



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

FLÁVIO MONTENEGRO DE ANDRADE

**MONITORAMENTO EM TEMPO REAL DA PRODUÇÃO DE BIODIESEL
UTILIZANDO UM ESPECTRÔMETRO NIR ULTRACOMPACTO E CONTROLE
ESTATÍSTICO MULTIVARIADO DE PROCESSOS**

Recife

2020

FLÁVIO MONTENEGRO DE ANDRADE

**MONITORAMENTO EM TEMPO REAL DA PRODUÇÃO DE BIODIESEL
UTILIZANDO UM ESPECTRÔMETRO NIR ULTRACOMPACTO E CONTROLE
ESTATÍSTICO MULTIVARIADO DE PROCESSOS**

Dissertação apresentada ao Programa de Pós-Graduação no Departamento de Química Fundamental da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Química.

Área de concentração: Química Analítica

Orientadora: Prof^a. Dr^a. Maria Fernanda Pimentel Avelar

Coorientador: Prof^o. Dr. Neirivaldo Cavalcante da Silva

Recife

2020

Catálogo na fonte
Bibliotecária Mariana de Souza Alves CRB4-2105

A553m Andrade, Flávio Montenegro de.
Monitoramento em tempo real da produção de biodiesel utilizando um espectrômetro NIR ultracompacto e controle estatístico multivariado de processos/ Flávio Montenegro de Andrade. – 2020.
94 f.: il., fig., tab.

Orientadora: Maria Fernanda Pimentel Avelar.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CCEN, Química. Recife, 2020.
Inclui referências e apêndices.

1. Química Analítica. 2. Biodiesel. 3. Espectroscopia NIR. 4. Reação de transesterificação em batelada. I. Avelar, Maria Fernanda Pimentel. (orientadora) II. Título.

543

CDD (22. ed.)

UFPE-CCEN 2020-160

FLÁVIO MONTENEGRO DE ANDRADE

MONITORAMENTO EM TEMPO REAL DA PRODUÇÃO DE BIODIESEL UTILIZANDO UM ESPECTRÔMETRO NIR ULTRACOMPACTO E CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS

Dissertação apresentada ao Programa de Pós-Graduação no Departamento de Química Fundamental da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Química.

Aprovada em: 13/03/2020

BANCA EXAMINADORA

Profa. Maria Fernanda Pimentel Avelar (Orientadora)

Universidade Federal de Pernambuco
Departamento de Engenharia Química

Prof. José Licarion Pinto Segundo Neto

Universidade Federal de Pernambuco
Departamento de Química Fundamental

Profa. Carolina Santos Silva

Universidade Federal de Pernambuco
Departamento de Engenharia Química

**Aos meus pais, meus maiores exemplos de
coragem, superação e altruísmo.**

**Aos meus avós maternos, Severino Montenegro
de Oliveira e Regina Maria de Oliveira, e
paternos, Manoel Gomes de Andrade e Maria
da Conceição de Andrade (*in memoriam*).**

Dedico.

AGRADECIMENTOS

Agradeço primeiramente a Deus pela saúde, pelo dom da vida, e por sempre guiar meus passos nessa jornada chamada vida.

Agradeço especialmente os meus pais Severino e Virgínia, e ao meu irmão Felipe pelo apoio incondicional, ao longo de toda a caminhada acadêmica, sem o qual eu não teria chegado tão longe. Visto que, mesmo perante a dificuldades, sempre se esforçaram para que eu pudesse seguir adiante.

Agradeço aos meus familiares, Almir, Antônio, Elinete, Enir, Isabela, Jacileide e Walter, bem como aos amigos André e Valmir por todo o apoio e contribuições durante o período de pós-graduação. Agradeço também aos demais familiares e amigos que tanto torceram para realização de mais este sonho.

Agradeço com carinho especial à minha namorada Amanda Suellen, por estar ao meu lado, pelos frequentes incentivos e principalmente pela compreensão dos momentos de sacrifício do convívio para dedicação a este trabalho.

Agradeço à Profa. Dra. Maria Fernanda Pimentel, primeiramente, por ter me proporcionado a oportunidade em ter sua valiosa orientação no mestrado, assim como por todo apoio, conselhos e contribuições que foram de suma importância para o desenvolvimento deste trabalho e para o meu crescimento pessoal e profissional.

Agradeço de maneira especial ao meu coorientador Prof. Dr. Neirivaldo Cavalcante e a Dra. Rafaela Sales pela imensa contribuição, bem como pela confiança, paciência e incentivo em todas as fases do trabalho, que muitas vezes apresentou obstáculos, dos quais sozinho, jamais seria capaz de superar.

Agradeço pelo suporte do Laboratório de Combustíveis (LAC) da UFPE, e a contribuição de todos seus integrantes dada a este trabalho direta e indiretamente.

Agradeço aos colegas de pós-graduação: Abinoan, Bárbara, Tatiane, Thiago, Humberto, Carolina, Francielson, Jéssica, Bruno, Aline e Erklaylle por todo apoio, conhecimentos compartilhados e bons momentos de conversas sempre regrados com café e descontração.

Agradeço ao CNPQ, pela concessão da bolsa de mestrado, à CAPES, ao INCTAA, NUQAAPE e FACEPE por todos os auxílios concedidos.

RESUMO

A presente dissertação descreve duas abordagens que utilizam diferentes metodologias aplicadas ao monitoramento da produção de Biodiesel a partir da espectroscopia na região do infravermelho próximo (NIR), associada à quimiometria. Ambas as abordagens descrevem o monitoramento da reação de transesterificação do óleo de algodão refinado, por rota etílica e catálise alcalina, utilizando um espectrômetro NIR ultracompacto *on-line* para o processo em batelada. Na primeira abordagem foi construído um modelo de calibração multivariada PLS a partir dos dados espectrais para prever o teor de etil-ésteres em amostras provenientes da reação de transesterificação. O modelo foi construído a partir de misturas sintéticas obtidas em diferentes concentrações, constituídas por uma mistura quaternária de óleo, glicerina, biodiesel e etanol, e de amostras provenientes de reações em batelada. O modelo apresentou desempenho satisfatório com valor da raiz do erro médio quadrático de predição (RMSEP) de 1,51% m/m e coeficiente de determinação de predição (R^2_{pred}) de 0,98, com a vantagem de se trabalhar com um equipamento portátil, relativo baixo custo e de fácil transporte. A segunda abordagem descreve a utilização da espectroscopia NIR para construção de cartas de controle estatístico multivariado a fim de capturar alterações promovidas na temperatura e concentração do catalisador na produção de biodiesel. Nesta abordagem foi realizada uma avaliação quanto a dois métodos de modelagem de dados de processos em batelada conhecido como: Nomikos-MacGregor (NM) e Wold-Kettaneh-Friden-Holmberg (WKFH), em que o primeiro pôde ser avaliado na condição pós-batelada (após o término da batelada), e o segundo pôde ser avaliado tanto no modo pós-batelada quanto em tempo real. O conjunto de calibração foi construído com 8 bateladas obtidas sob condições normais de operação (NOC) e o conjunto de predição foi composto por 2 bateladas NOC, uma batelada com alteração temporária da temperatura e uma batelada com alteração da concentração de catalisador. As cartas de distância ao modelo (DModX) e T^2 de Hotelling foram utilizadas para a avaliação dos modelos de controle estatístico multivariado, os quais apresentaram resultados semelhantes quanto à capacidade de detecção das falhas intencionalmente provocadas por simulação.

Palavras chave: Biodiesel. Espectroscopia NIR. Reação de transesterificação em batelada. Controle estatístico multivariado de processo. Monitoramento *on-line*. PAT.

ABSTRACT

The present work describes two approaches that use Near Infrared (NIR) spectroscopy and chemometrics, for the monitoring of Biodiesel production. Both approaches describe the on-line monitoring of transesterification reactions of cottonseed oil using a handheld NIR spectrometer (MicroNIR). The reactions were carried out in a batch system, using ethyl alcohol and alkaline catalysts. In the first approach, a Partial Least Square (PLS) multivariate calibration model was built based on the MicroNIR data to predict the ethyl-esters content of samples acquired from the transesterification reactions. The model was built using synthetic mixtures with different concentrations (composed of oil, glycerin, biodiesel and ethanol) and samples from the batch process. A satisfactory performance with root mean square error of prediction (RMSEP) of 1,51% w/w, and prediction determination coefficient (R^2_{pred}) of 0,98 was obtained. In addition, the use of a relatively low cost and easy transportation handheld spectrometer represented an advantage of this application. The second approach describes the use of NIR spectroscopy to build multivariate statistical control charts to detect disturbances in the temperature and catalyst content intentionally induced during biodiesel production. In this approach, two different methods to handle batch process data, referred to as Nomikos-MacGregor (NM) and Wold-Kettaneh-Friden-Holmberg (WKFH), were evaluated. While the former was used for off-line monitoring, the second method was employed for both off-line and on-line process monitoring. The training dataset was formed by eight batches under normal operating condition (NOC). The test data consisted of two NOC batches, one batch to which a disturbance in the temperature was induced, and other batch produced with a different catalyst content. Distance to the model (DmodX) and Hotelling's T^2 control charts were used to evaluate the performance of the models. In general, they presented a similar performance in relation to their ability of detect the simulated failures.

Keywords: Biodiesel. NIR spectroscopy. Batch transesterification reaction. Multivariate statistical process control. On-line monitoring. PAT.

LISTA DE FIGURAS

Figura 1 –	Evolução da produção de biodiesel (B100) no Brasil de 2009-2018.....	18
Figura 2 –	Regiões do espectro eletromagnético.....	24
Figura 3 –	Diagrama de bandas de absorção na região do infravermelho próximo.....	25
Figura 4 –	Resultados da derivação espectral. A linha verde corresponde ao erro aditivo e a vermelha ao multiplicativo.....	30
Figura 5 –	Projeção das amostras sobre os eixos de máxima variância (PCs)	32
Figura 6 –	Agrupamento dos dados oriundos do monitoramento de N bateladas.....	41
Figura 7 –	Desdobramento da matriz tridimensional por meio da abordagem NM.....	43
Figura 8 –	Desdobramento da matriz tridimensional por meio da abordagem WKFH....	44
Figura 9 –	Rearranjo dos escores do modelo PLS para construção dos gráficos de controle exemplificado para as duas primeiras VL.....	45
Figura 10 –	Sistema experimental utilizado para obtenção das bateladas.....	49
Figura 11 –	Ilustração do (a) equipamento portátil MicroNIR 1700 e (b) equipamento portátil acoplado ao acessório de transmitância.....	51
Figura 12 –	Representação do sistema operacional para aquisição dos espectros para as misturas sintéticas.....	53
Figura 13 –	Espectros NIR obtidos durante uma batelada típica de produção de biodiesel pela rota etílica (a) espectros brutos (ordem de aquisição espectral de baixo para cima). (b) último espectro adquirido da batelada.....	57
Figura 14 –	Espectros brutos de 54 misturas sintéticas (espectros em preto), e espectro médio dos 10 últimos espectros das bateladas de 1 a 11 (em vermelho) para a reação de transesterificação do óleo de algodão por rota etílica das bateladas.....	58
Figura 15 –	Espectros pré-processados com (a) SNV e (b) MSC.....	59
Figura 16 –	Espectros pré-processados com 1ª derivada com filtro de Savitzky-Golay e ajuste de polinômio de 2º grau utilizando (a) uma janela de 5 pontos, (b) uma janela de 7 pontos, (c) uma janela de 9 pontos.....	60
Figura 17 –	Faixas espectrais selecionadas pelo algoritmo <i>i</i> PLS (barras verdes) para modelos de EE.....	62
Figura 18 –	Parâmetros do modelo pré-processado com SNV para determinação do teor de EE (a) vetores de regressão; (b) valores preditos versus observados; (c) gráfico de influências.....	65

Figura 19 –	Parâmetros do modelo pré-processado com MSC para determinação do teor de EE (a) vetores de regressão; (b) valores preditos versus observados; (c) gráfico de influências.....	66
Figura 20 –	Figura 20 – Parâmetros do modelo pré-processado com 1ª derivada e janela de 5 pontos/ <i>i</i> PLS para determinação do teor de EE (a) vetores de regressão; (b) valores preditos versus observados; (c) gráfico de influências.....	66
Figura 21 –	Modelo PLS/ <i>i</i> PLS (1ª derivada de Savitzky-Golay com janela de 5 pontos) construído na faixa de concentração de 31 – 60% para teor de ésteres etílicos (a), e gráfico dos vetores de regressão (b).....	68
Figura 22 –	Espectros de uma batelada comum em condições normais de operação pré-processados com (a) MSC (b) SNV, (c) primeira derivada Savitzky-Golay..	71
Figura 23 –	Cartas de controle de (a) DModX, e (b) T ² de Hotelling, para as bateladas 4, 8, 11 e 12 submetidas a perturbações na temperatura e concentração do catalisador (NaOH)	73
Figura 24 –	Gráficos de controle de distância ao modelo para as bateladas de predição, utilizando-se a abordagem (a) NM e (b) WKFH. O símbolo (\$) à frente da descrição das bateladas para a abordagem WKFH indica que o gráfico foi construído a partir dos escores do modelo PLS, e não pelos dados originais..	76
Figura 25 –	Gráficos de controle de T ² de Hotelling para as bateladas de predição, utilizando-se a abordagem (a) NM e (b) WKFH. O símbolo (\$) à frente da descrição das bateladas para a abordagem WKFH indica que o gráfico foi construído a partir dos escores do modelo PLS, e não pelos dados originais..	77

LISTA DE TABELAS

Tabela 1 –	Classificação do biodiesel em função da origem da matéria-prima.....	20
Tabela 2 –	Faixas espectrais da radiação infravermelha.....	24
Tabela 3 –	Perturbações simuladas na preparação e durante as bateladas de predição.....	55
Tabela 4 –	Figuras de mérito para os modelos PLS para estimar o teor de EE construídos utilizando a faixa espectral completa.....	61
Tabela 5 –	Parâmetros dos modelos PLS para o teor de EE construídos com os intervalos espectrais selecionados pelo algoritmo <i>i</i> PLS.....	63
Tabela 6 –	Resultados para a conversão de etil-ésteres estimados pelo modelo PLS e o obtido por análise cromatográfica para amostras do conjunto de predição.....	70
Tabela 7 –	Resultados para as avaliações dos modelos multivariados em tempo real e pós-batelada.....	72

LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional de Petróleo, Gás Natural e Biocombustíveis
ASTM	<i>American Society for Testing and Materials</i>
CNPE	Conselho Nacional de Política Energética
CV	Validação Cruzada (<i>Cross-Validation</i>)
DModX	Distância ao modelo
EM	<i>European Standard</i>
FDA	<i>Food and Drug Administration</i>
FFA	Ácidos graxos livres (<i>free fatty Acids</i>)
FID	Detector de ionização de chama (<i>Flame Ionization Detector</i>)
FIR	Infravermelho distante (<i>Far Infrared</i>)
GC	Cromatografia Gasosa (<i>Gas Chromatography</i>)
HPLC	Cromatografia Líquida de Alta Eficiência (<i>High Performance Liquid Chromatography</i>)
iPLS	Mínimos Quadrados Parciais por Intervalos (<i>Interval Partial Least Squares</i>)
IR	Infravermelho (<i>Infrared</i>)
LV	Variáveis Latentes (<i>Latent Variables</i>)
LVF	Filtro variável linear (<i>Linear Variable Filter</i>)
MEMS	Sistemas microeletromecânicos (<i>Microelectromechanical Systems</i>)
MIR	Infravermelho Médio (<i>Middle Infrared</i>)
MSC	Correção de Espalhamento Multiplicativo (<i>Multiplicative Scatter Correction</i>)
MSPC	Controle Estatístico Multivariado de Processos (<i>Multivariate Statistical Process Control</i>)
NIR	Infravermelho Próximo (<i>Near Infrared</i>)
NOC	Condições Normais de Operação (<i>Normal Operating Condition</i>)
PAT	Tecnologias Analíticas de Processos (<i>Process Analytical Technology</i>)
PC	Componente Principal (<i>Principal Component</i>)
PCA	Análise de Componentes Principais (<i>Principal Component Analysis</i>)
PLS	Mínimos Quadrados Parciais (<i>Partial Least Squares</i>)
RMSE	Raiz Quadrada do Erro Médio Quadrático (Root Mean Square Error)
RMSEC	Raiz Quadrada do Erro Médio Quadrático de Calibração (<i>Root Mean Square Error of Calibration</i>)

RMSECV	Raiz Quadrada do Erro Médio Quadrático de Validação-Cruzada (<i>Root Mean Square Error of Cross-Validation</i>)
RMSEP	Raiz Quadrada do Erro médio Quadrático de Predição (<i>Root Mean Squared of Prediction</i>)
SNV	Variação Normal Padrão (<i>Standart Normal Variate</i>)
SPE	Erro Quadrático de Predição (<i>Squared Prediction Error</i>)
SPXY	<i>Sample Set Partioning based on Joint x-y Distances</i>
SDV	Desvio Padrão de Validação (<i>Standard Deviation of Validation</i>)

SUMÁRIO

1	INTRODUÇÃO	15
2	FUNDAMENTAÇÃO TEÓRICA	18
2.1	BIODIESEL: UM COMBUSTÍVEL RENOVÁVEL	18
2.1.1	Produção do biodiesel.....	19
2.2	TECNOLOGIAS ANALÍTICAS DE PROCESSO E A ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO PRÓXIMO	22
2.2.1	Tecnologias analíticas de processo.....	22
2.2.2	Espectroscopia na região do infravermelho próximo (NIR).....	23
2.3	QUIMIOMETRIA.....	27
2.3.1	Pré-processamento de dados espectroscópicos.....	27
2.3.1.1	Pré-processamento nas variáveis.....	28
2.3.1.2	Pré-processamento nas amostras.....	28
2.3.2	Análise de Componentes Principais (PCA)	31
2.3.3	Calibração multivariada	33
2.3.3.1	Regressão pelos mínimos quadrados parciais PLS	33
2.3.3.2	Validação de modelos de regressão multivariados e figuras de mérito	34
2.3.3.3	Seleção de variáveis pelo algoritmo <i>interval</i> PLS	38
2.3.3.4	Seleção de amostras pelo algoritmo SPXY	39
2.4	CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS EM BATELADA BASEADO EM VARIÁVEIS LATENTES	39
2.4.1	Abordagens de desdobramento da matriz de dados espectroscópicos	41
2.4.1.1	Desdobramento e análise da matriz tridimensional baseado na abordagem NM	42
2.4.1.2	Desdobramento e análise da matriz tridimensional baseado na abordagem WKFH	43
2.4.2	Diagnóstico de falhas em processos em batelada.....	45
2.4.3	Controle estatístico multivariado de processos como uma ferramenta de PAT	47
3	MATERIAIS E MÉTODOS	49
3.1	MONTAGEM DO SISTEMA REACIONAL	49
3.2	PRODUÇÃO DE BIODIESEL EM BATELADA	50
3.2.1	Aquisição dos espectros NIR.....	51
3.2.2	Purificação do biodiesel produzido.....	51
3.2.3	Análise cromatográfica	52
3.3	MISTURAS SINTÉTICAS	52

3.3.1	Tratamento dos dados e construção do modelo multivariado	53
3.4	CONTROLE ESTATÍSTICO MULTIVARIADO DO PROCESSO DE TRANSESTERIFICAÇÃO EM BATELADA	54
3.4.1	Bateladas em Condição Normal de Operação: Etapa de calibração	55
3.4.2	Etapa de predição.....	55
3.4.3	Desenvolvimento das cartas de controle multivariadas	55
3.5	SOFTWARES	56
4	RESULTADOS E DISCUSSÃO	57
4.1	CALIBRAÇÃO MULTIVARIADA	58
4.1.1	Pré-processamento dos espectros NIR.....	58
4.1.2	Construção do modelo de calibração multivariada para o teor de etil-ésteres (EE).....	60
4.1.3	Construção de modelos PLS utilizando a faixa espectral completa	60
4.1.4	Modelos PLS construídos utilizando o algoritmo de seleção de variáveis iPLS.	62
4.2	CONTROLE ESTATÍSTICO MULTIVARIADO DO PROCESSO DE TRANSESTERIFICAÇÃO EM BATELADA	70
4.2.1	Construção dos gráficos de controle multivariados de processo	71
4.2.1.1	Avaliação da detecção de falhas via análise em tempo real pela abordagem WKFH	72
4.2.1.2	Avaliação da detecção de falhas via análise pós-batelada.....	74
5	CONCLUSÃO	79
6	PERSPECTIVAS FUTURAS	81
	REFERÊNCIAS	82
	APÊNDICE A – TABELA DO PLANEJAMENTO DAS MISTURAS SINTÉTICAS	88
	APÊNDICE B – CORRELAÇÃO ENTRE AS FAIXAS ADOTAS PELO iPLS E OS RESPECTIVOS COMPRIMENTOS DE ONDA PARA DIFERENTES PRÉ- PROCESSAMENTOS UTILIZADOS	90
	APÊNDICE C – GRÁFICOS DE CONTROLE DO CONJUNTO DE CALIBRAÇÃO REFERENTE AS ABORDAGENS DE DESDOBRAMENTO NM E WKFH PARA ANÁLISE PÓS-BATELADA E EM TEMPO REAL	92

1 INTRODUÇÃO

O setor de transportes representa uma das maiores fontes antrópicas de emissão a nível mundial de gases que agravam o efeito estufa, cujo principal contribuinte é o dióxido de carbono (CO₂), o qual é responsável por 23% do total emitido em relação às fontes energéticas. Nesta linha, o transporte rodoviário contribui com cerca de 74%, em média, do total emitido desses gases (YANG et al., 2018).

Associadas ao cenário de poluição ambiental, a disponibilidade finita das fontes de combustíveis fósseis e o surgimento de regras mais rígidas quanto às diretrizes ambientais, motivaram a ampliação pela busca de combustíveis renováveis ao longo dos últimos anos. Nesta perspectiva, conforme Singh et al. (2019), o biodiesel apresenta-se como um biocombustível renovável, de emissão neutra de carbono. Além de ser um combustível biodegradável e atóxico, contribuindo assim para redução na emissão de poluentes, sendo capaz de aliar sustentabilidade ambiental e viabilidade econômica em sua utilização, por meio da geração de desenvolvimento econômico e social ao local de sua produção (VELICKOVIC et al., 2013).

Segundo Velickovic et al. (2013), o método mais utilizado para produção de biodiesel consiste na reação entre triacilgliceróis (compostos majoritários presentes na composição dos óleos vegetais ou gorduras animais), e um álcool de cadeia curta, em presença de catalisador. Em geral, a natureza do catalisador empregado pode ser: ácida, alcalina ou enzimática.

No Brasil, o biodiesel vem ganhando cada vez mais destaque na composição da matriz energética nacional. Conforme a Resolução nº16, de 29 de outubro de 2018, do Conselho Nacional de Política Energética (CNPE), até 2023, a alíquota de biodiesel adicionada ao diesel mineral poderá atingir até 15% em volume (BRASIL, 2018). A fonte de triacilgliceróis mais empregada para a produção de Biodiesel no Brasil atualmente é composta por óleos vegetais, conforme descrito pelo relatório anual da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP, 2019). O número expressivo e cada vez maior de trabalhos científicos voltados para a investigação de rotas alternativas, desenvolvimento de métodos de controle estatístico e, de maneira geral, o aumento da eficiência do processo de produção do biodiesel, mostra a importância desta matriz para o cenário energético mundial (MARJANOVIC et al., 2010; DIAS et al., 2013; SALES et al., 2016; GÜNAY; TÜRKER; TAPAN, 2019; SINGH et al., 2019).

O monitoramento do processo de produção do biodiesel é normalmente realizado por meio de técnicas de referência bem estabelecidas na literatura, como a cromatografia gasosa (GC, do inglês *Gas Chromatography*), e a cromatografia líquida de alta eficiência (HPLC, do

inglês *High Performance Liquid Chromatography*) (FREEDMAN; KWOLEK; PRYDE, 1986; KNOTHE, 2001; ARZAMENDI, 2006). O processo de amostragem, em alguns casos, pode ser complexo e as análises são normalmente demoradas. Por isso, estas técnicas, são aplicadas geralmente após o término da reação ou via *off-line*, por meio da retirada de alíquotas durante o processo reacional. Por isso, algumas limitações estão relacionadas com a obtenção de informações em tempo real, prejudicando a eficiência da tomada de decisão e intervenções no processo de modo a corrigi-lo, quando necessário.

A espectroscopia na região do infravermelho próximo (NIR, do inglês *Near infrared*) é uma técnica confiável, rápida, em geral de baixo custo, não destrutiva e que, comumente, não exige preparo de amostras (PASQUINI, 2003). Quando associada à quimiometria, a espectroscopia NIR torna possível a implementação de ferramentas analíticas, como o Controle Estatístico Multivariado de Processo (MSPC, do inglês *Multivariate Statistical Process Control*), para a realização do monitoramento da reação de transesterificação em tempo real. Além disso, também torna possível o desenvolvimento de modelos de calibração multivariada para acompanhamento em tempo real do teor de éster ao longo do processo de produção (RICHARD, et al., 2013; LIMA, 2014; SALES et al., 2019).

Um importante conjunto de ferramentas analíticas que tem o objetivo de garantir a qualidade de processos, insumos e produtos é denominado de Tecnologias Analíticas de Processo (PAT, do inglês *Process Analytical Technology*). PAT vem sendo aplicado como um sistema capaz de projetar, analisar e controlar o processo produtivo por meio de aferições periódicas de atributos de qualidade e desempenho de materiais durante o processamento. Dentre essas ferramentas, destacam-se: as ferramentas multivariadas para aquisição e análise de dados; os analisadores de processo; as ferramentas de controle de processo; e as ferramentas baseadas no conceito de melhoria contínua e gestão de conhecimento (FDA, 2004).

O desenvolvimento tecnológico associado às vantagens que envolve a utilização de sistemas microeletromecânicos (MEMS, do inglês *Microelectromechanical Systems*), nos últimos anos, contribuiu para a miniaturização de espectrômetros vibracionais, com consequente redução do custo e facilidade de transporte. Dentre os que ganharam escala comercial, destacam-se os espectrômetros raman, de infravermelho (IR, do inglês *Infrared*) e infravermelho próximo. Dentre os espectrofotômetros portáteis na região do NIR, o MicroNIR tem o seu tamanho reduzido principalmente devido ao desenvolvimento do filtro variável linear (LVF, do inglês *Linear Variable Filter*). Um componente formado por um filme fino dielétrico, disposto em uma direção específica, que é ligado diretamente a um conjunto de detectores lineares. Esse arranjo permite que cada pixel do detector responda a um diferente comprimento

de onda, resultando em um conjunto mecânico espectral ultracompacto e robusto (O'BRIEN et al., 2012).

Vários trabalhos descritos na literatura demonstraram resultados satisfatórios quanto à aplicação da espectroscopia NIR ao controle de qualidade de combustíveis (PIMENTEL et al., 2006; PAIVA et al., 2015; CORREIA et al., 2018), assim como para o monitoramento da reação de transesterificação para produção de biodiesel (RICHARD et al., 2013; KILLNER; ROHWEDDER; PASQUINI, 2011; LIMA, 2014; SALES et al., 2019). Entretanto, até o momento, não há relatos na literatura de trabalhos que utilizam um NIR ultracompacto, como o MicroNIR, para o monitoramento *on-line* da reação de transesterificação em batelada.

O MicroNIR, por ser miniaturizado, é um equipamento de menor resolução e com faixa espectral reduzida, porém, é capaz de coletar informação em intervalos de tempo bastante reduzidos. Portanto, o dispositivo demonstra potencial para realização do monitoramento da reação de transesterificação no modo *on-line*, a qual apresenta mudanças rápidas logo no início. Pelo exposto, este trabalho traz como proposta as seguintes abordagens:

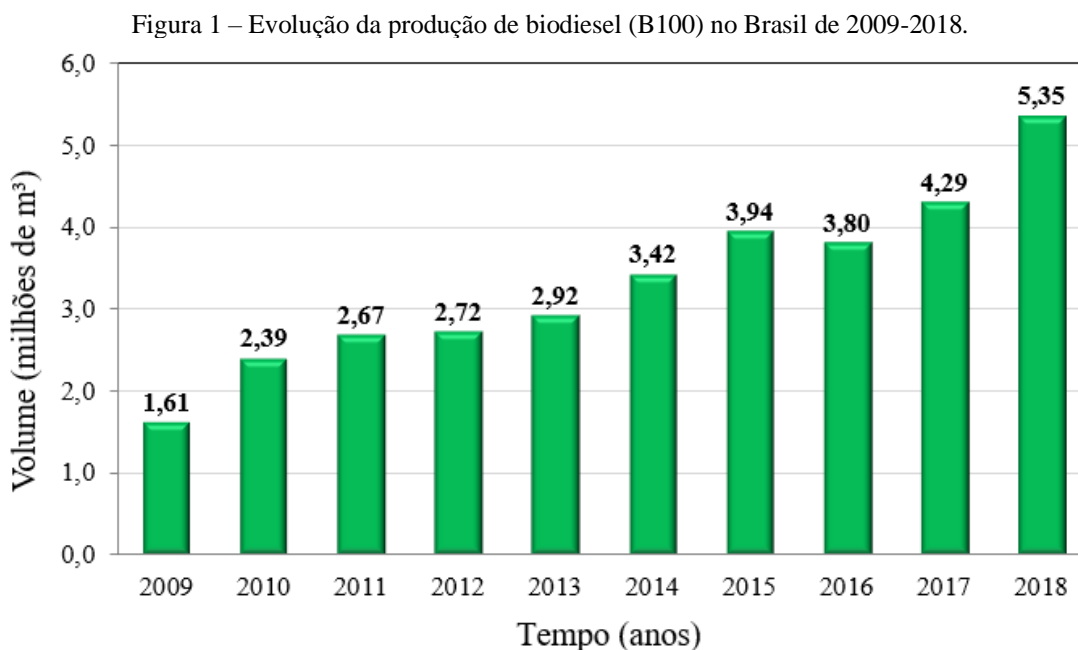
- a) Realizar o monitoramento da reação de transesterificação do óleo de algodão por rota etílica e catálise alcalina em batelada por meio da espectroscopia NIR no modo *on-line*, via utilização de um equipamento de infravermelho portátil (MicroNIR).
- b) Construir cartas de controle estatístico multivariadas de processos a partir de duas abordagens de desdobramento dos dados obtidos durante o monitoramento.
- c) Avaliar a eficiência das diferentes formas de construção das cartas de controle para detecção de falhas em novas bateladas em que perturbações foram intencionalmente efetuadas ao longo da reação.
- d) Construir um modelo de calibração multivariada por mínimos quadrados parciais (PLS, do inglês *Partial Least Square*) confiável, a partir dos dados espectrais do MicroNIR, a fim de realizar previsões do teor de éster em amostras provenientes da reação de transesterificação em batelada.

2 FUNDAMENTAÇÃO TEÓRICA

2.1 BIODIESEL: UM COMBUSTÍVEL RENOVÁVEL

Do ponto de vista legal (Lei nº 11.097, de 13 de janeiro de 2005), o biodiesel é definido como um biocombustível derivado de biomassa renovável que pode ser utilizado em motores a combustão interna que operam com ignição por compressão ou ainda, para geração de outro tipo de energia, que possa substituir total ou parcialmente, os combustíveis fósseis (BRASIL, 2005). Já a ANP, descreve o biodiesel em termos de composição química, como sendo um combustível composto de alquil ésteres de ácidos carboxílicos de cadeia longa, produzido a partir da transesterificação e/ou esterificação de matérias graxas, de gorduras de origem vegetal ou animal, e que atenda a especificação contida no regulamento técnico, presente na Resolução ANP nº 45, de 25 de agosto de 2014 (ANP, 2014).

Desde que o biodiesel começou a fazer parte da matriz energética nacional como aditivo obrigatório ao diesel mineral em 2008, houve um crescimento significativo da sua produção, como mostra a Figura 1 (ANP, 2019).



Fonte: Adaptado de ANP (2019).

Esse crescimento teve forte influência da demanda gerada pela exigência imposta pela Lei nº 11.097, de 13 de janeiro de 2005, que determinou a introdução de 2% em volume de biodiesel a ser incorporado ao diesel mineral, cuja mistura recebeu a denominação de (B2), em

caráter obrigatório a partir de 2008. Desde então a alíquota adicionada aumentou gradativamente até o patamar atual de 12% (B12), cuja regulação vem sendo feita pelo CNPE, amparado pela Lei nº 13.033, de 24 de setembro de 2014 e pela Lei nº 13.263, de 23 de março de 2016.

Essas medidas legais vieram como resposta à necessidade global de se buscar cada vez mais fontes renováveis de energia. Com o crescimento populacional e da indústria, associado à elevação no padrão de qualidade de vida das pessoas, espera-se um aumento no consumo de combustíveis fósseis por parte do setor de transportes da ordem de 60% até 2030, a nível mundial. Consequentemente, a emissão de gases causadores do efeito estufa também tende a aumentar, agravando a problemática ambiental já instalada atualmente. Nesse sentido, a utilização de fontes renováveis de energia tem mostrado ser uma saída atrativa para produção energética. O biodiesel por sua vez, tem demonstrado ser uma opção promissora para substituição total ou parcial do diesel mineral em motores do ciclo diesel, por ser um combustível renovável, biodegradável, devido à presença de moléculas oxigenadas que aceleram este processo, não tóxico e economicamente competitivo. Apresenta ainda viabilidade técnica para uso, visto que além de ser uma fonte energética também tem propriedades lubrificantes, reduzindo o desgaste dos motores. Não menos importante, é ambientalmente amigável, haja vista que é livre de compostos sulfurados e aromáticos, contribuindo para redução da emissão de gases poluentes (DEMIRBAS, 2009; YUSUF; KAMARUDIN; YAAKUB, 2011; BHUIYA et al., 2016).

2.1.1 Produção do biodiesel

Diferentes rotas para produção de biodiesel têm sido reportadas como: pirólise, métodos baseados em microemulsões, transesterificação catalítica e não catalítica, etc. Dentre os métodos citados, a transesterificação é o mais comum, o qual envolve a reação entre triacilgliceróis (principal componente de fontes de matéria-graxa, como óleos vegetais e gordura animal) com um álcool de cadeia curta, normalmente metanol ou etanol, na presença de um catalisador homogêneo ou heterogêneo, de natureza ácida, alcalina ou enzimática. Normalmente, reações de transesterificação envolvem o uso de catalisadores alcalinos como NaOH, KOH e/ou seus respectivos alcoóxidos (YUSUF; KAMARUDIN; YAAKUB, 2011; SINGH et al., 2019; BHUIYA et al., 2016).

O biodiesel pode ser produzido a partir de diversos tipos de matérias-primas, as quais têm sido reportadas pela literatura como: óleos vegetais, gordura animal, óleos de fritura, dentre

outros. Nesse sentido, Singh et al. (2019) classificou o biodiesel em 4 gerações com base no tipo de matéria-prima empregada para sua produção. O biodiesel de 4ª geração baseia-se em alteração biológica das fontes de matéria graxa, mas ainda está em fase inicial de pesquisas. Desse modo, na Tabela 1 estão descritas as principais fontes de ácidos graxos mais utilizadas para produção de biodiesel, as quais estão classificadas em 3 gerações.

Tabela 1 – Classificação do biodiesel em função da origem da matéria-prima.

1ª geração	2ª geração	3ª geração
óleos vegetais comestíveis	óleos vegetais não comestíveis	gordura animal, óleos residuais e outros
Soja	Algodão	Sebo bovino
Côco	Pinhão manso	Gordura avícola
Girassol	Linhaça	Microalga
Oliva	Neem	Banha de porco
Palma	Pongâmia	Óleo de fritura
Amendoim	Stillingia	
Colza		
Gergelim		
Amêndoa		
Mostarda		
Noz		

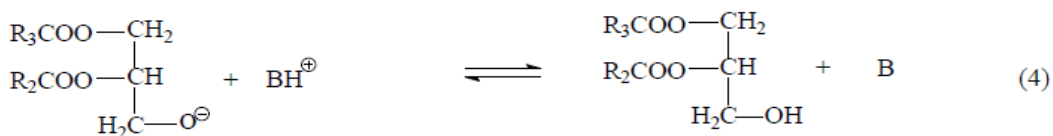
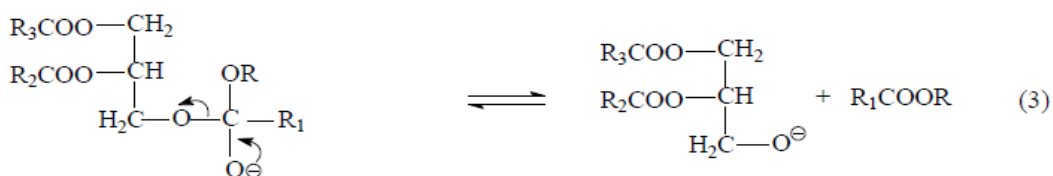
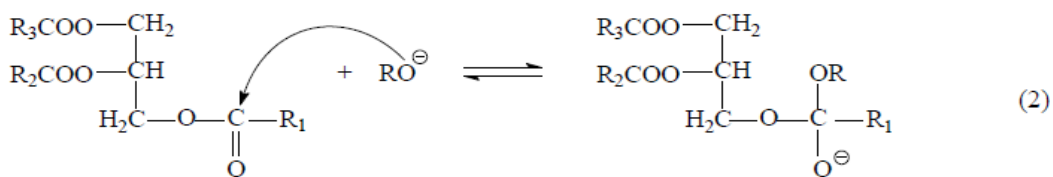
Fonte: Adaptado de Singh et al. (2019).

Em 2018 havia 51 plantas autorizadas pela ANP para produção de biodiesel no Brasil, das quais, 40 delas de fato produziram o biocombustível. A produção diária das plantas ativas naquele ano atingiu o patamar de 23.720 m³, utilizando como matéria-prima majoritária o óleo de soja (69,82%), que corresponde a uma fonte de 1ª geração, seguida de gordura animal (16,22%), óleo de algodão (0,93%) e outras fontes (13,03%), que incluem: óleo de palma, óleo de amendoim, óleo de nabo-forrageiro, óleo de girassol, óleo de mamona, óleo de sésamo, óleo de canola, óleo de milho, óleo de fritura usado e outros materiais graxos (ANP, 2019).

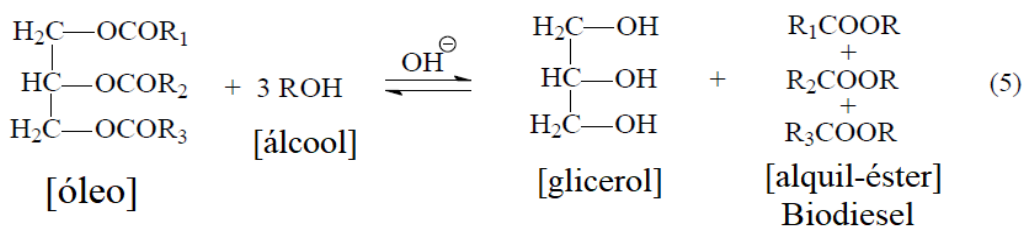
A escolha da matéria-prima para produção de biodiesel leva em consideração as condições geográficas, climáticas e econômicas, de modo particular em cada país. O uso de matérias-primas de 1ª geração para produção de biodiesel gera uma competição quanto ao destino da safra pós colheita com a cadeia alimentar, impactando fortemente nos custos para

aquisição da matéria-prima. Logo, é preferível a escolha de matérias-primas de 2ª ou 3ª geração, visto que em sua maioria ou são constituídas por subprodutos de uma atividade econômica já explorada ou não apresentam aproveitamento econômico direto como *commodity*.

A reação de transesterificação de óleos vegetais geralmente envolve catálise alcalina, em que um equilíbrio endotérmico ocorre na proporção estequiométrica de 1 mol de óleo para 3 mols de um álcool de cadeia curta, geralmente metanol ou etanol, cujo processo passa a ser chamado de metanólise ou etanólise, de acordo com o álcool utilizado, respectivamente. A reação ocorre em 3 etapas reversíveis, em que diacilgliceróis e monoacilgliceróis são obtidos como intermediários, glicerol como subproduto, e o alquil-éster como produto de interesse em cada etapa. De modo que para se obter altas conversões normalmente a razão álcool:óleo é mantida elevada. O mecanismo das etapas reacionais, bem como a reação global para a transesterificação alcalina pode ser evidenciada pelas Equação de 1 – 5. (GÜNAY; TÜRKER; TAPAN, 2019; THANH et al., 2012; DARNOKO; CHERYAN, 2000).



B = catalisador alcalino.



Um dos fatores principais a ser considerado para a escolha do tipo de catalisador em uma reação de transesterificação é a quantidade de água e ácidos graxos livres (FFA, do inglês *Free Fatty Acids*) na matéria-prima. Em relação aos óleos vegetais, aqueles que apresentam alta quantidade de FFA e água, principalmente para os de resíduo de fritura, a catálise ácida mostra-se mais apropriada, em detrimento à utilização de um catalisador alcalino diretamente, que promoveria altas taxas de saponificação, reduzindo o rendimento da conversão. Entretanto, a catálise ácida exige temperaturas relativamente maiores e mais tempo de reação como condições experimentais. Por outro lado, para óleos com baixa quantidade de FFA e água, a utilização direta da catálise alcalina mostra resultados vantajosos em termos de rendimento e tempo de reação (THANH et al., 2012).

Outros fatores determinantes são o tipo de álcool e a razão (álcool:óleo) utilizada na reação de transesterificação, geralmente metanol seguido do etanol são os álcoois mais comuns. O metanol apresenta maior polaridade, custo mais baixo que o etanol em outros países, e geralmente taxas de conversão rápidas, o que pode variar de acordo com o tipo de matéria-prima, a interação do mesmo com o catalisador e a temperatura utilizada. Por outro lado, o etanol tem baixa toxicidade e sua utilização tornaria toda a cadeia produtiva de biodiesel renovável, além de que em países como o Brasil, a utilização de etanol torna-se viável economicamente devido à alta produção de etanol de cana-de-açúcar (GHESTI et al., 2009; MENEGETTI et al., 2006).

Outra vantagem para o uso de etanol é observada quando o óleo de algodão é usado como substrato para produção de biodiesel por rota etílica. O etanol promove maior retenção de gossipol na fase éster, se comparado com metanol. O gossipol é um pigmento de natureza fenólica presente do caroço do algodão que apresenta propriedades antioxidantes, com potencial para aumentar a estabilidade à oxidação do biodiesel, atuando como retardante sobre sua degradação (JOSHI; TOLER; WALKER, 2008). Isto se torna interessante, particularmente no Brasil, devido à recente exigência de aditivação do biodiesel com antioxidantes antes da liberação para comercialização, conforme a Resolução nº 798, de 1 de agosto de 2019.

2.2 TECNOLOGIAS ANALÍTICAS DE PROCESSO E A ESPECTROSCOPIA NA REGIÃO DO INFRAVERMELHO PRÓXIMO

2.2.1 Tecnologias analíticas de processo

O desenvolvimento tecnológico acelerado vem transformando a forma de aquisição de informação que, atualmente, vem sendo obtida de modo mais rápido, amplo e contínuo. Assim, o desenvolvimento de ferramentas analíticas voltadas para auxiliar o processamento de dados vem se tornando uma demanda em comum entre a academia e a indústria. Neste sentido, as Tecnologias Analíticas de Processo (PAT) consistem em um conjunto de ferramentas com foco na aplicação de sistemas para análise, projeção e controle de processos, por meio da aquisição de medidas de parâmetros de qualidade e desempenho de materiais durante o processo produtivo. Dentre essas ferramentas, estão as ferramentas multivariadas para aquisição e análise de dados; os analisadores de processo; as ferramentas de controle de processo de melhoria contínua, cuja aplicação visa sempre a busca pela qualidade do produto final, entre outras (FDA, 2004; SANDEN et al., 2019).

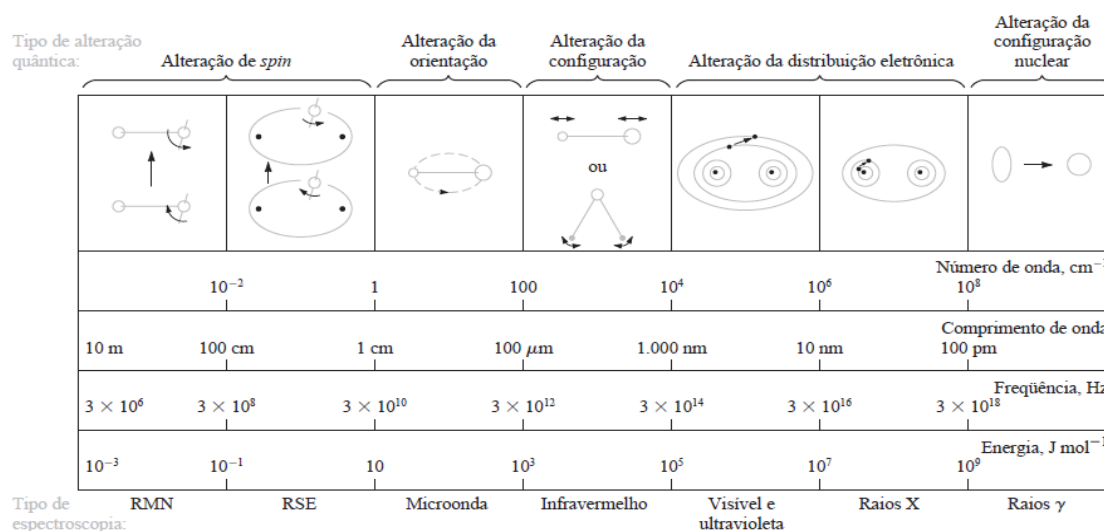
Trabalhos reportados pela literatura ressaltam a aplicação de ferramentas de análises multivariadas para aquisição e análise de dados como uma opção para avaliação de processos produtivos, e auxílio na tomada de decisão em tempo real, bem como para o desenvolvimento de métodos de controle de qualidade (ROUCHI et al., 2019; PAIVA et al., 2015; SALES et al., 2016; LIMA, 2014; CORREIA et al., 2018). Nesse contexto, métodos baseados em espectrometria vibracional, como Raman e NIR, tendem a ganhar destaque pela sua capacidade de aquisição de dados de forma rápida, não destrutiva e confiável, bem como pelo processo de miniaturização que eles vêm sofrendo, tornando-os cada vez mais compactos e versáteis (PASQUINI, 2003; O'BRIEN et al., 2012).

2.2.2 Espectroscopia na região do infravermelho próximo (NIR)

A espectroscopia estuda a interação da radiação eletromagnética com a matéria, de modo que os métodos espectroscópicos se baseiam na determinação da quantidade de radiação emitida ou absorvida por moléculas ou espécies atômicas na forma de fótons. As técnicas espectroscópicas apresentam alta versatilidade, podendo ser aplicadas tanto em análises qualitativas quanto para quantificação de espécies atômicas e moleculares. Conforme descrito pela Figura 2, essas técnicas podem ser classificadas de acordo com a região do espectro eletromagnético em que é feita a medida, cuja intensidade e tipo da alteração provocada a partir da interação da radiação com as espécies químicas, dependerá da energia característica de cada região. Observa-se que a intensidade da alteração sobre as espécies químicas é proporcional à quantidade de energia associada à frequência da radiação eletromagnética, tornando possível

acessar informações específicas por meio das técnicas espectroscópicas que utilizam faixas distintas do espectro.

Figura 2 – Regiões do espectro eletromagnético.



Fonte: Skoog et al. (2008).

A região do infravermelho no espectro eletromagnético apresenta uma subdivisão em 3 faixas, denominadas de infravermelho: próximo, médio (MIR, do inglês *Mid Infrared*) e distante (FIR, do inglês *Far Infrared*), cujos respectivos intervalos no espectro eletromagnético, se encontram descritos na Tabela 2.

Tabela 2 – Faixas espectrais da radiação infravermelha.

Faixa espectral	Comprimento de onda – λ (nm)	Número de onda (cm ⁻¹)
NIR	780 – 2.500	12.800 – 4.000
MID	2.500 – 50.000	4000 – 200
FAR	50.000 – 1.000.000	200 – 10

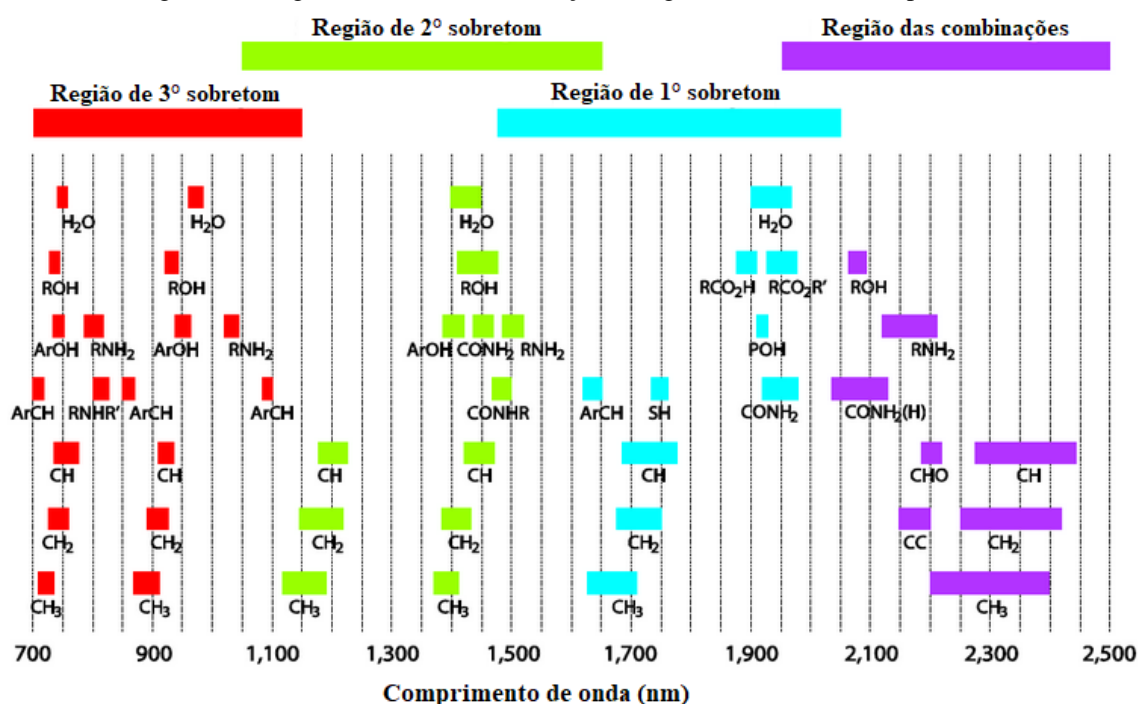
Fonte: Adaptado de Skoog, Holler e Crouch (2017).

A energia proveniente da radiação infravermelha não é energética o suficiente para causar transições eletrônicas, porém, pode induzir transições nos estados vibracionais e rotacionais das moléculas. Portanto, a espectroscopia na região do infravermelho é uma técnica de espectroscopia molecular. Para uma determinada ligação química ser capaz de absorver radiação na região do infravermelho, ela precisa sofrer variação no seu momento dipolo, em função dos seus movimentos naturais de vibração e/ou rotação, ou seja, é preciso respeitar a regra de seleção.

Tal condição considera que, para ligações químicas que apresentam momento dipolar diferente de zero, o movimento de aproximação e distanciamento dos átomos devido à vibração, gera um campo elétrico oscilante ao redor da ligação química, tornando-o capaz de interagir com a radiação eletromagnética incidente. Se a frequência da vibração natural da ligação coincidir com a frequência da radiação incidente, ocorre então a absorção de energia pela ligação química, o que resulta em uma variação na amplitude vibracional desta ligação. (SKOOG; HOLLER; CROUCH, 2017).

Na região do infravermelho próximo, a informação espectral é basicamente composta por bandas de absorção de sobretom e combinações das bandas de absorção fundamentais localizadas na faixa do infravermelho médio, conforme evidenciado na Figura 3 (WORKMAN; WEYER, 2012).

Figura 3 – Diagrama de bandas de absorção na região do infravermelho próximo.



Fonte: Adaptado de Zitting (2017).

Os sobretom correspondem a fenômenos de absorção de energia em que a molécula ao receber energia radiante é capaz de migrar do estado fundamental diretamente para níveis energéticos mais elevados, como o segundo, terceiro ou quarto estado excitado, correspondentes ao 1º, 2º e 3º sobretom respectivamente, considerando uma transição de número quântico vibracional maior que um ($\Delta v = \pm 2; \pm 3$ ou ± 4). Já as bandas de combinação ocorrem quando dois ou mais modos vibracionais são excitados ao mesmo tempo por um

mesmo fóton. A frequência das bandas de combinação é aproximadamente igual à soma ou à diferença de ambas as frequências fundamentais para os dois modos vibracionais (SKOOG; HOLLER; CROUCH, 2017; SILVA, 2017).

Segundo Pasquini (2003), a intensidade de absorção da radiação NIR é dependente do grau de anarmonicidade dos modos vibracionais e da diferença no momento dipolo presente nas ligações de espécies químicas. Estes aspectos são muito característicos de grupos funcionais que envolvem principalmente a presença de átomos como (C, O, N e S) ligados ao átomo de hidrogênio, como observado nas bandas de absorção apresentadas na Figura 3, para as ligações químicas ativas na região do infravermelho próximo. Esta condição viabiliza aplicações da espectroscopia NIR tanto em termos qualitativos como quantitativos para uma ampla faixa de compostos orgânicos, dentre eles os combustíveis.

Embora a intensidade de absorção de radiação NIR seja de 10 – 1.000 vezes menor do que na região do infravermelho médio, com o avanço tecnológico, a geração de fontes cada vez mais potentes associadas a detectores mais sensíveis promovem uma atenuação dessa deficiência, tornando as análises nessa faixa espectral confiáveis. Neste sentido, a espectroscopia NIR pode ser apontada como uma técnica com aplicação universal, visto que é não destrutiva, não invasiva, rápida e em geral requer pouco ou nenhum tratamento prévio de amostras.

Outra característica interessante desta técnica é a utilização de acessórios específicos junto aos equipamentos, tanto de bancada quanto portáteis, que permitem a aquisição de informação em tempo real diretamente do meio estudado. Nesta condição, destacam-se os modos de aquisição espectral *in-line* e *on-line*. No primeiro, um sensor analítico permanece diretamente em contato com o meio reacional, proporcionando maior sensibilidade a perturbações no sistema, gerando respostas mais representativas, além de tornar desnecessária uma etapa de amostragem. O segundo modo de aquisição de informação em tempo real, que cada vez mais é empregado, é o *on-line*, em que parte do volume do meio reacional é coletado, de forma contínua ou intermitente, por meio de uma linha de amostragem integrada ao meio reacional, de onde o fluxo é então condicionado em célula específica para coleta dos dados e processamento (TREVISAN; POPPI, 2006).

A grande quantidade de informação obtida por meio da rápida aquisição de espectros NIR, associada com a característica sobreposição de bandas de absorção no espectro, gera complexidade no processamento e interpretação das informações espectrais para fins analíticos qualitativos e quantitativos. Dessa forma, a utilização de quimiometria para extrair a informação

e correlacioná-la às características físicas e químicas da amostra é geralmente requerida (PASQUINI, 2018; WORKMAN; WEYER, 2012).

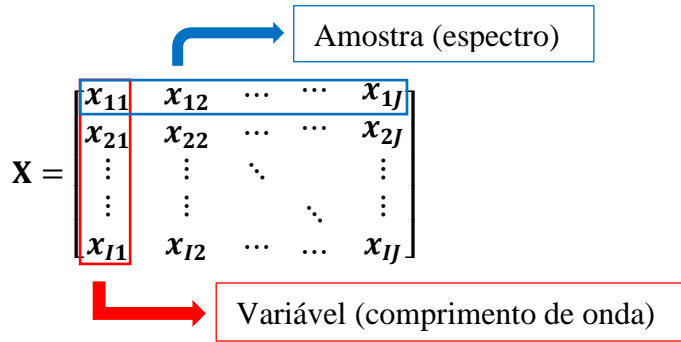
2.3 QUIMIOMETRIA

De modo genérico, a quimiometria pode ser definida como a aplicação de ferramentas matemáticas e estatísticas à química (KOWALSKI, 1975). A importância do uso dessas ferramentas está relacionada com o crescente desenvolvimento tecnológico e a inovação na instrumentação química, que trouxe desafios pelo fato da geração de dados em larga escala por sistemas, muitas vezes integrados, trazer complexidade à interpretação da resposta instrumental. Tais dificuldades são observadas tanto do ponto de vista matemático e estatístico, quanto para associação desses dados com as propriedades de interesse das amostras analisadas.

A aplicação de técnicas quimiométricas apresentam grandes vantagens, dentre as principais estão a otimização das condições para obtenção de medidas experimentais, a extração da maior quantidade possível de informação por análise química, bem como a correção de desvios associados ao processo de obtenção das medidas. Neste sentido, para correção dos efeitos de desvios de linha de base em espectros NIR, por exemplo, é necessária a aplicação de técnicas quimiométricas descritas como pré-processamentos. Esta etapa do tratamento dos dados é responsável por reduzir variações presentes nos espectros que não são diretamente correlacionadas com as propriedades de interesse mensuradas, a fim de tornar a análise química mais confiável (BRUNS; FAIGLE, 1984; BAPTISTA et al., 2008).

2.3.1 Pré-processamento de dados espectroscópicos.

Nesta seção serão introduzidas as técnicas mais comumente utilizadas para a correção de desvios associados aos espectros NIR, a fim de esclarecer sua aplicação ao longo da discussão do trabalho. Inicialmente, para fins de exemplificação, uma matriz de espectros (\mathbf{X}) foi considerada. Cada linha da matriz corresponde a uma amostra, neste caso, a um espectro NIR, e cada coluna corresponde à uma variável, ou seja, um comprimento de onda. Considerando que nas linhas são observadas I amostras e nas colunas, J variáveis, cada elemento dessa matriz é denotado por (x_{ij}) , conforme exemplificado na Equação 6 (FERREIRA, 2015).



$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1J} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2J} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ x_{I1} & x_{I2} & \cdots & \cdots & x_{IJ} \end{bmatrix} \quad (6)$$

2.3.1.1 Pré-processamento nas variáveis

A centragem na média é sem dúvida uma das principais técnicas de pré-processamento aplicado às variáveis. É calculada, para todas as J colunas da matriz \mathbf{X} , a partir da subtração de cada um dos valores que compõem a coluna pelo valor médio de cada variável, conforme descrito pela Equação 7.

$$x_{ij(cm)} = x_{ij} - \bar{x}_j \quad (7)$$

Em que: $(x_{ij(cm)})$ é o valor de cada elemento da variável centrada na média e (\bar{x}_j) é a média para cada coluna, calculada de acordo com Equação 8.

$$\bar{x}_j = \frac{1}{I} \sum_{i=1}^I x_{ij} \quad (8)$$

Este pré-processamento promove apenas uma translação de eixos para o valor médio de cada um deles, sem, no entanto, alterar sua estrutura (FERREIRA, 2015).

2.3.1.2 Pré-processamento nas amostras

Várias técnicas de pré-processamento são aplicadas aos espectros NIR para fins de redução da variabilidade em termos de espalhamento da radiação, efeito físico geralmente indesejado para avaliação da informação espectral. Dentre eles, a variação normal padrão (SNV, do inglês *Standart Normal Variate*), a correção de espalhamento multiplicativo (MSC, do inglês *Multiplicative Scatter Correction*), e a derivação espectral têm apresentado resultados positivos (FEARN et al., 2009).

A variação normal padrão é um dos métodos mais comuns para correção de desvios espectrais. SNV atua de modo individual, removendo a variação do deslocamento da linha de base, bem como efeitos de inclinação espectral, espectro por espectro. Tais efeitos são conhecidos como aditivo e multiplicativo, respectivamente. Dessa forma, os desvios gerados pelo espalhamento são eliminados sem haver alteração do perfil espectral, o que favorece o procedimento de identificação de bandas em etapas posteriores ao pré-processamento. Matematicamente, essa correção é efetuada através da subtração de todos os valores de cada espectro (x_i), pelo valor médio (\bar{x}_i) (uma centragem na média para as amostras), e posterior divisão pelo desvio padrão do respectivo espectro (s_i), conforme demonstrado na Equação 9 (BARNES; DANHOA; LISTER, 1989; FERREIRA, 2015).

$$x_{i\text{SNV}} = \frac{(x_i - \bar{x}_i)}{s_i} \quad (9)$$

Em que (\bar{x}_i) é denotado pela Equação 10:

$$\bar{x}_i = \frac{1}{J} \sum_{j=1}^J x_{ij} \quad (10)$$

A correção de espalhamento multiplicativo também é utilizada para correção de efeitos de espalhamento da radiação. Embora apresente resultados de correção espectral semelhantes ao SNV, essa técnica corrige cada espectro em função da média de todos os espectros da matriz de dados (ou outro espectro usado como referência), e não através do valor médio de cada espectro, como faz o SNV. Nesta técnica, é feita uma regressão de cada espectro em relação ao espectro de referência, geralmente o espectro médio do conjunto de calibração (x_m). Assim, cada espectro (x_i) pode ser descrito como uma função linear do espectro médio (x_m), com coeficiente angular (b_i) e coeficiente linear (a_i), ambos constantes para todos os comprimentos de onda, conforme a Equação 11.

$$x_i = a_i + b_i x_m + e_i \quad (11)$$

O espectro corrigido denotado por ($x_{i\text{MSC}}$) é obtido por meio da diferença entre a absorvância do espectro original (x_i) e o coeficiente linear (a_i), gerando a correção do efeito

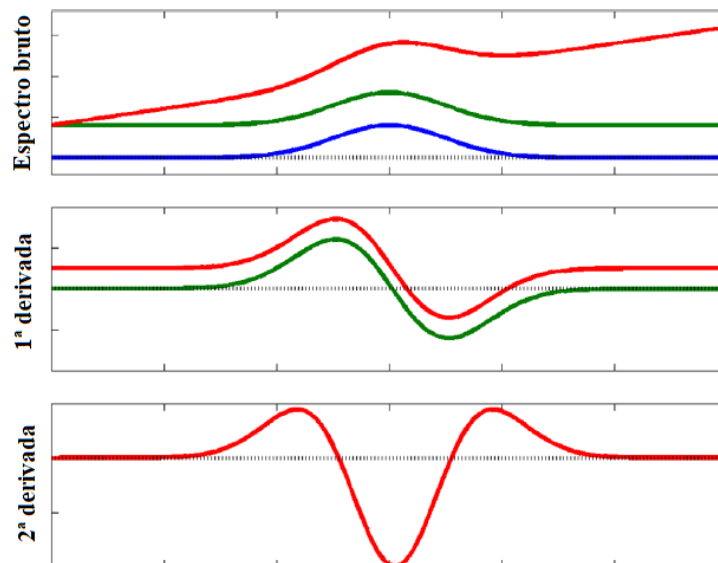
aditivo, e pela posterior divisão deste resultado pelo coeficiente angular (b_i), corrigindo agora o efeito multiplicativo. O resultado desta correção aplicado para a i -ésima amostra da matriz descrita na Equação 6, pode ser observada na Equação 12 (FERREIRA, 2015; SILVA, 2017).

$$x_{i \text{ MSC}} = \frac{x_i - a_i}{b_i} = x_m + \frac{e_i}{b_i} \quad (12)$$

Em que (e_i) é o resíduo do processo de modelagem do espectro original (x_i).

Outra estratégia de pré-processamento amplamente utilizada para correção de efeitos físicos em espectros NIR é a derivação espectral. Os métodos derivativos têm a habilidade de corrigir tanto o efeito aditivo, a partir do uso da primeira derivada, quanto o multiplicativo, que corrige além de deslocamento de linha de base, a inclinação presente nos espectros, conforme exemplificado na Figura 4.

Figura 4 – Resultados da derivação espectral. A linha verde corresponde ao erro aditivo e a vermelha ao multiplicativo.



Fonte: Adaptado de RINNAN; BERG; ENGELSEN (2009).

Conforme o grau de complexidade da derivada aplicada aumenta, há uma redução da intensidade do sinal e o surgimento de mais lobos como resultado intrínseco ao processo derivativo. Juntos, esses fatores contribuem para diminuição da razão sinal/ruído. Para correção desse problema, geralmente esses métodos são acompanhados de filtros de suavização, dentre eles um dos mais conhecidos e utilizados é o de Savitzky-Golay. Neste método, um polinômio

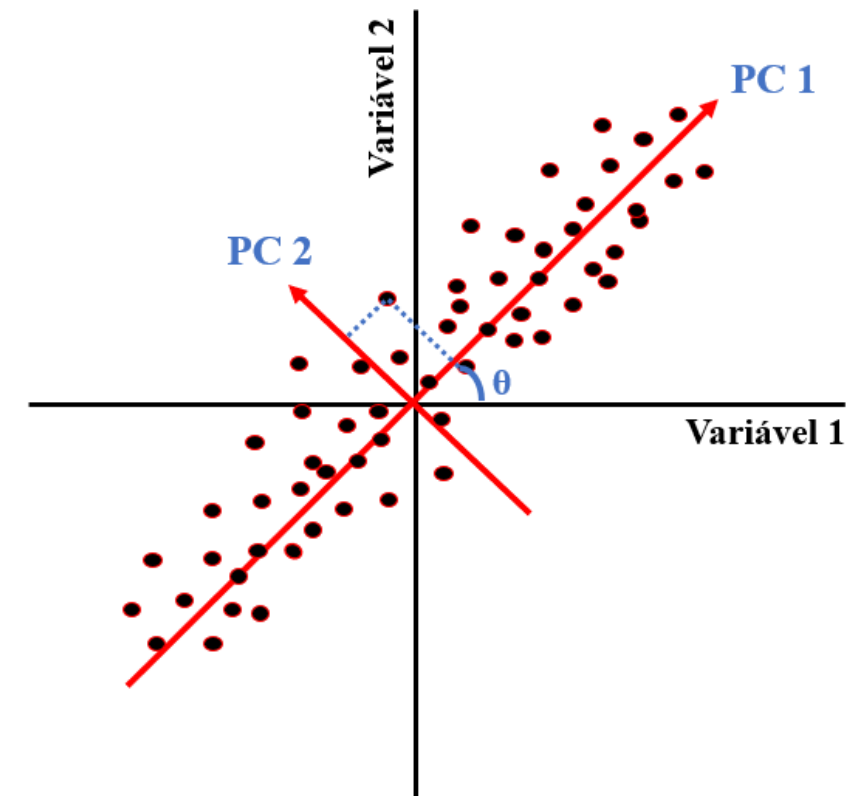
de grau n é ajustado para suavizar intervalos móveis ao longo de todo o espectro, chamados de janelas, cujo tamanho depende do número de pontos escolhidos. O valor central da janela é então substituído por uma média ponderada de seus valores, cujos pesos são obtidos pelo ajuste polinomial. Tanto o grau n do polinômio quanto o tamanho da janela devem ser ajustados, decisão que requer conhecimento sobre a natureza dos dados e das características técnicas do equipamento que gerou a resposta instrumental. A alteração do perfil espectral original devido ao pré-processamento a partir da derivada, no entanto, tem como principal desvantagem a dificuldade introduzida ao processo de atribuição de bandas, visto que após a correção, elas se tornam deslocadas em relação à posição original. (BARNES; DANHOA; LISTER, 1989; RINNAN; BERG; ENGELSEN, 2009; FERREIRA, 2015).

2.3.2 Análise de Componentes Principais (PCA)

A análise de componentes principais (PCA, do inglês *Principal Component Analysis*) é um dos métodos mais importantes da quimiometria. Frequentemente é utilizada como uma poderosa ferramenta para visualização da estrutura dos dados, bem como para reduzir a dimensionalidade do conjunto amostral. A PCA é classificada como um método de análise exploratória não supervisionado, que permite a interpretação de dados sem a necessidade de possuir conhecimento prévio sobre sua natureza. Este método torna viável o reconhecimento de padrões entre as amostras, de acordo com a apresentação de possíveis tendências, diferenças e/ou formação de agrupamentos naturais (SOUZA; POPPI, 2012; FERREIRA, 2015).

Inicialmente, as amostras que constituem o conjunto de dados original apresentam-se como pontos dispersos em um espaço de n dimensões. A PCA representa estes dados em um novo sistema de eixos ortogonais entre si, chamados de componentes principais (PC, do inglês *Principal Component*), que permitem a observação da informação de modo não correlacionado em cada PC. Desse modo, as amostras passam a estar distribuídas de acordo com a direção da sua máxima variância sobre essa nova forma de projeção. Cada PC resulta da combinação linear das variáveis originais. De modo que a primeira PC (PC1) é direcionada no sentido da maior variabilidade, a segunda componente (PC2), que é ortogonal à primeira, aponta no sentido da segunda maior variabilidade, explicando a variação que PC1 não foi capaz, e assim por diante. A representação das amostras sobre os eixos das PCs pode ser observada na Figura 5 (SOUZA; POPPI, 2012).

Figura 5 – Projeção das amostras sobre os eixos de máxima variância (PCs).



Fonte: O autor (2020).

Do ponto de vista matemático, a PCA decompõe a matriz \mathbf{X} (Equação 6), em duas outras matrizes, a de escores (\mathbf{T}) e a de *loadings* ou pesos (\mathbf{P}), mais uma matriz de erros (\mathbf{E}), conforme descrito na Equação 13.

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad (13)$$

Os escores representam as coordenadas das amostras no sistema de eixos formado a partir das componentes principais, em que essa relação no novo subespaço delimitado pelas PCs pode evidenciar informações acerca de semelhanças entre as amostras. Os pesos são obtidos a partir do cosseno do ângulo formado entre a componente principal e o eixo da variável original. Sua avaliação contribui para identificação do grau de importância de cada variável sobre a distribuição das amostras nas componentes principais.

Por outro lado, a matriz de erros (\mathbf{E}) corresponde ao desvio entre a projeção e as coordenadas originais, ou seja, nela estão contidas as informações que não foram descritas pelas PCs. Como cada componente principal apresenta um percentual de variância explicada, o número de PCs utilizado na modelagem é adotado de acordo com o máximo de informação

possível a ser capturada do conjunto de dados (SOUZA; POPPI, 2012; WOLD; ESBENSEN; GELADI, 1987).

2.3.3 Calibração multivariada

2.3.3.1 Regressão pelos mínimos quadrados parciais PLS

Para construção do modelo de regressão multivariada, a abordagem PLS utiliza as informações tanto da matriz de dados espectrais \mathbf{X} (Equação 6), quanto de um vetor \mathbf{y} ou matriz \mathbf{Y} , que contém as informações da propriedade de interesse que se deseja estabelecer uma correlação linear. Os modelos são denominados de PLS 1 quando existe apenas uma propriedade de interesse, a qual é descrita como um vetor \mathbf{y} , e PLS 2 quando há mais de uma propriedade em que se deseja correlacioná-las simultaneamente com a matriz de dados \mathbf{X} . Nesse caso, as propriedades de interesse são dispostas em uma matriz \mathbf{Y} . A relação entre os dois conjuntos de dados é promovida por meio de fatores, denominados de variáveis latentes (VL) (NAES et al., 2002; SOUZA et al., 2013).

Na etapa de modelagem pelo método dos mínimos quadrados parciais, a decomposição da matriz \mathbf{X} e da matriz \mathbf{Y} nas variáveis latentes segue um processo análogo ao descrito para a PCA, item 2.3.2. No entanto, nesse caso, as variáveis latentes promovem a maximização da covariância entre os escores da matriz \mathbf{X} e da matriz \mathbf{Y} , e o número de variáveis latentes utilizadas para o processo de modelagem é escolhido, em geral, de acordo com a minimização do erro de validação cruzada. Esta decomposição é descrita pelas Equações 14 e 15 (SOUZA et al., 2013; GELADI; KOWALKI, 1986).

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} = \sum t_h \mathbf{p}_h^t + \mathbf{E} \quad (14)$$

$$\mathbf{Y} = \mathbf{U}\mathbf{Q}^t + \mathbf{F} = \sum u_h \mathbf{q}_h^t + \mathbf{F} \quad (15)$$

Em que \mathbf{P} e \mathbf{Q} correspondem aos pesos e \mathbf{T} e \mathbf{U} são os escores da matriz \mathbf{X} e da matriz \mathbf{Y} , respectivamente. Assim como \mathbf{E} e \mathbf{F} são os resíduos da matriz \mathbf{X} e da matriz \mathbf{Y} , respectivamente. Uma vez que os escores de \mathbf{X} e \mathbf{Y} são correlacionados (\mathbf{T} e \mathbf{U} , respectivamente), é possível estabelecer uma relação linear entre as matrizes \mathbf{X} e \mathbf{Y} , representada pelos coeficientes de regressão.

Os escores da matriz \mathbf{X} e da matriz \mathbf{Y} são relacionados por meio do coeficiente \mathbf{b} para as \mathbf{h} variáveis latentes, conforme descrito na Equação 16.

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h \quad (16)$$

O agrupamento dos valores de \mathbf{b}_h em uma matriz diagonal \mathbf{B} , dá origem à matriz de coeficientes de regressão entre a matriz de escores \mathbf{U} da matriz \mathbf{Y} e os escores \mathbf{T} da matriz \mathbf{X} . A melhor relação linear entre a matriz \mathbf{X} e a matriz \mathbf{Y} é obtida através dos coeficientes de regressão, que também estabelecem uma relação com os *loadings* de \mathbf{X} e \mathbf{Y} , por meio de um fator de peso \mathbf{W} conforme descrito pela Equação 17 (NAES et al., 2002).

$$\mathbf{B} = \mathbf{W}(\mathbf{P}^t \mathbf{W})^{-1} \mathbf{Q} \quad (17)$$

Cada nova predição é então obtida a partir da Equação 18.

$$\hat{\mathbf{Y}} = \mathbf{X} \mathbf{B} \quad (18)$$

Após a construção do modelo é necessário testar sua capacidade preditiva. Esta etapa é chamada de validação, e consiste em verificar se a resposta do modelo corresponde a valores coerentes com a respectiva propriedade do analito, a partir de parâmetros e testes estatísticos que comprovem seu desempenho (FERREIRA et al., 1999).

2.3.3.2 Validação de modelos de regressão multivariados e figuras de mérito

Para avaliação da capacidade preditiva dos modelos de regressão multivariados, comumente são utilizados parâmetros estatísticos, conhecidos como figuras de mérito. Dentre elas, a raiz quadrada do erro médio quadrático (RMSE, do inglês *Root Mean Squared Error*) e o *bias* são comumente utilizadas. Esses parâmetros são calculados a partir dos valores preditos pelo modelo e os valores de referência. Já a comparação quanto ao desempenho entre os modelos construídos é realizada através dos testes de Fisher e *t-student* (NAES et al., 2002).

O parâmetro RMSE é demonstrado pela Equação 19, e corresponde à raiz quadrada da razão entre o somatório do quadrado da diferença entre o valor predito de cada amostra (\hat{y}_i) e o respectivo valor de referência (y_i), e o número (N) de amostras para um modelo. É importante

destacar que a unidade do erro calculado pela Equação 19 é a mesma das medidas originais realizadas, facilitando a avaliação do resultado.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (19)$$

Quando essa estimativa é realizada sobre o conjunto de dados de calibração utilizados para construção do modelo, um dos parâmetros utilizados é a raiz quadrada do erro médio quadrático de calibração (RMSEC, do inglês *Root Mean Squared Error of Calibration*), calculado a partir da Equação 20. Nesta etapa, as mesmas amostras utilizadas para construção do modelo são empregadas para testá-lo.

$$RMSEC = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{(N - h - 1)}} \quad (20)$$

Em que (\hat{y}_i) é o valor predito para cada amostra, tendo como respectivo valor de referência (y_i) , (N) é o número de amostras de calibração, e (h) é o número de variáveis latentes utilizadas pelo modelo.

Como o RMSEC estima apenas o erro correspondente às amostras de calibração, ou seja, sem qualquer informação do conjunto de predição, a estimativa dos erros dos coeficientes de regressão são desconsiderados, não sendo, portanto, um parâmetro confiável para qualquer estimativa sobre a predição. Nesse sentido, a fim de proporcionar maior confiança à verificação dos dados de calibração quanto à capacidade preditiva, faz-se necessária a utilização de alternativas mais eficazes como processos de validação e predição.

Um dos métodos mais utilizados para validação interna é a validação cruzada (CV, do inglês *Cross-Validation*). Nesse método, são realizadas sucessivas retiradas de amostras do conjunto de calibração, e a cada retirada, um modelo é ajustado para as amostras restantes, o qual é então testado para predição das amostras previamente removidas. Na sequência, aquelas primeiras amostras retiradas são devolvidas ao conjunto de calibração e o processo é repetido até que todas as amostras tenham sido removidas e tenha sido feita a sua predição. O número de amostras removidas de cada vez pode variar desde apenas uma, até a remoção de um grupo de amostras. Se o primeiro modo for escolhido, a técnica de validação cruzada é chamada de *leave-one-out*. Já no segundo caso, as amostras que compõem os grupos podem ser escolhidas

por blocos de amostras aleatórios, sendo descrita por *Contiguos Blocks*, ou em quantidades definidas, cujos blocos são formados a partir da retirada de posições especificadas previamente, sendo então chamado de *Venetian blinds*, ou ainda de modo personalizado *Custom*, cuja escolha é feita manualmente, além de outras variantes desses métodos (NAES et al., 2002).

Assim, a estimativa do erro para o método de validação cruzada é dada pela raiz quadrada do erro médio quadrático de validação cruzada (RMSECV, do inglês *Root Mean Squared Error of Cross-Validation*). O RMSECV é obtido pela Equação 21 (NAES et al., 2002).

$$RMSECV = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_{cv,i} - y_i)^2}{(N)}} \quad (21)$$

Já para a etapa de predição, o erro é calculado como a raiz quadrada do erro médio quadrático de predição (RMSEP, do inglês *Root Mean Squared of Prediction*), conforme descrito na Equação 22.

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{N_p} (\hat{y}_i - y_i)^2}{(N_p)}} \quad (22)$$

Em que (N_p) corresponde ao número de amostras do conjunto de predição, (\hat{y}_i) e (y_i) são: o valor predito pelo modelo e o valor de referência para cada amostra do conjunto de predição, respectivamente.

Nesta etapa por meio do RMSEP estima-se a capacidade preditiva do modelo em termos de exatidão entre as respostas previstas pelo modelo e as de referência. Essa resposta é avaliada para um conjunto de dados diferente daquele utilizado para construção do modelo de calibração, sendo uma referência de quão bem o modelo funciona. A desvantagem apresentada por este tipo de validação, é apenas a necessidade de reservar uma parte considerável do conjunto de dados original para esta finalidade, visto que a variabilidade dessas amostras, deixa de fazer parte dos dados de calibração (NAES et al., 2002).

A fim de avaliar o desempenho entre os modelos de predição construídos, o teste estatístico de Fisher, teste F, é uma das ferramentas normalmente utilizadas para comparação entre os valores de RMSEP (SILVA, 2017; LIMA, 2014), haja vista que ele é capaz de testar

hipóteses relativas às variâncias populacionais. A expressão matemática que descreve o teste é apresentada na Equação 23 (BRUNS; SCARMINIO; NETO, 2006).

$$\frac{S^2_A}{S^2_B} \cong \frac{\sigma^2_A}{\sigma^2_B} F_{\nu_A, \nu_B} \quad (23)$$

Em que, S^2_A e S^2_B correspondem às variâncias amostrais, σ^2_A e σ^2_B são as variâncias populacionais e ν_A e ν_B são os graus de liberdade para cada conjunto. Nesse sentido, através de um teste de hipóteses, é possível avaliar se as variâncias populacionais são de fato idênticas ou não para cada modelo construído. Para isso, faz-se necessário testar a hipótese de que as variâncias amostrais S^2_A e S^2_B são capazes de estimar a variância populacional. Assim, o teste de Fischer é aplicado a fim de obter o valor de F calculado (F_{calc}) a partir da Equação 24.

$$F_{calc} = \frac{S^2_A}{S^2_B} \quad (24)$$

O resultado da equação 24 é então comparado com o valor de referência tabelado (F_{tab}), levando em consideração os respectivos graus de liberdade ν_A e ν_B . Uma vez que (F_{calc}) é maior que (F_{tab}), não é possível afirmar que as variâncias são estatisticamente equivalentes, considerando um certo nível de confiança. Logo, pode-se afirmar que há diferença significativa entre os valores de RMSEP comparados.

Outro parâmetro utilizado para avaliação de modelos multivariados é a identificação da presença de erros sistemáticos, por meio do *bias*, o qual é calculado a partir da Equação 25 (NAES et al., 2002).

$$bias = \frac{\sum_{i=1}^{N_p} (\hat{y}_i - y_i)}{(N_p)} \quad (25)$$

O *bias* corresponde basicamente à diferença média entre o valor predito pelo modelo (\hat{y}_i) e o valor de referência (y_i) para o conjunto de predição com (N_p) amostras.

A presença de erros sistemáticos em modelos PLS é apontada, em geral, quando o valor do *bias* é significativo para o nível de 95% de confiança. Essa estimativa é obtida utilizando o teste *t-student*, conforme normalizado pela *American Society for Testing and Materials* (ASTM E1655-05), cuja expressão matemática está mostrada na Equação 26.

$$t_{calc} = \frac{|bias|\sqrt{N_p}}{SDV} \quad (26)$$

Em que (N_p) é o número de amostras do conjunto de predição, e SDV corresponde ao desvio padrão do erro de validação (SDV, do inglês *Standard Deviation of Validation*). O SDV pode ser calculado pela Equação 27.

$$SDV = \sqrt{RMSEP^2 - bias^2} \quad (27)$$

O valor de (t_{calc}) é então comparado com o valor tabelado (t_{tab}), e caso o valor de t calculado for maior que o valor de t crítico, o modelo apresenta *bias* significativo considerando v graus de liberdade para o nível de 95% de confiança.

Um outro parâmetro estatístico utilizado para avaliação de modelos multivariados é o coeficiente de determinação (R^2). Essa estatística representa o grau de ajuste do modelo aos dados, e está descrito pela Equação 28 (FERREIRA, 2015).

$$R^2 = 1 - \frac{\sum_{i=1}^l (y_i - \hat{y}_i)^2}{\sum_{i=1}^l (y_i - \bar{y}_i)^2} \quad (28)$$

Em que (y_i) é o valor de referência, (\hat{y}_i) é o valor estimado pelo modelo multivariado e (\bar{y}_i) é a resposta de referência média. O coeficiente de determinação (R^2) pode adotar valores no intervalo de $0 \leq R^2 \leq 1$, sendo estimado para cada etapa de modelagem, ou seja, serão obtidos valores de R^2 para os modelos de calibração e validação interna e predição.

2.3.3.3 Seleção de variáveis pelo algoritmo *interval* PLS

O algoritmo dos mínimos quadrados parciais por intervalos (*i*PLS, do inglês *Interval Partial Least Squares*) promove a construção iterativa de modelos locais PLS a partir de faixas de comprimentos de onda selecionadas do espectro inteiro. Este método de seleção de variáveis apresenta como principal objetivo fornecer uma visão geral das informações mais importantes do espectro, destacando aquelas regiões com maior relevância para construção do modelo PLS, em relação ao modelo desenvolvido com uso do espectro inteiro. Essa comparação é balizada pelo valor da raiz quadrada do erro médio quadrático de validação cruzada (RMSECV). Duas

importantes vantagens da utilização deste método são: a possibilidade de melhora na capacidade preditiva da regressão PLS, assim como a demanda de menor tempo de análise devido à avaliação de um número menor de variáveis analisadas. Esta segunda, em particular, é fundamental para utilização de dados obtidos através de medições on-line aplicadas a controle de processos em tempo real (NØRGAARD et al., 1999).

2.3.3.4 Seleção de amostras pelo algoritmo SPXY

Conforme descrito por Galvão et al. (2005), este método é utilizado para particionar as amostras do conjunto de dados originais em dois subconjuntos, o de calibração e o de predição. Para esta etapa, é necessário que os dados já tenham sido pré-tratados, e que amostras que apresentarem comportamento anômalo (*outliers*) tenham sido removidos. A divisão das amostras leva em consideração a variabilidade conjunta da matriz de dados espectroscópicos \mathbf{X} e do vetor \mathbf{y} que contém os valores de referência.

O algoritmo calcula a distância Euclidiana na matriz \mathbf{X} e na variável \mathbf{y} e seleciona iterativamente um par de amostras (p, q) que apresenta a maior distância entre si, $d_{xy}(p, q)$, de modo que para as novas iterações, a segunda maior distância mínima é tomada entre duas outras amostras, e assim sucessivamente. A distância $d_{xy}(p, q)$ corresponde, portanto, a soma das distâncias Euclidianas em \mathbf{X} e \mathbf{y} , divididas pelo valor máximo de cada conjunto, a fim de que uma distância normalizada seja obtida. O cálculo de $d_{xy}(p, q)$ está descrito na Equação 29.

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p, q \in [1, N]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p, q \in [1, N]} d_y(p, q)}; p, q \in [1, N] \quad (29)$$

2.4 CONTROLE ESTATÍSTICO MULTIVARIADO DE PROCESSOS EM BATELADA BASEADO EM VARIÁVEIS LATENTES

Processos produtivos convencionais em batelada têm a característica comum de apresentar duração finita. Basicamente ocorrem em três etapas gerais: introdução da matéria-prima, processamento dos componentes sob condições físicas controladas e descarga do produto. Em alguns casos, apenas ao final do processo, ensaios de aferição da qualidade do produto são realizadas. No entanto, variações normalmente ocorrem nas características dos produtos obtidos entre uma batelada e outra, podendo ser atribuídas a desvios nas variáveis de processo, à qualidade da matéria-prima empregada, e até mesmo a flutuações naturais do

processo. Nesse sentido, monitorar apenas a qualidade do produto de cada batelada torna-se uma maneira de avaliação parcial, visto que alterações que possam vir a ocorrer nas variáveis de processo ao longo da batelada podem alterar o seu desempenho, sendo detectadas muitas vezes só após uma ou algumas bateladas serem terminadas. Assim, a realização do monitoramento em tempo real das variáveis de processo e dos atributos de qualidade, torna-se necessária e dá origem a uma abordagem de avaliação robusta capaz de fornecer uma impressão digital da trajetória das variáveis de processo para cada batelada (NOMIKOS; MACGREGOR, 1995).

Nesse sentido, com o desenvolvimento acelerado da tecnologia, principalmente com a implantação da automação industrial, uma grande quantidade de medidas passou a ser coletada simultaneamente por diferentes instrumentos e compilada por computadores. Desse modo, a interpretação univariada dessa gama de informação vem se tornando cada vez mais inviável para interpretação e atuação corretiva em tempo real, quando necessária. Sendo assim, gráficos de controle estatístico multivariado de processos baseados em métodos de projeção multivariada como PCA e PLS vêm sendo empregados como uma ferramenta para reduzir a dimensionalidade das variáveis originais por meio da exploração da variabilidade e/ou estrutura de correlação entre estas. A informação correlacionada das variáveis originais é então projetada em um subespaço de baixa dimensionalidade definido pelas variáveis latentes. Portanto, esta é uma alternativa atrativa para aplicação em ambientes ricos em dados (KOURTI; MACGREGOR, 1996; FERRER, 2007; SALES, 2016; CATELANI et al., 2018).

Para aplicação da abordagem multivariada, é necessário conhecer o processo tecnicamente e obter um conjunto de dados de processo que seja representativo com relação ao comportamento esperado do processo estudado. Por isso, os dados históricos das variáveis de processo são muito valiosos durante a fase inicial do desenvolvimento de estratégias de MSPC. A condição padrão de um determinado processo, estabelecida levando-se em conta as especificações técnicas do processo e a qualidade do produto obtido, é denominada de condição normal de operação (NOC, do inglês *Normal Operating Condition*), e serve como referência para produção de outras bateladas. A partir dos dados de processos obtidos em bateladas sob condições normais de operação, gráficos ou cartas de controle podem ser construídos e utilizados para avaliar a dinâmica das variáveis de processo na produção de bateladas futuras.

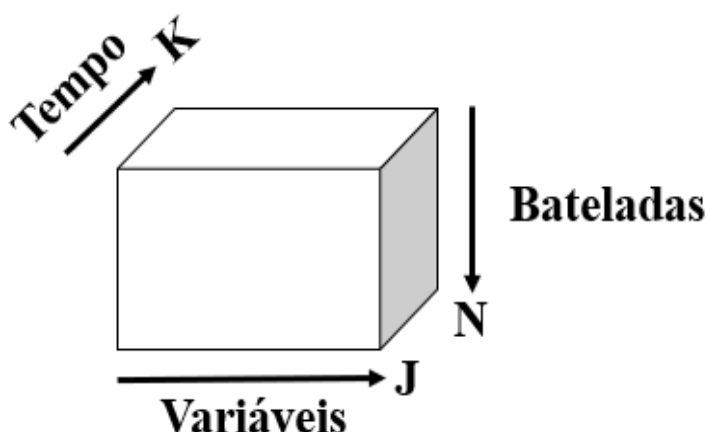
Assim, o desenvolvimento de cartas de controle baseadas nas estratégias MSPC normalmente ocorrem em duas etapas. Na etapa I, conhecida como etapa de calibração, um modelo multivariado é construído a partir de dados históricos provenientes de bateladas NOC. Nessa etapa são estipulados os limites de controle, levando em consideração a variação natural

entre as bateladas, bem como a correlação entre as variáveis. Na etapa II, a etapa de predição, os dados obtidos em novas bateladas são projetados no modelo multivariado desenvolvido com as bateladas NOC. Assim, a detecção de possíveis falhas nesse novo conjunto de bateladas pode ser realizada (NOMIKOS; MACGREGOR, 1995; FERRER, 2007).

2.4.1 Abordagens de desdobramento da matriz de dados espectroscópicos

Durante o monitoramento em tempo real do processo em batelada, J variáveis são medidas durante K intervalos de tempo. Considerando a produção de N bateladas, esses dados podem ser agrupados em uma matriz tridimensional ($N \times J \times K$), conforme a Figura 6. Ao longo da direção vertical estão as bateladas, no eixo horizontal as variáveis (dados de processo, dados espectroscópicos, etc.) e na terceira dimensão, os intervalos de tempo.

Figura 6 – Agrupamento dos dados oriundos do monitoramento de N bateladas.



Fonte: Adaptado de Nomikos e Macgregor (1995).

O tipo de informação extraída da matriz tridimensional e a interpretação dos dados presentes no arranjo da Figura 6 depende do tipo de tratamento aplicado a esta matriz. A matriz tridimensional pode ser manipulada de diferentes formas que podem ser adotadas de acordo com o objetivo da análise realizada. Para construção das cartas de controle através dos métodos de projeção multivariados, é necessário desdobrar esses dados organizados em 3 dimensões em uma matriz bidimensional, para que posteriormente seja possível aplicar métodos de projeção de dados, como PCA ou PLS (NOMIKOS; MACGREGOR, 1995).

Duas abordagens de desdobramento e modelagem dos dados são mais comumente utilizadas para extração da informação presente na matriz tridimensional. Uma delas foi

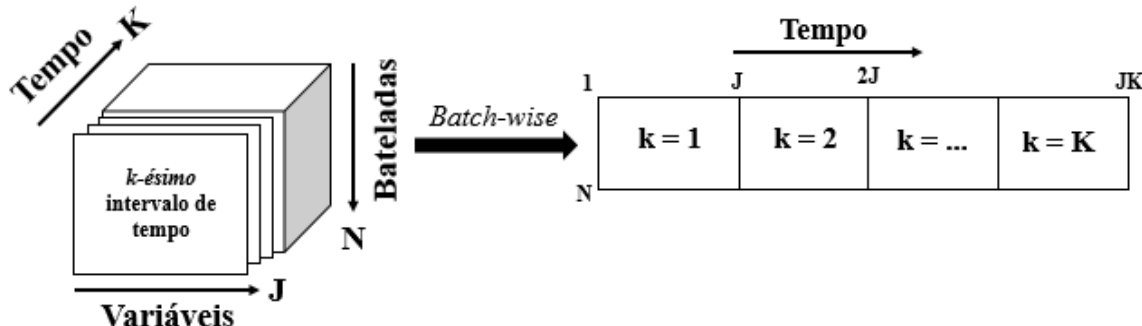
proposta por Nomikos e Macgregor, (1995), em que a matriz de dados da Figura 6 sofre inicialmente cortes verticais no sentido do eixo das bateladas. Este tipo de desdobramento é chamado de *batch-wise*, visto que o eixo das bateladas permanece inalterado para cada intervalo de tempo. A segunda abordagem foi proposta por Wold et al. (1998), e promove inicialmente o rearranjo da matriz de dados a partir de cortes horizontais da matriz tridimensional em relação ao eixo do tempo. Esse desdobramento é descrito como *variable-wise*, haja vista que o eixo das variáveis fica inalterado. Essas abordagens, descritas com mais detalhe a seguir, são apresentadas, respectivamente, como NM e WKFH (Wold, Kettaneh, Friden e Holmberg), em referência aos autores.

Ambas as abordagens são utilizadas tanto para análises de bateladas após o seu término (do inglês: *end of batch*) quanto para avaliação em tempo real (do inglês: *real time*). Na abordagem *end of batch*, é possível diagnosticá-las quanto ao desempenho esperado, através da comparação com as bateladas NOC. Por outro lado, na abordagem *real time*, é possível monitorar a batelada durante a sua evolução. Assim, quando verificado um desvio anormal significativo em relação às condições normais de operação, é possível tomar medidas corretivas a fim de conduzir o processo de volta para a condição desejável (GONZÁLEZ-MARTINEZ; CAMACHO; FERRER, 2018).

2.4.1.1 Desdobramento e análise da matriz tridimensional baseado na abordagem NM

Na abordagem NM, a matriz tridimensional da Figura 6 é inicialmente rearranjada a partir de cortes verticais ao longo do eixo das bateladas, como pode ser evidenciado na Figura 7. Cada camada correspondente a um intervalo de tempo é então agrupada lado a lado, tornando o eixo das bateladas inalterado (*batch-wise*). Após o rearranjo, é obtida então uma matriz bidimensional ($N \times JK$), na qual cada linha corresponde às informações de uma batelada inteira, possibilitando a observação da dinâmica do processo ao longo do tempo (NOMIKOS; MACGREGOR, 1995).

Figura 7 – Desdobramento da matriz tridimensional por meio da abordagem NM.



Fonte: Adaptado de Nomikos e Macgregor (1995).

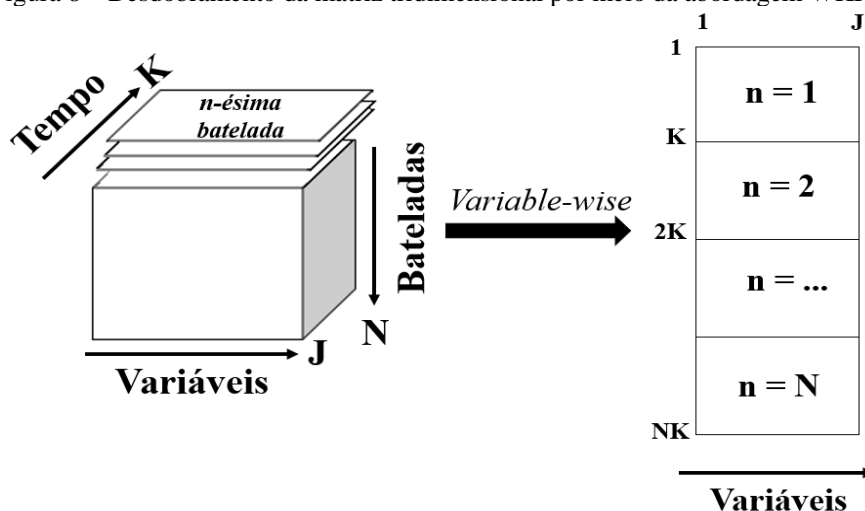
Conforme descrito por Nomikos e Macgregor (1995), após a operação de rearranjo da matriz tridimensional, a etapa de centragem dos dados na média é de suma importância para remoção da não-linearidade dos dados. Nesse sentido, por meio da subtração da média dos valores de cada coluna, é possível remover a trajetória média de cada variável. Em seguida, aplica-se a PCA aos dados pré-processados e obtidos sob condições normais de operação (bateladas que compõem o conjunto de calibração). Como resultado, cada elemento dos escores corresponde a uma única batelada completa e representa a variabilidade desta em relação às demais. Já o vetor de pesos fornece as direções de maior variabilidade do conjunto de dados.

Conforme discutido anteriormente, o monitoramento de um processo em batelada pode ser efetuado após seu término ou em tempo real. Para a estratégia de desdobramento NM, cada linha de matriz 2D representa as medidas em todos os tempos de uma determinada batelada. Assim, para aplicar o modelo a uma nova batelada (nova linha), é necessário esperar o fim de sua operação para que sejam adquiridas as medidas em todos os tempos. Entretanto, alguns métodos para realizar o preenchimento de valores futuros para as medidas, durante o andamento da batelada, podem ser utilizados para habilitar a utilização das cartas de controle em tempo real. Nomikos e Macgregor, (1995) descrevem algumas alternativas com esta finalidade, como: preencher as observações futuras com zeros, efetuar o preenchimento com valores que mantenham os desvios das trajetórias médias ao longo da batelada, e a aplicação do modelo PCA aos dados de referência (conjunto de calibração) para estimar as observações futuras. Todos esses métodos pressupõem que as observações futuras respeitam as trajetórias médias das bateladas do conjunto de calibração.

2.4.1.2 Desdobramento e análise da matriz tridimensional baseado na abordagem WKFH

Na abordagem WKFH, proposta por Wold et al. (1998), a matriz de dados da Figura 6 é inicialmente rearranjada a partir de cortes horizontais ao longo do eixo dos intervalos de tempo. Em seguida, as camadas são agrupadas uma abaixo da outra, de modo que o eixo das variáveis permanece inalterado (*variable-wise*), conforme ilustrado na Figura 8. Depois do rearranjo executado, uma matriz bidimensional ($NK \times J$) é obtida, em que cada coluna representa a intensidade do sinal referente a uma variável para todas as bateladas, neste caso, a matriz é composta pelos valores de absorbância para cada comprimento de onda (variáveis).

Figura 8 – Desdobramento da matriz tridimensional por meio da abordagem WKFH.



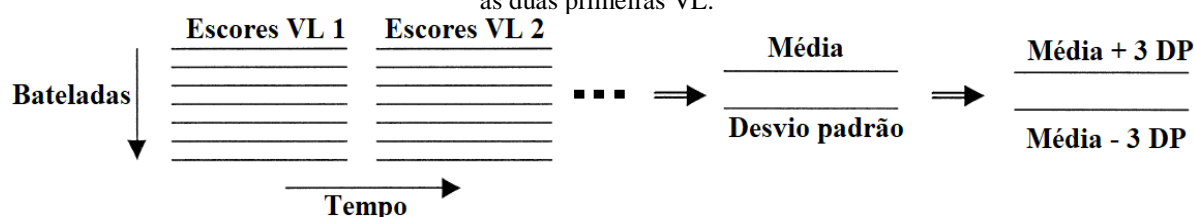
Fonte: Adaptado de Wold et al. (1998).

Diferentemente do que acontece na abordagem de desdobramento NM, na abordagem WKFH a centragem na média da matriz bidimensional remove apenas a grande média do conjunto de dados, mantendo a variabilidade relacionada com a trajetória média. Posteriormente, para realização do monitoramento de bateladas em tempo real, um modelo PLS é construído, conforme descrito no item 2.3.3.1. A construção do modelo PLS ocorre a partir dos dados da matriz bidimensional desdobrada e um vetor que indique o tempo local ou índice de maturidade para as bateladas, de modo que o vetor predito pelo modelo PLS indica a evolução da batelada.

Os escores do modelo PLS são então utilizados para construção das cartas de controle. Para isso, os vetores de escores para cada batelada, para cada variável latente, são agrupados lado a lado, formando uma matriz com N linhas (bateladas) e HK colunas. Em que H é o número de variáveis latentes utilizadas e K são os intervalos de tempo. Dessa matriz são estimados os limites de controle para cada intervalo de tempo de acordo com os valores da média e do desvio padrão calculados para cada coluna. Esta etapa está representada na Figura 9. Uma grande

vantagem dessa abordagem é o fato de não ser necessário a imputação de valores futuros para predição de bateladas em processamento, como exigido pela abordagem NM. Conforme descrito por Flores-Cerrillo e Macgregor, (2004), outro ponto positivo é o fato de a modelagem ser realizada a partir dos escores da matriz desdobrada, frente à utilização dos dados originais como faz a abordagem NM. Assim, reduz-se o tempo de trabalho computacional, visto que devido à alta correlação entre as variáveis de processo, estas podem ser representadas em um espaço de poucas dimensões, através dos métodos projeção.

Figura 9 – Rearranjo dos escores do modelo PLS para construção dos gráficos de controle exemplificado para as duas primeiras VL.



Fonte: Adaptado de Wold et al. (1998).

O monitoramento do processo em batelada além de efetuado em tempo real, conforme descrito acima, também ocorre no modo pós-batelada. Para este último, os escores provenientes do modelo PLS, desenvolvido anteriormente, são rearranjados de acordo com a abordagem NM (Figura 7), sendo submetidos posteriormente a uma PCA, conforme descrito por Wold et al. (1998).

2.4.2 Diagnóstico de falhas em processos em batelada

Uma vez construídas as cartas de controle baseadas em métodos de projeção PCA e/ou PLS sobre os dados de bateladas obtidas em condições normais de operação (conjunto de calibração), é possível monitorar as novas observações referentes às bateladas do conjunto de predição por meio da projeção dos escores dos novos dados pré-processados sobre o plano do modelo de referência construído. Nesse sentido, por meio dos dados referentes aos escores e os resíduos da modelagem, eventos anômalos podem ser detectados principalmente por meio de cartas de controle referentes a duas estatísticas, o erro quadrático de predição (SPE, do inglês *Squared Prediction Error*) e o T^2 de Hotelling. A primeira representa o quadrado da distância Euclidiana (distância perpendicular) de uma nova observação em relação ao plano do modelo construído com as bateladas NOC. Quando uma observação na carta da estatística SPE supera o limite de controle para um dado nível de confiança, observa-se que tal batelada apresenta um

comportamento diferente do apresentado pelo conjunto de calibração. Assim, é factível afirmar que houve uma quebra na estrutura de correlação das variáveis no modelo. O cálculo da estatística SPE pode ser evidenciado na Equação 30 (FERRER, 2007).

$$SPE = \mathbf{e}_i^t \mathbf{e}_i = (\mathbf{x}_i - \mathbf{x}_i^*)^t (\mathbf{x}_i - \mathbf{x}_i^*) \quad (30)$$

Em que (\mathbf{e}_i) é vetor de resíduo da i -ésima observação, (\mathbf{x}_i^*) é o valor predito pelo método de projeção adotado, a partir da observação (\mathbf{x}_i) . Outra forma de expressar a distância de novas observações frente ao plano do modelo é em termos da distância absoluta corrigida pelo desvio padrão, através da estatística DModX descrita pela Equação 31.

$$DModX = c \sqrt{\frac{SPE}{(J - h)}} \quad (31)$$

Em que (c) corresponde a um fator de correção, que depende do número de observações e do número de componentes. O mesmo apenas é considerado no cálculo da carta de DModX quando o número de observações do conjunto de calibração é pequeno. Já na etapa de predição, o cálculo de DModX é efetuado considerando $c = 1$. Como já mencionado anteriormente, (J) é o número de variáveis, e (h) é o número de variáveis latentes.

Por outro lado, segundo Kourti, (2005), a estatística T^2 de Hotelling avalia se a projeção de um novo vetor de medidas observado para as (J) variáveis de processo, está dentro do subespaço latente definido na etapa de predição. Esta avaliação leva em consideração a distância de Mahalanobis estimada do centro do subespaço latente até a projeção da nova observação. Portanto, quando uma nova observação ultrapassa o limite de controle estabelecido pela carta T^2 de Hotelling, é um indicativo de que as variáveis apresentam valores anômalos em relação às condições normais de operação. No entanto, ao contrário da carta DModX, isso não é um indicativo de que houve uma quebra de correlação entre as variáveis.

A estatística T^2 de Hotelling pode ser calculada através da Equação 32 (FERRER, 2007).

$$T^2 \text{ Hotelling} = \sum_{h=1}^H \frac{t_h^2}{s^2 t_h} \quad (32)$$

Em que \mathbf{t}_h podem ser os escores de um modelo PCA ou PLS construídos com às variáveis de processo, e $\mathbf{s}^2\mathbf{t}_h$ é a variância estimada da variável latente correspondente a \mathbf{t}_h .

2.4.3 Controle estatístico multivariado de processos como uma ferramenta de PAT

Vários trabalhos vêm descrevendo a aplicação do controle estatístico multivariado como uma ferramenta analítica de processo para detecção e análise de falhas em uma ampla faixa de processos em batelada, inclusive no monitoramento da produção de biodiesel.

Castelani et al. (2018) descreveram a utilização de MSPC associado à espectroscopia NIR *in-line* como uma ferramenta analítica de processo para realizar o monitoramento em tempo real do processo de torra do café em batelada. Neste estudo, a matriz tridimensional de dados foi rearranjada via *variable wise*, e na etapa de modelagem, uma PCA foi aplicada sobre a nova matriz de dados, mostrando-se uma estratégia eficaz para detecção de falhas para o sistema estudado.

Aguado et al. (2007) avaliaram a aplicação de MSPC ao processo de tratamento de efluentes em batelada. Neste trabalho, os autores realizaram uma comparação em relação ao desempenho do monitoramento nos modos pós-batelada e em tempo real. Para isto aplicaram a abordagem NM, e WKFH para o desdobramento da matriz de dados, de modo que o método de monitoramento mais eficaz foi observado quando utilizaram a abordagem NM para ambos os modos pós batelada e em tempo real em um esquema combinado entre os dois modos. As cartas de SPE, e T^2 de Hotelling foram empregadas para detecção de falhas, demonstrando resultados satisfatórios para monitoramento em estudo.

Kona et al. (2013) descreveram a utilização de MSPC associada a espectroscopia NIR *in-line*, para monitorar o processo em batelada de granulação e secagem em leito fluidizado de mistura de ativos utilizados pela indústria farmacêutica, a partir da obtenção de medidas de umidade e temperatura. Os autores utilizaram os gráficos dos escores, DModX e T^2 de Hotelling para realizar a detecção e análise de falhas a partir da abordagem WKFH para o monitoramento em tempo real. Esta estratégia de monitoramento foi capaz de identificar anomalias presentes no processo, contribuindo também para a compreensão das possíveis causas atribuíveis.

Li et al. (2016) descreveram a aplicação de MSPC associado a espectroscopia NIR *in-line* para monitorar o processo produtivo de medicamentos tradicionais chineses. Eles utilizaram o método *variable wise* para realizar o desdobramento da matriz tridimensional de dados. Em seguida, uma PCA foi aplicada sobre a nova matriz obtida, e o monitoramento do processo em batelada foi realizado a partir do gráfico dos escores e das estatísticas DModX e

T² de Hotelling. As cartas apresentaram desempenho satisfatório para detecção de diferentes tipos de desvios simulados sobre a temperatura e agitação do meio reacional nas bateladas de predição. Das quais, a carta DModX apresentou maior sensibilidade na detecção de falhas, e nas demais, tais desvios foram percebidas com algum atraso.

Lima et al. (2014) descreveram a utilização de cartas de controle multivariada de processo associado a espectroscopia NIR *in-line*, para monitorar a reação de transesterificação do óleo de soja por rota metílica e em batelada, para produção de biodiesel. A matriz de dados originais foi desdobrada apenas pela abordagem WKFH, em que a partir do gráfico dos escores foi possível detectar de modo eficiente a falha intencionalmente produzida a partir de alteração na temperatura durante a reação para avaliação do modelo construído.

Sales et al. (2016) descreveram a aplicação de MSPC para monitorar a reação de transesterificação do óleo de soja por rota metílica e em batelada para obtenção de biodiesel, utilizando a espectroscopia NIR *in-line*, empregando uma sonda de transflectância e um equipamento de bancada. Os autores avaliaram os modos de desdobramento da matriz de dados NM e WKFH para construção de cartas de controle, para fins comparativos em termos de desempenho na detecção e análise de falhas em tempo real e no modo pós batelada. Segundo os autores, ambas as abordagens apresentaram desempenho semelhantes para identificação dos desvios nas bateladas de predição para o modo pós batelada. Por outro lado, a abordagem NM na análise em tempo real demonstrou melhores resultados para detecção das perturbações provocadas nas bateladas de predição, provavelmente devido a esta abordagem ter a capacidade de capturar a dinâmica da reação.

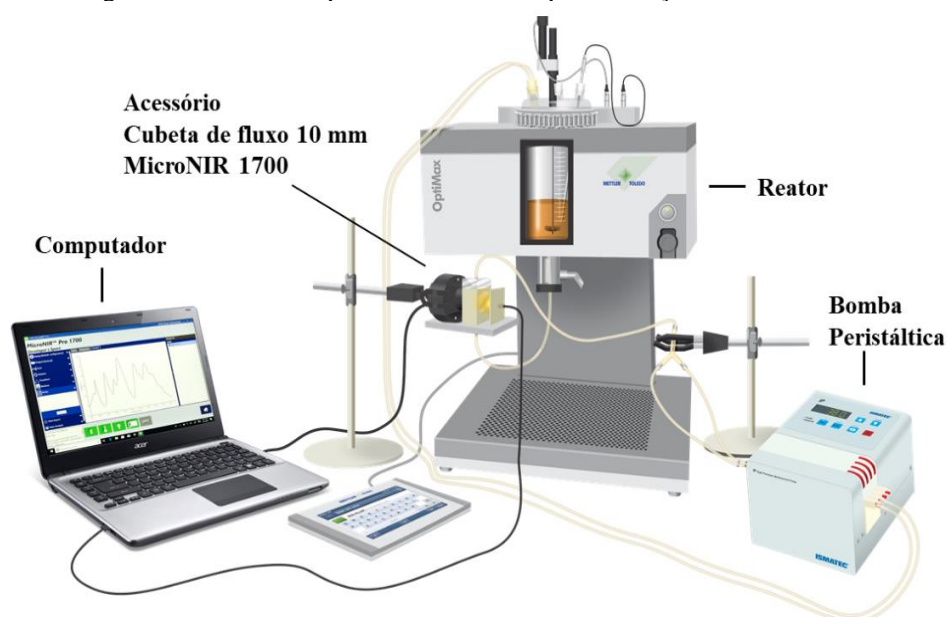
O presente trabalho está inserido neste contexto no sentido de demonstrar o potencial de um espectrômetro NIR ultracompacto, de menor faixa espectral, custo e resolução que um equipamento de bancada, para aplicação no monitoramento *on-line* da reação de transesterificação de biodiesel, fornecendo informações em tempo real sobre desvios no processo e auxiliando a tomada de decisão.

3 MATERIAIS E MÉTODOS

3.1 MONTAGEM DO SISTEMA REACIONAL

O sistema empregado para realização das reações de transesterificação foi composto por um reator equipado com controles integrados de temperatura e velocidade de agitação, modelo OptiMax™ 1001 (Mettler Toledo), utilizando um vaso de 1000 mL. Tais variáveis foram monitoradas *in-line* por meio do *software* da Mettler Toledo (iControl 5.2) com valores registrados a cada 2 s. Na saída do vaso reacional foi adaptada uma linha de amostragem em fluxo para aquisição espectral a partir do equipamento portátil utilizado: MicroNIR™ PRO 1700 (VIAVI Solutions Inc.), no modo *on-line*. O MicroNIR foi conectado a um outro computador para viabilização da aquisição dos espectros, sendo anexado em um acessório para análise por transmitância (PAIVA et al. 2015). Esse acessório suporta uma mini-lâmpada de tungstênio como fonte de luz, a qual é alimentada via conexão USB do computador, e dá suporte a uma cubeta, empregada como célula de residência do volume de amostragem para medida em fluxo contínuo dos espectros NIR pelo equipamento portátil. A linha de amostragem que conecta o sistema, composta por mangueiras de silicone, após a medida *on-line*, é novamente direcionada ao reator, recompondo o volume removido temporariamente por meio de uma bomba peristáltica (Ismatec IPC – 4) em vazão constante. Um esquema do sistema integrado descrito está representado na Figura 10.

Figura 10 – Sistema experimental utilizado para obtenção das bateladas.



Fonte: O autor (2020).

3.2 PRODUÇÃO DE BIODIESEL EM BATELADA

A reação de transesterificação foi promovida utilizando óleo de algodão refinado (Flor de algodão), etanol P.A (Neon – 98%) e hidróxido de sódio (NaOH) (Neon, de pureza 99,2%). Considerou-se a composição do óleo de algodão descrita por Pinto et al. (2005), uso de excesso de etanol na razão molar de 12:1 (álcool:óleo) e concentração de catalisador igual a 0,5% de hidróxido de sódio em massa em relação à massa de óleo; assim como temperatura de 50°C e a velocidade de agitação de 250 rpm. As condições reacionais empregadas foram adaptadas do trabalho de Joshi e colaboradores (JOSHI; TOLER; WALKER, 2008).

Com o auxílio de um béquer de 500 mL e um bastão de vidro, procedeu-se a pesagem de 245,24 g do óleo vegetal em balança semi-analítica (Mettler Toledo – PB3002-S), sendo reservado posteriormente. Na sequência, utilizando um béquer de 50 mL e com o auxílio de uma espátula de alumínio, pesou-se 1,2361 g de NaOH em balança analítica (Sartorius – BL210S), o qual foi dissolvido em etanol, e transferido para um balão volumétrico de 200 mL, sendo aferido na sequência com etanol até o menisco.

A seguir, o óleo pesado foi inserido no reator utilizando um funil simples para transferência, sendo acionada a agitação em sequência, bem como o aquecimento. Paralelamente, a solução de catalisador também foi aquecida até 50°C, sendo monitorada por um termômetro. Após estabilização da agitação e da temperatura no reator, bem como da temperatura da solução etílica, a solução catalisadora (etanol + hidróxido de sódio) foi introduzida ao reator, A seguir iniciou-se o monitoramento com a aquisição espectral por cerca de 60 minutos.

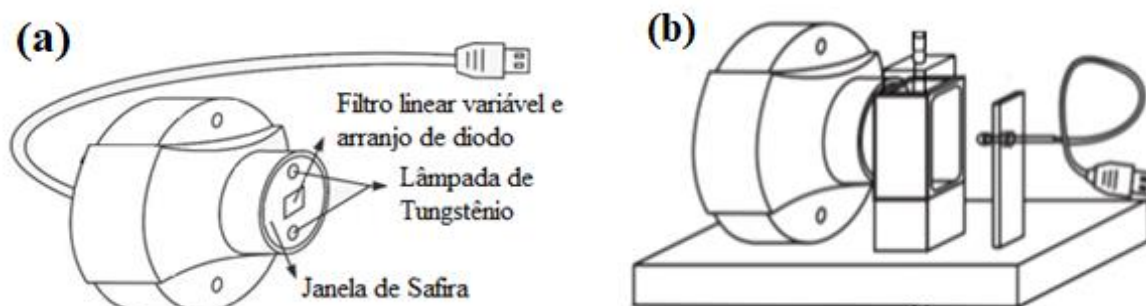
Foram realizadas 10 bateladas em condições estáveis, sem quaisquer alterações durante a reação, 1 batelada com alteração simulada na variável temperatura, correspondendo ao aumento de 5°C durante 5 minutos após estabilização da temperatura na mistura reacional, e 1 batelada com aumento da concentração do catalisador para 0,9%. Todas monitoradas *on-line* pelo MicroNIR. As 10 primeiras foram definidas como bateladas em condições normais de operação.

Os espectros coletados ao longo das bateladas foram utilizados para a construção de cartas de controle multivariadas e os últimos 10 espectros de 11 bateladas foram empregados para construção de um modelo de calibração multivariada, cujo procedimento detalhado está descrito no item 3.3. Os produtos de todas as bateladas foram submetidos ao processo de purificação, sendo posteriormente analisados por cromatografia gasosa para estimar o percentual de conversão de etil-ésteres.

3.2.1 Aquisição dos espectros NIR

O equipamento portátil empregado para aquisição dos espectros opera na faixa de 908 – 1676 nm, funciona via conexão USB junto a um computador e é composto por: 2 lâmpadas de tungstênio como fonte de radiação interna, um filtro variável linear (LVF), usado como elemento dispersivo, assim como um detector de 128 elementos (InGaAs), apresentando resolução de 12,5 nm em 1000 nm. A medida dos espectros durante a reação foi realizada por transmitância, mantendo a fonte do MicroNIR desligada e utilizando a lâmpada de tungstênio (1W e 5V) do acessório de transmitância, que é alimentada via conexão USB (PAIVA et al., 2015). O aparato para realização das medidas *on-line* de transmitância e os detalhes do espectrômetro estão mostrados na Figura 11. O *background* (“dark” 0% de transmitância) foi medido por meio da obstrução da passagem de radiação para o detector do instrumento; e o espectro de referência que corresponde a 100% de transmitância foi obtido com a cubeta de fluxo vazia, com caminho óptico de 10 mm. Utilizando-se o software MicroNIR™ PRO versão 2.2, os espectros foram adquiridos no modo contínuo, resultando em uma medida a cada 3 segundos em média durante 1 hora.

Figura 11 – Ilustração do (a) equipamento portátil MicroNIR 1700 e (b) equipamento portátil acoplado ao acessório de transmitância.



Fonte: Adaptado de Paiva et al. (2015).

3.2.2 Purificação do biodiesel produzido

Após o término de cada batelada, foram adicionados cerca de 250 mL de uma solução de ácido clorídrico ($\text{HCl} - 0,5 \text{ mol.L}^{-1}$) ao reator, sendo acionada a agitação por cerca de 30 segundos e desligada em seguida. Na sequência, aguardou-se a separação das fases e a fase inferior aquosa, rica em glicerina, foi removida na saída inferior do reator. O procedimento de lavagem foi repetido a 75°C , utilizando a mesma solução de HCL diluída na proporção de 1:1

(v/v), em seguida, a lavagem foi repetida por aproximadamente três vezes com água destilada, até o pH do meio se igualar ao pH da água destilada (BAPTISTA et al., 2008).

A fase rica em éster e óleo, foi submetida à agitação a 250 rpm e aquecimento à temperatura de 120°C, ainda no mesmo reator, por cerca de 1 hora e 15 minutos para evaporação da água proveniente da etapa de lavagem. Após o resfriamento, adicionou-se ao biodiesel uma quantidade de aproximadamente 10% da massa do biodiesel produzido de sulfato de sódio anidro (Na_2SO_4), para remoção da umidade restante. Na sequência, a mistura (biodiesel + Na_2SO_4) foi filtrada à vácuo, com auxílio de uma bomba à vácuo (Prismatec – 121), utilizando papel de filtro qualitativo (80g Unifil, diâmetro 12,5 cm, teor de cinzas máx 0,1%). (LIMA, 2014).

3.2.3 Análise cromatográfica

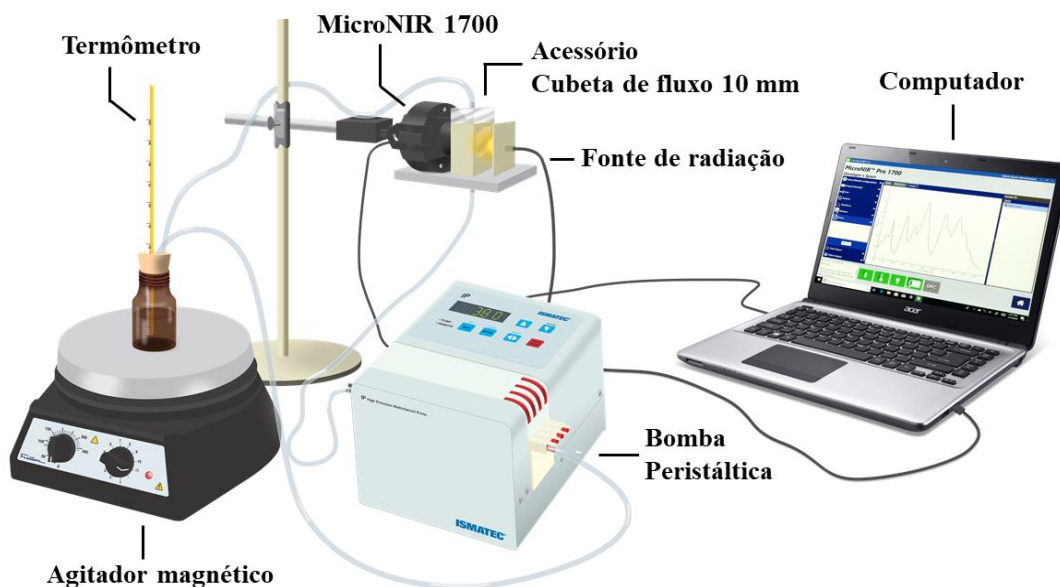
O biodiesel produzido nas bateladas, devidamente purificado, foi avaliado por uma metodologia padronizada para biodiesel, em que o teor de éster foi analisado em um cromatógrafo a gás (GC) modelo 17-A da SHIMADZU com injeção automática, e detector de ionização de chama (FID, do inglês *Flame ionization detector*), segundo a *European Standard* (EN-14103) (EUROPEAN STANDARD, 2003) utilizando o padrão interno C17 (metil heptadecanoato).

3.3 MISTURAS SINTÉTICAS

Um planejamento de misturas foi realizado para a preparação de misturas quaternárias sintéticas contendo glicerina, óleo, biodiesel e etanol. Durante a preparação, as massas foram pesadas em ordem crescente de volatilidade em frasco âmbar, com auxílio de uma balança analítica. A faixa de concentração dos componentes das amostras sintéticas em (% m/m) foram: 4 – 31% de óleo de algodão, 30 – 34% de etanol, 31 – 60% de biodiesel e 3 – 6% de glicerina.

Amostras em proporções estequiométricas também foram incorporadas ao planejamento de misturas, o qual está descrito no Apêndice A. Para adquirir os espectros das amostras sintéticas foi utilizado um sistema em fluxo de modo a reproduzir uma condição mais próxima possível das medidas realizadas no sistema integrado ao reator (Figura 12), seguindo o mesmo procedimento experimental descrito em 3.2.1 para obtenção do *background*. Uma representação do sistema experimental pode ser observada na Figura 12.

Figura 12 – Representação do sistema operacional para aquisição dos espectros para as misturas sintéticas.



Fonte: O autor (2020).

Como observado na Figura 12, os espectros NIR das misturas sintéticas também foram obtidos por transmitância, e em fluxo, utilizando uma cubeta de fluxo de caminho óptico igual 10 mm, porém, a aquisição espectral foi realizada manualmente, em que cada amostra foi analisada 3 vezes à temperatura de 50°C, condição similar as bateladas NOC, e 10 delas foram obtidas também à 60°C. Posteriormente, a média das triplicatas foi utilizada na construção de modelos de calibração.

3.3.1 Tratamento dos dados e construção do modelo multivariado

O conjunto de dados utilizados nos modelos de calibração multivariada é composto por 65 espectros, no total, sendo 54 espectros adquiridos das amostras sintéticas e 1 espectro de cada batelada. Os espectros das bateladas foram obtidos como a média dos últimos 10 espectros de cada batelada.

Os teores de éster de referência para todas as amostras foram determinados considerando-se a mistura quaternária de óleo de algodão, biodiesel, etanol e glicerina. No caso das amostras das bateladas o teor de éster foi determinado por cromatografia na mistura binária biodiesel/óleo, obtida após a purificação, e posteriormente convertido considerando a mistura quaternária por meio da relação estequiométrica utilizada nas reações.

Foram utilizadas diferentes estratégias de pré-processamento de dados, como: primeira derivada com filtro de Savitzky-Golay (considerando-se um polinômio de segunda ordem com

tamanho da janela de 5, 7 e 9 pontos), Correção de Espalhamento Multiplicativo e Variação Normal Padrão. Todos os modelos foram construídos com os dados centrados na média. A seleção de amostras para os conjuntos de calibração e predição foi realizada utilizando o algoritmo SPXY para as amostras sintéticas (GALVÃO et al., 2005). Adicionalmente, espectros de duas bateladas foram inseridos no conjunto de calibração. Espectros de outras 9 bateladas foram incluídos no conjunto de predição.

Para a construção do modelo de calibração multivariada, considerou-se 2 abordagens diferentes:

- a) Modelo PLS utilizando dos espectros NIR na faixa completa de aquisição.
- b) Modelo PLS construídos por faixas do espectro selecionadas pelo algoritmo de seleção de variáveis *i*PLS.

Os modelos construídos baseados nas estratégias acima foram avaliados de acordo com os erros de validação cruzada (RMSECV), de predição (RMSEP), assim como pelo número de variáveis latentes utilizado e pela melhor estratégia de pré-processamento capaz de minimizar as diferenças entre os espectros das bateladas e os espectros das misturas sintéticas. O número adequado de variáveis latentes para os modelos PLS construídos foi determinado utilizando-se o método de validação cruzada *venetian blinds*. Para a verificação da existência de diferença significativa entre os valores de RMSEP entre diferentes modelos, para o nível de confiança de 95%, foi realizado teste F. Neste teste, o valor de F calculado (F_{calc}) corresponde à relação entre os quadrados dos maiores e menores valores de RMSEP. Também foi avaliada a existência de erros sistemáticos significativos (*bias*) por meio do teste *t-student*, para o mesmo nível de confiança.

3.4 CONTROLE ESTATÍSTICO MULTIVARIADO DO PROCESSO DE TRANSESTERIFICAÇÃO EM BATELADA

O monitoramento da reação de transesterificação, foi realizado para as 10 bateladas produzidas sob condições normais de operação (1 – 10), 1 batelada com aumento de 5°C na temperatura por 5 minutos após estabilização inicial da mistura no vaso reacional. E mais 1 batelada com concentração de catalisador igual 0,9%, bateladas 11 e 12, respectivamente.

3.4.1 Bateladas em Condição Normal de Operação: Etapa de calibração

Para a construção das cartas multivariadas de controle, foram utilizadas 8 bateladas (1-3, 5-7, 9 e 10) (NOC), isto é, sem quaisquer alterações provocadas intencionalmente durante a reação, conforme as condições descritas no item 3.2. Nesta etapa procurou-se incluir toda a variabilidade natural presente ao longo do processo reacional, de modo a criar um conjunto de dados robusto e representativo desta condição.

3.4.2 Etapa de predição

O conjunto de dados de predição é formado por 2 bateladas produzidas em condições normais de operação (extraídas aleatoriamente do conjunto de calibração) e 2 bateladas que sofreram perturbações nas variáveis de processo ao longo de sua operação. Este conjunto de 4 bateladas foi utilizado para avaliar a capacidade das cartas de controle desenvolvidas em identificar falhas relacionadas às variações de temperatura e concentração de catalisador. Todas as perturbações nas variáveis de processo das bateladas produzidas fora de controle foram realizadas cerca de 90 segundos após a mistura dos reagentes no reator. Os tipos de perturbações simuladas para as bateladas da etapa de predição estão resumidos na Tabela 3. O tempo de duração de todas as bateladas foi cerca de 60 minutos. A perturbação da batelada 11 deu-se por meio de uma rampa de temperatura programada no próprio controlador do reator, cuja automação proporcionou um controle de início e fim das alterações.

Tabela 3 – Perturbações simuladas na preparação e durante as bateladas de predição.

Batelada	Perturbações
4	-
8	-
11	Aumento de temperatura do vaso em 5°C por cerca de 5 min.
12	Uso de 0,9% de catalisador em massa

Fonte: O autor (2020).

3.4.3 Desenvolvimento das cartas de controle multivariadas

Durante as reações, a faixa de comprimentos de onda em estudo é formada por 125 variáveis que foram adquiridas em 916 intervalos de tempo. Considerando ainda que foram produzidas 12 bateladas (tanto do conjunto de calibração quanto da etapa de predição), esse

conjunto de dados forma uma matriz com dimensões 12 x 916 x 125. A partir das abordagens de desdobramento NM e WKFW as matrizes puderam ser reorganizadas, em que à priori, fez-se uma comparação para análise pós-batelada para ambas as abordagens. Também foi realizado o modo de análise em tempo real apenas para a abordagem WKFW (NOMIKOS; MACGREGOR, 1995; WOLD et al., 1998).

Antes do arranjo tridimensional ser efetuado, e utilizando o software MATLAB®, as matrizes bidimensionais, cujo número de linhas é dado por $(N \times K)$ e J colunas foram submetidas a diferentes pré-processamentos: 1ª derivada com filtro de Savitzky-Golay (janela de 5 pontos, ajustado a um polinômio de 2º grau), MSC e SNV.

Após escolha do melhor pré-processamento, as matrizes de dados de cada batelada foram concatenadas para o desenvolvimento das cartas de controle multivariadas, que foram construídas a partir dos dados das bateladas do conjunto de calibração. Posteriormente, esses modelos foram empregados para a identificação de falhas nas bateladas que compõem a etapa de predição. O desempenho das cartas de controle foi avaliado utilizando as estatísticas distância ao modelo DModX e T^2 Hotelling (FERRER, 2007).

3.5 SOFTWARES

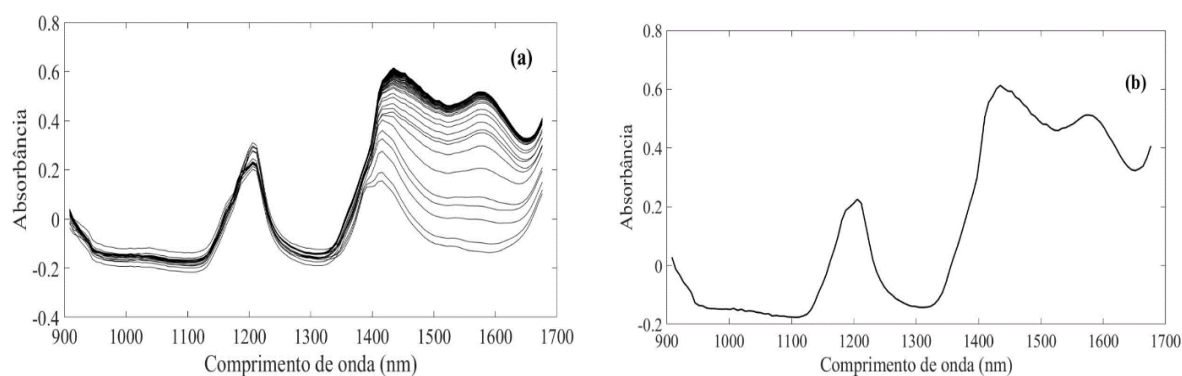
Todos os cálculos envolvendo pré-processamento de dados, seleção de amostras e variáveis e construção dos modelos de calibração multivariada foram realizados utilizando rotinas próprias implementadas no software MATLAB® (*Math Works, Inc., Natick, MA*) e o pacote de ferramentas PLS-Toolbox versão 8.6.2, em ambiente MATLAB. O planejamento de misturas foi realizado no software Minitab® versão teste. O desenvolvimento dos métodos de controle estatístico multivariado de processo foi realizado no software SIMCA versão 13.0.

4 RESULTADOS E DISCUSSÃO

Nas reações de transesterificação, durante os instantes iniciais da mistura da matéria-graxa, composta majoritariamente por triacilgliceróis, e um álcool de cadeia curta, há formação de uma emulsão no meio reacional, afetando diretamente a transferência de massa no meio (GÜNAY; TÜRKER; TAPAN, 2019). Como consequência dessa heterogeneidade observa-se também o espalhamento da radiação, que gera desvios na linha de base dos espectros NIR obtidos durante a reação em batelada. A Figura 13.a ilustra o espalhamento observado nos espectros adquiridos durante 60 minutos de monitoramento da batelada número 4, produzida em condições normais de operação. Estes são espectros brutos, típicos da reação de transesterificação do óleo de algodão por via etílica.

Como a espectroscopia NIR fundamenta-se na vibração molecular, os espectros NIR são afetados pela variação da temperatura. Para minimizar o efeito da variação de temperatura provocado durante a mistura dos reagentes no reator, todos os reagentes foram incorporados à mistura reacional à mesma temperatura, aproximadamente 50°C.

Figura 13 – Espectros NIR obtidos durante uma batelada típica de produção de biodiesel pela rota etílica (a) espectros brutos (ordem de aquisição espectral de baixo para cima). (b) último espectro adquirido da batelada.



Fonte: O autor (2020).

Na Figura 13.b, observa-se o espectro da mistura reacional de uma batelada produzida em condições normais de operação, após 60 minutos de operação. Nos espectros da Figura 13.b é possível observar bandas de absorção nas regiões de segundo sobretom (1120-1250 nm), bem como na região do primeiro sobretom das combinações (1300-1400 nm) devido ao estiramento das ligações C–H. Também destaca-se as regiões de terceiro e quarto sobretom da ligação C=O em torno de (1100-1200nm e 1400-1450 nm) respectivamente. Também pode-se considerar o

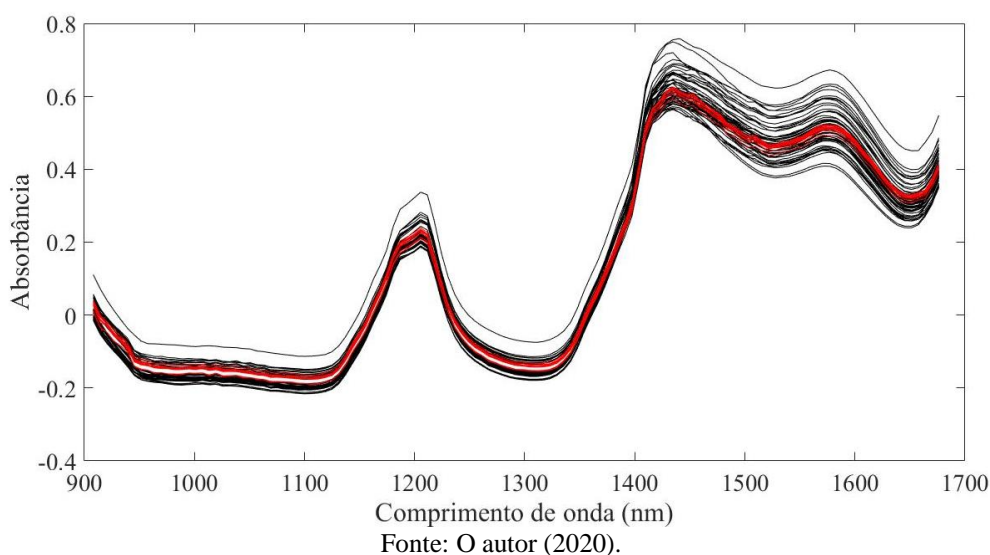
primeiro sobretudo correspondente à ligação O–H (1450-1600 nm), cujas bandas aparecem bem pronunciadas, visto que há excesso de etanol e glicerina (WORKMAN; WEYER, 2012).

4.1 CALIBRAÇÃO MULTIVARIADA

4.1.1 Pré-processamento dos espectros NIR

A Figura 14 apresenta os espectros brutos de 65 amostras, sendo 54 amostras sintéticas e 11 amostras da mistura reacional das bateladas de 1 a 11.

Figura 14 – Espectros brutos de 54 misturas sintéticas (espectros em preto), e espectro médio dos 10 últimos espectros das bateladas de 1 a 11 (em vermelho) para a reação de transesterificação do óleo de algodão por rota etílica das bateladas.

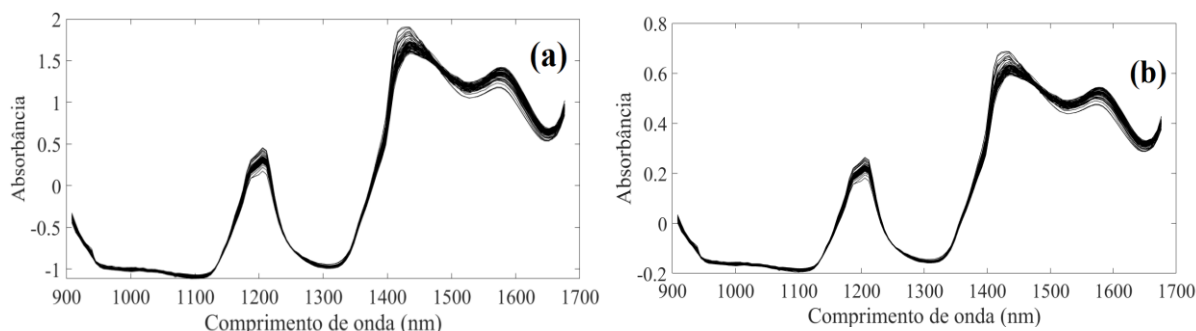


É possível observar alta semelhança entre o perfil espectral acima (Figura 14) com o espectro final de uma típica reação em batelada de transesterificação do óleo de algodão por rota etílica apresentado na Figura 13.b. Esta similaridade demonstra que o sistema de aquisição espectral das amostras sintéticas apresenta representatividade satisfatória quando comparada as condições de produção das bateladas no reator. Fatores como as diferentes temperaturas entre as misturas sintéticas (50 e 60° C) e a presença de amostras com concentração de óleo e/ou glicerina mais altas, tornado essas misturas mais turvas, contribuem para o espalhamento da radiação neste sistema. Este efeito também foi observado por Sales et al. (2019).

Para corrigir o efeito da variação de linha de base presente na matriz espectral representada na Figura 14, foram testados os pré-processamentos: SNV, MSC e 1ª derivada

com filtro de Savitzky Golay para um polinômio de 2º grau com janelas de 5, 7 e 9 pontos. Ambos os pré-processamentos SNV e MSC foram capazes de corrigir os efeitos de linha de base de forma satisfatória, apontando resultados semelhantes conforme demonstra a Figura 15.

Figura 15 – Espectros pré-processados com (a) SNV e (b) MSC.



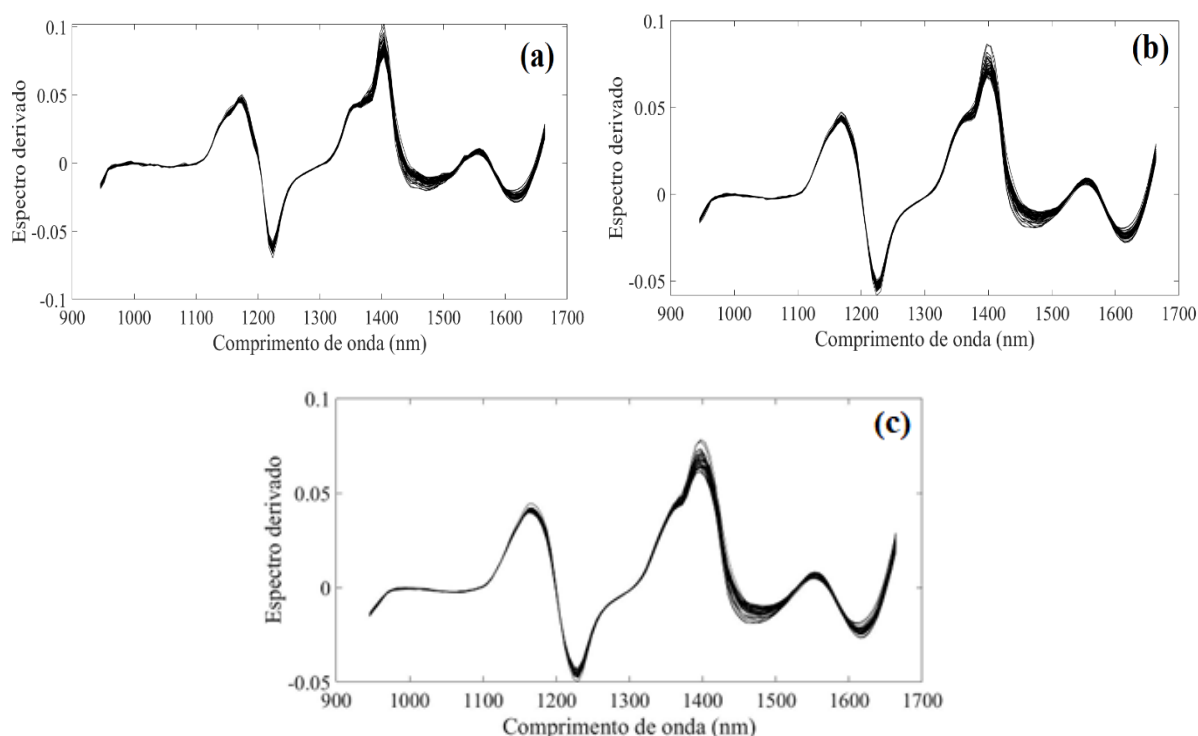
Fonte: O autor (2020).

A aplicação da primeira derivada com filtro de Savitzky-Golay utilizando janelas de 5,7 e 9 pontos foi capaz de eliminar o efeito da variação de linha de base, entretanto, apenas a derivada com janelas de 5 e 7 pontos foi capaz de preservar as principais bandas de absorção, (Figura 16). Observou-se que a utilização de janelas a partir de 9 pontos promove suavização excessiva em bandas presentes na região de (1100-1200 nm), cujo intervalo apresenta bandas importantes para a identificação da molécula de éster como visto no item 4, anteriormente. Dessa forma, essa suavização poderia mascarar uma fonte de informação relevante para o problema estudado.

Vale salientar que o método da derivação espectral gera um efeito de sobreposição cruzada nas bordas dos espectros derivados. Logo, para este pré-processamento, optou-se por cortar os intervalos dos respectivos comprimentos de onda (908-940 nm e 1670-1676 nm). Como esses intervalos correspondem a faixas muito pequenas, não houve perdas de informação química relevante para os componentes da mistura reacional, visto que também coincidem com a faixa de limite de aquisição do equipamento, apresentando uma baixa relação sinal ruído.

O pré-processamento mais adequado aos dados foi escolhido levando-se em consideração, além da discussão acima, o desempenho da capacidade preditiva dos modelos desenvolvidos, bem como a melhor correção das diferenças entre os espectros das amostras sintéticas e dos espectros provenientes das bateladas.

Figura 16 – Espectros pré-processados com 1ª derivada com filtro de Savitzky-Golay e ajuste de polinômio de 2º grau utilizando (a) uma janela de 5 pontos, (b) uma janela de 7 pontos, (c) uma janela de 9 pontos.



Fonte: O autor (2020).

4.1.2 Construção do modelo de calibração multivariada para o teor de etil-ésteres (EE)

Para construção do modelo de predição do teor de EE, as 65 amostras foram distribuídas em um conjunto de calibração e outro de predição. Essa seleção foi feita através da utilização do algoritmo SPXY que selecionou 40 amostras de calibração e 25 de predição. A seleção de amostras foi realizada conforme recomendado pela norma ASTM E1655-05, que dentre as recomendações previstas, estão as regras para seleção das amostras dos conjuntos de calibração e validação para construção de modelos de calibração multivariada a partir da espectroscopia NIR. Posteriormente foram construídos modelos de calibração a partir de abordagens descritas a seguir.

4.1.3 Construção de modelos PLS utilizando a faixa espectral completa

Os modelos de regressão PLS para determinação do teor de ésteres etílicos construídos a partir da utilização de diferentes pré-processamentos da matriz espectral, considerando a faixa de varredura completa do MicroNIR (908-1676 nm), com exceção do corte das bordas para os espectros derivados, são mostrados na Tabela 4.

Tabela 4 – Figuras de mérito para os modelos PLS para estimar o teor de EE construídos utilizando a faixa espectral completa.

Pré-proc.	Filtro de suav.	VL	R²_{cal}	RMSECV (%m)	bias_{cv}	R²_{pred}	RMSEP (%m)	bias_{pred}
SNV	-	5	0,97	2,24	-0,06	0,97	1,60	0,13
MSC	-	5	0,97	2,26	-0,05	0,97	1,62	0,25
(1ª Derivada)	Janela de 5 pts	5	0,97	2,37	-0,31	0,93	2,66	-0,66
Savitzky-Golay	Janela de 7 pts	7	0,98	1,43	-0,09	0,95	2,16	-0,21
	Janela de 9 pts	7	0,98	1,59	-0,12	0,95	2,22	-0,24

VL: variáveis latentes; pts = pontos; R²_{cal} e R²_{pred}: coeficientes de determinação para as etapas de calibração e predição, respectivamente; Pré-proc. = pré-processamento; suav. = suavização.

Fonte: O autor (2020).

De acordo com os dados da Tabela 4, pode-se observar que os modelos construídos após correção dos desvios espectrais, apresentaram valores de RMSECV entre 1,43 e 2,37%, e de RMSEP na faixa de 1,60 a 2,66%, sendo portanto menores que a reprodutibilidade do método de referência de 3,10% m/m. (EUROPEAN STANDARD, 2003). Os modelos construídos utilizando os pré-processamentos SNV e MSC apresentaram resultados muito semelhantes, desde a complexidade, visto que o número de variáveis latentes foi o mesmo para ambos, assim como os valores de RMSEP e RMSECV. Por outro lado, entre os modelos construídos utilizando a 1ª derivada como estratégia de pré-processamento, observou-se que o modelo que usou uma janela de 5 pontos foi o que apresentou menor complexidade. Visto que 5 variáveis latentes foram necessárias para construção do modelo, em detrimento de 7 LV como apresentado pelos modelos que usaram janelas de 7 e 9 pontos na derivada, respectivamente.

Em relação aos modelos construídos a partir de espectros derivados, embora não tenham apresentado diferença significativa para os valores de RMSEP entre si, segundo o teste F para 95% de confiança, apresentaram desempenho inferior em relação aos modelos construídos com os pré-processamentos MSC e SNV, de acordo com um teste F para os valores de RMSEP, com 95% de confiança. Logo, como os valores são muito próximos e não apresentam diferença significativa, os dois pré-processamentos apresentam desempenho equivalentes, para os modelos desenvolvidos com a faixa espectral completa. Não foram observados erros

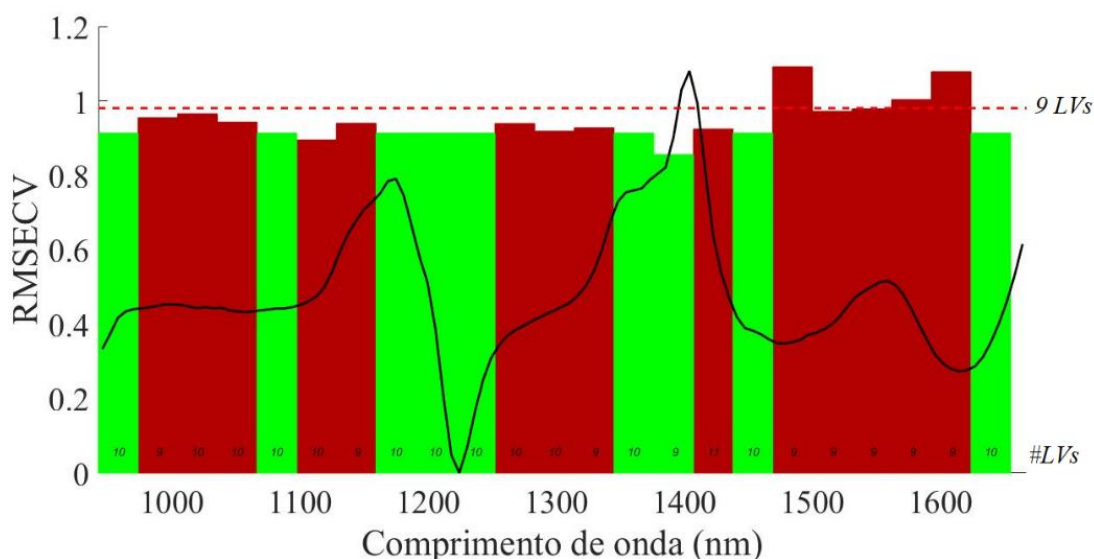
sistemáticos significativos (*bias*) para os modelos avaliados, de acordo com os resultados do teste *t-student* com significância de 0,05.

4.1.4 Modelos PLS construídos utilizando o algoritmo de seleção de variáveis *i*PLS

Na utilização do algoritmo *i*PLS, os espectros foram divididos em 23 intervalos iguais de 5 comprimentos de onda cada, conforme mostrado na Figura 17. Cada barra representa os valores de RMSECV para os modelos PLS locais construídos para cada intervalo, com o número de LVs necessárias. Esses resultados são balizados pelo RMSECV do modelo construído considerando toda a faixa espectral, representado pela linha tracejada horizontal.

A Figura 17 apresenta os resultados obtidos com a utilização do algoritmo *i*PLS nos espectros pré-processados com 1ª derivada com filtro de Savitzky-Golay e janela de 5 pontos com ajuste de polinômio de segundo grau. As barras destacadas em verde correspondem aos intervalos cujos modelos PLS apresentaram melhor desempenho. Observa-se que estas faixas além de apresentarem valores de RMSECV abaixo do erro esperado para o modelo construído com a faixa inteira (linha tracejada), também estão relacionados com os intervalos de comprimentos de onda em que aparecem bandas com informação química relevante. As quais são características de etil-ésteres (composto de interesse), como foi discutido no item 4.

Figura 17 – Faixas espectrais selecionadas pelo algoritmo *i*PLS (barras verdes) para modelos de EE.



As figuras de mérito para o modelo PLS construído para estimar o teor de EE utilizando o conjunto dos intervalos selecionados mostrados na Figura 17 (em verde), utilizando a 1ª

derivada com filtro de Savitzky-Golay e janela de 5 pontos com ajuste de polinômio de segundo grau, estão apresentadas na Tabela 5. Nessas condições o modelo apresentou valores de $R^2_{\text{cal}} = 0,98$; $\text{RMSECV} = 1,64\%$; $R^2_{\text{pred}} = 0,98$; $\text{RMSEP} = 1,46\%$ e $\text{bias} = 0,08$, utilizando 5 variáveis latentes. Para esta estratégia de pré-processamento, o modelo acima, obtido a partir da seleção de variáveis por meio do algoritmo *i*PLS apresentou figuras de mérito com valores significativamente menores do que o seu respectivo modelo construído com a faixa espectral completa, mostrado na Tabela 4.

Na Tabela 5, são apresentados de forma mais detalhada, os parâmetros referentes aos modelos PLS para determinação do teor de EE construídos a partir do uso do algoritmo *i*PLS como abordagem de seleção de variáveis, para as diferentes estratégias de pré-processamento adotadas, respectivamente às que foram apresentadas na Tabela 4.

Tabela 5 – Parâmetros dos modelos PLS para o teor de EE construídos com os intervalos espectrais selecionados pelo algoritmo *i*PLS.

Pré-proc.	Filtro suav	Faixa*	VL	R^2_{cal}	$\text{RMSECV} (\%_{\text{m}})$	bias_{cv}	R^2_{pred}	$\text{RMSEP} (\%_{\text{m}})$	$\text{bias}_{\text{pred}}$
SNV/ <i>i</i> PLS	-	[11:45 51:55 76:80 86:90 116:125]	5	0,98	1,93	-0,20	0,93	2,63	-0,43
MSC/ <i>i</i> PLS	-	[16:25 31:45 66:70 76:80 86:90 111:115 121:125]	7	0,99	1,58	-0,11	0,95	2,28	-0,20
(1ª Deriv.) Savitzky- Golay e <i>i</i> PLS	Jan. 5 pts	[1:5 21:25 36:50 66:75 81:85 111:115]	5	0,98	1,64	-0,20	0,98	1,46	0,08
	Jan. 7 pts	[41:50 56:70 76:85 91:95 111:115]	5	0,98	1,83	-0,09	0,94	2,37	-0,30
	Jan. 9 pts	[16:20 41:45 51:65 76:80 111:115]	6	0,97	1,80	0,15	0,95	2,11	0,01

*Os intervalos correspondem as posições dos comprimentos de onda selecionadas pelo algoritmo *i*PLS na matriz de comprimentos de onda (vide Apêndice B); pts = pontos.

Fonte: O autor (2020).

A partir da avaliação dos parâmetros resumidos na Tabela 5, os modelos pré-processados com SNV e MSC construídos a partir da abordagem de seleção de variáveis *i*PLS

demonstram de forma clara uma piora no desempenho em termos de valores absolutos, quando comparados ao uso da faixa espectral completa (Tabela 4). No entanto, há diferença significativa, segundo um teste F para 95% de confiança, apenas entre os valores de RMSEP para o modelo construído utilizando o pré-processamento SNV. Este modelo, portanto, apresentou melhor desempenho quando construído com a faixa espectral completa.

Já para o modelo obtido utilizando MSC como pré-processamento não houve diferença significativa para os valores de RMSEP entre a utilização da faixa completa ou a selecionada pelo *i*PLS, entretanto, com *i*PLS a modelagem exigiu mais variáveis latentes, tornando-o mais complexo. Ambos os modelos construídos com *i*PLS não apresentaram erros sistemáticos significativos (*bias*), de acordo com um teste *t*.

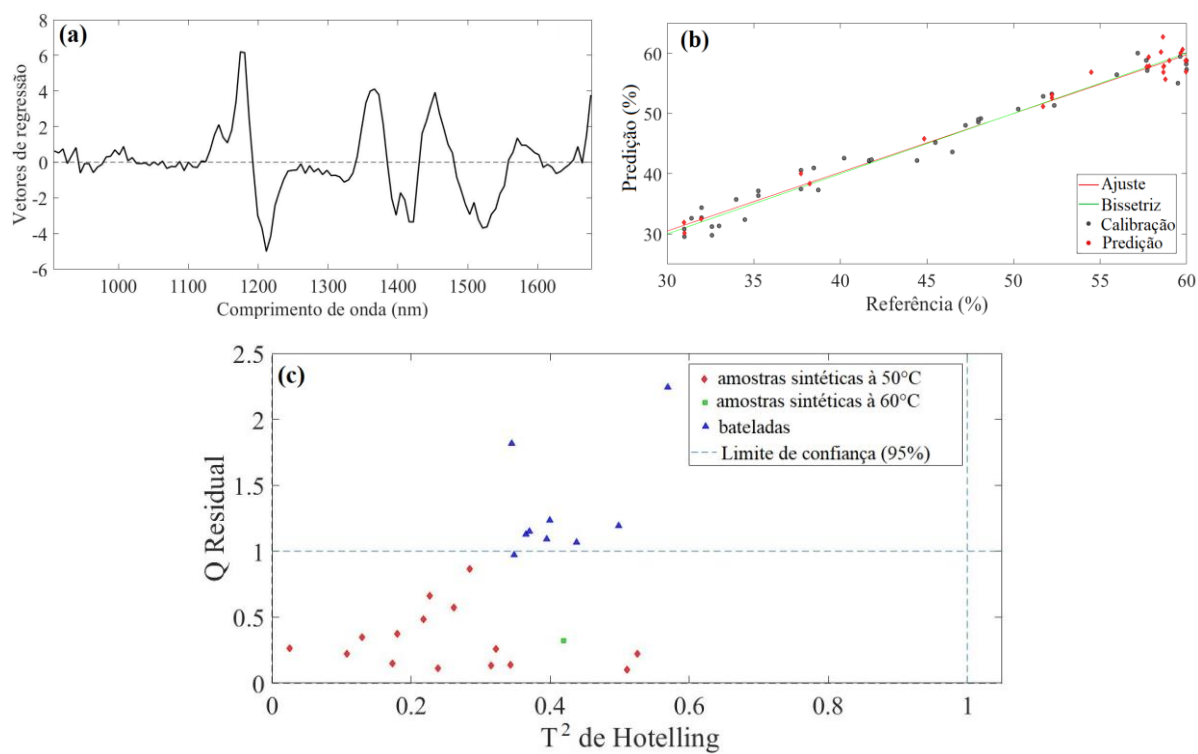
Em relação aos modelos construídos a partir do uso da 1ª derivada como pré-processamento, destaca-se aquele com janela de 5 pontos, que apresentou uma redução considerável nos valores absolutos de RMSECV e RMSEP. De acordo com um teste F, há diferença significativa, para 95% de confiança, comparando-se esse valor de RMSEP com os valores de RMSEP dos modelos utilizando janelas de 7 e 9 pontos respectivamente. Dessa forma, o modelo construído com 1ª derivada com filtro de Savitzky-Golay com janela de 5 pontos e seleção de variáveis por *i*PLS obteve melhor desempenho em relação aos demais modelos construídos com espectros derivados, quer utilizando a faixa espectral completa ou seleção de variáveis *i*PLS.

A partir da comparação das figuras de mérito, observou-se que para os modelos obtidos até o momento, aquele construído a partir da 1ª derivada com janela de 5 pontos e *i*PLS, e os construídos a partir da faixa espectral completa com os dados pré-processados por SNV e MSC, apresentaram desempenho semelhantes em termos de RMSEP para 95% de confiança, conforme teste F executado. Assim como, ambos não apresentam erros sistemáticos significativos (*bias*).

Nesse sentido, a fim de definir a abordagem mais adequada para o pré-processamento dos dados, avaliou-se de forma conjunta os gráficos de influência, o gráfico de valores preditos versus observados e os coeficientes de regressão (Figuras 18 – 20). Desta forma, é possível avaliar a eficiência de cada abordagem em relação às variáveis (comprimentos de onda) que deve coincidir com a informação química presente na matriz espectral. Bem como com as amostras, que por serem obtidas em sistemas experimentais diferentes, naturalmente apresentam variabilidade intrínseca a cada modo de aquisição. Entretanto, deseja-se observar comportamento semelhante entre elas, visto que para construção do modelo, todas devem

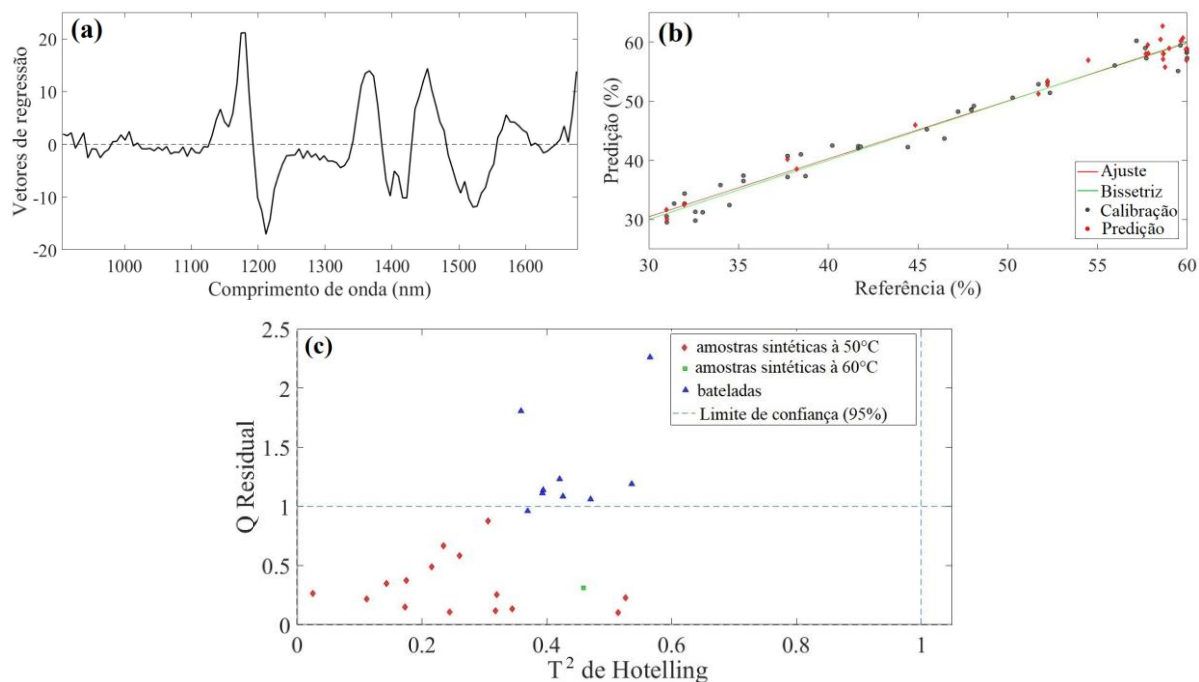
pertencer a um mesmo grupo amostral. Assim, espera-se que as estratégias de pré-processamento forneçam uma minimização das diferenças entre amostras sintéticas e bateladas.

Figura 18 – Parâmetros do modelo pré-processado com SNV para determinação do teor de EE **(a)** vetores de regressão; **(b)** valores preditos versus observados; **(c)** gráfico de influências.



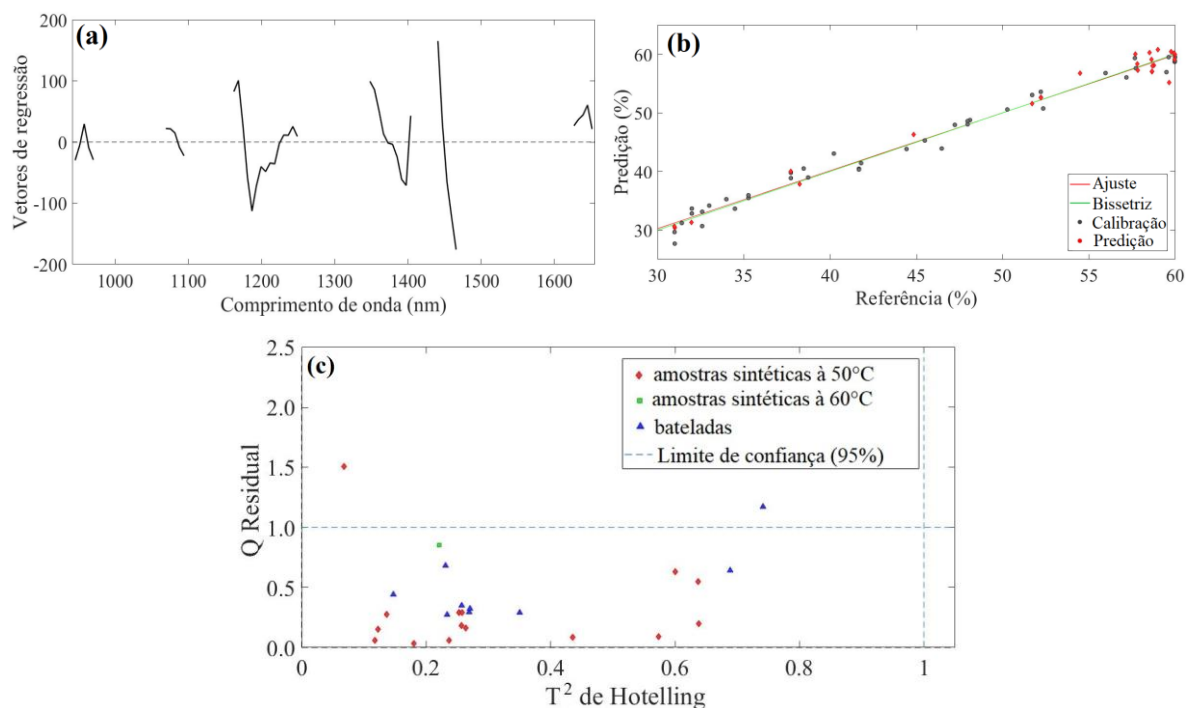
Fonte: O autor (2020).

Figura 19 – Parâmetros do modelo pré-processado com MSC para determinação do teor de EE (a) vetores de regressão; (b) valores preditos versus observados; (c) gráfico de influências.



Fonte: O autor (2020).

Figura 20 – Parâmetros do modelo pré-processado com 1ª derivada e janela de 5 pontos/iPLS para determinação do teor de EE (a) vetores de regressão; (b) valores preditos versus observados; (c) gráfico de influências.



Fonte: O autor (2020).

De acordo com as Figuras (18.a, 19.a e 20.a), é possível observar que embora os modelos sejam construídos por abordagens distintas de pré-processamentos, os respectivos vetores de regressão demonstraram, em geral, importância para as mesmas regiões do espectro. Entre elas, podemos destacar as regiões de terceiro e quarto sobreton da ligação C=O referente ao éster, composto de interesse que se forma ao longo da reação em torno de 1200 nm e 1450 nm respectivamente. Outras regiões relevantes nos gráficos dos vetores de regressão foram as de segundo sobreton (1120-1250 nm), bem como a região do primeiro sobreton das combinações (1300-1400 nm) do estiramento das ligações C-H. Estas regiões apresentaram intensidades relevantes em todos os respectivos gráficos, visto que os compostos presentes na reação de transesterificação são majoritariamente formados por ligações deste tipo. Devido à elevada razão álcool:óleo utilizada na reação, bem como pela presença de glicerina como subproduto, também é possível perceber a presença de uma região intensa em torno de 1450 nm nos gráficos dos vetores de regressão, referente ao primeiro sobreton da ligação O-H.

Por outro lado, o modelo obtido com a derivada/*i*PLS foi o único capaz de minimizar as diferenças entre os espectros das bateladas e os espectros das misturas sintéticas. Um reflexo disso pode ser observado nos gráficos de influência das Figuras (18.c, 19.c e 20.c). Por esse motivo o modelo com derivada foi considerado o modelo mais robusto, ou seja, que sofre menos com a diferença nas características das amostras. Além de demonstrar a manutenção da estrutura de correlação entre os dados devido à proximidade entre os scores das amostras sintéticas e das bateladas.

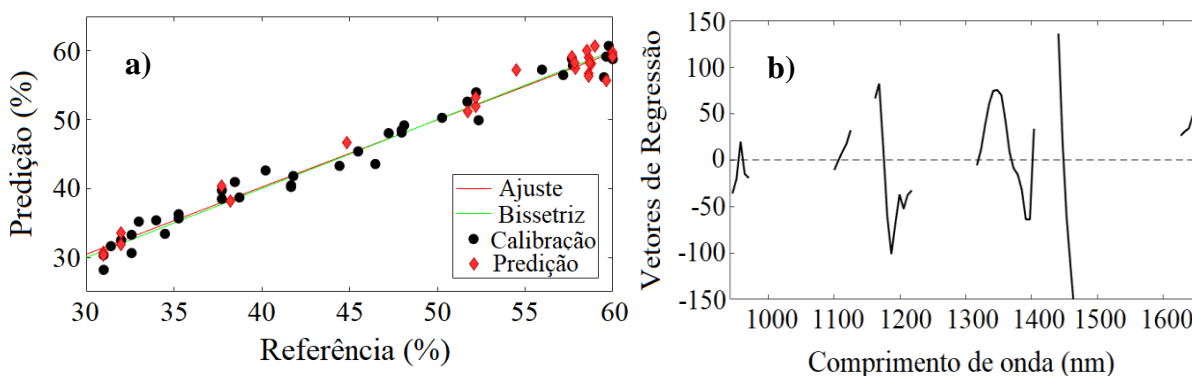
Tal abordagem foi capaz de demonstrar que há expressiva reprodutibilidade entre as respostas espectrais obtidas a partir da utilização de sistemas experimentais distintos (misturas sintéticas vs. bateladas). Além de obter melhores resultados em termos de valores absolutos para os erros de validação e predição para o teor de EE.

A fim de melhorar a representatividade do modelo que obteve o melhor desempenho analítico em relação às condições de temperatura utilizadas, substituiu-se 2 amostras sintéticas analisadas à 50°C do conjunto de predição por 2 amostras sintéticas analisadas à 60°C presentes no conjunto de calibração. Trazendo assim, ainda mais robustez para o modelo PLS quanto ao efeito da variação de temperatura.

Aplicando a mesma estratégia de pré-processamento associada a seleção de variáveis por *i*PLS, obteve-se um novo modelo PLS construído na faixa de 31 – 60% para o teor de EE que apresentou valores de $R^2_{\text{cal}} = 0,98$; $\text{RMSECV} = 1,63\%$; $R^2_{\text{pred}} = 0,98$; e $\text{RMSEP} = 1,51\%$ e $\text{bias} = 0,09$, utilizando 5 variáveis latentes. De acordo com as figuras de mérito, observa-se que este modelo também demonstrou resultados abaixo da reprodutibilidade do método de

referência, apresentando ainda desempenho um pouco melhor quanto à predição do teor de ésteres (EUROPEAN STANDARD, 2003). O modelo e o respectivo gráfico dos vetores de regressão estão descritos na Figura 21. É possível observar uma distribuição satisfatória dos pontos em torno da bissetriz (Figura 21.a), e o gráfico dos vetores de regressão (Figura 21.b) também destaca regiões de relevância química, coerente com os compostos presentes na reação conforme discutido anteriormente.

Figura 21 – Modelo PLS/iPLS (1ª derivada de Savitzky-Golay com janela de 5 pontos) construído na faixa de concentração de 31 – 60% para teor de ésteres etílicos (a), e gráfico dos vetores de regressão (b).



Fonte: O autor (2020).

Em comparação com a literatura, este trabalho mostrou melhor capacidade preditiva para o teor de EE que os resultados reportados por Sales et al. (2019). O modelo dos referidos autores apresentou RMSEP igual a 2,15% para a predição do teor de etil-ésteres em mistura quaternária, quando monitoraram a reação de transesterificação etílica do óleo de algodão, em modo contínuo e *on-line* numa coluna de destilação reativa, com equipamento NIR portátil semelhante ao utilizado neste trabalho. Também é válido destacar que o modelo PLS do presente trabalho apresentou valor de RMSEP menor, associado a uma faixa maior de concentração de EE que o obtido pelo referido autor, cuja faixa do modelo foi de 0 – 23% m/m,

Richard et al. (2013) realizaram o monitoramento da reação de etanolise do óleo de girassol via espectroscopia NIR. Os autores apresentaram dois modelos de predição do teor de EE na faixa de concentração de 0 – 100% para mistura binária (óleo-éster), que apresentaram valores de RMSEP iguais a 4,10% e 3,52%, quando os dados espectrais foram adquiridos por sondas de reflectância e transflectância, respectivamente. O modelo apresentado pelo presente trabalho demonstrou RMSEP menor, associado a uma faixa de concentração de éster mais restrita, e a utilização de um equipamento de menor resolução. Por outro lado, os resultados reportados pelos autores foram resultantes da aquisição de dados espectroscópicos

obtidos por um equipamento de bancada, no modo *on-line* e utilizando um microrreator em processo contínuo.

No trabalho reportado por Pinzi et al. (2012), que realizaram o monitoramento *off-line* da reação de transesterificação por rota metílica de 3 tipos de óleos vegetais diferentes: girassol, milho e de bagaço de azeitona, a partir da espectroscopia NIR utilizando um equipamento de bancada. Embora tenha obtido um modelo com apenas 3 variáveis latentes, o mesmo apresentou $RMSEP = 2,55\%$ para o teor de metil-éster, na faixa de concentração 86,5 – 97% m/m para mistura binária (óleo-éster). O referido autor agregou maior variabilidade ao respectivo modelo devido à utilização de diferentes óleos vegetais no processo reacional. No entanto, obteve menor capacidade preditiva, associado a uma menor faixa de concentração de éster, que o modelo PLS proposto pelo presente trabalho.

Quando comparado ao resultado obtido por Lima, (2014), que monitorou a reação de transesterificação do óleo de soja por rota metílica, para a faixa de concentração de 64,4 – 97,1% m/m para teor de ésteres metílicos (ME), em mistura binária, usando espectroscopia NIR com equipamento de bancada, no modo *in-line* e em batelada. Observa-se desempenho semelhante quanto às figuras de mérito, com $RMSEP$ de 1,60%, porém o modelo construído no presente trabalho exigiu o uso de uma variável latente a mais.

Killner, Rohwedder e Pasquini, (2011) que realizaram o monitoramento *on-line* da reação de transesterificação do óleo de soja pela rota metílica e em batelada, obtiveram modelos com valores de $RMSEP$ iguais a 0,74% e 1,27% de conversão para duas bateladas distintas na faixa de concentração de metil-ésteres entre 60 – 100% para a mistura binária (óleo-éster).

Vale salientar que além dos resultados descritos pelo presente trabalho terem sido obtidos no modo *on-line*, e a partir do uso de um equipamento portátil, que apresenta baixa resolução quando comparado a um equipamento de bancada, também foram utilizadas amostras coletadas em diferentes bateladas e amostras sintéticas. A utilização de amostras sintéticas para calibração simplifica bastante o processo, reduzindo a necessidade de produzir muitas bateladas em diferentes condições para aumentar a variabilidade da faixa de calibração, bem como redução do número de análises cromatográficas.

Para reforçar a confiança sobre os valores preditos pelo modelo, na Tabela 6 estão dispostos os valores de referência do teor de EE para as 9 amostras das bateladas presentes no conjunto de predição obtidos por cromatografia gasosa, conforme a norma europeia EN-14103, assim como seus respectivos valores preditos pelo modelo, em que os mesmos foram convertidos de (% m/m) de EE na mistura quaternária para (% m/m) na mistura binária

(éster/óleo), pela via estequiométrica. Tal conversão fez-se necessária porque os valores de referência foram obtidos nessa última condição.

Tabela 6 – Resultados para a conversão de etil-ésteres estimados pelo modelo PLS e o obtido por análise cromatográfica para amostras do conjunto de predição.

ID batelada	Teor de EE – GC* (%m/m)	Teor de EE – modelo PLS	Erro relativo
		(%m/m)	(%m/m)
1	91,83	92,35	0,57
2	93,36	87,46	6,32
3	91,93	91,43	0,55
4	91,87	89,03	3,09
5	90,66	90,04	0,69
6	90,42	92,68	2,50
7	90,58	91,28	0,78
8	91,67	93,95	2,49
9	91,87	88,31	3,87

*GC = Cromatografia gasosa.
Fonte: O autor (2020).

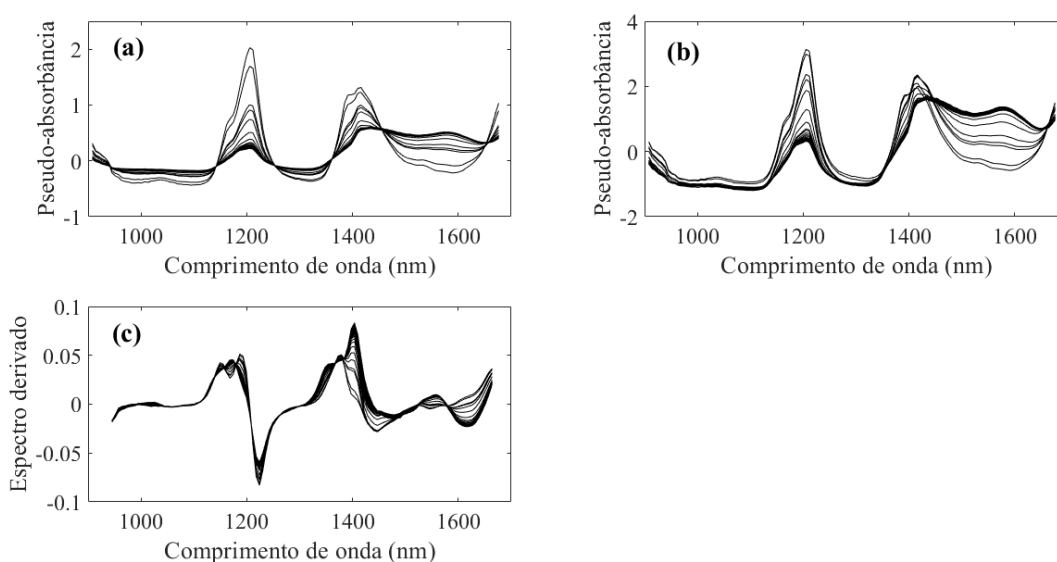
Pode-se observar que com exceção do resultado da batelada 2, todas as outras bateladas tiveram seus valores de conversão de EE estimados pelo modelo multivariado com erro relativo menor que 5%, sendo, portanto, comparáveis ao resultado obtido pelo método de referência. O erro relativo percentual foi calculado através do módulo da diferença entre o valor de referência e o predito pelo modelo, dividido pelo valor de referência, sendo esse resultado multiplicado por 100.

4.2 CONTROLE ESTATÍSTICO MULTIVARIADO DO PROCESSO DE TRANSESTERIFICAÇÃO EM BATELADA

Os espectros brutos de uma típica batelada em condições normais de operação, obtidos a partir do equipamento portátil MicroNIR, no modo *on-line* e sob aquisição contínua de espectros ao longo do tempo foram apresentados na Figura 13.a. Como observado nesta Figura, e discutido anteriormente, o efeito do espalhamento da luz provocado, entre outros fatores, pela turbidez nos momentos iniciais da reação gera uma grande variação de linha de base nos espectros. Faz-se, então, necessária a correção desses desvios por meio de técnicas de pré-

processamento, as quais tem por objetivo remover os efeitos físicos para que a informação química contida nos espectros possa ser interpretada adequadamente. Foram avaliadas como estratégias de correção as seguintes técnicas: 1ª derivada com filtro de suavização Savitzky-Golay e ajuste de polinômio de segundo grau associado a uma janela de 5 pontos, variação normal padrão (SNV), e correção do espalhamento multiplicativo (MSC). Os espectros pré-processados estão descritos na Figura 22 e os que utilizaram derivada tiveram as bordas cortadas devido ao efeito de sobreposição cruzada.

Figura 22 – Espectros de uma batelada comum em condições normais de operação pré-processados com (a) MSC (b) SNV, (c) primeira derivada Savitzky-Golay.



Fonte: O autor (2020).

De acordo com a Figura 22, observa-se que tanto o MSC quanto o SNV não foram capazes de corrigir os desvios de linha de base adequadamente. Por outro lado, os espectros derivados apresentaram redução significativa dos desvios originais, demonstrando maior eficiência na correção. A técnica da 1ª derivada com filtro de Savitzky-Golay foi então selecionada para desenvolvimento dos gráficos de controle apresentados a seguir.

4.2.1 Construção dos gráficos de controle multivariados de processo

A construção dos gráficos de controle multivariados foi realizada a partir da organização e avaliação dos dados espectrais conforme as abordagens NM e WKFH. Os parâmetros dos modelos multivariados construídos foram resumidos na Tabela 7, tanto para as avaliações em tempo real quanto para a análise pós-batelada.

Tabela 7 – Resultados para as avaliações dos modelos multivariados em tempo real e pós-batelada.

Abordagem	Modo de avaliação	Pré-processamento	Número de componentes	R² X (%)	Q² (%)
WKFH	Tempo real	Primeira derivada	5	96,4	77,3
WKFH	Pós-batelada		4	95,1	69,0
NM	Pós-batelada		2	70,0	36,0

Fonte: O autor (2020).

R²X representa o percentual de variância explicada pelo modelo, o qual está relacionado com a qualidade do ajuste aos dados experimentais. E o valor de Q² representa o percentual de variação prevista pelo modelo de acordo com a validação cruzada, e indica quão bem o modelo prevê dados novos (AGUADO et al., 2007).

Os modelos multivariados descritos neste trabalho foram construídos considerando que as bateladas do conjunto de calibração foram desenvolvidas em condições normais de operação, e sob controle. Esses modelos foram analisados previamente para identificação de possíveis eventos anômalos a partir das estatísticas DModX e T² de Hotelling. Os gráficos de controle para o conjunto de bateladas de calibração obtidas de acordo com as abordagens descritas na Tabela 7 podem ser observados no Apêndice C.

4.2.1.1 Avaliação da detecção de falhas via análise em tempo real pela abordagem WKFH

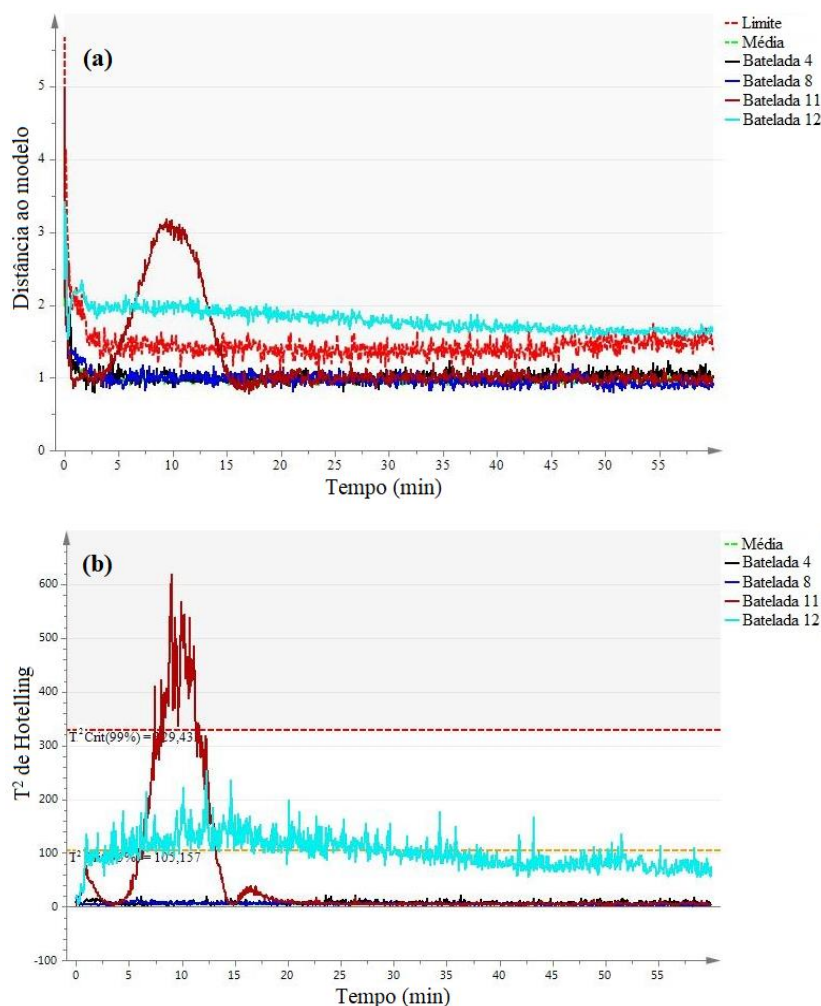
Para a construção de cartas de controle levando-se em consideração a abordagem WKFH, os dados das bateladas foram rearranjados e pré-processados com 1ª derivada. Em seguida, um modelo PLS foi desenvolvido considerando os dados gerados pelo monitoramento das bateladas do conjunto de calibração e o vetor com informação do tempo das bateladas. Este modelo PLS foi gerado utilizando-se cinco variáveis latentes, obtendo-se 96,4% da variação dos dados do conjunto de calibração, além de apresentar capacidade preditiva de 77,3% em relação ao conjunto de predição (Tabela 7).

O conjunto de calibração, composto por oito bateladas obtidas sob condições normais de operação, apresentou bateladas NOC dentro do limite de controle, tanto para as cartas de Distância ao modelo (DModX) quanto para de T² de Hotelling (Apêndice C), como esperado. A seguir, serão apresentadas as cartas de controle do conjunto de predição, as quais são compostas por duas bateladas NOC selecionadas aleatoriamente (4 e 8), uma batelada com

alteração simulada durante seu processo de monitoramento (11) e uma batelada com concentração de catalisador diferente (12).

Na batelada 11, o comando da alteração da temperatura ocorreu cerca de 2 minutos após a alimentação no reator. Foi então efetuado um aumento de 50°C para 55°C na temperatura, mantendo-se nessa temperatura por um período de 5 minutos, e posterior redução para temperatura normal de operação (50°C). Já na batelada 12, trabalhou-se com a concentração de catalisador igual 0,9%, valor superior à concentração em NOC (0,5%). As cartas de controle de DModX e de T^2 de Hotelling para a abordagem WKFH e análise em tempo real estão apresentados na Figura 23.

Figura 23 – Cartas de controle de (a) DModX, e (b) T^2 de Hotelling, para as bateladas 4, 8, 11 e 12 submetidas a perturbações na temperatura e concentração do catalisador (NaOH).



Fonte: O autor (2020).

Observa-se em ambas as cartas que as bateladas NOC (4 e 8) apresentaram-se dentro do limite de controle (95% de confiança), demonstrando coerência com sua origem. Também é

possível observar que em ambos os gráficos de controle, a variação de temperatura na batelada 11 (em vermelho), embora sutil, foi percebida de modo intenso pelo MicroNIR, mesmo com a aquisição espectral no modo *on-line*, em que a captura de informação dá-se a uma certa distância do vaso reacional, através de uma linha de amostragem (Figura 10). Esse fenômeno pode ser explicado devido à sensibilidade da espectroscopia NIR a variações de temperatura, segundo Workman e Weyer (2012). Essa alteração já foi suficiente para a batelada sair de controle temporariamente nas duas cartas. Destaca-se que, para a batelada 11, tanto a distância ao modelo quanto a estatística T^2 de Hotelling excederam o limite de controle para 95% de confiança durante a alteração, evidenciando uma quebra na estrutura de correlação das variáveis nesse período. No entanto, houve uma recuperação da trajetória da batelada em ambas as cartas após cessada a alteração e retorno para a temperatura normal da reação (50°C). Esse fato pode ser corroborado pela avaliação da qualidade do produto final da reação, visto que a determinação do teor de etil-ésteres por análise cromatográfica efetuado após o término da batelada e posterior purificação, apresentou teor de 90,40% (m/m) para mistura binária (éster-óleo). Este resultado é semelhante aos obtidos para as bateladas NOC (Tabela 6).

Outra batelada presente nos gráficos de controle descritos na Figura 23 foi a batelada 12, produzida com uma concentração do catalisador maior (0,9% em relação a massa de óleo). Observa-se que houve uma quebra na estrutura de correlação das variáveis, apontada pela carta DModX desde os instantes iniciais da reação. Conforme descrito por Marjanovic' et al., (2010), com o aumento da concentração de catalisador, há um aumento na taxa de reação, com efeito mais intenso principalmente nos instantes iniciais, alterando assim a composição da mistura. Essa mudança pôde ser detectada pela espectroscopia NIR, o que refletiu nos gráficos DmodX e T^2 de Hotelling. Particularmente para o gráfico de T^2 de Hotelling, embora a batelada 12 tenha apresentado oscilação da trajetória ao redor do limite de confiança de 95%, a alteração não foi suficiente para exceder o limite de 99% como aconteceu para as alterações na carta de DModX. Por outro lado, a alteração na concentração de catalisador não teve um impacto grande sobre o teor de etil-ésteres para esta batelada, a qual apresentou um rendimento de 88% (m/m), que não é muito diferente do teor de EE observado para bateladas NOC, descritos na Tabela 6.

4.2.1.2 Avaliação da detecção de falhas via análise pós-batelada

Na detecção de falhas via análise pós-batelada pôde ser realizada uma comparação entre as estatísticas DModX e T^2 de Hotelling para as duas abordagens de desdobramento (NM e WKFH). As cartas de controle construídas com a abordagem NM foram obtidas a partir de um

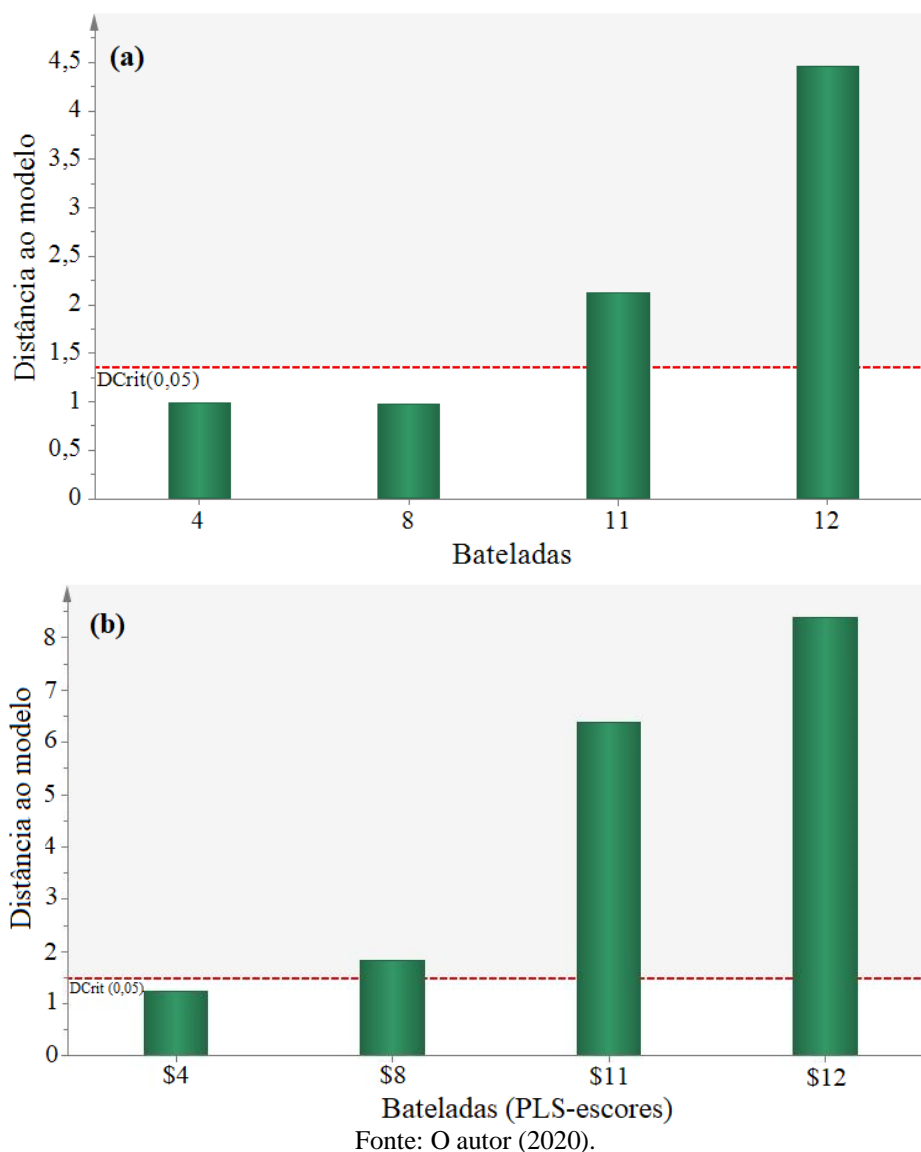
modelo PCA construído com os dados originais do processo, após rearranjo e pré-processamento com 1ª derivada. O modelo utilizou 2 variáveis latentes, obtendo 70% da variância dos dados do conjunto de calibração, e 36% de capacidade preditiva para o conjunto de predição (Tabela 7). O valor baixo apresentado pelos parâmetros descritos possivelmente está associado à baixa variabilidade apresentada pelas bateladas de calibração, especialmente pelo fato da abordagem NM analisar os dados de uma batelada completa como uma única amostra.

Por outro lado, os modelos construídos a partir da abordagem WKFH após pré-processamento com 1ª derivada, foram obtidos como resultado de um modelo PCA efetuado sobre os escores obtidos do modelo PLS desenvolvido pela análise em tempo real para esta mesma abordagem. O mesmo utilizou 4 variáveis latentes para obtenção de 95,1% da variância dos dados do conjunto de calibração e 69% de capacidade preditiva para o conjunto de predição (Tabela 7).

As cartas de controle da distância ao modelo para as bateladas do conjunto de predição estão descritas na Figura 24. Nesta carta, cada barra que excede o limite de controle, representa uma quebra do padrão na estrutura de correlação das variáveis, em relação aquelas utilizadas para a construção dos modelos do conjunto de calibração, segundo Ferrer, (2007).

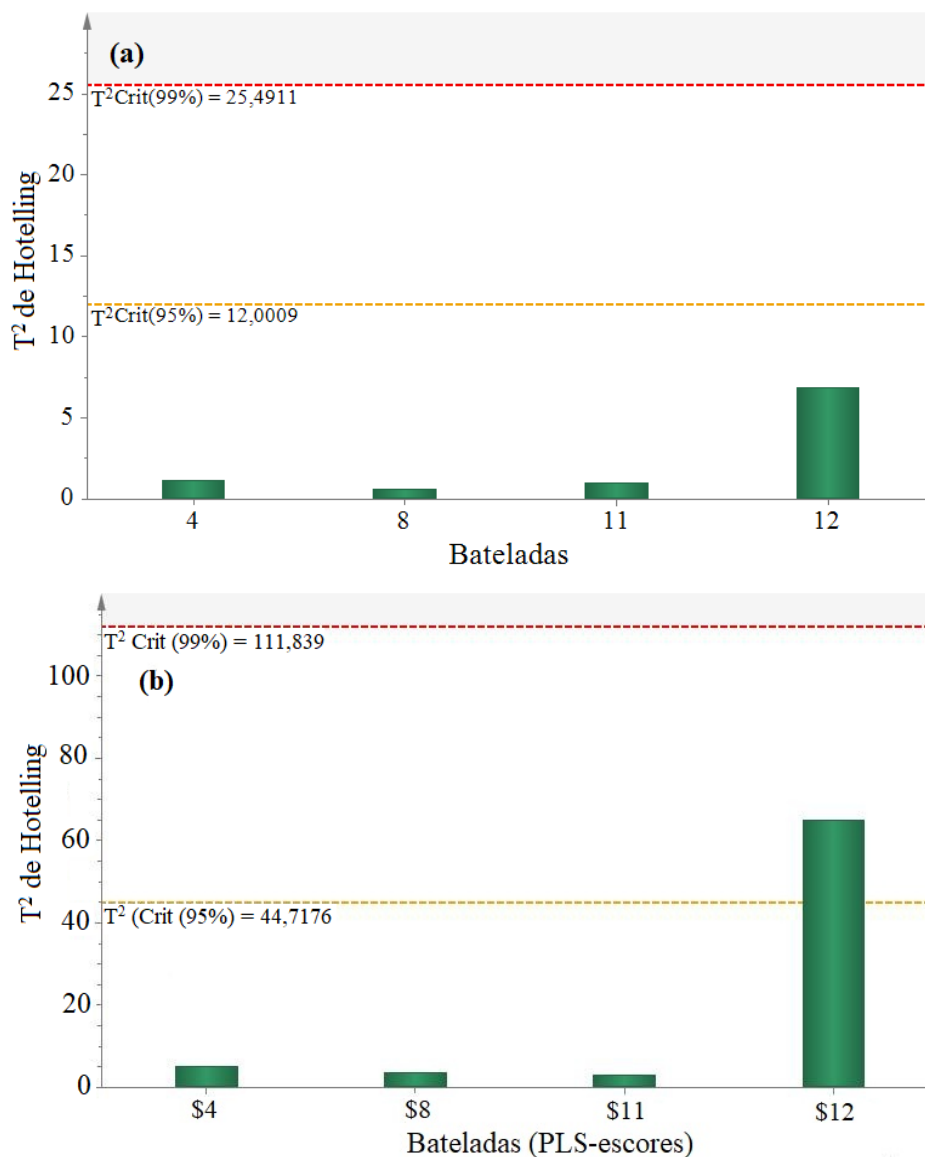
Considerando um nível de confiança de 95%, de acordo com a estatística DModX, as bateladas de predição (11 e 12) foram corretamente detectadas como bateladas fora de controle para as duas abordagens. Demonstrando que as estatísticas foram capazes de capturar a alteração temporária na temperatura (batelada 11) e o aumento da concentração do catalisador (batelada 12). Observa-se que a abordagem NM (Figura 24.a) também identificou corretamente as bateladas NOC presentes no conjunto de predição (bateladas 4 e 8), como bateladas dentro de controle. Esta condição não foi identificada para a batelada 8 quando utilizada a abordagem WKFH, a qual classificou a batelada 8 como uma batelada fora de controle para o nível de 0,05 de significância. Isto pode estar associado ao fato de o conjunto de calibração não contemplar totalmente a variabilidade do processo em NOC, visto que a distância desta batelada em relação ao plano do modelo não é tão grande, quando comparado com as bateladas em que houveram de fato alterações promovidas (11 e 12).

Figura 24 – Gráficos de controle de distância ao modelo para as bateladas de predição, utilizando-se a abordagem (a) NM e (b) WKFH. O símbolo (\$) à frente da descrição das bateladas para a abordagem WKFH indica que o gráfico foi construído a partir dos escores do modelo PLS, e não pelos dados originais.



Os gráficos de controle T^2 de Hotelling para as bateladas do conjunto de predição estão descritas na Figura 25. Para este gráfico, cada barra corresponde a distância de Mahalanobis de uma batelada em relação ao centro do plano que representa o modelo (FERRER, 2007). As bateladas NOC (bateladas 4 e 8) foram preditas corretamente, dentro de controle. No entanto, para a abordagem NM (Figura 25.a), as bateladas em que foram promovidas alterações (bateladas 11 e 12) não foram detectadas como bateladas fora de controle. Por outro lado, a batelada 12 (alteração na concentração do catalisador) ultrapassou o limite de confiança de 95% para a abordagem WKFH. Já a variação temporária da temperatura (batelada 11) não foi detectada por ambas as cartas.

Figura 25 – Gráficos de controle de T^2 de Hotelling para as bateladas de predição, utilizando-se a abordagem (a) NM e (b) WKFH. O símbolo (\$) à frente da descrição das bateladas para a abordagem WKFH indica que o gráfico foi construído a partir dos escores do modelo PLS, e não pelos dados originais.



Fonte: O autor (2020).

De acordo com a análise pós-batelada para o conjunto de predição, a partir das abordagens de desdobramento NM e WKFH, observou-se a partir do gráfico de DModX, que ambas as abordagens foram capazes de detectar corretamente as falhas intencionalmente provocadas (aumento temporário da temperatura – batelada 11 e aumento da concentração do catalisador – batelada 12), fato também observado por Sales, (2016) que monitorou a reação de transesterificação por rota metálica no modo *in-line* via cartas de controle multivariado, utilizando um equipamento NIR de bancada. Apenas a batelada 12 pôde ser detectada pela estatística T^2 de Hotelling com a abordagem WKFH.

Ambas as abordagens (NM e WKFH) foram capazes de detectar as bateladas produzidas em condições diferentes da NOC. Em geral, as diferenças impostas a essas bateladas geraram uma quebra na estrutura de correlação das variáveis e permitiram a detecção pela carta da distância ao modelo. Por fim, para o monitoramento *off-line*, ambas as abordagens apresentaram desempenho semelhante. Isso está relacionado ao fato de que, no modo *off-line*, a abordagem WKFH, assim como a abordagem NM, são capazes de capturar a dinâmica do processo.

5 CONCLUSÃO

A espectroscopia NIR associada à quimiometria mostrou desempenho satisfatório como uma ferramenta de PAT para o monitoramento da produção de biodiesel, através da reação de transesterificação em batelada do óleo de algodão por rota etílica e catálise alcalina. A utilização do espectrômetro NIR ultracompacto no modo *on-line* possibilitou a aquisição de dados em tempo real, de forma confiável, rápida, reprodutível e sem a necessidade de realização de tratamento nas amostras. As medidas foram obtidas com intervalo de tempo menor, cerca de 3 segundos, e por um equipamento de baixa resolução quando comparado a um equipamento NIR de bancada.

O modelo de calibração multivariada PLS construído para determinação do teor de ésteres (EE), a partir de misturas sintéticas e de amostras reais, apresentou capacidade preditiva satisfatória quando comparado com a literatura, com RMSEP = 1,51% m/m, considerando a mistura quaternária. Tal resultado foi obtido por meio do pré-processamento dos dados com 1ª derivada com filtro de Savitzky-Golay e tamanho de janela igual a 5 pontos, associado ao método de seleção de variáveis *i*PLS. A utilização de amostras sintéticas para a etapa de calibração trouxe simplicidade ao processo, reduzindo a necessidade de produzir muitas bateladas em condições distintas para aumentar a variabilidade da faixa de calibração, bem como a redução da quantidade de análises cromatográficas.

Já em relação ao controle estatístico multivariado de processos aplicado ao monitoramento da reação de transesterificação em batelada, de modo geral, as cartas de controle construídas de acordo com as abordagens NM e WKFH demonstraram resultados semelhantes quanto ao potencial em detectar falhas para a avaliação pós-batelada. A carta de distância ao modelo foi capaz de diagnosticar as falhas de forma mais eficiente, visto que a quebra na estrutura de correlação é refletida neste gráfico de controle.

Para a abordagem WKFH aplicada ao monitoramento em tempo real, quase todas as perturbações foram detectadas de modo satisfatório por ambas as cartas de controle. Este modo de monitoramento apresenta como principal vantagem para controle de processos industriais, a detecção de possíveis falhas durante o decorrer do processo, tornando viável a tomada de decisão para correção destas tão logo forem detectadas. Dessa forma, podem ser corrigidas antes do final do processo de fabricação, evitando custos com retrabalho e/ou perdas de matéria-prima e mão-de-obra.

De modo geral, Quase todas as falhas puderam ser detectadas por ambas as cartas de controle para as duas abordagens de organização e avaliação: NM e WKFH, com destaque para

a carta de distância ao modelo. Isto mostra que o sistema de monitoramento foi capaz de capturar mudanças importantes que ocorrem principalmente na fase inicial da reação de transesterificação.

Vale salientar que a sensibilidade da espectroscopia NIR frente à variação na temperatura da reação permitiu detectar uma mudança de apenas 5°C, configurando uma condição de alta sensibilidade para os métodos de monitoramento utilizados neste trabalho.

6 PERSPECTIVAS FUTURAS

Para a primeira abordagem do trabalho, relativa ao monitoramento da reação de transesterificação etílica em batelada do óleo de algodão, através de um modelo multivariado PLS, espera-se estender a faixa de concentração do modelo a fim de promover maior representatividade ao mesmo e testar outros aspectos relativos à modelagem, envolvendo mais métodos de pré-processamentos e seleção de variáveis.

Em relação ao monitoramento da reação por meio da aplicação da espectroscopia NIR com uma ferramenta de PAT para a construção de cartas de controle estatístico multivariado de processos (MSPC), espera-se:

- a) Produzir e monitorar mais bateladas em condições normais de operação a fim de atribuir ainda mais representatividade na variabilidade capturada pelos gráficos de controle a serem construídos;
- b) Promover outros tipos de alteração durante o monitoramento das reações, como alterações na temperatura ao longo da batelada, alteração da vazão de fluxo da linha de amostragem no retorno ao reator controlada via bomba peristáltica, alteração da velocidade de agitação, bem como simular defeito nos controladores desses parâmetros por meio de oscilações propositais, a fim de verificar a capacidade das cartas de controle em detectar essas falhas de processo;
- c) Comparar o desempenho de diferentes gráficos de controle para o monitoramento da reação de transesterificação em batelada no modo *on-line*, via utilização do MicroNIR, com o modo *in-line*, através do uso de equipamento de bancada com maior resolução, o qual já tem sido utilizado no monitoramento das bateladas já produzidas.

REFERÊNCIAS

AGUADO, D.; FERRER, A.; FERRER, J.; SECO, A. Multivariate SPC of a sequencing batch reactor for wastewater treatment. **Chemometrics and Intelligent Laboratory Systems**, v. 85, p. 82-93, 2007.

ANP – AGÊNCIA NACIONAL DE PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. **Anuário estatístico brasileiro do petróleo, gás natural e biocombustíveis de out/2019**. Disponível em: <<http://www.anp.gov.br/arquivos/central-conteudos/anuario-estatistico/2019/2019-anuario-versao-impressao.pdf>>. Acesso em 12 nov, 2019.

ANP – AGÊNCIA NACIONAL DE PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. **Resolução ANP N°798**. 1 ago, 2019 (DOU 02.8.2019). Disponível em: <<http://www.in.gov.br/web/dou/-/resolucao-n-798-de-1-de-agosto-de-2019-208544998>>. Acesso em: 05 jan, 2020.

ANP – AGÊNCIA NACIONAL DE PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS. **Resolução ANP N°45**. 25 ago, 2014 (DOU 26.8.2014). Disponível em: <<http://legislacao.anp.gov.br/?path=legislacao-anp/resol-anp/2014/agosto&item=ranp-45-2014>> . Acesso em: 05 jan, 2020.

AMERICAN STANDARD METHODS **ASTM E1655-05**. Standard Practices for Infrared Multivariate Quantitative Analysis. United States, 2005.

ARZAMENDI, G.; ARGUIÑARENA, E.; CAMPO, I.; GANDIA, L. M. Monitoring of biodiesel production: simultaneous analysis of the transesterification products using size-exclusion chromatography. **Chem Eng J**. v. 122, p. 31-40, 2006.

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

BARNES, R. J.; DHANOA, M. S.; LISTER, S. J. Standard Normal Variate Transformation and De-trending of Near-Infrared Diffuse Reflectance Spectra. **Applied Spectroscopy**. v. 43, n°. 5, p. 772-777, 1989.

BHUIYA, M. M. K.; RASUL, M.G.; KHAN, M.M.K.; ASHWATH, N.; AZAD. A.K. Prospects of 2nd generation biodiesel as a sustainable fuel – Part:1 selection of feedstocks, oil extraction techniques and conversion technologies. **Renewable and Sustainable Energy Reviews**, v. 55, p. 1109-1128, 2016.

BRASIL, Lei N°11.097, 13 jan., 2005 (DOU 14.1.2005). Disponível em:<<http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?jornal=1&pagina=8&data=14/01/2005>>. Acesso em 05 Jan, 2020.

BRASIL, Lei N°13.033, 24 set., 2014 (DOU 25.9.2014). Disponível em:<http://www.planalto.gov.br/ccivil_03/_Ato2011-2014/2014/Lei/L13033.htm>. Acesso em 05 Jan, 2020.

BRASIL, Lei N°13.263, 23 mar., 2016 (DOU 24.3.2016). Disponível em:<http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Lei/L13263.htm#art1>. Acesso em 05 Jan, 2020.

BRASIL. RESOLUÇÃO N°16, DE 29 DE OUTUBRO DE 2018. **Evolução da adição obrigatória de biodiesel ao óleo diesel vendido ao consumidor final, em qualquer parte do território nacional**, Brasília, DF, out 2018. Disponível em: <http://www.mme.gov.br/documents/10584/71068545/Resolucao_16_CNPE_29-10-18.pdf>. Acesso em: 02 out. 2019.

BRUNS, R. E.; FAIGLE, J. F. G. Quimiometria. **Química Nova**. v. 8, p 84-99, 1985.

BRUNS, R. E.; SCARMINIO, I. S.; NETO, B. B. Statistical design - Chemometrics, The Netherlands: Elsevier, 2006.

CATELANI, T. A.; SANTOS, J. R.; PÁSCOA, R. N. M. J.; PEZZA, L.; PEZZA, H. R.; LOPES, J. A. Real-time monitoring of a coffee roasting process with near infrared spectroscopy using multivariate statistical analysis: A feasibility study. **Talanta**. v. 179, p. 292-299, 2018.

CORREIA, R. M.; DOMINGOS, E.; CÁO, V. M.; ARAUJO, B. R. F.; SENA, S.; PINHEIRO, L. U.; FONTES, A. M.; AQUINO, L. F. M.; FERREIRA, E. C.; FILGUEIRAS, P. R.; ROMÃO, W. Portable near infrared spectroscopy applied to fuel quality control. **Talanta**. v. 176, p.26-33, 2018.

DARNOKO, D.; CHERYAN, M. Kinetics of palm oil transesterification in a batch reactor. **J. Am. Oil Chem. Soc.** v. 77, p. 1263-1267, 2000.

DEMIRBAS, A. Progress and recent trends in biodiesel fuels. **Energy Conversion and Management**, v. 50, p. 14-34, 2009.

DIAS, J. M.; ARAÚJO, J. M.; COSTA, J. F.; ALVIM-FERRAZ, M. C. M.; ALMEIDA, M.F. Biodiesel production from raw castor oil. **Energy**, v. 53, p. 58-66, 2013.

EUROPEAN STANDARD EN 14103. Fat and oil derivatives – Fatty Acid Methyl Esters (FAME) – Determination of ester and linolenic acid methyl ester contents. Brussels, Belgium: 2003.

FEARN, T.; RICCIOLI, C.; GARRIDO-VARO, A.; GUERRERO-GINEL, J. E. On the geometry of SNV and MSC. **Chemometrics and Intelligent Laboratory Systems**. v. 96, p. 22-26, 2009.

FERREIRA, M. M. C.; ANTUNES, A. M.; MELGO, M. S.; VOPE, P. L. O. Quimiometria I: Calibração Multivariada, um tutorial. **Química Nova**, v. 22, n°. 5, p. 724-731, 1999.

FERREIRA, M. M. C. **Quimiometria, Conceitos, Métodos e Aplicações**. 1ª ed., Campinas: Unicamp, 2015.

FERRER, A. Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. **Quality Engineering**, v. 19, p. 311-325, 2007.

FLORES-CERRILLO, J.; MACGREGOR, J. F. Control of batch product quality by trajectory manipulation using latent variable models. **Journal of Process Control**. v. 14, p. 539-553, 2004.

Food and Drug Administration (FDA). Guidance for Industry PAT – A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance. Pharmaceutical CGMPs, 2004. Disponível em <<https://www.fda.gov/media/71012/download>>. Acesso em: 05 Jan, 2020.

FREEDMAN B, KWOLEK WF, PRYDE EH. Quantitation in the analysis of transesterified soybean oil by capillary gas chromatography. **J Am Oil Chem Soc**. V. 63, nº10, p. 1370-1971, 1986.

GALVÃO, R. K. H.; ARAÚJO, M. C. U.; JOSE, G. E.; PONTES, M. J. C.; SILVA, E. C.; SALDANHA, T. C. B. A method for calibration and validation subset partitioning. **Talanta**, v. 67, p. 736-740, 2005.

GELADI, P.; KOWALSKI, B. R. Partial least-squares regression: a tutorial. **Analytica Chimica Acta**, v. 186, p. 1-17, 1986.

GHESTI, G. F.; MACEDO, J. L. D.; PARENTE, V. C. I.; DIAS, J. A.; DIAS, S. C. L. Synthesis, characterization and reactivity of Lewis acid/surfactant cerium trisdodecylsulfate catalyst for transesterification and esterification reactions. **Applied Catalysis A: General**. v. 355, p. 139-147, 2009.

GONZÁLEZ-MARTÍNEZ, J. M.; CAMACHO, J.; FERRER, A. MVBatch: A matlab toolbox for batch process modeling and monitoring. **Chemometrics and Intelligent Laboratory Systems**. v. 183, p. 122–133, 2018.

GÜNAY, M. E.; TÜRKER, L.; TAPAN, N. A. Significant parameters and technological advancements in biodiesel production systems. **Fuel**, v. 250, p. 27-41, 2019.

JOSHI, H. C.; TOLER, J.; WALKER, T. Optimization of cottonseed oil ethanolysis to produce biodiesel high in gossypol content. **Journal of the American Oil Chemists' Society**. v. 85, p. 357-363, 2008.

KILLNER, M. H. M.; ROHWEDDER, J. J. R.; PASQUINI, C. A PLS regression model using NIR spectroscopy for on-line monitoring of the biodiesel production reaction. **Fuel**, v. 90, p. 3268-3273, 2011.

KNOTHE G. Analytical methods used in the production and fuel quality assessment of biodiesel. **Trans ASAE**. v. 44, p. 193-200, 2001.

KONA, R.; Qu, H.; MATTES, R.; JANCSEK, B.; FAHMY, R. M.; HOAG, S. W. Application of in-line near infrared spectroscopy and multivariate batch modeling for process monitoring in fluid bed granulation. **International Journal of Pharmaceutics**, v. 452, p. 63-72, 2013.

KOURTI, T.; MACGREGOR, J. F. Multivariate SPC methods for process and product monitoring. **Journal of Quality Technology**, v. 28, nº. 4, p. 409-428, 1996.

KOURTI, T. Application of latent variable methods to process control and multivariate statistical process control in industry. **International Journal of Adaptive Control and Signal Processing**. v. 19, p. 213-246, 2005.

KOWALSKI, B. R. Chemometrics: Views and Propositions. **Journal of Chemical Information and Computer Sciences**. v. 15, n°4, p. 201-203, 1975.

LIMA, S. M.; SILVA, B. F. A.; PONTES, D. V.; PEREIRA, C. F.; STRAGEVITCH, L.; PIMENTEL, M. F. In-line monitoring of the transesterification reactions for biodiesel production using NIR spectroscopy, **Fuel**, v. 115, p. 46-53, 2014.

LIMA, Suzana Moreira. **Tecnologias analíticas de processo aplicadas à produção de biodiesel**. Tese de Doutorado. Recife, PE: UFPE, 2014.

LI, W.; HAN, H.; CHENG, Z.; ZHANG, Y.; LIU, S.; QU, H. A feasibility research on the monitoring of traditional Chinese medicine production process using NIR-based multivariate process trajectories. **Sensors and Actuators: B**. v. 231, p. 313-323, 2016.

MARJANOVIC', A. V.; STAMENKOVIC', O. S.; TODOROVIC', Z. B.; LAZIC', M. L.; VELJKOVIC', V. B. Kinetics of the base-catalyzed sunflower oil ethanolysis. **Fuel**, v. 89, p. 665-671, 2010.

MENEGHETTI, S. M. P.; MENEGHETTI M. R.; WOLF, C. R.; SILVA, E. C.; LIMA, G. E. S.; DE LIRA SILVA L.; SERRA, T. M.; CAUDURO, F.; OLIVEIRA, L. G. Biodiesel from castor oil: a comparison of ethanolysis versus methanolysis. **Energy & Fuels**. v. 20, p. 2262-2265, 2006.

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

NOMIKOS, P.; MACGREGOR, J. F. Multivariate SPC charts for monitoring batch processes. **Technometrics**, v. 37, p 41-59, 1995.

NØRGAARD. L.; SAUDLAND, A.; WAGNER, J.; NIELSEN, J. P.; MUNCK, L.; ENGELSEN, S. B. Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy. **Applied Spectroscopy**. v. 54, n°3, p. 413-419, 1999.

O'BRIEN, N. A.; HULSE, C. A.; FRIEDRICH, D. M.; MILLIGEN, F. J. V.; GUNTEN, M. K. V.; PFEIFER, F.; SIESLER, H. W. Miniature Near-Infrared (NIR) Spectrometer Engine For Handheld Applications. **Next-Generation Spectroscopic Technologies**. v. 8374, p. 1-8, 2012.

PAIVA, E. M.; ROHWEDDER, J. J. R.; PASQUINI, C.; PIMENTEL, M. F.; PEREIRA, C. F. Quantification of biodiesel and adulteration with vegetable oils in diesel/biodiesel blends using portable near-infrared spectrometer, **Fuel**, v. 160, p. 57-63, 2015.

PASQUINI, C. Near-Infrared spectroscopy: Fundamentals, practical aspects and analytical applications. **J. Braz. Chem. Soc.** v. 14, n. 2, p. 198-219, 2003.

PASQUINI, C. Near infrared spectroscopy: A mature analytical technique with new perspectives e A review. **Analytica Chimica Acta**. v. 1026, p. 8-36, 2018.

PIMENTEL M.F.; RIBEIRO, G. M. G. S.; CRUZ, R. S.; STRAGEVITCH, L.; PACHECO FILHO, J.G.A.; TEIXEIRA, L.S.G. Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration. **Microchemical Journal**. Ed. 2, v. 82, p. 201-206, 2006.

PINTO, A. C.; GUARIEIRO, L. L. N.; REZENDE, M. J. C.; RIBEIRO, N. M.; TORRES, E. A.; LOPES, W. A.; PEREIRA, P. A. P.; ANDRADE, J. B. Biodiesel: An Overview. **Journal of Brazilian Chemical Society**. v. 16, nº. 6B, p. 1313-1330, 2005.

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

RICHARD, R.; DUBREUIL, B.; THIEBAUD-ROUX, S.; PRAT, L. On-line monitoring of the transesterification reaction carried out in microreactors using near infrared spectroscopy. **Fuel**, v. 104, p. 318-325, 2013.

RINNAN, Å.; BERG, F. V. D.; ENGELSEN, S. B. Review of the most common pre-processing techniques for near-infrared spectra. **Trends in Analytical Chemistry**, v. 28, p. 1201-1222, 2009.

ROUCHI, M. B.; KHORRAMI, M. K.; GARMARUDI, A. B.; GUARDIA, M. Application of infrared spectroscopy as Process Analytics Technology (PAT) approach in biodiesel production process utilizing Multivariate Curve Resolution Alternative Least Square (MCR-ALS). **Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy**. v. 213, p. 347-353, 2019.

SALES, R. F.; VITALE, R.; LIMA, S. M.; PIMENTEL, M. F.; STRAGEVITCH, L.; FERRER, A. Multivariate statistical process control charts for batch monitoring of transesterification reactions for biodiesel production based on near-infrared spectroscopy. **Computers & Chemical Engineering**, v. 94, p. 343-353, 2016.

SALES, R.; SILVA, N. C.; SILVA, J. P.; FRANÇA, H. H.; PIMENTEL, M. F.; STRAGEVITCH, L. Handheld near-infrared spectrometer for on-line monitoring of biodiesel production in a continuous process. **Fuel**, v. 254, p. 1-8, 2019.

SANDEN, A.; SUHM, S.; RÜDT, M. HUBBUCH, J. Fourier-transform infrared spectroscopy as a process analytical technology for near real time in-line estimation of the degree of PEGylation in chromatography. **Journal of Chromatography A**. v. 1608, p. 1-6, 2019.

SILVA, Neirivaldo Cavalcante. **Uso da Espectroscopia NIR no desenvolvimento de um simulador para gasolina e na transferência de calibração entre instrumentos de bancadas e portátil**. Tese de Doutorado. Recife, PE: UFPE, 2017.

SINGH, D.; SHARMA, D.; SONI, S.L.; SHARMA, S.; KUMARIB, D. Chemical compositions, properties, and standards for different generation biodiesels: A review. **Fuel**, v. 253 p. 60-71, 2019.

SKOOG D. A.; HOLLER, F. J.; CROUCH, S. R. **Principles of Instrumental Analysis**. 7 ed., Boston: Cengage Learning, 2017.

SKOOG D. A.; WEST, D. M.; HOLLER, F. J.; CROUCH, S. R. Tradução Grassi M; revisão técnica Pasquini C. **Fundamentos de Química Analítica**. Tradução da 8ª edição norte-americana. São Paulo: Cengage Learning, 2008.

SOUZA, A. M.; BREITKREITZ, M. C.; FILGUEIRAS, P. R.; ROHWEDDER, J. R.; POPPI, R. J. Experimento didático de quimiometria para calibração multivariada na determinação de paracetamol em comprimidos comerciais utilizando espectroscopia no infravermelho próximo: um tutorial, parte II. **Química Nova**, v. 36, n°. 7, p. 1057-1065, 2013.

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

THANH, L.T.; OKITSU, K.; VAN BOI, L.; MAEDA, Y. Catalytic technologies for biodiesel fuel production and utilization of glycerol: a review. **Catalysts**, v 2. p. 191-222, 2012.

TREVISAN, M. G.; POPPI, R. J. Química analítica de Processos. **Química Nova**, v. 29, n°5, p. 1065-1071, 2006.

VELIC´KOVIC´, A. V.; STAMENKOVIC´, O.S.; TODOROVIC´, Z. B.; VELJKOVIC´, V. B. Application of the full factorial design to optimization of base-catalyzed sunflower oil ethanolysis. **Fuel**, v. 104, p. 433-442, 2013.

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

WOLD, S.; KETTANEH, N.; FRIDÉN, N.; HOLMBERG, A. Modelling and diagnostics of batch processes and analogous kinetic experiments. **Chemometrics and Intelligent Laboratory Systems**, v. 44, p. 331-340, 1998.

WORKMAN, J.; WEYER, L. **Practical Guide to Interpretive Near-Infrared Spectroscopy**. 2. ed. New York: CRC Press, 2012.

YANG, Z.; MOCK, P.; GERMAN, J.; BANDIVADEKAR, A.; LAH, O. On a pathway to decarbonization – A comparison of new passenger car CO₂ emission standards and taxation measures in the G20 countries. **Transportation Research Part D**, v. 64, p. 53-69, 2018.

YUSUF, N.N.A.N.; KAMARUDIN, S.K.; YAAKUB, Z. Overview on the current trends in biodiesel production. **Energy Conversion and Management**. v. 52, p. 2741-2751, 2011.

ZITTING, J. **Optical sorting technology for textile waste Development of an identification method with NIR spectroscopy**. Monografia. Mikkula, La: LAMK, 2017.

APÊNDICE A – TABELA DO PLANEJAMENTO DAS MISTURAS SINTÉTICAS

Amostra	Planejamento de misturas				Pureza bio (%)	Cálculo das massas				T (°C)	Massas reais (g)			
	Óleo	Etanol	Biodiesel	Glicerol		Óleo (g)	Etanol (g)	Biodiesel (g)	Glicerol (g)		Óleo (g)	Etanol (g)	Biodiesel (g)	Glicerol (g)
1	0,30	0,34	0,31	0,05	94,90	5,6668	6,8000	6,5332	1,0000	50	5,6701	6,8000	6,5402	1,0085
2	0,04	0,34	0,60	0,03	94,90	0,2006	8,3750	15,6744	0,7500	50	0,2076	8,3768	15,6848	0,7529
3	0,19	0,32	0,44	0,04	91,83	5,2021	11,2955	16,9434	1,5591	50	5,2292	11,2953	16,9500	1,5594
4	0,13	0,31	0,52	0,04	93,36	3,2181	10,8977	19,5796	1,3045	50	3,2296	10,8048	19,5801	1,3048
5/6	0,31	0,31	0,32	0,06	89,66	9,5584	10,8500	12,4916	2,1000	50/60	9,5677	10,8514	12,4956	2,1023
7	0,18	0,30	0,47	0,06	91,87	4,6848	10,5000	17,7152	2,1000	50	4,6919	10,5213	17,7226	2,1039
8/9	0,21	0,33	0,42	0,04	93,36	6,4175	11,4765	15,6255	1,4840	50/60	6,4206	11,4780	15,6292	1,4852
10	0,31	0,30	0,35	0,05	91,83	9,7757	10,5000	13,1493	1,5750	50	9,7772	10,5210	13,1566	1,5758
11	0,11	0,33	0,52	0,04	89,66	1,8851	11,4227	20,3876	1,3045	50	1,8876	11,4289	20,3930	1,3113
12/13	0,04	0,33	0,60	0,03	94,90	0,1939	8,2500	15,8061	0,7500	50/60	0,2000	8,2552	15,8082	0,7511
14	0,30	0,33	0,31	0,05	91,87	9,6565	11,6667	11,8102	1,8667	50	9,6698	11,6695	11,8107	1,8672
15	0,09	0,31	0,55	0,06	89,66	0,9957	10,7905	21,2748	1,9390	50	1,0019	10,7945	21,2765	1,9399
16	0,04	0,34	0,56	0,06	93,36	0,0060	11,9000	20,9940	2,1000	50	0,0168	11,9167	20,9958	2,1000
17/18	0,11	0,31	0,52	0,05	93,36	2,6931	10,8977	19,5796	1,8295	50/60	2,7002	10,9023	19,5803	1,8296
19	0,18	0,34	0,46	0,03	91,83	4,7082	11,9000	17,3418	1,0500	50	4,7159	11,9050	17,3493	1,0503
20	0,15	0,32	0,48	0,05	92,00	3,8599	11,1335	18,2951	1,7115	50	3,8655	11,1355	18,3171	1,7115
21	0,06	0,30	0,60	0,05	91,87	0,0666	10,5000	22,8584	1,5750	50	0,0700	10,6386	22,8630	1,5761
22	0,24	0,33	0,38	0,04	90,00	7,0229	11,6480	14,9606	1,3685	50	7,0309	11,6487	14,9653	1,3697
23	0,31	0,34	0,31	0,04	89,66	9,5987	11,9000	12,1013	1,4000	50	9,6006	11,9017	12,1027	1,4030
24/25	0,19	0,30	0,48	0,03	91,83	5,1553	10,5000	18,2947	1,0500	50/60	5,1700	10,5052	18,2952	1,0502
26	0,25	0,33	0,38	0,04	93,36	7,7666	11,5977	14,3311	1,3045	50	7,7696	11,5992	14,3338	1,3045

APÊNDICE A – CONTINUAÇÃO

27	0,31	0,34	0,31	0,03	89,66	9,5923	12,0260	12,2652	1,1165	50	9,6046	12,0265	12,2687	1,1175
28/29	0,07	0,30	0,60	0,03	91,87	0,5916	10,5000	22,8584	1,0500	50/60	0,5980	10,5080	22,8625	1,0519
30	0,06	0,30	0,58	0,06	92,00	0,3751	10,6225	21,9549	2,0510	50	0,3780	10,6250	21,9816	2,0517
31	0,31	0,34	0,32	0,03	90,00	9,6056	11,9000	12,4444	1,0500	50	9,6212	11,9060	12,4481	1,0523
32	0,25	0,33	0,38	0,04	90,00	7,2510	11,5977	14,6717	1,4795	50	7,2514	11,5980	14,6761	1,4815
33	0,04	0,34	0,59	0,03	94,90	0,2073	8,5000	15,5427	0,7500	50	0,2181	8,5030	15,5498	0,7506
34	0,31	0,32	0,34	0,03	89,66	9,4776	11,2000	13,2724	1,0500	50	9,4884	11,2000	13,2763	1,0511
35/36	0,27	0,34	0,35	0,04	91,87	8,4909	11,8195	13,4331	1,2530	50/60	8,4986	11,8195	13,4442	1,2535
37	0,04	0,33	0,59	0,05	94,90	0,2107	8,1875	15,4768	1,1250	50	0,2113	8,1892	15,4848	1,1254
38	0,20	0,34	0,42	0,04	91,87	5,6353	11,9000	15,9247	1,5400	50	5,6415	11,9000	15,9318	1,5402
39	0,18	0,30	0,47	0,05	91,83	4,9162	10,5000	18,0088	1,5750	50	4,9283	10,5030	18,0119	1,5759
40	0,31	0,30	0,33	0,06	91,83	9,8224	10,5000	12,5776	2,1000	50	9,8260	10,5024	12,5801	2,1003
41	0,18	0,32	0,45	0,05	90,00	4,6422	11,3050	17,4533	1,5960	50	4,6544	11,3355	17,4647	1,5968
42	0,25	0,31	0,39	0,05	90,00	7,2121	10,8977	15,0606	1,8295	50	7,2194	10,8977	15,0654	1,8295
43/44	0,31	0,32	0,33	0,04	92,00	9,8578	11,2000	12,4022	1,5400	50/60	9,8695	11,2133	12,4086	1,5405
45	0,06	0,32	0,60	0,03	92,00	0,0989	11,0250	22,8261	1,0500	50	0,0989	11,0348	22,8307	1,0512
46	0,04	0,30	0,60	0,06	93,36	0,0066	10,5175	22,3549	2,1105	50	0,0165	10,5235	22,3588	2,1105
47	0,25	0,31	0,40	0,04	90,00	7,1538	10,8977	15,6439	1,3045	50	7,1578	10,8999	15,6451	1,3049
48	0,05	0,31	0,60	0,04	93,36	0,2564	10,8500	22,4936	1,4000	50	0,2684	10,8710	22,4936	1,4011
49	0,31	0,33	0,31	0,05	89,66	9,5987	11,5500	12,1013	1,7500	50	9,6012	11,5540	12,1017	1,7514
50	0,11	0,33	0,50	0,05	91,83	2,4292	11,5977	19,1436	1,8295	50	2,4356	11,5620	19,1470	1,8306
51/52	0,11	0,33	0,52	0,04	91,87	2,3910	11,5977	19,7067	1,3045	50/60	2,4018	11,6039	19,7106	1,3056
53/54	0,24	0,33	0,38	0,05	91,83	7,1934	11,5977	14,3793	1,8295	50/60	7,1938	11,5990	14,3824	1,8297

Fonte: O autor (2020).

**APÊNDICE B – CORRELAÇÃO ENTRE AS FAIXAS ADOTAS PELO *i*PLS E OS
RESPECTIVOS COMPRIMENTOS DE ONDA PARA DIFERENTES PRÉ-
PROCESSAMENTOS UTILIZADOS**

Tabela B.1 – Correlação entre as faixas adotadas pelo *i*PLS e os respectivos comprimentos de onda para o pré-processamento com SNV e MSC.

Pré-processamento SNV/ MSC					
Posição da linha	Comprimento de onda (nm)	Posição da linha	Comprimento de onda (nm)	Posição da linha	Comprimento de onda (nm)
1	908	45	1181	89	1453
2	914	46	1187	90	1459
3	920	47	1193	91	1466
4	927	48	1199	92	1472
5	933	49	1205	93	1478
6	939	50	1212	94	1484
7	945	51	1218	95	1490
8	951	52	1224	96	1497
9	958	53	1230	97	1503
10	964	54	1236	98	1509
11	970	55	1243	99	1515
12	976	56	1249	100	1521
13	982	57	1255	101	1528
14	989	58	1261	102	1534
15	995	59	1267	103	1540
16	1001	60	1274	104	1546
17	1007	61	1280	105	1552
18	1013	62	1286	106	1559
19	1020	63	1292	107	1565
20	1026	64	1298	108	1571
21	1032	65	1305	109	1577
22	1038	66	1311	110	1583
23	1044	67	1317	111	1589
24	1051	68	1323	112	1596
25	1057	69	1329	113	1602
26	1063	70	1336	114	1608
27	1069	71	1342	115	1614
28	1075	72	1348	116	1620
29	1082	73	1354	117	1627
30	1088	74	1360	118	1633
31	1094	75	1366	119	1639
32	1100	76	1373	120	1645
33	1106	77	1379	121	1651
34	1113	78	1385	122	1658
35	1119	79	1391	123	1664
36	1125	80	1397	124	1670
37	1131	81	1404	125	1676
38	1137	82	1410		
39	1143	83	1416		
40	1150	84	1422		
41	1156	85	1428		
42	1162	86	1435		
43	1168	87	1441		
44	1174	88	1447		

Fonte: O autor (2020).

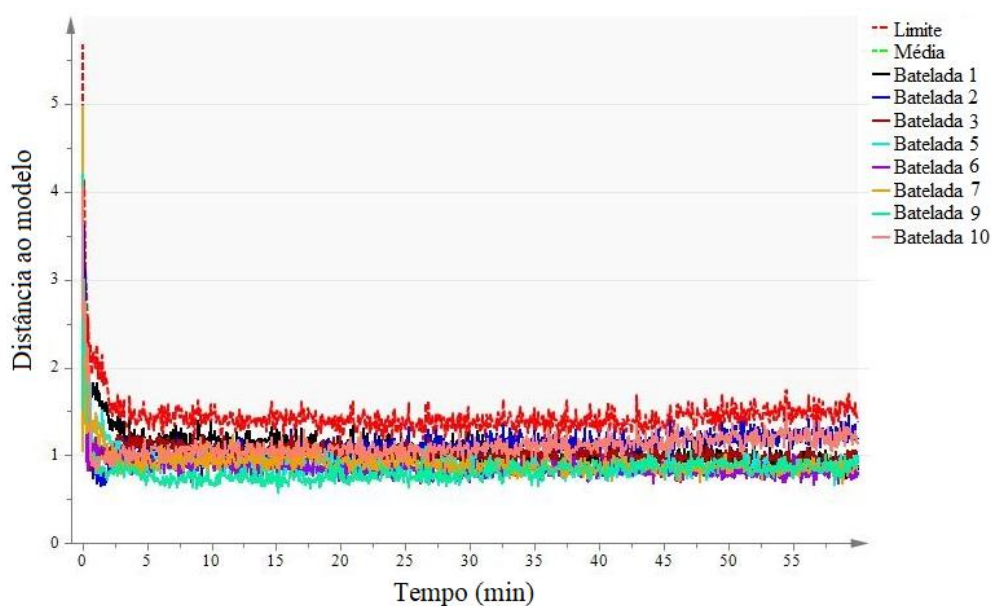
Tabela B.2 – Correlação entre as faixas adotadas pelo iPLS e os respectivos comprimentos de onda para o pré-processamento com 1ª derivada.

Pré-processamento 1ª Derivada					
Posição da linha	Comprimento de onda (nm)	Posição da linha	Comprimento de onda (nm)	Posição da linha	Comprimento de onda (nm)
1	945	45	1218	89	1490
2	951	46	1224	90	1497
3	958	47	1230	91	1503
4	964	48	1236	92	1509
5	970	49	1243	93	1515
6	976	50	1249	94	1521
7	982	51	1255	95	1528
8	989	52	1261	96	1534
9	995	53	1267	97	1540
10	1001	54	1274	98	1546
11	1007	55	1280	99	1552
12	1013	56	1286	100	1559
13	1020	57	1292	101	1565
14	1026	58	1298	102	1571
15	1032	59	1305	103	1577
16	1038	60	1311	104	1583
17	1044	61	1317	105	1589
18	1051	62	1323	106	1596
19	1057	63	1329	107	1602
20	1063	64	1336	108	1608
21	1069	65	1342	109	1614
22	1075	66	1348	110	1620
23	1082	67	1354	111	1627
24	1088	68	1360	112	1633
25	1094	69	1366	113	1639
26	1100	70	1373	114	1645
27	1106	71	1379	115	1651
28	1113	72	1385	116	1658
29	1119	73	1391	117	1664
30	1125	74	1397		
31	1131	75	1404		
32	1137	76	1410		
33	1143	77	1416		
34	1150	78	1422		
35	1156	79	1428		
36	1162	80	1435		
37	1168	81	1441		
38	1174	82	1447		
39	1181	83	1453		
40	1187	84	1459		
41	1193	85	1466		
42	1199	86	1472		
43	1205	87	1478		
44	1212	88	1484		

Fonte: O autor (2020).

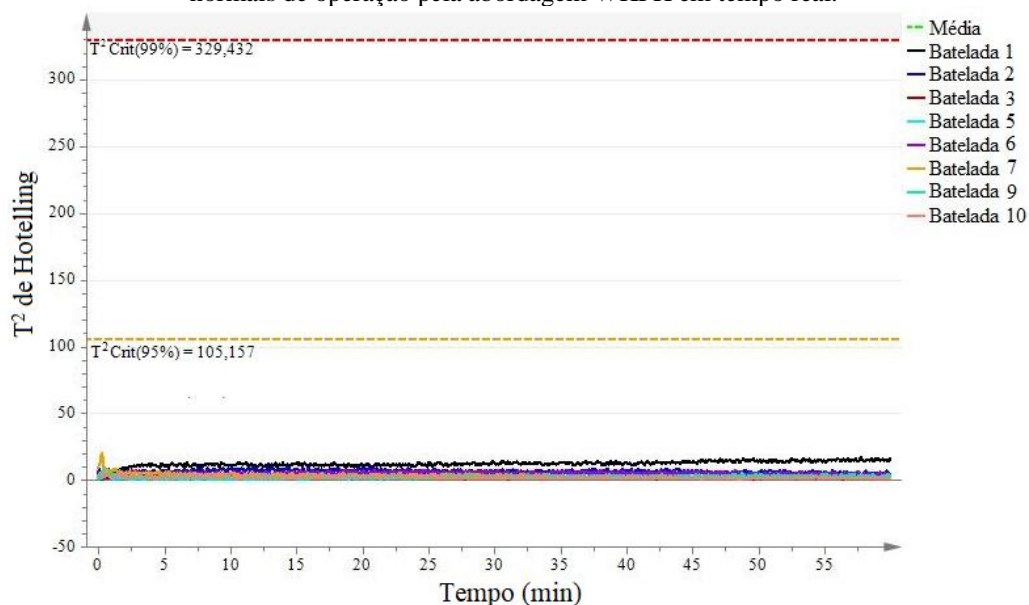
APÊNDICE C – GRÁFICOS DE CONTROLE DO CONJUNTO DE CALIBRAÇÃO REFERENTE AS ABORDAGENS DE DESDOBRAMENTO NM E WKFH PARA ANÁLISE PÓS-BATELADA E EM TEMPO REAL

Figura C1 – Gráfico da distância ao modelo para as bateladas do conjunto de calibração produzidas em condições normais de operação pela abordagem WKFH em tempo real.



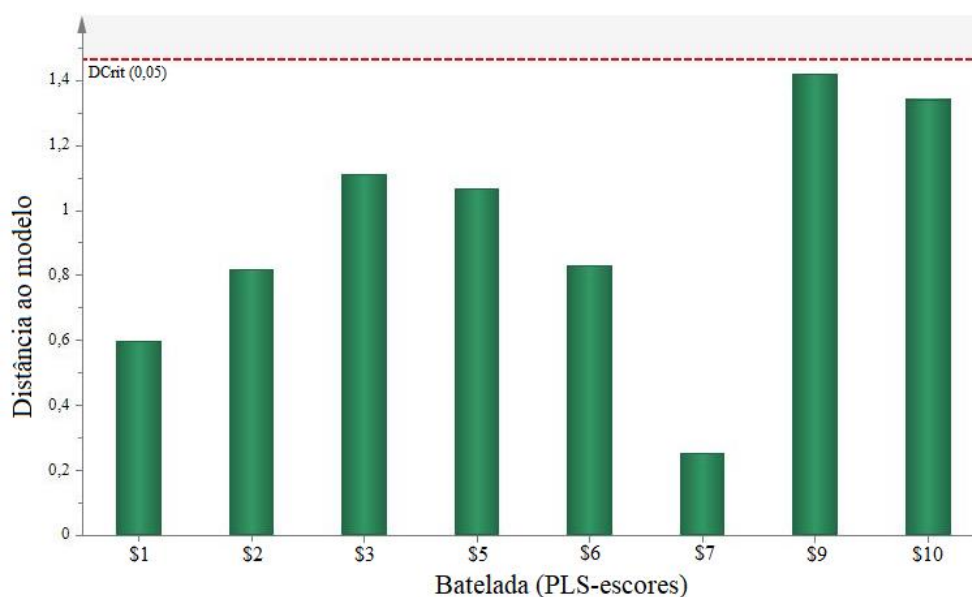
Fonte: O autor (2020).

Figura C2 – Gráfico de T^2 de Hotelling para as bateladas do conjunto de calibração produzidas em condições normais de operação pela abordagem WKFH em tempo real.



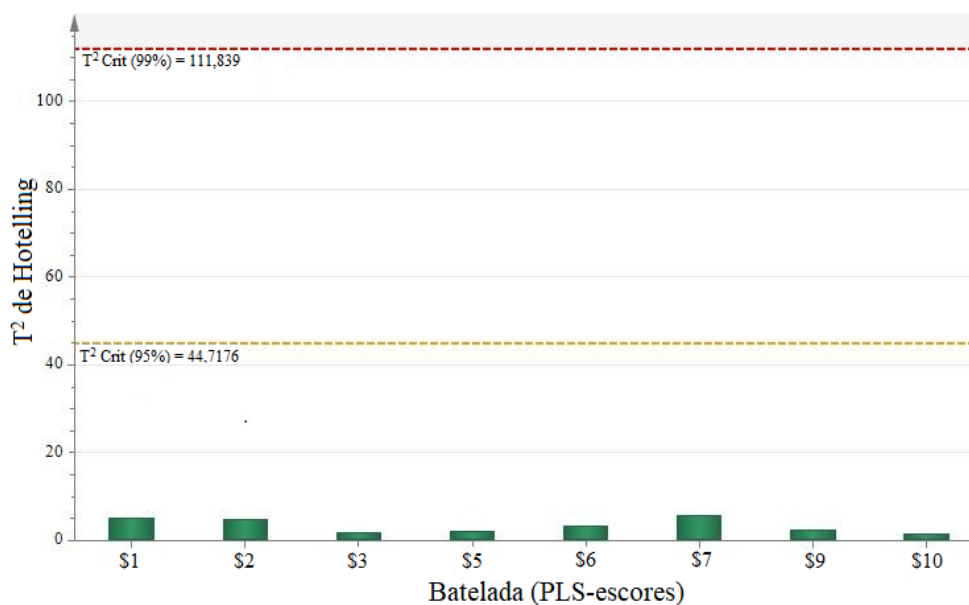
Fonte: O autor (2020).

Figura C3 – Gráfico de distância ao modelo para as bateladas do conjunto de calibração produzidas em NOC pela abordagem WKFH pós-batelada. O símbolo (\$) à frente da descrição das bateladas indica que o gráfico foi construído a partir dos escores do modelo PLS, e não pelos dados originais.



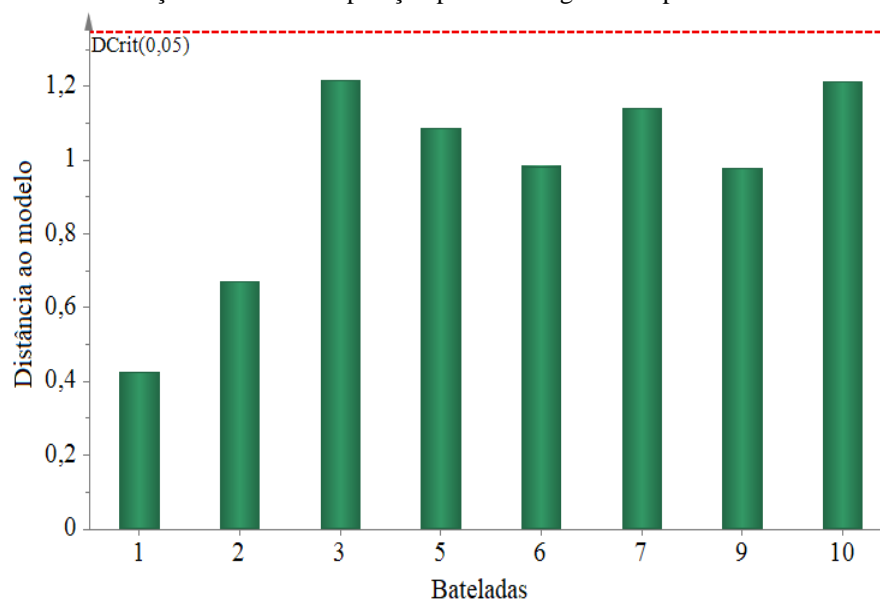
Fonte: O autor (2020).

Figura C4 – Gráfico de T^2 de Hotelling para as bateladas do conjunto de calibração produzidas em NOC pela abordagem WKFH pós-batelada. O símbolo (\$) à frente da descrição das bateladas indica que o gráfico foi construído a partir dos escores do modelo PLS, e não pelos dados originais.



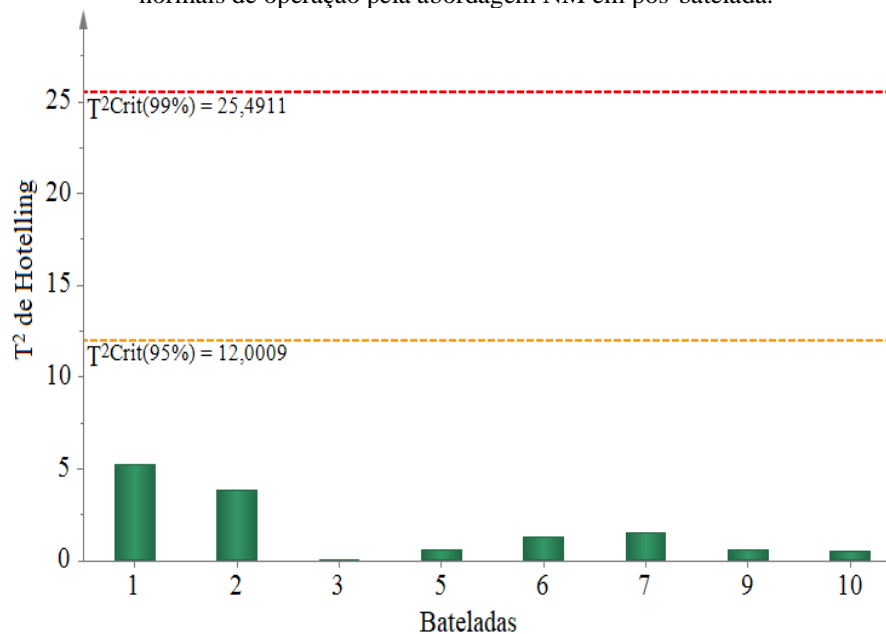
Fonte: O autor (2020).

Figura C5 – Gráfico da distância ao modelo para as bateladas do conjunto de calibração produzidas em condições normais de operação pela abordagem NM pós-batelada.



Fonte: O autor (2020).

Figura C6 – Gráfico de T^2 de Hotelling para as bateladas do conjunto de calibração produzidas em condições normais de operação pela abordagem NM em pós-batelada.



Fonte: O autor (2020).