



**UNIVERSIDADE ESTADUAL DA PARAÍBA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS AGRÁRIAS**

DAVID DOUGLAS DE SOUSA FERNANDES

**ESPECTROSCOPIA UV-VIS PARA AVALIAÇÃO DE BIODIESEL E MISTURAS
BODIESEL/DIESEL**

**CAMPINA GRANDE
2013**

DAVID DOUGLAS DE SOUSA FERNANDES

**ESPECTROSCOPIA UV-VIS PARA AVALIAÇÃO DE BIODIESEL E MISTURAS
BODIESEL/DIESEL**

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Agrárias da Universidade Estadual da Paraíba/Embrapa Algodão, como parte das exigências para obtenção do título de Mestre em Ciências Agrárias / Área de Concentração: Energias Renováveis e Biocombustíveis

Orientador: Prof. Dr. José Germano Vêras Neto

**CAMPINA GRANDE
2013**

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA CENTRAL-UEPB

F363e Fernandes, David Douglas de Sousa.
Espectroscopia UV_VIS para avaliação de biodiesel e misturas biodiesel/diesel. [manuscrito] / David Douglas de Sousa Fernandes. – 2013.

75 f. : il. color.

Digitado
Dissertação (Mestrado em Ciências Agrárias),
Centro de Ciências Humanas e Agrárias,
Universidade Estadual da Paraíba, 2013.

“Orientação: Prof. Dr. José Germano Vêras Neto,
Pró-Reitoria de Pós-Graduação e Pesquisa”

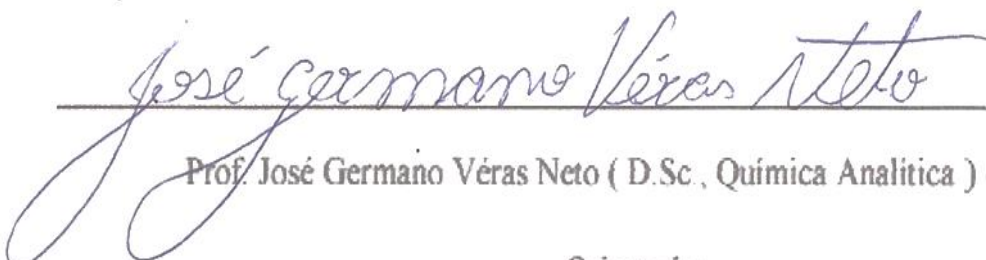
1. Fontes renováveis. 2. Biodiesel. 3.
Espectrometria. I. Título.

21. ed. CDD 581.3

ESPECTROSCOPIA UV-VIS PARA AVALIAÇÃO DE BIODIESEL E MISTURAS BIODIESEL/DIESEL

Dissertação apresentada ao Programa de Pós-Graduação em Ciências Agrárias da Universidade Estadual da Paraíba/Embrapa Algodão, como parte das exigências para obtenção do título de Mestre em Ciências Agrárias / Área de Concentração: Energias Renováveis e Biocombustíveis.

Aprovada em 18 de Fevereiro de 2013



Prof. José Germano Vêras Neto (D.Sc. , Química Analítica) - UEPB

Orientador



Prof. Marcelo Fabian Pistonesi (D.Sc. Química Analítica) UNS- AR



Prof. Carlos Henrique Salvino Gadelha Meneses (D.Sc Ciências Agrárias) - UEPB

A minha família, por sempre terem se dedicado e me ajudado ao máximo. Dedico.

AGRADECIMENTOS

A Deus, pela vida, pela saúde e discernimento para fazer este mestrado.

A minha Mãe/Pai, Mércia Oliveira de Sousa e a minha irmã, Danyhelem de Sousa Fernandes, por sempre se dedicarem ao máximo na hora que sempre precisei, eternamente agradecido.

A minha grande amiga, Herika Pereira Rodrigues e companheira que apoiou muito nos momentos difíceis.

Ao Prof. Dr. José Germano Vêras Neto e a Profa. Dra. Ana Cláudia Medeiros pela confiança depositada, por todos os seus ensinamentos, pelas oportunidades de sempre aprender um pouco mais e principalmente por construir uma amizade verdadeira.

A todos os amigos do Laboratório de Química Analítica e Quimiometria (LQAQ), Priscila (PC1), Gean (PC2), Marcelo (PC3), Clediano (PC4), Adriano (DF1), Valber (DF2), Alan (DF3), Katiane, Jéssika, Victor Hugo, Gildo, Ana Paula e Bruno pelos momentos únicos convividos e pela amizade construída.

A todos os amigos que fizeram parte Laboratório de Química Analítica e Quimiometria (LQAQ), Adenilton, Anna Luiza, Lorena, Everton e Welma, Odilon, Wellington e Janaina, por momentos inesquecíveis e pela amizade construída.

Aos meus amigos, Georgiano, Luzia Maria, Kátia Veloso e Eliane Souza por momentos inesquecíveis e pela amizade construída.

A todos os amigos do Laboratório de Desenvolvimento e Ensaio em Medicamentos (LABDEM), Deysiane, Ravelly, Thiago, Felipe, Fernando, Monik, Lianne, Elaine, Jocimar, Davy Magaiver, Fernanda e Aline.

Aos companheiros no mestrado, Genelício, Pollyne, Milena, Taíza, Gerkson, Monaliza, Emanuelle, Germana, Patrícia, Darlene e Klerisson pela amizade construída.

Aos meus grandes amigos da época da graduação, Stefane Rodrigo, Josué, Eduardo, Aglaison, Cosme e Flávio.

Aos professores do curso de Química Industrial, Licenciatura em Química e do Mestrado pelo incentivo e compreensão.

Ao Programa de Pós-Graduação em Ciências Agrárias da UEPB/EMBRAPA pela oportunidade. E a Capes pela bolsa de estudo concedida.

“Não há figura de mérito que consiga expressar a gratidão que tenho por todos vocês”.

“É melhor tentar e falhar,
que preocupar-se e ver a vida passar;
é melhor tentar, ainda que em vão,
que sentar-se fazendo nada até o final.

Eu prefiro na chuva caminhar,
que em dias tristes em casa me esconder.

Prefiro ser feliz, embora louco,
que em conformidade viver ...”

Martin Luther King

RESUMO

O biodiesel é um combustível proveniente de fontes renováveis e pode ser obtido desde óleos vegetais, gordura animal até óleos residuais. No Brasil a comercialização do biodiesel ocorre em forma de misturas, sendo adicionados 5% ao diesel (B5). Para ser comercializado o biodiesel deve obedecer a uma série de parâmetros estabelecidos em normas utilizadas pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP). Contudo metodologias alternativas estão sendo desenvolvidas com o propósito no controle e na qualidade do biodiesel. Objetivou-se com este trabalho a construção de modelos multivariados capazes em primeiro momento de classificar amostras de biodiesel produzidas a partir dos óleos de origens distintas, em segundo momento de prever o teor de biodiesel misturado ao diesel, e por fim classificar misturas biodiesel/diesel adulteradas com óleo de soja. Em todos os casos utilizou-se a espectroscopia de absorção molecular UV-Vis. Análise exploratória foi utilizada para determinar o comportamento das amostras em estudo, para tal utilizou-se a Análise por Componentes Principais (PCA) e a Análise de Agrupamentos Hierárquicos (HCA). Para classificar as amostras de biodiesel e de misturas biodiesel/diesel foram utilizadas técnicas de reconhecimento supervisionadas tais como a Modelagem Suave Independente de Analogias entre Classes (SIMCA) e Análise Discriminante Linear associada a técnica de seleção de variáveis Algoritmo das Projeções Sucessivas (SPA-LDA). Para prever a quantidade de biodiesel misturado ao diesel foi empregada seleção de variáveis Algoritmo das Projeções Sucessivas (SPA) e Stepwise (SW) associados a técnica de Regressão Linear Múltipla (MLR) e comparados os valores obtidos a técnica dos Mínimos Quadrados Parciais (PLS). Com base nos resultados obtidos é possível realizar a classificação das amostras de biodiesel a partir do óleo de origem e das amostras contaminadas com óleo de soja com 100% de êxito. Na predição do teor de biodiesel em misturas biodiesel/diesel os resultados foram mais ajustados utilizando o SPA-MLR, onde o erro médio de predição foi de 0,57 %.

Palavras-Chave: Fontes Renováveis, Controle de Qualidade, Calibração Multivariada, Seleção de Variáveis.

ABSTRACT

Biodiesel is a fuel derived from renewable sources and can be obtained from vegetable oils, animal fat until residual oils. In Brazil commercialization of biodiesel occurs in blends, being added 5 % to diesel (B5). Biodiesel to be marketed must follow a series of standards established by Brazilian National Agency of Petroleum, Natural Gas and Biofuels (ANP). However alternative methodologies are being developed for the purpose and quality control of biodiesel. The objective of this work is to build multivariate models, in first moment to classify samples of biodiesel produced from different oils, in the second moment to predict the amount of biodiesel blended with diesel, and finally to classify mixtures biodiesel/diesel adulterated with soybean oil. In all cases molecular UV-Vis spectroscopy was used. An exploratory analysis was used to determine the behavior of the samples under study, for this we used the Principal Component Analysis (PCA) and Hierarchical Cluster Analysis (HCA). To classify samples biodiesel and blends biodiesel/diesel recognition techniques were used such as Supervised Independent Modeling of Soft Analogies between Classes (SIMCA) and Linear Discriminate Analysis technique associated with variable selection Successive Projections Algorithm (SPA-LDA). To predict the amount of biodiesel blended with diesel was used variable selection Successive Projections Algorithm (SPA) and Stepwise (SW), techniques associated with Multiple Linear Regression (MLR) and compared the values with Partial Least Squares (PLS). Based on the results obtained, it is possible to carry out the classification of samples of biodiesel from the oil source and the samples contaminated with soybean oil with 100% success. For predicting biodiesel content in blends biodiesel/diesel results were further adjusted using the SPA-MLR, where the average error of prediction was 0.57 %.

Keywords: Renewable Sources, Quality Control, Multivariate Calibration, Selection of Variables.

LISTA DE FIGURAS

Figura1: Representação da reação de transesterificação	19
Figura 2: Radiação absorvente e a cor complementar.	22
Figura 3: Organização matricial dos dados multivariados de primeira ordem.	25
Figura 4: Espectros Brutos na faixa de 400 a 800 nm.....	36
Figura 5: Espectro derivado na faixa espectral de 400 a 800 nm.....	36
Figura 6: Modelo PCA para o conjunto de amostras antes do processo derivativo.....	37
Figura 7: Modelo PCA para o conjunto de amostras após o processo derivativo.	38
Figura 8: Dendograma para todas as amostras de biodiesel	39
Figura 9: Resultado Si/S0 versus Hi para o modelo de biodiesel de algodão	40
Figura 10: Resultado Si/S0 versus Hi para o modelo de biodiesel de girassol.	41
Figura 11: Resultado Si/S0 versus Hi para o modelo de biodiesel de soja.	41
Figura 12: Perfil Espectral na faixa de 441 a 631 nm.....	45
Figura 13: Resultado da regressão via PLS.	46
Figura 14: Comprimentos de onda selecionados do modelo APS-MLR.	47
Figura 15: Comprimentos de onda selecionados do modelo SW-MLR.	47
Figura 16: Resultado da regressão via APS-MLR.	48
Figura 17: Resultado regressão SW-MLR.	48
Figura 18: Espectro das 122 amostras sem tratamento.....	53
Figura 19: Espectros das 122 amostras pós off set.....	53
Figura 20: Gráfico de scores de PC1X PC2 para o conjunto de amostra.....	54
Figura 21: Gráfico da função de custo versus o número de onda selecionado pelo modelo APS LDA.....	56
Figura 22: Número de onda ideal do modelo APS-LDA.....	56
Figura 23: Gráfico de scores da função discriminante (DF1x DF2) para conjunto de amostra.....	57

LISTA DE TABELAS

Tabela 1: Resultados da regressão pelos modelos APS-MLR, SW-MLR e PLS.....	49
Tabela 2: Resumo do resultado da classificação do modelo SIMCA.....	55

LISTA DE ABREVIATURAS

- AG - Algoritmo Genético, do inglês *Genetic Algorithm*
- ANOVA – Análise de Variância, do inglês *Analysis of Variance*
- ANP - Agência Nacional de Petróleo, Gás Natural e Biocombustíveis
- APS - Algoritmo das Projeções Sucessivas, do inglês *Successive Projection Algorithm*
- ASTM - Sociedade Americana de Testes e Materiais, do inglês *American Society of Tests and Materials*
- B2 – Mistura de biodiesel em diesel a 2 % (v/v)
- B4 – Mistura de biodiesel em diesel a 4 % (v/v)
- B5 – Mistura de biodiesel em diesel a 5 % (v/v)
- CEN - Comitê Europeu de Normalização
- D – Diesel Puro
- DCLS – Mínimos Quadrados Clássico Direto, do inglês *Direct Classical Least Square*
- GUI – Interface Gráfica do Usuário, do inglês *Graphical User Interface*
- HCA - Análise de Agrupamentos Hierárquicos, do inglês *Hierarchical Cluster Analysis*
- ICLS - Mínimos Quadrados Clássico Inverso, do inglês *Indirect Classical Least Square*
- KOH – Hidróxido de Potássio
- KS – Algoritmo Kennard-Stone
- LDA - Análise Discriminante Linear, do inglês *Linear Discriminant Analysis*
- MLR - Regressão Linear Múltipla, do inglês *Multiple Linear Regression*
- NaOH – Hidróxido de Sódio
- NIPALS - Algoritmo dos Mínimos Quadrados Parciais Iterativos Não Lineares, do inglês *Nonlinear Iterative Partial Least Squares*
- OB5 - Mistura Biodiesel/Óleo em diesel a 5 % (v/v)
- OLS - Mínimos Quadrados Ordinários, do inglês *Ordinary Least Squares*
- OPEP - Organização dos Países Exportadores de Petróleo
- PC – Componente Principal, do inglês *Principal Components*
- PCA - Análise por Componentes Principais, do inglês *Principal Component Analysis*
- PCR - Regressão por Componentes principais, do inglês *Principal Components Regression*
- pH – Potencial Hidrogeniônico
- PLS - Mínimos Quadrados Parciais, do inglês *Partial Least Squares*
- Pro-Óleo - Plano de Produção de Óleos Vegetais para Fins Energéticos
- PROÁCOOL – Programa Nacional do Alcool.

RMSEC - Raiz Quadrada do Erro Médio Quadrático de Calibração, do inglês *Root Mean Square Error of Calibration*

RMSECV - Raiz Quadrada do Erro Médio Quadrático de Validação Cruzada, do inglês *Root Mean Square Error of Cross-Validation*

RMSEP - Raiz Quadrada do Erro Médio Quadrático de Predição, do inglês *Root Mean Square Error of Prediction*

RMSEV - Raiz Quadrada do Erro Médio Quadrático de Validação, do inglês *Root Mean Square Error of Validation*

RP - Reconhecimento de Padrão

R-square Coeficiente de determinação, do inglês *coefficient of determination*

SIMCA - Modelagem Suave Independente de Analogias entre Classes, do inglês *Soft Independent Modeling of Class Analogy*

SVD - Decomposição por Valores Singulares, do inglês *Singular Value Decomposition*

SW – Stepwise

UV-Vis – Ultravioleta/Visível

SUMÁRIO

1 INTRODUÇÃO	16
1.1 CARACTERIZAÇÃO GERAL DO PROBLEMA	16
1.2 OBJETIVO GERAL	17
1.3 OBJETIVOS ESPECÍFICOS	17
2 REVISÃO DA LITERATURA	18
2.1 BIODIESEL	18
2.1.1 Biodiesel brasileiro.....	20
2.2 ESPECTROMETRIA DE ABSORÇÃO MOLECULAR NO UV-VIS	21
2.3 QUIMIOMETRIA.....	22
2.3.1. Calibração univariada e multivariada	23
2.3.1.1 Organização de dados multivariados.....	24
2.3.1.2 Métodos Clássicos de Calibração.....	25
2.3.1.3 Métodos inversos de calibração	25
2.3.2 Técnicas de Reconhecimento de Padrões	28
2.3.2.1 Técnicas de Reconhecimento de Padrões não Supervisionada	29
2.3.2.2 Técnicas de Reconhecimento de Padrões Supervisionada	30
2.3.3 Seleção de Variáveis	31
2.3.3.1 Algoritmo das Projeções Sucessivas	32
2.3.3.2 Algoritmo Stepwise.....	33
3. CLASSIFICAÇÃO DE BIODIESEL NO VISÍVEL	35
3.1.1 Aquisição das amostras	35
3.1.2 Aquisição dos espectros e softwares empregados	35
ANEXO 1	42
4. DETERMINAÇÃO DO TEOR DE BIODIESEL EM MISTURAS BIODIESEL/DIESEL UTILIZANDO ESPECTROSCOPIA UV-VIS	44
4.1.1 Aquisição das Amostras	44
4.1.2 Aquisição dos espectros e softwares empregados	44
ANEXO 2	50
5. ESPECTROMETRIA UV-VIS NA DETECÇÃO DE ÓLEO VEGETAL COMO ADULTERANTE EM BIODIESEL/DIESEL	52
5.1.1 Aquisição das amostras	52
5.1.2 Aquisição dos espectros e softwares empregados	52
5.2.1 Espectro UV-Vis	52

5.2.2 Modelo PCA	54
5.2.3 Classificação SIMCA	54
5.2.4 Classificação APS- LDA	55
ANEXO 3	57
6. CONCLUSÃO	59
REFERÊNCIAS	60

1 INTRODUÇÃO

1.1 CARACTERIZAÇÃO GERAL DO PROBLEMA

A utilização do biodiesel como combustível alternativo, seja na substituição do diesel ou no uso fracionado em misturas, está recebendo grande atenção entre pesquisadores e formuladores de políticas devido a inúmeras vantagens ambientais, econômicas e sociais. Além disto, as preocupações com a diminuição das reservas de petróleo, a instabilidade política nos países produtores e a crescente demanda mundial de energia levantaram a necessidade de buscar alternativas para combustíveis renováveis. Diversas outras matrizes energéticas renováveis vêm se destacando, tais como bioetanol, óleos vegetais, gordura animal, energia eólica, e energia solar (ATADASHI et. al, 2012; BALABIN, 2011; BALABIN et al., 2011; SENGO et al., 2010).

O biodiesel apresenta características semelhantes ao diesel, como alto ponto de fulgor, excelente lubricidade e elevado número de cetanos, menor teor de aromáticos e enxofre e promover a diminuição na emissão dos gases geradores do efeito estufa. Além disso, há um forte interesse em países desenvolvidos na utilização de fontes de energias modernas, eficientes e renováveis. Por fim, em países em desenvolvimento, destaca-se a disponibilidade de terras agricultáveis, clima favorável e disponibilidade de mão de obra, tornando assim a produção de biocombustíveis vantajosa frente a outros países (BALAT, 2011; FERRÃO, 2011).

O Brasil, na esteira desse raciocínio, tem investido na construção de indústrias capazes de extrair esse potencial para a produção de biodiesel, reforçada por políticas públicas na área, por meio da inserção do biodiesel na matriz energética brasileira, decorrendo de maneira gradual e progressiva, estimulada pela Medida Provisória nº. 214, de 13/9/2004, e a partir de 13 de janeiro de 2005 com a implantação do Programa Nacional de Produção e Uso de Biodiesel (Lei nº 11.097). No Brasil, desde 2008, a comercialização do biodiesel ocorre na forma de mistura, inicialmente com adição de 2 % de biodiesel ao diesel (B2) e em 2010, B5. Esta proporção de biodiesel representa uma produção nacional diária de aproximadamente 16.344,25 m³/dia, segundo dados da Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) (ANP, 2012), para o ano de 2012. Com a produção de biodiesel tão expressiva, faz-se necessário fiscalizar a produção e a comercialização desse biocombustível, pois o mesmo apresenta facilidade na adulteração com outras matérias primas de menor valor agregado, tais como os óleos vegetais (por exemplo, soja e algodão).

A responsabilidade pela fiscalização da qualidade do biodiesel brasileiro está a cargo da ANP (MEDIDA PROVISÓRIA, 2004). Entretanto, em sua maioria os parâmetros estabelecidos pela ANP são harmonizadas a normas internacionais da Sociedade Americana de Testes e

Materiais (ASTM, do inglês *American Society for Testing and Materials*) e do Comitê Europeu de Normalização (CEN, do francês *Comité Européen de Normalisation*). Essas técnicas utilizam grandes quantidades de reagentes, além de serem laboriosas e utilizarem uma instrumentação analítica de valor muito elevado (MONTEIRO, 2008).

Entretanto, durante o processo de mistura do biodiesel ao diesel nas distribuidoras pode ocorrer erro nas proporções a serem adicionadas, falhas também podem ocorrer na adição de produtos de menor valor agregado, ou seja, a adição de óleo vegetal ao diesel ao invés do biodiesel. Nesse sentido é importante dispor de metodologias que consigam quantificar e classificar as amostras de biodiesel e suas respectivas misturas. Assim a construção de modelo de calibração e de reconhecimento de padrão associado à espectrometria de absorção molecular nas regiões ultravioleta e visível apresenta-se como técnica rápida, não destrutiva e com alta precisão que pode ser utilizada no controle e na qualidade de biodiesel (VÉRAS et al, 2010; VÉRAS et al, 2011).

1.2 OBJETIVO GERAL

Classificar o biodiesel através do óleo de origem e misturado com óleo de soja e, por fim, construir modelos de regressão para quantificação do teor de biodiesel em misturas biodiesel/diesel, utilizando espectroscopia molecular na região UV-Vis.

1.3 OBJETIVOS ESPECÍFICOS

- Avaliar o conjunto espectral das amostras em relação às técnicas de reconhecimento de padrões para classificação do biodiesel.
- Avaliar o conjunto espectral das amostras em relação às técnicas de calibração multivariada;
- Avaliar o conjunto espectral das amostras de mistura biodiesel/ diesel contaminada com óleos de soja na análise *screening*.

2. REVISÃO DA LITERATURA

2.1 BIODIESEL

O aproveitamento de óleos vegetais e gorduras com a finalidade de ser utilizados como combustível data do final do século XIX, quando Rudolph Diesel desenvolveu o motor de combustão interna e utilizou como combustível de teste óleo de amendoim (GOLDEMBERG, 2004). Um outro marco histórico aconteceu em 1937, quando a patente do pesquisador Charles George Chavanne intitulada “Processo de transformação de óleos de vegetais para uso como combustível” (*Procédé de transformation d’huiles végétales en vue de leur utilisation comme carburants*) foi solicitada (SUAREZ,2007).

Depois da Segunda Guerra Mundial, entretanto, os estudos sobre biodiesel foram deixados de lado, pelo crescimento e normalização do mercado de petróleo. Todavia, os estudos sobre biodiesel foram retomados de forma mais substancial a partir da década de 70 quando a Organização dos Países Exportadores de Petróleo (OPEP) determinou limitar a quantidade de petróleo explorado (VÉRAS, 2010). Com essa atitude os valores do preço do barril tiveram um aumento considerável, disparando assim uma crise mundial e colocando em pauta discussões sobre a prospecção de energias renováveis para substituição dos derivados do petróleo (SUAREZ, 2007; VERAS 2010; HUANG E WU, 2008), processo que ocorreu mais uma vez no final da década de setenta e aprofundou a crise.

Tentando diminuir a dependência dos combustíveis provenientes dos derivados do petróleo, vários países começaram a utilizar inúmeras matérias-primas na produção de biodiesel, como os óleos vegetais de soja (MORADI et al, 2013; QIU et al, 2011), algodão (NABI et al, 2009; PAPADOPOULOS et al. 2010), girassol, (GHANEI et al, 2011), canola (THAMSIROJ et al, 2009), dendê (MEKHILEF et al, 2011) e pinhão manso (ONG et al, 2011). Há também a utilização de gorduras de origem animal e óleos residuais (GÜRÜ et. al. 2009; TASHTOUSH et al, 2004), dentre outros.

O biodiesel apresenta vantagens ambientais quando comparado ao óleo diesel por ser livre de enxofre e com baixos níveis de emissão de monóxido de carbono e material particulado formado durante a combustão. Além disso, o gás carbônico emitido nos motores do ciclo diesel pela combustão do biodiesel é compensado quase que totalmente durante o cultivo das oleaginosas, que o absorvem do meio ambiente (CORDEIRO et al, 2011).

Entretanto, para ser tornar um combustível alternativo rentável, o biodiesel deve ser produzido de forma ecologicamente segura, econômico, e deve fornecer um ganho líquido de energia sobre as fontes usada para produzi-lo (THIRU, 2011).

Diante dos benefícios que o biodiesel pode trazer na redução das emissões dos gases que causam o efeito estufa é necessário que o combustível conquiste o mercado, seguindo um rigoroso controle de qualidade. Em virtude disso encontram-se na literatura diversos trabalhos com interesse no controle de qualidade e na produção de biodiesel e suas misturas com diesel (CHAROENCHAITRAKOOL, THIENMETHANGKON, 2011), no transporte e armazenamento e na busca de novas fontes biológicas (KARMAKAR, KARMAKAR, MUKHERJEE, 2010), e no planejamento e obtenção de novas rotas de síntese (DABDOUB, BRONZEL, RAMPIN, 2009), dentre outras (QI et al, 2010; SALOUA et al, 2010; FERRARI et al, 2005; LIMA et al, 2007; FERRARI, 2009).

Para a obtenção do biodiesel, várias metodologias estão dispostas no que concerne ao processo de produção, dentre as quais esterificação (HAYYAN et al, 2011) e craqueamento (LAKSMONO et al, 2013). Contudo o processo mais comumente adotado para a produção de biodiesel é a transesterificação (LIAN et al, 2012), que consiste em uma reação química na qual triglicerídeo reage com um álcool geralmente de cadeia curta, utilizando catalisador alcalino para produzir ésteres alquílicos (biodiesel) e glicerol (HALIM, 2009), **Figura 1**.

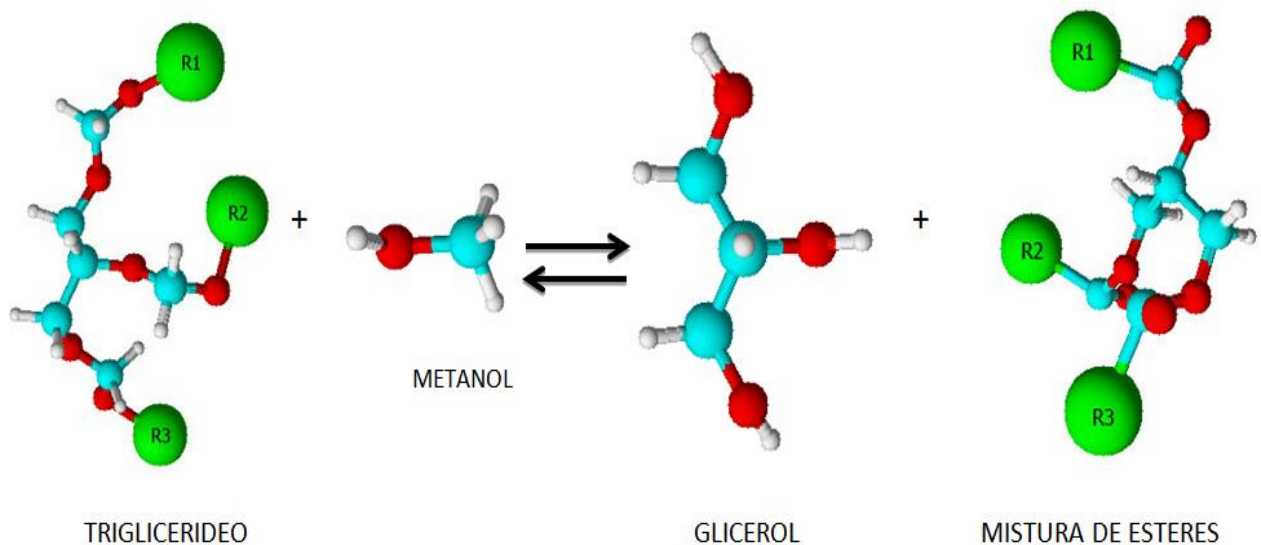


Figura 1: Representação da reação de transesterificação.

A reação apresenta um caráter reversível e assim, para evitar o deslocamento do equilíbrio químico para os reagentes, utiliza-se um excesso de álcool, aumentando o rendimento da reação (SINHA, 2008).

2.1.1 Biodiesel brasileiro

O uso de biocombustíveis no Brasil começou a ser discutido no período da crise mundial de petróleo (décadas de 70/80), em que o governo federal criou planos com intuito de utilizar fontes renováveis na matriz energética brasileira, tais como PROÁCOOL (GOLDEMBERG, 2004) e o PRO-ÓLEO (GARCIA, 2010). O Plano de Produção de Óleos Vegetais para Fins Energéticos (PRO-ÓLEO) tinha como propósito gerar excedentes de óleo vegetais para utilização em misturas com óleo diesel na proporção de 30 % e ao longo prazo a substituição total do óleo diesel.

Na década de 1980 o professor Expedito Parente, da Universidade Federal do Ceará, desenvolveu estudos intencionados a encontrar fontes alternativas de energia. Os estudos resultaram na criação de um novo combustível originário de óleos vegetais e com propriedades semelhantes ao óleo diesel convencional. Com esse trabalho o professor Expedito Parente tornou-se detentor da patente PI – 8007957 considerada a primeira patente, em termos mundiais, de biodiesel e de querosene vegetal para aviação (LIMA, 2004).

Porém, apenas com Medida Provisória nº. 214, de 13/9/2004, e com a implantação do Programa Nacional de Produção e Uso de Biodiesel (Lei nº 11.097), ocorrido em 13 de janeiro de 2005, que o biodiesel foi inserido na matriz energética brasileira. Comercialmente foi utilizado em forma de mistura a partir de 2008, onde a proporção usada era 2 % (v/v) biodiesel no diesel comum (B2). Nos anos seguintes, houve acréscimo na proporção de biodiesel misturado ao diesel e em 2010 a porcentagem foi a 5 % (B5), permanecendo até os dias atuais. A proporção necessária para corresponder ao consumo nacional diário de biodiesel é de aproximadamente 16.344,25 m³/dia, segundo dados da ANP (ANP, 2012, BRASIL, 2005) para o ano de 2012.

A produção de biodiesel no Brasil ocorre de forma progressiva, contando que o país possui disponibilidade de terras agricultáveis com boa capacidade produtiva, sendo explorada para produção de alimentos e bioenergia, trazendo benefícios econômicos e ambientais (PINTO et al., 2005). Um aspecto importante na produção de biodiesel é o apelo social, partindo do pressuposto de que a produção de oleaginosas apresenta aspectos familiares, o biodiesel torna-se uma alternativa importante para a possível diminuição da miséria no país, pela possibilidade da expansão do potencial produtivo da área agricultável, favorecendo a inclusão social, o desenvolvimento regional, a geração de emprego e renda, dentre outros fatores, principalmente na região do semiárido brasileiro (GARCIA et al, 2010; GERIS et al., 2007)

Mesmo com a descoberta de grande reserva de petróleo a cerca de 6 km de profundidade na camada do pré-sal (VÉRAS et al, 2012) o biodiesel continua sendo um produto importante. Além de renovável e usado com mecanismo de transferência de renda para agricultura familiar (GARCIA et al, 2010), o fator ambiental na utilização de biodiesel é muito bem visto, pois os resíduos provenientes de sua queima possuem níveis reduzidos de partículas, monóxido de carbono, óxidos de enxofre, hidrocarbonetos e fuligem (DU, 2004; REYES,2006).

Entretanto, os benefícios da utilização do biodiesel ao meio ambiente decorrem da fiscalização de toda cadeia produtiva, que no Brasil é gerenciada pela ANP, que utiliza parametrização que está fixada em algumas normas internacionais como da ASTM e do CEN. Porém tais técnicas utilizam grandes quantidades de reagentes além de serem laboriosas e utilizarem uma instrumentação analítica de valor muito elevado (SANTANA,2012; KUMAR, 2012; CALAND, 2012; MONTEIRO, 2008).

No que se refere às misturas biodiesel/diesel a ANP recomenda a utilização da EN 14078 para quantificar o teor de biodiesel em diesel, todavia a mesma emprega um único comprimento de onda na região do infravermelho médio (5730 nm), que corresponde ao máximo da banda de estiramento do grupo carbonila (ES, 2009). Como os óleos vegetais apresentam esta mesma banda, o método de referência é incapaz de detectar a presença dos mesmos. Além disso, os óleos vegetais possuem boa miscibilidade no óleo diesel, o que permite que possam ser utilizados como adulterantes nas misturas biodiesel/diesel.

Na literatura encontram-se trabalhos que utilizam metodologias alternativas capazes de quantificar e classificar misturas biodiesel/diesel, que são suficientemente rápidas, precisas e não invasivas. Dentre elas se destaca a espectrometria de absorção molecular nas regiões ultravioleta visível, por também ser viável economicamente e com largo espectro de aplicação (VÉRAS et al, 2011; VÉRAS et al, 2012).

2.2 ESPECTROMETRIA DE ABSORÇÃO MOLECULAR NO UV-VIS

Empregada a mais de 50 anos, a espectrometria molecular na região ultravioleta-visível (UV-Vis) vem sendo utilizada na quantificação de inúmeras espécies moleculares orgânicas e inorgânicas e também em amostras bioquímicas em diferentes tipos de materiais (FREITAS, 2006).

A técnica fundamenta-se nas medidas de absorção molecular em substâncias que são estimuladas a sofrer transições eletrônicas devido à absorção de energia quantizada na região UV-Vis (SKOOG, 2002). Para tal usa-se um feixe de luz branca incidente sobre uma amostra

(espécie molecular) que absorve luz e a radiação resultante emergente será detectável pela cor complementar da radiação absorvida (SKOOG, 2002), conforme ilustrado na **Figura 2**.

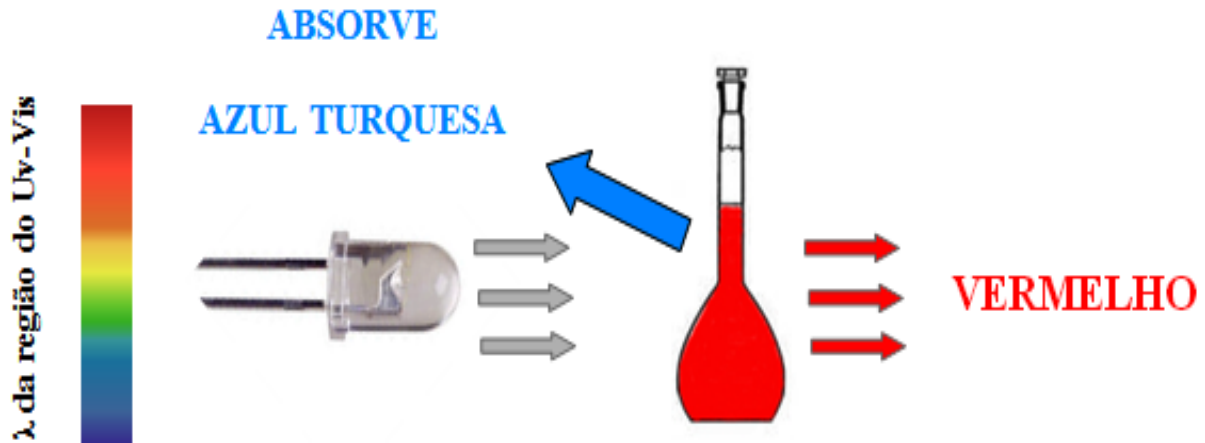


Figura 2: Radiação absorvente e a cor complementar.

A espectrometria de absorção molecular na região do UV-Vis destaca-se por apresentar, comparada a outras técnicas espectrométricas, uma instrumentação de baixo custo. Além disso, outras características são apresentadas abaixo confirmando a grande aplicabilidade dessa técnica (NETO, 2008).

Facilidade de operação – as medidas são realizadas de forma fácil e rápida, não destrutiva e não invasiva.

Ampla aplicabilidade em química, física, biologia, ciências agrárias, farmácia, geologia dentre outros, seja na identificação ou quantificação compostos químicos.

Grau de exatidão e aceitação – os erros relativos estão na faixa de 1 a 5 % em termos de concentração e

Possui alta relação sinal-ruído – os limites de detecção estão na faixa de 10^{-4} a 10^{-5} mol L⁻¹, podendo chegar a 10^{-7} mol L⁻¹.

2.3 QUIMIOMETRIA

Antigamente os químicos concentravam suas decisões numa pequena quantidade de dados que, na maioria das vezes eram obtidos de forma lenta e dispendiosa. Contudo o desenvolvimento da quimiometria deu-se com a evolução dos instrumentos computadorizados, que ocorreu por volta dos anos 60, onde a quantidade de dados aumentou de forma significativa

(FERREIRA et al, 1999), a obtenção tornou rápida e com menor esforço por parte dos analistas (SIMÕES, 2008).

A quimiometria é definida pela Sociedade Internacional de Quimiometria (ICS, do inglês *International Chemometrics Society*) como a disciplina da Química que utiliza métodos matemáticos e estatísticos para planejar ou selecionar experimentos de forma otimizada e extrair o máximo de informação química com a análise dos dados obtidos.

Considerada como um ramo aplicado à química analítica, a quimiometria tem um papel semelhante à biometria, sociometria, econometria, entre outros. A diferença da quimiometria para as demais ciências metrológicas está nas características dos dados de origem química, que possuem menor quantidade de variáveis influenciáveis, os quais em sua maioria são controláveis, de fácil repetição nas condições de trabalho desejadas (RAMOS et. al, 2001).

A quimiometria divide-se em algumas principais áreas tais como: processamento de sinais analíticos, planejamento e otimização de experimentos, calibração multivariada, modelagem de processos multivariados, métodos de inteligência artificial e reconhecimento e classificação de padrões (SIMÕES, 2008).

2.3.1. Calibração univariada e multivariada

Normalmente, os métodos analíticos procuram determinar a concentração de uma dada espécie química (analito) em uma amostra. Todavia essa grandeza não é mensurável diretamente e assim utiliza-se medidas indiretas para estimar tal valor (absorção de luz por moléculas, por exemplo) das amostras. No contexto da análise quantitativa instrumental, a propriedade mensurada guarda, geralmente, uma relação linear com a concentração do analito (SKOOG, 1992)

A calibração é definida como o processo que permite estabelecer a relação entre a resposta instrumental (sinal analítico) e uma determinada propriedade (física e/ou química) da amostra (concentração do analito) (BEEBE, 1998).

Quando na calibração uma única resposta (absorbância em um único comprimento de onda) é relacionada a uma única propriedade das amostras (concentração de um analito), denomina-se esse processo de calibração univariada ou de ordem zero (BARROS, 2002). Nestes casos, frequentemente utiliza-se o método dos Mínimos Quadrados Ordinários (BARROS, 2002) (OLS, do inglês *Ordinary Least Squares*) pra estimar os coeficientes de regressão do modelo de calibração.

A **Equação 1** estabelece a relação linear entre a variável aleatória (y , sinal analítico) e a variável assumida como não aleatória (x , concentração das amostras de calibração).

$$y_i = b_0 + b_1 x_i \quad (1)$$

A avaliação do modelo construído é realizada através da análise de variância (ANOVA, do inglês *Analysis of Variance*), no qual se verifica a significância da correlação linear entre x e y , e uma possível existência de falta de ajuste no modelo (INMETRO, 2003).

Comparadas aos modelos de calibração multivariada, a principal vantagem da calibração univariada é a simplicidade da matemática envolvida. Contudo, é necessário cautela na utilização do sinal medido, pois pode ocorrer que um único ponto não seja suficiente para descrever quantitativamente o modelo (GOMES, 2012).

Para superar o inconveniente indicado acima, utiliza-se calibração multivariada, em que ocorre calibração de múltiplas respostas (medidas de absorvância em vários comprimentos de onda) relacionadas a uma ou mais propriedades desconhecidas (concentração de um ou mais analitos) das amostras (PIMENTEL, 2008).

2.3.1.1 Organização de dados multivariados

Os modelos de calibração multivariada são organizados comumente em forma de matriz, **Figura 3**, em que a linha corresponde a uma amostra (objetos) e cada coluna contém a informação referente ao sinal analítico (variáveis).

Normalmente adota-se que a matriz que possui as respostas instrumentais (variáveis independentes) é denominada de matriz \mathbf{X} , e a matriz contendo parâmetros de referência (variáveis dependentes) é denotado por \mathbf{y} .

A calibração multivariada supera as limitações da calibração univariada, pois a mesma permite a determinação simultânea de analitos com maior sensibilidade e confiabilidade, reduzindo o tempo de análise. Outra vantagem é que pode ser realizada na presença de interferentes. Contudo, os interferentes devem também estar presentes na etapa de calibração, o que é conhecido como vantagem para a calibração de primeira ordem (FERREIRA, 1999; BRERETON, 2000).

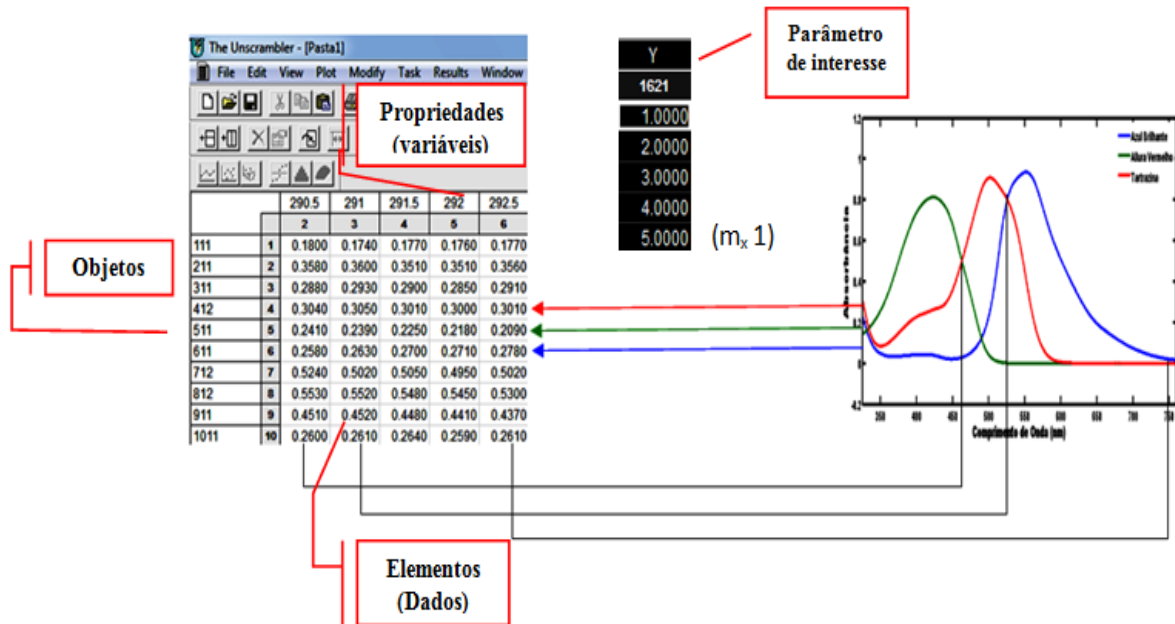


Figura 3: Organização matricial dos dados multivariados de primeira ordem.

2.3.1.2 Métodos Clássicos de Calibração

No método clássico de calibração existe uma relação de proporcionalidade entre o sinal analítico e a concentração (BEEBE,1998). Matematicamente, um modelo clássico pode ser expresso pela **Equação 2**:

$$\mathbf{X}=\mathbf{K}*\mathbf{Y} \quad (2)$$

Em que \mathbf{X} é a matriz de respostas instrumentais, \mathbf{Y} é a matriz das concentrações e \mathbf{K} é a matriz que contém o sinal puro de cada componente da mistura. Quando se obtém o valor de \mathbf{K} diretamente das medidas experimentais o método é denominado clássico direto (DCLS, do inglês *Direct Classical Least Square*). E quando o valor \mathbf{K} é estimado empregando \mathbf{X} e \mathbf{Y} o método é dito clássico indireto (ICLS, do inglês *Indirect Classical Least Square*) (BEEBE, 1998).

2.3.1.3 Métodos inversos de calibração

Nesses métodos é considerada a variável dependente como função do sinal analítico medido. Tais métodos superam alguns inconvenientes dos métodos clássicos por empregar a estrutura de variância/covariância da matriz \mathbf{X} na modelagem dos dados. Assim, é possível prever a variável dependente, como concentração, de um componente até mesmo se fontes adicionais de variações químicas e físicas estiverem presentes.

a) Regressão por Mínimos Quadrados Parciais

Wold et al em 1975 desenvolveram a Regressão por Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*). A modelagem PLS utiliza a informação contida na matriz de dados \mathbf{X} e a matriz resposta \mathbf{Y} (concentração, por exemplo) para obtenção das novas variáveis, que são denominadas de variáveis latentes, componentes ou fatores. As matrizes \mathbf{X} e \mathbf{Y} são decompostas em “ \mathbf{K} ” variáveis latentes como mostrado nas **Equações 3 e 4** (WOLD, 2001; BORGES, 2005).

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}_X = \sum_{k=1}^K \mathbf{t}_k \mathbf{p}_k' + \mathbf{E}_X \quad (3)$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{E}_Y = \sum_{k=1}^K \mathbf{u}_k \mathbf{q}_k' + \mathbf{E}_Y \quad (4)$$

Onde na **Equação 3**, \mathbf{P}' corresponde aos *loadings*, \mathbf{T} a matriz dos escores e \mathbf{E}_X é a matriz dos resíduos da matriz de dados \mathbf{X} .

Na **Equação 4**, \mathbf{U} corresponde a matriz dos escores, \mathbf{Q}' aos *loadings* e \mathbf{E}_Y corresponde aos resíduos da matriz resposta de \mathbf{Y} .

Entre os escores de \mathbf{X} e os escores de \mathbf{Y} , uma relação linear é estabelecida, **Equação 5**:

$$\mathbf{u}_k = \mathbf{b}_k \mathbf{t}_k \quad (5)$$

Onde \mathbf{b}_k é vetor dos coeficientes de regressão para cada um dos fatores. Para determinar a matriz dos coeficientes de regressão \mathbf{b} utiliza-se a **Equação 6**:

$$\mathbf{Y} = \mathbf{TBQ}' + \mathbf{E}_Y \quad (6)$$

Diversas formas de obter os parâmetros de um modelo PLS estão disponíveis na literatura (ANDERSSON, 2009), em que o mais conhecido e utilizado é o Algoritmo dos Mínimos Quadrados Parciais Iterativos Não-Lineares (NIPALS, do inglês *Nonlinear Iterative Partial Least Squares*), proposto por Wold (BRERETON, 2000). Todavia, a utilização de qualquer algoritmo para se obter a regressão PLS deveria levar ao mesmo resultado. Entretanto, estudos mostram que do ponto de vista numérico existem diferenças no resultado final (ANDERSSON, 2009), e a intensidade dessa diferença está associada à natureza dos dados, ao número de fatores PLS empregado e a precisão usada nos cálculos.

O método PLS pode ser utilizado para determinação de um ou de vários analitos na matriz de dados \mathbf{Y} , denominados de PLS1 e PLS2, respectivamente.

Um fato de que deve ser levado em consideração é que os modelos PLS não permitem uma interpretação físico-química direta dos resultados. A razão dessa dificuldade advém do fato de que nesta técnica se realiza a regressão no domínio dos dados transformados.

b) Regressão por Componentes Principais

A Regressão por Componentes Principais (PCR, do inglês *Principal Components Regression*) utiliza os fundamentos do cálculo da PCA (HIBBERT, 2009). O interessante nesse método de regressão é que não é necessária a utilização de seleção de variáveis prévia para contornar o problema de multicolinearidade dos dados, visto que a mesma faz uso de uma transformação ortogonal da matriz \mathbf{X} , de modo a obter um novo conjunto de variáveis linearmente independentes (VALDERRAMA, 2009). A matriz \mathbf{X} de alta dimensão é decomposta em duas matrizes menores, que recebem o nome de escores (\mathbf{T}) e pesos (*loadings*) (\mathbf{P}). No cálculo, a matriz \mathbf{E} representa o resíduo de \mathbf{X} (BRERETON, 2003), como mostrado na **equação 7**.

$$\mathbf{X} = \mathbf{T} * \mathbf{P}' + \mathbf{E} \quad (7)$$

Tendo calculado \mathbf{T} , é possível obter uma equação de regressão entre a propriedade a ser determinada \mathbf{y} e a matriz de escores \mathbf{T} . A relação existente entre essas duas propriedades pode ser descrita pela **equação 8**.

$$\mathbf{y} = \mathbf{T}_{(m \times k)} \mathbf{b}_{PCR} + \mathbf{F} \quad (8)$$

Onde k corresponde ao número de componentes principais empregados na obtenção dos coeficientes de regressão e \mathbf{F} aos resíduos não modelados.

Atualmente, alguns algoritmos vêm sendo utilizados para determinar o valor da matriz \mathbf{T} via PCA, como por exemplo o algoritmo NIPALS e a decomposição por valores singulares (SVD, do inglês *Singular Value Decomposition*) (SOARES, 2010).

O cálculo é realizado a partir da combinação linear das colunas de \mathbf{X} com a componente principal \mathbf{t}_1 , que por sua vez apresenta a maior variância explicada possível e \mathbf{p}_1 é escalonado para que a norma seja igual a 1. As outras componentes principais são calculadas usando o mesmo critério, porém \mathbf{t}_1 , \mathbf{t}_2 até \mathbf{t}_n são perfeitamente não correlacionadas entre si, assim com os

vetores \mathbf{p}_1 , \mathbf{p}_2 até \mathbf{p}_n . O cálculo é finalizado quando o número de componentes principais calculado é igual ao definido pelo usuário.

c) Regressão Linear Múltipla

Introduzida por Sternberg et al (1960), a Regressão Linear Múltipla (MLR, do inglês *Multiple Linear Regression*) é uma técnica que estabelece uma relação linear entre as matrizes da respostas instrumentais localizada em $\mathbf{X}_{(m \times n)}$ e a propriedade analisada alocada na matriz $\mathbf{Y}_{(m \times 1)}$ aplicando o métodos dos mínimos quadrados, como mostrado na **Equação 9**, em que o resíduo não modelado em \mathbf{y} é representado por \mathbf{E} (NUNES, 2008).

$$\mathbf{y} = \mathbf{X}\mathbf{b}_{\text{MLR}} + \mathbf{E} \quad (9)$$

O vetor \mathbf{b} é calculado na etapa de calibração utilizando os métodos dos mínimos quadrados ordinários, como na **equação 10**.

$$\mathbf{b}_{\text{MLR}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \quad (10)$$

A resolução da **Equação 4**, para obter o vetor dos coeficientes de regressão (\mathbf{b}), requer a inversão da matriz $(\mathbf{X}'\mathbf{X})$ e esta operação algébrica possui algumas limitações acerca dos dados das matrizes. Uma limitação do método é que o número de amostras deve ser igual ou superior ao número de variáveis. Caso contrário, o sistema torna-se indeterminado. Outra limitação é o fato que as variáveis (colunas de \mathbf{X}) devem ser idealmente vetores linearmente independentes. Caso contrário pode-se levar a uma matriz singular. Tais limitações são contornadas utilizando algoritmos de seleção de variáveis.

2.3.2 Técnicas de Reconhecimento de Padrões

Os seres humanos conseguem distinguir com facilidade as diferenças e semelhanças entre os objetos, seja ela através de formas ou cores. Essa característica fica limitada quanto maior for a quantidade de objetos a serem analisados. Semelhantemente quando tratamos de dados químicos, com um grande conjunto de dados, as técnicas de Reconhecimento de Padrão (RP) utilizam o mesmo conceito, onde se procura encontrar as similaridades e dissimilaridades no conjunto de amostras que foram submetidas a algum estudo. (GONZÁLEZ, 2007; HOPKE, 2003)

As técnicas de Reconhecimento de Padrão RP são divididas em supervisionada e não supervisionada, de acordo com a utilização *a priori* de informações sobre as amostras que constituem o conjunto para construção do modelo (BERRUETA, 2007; CORREIA, 2007; CANDOLFI, 1999).

2.3.2.1 Técnicas de Reconhecimento de Padrões não Supervisionada

Nas técnicas de RP não supervisionadas é avaliada a presença de agrupamentos sem a necessidade do conhecimento prévio do conjunto de amostras ou das classes. Essas técnicas utilizam apenas medidas de algumas propriedades com intuito de se observar agrupamentos naturais.

Na literatura são encontrados alguns algoritmos que estão relacionados com as técnicas de reconhecimento de padrão não supervisionado, destacando a análise de agrupamentos hierárquicos (HCA, do inglês *Hierarchical Cluster Analysis*) e análise de componentes principais (PCA, do inglês *Principal Component Analysis*) (BERRUETA, 2007; GONZÁLEZ, 2007).

a) Análise por Componentes Principais

A PCA é uma das técnicas de RP não supervisionadas mais utilizada, pois apresenta um resultado de fácil observação de agrupamentos das amostras que possuem propriedades semelhantes. Para tal, há a manipulação matemática da matriz de dados de forma que uma grande quantidade de variáveis seja reduzida, mantendo a maior quantidade de informação possível a fim de representar as variações presentes em um número menor de fatores (CORREIA et al, 2007). O tratamento matemático baseia-se na combinação linear das variáveis originais, que constituem um novo sistema de eixos ortogonais entre si, que se denominam de componentes principais (PC, do inglês *Principal Components*) (FERREIRA et al, 1999). O cosseno do ângulo entre o eixo da variável e o eixo da PC é denominado de pesos (*loadings*), e as coordenadas das amostras no novo sistema de eixos das PCs são chamadas de escores (NASCIMENTO, 2008).

O número máximo de PCs que é possível obter é igual ao número de variáveis. Porém, um número pequeno de componentes é responsável por grande parte da variabilidade total dos dados. Assim a PCA agrupa variáveis que estão altamente correlacionadas em novas variáveis, criando um conjunto que contém apenas as informações importantes e descartando as redundantes (PONTES, 2009).

b) Análise Hierárquica de Agrupamentos

A HCA tem a finalidade de agrupar as amostras que possuam similaridade entre si e separar as que possuem dissimilaridades. Para encontrar os agrupamentos é utilizado o cálculo de distâncias interpontos correspondentes as m amostras ou n variáveis no espaço n -dimensional de uma matriz de dados (BEEBE, 1998).

O dendograma é o gráfico utilizado na PCA para observação de similaridades e dissimilaridades entre os agrupamentos. Para calcular a similaridade entre duas amostras x e y utiliza-se a **equação 11**, onde d_{xy} é a distância x e y e d_{\max} é a distância máxima entre todas as amostras consideradas. As amostras são ditas similares se elas apresentarem valores de S_{xy} próximos da unidade (MOREIRA, 2001).

$$S_{xy} = 1 - \frac{d_{xy}}{d_{\max}} \quad (11)$$

Outro parâmetro que deve ser levado em consideração é a escolha do critério com que os subagrupamentos serão ligados e para isso diversas técnicas de conexão são utilizadas, como por exemplo *single linkage* (conexão simples) e *complete linkage* (conexão completa) são alguns exemplos.

2.3.2.2 Técnicas de Reconhecimento de Padrões Supervisionada

As técnicas de Reconhecimento de Padrões Supervisionadas utilizam o conhecimento prévio do conjunto de amostras ou das classes que o compõem para identificar as amostras desconhecidas (GONZÁLEZ, 2007; BERRUETA, 2007, BEEBE, 1998)

São encontrados diversos métodos de Reconhecimento de Padrões Supervisionado destacando a Modelagem Independente e Flexível Por Analogia de Classes (SIMCA, do inglês *Soft Independent Modeling of Class Analogy*) e a Análise Discriminante Linear (LDA, do inglês *Linear Discriminant Analysis*) (GONZÁLEZ, 2007; BERRUETA, 2007,)

a) Modelagem Suave Independente Por Analogias de Classes

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 principal 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 o permitam ser inserido 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 , **equação 12**, pela a variância da classe (MOREIRA, 2007), for menor que o F crítico.

$$z^2 = x^2 + y^2 \quad (12)$$

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).

b) Análise Discriminante Linear

Essa técnica estima a combinação linear entre duas ou mais funções discriminantes. Para tal, a discriminação é feita determinando os pesos das variáveis independentes do melhor conjunto de variáveis, de forma que haja a minimização da variância entre as amostras pertencente ao mesmo grupo e ocorra a maximização entre as amostras pertencentes a grupos distintos. (MOREIRA, 2007)

A análise discriminante linear assemelha-se com a PCA, pois ambas procuram reduzir a dimensionalidade do conjunto de dados, todavia a LDA conduz o cálculo com intuito de alcançar a máxima separação entre as classes avaliadas.

Comparado com alguns métodos classificatórios, a LDA possui sua funcionalidade restrita a conjuntos de dados de pequenas dimensões. Além disso, a capacidade de generalização de modelos LDA pode ser comprometida por problemas de colinearidade. Contudo, a utilização de algoritmos de seleção de variáveis tem sido utilizada com sucesso (PONTES, 2011; WANG, 2012).

2.3.3 Seleção de Variáveis

Em uma análise multivariada onde as dimensões do conjunto de dados são extensas ou apresentam poucas amostras ou muitas variáveis, a utilização e o desempenho de certos métodos de reconhecimento de padrões e de regressão pode ser prejudicada, pois muitas variáveis são irrelevantes ou redundantes. (PONTES, 2009)

As técnicas de seleção de variáveis baseiam-se no princípio de que um pequeno número de variáveis é capaz de gerar bom preditores e remover variáveis não informativas e minimizar a

multicolinearidade, que atrapalha as técnicas de reconhecimento de padrões e regressão (SOARES, 2010).

2.3.3.1 Algoritmo das Projeções Sucessivas

O Algoritmo das Projeções Sucessivas (SPA, do inglês *Successive Projections Algorithm*) trata de uma técnica que utiliza a seleção *forward*, partindo de uma variável inicial x_k e vai incorporando em cada iteração uma nova variável com a menor multicolinearidade possível em relação as já selecionadas (ARAÚJO et al, 2001). Originalmente foi idealizada para selecionar variáveis minimamente colineares em calibração multivariada baseada em MLR (ARAÚJO et al, 2001). Na literatura são encontrados diversos trabalhos que relatam a utilização de tal técnica (SOARES, 2013; LIU, 2009).

O SPA envolve essencialmente três etapas. Na primeira etapa, são selecionados subconjuntos de variáveis levando em consideração o critério de minimização da multicolinearidade, para tal, as operações para obter os conjuntos são aplicadas nas colunas da matriz de calibração. Na segunda fase, é escolhido o subconjunto que obtiver melhor resultado em relação ao critério que avalia a habilidade de previsão de um modelo MLR, a Raiz Quadrada do Erro Médio Quadrático de Validação (RMSEV: *Root Mean Square Error of Validation*) (ARAÚJO et al, 2001), de acordo com a **Equação 13**.

$$\mathbf{RMSEV} = \sqrt{\frac{1}{K_v} \sum_{k=1}^{K_v} (y_v^k - \hat{y}_v^k)^2} \quad (13)$$

onde: K_v é número de amostras do conjunto de validação e y_v^k e \hat{y}_v^k são os valores de referência e os valores previstos para o parâmetro de interesse nas amostras de validação.

Na terceira fase, o subconjunto escolhido é submetido a um procedimento de eliminação para determinar se alguma variável pode ser removida sem perda significativa da capacidade de predição.

O SPA possui uma versão aplicada a modelagem de métodos de reconhecimento de padrões supervisionados (GHASEMI-VARNAMKHAISTI, 2012; SOUTO, 2010), que ao invés de utilizar a função de custo RMSEV para guiar a classificação, utiliza o risco médio G da LDA, conforme a **Equação 14**.

$$\mathbf{G} = \frac{1}{K_v} \sum_{k=1}^{K_v} \mathbf{g}^k \quad (14)$$

onde g_k é o risco de uma classificação incorreta amostra de validação é definido na equação 15.

$$g_k = \frac{r^2(X_k, \mu_k)}{\min_{j \neq k} r^2(X_k, \mu_j)} \quad (15)$$

No numerador $r^2(X_k, \mu_k)$ é o quadrado da distância de Mahalanobis entre o objeto X_k com índice de classe e a média de sua classe. O denominador corresponde ao quadrado da distância de Mahalanobis entre o objeto X_k e o centro da classe errada mais próxima. O valor de g_k deverá assumir valores muito pequenos quanto a amostra estiver perto do centro da sua verdadeira classe e distante dos centros das demais classes.

2.3.3.2 Algoritmo Stepwise

O Stepwise é um método de seleção de variáveis baseado nos princípios de *Forward Selection* e *Backward Elimination* (ARAUJO, 2001).

Na rotina do método *Forward Selection*, o modelo parte de uma variável x , escolhida como tendo a maior correlação com y , e outras variáveis são adicionadas ao modelo posteriormente. Uma avaliação é realizada com base no teste F após cada variável ser adicionada, para tal o modelo utiliza o cálculo da soma quadrática residual do novo modelo para calcular um valor de F (F_{cal}). Assim a variável que fornecer o maior valor de F_{cal} permanece no modelo (SOARES, 2010).

O método *Backward Elimination* parte inicialmente com todas as variáveis disponíveis e vai retirando variáveis. A cada saída de variável, o algoritmo faz uma avaliação com base em um teste F e a variável que fornecer o menor valor de F_{cal} é removida do modelo. Em ambos os métodos (*Forward Selection* e *Backward Elimination*) o processo continua até que não haja variáveis com valores de F_{cal} maiores que um valor crítico ($F_{crítico}$) tabelado ou obtido empiricamente para um determinado nível de confiança e grau de liberdade. (GALVÃO et al, 2009)

Para evitar que cada variável seja recolocada no cálculo e assim reavaliada quando agrupada no modelo, o algoritmo inclui uma fase de inclusão seguida por uma fase de exclusão que é guiada através de testes F parciais para as x variáveis (SOARES, 2010).

No processo de inclusão, as variáveis restantes são submetidas uma a uma a um teste F parcial. Quando o valor do teste F for maior do que um valor crítico de entrada ($F_{entrada}$), a variável do valor correspondente é inserida ao modelo. De maneira análoga, no processo de

exclusão, cada uma das variáveis no modelo está sujeito a um teste F parcial, porém se o valor de F apresentar-se menor do que um valor crítico de saída ($F_{saída}$), o valor da variável correspondente é eliminado do modelo e as variáveis ainda estarão disponíveis para a seleção (GALVÃO, 2009).

3. CLASSIFICAÇÃO DE BIODIESEL NO VISÍVEL

3.1 METODOLOGIA

3.1.1 Aquisição das amostras

Para obtenção das amostras de biodiesel, foram adquiridas 24 amostras de óleos refinados sendo oito amostras de cada oleaginosa (soja, algodão e girassol), sendo cada amostra procedente de marcas e lotes diferentes. Os biodieseis dessas amostras de óleo foram obtidos em triplicata via transesterificação por rota etílica e hidróxido de potássio como catalisador, perfazendo um total de 72 amostras de biodiesel. O processo de purificação das amostras de biodiesel foi realizada utilizando água destilada e solução de HCl 0,1 mol/L, até alcançar pH próximo a 7,0. A lavagem do biodiesel tem a finalidade de retirar impurezas, tais como: o excesso de álcool, que dificulta a separação entre as fases biodiesel/glicerina e o excesso do catalisador hidróxido de potássio, o qual pode ocasionar a formação de sais de ácidos graxos (LÔBO, 2009).

3.1.2 Aquisição dos espectros e softwares empregados

As amostras foram submetidas a medidas espectrométricas de refletância difusa realizadas em triplicata em um espectrômetro Vis/NIR FOSS XDS MasterLab™. Selecionou-se a faixa de 400 a 780 nm, com uma resolução espectral de 0,5 nm para posterior pré-tratamento dos dados com seleção de faixa de variáveis mais significativas para classificação.

Os dados foram pré-processados com o programa computacional The Unscrambler 9.8 e para a construção dos modelos PCA e SIMCA. O programa computacional Statistica 9.0 foi utilizado para construir modelos HCA, para verificação dos agrupamentos dos tipos de biodiesel. O programa computacional Matlab 7.11.0.584 (R2010b) foi utilizado para seleção do conjunto de treinamento e teste baseado no algoritmo de Kennard-Stone, para construção do modelo SIMCA.

3.2 RESULTADOS E DISCUSSÃO

3.2.1 Dados Espectrais

A **Figura 4** apresenta os espectros do conjunto de amostras de biodiesel de soja, algodão e girassol na região de 400 a 800 nm. Os dados foram pré-processados utilizando primeira derivada pelo método de Savitzky-Golay, empregando um polinômio de segunda ordem e janela de 21 pontos, permitindo destacar os comprimentos de onda mais discriminativos na classificação das amostras, **Figura 5**. Os espectros estão destacados por cor para cada tipo de óleo utilizado e não é possível notar uma separação dos espectros que permita uma classificação sem um tratamento quimiométrico adequado.

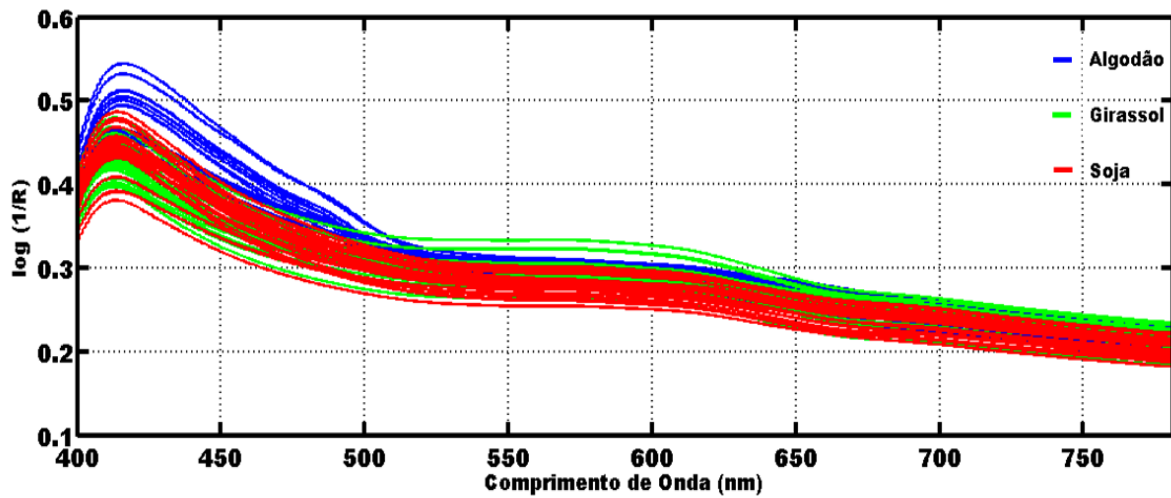


Figura 4: Espectros Brutos na faixa de 400 a 800 nm.

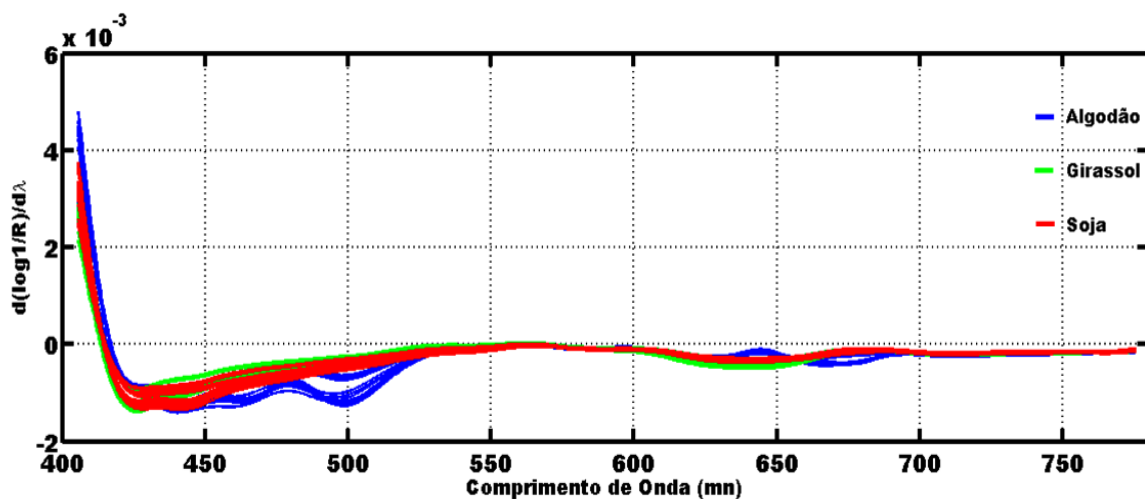


Figura 5: Espectro derivado na faixa espectral de 400 a 800 nm.

3.2.2 Modelo PCA

O modelo PCA foi construído com as 72 amostras de biodiesel, e para validar o modelo foi utilizada a validação cruzada completa, *leave-one-out*. A **Figura 6** apresenta o comportamento espacial do conjunto de amostras com os espectros brutos. É possível notar a sobreposição das amostras de biodiesel de girassol com biodiesel de soja, explicável pela semelhança na composição dos ácidos graxos entre as duas oleaginosas. As amostras do biodiesel de algodão estão afastadas das amostras de biodiesel de soja e de girassol, fato explicável, principalmente, pela diferença na quantidade dos ácidos mirístico, palmitoléico e palmítico (CODEX ALIMENTARIUS, 2011).

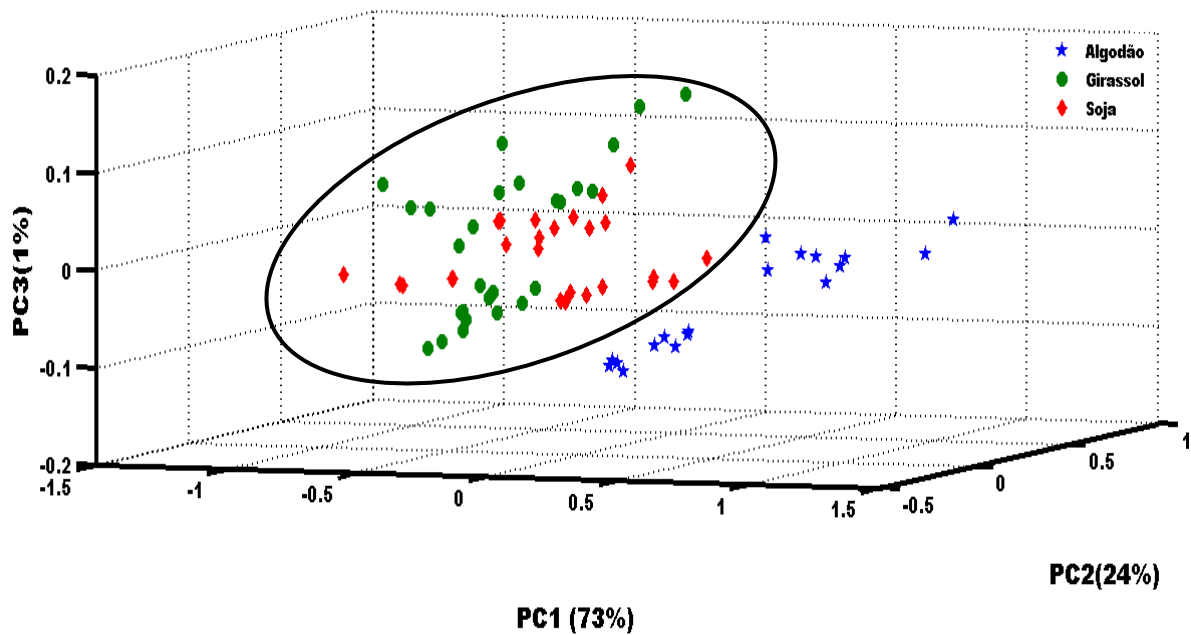


Figura 6: Modelo PCA para o conjunto de amostras antes do processo derivativo.

A **Figura 7** apresenta a distribuição de scores para as amostras de biodiesel estudadas após a utilização do filtro móvel de Savitzky-Golay, baseando-se em três PC's, em que é possível verificar a formação de três agrupamentos, associados com a o tipo de óleo utilizado na síntese das amostras de biodiesel. Assim, a supressão da sobreposição entre os três agrupamentos corrobora com a necessidade do pré-processamento aplicado às amostras e sugere que as informações, variáveis, que destacam os agrupamentos em questão estão em regiões do espectro encoberto por outras informações ou ruído. Entretanto, mesmo com o pré-processamento, a separação dos agrupamentos só é possível com técnicas quimiométricas, pois ainda permanece a sobreposição espectral e variações de sinal menos significativas provocam a separação dos agrupamentos.

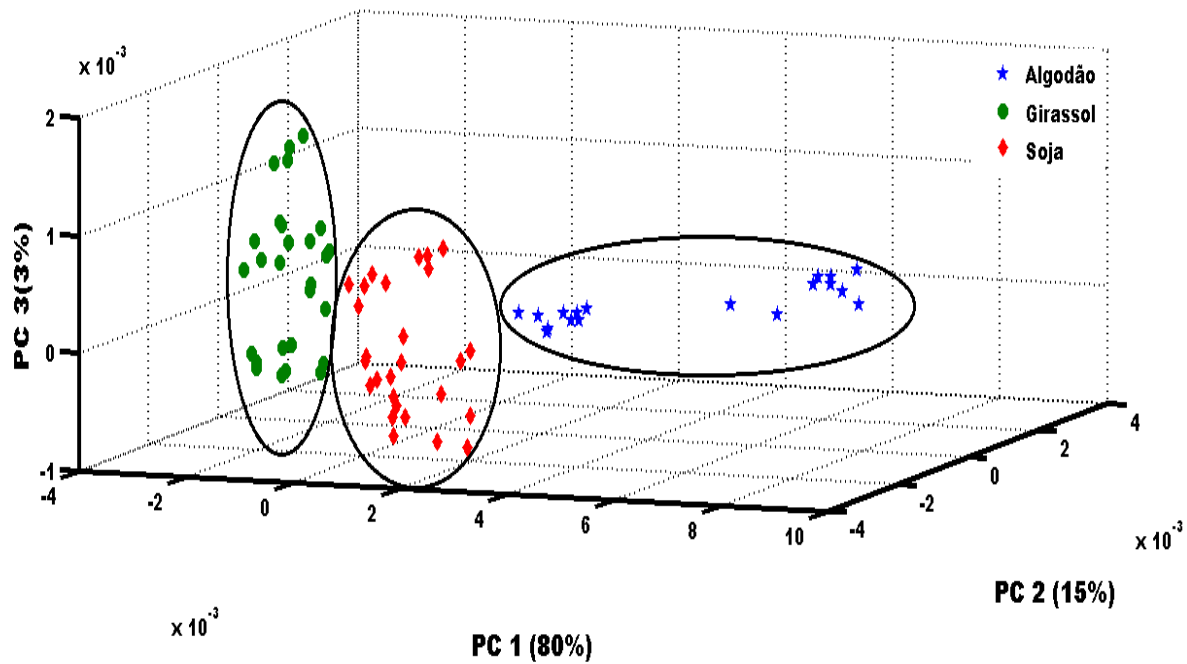


Figura 7: Modelo PCA para o conjunto de amostras após o processo derivativo.

3.2.3 Modelo HCA

A Análise HCA foi construída para todas as 72 amostras de biodiesel aplicando a técnica de ligação completa, *complete linkage*, que calcula a proximidade entre os agrupamentos das amostras por meio da regra dos vizinhos mais distantes e correlação de 1-Pearson R, que determina a extensão em que valores de duas variáveis são proporcionais entre si, para medida de dissimilaridades entre as amostras.

Agrupamentos HCA foram obtidos através do uso dessa técnica, **Figura 8**, baseando-se na distância de ligação máxima obtida. Todos os agrupamentos obtiveram faixas distância de ligações aproximadas, sendo estas: agrupamento do biodiesel de algodão (distância de ligação de 12,0 %), agrupamento do biodiesel de girassol (distância de ligação de 12,5 %), agrupamento do biodiesel de soja (distância de ligação de 11,2 %). Estas distâncias de ligação denotam a uniformidade das amostras pertencentes às mesmas oleaginosas. Por outro lado, nota-se que o grupo de amostras pertencentes ao biodiesel de soja é um pouco mais uniforme quando comparado às outras amostras de algodão e girassol.

Foi possível verificar que os agrupamentos das amostras de biodiesel se diferenciaram nas seguintes distâncias, com base na distância de ligação máxima obtida: agrupamento de soja para girassol (distância de ligação de 33,2 %), agrupamentos de soja e girassol para o agrupamento de algodão (distância de ligação de 100 %). Assim, o modelo HCA construído para as amostras de biodiesel corrobora com os resultados apresentados no modelo PCA, verificando

a presença de três agrupamentos também associados com o tipo de óleo utilizado na síntese das amostras de biodiesel. As características de destaque entre os agrupamentos de oleaginosas, da mesma forma do obtido no modelo PCA, estão associados aos ácidos graxos presentes que diferem de um óleo ao outro, conforme descrito no *codex alimentarius* (CODEX ALIMENTARIUS, 2011).

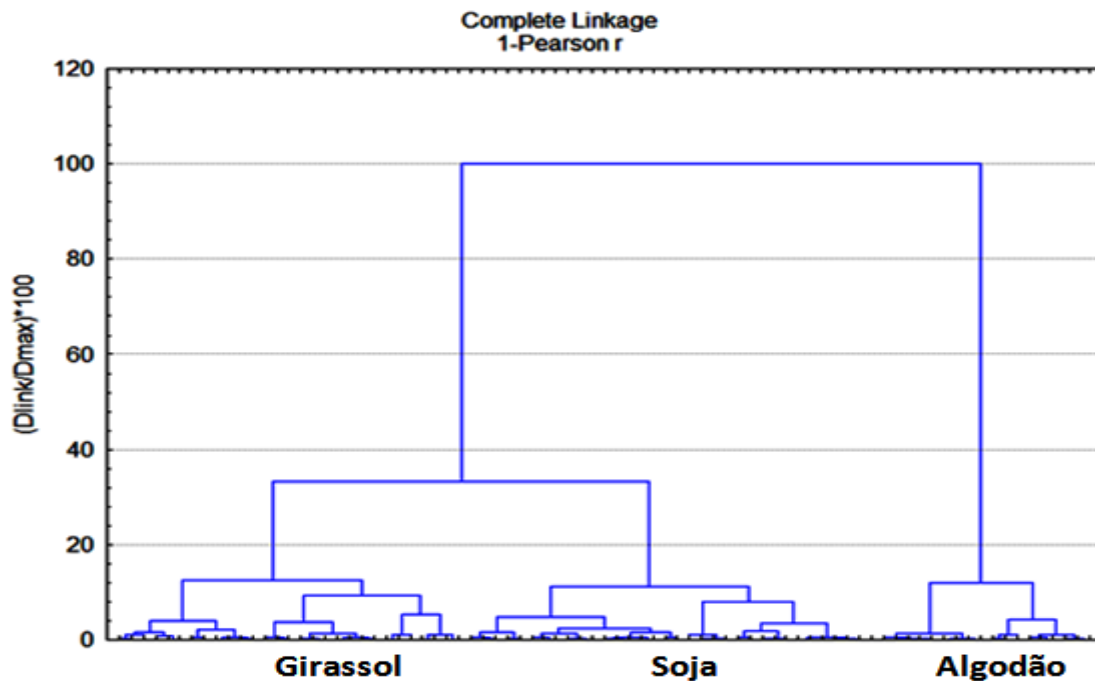


Figura 8: Dendrograma para todas as amostras de biodiesel.

3.2.4 Modelagem SIMCA

Observados os modelos PCA e HCA foi possível avaliar que um modelo utilizando técnicas de reconhecimento de padrões supervisionado seria uma alternativa viável para classificar amostras de biodiesel de óleos distintos.

O modelo de classificação SIMCA foi construído para a faixa espectral de 405 a 775 nm com os espectros pré-processados. O conjunto de amostras foi particionado, em conjuntos de treinamento e de teste, utilizando o algoritmo Kennard-Stone em ambiente de Matlab. O conjunto de treinamento foi construído com 44 amostras, em que 12 amostras de biodiesel de algodão 16 para girassol e 16 para soja. O conjunto de teste contém 28 amostras de biodiesel distribuídas em 6 amostras de algodão e 11 amostras de girassol e de soja. O modelo construído foi validado empregando validação cruzada completa. Os resultados obtidos na modelagem SIMCA apresentaram bom desempenho para todos os níveis de significância estatística avaliados (1%, 5%, 10% e 25%), classificando todas as amostras em suas respectivas classes.

Nos gráficos Si/S0 versus Hi dos modelos SIMCA para as amostras de biodiesel, **Figuras 9, 10 e 11**, é possível verificar as distâncias ortogonais das amostras dos conjuntos de teste, indicando os limites para classificação dentro de cada classe, a 95 % de confiança estatística.

Os gráficos indicam os dois limites utilizados para classificar uma amostra a um dado modelo. Ressalta-se a presença de uma amostra de girassol e uma de algodão, distantes de seus grupos que apresenta variabilidade das amostras nos modelos, não implicando, contudo, em erro de classificação dessas amostras.

É possível notar que uma maior similaridade entre as amostras de biodiesel de soja e de girassol nos modelos não-supervisionados PCA e HCA, também foram identificados nos modelos SIMCA construídos.

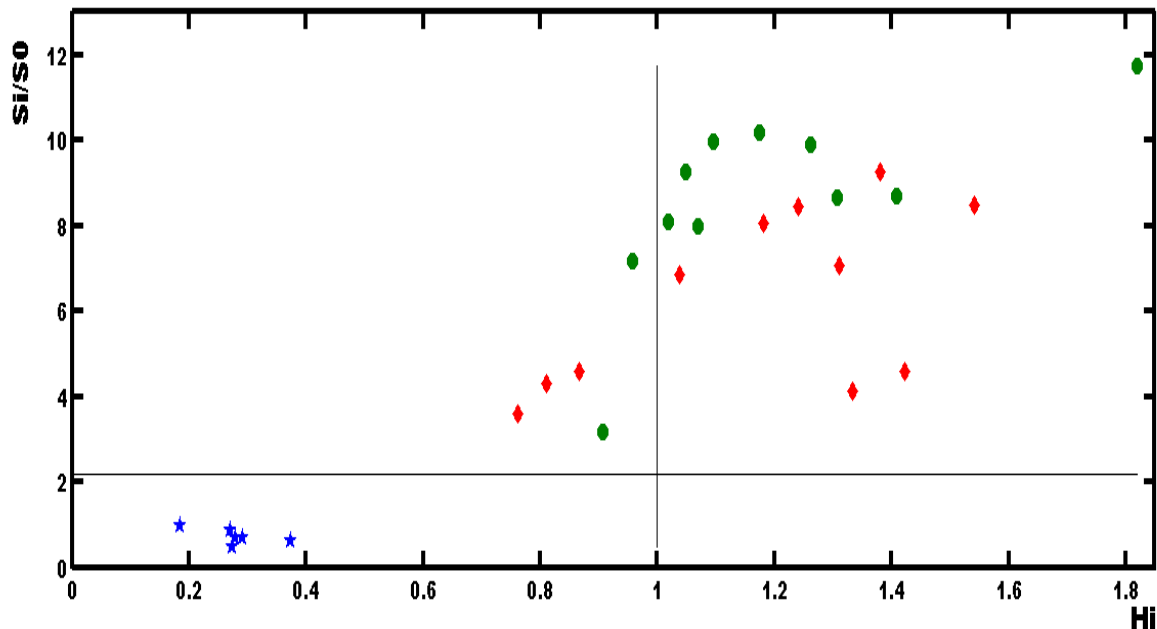


Figura 9: Resultado Si/S0 versus Hi para o modelo de biodiesel de algodão.

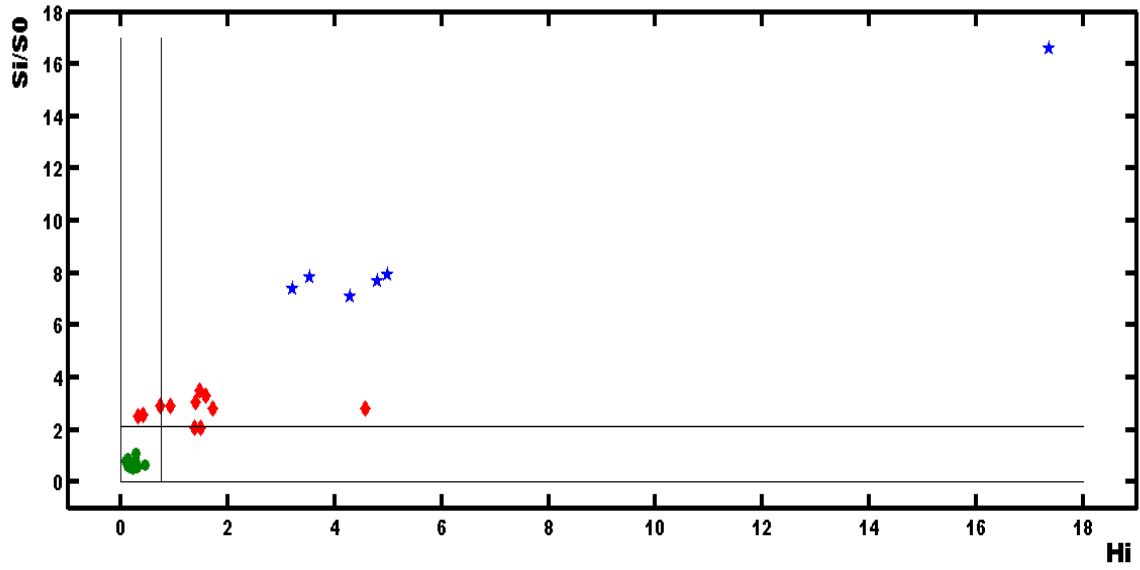


Figura 10: Resultado $Si/S0$ versus Hi para o modelo de biodiesel de girassol.

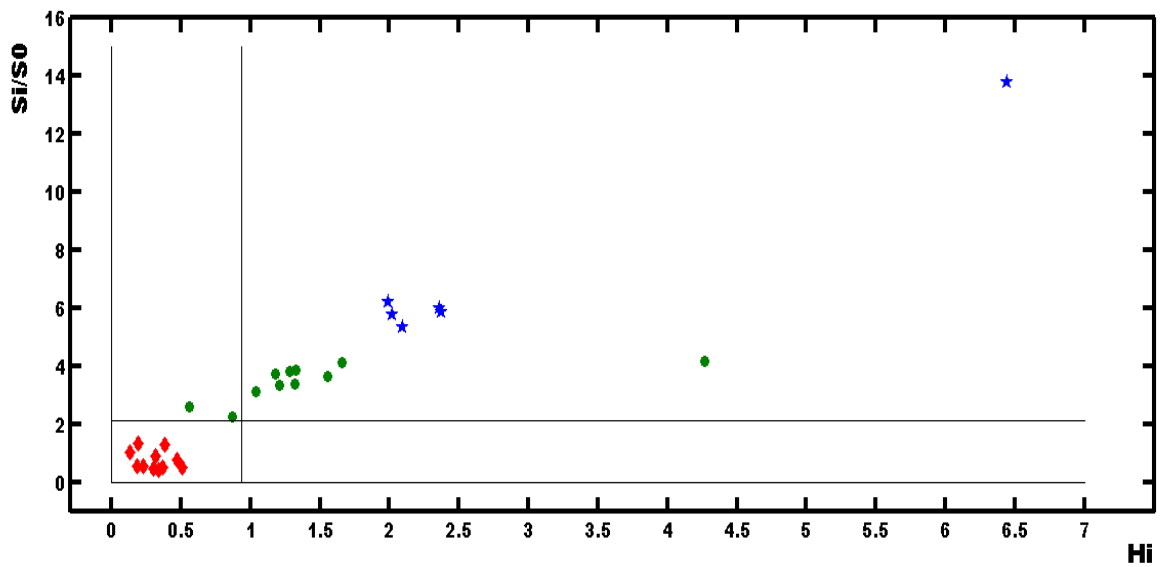


Figura 11: Resultado $Si/S0$ versus Hi para o modelo de biodiesel de soja.

ANEXO 1

CLASSIFICAÇÃO DE BIODIESEL NA REGIÃO DO VISÍVEL

Germano Veras*, Anna Luiza Bizerra de Brito, Adenilton Camilo da Silva, Priscila da Silva, Gean Bezerra da Costa e Lorena Cristina Nóbrega Félix

Departamento de Química, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, 58429-500 Campina Grande – PB, Brasil

David Douglas de Sousa Fernandes e Marcelo Marques de Fontes

Departamento de Ciências Agrárias, Universidade Estadual da Paraíba, 58.429-500 Campina Grande – PB, Brasil

Recebido em 17/3/11; aceito em 15/8/11; publicado na web em 26/9/11

BIODIESEL CLASSIFICATION IN THE VISIBLE REGION. Classification of biodiesel by oilseed type using pattern recognition techniques is described. The spectra of the samples were performed in the Visible region, requiring noise removal by use of a first derivative by the Savitzky-Golay method, employing a second-order polynomial and a window of 21 points. The characterization of biodiesel was performed using HCA, PCA and SIMCA. For HCA and PCA methods, one can observe the separation of each group of biodiesel in a spectral region of 405-500 nm. SIMCA model was used in a test group composed of 28 spectral measurements and no errors are obtained.

Keywords: biodiesel; classification; visible region.

INTRODUÇÃO

Nas últimas décadas cresceu o interesse por combustíveis renováveis, em contraposição aos derivados do petróleo,¹ de custo cada vez mais elevado por serem prospectados em áreas com acesso cada vez mais difícil e necessitando de maiores esforços tecnológicos. No Brasil, a produção de biocombustíveis surgiu como uma alternativa para diversificação da matriz energética, contribuindo para a preservação do meio ambiente,^{1,2} e para diminuir a dependência de uma fonte energética esgotável, o petróleo, apesar dos esforços em pesquisas de novos campos petrolíferos, como a descoberta de grandes reservas na camada do pré-sal.

No Brasil dois biocombustíveis se destacam: o bioálcool, produzido a partir da cana-de-açúcar, com o qual o país detém tecnologias avançadas em toda a cadeia produtiva; e o biodiesel, introduzido na matriz energética brasileira pela Medida Provisória nº. 214, de 13/9/2004.² Nesta norma, o biodiesel, combustível composto de monoalquilésteres de ácidos graxos de cadeia longa, é conceituado como “um combustível para motores a combustão interna com ignição por compressão, renovável e biodegradável, derivado de óleos vegetais ou de gorduras animais, que possa substituir parcial ou totalmente o óleo diesel de origem fóssil”.²

O biodiesel pode ser obtido por diferentes processos, dentre os quais se destaca a transesterificação de óleo vegetal ou gordura com um álcool de baixo peso molecular, como etanol ou metanol, através de uma reação catalisada por hidróxido de potássio ou de sódio.³

Apesar do diesel de petróleo comercializado ter 5% (v/v) de biodiesel,² este combustível carece de métodos de classificação de sua fonte de origem, que possa permitir um acompanhamento mais rigoroso dos padrões de qualidade de produção. Como os óleos são definidos pelos triacilgliceróis presentes e suas proporções e o método padrão de identificação e quantificação é a cromatografia gasosa, que é caro e demorado, uma alternativa viável seria desenvolver um método espectroscópico que possa ser utilizado para associar o espectro com as quantidades dos triacilgliceróis e suas proporções no óleo. Na esteira deste raciocínio, alguns trabalhos são encontrados

para reconhecimento de padrões de biodiesel por seu óleo de partida, sendo que os mesmos utilizam a região do infravermelho próximo.^{1,4} Entretanto, medidas espectroscópicas na região do NIR oferecem alguns inconvenientes, tais como, necessidade de correção de linha de base, custo relativamente elevado e dificuldade de aquisição de diodos emissores de luz e dos fototransistores como fototransdutores em toda a região do NIR, dentre outros.

Assim, a espectroscopia de absorção molecular na região UV-Vis, método baseado na interação da radiação eletromagnética na faixa espectral de 200 a 800 nm com a matéria, apresenta-se como uma alternativa viável para reconhecimento em padrões de biodiesel por permitir construção de instrumentos mais baratos e com componentes de fácil aquisição em lojas de eletrônicos mesmo em cidades de porte médio.⁵

Por outro lado, a complexidade de interpretação qualitativa dos espectros na região UV-Vis conduz à necessidade de utilização de ferramentas quimiométricas para reconhecimento de padrões. Diversos trabalhos foram publicados utilizando técnicas de reconhecimento de padrões, dentre os quais para amostras de vinhos,^{6,7} carne de porco⁸ e óleo de oliva,⁹ apesar da perspectiva conceitual de que os espectros na região do visível serem pouco discriminativos.¹⁰

As técnicas de reconhecimento de padrões podem ser divididas em duas classes: não supervisionadas e supervisionadas.¹¹ Das técnicas não supervisionadas, que têm a finalidade de verificar a formação natural de conjunto de amostras similares, ou agrupamentos, as de maior destaque são análise por componentes principais (PCA - *Principal Components Analysis*) e análise hierárquica de agrupamentos (HCA - *Hierarchical Clusters Analysis*). A técnica HCA baseia-se na similaridade e dissimilaridade entre amostras por meio do cálculo de distância entre elas.¹² A técnica PCA baseia-se na manipulação matemática dos dados originais para se obter novas variáveis, que são combinações lineares das variáveis originais e que apresentam ortogonalidade entre si, denominadas de componentes principais (PCs).¹² Em relação aos métodos supervisionados, que utilizam informação prévia sobre um conjunto de amostras conhecidas, a técnica mais utilizada é a modelagem suave independente de analogias entre classes (SIMCA - *Soft Independent Modeling of Class Analogies*), que usa PCA para cada classe definida e testa a distância das amostras

*e-mail: germano.veras@pq.cnpq.br

4. DETERMINAÇÃO DO TEOR DE BIODIESEL EM MISTURAS BIODIESEL/DIESEL UTILIZANDO ESPECTROSCOPIA UV-VIS

4.1 METODOLOGIA

4.1.1 Aquisição das Amostras

Para produção das amostras de biodiesel foram adquiridas 10 amostras de óleos refinados de soja de marcas e lotes diferentes, perfazendo 10 amostras de biodiesel. As amostras de biodiesel foram sintetizadas através de reação de transesterificação via rota metílica usando o hidróxido de potássio como catalisador (1,5 % m/v). A razão álcool óleo usada (1:8), e percentagem de catalisador foi relacionada com a massa de óleo, bem como o tempo (15 min) e a temperatura de aquecimento (45 °C).

Os biodieseis produzidos foram utilizados na preparação de misturas volumétricas com diesel. As amostras de diesel puro foram cedidas pela Petrobras Distribuidora, localizada na Cidade de Cabedelo no Estado da Paraíba. Um total de 100 amostras de misturas foi confeccionado nas devidas proporções 5% (B5), 10% (B10), 15% (B15), 20% (B20), 25% (B25), 30% (B30), 35% (B35), 40% (B40), 45% (B45) e 50% (B50), sendo 10 amostras para cada proporção.

4.1.2 Aquisição dos espectros e softwares empregados

Os espectros das misturas biodiesel/diesel foram obtidos em triplicata na faixa de 410 a 800 nm, com resolução de 1 nm, usando cubeta de quartzo com caminho óptico de 1 cm e espectrofotômetro Perkim Elmer lambda 750 com fonte de tungstênio e sistema de detecção de sulfeto de chumbo e tubo fotomultiplicador.

Os cálculos dos modelos PLS foram feitos no software The Unscrambler versão 9.8, os cálculos envolvendo o particionamento das amostras via Kennard Stones e os cálculos de regressão MLR com seleção de variáveis pelo SPA e Stepwise foram feitos em ambiente MatLab.

4.2 RESULTADOS E DISCUSSÃO

4.2.1 Aquisições dos dados Espectrais

Na **Figura 12** é apresentado o perfil espectral do conjunto de amostras de misturas biodiesel/diesel na região de 441 a 631nm. Vale ressaltar que a região de 632 a 800 nm não foi utilizada pois apresentou apenas sinal de linha de base, sendo feita, portanto, uma seleção de variáveis *a priori*.

Devido a não existência de ruído ou deslocamentos sistemáticos de linha de base na região espectral de trabalho, os espectros foram utilizados de forma bruta, ou seja, não foi realizado nenhum pré processamento. Nota-se a banda de absorção no visível com máximo em aproximadamente 530 nm, correspondente aos ésteres metílicos contendo duplas ligações conjugadas (BALABIN, 2011).

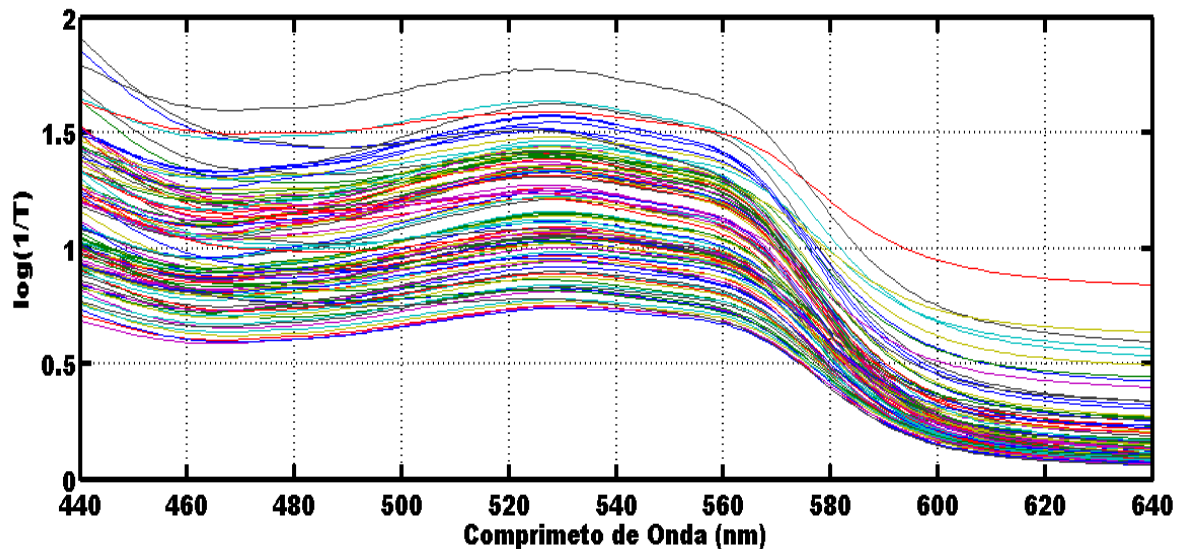


Figura 12: Perfil Espectral na faixa de 441 a 631 nm

Para a utilização do conjunto de amostras na técnicas de regressão, as amostras foram particionadas em conjunto de calibração, 60 amostras, e validação externa, 40 amostras, utilizando o algoritmo Kennard-Stone (KS), adaptado por Galvão et al (2005). O algoritmo KS utiliza cálculo de distâncias que garante que o conjunto de amostras de calibração possui uma maior dispersão em relação ao conjunto de predição.

4.2.2 Determinação do teor de biodiesel

A determinação do teor de biodiesel foi feita por três modelos de regressão: PLS empregando toda a faixa espectral, em que o número de fatores foi determinado empregando validação cruzada completa; e os cálculos MLR foram baseados em dois processos de seleção de variáveis, SPA e SW.

A **Figura 13** apresenta o gráfico do valor de referência *versus* o valor de predição para o modelo PLS. A regressão foi calculada utilizando duas variáveis latentes e nota-se que as amostras estão distribuídas em torno da bissetriz e que valores preditos são bem próximos dos valores de referência, sem presença de bias significativo.

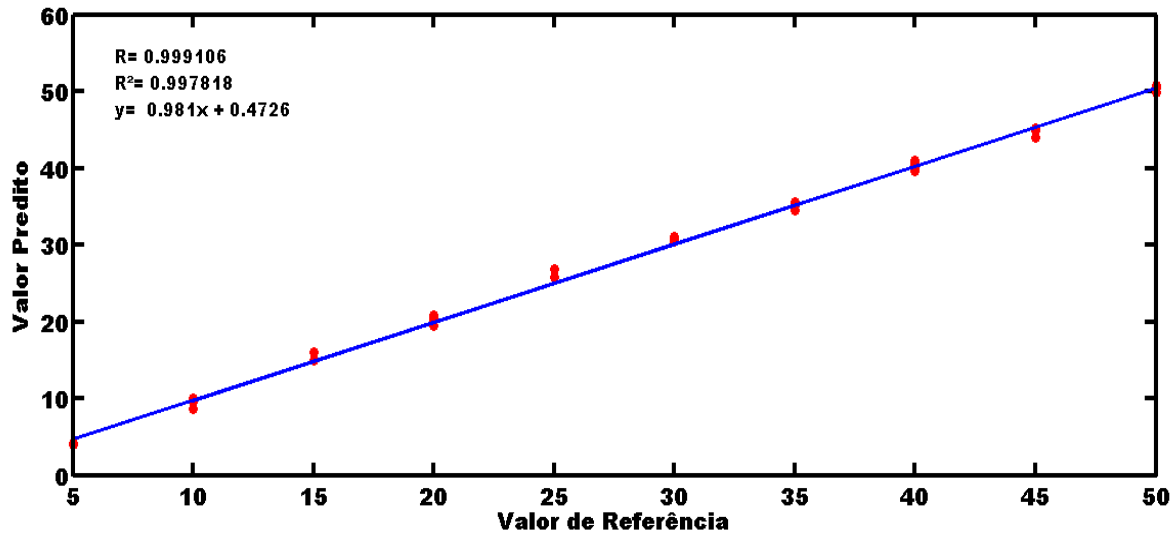


Figura 13: Resultado da regressão via PLS.

4.2.3 Seleção de Variáveis

Os algoritmos de seleção de variáveis empregados foram SPA e SW. Os resultados são apresentados nas **Figuras 14 e 15**, em que o espectro médio das amostras foi indexado aos comprimentos de onda selecionados por cada algoritmo para a faixa espectral de 441 a 631nm.

Para o Stepwise-MLR foi empregado o método misto com uma etapa *forward* seguido de uma etapa *backward*, partindo da variável de maior correlação com o parâmetro a ser determinado. O valor de alfa para inclusão ou exclusão foi mantido constante e igual a 0,05.

O SPA-MLR, utilizando validação cruzada completa, foi empregado utilizando a versão GUI- *Graphical User Interface* (PAIVA, 2012). A calibração e a previsão são realizadas em duas etapas. O modelo de calibração é construído com a introdução dos dados das matrizes de calibração (\mathbf{X}_{cal}) e de validação (\mathbf{X}_{val}), com os respectivos teores de misturas biodiesel em diesel (\mathbf{Y}_{cal} e \mathbf{Y}_{val}). Além disso, são informados os valores do número mínimo (N_1) e máximo de variáveis (N_2) que podem ser selecionadas. A etapa de predição é realizada com a inserção dos dados da matriz de calibração (\mathbf{X}_{cal}) e de previsão (\mathbf{X}_{pred}) e suas concentrações (\mathbf{Y}_{cal} e \mathbf{Y}_{pred}), bem como as variáveis selecionadas (*var_sel*) na etapa anterior.

Para a região trabalhada, o Stepwise se mostrou mais parcimonioso que o SPA, selecionando menor número de variáveis. Contudo, uma das variáveis selecionadas pelo SW está em uma região de baixa relação sinal-ruído, o que pode comprometer o poder de predição e generalização do modelo.

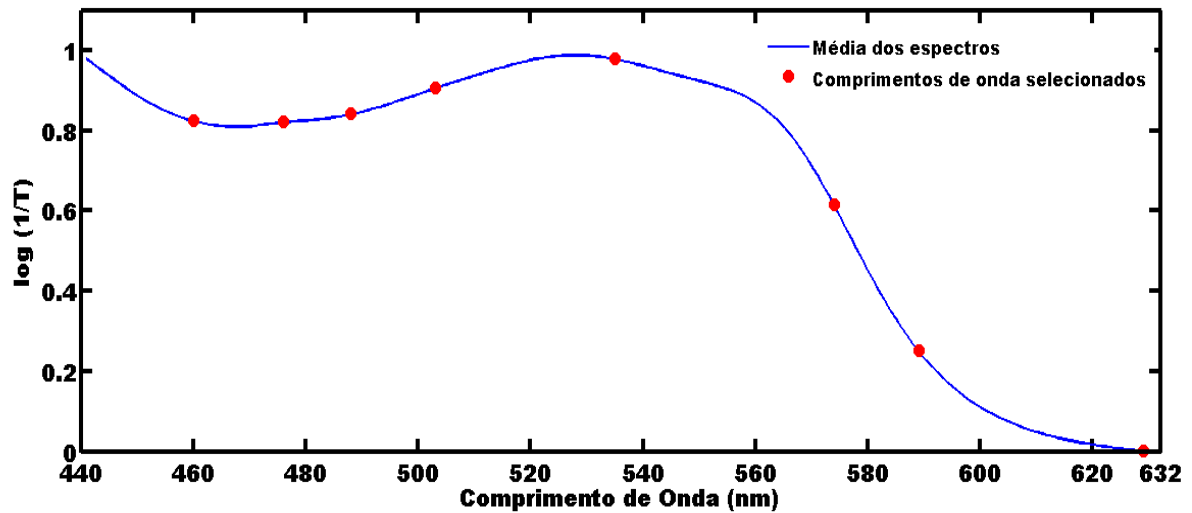


Figura 14: Comprimentos de onda selecionados do modelo SPA-MLR.

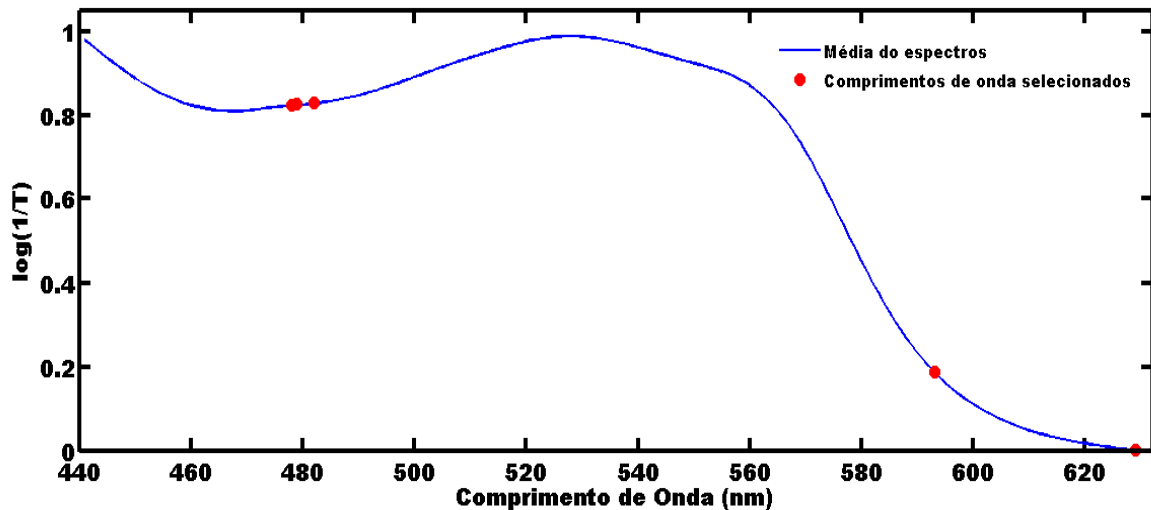


Figura 15: Comprimentos de onda selecionados do modelo SW-MLR.

Nas **Figuras 16 e 17** é apresentado o gráfico do valor predito *versus* o valor de referência para os modelos SPA-MLR e SW-MLR, respectivamente. Nenhum dos modelos apresentou bias considerável, seguindo a **equação 16**, com RMSEC, RMSECV e RMSEP em valores significativamente baixos para a faixa de mistura biodiesel/diesel considerada, conforme pode ser verificado na Tabela 1.

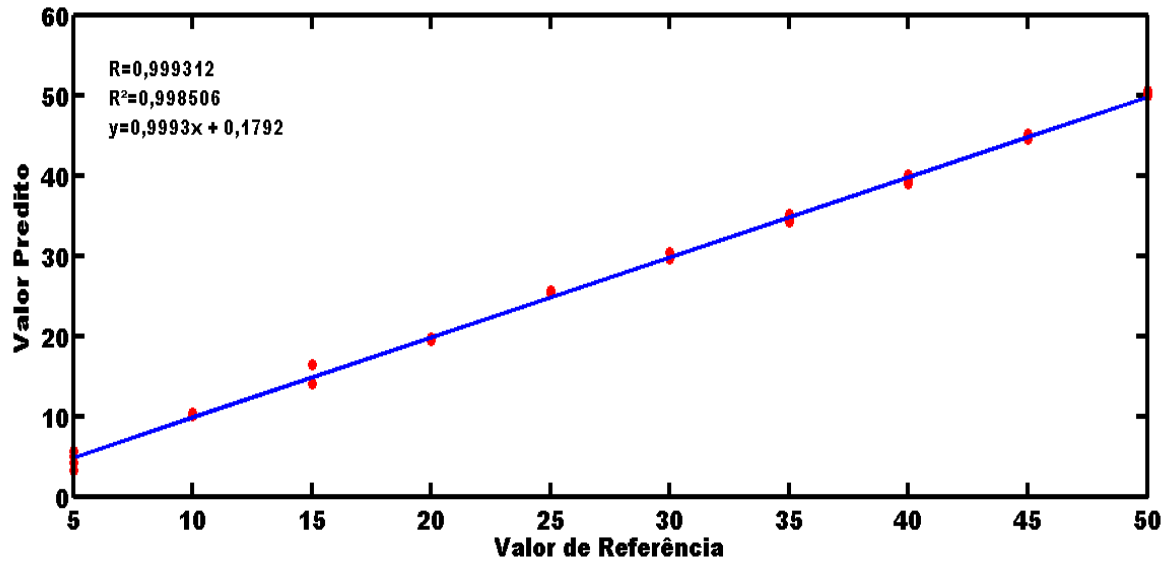


Figura 16: Resultado da regressão via SPA-MLR.

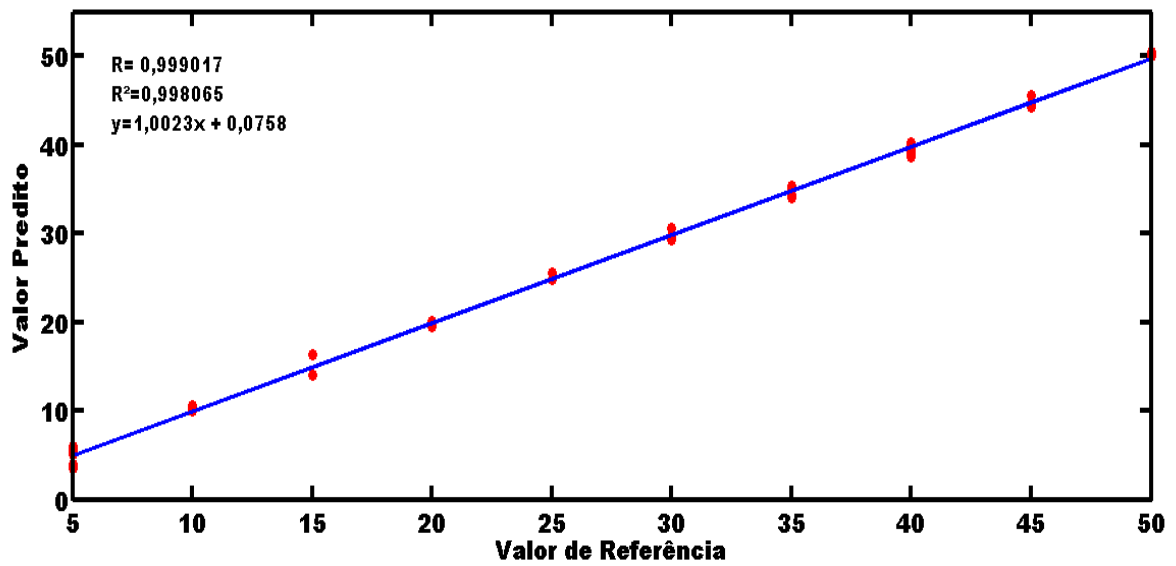


Figura 17: Resultado regressão SW-MLR.

Tabela 1: Resultados da regressão pelos modelos APS-MLR, SW-MLR e PLS.

Modelo	Parâmetros						
	RMSEC ^a (% v/v)	RMSECV ^a (% v/v)	R-square ^a	RMSEP ^b (% v/v)	bias ^b	SDV ^b	t ^b
PLS ^c	1.74	1.81	0.9949	0.66	0.14	0.70	1.22
SPA-MLR ^d	1.05	1.21	0.9944	0.57	0.16	0.65	1.55
SW-MLR ^d	1.00	1.10	0.9949	0.66	0.14	0.70	1.22

^a Conjunto de calibração

^b Conjunto de validação externo

^c Variáveis latentes usadas no modelo

^d Comprimento de onda usada no modelo

Para avaliar a robustez dos modelos de regressão, outros parâmetros estão apresentados na Tabela 1, como o Desvio Padrão de Validação (SDV) e o teste de significância estatística, com valores abaixo do valor crítico tabelado a 95 % de significância estatística (2,02). As equações 17 e 18 estão apresentados dos cálculos de SDV e teste t.

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (16)$$

$$SDV = \sqrt{\frac{\sum_{i=1}^n [(y_i - \hat{y}_i) - bias]^2}{n-1}} \quad (17)$$

$$t = \frac{|bias| \sqrt{n}}{SDV} \quad (18)$$

ANEXO 2



Determination of biodiesel content in biodiesel/diesel blends using NIR and visible spectroscopy with variable selection

David Douglas Sousa Fernandes^a, Adriano A. Gomes^b, Gean Bezerra da Costa^c, Gildo William B. da Silva^a, Germano Vêras^{a,b,*}

^a Programa de Pós-Graduação em Ciências Agrárias, Universidade Estadual da Paraíba, 58.429-500 Campina Grande – PB, Brazil

^b Departamento de Química, CCEN, Universidade Federal da Paraíba, 58.051-970, João Pessoa – PB, Brazil

^c Departamento de Química, Centro de Ciências e Tecnologia, Universidade Estadual da Paraíba, 58.429-500 Campina Grande – PB, Brazil

ARTICLE INFO

Article history:

Received 22 June 2011

Received in revised form

14 September 2011

Accepted 14 September 2011

Available online 19 September 2011

Keywords:

Biodiesel

Diesel

Near infrared spectroscopy

Visible spectroscopy

Variable selection

ABSTRACT

This work is concerned of evaluate the use of visible and near-infrared (NIR) range, separately and combined, to determine the biodiesel content in biodiesel/diesel blends using Multiple Linear Regression (MLR) and variable selection by Successive Projections Algorithm (SPA). Full spectrum models employing Partial Least Squares (PLS) and variables selection by Stepwise (SW) regression coupled with Multiple Linear Regression (MLR) and PLS models also with variable selection by Jack-Knife (Jk) were compared the proposed methodology. Several preprocessing were evaluated, being chosen derivative Savitzky-Golay with second-order polynomial and 17-point window for NIR and visible-NIR range, with offset correction. A total of 100 blends with biodiesel content between 5 and 50% (v/v) prepared starting from ten sample of biodiesel. In the NIR and visible region the best model was the SPA-MLR using only two and eight wavelengths with RMSEP of 0.6439% (v/v) and 0.5741 respectively, while in the visible-NIR region the best model was the SW-MLR using five wavelengths and RMSEP of 0.9533% (v/v). Results indicate that both spectral ranges evaluated showed potential for developing a rapid and nondestructive method to quantify biodiesel in blends with mineral diesel. Finally, one can still mention that the improvement in terms of prediction error obtained with the procedure for variables selection was significant.

© 2011 Elsevier B.V. All rights reserved.

5. ESPECTROMETRIA UV-VIS NA DETECÇÃO DE ÓLEO VEGETAL COMO ADULTERANTE EM BIODIESEL/DIESEL

5.1 METODOLOGIA

5.1.1 Aquisição das amostras

Um total de 122 amostras foi utilizado nesse estudo, sendo 32 amostras de misturas biodiesel/diesel (B5), 30 amostras de óleo diesel puro (D), fornecido pela Distribuidora Petrobras, localizada em Cabedelo, estado da Paraíba, e 60 amostras de misturas B5 adulteradas com óleo de soja cru (OB5), em que a percentagem de óleo variou de 0,5 a 2,5 % (v/v).

5.1.2 Aquisição dos espectros e softwares empregados

Os espectros do conjunto de amostras foram registrados na faixa de 430 a 850 nm, com resolução de 1 nm, em um espectrofotômetro Perkim Elmer modelo lambda 750 com fonte de tungstênio e sistema de detecção de sulfeto de chumbo e tubo fotomultiplicador.

Todos os cálculos empregados utilizaram a média dos espectros registrados em triplicata. Inicialmente o perfil de linha de base dos espectros foi corrigido empregando correção de *offset*. Na sequência, o conjunto de amostras foi particionado em conjuntos de treinamento e de teste, empregando o algoritmo Kernnard-Stone. As amostras de treinamento foram empregadas para construção de modelos SIMCA e modelos LDA com seleção de variáveis com o SPA. O conjunto de teste foi empregado para avaliar o desempenho dos mesmos. Os cálculos de pré-processamento, PCA e SIMCA foram feitos empregando o programa The Unscrambler, enquanto os cálculos envolvendo a modelagem SPA-LDA e partição do conjunto de dados foram conduzidas em ambiente MatLab 2010b.

5.2 RESULTADOS E DISCUSSÃO

5.2.1 Espectro UV-Vis

A **Figura 18** apresenta os espectros brutos das 122 amostras das misturas biodiesel/diesel, diesel e biodiesel/diesel/óleo há umas variações sistemáticas de linha de base. Cada um dos grupos (biodiesel/diesel, diesel e óleo/biodiesel/diesel) estão apresentados com cores distintas. É possível perceber a sobreposição dos espectros dos três grupos de amostras, indicando a impossibilidade de verificar a contaminação por inspeção visual das amostras e dos espectros.

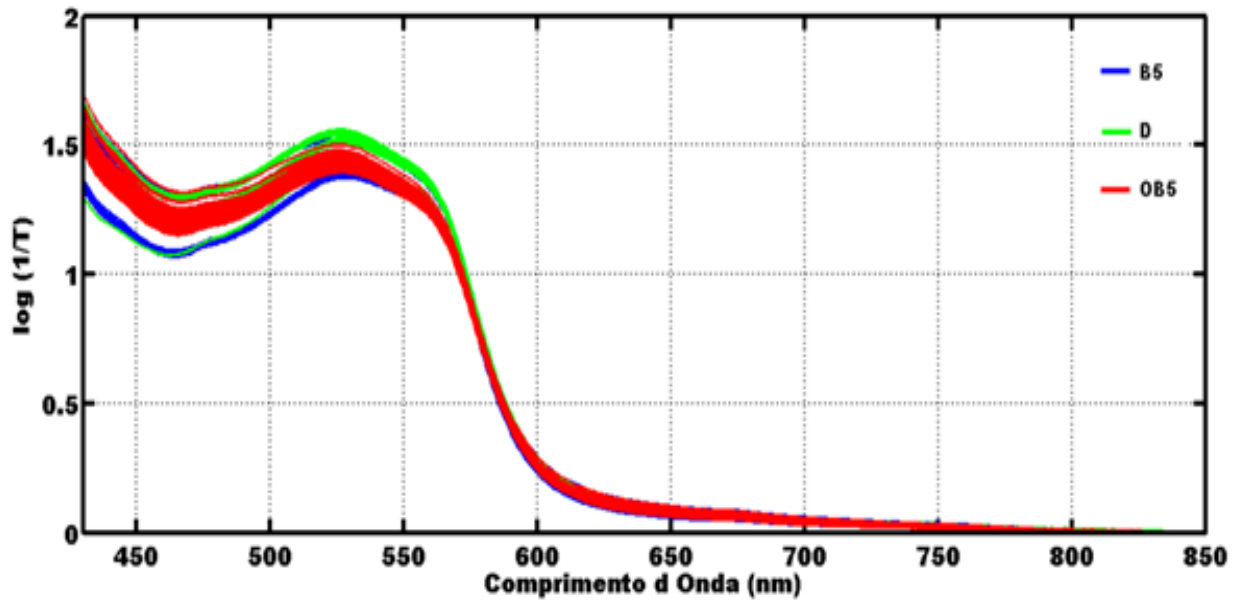


Figura 18: Espectro das 122 amostras sem tratamento.

Um pequeno deslocamento da linha de base foi percebido e para superar este inconveniente foi usado correção de *offset*. Tal técnica é adotada para reduzir as contribuições não desejadas dos sinais, que diminuem a capacidade de preditiva do modelo. A **Figura 19** apresenta o perfil do espectro pré-processados.

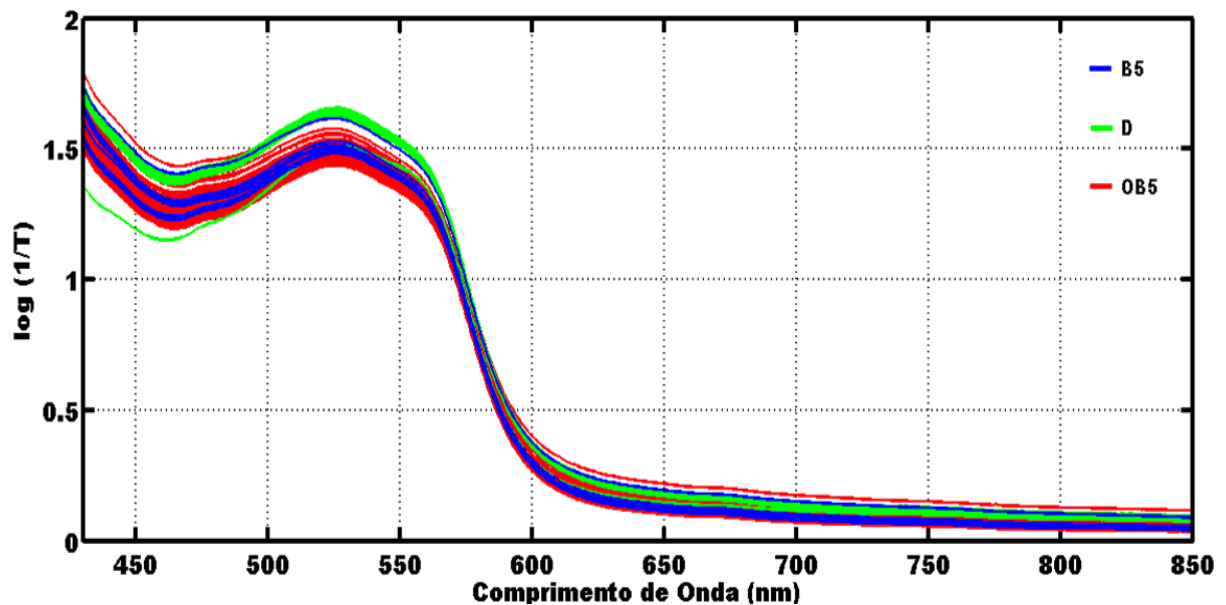


Figura 19: Espectros das 122 amostras pós off set.

5.2.2 Modelo PCA

A **Figura 20** apresenta o gráfico de scores PC2 × PC1 resultante da aplicação da PCA em todo conjunto de amostras. A percentagem de variância explicada está indicada em cada eixo. Como pode ser visto, não há nenhuma sobreposição entre as demais classes com as amostras de diesel puro, contudo as amostras de biodiesel/diesel e biodiesel/diesel adulteradas com óleo vegetal apresentam sobreposição, certamente devido às semelhanças químicas entre o óleo e o biodiesel.

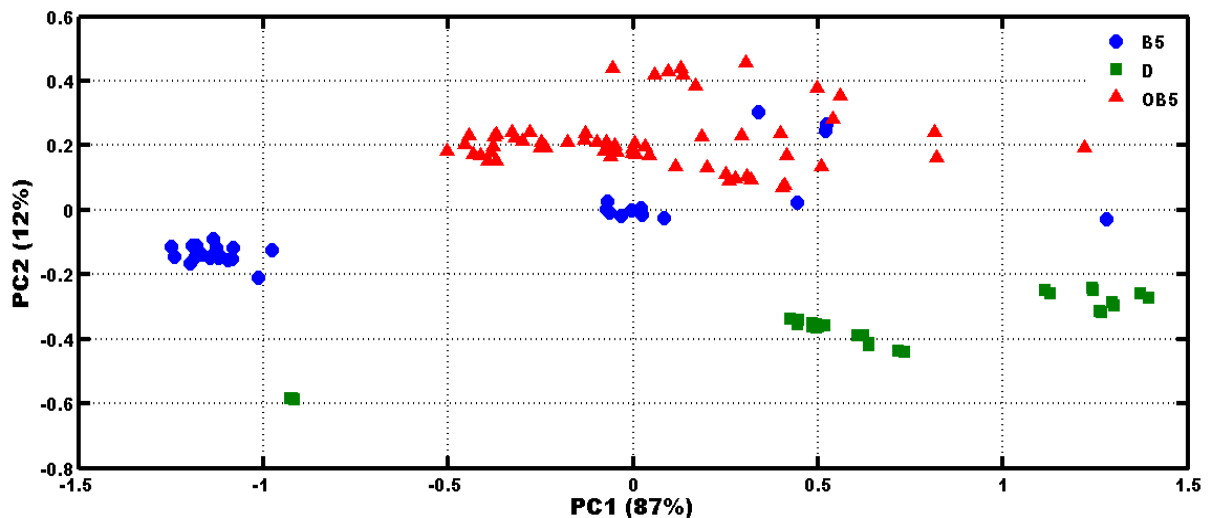


Figura 20: Gráfico de scores de PC1X PC2 para o conjunto de amostra.

Os resultados obtidos na fase de análise exploratória indicam que os espectros UV-Vis portam informação capaz de detectar adulteração em misturas biodiesel/diesel com óleo vegetal cru, ou mesmo a total ausência do biodiesel no diesel.

5.2.3 Classificação SIMCA

Um modelo SIMCA usando o espectro completo foi construído para cada classe em estudo: 5 % (v/v) de biodiesel em diesel (B5), diesel puro (D), misturas óleo/biodiesel/diesel (OB5), e os resultados em termos de erros de classificação para os conjuntos de treinamento validação e teste estão apresentados na **Tabela 2**, e o nível de significância adotado no teste F foi de 5%.

Os resultados obtidos com os modelos SIMCA são satisfatórios e estão em concordância com a etapa de análise exploratória. Nenhuma amostra de diesel foi classificada incorretamente, entretanto uma amostra do conjunto de treinamento da classe das misturas biodiesel/diesel adulterada foi classificada como sendo amostra de classe B5. Estes resultados são promissores e

atestam a potencialidade da espectrometria UV-Vis como ferramenta para o controle de qualidade de biodiesel.

Tabela 2: Resumo do resultado da classificação do modelo SICMA

Classes	Conjunto					
	Treinamento		Validação		Teste	
	I ^a	II ^b	I ^a	II ^b	I ^a	II ^b
BD (3)	0	0	0	0	0	0
D (4)	0	0	0	0	0	0
BDO (4)	0	1	0	0	0	0

^a Conjunto de calibração

^b Conjunto de validação externo

5.2.4 Classificação APS- LDA

O número ideal de variáveis para o SPA-LDA foi determinado a partir do mínimo da função G custo exibido na **Figura 21**. Como pode ser visto, um mínimo bem localizado é obtido com três variáveis. Estas variáveis correspondem aos comprimentos de onda 430, 563 e 672 nm, tal como indicado na **Figura 22**.

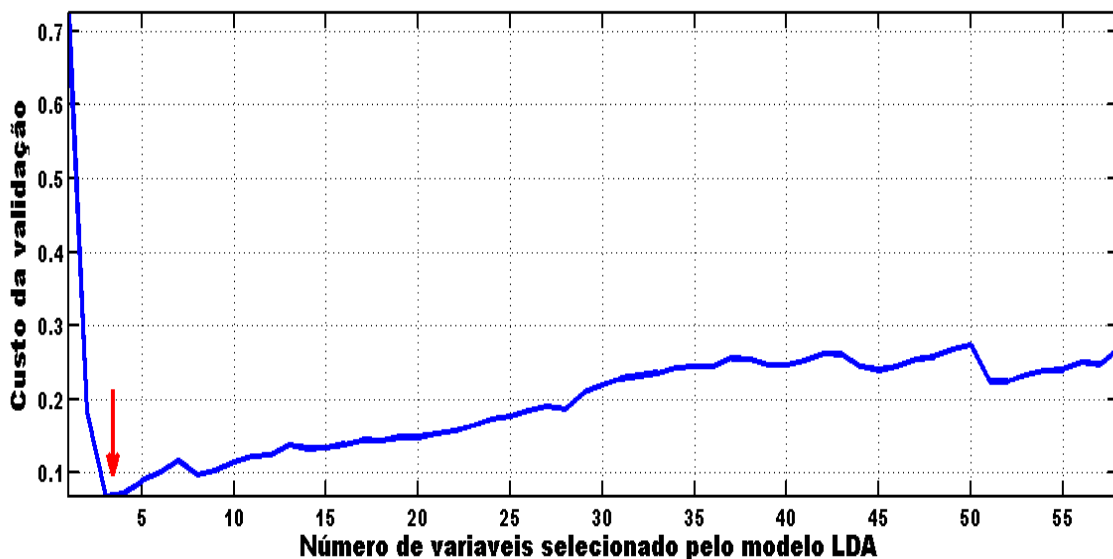


Figura 21: Gráfico da função de custo versus o número de onda selecionado pelo modelo APS-LDA.

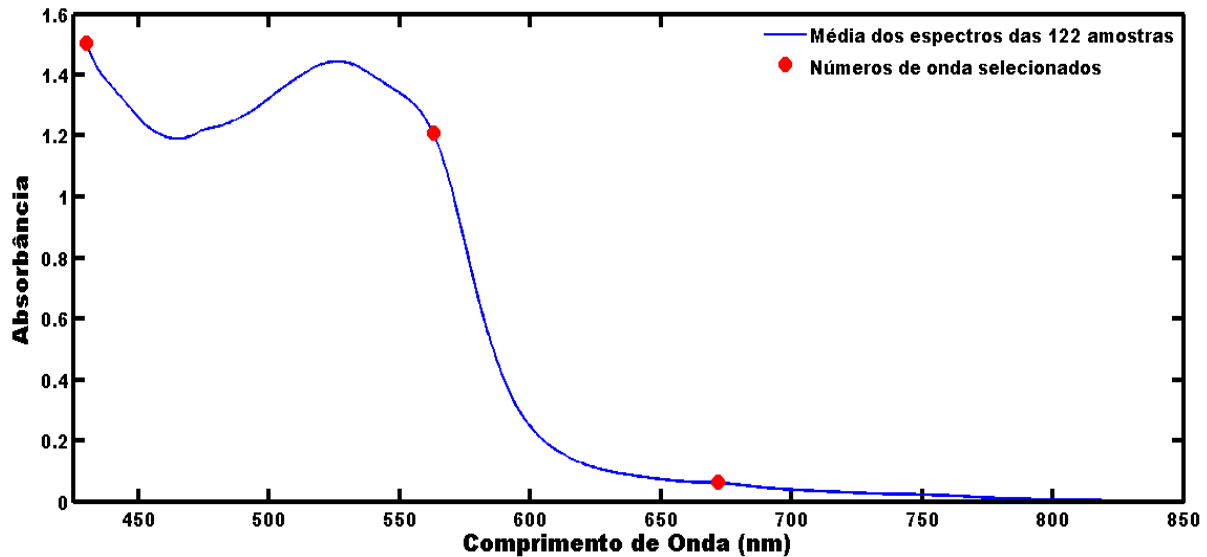


Figura 22: Número de onda ideal do modelo APS-LDA

O modelo LDA obtido com as 3 variáveis selecionadas por SPA foi utilizada para a classificação de um conjunto de teste e como resultado, todas as amostras foram classificadas corretamente. O bom desempenho de classificação do modelo SPA-LDA é ainda demonstrado na **Figura 23**, que apresenta os escores das duas primeiras funções discriminantes (DF2 *versus* DF1) para o conjunto de dados global. Como pode ser visto, a separação das três classes é muito melhor em comparação ao gráfico de escores de PC2 *versus* PC1, da **Figura 20**, pois havia leve sobreposição no modelo PCA das amostras pertencente ao conjunto de B5 com as amostras OB5.

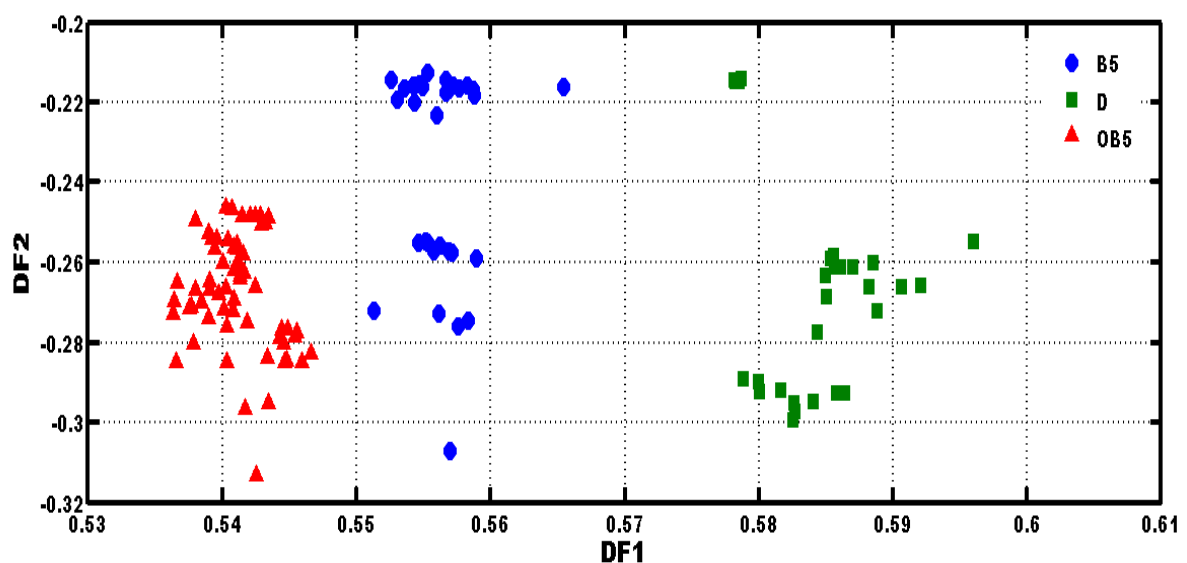


Figura 23: Gráfico de scores da função discriminante (DF1x DF2) para conjunto de amostras.

ANEXO 3

**UV-Vis spectrometric detection of soy oil adulterations in
biodiesel/diesel blends**

Journal:	<i>Analytical Methods</i>
Manuscript ID:	AY-ART-12-2012-026503
Article Type:	Paper
Date Submitted by the Author:	04-Dec-2012
Complete List of Authors:	<p>Fernandes, David Douglas; Universidade Estadual da Paraiba, Mestrado em Ciências Agrárias Gomes, Adriano; Federal University of Paraiba, Chemistry Fontes, Marcelo; Universidade Estadual da Paraiba, Mestrado em Ciências Agrárias Costa, Gean; Universidade Estadual da Paraiba, Química Almeida, Valber; Universidade Estadual da Paraiba, Química Porto, Katiane; Universidade Estadual da Paraiba, Química Uquino Araujo, Mario Cesar; Universidade Federal da Paraiba, Quimica Galvão, Roberto; Instituto Tecnológico de Aeronáutica, Divisão de Engenharia Eletrônica Veras, Germano; Universidade Estadual da Paraiba, Chemistry</p>

6. CONCLUSÃO

A classificação das amostras de biodiesel a partir dos óleos de origem ocorreu com 100% classificada corretamente na devida classe.

Com os resultados obtidos pode-se demonstrar que é possível determinar o teor de biodiesel em misturas biodiesel/diesel empregando a região do visível. Também é possível observar que o procedimento de seleção de variáveis aumenta o poder preditivo quando comparado com os modelos PLS.

O conjunto de amostra adulterado com óleos de soja empregando espectrometria UV-Vis e técnicas de reconhecimento de padrões. As duas modelagens investigadas (SIMCA e SPA-LDA) mostraram bom desempenho classificando corretamente todas as amostras do conjunto de teste, contudo no modelo SIMCA cometeu erros de classificação no conjunto de treinamento, que não ocorreu no modelo SPA-LDA. Não menos importante podemos mencionar a simplicidade do modelo SPA-LDA que emprega apenas três comprimentos de onda.

Pode-se concluir que a utilização de espectroscopia UV-Vis com métodos quimiométricos podem fornecer informações valiosas sobre a presença de óleo cru como adulterante em misturas biodiesel/diesel.

A metodologia proposta mostrou-se satisfatória para classificação de biodiesel, de forma rápida e não destrutiva, podendo ser empregada no controle de qualidade na forma de análise do tipo *screening*. A utilização de seleção de variáveis permite possibilidade de implementar esta metodologia na construção de fotômetros dedicados baratos e com componentes de fácil aquisição na classificação de biodiesel para suporte aos órgãos fiscalizadores, às empresas que comercializam e aos consumidores, aponta a relevância deste trabalho.

REFERÊNCIAS

AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS SUPERINTENDÊNCIA DE REFINO E PROCESSAMENTO DE GÁS NATURAL – SRP, BOLETIM MENSAL DE BIODIESEL, 2012.

ANDERSSON, M.A Comparison of nine PLS1 algorithms. **Journal Chemometrics**, v.23, p.518–529, 2009.

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 spectroscopic multicomponent analysis. **Chemometrics and Intelligent Laboratory Systems**, vol 65, pág 57, 2001.

ATADASHI, I.M; AROUA, M. K; AZIZ, A. R. A; SULAIMAN, N. M. N; The effects of water on biodiesel production and refining technologies: A review. **Renewable and Sustainable Energy Reviews**, v.16, p. 3456-3470, 2012.

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, p.2007–2015, 2011.

BALABIN, R. M; SMIRNOV, S. V. Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel. **Analytica Chimica Acta**, v.692, p.63–72, 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.

BARROS NETO, B., PIMENTEL, M. F., ARAÚJO, M. C. U. Recomendações para calibração em química analítica- Parte I. Fundamentos e calibração com um componente (Calibração Univariada). **Química Nova**, v.25, p.856 -865, 2002.

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

BERRUETA, L. A.; ALONSO-SALCES, R. M.; HÉBERGER, K., Supervised pattern recognition in food analysis. **Journal of Chromatography A**, v.1158, p.196–214, 2007.

BORGES NETO, W. **Parâmetros de qualidade de lubrificantes e óleos de oliva através de espectroscopia vibracional, calibração multivariada e seleção de variáveis**. 2006.55f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Estadual de Campinas/Instituto de Química.

BRERETON, R. G. **Chemometrics: data Analysis for the laboratory and chemical plant**. New York: John Wiley & Sons, 2003.

BRERETON, R. G. **Introduction to multivariate calibration in analytical chemistry**. **Analyst**, v.125, p.2125-2154, 2000.

CALAND, L. B.; SILVEIRA, E. L. C.; TUBINO, M. Determination of sodium, potassium, calcium and magnesium cations in biodiesel by ion chromatography. **Analytica Chimica Acta**, v.718, p.116– 120, 2012.

CANDOLFI, A.; MAESSCHALCK, R. D.; MASSART, D. L.; HAILEY, P. A.; HARRINGTON, A. C. E., Identification of pharmaceutical excipients using NIR spectroscopy and SIMCA. **Journal of Pharmaceutical and Biomedical Analysis**, v.19, p.923–935, 1999.

CHAROENCHAITRAKOOL, M.; THIENMETHANGKOON, J. Statistical optimization for biodiesel production from waste frying oil through two-step catalyzed process. **Fuel Processing Technology**, v.92, p.112-118, 2011.

CODEX ALIMENTARIUS COMMISSION - FAO/WHO. Codex Alimentarius . Fats, oils and related products. 4 ed. Roma: Secretariat of the Joint FAO/WHO Food Standards Programme, p.1-16. 2011.

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

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, p.481–487, 2007.

DABDOUB, M.; BRONZEL, J.; RAMPIN, M.A. Biodiesel: visão crítica do status atual e perspectivas na academia e na indústria. **Química Nova**, v.32, p.776-792, 2009.

DU, Y.Y.; Xu, D.H.; Liu, J. Comparative Study on Lipase-Catalyzed Transformation of Soybean Oil for Biodiesel Production With Different Acyl Acceptors. **Journal of Molecular Catalysis B: Enzymatic**, v.30, p.125-129, 2004.

FERRÃO, M. F; VIERA, M.D. S; PAZOS, R.E.P; FACHINI, D; GERBASE, A. E; MARDER, L., Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. **Fuel**, v.90, p.701-706, 2011.

FERRARI, R.A.; OLIVEIRA, V.S.; SCABIO, A. Biodiesel de soja – taxa de conversão em ésteres etílicos, caracterização físico-química e consumo em gerador de energia. **Química Nova**, v.28, p.19-23, 2005.

FERRARI, R. A.; SOUZA, W. L. Avaliação da estabilidade oxidativa de biodiesel de óleo de girassol com antioxidantes. **Química Nova**, v.32, p.106-111, 2009.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOLPE, L. O. Quimiometria I: calibração multivariada, um tutorial. **Química Nova**, v.22, p.724–731, 1999.

FREITAS, S. K. B.; **Uma metodologia para screening analysis de sucos cítricos utilizando um analisador automático em fluxo-batelada, espectrometria UV-VIS e técnicas quimiométricas.** 2006.55f Dissertação (Mestrado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba.

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 subset partitioning. **Talanta**, v.67, p.736–740, 2005.

GALVÃO, R. K. H. e ARAÚJO, M. C. U. variable selection. in: B. WALCZAK, R. T. FERRÉ e S. BROWN. *Linear regression modelling*. **Comprehensive chemometrics**, v.3, p.233–283, 2009.

GARCIA, J. R.; ROMEIRO, A. R. Desafios para a produção de biodiesel por produtores familiares no semiárido brasileiro. **Informações Econômicas**, v.40, p.5–17, 2010.

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, p.1369–1373, 2007.

GOLDEMBERG, J.; COELHO, S. T.; PLÍNIO, M.N.; LUCOND, O. Ethanol learning curve - the Brazilian experience. **Biomass and Bioenergy**, v.26, p.301–304, 2004.

GHANEI, R.; MORADI, G.R.; TAHERPOURKALANTARI, R.; ARJMANDZADEH, E. Variation of physical properties during transesterification of sunflower oil to biodiesel as an approach to predict reaction progress. **Fuel Processing Technology**, v.92, p.1593–1598, 2011.

GHASEMI-VARNAMKHAZI M.; MOHTASEBI, S. S.; RODRIGUEZ-MENDEZA, M. L.; 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.

GOMES, A. A. **Algoritmo das projeções sucessivas aplicado à seleção de variáveis em regressão PLS**. 2012.120f Dissertação (Mestrado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba.

GONZÁLEZ, A. G., Use and misuse of supervised pattern recognition methods for interpreting compositional data. **Journal of Chromatography A**, v.1158, p.215–225, 2007.

GÜRÜ, M.; ARTUKOGLU ,B. D.; KESKIN, A.; KOCA, A. Biodiesel production from waste animal fat and improvement of its characteristics by synthesized nickel and magnesium additive. **Energy Conversion and Management**, v.50, p.498–502, 2009.

HALIM, S.F.A.; KAMARUDDIN, H.; FERNANDO, W.J.N. Continuous Biosynthesis of Biodiesel from Waste Cooking Palm Oil in a Packed Bed Reactor: Optimization Using Response Surface methodology (RSM) and Mass Transfer Studies. **Bioresource Technology**, v.100, p.710–716, 2009.

HAYYAN, A.; MJALLI, F. S.; HASHIM, M. A.; HAYYAN, M.; ALNASHEF, I. M.; AL-ZAHRANI, S. M., AL-SAAD, M. A. Ethanesulfonic acid-based esterification of industrial acidic crude palm oil for biodiesel production. **Bioresource Technology**, v.102, p.9564–9570, 2011.

HIBBERT, D. B., et al. IUPAC project: A glossary of concepts and terms in chemometrics. **Analytica Chimica Acta**, v.642, p.3–5, 2009.

HOPKE, P. K. The evolution of chemometrics. **Analytica Chimica Acta**, v.500, p.365–377, 2003.

HUANG, Y.H.; WU, J.H. Analysis of biodiesel promotion in Taiwan. **Renewable and Sustainable Energy Reviews**, v.12, p.1176–1186, 2008.

INMETRO, *Orientação sobre validação de métodos de ensaios químicos*. DOQ-CGCRE/0008; 2003.

KARMAKAR, A.; KARMAKAR, S.; MUKHERJEE, S. Properties of various plants and animals feedstocks for biodiesel production. **Bioresource Technology**, v.101, p.7201–7210, 2010.

KUMAR, D.; KUMAR, G.; JOHARI, R.; KUMAR, P. Fast, easy ethanomethanolysis of *Jatropha curcus* oil for biodiesel production due to the better solubility of oil with ethanol in reaction mixture assisted by ultrasonication. **Ultrasonics Sonochemistry**, v.19, p.816–822, 2012.

LAKSMONO, N.; PARASCHIV, M.; LOUBAR, K.; TAZEROUT, M. Biodiesel production from biomass gasification tar via thermal/catalytic cracking. **Fuel Processing Technology**, v.106, p.776–783, 2013.

LIAN, S.; LI, H.; TANG, J.; TONG, D. Changwei Hu Integration of extraction and transesterification of lipid from jatropha seeds for the production of biodiesel. **Applied Energy**, v.98, p.540-547, 2012.

LIMA, J.R.O.; SILVA, R.B.; SILVA, C.C.M.; SANTOS, L.S.S.; SANTOS JR, J.R.; MOURA, E.M.; MOURA, C.V.R. Biodiesel de babaçu (*orbignya* sp.) obtido por via etanólica. **Química Nova**, v.30, p.600-603, 2007.

EUROPEAN STANDARD – EN 14078:2009, Liquid Petroleum Products – Determination of Fatty Methyl 170 Ester (FAME) Content in Middle Distillates – Infrared Spectrometry Method.

LEI Nº 11.097, 13/1/2005 - dispõe sobre a introdução do biodiesel na matriz energética brasileira; altera as Leis nºs 9.478, de 6/8/1997, 9.847, de 26/10/1999 e 10.636, de 30/12/2002; e dá outras providências; *Diário Oficial da União*, 14/1/2005, seção 1, p. 8.

LIMA, P. C. R. “O BIODIESEL E A INCLUSÃO SOCIAL”. **Câmara dos Deputados**, p. 1-33. 2004.

LIU, F.; HE, Y. Application of successive projections algorithm for variable selection to determine organic acids of plum vinegar. **Food Chemistry**, v.115, p.1430–1436, 2009.

LÔBO, I. P.; FERREIRA, S. L. C.. Biodiesel: parâmetros de qualidade e métodos analíticos. **Química Nova**, v.32, p.1596-1608, 2009.

MEDIDA PROVISÓRIA Nº 214.Nota descritiva: consultoria Legislativa da Câmara dos Deputados. Brasília, 9 de setembro de 2004.

MEKHILEF, S.; SIGA, S.; SAIDUR, R. A review on palm oil biodiesel as a source of renewable fuel. **Renewable and Sustainable Energy Reviews**, v.15, p.1937–1949, 2011.

MONTEIRO, M. R; AMBROZINA, A. R. P; LIÃO, L. M; FERREIRA, A. G. Critical review on analytical methods for biodiesel characterization. **Talanta**, v.77, p.593-605, 2008.

MORADI, G.R.; DEGHANI, S.; KHOSRAVIAN, F.; ARJMANDZADEH, A. The optimized operational conditions for biodiesel production from soybean oil and application of artificial neural networks for estimation of the biodiesel yield. **Renewable Energy**, v.50, p.915-920, 2013.

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

NABI, N.; RAHMAN, M.; AKHTER, S. Biodiesel from cotton seed oil and its effect on engine performance and exhaust emissions. **Applied Thermal Engineering**, v.29, p.2265-2270, 2009.

NASCIMENTO, E. C. L. **Um fotômetro microcontrolado LED-NIR, portátil e de baixo custo para análise screening de gasolinas tipo C**. 2008.110f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba.

NETO, J. G. V., **Um espectrômetro microcontrolado baseado em LED branco como fonte de radiação e mídia de CD como grade de difração**. 2008.116f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba.

NUNES, P. G. A. **Uma nova técnica para seleção de variáveis em calibração multivariada aplicada às espectrometrias UV-VIS e NIR**. 2008.106f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba.

ONG, H.C.; MAHLIAA, T.M.I.; MASJUKIA, H.H.; NORHASYIMAB, R.S. Comparison of palm oil, *Jatropha curcas* and *Calophyllum inophyllum* for biodiesel: A review. **Renewable and Sustainable Energy Reviews**, v.15, p.3501–3515, 2011.

PAIVA, H. M.; SOARES, S. F. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. A graphical user interface for variable selection employing the Successive Projections Algorithm. **Chemometrics and Intelligent Laboratory Systems**, v.118, p.260–266, 2012.

PAPADOPOULOS, C. E.; LAZARIDOU, A.; KOUTSOUMBA, A.; KOKKINOS, N.; CHRISTOFORIDIS, A.; NIKOLAOU, N. Optimization of cotton seed biodiesel quality (critical properties) through modification of its FAME composition by highly selective homogeneous hydrogenation. **Bioresource Technology**, v.101, p.1812-1819, 2010.

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, p.9-15, 1998.

PINTO, A.C.; GUARIEIRO, L.L. N.; REZENDE, M.J.C.; RIBEIRO, N. M.; TORRES, A. E.; LOPES, W.A.; PEREIRA, P.A.; ANDRADE, j. B. Biodiesel: an overview. **Journal of the Brazilian Chemical Society**, v.16, p.1313-1330, 2005.

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.

PONTES, M. J. C.; PEREIRA, C. F.; PIMENTEL, M. F., VASCONCELOS, F. V. C.; SILVA, A. G. B. Screening analysis to detect adulteration in diesel/biodiesel blends using near infrared spectrometry and multivariate classification. **Talanta**, v.85, p.2159–2165, 2011.

QI, D.H.; CHEN, H.; GENG, L.M.; BIAN, Y.Z.H.; REN, X.C.H. Performance and combustion characteristics of biodiesel–diesel–methanol blend fuelled engine. **Applied Energy**, v.87, p.1679-1686, 2010.

QIU, F.; LI, Y.; YANG D.; LI, X.; SUN P. Biodiesel production from mixed soybean oil and rapeseed oil. **Applied Energy**, v.88, p.2050–2055, 2011.

RAMOS, G. R.; ÁLVAREZ-COQUE, M. G. **Quimiometria**. Valle hermoso: Síntesis, 2001.

REYES, M.A.; SEPÚLVEDA, P. M. 10 Emissions and Power of a Diesel Engine Fueled With Crude and Refined Biodiesel from Salmon oil. **Fuel**, v.85, p.1714-1719, 2006.

RODRIGUEZ, A.M.G., TORRES, A.G.; CANO PAVON, J.M.; BOSCH OJEDA, C. Simultaneous determination of iron, cobalt, nickel and copper by UV-visible spectrophotometry with multivariate calibration. **Talanta**, v.47, p.463-470, 1998.

SALOUA, F.; SABER, C.; HEDI, Z. Methyl ester of [Maclura pomifera (Rafin.) Schneider] seed oil: Biodiesel production characterization. **Bioresource Technology**, v.101, p.3091-3096, 2010.

SANTANA, A.; MAÇAIRA, J.; LARRAYOZ, M. A. Continuous production of biodiesel from vegetable oil using supercritical ethanol/ carbon dioxide mixtures. **Fuel Processing Technology**, v.96, p.214–219, 2012.

SENGO, I; GOMINHO, J; D'OREY, L; MARTINS, M; DUARTE, E. A; PEREIRA, H; DIAS, S. F. Response surface modeling and optimization of biodiesel production from Cynara Cardunculus Oil. **European Journal of Lipid Science and Technology**, v.112, p.310-320, 2010.

SIMÕES, S.S. **Desenvolvimento de métodos validados para a determinação de captopril usando espectrometria NIRR e calibração multivariada**. 2008.84f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba.

SINHA, S.; AGARWAL, A.K.; GARG, S. Biodiesel development from rice bran oil: Transesterification process optimization and fuel characterization. **Energy Conversion and Management**, v.49, p.1248-1257, 2008.

SKOOG, D. A. E LEARY, J. J. **Principles of instrumental analysis**. 6. ed. New York: Saunders College Publishing, 1992.

SKOOG, D.A., HOLLER, F.J., NIEMAN, T.A. **Princípios de Análise Instrumental**. 5. ed. Rio de Janeiro: Bookman, 2002.

SOARES, S. F. C. **Um novo critério para seleção de variáveis usando o Algoritmo das Projeções Sucessivas**. 2010.107f Dissertação (Mestrado em Química) – Curso de Pós Graduação em Química, Universidade Federal da Paraíba.

SOARES, S. F. C.; GOMES, A. A.; FILHO, A. R. G.; ARAÚJO, M. C. U.; GALVÃO, R. K. H. The successive projections algorithm. **Trends in Analytical Chemistry**, v.42, p.84-96, 2013.

SOUTO, U. T. C. P.; PONTES, M. J. C.; SILVA, E. C.; GALVÃO, R. K. H.; ARAÚJO, M. C. U.; SANCHES, F. A. C.; CUNHA, F. A. S.; OLIVEIRA, M. S. R. UV-Vis spectrometric classification of coffees by SPA-LDA. **Food Chemistry**, v.119, p.368-371, 2010.

STERNBERG, J. C.; STILLS, H. S.; SCHWENDEMAN, R. H. Spectrophotometric analysis of multi-component systems using the least squares method in matrix form. **Analytical Chemistry**, v.32, p.84-90, 1960.

SUAREZ, Paulo A. Z.; MENEGHETTI, Simoni M. Plentz Aniversário do biodiesel em 2007: Evolução histórica e situação atual no Brasil. **Química Nova**, v.30, p.2068-2071, 2007.

TASHTOUSH, G. M.; AL-WIDYAN, M. I.; AL-JARRAH, M. M. Experimental study on evaluation and optimization of conversion of waste animal fat into biodiesel. **Energy Conversion and Management**, v.45, p.2697-2711, 2004.

THAMSIRIROJ, T.; MURPHY, J.D. Is it better to import palm oil from Thailand to produce biodiesel in Ireland than to produce biodiesel from indigenous Irish rape seed?. **Applied Energy**, v.86, p.595-604, 2009.

THIRU, M.; SANKH, S.; RANGASWAMY, V. Process for biodiesel production from *Cryptococcus curvatus*. **Bioresource Technology**, v.102, p.10163-10776, 2011.

VALDERRAMA, P. **Calibração multivariada de primeira e segunda ordem e figuras de mérito na quantificação de enantiômeros por espectroscopia**. 2009.230f Tese (Doutorado em Química) – Curso de Pós Graduação em Química, Universidade Estadual de Campinas/ Instituto de Química.

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

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

VERAS, G; FERNANDES, D. D. S.; GOMES, A. A.; Costa, G. B.; Silva, G. W. B. Determination of biodiesel content in biodiesel/diesel blends using NIR and visible spectroscopy with variable selection. **Talanta**, v.87, p.30–34, 2011.

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

WANG, S.; LIU, K.; YU, X.; WU, D.; He, Y. Application of hybrid image features for fast and non-invasive classification of raisin. **Journal of Food Engineering**, v.109, p.531–537, 2012.

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

WOLD, S., STÖSTRÖM, M., ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v.58, p.109-130, 2001.