espectral de  $9000\text{-}4000\text{ cm}^{-1}$  (1,0 mm), que inclui tanto as bandas de combinação quanto as bandas de sobretoms, destacada na **Figura 5.1a**.

## 5.2 Reconhecimento de padrões

Os métodos de reconhecimento de padrões que serão apresentados a seguir foram construídos utilizando-se os espectros NIR pré-processados das 132 amostras do conjunto de dados (incluindo-se 22 amostras de diesel puro) para cada um dos três caminhos ópticos estudados: 1,0 mm ( $9000\text{-}4000\text{ cm}^{-1}$ ); 10 mm ( $9000\text{-}6300\text{ cm}^{-1}$ ) e 50 mm ( $9000\text{-}7500\text{ cm}^{-1}$ ). Conforme foi descrito na **Seção 4.6.2** quatro amostras anômalas (*outliers*) foram excluídas do conjunto de dados, perfazendo um total de 150 amostras.

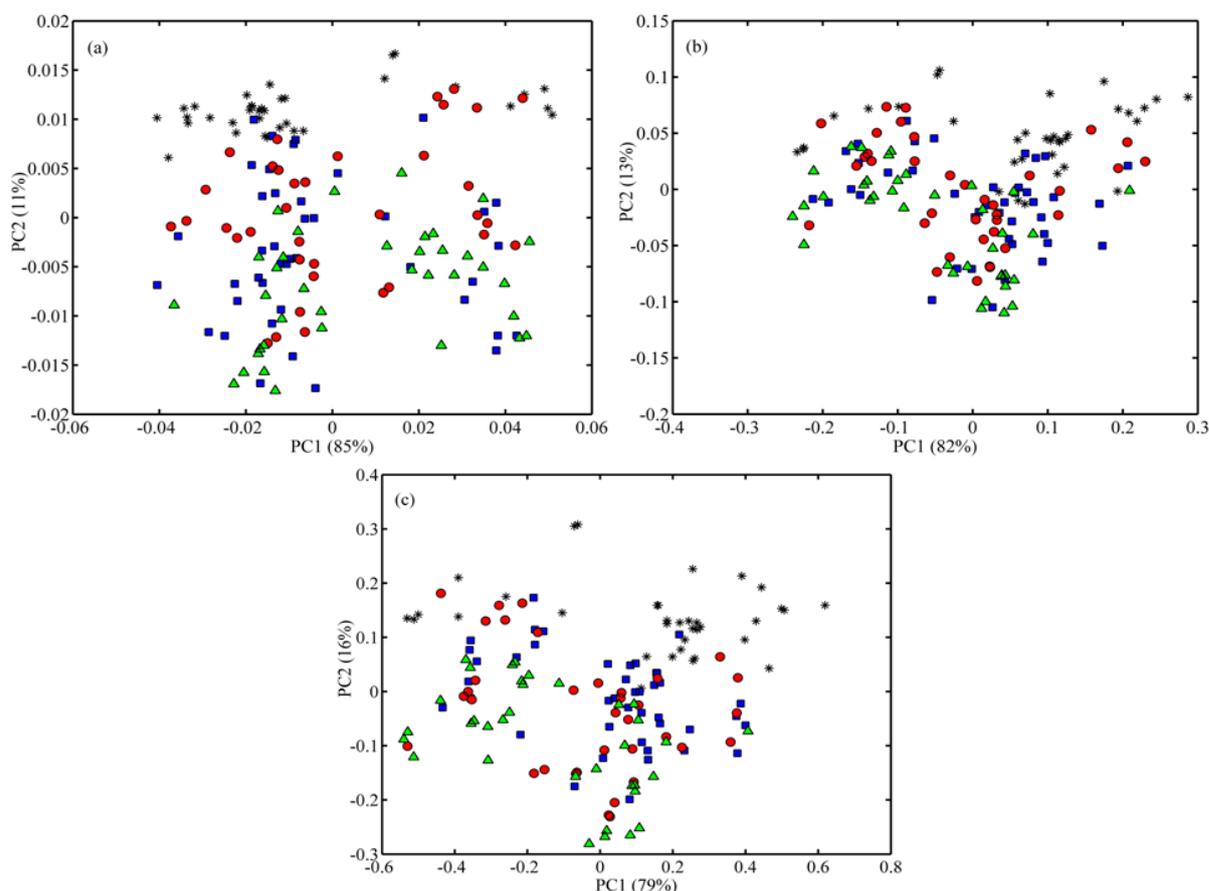
Em geral, os espectros NIR das 150 amostras apresentaram pequenas variações de linha de base para faixas espectrais selecionadas (**Figura 5.1**), e ampliando um dos espectros também observou-se pouco ruído aleatório. Por esse motivo, os métodos de reconhecimento de padrões foram obtidos utilizando-se os espectros NIR pré-processados.

Para o caminho óptico de 1,0 mm, foram utilizados os espectros NIR pré-processados com primeira derivada com filtro Savitzky–Golay e polinômio de segunda ordem com janela de 11 pontos. Para os caminhos ópticos de 10 e 50 mm foram utilizados os espectros NIR pré-processados

com correção linear de linha de base seguida de suavização com filtro média móvel e janela 5 de pontos.

### 5.2.1 Análise por Componentes Principais (PCA)

Nas **Figuras 5.2a-c** são apresentados os gráficos dos escores obtidos para PC2 *versus* PC1 para os caminhos ópticos de 1,0, 10 e 50 mm, respectivamente. Como pode ser observado há uma forte sobreposição entre as classes, não sendo possível obter uma discriminação efetiva das amostras de diesel puro (D) em relação às demais amostras (DB, DO e DBO). Apesar desses resultados, foram construídos modelos de classificação LDA/SPA para tentar detectar adulteração das misturas diesel/biodiesel por óleo vegetal.



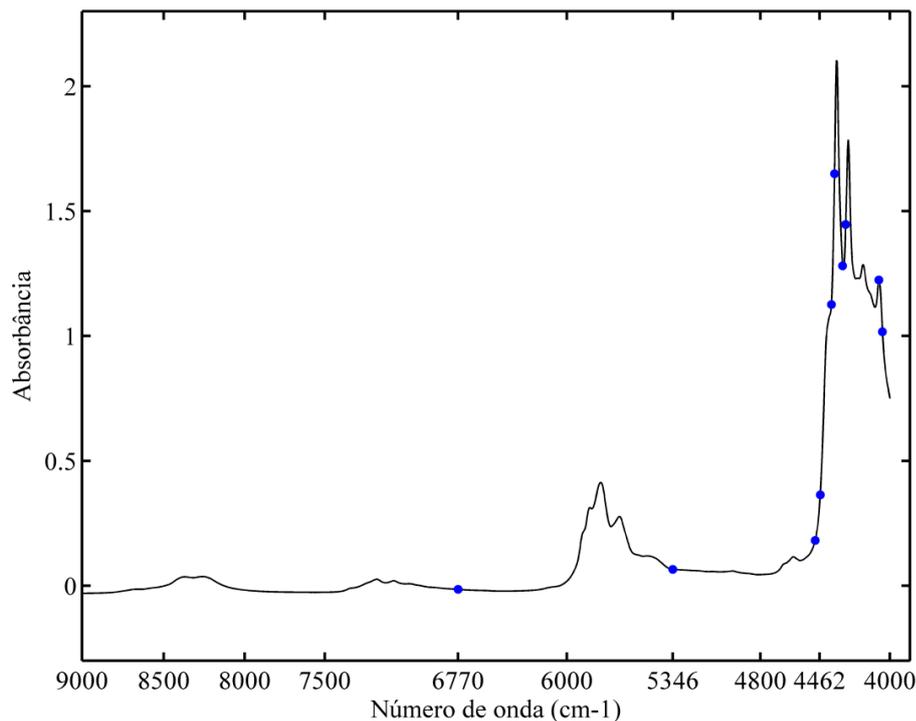
**Figura 5.2** – Gráfico dos escores da PC2  $\times$  PC1 para as 150 amostras do conjunto de dados nos caminhos ópticos: (a) 1,0 mm, (b) 10 mm e (c) 50 mm. (\*: D,  $\bullet$ : DO,  $\blacktriangle$ : DBO e  $\blacksquare$ : DB). O percentual da variância explicada por cada PC está indicado entre parênteses.

### 5.2.2 Modelos LDA/SPA

Os modelos LDA/SPA foram construídos utilizando os subconjuntos de amostras apresentados na **Tabela 4.2**. Como resultado, 10, 8 e 5 variáveis foram selecionadas pelo SPA para os caminhos ópticos de 1,0 mm, 10 mm e 50 mm, respectivamente.

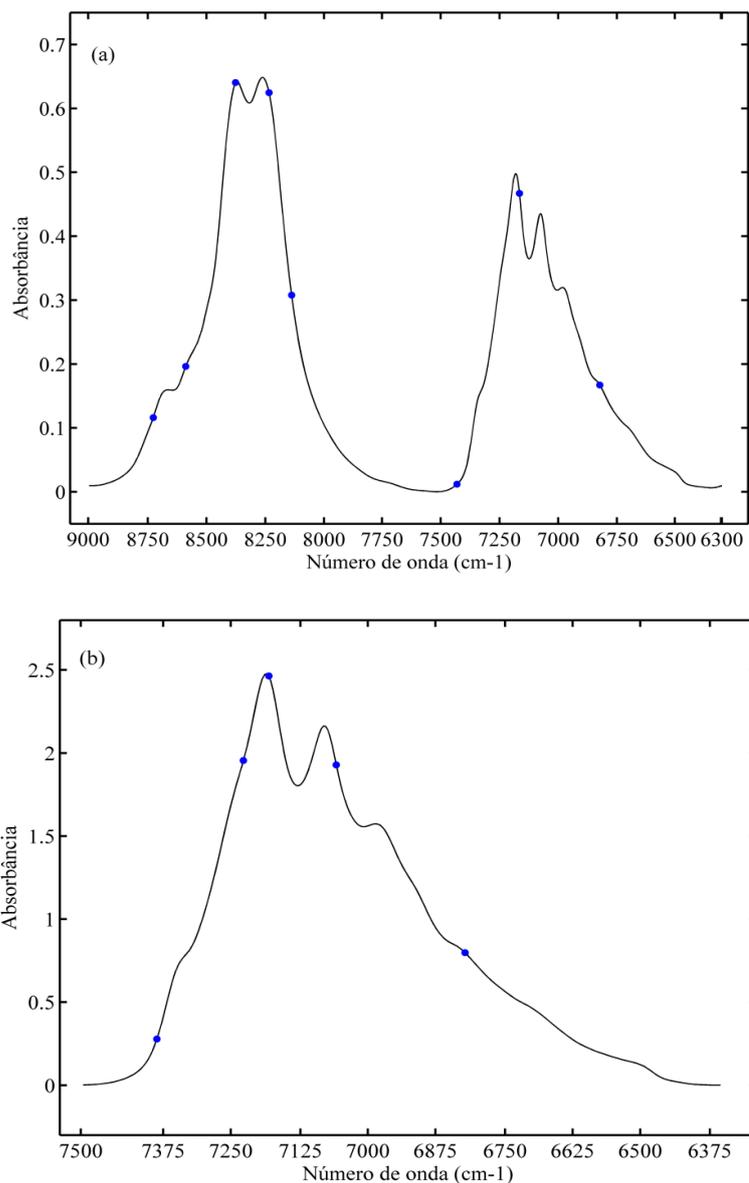
A **Figura 5.3** mostra o espectro NIR médio das amostras do conjunto de treinamento, com a indicação das dez variáveis selecionadas pelo SPA para a faixa de 9000 a 4000  $\text{cm}^{-1}$  (caminho óptico de 1,0 mm). Apesar do modelo LDA/SPA para essa região ter sido obtido utilizando-se os espectros derivativos, optou-se por indicar os números de onda selecionados no espectro original (sem pré-processamento) para facilitar a atribuição das bandas.

Pode-se observar que a maioria das variáveis selecionadas pelo SPA está associada às bandas de combinação dos modos de estiramento de ligações C-H (4800-4000  $\text{cm}^{-1}$ ) e às bandas de combinação dos modos de estiramento de ligações C-H com ligações C = O (4740-4370  $\text{cm}^{-1}$ ). As variáveis selecionadas em 5346  $\text{cm}^{-1}$  e 6770  $\text{cm}^{-1}$  podem ser associadas ao primeiro sobretom dos modos de estiramento de ligações C-H e ao primeiro sobretom das combinações de ligações C - H, respectivamente (WORKMAN, 2008).



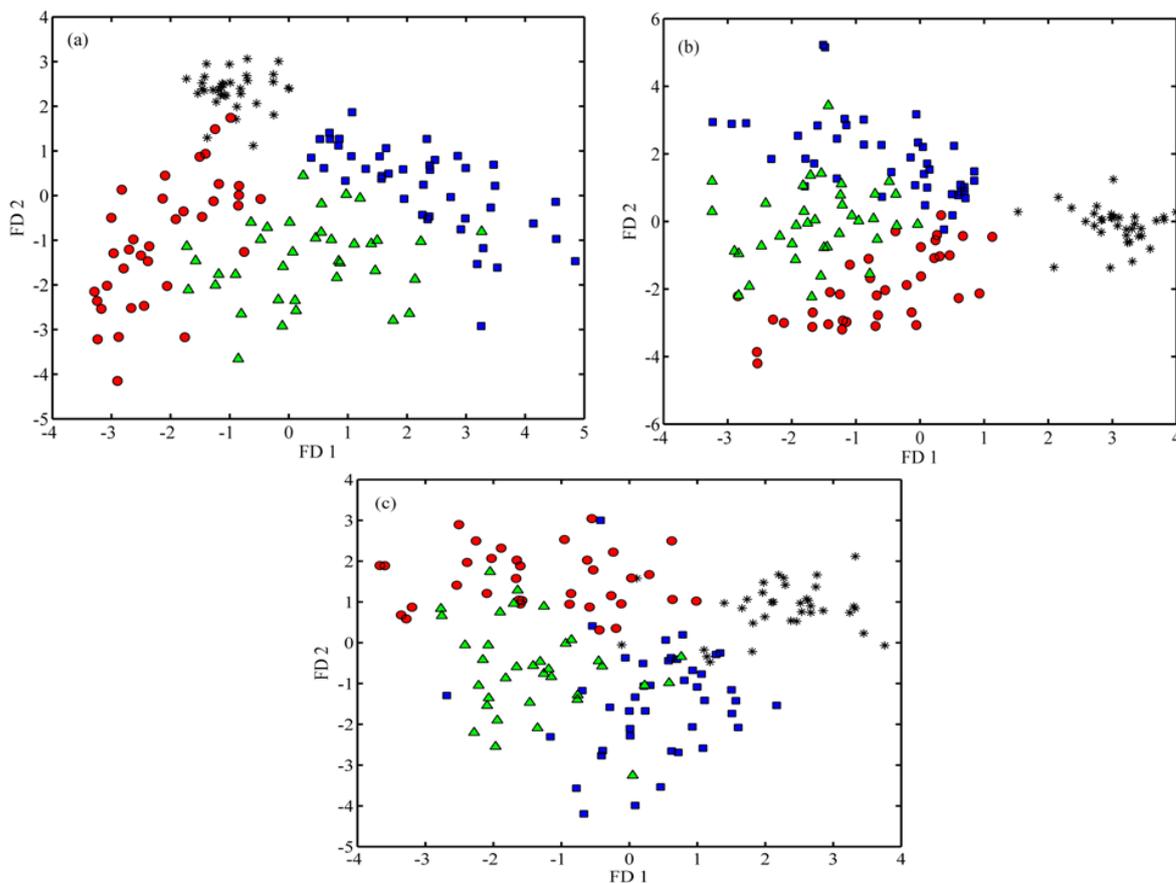
**Figura 5.3** - Variáveis espectrais selecionadas pelo SPA na construção dos modelos de classificação LDA utilizando-se o caminho óptico de 1,0 mm.

Para o caminho óptico de 10 mm, as variáveis selecionadas pelo SPA (**Figura 5.4a**) encontram-se distribuídas ao longo de toda faixa espectral (9000-6300  $\text{cm}^{-1}$ ), na qual estão incluídas as bandas associadas ao primeiro sobretom das combinações (7400-6900  $\text{cm}^{-1}$ ) e ao segundo sobretom (8910-8000  $\text{cm}^{-1}$ ) dos modos de estiramento das ligações C-H. As variáveis selecionadas pelo SPA para o caminho óptico de 50 mm (**Figura 5.4b**), também estão distribuídas ao longo da faixa de trabalho (7500-6300  $\text{cm}^{-1}$ ) correspondente ao primeiro sobretom das combinações (WORKMAN, 2008). Nesses casos, são mostrados os espectros médios pré-processados das amostras do conjunto de treinamento, visto que a técnica de pré-processamento utilizada (correção de linha de base com suavização por filtro média móvel e janela 5 pontos) alterou pouco o perfil original dos espectros.



**Figura 5.4** – Variáveis espectrais selecionadas pelo SPA na construção dos modelos de classificação LDA utilizando-se os caminhos ópticos de (a) 10 mm e (b) 50 mm.

As **Figuras 5.5a-c** mostram os gráficos dos escores das duas primeiras funções discriminantes (FD2 × FD1) para os três caminhos ópticos analisados 1,0, 10 e 50 mm, respectivamente. Como pode ser observado, houve uma melhor separação das classes quando comparado ao gráfico dos escores obtidos pela PCA (**Figuras 5.2a-c**).



**Figura 5.5** - Gráfico dos escores da FD2×FD1 para as 150 amostras do conjunto de dados nos caminhos ópticos: (a) 1,0 mm, (b) 10 mm e (c) 50 mm. (\*: D, ●: D0, ▲: DBO e ■: DB).

Os resultados obtidos pela LDA revelam uma discriminação mais efetiva das classes para o caminho óptico de 1,0 mm (**Figura 5.5a**), o qual foi obtido utilizando-se as variáveis selecionadas pelo SPA associadas principalmente à região das combinações (**Figura 5.3**), incluindo-se a região associada às bandas de combinação dos modos de estiramento de ligações C-H com ligações C = O ( $4740\text{-}4370\text{ cm}^{-1}$ ).

Apesar da região de sobretoms apresentar informação bastante sobreposta, os resultados obtidos pela LDA (**Figuras 5.5b-c**) demonstram que é possível separar misturas contendo diesel/biodiesel das misturas contendo diesel/óleo e diesel puro. Uma melhor discriminação das classes foi obtida para o caminho óptico de 10 mm ( $9000\text{-}6300\text{ cm}^{-1}$ ), que inclui às regiões espectrais de primeiro sobretom das combinações ( $7400\text{-}6900\text{ cm}^{-1}$ ) e segundo sobretom ( $8910\text{-}8000\text{ cm}^{-1}$ ) dos modos de estiramento das ligações C-H, em relação ao caminho óptico de 50 mm

(7500-6300  $\text{cm}^{-1}$ ), que corresponde apenas à região de primeiro sobretom das combinações (7400-6900  $\text{cm}^{-1}$ ).

Os modelos LDA obtidos com as variáveis selecionadas pelo SPA (**Figuras 5.5a-c**) foram aplicados à classificação das amostras do conjunto de teste. Na **Tabela 5.1** são apresentados os erros de classificação (em vermelho) obtidos pelos modelos LDA/SPA para as amostras do conjunto de teste ( $N^*$ ) utilizando-se os caminhos ópticos de 1,0, 10 e 50 mm.

**Tabela 5.1** – Erros de classificação obtidos pelos modelos LDA/SPA para as amostras do conjunto de teste utilizando-se os caminhos ópticos de 1,0, 10 e 50 mm.

Classe	$N^*$	1 mm $N^{**} = 10$				10 mm $N^{**} = 8$				50 mm $N^{**} = 5$			
		D	DB	DO	DBO	D	DB	DO	DBO	D	DB	DO	DBO
<b>D</b>	8	-	-	-	-	-	-	-	-	-	-	-	-
<b>DB</b>	10	-	-	-	-	-	-	-	<b>2</b>	-	-	<b>1</b>	<b>1</b>
<b>DO</b>	8	<b>2</b>	-	-	-	-	-	-	-	<b>1</b>	-	-	-
<b>DBO</b>	9	-	<b>1</b>	<b>1</b>	-	-	<b>2</b>	<b>1</b>	-	-	<b>1</b>	<b>1</b>	-

\*Número de amostras do conjunto de teste para cada classe.

\*\*Número de variáveis espectrais selecionadas pelo LDA/SPA para cada um dos caminhos ópticos.

Como resultado da aplicação do modelo para o caminho óptico de 1,0 mm (**Figura 5.5a**), quatro amostras do conjunto de teste ( $N^*=35$ ) foram classificadas incorretamente, dentre as quais: duas amostras de diesel/óleo (DO) classificadas como diesel puro (D), duas amostras de DBO classificadas uma como DB e outra como DO, totalizando um índice de acerto de 88%.

Para o caminho óptico de 10 mm, o modelo LDA/SPA (**Figura 5.5b**) classificou incorretamente cinco das trinta e cinco amostras do conjunto de teste, totalizando um índice de acerto de 86%. Das amostras classificadas incorretamente, tem-se: três amostras de DBO (duas para a classe DB e uma para a classe DO). Outras duas amostras de DB classificadas como pertencentes à classe DBO, mesmo não apresentando óleo vegetal em sua composição. Isso se deve ao fato de que no modelo

LDA cada amostra classificada incorretamente corresponde a um erro tipo I e a um erro tipo II. Isto é, se uma amostra não for incluída em sua classe verdadeira, a mesma será incluída em uma classe incorreta mais próxima, nesse caso DBO.

Já para o caminho óptico de 50 mm o modelo LDA/SPA (**Figura 5.5c**) classificou incorretamente: duas amostras de DB (uma para a classe DO e uma para a classe DBO), duas amostras de DBO (uma para a classe DB e uma para a classe DO), e uma amostra de DO foi classificada como D. Esse modelo apresentou um índice de acerto para o conjunto de teste semelhante ao modelo LDA/SPA para o caminho óptico de 10 mm, que foi de 86%.

Em geral, as amostras de diesel/biodiesel adulteradas com óleo vegetal (DBO) que foram classificadas incorretamente como DB pelos modelos LDA/SPA para os caminhos ópticos de: 1,0 mm (1 amostra), 10 mm (2 amostras) e 50 mm (1 amostra) apresentaram concentração de biodiesel de 6% e 8% v/v e uma concentração de óleo de 2% v/v. Segundo a legislação vigente (RESOLUÇÃO CNPE N° 6, 2009) essas amostras seriam consideradas não conformes, visto que apresentam teor de biodiesel acima de 5% v/v.

Os resultados obtidos indicam que a região de sobretons associada à estratégia LDA/SPA pode ser utilizada em estudos preliminares para detecção de adulteração de misturas diesel/biodiesel por óleos vegetais (*in natura*), a qual é uma técnica simples e rápida. Diante desses resultados, partiu-se então para determinação do teor de biodiesel em misturas diesel/biodiesel na presença de óleos vegetais utilizando modelos de calibração multivariada para faixas espectrais do NIR selecionadas na **Seção 5.1.**

### 5.3 Calibração multivariada

#### 5.3.1 Caminho óptico de 1,0 mm (9000-4000 $\text{cm}^{-1}$ )

A **Tabela 5.2** mostra os resultados obtidos para os modelos de calibração multivariada PLS, PLS/JK e MLR/SPA construídos utilizando-se 83 amostras de calibração e 45 amostras de validação externa a partir de seus respectivos espectros NIR (9000-4000  $\text{cm}^{-1}$ ) empregando-se diferentes estratégias de pré-processamento.

**Tabela 5.2** – Resultados obtidos para os modelos de calibração multivariada PLS, PLS-JK e MLR/SPA (caminho óptico: 1,0 mm) utilizando-se diferentes estratégias de pré-processamento dos espectros.

Modelo	Pré-processamento	R	RMSECV (% v/v)	N*	R	RMSEP (% v/v)
PLS	Sem	0,98	0,38	6	0,98	0,28
	LB / SMM (janela 5 pontos)	0,99	0,40	5	0,99	0,24
	LB / SSG (janela 11 pontos)	0,99	0,39	5	0,99	0,23
	1° derivada SG (janela 11 pontos)	0,99	0,40	4	0,99	0,31
PLS/JK	Sem	0,98	0,43	5	0,98	0,40
	LB / SMM (janela 5 pontos)	0,99	0,39	5	0,99	0,24
	LB / SSG (janela 11 pontos)	0,99	0,39	5	0,99	0,24
	1° derivada SG (janela 11 pontos)	0,99	0,27	4	0,99	0,23
MLR/SPA	Sem	0,99	0,07	13	0,99	0,09
	LB / SMM (janela 5 pontos)	0,99	0,06	13	0,99	0,12
	LB / SSG (janela 11 pontos)	0,99	0,06	15	0,99	0,09
	1° derivada SG (janela 11 pontos)	0,99	0,07	16	0,99	0,08

\*número de variáveis latentes (PLS) ou número de variáveis espectrais (MLR/SPA).

LB = Correção de linha de base; SMM = Suavização média móvel; SSG = Suavização Savitsky-Golay.

Para os modelos PLS o menor RMSEP (0,23% v/v) foi obtido a partir dos espectros NIR pré-processados com correção de linha de base e

suavização por polinômio móvel e filtro de SG (janela de 11 pontos). Comparando-se esse resultado com as demais estratégias de pré-processamento utilizando-se um teste  $F$  a 95% de confiança (**Tabela 5.3**), observa-se que não há diferença estatisticamente significativa entre essa estratégia e a correção de linha de base com suavização por filtro média móvel (janela de 5 pontos), cujo RMSEP é igual a 0,24% v/v.

**Tabela 5.3** – Resultados para o teste  $F$  (nível de confiança de 95%) obtidos na comparação das diferentes técnicas de pré-processamento empregadas na construção dos modelos (PLS, PLS/JK e MLR/SPA) para o caminho óptico de 1,0 mm.

Modelo	Menor RMSEP	$F_{cal}$		
		LB / SMM (janela 5 pts)	LB / SSG (janela 11pts)	1° derivada SG (janela 11pts)
PLS	LB / SSG (janela 5 pts)	1,09	-	1,82
PLS/JK	1° derivada SG (janela 11 pts)	1,09	1,09	-
MLR/SPA	1° derivada SG (janela 11 pts)	2,25	1,26	-

\* $F_{crit}(0,95; 44; 44) = 1,65$

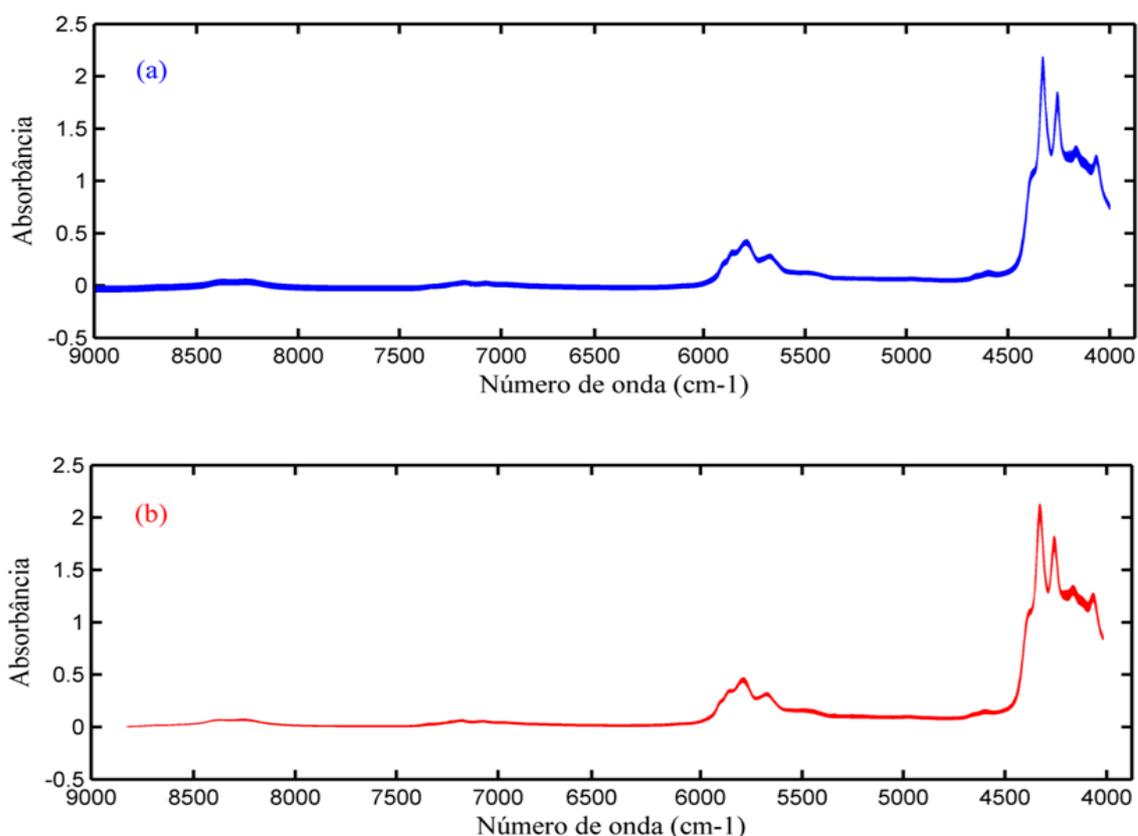
Pode-se observar também que o modelo PLS construído utilizando os espectros derivativos apresentou um maior valor do RMSEP (0,31% v/v), sendo esse estatisticamente diferente dos demais. Nesse caso, poderia-se optar pelas estratégias de pré-processamento: (1) LB/SMM (janela 5 pontos) ou (2) LB/SSG (janela 11 pontos).

Já para o modelo PLS/JK o menor RMSEP (0,23% v/v) foi obtido a partir dos espectros derivativos. Comparando-se esse valor com as demais estratégias de pré-processamento (**Tabela 5.3**) pode-se observar que não há diferença estatisticamente significativa entre os valores dos RMSEP's. Os resultados obtidos para o modelo PLS/JK são comparáveis aos descritos por PIMENTEL et al (2006), para determinação do teor de biodiesel em misturas com diesel adulteradas com óleo vegetal (0-5% v/v), a qual utilizou a faixa espectral do NIR de 12000-4000  $\text{cm}^{-1}$ .

Os modelos MLR/SPA apresentaram desempenho semelhante para as técnicas: (2) correção de linha de base com suavização por polinômio

móvel e filtro SG (janela 11 pontos) e (3) primeira derivada por polinômio de segunda ordem com filtro de SG (janela de 11 pontos). É importante citar que o modelo MLR/SPA sem pré-processamento apresentou um desempenho semelhante ( $RMSEP=0,09\%$  v/v) a técnica LB/SSG (janela de 11 pontos). Entretanto, optou-se por utilizar os dados pré-processados, visto que o SPA selecionou variáveis relacionadas a variações de linha de base para os dados originais que não estão relacionadas ao problema químico de interesse.

Desse modo, a correção de linha de base com polinômio móvel e filtro Savitzky-Golay filtro SG (janela de 11 pontos) pode ser considerado o pré-processamento mais adequado em relação à primeira derivada, visto que não há mudanças no perfil dos espectros pré-processados em relação aos espectros originais (**Figura 5.6**).



**Figura 5.6** – Espectros NIR das 128 amostras: (a) originais e (b) pré-processados por correção de linha de base e polinômio de SG com janela de 11 pontos (LB/SSG).

Avaliando-se então o desempenho dos modelos PLS, PLS/JK e MLR/SPA utilizando-se a estratégia de pré-processamento escolhida

(**Tabela 5.2**), observou-se que o modelo MLR/SPA apresentou o menor valor de RMSEP (0,09% v/v), o qual é estatisticamente diferente dos valores obtidos pelo modelo PLS (com todas as variáveis espectrais) e pelo modelo PLS construído com as variáveis selecionadas pelo algoritmo JK (**Tabela 5.4**). Esse resultado é inferior ao RMSEP estabelecido pela norma ABNT 15568:2008 para o teor de biodiesel em misturas com diesel nas faixas de concentração de 0% a 8% v/v de biodiesel (RMSEP = 0,10% v/v) e de 8% a 30% v/v (RMSEP = 1,0 % v/v) utilizando espectros MIR. Vale salientar, que os modelos desenvolvidos no presente estudo utilizaram concentrações de biodiesel e óleo vegetal variando de 0,0 a 10% v/v.

Os resultados obtidos apontam o MLR/SPA como sendo a melhor estratégia para determinação do teor de biodiesel em misturas com diesel adulteradas com óleo vegetal *in natura* para o caminho óptico de 1,0 mm.

**Tabela 5.4** – Resultados do teste *F* (nível de 95% de confiança) comparando-se os modelos de calibração (PLS, PLS/JK e MLR/SPA) construídos utilizando-se os espectros NIR (caminho óptico = 1,0 mm) pré-processados com correção de linha de base e polinômio móvel com filtro SG (janela de 11 pontos).

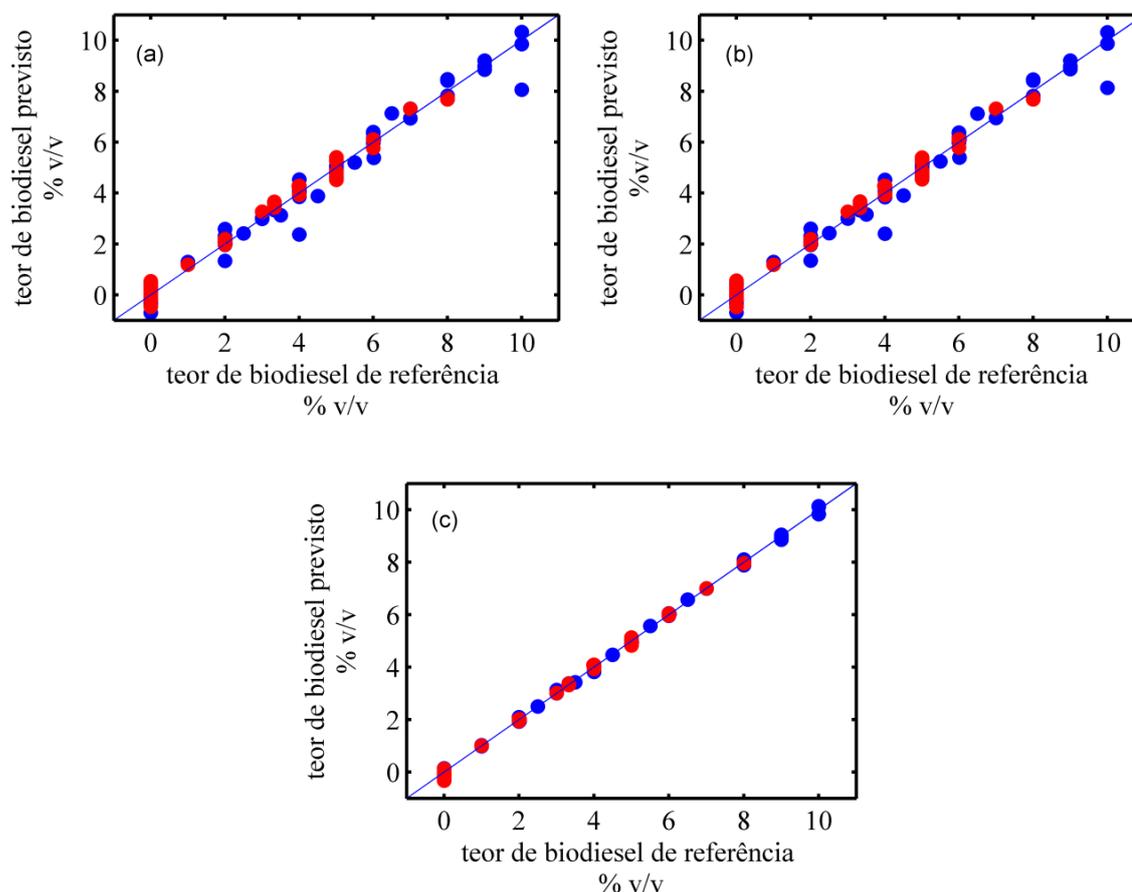
MLR/SPA × PLS	MLR/SPA × PLS/JK
6,53	7,11

$F_{crit}(44; 44; 0,95) = 1,65$

As **Figuras 5.7a-c** mostram os gráficos dos valores previstos *versus* os valores de referência para os conjuntos de calibração (azul) e validação externa (vermelho) obtidos pelos modelos PLS, PLS/JK e MLR/SPA respectivamente, para a determinação do teor de biodiesel utilizando o caminho óptico de 1,0 mm.

Como pode ser observado (**Figura 5.7**), não há presença de erro sistemático nos modelos (calibração e previsão), visto que, os valores encontram-se aleatoriamente distribuídos ao longo da linha bissetriz. Além disso, observa-se que o modelo MLR/SPA (**Figura 5.7c**) apresentou uma menor dispersão dos dados em relação à bissetriz conduzindo aos melhores resultados de previsão. No caso dos modelos PLS e PLS/JK

(**Figuras 5.7a-b**), esses apresentaram um comportamento bastante semelhante. Essas observações estão de acordo com os resultados obtidos para o RMSECV e RMSEP, mostrados na **Tabela 5.2**.

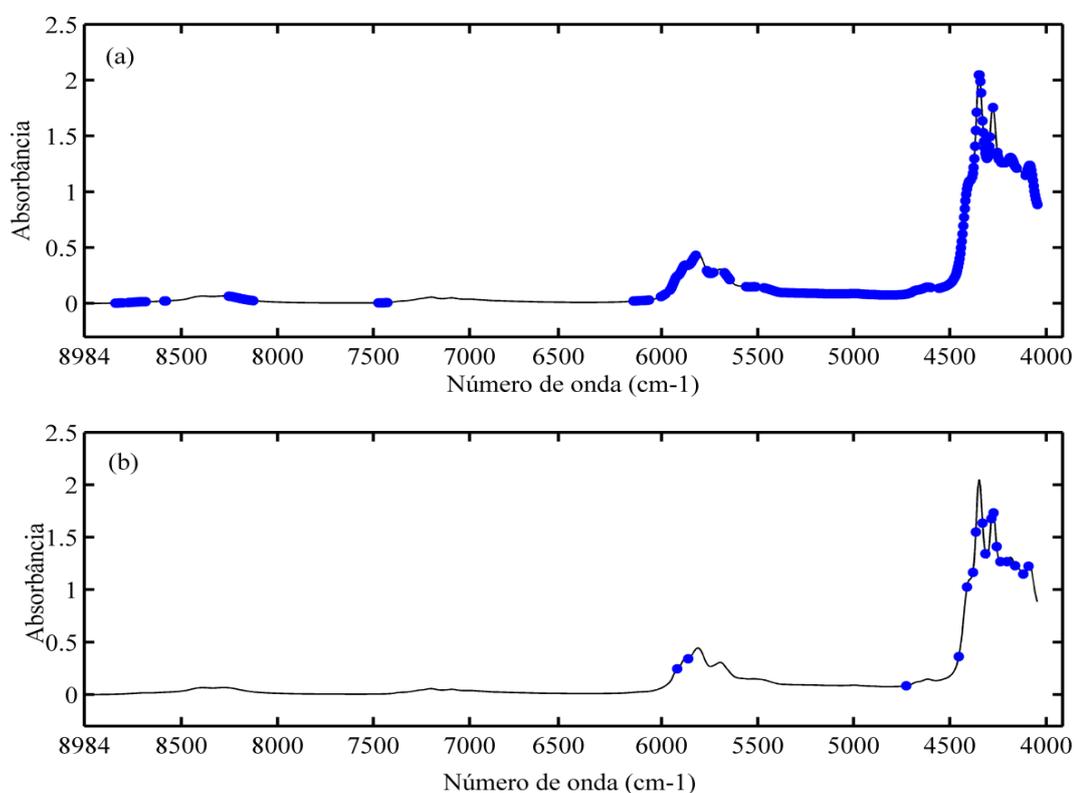


**Figura 5.7** – Gráfico dos valores previstos para teor de biodiesel versus valores de referência para os modelos: (a) PLS; (b) PLS/JK e (c) MLR/SPA utilizando o caminho óptico de 1,0 mm (●: amostras de calibração e ●: validação externa).

As variáveis espectrais selecionadas pelos algoritmos de Jack-Knife e SPA utilizadas na construção dos modelos PLS (**Figura 5.7b**) e MLR (**Figura 5.7c**) são mostradas nas **Figuras 5.8a e b**, respectivamente.

Em geral, as variáveis selecionadas para determinação do teor de biodiesel (caminho óptico de 1,0 mm) estão associadas às bandas de combinação do estiramento das ligações C-H ( $4800-4000\text{ cm}^{-1}$ ), incluindo-se a região de  $4740-4370\text{ cm}^{-1}$  (combinação de estiramento das ligações C-H e C=O). Além dessas, foram selecionadas variáveis na região de  $6300-5180\text{ cm}^{-1}$  associadas ao primeiro sobretom do estiramento de ligações C-H.

O PLS/JK utilizou também, variáveis selecionadas nas regiões de  $7500\text{ cm}^{-1}$  e  $8910\text{-}8000\text{ cm}^{-1}$  associadas ao primeiro sobretom das combinações e segundo sobretom do estiramento de ligações C-H, respectivamente (**Figuras 5.8a**). Resultados semelhantes foram obtidos por PIMENTEL et al, 2006. Como pode ser observado a estratégia MLR/SPA utilizou um número menor de variáveis para determinação do teor de biodiesel em relação ao PLS/JK (**Figura 5.8b**), sendo, portando um modelo mais parcimonioso.



**Figura 5.8** – Variáveis espectrais selecionadas pelas estratégias: (a) PLS/JK e (b) MLR/SPA para determinação do teor de biodiesel utilizando o caminho óptico de 1,0 mm.

### 5.3.2 Caminho óptico de 10 mm ( $9000\text{-}6300\text{ cm}^{-1}$ )

A **Tabela 5.5** mostra os resultados obtidos para os modelos de calibração multivariada PLS, PLS-JK e MLR/SPA construídos a partir de 84 amostras de calibração e 45 de validação externa para determinação do teor de biodiesel utilizando-se as diferentes estratégias de pré-processamento dos espectros NIR (caminho óptico de 10 mm).

A faixa espectral utilizada nesse estudo ( $9000\text{-}6300\text{ cm}^{-1}$ ) inclui as regiões do primeiro sobretom das combinações ( $7400\text{-}6900\text{ cm}^{-1}$ ) e segundo sobretom ( $8910\text{-}8000\text{ cm}^{-1}$ ) dos modos de estiramento de ligações C-H.

Para os modelos PLS o menor valor do RMSEP (0,41% v/v) foi obtido utilizando-se os espectros NIR derivativos. De acordo com o teste *F* (**Tabela 5.6**), esse valor é semelhante aos resultados obtidos para as demais estratégias de pré-processamento. Isso indica que o desempenho do modelo independe da estratégia de pré-processamento empregada.

**Tabela 5.5** – Resultados obtidos para os modelos de calibração multivariada PLS, PLS-JK e MLR/SPA (caminho óptico: 10 mm) utilizando-se diferentes estratégias de pré-processamento dos espectros.

Modelo	Pré-processamento	R	RMSECV (%v/v)	N*	R	RMSEP (%v/v)
PLS	Sem	0,99	0,38	9	0,99	0,45
	LB / SMM (janela 5 pontos)	0,99	0,44	7	0,99	0,44
	LB / SSG (janela 5 pontos)	0,99	0,41	7	0,99	0,42
	1° derivada SG (janela 7 pontos)	0,99	0,39	7	0,99	0,41
PLS/JK	Sem	0,98	0,42	8	0,98	0,48
	LB / SMM (janela 5 pontos)	0,99	0,37	7	0,99	0,38
	LB / SSG (janela 5 pontos)	0,99	0,36	7	0,99	0,39
	1° derivada SG (janela 7 pontos)	0,99	0,34	6	0,99	0,47
MLR/SPA	Sem	0,99	0,10	12	0,99	0,11
	LB / SMM (janela 5 pontos)	0,99	0,08	14	0,99	0,07
	LB / SSG (janela 5 pontos)	0,99	0,09	12	0,99	0,10
	1° derivada SG (janela 7 pontos)	0,99	0,26	13	0,99	0,36

\*número de variáveis latentes (PLS) ou número de variáveis espectrais (MLR/SPA).

LB = Correção de linha de base; SMM = Suavização média móvel; SSG = Suavização Savitsky-Golay.

Já para os modelos PLS/JK e MLR/SPA os menores RMSEP's (0,38% v/v e 0,07% v/v, respectivamente) foram obtidos utilizando-se os espectros NIR pré-processados com correção linear de linha de base e suavização por filtro média móvel (janela 5 pontos). De acordo com o teste  $F$  (**Tabela 5.6**), o RMSEP obtido para o modelo PLS/JK é semelhante aos resultados obtidos para as demais estratégias de pré-processamento, demonstrando que o desempenho desse modelo, assim como o PLS, independe da técnica de pré-processamento empregada.

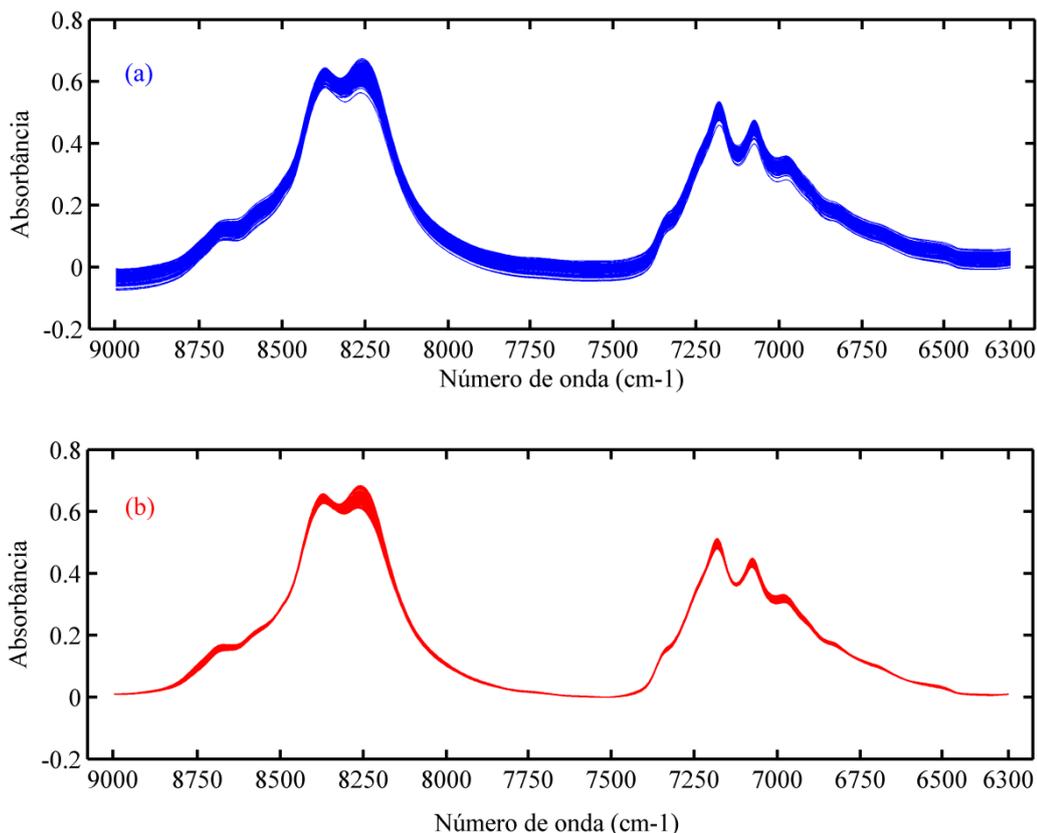
Os modelos MLR/SPA apresentaram valores de RMSEP estatisticamente diferentes para todas as técnicas de pré-processamento analisadas a 95% de confiança (**Tabela 5.6**). Nesse caso, a correção de linha de base com suavização por média móvel (janela de 5 pontos) pode ser considerada a melhor estratégia de pré-processamento, pois apresentou o menor RMSEP (0,07%v/v) para os modelos MLR/SPA. Além disso, é uma técnica simples, não havendo mudanças no perfil dos espectros pré-processados em relação aos espectros originais (**Figura 5.9**).

Diante dos resultados obtidos, escolheu-se a estratégia de pré-processamento correção linear de linha de base com suavização por filtro média móvel (janela de 5 pontos) para comparar o desempenho dos modelos de calibração multivariada desenvolvidos.

**Tabela 5.6**– Resultados para o teste  $F$  (nível de confiança de 95%) obtidos na comparação das diferentes estratégias de pré-processamento empregadas na construção dos modelos (PLS, PLS/JK e MLR/SPA) para o caminho óptico de 10 mm.

Modelo	Menor RMSEP	$F_{cal}$		
		LB / SMM (janela 5pts)	LB / SSG (janela 5pts)	1° derivada SG (janela 7pts)
PLS	1° derivada SG (janela 7 pts)	1,51	1,05	-
PLS/JK	LB / SMM (janela 5pts)	-	1,05	1,52
MLR/SPA	LB / SMM (janela 5pts)	-	2,04	26,45

$F_{crit}(0,95; 44; 44) = 1,65$



**Figura 5.9** - Espectros NIR das 129 amostras: (a) originais e (b) pré-processados com correção de linha de base e suavização por média móvel com janela de 5 pontos (LB/SMM).

Analisando-se os resultados obtidos pelos modelos PLS, PLS/JK e MLR/SPA, utilizando-se a estratégia de pré-processamento escolhida, pode-se observar que o menor valor de RMSEP (0,07% v/v) foi obtido com a estratégia MLR/SPA (**Tabela 5.5**). Esse resultado é inferior ao valor estabelecido pela norma ABNT 15568:2008 (RMSEP = 0,10 % v/v).

Os modelos PLS e PLS/JK apresentaram valores de RMSEP (0,44% v/v e 0,38% v/v, respectivamente) bem acima do valor estabelecido pela norma ABNT 15568:2008. O teste *F* revelou que esses dois modelos fornecem resultados estatisticamente diferentes do modelo MLR/SPA (**Tabela 5.7**).

O modelo MLR/SPA para o caminho óptico de 10 mm (9000-6300  $\text{cm}^{-1}$ ) apresentou o valor de RMSEP comparável ao modelo PLS descrito por OLIVEIRA et al (2006) para determinação do teor de biodiesel em misturas com óleo diesel (0-100% v/v). Entretanto, esse último utilizou a faixa espectral do NIR de 9000-4700  $\text{cm}^{-1}$  (caminho óptico = 2,0

mm) que abrange a região associada às bandas de estiramento das ligações C-H e C=O ( $4670\text{-}4700\text{ cm}^{-1}$ ) e não considera a presença de óleo vegetal nas misturas diesel/biodiesel. Devido a esses fatores, pode-se considerar que o modelo desenvolvido no presente trabalho (MLR/SPA) apresentou melhores resultados para determinação do teor de biodiesel.

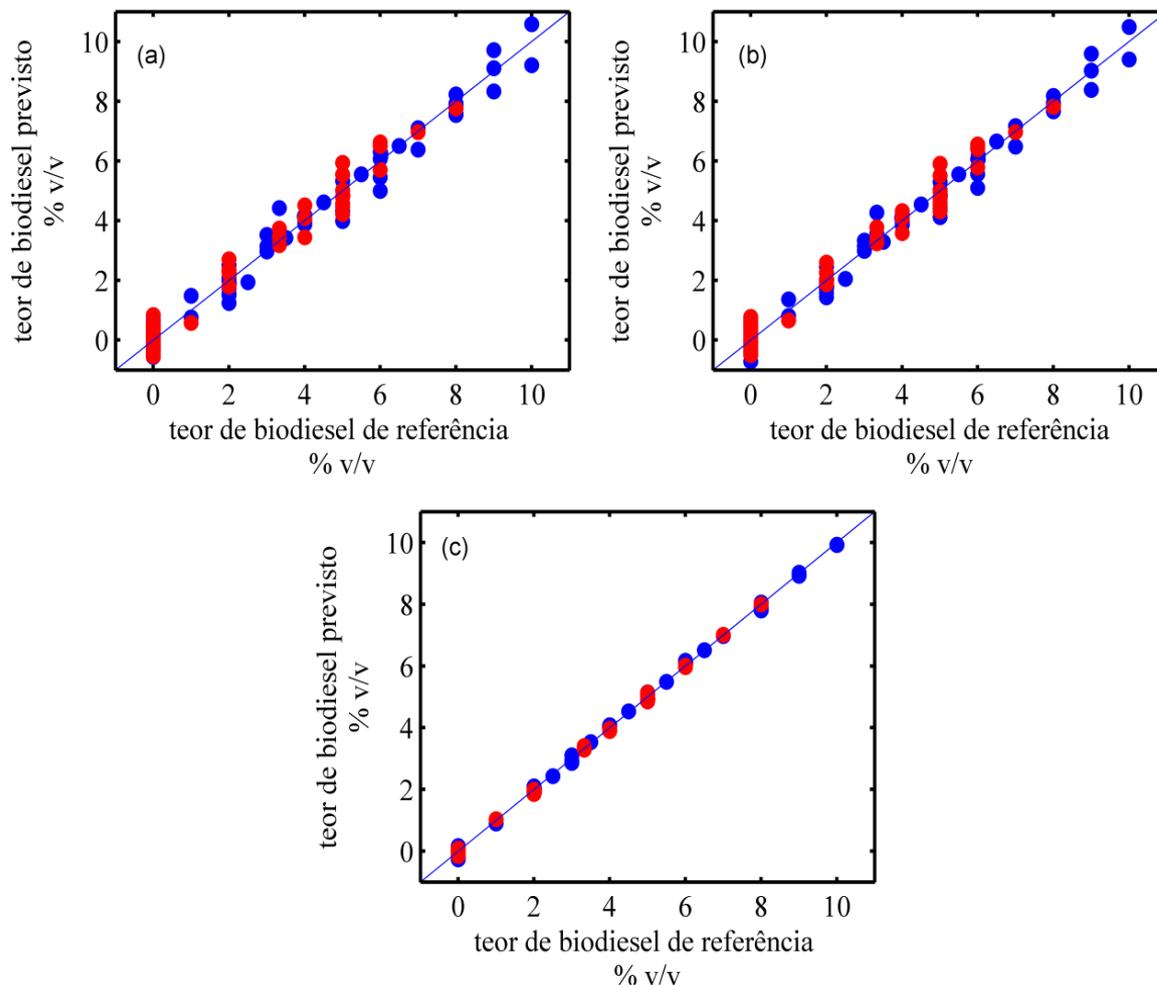
**Tabela 5.7**– Resultados do teste  $F$  (nível de 95% de confiança) comparando-se os modelos PLS, PLS/JK e MLR/SPA construídos utilizando-se os espectros NIR (caminho óptico = 10 mm) pré-processados com correção de linha de base e suavização por filtro média móvel (janela de 5 pontos).

MLR/SPA × PLS	MLR/SPA × PLS/JK
39,51	29,47

$$F_{\text{crit}}(44; 44; 0,95) = 1,65$$

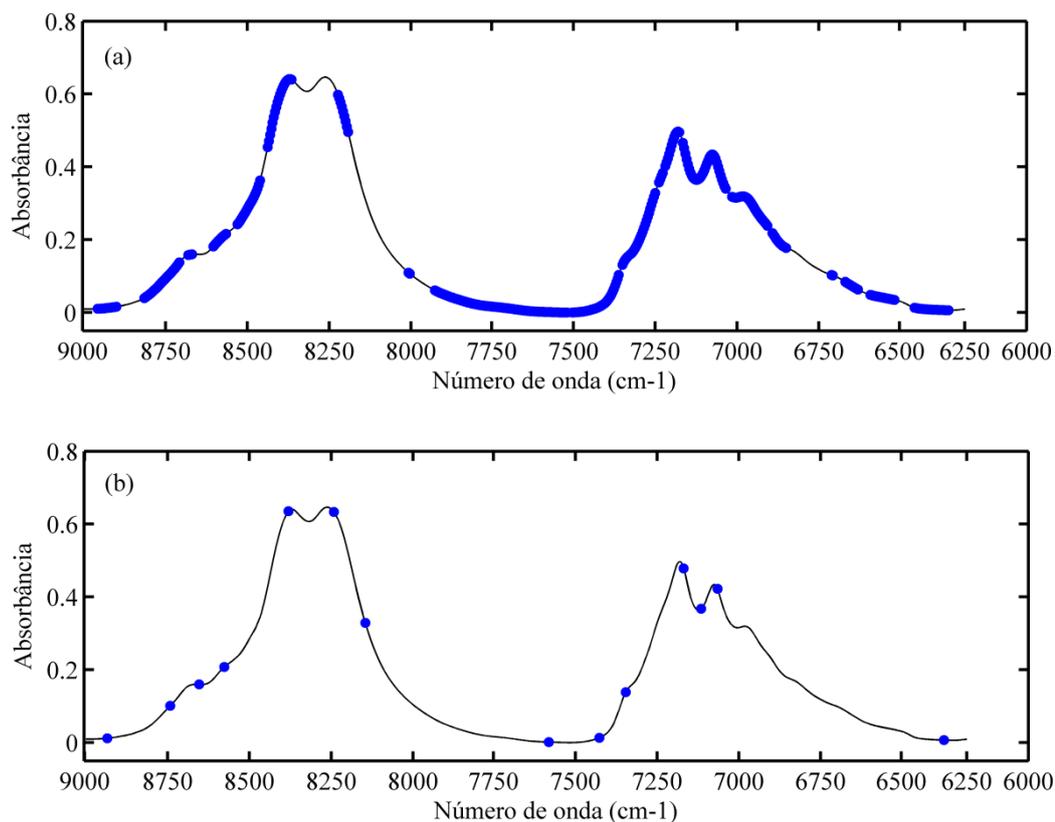
Nas **Figuras 5.10a-c** são mostrados os gráficos dos valores previstos *versus* os valores de referência para os conjuntos de calibração (azul) e validação externa (vermelho) obtidos pelos modelos PLS, PLS/JK e MLR/SPA para o caminho óptico de 10 mm.

Como pode ser observado o modelo MLR/SPA (**Figura 5.10c**) apresentou uma menor dispersão dos dados em relação à linha bissetriz, indicando um melhor desempenho tanto na etapa de calibração quanto na etapa de validação. Além disso, pode-se observar que os modelos PLS e PLS/JK (**Figuras 5.10a-b**) apresentaram uma dispersão dos dados bastante semelhantes para os conjuntos de calibração e validação externa. Essas observações mostram a concordância entre os erros obtidos nas etapas de calibração e validação externa (**Tabela 5.5**).



**Figura 5.10** – Gráfico dos valores previstos para teor de biodiesel versus valores de referência para os modelos: (a) PLS; (b) PLS/JK e (c) MLR/SPA utilizando o caminho óptico de 10 mm. (●: amostras de calibração e ●: validação externa).

As **Figuras 5.11a-b** mostram as variáveis espectrais selecionadas pelos algoritmos JK e SPA utilizadas na construção dos modelos PLS (**Figura 5.10b**) e MLR (**Figura 5.10c**) para determinação do teor de biodiesel utilizando caminho óptico de 10 mm.



**Figura 5.11**– Variáveis espectrais selecionadas pelas estratégias: (a) PLS/JK e (b) MLR/SPA para determinação do teor de biodiesel utilizando o caminho óptico de 10 mm.

Como pode ser observado na **Figura 5.11**, as variáveis selecionadas para determinação do teor de biodiesel encontram-se distribuídas ao longo de toda a faixa espectral ( $9000\text{-}6300\text{ cm}^{-1}$ ) que inclui as regiões de primeiro sobretom das combinações ( $7400\text{-}6900\text{ cm}^{-1}$ ) e segundo sobretom ( $8910\text{-}8000\text{ cm}^{-1}$ ) dos modos de estiramento de ligações C-H. O modelo PLS/JK utilizou um número muito maior de variáveis (**Figura 5.11a**) para determinação do teor de biodiesel em relação ao modelo MLR/SPA (**Figura 5.11b**), indicando que esse último é mais parcimonioso.

### 5.3.3 Caminho óptico de 50 mm ( $7500\text{-}6300\text{ cm}^{-1}$ )

Os resultados obtidos dos modelos PLS, PLS-JK e MLR/SPA para determinação do teor de biodiesel utilizando-se a faixa espectral de  $7500\text{-}6300\text{ cm}^{-1}$  (caminho óptico de 50 mm) a partir de 84 amostras de calibração e 45 de validação para os diferentes pré-processamentos são

mostrados na **Tabela 5.8**. A faixa espectral avaliada inclui apenas a região de primeiro sobretom das combinações do estiramento C-H (7400-6900  $\text{cm}^{-1}$ ).

**Tabela 5.8** – Resultados obtidos para os modelos de calibração multivariada PLS, PLS-JK e MLR/SPA (caminho óptico: 50 mm) utilizando-se diferentes estratégias de pré-processamento dos espectros.

Modelo	Pré-processamento	R	RMSECV (%v/v)	N	R	RMSEP (%v/v)
PLS	Sem	0,99	0,40	7	0,98	0,35
	LB / SMM (janela 5pts)	0,99	0,40	5	0,98	0,33
	LB / SSG (janela 5pts)	0,99	0,44	5	0,98	0,38
	1° derivada SG (janela 7pts)	0,99	0,44	6	0,99	0,43
PLS/JK	Sem	0,98	0,42	5	0,98	0,43
	LB / SMM (janela 5pts)	0,99	0,27	5	0,99	0,20
	LB / SSG (janela 5pts)	0,99	0,32	5	0,99	0,29
	1° derivada SG (janela 7pts)	0,99	0,38	5	0,99	0,44
MLR/SPA	Sem	0,99	0,17	10	0,99	0,18
	LB / SMM (janela 5pts)	0,99	0,10	11	0,99	0,12
	LB / SSG (janela 5pts)	0,99	0,12	10	0,99	0,13
	1° derivada SG (janela 7pts)	0,99	0,15	11	0,99	0,17

\*número de variáveis latentes (PLS) ou número de variáveis espectrais (MLR/SPA).

LB = Correção de linha de base; SMM = Suavização média móvel; SSG = Suavização Savitsky-Golay.

Para todos os modelos (PLS, PLS/JK e MLR/SPA) os menores RMSEP's (**Tabela 5.8**) foram obtidos utilizando-se os espectros pré-processados com correção linear de linha de base e suavização com filtro média móvel (janela de 5 pontos). Comparando-se esses resultados às demais estratégias de pré-processamento, utilizando-se teste *F* (**Tabela 5.9**), pode-se observar que para os modelos PLS e MLR/SPA não há diferença

estatisticamente significativa entre essa estratégia de pré-processamento e a correção de linear linha de base e suavização por polinômio móvel e filtro de SG (janela de 5 pontos). Ambos os modelos PLS e MLR/SPA apresentaram um maior valor de RMSEP utilizando-se os espectros derivativos.

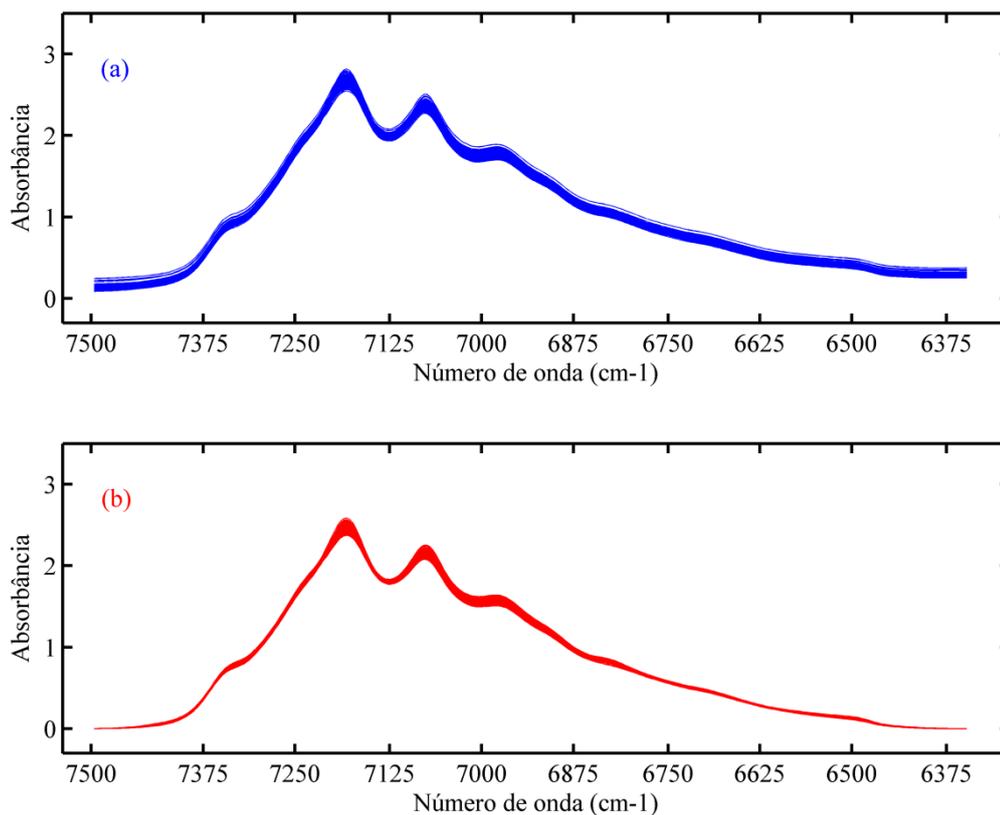
Já o modelo PLS/JK apresentou de fato um menor RMSEP para a estratégia LB/SMM (janela de 5 pontos), de acordo com o teste  $F$ , sendo, portanto, a melhor escolha entre as estratégias de pré-processamento avaliadas.

**Tabela 5.9**– Resultados para o teste  $F$  (nível de confiança de 95%) obtidos na comparação das diferentes estratégias de pré-processamento empregadas na construção dos modelos (PLS, PLS/JK e MLR/SPA) para o caminho óptico de 50 mm.

Modelo	Menor RMSEP	$F_{cal}$		
		LB / SMM (janela 5pts)	LB / SSG (janela 5pts)	1° derivada SG (janela 7pts)
PLS	LB / SMM (janela 5 pts)	-	1,32	1,70
PLS/JK	LB / SMM (janela 5 pts)	-	2,01	4,84
MLR/SPA	LB / SMM (janela 5 pts)	-	1,17	2,01

\* $F_{crit}(0,95; 44; 44) = 1,65$

Com base nos resultados obtidos, escolheu-se a estratégia de pré-processamento: correção de linha de base com suavização por média móvel (janela de 5 pontos) para correção das variações sistemáticas da linha de base dos espectros originais. Os espectros originais e pré-processados são mostrados nas **Figuras 5.12a-b**.



**Figura 5.12**– Espectros NIR das 129 amostras: (a) originais; (b) e pré-processados por correção de linha de base com suavização por média móvel com janela de 5pts (LB / SMM).

Comparando-se os resultados obtidos com os modelos PLS, PLS/JK e MLR/SPA utilizando-se a estratégia de pré-processamento escolhida (**Tabela 5.8**), verifica-se que o menor valor de RMSEP (0,12% v/v) foi obtido com a estratégia MLR/SPA. Esse resultado é comparável ao valor do RMSEP estabelecido pela norma ABNT 15568:2008 para o teor de biodiesel em misturas com diesel na faixa de concentração de 0 a 8% v/v de biodiesel (RMSEP = 0,10% v/v).

Para o caminho óptico de 50 mm, o modelo PLS/JK apresentou um melhor desempenho quando comparado ao modelo PLS (com todas as variáveis espectrais). Entretanto, esse modelo apresentou um valor de RMSEP (0,20 % v/v) acima do valor estabelecido pela norma ABNT 15568:2008, sendo estatisticamente diferente do modelo MLR/SPA a 95% de confiança (**Tabela 5.10**).

**Tabela 5.10** – Resultados do teste  $F$  (nível de 95% de confiança) comparando-se modelos PLS, PLS/JK e MLR/SPA construídos utilizando-se os espectros NIR (caminho óptico = 50 mm) pré-processados com correção de linha de base e suavização por filtro média móvel (janela de 5 pontos).

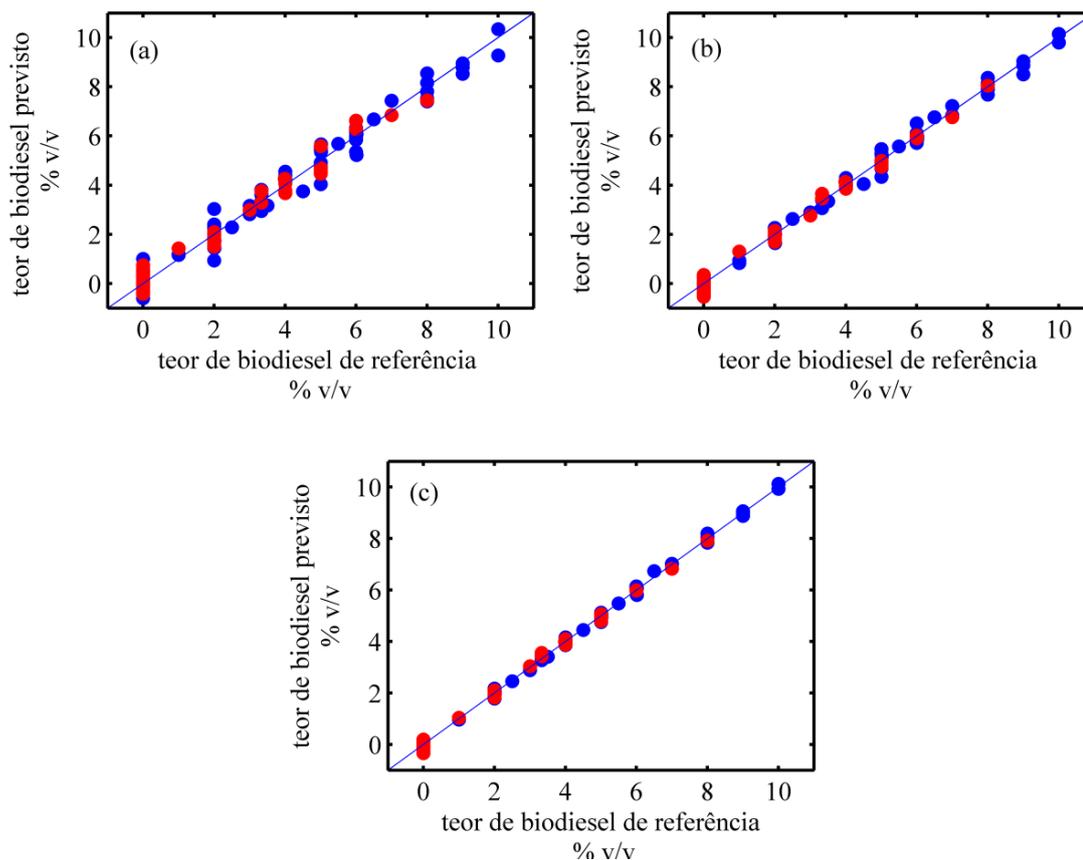
MLR/SPA × PLS	MLR/SPA × PLS/JK
7,56	2,78

$$F_{\text{crit}}(44; 44; 0,95) = 1,65$$

Esses resultados, também apontam o MLR/SPA como sendo a melhor estratégia para determinação do teor de biodiesel em misturas com diesel na presença de óleo vegetal *in natura*, o qual é mais simples e utiliza poucas variáveis espectrais para obtenção dos modelos de calibração.

As **Figuras 5.13a-c** mostram os gráficos dos valores previstos versus os valores de referência para os conjuntos de calibração (azul) e validação externa (vermelho) obtidos pelos modelos PLS, PLS/JK e MLR/SPA para o caminho óptico de 50 mm, respectivamente.

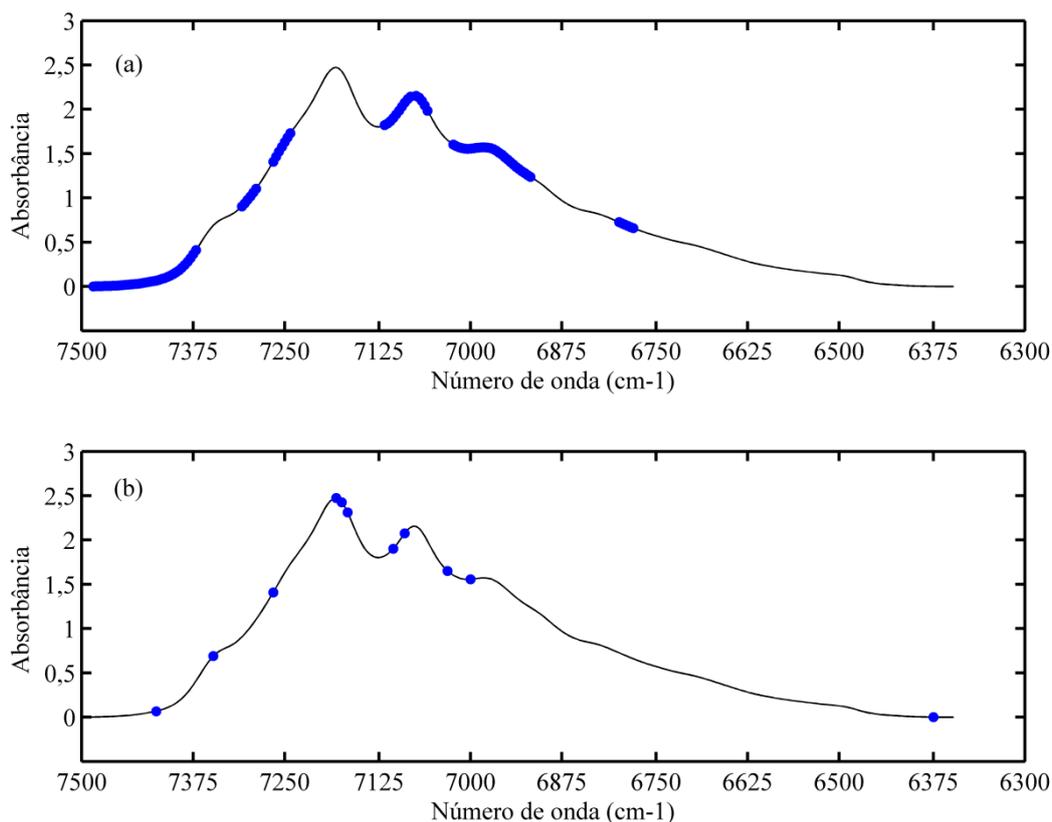
Como pode ser observado, os modelos construídos utilizando as estratégias de seleção de variáveis PLS/JK e MLR/SPA (**Figuras 5.13b-c**) apresentaram um melhor ajuste dos dados com relação ao modelo PLS (com todas as variáveis espectrais), **Figura 5.13a**. Entretanto, o modelo MLR/SPA apresentou um melhor resultado com relação ao PLS/JK. Essas observações estão de acordo com resultados mostrados na **Tabela 5.8**.



**Figura 5.13** – Gráfico dos valores previstos para teor de biodiesel versus valores de referência para os modelos: (a) PLS; (b) PLS/JK e (c) MLR/SPA utilizando o caminho óptico de 50 mm. (●: amostras de calibração e ●: validação externa).

As **Figuras 5.14a-b** mostram as variáveis espectrais selecionadas pelos algoritmos JK e SPA para construção dos modelos PLS e MLR (**Figuras 5.13b-c**) para determinação do teor de biodiesel utilizando o caminho óptico de 50 mm.

Como pode ser observada, a maioria das variáveis selecionadas encontra-se localizada na região espectral entre  $7500$  e  $6900\text{ cm}^{-1}$  associada ao primeiro sobretom das combinações de estiramento C-H (**Figura 5.14**).



**Figura 5.14** – Variáveis espectrais selecionadas pelas estratégias: (a) PLS/JK e (b) MLR/SPA para determinação do teor de biodiesel utilizando o caminho óptico de 50 mm.

### 5.3.4 Comparação dos resultados obtidos para os caminhos ópticos de 1,0, 10 e 50 mm

Os resultados obtidos para os caminhos ópticos avaliados ao longo desse estudo apontam o MLR/SPA como a melhor estratégia para determinação do teor de biodiesel em relação aos modelos PLS e PLS/JK. Comparando-se, então o desempenho dos modelos MLR/SPA para os diferentes caminhos ópticos estudados, verificou-se que o menor valor de RMSEP (0,07% v/v) foi obtido utilizando-se o caminho óptico de 10 mm (9000-6300  $\text{cm}^{-1}$ ). Essa faixa espectral inclui as regiões do primeiro sobretudo das combinações (7400-6900  $\text{cm}^{-1}$ ) e segundo sobretudo (8910-8000  $\text{cm}^{-1}$ ) dos modos de estiramento de ligações C-H.

Segundo o teste  $F$  (**Tabela 5.11**), não há diferença estatisticamente significativa entre o modelo MLR/SPA que utiliza essa região espectral em relação aquele que utiliza a região espectral das combinações (caminho

óptico de 1,0 mm). Sendo assim, ambas as regiões são apropriadas para previsão do teor de biodiesel em misturas diesel/biodiesel na presença de óleos vegetais.

O modelo MLR/SPA para o caminho óptico de 50 mm apresenta o valor de RMSEP (0,12 % v/v) estatisticamente diferente dos caminhos ópticos de 1,0 e 10 mm. Entretanto, esse valor é comparável ao estabelecido pela norma ABNT 15568:2008 (0,1 % v/v) para a faixa de concentração de biodiesel de 0 - 8% v/v.

**Tabela 5.11**– Resultados do teste  $F$  (nível de 95% de confiança) para os modelos MLR/SPA construídos utilizando-se os espectros NIR para os caminhos ópticos de 1,0; 10 e 50 mm.

$F_{cal}$		
<b>10 × 1,0 mm</b>	<b>10 × 50 mm</b>	<b>1,0 × 50 mm</b>
1,65	2,94	1,77

$F_{crit} (44; 44; 0,95) = 1,65$

Tais resultados são bastante promissores, visto que, a região espectral de sobretons tem sido pouco explorada até o momento na literatura para determinação do teor de biodiesel. As variáveis selecionadas pelo SPA nessa região espectral podem ser úteis para a construção de instrumentos de baixo custo baseados em LED's (Diodos Emissores de Luz) para uso no controle de qualidade de misturas diesel/biodiesel.

# **CONCLUSÃO**

## **Capítulo 6**

## 6 CONCLUSÃO

O presente trabalho investigou o uso da região de sobretons (NIR) para determinação do teor de biodiesel e classificação de misturas diesel/biodiesel na presença de óleo vegetal *in natura*. Para tanto, utilizou-se espectros NIR registrados em três diferentes caminhos ópticos (1,0, 10 e 50 mm).

Com base nos resultados obtidos, conclui-se que:

- (a) a região espectral de sobretons (caminhos ópticos de 10 e 50 mm) aliada à estratégia de classificação LDA/SPA, pode ser utilizada em estudos preliminares para detecção de adulteração de misturas diesel/biodiesel por óleos vegetais *in natura*, visto que apresentou bons resultados de classificação, com índice de acerto de 86% para os caminhos ópticos de 10 mm (9000-6300  $\text{cm}^{-1}$ ) e 50 mm (7500-6300  $\text{cm}^{-1}$ ). Tal estratégia consiste em uma análise rápida e não destrutiva;
- (b) a determinação do teor de biodiesel em misturas diesel/biodiesel na presença de óleos vegetais *in natura*, na faixa de 0,0 a 10%v/v, pode ser realizada utilizando-se a região espectral de sobretons associada à estratégia de regressão MLR/SPA, a qual é simples, rápida e utiliza poucas variáveis espectrais;
- (c) a região espectral de sobretons pode ser útil na construção instrumentos de baixo custo para o controle de qualidade de misturas diesel/biodiesel, considerando-se o baixo custo dos componentes óptico para essa região espectral.

# **PROPOSTAS FUTURAS**

## **Capítulo 7**

## 7 PROPOSTAS FUTURAS

Diante dos resultados obtidos no presente trabalho, apresentam-se as seguintes propostas para trabalhos futuros:

- (a) Determinar os demais parâmetros de qualidade (temperaturas de destilação T50% e T85%, massa específica e teor de enxofre) das misturas diesel/biodiesel;
- (b) Investigar o desempenho da região espectral de sobretons associada a ferramentas de calibração multivariada para determinação desses parâmetros de qualidade;
- (c) Avaliar o desempenho de modelos de calibração multivariada utilizando a região espectral de terceiro sobreton (caminho óptico de 100 mm) na determinação do teor de biodiesel em misturas diesel/biodiesel adulteradas com óleo vegetal, visto que esses espectros já foram registrados;
- (d) Construir modelos de calibração multivariada para determinação do teor de óleo vegetal (*in natura*) em mistura diesel/biodiesel utilizando-se a região espectral de sobretons;
- (e) Comparar o desempenho de modelos de calibração multivariada PLS frente aos parâmetros estabelecidos na norma ABNT 15568:2008 utilizando-se as amostras preparadas no presente trabalho para região espectral do MIR. Essa norma sugere o preparo de 160 misturas diesel/biodiesel (não adulteradas) para determinação do teor de biodiesel. Tal estudo pode demonstrar se é possível realizar a determinação desse parâmetro com um número menor de amostras de calibração;
- (f) Avaliar estratégias de transferência dos modelos de calibração multivariada desenvolvidos no presente trabalho;
- (g) Construir o fotômetro baseado em LED's para controle de qualidade de misturas diesel/biodiesel utilizando a região espectral de sobretons.

# **REFERÊNCIAS BIBLIOGRÁFICAS**

## **Capítulo 8**

## 8 REFERÊNCIAS BIBLIOGRÁFICAS

ALISKE, M. A.; ZAGONEL, G. F.; COSTA, B. J.; VEIGA, W.; SAUL, C. K. Measurement of biodiesel concentration in a diesel oil mixture. *Fuel*. **86: 1461, 2007.**

ANP. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. <<http://www.anp.gov.br>>. Acessado em: 13 de novembro de 2010.

ARAÚJO, M. C. U.; SALDANHA, T. C. B.; GALVÃO, R. K. H.; YONEYAMA, T.; CHAME, H. C.; VISANI, V. The successive projections algorithm for variable selection in spectroscopy multicomponent analysis. *Chemometrics and Intelligent Laboratory Systems*. **57: 65, 2001.**

Associação Brasileira de Normas Técnicas – ABNT, NBR 15568 - Biodiesel. Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio. Rio de Janeiro, 2008.

BAPTISTA, P.; FELIZARDO, P.; MENEZES, J. C.; CORREIA, M. J. N. Multivariate near infrared spectroscopy models for predicting the iodine value, CFPP, kinematic viscosity at 40° C and density at 15° C of biodiesel. *Talanta*. **77: 144, 2008.**

BARROS NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. 25 Anos de Quimiometria no Brasil. *Química Nova*. **29: 1401, 2006.**

BASHA, S. A.; GOPAL, K. R.; JEBARAJ, S. A review on biodiesel production, combustion, emissions and performance. *Renewable and Sustainable Energy Reviews*. **13: 1628, 2009.**

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B. *Chemometrics: a practical guide*. New York: A Wiley - Interscience Publication, 1998.

BLANCO, M.; VILLARROYA, I. NIR spectroscopy: a rapid-response analytical tool. *Trends in analytical chemistry*. **21: 240, 2002**.

Boletim Mensal da Qualidade dos Combustíveis Líquidos Automotivos Brasileiros - Outubro 2010 - Ano 9. Disponível em: <<http://www.anp.gov.br/?pg=33410>>. Acessado em: 09 de novembro de 2010.

BOUAID, A.; MARTINEZ, M.; ARACIL, J. A comparative study of the production of ethyl esters from vegetable oils as a biodiesel fuel optimization by factorial design. *Chemical Engineering Journal*. **134: 93, 2007**.

BREITKREITZ, M. C.; RAIMUNDO, I. M.; ROHWEDDER, J. J. R.; PASQUINI, C.; DANTAS, H. A.; JOSE, G. E. Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration. *The Analyst*. **9: 1204, 2003**.

British Standard - BS. European Standard, EN 14078 - Liquid petroleum products. Determination of fatty methyl ester (FAME) content in middle distillates. Infrared spectrometry method. London, 2009.

British Standard - BS. European Standard, EN ISO 14103 - Fat and oil derivatives. Fatty acid methyl esters (FAME). Determination of ester and linolenic acid methyl ester contents. London, 2003.

BROWN, J. C. R.; KEATES, A. C.; BROWN, A. S. Optimised determinations of water in ethanol by encoded photometric near infrared spectroscopy: a

---

special of sequential standard addition calibration. *Analytica Chimica Acta*. **690: 47, 2011.**

CERQUEIRA, E. O.; POPPI, R. J.; KUBOTA, L. T.; MELO, C. Utilização de filtro de transformada de Fourier para minimização de ruídos em sinais analíticos. *Química Nova*, **23: 690, 2000.**

CERQUEIRA, E. O. ; ANDRADE, J. C.; POPPI, R. J. Redes Neurais e suas aplicações em calibração multivariada. *Química Nova*. **24: 864, 2001.**

CORGOZINHO, C. N. C.; PASA, V. M. D.; BARBEIRA, P. J. S. Determination of residual oil in diesel oil by spectrofluorimetric and chemometric analysis. *Talanta*. **76: 479, 2008.**

CORREIA, P. R. M.; FERREIRA, M. M. C. Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos. *Química Nova*. **30: 481, 2007.**

COSTA FILHO, P. A.; POPPI, R. J. Aplicação de algoritmos genéticos na seleção de variáveis em espectroscopia no infravermelho médio. Determinação simultânea de glicose, maltose e frutose. *Química Nova*, **25: 46, 2002.**

DEMIRBAS, A. Comparison of transesterification methods for production from vegetable oils and fats. *Energy Conversion and Management*. **49: 125, 2008.**

DE LIRA, L. F. B. *Desenvolvimento de métodos analíticos para monitoramento da qualidade do biodiesel e suas misturas*. Recife, Programa de Pós-Graduação em Química, UFPE, 2010. Tese de doutorado, 145p.

DE LIRA, L. F. B.; ALBUQUERQUE, M. S.; PACHECO, J. G. A.; FONSECA, T. M.; CAVALCANTI, E. H. S.; STRAGEVITCH, L.; PIMENTEL, M. F. Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of biodiesel. *Microchemical Journal*. **96: 126, 2010.**

DE LIRA, L. F. B.; VASCONCELOS, F. V. C.; PEREIRA, C. F.; PAIM, A. P.; STRAGEVITCH, L.; PIMENTEL, M. F. Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*. **89: 405, 2010.**

DI NEZIO, M.S.; PISTONESI, M.F.; FRAGOSO, W.D.; PONTES, M.J.C.; GOICOECHEA, H.C; ARAUJO, M.C.U.; FERNÁNDEZ BAND, B.S. Successive projections algorithm improving the multivariate simultaneous direct spectrophotometric determination of five phenolic compounds in sea water. *Microchemical Journal*. **85: 194, 2007.**

ENCINAR, J. M.; GONZÁLES, J. F.; RODRÍGUEZ, R. A. Ethanolysis of used frying oil. Biodiesel preparation and characterization. *Fuel Processing Technology*. **88: 513, 2007.**

ESBENSEN, K. H. *Multivariate Data Analysis – in practice: an introduction to multivariate analysis and experimental design*. 5 ed. Aalborg University, Esbjerg: CAMO, 2002.

FANGRUI, M.; HANNA, M. A. Biodiesel production: a review. *Bioresource Technology*. **70: 1, 1999.**

FELÍCIO, C. C.; BRÁS, L. P.; LOPES, J. A.; CABRITA, L.; MENESES, J. C. Comparison of PLS algorithms in gasoline and gas oil parameter monitoring with MIR and NIR. *Chemometrics and Intelligent Laboratory Systems*. **78: 74, 2005.**

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S., VOLPE, P. L. O. Quimiometria I: calibração multivariada, um tutorial. *Química Nova*. **22: 724, 1999.**

FELIZARDO, P.; BAPTISTA, P.; MENEZES, J. C.; CORREIA, M. J. N. Multivariate near infrared spectroscopy models for predicting methanol and water content in biodiesel. *Analytica. Chimica. Acta*. **595: 107, 2007.**

FORINA, M.; LANTERI, S.; CASALE, M.; OLIVEROS, M. C. C. Stepwise orthogonalization of predictors in classification and regression techniques: an "old" technique revisited. *Chemometrics and Intelligent Laboratory Systems*. **87:252, 2007.**

GAMBARRA-NETO, F. F.; 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 square wave voltammetry with multivariate data analysis. *Talanta*, **77: 1660, 2009.**

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SALDANHA, T. C. B., VISANI, V.; PIMENTEL, M. F. Estudo comparativo sobre filtragem de sinais instrumentais usando transformadas de Fourier e Wavelet. *Química Nova*, **24: 874, 2001.**

GALVÃO, R. K. H.; ARAUJO, M. C. U.; JOSÉ, G. E.; PONTES, M. J. C.; SILVA, E. C.; SALDANHA, T. C. B. A method for calibration and validation subsetpartitioning. *Talanta*. **67: 736, 2005.**

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SILVA, E. C.; JOSÉ, G. E.; SOARES, S. F. C.; PAIVA, H. M. Cross-validation for the selection of spectral variables using the Successive projections algorithm. *Journal of Brazilian Chemical Society*. **18: 1580, 2007.**

GUARIEIRO, L. L. N.; PINTO, A. C.; AGUIAR, P. F.; RIBEIRO, N. M. Metodologia analítica para quantificar o teor de biodiesel na mistura biodiesel/diesel utilizando espectroscopia na região do infravermelho. *Química Nova*. **31: 421, 2008.**

GAYDOU, V.; KISTER, J.; DUPUY, N. Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil. *Chemometrics and Intelligent Laboratory Systems*. in press, 2010. <<http://dx.doi.org/10.1016/j.chemolab.2010.05.002>>.

GERIS, R.; SANTOS, N. A. C.; AMARAL, B. A.; MAIA, I. S.; CASTRO, V. D.; CARVALHO, J. R. M. Biodiesel de soja – Reação de transesterificação para aulas práticas de química orgânica. *Química Nova*. **30: 1369, 2007.**

HONORATO, F. A. *Previsão das propriedades das gasolinas do Nordeste empregando espectroscopia NIR/MID e transferência de calibração*. Recife, Programa de Pós-Graduação em Química, UFPE. 2006. Tese de doutorado, 106p.

HONORATO, F. A.; BARROS NETO, B.; MARTINS, M. N.; GALVÃO, R. K. H.; PIMENTEL, M. F. Transferência de calibração em métodos multivariados. *Química Nova*. **30: 1301, 2007.**

HONORATO, F. A.; NETO, B.B.; PIMENTEL, M. F.; STRAGEVITCH, L.; GALVÃO, R. K. H. Using principal component analysis to find the best calibration settings for simultaneous spectroscopic determination of several gasoline properties. *Fuel*. **87: 3706, 2008.**

KENNARD, R.W.; STONE, L.A. Computer-aided design of experiments. *Technometrics*. **11: 137, 1969.**

KNOTHE, G.; STEIDLEY, K. R. Kinematic viscosity of biodiesel fuel components and related compounds. Influence of compound structure and comparison to petrodiesel fuel components. *Fuel*. **84: 1059, 2005**.

KNOTHE, G.; GERPEN, J. V.; KRAHL, J.; RAMOS, L.P. *Manual do biodiesel*. Traduzido do original "*The Biodiesel Handbook*" por Luiz Pereira Ramos. São Paulo: Edgard Blucher, 2006.

Lei N° 11.097, de janeiro de 2003. Dispõe sobre criação do Programa Nacional de Produção e Uso de Biodiesel e sobre a adição de biodiesel ao óleo diesel. Disponível em: <<http://nxt.anp.gov.br/NXT/gateway.dll/leg/leis/2005/lei%2011.097%20-%202005.xml?>>>. Acessado em: 13 de novembro de 2010.

LEUNG, D. Y. C.; WU, X.; LEUNG, M.K.H. A review on biodiesel production using catalyzed transesterification. *Applied Energy*. **87: 1083, 2010**.

LEARDI, R. Genetic algorithms in chemometrics and chemistry: a review. *Journal of Chemometrics*. **15: 559, 2001**.

LÔBO, I. P.; FERREIRA, S. L. C. Biodiesel: Parâmetros de qualidade e métodos analíticos. *Química Nova*. **32: 1596, 2009**.

MAHAMUNI, N. N.; ADEWUYI, Y. G. Fourier transform infrared spectroscopy (FTIR) method to monitor soy biodiesel and soybean oil in transesterification reactions, petrodiesel/biodiesel blends, and blend adulteration with soy oil. *Energy and Fuels*. **23: 3773, 2009**.

MARCHETTI, J. M.; MIGUEL, V. U.; ERRAZU, A. F. Possible methods for biodiesel production. *Renewable and Sustainable Energy Reviews*. **11: 1300, 2007**.

MEHER, L. C.; SAGAR, D. V.; NAIK, S. N. Technical aspects of biodiesel production by transesterification - a review. *Renewable and Sustainable Energy Reviews*. **10: 248, 2006.**

McCLURE, W. F. 204 Years of near infrared technology: 1800-2003. *J. Near Infrared Spectrosc.* **11: 487, 2003.**

MONTEIRO, M. R.; AMBROZIN, A. R. P., LIÃO, L. M., FERREIRA, A. G. Determination of biodiesel blend levels in different diesel samples by  $^1\text{H}$  NMR. *Fuel*. **88: 691, 2009.**

MOREIRA, E. D.; PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. Near infrared reflectance spectrometry classification of cigarettes using the successive projections algorithm for variable selection. *Talanta*. **79: 1260, 2009.**

MOTA, C. J. A.; SILVA, C. X. A.; GONÇALVES, V. L. C. Gliceroquímica: Novos produtos e processos a partir da glicerina de produção de biodiesel. *Química Nova*. **32: 639, 2009.**

NAES, T.; ISAKSSON, T.; FEARN, T.; DAVIES, T. *A user-friendly guide to Multivariate Calibration and Classification*. Chichester-UK: NIR Publications, 2002.

NASCIMENTO, E. C. L. *Um fotômetro microcontrolado LED-NIR, portátil e de baixo custo para análise screening de gasolinas tipo C*. João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2008. Tese de doutorado, 110p.

NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval partial least-squares regression (i-PLS): a comparative chemometric study with an example from near infrared spectroscopy. *Applied Spectroscopy*. **54: 413, 2000.**

OLIVEIRA, F. C.; BRANDÃO, C. R. R.; RAMALHO, H. F.; COSTA, L. A. F.; SUAREZ, P. A. Z.; RUBIM, J. C. Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy. *Analytica Chimica Acta*. **587: 194, 2007**.

OLIVEIRA, J. S.; MONTALVÃO, R.; DAHER, L.; SUAREZ, P.; RUBIM, J. C. Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies. *Talanta*. **69: 1278, 2006**.

PASQUINI, C. Near infrared spectroscopy: fundamentals, practical aspects and analytical applications. *Journal Brazilian Chemical Society*. **14: 198, 2003**.

PERALTA-ZAMORA, P.; CORNEJO-PONCE, L., NAGATA, N., POPPI, R. J. Alternativas quimiométricas para a resolução de problemas analíticos clássicos. Determinação espectrofotométrica de misturas de zircônio e háfnio. *Química Nova*. **20: 469, 1997**.

PEREIRA, C. F. *Uso da espectroscopia de correlação bidimensional (2D) e construção e avaliação de um espectropolarímetro para região do infravermelho próximo (NIR)*. Campinas, Programa de Pós-Graduação em Química, Unicamp, 2006. Tese de doutorado, 124p.

PIMENTEL, M. F.; RIBEIRO, G. M. G. S.; CRUZ, R. S.; STRAGEVITCH, L.; PACHECO FILHO, J. G. A.; TEIXEIRA, L. S. G. Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. *Microchemical Journal*. **82: 201, 2006**.

PISTONESI, M. F.; DI NEZIO, M.; CENTURIÓN, M. E.; LISTA, A. G.; FRAGOSO, W. D.; PONTES, M. J. C.; ARAÚJO, M. C. U.; FERNÁNDEZ BAND. Simultaneous determination of hydroquinone, resorcinol, phenol, m-cresol and p-cresol in untreated air samples using spectrofluorimetry

and a custom multiple linear regression-successive projection algorithm. *Talanta*. **83: 320, 2010.**

PINTO, A. C.; GUARIEIRO, L. L. N.; REZENDE, M. J. C.; RIBEIRO, N. M.; TORRES E. A.; LOPES, W. A.; PERREIRA, P. A. P.; ANDRADE, J. B. Biodiesel: An Overview. *Journal Brazilian Chemists Society*. **16: 1313, 2005.**

PONTES, M. J. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; MOREIRA, P. N. T.; PESSOA NETO, O. D.; JOSÉ, G. E., SALDANHA, T. C. B. The successive projections algorithm for spectral variable selection in classification problems. *Chemom. Intell. Lab. Syst.* **78: 11, 2005.**

PONTES, M. J. C. *Algoritmo das Projeções Sucessivas para seleção de variáveis espectrais em problemas de classificação*. João Pessoa, Programa de Pós-Graduação em Química, UFPB. 2009. Tese de doutorado, 144p.

PONTES, M. J. C.; CORTEZ, J.; GALVÃO, R. K. H.; PASQUINI, C.; COELHO, R. M.; CHIBA, M. K.; ABREU, M. F.; MADARI, B. E. Classification of Brazilian soils by using LIBS and variable selection in the wavelet domain. *Analytica Chimica Acta*. **642: 12, 2009.**

PONTES, M. J. C.; ROCHA, A. M. J.; PIMENTEL, M. F.; PEREIRA, C. F. Determining the quality of insulating oils using near infrared spectroscopy and wavelength selection. *Microchemical Journal*. *in press*, 2011. <<http://dx.doi.org/10.1016/j.microc.2011.02.010>>.

Portal do BIODIESEL. Disponível em <<http://www.biodiesel.gov.br>>. Acessado em: 09 de novembro de 2010.

RAMADHAS, A. S.; JAYARAJ, S.; MURALEEDHARAN, C. Biodiesel production from high FFA rubber seed oil. *Fuel*. **84: 335, 2005**.

REIS, C; ANDRADE, J. C. Planejamento experimental para misturas usando cromatografia em papel. *Química Nova*. **19: 3, 1996**.

RESOLUÇÃO ANP N° 07. ANP – Agência Nacional do Petróleo, Gás Natural e Bicombustíveis, 19 de Mar. 2008. Disponível em: <[http://nxt.anp.gov.br/NXT/gateway.dll/leg/resolucoes\\_anp/2008/mar%C3%A7o/ranp%207%20%202008.xml?f=templates\\$fn=document-frame.htm\\$3.0\\$q=\\$x=\\$nc=7834](http://nxt.anp.gov.br/NXT/gateway.dll/leg/resolucoes_anp/2008/mar%C3%A7o/ranp%207%20%202008.xml?f=templates$fn=document-frame.htm$3.0$q=$x=$nc=7834)>. Acessado em: 13 de novembro de 2010.

RESOLUÇÃO ANP N° 42. ANP – Agência Nacional do Petróleo, Gás Natural e Bicombustíveis, 16 de Dez. 2009. Disponível em: <[http://nxt.anp.gov.br/NXT/gateway.dll/leg/resolucoes\\_anp/2009/dezembro/ranp%2042%20-%202009.xml](http://nxt.anp.gov.br/NXT/gateway.dll/leg/resolucoes_anp/2009/dezembro/ranp%2042%20-%202009.xml)>. Acessado em: 20 de novembro de 2010.

RESOLUÇÃO CNPE N° 6. CNPE – Conselho Nacional de Política Energética, 16 de Set. 2009. Disponível em: <[http://nxt.anp.gov.br/NXT/gateway.dll/leg/folder\\_resolucoes/resolucoes\\_cnpe/2009/rcnpe%206%20%202009.xml?f=templates\\$fn=document-frame.htm\\$3.0\\$q=\\$x=\\$nc=9975](http://nxt.anp.gov.br/NXT/gateway.dll/leg/folder_resolucoes/resolucoes_cnpe/2009/rcnpe%206%20%202009.xml?f=templates$fn=document-frame.htm$3.0$q=$x=$nc=9975)>. Acessado em: 20 de novembro de 2010.

SENA, M. M.; POPPI, R. J.; FRIGHETTO, R. T. S.; VALARINI, P. J. Avaliação do uso de métodos quimiométricos em análise de solos. *Química Nova*. **23: 547, 2000**.

SCHUCHARDT, U.; RINALDI, R.; GARCIA, C.; MARCINIUK, L. L.; ROSSI, A. V. Síntese de biodiesel: Uma proposta contextualizada de experimento para laboratório de química geral. *Química Nova*. **30: 1374, 2007**.

SIMÕES, S. S. *Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIRR e calibração multivariada*. João Pessoa, Programa de Pós-Graduação em Química, UFPB, 2008. Tese de doutorado, 98p.

SKOOG, D. A.; HOLLER, F. J.; NEIMAN, T. A. *Princípios de Análise Instrumental*. 5. Ed. Porto Alegre: Bookman, 2006.

SOARES, I. P.; REZENDE, T. F.; SILVA, R. C.; CASTRO, E. V. R.; FORTES, I. C. P. Multivariate calibration by variable selection for blends of raw soybean oil/biodiesel from different sources using Fourier transform infrared spectroscopy (FTIR) spectra data. *Energy Fuels*. **22: 2079, 2008.**

SOUTO, U. T. C.; 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 Chemistry*. **119: 368, 2010.**

SUAREZ, P. A. Z.; MENEGHETTI, S. M. P. 70º Aniversário do biodiesel em 2007: evolução histórica e situação atual no Brasil. *Química Nova*. **30: 2068, 2007.**

SUAREZ, P. A. Z.; SANTOS, A. L. F.; RODRIGUES, J. P.; ALVES, M. B. Bicom bustíveis a partir de óleos e gorduras: desafios tecnológicos para viabilizá-los. *Química Nova*. **32: 768, 2009.**

TAPANES, N. C. O.; ARANDA, D. A. G.; CARNEIRO, J. W. M.; ANTUNES, O. A. C. Transesterification of *Jatropha curcas* oil glycerides: theoretical and experimental studies of biodiesel reaction. *Fuel*. **87: 2286, 2008.**

WORKMAN, J.; WEYER, L. *Practical guide to interpretative near infrared spectroscopy*. New York: CRC Press, 2008.

ZAGONEL, G. F.; PERALTA-ZAMORA, P. R.; RAMOS, L. P. Multivariate monitoring of soybean oil ethanolysis by FTIR. *Talanta*. **63: 1021, 2004.**

**ANEXO**

**Anexo 1**– Concentrações de diesel, biodiesel e óleo (% v/v) contidas nas 33 misturas do planejamento.

<b>Amostra</b>	<b>Diesel (%v/v)</b>	<b>Biodiesel (%v/v)</b>	<b>Óleo (%v/v)</b>
1	100	0	0
2	99	1	0
3	98	2	0
4	99	0	1
5	98	0	2
6	97	3	0
7	96	4	0
8	96	2	2
9	97	0	3
10	96	0	4
11	95	5	0
12	94	6	0
13	94	4	2
14	94	2	4
15	95	0	5
16	94	0	6
17	93	7	0
18	92	8	0
19	92	6	2
20	92	4	4
21	92	2	6
22	93	0	7
23	92	0	8
24	91	9	0
25	90	10	0
26	90	8	2
27	90	6	4
28	90	4	6
29	90	2	8
30	91	0	9
31	90	0	10
32	93,33	3,33	3,33
33	93,33	3,33	3,33