



**Universidade do Estado do Rio de Janeiro**

Centro de Tecnologia e Ciências

Instituto de Química

Juliana Verdan da Silva

**Uso da calibração multivariada para a predição de propriedades físico-químicas de misturas de óleo de soja e biodiesel**

Rio de Janeiro

2015

Juliana Verdan da Silva

**Uso da calibração multivariada para a predição de propriedades físico-químicas de misturas de óleo de soja e biodiesel**

Dissertação apresentada, como requisito parcial para a obtenção do título de Mestre, ao Programa de Pós-Graduação em Engenharia Química, da Universidade do Estado do Rio de Janeiro. Área de concentração: Processos Químicos, Petróleo e Meio Ambiente.

Orientadoras: Prof.<sup>a</sup> Dr.<sup>a</sup> Lucia Regina Raddi de Araújo  
Prof.<sup>a</sup> Dr.<sup>a</sup> Cristiane Assumpção Henriques

Rio de Janeiro

2015

CATALOGAÇÃO NA FONTE  
UERJ/REDE SIRIUS/CTC/Q

S586 Silva, Juliana Verdan da  
Uso da calibração multivariada para a predição de propriedades físico-químicas de misturas de óleo de soja e biodiesel./ Juliana Verdan da Silva. – 2015.  
95 f.

Orientador: Lucia Regina Raddi de Araújo.  
Cristiane Assumpção Henriques  
Dissertação (mestrado) – Universidade do Estado do Rio de Janeiro, Instituto de Química.

1. Biodiesel – Teses. 2. Calibração multivariada – Teses. 3. Óleo de soja– Teses. I. Araújo, Lucia Regina Raddi de. II. Henriques, Cristiane Assumpção. III Universidade do Estado do Rio de Janeiro. Instituto de Química. IV. Título.

CDU 662.756.3

Autorizo, apenas para fins acadêmicos e científicos, a reprodução total ou parcial desta tese.

---

Assinatura

---

Data

Juliana Verdan da Silva

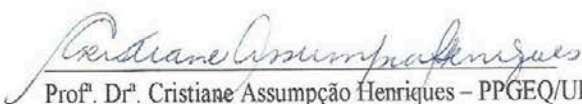
**Uso da calibração multivariada para a predição de propriedades físico-químicas de misturas de óleo de soja e biodiesel**

Dissertação apresentada, como requisito parcial para a obtenção do título de Mestre, ao Programa de Pós-Graduação em Engenharia Química, da Universidade do Estado do Rio de Janeiro. Área de concentração: Processos Químicos, Petróleo e Meio Ambiente.

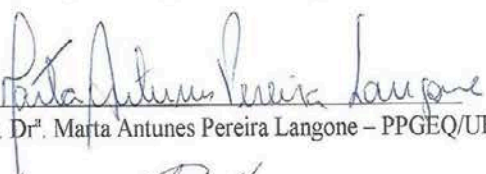
Aprovada em 8 de Janeiro 2015

Banca Examinadora:

  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Lucia Regina Raddi de Araújo – PPGEQ/UERJ

  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Cristiane Assumpção Henriques – PPGEQ/UERJ

  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Cynthia Fraga Scofield – IQ/UERJ

  
Prof.<sup>ª</sup> Dr.<sup>ª</sup> Marta Antunes Pereira Langone – PPGEQ/UERJ

  
Dr. Márcio de Figueiredo Portilho – CENPES/PETROBRAS

Rio de Janeiro  
2015

## **DEDICATÓRIA**

Este trabalho eu dedico primeiramente a Deus por tudo o quanto ele tem feito em minha vida e por ter me capacitado a conquistar mais um sonho. Aos meus pais Antônio e Fátima que sempre acreditaram nos meus sonhos me incentivando e apoiando.

## AGRADECIMENTOS

Primeiramente a Deus por tudo que tens feito e que ainda há de fazer em minha vida.

Pela alegria da sua presença em todos os momentos, mesmo quando tudo parecia difícil o Senhor me sustentou e me capacitou para terminar esse trabalho.

Por mais um conhecimento obtido estudando e aprendendo com toda a dedicação.

A minha família que sempre apoiou nas minhas decisões e o sonho de fazer o mestrado com o objetivo de trabalhar e contribuir para a vida científica.

Aos meus amigos que me compreenderam sem restrições com esta decisão tomada sempre me incentivando me dizendo palavras de conforto mesmo sem obter nenhum conhecimento sobre o assunto.

As minhas orientadoras Lúcia Regina Raddi e Cristiane Assumpção Henriques por me aceitarem para orientação, pela paciência e confiança.

Ao Programa de Pós – Graduação em Engenharia Química da Universidade do Estado do Rio de Janeiro (PPGEQ-UERJ).

Aos colaboradores que me ajudaram com as análises no laboratório me acompanhando e me explicando sobre cada equipamento : Jéssica Cruz de Luca (FT-IR), Paula Moraes Veiga (CG), Márcia Rita Maia e Wallace Magalhães Antunes (HPLC).

Pelo imenso auxílio de pessoas que me ajudaram na parte de quimiometria. Foi muito gratificante ter aprendido sobre um assunto extenso e imensamente importante para a confiabilidade dos resultados.

Ao professor Dr Aderval Luna do Programa de Pós – Graduação em Engenharia Química da Universidade do Estado do Rio de Janeiro (PPGEQ-UERJ).

Aos meus amigos Arnaldo e Igor que foram fundamentais para o meu aprendizado. Sem eles eu não conseguiria dar continuidade ao trabalho. Sou muito grata a Deus por ter conhecido vocês e muito obrigada pela força.

Ao Pesquisador do Instituto Nacional de Metrologia, Qualidade e Tecnologia (Inmetro) D.Sc .Werickson Fortunato de Carvalho Rocha que me ajudou a analisar os resultados fornecendo sugestões para a melhoria do trabalho.

Ao Professor da Universitat Rovira i Virgili, Espanha, Dr Joan Ferré que foi imprescindível com suas interpretações, sugestões e explicações sobre os resultados do trabalho e a aplicabilidade das ferramentas de quimiometria. Foi de fundamental importância o curso por ele ministrado no Inmetro que contribuiu muito para o meu aprendizado.

A Capes pela bolsa de mestrado.

Aprender é a única coisa de que a mente nunca se cansa, nunca tem medo e nunca se arrepende.

*Leonardo da Vinci*

## RESUMO

SILVA, Juliana Verdan. *Uso da calibração multivariada para a predição de propriedades físico-químicas de misturas de óleo de soja e biodiesel*. 2015. 95 f. Dissertação (Mestrado em Engenharia Química) – Instituto de Química, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2015.

O objetivo deste trabalho foi estabelecer um modelo empregando-se ferramentas de regressão multivariada para a previsão do teor em ésteres metílicos e, simultaneamente, de propriedades físico-químicas de misturas de óleo de soja e biodiesel de soja. O modelo foi proposto a partir da correlação das propriedades de interesse com os espectros de reflectância total atenuada no infravermelho médio das misturas. Para a determinação dos teores de ésteres metílicos foi utilizada a cromatografia líquida de alta eficiência (HPLC), podendo esta ser uma técnica alternativa aos métodos de referência que utilizam a cromatografia em fase gasosa (EN 14103 e EN 14105). As propriedades físico-químicas selecionadas foram índice de refração, massa específica e viscosidade. Para o estudo, foram preparadas 11 misturas com diferentes proporções de biodiesel de soja e de óleo de soja (0-100 % em massa de biodiesel de soja), em quintuplicata, totalizando 55 amostras. A região do infravermelho estudada foi a faixa de 3801 a 650  $\text{cm}^{-1}$ . Os espectros foram submetidos aos pré-tratamentos de correção de sinal multiplicativo (MSC) e, em seguida, à centralização na média (MC). As propriedades de interesse foram submetidas ao autoescalamento. Em seguida foi aplicada análise de componentes principais (PCA) com a finalidade de reduzir a dimensionalidade dos dados e detectar a presença de valores anômalos. Quando estes foram detectados, a amostra era descartada. Os dados originais foram submetidos ao algoritmo de Kennard-Stone dividindo-os em um conjunto de calibração, para a construção do modelo, e um conjunto de validação, para verificar a sua confiabilidade. Os resultados mostraram que o modelo proposto por PLS2 (Mínimos Quadrados Parciais) foi capaz de se ajustar bem os dados de índice de refração e de massa específica, podendo ser observado um comportamento aleatório dos erros, indicando a presença de homocedasticidade nos valores residuais, em outras palavras, o modelo construído apresentou uma capacidade de previsão para as propriedades de massa específica e índice de refração com 95% de confiança. A exatidão do modelo foi também avaliada através da estimativa dos parâmetros de regressão que são a inclinação e o intercepto pela Região Conjunta da Elipse de Confiança (EJCR). Os resultados confirmaram que o modelo MIR-PLS desenvolvido foi capaz de prever, simultaneamente, as propriedades índice de refração e massa específica. Para os teores de ésteres metílicos determinados por HPLC, foi também desenvolvido um modelo MIR-PLS para correlacionar estes valores com os espectros de MIR, porém a qualidade do ajuste não foi tão boa. Apesar disso, foi possível mostrar que os dados podem ser modelados e correlacionados com os espectros de infravermelho utilizando calibração multivariada.

Palavras chave: Calibração multivariada. Mistura biodiesel-óleo de soja. Espectros MIR-ATR. Massa específica. Índice de refração.



## ABSTRACT

SILVA, Juliana Verdan. *Use of multivariate calibration for prediction of physicochemical properties of soybean oil - biodiesel blends*. 2015. 95 f. Dissertação (Mestrado em Engenharia Química) – Instituto de Química, Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2015.

In the present work, a model that uses multivariate regression tools was proposed to predict both contents of methyl esters and physical-chemical properties of soybean oil - soybean biodiesel blends. The model was proposed from the correlation of the properties of interest with the attenuated total reflectance infrared spectra of the samples (ATR/MID-FTIR). The composition of the blends (methyl esters content) was determined by high performance liquid chromatography (HPLC), which can be seen as an alternative technique to the standard reference methods based on gas chromatography (EN 14103 and EN 14105). The selected physicochemical properties were refractive index, density, and viscosity. For the study, 11 mixtures were prepared with different proportions of soybean biodiesel and soybean oil (0-100 % wt of soybean biodiesel) in five replications totalizing 55 samples. The infrared spectra were acquired in the range 3801-650  $\text{cm}^{-1}$ . The spectra were submitted to the multiplicative signal correction (MSC) and then to mean centering (MC) preprocessing. The properties of interest were submitted to auto scale. It was then applied principal component analysis (PCA) for the purpose of reducing the dimensionality of the data, and detected the presence of outliers. When the outliers were detected, the samples were discarded. The original data were submitted to Kennard-Stone algorithm dividing them into a calibration set to build the model and the validation to verify its reliability. The results showed that the model proposed by PLS2 (Partial Least Squares) was able to adjust well to the refractive index data and density, can be observed a random behavior of errors, indicating the presence of homoscedasticity in residual values, in other words, the constructed model presented a forecast of capacity for specific mass properties and refractive index with 95% confidence. The accuracy of the model was evaluated by estimating the regression parameters which are the slope and the intercept by EJCR (Joint Region Confidence Ellipse). The results confirmed that MIR-developed PLS model was able to predict both the refractive index and specific gravity properties. For contents of methyl esters via liquid chromatography the model showed a lower adjustment of the data. Nevertheless, it was possible to demonstrate that the results obtained from methyl esters by HPLC analytical method can be modeled and correlate them with infrared spectra using multivariate calibration.

Keywords: Multivariate calibration. mixing soybean biodiesel-oil MIR-ATR spectra. density. refractive index.

## LISTA DE FIGURAS

Figura 1 -	Reação de transesterificação de triacilgliceróis (triglicerídeos), onde R representa a cadeia carbônica dos ácidos graxos.....	18
Figura 2 -	Capacidade instalada de produção de biodiesel no Brasil.....	20
Figura 3 -	Representação das matrizes de dados X e Y na calibração multivariada.....	29
Figura 4 -	Formação da Variável Latente.....	34
Figura 5 -	Elipse de confiança.....	37
Figura 6 -	Material utilizado no preparo das amostras.....	44
Figura 7 -	Espectrômetro de IV com acessório de Reflectância Total Atenuada (ATR).....	46
Figura 8 -	Cromatógrafo Varian CP3800.....	47
Figura 9 -	Cromatógrafo Varian 920LC.....	48
Figura 10 -	Densímetro DM40/Modulo Índice de Refração RX40 - Metter Toledo.....	49
Figura 11 -	Viscosímetro.....	50
Figura 12 -	Espectros de Infravermelho.....	56
Figura 13 -	Espectros de ATR/FTIR com 47 amostras das misturas de biodiesel e óleo de soja.....	61
Figura 14 -	Resultado da análise de PCA (rendimento de variância capturada <i>versus</i> PCs) na matriz com 46 amostras de MIR.....	63
Figura 15 -	Gráfico dos resíduos <i>versus</i> escores ( <i>Hotelling T<sup>2</sup> versus</i> Q resíduos) na matriz com 46 amostras de MIR.....	64
Figura 16 -	Gráfico dos escores (PC1 <i>versus</i> PC2) na matriz com 46 amostras de MIR.....	64
Figura 17 -	Gráfico dos resíduos ( <i>Hotelling T<sup>2</sup> versus</i> Q resíduos) para índice de refração, massa específica e teor de ésteres por HPLC para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.....	67
Figura 18 -	Curva de regressão utilizando PLS2 para os valores de índice de refração (preditos <i>versus</i> medidos) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.....	67

Figura 19 -	Curva de regressão utilizando PLS2 para os valores de massa específica (preditos <i>versus</i> medidos) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.....	68
Figura 20 -	Gráfico de resíduos utilizando PLS2 para índice de refração dos valores (preditos <i>versus</i> residuais) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.....	68
Figura 21 -	Gráfico de resíduos utilizando PLS2 para massa específica dos valores (preditos <i>versus</i> residuais) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.....	69
Figura 22 -	Curva de regressão utilizando PLS2 para o teor de ésteres metílicos determinado por HPLC dos valores (preditos <i>versus</i> medidos) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.....	69
Figura 23 -	Gráfico de resíduos utilizando PLS2 para o teor de ésteres metílicos determinado por HPLC dos valores (preditos <i>versus</i> residuais) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.....	70
Figura 24 -	Curva de regressão utilizando OLS para os valores de índice de refração ( preditos pelo método PLS2 <i>versus</i> medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.....	71
Figura 25 -	Curva de regressão utilizando OLS para os valores de massa específica (preditos pelo método PLS2 <i>versus</i> medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.....	71
Figura 26 -	Curva de regressão utilizando OLS para os valores de teor de ésteres determinados por HPLC (preditos pelo método PLS2 <i>versus</i> medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.....	72
Figura 27 -	Gráfico de erros absolutos utilizando OLS para índice de refração dos valores (residuais <i>versus</i> medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.....	73
Figura 28 -	Gráfico de erros absolutos utilizando OLS para massa específica dos valores (residuais <i>versus</i> medidos de referência) para o conjunto de	

	validação contendo 16 amostras de MIR.....	73
Figura 29 -	Gráfico de erros absolutos utilizando OLS para o teor de ésteres metílicos por HPLC dos valores (residuais <i>versus</i> medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.....	74
Figura 30 -	Gráfico da região conjunta da elipse de confiança para inclinação e intercepto da regressão de valores (preditos <i>versus</i> valores de referência) usando mínimos quadrados ordinários (OLS) para os dados de índice de refração para o conjunto de validação.....	75
Figura 31 -	Gráfico da região conjunta da elipse de confiança para inclinação e intercepto da regressão de valores (preditos <i>versus</i> valores de referência) usando mínimos quadrados ordinários (OLS) para os dados de índice de refração para o conjunto de validação.....	75
Figura 32 -	Gráfico da região conjunta da elipse de confiança para inclinação e intercepto da regressão de valores (preditos <i>versus</i> valores de referência) usando mínimos quadrados ordinários (OLS) para os dados de ésteres metílicos por HPLC para o conjunto de validação.....	76

## LISTA DE TABELAS

Tabela 1 -	Algumas especificações para o biodiesel no Brasil.....	85
Tabela 2 -	Medições volumétricas e mássicas de biodiesel e óleo de soja das misturas.....	45
Tabela 3 -	Valores de massa específica obtidos a 20°C (g/cm <sup>3</sup> ).....	51
Tabela 4 -	Valores do índice de refração obtidos a 40°C .....	52
Tabela 5 -	Teor de ésteres metílicos (rendimento simulado) por HPLC.....	53
Tabela 6 -	Teor de ésteres metílicos (rendimento simulado por CG) .....	54
Tabela 7 -	Valores das viscosidades cinemáticas à 40 °C (mm <sup>2</sup> /s).....	55
Tabela 8 -	Atribuição das bandas no espectro de infravermelho médio das amostras estudadas na região entre 3600 – 650 cm <sup>-1</sup> .....	55
Tabela 9 -	Testes Estatísticos de Grubbs e Dixon para o nível 4 (amostras com 70% de biodiesel).....	58
Tabela 10 -	Testes Estatísticos de Grubbs e Dixon para o teor de ésteres por HPLC para o nível 2 com 90% de biodiesel referente à amostra 06.....	60
Tabela 11 -	Resultados estatísticos da análise de componentes principais com 46 amostras de MIR fornecidos pelo programa PLS_Toolbox.....	62
Tabela 12	Resultados estatísticos da regressão por PLS com 46 amostras fornecidos pelo programa PLS_Toolbox para o conjunto de calibração e de validação.....	66
Tabela 13 -	Valores dos coeficientes de inclinação e intercepto para as propriedades estudadas.....	72
Tabela 14 -	Valores estimados para as Figuras de Mérito.....	91

## LISTA DE ABREVIATURAS E SIGLAS

AGL	Ácidos Graxos Livres
ANP	Agência Nacional de Petróleo, Gás Natural e Biocombustíveis
CG	Cromatografia Gasosa
EJCR	Região Conjunta da Elipse de Confiança (do inglês, <i>Elliptical Joint Confidence Region</i> )
FT	Transformada de Fourier (em inglês: <i>Fourier Transform</i> )
G	Estatística de Grubbs
H <sub>0</sub>	Hipótese Nula
H <sub>a</sub>	Hipótese Alternativa
HPLC	Cromatografia Líquida de Alta Eficiência (em inglês, <i>High Performance Liquid Chromatography</i> )
IV	Infravermelho
KS	Kennard – Stone
MC	Centralização na Média (em inglês, <i>Mean Center</i> )
MIR	Infravermelho Médio (em inglês, <i>Middle Infrared</i> )
MSC	Correção de Sinal Multiplicativo (em inglês, <i>Multiply Scatter Correction</i> )
PC	Componentes Principais (do inglês, <i>Principal Component</i> )
PCA	Análise de Componentes Principais (em inglês, <i>Principal Component Analysis</i> )
PLS	Regressão por Mínimos Quadrados Parciais (em inglês, <i>Partial Least Square</i> )
Q	Estatística de Dixon
RMSEC	Raiz quadrada do Erro Quadrático Médio de Calibração (em inglês, <i>Root Mean Square Error of Calibration</i> )
RMSECV	Raiz quadrada do Erro Quadrático Médio de Validação Cruzada (em inglês, <i>Root Mean Square Error of Cross-Validation</i> )
RMSEP	Raiz quadrada do Erro Quadrático Médio de Predição (em inglês, <i>Root Mean Square Error of Prediction</i> )
VL	Variáveis Latentes
W	Estatística de Shapiro Wilk

## SUMÁRIO

	<b>INTRODUÇÃO</b> .....	15
	<b>OBJETIVOS</b> .....	17
1.	<b>REVISÃO BIBLIOGRÁFICA</b> .....	18
1.1	<b>Biodiesel</b> .....	18
1.2	<b>Panoramas Mundial e Nacional</b> .....	19
1.3	<b>Caracterização do biodiesel</b> .....	20
1.3.1	<u>Massa Específica</u> .....	20
1.3.2	<u>Viscosidade cinemática a 40°C</u> .....	21
1.3.3	<u>Teor de ésteres</u> .....	22
1.3.4	<u>Índice de refração</u> .....	22
1.3.5	<u>Métodos analíticos para o biodiesel</u> .....	22
1.4	<b>Quimiometria</b> .....	24
1.4.1	<u>Testes de hipóteses nas propriedades do biodiesel</u> .....	24
1.4.2	<u>Nível de significância e p-valor</u> .....	25
1.4.3	<u>Testes de outliers</u> .....	26
1.4.3.1	Teste de Grubbs.....	26
1.4.3.2	Teste de Dixon.....	27
1.4.4	<u>Teste de normalidade de Shapiro-Wilk</u> .....	27
1.4.5	<u>Algoritmo Kennard – Stone</u> .....	28
1.4.6	<u>Calibração Multivariada</u> .....	29
1.4.7	<u>Métodos de pré-processamento</u> .....	30
1.4.7.1	Correção de sinal multiplicativo (MSC).....	31
1.4.7.2	Centralização na Média (MC).....	31
1.4.7.3	Autoescalamento.....	32
1.4.8	<u>Método de classificação</u> .....	32
1.4.8.1	Análise de componentes principais (PCA).....	32
1.4.9	<u>Métodos de calibração</u> .....	33
1.4.9.1	Regressão por mínimos quadrados parciais (PLS).....	33
1.4.10	<u>Elipse de Confiança</u> .....	35
1.5	<b>Uso de ferramentas quimiométricas e da espectrometria no infravermelho para a predição de propriedades</b> .....	37

2	<b>METODOLOGIA</b> .....	44
2.1	<b>Preparo das misturas de biodiesel de soja e óleo de soja</b> .....	44
2.2	<b>Aquisição dos espectros de MIR</b> .....	46
2.3	<b>Quantificação dos ésteres por cromatografia em fase gasosa (CG) e cromatografia líquida de alta eficiência (HPLC)</b> .....	46
2.4	<b>Determinação da massa específica e do índice de refração</b> .....	48
2.5	<b>Determinação da viscosidade</b> .....	49
2.6	<b>Softwares utilizados no trabalho</b> .....	50
3	<b>RESULTADOS E DISCUSSÃO</b> .....	51
3.1	<b>Medições das propriedades físico-químicas</b> .....	51
3.2	<b>Espectroscopia no infravermelho: interpretação dos espectros</b> .....	55
3.3	<b>Testes de <i>outliers</i> e normalidade</b> .....	57
3.4	<b>Pré-processamento</b> .....	61
3.5	<b>Análise de Componentes Principais (PCA)</b> .....	62
3.6	<b>Regressão por mínimos quadrados parciais (PLS)</b> .....	65
3.7	<b>Regressão por mínimos quadrados ordinários (OLS)</b> .....	70
	<b>CONCLUSÃO</b> .....	77
	<b>SUGESTÕES</b> .....	78
	<b>REFERÊNCIAS</b> .....	79
	<b>APÊNDICE A - Especificações para o biodiesel no Brasil</b> .....	85
	<b>APÊNDICE B - Figuras de mérito</b> .....	88
	<b>APÊNDICE C - Cromatograma por CG</b> .....	92
	<b>APÊNDICE D - Cálculo para quantificar o teor de ésteres metílicos por CG</b> ....	93
	<b>APÊNDICE E - Cromatograma por HPLC</b> .....	94
	<b>APÊNDICE F – Cálculo para quantificar o teor de ésteres metílicos por HPLC</b> .....	95



## INTRODUÇÃO

A crescente preocupação com a preservação do meio ambiente, aliada à instabilidade nos preços do petróleo e às incertezas quanto ao fornecimento do mesmo em função da instabilidade política das regiões produtoras, tem impulsionado a indústria a desenvolver combustíveis alternativos a partir de recursos renováveis e processos ambientalmente não agressivos. O biodiesel, uma mistura de ésteres de ácidos graxos obtidos a partir de óleos vegetais ou gorduras animais, é um combustível alternativo interessante, devido às suas propriedades similares às do diesel obtido a partir do petróleo.

Os primeiros motores diesel inventados por Rudolf Diesel eram movidos a óleos vegetais, porém, com o advento de óleos de origem fóssil obtidos a menor custo, este tipo de biocombustível não se mostrou, na época, economicamente viável. Desde 1930, os motores diesel foram aprimorados para o uso da fração diesel obtida a partir do petróleo, que consiste basicamente de hidrocarbonetos saturados. Por essa razão, os motores diesel modernos não funcionam adequadamente usando óleos vegetais puros como combustíveis, devido aos problemas de viscosidade elevada, formação de depósitos nos sistemas de injeção e propriedades pouco satisfatórias para partida a frio (LANCASTER, 2002).

Com o aumento do interesse internacional por combustíveis renováveis, algumas soluções foram investigadas para resolver o problema das propriedades dos óleos vegetais de modo a permitir seu uso como combustível renovável (KNOTHE et. al., 2006). A solução que se mostrou mais promissora foi a de submeter os óleos vegetais ou as gorduras ao processo de transesterificação com alcoóis de cadeia curta, pois o produto deste processo, uma mistura de ésteres alquílicos de ácidos graxos, chamada de Biodiesel, pode ser adicionado ao diesel convencional ou até substituí-lo por completo, conforme tem sido testado ultimamente<sup>1</sup>.

Atualmente, a produção comercial de biodiesel ocorre principalmente pela reação de transesterificação de óleos vegetais ou gorduras animais com álcoois de cadeia curta (geralmente metanol), na presença de catalisadores básicos (usualmente KOH, NaOH, metóxido de sódio ou potássio), que são solúveis no álcool utilizado, para produzir ésteres alquílicos de ácidos graxos (KNOTHE, 2001; KNOTHE et. al., 2006). Durante o processo de transesterificação, além dos ésteres alquílicos também são formados intermediários como mono e diacilgliceróis. Além destes gliceróis parcialmente reagidos, podem-se ter triacilgliceróis não reagidos, bem como glicerol, ácidos graxos livres (AGL), álcool e resíduos

---

<sup>1</sup> Disponível em < <http://www.biodieselbr.com/biodiesel/definicao/o-que-e-biodiesel.htm> > Acessado em 3 de Maio de 2014.

,do catalisador, os quais podem contaminar o produto final. A presença destes contaminantes no biodiesel pode levar a problemas operacionais severos no motor, além da formação de depósitos no filtro e deterioração do combustível (KNOTHE et. al., 2006).

Embora já existam processos comerciais implantados para produção de biodiesel via catálise básica em meio homogêneo ou heterogêneo, muitos são os desafios ainda existentes e que precisam ser enfrentados pela comunidade científica. Dentre estes destacam-se a necessidade de utilização de matéria prima de menor custo e que não seja utilizada na alimentação humana e a substituição dos catalisadores homogêneos pelos heterogêneos, que oferecem inúmeras vantagens sobre os homogêneos clássicos, tais como pouca ou nenhuma corrosão; fácil separação; poucos problemas com rejeitos; fácil manuseio e possibilidade de reutilização. Além disso, esta classe de catalisadores pode ser usada em reatores de leito fixo, possibilitando a produção do biodiesel em sistemas contínuos. O profundo conhecimento do efeito dessas modificações nos processos de produção sobre a qualidade do biodiesel final é muito importante para garantir a obtenção de um produto dentro das especificações exigidas por lei.

A qualidade do biodiesel é usualmente determinada através de vários métodos analíticos que têm sido propostos para estabelecer o teor de óleo vegetal em biodiesel, determinar a presença de substâncias indesejáveis como glicerol (livre e total), mono, di e triglicerídeos, além de determinar uma série de propriedades físico-químicas (BATTISTI, 2009).

No Brasil, a normalização dos padrões para o biodiesel é estabelecida por um conjunto de especificações reguladas pela ANP (Agência Nacional de Petróleo, Gás Natural e Biocombustíveis) que garantem a sua qualidade. A técnica analítica recomendada pela resolução ANP nº 14 de 2012 para a determinação do teor de ésteres metílicos no biodiesel é a cromatografia em fase gasosa (CG) sem derivatização da amostra (padronizada pela norma EN14103). Já a cromatografia em fase gasosa (CG) com derivatização da amostra (padronizada pela norma EN14105) é utilizada para a determinação de glicerol, mono-, di- e triacilglicerídeos.

No método de CG derivativo as amostras passam por uma etapa de tratamento que é feita com *N*, *O*-bis (trimetilsilil) trifluoracetamida (BSTFA) para converter grupamentos hidroxílicos nos trimetilsilil-derivados (TMS) utilizados para a quantificação de ésteres totais, bem como a determinação simultânea de glicerol juntamente com mono-, di- e triacilgliceróis, por isso, o procedimento derivativo (EN14105) é demorado e dispendioso (VALDEZ et. al., 2012; KHOTHE et. al., 2006).

A cromatografia líquida de alta eficiência (HPLC) também permite a determinação dos produtos intermediários parcialmente transesterificados, além dos ésteres metílicos de ácidos graxos e tem como vantagem a ausência da etapa de tratamento, o que pode reduzir o tempo e os custos analíticos. No entanto, ainda não existe uma norma oficial recomendada pela resolução ANP nº 14 de 2012 que aplique a HPLC para a análise de biodiesel, o que justifica um estudo mais aprofundado sobre este assunto.

Nesse contexto, o desenvolvimento de metodologias analíticas baseadas em técnicas que permitam determinar a composição das amostras de biodiesel de um modo rápido e confiável tem grande interesse, não apenas para a certificação do produto como também para o acompanhamento das reações de síntese de biodiesel. Nesse último caso, a avaliação do desempenho de novos sistemas catalíticos e da utilização de matérias primas (oleaginosas) alternativas seria facilitada pela redução do tempo e dos custos analíticos.

## **OBJETIVOS**

Este trabalho teve como objetivo avaliar a aplicação de uma metodologia baseada na técnica de espectroscopia no infravermelho médio (MIR), combinada com a calibração multivariada, visando a determinação do teor de ésteres metílicos em misturas biodiesel/óleo de soja bem como das propriedades físicas: massa específica, índice de refração e viscosidade dessas misturas. Espera-se, assim, disponibilizar uma metodologia para monitoramento da qualidade do biodiesel através de uma determinação rápida, fácil, segura e não destrutiva. Tal informação deverá ser conseguida através da correlação dos espectros de infravermelho com os correspondentes valores das propriedades de interesse e que tem sido empregado amplamente na literatura (LIRA et. al., 2010; PIMENTEL et. al., 2006; FERREIRA et. al., 1999; VASCONCELOS, 2012; ROCHA; NOGUEIRA; VAZ, 2012).

Neste trabalho, as propriedades físico-químicas monitoradas foram a massa específica, a viscosidade cinemática e o índice de refração. A composição das misturas modelo de biodiesel de soja e óleo de soja, expressa em termos de % em massa, foi determinada pela técnica de HPLC, utilizada como alternativa à cromatografia em fase gasosa com derivatização da amostra (norma EN 14105). Os teores de ésteres metílicos medidos por HPLC foram comparados com os valores obtidos pelo método de referência estabelecido pela norma

EN 14103.

## 1. REVISÃO BIBLIOGRÁFICA

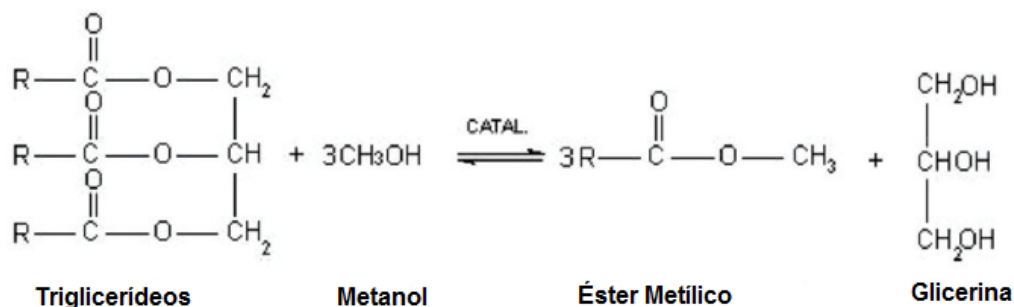
### 1.1. Biodiesel

O alto consumo de produtos derivados do petróleo, como a gasolina e o diesel, e a necessidade em reduzir o aquecimento global causado pelas emissões gasosas têm atraído a atenção de mercados do mundo inteiro para a utilização de combustíveis alternativos que sejam renováveis e que resultem em um menor impacto ambiental. A energia proveniente da biomassa é apontada como uma grande opção de fonte energética e o biodiesel já se destacam como uma alternativa viável (MARQUES et. al., 2010).

O biodiesel é definido, segundo a ANP, como um combustível composto de alquil-ésteres de ácidos carboxílicos de cadeia longa, produzido a partir da transesterificação e ou/esterificação de matérias graxas, de gorduras de origem vegetal ou animal, e que atenda às especificações contidas no Regulamento Técnico ANP nº 14/2012.

O processo global da reação de transesterificação, como é mostrado na Figura 1, é uma sequência de três reações reversíveis e consecutivas, em que os triglicerídeos são primeiramente transformados em diglicerídeos, em seguida são transformados em monoglicerídeos, os quais são finalmente convertidos nos ésteres de ácidos graxos, denominados Biodiesel (MARQUES et. al., 2011; KNOTHE, 2001; KNOTHE et. al., 2006).

Figura 1 - Reação de transesterificação de triacilgliceróis (triglicerídeos), onde R representa a cadeia carbônica dos ácidos graxos.



Fonte: MARQUES, 2011 adaptada.

## 1.2. Panoramas Mundial e Nacional

Apesar de existirem várias fontes de energia alternativas, os combustíveis fósseis ainda constituem a principal parcela do consumo de energia no mundo e a previsão é de aumento de 1,7% ao ano, de 2000 a 2030, alcançando 15,3 bilhões de toneladas equivalentes de petróleo. A matriz energética mundial tem participação total de 80% de fontes de carbono fóssil, sendo 36% de petróleo, 23% de carvão e 21% de gás natural. O Brasil se destaca entre as economias industrializadas pela elevada participação das fontes renováveis em sua matriz energética. Isso se explica por alguns privilégios da natureza, como uma bacia hidrográfica contando com vários rios de planalto, fundamental à produção de eletricidade (14%), e o fato de ser o maior país tropical do mundo, um diferencial positivo para a produção de energia de biomassa (23%)<sup>2</sup>.

A elevada demanda global de energia não renovável poderá aumentar o uso de recursos energéticos fósseis, cujos preços sofrem com a instabilidade política nas áreas produtoras.

Uma das opções promissoras de energia renovável que tem atraído muita atenção é o biodiesel, que apresenta praticamente as mesmas propriedades que as do diesel derivado do petróleo (ASHNANI et. al. 2014).

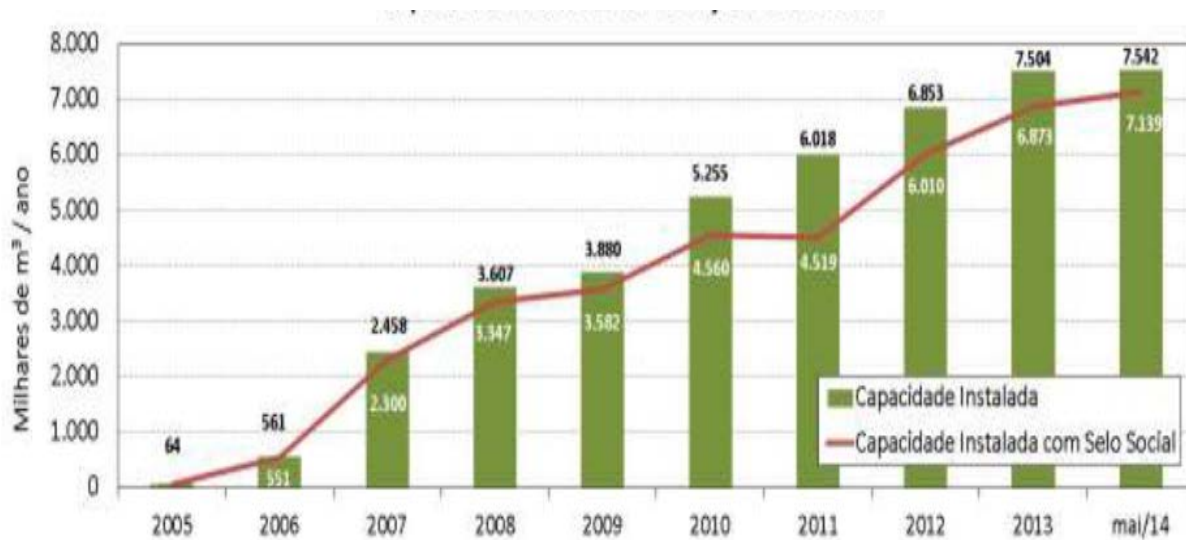
A produção de biodiesel atingiu 17,6 bilhões de litros ao longo dos últimos 20 anos e está previsto que este aumento seja muito mais rápido nas próximas décadas crescendo até 42 bilhões de litros. A União Européia, Argentina, Brasil, Malásia e EUA são responsáveis por cerca de 93% da produção de biodiesel no mundo (AYTAV; KOCAR, 2013).

De acordo com os dados fornecidos pela ANP, a produção de biodiesel em maio de 2014 foi de 243 mil m<sup>3</sup>. No acumulado do ano, a produção atingiu 1.253 mil m<sup>3</sup>, um acréscimo de 7,7% em relação ao mesmo período de 2013 (1.163 mil m<sup>3</sup>). A capacidade instalada, ou seja, autorizada para a comercialização, em maio de 2014, ficou em 7.542 mil m<sup>3</sup>/ano (629 mil m<sup>3</sup>/mês). Dessa capacidade, 95% são referentes às empresas detentoras do Selo Combustível de acordo com o Ministério de Minas e Energia - Boletim Mensal dos Combustíveis Renováveis (MME, 2014), como ilustrado na Figura 2.

---

<sup>2</sup> Disponível em <<http://www.biodieselbr.com/energia/alternativa/agro-energia.htm>> Acessado em 30 de Maio de 2014.

Figura 2 - Capacidade instalada de produção de biodiesel no Brasil.



Fonte: Ministério de Minas e Energia - Boletim Mensal dos Combustíveis Renováveis, 06/2014.

### 1.3. Caracterização do biodiesel

Para a certificação da qualidade do biodiesel, a amostra deve atender a limites estabelecidos para 23 propriedades, por exemplo, acidez, massa específica, viscosidade, índice de saponificação. Devido a não existência de restrições em relação à matéria-prima graxa ou ao álcool utilizado na sua produção, algumas propriedades não possuem valor especificado como, por exemplo, o índice de iodo e o número de cetano que precisam somente ter seus valores anotados. Por outro lado, outras propriedades não especificadas no Regulamento Técnico ANP nº 14/2012 também podem fornecer informações importantes que possibilitem uma avaliação criteriosa sobre as matérias-primas mais adequadas à produção do biodiesel.

A Tabela 1 (no APÊNDICE A) apresenta as especificações do biodiesel que foram estudadas neste trabalho.

#### 1.3.1 Massa específica

A massa específica relaciona a massa presente por unidade de volume de combustível e a finalidade desta especificação é a de verificar a presença de impurezas no combustível. Para o biodiesel, a massa específica é uma propriedade fluidodinâmica que está relacionada

com o funcionamento de motores de injeção por compressão (motores diesel). Tal propriedade tem uma grande influência na circulação e injeção do combustível (ALMEIDA, 2011). Como a bomba injetora alimenta volumes constantes de combustível para o motor, variando-se a massa específica do biodiesel, a massa de combustível injetada no motor variará, podendo afetar o desempenho, consumo, potência e aquecimento de válvulas. O aumento da massa específica provoca o aumento das emissões de monóxido de carbono, de hidrocarbonetos e de material particulado (KNOTHE, 2005).

A composição química da matéria-prima que constitui o biodiesel é responsável pela sua massa específica. Quanto mais longa for a cadeia carbônica do alquiléster, maior será sua massa, entretanto, a presença de insaturações na molécula tende a diminuir o valor da massa específica. A massa específica do biodiesel é também influenciada pela presença de impurezas como, por exemplo, subprodutos do processo ou substâncias adulterantes (LÔBO; FERREIRA; CRUZ, 2009).

Segundo a Resolução ANP nº 14 de 2012, para a certificação da qualidade do biodiesel, este deve ter uma massa específica, a 20 °C, na faixa entre 850 e 900 kg/m<sup>3</sup>.

### 1.3.2 Viscosidade cinemática a 40°C

A viscosidade cinemática é a medida de resistência ao fluxo de um fluido sob a ação da gravidade, sendo definida como o quociente entre a viscosidade dinâmica do fluido pela sua massa específica. A viscosidade é um requisito fundamental para o desempenho da combustão no motor. Viscosidades elevadas provocam a má nebulização, a formação de depósitos e o aumento da quantidade de emissões. Viscosidades baixas provocam a má lubrificação ocasionando danos ao sistema de injeção. Pode ser responsável também por perda de potência do motor devido a fugas de combustível ocasionadas por uma baixa viscosidade (FARAH, 2006). A viscosidade também pode ser usada para avaliar a evolução do nível de oxidação do combustível, pois os mecanismos das reações de oxidação geram polímeros que aumentam a viscosidade do produto (WAYNICK, 2005).

A viscosidade do biodiesel é influenciada por fatores como o comprimento da cadeia, posição, número e natureza das duplas ligações da matéria-prima usada para sua produção (KNOTHE; STEIDLEY, 2005).

De acordo com a Resolução ANP nº 14 de 2012, para a certificação da qualidade do biodiesel, este deve ter uma viscosidade cinemática, a 40°C, na faixa entre 3,0 e 6,0 mm<sup>2</sup>/s.

### 1.3.3 Teor de ésteres

Tem por finalidade especificar a quantidade mínima de ésteres (metílicos ou etílicos) de ácidos graxos presentes no biodiesel B100. Quanto maior o teor de ésteres, maior a pureza do biodiesel, garantindo que houve um teor satisfatório da reação, indicando também menor presença de contaminantes no produto (KNOTHE; GERPEN; KRAHL, 2005).

A quantidade de ésteres metílicos de ácido linolênico é controlada pela norma européia EN 15215 devido à maior propensão à oxidação desse éster, por possuir três insaturações em sua estrutura (KNOTHE; GERPEN; KRAHL, 2005).

No presente trabalho, o teor de ésteres foi utilizado para simular a conversão do óleo de soja (constituído por triglicerídeos) em biodiesel (mistura de ésteres metílicos), determinando-se o teor total de ésteres metílicos contidos nas misturas preparadas de biodiesel de soja e óleo de soja.

Segundo a Resolução ANP nº 14 de 2012, para a certificação da qualidade do biodiesel, este deve ter um teor mínimo de 96,5 % massa de ésteres.

### 1.3.4 Índice de refração

O índice de refração é um parâmetro físico importante para garantir a qualidade dos óleos, sendo uma relação entre a velocidade da luz no vácuo e no meio de interesse (ALMEIDA et. al., 2011).

O índice de refração não consta no conjunto de especificações reguladas pela ANP, porém esta propriedade possui íntima relação com o grau de saturação das ligações, afetando outras propriedades importantes do biodiesel, como o índice de iodo e a estabilidade oxidativa (LÔBO; FERREIRA; CRUZ, 2009).

O índice de refração tende a aumentar com o grau de insaturação dos ácidos graxos constituintes dos triglicerídeos e pode ser afetado por outros parâmetros, tais como, teor de ácidos graxos livres, oxidação e tratamentos térmicos (ALMEIDA et. al., 2011).

### 1.3.5 Métodos analíticos para o biodiesel

Durante o processo de transesterificação, gliceróis intermediários como mono- e diacilgliceróis são formados, sendo que uma pequena quantidade destes pode permanecer retida no produto final (ésteres metílicos ou quaisquer outros alquil-ésteres). Além destes,



triacilglicerídeos não reagidos bem como glicerol, ácidos graxos livres (AGL), álcool e catalisador residual podem contaminar o produto final. Por isso, há exigências das especificações para determinar a qualidade do combustível sendo de grande importância para a comercialização do biodiesel (KNOTHE et. al., 2006).

Diversos trabalhos informam que os métodos cromatográficos são essenciais para a análise da qualidade do biodiesel (SHAH; SAHARMA; GUPTA, 2004; STAVARACHE et. al., 2005; CUNHA; OLIVEIRA, 2006; MADRAS; KOLLUR; KUMAR, 2004; ZAGONEL; ZAMORA; RAMOS, 2004; SANDRA et. al., 2002).

Os métodos de cromatografia gasosa recomendados pela ANP nº 14 de 2012 são padronizados pelas normas EN14103 (determinação de ésteres) e EN14105 (determinação de glicerol livre e total e dos teores de mono-, di- e triacilgliceróis). O primeiro trabalho sobre o uso de CG derivativa capilar discorreu sobre a quantificação de ésteres totais, bem como de mono-, di- e triacilgliceróis. As amostras foram tratadas com *N, O-bis* (trimetilsilil) trifluoracetamida (BSTFA) para converter grupamentos hidroxílicos nos trimetilsilil-derivados (TMS). Este tipo de transformação levou aos métodos padronizados ASTM D6584 e EN14105 utilizados para a quantificação simultânea de glicerol juntamente com mono-, di- e triacilgliceróis (VALDEZ et. al., 2012; KNOTHE, 2006).

Uma técnica alternativa, como a cromatografia líquida de alta eficiência (HPLC), pode apresentar uma viabilidade de uso na determinação quantitativa de compostos parcialmente transesterificados, além dos teores de ésteres metílicos (biodiesel) em forma de porcentagem. A vantagem do método de HPLC está relacionada ao fato de que o procedimento de derivatização é muito dispendioso e demorado, uma vez que precisa ser feito um tratamento nas amostras com o reagente *N, O-bis* (trimetilsilil) trifluoracetamida (BSTFA) para converter grupamentos hidroxílicos nos trimetilsilil-derivados (TMS). Sendo assim a HPLC pode ser utilizada como um método alternativo, reduzindo o tempo de análise e custos em comparação ao método de CG derivativo (BATTISTI et. al., 2009; KNOTHE, 2006).

Já o método padronizado como EN14103 por cromatografia gasosa não derivativa é utilizado para a identificação apenas dos ésteres metílicos segundo a resolução ANP nº 14 de 2012. Neste trabalho, os resultados obtidos por este método de CG não derivativo foram utilizados para validar o modelo proposto com os resultados obtidos por HPLC.

## 1.4 Quimiometria

A quimiometria é uma área da química que utiliza métodos matemáticos, estatísticos e computacionais para projetar, selecionar e aperfeiçoar procedimentos de medições e experimentos e obter o máximo de informação sobre o sistema (NETO; PIMENTEL; ARAÚJO, 2002).

Entre as diversas subáreas da quimiometria, destaca-se a calibração multivariada, que busca estabelecer um modelo que relacione um conjunto de medições (ex: IV, UV/Vis, RMN) com uma determinada propriedade (ex: concentração, massa específica, etc.) (SOUZA; POPPI, 2012).

A definição sobre calibração multivariada será explicada na seção 1.4.6, pois as próximas seções (1.4.2 a 1.4.4) abordam os testes de hipóteses para o tratamento dos dados de entrada. A realização desses testes preliminarmente à calibração dos dados é uma prática em quimiometria para a determinação de possíveis anomalias e verificação da normalidade da distribuição. A seção 1.4.5 também aborda uma técnica muito utilizada em quimiometria antes da calibração dos dados, o algoritmo de Kennard-Stone, o qual tem a finalidade de validar o modelo proposto.

### 1.4.1 Testes de hipóteses nas propriedades do biodiesel

Quando se trabalha com medições de parâmetros e se pretende validá-los, a análise estatística dos resultados obtidos é uma ferramenta importantíssima para esta validação (NORMANDO; TJÄDERHANE; QUINTÃO, 2010).

Quando se efetuam repetidas medições de determinada grandeza em um equipamento calibrado, não se consegue chegar ao mesmo resultado todas às vezes, obtendo-se, ao contrário, um conjunto de valores que oscilam, de modo aproximadamente simétrico, em torno do verdadeiro valor (CORREA, 2003). Na maioria dos trabalhos, a pressuposição estatística refere-se a premissas como linearidade nos parâmetros, normalização dos dados (ausência de *outliers*), ausência de autocorrelação e heterocedasticidade, estes dois últimos nos dados residuais. Todas essas premissas citadas são validadas quando o conjunto de dados assume, geralmente, uma distribuição normal (PONTES, 2005; GUJARATI, 2006).

Os modelos de regressão linear pressupõem que os valores residuais (diferença entre os reais valores observados e os valores ajustados pelo modelo) assumem a mesma distribuição normal (KORZENOWSKI, 2009).

Em outras palavras, quando se confirma a correlação linear entre duas variáveis X e Y pode se tentar prever o comportamento de uma delas em função da variação da outra. No caso, os valores de Y são aleatórios, pois dependem não apenas de X, mas de outras variáveis que não estão representadas no modelo. Estas variáveis são consideradas no modelo através de um termo aleatório denominado “erro - U”(FREITAS, 2003).

Desta forma pode-se considerar que o modelo relacionando as variáveis X e Y é representado por uma equação do tipo:

$$Y = \hat{a} + \hat{b}X + U \quad (\text{Equação 1})$$

De forma ideal, o termo “U” deve ser pequeno e independente de X, ou seja, U é uma variável aleatória com distribuição supostamente normal, com média zero e variância constante, de modo que se possa modificar X, sem modificar “U” e determinar o que acontecerá em Y.

Com isso, a hipótese da variância dos erros ser a mesma para todos os valores de X, denomina-se homocedasticidade que significa “igualmente disperso” (FREITAS, 2003; GUJARATI, 2006).

Para verificar esta distribuição supostamente normal, são realizados testes de hipóteses, os quais informam que a hipótese nula  $H_0$  é a alegação inicialmente assumida como verdadeira para o valor de determinado parâmetro. A hipótese alternativa  $H_1$  é a afirmação contraditória a  $H_0$  (SCUDINO, 2008). A rejeição da hipótese nula  $H_0$  ocorrerá somente se a evidência da amostra sugerir que ela seja falsa. Se a amostra não contradiz a alegação da  $H_0$ , continua-se acreditando na verdade da hipótese nula, ou seja, aceita-se a  $H_0$  (SCUDINO, 2008).

#### 1.4.2 Nível de significância e p-valor

O teste de hipótese depende fundamentalmente do nível de significância ( $\alpha$ ) com o qual se quer trabalhar (GUJARATI, 2006). A escolha do nível de significância varia em função das exigências da área em que os testes estatísticos são aplicados, mas na prática científica e na engenharia em geral, o uso do valor de  $\alpha = 0,05$  é usado na maioria das situações a menos que haja alguma informação ou exigência que indique que essa não é uma escolha apropriada (MONTGOMERY; RUNGER, 2012).

Se em uma determinada análise estatística for usado um nível de significância  $\alpha = 0,05$ , isto significa que existem 5 possibilidades em 100 da hipótese  $H_0$  ser rejeitada. Do mesmo modo pode-se dizer que existe uma confiança de 95 % de que se tome uma decisão correta, ou seja, aceitar-se a  $H_0$ .

Na resposta dos testes de hipóteses é possível verificar a probabilidade efetiva para o valor calculado, denominada p-valor, comparando-o com o nível de significância previamente escolhido. O p-valor é a menor probabilidade a qual a hipótese nula pode ser rejeitada (GUJARATI, 2006). Portanto, quando o p-valor é menor do que  $\alpha$ , implica na rejeição da hipótese  $H_0$  para aquele valor de  $\alpha$ , e o valor p achado é dito significativo. Ou, se p-valor for maior do que  $\alpha$ , implica na aceitação da  $H_0$  para aquele valor de  $\alpha$ , e o valor p é dito não significativo. Sendo assim, em estudos de testes de hipóteses, os resultados deverão fazer referência ao nível de significância e aos p-valores utilizados (SCUDINO, 2008).

#### 1.4.3 Testes de outliers

Neste trabalho, foram aplicados nos dados de propriedades (dados iniciais) os testes para identificação de valores anômalos (*outliers*), sendo eles os testes de Grubbs (estatística G) e Dixon (estatística Q) e, posteriormente, o teste de normalidade para verificar se os valores medidos seguiam uma distribuição normal como pressuposto pelos modelos de regressão linear, utilizando o teste de Shapiro-Wilk (estatística W). Quando um conjunto de resultados apresentarem um ou possivelmente mais valores que diferem significativamente dos demais, tais medições são chamadas de *outliers*. No entanto, mesmo que esses valores pareçam ser discrepantes, eles devem permanecer no conjunto e por meio de testes estatísticos verifica-se se eles devem ou não ser rejeitados, pois os valores finais da média e desvio-padrão dependerão da rejeição ou não desses valores atípicos (MILLER; MILLER, 2005).

##### 1.4.3.1 Teste de Grubbs

O teste de Grubbs, também chamado de teste G, é o mais comumente usado para a verificação de *outliers* e recomendado pela *International Union of Pure and Applied Chemistry* (IUPAC, 1988). O teste compara o desvio do valor suspeito a partir da média e do desvio-padrão das amostras e este valor suspeito é aquele mais distante da média.

Através do teste de Grubbs (Equação 2) é possível testar a  $H_0$  (que supõe que não há *outliers* nos resultados).

$$G_{\text{calc}} = \frac{\text{Valor}_{\text{suspeito}} - \bar{x}_i}{s_i} \quad (\text{Equação 2})$$

onde,  $G_{\text{calc}}$  é o valor calculado pelo teste,  $\bar{x}_i$  é a média de cada nível (conjunto de replicatas) e  $s_i$  é o desvio-padrão de cada nível.

Para  $p = 0,05$ , os valores críticos para G são dados em tabelas estatísticas adequadas. Se o valor calculado G exceder ao valor crítico, este valor suspeito é rejeitado (ou seja, há *outliers* nos resultados). Neste caso, a  $H_0$  é rejeitada, ou seja, aceita-se a hipótese  $H_1$  que diz que há diferença entre os valores observados e conhecidos (MILLER; MILLER, 2005).

#### 1.4.3.2 Teste de Dixon

O teste de Dixon, também chamado teste Q, é mais um teste usado na identificação de *outliers* devido ao seu cálculo simples e quando se tem um número pequeno de amostras, entre 3-25 (ANDRIOTTI, 2005). O teste de Dixon apresenta três procedimentos de cálculos estatísticos dependendo do tamanho da amostragem (entre 3-7, 8-11 e 12-25) (ANDRIOTTI, 2005). O teste avalia uma medição suspeita comparando-a com a diferença de uma medição mais próxima entre as outras medições (MILLER; MILLER, 2005).

O teste de estatística Q é calculado através da (Equação 3).

$$Q_{\text{calculado}} = \frac{\text{Valor}_{\text{suspeito}} - \text{Valor}_{\text{mais próximo}}}{\text{Valor}_{\text{maior}} - \text{Valor}_{\text{menor}}} \quad (\text{Equação 3})$$

Este teste assume que a população tem distribuição normal. Para  $p = 0,05$ , os valores críticos de Q são dados em tabelas adequadas. Se o valor de Q calculado exceder o valor crítico, o valor suspeito é rejeitado (MILLER; MILLER, 2005).

#### 1.4.4 Teste de normalidade de Shapiro-Wilk

Existem diversos testes estatísticos para se avaliar a normalidade de um conjunto de dados e um dos mais usados é o teste de Shapiro-Wilk que se baseia na proporção de duas

hipóteses. O teste de Shapiro-Wilk é calculado pela estatística  $W$  da seguinte forma (SCUDINO, 2008; KIM, 2011).

$$W = \frac{b^2}{\sum_{i=1}^n (x_{(i)} - \bar{x})^2} \quad (\text{Equação 4})$$

em que  $x_i$  são os valores das amostras ordenados ( $x_1$  é o menor valor); a constante  $b$  é determinada da seguinte forma:

$$b = \sum_{i=1}^{n/2} a_{n-i+1} \cdot (x_{(n-i+1)} - x_{(i)}) \text{ se } n \text{ é par}$$

$$\sum_{i=1}^{(n+1)/2} a_{n-i+1} \cdot (x_{(n-i+1)} - x_{(i)}) \text{ se } n \text{ é ímpar}$$

em que  $a_{n-i+1}$  são as constantes geradas pelas médias, variâncias e covariâncias da ordem estatística de uma amostra de tamanho  $n$  e uma distribuição normal.

Para realizar o teste de Shapiro Wilk (FREITAS, 2003) deve-se formular as hipóteses da seguinte maneira:

$H_0$ : A amostra provém de uma distribuição normal;

$H_1$ : A amostra não provém de uma distribuição normal.

#### 1.4.5 Algoritmo Kennard - Stone (KS)

Na modelagem multivariada, geralmente é feita a divisão das amostras originais em dois conjuntos, um para a calibração e o outro para a validação do modelo gerado através do algoritmo Kennard-Stone, pois é um algoritmo bem estabelecido na literatura (KNUZ, et. al., 2010; ANDRIES, et. al., 2009; VALDERRAMA, 2005).

O algoritmo de Kennard-Stone está baseado na distância euclidiana e é capaz de selecionar as amostras de maneira uniforme a partir das informações de toda área espectral (ANDRIES et. al., 2009).

Primeiramente, o algoritmo seleciona um par de amostras dentro do conjunto original que apresente uma maior distância euclidiana. A partir daí, é calculada a menor distância das amostras restantes em relação às amostras selecionadas; em seguida, o algoritmo informa que a amostra selecionada para o conjunto de calibração é aquela que apresentou a maior distância

calculada e assim sucessivamente até atingir o número de amostras desejadas (GROOT et. al., 1999).

#### 1.4.6 Calibração Multivariada

Os avanços tecnológicos dos instrumentos analíticos levaram ao desenvolvimento de novos procedimentos para analisar um determinado conjunto de dados gerados, de modo a possibilitar a determinação de propriedades de interesse dentro desse sistema relativamente complexo.

O termo calibração descreve a construção de um modelo que relacione medições instrumentais e valores da propriedade de interesse realizados com padrões, através de uma série de operações matemáticas (VALDERRAMA; BRAGA; POPPI, 2009).

O objetivo da construção de um modelo é prever propriedades desconhecidas a partir de respostas obtidas de novas amostras. De um modo geral, o processo de calibração é dividido em duas etapas:

- modelagem, que estabelece uma relação matemática entre X e Y no conjunto de calibração.
- validação, que otimiza a relação para uma melhor descrição do analito de interesse (TEÓFILO, 2013; FERREIRA et. al., 1999).

A transformação matemática e, posteriormente, a criação do modelo são organizadas da seguinte maneira: os dados espectrais (variáveis independentes, sendo uma para cada amostra) são indicados por uma matriz X de dimensão (n x m), onde cada linha representa uma amostra  $\underline{n}$  e cada coluna as variáveis  $\underline{m}$  (comprimento ou número de onda), como visto na Figura 3. O vetor Y, de dimensão n, representa o parâmetro de interesse (FERREIRA et. al., 1999).

Figura 3 - Representação das matrizes de dados X e Y na calibração multivariada.

$$\mathbf{X} = \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{vmatrix}$$

**Variáveis Independentes**

$$\mathbf{y} = \begin{vmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{vmatrix}$$

**Variáveis Dependentes**

Uma vez construído o modelo, este deve ser validado. A validação pode ser de dois tipos: validação interna (também chamada de validação cruzada) ou validação externa (PARREIRA, 2003).

A validação interna baseia-se na divisão das amostras em dois subgrupos, um de calibração e o outro de validação. O modelo é construído usando as amostras do conjunto de calibração. O método comumente utilizado na validação interna é o método conhecido como *leave-one-out* ou “deixe um de fora”. Para cada amostra removida, um modelo é construído com as amostras restantes (n-1), sendo o procedimento repetido n vezes (o número de amostras do conjunto de calibração), porém somente o melhor modelo é proposto pelo *software* (WOLD; SJOSTROM; ERIKSSON, 2001).

A validação externa é utilizada quando já se tem um modelo proposto para um conjunto de calibração e, então, adicionamos novas amostras desconhecidas para a previsão da propriedade de interesse.

No presente trabalho, devido ao pequeno número de amostras, o modelo foi validado apenas pelo método *leave-one-out*. Para que a análise multivariada tenha sucesso é de fundamental importância a aplicação de métodos de pré-processamento e a análise exploratória de dados pelo método de PCA (análise de componentes principais).

#### 1.4.7 Métodos de pré-processamento

A obtenção dos dados espectrais pode ser comprometida por agentes como desgaste do equipamento, instabilidade da corrente elétrica, condições ambientais, sensibilidade de acessórios de detecção, entre outros. Em consequência disto, o espectro pode apresentar informações que não são decorrentes do material que está sendo analisado, mas sim de interferências de origem instrumental-analítica.

A etapa de pré-processamento é essencial para a correta interpretação de dados multiplicativos (BRERETON, 2003). Os principais objetivos das técnicas de pré-processamento são eliminar as informações irrelevantes do ponto de vista químico (ex: ruídos) e tornar a matriz de dados adequada para a análise exploratória (SOUZA; POPPI, 2012).

A maneira mais simples e rotineira em laboratórios é a média de sinal, levando a uma redução de ruído e a outra é a correção na linha de base devido aos efeitos de espalhamento de luz nos dados espectrais (WEHRENS, 2011). No presente trabalho, foram aplicados, nos dados espectrais, os pré-processamentos de correção de sinal multiplicativo (MSC) e, em



seguida, a centralização dos dados na média (MC), e nos dados das propriedades, aplicou-se o autoescalamamento.

#### 1.4.7.1 Correção de sinal multiplicativo (MSC)

Em espectroscopia de infravermelho, a superfície da amostra pode influenciar nas medições e conduzir a uma dispersão da luz acarretando em uma linha de base não constante. Para a correção de linha de base em espectroscopia, usa-se o método de correção de sinal multiplicativo (do inglês, *Multiply Scatter Correction* - MSC) (WEHRENS, 2011).

O objetivo deste método é remover o efeito do espalhamento através da linearização de cada espectro por um espectro ideal. O espectro ideal é o espectro médio do conjunto de dados que está sendo analisado. Em seguida, utiliza-se uma regressão linear para calcular o coeficiente angular e linear entre o espectro ideal e o espectro que se quer corrigir. Posteriormente, o espectro corrigido é obtido, subtraindo-se cada ponto do espectro pelo valor do coeficiente linear e dividindo pelo valor do coeficiente angular (SOUZA et. al., 2012).

#### 1.4.7.2 Centralização na média (MC)

A centralização dos dados na média (do inglês, *Mean Center* – MC) deve ser utilizada se houver problemas de linha de base, se apenas uma pequena região dos dados é utilizada e/ou se os dados forem muito diferentes entre si (CORGOZINHO, 2009). Baseia-se em calcular a média das intensidades para cada comprimento de onda, ou seja, para cada coluna obtêm-se a média e, em seguida, subtrai-se este valor de cada variável dessa mesma coluna. Desta forma, as possíveis diferenças nas intensidades de cada variável são mais fáceis de serem vistas (SOUZA; POPPI, 2012; SOUZA et. al., 2012).

$$X_{ij} \text{ (centrado na média)} = X_{ij} \text{ (original)} - \bar{X}_j \text{ (média)} \quad \text{(Equação 5)}$$

onde  $X_{ij}$ , corresponde ao valor centrado na média para a  $i$ -ésima amostra e  $j$ -ésima variável,  $X_{ij}$  (original) é o valor da variável  $j$  na amostra  $i$ ,  $\bar{X}_j$  é a média da  $j$ -ésima coluna calculada pela Equação 6:

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad \text{(Equação 6)}$$

### 1.4.7.3 Autoescalamento

O autoescalamento é um pré-tratamento dos dados, sendo um procedimento de rotina na construção dos modelos nos estudo das relações quantitativas entre a estrutura química e a atividade biológica ou alguma propriedade físico-química. Quando as variáveis têm diferentes unidades ou quando a faixa de variação dos dados é grande, o que ocorre com frequência nos estudos de propriedades físico-químicas, se recomenda o autoescalamento das variáveis. Com este procedimento, a influência de uma variável dominante é minimizada em cálculos posteriores. O autoescalamento implica em subtrair de cada elemento de uma coluna da matriz de dados o valor médio da respectiva coluna e dividir o resultado pelo desvio padrão da mesma, de acordo com a Equação 7 (MARTIN; FERREIRA, 2013; CARVALHO, 2005).

$$X_{ij(\text{as})} = \frac{x_{ij} - \bar{x}_j}{s_j} \quad (\text{Equação 7})$$

onde  $x_i$  e  $x_{(i\text{as})}$  são, respectivamente, os valores da  $j$ -ésima variável, do  $i$ -ésimo composto antes e depois do autoescalamento;  $\bar{x}_j$  é o valor médio da  $j$ -ésima variável e  $s_j$  é o seu desvio padrão. Este tratamento é aplicado à matriz dos descritores e ao vetor contendo as propriedades físico-químicas.

### 1.4.8 Método de classificação

#### 1.4.8.1. Análise de componentes principais (PCA)

A base fundamental da maioria dos métodos modernos para tratamentos de dados multivariados é o PCA (FERREIRA et. al., 1999). É um método exploratório dos dados porque ele permite a separação de informações relevantes das redundantes dentro de um conjunto de dados originais, ou seja, é utilizado para extrair informações e detectar tendências nos dados, baseados nas medições multivariadas das amostras (SENA et. al., 2000; SOUZA; POPPI, 2012).

A PCA (do inglês, *Principal Component Analysis*) é um método que permite reduzir a dimensionalidade de dados, ou seja, diminuir o número de variáveis espectrais do conjunto original formando um novo sistema (SOUZA; POPPI, 2012; CÔCCO, 2008). Essa redução da dimensionalidade dos dados ocorre a partir de combinações lineares das variáveis originais

formando novos eixos chamados componentes principais (PCs) que são ortogonais entre si permitindo a ausência de colinearidade nos mesmos (TEÓFILO, 2013).

A PCA decompõe uma matriz de dados  $X$  (onde as  $n$  linhas são as amostras e as  $m$  colunas são as variáveis) obtendo as novas matrizes chamadas “scores”,  $T$  e “loadings”,  $P^T$ , mais uma matriz de erros ( $E$ ). Os “scores” fornecem a composição das PCs em relação aos objetos (amostras) enquanto os “loadings” fornecem essa mesma composição em relação às variáveis (SENA et. al., 2000). A equação de decomposição da matriz de dados  $X$  pode ser mostrada na forma matricial da Equação 8.

$$X = TP^T + E \quad (\text{Equação 8})$$

Cada PC é construída a partir das combinações lineares das variáveis originais cuja principal característica, além da ortogonalidade, é que são obtidas em ordem decrescente de máxima variância, ou seja, a PC1 detém mais informações estatísticas que a PC2, que por sua vez tem mais informação estatística que a PC3 e, assim, sucessivamente (MOITA NETO; MOITA, 1998; FERREIRA et. al., 1999).

A análise de componentes principais também é usada para conjecturar quais variáveis originais apresentam maior importância, ou seja, as variáveis originais de maior peso (loadings) (MOITA NETO; MOITA, 1998).

Após a otimização do conjunto espectral e das propriedades de interesse, pode-se fazer a modelagem dos dados pelos métodos de calibração.

#### 1.4.9 Métodos de calibração

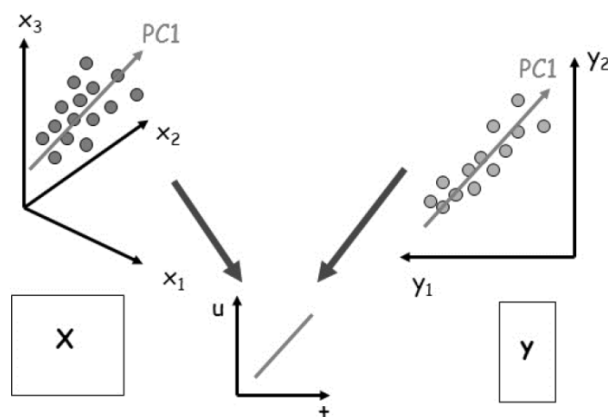
Existem vários métodos matemáticos para realizar a calibração e o usado neste trabalho foi a Regressão por Mínimos Quadrados Parciais (PLS), que é o método mais utilizado em calibração multivariada.

##### 1.4.9.1 Regressão por mínimos quadrados parciais (PLS)

O PLS (do inglês, *Partial Least Square*) é um método que utiliza a técnica de análise de componentes principais (PCA) para a redução da dimensionalidade dos dados originais para posteriormente correlacionar os espectros (matriz  $X$ ) com as propriedades de interesse (matriz  $Y$ ) (SOUZA et. al., 2013).

No método de regressão por PLS, modelam-se duas matrizes de dados X e Y buscando a máxima covariância entre elas, de modo que cada PC sofra uma pequena modificação em seus ângulos deixando de ser ortogonais e passando a ser chamada de variáveis latentes, (VLs), (Figura 4).

Figura 4 - Formação da Variável Latente.



Fonte: FONSECA, 2011.

A regressão por PLS é capaz de modelar, simultaneamente, muitas variáveis de resposta, mas quando se correlaciona com apenas uma propriedade de interesse, Y é um vetor e, neste caso, o método é denominado PLS1. Quando se tem mais de uma propriedade, Y é uma matriz e, nesse caso, o método é denominado PLS2.

O objetivo da regressão por PLS é construir um modelo linear capaz de prever uma ou mais propriedades físico-químicas (Y) a partir de espectros medidos (X). Nas Equações 9 e 10, a seguir, a matriz X é decomposta por PCA e, posteriormente, incorpora-se a informação de Y para melhorar a correlação linear entre ambas as matrizes, de modo que são modeladas em variáveis latentes, formando dois grupos de matrizes, uma de escores T e U e outra de *loadings* (pesos) P e Q e duas matrizes de resíduos E e F de X e Y, respectivamente. As matrizes de resíduos são constituídas pelas variáveis latentes que não fazem parte da modelagem (SOUZA et. al., 2013; FONSECA, 2011).

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E} \quad (\text{Equação 9})$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \quad (\text{Equação 10})$$

O modelo consiste em relacionar, linearmente, os escores do bloco X (matriz T) com os escores do bloco Y (matriz U) para cada VL, como visto na (Equação 11).

$$u_h = b_h t_h \quad (\text{Equação 11})$$

onde,  $b_h$  é o coeficiente de regressão linear do modelo para cada (VL).

O número de variáveis latentes a ser utilizado no modelo é normalmente determinado pelo procedimento de validação cruzada com o método *leave-one-out* (BASSBASI et. al., 2014). O processo da variância nas VLs é obtido de forma decrescente; sendo assim, a maior variância é explicada pela primeira VL, a segunda VL terá uma menor variância explicada e, assim, sucessivamente (FONSECA, 2011).

#### 1.4.10 Elipse de confiança

De acordo com FRANCO et. al., 2002, o uso da quimiometria para a validação de um método de regressão garante a qualidade de um método analítico, isto é, se a qualidade do processo de medição e dos dados produzidos não é suficientemente adequada, a informação química pode ser incerta ou até mesmo errada. Uma das características mais importantes de um método analítico é a sua exatidão que é avaliada através da comparação dos valores de concentração obtidos pelo método proposto (o método usado na modelagem) com os valores fornecidos pelo de referência (método analítico).

Algumas vezes, o erro médio de previsão (definido no APÊNDICE B) da propriedade de interesse é referido como uma estimativa da exatidão e utilizado para comparar os resultados obtidos em trabalhos com calibração multivariada, através do erro médio quadrático de previsão (RMSEP – *Root Mean Squares Error of Prediction*), entretanto, o RMSEP é um parâmetro global que incorpora tanto erros sistemáticos quanto aleatórios (VALDERRAMA; BRAGA; POPPI, 2009).

De acordo com SKOOG et al. (2009), as análises químicas são afetadas por dois tipos de erros que são causados por padronizações ou calibrações malfeitas ou variações aleatórias e incertezas nos resultados. O primeiro é chamado erro aleatório, que afeta a precisão e faz com que os dados se distribuam de forma mais ou menos simétrica em torno do valor médio e o segundo tipo de erro, denominado erro sistemático, afeta a exatidão e faz com que a média de um conjunto de dados seja diferente do valor aceito. Os termos que descrevem a precisão

são o desvio-padrão, a variância ou o coeficiente de variação e todos são uma função de quanto um resultado individual  $x_i$  difere da média. A exatidão indica a proximidade da medida do valor verdadeiro, ou aceito, e é expressa em termos do erro absoluto ou erro relativo.

Devido à presença destes erros, o método mais indicado para validação é a execução de uma regressão do valor estimado pelo modelo proposto *versus* o valor medido e a estimação do coeficiente angular e linear, incluindo as considerações da região conjunta da elipse de confiança usando mínimos quadrados ordinários (OLS) (VALDERRAMA et. al., 2007; VALDERRAMA; BRAGA; POPPI, 2009).

A regressão linear por mínimos quadrados ordinários (OLS), (em inglês, *Ordinary Least Square*) é frequentemente realizada sem ponderação, isto é, dando um peso igual a todos os pontos de dados experimentais (FRANCO et. al., 2002). O método de mínimos quadrados ordinários fornece estimativas únicas dos coeficientes de regressão, para qualquer amostra ou conjunto de dados, proporcionando um menor valor possível do somatório dos erros elevado ao quadrado ( $\sum \hat{u}_i^2$ ), nos quais esses erros apresentam a mesma variância, ou seja, se distribuem simetricamente representando a premissa de homocedasticidade (GUJARATI, 2006). Os valores estimados do coeficiente de inclinação ( $\hat{b}$ ) e intercepto ( $\hat{a}$ ) são calculados pelas equações a seguir (FRANCO et. al., 2002).

$$\hat{b} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2} \quad (\text{Equação 12})$$

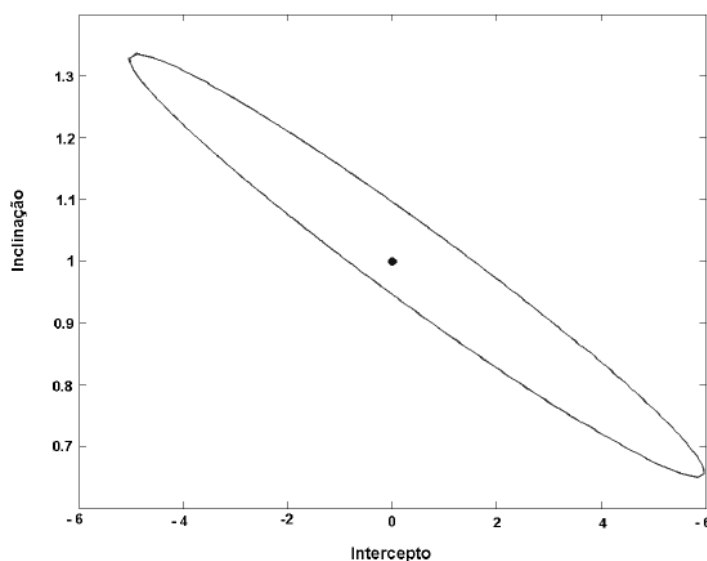
$$\hat{a} = \bar{y} - \hat{b}\bar{x} \quad (\text{Equação 13})$$

onde  $\bar{x}$  e  $\bar{y}$  são os valores médios dos dados de x e y, respectivamente, e N é o número de dados.

Na comparação de resultados entre dois métodos distintos, a exatidão pode ser acessada por meio da comparação dos valores obtidos para a inclinação e o intercepto de uma reta ajustada entre os valores de referência e os estimados pelo modelo. Isto pode ser mais bem observado através da região conjunta de elipse de confiança (EJCR), como mostra o exemplo hipotético da Figura 5. Caso os intervalos de confiança para esses parâmetros contenham os seus valores esperados iguais a 1 e 0, para a inclinação e o intercepto,

respectivamente, os modelos podem ser considerados estatisticamente equivalentes no nível de confiança considerado, como no exemplo. Valores estimados para a inclinação e intercepto fora de seus intervalos de confiança indicam erros sistemáticos proporcionais e constantes, respectivamente (VALDERRAMA; BRAGA; POPPI, 2009).

Figura 5 - Elipse de confiança.



Fonte: VALDERRAMA, 2009.

O intervalo de confiança é uma estimativa de intervalo para um parâmetro de uma população como média e a variância, por exemplo, na qual esta estimativa de intervalo é precisa sendo expressa pelo comprimento do intervalo, ou seja, quanto mais curto o intervalo mais precisa é a estimação. Da mesma forma, podem ser obtidos intervalos de confiança para os parâmetros de regressão, ou seja, para os coeficientes individuais de regressão, angular e linear. O desenvolvimento de um procedimento para obtenção de intervalos de confiança requer que os erros sejam normais e independentemente distribuídos com média zero e variância constante (MONTYGOMERY; RUNGER, 2012).

### **1.5 Uso de ferramentas quimiométricas e da espectrometria no infravermelho para a predição de propriedades**

LIRA et al. (2010) utilizaram modelo de calibração multivariada por mínimos quadrados parciais (PLS) combinado com espectroscopia nas regiões de MIR (infravermelho médio) e NIR (infravermelho próximo) usando dois tipos de equipamentos: o primeiro foi um

espectrômetro FTIR convencional (Perkin-Elmer, modelo Spectrum GX spectrometer) nas regiões MIR e NIR e o segundo um instrumento FTIR portátil (Irox Diesel) somente na região MIR para prever a qualidade dos parâmetros (propriedades), tais como densidade, teor de enxofre e temperatura de destilação, de misturas de diesel/biodiesel e misturas de diferentes tipos de biodiesel. Os resultados de erros de predição dos modelos desenvolvidos usando o instrumento convencional para ambas as regiões MIR e NIR foram comparados com a reprodutibilidade do método padrão correspondente. Foi aplicado um teste F para avaliar as diferenças entre os modelos desenvolvidos para ambas as regiões no espectrômetro convencional e os resultados mostraram que não havia diferenças significativas entre os modelos. Sendo assim ambos os modelos nas regiões MIR e NIR mostraram-se eficientes para a previsão das propriedades investigadas nas misturas diesel/biodiesel e nas misturas de biodiesel, exceto para a densidade na região NIR, ou seja, o modelo não forneceu um valor previsto de acordo com a norma nas misturas diesel/biodiesel e para o ponto de destilação T85% na região espectral NIR das misturas de biodiesel. Foi feita uma comparação dos modelos PLS usando somente a região espectral MIR com o instrumento FTIR portátil e os resultados foram similares aos modelos obtidos pelo instrumento convencional, podendo-se concluir que os dois instrumentos são eficientes para prever as propriedades em análise de ambos os conjuntos de calibração (misturas com tipos diferentes de biodiesel e misturas de biodiesel/diesel).

VASCONCELOS et al. (2012) avaliaram o uso de espectroscopia no infravermelho próximo (NIR) e médio (MIR) para determinar o teor de biodiesel em misturas de diesel/biodiesel, considerando a presença de óleo vegetal como contaminante. O nível máximo e mínimo nas misturas contendo biodiesel e óleo foi 0 – 10% (v/v) numa razão de éster/óleo numa faixa de concentração de 0,25 a 6 totalizando um conjunto com 132 amostras. Os métodos de calibração multivariada utilizados foram regressão linear múltipla com seleção de variáveis pelo algoritmo de projeções sucessivas (MLR/SPA), método por mínimos quadrados parciais com seleção de variáveis com o algoritmo de Jack- Knife (PLS/JK) e o método por mínimos quadrados parciais sem seleção de variáveis (PLS) para assegurar a eficiência dos métodos com SPA e JK. Todos os modelos foram avaliados através do erro médio quadrático de predição (RMSEP). A faixa espectral NIR utilizada foi de 12.000 a 6.300  $\text{cm}^{-1}$  com os caminhos ópticos de 10, 20 e 50 mm e os modelos com seleção de variáveis apresentaram coeficientes de determinação significativos em comparação com o modelo PLS sem seleção de variáveis. Além disso, foi avaliado o desempenho dos modelos nas regiões NIR de 9000 a 4000  $\text{cm}^{-1}$  com caminho óptico de 1 mm e MIR com acessório de reflectância



total atenuada (ATR). Os autores concluíram que a melhor estratégia de modelagem é por MLR/SPA tanto na região NIR de 12.000 – 6.300  $\text{cm}^{-1}$  com os caminhos ópticos de 10, 20 e 50 mm como na região de NIR de 9000 - 4000  $\text{cm}^{-1}$  com caminho óptico de 1 mm, inclusive utilizando a espectroscopia MIR que também é capaz de determinar o teor de biodiesel contido nas misturas de diesel/biodiesel considerando a presença de óleo vegetal. O método de calibração por MLR/SPA na região MIR é mais simples, rápido e utiliza um menor número de variáveis latente sendo o mais indicado para a determinação de biodiesel nas misturas.

PIMENTEL et al. (2006) desenvolveram um modelo multivariado usando mínimos quadrados parciais (PLS) baseado em espectroscopia de infravermelho médio – MIR (4.000 – 600  $\text{cm}^{-1}$  com o acessório de reflectância total atenuada – ATR) e próximo – NIR (12.000 – 4.000  $\text{cm}^{-1}$  com uma célula de comprimento óptico de 1 mm) para a determinação de biodiesel em misturas de biodiesel/diesel considerando a presença de óleos vegetais como contaminantes. Os níveis máximo e mínimo de ambos os fatores biodiesel e óleo foi entre 0 - 5% (v/v) com nove níveis de concentração (0,5 – 4,5% (v/v)) numa razão de éster/óleo de 0,2 a 8 totalizando 26 misturas com 26 amostras diferentes de diesel. As amostras foram preparadas usando óleo de soja, óleo de rícino e óleo de fritura usado, misturados com seus respectivos ésteres e diferentes amostras de diesel obtidas de estações de serviço na região metropolitana de Salvador, Brasil. Foram aplicados nos dados espectrais, como pré-tratamento, a primeira derivada e a análise de PCA. Os resultados mostram que o modelo proposto na região NIR parece ser melhor do que o modelo na região MIR devido aos valores do erro médio quadrático de predição (RMSEP) obtidos, 0,18 e 0,25, respectivamente. Porém, foi feito um teste F indicando que os valores de RMSEP não são estatisticamente significativos em um nível de confiança de 95% para essas diferenças. Deste modo, os autores concluíram que ambos os modelos para os espectros MIR e NIR são adequados para a previsão do teor de biodiesel nas misturas de diesel de petróleo, na presença de possíveis óleos como contaminantes originados do mau controle do processo de produção de biodiesel ou, ainda, a partir da adição ilegal.

ROCHA (2012) propôs a construção de um modelo na região NIR para prever a concentração de biodiesel em misturas de diesel/biodiesel numa faixa de 2 a 90 %. O modelo foi construído usando mínimos quadrados parciais (PLS) com 3 variáveis latentes para um menor valor de erro médio quadrático de validação cruzada (RMSECV) que explicou 99,9% das variáveis X e 98,9% de Y. A validação do modelo foi feita através dos valores de erro médio quadrático de calibração (RMSEC), RMSECV e erro médio quadrático de predição (RMSEP) e o coeficiente de correlação (R), e os resultados mostraram que o modelo ajustou

bem os dados. A distribuição dos erros indicou um comportamento linear porque a maioria dos erros observados foi abaixo de 5 %, ou seja, o modelo foi capaz de prever com 95% de confiança a concentração de biodiesel, sendo o modelo considerado apropriado para estimar a concentração de biodiesel nas misturas. Tal afirmação foi comprovada pela ausência de erros sistemáticos neste modelo de calibração que foi avaliado através da região conjunta da elipse de confiança estimando o ponto ideal que mostrou um valor de coeficiente angular (slope) = 0,9711 e de coeficiente linear (intercepto) = 0,0133. O autor estimou algumas figuras de mérito para validar o modelo PLS e os resultados para o limite de detecção e quantificação indicaram que o método NIR-PLS foi capaz de detectar e quantificar concentrações de biodiesel maiores que 0,2 % e 0,6 %, respectivamente, sendo apropriado para a faixa de concentração de biodiesel estudada que é de 2- 90%. O modelo apresentou um baixo valor de sensibilidade (0,0098%), ou seja, sua sensibilidade de sinal analítico é capaz de distinguir uma diferença de concentração nas amostras de biodiesel inferior a 0,1 %. Como as concentrações de biodiesel nas amostras estudadas diferiram em pelo menos 0,2%, a adequação do modelo foi comprovada. Outro parâmetro de validação foi a seletividade do modelo, a qual informa a quantidade de sinal original do analito de cada amostra usada no modelo e que está exclusivamente relacionada com a propriedade de interesse. O valor foi igual a 0,0141, o que significa que cerca de 98,5% do sinal analítico foi utilizado na modelagem dos dados.

SOUZA et. al. (2014) utilizaram a espectroscopia de infravermelho médio (MIR) utilizando o acessório de reflectância total atenuada (ATR) combinada com a calibração multivariada por meio da regressão de mínimos quadrados parciais (PLS) para quantificar a presença de adulterante, como o óleo lubrificante automotivo (RAL). As amostras de diesel S-10 brasileiro foram preparadas numa faixa de concentração de 1,75 -30,0 % em massa (w/w). O modelo foi validado de acordo com o método padrão ASTM E1655-05 e figuras de mérito também foram estimadas para validar o modelo. O modelo PLS construído mostrou um bom ajuste para a análise quantitativa de RAL nas amostras de diesel S-10 brasileiro, apresentando um coeficiente de correlação igual a 0,9989 e um RMSEP de 0,40% (w/w). Os autores utilizaram a regressão por mínimos quadrados ordinários incluindo a consideração da região conjunta da elipse de confiança. O resultado da região conjunta da elipse de confiança (EJCR) indicou que o modelo PLS não mostrou diferença significativa entre os valores de concentração preparada e as concentrações previstas pelo modelo PLS, apresentando 95% de confiança. Portanto, o modelo PLS foi capaz de quantificar o adulterante RAL em amostras de diesel S-10 brasileiro dentro da faixa estudada.

ALVES e POPPI (2013) desenvolveram modelos de calibração multivariada aplicada em espectroscopia de infravermelho próximo (NIR) para determinação de biodiesel contido em misturas de diesel/biodiesel. Os modelos de calibração foram desenvolvidos nas faixas espectrais (i)  $4.400 - 6.200 \text{ cm}^{-1}$  e (ii)  $4.400 - 4.600 \text{ cm}^{-1}$ , comparando qual modelo se ajustou melhor aos dados: mínimos quadrados parciais (PLS) ou máquinas de vetores de suporte (SVM), para um teor de  $0 - 35\%(\text{v/v})$  e  $0 - 100\%(\text{v/v})$  de biodiesel contido em misturas com diesel. Os resultados de erro médio quadrado de predição foram comparados e o modelo SVM apresentou um RMSEP (0,12/0,11) e o modelo PLS apresentou um RMSEP (0,76/0,13) nas regiões (i) e (ii), respectivamente. Os valores de RMSEP dos modelos em ambas as regiões foram menores do que o mínimo de reprodutibilidade exigido pelo método ASTM para determinação do biodiesel. Os autores concluíram que na região (ii) ambos os modelos apresentaram resultados similares de RMSEP (0,11-SVM/0,13-PLS) com uma distribuição residual similar e aleatória. A similaridade dos resultados de RMSEP na região (ii) foi devida à presença da banda de absorção relacionada com o modo de vibração do grupo metil terminal dos ésteres metílicos. Na região (i) o modelo SVM apresentou um RMSEP melhor que o do PLS devido à distribuição residual que apresentou valores menores do que aqueles obtidos na região (ii) e aleatórios, diferentemente dos resíduos obtidos por PLS que mostraram-se tendenciosos. Os autores utilizaram a regressão por mínimos quadrados ordinários (OLS) incluindo a consideração da região conjunta da elipse de confiança para obter mais informações da exatidão dos modelos. Os resultados das elipses do modelo PLS e SVM não apresentaram diferenças significativas na região (ii) na faixa de concentração de  $0 - 35\%(\text{v/v})$  de biodiesel nas misturas com diesel. Na região (i), o resultado da elipse do modelo SVM apresentou um menor tamanho, mostrando que o modelo SVM se ajusta melhor aos dados, sendo mais eficiente na determinação de biodiesel do que o modelo PLS. Na faixa de concentração de  $0 - 100\%(\text{v/v})$ , os resultados das elipses do modelo PLS e SVM não apresentaram diferenças significativas na região (ii) e o modelo SVM na região (i) foi o de melhor ajuste e eficiência na determinação de biodiesel em misturas com diesel. Sendo assim, esses resultados de EJCRC (elipses de confiança) confirmam que o melhor modelo de previsão para determinar as concentrações de biodiesel na faixa entre  $0 - 35\%$  e  $0 - 100\%(\text{v/v})$  através da análise de regressão utilizando máquinas de vetores de suporte (SVM).

Em outras áreas de pesquisas como, por exemplo, farmacêutica, alimentícia, cosmética pode-se encontrar autores que estão aderindo à combinação da técnica espectroscópica combinada com métodos quimiométricos para a previsão de alguma propriedade de interesse não sendo aplicados apenas para combustíveis.

MORGANO et. al. (2005) desenvolveram uma metodologia para a determinação do teor de proteína total em amostras de café cru, utilizando os espectros de reflectância no infravermelho próximo (NIR) e aplicação de regressão por mínimos quadrados parciais (PLS), pois as análises convencionais consomem muito tempo e geram muitos resíduos químicos, como o método de Kjeldhal. Para a construção do modelo de regressão foram usadas 53 amostras de café cru, da variedade arábica, procedentes de diferentes regiões produtoras de café dos estados de São Paulo, Minas Gerais e Bahia. Os espectros originais na região NIR sofreram diferentes transformações e pré-tratamentos matemáticos, como a transformação Kubelka-Munk; correção de sinal multiplicativa (MSC); alisamento (SPLINE); derivada primeira; média móvel e o pré-tratamento dos dados escalados pela variância. O conjunto de dados foi dividido em 29 amostras para o conjunto de calibração e 24 amostras usadas na validação externa do modelo. Os modelos propostos foram por PLS sem e com seleção de variáveis, modelos I e II, respectivamente e os resultados mostraram menores valores de erro e um ligeiro aumento no valor do coeficiente de correlação linear ( $r$ ) para o modelo II (com a seleção de variáveis), sendo possível obter um modelo mais parcimonioso e mais ajustado indicando uma boa capacidade de previsão dos teores de proteína total pelo método proposto para as amostras de café cru.

BRAGA e POPPI (2004) tiveram como objetivo a determinação de figuras de mérito em modelos de calibração multivariada construídos pelo método de Mínimos Quadrados Parciais (PLS) e exemplificar sua utilização na validação de um modelo de calibração para a determinação de pureza polimórfica de carbamazepina em misturas binárias dos polimorfos I e III, construído com base em dados de reflectância difusa na região do infravermelho próximo (NIR). A carbamazepina (CBZ) é um conhecido princípio ativo de medicamentos antiepiléticos, que apresenta quatro formas polimórficas, três anidras (formas I, II e III) e uma di-hidratada (IV). O medicamento produzido à base de CBZ é feito com a forma III que é absorvida pelo organismo mais rapidamente, devido à sua solubilidade, sendo que a forma I é tida como impureza na matéria-prima que é adquirida pela indústria. O modelo de calibração PLS foi construído para a determinação quantitativa da forma III da CBZ em misturas binárias das formas I e III, entre a faixa de concentração de 80 a 100% (m/m), do polimorfo III em relação ao I. Para construção do modelo, foram preparadas 81 amostras na concentração entre 80% a 100%. Após a detecção de amostras anômalas, segundo os testes descritos na norma E1655 da ASTM17, foram usadas para a construção do modelo 38 amostras na etapa de calibração e 39 para validação. Como pré-processamento utilizou a 1ª derivada das absorvâncias, seleção da melhor faixa espectral e dados centrados na média. O

modelo PLS construído utilizou três variáveis latentes, que correspondem a mais de 95 % da variância dos dados originais (espectrais e concentrações de referência). O maior erro percentual observado para as amostras de validação foi de 1,7 %, com valores absolutos seguindo aproximadamente uma distribuição normal. Os resultados das figuras de mérito determinadas para o modelo PLS demonstraram que modelos de calibração multivariada podem ser validados de forma a atender às exigências impostas por órgãos de fiscalização e padronização.

## 2. METODOLOGIA

### 2.1 Preparo das misturas de biodiesel de soja e óleo de soja

Foram preparadas 11 misturas de biodiesel de soja e óleo de soja em diferentes proporções de 0-100%, em quintuplicata, totalizando 55 amostras. Foram utilizadas duas buretas de 25 mL e uma balança analítica (Figura 6) para medições volumétricas e mássicas, respectivamente, de óleo de soja e do biodiesel original. Primeiramente, mediu-se o volume de biodiesel e, em seguida, a sua massa. O mesmo procedimento foi realizado para o óleo, obtendo-se um volume total de 30 mL para cada mistura. O biodiesel, com atribuição nominal B100, foi sintetizado e cedido pelo CENPES/PETROBRAS. O óleo de soja utilizado foi um produto comercial da marca SOYA. As misturas preparadas foram armazenadas em frascos escuros e mantidas em local escuro, à temperatura ambiente.

Figura 6 - Material utilizado no preparo das amostras.



Fonte: O autor, 2014.

Tabela 2 - Medições volumétricas e mássicas de biodiesel e óleo de soja das misturas

Amostras/Solução	Biodiesel (% em volume)	Volume de biodiesel (mL)	Volume de óleo de soja (mL)	Massa de biodiesel (g)					Massa de óleo de soja (g)				
01/02/03/04/05	100	30	0	26,4337	26,4548	26,4251	26,3688	26,3664	0	0	0	0	0
06/07/08/09/10	90	27	3	24,0499	23,7471	23,7249	23,731	23,7563	2,7599	2,7086	2,6945	2,6935	2,7094
11/12/13/14/15	80	24	6	21,0741	21,0578	21,0552	21,0654	21,0717	5,343	5,3926	5,3483	5,3411	5,3587
16/17/18/19/20	70	21	9	18,3937	18,3872	18,3858	18,3788	18,3811	8,0144	8,0185	8,3815	8,0389	8,0031
21/22/23/24/25	60	18	12	15,7202	15,758	15,7175	15,779	15,7292	10,6901	10,7693	10,6639	10,777	10,6667
26/27/28/29/30	50	15	15	12,0818	12,0741	12,0717	12,0794	12,2575	13,3342	13,3544	13,3782	13,348	13,489
31/32/33/34/35	40	12	18	9,4742	9,5234	9,5174	9,499	9,4976	16,0811	16,0819	16,1236	16,0775	16,0428
36/37/38/39/40	30	9	21	7,8474	7,8535	7,8526	7,8266	7,8336	18,7558	18,7561	18,7641	18,7734	18,7646
41/42/43/44/45	20	6	24	5,257	5,2568	5,2597	5,222	5,2624	21,4661	21,4405	21,5152	21,4398	21,4383
46/47/48/49/50	10	3	27	2,6339	2,6467	2,6457	2,6535	2,6283	24,1243	24,1584	24,1526	24,1417	24,1242
51/52/53/54/55	0	0	30	0	0	0	0	0	26,8376	26,7959	26,8598	26,7809	26,8214

Fonte: O autor, 2014.

## 2.2 Aquisição dos espectros de MIR

Os espectros na região do infravermelho médio ( IR -7.800 a 450  $\text{cm}^{-1}$ ) foram obtidos em um espectrofotômetro de infravermelho com Transformada de Fourier (FT-IR) Perkin Elmer modelo Frontier (Figura 8), equipado com um detector de sulfato de triglicina deuterado (DTGS). Para as análises foi utilizado um acessório de reflectância total atenuada horizontal (ATR) com cristal de quartzo, com uma resolução de 4  $\text{cm}^{-1}$ , intervalo de dados de 1  $\text{cm}^{-1}$  e 20 varreduras. As amostras preparadas foram analisadas, colocando-se 0,2 mL da mistura no porta-amostra, obtendo-se o seu espectro de IV.

Figura 7 - Espectrômetro de IV com acessório de Reflectância Total Atenuada (ATR).



Fonte: O autor, 2014.

## 2.3 Quantificação dos ésteres por cromatografia em fase gasosa (CG) e cromatografia líquida de alta eficiência (HPLC)

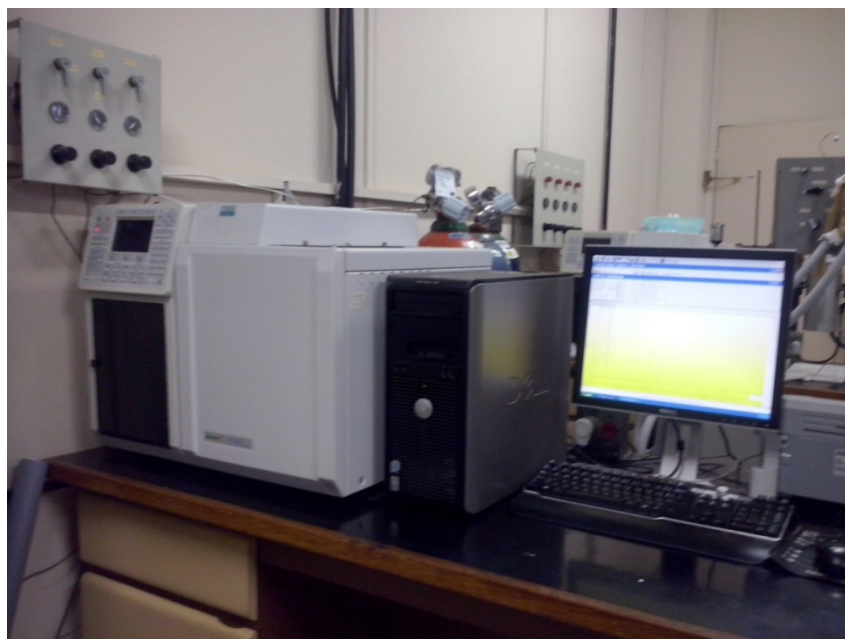
O cromatógrafo a gás utilizado para a quantificação de ésteres metílicos presentes nas misturas de óleo de soja e biodiesel de soja foi o CG Varian CP3800 (Figura 8) com detector de ionização de chama. A temperatura da coluna, CP WAX 52 CB com 30 m de comprimento, foi programada para 220 °C durante os 15 primeiros minutos e, em seguida, foi aumentada até 240°C a uma taxa de 10°C/min e permaneceu em 240°C por 10 minutos. As temperaturas do injetor e do detector foram mantidas em



250 °C. O volume das amostras injetadas foi de 1 µL, sendo cada amostra previamente dissolvida na proporção de 0,1 g de amostra para 1 mL do padrão (solução de heptadecanoato de metila em n-heptano com concentração 10g/L). Para a quantificação do teor de ésteres metílicos presentes na mistura, seguiu-se a norma EN14103. O teor de ésteres metílicos obtido por esta técnica é o teor de biodiesel de soja obtido, a partir de uma simulação de conversão do óleo de soja, já que os ésteres metílicos são oriundos exclusivamente das amostras de biodiesel.

Um cromatograma típico obtido por análise de Cromatografia Gasosa (CG) pode ser observado no APÊNDICE C. Os teores de ésteres metílicos foram quantificados de acordo com os cálculos descritos no APÊNDICE D, baseados na norma padronizada pela ANP (EN 14103).

Figura 8 - Cromatógrafo Varian CP3800.

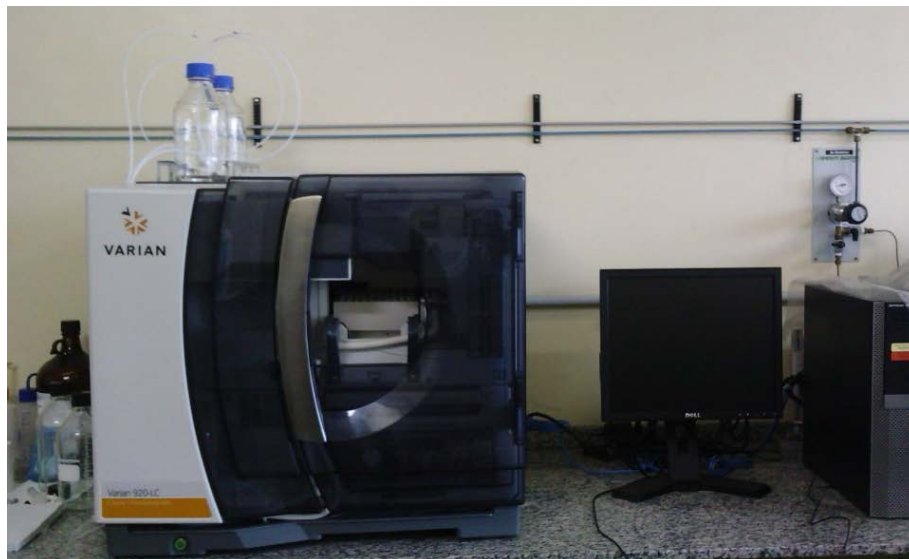


Fonte: O autor, 2014.

Para as medidas por cromatografia em fase líquida de alta eficiência (HPLC) foi utilizado um cromatógrafo Varian 920LC (Figura 10), com detector UV a 205 nm. Usou-se uma coluna Pursuit 5µm C18 250 x 4,6 mm operando à temperatura constante de 26 °C. Como fase móvel foram utilizados metanol (A) e mistura isopropanol/n-hexano(B) na proporção de 5/4 com uma vazão de 1 mL/min. A quantidade de 100 mg de cada amostra foi dissolvida em 5 mL da fase móvel (eluente A/ eluente B 50/50). A

identificação dos ésteres metílicos (EsMAG) foi baseada no trabalho de ANDRADE (2011).

Figura 9 - Cromatógrafo Varian 920LC.



Fonte: O autor, 2014.

Um cromatograma típico obtido por análise de Cromatografia Líquida de Alta Eficiência (HPLC) pode ser observado no APÊNDICE E. Os teores de ésteres metílicos foram calculados de acordo com o procedimento descrito no APÊNDICE F, baseado na tese de ANDRADE (2011).

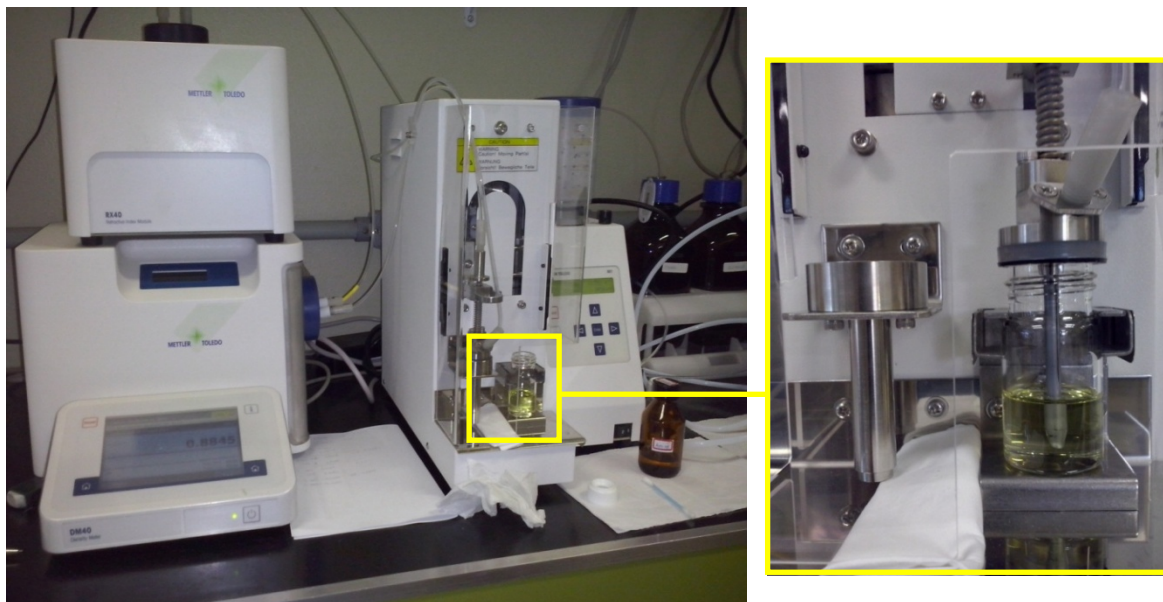
#### **2.4 Determinação da massa específica e do índice de refração**

Para a determinação das massas específicas e dos índices de refração das amostras foi utilizado o densímetro DM40 com Módulo de Índice de Refração RX40 - Mettler Toledo (Figura 10), sendo o sistema programado de forma a informar as duas medidas simultaneamente. As massas específicas foram determinadas à temperatura de 20°C e os índices de refração a 40°C.

Primeiramente, foram colocados 10 mL de amostra no frasco de vidro, sendo o equipamento ajustado à 20°C, ocorrendo a sucção da amostra até o densímetro. Em seguida, o sistema é aquecido à 40°C e a amostra segue para o módulo RX40, obtendo-se, ao final, ambas as medidas. Antes da análise de uma nova amostra, o equipamento é

programado para realizar a limpeza automaticamente, utilizando etanol e acetona. Na sequência, o equipamento é resfriado à 20°C, podendo-se iniciar a próxima medida.

Figura 10 - Densímetro DM40/Modulo Índice de Refração RX40 - Mettler Toledo.



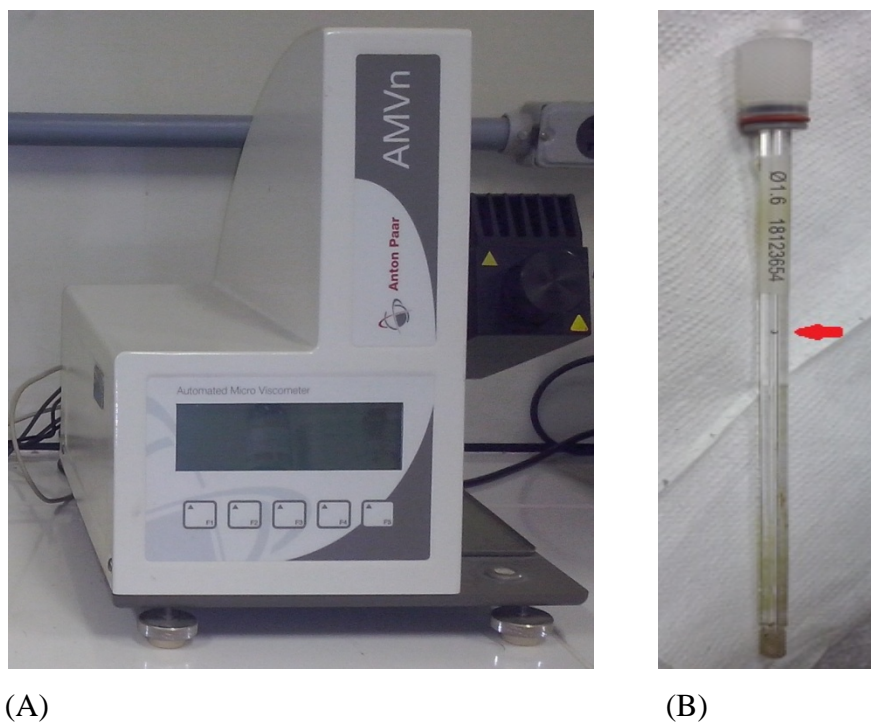
Fonte: O autor, 2014.

## 2.5. Determinação da viscosidade

O equipamento utilizado para as medidas de viscosidade foi o Viscosímetro Automático Micro/Anton Paar AMVn e capilar  $\varnothing 1.6$  (18123654). Este tipo de equipamento também é conhecido como viscosímetro de bolas. Ele determina o tempo de queda de bola (uma esfera metálica) indicada com a seta vermelha no interior do capilar (Figura 11). O capilar fica inserido no Pietter (caixa preta) sob um ângulo de 60°. Quando se inicia a análise, essa caixa aquece o capilar a 40°C e o gira até 60° no sentido horário e anti-horário, de forma a se medir o tempo decorrido para que a esfera metálica deslize pelo capilar. Depois do tempo medido, este valor é substituído na equação de calibração do equipamento, calculando-se, assim, a viscosidade da amostra. As misturas contendo um maior teor de óleo (80-100%, amostras 41 a 55) apresentaram erros na leitura devido à alta viscosidade das mesmas. Foram feitas tentativas para se encontrar uma única condição que fosse adequada para a realização de todas as análises, o que não foi possível, devido às diferenças de viscosidade entre elas. Sendo assim, os

valores de viscosidade correspondentes às amostras 41 a 55 não foram utilizados na construção do modelo.

Figura 11 - Viscosímetro.



Legenda: A)Viscosímetro Automated Micro/Anton Paar AMVn. B) Capilar.

Fonte: O autor, 2014.

## 2.6 Softwares utilizados no trabalho

O *software* utilizado para a inferência estatística (distribuição dos dados e anomalias das amostras) foi o programa R versão 3.0.1 e para as análises quimiométricas (PCA, pré-processamento, PLS2) foi utilizado o programa Matlab R2008b com a ferramenta PLS\_Toolbox versão 6.5.2.

## 2. RESULTADOS E DISCUSSÃO

### 3.1 Medições das propriedades físico-químicas

Como citado anteriormente (seção 2.1), foram preparadas 11 misturas contendo biodiesel de soja e óleo de soja em proporções de 0 a 100%, em quintuplicata, totalizando 55 amostras. Cada frasco foi pesado e tarado, sendo medidos as massas e os volumes das misturas como pode observado na Tabela 2. As medições das propriedades podem ser encontradas nas Tabelas 3 a 7.

Podemos observar que os valores de massa específica, à 20°C (Tabela 3) nas diferentes proporções de biodiesel, apresentaram uma boa reprodutibilidade. Nota-se, ainda, que o aumento gradual de óleo de soja na mistura provoca o aumento da massa específica da amostra, coerente com o maior valor da massa específica do óleo de soja puro ( $\approx 0,92 \text{ g/cm}^3$ ) comparado ao do biodiesel puro ( $\approx 0,88 \text{ g/cm}^3$ ). A partir de 50% de adição de óleo de soja ao biodiesel (amostras 26-50), as amostras ficam fora do limite estabelecido pela resolução ANP (limite máximo igual a  $900 \text{ Kg/m}^3$  ou  $0,900 \text{ g/cm}^3$ ).

Tabela 3 - Valores de massa específica obtidos a 20°C ( $\text{g/cm}^3$ )

Amostra/Solução	(%) Biodiesel	Massa específica ( $\text{g/cm}^3$ )				
01/02/03/04/05	100	0,8834	0,8834	0,8834	0,8834	0,8834
06/07/08/09/10	90	0,8870	0,8870	0,8870	0,8870	0,8870
11/12/13/14/15	80	0,8905	0,8906	0,8905	0,8906	0,8906
16/17/18/19/20	70	0,8908	0,8924	0,8915	0,8922	0,8924
21/22/23/24/25	60	0,8879	0,8980	0,8879	0,8980	0,8879
26/27/28/29/30	50	0,9016	0,9016	0,9016	0,9016	0,9016
31/32/33/34/35	40	0,9053	0,9052	0,9052	0,9052	0,9052
36/37/38/39/40	30	0,9088	0,9088	0,9089	0,9089	0,9089
41/42/43/44/45	20	0,9115	0,9115	0,9116	0,9116	0,9116
46/47/48/49/50	10	0,9161	0,9162	0,9162	0,9162	0,9162
51/52/53/54/55	0	0,9198	0,9199	0,9199	0,9199	0,9199

Fonte: O autor, 2014.

Na Tabela 4, podemos observar a mesma tendência observada para a massa específica, ou seja, os valores do índice de refração nas quintuplicatas são bem semelhantes e, à medida que a quantidade de biodiesel é reduzida, maiores são os índices de refração, resultado coerente com os valores do óleo e do biodiesel puros. O índice de refração não é um parâmetro de certificação da qualidade do biodiesel, mas devido ao equipamento utilizado nas medições de massa específica ser acoplado ao refratômetro, decidiu-se por obter ambas as medidas, uma vez que o índice de refração está diretamente relacionado com a estrutura molecular.

Tabela 4 - Valores do índice de refração obtidos a 40°C

Amostra/Solução	Biodiesel (%)	Índice de Refração				
01/02/03/04/05	100	1,4575	1,4575	1,4575	1,4575	1,4575
06/07/08/09/10	90	1,4592	1,4592	1,4592	1,4592	1,4592
11/12/13/14/15	80	1,4609	1,4609	1,4609	1,4609	1,4609
16/17/18/19/20	70	1,4618	1,4628	1,4622	1,4626	1,4626
21/22/23/24/25	60	1,4642	1,4642	1,4642	1,4642	1,4642
26/27/28/29/30	50	1,4659	1,4659	1,4659	1,4659	1,4659
31/32/33/34/35	40	1,4676	1,4676	1,4676	1,4676	1,4676
36/37/38/39/40	30	1,4693	1,4693	1,4693	1,4693	1,4693
41/42/43/44/45	20	1,4709	1,4709	1,4710	1,4709	1,4709
46/47/48/49/50	10	1,4726	1,4726	1,4726	1,4726	1,4726
51/52/53/54/55	0	1,4743	1,4743	1,4743	1,4743	1,4743

Fonte: O autor, 2014.

Na Tabela 5, podemos observar que o teor de ésteres metílicos obtido por HPLC diminui à medida que se aumenta o teor de óleo de soja presente na mistura, resultado coerente já que os ésteres metílicos são oriundos da amostra de biodiesel. Pode ser observado nessas medições que a amostra de biodiesel tida como B100, possui apenas

entre 80 e 85 % de ésteres metílicos. Outro ponto de destaque é que a reprodutibilidade das medidas não é tão elevada quanto à observada para as medidas de massa específica e índice de refração, com diferenças que chegam a aproximadamente 30% (segunda série, amostras 06 e 10). Este resultado sugere que a simplificação feita na análise do cromatograma (considerar que as áreas dos picos são diretamente proporcionais à concentração) não foi adequada para estimar o percentual de ésteres metílicos. Essas variações serão mais bem discutidas após a modelagem dos dados obtidos.

Tabela 5 - Teor de ésteres metílicos ( simulado) por HPLC

Amostra/Solução	Biodiesel (%)	Teor de ésteres metílicos (forma de %)				
01/02/03/04/05	100	82.7	83.5	87.9	89.4	89.2
06/07/08/09/10	90	51.0	69.6	73.7	71.0	74.9
11/12/13/14/15	80	64.8	58.2	61.5	64.2	61.8
16/17/18/19/20	70	47.6	53.2	52.4	54.0	50.9
21/22/23/24/25	60	47.3	44.0	43.7	47.6	47.5
26/27/28/29/30	50	40.7	41.8	37.1	37.9	37.7
31/32/33/34/35	40	31.3	33.7	33.9	35.0	34.0
36/37/38/39/40	30	25.2	28.1	26.3	27.2	27.8
41/42/43/44/45	20	19.1	20.8	18.9	20.6	20.5
46/47/48/49/50	10	12.2	11.0	12.7	11.6	11.5

Fonte: O autor, 2014.

Na Tabela 6, observa-se que os teores de ésteres metílicos obtidos por cromatografia em fase gasosa apresentam, em geral, valores semelhantes nas réplicas para cada nível, sendo este método recomendado pela resolução ANP nº 14 de 2012 para identificar os ésteres metílicos formados a partir de triglicerídeos através do método referencial EN14103. Comparando-se as Tabelas 6 e 7, observa-se que nos percentuais de 100, 80 e 40% de biodiesel os valores obtidos por HPLC estão próximos dos valores encontrados por CG e nas amostras contendo 30, 20 e 10% de biodiesel a quantificação por HPLC foi maior que pelo método de referência (CG). Este fato indica que a quantificação dos ésteres metílicos por HPLC não deve ser totalmente descartada, precisando, no entanto, de um refinamento no tratamento dos dados.

Tabela 6 - Teor de ésteres metílicos (simulado) por CG

Amostra/Solução	Biodiesel (%)	Teor de ésteres metílicos (forma de %)				
01/02/03/04/05	100	82,4	85,6	84,5	84,0	83,6
06/07/08/09/10	90	81,1	81,8	83,2	83,4	78,7
11/12/13/14/15	80	66,5	67,4	68,3	64,1	64,6
16/17/18/19/20	70	60,5	62,3	60,0	61,1	61,6
21/22/23/24/25	60	55,6	55,9	51,0	55,9	53,9
26/27/28/29/30	50	41,9	41,1	41,4	43,4	41,0
31/32/33/34/35	40	32,0	33,6	34,6	34,2	34,3
36/37/38/39/40	30	25,4	25,1	25,2	25,2	26,3
41/42/43/44/45	20	17,3	17,6	17,1	17,2	17,4
46/47/48/49/50	10	8,8	8,0	8,0	8,5	8,3

Fonte: O autor, 2014.

Na Tabela 7, observa-se que os valores de viscosidade cinemática aumentam com a diminuição do teor de biodiesel nas misturas e que, a partir de 30% de óleo de soja, as amostras ficam fora do limite estabelecido pela ANP (entre 3,0 e 6,0 mm<sup>2</sup>/s).



Tabela 7 - Valores das viscosidades cinemáticas à 40 °C (mm<sup>2</sup>/s)

Amostra/Solução	Biodiesel (%)	Viscosidade Cinemática à 40 °C (mm <sup>2</sup> /s)				
01/02/03/04/05	100	4,392	4,360	4,387	4,380	4,389
06/07/08/09/10	90	5,364	5,343	5,347	5,368	5,352
11/12/13/14/15	80	5,777	5,776	5,777	5,776	5,776
16/17/18/19/20	70	8,340	8,055	8,497	8,061	7,983
21/22/23/24/25	60	9,289	9,311	9,306	9,328	9,306
26/27/28/29/30	50	11,45	11,37	11,34	11,34	11,36
31/32/33/34/35	40	13,85	13,85	13,89	13,84	13,89
36/37/38/39/40	30	16,96	16,95	16,91	16,90	16,98

Fonte: O autor, 2014.

### 3.2 Espectroscopia no infravermelho: interpretação dos espectros

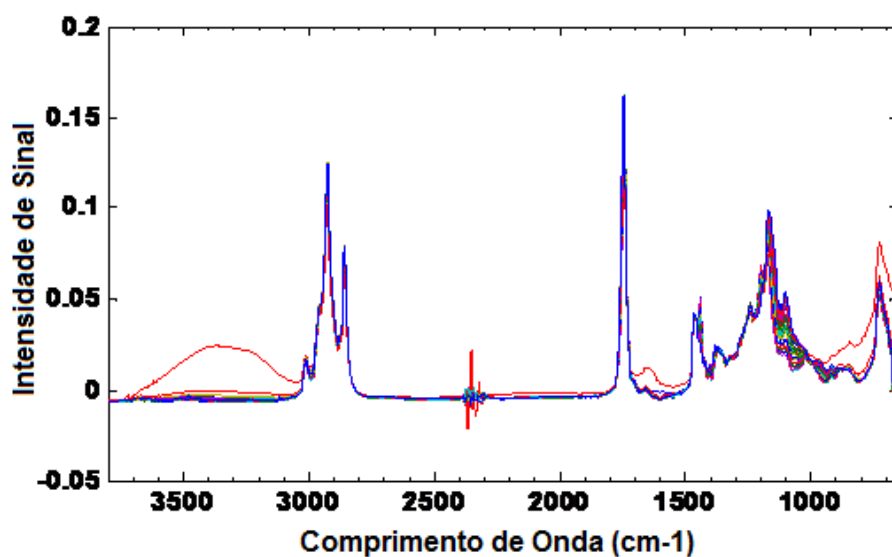
A Figura 12 mostra os espectros de FTIR/ATR obtidos para as 50 amostras correspondentes ao biodiesel puro e às misturas de biodiesel e óleo de soja. A faixa de números de onda trabalhada nos espectros de infravermelho foi de 3.801 – 650 cm<sup>-1</sup>, totalizando 3152 variáveis. As atribuições de algumas bandas presentes nestes espectros de infravermelho médio são apresentadas na Tabela 8 (SILVERSTEIN, 2000).

Tabela 8 – Atribuição das bandas no espectro de infravermelho médio das amostras estudadas na região entre 3600 – 650 cm<sup>-1</sup>

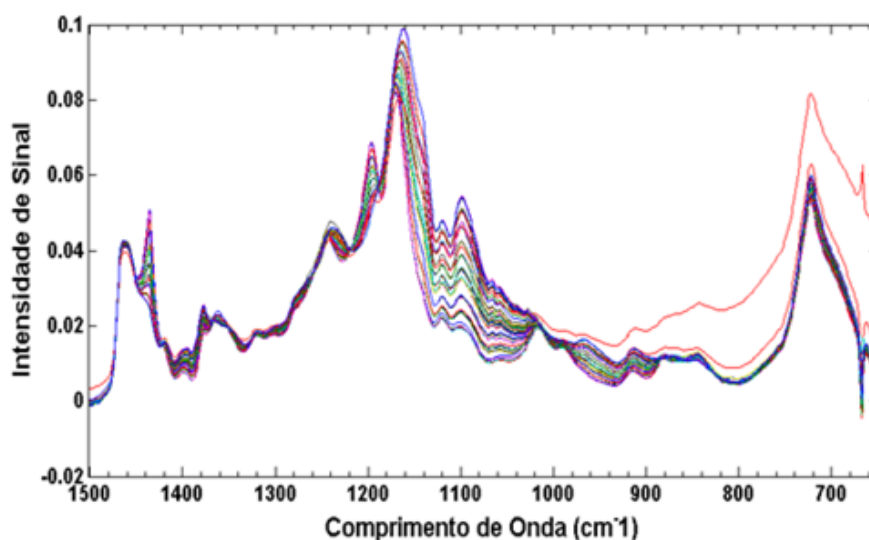
Banda	Atribuição
3010	Estiramento H – C = (carbono com ligação dupla)
2925	Estiramento assimétrico H – C (carbono saturado)
2855	Estiramento simétrico H – C (carbono saturado)
1742-1744	Estiramento C = O (carbonila) de éster
1300 – 1100	Estiramentos da ligação C-O de ésteres

Fonte: SILVERSTEIN, 2000.

Figura 12 - Espectros de Infravermelho



(A)



(B)

Legenda: Espectros de ATR/FTIR das 50 amostras correspondentes ao biodiesel puro e as misturas de biodiesel e óleo de soja na faixas de (A) 3.801 a 650  $\text{cm}^{-1}$  e (B) ampliado na região de 1.500 – 1000  $\text{cm}^{-1}$ .

Fonte: O autor, 2014.

Podemos observar na (Figura 12 - A) dois espectros com um perfil espectral diferente dos demais, na região de 3650 - 3590  $\text{cm}^{-1}$  que corresponde à região de absorção de grupos hidroxila - OH. A presença de banda de absorção relacionada a esses grupos foi provavelmente decorrente da limpeza (solução de EXTRAN, em seguida água e após etanol anidro) feita no cristal de quartzo do acessório ATR do equipamento de infravermelho, entre cada uma das análises.

Para os espectros representados na (Figura 12 – A), na faixa entre 3600 – 1500  $\text{cm}^{-1}$ , são observadas as bandas características dos grupos funcionais presentes tanto nas moléculas de triglicerídeos quanto de ésteres metílicos (biodiesel). As principais diferenças entre os espectros são observadas na região entre 1500 – 1000  $\text{cm}^{-1}$  (Figura 12 – B), onde, além das deformações angulares associadas aos grupos  $\text{CH}_2$  e  $\text{CH}_3$ , encontram-se as bandas correspondentes às vibrações da função éster. Este grupamento apresenta diferenças nas moléculas dos triglicerídeos (óleo) e dos ésteres metílicos de ácidos graxos (biodiesel), sendo representado por  $\text{R}_1 - \text{C}(=\text{O}) - \text{OR}$  e por  $\text{R}_1 - \text{C}(=\text{O}) - \text{O} - \text{CH}_3$ , respectivamente. Nessa região, as bandas correspondentes à deformação angular simétrica do grupo  $\text{CH}_2$  (metileno) e à deformação angular assimétrica do grupo  $\text{CH}_3$  (metil), aparecem sobrepostas em 1460-1464  $\text{cm}^{-1}$ , enquanto a deformação angular simétrica do grupo  $\text{CH}_3$  é observada a 1378-1362  $\text{cm}^{-1}$  (SILVERSTEIN, 2000).

As diferenças mais significativas são observadas nas posições das bandas associadas às vibrações  $\text{C} - \text{C}(=\text{O}) - \text{O}$  e  $-\text{O} - \text{C}$  -. Os ésteres metílicos de ácidos graxos apresentam um padrão com três bandas próximas em 1250, 1200 e 1175  $\text{cm}^{-1}$ , sendo a última delas a mais intensa. No caso dos triglicerídeos, observam-se três bandas próximas a 1250, 1160 e 1100  $\text{cm}^{-1}$  sendo aquela a 1160  $\text{cm}^{-1}$  a mais intensa. A mudança gradativa nas posições e intensidades dessas bandas pode ser observada com o aumento no teor de ésteres na mistura. Ainda nessa mesma região, a banda a 1436  $\text{cm}^{-1}$ , ausente no triglicerídeo puro, torna-se mais intensa à medida que aumenta a quantidade de éster metílico de ácido graxo na mistura e corresponde à deformação angular do grupo metil ligado ao oxigênio da carboxila.

Uma banda comum a todas as amostras é observada em 722  $\text{cm}^{-1}$  e corresponde à deformação angular do tipo balanço do grupo  $\text{RHC} = \text{CHR}$  (SILVERSTEIN, 2000).

### 3.3 Testes de *outliers* e normalidade

Testes estatísticos foram aplicados nos valores medidos das propriedades de interesse, massa específica, índice de refração e teor de ésteres metílicos para avaliar e determinar possíveis anomalias e verificar a normalidade da distribuição dos dados. Os dados de viscosidade não foram submetidos a estes testes, devido a limitações do equipamento que impediram a determinação da viscosidade em condições idênticas para todas as amostras preparadas.

Os testes aplicados para a detecção de amostras anômalas (*outliers*) foram os testes de Grubbs e Dixon e, para avaliar a normalidade, o teste de Shapiro-Wilk. Os testes foram aplicados em 50 amostras, pois as cinco últimas misturas (amostras 51-55) contêm apenas óleo de soja e estas não foram analisadas na identificação de ésteres metílicos por CG, seguindo a norma EN14103 recomendada pela ANP, uma vez que as amostras estão constituídas apenas de triacilglicerídeos, não havendo a presença de ésteres metílicos.

Os testes de *outliers* (Grubbs – estatística G e Dixon – estatística Q) foram aplicados nas amostras (01 a 50), avaliando as medições de massa específica, índice de refração e teor de ésteres metílicos por HPLC. Os testes de hipóteses foram feitos utilizando o programa R versão 3.0.1 e aplicados em todos os conjuntos de réplicas (níveis) de cada parâmetro. O programa R calcula os testes de Grubbs e Dixon, porém não fornece nenhum resultado quando os níveis analisados não apresentam anomalias nas medições, apenas quando algum valor suspeito no nível é identificado. O *software* fornece o valor calculado (G, para o teste de Grubbs, e Q, para o teste de Dixon), o valor crítico U para o teste de Grubbs (não fornecido para o teste de Dixon) e o p-valor.

Nas propriedades massa específica e índice refração, o *software* identificou um valor atípico no nível 4 (Tabela 9) para ambas as propriedades. O nível 4 refere-se às amostras 16 a 20, correspondente à concentração de 70% de biodiesel. Pode-se observar nas (Tabelas 3 e 4 – massa específica e índice de refração, respectivamente) que, exceto para o nível 4, não há grandes diferenças nas medições dos demais níveis. No nível 5 há algumas diferenças nas medições das amostras, porém o programa R não identificou nenhuma amostra discrepante neste nível.

Tabela 9 - Testes Estatísticos de Grubbs e Dixon para o nível 4 (amostras com 70% de biodiesel)

	Teste de Grubbs	Teste de Dixon
Massa específica	G = 1,4592	Q = 0,375
	U = 0,3346	-
	p-valor = 0,2308	p-valor = 0,5943
Índice de refração	G = 1,5000	Q = 0,4
	U = 0,2969	-
	p-valor = 0,1899	p-valor = 0,5258

G e Q = valor calculado U = valor crítico

Fonte: O autor, 2014.

Embora, Miller e Miller (2005) afirmem que o valor em análise deve ser rejeitado quando a estatística calculada exceder o valor crítico (U), Montgomery e Runger (2012) afirmam que, quando o p-valor é conhecido, ele tem um maior peso de informação de quão significativos são os dados. Neste caso, o p-valor (que conduz à rejeição da hipótese nula  $H_0$ ), é comparado com o nível de significância pré-definido  $\alpha=0,05$ . O p-valor seria a probabilidade de que o valor discrepante ou a variável aleatória analisada seja maior que a estatística calculada, por isso, quando o p-valor é menor que 0,05 o valor é rejeitado.

Na Tabela 9, observa-se que na estatística G (Grubbs), os p-valores de ambas as propriedades não conduzem à rejeição da  $H_0$  devido aos p-valores (0,2308 e 0,1899) serem maiores que o nível de significância estimado que é de  $\alpha=0,05$ . Sendo assim, o p-valor para as propriedades de massa específica e índice de refração não são significativos, isto é, não há evidências suficientes para provar que a  $H_0$  seja incorreta e, por isso, não se rejeita o valor discrepante identificado pelo programa R no nível 4 da mistura.

Na estatística Q (Dixon), como o programa R não informa o valor crítico e, como foi seguida a definição de Montgomery e Runger (2012) baseada no p-valor fornecido, também não se rejeitou a  $H_0$  assim como foi feito na estatística (G), mantendo-se esses valores no conjunto de dados originais.

Para os resultados de teor de ésteres metílicos por HPLC, em todos os níveis o *software* identificou amostras atípicas, mas isto era esperado uma vez que os resultados mostravam variações significativas nos valores para um mesmo nível, diferentemente das medições de massa específica e índice de refração, como pode ser mostrado na Tabela 10.

Ainda na Tabela 10, pode-se observar que ambos os testes, G e Q, detectaram a mesma amostra anômala no nível 2, correspondente à concentração de 90% de biodiesel. Neste caso, o valor discrepante referente à amostra 06 deve ser rejeitado, pois os p-valores = 0,01079 (G) e 0,02054 (Q) são menores que  $\alpha=0,05$ , concluindo-se que o p-valor é significativo, ou seja, há evidências de que as medições neste nível 2 não são da mesma população e  $H_0$  deve ser rejeitada (MONTGOMERY; RUNGER, 2012), o que levou à remoção da amostra 06 do conjunto de dados originais.

Tabela 10 - Testes Estatísticos de Grubbs e Dixon para o teor de ésteres por HPLC para o nível 2 com 90% de biodiesel referente à amostra 06

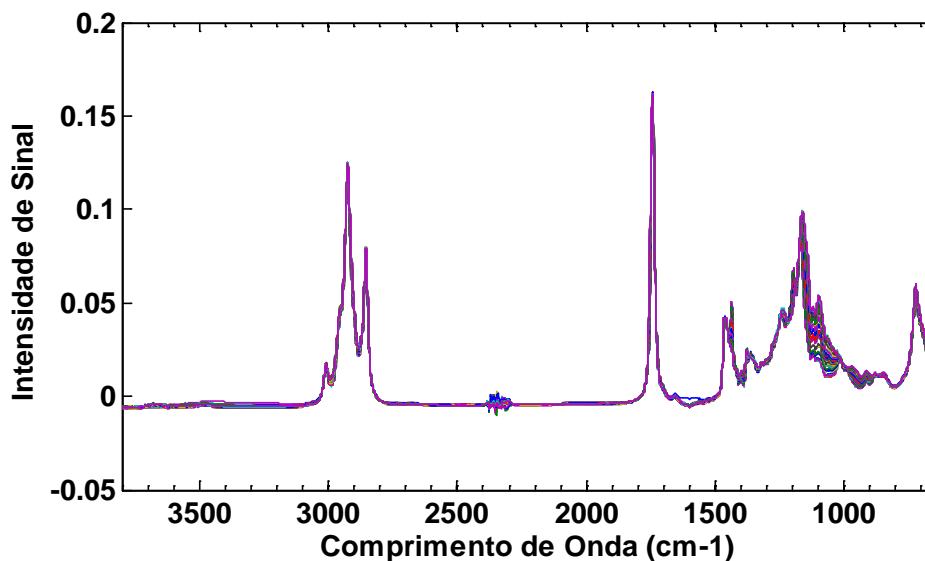
	Teste de Grubbs	Teste de Dixon
Teor de ésteres metílicos	G = 1,7468 U = 0,0465	Q = 0,7782 -
	p-valor = 0,01079	p-valor = 0,02054
	G e Q = valor calculado	U = valor crítico

Fonte: O autor, 2014.

Em seguida, foi aplicado o teste de Shapiro-Wilk (estatística W) para avaliação da normalidade dos dados na matriz com 49 amostras. Exatamente como no caso das estatísticas G e Q, o programa R só fornece os resultados calculados para o teste de Shapiro-Wilk caso o nível analisado não siga uma distribuição normal. Neste caso, o *software* não identificou a presença de distribuição não normal de dados nos níveis analisados. Pode-se concluir que, pelo teste de Shapiro-Wilk, todos os níveis apresentaram uma distribuição normal.

Em seguida, analisando os espectros brutos ainda na Figura 12, pelo ponto de vista químico podemos observar que dois espectros apresentam uma divergência em relação aos demais na intensidade de sinal na região de 3650 - 3590  $\text{cm}^{-1}$  que corresponde à região de absorção de grupos hidroxila - OH. A presença de banda de absorção relacionada a esses grupos foi provavelmente decorrente da limpeza (solução de EXTRAN, em seguida água e após etanol anidro) feita no cristal de quartzo do acessório ATR do equipamento de infravermelho, entre cada uma das análises. Estes dois espectros divergentes correspondem às amostras 31 e 38 e foram removidos do conjunto de dados espectrais (matriz X) gerando uma nova matriz com 47 amostras. Da mesma forma, os dados das propriedades referentes a essas duas amostras também foram removidos do conjunto de propriedades (matriz Y). O perfil espectral contendo a matriz com 47 amostras pode ser visto na Figura 13.

Figura 13 - Espectros de ATR/FTIR com 47 amostras das misturas de biodiesel e óleo de soja.



Fonte: O autor, 2014.

### 3.4 Pré-processamento

Os espectros brutos de IV podem apresentar informações que não são decorrentes somente do material que está sendo analisado, mas também de interferências de origem instrumental-analítica. Por isso, de acordo com Rinnan (2009), os dados espectrais sempre devem ser pré-processados antes da modelagem, já que não é possível a obtenção de espectros perfeitos.

A matriz de dados X contendo 47 amostras foi submetida a dois tipos de pré-processamentos: primeiramente, a correção de sinal multiplicativo (MSC) que tem por finalidade corrigir a linha de base devido aos efeitos de espalhamento de luz nos dados espectrais e, em seguida, a centralização dos dados na média (MC) onde são calculadas as médias de cada coluna de variáveis e subtraídas das variáveis daquela coluna (CORGOZINHO, 2009).

Na matriz Y, o pré-tratamento empregado foi o autoescalamento que é o mais adequado para dados que estão em escalas muito diferentes e se quer atribuir a mesma significância a todos (CORGOZINHO, 2009).

### 3.5 Análise de Componentes Principais (PCA)

Após os pré-processamentos MSC e MC, o conjunto espectral com 47 amostras foi submetido à análise de componentes principais (PCA) e identificou-se uma amostra anômala que corresponde à amostra 01. Em seguida, esta amostra foi removida da matriz X e Y para a posterior modelagem. Na Tabela 11 podemos observar os resultados estatísticos da PCA com a matriz espectral contendo 46 amostras fornecido pelo programa PLS\_Toolbox.

O conjunto de dados contendo 46 amostras pré-tratadas por correção de sinal multiplicativo (MSC) e centralização dos dados na média (MC) apresentou um menor valor de erro médio quadrático de calibração (RMSEC) e validação cruzada (RMSECV) pelo método *leave-one-out* (LOO) utilizando duas componentes principais (PCs) que explicou 97,58% da variabilidade dos dados com as 46 amostras para a primeira componente principal (PC) e 0,76% para a segunda PC.

Tabela 11 - Resultados estatísticos da análise de componentes principais com 46 amostras de MIR fornecidos pelo programa PLS\_Toolbox

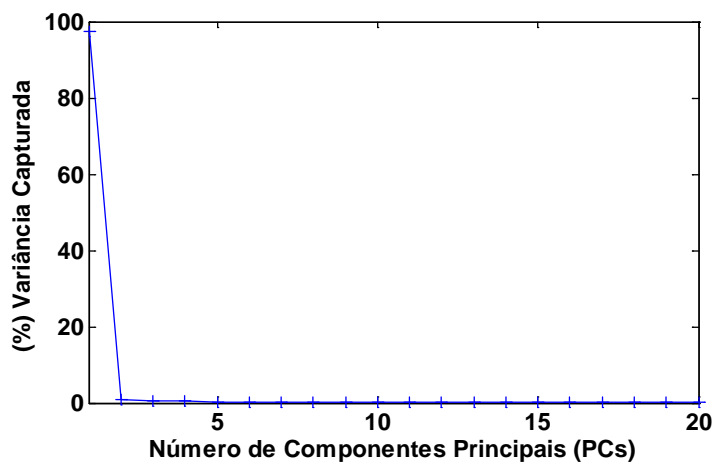
PCA	
Amostras de Calibração	46
Pré-processamento	MSC + MC
Validação Cruzada	LOO
Número de PC	2
RMSEC	0,00025
RMSECV	0,2953
% Variância Capturada	97,58 / 0,76

Fonte: O autor, 2014.

Na Figura 14, temos o gráfico de rendimento de variância capturada *versus* o número de componentes principais e podemos observar que acima de duas PCs não há ganho significativo na variância capturada.



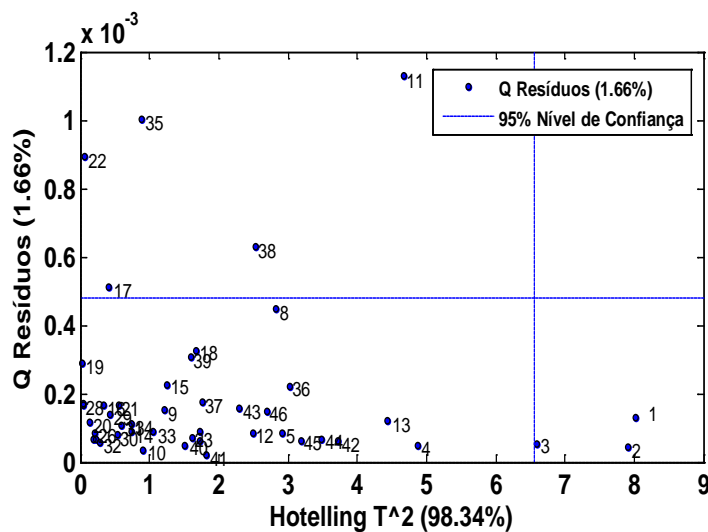
Figura 14 - Resultado da análise de PCA (% de variância capturada *versus* PCs) na matriz com 46 amostras de MIR.



Fonte: O autor, 2014.

O gráfico de  $T^2$  *versus* Q-resíduos (Figura 15) foi utilizado para avaliar o novo modelo após a remoção da amostra 01, pois, a princípio, na primeira rodada da PCA com 47 amostras foi possível identificar esta amostra como um *outlier* devendo ser removida por não se encontrar dentro do nível de confiança de 95% apresentando valores em  $T^2$  e Q elevados. Como pode ser visto na Figura 15, a segunda rodada da PCA com 46 amostras não apontou a existência de *outlier* (amostra anômala), pois a maioria das amostras está dentro do nível de confiança de 95% em  $T^2$ . Pode ser observado que as amostras 11, 17, 22, 35, 38 possuem resíduos altos, porém separadamente elas não influenciam negativamente no modelo devido ao baixo valor de  $T^2$ . Portanto, para uma amostra ser removida do conjunto de dados, ambos os valores de Q e  $T^2$  devem ser elevados, indicando que a mesma apresenta uma influência negativa para o modelo. Verifica-se que a quantidade de resíduos é de 1,66 %, o que significa que 98,34 % das informações com duas componentes principais foram utilizadas na calibração, ou seja, a remoção da amostra 01 do conjunto de dados espectrais de MIR aumentou a porcentagem de variância capturada com duas PCs. Caso a amostra 01 não fosse removida, a significância do modelo seria comprometida, pois deve-se buscar o menor número possível de PCs que gere um menor erro de validação cruzada (RMSECV) e que contenha as informações mais significativas das variáveis. Caso contrário, se há variáveis com informações irrelevantes, o modelo precisará de mais PCs e, conseqüentemente, um aumento no valor do erro de validação.

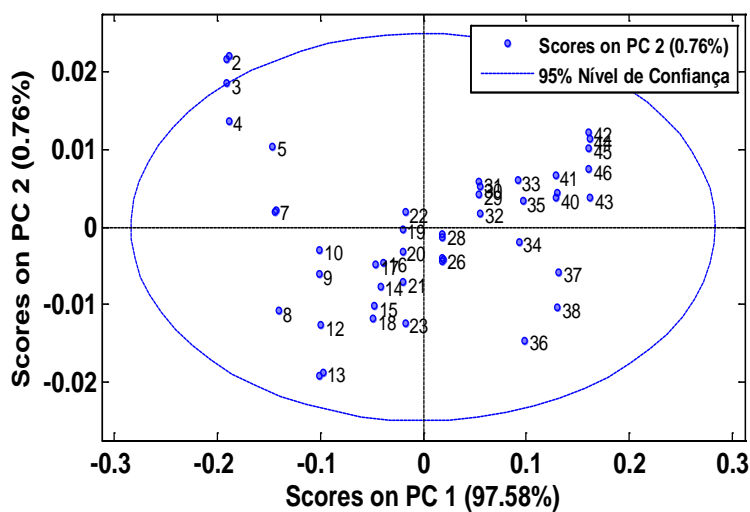
Figura 15 - Gráfico dos resíduos *versus* escores (*Hotelling T<sup>2</sup> versus* Q resíduos) na matriz com 46 amostras de MIR.



Fonte: O autor, 2014.

A Figura 16 apresenta o gráfico de escores (PC1 x PC2) para o conjunto com 46 dados, onde pode ser visto que a PC1 explica 97,58% da variabilidade dos dados enquanto que a PC2 explica 0,76% desta variabilidade. Podemos observar que a amostra 02 está fora do intervalo de confiança no segundo quadrante do lado superior esquerdo do gráfico, mas a mesma possui um valor baixo de resíduos (Figura 15). Como mencionado anteriormente, para uma amostra ser removida do conjunto de dados ambos os valores de  $Q$  e  $T^2$  devem ser elevados, pois a mesma apresentaria uma influência negativa para o modelo. Como esse não é o caso da amostra 02, conclui-se que os dados são homogêneos, ou seja, não há amostras atípicas no conjunto de dados.

Figura 16 - Gráfico dos escores (PC1 *versus* PC2) na matriz com 46 amostras de MIR.



Fonte: O autor, 2014.

### 3.6 Regressão por mínimos quadrados parciais (PLS)

Uma vez realizada a análise estatística seguida da análise exploratória de dados através da PCA na matriz original com 50 amostras, chegou-se a uma matriz final com 46 amostras. A partir daí, iniciou-se a etapa para buscar a correlação entre os espectros MIR e as propriedades de interesse, massa específica, índice de refração e porcentagem de ésteres metílicos estabelecida por HPLC.

O método de calibração utilizado foi a regressão por Mínimos Quadrados Parciais (PLS2), sendo o conjunto com os 46 espectros de infravermelho médio (MIR) submetido a um algoritmo denominado *Kennard-Stone*, que tem como objetivo dividir o conjunto de dados em um conjunto de calibração e outro de validação.

O modelo é construído pela regressão por PLS2, no qual o conjunto de calibração utilizou uma matriz  $\mathbf{X}$  contendo 30 espectros de MIR e uma matriz  $\mathbf{Y}$  com as três propriedades de interesse. Uma vez construído o modelo, o próprio programa PLS-Toolbox fornece as incertezas de previsão de cada modelo, gerando os erros médios de calibração (RMSEC), e faz a validação por *leave-one-out*, verificando a viabilidade do modelo proposto com as amostras de calibração, gerando os erros médios de validação-cruzada (RMSECV). Em seguida, o melhor modelo construído, aquele com o menor valor de RMSECV, é testado com as amostras do conjunto de validação, que contém 16 espectros de MIR também submetidos a pré-processamentos com MSC e MC. Este teste usando as amostras de validação é feito para verificar se realmente o modelo ajusta bem os dados através dos valores dos erros médios de predição (RMSEP). O programa PLS\_Toolbox forneceu os resultados estatísticos destes erros de calibração, validação-cruzada e predição, os quais são apresentados na Tabela 12.

Tabela 12 - Resultados estatísticos da regressão por PLS com 46 amostras fornecidos pelo programa PLS\_Toolbox para o conjunto de calibração e de validação.

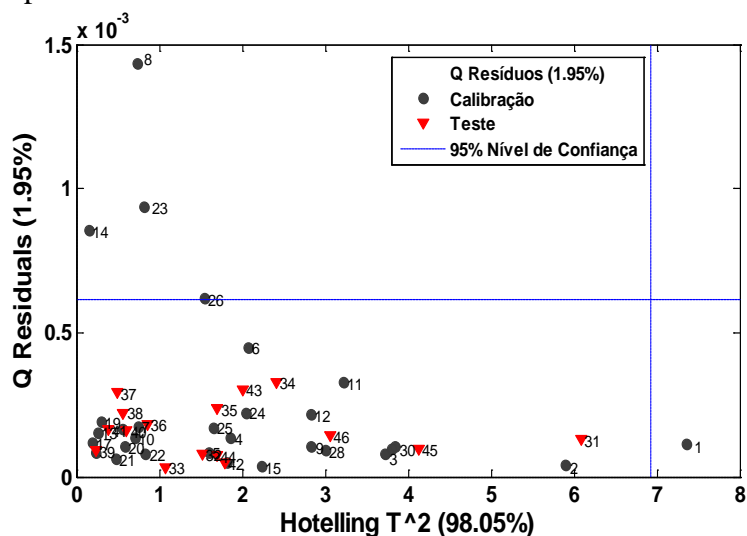
PLS2			
Amostras de Calibração e Validação	30 / 16		
Pré-processamento	MSC + MC		
Validação Cruzada	LOO		
Número de VL	2		
% Variância Capturada	97,37% de <b>X</b> e 97,72% de <b>Y</b>		
	Índice de Refração	Massa Específica	Teor de ésteres por HPLC
RMSEC	0,0002	0,0018	2,5313
RMSECV	0,0003	0,0021	2,9047
RMSEP	0,0002	0,0032	2,2589
R <sup>2</sup> Cal	0,9964	0,9721	0,9874
R <sup>2</sup> CV	0,9954	0,9632	0,9834
R <sup>2</sup> Pred	0,9959	0,9209	0,9899

Fonte: O autor, 2014.

O *software* utilizando duas variáveis latentes (VLs) construiu um modelo que explicou 97,37 % da variável **X** e 97,72 % da variável **Y** para a determinação do índice de refração, massa específica e teor de ésteres, apresentando o menor erro médio quadrático de calibração (RMSEC), validação cruzada (RMSECV) e de predição (RMSEP) para os três parâmetros. Foram calculados, também, os valores de coeficientes de determinação para os dados de calibração, validação cruzada e de predição.

Na Figura 17, o gráfico de  $T^2$  versus  $Q$  resíduos mostrou que em ambos os conjuntos de calibração e validação não há amostras atípicas, pois todas estão dentro do nível de confiança de 95% e a maior parte das amostras apresentam, ao mesmo tempo, baixos valores de resíduos e de *hotelling*. Ainda se observa que as amostras 8, 14 e 23 possuem resíduos altos, porém separadamente elas não influenciam no modelo devido ao baixo valor de  $T^2$ . Verifica-se que a quantidade de resíduos é de 1,95%, o que significa que 98,05% das informações foram explicadas com duas variáveis latentes na construção do modelo.

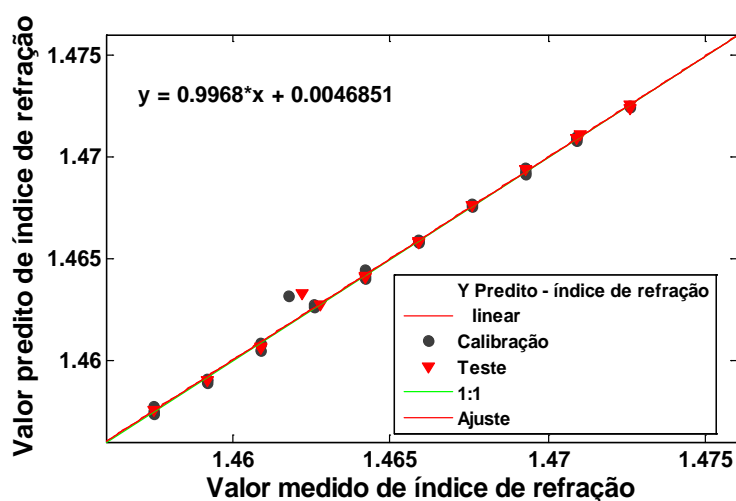
Figura 17 - Gráfico dos resíduos (*Hotelling T<sup>2</sup> versus Q* resíduos) para índice de refração, massa específica e teor de ésteres determinado por HPLC para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.



Fonte: O autor, 2014.

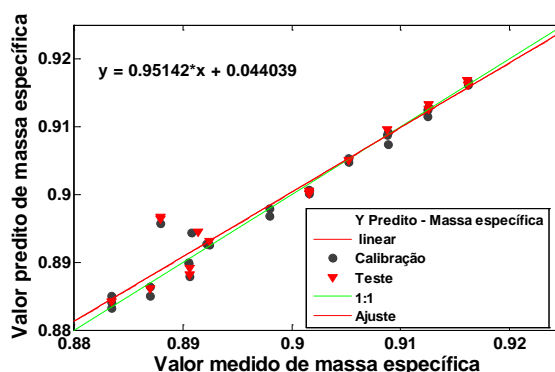
As Figuras 18 e 19 correspondem às curvas de regressão que mostram os valores preditos pelo modelo PLS2 em função dos valores medidos para o índice de refração e massa específica, respectivamente. Pode ser observado que, nas duas curvas de regressão, todos os pontos estão distribuídos em torno da reta, não havendo desvios significativos.

Figura 18 - Curva de regressão utilizando PLS2 para os valores de índice de refração (preditos *versus* medidos) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.



Fonte: O autor, 2014.

Figura 19 - Curva de regressão utilizando PLS2 para os valores de massa específica (preditos *versus* medidos) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.

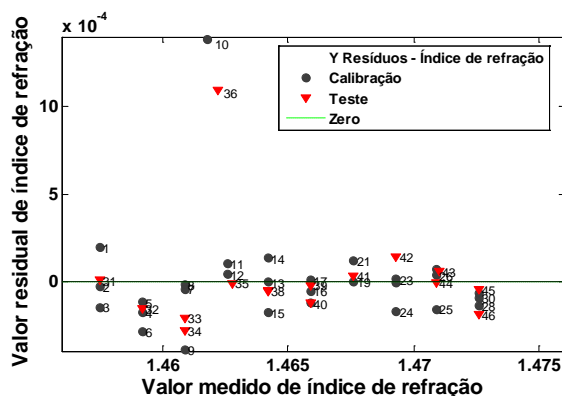


Fonte: O autor, 2014.

As Figuras 20 e 21 correspondem aos gráficos de resíduos em função dos valores ajustados (valores preditos) para índice de refração e massa específica, respectivamente. Este tipo de gráfico é a técnica visual mais utilizada para analisar o comportamento dos dados na regressão linear verificando as suposições dos resíduos para ambos os conjuntos de calibração e validação.

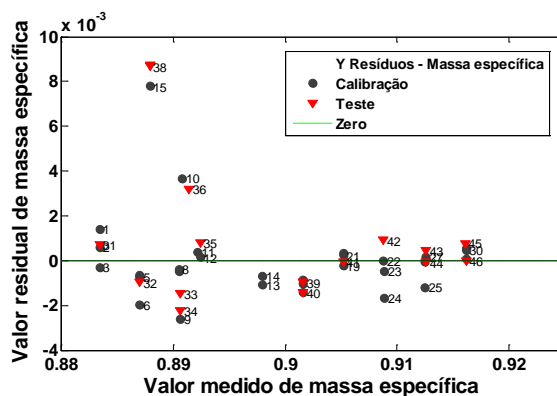
Os gráficos de resíduos, Figuras 20 e 21, mostram que não há erros significativos nas previsões do modelo PLS2 para os dois primeiros parâmetros estudados, porque os erros observados ficaram abaixo de 5%, ou seja, os resultados para o índice de refração e massa específica indicaram a pressuposição estatística para modelos de regressão no que se referem à aleatoriedade dos resíduos (homocedasticidade), todos com 95% de confiança.

Figura 20 - Gráfico de resíduos utilizando PLS2 para índice de refração dos valores (medidos *versus* residuais) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.



Fonte: O autor, 2014.

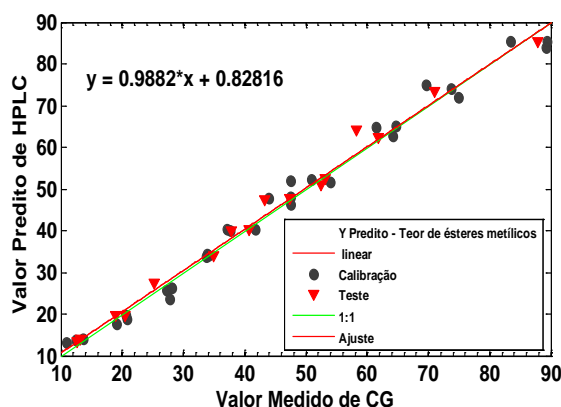
Figura 21 - Gráfico de resíduos utilizando PLS2 para massa específica dos valores (medidos *versus* residuais) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.



Fonte: O autor, 2014.

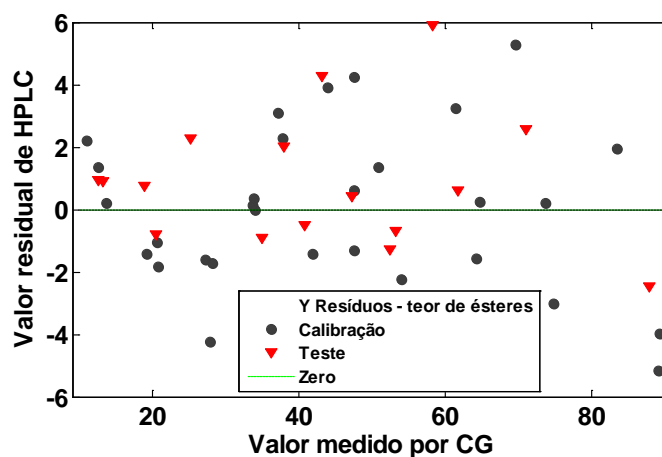
No entanto, no caso da curva de regressão que mostra os valores preditos pelo modelo PLS2 para o teor de ésteres determinado por HPLC, a qualidade do ajuste dos pontos não foi tão boa quanto à observada para o índice de refração e a massa específica. Como indicado na Figura 22, são observados desvios significativos de algumas amostras tanto para o conjunto de calibração como para o de validação. Tal afirmação pode ser corroborada pela não aleatoriedade dos dados residuais apresentando tendências no gráfico dos resíduos (Figura 23). Neste gráfico, pode se observar uma tendência nos erros superestimando o teor de ésteres para as amostras com maior teor de biodiesel, ou seja, a variância no teor de ésteres predito aumenta quando se aumenta o teor de biodiesel, indicando a presença de heterocedasticidade nos dados experimentais.

Figura 22 - Curva de regressão utilizando PLS2 para o teor de ésteres metílicos determinado por HPLC dos valores (preditos *versus* medidos) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.



Fonte: O autor, 2014.

Figura 23 - Gráfico de resíduos utilizando PLS2 para o teor de ésteres metílicos determinado por HPLC dos valores (preditos *versus* residuais) para os conjuntos de calibração e validação contendo 30 e 16 amostras de MIR, respectivamente.



Fonte: O autor, 2014.

### 3.7. Regressão por mínimos quadrados ordinários (OLS)

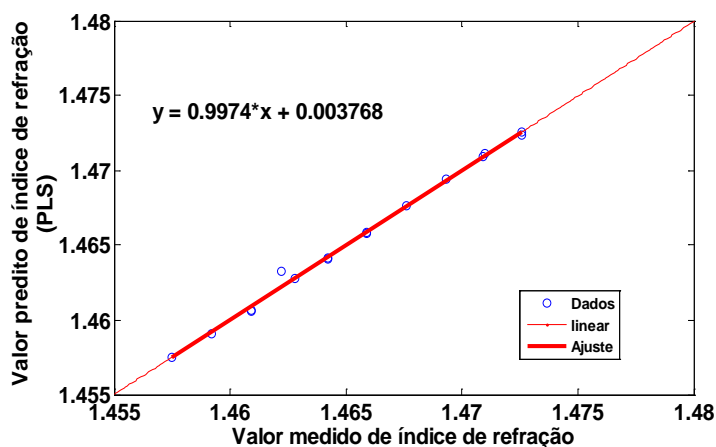
Para obter mais informações sobre a exatidão do modelo PLS2 proposto anteriormente, utilizou-se o método de regressão linear por mínimos quadrados ordinários (OLS) que é avaliado através da comparação dos valores obtidos pelo método proposto (o método usado na modelagem) com os valores medidos (método analítico). Esta metodologia verifica se os coeficientes são estatisticamente iguais ou diferentes da estimativa ideal 1 e 0, respectivamente através da região conjunta da elipse de confiança (EJCR). Além disso, o método por OLS calcula um menor valor possível de erros apresentando uma distribuição normal com média zero e variância constante.

Os resultados fornecidos pela regressão OLS informam os valores dos coeficientes através da curva de regressão encontrada para os três parâmetros de interesse (índice de refração, massa específica e teor de ésteres determinado por HPLC), os gráficos de erros absolutos e a região conjunta da elipse de confiança.

Nas Figuras 24 e 25 podemos observar as curvas de regressão por OLS para o conjunto de validação constituído por 16 amostras para o índice de refração e massa específica. Estes gráficos informam a qualidade de ajuste do modelo representado pelos valores de referência em função dos valores preditos pelo modelo PLS2 que fornece a equação da reta e os valores dos coeficientes lineares e angulares para ambas as propriedades.

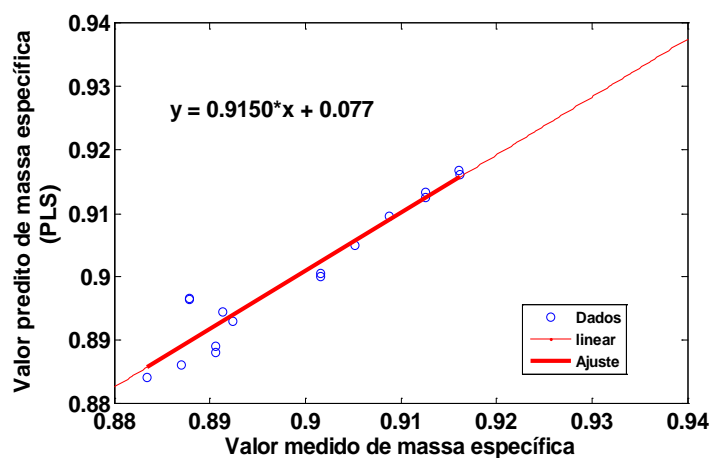


Figura 24 - Curva de regressão utilizando OLS para os valores de índice de refração (preditos pelo método PLS2 *versus* medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.



Fonte: O autor, 2014.

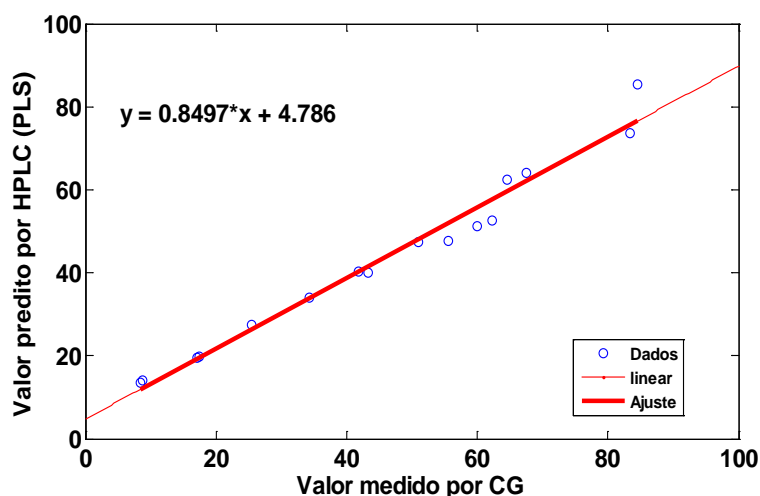
Figura 25 - Curva de regressão utilizando OLS para os valores de massa específica (preditos pelo método PLS2 *versus* medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.



Fonte: O autor, 2014.

Já na curva de regressão por OLS para o teor de ésteres determinado por HPLC, Figura 26, pode ser observado que os dados experimentais não foram tão bem ajustados pelo modelo apresentando valores de coeficientes bem distantes do ponto ideal (1 e 0).

Figura 26 - Curva de regressão utilizando OLS para os valores de teor de ésteres determinados por HPLC (preditos pelo método PLS2 *versus* medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.



Fonte: O autor, 2014.

Na Tabela 13 podem ser observados os valores dos coeficientes angulares e lineares e seus desvios associados às curvas de regressão para as propriedades de interesse (índice de refração, massa específica e teor de ésteres metílicos) utilizando OLS mostradas nas Figuras 24, 25 e 26.

Tabela 13 - Valores dos coeficientes de inclinação e intercepto para as propriedades estudadas.

Propriedades	Parâmetros da Regressão	
	Inclinação	Intercepto
Índice de Refração	0,9974 ( $\pm 0,0171$ )	0,0038 ( $\pm 0,0251$ )
Massa Específica	0,9150 ( $\pm 0,0716$ )	0,0774 ( $\pm 0,0644$ )
Teor de ésteres por HPLC	0,8497 ( $\pm 0,0365$ )	4,7855 ( $\pm 1,8716$ )

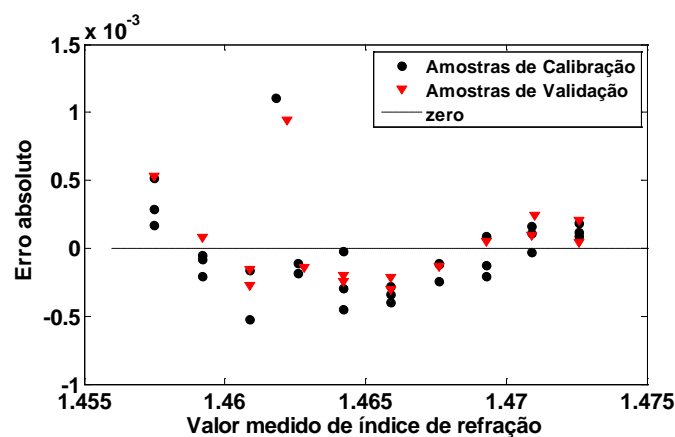
Fonte: O autor, 2014.

Os resultados dos gráficos de erros absolutos podem ser analisados através da inspeção visual, da mesma maneira que foram avaliados os gráficos de resíduos para verificar a aleatoriedade nos erros. Neste caso, podemos observar as semelhanças na distribuição dos erros absolutos pela regressão OLS, Figuras 27, 28 e 29, para índice de refração, massa específica e teor de ésteres por HPLC, respectivamente, quando

comparados com os gráficos de resíduos do modelo PLS2 (Figuras 20, 21 e 23). Porém, os gráficos de erros absolutos não são suficientes para garantir a homocedasticidade nos dados.

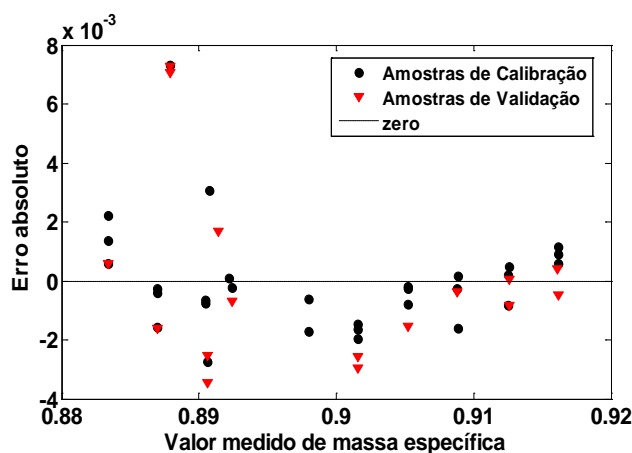
Nas Figuras 27 e 28 são apresentados os gráficos de erros absolutos do índice de refração e massa específica, respectivamente, obtidos pela regressão linear por mínimos quadrados ordinários (OLS). Pode ser observada uma distribuição aleatória dos dados semelhante a apresentada pelos gráficos de resíduos das mesmas propriedades obtidos por mínimos quadrados parciais (PLS2) (Figuras 20 e 21).

Figura 27 - Gráfico de erros absolutos utilizando OLS para índice de refração dos valores (residuais *versus* medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.



Fonte: O autor, 2014.

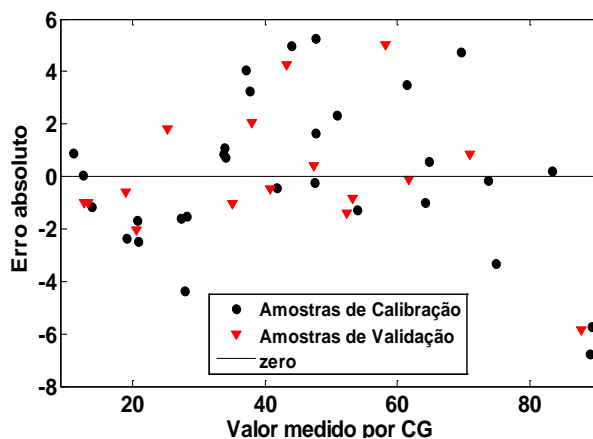
Figura 28 - Gráfico de erros absolutos utilizando OLS para massa específica dos valores (residuais *versus* medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.



Fonte: O autor, 2014.

Na Figura 29, o gráfico de erros absolutos para o teor de ésteres determinado por HPLC mostra uma tendência semelhante à observada pelos resultados mostrados no gráfico de resíduos obtido pelo modelo PLS2 (Figura 23). Como a tendência nestes dados foi mantida tanto no gráfico de resíduo por PLS2 como no gráfico de erros absolutos por OLS (Figura 29), isto evidencia a presença de heterocedasticidade, ou seja, não há uma distribuição supostamente normal com média zero e variância constante na variável aleatória U (erro). Tal afirmação pode ser mais bem observada analisando a região conjunta de elipse de confiança (EJCR).

Figura 29 - Gráfico de erros absolutos utilizando OLS para o teor de ésteres metílicos por HPLC dos valores (residuais *versus* medidos de referência) para o conjunto de validação contendo 16 amostras de MIR.



Fonte: O autor, 2014.

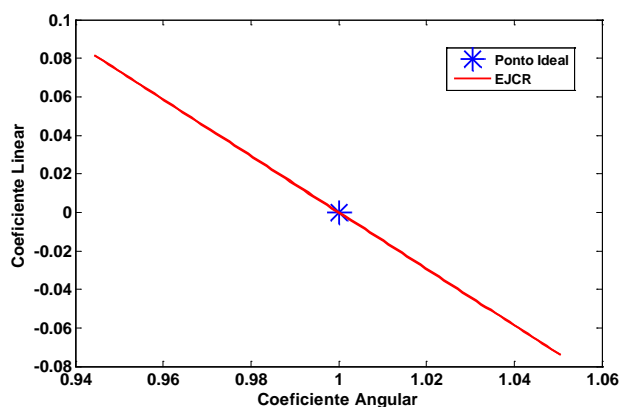
A exatidão do modelo proposto por PLS2 pode ser comparada por meio dos valores obtidos para a inclinação e o intercepto de uma reta ajustada por OLS e isto pode ser mais bem observado através da região conjunta de elipse de confiança (EJCR). Este intervalo conjunto (elipse de confiança) informa uma estimativa simultânea dos valores verdadeiros para os parâmetros da regressão, podendo prever novas observações com 95% de confiança nos resultados.

As Figuras 30, 31 e 32 mostram as regiões de elipse de confiança para índice de refração, massa específica e teor de ésteres determinado por HPLC, respectivamente. Nos dois primeiros gráficos, Figuras 30 e 31, pode-se observar que o ponto ideal se encontra dentro do intervalo conjunto, este por sua vez apresenta um intervalo de previsão bem estreito indicando uma maior precisão na estimação dos parâmetros. Além

disso, é um indicativo de ausência de erros sistemáticos nos dados de índice de refração e massa específica, pois os intervalos de previsão estimados simultaneamente para os parâmetros da regressão de tais propriedades contem os valores ideais de 1 e 0 para inclinação e intercepto, respectivamente.

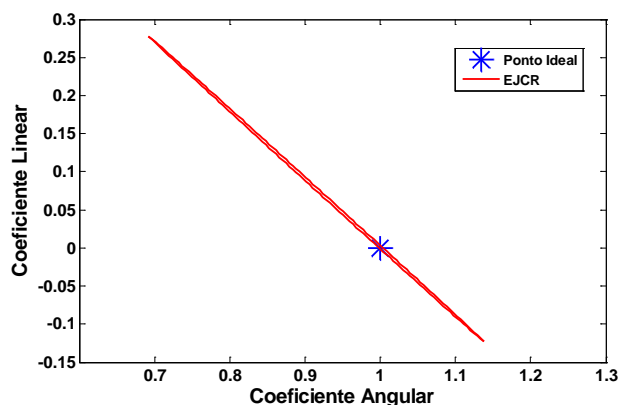
Os resultados da EJCR confirmam alguns indicativos obtidos para a adequação do modelo por PLS2 para o índice de refração e massa específica através do bom ajuste dos dados experimentais pela reta de regressão por OLS, a semelhança na aleatoriedade dos resíduos através da inspeção visual nos gráficos de erros absolutos e nos gráficos de resíduos.

Figura 30 - Gráfico da região conjunta da elipse de confiança para inclinação e intercepto da regressão de valores (preditos *versus* valores de referência) usando mínimos quadrados ordinários (OLS) para os dados de índice de refração para o conjunto de validação.



Fonte: O autor, 2014.

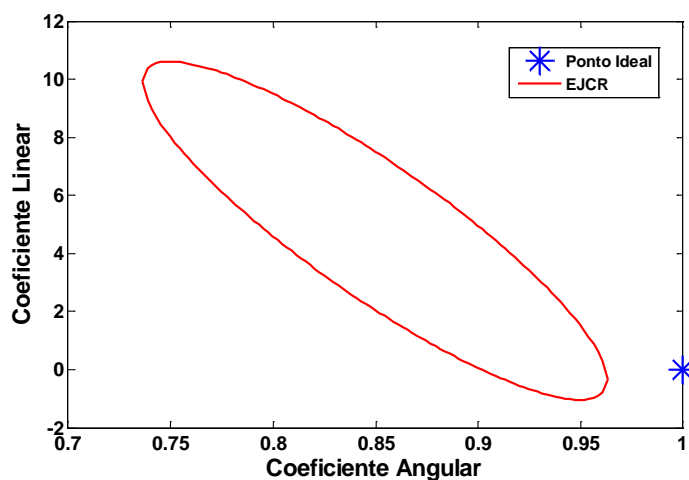
Figura 31 - Gráfico da região conjunta da elipse de confiança para inclinação e intercepto da regressão de valores preditos *versus* valores de referência usando mínimos quadrados ordinários para os dados de massa específica para o conjunto de validação.



Fonte: O autor, 2014.

Na Figura 32 podemos observar que o ponto ideal não se encontra dentro do intervalo conjunto. Este por sua vez apresenta um intervalo de previsão mais largo, ou seja, menor precisão para a estimativa dos parâmetros de regressão, indicando a presença de erros sistemáticos nos dados de teor de ésteres metílicos determinados por HPLC, pois o intervalo de previsão estimado simultaneamente para os parâmetros da regressão não contem os valores ideais de 1 e 0 para inclinação e intercepto, respectivamente. O resultado da EJCR para o teor de ésteres metílicos por HPLC explica a tendência observada no gráfico de resíduos por PLS2 e no gráfico de erros absolutos, indicando que o modelo por PLS2 prevê os valores de teor de ésteres metílicos com um nível de confiança inferior àquele estabelecido para a massa específica e o índice de refração.

Figura 32 - Gráfico da região conjunta da elipse de confiança para inclinação e intercepto da regressão de valores (preditos *versus* valores de referência) usando mínimos quadrados ordinários (OLS) para os dados de ésteres metílicos por HPLC para o conjunto de validação.



Fonte: O autor, 2014.

## CONCLUSÃO

O modelo MIR-PLS desenvolvido mostrou-se capaz de prever as propriedades índice de refração e massa específica devido à aleatoriedade nos valores residuais referindo-se à premissa de homocedasticidade, e pelos resultados satisfatórios obtidos para todos os valores estimados das figuras de mérito. Além disto, a validação do modelo proposto foi confirmada pela região conjunta de elipse de confiança, na qual pode ser observada uma ausência de erros sistemáticos. Os limites de confiança estimados pela regressão usando mínimos quadrados ordinários (OLS), para ambas as propriedades, contêm os valores esperados de 1 e 0 para coeficiente angular e linear, respectivamente. Com isso, podemos concluir que o modelo apresenta uma boa capacidade preditiva para a massa específica e o índice de refração, com 95% de confiança, a partir dos espectros de MIR-ATR.

Porém, para o teor de ésteres metílicos medido por HPLC, o modelo desenvolvido, apesar dos resultados satisfatórios obtidos para todos os valores estimados das figuras de mérito, não foi capaz de ajustar esses dados experimentais, apresentando uma tendência nos valores residuais (gráficos de erros absolutos), a qual indicou a ausência de homocedasticidade. A presença da heterocedasticidade nos dados pode ser corroborada, também, através da análise da região conjunta de elipse de confiança em que os dados de HPLC apresentaram erros sistemáticos. Os valores esperados de 1 e 0 para os coeficientes angular e linear, respectivamente, encontraram-se fora do intervalo de confiança estimado para os coeficientes de regressão usando mínimos quadrados ordinários (OLS). Sendo assim, conclui-se que a capacidade preditiva do modelo para o teor de ésteres metílicos foi inferior à observada para a massa específica e o índice de refração.

O uso da cromatografia líquida de alta eficiência (HPLC) para a determinação do teor de ésteres metílicos mostrou-se promissor para determinar por uma única metodologia tanto os teores de ésteres metílicos, bem como os teores de mono-, di- e triglicerídeos. Neste trabalho foi possível demonstrar que os resultados obtidos de ésteres metílicos pelo método analítico HPLC pode ser modelado e correlacioná-los com espectros de infravermelho utilizando calibração multivariada. Com isso um modelo que ajuste bem os dados pode ser capaz de prever novas observações, ou seja, novas amostras com valores de ésteres metílicos desconhecidos.

## SUGESTÕES

Aprimorar a metodologia de determinação da composição química das amostras de biodiesel por HPLC, dado o potencial de técnica de fornecer resultados confiáveis com menores tempos e custos analíticos quando comparada á técnica de cromatografia em fase gasosa com a derivatização da amostra.

Preparar um número maior de misturas modelo biodiesel de soja/óleo de soja com diferentes composições, de modo a tornar possível a separação das mesmas em três conjuntos: o de calibração, o de validação e o de teste.

Avaliar a viabilidade da utilização da metodologia para a certificação de amostras de biodiesel, preparando o mesmo número de misturas, porém limitando sua composição na faixa entre 90 e 100 % em massa de biodiesel.



**REFERÊNCIAS**

ALMEIDA, et. al. Caracterizações físico-químicas de óleos vegetais utilizados para produção de biodiesel com metodologias alternativas simples. In. XXXI Encontro Nacional de Engenharia de Produção. 2011, Belo Horizonte, MG.

ALVES, J. C. L.; POPPI, R. J. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). *Talanta*, v. 104, p. 155 - 161, 2012.

ANDRADE, D. F. *Caracterização de Biodiesel através da separação dos constituintes por Cromatografia Líquida*. 2011. 166f. Tese (Doutorado em Tecnologia de Processos Químicos e Bioquímicos) – Escola de Química, Universidade Federal do Rio de Janeiro, 2011.

ANDRIES, J. P. M., et. al. Strategy for reduced calibration sets to develop quantitative structure–retention relationships in high-performance liquid chromatography. *Analytica Chimica Acta*, v. 652, p. 180-188, 2009.

ANDRIOTTI, J. L. S. Técnicas estatísticas aplicáveis a tratamento de informações oriundas de procedimentos laboratoriais. Porto Alegre: CPRM, 2005.

ASHNANI, M. H. M., et. al. A source of renewable energy in Malaysia, why biodiesel? *Renewable and Sustainable Energy Reviews*, v. 35, p.244-257, 2014.

AYTAV, E.; KOCAR, G. Biodiesel from the perspective of Turkey: Past, present and future. *Renewable and Sustainable Energy Reviews*, v. 25, p. 335-350, 2013.

BASSBASI, M. et. al. FTIR-ATR determination of solid non fat (SNF) in raw milk using PLS and SVM chemometric methods. *Food Chemistry*, v. 146, p. 250-254, 2014.

BATTISTI, A. P., et. al. Determinação da rendimento do óleo em biodiesel por RMN<sup>1</sup>H e HPLC na reação de transesterificação. In. VIII Congresso Brasileiro de Engenharia Química em Iniciação Científica. 2009, Uberlândia, MG.

BOTELHO, B. G.; MENDES, B. A. P.; SENA, M. M. Implementação de um método robusto para o controle fiscal de umidade em queijo minas artesanal abordagem metrológica multivariada. *Química Nova*, v. 36, n. 9, p. 1416-1422, 2013.

BRAGA, J. W. B.; POPPI, R. J. Validação de modelos de calibração multivariada: uma aplicação na determinação de pureza polimórfica de carbamazepina por espectroscopia no infravermelho próximo. *Química Nova*, v. 27, n. 6, p. 1004-1011, 2004.

BRERETON, R. G. Chemometrics: Data Analysis for the Laboratory and Chemical Plant. University of Bristol, UK: John Wiley & Sons, 2003.

CARVALHO, M. J. S. *Quimiometria e espectroscopia no infravermelho próximo (NIR) e médio (MIR) aplicadas na análise de óleos lubrificantes*. 2005. 124f. Dissertação

(Mestrado em Ciências em Químicas), Instituto Militar de Engenharia, Rio de Janeiro, 2005.

CÔCCO, L. C. *Previsão de propriedades físico-químicas e composição química da gasolina a partir de espectros no infravermelho*. 2008. 873f. Tese (Doutorado em Engenharia de Processos Térmicos e Químicos), Universidade Federal do Paraná, Curitiba, 2008.

CORGOZINHO, C. N. C. *Utilização de técnicas espectrofluorimétricas e quimiométricas na análise de biodiesel e de suas misturas com óleo diesel*. 2009. 164 f. Tese (Doutorado em Ciências) – Instituto de Ciências Exatas, Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

CORREA, S. M. B. B. *Probabilidade Estatística*. 2. ed. Belo Horizonte: PUC Minas Virtual, 2003.

CUNHA, S. C.; OLIVEIRA, M. B. P. P. Discrimination of vegetable oils by triacylglycerols evaluation of profile using HPLC/ELSD. *Food Chemistry*, v. 95, p. 518-524, 2006.

ENERGIA. Panorama energético atual e perspectivas futuras. Disponível em <<http://www.biodieselbr.com/energia/alternativa/agro-energia.htm>> Acesso em 3 de Maio de 2014.

FARAH, M. A.; ARAÚJO, M. A. S.; PINTO, R. R. C. *Biodiesel*. Rio de Janeiro: Universidade Petrobras, 2006.

FERREIRA, M. M. C. et. al. Quimiometria 1: Calibração Multivariada, um tutorial. *Química Nova*, v.22, n.5, p.724-731, 1999.

FONSECA, F. S. *Espectrometria no infravermelho médio e calibração multivariada por PLS para a quantificação de biodiesel metílico e etílico de soja, milho e girassol em misturas com diesel*. 2011. 77f. Dissertação (Mestrado em Química) – Instituto de Química, Universidade Federal de Uberlândia, Uberlândia, 2011.

FRANCO, V.G. et. al. Teaching chemometrics with a bioprocess: Analytical Methods comparison using bivariate linear regression. *Chem. Educator*, v. 7, p. 265-269, 2002.

FREITAS, S. M. *Metodologia estatística para validação de métodos analíticos aplicável à metrologia em Química*. 2003. 106f. Dissertação (Mestrado em Metrologia para Qualidade Industrial) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2003.

GROOT, P. J., et. al. Selecting a representative training set for the classification of demolition waste using remote NIR sensing. *Analytica Chimica Acta*, v. 392, p. 67-75, 1999.

GUJARATI, D. N. *Econometria Básica*. 4. ed. Rio de Janeiro: Elsevier, 2006.

IUPAC. *Protocol for the design, conduct and interpretation of method-performance studies*. [S.l.], 1988.

KIM, N. The limit distribution of a modified Shapiro–Wilk statistic for normality to type II censored data. *Journal of the Korean Statistical Society*, v. 40, p. 257-266, 2011.

KNOTHE, et. al., *Manual de Biodiesel*. 1<sup>o</sup>ed. São Paulo: Blucher, 2006.

KNOTHE, G. Dependence of biodiesel fuel properties on the structure of fatty acid alkyl esters. *Fuel Processing Technology*, v. 86, p. 1059-1070, 2005.

KNOTHE, G. Determining the blend level of mixtures of biodiesel with conventional diesel fuel by fiber-optic near-infrared spectroscopy and <sup>1</sup>H nuclear magnetic resonance spectroscopy. *Jaocs*, v. 78, n. 10, 2001.

KNOTHE, G.; GERPEN, J. V.; KRAHL, J. *The Biodiesel Handbook*. Champaign: AOCS Publishing, 2005.

KNOTHE, G.; STEIDLEY, K. R. Kinematic viscosity of biodiesel fuel components and related compounds. Influence of compound structure and comparison to petrodiesel fuel components. *Fuel*, v. 84, p.1059-1065, 2005.

KNUZ, M. R., et. al. Impact of standardization sample design on tikhonov regularization variants for spectroscopic calibration maintenance and transfer. *J. Chemometrics*, v. 24, p. 218-229, 2010.

KORZENOWSKI, A. L. *Premissas e suposições para construção de gráficos de controle: um framework para verificação*. 2009. 102 f. Dissertação (Engenharia de Produção) – Sistemas de Qualidade, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.

LIRA, L. F. B. et. al., Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*, v. 89, p. 405-409, 2010.

LÔBO, I. V.; FERREIRA, S. L. C.; CRUZ, R. S. Biodiesel: parâmetros de qualidade e métodos analíticos. *Química Nova*, v. 32, n. 6, 1596-1608, 2009.

MADRAS, G.; KOLLURU, C.; KUMAR, R. Synthesis of biodiesel in supercritical fluids. *Fuel*, v. 83, p. 2029-2033, 2004.

MARQUES, et. al. Metodologia analítica para glicerol em biodiesel. *Cadernos de Pesquisa*, São Luís, v. 18, n. especial, dez. 2011

MARTINS, J. P. FERREIRA, M. M. QSAR modeling: um novo pacote computacional open source para gerar e validar modelos QSAR. *Química Nova*, v. 36, n. 4, p. 554-560, 2013.

MILLER, J. N.; MILLER, J. C. *Statistics and chemometrics for analysis chemistry*. 5. ed. [S. l.]: Pearson, 2005.

Ministério de Minas e Energia (MME) - Boletim Mensal dos Combustíveis Renováveis, 06/2014.

MOITA NETO, J. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. *Química Nova*, v. 21, n.4, 1998.

MONTGOMERY, D. C.; RUNGER, G. C. Estatística aplicada e probabilidade para engenheiros. 5. ed. Rio de Janeiro: LTC, 2012.

MORGANO, M. A. et. al. Determinação de proteína em café cru por espectroscopia NIR e regressão PLS. *Ciências e Tecnologia em Alimentos*, Campinas, v. 25, n. 1, p. 25-31, jan./mar. 2005.

NETO, B. B.; PIMENTEL, M. F.; ARAÚJO, M. C. U. Recomendações para calibração em química analítica - Parte 1. Fundamentos e calibração com um componente (calibração univariada). *Química Nova*, v. 25, n. 5, p. 856-865, 2002.

NORMANDO, D.; TJaDERHANE, L.; QUINTÃO, C. C. A. A escolha do teste estatístico – um tutorial em forma de apresentação em PowerPoint. *Dental Press Journal of Orthodontics*, v. 15, n. 1, p. 101-106, jan./fev 2010.

O QUE É BIODIESEL? Disponível em <<http://www.biodieselbr.com/biodiesel/definicao/o-que-e-biodiesel.htm>> Acesso em 3 de Maio de 2014.

PARREIRA. T. F. *Utilização de métodos quimiométricos em dados de natureza multivariada*. 2003. 106f. Dissertação (Mestrado em Química) – Departamento de Físico-Química, Universidade Estadual de Campinas, Campinas, SP, 2003.

PIMENTEL, M. F., et. al. Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. *Microchemical Journal*, v. 82, p. 201-206, 2006.

PONTES, A. C. F. *Análise de variância multivariada com a utilização de testes não-paramétricos e componentes principais baseados em matrizes de postos*. 2005. 117f. Tese (Doutorado em Agronomia) – Estatística e Experimentação Agronômica, Universidade de São Paulo, Piracicaba, 2005.

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO DE JANEIRO. (PUC-RIO). 3 *Modelagem da Curva de Calibração*. 22f. Disponível em: [http://www2.dbd.puc-rio.br/pergamum/tesesabertas/0116460\\_03\\_cap\\_03.pdf](http://www2.dbd.puc-rio.br/pergamum/tesesabertas/0116460_03_cap_03.pdf) Acesso em 7 de agosto de 2014.

Resolução ANP - Agência Nacional de Petróleo, Gás Natural e Biocombustíveis nº 14, de 11.5.2012 - DOU 18.5.2012. Disponível em [www.anp.gov.br](http://www.anp.gov.br) Acesso em junho de 2014.

ROCHA, W. F. C.; NOGUEIRA, R.; VAZ, B. G. Validation of model of multivariate calibration: an application to the determination of biodiesel blend levels in diesel by near-infrared spectroscopy. *Journal of Chemometrics*, v. 26, p. 456-461, 2012.

SANDRA, P., et. al. Characterization of triglycerides in vegetable oils by silver-ion packed-column supercritical fluid chromatography coupled to mass spectroscopy with atmospheric pressure chemical ionization and coordination ion spray. *Journal of Chromatography*, v. 974, p. 231-241, 2002.

SHAH, S.; SAHARMA, S.; GUPTA, M. N. Biodiesel Preparation by Lipase-Catalyzed Transesterification of Jatropha Oil. *Energy & Fuels*, v. 18, p. 154-159, 2004.

SCUDINO, P. A. *A utilização de alguns testes estatísticos para análise da variabilidade do preço do mel nos municípios de Angra dos Reis e Mangaratiba, estado do Rio de Janeiro*. 2008. 51 f. Trabalho de Conclusão de Curso (Licenciado e Bacharel em Matemática) – Instituto de Ciências Exatas, Universidade Federal Rural do Rio de Janeiro, Seropédica, 2008.

SENA, M. M. et. al. Avaliação do uso de métodos quimiométricos em análise de solos. *Química Nova*, v. 23, n. 4, p.547 - 556, 2000.

SILVERSTEIN, R. M.; WEBSTER, F. X.; Identificação Espectrofotométrica de Compostos Orgânicos, LTC - Livros Técnicos e Científicos Editora S.A., Rio de Janeiro, 6<sup>o</sup> edição, 2000.

SKOOG, et al. Fundamentos de Química Analítica, tradução da 8. ed. São Paulo: Cengage Learnin, 2009.

SOUZA, A. M., et. al. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. *Química Nova*, v. 36, n. 7, p. 1057-1065, 2013.

SOUZA, L. M., et. al. Quantification of residual automotive lubricant oil as an adulterant in Brazilian S-10 diesel using MIR spectroscopy and PLS. *Fuel*, v. 130, p. 257-262, 2014.

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

SOUZA, A. M., et. al. Proposta de tutorial de quimiometria utilizando técnicas modernas para a análise de solos. In: VI Simposio Brasileiro de Educacao em Solos, 2012. Sobral, CE: *Revista Homem, Espaço e Tempo*.

STAVARACHE, C. et. al. Fatty acid methyl esters from vegetable oil by means of ultrasonic energy. *Ultrasonics Sonochemistry*, v. 12, p. 367-372, 2005.

TEÓFILO, R. F. Métodos Quimiométricos: Uma Visão Geral - Conceitos básicos de quimiometria. 2013. 118f. Universidade Federal de Viçosa, Viçosa, 2013.

VALDERRAMA, P. *Avaliação de figuras de mérito em calibração multivariada na determinação de parâmetros de controle de qualidade em indústria alcooleira por*

*espectroscopia no infravermelho próximo*. 2005. 135f. Dissertação de Mestrado, Universidade Estadual de Campinas, SP, 2005.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. *Química Nova*, v. 32, n. 5, p. 1278-1287, 2009.

VALDERRAMA, et. al. Variable selection, outlier detection, and figures of merit estimation in a partial least-square regression multivariate calibration model. A case study for the determination of quality parameters in the Alcohol Industry by Near-Infrared Spectroscopy. *Journal of Agricultural and Food Chemistry*, v. 55, p. 8331-8338, 2007.

VALDEZ, H. C., et. al. Determinação de glicerol livre e total em amostras de biodiesel por método enzimático com detecção colorimétrica. *Química Nova*, v. 35, n. 3, p. 601-607, 2012.

VASCONCELOS, F. V. C. Using near-infrared overtone regions to determine biodiesel content and adulteration of diesel/biodiesel blends with vegetable oils. *Analytica Chimica Acta*, v. 716, p. 101-107, 2012.

VASCONCELOS, F. V. C. *Uso da região espectral de sobretons para determinação do teor de biodiesel e classificação de misturas diesel/biodiesel adulteradas com óleo vegetal*. 2011, 112f. Dissertação (Mestrado em Química) – Universidade Federal da Paraíba, João Pessoa, 2011.

WAYNICK, J. A. Characterization of biodiesel oxidation and oxidation products. Technical Literature Review – National Renewable Energy Laboratory (NREL)/TP-540-39096, 2005.

WEHRENS, R. *Chemometrics with R: Multivariate data analysis in the natural sciences and life sciences*. London New York : *Springer Heidelberg Dordrecht*, 2011.

WISE, M.B., et. al. 2006....Chemometrics Tutorial for PLS\_Toolbox and Solo.

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

ZAGONEL, G. F.; ZAMORA, P. P.; RAMOS, L. P. Multivariate monitoring of soybean oil ethanolysis by FTIR. *Talanta*, v. 63, p. 1021-1025, 2004.

## APÊNDICE A – Especificações para o biodiesel no Brasil

Tabela 1 - Algumas especificações para o biodiesel no Brasil

Característica	Unidade	Limite	Método		
			ABNT NBR	ASTM D	EN/ISO
Aspecto	-	LII (1)	-	-	-
Massa específica a 20°C	kg/m <sup>3</sup>	850 a 900	7148	1298	EN ISO 3675
			14065	4052	-
					EN ISO 12185
Viscosidade Cinemática a 40°C	mm <sup>2</sup> /s	3,0 a 6,0	10441	445	EN ISO 3104
Teor de água, máx.	mg/kg	(2)	-	6304	EN ISO 12937
Contaminação Total, máx.	mg/kg	24	-	-	EN ISO 12662
					NBR 15995
Ponto de fulgor, mín. (3)	°C	100,0	14598	93	EN ISO 3679
Teor de éster, mín	% massa	96,5	15764	-	EN 14103
Resíduo de carbono, máx. (4)	% massa	0,050	15586	4530	-

---

Cinzas sulfatadas, máx.	% massa	0,020	6294	874	EN ISO 3987
Enxofre total, máx.	mg/kg	10	15867	5453	EN ISO 20846
					EN ISO 20884
Sódio + Potássio, máx.	mg/kg	5	15554		EN 14108
			15555		EN 14109
			15553	-	EN 14538
			15556		
Cálcio + Magnésio, máx.	mg/kg	5	15553	-	EN 14538
			15556		
Fósforo, máx.	mg/kg	10	15553	4951	EN 14107
Corrosividade ao cobre, 3h a 50°C, máx.	-	1	14359	130	EN ISO 2160
Número Cetano (5)	-		-	613	EN ISO 5165
		Anotar		6890 (6)	
Ponto de entupimento de filtro a frio, máx.	°C	(7)	14747	6371	EN 116
Índice de acidez, máx.	mg KOH/g	0,50	14448	664	EN 14104 (8)
			-	-	

---



---

Glicerol livre, máx.	% massa	0,02	15344		
			15908	6584 (8)	EN 14105 (8)
				-	EN 14106 (8)
Glicerol total, máx. (9)	% massa	0,25	15344	6584 (8)	EN 14105 (8)
			15908	-	
Monoacilglicerol, máx.	% massa	0,80	15342 (8)		
			15344	6584 (8)	EN 14105 (8)
			15908		
Diacilglicerol, máx.	% massa	0,20	15342 (8)		
			15344	6584 (8)	EN 14105 (8)
			15908		
Triacilglicerol, máx.	% massa	0,20	15343		
				6584 (8)	EN 14105 (8)
Metanol e/ou Etanol, máx.	% massa	0,20	-	-	EN 14110 (8)
Índice de Iodo	g/100g	Anotar	-	-	EN 14111 (8)

---

Fonte: ANP, 2014

## APÊNDICE B - Figuras de mérito

Quando o procedimento analítico é proposto, é necessário investigar se o método apresenta um desempenho adequado nas condições em que o mesmo será aplicado. Para isto, é feita a validação através da determinação de figuras de mérito (ROCHA; NOGUEIRA; VAZ, 2012; VALDERRAMA; BRAGA; POPPI, 2009).

A validação é um método que estabelece estudos sistemáticos que atendem ao propósito e a determinadas especificações impostas por indústrias ou por algum órgão de fiscalização (VALDERRAMA; BRAGA; POPPI, 2009).

No presente trabalho, determinaram-se alguns parâmetros tais como, exatidão, sensibilidade, sensibilidade analítica e o inverso da sensibilidade analítica.

### 1. Exatidão

Os erros informam o grau de concordância entre o valor de referência e o valor encontrado pelo modelo de calibração considerado como uma estimativa de precisão. Em calibração multivariada, estes são geralmente expressos como a raiz quadrada do erro médio de calibração (RMSEC), a raiz quadrada do erro médio de validação cruzada (RMSECV) e raiz quadrada do erro médio de predição (RMSEP) (ROCHA; NOGUEIRA; VAZ, 2012; BOTELHO; MENDES; SENA, 2013; WISE, et. al., 2006).

Em todas as medidas de erros tenta-se estimar o desvio médio a partir dos dados. O erro quadrático médio de calibração (RMSEC) é uma medida de como o modelo ajusta os dados do conjunto de calibração, como visto na equação 14 (WISE, et. al., 2006).

$$\text{RMSEC} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (\text{Equação 14})$$

onde  $\hat{y}_i$  e  $y_i$  são os valores previstos pelo modelo de calibração e os valores medidos, respectivamente;  $n$  é o número todas as amostras do conjunto de calibração.

O erro quadrático médio de validação cruzada (RMSECV) mede a capacidade de previsão das amostras que não foram incluídas na construção do modelo, comumente utilizando o método (*leave-one-out*), de acordo com a equação 15.

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-1}} \quad (\text{Equação 15})$$

onde  $\hat{y}_i$  e  $y_i$  são os valores previstos pelo modelo de calibração e os valores medidos, respectivamente;  $n-1$  é o número de amostras do conjunto de calibração menos a amostra não incluída na construção do modelo

O melhor modelo escolhido é aquele com menor valor de RMSECV e alternativamente, pode-se designar um conjunto de amostras que têm valores de  $y_i$  conhecido para a previsão, de modo que tal avaliação pode ser feita diretamente, neste caso, pode-se calcular um erro quadrático médio de previsão (RMSEP), quando o modelo é aplicado a novos dados que é o conjunto de validação. A equação 16 está de acordo com BRERETON (2003).

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-a-1}} \quad (\text{Equação 16})$$

onde  $\hat{y}_i$  e  $y_i$  são os valores previstos pelo modelo de validação e os valores medidos, respectivamente;  $n$  é o número todas as amostras do conjunto de validação;  $a$  é o número de variáveis latentes.

## 2. Sensibilidade

A sensibilidade corresponde à fração do sinal responsável pelo acréscimo de uma unidade de concentração à propriedade de interesse e é determinada através da (Equação 17).

$$\hat{S\acute{E}N} = \frac{1}{\|b_k\|} \quad (\text{Equação 17})$$

em que,  $b_k$  é o vetor dos coeficientes de regressão estimados pelo PLS.

### 3. Sensibilidade analítica

A sensibilidade analítica apresenta a sensibilidade do método em termos de unidade de concentração que é utilizada, sendo definida como a razão entre a sensibilidade e o desvio-padrão do sinal de referência de acordo com a (Equação 18).

$$\gamma = \frac{s\hat{E}N}{|\delta x|} \quad (\text{Equação 18})$$

em que,  $s\hat{E}N$  é obtido através das equações 18 e  $\delta x$  é o desvio padrão do sinal de referência estimado através do desvio padrão do valor de NAS para 15 espectros do sinal de referência. O inverso desse parâmetro, ou seja,  $\gamma^{-1}$ , permite estabelecer a menor diferença de concentração entre amostras, que pode ser distinguida pelo método (VALDERRAMA; BRAGA; POPPI, 2009).

O sinal analítico líquido (NAS) é definido com a parte do sinal analítico exclusivamente relacionado com a propriedade de interesse e ortogonal ao espaço dos interferentes (BOTELHO; MENDES; SENA, 2013).

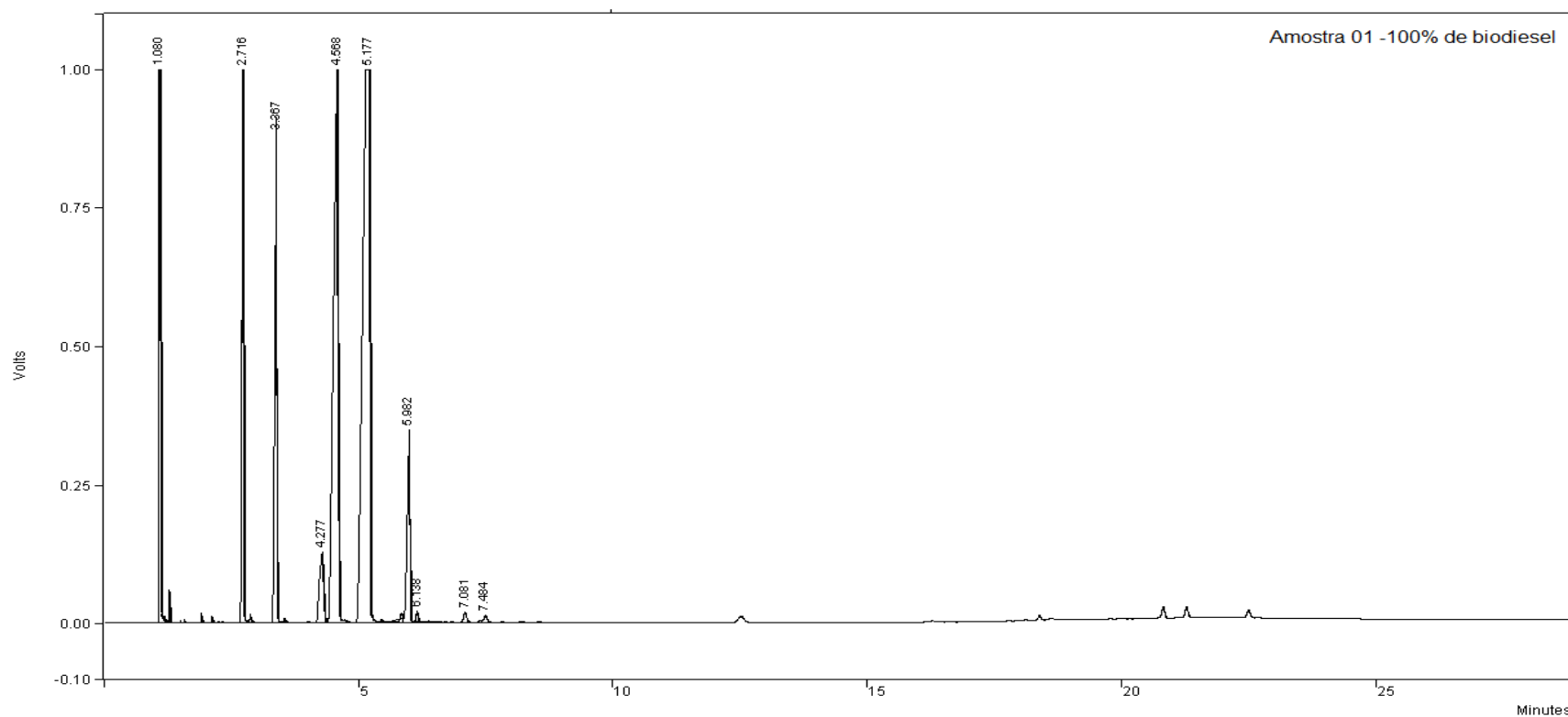
A exatidão do modelo foi avaliada através dos valores de RMSEC, RMSECV e RMSEP e os coeficientes de determinação. Os resultados obtidos na Tabela 14 demonstram uma baixa dispersão nos resultados dos erros e uma boa concordância entre si para as propriedades de índice de refração e massa específica indicando uma boa exatidão do modelo para tais propriedades, por conseguinte os resultados dos erros para o teor de ésteres metílicos apresentou grande dispersão, ou seja, o modelo prever com baixa exatidão tal parâmetro. O modelo apresentou um baixo valor de sensibilidade para todos os parâmetros estudados e este resultado que tem uma correlação direta com a concentração, mostrou que o modelo foi capaz de distinguir entre amostras com diferenças de concentrações abaixo de 0,1%.

Tabela 14 - Valores estimados para as Figuras de Mérito

Figuras de Mérito		Índice de Refração	Massa Específica	Teor de ésteres metílicos
Exatidão	RMSEC	0,0003	0,0018	2,5313
	RMSECV	0,0003	0,0021	2,9048
	RMSEP	0,0002	0,0033	2,2590
Sensibilidade		0,0917	0,0592	0,0938
Sensibilidade Analítica		1,0e+003 *1,0992	1,0e+003 *0,7098	1,0e+003 *1,1235
Sensibilidade Analítica <sup>-1</sup>		0,0009	0,0014	0,0009
	Coef. Det. (R <sup>2</sup> )	0,996475	0,972106	0,987465

Fonte: O autor, 2014.

## APÊNDICE C – Cromatograma por CG



Fonte: O autor, 2014.

Tempo de Retenção (minutos)	1.080	3.367	2.718	4.277	4.568	5.177	5.982
	Solvente	Padrão (Heptadecanoato de metila em n- Heptano)	C16:0 (Palmítico)	C18:0 (Esteárico)	C18:1 (Oleico)	C18:2 (Linoleico)	C18:3 (Linolênico)

Fonte: O autor, 2014.

**APÊNDICE D** - Cálculo para quantificar o teor de ésteres metílicos por CG

A determinação de ésteres metílicos por cromatografia gasosa calculada através da seguinte fórmula: sendo expressa em porcentagem.

$$C = \frac{(\Sigma A) - A_{EI}}{A_{EI}} \times \frac{C_{EI} \times V_{EI}}{M} \times 100$$

Onde:

$\Sigma A$  = área total de picos de ésteres metílicos

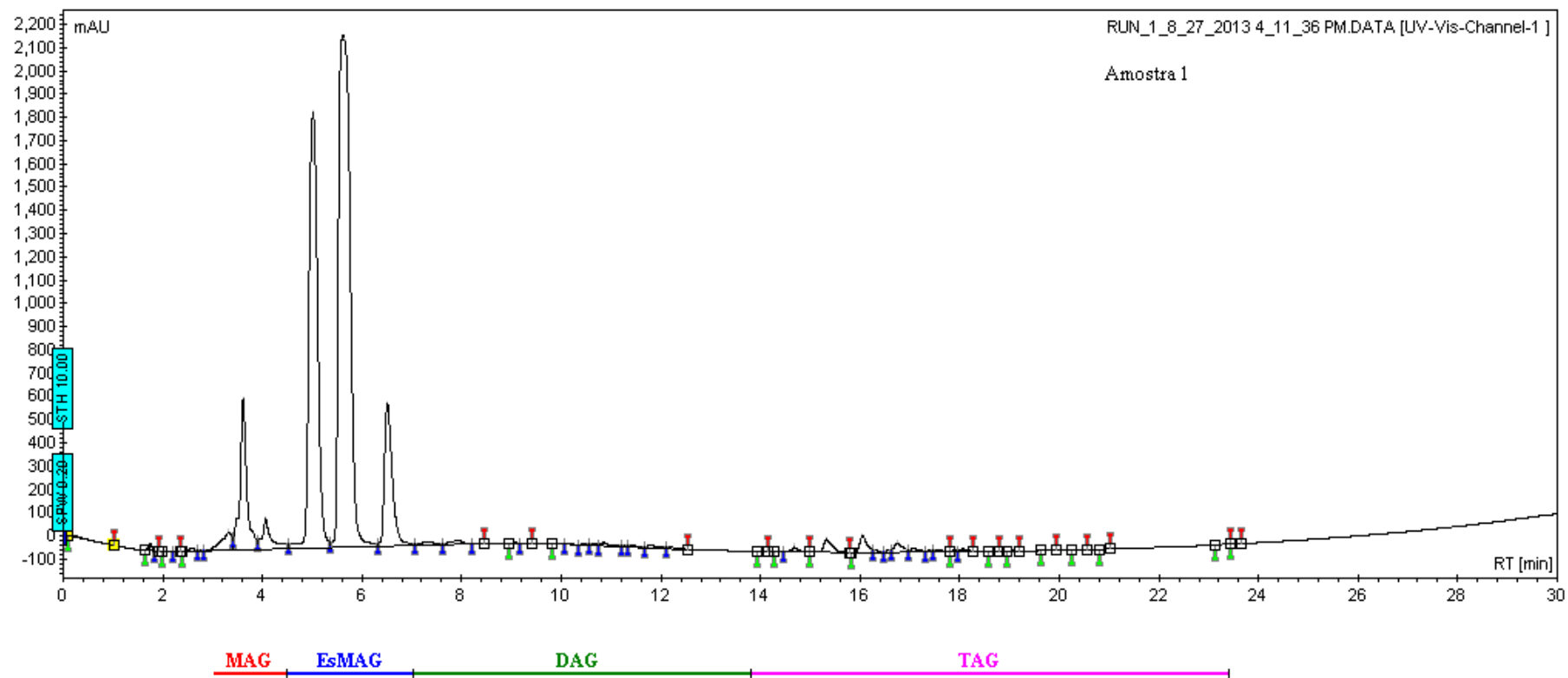
$A_{EI}$  = área do pico correspondente ao padrão ( heptadecanoato de metila)

$C_{EI}$  = concentração em miligramas/mililitros, da solução de padrão ( heptadecanoato de metila)

$V_{EI}$  = volume em mililitros da solução de padrão ( heptadecanoato de metila)

$M$  = massa da amostra em miligramas

## APÊNDICE E – Cromatograma por HPLC





## APENDICE F – Cálculo para quantificar o teor de ésteres metílicos por HPLC

A determinação de ésteres metílicos por cromatografia à líquido é primeiramente calculando o tempo de retenção relativo ( $t_{rr}$ ) que é definido como o tempo de retenção calculado em relação ao pico principal (ou de referência). O  $t_{rr}$  de cada componente identificado no cromatograma obtido por HPLC é calculado através da média de 20 determinações, segundo a equação 1:

$$t_{rr} = \frac{t_r}{t_p} \quad (\text{Equação 1})$$

Onde:  $t_r$  e  $t_p$  referem-se, respectivamente, aos tempos de retenção, em min, de cada componente e do pico de referência em cada grupo (isômeros sn-2 e sn-1 da monolinoleína para MAG; éster metílico do ácido linoléico para EsMAG; isômeros sn-1,2 e sn-1,3 da oleoil-linoleoil-glicerol para DAG e o par trilinoleína/oleoil-linoleoil-linolenoil-glicerol para TAG). Em seguida é calculada a área corrigida ( $A_c$ ) de cada componente identificado no cromatograma obtido por HPLC através da equação 2:

$$A_c = \frac{A}{NLD} \quad (\text{Equação 2})$$

Onde:

$A_c$  e  $A$  referem-se, respectivamente, as áreas, em mAU x min, corrigida e obtida diretamente de cada componente. As áreas corrigidas de cada classe de componentes (EsMAG, MAG, DAG e TAG) foram obtidas através da soma da  $A_c$  dentro de cada classe. A quantificação do teor de ésteres metílicos por HPLC considera os valores de área corrigida de cada classe encontrada na HPLC através da equação 3.

$$C_{HPLC} = 100 \times \frac{A_c \text{ EsMAG}}{A_c \text{ EsMAG} + A_c \text{ MAG} + (2 \times A_c \text{ DAG}) + (3 \times A_c \text{ TAG})} \quad (\text{Equação 3})$$

Para quantificar o teor de biodiesel por esta técnica, optou-se por somar as áreas obtidas nos cromatogramas referentes à faixa de componentes identificados como EsMAG.