



**UNIVERSIDADE ESTADUAL DA PARAÍBA  
CAMPUS I  
CENTRO DE CIÊNCIAS E TECNOLOGIA - CCT  
CURSO DE GRADUAÇÃO E QUÍMICA INDUSTRIAL**

**ELLEN DE SOUZA ALMEIDA DUARTE**

**CLASSIFICAÇÃO DE MISTURAS BIODIESEL/DIESEL DE ACORDO COM A  
ROTA DE SÍNTESE USANDO ESPECTROSCOPIA NIR**

**CAMPINA GRANDE  
2017**

**ELLEN DE SOUZA ALMEIDA DUARTE**

**CLASSIFICAÇÃO DE MISTURAS BIODIESEL/DIESEL DE ACORDO COM A  
ROTA DE SÍNTESE USANDO ESPECTROSCOPIA NIR**

Trabalho de Conclusão de Curso apresentado ao curso de Graduação em Química Industrial da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Química Industrial.

**Área de concentração:** Química Analítica.

**Orientador:** Prof. Dr. José Germano Vêras Neto.

**CAMPINA GRANDE  
2017**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

D812c Duarte, Ellen de Souza Almeida.  
Classificação de misturas biodiesel/diesel de acordo com a rota de síntese usando espectroscopia NIR [manuscrito] : / Ellen de Souza Almeida Duarte. - 2017.  
37 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Química Industrial) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2017.

"Orientação : Prof. Dr. José Germano Vêras Neto, Departamento de Química - CCT."

1. Biodiesel. 2. Diesel. 3. Espectroscopia NIR.

21. ed. CDD 662.88

ELLEN DE SOUZA ALMEIDA DUARTE

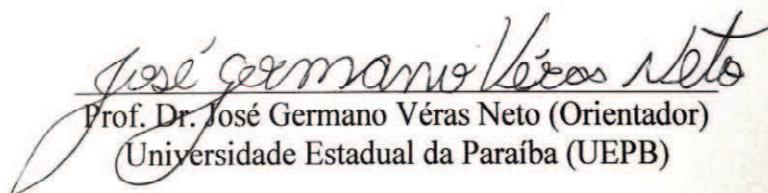
**CLASSIFICAÇÃO DE MISTURAS BIODIESEL/DIESEL DE ACORDO COM A  
ROTA DE SÍNTESE USANDO ESPECTROSCOPIA NIR**

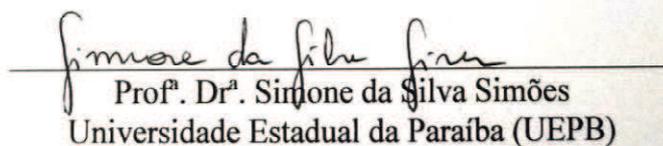
Trabalho de Conclusão de Curso apresentado ao curso de Graduação em Química Industrial da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Química Industrial.

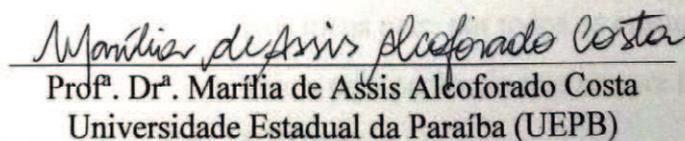
Área de concentração: Química Analítica.

Aprovada em: 14/12/2017.

**BANCA EXAMINADORA**

  
Prof. Dr. José Germano Vêras Neto (Orientador)  
Universidade Estadual da Paraíba (UEPB)

  
Prof.<sup>a</sup> Dr.<sup>a</sup> Simone da Silva Simões  
Universidade Estadual da Paraíba (UEPB)

  
Prof.<sup>a</sup> Dr.<sup>a</sup> Marília de Assis Alcoforado Costa  
Universidade Estadual da Paraíba (UEPB)

Aos meus pais, por todos os ensinamentos acerca da vida e pela paciência que sempre tiveram comigo,  
DEDICO.

*" Diga-me e eu esquecerei,  
ensina-me e eu poderei lembrar,  
envolva-me e eu aprenderei."*  
**Benjamin Franklin**

## RESUMO

Este trabalho teve como objetivo empregar técnicas de reconhecimento de padrões supervisionadas: Modelagem Suave Independente Por Analogias de Classes (SIMCA), K-vizinhos mais próximos (KNN), Algoritmo das Projeções Sucessivas por Análise Discriminante Linear (SPA-LDA), e Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA) aliadas à espectroscopia do infravermelho próximo (NIR) como um método analítico para a classificação de misturas biodiesel/diesel (B7) através de sua rota de síntese (metílica e etílica). Um total de 180 amostras de misturas biodiesel/diesel B7 foram utilizadas. O modelo PLS-DA apresentou 100% de classificação correta na discriminação das amostras com respeito à rota de síntese, verificou-se a avaliação do modelo por meio de valores de sensibilidade e especificidade para cada parâmetro, na classe de interesse. Portanto, a metodologia proposta é rápida, uma vez que permite a classificação das misturas biodiesel/diesel pela a sua rota de síntese. Por conseguinte, ele pode ser usado no controle de qualidade deste tipo de biocombustível.

**Palavras-Chave:** biodiesel, diesel, espectroscopia NIR.

## ABSTRACT

This work aims to use supervised standards recognition techniques: Independent Modeling by Class Analogies (SIMCA), K-nearest neighbors (KNN), Algorithm of Successive Projections by Linear Discriminant Analysis (SPA-LDA) and Discriminant Analysis by Partial Least Squares (PLS-DA) allied to Near Infrared Spectroscopy (NIR) as an analytical method for the classification of biodiesel / diesel mixtures (B7) through its route of synthesis (methyl and ethyl). A total of 180 samples of B7 biodiesel / diesel blends with. The PLS-DA model presented 100% correct classification in the discrimination of the samples with respect to the route of synthesis, an evaluation of the model was verified through values of sensitivity and specificity for each parameter, in the interest class. Please, a proposed methodology is fast, since it allows a classification of biodiesel / diesel blends by its own route of synthesis. Therefore, it can be used without quality control of this type of biofuel.

**Keywords:** biodiesel, diesel, NIR spectroscopy.

## LISTA DE ILUSTRAÇÕES

Figura 1 - Espectro de radiação eletromagnética e distribuição do Infravermelho.....	15
Figura 2 – Representação esquemática da análise discriminante linear.....	20
Figura 3a - Espectros Brutos das amostras de misturas biodiesel/diesel.....	24
Figura 3b- Espectros derivados das amostras de misturas biodiesel/diesel.....	25
Figura 4 - Gráfico de scores (PC1 x PC2 x PC3).....	26
Figura 5 - Característica do receptor de operação (ROC).....	30

## LISTA DE TABELAS

Tabela 1 – Classificação quanto à rota de síntese – SIMCA.....	27
Tabela 2 – Classificação quanto à rota de síntese – KNN.....	28
Tabela 3 – Classificação quanto à rota de síntese – SPA-LDA.....	28
Tabela 4 – Classificação quanto à rota de síntese – PLS-DA.....	29

## LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
HCA	Conselho Nacional de Desenvolvimento Científico e Tecnológico
KNN	Sistema Integrado de Automação de Bibliotecas.
LDA	Análise Discriminante Linear
LV	Variáveis Latentes
NIR	Infravermelho Próximo
PLS-DA	Análise Discriminante pelos Mínimos Quadrados Parciais
ROC	Característica do Receptor de Operação
SIMCA	Modelagem Flexível Independente por Analogias de Classes
SPA	Algoritmo das Projeções Sucessivas
SPA-	Análise Discriminante Linear associado ao Algoritmo das Projeções
LDA	Sucessivas

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>11</b>
<b>1.2</b>	<b>OBJETIVOS .....</b>	<b>13</b>
1.2.1	Objetivo geral .....	13
1.2.2	Objetivos específicos .....	13
<b>2</b>	<b>REVISÃO DE LITERATURA .....</b>	<b>14</b>
2.1	Biodiesel. ....	14
2.2	Misturas biodiesel/diesel.....	14
2.3	Espectroscopia no Infravermelho Próximo (NIR).....	15
2.4	Quimiometria.....	16
2.5	Técnicas de reconhecimento de padrões.....	17
2.5.1	Técnicas de reconhecimento de padrões não supervisionadas.....	17
2.5.1.2	Análise por Componentes Principais (PCA).....	17
2.5.2	Técnicas de reconhecimento de padrões supervisionadas.....	18
2.5.2.1	Modelagem Flexível Independente por Analogias de Classes (SIMCA).....	18
2.5.2.2	K- vizinhos mais próximos (KNN).....	18
2.5.2.3	Análise Discriminante Linear (LDA).....	19
2.6	Ferramentas de diagnóstico da classificação.....	20
2.6.1	Taxa de Falso Positivo.....	20
2.6.2	Taxa de Falso Negativo.....	21
2.6.3	Taxa de Sensibilidade.....	21
2.6.4	Taxa de Especificidade.....	21
2.6.5	Precisão Global.....	21
2.6.6	Característica de Operação do Receptor (ROC).....	22
<b>3</b>	<b>METODOLOGIA .....</b>	<b>23</b>
3.1	Reagentes e amostras de biodiesel.....	23
3.2	Instrumentos e programas computacionais.....	23
<b>4</b>	<b>RESULTADOS E DISCUSSÃO .....</b>	<b>24</b>
4.1	Análise exploratória.....	25
4.2	Classificação.....	26
4.3	Modelagem Flexível Independente por Analogias de Classes (SIMCA).....	27
4.4	KNN- vizinhos mais próximos.....	27
4.5	Análise Discriminante Linear associado ao Algoritmo das Projeções Sucessivas (SPA- LDA).....	28
4.6	Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA).....	29
<b>5</b>	<b>CONCLUSÃO.....</b>	<b>31</b>
	<b>REFERÊNCIAS .....</b>	<b>32</b>

## 1 INTRODUÇÃO

Objetivando reduzir o uso de combustíveis fósseis, o mundo tem buscado cada vez mais fontes alternativas de energia. Uma destas alternativas reside na obtenção do biodiesel, que é extraído de óleos vegetais ou de gordura animal. Como consequência, existe o benefício da redução de custos e principalmente, e da emissão de gases poluentes na atmosfera. Neste sentido, esta relação custo/benefício constitui-se como importante vantagem na utilização do biodiesel. (LEE et al, 2011; VASUDEVAN, 2010; MURUGESAN et al, 2009).

Os óleos vegetais tiveram seu uso testado desde o século XIX, como alternativa para substituição ao óleo diesel (motores de ciclo Otto). Seu uso produziu resultados satisfatórios no próprio motor diesel. (FERREIRA, 2005)

Os principais métodos para a produção de biodiesel incluem micro emulsões, craqueamento térmico e transesterificação. (TAKASE et al, 2015; ARANSIOLA et al, 2014) A transesterificação é o processo de transformação de um éster carboxílico em outro, através da troca do grupo RO- presente no éster original, por outro grupo semelhante proveniente de um álcool na presença de um catalisador ou não (condições supercríticas), para dar origem à glicerina e a mono-álquil ésteres (biodiesel) (ABBASZAADEH et al, 2012; BALAT, 2011; YUSUF et al, 2010, KNOTHE, 2006).

Com relação ao agente de transesterificante, o processo reacional ocorre preferencialmente com álcoois simples, tais como metanol, etanol, propanol e butanol (GERIS et al, 2007). No entanto, o metanol e o etanol são os álcoois mais utilizados. De acordo com Aransiola, et al. 2014, o metanol é mais frequentemente utilizado por razões de natureza física e química (cadeia curta e polaridade), além disso, permite a separação simultânea do glicerol. A mesma reação usando etanol é mais complicada, pois requer um álcool anidro, bem como um óleo com baixo teor de água para levar à separação do glicerol.

Contudo, vale a pena ressaltar que os combustíveis renováveis são na sua maioria provenientes de atividades agrícolas, como cana de açúcar, oleaginosas e biomassa florestal, entre outros. Apesar de seu alto potencial na demanda por geração de energia estacionária, os principais biocombustíveis produzidos em grande escala são o biodiesel, metanol e etanol, demonstrando ser a solução economicamente mais viável e amplamente implementada para substituir os combustíveis fósseis convencionais (CREMONEZ et al, 2015).

O biodiesel utilizado em vários países da Europa e nos Estados Unidos são ésteres produzidos principalmente pela rota metílica. Ressalte-se, no entanto, que o metanol tem uma toxicidade muito elevada, ele traz malefícios à saúde, podendo causar, inclusive, cegueira e

câncer. O etanol apresenta a vantagem de não ser tóxico, ser biodegradável e ser produzido a partir de fontes renováveis.

Nos últimos anos diversos métodos foram propostos para identificar o tipo de óleo utilizado na produção de biodiesel e de suas misturas (VÉRAS et al., 2010; BALABIN et al., 2011; VÉRAS et al., 2012; SILVA et al., 2012; INSAUSTI et al., 2012; COSTA GB et al., 2015). Para a classificação pela rota de síntese Mazivila et al. (2015), empregou Análise Discriminante dos Mínimos Quadrados (PLS-DA, do inglês *Partial Least Square Discriminant Analysis*), aliada a Espectroscopia no Infravermelho Médio (MIR, do inglês "*Medium Infrared Spectroscopy*") como um método de análise para a classificação simultânea de biodiesel a partir de diferentes óleos (soja e utilizado óleo de fritura) e rotas (metílica e etílica).

Por outro lado o método de Espectroscopia no Infravermelho Próximo (NIR, do inglês *Near Infrared Spectroscopy*) possui algumas vantagens que motivam sua utilização na caracterização de processos, dentre os quais podem ser citadas a velocidade da medida quando comparada a outros processos (usualmente as medidas são tomadas em menos de 1 segundo), sendo uma técnica não destrutiva e que na grande maioria dos casos não necessita de preparação da amostra. Além destas vantagens, esta técnica é considerada altamente versátil. De forma genérica, a espectroscopia NIR é sensível a qualquer fator que afete a massa atômica, força de ligação, momento de polo ou simetria da molécula.

Para tanto, este trabalho visa classificar os diferentes tipos de biodiesel em relação ao álcool utilizada no processo de transesterificação (metílico e etílico) com uma combinação apropriada de espectroscopia NIR e técnicas de reconhecimento de padrões supervisionado, como: Modelagem Flexível Independente por Analogia de Classe (SIMCA, do inglês *Soft Independent Modeling by Class Analogy*), K-ésimo Vizinhos mais Próximos (KNN, do inglês *K-nearest Neighbors*) e Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA). O Algoritmo das Projeções Sucessivas (SPA, do inglês *Successive Projections Algorithm*) método de seleção de variáveis em combinação com modelos Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*) também foi investigado. (GHASEMI-VARNAMKHASTI et al., 2012)

## 1.2 OBJETIVOS

### 1.2.1 Objetivo geral

Desenvolver uma metodologia simples, rápida, de baixo custo e não destrutiva baseada na utilização da espectroscopia no infravermelho próximo (NIR) e técnicas de reconhecimento de padrões supervisionadas para classificar amostras de biodiesel/diesel (B7) em termos de sua rota de síntese, metílica ou etílica.

### 1.2.2 Objetivos específicos

- Utilizar dos espectros NIR, como informação analítica.
- Utilizar técnicas de reconhecimento de padrões supervisionados SIMCA, KNN, e PLS-DA.
- Classificar misturas biodiesel/diesel B7 a partir da sua rota de síntese utilizando espectros NIR e técnicas de reconhecimento de padrões supervisionados.

## 2 REVISÃO DE LITERATURA

### 2.1 Biodiesel

O biodiesel é um combustível biodegradável oriundo de matriz renovável, menos tóxico se comparado a combustíveis fósseis, tais como o diesel e a gasolina, podendo ser produzido a partir de uma reação de transesterificação de óleo vegetal ou gordura animal. É um combustível para uso em motores a combustão por compressão, podendo substituir parcial ou totalmente o óleo diesel (SENGO et al., 2010).

Uma gama de matérias-primas vem sendo utilizada para a produção de biodiesel, dentre elas soja, girassol, amendoim, mamona, dendê e canola (GERIS et al., 2007). Existem também espécies menos conhecidas sendo considerado para a produção de biodiesel, incluindo macaúba, buriti e licuri (BERGMANN et al., 2013).

Físico-quimicamente semelhante ao óleo diesel, o biodiesel é perfeitamente miscível e pode ser utilizado puro ou misturado ao diesel em quaisquer proporções em motores do ciclo diesel. As misturas binárias de biodiesel e óleo diesel de petróleo são designadas pela abreviação BX, onde X é a proporção em volume de biodiesel na mistura (COSTA NETO et al., 2000).

No cenário ambiental, o biodiesel apresenta vantagens claras sobre os combustíveis convencionais por serem provenientes de fontes renováveis, biodegradáveis, não tóxicos e reduzem significativamente a emissão de CO<sub>2</sub> e SOX durante o processo de combustão (CORDEIRO et al., 2011).

### 2.2 Misturas biodiesel/diesel

Em decorrência da crescente demanda por combustíveis de fontes renováveis e do potencial brasileiro em atender parte expressiva dessas necessidades (BALABIN e SAFIEVA, 2011), o Governo Federal lançou em 2004 o Programa Nacional de Produção e Uso do Biodiesel. Este programa interministerial teve como objetivo a implantação, de forma sustentável, da cadeia de produção e uso do biodiesel, visando geração de emprego e renda na agricultura familiar.

Em 2005, com a publicação da Lei Federal N° 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 esta lei, o percentual de biodiesel adicionado ao diesel aumentaria para 5% apenas em 2013 (BAKEAS et al., 2011). Entretanto, essa obrigatoriedade foi antecipada e em janeiro de 2010 a mistura B5 já estava sendo comercializada em todo o

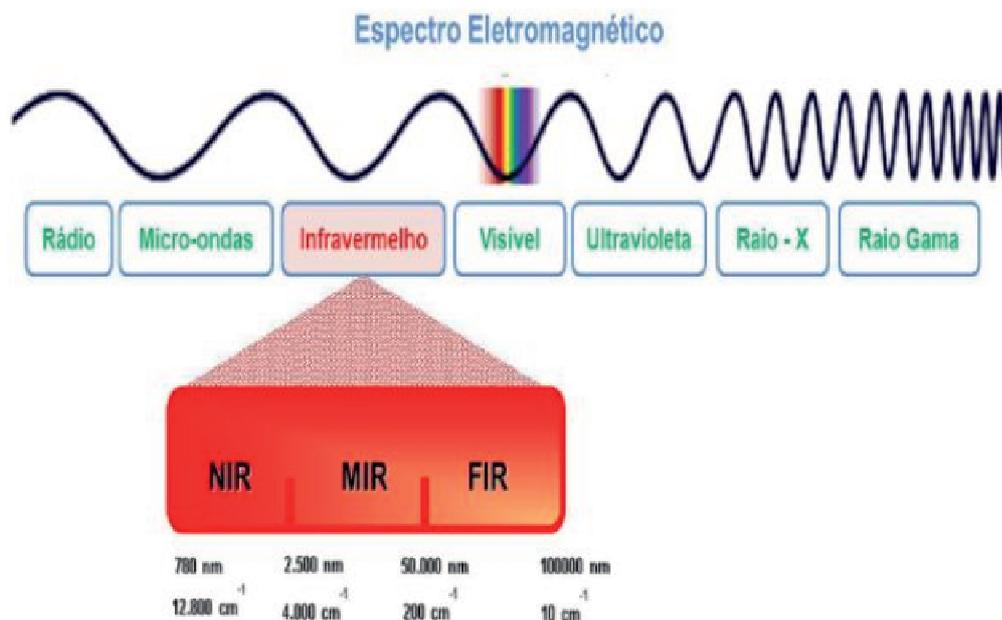
território brasileiro. Em 2015, a mistura passou a ser B7. Com a expansão da capacidade produtiva do biodiesel no Brasil, há uma tendência ao longo dos próximos anos da comercialização de misturas com teores cada vez mais elevados de biodiesel, mediante a autorização da ANP, podendo chegar a B20.

Apesar das normas de avaliação de qualidade do biodiesel estar estabelecida surge a necessidade de desenvolvimento de métodos expeditos, economicamente viáveis, que sejam não invasivos e não destrutivos para amostras, podendo ser feitas medidas *on-line*, *in-line* e *in situ*. Desta forma, a espectroscopia no infravermelho próximo surge alternativa viável ao desenvolvimento desses métodos.

### 2.3 Espectroscopia no Infravermelho Próximo (NIR)

A espectroscopia na região do infravermelho é uma técnica analítica que lida com os modos vibracionais das moléculas (RUBINSON e KENNETH, 2001), que utiliza a faixa de comprimentos de onda adjacente ao infravermelho médio e se estende até a região do visível (Figura 1). A região NIR do espectro eletromagnético pode ser definida como a faixa de comprimentos de onda de 780 a 2500nm. As bandas de absorção mais proeminentes que ocorrem nesta região estão relacionadas à sobretons e combinações de vibrações fundamentais de grupos funcionais –CH, –NH, –OH e –SH (PASQUINI, 2003).

**Figura 1** – Espectro de radiação eletromagnética e distribuição do Infravermelho



**Fonte:** Adaptado de BART, 2007.

A radiação no infravermelho próximo absorvida por uma molécula, faz seus átomos vibrarem com maior amplitude, de maneira similar a um oscilador diatômico. O espectro NIR contém informações relacionadas com a diferença entre as forças de ligações químicas, espécies químicas presentes, eletronegatividade e a ligação de hidrogênio. A luz pode interagir com a amostra por meio de reflexão, refração, absorção, espalhamento, difração e transmissão. A perda no sinal pode ocorrer por refletância difusa especular, espalhamento interno e uma completa absorção (SCAFI et al., 2001).

O uso da espectrometria NIR apresenta vantagens em relação às demais técnicas que exploram o espectro eletromagnético vibracional uma vez que as medidas nessa região exigem sistemas ópticos e detectores facilmente disponíveis, simplicidade relativa na instrumentação e a grande quantidade de compostos orgânicos possíveis de serem analisados. Em função de tais características, o uso da espectrometria NIR atualmente é uma realidade nos mais variados setores da cadeia produtiva, sendo empregada diretamente na linha de produção, no controle de qualidade de matérias primas e produtos, tais como produtos da indústria de fármacos (BLANCO e ALCALA, 2006; LANGERON et al., 2007), de polímeros (QU et al., 2006), de produtos petroquímicos (PRA E CHAMPAGNE, 2005) e agrícolas em geral (SZIDLOWSKI-ZANIER et al., 2003).

Apesar do grande potencial da espectroscopia no infravermelho próximo, o tratamento dos espectros requer procedimentos matemáticos e estatísticos adequados, suportados pela quimiometria.

#### 2.4 Quimiometria

A Quimiometria pode ser definida como a utilização de métodos matemáticos e estatísticos para o tratamento de dados químicos de forma a extrair uma maior quantidade de informações e melhores resultados analíticos (BLANCO e ALCALA, 2006). A diferença da Quimiometria para as demais ciências metrológicas reside nas características dos dados serem de origem química, que possuem menor quantidade de variáveis, os quais em sua maioria são controláveis, de fácil repetição nas condições de trabalho desejadas, etc.

A Quimiometria está dividida em algumas áreas principais muito pesquisadas e aplicadas atualmente, tais como processamento de sinais analíticos, planejamento e otimização de experimentos, classificação e reconhecimento de padrões, calibração multivariada, monitoramento e modelagem de processos multivariados e métodos de inteligência artificial (PIMENTEL et al., 1998). A Quimiometria pode ser aplicada em todas

as etapas de uma análise no sentido de melhorar o rendimento e dar melhor qualidade nos resultados.

## 2.5 Técnicas de reconhecimento de padrões

São técnicas responsáveis por identificar as semelhanças e diferenças entre as amostras, a partir do grau de similaridade entre suas propriedades físicas e químicas, comparando-as entre si. As técnicas quimiométricas de reconhecimento de padrões (RP) promovem os agrupamentos dos objetos, possibilitando, inclusive, a identificação de amostras anômalas (do inglês *outliers*). As técnicas de reconhecimento de padrões podem ser divididas entre as não supervisionadas e as supervisionadas (MASSART et al., 1997).

### 2.5.1 Técnicas de reconhecimento de padrões não supervisionadas

Os métodos de reconhecimento de padrões não supervisionados são utilizados quando não se tem informação anterior sobre as classes as quais pertencem os objetos, ou seja, uma avaliação prévia dos dados sem sofrer influência do rótulo das amostras é realizada. Dentre estes, os mais conhecidos são: Análise por Componentes Principais (PCA, do inglês *Principal Component Analysis*) e Análise Hierárquica de Agrupamentos (HCA, do inglês *Hierarchical Cluster Analysis*) (CORREIA e FERREIRA, 2007).

#### 2.5.1.2 Análise de Componentes Principais (PCA)

A Análise por Componentes Principais (PCA) é um dos métodos mais comumente empregados na análise de informações (FERREIRA, 2002).

O método PCA pode ser aplicado nas variáveis da matriz X que apresentam alto grau de colinearidade. Como consequência, informações redundantes e a pequena variabilidade de ruídos podem ser removidas (MARTENS e TORMOD, 1989).

Objetiva-se representar o conjunto de dados com um número reduzido de novas variáveis, chamadas componentes principais, variáveis latentes ou fatores (THOMAS, 1994).

Neste sentido, a natureza multivariada dos dados pode ser visualizada em poucas dimensões. A primeira componente principal (PC1) representa a máxima variância associada a todas as combinações lineares e cada componente sucessiva (PC2, PC3, PC4, PCn) representa a variância não representada por PC1, e assim sucessivamente (MARTENS e NAES, 1996).

. A Análise por Componentes Principais (PCA) é uma técnica muito utilizada e serve de base para a maioria das demais técnicas quimiométricas multivariadas, inclusive as técnicas de reconhecimento de padrões supervisionado.

### 2.5.2 Técnicas de reconhecimento de padrões supervisionadas

Os métodos supervisionados são utilizados quando se tem informação prévia sobre as classes e esta é usada para a construção do modelo. Dentre estas técnicas destacam-se: Modelagem Flexível Independente por Analogia de Classe (SIMCA), K-ésimo Vizinhos mais Próximos (KNN), Análise Discriminante Linear associado ao Algoritmo das Projeções Sucessivas (SPA-LDA) e Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA).

#### 2.5.2.1 Modelagem Independente e Flexível por Analogia de Classe (SIMCA)

O método SIMCA foi proposto por Wold (PONTES, 2009), sendo utilizado para classificação de amostras em conjuntos de dados com alta dimensionalidade. O SIMCA faz uso de componentes principais para localizar os objetos no espaço multidimensional. Em outras palavras, o espaço multidimensional fica delimitado através da construção do modelo PCA para cada categoria de amostras. Uma amostra será classificada como pertencente a uma dada classe previamente modelada se possuir características que a permitam ser inserida no espaço multidimensional do modelo construído (BEEBE, 1998; WOLD, 1976).

Matematicamente a amostra será considerada pertencente à classe se o valor do F calculado, razão entre o valor de  $z^2$ , pela a variância da classe (MOREIRA, 2007), for menor que o F crítico.

$$z^2 = x^2 + y^2$$

Onde  $z^2$  é igual à soma das distâncias entre a amostra desconhecida e o eixo da PC ( $y^2$ ), e a distância entre a projeção da amostra desconhecida na direção da PC e a fronteira da classe ( $x^2$ ).

#### 2.5.2.2 K-ésimo Vizinhos mais Próximos (KNN)

O classificador K-NN foi descrito por Cover e Hart (1967). É um método de classificação não paramétrico muito popular devido a sua simplicidade e razoável eficiência.

Seu desempenho depende fortemente da métrica usada para o cálculo da distância entre os pontos.

Supondo um conjunto D de amostras utilizadas no treinamento. Cada elemento de D é dado por  $(x_1, x_2, \dots, x_n, c)$ , sendo c a classe à qual o conjunto de características  $(x_1, x_2, \dots, x_n)$  pertence. Considerando uma nova amostra não classificada  $Y=(y_1, y_2, \dots, y_n)$ , para ser classificada calcula-se as distâncias de Y em relação a todas as amostras utilizadas no treinamento e considerando as k amostras de treinamento mais próximas a Y. A amostra Y será classificada na classe que ela está mais próxima.

A distância pode ser calculada por diversas métricas

- Euclidiana:  $d_{st}^2 = (x_s - y_t)(x_s - y_t)'$
- Mahalanobis:  $d_{st}^2 = (x_s - y_t)C^{-1}(x_s - y_t)'$
- Minkowski:  $d_{st}^2 = \sqrt[p]{\sum_{j=1}^n |x_j - y_j|^p}$

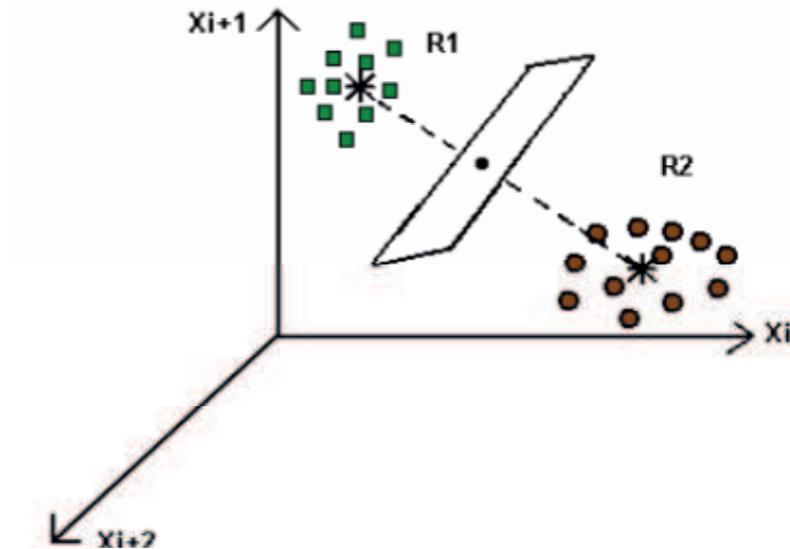
De forma geral é preciso fazer a normalização das características de entrada de modo que todas apresentem o mesmo intervalo de variação, que poderia influenciar no cálculo da distância.

### 2.5.2.3 Análise Discriminante Linear (LDA)

A Análise Discriminante Linear (LDA) é um dos métodos de reconhecimento de padrão mais utilizados, é uma técnica que consiste na determinação de funções discriminantes lineares as quais maximizam a variância entre as classes e minimizam a variância dentro das classes (MASSART, 1997). Uma das melhores formas de se obter uma função linear discriminante é ilustrada na Figura 2.

Ao contrário do SIMCA (MELLINGER, 1987), método de classificação consolidado para trabalhar em altas dimensões, a Análise Discriminante Linear (LDA) (CORGOZINHO et al., 2008) limita-se ao uso de dados com baixa dimensionalidade. Dessa forma, torna-se necessário a utilização de métodos de seleção de variáveis que são úteis para aumentar a capacidade preditiva dos modelos, eliminar variáveis não informativas e reter aquelas que, provavelmente, resulta em uma boa exatidão. Importante ressaltar que em LDA, se a amostra apresentar erro do Tipo I, conseqüentemente ocorrerá erro do tipo II (BRUNS e FAIGLE, 1985).

**Figura 2** - Representação esquemática da Análise Discriminante Linear.



Fonte: COSTA, 2015.

## 2.6 Ferramentas de diagnóstico da classificação

Para avaliar a eficiência de um classificador, se faz necessária a utilização de ferramentas de diagnóstico. A ferramenta de diagnóstico mais eficaz para isso é a matriz de confusão. A utilização dessa matriz consiste na comparação entre os dados que são utilizados no teste e os que são utilizados para o treinamento, indicando o percentual de acerto e erro de classificação. Na matriz de confusão, os dados são organizados em uma tabulação cruzada entre a classe predita pelo classificador e a classe real das amostras. A diagonal principal representa os acertos e os demais dados fazem referência aos erros de classificação (RODRIGUES et al, 2017).

Portanto, para ser considerado um bom classificador, a diagonal principal deverá apresentar altos valores enquanto os demais elementos da matriz de confusão deverão apresentar baixos valores (RODRIGUES et al, 2017).

### 2.6.1 Taxa de Falso Positivo

A taxa de falso positivo, erro do tipo II, representa a porcentagem de amostras que não pertencem a uma dada classe e que foram incorretamente classificadas como pertencentes em relação a todas as amostras definidas como não pertencentes a esta classe. A taxa de falso positivo pode ser calculada pela relação matemática abaixo:

$$TFP = \frac{FP}{FP + VN} * 100$$

Na qual FP é o número de falsos positivos e VN é o número de verdadeiros negativos.

### 2.6.2 Taxa de Falso Negativo

A Taxa de Falso Negativo, erro do tipo I, é a porcentagem de amostras não pertencentes a uma dada classe que foram incorretamente classificadas como não pertencentes em relação a todas as amostras definidas como pertencentes a esta classe. A Taxa de Falso Negativo é dada pela seguinte relação matemática:

$$TFN = \frac{FN}{FN + VP} * 100$$

Na qual FN é o número de falsos negativos e VP é o número de verdadeiros positivos.

### 2.6.3 Taxa de Sensibilidade

Consiste na porcentagem de amostras pertencentes a uma dada classe corretamente classificadas em relação a todas as amostras definidas como pertencentes a esta classe. Calcula-se através da relação matemática:

$$TSN = \frac{VP}{VP + FN} * 100$$

Na qual VP é o número de verdadeiros positivos e FN é o número de falsos negativos.

### 2.6.4 Taxa de Especificidade

Consiste na porcentagem de amostras não pertencentes a uma dada classe que foram corretamente classificadas em relação a todas as amostras definidas como não pertencentes a esta classe. Pode ser calculado pela relação matemática:

$$TES = \frac{VN}{VN + FP} * 100$$

Na qual VN é o número de verdadeiros negativos e FP é o número de falsos positivos.

### 2.6.5 Precisão Global

Consiste na porcentagem de todas as amostras corretamente classificadas em relação a todas as amostras. É dada pela relação:

$$PGL = \frac{VN + VP}{VP + VN + FP + FN} * 100$$

Na qual VN é o número de verdadeiros negativos, VP é o número de verdadeiros positivos, FN é o número de falsos negativos e FP é o número de falsos positivos

### 2.6.6 Característica de Operação do Receptor (ROC)

A curva ROC foi desenvolvida no contexto da detecção de sinais eletrônicos e problemas com radares, durante a segunda guerra mundial. Seu objetivo era quantificar a habilidade dos operadores dos radares (chamados originalmente de receiver operators) em distinguir um sinal de um ruído. Esta habilidade era chamada Receiver Operating Characteristic (ROC). Na década de 60, curvas ROC foram utilizadas em psicologia experimental e, nos anos 70, a metodologia amplamente se disseminou em vários ramos da pesquisa. (MARTINEZ, LOUSADA-NETO, PEREIRA, 2003).

Para SILVA (2006), ROC é uma técnica para visualizar, avaliar, organizar e selecionar classificadores baseado em seu desempenho. Para realizar estas análises, gráficos ROC podem mostrar o limiar entre taxas de acertos e alarmes falsos (taxas de erros) dos classificadores.

A área sob a curva ROC é uma medida resumo usual do desempenho de um teste, já que ela é estimada levando-se em consideração todas as Se e Es relativas a cada um dos valores  $t_0$  estipulados. Quanto maior a capacidade do teste em discriminar os indivíduos segundo estes dois grupos, mais a curva se aproximaria do canto superior esquerdo do gráfico, e a área sob a curva seria próxima de 1 (MARTINEZ, LOUSADA-NETO, PEREIRA, 2003)

### 3 METODOLOGIA

Os ensaios foram conduzidos no Laboratório de Química Analítica e Quimiometria (LQAQ) pertencente à Universidade Estadual da Paraíba (UEPB), Campus I, Campina Grande - PB.

#### 3.1 Reagentes e amostras de biodiesel

As amostras de biodiesel foram obtidas a partir de óleos. A reação de transesterificação via rota metílica e etílica se processou em temperatura ambiente (25°C), com catalise homogênea de 1% em relação a massa do óleo, com uma razão molar de 1:6 de óleo/álcool. O diesel puro foi cedido pela Petrobras (Cabedelo, Paraíba, Brasil). Foram preparadas misturas biodiesel/diesel utilizando 93% de diesel e 7% de biodiesel (B7). Sendo 60 amostras via rota metílica, 60 amostras via rota etílica e 60 amostras de misturas etílica e metílica, perfazendo um total de 180 amostras.

#### 3.2 Instrumentos e programas computacionais

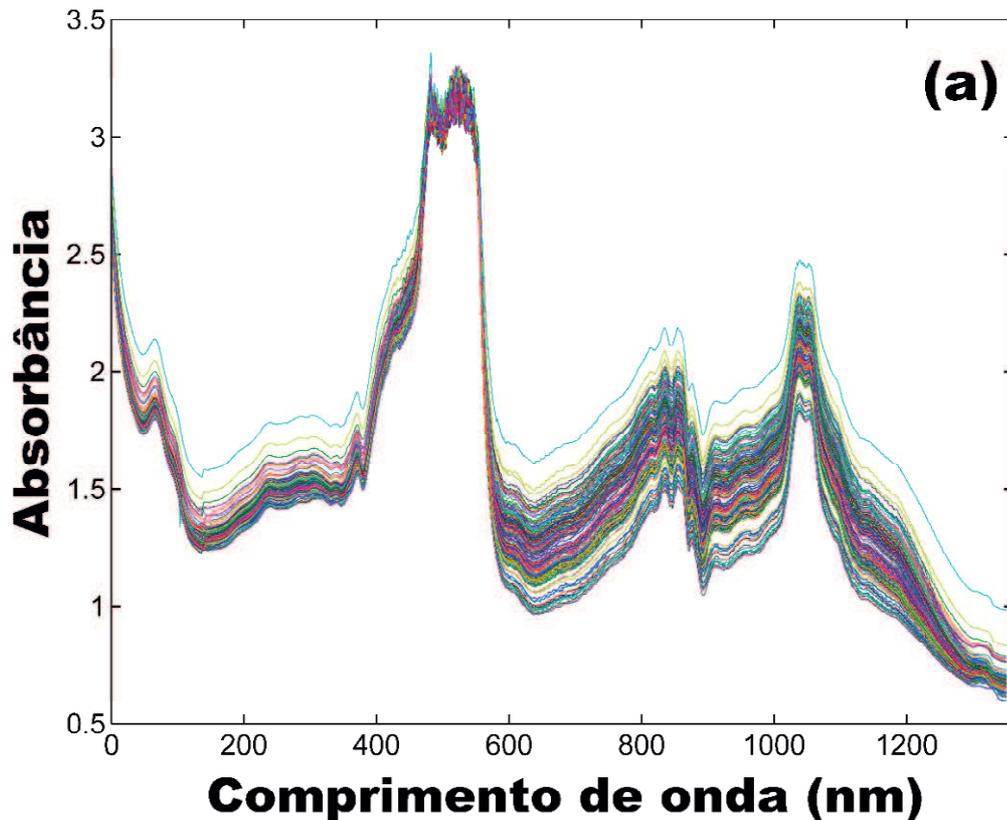
Os espectros das misturas foram registrados em triplicata no intervalo de 900 a 2300 nm usando um espectrofotômetro Lambda 750 Vis-NIR da Perkin Elmer, com cubeta de vidro de 1 cm de caminho óptico. Cada espectro foi medido como uma média de 32 varreduras a uma resolução de 1 nm. Os espectros foram derivados e pré-processados empregando o algoritmo de Savitzky-Golay com janela de 21 pontos e de segunda ordem polinomial. Análise de Componentes Principais foi utilizada para avaliar o poder discriminativo. O conjunto de amostras foi repartido em treinamento (75%) e teste (25%) pela aplicação do algoritmo Kernnard-Stone. (Kernnard-Stone, 1969)

PCA foi calculado utilizando o software Unscrambler 9.8. KS, SIMCA, KNN, PLS-DA e SPA-LDA foram realizados em Matlab 2011b.

#### 4 RESULTADOS E DISCUSSÃO

Os espectros apresentados na Figura 3a são das 180 amostras de misturas de biodiesel/diesel sendo 60 amostras de misturas com álcoois etílico e metílico, 60 amostras com apenas álcool etílico e 60 amostras com álcool metílico, na região de 900 a 2300 nm.

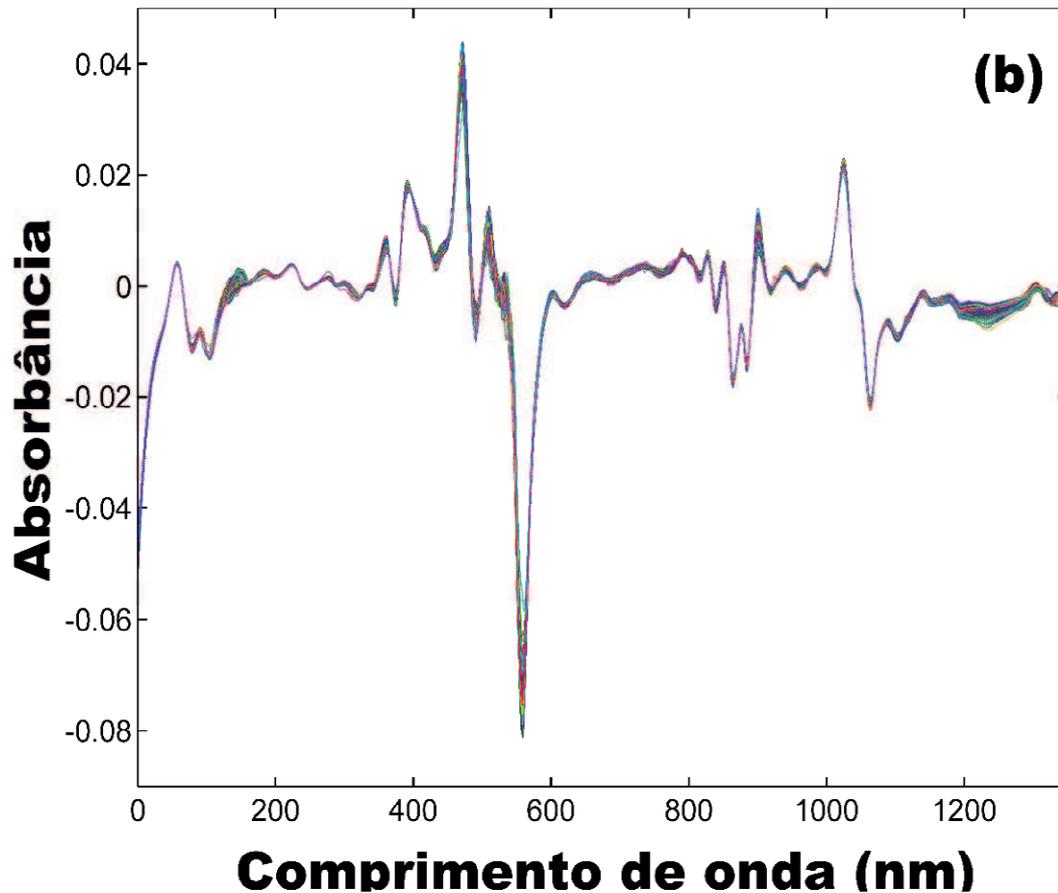
**Figura 3a** – Espectros Brutos das amostras de misturas biodiesel/diesel



Fonte: Autor.

Os dados foram pré-processados por primeira derivada pelo método de Savitzky-Golay, empregando um polinômio de segunda ordem e janela de 21 pontos (Figura 3b).

Figura 3b – Espectros derivados das amostras de misturas biodiesel/diesel

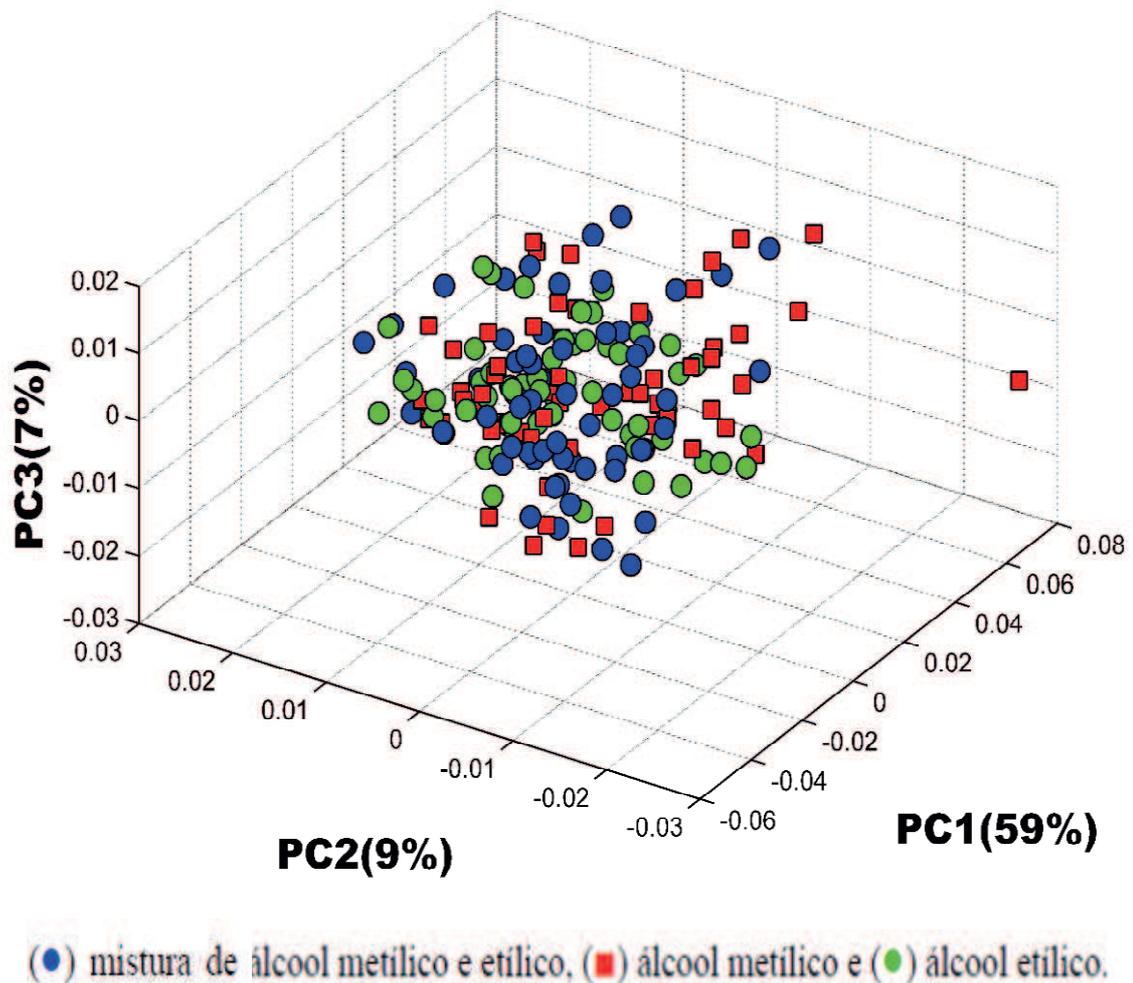


Fonte: Autor.

#### 4.1 Análise exploratória

A fim de avaliar a capacidade discriminante de espectros NIR, tanto em termos do tipo de álcool utilizado na reação de transesterificação uma análise de componentes principais foi realizada. A Figura 3 mostra os gráficos de scores (PC1 x PC2 x PC3) resultantes da aplicação das PCA para os espectros derivado. O estudo foi realizado tendo em conta três classes de mistura biodiesel/diesel para a discriminação da rota de síntese: (●) mistura de álcool metílico e etílico, (■) álcool metílico e (●) álcool etílico.

**Figura 4** – Gráfico de scores (PC1 x PC2 x PC3).



**Fonte:** Autor.

Utilizando as projeções em 3D os gráficos de escores apresentam forte sobreposição entre as amostras e nenhuma tendência de separação. Isso faz com que seja necessária a utilização de métodos classificatórios supervisionados para contornar este problema e, conseqüentemente classificar as amostras de biodiesel/diesel de acordo com a sua rota de síntese.

#### 4.2 Classificação

A construção dos modelos multivariados de classificação foi realizada utilizando um conjunto de treinamento (75% das amostras estudadas). Cada modelo foi validado utilizando a técnica de validação cruzada. Um conjunto de teste (25% das amostras estudadas) foi então usado para a avaliação final de dados e comparação com os modelos de classificação. O

desempenho dos modelos foi avaliado pela precisão, que é definida utilizando a proporção de amostras no conjunto de teste corretamente afetados em suas respectivas classes. Na próxima seção serão discutidos os resultados dos modelos estudados: SIMCA, KNN, SPA-LDA e PLS-DA.

#### 4.3 Modelagem Independente e Flexível por Analogia de Classe (SIMCA)

O resultado obtido pelo SIMCA foi 72,59 e 88.89% de precisão de classificação nos conjuntos de treinamento e teste, respectivamente, para a classificação com relação ao tipo de álcool (metílico e etílico) usado na reação de transesterificação. Para alcançar este resultado, o número ideal de componentes principais utilizados para cada classe era: 2 PC para misturas biodiesel/diesel com álcool metílico e etílico, 2 PC para misturas biodiesel/diesel com álcool metílico e um PC para misturas biodiesel/diesel com álcool etílico. Os erros de classificação para cada uma das classes do modelo SIMCA estão descritos na Tabela 1.

**Tabela 1** – Classificação quanto à rota de síntese - SIMCA

<b>Modelagem Suave Independente Por Analogias de Classes - (SIMCA)</b>						
<b>Classe</b>	<b>Treinamento</b>			<b>Teste</b>		
	<b>Et/Mt</b>	<b>Me</b>	<b>Et</b>	<b>Et/Mt</b>	<b>Me</b>	<b>Et</b>
<b>Et/Me</b>	33	8	4	14	-	1
<b>Me</b>	7	33	5	-	15	-
<b>Et</b>	3	10	32	-	4	11
<b>Sensibilidade</b>	73.3	73.3	71.1	93.3	100	73.3
<b>Especificidade</b>	88.8	80	90	100	86.6	96.6
<b>Classificação</b>	72.59%			88.89%		

Fonte: Autor.

#### 4.4 KNN - Vizinhos mais próximos

O melhor resultado para o modelo KNN foi obtido utilizando 7 para vizinhos mais próximos, atingindo assim uma exatidão de classificação de 80.74% e 91.1%. Os erros de classificação para cada uma das classes do modelo KNN estão descritos na Tabela 2.

**Tabela 2** – Classificação quanto à rota de síntese - KNN

<b>K- vizinhos mais próximos - KNN</b>						
<b>Classe</b>	<b>Treinamento</b>			<b>Test</b>		
	<b>Et/Mt</b>	<b>Me</b>	<b>Et</b>	<b>Et/Mt</b>	<b>Me</b>	<b>Et</b>
<b>Et/Me</b>	37	2	6	13	-	2
<b>Me</b>	4	32	9	-	14	1
<b>Et</b>	-	5	40	-	1	14
<b>Sensibilidade</b>	82.2	71.1	88.89	86.67	93.33	93.33
<b>Especificidade</b>	95.56	92.22	83.33	100	96.67	90
<b>Classificação</b>	80.74%			91.1%		

Fonte: Autor.

4.5 Análise Discriminante Linear associado ao Algoritmo das Projeções Sucessivas (SPA-LDA).

. O SPA-LDA utilizando 12 variáveis alcançou um resultado 90.37 e 93.3% de precisão da classificação nos conjuntos de treinamento e teste, respectivamente, para a classificação da rota de síntese. Os erros de classificação para cada uma das classes do modelo SPA-LDA estão descritos na Tabela 3.

**Tabela 3** – Classificação quanto à rota de síntese – SPA-LDA

<b>Análise Discriminante Linear associado ao Algoritmo das Projeções Sucessivas (SPA- LDA)</b>						
<b>Classe</b>	<b>Treinamento</b>			<b>Test</b>		
	<b>Et/Mt</b>	<b>Me</b>	<b>Et</b>	<b>Et/Mt</b>	<b>Me</b>	<b>Et</b>
<b>Et/Me</b>	40	2	3	14	-	1
<b>Me</b>	6	38	1	2	13	-
<b>Et</b>	1	-	44	-	-	15
<b>Sensibilidade</b>	88.89	84.44	97.78	93.3	86.67	100
<b>Especificidade</b>	92.22	97.78	95.56	93.3	100	96.67
<b>Classificação</b>	90.37%			93.33%		

Fonte: Autor.

#### 4.6 Análise Discriminante pelo método Dos Mínimos Quadrados Parciais (PLS-DA)

O modelo PLS-DA utilizando 19 variáveis latentes apresentou o melhor resultado em termos das figuras de mérito avaliadas: 100% de classificação correta (precisão), 100% de sensibilidade e 100% de especificidade para cada parâmetro na classe de interesse. O valor de limiar é previsto entre 0 e 1, com base no teorema de Bayes, a fim de minimizar os erros na predição da classe através de um número adequado de variáveis latentes (LV). Tabela 4.

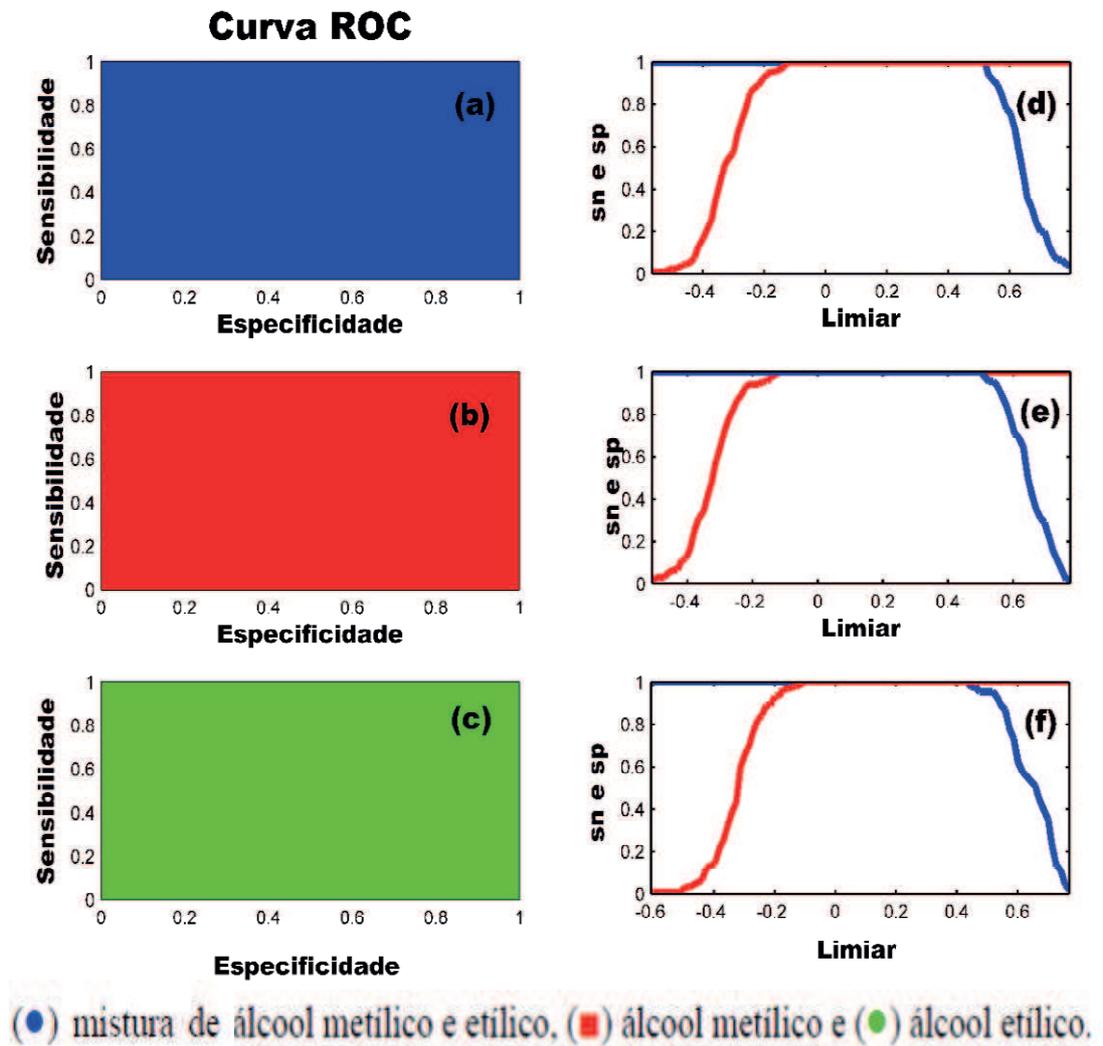
**Tabela 4** – Classificação quanto à rota de síntese – PLS-DA

<b>Análise Discriminante pelos Mínimos Quadrados Parciais (PLS-DA)</b>						
Class	Treinamento			Test		
	Et/Mt	Me	Et	Et/Mt	Me	Et
<b>Et/Me</b>	45	-	-	15	-	-
<b>Me</b>	-	45	-	-	15	-
<b>Et</b>	-	-	45	-	-	15
<b>Sensibilidade</b>	100	100	100	100	100	100
<b>Especificidade</b>	100	100	100	100	100	100
<b>Classificação</b>	100%			100 %		

Fonte: Autor.

O número de variáveis latentes escolhidas para os modelos PLS-DA seguiu o critério do menor erro de previsão em validação cruzada e avaliação da variância explicada no X e blocos y. O limiar ótimo para cada classe foi determinado utilizando o ponto em que a sensibilidade e especificidade têm o mesmo valor numérico de acordo com as características do receptor de operação (ROC) Figura 5.

Figura 5 – Característica do receptor de operação (ROC).



Fonte: Autor

A Figura 5abc corrobora com as informações apresentadas na Tabela 4 e indicam a classificação das amostras, permitindo 100% de acertos na previsão das amostras em cada classe das misturas biodiesel/diesel com respeito ao álcool utilizado na reação de transesterificação.

Deste modo, os resultados aqui apresentados, indicam que a associação da espectroscopia no infravermelho próximo (NIR) e PLS-DA é uma ferramenta adequada para a classificação de amostras de misturas biodiesel/diesel de acordo com a rota de síntese, além de apresentar uma vantagem extra, ser uma técnica bastante simples e rápida, o que permite seu uso como ferramenta capaz de contribuir para o controle fiscal de comercialização destes combustíveis.

## 5 CONCLUSÃO

Neste trabalho foi apresentada uma metodologia baseada na utilização de Espectroscopia NIR e técnicas de reconhecimento de padrões para classificação de blendas de biodiesel/ diesel de acordo com a sua rota de síntese.

Apesar do bom desempenho dos modelos de classificação SIMCA, KNN e SPA-LDA, para o problema de classificação abordado no presente estudo, os modelos PLS-DA forneceram melhores resultados em termos de taxa de classificação correta (precisão), sensibilidade e especificidade para ambos os conjuntos de treinamento e teste, respectivamente, o que indica que as variáveis selecionadas pelo algoritmo contém informação adequada para classificação das blendas de biodiesel/diesel de acordo com o álcool de origem.

A partir do estudo foi possível perceber que espectroscopia NIR associadas às ferramentas quimiométricas empregadas obtiveram uma boa taxa de classificação correta para a classificação de blendas biodiesel/diesel de acordo com a rota de síntese. O melhor resultado foi para o modelo PLS-DA, que apresentou uma taxa de classificação correta, especificidade e seletividade de 100%. Portanto, verifica-se que a utilização de PLS-DA acoplada com espectroscopia NIR apresenta-se como uma alternativa viável para o controle de qualidade de biodiesel em termos de seus álcoois empregados na síntese e está de acordo com os princípios da Química Verde.

## REFERÊNCIAS

ABBASZAADEH; A., GHOBADIAN; B., OMIDKHAH; M.R., NAJAFI;G. Current biodiesel production technologies: A comparative review. **Energy Conversion and Management**, v. 63, p. 138–148, 2012.

ARANSIOLA; E.F., OJUMU; T.V., OYEKOLA; O.O., MADZIMBAMUTO; T.F., IKHU-OMOREGBE; D.I.O. A review of current technology for biodiesel production: State of the art. **Biomass and Bioenergy**, n. 61, p. 276 – 297, 2014.

BALABIN, R. M; SAFIEVA, R. Z; LOMAKINA, E.I; Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. **Fuel**, v. 90, n.1, p. 2007–2015, 2011.

BALABIN, R.M.; SAFIEVA, R.Z. Biodiesel classification by base stock type (vegetable oil) using near infrared spectroscopy data, **Analytica Chimica Acta**, v.89, n.1, p. 190–197, 2011.

BALAT , M. Potential alternatives to edible oils for biodiesel production – A review of current work. **Energy Conversion and Management**, v. 52, p. 1479–1492, 2011.

BAKEAS, E.; KARAVALAKIS, G.; STOURNAS, S. Biodiesel emissions profile in modern diesel vehicles. Part 1: effect of biodiesel origin on the criteria emissions. **Science Total Environmental**, v.409, n.1, p.1670–1676, 2011.

BARKER,M.; RAYENS, W. Partial least square for discrimination. **Journal of Chemometrics**, v. 17, p 166 – 173, 2003.

BARTH, A., Infrared Spectroscopy of Proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*,v.1767, p. 1073–1101, 2007.

BASSBASI, M.; PLATIKANOV, S.; TAULER, R.; OUSSAMA, A. FTIR-ATR determination of solid non fat (SNF) in raw milk using PLS and SVM chemometric methods. **Food Chemistry**, v. 146, n. 0, p. 250-254, 2014

BERGMANN, J.C.; TUPINAMBÁ, D.D.; COSTA, O.Y.A.; ALMEIDA, J.R.M.; BARRETO, C.C.; QUIRINO, B.F. Biodiesel production in Brazil and alternative biomass feedstocks. **Renewable and Sustainable Energy Reviews**, v.21, n.1, p.411-420, 2013.

BEEBE, K.R.; PELL, R.J.; SEASHOLTZ, B. **Chemometrics – A Practical Guide**. New York: Wiley, 1998.

BLANCO, M.; ALCALA, M. Simultaneous quantitation of five active principles in a pharmaceutical preparation: Development and validation of a near infrared spectroscopic method. **European Journal of Pharmaceutical Sciences**, v.27, n.1, p.280-284, 2006.

BRUNS, R. E.; FAIGLE, J. F. G., Quimiometria. **Química Nova**, n. 8, p. 84, 1985.

CORDEIRO, C. S.; SILVA, F. R.; WYPYCH, F.; RAMOS, L. P. Catalisadores heterogêneos para a produção de monoésteres graxos (Biodiesel). **Química Nova**, v.3, n.1, p. 477-486, 2011.

CORGOZINHO, C. N. C.; PASA, V. M. D.; BARBEIRA, P. J. S.; **Talanta**. v. 76, p. 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**, v.30, n.2, p.481-487, 2007.

COSTA; G. B. A, FERNANDES; D.D.S., ALMEIDA; V.E., ARAÚJO; T.S.P. B., MELO; J.P. DINIZ; P.H.G.D., VÉRAS; G. Digital image-based classification of biodiesel. **Talanta**, n.139, p. 50-55, 2015.

COSTA; G. B. **Classificação multivariada de biodiesel B100 e B5 usando imagens digitais**. 2015. 80 f. Dissertação (Mestrado em Ciências Agrárias) – Universidade Estadual da Paraíba, Campina Grande, 2015.

COSTA NETO, P. R.; ROSSI, L. F. S.; ZAGONEL, G. F.; RAMOS L. P. Produção de Biocombustível Alternativo ao Óleo Diesel através da Transesterificação de Óleo de Soja Usado em Frituras. **Química Nova**, v.23, n.1, p.532-537, 2000.

COVER, T.M., HART, P.E. Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, v. 13, p. 21-27. 1967.

CREMONEZ, P.A.; FEROLDI, M.; FEIDEN, A.; TELEKEN, J.G.; GRIS, D.J.; DIETER, J.; ROSSI, E.; ANTONELLI, J. Currents cenário and prospects of use of liquid biofuels in South America. **Renewable and Sustainable Energy Reviews**, v. 43, p. 352–362, 2015.

FERREIRA, M.C. Multivariate QSAR. **Journal Brazilian Chemical Society**, São Paulo, v.13, n.6, p.742-753, 2002.

FERRARI, R. A. (s.d.). Biodiesel de Soja - Taxa de conversão em ésteris etílicos, caracterização físico química em gerador de energia. **Química Nova**, v. 28, n. 1, p. 19-23, 2005.

GHASEMI-VARNAMKHAJASTI, G.; MOHTASEBI, S.S.; MENDEZ, M.L.R, GOMES, A.A.; ARAÚJO, M.C.U.; GALVÃO, R.K.H. Screening analysis of beer ageing using near infrared spectroscopy and the Successive Projections Algorithm for variable selection. **Talanta**, v. 89, p. 286–291, 2012.

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**, v.30, n.3, p. 1369-1373, 2007.

INSAUSTI, M.; GOMES, A.A.; CRUZ, F.V.; PISTONESI, M.F.; ARAUJO, M.C.U.; GALVÃO, R.K.H.; PEREIRA, C.F.; BAND, B.S.F. Screening analysis of biodiesel feedstock using UV–vis, NIR and synchronous fluorescence spectrometries and the successive projections algorithm **Talanta**, v. 97, n.1, p.579-583, 2012.

KENNARD, R.W.; STONE, L. A. Computer Aided Design of Experiments. **Technometrics**, v. 11, p.137-148, 1969.

KNOTHE; G. Analyzing Biodiesel: Standards and Other Methods. **The Journal of the American Oil Chemists' Society**, v. 83, n. 10, p 823 – 833, 2006.

LANGERON, Y.; DOUSSOT, M.; HEWSON, D. J.; DUCHENE, J.; Classifying NIR spectra of textile products with kernel methods. **Engineering Applications of Artificial Intelligence**, v. 20, p. 415, 2007.

LEE; H.V., YUNUS; R., JUAN; J.C., TAUFIQ-YAP; Y.H. Process optimization design for jatropha-based biodiesel production using response surface methodology. **Fuel Processing Technology**, v.92, p. 2420-2428, 2011.

MARTENS H.; TORMOD N. **Multivariate Calibration**, John Wiley & Sons, New York, 1996.

MARTINEZ, E.Z.; LOUZADA-NETO, F.; PEREIRA, B.B. A curva ROC para testes diagnósticos. **Cadernos Saúde Coletiva**, v. 11, n. 1, p 7-37, 2003.

MASSART, D.L.; VANDEGINSTE, B.G.M.; BUYDENS, S.J.; LEWI, P.J.; SMEYERS-VERBEKE, J. **Handbook of Chemometrics and Qualimetrics**: Parte B, Elsevier, 1997.

MAZIVILA, S.J.; MITSUTAKE, H.; SANTANA, F.B.; GONTIJO, L.C.; SANTOS, D.Q.; NETO, W.B. Fast Classification of Different Oils and Routes Used in Biodiesel Production Using Mid Infrared Spectroscopy and PLS-DA. **Journal of the Brazilian Chemical Society**, v. 26, p. 642 -648, 2015.

MELLINGER, M. Multivariate data analysis: its methods. **Chemometrics and Intelligent Laboratory Systems**, v. 2, n. 1-3, p. 29-36, 1987.

MOREIRA, E. D. T. **Classificação de Cigarros Usando Espectrometria NIRR e Métodos Quimiométricos de Análise**. 2007.83f Dissertação (Mestrado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba. 2007.

MURUGESAN, A.; UMARANI, C.; CHINNUSAMY, T.R.; KRISHNAN, M.; SUBRAMANIAN, R.; NEDUZCHEZHAIN, N. Production and analysis of bio-diesel from non-edible oils—A review. **Renewable and Sustainable Energy Reviews**, n. 13, p. 825–834, 2009.

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

PIMENTEL, M. F.; SALDANHA, T. C. B.; ARAÚJO, M. C. U. Effects of experimental design on calibration curve precision in routine analysis, **Journal of Automatic Chemistry**, v.20, n.1, p.9-11, 1998.

PONTES, M. J. C. **Algoritmo das projeções sucessivas para a seleção de variáveis espectrais em problemas de classificação**. 2009.123f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Federal de João Pessoa. 2009.

PRA, L. D.; CHAMPAGNE, S. D. A comparative study of the electronic structure and spectroelectrochemical properties of electrosynthesized polyaniline films and nanotubes. **Thin Solid Films**, p. 479:321, 2005.

QU, H.B.; YANG, H.L.; CHENG, Y.Y. Fast and nondestructive discrimination of donkeyhide glue by near-infrared spectroscopy, **Spectroscopy and spectral analysis**, v. 26 p. 60, 2006.

RODRIGUES, D. A.; SANTOS, G. P.; FERNANDES, M. C.; SANTOS, J. C.; FREITAS, F. N. C.; FILHO, P. P. R. Classificação automática do tipo de ferro fundido utilizando reconhecimento de padrões em imagens de microscopia. **Revista Matéria**, v. 22, n. 3, 2017.

RUBINSON, K.A. Análisis Instrumental, **Prentice Hall**, 2001.

SCAFI, S.H.F., PASQUINI, C.; Identification of Counterfeit Drugs Using Near- infrared Spectroscopy for Identification of Falsified, **The Analyst**, v.126, p. 2218-2224, 2001.

SENGO, I.; GOMINHO, J.; D'OREY, L.; MARTINS, M.; DUARTE, E. A.; PEREIRA, H.; DIAS, S. F. Response surface model and optimization of biodiesel production from Cynara Cardunculus Oil. **European Journal of Lipid Science and Technology**, 2010.

SILVA JUNIOR, G.B.; DAHER, E.F.; MOTA, R.M.S.; MENEZES, F.A. Risk factors for death among critically ill patients with acute renal failure. **São Paulo Medical Journal**, v. 124, n. 5, 2006.

SILVA, G.W.B; GOMES, A.A; SILVA, P; COSTA, G.B; FERNANDES, D.D.S; FONTES, M.M; VERA, S.G. Biodiesel/Diesel Blends Classification with Respect to Base Oil Using NIR Spectrometry and Chemometrics Tools. **Journal of the American Oil Chemists' Society**, v.89, n.1, p.1165-1171, 2012.

SZYDLOWSKI-ZANIER, N.; BERGER, M.; WAHL, F.; GUILLAUME, D. Performance of a near infrared spectrometer equipped with an autosampling accessory. **Journal of near infrared spectroscopy**, 11:83, 2003.

TAKASE; M., ZHAO; T., ZHANG; M., CHEN; Y., LIU; L., YANG; L., WU; X. An expatiated review of neem, jatropha, rubber and karanja as multipurpose non-edible biodiesel resources and comparison of their fuel, engine and emission properties. **Renewable and Sustainable Energy Reviews**, v. 43, p. 495–520, 2015.

THOMAS, E.V. A primer on multivariate calibration. **Analytical Chemistry**, v.66, p.795A-804A 1994.

VASUDEVAN; P. T., FU; B., Environmentally Sustainable Biofuels: Advances in Biodiesel Research. **Waste Biomass Valor**, v.1; p. 47- 63, 2010.

VERAS, G.; GOMES, A. A.; SILVA, A. C.; BRITO, A. L. B.; ALMEIDA, P. B. A.; MEDEIROS, E.P. Classification of biodiesel using NIR spectrometry and multivariate techniques. **Talanta**, v.83, n.1, p. 565-568, 2010.

VERAS, G.; BRITO, A.L.B.; SILVA, A.C.; COSTA, G.B.; FELIX, C.N.; FERNANDES, D.D.S.; FONTES, M.M. Classificação de biodiesel na região do visível. **Química Nova**, v.35, n. 2, p. 315–318, 2012.

YUSUF; N.N.A.N., KAMARUDIN; S.K., YAAKUB; Z. Overview on the current trends in biodiesel production. **Energy Conversion and Management**, v. 52, p. 2741–2751, 2010.

WOLD, S. Pattern recognition by means of disjoint principal component models, *Pattern Recognition, handbook*. Capítulo 33. p. 203 – 241. 1976.