



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

MARIA TERESA CARVALHO FERREIRA

USO DE FT-MIR E CALIBRAÇÃO MULTIVARIADA POR MCR-ALS e SVR  
NA DETERMINAÇÃO DO TEOR DE BIOQUEROSENE DE MACAÚBA E  
PALMISTE EM MISTURAS COM QUEROSENE DE AVIAÇÃO

UBERLÂNDIA

2021

MARIA TERESA CARVALHO FERREIRA

USO DE FT-MIR E CALIBRAÇÃO MULTIVARIADA POR MCR-ALS e SVR  
NA DETERMINAÇÃO DO TEOR DE BIOQUEROSENE DE MACAÚBA E  
PALMISTE EM MISTURAS COM QUEROSENE DE AVIAÇÃO

Dissertação apresentada ao Programa de  
Pós-graduação em Química do Instituto de  
Química da Universidade Federal de  
Uberlândia, como requisito para obtenção  
de Título de Mestre em Química

Área de Concentração: Química Analítica

Orientador: Prof. Dr. Waldomiro Borges Neto

UBERLÂNDIA

2021

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

F383 Ferreira, Maria Teresa Carvalho, 1987-  
2021 Uso do FT-MIR e calibração multivariada por MCR-ALS e SVR na determinação do teor de bioquerosene de macaúba e palmiste em misturas com querosene de aviação [recurso eletrônico] / Maria Teresa Carvalho Ferreira. - 2021.

Orientador: Waldomiro Borges Neto.  
Dissertação (Mestrado) - Universidade Federal de Uberlândia, Pós-graduação em Química.  
Modo de acesso: Internet.  
Disponível em: <http://doi.org/10.14393/ufu.di.2021.44>  
Inclui bibliografia.

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

CDU: 54

Bibliotecários responsáveis pela estrutura de acordo com o AACR2:

Gizele Cristine Nunes do Couto - CRB6/2091



## UNIVERSIDADE FEDERAL DE UBERLÂNDIA

Coordenação do Programa de Pós-Graduação em Química  
Av. João Naves de Ávila, 2121, Bloco 5I - 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:	Dissertação de Mestrado Acadêmico, 320, PPGQUI				
Data:	oito de janeiro de dois mil e vinte e um	Hora de início:	[09:00]	Hora de encerramento:	[12:10]
Matrícula do Discente:	11912QMI009				
Nome do Discente:	Maria Teresa Carvalho Ferreira				
Título do Trabalho:	Uso de FT-MIR e calibração multivariada por MCR-ALS e SVR na determinação do teor de bioquerosene de macaúba e palmiste em misturas com querosene de aviação				
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, Plataforma Google Meet, link: <https://meet.google.com/dxg-uxzb-fat>, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Química, assim composta: Professores Doutores: Regina Massako Takeuchi, da Universidade Federal de Uberlândia; Frederico Garcia Pinto, da Universidade Federal de Viçosa e Waldomiro Borges Neto, orientador da candidata.

Iniciando os trabalhos o presidente da mesa, Dr. Waldomiro Borges Neto, apresentou a Comissão Examinadora e a candidata, 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(às) 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):

Aprovada.

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

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



Documento assinado eletronicamente por **Waldomiro Borges Neto, Professor(a) do Magistério Superior**, em 08/01/2021, às 12:15, 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 08/01/2021, às 12:18, 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 **Regina Massako Takeuchi, Professor(a) do Magistério Superior**, em 21/01/2021, às 10:40, 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](https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0), informando o código verificador **2434690** e o código CRC **89CFE99E**.

---

*Aos meus pais, Ione e Márcio,  
às minhas irmãs Isabela e  
Daniela, ao meu irmão Marcinho,  
pelas várias vezes que suportaram  
minha ausência e por sempre  
acreditarem em mim. Sem o amor,  
carinho, incentivo, confiança e  
paciência de vocês nada disso  
seria possível*

*dedico*

## AGRADECIMENTOS

Agradeço a Deus por iluminar meu caminho e ter me dado sabedoria e sempre estar à frente de cada passo, guiando por bons e maus caminhos, onde esses maus caminhos foram essenciais para um crescimento como ser humano e profissional no qual me proporcionaram momentos magníficos para me dar condições suficientes de mais uma etapa concluída.

Ao meu orientador, Prof. Dr. Waldomiro Borges Neto, pela sua orientação, dedicação, paciência, preciosos ensinamentos. Um exemplo de pessoa e profissional. Tenho um orgulho enorme de ser sua orientanda.

Aos meus colegas do Laboratório de Quimiometria do Triângulo: Ademar Domingos Viagem Máquina, Baltazar Vasco Siteo, Edvando Souza Teles, José Eduardo Buiate, Eloíza Guimarães, Lucas Gustavo da Costa, Raquel Boaventura de Moraes, Tathiana Dienifer Godinho, Hery Mitsutake e Felipe Bachion de Santana, pelo apoio e a companhia, pelos momentos de trabalho e diversão que serão inesquecíveis.

Aos professores, funcionários e colegas do Instituto de Química.

A minha grande amiga Dayane, irmã que escolhi, por toda amizade, por me incentivar e sempre me mostrar o lado bom da vida, mas também sempre puxar minha orelha. Por em determinados momentos confiar mais em mim que eu mesma. Você é um Grande exemplo de determinação e coragem.

Ao meu amigo Arthur, por sempre me ajudar nesse mestrado.

Ao meu namorado Leandro, pelo incentivo, paciência e companheirismo, te gosto!

Enfim, a minha família pelo amor, compreensão, apoio e incentivo em todos os momentos da minha vida.

## RESUMO

O querosene de aviação, comercialmente conhecido como QAV é um combustível derivado de fontes de energias não renováveis usado principalmente no setor aéreo. De acordo com dados fornecidos pela International Air Transport Association – IATA, o uso deste combustível por este setor gera aproximadamente 2% das emissões de dióxido de carbono para a atmosfera. Essas emissões representam uma preocupação global no século XXI e podem aumentar com a expansão da indústria de aviação. Assim, torna-se necessário o uso de combustível ambientalmente sustentável como alternativa ao combustível fóssil. O bioquerosene de aviação (BioQAV) é um combustível derivado de fontes de origem natural podendo ser animal ou vegetal. Neste trabalho foi utilizado amostras de bioquerosene de macaúba e palmiste fornecidas pelo Laboratório de Biocombustíveis do Instituto de Química da Universidade Federal de Uberlândia, para a produção da mistura de bioquerosene/querosene e da análise por espectroscopia no infravermelho médio aliada aos métodos de Resolução de Curvas Multivariadas por Quadrados Mínimos Alternados (MCR-ALS) e Regressão por Vetores de Suporte (SVR) para determinar os seus teores em misturas com querosene na faixa de concentração de 1,00 a 70,00 (% v/v). Para os modelos de Regressão por Vetores de Suporte (SVR) das misturas de bioquerosene/querosene, os resultados para o bioquerosene de macaúba (BioQAVM) obtiveram valores de RMSEP = 0,81 (% v/v), RMSEC = 0,06 (% v/v), C = 997,23,  $\varepsilon$  = 0,064 e R = 0,99953, para o bioquerosene de palmiste (BioQAVP) foram obtidos o valores de RMSEP = 0,78 (% v/v), RMSEC = 0,20 (% v/v), C = 915,08,  $\varepsilon$  = 0,063 e R = 0,99993. Os resultados apontam o SVR como a melhor ferramenta para calibração multivariada aplicada a dados de bioquerosene obtidos através da espectroscopia na região do infravermelho médio, para quantificar de forma mais rápida, direta e com menor custo, os teores de bioquerosene de macaúba e palmiste nas amostras analisadas. Para os modelos MCR-ALS, a recuperação dos espectros é uma ferramenta que proporciona significado e interpretação química dos resultados. Pode-se observar na que o número ótimo de componentes principais para descrever os modelos foram 3 componentes, na prática, o posto de uma matriz corresponde ao número de espécies químicas presentes na mistura.

**Palavras-chave:** BioQAV. infravermelho médio. controle de qualidade. biocombustível. Quimiometria.

## ABSTRACT

Aviation kerosene, commercially known as QAV, is a fuel derived from non-renewable energy sources used mainly in the airline industry. According to data provided by International Air Transport Association - IATA, the use of this fuel by this sector generates approximately 2% of carbon dioxide emissions to the atmosphere. These emissions represent a global concern in the 21st century and may increase with the expansion of the aviation industry. Thus, it is necessary to use environmentally sustainable fuel as an alternative to fossil fuel. Aviation biokerosene (BioQAV) is a fuel derived from sources of natural origin and can be animal or vegetable. In this work, the oilseeds macaúba and palmiste were used, which are not used in human consumption for the production of biokerosene and the analysis by spectroscopy in the medium infrared combined with the methods of Resolution of Multivariate Curves by Alternating Minimum Squares (MCR-ALS) and Regression by Support Vectors (SVR) to determine its contents in mixtures with kerosene in the concentration range from 1.00 to 70.00 (% v / v). For Support Vector Regression (SVR) models of biokerosene / kerosene mixtures, the results for macaúba biokerosene (BioQAVM) obtained values of RMSEP = 0.81 (% v / v), RMSEC = 0.06 (% v / v), C = 997.23,  $\epsilon$  = 0.064 and R = 0.999953, for palm kernel biokerosene (BioQAVP) the values of RMSEP = 0.78 (% v / v), RMSEC = 0, 20 (% v / v), C = 915.08,  $\epsilon$  = 0.063 and R = 0.999993. The results indicate the SVR as the best tool for multivariate calibration applied to biokerosene data obtained through spectroscopy in the medium infrared region, in order to quantify more quickly, directly and at a lower cost, the biokerosene contents of macauba and palm kernel in the samples. analyzed. For the MCR-ALS models, the recovery of the spectra is a tool that provides meaning and chemical interpretation of the results. It can be seen that the optimal number of main components to describe the models were 3 components. Thus, the rank of the data matrix for each BioQAV sample is 3, that is, the spectra of the samples can be obtained through a linear combination of three individual spectra. In practice, the rank of a matrix corresponds to the number of chemical species present in the mixture.

**Keywords:** BioQAV. mid infrared. quality control. Biofuel. Chemometrics.

## LISTA DE FIGURAS

Figura 1: Comparação do ciclo de vida do CO <sub>2</sub> entre combustível fóssil e biocombustíveis. ....	17
Figura 2: Processo convencional de obtenção de querosene de aviação. ....	19
Figura 3: Classes de hidrocarbonetos no querosene de aviação fóssil. ....	20
Figura 4: Evolução da produção e importação de querosene de aviação no Brasil. ....	22
Figura 5: Principais matérias-primas utilizadas para produção de biodiesel no Brasil ..	24
Figura 6: A) Planta adulta da palmeira de macaúba, B) Cachos do fruto, C) Fruto da macaúba. ....	25
Figura 7: A) Planta adulta da Palmiste, B) Frutos no período de colheita, C) Frutos. ...	26
Figura 8: Processo de obtenção do bioquerosene. ....	28
Figura 9: Representação da análise por ATR horizontal. ....	33
Figura 10: Espectros originais da mistura de bioquerosene/querosene no teor de 1,00 a 70,00% (v/v) de macaúba. ....	34
Figura 11: Representação da construção da matriz <b>X</b> para calibração multivariada. ....	36
Figura 12: Representação de um conjunto de dados com hiperplano de separação ótima. ....	37
Figura 13: Hiperplano de separação e margem. ....	38
Figura 14: Ilustração de um problema de classificação binária linearmente inseparável. ....	39
Figura 15: Problema não linear separável em um problema linearmente separável. ....	41
Figura 16: Representações gráficas das funções de Kernel mais utilizadas no SVM. ...	42
Figura 17: Ilustração do processo de penalização da função de perda $\epsilon$ -insensível para SVM com Kernel linear. ....	44

Figura 18: Representação de decomposição do MCR para um sistema de dois componentes. ....	45
Figura 19: Perfis antes (cor cinza) e após (cor rosa) a aplicação das restrições de não negatividade, unimodalidade, sistema fechado e modelo pré definido.....	47
Figura 20: Restrição de correção feito pelo toolbox MCR-ALS GUI 2.0 .....	49
Figura 21: Espectros de MIR das misturas de bioquerosene/querosene de (a) macaúba e (b) palmiste de 1,00 a 70,00% (v/v) sobrepostos.....	51
Figura 22: Espectros de MIR corrigidos da mistura de bioquerosene/querosene de (a) macaúba e (b) palmiste.....	53
Figura 23: Espectros MIR dos bioquerosene puro de macaúba (vermelho), bioquerosene de palmiste (azul) e do QAV puro (verde). ....	54
Figura 24: Modelos de regressão por SVR (A) BioQAVM e (B)BioQAVP .....	55
Figura 25: SVD para determinação do número de componentes do sistema.....	58
Figura 26: Curvas de calibração para os bioquerosene: (a) macaúba e (b) palmiste.....	59
Figura 27: Espectros da mistura BioQAVM/QAV, (a) recuperado pelo MCR-ALS e (b) experimental .....	61
Figura 28: Espectros da mistura BioQAVP/QAV, (a) recuperado pelo MCR-ALS e (b) experimental .....	62
Figura 27: Regressão dos valores reais pelos previstos das misturas de bioquerosene de (a) macaúba e (b) palmiste. ....	66

## LISTA DE TABELAS

Tabela 1: Distribuição dos hidrocarbonetos na composição do querosene de aviação. .	20
Tabela 2: Região espectrais do infravermelho .....	32
Tabela 3: Valores dos parâmetros e dos erros para as amostras de BioQAV. ....	57
Tabela 4: Resultados dos parâmetros do modelo SVR para BioQAVM e BioQAVP....	57
Tabela 5: Valores de $R^2$ , LOF, RMSEP e RMSEC. ....	64
Tabela 7: Parâmetros de exatidão dos modelos PLS construídos.....	67
Tabela 8: Comparação entre os modelos PLS, SVR e MCR-ALS.....	68
Tabela 9: Resultados obtidos pelo teste F.....	68

## LISTA DE ABREVIATURAS E ACRÔNIMOS

<b>NBR-ABNT</b>	Norma Brasileira da Associação Brasileira de Normas Técnicas
<b>ANP</b>	Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
<b>ASTM</b>	Sociedade Americana de Testes e Materiais (do inglês, <i>American Society of Testing and Materials</i> )
<b>ATR</b>	Reflectância Total Atenuada (do inglês, <i>Attenuated Total Reflectance</i> )
<b>BioQAV</b>	Bioquerosene de Aviação
<b>BioQAVM</b>	Bioquerosene de Aviação de Macaúba
<b>BioQAVP</b>	Bioquerosene de Aviação de Palmiste
<b>FT-MIR</b>	Infravermelho Médio com Transformada de Fourier (do inglês, <i>Fourier Transform Mid Infrared</i> )
<b>FT-SPK</b>	Querosene parafínico sintético obtido pelo processo de Fischer-Tropsch (do inglês, <i>Fischer-Tropsch Synthesized Paraffinic Kerosene</i> )
<b>MCR-ALS</b>	Resolução de Curvas Multivariadas por Quadrados Mínimos Alternados, (do inglês, <i>Multivariate Curve Resolution-Alternating Least Squares</i> )
<b>MIR</b>	Infravermelho Médio (do inglês, <i>Mid Infrared</i> )
<b>PCA</b>	Análise de Componentes Principais (do inglês, <i>Principal Components Analysis</i> )
<b>RMSEC</b>	Erro Quadrático Médio de Calibração (do inglês, <i>Root Mean Square Error of Calibration</i> )
<b>RMSECV</b>	Erro Quadrático Médio de Validação Cruzada (do inglês, <i>Root Mean Square Error of Cross-Validation</i> )
<b>RMSEP</b>	Erro Quadrático Médio de Previsão (do inglês, <i>Root Mean Square Error of Prediction</i> )
<b>SVM</b>	Máquina de Vetores de Suporte (do inglês, <i>Support Vector Machines</i> )
<b>SVR</b>	Regressão por Vetores de Suporte (do inglês, <i>Support Vector Regression</i> )
<b>VL</b>	Variável Latente

# Sumário

<b>1</b>	<b>Introdução</b> .....	16
<b>2</b>	<b>Objetivos</b> .....	19
<b>2.1</b>	<b>Objetivos Específicos</b> .....	19
<b>3</b>	<b>Revisão Bibliográfica</b> .....	19
<b>3.1</b>	<b>Querosene de Aviação</b> .....	19
<b>3.2</b>	<b>Matéria-prima para Produção de Bioquerosene</b> .....	22
<b>3.3</b>	<b>Bioquerosene de Macaúba</b> .....	24
<b>3.4</b>	<b>Bioquerosene de Palmiste</b> .....	26
<b>3.5</b>	<b>Métodos para Produção de Bioquerosene</b> .....	27
3.5.1	Processos químicos .....	28
3.5.2	Processos Bioquímicos.....	29
3.5.3	Processos Termoquímicos .....	29
<b>3.6</b>	<b>Exigências de qualidade do Bioquerosene</b> .....	30
<b>4</b>	<b>Espectroscopia na região do Infravermelho</b> .....	31
4.1	Reflectância Total Alternada.....	33
<b>5</b>	<b>Métodos Quimiométricos</b> .....	34
<b>5.1</b>	<b>Máquina de Vetores de Suporte (SVM)</b> .....	36
5.1.2	Margem Suave.....	39
5.1.3	Função <i>Kernel</i> .....	40
5.1.4	Regressão por Vetores de Suporte (SVR).....	42
5.1.5	$\epsilon$ -tubo e a função de perda $\epsilon$ -insensível .....	43
<b>5.2</b>	<b>Resolução Multivariada de Curvas com Mínimos Quadrados Alternados(MCR-ALS)</b> 44	
5.2.1	Imposição de Restrições .....	46
5.2.2	MCR-ALS com restrição de correlação .....	48
<b>6</b>	<b>Procedimento Experimental</b> .....	50
6.1	Preparo das misturas de bioquerosene/querosene.....	50
6.2	Obtenção dos espectros de MIR .....	50
6.3	Pré-Tratamento dos dados .....	51

<b>7</b>	<b>Resultados e Discussões</b> .....	51
7.1	Espectros MIR de bioquerosene e querosene.....	51
7.2	Modelos SVR .....	55
7.3	Modelos MCR-ALS .....	58
7.4	Modelo PLS e comparação entre SVR .....	65
<b>8</b>	<b>Conclusão</b> .....	69
<b>9</b>	<b>Referências</b> .....	71

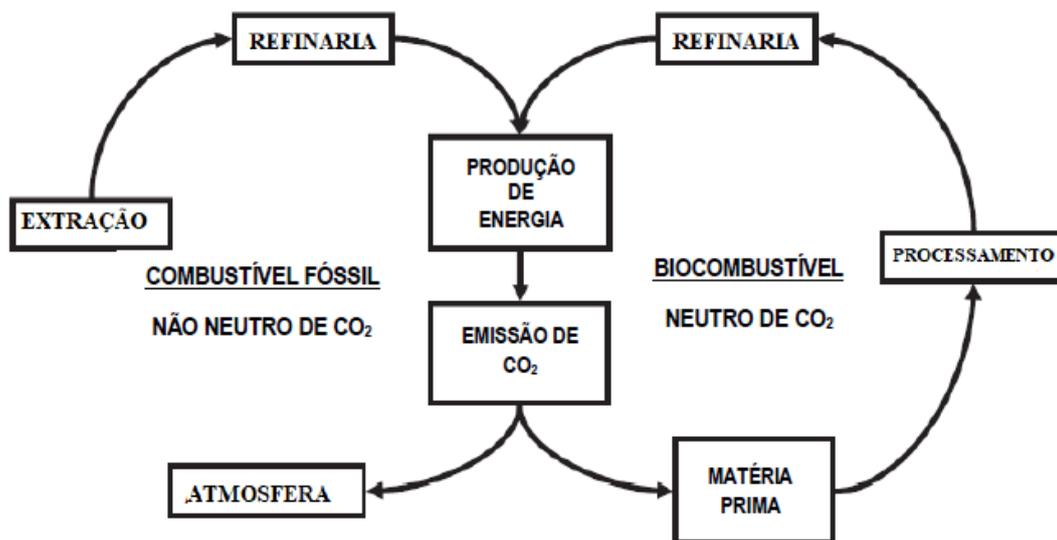
## 1 Introdução

Nos últimos anos, o interesse pela produção de biocombustíveis tem ganhado destaque no cenário da aviação mundial, e conseqüentemente impulsionando a busca por alternativas mais sustentáveis visando a diminuição das emissões dos gases do efeito estufa (GEE) na atmosfera (GELLER, 2003). Nesse sentido, as companhias aéreas, fabricantes de aeronaves, empresas químicas e os governos estão agindo para buscar atender às necessidades de abastecimento (JANSEN et al., 2013).

De acordo com dados fornecidos pela International Air Transport Association – IATA, estima-se que este setor gera aproximadamente 2% das emissões de dióxido de carbono (CO<sub>2</sub>) para a atmosfera, podendo atingir um nível de 3% até 2030. Essas emissões representam uma preocupação global no século XXI podem aumentar com a expansão da indústria de aviação. Desta maneira, uma das estratégias para atingir as metas de redução inclui-se o uso de combustível ambientalmente sustentável como alternativa ao combustível fóssil (CORTEZ et al., 2013; GELLER, 2003).

Neste contexto, destacam-se os biocombustíveis definidos como não concorrente com a produção de alimentos, além de não contribuírem com o desmatamento são produzidos de forma sustentável no qual reduzem as emissões de CO<sub>2</sub> em relação aos combustíveis fósseis (THEGARID et al., 2014). Do ponto de vista do ciclo do CO<sub>2</sub>, o dióxido de carbono absorvido pelas plantas durante o crescimento da biomassa é equivalente à quantidade produzida quando o combustível é queimado num motor de combustão, quando será devolvido para a atmosfera. Esse processo permite que o biocombustível receba a condição de CO<sub>2</sub> neutro durante o seu ciclo de vida (CHIONG et al., 2018; LLAMAS et al., 2012). A Figura 1 compara o ciclo de emissão de CO<sub>2</sub> entre combustíveis fósseis e biocombustíveis.

Figura 1: Comparação do ciclo de vida do CO<sub>2</sub> entre combustível fóssil e biocombustíveis.



Fonte: Adaptado de (CHIONG et al., 2018).

No entanto, as emissões de gases produzidos durante a produção dos biocombustíveis, a partir do equipamento necessário para cultivar a colheita, o transporte de matérias-primas, refino do combustível, dentre outros, devem ser considerados. Quando estes elementos são contabilizados, espera-se, ainda uma redução das emissões globais de dióxido de carbono de até 80% em relação aos combustíveis fósseis. Além disso, os biocombustíveis possuem quantidades menores de impurezas (tais como enxofre), o que permite ainda uma maior redução na emissão de dióxido de enxofre e de fuligem (ENVIRO. AERO, 2011).

No mercado Internacional o bioquerosene de aviação é definido pela Norma D7566 da American Society for Testing and Materials International (ASTM *International*) que contém as especificações padrões para biocombustíveis (ANP- AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCMBUSTÍVEIS BRASIL, 2013; DOS SANTOS, 2015). No Brasil, a regulamentação fica por conta da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis na Lei nº 12.490/2011 e, também na resolução da ANP, nº 63 de 5.12.2014, como substância derivada de biomassa renovável que pode ser usada no consumo em turbinas de aeronaves (ANAC- AGÊNCIA NACIONAL DE

AVIAÇÃO CIVIL, 2019; ANP- AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS BRASIL, 2014).

Os substratos de diferentes espécies de oleaginosas no Brasil, possuem potencial para serem utilizadas como matérias primas na produção de bioquerosene, contudo vem se destacando oleaginosas que em sua composição possuem elevados teores de ácidos graxos cujo tamanho da cadeia carbônica seja semelhante à faixa do querosene comercial (LLAMAS et al., 2012; UBRABIO- UNIÃO BRASILEIRA DO BIODIESEL E BIOQUEROSENE, 2012).

Nesse sentido, as possibilidades de matérias-primas no Brasil são inúmeras, principalmente devido ao grande potencial de cultivo no solo. Os óleos de Macaúba e de Palmiste apresentam sua composição próxima a essa faixa do querosene comercial, sendo assim uma possível fonte de matéria-prima para obtenção de bioquerosene ( GARCIA, 2006). Estes fatos ditaram a escolha destes dois tipos de óleos (Macaúba e Palmiste) como objeto de estudo deste trabalho.

Diante desse contexto, faz-se necessário o desenvolvimento de metodologias que forneçam respostas rápidas e eficientes para monitorar o teor do bioquerosene de Macaúba e Palmiste em mistura como querosene de aviação. No entanto, a espectroscopia no infravermelho médio é uma das técnicas espectroscópicas vibracionais que tem se mostrado promissoras em análises químicas, por ser um método de baixo custo, que proporciona celeridade nas análises, por ser de natureza não destrutiva, não exigindo o tratamento de amostras, permitindo, assim, a análise *in situ* com o auxílio de equipamentos portáteis (BORIN; POPPI, 2004; PIMENTEL et al., 2006; KAROUI; DOWNEY; BLECKER, 2010).

## 2 Objetivos

O projeto tem como objetivo geral, desenvolver metodologias analíticas para determinar o teor de bioquerosene de Macaúba e Palmiste em misturas com o querosene de aviação usando espectroscopia no infravermelho médio e calibração multivariada por MCR-ALS e SVR que possam ser implementadas em análises de rotina e controle de qualidade desses combustíveis.

### 2.1 Objetivos Específicos

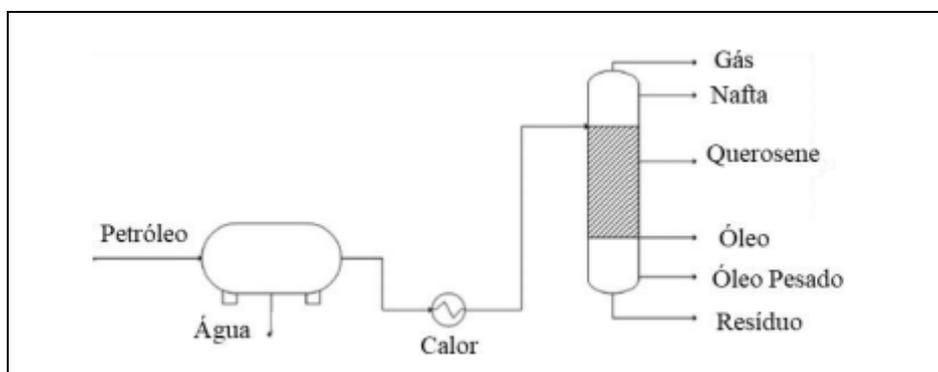
- ✓ Construir modelos de MCR-ALS e SVR para determinar o teor de bioquerosene de Macaúba e Palmiste em misturas com querosene;
- ✓ Comparar os modelos construídos com o modelo PLS.

## 3 Revisão Bibliográfica

### 3.1 Querosene de Aviação

O transporte aéreo sofreu um grande marco, após a Segunda Guerra Mundial houve uma intensificação do uso de aviões com o aperfeiçoamento das turbinas. Um novo combustível, querosene de aviação substituiu o uso da gasolina por apresentar vantagens como excelente relação peso/potência e elevado empuxo. O querosene de aviação é um líquido oleoso de coloração amarelo pálido ou incolor, com odor característico, obtido a partir da destilação do petróleo bruto, como ilustrado na Figura 2 (MONTEIRO, 2020).

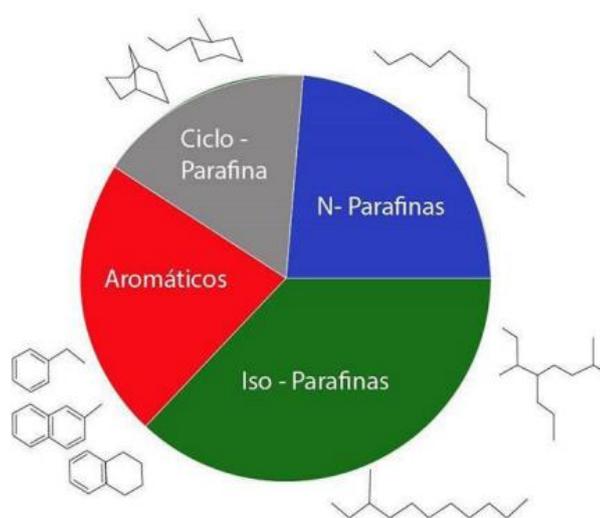
Figura 2: Processo convencional de obtenção de querosene de aviação.



Fonte: adaptado (LIU; YAN; CHEN, 2013).

É composto por uma mistura de hidrocarbonetos de petróleo, destacando-se as isoparafinas, n-parafinas, ciclo-parafina, aromáticos e olefinas como apresentado na Figura 3. Entre 75 a 85% do volume é composto por parafinas (iso- e ciclo-parafinas) e o restante por aromáticos (menos de 25%) e aletinas (1%) (BROOKS et al., 2016). As moléculas apresentam cadeias entre 8 a 16 átomos de carbono e possui faixa de destilação ASTM D-86 compreendida entre 130 a 300 °C.

Figura 3: Classes de hidrocarbonetos no querosene de aviação fóssil.



Fonte: adaptado (BROOKS et al., 2016).

Na Tabela 1, observa-se uma distribuição dos percentuais de cada hidrocarboneto, os compostos sulfonados, oxigenados e nitrogenados presentes no querosene dependem da origem do petróleo e dos processos de tratamento empregados (SPEIGHT; EL-GENDY, 2018).

Tabela 1: Distribuição dos hidrocarbonetos na composição do querosene de aviação.

<b>Tipos de Hidrocarbonetos</b>	<b>% Volume</b>
Isoparafina	40-85
Parafinas	33-61
Naftênicos	33-45
Aromáticos	12-25
Olefinas	0,5-5

O querosene de aviação, também conhecido como QAV, apresenta algumas propriedades, como forte poder de solvência, taxa lenta de evaporação, baixo ponto de congelamento, além de ser insolúvel em água. Para que esse derivado de petróleo apresente características adequadas à geração de energia para motores de turbina, diversos critérios físico-químicos são requeridos durante a sua produção, que incluem desde fluidez (escoamento), estabilidade (estocagem) até a adequada combustão para esses motores (BRASIL, 2013). Há diversos tipos de querosene de aviação comercializados no mundo. As diferenças se dão principalmente pela região de utilização, ponto de fulgor e ponto de congelamento havendo adequações do produto às questões de segurança e necessidades específicas, como por exemplo, uso militar (CGEE, 2010).

A expressiva evolução da indústria aeronáutica determina atualizações sistemáticas nas especificações dos combustíveis de aviação, caracterizando-as como as mais dinâmicas em relação a todos os produtos derivados de petróleo. As exigências de segurança são fatores importantes para a produção, controle e manuseio do QAV, uma vez que desvios de qualidade podem resultar em acidentes aéreos. A padronização mundial destas especificações é interessante para todos os países, de forma a garantir a segurança e facilitar as negociações de mercado (ANP- AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (BRASIL), 2013). A especificação do querosene de aviação no Brasil é determinada pela Resolução ANP nº 37/2009.

No Brasil possui um crescimento do consumo de QAV de forma superior à sua capacidade de produção, e o conseqüentemente o aumento de emissões dos gases do efeito estufa. Observando na Figura 4, no ano de 2018 foram produzidos 6,3 milhões m<sup>3</sup>, um aumento de 65% em uma década. Entretanto, 2019 foi marcado por uma queda de 4,8% em relação ao consumo do ano anterior, produzindo cerca de 6 milhões. Na Figura 4, encontra-se a evolução nacional de produção e importação de QAV (MONTEIRO, 2020).

Figura 4: Evolução da produção e importação de querosene de aviação no Brasil.



Fonte: (ANP,2020).

Diante do contexto, uma alternativa viável para se alcançar as metas e reduzir a preocupação com aumento das emissões é a produção de biocombustível de origem renovável com especificações semelhantes ao derivado de petróleo. Nesse cenário o Brasil se destaca pelo seu potencial de produção de biomassa devido a disponibilidade territorial e condições climáticas adequadas.

### 3.2 Matéria-prima para Produção de Bioquerosene

Os novos combustíveis para a aviação visam melhorar o balanço de emissões de gases do efeito estufa (GEE) por meio do seu ciclo de vida e devem ter uma série de critérios analisados sobre seus efeitos indiretos, de forma que atendam aos requisitos básicos ambientais, desde a lavoura (matéria-prima) até sua utilização final (“INTERNATIONAL CIVIL AVIATION ORGANIZATION”, 2010).

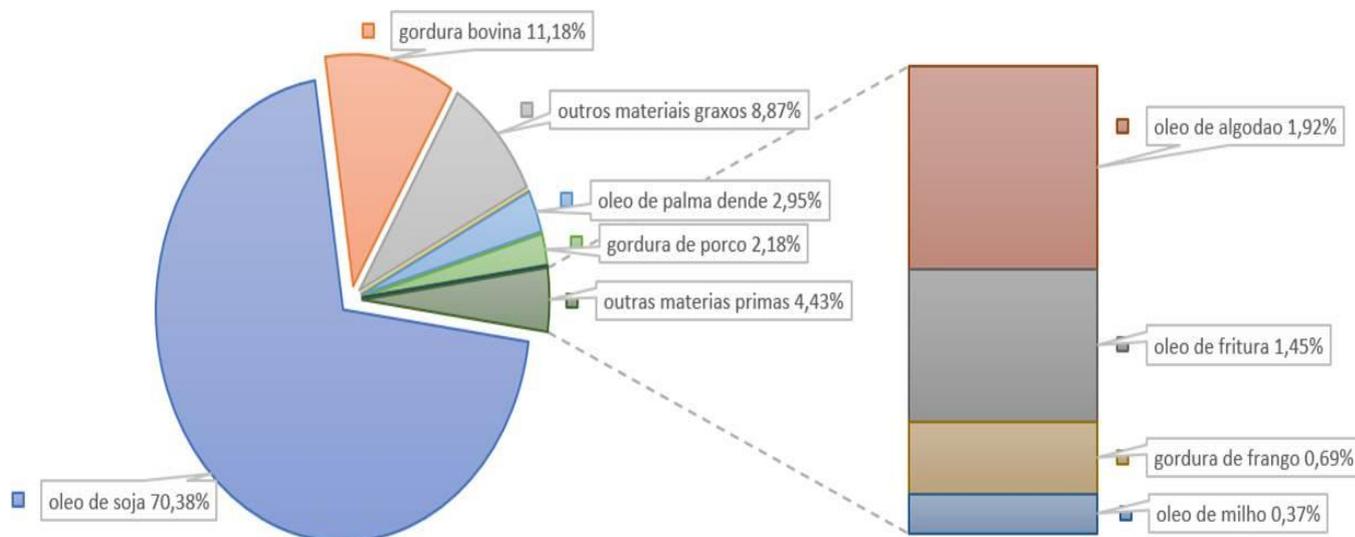
Entre os requisitos básicos da matéria-prima e de seu cultivo, estão: não concorrer com o setor de alimentos; ser cultivada em terras não utilizadas para produção de alimentos e terras marginais; não interferir nos ecossistemas naturais e ser produzida de forma que o solo e a água não sejam sobrecarregados; não necessitar de grandes quantidades de insumos agrícolas; proporcionar redução de emissões e conteúdo energético maior ou igual comparados ao combustível tradicional; não ameaçar a biodiversidade; e agregar valor socioeconômico às comunidades locais (VELÁZQUEZ; KUBOTANI; VELÁZQUEZ, 2012).

O único recurso natural de carbono renovável conhecido que possui porte suficiente para substituir os combustíveis fósseis é a biomassa. Dentro das matérias-primas incluídas como biomassa temos todos os organismos baseados em água e terra, vegetações e árvores, ou biomassa virgem; e todos os resíduos sólidos urbanos, biosólidos provenientes de esgotos, resíduos animais, florestais, agrícolas e industriais. A biomassa é renovável no sentido de que somente um curto período de tempo é necessário para sua formação, ao contrário dos combustíveis fósseis, que precisam de milhões de anos para sua formação. Além disso, estima-se que o carbono de biomassa terrestre permanente do mundo, ou seja, aquele acima da superfície disponível para uso energético, é aproximadamente 100 vezes maior que o consumo de energia mundial. Esta biomassa está concentrada principalmente sob a forma de biomassa florestal, contendo cerca de 80 a 90% do carbono de biomassa total de superfície (KLASS, 2004).

No Brasil, diferentes espécies de oleaginosas possuem potencial para serem utilizadas como matérias-primas na produção de bioquerosene, contudo vem se destacando oleaginosas cujos os substratos precursores dos ésteres metílicos ou etílicos, cujo tamanho da cadeia parafínica se assemelha à faixa de destilados equivalentes ao querosene de aviação fóssil, compreendem as frações de óleos vegetais ou de gorduras animais com cadeia linear ou ramificada que contenham entre 8 e 16 carbonos (LLAMAS et al., 2012).

Na Figura 5, são mostradas as principais matérias-primas utilizadas na produção de biodiesel no Brasil segundo dados de boletim mensal da ANP (Agência Nacional do Petróleo) de março de 2019. A mais utilizada atualmente ainda é o óleo de soja com 70,39%, isso ocorre devido às extensas áreas de cultivo nas regiões Centro-Oeste, Nordeste e Sul do Brasil. Em seguida, a gordura Bovina com 11,18% devido ao seu baixo custo e por conferir ao biocombustível uma boa estabilidade oxidativa. (ANP, 2019; BERGMANN et al., 2013; RINCÓN, JARAMILLO, CARDONA, 2014).

Figura 5: Principais matérias-primas utilizadas para produção de biodiesel no Brasil



Fonte: Adaptada (ANP,2019).

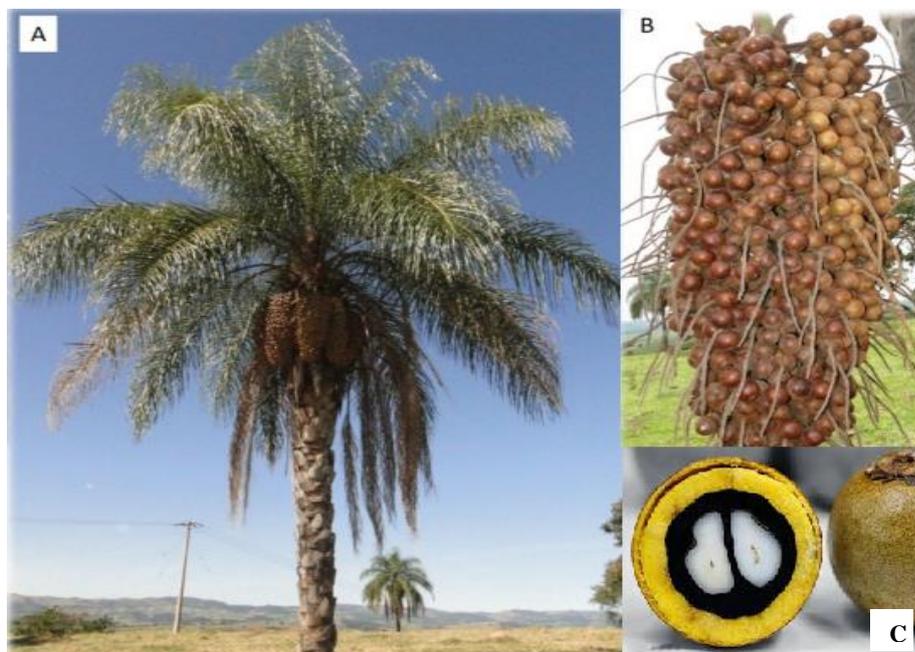
O óleo de soja é a principal matéria-prima para a produção de biodiesel no Brasil, por apresentar razões principalmente econômicas, mas também tradicionais. Essa oleaginosa possui sua cadeia produtiva bem estruturada no país, onde é uma cultura adaptada a todo o território nacional, sendo seu ciclo curto, de 4 a 5 meses. Seu óleo pode ser utilizado tanto para consumo humano, quanto para produzir biodiesel ou para usos na indústria química, além de produzir o farelo protéico utilizado em rações para animais produtores de carne (MAPA, 2016). Com desvantagem por ser competitivo com o mercado de alimentos, além de representar cerca de 70% do custo de produção do biodiesel, o que torna necessário algum subsídio, como por exemplo, isenção de imposto, assim a busca por outras matérias-primas tem se destacado (GONTIJO, 2016).

### 3.3 Bioquerosene de Macaúba

A macaúba é uma palmeira que apresenta nome científico *Acrocomia Aculeata*, pertencente à família *Arecaceae*. Uma planta rústica nativa muito comum em grande parte do Brasil que pode alcançar até 25 metros de altura e possui espinhos longos e pontiagudos na base de suas folhas ou ao longo do caule e suas flores se agrupam em

cachos (MORAES, 2020). Na Figura 6 pode-se observar as fotos da planta, cachos e frutos da macaúba.

Figura 6: A) Planta adulta da palmeira de macaúba, B) Cachos do fruto, C) Fruto da macaúba.



Fonte: Adaptado de (COLOMBO et al., 2018).

É uma palmeira oleaginosa com significativo potencial para a produtividade de óleo e não explorada para fins alimentícios, com capacidade de produção de 6,2 ton/ha. A planta se destaca frente a outras oleaginosas devido ao seu grande potencial na produção de óleo, capaz de produzir até seis vezes a mais que a soja (LANES; MOTOIKE; COSTA, 2014; MELO, 2012). Facilmente adaptável e já muito conhecida por agricultores. É comum ver os frutos serem comidos pelo gado ou por pessoas interessadas na amêndoa da planta, que tem cheiro doce e gosto de coco (MEDEIROS, 2020).

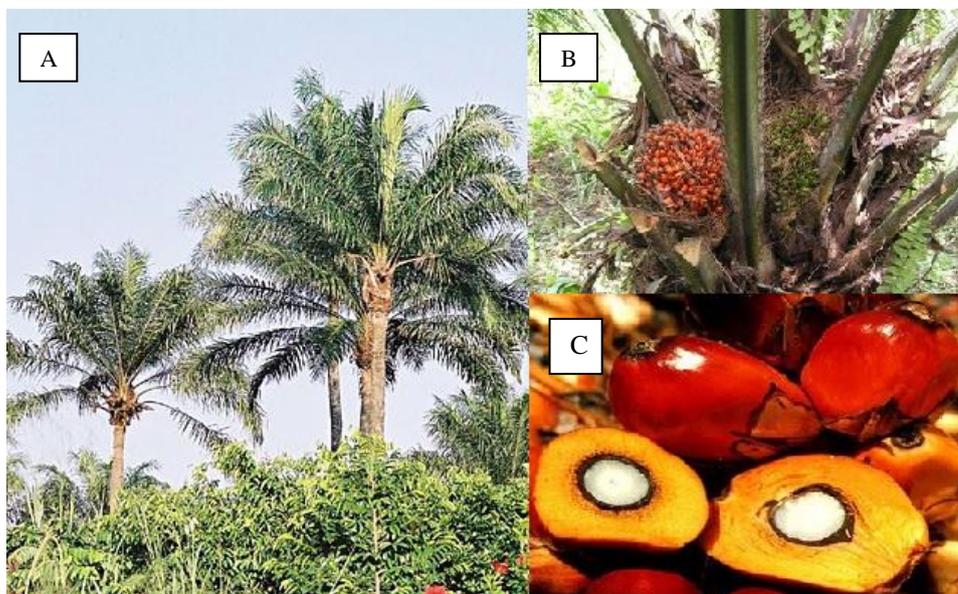
A macaúba produz dois óleos de qualidades diferentes. O da amêndoa, bem no centro do fruto e em menor quantidade conforme a Figura 6C. Já o óleo para biocombustível é extraído da parte mole, o mesocarpo (a parte amarela da Figura 6C), que precisa ser processado de maneira correta para melhor aproveitamento do óleo. A amêndoa da macaúba possui alto potencial para a produção de óleo visto que a maior quantidade de óleo se concentra nessa área, de 40 a 50%, enquanto a polpa apresenta

apenas 27% todo óleo extraído. Além do mais, a umidade na amêndoa apresenta-se em uma menor porcentagem (DA SILVA, 2019).

### 3.4 Bioquerosene de Palmiste

O palmiste é uma palmeira com nome científico *Elaeisguineensis*, de origem do continente africano sendo encontrada na Costa Ocidental da África (Golfo da Guiné), se destaca devido ao grande potencial para produção de óleo (MARZULLO, 2007). Esta palmácea pode ser cultivada em terras degradadas, sendo seu cultivo uma solução apropriada para produção de biocombustíveis no Brasil. É uma cultura que pode crescer até uma altura de 20 metros quando atinge o fim de sua vida econômica, por volta de 25 anos (FAHRENKROG; POKAR; NOHL, 2017). Na Figura 7 observam-se as fotos da planta e frutos da Palmiste.23

Figura 7: A) Planta adulta da Palmiste, B) Frutos no período de colheita, C) Frutos.



Fonte: Adaptado de (“Dendê (*Elaeisguineensis*)”, 2013).

A produção do óleo de palma tem se tornado bastante atrativa em virtude do seu baixo custo de produção e de elevada produtividade de óleo por hectare (EQUATORIANO, 2006). Podendo alcançar em uma mesma área até 10 vezes mais produtividade de óleo quando comparada à cultura de soja. Mostrando-se uma interessante possibilidade na produção de combustível aéreo por ser estável à oxidação devido à presença de antioxidante naturais e rico em ácido láurico (C12) (da SILVA, 2019; SEMI-BATELADA, 2009)

O óleo de palmiste é extraído da semente do fruto da palma, sendo que a sua composição difere significativamente da composição do óleo de palma, apresentando aproximadamente 80% de ácidos graxos saturados e predominantemente ácidos graxos de cadeia curta, destacando-se assim como uma matéria-prima viável à produção de bioquerosene. Do fruto é possível a extração de dois tipos de óleos: o óleo dendê, também conhecido como óleo de palma, extraído da parte externa do fruto (polpa), nomeado mesocarpo, e o óleo de palmiste, obtido da amêndoa localizada no interior do fruto (MARZULLO, 2007).<sup>24</sup>

### **3.5 Métodos para Produção de Bioquerosene**

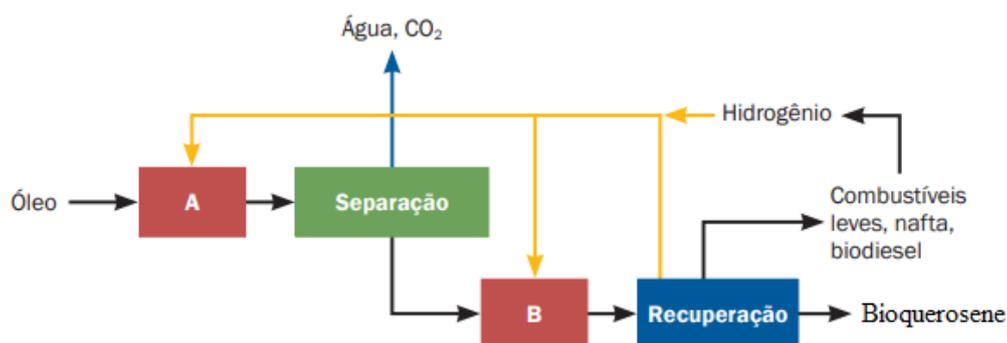
Existem vários métodos para transformar matéria-prima em bioquerosene. Dentre eles, se destacam três categorias: processos químicos para lipídeos, processos bioquímicos e termoquímicos para biomassa. Os processos químicos para lipídeos e processos térmicos já possuem rotas tecnológicas e receberam certificação de aprovação pela ASTM, em uma mistura de até 50% para uso em combustível de aviação (DOS SANTOS, 2015; JANSEN, 2013; LIU; YAN; CHEN, 2013).

O bioquerosene pode ser obtido através de algumas rotas de produção, dentre estas estão: processos químicos, utilizando oleaginosas vegetais ou gorduras animais que passam por processos de transesterificação; processos bioquímicos, através de leveduras ou bactérias modificadas que metabolizam matérias açucaradas e produzem uma gama de hidrocarbonetos; processos termoquímicos, produzido por processos de craqueamento catalítico e gaseificação e síntese catalítica (CHIARAMONTI et al., 2014).

### 3.5.1 Processos químicos

Consiste num processo de transesterificação, onde os triglicerídeos são convertidos em ésteres metílicos ou etílicos e glicerina através de uma reação na presença de álcool (metanol ou etanol) e catalisador como bases fortes, ácidos ou enzimas (SCHUCHARDT; SERCHELI; MATHEUS, 1998). Considerando as especificações exigidas, é necessário que haja uma etapa posterior de separação e purificação dos ésteres, obtendo as frações mais adequadas para utilização em turbinas (PARENTE, 2009). Esta etapa de purificação ocorre de acordo com o observado na Figura 8.

Figura 8: Processo de obtenção do bioquerosene.



Fonte: (CGEE, 2010).

Na etapa identificada na Figura 8 como A, ocorre a remoção de oxigênio através da adição de hidrogênio (descarbonilação). Em seguida, durante a etapa B o hidrocarboneto proveniente da etapa A é isomerado e quebrado para que seu número de carbono seja reduzido (hidrogenação), ficando com estrutura semelhante ao QAV (querosene fóssil) em relação a propriedades de ponto de congelamento e centelha.

Este processo (554\_HIDROCARBONETOS) foi desenvolvida por pesquisadores da Universidade Estadual de Campinas (UNICAMP) no ano de 2008 (FILHO, R.M., MACIEL, M. R. W., BATISTELLA, C. B., SILVA, 2008). Os biocombustíveis

produzidos podem ser utilizados em *blends* com o querosene de origem fóssil ou puro, posterior à adaptação dos motores.

### 3.5.2 Processos Bioquímicos

Esta rota utiliza matérias-primas constituídas por carboidratos (açúcares), onde há a conversão desses em hidrocarbonetos através do uso de leveduras *Saccharomyces Cerevisiae* geneticamente modificadas. Os microrganismos consomem os açúcares do meio e produzem isoprenóides, como farnesene e pinene, ao invés do etanol. O produto dessa conversão passará por processo de hidrogenação, obtendo-se o bioquerosene (CGEE, 2010).

### 3.5.3 Processos Termoquímicos

Através da rota termoquímica o bioquerosene pode ser produzido a partir de craqueamento catalítico ou gaseificação e síntese catalítica.

- Craqueamento Catalítico: No craqueamento, processo de decomposição térmica do óleo vegetal na presença de hidrogênio e catalisador, eliminam-se os produtos oxigenados obtendo-se uma mistura de hidrocarbonetos, que passa por uma destilação, gerando frações semelhantes ao querosene de aviação (OLIVEIRA; SUAREZ; SANTOS, 2008).

- Gaseificação e síntese catalítica: Nessa rota, através de temperaturas elevadas (aproximadamente 1000 °C), ocorre à gaseificação da biomassa sólida, convertendo-a quase completamente em uma mistura de gases como: CO<sub>2</sub>, CO, H<sub>2</sub> e vapores de água. Após a purificação desses gases, estes podem ser utilizados para a síntese de hidrocarbonetos líquidos pelo processo Fischer-Tropsch, na presença de catalisadores como cobalto e ferro. Conforme a concepção adotada e os fluxos dos processos, é possível que se obtenha hidrocarbonetos similares ao querosene de aviação (DATAR et al., 2004).

### 3.6 Exigências de qualidade do Bioquerosene

Para a especificação do querosene de aviação alternativo, são adotadas as normas ASTM D7566, que regulamentam o querosene alternativo: (i) SPK-HEFA, obtido pela hidrogenação e desoxigenação de ésteres e ácidos graxos e ácidos livres, com o objetivo de remover essencialmente o oxigênio; (ii) querosene alternativo SPK-FT, obtido da reação de Fischer-Tropsch, usando catalisadores baseados em ferro ou cobalto, (iii) querosene de aviação alternativo de isoparafinas sintéticas (SIP), a partir dos açúcares fermentados hidroprocessados, produzido por hidroprocessamento e fracionamento farneseno, derivados da fermentação de açúcares.

No Brasil, o órgão regulador responsável pela especificação do QAV é a Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP). A especificação do querosene de aviação no Brasil é determinada pela Resolução ANP nº 37/2009.

As exigências de qualidade do QAV para uso em turbinas aeronáuticas são as seguintes:

- Proporcionar máxima autonomia de voo;
- Vaporizar-se adequadamente no interior da câmara de combustão da turbina, proporcionando “chama limpa”, minimizando assim a formação de fuligem;
- Proporcionar partidas fáceis e seguras, e ter facilidade de reacendimento;
- Minimizar a formação de resíduos e cinzas por combustão;
- Escoar facilmente em baixa temperatura;
- Ser estável química e termicamente;
- Não ser corrosivo aos materiais da turbina;
- Minimizar a tendência à solubilização de água;
- Ter aspecto límpido, indicando ausência de sedimentos e de alteração de cor;
- Não apresentar água livre, evitando o desenvolvimento de microrganismos e a obstrução de filtros; oferecer segurança no manuseio e na estocagem.

A especificação brasileira do QAV é determinada pela Agência Nacional de Petróleo (ANP), resolução 37 de 2009 (BRASIL, 2009), sendo totalmente compatível com o JET A-1 (FARAH, 2012). Resumidamente, as principais características físico-químicas dos querosenes de aviação, particularmente para os combustíveis de aviação comercial envolvem os seguintes aspectos (GONÇALVES, 2011; BRASIL, 2009):

- Composição e caracterização química (teor de aromáticos, teor de enxofre, índice de acidez);
- Propriedades de escoamento, como viscosidade cinemática, densidade, pressão de vapor e ponto de congelamento;
- Propriedades térmicas, como conteúdo energético do combustível em relação a sua massa específica, ponto de fulgor e estabilidade térmica (formação de goma).

#### 4 Espectroscopia na região do Infravermelho

A espectroscopia compreende o estudo da interação da radiação eletromagnética com a matéria, tendo como um de seus principais objetivos a determinação dos níveis de energia e transições de espécies atômicas e moleculares (SKOOG; HOLLER; NIEMAN, 2002). O espectro na região do Infravermelho Médio (MIR) é a faixa de frequência responsável pelas transições vibracionais e rotacionais das ligações das moléculas, faixa esta que compreende a região de  $4.000$  a  $200\text{ cm}^{-1}$  (MÜLLER et al., 2011). Dentro desta região, existe uma faixa espectral específica de  $1200$  a  $700\text{ cm}^{-1}$ , conhecida como região de impressão digital ou *fingerprint*, em que pequenas diferenças na estrutura e na constituição de uma molécula resultam em mudanças significativas nas intensidades ou perfis das bandas de absorção e por esse fato são usados como parâmetro de identificação ou distinção entre compostos (BARBOSA, 2008).

Com o aprimoramento da técnica, a Espectroscopia no Infravermelho Médio com a Transformada de Fourier (do inglês, *Fourier Transform Mid Infrared*) foi desenvolvida com a finalidade de superar algumas limitações encontradas nos instrumentos dispersivos como, por exemplo, o longo processo de varredura. O diferencial é a formação da onda evanescente no seu interior, gerada a parte da reflexão da luz incidida e absorção por parte da amostra na superfície, que em seguida sofre uma atenuação do feixe, passado pelo detector assim gerando o espectro (MÜLLER et al., 2011).

As análises espectroscópicas são baseadas na medida da quantidade radiação produzida ou absorvida por moléculas ou espécies químicas de interesse e podem ser classificadas de acordo com a região do espectro eletromagnético envolvida na medida.

As transições eletrônicas são situadas na região do ultravioleta ou visível, as vibracionais na região do infravermelho e as rotacionais na região de micro-ondas e, em casos particulares, também na região no Infravermelho distante. Do ponto de vista da aplicação e instrumentação, o espectro infravermelho é dividido em radiação no Infravermelho próximo (NIR), médio (MIR) e distante (FIR). A Tabela 2 apresenta os intervalos aproximados correspondente para cada região.

Tabela 2: Região espectrais do infravermelho

<b>Região</b>	<b>Número de onda (<math>\nu</math>), <math>\text{cm}^{-1}</math></b>	<b>Comprimento de onda (<math>\lambda</math>), nm</b>	<b>Frequência (<math>\nu</math>), Hz</b>
<b>Próximo (NIR)</b>	12.800 a 4.000	780 a 2.500	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
<b>Médio (MIR)</b>	4.000 a 200	2.500 a 5.000	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
<b>Distante (FIR)</b>	200 a 10	5.000 a 100.000	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$

Fonte: (SKOOG; HOLLER; NIEMAN, 2002).

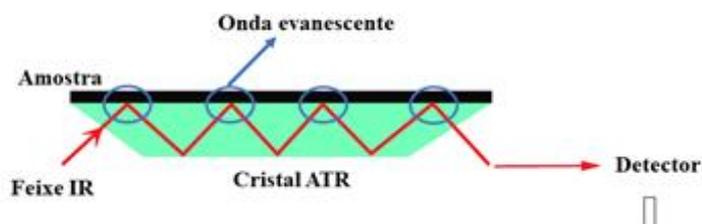
Para absorver radiação infravermelha, uma molécula sofre uma variação no momento de dipolo como consequência do movimento vibracional ou rotacional. O campo elétrico alternado da radiação interage com a molécula e causa variações na amplitude de um de seus movimentos. O contato da molécula com a radiação infravermelha geralmente não provoca transições eletrônicas e por não ser energética o suficiente, a absorção desta radiação está muito restrita a espécies moleculares que têm diferenças de energia pequenas entre vários estados vibracionais e rotacionais (SKOOG, ; HOLLER; NIEMAN, 2002).

Além disso, no MIR pode-se empregar o acessório de Reflexão Total Atenuada (ATR) em que os espectros de absorção são obtidos rapidamente em uma variedade de tipos de amostras com um mínimo de preparação. Para amostras líquidas, pode ser adicionado diretamente no acessório ATR, como por exemplo, amostras de biodiesel em diesel. Soluções aquosas podem também ser usadas, desde que o cristal não seja solúvel em água, utilizando-se este tipo de acessório a amostra pode ser preservada, ou seja, utilizando-se o ATR a técnica tem a vantagem de ser não destrutiva (HOLLER, 2009).

#### 4.1 Reflectância Total Alternada

A técnica de Reflectância Total Atenuada (ATR, do inglês, *Attenuated Total Reflectance*), baseia-se no fato de que ocorra reflexão interna total na interface dos dois meios conforme ilustrado na Figura 9. Essa radiação denominada onda evanescente pode ser parcialmente absorvida, colocando-se uma amostra em contato com meio mais denso. Se o meio menos denso absorve essa onda evanescente, ocorre atenuação do feixe nos comprimentos de onda das bandas de absorção. Esse meio mais denso é denominado prisma ou elemento de reflexão interna (BARBOSA, 2008).

Figura 9: Representação da análise por ATR horizontal.



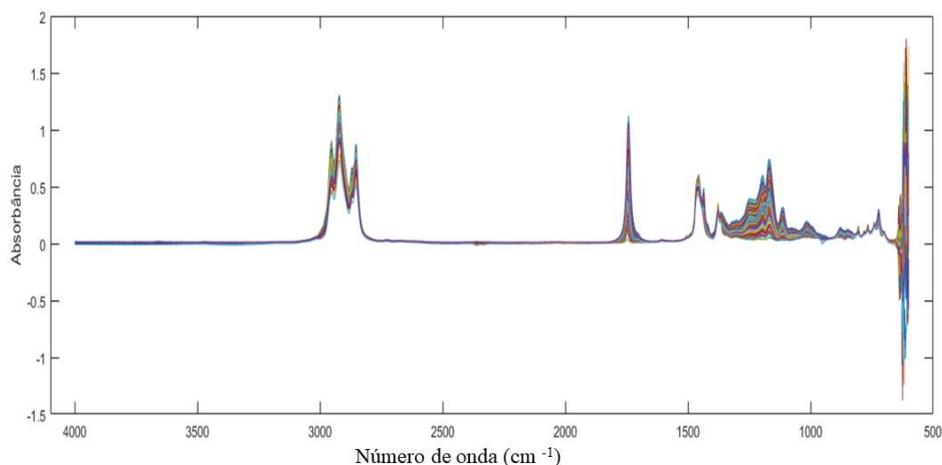
Fonte: adaptado (BARBOSA, 2008).

A espectroscopia no infravermelho médio oferece vantagens para o controle de qualidade e vem conquistando espaço nos laboratórios analíticos. Tais vantagens consistem em: velocidade, facilidade devido ao fato de que a amostra não necessita tratamento prévio, pequena quantidade de amostra utilizada (aproximadamente 0,50 mL), além de ser uma técnica de alta seletividade utilizada para determinações tanto qualitativas quanto quantitativas quando associada à ferramentas quimiométricas (SOARES et al., 2008).

As análises quantitativas e qualitativas a partir de dados de Espectroscopia MIR envolvem o uso de métodos quimiométricos (DE SOUZA et al., 2015). Isso ocorre devido à complexidade dos dados obtidos a partir destes espectros, os quais visualmente, isto é, sem qualquer tratamento, é difícil extrair informações sobre a qualidade das amostras.

Ao observar a Figura 10, os espectros de MIR obtidos durante este trabalho de bioquerosene/querosene no teor de 1,00 a 70,00% (v/v) é notório que não é possível distinguir os espectros sobrepostos com relação às suas respectivas concentrações.

Figura 10: Espectros originais da mistura de bioquerosene/querosene no teor de 1,00 a 70,00% (v/v) de macaúba.



Fonte: A autora.

Assim, é necessário que a Espectroscopia no Infravermelho seja auxiliada por métodos quimiométricos, uma vez que os espectros apresentam grande semelhança e muitas informações, tornando-se assim, imprescindível a associação a Quimiometria (FERRÃO et al., 2011).

## 5 Métodos Quimiométricos

Na primeira metade da década de 70 surgiu dentro da química analítica uma área de pesquisa que buscava extrair de uma grande quantidade de dados químicos complexos resultados analíticos interpretáveis, não podendo dessa maneira serem explicados por modelos teóricos da química clássica. Foi então que um certo número de químicos, como S. Wold (*Umea University* – Suécia), se intitularam como quimiométricos dando

origem a uma nova área da química analítica (BORGES NETO, 2005; BRERETON, 2000).

A quimiometria é definida como a aplicação de métodos matemáticos, estatísticos e de lógica formal para o tratamento de dados químicos de forma a extrair uma maior quantidade de informação e otimização dos resultados analíticos (MASSART et al., 1998). Os métodos quimiométricos são utilizados hoje para análises de dados, com a finalidade específica dentro de um estudo químico como a otimização de um processo, classificação de dados, modelagem e monitoramento de processos multivariados, construção de modelos de classificação e regressão e desenvolvimento de inteligência artificial, entre outros (BRERETON, 2000).

A química analítica quantitativa teve grande impacto com o desenvolvimento das técnicas de calibração multivariada. Desta é possível estimar uma propriedade de interesse a partir de outras medições, que normalmente são espectros obtidos por procedimentos analíticos mais simples, rápidos, menos onerosos e que dependem de menor quantidade de amostras (FILGUEIRAS, 2014).

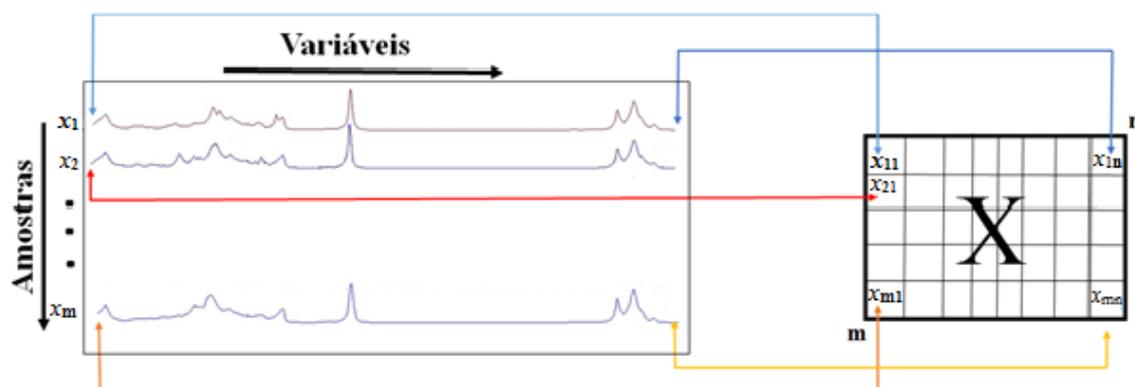
A calibração multivariada é definida como uma série de processos que permite estabelecer a relação entre a resposta instrumental e valores para uma propriedade de interesse correspondente (BRERETON, 2000). O modelo de calibração é uma função matemática ( $f$ ) que relaciona dois grupos de variáveis, sendo uma delas denominada dependente ( $y$ ) e a outra denominada independente ( $x$ ) conforme demonstrado pela Equação (1).

$$y = f(x) = w \cdot x + b \quad \text{Equação (1)}$$

Os espectros são organizados na matriz de dados  $\mathbf{X}$  de forma que cada amostra represente um vetor linha. Cada variável espectral (número de onda, deslocamento químico, etc.) é emparelhada na mesma coluna, assim todos os espectros devem ser medidos com mesma resolução espectral. Para  $n$  amostras com  $m$  variáveis espectrais, a matriz de dados terá dimensão  $\mathbf{X}_{(n,m)}$  ( $n$  linhas por  $m$  colunas). A Figura 11 ilustra como uma matriz de dados  $\mathbf{X}$  de dimensão  $n \times m$ , ou seja, “ $n$ ” espectros e “ $m$ ” variáveis pode

ser construída a partir de um vetor de resposta instrumental para calibração multivariada de primeira ordem.

Figura 11: Representação da construção da matriz  $\mathbf{X}$  para calibração multivariada.



Fonte: (MÁQUINA, 2017).

A calibração é dividida em dois passos: o primeiro passo é a modelagem, em que se estabelece uma relação matemática entre a Matriz  $\mathbf{X}$  e o vetor  $\mathbf{y}$ ; o segundo passo é a validação que consiste em verificar se o modelo é ou não realístico. Sempre que um método é proposto, existe a necessidade de averiguar se ele é adequado para as condições nas quais será aplicado, o qual assegura a confiabilidade dos resultados. (FERREIRA, 2015). A validação pode ser atestada pela determinação de parâmetros conhecidos como figuras de mérito recomendados pela IUPAC. As principais figuras de méritos são: exatidão, ajuste do modelo, linearidade, sensibilidade, sensibilidade analítica, seletividade, razão sinal/ruído, teste para erro sistemático “bias”, limite de detecção e quantificação e intervalos de confiança (GUIMARÃES, 2014).

### 5.1 Máquina de Vetores de Suporte (SVM)

O algoritmo Máquina de Vetores de Suporte (SVM, do inglês *Support Vector Machine*) é um método de aprendizagem estatístico supervisionado desenvolvido por Vapnik nos anos 90 (CORTES, C.; VAPNIK, 1995), originalmente desenvolvido para problemas de classificação de (ALVES; POPPI, 2013; BRUDZEWSKI et al., 2006) e posteriormente estendida para a regressão por vetores de suporte (ALVES, 2012a; FILGUEIRAS, 2014). O objetivo principal do SVM é encontrar um hiperplano como

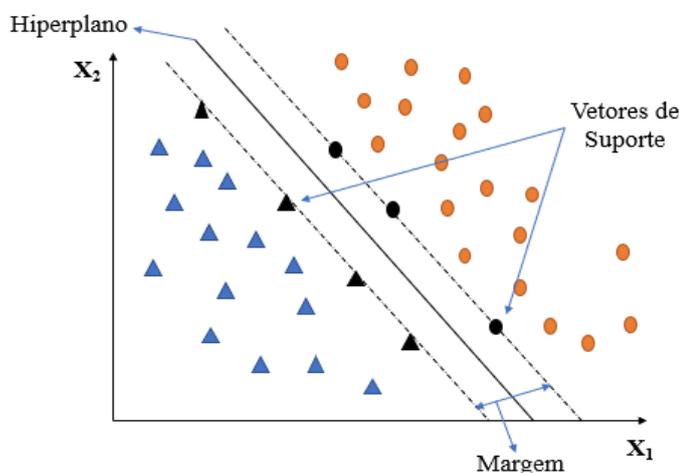
superfície de decisão de tal forma que a margem de separação entre os exemplos positivos e negativos seja máxima (MARETTO, 2011)

Considerando um conjunto de treinamento  $\mathbf{x}_i$ , assumimos que os subconjuntos representados por  $d_i$  é a resposta desejada,  $d_i = \{+1, -1\}$ , que representa as classes linearmente separáveis. Para um efeito de simplificação, a Equação (2) que separa as duas classes usando um modelo linear pode ser definida da seguinte forma:

$$f(x) = \mathbf{w} \cdot \mathbf{x} + b \quad \text{Equação (2)}$$

sendo  $\mathbf{w} \in R^m$  um vetor de pesos ajustados, que multiplicado pelo vetor amostral  $\mathbf{x}$  e somado à constante  $b$  define o hiperplano que separa as classes e permite identificar a classe ao qual a amostra pertence (FILGUEIRAS, 2014). A Figura 12 ilustra o hiperplano de separação para um conjunto de treinamento linearmente separável, com os vetores de suporte e margem.

Figura 12: Representação de um conjunto de dados com hiperplano de separação ótima.

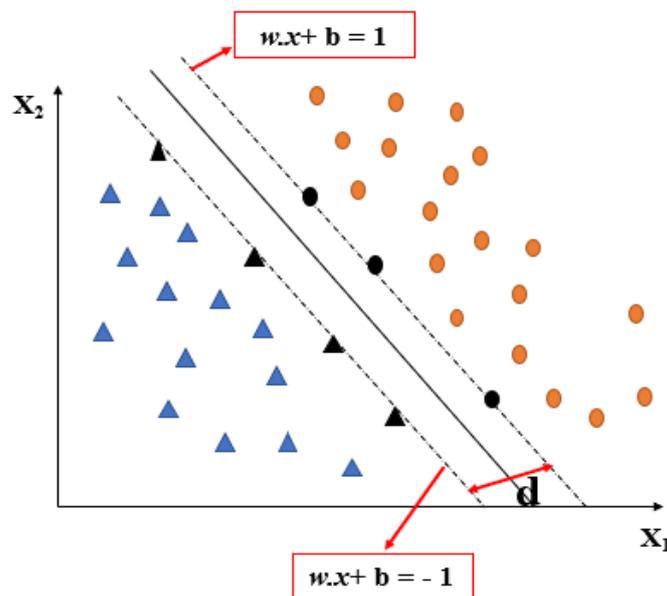


Fonte: A autora.

A distância entre o hiperplano e o ponto mais próximo de cada classe é nomeada de margem. O modelo SVM primeiramente classifica as classes, por meio de cada ponto pertencente a cada uma das classes, e em seguida ajusta o modelo de maneira a maximizar a margem entre o hiperplano de separação e os vetores de suporte. Assim, as amostras nas quais  $\mathbf{w} \cdot \mathbf{x} + b = 1$  pertencem a classe +1, enquanto as amostras que tiverem

$wx + b = -1$  correspondem a classe -1, conforme Figura 13, os vetores de suporte do modelo SVM são amostras que demarcam a fronteira entre as classes, sendo assim as mais importantes para a classificação. O modelo SVM é construído apenas com as amostras que são vetores de suporte (RUSCHEL, 2017; FILGUEIRAS, 2014).

Figura 13: Hiperplano de separação e margem.



Fonte: A autora.

Tem-se que  $w x_1 + b = +1$  e  $w x_2 + b = -1$ . A diferença entre essas equações fornece  $w (x_1 - x_2) = 2$  (HEARST et al., 1998). Substituindo esse resultado na Equação 3.

$$\frac{2 (x_1 - x_2)}{\|w\| \|x_1 - x_2\|} \quad \text{Equação (3)}$$

Como deseja-se obter o comprimento do vetor projetado a Equação 4, obtendo:

$$\frac{2}{\|w\|} \quad \text{Equação (4)}$$

Essa é a distância  $d$ , ilustrada na Figura 13, entre os hiperplanos  $w x_1 + b = +1$  e  $w x_1 + b = -1$ , paralelos ao hiperplano separador. Assim,  $\frac{1}{\|w\|}$  é a distância mínima entre o hiperplano separador e os dados de treinamento. Essa distância é definida como a margem geométrica do classificador linear (CAMPBELL, 2002). A partir das

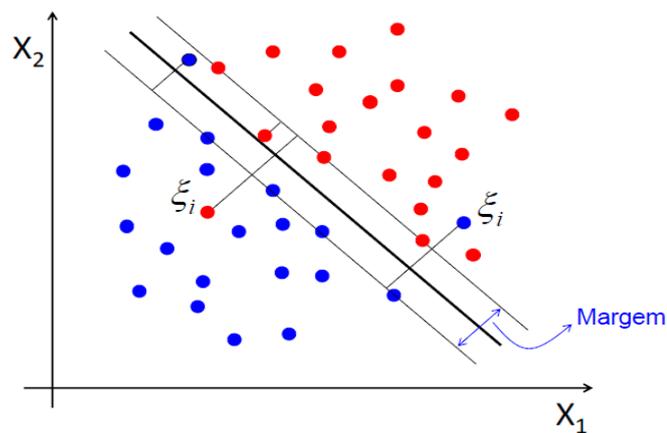
considerações anteriores, verifica-se que a maximização da margem de separação dos dados em relação a  $wx + b = 0$  pode ser obtida pela minimização de  $\|w\|$ , conforme Equação 5 (BURGES, 1998). Dessa forma, recorre-se ao seguinte problema de otimização (HOFMANN. T, SCHOLKOPF. B, SMOLA, 2008):

$$\text{Minimizar: } \frac{1}{2} \|w\|^2 \quad \text{Equação (5)}$$

### 5.1.2 Margem Suave

Em situações reais, é difícil encontrar aplicações cujos dados sejam linearmente separáveis. Isso se deve a diversos fatores, entre eles a presença de ruídos e *outliers* nos dados ou à própria natureza do problema. Para contornar esse problema, Cortes e Vapnik (1995), introduziram o conceito de variável de folga  $\xi_i$ , no qual admitem que algumas amostras possam ter um erro associado à sua classificação (FILGUEIRAS, 2014; LORENA; CARVALHO, [s.d.]). Este erro é proporcional a distância da amostra ao hiperplano definido pelos vetores de suporte da classe à qual pertence Figura 14.

Figura 14: Ilustração de um problema de classificação binária linearmente inseparável.



Fonte: (FILGUEIRAS, 2014).

A margem para essa nova situação mantém a estrutura de classificação linear, mas com uma certa tolerância as violações da hipótese de separabilidade, as SVMs obtidas neste caso também podem ser referenciadas como SVMs com margens suaves.

Um erro no conjunto de treinamento é indicado por um valor de  $\xi_i$  maior que 1. Logo, a soma dos  $\xi_i$  representa um limite no número de erros de treinamento. Para levar em consideração esse termo, minimizando assim o erro sobre os dados de treinamento, a função objetivo da Equação 5 é reformulada como (CAMPBELL, 2002):

$$\text{minimize: } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i \quad \text{Equação (6)}$$

A constante  $C$  é conhecida como “parâmetro de margem” e estabelece a importância relativa das duas parcelas do lado direito da igualdade da Equação 6, nesse processo de minimização, ou seja, a constante  $C$  controla os erros de classificação (BONASSO, 2013).

Para um baixo valor de  $C$  admite-se que muitas amostras possam ser classificadas com erros, gerando uma função de classificação mais suave. Caso o valor de  $C$  seja alto, a função de classificação torna-se menos suave para que as amostras não sejam erroneamente classificadas, podendo gerar um maior número de vetores de suporte. Assim o valor de  $C$  deve ser ajustado para cada problema. Mesmo com o problema linear ajustado para o caso onde é admitido erros de classificação, problemas reais podem não ser lineares. Assim, uma transformação é aplicada aos dados de entrada do ajuste de uma função não linear (FILGUEIRAS, 2014).

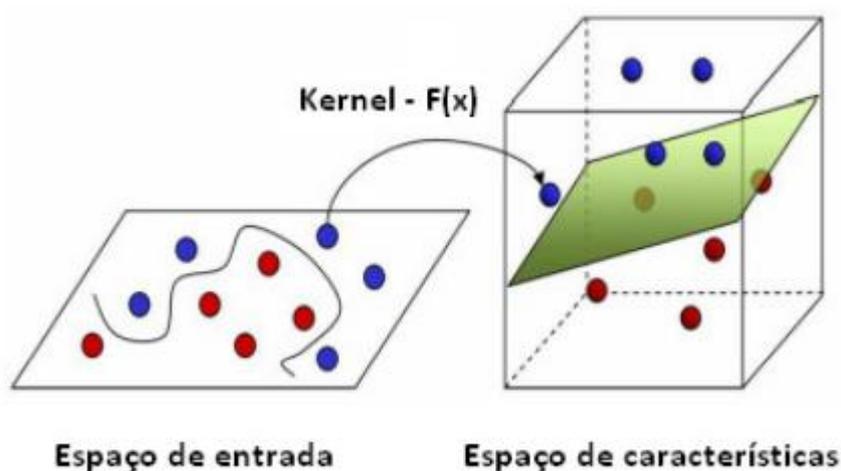
### 5.1.3 Função *Kernel*

Muitos conjuntos de dados utilizados para resolver problemas em química tem característica não linear, enquanto o algoritmo do SVM está baseado em uma máquina de aprendizado linear com a utilização de margem de separação. Assim, torna-se necessário a utilização de uma técnica para mapeamento não linear de modo a tornar possível o tratamento dos conjuntos de dados com SVM (ALVES, 2012b; HOFMANN. T, SCHOLKOPF. B, SMOLA, 2008).

As funções *Kernel* têm a finalidade de projetar os vetores de características de entrada em um espaço de característica de alta dimensão para classificação de problemas que se encontram em espaços não linearmente separáveis. Isso é feito, pois a medida que se aumenta o espaço da dimensão do problema, aumenta também a probabilidade desse problema se tornar linearmente separável em relação a um espaço de baixa dimensão(GONÇALVES, 2008).

A Figura 15 mostra o processo de transformação de um domínio não linearmente separável, em um problema linearmente separável através do aumento da dimensão, onde é feito um mapeamento por uma função de *Kernel*.

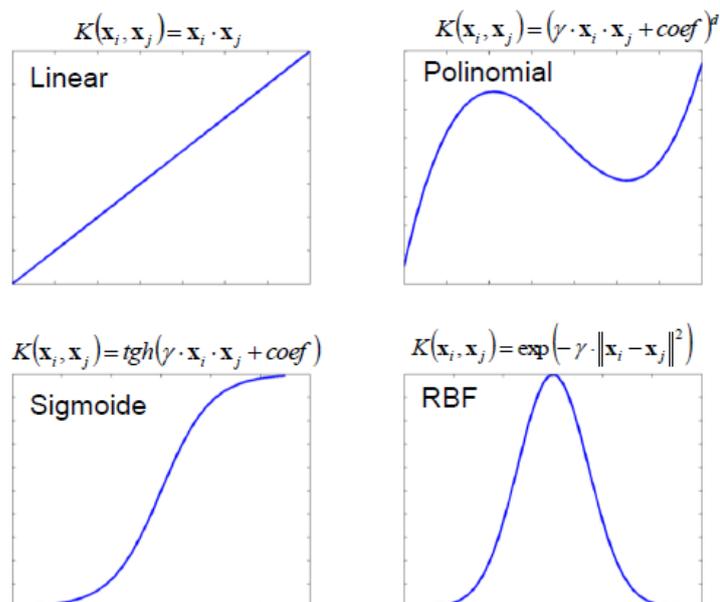
Figura 15: Problema não linear separável em um problema linearmente separável.



Fonte: (RABELO, 2008).

As funções Kernel podem ser usadas para facilitar a separação das margens, pois realizam um mapeamento implícito dos padrões para outro espaço de atributos onde a separabilidade é maior(NEGRI, 2013).Segundo (HOFMANN. T, SCHOLKOPF. B, SMOLA, 2008) as funções mais utilizadas são a linear, polinomial, função de base radial (RBF do inglês, *radial basis function*) e sigmoide. Um esboço é apresentado na Figura 16, onde cada uma destas funções apresenta parâmetros que devem ser determinados pelo usuário.

Figura 16: Representações gráficas das funções de Kernel mais utilizadas no SVM.



Fonte: (FILGUEIRAS, 2014).

Nesse estudo, para o desenvolvimento dos modelos de regressão com SVM foi utilizado a função *Kernel* RBF, por se tratar do método de *Kernel* mais simples e rápido, fornecendo bons resultados quando apenas dois parâmetros necessitam ser selecionados.

#### 5.1.4 Regressão por Vetores de Suporte (SVR)

Para problemas de regressão foi criada uma adaptação ou extensão do método SVM. Tal extensão se dá pela inclusão de uma função perda. O SVM foi originalmente desenvolvido para resolver problemas de classificação. Ao contrário do SVM, o SVR propõe determinar um hiperplano ótimo em que as amostras de treinamento estejam tão próximas quanto possível, não importando em qual dos lados da superfície os pontos se localizem e sim que a distância para a superfície seja a mínima possível (LIMA, 2004). Assim, no SVR dois conceitos importantes devem ser primeiramente comentados, o  $\epsilon$ -tubo e a função perda  $\epsilon$ -insensível (ALVES, 2012).

### 5.1.5 $\varepsilon$ -tubo e a função de perda $\varepsilon$ -insensível

No  $\varepsilon$ -SVR busca-se encontrar uma função que produza saídas contínuas para todos os dados de treinamento que desviem no máximo de  $\varepsilon$  do seu rótulo desejado e ao mesmo tempo permite que o  $\varepsilon$ -tubo seja o mais delgado possível. Essa função também deve ser o mais uniforme e regular possível.

A teoria do aprendizado estatístico pode também ser aplicada para problemas de regressão pela introdução de uma nova função de perda, de modo a evitar a modelagem de erro experimental e o sobreajuste dos modelos: a função de perda  $\varepsilon$ -insensível. Assim, na regressão com vetores de suporte (SVR) dois conceitos importantes devem ser primeiramente comentados,  $\varepsilon$ -tubo e a função de perda  $\varepsilon$ -insensível. Os dados originais são primeiramente mapeados não linearmente em um espaço de características de alta dimensão e então uma função linear é ajustada para aproximar a função latente entre  $\mathbf{X}$  e  $\mathbf{y}$ .

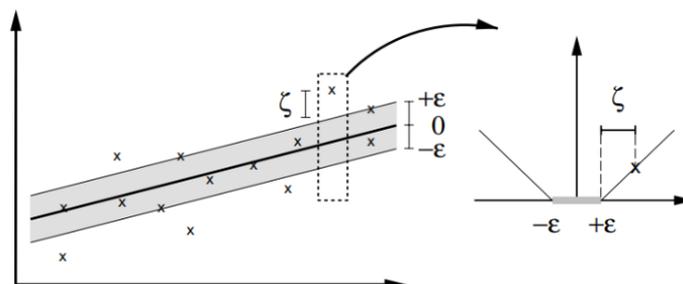
A função de perda é denominada  $\varepsilon$ -insensível descrita por:

$$|\xi|_{\varepsilon} = 0, \quad \text{se } |\xi| \leq \varepsilon$$

$$|\xi|_{\varepsilon} = |\xi| - \varepsilon, \quad \text{caso contrário} \quad \text{Equação (8)}$$

O  $\varepsilon$ -tubo, apresentado na Figura 17 mostra a representação gráfica da situação proposta pela Equação (8) na qual pode ser obtido pelo movimento da linha sólida com um deslocamento de  $\varepsilon$  para cima e para baixo. Aqui  $\varepsilon$  é um número positivo pré-definido.

Figura 17: Ilustração do processo de penalização da função de perda  $\epsilon$ -insensível para SVM com Kernel linear.



Fonte: (SMOLA; SCHOLKOPF, 2004).

De fato, essa restrição define um tubo de raio  $\epsilon$  ao redor da função de regressão hipotética, de forma que se um ponto está posicionado dentro do tubo a função de perda é igual a 0, enquanto se um ponto está posicionado fora do tubo, a perda é proporcional a distância euclidiana entre o ponto (ALVES, 2012).

## 5.2 Resolução Multivariada de Curvas com Mínimos Quadrados Alternados (MCR-ALS)

A Resolução Multivariada de Curvas (MCR do inglês, *Multivariate Curve Resolution*) é um método quimiométrico de processamento de sinais analíticos que tem o intuito de resolver misturas de sinais. Ou seja, permite adquirir informações sobre componentes puros de uma mistura (RAVAZZI, 2019).

Em 1993, Romà Tauler e colaboradores implementaram um método de resolução multivariada de curvas aplicando o algoritmo mínimos quadrados alternados (ALS do inglês, *Alternating Least Squares*). Para a aplicação do MCR-ALS, o sinal analítico deve obedecer uma relação semelhante à lei de Beer-Lambert, ou seja os dados devem ter uma relação linear com a concentração (MARÇO et al., 2014).

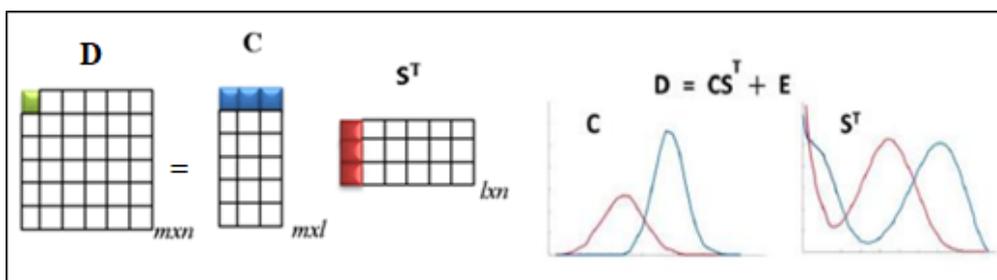
O método do MCR-ALS utiliza como processo a decomposição bilinear da matriz de dados  $\mathbf{D}$  em três conjuntos de dados, matriz de dados  $\mathbf{C}$  que representa o perfil de concentração, matriz  $\mathbf{S}^T$  que representa o perfil espectral desses compostos puros e a

matriz de resíduos  $\mathbf{E}$  (GOICOCHEA; OLIVIERI; TAULER, 2010). Conforme expresso na Equação 9.

$$\mathbf{D} = \mathbf{C}\mathbf{S}^T + \mathbf{E} \quad (9)$$

A Equação 9 pode ser representada pelo esquema da Figura 18 pode ser observada visualmente a decomposição da matriz, onde para cada elemento da matriz  $\mathbf{X}$  existe um produto correspondente entre um vetor linha da matriz  $\mathbf{C}$  e um vetor coluna  $\mathbf{S}^T$  (MARÇO et al., 2014).

Figura 18: Representação de decomposição do MCR para um sistema de dois componentes.



Fonte: adaptado(DE JUAN; JAUMOT; TAULER, 2014).

Para que o MCR-ALS utilize a resolução iterativa da Equação 9, é necessário estimar o posto (*rank*) da matriz de dados, ou seja, do número de componentes presentes na mistura e que produzem o sinal analítico. No caso de dados matriciais, o posto corresponde aos números de linhas linearmente independentes que compõe a matriz de dados. O número de componentes é estimado com base no conhecimento do sistema em estudo ou a partir dos resultados da decomposição em valores singulares (SVD do inglês, *Singular Value Decomposition*) da matriz de dados. Na estimativa por SVD, o número de espécies é aproximado para o número de valores singulares acima do valor singular correspondente ao nível do ruído instrumental dos dados (DE JUAN; JAUMOT; TAULER, 2014; RAVAZZI, 2019).

Para que o processo de otimização de matrizes  $\mathbf{C}$  e  $\mathbf{S}^T$  tenha início, é necessário o conhecimento inicial dos sinais ou das concentrações dos componentes puros presentes

na amostra. Entretanto, nem sempre esta informação está disponível e, portanto, deve-se recorrer a métodos apropriados para essa finalidade. Entre esses métodos, destacam-se a Análise de Fatores Evolucionários (EFA do inglês, *Evolving Factor Analysis*) e a aproximação da variável pura (do inglês, *pure*) (DE JUAN; TAULER, 2006; GOICOECHEA; OLIVIERI; TAULER, 2010; MARÇO et al., 2014).

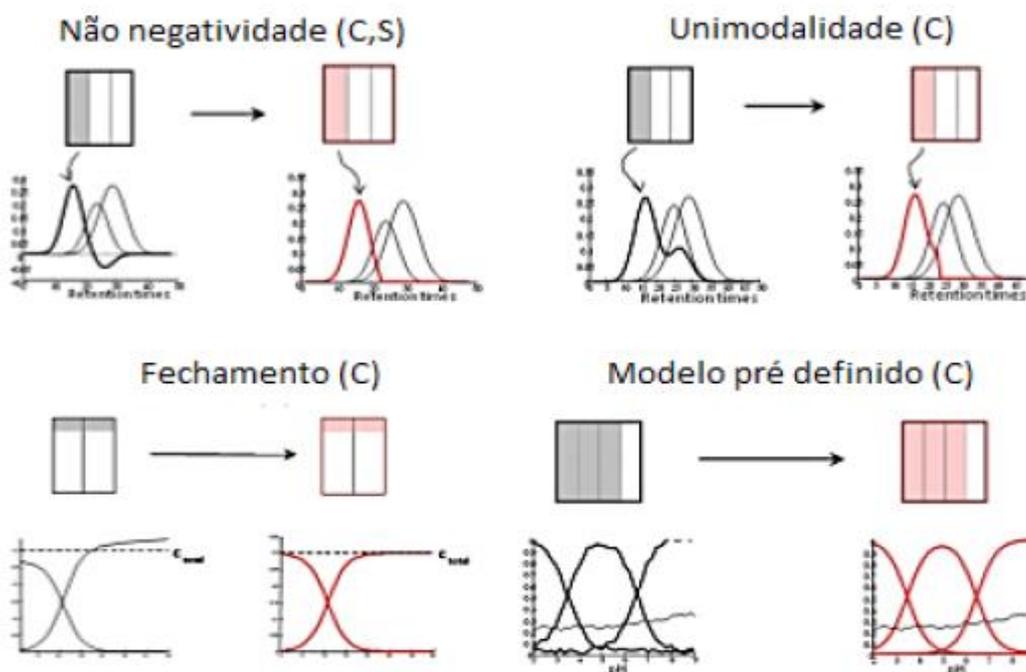
O MCR-ALS é um tipo de MCR no qual o modelo é resolvido utilizando-se o algoritmo de mínimos quadrados alternados (ALS), que iterativamente ajusta as matrizes  $\mathbf{C}$  e  $\mathbf{S}^T$  ao conjunto de dados  $\mathbf{D}$ . O processo de otimização iterativo continua até que o critério de convergência seja atingido, ou seja, quando a variação de resultados entre iterações consecutivas apresentar valores abaixo de um valor limite predeterminado, ou quando um certo número de iterações for excedido (GOICOECHEA; OLIVIERI; TAULER, 2010; JAYARAMAN et al., 2012).

Para tentar diminuir a ambiguidade rotacional é considerada uma das principais causas de soluções incorretas obtidas por MCR-ALS, são aplicado as chamadas “restrições”, que são condições matemáticas impostas ao algoritmo para tentar diminuir ou até suprir esse problema (RAVAZZI, 2019).

### 5.2.1 Imposição de Restrições

Desta maneira, é possível impor restrições ao sistema baseado em conhecimentos prévios ou em critérios que garantirão resultado final. Assim, as restrições são a não-negatividade, a unimodalidade, a seletividade ou posto local e o sistema fechado ou balanço de massa, podem ser aplicadas e auxiliam o algoritmo ALS a encontrar uma solução matemática que facilite a interpretação dos problemas (CORDEIRO DANTAS; LAURENTINO ALVES; POPPI, 2017). Algumas dessas restrições são apresentadas na Figura 19.

Figura 19: Perfis antes (cor cinza) e após (cor rosa) a aplicação das restrições de não negatividade, unimodalidade, sistema fechado e modelo pré definido.



Fonte: (DE JUAN; TAULER, 2006).

A não-negatividade pode ser aplicada tanto nas concentrações quanto no sinal analítico e permite somente soluções maiores ou iguais a zero, já a unimodalidade admite soluções com a presença de apenas um ponto máximo no perfil. Essas duas restrições podem ser impostas às matrizes  $\mathbf{C}$  ou  $\mathbf{S}^T$ , ou as duas, e podem ser implementadas para um componente isolado ou até mesmo para todos (DE JUAN; TAULER, 2006).

As restrições de fechamento, e de modelos pré definidos são aplicadas essencialmente aos perfis de concentração. A restrição de fechamento estabelece que o somatório das concentrações relativas permaneça constante durante o processo de otimização (DE JUAN; TAULER, 2006).

### 5.2.2 MCR-ALS com restrição de correlação

A restrição de correlação, utilizada nesse trabalho tem como objetivo a criação de modelos internos de calibração univariada baseados nos valores de concentração calculados, e os valores de referência das amostras conhecidas (DEBUS et al., 2017). Deste modo, a partir desses modelos de calibração univariada, as concentrações para as amostras desconhecidas são estimados, minimizando assim o efeito de ambiguidade rotacional (RAVAZZI, 2019).

Em cada etapa iterativa, é realizada uma regressão entre os valores de concentração obtidos pelo MCR-ALS ( $C_{als}^{cal}$ ) e os valores de referência presente na amostra ( $C_{als}^{val}$ ) para cada componente das amostras, gerando um modelo descrito pela Equação 10, onde  $b$  é o coeficiente angular e  $b_0$  o coeficiente linear (ROCHA et al., 2014).

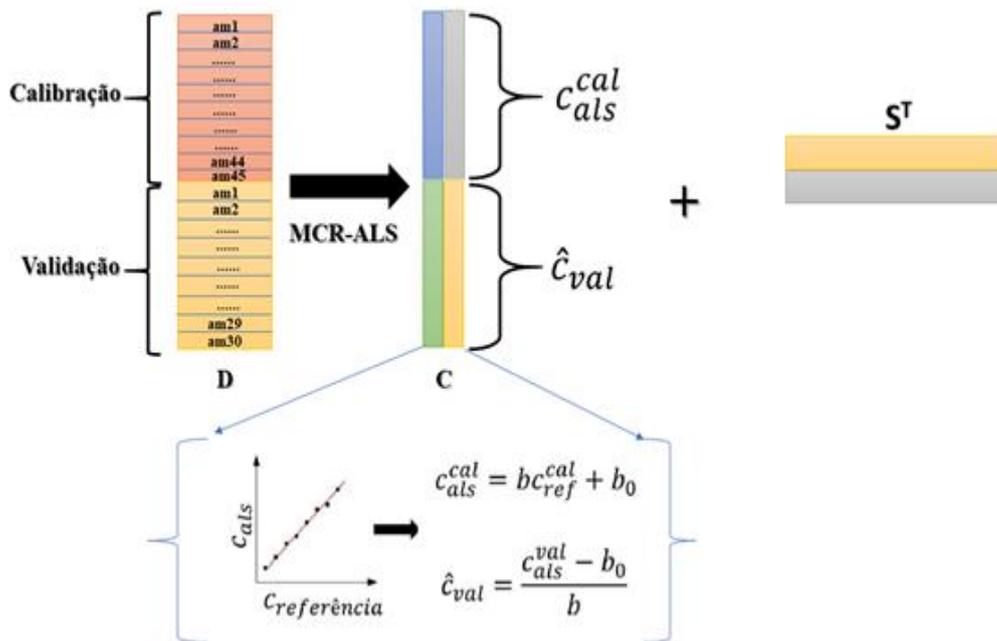
$$C_{als}^{cal} = b \cdot C_{als}^{val} + b_0 \quad \text{Equação (10)}$$

Obtido os valores de  $b$  e  $b_0$ , pode-se prever os valores das concentrações das amostras nos conjuntos de calibração e de validação através da Equação (11).

$$\hat{C}_{val} = \frac{C_{als}^{cal} - b}{a} \quad \text{Equação (11)}$$

A Figura 20 representa esquematicamente o processo de otimização pelo ALS utilizando a restrição de correlação tendo como base uma matriz  $\mathbf{D}$  formada pelos conjuntos de calibração e validação.

Figura 20: Restrição de correção feito pelo toolbox MCR-ALS GUI 2.0



Fonte: adaptado(RAVAZZI, 2019).

As figuras de mérito são utilizadas para avaliar a qualidade e a confiabilidade dos resultados obtidos por MCR-ALS, são utilizados alguns parâmetros que comparam a matriz de dados originais  $D$  com uma nova matriz  $D^*$ , que podem ser calculados por dois parâmetros: a porcentagem de falta de ajuste (%LOF – *Lack of Fit*) e a porcentagem de variância explicada ( $R^2$ ) (JAYARAMAN et al., 2012; MARÇO et al., 2014). Esses parâmetros são determinados pelas Equações 12 e 13, respectivamente.

$$lackoffit(\%) = 100 \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \quad \text{Equação (12)}$$

$$R^2 = 100 \sqrt{\frac{\sum_{i,j} d_{ij}^2 - \sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \quad \text{Equação (13)}$$

onde  $d_{ij}$  é um elemento da matriz de dados experimentais e  $e_{ij}$  é o valor residual relacionado obtido da diferença entre os dados experimentais (matriz  $D$ ) e os dados reproduzidos (produto da matriz  $CST$  obtido por MCR-ALS). Como o desejado é que

essas matrizes sejam mais semelhantes, espera-se que baixos valores de %LOF e valores de  $R^2$  próximos a 1 sejam obtidos.

## 6 Procedimento Experimental

### 6.1 Preparo das misturas de bioquerosene/querosene

As amostras foram preparadas à temperatura ambiente ( $26 \pm 2$  °C) utilizando uma balança analítica marca (Santorius, modelo BP211D). Para esse procedimento, foi adicionada uma alíquota de bioquerosene de massa conhecida em um frasco âmbar de 25,00 mL, posteriormente da adição do querosene e em seguida homogeneizado em agitador de tudo marca (Phoenix, modelo AP56). As amostras de misturas bioquerosene/querosene foram preparadas na faixa de concentração de 1,00 a 70,00 (% v/v). Foram preparadas 75 amostras de misturas bioquerosene/querosene, sendo 45 amostras para o conjunto de calibração e 30 amostras para o conjunto de previsão. Ambos os conjuntos de amostras foram preparados de forma bem distribuída ao longo da faixa de concentração avaliada, em seguida foram feitos os registros dos espectros MIR.

### 6.2 Obtenção dos espectros de MIR

Os espectros de absorvância no infravermelho médio foram obtidos em quintuplicatas, utilizando o espectrômetro Shimadzu – IRPrestige-21, pertencente ao Laboratório Multiusuário, do Instituto de Química, da Universidade Federal de Uberlândia, equipado com acessório Horizontal de Reflectância Total Atenuada (HATR), com cristal de ZnSe na região de 4000 a 600  $\text{cm}^{-1}$ , resolução de 4  $\text{cm}^{-1}$  e 16 varreduras.

### 6.3 Pré-Tratamento dos dados

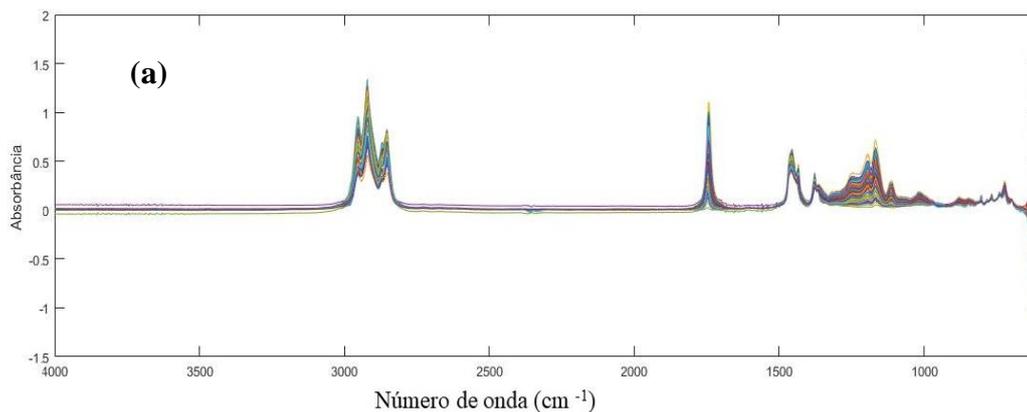
O programa utilizado para a construção dos modelos foi o MATLAB<sup>®</sup> versão R2017<sup>a</sup> software (Math Works Inc., Natick, EUA) juntamente com o pacote PLS\_Toolbox versão 8.8.1 (Eigenvector Research, Inc., Manson, EUA) e pacote MCR-ALS GUI 2.0. Para o desenvolvimento dos modelos multivariados, as linhas de bases foram corrigidas para os intervalos de 2500 a 1850 $\text{cm}^{-1}$  e de 4000 a 3300 $\text{cm}^{-1}$ .

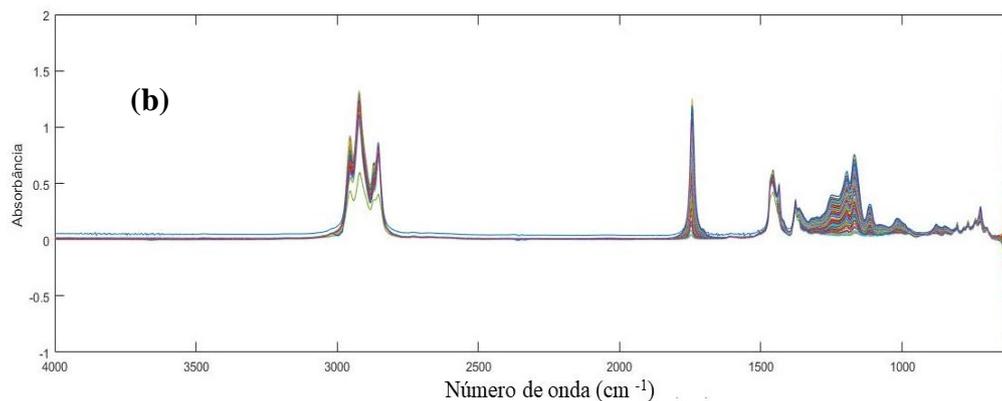
## 7 Resultados e Discussões

### 7.1 Espectros MIR de bioquerosene e querosene

Os espectros de MIR da mistura de bioquerosene/querosene foram obtidos em quintuplicada, totalizando 375 espectros de cada mistura de bioquerosene/querosene na faixa de 1,00 a 70,00 (% v/v). Esses dados obtidos foram sobrepostos usando comandos no ambiente MATLAB<sup>®</sup> sem nenhum pré tratamento conforme a Figura 21 (a) BioQAVM/QAV e (b) BioQAVP/QAV.

Figura 21: Espectros de MIR das misturas de bioquerosene/querosene nas proporções de 1 a 70 (% v/v) (a) BioQAVM/QAV (b) BioQAVP/QAV.



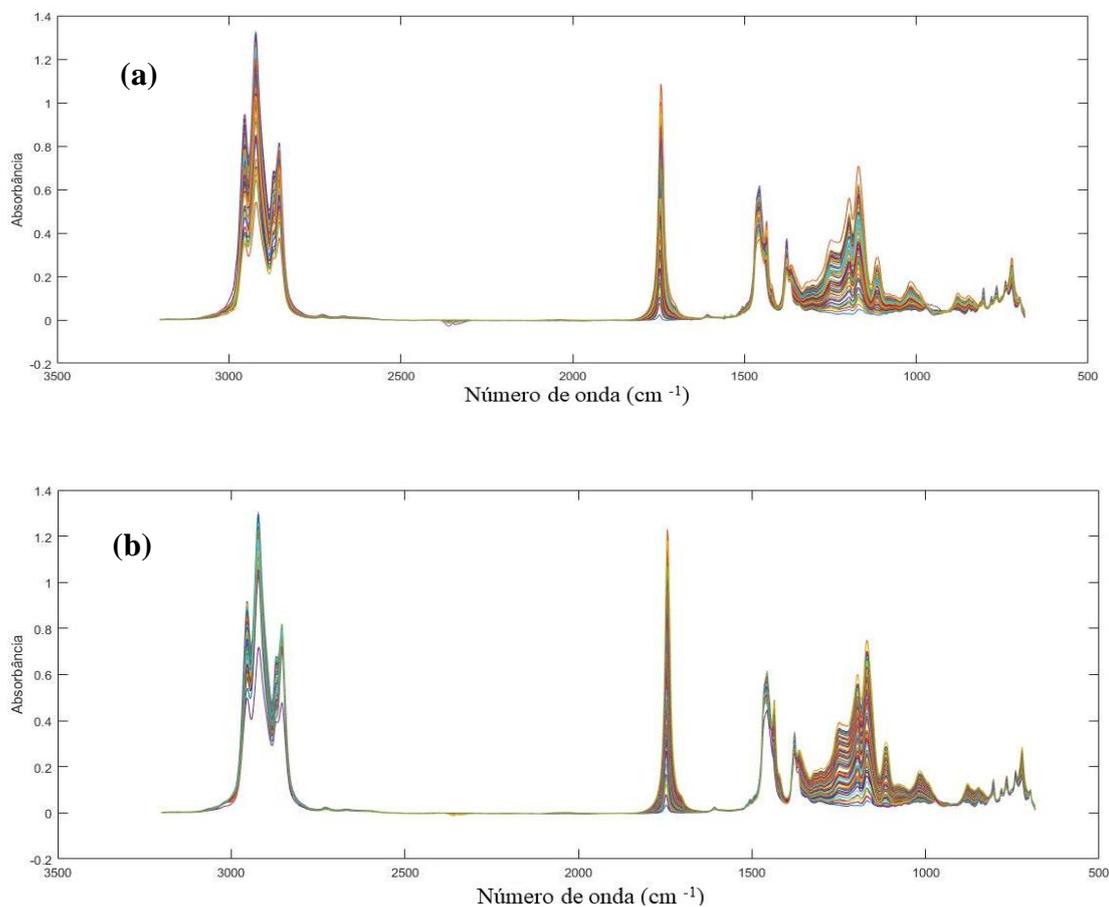


Fonte: A autora.

Ao observar a Figura 21 (a) e (b), não é possível distinguir os espectros sobrepostos com relação às suas respectivas concentrações. Deste modo, para facilitar a interpretação e a obtenção de informações, é necessário aplicar métodos quimiométricos, uma vez que os espectros apresentam grande semelhança. Percebe-se que a região inicial (menor número de onda) dos espectros possui grande quantidade de ruídos que são inerentes ao equipamento. Para que esses ruídos não influenciem na construção do modelo é necessário que se faça um corte nos espectros nas faixas iniciais a fim de eliminar essas informações interferentes. Da mesma forma, na região final dos espectros não há informações relevantes sobre a absorvância relativa aos números de onda. Sendo assim, essa região também é cortada minimizando assim possíveis erros na construção do modelo.

A região central dos espectros (números de onda intermediários) é uma região com absorvância aproximadamente zero. A fim de garantir essa aproximação, faz-se uma correção nas linhas de base dos espectros através do método de *baseline*. A Figura 22 mostra o resultado dos espectros gerados após os cortes e correções por *baseline*.

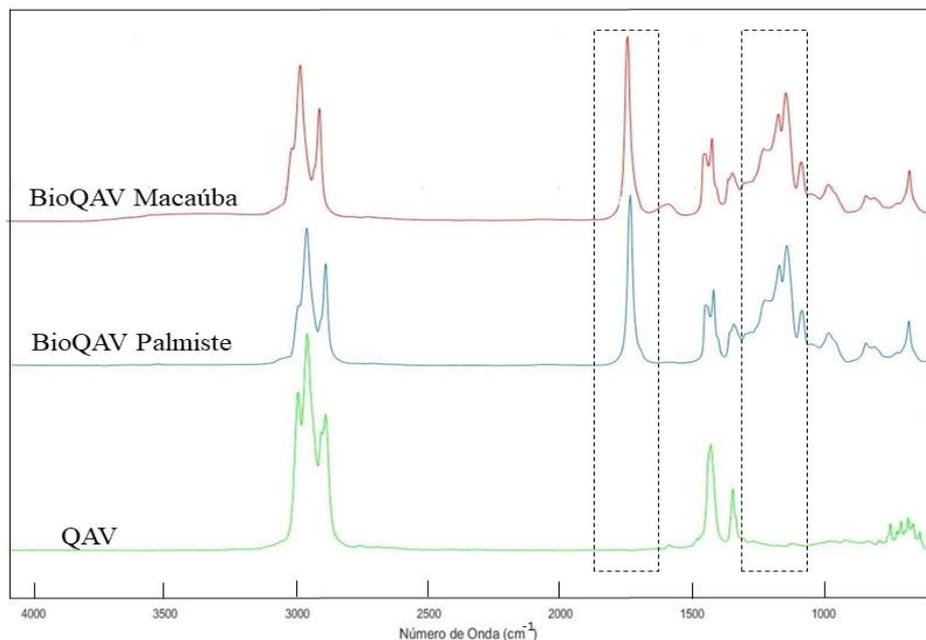
Figura 22: Espectros de MIR corrigidos da mistura de bioquerosene/querosene de (a) macaúba e (b) palmiste.



Fonte: A autora.

A Figura 23 apresenta os espectros MIR de bioquerosene de macaúba, palmiste e do querosene de aviação puro empregados nesse estudo. Fazendo uma avaliação detalhada dos espectros verifica-se principalmente a diferença entre os espectros dos BioQav e do QAV.

Figura 23: Espectros MIR dos bioquerosene puro de macaúba (vermelho), bioquerosene de palmiste (azul) e do QAV puro (verde).



Fonte: A autora.

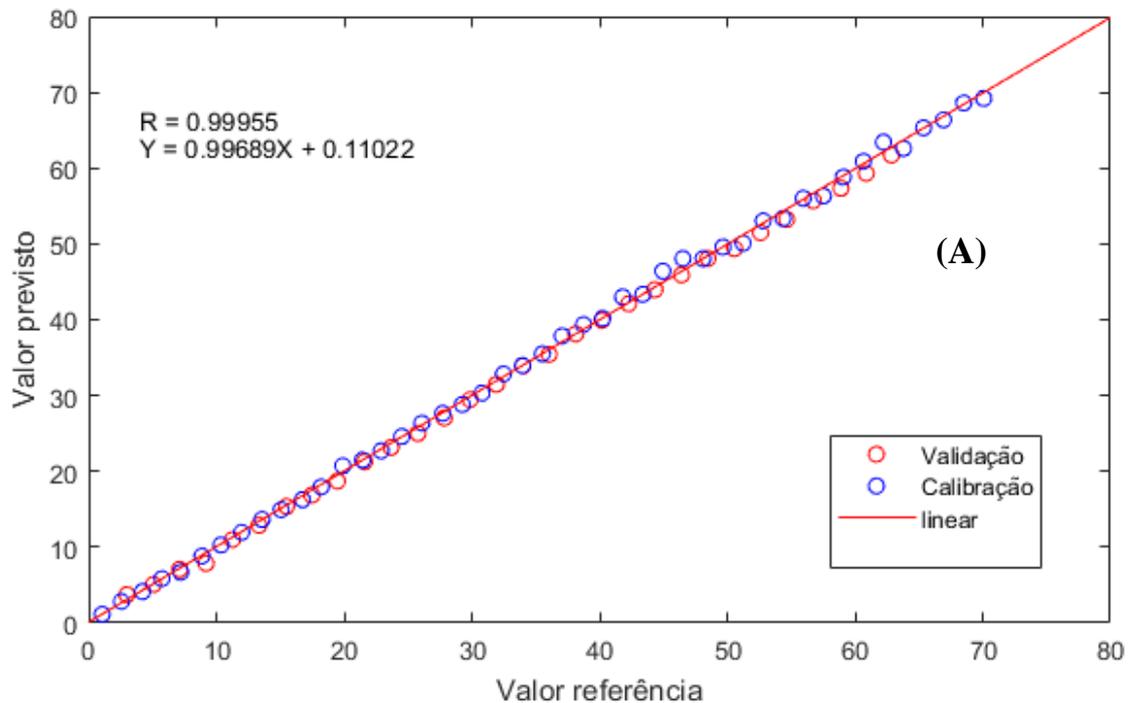
No espectro do QAV verifica-se bandas de absorção correspondente as três regiões espectrais características: a região entre 2800 cm<sup>-1</sup> e 3000 cm<sup>-1</sup> atribuídas às vibrações de estiramento das ligações C-H dos grupamentos de metila e metileno; bandas de intensidade intermediarias na região de 1250 cm<sup>-1</sup> a 1500 cm<sup>-1</sup> das vibrações de deformação angular das ligações C-H do grupo metila e metileno, outra banda de baixa intensidade, porém relevante na região de 730 cm<sup>-1</sup> atribuída às vibrações de deformação angular assimétrica resultante das deformações C-H do grupo metileno (SKOOG, D. A.; HOLLER, F. J.; NIEMAN, 2002).

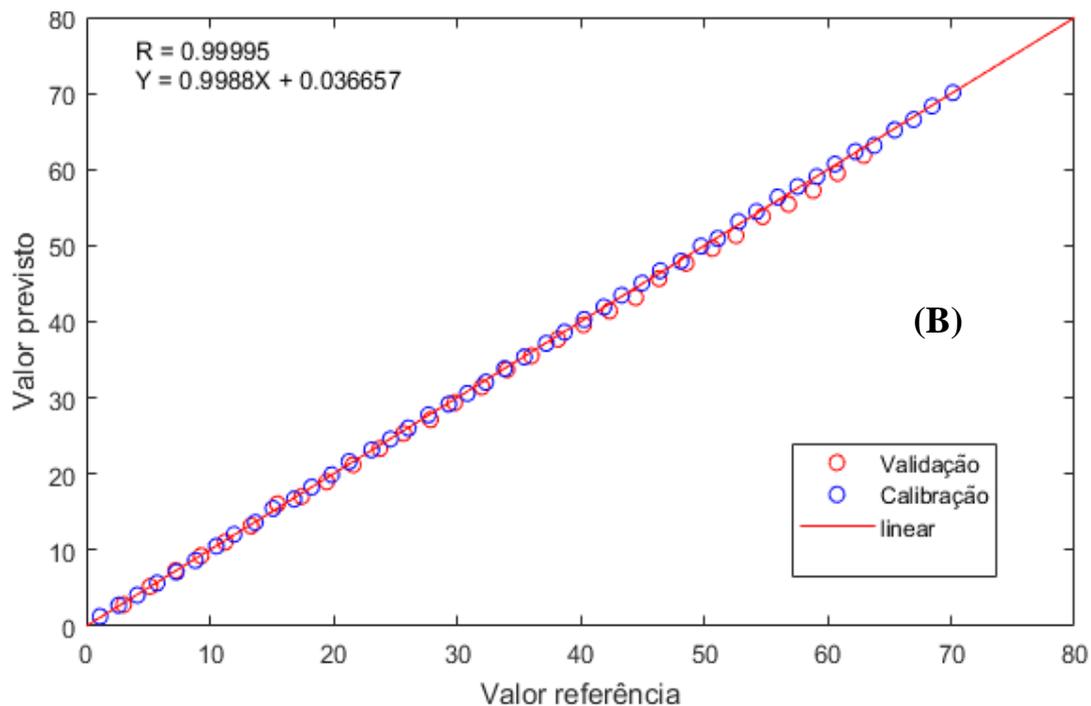
Os bioquerosenes apresentam em comum padrões de absorção na faixa de 1700 cm<sup>-1</sup> a 1750 cm<sup>-1</sup>, da carbonila (C=O) e a presença de bandas na região de 1000 cm<sup>-1</sup> a 1300 cm<sup>-1</sup> atribuídas às vibrações de estiramento da ligação C-O (BARBOSA, 2008).

## 7.2 Modelos SVR

As curvas de calibração multivariadas dos modelos SVR para os BioQAV de macaúba e palmiste são apresentados na Figura 24, onde podemos observar que as amostras de previsão estão dentro da faixa de concentração para os quais os modelos foram desenvolvidos; isto é, não houve uma extrapolação dos modelos. Através do coeficiente de correlação (R) é possível perceber uma boa relação linear entre os dados de referência e os calculados pelo modelo.

Figura 24: Modelos de regressão por SVR (A) BioQAVM e  
(B)BioQAVP





Fonte: A autora.

A linearidade é a capacidade de um método analítico obter resultados proporcionais à concentração da espécie em análise em que é possível relacionar uma medida através do conhecimento de uma dada concentração. O ajuste dos modelos foi verificado com base nos gráficos dos valores reais *versus* previstos. Conforme mostrado na Figura 24, as amostras de calibração e de previsão estão próximas à regressão e bem distribuídas ao longo da reta.

De acordo com os valores dos coeficientes angulares e linear, observa-se que o ajuste para o modelo SVR tem uma boa correlação entre os dados de referência e os dados calculados pelo modelo uma vez que quanto mais próximo de 1 para a inclinação, e mais próximo de 0 para o intercepto, melhor é a concordância entre os valores. O coeficiente de regressão indica o quanto a reta pode ser adequada com o modelo matemático. A Tabela 3 apresenta o resumo dos valores ótimos para o parâmetro de margem ( $C$ ),  $\epsilon$  é um parâmetro de “tolerância”, Erro Quadrático Médio de Calibração (RMSEC) e Erro Quadrático Médio da Previsão (RMSEP), para os modelos de BioQAVM e BioQAVP.

Tabela 3: Valores dos parâmetros e dos erros para as amostras de BioQAV.

Amostras	Parâmetro $C$	Parâmetro $\epsilon$	RMSEC (% v/v)	RMSEP (% v/v)	R
BioQAVmacaúba	997,23	0,064	0,63	0,81	0,99953
BioQAV palmiste	915,08	0,063	0,20	0,78	0,99993

Pode-se observar um alto valor da constante  $C$  para ambas amostras, essa constante controla os erros de classificação do algoritmo SVM, com esse valor alto de  $C$  a função de classificação torna-se menos suave para que as amostras sejam classificadas erroneamente, podendo gerar um maior número de vetores de suporte.

Os valores de RMSEP e RMSEC obtidos apresentam valores pequenos de forma que fornece maior credibilidade aos modelos. Os valores de RMSEC foram menores que o RMSEP, transmitindo assim uma confiabilidade na calibração, sendo que os valores de calibração devem ser menores que os de previsão, caso contrário, os modelos serão tendenciosos. Os modelos apresentaram uma boa linearidade com valores de R acima de 0,999.

Os resultados dos parâmetros estatísticos do modelo SVR são apresentados na Tabela 4. O modelo SVR apresenta menores erros de previsão (RMSEP) e maiores coeficiente de determinação ( $R^2_p$ ), indicando que o sistema em estudo se comporta de forma linear. A boa relação entre os valores previstos pelo modelo SVR e medidos pelos métodos de referência é observado nos coeficientes de determinação: 0,9979 para validação cruzada do BioQAVM e 0,9990 para o BioQAVP.

Tabela 4: Resultados dos parâmetros do modelo SVR para BioQAVM e BioQAVP.

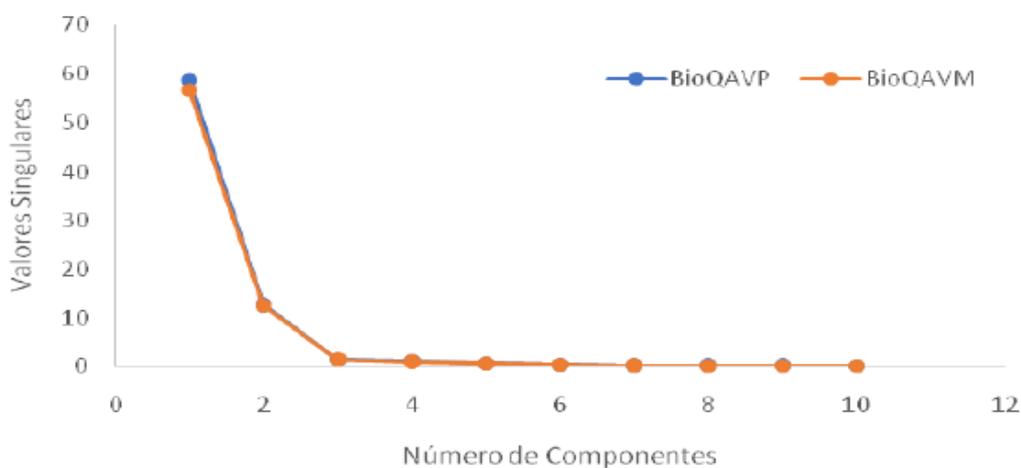
Parâmetros	BioQAVM	BioQAVP
$R^2_{cv}$	0,9979	0,9990
$R^2_p$	0,9982	0,9996

Viés-cv	-0,13	0,02
Viés-p	0,14	0,01
RMSECV	0,98	0,68
RMSEP	0,81	0,78

### 7.3 Modelos MCR-ALS

Para a escolha do número de componentes do sistema utilizou-se o algoritmo SVD (do inglês, *Singular Value Decomposition*), que estima o posto da matriz por meio da percentagem de variância explicada, e as estimativas iniciais dos perfis espectrais foram gerados pelo algoritmo PURE (Método de Detecção da Variável Pura). Foram usados simultaneamente as restrições de correlação e de não-negatividade nas concentrações dos espectros. O número de componentes das misturas foi obtido através do algoritmo SVD, sendo que a escolha foi feita a partir da convergência nos valores de variância. Na Figura 25 tem-se um exemplo de como foi definido o número de componentes. É possível notar que a partir do segundo termo os valores convergem para zero e por isso foram escolhidos 3 fatores. O mesmo procedimento foi realizado com dois conjuntos de amostras de bioquerosene avaliados.

Figura 25: SVD para a determinação do número de componentes do sistema

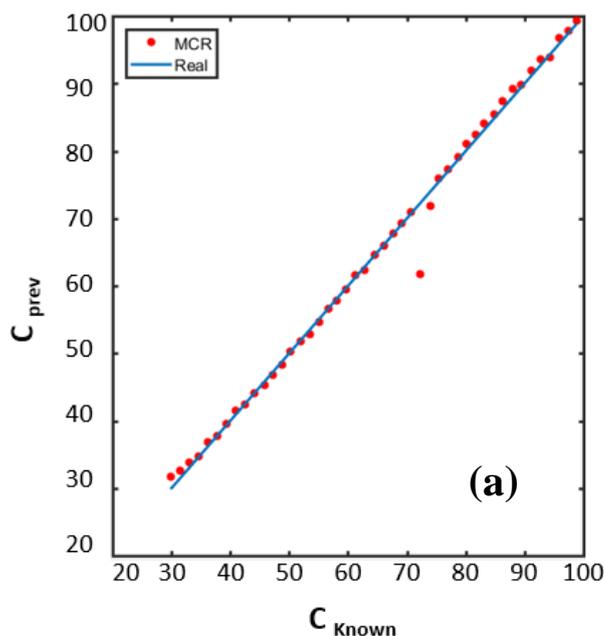


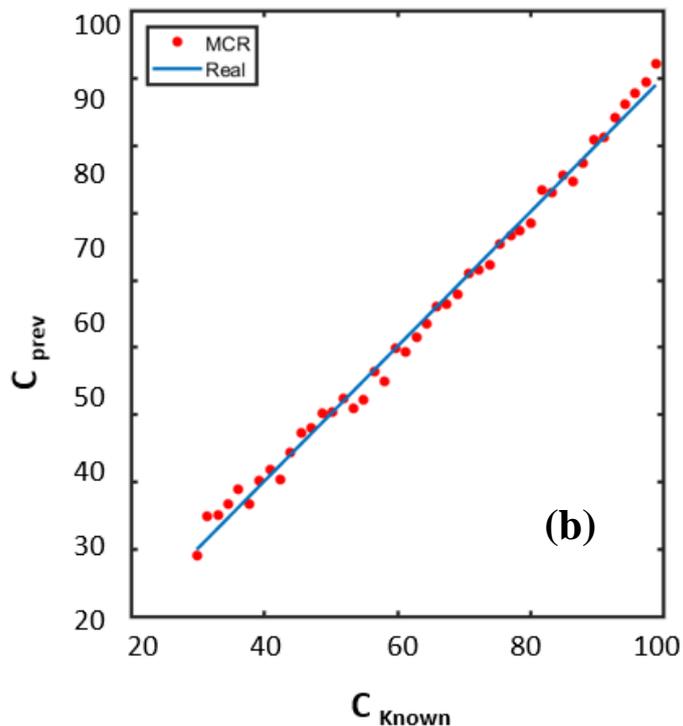
Pode-se observar na Figura 25 que o número ótimo de componentes principais para descrever os modelos são 3 componentes, já que após a terceira componente a percentagem de variância explicada é praticamente constante. Assim será considerado que o posto da matriz de dados para cada amostra de BioQAV é 3, ou seja, os espectros das amostras podem ser obtidos através de combinação linear de três espectros individuais.

Os modelos foram construídos a fim de determinar por FTIR-ATR as concentrações e os perfis espectrais dos componentes presentes das amostras de BioQAV. Os escores do MCR-ALS puderam ser relacionados diretamente com as concentrações das amostras, sendo que a partir de uma regressão linear é possível determinar a equação relacionada. A principal vantagem desse método, é que através dos espectros de FT-MIR foi possível estimar os teores de bioquerosene e querosene utilizando apenas o modelo de calibração.

Na Figura 26, estão apresentados os modelos de calibração pelo MCR-ALS para realizar as determinações quantitativas, foram feitas as regressões lineares entre as intensidades relativas  $C$  e as concentrações de referência das amostras de calibração, para todos os compostos estudados.

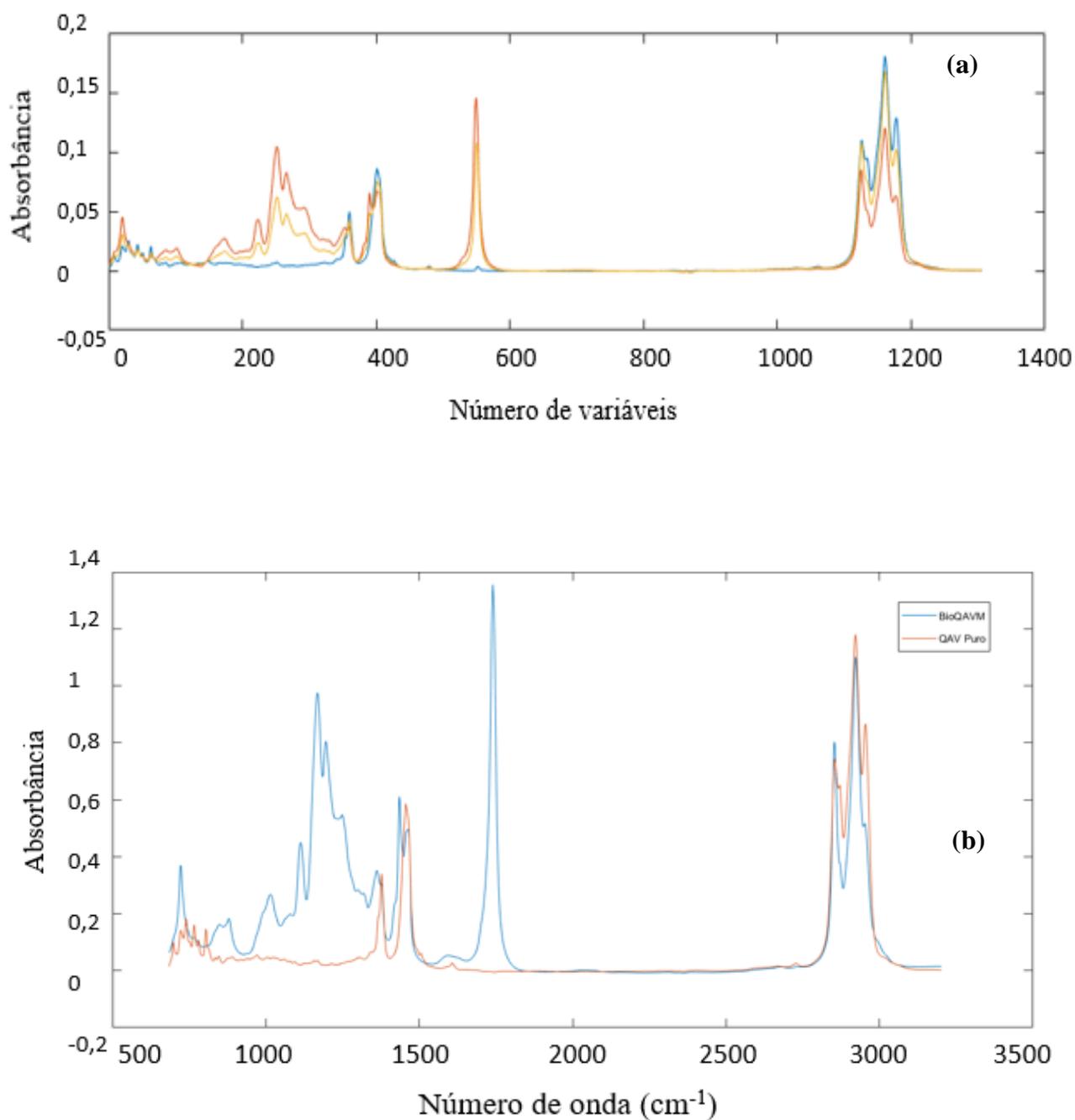
Figura 26: Curvas de calibração para os bioquerosene: (a) macaúba e (b) palmiste.





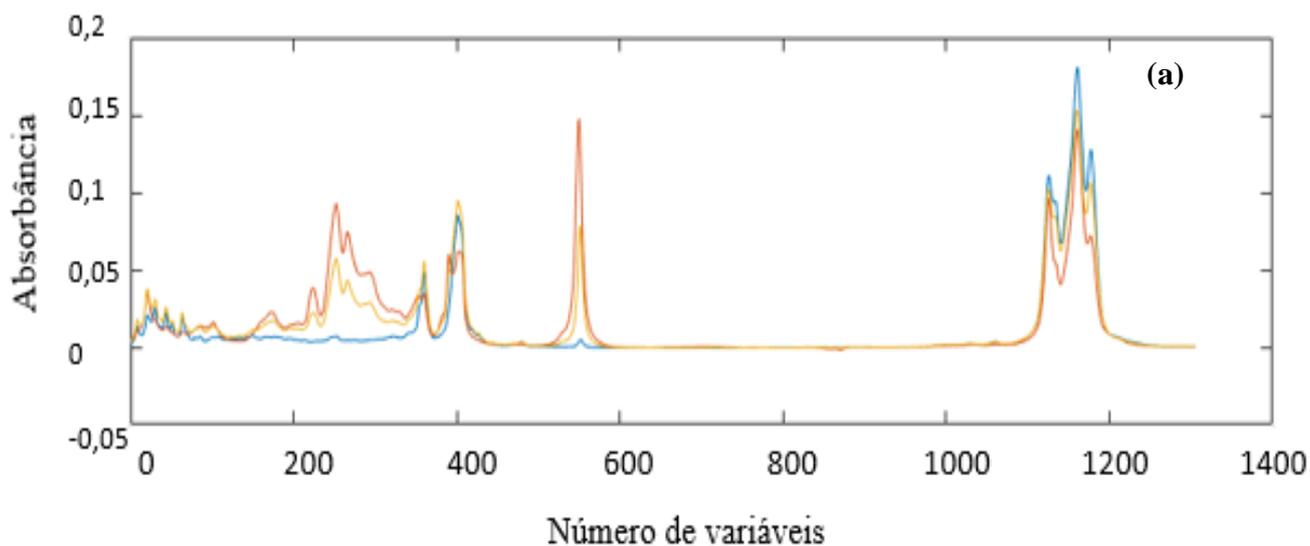
Para conduzir uma análise de MCR-ALS, o número de componentes presentes no sistema analisado e as estimativas iniciais dos perfis de concentração ou espectrais devem ser fornecidas. A recuperação dos espectros é uma ferramenta que proporciona significado e interpretação química dos resultados. Assim, além da quantificação é possível a identificação do adulterante através do espectro recuperado, o que não pode ser realizado utilizando outros métodos quimiométricos como por exemplo, PLS. Na Figura 27 (a) observa-se o espectro da mistura recuperado pelo modelo do MCR-ALS, 27(b) espectro puro de BioQAVM e QAV puro, obtido experimentalmente.

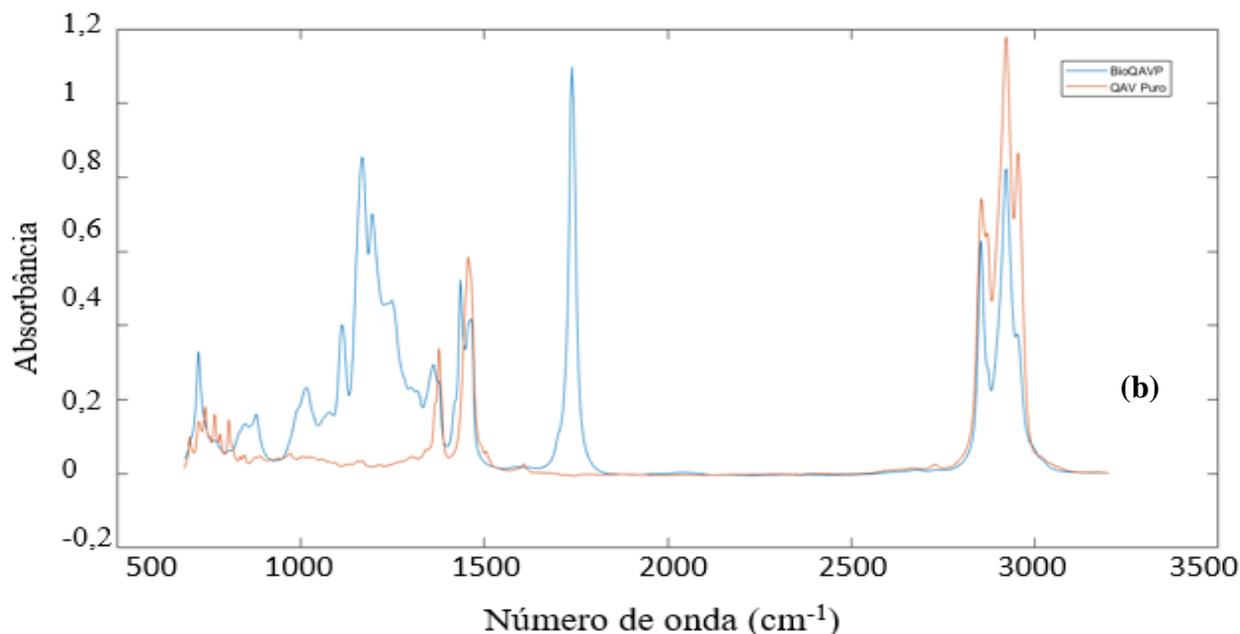
Figura 27: Espectros da mistura BioQAVM/QAV, (a) Perfil da mistura recuperado pelo MCR-ALS e (b) experimental



Fazendo uma comparação visual entre os espectros da Figura 27 (a) com a (b), observa-se bandas de absorção correspondente as três regiões espectrais características: às vibrações de estiramento das ligações C-H dos grupamentos de metila e metileno; bandas de intensidade das vibrações de deformação angular das ligações C-H do grupo metila e metileno, outra banda de baixa intensidade das vibrações de deformação angular assimétrica resultante das deformações C-H do grupo metileno. Os bioquerosenes apresentam em comum padrões de absorção da carbonila (C=O) e a presença de bandas atribuídas às vibrações de estiramento da ligação C-O (BARBOSA, 2008).

Figura 28: Espectros da mistura BioQAVP/QAV, (a) Perfil da mistura recuperado pelo MCR-ALS e (b) experimental





Observando a Figura 28 (a) e (b), o mesmo ocorre para os espectros da mistura de bioquerosene de palmiste/ querosene, estiramento das ligações C-H, deformação angular das ligações C-H e deformação angular assimétrica resultante das deformações C-H. Os bioquerosenes apresentam em comum padrões de absorção da carbonila (C=O) e a presença de bandas atribuídas às vibrações de estiramento da ligação C-O (BARBOSA, 2008).

A otimização do MCR-ALS, não pode ser concluída uma vez que o algoritmo detectou deficiência de posto (*rank*) na matriz de dados. Na prática, o posto de uma matriz corresponde ao número de espécies químicas presentes na mistura. Em uma matriz de espectros esse número correspondente ao número de linhas e colunas linearmente independentes, ou seja, o número de vetores que não podem ser escritos como uma combinação linear dos outros. Uma matriz é considerada de posto completo se o posto estimado for igual ao número de componentes químicos presentes na mistura. No caso de deficiência de posto, isso significa que o posto previamente estimado neste caso igual a dois, referente a mistura de bioquerosene com o querosene, é diferente do número de espécies químicas que o algoritmo conseguiu identificar.

Na tentativa de quebrar a deficiência de posto, foi acrescentado à matriz de interesse os espectros puros, assim, foi fornecido os perfis espectrais ao invés de usar as estimativas do PURE. Isso nos leva a concluir que como o bioquerosene foi obtido a partir do biodiesel, é possível que na esterificação nem todo biodiesel tenha sido

convertido a bioquerosene. Assim sendo, um resíduo do biodiesel original pode estar presente no bioquerosene, e mesmo em concentração pequena foi percebida pelo MCR-ALS, o que explica o terceiro perfil espectral decomposto pelo modelo.

Como pode ser visto na Tabela 5, que também apresenta as figuras de mérito calculadas após a execução da segunda etapa do MCR-ALS, apresenta os resultados de avaliação dos modelos, sendo que ao final do processo iterativo, já são calculados os valores de RMSEC,  $R^2$  e LOF. A avaliação do modelo não é realizada somente com a observação do ajuste linear dos resultados.

Tabela 5: Valores de  $R^2$ , LOF, RMSEP e RMSEC.

Parâmetros	BioQAVM	BioQAVP
$R^2$	0,9995	0,9997
LOF	2,16	1,42
RMSEP (%)	2,39	1,55
RMSEC (%)	1,85	1,39

A variância explicada pelos modelos alcançou valores maiores do que 99,99% em todos os casos. A falta de ajuste também apresentou valores satisfatórios, mantendo-se abaixo para o BioQAVM de 2,16 e 1,42 para o BioQAVP. Em relação à aplicação da restrição de correlação, os valores de concentração obtidos por MCR-ALS ajustam-se bem aos dados de referência, estando os valores de RMSEC baixo para os dois bioquerosene. Esses resultados confirmam a adequação do modelo de calibração do MCR-ALS para o conjunto de dados proposto, mesmo ocorrendo a ocorrência de um terceiro perfil. Os valores de LOF obtidos foram baixos, isso indica uma boa recuperação das matrizes  $C$  e  $S^T$ .

Dessa forma, apesar dos inconvenientes inerentes a natureza do sistema analisado em relação ao método MCR-ALS, as estratégias aplicadas permitiram de forma

satisfatória, a recuperação dos perfis de concentração do bioquerosene de macaúba e palmiste em unidades reais de concentrações.

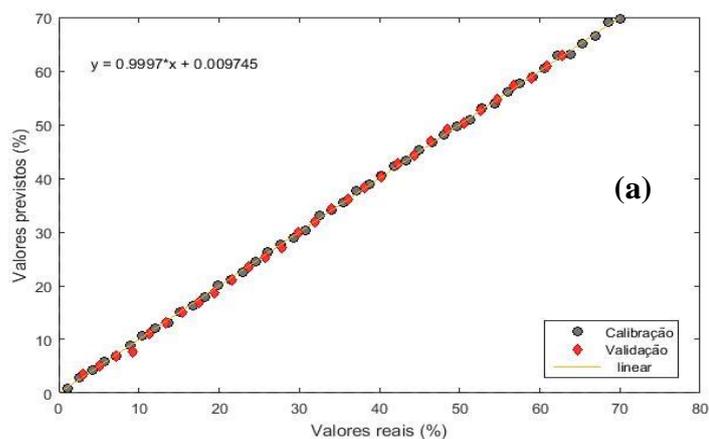
#### 7.4 Modelo PLS e comparação entre SVR

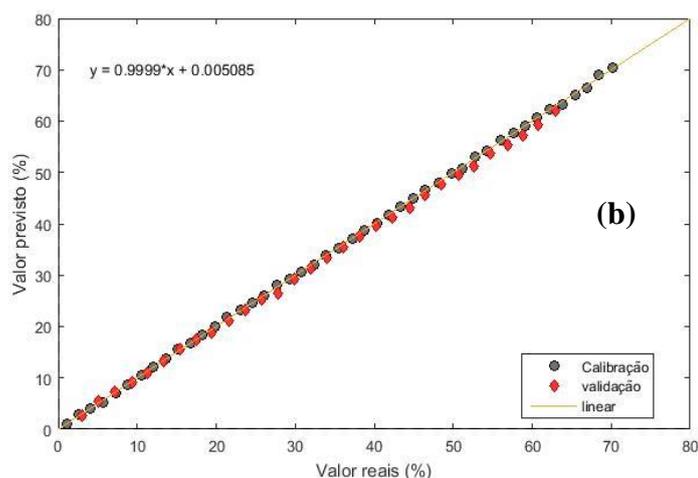
Na literatura há alguns trabalhos que fazem a comparação entre o método dos quadrados mínimos parciais (PLS) e o SVR (ALVES, 2012a), para verificar qual método é mais eficiente na quantificação, demonstram em seu trabalho a aplicação do SVR aliada à espectroscopia de infravermelho próximo para resolver problemas de regressão associados a determinação de parâmetros de qualidade de óleo diesel para um sistema otimizador de misturas em uma refinaria de petróleo. Utilizando o SVR e os resultados obtidos foram comparados com os resultados obtidos pelo algoritmo PLS. Uma otimização paramétrica usando um algoritmo genérico foi realizada para a escolha dos parâmetros nos modelos de regressão de SVR. Os melhores modelos usando SVR apresentavam um Kernel RBF e espectros pré processados com correção de linha de base e dados centrados na média. Os valores obtidos de RMSEP com os modelos SVR é 0,453 para ponto de fulgor e número de cetano. O SVR forneceu resultados significativamente melhores quando comparado com o PLS, especialmente para dados com alta relação não lineares ou complexos. O melhor resultado para o modelo PLS foi obtido usando 3 VL's para explicar uma variância de 99,12%, os resultados apresentam o RMSEC = 4,21 RMSEP = 3,77 e R = 0,698, enquanto no SVR (o melhor modelo) foram obtidos os resultados usando a função de *Kernel* RBF, RMSEC = 1,99, RMSEP = 1,98 e R = 0,936. Para o número de cetano os valores para o modelo PLS os resultados apresentam o RMSEC = 0,74, RMSEP = 0,56 e R = 0,894, enquanto no SVR os resultados de RMSEC = 0,76, RMSEP = 0,45 e R = 0,895.

Os resultados mostram que SVR fornece os melhores modelos de regressão em relação ao PLS, uma vez que o SVR pode modelar linear e não linear presentes nos conjuntos de dados. Por outro lado, o desenvolvimento de modelos SVR requer um maior tempo de trabalho, para obter a otimização paramétrica, que é plenamente justificado pela obtenção de modelos mais eficientes.

Deste modo, os resultados obtidos também foram comparados com o algoritmo PLS a fim de verificar a linearidade da regressão. Os espectros foram pré processados com correção de linha de base e dados centrados na média, o modelo matemático é apresentado na Figura 27 onde se compara os valores reais de concentração pelos valores previstos dos modelos de BioQAVM e BioQAVP, pode-se verificar que os gráficos apresentam uma boa linearidade dos modelos de PLS.

Figura 29: Regressão dos valores reais pelos previstos das misturas de bioquerosene de (a) macaúba e (b) palmiste.





Fonte: a autora.

Para avaliar os modelos desenvolvidos são capazes de prever corretamente os valores das amostras, ou seja, para avaliar a eficiência dos modelos foram analisados a exatidão que é determinada pelos parâmetros RMSEC, RMSEP e RMSECV. A Tabela 7 mostra os parâmetros de exatidão da mistura de bioquerosene/querosene com os valores de variáveis latentes (VL), RMSEC, RMSEP, RMSECV e R.

Tabela 6: Parâmetros de exatidão dos modelos PLS construídos.

Amostras	VL	RMSEC	RMSEP	RMSECV	R
BioQAVM	6	0,34	0,45	0,52	0,9995
BioQAVP	4	0,24	0,85	0,54	0,9997

Nestas condições, avaliando a exatidão dos modelos da Tabela 8, através dos valores de RMSEC, RMSEP e RMSECV, podemos observar que em todos os modelos os valores de erros foram satisfatórios, ou seja, apresentam uma pequena diferença entre eles indicando que não há sobreajuste ou subajuste (GONTIJO, 2016). Para uma melhor

análise entre os modelos de PLS, SVR e MCR-ALS, foi criada a Tabela 8 com os valores dos parâmetros de RMSEC, RMSEP e R.

Tabela 7: Comparação entre os modelos PLS, SVR e MCR-ALS

Amostra	PLS			SVR			MCR-ALS		
	RMSEC	RMSEP	R	RMSEC	RMSEP	R	RMSEC	RMSEP	R
BioQAVM	0,33	0,45	0,9997	0,63	0,81	0,9995	1,85	2,39	0,9995
BioQAVP	0,24	0,85	0,9998	0,20	0,78	0,9999	1,39	1,55	0,9997

Observando os valores obtidos para os três modelos, os valores de RMSEC foram menores em relação ao RMSEP dos dois bioquerosene. Sendo que os valores para os modelos PLS foram mais baixos enquanto que para os modelos MCR-ALS foram mais altos. Em relação aos modelos do MCR-ALS os valores de RMSEC foram menores do que o RMSEP. Os modelos apresentam uma boa linearidade com valores de R acima de 0,99.

Os modelos gerados por análises multivariadas podem ser comparados utilizando os valores de RMSEP gerados e a distribuição de Fisher-Snedecor (teste F). Se o valor de  $F_{cal}$  for maior que o valor tabelado da distribuição de Fisher-Snedecor com graus de liberdade igual ao número de amostras de predição e um nível de significância de 5%, temos que os modelos comparados possuem diferença estatística. Os modelos gerados por SVR, MCR-ALS e PLS foram através da estatística  $F$ , como pode ser observado na Tabela 9.

Tabela 8: Resultados obtidos pelo teste F

Amostras	RMSEP		Teste F
BioQAVM	SVR	0,81	3,66

			(SVR/MCR-ALS)
	MCR-ALS	1,55	11,86 (MCR-ALS/PLS)
	PLS	0,45	3,24 (SVR/PLS)
BioQAVP	SVR	0,75	3,43 (SVR/MCR-ALS)
	MCR-ALS	1,39	2,67 (MCR-ALS/PLS)
	PLS	0,85	1,28 (SVR/PLS)

\*Em relação ao valor tabelado. ( $F(0.5;40;40) = 1,693$ )

A Tabela 9, resume uma comparação dos modelos obtidos nesse trabalho. Observando o valor de  $F$  tabelado e os valores obtidos entre os modelos, pode-se notar que entre SVR e PLS do BioQAVP não tem diferença estatística entre os modelos, o valor calculado pelo teste  $F$  de 1,28 é próximo ao valor tabelado de 1,693, portanto eles possuem a mesma eficiência de previsão. Entretanto outros modelos SVR/MCR-ALS, MCR-ALS/PLS, SVR/PLS, SVR/MCR-ALS, MCR-ALS/PLS possuem diferença estatística quando comparado ao valor de tabelado.

## 8 Conclusão

Foi possível desenvolver métodos analíticos através da técnica de espectroscopia no infravermelho médio aliada a calibração multivariada, capazes de quantificar de maneira satisfatória o teor de bioquerosenes presentes nas misturas com querosene.

Os parâmetros dos modelos SVR obtidos para ambas misturas de bioquerosenes tiveram um valor alto da constante  $C$  e  $\epsilon$  baixos, com esses valores o algoritmo SVR torna-se menos suave para que as amostras sejam modeladas erroneamente.

Os valores de RMSEC foram menores que o RMSEP, transmitindo assim uma confiabilidade na calibração, sendo que os valores de calibração devem ser menores que a previsão, caso contrário, os modelos serão tendenciosos. Os modelos apresentaram uma boa linearidade com valores de  $R$  acima de 0,999.

Além disso, o método de regressão por vetores suporte apresenta resultados semelhantes ao método de calibração multivariada por PLS, que é amplamente aplicado em química analítica, mostrando que o método SVR apresenta-se como uma alternativa para tal quantificação.

No estudo da aplicação de MCR-ALS, algumas estratégias foram adotadas com o objetivo de melhorar a obtenção dos perfis por MCR-ALS, como as restrições de não-negatividade e restrição de correlação, porém o número ótimo de componentes principais fornecidas pelo modelo foram 3, no qual se esperava duas componentes (BioQAV e QAV). A variância explicada pelos modelos alcançou valores maiores do que 99,99% em todos os casos. A falta de ajuste também apresentou valores bastante satisfatórios, mantendo-se baixo para o BioQAVM de 2,16 e 1,42 para o BioQAVP. Os valores de RMSEC foram menores em relação ao RMSEP dos dois bioquerosene.

## 9 Referências

ALVES, J. C. L. **Máquina de vetores de suporte aplicada a dados de espectroscopia NIR de combustíveis e lubrificantes para o desenvolvimento de modelos de regressão e classificação.** [s.l.: s.n.].

ALVES, J. C. L. **Máquina de vetores de suporte aplicada a dados de espectroscopia NIR de combustíveis e lubrificantes para o desenvolvimento de modelos de regressão e classificação.** [s.l.] Universidade Estadual de Campinas, 2012b.

ALVES, J. C. L.; POPPI, R. J. Biodiesel content determination in diesel fuel blends using near infrared (NIR) spectroscopy and support vector machines (SVM). **Talanta**, v. 104, p. 155–161, 2013. <https://doi.org/10.1016/j.talanta.2012.11.033>

ANAC- **AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL.** Disponível em: <<https://www.anac.gov.br/assuntos/paginas-tematicas/meio-ambiente/combustiveis-sustentaveis-para-a-aviacao>>. Acesso em: 27 fev. 2020.

ANP- **AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (BRASIL).** Disponível em: <<http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2014/dezembro&item=rانp-63-2014>>. Acesso em: 27 fev. 2020.

ANP- **AGÊNCIA NACIONAL DO PETRÓLEO GÁS NATURAL E BIOCOMBUSTÍVEIS (BRASIL).** Disponível em: <<https://www.diariodasleis.com.br/legislacao/federal/223751-especificaues-dos-querosenes-de-aviauuo-alternativos-e-de-suas-misturas-com-o-querosene-de-aviauuo-qav-1- ficam-estabelecidas-por-meio-da-presente-resoluuuo-as-especificau.html>>. Acesso em: 20 fev. 2020.

BARBOSA, L. C. D. A. **Espectrometria no Infravermelho na caracterização de compostos orgânicos.** Viçosa: [s.n.].

BOGRES NETO, W. Parâmetros de qualidade de lubrificantes e óleo de oliva através de espectroscopia vibracional, calibração multivariada e seleção de variáveis. p. 130, 2005.

BONESSO, D. **Estimação dos Parâmetros do Kernel em um Classificador SVM na Classificação de Imagens Hiperespectrais em uma Abordagem Multiclasse.** [s.l.: s.n.].

BORIN, A.; POPPI, R. J. Multivariate quality control of lubricating oils using fourier transform infrared spectroscopy. **Journal of the Brazilian Chemical Society**, v. 15, n. 4, p. 570–576, 2004. <https://doi.org/10.1590/S0103-50532004000400020>

BRASIL. **Agência Nacional de Petróleo, Gás Natural e Biocombustíveis-ANP. Resolução ANP nº20. Regulamenta as especificações dos Querosene de aviação Alternativos e suas misturas com o Querosene de Aviação (QAV-1).**

BRERETON, R. G. Introduction to multivariate calibration in analytical chemistry. **Analyst**, v. 125, n. 11, p. 2125–2154, 2000. <https://doi.org/10.1039/b003805i>

BROOKS, K. P. et al. **Low-Carbon Aviation Fuel Through the Alcohol to Jet**

- Pathway**. [s.l.] Elsevier Inc., 2016. <https://doi.org/10.1016/B978-0-12-804568-8.00006-8>
- BRUDZEWSKI, K. et al. Gasoline quality prediction using gas chromatography and FTIR spectroscopy: An artificial intelligence approach. **Fuel**, v. 85, n. 4, p. 553–558, 2006. <https://doi.org/10.1016/j.fuel.2005.07.019>
- BURGES, C. J. C. A Tutorial on Support Vector Machines for Pattern Recognition. **Data Mining and Knowledge Discovery**, v. 2, p. 121–167, 1998. <https://doi.org/10.1023/A:1009715923555>
- C, C. Kernel methods: a survey of current techniques. **Journal of Natural Gas Science and Engineering**, v. 48, n. October, p. 103237, 2002.
- CARLA FELIPPI CHIELLA RUSCHEL. **Aplicação de ferramentas quimiométricas e técnicas espectroscópicas na análise de combustíveis**. [s.l.] Universidade Federal do Rio Grande do Sul, 2017.
- CGEE. Biocombustíveis aeronáuticos: Progressos e desafios. **Série Documentos Técnicos**, v. 08, p. 1–55, 2010.
- CHIARAMONTI, D. et al. Sustainable bio kerosene: Process routes and industrial demonstration activities in aviation biofuels. **Applied Energy**, v. 136, p. 767–774, 2014. <https://doi.org/10.1016/j.apenergy.2014.08.065>
- CHIONG, M. C. et al. Liquid biofuels production and emissions performance in gas turbines: A review. **Energy Conversion and Management**, v. 173, n. August, p. 640–658, 2018. <https://doi.org/10.1016/j.enconman.2018.07.082>
- COLOMBO, C. A. et al. The oil & protein crop supply chain in south america la filière oléoprotéagineuse en amérique du sud. **EDP Sciences**, v. 25, n. 1, p. 9, 2018. <https://doi.org/10.1051/ocl/2018007>
- CORDEIRO DANTAS, W. F.; LAURENTINO ALVES, J. C.; POPPI, R. J. MCR-ALS with correlation constraint and Raman spectroscopy for identification and quantification of biofuels and adulterants in petroleum diesel. **Chemometrics and Intelligent Laboratory Systems**, v. 169, p. 116–121, 2017. <https://doi.org/10.1016/j.chemolab.2017.04.002>
- CORTES, C.; VAPNIK, V. N. Support-Vector Networks. **Machine Learning**, v. 20, p. 273–297, 1995. <https://doi.org/10.1007/BF00994018>
- CORTEZ, L. A. B. et al. **Plano de voo para biocombustíveis de aviação no Brasil: Plano de ação Fapesp**. [s.l.: s.n.].
- DA SILVA, J. Q. **Biodieseis leves derivados dos óleos da macaúba e do palmiste: propriedade de misturas com o querosene na perspectiva de um combustível alternativo de aviação**. [s.l.] Universidade Federal de Uberlândia, 2019.
- DATAR, R. P. et al. Fermentation of biomass-generated producer gas to ethanol. **Biotechnology and Bioengineering**, v. 86, n. 5, p. 587–594, 2004. <https://doi.org/10.1002/bit.20071>
- DE JUAN, A.; JAUMOT, J.; TAULER, R. Multivariate Curve Resolution (MCR). Solving the mixture analysis problem. **Analytical Methods**, v. 6, n. 14, p. 4964–4976, 2014. <https://doi.org/10.1039/C4AY00571F>

- DE JUAN, A.; TAULER, R. Multivariate Curve Resolution (MCR) from 2000: Progress in concepts and applications. **Critical Reviews in Analytical Chemistry**, v. 36, n. 3–4, p. 163–176, 2006. <https://doi.org/10.1080/10408340600970005>
- 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>
- DEBUS, B. et al. Three-point multivariate calibration models by correlation constrained MCR-ALS: A feasibility study for quantitative analysis of complex mixtures. **Talanta**, v. 163, p. 39–47, 2017. <https://doi.org/10.1016/j.talanta.2016.10.081>
- DOS SANTOS, F. I. V. Bioquerosene de Aviação: Panorama e perspectivas do biocombustível. **Universidade Estadual de Campinas**, p. 1–47, 2015.
- EMBRAPA- EMPRESA BRASILEIRA DE PESQUISA AGROPECUÁRIA.** Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/3344909/artigo-materias-primas-oleaginosas-para-a-producao-de-bioquerosene--oportunidades-e-desafios>>. Acesso em: 27 fev. 2020.
- EMBRAPA-EMPRESA BRASILEIRA DE PESQUISA.** Disponível em: <<https://www.embrapa.br/busca-de-noticias/-/noticia/2329636/macaba-e-materia-prima-promissora-para-biodiesel>>. Acesso em: 27 fev. 2020.
- ENVIRO. AERO. Beginner's Guide to Aviation Biofuels. v. Edition 2, 2011.
- EQUATORIANO, D. O Mercado Brasileiro Para Óleo De Plama ( Azeite De Dendê) Equatoriano. 2006.
- FAHRENKROG, J.; POKAR, J.; NOHL, G. Therapie einer tragenden Stute mit dekompensierter Herzinsuffizienz. **Praktische Tierarzt**, v. 98, n. 11, p. 1154–1164, 2017.
- FERNANDA PIMENTEL, M. et al. Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. **Microchemical Journal**, v. 82, n. 2, p. 201–206, 2006. <https://doi.org/10.1016/j.microc.2006.01.019>
- FERRÃO, M. F. et al. Simultaneous determination of quality parameters of biodiesel/diesel blends using HATR-FTIR spectra and PLS, iPLS or siPLS regressions. **Fuel**, v. 90, n. 2, p. 701–706, 2011. <https://doi.org/10.1016/j.fuel.2010.09.016>
- FERREIRA, M. M. C. **Quimiometria: conceitos, métodos e aplicações.** Campinas-SP: Editora da Unicamp, 2015.
- FILGUEIRAS, P. R. **Regressão Por Vetores De Suporte Aplicado Na Determinação De Propriedades Físico-Químicas De Petróleo E Biocombustíveis.** [s.l: s.n.].
- FILHO, R.M., MACIEL, M. R. W., BATISTELLA, C. B., SILVA, N. L. DE. **Processo de Produção de Bioquerosene**, 2008.
- GARCIA, C. M. Transesterificação de óleos vegetais. p. 120, 2006.
- GELLER, H. S. **Revolução energética: políticas para um futuro sustentável.** Rio de Janeiro: [s.n.].

GOICOECHEA, H. C.; OLIVIERI, A. C.; TAULER, R. Application of the correlation constrained multivariate curve resolution alternating least-squares method for analyte quantitation in the presence of unexpected interferences using first-order instrumental data. **Analyst**, v. 135, n. 3, p. 636–642, 2010. <https://doi.org/10.1039/b922547a>

GONÇALVES, A. R. **Máquina de Vetores Suporte**. Disponível em: <<http://www.dca.fee.unicamp.br/~andreric/arquivos/pdfs/svm.pdf>, 2010>. Acesso em: 15 ago. 2020.

GONTIJO, L. C. **Uso de espectrometria no infravermelho médio calibração multivariada e seleção de variáveis por intervalos na quantificação de biodieséis em misturas com diesel**. [s.l.] Universidade Federal de Uberlândia, 2016.

GUIMARÃES, E. **Aplicação de espectroscopia no infravermelho médio e calibração multivariada para quantificação de adulterantes em biodieséis de óleo de soja e residual e suas misturas com diesel**. [s.l.] Universidade Federal de Uberlândia, 2014.

HEARST, M. A. et al. Support vector machines. **IEEE Intelligent Systems and their Applications**, v. 13, n. 4, p. 18–28, 1998. <https://doi.org/10.1109/5254.708428>

HOFMANN, T, SCHOLKOPF, B, SMOLA, A. Kernel methods in machine learning. In: **The annals of statistics**. V.36 ed. [s.l.: s.n.]. p. 50. <https://doi.org/10.1214/009053607000000677>

HOLLER, F. J. **Pincípios de análise instrumental**. Porto Alegre: [s.n.].

INTERNATIONAL CIVEL AVIATION ORGANIZATION. **Environmental report 2010**, v. 65, p. 6, 2010.

JANSEN, B. R. J. et al. The influence of experiencing success in math on math anxiety, perceived math competence, and math performance. **Learning and Individual Differences**, v. 24, p. 190–197, 2013. <https://doi.org/10.1016/j.lindif.2012.12.014>

JANSEN, R. A. **Second Generation Biofuels and Biomass: Essential Guide for Investors**. Wiley-VCH: [s.n.].

JAYARAMAN, A. et al. Study of the photodegradation of 2-bromophenol under UV and sunlight by spectroscopic, chromatographic and chemometric techniques. **Journal of Chromatography B: Analytical Technologies in the Biomedical and Life Sciences**, v. 910, p. 138–148, 2012. <https://doi.org/10.1016/j.jchromb.2012.03.038>

KAROUI, R.; DOWNEY, G.; BLECKER, C. Mid-infrared spectroscopy coupled with chemometrics: A tool for the analysis of intact food systems and the exploration of their molecular structure-quality relationships-A review. **Chemical Reviews**, v. 110, n. 10, p. 6144–6168, 2010. <https://doi.org/10.1021/cr100090k>

LANES, E. C. M. DE; MOTOIKE, S. Y.; COSTA, P. M. DE A. Funding: Reassess Chinese research awards. **Nature**, v. 511, n. 7507, p. 31, 2014. <https://doi.org/10.1038/511031b>

LIMA, C. A. DE M. **Comitê de Máquina: uma abordagem unificada empregando máquinas de vetores-suporte**. [s.l.] Universidade Estadual de Campinas, 2004.

LIU, G.; YAN, B.; CHEN, G. Technical review on jet fuel production. **Renewable and Sustainable Energy Reviews**, v. 25, p. 59–70, 2013.

<https://doi.org/10.1016/j.rser.2013.03.025>

LLAMAS, A. et al. Biokerosene from babassu and camelina oils: Production and properties of their blends with fossil kerosene. **Energy and Fuels**, v. 26, n. 9, p. 5968–5976, 2012. <https://doi.org/10.1021/ef300927q>

LORENA, A. C.; CARVALHO, A. C. P. L. F. DE. **Uma introdução às Support Vector Machine**. Disponível em: <<https://pdfs.semanticscholar.org/8469/81890b81ae5d9ff5cedfcd99150a8bde13.pdf>>. Acesso em: 13 ago. 2020.

MAPA- **Ministério da Agricultura**. Disponível em: <<https://www.gov.br/agricultura/pt-br/vegetal/culturas/soja>>. Acesso em: 22 fev. 2020.

MÁQUINA, A. D. V. **Espectrometria no infravermelho médio e métodos quimiométricos PLS-DA e PLS: classificação e previsão do teor de biosiel na mistura de biosiesel/diesel de mafurra, moringa e algodão**. [s.l.] Universidade Federal de Uberlândia, 2017.

MARÇO, P. H. et al. Resolução multivariada de curvas com mínimos quadrados alternantes: Descrição, funcionamento e aplicações. **Química Nova**, v. 37, n. 9, p. 1525–1532, 2014.

MARETTO, D. A. **Aplicação de máquinas de vetores suporte para desenvolvimento de modelos de classificação e calibração multivariada em espectroscopia no infravermelho**. [s.l.] Universidade Estadual de Campinas, 2011.

MARZULLO, R. C. M. Análise de Ecoeficiência dos Óleos Vegetais Oriundos da Soja e Palma , visando a Produção de Biodiesel. p. 303, 2007.

MEDEIROS, L. **Macaúba pode ser alternativa para bioquerosene e solução para pastagem degradadas no Brasil**.

MELO, P. G. **UNIVERSIDADE FEDERAL DE UBERLÂNDIA PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA “PRODUÇÃO E CARACTERIZAÇÃO DE BODIESEIS OBTIDOS A PARTIR DA OLEAGINOSA MACAÚBA (Acrocomia aculeata)”**. [s.l.] Universidade Federal de Uberlândia, 2012.

MONTEIRO, R. S. **Avaliação do potencial do óleo de microalgas como matéria-prima na produção de hidrocarbonetos para bioquerosene de aviação Projeto Final de Curso em Engenharia Química como matéria-prima na produção de**. [s.l.] Universidade Federal do Rio de Janeiro, 2020.

MORAES, P. L. **Mundo Educação. Biologia. Vegetais Comestíveis**. Disponível em: <<https://mundoeducacao.uol.com.br/biologia/macauaba.htm>>. Acesso em: 22 maio. 2020.

MÜLLER, A. L. H. et al. Attenuated total reflectance with Fourier transform infrared spectroscopy (ATR/FTIR) and different PLS algorithms for simultaneous determination of clavulanic acid and amoxicillin in powder pharmaceutical formulation. **Journal of the Brazilian Chemical Society**, v. 22, n. 10, p. 1903–1912, 2011. <https://doi.org/10.1590/S0103-50532011001000011>

NEGRI, R. G. **MÁQUINA DE VETORES DE SUPORTE ADAPTATIVA AO CONTEXTO : FORMALIZAÇÃO E APLICAÇÕES** EM Tese de Doutorado do

**Curso de Siqueira Sant ' Anna , aprovada em.** [s.l.] Instituto nacional de pesquisas espaciais, 2013.

OLIVEIRA, F. C. C.; SUAREZ, P. A. Z.; SANTOS, W. L. P. Biodiesel: Possibilidades e Desafios. **Química Nova na Escola**, p. 3–8, 2008.

PARENTE, E. **Lipofuels: Biodiesel and Biokerosene.** Disponível em: <www.inmetro.gov.br/metcientifica/.../Lipofuels.ppt>. Acesso em: 15 jun. 2020.

RABELO, L. D. T. **Avaliação automática do resultado estético do tratamento conservador do cancro da mama.** [s.l.] Faculdade de Engenharia da Universidade do Porto, 2008.

RAVAZZI, C. G. **Identificação e quantificação de adulterantes de Whey protein concentrado empregando espectroscopia no infravermelho próximo e resolução multivariada de curvas.** [s.l: s.n.].

ROCHA, R. et al. Talanta Application of correlation constrained multivariate curve resolution alternating least-squares methods for determination of compounds of interest in biodiesel blends using NIR and UV – visible spectroscopic data. **Talanta**, v. 125, p. 233–241, 2014. <https://doi.org/10.1016/j.talanta.2014.02.073>

SCHUCHARDT, U.; SERCHELI, R.; MATHEUS, R. Transesterification of Vegetable Oils : a Review General Aspects of Transesterification Transesterification of Vegetable Oils Acid-Catalyzed Processes Base-Catalyzed Processes. **J. Braz. Chem. Soc.**, v. 9, n. 1, p. 199–210, 1998. <https://doi.org/10.1590/S0103-50531998000300002>

SEMI-BATELADA, E. M. S. **Partir Do Triacilglicerol Do Óleo De Palma**, 2009.

SKOOG, D. A.; HOLLER, F. J.; NIEMAN, T. A. **Princípios de Análise Instrumental.** 5ª Edição ed. São Paulo: [s.n.].

SMOLA, A. J.; SCHOLKOPF, B. A tutorial on support vector regression \* ". p. 199–222, 2004. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>

SOARES, I. P. et al. Multivariate calibration by variable selection for blends of raw soybean oil/biodiesel from different sources using Fourier transform infrared spectroscopy (FTIR) spectra data. **Energy and Fuels**, v. 22, n. 3, p. 2079–2083, 2008. <https://doi.org/10.1021/ef700531n>

THEGARID, N. et al. Second-generation biofuels by co-processing catalytic pyrolysis oil in FCC units. **Applied Catalysis B: Environmental**, v. 145, p. 161–166, 2014. <https://doi.org/10.1016/j.apcatb.2013.01.019>

**UBRABIO- UNIÃO BRASILEIRA DO BODIESEL E BIOQUEROSENE.** Disponível em: <http://www.ubrabilio.com.br/1891/noticias/plataformamameiradebioqueroseneseradiscuti da embh\_244120>. Acesso em: 19 fev. 2020.

VANDEGINSTE, B. G. M.; MASSART, D. L.; BUYDENS, L. M. C.; JING, S.; LEWI, P. J.; SMEYERS-VERBEKE, J. **Data Handling in Science and Technology: Part A.** Amsterdã: [s.n.].

VELÁZQUEZ, R. S. G.; KUBOTANI, R. T.; VELÁZQUEZ, S. M. S. G. Novos Combustíveis para a Aviação – Um Estudo de Caso. **Revista Mackenzie de**

**Engenharia e Computação**, v. 12, p. 77–93, 2012.