



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

ADEMAR DOMINGOS VIAGEM MÁQUINA

**DETERMINAÇÃO DO TEOR DE BIODIESEL DE MAFURRA E CRAMBE EM
MISTURAS COM DIESEL POR RMN DE ^1H E REGRESSÃO MULTIVARIADA POR
OPLS E OPLS-DA**

**UBERLÂNDIA
Maio de 2020**

ADEMAR DOMINGOS VIAGEM MÁQUINA

**DETERMINAÇÃO DO TEOR DE BIODIESEL DE MAFURRA E CRAMBE EM
MISTURAS COM DIESEL POR RMN DE ^1H E REGRESSÃO MULTIVARIADA POR
OPLS E OPLS-DA**

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

Área de Concentração: Química Analítica

Orientador: Prof. Dr. Waldomiro Borges Neto

UBERLÂNDIA

Maio de 2020

Ficha Catalográfica Online do Sistema de Bibliotecas da UFU
com dados informados pelo(a) próprio(a) autor(a).

M478 Máquina, Ademar Domingos Viagem, 1987-
2020 Determinação do teor de biodiesel de Mafurra e Crambe em misturas com diesel por RMN de 1H e regressão multivariada por OPLS e OPLS-DA [recurso eletrônico] / Ademar Domingos Viagem Máquina. - 2020.

Orientador: Waldomiro Borges Neto.
Tese (Doutorado) - Universidade Federal de Uberlândia, Pós-graduação em Química.
Modo de acesso: Internet.
Disponível em: <http://doi.org/10.14393/ufu.te.2020.453>
Inclui bibliografia.
Inclui ilustrações.

1. Química. I. Borges Neto, Waldomiro ,1970-, (Orient.). II. Universidade Federal de Uberlândia. Pós-graduação em Química. III. Título.

CDU: 54



UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Química
Av. João Naves de Ávila, 2121, Bloco 51 - Bairro Santa Mônica, Uberlândia-MG, CEP
38400-902
Telefone: (34) 3239-4385 - www.cpgquimica.iq.ufu.br - cpgquimica@ufu.br



ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Química			
Defesa de:	Tese de Doutorado, 105, PPGQUI			
Data:	vinte e nove de maio de dois mil e vinte	Hora de início: [14:00mm]	Hora de encerramento: [18:00]	
Matrícula do Discente:	11813QMI001			
Nome do Discente:	Ademar Domingos Viagem Máquina			
Título do Trabalho:	Determinação do teor de biodiesel de mafurra e crambe em misturas com diesel por RMN de 1H e Regressão Multivariada por OPLS e OPLS-DA			
Área de concentração:	Química			
Linha de pesquisa:	Instrumentação Analítica e Preparo de Amostras			
Projeto de Pesquisa de vinculação:	Controle de qualidade de biocombustíveis.			

Reuniu-se por meio de webconferência, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Química, assim composta: Professores Doutores: Jefferson Luis Ferrari e Sidnei Gonçalves da Silva, da Universidade Federal de Uberlândia; Roberta Eliane Santos Froes, da Universidade Federal de Ouro Preto; Frederico Garcia Pinto, da Universidade Federal de Viçosa e Waldomiro Borges Neto, orientador do candidato.

Iniciando os trabalhos o presidente da mesa, Dr. Waldomiro Borges Neto, apresentou a Comissão Examinadora e o candidato, agradeceu a presença do público, e concedeu ao Discente a palavra para a exposição do seu trabalho. A duração da apresentação do Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(as) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

[Aprovado].

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação

interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.



Documento assinado eletronicamente por **Waldomiro Borges Neto**, **Professor(a) do Magistério Superior**, em 29/05/2020, às 17:44, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Sidnei Gonçalves da Silva**, **Professor(a) do Magistério Superior**, em 29/05/2020, às 17:45, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Jefferson Luis Ferrari**, **Professor(a) do Magistério Superior**, em 29/05/2020, às 17:46, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Frederico Garcia Pinto**, **Usuário Externo**, em 29/05/2020, às 17:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



Documento assinado eletronicamente por **Roberta Eliane Santos Froes**, **Usuário Externo**, em 29/05/2020, às 17:47, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do [Decreto nº 8.539, de 8 de outubro de 2015](#).



A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **2062235** e o código CRC **3CBEFA3C**.

Dedico este trabalho aos meus pais Domingos
Viagem Máquina (in memorian) e Belia
Domingos Praújo Máquina.

À minha esposa Agnes Mariza Paula
Cordeiro Máquina.

Às minhas filhas Cynara Domingas Ademar
Máquina e Yunaika Paula Ademar
Máquina.

AGRADECIMENTOS

Agradeço:

À Deus Misericordioso, que em todos os dias da minha vida me deu força interior para superar as dificuldades e nunca desistir daquilo que acredito.

À minha família, que soube viver sem a minha presença física durante o período de formação.

Ao meu orientador, Prof. Dr. Waldomiro Borges Neto, pelo devido acompanhamento acadêmico e não só, mas como também pelo convívio, apoio, compreensão e, principalmente, pela confiança depositada em meu trabalho.

Aos colegas do Laboratório de Quimiometria do Triângulo: Baltazar Vasco Sítioe, Edvando Souza Teles, Eloíza Guimarães, Felipe Bachion de Santana, Hery Mitsutake, José Eduardo Buiatte, Letícia Maria de Souza, Lucas Caixeta Gontijo, Lucas Gustavo da Costa, Maria Teresa Carvalho Ferreira, Raquel Boaventura de Moraes e Tathiana Dienifer Godinho pela ajuda prestada durante os trabalhos acadêmicos.

Ao Programa do Convênio do Conselho Nacional de Desenvolvimento Científico e Tecnológico do Brasil, processo número 190025/2017-4, pelo suporte financeiro.

A Financiadora de Estudos e Projetos (Finep), pelo financiamento do equipamento de RMN de ^1H .

À Universidade Federal de Uberlândia e aos seus funcionários que forneceram toda a estrutura física e tecnológica para a realização deste trabalho, principalmente à secretaria do programa de pós-graduação em química, Mayta Mamede Negreto Peixoto, pelo acompanhamento em todos os processos técnico-administrativos.

À Universidade Púnguè pelo afastamento para a realização do meu doutoramento.

À TRANSPETRO S.A. (Uberlândia, Estado de Minas Gerais, Brasil) pelo fornecimento do diesel utilizado para a preparação das amostras.

A todos os meus colegas de serviço e alunos pela convivência profissional e apoio moral durante os meus estudos.

Finalmente, ressalto que, mencionar nomes de todos os que contribuíram para fazer deste trabalho uma realidade é uma tarefa difícil. Portanto, agradeço aos demais que não os mencionei aqui, mas que direta ou indiretamente contribuíram para o sucesso deste trabalho.

RESUMO

A produção e o uso do biodiesel estão sendo cada vez mais potencializados em vários países devido à possibilidade de escassez dos combustíveis fósseis a médio ou longo prazo e a necessidade de reduzir os impactos ambientais causado pelo seu processo de produção. O Brasil é um dos exemplos dos países que mais produz e usa biodiesel, tendo atingido 5.350.036 m³ de biodiesel em 2018 contra 4.291.294 m³ do ano anterior, favorecendo desta forma para evolução da sua percentagem adicionada ao diesel. Esta evolução representa preocupações em relação a sua comercialização com teor que não é o exigido pelas normas. Por esta razão, neste trabalho foram desenvolvidos quatro métodos para monitorar o teor B10 de biodiesel de mafurra e crambe em misturas com diesel usando espectroscopia de ressonância magnética nuclear de hidrogênio (RMN de ¹H) combinada com a regressão multivariada por projeções ortogonais à estrutura latente (OPLS) e análise discriminante por projeções ortogonais à estrutura latente (OPLS-DA). A eficiência dos métodos desenvolvidos com base na regressão multivariada por OPLS foi analisada com base nas figuras de mérito e no ajuste dos modelos por meio da correlação dos valores medidos e previstos dos conjuntos de calibração e previsão. Os resultados das figuras de mérito foram concordantes com os requisitos estabelecidos na norma ASTM E1655-05. Uma alta correlação entre os valores medidos e previstos foi evidente em todos os modelos OPLS, com um coeficiente de correlação (R^2) superior a 0,99. A eficiência dos métodos desenvolvidos com base na regressão multivariada por OPLS-DA foi analisada com base nos parâmetros de taxa de falso positivos e falso negativos, sensibilidade, especificidade e coeficiente de correlação de Matthew, onde não foi notada a presença de amostras falsos positivos e nem falsos negativos, consequentemente, os parâmetros de sensibilidade, especificidade e coeficiente de correlação de Matthew foram iguais a 1, o que significa que os modelos apresentaram uma classificação 100% correta das amostras de B10 (10% de biodiesel e 90% de diesel puro) e BX (teor de biodiesel menor e maior que B10) utilizadas para treinamento e teste. A alta eficiência demonstrada pelos modelos OPLS e OPLS-DA no monitoramento do teor B10 do biodiesel de mafurra e crambe em mistura com diesel, sugerem que os métodos analíticos desenvolvidos são ideais, eficientes e adequados para serem usados pelas agências de inspeção no controle da qualidade desse combustível.

Palavras-chave: Biocombustível. RMN de ¹H. Calibração Multivariada. Métodos de Classificação. Quimiometria.

ABSTRACT

The production and the use of biodiesel is being increasingly strengthened in several countries due to the possibility of a shortage of fossil fuels in the medium or long term and the need to reduce the environmental impacts caused by its production process. Brazil is one of the examples of the countries that most produce and use biodiesel, having reached 5,350,036 m³ of biodiesel in 2018 against 4,291,294 m³ of the previous year, thus favoring the evolution of its percentage added to diesel. This evolution represents concerns in relation to its commercialization with content that is not what is required by the standards. For this reason, in this work, four methods were developed to monitor the B10 content of mafurra and crambe biodiesel in mixtures with diesel using hydrogen nuclear magnetic resonance spectroscopy (¹H NMR) combined with the multivariate regression by orthogonal projections to the latent structure (OPLS) and orthogonal projections on the latent structure-discrimination analysis (OPLS-DA). The efficiency of the methods developed based on the multivariate regression by OPLS was analyzed based on the figures of merit and the fit of the models through the correlation of the measured and predicted values of the calibration and forecast sets. The results of the figures of merit were in agreement with the requirements established in the standard ASTM E1655-05. A high correlation between measured and predicted values was evident in all OPLS models, with a correlation coefficient (R^2) greater than 0.99. The efficiency of the methods developed based on multivariate regression by OPLS-DA was analyzed based on the parameters of false positive and false negative rates, sensitivity, specificity and Matthew's correlation coefficient, where the presence of false positive and false negative samples was not noticed, consequently, the parameters of sensitivity, specificity and Matthew's correlation coefficient were equal to 1, which means that the models presented a 100% correct classification of the B10 samples (10% biodiesel and 90% pure diesel) and BX (biodiesel content less and greater than B10) used in the training and test sets. The high efficiency demonstrated by the OPLS and OPLS-DA models in monitoring the B10 content of biodiesel from mafurra and crambe mixed with diesel, suggests that the analytical methods developed are ideal, efficient and suitable for use by quality control inspection agencies of that fuel.

Keywords: Biofuel. ¹H NMR. Multivariate calibration. Classification methods. Chemometrics.

LISTA DE FIGURAS

Figura 1 –	Evolução da percentagem do teor de biodiesel no diesel comercializado no Brasil.....	22
Figura 2 –	Evolução da percentagem do teor de biodiesel no diesel comercializado em Moçambique.....	24
Figura 3 –	Árvore, frutos e sementes de mafurra.....	26
Figura 4 –	Plantas, frutos e sementes de crambe.....	27
Figura 5 –	Mecanismo da reação de hidrólise de triacilgliceróis na presença de umidade.....	28
Figura 6 –	Mecanismos de saponificação de triglycerídeos.....	29
Figura 7 –	Mecanismo de reação de esterificação via catálise ácida.....	29
Figura 8 –	Reação da formação de alcóxido.....	30
Figura 9 –	Mecanismo de reação de transesterificação via catálise alcalina.....	31
Figura 10 –	Índice de não conformidade do diesel comercializado no Brasil.....	32
Figura 11 –	Relação de Boltzmann com a) ausência de um campo magnético externo e b) presença de um campo magnético externo.....	35
Figura 12 –	Processo de Ressonância Magnética Nuclear.....	37
Figura 13 –	Construção da matriz X a partir dos dados espectrais.....	39
Figura 14 –	Construção do modelo de regressão por OPLS.....	41
Figura 15 –	Gráfico de <i>leverage</i> versus Q Residuals.....	43
Figura 16 –	Processo de previsão de amostras de validação pelo modelo OPLS.....	44
Figura 17 –	Representação da matriz de dados X e o vetor y usados no OPLS-DA.....	49
Figura 18 –	Representação gráfica de estimativa de classes para o modelo OPLS-DA.....	51
Figura 19 –	Representação gráfica de dispersão dos escores para o modelo OPLS-DA.....	52
Figura 20 –	Etapas de produção de biodiesel metílico de mafurra e crambe.....	55
Figura 21 –	Especetros de RMN de ^1H não corrigidos de 70 amostras de misturas diesel/biodiesel de (a) mafurra e (b) crambe.....	60
Figura 22 –	Especetros de RMN de ^1H corrigidos de 70 amostras de misturas diesel/biodiesel de (a) mafurra e (b) crambe.....	61
Figura 23 –	Especetros de RMN de ^1H de B10 e BX.....	62
Figura 24 –	Variância acumulada capturada nas matrizes X e Y versus variável latente do modelo OPLS de mafurra.....	63

Figura 25 –	Variância acumulada capturada nas matrizes X e Y <i>versus</i> variável latente do modelo OPLS de crambe.....	63
Figura 26 –	Correlação dos valores medidos e previstos dos conjuntos de calibração e previsão do modelo OPLS de mafurra.....	66
Figura 27 –	Correlação dos valores medidos e previstos dos conjuntos de calibração e previsão do modelo OPLS de crambe.....	67
Figura 28 –	Valores de <i>leverage</i> versus valores <i>Q residuals</i> com 95% de confiança para o modelo OPLS de mafurra.....	68
Figura 29 –	Valores de <i>leverage</i> versus valores <i>Q residuals</i> com 95% de confiança para o modelo OPLS de crambe.....	68
Figura 30 –	Valores de <i>leverage</i> versus valores <i>Q residuals</i> com 95% de confiança para o modelo OPLS-DA de mafurra.....	70
Figura 31 –	Valores de <i>leverage</i> versus valores <i>Q residuals</i> com 95% de confiança para o modelo OPLS-DA de crambe.....	70
Figura 32 –	Parâmetros R^2 e Q^2 do modelo OPLS-DA de mafurra.....	72
Figura 33 –	Parâmetros R^2 e Q^2 do modelo OPLS-DA de crambe.....	72
Figura 34 –	Curva ROC do modelo OPLS-DA de mafurra para (a) sensibilidade <i>versus</i> especificidade do BX; (b) sensibilidade <i>versus</i> especificidade do B10.....	75
Figura 35 –	Curva ROC do modelo OPLS-DA de crambe para (a) sensibilidade <i>versus</i> especificidade do BX; (b) sensibilidade <i>versus</i> especificidade do B10.....	76
Figura 36 –	Valores estimados de classe para conjunto de treinamento e de teste das amostras B10 e BX do modelo OPLS-DA de mafurra.....	77
Figura 37 –	Valores estimados de classe para conjunto de treinamento e de teste das amostras B10 e BX do modelo OPLS-DA de crambe.....	77
Figura 38 –	Seleção do <i>threshold</i> pelo teorema da Bayer para amostras B10 e BX do modelo OPLS-DA de mafurra.....	78
Figura 39 –	Seleção do <i>threshold</i> pelo teorema da Bayer para amostras B10 e BX do modelo OPLS-DA de crambe.....	78
Figura 40 –	Escores para o modelo OPLS-DA de mafurra em (a) calibração e (b) previsão.....	79
Figura 41 –	Escores para o modelo OPLS-DA de crambe em (a) calibração e (b) previsão.....	80

Figura 42 –	Gráfico dos pesos de VL-1, VL-2, VL-3 <i>versus</i> variáveis do modelo OPLS-DA de mafurra.....	81
Figura 43 –	Gráfico dos pesos de VL-1, VL-2, VL-3 <i>versus</i> variáveis do modelo OPLS-DA de crambe.....	82
Figura 44 –	Gráfico da Soma do Quadrado Y (SSQ_Y) <i>versus</i> correlação da matriz Y para o modelo OPLS-DA de mafurra.....	83
Figura 45 –	Gráfico da Soma do Quadrado Y (SSQ_Y) versus correlação da matriz Y para o modelo OPLS-DA de crambe.....	84

LISTA DE TABELAS

Tabela 1 –	Quantidade de matéria-prima usada para produção do biodiesel nos últimos 5 anos.....	25
Tabela 2 –	Propriedades atômicas e magnéticas de alguns núcleos atômicos.....	36
Tabela 3 –	Equações usadas para calcular as figuras de mérito dos modelos OPLS....	45
Tabela 4 –	Condições experimentais.....	57
Tabela 5 –	Parâmetros e valores de exatidão dos modelos OPLS.....	64
Tabela 6 –	Resultados das figuras de mérito dos modelos OPLS de mafurra e de crambe.....	65
Tabela 7 –	Probabilidade de insignificância dos modelos OPLS de mafurra e de crambe com 2 e 3 VL, respectivamente.....	69
Tabela 8 –	Parâmetros de classificação obtidos pelos modelos OPLS-DA de mafurra e de crambe.....	71
Tabela 9 –	Matriz de confusão dos modelos OPLS-DA de mafurra e de crambe.....	73
Tabela 10 –	Tabela de confusão dos modelos OPLS-DA de mafurra e de crambe.....	74
Tabela 11 –	Regiões espectrais e suas atribuições.....	81
Tabela 12 –	Probabilidade de insignificância dos modelos OPLS-DA de mafurra e de crambe com 3 VLs.....	83

LISTA DE ABREVIATURAS E SIGLAS

ABNT NBR	Norma Brasileira da Associação Brasileira de Normas Técnicas
ANP	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
ASTM	Sociedade Americana de Testes e Materiais (do inglês, <i>American Society of Testing and Materials</i>)
ATR	Refletância Total Atenuada (do inglês, <i>Attenuated Total Reflectance</i>)
B100	Biodiesel Puro
B10	Mistura Biodiesel/Diesel com 10% de Biodiesel
BX	Mistura Biodiesel/Diesel com X% de Biodiesel
BMMA	Biodiesel metílico de mafurra
BMC	Biodiesel metílico de crambe
EN	Norma Europeia (do inglês, <i>European Norm</i>)
FT	Transformada de Fourier (do inglês, <i>Fourier Transform</i>)
FT-MIR	Infravermelho Médio com Transformada de Fourier (do inglês, <i>Fourier Transform Mid Infrared</i>)
HCA	Análise de Agrupamentos Hierárquicos (do inglês, <i>hierarchical cluster analysis</i>)
KNN	K-Vizinho mais Próximo (do inglês, <i>K-nearest neighbors</i>)
MIR	Infravermelho Médio (do inglês, <i>Mid Infrared</i>)
NAS	Sinal Analítico Líquido (do inglês, <i>Net Analyte Signal</i>)
OPLS	Projeções ortogonais à estrutura latente (do inglês, <i>orthogonal projections to latent structure</i>)
OPLS-DA	Análise discriminante por projeções ortogonais à estrutura latente (do inglês, <i>orthogonal projections on the latent structure-discrimination analysis</i>)
OSC	Correção Ortogonal do Sinal (do inglês, <i>Orthogonal Signal Correction</i>)
PCA	Análise de Componentes Principais (do inglês, <i>Principal Components Analysis</i>)
PLS	Quadrados Mínimos Parciais (do inglês, <i>Partial Least Squares</i>)
PLS-DA	Análise Discriminante por Quadrados Mínimos Parciais (do inglês, <i>Partial Least Squares – Discriminant Analysis</i>)
PNPB	Programa Nacional de Produção e Uso do Biodiesel
RMN	Ressonância Magnética Nuclear (do inglês, <i>Nuclear Magnetic Resonance</i>)

RMN de ^1H	Ressonância Magnética Nuclear de Hidrogênio (do inglês, <i>Hydrogen Nuclear Magnetic Resonance</i>)
RMSEC	Erro Quadrático Médio de Calibração (do inglês, <i>Root Mean Square Error of Calibration</i>)
RMSECV	Erro Quadrático Médio de Validação Cruzada (do inglês, <i>Root Mean Square Error of Cross-Validation</i>)
RMSEP	Erro Quadrático Médio de Previsão (do inglês, <i>Root Mean Square Error of Prediction</i>)
ROC	Característica de Operação do Receptor (ROC – do inglês, <i>receiver operating characteristics</i>)
SADC	Comunidade para o Desenvolvimento da África Austral (do inglês, Southern Africa Development Community)
SDV	Desvio Padrão da Validação (do inglês, Standard Deviation of Validation)
SIMCA	Modelagem Independente Suave por Analogia de Classe (do inglês, <i>soft independent modelling of class analogy</i>)
SSQ_y,c	Soma do Quadrado Y da calibração (do inglês, <i>Sum of Y Square of calibration</i>)
SSQ_y,cv	Soma do Quadrado Y da validação cruzada (do inglês, <i>Sum of Y Square of cross-validation</i>)
TAG	Triacilglicerol
VL	Variável Latente

SUMÁRIO

1. INTRODUÇÃO.....	17
2. BIODIESEL.....	21
2.1. Cooperação entre Moçambique e Brasil na área dos biocombustíveis.....	22
2.2. Matérias-primas.....	24
2.3. Produção de biodiesel.....	27
2.4. Teor de biodiesel em diesel e normas para seu monitoramento.....	31
3 ESPECTROSCOPIA DE RESSONÂNCIA MAGNÉTICA NUCLEAR E MÉTODOS QUIMIOMÉTRICOS.....	35
3.1. ESPECTROSCOPIA DE RESSONÂNCIA MAGNÉTICA NUCLEAR.....	35
3.1.1. Mecanismo de absorção de energia na RMN.....	36
3.1.2. Espectrômetro de RMN.....	38
3.2. MÉTODOS QUIMIOMÉTRICOS.....	38
3.2.1. Projeções ortogonais à estrutura latente (OPLS).....	40
3.2.1.1. Detecção de Amostras Anômalas (<i>Outlier</i>).....	42
3.2.1.2. Validação dos modelos OPLS.....	43
3.2.2. Métodos de classificação.....	47
3.2.2.1 Análise discriminante por projeções ortogonais à estrutura latente (OPLS-DA).....	48
3.2.2.1.1 Análise da eficiência dos Modelos OPLS-DA.....	50
3.2.2.1.2 Representação gráfica do resultado do modelo OPLS-DA.....	51
4 Objetivos.....	53
4.1. Objetivo Geral.....	53
4.1.1. Objetivos Específicos.....	53
5 PROCEDIMENTO EXPERIMENTAL.....	54
5.1. Produção do Biodiesel.....	54
5.2. Preparação das Amostras.....	56
5.3. Obtenção dos Espectros de RMN de ^1H.....	56
5.4. Construção dos modelos OPLS e OPLS-DA.....	57
6 RESULTADOS E DISCUSSÃO.....	59
6.1. Caracterização dos Espectros de RMN de ^1H.....	59
6.2. Modelos OPLS.....	62

6.3. Modelos OPLS-DA.....	69
7 CONCLUSÕES.....	85
REFERÊNCIAS.....	86

1. INTRODUÇÃO

A possibilidade de escassez dos combustíveis fósseis a médio ou longo prazo e a necessidade de reduzir os impactos ambientais causado pelo seu processo de produção, faz com que seja necessário potencializar as fontes de energia renováveis como alternativa a dependência por combustíveis fósseis (PIRES DE OLIVEIRA; CAIRES, 2019). No caso do diesel, várias pesquisas apontam a possibilidade de sua substituição parcial ou completa pelo biodiesel por ser um produto de fontes renováveis que apresenta propriedades similares ao diesel, e ser livre de compostos aromáticos e de enxofre (JANAUN; ELLIS, 2010).

O biodiesel pode ser produzido a partir de reações de esterificação e transesterificação de uma matéria-prima que contenha ácidos graxos livres e ou triglicerídeos, como óleos vegetais, óleos usados, gorduras animais, entre outros (JANAUN; ELLIS, 2010). O seu uso tem se destacado na União Europeia, Estados Unidos, Brasil e Argentina com 43%, 15%, 13% e 7% da produção mundial, respectivamente (NAYLOR; HIGGINS, 2018). A produção do biodiesel nestes países é feita na sua maioria a partir de matéria-prima de primeira geração, isto é, dos óleos que também são utilizados na alimentação humana, tais como de soja que em 2018 representou 69,82% da matéria-prima usada para produção de 5.35 bilhões de litros de biodiesel no Brasil e 100% na Argentina, colza 49% na União Europeia, palma 14% na União Europeia, entre outros (BRASIL, 2019; NAYLOR; HIGGINS, 2018). Esses dados mostram que a produção de biodiesel pode contribuir para o aumento do custo do óleo vegetal por estabelecer uma concorrência ao uso do óleo como alimento. Assim, como forma de não incentivar o seu uso para garantir a segurança alimentar, a Diretiva 2012/0288/CE da Comissão Europeia estabelece que a partir de 2020, o biodiesel produzido com base na matéria-prima de primeira geração não será subsidiado (EUROPEAN COMMISSION DIRECTIVE 2012/0288 (COD)C7-0337/12, 2012).

Portanto, ao se materializar a Diretiva 2012/0288/CE (EUROPEAN COMMISSION DIRECTIVE 2012/0288 (COD)C7-0337/12, 2012), o uso de matéria-prima de segunda geração (óleos que não são usados na alimentação humana, tais como de: crambe (*Crambe abyssinica*) (TENFEN WAZILEWSKI et al., 2013), mafurra (*Trichilia emetica*) (KOMANE; OLIVIER; VILJOEN, 2011), fritura usado (MAZIVILA et al., 2015a), jatropha (*Jatropha curcas L.*) (PANDEY et al., 2012), karanja (PATEL; SANKHAVARA, 2017), entre outros) constituirá destaque em vários países, principalmente crambe e mafurra por serem oleaginosas promissoras na produção de biodiesel, pois a partir das suas sementes pode se obter até 35% e 65% de um óleo, respectivamente, que possui alto teor de ácidos graxos de alto peso

molecular, o que confere melhor qualidade ao biodiesel por apresentar boa estabilidade oxidativa quando comparado ao óleo de soja (TENFEN WAZILEWSKI et al., 2013).

A comercialização do biodiesel em vários países é feita em mistura com diesel. Esta mistura pode ser feita em diferentes percentagens, de acordo com a legislação de cada país, mas no geral ela começou a ser feita em pequena percentagem B2 (2% de biodiesel e 98% de diesel). Devido ao seu sucesso tem evoluído para percentagens mais elevadas. Por exemplo, na França a mistura evoluiu de B5 (5% de biodiesel e 95% de diesel) para B30 (30% de biodiesel e 70% de diesel); nos EUA, de B2 para B20; em Moçambique, de B3 para B7,5; no Brasil, de B2 para B10, com perspectiva de evoluir para B15 em 2023 (BAMBO, 2014; BRASIL, 2014; MAZIVILA et al., 2015b). Esta evolução além de representar somente a esperança de haver uma substituição total do diesel pelo biodiesel a médio ou longo prazo, também representa preocupações em relação a sua comercialização com teores fora do que é exigido pelas normas. O Brasil é um exemplo disso porque mesmo com a exigência de adição do biodiesel ao diesel em pequenas percentagens, a característica teor de biodiesel sempre representou mais de 40% das não conformidades observadas para o diesel comercializado no território brasileiro (ANP, 2019). Com isso, o desenvolvimento de métodos analíticos eficientes, rápidos e de amostragem direta é uma necessidade constante.

Existem pesquisas que relatam o uso de uma técnica analítica associada a algum tipo de análise estatística de dados, uni e/ou multivariada, como sendo bastante eficazes para previsão. Como são os casos da espectroscopia MIR associados à calibração multivariada por: Quadrados Mínimos Parciais (PLS) para identificar e quantificar o adulterante (óleo de soja) em misturas de diesel/biodiesel de gordura animal (CÂMARA et al., 2017) e o teor de biodiesel metílico de Algodão em mistura com diesel (BUIATTE et al., 2016); Análise Discriminante por Quadrados Mínimos Parciais (PLS-DA) para discriminar amostras de diesel dos quatro principais fornecedores de produtos petrolíferos em Marrocos (BARRA et al., 2020) e para realizar o controle de qualidade de biodiesel de soja e pinhão manso, discriminando as amostras de acordo com as matérias-primas utilizadas no seu processo de produção (óleo e álcool) (MAZIVILA et al., 2015b).

Para determinação do teor de biodiesel em mistura com o diesel, a norma brasileira ABNT NBR 15568 recomenda o uso da espectroscopia no infravermelho médio (MIR) e calibração multivariada por PLS (ABNT NBR-15568, 2008). Como alternativa à Espectroscopia no MIR, a Ressonância Magnética Nuclear de Hidrogênio (RMN de ^1H) combinada com calibração multivariada é bastante utilizada em diversas análises (FAN et al., 2018; JOVIĆ et al., 2019; MONAKHOVA; DIEHL, 2015; VIEIRA et al., 2019) por envolver

geralmente os núcleos dos átomos de hidrogênio no processo de absorção, o que torna possível determinar o número de cada um dos diferentes tipos de prótons não equivalentes (PAVIA et al., 2010). Outras vantagens da Ressonância Magnética Nuclear sobre as técnicas ópticas são: oferece informações sobre o número de átomos magneticamente distintos do isótopo estudado; as fórmulas estruturais, moleculares e geométricas de compostos orgânicos; realiza simultaneamente análises qualitativas e quantitativas (PAVIA et al., 2010; SHIMAMOTO; TUBINO, 2016).

As regressões multivariadas por PLS e PLS-DA usadas para calibração e classificação, respectivamente, foram estendidas para projeções ortogonais à estrutura latente (OPLS) e análise discriminante por projeções ortogonais à estrutura latente (OPLS-DA) com objetivo de melhorar os resultados dos modelos porque esses métodos conseguem remover a variação ortogonal-Y na matriz de dados. Esta realidade é evidenciada em alguns trabalhos recentemente publicados pelos autores Silva et al. (SILVA et al., 2017) em que fizeram uma comparação entre os métodos OPLS e PLS no controle estatístico de processos de co-cristalização; Souihi et al. (SOUIHI et al., 2015) ao mostrar novas oportunidades que o método OPLS oferece no monitoramento de lotes.

A presente tese está dividida em seis capítulos, sendo:

O primeiro capítulo (Introdução), além do que foi descrito acima, também descreve de forma resumida os restantes capítulos.

O segundo capítulo (Biodiesel) apresenta as vantagens do uso do biodiesel; cooperação entre Moçambique e Brasil na área dos biocombustíveis; evolução da percentagem do teor de biodiesel no diesel comercializado no Brasil e em Moçambique; as matérias-primas utilizadas para produção de biodiesel no Brasil e em Moçambique; a produção do biodiesel e índice de não conformidade com relação à natureza do diesel comercializado no território brasileiro.

O terceiro capítulo (Espectroscopia de Ressonância Magnética Nuclear e Métodos Quimiométricos) descreve o princípio; vantagens da espectroscopia de RMH para o controle de qualidade de biocombustíveis e mecanismo de absorção de energia na RMN; contextualiza os métodos quimiométricos usados (OPLS e OPLS-DA), a partir do princípio geral de análise multivariada; as áreas de aplicação da Quimiometria e, até a organização dos dados espectrais.

O quarto capítulo (Objetivo) apresenta os objetivos da tese.

O quinto capítulo (Procedimento Experimental) descreve o processo de produção de biodiesel, preparo de amostras dos conjuntos de calibração e previsão, aquisição dos respectivos dados espectrais de RMH ^1H , a construção e validação dos modelos OPLS e OPLS-DA.

O sexto capítulo (Resultados e Discussão) apresenta a caracterização dos espectros das amostras e análise dos resultados obtidos em cada um dos modelos OPLS e OPLS-DA de biodiesel metílico de mafurra (BMMa) e de crambe (BMC).

O sétimo capítulo (Conclusões) sumariza as principais constatações que demonstram o alcance dos objetivos do trabalho.

Todas as referências usadas para a produção desta tese são apresentadas depois das conclusões.

2. BIODIESEL

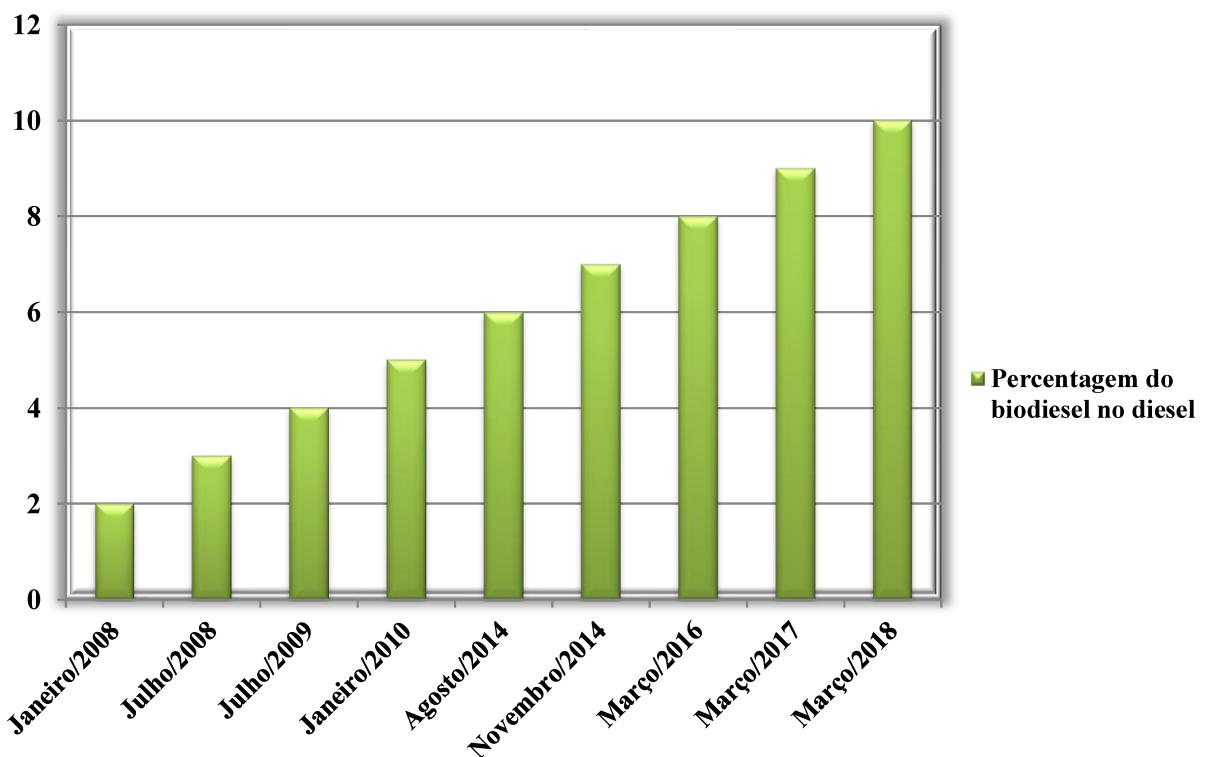
O biodiesel é um combustível composto por alquil ésteres de ácidos carboxílicos de cadeia longa, produzido a partir da transesterificação e/ou esterificação de matérias-graxas, de gorduras de origem vegetal ou animal (ANP - AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (BRASIL), 2012). Este combustível apresenta propriedades similares com diesel, razão pela qual é usado em vários países na mistura diesel/biodiesel.

No Brasil, o biodiesel foi introduzido na matriz energética mediante a publicação da Lei nº 11.097, no Diário Oficial da União, em 2005 (BRASIL, 2005). O país se destaca como um dos potenciais produtores e consumidores de biodiesel devido à criação do Programa Nacional de Produção e Uso do Biodiesel (PNPB) como ação estratégica e prioritária para implantar de forma sustentável a produção e uso do biodiesel, com ênfase no desenvolvimento regional (BRASIL, 2003). O uso de biodiesel proporciona algumas vantagens (BALAT; BALAT, 2009; QUINTELLA et al., 2009):

- *Ecológicas:* os gases de combustão emitidos pelos motores que operam com biodiesel não contêm óxidos de enxofre, principal causador da chuva ácida e irritações das vias respiratórias. Além disso, o ciclo de produção agrícola que origina as matérias-primas para o biodiesel capta CO₂ da atmosfera durante o período de crescimento, auxiliando o controle do efeito estufa.
- *Macroeconômicas:* a demanda por produtos agrícolas gera oportunidades de emprego e renda para a população rural. Além disso, o aproveitamento interno dos óleos vegetais e gorduras animais fortalecem o agronegócio e a agropecuária.
- *Financeiras:* redução da importação de diesel.

A produção de biodiesel no Brasil aumenta anualmente, tendo atingido 5.350.036 m³ em 2018 contra 4.291.294 m³ do ano anterior, disponibilizando desta forma mais 24,7% de biodiesel no mercado interno (BRASIL, 2018a, 2019). Este aumento favorece a evolução da percentagem do teor de biodiesel no diesel, como se pode ver na Figura 1, a adição obrigatória de biodiesel no diesel foi feita de forma gradual a partir do ano 2008 com 2% até 10% em 2019 e, com perspectiva de ser 15% em 2023.

Figura 1 – Evolução da percentagem do teor de biodiesel no diesel comercializado no Brasil.



Fonte: Adaptado de (BRASIL, 2014).

2.1. Cooperação entre Moçambique e Brasil na área dos biocombustíveis

Moçambique, país localizado no sudeste do Continente Africano, banhado pelo oceano Índico a leste e que faz fronteira com a Tanzânia ao norte; Malawi e Zâmbia a noroeste; Zimbábwe a oeste; Suazilândia e África do Sul a sudoeste possui uma abundância relativa de terras com condições agroclimáticas favoráveis para o cultivo de várias matérias-primas usadas para a produção de biocombustíveis no Brasil (ARNDT et al., 2010; BATIDZIRAI; FAAIJ; SMEETS, 2006; CUMBE, 2007). Por esta razão, um acordo de cooperação na área de biocombustíveis foi firmado por Moçambique e Brasil em 2007, com os seguintes objetivos (BAMBO, 2014):

- Promover trocas de missões técnico-empresariais, em datas de conveniência mútua;
- Apoiar o desenvolvimento de programa de produção e uso de biocombustíveis em Moçambique;

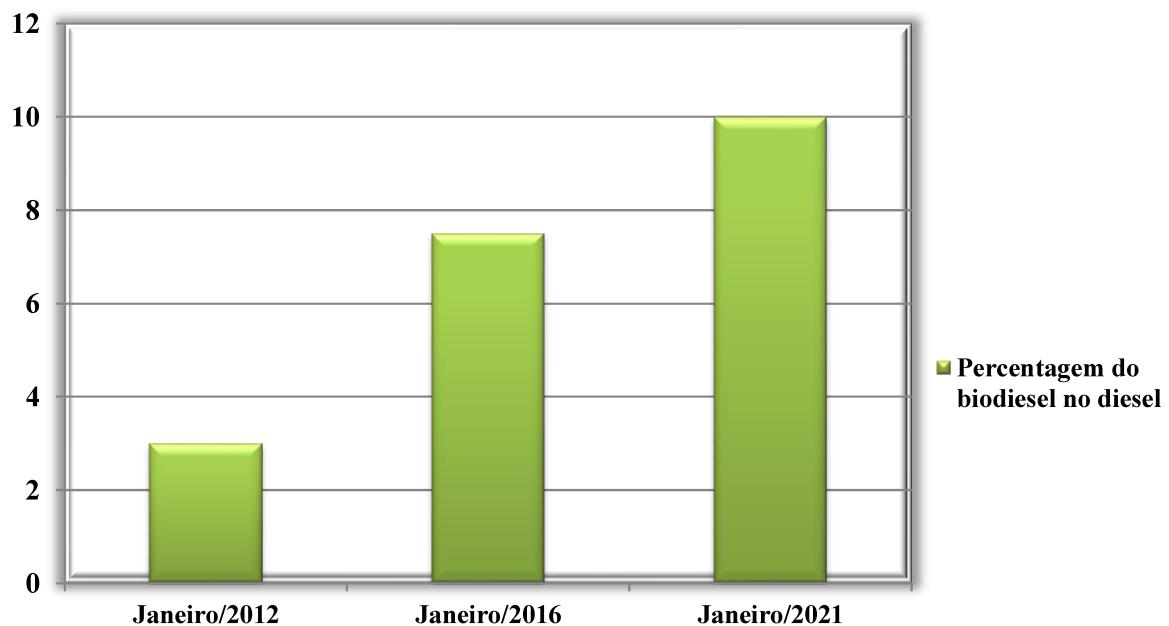
- Elaborar, conjuntamente, projetos de cooperação técnica para capacitar recursos humanos do setor de biocombustíveis em Moçambique;
- Identificar parcerias com outros países e organismos internacionais interessados em apoiar projetos específicos para a implantação do programa de desenvolvimento dos biocombustíveis em Moçambique.

Dois anos depois do acordo, o Governo de Moçambique introduziu os biocombustíveis na sua matriz energética com a aprovação da Política e Estratégias de Biocombustíveis no dia 24 de Março de 2009, tendo em conta que os mesmos proporcionariam os seguintes benefícios (MOÇAMBIQUE, 2009):

- A substituição gradual dos combustíveis fósseis, sem nenhuma ou com uma adaptação tecnológica mínima do equipamento, com impactos positivos na balança de pagamentos e na redução da dependência externa do país;
- A geração de emprego e auto-emprego na agricultura e no processamento agroindustrial, particularmente nas zonas rurais, gerando ocupação para a abundante, disciplinada e treinável mão de obra moçambicana;
- A possibilidade de produção de biocombustíveis seguindo um modelo de desenvolvimento agrícola apropriado à prevenção da predominância excessiva de monoculturas e que assegure, ao mesmo tempo, a produção de outras culturas;
- Melhor uso das infraestruturas existentes, incluído os portos das águas profundas, ferrovias, estradas, oleodutos e gasodutos;
- Melhor aproveitamento da água e recursos climáticos favoráveis para produzir biocombustíveis em grande escala para o consumo interno e exportação;
- Melhor aproveitamento das oportunidades oferecidas pelo mercado regional da Comunidade para o Desenvolvimento da África Austral (SADC).

O teor obrigatório de biodiesel na mistura com diesel a ser comercializado em território moçambicano foi estabelecido no Regulamento de Biocombustíveis e suas Misturas com combustíveis fósseis, aprovado em 2011 pelo Decreto nº 58/2011 (MOÇAMBIQUE, 2011). Para Moçambique, a percentagem do teor biodiesel no diesel evoluiu de 3% em 2012 até 7,5% em 2020, com perspectiva de ser 10% em 2021, como se pode ver na Figura 2.

Figura 2 – Evolução da percentagem do teor de biodiesel no diesel comercializado em Moçambique.



Fonte: Adaptado de (MOÇAMBIQUE, 2011).

2.2. Matérias-primas

As principais matérias-primas usadas para a produção de biodiesel são: óleos vegetais, gorduras animais, óleos de frituras residuais e matérias graxas de alta acidez. No entanto, a escolha da matéria-prima depende de alguns fatores como a geografia, o clima, a economia e a tecnologia da planta de produção. Devido a estes fatores, vários países têm optado em utilizar óleos vegetais que mais se adaptam a sua realidade, como por exemplo, nos Estados Unidos o óleo mais utilizado é o de soja; na Europa é óleo de colza; nos países tropicais a preferência recai para o óleo de soja e de palma (KNOTHE et al., 2006).

Em Moçambique, existem vários óleos vegetais disponíveis para produção de biodiesel, como: óleo de pinhão manso, coco, girassol, amendoim, soja, dentre outros. Mas, levando em conta aspectos como a sustentabilidade, a avaliação dos seus rendimentos, custos de produção, impactos socioeconômicos e ambientais, atualmente o óleo de pinhão manso é mais utilizado e incentivado pelo Governo que orienta aos produtores locais dessa matéria-prima a entregar sua produção exclusivamente aos titulares licenciados de produção,

armazenagem e distribuição de biocombustíveis, para consequente introdução na cadeia produtiva do biodiesel.

No Brasil, as matérias-primas comuns, empregadas na produção de biodiesel são: óleos de soja, algodão, palma e de fritura, gordura bovina, gordura de porco e de frango. A quantidade de cada matéria-prima usada nos últimos 5 anos é apresentada na Tabela 1, na qual se observa que a produção do biodiesel no Brasil sempre foi feita na sua maioria a partir de matéria-prima de primeira geração (óleo de soja que também é utilizado na alimentação humana), que atingiu 3,7 bilhões de litros de óleo de soja, representando 69,82% da matéria-prima usada para produção de 5,35 bilhões de litros de biodiesel no ano 2018. Esse dado mostra que a produção de biodiesel pode contribuir para o aumento do custo do óleo por estabelecer uma concorrência ao uso do óleo como alimento. Por este motivo, pesquisas visando à introdução de matérias-primas alternativas aos óleos vegetais alimentares na matriz brasileira tornam-se importantes.

Culturas como a mafurra (*Trichilia emetica*) e o crambe (*Crambe abyssinica*) que são muito produzidas em Moçambique e no Brasil, respectivamente, constituem alvo de estudo neste trabalho por serem oleaginosas que de uma forma geral os seus óleos não são usados como alimentos (KOMANE; OLIVIER; VILJOEN, 2011; TENFEN WAZILEWSKI et al., 2013).

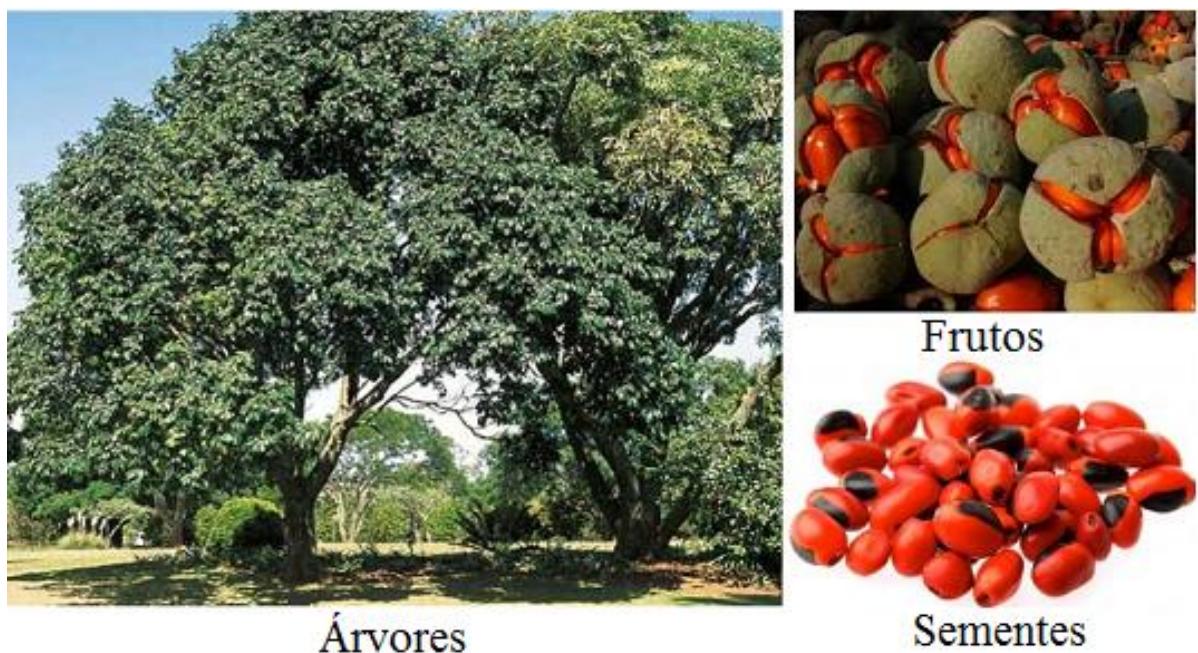
Tabela 1 – Quantidade de matéria-prima usada para produção do biodiesel nos últimos 5 anos no Brasil

Matéria-prima (m ³)	2014	2015	2016	2017	2018
Óleo de soja	2.625.558	3.061.027	3.020.819	3.072.446	3.703.066
Óleo de palma	1.026	3.336	15.534	36.436	70.710
Óleo de algodão	76.792	78.840	39.628	12.426	49.175
Óleo de amendoim	0	0	0	0	5.349
Óleo de canola	0	158	7.234	11.762	4.059
Sebo	644.382	712.670	576.429	586.761	704.012
Outros materiais graxos	14.119	39.656	85.428	383.164	523.993
Gordura de frango	15.831	1.797	6.924	46.981	43.772
Gordura de porco	15.648	24.452	38.957	87.193	112.411
Óleo de fritura usado	22.110	16.935	26.101	52.182	87.086

Fonte: (BRASIL, 2019)

A árvore da mafurra (Figura 3) é uma planta da família das meliáceas de folha perene, tem uma preferência por áreas com uma elevada precipitação pluviométrica e é abundante ao longo de áreas costeiras. As suas sementes contêm um teor de óleo de 55-65% (KOMANE; OLIVIER; VILJOEN, 2011).

Figura 3 – Árvores, frutos e sementes de mafurra.



Fonte: O autor.

O crambe (Figura 4) é uma planta herbácea perene com altura entre trinta a oitenta centímetros e tem origem na região mediterrânea. É muito resistente às secas e geadas, possui um ciclo de cultivo muito reduzido, o que proporciona baixos custos de produção. É uma oleaginosa promissora na produção de biodiesel, pois a partir das suas sementes pode se obter até 35% de óleo que possui alto teor de ácidos graxos de alto peso molecular, o que confere melhor qualidade ao biodiesel, além de apresentar melhor estabilidade oxidativa quando comparado ao óleo de soja (SITOÉ et al., 2016).

Figura 4 – Plantas, frutos e sementes de crambe.



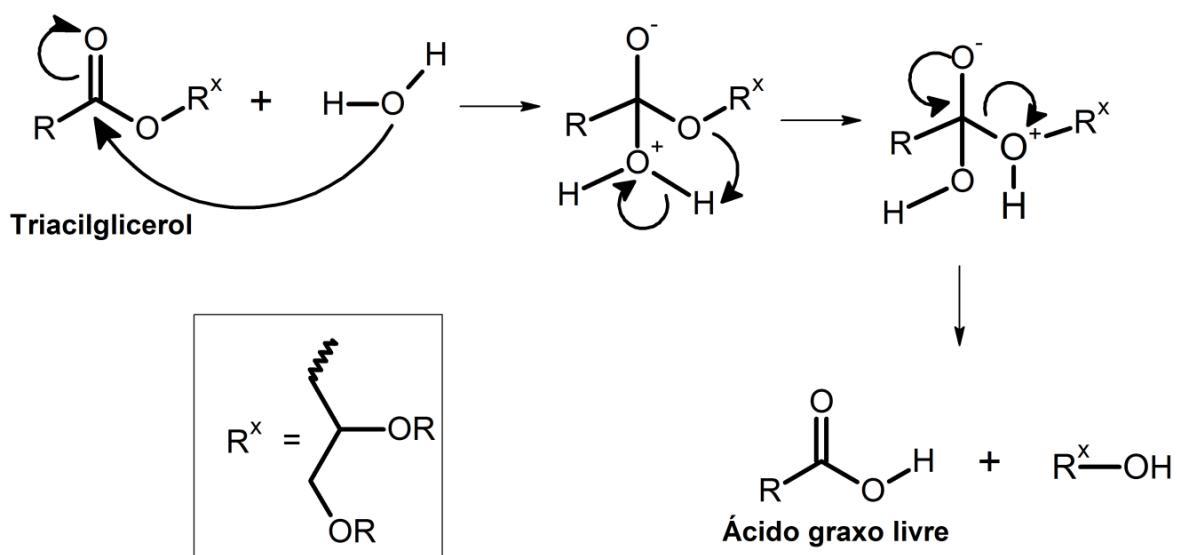
Fonte: Adaptado de (“Crambe hispanica – Wikipédia, a enciclopédia livre”, 2014).

2.3. Produção de biodiesel

Os processos mais usados para a produção do biodiesel são a transesterificação e a esterificação. Para tal, os óleos vegetais são submetidos às respectivas reações químicas, obedecendo às seguintes etapas:

Na primeira etapa é feita o pré-tratamento da matéria-prima, filtrando os óleos para remover contaminantes sólidos e eliminar o máximo de água para evitar a ocorrência da hidrólise dos ésteres alquílicos sintetizados a três moléculas de ácidos graxos. Da mesma forma, a reação de triacilgliceróis (TAG) com água pode formar ácidos graxos livres (KNOTHE et al., 2006). A equação da reação de formação dos ácidos graxos livres é mostrada na Figura 5.

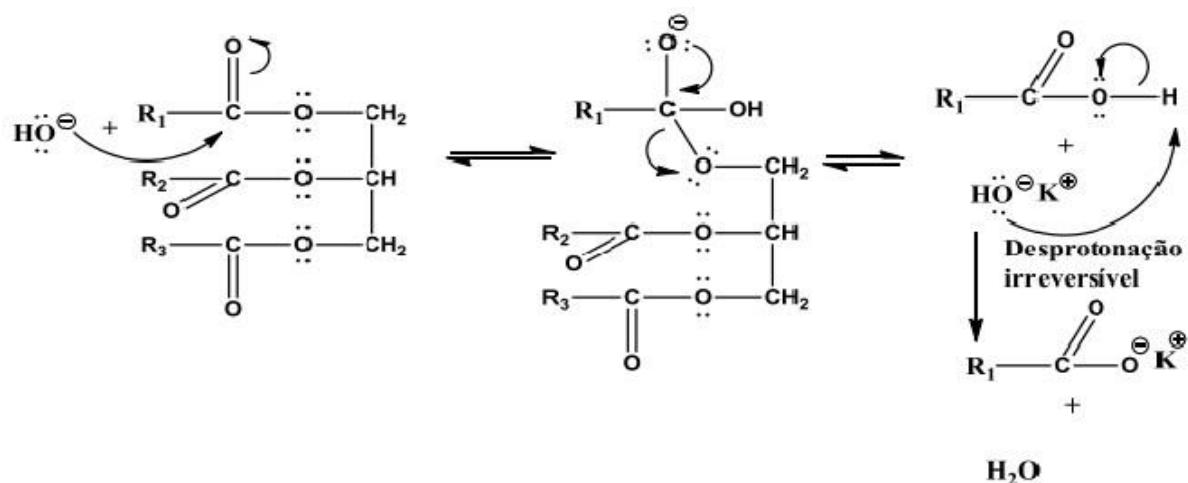
Figura 5 – Mecanismo da reação de hidrólise de triacilgliceróis na presença de umidade.



Fonte: (CLAYDEN; GREEVES; WARREN, 2012).

Na segunda etapa é feita a determinação e o tratamento dos ácidos graxos livres. Até o nível de 5% de ácidos graxos livres, a reação ainda pode ser realizada com catalisadores alcalinos porque os sabões formados durante a reação são facilmente removidos com o glicerol ou são eliminados durante a etapa de lavagem aquosa, mas para tal uma quantidade adicional de catalisador deve ser utilizada para compensar a perda para reações de saponificação (KNOTHE et al., 2006). Quando a concentração de ácidos graxos livres for superior a 5%, não é recomendado o uso de catalisadores alcalinos porque a sua reação com os ácidos graxos livres (Figura 6) forma sabões que inibem a separação de fases entre o glicerol e os ésteres metílicos e contribuem para a formação de emulsões durante a lavagem aquosa (LEUNG; WU; LEUNG, 2009).

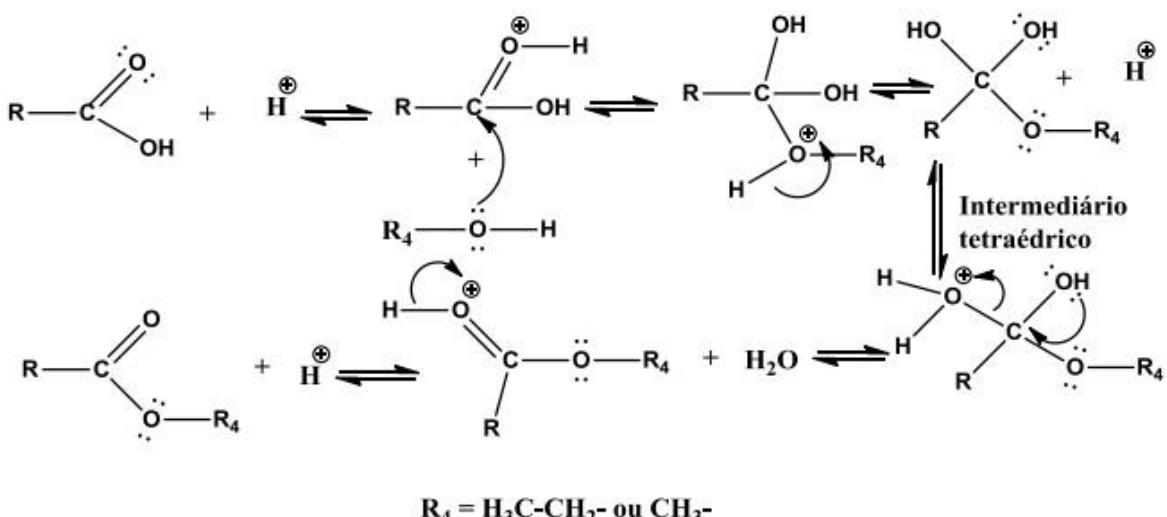
Figura 6 – Mecanismos de saponificação de triglicerídeos.



Fonte: (CLAYDEN; GREEVES; WARREN, 2012).

Para evitar o processo de saponificação nos óleos que contém ácidos graxos livres superiores a 5%, é necessária a utilização de um catalisador ácido como o ácido sulfúrico para converter os ácidos graxos livres em ésteres metílicos, como demonstra o mecanismo da reação apresentado na Figura 7.

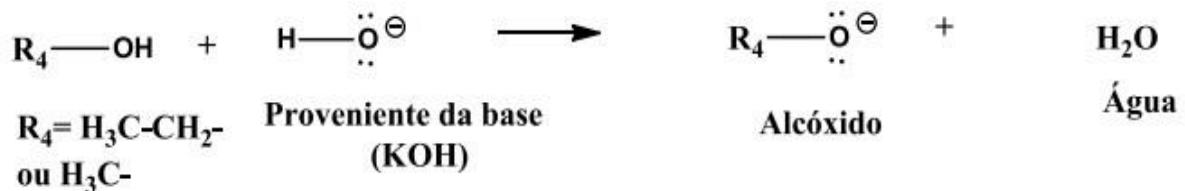
Figura 7 – Mecanismo de reação de esterificação via catálise ácida.



Fonte: (CLAYDEN; GREEVES; WARREN, 2012).

Na terceira etapa é realizada a reação de transesterificação. Nesta reação, óleos vegetais reagem na presença de um catalisador (ácido ou básico) com um álcool de cadeia curta, geralmente metanol ou etanol (KNOTHE et al., 2006). No entanto, os catalisadores alcalinos (hidróxido de sódio e de potássio; ou os alcóxidos correspondentes) são mais viáveis financeiramente e proporcionam processos muito mais rápidos que catalisadores ácidos (KEE LAM; TEONG LEE; RAHMAN MOHAMED, 2010). As soluções de alcóxidos são obtidas adicionando lentamente a base padronizada a um excesso do álcool correspondente (Figura 8). Estas soluções são mais utilizadas que os respectivos hidróxidos porque não permitem ocorrer a reação de formação de água, assegurando que o processo de transesterificação permaneça livre de água tanto quanto possível (KEE LAM; TEONG LEE; RAHMAN MOHAMED, 2010). Uma das desvantagens do uso do etanol em relação ao metanol é o de promover maior dispersão da glicerina, co-produto na produção de biodiesel, o que torna mais difícil a separação das fases, proporcionando maior custo de produção e menor rendimento da mesma.

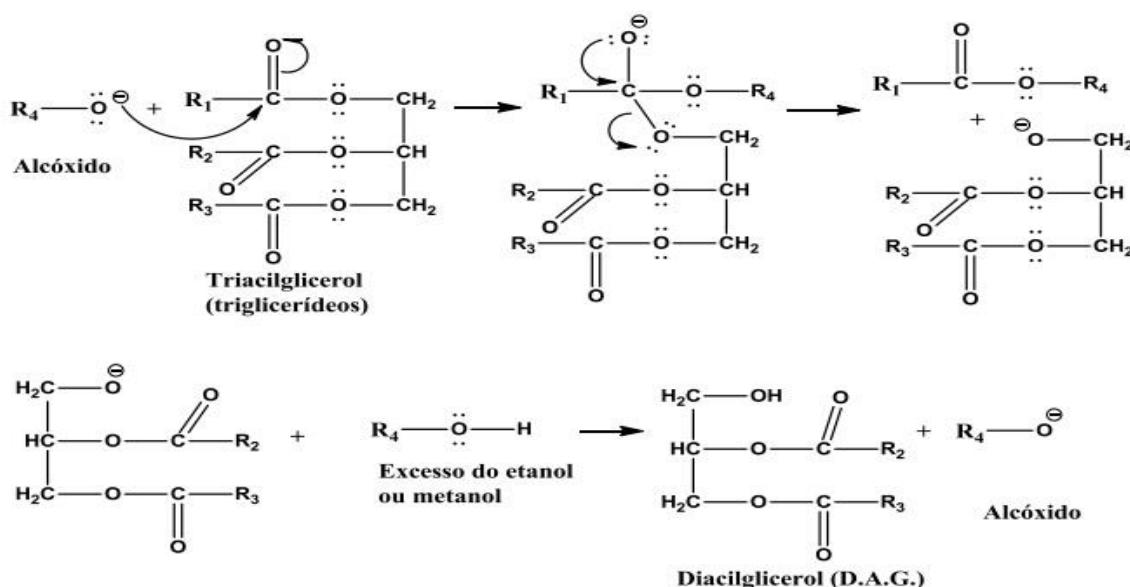
Figura 8 – Reação da formação de alcóxido.



Fonte: (CLAYDEN; GREEVES; WARREN, 2012).

Para este trabalho, a transesterificação foi realizada reagindo os óleos vegetais com o metanol, na presença de alcóxido de potássio como catalisador, como demonstra o mecanismo da reação apresentado na Figura 9 (DONG et al., 2013). A produção de biodiesel de mafurra e crambe foi feita devido à dificuldade de encontrá-los no comércio.

Figura 9 – Mecanismo de reação de transesterificação via catálise alcalina.



Fonte: (CLAYDEN; GREEVES; WARREN, 2012).

Na quarta e última etapa é feita separação do biodiesel por decantação e em seguida é feita a sua purificação usando o processo de lavagem com água quente para remover os produtos gerados durante a sua produção, entre eles estão: glicerol e água residual. Por fim, é feito o aquecimento para a secagem e a remoção da umidade.

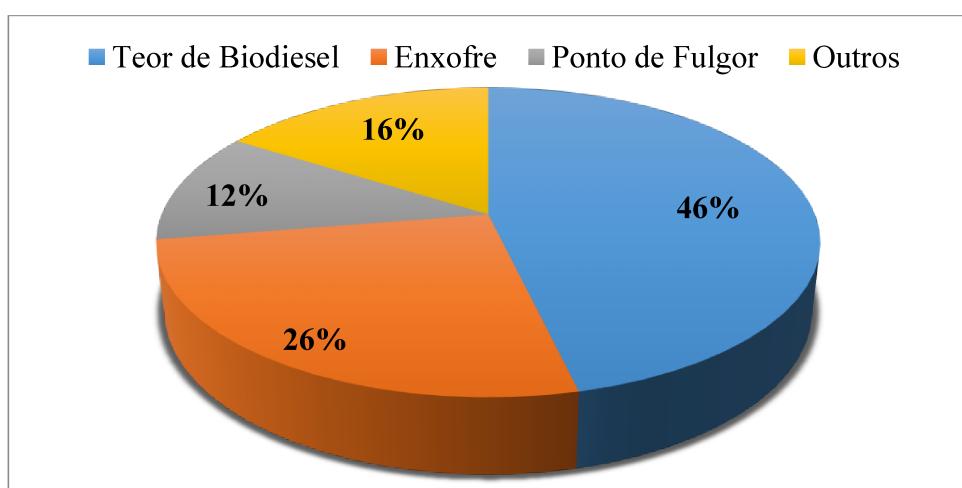
2.4. Teor de biodiesel em diesel e normas para seu monitoramento

O biodiesel é miscível com o diesel em qualquer proporção porque apresentam propriedades físico-químicas semelhantes, como por exemplo, a cor, viscosidade e densidade. Por esta razão, muitos países têm optado pelo uso de misturas binárias biodiesel/diesel, ao invés do biodiesel puro devido às seguintes vantagens: a possibilidade do aproveitamento da rede de distribuição de combustível já existente e a utilização desta mistura nos motores atuais sem qualquer modificação em uma percentagem de até 20,0% (v/v) (MOFIJUR et al., 2013). No caso do Brasil, desde 23 de março de 2018, esta mistura passou a ser comercializada obrigatoriamente na seguinte proporção: 10% de biodiesel para 90% de diesel puro, formando um combustível denominado de B10 (BRASIL, 2014).

Atualmente, a produção do biodiesel no Brasil é subsidiada pelo governo e a sua comercialização é feita por meio de editais de leilões como forma de garantir que o biodiesel

arrematado seja destinado à mistura com o óleo diesel nas condições previstas pela legislação vigente (BRASIL, 2018b). De janeiro a dezembro de 2018, a ANP realizou seis leilões para a compra de biodiesel pelas distribuidoras de combustível, totalizando 64 desde o início do programa. Mesmo assim, a comercialização da mistura diesel/biodiesel com teores fora do que é exigido pela norma ainda é feita em grande percentagem. Este cenário é comprovado com os dados estatísticos de monitoramento da qualidade dos combustíveis líquidos automotivos do ano 2019, em que mostram que a característica teor de biodiesel representou uma média de 46,3% das não conformidades observadas para o diesel comercializado. As outras características são devido ao ponto de fulgor, teor de enxofre e outros, como se pode ver na Figura 10.

Figura 10 – Índice de não conformidade do diesel comercializado no Brasil.



Fonte: (ANP - AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (BRASIL), 2019).

Diante do exposto, se faz necessário o desenvolvimento de métodos que forneçam resposta rápida, precisas e exatas no monitoramento quantitativo e qualitativo do teor B10 na mistura biodiesel/diesel. Para tal, o método desenvolvido deve apresentar eficiência igual ou melhor que a de um método padrão, para determinar o teor de biodiesel na mistura com diesel recomenda-se as seguintes normas:

- Norma Européia EN 14078 (do inglês, European Standard) (EUROPEAN STANDARD DIN EN 14078, 2004);
- Associação Brasileira de Normas Técnicas ABNT NBR 15568 (ABNT NBR-15568, 2008);

- Sociedade Americana de Testes e Materiais ASTM D7371-14 (do inglês, American Society for Testing and Materials) (ASTM STANDARD D7371, 2014);
- Sociedade Americana de Testes e Materiais ASTM E1655-05 (do inglês, American Society for Testing and Materials) (ASTM STANDARD E1655-05, 2012).

A Norma Européia EN 14078 envolve diluição das amostras em hexano e utiliza a Espectroscopia na região do MIR, baseando na obtenção de uma curva de calibração (univariada) que emprega o sinal de absorção associado à vibração de estiramento da carbonila ($C=O$) presente nos ésteres ($\sim 1745\text{ cm}^{-1}$). Assim, é impossível ter um modelo linear único e simples, abrangendo toda a gama de combinações possíveis.

A norma ABNT NBR 15568 sugere a determinação do teor de biodiesel metílico e/ou etílico em óleo diesel usando modelo de calibração multivariada, baseado no método quimiométrico de regressão por Quadrados Mínimos Parciais (PLS), que emprega medidas de absorvância no MIR nas regiões espectrais de 1100 cm^{-1} a 1200 cm^{-1} e de 1735 cm^{-1} a 1750 cm^{-1} , atribuídas ao estiramento das ligações C-O e C=O de ésteres, respectivamente. Para tal, são apresentados dois modelos de regressão, em que um é para a faixa de concentração de biodiesel de 0% (v/v) a 8% (v/v) e outro para a faixa de 8% (v/v) a 30% (v/v).

A ASTM D7371-14 também utilizam a Espectrometria no MIR, através de medições de Reflectância Total Atenuada (ATR) aliada ao método quimiométrico de regressão por Quadrados Mínimos Parciais (PLS). Esta norma exige que um número maior de padrões de calibração seja preparado, mas permite uma análise muito mais rápida que a norma Européia EN 14078, pois não há necessidade de diluição das amostras. No entanto, para a obtenção da exatidão necessária é recomendada a utilização de dois ou três modelos analíticos. Assim, a norma 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).

Para a validação do modelo de calibração é recomendado seguir as orientações da norma ASTM E1655-05 que descreve um guia prático de procedimentos para coleta e tratamento de dados na quantificação de analitos quando se utiliza a Espectrometria no Infravermelho.

Portanto, pesquisadores têm desenvolvido métodos de quantificação visando obter resultados mais satisfatórios e adequações às Normas estabelecidas, como são os casos: Sitoé et al. quantificaram biodiesel metílico de jatropha (pinhão manso) em misturas com diesel utilizando espectroscopia no MIR e métodos de seleção de variável por intervalo, seguindo as orientações da ASTM E1655-05 (SITOÉ et al., 2020); Portela et al. fizeram a quantificação

do biodiesel de soja e de gordura animal no diesel de petróleo por RMN de ^1H , avaliando abordagens univariada e multivariada (PORTELA et al., 2016). Buiatte et al. identificaram e quantificaram o teor de biodiesel metílico de algodão em mistura com diesel usando espectrometria no MIR e métodos quimiométrico PLS e cartas de controle multivariadas, de acordo com a norma ABNT NBR 15568, no qual as figuras de mérito usadas para validação dos modelos foram determinadas com base nas diretrizes da ASTM E1655-05 (BUIATTE et al., 2016). Faraguna et al. desenvolveram um método para determinação de diferentes teores de biodiesel no combustível diesel usando FT-MIR, com base num procedimento bem conhecido descrito nas normas ASTM D7371-14 e EN 14078 (FARAGUNA; RACAR; JUKIĆ, 2019).

Os modelos apresentados neste trabalho foram desenvolvidos e validados tendo como base os princípios das normas ABNT NBR 15568, ASTM D7371 e ASTM E1655-05, com destaque para:

- Uso da regressão multivariada por OPLS ao invés do PLS para construir um único modelo de calibração multivariada para a faixa de concentração de 1-30% (v/v) a partir de espectros de RMN de ^1H das misturas biodiesel/diesel de mafurra e crambe;
- Uso de pequenas quantidades de amostra (70 amostras);
- A não seleção de regiões espectrais para a construção dos modelos.

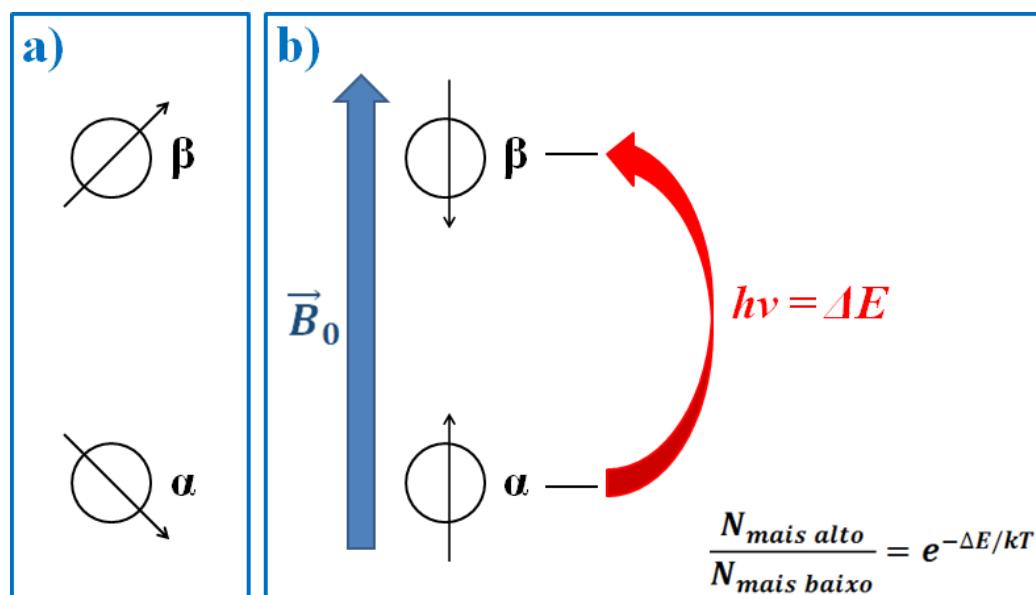
3.1. ESPECTROSCOPIA DE RESSONÂNCIA MAGNÉTICA NUCLEAR

A Ressonância Magnética Nuclear (RMN) é uma técnica espectroscópica baseada na medida de absorção de radiação eletromagnética na região de radiofrequência aproximadamente entre 4 a 900 MHz, através da interação do momento angular magnético $\vec{\mu}$ de um núcleo atômico com um campo magnético \vec{B}_0 aplicado (HOLLER; SKOOG; CROUCH, 2009). Esta técnica se fundamenta nas propriedades magnéticas dos núcleos atômicos presentes em uma determinada amostra (GRUTZNER, 2005), atendendo e considerando que os núcleos atômicos possuem um spin nuclear, I , ao qual está associado o vetor momento angular magnético, $\vec{\mu}$. A determinação do número de estados de energia (n) possíveis para os spins é feita pela Equação 1:

$$n = (2I + 1) \quad (1)$$

Como por exemplo, para o núcleo de ^1H que apresenta spin $\frac{1}{2}$, existem dois níveis de energia com um pequeno excesso de população dos hidrogênios no estado de energia mais baixa ($N_\alpha > N_\beta$), de acordo com a distribuição de Boltzmann. Os estados identificados como α e β , ($\frac{1}{2}$ e $-\frac{1}{2}$) respectivamente, apresentados na Figura 11, são degenerados quando não há um campo magnético externo (GRUTZNER, 2005).

Figura 11 – Relação de Boltzmann com **a)** ausência de um campo magnético externo e **b)** presença de um campo magnético externo.



Fonte: o autor.

A Ressonância Magnética Nuclear diferencia-se da Espectroscopia no Ultravioleta, no Visível e no Infravermelho por envolver os núcleos dos átomos no processo de absorção. Vários núcleos podem ser estudados pela técnica de RMN, mas os mais comumente disponíveis são hidrogênio e carbono. Por exemplo, quando se estudam núcleos de hidrogênio (prótons), como no caso deste trabalho, é possível determinar o número de cada um dos diferentes tipos de prótons não equivalentes (PAVIA et al., 2010).

A Ressonância Magnética Nuclear apresenta algumas vantagens sobre as técnicas ópticas, a saber: oferece informações sobre o número de átomos magneticamente distintos do isótopo estudado; é possível extrair informação sobre as fórmulas moleculares, estruturais e geométricas de compostos orgânicos; as análises qualitativas e quantitativas podem ser realizadas simultaneamente; as medições podem ser realizadas rapidamente (por um núcleo abundante, tais como ^1H), que não requer isolamento da substância a analisar, quando presente em misturas (PAVIA et al., 2010; SHIMAMOTO; TUBINO, 2016).

3.1.1. Mecanismo de absorção de energia na RMN

Para que seja possível o fenômeno de ressonância magnética nuclear, o spin nuclear deve ser diferente de zero ($I \neq 0$) para se ter pelo menos dois estados. A Tabela 2 apresenta alguns exemplos de núcleos com os respectivos estados, no qual se pode ter informação em relação sua atividade na espectroscopia de ressonância magnética nuclear (GRUTZNER, 2005).

Tabela 2 – Propriedades atômicas e magnéticas de alguns núcleos atômicos

Elemento	$^1\text{H}^1$	$^1\text{H}^2$	$^6\text{C}^{12}$	$^6\text{C}^{13}$	$^7\text{N}^{14}$	$^8\text{O}^{16}$	$^8\text{O}^{17}$	$^9\text{F}^{19}$	$^{15}\text{P}^{31}$	$^{17}\text{Cl}^{35}$
I	1/2	1	0	1/2	1	0	5/2	1/2	1/2	3/2
n	2	3	1	2	3	1	6	2	2	4

I – número quântico de spin nuclear; n – número de estados de spin na presença de \vec{B}_0

Fonte: (PAVIA et al., 2010)

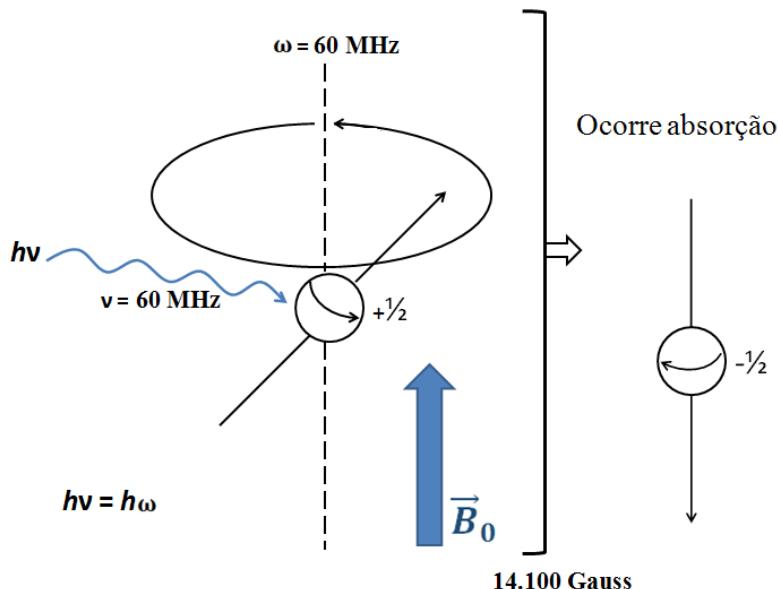
O fenômeno de ressonância magnética nuclear ocorre quando núcleos alinhados com um campo aplicado são induzidos a absorver energia e a mudar a orientação de *spin* em

relação ao campo aplicado. A absorção de energia é um processo quantizado, e a energia absorvida deve ser igual à diferença de energia entre os dois estados envolvidos, de acordo com a Equação 2. Essa diferença de energia é uma função da intensidade do campo aplicado \vec{B}_0 e, quanto mais forte for o campo magnético aplicado, maior será a diferença entre os estados de *spin* possíveis (PAVIA et al., 2010).

$$E_{\text{absorvida}} = (E_{-\frac{1}{2} \text{ estado}} - E_{+\frac{1}{2} \text{ estado}}) = h\nu \quad (2)$$

Quando o campo magnético é aplicado, o núcleo começa a mudar de direção sobre seu próprio eixo de rotação com frequência angular ω . A frequência com que um próton muda de direção é diretamente proporcional à intensidade do campo magnético aplicado: quanto mais intenso o campo magnético, maior será a velocidade (frequência angular ω) da precessão (PAVIA et al., 2010). Como o núcleo tem uma carga, a precessão gera um campo elétrico oscilatório de mesma frequência. Se as ondas de radiofrequência dessa frequência forem fornecidas ao próton que está precessando, pode haver absorção de energia, isto é, quando a frequência do componente do campo elétrico oscilatório da radiação que está entrando equivale à freqüência do campo elétrico gerado pelo núcleo que está precessando, os dois campos podem se acoplar, e será possível transferir energia da radiação para o núcleo, causando assim uma mudança de *spin*. Como se pode verificar na Figura 12, a absorção ocorre quando $\nu = \omega$ (PAVIA et al., 2010).

Figura 12 – Processo de Ressonância Magnética Nuclear.



Fonte: Adaptado de (PAVIA et al., 2010).

3.1.2. Espectrômetro de RMN

Inicialmente os espectrômetros de alta resolução eram todos do tipo de Onda Contínua (CW), usando ímãs permanentes ou eletroímãs para gerar o campo magnético (HOLLER; SKOOG; CROUCH, 2009). Estes espectrômetros operam pela excitação dos núcleos do isótopo que é observado, um tipo de cada vez. No caso de núcleos de ^1H , cada tipo distinto de próton é excitado individualmente, e seu pico de ressonância é observado e registrado de maneira independente em relação aos outros (PAVIA et al., 2010). No entanto, esse tipo de instrumento foi totalmente substituído pelos espectrômetros com Transformada de Fourier (FT), equipados com ímãs supercondutores para gerar o campo magnético (HOLLER; SKOOG; CROUCH, 2009). Estes espectrômetros utilizam uma descarga de energia de potência alta, mas curta, chamada de *pulso*, que excita simultaneamente todos os núcleos magnéticos de uma molécula. Em uma molécula orgânica, por exemplo, todos os núcleos de ^1H são induzidos a passar por ressonância ao mesmo tempo (PAVIA et al., 2010).

Normalmente, os espectros de RMN são referenciados a um sinal-padrão com deslocamento pré-definido, como por exemplo, o tetrametil-silano (TMS) e o ácido trimetilsililpropiônico (TSP) são mais usados como referência de deslocamento químico nos espectros de RMN de ^1H e ^{13}C porque apresentam um sinal atribuído aos grupos metilas. Neste trabalho, foi usado tetrametil-silano (TMS) como referência de deslocamento químico para analisar os espectros de RMN de ^1H das misturas de diesel/biodiesel de mafurra e crambe.

3.2. MÉTODOS QUIMIOMÉTRICOS

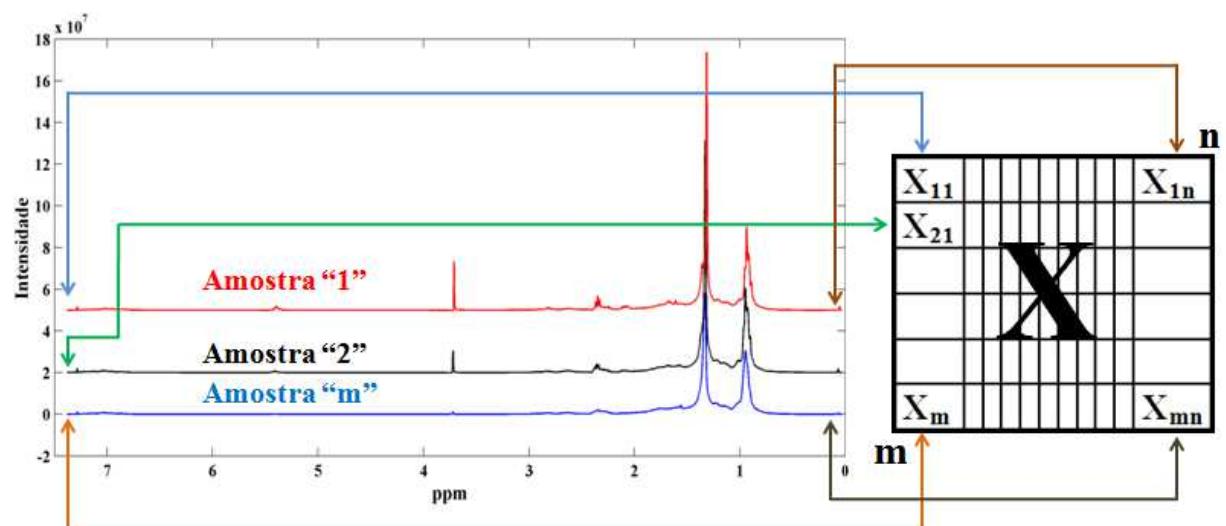
A introdução de microprocessadores e microcomputadores no laboratório químico tornou as técnicas instrumentais mais sofisticadas, que geram grande quantidade de informação química (variáveis ou características) para um pequeno ou grande número de amostras, deixando os dados mais complexos do ponto de vista matemático e estatístico para relacioná-los com a(s) propriedade(s) de interesse. No entanto, a necessidade de extrair ao máximo a informação química útil que de outra forma estaria soterrada na avalanche de dados complexos produzidos pela moderna instrumentação fez com que vários estudos relacionados com o tratamento destes dados fossem desenvolvidos (FERREIRA, 2015). Dentre vários, se destaca os estudos desenvolvidos pelos autores Pearson em 1901 (KARL PEARSON, 1901),

ao propor a análise de componentes principais que 30 anos depois foi desenvolvida por Hotelling (HOTELLING, 1933). Nos anos 70, esta técnica começou a ser usada como método de classificação em química, dando origem a uma área de química que se chama Quimiometria.

A Quimiometria é definida como a aplicação de métodos matemáticos, estatísticos e de lógica formal para o tratamento de dados químicos. 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 (MASSART et al., 1998).

Para realizar os procedimentos multivariados, os dados devem ser organizados em um arranjo ordenado de linhas e colunas, constituindo uma matriz \mathbf{X} , em que cada linha corresponde a uma amostra “ m ” e cada coluna corresponde a uma variável “ n ”, na qual $m = 1, 2, 3, \dots, m$ e $n = 1, 2, 3, \dots, n$ (FERREIRA, 2015), como se pode ver no exemplo de construção de uma matriz de dados \mathbf{X} com dimensão $(m \times n)$ a partir de um vetor de respostas instrumentais (Figura 13). Depois disso, os dados dos espectros de RMN de ^1H são alinhados usando uma rotina desenvolvida pelo autor, aplicando a função Interval-correlation-shift (icoshifit) (SAVORANI; TOMASI; ENGELSEN, 2010) no software MATLAB.

Figura 13 – Construção da matriz \mathbf{X} a partir dos dados espetrais.



Fonte: O autor.

3.2.1. Projeções ortogonais à estrutura latente (OPLS)

A regressão multivariada por OPLS é uma extensão da regressão multivariada por PLS que se constrói quando a matriz \mathbf{X} é correlacionada com matriz \mathbf{Y} (que contém a(s) propriedade(s) de interesse) através de operações matemáticas para se obter Variáveis Latentes (VLs) e Coeficientes de Regressão usados para atingir a covariância máxima entre os espectros e as concentrações das espécies de interesse e determinar o valor da concentração de cada perfil espectral, respectivamente (MASSART et al., 1998). O objetivo deste processo é encontrar um pequeno número A de fatores relevantes que são preditivos para \mathbf{Y} e que utilizam \mathbf{X} com eficiência. Para tal, a matriz \mathbf{X} é decomposta em um conjunto de fatores ortogonais que são usados para ajustar \mathbf{Y} , de acordo com a Equação 3. A matriz \mathbf{Y} é decomposta de acordo com a Equação 4:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E}_X \quad (3)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{E}_Y \quad (4)$$

Em que, \mathbf{X} e \mathbf{Y} são matrizes que contêm dados instrumentais de medição e resposta (concentração), respectivamente; \mathbf{T} e \mathbf{U} são as escores para as duas matrizes de dados; \mathbf{P} e \mathbf{Q} são os respectivos pesos, h é o número de variável latente (VL), \mathbf{E}_X e \mathbf{E}_Y são os respectivos resíduos.

A relação linear entre as duas matrizes é estabelecida pela correlação dos escores de \mathbf{X} e \mathbf{Y} para cada VL, de acordo com a Equação 5:

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h + \mathbf{E} \quad (5)$$

Em que, \mathbf{u}_h é uma matriz que contém as propriedades de todas as amostras (concentração), \mathbf{b}_h é um vetor que contém os parâmetros do modelo, \mathbf{t}_h é uma matriz que contém os dados de medição instrumental (espectro) para as amostras de calibração e \mathbf{E} é uma matriz que representa ruído.

Matrizes complexas, por exemplo, de espectros de RMN de ^1H de misturas de diesel/biodiesel, quando correlacionada com as informações da matriz \mathbf{Y} apresentam forte variação sistemáticas ortogonais (não correlacionados) a \mathbf{Y} . Estas variações desajustam o \mathbf{Y} , no entanto, devem ser corrigidas para melhorar a capacidade preditiva do modelo. Para a correção destas variações, o modelo PLS é estendido para OPLS que integra o filtro de Correção Ortogonal do Sinal (OSC).

O OPLS usa informações na matriz \mathbf{Y} para decompor a matriz \mathbf{X} em três partes distintas, conforme descrito na Equação 6, em que \mathbf{T}_p é a matriz de escores preditiva para \mathbf{X} , \mathbf{P}_p é a matriz de peso preditiva para \mathbf{X} , \mathbf{T}_o é a matriz de escores Y-ortogonal correspondente, \mathbf{P}_o é a matriz de peso dos componentes Y-ortogonais, e \mathbf{E} é a matriz residual de \mathbf{X} (MAX et al., 2006; TRYGG; WOLD, 2002).

$$\mathbf{X} = \mathbf{T}_p \mathbf{P}_p^T + \mathbf{T}_o \mathbf{P}_o^T + \mathbf{E} \quad (6)$$

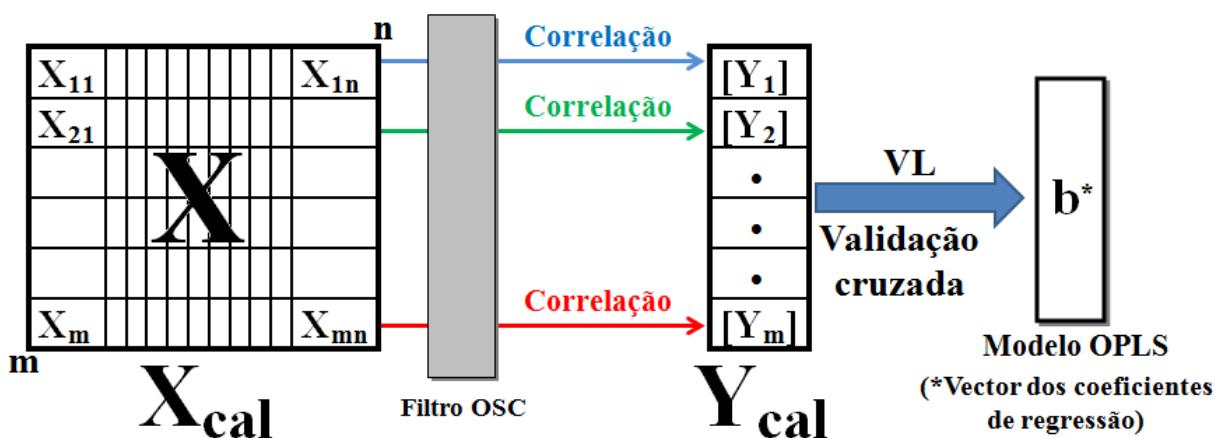
Para previsão, a variação ortogonal-Y é removida da matriz de dados \mathbf{X} , de acordo com a Equação 7:

$$\mathbf{X}_p = \mathbf{X} - \mathbf{T}_o \mathbf{P}_o^T \quad (7)$$

Em seguida, a matriz Y prevista é estimada usando o \mathbf{X}_p atualizado e os componentes preditivos do modelo OPLS, que são estimados a partir do conjunto de dados de calibração (MAX et al., 2006).

Na calibração do modelo OPLS (Figura 14) é escolhido um número de VL que proporcione o menor erro de previsão e que esteja de acordo com a norma ASTM E1655-05, ou seja, que a diferença entre os valores de referência e os valores previstos pelo modelo seja o menor possível. Assim, têm-se os vetores dos coeficientes de regressão.

Figura 14 – Construção do modelo de regressão por OPLS.



Fonte: O autor.

O número de VL é determinado através da aplicação de um método 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. Dentre vários, os mais usados para este tipo de calibração são: validação cruzada pelo método *leave-one-out*, em que uma amostra é deixada de fora no processo de construção do modelo e a seguir essa amostra é prevista pelo modelo construído. Na sequência, esta amostra retorna ao conjunto de calibração, e outra é retirada e prevista pelo modelo construído pelas restantes, e esse processo é repetido até que todas as amostras tenham sido retiradas do modelo uma vez (MASSART et al., 1998); o outro é a validação cruzada pelo critério da veneziana que é realizada em blocos de amostras, isto é, 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, seguindo o mesmo processo que o do método *leave-one-out*. Em ambos os casos, o processo possibilita a determinação do Erro Quadrático Médio de Validação Cruzada (RMSECV), de acordo com a Equação 8.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (8)$$

Onde y_i , \hat{y}_i e n são, os valores reais, valores previstos pelo modelo PLS e número de amostras de calibração, respectivamente.

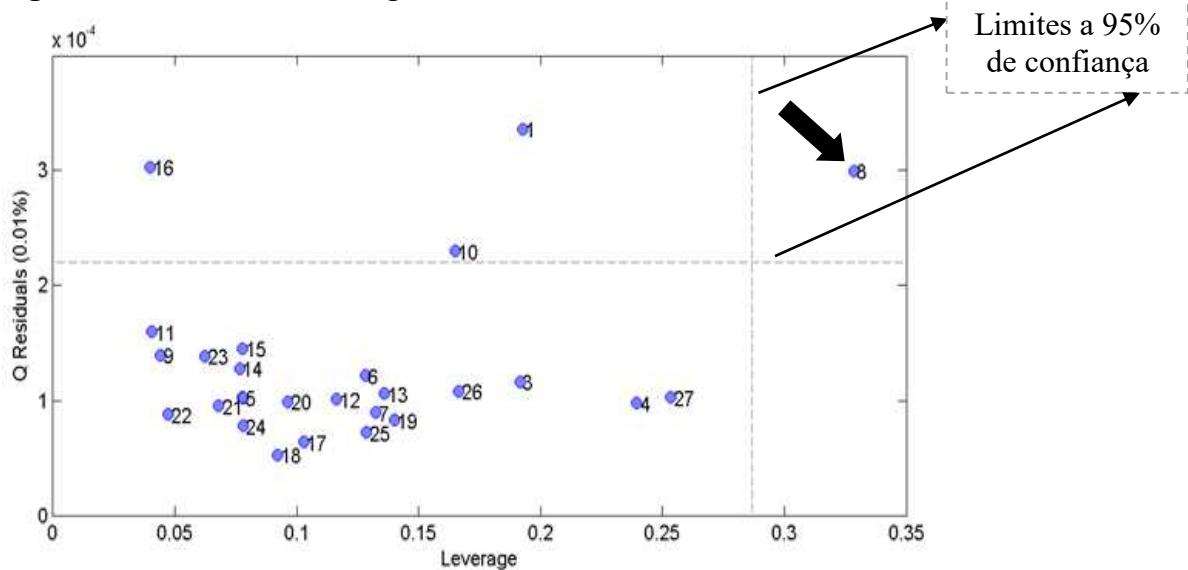
No entanto, para este trabalho foi usada a validação cruzada pelo critério da veneziana por ser o mais recomendado para um conjunto de dados acima de 20 amostras porque não apresentam resultados tendenciosos e demandam menos tempo de construção, diferentemente do método *leave-one-out*.

3.2.1.1. Detecção de Amostras Anômalas (Outlier)

Amostra anômala, também chamada de *outlier*, é aquela que apresenta um perfil diferente das restantes. Estas amostras afetam a qualidade global do modelo quimiométrico e ocorrem devido aos erros de digitação e experimentais, como por exemplo, na determinação da propriedade de interesse por um método de referência; e os erros instrumentais, quando o espectro da amostra em questão é de baixa qualidade (VALDERRAMA; BRAGA; POPPI, 2009). Portanto, para assegurar que as amostras formem um conjunto homogêneo que não afeta a qualidade global do modelo, é necessário retirar as amostras anômalas do conjunto de calibração.

A identificação das amostras anômalas é feita analisando o gráfico dos valores de leverage (o grau que uma amostra está distante da média do conjunto de dados) versus valores de resíduos (a parte não modelada nos dados espectrais), conforme mostra a Figura 15 (MASSART et al., 1998). Nesta figura, se observa que apenas a amostra nº 8 encontra-se fora dos limites a 95% de confiança (linhas tracejadas), isto é, apresenta um elevado valor de leverage e de resíduos. Portanto, ela deve ser retirada do conjunto de calibração, pois sua permanência implica no aumento de informação errônea, podendo diminuir a eficiência de previsão caso tal informação seja correlacionada com as amostras do conjunto de previsão. Caso seja necessária a informação sobre esta amostra, o analista deve refazer a sua medição até que esta fique dentro dos limites a 95% de confiança.

Figura 15 – Gráfico de leverage versus Q Residuals.



Fonte: O autor.

3.2.1.2. Validação dos modelos OPLS

Nesta etapa, é preparado outro conjunto de amostras na mesma faixa de concentração das amostras de calibração e faz-se uma correlação entre essa nova matriz \mathbf{X} e o vetor de regressão, prevendo-se então a concentração dessas amostras conforme é mostrado na Figura 16 (WOLD; SJÖSTRÖM; ERIKSSON, 2001).

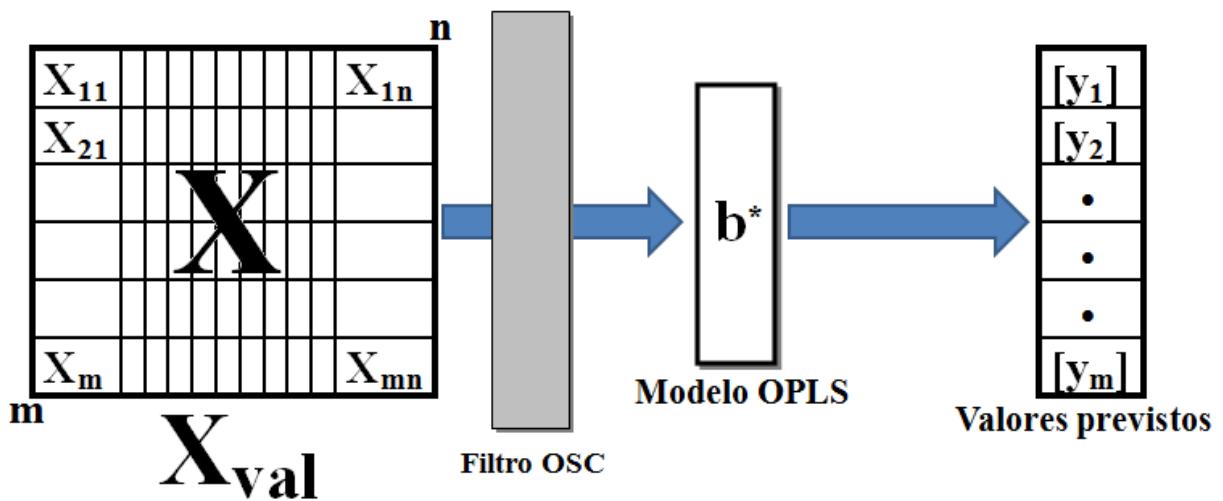
A validação dos modelos OPLS é realizada com base na análise sinal-ruído e na determinação das seguintes figuras de méritos: seletividade, sensibilidade, sensibilidade analítica, limite de detecção, limite de quantificação, teste para erros sistemáticos ($bias$ e t_{bias}),

de acordo com as equações apresentadas na Tabela 3 (DE CARVALHO ROCHA; NOGUEIRA; VAZ, 2012; FERREIRA; BRAGA; SENA, 2013; SILVA et al., 2012; VALDERRAMA; BRAGA; POPPI, 2009). A qualidade dos modelos OPLS é analisada com base no valor de R^2 e Q^2 , que indicam a correlação entre os valores observados e previstos para a resposta estudada e a correlação entre a resposta prevista observada e validada cruzada, calculados de acordo com as Equações 9 e 10, respectivamente (SOUIHI et al., 2015).

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{ical})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (9)$$

$$Q^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_{icv})^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (10)$$

Figura 16 – Processo de previsão de amostras de validação pelo modelo OPLS.



Fonte: O autor.

Tabela 3 – Equações usadas para calcular as figuras de mérito dos modelos OPLS*

Figura de mérito	Equação
Seletividade	$\hat{SEL}_i = \frac{n_{\text{âs}}}{\ \mathbf{x}_i\ }$
Sensibilidade	$\hat{SEN} = \frac{1}{\ \mathbf{b}_k\ }$
Sensibilidade analítica	$\gamma = \frac{\hat{SEN}}{\ \delta_x\ }$
Limite de detecção	$LD = 3.3\delta_x \frac{1}{\hat{SEN}}$
Limite de quantificação	$LQ = 10\delta_x \frac{1}{\hat{SEN}}$
bias	$bias = \frac{\sum_{i=1}^{n_{\text{val}}} (y_i - \hat{y}_i)}{n_{\text{val}}}$
SDV	$SDV = \sqrt{\frac{\sum[(y_i - \hat{y}_i) - bias]^2}{n_{\text{val}} - 1}}$
t_{bias}	$t_{\text{bias}} = \frac{ bias \sqrt{n_{\text{val}}}}{SDV}$

Fonte: O autor.

Onde, y_i e \hat{y}_i correspondem aos valores de referência e aos previstos pelo modelo, respectivamente; n_{val} é o número de amostras de validação; $n_{\text{âs}}_i$ é o valor escalar do sinal analítico líquido para a amostra i ; $\|\mathbf{x}_i\|$ é a norma de cada espectro; \mathbf{b}_k é o vetor dos coeficientes de regressão para a espécie de interesse k estimada pelo modelo de calibração multivariada; δ_x é o desvio padrão do sinal de referência.

A relação sinal-ruído foi analisada, tendo em conta que, todo espectro medido experimentalmente possa ser representado pela soma de dois vetores ortogonais entre si, onde um deles é vetor **NAS**, que é exclusivo do constituinte de interesse (sinal analítico), e outro é o vetor interferente, que contém as informações de todos os outros constituintes presentes (LORBER, 1986). No entanto, o sinal analítico líquido (NAS) (\mathbf{r}_k^*) foi calculado para analitos decompondo a matriz \mathbf{X} das variáveis preditoras do conjunto de calibração em duas outras: \mathbf{X}_k (contribuição exclusiva do constituinte de interesse) e \mathbf{X}_{-k} (contribuição dos interferentes), de acordo com a Equação 11 (FERREIRA, 2015; LORBER; FABER; KOWALSKI, 1997).

$$\mathbf{r}_k^* = \mathbf{P}_{\text{NAS},k} \mathbf{x}_i = [\mathbf{I} - \mathbf{X}_{-k} (\mathbf{X}_{-k})^+] \mathbf{x}_i \quad (11)$$

Em que, a matriz \mathbf{P}_{NAS} (de dimensões JxJ) faz a projeção dos espectros no espaço ortogonal, no qual se encontra o vetor NAS ; k identifica o analito de interesse; \mathbf{x}_i representa o espectro de uma dada amostra; \mathbf{I} é uma unidade de matriz adequadamente dimensionada; \mathbf{X}_{-k} é uma matriz dos sinais espectrais gerados por todos os outros analitos, com exceção de k ; $(\mathbf{X}_{-k})^+$ é a pseudo-inversa de \mathbf{X}_{-k} , normalmente calculado pela decomposição em valores singulares usando A fatores. A concentração de k na amostra desconhecida foi obtida a partir do espectro de RMN de ^1H de \mathbf{r} , usando a Equação 12 (DE CARVALHO ROCHA; NOGUEIRA; VAZ, 2012).

$$Y_{\text{un},k} = \frac{\mathbf{s}_k^T \mathbf{P}_{\text{NAS},k} \mathbf{r}}{\mathbf{s}_k^T \mathbf{P}_{\text{NAS},k} \mathbf{s}_k} = \frac{\mathbf{s}_k^T \mathbf{P}_{\text{NAS},k} \mathbf{P}_{\text{NAS},k} \mathbf{r}}{\mathbf{s}_k^T \mathbf{P}_{\text{NAS},k} \mathbf{P}_{\text{NAS},k} \mathbf{s}_k} = \frac{(\mathbf{s}_k^*)^T \mathbf{r}_k^*}{\|\mathbf{s}_k^*\|^2} \quad (12)$$

Seletividade

É a medida do grau de sobreposição entre o sinal da espécie de interesse e os interferentes presentes na amostra indicando, também, a parte do sinal que é perdida por essa sobreposição (DE SOUZA et al., 2015; VALDERRAMA; BRAGA; POPPI, 2009).

Sensibilidade

Corresponde à fração do sinal responsável pelo acréscimo de uma unidade de concentração à propriedade de interesse (DE CARVALHO ROCHA; NOGUEIRA; VAZ, 2012; SILVA et al., 2012).

Sensibilidade analítica

Apresenta a sensibilidade do método em termos da unidade de concentração que é utilizada, sendo definida como a razão entre a sensibilidade e o desvio padrão do sinal de referência (DE CARVALHO ROCHA; NOGUEIRA; VAZ, 2012; VALDERRAMA; BRAGA; POPPI, 2009).

Limite de detecção (LD)

Equivale à menor concentração da substância de interesse que pode ser detectada, mas não necessariamente quantificada (LORBER; FABER; KOWALSKI, 1997).

Limite de quantificação (LQ)

Representa a menor concentração da substância de interesse que pode ser medida com uma incerteza máxima de 10% (SILVA et al., 2012; VALDERRAMA; BRAGA; POPPI, 2009).

Teste para erro sistemático

Para este teste, primeiro é calculado o bias (a diferença entre a média da população e o valor verdadeiro) para o conjunto de validação (VALDERRAMA; BRAGA; POPPI, 2009). A seguir, é calculado o desvio padrão dos erros de validação (SDV) (FERREIRA; BRAGA; SENA, 2013). Por fim, é determinado o valor de t_{bias} (DE CARVALHO ROCHA; NOGUEIRA; VAZ, 2012). Caso o valor de t_{bias} resultar em maior valor 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 STANDARD E1655-05, 2012).

3.2.2. Métodos de classificação

Os métodos de classificação são chamados de métodos de reconhecimento de padrões por causa da determinação da propriedade de interesse ao atribuir uma amostra à sua respectiva classe. No entanto, são divididos em dois tipos: métodos não supervisionados e métodos supervisionados (FERREIRA, 2015).

Os métodos não supervisionados de conhecimento de padrões são usados basicamente para a análise exploratória dos dados. Os mais usados são: Análise de Componentes Principais (PCA) e Análise de Agrupamentos Hierárquicos (HCA) (FERREIRA, 2015).

Os métodos supervisionados de reconhecimento de padrões são usados quando a propriedade de interesse é categórica (discreta), isto é, cada uma das amostras é descrita por um conjunto de medidas experimentais, denominada de “padrão” e são classificadas de acordo com uma propriedade de interesse (bom/ruim; falso/verdadeiro; ativo/não ativo, etc.). Os mais utilizados na química são: A Modelagem Independente e Flexível por Analogia de Classe (SIMCA), K-ésimo Vizinhos mais Próximos (KNN) e Análise Discriminante por Quadrados Mínimos Parciais (PLS-DA) (FERREIRA, 2015). Estes métodos são aplicados a

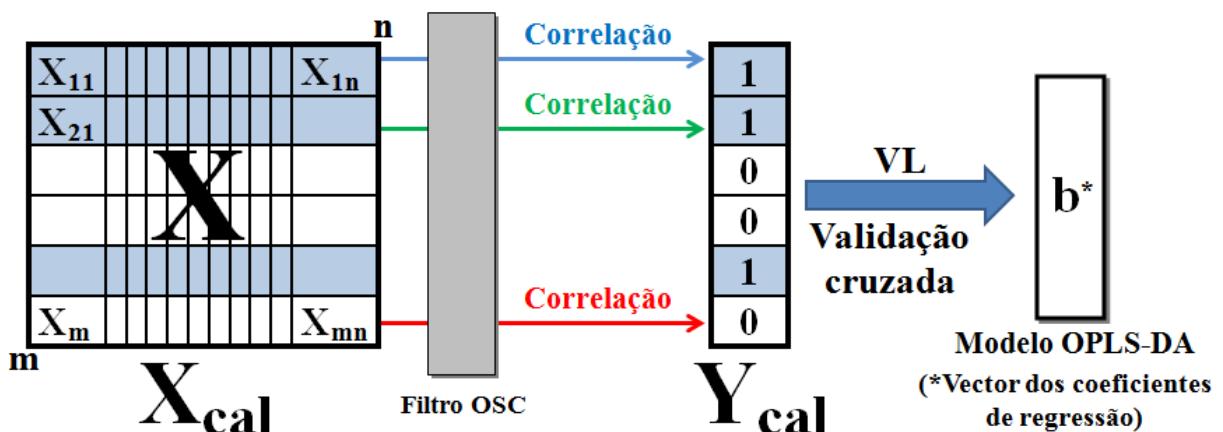
uma ampla variedade de dados químicos com diversas finalidades, tais como: identificação de perfis e impressões digitais, autenticação, detecção de falsificação, avaliação da qualidade de vários produtos alimentares, interpretação de dados, etc (MASSART et al., 1998).

3.2.2.1. Análise discriminante por projeções ortogonais à estrutura latente (OPLS-DA)

O método OPLS-DA é uma variação do algoritmo de regressão por (OPLS), onde as variáveis da matriz \mathbf{X} (dados espectrais) são relacionadas com a propriedade de interesse categórica (discreta) contidas em um vetor \mathbf{y} , como mostra a Figura 17 (WOLD; SJÖSTRÖM; ERIKSSON, 2001). O vetor \mathbf{y} é constituído por valores inteiros (0 e 1), onde 0 é usado para classe não de interesse e 1 para a classe de interesse (SERRANO-LOURIDO et al., 2012). Idealmente, os valores previstos pelo modelo OPLS-DA seriam 0 ou 1, mas na prática os valores previstos são próximos de 0 e 1. Tendo em conta que os valores previstos não são 0 ou 1, faz-se necessário o cálculo de um valor limite para separar as classes, denominado de “*threshold*” representado normalmente em gráfico por uma linha tracejada (WISE et al., 2006).

O valor de limiar “*threshold*” é previsto entre 0 e 1, com base no teorema de Bayes, segundo o qual, o limiar assume que os valores de \mathbf{y} previstos seguem uma distribuição semelhante ao que vai ser observado para as amostras futuras. Quando as duas distribuições estimadas se cruzam, é selecionado um limite, isto é, um valor de \mathbf{y} em que o número de falsos positivos e falsos negativos deve ser minimizado para previsões futuras (ALMEIDA et al., 2013). Quando a amostra apresentar um valor previsto acima do *threshold*, ela é considerada como pertencente à classe 1 e as amostras abaixo deste valor limite são pertencentes à classe 0.

Figura 17 – Representação da matriz de dados X e o vetor y usados no OPLS-DA.



Fonte: O autor.

As variáveis latentes dos modelos OPLS-DA são determinadas no processo de validação cruzada pelo critério da veneziana das amostras do conjunto de treinamento, como acontece para os modelos OPLS. Portanto, a escolha da quantidade de variáveis latentes na construção dos modelos OPLS-DA deve ser feita obedecendo a norma ASTM E1655-05, segundo o qual, o número máximo de variáveis latentes permitido para um conjunto de amostras de treinamento e teste são determinadas pelas Equações 13 e 14, respectivamente (ASTM STANDARD E1655-05, 2012). Além disso, estas variáveis latentes devem descrever a covariância nas amostras das matrizes e ter a correlação máxima com a classe de valores conhecidos, dando menos peso para a variância do ruído (BARKER; RAYENS, 2003). Este processo também possibilita a determinação do RMSECV.

$$K = \frac{n_{cal}-6}{6} \quad (13)$$

$$K = \frac{n_{prev}}{4} \quad (14)$$

Em que, K é o número de variável latente; n_{cal} é o número de amostras no conjunto de treinamento e n_{prev} é o número de amostras no conjunto de teste.

3.2.2.1.1. Análise da eficiência dos Modelos OPLS-DA

A eficiência dos modelos OPLS-DA é analisada com base nos critérios de estatísticas de respostas verdadeiras, como: taxa de falsos positivos e falsos negativos, sensibilidade, especificidade, eficiência e coeficiente de correlação de Matthew. A taxa de falsos positivos (**FP**) corresponde a probabilidade de uma amostra negativa ser classificada como positiva, calculada de acordo com a Equação 15 e a taxa de falsos negativos (**FN**) corresponde a probabilidade de uma amostra positiva ser classificada como negativa, calculada de acordo com a Equação 16 (ALMEIDA et al., 2013).

$$\mathbf{FP} = \frac{fp}{fp+vn} \times 100 \quad (15)$$

$$\mathbf{FN} = \frac{fn}{fn+vp} \times 100 \quad (16)$$

Em que, **fp** é o número de amostras falso positivas **vn** é o número de amostras verdadeira negativas, **fn** é o número de amostras falso negativas e **vp** é o número de amostras verdadeira positivas.

Os parâmetros sensibilidade e especificidade do modelo OPLS-DA são determinados com base nas Equações (17) e (18) para verificar a capacidade que o modelo tem em classificar corretamente as amostras verdadeiras positivas e verdadeiras negativas, respectivamente (XU et al., 2012). O coeficiente de correlação de Matthew, calculado de acordo com a Equação 19, pode apresentar valores entre -1 e +1, onde o valor +1 representa uma classificação perfeita, zero representa uma classificação aleatória e -1 uma classificação inversa. A eficiência do modelo OPLS-DA é calculado de acordo com a Equação 20.

$$\mathbf{Sensibilidade} = \frac{vp}{vp+fn} \quad (17)$$

$$\mathbf{Especificidade} = \frac{vn}{vn+fp} \quad (18)$$

$$\mathbf{Coeficiente\ de\ Matthew} = \frac{(vp \times vn) - (fp \times fn)}{\sqrt{(vp+fp) \times (vp+fn) \times (vn+fp) \times (vn+fn)}} \quad (19)$$

$$\mathbf{Eficiência} = \frac{vp+vn}{vp+vn+fp+fn} \quad (20)$$

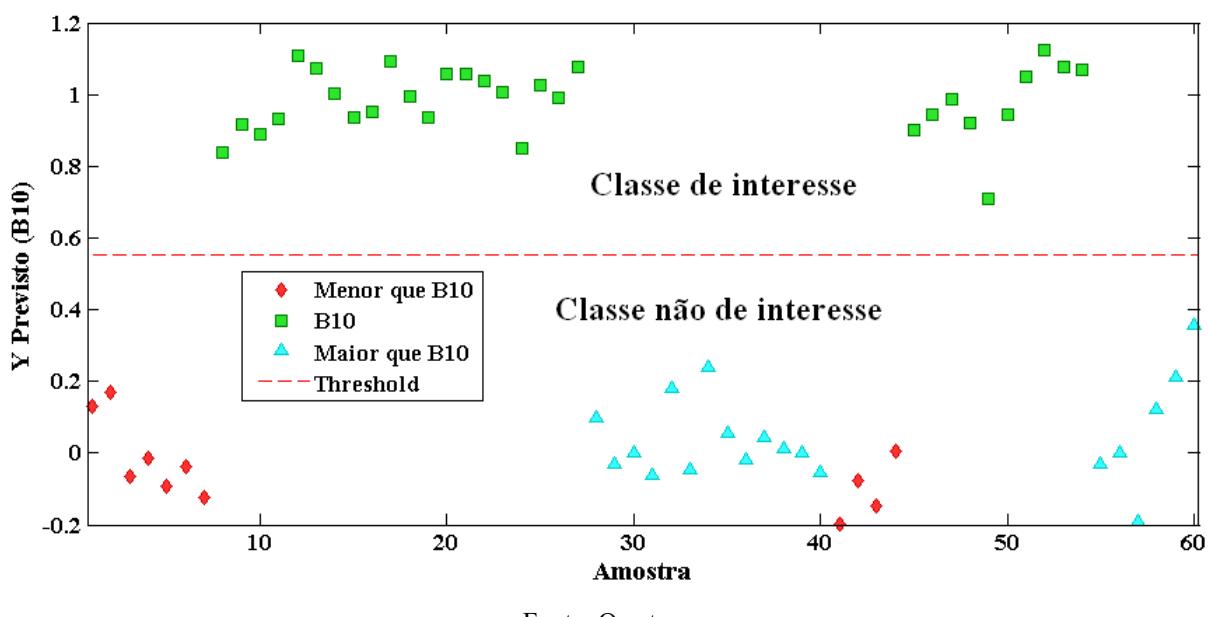
Finalmente, o ajuste excessivo do modelo OPLS-DA é analisado com base nos resultados dos testes de permutação.

3.2.2.1.2. Representação gráfica do resultado do modelo OPLS-DA

Geralmente, a classificação das amostras dos conjuntos de treinamento e de teste pode ser observada em dois gráficos diferentes:

- O gráfico de estimativas de classes, onde as amostras posicionadas acima da linha tracejada (*threshold*) são classificadas como pertencentes à classe 1 (classe de interesse) e, as amostras posicionadas abaixo do valor de *threshold* são classificadas como pertencentes à classe 0, ou seja, a classe que não é de interesse. A Figura 18 mostra um exemplo do gráfico de estimativas para um modelo OPLS-DA, em que nele pode se observar que as amostras do conjunto de treinamento e de teste (■) que pertencem à classe de interesse estão acima do *threshold*, separando-se das amostras do conjunto de treinamento e de teste (◆ e ▲) da classe não de interesse que estão abaixo do *threshold*;

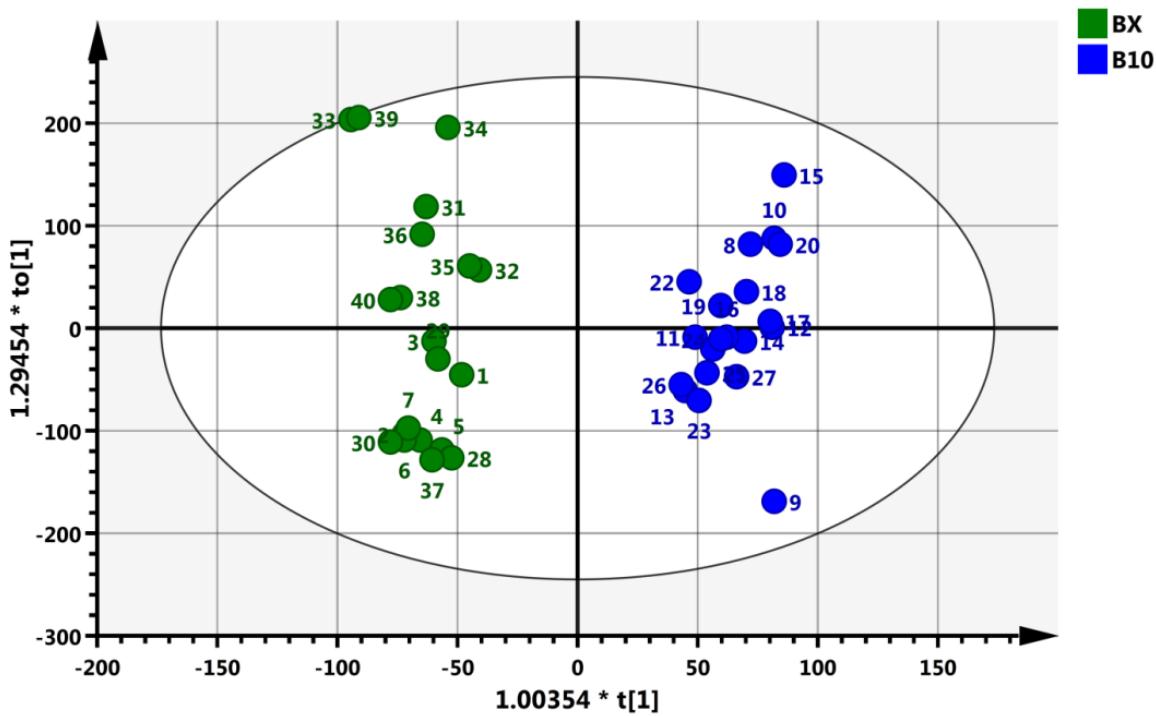
Figura 18 – Representação gráfica de estimativa de classes para o modelo OPLS-DA.



Fonte: O autor.

- O gráfico de dispersão dos escores $t[1]$ versus $t_0[1]$. Este gráfico é importante na análise de um problema de separação de duas classes (por exemplo, grupo de controle versus grupo de tratamento ou tipo de produto A versus tipo de produto B, etc.) porque proporcionam sempre um componente preditivo $t[1]$ e nenhum, um ou mais componentes ortogonais $t_0[1]$. A Figura 19 mostra um exemplo deste tipo de gráfico, em que nele pode se observar separação das amostras de BX e B10.

Figura 19 – Representação gráfica de dispersão dos escores para o modelo OPLS-DA.



Fonte: O autor.

4.1. Objetivo Geral

- Desenvolver métodos analíticos para determinar o teor de biodiesel metílico de mafurra e crambe em mistura com diesel usando a Espectroscopia de Ressonância Magnética Nuclear de Hidrogênio aliada aos métodos quimiométricos OPLS e OPLS-DA.

4.1.1. *Objetivos Específicos*

- Construir modelos OPLS para prever o teor de biodiesel metílico de mafurra e crambe em mistura com diesel;
- Construir modelos OPLS-DA para classificar teor de biodiesel metílico de mafurra e crambe em mistura com diesel;
- Validar os modelos OPLS construídos, de acordo com as normas ASTM E1655-05 e ABNT NBR 15568;
- Validar os modelos OPLS-DA construídos, através dos valores de sensibilidade e especificidade.

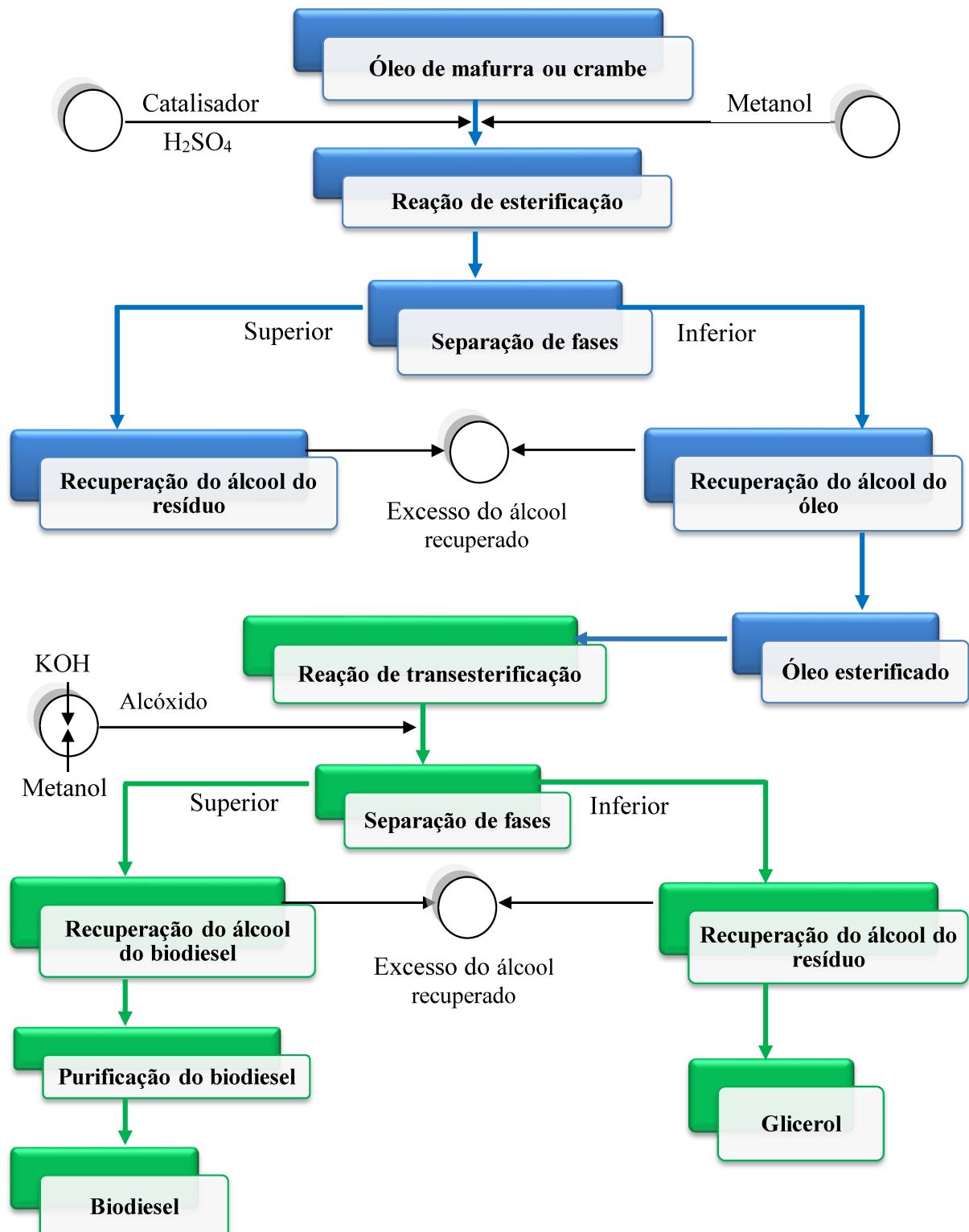
5.1. Produção do Biodiesel

Os óleos brutos de mafurra e crambe foram obtidos em Moçambique e no Brasil, respectivamente. Estes óleos foram usados para a produção do biodiesel metílico de mafurra (BMMa) e crambe (BMC), respectivamente, no Laboratório de Biocombustíveis do Instituto de Química, da Universidade Federal de Uberlândia, seguindo duas etapas.

Na primeira etapa, se realizou o processo de esterificação com o objetivo de reduzir o alto índice de acidez e aumentar a eficiência do processo de conversão dos ácidos graxos livres em ésteres (DONG et al., 2013; RAMOS et al., 2015). Para tal, se adicionou 3,0 g de ácido sulfúrico concentrado à 100,0 g de álcool metílico e, em seguida se adicionou 100,0 g do óleo (mafurra ou crambe). A mistura foi aquecida a uma temperatura de 60 °C sob agitação magnética a 400 rpm e sob refluxo durante 3 horas. Em seguida, colocou-se a mistura num funil separador por 1 hora de repouso para decantação das fases: inferior (óleo esterificado) e superior (misturas de álcool e ácido sulfúrico). Por fim, o óleo separado foi transferido para um balão de fundo redondo para a destilação a vácuo usando o evaporador rotatório a 90 rpm e 80 °C durante 1 hora para remover o álcool em excesso (MÁQUINA et al., 2017; SITOE et al., 2017).

Na segunda etapa, se realizou o processo de transesterificação do óleo esterificado, usando a proporção de 100,0 g de cada óleo e 30,0 g de metanol na presença de 1,0 g de KOH como catalisador. Para tal, se adicionou em primeiro lugar, o álcool metílico no hidróxido de potássio para formar o alcóxido. Após a formação do alcóxido, se adicionou o óleo para dar início a reação de transesterificação (MÁQUINA et al., 2017). Esta reação foi mantida sob agitação magnética a 400 rpm e à temperatura ambiente durante 1 h. Após este tempo, colocou-se a solução resultante num funil de separação por 24 h de repouso para se separar as fases: inferior (glicerol) e superior (biodiesel). Após a separação, o biodiesel obtido foi purificado por lavagem com água destilada a 90 °C e, em seguida, seco utilizando um evaporador rotativo durante 1 h, à 88 rpm e 90 °C (MÁQUINA et al., 2017; SITOE et al., 2017). A Figura 20 apresenta um esquema que resume as etapas de produção de biodiesel metílico de mafurra e crambe.

Figura 20 – Etapas de produção de biodiesel metílico de mafurra e crambe.



Fonte: O autor.

5.2. Preparação das Amostras

O diesel puro usado no preparo das amostras foi fornecido pela empresa brasileira Transpetro S/A. As amostras foram pesadas à temperatura ambiente ($25 \pm 2^{\circ}\text{C}$) utilizando uma balança analítica (Sartorius, modelo BP211D).

Para a construção dos modelos OPLS, preparou-se 70 amostras para cada tipo de biodiesel (BMMa e BMC), por adição do biodiesel B100 ao diesel puro, em um intervalo de concentração de 1,00 a 30,00% (v/v) usando um incremento de 0,30%. Para cada modelo OPLS, foram utilizadas 46 amostras no conjunto de calibração e 24 amostras no conjunto de previsão.

Para a construção dos modelos OPLS-DA, Preparou-se dois lotes de amostras para cada tipo de biodiesel (BMMa, BMC) pela adição do biodiesel B100 ao diesel puro, sendo:

- ✓ Um lote de 30 amostras de B10 usando a proporção de 10% de biodiesel para 90% de diesel puro;
- ✓ Um lote de 30 amostras de diferentes percentuais BX na faixa de concentração de 2 a 23% (v/v), sendo: 11 amostras com concentrações menores que B10 e 19 amostras acima de B10.

No entanto, o conjunto de treinamento de cada modelo OPLS-DA foi constituído por 20 amostras de B10, 7 amostras com concentração menor que B10 e 13 amostras com concentração maior que B10, totalizando 40 amostras neste conjunto. E o restante das amostras foi usado para construir o conjunto de teste, isto é, 10 amostras de B10, 4 amostras com concentração menor que B10 e 6 amostras com concentração maior que B10, totalizando 20 amostras neste conjunto.

5.3. Obtenção dos espectros de RMN de ^1H

Todos os espectros de RMN de ^1H foram obtidos num espetrômetro de RMN da marca Bruker, modelo Avance III 400 MHz. Para obter o espetro, 100 μL de cada amostra foi dissolvido em 400 μL de CDCl_3 (clorofórmio deuterado), totalizando um volume de 500 μL de solução. A Tabela 4 apresenta as condições experimentais utilizadas na obtenção dos espetros.

Tabela 4 – Condições experimentais

Parâmetros	Condições experimentais
Programa de pulso	zg30
Largura espectral	<ul style="list-style-type: none"> • Esquerda 16,19 ppm; • Direita -3,84 ppm
Tamanho espectral	65536 pontos
Pulso 90°	11,75 µs
Tempo de espera	1s
Varreduras	16

Fonte: O autor.

5.4. Construção dos modelos OPLS e OPLS-DA

Este trabalho foi desenvolvido usando duas áreas: calibração multivariada por OPLS, 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 para posteriormente tornar possível a quantificação dessa(s) propriedade(s) em novas amostras desconhecidas; a outra área foi calibração multivariada por OPLS-DA, um método de reconhecimento de padrões e classificação de dados, em que a propriedade de interesse é categórica (discreta), isto é, cada uma das amostras é descrita por um conjunto de medidas experimentais, denominado de “padrão” e são classificadas de acordo com uma propriedade de interesse (bom/ruim; falso/verdadeiro; ativo/não ativo, etc.) (FERREIRA, 2015).

O Software MATLAB, versão R2018b (Mathworks, Inc.); o PLS_Toolbox, versão 8.8 (Eigenvector Research) e o software SIMCA, versão 14.1, foram utilizados para pré-processar os dados e desenvolver os modelos OPLS e OPLS-DA de BMa e BMC.

Para realizar os procedimentos multivariados na construção dos modelos OPLS, os dados foram organizados em uma matriz ordenada de linhas e colunas, constituindo uma matriz **X**, onde cada linha corresponde a uma amostra "**m**" e cada coluna corresponde a uma variável "**n**", na qual **m** = 1,2,3, ..., 70 e **n** = 1,2,3, ..., 65536. Uma matriz **Y** foi criada contendo valores de concentração (de 1 a 30,00% (v/v)) das amostras. Na construção dos modelos OPLS-DA, a matriz **X** foi constituído por **m** = 1,2,3, ..., 60 e **n** = 1,2,3, ..., 65536 e,

um vetor y foi criado contendo valores inteiros discretos 0 e 1 para indicar a classe não de interesse e de interesse, respectivamente.

Nos dois modelos, os dados foram alinhados usando uma rotina desenvolvida pelo autor, no Laboratório de Quimiometria do Triângulo, aplicando a função Interval-correlation-shift (icoshift) (SAVORANI; TOMASI; ENGELSEN, 2010). Em seguida, foram centrados na média e pré-processados usando o filtro OSC, com o objetivo de minimizar as variações sistemáticas indesejáveis para produzir modelos menos complexos e com melhor poder preditivo (PERS et al., 2008; WESTERHUIS; DE JONG; SMILDE, 2001). A validação cruzada nos dois modelos foi realizada pelo critério da veneziana, com 14 blocos de divisão de dados e com duas amostras por bloco.

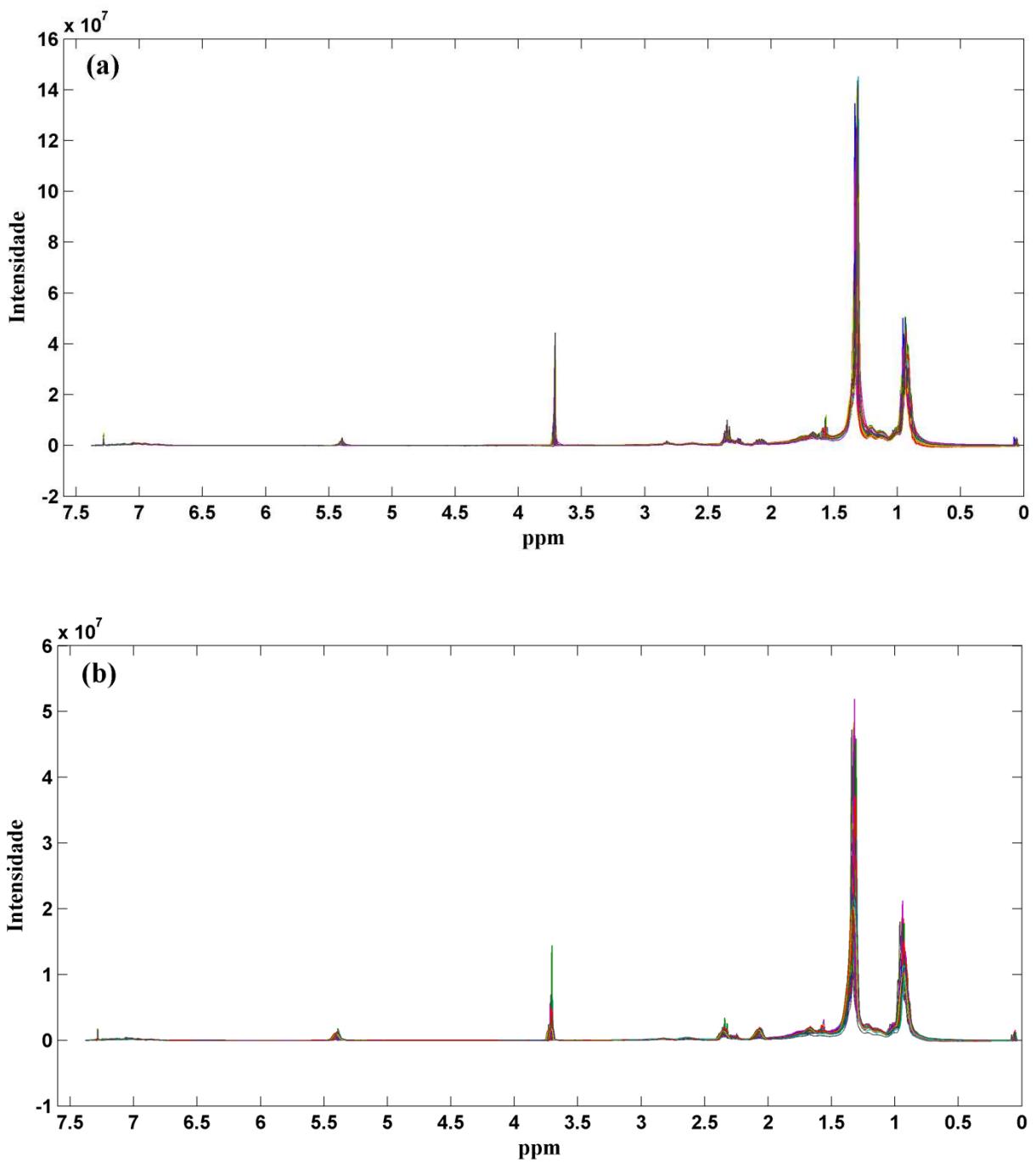
6.1. Caracterização dos Espectros de RMN de ^1H

A Figura 21 mostra espectros de RMN de ^1H de 70 amostras de mistura diesel/biodiesel de a) mafurra e b) crambe, na faixa de concentração de 1 à 30% (v/v) antes de serem alinhados, onde pode se notar pequenos desvios nos deslocamentos químicos causados pelo ambiente ao qual os prótons são submetidos. Esses pequenos desvios foram corrigidos, alinhando os espectros com o algoritmo *Icoshift* para se desenvolver modelos que não exijam maior número de variáveis latentes para explicar a mesma variação causada pelo desvio de uma banda espectral.

A Figura 22 mostra os mesmos espectros já corrigidos, onde pode se notar deslocamentos químicos bem definidos, como: uma banda na região de 0 ppm atribuída à tetrametilsilano [TMS; $(\text{CH}_3)_4\text{Si}$] usado como referência nas medições de RMN de ^1H para definir a posição de uma absorção (MCMURRY, 2016); um conjunto de bandas na região de 0,6 à 3,0 ppm atribuídas à hidrogênios alifáticos, precisamente, um triplete perto de 0,8 ppm relacionado à hidrogênios de grupo metila terminal, um sinal forte em 1,3 ppm de cadeia de carbono dos metilenos, um multiplete em 1,6 ppm de prótons metilênicos β -carbonílicos, um sinal associado à insaturação em 2,1 ppm atribuída a alílico, um triplete em 2,3 ppm relacionados com prótons metilênicos α -carbonílicos, sinal associado à insaturação em 2,8 ppm atribuída à bis-alílicos (MCMURRY, 2016; MONTEIRO et al., 2009); bandas características do biodiesel com baixa absorção em 3,7 ppm que surge a partir da fração de éster metílico, atribuído à prótons metoxílicos e em 5,4 ppm que é atribuído à prótons olefínicos; as bandas características do diesel na região de 6,6 à 7,3 ppm atribuídas à gama de hidrogênios aromáticos (PAVIA et al., 2010; PORTELA et al., 2016).

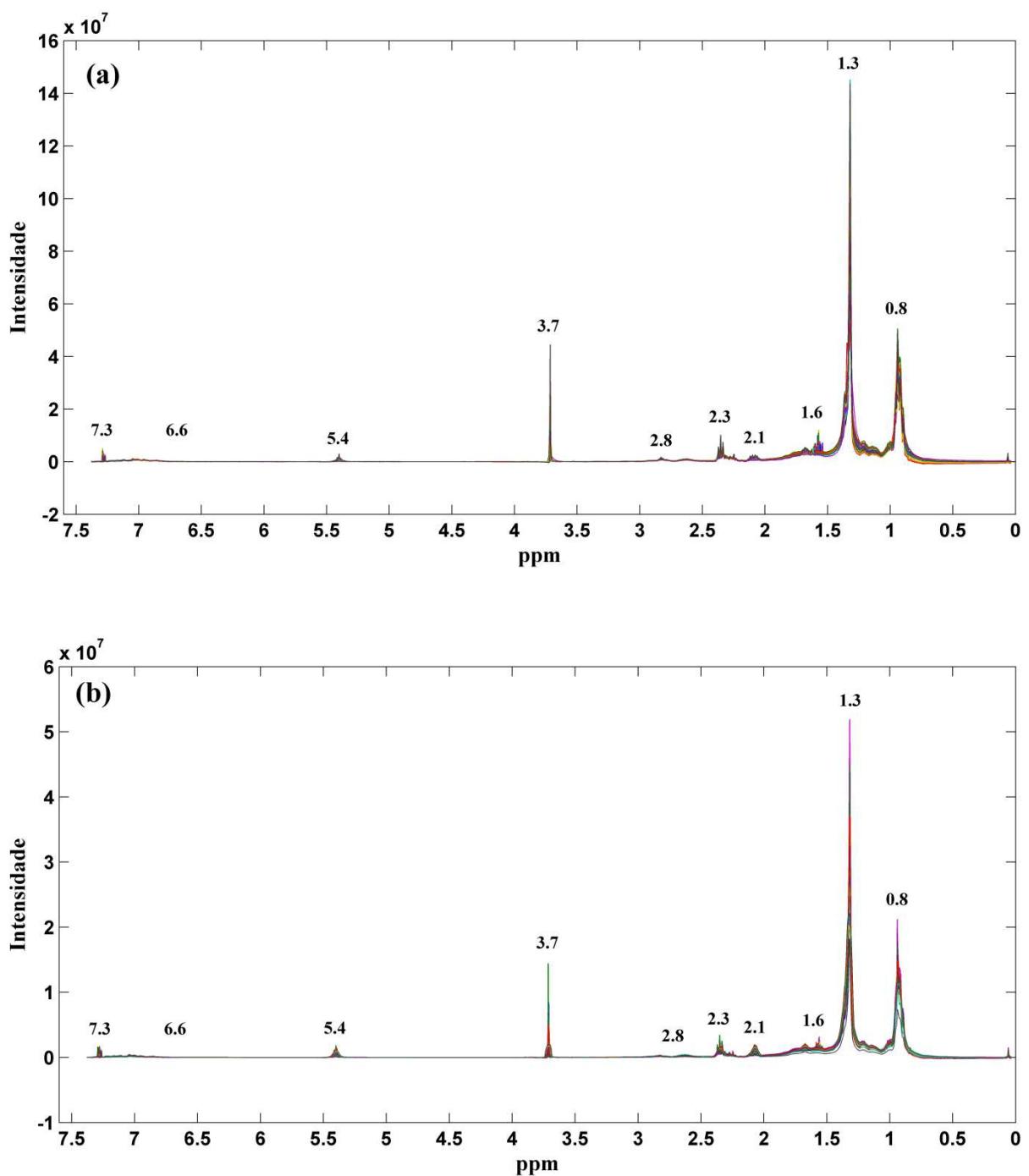
No entanto, as semelhanças observadas nos espectros de RMN de ^1H das 70 amostras dificultam o monitoramento visual do teor do biodiesel metílico de mafurra e de crambe em mistura com diesel, ou seja, é difícil atribuir um valor de concentração a cada perfil espectral. Como alternativa ao monitoramento visual, os métodos quimiométricos são eficientes para monitorar o teor de biodiesel em mistura com diesel quando a intensidade das suas bandas varia com o aumento ou diminuição da quantidade do biodiesel na mistura.

Figura 21 – Espectros de RMN de ^1H não corrigidos de 70 amostras de misturas diesel/biodiesel de (a) mafurra e (b) crambe.



Fonte: O autor.

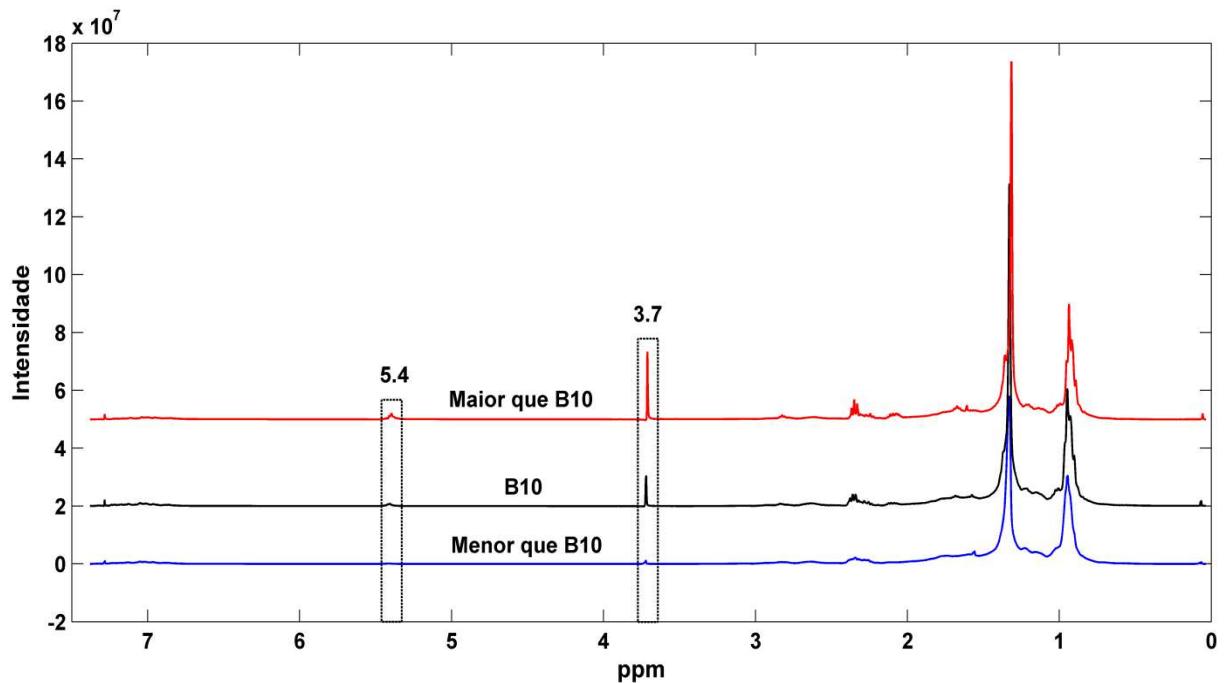
Figura 22 – Espectros de RMN de ^1H corrigidos de 70 amostras de misturas diesel/biodiesel de (a) mafurra e (b) crambe.



Fonte: O autor.

Neste trabalho, a intensidade das bandas aumentou com o aumento da quantidade do biodiesel na mistura, como se pode ver na Figura 23, as bandas características do biodiesel que aparecem em 3,7 e 5,4 ppm nos espectros de BX (menor que B10), B10 e BX (maior que B10), apresentam intensidades baixa, média e alta respectivamente, devido ao aumento dos prótons que absorvem nestes deslocamentos químicos.

Figura 23 – Espectros de RMN de ^1H de B10 e BX.

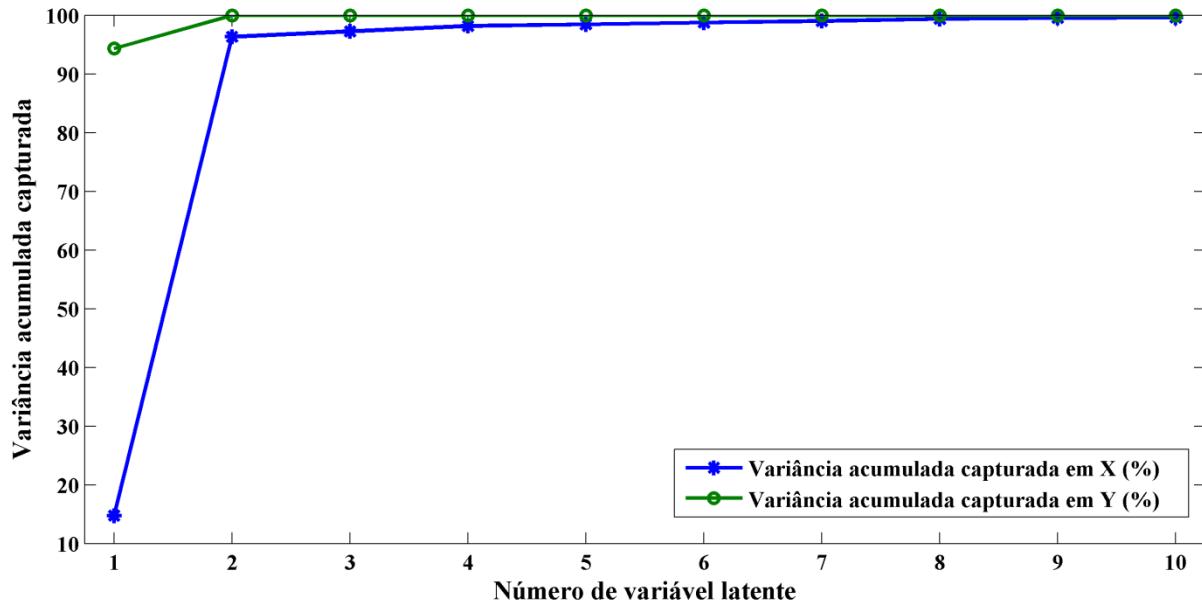


Fonte: O autor.

6.2. Modelos OPLS

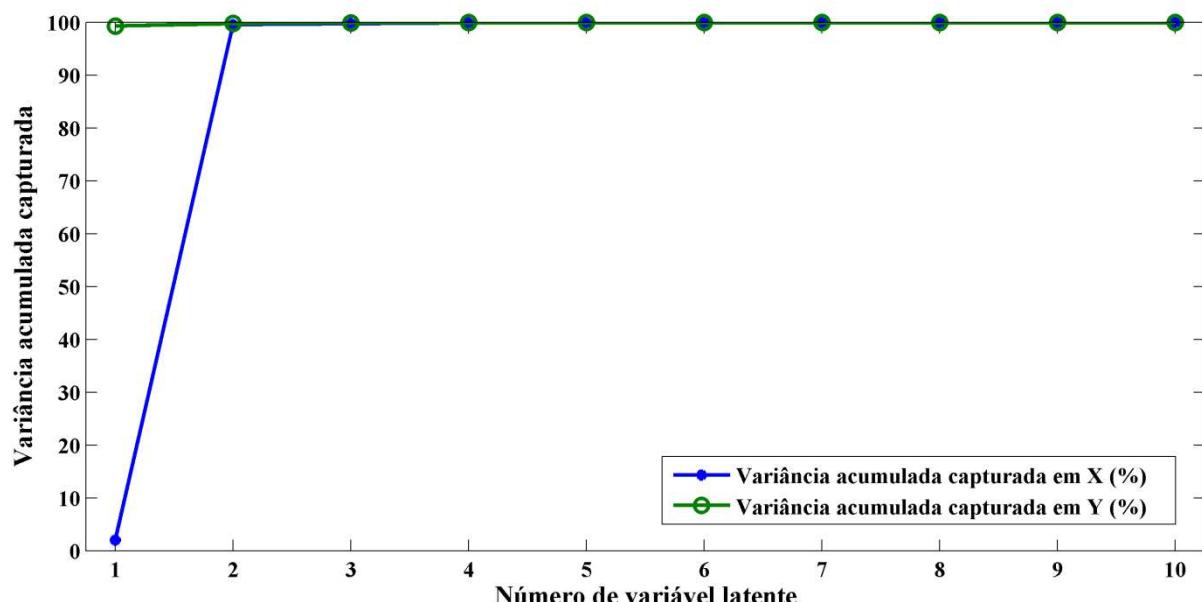
As Figuras 24 e 25 mostram a variância acumulada capturada nas matrizes **X** e **Y** obtidos em função da quantidade da variável latente escolhida para construção dos modelos OPLS de mafurra e crambe, respectivamente. No caso do modelo OPLS de mafurra, as 2 variáveis latentes escolhidas proporcionaram menor erro e maior variância capturada, representando 96,33% e 99,97% da variância explicada nas matrizes **X** e **Y**, respectivamente. Para o modelo OPLS de crambe, as 3 variáveis latentes escolhidas também proporcionaram menor erro e maior variância capturada, representando 99,78% e 99,89% da variância explicada nas matrizes **X** e **Y**, respectivamente.

Figura 24 – Variância acumulada capturada nas matrizes X e Y versus variável latente do modelo OPLS de mafurra.



Fonte: O autor.

Figura 25 – Variância acumulada capturada nas matrizes X e Y versus variável latente do modelo OPLS de crambe.



Fonte: O autor.

Os resultados apresentados na Tabela 5 mostram que a quantidade de amostras usadas para constituir os conjuntos de calibração e previsão dos modelos OPLS de BMMa e BMC estão de acordo com as exigências da norma ASTM E1655-55, segundo o qual 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 dos modelos OPLS.

Os valores de Erro Quadrático Médio de Validação Cruzada (RMSECV), Erro Quadrático Médio de Calibração (RMSEC) e Erro Quadrático Médio de Previsão (RMSEP) dos dois modelos OPLS são relativamente baixos e considerados aceitáveis de acordo com a norma ABNT NBR 15568, a qual estabelece que o RMSEP não deve ser superior a 1%. Os dois modelos não estão sobreajustados porque os valores de RMSEC são inferiores que os valores de RMSEP, indicando uma boa concordância entre os dois parâmetros (LIU; HE; WANG, 2008). Estes resultados estão próximos dos valores obtidos em alguns modelos de quantificação publicados na literatura (SHI et al., 2019; SINGH; KUMAR; BLÜMICH, 2019).

Tabela 5 – Parâmetros e valores de exatidão dos modelos OPLS

MODELO OPLS	BMMa	BMC
Nº amostras de calibração	46	46
*Mínimo de amostras de calibração	18	24
Nº amostras de previsão	24	24
*Mínimo de amostras de previsão	8	12
VLs	2	3
RMSECV	1,04	0,56
RMSEC	0,10	0,28
RMSEP	0,78	0,66

*Conforme norma ASTM E1655-05.

Fonte: O autor.

A Tabela 6 apresenta resultados das figuras de mérito, calculadas para os modelos OPLS de mafurra e de crambe, na qual se observa que os valores do t_{bias} foram menores que os de $t_{crítico}$ nos dois modelos para graus de liberdade (GL) igual ao número de amostras de previsão, o que equivale dizer que os modelos não apresentaram erros sistemáticos.

O valor do LQ obtido nos modelos OPLS de mafurra e de crambe é superior ao LD, demonstrando uma boa concordância entre os parâmetros porque a concentração mais baixa da substância de interesse que é medida com uma incerteza máxima de 10% está dentro do que pode ser detectado.

A sensibilidade expressa um aumento na fração do sinal quando a concentração do analito de interesse variar em um valor unitário da grandeza medida, neste caso, o seu valor foi estimado em $5,73 \times 10^7$ e $3,66 \times 10^7$ para os modelos OPLS de mafurra e de crambe, respectivamente.

A sensibilidade analítica do modelo OPLS de mafurra foi maior que a do modelo OPLS de crambe, mostrando que a influência dos resíduos na previsão de amostras desconhecidas é menor no modelo OPLS de mafurra. O inverso da sensibilidade analítica mostra que os dois modelos OPLS são capazes de distinguir diferenças entre amostras com uma variação de $4,03 \times 10^{-4}$ e $6,33 \times 10^{-4}$, respectivamente.

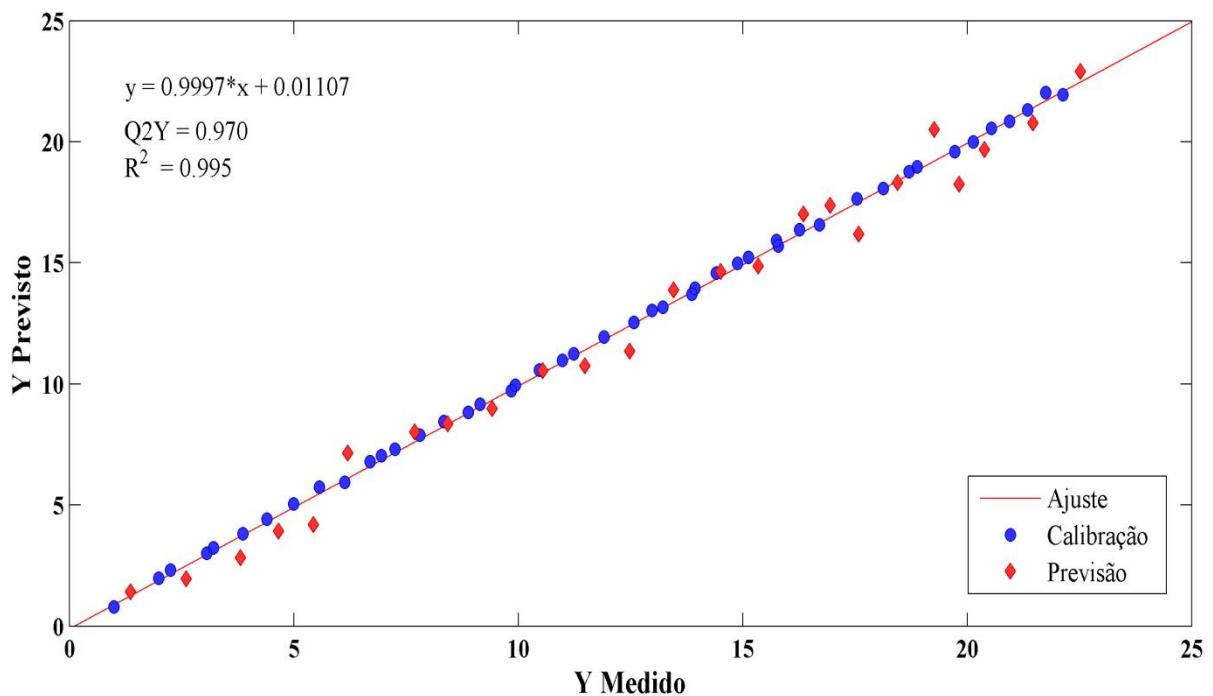
Tabela 6 – Resultados das figuras de mérito dos modelos OPLS de mafurra e de crambe

Figuras de mérito	Valores do modelo OPLS	
	BMMa	BMC
Seletividade	0,6956	0,7217
Sensibilidade / % (v/v) ⁻¹	$5,73 \times 10^7$	$3,66 \times 10^7$
Sensibilidade analítica / % (v/v)	$2,47 \times 10^3$	$1,58 \times 10^3$
Inverso da sensibilidade analítica / (v/v) ⁻¹	$4,03 \times 10^{-4}$	$6,33 \times 10^{-4}$
Limite de detecção / % (v/v)	0,0013	0,0021
Limite de quantificação / % (v/v)	0,0040	0,0063
t_{bias}	0,3241	0,3512
$t_{critico}$	2,0639	2,0639
R ²	0,995	0,997
Q ²	0,970	0,996

Fonte: O autor.

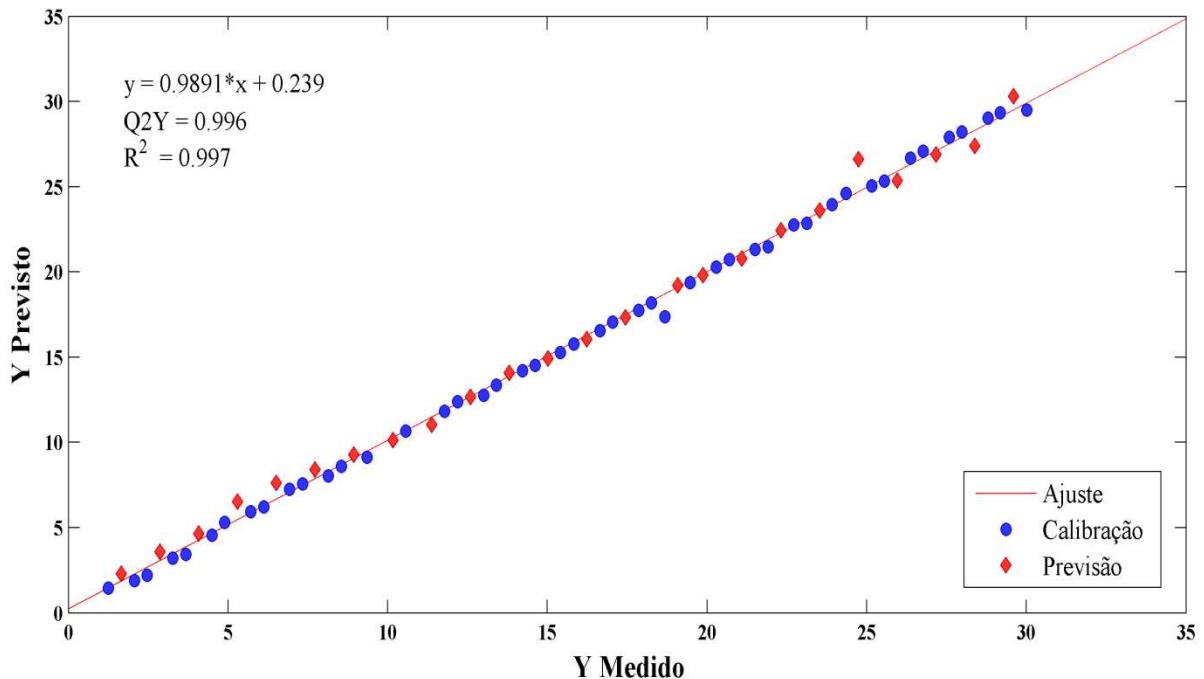
A avaliação do ajuste dos modelos OPLS de mafurra e de crambe por meio da correlação dos valores medidos e previstos dos conjuntos de calibração e previsão é mostrada nas Figuras 26 e 27, em que uma alta correlação entre os dois valores foi evidente nos dois modelos, com o coeficiente de correlação (R^2) superior a 0,99. O valor de Q^2 nos dois modelos foi maior que 0,5, o que equivale dizer que os modelos OPLS de mafurra e de crambe são robustos e apresentam boa capacidade preditiva.

Figura 26 – Correlação dos valores medidos e previstos dos conjuntos de calibração e previsão do modelo OPLS de mafurra.



Fonte: O autor.

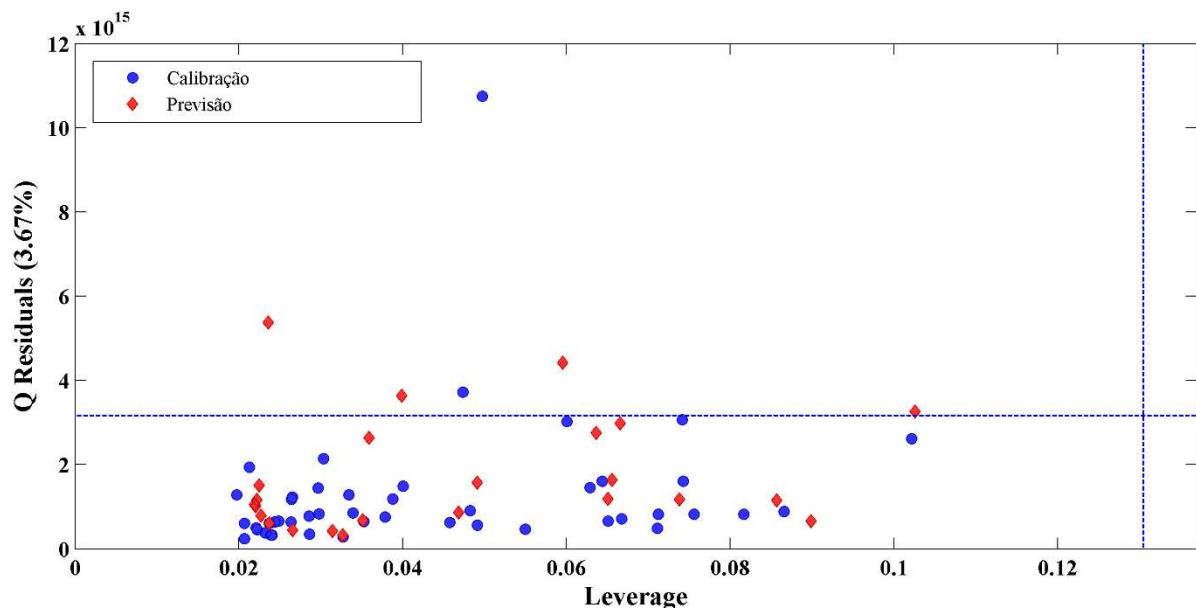
Figura 27 – Correlação dos valores medidos e previstos dos conjuntos de calibração e previsão do modelo OPLS de crambe.



Fonte: O autor.

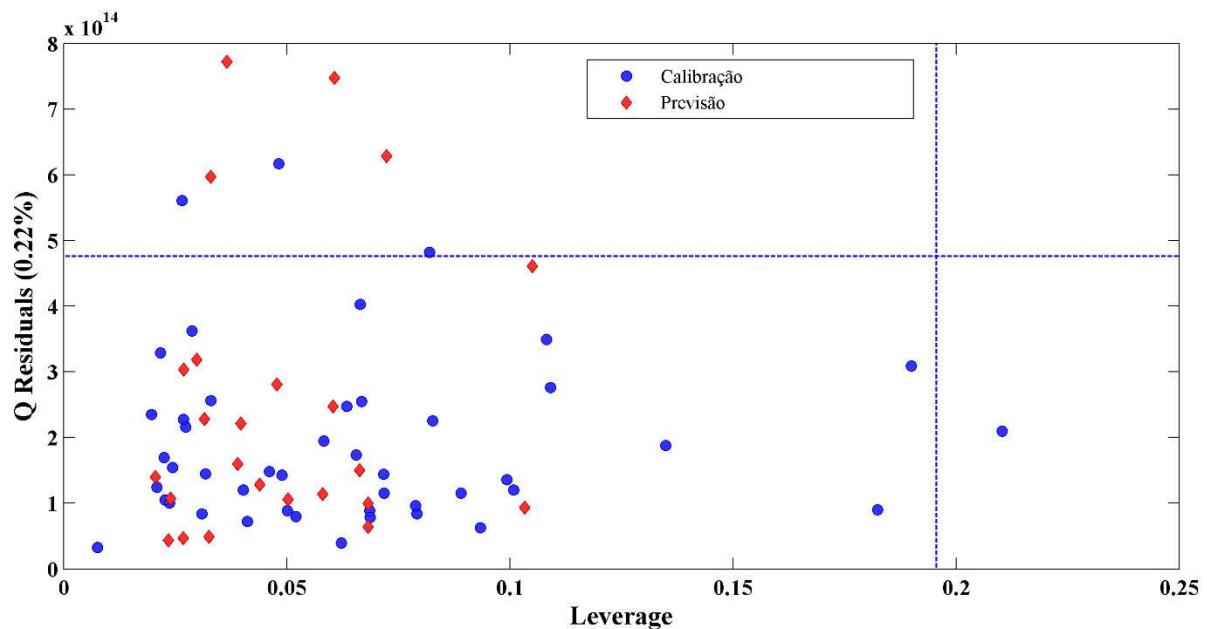
As Figuras 28 e 29 apresentam os valores de *leverage* (o grau que uma amostra está distante da média do conjunto de dados) *versus* valores de resíduos (a parte não modelada nos dados espectrais) que foram analisados para identificar amostras anômalas nos modelos OPLS de mafurra e de crambe, respectivamente. Nestas figuras, se observa que existem algumas amostras que apresentaram altos valores de resíduos, isto é, estão fora do limite de resíduos. Estas amostras não são consideradas de anômalas porque apresentam valores de *leverage* dentro do seu limite. As amostras que apresentam altos valores de *leverage* também não são consideradas de anômalas porque apresentam valores de resíduos baixos.

Figura 28 – Valores de leverage versus valores Q residuals com 95% de confiança para o modelo OPLS de mafurra.



Fonte: O autor.

Figura 29 – Valores de leverage versus valores Q residuals com 95% de confiança para o modelo OPLS de crambe.



Fonte: O autor.

A Tabela 7 apresenta os resultados dos testes de permutação realizados nos modelos OPLS de mafurra e de crambe para verificar o ajuste excessivo dos modelos. Estes resultados mostram que os dois modelos não estão super ajustados porque os valores de Wilcoxon (teste de classificação assinado em pares Wilcoxon), Sign Test (teste de classificação assinada em pares) e Rand t-test (Teste-t de randomização) são menores que 0,05.

Tabela 7 – Probabilidade de insignificância dos modelos OPLS de mafurra e de crambe com 2 e 3 VL, respectivamente

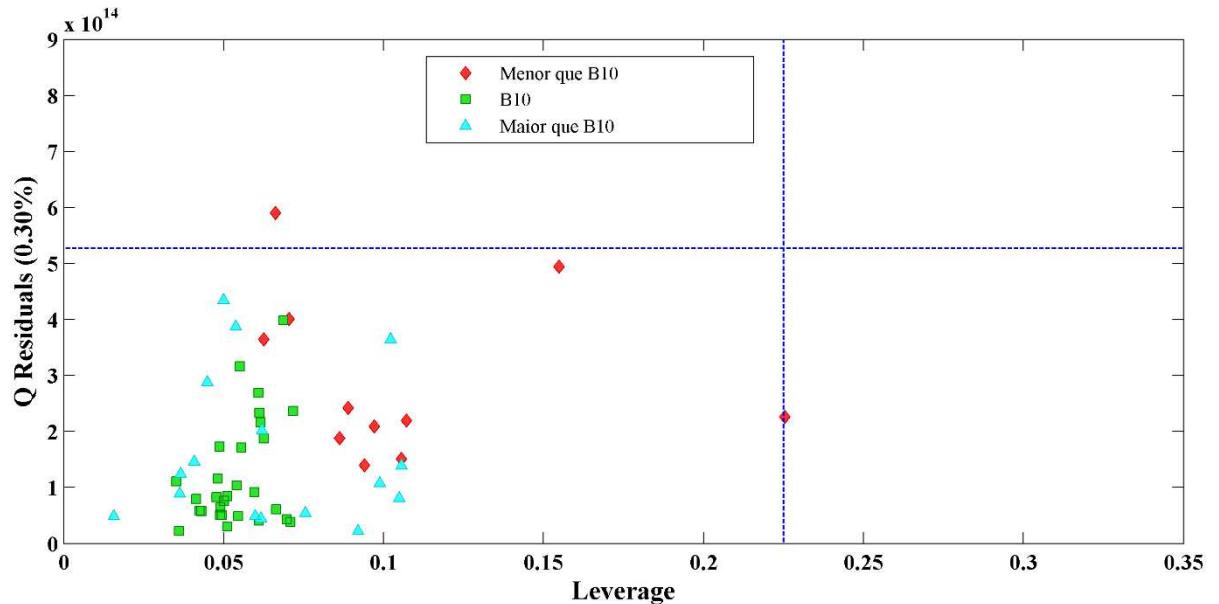
	Wilcoxon		Sign Test		Rand t-test	
	BMMA	BMC	BMMA	BMC	BMMA	BMC
Coluna Y: 1						
Auto-Previsão (RMSEC)	0,000	0,000	0,000	0,000	0,005	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,000	0,000	0,005	0,005
Coluna Y: 2						
Auto-Previsão (RMSEC)	0,000	0,000	0,000	0,000	0,005	0,007
Validação cruzada (RMSECV)	0,000	0,000	0,000	0,000	0,005	0,009

Fonte: O autor.

6.3. Modelos OPLS-DA

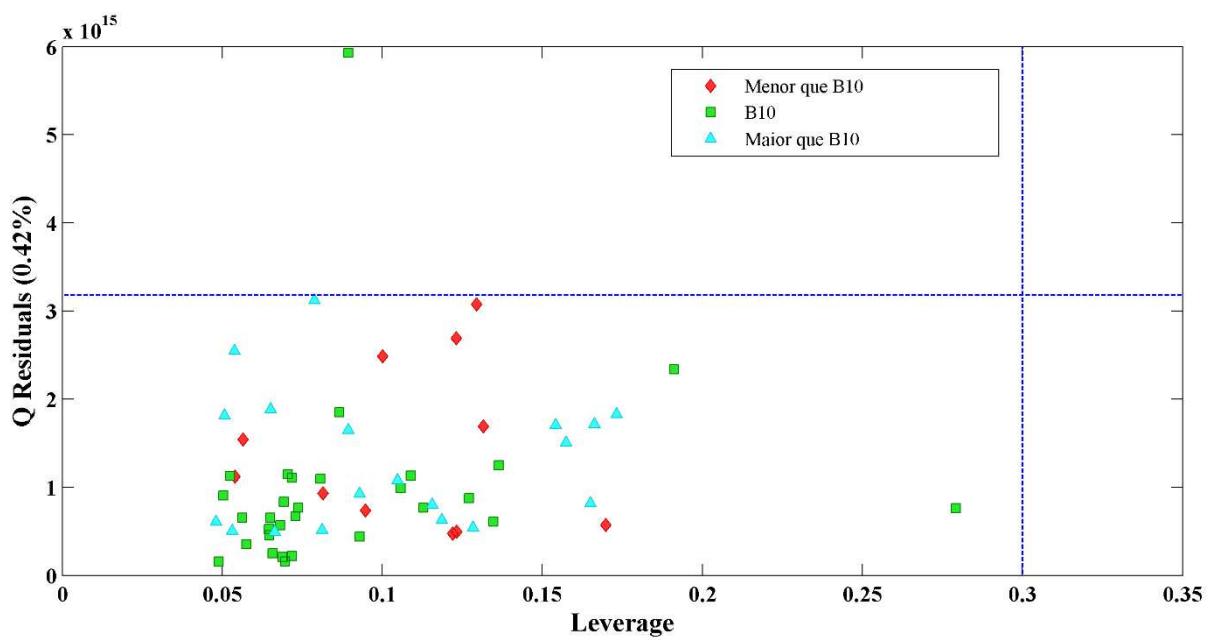
Os modelos OPLS-DA de mafurra e de crambe foram construídos com três variáveis latentes (VLs) para classificar o teor B10 de BMMA e BMC em mistura com o diesel, respectivamente. Estas VLs para o modelo OPLS-DA de mafurra representaram 99,70 e 96,85% da variância explicada nas matrizes **X** e **Y**. As três VLs para o modelo OPLS-DA de crambe representaram 99,34 e 98,45% da variância explicada nas matrizes **X** e **Y**. Nenhuma amostra anômala foi verificada nos dois modelos, como se pode ver nas Figuras 30 e 31, na qual apresentam os valores de *leverage* versus valores *Q residuals* com 95% de confiança para os modelos OPLS-DA de mafurra e de crambe.

Figura 30 – Valores de leverage versus valores Q residuals com 95% de confiança para o modelo OPLS-DA de mafurra.



Fonte: O autor.

Figura 31 – Valores de leverage versus valores Q residuals com 95% de confiança para o modelo OPLS-DA de crambe.



Fonte: O autor.

Os modelos OPLS-DA de mafurra e de crambe foram validados usando as amostras do conjunto de testes e produziram os resultados dos parâmetros de classificação apresentados nas Tabelas 8, 9 e 10. Os valores de RMSEC e RMSEP dos dois modelos são considerados aceitáveis de acordo com a norma brasileira ABNT NBR 15568 (ABNT NBR-15568, 2008) e melhores que dos modelos PLS-DA desenvolvidos com dados de outra técnica pelos pesquisadores Máquina et al. (MÁQUINA et al., 2019), que obtiveram 0,12 e 0,16% (v/v) de RMSEC e RMSEP, respectivamente, ao classificar o teor de biodiesel de algodão em misturas de diesel usando espectroscopia no MIR e PLS-DA; Mazivia et al. (MAZIVILA et al., 2015b), que obtiveram 0,13 e 0,19% (v/v) de RMSEC e RMSEP, respectivamente, ao classificar amostras de mistura B5 de diesel/biodiesel de soja e pinhão manso usando espectroscopia no MIR e PLS-DA.

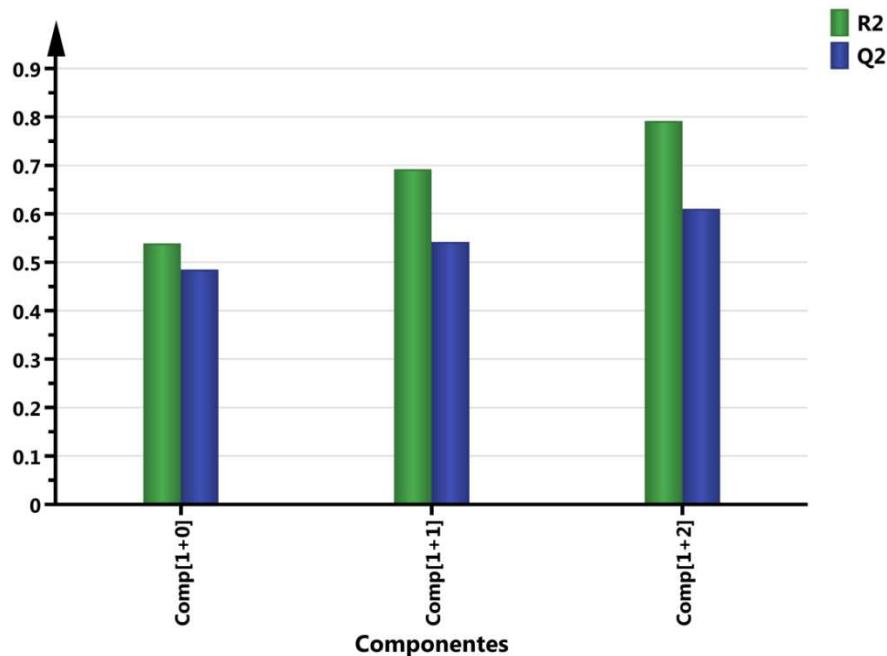
Tabela 8 – Parâmetros de classificação obtidos pelos modelos OPLS-DA de mafurra e de crambe

Parâmetros	Valores do modelo	
	OPLS-DA	
	BMMa	BMC
RMSECV (% v/v)	0,32	0,19
RMSEC (% v/v)	0,08	0,06
RMSEP (% v/v)	0,14	0,07
Especificidade (Cal, CV, Prev)	1,0	1,0
Sensibilidade (Cal, CV, Prev)	1,0	1,0
R ²	0,792	0,956
Q ²	0,610	0,909

Fonte: O autor.

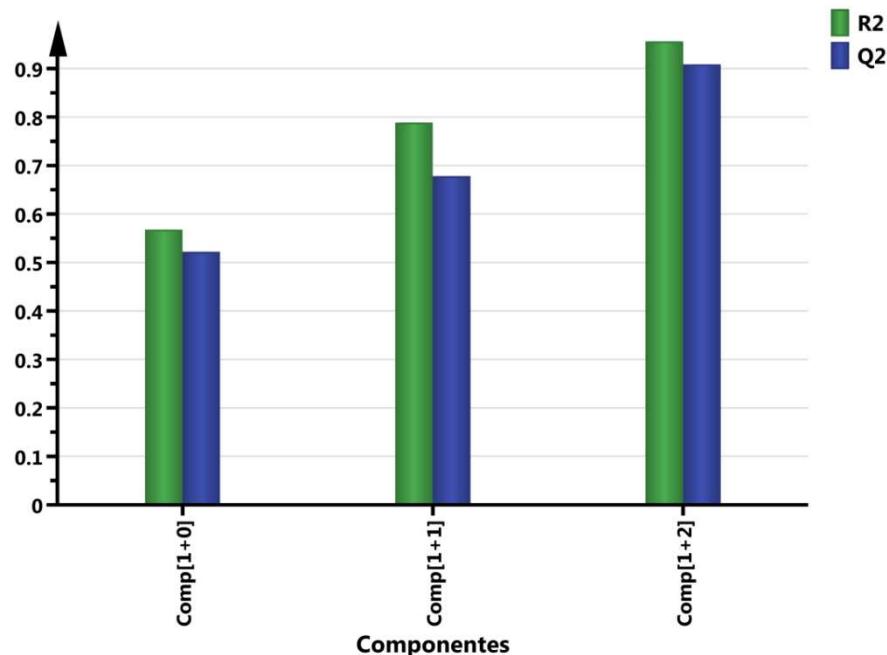
As Figuras 32 e 33 mostram os valores de R² e Q² que foram analisados para avaliar o ajuste dos modelos OPLS-DA de mafurra e de crambe, respectivamente. Os valores destes parâmetros adquiridos com 3 variáveis latentes foram superiores que 0,5, o que equivale a dizer que os modelos OPLS-DA de mafurra e de crambe são robustos e apresentam boa capacidade de classificação.

Figura 32 – Parâmetros R^2 e Q^2 do modelo OPLS-DA de mafurra.



Fonte: O autor.

Figura 33 – Parâmetros R^2 e Q^2 do modelo OPLS-DA de crambe.



Fonte: O autor.

A Tabela 9 apresenta uma matriz de confusão dos modelos OPLS-DA de mafurra e de crambe, na qual se observa que as taxas de *vp* e *vn* foi de 1 e as taxas de *fp* e *fn* foi de 0 para os conjuntos de treinamento e de teste das amostras de B10 e BX. Estes valores foram usados para determinar a sensibilidade e especificidade dos modelos OPLS-DA de mafurra e de crambe, com base nas Equações 15 e 16, respectivamente, no qual se obteve valor de 1 para os dois parâmetros de avaliação da eficiência dos dois modelos, significando que os modelos classificaram corretamente todas as amostras verdadeiras positivas e verdadeiras negativas, respectivamente.

Todas as informações obtidas sobre sensibilidade e especificidade foram combinadas para determinar a eficiência dos dois modelos, de acordo com a Equação 18, tendo obtido resultado de 1, demonstrando que os dois modelos apresentam um bom ajuste e alto desempenho.

O Coeficiente de Correlação de Matthew's dos dois modelos resumiu a qualidade da tabela de matriz de confusão em um único valor numérico igual a 1 para os processos de treinamento e teste, o que representa uma classificação perfeita.

Tabela 9 – Matriz de confusão dos modelos OPLS-DA de mafurra e de crambe

		<i>vp</i>	<i>fp</i>	<i>vn</i>	<i>fn</i>	N
Treinamento	Classe 1 (B10)	1	0	1	0	20
	Classe 2 (BX)	1	0	1	0	20
	Eficiência = 1					
	Coeficiente de Correlação Matthew's = 1					
Teste	Classe 1 (B10)	1	0	1	0	10
	Classe 2 (BX)	1	0	1	0	10
	Eficiência = 1					
	Coeficiente de Correlação Matthew's = 1					

Fonte: O autor.

Os resultados apresentados na Tabela 9 permitiram que:

- No processo de treinamento dos dois modelos, todas as 20 amostras de B10 fossem previstas na classe 1 e nenhuma amostra de BX fosse prevista na classe 1, assim como todas as 20 amostras de BX fossem previstas na classe 2 e nenhuma amostra de B10 fosse prevista na classe 2, como pode-se observar na tabela de confusão (Tabela 10);
- No processo de teste dos dois modelos, todas as 10 amostras de B10 fossem previstas na classe 1 e nenhuma amostra de BX fosse prevista na classe 1, assim como todas as 10 amostras de BX fossem previstas na classe 2 e nenhuma amostra de B10 fosse prevista na classe 2, como pode-se observar na tabela de confusão (Tabela 10).

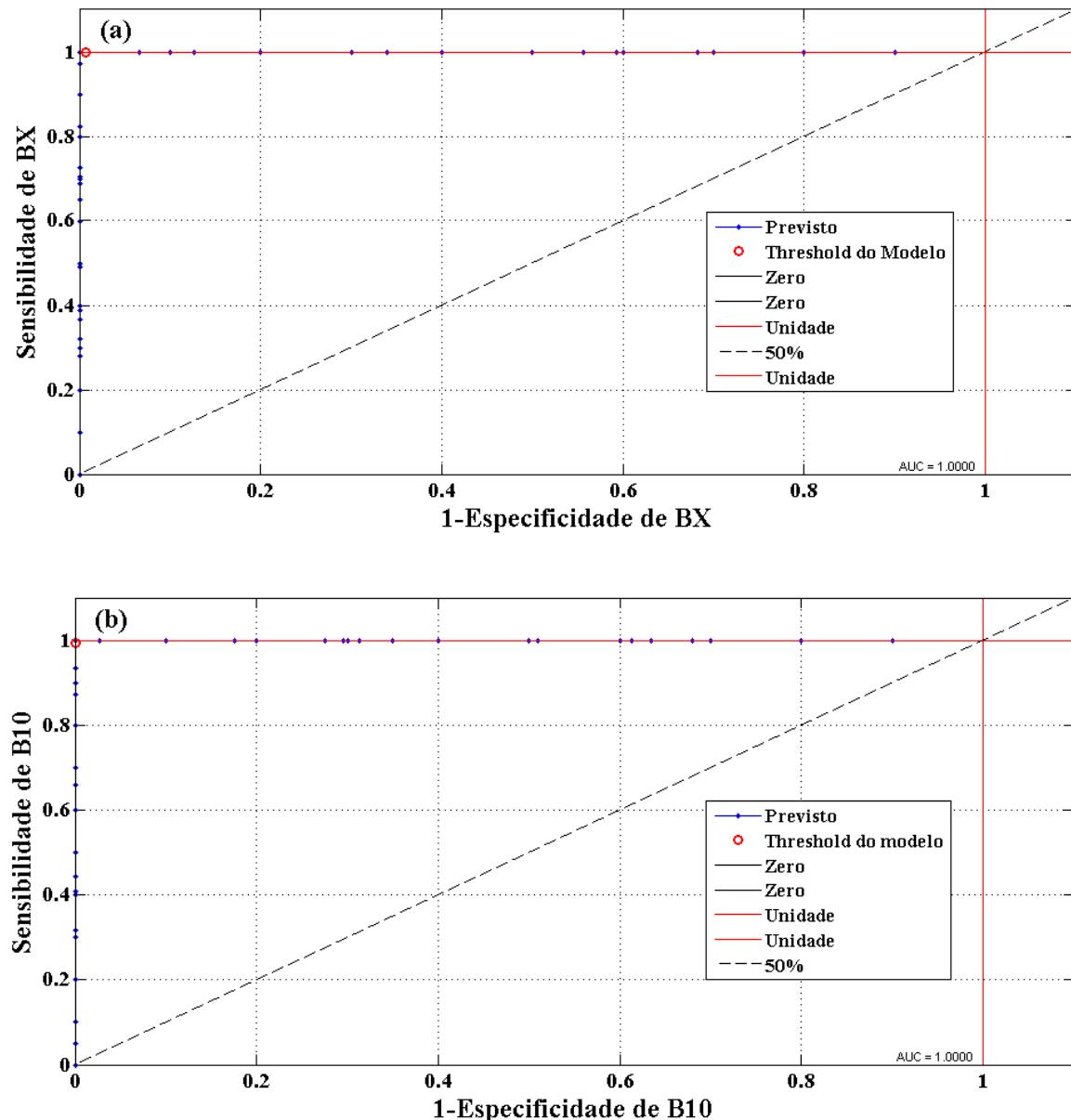
Tabela 10 – Tabela de confusão dos modelos OPLS-DA de mafurra e de crambe

		Classe 1	Classe 2
Treinamento	Previsto como Classe 1 (B10)	20	0
	Previsto como Classe 2 (BX)	0	20
	Previsto como sem classe	0	0
Teste	Previsto como Classe 1 (B10)	10	0
	Previsto como Classe 2 (BX)	0	10
	Previsto como sem classe	0	0

Fonte: O autor.

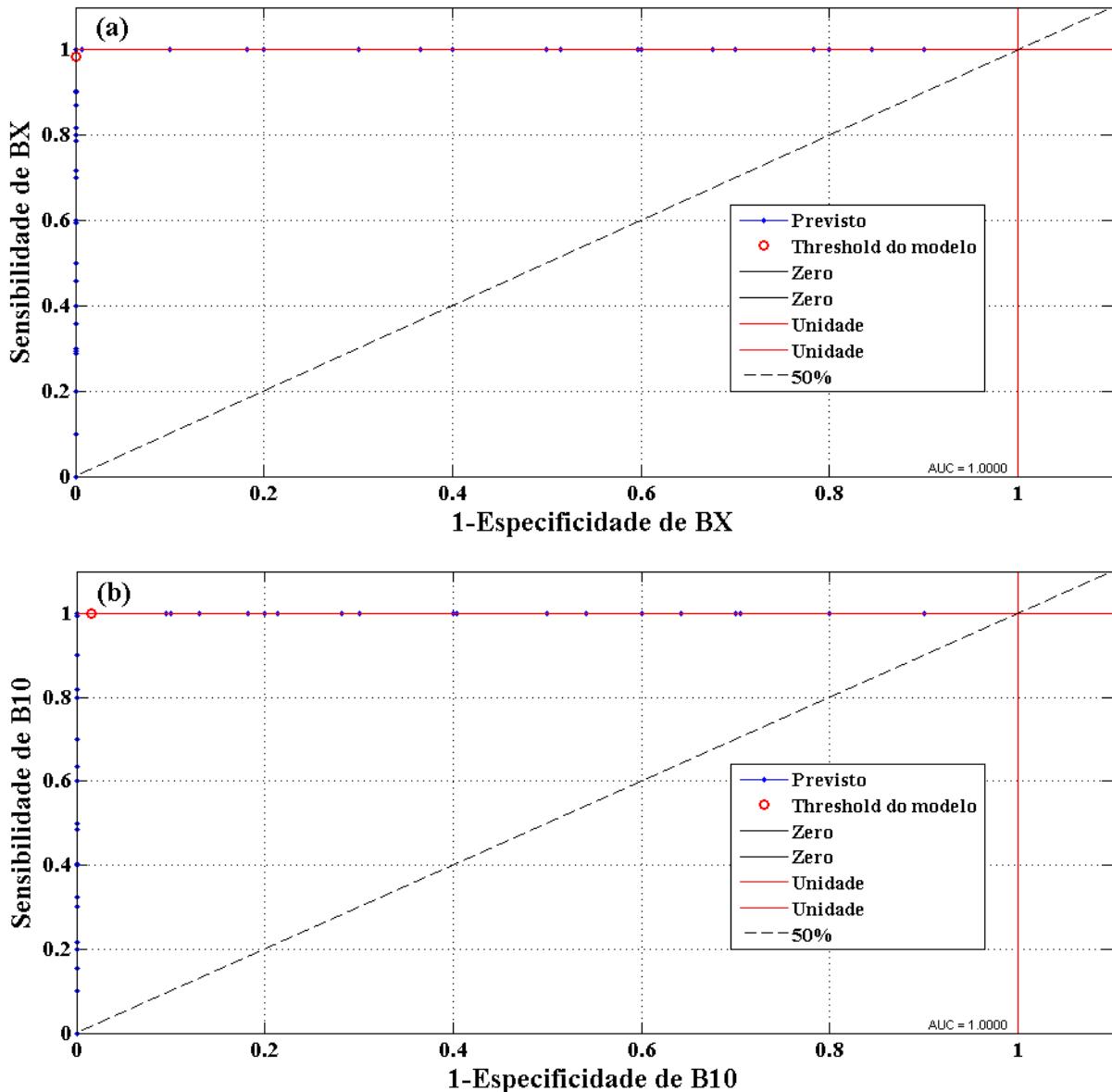
O valor da sensibilidade e especificidade também é observado na Curva de características operacionais do receptor (ROC curve) dos modelos OPLS-DA de mafurra e de crambe, apresentados nas Figuras 34 e 35, respectivamente, que mostram a sensibilidade e especificidade como funções do *threshold* na separação das classes de B10 e BX.

Figura 34 – Curva ROC do modelo OPLS-DA de mafurra para (a) sensibilidade *versus* especificidade do BX; (b) sensibilidade *versus* especificidade do B10.



Fonte: O autor.

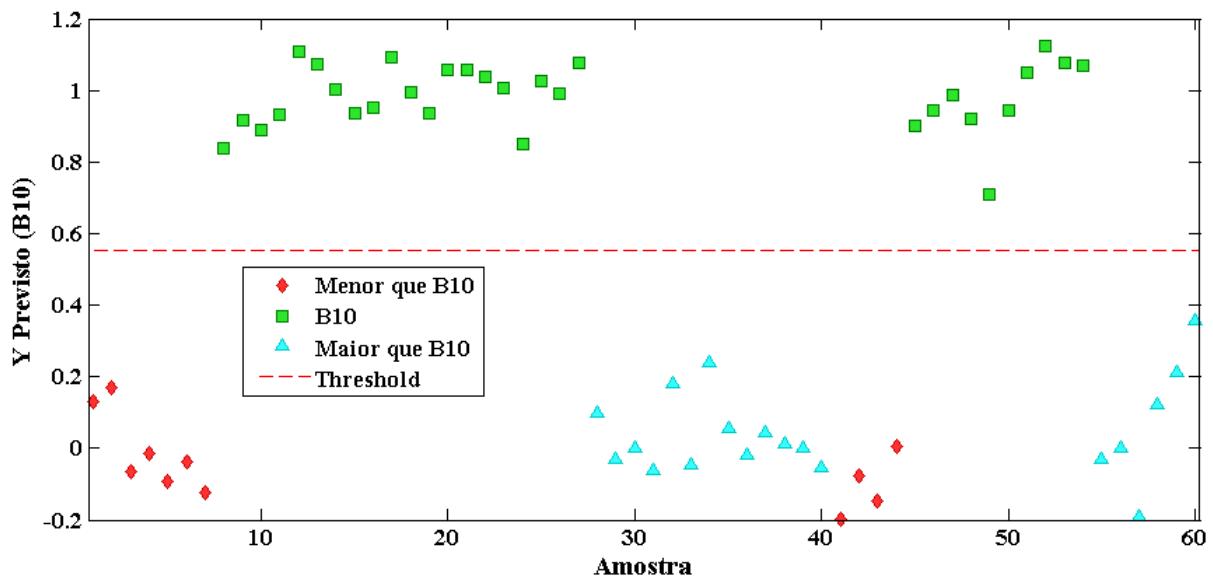
Figura 35 – Curva ROC do modelo OPLS-DA de crambe para (a) sensibilidade *versus* especificidade do BX; (b) sensibilidade *versus* especificidade do B10.



Fonte: O autor.

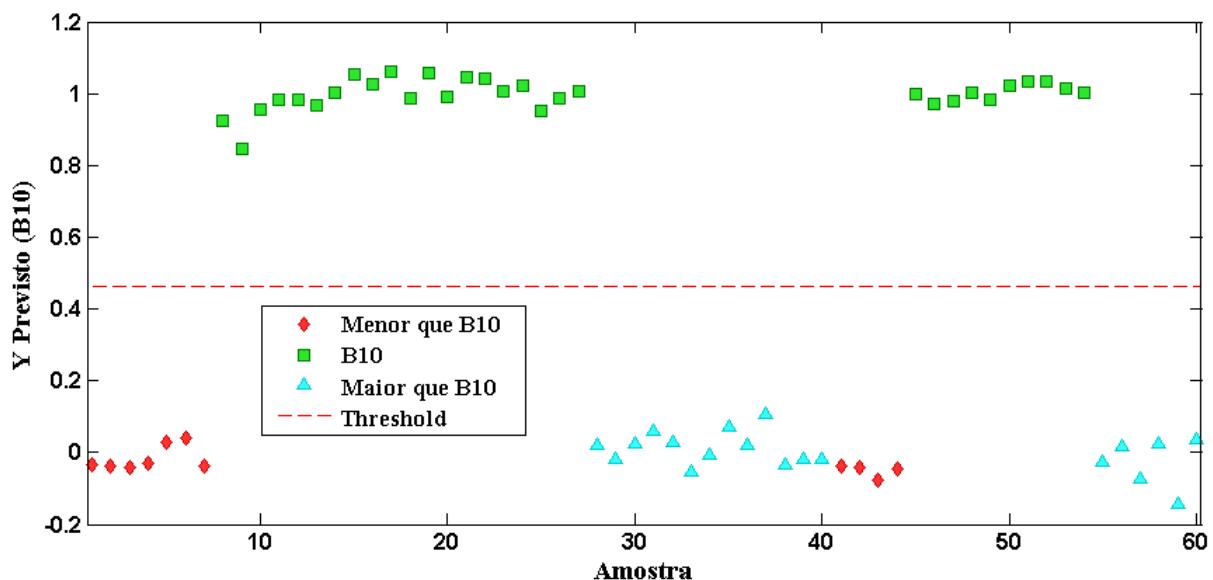
Os resultados apresentados nas Tabelas 9 e 10 mostram que os modelos OPLS-DA de mafurra e de crambe possuem alta eficiência na classificação das amostras B10 e BX do BMMA e BMC. Essa eficiência de classificação é observada nas Figuras 36 e 37, nas quais se observa que as amostras B10 dos conjuntos de treinamento e de teste estão acima do *threshold*, separando-se das amostras BX dos conjuntos de treinamento e de teste que estão abaixo do *threshold*.

Figura 36 – Valores estimados de classe para conjunto de treinamento e de teste das amostras B10 e BX do modelo OPLS-DA de mafurra.



Fonte: O autor.

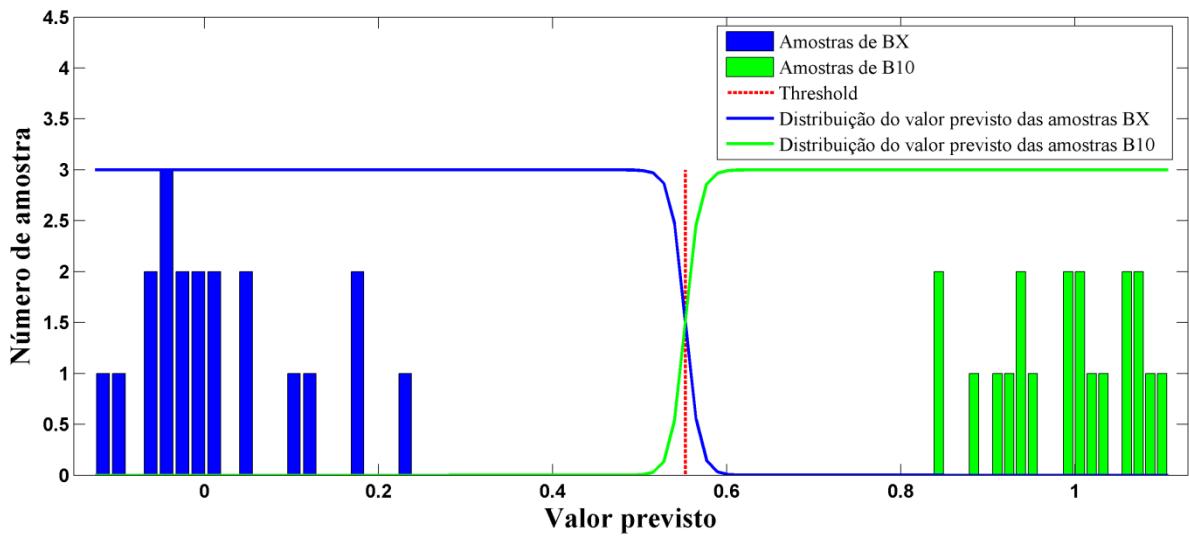
Figura 37 – Valores estimados de classe para conjunto de treinamento e de teste das amostras B10 e BX do modelo OPLS-DA de crambe.



Fonte: O autor.

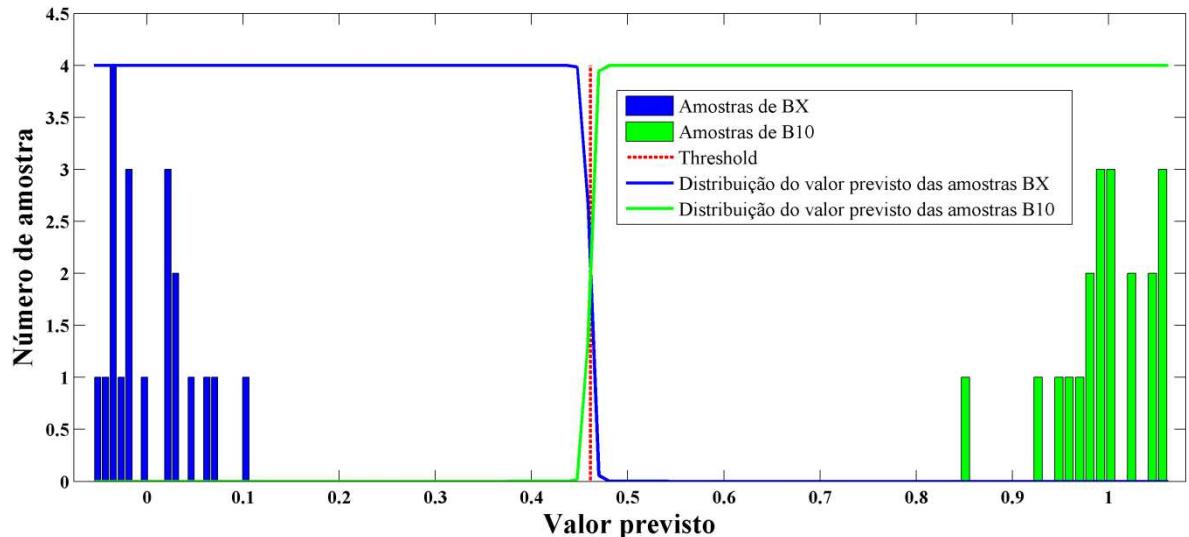
As Figuras 38 e 39 mostram a seleção do *threshold* em 0,55 e 0,41 para os modelos OPLS-DA de mafurra e de crambe, respectivamente, quando as distribuições estimadas da classe 1 (B10) e classe 2 (BX) se cruzam. O limite selecionado é o valor de *y* previsto, em que o número de falsos positivos e falsos negativos é minimizado para previsões futuras.

Figura 38 – Seleção do *threshold* pelo teorema da Bayer para amostras B10 e BX do modelo OPLS-DA de mafurra.



Fonte: O autor.

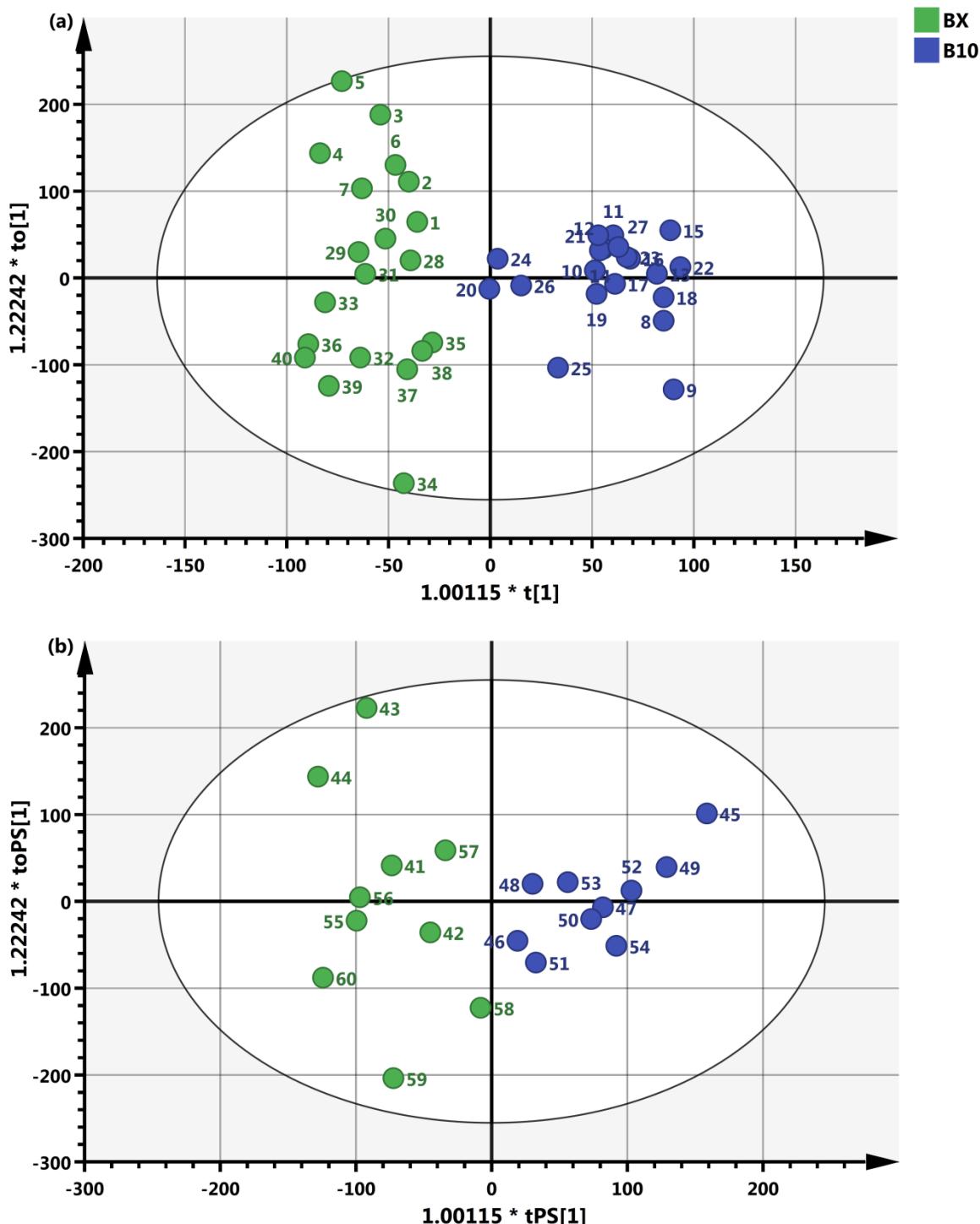
Figura 39 – Seleção do *threshold* pelo teorema da Bayer para amostras B10 e BX do modelo OPLS-DA de crambe.



Fonte: O autor.

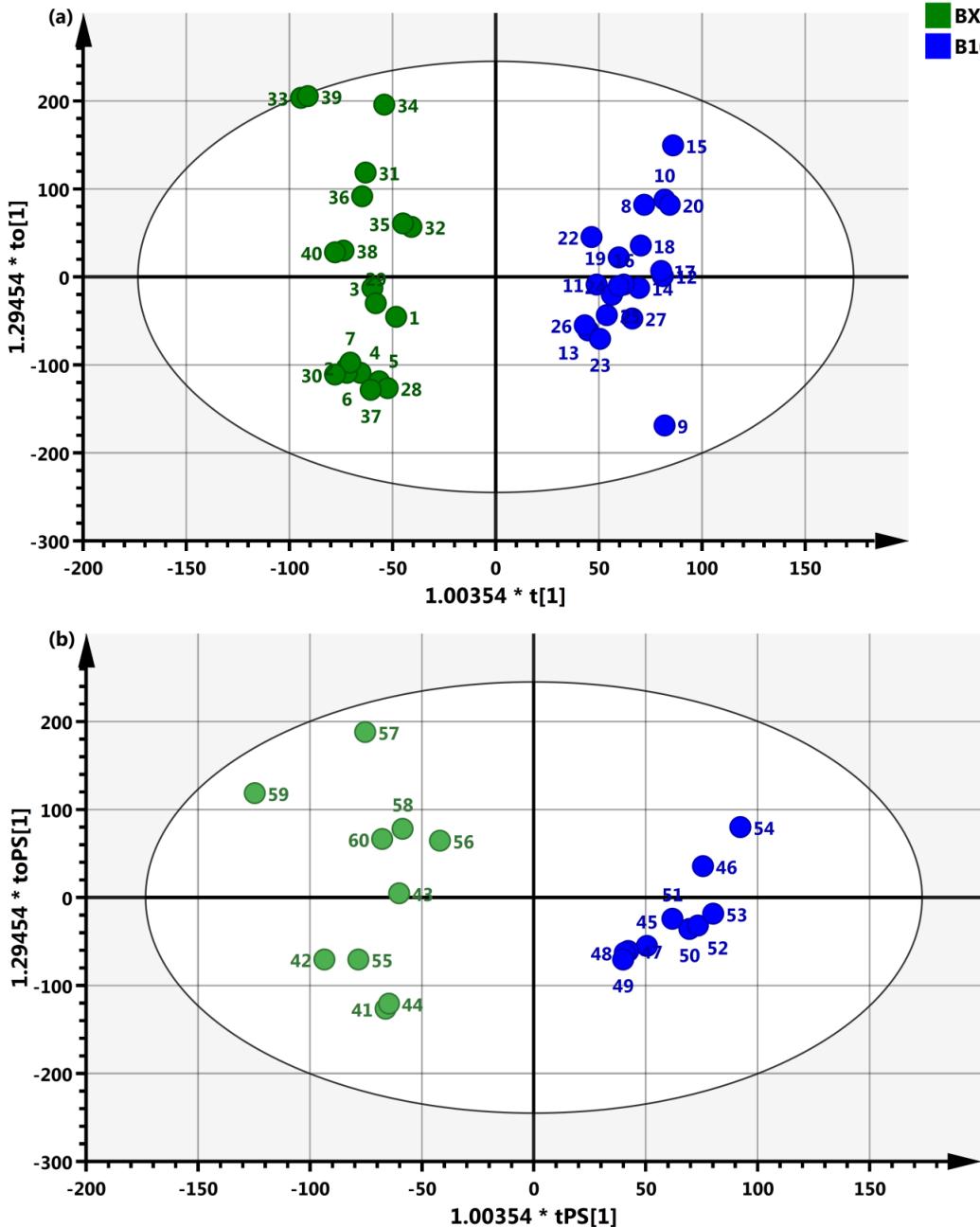
As Figuras 40 e 41 apresentam gráficos de escores para os modelos OPLS-DA de mafurra e de crambe, nos conjuntos de (a) calibração e (b) previsão, que efetivamente separaram a direção preditiva em $t[1]$ da direção ortogonal-Y em $t_0[1]$, proporcionando uma boa separação entre amostras de BX e B10.

Figura 40 – Escores para o modelo OPLS-DA de mafurra em (a) calibração e (b) previsão.



Fonte: O autor.

Figura 41 – Escores para o modelo OPLS-DA de crambe em (a) calibração e (b) previsão.



Fonte: O autor.

A análise discriminante foi realizada com base no perfil dos espectros de RMN de ^1H das misturas B10/BX de BMMA e de BMC. As regiões espectrais que contribuíram para a diferenciação entre as classes são observadas nas Figuras 42 e 43 para BMMA e BMC. Os pesos das variáveis latentes mostraram que as três VLs aplicadas ao modelo OPLS-DA de mafurra descreveram 99,70% das variáveis que influenciaram a discriminação de classe, em que toda a variação definida por classe de interesse está na VL-1 (98,91%) e da classe não de

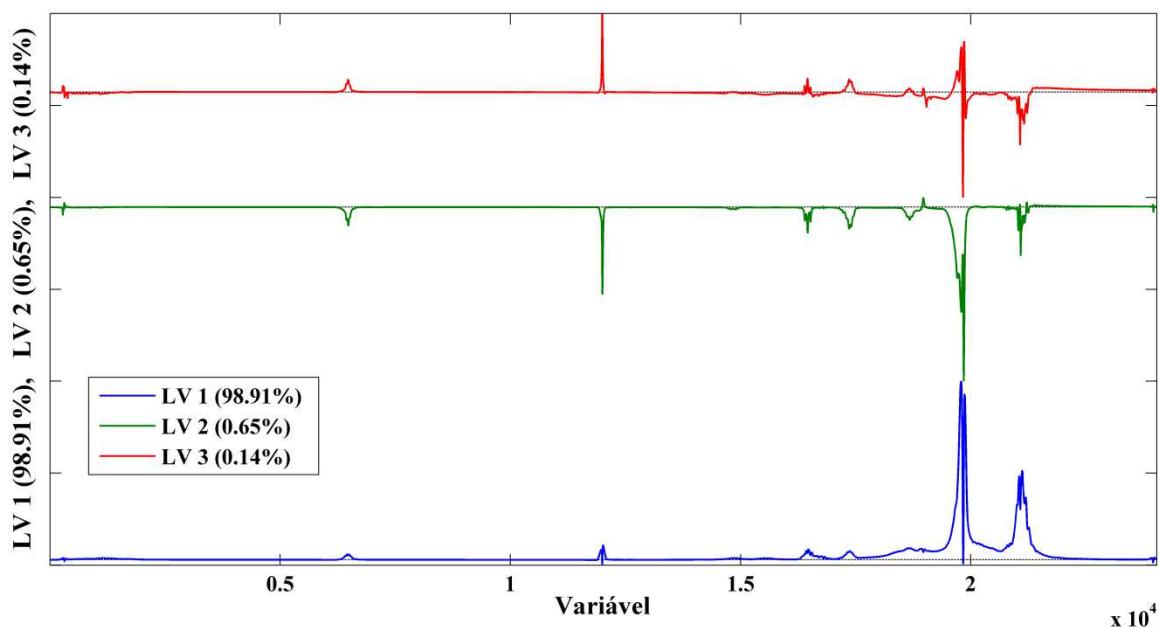
interesse estão nas VL-2 (0,65%) e VL-3 (0,14%). Para o modelo OPLS-DA de crambe, as três VLs descreveram 99,34% das variáveis que influenciaram a discriminação de classe, em que toda a variação definida por classe de interesse está na VL-1 (97,23%) e da classe não de interesse estão nas VL-2 (1,95%) e VL-3 (0,16%). Portanto, as variáveis que diferenciam B10 de BX são restritas ao VL-1, encontradas nas regiões espectrais que são apresentadas na Tabela 11.

Tabela 11 – Regiões espectrais e suas atribuições

Deslocamento químico (ppm)	Atribuição
0.8	Hidrogênios metílicos terminais
1.3	Metilenos da cadeia do carbono
1.6	Multipletos dos β -carbonil metilenos
2.1	Insaturação atribuída ao alílico
2.3	Tripleto relacionado à α -carbonil metilenos
2.8	Insaturação atribuída ao bis- alílicos
3.7	Prótons metoxi

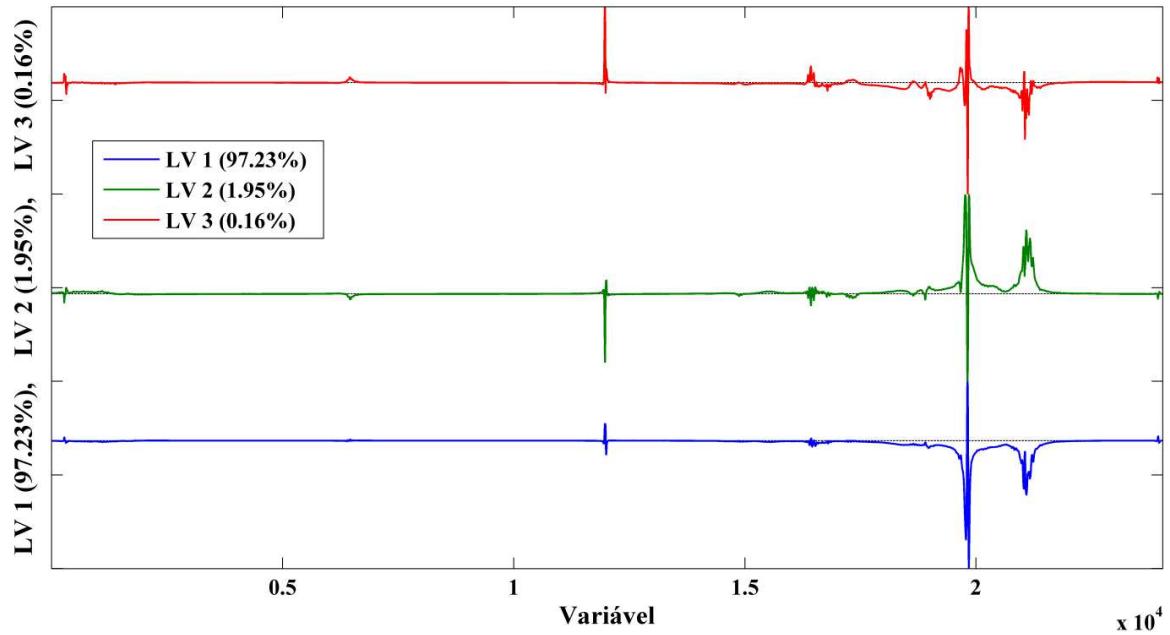
Fonte: O autor.

Figura 42 – Gráfico dos pesos de VL-1, VL-2, VL-3 *versus* variáveis do modelo OPLS-DA de mafurra.



Fonte: O autor.

Figura 43 – Gráfico dos pesos de VL-1, VL-2, VL-3 *versus* variáveis do modelo OPLS-DA de crambe.



Fonte: O autor.

A Tabela 12 apresenta os resultados dos testes de permutação realizados nos modelos OPLS-DA de mafurra e de crambe para verificar o ajuste excessivo dos modelos. Estes resultados mostram que os dois modelos não estão super ajustados porque os valores de Wilcoxon (teste de classificação assinado em pares Wilcoxon), Sign Test (teste de classificação assinada em pares) e Rand t-test (Teste-t de randomização) são menores que 0,05.

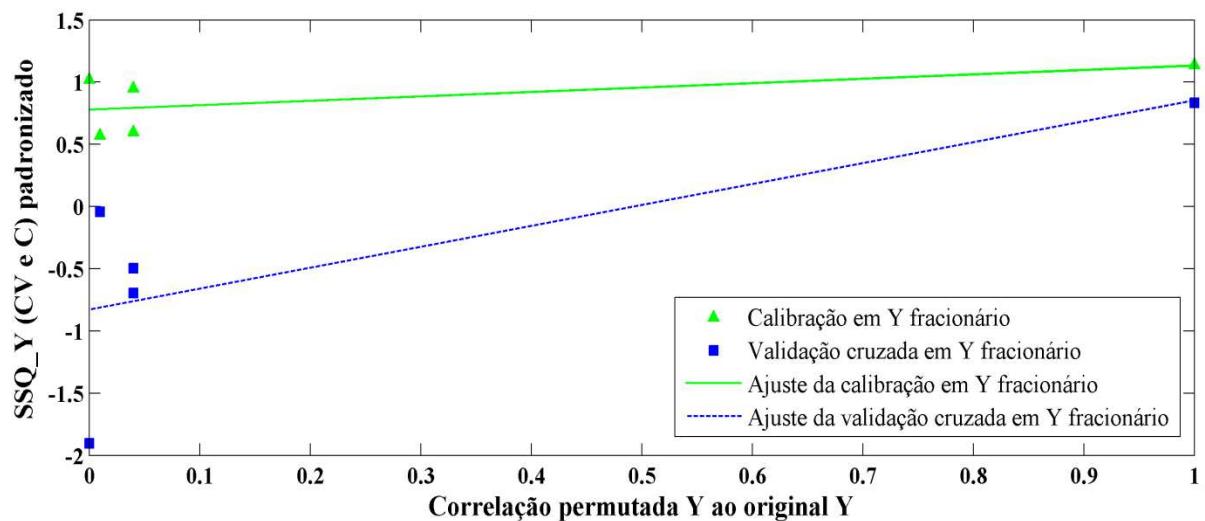
A variância capturada no \mathbf{y} fracionário para calibração e validação cruzada *versus* a correlação da matriz \mathbf{Y} permutado com a matriz \mathbf{Y} original foram analisadas nos modelos OPLS-DA de mafurra e de crambe, como se observa nas Figuras 44 e 45, respectivamente, em que a Soma do Quadrado \mathbf{Y} da calibração ($SSQ_{Y,C}$) aumentou para o valor próximo de "1" e a Soma do Quadrado \mathbf{Y} da validação cruzada ($SSQ_{Y,CV}$) foi aproximadamente o mesmo que $SSQ_{Y,C}$. Assim, prova que os dois modelos OPLS-DA não estão super ajustado, pois além dos valores de validação cruzada e de calibração serem relativamente próximos um do outro, eles são significativamente menores que os resultados para a matriz \mathbf{Y} não permutado (no lado direito do gráfico).

Tabela 12 – Probabilidade de insignificância dos modelos OPLS-DA de mafurra e de crambe com 3 VLs

	Wilcoxon		Sign Test		Rand t-test	
	BMMa	BMC	BMMa	BMC	BMMa	BMC
Coluna Y: 1						
Auto-Previsão (RMSEC)	0,000	0,000	0,001	0,000	0,005	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,000	0,000	0,005	0,005
Coluna Y: 2						
Auto-Previsão (RMSEC)	0,000	0,000	0,001	0,000	0,005	0,005
Validação cruzada (RMSECV)	0,000	0,000	0,000	0,000	0,005	0,005

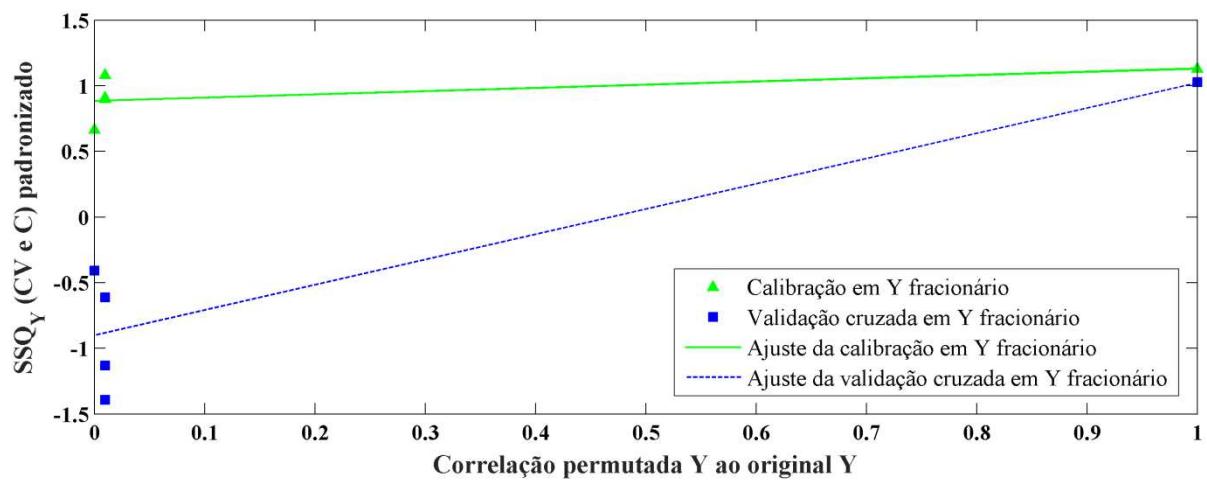
Fonte: O autor.

Figura 44 – Gráfico da Soma do Quadrado Y (SSQ_Y) versus correlação da matriz Y para o modelo OPLS-DA de mafurra.



Fonte: O autor.

Figura 45 – Gráfico da Soma do Quadrado Y (SSQ_Y) versus correlação da matriz Y para o modelo OPLS-DA de crambe.



Fonte: O autor.

7. CONCLUSÕES

A aplicação da regressão multivariada por OPLS e OPLS-DA nos dados espectrais de RMN de ^1H de misturas diesel/biodiesel de mafurra e de crambe permitiu o desenvolvimento de métodos eficientes capazes de determinar o teor do biodiesel nas misturas com diesel.

Todos os métodos desenvolvidos no trabalho foram validados de acordo com as normas ABNT NBR 15568, ASTM D7371 e ASTM E1655-05.

A eficiência dos métodos desenvolvidos com base na regressão multivariada por OPLS foi analisada com base nas figuras de mérito e no ajuste dos modelos. Os resultados das figuras de mérito foram concordantes com os requisitos estabelecidos na norma ASTM E1655-05 e o ajuste dos modelos mostrou uma alta correlação existente entre valores de concentração real e previsto dos conjuntos de calibração e previsão, com um coeficiente de correlação superior que 0,99 e com erros relativamente baixos.

A eficiência dos métodos desenvolvidos com base na regressão multivariada por OPLS-DA foi analisada com base nos parâmetros de taxa de falso positivos e falso negativos, sensibilidade, especificidade e coeficiente de correlação de Matthew, onde não foi notada a presença de amostras falsos positivos e nem falsos negativos, consequentemente, os parâmetros de sensibilidade, especificidade e coeficiente de correlação de Matthew foram iguais a 1. Esses resultados levaram a uma classificação 100% correta das amostras de B10 e BX utilizadas nos conjuntos de treinamento e teste, demonstrando uma alta eficiência dos modelos OPLS-DA no monitoramento do teor do biodiesel metílico de mafurra e de crambe quando misturado com diesel em várias proporções.

A alta eficiência apresentada nos modelos OPLS e OPLS-DA para quantificar e classificar amostras em relação ao teor do biodiesel metílico de mafurra e crambe em mistura com diesel sugere que os métodos analíticos desenvolvidos são viáveis e podem ser usados pelas agências reguladoras para o controle quantitativo e qualitativo destes combustíveis.

REFERÊNCIAS

- ABNT NBR-15568. **Biodiesel -Determinação do teor de biodiesel em óleo diesel por espectroscopia na região do infravermelho médio.** Brasil, 2008. Disponível em: <https://www.target.com.br/produtos/normas-tecnicas/40641/nbr15568-biodiesel-determinacao-do-teor-de-biodiesel-em-oleo-diesel-por-espectroscopia-na-regiao-do-infravermelho-medio>
- ALMEIDA, M. R. et al. Classification of Amazonian rosewood essential oil by Raman spectroscopy and PLS-DA with reliability estimation. **Talanta**, v. 117, p. 305–311, 2013. <https://doi.org/10.1016/j.talanta.2013.09.025>
- ANP. **Boletim de Monitoramento da Qualidade dos Combustíveis.** in: PMQC. Rio de Janeiro, 2019. Disponível em: <http://www.anp.gov.br/arquivos/central-conteudos/boletins/pmqc/201910-boletim-pmqc.pdf>. Acesso em: 11 dez. 2019
- ANP - AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (BRASIL). **Resolução ANP no. 14, de 11 maio 2012,** 2012. Disponível em: <http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2012/maio&item=ramp-14--2012&export=pdf>. Acesso em: 18 fev. 2020
- ANP - AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (BRASIL). **Boletim de monitoramento da qualidade dos combustíveis,** 2019. Disponível em: http://www.anp.gov.br/images/QUALIDADE/BOLETIM/2019_JAN_Boletim_PMQC_CO-P2-A2.pdf
- ARNDT, C. et al. Biofuels, poverty, and growth: A computable general equilibrium analysis of Mozambique. **Environment and Development Economics**, v. 15, n. 1, p. 81–105, 2010. <https://doi.org/10.1017/S1355770X09990027>
- ASTM STANDARD D7371. **Standard Test Method for Determination of Biodiesel (Fatty Acid Methyl Esters) Content in Diesel Fuel Oil Using Mid Infrared Spectroscopy (FTIR-ATR-PLS Method)** West Conshohocken, PA: ASTM International; 1991, www.astm.org. United States, 2014.
- ASTM STANDARD E1655-05. **Standard Practices for Infrared Multivariate Quantitative Analysis.** United States, 2012. Disponível em: WWW.compass.astm.org/Standards/HISTORICAL/E1655-05R12.htm
- BALAT, M.; BALAT, H. Recent trends in global production and utilization of bio-ethanol fuel. **Applied Energy**, v. 86, n. 11, p. 2273–2282, nov. 2009. <https://doi.org/10.1016/j.apenergy.2009.03.015>
- BAMBO, T. F. **Cooperação Sul-Sul: o acordo Brasil-Moçambique na área de biocombustíveis.** São Paulo: Universidade de São Paulo, 31 mar. 2014. Disponível em: https://teses.usp.br/teses/disponiveis/101/101131/tde-26052014-111753/publico/Tome_Fernando_Bambo.pdf

BARKER, M.; RAYENS, W. Partial least squares for discrimination. **Journal of Chemometrics**, v. 17, n. 3, p. 166–173, 2003. <https://doi.org/10.1002/cem.785>

BARRA, I. et al. Discrimination of diesel fuels marketed in Morocco using FTIR, GC-MS analysis and chemometrics methods. **Talanta**, v. 209, 1 mar. 2020. <https://doi.org/10.1016/j.talanta.2019.120543>

BATIDZIRAI, B.; FAAIJ, A. P. C.; SMEETS, E. Biomass and bioenergy supply from Mozambique. **Energy for Sustainable Development**, v. 10, n. 1, p. 54–81, 2006. [https://doi.org/10.1016/S0973-0826\(08\)60507-4](https://doi.org/10.1016/S0973-0826(08)60507-4)

BRASIL. DECRETO S/Nº, de 23 de Dezembro de 2003 - DOU 24.12.2003: Institui a Comissão Executiva Interministerial encarregada da implantação das ações direcionadas à produção e ao uso de óleo vegetal - biodiesel como fonte alternativa de energia, 2003. Disponível em:
http://www.planalto.gov.br/ccivil_03/DNN/2003/Dnn10093.htm

BRASIL. LEI Nº 11.097, DE 13.1.2005. Dispõe sobre a introdução do biodiesel na matriz energética brasileira; altera as Leis nºs 9.478, de 6 de agosto de 1997, 9.847, de 26 de outubro de 1999 e 10.636, de 30 de dezembro de 2002; e dá outras providências, 2005. Disponível em: <http://legislacao.anp.gov.br/?path=legislacao-federal/leis/2005&item=lei-11.097--2005>

BRASIL. Lei no. 13.033, de 24 set. 2014: Dispõe sobre a adição obrigatória de biodiesel ao óleo diesel comercializado com o consumidor final; altera as Leis nº 9.478, de 6 de agosto de 1997, e 8.723, de 28 de outubro de 1993; revoga dispositivos da Lei no 11.097, [s.l.: s.n.]. Disponível em: <https://www.cosif.com.br/captura.asp?arquivo=lo13033>. Acesso em: 21 jun. 2019.

BRASIL. BALANÇO ENERGÉTICO NACIONAL 2018: Ano base 2017. Rio de Janeiro: Copyright © 2019 EPE – Empresa de Pesquisa Energética, 2018a. Disponível em: http://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-303/topico-419/BEN2018__Int.pdf

BRASIL. Análise de Conjuntura dos Biocombustíveis Ano 2018. Rio de Janeiro: Copyright © 2017 EPE, 2018b. Disponível em: http://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-402/An%C3%A1lise_de_Conjuntura_Ano%202018.pdf

BRASIL, R. F. DO. BALANÇO ENERGÉTICO NACIONAL 2019: Ano base 2018. Rio de Janeiro: Copyright © 2019 EPE – Empresa de Pesquisa Energética, 2019. Disponível em: <http://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-377/topico-494/BEN%202019%20Completo%20WEB.pdf>

BUIATTE, J. E. et al. Qualitative and quantitative monitoring of methyl cotton biodiesel content in biodiesel/diesel blends using MIR spectroscopy and chemometrics tools. **Journal of the Brazilian Chemical Society**, v. 27, n. 1, p. 84–90, 2016. <https://doi.org/10.5935/0103-5053.20150251>

CÂMARA, A. B. F. et al. MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-diesel blends. **Fuel**, v. 210, p. 497–506, 2017. <https://doi.org/10.1016/j.fuel.2017.08.072>

CLAYDEN, J.; GREEVES, N.; WARREN, S. **Organic Chemistry**. 2^a ed. New York: Oxford University Press, 2012.

Crambe hispanica – Wikipédia, a encyclopédia livre. Disponível em: https://pt.wikipedia.org/wiki/Crambe_hispanica. Acesso em: 29 mar. 2020.

CUMBE, Â. N. F. **O Património Geológico de Moçambique: Proposta de Metodologia de Inventariação, Caracterização e Avaliação**. [s.l.] UNIVERSIDADE DO MINHO, 2007. Disponível em: http://www.dct.uminho.pt/mest/pgg/docs/tese_cumbe.pdf

DE CARVALHO ROCHA, W. F.; NOGUEIRA, R.; VAZ, B. G. Validation of model of multivariate calibration: An application to the determination of biodiesel blend levels in diesel by near-infrared spectroscopy. **Journal of Chemometrics**, v. 26, n. 8–9, p. 456–461, 2012. <https://doi.org/10.1002/cem.2420>

DE SOUZA, L. M. et al. Quantification of adulterations in extra virgin flaxseed oil using MIR and PLS. **Food Chemistry**, v. 182, p. 35–40, 2015. <https://doi.org/10.1016/j.foodchem.2015.02.081>

DONG, T. et al. Two-step in situ biodiesel production from microalgae with high free fatty acid content. **Bioresource Technology**, v. 136, p. 8–15, 2013. <https://doi.org/10.1016/j.biortech.2013.02.105>

EUROPEAN COMMISSION DIRECTIVE 2012/0288 (COD)C7-0337/12. **Proposal for a Directive of the European Parliament and of the Council amending Directive 98/70/EC relating to the quality of petrol and diesel fuels and amending Directive 2009/28/EC on the promotion of the use of energy from renewable sources**. Brussels: [s.n.]. Disponível em: [http://www.europarl.europa.eu/meetdocs/2009_2014/documents/com/com_com\(2012\)0595/com_com\(2012\)0595_en.pdf](http://www.europarl.europa.eu/meetdocs/2009_2014/documents/com/com_com(2012)0595/com_com(2012)0595_en.pdf)

EUROPEAN STANDARD DIN EN 14078. **Liquid petroleum products - Determination of fatty acid methyl ester (FAME) content in middle distillates - Infrared spectrometry method** Berlin: Deutsches Institut für Normung, 2004. Disponível em: <https://publishing.energyinst.org/topics/renewables/biofuels/ip-579-liquid-petroleum-products-determination-of-fatty-acid-methyl-ester-fame-content-in-middle-distillates-infrared-spectrometry-method>

FAN, S. et al. Classification of Chinese wine varieties using ¹H NMR spectroscopy combined with multivariate statistical analysis. **Food Control**, v. 88, p. 113–122, 1 jun. 2018. <https://doi.org/10.1016/j.foodcont.2017.11.002>

FARAGUNA, F.; RACAR, M.; JUKIĆ, A. Test method for determination of different biodiesels (fatty acid alkyl esters) content in diesel fuel using FTIR-ATR. **Renewable Energy**, p. 1231–1235, 1 abr. 2019. <https://doi.org/10.1016/j.renene.2018.09.010>

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. 109, p. 158–164, 2013. <https://doi.org/10.1016/j.microc.2012.03.008>

FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações**. Campinas-SP: Editora da Unicamp, 2015. <https://doi.org/10.7476/9788526814714>

GRUTZNER, J. B. NUCLEAR MAGNETIC RESONANCE SPECTROSCOPY | Principles. In: **Encyclopedia of Analytical Science**. [s.l.] Elsevier, 2005. v. 11, p. 211–237. <https://doi.org/10.1016/B0-12-369397-7/00405-2>

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

HOTELLING, H. Analysis of a complex of statistical variables into principal components. **Journal of Educational Psychology**, v. 24, n. 6, p. 417–441, 1933. <https://doi.org/10.1037/h0071325>

JANAUN, J.; ELLIS, N. Perspectives on biodiesel as a sustainable fuel. **Renewable and Sustainable Energy Reviews**, v. 14, p. 1312–1320, 2010. <https://doi.org/10.1016/j.rser.2009.12.011>

JOVIĆ, O. et al. 1 H NMR adulteration study of hempseed oil with full chemometric approach on large variable data. **Chemometrics and Intelligent Laboratory Systems**, v. 185, p. 41–46, 15 fev. 2019. <https://doi.org/10.1016/j.chemolab.2018.12.012>

KARL PEARSON. On lines and planes of closest fit to systems of points in space. **Philosophical Magazine**, v. 2, p. 559–572, 1901. <https://doi.org/10.1080/14786440109462720>

KEE LAM, M.; TEONG LEE, K.; RAHMAN MOHAMED, A. Homogeneous, heterogeneous and enzymatic catalysis for transesterification of high free fatty acid oil (waste cooking oil) to biodiesel: A review. **Biotechnology Advances**, v. 28, p. 500–518, 2010. <https://doi.org/10.1016/j.biotechadv.2010.03.002>

KNOTHE, G. et al. **Manual de Biodiesel**. 1^a ed. São Paulo: Editora Blucher, 2006.

KOMANE, B. M.; OLIVIER, E. I.; VILJOEN, A. M. Trichilia emetica (Meliaceae) - A review of traditional uses, biological activities and phytochemistry. **Phytochemistry Letters**, v. 4, n. 1, p. 1–9, 2011. <https://doi.org/10.1016/j.phytol.2010.11.002>

LEUNG, D. Y. C.; WU, X.; LEUNG, M. K. H. A review on biodiesel production using catalyzed transesterification. **Applied Energy**, v. 87, p. 1083-1095, 2009. <https://doi.org/10.1016/j.apenergy.2009.10.006>

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**, v. 610, p. 196–204, 2008. <https://doi.org/10.1016/j.aca.2008.01.039>

LORBER, A. Error Propagation and Figures of Merit for Quantification. **Analytical Chemistry**, v. 58, n. 8, p. 1167–1172, 1986. <https://doi.org/10.1021/ac00297a042>

LORBER, A.; FABER, K.; KOWALSKI, B. R. Net Analyte Signal Calculation in Multivariate Calibration. **Analytical Chemistry**, v. 69, n. 1–2, p. 1620–1626, 1997. <https://doi.org/10.1021/ac960862b>

MÁQUINA, A. D. V. et al. Fast quantitative and qualitative monitoring of mafurra biodiesel content using fourier transform mid-infrared spectroscopy, chemometric tools, and variable selection. **Energy and Fuels**, v. 31, n. 1, p. 571–577, 2017. <https://doi.org/10.1021/acs.energyfuels.6b02079>

MÁQUINA, A. D. V. et al. Quantification and classification of cotton biodiesel content in diesel blends, using mid-infrared spectroscopy and chemometric methods. **Fuel**, v. 237, n. September 2018, p. 373–379, 2019. <https://doi.org/10.1016/j.fuel.2018.10.011>

MASSART, D. L. et al. **Handbook of Chemometrics and Qualimetrics, Volume 20B.** Amsterdã: Elsevier Science, 1998.

MAX, B. et al. OPLS discriminant analysis : combining the strengths of PLS-DA and SIMCA classification y. **Journal of Chemometrics**, v. 20, p. 341–351, 2006. <https://doi.org/10.1002/cem.1006>

MAZIVILA, S. J. et al. Fast Classification of Different Oils and Routes Used in Biodiesel Production Using Mid Infrared Spectroscopy and PLS2-DA. **Journal of the Brazilian Chemical Society**, v. 26, n. 4, p. 642–648, 2015a. <https://doi.org/10.5935/0103-5053.20150020>

MAZIVILA, S. J. et al. Discrimination of the type of biodiesel/diesel blend (B5) using mid-infrared spectroscopy and PLS-DA. **Fuel**, v. 142, p. 222–226, fev. 2015b. <https://doi.org/10.1016/j.fuel.2014.11.014>

MCMURRY, J. **Química Orgânica**. 9^a ed. ed. São Paulo: Cengage Learning, 2016.

MOÇAMBIQUE. **Boletim DA República: Estratégia do Sector de Energia**. CONSELHO DE MINISTROS. Maputo: Série-Número I, p. 20-22, 2009. Disponível em: <https://www.portaldogoverno.gov.mz/por/Governo/Legislacao/Boletins-da-Republica/Boletins-da-Republica-2009>

MOÇAMBIQUE. **Boletim Da República: Regulamento de Biocombustíveis e suas Misturas**. CONSELHO DE MINISTROS. Maputo: I Série-Número, p. 36-39, 2011. Disponível em: <http://extwprlegs1.fao.org/docs/pdf/moz119717.pdf>

MOFIJUR, M. et al. Effect of biodiesel from various feedstocks on combustion characteristics, engine durability and materials compatibility: A review. 2013. <https://doi.org/10.1016/j.rser.2013.07.051>

MONAKHOVA, Y. B.; DIEHL, B. W. K. Combining ^1H NMR spectroscopy and multivariate regression techniques to quantitatively determine falsification of porcine heparin with bovine species. **Journal of Pharmaceutical and Biomedical Analysis**, v. 115, p. 543–551, 10 nov. 2015. <https://doi.org/10.1016/j.jpba.2015.08.017>

MONTEIRO, M. R. et al. Determination of biodiesel blend levels in different diesel samples by ^1H NMR. **Fuel**, v. 88, n. 4, p. 691–696, 2009. <https://doi.org/10.1016/j.fuel.2008.10.010>

NAYLOR, R. L.; HIGGINS, M. M. The rise in global biodiesel production: Implications for food security. **Global Food Security**, v. 16, p. 75–84, mar. 2018. <https://doi.org/10.1016/j.gfs.2017.10.004>

PANDEY, V. C. et al. Jatropha curcas: A potential biofuel plant for sustainable environmental development. **Renewable and Sustainable Energy Reviews**, v. 16, n. 5, p. 2870–2883, jun. 2012. <https://doi.org/10.1016/j.rser.2012.02.004>

PATEL, R. L.; SANKHAVARA, C. D. Biodiesel production from Karanja oil and its use in diesel engine: A review. **Renewable and Sustainable Energy Reviews**, v. 71, p. 464–474, maio 2017. <https://doi.org/10.1016/j.rser.2016.12.075>

PAVIA, D. L. et al. **Introdução à Espectroscopia. Tradução da 4^a edição norte-americana.** 4^a ed. São Paulo: CENGAGE Learning, 2010.

PERS, T. H. et al. Prediction of fat oxidation capacity using ^1H -NMR and LC-MS lipid metabolomic data combined with phenotypic data. **Chemometrics and Intelligent Laboratory Systems**, v. 93, n. 1, p. 34–42, ago. 2008. <https://doi.org/10.1016/j.chemolab.2008.03.008>

PIRES DE OLIVEIRA, I.; CAIRES, A. R. L. Molecular arrangement in diesel/biodiesel blends: A Molecular Dynamics simulation analysis. **Renewable Energy**, v. 140, p. 203–211, set. 2019. <https://doi.org/10.1016/j.renene.2019.03.061>

PORTELA, N. A. et al. Quantification of biodiesel in petroleum diesel by ^1H NMR: Evaluation of univariate and multivariate approaches. **Fuel**, v. 166, p. 12–18, 2016. <https://doi.org/10.1016/j.fuel.2015.10.091>

QUINTELLA, C. M. et al. Cadeia do biodiesel da banCada à indústria: uma visão geral Com prospeCção de tarefas e oportunidades para p&d&i. **Química Nova**, v. 32, n. 3, p. 793–808, 2009. <https://doi.org/10.1590/S0100-40422009000300022>

RAMOS, L. P. et al. ESTERIFICAÇÃO E TRANSESTERIFICAÇÃO SIMULTÂNEAS DE ÓLEOS ÁCIDOS UTILIZANDO CARBOXILATOS LAMELARES DE ZINCO COMO CATALISADORES BIFUNCIONAIS. **Química Nova**, v. 38, n. 1, p. 46–54, 2015. <http://dx.doi.org/10.5935/0100-4042.20140274>

SAVORANI, F.; TOMASI, G.; ENGELSEN, S. B. icoshift: A versatile tool for the rapid alignment of 1D NMR spectra. **Journal of Magnetic Resonance**, v. 202, n. 2, p. 190–202, fev. 2010. <https://doi.org/10.1016/j.jmr.2009.11.012>

SERRANO-LOURIDO, D. et al. Classification and characterisation of Spanish red wines according to their appellation of origin based on chromatographic profiles and chemometric data analysis. **Food Chemistry**, v. 135, n. 3, p. 1425–1431, 2012.
<https://doi.org/10.1016/j.foodchem.2012.06.010>

SHI, T. et al. ¹H NMR combined with PLS for the rapid determination of squalene and sterols in vegetable oils. **Food Chemistry**, v. 287, p. 46–54, 30 jul. 2019.
<https://doi.org/10.1016/j.foodchem.2019.02.072>

SHIMAMOTO, G. G.; TUBINO, M. Quantification of methanol in biodiesel through ¹H nuclear magnetic resonance spectroscopy. **Fuel**, v. 175, p. 99–104, 2016.
<https://doi.org/10.1016/j.fuel.2016.02.019>

SILVA, A. F. T. et al. Statistical process control of cocrystallization processes: A comparison between OPLS and PLS. **International Journal of Pharmaceutics**, v. 520, n. 1–2, p. 29–38, 30 mar. 2017.
<https://doi.org/10.1016/j.ijpharm.2017.01.052>

SILVA, M. A. M. et al. Development and analytical validation of a multivariate calibration method for determination of amoxicillin in suspension formulations by near infrared spectroscopy. **Talanta**, v. 89, p. 342–351, 2012.
<https://doi.org/10.1016/j.talanta.2011.12.039>

SINGH, K.; KUMAR, S. P.; BLÜMICH, B. Monitoring the mechanism and kinetics of a transesterification reaction for the biodiesel production with low field ¹H NMR spectroscopy. **Fuel**, v. 243, p. 192–201, 1 maio 2019.
<https://doi.org/10.1016/j.fuel.2019.01.084>

SITOÉ, B. V. et al. 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**, v. 30, p. 1062–1070, 2016.
<https://doi.org/10.1021/acs.energyfuels.5b02489>

SITOÉ, B. V. et al. Monitoring of biodiesel content and adulterant presence in methyl and ethyl biodiesels of jatropha in blends with mineral diesel using MIR spectrometry and multivariate control charts. **Fuel**, v. 191, p. 290–299, 2017.
<https://doi.org/10.1016/j.fuel.2016.11.078>

SITOÉ, B. V. et al. Quantification of Jatropha methyl biodiesel in mixtures with diesel using mid-infrared spectrometry and interval variable selection methods. **Analytical Letters**, v. 53, n. 4, p. 589–605, 3 mar. 2020.
<https://doi.org/10.1080/00032719.2019.1659805>

SOUIHI, N. et al. OPLS in batch monitoring - Opens up new opportunities. **Analytica Chimica Acta**, v. 857, p. 28–38, 1 fev. 2015.
<https://doi.org/10.1016/j.aca.2014.12.003>

TENFEN WAZILEWSKI, W. et al. Study of the methyl crambe (*Crambe abyssinica* Hochst) and soybean biodiesel oxidative stability. **Industrial Crops and Products**, v. 43, p. 207–212, 2013. <https://doi.org/10.1016/j.indcrop.2012.07.046>

TRYGG, J.; WOLD, S. Orthogonal projections to latent structures (O-PLS). **Journal of Chemometrics**, v. 16, n. 3, p. 119–128, 2002. <https://doi.org/10.1002/cem.695>

VALDERRAMA, P.; BRAGA, J. W. B.; POPPI, R. J. Estado da arte de figuras de mérito em calibração multivariada. **Química Nova**, v. 32, n. 5, p. 1278–1287, 2009.
<https://doi.org/10.1590/S0100-40422009000500034>

VIEIRA, A. P. et al. Determination of physicochemical properties of petroleum using ^1H NMR spectroscopy combined with multivariate calibration. **Fuel**, v. 253, p. 320–326, 1 out. 2019. <https://doi.org/10.1016/j.fuel.2019.05.028>

WESTERHUIS, J. A.; DE JONG, S.; SMILDE, A. K. Direct orthogonal signal correction. **Chemometrics and Intelligent Laboratory Systems**, v. 56, n. 1, p. 13–25, abr. 2001.
[https://doi.org/10.1016/S0169-7439\(01\)00102-2](https://doi.org/10.1016/S0169-7439(01)00102-2)

WISE, B. M. et al. **PLS_Toolbox 4.0 for use with MATLABTM, Eigenvector Research, Inc., 3905 West Eaglerock Drive, Wenatchee, WA 98801. USA.** [s.l.: s.n.]. Disponível em: www.eigenvector.com. Acesso em: 3 jul. 2019.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. PLS-regression: a basic tool of chemometrics. **Chemometrics and Intelligent Laboratory Systems**, v. 58, n. 2, p. 109–130, out. 2001.
[https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)

XU, L. et al. Rapid discrimination of pork in Halal and non-Halal Chinese ham sausages by Fourier transform infrared (FTIR) spectroscopy and chemometrics. **Meat Science**, v. 92, n. 4, p. 506–510, dez. 2012.
<https://doi.org/10.1016/j.meatsci.2012.05.019>