



Universidade do Estado do Rio de Janeiro

Centro de Tecnologia e Ciências

Instituto de Química

Camilla Lima Cunha

**Estudo da previsão de propriedades do biodiesel utilizando
espectros de infravermelho e calibração multivariada**

Rio de Janeiro

2014

Camilla Lima Cunha

Estudo da previsão de propriedades do biodiesel utilizando espectros de infravermelho e calibração multivariada

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.

Orientadores: Prof. Dr. Aderval Severino Luna
Prof. Dr. Márcio Luis Lyra Paredes
Prof. Dr. Alexandre Rodrigues Torres

Rio de Janeiro

2014

CATALOGAÇÃO NA FONTE
UERJ/REDE SIRIUS/CTC/Q

C972 Cunha, Camilla Lima.
Estudo da previsão de propriedades do biodiesel utilizando espectros de infravermelho e calibração multivariada. / Camilla Lima Cunha. – 2014.
179 f.

Orientador: Aderval Severino Luna.
Márcio Luis Lyra Paredes.
Alexandre Rodrigues Torres.

Dissertação (mestrado) – Universidade do Estado do Rio de Janeiro, Instituto de Química.

1. Biodiesel – Teses. 2. Calibração multivariada – Teses. 3. Espectros de infravermelho – Teses. I. Luna, Aderval Severino. II. Paredes, Márcio Luis Lyra. III. Torres, Alexandre Rodrigues IV Universidade do Estado do Rio de Janeiro. Instituto de Química. V. Título.

CDU 662.756.3

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

Assinatura

Data

Camilla Lima Cunha

Estudo da previsão de propriedades do biodiesel utilizando espectros de infravermelho e calibração multivariada

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

Aprovada em 25 de fevereiro de 2014.

Banca Examinadora:

Prof. Dr. Aderval Severino Luna (Orientador)
Instituto de Química - UERJ

Prof. Dr. Márcio Luis Lyra Paredes (Orientador)
Instituto de Química - UERJ

Prof. Dr. Alexandre Rodrigues Torres (Orientador)
Faculdade de Tecnologia - UERJ

Prof. Dra. Lucia Regina Raddi de Araújo
Instituto de Química - UERJ

Prof. Dr. Ricardo Jorgensen Cassella
Instituto de Química - UFF

Dr. Werickson Fortunato de Carvalho Rocha
Divisão de Metrologia Química - INMETRO

Rio de Janeiro

2014

AGRADECIMENTOS

A Deus por ter me dado forças e iluminado meu caminho para que eu pudesse concluir mais essa realização na minha vida.

Aos meus familiares, em especial, a minha avó Maria Aparecida, minha mãe Sandra e a minha irmã Marianne, pelo apoio, incentivo e, exemplo de persistência e amor.

Aos Prof. Dr. Alexandre R. Torres, Prof. Dr. Aderval S. Luna e Prof. Dr. Márcio L. L. Paredes pela paciência e dedicação ao meu aprendizado e à realização desse trabalho.

Ao meu namorado Rafael pelo carinho, apoio, paciência e companheirismo durante todo o mestrado.

Aos meus sogros, Selma e Romualdo, por todo o apoio, carinho e paciência, a minha cunhada e amiga Gabriela pela paciência, pelos ensinamentos compartilhados e, principalmente, pela amizade.

Aos meus amigos do laboratório da FAT-UERJ, Luciana, Rejane, Érica, Mariana, Nilson, José Luís, Sônia, Wolney, Volnei, Sara, Carolina e Sr. Antônio pelas palavras de apoio, incentivo e por sempre me receberem de braços abertos.

Ao Prof. Dr. Sérgio Machado pelos conhecimentos compartilhados e análises de cromatografia à gás.

À Prof. Deborah Vargas e a Jéssica pela ajuda com o equipamento e acessórios de infravermelho.

Aos professores da FAT-UERJ, em especial, Prof. Dr^a. Fernanda e Prof. Dr. Leonardo que muito me incentivaram com palavras de apoio e perseverança.

Aos meus tios, tias, primos e avós de coração, em especial, tia Carmita, tia Núbia, Marco, tia Rose, tio Raimundo, Rayme, Cecília e Geraldo que sempre me receberam com abraços carinhosos e me acolheram durante este período.

Aos amigos, Gesiene Oliveira, Natali Verri, Isabela Meireles, Anne Rose, Alanna Pittas, Silas Romanha, Narã Vetter, Danilo Honda, que cada um, a sua maneira, foram e são pessoas importantes na minha vida, que mesmo longe sempre estarão presentes em minhas melhores lembranças e em minhas orações.

Aos alunos do Programa de Pós-graduação em Engenharia Química da UERJ, em especial, Igor, Arnaldo, Deivisson e Patrícia.

Ao corpo docente e aos funcionários do Programa de Pós-graduação em Engenharia Química da UERJ, especialmente ao grande amigo Marco pelo suporte e apoio.

À CAPES (Coordenação de Aperfeiçoamento Pessoal de Nível Superior) pela assistência financeira.

Ao CENPES (Centro de Pesquisas e Desenvolvimento Leopoldo A. Miguez de Mello) pela bolsa de projeto.

RESUMO

CUNHA, Camilla Lima. *Estudo da previsão de propriedades do biodiesel utilizando espectros de infravermelho e calibração multivariada*. 2014. 184f. Dissertação (Mestrado em Engenharia Química) – Instituto de Química – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2014.

O biodiesel tem sido amplamente utilizado como uma fonte de energia renovável, que contribui para a diminuição de demanda por diesel mineral. Portanto, existem várias propriedades que devem ser monitoradas, a fim de produzir e distribuir biodiesel com a qualidade exigida. Neste trabalho, as propriedades físicas do biodiesel, tais como massa específica, índice de refração e ponto de entupimento de filtro a frio foram medidas e associadas a espectrometria no infravermelho próximo (NIR) e espectrometria no infravermelho médio (Mid-IR) utilizando ferramentas quimiométricas. Os métodos de regressão por mínimos quadrados parciais (PLS), regressão de mínimos quadrados parciais por intervalos (iPLS), e regressão por máquinas de vetor de suporte (SVM) com seleção de variáveis por Algoritmo Genético (GA) foram utilizadas para modelar as propriedades mencionadas. As amostras de biodiesel foram sintetizadas a partir de diferentes fontes, tais como canola, girassol, milho e soja. Amostras adicionais de biodiesel foram adquiridas de um fornecedor da região sul do Brasil. Em primeiro lugar, o pré-processamento de correção de linha de base foi usado para normalizar os dados espectrais de NIR, seguidos de outros tipos de pré-processamentos que foram aplicados, tais como centralização dos dados na média, 1ª derivada e variação de padrão normal. O melhor resultado para a previsão do ponto de entupimento de filtro a frio foi utilizando os espectros de Mid-IR e o método de regressão GA-SVM, com alto coeficiente de determinação da previsão, $R^2_{\text{Pred}}=0,96$ e baixo valor da Raiz Quadrada do Erro Médio Quadrático da previsão, RMSEP ($^{\circ}\text{C}$)= 0,6. Para o modelo de previsão da massa específica, o melhor resultado foi obtido utilizando os espectros de Mid-IR e regressão por PLS, com $R^2_{\text{Pred}}=0,98$ e RMSEP (g/cm^3)= 0,0002. Quanto ao modelo de previsão para o índice de refração, o melhor resultado foi obtido utilizando os espectros de Mid-IR e regressão por PLS, com excelente $R^2_{\text{Pred}}=0,98$ e RMSEP= 0,0001. Para esses conjuntos de dados, o PLS e o SVM demonstraram sua robustez, apresentando-se como ferramentas úteis para a previsão das propriedades do biodiesel estudadas.

Palavras-chave: Biodiesel. PLS. SVM. NIR. Mid-IR. Massa específica. Índice de refração. Ponto de entupimento de filtro a frio.

ABSTRACT

CUNHA, Camilla Lima. *Study of prediction of biodiesel properties using infrared spectra and multivariate calibration*. 2014. 184f. Dissertação (Mestrado em Engenharia Química) – Instituto de Química – Universidade do Estado do Rio de Janeiro, Rio de Janeiro, 2014.

Biodiesel has been widely used as a renewable energy source which contributes to the mineral diesel decrease demand. Therefore, there are several properties that must be monitored in order to produce and distribute biodiesel with the required quality. In this work, the biodiesel physical properties such as specific mass, refractive index and cold filter plugging point were measured and associated with near infrared spectroscopy (NIR) and mid-Infrared spectroscopy (mid-IR) spectra using chemometric tools. The Partial Least Squares Regression (PLS), Interval Partial Least Squares Regression (iPLS), and Support Vector Machines Regression (SVM) with variable selection by Genetic Algorithm (GA) methods were used to model the aforementioned properties. The biodiesel samples were synthesized from different sources such as canola, sunflower, corn, and soybean. Additional biodiesel samples were purchased from a Brazil South Region supplier. Firstly, the preprocessing baseline correction was used to normalize the NIR spectral data, following others preprocessing types were applied in such as the mean center, the first derivative and standard normal variate. The best result for predicting the cold filter plugging point was using Mid-IR spectra and GA-SVM regression method, with high coefficient determination of prediction, $R^2_{Pred} = 0.94$ and low value of the Root Mean Square Error of Prediction, $RMSEP (^{\circ}C) = 0.7$. For the specific mass prediction model, the best result was obtained using the Mid-IR spectrums and PLS regression, with the $R^2_{Pred} = 0.98$ and $RMSEP (g/cm^3) = 0.0002$. As for a prediction model for the refractive index, the best result was obtained using the Mid-IR spectrums and PLS regression, with the $R^2_{Pred} = 0.98$ and $RMSEP = 0.0001$. For these datasets, the PLS and SVM models demonstrated their robustness, presenting themselves as useful tools for the biodiesel properties prediction studied.

Keywords: Biodiesel. PLS. SVM. NIR. Mid-IR. Specific mass. Refractive index. Cold filter plugging point.

LISTA DE ILUSTRAÇÕES

Figura 1 - Reação de transesterificação de triglicerídeos	25
Figura 2 – Reator contendo massa reacional.....	43
Figura 3 – Fluxograma do processo de síntese do biodiesel.	44
Figura 4 – Etapas de purificação do biodiesel.....	46
Figura 5 – Etapas de purificação da glicerina, recuperação e desidratação do álcool metílico.....	50
Figura 6 – Densímetro (DM40) acoplado ao refratômetro (RX40) com amostrador automático SC1.....	53
Figura 7 – Equipamento para ensaio do ponto de entupimento de filtro a frio marca TANAKA modelo AFP-102.	54
Figura 8 – Ponto de fulgor da marca TANAKA modelo APM-8	55
Figura 9 – Cromatógrafo de fase gasosa Varian (modelo 450) acoplado a um espectrômetro de massa Varian (modelo 220).....	56
Figura 10 – Espectrofotômetro de infravermelho com acessório ATR	57
Figura 11 – Espectrofotômetro de infravermelho com acessório NIRA.....	58
Figura 12 – Esquema representativo da aplicação de pré-processamento centralização dos dados na média aplicado em espectros NIR. Fonte: Sauer (2007).	60
Figura 13 – Esquema representativo da aplicação de pré-processamento de cálculo de primeira derivada em espectros NIR. Fonte: Sauer (2007).....	61
Figura 14 – Fluxograma do estudo quimiométrico utilizado nesta dissertação	77
Figura 15 – Exemplo representativo do gráfico de Hotelling T2 versus resíduos estudatizados com localização das amostras.....	82
Figura 16 – Resultados das medidas de PEFF para mistura binária de biodiesel de milho com biodiesel de soja	85
Figura 17 – Resultados das medidas de PEFF para mistura binária de biodiesel de canola com biodiesel de soja	85
Figura 18 – Resultados das medidas de PEFF para mistura binária de biodiesel de girassol com biodiesel de soja	86
Figura 19 – Resultados das medidas de PEFF para mistura binária de biodiesel de soja com biodiesel da distribuidora da região sul.....	86
Figura 20 – Resultados das medidas de PEFF para misturas ternárias de biodiesel de milho (A) com biodiesel de canola (B) e biodiesel de soja (C).....	87
Figura 21 – Resultados das medidas de PEFF para mistura ternária de biodiesel de milho (A) com biodiesel de girassol (B) e biodiesel de soja (C).....	88
Figura 22 – Resultados das medidas de PEFF para mistura ternária de biodiesel de girassol (A) com biodiesel de canola (B) e biodiesel de soja (C).....	88

Figura 23 – Resultados das medidas de massa específica para mistura binária de biodiesel de milho com biodiesel de soja	92
Figura 24 – Resultados das medidas de massa específica para mistura binária de biodiesel de canola com biodiesel de soja	92
Figura 25 – Resultados das medidas de massa específica para mistura binária de biodiesel de girassol com biodiesel de soja.....	93
Figura 26 – Resultados das medidas de massa específica para mistura binária de biodiesel de canola com biodiesel fornecido por distribuidora da região sul.....	93
Figura 27 – Resultados das medidas de massa específica para misturas ternárias de biodiesel de milho (A) com biodiesel de canola (B) e biodiesel de soja (C).....	94
Figura 28 – Resultados das medidas de massa específica para mistura ternária de biodiesel de milho (A) com biodiesel de girassol (B) e biodiesel de soja (C).....	94
Figura 29 – Resultados das medidas de massa específica para mistura ternária de biodiesel de girassol (A) com biodiesel de canola (B) e biodiesel de soja (C).....	95
Figura 30 – Resultados das medidas de Índice de Refração para mistura binária de biodiesel de milho com biodiesel de soja.....	99
Figura 31 – Resultados das medidas de índice de refração para mistura binária de biodiesel de canola com biodiesel de soja.....	99
Figura 32 – Resultados das medidas de índice de refração para mistura binária de biodiesel de girassol com biodiesel de soja.....	100
Figura 33 – Resultados das medidas de índice de refração para mistura binária de biodiesel de canola com biodiesel fornecido por distribuidora da região sul	100
Figura 34 – Resultados de índice de refração para misturas ternárias de biodiesel de milho (A) com biodiesel de canola (B) e biodiesel de soja (C).....	101
Figura 35 – Resultados de índice de refração para mistura ternária de biodiesel de milho (A) com biodiesel de girassol (B) e biodiesel de soja (C).....	101
Figura 36 – Resultados índice de refração para mistura ternária de biodiesel de girassol (A) com biodiesel de canola (B) e biodiesel de soja (C).....	102
Figura 43 – Espectros brutos de infravermelho próximo (NIR).....	105
Figura 44 – Espectros NIR após utilizar o pré-tratamento de ajuste de linha de base	106
Figura 45 – Espectros brutos de infravermelho médio (Mid-IR)	107
Figura 46 – Espectros brutos de infravermelho médio (Mid-IR) após a retirada da amostra atípica	108
Figura 47 – Gráfico de autovalores versus número de componentes principais (PC) para o banco de dados de NIR	111

Figura 48 – Gráfico dos resíduos versus escores (Q Resíduos versus Hotelling T2) para o banco de dados de NIR	112
Figura 49 – Gráfico dos escores (PC1 versus PC2) para o banco de dados de NIR	112
Figura 50 – Gráfico de autovalores versus número de componentes principais (PC) para o banco de dados de Mid-IR	114
Figura 51 – Gráfico dos escores versus resíduos (Q Resíduos versus Hotelling T2) para o banco de dados de Mid-IR	115
Figura 52 – Gráfico dos escores (PC2 versus PC1) para o banco de dados de Mid-IR.....	116
Figura 53 – Gráfico do RMSEC versus número de variáveis latentes para o índice de refração com espectros NIR	118
Figura 54 – Gráfico do RMSEC versus número de variáveis latentes para o índice de refração com espectros Mid-IR	118
Figura 55 – Gráfico dos resíduos estudentizados versus Hotelling T2 para o índice de refração com espectros NIR	119
Figura 56 – Gráfico dos resíduos estudentizados versus Hotelling T2 para o índice de refração com espectros Mid-IR.....	120
Figura 57 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para o índice de refração com espectros NIR	121
Figura 58 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para o índice de refração com espectros Mid-IR.....	121
Figura 59 – Gráfico do Y previsto versus Y medido para o índice de refração com espectros NIR.....	122
Figura 60 – Gráfico do Y previsto versus Y medido para o índice de refração com espectros Mid-IR	122
Figura 61 – Gráfico de RMSEC versus número de variáveis latentes para a massa específica com espectros NIR	125
Figura 62 – Gráfico de RMSEC versus número de variáveis latentes para a massa específica com espectros Mid-IR.....	125
Figura 63 – Gráfico dos resíduos estudentizados versus Hotelling T2 para a massa específica com espectros NIR	126
Figura 64 – Gráfico dos resíduos estudentizados versus Hotelling T2 para a massa específica com espectros Mid-IR.....	127
Figura 65 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para a massa específica com espectros NIR	128
Figura 66 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para a massa específica com espectros Mid-IR.....	128
Figura 67 – Gráfico do Y previsto versus Y medido para a massa específica com espectros NIR.....	129
Figura 68 – Gráfico do Y previsto versus Y medido para a massa específica com espectros Mid-IR	129
Figura 69 – Gráfico do RMSEC versus número de variáveis latentes para o PEFF com espectros NIR.....	131

Figura 70 – Gráfico do RMSEC versus número de variáveis latentes para o PEFF com espectros Mid-IR	132
Figura 71 – Gráfico dos resíduos estudentizados versus Hotelling T2 para o PEFF com espectros NIR.....	133
Figura 72 – Gráfico dos resíduos estudentizados versus Hotelling T2 para o PEFF com espectros Mid-IR	133
Figura 73 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para o PEFF com espectros NIR.....	134
Figura 74 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para o PEFF com espectros Mid-IR	135
Figura 75 – Gráfico do Y previsto versus Y medido para o PEFF com espectros NIR.....	135
Figura 76 – Gráfico do Y previsto versus Y medido para o PEFF com espectros Mid-IR	136
Figura 77 – Gráfico dos intervalos sugeridos para a realização do PLS para o PEFF com espectros NIR.....	137
Figura 78 – Gráfico dos intervalos sugeridos para a realização do PLS para o PEFF com espectros Mid-IR	137
Figura 79 – Gráfico do RMSEC versus número de variáveis latentes para o PEFF com espectros NIR (iPLS).....	139
Figura 80 – Gráfico do RMSEC versus número de variáveis latentes para o PEFF com espectros Mid-IR (iPLS)	139
Figura 81 – Gráfico dos resíduos estudentizados versus Hotelling T2 para o PEFF com espectros NIR (iPLS).....	140
Figura 82 – Gráfico dos resíduos estudentizados versus Hotelling T2 para o PEFF com espectros Mid-IR (iPLS)	141
Figura 83 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para o PEFF com espectros NIR (iPLS).....	142
Figura 84 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para o PEFF com espectros Mid-IR (iPLS)	143
Figura 85 – Gráfico do Y previsto versus Y medido para o PEFF com espectros NIR (iPLS).....	144
Figura 86 – Gráfico do Y previsto versus Y medido para o PEFF com espectros Mid-IR (iPLS)	144
Figura 87 –Número de ondas (cm-1) selecionados pelo algoritmo genético (GA) para compor a matriz dos espectros NIR.....	145
Figura 88 – Número de ondas (cm-1) selecionados pelo algoritmo genético (GA) para compor a matriz dos espectros Mid-IR	145
Figura 89 – Região de busca dos parâmetros de otimização (Épsilon, Gamma e Cost) para o PEFF com espectros NIR	148
Figura 90 – Região de busca dos parâmetros de otimização (Épsilon, Gamma e Cost) para o PEFF com espectros Mid-IR.....	148
Figura 91 – Gráfico dos valores previstos versus medidos para o PEFF (°C) com espectros NIR.....	149

Figura 92 – Gráfico dos valores previstos versus medidos para o PEF (°C) com espectros Mid-IR	150
--	-----

LISTA DE TABELAS

Tabela 1 – Adaptado do Boletim Mensal do Biodiesel do Mês de Agosto (ANP, 2013)	23
Tabela 2 – Composição dos óleos vegetais em ácidos graxos	25
Tabela 3 – Quadro resumo dos autores que estudaram e trabalharam com a produção do biodiesel	29
Tabela 4 – Padrões de qualidade do biodiesel (adaptado de ANP, 2012)	32
Tabela 5 – Reagentes utilizados na produção do biodiesel	41
Tabela 6 – Composição das misturas ternárias	47
Tabela 7 – Composição das misturas binárias	48
Tabela 8 – Composição das misturas quaternárias	49
Tabela 9 – Resultados das medidas de PEFf dos cinco tipos de biodiesel puro comparado com os resultados da literatura (obtidos por outros autores)	83
Tabela 10 – Resultados das medidas de PEFf para misturas ternárias entre os quatro tipos de biodiesel	89
Tabela 11 – Resultados de PEFf para misturas quaternárias entre os quatro tipos de biodiesel.....	90
Tabela 12 – Resultados das medidas de massa específica (ρ) dos cinco tipos de biodiesel puro comparado com os resultados da literatura	91
Tabela 13 – Resultados das medidas de massa específica (ρ) para misturas ternárias entre os quatro tipos de biodiesel.....	96
Tabela 14 – Resultados das medidas de massa específica (ρ) para misturas quaternárias entre os quatro tipos de biodiesel.....	97
Tabela 15 – Resultados das medidas de índice de refração dos cinco tipos de biodiesel puro	98
Tabela 16 – Resultados das medidas de índice de refração para misturas ternárias entre os quatro tipos de biodiesel.....	103
Tabela 17 – Resultados das medidas de índice de refração para misturas quaternárias entre os quatro tipos de biodiesel.....	104
Tabela 18 – Identificação das regiões espectrais (NIR) das amostras de biodiesel	106
Tabela 19 – Identificação das regiões espectrais (Mid-IR) das amostras de biodiesel	108
Tabela 20 – Comparativo dos resultados obtidos pela PCA utilizando diferentes pré-processamentos nos espectros NIR	110
Tabela 21 – Comparativo dos resultados obtidos pelo PCA utilizando diferentes pré-processamentos nos espectros Mid-IR.....	114
Tabela 22 – Quadro comparativo dos parâmetros de mérito entre os dois bancos de dados (NIR e Mid-IR) obtidos a previsão de índice de refração usando a técnica PLS	123

Tabela 23 – Quadro Comparativo dos parâmetros de mérito entre os dois bancos de dados (NIR e Mid-IR) obtidos a previsão de massa específica usando a técnica PLS	130
Tabela 24 – Intervalos selecionados pelo iPLS ao conjunto de espectros NIR comparados aos intervalos identificados por Workman e Weyer (2008).....	138
Tabela 25 – Intervalos selecionados pelo iPLS ao conjunto de espectros Mid-IR comparados aos intervalos identificados por Silverstein e Webster (2000).....	138
Tabela 26 – Intervalos selecionados pelo GA ao conjunto de espectros NIR comparados aos intervalos identificados por Workman e Weyer (2008).....	146
Tabela 27 – Intervalos selecionados pelo GA ao conjunto de espectros Mid-IR comparados aos intervalos identificados por Silverstein e Webster (2000).....	147
Tabela 28 – Divisão das regiões dos espectros NIR e Mid-IR para aplicação do algoritmo genético (GA)	150
Tabela 29 – Parâmetros dos modelos obtidos por SVM-R para cada uma das cinco regiões e para o espectro completo de NIR.....	151
Tabela 30 – Parâmetros dos modelos obtidos por SVM-R para cada uma das cinco regiões e para o espectro completo de Mid-IR	151
Tabela 31 – Parâmetros do modelo de PLS, IPLS e SVM obtido dos bancos de dados de NIR e Mid-IR para o PEFF.....	154
Tabela 32 – Resultados para o teste F (nível de confiança de 95%) obtidos na comparação dos diferentes modelos de previsão do PEFF utilizando espectros de NIR e Mid-IR	155

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
ATR	Reflectância total atenuada
ASTM	American Society for Testing and Materials
biPLS	Mínimos Quadrados Parciais por exclusão de intervalos
CETESB	Companhia de Tecnologia de Saneamento Ambiental
CG-EM	Cromatógrafo à gás acoplado ao espectrômetro de massas
CLS	Classical Least Squares Regression (Regressão por mínimos quadrados Clássica)
EN	European Committee for Standardization
FIR	Far Infrared (Infravermelho Afastado)
FTIR	Espectrometria no infravermelho médio com transformada de Fourier
FTIR-ATR	Espectrometria de infravermelho por Transformada de Fourier com Reflexão Total Atenuada
GA	Genetic Algorithm (Algoritmo Genético)
ILS	Inverse Least-Square (Regressão dos Mínimos Quadrados Inversa)
iPCA	Interval Principal Component Analysis (Análise por Componentes Principais por intervalos)
iPLS	Interval Partial Least Squares (Mínimos Quadrados Parciais por intervalos)
JK	Algoritmo Jack-Knife
KS	Algoritmo Kennard-Stone
LDA	Linear Discriminant Analysis (Análise Discriminante Linear)
LS-SVM	Least Square - Support Vector Machine (Mínimos Quadrados Parciais por Máquinas de Vetores Suporte)
MC	Mean Center (Centralização dos Dados na Média)
Mid-IR	Medium-Infrared (Infravermelho Médio)
Mid-IRS	Medium-Infrared Spectroscopy (Espectrometria de Infravermelho Médio)
MLR	Multiple Linear Regression (Regressão Linear Múltipla)
MSC	Multiplicative Signal Correction (Correção de Sinal Multiplicativo)
MSE	Mean Square Error (Erro Médio Quadrático)
NBR	Normas Brasileiras
NIR	Near-Infrared (Infravermelho Próximo)
NIRS	Near-Infrared Spectroscopy (Espectrometria no Infravermelho Próximo)
NMR	Nuclear Magnetic Resonance (Ressonância Magnética Nuclear)
OSC	Orthogonal Signal Correction (Correção do Sinal Ortogonal)
PC	Principal Component (Componente Principal)
PCA	Principal Component Analysis (Análise por Componentes Principais)
PCR	Principal Component Regression (Regressão por Componentes Principais)
PEFF	Ponto de Entupimento de Filtro a Frio
PLS	Partial Least Square (Mínimos Quadrados Parciais)
PN	Ponto de Névoa
R	Coefficiente de Correlação

R^2	Coeficiente de Determinação
R^2_{Cal}	Coeficiente de Determinação da Calibração
R^2_{CV}	Coeficiente de Determinação da Validação Cruzada
R^2_{Pred}	Coeficiente de Determinação da Previsão
R_{Cal}	Coeficiente de Correlação da Calibração
R_{CV}	Coeficiente de Correlação da Validação Cruzada
R_{Pred}	Coeficiente de Correlação da Previsão
RMSE	Root Mean Square Error (Raiz Quadrada do Erro Médio Quadrático)
RMSEC	Root Mean Square Error Calibration (Raiz Quadrada do Erro Médio Quadrático de Calibração)
RMSECV	Root Mean Square Error Cross-Validation (Raiz Quadrada do Erro Médio Quadrático com Validação Cruzada)
RMSEP	Root Mean Square Error Prediction (Raiz Quadrada do Erro Médio Quadrático de Previsão)
RANP	Resolução da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
SEV	Erro padrão de validação
siPLS	Mínimos Quadrados Parciais por sinergismo de intervalos
SNV	Standard Normal Variate (Variação Normal Padronizada)
SPA	Successive Projections Algorithm (Algoritmo das Projeções Sucessivas)
SVM	Support Vector Machine (Máquinas de Vetores Suporte)
SVM-R	Support Vector Machine Regression (Regressão por Máquinas de Vetores Suporte)
VL	Variável Latente

SUMÁRIO

	INTRODUÇÃO	19
1	REVISÃO BIBLIOGRÁFICA	22
1.1	Biodiesel	22
1.1.1	<u>Matéria-prima</u>	22
1.1.2	<u>Produção do Biodiesel</u>	24
1.1.3	<u>Purificação do Biodiesel</u>	26
1.1.4	<u>Produção de biodiesel por outros autores</u>	26
1.2	Propriedades do biodiesel	30
1.2.1	<u>Parâmetros de qualidade do biodiesel</u>	30
1.2.2	<u>Ponto de Entupimento de Filtro a Frio</u>	33
1.2.3	<u>Índice de Refração</u>	33
1.2.4	<u>Massa Específica</u>	34
1.2.5	<u>Ponto de Fulgor</u>	35
1.3	Quimiometria	35
1.4	Pesquisas na área de previsão de propriedades de biodiesel utilizando ferramentas quimiométricas e espectros de infravermelho	37
2	METODOLOGIA	41
2.1	Produção do biodiesel	41
2.1.1	<u>Preparo da matéria-prima</u>	42
2.1.2	<u>Reação de transesterificação</u>	42
2.1.3	<u>Separação de fases</u>	43
2.1.4	<u>Purificação do biodiesel</u>	44
2.1.5	<u>Estocagem adequada do biodiesel</u>	46
2.2	Preparo das Misturas de Biodiesel	47
2.3	Recuperação dos Resíduos	49
2.3.1	<u>Destilação da Glicerina</u>	49
2.3.2	<u>Recuperação e desidratação do álcool metílico</u>	49
2.3.3	<u>Recuperação do Sulfato de Sódio e Cloreto de Potássio</u>	50

2.4	Medição das Propriedades	52
2.4.1	<u>Massa Específica e Índice de Refração</u>	52
2.4.2	<u>Ponto de Entupimento de Filtro a Frio</u>	53
2.4.3	<u>Ponto de Fulgor</u>	54
2.5	Avaliação de Composição por Cromatografia	55
2.6	Medição dos espectros de infravermelho	56
2.6.1	<u>Espectros de infravermelho médio</u>	57
2.6.2	<u>Espectros de Infravermelho Próximo</u>	58
2.7	Tratamento dos Espectros	58
2.7.1	<u>Suavização</u>	59
2.7.2	<u>Centralização na média</u>	59
2.7.3	<u>Correção de linha de base</u>	60
2.7.4	<u>Derivada de 1ª ordem</u>	61
2.7.5	<u>Correção de sinal multiplicativo</u>	61
2.7.6	<u>Variação normal padronizada</u>	62
2.7.7	<u>Correção de sinal ortogonal</u>	62
2.7.8	<u>Autoescalamento</u>	64
2.8	Amostras Atípicas	64
2.9	Avaliação dos Componentes Principais	65
2.9.1	<u>PCA</u>	65
2.10	Algoritmo KS	66
2.11	Técnicas de Calibração Multivariada	67
2.11.1	<u>Mínimos quadrados parciais (PLS)</u>	67
2.11.2	<u>Máquinas de Vetores Suporte</u>	69
2.11.3	<u>Máquinas de Vetores Suporte por Mínimos Quadrados Parciais</u>	73
2.12	Técnicas de Seleção de Variáveis	74
2.12.1	<u>Mínimos Quadrados Parciais por Intervalos (iPLS)</u>	74
2.12.2	<u>Algoritmo Genético (GA)</u>	74
2.13	Resumo do estudo quimiométrico	76
2.14	Ferramentas e conceitos estatísticos	79
2.14.1	<u>Parâmetros estatísticos</u>	79
2.14.2	<u>Hotelling T^2, resíduos padronizados e estudentizados</u>	81
3	RESULTADOS E DISCUSSÕES	83

3.1	Caracterização do Biodiesel	83
3.1.1	<u>Ponto de Entupimento de Filtro a Frio</u>	83
3.1.2	<u>Massa Específica</u>	91
3.1.3	<u>Índice de Refração</u>	98
3.1.4	<u>Ponto de Fulgor</u>	104
3.2	Obtenção dos espectros	105
3.3	Análise de Componentes Principais	109
3.3.1	<u>Banco de dados dos espectros de infravermelho próximo (NIRS)</u>	109
3.3.2	<u>Banco de dados dos espectros de infravermelho médio (Mid-IR)</u>	113
3.4	Modelos de Previsão das Propriedades	116
3.4.1	<u>Modelos de Previsão para Índice de Refração</u>	117
3.4.1.1	Regressão por PLS	117
3.4.2	<u>Modelos de Previsão para Massa Específica</u>	124
3.4.2.1	Regressão por PLS	124
3.4.3	<u>Modelos de Previsão para Ponto de Entupimento (PEFF)</u>	131
3.4.3.1	Regressão por PLS	131
3.4.3.2	Seleção de Variáveis por iPLS seguido pela Regressão por PLS.....	136
3.4.3.3	Regressão por SVM com seleção de variáveis por GA.....	144
3.5	Recuperação dos Resíduos	156
4	CONCLUSÕES E PROPOSTAS FUTURAS	157
4.1	Conclusões	157
4.2	Propostas Futuras	158
	REFERÊNCIAS	160
	ANEXO: Informação da composição e nomeação das amostras de biodiesel nos conjuntos espectrais de NIR e Mid-IR	174

INTRODUÇÃO

A Resolução N° 14 de 2012 da ANP (Agência Nacional de Petróleo, Gás Natural e Biocombustíveis) define o biodiesel como um combustível renovável, biodegradável e para uso em motores a combustão interna com ignição por compressão. O biodiesel tem a sua composição baseada em ésteres alquílicos de ácidos graxos, com cadeia longa, derivados de óleos vegetais ou de gorduras animais (ANP, 2012).

Segundo Fangrui *et al.* (1999), o biodiesel apresenta algumas vantagens quando comparado ao diesel por ser biodegradável, praticamente isento de enxofre, e por sua combustão ser mais eficiente que a do diesel. Alguns estudos indicaram que o uso do biodiesel, além de reduzir a emissão de gases associados ao efeito estufa (CO₂ e CO), ocasiona uma redução de particulados e óxidos de enxofre (FANGRUI *et al.* (1999); SUAREZ *et al.* (2009)).

O biodiesel vem sendo empregado como uma fonte de energia renovável e isto contribui para a diminuição da demanda por diesel mineral, que é importado pelo Brasil. Para isto, o biodiesel deve apresentar atributos de qualidade que devem ser monitorados para que se possa produzir e distribuir biodiesel com a qualidade especificada.

Neste contexto este estudo tem como objetivo apresentar modelos de previsão do ponto de entupimento, massa específica e índice de refração, de origens distintas, baseados em técnicas espectrométricas acopladas com ferramentas quimiométricas. Estas técnicas instrumentais são não destrutivas, apresentam versatilidade e são capazes de fazer a previsão destas propriedades quando acopladas com técnicas quimiométricas.

Motivação do Estudo

Segundo a norma ANP (ANP, 2012), diversas propriedades são utilizadas para atender aos critérios de avaliação da qualidade do biodiesel. Dentre estas propriedades estão: o ponto de entupimento de filtro a frio (PEFF), a massa específica a 20°C e a composição química (teor de ésteres metílicos). Embora o índice de refração não seja considerado pela ANP um parâmetro de qualidade do

biodiesel, esta propriedade foi incluída neste estudo por ser uma importante propriedade física.

O estudo da previsão do ponto de entupimento de filtro a frio é importante por causa do apelo termodinâmico da não aditividade e da diversidade dos óleos utilizados na produção do biodiesel. Quando a temperatura se reduz, o biodiesel tende a solidificar-se parcialmente ou a perder sua fluidez, levando à interrupção do fluxo do combustível e entupimento do sistema de filtração, ocasionando problemas na partida do motor.

A massa específica é uma propriedade muito importante para conferir qualidade ao biodiesel, visto que a resolução brasileira determina como 30 dias o prazo de comercialização do biodiesel a partir da produção do mesmo. Ultrapassado este prazo, a recomendação é realizar uma nova medida da massa específica a 20 °C, e se houver uma diferença superior a 3,0 kg/m³ em relação ao valor da primeira determinação contida no certificado, todos os parâmetros de qualidade da resolução deverão ser reavaliados (ANP, 2012).

O índice de refração é uma propriedade físico-química muito importante na avaliação da qualidade de óleos, além do que permite fazer uma inferência do grau de saturação do mesmo. Esta propriedade pode ser bastante útil na avaliação da qualidade do biodiesel.

Objetivo Geral

Este trabalho tem como objetivo avaliar o uso dos espectros de infravermelho médio (Mid-IR) e próximo (NIR) acoplados com ferramentas quimiométricas para prever o ponto de entupimento de filtro a frio, a massa específica e o índice de refração do biodiesel puro e de suas misturas.

Objetivos Específicos

Podem ser citados como objetivos específicos desse trabalho:

- Construir modelos de calibração multivariada utilizando diferentes técnicas de regressão para determinação das propriedades das diferentes misturas entre os óleos de biodiesel.

- Avaliar o desempenho dos modelos construídos utilizando-se os métodos de regressão de mínimos quadrados parciais (*Partial Least Squares*, PLS), utilizando o algoritmo de Kennard-Stone (KS) como método de seleção de amostras, para compor as matrizes de calibração e validação.
- Utilizar a regressão de mínimos quadrados parciais por intervalos (*Interval Partial Least Squares*, iPLS) para a seleção das variáveis e a regressão por método linear;
- Avaliar desempenho dos modelos construídos a partir do método máquinas de vetores de suporte (*Support Vector Machine*, SVM), utilizando o algoritmo genético (GA) para a seleção de variáveis.

Delimitação do Trabalho

O presente estudo está limitado a análise estatística dos dados, através do uso das ferramentas de estatística multivariada. Os modelos propostos são específicos para os bancos de dados obtidos através de medidas espectrais no infravermelho próximo e médio, além das medições das propriedades, desta forma, a previsão é específica para as propriedades estudadas.

1 REVISÃO BIBLIOGRÁFICA

1.1 Biodiesel

O biodiesel é importante por ser um combustível renovável, com diversidade de matérias-primas e biodegradável, contribuindo para uma menor emissão de monóxido de carbono, particulados e outros poluentes (LÔBO *et al.*, 2009; CANDEIA, 2008; SILVA e FREITAS, 2008).

O biodiesel é uma mistura de ésteres alquílicos de ácidos graxos de cadeia linear de hidrocarbonetos contendo de 14 a 20 átomos de carbono, obtido da transesterificação dos triglicerídeos de óleos e gorduras com álcoois de cadeia curta (LÔBO *et al.*, 2009 e CARRARETTO *et al.*, 2004). Esses ésteres podem ser metílicos ou etílicos, atualmente no Brasil é utilizado o metanol para a produção do éster metílico. Dentre as vantagens de utilizar o metanol estão o maior favorecimento da reação, fácil dissolução do catalisador básico, a menor razão molar entre o óleo e o álcool, porém apresenta a desvantagem de ser tóxico e de origem não renovável. Já o etanol, além de ter produção consolidada no Brasil, é menos tóxico, é renovável e produz biodiesel com número de cetano maior e melhor lubricidade (LEUNG e GUO, 2006; INNOCENTINI, 2007; MARCHETTI *et al.*, 2007; DEMIBRAS, 2008; LEUNG *et al.*, 2010). A principal desvantagem do etanol está no fato de promover uma maior dispersão da glicerina no biodiesel, dificultando muito sua separação. Para a obtenção de maiores rendimentos na reação de transesterificação costuma-se utilizar excesso de álcool para facilitar a remoção da glicerina.

1.1.1 Matéria-prima

Segundo Pinto *et al.* (2005), a utilização de óleos vegetais como combustível teve início com o desenvolvimento dos motores diesel no final do século XIX por Rudolph Diesel, no entanto, foi no início do século XX que o diesel começou a ser utilizado nesse tipo de motor em consequência do seu baixo custo e dificuldades encontradas na utilização de óleos vegetais. Dentre as dificuldades enfrentadas na utilização de oleaginosas no motor a diesel, os autores citaram a redução na potência, redução na eficiência térmica do motor, elevada viscosidade e menor

volatilidade que resultam na formação de depósitos de carbono nos motores devido a combustão incompleta.

As principais fontes de produção de biodiesel são os óleos vegetais e a gordura animal. De acordo com a ANP, em agosto de 2013, as principais matérias-primas utilizadas na produção de biodiesel foram o óleo de soja e a gordura bovina (ANP, 2013). Em 2013, até o mês de julho, o óleo de soja foi a matéria-prima mais utilizada na produção de biodiesel no Brasil, respondendo por 80,4% da produção de biodiesel, enquanto que a gordura bovina por 16,0% (ANP, 2013). Quando analisadas quanto à região, as principais matérias-primas utilizadas na produção do biodiesel podem sofrer variações devidas à disponibilidade da oleaginosa ou da gordura animal, e as características físico-químicas requeridas para o combustível em determinada região, conforme observado na Tabela 1.

Tabela 1 – Adaptado do Boletim Mensal do Biodiesel do Mês de Agosto (ANP, 2013)

Matéria-Prima	Região				
	Norte	Nordeste	Centro-Oeste	Sudeste	Sul
Óleo de Soja	-	65,3%	88,05%	52,5%	85,2%
Gordura Bovina	56,5%	28,1%	9,38%	34,6%	14,4%
Óleo de Algodão	-	6,6%	0,69%	0,6%	-
Outros Materiais Graxos	43,5%	-	0,55%	3,8%	-
Óleo de Fritura usado	-	-	1,34%	8,4%	0,2%
Gordura de Porco	-	-	-	-	0,2%
Gordura de Frango	-	-	-	-	0,01%
Óleo de Macaúba	-	-	-	0,03%	-

Segundo Shahid e Jamal (2007) e Pousa *et al.* (2007), o uso de óleo de girassol, óleo de algodão, óleo de soja, óleo de palma, óleo de amendoim e óleo de canola para síntese de biodiesel são excelentes alternativas para a produção do biodiesel. Conceição *et al.* (2007) indicam o uso de óleo de mamona como precursor para o biodiesel, justificando sua utilização em função dos baixos custos de obtenção da mamona em relação a outras fontes vegetais.

Moraes (2008) estudou a utilização de biodiesel proveniente de sebo animal em motores a diesel. O autor relatou em seus estudos conversões de até 95% de

ésteres metílicos, dando origem a um biodiesel mais estável que os de origem vegetal em relação à oxidação.

O uso contínuo de óleos vegetais em motores a combustão interna, apesar de ser benéfico do ponto de vista energético, apresenta diversos problemas, tais como: combustão incompleta, formação de depósitos de carbono nos sistemas de injeção e contaminação do óleo lubrificante. Para Knothe e Steidley (2005) esses problemas podem ter como causas a polimerização dos triglicerídeos, a baixa volatilidade e, principalmente, a alta viscosidade dos óleos.

1.1.2 Produção do Biodiesel

A literatura apresenta diversas abordagens para viabilizar a utilização de óleos vegetais como combustíveis, dentre as quais se destacam as etapas de craqueamento, transesterificação, microemulsão e esterificação (SUAREZ *et al.*, 2009). Dentre os processos citados, o mais comum é a transesterificação, cujo processo leva à obtenção de ésteres alquílicos de óleo vegetal ou de gordura animal, conhecido como biodiesel. Este possui propriedades físico-químicas semelhantes às do diesel mineral podendo ser utilizado puro ou em misturas, exigindo pouca ou nenhuma alteração do motor (SCHUCHARDT *et al.*, 2007; KNOTHE *et al.*, 2006).

Na reação de transesterificação, os triglicerídeos oriundos de óleos vegetais ou gorduras animais reagem com um álcool de cadeia curta (etanol ou metanol) na presença de um catalisador (básico, ácido, enzimático ou heterogêneo), produzindo enfim a glicerina e o biodiesel, como pode ser observado na Figura 1 (GERIS *et al.*, 2007 e TAPANES *et al.*, 2008). Segundo alguns autores da literatura (MARCHETTI *et al.*, 2007; DEMIBRAS, 2008; LEUNG *et al.*, 2010), dentre os estudos de otimização realizados no processo de obtenção do biodiesel, o mais recomendado em escala industrial é a transesterificação alcalina homogênea.

Na Figura 1, o radical R_4 é o CH_3 proveniente do álcool metílico (CH_3-OH). Os radicais R_1 , R_2 e R_3 não são apresentados diretamente porque dependem da oleaginosa utilizada como matéria-prima, como por exemplo, o óleo de canola, girassol, milho e soja apresentam diferentes ácidos graxos em suas composições, conforme mostrado na Tabela 2 (GARCIA, 2006; CORDEIRO, 2003; ALANDER *et al.*, 2007). De acordo com a Tabela 2, os óleos de soja, milho, girassol e canola são

compostos principalmente por ácido palmítico, esteárico, oleico, linoleico e linolênico, compondo aproximadamente 97,8%, 99,1%, 98,3% e 100% da massa total, respectivamente.

Figura 1 - Reação de transesterificação de triglicerídeos

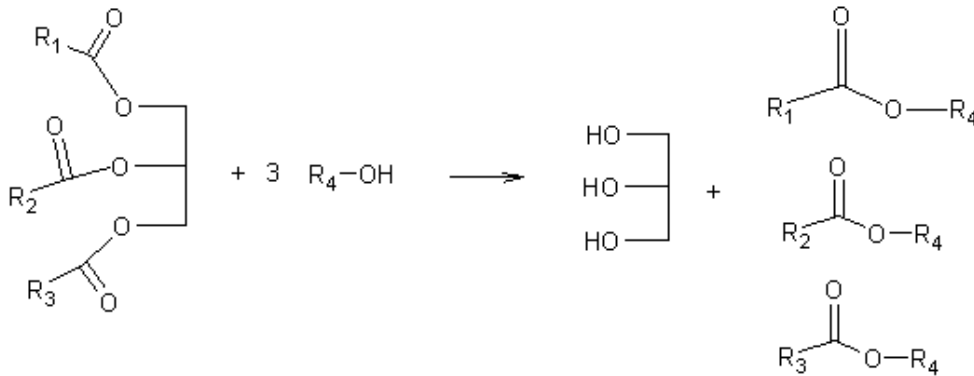


Tabela 2 – Composição dos óleos vegetais em ácidos graxos

Sigla	Nome usual do ácido	Ácido graxo			Composição do óleo vegetal (% massa)			
		Nome IUPAC	Fórmula molecular	Massa Molar (g/mol)	Soja	Milho	Girassol	Canola
C12:0	Láurico	dodecanóico	C ₁₁ H ₂₃ COOH	200,32	0,1	-	-	-
C14:0	Mirístico	tetradecanóico	C ₁₃ H ₂₇ COOH	228,38	0,2	0,2	0,1	-
C16:0	Palmítico	hexadecanóico	C ₁₅ H ₃₁ COOH	256,43	11,0	13,0	5,5	4,8
C16:1	Palmitoléico	cis-9-hexadecenóico	C ₁₅ H ₂₉ COOH	254,41	0,2	-	0,1	-
C17:0	Heptadecanóico	heptadecanóico	C ₁₆ H ₃₃ COOH	270,45	-	-	-	-
C18:0	Esteárico	octodecanóico	C ₁₇ H ₃₅ COOH	284,48	4,2	2,5	4,7	2,6
C18:1	Oléico	cis-9-octadecenóico	C ₁₇ H ₃₃ COOH	282,47	21,8	30,5	19,5	68,8
C18:2	Linoléico	cis,cis-9,12-octadecadienóico	C ₁₇ H ₃₁ COOH	280,45	53,3	52,1	68,5	17,8
C18:3	Linolênico	cis,cis,cis-9,12,15-octadecatrienóico	C ₁₇ H ₂₉ COOH	278,44	7,5	1,0	0,1	6,0
C20:0	Araquídico	eicosanóico	C ₁₉ H ₃₉ COOH	312,53	0,3	0,5	0,3	-
C20:1	Gadoléico	cis-9-eicosenóico	C ₁₉ H ₃₇ COOH	310,51	0,2	0,2	0,1	-
C22:0	Behênico	docosanóico	C ₂₁ H ₄₃ COOH	340,58	0,5	-	0,9	-
C22:1	Erúcico	cis-13 docosenóico	C ₂₁ H ₄₁ COOH	338,57	0,3	-	-	-
C24:0	Lignocérico	tetracosanóico	C ₂₃ H ₄₇ COOH	368,64	0,4	-	0,2	-

Para Meher *et al.* (2006), Bouiad *et al.* (2007) e Encinar *et al.* (2007), alguns fatores afetam a reação de transesterificação, como a razão molar álcool/óleo, o tipo de catalisador, o tipo de álcool, a temperatura da reação, a velocidade de agitação e também a qualidade da matéria-prima utilizada para produção do biodiesel.

1.1.3 Purificação do Biodiesel

Após o processo de transesterificação, a massa reacional é constituída de duas fases que podem ser separadas por decantação ou centrifugação. A fase mais leve é composta de uma mistura de ésteres metílicos ou etílicos, enquanto que a fase mais pesada é composta de glicerina bruta, álcool e impurezas que podem ser provenientes da matéria-prima. Geris *et al.* (2007) afirmam que o processo de purificação dos ésteres consiste basicamente na separação da glicerina, evaporação do excesso de álcool e lavagem ácida para a neutralização de resíduos de catalisador, seguido pelo processo de secagem, obtendo como produto final o biodiesel. Enquanto o coproduto dessa reação, a glicerina, se for devidamente purificado, poderá ser reaproveitado em indústrias farmacêuticas e de cosméticos, conforme foram observados por Mota *et al.* (2009).

1.1.4 Produção de biodiesel por outros autores

O biodiesel é sintetizado em escala laboratorial, em plantas pilotos ou até mesmo em escala semi-industrial, utilizando variadas fontes de oleaginosas, diferentes alcoóis de cadeia curta, distintos catalisadores (ácido, básico, heterogêneos), fazendo uso de algumas etapas de preparo da matéria-prima e de purificação de biodiesel.

Lima *et al.* (2007) produziram biodiesel de babaçu utilizando álcool metílico e etílico e catalisador básico, o hidróxido de sódio (NaOH). Não realizaram nenhuma etapa de preparação da matéria-prima antes de iniciar a reação, no entanto, depois de reagido e separado da glicerina, submeteram o biodiesel a sucessivas lavagens com água destilada e aquecimento a 110°C.

Alves *et al.* (2006) produziram biodiesel de mamona em escala semi-industrial, realizando a reação de transesterificação com álcool etílico e como catalisador básico, o hidróxido de potássio (KOH). Prepararam previamente as

matérias-primas, realizando as etapas de correção da acidez do óleo de mamona e a desidratação do álcool. Contudo, os autores não mencionaram a realização de etapas de purificação.

Gama *et al.* (2010) e Perin *et al.* (2006) produziram biodiesel de girassol e mamona, respectivamente. As sínteses foram realizadas pela rota metílica e os respectivos catalisadores heterogêneos escolhidos foram K_2CO_3/Al_2O_3 com KOH, e KOH/Al_2O_3 e KOH/SiO_2 , respectivamente. Os autores optaram pelo preparo prévio dos catalisadores, sendo que Gama *et al.* (2006) realizaram previamente a extração do óleo de girassol a partir da semente do mesmo. Dentre as etapas de purificação do biodiesel, a lavagem do biodiesel com solução saturada de NaCl e o posterior aquecimento foram escolhidas por Gama *et al.* (2010) como passos importantes no processo de retirada de impurezas do combustível.

Brito (2011) e Melo (2009) sintetizaram biodiesel de soja, as reações ocorreram via rota metílica e o catalisador utilizado foi o KOH. Os autores optaram por diferentes etapas de purificação do biodiesel, Brito (2011) realizou lavagens com solução de cloreto de amônio, com água destilada e aquecimento a 100°C, enquanto que Melo (2009) optou por sucessivas lavagens com água destilada e secagem à vácuo.

Silva Filho (2010) e Freire (2009) produziram biodiesel etílico de óleos e gorduras residuais, e pinhão manso, respectivamente. As reações de transesterificação ocorreram com catalisador básico KOH. Silva Filho (2010) preparou o óleo com as etapas de filtração à vácuo, homogeneização e secagem sob vácuo a 80°C. Para a purificação do biodiesel foram realizadas lavagens com água destilada, solução de H_3PO_4 0,5 mol.L⁻¹, com água destilada novamente e por fim o aquecimento a 100°C.

Candeia (2008) sintetizou biodiesel de soja utilizando a rota etílica e metílica na reação de transesterificação e KOH como catalisador, sem realizar processo de preparo da matéria-prima. Para a purificação do biodiesel, o autor utilizou lavagens com solução HCl 0,1 mol.L⁻¹, água destilada, aquecimento, adição de Na_2SO_4 sólido e, por fim, uma filtração.

Cunha (2008) produziu biodiesel metílico de sebo bovino, óleo de frango e soja em escala laboratorial, utilizando KOH como catalisador. As etapas de purificação do biodiesel utilizadas foram aquecimento a 110°C, utilização do adsorvente Magnesol (silicato de magnésio sintético e amorfo) e filtração. O

Magnesol foi aplicado para remover contaminantes do biodiesel como água, sabão e glicerina livre.

Santos (2008) produziu biodiesel de babaçu utilizando álcoois metílico e etílico, e como catalisador básico, o hidróxido de potássio (KOH). Não foi realizada nenhuma etapa prévia de preparo da matéria-prima. Após a etapa de separação de fases, o biodiesel foi submetido a lavagens com água acidificada, água destilada e a um aquecimento.

Christoff (2006) sintetizou biodiesel de óleo de fritura utilizando a rota etílica e metílica na reação de transesterificação e NaOH como catalisador. Antes de iniciar a reação, foi realizado o aquecimento e filtração do óleo. Para a purificação do biodiesel, optou-se pela evaporação do álcool, seguido de lavagens com água destilada e aquecimento.

Dantas (2006) produziu biodiesel de milho pelas rotas metílica e etílica, utilizando o catalisador básico KOH. Para purificação do biodiesel realizou lavagens com solução de HCl $0,1 \text{ mol.L}^{-1}$, água destilada e aquecimento a 100°C .

Na Tabela 3 são apresentadas de forma resumida as matérias-primas, as etapas de preparo das matérias-primas e as etapas de purificação do biodiesel, utilizadas pelos autores supracitados.

Tabela 3 – Quadro resumitivo dos autores que estudaram e trabalharam com a produção do biodiesel

Matéria-prima			Ajuste Matéria-prima	Purificação do biodiesel	Referência
Óleo	Álcool	Catalisador			
Mamona	Etílico	KOH	Correção de acidez do óleo e desidratação do álcool	-	Alves <i>et al.</i> , 2006
Óleo de fritura	Etílico/Metílico	NaOH	Aquecimento e filtração do óleo;	Evaporação do álcool, lavagens com água destilada e aquecimento	Christoff, 2007
soja	Etílico/Metílico	KOH	-	Lavagens com solução HCl (0,1M), água destilada, aquecimento, adição de Na ₂ SO ₄ e filtração	Candeia, 2008
Sebo Bovino, óleo de frango e soja	Metílico	KOH	-	Aquecimento a 110°C, adição de magnésol e filtração	Cunha, 2008
Babaçu	Etílico/Metílico	KOH	-	Lavagens com água acidificada, água destilada e aquecimento	Santos, 2008
Soja e mamona	Metílico	KOH	-	3 Lavagens com água destilada e secagem à vácuo	Melo, 2009
girassol	Metílico	Heterogêneo K ₂ CO ₃ /Al ₂ O ₃ com KOH	Preparação do catalisador e extração do óleo de girassol	Lavagem com solução saturada de NaCl e aquecimento	Gama <i>et al.</i> , 2010
Mamona	Metílico	Heterogêneo KOH/Al ₂ O ₃ e KOH/SiO ₂	Preparação do catalisador	-	Perin <i>et al.</i> , 2006
Babaçu	Etílico/Metílico	NaOH	-	2 Lavagens com água destilada e aquecimento	Lima <i>et al.</i> , 2007
Soja	Metílico	KOH	-	Lavagens com solução de cloreto de amônio, com água destilada e aquecimento a 100°C	Brito, 2011
Milho	Etílico/Metílico	KOH	-	Lavagens com solução de HCl (0,1M), água destilada e aquecimento a 100°C	Dantas, 2006
Óleos e gorduras residuais	Etílico	KOH	Filtração à vácuo, homogeneização e secagem sob vácuo a 80°C	Lavagens com água destilada, solução de H ₃ PO ₄ (0,5M), com água destilada e aquecimento a 100°C	Silva Filho, 2010
Pinhão Manso	Etílico	KOH/NaOH	-	Lavagem com água destilada, sulfato de sódio anidro e aquecimento	Freire, 2009

1.2 Propriedades do biodiesel

Dentre várias propriedades do biodiesel que podem ser utilizadas para confirmar e avaliar a qualidade do biodiesel, algumas foram consideradas parâmetros imprescindíveis e por este motivo estão presentes na resolução da ANP 14/12.

1.2.1 Parâmetros de qualidade do biodiesel

A ANP foi implantada pelo Decreto nº 2.455, de 14 de janeiro de 1998, para regular as atividades que integram a indústria do petróleo, do gás natural e dos biocombustíveis no Brasil (ANP, 1998). Os critérios de avaliação adotados pela ANP para analisar a qualidade do biodiesel brasileiro são regulamentados pela resolução Nº 14 de 2012 (RANP 14/12). Esta resolução mantém os critérios de avaliação da qualidade do biodiesel no Brasil mais restritivos. Os padrões de qualidades presentes nesta resolução foram baseados nas normas ASTM D6751 (*American Society for Testing and Materials*) e EN 14214 (*European Committee for Standardization*). A mistura óleo diesel/biodiesel tem sua especificação estabelecida pela resolução ANP 15/2006.

Segundo Lôbo *et al.* (2009) para preservar e garantir a qualidade do biodiesel é necessário estabelecer padrões de qualidade, fixando os teores limites dos contaminantes que não prejudiquem a qualidade das emissões, o desempenho, a integridade do motor, a segurança no transporte e o manuseio. A possível degradação do produto também precisa ser monitorada durante o processo de estocagem.

Propriedades como viscosidade cinemática, ponto de fulgor e cinzas sulfatadas, embora tenham origem na normatização do diesel mineral, fornecem resultados muito elucidativos acerca da qualidade do biodiesel. Outros componentes também podem ser encontrados como contaminantes no biodiesel, provenientes da matéria prima, como fósforo, enxofre, magnésio e cálcio (LÔBO *et al.*, 2009). A quantidade de contaminantes no biodiesel é resultado da eficiência do processo de produção, podendo estar presentes em pequenas ou grandes quantidades: glicerina

livre, glicerídeos não reagidos, álcool residual, sabões, resíduos de catalisadores e água. Durante a armazenagem do biodiesel, a absorção de umidade e processos de degradação oxidativa podem contribuir para a formação de água, peróxidos e ácidos carboxílicos de cadeia curta (LÔBO *et al.*, 2009).

No trabalho de Lôbo *et al.* (2009), os métodos analíticos para avaliação da qualidade do biodiesel foram separados em quatro grupos:

- (i) métodos para determinação de contaminantes provenientes da matéria prima (teor de fósforo e enxofre, corrosividade ao cobre, cálcio e magnésio);
- (ii) métodos para avaliação do processo produtivo (aspecto, contaminação total, glicerina livre e total, resíduo de carbono, teor de éster, ponto de fulgor, metanol e etanol, cinzas sulfatadas, sódio e potássio);
- (iii) métodos para avaliação das propriedades inerentes às estruturas moleculares (massa específica, viscosidade cinemática a 40°C, índice de iodo, destilação 90% volume recuperado, número de cetanos, ponto de névoa, ponto de entupimento de filtro a frio e ponto de fluidez);
- (iv) métodos para monitoramento da qualidade do biodiesel durante o processo de estocagem (estabilidade à oxidação a 110°C, água e sedimentos, índice de acidez).

Na Tabela 4 são apresentados e comparados os padrões de qualidade do biodiesel segundo as normas ANP 14/2012, EN 14214 e ASTM D6751.

Tabela 4 – Padrões de qualidade do biodiesel (adaptado de ANP, 2012)

CARACTERÍSTICA	UNIDADE	Brasil (ANP 07/2008)	EU (EN 14214)	EUA (ASTM D6751)
Aspecto	---	Límpido e isento de impurezas	---	---
Massa específica *	kg/m ³	850-900 a 20 °C	860-900 a 15 °C	---
Viscosidade cinemática a 40 °C *	mm ² /s	3,0-6,0	3,5-5,0	1,9-6,0
Água e sedimentos, máx. *	% volume	---	---	0,05
Ponto de fulgor, mín. *	°C	100	120	130
Destilação; 90% vol. recuperados, máx. *	°C	---	---	360
Resíduo de carbono dos, máx. *	% massa	Em 100% da amostra 0,050	10% residual da destilação 0,30	Em 100% da amostra 0,050
Cinzas sulfatadas, máx. *	% massa	0,020	0,020	---
Enxofre total, máx. *	mg/kg	50	10	15
Corrosividade ao cobre, 3 h a 50 °C, máx. *	-	1	1	3
Número de cetanos *	-	Anotar	51 (mín.)	47 (mín.)
Ponto de entupimento de filtro a frio, máx. *	°C	19	Por região	---
Ponto de fluidez (<i>pour point</i> - PP) *	°C	---	Por região	---
Ponto de névoa (<i>cloud point</i> - CP) *	°C	---	---	Registrar
Sódio + Potássio, máx	mg/kg	5	5	---
Cálcio + Magnésio, máx	mg/kg	5	5	---
Fósforo, máx	mg/kg	10	10	10
Contaminação total, máx	mg/kg	24	24	---
Teor de éster, min.	% massa	96,5	96,5	---
Índice de acidez, máx.	mg KOH/g	0,5	0,5	0,5
Glicerina livre, máx.	% massa	0,02	0,02	0,02
Glicerina total, máx.	% massa	0,25	0,25	0,24
Monoglicérides	% massa	Anotar	0,8 (máx)	---
Diglicérides	% massa	Anotar	0,2 (máx)	---
Triglicérides	% massa	Anotar	0,2 (máx)	---
Metanol ou Etanol, máx.	% massa	0,2	0,20	---
Índice de iodo	g I ₂ /100 g	Anotar	120 (máx)	---
Estabilidade à oxidação a 110 °C, mín	h	6	6	---
Água, máx	mg/kg	500	500	500
Ácido linolênico	% massa	---	12 (máx)	---
Ésteres metílicos com mais que 4 insaturações	% massa	---	1 (máx)	---

* Parâmetros típicos da normatização do diesel mineral.

1.2.2 Ponto de Entupimento de Filtro a Frio

Quando o biodiesel líquido é resfriado, os componentes com um ponto de fusão mais elevado começam a se cristalizar e saem da solução líquida. Inicialmente, eles formam pequenos cristais, que tornam a aparência do combustível nebulosa a olho nu. A temperatura em que ocorre essa névoa pela primeira vez é chamada de ponto de névoa (PN). O arrefecimento contínuo faz com que os cristais se acumulem e se aglomerem, aumentando assim a viscosidade do biodiesel (WANG *et al.*, 2012 e MEJIA *et al.*, 2013). Quando essas estruturas se tornam suficientemente grandes, podem causar a obstrução do filtro, que é chamado de ponto de entupimento de filtro a frio (PEFF). Portanto, o ponto de entupimento de filtro a frio é a temperatura em que o combustível perde a filtrabilidade quando resfriado. Segundo a resolução Nº 14 de 2012 da ANP o método ABNT NBR 14747 é o utilizado para determinação do ponto de entupimento de filtro a frio, determinando como aceitável a temperatura máxima de 19°C e aplicável para as regiões Sul, Sudeste, Centro-Oeste e para o estado da Bahia, devendo ser anotado para as demais regiões (LÔBO *et al.*, 2009). A RANP 14/12 indica como alternativa o método ASTM D6371.

1.2.3 Índice de Refração

O índice de refração está fundamentado como um parâmetro físico importante para garantir a qualidade dos óleos, mas pode ser usado também como um parâmetro auxiliar para garantir a qualidade do biodiesel, embora esta propriedade físico-química não tenha sido incluída pela ANP na resolução Nº 14 de 2012. Segundo Moretto e Fett (1998), o índice de refração é a relação existente entre a velocidade da luz no vácuo e no meio (substância em exame), como apresentado na Equação 1, variando na razão inversa da temperatura e tende a aumentar com o grau de insaturação dos ácidos graxos constituintes dos triglicerídeos. Assim quanto mais insaturado for o éster, maior será o índice de refração do mesmo.

$$n = \frac{c}{v} \quad (1)$$

Onde c é a velocidade da luz no vácuo ($c=3 \cdot 10^8$ m/s) e v é a velocidade da luz no meio. De modo geral, a velocidade da luz nos meios materiais é menor que c e assim, em geral, tem-se $n > 1$.

1.2.4 Massa Específica

A massa específica de uma substância é definida através da razão entre a massa de uma porção compacta e homogênea dessa substância e o volume ocupado por ela. Então, matematicamente, a massa específica é expressa pela Equação 2 (Milli *et al.*, 2011).

$$\rho = \frac{m}{V} \quad (2)$$

Onde m é a massa da porção de substância e V é o volume ocupado por ela. No Sistema Internacional de Unidades a unidade de massa específica é kg/m^3 .

O biodiesel quando comparado ao diesel mineral, apresenta um valor de massa específica maior. A massa específica do biodiesel está relacionada diretamente com a estrutura molecular das moléculas que o compõe. Então, quanto maior for o comprimento da cadeia carbônica do éster alquílico, maior será a massa específica do combustível, contudo esta será menor quanto maior for o número de insaturações presentes na molécula. Para Lôbo *et al.* (2009), a presença de impurezas no biodiesel pode influenciar no valor da massa específica do mesmo. Como exemplo de impurezas, é possível encontrar resíduo de álcool, glicerina e outras substâncias adulterantes.

A especificação da massa específica do biodiesel está na faixa de 860-900 kg/m^3 , de acordo com a norma EN 142114 (*European Committee for Standardization*, 2009). A RANP 14/12, fixa uma faixa de valores de massa específica entre 850 a 900 kg/m^3 . Além dos métodos indicados pela norma européia, a ANP permite a aplicação dos métodos ASTM D1298 e NBR 7148 (hidrômetros de vidro) e os métodos ASTM D4052 e NBR 14065 (decímetros digitais).

A resolução brasileira determinou também que o biodiesel produzido tem um prazo máximo de 30 dias, a contar da data de certificação, para ser comercializado.

Ultrapassado este prazo, o biodiesel deve ser submetido a uma nova determinação da massa específica a 20°C, onde, havendo diferença inferior a 3,0 kg/m³ em relação ao valor da primeira determinação contida no certificado, deverão ser novamente realizadas as determinações de índice de acidez, teor de água, e estabilidade à oxidação a 110 °C. No entanto, caso a diferença seja superior a 3,0 kg/m³, todos os parâmetros de qualidade da resolução deverão ser reavaliados (LÔBO *et al.*, 2009).

1.2.5 Ponto de Fulgor

Ponto de fulgor é a menor temperatura na qual um líquido combustível se vaporiza em quantidades suficientes para formar com o ar uma mistura capaz de inflamar-se momentaneamente quando aplicada uma chama sobre a mesma (Costa Neto *et al.*, 2000). Segundo o informado pela CETESB (2008), quando removida a fonte externa de calor, a combustão é interrompida, pois os vapores liberados a temperatura de fulgor não são suficientes para manter a combustão. Para Alptekin e Canacki (2009), o ponto de fulgor é uma das mais importantes vantagens do biodiesel, já que seu ponto de fulgor é superior ao do diesel mineral, e quando misturado ao diesel torna o combustível mais seguro durante o transporte, manuseio e estocagem.

A RANP 14/12 adotou o método de determinação do ponto de fulgor pelo equipamento de vaso fechado Pensky-Martens (NBR 14598), determinando como aceitável a temperatura mínima de fulgor de 100°C. Esta resolução indica como alternativa o método ASTM D93.

1.3 **Quimiometria**

A Quimiometria é uma área especificamente destinada à análise de dados químicos de natureza multivariada. Como citado por Kowalski *apud* Pimentel *et al.* (1999) "as ferramentas quimiométricas são veículos que podem auxiliar os químicos a se moverem mais eficientemente na direção do maior conhecimento". Com a utilização da quimiometria é possível planejar ou selecionar o número de experimentos ótimo, de forma que estes possam fornecer o máximo de informação química com a análise dos dados obtidos.

Através do *site* de discussão sobre quimiometria (Mundo da Quimiometria, 2010), foi possível elaborar o seguinte histórico:

- 1971 – O Professor Svante Wold (UMEA University, Suécia) cria a palavra *Kemometri* (sueco), em inglês: *Chemometrics*, em português: Quimiometria.
- 1972 – O primeiro grupo (*Research Group of Chemometrics*) foi criado pelo Prof. Svante Wold.
- 1973 – O primeiro artigo com o termo *chemometrics* é publicado pelo grupo de quimiometria do Prof. Svante Wold.
- 1974 – O Prof. Svante Wold cria a primeira definição de quimiometria: “A arte de extrair a informação química relevante de dados produzidos em experimentos químicos é chamada de quimiometria em analogia a biometria e econometria.” Neste mesmo ano a Sociedade Internacional de Quimiometria foi fundada pelo Prof. Svante Wold (UMEA University) e Bruce R. Kowalski (University of Washington).
- 1975 – o segundo artigo contendo o termo *chemometrics* em seu título é publicado pelo Grupo de Pesquisa em Quimiometria do Prof. Bruce R. Kowalski, este artigo define quimiometria como uma nova disciplina da Química.
- 1976 – acontece o primeiro Simpósio de Quimiometria, ‘*Chemometrics: Theory and Applications*’, em São Francisco, USA.
- 1977 – é publicado o primeiro livro sobre *chemometrics* (quimiometria), resultado do *first Symposium on Chemometrics: Theory and Applications*.
- 1978 – ocorre “*The First International Conference on Chemometrics in Analytical Chemistry*” – CAC I, em Petten, Noruega (15 – 17 de Setembro).
- 1980 – o Prof. Kowalski ministrou o primeiro curso de quimiometria no Brasil, no Instituto de Química da Unicamp.
- 1986 – é fundada a primeira revista de quimiometria, *Chemometrics and Intelligent Laboratory Systems*, da Elsevier.

Desde então inúmeros estudos com abordagem quimiométrica têm sido desenvolvidos e aprimorados no Brasil e no mundo.

1.4 Pesquisas na área de previsão de propriedades de biodiesel utilizando ferramentas quimiométricas e espectros de infravermelho

Viera (2011) avaliou o uso da espectroscopia no infravermelho médio por reflexão total atenuada (FTIR-ATR) para qualificação e quantificação de adulterantes em misturas de biodiesel/diesel na faixa de 4000 a 650 cm^{-1} . A autora utilizou a análise de componentes principais (PCA) para avaliar o comportamento dos dados e remover possíveis amostras atípicas, dentro de um conjunto de 81 amostras de misturas binárias contendo biodiesel e diesel (interior e metropolitano) e 60 amostras de misturas ternárias contendo biodiesel, diesel (interior e metropolitano) e adulterantes (óleo de girassol bruto, óleo de soja degomado e óleo residual de fritura). Para selecionar e avaliar as faixas espectrais obtidas por FTIR-ATR foi utilizado o algoritmo análise de componentes principais por intervalos (iPCA). De acordo com os resultados obtidos por Viera (2011), foi possível visualizar a separação dos grupos formados pelos óleos vegetais (adulterantes) das misturas contendo biodiesel e separar as amostras contendo diesel interior e metropolitano. Os mesmos espectros FTIR-ATR foram submetidos à análise de regressão por mínimos quadrados parciais (PLS) para quantificar a presença de adulterantes. As técnicas de seleção de variáveis utilizadas foram mínimos quadrados parciais por intervalos (iPLS), mínimos quadrados parciais por sinergismo de intervalos (siPLS) e mínimos quadrados parciais por exclusão de intervalos (biPLS). O melhor modelo de previsão de adulterantes do biodiesel foi obtido utilizando o siPLS para seleção de variáveis, o PLS como método de regressão e o autoescalamento como pré-processamento, com sete variáveis latentes apresentando coeficiente de determinação da previsão (R^2_{Pred}) e valor da raiz quadrada do erro médio quadrático de previsão (RMSEP) iguais a 0,991 e 0,424 (% v/v), respectivamente. Deste modo foi demonstrada a aplicabilidade do espectro MIR associado à estratégia de regressão por PLS para quantificação de adulterantes em misturas de biodiesel/diesel, na faixa de concentração de 0,5 a 30% (em volume) para o biodiesel e adulterantes.

No trabalho de Vasconcelos (2011) foi avaliado o uso de espectro NIR para classificação e determinação do teor de biodiesel em misturas diesel/biodiesel adulteradas com óleo vegetal (*in natura*), com amostras fornecidas pela TRANSPETRO. A análise de componentes principais (PCA) e a análise

discriminante linear (LDA) com seleção de variáveis pelo algoritmo das projeções sucessivas (SPA) foram utilizadas na identificação de adulteração de misturas diesel/biodiesel com óleo vegetal. Os modelos de calibração multivariada utilizando regressão por mínimos quadrados parciais por intervalos (iPLS), regressão por mínimos quadrados parciais com coeficientes de regressão selecionados pelo algoritmo Jack-Knife (PLS-JK) e regressão linear múltipla com seleção de variáveis pelo algoritmo das projeções sucessivas (MLR-SPA) foram empregados para determinação do teor de biodiesel. Os resultados obtidos demonstraram que o espectro NIR aliado à estratégia de classificação LDA-SPA pode ser utilizada em estudos preliminares para detecção de adulteração de misturas diesel/biodiesel por óleos vegetais *in natura* com índice de acerto de 86% para os caminhos óticos de 10 e 50 mm, respectivamente. Além disso, foi demonstrada a aplicabilidade do espectro NIR associado à estratégia de regressão MLR-SPA para determinação do teor de biodiesel em misturas diesel/biodiesel na presença de óleos vegetais *in natura*, na faixa de 0,0 a 10,0% v/v.

Monteiro *et al.* (2009) tiveram como alvo de seu estudo a avaliação da qualidade das misturas de biodiesel-diesel e a previsão da concentração de óleo e biodiesel no diesel utilizando espectrometria NMR-1H. Foram analisadas 46 amostras de mistura de óleo-diesel, biodiesel-diesel e óleo-biodiesel-diesel, sendo que 36 amostras pertenciam ao conjunto de calibração, restando as demais amostras para o conjunto de validação. Uma análise exploratória foi realizada, empregando análise de componentes principais (PCA) e as amostras atípicas foram retiradas do conjunto de dados. Os modelos de calibração multivariada PCR e PLS foram desenvolvidos para determinação do teor de óleo e biodiesel no diesel mineral. Para as amostras mais concentradas os valores previstos foram bastante semelhantes aos de referências, devido aos baixos erros relativos apresentados. O contrário aconteceu com as amostras com baixas concentrações, que apresentaram valores elevados de desvio. Os modelos obtidos por PCR e PLS foram úteis para a previsão do teor de biodiesel em misturas de biodiesel-diesel, particularmente adequados para a previsão de concentrações superiores a 2%. Para a previsão do teor de biodiesel o modelo por PLS obteve coeficiente de correlação de validação (R_{cv}) igual a 0,98 e erro padrão de validação (SEV) igual a 0,82 (% v/v).

Baptista *et al.* (2008(a)) descreveram a utilização de espectroscopia de infravermelho próximo (NIR) para determinar o teor de ésteres de biodiesel, assim

como o teor em ésteres metílicos de ácido linolênico (C18:3), mirístico (C14:0), palmítico (C16:0), esteárico (C18:0), oleico (C18:1) e linoleico (C18:2). A Análise dos componentes principais (PCA) foi utilizada para a análise qualitativa dos espectros, enquanto a regressão parcial dos mínimos quadrados (PLS) foi utilizada para determinar o teor total e individual dos ésteres. Alguns métodos de pré-processamento (correção do espalhamento multiplicativo, variação normal padronizada, linha de base, derivada de 1^a e 2^a ordem) foram aplicadas aos espectros. Para o modelo de teor de éster, os melhores resultados foram obtidos com o modelo PLS, com seis variáveis latentes, que forneceu um coeficiente de determinação de previsão (R^2_{Pred}) 0,930 e RMSEP igual a 0,9%. Com um conjunto de calibração contendo 182 amostras e um conjunto de validação contendo 93 amostras, Baptista *et al.* (2008(a)) conseguiram prever o teor do éster C18:3 utilizando o PLS com três variáveis latentes, com R^2_{Pred} igual 0,991 e RMSEP de 0,18%. Os melhores modelos de previsão para ésteres metílicos de ácido mirístico (C14:0), oléico (C18:1) e linoléico (C18:2) foram obtidos utilizando o modelo PLS com seis variáveis latentes, que forneceu R^2_{Pred} de 0,993; 0,914 e 0,921 e RMSEP 0,02%, 1,79% e 2,50%, respectivamente. O melhor resultado para a previsão do éster metílico de ácido esteárico (C18:0) foi obtido utilizando o modelo PLS com seis variáveis latentes, R^2_{Pred} de 0,852 e RMSEP 0,22%. Para o éster metílico de ácido palmítico (C16:0), o melhor resultado foi alcançado utilizando o modelo PLS com quatro variáveis latentes, R^2_{Pred} de 0,957 e RMSEP 0,79% (BAPTISTA *et al.*, 2008(a)).

No trabalho de Baptista *et al.* (2008(b)) foi avaliado a utilização de espectroscopia de infravermelho próximo (NIR) para a previsão das seguintes propriedades do biodiesel: o índice de iodo, o ponto de entupimento do filtro a frio (PEFF), a viscosidade cinemática à temperatura de 40°C e a densidade a 15°C. A análise dos componentes principais (PCA) foi utilizada para a realização de uma análise exploratória dos dados e a regressão por mínimos quadrados parciais (PLS) para a previsão das propriedades. Foram preparadas 71 amostras de biodiesel produzidos a partir de misturas de óleo de soja, palma, canola e os óleos usados em frituras, sendo que 49 amostras foram destinadas ao conjunto de calibração e 22 amostras para validação. O modelo de previsão do PEFF do biodiesel foi construído utilizando o PLS com duas variáveis latentes. Foram obtidos um R^2_{Pred} de 0,951 e RMSEP 1,0°C para esta propriedade. O melhor modelo de previsão para a

densidade a 15°C foi obtido por PLS e fez uso de seis variáveis latentes. Este modelo apresentou um R^2_{Pred} de 0,999 e RMSEP de 0,9 kg/m³.

O conjunto de dados usado por Lira *et al.* (2010(a)) consistiu de misturas preparadas utilizando ésteres metílicos de soja, óleo de rícino, óleo de girassol, óleo de sementes de algodão e óleo de canola, bem como amostras de diesel de diferentes regiões do Brasil (Pernambuco, Alagoas e Sergipe). Os autores prepararam um total de 161 amostras para o estudo da previsão de densidade, teor de enxofre e temperaturas de destilação, com suas respectivas medições espectrais no NIR e Mid-IR. Os dados foram divididos em dois tipos de conjuntos, de calibração (80% das amostras) e de validação (20% das amostras). Todos os modelos foram obtidos utilizando a regressão dos mínimos quadrados parciais (PLS) e os melhores resultados foram obtidos com o emprego da primeira derivada após a suavização dos sinais espectrais. Para realizar a previsão da densidade utilizando os dados espectrais de NIR, foram utilizadas oito variáveis latentes para geração do modelo PLS. Os autores obtiveram um R^2_{Pred} de 0,99 e RMSEP de 0,56 kg/m³. Para realizar a previsão da densidade utilizando os dados espectrais de Mid-IR, foram utilizadas oito variáveis latentes para geração do modelo PLS. Os autores obtiveram um R^2_{Pred} igual a 0,97 e RMSEP de 1,60 kg/m³.

Pereira (2007) propôs uma nova metodologia para a determinação simultânea do índice de acidez, índice de refração e viscosidade em quatro tipos de óleos vegetais comestíveis (soja, milho, girassol e canola). Para este propósito, o autor empregou a combinação da espectrometria NIR com um modelo por regressão linear múltipla com seleção de variáveis pelo algoritmo das projeções sucessivas (MLR-SPA) e regressão por mínimos quadrados parciais por intervalos (iPLS). Os resultados obtidos mostraram que os modelos MLR-SPA foram mais eficientes e robustos, visto que apresentaram baixos valores de RMSEP 0,009 % m/m; 0,0003 e 0,8 mPa.s, respectivamente e bons valores para os coeficientes de correlação (0,94, 0,98 e 0,96) para acidez, índice de refração e viscosidade, respectivamente.

2 METODOLOGIA

2.1 Produção do biodiesel

Para a síntese do biodiesel obtido pela reação de transesterificação e para o processo de purificação foram empregados os reagentes apresentados na Tabela 5.

Neste estudo, o processo de produção do biodiesel foi dividido em quatro principais etapas: (i) preparo da matéria prima, (ii) reação de transesterificação, (iii) separação de fases e (iv) purificação do biodiesel. No entanto, três etapas complementares de recuperação de resíduos foram consideradas e realizadas, (v) a recuperação e desidratação do resíduo de álcool metílico, (vi) destilação da glicerina e (vii) recuperação dos sais Na_2SO_4 e KCl utilizados no processo de purificação.

Tabela 5 – Reagentes utilizados na produção do biodiesel

Reagentes	Fonte	Pureza (%)	Processo
Óleo de Soja Refinado	Marca Liza	-	Síntese
Óleo de Canola Refinado	Marca Liza	-	Síntese
Óleo de Girassol Refinado	Marca Liza	-	Síntese
Óleo de Milho Refinado	Marca Liza	-	Síntese
Álcool Metílico	VE TEC	99,8	Síntese
Hidróxido de Potássio	VE TEC	85,0	Síntese
Ácido Clorídrico	VE TEC	37,0	Purificação
Cloreto de Potássio	VE TEC	99,0	Purificação
Sulfato de Sódio	VE TEC	99,0	Purificação

Foram produzidos 10,5 litros de biodiesel de soja, 5 litros de biodiesel de canola, 4 litros de biodiesel de girassol e 4 litros de biodiesel de milho, totalizando 23,5 litros de biodiesel a ser utilizado na composição das misturas. Foram produzidos também 1 litro de biodiesel misto (25% v/v de óleo de soja, 25% v/v de óleo de canola, 25% v/v de girassol e 25% v/v de milho) e 6 litros de biodiesel de soja adicionais, que foram utilizados nos testes de validação da metodologia e de qualificação do equipamento de ponto de entupimento de filtro a frio. No teste de

qualificação do equipamento, os resultados dos testes foram coletados por um período de tempo, de forma a comprovar a consistência dos mesmos.

2.1.1 Preparo da matéria-prima

Os óleos refinados da marca Liza foram submetidos a um aquecimento a 45°C, por 20 minutos, antes de iniciar a reação de transesterificação, para favorecer a cinética da reação e assim obter uma síntese de boa qualidade.

2.1.2 Reação de transesterificação

A síntese, para obtenção do éster metílico, foi processada com razão molar de 6 mols de álcool metílico para 1 mol de óleo de soja, com 1% do catalisador básico (KOH).

Para a produção de metóxido de potássio (alcoolato) dissolveu-se hidróxido de potássio (KOH) em metanol com auxílio de agitação e controle de temperatura (45°C) até completa dissolução do KOH. O alcoolato produzido para a reação de transesterificação têm por finalidade alcançar um melhor rendimento da produção do éster.

A reação de transesterificação foi realizada em um reator com capacidade volumétrica de 2 litros, com agitação mecânica (280 rpm) e temperatura controlada de 45°C, conforme mostrado na Figura 2. No reator foi adicionado o óleo refinado (soja, milho, canola ou girassol) medido em uma proveta e aquecido em banho-maria, sob agitação constante até atingir a temperatura de 45°C. Em seguida, adicionou-se a solução de metóxido de potássio (alcoolato) recentemente preparada, mantendo a mistura reacional por 30 minutos a 45° C sob agitação. Após a adição do alcoolato foi observado mudança no aspecto da solução que se tornou mais densa e a coloração também sofreu alterações, tornando-se mais clara.

Após a transesterificação a solução foi transferida para um funil de separação para permitir a decantação e separação das fases: superior contendo biodiesel e inferior composta de glicerol, sabões, excesso de base e álcool (tempo de espera de separação de fases: 40 minutos). Recolheu-se a fase inferior contendo a glicerina e impurezas e o volume de biodiesel (fase superior) foi mantido no funil, onde ocorreu as etapas de purificação do biodiesel.

Figura 2 – Reator contendo massa reacional

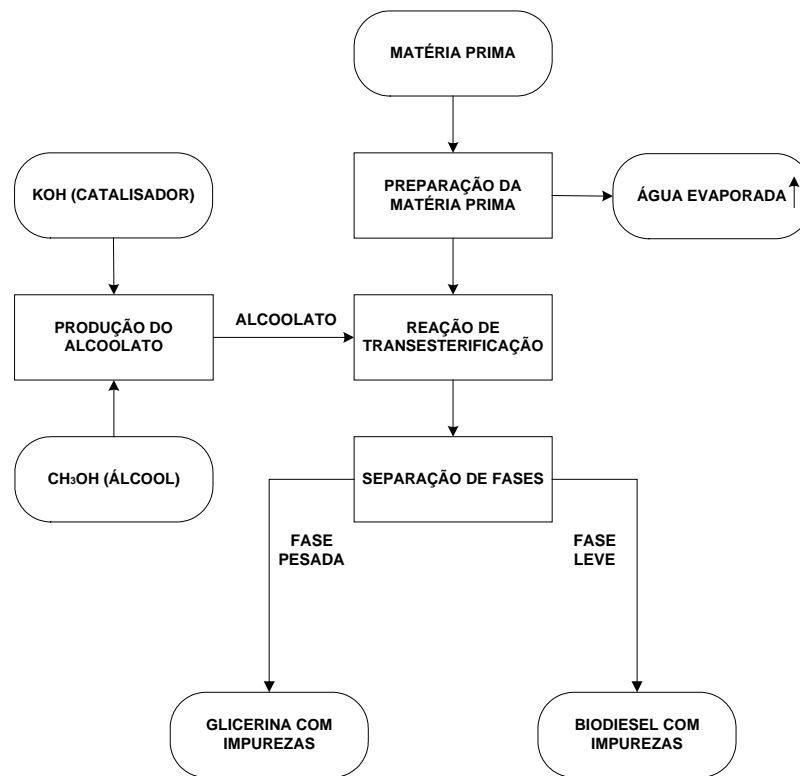


2.1.3 Separação de fases

A massa reacional produzida pela reação de transesterificação foi mantida no reator para decantar, e assim separar as fases. Após o tempo de 20 minutos, foi possível observar duas fases bem distintas no reator: uma fase rica em ésteres (menos densa e mais clara), e outra em glicerina (mais densa e mais escura). Após 90 minutos de repouso, a glicerina foi retirada pela abertura da válvula de fundo do reator e o biodiesel foi mantido no reator para as etapas de purificação.

As etapas de preparo da matéria-prima, reação de transesterificação e separação de fases podem ser observadas no fluxograma apresentado na Figura 3.

Figura 3 – Fluxograma do processo de síntese do biodiesel.



2.1.4 Purificação do biodiesel

A fase mais leve, ou seja, a fase que contém o biodiesel e impurezas, foi submetido ao processo de sucessivas lavagens, com o objetivo de purificar os ésteres presentes, removendo impurezas e álcalis residuais.

A primeira lavagem foi feita neutralizando o meio com uma solução de HCl a 0,5%v/v, mantendo uma agitação constante (280 rpm) por 10 minutos. Posteriormente, o sistema foi deixado em repouso de 90 minutos, sucedido por uma separação de fases, com o biodiesel compondo a fase superior e permanecendo no reator. A fase inferior, contendo solução de HCl mais impurezas foi descartada do processo e armazenada em frascos adequados para posterior tratamento desse resíduo.

Para a segunda lavagem foi adicionada uma solução saturada de cloreto de potássio (KCl), mantendo agitação constante (280 rpm) por 10 minutos, seguido de 90 minutos de repouso. A adição de solução saturada de KCl no processo de purificação do biodiesel, tem como objetivo a quebra da emulsão formada entre o

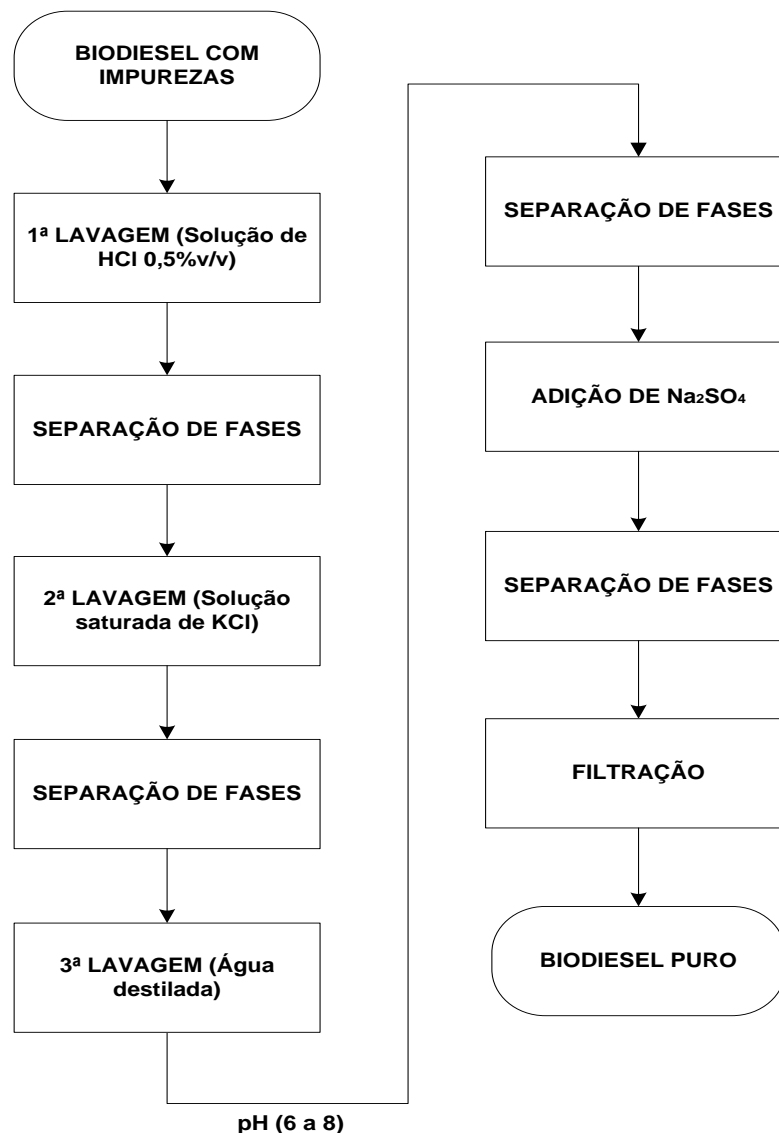
biodiesel e alguns resíduos ainda presentes no mesmo. Com a adição do sal, aumenta-se o “*salting-out effect*”, ou seja, a capacidade de solvatação com um eletrólito forte (no caso, o cloreto de potássio) e aumenta-se a tensão na interface entre as fases. A retirada de água também é facilitada neste processo. A fase inferior da separação de fases contendo solução de KCl mais impurezas, foi descartada e armazenada para posterior recuperação do KCl.

A terceira lavagem foi realizada com a adição de água destilada e agitação constante (280 rpm) por 10 minutos. Após 90 minutos de repouso foi observado o biodiesel compondo a fase superior e água leitosa com impurezas na fase inferior. Essa lavagem permitiu classificar o biodiesel como de boa qualidade, visto que, quando o combustível não é de boa qualidade a separação costuma ser lenta ou formar emulsão, com uma camada fina entre o biodiesel e a água. A neutralização do catalisador foi confirmada medindo o pH da água de lavagem, que apresentou resultado entre 6 e 8 em todos os lotes de biodiesel produzido.

A quarta etapa de purificação foi realizada com a adição de sulfato de sódio anidro (Na_2SO_4) para remover qualquer indício de traços de água presente no biodiesel, sucedida por uma separação de fases. Em seguida, o biodiesel foi submetido a uma etapa de filtração em funis preparados com duas camadas de algodão. Após a filtração, o éster foi armazenado em frascos adequadamente limpos.

As etapas de purificação descritas nos parágrafos anteriores estão representadas na Figura 4.

Figura 4 – Etapas de purificação do biodiesel



2.1.5 Estocagem adequada do biodiesel

Cada lote de biodiesel filtrado foi armazenado em frascos com tampa de rosca e capacidade de 1 litro. Os frascos foram preenchidos com biodiesel até quase o gargalo, para minimizar a quantidade de ar dentro dos frascos, e mantidos sob-refrigeração e ao abrigo da luz. Todos esses cuidados foram adotados para evitar a degradação e estender o prazo de validade do biodiesel. Foram retiradas amostras de cada lote de biodiesel. Lotes de mesma origem de biodiesel foram unificados, como por exemplo, todos os lotes de biodiesel de soja (10,5 litros) foram misturados

formando um lote unificado de biodiesel de soja. Esta unificação dos lotes foi feita também para os demais tipos de biodiesel estudados, tais como o de milho (4 litros), canola (5 litros) e girassol (4 litros).

2.2 Preparo das Misturas de Biodiesel

As misturas englobando as quatro diferentes origens de biodiesel produzidas foram preparadas em triplicata. Sendo assim, como foram estabelecidas 36 misturas distintas, foi necessário o preparo de 108 misturas no total. Nas Tabelas 6, 7 e 8 é apresentada a composição das misturas ternárias, binárias e quaternárias entre os óleos de biodiesel. Para compor as misturas foram utilizados 2 litros do biodiesel cedido por uma distribuidora da região sul. O biodiesel de canola, por apresentar valor de ponto de entupimento mais distinto entre os tipos de biodiesel produzidos neste estudo, foi o escolhido para compor a mistura com o biodiesel cedido por uma distribuidora da região sul.

Embora as misturas tenham sido realizadas em percentagem volumétrica, os valores em percentagem mássica também foram conservados para posterior utilização.

Tabela 6 – Composição das misturas ternárias

Misturas Ternárias					
Biodiesel_1	%v/v_1	Biodiesel_2	%v/v_2	Biodiesel_3	%v/v_3
Soja	50	Milho	40	Canola	10
Soja	40	Milho	30	Canola	30
Soja	50	Milho	10	Canola	40
Soja	50	Milho	40	Girassol	10
Soja	40	Milho	30	Girassol	30
Soja	50	Milho	10	Girassol	40
Soja	50	Canola	40	Girassol	10
Soja	40	Canola	30	Girassol	30
Soja	50	Canola	10	Girassol	40

Tabela 7 – Composição das misturas binárias

Misturas Binárias			
Biodiesel_1	%v/v_1	Biodiesel_2	%v/v_2
Soja	10	Milho	90
Soja	30	Milho	70
Soja	50	Milho	50
Soja	70	Milho	30
Soja	90	Milho	10
Soja	10	Girassol	90
Soja	30	Girassol	70
Soja	50	Girassol	50
Soja	70	Girassol	30
Soja	90	Girassol	10
Soja	10	Canola	90
Soja	30	Canola	70
Soja	50	Canola	50
Soja	70	Canola	30
Soja	90	Canola	10
Canola	10	Região Sul	90
Canola	30	Região Sul	70
Canola	50	Região Sul	50
Canola	70	Região Sul	30
Canola	90	Região Sul	10

Tabela 8 – Composição das misturas quaternárias

Misturas Quaternárias							
Biodiesel_1	%v/v_1	Biodiesel_2	%v/v_2	Biodiesel_3	%v/v_3	Biodiesel_4	%v/v_4
Soja	25	Milho	25	Canola	25	Girassol	25
Soja	40	Milho	40	Canola	10	Girassol	10
Soja	40	Milho	10	Canola	40	Girassol	10
Soja	40	Milho	10	Canola	10	Girassol	40
Soja	10	Milho	40	Canola	40	Girassol	10
Soja	10	Milho	40	Canola	10	Girassol	40
Soja	10	Milho	10	Canola	40	Girassol	40

2.3 Recuperação dos Resíduos

Os resíduos gerados na produção do biodiesel, principalmente a glicerina, excesso de álcool metílico, o cloreto de potássio (KCl) e o sulfato de sódio (Na_2SO_4), foram recuperados como apresentado nos itens a seguir.

2.3.1 Destilação da Glicerina

A corrente de glicerina, fase mais densa proveniente do processo de transesterificação, contém impurezas e pode ser purificada para obter um valor de mercado mais favorável. O processo de purificação da glicerina consiste na remoção do álcool, no caso o metílico, seguida de uma destilação a vácuo, o que gera uma glicerina límpida e transparente. O produto de calda da destilação a vácuo, na faixa de 10% a 15% da massa da glicerina bruta, neste estudo foi denominado de resíduo glicérico. A Figura 5 ilustra as etapas de recuperação da glicerina advinda do processo de produção do biodiesel.

2.3.2 Recuperação e desidratação do álcool metílico

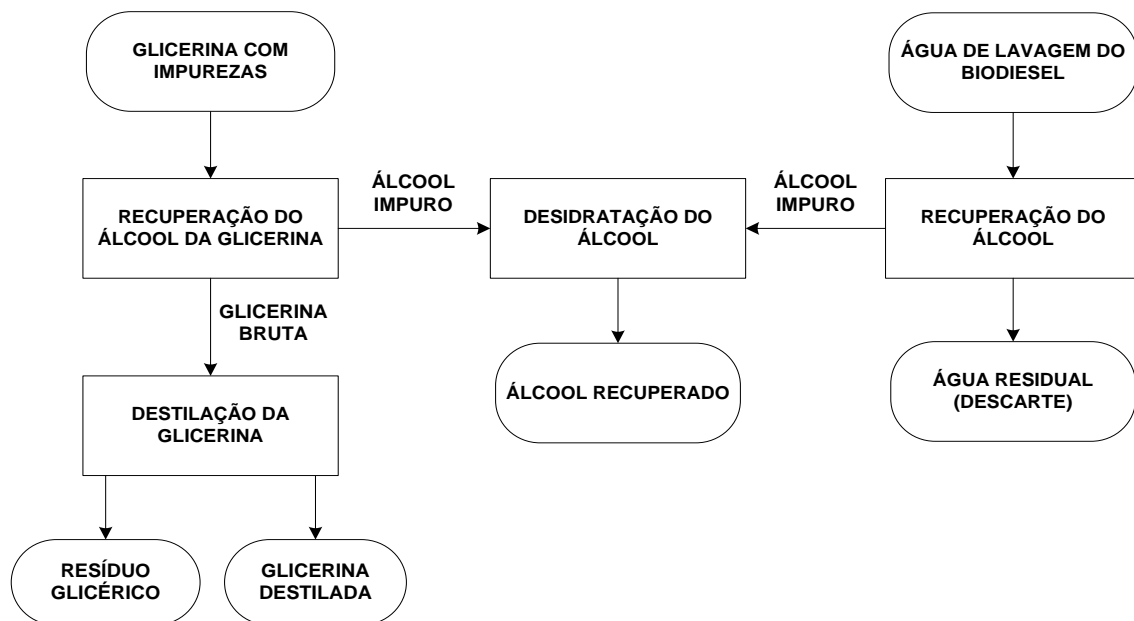
A fase de fundo, gerada na separação de fases após a reação de transesterificação, contém basicamente glicerina com excesso de metanol e

presença de impurezas. O álcool metílico, removido no processo de destilação da glicerina, foi submetido novamente a um processo de destilação, o qual serviu para desidratar o álcool, deixando-o apto para ser utilizado novamente no processo.

A fase de fundo, gerada na etapa de purificação do biodiesel com água destilada, contém basicamente água e excesso de álcool metílico. Estes podem ser removidos através de um processo de evaporação. Os vapores do álcool e água são liquefeitos em um condensador, onde o álcool residual é recuperado na fase de topo e encaminhado para um processo de desidratação, normalmente realizada por destilação, conforme processo proposto por Parente (2003). A Figura 5 apresenta de modo simplificado, as etapas de recuperação e desidratação do álcool metílico proveniente do processo de produção do biodiesel.

O álcool metílico recuperado pode entrar novamente na síntese do biodiesel como matéria-prima.

Figura 5 – Etapas de purificação da glicerina, recuperação e desidratação do álcool metílico



2.3.3 Recuperação do Sulfato de Sódio e Cloreto de Potássio

A recuperação do resíduo de sulfato de sódio (Na_2SO_4) oriundo da processo de purificação do biodiesel divide-se em duas principais etapas: recuperação do

Na_2SO_4 retido no algodão (no processo de filtração) e a extração do biodiesel da solução salina.

O algodão contendo o sulfato de sódio foi transferido para um béquer, adicionou-se água destilada e deixou-se em repouso durante 12 horas, para separar o biodiesel e dissolver o sulfato de sódio que ficou retido no algodão. Parte do sulfato de sódio (Na_2SO_4) precipitou e a outra parte ficou dissolvida na solução. Com a ajuda de uma pinça, os algodões foram espremidos na parede do béquer para retirada da parte úmida e posteriormente descartados.

A fase líquida da solução (biodiesel e solução de sulfato de sódio dissolvido) foi transferida para um funil de separação, agitada e deixada em repouso até a separação das fases. A fase inferior, contendo solução de Na_2SO_4 com impurezas, foi recolhida em um béquer e a fase superior, contendo o biodiesel, foi recolhida em outro béquer.

A parte sólida de sulfato de sódio (sulfato de sódio precipitado) foi transferida para um béquer, solubilizado em água destilada levemente aquecida para facilitar a dissolução do sal. Após a dissolução, a solução foi transferida para um funil de separação, agitada e deixada em repouso até a separação das fases. A fase inferior, contendo solução de Na_2SO_4 com impurezas, foi recolhida em um béquer e a fase superior, contendo o biodiesel, foi recolhida em outro béquer.

Todas as fases aquosas contendo Na_2SO_4 foram transferidas para um único funil de separação e lavadas por quatro vezes com 50 ml de clorofórmio. A fase mais densa (ou inferior) contém o biodiesel dissolvido no clorofórmio e a fase mais leve (ou superior) contém uma solução turva de Na_2SO_4 . A fase inferior foi recolhida e adequadamente armazenada em um frasco limpo para posterior recuperação do clorofórmio.

A fase aquosa superior foi submetida a uma filtração em um funil com algodão para um béquer, adicionado 3 g de carvão ativo. Depois, a solução foi aquecida até alcançar a ebulição por 3 a 5 minutos. O carvão ativo auxiliou na retenção das impurezas e o biodiesel. A solução foi filtrada em um funil com algodão para uma travessa de vidro (tipo pirex), colocada em estufa a 110°C para evaporar a água e obter as placas de sulfato de sódio recristalizado. As placas de sulfato de sódio foram maceradas, utilizando o almofariz (gral e pistilo), até a obtenção de cristais. Novamente foi levado à estufa a 110°C por 3 horas para retirada da umidade residual e o Na_2SO_4 está pronto para o consumo. O mesmo processo pode ser

utilizado na recuperação do resíduo de cloreto de potássio, que foi utilizado na segunda lavagem da purificação do biodiesel.

O biodiesel com impurezas, proveniente do processo de recuperação do sulfato de sódio, foi recuperado lavando-o por três vezes com água destilada e descartando a fase aquosa. Ao biodiesel lavado foi acrescentado 50 g de sulfato de sódio anidro para absorção da água residual, seguido de uma filtração com algodão para a retirada do sulfato de sódio.

2.4 Medição das Propriedades

As propriedades físico-químicas medidas no presente estudo são massa específica, índice de refração, ponto de entupimento de filtro a frio (PEFF) e ponto de fulgor.

2.4.1 Massa Específica e Índice de Refração

Os ensaios de massa específica e índice de refração foram realizados simultaneamente no densímetro DM40 acoplado ao refratômetro RX40 com amostrador automático SC1, equipamentos da marca Mettler Toledo conforme apresentado na Figura 6. O procedimento de análise baseia-se na adição de aproximadamente 10 mL da amostra no frasco que acompanha o equipamento, que deve ser acomodado adequadamente no amostrador automático. Seleciona-se o método de análise e iniciam-se as medições. O método de análise foi construído com as seguintes características: temperatura de amostragem de 20°C; com configuração do amostrador SC1 ajustado para operar com velocidade automática, taxa de enchimento de 120% e máxima duração do enchimento de 30 segundos; a limpeza da linha foi ajustada para realizar a lavagem com dois solventes (álcool e acetona).

Para ambas as propriedades (massa específica e índice de refração) os resultados foram apresentados com 4 casas decimais, sendo que a unidade resposta para a massa específica foi g/cm^3 .

Figura 6 – Densímetro (DM40) acoplado ao refratômetro (RX40) com amostrador automático SC1



2.4.2 Ponto de Entupimento de Filtro a Frio

O ponto de entupimento de filtro a frio (PEFF) é a temperatura, em °C, na qual um determinado volume de amostra deixa de passar através de um filtro metálico padrão, num período de tempo específico, quando resfriado sob certas condições.

Os ensaios de PEFF foram realizados no equipamento da marca TANAKA, modelo AFP-102, conforme apresentado na Figura 7. O método baseia-se em resfriar, com uma taxa de 1°C/min, um volume de 45 ml de amostra, que é succionado para uma pipeta através de um filtro de malha metálico padronizado, sob vácuo controlado. Esse procedimento é repetido tantas vezes quanto possível até que a quantidade de cristais que se separam da solução seja suficiente para interromper ou reduzir a circulação da amostra pelo filtro. Alternativamente, este procedimento é repetido se o tempo necessário para preencher a pipeta ultrapasse 60 segundos ou ainda que a amostra deixe de retornar completamente para o recipiente de ensaio, antes de ser resfriado mais 1°C. A temperatura na qual a última filtração foi iniciada é o ponto de entupimento de filtro a frio.

Figura 7 – Equipamento para ensaio do ponto de entupimento de filtro a frio marca TANAKA modelo AFP-102.



2.4.3 Ponto de Fulgor

Os ensaios de ponto de fulgor foram determinados em um analisador da marca TANAKA, modelo APM-8, em vaso fechado, conforme apresentado na Figura 8. É um equipamento automático que possui um sistema de ignição a gás com acendimento automático e esfriamento automático com ar forçado por ventoinha. Cada ensaio foi realizado com 75 ml de amostra e o método utilizado foi o ASTM C-normal. O método baseia-se em aquecer a amostra sob agitação, com uma taxa de aquecimento de 5 a 6°C/min, até que se alcance uma temperatura 20°C antes da temperatura sugerida como temperatura expectativa. Neste ponto, se inicia a aplicação da chama no vaso a cada 2°C até o momento de atingir o ponto de fulgor, quando ocorre a liberação de um chama. A procura do ponto do fulgor se estende até 20°C depois da temperatura expectativa, temperatura a qual se espera que seja a do ponto de fulgor.

Figura 8 – Ponto de fulgor da marca TANAKA modelo APM-8



2.5 Avaliação de Composição por Cromatografia

O método de padronização externa foi utilizado para comparar a área dos ésteres (oleato e palmitato de metila) a serem quantificados nas amostras com as áreas obtidas com soluções de concentrações conhecidas preparadas a partir de um padrão contendo oleato e palmitato de metila. Soluções de oleato de metila e palmitato de metila foram preparadas em diversas concentrações. Posteriormente, foram obtidos os cromatogramas correspondentes a cada uma amostras, seguida da construção do gráfico relacionando a área obtida com a concentração do analito. Após a construção dos gráficos foi possível obter uma equação da curva analítica resultante para o oleato de metila e outra para o palmitato de metila, o que permitiu inferir a concentração destes ésteres na amostra a partir da área dos ésteres obtidos no cromatograma resultante de uma injeção de uma amostra separada.

As análises foram realizadas em um cromatógrafo de fase gasosa Varian (modelo 450) acoplado a um espectrômetro de massa Varian (modelo 220), com analisador do tipo armadilha de íons e o método de ionização por impacto de elétrons (Figura 9). As injeções foram realizadas no modo “*Splitless*” utilizando uma seringa de 10 μ L e um injetor automático combiPAL. Para separação dos analitos foi

empregado uma coluna capilar de sílica fundida da marca SUPELCO modelo Omegawax™250 (comprimento: 30m, diâmetro: 0,25mm e espessura do filme: 0,25mm), cujo limite de temperatura é 300°C.

A temperatura do forno da coluna iniciou-se a 60°C com uma rampa de aquecimento de 5 °C/min até alcançar uma temperatura de 240°C. A injeção foi realizada no modo “*splitless*” com volume de injeção de 1 µL, a uma velocidade de injeção de 5 µL/min e com a temperatura de 280°C.

Todas as amostras puras e as misturas de biodiesel foram preparadas, realizando duas diluições sucessivas em acetonitrila (100 µL da amostra diluída em um balão de 10 mL de solvente).

Figura 9 – Cromatógrafo de fase gasosa Varian (modelo 450) acoplado a um espectrômetro de massa Varian (modelo 220).



2.6 Medição dos espectros de infravermelho

Foram realizadas medições de espectros de infravermelho médio e próximo dos óleos de biodiesel e suas misturas.

2.6.1 Espectros de infravermelho médio

Os espectros de infravermelho médio (Mid-IR) das misturas entre os diferentes óleos de biodiesel foram adquiridos na faixa de 7.800 a 450 cm^{-1} , com resolução de 4 cm^{-1} , intervalos de dados de 1 cm^{-1} e 20 varreduras, utilizando um acessório de reflectância total atenuada (ATR) horizontal com cristal de seleneto de zinco (ZnSe) da PIKE Technologies acoplado a um espectrofotômetro de infravermelho da marca Perkin Elmer modelo FT-IR/NIR Spectrometer Frontier (Figura 10).

As amostras das misturas dos diferentes óleos de biodiesel foram analisadas colocando-se $0,2\text{ mL}$ de cada mistura no porta-amostra (cristal). Foram obtidos os espectros de infravermelho médio das misturas e do biodiesel puro.

Para a limpeza da superfície do cristal foi usado lenços de papel umedecido em água destilada, seguido por lenço de papel umedecido em álcool etílico e após a evaporação do álcool etílico foi adquirido o espectro do branco para certificar-se da ausência de resíduos e ou de contaminantes no cristal.

Figura 10 – Espectrofotômetro de infravermelho com acessório ATR



2.6.2 Espectros de Infravermelho Próximo

Os espectros de NIR foram adquiridos na faixa de 10.000 a 4.000 cm^{-1} , com resolução de 4 cm^{-1} , intervalos de dados de 1 cm^{-1} , 20 varreduras e espessura do filme da amostra de 0,224 mm. Utilizou-se um acessório para transfectância no infravermelho próximo (NIRA) com cristal de seleneto de zinco (ZnSe) da Perkin-Elmer acoplado a um espectrofotômetro de infravermelho, modelo FT-IR/NIR Spectrometer Frontier, do mesmo fabricante (Figura 11).

O porta-amostra utilizado nas medições consiste de uma pequena placa de Petri e um limitador da espessura do filme da amostra. A limpeza do porta-amostra foi realizada com uma lavagem com água e detergente, seguida de um enxague com água destilada, álcool etílico e colocado em estufa para secar.

Figura 11 – Espectrofotômetro de infravermelho com acessório NIRA



2.7 **Tratamento dos Espectros**

Os espectros podem apresentar variações sistemáticas que não são explicadas pela concentração do analito em questão, como ruídos aleatórios, espalhamento de luz e desvios da linha de base, conforme relatado por Menezes *et al.* (2009) e Zhao *et al.* (2010). Assim, o pré-tratamento do espectro pode trazer

melhorias na interpretação e na utilização do mesmo, além de reduzir a complexibilidade do modelo, tornando-o muitas vezes mais robusto e confiável contra variações indesejáveis em novas amostras.

2.7.1 Suavização

O filtro de suavização Savitzky-Golay é um tipo de filtro descrito pela primeira vez em 1964 por Abraham Savitzky e Marcel J. E. Golay, sendo concebido para analisar os picos de absorção no espectro e remover os ruídos (SAVITZKY e GOLAY, 1964). Os espectros medidos geralmente podem apresentar ruídos aleatórios, tanto no início quanto no fim dos espectros. O principal objetivo da suavização é reduzir o nível de ruído mantendo detalhes importantes do espectro.

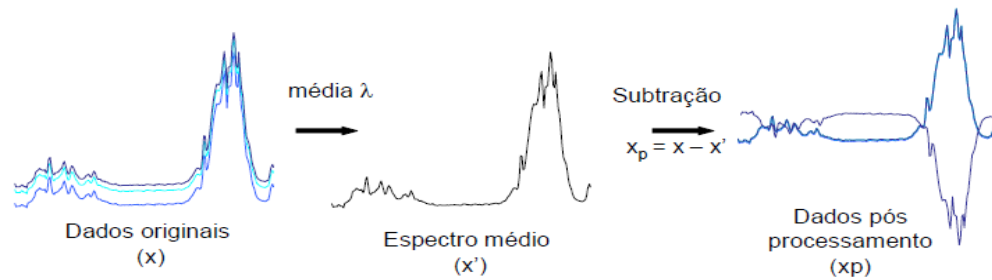
O método executa essencialmente uma regressão polinomial local, de grau k , relativa a uma série de valores, de pelo menos $k+1$ pontos, que são tratados como sendo igualmente espaçados na série, para determinar o valor de suavização para cada ponto. A principal vantagem da suavização é que ela tende a preservar as características de distribuição, tais como larguras, máximos e mínimos relativos, os quais são geralmente "achatados" por outras técnicas de nivelamento (MADDEN, 1978; BRERETON, 2003).

2.7.2 Centralização na média

Para dados espectrais, a centralização ou remoção dos dados na média é o mais utilizado, podendo ser combinada com outros tipos de pré-processamento.

Centralizar os dados na média é gerar um novo conjunto de dados (x'), que exhibe o resultado do cálculo da diferença de cada dado em relação à média do conjunto original (x), conforme Equação 3. Esta operação de centralização da média remove a intensidade absoluta da informação de cada variável, tornando-se possível focar-se nas variações da resposta (Figura 12). O objetivo da remoção da média é eliminar dos dados o valor da intensidade de cada variável revelando flutuações dos dados em torno do valor médio (NAES *et al.*, 2002).

Figura 12 – Esquema representativo da aplicação de pré-processamento centralização dos dados na média aplicado em espectros NIR. Fonte: Sauer (2007).



$$x_p = x_i - x' \quad (3)$$

2.7.3 Correção de linha de base

A primeira etapa, antes de aplicar a maioria dos métodos quimiométricos ou de utilizar outros pré-processamentos, é muitas vezes a correção da linha de base, especialmente quando se utiliza instrumentos mais antigos. A razão para isto é que as abordagens quimiométricas olham a variação acima de uma linha de base, assim se a correção de linha de base não for realizada, ruídos indesejáveis vão dificultar a análise dos espectros e a construção de modelos. Esse ajuste pode envolver simplesmente a média, ou então se aplica um melhor ajuste linear ou polinomial. Às vezes, a linha de base é calculada ao longo de todo o espectro, em outros casos, separadamente para cada região espectral. Desta forma, obtém-se um modelo matemático simples, e, em seguida é necessário subtrair a linha de base a partir de toda a região de interesse (BRERETON, 2003).

A correção de linha de base é utilizada para ajustar o espectro, de modo que não apresente valores negativos. Desta forma, para cada amostra o valor do menor ponto no espectro é subtraído de todas as variáveis, obrigando que o menor valor do conjunto final seja zero e os demais assumam sinal positivo (GAYDOU *et al.*, 2010; BRERETON, 2003). Após a realização da correção da linha de base, um sinal com um pico bem definido é obtido, lembrando que ao final do processo o número de variáveis originais não é alterado (TEÓFILO, 2013). Este pré-processamento é aplicado conforme demonstrado na Equação 4. Onde x é uma variável e X são todas as variáveis selecionadas para determinada amostra.

$$f(x) = x - \min(X) \quad (4)$$

2.7.4 Derivada de 1ª ordem

O cálculo de primeira derivada provoca significativas mudanças na informação processada, podendo apresentar um efeito positivo, ao evidenciar um sinal analítico de interesse, ou negativo, ao evidenciar ruído instrumental. A derivação também tem como efeito o deslocamento ou alinhamento de linha de base, como pode ser observado na Figura 13.

Figura 13 – Esquema representativo da aplicação de pré-processamento de cálculo de primeira derivada em espectros NIR. Fonte: Sauer (2007).



2.7.5 Correção de sinal multiplicativo

A correção de sinal multiplicativo (*Multiplicative Signal Correction*, MSC) é usada em aplicações onde existem variações multiplicativas entre os perfis de resposta das amostras, removendo a influência de efeitos físicos nos espectros, tais como o tamanho de partícula, a rugosidade e opacidade, os quais não trazem informações químicas sobre as amostras e introduz variações espectrais, como o deslocamento da linha de base. Segundo Naes *et al.* (2002) estas variações podem ser causadas por diferenças no percurso óptico das amostras ou em consequência de processos de dispersão de luz. O modelo de pré-processamento MSC assume que qualquer espectro da amostra pode ser estimado como um múltiplo de um espectro de referência mais um valor aditivo, sendo muitas vezes o espectro de referência considerado como o espectro médio do conjunto de amostras.

Segundo Souza e Poppi (2012), para fazer a correção, o método MSC assume que cada espectro é determinado pelas características químicas da amostra somadas às características físicas indesejadas, obedecendo a Equação 5. Onde x_{ik} (original) e x_{ik} (transformado) são os valores de absorvância antes e depois de correção com o MSC em cada comprimento de onda, a_i e b_i são constantes estimadas a partir de uma regressão por mínimos quadrados de um espectro individual x_{ik} contra um espectro médio do conjunto de calibração em todos os comprimentos de onda, sendo descrito pela Equação 6 (RINNAN *et al.*, 2009). Neste caso, e_{ik} corresponde a todos os outros efeitos nos espectros que não foram modelados.

$$x_{ik}(\text{transformado}) = \frac{x_{ik}(\text{original}) - a_i}{b_i} \quad (5)$$

$$x_{ik} = a_i + b_i \bar{x} + e_{ik} \quad (6)$$

2.7.6 Varição normal padronizada

A variação normal padronizada (*Standard Normal Variate*, SNV) é utilizada em muitas aplicações espectrométricas. Este pré-processamento realiza um ajuste aditivo e multiplicativo tal como o MSC, embora os fatores de correção sejam determinados de modo diferente. Para cada espectro da amostra o desvio ajustado é simplesmente a média de todos os valores sobre todas as variáveis e o ajuste multiplicativo é o desvio padrão dos valores de todas as variáveis. O SNV melhora a precisão de previsão, porém não simplifica o modelo e nem reduz as interferências sistemáticas (NAES *et al.*, 2002).

2.7.7 Correção de sinal ortogonal

A correção de sinal ortogonal (*Orthogonal Signal Correction*, OSC), foi desenvolvida para reduzir o número de fatores em regressões multivariadas, mas passou também a ser utilizada na transferência de calibração entre instrumentos,

visto que sua utilização evita perda de informação durante a etapa do pré-processamento dos espectros (WOLD *et al.*, 1998; PIERNA *et al.*, 2001; SJOBLÖM *et al.*, 1998). O principal objetivo da OSC é reduzir a variação na matriz que contém os espectros (\mathbf{X}) que não seja ortogonal a matriz das propriedades (\mathbf{Y}) a serem preditas no modelo, simplificando assim o modelo. Quando utilizada com calibração multivariada, a OSC é calculada apenas com as amostras de calibração.

Segundo Wold *et al.* (1998) para realizar o cálculo das componentes OSC, a primeira etapa é ortogonalizar a primeira componente principal (PC1) de \mathbf{X}_t em relação a \mathbf{Y} , obedecendo a Equação 7, onde \mathbf{S} representa o vetor dos escores das amostras em PC1. Esta etapa tem como objetivo obter a maior variância possível nos dados que não esteja correlacionada com a informação química.

$$\mathbf{S}_{ort} = \mathbf{S} - \mathbf{Y}(\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{S} \quad (7)$$

Honorato *et al.* (2007) descreveram que a partir de uma regressão PLS de \mathbf{S}_{ort} contra \mathbf{X}_t , é possível obter o vetor dos pesos \mathbf{W}_{ort} que minimiza a covariância entre \mathbf{X}_t e \mathbf{Y} , visto que este vetor quase não contém informação química relevante (Equação 8). Os pesos então são utilizados para atualizar o vetor ortogonal dos escores (\mathbf{S}_{ort}), que será utilizado no cálculo do vetor correspondente dos pesos (\mathbf{P}_{ort}), conforme apresentado na Equação 9.

$$\mathbf{S}_{ort} = \mathbf{X}_t \mathbf{W}_{ort} \quad (8)$$

$$\mathbf{P}_{ort} = \frac{\mathbf{X}_t^T \mathbf{S}_{ort}}{\mathbf{S}_{ort}^T \mathbf{S}_{ort}} \quad (9)$$

A primeira componente OSC é subtraída de \mathbf{X}_t , fornecendo a matriz de resíduos \mathbf{E}_t . Esta última possui uma melhor relação sinal/ruído, devendo ser usada para substituir \mathbf{X}_t e o procedimento é repetido tantas vezes até que o número desejável de componentes OSC seja removido (HONORATO *et al.*, 2007).

2.7.8 Autoescalamento

O propósito do autoescalamento dos dados originais é representar cada observação em termos de variações características do sistema, fazendo com que todas as variáveis possam desempenhar influências equitativas nos resultados além de torná-las adimensionais.

Para Lyra *et al.* (2010) e Moita Neto e Moita (1998), a matriz de dados contendo os valores das propriedades dos elementos podem apresentar variáveis com significados físicos, magnitudes e unidades distintas, que não podem ser combinados sem antes passar por uma etapa de padronização. Quando este problema é identificado, uma maneira de resolver este problema mantendo a informação estatística dos dados, é submeter a matriz a uma etapa de pré-processamento e o pré-processamento mais indicado nestes casos é o autoescalamento, onde a matriz é centrada na média dos valores subtraindo-se o valor de cada elemento da matriz da média de cada variável (coluna) e, em seguida, é normalizada pelo desvio padrão dividindo-se o valor de cada elemento centrado na média pelo desvio padrão da variável.

2.8 **Amostras Atípicas**

Na construção de um modelo podem existir algumas amostras que não se ajustam adequadamente ao modelo devido a fenômenos inevitáveis que podem acompanhar uma análise quantitativa como erros de operação, ruídos aleatórios, alterações nas respostas de equipamentos, dentre outras característica (MARTENS & NAES, 1989). Estas amostras são denominadas atípicas e sua influência pode comprometer o desempenho da calibração multivariada. A identificação e eliminação dessas amostras contribuem para a qualidade do modelo, porém é necessário ter critério no descarte de amostras, visto que algumas podem conter informações pertinentes ao sistema estudado.

Neste estudo foi necessário retirar a amostra Am158 do conjunto de espectros de Mid-IR, que é uma amostra referente a uma mistura quaternária entre quatro biodiesel com origens diferentes (10% v/v de biodiesel de soja, 10%v/v de biodiesel de milho, 40%v/v de biodiesel de canola e 40% v/v de biodiesel de

girassol), devido a discrepância de comportamento quando comparada às outras duas amostras de mesma composição.

2.9 Avaliação dos Componentes Principais

O método PCA não é um método de regressão, no entanto é uma ferramenta muito utilizada que permite uma análise prévia dos dados.

2.9.1 PCA

A análise de componentes principais (*Principal Component Analysis*, PCA) é um método não supervisionado de análise exploratória, que tem como objetivos visualizar a estrutura dos dados, encontrar similaridades entre amostras, detectar amostras atípicas e principalmente reduzir a dimensionalidade dos dados em poucos componentes principais (*Principal Components*, PC), que representam o conjunto de dados original utilizando uma combinação de variáveis que descrevem a tendência dos dados (FERRER-RIQUELME (2009); DE SOUZA e POPPI (2012)).

No PCA a matriz \mathbf{X} é decomposta em um produto de duas matrizes, pesos (\mathbf{P}) e escores (\mathbf{T}), mais uma matriz de erros (\mathbf{E}), como apresentado na Equação 10. Onde \mathbf{X} é a matriz ($m \times n$), \mathbf{T} é a matriz de escores ($m \times N^{\circ}\text{PCs}$), \mathbf{P}^t é a matriz transposta dos pesos ($N^{\circ}\text{PCs} \times n$) e \mathbf{E} é a matriz dos erros ($m \times n$).

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^t + \mathbf{E} \quad (10)$$

Os componentes principais (PC) são os representantes do conjunto de dados originais, este é definido por um par de pesos e escores. Os componentes principais descrevem o máximo de variância e são calculados em ordem de importância. Eles podem classificar amostras, identificar espectros de substâncias e determinar as variáveis mais significativas do processo. Além disso, os componentes principais são ortogonais entre si e por isso, não são correlacionados.

Segundo De Souza e Poppi (2012), cada componente principal é constituído pela combinação linear das variáveis originais e os coeficientes dessa combinação

são chamados de pesos. Os pesos são calculados como sendo os cossenos dos ângulos entre as variáveis originais e os componentes principais (PC), representando, deste modo, o quanto cada variável original contribui para determinado PC.

O primeiro componente principal (PC1) é traçado no sentido da maior variação no conjunto de dados e o segundo (PC2) é traçado ortogonalmente ao primeiro, com a finalidade de explicar a maior porcentagem da variação não explicada pela PC1 e assim por diante para outros componentes principais que sejam necessários para descrever o comportamento dos dados, conforme observado por De Souza e Poppi (2012).

A avaliação do gráfico dos escores pode, principalmente, informar a presença de relações de similaridade entre as amostras e a avaliação do gráfico dos pesos permite entender quais variáveis mais contribuem para os agrupamentos observados no gráfico dos escores. Desta forma, a análise conjunta do gráfico de escores e pesos, permite a identificação das variáveis que são responsáveis pelas diferenças observadas entre as amostras. O número de componentes principais a ser utilizado no modelo PCA pode ser determinada pela porcentagem de variância explicada. Assim, seleciona-se um número de componentes de tal maneira que a maior porcentagem da variação presente no conjunto de dados originais seja capturada (DE SOUZA e POPPI, 2012). Para este estudo os pesos estão relacionados aos números de ondas (variáveis) e os escores estão relacionados às amostras.

Neste estudo o PCA está sendo utilizado como um método exploratório que auxilia na elaboração de hipóteses gerais a partir do comportamento dos dados coletados, contrastando com estudos direcionados nos quais hipóteses prévias são testadas (classificação), sendo capaz de separar as informações importantes das redundantes e aleatórias, o que facilita na identificação e retirada de amostras atípicas.

2.10 Algoritmo KS

O algoritmo de Kennard-Stone (KS) é o mais conhecido algoritmo para seleção de amostras entre os químicos. O algoritmo KS começa selecionando as duas amostras com a maior distância Euclidiana entre si no espaço X. Para cada

uma das amostras restantes, calcula-se a distância mínima com respeito às amostras já selecionadas. Feito isso, a amostra com a maior distância mínima é retida, e o procedimento é repetido até que um determinado número de amostras seja selecionado (KENNARD e STONE, 1969).

O algoritmo de Kennard-Stone (KS) foi utilizado na seleção das amostras do conjunto de validação e de calibração para o banco de dados espectrais de NIR e Mid-IR. A partir do conjunto original dos dados, o algoritmo divide o total de amostras em dois conjuntos: calibração que é responsável pela construção do modelo de regressão e o de validação que tem como objetivo avaliar a capacidade preditiva do modelo.

O conjunto de calibração e validação dos espectros NIR formados apresentaram em tamanho as respectivas matrizes, 100 x 6001 e 48 x 6001 (amostras x variáveis), sendo que o conjunto inicial dos espectros NIR apresentavam 148 amostras para 6001 variáveis. Para os conjuntos de calibração e validação dos espectros Mid-IR as matrizes construídas foram 100 x 3321 e 48 x 3321, sendo que o conjunto inicial apresentavam 148 amostras com 3321 variáveis cada.

2.11 Técnicas de Calibração Multivariada

As técnicas utilizadas neste estudo são os métodos de regressão de 1ª ordem com medições espectrométricas para a previsão do PEFF, massa específica, índice de refração e composição das misturas de biodiesel. Os métodos lineares e não lineares de 1ª ordem mais comumente utilizados são, a regressão por mínimos quadrados parciais (PLS) e máquinas de vetor suporte (SVM).

2.11.1 Mínimos quadrados parciais (PLS)

Os modelos de calibração por PLS são uns dos mais utilizados e foi proposto inicialmente por Herman Wold (WOLD, 1982) como uma técnica de análise de dados multivariados utilizada para relacionar uma ou mais variáveis resposta (Y) com diversas variáveis independentes (X). O PLS permite identificar fatores, ou seja, combinações lineares das variáveis X, que melhor modelam as variáveis dependentes Y. Ademais, o PLS permite trabalhar com conjuntos de dados onde

haja variáveis altamente correlacionadas e que apresentam ruído aleatório considerável (TEÓFILO, 2013).

De acordo com Brereton (2003), modelos PLS podem ser muito robustos, desde que as amostras futuras contenham características semelhantes aos dados originais, mas as previsões são essencialmente estatísticas. Como pode haver diferentes características espectrais, não há garantia de que o modelo terá um bom desempenho nos novos dados, por isso é sempre importante estar ciente das limitações do método, principalmente para lembrar que o uso do PLS não pode compensar experimentos mal planejados ou dados experimentais inadequados.

Ainda segundo Brereton (2003), uma característica importante do PLS é que ele leva em consideração os erros em ambas as matrizes, dos espectros e das propriedades. Em estudos que envolvem química, muitas vezes há erros significativos no preparo da amostra, como na precisão de pesagens e diluições, e assim a variável independente também pode apresentar erros. As calibrações clássica e inversa forçam o usuário a escolher qual variável contém o erro, enquanto o PLS assume que os erros estão igualmente distribuídos em ambas as matrizes.

O PLS considera a matriz \mathbf{X} ($m \times n$) para o eixo x e o vetor \mathbf{Y} ($n \times 1$) no eixo y na construção do modelo.

Em termos físicos o PLS assume que existem erros em ambas as matrizes, os quais são de igual importância, como mostrado nas Equações 11 e 12. Segundo Aleme (2008, p. 27) “No método PLS cada variável latente (VL) da matriz \mathbf{X} é modificada de modo que a covariância entre ela e o vetor da matriz \mathbf{Y} seja maximizado”.

$$\mathbf{X} = \mathbf{S} \cdot \mathbf{L}^t + \mathbf{E} \quad (11)$$

$$\mathbf{Y} = \mathbf{A} \cdot \mathbf{V}^t + \mathbf{R} \quad (12)$$

Onde, \mathbf{X} e \mathbf{Y} são as matrizes decompostas, \mathbf{S} e \mathbf{A} são as matrizes dos escores ($m \times N^{\circ}VL$), \mathbf{L}^t e \mathbf{V}^t são matrizes transpostas dos pesos ($N^{\circ}VL \times n$) e \mathbf{E} e \mathbf{R} são as matrizes dos erros das matrizes \mathbf{X} e \mathbf{Y} , respectivamente (BRERETON, 2003).

2.11.2 Máquinas de Vetores Suporte

As Máquinas de Vetores de Suporte (*Support Vector Machines*, SVM) constituem uma técnica de aprendizado de máquina desenvolvida por Vapnik e foi originalmente desenvolvida para resolver problemas de reconhecimento de padrões. Segundo Vapnik (1995) e Vapnik e Chervonenkis (1971), as SVMs têm como base a teoria de aprendizado estatístico, que estabelece uma série de princípios que devem ser obedecidos na obtenção de classificadores com boa generalização, isto é, com boa capacidade de prever corretamente a classe de novos dados do mesmo domínio em que o aprendizado foi realizado.

Segundo Bishop (2006), Vapnik (1999), Steinwart e Christmann (2008(a) e 2008(b)) a técnica SVM tem sido usada para resolver problemas de classificação e regressão, podendo ser usada também para modelar processos não lineares. Devido a inclusão de funções de perda em sua estrutura, a SVM ampliou a utilidade de sua aplicação nas áreas de regressão não linear e previsões de séries temporais (WANG *et al.*, 2008(b); LU e WANG, 2005; VAPNIK, 1995; MÜLLER *et al.*, 1997).

De um modo geral, para gerar um modelo de regressão é necessário que o conjunto de dados seja dividido em dois conjuntos, calibração e validação. As SVM utilizam o princípio de inferência chamado indução, no qual o aprendizado pode ser do tipo supervisionado ou não supervisionado. Para Haykin (1999), no aprendizado supervisionado tem-se a figura de um professor, o qual apresenta o conhecimento do ambiente por um conjunto de exemplos na forma de entrada e saída desejada. O algoritmo das SVM é construído a partir do conjunto de calibração, sendo capaz de produzir resultados para novos dados de entradas. No entanto, no aprendizado não supervisionado não existe conhecimento prévio do comportamento dos dados, o algoritmo aprende a agrupar as entradas segundo a similaridade entre os dados.

O tipo de validação cruzada utilizada na técnica SVM é a K-dobra (*K-fold*), a qual a amostra é aleatoriamente dividida em K sub-amostras, onde uma única amostra é mantida no conjunto de validação para testar o modelo e, as K-1 sub-amostras restantes são mantidas no conjunto de calibração. O processo de validação cruzada é repetido K vezes, com cada uma das K sub-amostras sendo utilizada uma única vez como informação de validação. Com os resultados das K dobras pode então ser calculada a média para produzir uma única estimativa. A

grande vantagem deste método repetido de sub-amostragem aleatória é que todas as observações são usadas tanto para o conjunto de calibração quanto para o conjunto de validação, e cada observação é utilizada para uma única validação. A validação cruzada mais utilizada é a com 10 dobras (SANCHEZ *et al.*, 2011; EFRON e TIBSHIRANI, 1997; PICARD e COOK, 1984).

As SVM podem ser lineares com margens rígidas ou não lineares com margens suaves. As SVM lineares de margens rígidas são utilizadas para resolver problemas lineares, que são de fácil resolução. As SVM não lineares com margens suaves são a extensão das SVMs lineares com margens rígidas e lidam com conjuntos de calibração mais gerais que podem apresentar ruídos e pontos anômalos, devido à própria natureza dos dados que podem ser não lineares. A diferença entre as SVM não lineares e as SVM lineares é que para a primeira, algumas restrições antes válidas são ignoradas e variáveis de folgas são incluídas no problema de otimização (LORENA e CARVALHO, 2007; SMOLA *et al.*, 1999; SMOLA e SCHOLKOPF, 2002).

Para Monard e Baranauskas (2003), mesmo que sejam utilizados conjuntos de dados diferentes para a calibração e validação, o modelo ainda pode estar sujeito a ocorrência de super-ajustamento (*overfitting*) ou sub-ajustamento (*underfitting*). Quando o modelo se especializa nos dados utilizados em seu conjunto de calibração, apresentando uma baixa taxa de acerto quando confrontado com novos dados, tem-se uma indicação de um super-ajustamento (*overfitting*). Deste modo, o super-ajustamento ocorre quando um classificador classifica corretamente todos os dados do conjunto de calibração, incluindo possíveis ruídos ou pontos que, na verdade, pertencem à outra classe, por ser muito específico para este determinado conjunto de calibração. Assim, o algoritmo criado torna-se mais propenso a errar a previsão de novos dados. É também possível que apresentem uma baixa taxa de acerto mesmo no subconjunto de treinamento, demonstrando uma condição de sub-ajustamento (*underfitting*) e pode ocorrer quando os exemplos de calibração disponíveis são pouco representativos ou quando o modelo obtido é muito simples.

Segundo Wang *et al.* (2008(a)), a aproximação por regressão direciona a um problema de estimação de uma função com base em dados de um determinado conjunto $G = \{(x_i, y_i)\}_i$ (x_i representa os vetores de entrada, y_i os valores desejados), que é produzido de ϕ . A regressão por SVM aproxima a função na forma da Equação 13.

$$f(x) = \sum_{i=1}^l w_i \Phi_i(x) + b \quad (13)$$

Onde $\{\phi_i(x)\}_{i=1}^l$ representa as variáveis de entrada, $\{w_i\}_{i=1}^l$ e b são os coeficientes. Estes podem ser estimados pela minimização da função risco apresentada na Equação 14, onde $L_\varepsilon(y, f(x))$ é descrito pela Equação 14, com o auxílio da Equação 15.

$$R(C) = C \frac{1}{l} \sum_{i=1}^l L_\varepsilon(y_i, f(x_i)) + \frac{1}{2} \|w\|^2 \quad (14)$$

$$L_\varepsilon(y_i, f(x_i)) = \begin{cases} |y - f(x)| - \varepsilon & |d - y| \geq \varepsilon \\ 0 & \text{outros} \end{cases} \quad (15)$$

Sendo ε um parâmetro previsto e $L_\varepsilon(y, f(x))$ uma função perda ε -indiferente, que não penaliza erros menores que ε . O termo $\frac{1}{2} \|w\|^2$ é utilizado como a medida do nivelamento da função e C é a constante reguladora que determina a troca entre o erro de calibração e o nivelamento do modelo. Introduzindo as variáveis de folga ζ , ζ^* tem-se a Equação 16 e 17 a minimizar, sujeitas as restrições apresentadas nas Equações 18 e 19 (WANG *et al.*, 2008(a)).

$$R(w, \zeta^*) = \frac{1}{2} \|w\|^2 + C^* \sum_{i=1}^n (\zeta_i + \zeta_i^*) \quad (16)$$

$$w \phi(x_i) + b - y_i \leq \varepsilon + \zeta_i \quad (17)$$

$$y_i - w \phi(x_i) - b \leq \varepsilon + \zeta_i \quad (18)$$

$$\zeta, \zeta^* \geq 0 \quad (19)$$

Desta forma, reescrevendo a Equação 13 de forma explícita, obtemos a Equação 20 (WANG *et al.*, 2008(a)).

$$f(x, \alpha_i, \alpha_i^*) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (20)$$

Presentes na Equação 20 estão os multiplicadores Lagrangianos, α_i e α_i^* , que satisfazem a igualdade: $\alpha_i \alpha_i^* = 0$, $\alpha_i \geq 0$, $\alpha_i^* \geq 0$. Fazendo $l = 1, \dots, l$, pode ser obtido pela maximização da forma dual da Equação 21, considerando as restrições apresentadas nas Equações 22 e 23, (VAPNIK, 1998; PLATT, 1998; YEGANEH *et al.* (2012)).

$$\begin{aligned} \phi(\alpha, \alpha^*) = & \sum_{i=1}^l y_i (\alpha_i - \alpha_i^*) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) \\ & - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) K(\alpha_i, \alpha_j) \end{aligned} \quad (21)$$

$$\sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \quad (22)$$

$$\begin{aligned} 0 & \leq \alpha_i \leq C \\ 0 & \leq \alpha_i^* \leq C \end{aligned} \quad (23)$$

Sabendo que $K(x_i, x_j) = \phi^T(x) \phi(x)$ no espaço, chamada de função Kernel. Para Schölkopf e Smola (2002) e Vapnik (1998), a utilidade das funções Kernels está na simplicidade de seu cálculo e em sua capacidade de representar espaços abstratos. Dentre as funções *Kernels*, as mais utilizadas são as lineares, quadráticas e RBF (*Radial Basis Function*), sendo que para os dados não lineares o *Kernel* RBF é o mais indicado. Em relação à regressão, dois tipos de vetores suporte são os mais utilizados, o Epsilon-SVR e o Nu-SVR. Durante o cálculo das SVM, o algoritmo faz uma busca sobre as faixas de valores existentes utilizando a validação cruzada

para selecionar as melhores faixas de valores para estes parâmetros e, então o modelo é construído.

Segundo Wise *et al.* (2006), dentre os parâmetros que se destacam na SVM, estão os valores de cost, gamma e épsilon. O parâmetro cost representa a penalidade associada a erros maiores que épsilon, sendo que o aumento do valor do custo (cost) faz com que o ajuste seja mais próximo e, portanto, mais apropriado para os dados de calibração. O parâmetro Kernel gamma controla a forma de separação do hiperplano, sendo que o aumento do gamma geralmente aumenta o número de vetores de suporte. No treinamento da função de regressão não há nenhuma penalidade associada aos pontos que estão previstos dentro do valor real da distância de epsilon, sendo que a diminuição do valor de epsilon faz com que o ajuste seja mais apropriado aos dados de calibração.

Neste estudo, os modelos de SVM apresentados, são os que fornecerem os melhores valores dos parâmetros comparativos de adequação do modelo (R^2_{CV} e RMSECV), para realizar comparação com a técnica PLS e iPLS-PLS. Os pré-processamentos utilizados para as SVM foram praticamente os mesmos utilizados na realização do PLS e iPLS-PLS, sendo utilizada 10 dobras (K-folds) para a validação cruzada.

2.11.3 Máquinas de Vetores Suporte por Mínimos Quadrados Parciais

Para Ferrão *et al.* (2007) as máquinas de vetores de suporte (*Support Vector Machine*, SVM) são máquinas de aprendizagem que podem ser usadas para regressão ou classificação. A regressão de mínimos quadrados por máquinas de vetor de suporte (*Least Square - Support Vector Machine*, LS-SVM) é uma modificação das SVMs que, ao invés de usarem uma programação quadrática, usam um sistema de equações lineares e uma função custo por mínimos quadrados, proporcionando menor exigência computacional em relação às SVM sem perda na qualidade das soluções, uma vez que ambas são baseadas nos mesmos princípios. A principal vantagem que tem sido relatada na aplicação das LS-SVM é a elevada capacidade de generalização dos modelos construídos. Porém, a desvantagem da LS-SVM reside no fato de que todos os dados de treinamento são considerados vetores de suporte, diferentemente das SVMs que detectam apenas uma pequena fração deste conjunto durante a fase de treinamento (BORIN, 2007).

2.12 Técnicas de Seleção de Variáveis

Técnicas de seleções de variáveis são utilizadas para auxiliar na escolha da melhor região espectral a ser utilizada na geração de modelo multivariado. Esta escolha é muito importante, pois permite minimizar os erros de previsão de forma a obter modelos mais robustos e de simples interpretação.

Dentre os métodos heurísticos mais empregados encontra-se o algoritmo genético e dentre os métodos determinísticos destaca-se o método de mínimos quadrados parciais por intervalo (*Interval Partial Least-Squares – iPLS*) (ZOU *et al.*, 2007 (a); ZOU *et al.*, 2007(b)).

Dentre as técnicas de seleções de variáveis, neste estudo foram utilizadas o iPLS e o algoritmo genético (*Genetic Algorithm, GA*).

2.12.1 Mínimos Quadrados Parciais por Intervalos (iPLS)

Segundo Silva *et al.* (2007) o método de regressão de mínimos quadrados parciais por intervalos indica o conjunto de variáveis espectrais baseando-se em critérios objetivos, como por exemplo, avaliar a correlação dos intervalos selecionados com o método de referência a partir do coeficiente de determinação (R^2), ou avaliar a raiz quadrada dos erros médios de previsão (RMSEP).

Segundo Nørgaard *et al.* (2000), o método iPLS é uma extensão do PLS, no qual o algoritmo desfragmenta o conjunto de dados espectrais em um determinado número de intervalos, construindo um modelo PLS para cada intervalo e expondo os resultados em um gráfico para simplificar a comparação. O método é planejado para dar uma visão completa dos dados e costuma ser bastante útil na interpretação de quais sinais do espectro são mais relevantes, contendo informações químicas importantes sobre a amostra, na construção de um bom modelo de calibração.

2.12.2 Algoritmo Genético (GA)

O algoritmo genético é um método de inteligência artificial e tem sido bastante aplicado na seleção de comprimentos ou números de ondas em aplicações

espectrométricas quantitativas. Segundo Cartwright (1995) e Lacerda & Carvalho (1999), o algoritmo genético atua modelando matematicamente a teoria de Darwin, sendo empregado por analogia ao que ocorre no mundo biológico, onde as condições experimentais (cromossomos), que levam as melhores respostas (indivíduos mais adaptados ao ambiente) tem maior chance de serem selecionados (sobreviver), sendo transmitidos às novas gerações pela reprodução. Assim, a otimização das respostas (evolução da espécie) é alcançada por meio de recombinação das variáveis (cruzamento de cromossomos) e de algumas modificações randômicas (mutações de genes).

Segundo Lucas (2002), de forma resumida, o algoritmo genético gera uma população de possíveis respostas para o problema a ser tratado para depois submetê-la ao processo de evolução, constituído pelas etapas de avaliação, seleção, cruzamento, mutação, atualização e finalização. Na etapa de avaliação, a aptidão das soluções é avaliada ao realizar uma análise de quão bem elas respondem ao problema proposto. Para a etapa de seleção, os indivíduos são selecionados para a reprodução, onde a probabilidade de uma dada solução ser selecionada é proporcional à sua aptidão. Cruzamento é a etapa em que as características das soluções escolhidas são recombinadas, gerando novos indivíduos. A etapa de mutação compreende o acréscimo da variedade à população, devido às alterações das características dos indivíduos resultantes do processo de reprodução. A penúltima etapa é a atualização, onde os indivíduos criados nesta geração são inseridos na população. A última etapa de finalização compreende a verificação se as condições de encerramento da evolução foram alcançadas, retornando para a etapa de avaliação em caso negativo e encerrando a execução em caso positivo.

Embora haja uma variedade de métodos de seleção de variáveis, no entanto o GA é um algoritmo que fornece um método direto baseado na sobrevivência das variáveis mais aptas como uma aproximação para modelar os dados. A técnica de seleção natural para selecionar variáveis em calibração multivariada foi descrita por Wise *et al.* (2003), como sendo uma dada matriz \mathbf{X} de variáveis independentes e um vetor \mathbf{Y} de variável dependente, o algoritmo escolhe um subconjunto de variáveis em \mathbf{X} e, usando algum método de regressão multivariada e alguma estratégia de validação cruzada, é determinado o valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV) para o subconjunto de variáveis. O algoritmo

genético usa este conceito de forma iterativa até encontrar os subconjuntos que proporcionam o menor valor de RMSECV.

2.13 Resumo do estudo quimiométrico

O estudo quimiométrico proposto a esta dissertação inclui as etapas descritas no fluxograma apresentado na Figura 14. No fluxograma da Figura 14 foi incluído desde a etapa de construção das matrizes de espectros brutos de NIR e Mid-IR, o pré-tratamento desses espectros, análise exploratória utilizando PCA para detecção de amostras atípicas até o segundo modelo conforme ao estabelecido pela norma ASTM E1655-05 (indicada com um (*) na Figura 14), utilização de algoritmo de seleção de amostras (algoritmo KS) e algoritmos de seleção de variáveis (iPLS e GA), seguido pela construção de modelos preditivos para as propriedades estudadas e finalizando com a comparação dos resultados dos modelos preditivos utilizando o teste F.

Figura 14 – Fluxograma do estudo quimiométrico utilizado nesta dissertação

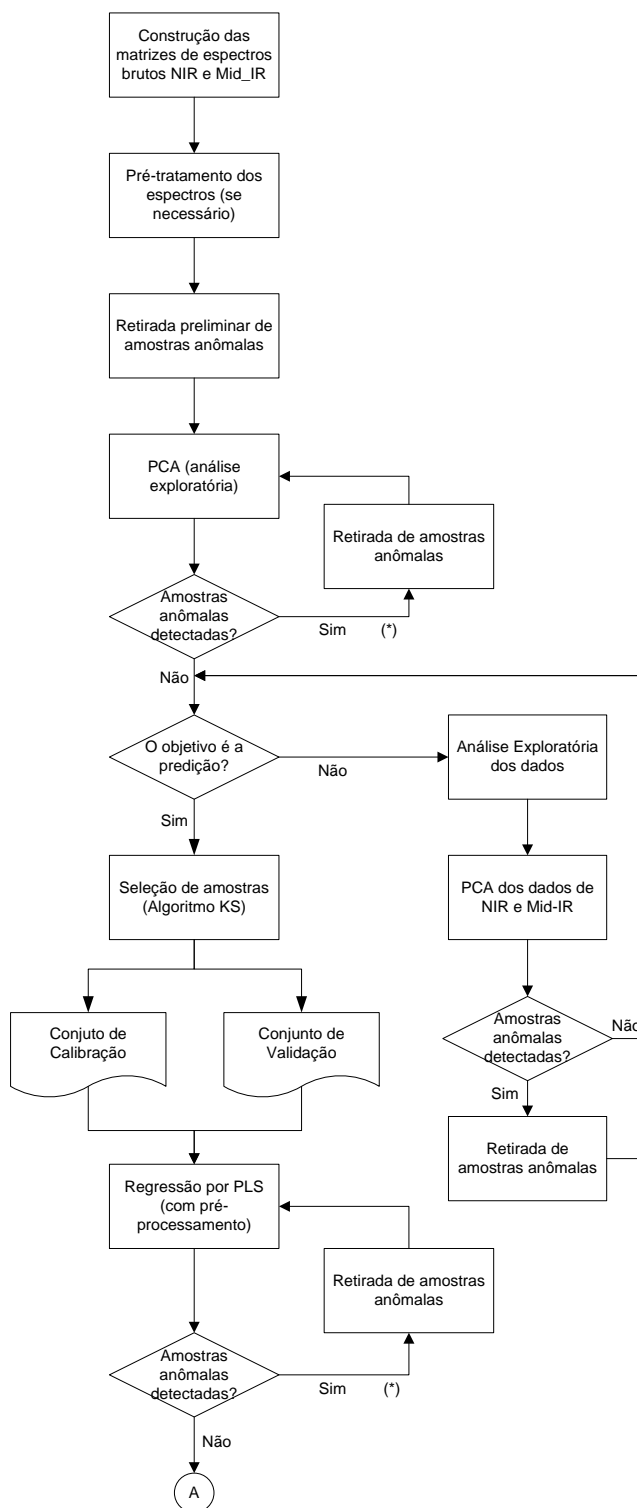
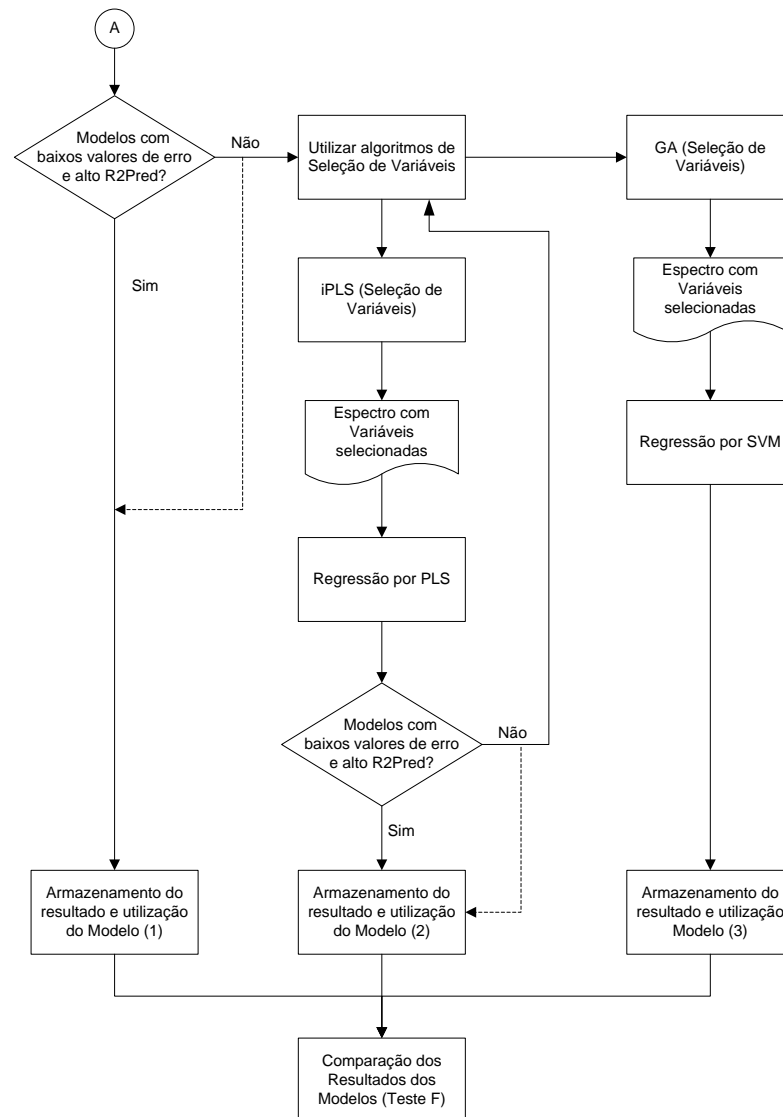


Figura 14 – Fluxograma do estudo quimiométrico utilizado nesta dissertação
(continuação)



A retirada sucessiva de amostras atípicas, deve atender as determinações da norma ASTM E1655-05, onde a construção de um novo modelo a partir da retirada de um conjunto de amostras atípicas de um modelo anterior, este novo modelo pode apresentar outras amostras atípicas não detectadas no modelo anterior e isto pode ocorrer incessantemente a cada novo modelo construído, sendo este comportamento classificado como efeito bola de neve (do inglês, *snowballing effect*). Neste caso, a norma aconselha detectar e remover a amostra atípica somente até o segundo modelo (ASTM, 2005).

2.14 Ferramentas e conceitos estatísticos

2.14.1 Parâmetros estatísticos

Neste estudo são utilizados alguns parâmetros estatísticos, que são muito importantes na avaliação e comparação dos modelos, tais como, os valores de RMSEC, RMSECV, RMSEP, R_{cal} , R_{cv} e R_{pred} .

Segundo Teófilo (2013), as amostras previstas para os modelos durante a validação cruzada são comparadas com seus valores conhecidos e que parâmetros estatísticos são calculados para avaliar em que situação o modelo apresentou maior capacidade de previsão.

Um dos parâmetros utilizados nesta dissertação foi a raiz do erro quadrático médio da validação da calibração (RMSEC) apresentada pela Equação 24 e o coeficiente de correlação da calibração (R_{cal}) apresentada pela Equação 25. Ambos obtidos a partir dos valores estimados pela validação cruzada e os valores conhecidos.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^{I_c} (y_i - \hat{y})^2}{I_c}} \quad (24)$$

$$R_{cal} = \frac{\sum_{i=1}^{I_c} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_c} (\hat{y}_i - \bar{\hat{y}})^2 (y_i - \bar{y})^2}} \quad (25)$$

Em que \hat{y}_i é o valor estimado para a amostra i , não incluída na construção do modelo, $\bar{\hat{y}}$ é a média dos valores em \hat{y} e \bar{y} é a média dos valores em y e I_c é o número de amostras da calibração.

Outros parâmetros usados nesta dissertação foi a raiz do erro quadrático médio da validação cruzada (RMSECV) apresentada pela Equação 26 e o coeficiente de correlação de validação cruzada (R_{cv}) apresentada pela Equação 27. Ambos os parâmetros foram obtidos a partir dos valores estimados pela validação cruzada (blocos contíguos) e os valores conhecidos. De acordo com Wise *et al.* (2006) e Ballabio e Consonni (2013):

$$RMSECV = \sqrt{\frac{\sum_{i=1}^s (y_i - \hat{y})^2}{s}} \quad (26)$$

$$R_{cv} = \frac{\sum_{i=1}^s (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^s (\hat{y}_i - \bar{\hat{y}})^2 (y_i - \bar{y})^2}} \quad (27)$$

Em que \hat{y}_i é o valor estimado para a amostra i , não incluída na construção do modelo, s é o número de segmentos ou divisões no conjunto de dados e \bar{y} é a média dos valores em y .

Na etapa de validação externa, um conjunto de variáveis independentes é fornecido ao modelo com h componentes para verificar o seu poder de previsão (TEÓFILO, 2013). A habilidade de previsão do modelo foi calculada nesta dissertação usando erro quadrático médio de previsão (RMSEP) conforme é apresentado na Equação 28 e o coeficiente de correlação de previsão (R_{pred}) apresentada na Equação 29.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{I_p} (y_i - \hat{y}_i)^2}{I_p}} \quad (28)$$

$$R_{pred} = \frac{\sum_{i=1}^{I_p} (\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{I_p} (\hat{y}_i - \bar{\hat{y}})^2 (y_i - \bar{y})^2}} \quad (29)$$

Em que \hat{y}_i é o valor estimado pelo modelo final para a amostra i , I_p é o número de amostra no conjunto de previsão e \bar{y} é a média dos valores em y . Uma vez o modelo ter sido construído e validado, ele pode ser usado para previsões de novas amostras.

2.14.2 Hotelling T^2 , resíduos padronizados e estudentizados

Por definição, o resíduo padronizado (d_i) corresponde ao resíduo bruto dividido pelo erro padrão estimado dos resíduos (Equação 30). Se os erros têm distribuição normal, então aproximadamente 95% dos resíduos padronizados devem estar contidos no intervalo de $-2 \leq d_i \leq 2$. Resíduos fora desse intervalo podem indicar a presença de amostras atípicas. Desta forma, se os pontos estiverem distribuídos dentro do intervalo $[-2,2]$, considera-se que o modelo está bem ajustado e é esperado que menos de 5% dos resíduos se posicionem fora desse intervalo.

$$d_i = \frac{e_i}{\sqrt{QM_E}} \quad (30)$$

Onde o e_i corresponde ao resíduo bruto e QM_E é o erro padrão estimado dos resíduos.

O resíduo estudentizado é uma medida muito mais adequada de atipicidade, onde o termo estudentizado está relacionado com a distribuição t que é mais adequada para a medida de variável aleatória (Barthus *et al.*, 2005) e o intervalo entre -3 e +3 deve ser utilizado para detectar a presença de observações atípicas.

Segundo Miranda (2008) e Kourti (2002) a estatística de *Hotelling* (T^2) é utilizada para monitorar os componentes do modelo PCA ou PLS e pode ser usada para analisar a existência de amostras atípicas (valores aberrantes ou atípicos). A Equação 31 é usada para obter a estatística para uma amostra i correspondente a um componente principal ou variável latente t_i para o PCA ou PLS, respectivamente, para as primeiras n variáveis latentes.

$$T_n^2 = \sum_{i=1}^n \frac{t_i^2}{S_{t_i}^2} \quad (31)$$

Onde S_{t_i} é a variância estimada da variável latente correspondente.

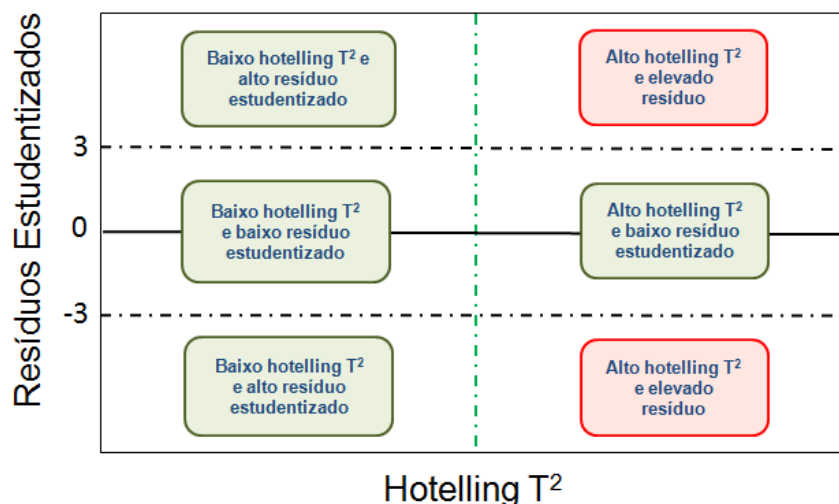
Assim, um valor de *Hotelling* T^2 baixo para as amostras indica que estão situadas próxima do centro dos dados e conseqüentemente apresentam pouca influência na construção do modelo e devem ser mantidas. Amostras com resíduo alto, mas com

espectro similar à média do conjunto de calibração (*Hotelling T²* baixo) podem indicar um erro no valor propriedade medida ou simplesmente um erro de digitação. Para amostras com valores alto de *Hotelling T²* (muito diferente da média dos espectros), porém com resíduo estudentizado baixo significa que elas têm uma grande importância no modelo, mas que não são atípicas. *Hotelling T²* e resíduos elevados indicam que as amostras não foram modeladas de forma adequada pelo PLS e, portanto são diferentes da média dos espectros e que provavelmente, são amostras atípicas e devem ser retiradas do conjunto de dados (Barthus *et al.*, 2005).

Para decidir se amostras com alto valor de *Hotelling T²* devem ser retiradas ou mantidas no modelo, é recomendável fazer uma análise cuidadosa de cada amostra, avaliando o *Hotelling T²* em análise conjunta com os resíduos estudentizados.

A interpretação do gráfico de resíduos estudentizados *versus Hotelling T²* pode ser realizada mais facilmente ao levar em consideração a localização das amostras no gráfico, conforme apresentado na Figura 15. Como pode ser observado na Figura 15, as informações sobre a interpretação do gráfico foram sintetizadas e a localização das amostras atípicas foram identificadas em vermelho, mais precisamente nos cantos inferior e superior direito do gráfico.

Figura 15 – Exemplo representativo do gráfico de *Hotelling T²* *versus* resíduos estudentizados com localização das amostras



3 RESULTADOS E DISCUSSÕES

3.1 Caracterização do Biodiesel

Para a caracterização do biodiesel foram realizadas medições das propriedades: ponto de entupimento de filtro a frio, massa específica, índice de refração, assim como foi determinada a composição química por cromatografia gasosa. Além das propriedades citadas, foram realizadas medidas de ponto de fulgor e destilação D86, sendo esta última usada como complemento do presente estudo.

3.1.1 Ponto de Entupimento de Filtro a Frio

A Tabela 9 mostra os resultados das medidas de ponto de entupimento (PEFF) para os cinco tipos de biodiesel puro e compara com os resultados relatados na literatura. A diferença apresentada nas medidas de PEFF podem ser explicadas pelas diferenças na síntese e nas etapas de purificação, além da qualidade/especificações da matéria-prima.

Tabela 9 – Resultados das medidas de PEFF dos cinco tipos de biodiesel puro comparado com os resultados da literatura (obtidos por outros autores)

Experimental		Literatura	
Descrição	Média PEFF (°C)	PEFF (°C)	Referência
Biodiesel de Soja	-5,0	-4,0	Sanford et al., 2009
Biodiesel de Canola	-6,6	-8,8	Giakoumis, 2013
Biodiesel de Milho	-2,6	-3,0	Sanford et al., 2009
Biodiesel de Girassol	-1,6	-3,0	Sanford et al., 2009
Biodiesel Região Sul	6,0	-	-

Para as misturas binárias os resultados são apresentados graficamente, o que torna possível avaliar a variação nos resultados de PEFF em função da alteração na composição da mistura entre dois tipos diferentes de biodiesel. Nas

Figuras 16 a 19 são mostrados os resultados de PEFF (°C) das misturas binárias de biodiesel de milho com biodiesel de soja, biodiesel de canola com biodiesel de soja, biodiesel de girassol com biodiesel de soja e biodiesel de canola com biodiesel fornecido por distribuidora da região sul, respectivamente. Nos gráficos, a barra de erro (I) utilizada corresponde à margem de erro das medições das amostras, sendo calculada a partir da distribuição *t* de Student com 95% de confiança. Nos gráficos também é possível observar os valores individuais de PEFF (°C) de cada amostra (▪) e a média das amostras (•). As Figuras 16, 18 e 19 apresentam o comportamento esperado, ou seja, conforme adição do segundo tipo de biodiesel no biodiesel base (soja ou distribuidora da região sul), o ponto de entupimento sofre alterações, aproximando-se do valor de PEFF do biodiesel que está sendo adicionado. Na Figura 17, para a composição de 90% (v/v) em biodiesel de canola e 10% (v/v) de biodiesel de soja, a variação da temperatura de entupimento obtida pelas análises não foram as esperadas, visto que, que a temperatura entre os puros são muito próximas. O desvio observado pode ser atribuído a precisão do equipamento que é programado segundo a indicação da norma ASTM D6371 para indicar a temperatura com uma incerteza de 1,0°C, sendo que a limitação da medida não é o sensor de temperatura, mas a determinação do ponto no qual o entupimento do filtro foi realmente efetivo, verificado a partir da diferença de pressão. Desta forma, a repetibilidade do método é de 1,76°C, para um nível de confiança de 95%, e a reprodutibilidade de $0,102 \cdot (25 - X)^\circ\text{C}$, que na faixa de 5°C retorna um valor de 2,04°C.

Figura 16 – Resultados das medidas de PEFf para mistura binária de biodiesel de milho com biodiesel de soja

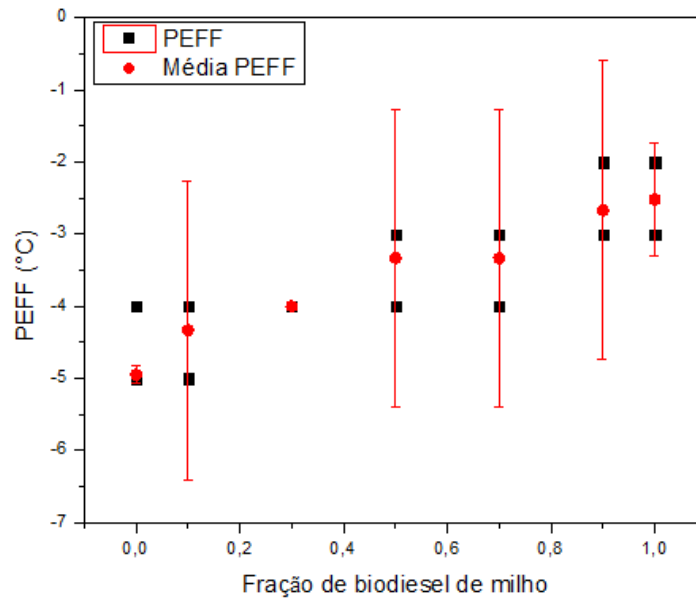


Figura 17 – Resultados das medidas de PEFf para mistura binária de biodiesel de canola com biodiesel de soja

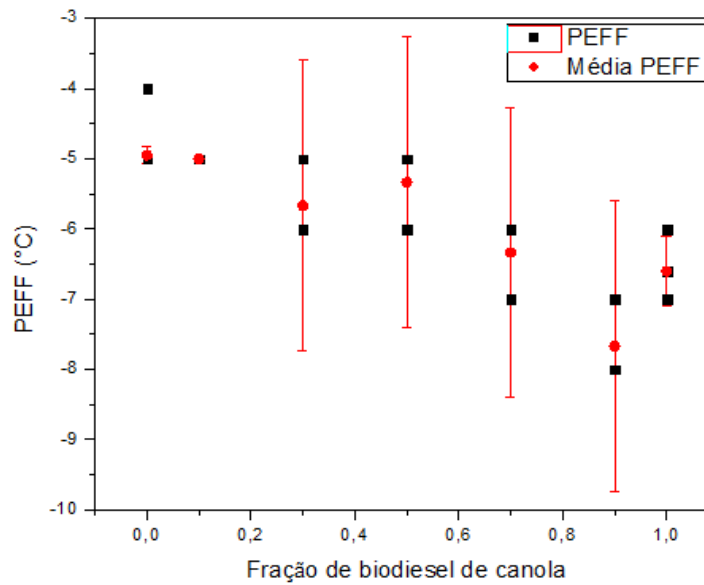


Figura 18 – Resultados das medidas de PEFF para mistura binária de biodiesel de girassol com biodiesel de soja

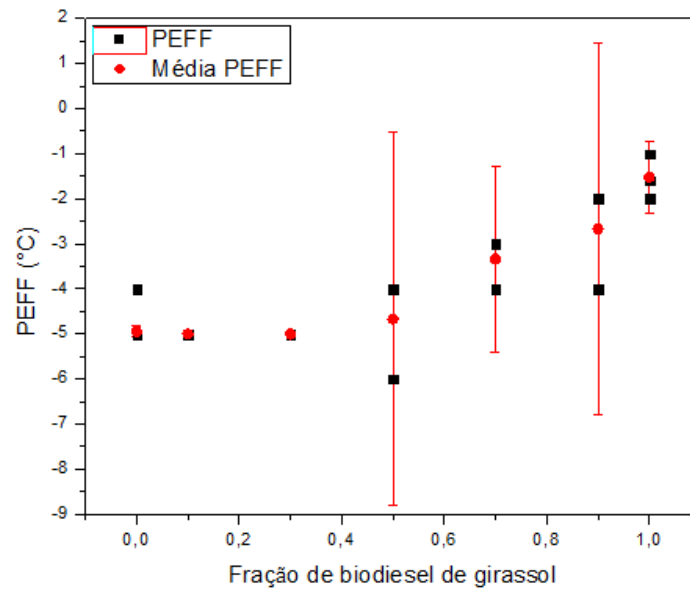
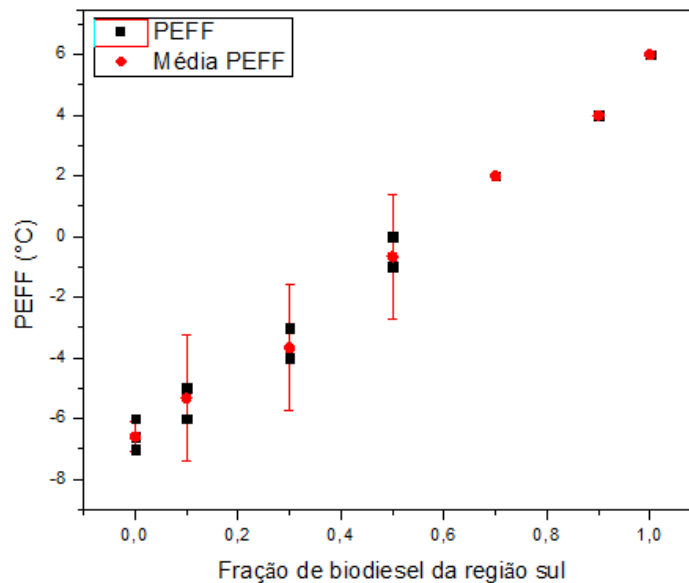


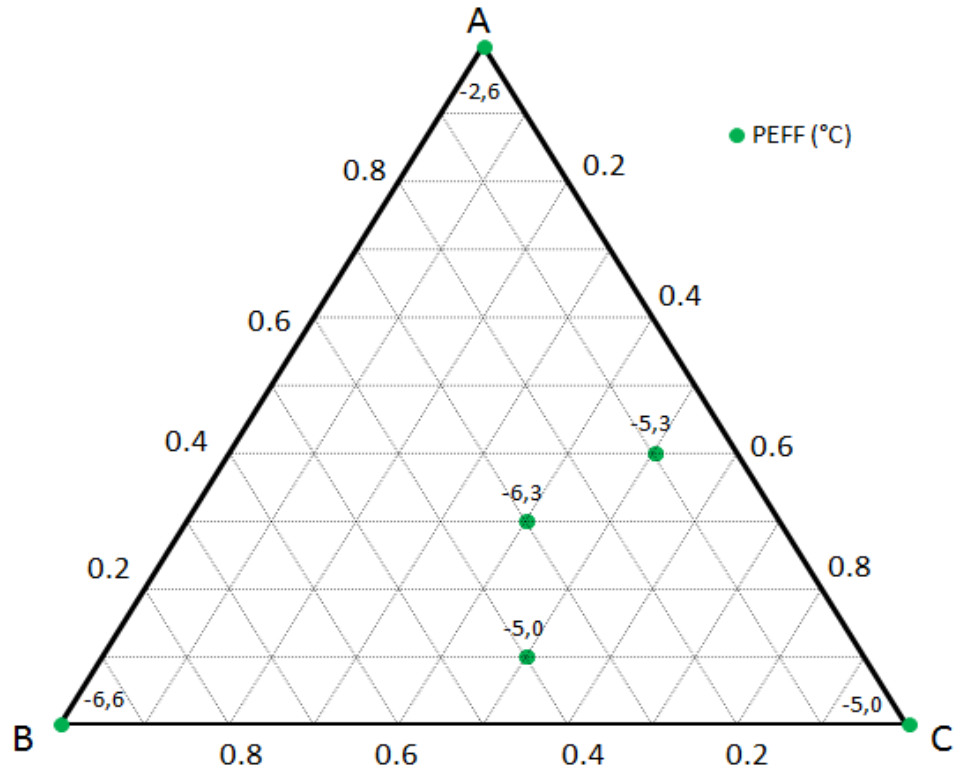
Figura 19 – Resultados das medidas de PEFF para mistura binária de biodiesel de soja com biodiesel da distribuidora da região sul



Os resultados de PEFF para as misturas ternárias são apresentados na Tabela 10 e em diagramas ternários (Figuras 20 a 22), o que torna possível avaliar a variação dos resultados de PEFF em função da alteração na composição da mistura entre três tipos diferentes de biodiesel. Nas Figuras 20 a 22 são mostrados os resultados médios de PEFF (°C) das misturas ternárias de biodiesel de milho-

canola-soja, biodiesel de milho-girassol-soja e biodiesel de girassol-canola-soja, respectivamente, nas proporções indicadas nos diagramas.

Figura 20 – Resultados das medidas de PEFF para misturas ternárias de biodiesel de milho (A) com biodiesel de canola (B) e biodiesel de soja (C)



A Tabela 10 apresenta os resultados das medidas de PEFF para as misturas ternárias, onde a coluna identificada como amostra representa a identificação das amostras, a coluna denominada composição representa a composição volumétrica das misturas entre os diferentes tipos de biodiesel utilizados e, também são apresentados os intervalos estimados das medições para as amostras de cada composição.

Figura 21 – Resultados das medidas de PEFF para mistura ternária de biodiesel de milho (A) com biodiesel de girassol (B) e biodiesel de soja (C)

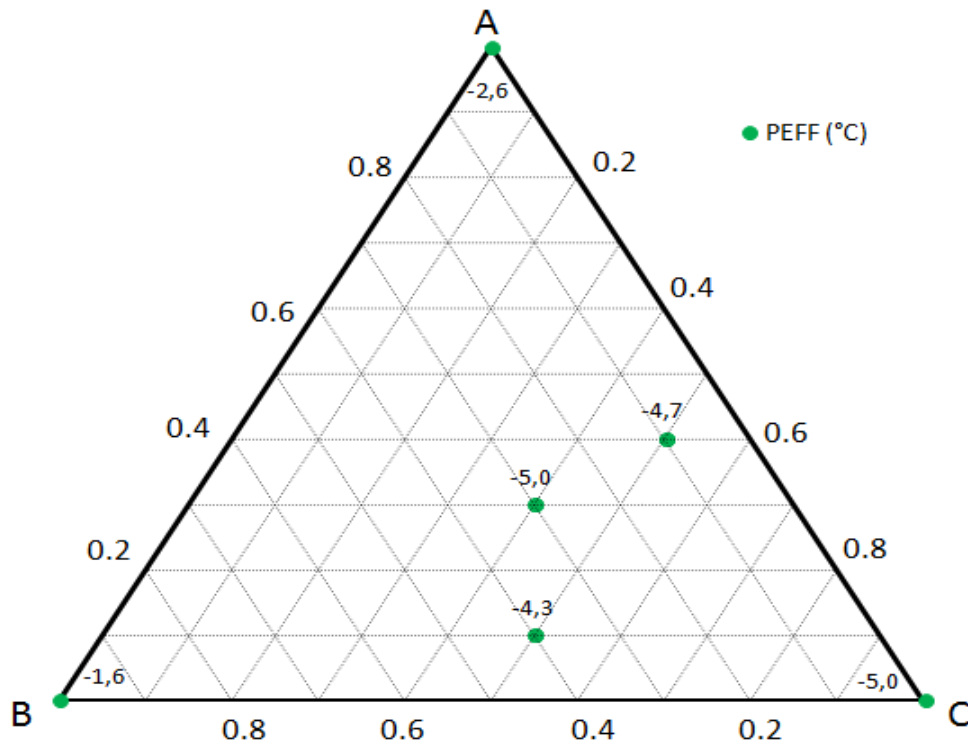


Figura 22 – Resultados das medidas de PEFF para mistura ternária de biodiesel de girassol (A) com biodiesel de canola (B) e biodiesel de soja (C)

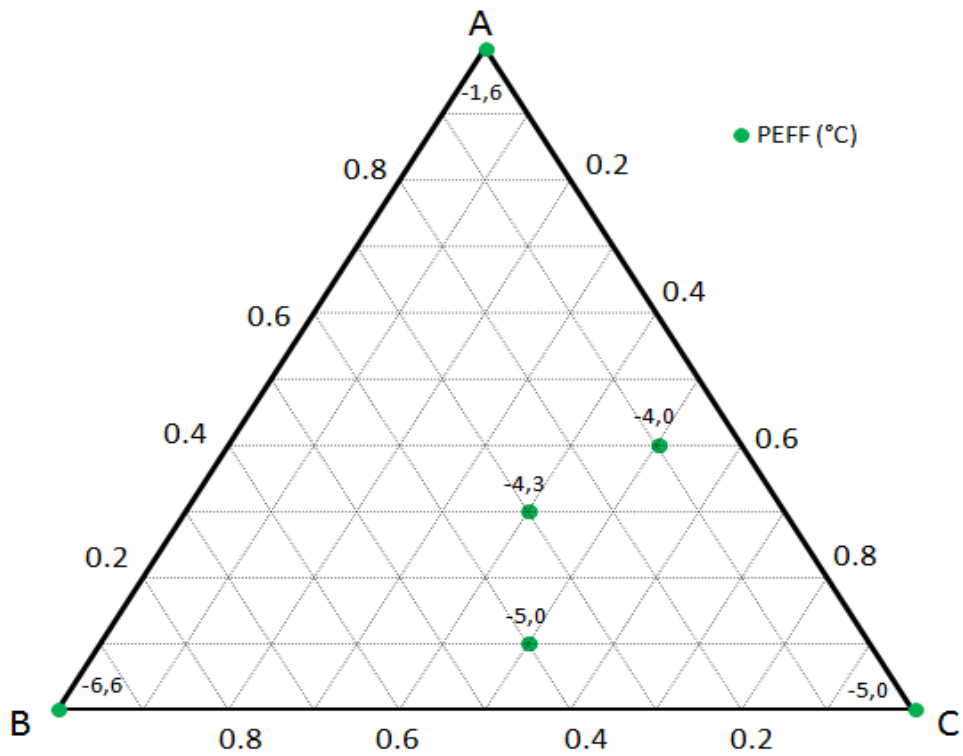


Tabela 10 – Resultados das medidas de PEFf para misturas ternárias entre os quatro tipos de biodiesel

Amostra	Composição das misturas (% volumétrica)				PEFF (°C)	Intervalo de confiança de 95% (°C)
	Biodiesel de canola	Biodiesel de girassol	Biodiesel de milho	Biodiesel de soja		
Am112	10	0	40	50	-6,0	
Am113	10	0	40	50	-5,0	-5,3 ± 2,0
Am114	10	0	40	50	-5,0	
Am115	30	0	30	40	-7,0	
Am116	30	0	30	40	-6,0	-6,3 ± 2,0
Am117	30	0	30	40	-6,0	
Am118	40	0	10	50	-5,0	
Am119	40	0	10	50	-4,0	-5,0 ± 3,6
Am120	40	0	10	50	-6,0	
Am121	0	10	40	50	-5,0	
Am122	0	10	40	50	-5,0	-4,7 ± 2,0
Am123	0	10	40	50	-4,0	
Am124	0	30	30	40	-5,0	
Am125	0	30	30	40	-5,0	-5,0 ± 0,0
Am126	0	30	30	40	-5,0	
Am127	0	40	10	50	-5,0	
Am128	0	40	10	50	-4,0	-4,3 ± 2,0
Am129	0	40	10	50	-4,0	
Am130	40	10	0	50	-5,0	
Am131	40	10	0	50	-5,0	-5,0 ± 0,0
Am132	40	10	0	50	-5,0	
Am133	30	30	0	40	-4,0	
Am134	30	30	0	40	-5,0	-4,3 ± 2,0
Am135	30	30	0	40	-4,0	
Am136	10	40	0	50	-4,0	
Am137	10	40	0	50	-4,0	-4,0 ± 0,0
Am138	10	40	0	50	-4,0	

Para as misturas quaternárias os resultados são apresentados em forma de tabela, permitindo observar a variação dos resultados de PEFf em função da

alteração na composição da mistura entre quatro diferentes tipos de biodiesel. Na Tabela 11 são exibidos os resultados médios de PEFF (°C) das misturas quaternárias de biodiesel de canola, girassol, milho e soja, onde a coluna identificada como amostra representa a identificação das amostras, a coluna denominada composição representa a percentagem volumétrica das misturas entre os quatro diferentes tipos de biodiesel utilizados e, também são apresentados os intervalos estimados das medições para as amostras de cada composição.

Tabela 11 – Resultados de PEFF para misturas quaternárias entre os quatro tipos de biodiesel

Amostra	Composição das misturas (% volumétrica)				PEFF (°C)	Intervalo de confiança de 95% (°C)
	Biodiesel de canola	Biodiesel de girassol	Biodiesel de milho	Biodiesel de soja		
Am139	25	25	25	25	-6,0	
Am140	25	25	25	25	-6,0	-6,0 ± 0,0
Am141	25	25	25	25	-6,0	
Am142	10	10	40	40	-5,0	
Am143	10	10	40	40	-6,0	-5,3 ± 2,0
Am144	10	10	40	40	-5,0	
Am145	40	10	10	40	-6,0	
Am146	40	10	10	40	-6,0	-6,0 ± 0,0
Am147	40	10	10	40	-6,0	
Am148	10	40	10	40	-4,0	
Am149	10	40	10	40	-5,0	-4,7 ± 2,0
Am150	10	40	10	40	-5,0	
Am151	40	10	40	10	-7,0	
Am152	40	10	40	10	-7,0	-7,0 ± 0,0
Am153	40	10	40	10	-7,0	
Am154	10	40	40	10	-6,0	
Am155	10	40	40	10	-6,0	-6,0 ± 0,0
Am156	10	40	40	10	-6,0	
Am157	40	40	10	10	-7,0	
Am158	40	40	10	10	-5,0	-6,0 ± 3,6
Am159	40	40	10	10	-6,0	

3.1.2 Massa Específica

Os resultados das medidas de massa específica a 20°C, para os cinco tipos de biodiesel puro, foram comparados com os resultados relatados na literatura e são mostrados na Tabela 12.

Tabela 12 – Resultados das medidas de massa específica (ρ) dos cinco tipos de biodiesel puro comparado com os resultados da literatura

Experimental		Literatura	
Descrição	Média $\rho(\text{g/cm}^3)$	$\rho(\text{g/cm}^3)$	Referência
Biodiesel de Soja	0,8815	0,8820	Hoekman <i>et al.</i> , 2012
Biodiesel de Canola	0,8791	0,8830	Hoekman <i>et al.</i> , 2012
Biodiesel de Milho	0,8798	0,8830	Hoekman <i>et al.</i> , 2012
Biodiesel de Girassol	0,8814	0,8780	Hoekman <i>et al.</i> , 2012
Biodiesel Região Sul	0,8777	-	-

Os resultados para as misturas binárias são apresentados graficamente, facilitando a avaliação de variação dos resultados de massa específica em função da alteração na composição da mistura entre dois diferentes tipos de biodiesel. Nas Figuras 23 a 26 são mostrados os resultados de massa específica (g/cm^3) das misturas binárias de biodiesel de milho com biodiesel de soja, biodiesel de canola com biodiesel de soja, biodiesel de girassol com biodiesel de soja e biodiesel de canola com biodiesel fornecido por distribuidora da região sul, respectivamente. Nos gráficos, apresentados nas Figura 23 a 26, a barra de erro (I) utilizada corresponde à margem de erro das medições das amostras, sendo calculada a partir da distribuição *t* de Student com 95% de confiança. Nos gráficos também é possível observar os valores individuais de massa específica (g/cm^3) de cada amostra (■) e a média das amostras (•).

As Figuras 23, 24 e 26 demonstram o comportamento esperado, ou seja, conforme adição do segundo tipo de biodiesel no biodiesel base (soja ou distribuidora da região sul) a massa específica do biocombustível sofre alterações na composição, aproximando-se do valor de massa específica do biodiesel adicionado. No entanto, na Figura 25, como a massa específica do biodiesel de girassol e

biodiesel de soja são muito próximas, 0,8814 e 0,8815 g/cm³ respectivamente, não era esperado uma variação significativa entre as massas específicas das misturas, visto que a repetibilidade do método de medição da massa específica é de $1,6 \cdot 10^{-4}$ g/cm³, para um nível de confiança de 95%, e a reprodutibilidade de $5,2 \cdot 10^{-4}$ g/cm³. A estimativa intervalar foi utilizada para descrever o intervalo de confiança que representa a amplitude de valores com grau de confiança de 95% de conter o verdadeiro valor da população.

Figura 23 – Resultados das medidas de massa específica para mistura binária de biodiesel de milho com biodiesel de soja

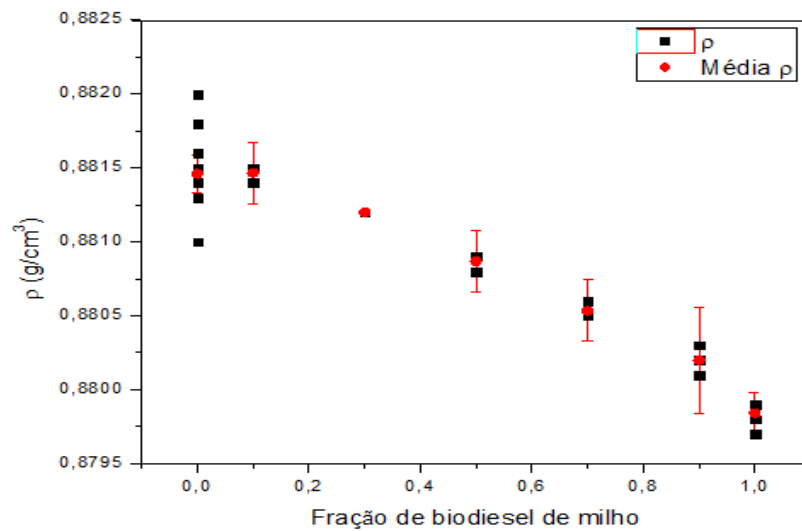


Figura 24 – Resultados das medidas de massa específica para mistura binária de biodiesel de canola com biodiesel de soja

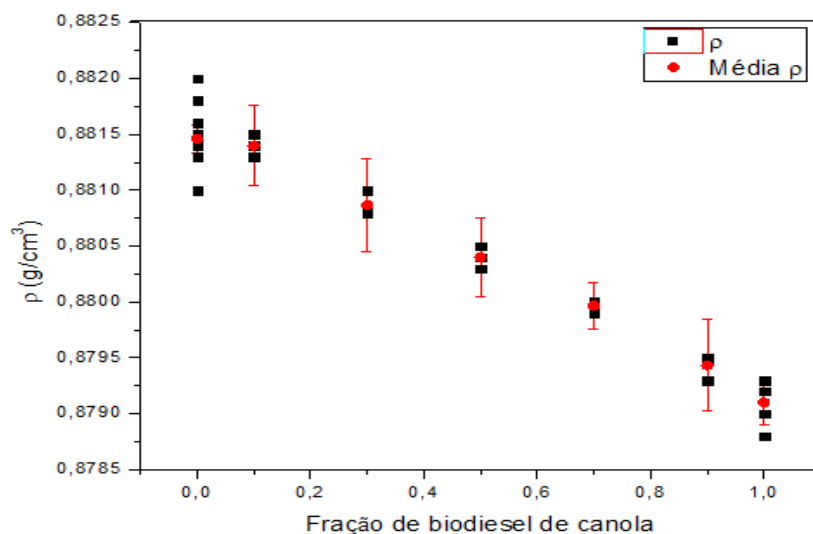


Figura 25 – Resultados das medidas de massa específica para mistura binária de biodiesel de girassol com biodiesel de soja

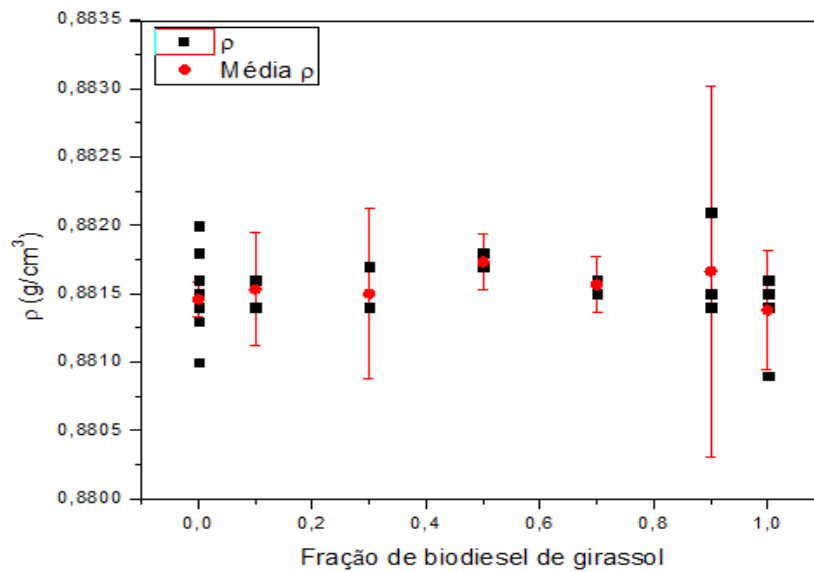
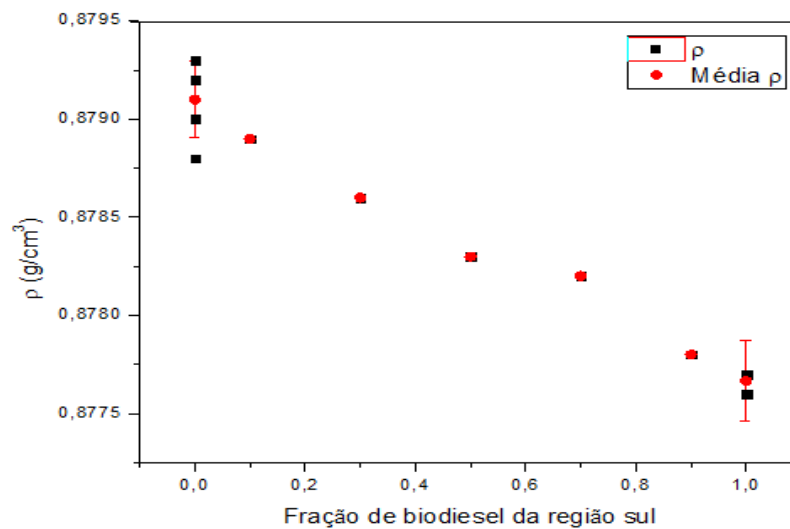


Figura 26 – Resultados das medidas de massa específica para mistura binária de biodiesel de canola com biodiesel fornecido por distribuidora da região sul



Os resultados de massa específica para as misturas ternárias são apresentados na Tabela 13 e em diagramas ternários (Figuras 27 a 29), que permitem a avaliação da variação dos resultados de massa específica em função da alteração na composição da mistura entre três diferentes tipos de biodiesel. Nas Figuras 27 a 29 são apresentados os resultados médios de massa específica (g/cm³) das misturas ternárias de biodiesel de milho-canola-soja, biodiesel de milho-girassol-soja e biodiesel de girassol-canola-soja, respectivamente, nas proporções indicadas

nos diagramas. Como as medidas de massa específica dos puros (biodiesel canola, girassol, milho e soja) são muito similares, ocorre que o comportamento observado se aproxima ao esperado, ou seja, os dados apresentam pouca variação.

Figura 27 – Resultados das medidas de massa específica para misturas ternárias de biodiesel de milho (A) com biodiesel de canola (B) e biodiesel de soja (C)

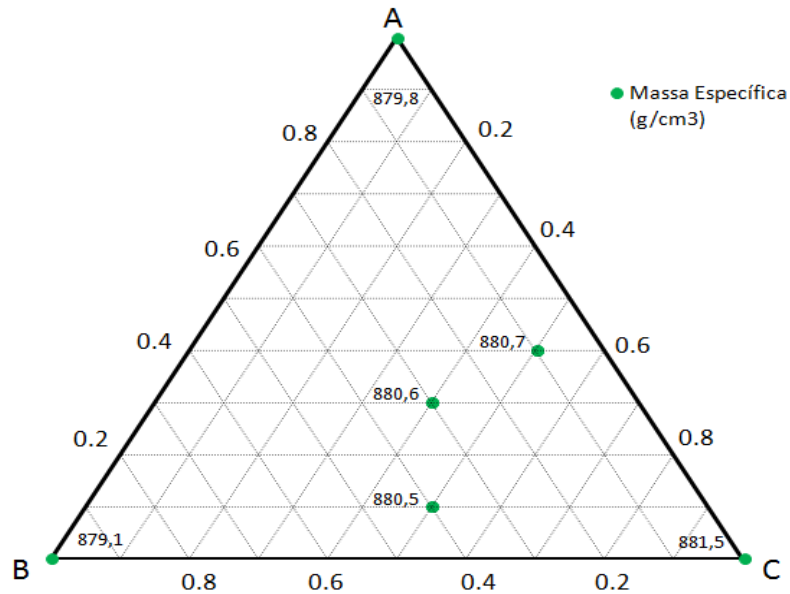


Figura 28 – Resultados das medidas de massa específica para mistura ternária de biodiesel de milho (A) com biodiesel de girassol (B) e biodiesel de soja (C)

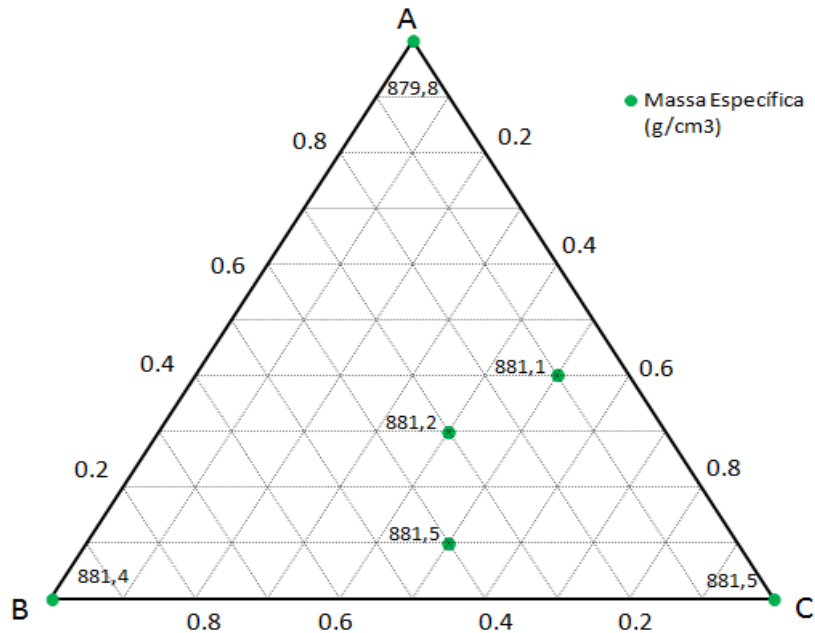
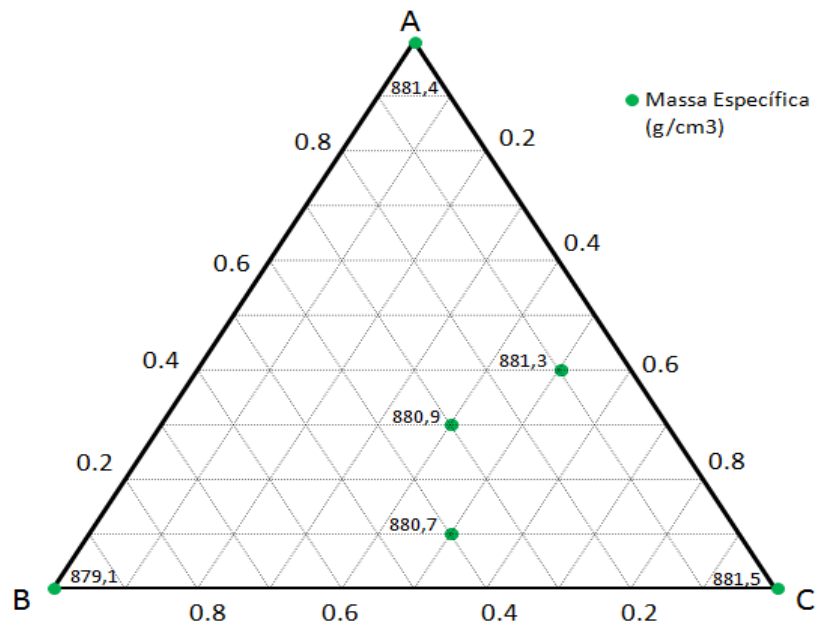


Figura 29 – Resultados das medidas de massa específica para mistura ternária de biodiesel de girassol (A) com biodiesel de canola (B) e biodiesel de soja (C)



A Tabela 13 apresenta os resultados das medidas de massa específica para as misturas ternárias, onde a coluna identificada como amostra representa a identificação das amostras, a coluna denominada composição representa a percentagem volumétrica das misturas entre os diferentes tipos de biodiesel utilizados e, também são apresentados os intervalos estimados com 95% de confiança das medições para as amostras de cada composição.

Para as misturas quaternárias os resultados são apresentados no formato de tabela, onde se observa a variação dos resultados de massa específica em função da alteração na composição da mistura entre quatro diferentes tipos de biodiesel. Na Tabela 14 são exibidos os resultados médios das medidas de massa específica (g/cm³) das misturas quaternárias de biodiesel de canola, girassol, milho e soja, onde a coluna identificada como amostra representa a identificação das amostras, a coluna denominada composição representa a fração volumétrica das misturas entre os diferentes tipos de biodiesel utilizados e, também são apresentados os intervalos estimados com 95% de confiança das medições para as amostras de cada composição.

Tabela 13 – Resultados das medidas de massa específica (ρ) para misturas ternárias entre os quatro tipos de biodiesel

Amostra	Composição das misturas (% volumétrica)				Massa Específica (g/cm ³)	Intervalo de confiança de 95% (g/cm ³)
	Biodiesel de canola	Biodiesel de girassol	Biodiesel de milho	Biodiesel de soja		
Am112	10	0	40	50	0,8806	
Am113	10	0	40	50	0,8808	0,8807 ± 0,0004
Am114	10	0	40	50	0,8806	
Am115	30	0	30	40	0,8806	
Am116	30	0	30	40	0,8806	0,8806 ± 0,0002
Am117	30	0	30	40	0,8805	
Am118	40	0	10	50	0,8805	
Am119	40	0	10	50	0,8805	0,8805 ± 0,0000
Am120	40	0	10	50	0,8805	
Am121	0	10	40	50	0,8811	
Am122	0	10	40	50	0,8811	0,8811 ± 0,0000
Am123	0	10	40	50	0,8811	
Am124	0	30	30	40	0,8813	
Am125	0	30	30	40	0,8811	0,8812 ± 0,0004
Am126	0	30	30	40	0,8811	
Am127	0	40	10	50	0,8814	
Am128	0	40	10	50	0,8814	0,8815 ± 0,0004
Am129	0	40	10	50	0,8816	
Am130	40	10	0	50	0,8808	
Am131	40	10	0	50	0,8805	0,8807 ± 0,0006
Am132	40	10	0	50	0,8807	
Am133	30	30	0	40	0,8809	
Am134	30	30	0	40	0,8811	0,8809 ± 0,0006
Am135	30	30	0	40	0,8808	
Am136	10	40	0	50	0,8813	
Am137	10	40	0	50	0,8812	0,8813 ± 0,0002
Am138	10	40	0	50	0,8813	

Tabela 14 – Resultados das medidas de massa específica (ρ) para misturas quaternárias entre os quatro tipos de biodiesel

Amostra	Composição das misturas (% volumétrica)				Massa Específica (g/cm ³)	Intervalo de confiança de 95% (g/cm ³)
	Biodiesel de canola	Biodiesel de girassol	Biodiesel de milho	Biodiesel de soja		
Am139	25	25	25	25	0,8806	
Am140	25	25	25	25	0,8807	0,8806 ± 0,0002
Am141	25	25	25	25	0,8806	
Am142	10	10	40	40	0,8809	
Am143	10	10	40	40	0,8808	0,8809 ± 0,0002
Am144	10	10	40	40	0,8809	
Am145	40	10	10	40	0,8805	
Am146	40	10	10	40	0,8805	0,8805 ± 0,0000
Am147	40	10	10	40	0,8805	
Am148	10	40	10	40	0,8812	
Am149	10	40	10	40	0,8812	0,8812 ± 0,0000
Am150	10	40	10	40	0,8812	
Am151	40	10	40	10	0,8801	
Am152	40	10	40	10	0,8801	0,8801 ± 0,0000
Am153	40	10	40	10	0,8801	
Am154	10	40	40	10	0,8809	
Am155	10	40	40	10	0,8809	0,8809 ± 0,0002
Am156	10	40	40	10	0,8808	
Am157	40	40	10	10	0,8805	
Am158	40	40	10	10	0,8806	0,8806 ± 0,0006
Am159	40	40	10	10	0,8808	

3.1.3 Índice de Refração

Os resultados das medidas de índice de refração a 20°C, para os cinco tipos de biodiesel puro são exibidos na Tabela 15.

Tabela 15 – Resultados das medidas de índice de refração dos cinco tipos de biodiesel puro

Experimental	
Descrição	Média do Índice Refração
Biodiesel de Soja	1,4570
Biodiesel de Canola	1,4555
Biodiesel de Milho	1,4559
Biodiesel de Girassol	1,4573
Biodiesel Região Sul	1,4535

Para as misturas binárias os resultados são apresentados graficamente, o que permite avaliar a variação dos resultados de índice de refração em função da alteração na composição da mistura entre dois diferentes tipos de biodiesel. Nas Figuras 30 a 33 são mostrados os resultados de índice de refração das misturas binárias de biodiesel de milho com biodiesel de soja, biodiesel de canola com biodiesel de soja, biodiesel de girassol com biodiesel de soja e biodiesel de canola com biodiesel fornecido por distribuidora da região sul, respectivamente. Nos gráficos, a barra de erro (I) utilizada corresponde à margem de erro das medições das amostras, sendo calculada a partir da distribuição *t* de Student com 95% de confiança. Nos gráficos também é possível observar os valores individuais de índice de refração de cada amostra (■) e a média das amostras (●). As Figuras 30, 31, 32 e 33 demonstram o comportamento esperado, ou seja, conforme adição do segundo tipo de biodiesel no biodiesel base (soja ou distribuidora da região sul) o índice de refração do biocombustível sofre alterações na composição, aproximando-se do valor especificado do índice de refração do biodiesel adicionado. No entanto, na Figura 32, como a média do índice de refração do biodiesel de girassol e do biodiesel de soja são bastante próximos, 1,4573 e 1,4570 respectivamente, não era

esperado uma variação significativa entre os índices de refrações das mistura, visto que a repetibilidade do método de medição do índice de refração é de $5,0 \cdot 10^{-5}$ e a reprodutibilidade é de $1,0 \cdot 10^{-4}$ para um nível de confiança de 95%.

Figura 30 – Resultados das medidas de Índice de Refração para mistura binária de biodiesel de milho com biodiesel de soja

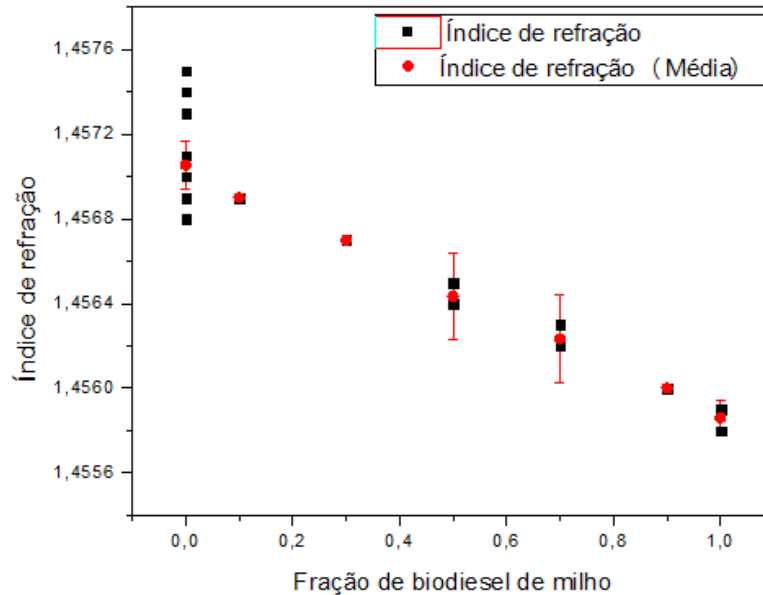


Figura 31 – Resultados das medidas de índice de refração para mistura binária de biodiesel de canola com biodiesel de soja

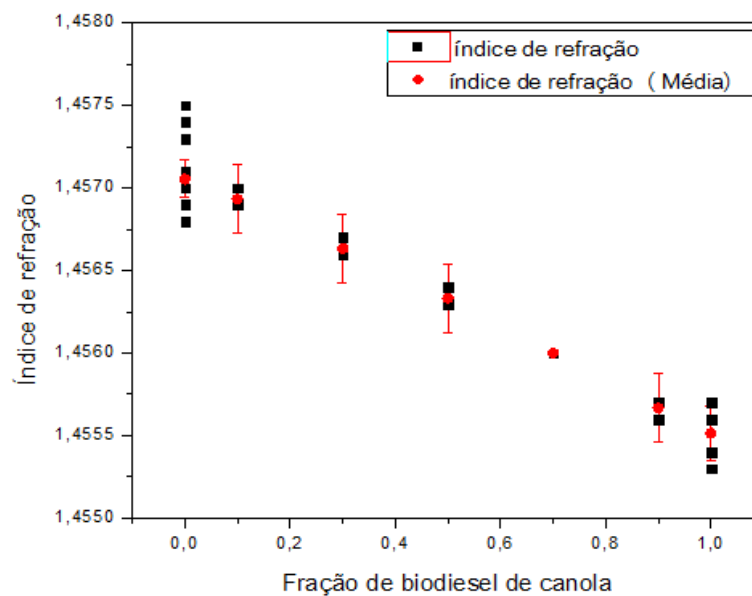


Figura 32 – Resultados das medidas de índice de refração para mistura binária de biodiesel de girassol com biodiesel de soja

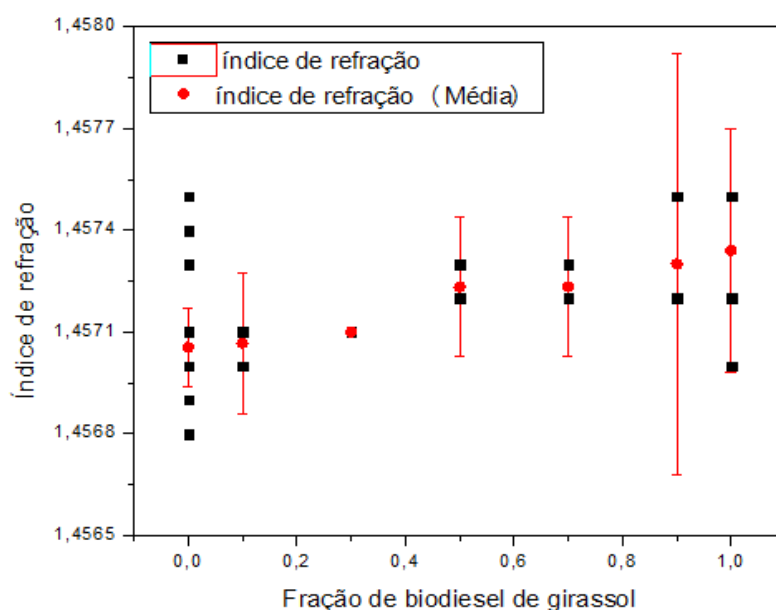
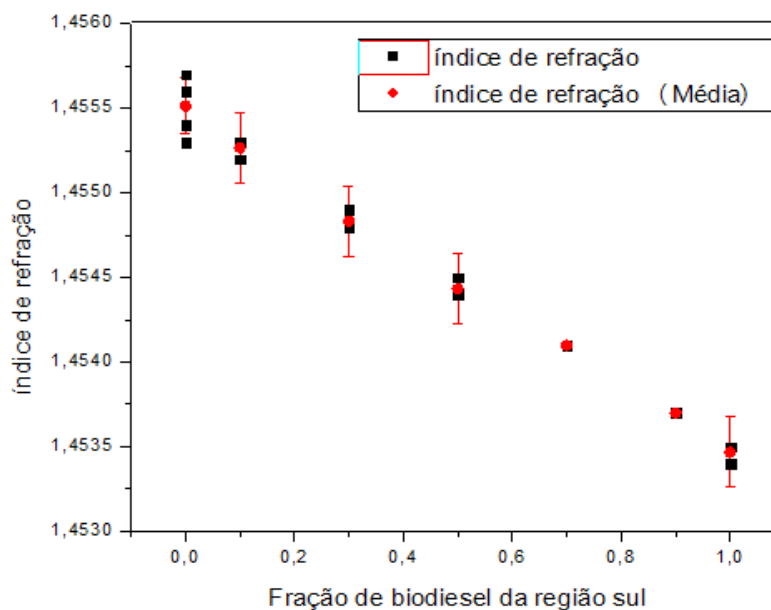


Figura 33 – Resultados das medidas de índice de refração para mistura binária de biodiesel de canola com biodiesel fornecido por distribuidora da região sul



Os resultados de índice de refração para as misturas ternárias são apresentados na Tabela 16 e em diagramas ternários (Figuras 34 a 36), o que torna possível avaliar a variação dos resultados de índice de refração em função da alteração na composição da mistura entre três diferentes tipos de biodiesel. Nas Figuras 34 a 36 são mostrados os resultados médios de índice de refração das

misturas ternárias de biodiesel de milho-canola-soja, biodiesel de milho-girassol-soja e biodiesel de girassol-canola-soja, respectivamente, nas proporções indicadas nos diagramas.

Figura 34 – Resultados de índice de refração para misturas ternárias de biodiesel de milho (A) com biodiesel de canola (B) e biodiesel de soja (C)

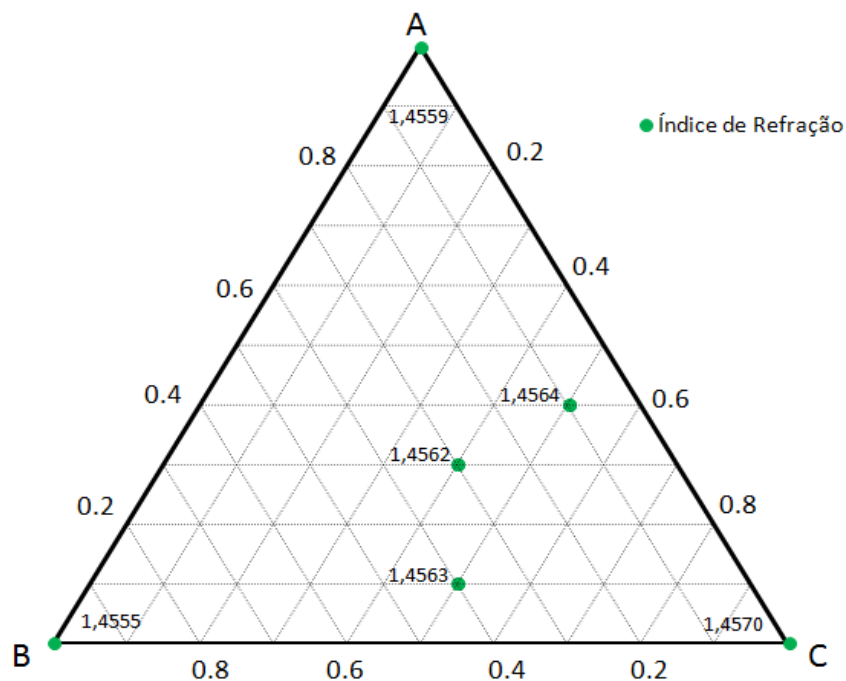


Figura 35 – Resultados de índice de refração para mistura ternária de biodiesel de milho (A) com biodiesel de girassol (B) e biodiesel de soja (C)

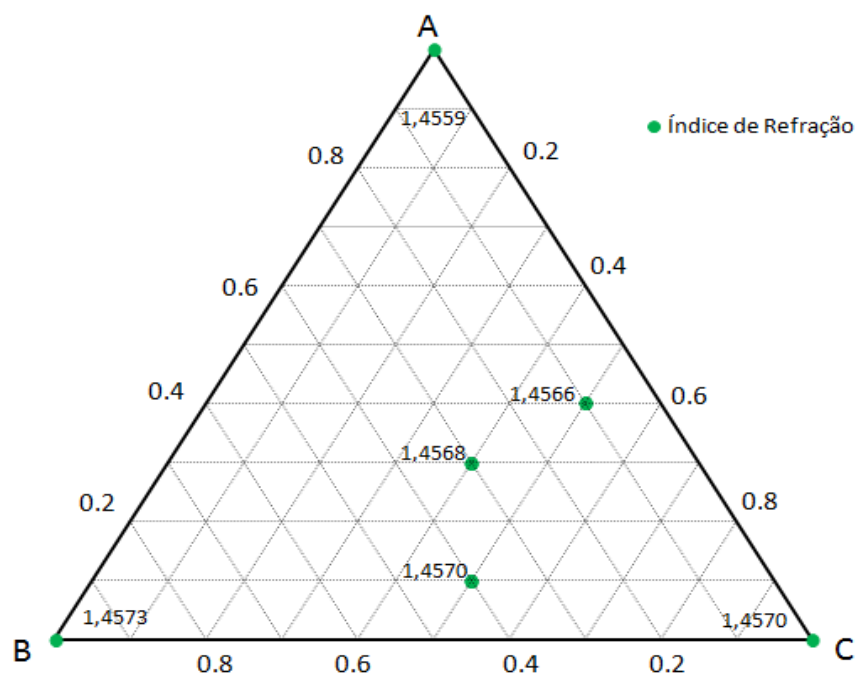
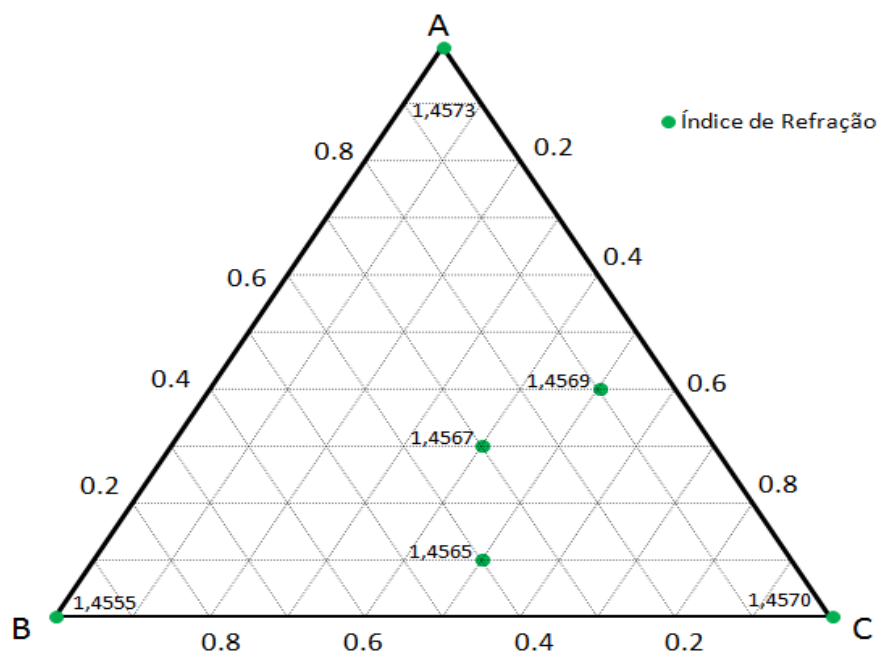


Figura 36 – Resultados índice de refração para mistura ternária de biodiesel de girassol (A) com biodiesel de canola (B) e biodiesel de soja (C)



Para as misturas ternárias e quaternárias (Tabelas 16 e 17) os resultados são apresentados no formato de tabela, onde se observa a variação dos resultados de índice de refração em função da alteração na composição da mistura entre quatro e três diferentes tipos de biodiesel, respectivamente. Na Tabela 17 são exibidos os resultados médios das medidas de índice de refração das misturas quaternárias de biodiesel de canola, girassol, milho e soja, onde a coluna identificada como amostra representa a identificação das amostras, a coluna denominada composição representa a fração volumétrica das misturas entre os diferentes tipos de biodiesel utilizados e, também são apresentados os intervalos estimados com 95% de confiança das medições para as amostras de cada composição.

Tabela 16 – Resultados das medidas de índice de refração para misturas ternárias entre os quatro tipos de biodiesel

Amostra	Composição das misturas (% volumétrica)				Índice de Refração	Intervalo de confiança de 95%
	Biodiesel de canola	Biodiesel de girassol	Biodiesel de milho	Biodiesel de soja		
Am112	10	0	40	50	1,4564	
Am113	10	0	40	50	1,4564	1,4564 ± 0,0000
Am114	10	0	40	50	1,4564	
Am115	30	0	30	40	1,4562	
Am116	30	0	30	40	1,4563	1,4562 ± 0,0004
Am117	30	0	30	40	1,4562	
Am118	40	0	10	50	1,4563	
Am119	40	0	10	50	1,4563	1,4563 ± 0,0000
Am120	40	0	10	50	1,4563	
Am121	0	10	40	50	1,4566	
Am122	0	10	40	50	1,4566	1,4566 ± 0,0000
Am123	0	10	40	50	1,4566	
Am124	0	30	30	40	1,4568	
Am125	0	30	30	40	1,4568	1,4568 ± 0,0000
Am126	0	30	30	40	1,4568	
Am127	0	40	10	50	1,4570	
Am128	0	40	10	50	1,4570	1,4570 ± 0,0000
Am129	0	40	10	50	1,4570	
Am130	40	10	0	50	1,4565	
Am131	40	10	0	50	1,4564	1,4565 ± 0,0002
Am132	40	10	0	50	1,4565	
Am133	30	30	0	40	1,4567	
Am134	30	30	0	40	1,4568	1,4567 ± 0,0004
Am135	30	30	0	40	1,4566	
Am136	10	40	0	50	1,4569	
Am137	10	40	0	50	1,4569	1,4569 ± 0,0000
Am138	10	40	0	50	1,4569	

Tabela 17 – Resultados das medidas de índice de refração para misturas quaternárias entre os quatro tipos de biodiesel

Amostra	Composição das misturas (% volumétrica)				Índice de Refração	Intervalo de confiança de 95%
	Biodiesel de canola	Biodiesel de girassol	Biodiesel de milho	Biodiesel de soja		
Am139	25	25	25	25	1,4564	
Am140	25	25	25	25	1,4565	1,4564 ± 0,0002
Am141	25	25	25	25	1,4564	
Am142	10	10	40	40	1,4565	
Am143	10	10	40	40	1,4564	1,4565 ± 0,0004
Am144	10	10	40	40	1,4566	
Am145	40	10	10	40	1,4563	
Am146	40	10	10	40	1,4563	1,4563 ± 0,0000
Am147	40	10	10	40	1,4563	
Am148	10	40	10	40	1,4568	
Am149	10	40	10	40	1,4568	1,4568 ± 0,0002
Am150	10	40	10	40	1,4569	
Am151	40	10	40	10	1,4560	
Am152	40	10	40	10	1,4560	1,4560 ± 0,0000
Am153	40	10	40	10	1,4560	
Am154	10	40	40	10	1,4566	
Am155	10	40	40	10	1,4565	1,4565 ± 0,0002
Am156	10	40	40	10	1,4565	
Am157	40	40	10	10	1,4564	
Am158	40	40	10	10	1,4564	1,4564 ± 0,0002
Am159	40	40	10	10	1,4565	

3.1.4 Ponto de Fulgor

Para o ponto de fulgor não foi possível obter resultados adequados para o biodiesel, pois o equipamento indicado pela norma não foi capaz de determinar uma medida confiável para o biodiesel, pois os resultados fornecidos pelo equipamento apenas garantem que os pontos de fulgor são maiores que 130°C. O problema neste caso é que a norma indica um equipamento de vaso fechado que opera numa faixa

de temperatura baixa, ideal para análises de diesel, mas que não apresenta a mesma eficiência na análise de biodiesel, visto que para o biodiesel a faixa de temperatura de trabalho do equipamento deve ser mais alta, devido ao elevado ponto de fulgor característico desse combustível.

3.2 Obtenção dos espectros

Neste trabalho foram obtidos espectros de infravermelho próximo e médio para biodiesel puro de cinco origens distintas e suas misturas (binárias, ternárias e quaternárias).

Na Figura 43, são apresentados os espectros brutos obtidos a partir de medidas de infravermelho próximo (NIR) e na Figura 44 são apresentados os mesmos espectros após o ajuste de linha de base. A matriz com os espectros NIR apresenta o sinal em absorvância de medidas de 6001 números de onda (10000 a 4000 cm^{-1}) para 149 amostras. Logo, torna-se possível identificar que o tamanho da matriz dos espectros é 149 x 6001 (amostras *versus* variáveis).

Figura 37 – Espectros brutos de infravermelho próximo (NIR)

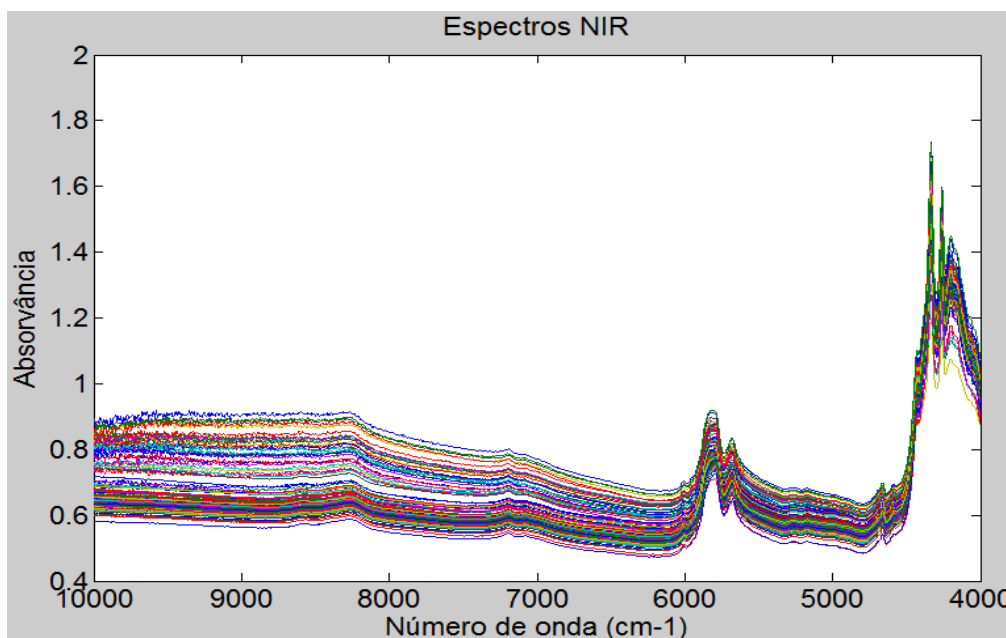
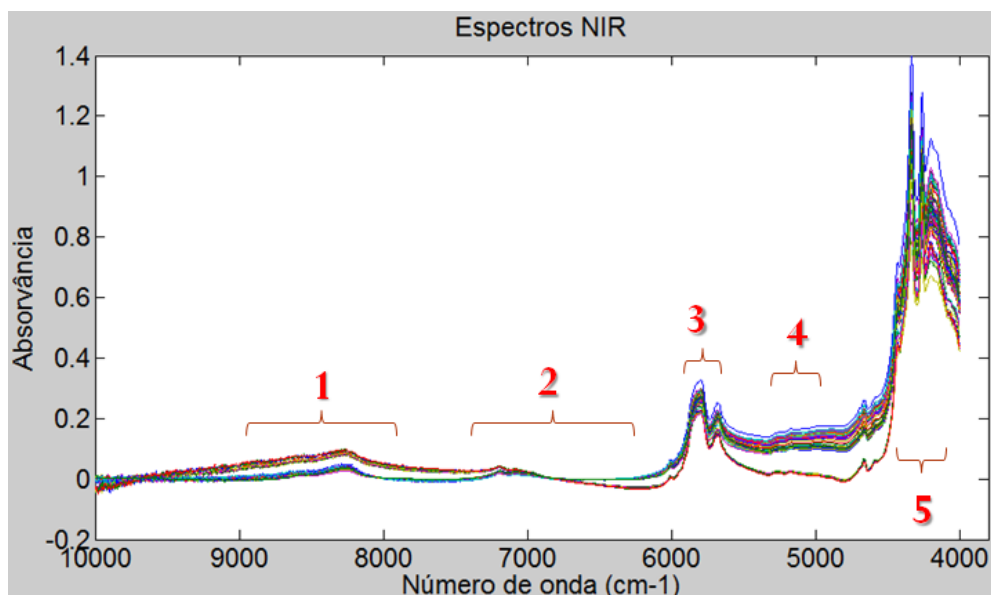


Figura 38 – Espectros NIR após utilizar o pré-tratamento de ajuste de linha de base



As regiões espectrais assinaladas na Figura 44 foram identificadas (Tabela 18), evidenciando a presença combinação de bandas de deformações axiais das ligações C-H, O-H, além de deformações axiais da ligação C=O de derivados carbonilados.

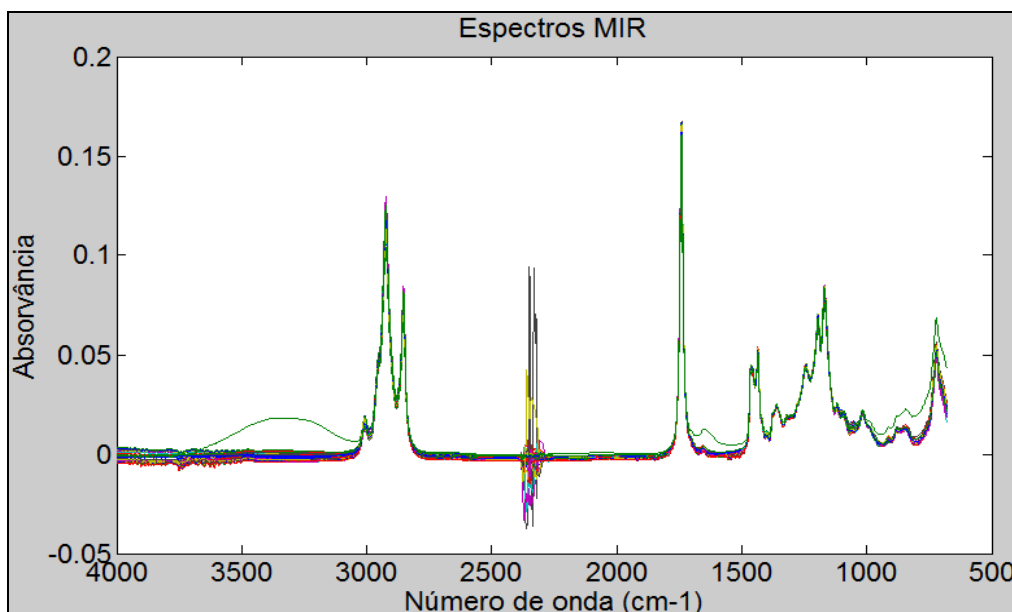
Tabela 18 – Identificação das regiões espectrais (NIR) das amostras de biodiesel

Banda	Nº de onda (cm ⁻¹)*	Grupo provável	Classe das substâncias
1	9000-8000	3vC-H	Alcanos (CH ₃ /CH ₂)
2	7500-6150	2vO-H	Ácido Carboxílico, álcoois (O-H)
3	5750-5700	2vC-H	Alcanos (CH ₃ /CH ₂)
4	5350-4900	3vC=O	Derivados carbonilados
5	4370-4260	1ª Região de Combinação de C-H	Alcanos (CH ₃ /CH ₂)

*(Workman e Weyer, 2008)

Os espectros originais de infravermelho médio (Mid-IR), obtidos para as 149 amostras, são apresentados na Figura 45. Nesta figura é apresentado o espectro no intervalo de 4001 a 681 cm⁻¹. Com isso o tamanho da matriz de espectros Mid-IR é 149 x 3321 (amostras *versus* variáveis).

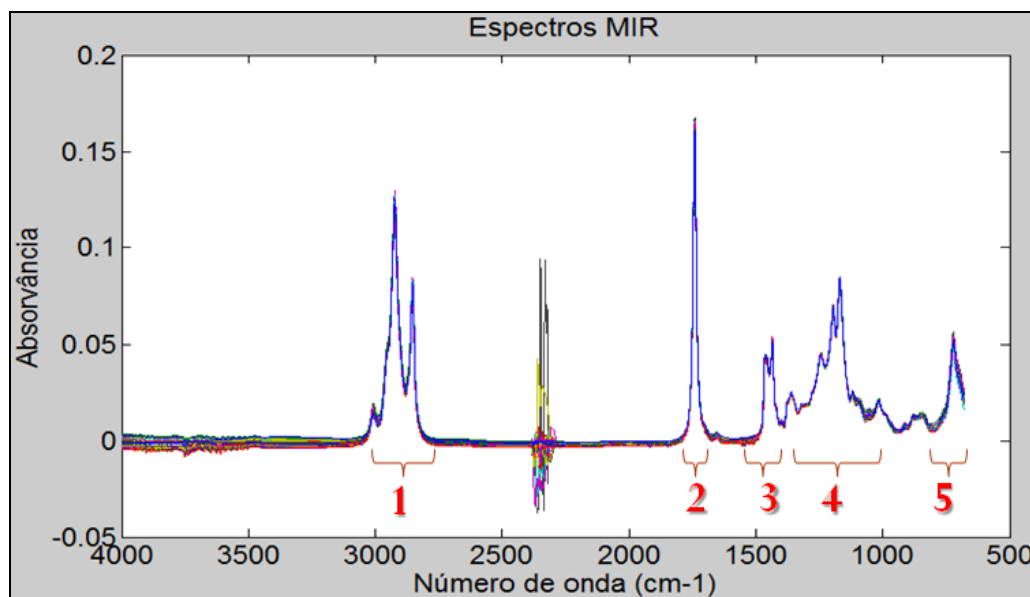
Figura 39 – Espectros brutos de infravermelho médio (Mid-IR)



Como pode ser visto na Figura 45, há uma amostra (Am158) que apresenta comportamento distinto das amostras similares que compõem o conjunto de amostras triplicatas, sendo assim foi necessário retirá-la do conjunto de dados. Na Figura 46 são apresentados os espectros após a remoção da amostra Am158, reduzindo a matriz de 149 x 3321 para 148 x 3321 (amostras *versus* variáveis).

A região espectral entre 2450 a 2300 cm⁻¹, apresentada nas Figura 45 e 46, foi identificada como sendo sinal de resposta do cristal de seleneto de zinco do acessório de reflectância total atenuada (ATR) e não de bandas espectrais das amostras de biodiesel. Por este motivo, optou-se pela retirada desta região do conjunto de dados espectrais de Mid-IR, reduzindo a matriz de 148 x 3321 para 148 x 3170 (amostras *versus* variáveis).

Figura 40 – Espectros brutos de infravermelho médio (Mid-IR) após a retirada da amostra atípica



As regiões espectrais assinaladas na Figura 46 foram identificadas (Tabela 19), evidenciando a presença de bandas de deformações axiais e angulares da ligação C-H, e as deformações axiais das ligações C=O e C-O, respectivamente.

Tabela 19 – Identificação das regiões espectrais (Mid-IR) das amostras de biodiesel

Região	Nº de onda (cm ⁻¹)*	Grupo provável	Classe dos Compostos
1	3000-2840	vC-H	Alcanos
2	1750-1730	vC=O	Éster
3	1500-1400	δC-H	Alcanos
4	1300-1000	vC-O	Ácidos carboxílicos, ésteres
5	750-700	δC-H	(CH ₂) _n ; n ≥ 4

*(Silverstein e Webster, 2000)

Neste estudo, diferentes pré-processamentos foram utilizados na matriz espectral durante a construção dos modelos, como suavização (*Smoothing*), centralização na média (*Mean Center*), derivada de 1ª ordem, variação normal padronizada (SNV), correção de sinal multiplicativo (MSC), correção do sinal ortogonal (OSC), além da correção de linha de base (realizado nos espectros NIR). Na matriz das propriedades estudadas (PEFF, massa específica e índice de refração) foi utilizado o autoescalamento.

3.3 Análise de Componentes Principais

Na análise de componentes principais (PCA) os dados são projetados em dimensões menores, utilizando combinações lineares das variáveis originais, formando as componentes principais (PCs). Como resultados desta técnica estatística são obtidos gráficos de distribuição das amostras (escores) e importância das variáveis (pesos).

Após a retirada de uma amostra atípica, que apresentou uma contaminação evidente no espectro Mid-IR, utilizou-se o PCA para uma análise exploratória dos dados com os dados espectrais das demais amostras. O estudo envolvendo a análise de componentes principais foi realizado para o banco de dados contendo os espectros de infravermelho próximo (NIR) e médio (Mid-IR) com o objetivo de realizar uma análise exploratória dos dados.

3.3.1 Banco de dados dos espectros de infravermelho próximo (NIRS)

Para o banco de dados de espectros de infravermelho próximo, o estudo foi realizado utilizando combinações de pré-processamentos, além do já utilizado correção de linha de base. Os pré-processamentos testados para a matriz dos espectros de NIRS (149 amostras x 6001 variáveis) foram centralização dos dados na média (*Mean Center*, MC), 1ª derivada, variação normal padronizada (SNV) e correção do sinal multiplicativo (MSC). Na etapa de validação do modelo, utilizou-se o procedimento da validação cruzada com a seleção de blocos contíguos (com 10 divisões), a qual se mostrou adequado para o tamanho dos bancos.

Para facilitar a escolha do pré-processamento mais adequado foi construída a Tabela 20 para comparar os resultados obtidos a partir dos modelos de PCA obtidos com os espectros NIR. Dentre os pré-processamentos testados, o modelo de PCA que proporcionou o melhor resultado foi a 1ª derivada do espectro, visto que apresentaram baixos valores de RMSEC (raiz quadrada do erro médio quadrático de calibração) e RMSECV (raiz quadrada do erro médio quadrático de validação cruzada) para os dados de espectros de infravermelho próximo.

Tabela 20 – Comparativo dos resultados obtidos pela PCA utilizando diferentes pré-processamentos nos espectros NIR

Pré-processamentos	Figuras de Mérito			
	RMSEC	RMSECV	PCs	Variância Capturada (%)
Matriz X (espectros)				
-	0,0038	15,722	2	99,97
MC	0,0034	1,244	2	97,48
1ª Derivada	0,0001	0,1343	2	99,61
MSC	0,0033	15,632	2	99,98
SNV	0,0163	62,799	2	99,97

A escolha do número de componentes principais pode ser feita baseando-se nos autovalores, nos valores de RMSEC (raiz quadrada do erro médio quadrático de calibração) ou RMSECV (raiz quadrada do erro médio quadrático de validação cruzada). No entanto, neste estudo foi adotado o autovalor como o critério para avaliação do número de componentes principais.

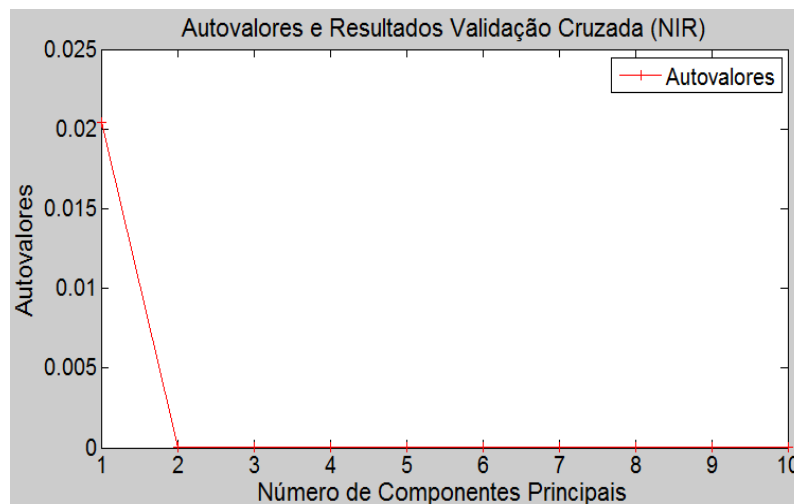
Após construção do modelo utilizando o pré-processamento 1ª derivada foi identificado à presença de uma amostra de comportamento atípico no conjunto dados, a amostra 16 (Am036). Posteriormente, foi gerado um novo modelo a partir da retirada desta amostra atípica, neste segundo modelo também foi identificada uma amostra apresentando comportamento atípico.

Deve-se ressaltar que quando é gerado um novo modelo a partir da retirada da amostra atípica de um conjunto, o novo modelo pode apresentar amostras atípicas não detectadas inicialmente no modelo anterior, isto pode ocorrer incessantemente a cada novo modelo construído. Este comportamento é explicado pela norma ASTM E1655-05, que classifica esta ocorrência como “*snowballing effect*”. Neste caso, a norma aconselha detectar e remover a amostra atípica somente até o segundo modelo (ASTM, 2005). De acordo com o que aconselha a norma ASTM E1655-05, para este estudo foi possível retirar a amostra 15 (Am035), visto que ela se enquadra como sendo a amostra atípica pertencente ao segundo modelo.

Ao retirar as amostras 16 (Am036) e 15 (Am035), referentes ao primeiro e segundo modelo de PCA, do conjunto de espectros de NIR e refazer a análise dos componentes principais (PCA), utilizando o pré-processamento 1ª derivada, foi

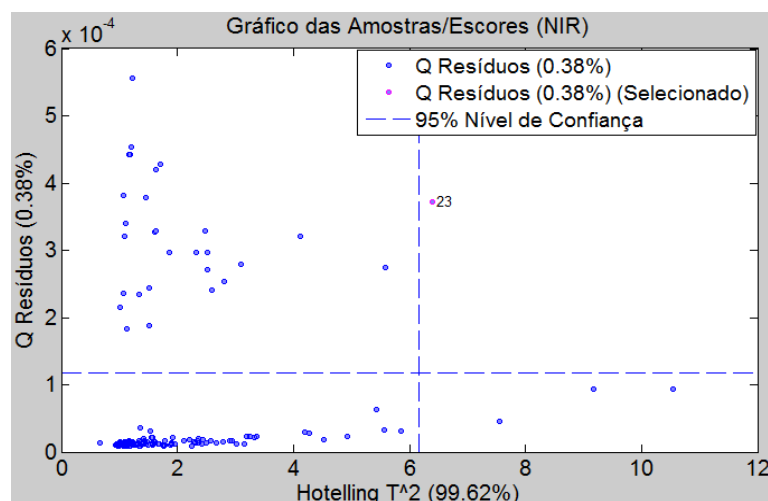
gerado o gráfico da Figura 47. Mais uma vez, o modelo do PCA com dois PCs mostrou-se adequado para a descrição do banco de dados.

Figura 41 – Gráfico de autovalores *versus* número de componentes principais (PC) para o banco de dados de NIR



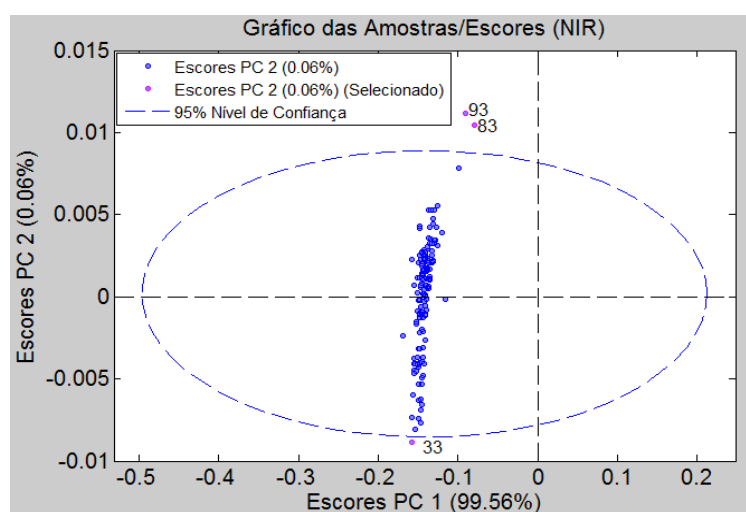
Após a construção do novo modelo da PCA, foi gerado o gráfico dos resíduos *versus* escores (Q Resíduos *versus* Hotelling T^2), conforme mostrado na Figura 48. Observa-se a presença de uma nova amostra atípica, a amostra 23 (Am043), referente ao biodiesel puro de canola, que está situada no quadrante superior direito do gráfico de resíduos *versus* escores, ou seja, apresenta alta importância no modelo associado a elevado valor de resíduo. No entanto, conforme a norma ASTM E1655-05 é recomendado que esta amostra seja mantida no modelo, devido à ocorrência do efeito bola de neve (do inglês, *snowballing effect*), como explicado anteriormente. A Figura 48 apresenta o gráfico dos resíduos *versus* escores (Q Resíduos *versus* Hotelling T^2) para o banco de dados de NIR.

Figura 42 – Gráfico dos resíduos *versus* escores (Q Resíduos *versus* Hotelling T^2) para o banco de dados de NIR



A Figura 49 apresenta os gráficos dos escores (PC1 *versus* PC2) para o banco de dados de NIR e como pode ser observado, o PC2 exerce maior influência sobre o comportamento das amostras enquanto o PC1 é responsável por explicar 99,56% da variabilidade dos dados. Ao avaliar a Figura 49, se observa que existem três amostras fora do intervalo de confiança, as amostras 33 (Am061), 83 (Am113) e 93 (Am123), referentes ao biodiesel fornecido por distribuidora da região sul. As últimas duas amostras são misturas ternárias, mas que não apresentam qualquer anormalidade que justifique a sua retirada do banco de dados. Desta forma as amostras 33, 83 e 93 foram mantidas no atual conjunto.

Figura 43 – Gráfico dos escores (PC1 *versus* PC2) para o banco de dados de NIR



Os resultados obtidos neste estudo podem ser comparados a outros encontrados na literatura, tais como os de Baptista *et al.* (2008(a)) e Pinzi *et al.* (2012). Baptista *et al.* (2008(a)) realizaram, por meio da PCA, uma análise qualitativa dos espectros de NIR de biodiesel. Neste estudo, dois PCs conseguiram capturar 90,61 % da variância total, sendo que o PC1 capturou 80,22% e o PC2 10,39%. Ao mesmo tempo, Pinzi *et al.* (2012) estudaram a similaridade entre as amostras de biodiesel utilizando espectros de infravermelho próximo e modelos de PCA, conseguindo com dois PCs explicar 49% da variância. Neste caso, o PC1 capturou 46% da variância e o PC2 3%.

Neste estudo realizado, o PC1 conseguiu explicar mais de 99% da variância do banco de dados do NIR, enquanto o segundo componente principal explicou uma variância inferior a 1% dos dados originais.

3.3.2 Banco de dados dos espectros de infravermelho médio (Mid-IR)

Para o banco de dados de infravermelho médio (Mid-IR), utilizou-se uma matriz com 148 amostras e 3170 variáveis. Este estudo foi realizado utilizando as mesmas combinações de pré-processamentos testados para a matriz dos espectros de NIR. Na etapa de validação do modelo, utilizou-se o procedimento da validação cruzada com a seleção de blocos contíguos (com 10 divisões), a qual se mostrou adequada para o tamanho dos bancos.

Para facilitar a apresentação dos modelos PCA gerados a partir de diferentes pré-processamentos, foi construída a Tabela 21 para comparar os resultados obtidos a partir dos espectros Mid-IR.

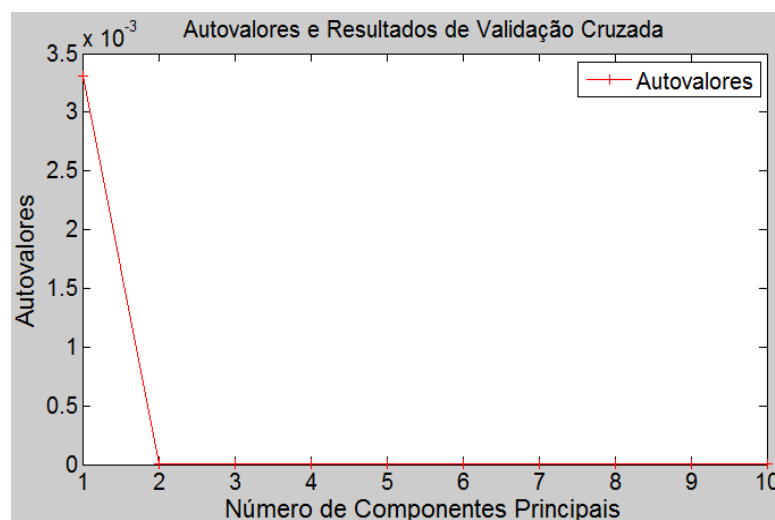
Tabela 21 – Comparativo dos resultados obtidos pelo PCA utilizando diferentes pré-processamentos nos espectros Mid-IR

Pré-processamentos	Figuras de Mérito				
	Matriz X (espectros)	RMSEC	RMSECV	PCs	Variância Capturada (%)
-		0,0009	0,3462	2	99,83
MC		0,0006	0,0548	3	78,45
1ª Derivada		0,0000	0,0354	2	99,97
MSC		0,0008	0,3535	2	99,85
SNV		0,0416	15,051	2	99,83

Dentre os pré-processamentos testados, o modelo de PCA que proporcionou o melhor resultado foi a 1ª derivada do espectro, visto que apresentou baixos valores de RMSEC e RMSECV para os espectros de infravermelho médio (Mid-IR).

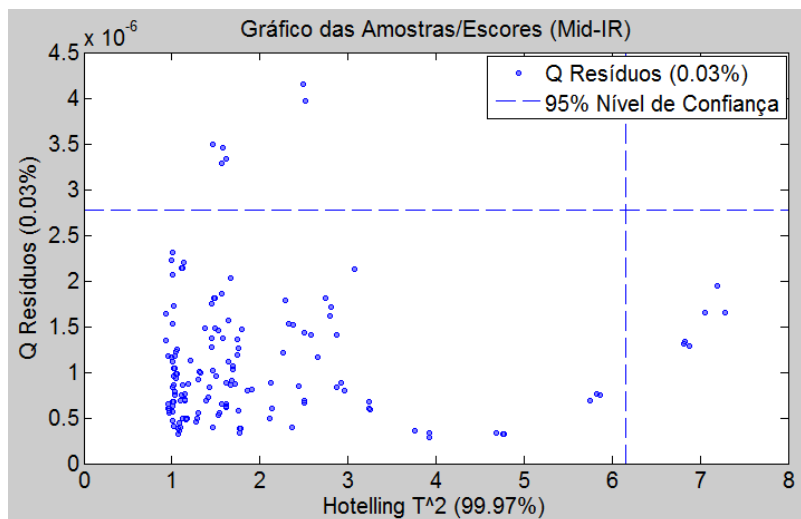
A Figura 50 foi obtida na análise do PCA usando os espectros de Mid-IR, após o pré-processamento da 1ª. derivada. Observando-se os autovalores no gráfico, confirma-se que o número de componentes principais (PCs) a serem utilizados na PCA para o banco de dados de Mid-IR são dois PCs, visto que o modelo explica 99,81% da variância dos dados e apresenta baixos valores de RMSEC e RMSECV.

Figura 44 – Gráfico de autovalores *versus* número de componentes principais (PC) para o banco de dados de Mid-IR



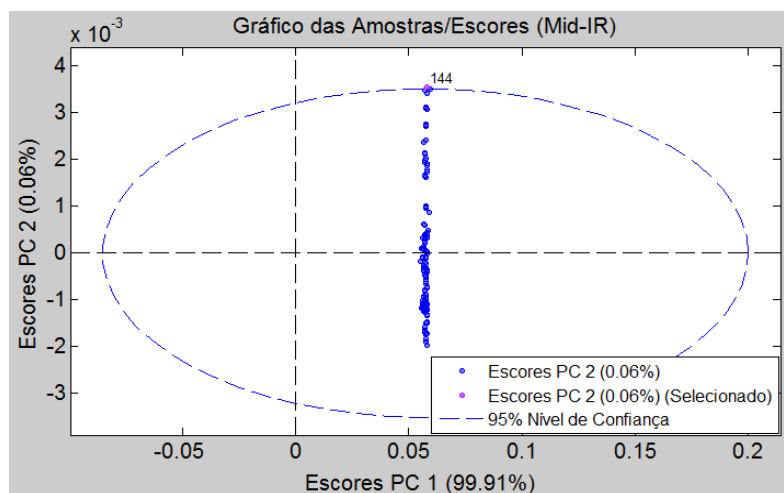
Após a construção do modelo PCA para o banco de espectros de Mid-IR, foi gerado o gráfico dos Q Resíduos *versus* Hotelling T^2 , conforme mostrado na Figura 51. Esta figura mostra que todas as amostras possuem, ao mesmo tempo, valores baixos para Hotelling T^2 e Resíduos, significando assim uma boa homogeneidade dos conjuntos de dados e ausência de amostras atípicas.

Figura 45 – Gráfico dos escores *versus* resíduos (Q Resíduos *versus* Hotelling T^2) para o banco de dados de Mid-IR



A Figura 52 apresenta os gráficos dos escores (PC2 *versus* PC1) para o banco de dados de Mid-IR e como pode ser observado, o PC2 exerce maior influência sobre o comportamento das amostras, mas o PC1 que é responsável por explicar 99,35% da variabilidade dos dados. O PC2 explica 0,47% da variabilidade, o que provavelmente pode ser atribuído aos ruídos dos espectros. Ainda avaliando-se a Figura 52, observa-se que há uma amostra fora do intervalo de 95% de confiança para os espectros de Mid-IR. Esta amostra 144 (Am174) é uma mistura binária de biodiesel de canola e da região sul, mas este fato não indica qualquer anormalidade no banco de dados. Portanto, esta amostra foi mantida no banco de dados.

Figura 46 – Gráfico dos escores (PC2 versus PC1) para o banco de dados de Mid-IR



De posse das informações obtidas com os modelos de PCA, onde foram detectadas e retiradas duas amostras atípicas, 15 (Am035) e 16 (Am036), do conjunto de espectros de NIR e a ausência de amostras atípicas para banco de espectros de Mid-IR, as amostras foram então divididas em dois conjuntos, de calibração e de validação, utilizando o algoritmo de seleção de amostras Kennard-Stone (KS).

Após o emprego do KS, para os espectros de NIR, o conjunto de calibração ficou com 100 amostras e o conjunto de validação 47 amostras, enquanto para os espectros Mid-IR, o conjunto de calibração ficou com 100 amostras e o conjunto de validação 48 amostras, conforme apresentado no Anexo. Após a construção dos conjuntos de calibração e validação, deu-se seguimento a construção de modelos preditivos para as propriedades estudadas, utilizando os espectros NIR e Mid-IR e ferramentas de calibração multivariada de primeira ordem: a regressão por mínimos quadrados parciais (PLS) e a regressão por máquinas de vetores de suporte (SVM).

3.4 Modelos de Previsão das Propriedades

Neste estudo foi utilizado o método de regressão por mínimos quadrados parciais (PLS) para a obtenção de modelos de calibração multivariada para a previsão do índice de refração, massa específica e ponto de entupimento usando os espectros de infravermelho próximo e médio de amostras de biodiesel e suas

misturas. Para o ponto de entupimento foi testado também o algoritmo iPLS, que seleciona as variáveis seguido de regressão por mínimos quadrados parciais, foi aplicado também a regressão por máquinas de vetores suporte com seleção de variáveis realizada por algoritmo genético (GA-SVM). Como o ponto de entupimento é uma propriedade que apresenta comportamento não linear, o método SVM foi escolhido por ser também um método não linear, que por sua vez, pode ser mais adequado ao modelo de previsão desta propriedade. A seguir é apresentado o desenvolvimento deste estudo para cada uma das propriedades físico-químicas citadas anteriormente. Para todos os modelos obtidos abaixo foi escolhida a validação cruzada com blocos contíguos (com 10 divisões) e testados vários pré-processamentos na matriz dos espectros (matriz \mathbf{X}), tais como: centralização dos dados na média, primeira derivada, variação normal padrão (SNV) e correção do espalhamento multiplicativo (MSC). Para o vetor de cada propriedade (vetor \mathbf{Y}) foi aplicado o autoescalamento.

3.4.1 Modelos de Previsão para Índice de Refração

Para realizar a previsão da propriedade índice de refração foi utilizado o método de regressão por mínimos quadrados parciais (PLS).

3.4.1.1 Regressão por PLS

Para este estudo foi adotado o RMSEC (raiz quadrada do erro médio quadrático de calibração) como parâmetro de escolha do número de variáveis latentes para construção do modelo. O RMSEC de um modelo é o valor que permite visualizar o número ideal mínimo de variáveis latentes para a construção de um modelo consistente. Por exemplo, as Figuras 53 e 54 foram obtidas utilizando como pré-processamento 1ª derivada e centralização dos dados na média (MC) para os conjuntos espectrais de NIR e Mid-IR, respectivamente.

Figura 47 – Gráfico do RMSEC *versus* número de variáveis latentes para o índice de refração com espectros NIR

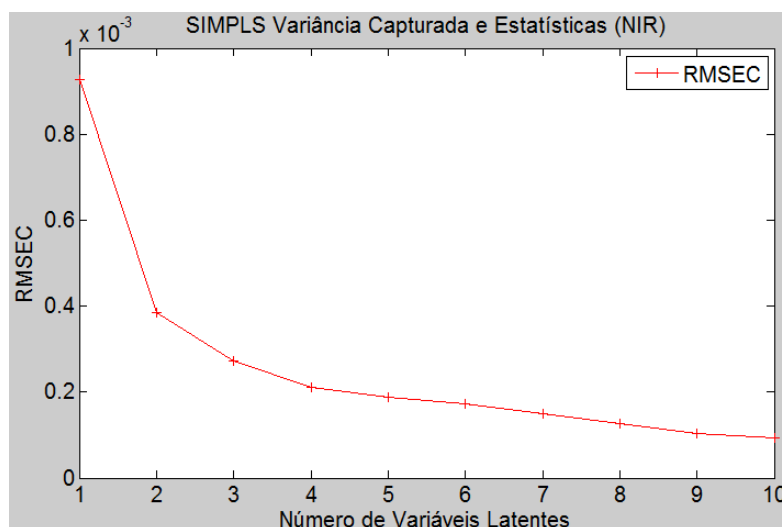
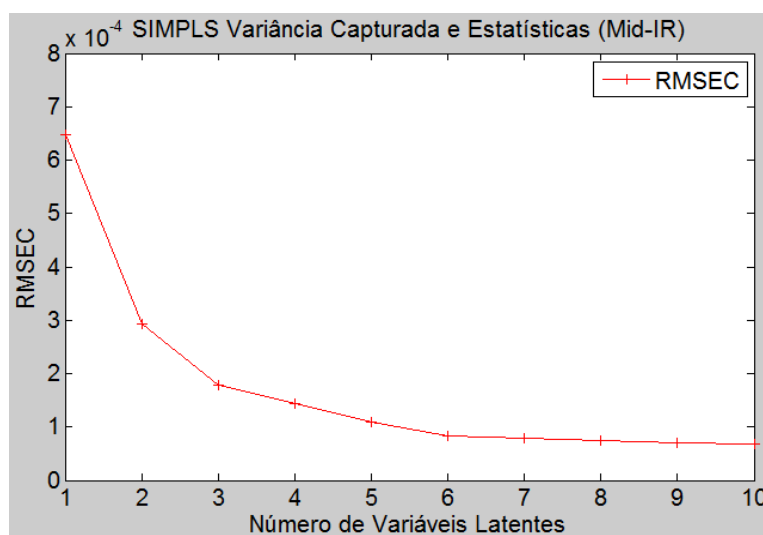


Figura 48 – Gráfico do RMSEC *versus* número de variáveis latentes para o índice de refração com espectros Mid-IR



Quando se observa o valor de RMSEC nos gráficos das Figuras 53 e 54, pode-se afirmar que com três variáveis latentes é possível construir um modelo que correlacione o índice de refração com os dados espectrais de NIR e Mid-IR, apresentando baixos RMSEC (0,0003; 0,0002), RMSECV (0,0003; 0,0002) e RMSEP (0,0003; 0,0001), respectivamente. Dentre os pré-processamentos testados, o melhor resultado foi obtido com a utilização da 1ª derivada para a matriz dos espectros de NIR, centralização dos dados na média para os espectros Mid-IR e autoescalamiento para o vetor contendo os dados de índice de refração. Nestes

modelos foi possível explicar 99,52% e 89,55% da variância dos dados do Bloco X (matriz dos espectros) e 91,38% e 95,93% para o Bloco Y (vetor da propriedade), para os conjuntos de dados de NIR e Mid-IR, respectivamente.

Para o conjunto de dados de Mid-IR, o modelo apresentado foi construído na 3ª tentativa após a retirada de duas amostras atípicas, amostras 6 (Am026) e 9 (Am029).

Após construído os modelos por PLS para o banco de espectros de NIR e Mid-IR, foi gerado o gráfico resíduos estudentizados *versus* *Hotelling T*², conforme mostrado nas Figuras 55 e 56. O gráfico apresentado na Figura 55 mostra que grande parte das amostras possuem, simultaneamente, valores baixos para *Hotelling T*² e resíduos estudentizados (intervalo aceitável entre -3 e +3). A amostra 73 (Am135) que apresenta alto valor de resíduo não pode ser considerada atípica, pois apresenta baixo valor de *Hotelling T*² e o mesmo ocorre com as amostras 4 (Am024), 15 (Am037), 17 (Am039) e 35 (Am069) que embora apresentem um valor de *Hotelling* mais alto, também não são consideradas amostras atípicas devido ao baixo valor de resíduo estudentizado. Deste modo, nenhuma das amostras do conjunto de dados de espectros de NIR devem ser excluídas, significando assim uma boa homogeneidade do conjunto de dados e ausência de amostras atípicas.

Para o banco de espectros de Mid-IR, o gráfico dos resíduos estudentizados *versus* *Hotelling T*² (Figura 56), mostra uma boa homogeneidade dos conjuntos de dados e não apresenta amostras atípicas, assim como na interpretação da Figura 55, conforme foi mencionado anteriormente.

Figura 49 – Gráfico dos resíduos estudentizados *versus* *Hotelling T*² para o índice de refração com espectros NIR

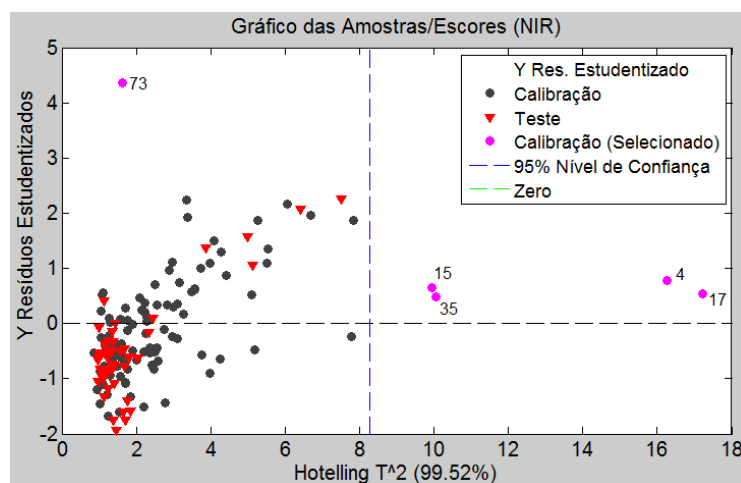
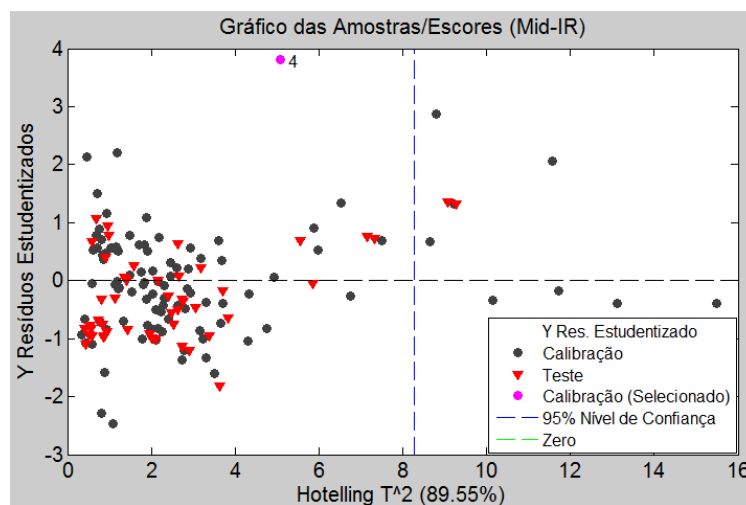


Figura 50 – Gráfico dos resíduos estudentizados *versus* Hotelling T^2 para o índice de refração com espectros Mid-IR



O gráfico dos escores permite observar a presença de agrupamentos ou tendências e é de costume mostrar-se apenas a representação da variável latente 1 *versus* variável latente 2, mas é possível obter um gráfico para qualquer par de variáveis latentes que estejam sendo utilizados na construção do modelo.

Uma região de confiança com 95% pode ser vista nas Figuras 57 e 58 como uma ferramenta para a detecção de amostras com comportamentos anômalos. Ao trabalhar com um intervalo de 95% significa que é esperado que 95% das amostras fiquem contidas dentro deste intervalo de confiança e é aceitável que 5% se encontrem fora deste limite (Erro do tipo I no teste de hipóteses). Ao analisar o gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o banco de NIR (Figura 57) percebe-se que cinco amostras (5%) do conjunto de calibração e uma amostra (2,1%) do conjunto de validação encontram-se fora do limite de confiança de 95%, o que significa uma percentagem aceitável de amostras fora do intervalo de confiança. O gráfico dos escores da variável latente 1 *versus* escores da variável latente 2, para o banco contendo os espectros de Mid-IR, é apresentado na Figura 58 e observa-se que três amostras (3%) do conjunto de calibração permanecem fora do intervalo de confiança de 95%, sendo um comportamento aceitável devido ao nível de confiança. A variável latente 1 é responsável por explicar 99,33% e 65,67% dos dados no modelo de previsão do índice de refração utilizando espectros NIR e Mid-IR, respectivamente.

Figura 51 – Gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o índice de refração com espectros NIR

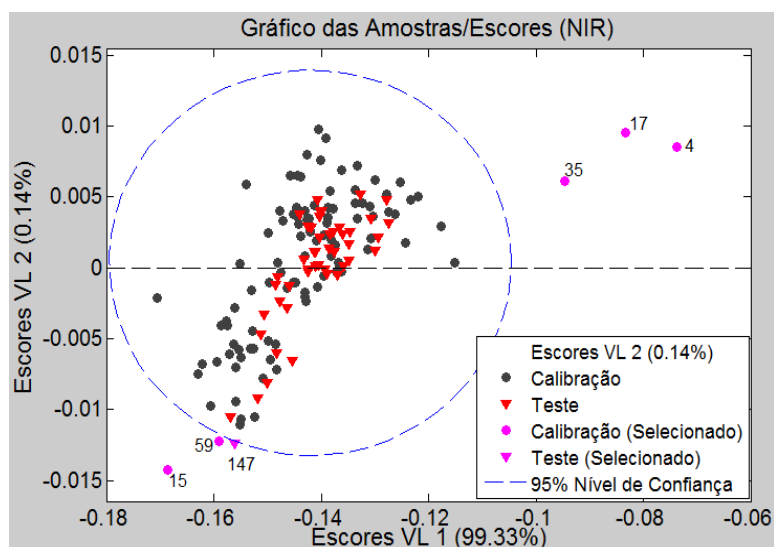
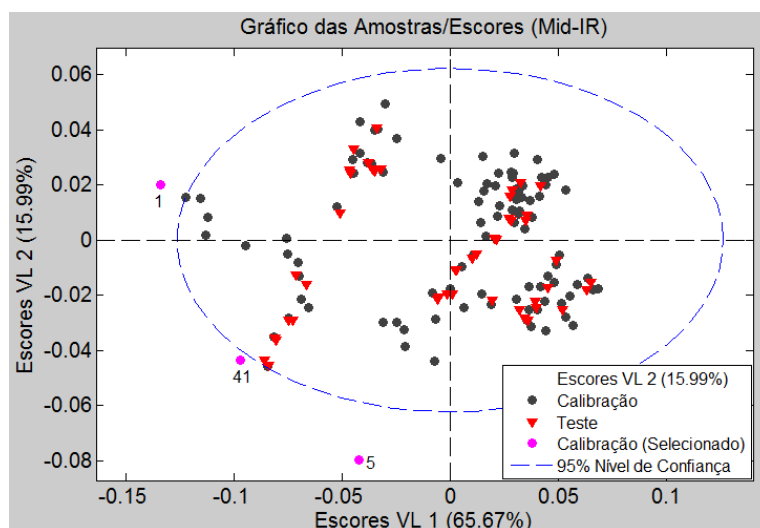


Figura 52 – Gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o índice de refração com espectros Mid-IR



Dos gráficos apresentados nas Figuras 59 e 60, podem ser obtidas informações relevantes sobre os modelos, tais como, o comportamento do valor do Y (índice de refração) que foi medido contra o valor do Y (índice de refração) que foi previsto, o coeficiente e a equação do modelo de previsão do índice de refração utilizando espectros NIR e Mid-IR.

O desempenho do modelo de calibração multivariada pode ser avaliado a partir dos valores de RMSEC, RMSECV, RMSEP e coeficiente de determinação

(R^2), sendo que modelos aceitáveis devem apresentar erros baixos, RMSEC, RMSECV e RMSEP. Avaliando o ajuste de modelos (Figuras 59 e 60), ao longo de uma correlação entre os valores de referência e os valores calculados com os modelos de previsão do conjunto, verificou-se um excelente coeficiente de determinação da predição, $R^2_{Pred} = 0,96$ e $0,98$, para o modelo de previsão do índice de refração utilizando espectros de NIR e Mid-IR, com três variáveis latentes e RMSEP de $0,0003$ e $0,0001$, respectivamente.

Figura 53 – Gráfico do Y previsto versus Y medido para o índice de refração com espectros NIR

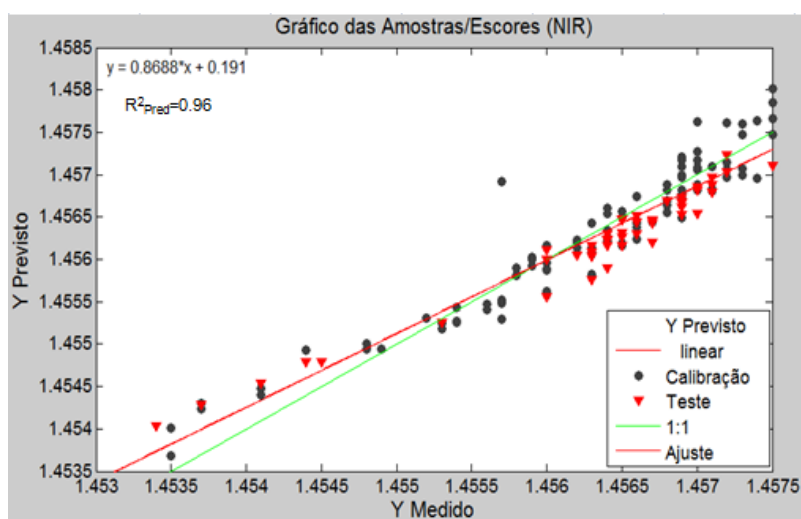
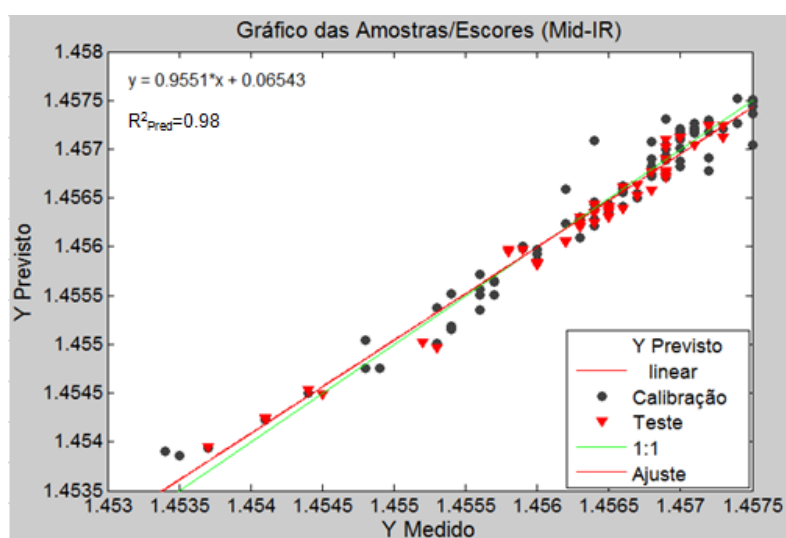


Figura 54 – Gráfico do Y previsto versus Y medido para o índice de refração com espectros Mid-IR



Não foram encontrados estudos sobre modelos de previsão do índice de refração do biodiesel e misturas de biodiesel utilizando espectros de NIR e ou Mid-IR. Isto pode ser atribuído ao fato de que estamos lidando com um novo modo de estimar estes parâmetros para o biodiesel usando a quimiometria. No entanto, vale ressaltar que esta técnica foi empregada em óleos vegetais em um trabalho de nosso grupo de pesquisa.

Os resultados obtidos neste estudo podem ser comparados a outros estudos similares encontrados na literatura, tais como o de Pereira (2007) que desenvolveu uma metodologia para a determinação índice de refração em quatro tipos de óleos vegetais (soja, milho, girassol e canola) utilizando espectros de NIR e com três variáveis latentes conseguiram construir um modelo por PLS com RMSEP igual a 0,0006 e R^2_{Pred} de 0,83. Neste estudo, a previsão do índice de refração para misturas de biodiesel foi realizada utilizando espectros de NIR e Mid-IR, que com três variáveis latentes foram construídos modelos por PLS com RMSEP igual a 0,0003 e 0,0001 e R^2_{Pred} 0,96 e 0,98, respectivamente.

O quadro comparativo da Tabela 22 mostra os resultados dos parâmetros de mérito dos bancos de dados, com os valores de RMSEC, RMSECV e RMSEP e dos coeficientes de determinação da calibração, da validação cruzada e da previsão, respectivamente R^2_{Cal} , R^2_{CV} e R^2_{Pred} . Aparentemente, a regressão por PLS para modelar as previsões de índice de refração para o banco de Mid-IR apresentaram resultado superior aos encontrados utilizando o banco de NIR, visto que os erros são menores e os coeficientes de determinação são maiores.

Tabela 22 – Quadro comparativo dos parâmetros de mérito entre os dois bancos de dados (NIR e Mid-IR) obtidos a previsão de índice de refração usando a técnica PLS

	NIR	Mid-IR
RMSEC	0,0003	0,0002
RMSECV	0,0003	0,0002
RMSEP	0,0003	0,0001
R^2_{cal}	0,91	0,96
R^2_{cv}	0,91	0,95
R^2_{Pred}	0,96	0,98

Para afirmar que o modelo apresentado pelo banco de espectros de Mid-IR é diferente do modelo apresentado para banco de NIR, foi realizado o teste F para comprovar estatisticamente se existe ou não diferença entre os modelos apresentados. Para a execução deste teste é necessário o valor de F crítico ($F_{\text{crít.}, 0,05, \nu_1, \nu_2}$) tabelado em função do grau de liberdade do numerador (ν_1) e denominador (ν_2), da Equação 36 e conforme a Equação 36. As letras i e j representam, respectivamente, os modelos que possuem os maiores e menores valores de RMSEP (DOMINGOS *et al.*, 2013).

$$F_{\text{calculado}} = \frac{RMSEP_i^2}{RMSEP_j^2} \quad (36)$$

Para os bancos de dados de espectros de NIR *versus* espectros Mid-IR, $F_{\text{calc.}} = 9,00 > F_{\text{crít.}, 0,05, \nu_1, \nu_2} = 1,64$, portanto pode-se afirmar que estes modelos são diferentes. Isto significa que o modelo utilizando espectros Mid-IR tem melhor desempenho do que o modelo utilizando NIR e apresenta diferença estatisticamente com 95% de confiança comprovada pelo teste F entre os modelos.

3.4.2 Modelos de Previsão para Massa Específica

Para realizar a previsão da propriedade massa específica foi utilizado o métodos de regressão por mínimos quadrados parciais (PLS).

3.4.2.1 Regressão por PLS

O número de variáveis latentes para o modelo, como observado anteriormente, foi escolhido baseando-se no valor de RMSEC (raiz quadrada do erro médio quadrático de calibração), valor o qual permitem visualizar o número ideal mínimo de variáveis latentes para a construção de um modelo com baixos valores de erros. Utilizando o pré-processamento 1ª derivada para o conjunto espectral de NIR e centralização dos dados na média (MC) para os espectros de Mid-IR, foram obtidas as Figuras 61 e 62, respectivamente.

Figura 55 – Gráfico de RMSEC *versus* número de variáveis latentes para a massa específica com espectros NIR

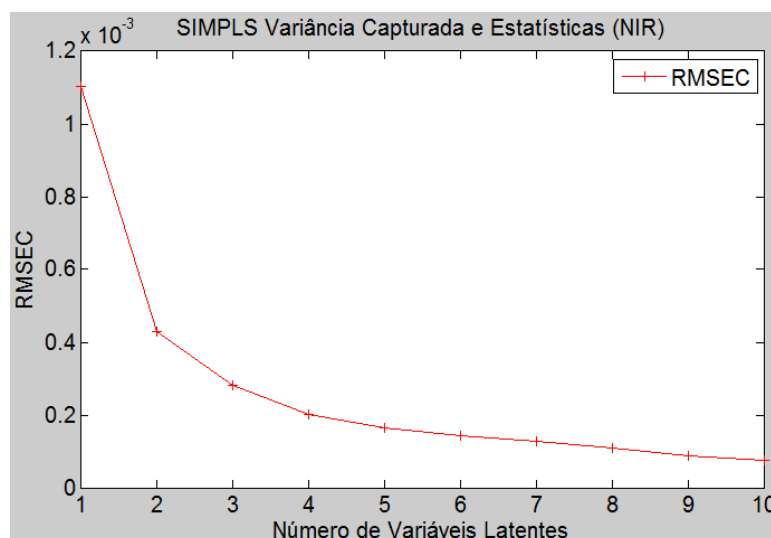
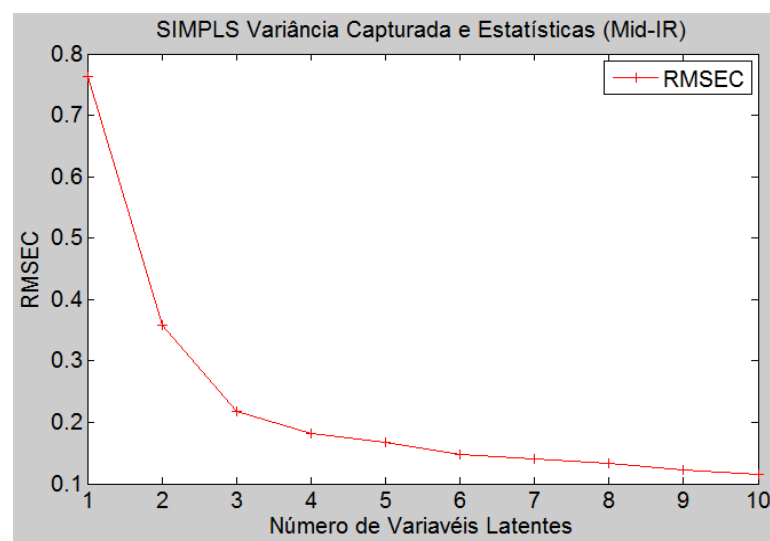


Figura 56 – Gráfico de RMSEC *versus* número de variáveis latentes para a massa específica com espectros Mid-IR



Quando se observam os valores de RMSEC nos gráficos das Figuras 61 e 62, observa-se que com três variáveis latentes é possível construir modelos que correlacionem a massa específica com os dados espectrais de NIR e Mid-IR, apresentando baixos valores de RMSEC (0,0003; 0,0002), RMSECV (0,0003; 0,0002) e RMSEP (0,0003; 0,0002), respectivamente. Dentre os pré-processamentos testados, o melhor resultado foi obtido com a utilização da 1ª derivada para a matriz dos espectros de NIR, centralização dos dados na média

(MC) para os espectros de Mid-IR e autoescalamiento para o vetor contendo os dados de massa específica. Nestes modelos foi possível explicar 99,52% e 89,34% da variância dos dados do Bloco **X** (matriz dos espectros) e 93,48% e 95,82% para o Bloco **Y** (vetor da propriedade), para os conjuntos de dados dos espectros de NIR e Mid-IR, respectivamente.

Depois de construído o modelo por PLS para o banco de espectros de NIR e Mid-IR, foi gerado o gráfico de resíduos estudentizados *versus* *Hotelling T*², conforme mostrado nas Figuras 63 e 64. Os gráficos mostram que todas as amostras possuem, simultaneamente, valores baixos para *Hotelling T*² e resíduos estudentizados, significando assim uma boa homogeneidade do conjunto de dados e ausência de amostras atípicas para os modelos. Sendo assim, todas as amostras desse conjunto devem permanecer no modelo.

Figura 57 – Gráfico dos resíduos estudentizados *versus* *Hotelling T*² para a massa específica com espectros NIR

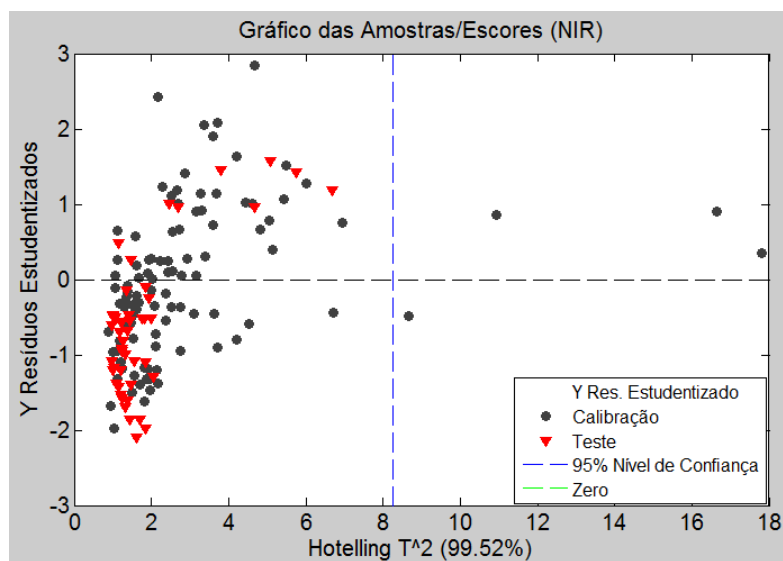
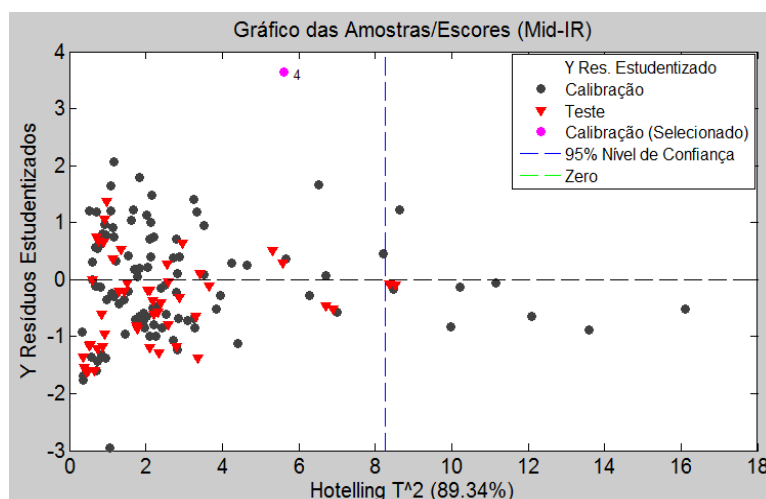


Figura 58 – Gráfico dos resíduos estudentizados *versus* Hotelling T^2 para a massa específica com espectros Mid-IR



As Figuras 65 e 66 mostram o gráfico dos escores. Estes apresentam uma região de confiança com 95% que pode ser usada como uma ferramenta de indicação de amostras com comportamentos atípicos. Ao trabalhar com um intervalo de 95% significa que é esperado que 95% das amostras fiquem contidas dentro deste intervalo de confiança e é aceitável que 5% se encontrem fora deste limite. Ao analisar o gráfico dos escores da variável latente 2 versus escores da variável latente 1 para os bancos de dados dos espectros de NIR e Mid-IR, (Figuras 65 e 66), percebe-se que quatro e três amostras (4% e 3%) dos conjuntos de calibração do espectros NIR e Mid-IR, respectivamente, encontram-se fora do limite de confiança de 95%, o que significa isto representa uma percentagem aceitável de amostras fora do intervalo de confiança.

Figura 59 – Gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para a massa específica com espectros NIR

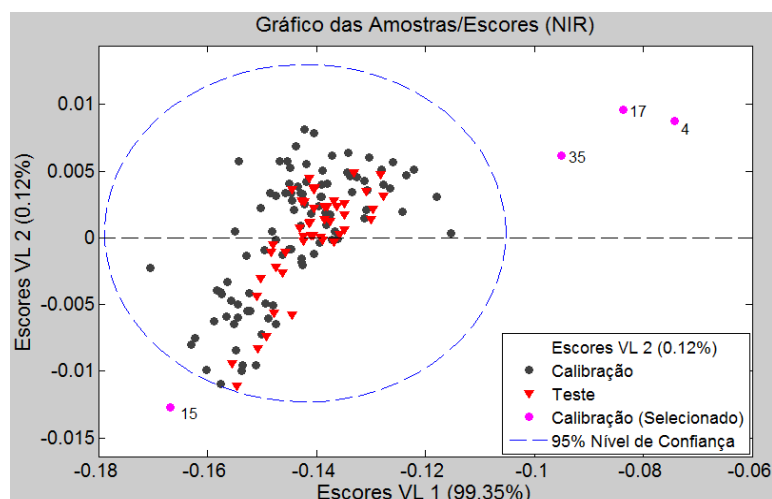
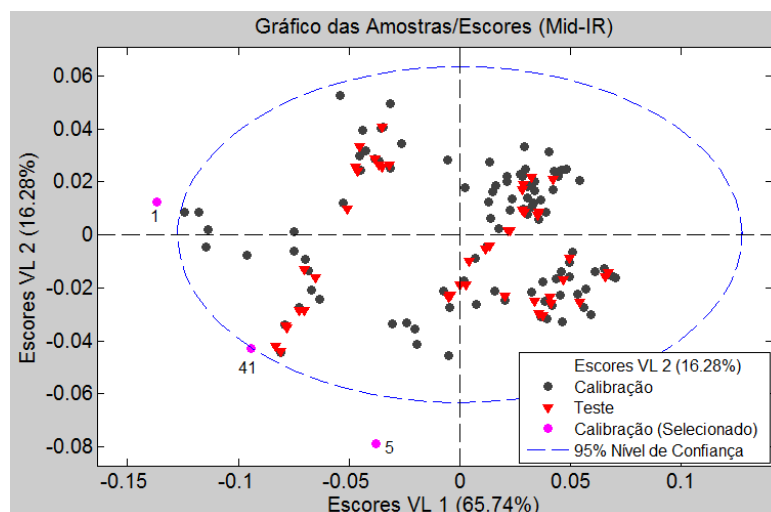


Figura 60 – Gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para a massa específica com espectros Mid-IR



Dos gráficos apresentados nas Figuras 67 e 68 (Y previsto versus Y medido), podem ser obtidas informações relevantes sobre os modelos para previsão da massa específica utilizando espectros NIR e Mid-IR. O desempenho de um modelo de calibração multivariada pode ser avaliado a partir de RMSEC, RMSECV e RMSEP e coeficiente de determinação (R^2), devendo os modelos apresentar valores de erros baixos. Assim, avaliando o ajuste dos modelos (Figuras 67 e 68), no gráfico de dispersão entre os valores de referência e os valores calculados com os modelos de previsão do conjunto, verificou-se um excelente coeficiente de determinação,

$R^2_{Pred}=0,95$ e $0,98$, para o modelo de previsão da massa específica utilizando espectros de NIR e Mid-IR, com três variáveis latentes e RMSEP de $0,0003$ e $0,0002$, respectivamente.

Figura 61 – Gráfico do Y previsto versus Y medido para a massa específica com espectros NIR

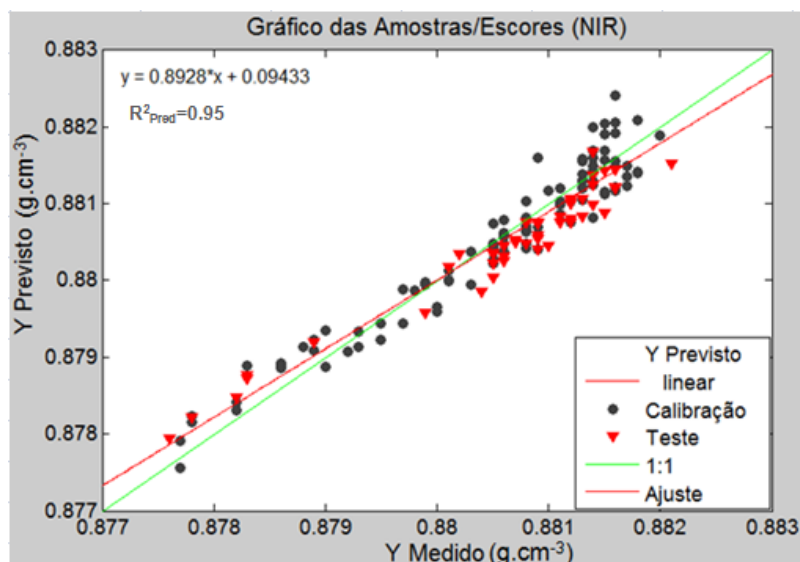
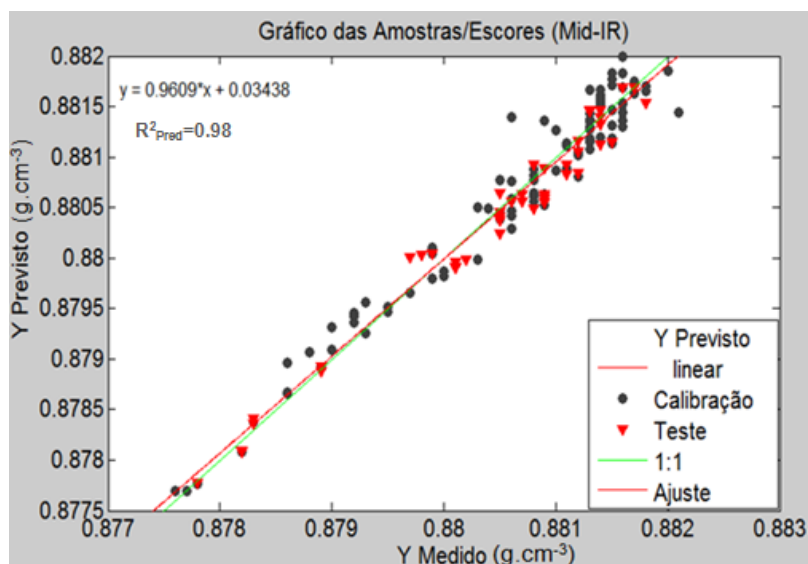


Figura 62 – Gráfico do Y previsto versus Y medido para a massa específica com espectros Mid-IR



Os resultados obtidos neste estudo podem ser comparados a outros estudos similares encontrados na literatura, tais como o de Baptista *et al.* (2008(b)) e Ferrão *et al.* (2011). Baptista *et al.* (2008(b)) desenvolveram uma metodologia para a

determinação da massa específica de misturas de biodiesel produzidos a partir de misturas de soja, palma, colza e óleo de fritura utilizando NIRS, no intervalo entre 9.000 a 4.500 cm^{-1} , e com seis variáveis latentes. Os autores construíram um modelo PLS com RMSEP (Kg/m^3) igual a 0,90 e R^2_{Pred} de 0,99. Ferrão *et al.* (2011) obtiveram um modelo de previsão de PLS, com oito variáveis latentes, para a massa específica com RMSEP (Kg/m^3) igual a 0,73 e R^2_{Pred} igual a 0,99, utilizando dados de NIRS.

Neste estudo, a previsão da massa específica para misturas de biodiesel foi realizada utilizando modelos PLS com espectros de NIR e Mid-IR, com três variáveis latentes, com RMSEP (g/cm^3) igual a 0,0003 e 0,0002 e R^2_{Pred} igual a 0,95 e 0,98, respectivamente.

O quadro comparativo da Tabela 23 mostra os resultados dos parâmetros de mérito dos bancos de dados de dados de espectros de NIR e Mid-IR, com os valores de RMSEC, RMSECV e RMSEP e dos coeficientes de determinação da calibração, da validação cruzada e da previsão, respectivamente R^2_{Cal} , R^2_{CV} e R^2_{Pred} . Novamente, a técnica PLS demonstra ter tido mais êxito em modelar as previsões de massa específica para o banco de espectros Mid-IR, pois os valores de RMSEC, RMSECV e RMSEP são os menores e os valores de R^2_{Cal} , R^2_{CV} e R^2_{Pred} são os maiores. Os resultados utilizando os espectros NIR foram ligeiramente inferiores quando comparados aos obtidos utilizando espectros Mid-IR.

Tabela 23 – Quadro Comparativo dos parâmetros de mérito entre os dois bancos de dados (NIR e Mid-IR) obtidos a previsão de massa específica usando a técnica PLS

	NIR	Mid-IR
RMSEC	0,0003	0,0002
RMSECV	0,0003	0,0002
RMSEP	0,0003	0,0002
R^2_{cal}	0,93	0,96
R^2_{cv}	0,92	0,95
R^2_{Pred}	0,95	0,98

Para avaliar a afirmação que o modelo apresentado pelo banco de espectros de Mid-IR é superior ao modelo apresentado para banco de NIR, foi feito o teste F para comprovar estatisticamente se existe ou não diferença entre os modelos

apresentados. Para o banco de espectros NIR *versus* espectros Mid-IR, $F_{\text{calc}} = 2,25 > F_{\text{crít.}, 0,05, 1, 1} = 1,64$, portanto pode-se afirmar que estes modelos são diferentes. Isto significa que o modelo utilizando espectros Mid-IR é ligeiramente superior do que o modelo utilizando NIR e apresenta diferença estatisticamente com 95% de confiança comprovada pelo teste F entre os modelos.

3.4.3 Modelos de Previsão para Ponto de Entupimento (PEFF)

Para realizar a previsão da propriedade ponto de entupimento de filtro a frio, foram utilizados os métodos de regressão por mínimos quadrados parciais (PLS), regressão por mínimos quadrados por intervalos (iPLS) e, por fim, a regressão por máquinas de vetor suporte (SVM-R) com seleção de variáveis por algoritmo genético (GA).

3.4.3.1 Regressão por PLS

O número de variáveis latentes para construção do modelo, como observado anteriormente, foi escolhido baseando-se no valor de RMSEC (raiz quadrada do erro médio quadrático de calibração). Utilizando o pré-processamento centralização dos dados na média (MC) para o conjunto espectral de NIR e 1ª derivada para os espectros de Mid-IR foram obtidas as Figuras 69 e 70, respectivamente.

Figura 63 – Gráfico do RMSEC *versus* número de variáveis latentes para o PEFF com espectros NIR

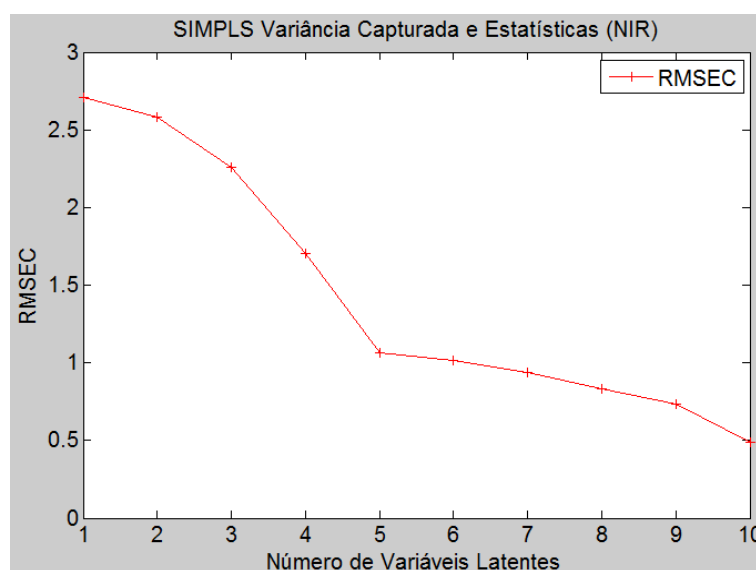
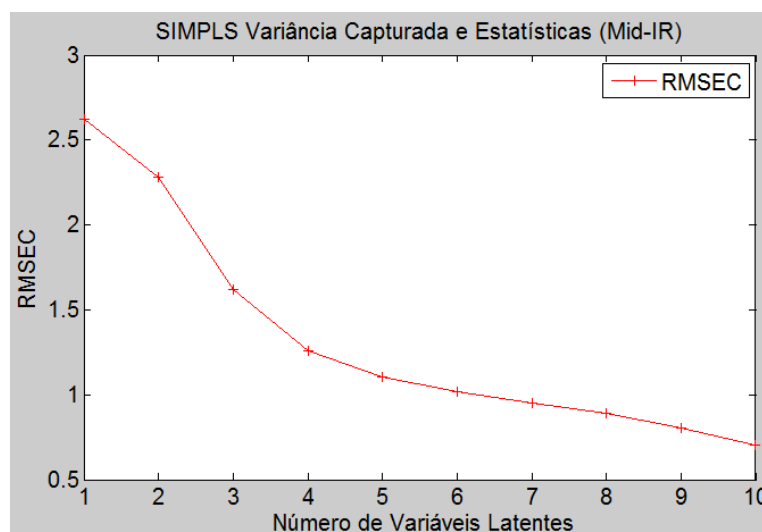


Figura 64 – Gráfico do RMSEC versus número de variáveis latentes para o PEFF com espectros Mid-IR



Quando se observa o valor de RMSEC nos gráficos das Figuras 70 e 71, pode-se afirmar que com cinco e quatro variáveis latentes é possível construir um modelo que correlacione o ponto de entupimento de filtro a frio (PEFF) com os dados espectrais de NIR e Mid-IR, apresentando baixos RMSEC (1,1; 1,3), RMSECV (1,2; 1,4) e RMSEP (1,0; 0,9), respectivamente. Dentre os pré-processamentos testados, o melhor resultado foi obtido com a utilização da centralização dos dados na média para a matriz dos espectros de NIR, 1ª derivada para os espectros de Mid-IR e autoescalamamento para o vetor contendo os dados de PEFF, sendo possível explicar 99,35% e 99,98% da variância dos dados do Bloco **X** (matriz dos espectros) e 84,82% e 77,03% para o Bloco **Y** (vetor da propriedade), para os conjuntos de dados de NIR e Mid-IR, respectivamente.

Para o conjunto de dados de NIR, o modelo apresentado foi construído na 3ª tentativa após a retirada de dezessete amostras atípicas, amostras 1 (Am021), 2 (Am022), 4 (Am024), 6 (Am026), 23 (Am045), 28 (Am058), 30 (Am060) e 34 (Am066), referentes ao biodiesel puro de soja, amostras 9 (Am029) e 11 (Am031), referentes ao biodiesel puro de milho, amostra 13 (Am033), referente ao biodiesel puro de canola, amostra 17 (Am039), referente ao biodiesel puro de girassol, amostra 32 (Am062), referente a mistura de biodiesel da região sul e de soja, amostras 35 (Am069), 37 (Am071) e 42 (Am081), referentes a misturas binárias de

biodiesel de soja e milho, e amostra 43 (Am083), referente a amostra binária de biodiesel de girassol e soja.

Após construído o modelo por PLS para o banco de espectros de NIR e Mid-IR, foi gerado o gráfico dos resíduos estudentizados *versus* Hotelling T^2 , conforme mostrado nas Figuras 71 e 72, respectivamente. Estes gráficos mostram uma boa homogeneidade dos conjuntos de dados e não apresentam amostras atípicas.

Figura 65 – Gráfico dos resíduos estudentizados *versus* Hotelling T^2 para o PEFf com espectros NIR

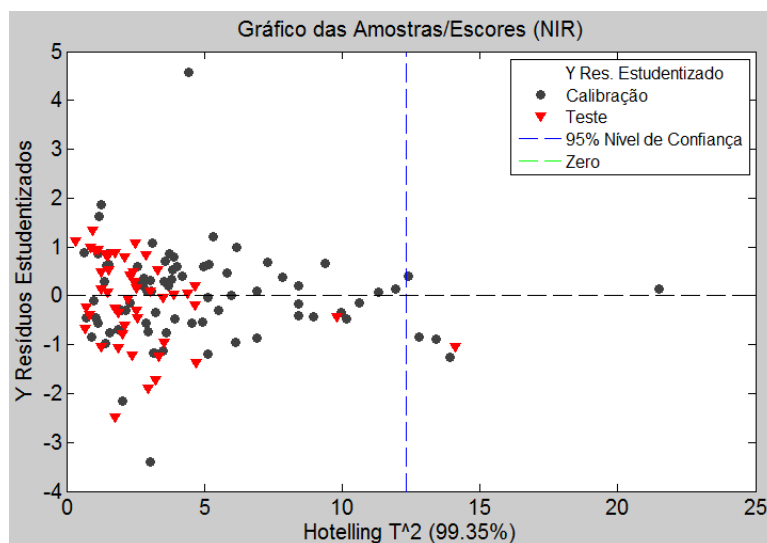
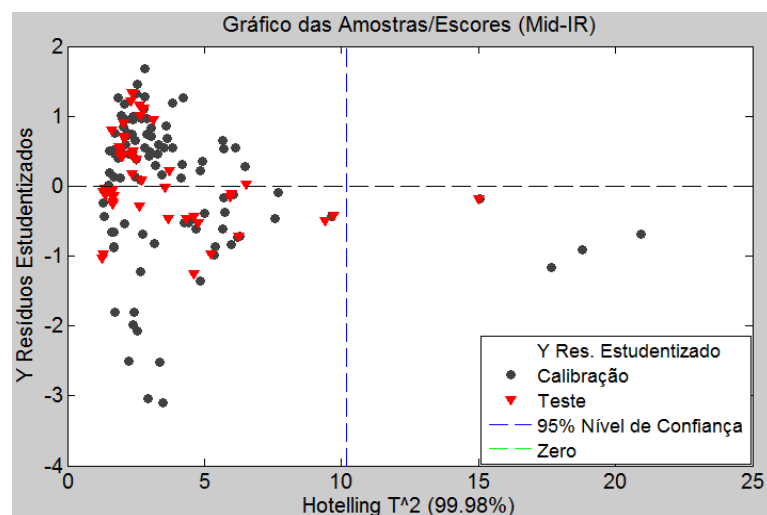


Figura 66 – Gráfico dos resíduos estudentizados *versus* Hotelling T^2 para o PEFf com espectros Mid-IR



O gráfico dos escores permite observar a presença de agrupamentos ou tendências. Uma região de confiança com 95% pode ser vista nas Figuras 73 e 74

como uma ferramenta para a indicação de amostras com comportamentos atípicos. Ao analisar o gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o conjunto de espectros de NIR (Figura 73), percebe-se que 6 amostras (7,7%) do conjunto de calibração do banco de espectros estão fora do limite de confiança de 95%. Como a percentagem aceitável de amostras fora do intervalo é de 5% (Erro do tipo I no teste de hipóteses), as amostras do conjunto de calibração do banco de espectros de NIR apresentam amostras com comportamento atípico acima do aceitável, porém a remoção dessas amostras não é aconselhada pela norma ASTM E1655-05, que permite a retirada de amostras atípicas somente até o segundo modelo.

O gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o banco contendo os espectros de Mid-IR, é apresentado na Figura 74. Observa-se que cinco amostras (5,0%) do conjunto de calibração e quatro amostras (8,3%) do conjunto de validação estão situadas fora do intervalo de confiança de 95%. Para o conjunto de validação dos espectros Mid-IR, as amostras 141 (Am160), 142 (Am162), referentes a misturas binárias entre biodiesel de canola e da região sul, amostras 144 (Am165) e 146 (Am167), referentes a amostras de biodiesel da região sul e canola, respectivamente, foram mantidas no modelo, visto que a retirada das mesmas do conjunto poderia causar perda de informações importantes ao modelo, devido à escassez de amostras com essa informação de composição específica. Por este motivo, a permanência dessas amostras no modelo foi mantida.

Figura 67 – Gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o PEFF com espectros NIR

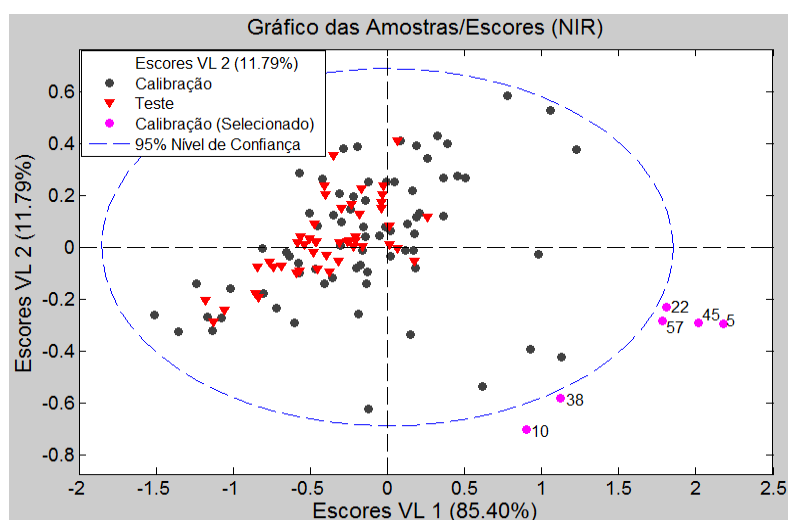
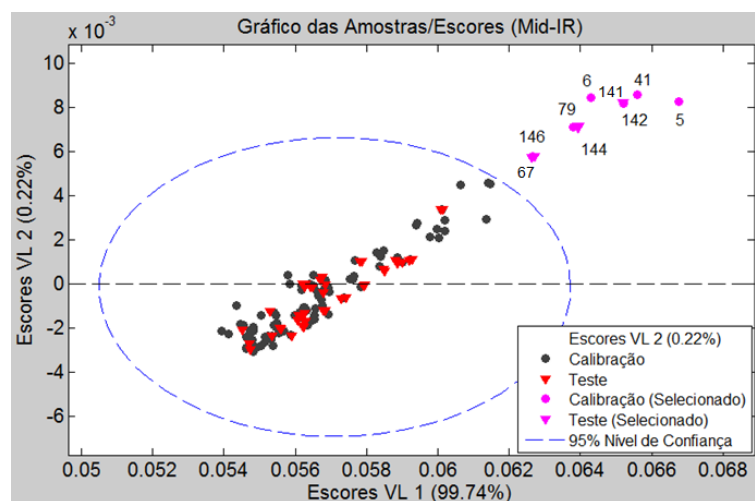


Figura 68 – Gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o PEFF com espectros Mid-IR



O desempenho do modelo de calibração multivariada pode ser avaliado a partir do valor de RMSEC, RMSECV e RMSEP e coeficiente de determinação da previsão (R^2), devendo os modelos apresentar valores de erros baixos. Assim, avaliando o ajuste dos modelos (Figuras 75 e 76), no gráfico de dispersão entre os valores calculados com os modelos de previsão do conjunto e os valores de referência, verificou-se bons coeficientes de determinação, $R^2_{Pred} = 0,86$ e $0,89$, para o modelo de previsão do ponto de entupimento utilizando espectros de NIR e Mid-IR, com cinco e quatro variáveis latentes e RMSEP de 1,0 e 0,9, respectivamente.

Figura 69 – Gráfico do Y previsto *versus* Y medido para o PEFF com espectros NIR

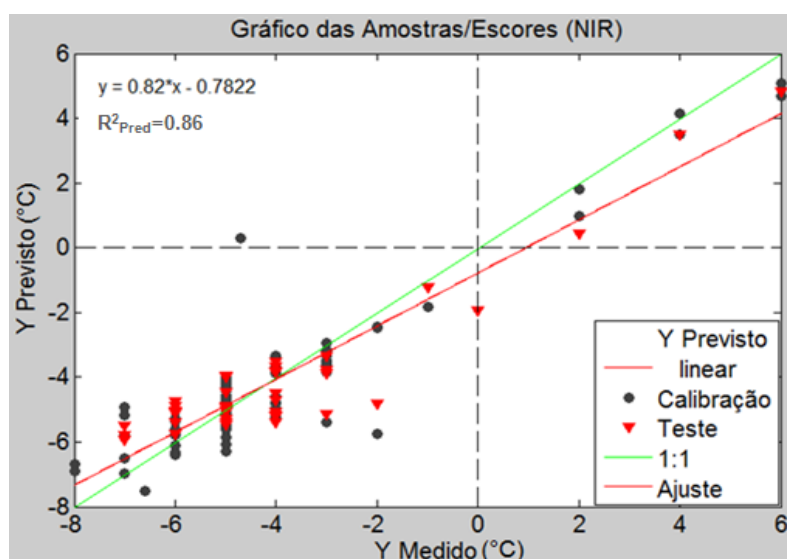
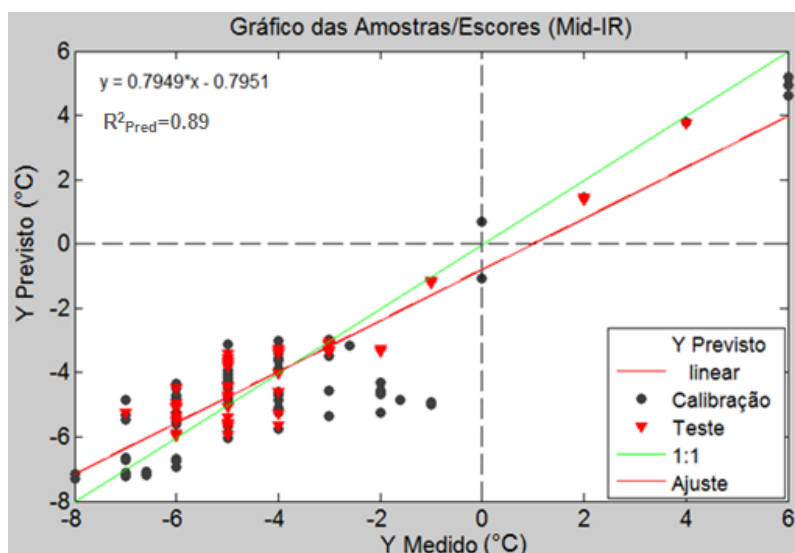


Figura 70 – Gráfico do Y previsto versus Y medido para o PEFF com espectros Mid-IR



Os resultados obtidos neste estudo podem ser comparados a outros estudos similares encontrados na literatura, tais como o de Baptista *et al.* (2008(b)). Baptista *et al.* (2008(b)) desenvolveram uma metodologia para a determinação do ponto de entupimento de filtro a frio (PEFF) de misturas de biodiesel produzidos a partir de misturas de soja, palma, colza e óleo de fritura utilizando espectros de NIR, no intervalo entre 9.000 a 4.500 cm^{-1} . Foi construído um modelo PLS, com 3 variáveis latentes, com RMSEP ($^{\circ}\text{C}$) igual a 1,1 e R^2_{Pred} de 0,93.

Neste estudo, a previsão do ponto de entupimento de filtro a frio (PEFF) para misturas de biodiesel foi realizada utilizando espectros de NIR e Mid-IR. Neste caso, foram construídos modelos PLS, com cinco e quatro variáveis latentes, com RMSEP ($^{\circ}\text{C}$) igual a 1,0 e 0,9 e R^2_{Pred} 0,86 e 0,89, respectivamente.

3.4.3.2 Seleção de Variáveis por iPLS seguido pela Regressão por PLS

Na construção do modelo iPLS, algumas opções foram selecionadas para efetuar o cálculo desse modelo, como o número de intervalos: 10, o tamanho do intervalo (variáveis): 60 e número máximo de variáveis latentes: 8. Os intervalos sugeridos pelo *software* para a realização do PLS, com menor valor de RMSECV, foram: 9039 - 8980, 8679 - 8620, 6819 - 6700, 5859 - 5800, 5739 - 5.560, 4719 - 4660 e 4119 - 4060 cm^{-1} para o conjunto de espectros de NIR (matriz reduzida a 600

variáveis) e 3400 - 3341, 2980 - 2861, 1960 - 1901, 1420 - 1301 e 1240 - 1001 cm^{-1} , para o Mid-IR, (matriz reduzida a 600 variáveis), conforme os intervalos apresentados em verde nas Figuras 77 e 78, respectivamente.

Figura 71 – Gráfico dos intervalos sugeridos para a realização do PLS para o PEFF com espectros

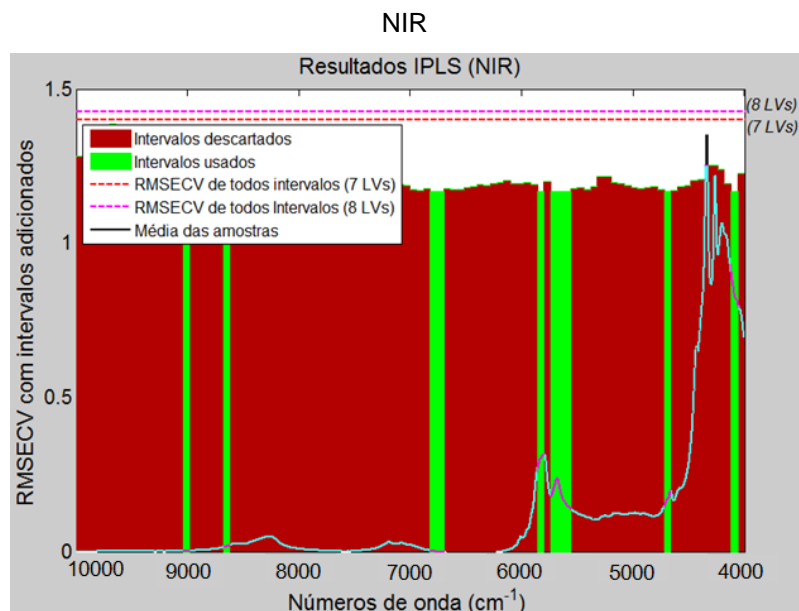
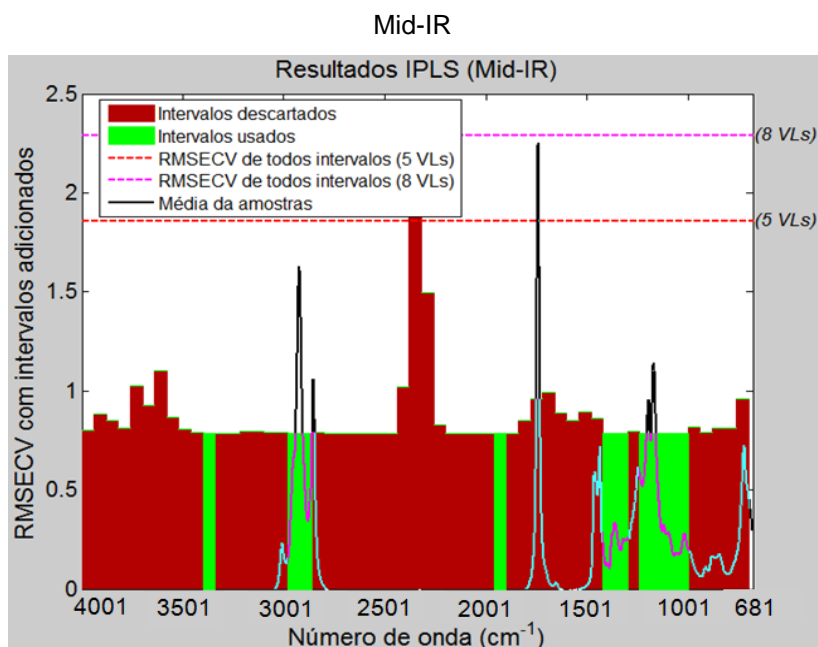


Figura 72 – Gráfico dos intervalos sugeridos para a realização do PLS para o PEFF com espectros



Os intervalos selecionados pelo iPLS para os conjuntos de dados de NIR e Mid-IR foram comparados com os intervalos identificados por Workman e Weyer

(2008) e Silverstein e Webster (2000), respectivamente, conforme apresentado nas Tabelas 24 e 25. Para os conjuntos NIR e Mid-IR, o modelo iPLS conseguiu selecionar 70% e 80% das variáveis (números de onda) com informações químicas relevantes ao biodiesel.

Tabela 24 – Intervalos selecionados pelo iPLS ao conjunto de espectros NIR comparados aos intervalos identificados por Workman e Weyer (2008)

Intervalo selecionado pelo iPLS	Nº de onda (cm ⁻¹) (literatura)	Grupo provável	Classe dos compostos
9039-8980	9000-8000	3vC-H	Alcanos (CH ₃ /CH ₂)
8679-8620	9000-8000	3vC-H	Alcanos (CH ₃ /CH ₂)
6819-6700	7500-6150	2vO-H	Ácido carboxílico, álcoois (O-H)
5859-5800	-	-	-
5739-5560	5750-5700	2vC-H	Alcanos (CH ₃ /CH ₂)
4719-4660	-	-	-
4119-4060	-	-	-

Tabela 25 – Intervalos selecionados pelo iPLS ao conjunto de espectros Mid-IR comparados aos intervalos identificados por Silverstein e Webster (2000)

Intervalo selecionado pelo iPLS	Nº de onda (cm ⁻¹) (literatura)	Grupo provável	Classe dos compostos
3400-3341	-	-	-
2980-2861	3000-2840	vC-H	Alcanos
1960-1901	-	-	-
1420-1301	1500-1400	δC-H	Alcanos
1240-1001	1300-1000	vC-O	Ácidos carboxílicos, ésteres

O número de variáveis latentes para o modelo foi escolhido baseando-se nos valor de RMSEC. Neste estudo utilizando iPLS foi utilizado a centralização dos dados na média (MC) para o conjunto espectral de NIR e 1ª derivada para os espectros de Mid-IR, onde foram obtidas as Figuras 79 e 80, respectivamente.

Figura 73 – Gráfico do RMSEC *versus* número de variáveis latentes para o PEFF com espectros NIR (iPLS)

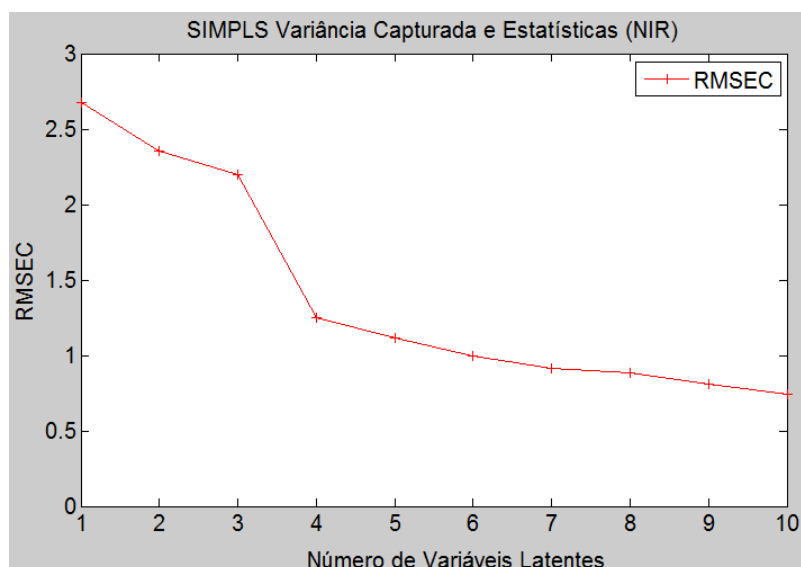
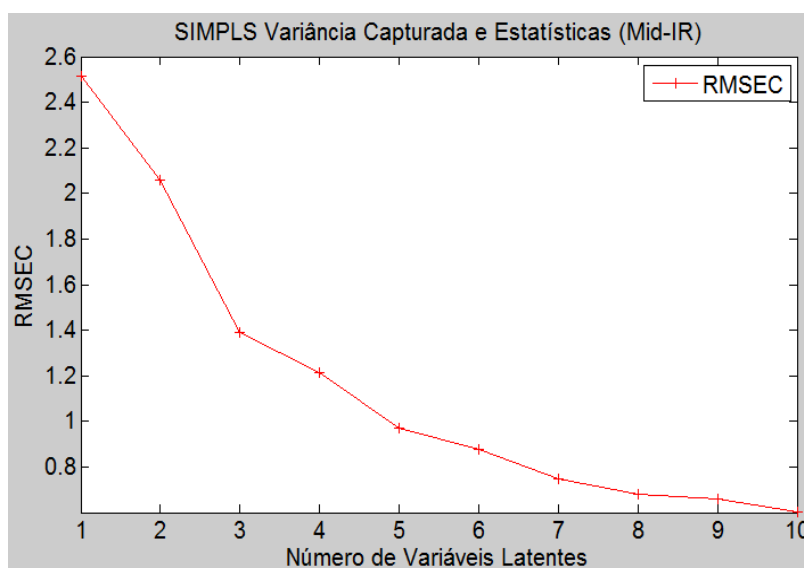


Figura 74 – Gráfico do RMSEC *versus* número de variáveis latentes para o PEFF com espectros Mid-IR (iPLS)



Quando se observam os valores de RMSEC nos gráficos das Figuras 79 e 80, observa-se que com quatro e cinco variáveis latentes é possível construir modelos que correlacionem o ponto de entupimento com os dados espectrais de NIR e Mid-IR, apresentando baixos valores de RMSEC (1,2; 1,0), RMSECV (1,4; 1,1) e, principalmente, baixo RMSEP (1,1; 0,9), respectivamente. Dentre os pré-processamentos testados, o melhor resultado foi obtido com a utilização a

centralização dos dados na média (MC) para a matriz dos espectros de NIR, 1ª derivada para os espectros de Mid-IR e autoescalamento para o vetor contendo os dados de PEFF. Nestes casos, foi possível explicar 99,75% e 99,99% da variância dos dados do Bloco X (matriz dos espectros) e 78,75% e 85,13% para o Bloco Y (vetor da propriedade), para os conjuntos de dados de NIR e Mid-IR, respectivamente.

Para o conjunto de dados de NIR, o modelo de previsão apresentado foi construído na 3ª tentativa após a retirada de dezessete amostras atípicas: amostras 1 (Am021), 2 (Am022), 4 (Am024), 6 (Am026), 9 (Am029), 11 (Am031), 13 (Am033), 17 (Am039), 23 (Am045), 28 (Am058), 32 (Am062), 34 (Am066), 35 (Am069), 37 (Am071), 30 (Am060), 36 (Am070) e 42 (Am081) do conjunto de calibração dos espectros NIR. Posteriormente, foi refeita a regressão por mínimos quadrados parciais por intervalos (iPLS) utilizando a centralização dos dados na média (MC).

Depois de construído o modelo por iPLS para o banco de espectros de NIR e Mid-IR, foi gerado o gráfico de resíduos estudentizados *versus* Hotelling T^2 , conforme mostrado nas Figuras 81 e 82. Estes gráficos mostram que todas as amostras possuem, simultaneamente, valores baixos para Hotelling T^2 e resíduos estudentizados, significando assim uma boa homogeneidade do conjunto de dados e ausência de amostras atípicas para os modelos. Sendo assim, todas as amostras desse conjunto devem permanecer no modelo.

Figura 75 – Gráfico dos resíduos estudentizados *versus* Hotelling T^2 para o PEFF com espectros NIR (iPLS)

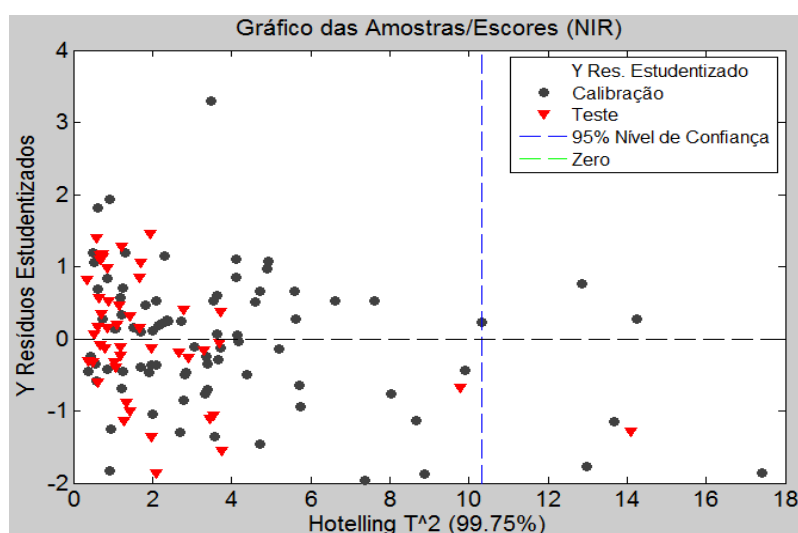
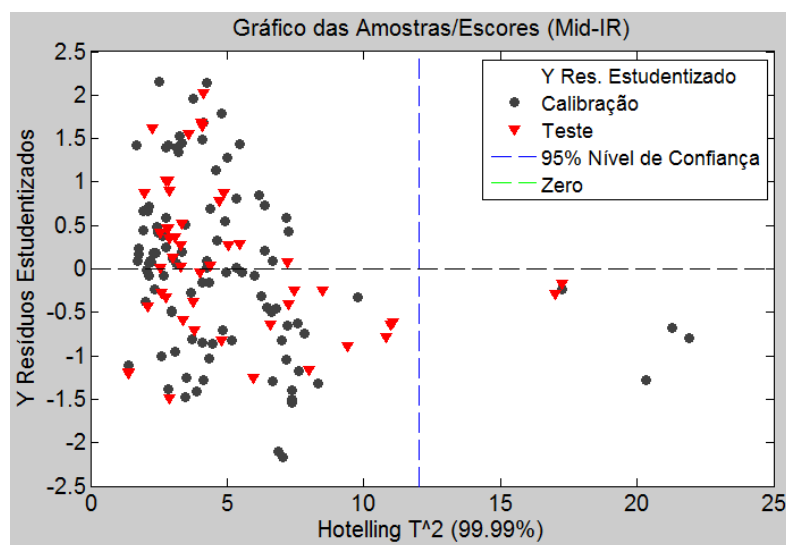


Figura 76 – Gráfico dos resíduos estudentizados *versus* Hotelling T^2 para o PEFF com espectros Mid-IR (iPLS)



Ao analisar o gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o banco de NIR, (Figura 83), percebe-se que cinco amostras (6%) do conjunto de calibração encontram-se fora do limite de confiança de 95%, o que significa uma percentagem que excede o valor aceitável de amostras fora do intervalo de confiança para o conjunto de calibração, deste modo as cinco amostras deveriam ser retiradas do modelo. No entanto, a remoção dessas amostras não é aconselhada pela norma ASTM E1655-05, que permite a retirada de amostras atípicas somente até o segundo modelo, por causa do efeito bola de neve.

O gráfico dos escores da variável latente 2 *versus* escores da variável latente 1, para o banco contendo os espectros de Mid-IR, é apresentado na Figura 84 e observa-se que cinco amostras (5,0%) do conjunto de calibração e seis amostras (12,5%) do conjunto de validação estão situadas fora do intervalo de confiança de 95%. Desta forma, se fez necessário avaliar a situação de cada uma das amostras dos conjuntos e decidir quais devem ser retiradas do modelo. Para o conjunto de calibração, verificou-se que as amostras atípicas 6 (Am026: biodiesel de soja), 11 (Am034: biodiesel de canola), 16 (Am039: biodiesel de girassol), amostra 58 (Am109: mistura de biodiesel de soja e biodiesel de canola 90:10) e amostra 90 (Am163: mistura de biodiesel de canola e biodiesel da região sul 30:70), podem ser mantidas no modelo, visto que, 5% de amostras atípicas é uma percentagem

aceitável para o limite de confiança de 95% e por este motivo essas amostras podem ser mantidas no conjunto.

No entanto, para o conjunto de validação a situação tornou-se mais complicada, visto que as amostras 138 (Am152), 139 (Am153) e 140 (Am155), referentes às misturas quaternárias de biodiesel de canola, milho, girassol e soja são amostras que compõem um conjunto de triplicatas, do mesmo modo que as amostras 141 (Am160), 142 (Am162) e 143 (Am165), referentes às misturas binárias de biodiesel de canola e da região sul. A retirada dessas amostras do conjunto pode causar perda de informações importantes ao modelo, já que esses dois conjuntos de amostras triplicatas são as únicas que carregam a informação de sua composição no conjunto de validação, além de apresentarem comportamentos similares quando comparados aos seus iguais (triplicatas). Por estes motivos, essas amostras foram mantidas no modelo.

Figura 77 – Gráfico dos escores da variável latente 2 versus escores da variável latente 1 para o PEFF com espectros NIR (iPLS)

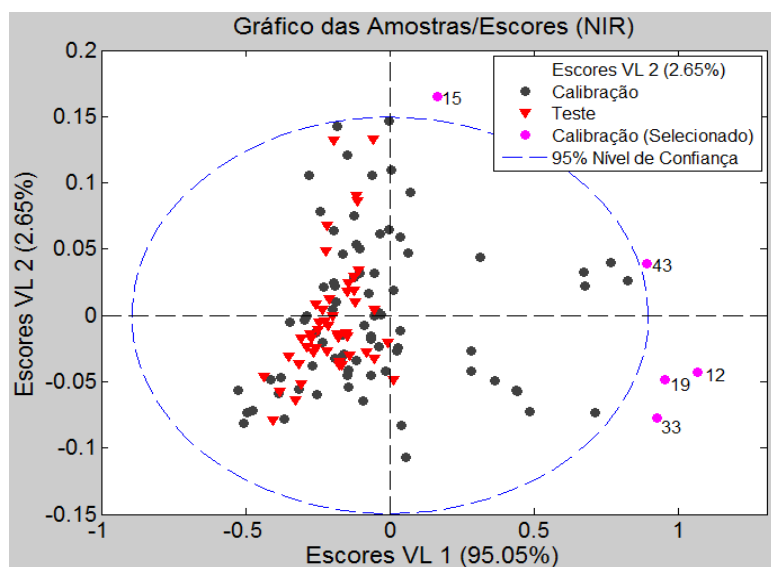
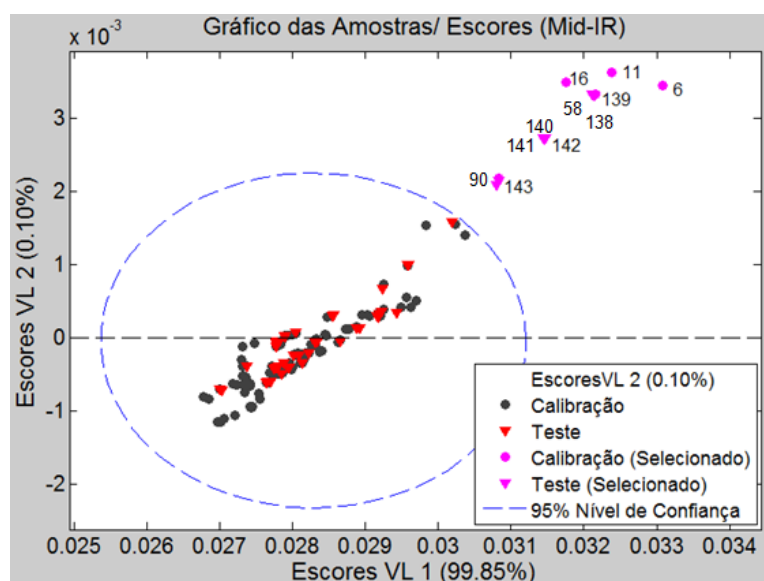


Figura 78 – Gráfico dos escores da variável latente 2 *versus* escores da variável latente 1 para o PEFF com espectros Mid-IR (iPLS)



Dos gráficos apresentados nas Figuras 85 e 86 (Y previsto *versus* Y medido), podem ser obtidas informações importantes sobre os modelos para previsão do ponto de entupimento utilizando espectros NIR e Mid-IR. O desempenho de um modelo de calibração multivariada é avaliado a partir dos valores de RMSEC, RMSECV, RMSEP e coeficiente de determinação (R^2), devendo os modelos apresentar valores de erros baixos. Os modelos de previsão para o ponto de entupimento, com dados espectrais de NIR e Mid-IR usando iPLS com quatro e cinco variáveis latentes, apresentou um bom coeficiente de determinação, $R^2_{Pred} = 0,82$ e $0,90$, e RMSEP (°C) de $1,1$ e $0,9$, respectivamente.

Figura 79 – Gráfico do Y previsto versus Y medido para o PEFf com espectros NIR (iPLS)

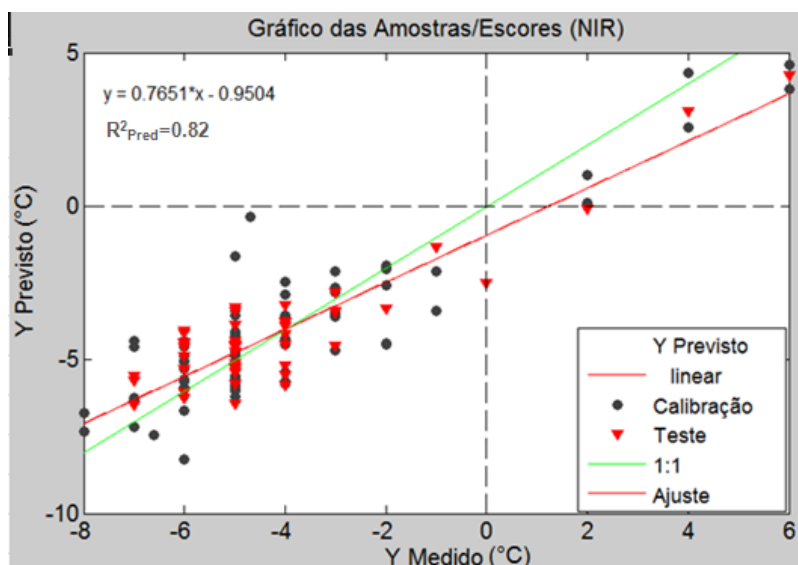
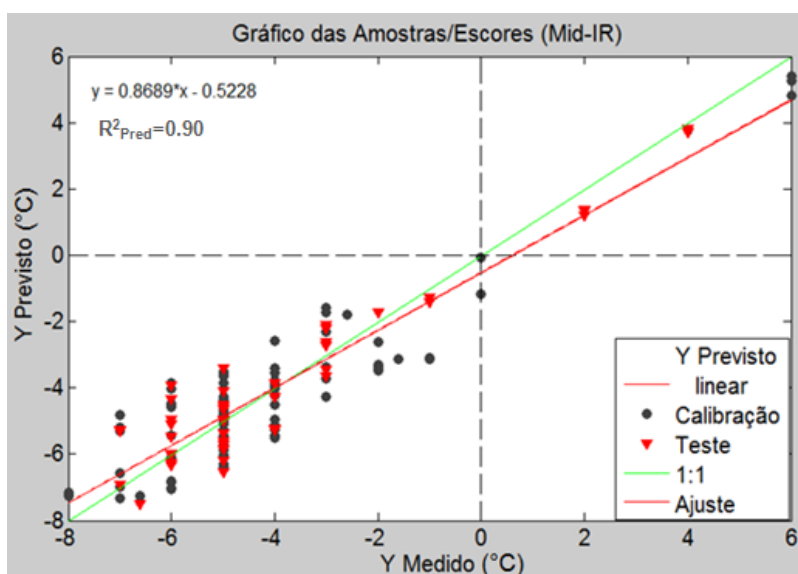


Figura 80 – Gráfico do Y previsto versus Y medido para o PEFf com espectros Mid-IR (iPLS)



3.4.3.3 Regressão por SVM com seleção de variáveis por GA

Foi usado o algoritmo genético (GA) para a seleção de variáveis dos espectros de NIR e Mid-IR, enquanto que a variação normal padronizada (SNV) foi empregada para fazer o pré-processamento na matriz dos espectros. Foram utilizadas as opções padrões do algoritmo (tamanho da população: 64; largura da janela: 50; % de termos iniciais: 30; gerações máximas: 100; % de convergência: 50;

taxa de mutação: 0,005). O GA selecionou 1401 e 600 variáveis dos espectros de NIR e Mid-IR, respectivamente, conforme pode ser observado nas Figuras 87 e 88.

Figura 81 –Número de ondas (cm^{-1}) selecionados pelo algoritmo genético (GA) para compor a matriz dos espectros NIR

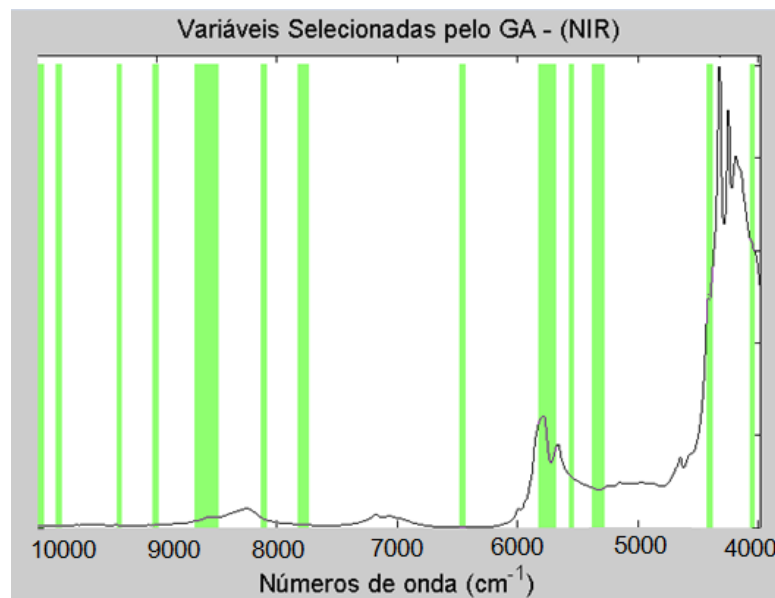
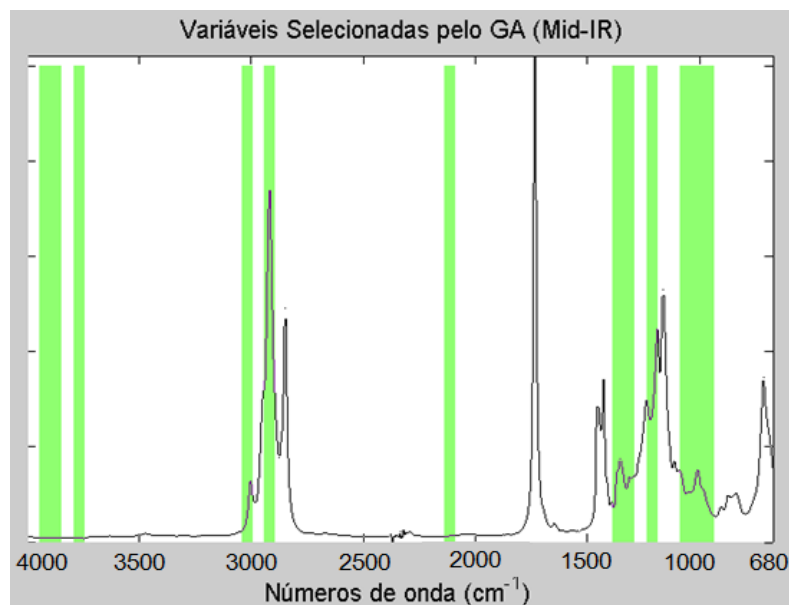


Figura 82 – Número de ondas (cm^{-1}) selecionados pelo algoritmo genético (GA) para compor a matriz dos espectros Mid-IR



Os intervalos de números de onda selecionados pelo GA para os conjuntos de dados de NIR e Mid-IR foram comparados com os intervalos identificados por Workman e Weyer (2008) e Silverstein e Webster (2000), respectivamente, conforme

apresentado nas Tabelas 26 e 27. Para os conjuntos NIR e Mid-IR, o algoritmo genético (GA) conseguiu selecionar 61% e 58% de variáveis (números de onda) com informações químicas pertinentes ao biodiesel.

Com as variáveis selecionadas pelo GA foram construídas novas matrizes para os espectros de NIR e Mid-IR, que foram utilizadas na construção de modelos por regressão SVM (SVM-R).

Tabela 26 – Intervalos selecionados pelo GA ao conjunto de espectros NIR comparados aos intervalos identificados por Workman e Weyer (2008)

Intervalo selecionado pelo GA	Nº de onda (cm ⁻¹) (literatura) *	Grupo provável	Classe dos compostos
9350-9301	-	-	-
9150-9101	-	-	-
8700-8651	9000-8000	3vC-H	Alcanos (CH ₃ /CH ₂)
8550-8451	9000-8000	3vC-H	Alcanos (CH ₃ /CH ₂)
8400-8351	9000-8000	3vC-H	Alcanos (CH ₃ /CH ₂)
8150-8101	9000-8000	3vC-H	Alcanos (CH ₃ /CH ₂)
7650-7551	-	-	-
6950-6901	7500-6150	2vO-H	Ácido carboxílico, álcoois (O-H)
6800-6551	7500-6150	2vO-H	Ácido carboxílico, álcoois (O-H)
6500-6401	7500-6150	2vO-H	Ácido carboxílico, álcoois (O-H)
6100-6051	-	-	-
5950-5901	-	-	-
5750-5601	5750-5700	2vC-H	Alcanos (CH ₃ /CH ₂)
5300-5251	5350-4900	3vC=O	Derivados carbonilados
4700-4651	-	-	-
4600-4501	-	-	-
4450-4401	-	-	-
4100-4051	-	-	-
4000	-	-	-

Tabela 27 – Intervalos selecionados pelo GA ao conjunto de espectros Mid-IR comparados aos intervalos identificados por Silverstein e Webster (2000)

Intervalo selecionado pelo GA	Nº de onda (cm ⁻¹) (literatura) *	Grupo provável	Classe dos compostos
3950-3851	-	-	-
3800-3751	-	-	-
3050-3001	-	-	-
2950-2901	3000-2840	νC-H	Alcanos
2150-2101	-	-	-
1450-1351	1500-1400	δC-H	Alcanos
1250-1200	1300-1000	νC-O	Ácidos carboxílicos, ésteres
1100-951	1300-1000	νC-O	Ácidos carboxílicos, ésteres

Para a técnica SVM-R foi utilizado o pré-processamento SNV para matriz dos espectros NIR e Mid-IR conforme mencionado anteriormente, autoescalamento para a matriz da propriedade (PEFF) e a seleção de n-fold= 10 para a validação cruzada. O vetor suporte de regressão selecionado foi o épsilon que comparado ao vetor suporte de regressão Nu apresentou os melhores resultados para a previsão do PEFF em função das variáveis (absorvância medida em função do número de onda).

A técnica SVM-R apresenta uma região de busca dos valores ótimos dos parâmetros Épsilon, Gamma e Cost. Os pontos identificados nas Figuras 89 e 90 e assinalado com um x é o ponto ótimo para os parâmetros, onde se obtém o menor erro médio quadrático (MSE). Para o banco de espectros de NIR (Figura 89), os valores de Épsilon, Gamma e Cost são 0,01; 3,16 e 10, respectivamente. Para o banco de espectros de Mid-IR (Figura 90) são 0,1; 10 e 1.

Figura 83 – Região de busca dos parâmetros de otimização (Épsilon, Gamma e Cost) para o PEFf com espectros NIR

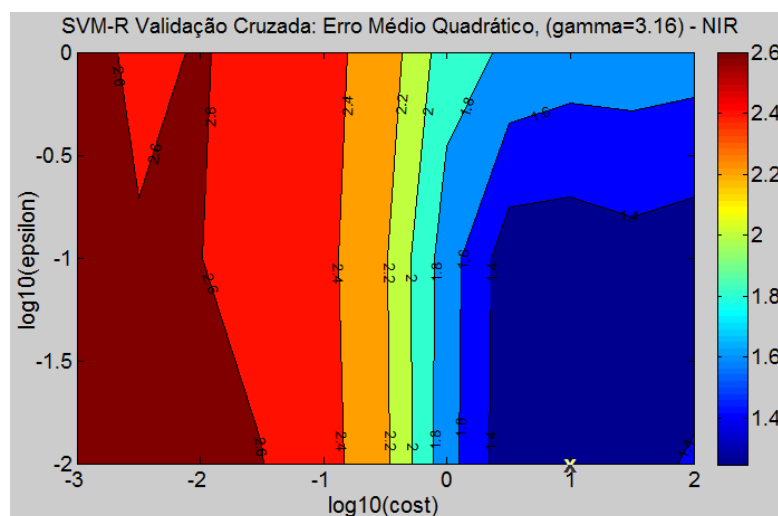
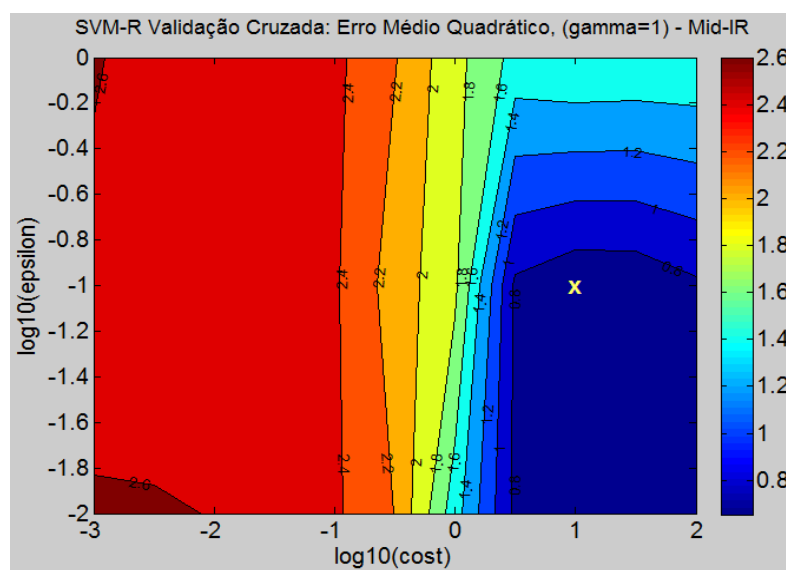


Figura 84 – Região de busca dos parâmetros de otimização (Épsilon, Gamma e Cost) para o PEFf com espectros Mid-IR

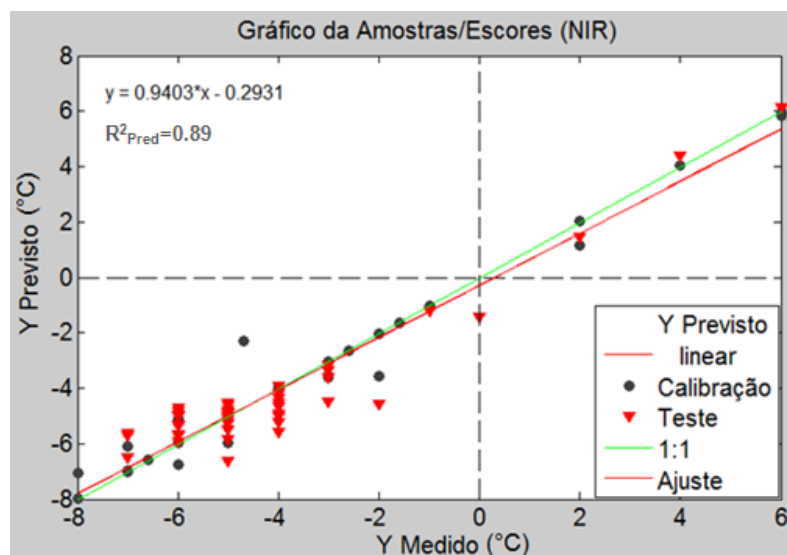


Os gráficos dos valores previstos *versus* medidos usando-se os dados dos espectros de NIR e Mid-IR são apresentados nas Figuras 91 e 92, respectivamente. A Figura 91, banco de espectros de NIR, apresenta o gráfico dos valores previstos *versus* medidos. A linha verde representa a reta ideal e a linha vermelha representa o modelo ajustado. A inclinação da reta do modelo é influenciada pela distribuição das amostras, quanto menor o ângulo entre as duas retas, mais próximos são os

valores previstos dos observados. Quanto mais próximo de 1 é o coeficiente de determinação mais próximos estão os valores previstos dos medidos.

O gráfico da Figura 91, para o banco de espectros de NIR, mostra que o distanciamento entre os valores observados (medidos) e os previstos cresce de acordo com os valores mais altos. Para valores baixos a previsão é melhor, visto que há uma maior proximidade entre os valores observados e previstos. Isto significa que embora o modelo apresente uma boa aproximação, o mesmo tende a apresentar melhores previsões para valores mais baixos de PEFF. No entanto, ao analisar o gráfico da Figura 92 para o banco de espectros Mid-IR, os valores observados e previstos estão quase sobrepostos o que evidencia uma excelente aproximação do modelo obtido com os dados espectrais.

Figura 85 – Gráfico dos valores previstos versus medidos para o PEFF (°C) com espectros NIR



Além de utilizar os espectros completos de NIR e Mid-IR na seleção de variáveis pelo GA, foi testado em separado a utilização do GA para cada uma das cinco regiões espectrais que contém informações químicas referentes ao biodiesel, conforme identificado por Workman e Weyer (2008) para os espectros de NIR e Silverstein e Webster (2000) para os espectros de Mid-IR, conforme apresentado na Tabela 28. Esta divisão visou reduzir o número de variáveis selecionadas pelo algoritmo genético, simplificar o modelo e reduzir o tempo computacional.

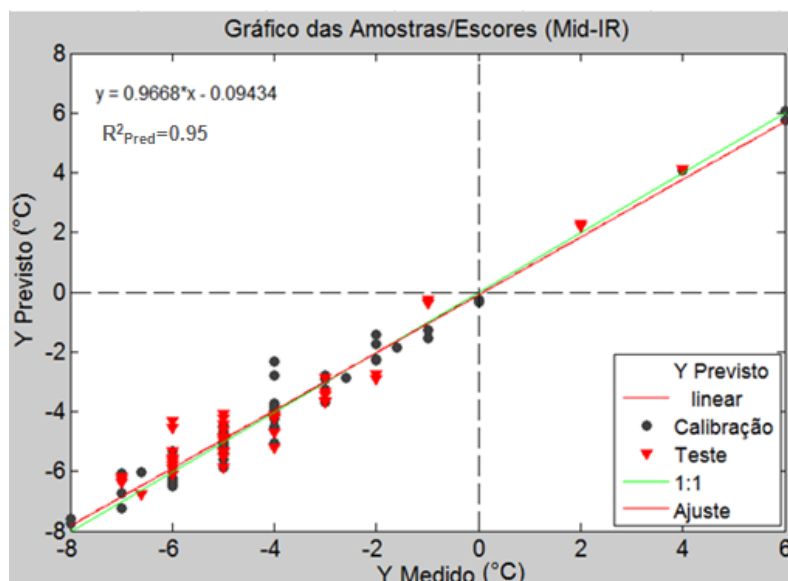
Figura 86 – Gráfico dos valores previstos *versus* medidos para o PEFF (°C) com espectros Mid-IR

Tabela 28 – Divisão das regiões dos espectros NIR e Mid-IR para aplicação do algoritmo genético (GA)

	NIR	Mid-IR
Região	Nº de onda (cm ⁻¹)	Nº de onda (cm ⁻¹)
1	9000-8000	3000-2840
2	7500-6150	1750-1730
3	5750-5700	1500-1400
4	5350 e 4900	1300-1000
5	4370 e 4260	750-700

As Tabelas 29 e 30 mostram, respectivamente, os resultados dos parâmetros dos modelos dos bancos de dados de espectros de NIR e Mid-IR, os valores de RMSEC, RMSECV, RMSEP, os coeficientes de determinação da calibração, validação cruzada e previsão (respectivamente R^2_{Cal} , R^2_{CV} e R^2_{Pred}), para os cinco intervalos e o espectro todo. Para o banco de NIR, o método GA-SVM obteve mais êxito ao modelar a previsão de ponto de entupimento com o espectro completo quando comparado aos modelos obtidos para cada região espectral, pois os erros são menores e os coeficientes de determinação são maiores (Tabela 29).

Como pode ser observado na Tabela 30, para os espectros de Mid-IR, o método GA-SVM obteve resultados melhores e similares ao modelar a previsão de

ponto de entupimento utilizando a 4ª região espectral, 1ª região ou o espectro completo. Estes modelos foram superiores quando comparado aos modelos obtidos da 2ª, 3ª e 5ª região espectral, pois os erros são menores e os coeficientes de determinação são maiores.

Tabela 29 – Parâmetros dos modelos obtidos por SVM-R para cada uma das cinco regiões e para o espectro completo de NIR

	NIR					
	1ª Região	2ª Região	3ª Região	4ª Região	5ª Região	Espectro completo
RMSEC	0,1	0,3	1,3	1,1	0,9	0,4
RMSECV	1,8	2,3	1,5	1,9	1,2	1,3
RMSEP	1,7	2,0	1,1	1,8	1,0	0,8
R^2_{cal}	0,99	0,99	0,75	0,82	0,87	0,98
R^2_{cv}	0,57	0,21	0,66	0,44	0,77	0,75
R^2_{Pred}	0,73	0,52	0,84	0,60	0,85	0,89

Tabela 30 – Parâmetros dos modelos obtidos por SVM-R para cada uma das cinco regiões e para o espectro completo de Mid-IR

	Mid-IR					
	1ª Região	2ª Região	3ª Região	4ª Região	5ª Região	Espectro completo
RMSEC	0,5	2,3	0,7	0,5	0,68	0,4
RMSECV	0,8	2,4	0,9	0,8	0,87	0,7
RMSEP	0,6	2,5	0,9	0,6	0,9	0,6
R^2_{cal}	0,95	0,20	0,92	0,96	0,93	0,97
R^2_{cv}	0,91	0,09	0,87	0,91	0,88	0,93
R^2_{Pred}	0,95	0,30	0,90	0,96	0,91	0,95

O quadro comparativo da Tabela 31 mostra os resultados dos parâmetros de mérito dos bancos de dados de espectros de NIR e Mid-IR, os valores de RMSEC, RMSECV, RMSEP, os coeficientes de determinação da calibração, validação

cruzada e previsão (respectivamente R^2_{Cal} , R^2_{CV} e R^2_{Pred}) utilizando os métodos lineares (PLS e iPLS) e não-lineares (SVM-R) na previsão do PEFF.

Aparentemente, para os dois bancos de dados (NIR e Mid-IR) o método GA-SVM obteve melhores modelos de previsão de ponto de entupimento quando comparado aos demais modelos, pois os erros são menores e os coeficientes de determinação são maiores. No entanto, a capacidade de previsão dos modelos deve ser avaliada utilizando o teste F.

A capacidade preditiva dos modelos foi avaliada através do RMSEP, obtido para o conjunto de validação externa. Um teste F, em um nível de confiança de 95%, foi utilizado para avaliar se existem diferenças estatisticamente significativas entre os valores de RMSEP dos modelos PLS, iPLS e SVM, sendo este último com seleção de variáveis por GA. Os valores de F calculado ($F_{\text{calc.}}$) foram obtidos calculando-se a razão entre os valores quadráticos do maior e menor RMSEP obtidos pelos modelos. Esses valores de $F_{\text{calc.}}$ foram comparados com os valores de F crítico ($F_{\text{crít.}, 0,05, 1, 2}$) tabelado, em função do grau de liberdade do numerador (1) e denominador (2) da Equação 36, anteriormente apresentada.

Na Tabela 32 são apresentados valores de F críticos e F calculados para o teste F (com 95% de confiança) obtidos na comparação dos diferentes modelos de previsão do PEFF utilizando espectros de NIR e Mid-IR. O valor em vermelho está representando quando o valor de F calculado é maior que o valor de F crítico.

Ao comparar os modelos, quando $F_{\text{calc.}} > F_{\text{crít.}, 0,05, 1, 2}$, pode-se afirmar que estes modelos são diferentes. Isto significa que o modelo com menor RMSEP é melhor do que o com maior RMSEP e apresenta diferença estatisticamente significativa comprovada entre os modelos.

Os modelos de previsão do PEFF utilizando o método GA-SVM e a técnica espectrométrica Mid-IR, ou simplesmente GA-SVM (Mid-IR), tanto para o espectro completo quanto para 1ª e 4ª região, apresentam desempenho superior quando comparados aos outros sete modelos. Estes modelos são estatisticamente iguais entre si e diferentes dos demais modelos propostos neste estudo. Uma vantagem dos modelos construídos a partir dos dados espectrais de Mid-IR da 1ª e 4ª região é poder trabalhar com um conjunto de dados reduzido.

Dentre os cinco modelos construídos com os espectros de NIR, o que se destaca com o melhor resultado, ou seja, com menor RMSEP, é o que utiliza o

método SVM, seguido pelos modelos PLS e SVM (5ª Região). Os três modelos são bastante adequados na previsão do ponto de entupimento (PEFF). No entanto, estes modelos não superam nenhum dos modelos que utilizam os espectros Mid-IR.

Para ambos os espectros, os melhores resultados foram obtidos com o método não linear utilizado neste estudo, a SVM. Este comportamento já era esperado devido ao comportamento não linear do PEFF.

Tabela 31 – Parâmetros do modelo de PLS, iPLS e SVM obtido dos bancos de dados de NIR e Mid-IR para o PEFf

Parâmetros do Modelo	NIR					Mid-IR				
	PLS	iPLS	SVM	SVM 3ª Região	SVM 5ª Região	PLS	iPLS	SVM	SVM 1ª Região	SVM 4ª Região
RMSEC	1,1	1,2	0,4	1,3	0,9	1,3	1,0	0,4	0,5	0,5
RMSECV	1,2	1,4	1,3	1,5	1,2	1,4	1,1	0,7	0,8	0,8
RMSEP	1,0	1,1	0,8	1,1	1,0	0,9	0,9	0,6	0,6	0,6
R^2_{Cal}	0,85	0,79	0,98	0,75	0,87	0,77	0,85	0,97	0,95	0,96
R^2_{CV}	0,80	0,73	0,75	0,66	0,77	0,72	0,81	0,93	0,91	0,91
R^2_{Pred}	0,86	0,82	0,89	0,84	0,85	0,89	0,90	0,95	0,95	0,96

Tabela 32 – Resultados para o teste F (nível de confiança de 95%) obtidos na comparação dos diferentes modelos de previsão do PEFf utilizando espectros de NIR e Mid-IR

Modelos		NIR					Mid-IR					
		PLS	iPLS	SVM	SVM 3ª Região	SVM 5ª Região	PLS	iPLS	SVM	SVM 1ª Região	SVM 4ª Região	
	$F_{crit., 0,05, \square 1,}$	1,64	1,64	1,64	1,64	1,64	1,64	1,64	1,64	1,64	1,64	1,64
PLS (NIR)	$F_{calc.}$	-	1,21	1,56	1,21	1	1,23	1,23	2,78	2,78	2,78	2,78
iPLS (NIR)	$F_{calc.}$	1,21	-	1,89	1	1,21	1,49	1,49	3,36	3,36	3,36	3,36
SVM (NIR)	$F_{calc.}$	1,56	1,89	-	1,89	1,56	1,27	1,27	1,78	1,78	1,78	1,78
SVM (NIR) - 3ª Região	$F_{calc.}$	1,21	1	1,89	-	1,21	1,49	1,49	3,36	3,36	3,36	3,36
SVM (NIR) - 5ª Região	$F_{calc.}$	1	1,21	1,56	1,21	-	1,23	1,23	2,78	2,78	2,78	2,78
PLS (Mid-IR)	$F_{calc.}$	1,23	1,49	1,27	1,49	1,23	-	1	2,25	2,25	2,25	2,25
iPLS (Mid-IR)	$F_{calc.}$	1,23	1,49	1,27	1,49	1,23	1	-	2,25	2,25	2,25	2,25
SVM (Mid-IR)	$F_{calc.}$	2,78	3,36	1,78	3,36	2,78	2,25	2,25	-	1	1	1
SVM (Mid-IR) - 1ª Região	$F_{calc.}$	2,78	3,36	1,78	3,36	2,78	2,25	2,25	1	-	1	1
SVM (Mid-IR) - 4ª Região	$F_{calc.}$	2,78	3,36	1,78	3,36	2,78	2,25	2,25	1	1	-	-

3.5 Recuperação dos Resíduos

Dos resíduos gerados neste estudo estão a glicerina, álcool metílico, álcool etílico, sulfato de sódio e cloreto de potássio. O resíduo de sulfato de sódio foi totalmente recuperado com um rendimento médio de 80%, os resíduos de cloreto de potássio, glicerina e álcool metílico estão sendo recuperados conforme a disponibilidade e recursos do laboratório. No entanto, o resíduo de álcool etílico que vem sendo utilizado em larga escala na limpeza dos equipamentos e vidrarias tem sido recuperado continuamente.

4 CONCLUSÕES E PROPOSTAS FUTURAS

4.1 Conclusões

Neste trabalho foi apresentada a aplicação das técnicas de estatística multivariada para a análise exploratória e previsão do índice de refração, massa específica e ponto de entupimento de filtro a frio, a partir de espectros de infravermelho próximo e médio de misturas de diferentes tipos de biodiesel (soja, canola, milho, girassol e da região sul). Com a utilização da técnica PCA foi realizada uma análise exploratória dos dados, onde foi possível visualizar a distribuição das amostras, o comportamento e principalmente, fazer a identificação de amostras atípicas para os dois bancos (NIR e Mid-IR). O modelo apresentado pela ferramenta PCA mostrou-se consistente e permitiu a remoção de amostras que apresentavam comportamentos atípicos.

A fim de obter um modelo de previsão adequado para as propriedades utilizando os dois bancos de dados espectrais em questão, foram utilizados os métodos PLS, iPLS e GA-SVM para prever o ponto de entupimento de filtro a frio (PEFF) e PLS para prever índice de refração e massa específica de misturas de biodiesel. Pode se concluir que para a previsão do índice de refração a regressão por PLS obteve tanto êxito no modelo previsto para o banco de espectros NIR quanto para os de Mid-IR, pois os resultados encontrados são muito similares entre si, com pequenas diferenças apresentadas entre os valores dos parâmetros escolhidos para a comparação. No entanto, ao comparar os modelos utilizando o teste F foi possível identificar que os espectros de Mid-IR proporcionaram a construção de um modelo mais adequado, ou seja, com baixo valor de erro e maior coeficiente de determinação.

Para a previsão da massa específica a regressão por PLS utilizando o banco de espectros de Mid-IR apresentou resultados melhores que os encontrados com o banco de espectros NIR, sendo comprovado pelo teste F que o modelo gerado por PLS utilizando o banco de espectros de Mid-IR é o mais adequado.

Dentre os modelos de previsão do PEFF construídos a partir do banco de espectros de NIR, o modelo GA-SVM para o espectro completo pode ser citado como o modelo que apresentou os melhores resultados quando comparados aos

modelos PLS, iPLS e SVM para as 3^a e 5^a regiões visto que para o modelo obtido por GA-SVM para o espectro completo observa-se menor erro de previsão e maior coeficiente de determinação da previsão. Embora o modelo PLS não tenha obtido o melhor resultado, ele foi considerado, pelo teste F, adequado para a previsão do PEFF. No entanto, para os modelos construídos utilizando o banco de espectros de Mid-IR, os modelos GA-SVM destacam-se como os modelos que apresentaram os melhores resultados se comparados aos modelos que utilizaram regressão por PLS e iPLS, visto que tanto o modelo obtido por GA-SVM para o espectro completo quanto os modelos para as 1^a e 4^a regiões apresentaram menores erros de previsão e maiores coeficientes de determinação da previsão. De um modo geral, os modelos de previsão do PEFF que utilizaram o banco de espectros de Mid-IR foram considerados mais adequados que os modelos que utilizaram espectros de NIR, uma prova desta afirmação, é que os melhores resultados foram obtidos com os modelos GA-SVM (Mid-IR), o que confirma também que o método não linear é mais eficiente para propriedades que apresentam comportamento não linear como, por exemplo, o ponto de entupimento.

Analisando todos os resultados obtidos, verificou-se que a metodologia desenvolvida é eficaz para determinar as propriedades físico-químicas estudadas de forma rápida e não destrutível, sendo possível utilizá-la para controle da qualidade do biodiesel em ambiente industrial.

4.2 Propostas Futuras

Pretende-se, como continuidade desse trabalho:

- Ampliar o modelo desenvolvido para a determinação de outras propriedades do biodiesel, tais como: índice de iodo, índice de acidez, índice de saponificação, teor de água, ponto de fulgor, teor de ésteres, corrosividade ao cobre, viscosidade, condutividade e teor de cinzas, bem como acrescentar outros tipos de biodiesel (palma e sebo bovino) e misturas com diesel.
- Realizar a técnica de fusão dos dados espectrais NIR e Mid-IR para realizar a previsão das propriedades, bem como acrescentar outras técnicas (ANN-PLS,

SPA-MLR e UVE-PLS) que permitam obter modelos adequados para as propriedades estudadas.

REFERÊNCIAS

AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCMBUSTÍVEIS (ANP). Boletim mensal de biodiesel. Disponível em: <http://www.ubrabilio.com.br/1891/Noticias/BoletimMensalDoBiodieselAnpAgosto2013_194935/>. Acesso em: 27 de set. 2013.

AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCMBUSTÍVEIS (ANP). Decreto nº 2.455, de 14 de janeiro de 1998. Disponível em: <http://www.planalto.gov.br/ccivil_03/decreto/D2455.htm/>. Acesso em: 17 de nov. 2013.

AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCMBUSTÍVEIS (ANP). Resolução ANP nº. 14 de maio de 2012. Disponível em: <<http://www.anp.gov.br/>>. Acesso em: 21 de maio 2012.

ALANDER, J.; ANDERSON, A. C.; BAGGE, C.; BRINGSARVE, K.; HJORTH, M.; JOHANSSON, M.; GRANROTH, B.; NORBERG, S.; PEDERSEN, M.; PERSSON, M.; WENNERMARK, B.; WENNERMARK, M. *Handbook of Vegetable Oils and Fats*. 2. ed. Aarhus Karlshamn AB. Sweden, 2007.

ALEME, H. G. *Utilização de curvas de destilação e métodos quimiométricos na previsão da origem e na determinação do teor alcoólico e massa específica da gasolina automotiva*. 2008. 81 f. Dissertação (Mestrado em Química) - Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2008.

_____. *Determinação de parâmetros físico-químicos do óleo diesel a partir de curvas de destilação utilizando técnicas quimiométricas*. 2011. 149 f. Tese (Doutorado em Ciências) - Instituto de Química - Universidade Federal de Minas Gerais, Belo Horizonte, 2011.

_____.; BARBEIRA, P. J. S. *Comparação entre métodos quimiométricos de calibração para a determinação do teor de biodiesel no óleo diesel usando curvas de destilação*. In: Encontro Nacional de Química Analítica, 16., Campos do Jordão, São Paulo, 2011.

ALEME, H. G.; CORGOZINHO, C. N. C.; BARBEIRA, P. J. S. Diesel oil discrimination by origin and type using physicochemical properties and multivariate analysis. *Fuel*, v. 89, p. 3151–3156, maio 2010.

AL-GHOUTI, M. A.; AL-DEGSB, Y. S.; AMERA, M. Determination of motor gasoline adulteration using FTIR spectroscopy and multivariate calibration. *Talanta*, v. 76, n. 5, p. 1105–1112, set. 2008.

ALPTEKIN, E.; CANAKCI, M. Characterization of the key fuel properties of methyl ester–diesel fuel blends. *Fuel*, v. 88, p. 75–80, jan. 2009.

ALVES, C. T.; SANTOS, D. C.; TORRES, E. A. *Produção do biodiesel de mamona em escala semi-industrial na planta piloto da Universidade Federal da Bahia*. In: CONGRESSO BRASILEIRO DA MAMONA, 2., 2006, Aracaju, Sergipe, p. 1-6, 2006.

ARDREY, R. E. *Liquid chromatography-Mass Spectrometry: an introduction*. 1. ed. Huddersfield: John Wiley & Sons, p. 276, 2003.

ASTM International, Annual Book of ASTM standards. Standards practices for infrared, multivariate, quantitative analysis, E1655 ASTM International, West Conshohocken, 2005.

BAIANU, I. C.; YOU, T.; COSTESCU, D. M. Novel Techniques for Microspectroscopy and Chemical Imaging Analysis of Soybean Seeds and Embryos. *T. Soy2002 Conference Urbana*, Illinois, 2002.

BAIANU I. C.; COSTESCU D. M.; HOFMANN, N.; KORBAN, S. S. Near Infrared Microspectroscopy, Chemical Imaging and NMR Analysis of Oil in Developing and Mutagenized Soybean Embryos in Culture. Champaign: AOCS Meeting, Analytical Division, 2003.

BALLABIO, D.; CONSONNI, V. Classification tools in chemistry. Part 1: linear models. *PLS_DA. Anal. Methods*, v. 5, n. 2, p. 3790-3798, jun. 2013.

BAPTISTA P.; FELIZARDO, P.; MENEZES, J. C.; CORREIA, M. J. N. Multivariate near infrared spectroscopy models for predicting the methyl esters content in biodiesel. *Analytica Chimica Acta*, v. 607, p. 153–159, dez. 2008(a).

_____. Multivariate near infrared spectroscopy models for predicting the iodine value, CFPP, kinematic viscosity at 40 °C and density at 15 °C of biodiesel. *Talanta*, v. 77, p. 144–151, jun. 2008(b).

BARTHUS, R. C.; MAZO, L. H.; POPPI, R. J. Determinação simultânea de NADH e ácido ascórbico usando voltametria de onda quadrada com eletrodo de carbono vítreo e calibração multivariada. *Ecl. Quím.*, São Paulo, v. 30, n. 4, p.: 51-58, out. 2005.

BASIC CHEMOMETRIC TECHNIQUES IN ATOMIC SPECTROSCOPY, editado por Andrade-Garda, J. M., Cambridge: *RSC Publishing*, p. 215, 2009.

BISHOP, M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006.

BORIN, A. *Aplicações de máquinas de vetores de suporte por mínimos quadrados (LS-SVM) na quantificação de parâmetros de qualidade de matrizes lácteas*. 2007. 122 f. Tese (Doutorado em Química), Universidade Estadual de Campinas. Campinas, São Paulo, 2007.

_____.; POPPI, R. J. Application of mid infrared spectroscopy and iPLS for the quantification of contaminants in lubricating oil. *Vib. Spectrosc.*, v. 37, n. 1, p. 27–32, jan. 2005.

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

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

BRITO, J. Q. A. *Obtenção de biodiesel via rota etílica auxiliado por ondas ultrassônicas*. 2011. 67 f. Dissertação (Mestrado em Química) - Universidade Federal da Bahia, Instituto de Química, Salvador, BA, 2011.

CANDEIA, R. A. *Biodiesel de Soja: síntese, degradação e misturas binárias*. 2008. 132 f. Tese (Doutorado em Química) – Universidade Federal da Paraíba, João Pessoa, PB, 2008.

CARNEIRO, H. S. P.; MEDEIROS, A. R. B.; OLIVEIRA, F. C. C.; AGUIAR, G. H. M.; RUBIM, J. C.; SUAREZ, P. A. Z. Determination of Ethanol Fuel Adulteration by Methanol Using Partial Least-Squares Models Based on Fourier Transform Techniques. *Energy & Fuels*, v. 22, n. 4, p. 2767–2770, maio 2008.

CARRARETTO, C.; MACOR, A.; MIRANDOLA, A.; STOPPATO, A.; TONON, S. Biodiesel as alternative fuel: Experimental analysis and energetic evaluations. *Energy*, v. 29, p. 2195-2211, out. 2004.

CARTWRIGHT, H. M. *Applications of Artificial Intelligence in Chemistry*. New York: Oxford Science Publications, 1995.

CETESB. Companhia de Tecnologia de Saneamento Ambiental. Riscos ao fogo. Disponível em: <http://www.cetesb.sp.gov.br/emergencia/produtos/RISCO_HELP.htm>. Acesso em: 20 de ago. 2012.

CHRISTMANN, A.; STEINWART, I. Consistency of kernel based quantile regression. *Appl. Stoch. Models Bus. Ind.*, v. 24, n. 2, p. 171-183, jan. 2008.

CHRISTOFF, P. *Produção de biodiesel a partir do óleo residual de fritura comercial. Estudo de caso: guaratuba, litoral paranaense*. 2006. 82 f. Dissertação (Mestrado em Tecnologias energéticas) - Instituto de Tecnologia para o Desenvolvimento - LACTEC, Curitiba, PR, 2006.

COATES, J. A review of current new technology: Used in instrumentation for Industrial vibrational spectroscopy. *Spectroscopy*, v. 14, n. 10, p. 21-34, jun. 1999.

CONCEIÇÃO, M. M.; CANDEIA, R. A.; SILVA, F. C.; BEZERRA, A. F.; FERNANDES Jr, V. J.; SOUZA, A. G. Thermoanalytical characterization of castor oil biodiesel. *Renew. Sustain. Energy Rev.*, v. 11, n. 5, p. 964-975, jun. 2007.

CORDEIRO, C. S. *Síntese e caracterização de biodiesel etílico para utilização em motores ciclo-diesel*. 2003. 104 f. Dissertação (Mestrado em engenharia de materiais) - Universidade Tecnológica Federal do Paraná, Brasil, 2003.

COSTA NETO, P. R.; ROSSI, L. F. S.; ZAGONEL, G. F.; RAMOS, L. P. Produção de biocombustível alternativo ao óleo diesel através da transesterificação de óleo de soja usado em frituras. *Química Nova*, v. 23, n. 4, p. 531- 537, jul. 2000.

COULTHUP, D. W. Introduction to Infrared and Raman Spectroscopy. 2. ed. New York: Academic Press, 1975.

CUNHA, M. E. *Caracterização de biodiesel produzido com misturas binárias de sebo bovino, óleo de frango e óleo de soja*. 2008. 72 f. Dissertação (Mestrado em Química) - Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2008.

DANTAS, M. B.; *Obtenção, caracterização e estudo termoanalítico de biodiesel de milho*. 2006. 115 f. Dissertação (Mestrado em Química) – Universidade Federal da Paraíba, João Pessoa, PB, 2006.

DE 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 I. *Química Nova*, v. 35, n. 1, p. 223-229, jul. 2012.

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

DOMINGOS, E.; PEREIRA, T. M. C.; FILGUEIRAS, P. R.; BUENO, M. I. M. S.; CASTRO, E. V. R.; GUIMARÃES, R. C. L.; SENA, G. L.; ROCHA, W. F. C.; ROMÃO, W. Monitoring the polyamide 11 degradation by thermal properties and X-ray fluorescence spectrometry allied to chemometric methods. *X Ray Spectrom.* v. 42, p. 79 - 86, mar. 2013.

EFRON, B.; TIBSHIRANI, R. Improvements on cross-validation: the 0.632 + bootstrap method, *J. Amer. Statist. Assoc.*, v. 92, n. 438, p. 548–560, jun. 1997.

EIDE, I.; ZAHLESEN, K. Chemical Fingerprinting of Biodiesel Using Electrospray Mass Spectrometry and Chemometrics: Characterization, Discrimination, Identification, and Quantification in Petrodiesel. *Energy & Fuels*, v. 21, p. 3702–3708, jul. 2007.

ENCINAR, J. M.; GONZÁLES, J. F.; RODRÍGUEZ, R. A. Ethanolysis of used frying oil. Biodiesel preparation and characterization. *Fuel Processing Technology*, v. 88, n. 5, p. 513, maio 2007.

EUROPEAN COMMITTEE FOR STANDARDIZATION. Standard EN 14214. Automotive fuels – fatty acid methyl esters (FAME) for diesel engines – requirements and test methods, 2009.

FANGRUI, M.; HANNA, M. A. Biodiesel production: a review. *Bioresource Technology*, v. 70, n. 1, p. 1-15, out. 1999.

FERNANDES, D. D. S.; COSTA, G. B.; GOMES, A. A.; BRITO, A. L. B.; ALMEIDA, V. E.; VERÁS, G. Classificação de misturas de biodiesel/diesel com respeito ao tipo de óleo de origem do biodiesel. In: ENCONTRO NACIONAL DE QUÍMICA

ANALÍTICA, 16., Campos do Jordão, São Paulo, 2011. *Resumos...* Campos do Jordão (SP), 2011.

FERRÃO, M. F.; GODOY, S. C.; GERBASE, A. E.; MELLO, C.; FURTADO, J. C.; PETZHOLD, C. L. Nondestructive method for determination of hydroxyl value of soybean polyol by LS-SVM using HATR/FT-IR. *Anal Chim Acta*, v. 595, n. 9, p.114–119, jul. 2007.

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

FERRER-RIQUELME, A. J. Statistical control of measures and process. Separata de: BROWN S. D.; TAULER, R.; WALCZAK, B. *Comprehensive chemometrics: chemical and biochemical data analysis*. Oxford: Elsevier Science, 2009, p. 97–126.

FREIRE, L. M. S. *Otimização da síntese etanólica do biodiesel de pinhão manso (Jatropha curcas L.) e determinação de seus parâmetros fluidodinâmicos*. 2009. 80 f. Dissertação (Mestrado em Química analítica) - Universidade Federal da Paraíba, João Pessoa, PB, 2009.

GAMA, P. E.; GIL, R. A. S. S.; LACHTER, E. R. Produção de biodiesel através de transesterificação in situ de sementes de girassol via catálise homogênea e heterogênea. *Química Nova*, v. 33, n. 9, p. 1859-1862, set. 2010.

GARCIA, C. M. *Transesterificação de óleos vegetais*. 2006. 120 f. Dissertação (Mestrado em Química) - Universidade Estadual de Campinas, Instituto de Química, Campinas, SP, 2006.

GAYDOU, V.; KISTER, J.; DUPUY, N. Evaluation of multiblock NIR/MIR PLS predictive models to detect adulteration of diesel/biodiesel blends by vegetal oil. *Chemometrics and Intelligent Laboratory Systems*, v.106, v.2, n. 15, p. 190 – 197, abril 2010.

GERIS, R.; SANTOS, N. A. C.; AMARAL, B. A.; MAIA, I. S.; CASTRO, V. D.; CARVALHO, J. R. M. Biodiesel de soja – Reação de transesterificação para aulas práticas de química orgânica. *Química Nova*, v. 30, n. 5, p. 1369, set. 2007.

GIAKOUMIS, E. G. A statistical investigation of biodiesel physical and chemical properties, and their correlation with the degree of unsaturation. *Renewable Energy*, v. 50, p. 858-878, fev. 2013.

HAYKIN, S. *Neural Networks: A Comprehensive Foundation*. 2. ed., New Jersey: Prentice-Hall, 1999.

HELPER, G. A.; FERRÃO, M. F.; FERREIRA, C. V.; HERMES, N. Aplicação de métodos de análise multivariada no controle qualitativo de essências alimentícias empregando espectroscopia no infravermelho médio. *Ciênc. Tecnol. Aliment.*, Campinas, v. 26, n. 4, p. 779-786, out. 2006.

HOEKMAN, S. K.; BROCH, A.; ROBBINS, C.; CENICEROS, E.; NATARAJAN, M. Review of biodiesel composition, properties, and specifications. *Renewable and Sustainable Energy Reviews*, v.16, n. 1, p.143– 169, jan. 2012.

HONORATO, F. A.; BARROS NETO, B.; MARTINS, M. N.; GALVÃO, R. K. H.; PIMENTEL, M. F. Transferência de calibração em métodos multivariados. *Quim. Nova*, v. 30, n. 5, set. 2007.

INNOCENTINI, M. D. M. O Processo de produção industrial do biodiesel. In: JORNADA CIENTIFICA DA UFSCAR, 7., 2007, São Carlos (SP). Estudo desenvolvido pelo LADEBIO da Universidade de Ribeirão Preto – UNEARP. São Carlos: UFSCAR, 2007.

KENNARD, R. W.; STONE, L. A.; Computer aided design of experiments. *Technometrics*, v. 11, p. 137-148, fev. 1969.

KNOTHE, G.; STEIDLEY, K. R. Kinematic viscosity of biodiesel fuel components and related compounds. Influence of compound structure and comparison to petrodiesel fuel components. *Fuel*, v. 84, n. 9, p. 1059-1065, jun. 2005.

KNOTHE, G.; GERPEN, J. V.; KRAHL, J.; RAMOS, L. P. *Manual do biodiesel*. Traduzido do original "*The Biodiesel Handbook*" por Luiz Pereira Ramos. São Paulo: Editora Edgard Blucher, 1. ed., 2006. p. 29-42.

KOURTI, T. Process analysis and abnormal situation detection: from theory to practice. *IEEE control Systems* v. 22, n. 5, p. 10-25, out. 2002.

LACERDA, E. G. M.; CARVALHO, A. C. P. L. Introdução aos algoritmos genéticos. Separata de: GALVÃO, C. O.; VALENÇA, M. J. S. (Ed.). *Sistemas inteligentes: aplicações a recursos hídricos e ciências ambientais*. Porto Alegre: Universidade/UFRGS: Associação Brasileira de Recursos Hídricos, (Coleção ABRH de Recursos Hídricos; 7), 1999. p. 99-150.

LEITE, J. G. *Aplicação das Técnicas de espectroscopia FTIR e de micro espectroscopia confocal Raman à preservação do patrimônio*. 2008. 73 f. Tese (Doutorado em Ciências) - Faculdade de Engenharia da Universidade do Porto, 2008.

LEUNG, D. Y. C.; WU, X.; LEUNG, M.K.H. A review on biodiesel production using catalyzed transesterification. *Applied Energy*, v. 87, n. 4, p. 1083-1095, abr. 2010.

LEUNG, D. Y. C.; GUO, Y. Transesterification of neat and used frying oil: optimization for biodiesel production. *Fuel Processing Technology*, v. 87, n. 10, p. 883 - 890, out. 2006.

LIMA, J. R. O.; SILVA, R. B.; SILVA, C. M. Biodiesel de babaçu (*Orignya* sp.) obtido por via etanólica. *Química Nova*, v. 30, n. 3, p. 600-603, jan. 2007.

LIRA, L. F. B.; VASCONCELOS, F. V. C.; PEREIRA, C. F.; PAIM, A. P. S.; STRAGEVITCH, L.; PIMENTEL, M. F. Prediction of properties of diesel/ biodiesel blends by infrared spectroscopy and multivariate calibration. *Fuel*, v. 89, n. 2, p. 405-409, fev. 2010 (a).

LIRA, L. F. B.; ALBUQUERQUE, M. S.; PACHECO, J. G. A.; FONSECA, T. M.; CAVALCANTI, E. H. S.; STRAGEVITCH, L.; PIMENTEL, M. F. Infrared spectroscopy and multivariate calibration to monitor stability quality parameters of biodiesel. *Microchemical Journal*, v. 96, p.126–131, mar. 2010(b).

LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. Biodiesel: Parâmetros de qualidade e métodos analíticos. *Química Nova*, v. 32, n. 6, p. 1596, jul. 2009.

LOPES, J. *Supervisão e diagnóstico de processos farmacêuticos com métodos inteligentes de análise de dados*. Universidade Técnica de Lisboa. Lisboa: UTL, 2001. p. 311.

LORENA, A. C.; CARVALHO, A. C. P. L. F. Uma Introdução às *Support Vector Machines*. RITA, v. 14 , n. 2, p. 43-67, jul. 2007.

LOURENÇO, L. M. *Desenvolvimento de método de determinação de glicerol livre em biodiesel utilizando a técnica de varredura sucessiva de potencial*. 2009. 120 f. Tese (doutorado em Química) - Instituto de Química da Universidade Estadual Paulista, Araraquara, 2009.

LU, W.; WANG, W. Potencial assessment of the “support vector machine” method in forecasting ambient air pollutant trends. *Chemosphere*, v. 59, n. 5, p. 693-701, dez. 2005.

LUCAS, D. C. Algoritmos Genéticos: uma introdução (apostila), p. 1-47 , 2002.

LYRA, W. S.; DA SILVA, E. C.; ARAÚJO, M. C. U.; FRAGOSO, W. D.; VERAS, G. Classificação periódica: um exemplo didático para ensinar análise de componentes principais. *Química Nova*, v. 33, n. 7, p.1594-1597, jul. 2010.

MADDEN, H. H. Comments on the Savitzky-Golay Convolution Method for Least-Squares Fit Smoothing and Differentiation of Digital Data. *Analytical Chemistry*, v. 50, n. 9, p. 1383-1386, ago. 1978.

MATERNIS, H.; NAES, T. *Multivariate Calibration*. New York: John Wiley & Sons, p. 419, 1989.

MARCHETTI, J. M.; MIGUEL, V. U.; ERRAZU, A. F. Possible methods for biodiesel production. *Renewable and Sustainable Energy Reviews*, v. 11, n. 6, p. 1300, ago. 2007.

MEHER, L. C.; SAGAR, D. V.; NAIK, S. N. Technical aspects of biodiesel production by transesterification - a review. *Renewable and Sustainable Energy Reviews*, v. 10, n. 3, p. 248, jun. 2006.

MEJIA, J. D.; SALGADO, N.; ORREGO, C. E. Effect of blends of Diesel and Palm-Castor biodiesels on viscosity, cloud point and flash point. *Industrial Crops and Products*, v. 43, p. 791-797, maio 2013.

MELO, M. A. R. *Monitoramento da estabilidade oxidativa no armazenamento de biodiesel metílico de soja/mamona e blendas em recipientes de vidro*. 2009. 93 f. Dissertação (Mestrado em Química) - Instituto de Química da Universidade Federal da Paraíba, João Pessoa, PB, 2009.

MENEZES, J. C.; FERREIRA, A. P.; RODRIGUES, L. O.; BRÁS, L. P.; ALVES, T. P. Chemometrics role within the PAT context: examples from primary pharmaceutical manufacturing. Separata de: BROWN, S. D.; TAULER, R.; WALCZAK, B. *Comprehensive chemometrics: chemical and biochemical data analysis*. Oxford: Elsevier Science, 2009. p. 313–355.

MILLI, B. B.; GRIPA, D. C.; SIMONELLI, G.; OST, M.; MARTINS, D. Produção de biodiesel a partir da mistura de sebo bovino com óleo vegetal. *Enciclopédia biosfera, Centro Científico Conhecer - Goiânia*, v.7, n.12, p. 1–26, 2011.

MIRANDA, M. C. A. G. V. A. *Aplicação de espectroscopia de FT-IR para a serotipagem e avaliação da susceptibilidade à penicilina em Streptococcus pneumoniae*. 2008. 182 f. Dissertação (Mestrado em Microbiologia Clínica) – Faculdade de Medicina de Lisboa, Universidade de Lisboa, Lisboa, Portugal, 2008.

MOITA NETO, J. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. *Química Nova*, v. 21, n. 4, p. 467-469, out. 1998.

MONARD, M. C.; BARANAUSKAS, J. A. Conceitos de aprendizado de máquina. Separata de: Rezende, S. O. (Ed.) *Sistemas Inteligentes - Fundamentos e Aplicações*, 1. ed., Barueri SP: Editora Manole, 2003. p. 89–114.

MONTEIRO, M. R.; AMBROZIN, A. R. P.; SANTOS, M. S.; BOFFO, E. F.; PEREIRA-FILHO, E. R.; LIÃO, L. M.; FERREIRA, A. G. Evaluation of biodiesel–diesel blends quality using ¹H NMR and chemometrics. *Talanta*, v. 78, n. 3, p. 660-664, maio 2009.

MORAES, M. S. A. *Biodiesel de sebo: avaliação de propriedades e testes de consumo em motor a diesel*. 2008. 105 f. Dissertação (Mestrado em Química) – Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2008.

MORETTO, E.; FETT, R. *Tecnologia de óleos e gorduras vegetais na indústria de alimentos*. São Paulo: Editora e Livraria Varela LTDA, 1998.

MOTA, C. J. A.; SILVA, C. X. A.; GONÇALVES, V. L. C. Gliceroquímica: Novos produtos e processos a partir da glicerina de produção de biodiesel. *Química Nova*, v. 32, n. 3, p. 639-648, mar. 2009.

MULLER, K. R.; SMOLA, A.; RATSCH, G.; SCHOLKOPF, B.; KOHLMORGEN, J.; VAPNIK, V. Predicting time series with support vector machines. Separata de: GERSTNER, W.; GERMOND, A.; HASLER, M.; MICORD, J. D.(Eds.). *Artificial Neural Networks*, Berlin, 1997. p. 999–1004.

MUNDO DA QUIMIOMETRIA (2010). Disponível em: < <http://mundodaquimiometria.blogspot.com/> >. Acesso em: 20 de fev. 2012.

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

NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L. Interval Partial Least Square Regression (iPLS): A comparative Chemometric Study with an Example from near-infrared spectroscopy. *Applied Spectroscopy*, v. 54, n. 3, p. 413-419, mar. 2000.

OLIVEIRA, F. C. C. Modelos de calibração multivariada associados à espectroscopia vibracional para análise de misturas diesel- óleos vegetais. 2006. 108 f. Dissertação (Mestrado em química) - Instituto de química da UNB, Brasília, 2006.

PARENTE, E. J. S. *Biodiesel: Uma Aventura Tecnológica num País Engraçado*. 1. ed. Fortaleza: Unigrafica, 2003.

PEREIRA, A. F. C. *Determinação simultânea de acidez, índice de refração e viscosidade em óleos vegetais usando espectrometria NIR, calibração multivariada e seleção de variáveis*. 2007. 59 f. Dissertação (Mestrado em química) - UFPB, João Pessoa, Paraíba, 2007.

_____.; PONTES, M. J. C.; GAMBARRA NETO, F. F.; SANTOS, S. R. B.; GALVÃO, R. K. H.; ARAÚJO, M. C. U. NIR spectrometric determination of quality parameters in vegetable oils using iPLS and variable selection. *Food Research International*, v.41, n. 4, p. 341–348, maio 2008.

PERIN, G.; SACHINI, M.; ÁLVARO, G.; WESTPHAL, E.; JACOB, R. G.; LENARDÃO, E. J.; ANTHONISEN, D. G.; SILVA, S. D. A. Transesterificação do óleo de mamona via catálise heterogênea. In: CONGRESSO BRASILEIRO DE MAMONA, 2., 2006, Sergipe. *Resumos...Sergipe*: Embrapa, 2006. p. 1-7.

PICARD, R.; COOK, D. Cross-validation of regression models. *J. Amer. Statist. Assoc.*, v. 79, n. 387, p. 575–583, set. 1984.

PIERNA, J. A. F.; MASSART, D. L.; NOORD, O. E.; RICOUX, P. H. Direct Orthogonalization: some case studies. *Chemom. Intell. Lab. Syst.*, v. 55, n. 1-2, p. 101-108, jan. 2001.

PINTO, A. C.; GUARIEIRO, L. L. N.; REZENDE, M. J. C.; RIBEIRO, N. M.; TORRES E. A.; LOPES, W. A.; PERREIRA, P. A. P.; ANDRADE, J. B. Biodiesel: An Overview. *Journal Brazilian Chemists Society*, v. 16, n. 6b, p. 1313-1330, dez. 2005.

PINZI, S.; ALONSO, F.; OLMO, J. G.; DORADO, M. P. Near infrared reflectance spectroscopy and multivariate analysis to monitor reaction products during biodiesel production. *Fuel*, v. 92, n. 1, p. 354–359, fev. 2012.

PLATT, L. Fast training of SVM using sequential optimization. Separata de: SCHÜLKOPF, B.; BURGESS, B.; SMOLA, A. (Eds.). *Advances in Kernel Methods-support Vector Learning*. Cambridge: MIT Press, 1998. p. 185-208.

POUSA, G. P. A. G.; SANTOS, A. L. F.; SUAREZ, P. A. Z. History and policy of biodiesel in Brazil. *Energy Policy*, v.35, n. 11, p. 5393-5398, nov. 2007.

PRATES, R. G. D.; AUGUSTI, R.; FORTES, I. C. P. Electrospray Ionization Mass Spectrometry and Multivariate Calibration Analysis: A Combined Protocol To Quantify Biodiesel in Blends with Petrodiesel. *Energy Fuels*, v. 24, p. 3183–3188, mar. 2010.

RESOLUÇÃO ANP N° 07. ANP – Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, 19 de Mar. 2008. Disponível em: < http://nxt.anp.gov.br/nxt/gateway.dll/leg/resolucao_anp/2008/mar%C3%A7o/ranp%207%20-%202008.xml>. Acesso em: 20 de fev. 2012.

RINNAN, A.; VAN DEN BERG, F.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. *Trends Analytical Chemistry*, v. 28, n. 10, p.1201-1222, jul. 2009.

SÁNCHEZ, A. S.; GARCIA - NIETO, P. J.; FERNANDEZ, P. R.; DEL COZ DÍAZ, J. J.; IGLESIAS-RODRÍGUEZ, F. J. Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). *Mathematical and Computer Modelling*, v. 54, n. 5-6, p. 1453-1466, abr. 2011.

SANFORD, S.; WHITE, J. M.; SHAH, P. S.; WEE, C.; VALVERDE, A.; MEITZNER, G. *Feedstock and biodiesel characteristics report*, 2009.

SANTOS, N. A. *Propriedades termo-oxidativas e de fluxo do biodiesel de babaçu (Orbignya phalerata)*. 2008. 129 f. Dissertação (Mestrado em Físico-química) - Universidade Federal da Paraíba, Instituto de Química, João Pessoa, PB, 2008.

SAUER, E. *Desenvolvimento de metodologias multivariadas para análise de queijos por espectroscopia drift*. 2007. 134 f. Tese de doutorado, Universidade Federal do Paraná, Curitiba, 2007.

SAVITZKY, A.; GOLAY, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, v. 36, n. 8, p. 1627–1639, jul. 1964.

SCHÖLKOPF, B.; SMOLA, A. J. *Learning with Kernels*. Cambridge: MIT Press, MA, 2002.

SCHUCHARDT, U.; RINALDI, R.; GARCIA, C.; MARCINIUK, L. L.; ROSSI, A. V. Síntese de biodiesel: Uma proposta contextualizada de experimento para laboratório de química geral. *Química Nova*, v. 30, n. 5, p. 1374, set. 2007.

SHEN F, YING Y, LI B, ZHENG Y, HU J. Prediction of sugars and acids in Chinese rice wine by mid-infrared spectroscopy. *Food Research International*, v. 44, n. 5, p. 1521-27, jun. 2011.

SILVA, F. E. B.; SABIN, G. P.; MÜLLER, E. I.; PARISOTTO, G.; FLORES, E. M. M.; FERRÃO, M. F. Otimização de métodos de regressão multivariada para quantificação de sulfametoxazol e Trimetoprima em medicamentos. In: ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 27., 2007, Foz do Iguaçu. *Resumos...* Foz do Iguaçu: ABEPRO, 2007.

SILVA, P. R. F.; FREITAS, T. F. S. Biodiesel: o ônus e o bônus de produzir combustível. *Ciência Rural*, Santa Maria, v. 38, n. 3, p. 843-851, maio 2008.

SILVA FILHO, J. B. *Produção de biodiesel etílico de óleos e gorduras residuais (OGR) em reator químico de baixo custo*. 2010. 73 f. Dissertação (Mestrado em Engenharia de materiais) – Universidade Tecnológica Federal do Paraná, Curitiba, PR, 2010.

SILVERSTEIN, R. M.; WEBSTER, F. X. *Identificação Espectrofotométrica de Compostos Orgânicos*. 6. ed. Rio de Janeiro: LTC, 2000.

SJOBLOM, J.; SVENSSON, O.; JOSEFSON, M.; KULLBERG, H.; WOLD, S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemom. Intell. Lab. Syst.*, v. 44, n. 1-2, p. 229-244, dez. 1998.

SKOOG, D. A.; LEARY, J. J. *Principles of instrumental analysis*. 4. ed., Saunders College, 1992. p. 131.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. *Princípios de análise instrumental*. 5. ed. Porto Alegre: Bookman, 2002. p. 836.

SHAHID, E. M.; JAMAL, Y. A review of biodiesel as vehicular fuel. *Renewable Sustainable Energy Reviews*, v. 12, n. 9, p. 2484-2494, dez. 2008.

SMOLA, A. J.; BARLETT, P.; SCHÖLKOPF, B.; SCHUURMANS, D. Introduction to large margin classifiers. In: _____. *Advances in Large Margin Classifiers*. Cambridge: MIT press, 1999. p. 1–28.

SMOLA, A. J.; SCHÖLKOPF, B. A. *Learning with Kernels*. Cambridge: MIT Press, MA, 2002.

STEINWART, I.; CHRISTMANN, A. *Support Vector Machines*. New York: Springer, 2008(a).

_____. How SVMs can estimate quantiles and the median. In: _____. *Advances in Neural Information Processing Systems*, Cambridge, MA: MIT Press, 2008(b). v.20, p. 305-312.

SUAREZ, P. A. Z.; SANTOS, A. L. F.; RODRIGUES, J. P.; ALVES, M. B. Bicomcombustíveis a partir de óleos e gorduras: desafios tecnológicos para viabilizá-los. *Química Nova*, v. 32, n. 3, p. 768-775, mar. 2009.

TAMIME, A. Y.; ROBINSON, R. K. *Yoghurt: Science and technology*, 2. ed. Abington

Hall: Woodhead Publishing Limited, 1999. p. 285.

TAPANES, N. C. O.; ARANDA, D. A. G.; CARNEIRO, J. W. M.; ANTUNES, O. A. C. Transesterification of *Jatropha curcas* oil glycerides: theoretical and experimental studies of biodiesel reaction. *Fuel*, v. 87, n. 10-11, p. 2286-2295, ago. 2008.

TAVARES, C. R. O. *Síntese do glifosato marcado com nitrogênio-15*. 2005. 74 f. Tese (doutorado em ciências) - Centro de Energia Nuclear na Agricultura da Universidade de São Paulo, São Paulo, 2005.

TEIXEIRA, L. S. G.; OLIVEIRA, F. S.; DOS SANTOS, H. C.; CORDEIRO, P. W. L.; ALMEIDA, S. Q. Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline. *Fuel*, v. 87, n. 3, p. 346-352, mar. 2008.

TEÓFILO, R. F. 2013. *Métodos Quimiométricos: Uma Visão Geral - Conceitos básicos de quimiometria*. Universidade Federal de Viçosa, Viçosa, v. 1, 118 p.

VAPNIK, V. N. *The nature of Statistical learning theory*. New York: Springer-Verlag, 1995.

_____. *Statistical Learning Theory*. Huddersfield: John Wiley and Sons, 1998.

_____. *The Nature of Statistical Learning Theory*. New York: Springer, 1999.

_____.; CHERVONENKIS, A. Y. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, v. 16, n. 2, p. 264-280, maio 1971.

VASCONCELOS, A. F. F. *Conformidade de misturas binárias de biodieseis etílicos quanto aos parâmetros reológicos, fluidodinâmicos, oxidativos e carburantes*. 2009. 116 f. Tese (doutorado em química) - UFPB/CCEN, João Pessoa, Paraíba, 2009.

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. 93 f. Dissertação (mestrado em Química Analítica) - UFPB/CCEN, João Pessoa, Paraíba, 2011.

VÉKEY, K.; Mass spectrometry and mass-selective detection in chromatography. *J. Chromatogr.*, v. 921, n. 2, p. 227-236, jul. 2001.

VIERA, M. S. *Qualificação e quantificação de adulterantes em misturas de biodiesel/diesel empregando espectroscopia no infravermelho por reflexão total atenuada e calibração multivariada*. 2011. 109 f. Dissertação (Mestrado em Sistemas e Processos Industriais) – Universidade de Santa Cruz do Sul, Santa Cruz do Sul, Rio Grande do Sul, 2011.

WANG, W.; MEN, C.; LU, W. Online prediction model based on support vector machine. *Neurocomputing*, v. 71, n. 4-6, p. 550-558, jan. 2008(a).

WANG, H.; TANG, H.; WILSON, J.; SALLEY, S. O.; NG, K. Y. S. Total Acid Number Determination of Biodiesel and Biodiesel Blends. *J. Am. Oil Chem. Soc.*, v. 85, n. 11, p. 1083-1086, ago. 2008(b).

WANG, L. B.; YU, H. Y.; HE, X. H.; LIU, R. Y. Influence of fatty acid composition of woody biodiesel plants on the fuel properties. *J. Fuel Chem. Technol.*, v. 40, n.4, p. 397-404, abr. 2012.

WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M. *PLS_Toolbox 3.01*, Eigenvector Research, Inc.: Manson, 2003.

WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M.; WINDIG, W.; KOCH, R. S. *Chemometrics Tutorial for PLS_Toolbox and Solo*. Eigenvector, 2006.

WISE, B. M.; GALLAGHER, N. B.; BRO, R.; SHAVER, J. M.; WINDIG, W.; KOCH, R. S. *PLS Toolbox version 4.0 for use with Matlab, 2006*, Eigenvector research, Inc. Wenatchee, WA, USA. Disponível em: <<http://www.eigenvector.com>>. Acesso em fev. 2013.

WOLD, H. Soft modeling: The basic design and some extensions. Separata de: JORESKOG, K. G.; WOLD, H. (Ed.). *Systems under indirect observation*. Amsterdam: North-Holland, 1982. v.2, p. 1- 53.

WOLD, S.; ANTTI, H.; LINDGREN, F.; OHMAN, J. Orthogonal signal correction of near-infrared spectra. *Chemom. Intell. Lab. Syst.*, v. 44, n. 1-2, p.175-185, dez. 1998.

WORKMAN, J., WEYER, L. *Practical Guide to Interpretative Near Infrared Spectroscopy*. Boca Raton, USA: CRC Press, 2008.

YAN, B.; GREMLICH, H. U.; MOSS, S.; COPPOLA, G. M.; SUN, Q.; LIU, L. A Comparison of Various FTIR and FT Raman Methods: Applications in the Reaction Optimization Stage of Combinatorial Chemistry. *J. Comb. Chem.*, v. 1, n. 1, p. 46-54, dez. 1999.

YEGANEH, B.; MOTLAGH, M. S. P.; RASHIDI, Y.; KAMALAN, H. Prediction of CO concentrations based on a hybrid Partial Least Square and Support Vector Machine model. *Atmospheric Environment*, v. 55, p. 357-365, ago. 2012.

ZHAO, C.; GAO, F.; WANG, F. Phase-based joint modeling and spectroscopy analysis for batch processes monitoring. *Ind. Eng. Chem. Res.*, v. 49, n. 2, p. 669–681, dez. 2010.

ZHAO, C.; GAO, F. Multiblock-Based Qualitative and Quantitative Spectral Calibration Analysis. *Ind. Eng. Chem. Res.*, v. 49, n. 18, p. 8694–8704, nov. 2010.

ZOU, X.; ZHAO, J.; HUANG, X.; LI, Y. Use of FT-NIR spectrometry in non-invasive measurements of soluble solid contents (SSC) of Fuji apple based on different PLS models. *Chemometrics and Intelligent Laboratory Systems*, v. 87, n. 1, p. 43-51, maio 2007(a).

ZOU, X.; LI, Y.; ZHAO, J. Using genetic algorithm interval partial least squares selection of the optimal near infrared wavelength regions for determination of the soluble solids content of Fuji apple. *Journal of Near Infrared Spectroscopy*, v. 15, n. 3, p. 153-159, jan. 2007(b).

ANEXO: Informação da composição e nomeação das amostras de biodiesel nos conjuntos espectrais de NIR e Mid-IR

Amostra	Composição do biodiesel (fração volumétrica)					Massa Específica	Índice Refração	PEFF	NIR		Mid-IR	
									Numeração no conjunto		Numeração no conjunto	
	Canola	Girassol	Milho	Soja	Região Sul	g.cm ⁻³	-	°C	Calibração	Validação	Calibração	Validação
Am021	0	0	0	1	0	0,8820	1,4575	-5,0	1	-	1	-
Am022	0	0	0	1	0	0,8813	1,4568	-5,0	2	-	2	-
Am023	0	0	0	1	0	0,8813	1,4569	-5,0	3	-	3	-
Am024	0	0	0	1	0	0,8814	1,4570	-5,0	4	-	4	-
Am025	0	0	0	1	0	0,8814	1,4570	-5,0	5	-	5	-
Am026	0	0	0	1	0	0,8813	1,4568	-5,0	6	-	6	-
Am027	0	0	0	1	0	0,8813	1,4569	-5,0	7	-	7	-
Am028	0	0	0	1	0	0,8816	1,4571	-5,0	8	-	8	-
Am029	0	0	1	0	0	0,8799	1,4559	-3,0	9	-	9	-
Am030	0	0	1	0	0	0,8797	1,4558	-2,0	10	-	-	101
Am031	0	0	1	0	0	0,8798	1,4558	-3,0	11	-	-	102
Am032	0	0	1	0	0	0,8799	1,4559	-2,0	12	-	-	103
Am033	1	0	0	0	0	0,8793	1,4556	-6,0	13	-	10	-
Am034	1	0	0	0	0	0,8792	1,4557	-7,0	14	-	11	-
Am035	1	0	0	0	0	0,8792	1,4556	-6,0	x	x	12	-
Am036	1	0	0	0	0	0,8792	1,4556	-7,0	x	x	13	-
Am037	0	1	0	0	0	0,8809	1,4570	-2,0	15	-	14	-
Am038	0	1	0	0	0	0,8815	1,4575	-1,0	16	-	15	-
Am039	0	1	0	0	0	0,8815	1,4575	-2,0	17	-	16	-
Am040	0	1	0	0	0	0,8816	1,4575	-1,0	18	-	17	-
Am041	0	0	0	1	0	0,8815	1,4570	-5,0	19	-	18	-
Am042	0	0	1	0	0	0,8799	1,4559	-2,6	20	-	19	-
Am043	1	0	0	0	0	0,8790	1,4553	-6,6	21	-	20	-
Am044	0	1	0	0	0	0,8814	1,4572	-1,6	22	-	21	-

Amostra	Composição do biodiesel (fração volumétrica)					Massa Específica	Índice Refração	PEFF	NIR		Mid-IR	
									Numeração no conjunto		Numeração no conjunto	
	Canola	Girassol	Milho	Soja	Região Sul	g.cm ⁻³	-	°C	Calibração	Validação	Calibração	Validação
Am045	0	0	0	1	0	0,8813	1,4569	-5,0	23	-	-	104
Am046	0	0	0	1	0	0,8814	1,4569	-5,0	24	-	-	105
Am047	0	0	0	1	0	0,8813	1,4569	-5,0	25	-	22	-
Am049	0	0	0	1	0	0,8816	1,4574	-5,0	26	-	23	-
Am057	0	0	0	1	0	0,8816	1,4573	-5,0	27	-	-	106
Am058	0	0	0	1	0	0,8818	1,4573	-5,0	28	-	-	107
Am059	0	0	0	1	0	0,8814	1,4570	-5,0	29	-	24	-
Am060	0	0	0	1	0	0,8818	1,4574	-4,0	30	-	25	-
Am061	0	0	0	0	1	0,8777	1,4535	6,0	31	-	26	-
Am062	0	0	0	0,5	0,5	0,8797	1,4554	-4,7	32	-	27	-
Am064	0	0	0	1	0	0,8810	1,4569	-5,0	33	-	28	-
Am066	0	0	0	1	0	0,8813	1,4571	-5,0	34	-	-	108
Am067	0	0	0,9	0,1	0	0,8801	1,4560	-3,0	-	101	-	109
Am068	0	0	0,9	0,1	0	0,8802	1,4560	-3,0	-	102	-	110
Am069	0	0	0,9	0,1	0	0,8803	1,4560	-3,0	35	-	29	-
Am070	0	0	0,7	0,3	0	0,8806	1,4563	-4,0	36	-	30	-
Am071	0	0	0,7	0,3	0	0,8805	1,4562	-3,0	37	-	-	111
Am072	0	0	0,7	0,3	0	0,8805	1,4562	-3,0	38	-	-	112
Am073	0	0	0,5	0,5	0	0,8809	1,4564	-4,0	-	103	31	-
Am074	0	0	0,5	0,5	0	0,8808	1,4564	-3,0	39	-	-	113
Am075	0	0	0,5	0,5	0	0,8809	1,4565	-3,0	-	104	-	114
Am076	0	0	0,3	0,7	0	0,8812	1,4567	-4,0	40	-	32	-
Am077	0	0	0,3	0,7	0	0,8812	1,4567	-4,0	-	105	-	115
Am078	0	0	0,3	0,7	0	0,8812	1,4567	-4,0	-	106	-	116
Am079	0	0	0,1	0,9	0	0,8814	1,4569	-4,0	-	107	-	117
Am080	0	0	0,1	0,9	0	0,8815	1,4569	-5,0	41	-	-	118

Amostra	Composição do biodiesel (fração volumétrica)					Massa Específica g.cm ⁻³	Índice Refração -	PEFF °C	NIR		Mid-IR	
	Canola	Girassol	Milho	Soja	Região Sul				Numeração no conjunto		Numeração no conjunto	
									Calibração	Validação	Calibração	Validação
Am081	0	0	0,1	0,9	0	0,8815	1,4569	-4,0	42	-	33	-
Am082	0	0,9	0	0,1	0	0,8821	1,4575	-2,0	-	108	34	-
Am083	0	0,9	0	0,1	0	0,8815	1,4572	-2,0	43	-	35	-
Am084	0	0,9	0	0,1	0	0,8814	1,4572	-4,0	-	109	36	-
Am085	0	0,7	0	0,3	0	0,8816	1,4572	-3,0	-	110	37	-
Am086	0	0,7	0	0,3	0	0,8816	1,4573	-3,0	44	-	38	-
Am087	0	0,7	0	0,3	0	0,8815	1,4572	-4,0	-	111	39	-
Am088	0	0,5	0	0,5	0	0,8818	1,4573	-4,0	45	-	40	-
Am089	0	0,5	0	0,5	0	0,8817	1,4572	-4,0	46	-	41	-
Am090	0	0,5	0	0,5	0	0,8817	1,4572	-6,0	47	-	-	119
Am091	0	0,3	0	0,7	0	0,8817	1,4571	-5,0	48	-	42	-
Am092	0	0,3	0	0,7	0	0,8814	1,4571	-5,0	-	112	43	-
Am093	0	0,3	0	0,7	0	0,8814	1,4571	-5,0	-	113	44	-
Am094	0	0,1	0	0,9	0	0,8814	1,4570	-5,0	-	114	45	-
Am095	0	0,1	0	0,9	0	0,8816	1,4571	-5,0	49	-	46	-
Am096	0	0,1	0	0,9	0	0,8816	1,4571	-5,0	-	115	47	-
Am097	0,9	0	0	0,1	0	0,8795	1,4557	-7,0	50	-	48	-
Am098	0,9	0	0	0,1	0	0,8793	1,4556	-8,0	51	-	49	-
Am099	0,9	0	0	0,1	0	0,8795	1,4557	-8,0	52	-	50	-
Am100	0,7	0	0	0,3	0	0,8800	1,4560	-6,0	53	-	51	-
Am101	0,7	0	0	0,3	0	0,8800	1,4560	-6,0	54	-	52	-
Am102	0,7	0	0	0,3	0	0,8799	1,4560	-7,0		116	53	-
Am103	0,5	0	0	0,5	0	0,8804	1,4563	-5,0		117	54	-
Am104	0,5	0	0	0,5	0	0,8803	1,4563	-5,0	55	-	55	-
Am105	0,5	0	0	0,5	0	0,8805	1,4564	-6,0	-	118	-	120

Amostra	Composição do biodiesel (fração volumétrica)					Massa Específica	Índice Refração	PEFF	NIR		Mid-IR	
									Numeração no conjunto		Numeração no conjunto	
	Canola	Girassol	Milho	Soja	Região Sul	g.cm ⁻³	-	°C	Calibração	Validação	Calibração	Validação
Am106	0,3	0	0	0,7	0	0,8808	1,4566	-6,0	56	-	56	-
Am107	0,3	0	0	0,7	0	0,8808	1,4566	-5,0	57	-	-	121
Am108	0,3	0	0	0,7	0	0,8810	1,4567	-6,0	-	119	57	-
Am109	0,1	0	0	0,9	0	0,8813	1,4569	-5,0	-	120	58	-
Am110	0,1	0	0	0,9	0	0,8814	1,4569	-5,0	58	-	-	122
Am111	0,1	0	0	0,9	0	0,8815	1,4570	-5,0	-	121	59	-
Am112	0,1	0	0,4	0,5	0	0,8806	1,4564	-6,0	59	-	60	-
Am113	0,1	0	0,4	0,5	0	0,8808	1,4564	-5,0	60	-	61	-
Am114	0,1	0	0,4	0,5	0	0,8806	1,4564	-5,0	61	-	62	-
Am115	0,3	0	0,3	0,4	0	0,8806	1,4562	-7,0	-	122	63	-
Am116	0,3	0	0,3	0,4	0	0,8806	1,4563	-6,0	-	123	64	-
Am117	0,3	0	0,3	0,4	0	0,8805	1,4562	-6,0	62	-	65	-
Am118	0,4	0	0,1	0,5	0	0,8805	1,4563	-5,0	-	124	-	123
Am119	0,4	0	0,1	0,5	0	0,8805	1,4563	-4,0	-	125	66	-
Am120	0,4	0	0,1	0,5	0	0,8805	1,4563	-6,0	63	-	67	-
Am121	0	0,1	0,4	0,5	0	0,8811	1,4566	-5,0	-	126	68	-
Am122	0	0,1	0,4	0,5	0	0,8811	1,4566	-5,0	-	127	69	-
Am123	0	0,1	0,4	0,5	0	0,8811	1,4566	-4,0	64	-	-	124
Am124	0	0,3	0,3	0,4	0	0,8813	1,4568	-5,0	65	-	70	-
Am125	0	0,3	0,3	0,4	0	0,8811	1,4568	-5,0	66	-	71	-
Am126	0	0,3	0,3	0,4	0	0,8811	1,4568	-5,0	67	-	72	-
Am127	0	0,4	0,1	0,5	0	0,8814	1,4570	-5,0	68	-	73	-
Am128	0	0,4	0,1	0,5	0	0,8814	1,4570	-4,0	69	-	-	125
Am129	0	0,4	0,1	0,5	0	0,8816	1,4570	-4,0	-	128	74	-
Am130	0,4	0,1	0	0,5	0	0,8808	1,4565	-5,0	70	-	75	-

Amostra	Composição do biodiesel (fração volumétrica)					Massa Específica	Índice Refração	PEFF	NIR		Mid-IR	
									Numeração no conjunto		Numeração no conjunto	
	Canola	Girassol	Milho	Soja	Região Sul	g.cm ⁻³	-	°C	Calibração	Validação	Calibração	Validação
Am131	0,4	0,1	0	0,5	0	0,8805	1,4564	-5,0	71	-	-	126
Am132	0,4	0,1	0	0,5	0	0,8807	1,4565	-5,0	-	129	-	127
Am133	0,3	0,3	0	0,4	0	0,8809	1,4567	-4,0	-	130	-	128
Am134	0,3	0,3	0	0,4	0	0,8811	1,4568	-5,0	72	-	-	129
Am135	0,3	0,3	0	0,4	0	0,8808	1,4566	-4,0	73	-	76	-
Am136	0,1	0,4	0	0,5	0	0,8813	1,4569	-4,0	74	-	77	-
Am137	0,1	0,4	0	0,5	0	0,8812	1,4569	-4,0	-	131	-	130
Am138	0,1	0,4	0	0,5	0	0,8813	1,4569	-4,0	-	132	78	-
Am139	0,25	0,25	0,25	0,25	0	0,8806	1,4564	-6,0	75	-	79	-
Am140	0,25	0,25	0,25	0,25	0	0,8807	1,4565	-6,0	-	133	-	131
Am141	0,25	0,25	0,25	0,25	0	0,8806	1,4564	-6,0	-	134	-	132
Am142	0,1	0,1	0,4	0,4	0	0,8809	1,4565	-5,0	76	-	-	133
Am143	0,1	0,1	0,4	0,4	0	0,8808	1,4564	-6,0	-	135	80	-
Am144	0,1	0,1	0,4	0,4	0	0,8809	1,4566	-5,0	-	136	-	134
Am145	0,4	0,1	0,1	0,4	0	0,8805	1,4563	-6,0	77	-	81	-
Am146	0,4	0,1	0,1	0,4	0	0,8805	1,4563	-6,0	78	-	-	135
Am147	0,4	0,1	0,1	0,4	0	0,8805	1,4563	-6,0	79	-	-	136
Am148	0,1	0,4	0,1	0,4	0	0,8812	1,4568	-4,0	80	-	82	-
Am149	0,1	0,4	0,1	0,4	0	0,8812	1,4568	-5,0	-	137	-	137
Am150	0,1	0,4	0,1	0,4	0	0,8812	1,4569	-5,0	-	138	83	-
Am151	0,4	0,1	0,4	0,1	0	0,8801	1,4560	-7,0	81	-	84	-
Am152	0,4	0,1	0,4	0,1	0	0,8801	1,4560	-7,0	82	-	-	128
Am153	0,4	0,1	0,4	0,1	0	0,8801	1,4560	-7,0	83	-	-	139
Am154	0,1	0,4	0,4	0,1	0	0,8809	1,4566	-6,0	84	-	85	-
Am155	0,1	0,4	0,4	0,1	0	0,8809	1,4565	-6,0	-	139	-	140

Amostra	Composição do biodiesel (fração volumétrica)					Massa Específica	Índice Refração	PEFF	NIR		Mid-IR	
									Numeração no conjunto		Numeração no conjunto	
	Canola	Girassol	Milho	Soja	Região Sul	g.cm ⁻³	-	°C	Calibração	Validação	Calibração	Validação
Am156	0,1	0,4	0,4	0,1	0	0,8808	1,4565	-6,0	-	140	86	-
Am157	0,4	0,4	0,1	0,1	0	0,8805	1,4564	-7,0	-	141	87	-
Am158	0,4	0,4	0,1	0,1	0	0,8806	1,4564	-5,0	85	-	x	x
Am159	0,4	0,4	0,1	0,1	0	0,8808	1,4565	-6,0	86	-	88	-
Am160	0,1	0	0	0	0,9	0,8778	1,4537	4,0	87	-	-	141
Am161	0,1	0	0	0	0,9	0,8778	1,4537	4,0	88	-	89	-
Am162	0,1	0	0	0	0,9	0,8778	1,4537	4,0	-	142	-	142
Am163	0,3	0	0	0	0,7	0,8782	1,4541	2,0	89	-	90	-
Am164	0,3	0	0	0	0,7	0,8782	1,4541	2,0	90	-	-	143
Am165	0,3	0	0	0	0,7	0,8782	1,4541	2,0	-	143	-	144
Am166	0,5	0	0	0	0,5	0,8783	1,4544	-1,0	91	-	-	145
Am167	0,5	0	0	0	0,5	0,8783	1,4545	-1,0	-	144	-	146
Am168	0,5	0	0	0	0,5	0,8783	1,4544	0,0	-	145	91	-
Am169	0,7	0	0	0	0,3	0,8786	1,4549	-3,0	92	-	92	-
Am170	0,7	0	0	0	0,3	0,8786	1,4548	-4,0	93	-	93	-
Am171	0,7	0	0	0	0,3	0,8786	1,4548	-4,0	94	-	94	-
Am172	0,9	0	0	0	0,1	0,8789	1,4553	-6,0	-	146	-	147
Am173	0,9	0	0	0	0,1	0,8789	1,4553	-5,0	95	-	95	-
Am174	0,9	0	0	0	0,1	0,8789	1,4552	-5,0	96	-	-	148
Am175	0	0	0	0	1	0,8776	1,4534	6,0	-	147	96	-
Am176	0	0	0	1	0	0,8816	1,4570	-5,0	97	-	97	-
Am177	1	0	0	0	0	0,8788	1,4554	-7,0	98	-	98	-
Am180	1	0	0	0	0	0,8790	1,4554	-6,6	99	-	99	-
Am181	0	0	0	0	1	0,8777	1,4535	6,0	100	-	100	-