



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA
LABORATÓRIO DE QUIMIOMETRIA DO TRIÂNGULO

LUCAS CAIXETA GONTIJO

**USO DE ESPECTROMETRIA NO INFRAVERMELHO MÉDIO,
CALIBRAÇÃO MULTIVARIADA E SELEÇÃO DE VARIÁVEIS POR
INTERVALOS NA QUANTIFICAÇÃO DE BODIESEIS EM MISTURAS
COM DIESEL**

UBERLÂNDIA

2016

LUCAS CAIXETA GONTIJO

**USO DE ESPECTROMETRIA NO INFRAVERMELHO MÉDIO,
CALIBRAÇÃO MULTIVARIADA E SELEÇÃO DE VARIÁVEIS POR
INTERVALOS NA QUANTIFICAÇÃO DE BIODIESEIS EM MISTURAS
COM DIESEL**

Tese apresentada ao Programa de Pós-Graduação em Química da Universidade Federal de Uberlândia, como requisito parcial para obtenção de título de Doutor em Química.

Área de concentração: Química Analítica

Orientador: Prof. Dr. Waldomiro Borges Neto

UBERLÂNDIA

2016

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

G641u Gontijo, Lucas Caixeta, 1982-
2016 Uso de espectrometria no infravermelho médio, calibração
multivariada e seleção de variáveis por intervalos na quantificação de
biodieseis em misturas com diesel / Lucas Caixeta Gontijo. - 2016.
104 f. : il.

Orientador: Waldomiro Borges Neto.
Tese (doutorado) - Universidade Federal de Uberlândia, Programa
de Pós-Graduação em Química.
Inclui bibliografia.

1. Química - Teses. 2. Biodiesel - Teses. 3. Espectroscopia de
infravermelho - Teses. 4. Calibração - Teses. I. Borges Neto, Waldomiro.
II. Universidade Federal de Uberlândia. Programa de Pós-Graduação em
Química. III. Título.



**Ata da defesa de TESE DE DOUTORADO junto ao Programa de Pós-Graduação em Química,
do Instituto de Química da Universidade Federal de Uberlândia
DEFESA DE TESE DE DOUTORADO EM QUÍMICA, NÚMERO 48/PPQUI.**

DATA: 18/03/2016

DISCENTE: Lucas Caixeta Gontijo

MATRÍCULA: 11213QMI007

TÍTULO DO TRABALHO: Uso de espectrometria no infravermelho médio, calibração multivariada e seleção de variáveis por intervalos para quantificação de biodieseis em misturas com diesel.

ÁREA DE CONCENTRAÇÃO: Química

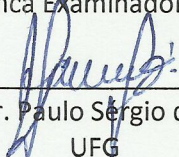
LINHA DE PESQUISA: Espectroanalítica Aplicada

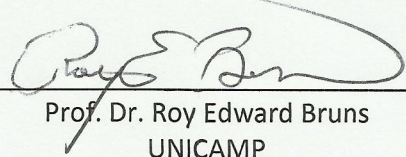
PROJETO DE PESQUISA DE VINCULAÇÃO: Uso de espectrometria no infravermelho médio e métodos quimiométricos de seleção de variáveis para quantificação de biodieseis de soja, algodão e pinhão manso em diesel.

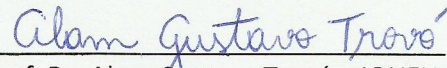
Às quatorze horas do dia dezoito de março do ano dois mil e dezesseis, no Auditório Prof. Manuel Gonzalo Hernández Terrones, Bloco 5I, Campus Santa Mônica, reuniu-se a Banca Examinadora composta pelos Professores Doutores Paulo Sérgio de Souza, da Universidade Federal de Goiás, Roy Edward Bruns, da Universidade Estadual de Campinas, Alam Gustavo Trovó e Alex Domingues Batista, da UFU e Waldomiro Borges Neto, orientador(a) e presidente da mesa. Iniciando os trabalhos, o(a) presidente da mesa apresentou o(a) candidato(a) e a Banca Examinadora, agradeceu a presença do público e discorreu sobre as normas e critérios para a realização desta sessão, baseadas na Norma Regulamentar nº 03/2013/PPQUI. Em seguida, o(a) presidente da mesa concedeu a palavra ao(a) candidato(a) para a exposição do seu trabalho e, em sequência, aos examinadores, em ordem sucessiva, para arguir o(a) apresentador(a). A duração da apresentação e o tempo de arguição e resposta deram-se conforme as normas do Programa. Ultimada a arguição, desenvolvida dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu os conceitos finais e A provou o(a) candidata(o). Por sugestão da Banca Examinadora, o título do trabalho será

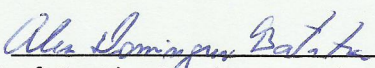
mantido conforme o original

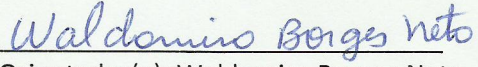
Esta defesa de Tese de Doutorado é parte dos requisitos necessários à obtenção do título de Doutor. O competente diploma será expedido após cumprimento do estabelecido nas normas do Programa, legislação e regulamentação internas da UFU. As correções observadas pelos examinadores deverão ser realizadas no prazo máximo de 30 dias. Nada mais havendo a tratar, deu-se por encerrada a sessão às 17 horas e 25 minutos e lavrada a presente ata, que após lida e aprovada, foi assinada pela Banca Examinadora.


Prof. Dr. Paulo Sérgio de Souza
UFU


Prof. Dr. Roy Edward Bruns
UNICAMP


Prof. Dr. Alam Gustavo Trovó – IQUFU


Prof. Dr. Alex Domingues Batista – IQUFU


Professor(a) Orientador(a): Waldomiro Borges Neto – IQUFU

Dedico este trabalho primeiramente a Deus, por ser essencial em minha vida. Ao meu pai Carolino (*in memoriam*) e minha mãe Juversina que foram meus primeiros educadores/orientadores. À minha esposa Núbia e aos meus filhos Lukian e Nubiany pois suas presenças significam segurança e certeza de que não estou sozinho nesta caminhada.

AGRADECIMENTOS

É difícil agradecer todas as pessoas que de algum modo, nos momentos serenos e ou apreensivos, fizeram ou fazem parte da minha vida, por isso primeiramente agradeço a todos de coração.

Ao prof. Dr. Waldomiro Borges Neto pelo convívio, pelo apoio, compreensão e amizade. Obrigado pela orientação segura e competente e, principalmente, pela confiança depositada em meu trabalho.

Aos meus pais, Carolino Caixeta Gontijo (*in memorian*) e Juversina Geralda dos Santos Gontijo que iluminaram os caminhos obscuros com afeto e dedicação para que os trilhasse sem medo e cheio de esperança; que se doaram inteiros e renunciaram aos seus sonhos, para que, muitas vezes, pudesse realizar o meu; Muito obrigado, AMO VOCÊS.

À Núbia Rosa Martins Gontijo, minha esposa, pessoa que amo partilhar a vida. Obrigado pelo carinho, a paciência e por sua capacidade de me trazer paz na correria de cada ano de doutoramento.

Aos meus filhos, Lukian Martins Gontijo e Nubiany Martins Gontijo, que embora não tenham conhecimento disto, mas iluminam de maneira muito especial os meus pensamentos, os meus dias, me levando a buscar mais conhecimentos.

Agradeço aos meus irmãos, Valéria Maria Caixeta e Luciano dos Santos Gontijo, que por mais difícil que fossem as circunstâncias, sempre tiveram ao meu lado me dando apoio.

Agradeço ao prof. Dr. Douglas Queiroz Santos pelas conversas formais e informais, pelos seus conselhos, confiança, presteza e amizade. Pessoas como você tem que “DAR VALOR”.

Aos amigos do Laboratório, Hery Mitsutake, Felipe Bachion de Santana, Eloiza Guimarães, Letícia Maria Souza, Carlos Tadeu Zanetti, Sarmento Júnior Mazivila, Baltazar Vasco Siteo, Ademar Viagem Máquina, José Eduardo Buiatte e Edvando Souza Teles (que equipe!!!) pelas discussões e pelo apoio nos momentos difíceis. Se contar com as festas (momentos fáceis)..... nooossa!!!! Valeu pessoal.

Aos companheiros de pesca, amigos e parentes, pelos momentos de descontração, isto permitiu renovar as energias na busca de realizar um bom trabalho.

À todos os amigos de trabalho do Instituto Federal Goiano – Campus Urutaí, aqui não cometerei o erro de cair no esquecimento, por isso não mencionarei os nomes.

Ao Instituto Federal Goiano – Campus Urutaí pela concessão da liberação total para realizar este curso de doutoramento.

À Universidade Federal de Uberlândia e ao Instituto de Química pelo apoio e por fornecerem toda a estrutura física e tecnológica para a realização deste trabalho.

Aos professores doutores Alam Gustavo Trovó, Sidnei Gonçalves da Silva e Alex Domingues Batista, pelas contribuições e questionamentos realizados durante o Exame de Qualificação.

Aos professores doutores Roy Edward Bruns e Paulo Sérgio Souza, pelas contribuições e questionamentos realizados durante a defesa desta Tese.

À TRANSPETRO S/A e CARAMURU S/A pelo fornecimento das amostras de diesel e biodiesel metílico de soja.

Aos professores, funcionários e colegas do Instituto de Química, que contribuíram direta ou indiretamente para a minha formação acadêmica.

À secretária da pós-graduação, Mayta Mamede Negreto Peixoto pelo atendimento cordial e eficiente.

À FAPEG/CAPES pela bolsa fornecida durante a realização deste projeto.

À todos aqueles que, de alguma forma, contribuíram para a realização deste trabalho.

E finalmente agradeço a Deus, por proporcionar estes agradecimentos a todos que tornaram minha vida mais afetuosa, além de ter me dado uma família maravilhosa e amigos sinceros. Deus, que a mim atribuiu alma e missões pelas quais já sabia que eu iria batalhar e vencer, agradecer é pouco. Por isso lutar, conquistar, vencer e até mesmo cair e perder, e o principal, viver é o meu modo de agradecer sempre.

MUITO OBRIGADO!!!

VALEU!!!

RESUMO

No Brasil, a partir de novembro de 2014, é exigido por lei a adição de 7% (v/v) de biodiesel ao diesel sendo esta mistura denominada de B7. De acordo com o boletim emitido pela Agência Nacional de Petróleo, Gás Natural e Biocombustíveis, em 2015, 33,6% do total das amostras de diesel comercial analisadas apresentaram teor de biodiesel fora das especificações de qualidade. Assim, o presente estudo teve por objetivo desenvolver e validar metodologias para quantificar biodieseis metílicos e etílicos, provenientes de óleos de soja, fritura usado, algodão e pinhão manso, em misturas binárias com diesel, empregando a técnica de Espectrometria no Infravermelho Médio (680 a 3100 cm^{-1}) e calibração multivariada por Quadrados Mínimos Parciais (PLS). Além disso, apresenta um estudo da aplicação de métodos de seleção de variáveis por intervalos como os Quadrados Mínimos Parciais por Intervalos (iPLS), Quadrados Mínimos Parciais por Exclusão de Intervalos (biPLS) e Quadrados Mínimos Parciais por Intervalos Sinérgicos (siPLS). Desta maneira, os modelos construídos a partir dos intervalos espectrais que proporcionaram, estatisticamente, menor valor de Erro Quadrático Médio de Previsão (RMSEP) e menor número de variáveis, em comparação ao modelo de espectro total, foram validados para determinação do respectivo tipo de biodiesel em mistura com diesel. Os resultados dos modelos PLS global, para todos os tipos de biodieseis, demonstram excelente correlação entre os valores reais e previstos do conjunto de previsão ($R > 0,99$), na faixa de concentração de 1,00 a 30,00% (v/v), valores de RMSEP entre 0,0473 e 0,1020% (v/v) e não apresentaram erros sistemáticos conforme a norma ASTM E1655. Os modelos PLS global propostos atenderam as exigências de exatidão requerida pela ABNT NBR15568 (RMSEP < 0,1%) além de utilizar apenas um modelo de calibração contra dois da respectiva norma. Em relação aos métodos de seleção de variáveis por intervalos estes foram eficientes na eliminação de variáveis não informativas melhorando significativamente a previsão de amostras de validação, se comparado ao desempenho do modelo de calibração multivariada por PLS, com exceção aos modelos contendo biodieseis provenientes de óleo de fritura usado e biodiesel metílico de pinhão manso. Portanto, as metodologias desenvolvidas apresentam como uma alternativa viável às análises usuais para a determinação do teor de biodiesel em misturas com diesel.

Palavras-chave: Biodiesel. Infravermelho. Calibração Multivariada. PLS. iPLS. Seleção de Variáveis.

ABSTRACT

In Brazil, since November 2014, it is required by law an addition of 7% (v/v) of biodiesel to diesel, and this mixture is called B7. According to the report issued by National Brazilian Agency of Petroleum, Natural Gas and Biofuels in 2015, 33.6% of commercial diesel samples examined presented biodiesel content outside quality specifications. Therefore, this work aims to develop and validate methodologies to quantify methyl and ethyl biodiesel from soybeans, used frying oil, cottonseed and jatropha into binary mixtures with diesel, using Mid Infrared Spectroscopy (680 to 3100 cm^{-1}) and multivariate calibration by Partial Least Squares (PLS). Furthermore, we present a study of application of variable selection methods for intervals as Interval Partial Least Squares (iPLS), Backward Interval Partial Least Squares (biPLS) and Synergy Interval Partial Least Squares (siPLS). Thus, the models built from spectral intervals, which provided, statistically, lower Root Mean Square Error of Prediction (RMSEP) and fewer variables compared to the model of full spectrum, were validated to determine the respective type of biodiesel mixed with diesel. The results of the global PLS models, to all kinds of biodiesels, demonstrated excellent correlation between actual and predicted values of the prediction set ($R > 0.99$), in the range of concentration of 1.00 to 30.00% (v/v), values of RSMEP between 0.0473 and 0.1020% (v/v) and did not show systematic errors according to ASTM E1655 norm. The proposed global PLS models met the requirements of accuracy demanded by ABNT NBR15568 ($\text{RMSEP} < 0.1\%$) besides using only one calibration model against two of the respective norm. In relation to Variable Selection methods by intervals, were efficient in the elimination of non-informative variables significantly improving the prediction of validation samples if compared to multivariate calibration by PLS models, with the exception of the models containing biodiesels from used frying oil and jatropha methyl biodiesel. Therefore, the developed methodologies are presented as a viable alternative to the usual analysis for determining the biodiesel content in mixtures as diesel.

Keywords: Biodiesel. Infrared. Multivariate Calibration. PLS. iPLS. Variable Selection.

LISTA DE SIGLAS

ABIOVE	Associação Brasileira das Indústrias de Óleos Vegetais
ABNT	Associação Brasileira de Normas Técnicas
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
ASTM	American Society for Testing and Materials
ATR	Reflectância Total Atenuada (do inglês, Attenuated Total Reflectance)
B7	7% de biodiesel e 93% de diesel em volume
BEA	Biodiesel Etílico de Algodão
BEPM	Biodiesel Etílico de Pinhão Manso
BES	Biodiesel Etílico de Soja
BESR	Biodiesel Etílico de Soja Residual
biPLS	Quadrados Mínimos Parciais por Exclusão de Intervalos (do inglês, Backward Interval Partial Least Squares)
BMA	Biodiesel Metílico de Algodão
BMPM	Biodiesel Metílico de Pinhão Manso
BMS	Biodiesel Metílico de Soja
BMSR	Biodiesel Metílico de Soja Residual
PC	Componente Principal (do inglês, Principal Component)
EN	Norma Europeia
GA	Algoritmo Genético (do inglês, Genetic Algorithm)
GC-MS	Cromatografia Gasosa acoplada com Espectrometria de Massas (do inglês, Gas Chromatography–Mass Spectrometry)
GL	Graus de Liberdade
iPLS	Quadrados Mínimos Parciais por Intervalos (do inglês, Interval Partial Least Squares)
IUPAC	União Internacional de Química Pura e Aplicada (do inglês, International Union of Pure and Applied Chemistry)
LC-MS	Cromatografia Líquida acoplada com Espectrometria de Massas (do inglês, Liquid Chromatography–Mass Spectrometry)
MCR-ALS	Resolução de Curvas Multivariadas por Quadrados Mínimos Alternados, (do inglês, Multivariate Curve Resolution-Alternating Least Squares)

MIR	Infravermelho Médio (do inglês, Middle Infrared)
MLR	Regressão Linear Múltipla (do inglês, Multiple Linear Regression)
NIR	Infravermelho Próximo (do inglês, Near Infrared)
NMR	Ressonância Magnética Nuclear (do inglês, Nuclear Magnetic Resonance)
PCA	Análise de Componentes Principais (do inglês, Principal Components Analysis)
PCR	Regressão por Componentes Principais (do inglês, Principal Components Regression)
PLS	Quadrados Mínimos Parciais (do inglês Partial Least Squares)
RMSEC	Erro Quadrático Médio da Calibração (do inglês, Root Mean Square Error of Calibration)
RMSECV	Erro Quadrático Médio da Validação Cruzada (do inglês, Root Mean Square Error of Cross Validation)
RMSEP	Erro Quadrático Médio da Previsão (do inglês, Root Mean Square Error of Prediction)
SDV	Desvio Padrão da Validação (do inglês, Standard Deviation of Validation)
siPLS	Quadrados Mínimos Parciais por Intervalos Sinérgicos (do inglês, Synergy Interval Partial Least Squares)
SPA	Algoritmo das Projeções Sucessivas (do inglês, Successive Projection Algorithm)
SVM	Máquinas de Vetores de Suporte (do inglês, Support Vector Machines)
UVE	Eliminação de Variáveis não Informativas (do inglês, Uninformative Variable Elimination)
VL	Variável Latente.

SUMÁRIO

1. INTRODUÇÃO	17
1.1 Biodiesel e matérias-primas	18
1.2 Teor, não conformidades e normas	20
1.3 Espectrometria no Infravermelho Médio	22
1.4 Quimiometria	24
1.4.1 <i>Análise de Componentes Principais (PCA)</i>	26
1.4.2 <i>Regressão por Quadrados Mínimos Parciais (PLS)</i>	29
1.4.3 <i>Seleção de variáveis por intervalos</i>	32
1.5 Validação do modelo PLS	37
1.5.1 <i>Exatidão</i>	37
1.5.2 <i>Linearidade</i>	38
1.5.3 <i>Teste para Erro Sistemático</i>	38
1.6 Comparação dos modelos	39
1.7 Aplicações da Espectroscopia no Infravermelho combinada com os métodos multivariados para análise de amostras de biodiesel/diesel	40
1.8 Publicações do Laboratório de Quimiometria do Triângulo (LQT) utilizando Espectrometria no Infravermelho Médio aliada a métodos multivariados para análise de amostras de biodiesel/diesel	43
2 OBJETIVOS	46
2.1 Objetivo Geral	47
2.2 Objetivos Específicos	47
3 PROCEDIMENTO EXPERIMENTAL	48
3.1 Amostras de biodieseis e diesel	49

3.2	Preparo das misturas biodiesel/diesel	49
3.3	Obtenção dos espectros MIR	49
3.4	Construção dos modelos de calibração multivariada	50
4	RESULTADOS E DISCUSSÃO	53
4.1	Espectros MIR de biodiesel e diesel	54
4.2	Modelos PLS global	56
4.3	Modelos de seleção de variáveis por intervalos (iPLS, biPLS e siPLS)	72
4.3.1	<i>Modelos iPLS, biPLS e siPLS na quantificação de BMS em mistura com diesel</i> .	72
4.3.2	<i>Modelos iPLS, biPLS e siPLS na quantificação de BES em mistura com diesel</i> .	79
4.3.3	<i>Modelos iPLS, biPLS e siPLS na quantificação de BMSR e BESR em mistura com diesel</i> .	83
4.3.4	<i>Modelos iPLS, biPLS e siPLS na quantificação de BMA e BEA em mistura com diesel</i>	84
4.3.5	<i>Modelos iPLS, biPLS e siPLS na quantificação de BMPM e BEPM em mistura com diesel</i> .	88
5	CONCLUSÕES	95
	REFERÊNCIAS BIBLIOGRÁFICAS	97

LISTA DE TABELAS

Tabela 1.	Exemplo de Tabela de Exclusão obtido pelo algoritmo biPLS para uma divisão de região espectral em 20 intervalos.	35
Tabela 2.	Exemplo de resultados de saída fornecido pelo algoritmo siPLS para uma combinação entre 2 intervalos	37
Tabela 3.	Número de amostras nos conjuntos de calibração e previsão, número de VLs, variância capturada e valores de exatidão para os modelos PLS global.	60
Tabela 4.	Valores de $t_{crítico}$ e $t_{calculado}$ para avaliação da presença de erro sistemático dos modelos PLS.	70
Tabela 5.	Comparação de resultados de modelos PLS global que empregaram a Espectrometria no Infravermelho Médio	71
Tabela 6.	Valores de RMSEP, número de variáveis e intervalos obtidos a partir dos modelos iPLS para quantificação de BMS em mistura com diesel	72
Tabela 7.	Valores de RMSEP, número de variáveis e intervalos obtidos a partir dos modelos biPLS para quantificação de BMS em mistura com diesel	74
Tabela 8.	Valores de RMSECV para cada modelo siPLS avaliado na quantificação de BMS em mistura com diesel.	75
Tabela 9.	Valores de RMSEP, número de variáveis e intervalos obtidos a partir dos modelos siPLS para quantificação de BMS em mistura com diesel.	76
Tabela 10.	Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMS em mistura com diesel.	77
Tabela 11.	Parâmetros de exatidão e verificação da presença de erro sistemático do modelo iPLS35 na quantificação de BMS	78
Tabela 12.	Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BES em mistura com diesel	80
Tabela 13.	Parâmetros de exatidão e verificação da presença de erro sistemático do modelo si2PLS25 na quantificação de BES	81
Tabela 14.	Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMSR e BESR em mistura com diesel	83
Tabela 15.	Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMA e BEA em mistura com diesel	84

Tabela 16.	Parâmetros de exatidão e verificação da presença de erro sistemático dos modelos si2PLS35 e si3PLS35 na quantificação de BMA e BEA, respectivamente	86
Tabela 17.	Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMPM e BEPM em mistura com diesel	89
Tabela 18.	Parâmetros de exatidão e verificação da presença de erro sistemático do modelo si3PLS30 na quantificação de BEA	90
Tabela 19.	Resultados de diferentes técnicas e métodos quimiométricos obtidos na quantificação de biodiesel em misturas com diesel	92

LISTA DE FIGURAS

Figura 1.	Matérias-primas utilizadas na produção de biodiesel no território brasileiro	19
Figura 2.	Índice de não-conformidade com relação a natureza do diesel	21
Figura 3.	Representação da ATR	24
Figura 4.	Organização dos dados para a construção da matriz X	26
Figura 5.	Representação da decomposição da matriz X nas matrizes de escores (T), pesos (P) e resíduos (E)	27
Figura 6.	Representação de duas Componentes Principais	28
Figura 7.	Rotação no eixo da PC (Variável Latente)	30
Figura 8.	Gráfico do modelo iPLS do número de intervalos versus RMSECV. A linha tracejada é o RMSECV para o modelo global e os números em itálico são as VLs indicadas no modelo por intervalos	33
Figura 9.	(a) espectrômetro FT-MIR (Perkin Elmer, modelo Spectrum Two) e (b)ATR horizontal (Pike Technologies)	50
Figura 10.	Gráfico de <i>Leverage</i> versus resíduos de Student na identificação de amostras anômalas	51
Figura 11.	(a) Espectros MIR dos biodieseis puros e do diesel ($600-4000\text{ cm}^{-1}$); (b) Destaque para a região de estiramento da carbonila ($1700 - 1780\text{ cm}^{-1}$)	55
Figura 12.	Espectros MIR das amostras de biodieseis metílicos dos conjuntos de calibração e previsão obtidos na faixa de trabalho de 0,25 a 30% (v/v) de biodiesel em mistura com diesel. (a) BMS; (b) BMSR; (c) BMA; (d) BMPM	56
Figura 13.	Espectros MIR das amostras de biodieseis etílicos dos conjuntos de calibração e previsão obtidos na faixa de trabalho de 0,25 a 30% (v/v) de biodiesel em mistura com diesel. (e) BES; (f) BESR; (g) BEA e (h) BEPM	57
Figura 14.	Gráfico de <i>leverage</i> versus Q <i>residuals</i> com 95% de confiança para os modelos PLS global dos biodieseis metílicos: (a) BMS, (b) BMSR, (c) BMA, (d) BMPM	58
Figura 15.	Gráfico de <i>leverage</i> versus Q <i>residuals</i> , com 95% de confiança, para os modelos PLS global dos biodieseis etílicos: (e) BES, (f) BESR, (g) BEA e (h) BEPM	59
Figura 16.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BMS	62
Figura 17.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BES.	63
Figura 18.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BMSR.	64

Figura 19.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BESR	65
Figura 20.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BMA	66
Figura 21.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BEA	67
Figura 22.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BPPM	68
Figura 23.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BEPM	69
Figura 24.	Espectro MIR dividido em 35 intervalos para a determinação de BMS em mistura com diesel. As barras correspondem aos valores de RMSECV para cada intervalo e a linha tracejada para o modelo PLS global. Os números internos apresentados no interior de cada intervalo correspondem ao número de variáveis latentes sugerido pelo algoritmo	73
Figura 25.	Intervalo selecionado ($1166 - 1327 \text{ cm}^{-1}$) para construção do modelo PLS para quantificação de BMS em mistura com diesel obtida a partir do algoritmo biPLS15.	75
Figura 26.	Intervalos selecionados ($875-972$, $1166-1263$, $1747-1844 \text{ cm}^{-1}$) para construção do modelo PLS para quantificação de BMS em mistura com diesel obtida a partir do algoritmo si3PLS25	77
Figura 27.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo iPLS35 na quantificação de BMS	79
Figura 28.	Intervalos selecionados ($1069-1166$ e $1748-1845 \text{ cm}^{-1}$) para construção do modelo PLS para quantificação de BES em mistura com diesel obtidas a partir do algoritmo si2PLS25	81
Figura 29.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si2PLS25 na quantificação de BES	82
Figura 30.	Intervalos selecionados ($1166-1234$ e $1235-1304 \text{ cm}^{-1}$) para construção do modelo PLS para quantificação de BMA em mistura com diesel obtidas a partir do algoritmo si2PLS35.	85
Figura 31.	Intervalos selecionados ($1166-1235$, $1444-1513$ e $1720-1789 \text{ cm}^{-1}$) para construção do modelo PLS para quantificação de BEA em mistura com diesel obtidas a partir do algoritmo si3PLS35.	86
Figura 32.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si2PLS35 na quantificação de BMA	87
Figura 33.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si3PLS35 na quantificação de BEA	88

Figura 34.	Intervalos selecionados (1004-1085, 1166-1247 e 1248-1329 cm^{-1}) para construção do modelo PLS para quantificação de BEPM em mistura com diesel obtidas a partir do algoritmo si3PLS30	90
Figura 35.	(a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si3PLS30 na quantificação de BEPM.	91

1. INTRODUÇÃO

1. INTRODUÇÃO

1.1. Biodiesel e matérias-primas

A maioria do consumo de energia do mundo vem do petróleo, carvão e gás natural, mas estas fontes são limitadas por causa de sua origem fóssil. Portanto, a busca por fontes de energia alternativas a partir de fontes renováveis é essencial. Nesta perspectiva, o biodiesel vem se destacando no Brasil e em muitos países devido principalmente a questões econômicas, sociais e ambientais (PORTE et al.; 2010; FAZAL; HASEEB; MASJUKI, 2011). Questões econômicas, como a redução de importações de óleo diesel; questões sociais, como incentivo a agricultura familiar ou dos pequenos produtores e, questão ambiental devido a crescente atenção para o impacto ambiental causado pelo uso de recursos energéticos fósseis e o esforço para diminuir gases poluentes atmosféricos (ALVES; POPPI, 2013).

O biodiesel pode ser definido como uma mistura de mono-alquil ésteres derivado de fontes renováveis, como óleos vegetais e gorduras animais, que pode ser obtido através da reação de transesterificação, no qual ocorre a transformação de triglicerídeos em moléculas menores de ésteres (ENCINAR et al., 2002).

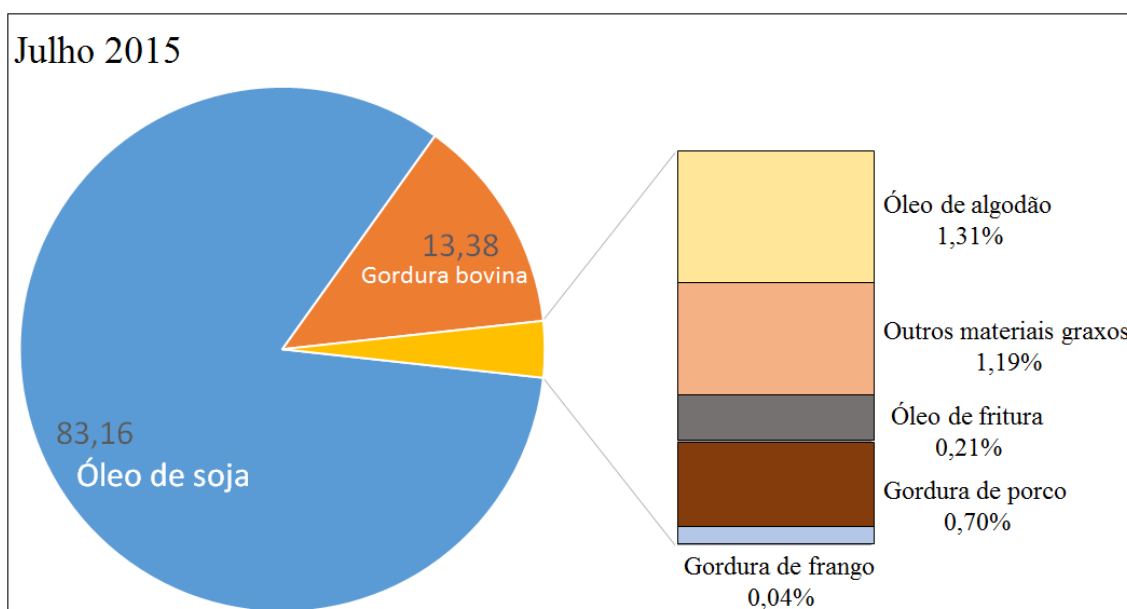
Este biocombustível possui algumas características que representam vantagens sobre os combustíveis derivados do petróleo, tais como, livre de enxofre e de compostos aromáticos, alto número de cetano, teor médio de oxigênio, maior ponto de fulgor, menor índice de emissão de gases, além de ser proveniente de fontes renováveis (FERRARI; OLIVEIRA; SCABIO, 2005).

A maior parte do biodiesel atualmente produzido no mundo deriva do óleo de soja, utilizando metanol e catalisador alcalino, porém, todos os óleos vegetais, enquadrados na categoria de óleos fixos ou triglicerídeos, podem ser transformados em biodiesel. Fatores como a geografia, o clima e a economia determinam o óleo vegetal de maior interesse para uso potencial nos biocombustíveis. Assim, nos Estados Unidos, por exemplo, o óleo de soja é considerado como matéria-prima primordial e, nos países tropicais, é o óleo de palma. Na Alemanha, o óleo de colza é utilizado na produção de biodiesel, o qual é distribuído de forma pura, isento de qualquer mistura ou aditivação (FERRARI; OLIVEIRA; SCABIO, 2005).

No Brasil, os óleos vegetais mais comuns, empregados na produção de biodiesel são soja, milho, amendoim, algodão, babaçu e palma. Porém, a soja se destaca no cenário nacional devido à sua alta produtividade (ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE

ÓLEOS VEGETAIS - ABIOVE, 2015). Ela apresenta teor de óleo compreendido entre 20 e 22%, alto teor de proteína (40 a 42% nas variedades difundidas) e fácil adaptação aos diversos tipos de clima e fotoperíodo, características essas que levaram à formação de um complexo industrial destinado ao seu processamento (MACIEL et al., 2005). Por isso, a soja é a matéria-prima mais empregada para a produção de biodiesel no Brasil (83,16%) seguidas das matérias-primas gordura bovina (13,38%) e óleo de algodão (1,31%) (AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS - ANP, 2015a), conforme demonstra a Figura 1.

Figura 1 – Matérias-primas utilizadas na produção de biodiesel no território brasileiro



Fonte: ANP (2015a).

No entanto, apesar do óleo de soja ser a principal matéria-prima para produção de biodiesel este apresenta como desvantagem ser competitivo com o mercado de alimentos, além de representar cerca de 70% do custo de produção do biodiesel, o que torna necessário algum subsídio, como por exemplo, isenções de impostos e incentivos fiscais para sua produção (HUANG; CHANG, 2010; PADULA et al., 2012; FAROOQ; RAMLI; SUBBARAO, 2013). Neste sentido a busca por outras matérias-primas tem se destacado, como por exemplo, óleos residuais e oleaginosas não comestíveis. A utilização de óleos residuais, como por exemplo, óleo de fritura usado, torna-se uma alternativa para a diminuição dos custos de produção. Além disso, evita-se a competição com a indústria de alimentos e resolve-se os problemas associados com a eliminação deste resíduo (FAROOQ; RAMLI; SUBBARAO, 2013). Ou seja, dando uma

destinação adequada a este resíduo resolve-se um problema ambiental, que, em geral, é descartado inadequadamente, impactando o solo e o lençol freático, e consequentemente a biota desses sistemas. Estudos indicam que cada litro de óleo despejado no esgoto pode poluir mais de 25 mil litros de água, e que o custo de tratamento de efluentes nas unidades de tratamento de esgotos podem aumentar em cerca de 40% devido a presença de óleo residual (COSTA NETO et al., 2000; COMPANHIA DE SANEAMENTO BÁSICO DO ESTADO DE SÃO PAULO - SABESP, 2010).

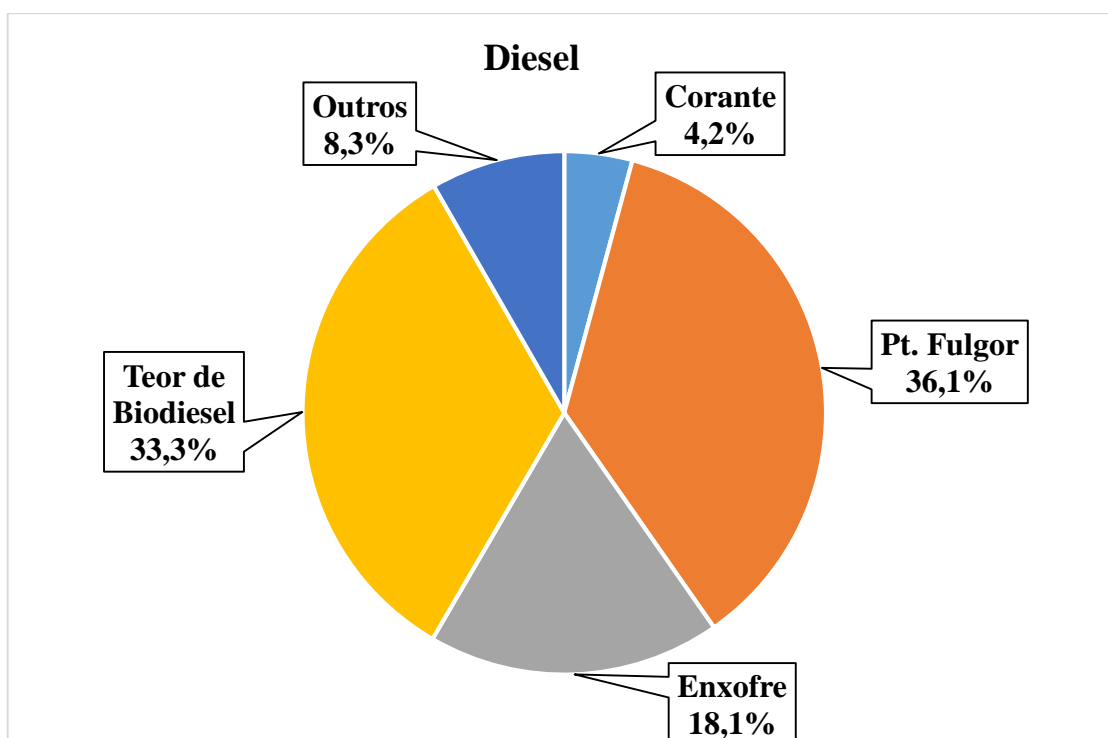
Dentre as oleaginosas não comestíveis destaca-se o óleo de algodão e pinhão manso. A utilização de óleo de algodão para produção de biodiesel no Brasil corresponde a terceira fonte de matéria-prima, sendo sua atividade principal a produção de fibras têxteis naturais e sintéticas. A torta resultante do processo de esmagamento das sementes é empregada na produção de ração animal por ser proteica e energética. Por outro lado, o emprego do óleo de algodão para fins culinários fica inviabilizado devido a presença de gossipol (pigmento tóxico presente nas sementes) (CHIAVEGATO et al., 2009). Portanto, por se tratar de um subproduto dessa cadeia produtiva o óleo de algodão pode ser destinado a produção de biodiesel, não sendo seu custo tão afetado pelo mercado alimentício como no caso dos óleos de soja, canola, milho, girassol, dentre outras oleaginosas. Outra matéria-prima em ascensão na produção de biodiesel no Brasil é o pinhão manso (*Jatropha curcas* L.), que como no caso do algodão não se utiliza como alimento, mas que tem a capacidade de fornecer até 50% de óleo em suas sementes mostrando-se uma espécie promissora na produção de biodiesel. O pinhão manso apresenta outras vantagens, como por exemplo, baixo custo de produção, capacidade de produzir em solos pouco férteis e arenosos, além de alta produtividade e excelente qualidade do óleo (KOH; GHAZI, 2011; OLIVEIRA et al., 2010). Por possuir um elevado teor de ácidos graxos saturados e um baixo teor de ácidos graxos insaturados, o biodiesel produzido a partir desta oleaginosa apresenta boa estabilidade à oxidação e menor formação de hidroperóxidos, resultando na produção de um biodiesel de excelente qualidade (KNOTHE et al., 2006; SARIN et al., 2010). Assim, a produção de biodiesel a partir de fontes alternativas é uma realidade e o controle de qualidade dos produtos obtidos se fazem necessários, principalmente no que tange ao teor de biodiesel em misturas com diesel.

1.2. Teor, não conformidades e normas

Devido as propriedades físico-químicas do biodiesel serem semelhantes ao diesel, as misturas destes são usadas como combustíveis, as quais são denominadas de BX, onde X indica

a porcentagem volumétrica de biodiesel na mistura. Desde novembro de 2014, o Brasil obriga a adição de 7,0% (v/v) de biodiesel ao diesel, mistura denominada de B7 (ANP, 2014). Uma das vantagens no uso desta mistura é a possibilidade de aproveitamento da rede de distribuição de combustível já existente, e o combustível pode ser utilizado nos motores atuais sem qualquer modificação em uma porcentagem de até 20,0% (v/v) (MOFIJUR et al., 2013). No entanto, em análises mensais descobriu-se que parte significativa das misturas biodiesel/diesel vendidas no país estavam fora das especificações de qualidade, onde 33,3% das não-conformidades estavam relacionadas ao teor de biodiesel nas misturas (Figura 2). As outras são devido aos corantes, aspectos, ponto de fulgor, teor de enxofre e outros (ANP, 2015b).

Figura 2 – Índice de não-conformidade com relação a natureza do diesel



Fonte: ANP (2015b).

Com a publicação desta resolução e os altos índices de não conformidades em relação ao teor de biodiesel em mistura com diesel torna-se necessário intensificar os estudos de desenvolvimento de metodologias capazes de quantificar de maneira satisfatória o teor de biodiesel na mistura binária com diesel em concentrações principalmente em torno dos 7% (v/v).

Existem três métodos padrão para quantificar biodiesel em misturas com diesel: Norma Européia EN 14078 (2004) (do inglês, European Standard), Associação Brasileira de Normas

Técnicas (ABNT) NBR 15568 (2008) e Sociedade Americana de Testes e Materiais (ASTM) D7371 (2007) (do inglês, American Society for Testing and Materials). EN 14078 utiliza a Espectrometria na região do Infravermelho Médio (MIR) e o método baseia na obtenção de uma curva de calibração (univariada) e emprega o sinal de absorção associado à vibração de estiramento da carbonila (C=O) presente nos ésteres ($\sim 1745\text{ cm}^{-1}$). A ASTM D7371 e ABNT NBR15568 também utilizam a Espectrometria no Infravermelho Médio, através de medições de transmitância ou Reflectância Total Atenuada (ATR) aliada ao método quimiométrico por Quadrados Mínimos Parciais (PLS). No entanto, devido à dificuldade na obtenção da exatidão necessária a utilização de dois ou três modelos analíticos é recomendada. O método ASTM sugere o desenvolvimento de três modelos dependendo da faixa de concentração, ou seja, 0-10% (v/v), 10-30% (v/v) e 30-100% (v/v). A reprodutibilidade é especificada de acordo com o teor de biodiesel na amostra e varia 0,76-1,66% (v/v) para amostras com 1% e 20% (v/v) de biodiesel, respectivamente (ASTM, 2007). O método ABNT NBR sugere o desenvolvimento de dois modelos analíticos um para intervalo de 0-8% (v/v) e o outro para o intervalo de 8-30% (v/v). Esses modelos devem ter Erros Quadráticos Médio de Previsão (RMSEP) que não sejam maior que 0,1% (v/v) e 1% (v/v), respectivamente. Além disso, deve ser selecionado as regiões do espectro características das ligações C-O ($1100\text{-}1200\text{ cm}^{-1}$) e C=O ($1735\text{-}1750\text{ cm}^{-1}$) presente nos ésteres (ABNT, 2008).

Verifica-se que são encontrados diferentes procedimentos para análise de misturas biodiesel/diesel, mas todas utilizam a Espectrometria no Infravermelho Médio. Além destas normas também é recomendado seguir as orientações da norma ASTM E1655-05 (2012). Esta norma descreve um guia prático de procedimentos para coleta e tratamento de dados na quantificação de analitos sempre quando se utiliza a Espectrometria no Infravermelho. Além disso, esta norma também descreve critérios para validação do modelo de calibração.

1.3. Espectrometria no Infravermelho Médio

A Espectrometria no Infravermelho Médio (MIR) refere-se a transições vibracionais e rotacionais nas ligações das moléculas. A região de $4.000\text{ a }200\text{ cm}^{-1}$ é considerada a mais importante para a caracterização de compostos orgânicos (BARBOSA, 2007). Como cada tipo de ligação tem sua frequência característica de vibração, e como duas ligações idênticas em diferentes compostos estão em dois ambientes diferentes, os espectros de duas moléculas diferentes nunca serão exatamente idênticos. É essa característica dos espectros no infravermelho que permitem que eles sejam utilizados para a comparação entre duas substâncias

para descobrir se são as mesmas ou diferentes (PAVIA et al., 2010). Em específico, a região conhecida como impressão digital, a faixa de 1.200 a 600 cm^{-1} , é muito útil na análise dos espectros, uma vez que pequenas diferenças na estrutura e constituição das moléculas mudam significativamente o perfil nessa região. Consequentemente, uma grande semelhança entre dois espectros nessa região constitui forte evidência da identidade dos compostos que os produziram (HOLLER; SKOOG; CROUCH, 2009).

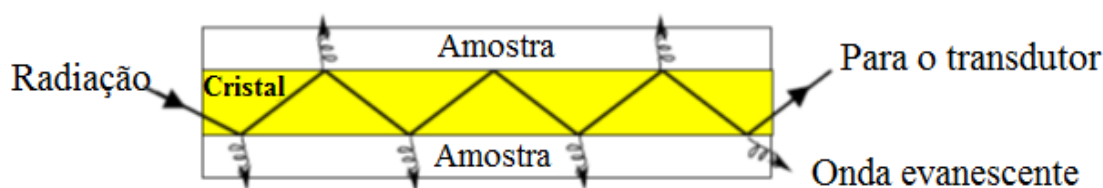
A primeira vantagem da Espectrometria MIR é que trata-se de um método quase universal; muitas moléculas, sejam sólidos, líquidos, gases, semi-sólidos, pós, polímeros, substâncias orgânicas, inorgânicas, materiais biológicos, substâncias puras e misturas apresentam intensas bandas de absorção nessa região. A segunda vantagem é que os espectros são ricos em informação, uma vez que as bandas fornecem as estruturas das moléculas em uma amostra, enquanto que suas intensidades fornecem as concentrações, e as larguras das bandas são sensíveis a matriz química da amostra, incluindo pH e ligações de hidrogênio. A terceira vantagem consiste na rapidez e facilidade de se obter um espectro de infravermelho. Embora o tempo de medição varie conforme a natureza e técnica de amostragem, em geral os espectros são medidos em cinco minutos ou menos. A quarta vantagem está no custo relativamente baixo em relação as outras técnicas instrumentais, tais como Ressonância Magnética Nuclear (NMR), Cromatografia Gasosa – Espectrometria de Massa (GC-MS) e Cromatografia Líquida – Espectrometria de Massa (LC-MS) (SMITH, 2011).

Além disso, no MIR pode-se empregar o acessório de Reflexão Total Atenuada (ATR) em que os espectros de absorção são obtidos rapidamente em uma variedade de tipos de amostras com um mínimo de preparação. Assim, linhas, fios, tecidos, fibras, pós, pastas, suspensões, polímeros, borrachas e outros materiais podem ser analisados pressionando as amostras sobre o cristal ATR. Para amostras líquidas, o líquido pode ser adicionado diretamente no acessório ATR, como por exemplo, amostras de biodiesel em diesel. Soluções aquosas podem também ser usadas, desde que o cristal não seja solúvel em água (SKOOG; HOLLER; NIEMAN, 2002). Utilizando-se este tipo de acessório a amostra pode ser preservada, ou seja, utilizando-se o ATR a técnica tem a vantagem de ser não destrutiva.

A ATR baseia-se no fato de que quando um feixe de radiação passa de um meio mais denso (cristal de ATR) para um menos denso (amostra) ocorre uma reflexão. A fração do feixe incidente que é refletida aumenta com o ângulo de incidência; além de certo ângulo crítico, a reflexão é total. No processo de reflexão, o feixe se comporta como se penetrasse um pouco no meio menos denso antes que a reflexão ocorra. A profundidade de penetração, que varia de uma fração até vários comprimentos de onda, depende do comprimento de onda da radiação

incidente, dos índices de refração dos dois materiais e do ângulo do feixe incidente em relação à interface. A radiação penetrante é chamada onda evanescente. Se o meio menos denso absorve essa onda evanescente, ocorre atenuação do feixe nos comprimentos de onda das bandas de absorção (SKOOG; HOLLER; NIEMAN, 2002). Esse fenômeno é conhecido como Refletância Total Atenuada (ATR), representado na Figura 3.

Figura 3 – Representação da ATR



Fonte: Borges Neto (2005).

Os espectros de ATR são semelhantes, mas não iguais aos espectros comuns de absorção. Em geral, enquanto as mesmas bandas são observadas, suas intensidades relativas diferem. As absorções, embora dependam do ângulo de incidência, são independentes da espessura da amostra, uma vez que a radiação penetra apenas alguns micrômetros na mesma (SKOOG; HOLLER; NIEMAN, 2002).

Assim, devido a estas vantagens a Espectrometria MIR tem sido aplicada no controle de qualidade de vários produtos, como alimentos (XU et al., 2012; FERRAND et al., 2011), bebidas (FRIEDEL; PATZ; DIETRICH, 2013), fármacos (PIANTAVINI et al., 2014), inclusive combustíveis em geral (ANZANELLO et al., 2015; MAZIVILA et al., 2015a; SOUZA et al., 2015). No entanto, esta aplicabilidade só aumentou devido ao desenvolvimento dos métodos quimiométricos, pois os dados obtidos a partir dos espectros MIR são complexos, os quais visualmente, isto é, sem qualquer tratamento matemático e estatístico, é difícil extrair informação útil.

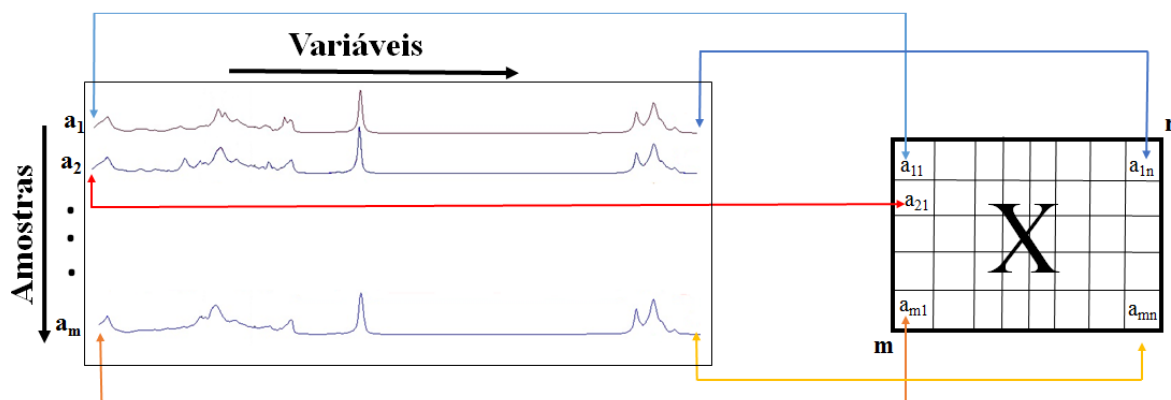
1.4. Quimiometria

A quimiometria é definida como a aplicação de métodos matemáticos, estatísticos e de lógica formal para o tratamento de dados químicos (MASSART et al., 1998). Ela é dividida de acordo com a área em que é aplicada, sendo as principais: processamento de sinais analíticos, planejamento e otimização de experimentos, reconhecimento de padrões e classificação de

dados, calibração multivariada, monitoramento e modelagem de processos multivariados, métodos de inteligência artificial, entre outros (VANDEGINSTE et al., 1998). Neste trabalho, foi dado foco na área de calibração multivariada, que tem por objetivo a construção de um modelo matemático que correlacione a resposta de um dado instrumento analítico com alguma(s) propriedade(s) de interesse, de um conjunto de amostras conhecidas, sendo essa etapa denominada de calibração. A partir do modelo construído torna-se então possível a quantificação dessa(s) propriedade(s) em novas amostras onde elas são desconhecidas, etapa denominada de validação. Dentre os métodos de calibração existentes, sem dúvida, os mais difundidos são ainda os métodos de calibração univariada que também são conhecidos como calibração de ordem zero, ou seja, tem-se apenas uma medida instrumental para cada uma das amostras de calibração, isto é, para cada amostra tem-se apenas um escalar. No entanto, a aplicação da calibração univariada é restrita, visto que, quando a amostra não é livre de interferentes e a medida é realizada diretamente por uma metodologia instrumental, isso provavelmente provocará desvios na determinação da propriedade de interesse e a aplicação deste método de calibração torna-se inviável (CHUI; ZUCCHINI; LICHTIG, 2001).

Em calibração multivariada, mais de uma resposta instrumental é relacionada com a propriedade de interesse. Esses métodos de calibração possibilitam a análise mesmo na presença de interferentes, desde que esses interferentes estejam presentes nas amostras utilizadas para a construção do modelo de calibração. Outras possibilidades apresentadas por esse tipo de calibração são determinações simultâneas e análises mesmo sem resolução. Isso faz com que os modelos de calibração multivariada sejam uma alternativa quando os métodos univariados não são aplicáveis. Neste tipo de calibração a resposta instrumental é representada na forma de matriz, enquanto a propriedade de interesse, determinada por uma metodologia padrão, é representada por um vetor (MARTENS; NAES, 1996). A Figura 4 ilustra como uma matriz de dados \mathbf{X} de dimensão $m \times n$, ou seja, m objetos (espectros) e n variáveis (números de onda), pode ser construída a partir de um vetor de respostas instrumentais.

Figura 4 – Organização dos dados para a construção da matriz X



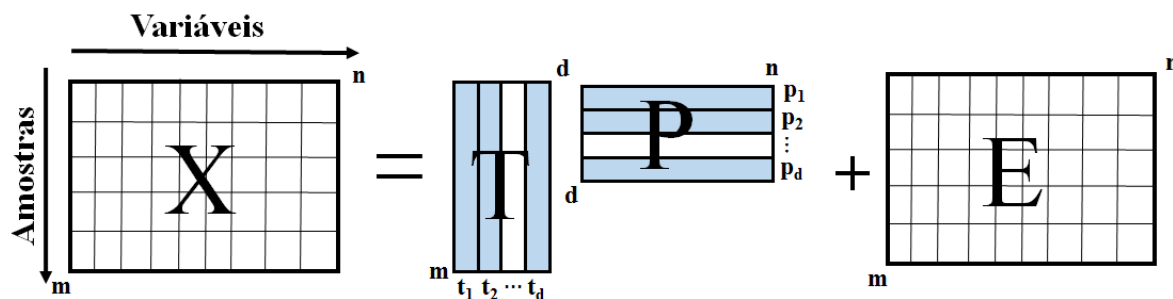
Fonte: O autor.

Entre os métodos de calibração multivariada pode-se destacar os métodos de regressão, tais como, Regressão por Componentes Principais (PCR), Quadrados Mínimos Parciais (PLS) e a Regressão Linear Múltipla (MLR) (MASSART et al., 1998). Nesse trabalho, usou-se a regressão por PLS, porém é necessário introduzir o conceito de Análise de Componentes Principais (PCA), uma vez que nessa regressão, utiliza-se a PCA como a técnica de ortogonalização baseada em uma mudança da base vetorial.

1.4.1. Análise de Componentes Principais (PCA)

Uma das principais formas de análise de dados multivariados é a Análise por Componentes Principais (PCA). Este método tem por objetivo a redução da dimensionalidade do conjunto de dados original, preservando a maior quantidade de informação (variância) possível, que neste caso, é válida somente para transformações lineares. Essa redução é obtida por meio do estabelecimento de novas variáveis ortogonais entre si, denominadas Componentes Principais (PCs) (CORREIA; FERREIRA, 2007). Assim, a matriz X é decomposta em duas matrizes menores de escores (T) e de pesos (P) e mais uma matriz de resíduos (E) que indica a parte que não foi modelada, como representado na Figura 5.

Figura 5 – Representação da decomposição da matriz **X** nas matrizes de escores (**T**), pesos (**P**) e resíduos (**E**)



Fonte: Adaptado de Otto(1999).

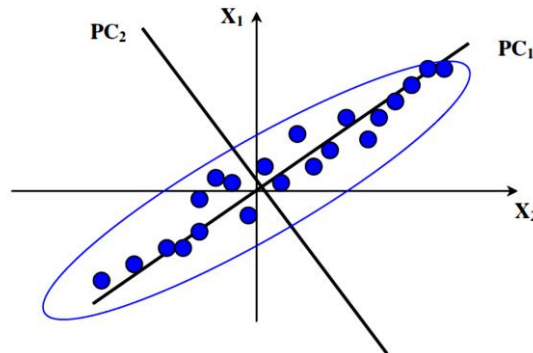
Neste caso, **X** é a matriz original constituída de “m” linhas (amostras) e “n” colunas (variáveis); **T** é a matriz de escores com “m” linhas e “d” colunas (número de componentes principais); **P** é a matriz de pesos com dimensões “d” e “n” e **E** é a matriz de resíduos com “m” linhas e “n” colunas. Assim, a equação básica para PCA é representada pelas Equações 1 ou 2 (WOLD; ESBERSEN; GELADI, 1987).

$$X = t_1 p_1^T + t_2 p_2^T + \dots + t_d p_d^T + E \quad (1)$$

$$X = TP^T + E \quad (2)$$

As componentes principais são determinadas baseadas no critério de variância máxima. Cada componente principal subsequente descreve um máximo de variância contida nos dados, sendo que a primeira componente principal determina a maior variância dos dados. Na segunda componente principal, há menos informação do que na primeira e, assim, sucessivamente. A Figura 6 representa graficamente duas componentes principais no caso de 2 variáveis (x_1 e x_2) em que a maior variância dos dados está sendo explicada pela primeira componente principal (PC1).

Figura 6 – Representação de duas Componentes Principais



Fonte: Guimarães (2014).

O número de componentes principais a ser utilizado no modelo PCA é determinado pela percentagem da variância explicada. Assim, seleciona-se um número de componentes de tal maneira que a maior porcentagem da variância presente no conjunto de dados originais seja capturada (WOLD; ESBENSEN; GELADI, 1987).

A análise de PCA é feita geralmente usando o algoritmo *Nonlinear Iterative Partial Least Squares* (NIPALS). NIPALS é um método comumente utilizado para o cálculo das componentes principais de um conjunto de dados, onde os vetores dos pesos e dos escores são calculados iterativamente, um de cada vez, em ordem decrescente de importância. O processo iterativo, para a PC1, é inicializado com uma primeira estimativa de escores, que pode ser a coluna de \mathbf{X} que tem maior variância.

Utilizando estes escores, calculam-se os pesos através da Equação 3.

$$\mathbf{p} = \frac{\mathbf{t}\mathbf{X}}{\mathbf{t}^T\mathbf{t}} \quad (3)$$

Os pesos são normalizados para comprimento igual a 1. Após isso, os escores são calculados pela Equação 4.

$$\mathbf{t} = \frac{\mathbf{X}\mathbf{p}}{\mathbf{p}^T\mathbf{p}} \quad (4)$$

Esses valores de escores são comparados com os anteriores e se forem diferentes (dentro de um critério pré-estabelecido), os pesos são novamente calculados como mostrado acima. Esse processo continua até que os escores sejam semelhantes ou certos números de iterações tenham sido realizadas. Após a convergência, o produto calculado a partir da Equação 2 é subtraído de \mathbf{X} , obtendo-se o resíduo \mathbf{E} através da Equação 5.

$$\mathbf{E} = \mathbf{X} - \mathbf{t}\mathbf{p}^T \quad (5)$$

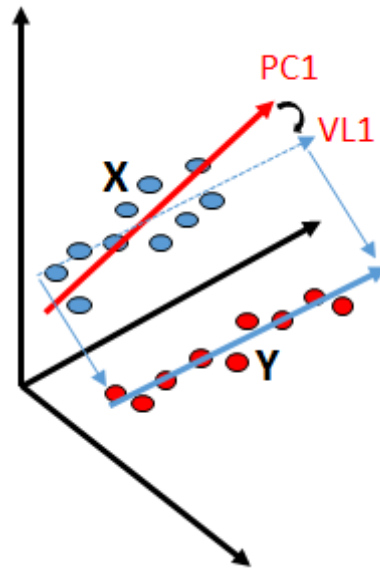
O processo continua para a próxima PC, substituindo \mathbf{E} por \mathbf{X} .

1.4.2. Regressão por Quadrados Mínimos Parciais (PLS)

O PLS é o método de regressão mais utilizado para a construção de modelos de calibração multivariada. Este método não requer um conhecimento exato de todos os componentes presentes nas amostras, podendo realizar a previsão de amostras mesmo na presença de interferentes. Porém estes interferentes também devem estar presentes por ocasião da construção do modelo (BEEBE; KOWALSKI, 1987).

A decomposição da matriz \mathbf{X} realizada pela PCA é feita de forma independente da matriz \mathbf{Y} . Enquanto que para o método de regressão PLS a informação de \mathbf{Y} é incorporada, de forma que cada PC do modelo sofre uma pequena modificação para buscar a máxima covariância entre \mathbf{X} e \mathbf{Y} (variável latente). Assim, o modelo PLS é obtido através de um processo iterativo, no qual se otimiza ao mesmo tempo a projeção das amostras sobre o(s) peso(s), para a determinação dos escores, e o ajuste por uma função linear dos escores da matriz \mathbf{X} aos escores da matriz \mathbf{Y} de modo a minimizar os desvios. Essa otimização simultânea ocasiona pequenas mudanças nas direções dos pesos, de modo que, rigorosamente eles perdem a ortogonalidade, levando a pequenas redundâncias de informação. Porém, são essas pequenas redundâncias que otimizam a relação linear entre os escores, e estas distorções da ortogonalidade entre as componentes principais no PLS fazem com que não sejam mais componentes principais (que são ortogonais) e, sim, Variáveis Latentes (VLs) (VANDEGINSTE et al., 1998; BEEBE; KOWALSKI, 1987), conforme representação observada na Figura 7.

Figura 7 – Rotação no eixo da PC (Variável Latente)



Fonte: O autor.

A PLSR estende o conceito do modelo inverso (propriedade como função da medida instrumental) trocando as variáveis originais por um subconjunto truncado das VLs dos dados originais. Considerando um caso geral para a determinação de mais de uma espécie de interesse, logo \mathbf{Y} é uma matriz de dimensão $(m \times z)$, em que m é o número de amostras e z é o número de colunas de \mathbf{Y} , tem-se a decomposição de ambas as matrizes \mathbf{X} de dimensão $(m \times n)$ e \mathbf{Y} em suas matrizes de escores e pesos, conforme Equações 6 e 7.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E}_X \quad (6)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{E}_Y \quad (7)$$

Onde, \mathbf{X} é a matriz de dados (medida instrumental), \mathbf{Y} é a matriz de resposta (concentração, por exemplo), \mathbf{T} e \mathbf{U} são os escores para as duas matrizes de dados, \mathbf{P} e \mathbf{Q} são os respectivos pesos, h é número de VLs, \mathbf{E}_X e \mathbf{E}_Y são os respectivos resíduos, ou seja, as matrizes que contem a parte não modelada (MILLER; MILLER; 2010).

Uma relação linear é, então, estabelecida entre os escores de \mathbf{X} e os escores de \mathbf{Y} para cada VL ($h=1,2, \dots, A$) conforme Equação 8 (OTTO, 1999).

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h \quad (8)$$

Em que, \mathbf{b}_h é o vetor de coeficientes de regressão do modelo linear para cada VL, obtido através da Equação 9.

$$\mathbf{b}_h = \frac{\mathbf{u}_h^T \mathbf{t}_h}{\mathbf{t}_h^T \mathbf{t}_h} \quad (9)$$

O melhor modelo deverá minimizar as matrizes de resíduos \mathbf{E}_X e \mathbf{E}_Y e ao mesmo tempo, obter uma relação linear entre \mathbf{t} e \mathbf{u} .

A decomposição pode ser realizada da mesma maneira como demonstrado para PCA, por exemplo, empregando o algoritmo NIPALS. No final do processo a variância explicada pela primeira VL será maior que a variância explicada pela segunda VL e a terceira VL explicará uma variância menor que a segunda VL, e assim sucessivamente até que o número de VLs seja definido e o algoritmo, geralmente, convergido rapidamente (GELADI; KOWALSKI, 1986).

Para a determinação do número correto de VL o método mais utilizado é o de validação cruzada, o qual se baseia na habilidade de previsão de um modelo construído por parte de um conjunto de dados seguido pela previsão do restante do conjunto de dados, que é realizada pelo modelo construído. A validação cruzada pode ser realizada em blocos, ou seja, um número determinado de amostras é deixado de fora no processo de construção do modelo e a seguir essas amostras são previstas pelo modelo construído, ou ainda por “*leave-one-out*” (deixe uma de fora por vez), em que uma amostra é deixada de fora no processo de construção do modelo e a seguir essa amostra é prevista pelo modelo construído. Em ambos os casos, o processo é repetido até que todas as amostras tenham sido previstas. Assim, o Erro Quadrático Médio de Validação Cruzada (RMSECV) é calculado de acordo com a Equação 10 (GELADI; KOWALSKI, 1986; BRERETON, 2000).

$$\text{RMSECV} = \sqrt{\frac{\sum_{i=1}^{n-1} (\hat{y}_i - y_i)^2}{n}} \quad (10)$$

Onde, \hat{y}_i e y_i são, respectivamente, os valores previstos e de referência para a propriedade de interesse.

O cálculo é realizado para o número de componentes de 1 até A, e os resultados de RMSECV são mostrados em um gráfico de RMSECV contra o número de VLs. O

comportamento típico para esses gráficos é a observação de um mínimo ou um patamar, que indica a melhor dimensionalidade do modelo de regressão, ou seja, o melhor número de VLs que produziu um baixo erro de previsão sem perda significativa da variância dos dados. Assim, esta etapa de escolha das VLs é importante. Em um modelo PLS deve ser escolhido um número mínimo de VLs que explica o máximo da variância dos dados, minimizando a influência do ruído. Assim, o número ótimo de VLs a ser usado para descrever um modelo será o número de VLs para o qual um erro mínimo na previsão seja obtido (GEMPERLINE, 2006). No entanto, segundo Gomez-Carracedo et al. (2007), um modelo que contenha número de VLs inferior ao ideal resultará em subajuste e superior ao ideal em sobreajuste. A decisão sobre um número de VLs superior ao ideal ocorre com mais frequência ocasionando resultados satisfatórios para a previsão da propriedade de amostras do conjunto de calibração, porém, não adequados para a previsão de amostras de um conjunto validação ou previsão devido à informação não necessária incluída no modelo.

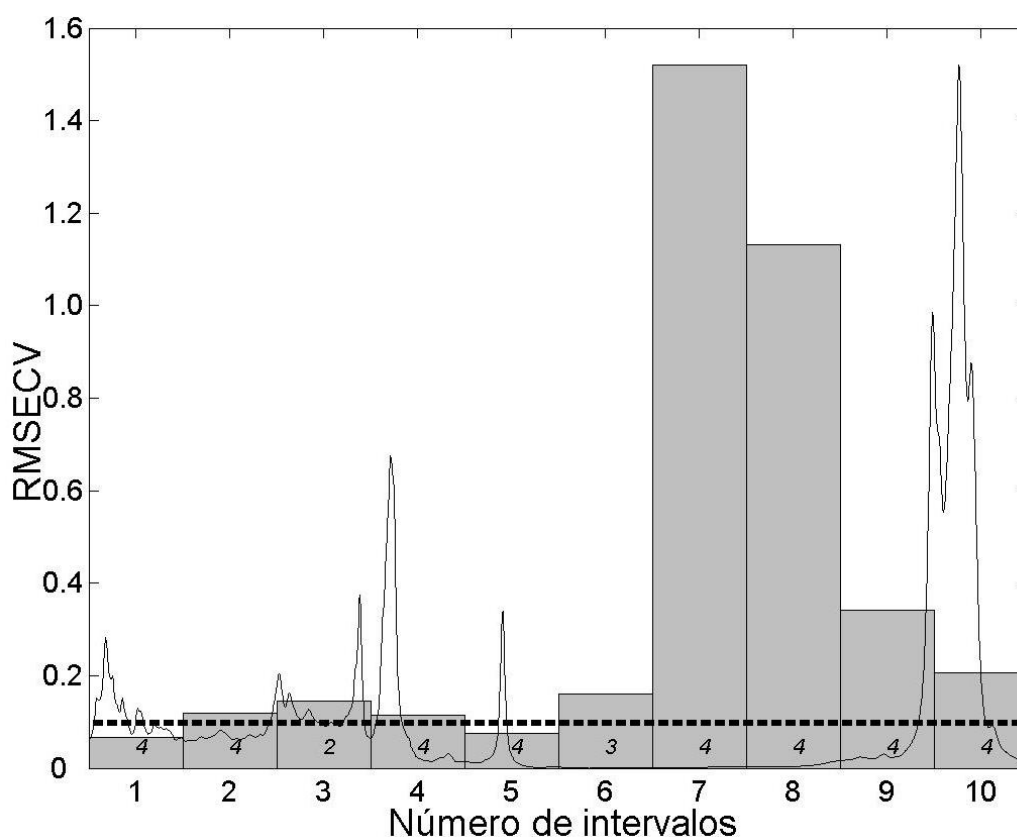
1.4.3. Seleção de variáveis por intervalos

Além do emprego de método de calibração multivariada por PLS existe a possibilidade de implementação de métodos de seleção de região espectral para melhorar significativamente o desempenho dos métodos de calibração de espectros totais. Esses métodos selecionam regiões específicas do espectro (um comprimento de onda ou um conjunto de comprimentos de onda) em que a colinearidade não é tão importante, enquanto gera modelos mais estáveis, robustos e mais simples de interpretar. Na prática, a filosofia está baseada na identificação de um subconjunto dos dados inteiros que produzirão erros de previsão mais baixos. Assim, em espectros vibracionais, por exemplo, os comprimentos de onda que apenas induzem a ruídos, informações irrelevantes ou não-linearidades, podem ser eliminados (FERNANDES et al., 2011; BALABIN; SMIRNOV, 2011).

Os métodos de seleção de variáveis existentes se diferem com relação ao procedimento realizado na seleção da região espectral. Dentre os métodos que estão sendo utilizados pode-se destacar o método de Quadrados Mínimos Parciais por Intervalos (iPLS – do inglês, Interval Partial Least Square) (NØRGAARD et al., 2000; JIANG et al., 2012; PIAVAVINI et al., 2014), o método de Eliminação de Variáveis não Informativas por Quadrados Mínimos Parciais (UVE-PLS – do inglês, Elimination of Uninformative Variables in Partial Least Square) (CENTNER; MASSART, 1996; HUANGA et al., 2013) e Algoritmo Genético (GA – do inglês, Genetic Algorithm) (COSTA FILHO; POPPI, 1999; FRIEDEL; PATZ; DIETRICH, 2013).

O algoritmo iPLS foi proposto por Nørgaard et al.(2000). O método iPLS é uma extensão iterativa para o PLS, que desenvolve modelos locais PLS em subintervalos equidistantes de toda a região do espectro, conforme demonstra a Figura 8.

Figura 8 – Gráfico do modelo iPLS do número de intervalos versus RMSECV. A linha tracejada é o RMSECV para o modelo global e os números em *itálico* são as VLs indicadas no modelo por intervalos



Fonte: O autor.

Seu principal objetivo é prever informação relevante do espectro global nas diferentes subdivisões, de modo que as regiões espectrais cujas variáveis se apresentam como supostamente de menor relevância e/ou interferentes, podem ser removidas, e um novo modelo PLS é construído a partir das variáveis selecionadas (NØRGAARD et al., 2000).

A sensibilidade do algoritmo PLS para variáveis não informativas é salientada pelos gráficos do iPLS. Os modelos do iPLS são desenvolvidos em subintervalos do espectro com igual peso, e a previsão destes modelos locais e o modelo global é comparado. A comparação é baseada no parâmetro de validação RMSECV, mas outros parâmetros são também utilizados como coeficiente de correlação e inclinação da reta do gráfico dos valores reais e previstos pelo

modelo (BORGES NETO, 2005). No caso do exemplo da Figura 8 o espectro total foi dividido em 10 intervalos equidistantes. As barras apresentam a estimativa do RMSECV e o número em itálico dentro de cada barra corresponde ao número de VLS utilizadas para esta estimativa. E, a linha tracejada representa o valor de RMSECV para o modelo global, ou seja, quando se utiliza todas as variáveis espectrais (espectro completo). Este gráfico do iPLS apresenta a vantagem de demonstrar os intervalos que apresentam valores de RMSECV menor do que o modelo global, neste caso, os intervalos 1 e 5.

Assim, a(s) região(ões) espectral(is) com valores de parâmetro de validação menor que o modelo global é(são) selecionada(s) para a construção de modelos PLS que são comparados com a eficiência do modelo PLS global. Amostras e/ou medidas anômalas detectadas pelo PLS devem ser geralmente removidas antes da aplicação do iPLS.

Além do iPLS foram desenvolvidos métodos de combinação de intervalos, como por exemplo, Quadrados Mínimos Parciais por Exclusão de Intervalos (biPLS – do inglês Backward Interval Partial Least Squares) (LEARDI; NØRGAARD, 2004) e Quadrados Mínimos Parciais por Intervalos Sinérgicos (siPLS – do inglês Sinergy Partial Least Squares) (MUNCK et al., 2001).

No método biPLS, a região espectral é dividida em intervalos equidistantes como no iPLS, mas com o objetivo de selecionar um único intervalo como no iPLS ou uma combinação de intervalos, que apresentem uma melhor correlação entre as variáveis selecionadas e a resposta de interesse, quando comparado ao modelo PLS global. No algoritmo biPLS, depois que a região espectral for dividida, o intervalo que apresentar o maior valor de RMSECV, será retirado e um novo modelo é construído com os outros intervalos, fornecendo uma estimativa do valor de RMSECV para os intervalos restantes. Este procedimento é continuado até que reste apenas um intervalo. Os resultados de saída do algoritmo são sumarizados em uma tabela denominada de Tabela de Exclusão (Tabela 1).

Tabela 1 – Exemplo de Tabela de Exclusão obtida pelo algoritmo biPLS para uma divisão de região espectral em 20 intervalos

Número de intervalos	Intervalo	RMSECV	Número de variáveis
20	19	0,0923	2421
19	18	0,0823	2300
18	20	0,0817	2179
17	14	0,0766	2058
16	2	0,0761	1937
15	16	0,0757	1816
14	15	0,0756	1695
13	13	0,0756	1574
12	11	0,0756	1453
11	12	0,0756	1332
10	17	0,0757	1211
9	10	0,0759	1090
8	8	0,0763	969
7	1	0,0797	848
6	7	0,0786	726
5	3	0,0851	605
4	9	0,0846	484
3	6	0,0826	363
2	4	0,0640	242
1	5	0,0782	121

Fonte: O autor.

Os detalhes da interpretação desta Tabela de Exclusão estão apresentados a seguir:

- i) os dados foram divididos em 20 intervalos (conforme coluna 1), e a primeira linha de resultados apresenta a estimativa do valor de RMSECV quando todos os intervalos estão inseridos no modelo (nesse caso, PLS global);
- ii) na primeira linha de resultados também observa-se a indicação do intervalo com maior valor de RMSECV, ou seja, o intervalo 19. Portanto, a segunda coluna da tabela refere-se ao número do intervalo que apresenta maior valor de RMSECV, ou seja, o intervalo 19 seguido dos intervalos 18, 20 e assim sucessivamente;

- iii) na segunda linha de resultados tem-se uma estimativa do valor de RMSECV (0,0823) quando o intervalo 19 é excluído. Portanto, nas linhas subsequentes são demonstradas todas as estimativas de RMSECV à medida que os intervalos vão sendo excluídos;
- iv) por fim, a estimativa do menor valor de RMSECV indica a melhor combinação de intervalos. Neste caso, ocorreu a partir da combinação de dois intervalos, ou seja, quando os intervalos 4 e 5 estão presentes no modelo (RMSECV = 0,0640);
- v) é importante salientar como verificar as variáveis referentes aos intervalos 4 e 5, ou seja, após o número do intervalo 3 (na coluna número de variáveis) inicia-se as variáveis correspondentes ao intervalo 4 que neste caso vai de 364 até 484.

No algoritmo siPLS é realizado uma combinação de todos os intervalos possíveis. Assim, se um conjunto de dados é dividido em 10 intervalos e uma combinação de dois a dois seja escolhida tem-se uma estimativa de 45 modelos PLS. Cada uma das combinações de intervalos fornece uma estimativa do valor de RMSECV. Assim, a combinação que fornece menor valor de RMSECV é escolhida e um novo modelo PLS é construído a partir da combinação destes intervalos. No algoritmo siPLS é possível realizar combinações entre 2, 3 e 4 intervalos. O algoritmo fornece como saída uma tabela de resultados sumarizando o número de VLs, os intervalos a serem combinados e a estimativa do valor de RMSECV para a combinação destes intervalos. O próprio algoritmo já ordena de forma crescente os valores de RMSECV, conforme exemplificado na Tabela 2.

Tabela 2 – Exemplo de resultados de saída fornecido pelo algoritmo siPLS para uma combinação entre 2 intervalos

VLs	Intervalos	RMSECV
4	[1 5]	0.05812
4	[3 5]	0.05901
4	[2 8]	0.06124
4	[1 2]	0.06174
4	[2 6]	0.06354
4	[5 9]	0.06453
4	[2 4]	0.06523
4	[2 5]	0.06664
4	[5 10]	0.06664
4	[2 7]	0.06816

Fonte: O autor.

1.5. Validação do modelo PLS

Sempre que um método é proposto, existe a necessidade de averiguar se ele é adequado para as condições nas quais será aplicado, procedimento denominado de validação, o qual assegura a confiabilidade dos resultados. Com a crescente utilização de modelos multivariados, fez-se necessário o desenvolvimento de guias, normas e trabalhos científicos que enfocam a validação de modelos de calibração multivariada (VALDERRAMA; BRAGA; POPPI, 2009). Alguns documentos apresentam caráter geral, enquanto que outros, como por exemplo, a norma ASTM E1655-05 (2012), descrevem especificamente o desenvolvimento e validação de modelos quantitativos multivariados quando se utiliza a Espectrometria no Infravermelho. Assim, para a validação da metodologia desenvolvida neste trabalho seguiu-se as orientações desta norma.

1.5.1. Exatidão

A exatidão expressa o grau de concordância entre o valor estimado ou medido e o valor considerado como verdadeiro ou de referência. Comumente em aplicações com calibração multivariada a exatidão é estimada através do Erro Quadrático Médio de Calibração (RMSEC) e do Erro Quadrático Médio de Previsão (RMSEP), os quais são calculados de acordo com a

Equação 14 (VALDERRAMA; BRAGA; POPPI, 2009). RMSEC é o RMSE calculado a partir das amostras de calibração, isto é, uma medida do erro na modelagem e RMSEP das amostras do conjunto de previsão, ou seja, amostras utilizadas para validação externa do modelo de calibração.

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

Em que n é o número de amostras, y_i o valor de referência e \hat{y}_i é o valor previsto pelo modelo de calibração.

1.5.2. Linearidade

A linearidade, isto é, a capacidade do modelo de fornecer resultados diretamente proporcionais a concentração do analito de interesse, é verificada a partir da análise conjunta de dois parâmetros: o ajuste e o gráfico de resíduos (PASSARI et al., 2011). O grau de ajuste do modelo pode ser estimado a partir da correlação entre os valores de referência e os valores estimados da propriedade de interesse pelo modelo para as amostras de calibração e validação. Isso é feito determinando-se, por quadrados mínimos, a reta que melhor se ajusta aos valores de referência e os estimados pelo modelo. Aliado ao gráfico de ajuste, a verificação da distribuição aleatória dos valores de resíduos para as amostras de calibração e validação é um indicativo da linearidade do modelo construído (VALDERRAMA; BRAGA; POPPI, 2009).

1.5.3. Teste para Erro Sistemático

A exatidão avaliada a partir do RMSEP expressa a exatidão média do modelo, isto é, inclui tanto os erros aleatórios quanto sistemáticos e, portanto, não pode ser utilizado como único parâmetro para avaliar a exatidão do modelo. De acordo com a definição da União Internacional de Química Pura e Aplicada (IUPAC), *bias* é definido como a diferença entre a média da população e o valor verdadeiro, e esse erro afeta a precisão e exatidão do método. A norma ASTM E1655-05 aborda a investigação de erros sistemáticos em modelos de calibração multivariada através do teste t-Student aplicado nas amostras de validação com 95% de confiança (ASTM, 2012). Para este teste, primeiro é calculado a média do *bias* para o conjunto de validação usando a Equação 15.

$$bias = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)}{n} \quad (15)$$

Onde, n corresponde ao número de amostras do conjunto de validação. A seguir, o desvio padrão dos erros de validação (SDV) é obtido pela Equação 16.

$$SDV = \sqrt{\frac{\sum [(y_i - \hat{y}_i) - bias]^2}{n-1}} \quad (16)$$

Por fim, o valor de t_{bias} é dado pela Equação 17.

$$t_{bias} = \frac{|bias|\sqrt{n}}{SDV} \quad (17)$$

Caso o valor de t_{bias} apresentar resultado maior do que o valor de $t_{crítico}$ para n graus de liberdade e 95% de confiança, isso é uma evidência de que erros sistemáticos presentes no modelo multivariado são significativos. No entanto, se o valor de t_{bias} apresentar valor menor do que $t_{crítico}$, então, o erro sistemático no modelo pode ser considerado desprezível (ASTM, 2012).

1.6. Comparação dos modelos

Para os modelos multivariados o teste F pode ser utilizado na comparação dos erros de previsão de modelos distintos, conforme Equação 18 (FILGUEIRAS; ALVES; POPPI, 2014).

$$F = \left(\frac{RMSEP_1}{RMSEP_2} \right)^2 \quad (18)$$

Onde, $RMSEP_1 > RMSEP_2$. O valor calculado de F (F_{calc}) é comparado com o valor da distribuição de Fisher-Snedecor (estatística F) com graus de liberdade igual ao número de amostras de predição e um nível de significância de 5%. Se o valor tabelado da estatística F é inferior a F_{calc} , não há nenhuma evidência estatística de que os valores seguem a distribuição normal e, portanto o método com $RMSEP_2$ apresenta uma melhor exatidão.

1.7. Aplicações da Espectrometria no Infravermelho combinada com os métodos multivariados para análise de amostras de biodiesel/diesel

O primeiro trabalho que descreve a utilização da Espectrometria no Infravermelho para determinação de biodiesel em mistura com diesel foi publicado por Knothe (2001). O autor empregou a Espectroscopia no Infravermelho Próximo (NIR) especificamente os números de onda 6005 cm^{-1} e 4800 a 4600 cm^{-1} para quantificar biodiesel metílico de soja em mistura com diesel na faixa de concentração de 1 a 95% (v/v). Neste trabalho foram empregados 38 amostras para o conjunto de calibração e 9 amostras para o conjunto de previsão, empregando o método de Regressão por Quadrados Mínimos Parciais. Os desvios máximos entre os valores estimados e medidos ficaram no intervalo de 1-1,5%, consideravelmente mais baixos e dentro do intervalo de erro experimental da preparação da amostra. O autor também destacou que a utilização da banda da carbonila (1740 a 1750 cm^{-1}) na Espectrometria MIR apresenta a desvantagem dos espectros do óleo vegetal e biodiesel, nesta região, serem praticamente idênticos tornando praticamente impossível distinguir se o diesel foi misturado com biodiesel ou óleo vegetal.

Neste sentido, Pimentel et al. (2006) utilizaram a Espectrometria no Infravermelho Médio para determinar o teor de biodieseis metílicos, obtidos a partir de óleo de soja, fritura usado e mamona, em misturas com diesel mesmo na presença de óleo vegetal. Para isto propuseram a identificação da adulteração empregando análise de componentes principais (PCA) aplicando a região de absorção da carbonila de ésteres (1700 a 1800 cm^{-1}). Assim, foi possível resolver o problema da sobreposição dos espectros no infravermelho por causa das semelhanças dos óleos vegetais e os seus respectivos ésteres. Portanto, empregando em conjunto a PCA e PLS, os modelos demonstraram-se apropriados para prever teor de biodiesel em misturas com diesel na faixa de concentração de 0-5% (v/v), considerando a possível presença de óleo vegetal como contaminante.

Por outro lado, Aliske et al. (2007) utilizaram a área e altura da banda da carbonila para quantificar tanto ésteres metílicos quanto etílicos na faixa de concentração de 0 a 100% de biodiesel em diesel empregando calibração univariada. Para isto avaliaram os biodieseis provenientes de diferentes matérias primas, ou seja, girassol, soja, palma e óleo de fritura usado. Os autores demonstraram que ocorreram excelente sobreposição da banda da carbonila dos seus respectivos ésteres e, portanto a altura e a área da banda apresentaram-se proporcionais à concentração do éster.

Mahamuni e Adewuyi (2009) usaram MIR para a determinação de biodiesel na mistura biodiesel/diesel. Nos espectros MIR do biodiesel e diesel foram observadas diferenças nas

regiões de 1700-1800 cm^{-1} , 1000-1300 cm^{-1} e outras. No entanto, os autores utilizaram apenas a região de 1427-1472 cm^{-1} para a quantificação de biodiesel, pois nesta região foram encontradas as diferenças entre biodiesel, diesel e óleo vegetal. Assim, a presença de pequenas quantidades de óleo não podem interferir com a análise de biodiesel. O modelo de calibração desenvolvido funcionou muito bem na previsão dos valores reais e o erro médio foi abaixo de 0,1194.

Vasconcelos et al. (2012) avaliaram a utilização de regiões de sobretons do NIR para determinar o teor de biodiesel e potenciais adulteração por óleos vegetais na mistura biodiesel/diesel. Nesta proposta, os espectros NIR (12000-6300 cm^{-1}) foram obtidos usando três diferentes caminhos ópticos (10, 20 e 50 nm) e duas estratégias de regressão com seleção de variáveis: PLS com algoritmo Jack-Knife e Regressão Linear Múltipla (MLR) com o Algoritmo das Projeções Sucessivas (SPA). Os resultados demonstraram o potencial de regiões sobretons em prever o teor de biodiesel em misturas biodiesel/diesel, no intervalo de 0,0% a 10,0% (v/v), considerando a possível presença de óleo vegetal como contaminante. Em geral, os resultados apontaram SPA/MLR como a melhor estratégia de modelação para esta finalidade, utilizando 10 e 20 mm de caminho óptico, o que é mais simples e utiliza um subconjunto de números de onda (por exemplo, 9-15 variáveis espectrais). Além disso, os resultados também foram melhores do que os obtidos para os modelos que incluem as bandas de absorção de combinação do NIR e regiões espectrais do MIR.

Na perspectiva de métodos de seleção de variáveis Fernandes et al. (2011) avaliaram a utilização combinada e separadamente do uso do Visível e NIR para determinar o teor de biodiesel em misturas com diesel empregando a MLR e o método de seleção de variáveis por SPA. Um total de 100 misturas foram preparadas na faixa de 5 a 50% (v/v) a partir de 10 amostras de biodiesel. No NIR e na região do visível o melhor modelo foi SPA-MLR usando apenas 2 e 8 variáveis com RMSEP de 0,6439% (v/v) e 0,5741, respectivamente. Enquanto, para o modelo visível-NIR o melhor modelo foi SW-MLR usando 5 variáveis e RMSEP de 0,9533% (v/v).

Balabin e Smirnov (2011) compararam a performance de 16 diferentes métodos de seleção de variáveis na previsão de propriedades de biodiesel, incluindo densidade, viscosidade, teores de água e metanol empregando a Espectrometria no NIR. No entanto, não avaliaram a determinação do teor de biodiesel em misturas com diesel. Ferrão et al. (2011) também avaliaram parâmetros de qualidade como densidade, teor de enxofre, ponto de fulgor e teor de biodiesel metílico de soja, porém empregando Espectrometria no MIR e os métodos de seleção de variáveis por intervalos (iPLS e siPLS). O melhor resultado para teor de biodiesel em mistura

com diesel foi o modelo si4PLS8, intervalos de números de onda de 650 a 1909 cm^{-1} e 2746 a 3165 cm^{-1} para a faixa de concentração de 0,2% a 30% (v/v) de biodiesel na mistura.

Em relação a utilização de outras ferramentas quimiométricas pode-se destacar os trabalhos de Alves e Poppi (2013) e Oliveira et al. (2014). O primeiro trabalho verificou a potencialidade do algoritmo de Máquinas de Vetores de Suporte (SVM) aplicados a dados de Espectrometria no NIR no desenvolvimento de modelos de calibração multivariada para determinação do teor de biodiesel em misturas com diesel. A faixa de concentração estudada foi de 0 a 35%, já que na prática esta mistura é usada em até 30% (v/v). No entanto, os autores discutiram a possibilidade de usar modelos de calibração na faixa de 0 a 100% devido à dificuldade de obter exatidão adequada. Os modelos SVM foram comparados com modelos PLS. O melhor resultado foi obtido pelo modelo SVM empregando a faixa espectral 4400-4600 cm^{-1} , apresentando RMSEP de 0,11% no modelo de calibração com faixa de 0 a 35%. O segundo trabalho descreve a aplicação do método de Resolução de Curvas Multivariadas por Quadrados Mínimos Alternados (MCR-ALS) na quantificação de biodiesel em diesel empregando a Espectrometria NIR. A faixa de concentração avaliada foi de 0 a 25% e os resultados demonstraram excelente exatidão para esta determinação com a vantagem do MCR-ALS apresentar informação qualitativa sobre os compostos analisados.

Pinho et al. (2014) avaliaram a utilização da norma EN 14078 para determinação de biodiesel em misturas com diesel vendidas no mercado brasileiro. Assim, eles concluíram que o método é exato e preciso quando aplicado a uma única matéria-prima o que não condiz com a realidade brasileira. Assim, os autores sugeriram que a composição exata deve ser conhecida para construir um gráfico de calibração apropriada, o que é quase impossível devido as peculiaridades no mercado brasileiro de biodiesel. Além disso, os autores concluíram que deve-se garantir que a amostra não sofra oxidação durante o transporte ou estocagem antes do uso da EN 14078 porque os compostos oxigenados produzidos durante a degradação do combustível afetará a exatidão do resultado. Assim, possivelmente devido as características do Programa Nacional de Produção e uso de Biodiesel (PNPB), o método mais apropriado a ser usado na determinação de quantidade de biodiesel no diesel não pode considerar apenas uma variável, como usado pela EN 14078. Métodos multivariados que usam o PLS e Redes Neurais, mostraram serem capazes de determinar com exatidão não apenas a quantidade de biodiesel, bem como o tipo de matéria-prima usado na sua produção (OLIVEIRA et al., 2006). Além disso, os autores acreditam que devido as características da produção de biodiesel e o uso de políticas no Brasil, o país pode necessitar desenvolver seus próprios métodos padrões, provavelmente diferentes do usado na Europa e Estados Unidos onde há uma homogeneidade da matéria-prima.

1.8. Publicações do Laboratório de Quimiometria do Triângulo (LQT) utilizando Espectrometria no Infravermelho Médio aliada a métodos multivariados para análise de amostras de biodiesel/diesel

Fazendo-se um levantamento dos trabalhos publicados pelo LQT observa-se, até o presente momento, 12 publicações que utilizaram a Espectrometria no Infravermelho Médio aliada aos métodos multivariados em análises de amostras de biodiesel/diesel. Dentre estes trabalhos são utilizados tanto métodos para análises quantitativas quanto qualitativas. O método de regressão por Quadrados Mínimos Parciais (PLS) tem sido o mais utilizado, perfazendo um total de 7 aplicações. Em relação aos métodos de classificação são encontrados 6, sendo 3 que utilizam o método de Análise Discriminante por Quadrados Mínimos Parciais (PLS-DA) e 3 por Cartas de Controle Multivariadas. Em uma das publicações utilizaram métodos qualitativos e quantitativos.

Gontijo et al. (2014a) aplicaram a Espectrometria MIR aliada ao PLS para quantificar biodieseis metílicos e etílicos provenientes de óleo de soja, na faixa de concentração de 1,00 a 30,00% (v/v) em amostras de misturas biodiesel/diesel, seguindo as orientações da ASTM E1655. Os modelos PLS apresentaram baixos valores de RMSEP 0,0792% (v/v) e 0,1050% (v/v), respectivamente para os modelos contendo biodiesel metílico e etílico de soja. Os mesmos autores também desenvolveram e validaram modelos PLS para quantificar biodieseis metílicos e etílicos provenientes de óleo de fritura usado em amostras de misturas biodiesel/diesel empregando dados de Espectrometria no Infravermelho Médio. A exatidão dos modelos PLS mostraram excelente desempenho com valores de RMSEC menor que 0,07%, RMSECV menor que 0,09% e RMSEP menor que 0,09% (v/v). Além disso, a exatidão foi comprovada pela avaliação da região de elipse de confiança. O método foi validado de acordo com as diretrizes internacionais e nacionais pela estimativa de figuras de mérito, tais como exatidão, precisão, linearidade, seletividade, sensibilidade analítica, limites de detecção e quantificação e erro sistemático. Uma excelente correlação foi observada entre os valores reais e previstos do conjunto de previsão para ambos os modelos ($R > 0,9999$), além disso, os modelos não apresentaram erros sistemáticos conforme a norma ASTM E1655 (GONTIJO et al., 2014b). Outro trabalho publicado pelo grupo alia a utilização de dois métodos: um de classificação (Cartas de Controle Multivariadas) e outro de quantificação (PLS) (BUIATTE et al., 2016). Na quantificação foi construído um modelo PLS para determinar o teor de biodiesel metílico de algodão em misturas binárias com diesel na faixa de concentração de 1,00 a 30,00% (v/v). Este modelo obteve valor de RMSEP igual a 0,05% (v/v) sendo validado a partir de

figuras de mérito multivariadas conforme diretrizes da ASTM E1655 e IUPAC.

Já Guimarães et al. (2014) utilizaram o PLS para quantificar etanol em amostras de biodieseis etílicos B100 provenientes de óleos de soja e de fritura usado. Os modelos obtidos foram eficientes na determinação da concentração de etanol que variaram entre 0,14 e 1,00% (m/m). Em ambos os modelos PLS, observou-se baixos valores de RMSEP (0,02% m/m), excelente correlação entre os valores medidos e preditos no conjunto de previsão ($R > 0,99$) e não houve a presença de erros sistemáticos conforme ASTM E1655. Além disso, os métodos foram validados de acordo com orientações internacionais e nacionais pela estimativa de figuras de mérito. Os mesmos autores também desenvolveram e validaram metodologias utilizando PLS para quantificar metanol em biodieseis metílicos (B100) provenientes de óleo de soja (GUIMARÃES et al., 2015a) e óleo de soja em misturas B5 de biodieseis metílico e etílico de óleo de soja em diesel (GUIMARÃES et al., 2015b).

No mesmo sentido da quantificação de adulterantes, Souza et al. (2014) desenvolveram um modelo PLS para quantificar óleo lubrificante automotivo usado, na faixa de concentração de 1,75 a 30,00% (m/m), em um conjunto de amostras de diesel brasileiro S-10. O modelo apresentou RMSEP igual a 0,40% (m/m) e foi validado pela estimativa de figuras de mérito e orientações da norma ASTM E1655.

Em relação ao método de classificação por PLS-DA foram publicados três artigos sendo: i) identificação da matéria prima de origem do biodiesel, ou seja, o tipo de óleo e álcool utilizado em sua produção (MAZIVILA et al., 2015a); ii) classificação do tipo de biodiesel tanto metílico quanto etílico provenientes de óleos de soja, óleo de fritura usado e pinhão-mansão em misturas B5 (MAZIVILA et al., 2015b) e iii) identificação da presença de adulterantes (óleos de soja, fritura usado, automotivo usado e gasolina) em misturas B5 composta de biodiesel metílico de soja (MAZIVILA et al., 2015c). Em todos os modelos PLS-DA os autores obtiveram 100% de classificação correta das amostras.

Outro método de classificação utilizado pelo grupo de pesquisadores do LQT foram as Cartas de Controle Multivariadas baseadas no sinal analítico líquido (NAS). Neste sentido, Mitsutake et al. (2015) avaliaram o controle de qualidade de misturas B5 formadas a partir de biodiesel metílico de soja referentes ao teor de biodiesel e adulteração por óleos vegetais, óleo lubrificante automotivo residual, querosene e gasolina. Os resultados obtidos mostram uma excelente distinção entre as amostras dentro e fora das especificações de qualidade, com 91% e 100% de classificação correta. Site et al. (2016) aplicaram as Cartas de Controle Multivariadas para o monitoramento dos teores de biodieseis, provenientes de pinhão-mansão e crambe, em misturas B7 com o diesel.

Baseando-se nos resultados obtidos nos trabalhos desenvolvidos no grupo do LQT, e ressaltando que são coerentes aos encontrados na literatura científica, pode-se afirmar que o uso da técnica de Espectrometria na região do Infravermelho Médio aliada aos métodos quimiométricos apresenta-se como uma solução viável no que tange às análises na área de biocombustíveis. Em resumo, esta aplicabilidade, tem se dado graças às vantagens da técnica como: rapidez na obtenção de espectros, exigência de mínima preparação de amostras, volume muito pequeno da amostra, instrumentação com portabilidade, dentre outras.

2. OBJETIVOS

2. OBJETIVOS

2.1. Objetivo geral

O objetivo geral deste trabalho foi utilizar a Espectrometria no Infravermelho Médio aliada a calibração multivariada e métodos de seleção de variáveis por intervalos para quantificar biodieseis em misturas com diesel.

2.2. Objetivos específicos

A realização deste trabalho envolveu os seguintes objetivos específicos:

1 – Construir e validar modelos de calibração multivariada por Quadrados Mínimos Parciais (PLS) empregando a faixa espectral no MIR, compreendida entre 680 e 3100 cm^{-1} , para quantificação dos biodieseis metílicos e etílicos de soja, fritura usado, algodão e pinhão manso em misturas binárias com diesel na faixa de concentração de 1,00 a 30,00% (v/v).

2 – Avaliar a utilização dos métodos de seleção de variáveis por intervalos iPLS, biPLS e siPLS para a quantificação dos biodieseis em misturas com diesel, comparando com base nos valores de RMSEP e número de variáveis, suas eficiências com os respectivos modelos PLS globais.

3 – Validar os modelos de seleção de variáveis por intervalos que apresentarem melhor eficiência.

3. PROCEDIMIENTO EXPERIMENTAL

3. PROCEDIMENTO EXPERIMENTAL

3.1. Amostras de biodieseis e diesel

O biodiesel metílico de soja foi doado pela empresa Caramuru S/A (Ipameri, Goiás, Brasil). Os biodieseis metílicos e etílicos de óleo de fritura usado, algodão e pinhão manso, além do biodiesel etílico de soja, foram produzidos e caracterizados pelo Laboratório de Quimiometria do Triângulo (LQT), no Instituto de Química da Universidade Federal de Uberlândia. O diesel S-10 puro (isento de biodiesel) foi cedido pela TRANSPETRO S/A (Uberlândia, Minas Gerais, Brasil). O óleo de fritura usado foi obtido em restaurantes locais. O óleo de algodão, usado na síntese do biodiesel, foi adquirido da indústria Triângulo Alimentos (Itápolis, São Paulo, Brasil), enquanto que o óleo bruto de pinhão manso foi adquirido da indústria Bioauto MT Agroindustrial (Nova Mutum, Mato Grosso, Brasil).

3.2. Preparo das misturas biodiesel/diesel

As amostras foram preparadas à temperatura ambiente ($25 \pm 2^\circ\text{C}$) utilizando-se uma balança analítica (Sartorius, modelo BP211D). Para esse procedimento, foi adicionada uma alíquota de biodiesel de massa conhecida em um frasco âmbar de 25 mL, seguida da adição do diesel S-10 e posterior homogeneização em Vórtex (Phoenix, modelo AP56) por 30 segundos. O registro dos espectros MIR das amostras foi realizado após 48 horas de repouso. As frações mássicas (% m/m) dos componentes das misturas foram convertidas em frações volumétricas (% v/v) através da densidade, com a finalidade de expressar os resultados para teor de biodiesel na mesma unidade de medida referenciada nas normas EN14078 e ABNT NBR 15568.

As amostras foram preparadas na faixa de concentração de 0,25% a 30,00% (v/v). Para cada tipo de biodiesel foram preparadas 77 amostras de misturas biodiesel/diesel, sendo 49 amostras para o conjunto de calibração e 28 amostras para o conjunto de previsão. Ambos os conjuntos de amostras foram preparados de forma bem distribuída ao longo da faixa de concentração avaliada.

3.3. Obtenção dos espectros MIR

Os espectros MIR foram obtidos em triplicata, na região de 4.000 a 600 cm^{-1} , resolução de 4 cm^{-1} e 16 varreduras, utilizando um espectrômetro FTMIR, marca Perkin Elmer, modelo Spectrum Two (Figura 9a) com acessório ATR horizontal de ZnSe (Pike Technologies). Após o

registro do espectro de cada amostra, o ATR horizontal (Figura 9b) foi limpo com álcool isopropílico (marca Synth, grau P.A).

Figura 9 – (a) espectrômetro FT-MIR (Perkin Elmer, modelo Spectrum Two) e (b) ATR horizontal (Pike Technologies)



(a)



(b)

Fonte: O autor.

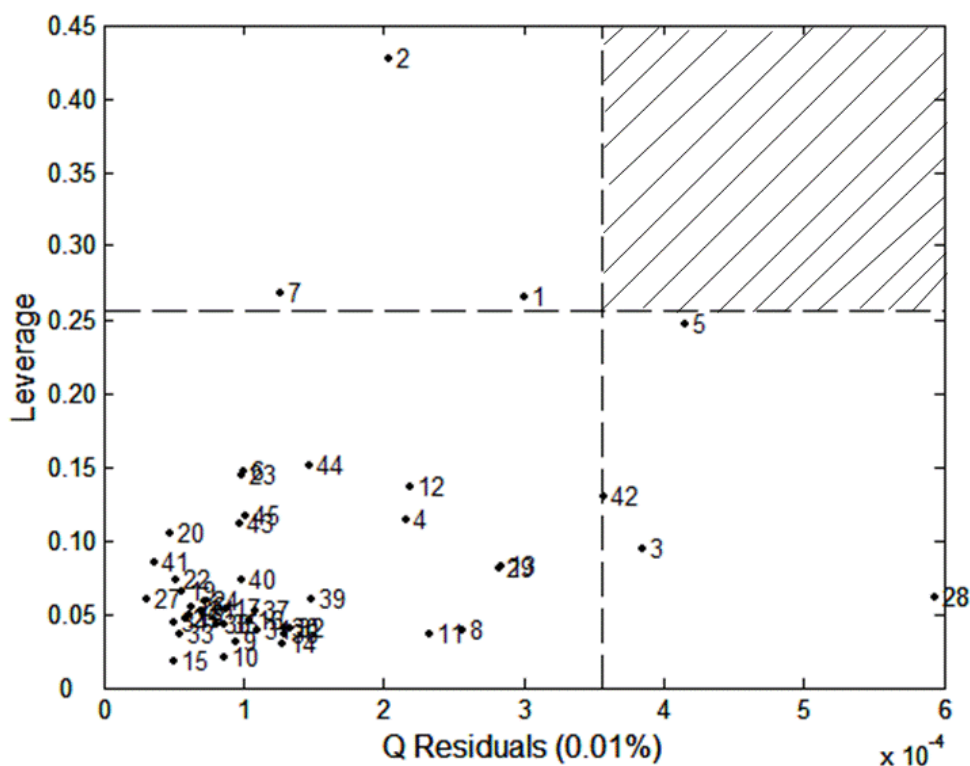
Os espectros foram corrigidos pelo método de *baseline* aplicado nas regiões de 2570 a 1850 cm^{-1} e 4000 a 3200 cm^{-1} (regiões sem bandas de absorção). Após esta correção da linha de base, a região abaixo de 680 cm^{-1} e acima de 3100 cm^{-1} foram descartadas, por apresentarem grande quantidade de ruído e nenhuma banda de absorção, respectivamente.

3.4. Construção dos modelos de calibração multivariada

Foram construídos oito modelos PLS empregando a faixa espectral compreendida entre 680 cm^{-1} e 3100 cm^{-1} para quantificar os 8 tipos de biodiesel em misturas com diesel, sendo: biodiesel metílico de soja (BMS) e etílico de soja (BES); biodiesel metílico (BMSR) e etílico (BESR) de óleo de fritura usado; biodiesel metílico (BMA) e etílico (BEA) de algodão; biodiesel metílico (BMPM) e etílico (BEPM) de pinhão-mansão. Os dados foram centrados na média e o método *leave-one-out* foi empregado para a validação interna. O número de VLs foi selecionado em comparação conjunta da porcentagem de variância explicada em **X** (matriz de espectros) e **y** (concentração) e o gráfico de Erro Quadrático Médio de Validação Cruzada (RMSECV) versus VL.

A detecção de amostras anômalas (*outlier*) foi realizada através do gráfico de influência ou *leverage* (grau que uma amostra está distante da média do conjunto de dados) e dos resíduos de Student ou *Q Residuals* (a parte não modelada nos dados espectrais ou nos valores de concentração analítica), a nível de significância de 5% (VALDERRAMA; BRAGA; POPPI, 2007). Amostras fora do limite de confiança de 95% (região achurada na Figura 10) são consideradas amostras anômalas. Além disso, amostras anômalas também foram detectadas a partir da comparação entre altos valores de erros absolutos nas amostras individuais em relação ao valor de erro quadrático médio de calibração (RMSEC), ou seja, se uma amostra tem uma diferença entre o valor de referência e o valor estimado maior que três vezes o RMSEC, esta deverá ser excluída (FERREIRA; BRAGA; SENA, 2012).

Figura 10 – Gráfico de *Leverage* versus resíduos de Student na identificação de amostras anômalas



Fonte: O autor.

Para construção dos modelos de seleção de variáveis por intervalos foram utilizados os algoritmos iPLS, biPLS e siPLS empregando rotinas próprias modificadas a partir das demonstrações contidas no iToolbox 1.0 (NØRGAARD, 2004). Nestes algoritmos foram avaliadas a divisão da faixa espectral completa (680 a 3100 cm^{-1}) em 5, 10, 15, 20, 25, 30 e 35 intervalos. Para os resultados de saída do algoritmo iPLS foi selecionado e construído um

modelo PLS, dentro de cada divisão, com o intervalo que apresentasse menor valor de RMSECV se comparado com o valor de RMSECV do modelo PLS global, ou seja, empregando a faixa espectral completa. Em relação aos algoritmos biPLS e siPLS construiu-se apenas o modelo PLS com a combinação de intervalos que apresentassem menor valor de RMSECV. Para o algoritmo siPLS foram avaliadas as combinações entre 2, 3 e 4 intervalos. O melhor intervalo, entre cada uma das combinações, foi selecionado com base no menor valor de RMSECV.

As figuras de mérito empregadas para validação dos modelos PLS estão descritas no item 1.5. O programa utilizado para a construção dos modelos foi Matlab versão 6.1 (MathWorks Inc.) juntamente com PLS_Toolbox 3.5 e iToolbox 1.0 (Eigenvector Research Inc.).

4. RESULTADOS E DISCUSSÃO

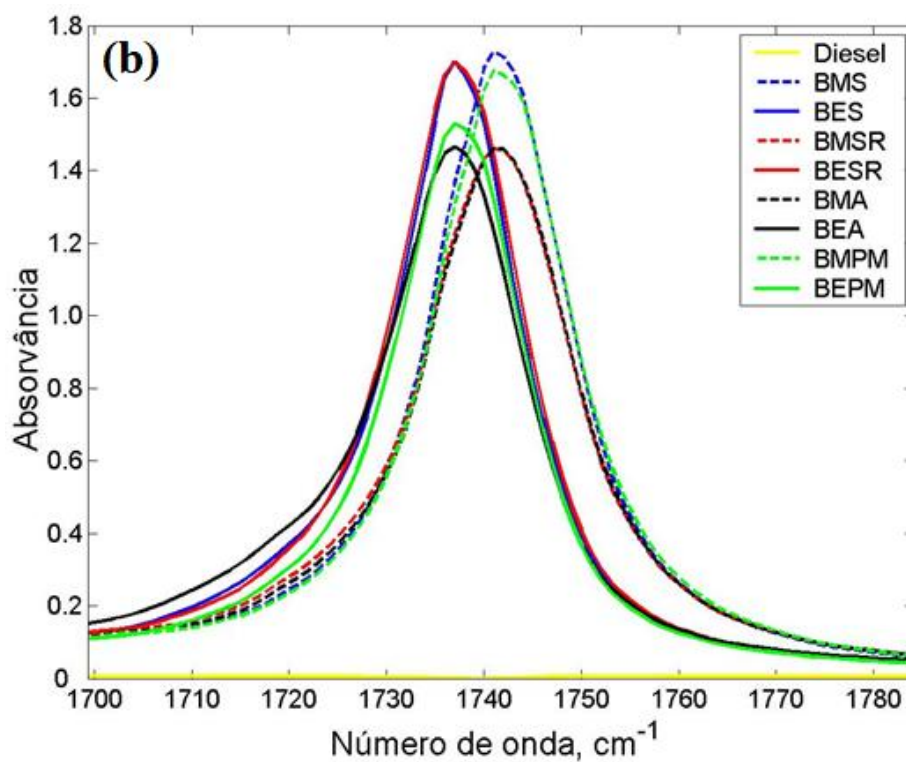
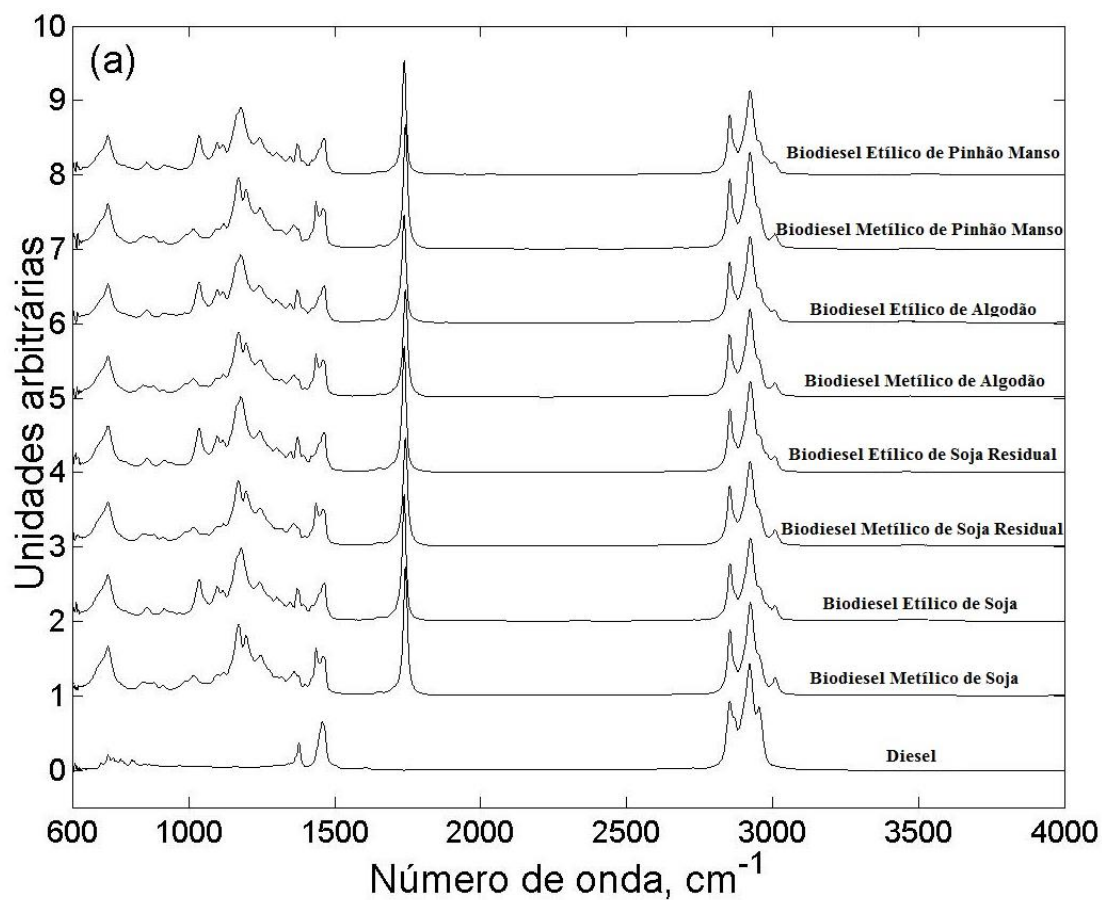
4. RESULTADOS E DISCUSSÃO

4.1. Espectros MIR de biodiesel e diesel

A Figura 11a apresenta os espectros MIR dos diferentes tipos de biodieseis empregados neste estudo e o espectro do diesel. Fazendo uma avaliação detalhada dos espectros verifica-se principalmente diferenças entre os espectros dos biodieseis e do diesel. No espectro do diesel verifica-se bandas de absorção correspondentes aos modos vibracionais característico de alcanos normais. São três regiões espectrais de absorção significativas do diesel: a região entre 2840 cm^{-1} e 3000 cm^{-1} atribuída às vibrações de estiramento das ligações C-H dos grupamentos de metila e metileno; bandas de intensidade intermediária na região de 1300 cm^{-1} a 1500 cm^{-1} das vibrações de deformação angular das ligações C-H do grupo metila e metileno e, outra banda de baixa intensidade, porém relevante na região de 720 cm^{-1} atribuída às vibrações de deformação angular assimétrica resultante das deformações C-H do grupamento metileno (SKOOG; HOLLER; NIEMAN, 2002). Os biodieseis apresentam em comum padrões de absorção na banda ao redor de 3010 cm^{-1} relacionada às vibrações de deformação axial da ligação C-H da ligação dupla; a presença de absorção na faixa de 1700 cm^{-1} a 1750 cm^{-1} , da carbonila (C = O) e a presença de bandas na região de 1000 cm^{-1} a 1300 cm^{-1} atribuídas às vibrações de estiramento da ligação C-O. Os ésteres metílicos possuem absorção com três bandas na região de 1165 cm^{-1} a 1265 cm^{-1} , de modo que na mesma região para o éster etílico sobressaem duas bandas.

Na Figura 11b, foi destacado a região de estiramento da carbonila, 1765 a 1720 cm^{-1} para demonstrar que biodiesel produzidos a partir de diferentes matérias primas e diferentes álcoois (metanol ou etanol), apresentam diferenças de absorção nessa região, onde os biodieseis metílicos apresentam bandas deslocadas para a direita em relação aos biodieseis etílicos, provavelmente devido ao aumento da doação da densidade eletrônica do substituinte etil. Além disso, existem diferenças nas intensidades da banda da carbonila. Estas diferenças contribuem como fontes de erro na determinação do teor de biodiesel em mistura com diesel caso o modelo de calibração seja construído com uma única fonte de matéria-prima, por exemplo, biodiesel metílico de soja (PINHO et al., 2014).

Figura 11 – (a) Espectros MIR dos biodieseis puros e do diesel (600-4000 cm^{-1});
(b) Destaque para a região de estiramento da carbonila (1700 – 1780 cm^{-1})

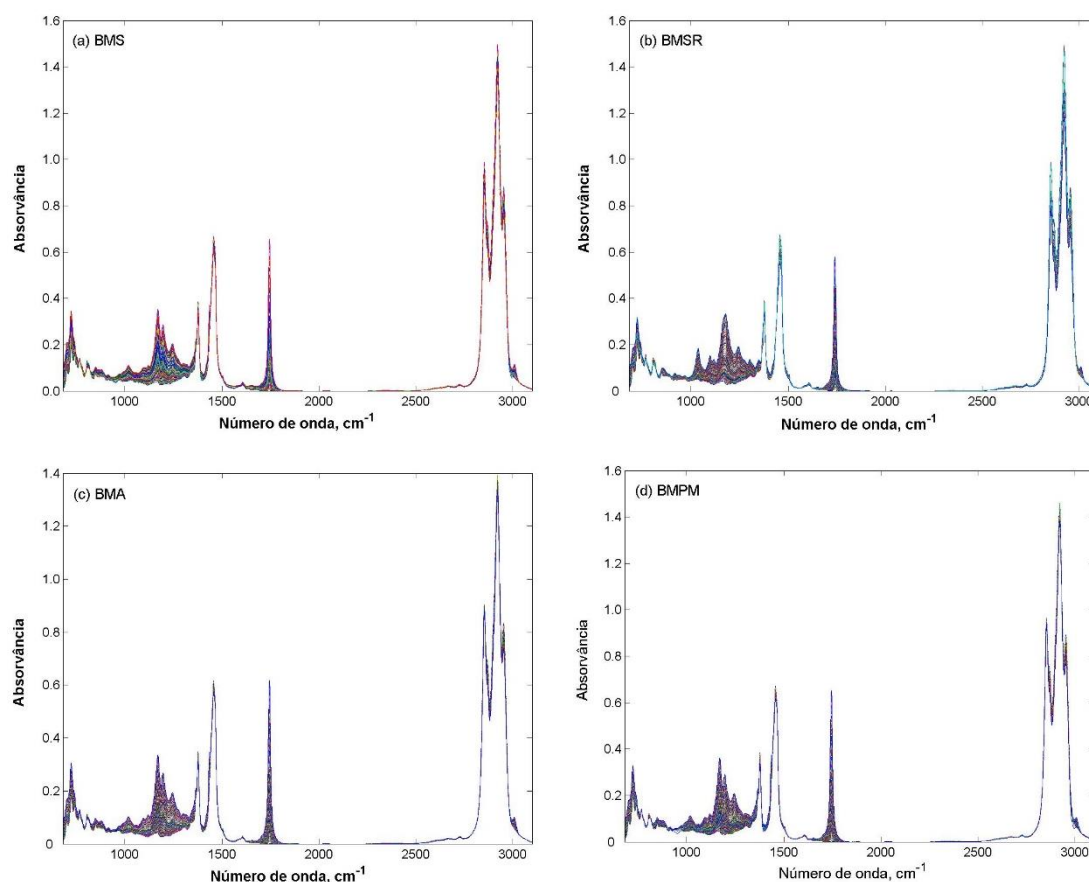


Fonte: O autor.

4.2. Modelos PLS global

As Figuras 12 e 13 apresentam os espectros obtidos das amostras de calibração e previsão para construção de cada modelo PLS. Verifica-se que em todos os conjuntos espectrais que as regiões de maior variação ocorrem na região de impressão digital de 1000-1300 cm^{-1} e na variação do grupo carbonila de 1735-1750 cm^{-1} . As regiões da norma ABNT NBR 15568 estão justamente entre estes intervalos. Mas, pequenas variações também ocorrem na intensidade do sinal do diesel visto que com o aumento da concentração de biodiesel ocorre uma diluição do diesel. E, a utilização desta e outras regiões podem tornar o modelo de calibração multivariada com performance melhor ou no mínimo igual ao estabelecido por esta norma. Desta maneira avaliou-se inicialmente a performance dos modelos PLS empregando a faixa espectral compreendida entre 680 e 3100 cm^{-1} (denominada de espectro total ou global).

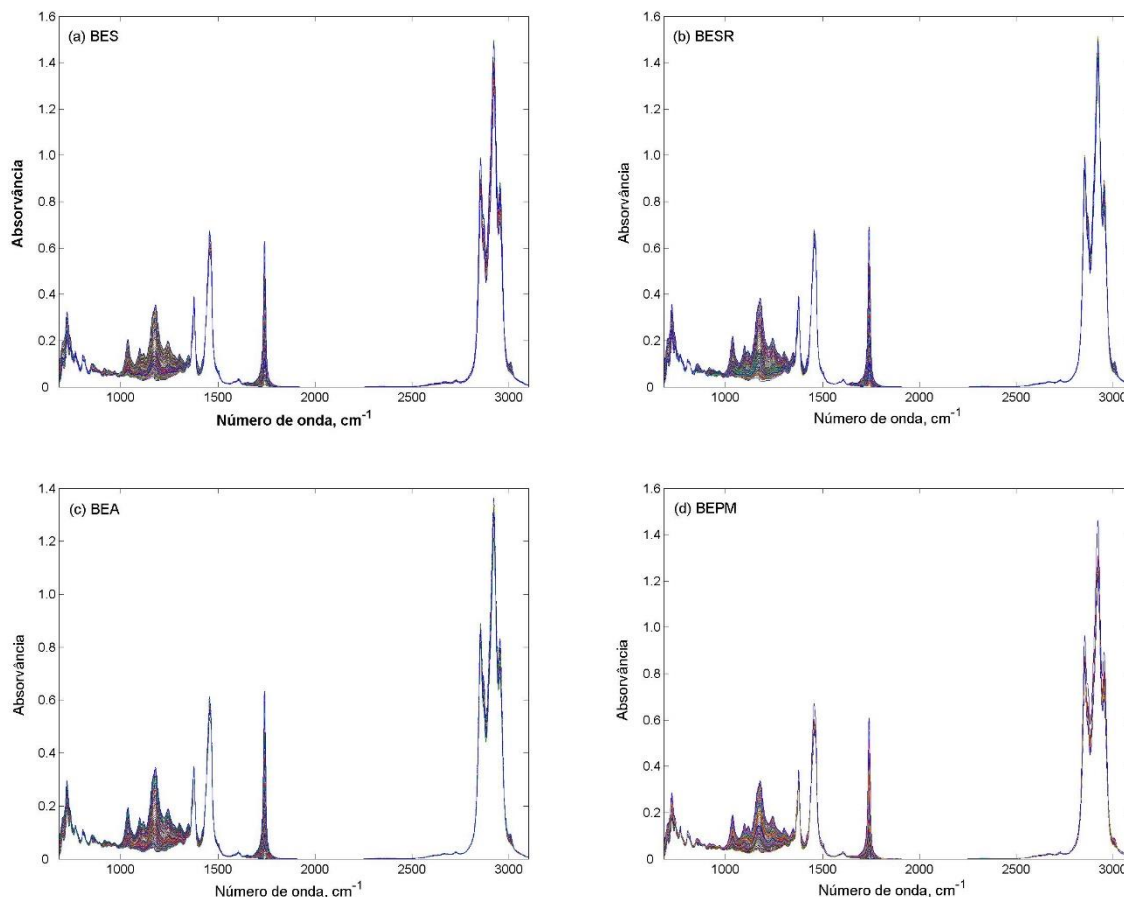
Figura 12 – Espectros MIR das amostras de biodieseis metílicos dos conjuntos de calibração e previsão obtidos na faixa de trabalho de 0,25 a 30% (v/v) de biodiesel em mistura com diesel. (a) BMS; (b) BMSR; (c) BMA; (d) BMPM



Fonte: O autor.

Figura 13 – Espectros MIR das amostras de biodieises etílicos dos conjuntos de calibração e previsão obtidos na faixa de trabalho de 0,25 a 30% (v/v) de biodiesel em mistura com diesel.

(a) BES; (b) BESR; (c) BEA e (d) BEPM



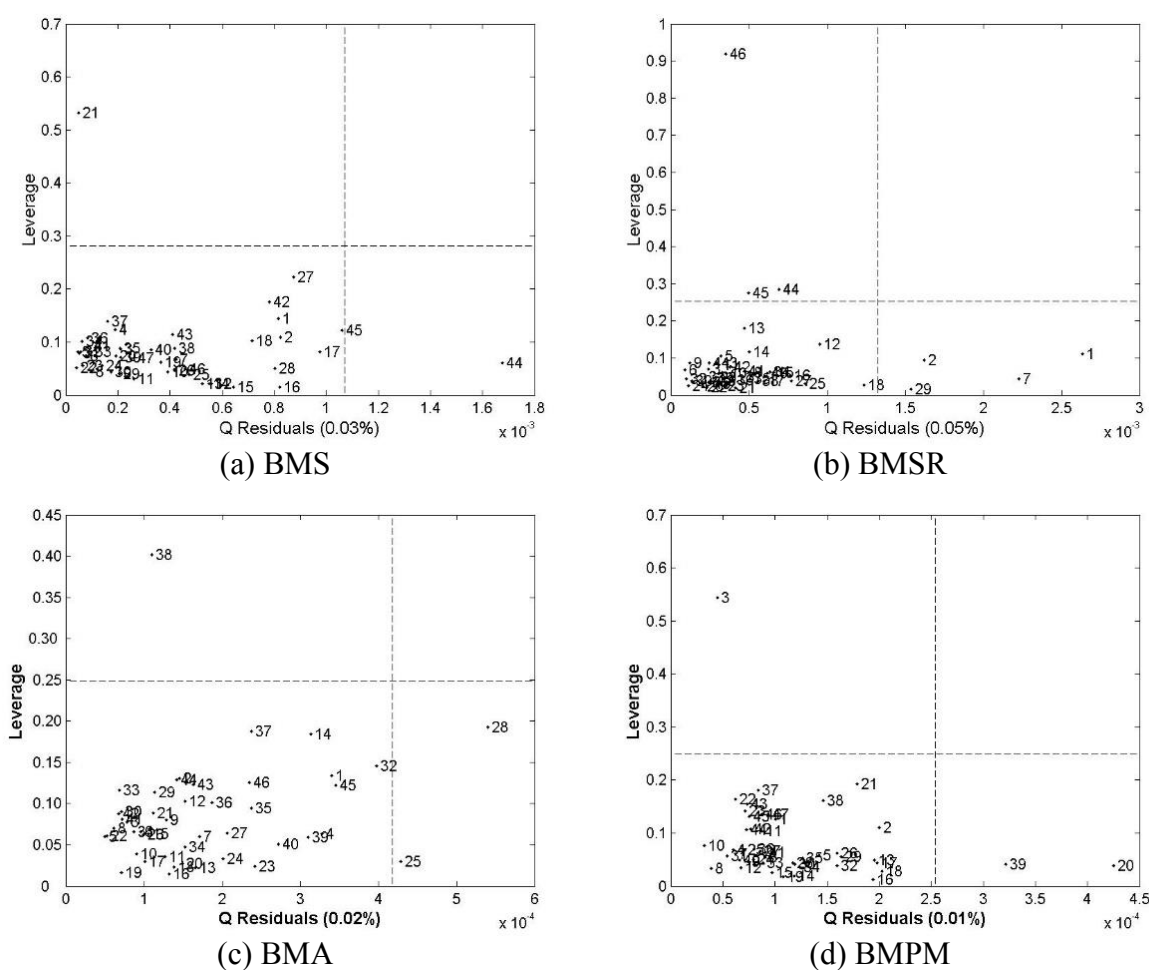
Fonte: O autor.

Um passo importante na construção de modelos de calibração multivariada é a identificação de possíveis amostras anômalas uma vez que estas amostras podem afetar a qualidade final dos modelos (FERNANDES et al., 2011). As Figuras 14 e 15 apresentam o gráfico *Q residuals* contra *leverage* para os 8 modelos PLS. As linhas horizontais representam os limites de resíduos, a 95% confiança, enquanto que a linha vertical representa a influência crítica do valor de leverage, definido como $3k/n$, onde k é o número de fatores e n o número de amostras (FERREIRA et al., 1999). Verifica-se na Figura 14 que as amostras 21, 46, 38 e 3, dos modelos BMS, BMSR, BMA e BMPM, respectivamente apresentam altos valores de leverage, mas apesar destas amostras estarem distante do centro do modelo apresentam resíduos baixos, não prejudicando a qualidade do modelo. Na Figura 15 observa-se o mesmo comportamento para as amostras 44 e 42 dos modelos BESR e BEA, respectivamente e, as amostras de número 1, 2 e 7 para o modelo BEPM. Por outro lado, também são observados amostras com valores de

resíduos fora dos limites estatísticos em todos os modelos, porém todas estas amostras apresentam valores de leverage baixo e, portanto, não são consideradas amostras anômalas. Assim, com base nos gráficos nenhuma das amostras foram consideradas amostras anômalas em toda a região espectral avaliada, ou seja, nenhuma amostra apresentou altos valores de leverage e altos valores de resíduos.

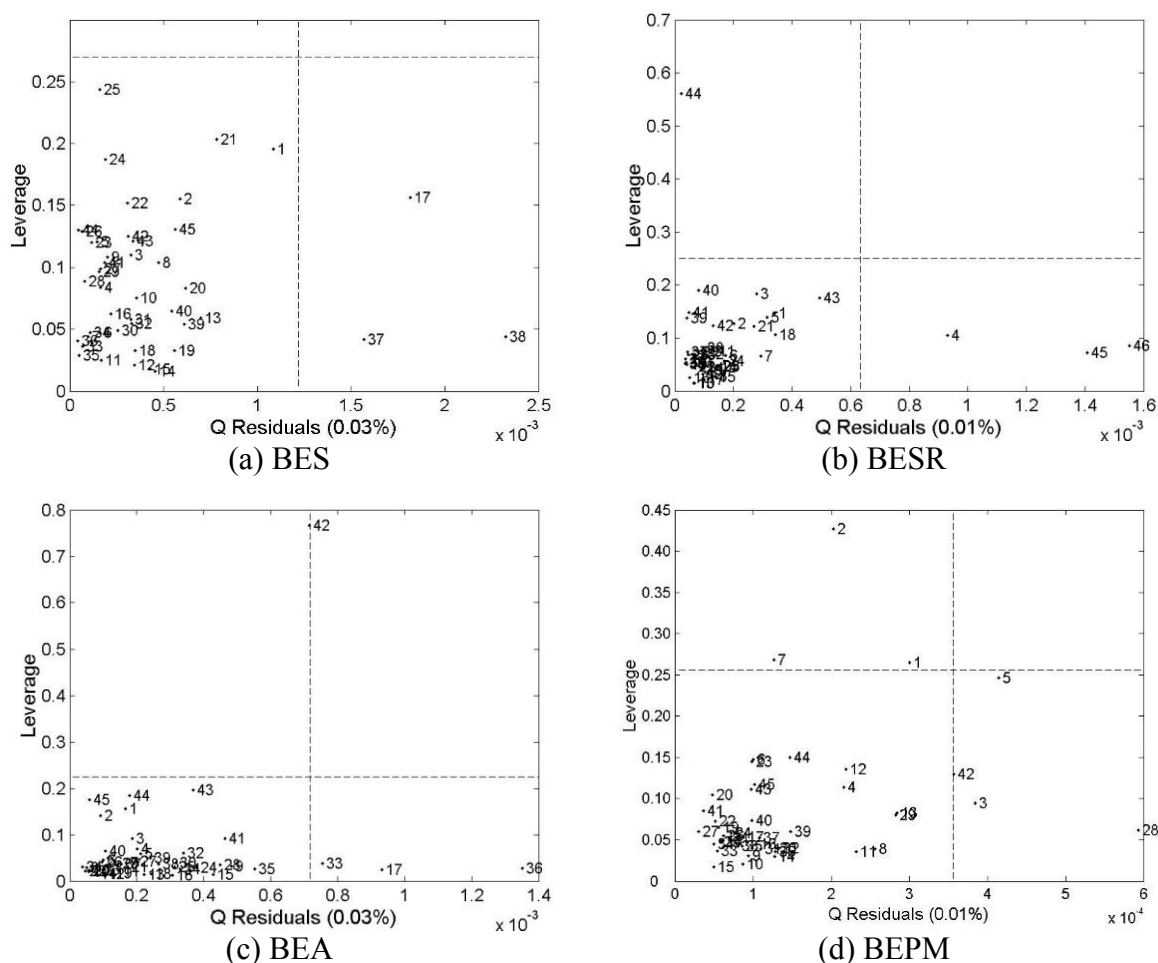
Figura 14 – Gráfico de *leverage* versus *Q residuals* com 95% de confiança para os modelos

PLS global dos biodieseis metílicos: (a) BMS, (b) BMSR, (c) BMA, (d) BPM



Fonte: O autor.

Figura 15 – Gráfico de *leverage* versus *Q residuals*, com 95% de confiança, para os modelos PLS global dos biodiesel etílicos: (a) BES, (b) BESR, (c) BEA e (d) BEPM



Fonte: O autor.

Porém, amostras anômalas foram detectadas a partir de altos erros absolutos das amostras individuais em comparação com o valor de RMSEC. Estas amostras anômalas ocorreram com amostras de baixa concentração deixando os modelos com faixa de trabalho de 1,00 a 30,00% (v/v).

Após a remoção das amostras anômalas torna-se necessário verificar se o número de amostras utilizadas na construção dos modelos PLS seguem as exigências da norma ASTM E1655 (ASTM, 2012). Esta norma estabelece que para o conjunto de calibração o número mínimo de amostras deve ser $6(k+1)$ para os dados centrados na média e, para o conjunto de previsão, $4k$. O valor de k refere-se ao número de VLs escolhido para construção do modelo PLS. Observa-se na Tabela 3 que tanto no conjunto calibração quanto de previsão o número de amostra segue as exigências desta norma.

Tabela 3 – Número de amostras nos conjuntos de calibração e previsão, número de VLs, variância capturada e valores de exatidão para os modelos PLS global

MODELO	BMS	BMSR	BMA	BMPM	BES	BESR	BEA	BEPM
Nº amostras de calibração	47 (30)*	46 (30)	46 (30)	47 (24)	45 (30)	46 (30)	45 (24)	45 (30)
Nº amostras de previsão	26 (20)	27 (20)	27 (20)	27 (20)	27 (20)	27 (20)	25 (20)	27 (20)
VLs	4	4	4	3	4	4	3	4
Variância em X (%)	99,97	99,95	99,98	99,96	99,97	99,99	99,97	99,99
Variância em y (%)	99,99	99,99	100,00	100,00	100,00	100,00	100,00	100,00
RMSECV	0,0898	0,1217	0,0516	0,0717	0,0658	0,0461	0,0866	0,0723
RMSEC	0,0761	0,0938	0,0418	0,0463	0,0523	0,0371	0,0612	0,0590
RMSEP	0,0916	0,1020	0,0473	0,0602	0,0974	0,0483	0,0751	0,0723

* Os números entre parênteses referem-se ao número mínimo de amostras exigidas pela norma ASTM E1655-05 (ASTM, 2012).

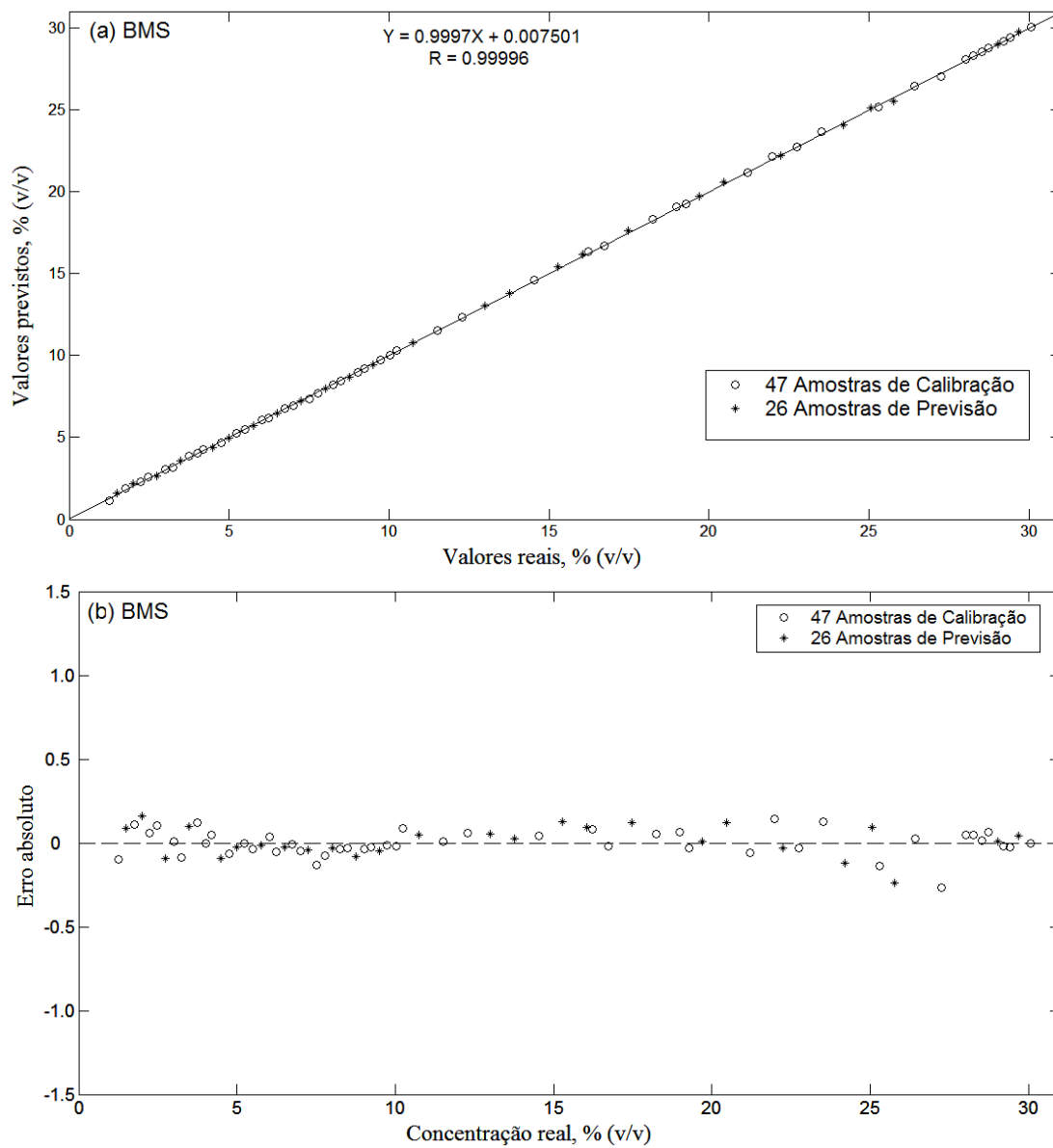
Fonte: O autor.

Os modelos PLS foram construídos com quatro VLs, com exceção aos modelos de BEA e BMPM que foram construídos com 3 VLs. A Tabela 3 também apresenta os valores de variância capturada para os blocos **X** e **y**.

Nestas condições, avaliando a exatidão dos modelos (Tabela 3), através dos valores de RMSEC, RMSECV e RMSEP, pode-se observar que em todos os modelos os valores de erros foram satisfatórios, ou seja, estão dentro da exatidão requerida pela norma ABNT NBR 15568 (ABNT, 2008), a qual especifica que o valor máximo de RMSEP permitido é de 0,1% (v/v) e 1% (v/v) para os modelos construídos de 0,00 a 8,00% (v/v) e de 8,00 a 30,00% (v/v), respectivamente. Além disso, devido a exigência futura no Brasil de adição de 10% (v/v) de biodiesel ao diesel a ABNT NBR 15568 aceita uma exatidão de até 1% para esse teor de biodiesel, portanto os modelos propostos apresentam exatidão na ordem de 10 vezes menor que o recomendado. Também deve-se observar que os valores de RMSEC e RMSEP, apresentam uma pequena diferença entre eles indicando que não há sobreajuste ou subajuste (BRAGA; SANTOS JUNIOR; MARTINS, 2014; LIU; HE; WANG, 2008).

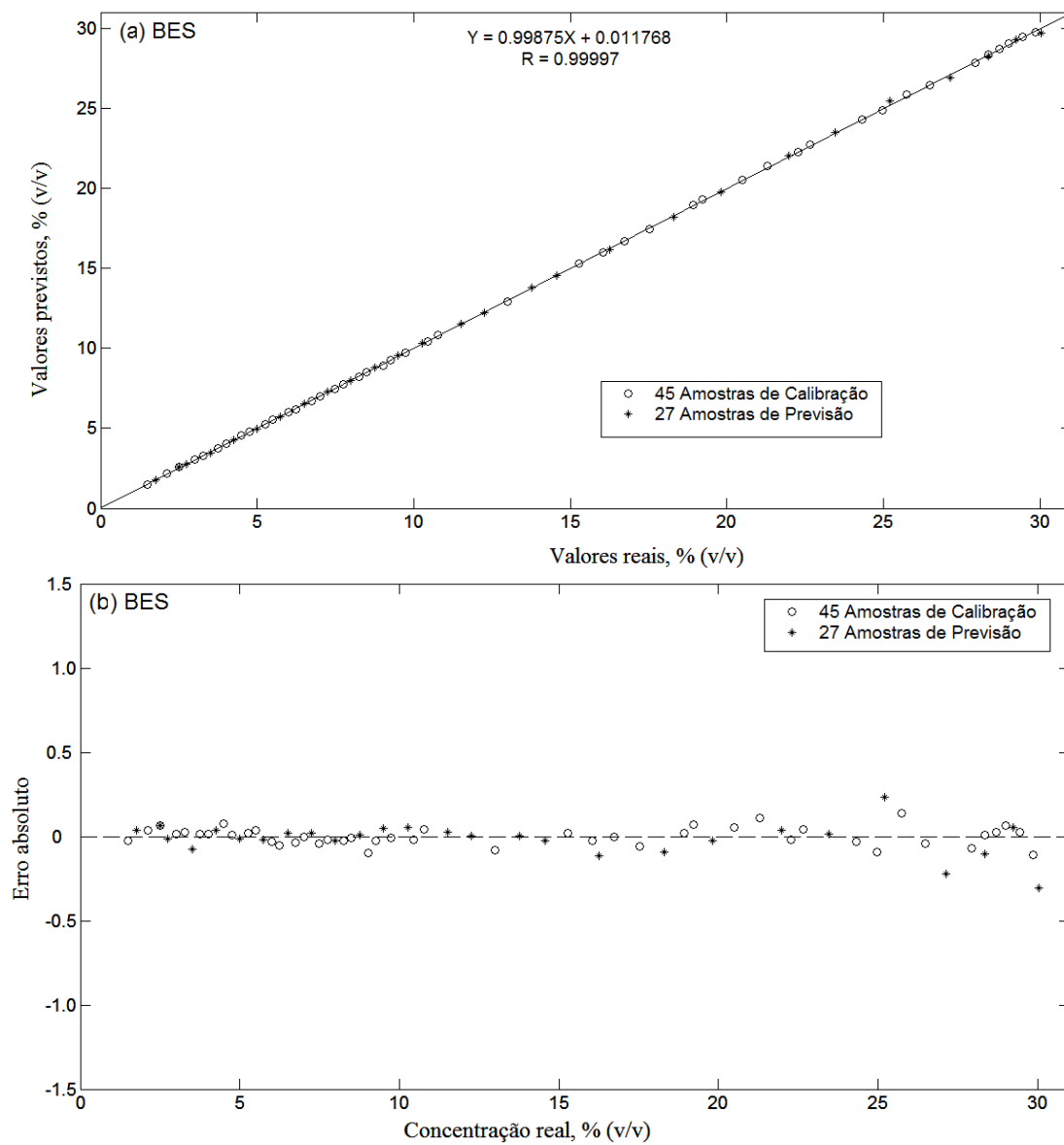
A linearidade dos modelos foi verificada com base no gráfico dos valores reais de concentração de biodiesel versus valores previstos pelos modelos em avaliação conjunta com o gráfico de resíduos. Assim, conforme apresentado nas Figuras 16 até 23, as amostras de calibração (círculos) e de previsão (asteriscos) estão próximas à regressão e bem distribuídas ao longo da reta. De acordo com os valores dos coeficientes de correlação (R), observa-se que o ajuste para o modelo PLS tem uma boa correlação entre os dados de referência e os dados calculados pelo modelo PLS, uma vez que quanto mais próximo de 1 para a inclinação, e mais próximo de 0 para o intercepto, melhor é a concordância entre os valores. No entanto, esta linearidade deve ser evidenciada pela distribuição aleatória dos resíduos. As Figuras 16 até 23 também apresentam os gráficos de resíduos para os modelos. Verifica-se que em todos os modelos os resíduos estão distribuídos aleatoriamente ao redor da linha no valor zero. Portanto, os modelos PLS apresentaram comportamento linear ao longo da faixa de trabalho de 1,00% a 30,00% (v/v).

Figura 16 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BMS



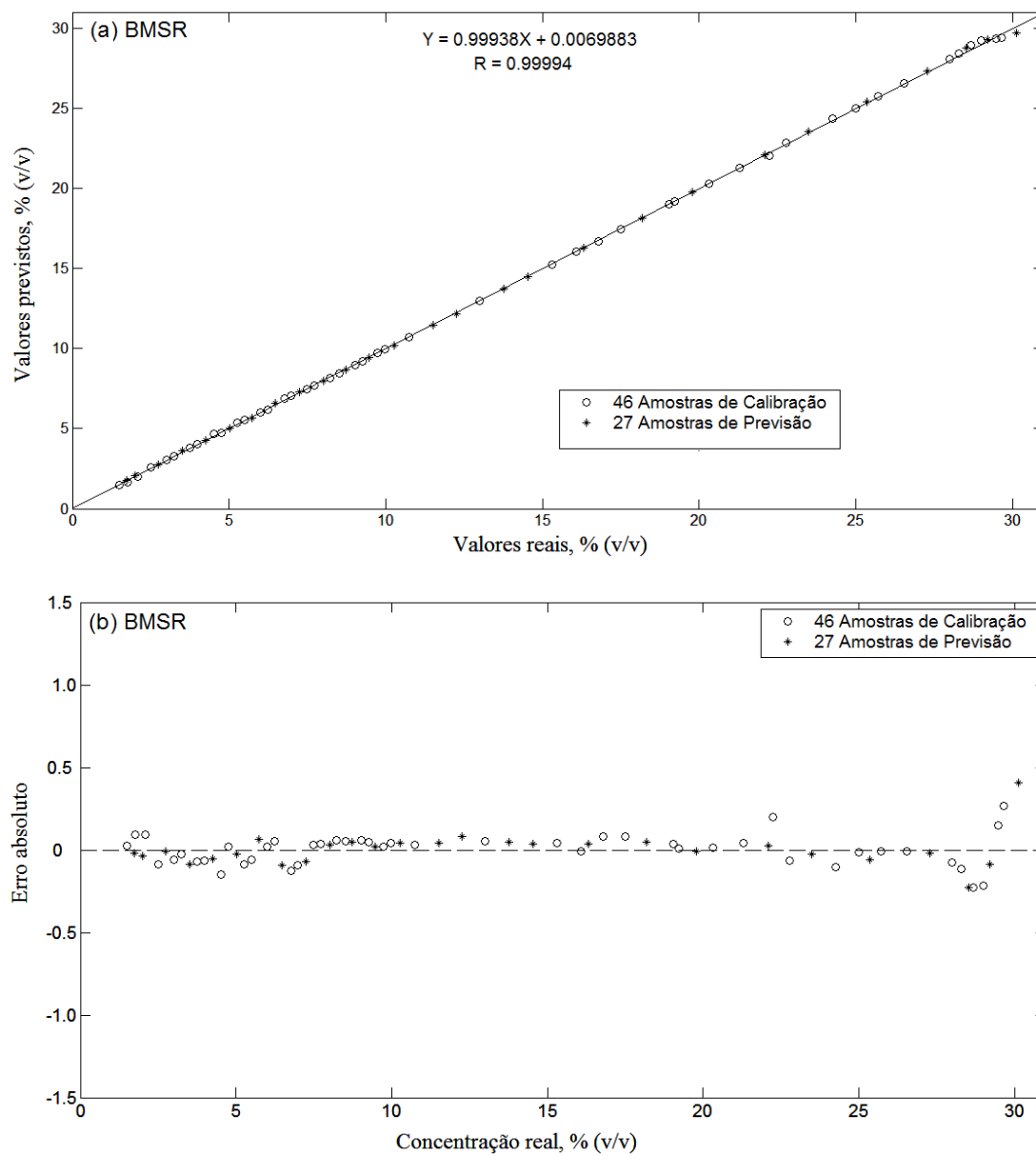
Fonte: O autor.

Figura 17 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BES



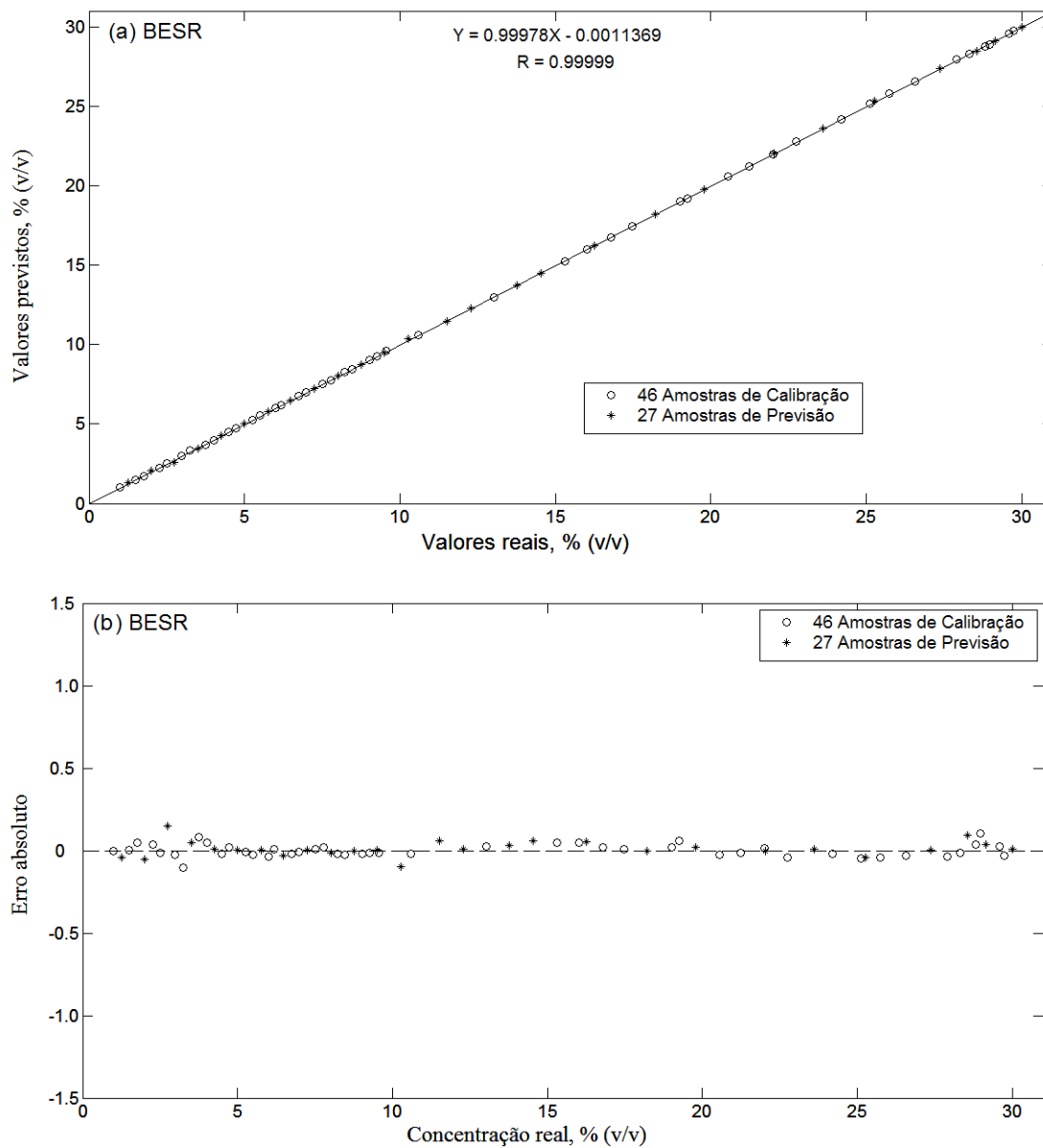
Fonte: O autor.

Figura 18 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BMSR



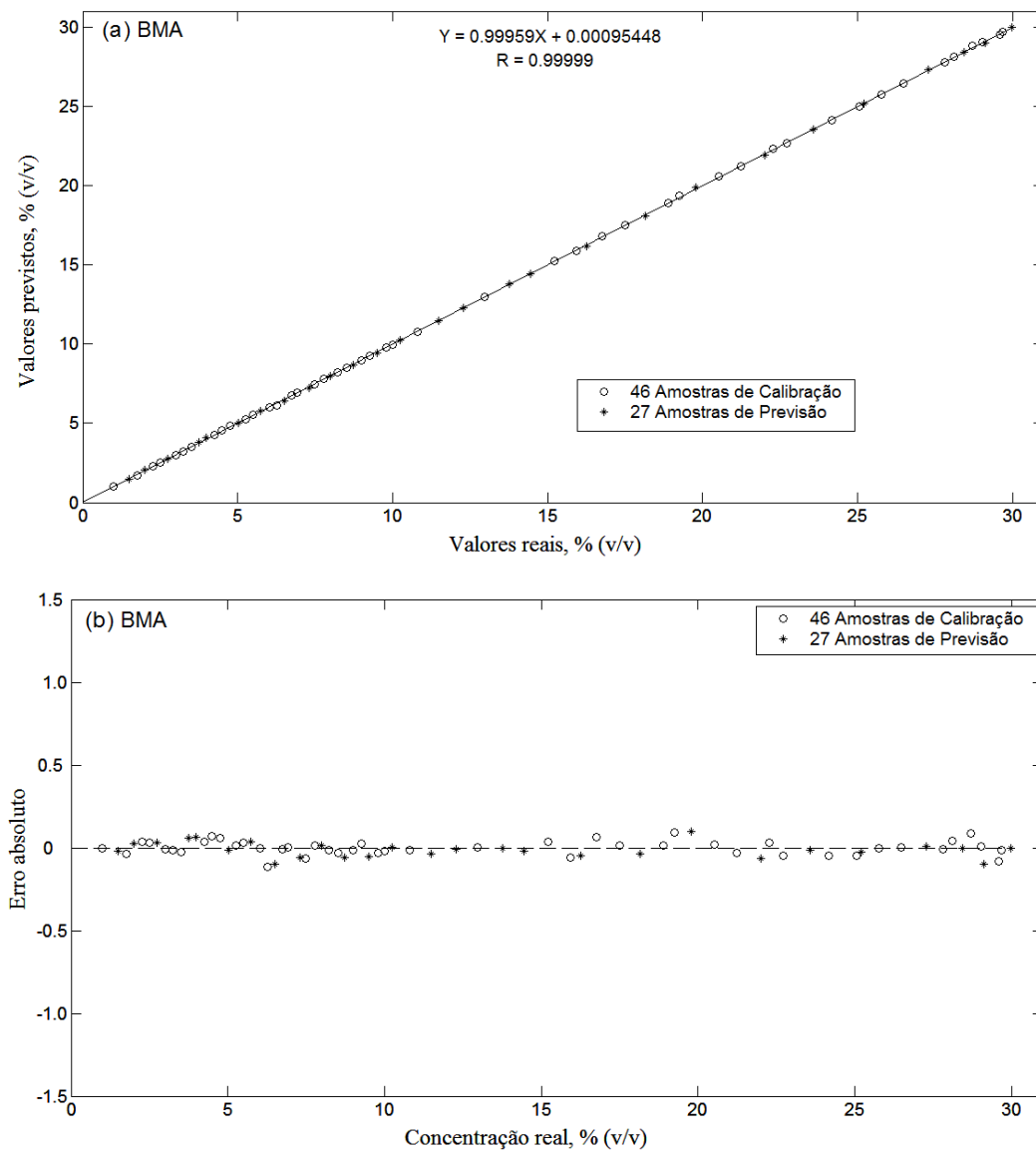
Fonte: O autor.

Figura 19 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BESR



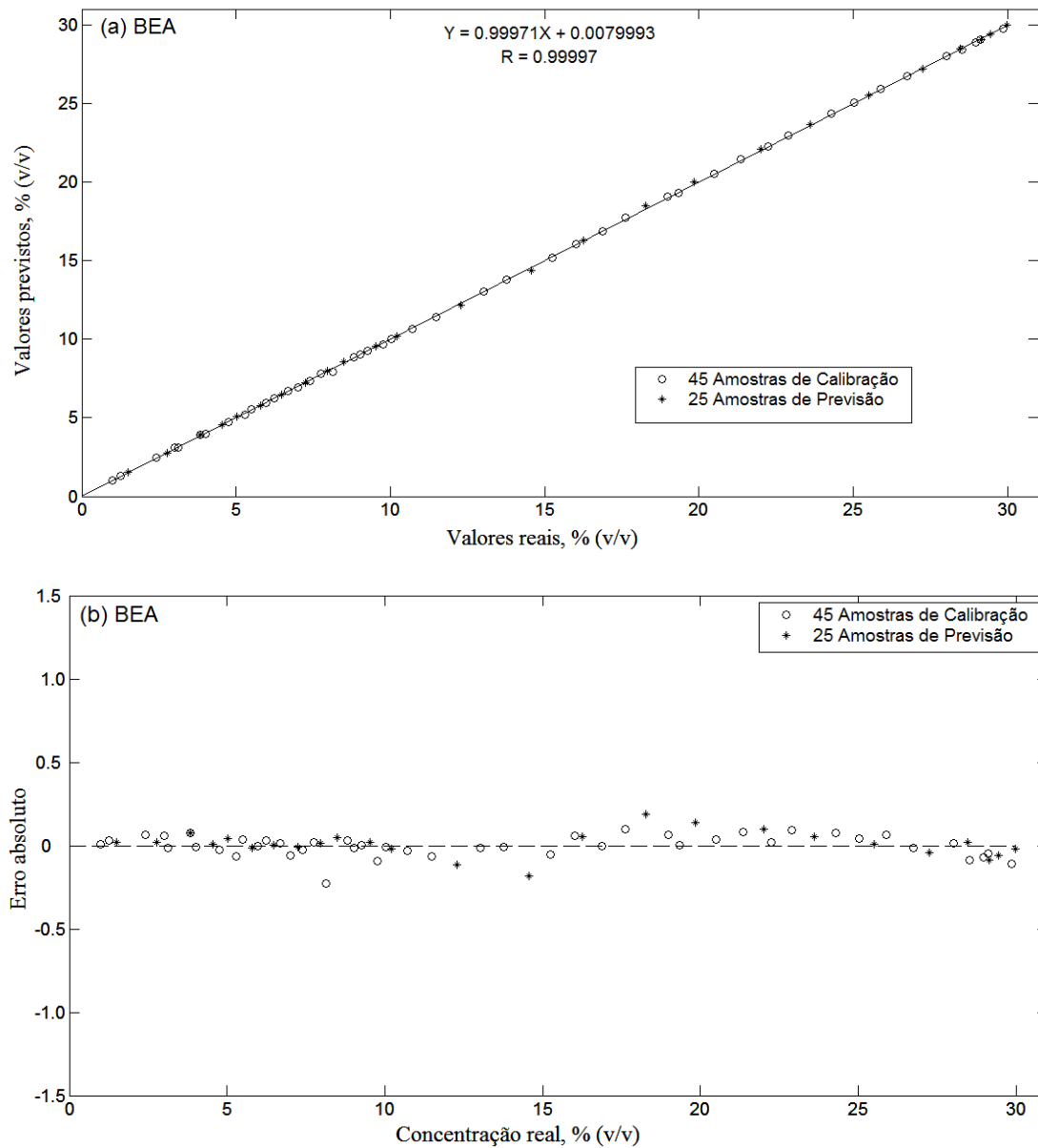
Fonte: O autor.

Figura 20 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BMA



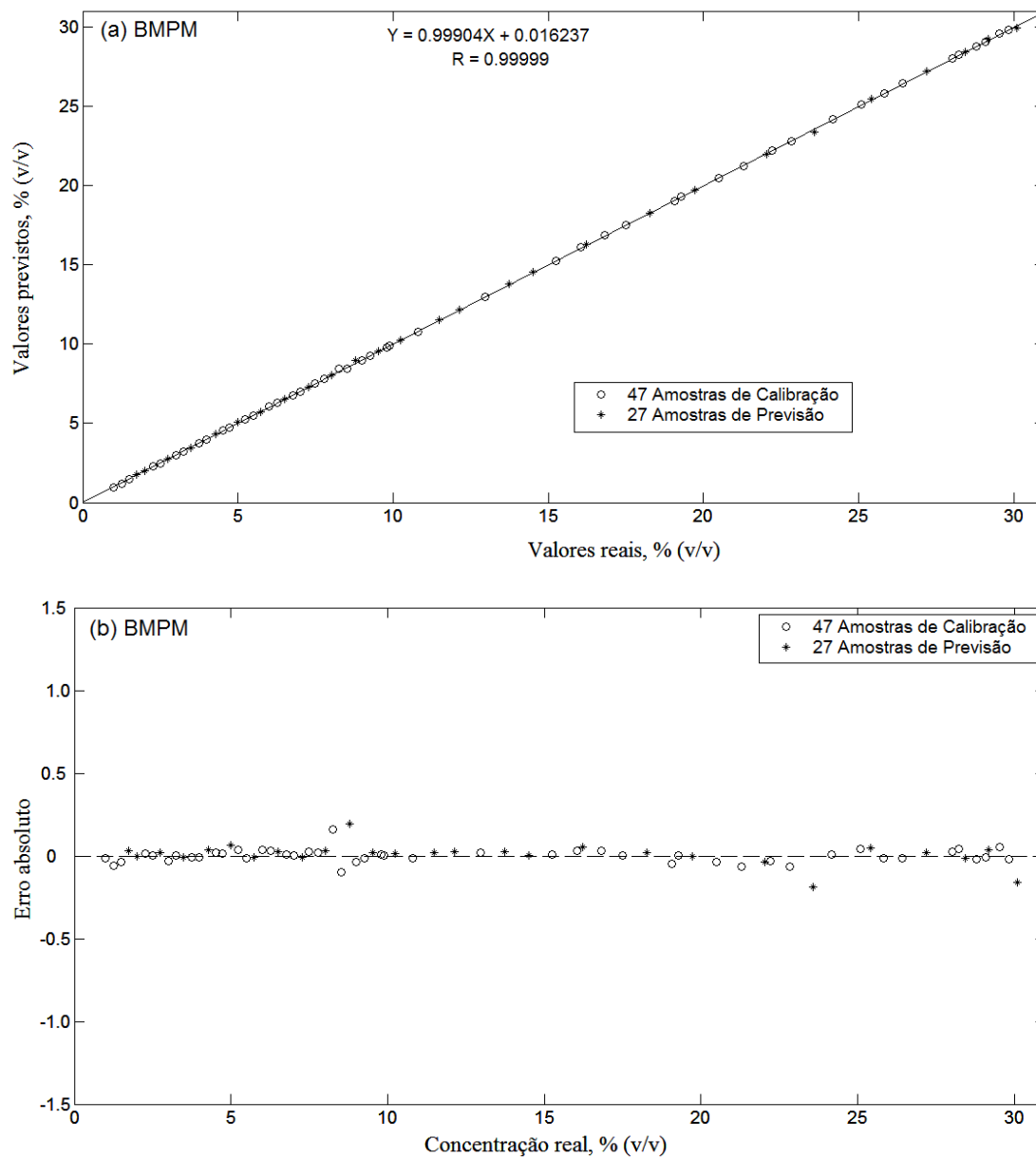
Fonte: O autor.

Figura 21 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BEA



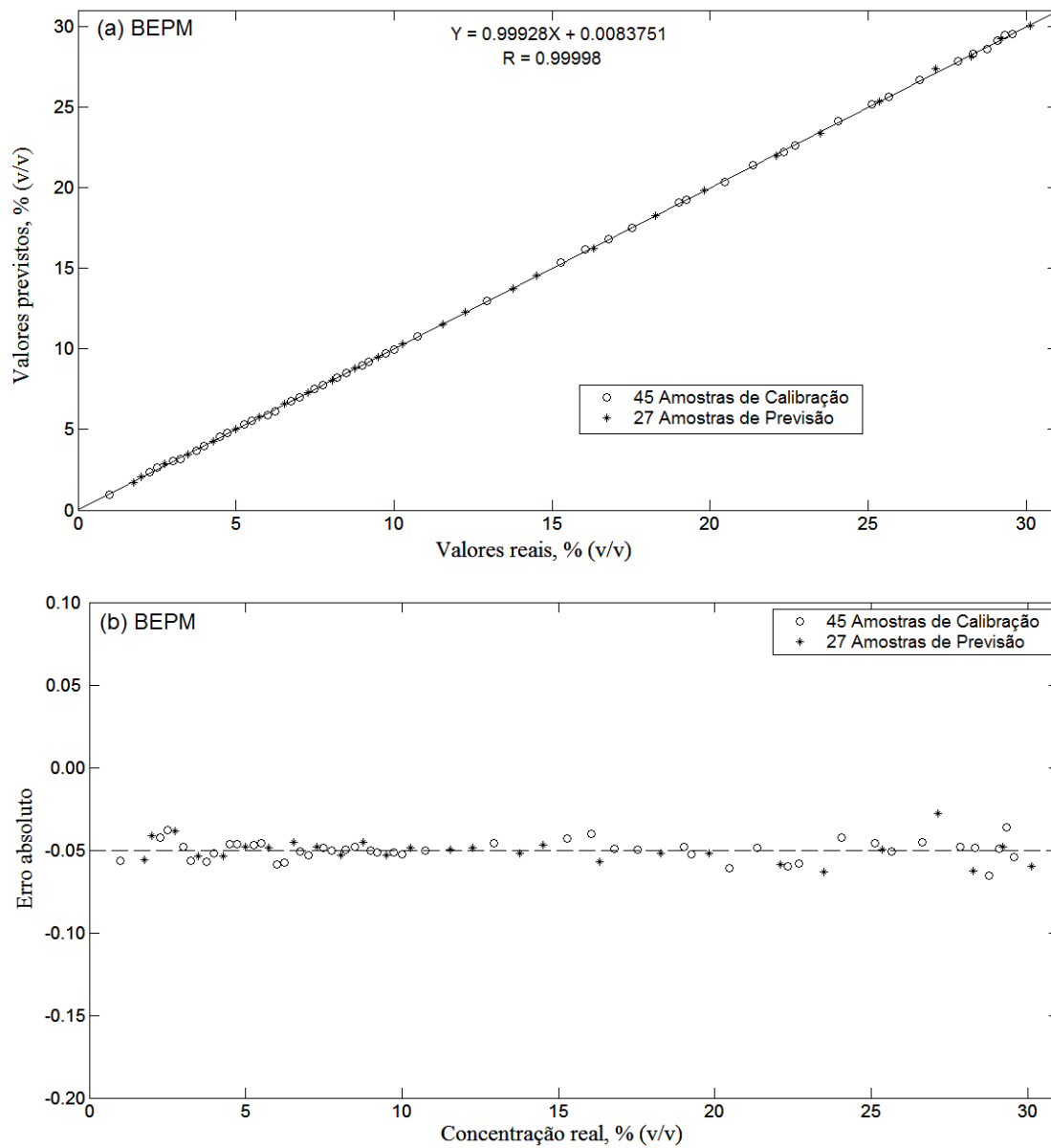
Fonte: O autor.

Figura 22 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BMPM



Fonte: O autor.

Figura 23 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo PLS global contendo BEPM



Fonte: O autor.

Em relação aos erros sistemáticos, a norma ASTM E1655-05, sugere a realização de teste-*t* para avaliar se o “*bias*” incluso no modelo é significativo. Em todos os modelos PLS verificou-se que $t_{calculado}$ é menor do que $t_{critico}$ (Tabela 4), para graus de liberdade (GL) igual ao número de amostras de previsão. Logo, todos os modelos não apresentaram erros sistemáticos e estão de acordo com as recomendações desta norma.

Tabela 4 – Valores de $t_{critico}$ e $t_{calculado}$ para avaliação da presença de erro sistemático dos modelos PLS

MODELO	BMS	BMSR	BMA	BMPM	BES	BESR	BEA	BEPM
$t_{critico}$	2,0555	2,0518	2,0518	2,0518	2,0518	2,0518	2,0595	2,0518
$t_{calculado}$	0,5359	0,3453	1,3750	0,7079	0,7696	1,2015	0,7123	0,2883

Fonte: O autor.

A Tabela 5, resume uma comparação dos modelos PLS global obtidos neste trabalho com alguns de caráter semelhante encontrados na literatura científica. Na construção do modelo PLS com 3 VLs, Knothe (2001) utilizou 38 amostras para o conjunto de calibração e apenas 9 para previsão e, conforme exigência da ASTM E1655, o número de amostras seria de 24 no conjunto de calibração e no mínimo 20 amostras para o conjunto de previsão. Pimentel et al., (2006) utilizaram 10 VLs na construção do modelo PLS o que exigiria 66 amostras no conjunto de calibração e 40 amostras no conjunto de previsão. Portanto, ambos não atenderam às exigências da ASTM E1655 em relação ao número de amostras utilizadas na construção dos modelos PLS global. Já os resultados de Ferrão et al., (2011) e Vasconcelos et al., (2012) apresentaram em conformidade com relação ao número de amostras tanto de calibração quanto de previsão, mas os valores de RMSEP para teores abaixo de 8% (v/v) não atendem aos requisitos da norma brasileira ABNT NBR 15568 que exige valor de RMSEP menor que 0,1% (v/v). Guerrero et al., (2013) avaliaram outro tipo de biodiesel, ou seja, biodiesel metílico de palma. No entanto, não seguiram os critérios de validação do modelo conforme exigências da ASTM E1655, mas propuseram uma faixa mais ampla para determinação de biodiesel de palma em diesel. No caso do presente estudo, os modelos PLS global propostos apresentaram-se em conformidade com as orientações da norma ASTM E1655-05, em relação ao número de amostras de calibração e validação e sem a presença de erros sistemáticos. Além disso, os modelos exibiram boa exatidão, estando de acordo com o exigido pela norma padrão ABNT NBR 15568, ou seja, valor de RMSEP menor do que 0,1% (v/v).

Tabela 5 – Comparação de resultados de modelos PLS global que empregaram a Espectrometria no Infravermelho Médio

Tipo de biodiesel	Método Quimiométrico	Intervalo espectral	Faixa de concentração	RMSEP	Validação ^(e)	Nº de Amostras ^(f)	Referência
BMS	PLS(3) ^(c)	650-4000	1 a 95%	----- ^(d)	X	C=38; P=9	KNOTHE (2001)
BMS, BMSR e BMM ^(a)	PLS(10)	650-4000	0 a 5%	0,25%	X	C=26; P=13	PIMENTEL et al., (2006)
BMS	PLS(5)	650-4000	0,2 a 30%	0,32	X	C=57; P=28	FERRÃO et al., (2011)
BMS	PLS(3)	650-4000	0 a 10%	0,21	X	C=84; P=45	VASCONCELOS et al., (2012)
BMP ^(b)	PLS(6)	651-4001	0-100%	0,74	X	C=57; P=19	GUERRERO et al., (2013)
BMS	PLS(4)	680-3100	1 a 30%	0,09	V	C=47; P=26	Presente estudo

(a) Biodiesel Metílico de Mamona; (b) Biodiesel Metílico de Palma; (c) o número entre parênteses refere-se ao número de VL utilizadas na construção do modelo PLS; (d) valor não informado pelo autor; (e) Validação conforme exigências da ASTM E1655, X = não realizaram a validação conforme esta norma e V = verificaram a validação; (f) C = número de amostras do conjunto de calibração e P = número de amostras do conjunto de previsão.

Fonte: o autor

4.3. Modelos de seleção de variáveis por intervalos (iPLS, biPLS e siPLS)

4.3.1. Modelos iPLS, biPLS e siPLS na quantificação de BMS em mistura com diesel

A Tabela 6 apresenta os resultados para os modelos iPLS obtidos na quantificação de BMS em mistura com diesel no qual o espectro total foi dividido em até 35 intervalos equidistantes em incrementos que variaram em 5 unidades. Os modelos foram avaliados em função dos valores de RMSEP, número de VLs e número de variáveis (NV).

Tabela 6 – Valores de RMSEP, número de variáveis e intervalos obtidos a partir dos modelos iPLS para quantificação de BMS em mistura com diesel

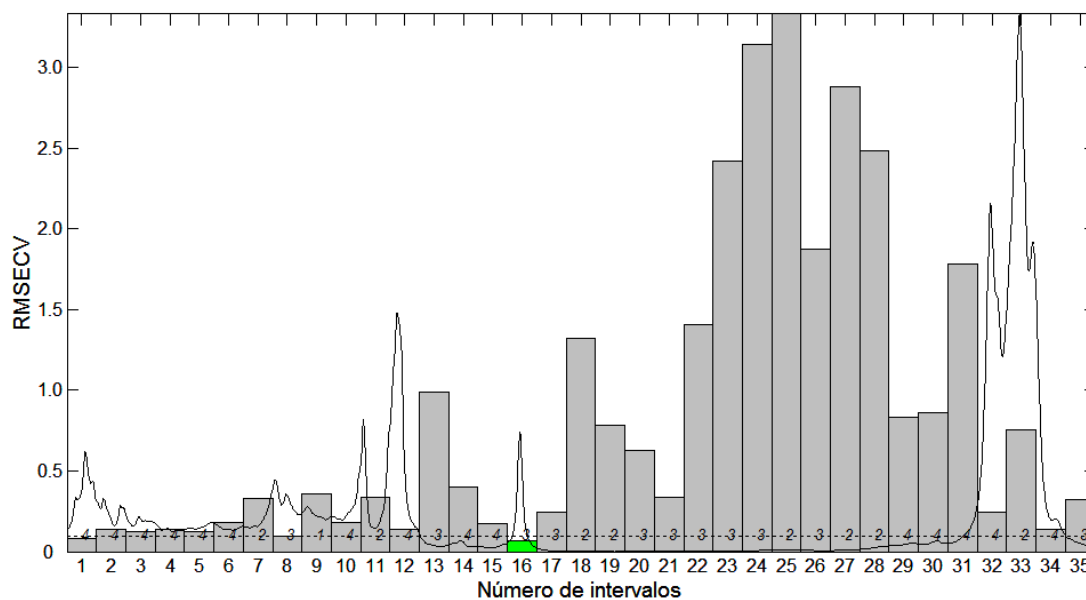
Modelo	NV	Intervalo (cm ⁻¹)	VLs	RMSEP	Valor de F*	Valor de F**
PLS global	2420	680-3100	4	0,0916	-----	3,8473
iPLS5	484	1650-2134	4	0,0627	2,1343	1,8026
iPLS10	242	1650-1892	4	0,0593	2,3861	1,6124
iPLS15	161	1166-1327	4	0,0467	3,8473	-----
iPLS20	121	1650-1771	4	0,0507	3,2642	1,1786
iPLS25	97	680-777	4	0,0530	2,9870	1,2880
iPLS30	81	680-761	4	0,0562	2,6566	1,4482
iPLS35	69	1722-1791	4	0,0488	3,5233	1,0920

* em relação ao modelo PLS global; ** em relação ao menor valor de RMSEP; ($F_{(0,05;25;25)} = 1,9555$).

Fonte: O autor

Aplicando-se o teste F, com a finalidade de comparar o modelo PLS global com os modelos de seleção de variáveis por intervalos, verificou-se que todos os modelos iPLS apresentaram diferença estatística entre os valores de RMSEP, consequentemente, todos os modelos iPLS apresentaram valores de erros de previsão estatisticamente melhores do que o modelo PLS global. Avaliando a diferença entre os modelos iPLS observou-se que não há diferença estatística entre eles. Como os modelos foram construídos com o mesmo número de VLs, decidiu-se pelo modelo que apresentou menor número de variáveis, neste caso, o modelo iPLS35. Este modelo utilizou o intervalo 16, conforme apresentado na Figura 24, que corresponde à região entre 1722 e 1791 cm⁻¹, referente a absorção da banda da carbonila (C=O).

Figura 24 – Espectro MIR dividido em 35 intervalos para a determinação de BMS em mistura com diesel. As barras correspondem aos valores de RMSECV para cada intervalo e a linha tracejada para o modelo PLS global. Os números internos apresentados no interior de cada intervalo correspondem ao número de variáveis latentes sugerido pelo algoritmo



Fonte: O autor

Na Tabela 7 estão apresentados os resultados obtidos para os modelos biPLS.

Tabela 7 – Valores de RMSEP, número de variáveis e intervalos obtidos a partir dos modelos biPLS para quantificação de BMS em mistura com diesel

Modelo	NV	Intervalo(s) (cm ⁻¹)	VLs	RMSEP	Valor de F*	Valor de F**
PLS global	2420	680-3100	4	0,0916	-----	4,4546
biPLS5	1452	680-2132	4	0,0657	1,9438	2,2917
biPLS10	242	1166-1408	4	0,0434	4,4546	-----
biPLS15	161	1166-1327	4	0,0467	3,8473	1,1579
biPLS20	242	1044-1286	4	0,0499	3,3697	1,3220
biPLS25	938	680-1456; 1554-1635; 1747-1828	4	0,0540	2,8774	1,5481
biPLS30	243	924-1005; 1166-1247; 1732-1813	4	0,0531	2,9758	1,4970
biPLS35	207	1029-1098; 1166-1235; 1721-1790	4	0,0459	3,9826	1,1185

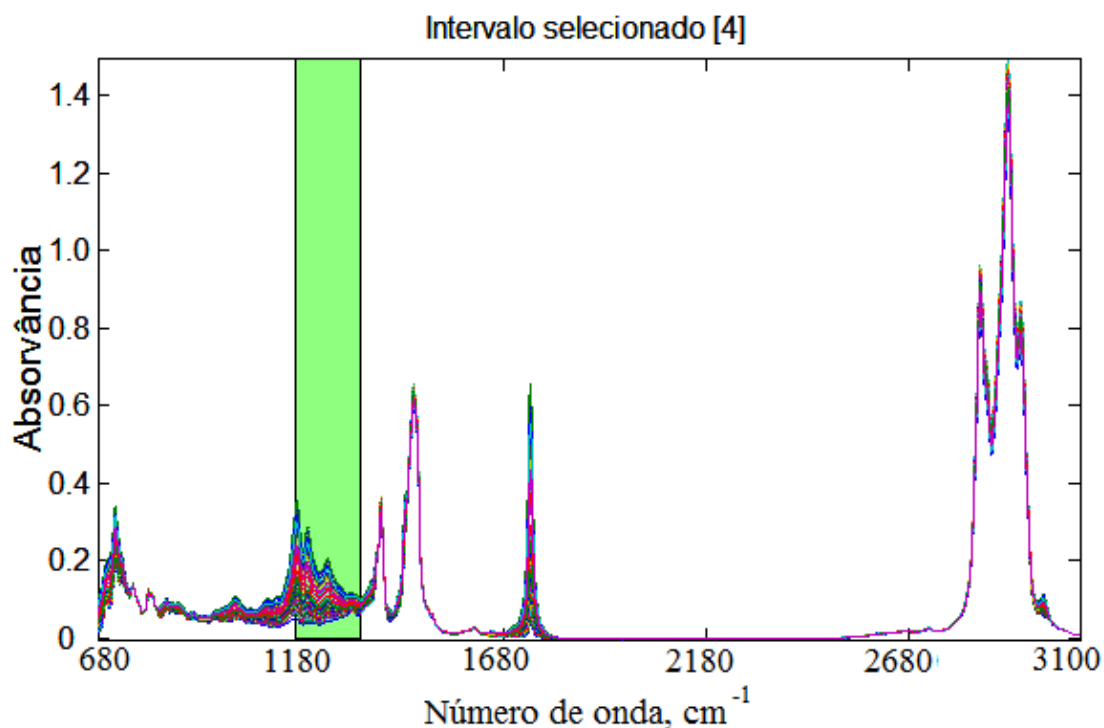
* em relação ao modelo PLS global; ** em relação ao menor valor de RMSEP;

(F_(0,05;25;25) = 1,9555).

Fonte: O autor.

Verifica-se, na Tabela 7, que todos os modelos biPLS, com exceção ao modelo biPLS5, apresentaram valores de erro de previsão estatisticamente melhores do que o modelo PLS global. No entanto, entre estes modelos biPLS, verificou-se que não há diferença estatística no parâmetro RMSEP. Assim, escolheu-se entre os demais modelos biPLS, o modelo com o menor número de variáveis, ou seja, biPLS15. A Figura 25 destaca a região selecionada por este modelo, que refere-se a bandas de 1000 cm⁻¹ a 1300 cm⁻¹ atribuídas às vibrações de estiramento da ligação C-O.

Figura 25 – Intervalo selecionado ($1166 - 1327 \text{ cm}^{-1}$) para construção do modelo PLS para quantificação de BMS em mistura com diesel obtida a partir do algoritmo biPLS15



Fonte: O autor.

Na Tabela 8 observa-se os resultados da estimativa dos valores de RMSECV para os modelos siPLS. Baseado no menor valor de RMSECV (valores em negrito) foi construído um modelo PLS em cada uma das combinações, ou seja, 2, 3 e 4.

Tabela 8 – Valores de RMSECV para cada modelo siPLS avaliado na quantificação de BMS em mistura com diesel

Modelo	RMSECV		
	Combinação 2	Combinação 3	Combinação 4
siPLS5	0,08502	0,07591	-----
siPLS10	0,06577	0,07249	0,07960
siPLS15	0,06008	0,05741	0,05801
siPLS20	0,06401	0,06042	0,05979
siPLS25	0,06254	0,05585	0,05406
siPLS30	0,06424	0,05987	0,05700
siPLS35	0,06075	0,05711	0,05559

Fonte: O autor.

Os resultados para os modelos siPLS com menor valor de RMSECV, dentro de cada combinação, estão apresentados na Tabela 9.

Tabela 9 – Valores de RMSEP, número de variáveis e intervalos obtidos a partir dos modelos siPLS para quantificação de BMS em mistura com diesel.

Modelo	NV	Intervalo(s) (cm^{-1})	VLs	RMSEP	Valor de F*	Valor de F**
PLS global	2420	680-3100	4	0,0916	-----	4,5804
si2PLS15	322	1166-1327; 2297-2458	4	0,0428	4,5804	-----
si3PLS25	291	875-972; 1166-1263; 1747-1844	4	0,0429	4,5591	1,0047
si4PLS25	388	875-972; 1166-1263; 1263-1360; 1747-1844	4	0,0467	3,8473	1,1905

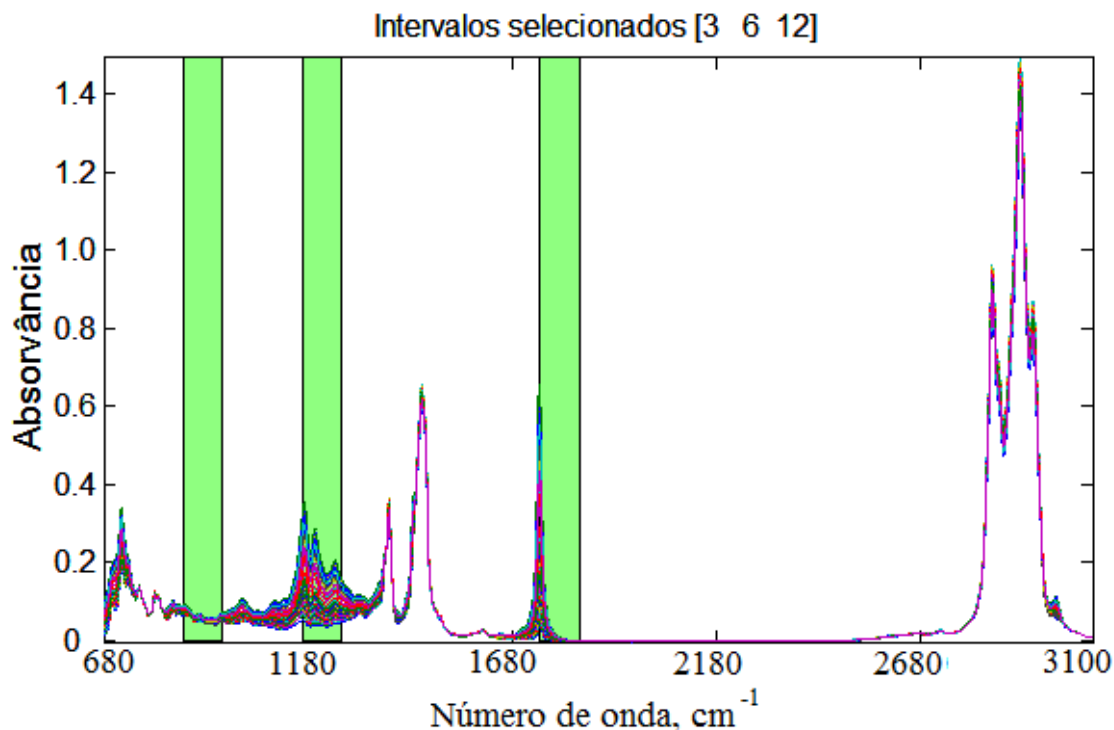
* em relação ao modelo PLS global; ** em relação ao menor valor de RMSEP;

($F_{(0,05;25;25)} = 1,9555$).

Fonte: o autor

Observa-se que todos os modelos siPLS apresentaram performance melhor do que o modelo PLS global. Entre os modelos siPLS verificou-se que não há diferença estatística. O modelo com o menor número de variáveis foi o si3PLS25. A Figura 26 ilustra as regiões selecionadas por este modelo que referem-se a absorção da carbonila (intervalo 12 de 1747-1844 cm^{-1}), do conjunto de bandas associadas com vibrações assimétrica das ligações C-C(=O)-O e O-C-C (intervalo 6 de 1166-1263 cm^{-1}); e sinal de baixa intensidade provavelmente referente a deflexões fora do plano do grupo C-O (intervalo 3 de 875-972 cm^{-1}).

Figura 26 – Intervalos selecionados (875-972, 1166-1263, 1747-1844 cm^{-1}) para construção do modelo PLS para quantificação de BMS em mistura com diesel obtida a partir do algoritmo si3PLS25



Fonte: O autor

A Tabela 10, resume uma comparação entre os melhores modelos dentro de cada um dos métodos de seleção de variáveis. Como pode ser observado não há diferença estatística entre os modelos de seleção de variáveis. Como todos foram construídos com o mesmo número de VLs foi escolhido o modelo iPLS35 por possuir o menor número de variáveis.

Tabela 10 – Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMS em mistura com diesel

Modelo	NV	Intervalo(s) (cm^{-1})	VLs	RMSEP	Valor de F*
iPLS35	69	1722-1791	4	0,0488	1,2940
biPLS15	161	1166-1327	4	0,0467	1,1850
si3PLS25	291	875-972; 1166-1263; 1747-1844	4	0,0429	-----

* em relação ao menor valor de RMSEP, $F_{(0,05;25;25)} = 1,9555$.

Fonte: O autor

Na literatura, Ferrão et al. (2011) encontrou que o melhor modelo para quantificar biodiesel metílico de soja foi o si4PLS8, composto por 1678 variáveis ($650\text{-}1909\text{ cm}^{-1}$ e $2746\text{-}3165\text{ cm}^{-1}$), construído com 5 VLs, o qual apresentou RMSEP de 0,16% (v/v). Portanto, neste estudo, obteve-se um modelo que indicou um subconjunto de dados com performance mais eficiente do que o modelo PLS global e o descrito na literatura.

Devido ao modelo iPLS35 apresentar exatidão melhor do que o modelo PLS global prosseguiu-se no estudo da validação deste modelo. A Tabela 11 apresenta os resultados dos parâmetros de exatidão, juntamente com a avaliação do teste t para verificação da presença de erro sistemático. A exatidão atende à exigência da norma brasileira ABNT NBR 15568 por apresentar valor de RMSEP menor que 0,1% e não apresentou erro sistemático no intervalo avaliado, ou seja, t_{bias} menor que $t_{\text{crítico}}$.

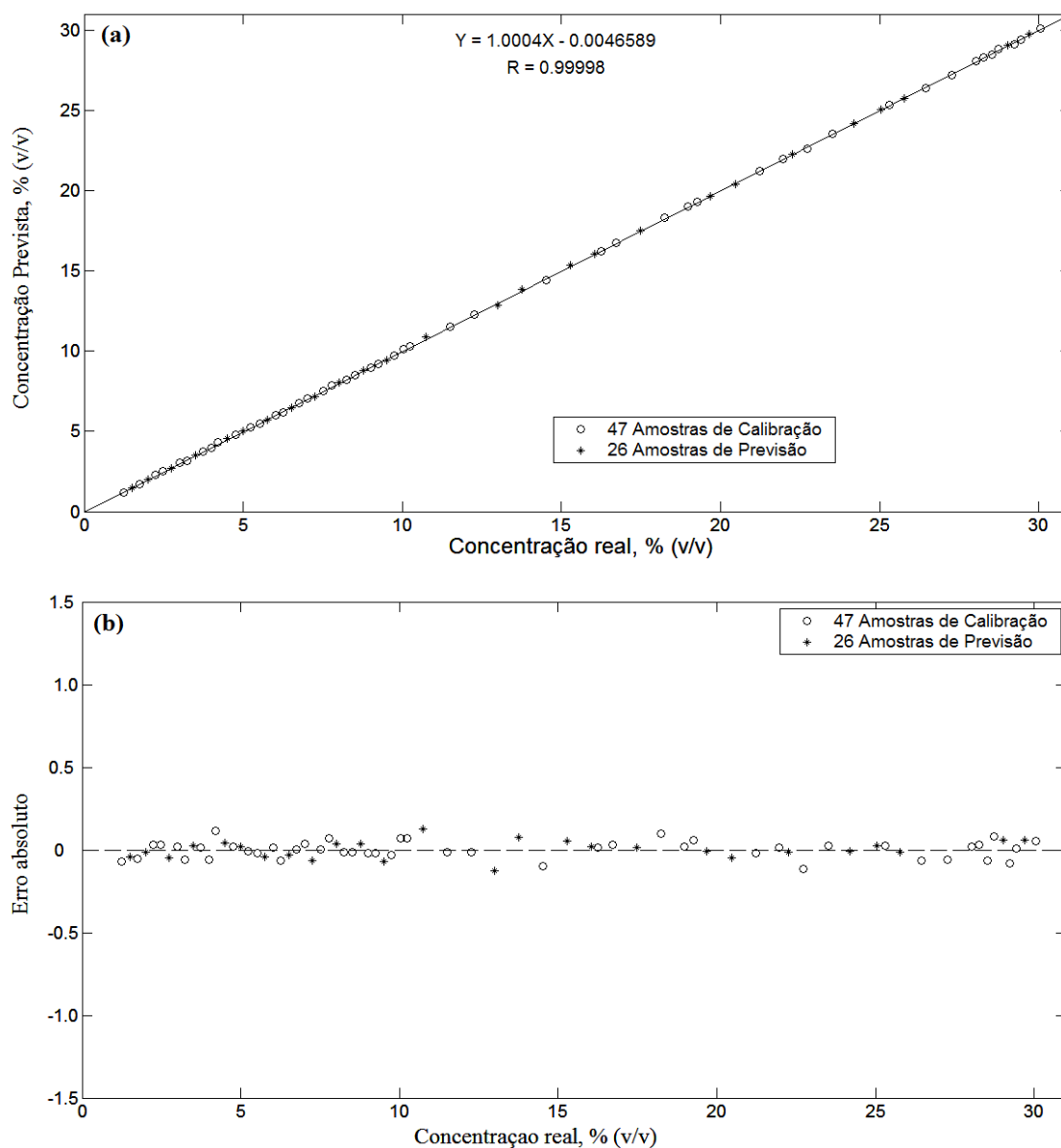
Tabela 11 – Parâmetros de exatidão e verificação da presença de erro sistemático do modelo iPLS35 na quantificação de BMS

Parâmetro	Valor
RMSECV	0,0677
RMSEC	0,0575
RMSEP	0,0488
t_{bias}	0,2307
$t_{\text{crítico}}$	2,0555

Fonte: O autor.

A linearidade e o ajuste estão apresentados na Figura 27. Verifica-se uma excelente correlação entre os valores medidos e os previstos pelo modelo iPLS além da distribuição aleatório dos resíduos, indicando que o modelo iPLS35 apresenta comportamento linear.

Figura 27 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo iPLS35 na quantificação de BMS



Fonte: O autor.

4.3.2. Modelos iPLS, biPLS e siPLS na quantificação de BES em mistura com diesel

Os procedimentos para a verificação do melhor modelo de seleção de variáveis por intervalos para os demais tipos de biodiesel foram realizados da mesma forma como descrito, passo a passo, na avaliação do BMS. Portanto, a partir deste momento serão descritos apenas o melhor modelo dentro de cada método de seleção de variáveis por intervalos (iPLS, biPLS e siPLS).

A Tabela 12 apresenta os resultados para os modelos PLS global, iPLS, biPLS e siPLS obtidos na quantificação de BES em mistura com diesel.

Tabela 12 – Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BES em mistura com diesel

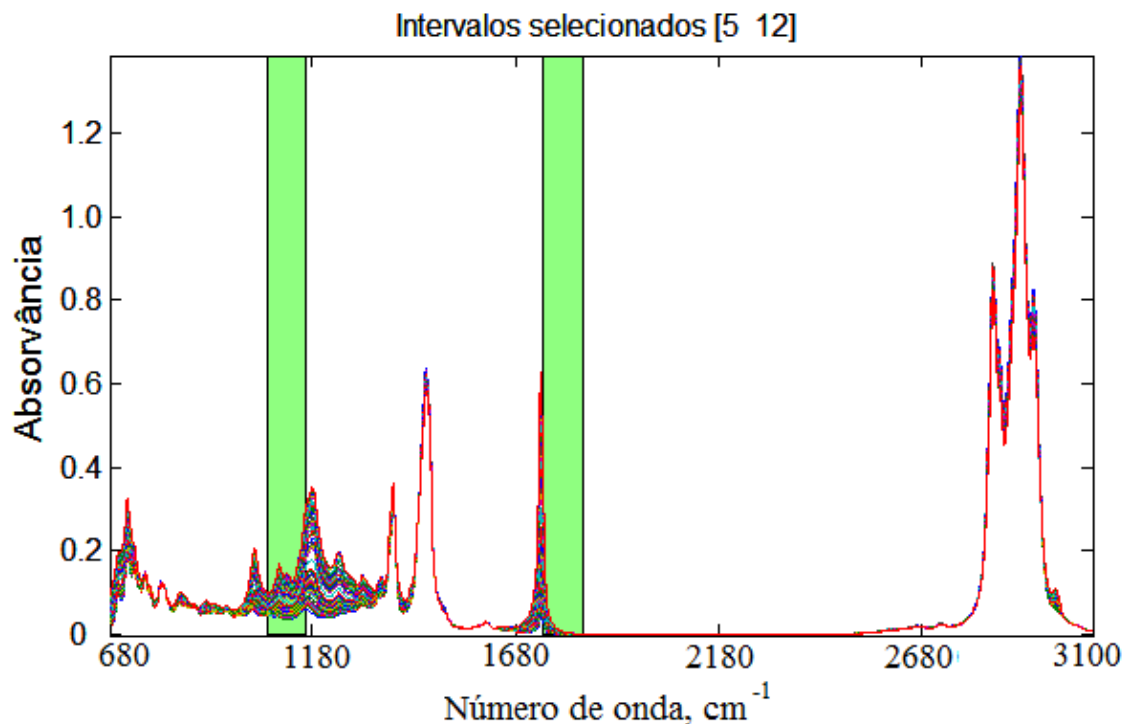
Modelo	NV	Intervalo(s) (cm ⁻¹)	VLs	RMSEP	Valor de F*
PLS global	2420	680-3100	4	0,0974	-----
iPLS25	97	1263-1360	3	0,0784	1,5434
biPLS30	405	762-843; 1166-1247; 1409-1490; 1731-1812; 2780-2861	4	0,0812	1,4388
si2PLS25	194	1069-1166; 1748-1845	3	0,0599	2,6440

* em relação ao modelo PLS global, $F_{(0,05;26;26)} = 1,9292$.

Fonte: O autor.

Verifica-se na Tabela 12 que apenas o modelo si2PLS25 apresentou diferença estatística em relação ao valor de RMSEP do modelo PLS global. A Figura 28 ilustra os intervalos selecionados pelo algoritmo si2PLS25. Estes intervalos foram a combinação dos intervalos 5 e 12 que correspondem às regiões de 1069-1166 cm⁻¹ e 1748-1845 cm⁻¹ (Figura 28), referentes a vibrações de ligações C-O e C=O, respectivamente. Estas regiões são semelhantes às regiões sugeridas pela norma ABNT NBR15568, principalmente a região 5. É importante ressaltar que esta norma utiliza as mesmas regiões para quantificar qualquer tipo de biodiesel, mas aplicando os métodos de seleção de variáveis percebe-se até o momento que dependendo do tipo de biodiesel há sugestões de regiões ou intervalos diferentes indicados pelo algoritmo si2PLS25.

Figura 28 – Intervalos selecionados (1069-1166 e 1748-1845 cm^{-1}) para construção do modelo PLS para quantificação de BES em mistura com diesel obtidas a partir do algoritmo si2PLS25



Fonte: O autor.

Os parâmetros de exatidão obtidos do modelo si2PLS25 na quantificação de BES em mistura com diesel estão apresentados na Tabela 13. Verifica-se que os valores de RMSEP atendem aos requisitos da ABNT NBR 15568, ou seja, RMSEP menor que 0,1% e, que neste modelo não há a presença de erro sistemático ($t_{\text{bias}} < t_{\text{crítico}}$).

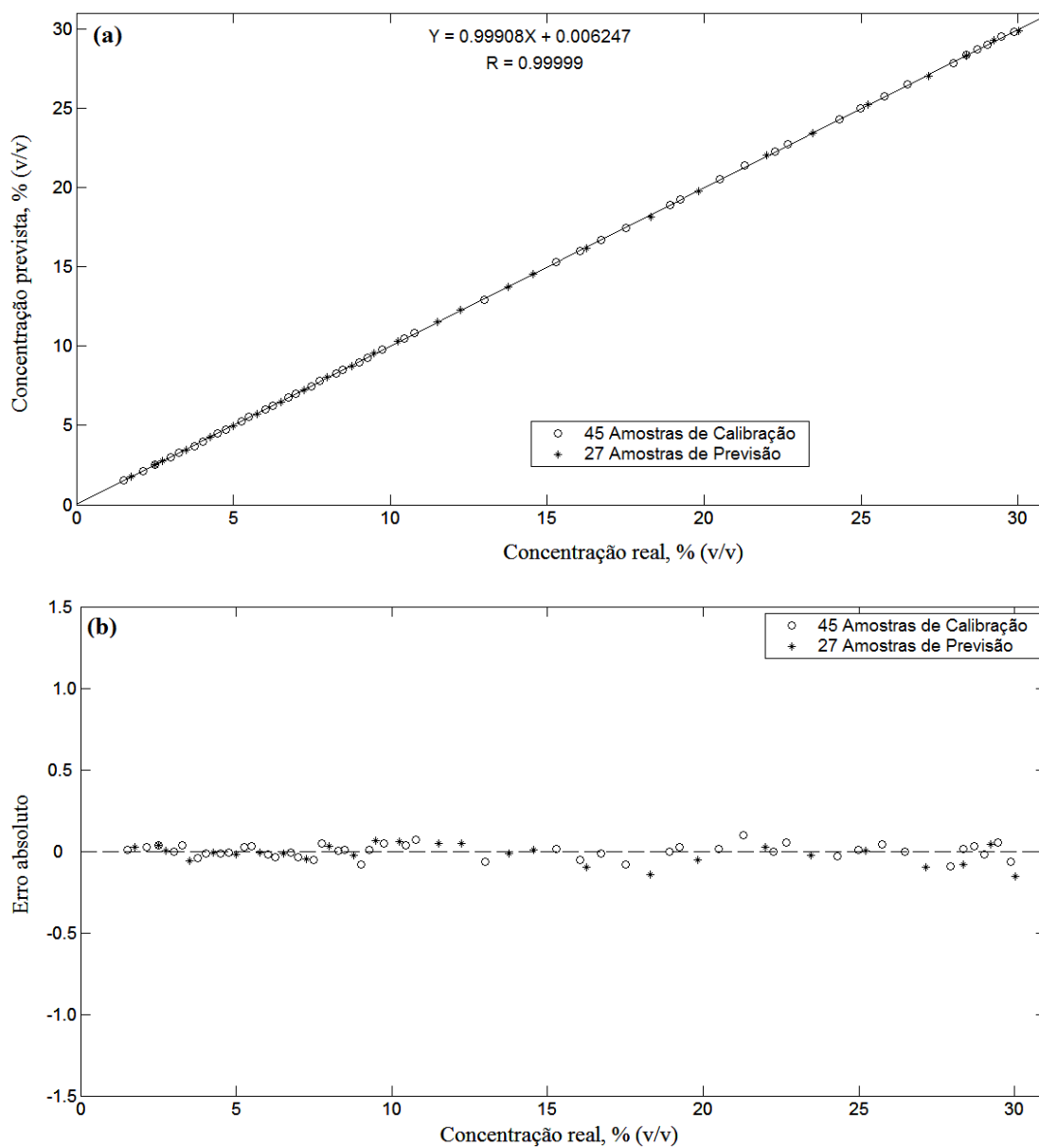
Tabela 13 – Parâmetros de exatidão e verificação da presença de erro sistemático do modelo si2PLS25 na quantificação de BES

Parâmetro	Valor
RMSECV	0,0479
RMSEC	0,0416
RMSEP	0,0599
t_{bias}	1,5188
$t_{\text{crítico}}$	2,0518

Fonte: O autor.

Em relação à linearidade (Figura 29a) verificou-se uma excelente correlação entre os valores medidos e os previstos pelo modelo si2PLS25 além da distribuição aleatória dos resíduos (Figura 29b), indicando que o modelo si2PLS25 apresenta comportamento linear, consequentemente, pode ser aplicado na quantificação de BES em mistura com diesel.

Figura 29 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si2PLS25 na quantificação de BES



Fonte: O autor.

4.3.3. Modelos iPLS, biPLS e siPLS na quantificação de BMSR e BESR em mistura com diesel

Os resultados para os modelos PLS global e os melhores modelos de seleção de variáveis por intervalos iPLS, biPLS e siPLS obtidos na quantificação de BMSR e BESR estão apresentados na Tabela 14.

Tabela 14 – Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMSR e BESR em mistura com diesel

Tipo de biodiesel	Modelo	NV	Intervalo(s) (cm ⁻¹)	VLs	RMSEP	Valor de F*
BMSR	PLS global	2420	680-3100	4	0,1020	-----
	iPLS35	97	1722-1791	4	0,1089	1,1399
	biPLS**	-----	-----	--	-----	-----
	si2PLS25	194	1166-1263; 1748-1845	4	0,0771	1,7502
BESR	PLS global	2420	680-3100	4	0,0483	-----
	iPLS30	81	1005-1086	4	0,0633	1,7176
	biPLS30	405	762-843; 1166-1247; 1409-1490; 1731-1812; 2780-2861	4	0,0349	1,9153
	si3PLS25	291	972-1166; 1746-1843	4	0,0565	1,3684

* em relação ao modelo PLS global, $F_{(0,05;26;26)} = 1,9292$; ** Não foram encontradas combinações de intervalos, os resultados foram iguais ao iPLS.

Fonte: O autor.

Para estes tipos de biodieseis não foram encontrados intervalos que apresentassem melhoras significativas no parâmetro de exatidão. Desta maneira, os métodos de seleção de variáveis por intervalos não encontraram intervalos que justificassem sua escolha. Portanto, o melhor modelo foi o PLS global, cuja validação está apresentada no item 4.2.

4.3.4. Modelos iPLS, biPLS e siPLS na quantificação de BMA e BEA em mistura com diesel

A Tabela 15 apresenta os resultados para os modelos PLS global e os melhores modelos de seleção de variáveis por intervalos iPLS, biPLS e siPLS obtidos na quantificação de BMA e BEA.

Tabela 15 – Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMA e BEA em mistura com diesel

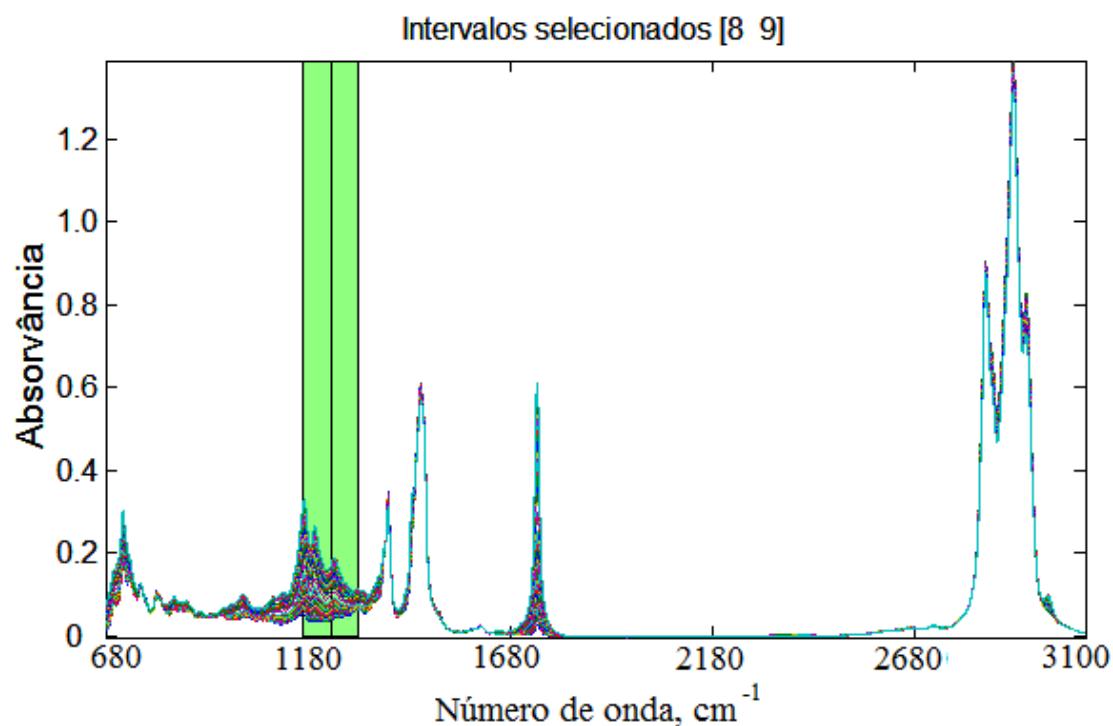
Tipo de biodiesel	Modelo	NV	Intervalo(s) (cm^{-1})	VLs	RMSEP	Valor de F*
BMA	PLS global	2420	680-3100	4	0,0473	-----
	iPLS20	121	1166-1287	4	0,0456	1,0760
	biPLS25	582	680-971; 1166-1263; 1360-1457; 1748-1845	3	0,0376	1,5825
	si2PLS35	138	1166-1304	3	0,0463	1,0437
BEA**	PLS global	2420	680-3100	3	0,0751	-----
	iPLS25	97	1263-1360	3	0,0546	1,8919
	biPLS25	388	1166-1457; 1746-1843	3	0,0460	2,6654
	si3PLS35	207	1166-1235; 1444-1513; 1720-1789	3	0,0394	3,6332

* em relação ao modelo PLS global, $F_{(0,05;26;26)} = 1,9292$; ** O valor de F para BEA é $F_{(0,05;24;24)} = 1,9838$

Fonte: O autor.

Avaliando-se os resultados apresentados na Tabela 15 verifica-se que para os modelos contendo BMA não existem diferenças estatísticas para o parâmetro RMSEP. No entanto, os modelos biPLS e siPLS foram construídos com uma dimensão a menos do que os modelos PLS global e iPLS, consequentemente, são modelos teoricamente mais simples. Aplicando-se o teste F entre os modelos biPLS25 e si2PLS35 verificou-se que não há diferença estatística entre os valores de RMSEP. Assim, como o modelo si2PLS35 apresentou menor número de variáveis optou-se como o melhor modelo. A Figura 30 ilustra os intervalos selecionados 8 e 9 pelo modelo si2PLS35 correspondentes ao conjunto de bandas associadas com as vibrações assimétrica das ligações C-C(=O)-O e O-C-C ($1166-1304 \text{ cm}^{-1}$).

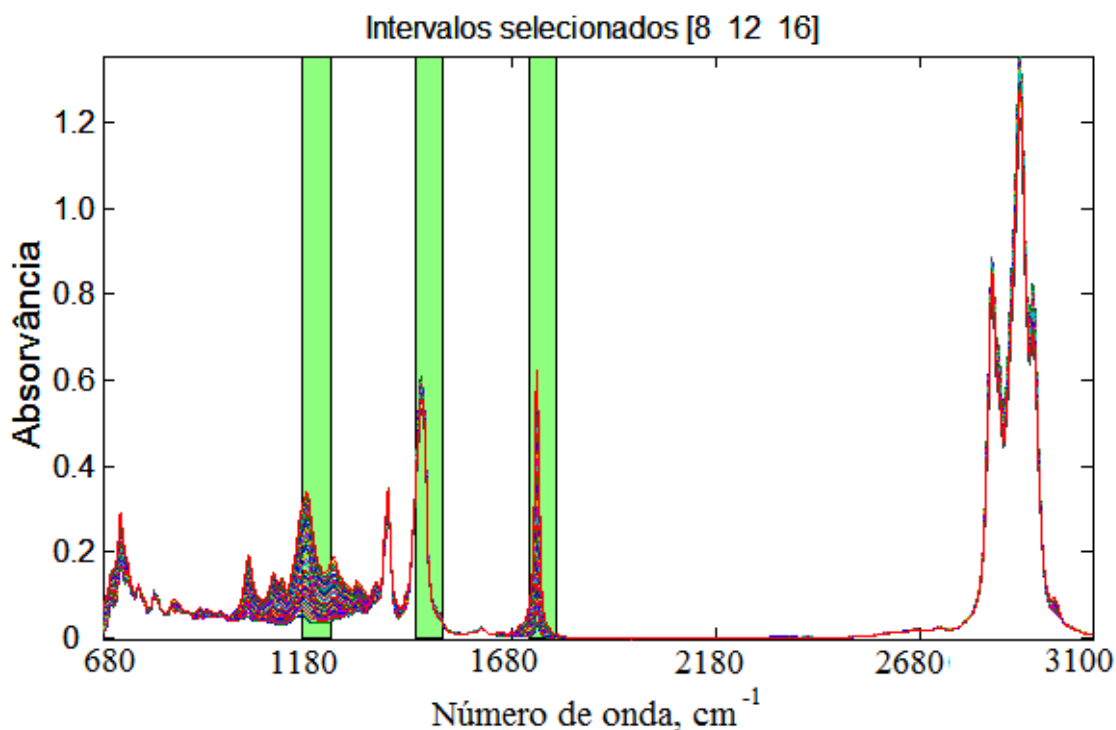
Figura 30 – Intervalos selecionados (1166-1234 e 1235-1304 cm^{-1}) para construção do modelo PLS para quantificação de BMA em mistura com diesel obtidas a partir do algoritmo si2PLS35



Fonte: O autor.

No caso dos modelos contendo BEA verifica-se que todos foram construídos com o mesmo número de VLs, mas os modelos biPLS25 e si3PLS35 apresentaram melhores valores de RMSEP ($F_{\text{calculado}} > F_{\text{tabelado}}$). Entre estes modelos não há diferença estatística nos valores de RMSEP, mas o modelo si3PLS35 contém menor número de variáveis. A Figura 31 demonstra os intervalos selecionados pelo modelo si3PLS35. Estes intervalos referem-se a região de vibração de deformação axial da ligação C-O (intervalo 8 de 1166-1235 cm^{-1}); estiramento assimétrico da ligação C-H (intervalo 12 de 1444-1513 cm^{-1}) e banda da carbonila (intervalo 16 de 1720-1789 cm^{-1}).

Figura 31 – Intervalos selecionados (1166-1235, 1444-1513 e 1720-1789 cm^{-1}) para construção do modelo PLS para quantificação de BEA em mistura com diesel obtidas a partir do algoritmo si3PLS35



Fonte: O autor.

Na Tabela 16 estão apresentados os parâmetros de exatidão e os dados para verificação da presença de erro sistemático para os modelos si2PLS35 e si3PLS35, respectivamente, para os modelos contendo BMA e BEA. Verifica-se que ambos os modelos cumprem com o exigido pela norma ABNT NBR 15568 e não apresentam erros sistemáticos ($t_{\text{bias}} < t_{\text{crítico}}$).

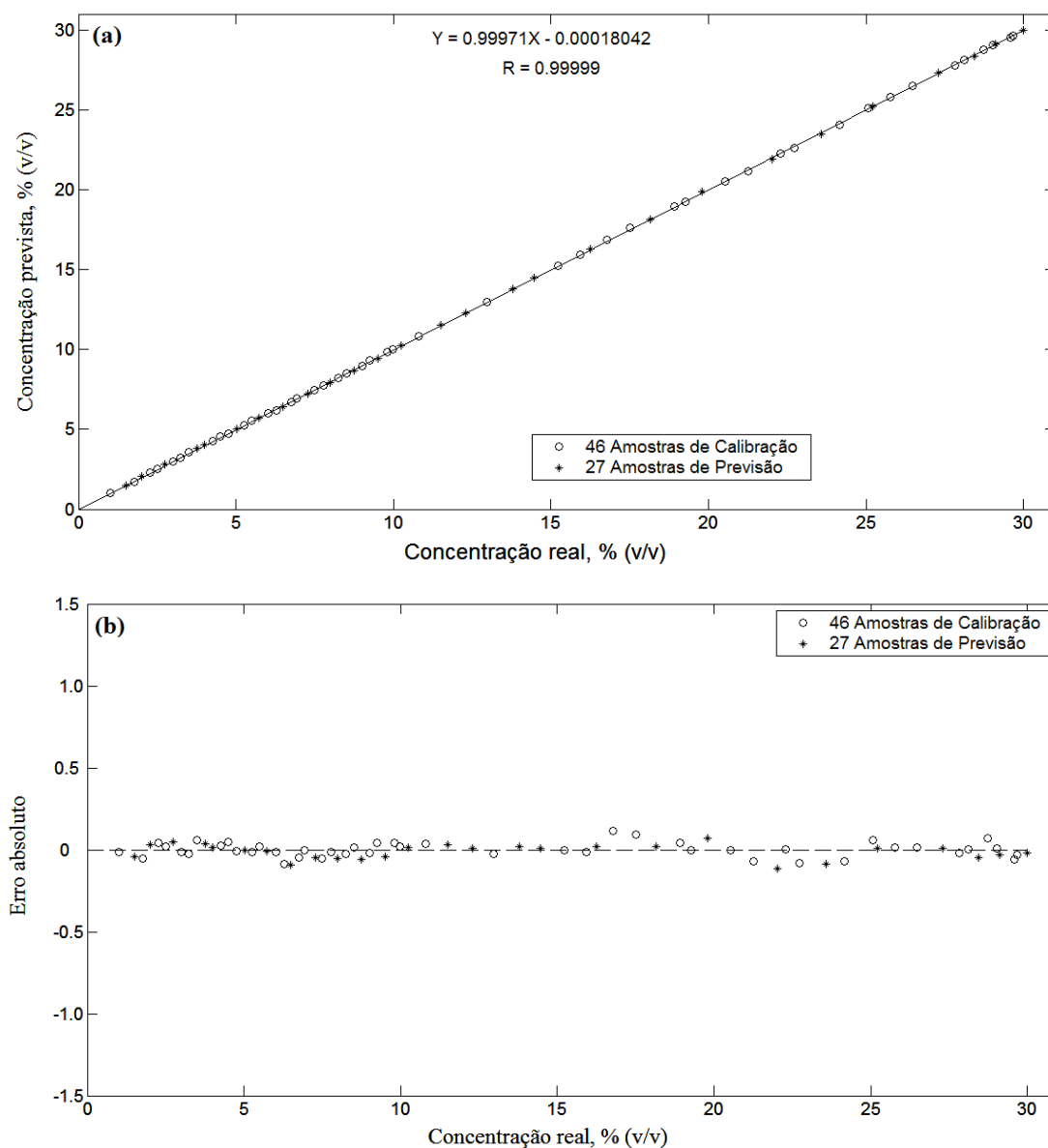
Tabela 16 – Parâmetros de exatidão e verificação da presença de erro sistemático dos modelos si2PLS35 e si3PLS35 na quantificação de BMA e BEA, respectivamente

Parâmetro	BMA	BEA
	si2PLS35	si3PLS35
RMSECV	0,0494	0,0539
RMSEC	0,0435	0,0356
RMSEP	0,0463	0,0394
t_{bias}	1,2548	1,3777
$t_{\text{crítico}}$	2,0518	2,0595

Fonte: O autor.

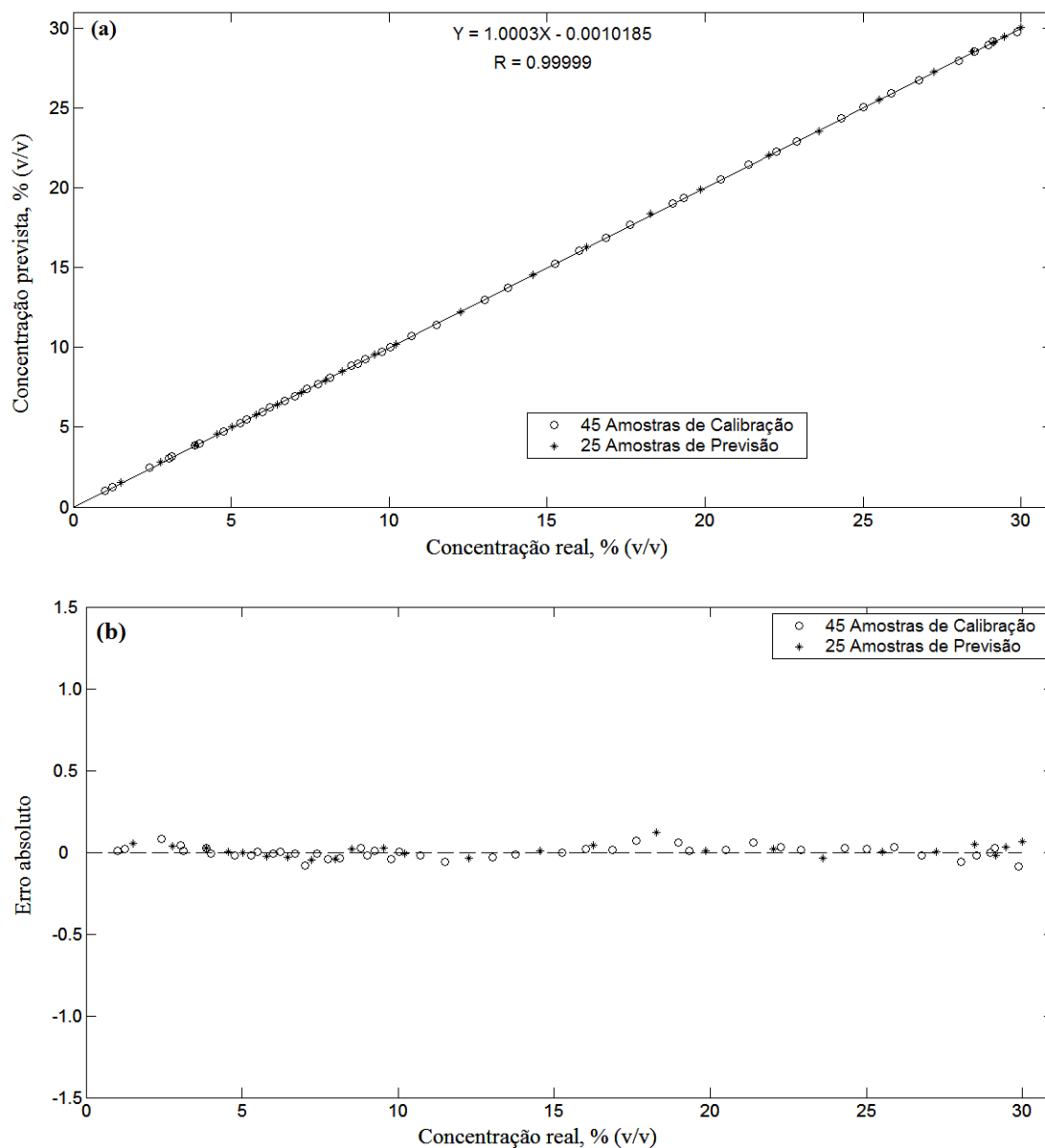
Além disso, ambos os modelos apresentaram comportamento linear devido a excelente correlação entre os valores reais e previsto e distribuição aleatória dos resíduos (Figuras 32 e 33).

Figura 32 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si2PLS35 na quantificação de BMA



Fonte: O autor.

Figura 33 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si3PLS35 na quantificação de BEA



Fonte: O autor.

4.3.5. Modelos iPLS, biPLS e siPLS na quantificação de BMPM e BEPM em mistura com diesel

Os resultados da aplicação dos métodos de seleção de variáveis por intervalos na quantificação de BMPM e BEPM estão apresentados na Tabela 17.

Tabela 17 – Valores de RMSEP, número de variáveis e intervalos obtidos dentre os melhores modelos de seleção de variáveis por intervalos para quantificação de BMPM e BEPM em mistura com diesel

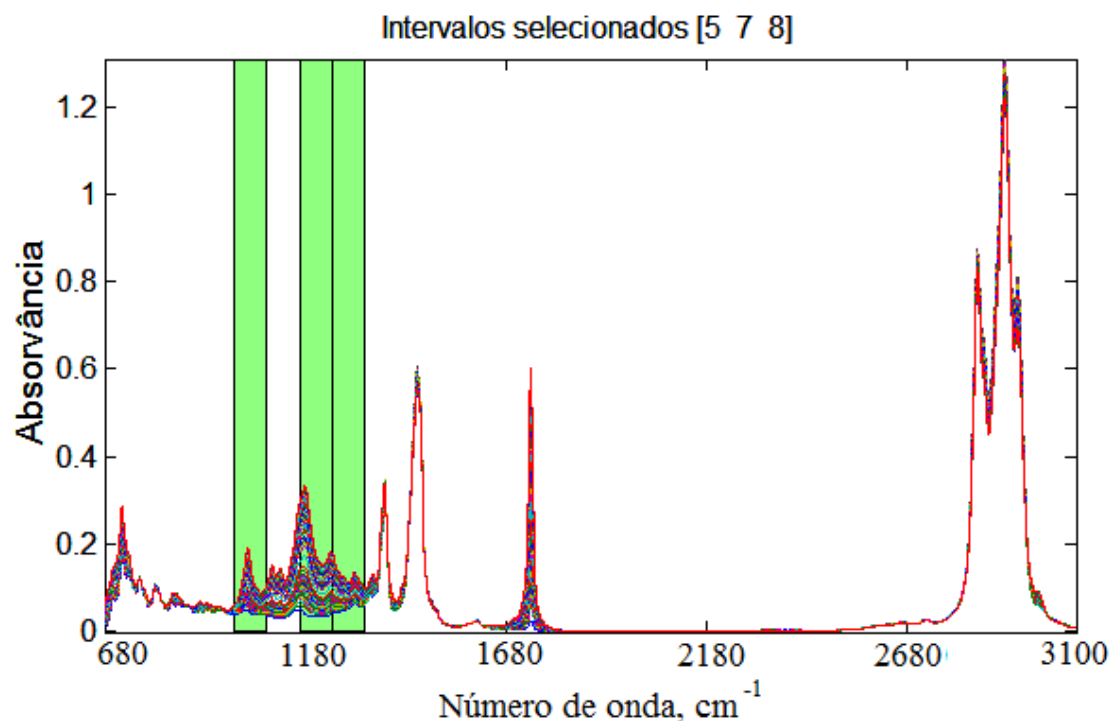
Tipo de biodiesel	Modelo	NV	Intervalo(s) (cm ⁻¹)	VLs	RMSEP	Valor de F*
BMPM	PLS global	2420	680-3100	3	0,0602	-----
	iPLS20	121	680-801	3	0,0571	1,1115
	biPLS25	679	680-971; 1069-1263; 1457-1554; 1748-1845	3	0,0439	1,8805
	si3PLS25	291	875-972; 1166-1263; 1748-1845	4	0,0523	1,3249
BEPM	PLS global	2420	680-3100	4	0,0723	-----
	iPLS30	81	1005-1086	4	0,0546	1,7534
	biPLS25	485	680-874; 1166-1360; 1747-1844	4	0,0349	4,2917
	si3PLS30	243	1004-1085, 1166-1247; 1248-1329	4	0,0402	3,2346

* em relação ao modelo PLS global, $F_{(0,05;26;26)} = 1,9292$;

Fonte: O autor.

Avaliando os resultados apresentados na Tabela 17 verifica-se que a aplicação dos métodos de seleção de variáveis na quantificação de BMPM não ocorreram melhoras significativas no parâmetro de RMSEP, portanto, não há seleção de intervalos que justifiquem a sua utilização. Por outro lado, na quantificação de BEPM verifica-se que os modelos biPLS25 e si3PLS30 apresentaram valores de RMSEP estatisticamente melhores do que o modelo PLS global. Entre estes dois modelos não há diferença estatística, portanto, escolheu-se o modelo com o menor número de variáveis como o de melhor performance para realizar a sua validação, ou seja, si3PLS30. A Figura 34 ilustra os intervalos 5, 7 e 8 selecionados por este modelo, relacionadas às vibrações de deformação axial da ligação C-O.

Figura 34 – Intervalos selecionados (1004-1085, 1166-1247 e 1248-1329 cm^{-1}) para construção do modelo PLS para quantificação de BEPM em mistura com diesel obtidas a partir do algoritmo si3PLS30



Fonte: O autor.

Os parâmetros de exatidão e os dados para verificação da presença de erro sistemático para o modelo si3PLS30 estão apresentados na Tabela 18. Verifica-se que o modelo atende com o exigido pela norma ABNT NBR 15568 e não apresenta erro sistemático ($t_{\text{bias}} < t_{\text{crítico}}$).

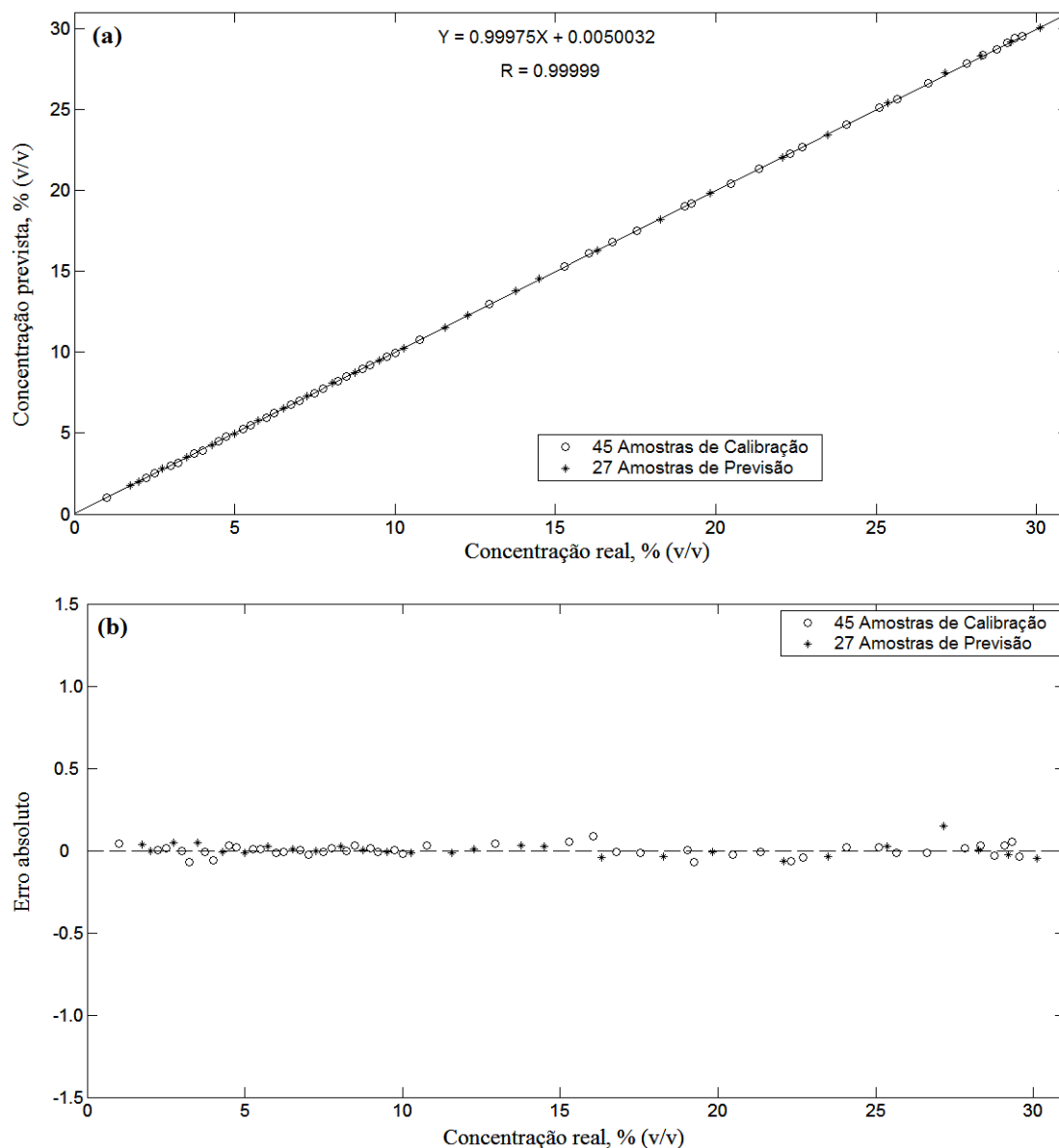
Tabela 18 – Parâmetros de exatidão e verificação da presença de erro sistemático do modelo si3PLS30 na quantificação de BEA

Parâmetro	Valor
RMSECV	0,0384
RMSEC	0,0328
RMSEP	0,0402
t_{bias}	0,5104
$t_{\text{crítico}}$	2,0518

Fonte: O autor.

A Figura 35 demonstra que o modelo si3PLS30 apresentou comportamento linear visto que há uma excelente correlação entre os valores reais e previsto e distribuição aleatória dos resíduos.

Figura 35 – (a) Gráfico de ajuste e (b) gráfico de resíduo do modelo si3PLS30 na quantificação de BEPM



Fonte: O autor.

A Tabela 19 resume resultados de outras técnicas e de diferentes métodos quimiométricos que também tiveram como objetivo a quantificação de biodiesel em mistura com diesel.

Tabela 19 – Resultados de diferentes técnicas e métodos quimiométricos obtidos na quantificação de biodiesel em misturas com diesel

Tipo de biodiesel	Técnica	Método Quimiométrico	Número de Variáveis	Faixa de concentração	RMSEP	Validação	Nº de Amostras	Referência
BMS	NIR	SPA-MLR	2	5-50%	0,64	✓	C=60	FERNANDES et al., (2011)
	VIS	SPA-MLR	8		0,57		P=40	
	VIS-NIR	SW-MLR	5		0,95			
BMS	NIR	SPA-MLR	15	0-10%	0,09	✓	C=84	VASCONCELOS et al., (2012)
	MIR	SPA-MLR	9		0,18		P=45	
BMS	MIR	si4PLS8 (5)	1678	0,2-30%	0,16	X	C=57 P=28	FERRÃO et al., (2011)
Diversos	VIS	Univariado (435 nm)	1	0-5; 5-10; 11-15 e 16-20%	-----	-----	-----	SILVA et al., (2015)
BMS	Fluorescência	PLS(6)	529	0-15%	0,37	✓	C=40	INSAUSTI et al., (2013)
		SPA-MLR	14		0,58		P=30	
Diversos	¹⁴ C	Univariado	1	2-20%	-----	-----	-----	NORTON; CLINE; THOMPSON (2012)
BMPM	CG-MS	PLS(5)	Não informado	10-30%	1,4	X	C=26	SCHALE; LE;
BMS		PLS(4)			1,2		P=10	PIERCE (2012)

Observa-se que Ferrão et al., (2011) também aplicaram os métodos de seleção de variáveis por intervalos (iPLS e siPLS) para quantificação de biodiesel metílico de soja em mistura com diesel. No entanto, o melhor modelo foi o si4PLS8 totalizando 1678 variáveis. Estas variáveis foram correspondentes, na região do MIR, a todos os intervalos que apresentaram sinal, ou seja, de 650 a 1909 cm^{-1} e 2746 a 3165 cm^{-1} . No caso do presente estudo, para o modelo BMS, foi possível uma redução de 2420 variáveis para 69 variáveis correspondentes a região de absorção do grupo carbonila.

Vasconcelos et al., (2012) empregando MIR porém aplicando Algoritmo das Projeções Sucessivas (SPA) conseguiram propor um modelo com 9 variáveis o qual aplicando a Regressão Linear Múltipla (MLR) obteve RMSEP de 0,18% (v/v). Porém do ponto de vista da exigência da norma brasileira para a faixa de 0 a 10,00% (v/v) este valor deveria ser menor que 0,1% (v/v). Fernandes et al., (2011) avaliaram a utilização da região no Infravermelho Próximo, conseguindo uma excelente redução de variáveis aplicando SPA e Stepwise (SW) para uma faixa de concentração maior que 5,00%. Se comparado a norma brasileira, para o modelo com concentração maior que 8,00% cuja exigência é de RMSEP maior que 1%, sua metodologia apresentou bom desempenho.

Na literatura também pode ser encontrado a quantificação de biodiesel em diesel por outras técnicas como a espectrofotometria no visível (SILVA et al., 2015), fluorescência (INSAUSTI et al., 2013), ^{14}C (NORTON; CLINE; THOMPSON, 2012), CG-MS (SCHALE; LE; PIERCE, 2012), entre outras (MONTEIRO et al., 2009; ALEME; BARBEIRA, 2012).

Silva et al., (2015) propuseram a quantificação de biodiesel em diesel através da medida de absorção em 435 nm do complexo formado entre hidroxamato de sódio (obtido da reação entre o éster e hidrocloreto de hidroxilamina – H_2NOHHCl), em meio ácido e íons férrico. Assim, foi avaliado a quantificação de biodieseis provenientes de diferentes matérias primas (milho, amendoim, pinhão-manso, óleo de fritura usado, soja, girassol e mamona) sendo que apenas mamona apresentou diferença na quantificação para o mesmo teor que as demais oleaginosas. Mas por outro lado se comparado à Espectrometria no Infravermelho há gasto de reagentes específicos, o procedimento é mais oneroso, além da falta de exatidão para uma faixa mais ampla, ou seja, exige-se a construção de 4 curvas analíticas no intervalo entre 0 e 20,00% (v/v).

Na mesma ideia, porém empregando outra técnica, Norton, Cline e Thompson (2012) usaram a análise de ^{14}C para quantificar biodieseis de diferentes matérias primas (soja, canola, sebo e coco) em diesel e compararam a metodologia proposta com a norma americana ASTM D7371. Eles verificaram que a utilização da referida norma apresentou erros significativos para

a quantificação de biodieseis provenientes de diferentes fontes, principalmente do óleo de coco. Por outro lado, a metodologia proposta pelos autores embora apresentando excelentes resultados para diferentes matérias primas, utiliza uma técnica onerosa e o tempo de análise para esta quantificação foi estimado em 6 horas.

Schale, Le e Pierce (2012) quantificaram e classificaram diferentes biodieseis empregando CG-MS aliado a quimiometria. Os modelos PLS para biodiesel metílico de pinhão manso e biodiesel metílico de soja apresentaram RMSEP de 1,40 e 1,20% (v/v), respectivamente. Isto demonstra que esta técnica não conseguiu exatidão adequada para a faixa de concentração de biodiesel de 10,00 a 30,00% em relação às exigências da norma brasileira ABNT NBR 15568. Além disso, esta técnica quando comparada ao MIR tem elevado custo e não permite a análise in loco devido às exigências de estruturas para instalação do mesmo. Insausti et al., (2013) utilizaram a Espectrofotometria de Fluorescência. Porém, os modelos propostos tanto globais quanto empregando seleção de variáveis (SPA-MLR) não apresentaram exatidão requerida pela ABNT NBR 15568.

Assim diante do exposto a técnica MIR aliada aos métodos quimiométricos, quando comparada a outras técnicas, apresenta vantagens expressivas para possibilitar as análises necessárias para o controle de qualidade e fiscalização destes tipos de combustíveis, atendendo as exigências das normas nacionais e internacionais.

5. CONCLUSÕES

5. CONCLUSÕES

A metodologia desenvolvida utilizando toda a região espectral (680 a 3100 cm^{-1}) no MIR combinada com a calibração multivariada por PLS mostrou-se eficiente para a determinação dos teores de biodieseis metílico e etílicos provenientes de diversas matérias primas em misturas binárias com diesel na faixa de concentração de $1,00$ a $30,00\%$ (v/v). Além disso, os modelos propostos foram validados conforme orientações da norma ASTM E1655 e atenderam às exigências da norma ABNT NBR 15568 em relação ao parâmetro de exatidão, apresentando valor de RMSEP na ordem de 10 vezes menor do que o exigido. Portanto, os modelos PLS foram construídos e validados os quais apresentaram boa exatidão ($\text{RMSEP} < 0,1\%$ v/v), excelente linearidade e não apresentaram erros sistemáticos.

Em relação a avaliação da aplicação dos métodos de seleção de variáveis por intervalos verificou-se que na maioria dos casos estudados (com exceção aos modelos contendo biodieseis provenientes de óleo de fritura usado e biodiesel metílico de pinhão manso), foram eficientes na eliminação de variáveis não informativas, melhorando a previsão de amostras de validação, se comparado ao desempenho do modelo de calibração multivariada por PLS. Todos os modelos de seleção de variáveis que apresentaram estatisticamente erros de previsão menores do que o modelo PLS global foram validados.

De forma geral, este trabalho demonstrou que para cada tipo de biodiesel existem intervalos diferentes que fornecem melhores valores de exatidão. A ABNT (2008) sugere a utilização de dois intervalos, mas foi verificado que dependendo da matéria prima de origem do biodiesel pode ser atribuída diferentes regiões espectrais. Portanto, como relatado por Pinho et al. (2014), devido ao Brasil produzir biodiesel a partir de diversas matérias primas, sugere-se aplicar primeiramente metodologia de classificação do tipo de biodiesel e, posterior aplicação de metodologia de quantificação.

Finalmente, a técnica de Espectrometria no Infravermelho Médio aliada aos métodos de calibração multivariada e seleção de variáveis por intervalos, mostrou-se promissora em relação às análises de determinação do teor de biodiesel em misturas com diesel. Assim, acredita-se que este trabalho indica caminhos importantes aos órgãos brasileiros no que tange ao desenvolvimento de metodologias padrão, uma vez que, no Brasil, não há uma homogeneidade da matéria prima como ocorre na Europa e Estados Unidos.

REFERÊNCIAS BIBLIOGRÁFICAS

- AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP. **Boletim mensal do biodiesel**. 2015a. Disponível em: <www.anp.gov.br/?dw=73583>. Acesso em: 08 ago. 2015.
- AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP. **Resolução ANP no. 45, de 26 de agosto 2014**. 2014. Disponível em: <[http://nxt.anp.gov.br/NXT/gateway.dll/leg/resolucoes_anp/2014/agosto/ranp%2045%20-%202014.xml?fn=document-frameset.htm\\$f=templates\\$3.0](http://nxt.anp.gov.br/NXT/gateway.dll/leg/resolucoes_anp/2014/agosto/ranp%2045%20-%202014.xml?fn=document-frameset.htm$f=templates$3.0)>. Acesso em: 19 ago. 2015.
- AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP. **Monitoramento da qualidade dos combustíveis líquidos automotivos**. 2015b. Disponível em: <www.anp.gov.br/?dw=74509>. Acesso em: 14 abr. 2015.
- ALEME, H. G.; BARBEIRA, P. J. S. Determination of Biodiesel Content in Diesel Using Distillation Curves and Multivariate Calibration, **Energy Fuels**, v. 26, p. 5769–5774, 2012.
- ALISKE, M. A.; ZAGONEL, G. F.; COSTA, B. J.; VEIGA, W.; SAUL, C. K. Measurement of biodiesel concentration in a diesel mixture. **Fuel**, v. 86, p. 1461-1464, 2007.
- ALVES, J. C. L.; POPPI, R. J. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). **Talanta**, v. 104, p. 155-161, 2013.
- AMERICAN SOCIETY FOR TESTING AND MATERIALS – ASTM. **D7371**: Standard test method for the determination of biodiesel (fatty acid methyl esters) content in diesel fuel oil using mid-infrared spectroscopy (FTIR-ATR-PLS method). West Conshohocken, 2007.
- AMERICAN SOCIETY FOR TESTING AND MATERIALS – ASTM. **E1655-05**: Standard practices for infrared multivariate quantitative analysis. West Conshohocken, 2012.
- ANZANELLO, M. J.; FU, K.; FOGLIATTO, F. F.; Ferrão, M. HATR–FTIR wavenumber selection for predicting biodiesel/diesel blends flash point. **Chemometrics and Intelligent Laboratory Systems**, v. 145, p. 1-6, 2015.
- ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS - ABIOVE, Estatística Mensal do Complexo Soja - Agosto de 2015. Disponível em: <<http://www.abiove.org.br/site/index.php?page=estatistica&area=NC0yLTE=>> Acesso em 22 nov. 2015.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS – ABNT. **NBR 15568**: Biodiesel – Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio. Rio de Janeiro, 2008.
- BALABIN, R. M.; SMIRNOV, S. V. Variable selection in near-infrared spectroscopy: Benchmarking of feature selection methods on biodiesel data. **Analytica Chimica Acta**, v. 692, p. 63-72, 2011.

BARBOSA, L. C. A. **Espectroscopia no infravermelho na caracterização de compostos orgânicos**. Viçosa: Ed. UFV, 2007.

BEEBE, K. R.; KOWALSKI, B. R. An Introduction to Multivariate Calibration and Analysis, **Analytical Chemistry**, v. 59, n. 17, p. 1007A-1017A, 1987.

BORGES NETO, W. Parâmetros de qualidade de lubrificantes e óleo de oliva através de espectroscopia vibracional, calibração multivariada e seleção de variáveis. In: Dissertações e teses defendidas na UNICAMP, Campinas – SP, Brasil, 2005.

BRAGA, J. W. B.; SANTOS JUNIOR, A. A.; MARTINS, I. S. Determination of viscosity index in lubricant oils by infrared spectroscopy and PLSR. **Fuel**, v. 120, p. 171-178, 2014.

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

BUIATTE, J. E.; GUIMARÃES, E.; MITSUTAKE, H.; GONTIJO, L. C.; SANTOS, D. Q.; BORGES NETO, W. Qualitative and Quantitative Monitoring of Methyl Cotton Biodiesel Content in Biodiesel/Diesel Blends Using MIR Spectroscopy and Chemometrics Tools, **Journal of Brazilian Chemical Society**, v. 27, n. 1, p. 84-90, 2016.

CENTNER, V.; MASSART, D. Elimination of uninformative variables for multivariate calibration. **Analytical Chemistry**, v. 68, n. 21, p. 3851-3858, 1996.

CHIAVEGATO, E. J. et. al. Guia do Algodão - Tecnologia no campo para uma indústria de qualidade. Agosto 2009. Disponível em: <www.cib.org.br/pdf/guia_algodao_ago09.pdf>. Acesso em 19 ago. 2015.

CHUI, Q. S. H.; ZUCCHINI, R. R.; LICHTIG, J. Qualidade de medições em química analítica. Estudo de caso: determinação de cádmio por espectrofotometria de absorção atômica com chama. **Química Nova**, v. 24, p. 374-380, 2001.

COMPANHIA DE SANEAMENTO BÁSICO DO ESTADO DE SÃO PAULO – SABESP. Programa de Reciclagem de Óleo de Fritura da Sabesp, São Paulo, 2010. Disponível em: <http://site.sabesp.com.br/uploads/file/asabesp_doctos/programa_reciclagem_oleo_completo.pdf>. Acesso em: 19 ago. 2015.

CORREIA, P. R. M.; FERREIRA, M. M. C.; Reconhecimento de padrões por métodos não supervisionados: explorando procedimentos quimiométricos para tratamento de dados analíticos, **Química Nova**, v. 30, p. 481-487, 2007.

COSTA FILHO, P. A.; POPPI, R. J. Algoritmo genético em química. **Química Nova**, v. 22, n. 3, p. 405-411, 1999.

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, 2000.

ENCINAR, J. M.; GONZÁLEZ, J. F.; RODRÍGUEZ, J. J.; TEJEDOR, A. Biodiesel fuels

from vegetable oils: Transesterification of *Cynara cardunculus* L. Oils with ethanol. **Energy & Fuels**, v. 16, p. 443-450, 2002.

EUROPEAN STANDARD – DIN. **EN 14078**: determination of fatty acid methyl esters (FAME) in middle distillates - infrared spectroscopy method. Berlin: Deutsches Institut für Normung e.V.; 2004.

FAROOQ, M.; RAMLI, A.; SUBBARAO, D. Biodiesel production from waste cooking oil using bifunctional heterogeneous solid catalysts. **Journal of Cleaner Production**, v. 59, p. 131-140, 2013.

FAZAL, M. A.; HASEEB, A. S. M. A.; MASJUKI, H. H. Biodiesel feasibility study: an evaluation of material compatibility; performance; emission and engine durability. **Renewable Sustainable Energy Review**, v. 15, p. 1314-1324, 2011.

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

FERRAND, M.; HUQUET, B.; BARBEY, S.; BARILLET, F.; FAUCON, F.; LARROQUE, H.; LERAY, O.; TROMMENSCHLAGER, J.M.; BROCHARD, M. Determination of fatty acid profile in cow's milk using mid-infrared spectrometry: Interest of applying a variable selection by genetic algorithms before a PLS regression. **Chemometrics and Intelligent Laboratory Systems**, v. 106, p. 183–189, 2011.

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

FERRARI, R. A.; OLIVEIRA, V. S.; SCABIO, A. Biodiesel de soja – taxa de conversão em ésteres etílicos, caracterização físico-química e consumo em gerador de energia. **Química Nova**, v. 28, n. 1, p. 19-23, 2005.

FERREIRA, M. H.; BRAGA, J. W. B.; SENA, M. M. Development and validation of a chemometric method for direct determination of hydrochlorothiazide in pharmaceutical samples by diffuse reflectance near infrared spectroscopy. **Microchemical Journal**, v. 89, p. 342-351, 2012.

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

FILGUEIRAS, P. R.; ALVES, J. C. L.; POPPI, R. J. Quantification of animal fat biodiesel in soybean biodiesel and B20diesel blends using near infrared spectroscopy and synergy interval support vector regression. **Talanta**, v. 119, p. 582–589, 2014.

FRIEDEL, M.; PATZ, Claus-Dieter.; DIETRICH, H. Comparison of different measurement techniques and variable selection methods for FT-MIR in wine analysis. **Food Chemistry**, v. 141, p. 4200-4207, 2013.

GELADI, P.; KOWALSKI, B. R. Partial Least Squares Regression: A Tutorial, **Analytica**

Chimica Acta, v. 185, p.1-17, 1986.

GEMPERLINE, P. **Practical guide to chemometrics**, 2nd ed.; CRC Press Taylor & Francis, New York, 2006.

GOMEZ-CARRACEDO, M. P.; ANDRADE, J. M.; RUTLEDGE, D. N.; FABER, N. M.; Selecting the optimum number of partial least squares components for the calibration of attenuated total reflectance-mid-infrared spectra of undersigned kerosene samples, **Analytica Chimica Acta**, v. 585, p. 253-265, 2007.

GONTIJO, L. C.; GUIMARÃES, E.; MITSUTAKE, H.; SANTANA, F. B.; SANTOS, D. Q.; BORGES NETO, W. Quantification of soybean biodiesels in diesel blends according to ASTM E1655 using mid-infrared spectroscopy and multivariate calibration, **Fuel**, v. 117, p. 1111-1114, 2014a.

GONTIJO, L. C.; GUIMARÃES, E.; MITSUTAKE, H.; SANTANA, F. B.; SANTOS, D. Q.; BORGES NETO, W. Development and Validation of PLS Models for Quantification of Biodiesels Content from Waste Frying Oil in Diesel by HATR-MIR, **Revista Virtual de Química**, v. 6, n. 5, p. 1517-1528, 2014b.

GUERRERO, A.; ANGUEBES, F.; CASTELÁN, M.; MORALES, V.; GARCÍA, R.; CÓRDOVA, A. V.; ZAVALA, J. C. FTIR-ATR and Multivariate Calibration for the Prediction of Biodiesel Concentration in Petrodiesel Blends, **American Journal of Analytical Chemistry**, v. 4, p. 343-347, 2013.

GUIMARÃES, E. Aplicação de Espectroscopia no Infravermelho Médio e Calibração Multivariada para quantificação de adulterantes em biodieséis de óleo de soja e residual e suas misturas com diesel. In: Dissertações e teses defendidas na UFU, Uberlândia – MG, Brasil, 2014.

GUIMARÃES, E.; GONTIJO, L. C.; MITSUTAKE, H.; SANTANA, F. B.; SANTOS, D. Q.; BORGES NETO, W. Quantification of Ethanol in Biodiesels Using Mid-Infrared Spectroscopy and Multivariate Calibration, **Industrial & Engineering Chemistry Research**, v. 53, p. 13575-13580, 2014.

GUIMARÃES, E.; MITSUTAKE, H.; GONTIJO, L. C.; SANTANA, F. B.; SANTOS, D. Q.; BORGES NETO, W. Infrared Spectroscopy and Multivariate Calibration for Quantification of Soybean Oil as Adulterant in Biodiesel Fuels, **Journal American Oil Chemical Society**, v. 92, p. 777-782, 2015a.

GUIMARÃES, E.; SANTANA, F. B.; GONTIJO, L. C.; MITSUTAKE, H.; SANTOS, D. Q.; BORGES NETO, W. Aplicação das Figuras de Mérito Multivariadas na Validação de Metodologias em Análises de Biocombustíveis empregando Espectroscopia no Infravermelho Médio e PLS, **Revista Virtual de Química**, v. 7, n. 6, p. 2242-2254, 2015b.

HOLLER, F. J.; SKOOG, D. A.; CROUCH, S. R. **Princípios de Análise Instrumental**. Tradução da 6 ed. Porto Alegre: Ed. Bookman, 2009.

HUANG, Y. P.; CHANG, J. I. Biodiesel production from residual oils recovered from spent bleaching earth. **Renewable Energy**, Oxford, v. 35, n. 1, p. 269-274, 2010.

HUANGA, Z.; SHAB, S.; RONGA, Z.; CHENA, J.; HEA, Q.; KHANA, D. M.; ZHUA, S. Feasibility study of near infrared spectroscopy with variable selection for non-destructive determination of quality parameters in shell-intact cottonseed. **Industrial Crops and Products**, v. 43, p. 654-660, 2013.

INSAUSTI, M.; ROMANO, C.; PISTONESI, M. F.; FERNANDEZ BAND, B. S. Simultaneous determination of quality parameters in biodiesel/diesel blends using synchronous fluorescence and multivariate analysis, **Microchemical Journal**, v. 108, p. 32-37, 2013.

JIANG, H.; LIU, G.; MEI, C.; YU, S.; XIAO, X.; DING, Y. Measurement of process variables in solid-state fermentation of wheat straw using FT-NIR spectroscopy and synergy interval PLS algorithm. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 97, p. 277-283, 2012.

KNOTHE, G. Determining the blend level of mixtures of biodiesel with conventional diesel fuel by fiber-optic near-infrared spectroscopy and ^1H nuclear magnetic resonance spectroscopy. **Journal American Oil Chemical Society**, v. 78, n. 10, p. 1025-1028, 2001.

KNOTHE, G.; KRAHL, J.; GERPEN, J. V.; RAMOS, L. P. **Manual de Biodiesel**. São Paulo: Edgar Blucher, 2006.

KOH, M. Y.; GHAZI, T. I. M. A review of biodiesel production from *Jatropha curcas* L. Oil. **Renewable and Sustainable Energy Reviews**, v. 15, p. 2240-2251, 2011.

LEARDI, R.; NØRGAARD, L. Sequential application of backward interval partial least squares and genetic algorithms for the selection of relevant spectral regions. **Journal Chemometrics**, v. 18, p. 486-497, 2004.

LIU, F.; HE, Y.; WANG, L. Comparison of calibrations for the determination of soluble solids content and pH of rice vinegars using visible and short-wave near infrared spectroscopy. **Analytica Chimica Acta**, Amsterdã, v. 610, n. 2, p. 196-204, 2008.

MACIEL, C. D. G.; POLETINE, J. P.; PEREIRA, J. C.; MONDINI, M. L. Avaliação da qualidade de semente de soja (*Glycine max* L. Merrill) cultivar IAC - 18. **Revista Científica Eletrônica de Agronomia**, v. 7, p. 1-14, 2005.

MAHAMUNI, N. N.; ADEWUYI, Y. G. Fourier transform infrared spectroscopy (FTIR) method to monitor soy biodiesel and soybean oil in transesterification reactions, petrodiesel-biodiesel blends, and blend adulteration with soy oil. **Energy & Fuels**, v. 23, n. 7, p. 3773-3782, 2009.

MARTENS, H.; NAES, T. **Multivariate calibration**. New York: Wiley, 1996.

MASSART, D. L.; VANDEGINSTE, B. G. M.; BUYDENS, L. M. C.; JONG, S. D.; LEWI, P. J.; SMEYERS-VERBEKE, J. **Data Handling in Science and Technology: Part A**, Amsterdã: Ed. Elsevier, 1998.

MAZIVILA, S. J.; MITSUTAKE, H.; SANTANA, F. B.; GONTIJO, L. C.; SANTOS, D. Q.;

- BORGES NETO, W. Fast Classification of Different Oils and Routes Used in Biodiesel Production Using Mid Infrared Spectroscopy and PLS2-DA, **Journal of Brazilian Chemical Society**, v. 26, n. 4, p. 642-648, 2015a.
- MAZIVILA, S. J.; SANTANA, F. B.; MITSUTAKE, H.; GONTIJO, L. C.; SANTOS, D. Q.; BORGES NETO, W. Discrimination of the type of biodiesel/diesel blend (B5) using mid-infrared spectroscopy and PLS-DA, **Fuel**, v. 142, p. 222–226, 2015b.
- MAZIVILA, S. J.; GONTIJO, L. C.; SANTANA, F. B.; MITSUTAKE, H.; SANTOS, D. Q.; BORGES NETO, W. Fast detection of adulterants/contaminants in biodiesel/diesel blend (B5) employing mid-infrared spectroscopy and PLS-DA. **Energy & Fuels**, v. 29, p. 227–232, 2015c.
- MILLER, J. N.; MILLER, J. C. **Statistics and chemometrics for analytical chemistry**. 6 ed. Essex, Ed. Pearson, 2010.
- MITSUTAKE, H.; GUIMARÃES, E.; FREITAS, H. C.; GONTIJO, L. C.; SANTOS, D. Q.; BORGES NETO, W. Multivariate control charts based on NAS and mid-infrared spectroscopy for quality control of B5 blends of methyl soybean biodiesel in diesel, **Journal of Chemometrics**, v. 29, p. 411–419, 2015.
- MOFIJUR, M.; MASJUKI, H. H.; KALAM, M. A.; ATABANI, A. E.; SHAHABUDDIN, M.; PALASH, S. M.; HAZRAT, M. A. Effect of biodiesel from various feedstocks on combustion characteristics, engine durability and materials compatibility: A review. **Renewable & Sustainable Energy Reviews**, v. 28, p. 441-455, 2013.
- MONTEIRO, M. R.; AMBROZINA, 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, p. 660–664, 2009.
- MUNCK, L.; PRAM NIELSEN, J.; MØLLER, B.; JACOBSEN, S.; SØNDERGAARD, I.; ENGELSEN, S.B.; NØRGAARD, L.; BRO, R. Exploring the phenotypic expression of a regulatory proteome-altering gene by spectroscopy and chemometrics. **Analytica Chimica Acta**, v. 446, p. 171-186, 2001.
- NØRGAARD, L. The iToolbox for MATLAB, KVL, Denmark, July 2004. Disponível em: <<http://www.models.life.ku.dk>>. Acesso em 23 nov. 2015.
- NØRGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. 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, 2000.
- NORTON, G. A.; CLINE, A. M.; THOMPSON, G. C. Use of radiocarbon analyses for determining levels of biodiesel in fuel blends – Comparison with ASTM D7371 for FAME, **Fuel**, v. 96, p. 284-290, 2012.
- OLIVEIRA, A. D. et. al Biodiesel do óleo de pinhão manso degomado por esterificação. In: IV Congresso Brasileiro de Mamona e I Simpósio Internacional de Oleaginosas Energéticas, João Pessoa – PB, 2010. Disponível em: <www.cbmamona.com.br/pdfs/BID-12.pdf>. Acesso

em: 19 ago. 2015.

OLIVEIRA, J. S.; MONTALVÃO, R.; DAHER, L.; SUAREZ, P. A. Z.; Rubim, J. C. Determination of methyl ester contents in biodiesel blends by FTIR-ATR and FTNIR spectroscopies. **Talanta**, v. 69, p. 1278–1284, 2006.

OLIVEIRA, R. R.; LIMA, K. M. G.; TAULER, R.; JUAN, A. Application of correlation constrained multivariate curve resolution alternating least-squares methods for determination of compounds of interest in biodiesel blends using NIR and UV-visible spectroscopic data. **Talanta**, v. 125, p. 233-241, 2014.

OTTO, M. **Chemometrics – Statistics and Computer application in Analytical Chemistry**, Weinheim: Wiley, 1999.

PADULA, A. D.; SANTOS, M. S.; FERREIRA, L.; BORENSTEIN, D. The emergence of the biodiesel industry in Brazil: Current figures and future prospects. **Energy Policy**, v. 35, p. 395-405, 2012.

PASSARI, L. M. Z. G.; SOARES, P. K.; BRUNS, R. E.; SCARMINIO, I. S. Estatística aplicada à química: dez dúvidas comuns. **Química Nova**, v. 34, n.5, p. 888-892, 2011.

PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S.; VYVYAN, J. R. **Introdução a espectroscopia**. Tradução da 4 ed. São Paulo: Ed. Cengage Learning, 2010.

PIANTAVINI, M. S.; PONTES, F. L. D.; UBER, C. P.; STREMEL, D. P.; SENA, M. M.; PONTAROLO, R. Chemometric quality inspection control of pyrantel pamoate, febantel and praziquantel in veterinary tablets by mid infrared spectroscopy. **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**, v. 125, p. 396–403, 2014.

PIMENTEL, M. F.; RIBEIRO, G. M. G. S.; CRUZ, R. S.; STRAGEVITCH, L.; PACHECO FILHO, J. G. A.; TEIXEIRA, L. S. G. Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. **Microchemical Journal**, v. 82, p. 201-206, 2006.

PINHO, D. M. M.; SANTOS, V. O.; SANTOS, V. M. L.; OLIVEIRA, M. C. S.; SILVA, M. T.; PIZA, P. G. T.; PINTO, A. C.; REZENDE, M. J. C.; SUAREZ, P. A. Z. Evaluating the use of EN 14078 for determination of biodiesel in diesel blends sold in the Brazilian market. **Fuel**, v. 136, p. 136-142, 2014.

PORTE, A. F.; SCHNEIDER, R. C. S.; KAERCHER, J. A.; KLAMT, R. A.; SCHMATZ, W. L.; SILVA, W. L. T.; SeveroFilho, W. A. Sunflower biodiesel production and application in family farms in Brazil. **Fuel**, v. 89, p. 3718-3724, 2010.

SARIN, A.; ARORA, R.; SINGH, N. P.; SARIN, R.; MALHOTRA, R. K. Blends of biodiesels synthesized from non-edible and edible oils: Influence on the OS (oxidation stability). **Energy**, v. 35, n. 8, p. 3449-3453, 2010.

SCHALE, S. P.; LE, T. M.; PIERCE, K. M. Predicting feedstock and percent composition for blends of biodiesel with conventional diesel using chemometrics and gas chromatography-mass spectrometry, **Talanta**, v. 94, p. 320-327, 2012.

SITOE, B. V.; MITSUTAKE, H.; GUIMARÃES, E.; GONTIJO, L. C.; SANTOS, D. Q.; BORGES NETO, W. Quality Control of Biodiesel Content of B7 Blends of Methyl Jatropha and Methyl Crambe Biodiesels Using Mid-Infrared Spectroscopy and Multivariate Control Charts Based on Net Analyte Signal, **Energy Fuels**, 2016. No prelo. doi 10.1021/acs.energyfuels.5b02489

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

SILVA, M. A. A.; CORREA, R. A.; TAVARES, M. G. O.; ANTONIOSI FILHO, N. R. A new spectrophotometric method for determination of biodiesel content in biodiesel/diesel blends, **Fuel**, v. 143, p. 16–20, 2015.

SMITH, B.C. **Fundamentals of Fourier transform infrared spectroscopy**. 2 ed. Boca Raton: CRC Press, 2011.

SOUZA, L. M.; SANTANA, F. B.; GONTIJO, L. C.; MAZIVILA, S. J.; BORGES NETO, W. Quantification of adulterations in extra virgin flaxseed oil using MIR and PLS. **Food Chemistry**, v. 182, p. 35–40, 2015.

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Variable selection, outlier detection, and figures of merit estimation in a partial least-squares regression multivariate calibration model. A case study for the determination of quality parameters in the alcohol industry by near-infrared spectroscopy. **Journal of Agricultural and Food Chemistry**, v. 55, n. 21, p. 8331-8338, 2007.

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

VANDEGINSTE, B. G. M.; MASSART, D. L.; BUYDENS, L. M. C.; JING, S.; LEWI, P. J.; SMEYERS-VERBEKE, J. **Handbook of Chemometrics and Qualimetrics: Part B**, Amsterdã: Ed. Elsevier, 1998.

VASCONCELOS, F. V. C.; SOUZA JR, P. F. B.; PIMENTEL, M. F.; PONTES, M. J. C.; PEREIRA, C. F. Using near-infrared overtone regions to determine biodiesel content and adulteration of diesel/biodiesel blends with vegetable oils. **Analytica Chimica Acta**, v. 716, p. 101-107, 2012.

WOLD, S.; ESBENSEN, K.; GELADI, P. **Chemometrics and Intelligent Laboratory Systems**, Amsterdã, v. 2, p. 37-52, 1987.

XU, L.; CAI, C. B.; CUI, H. F.; YE, Z. H.; YU, X. P. Rapid discrimination of pork in Halal and non-Halal Chinese ham sausages by Fourier transform infrared (FTIR) spectroscopy and chemometrics. **Meat Science**, v. 92, p. 506–510, 2012.