



UNIVERSIDADE FEDERAL DE PERNAMBUCO  
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS  
DEPARTAMENTO DE ENGENHARIA QUÍMICA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA QUÍMICA

RAFAELLA DE FIGUEIREDO SALES

**APLICAÇÃO DE TECNOLOGIAS ANALÍTICAS DE PROCESSO PARA  
MONITORAMENTO NA PRODUÇÃO DE BIOCOMBUSTÍVEIS E METALIZAÇÃO  
INDUSTRIAL DE FILMES POLIMÉRICOS**

Recife

2020

RAFAELLA DE FIGUEIREDO SALES

**APLICAÇÃO DE TECNOLOGIAS ANALÍTICAS DE PROCESSO PARA  
MONITORAMENTO NA PRODUÇÃO DE BIOCOMBUSTÍVEIS E METALIZAÇÃO  
INDUSTRIAL DE FILMES POLIMÉRICOS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutora em Engenharia Química.

Área de concentração: Engenharia de Processos Químicos e Bioquímicos.

Orientadores: Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Fernanda Pimentel Avelar.

Prof. Dr. Luiz Stragevitch.

Coorientador: Prof. Dr. Neirivaldo Cavalcante da Silva.

Recife

2020

Catalogação na fonte  
Bibliotecária Margareth Malta, CRB-4 / 1198

S163a Sales, Rafaella de Figueiredo.  
Aplicação de tecnologias analíticas de processo para monitoramento na produção de biocombustíveis e metalização industrial de filmes poliméricos / Rafaella de Figueiredo Sales - 2020.  
122 folhas, il., gráfs., tabs.

Orientadores: Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Fernanda Pimentel Avelar.  
Prof. Dr. Luiz Stragevitch.

Coorientador: Prof. Dr. Neirivaldo Cavalcante da Silva.  
Tese (Doutorado) – Universidade Federal de Pernambuco. CTG.  
Programa de Pós-Graduação em Engenharia Química, 2020.  
Inclui Referências e Apêndices.

1. Engenharia Química. 2. Calibração multivariada. 3.  
Monitoramento estatístico de processo. 4. Quimiometria de processo. 5.  
Tecnologia analítica de processo. I. Avelar, Maria Fernanda Pimentel  
(Orientadora). II. Stragevitch, Luiz (Orientador). III. Silva, Neirivaldo  
Cavalcante da (Coorientador). IV. Título.

UFPE

660.2 CDD (22. ed.)

BCTG/2021-50

RAFAELLA DE FIGUEIREDO SALES

**APLICAÇÃO DE TECNOLOGIAS ANALÍTICAS DE PROCESSO PARA  
MONITORAMENTO NA PRODUÇÃO DE BIOCOMBUSTÍVEIS E METALIZAÇÃO  
INDUSTRIAL DE FILMES POLIMÉRICOS**

Tese apresentada ao Programa de Pós-Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Doutora em Engenharia Química.

Aprovada em: 23/07/2020.

**BANCA EXAMINADORA**

---

Profa. Dra. Maria Fernanda Pimentel Avelar (Orientadora)  
Universidade Federal de Pernambuco

---

Prof. Dr. Luiz Stragevitch (Orientador)  
Universidade Federal de Pernambuco

---

Prof. Dr. Neirivaldo Cavalcante da Silva (Coorientador)  
Universidade Federal do Pará

---

Prof. Dr. Adriano de Araújo Gomes (Examinador Externo)  
Universidade Federal do Rio Grande do Norte

---

Profa. Dra. Celmy Maria Bezerra de Menezes Barbosa (Examinadora Interna)  
Universidade Federal de Pernambuco

---

Prof. Dr. Heronides Adonias Dantas Filho (Examinador Externo)  
Universidade Federal do Pará

---

Prof. Dr. José Licarion Pinto Segundo Neto (Examinador Externo)  
Universidade Federal de Pernambuco

---

Profa. Dra. Marta Maria Menezes Bezerra Duarte (Examinadora Interna)  
Universidade Federal de Pernambuco

## AGRADECIMENTOS

A Deus, por estar sempre presente em minha vida, me protegendo e guiando nessa jornada.

A minha orientadora, Prof.<sup>a</sup> Dr.<sup>a</sup> Maria Fernanda Pimentel, à qual sou imensamente grata pela excelente orientação, pela dedicação ao trabalho e pela confiança depositada em mim. Além de ser um exemplo de profissional, ela teve um papel fundamental na minha formação acadêmica.

Ao meu orientador, Prof. Dr. Luiz Stragevitch, por todas as contribuições e discussões, que foram fundamentais.

Ao meu coorientador Prof. Dr. Neirivaldo Cavalcante, por todo o apoio e pelas valiosas contribuições para o desenvolvimento desse trabalho, desde a parte experimental até os ensinamentos teóricos.

Ao Prof. Dr. Frans Van den Berg, pela valiosa contribuição dada a esse trabalho durante o período sanduíche na Universidade de Copenhague, Dinamarca, e ao Prof. Dr. Jose Manuel Amigo, por possibilitar o doutorado sanduíche e me orientar durante esse período.

Ao Prof. Dr. Jarbas Rohewedder, pela confecção do acessório utilizado para as medidas de transmitância e pelas discussões.

A Lucas Del Prette e Priscilla Figueirêdo, por todo o suporte e pelas discussões relacionadas aos dados coletados da empresa.

Ao aluno de iniciação científica, Higor França, pelo auxílio no desenvolvimento experimental desse trabalho, ao Dr. Josivan Pedro da Silva, pela ajuda na operação da coluna de destilação reativa e aos professores Dr. José Geraldo Pacheco Filho e Dr. José Marcos da Silva, por permitirem a utilização da coluna.

Ao CETENE, por ter fornecido biodiesel e óleo de algodão, e a Marcélio Alves, por ter fornecido biodiesel.

A todos os amigos e colegas de trabalho do grupo de pesquisa do Laboratório de Combustíveis (LAC) da UFPE, por terem me auxiliado em inúmeras situações e me proporcionado diversos momentos de descontração, tornando a rotina sempre mais leve.

Aos integrantes de todos os setores do LAC, que de alguma forma contribuíram para o desenvolvimento deste trabalho e tornaram o ambiente de trabalho descontraído e acolhedor.

Ao Programa de Pós-graduação em Engenharia Química da UFPE.

A FACEPE, pela concessão da bolsa de doutorado, ao CNPq, pela concessão da bolsa de doutorado sanduíche e a NUQAAPE e ao INCTAA, por todos os auxílios concedidos.

Por fim, a toda minha família, que sempre me apoia e me transmite amor e cuidado, em especial aos meus avós queridos, aos meus pais e à minha irmã. Ao meu noivo, que preenche minha vida com leveza, alegria e amor. Aos meus amigos, que me apoiam e torcem pela minha felicidade.

## RESUMO

À medida que os processos industriais se tornam mais complexos, torna-se essencial o desenvolvimento de novas metodologias capazes de monitorá-los de forma eficiente. Nesse contexto, métodos quimiométricos podem ser utilizados para o tratamento de dados provenientes desses processos, visando extrair informações relevantes que auxiliem na sua compreensão e monitoramento. Dentre as ferramentas quimiométricas aplicadas a processos, podem-se destacar as técnicas de calibração multivariada e os métodos de monitoramento estatístico de processos (*Statistical Process Monitoring* - SPM). No presente trabalho, essas ferramentas foram utilizadas em duas abordagens. Na primeira, modelos de regressão por mínimos quadrados parciais (*Partial Least-squares* - PLS) foram desenvolvidos, com base em dados espectroscópicos, para o monitoramento da produção de biodiesel em uma coluna de destilação reativa. Um espectrômetro portátil no infravermelho próximo, o MicroNIR, foi utilizado para o monitoramento *on-line* da mistura reacional na base da coluna. Além disso, misturas sintéticas foram empregadas na construção dos modelos com o intuito de ampliar a faixa de concentração dos componentes modelados: etanol, glicerol e éster alquílico (biodiesel). Resultados satisfatórios foram obtidos em relação aos valores da raiz do erro quadrático médio de predição (*Root Mean Square Error of Prediction* - RMSEP). A segunda abordagem envolveu o desenvolvimento de uma estratégia de SPM para monitorar o processo industrial de metalização a vácuo de filmes de poliéster, com base nos dados de densidade óptica (DO) do filme metalizado adquiridos ao longo da produção. Esses dados representam uma medida indireta da espessura do revestimento e são utilizados no controle do processo. Assim, esse trabalho visou possibilitar uma estratégia de monitoramento eficiente, através de uma análise mais elaborada dos dados de DO já coletados. Após uma análise exploratória desses dados, envolvendo um estudo de correlação e tendência, foi proposta a aplicação de uma carta de controle da média móvel exponencialmente ponderada (*Exponentially weighted moving average* - EWMA) para o monitoramento da tendência. Além disso, após diferenciação dos dados, utilizou-se uma carta EWMA e uma carta da amplitude para o monitoramento da média e da variabilidade do processo, respectivamente. Por fim, a avaliação dos resultados permitiu uma análise coerente dos dados, indicando a viabilidade da estratégia sugerida para o monitoramento do processo.

Palavras-chave: Calibração multivariada. Monitoramento estatístico de processo. Quimiometria de processo. Tecnologia analítica de processo.

## ABSTRACT

As industrial processes become more complex, the development of new methodologies capable of efficiently monitoring them are essential. Chemometric methods can be used in the analysis of the data of these processes to extract relevant information that will facilitate understanding and monitoring of this data. In the present work, two chemometric tools, multivariate calibration and statistical process monitoring (SPM) were used for two different approaches. In the first one, partial least squares (PLS) regression models were developed, based on spectroscopic data, for monitoring biodiesel production in a reactive distillation column. A portable spectrometer in the near infrared region was used for the online monitoring of the reaction mixture at the bottom of the column. In addition, synthetic blends were used in the development of the models in order to increase the concentration range of the components: ethanol, glycerol and alkyl ester (biodiesel). Satisfactory results were obtained in relation to the root mean square error of prediction (RMSEP) values. The second approach involved the development of a SPM strategy to monitor the industrial vacuum metallization process of polyester films, based on the optical density (DO) data of the metallized film acquired during production. These data represent an indirect measure of the coating thickness and are used in the process control. The aim of this present work was to suggest an efficient monitoring strategy through a further elaboration analysis of the DO data already collected. After an exploratory data analysis, involving correlation and trend investigations, the application of an exponentially weighted moving average (EWMA) control chart was considered for trend monitoring. In addition, an EWMA chart and a range control chart on the differenced data were used to monitor process mean and variability, respectively. Finally, the evaluation of the results enabled a coherent analysis of the data, indicating the viability of the suggested strategy for monitoring the process.

**Keywords:** Multivariate calibration. Process analytical technology. Process chemometrics. Statistical process monitoring.

## LISTA DE FIGURAS

Figura 1 – Esquema simplificado do funcionamento do MicroNIR, apresentando a dispersão da radiação incidente por parte do Filtro Linear Variável e a detecção dessa radiação pelo arranjo de fotodiodos, para aquisição espectral na região entre 908 a 1676 nm (a) e espectrômetro portátil MicroNIR Pro (b). ....	24
Figura 2 – Representação de um conjunto de dados aleatórios descritos originalmente pelas variáveis 1 e 2, utilizando as componentes principais PC1 e PC2. As retas tracejadas representam a projeção das amostras nos eixos das PC's (escores), enquanto as curvas tracejadas representam os ângulos entre a PC e a variável original ( <i>loadings</i> ). ....	28
Figura 3 – Comparação entre as abordagens de monitoramento estatístico univariada (limites de controle em vermelho – linha contínua) (a) e multivariada (elipse de controle em vermelho – linha tracejada) (b) para o monitoramento de duas variáveis aleatórias dependentes.....	43
Figura 4 – Comparação entre o cálculo das cartas de controle de Shewhart, EWMA e CUSUM, do ponto de vista do peso aplicado às amostras presente e passadas. ....	46
Figura 5 – Esquema da coluna de destilação reativa e do monitoramento <i>on-line</i> com o MicroNIR.....	58
Figura 6 – Espectros adquiridos utilizando cubetas de quartzo de diferentes caminhos ópticos.....	61
Figura 7 – Espectros médios, pré-processados com SNV, de misturas sintéticas de mesma composição (97 e 98 – APÊNDICE C), adquiridos, respectivamente, na temperatura ambiente ( $22 \pm 2^\circ\text{C}$ ) (em azul) e na faixa de temperatura entre 50 e 60°C (em vermelho).....	64
Figura 8 – Espectros médios das 108 misturas sintéticas preparadas (em cinza) e das 12 amostras retiradas da coluna de destilação reativa (em preto), antes da exclusão dos outliers.....	65
Figura 9 – Resíduos studentizados vs. <i>leverage</i> para as amostras de calibração (em cinza) e de predição (misturas sintéticas em azul e amostras da coluna em vermelho) para o modelo para quantificação de éster alquílico construído com os dados pré-processados com EMSC. ....	68

Figura 10 – Espectros médios das misturas sintéticas (em cinza) e das amostras da coluna (em preto) pré-processados com EMSC, após a exclusão de outliers.....	70
Figura 11 – Gráficos dos valores preditos vs. referência (a, c e e) para o conjunto de calibração (em cinza) e o conjunto de predição, composto de misturas sintéticas (em azul) e amostras da coluna (em vermelho); e os VIPs escores (b, d e f) dos modelos PLS para quantificar etanol (a e b), éster alquílico (c e d) e glicerol (e e f).....	71
Figura 12 – Evaporadores dispostos de forma alternada dentro da metalizadora.....	80
Figura 13 – Representação esquemática simplificada do processo de metalização de um rolo de filme de poliéster, da aquisição dos dados de DO e da geração da matriz para cada corrida.....	81
Figura 14 – Perfil da velocidade do filme (a) e perfis de DO adquiridos ao longo dos 18 evaporadores (b) para a corrida A2.....	86
Figura 15 – Mapas de correlação para as corridas A1 a A8. As escalas de cores estão relacionadas com os coeficientes de correlação entre as variáveis.....	87
Figura 16 – ACF das variáveis 1 (em azul) e 2 (em preto) e CCF dessas variáveis (em magenta), para a corrida A2 (a); e ACF das variáveis 1 (em azul) e 15 (em verde) e correlação cruzada dessas variáveis (em magenta), para a mesma corrida (b). Em vermelho, limites com 99% de confiança para ACF e CCF.....	89
Figura 17 – ACF das variáveis 1 (em azul) e 2 (em preto) e CCF dessas variáveis (em magenta), para a corrida A5 (a); e ACF das variáveis 1 (em azul) e 15 (em verde) e correlação cruzada dessas variáveis (em magenta), para a mesma corrida (b). Em vermelho, limites com 99% de confiança para ACF e CCF.....	90
Figura 18 – Cartas de controle EWMA para o monitoramento da tendência dos dados (os símbolos conectados por linhas finas representam os dados originais enquanto as linhas grossas mostram a tendência) para os barcos de evaporação 1, 2 e 15 das corridas A2 (a) e A5 (b).....	91
Figura 19 – Carta de controle EWMA nos dados diferenciados para o barco de evaporação 1 da corrida A2 (a) e o barco 15 da corrida A5 (b); e carta de controle da amplitude (R) nos dados diferenciados para o barco de evaporação 1 da corrida A2 (c) e o barco 15 da corrida A5 (d). As linhas contínuas em vermelho representam os limites superior e inferior (LSC / LIC) de controle e a linha tracejada em vermelho o limite central (LC) para os gráficos de controle.....	93

Figura 20 – Mapas de correlação para as corridas B1 a B12. As escalas de cores estão relacionadas com os coeficientes de correlação entre as variáveis.....	95
Figura 21 – Estratégia de monitoramento proposta para a corrida B2: carta EWMA para monitoramento da tendência dos dados (símbolos conectados por uma linha fina representam os dados originais enquanto a linha grossa mostra a tendência) (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 1.....	96
Figura 22 – Valores de CPk para cada barco de evaporação da corrida A2. A linha tracejada é o limite unitário para o CPk.....	97
Figura 23 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 4 da corrida A2.....	98
Figura 24 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 5 da corrida A2.....	99
Figura 25 – Valores de CPk para cada barco de evaporação da corrida A7. A linha tracejada é o limite unitário para o CPk. ....	100
Figura 26 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 3 da corrida A7.....	100
Figura 27 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 1 da corrida A7.....	101

## **LISTA DE TABELAS**

Tabela 1 – Número de misturas sintéticas e amostras da coluna aplicadas às etapas de calibração e predição para o desenvolvimento de modelos PLS para quantificar etanol, éster alquílico e glicerol, antes da exclusão de <i>outliers</i> .....	59
Tabela 2 – Teores de etanol e éster alquílico, em %(m/m), das amostras retiradas da coluna, obtidos por cromatografia gasosa. ....	62
Tabela 3 – Resultados para as etapas de calibração e de predição dos modelos de regressão PLS para quantificação dos teores de etanol, éster alquílico e glicerol .....	66
Tabela 4 – Resultados para a predição das misturas sintéticas e amostras da coluna para os modelos PLS desenvolvidos com EMSC, para quantificação de etanol, éster alquílico e glicerol.....	73
Tabela 5 – Comparaçao entre os resultados dos rendimentos de éster alquílico das amostras da coluna estimados a parti dos modelos PLS e adquiridos por cromatografia.....	75

## LISTA DE ABREVIATURAS E SIGLAS

ANP	Agência Nacional de Petróleo, Gás Natural e Biocombustíveis
ASTM	<i>American Society for Testing and Materials</i>
CEP	Controle Estatístico de Processo
CUSUM	Soma cumulativa ( <i>Cumulative Sum</i> )
DOE	<i>Design of Experiment</i>
EWMA	Média móvel exponencialmente ponderada ( <i>Exponentially Weighted Moving Average</i> )
EMSC	Correção de Espalhamento Multiplicativo Estendido ( <i>Extended Multiplicative Scatter Correction</i> )
FDA	<i>Food and Drug Administration</i>
FIR	Infravermelho distante ( <i>Far Infrared</i> )
FT-NIR	Infravermelho Próximo por Transformada de Fourier
LIC	Limite inferior de controle
LIE	Limite inferior de especificação
LSC	Limite superior de controle
LSE	Limite superior de especificação
LVF	Filtro Linear Variável ( <i>Linear Variable Filter</i> )
MIR	Infravermelho Médio ( <i>Middle Infrared</i> )
MSC	Correção de Espelhamento Multiplicativo ( <i>Multiplicative Scatter Correction</i> )
MSPM	Monitoramento estatístico multivariado de processo ( <i>Multivariate Statistical Process Monitoring</i> )
NIR	Infravermelho Próximo ( <i>Near Infrared</i> )
OSC	Correção do sinal ortogonal ( <i>Orthogonal Signal Correction</i> )
PAT	Tecnologia Analítica de Processo ( <i>Process Analytical Technology</i> )
PC	Componente Principal ( <i>Principal Component</i> )
PC1	Primeira componente principal
PC2	Segunda componente principal
PCA	Análise de Componentes Principais ( <i>Principal Component Analysis</i> )
PLS	Mínimos Quadrados Parciais ( <i>Partial Least-squares</i> )

RMSE	Raiz do Erro Quadrático Médio ( <i>Root Mean Square Error</i> )
RMSEC	Raiz do Erro Quadrático Médio de Calibração ( <i>Root Mean Square Error of Calibration</i> )
RMSECV	Raiz do Erro Quadrático Médio de Validação Cruzada ( <i>Root Mean Square Error of Cross-validation</i> )
RMSEP	Raiz do Erro Quadrático Médio de Predição ( <i>Root Mean Square Error of Prediction</i> )
SPM	Monitoramento estatístico de processo ( <i>Statistical Process Monitoring</i> )
SPXY	<i>Sample Set Partitioning based on Joint x-y Distances</i>
VIP	Importância das variáveis na projeção ( <i>Variable Importance in Projection</i> )
VL	Variáveis latentes

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO .....</b>	<b>17</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA.....</b>	<b>19</b>
2.1	TECNOLOGIA ANALÍTICA DE PROCESSO .....	19
2.2	ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO .....	20
2.2.1	<b>Espectroscopia no infravermelho próximo como uma ferramenta da tecnologia analítica de processo.....</b>	<b>22</b>
2.2.2	<b>Miniaturização dos equipamentos de infravermelho próximo .....</b>	<b>23</b>
2.3	QUIMIOMETRIA.....	25
2.3.1	<b>Análise de Componentes Principais.....</b>	<b>26</b>
2.3.2	<b>Técnicas de pré-processamento aplicadas a dados espectroscópicos .....</b>	<b>28</b>
2.3.2.1	Pré-processamento de variáveis .....	29
2.3.2.2	Pré-processamento de amostras.....	30
2.3.3	<b>Aplicação das ferramentas quimiométricas a processos industriais .....</b>	<b>32</b>
2.4	CALIBRAÇÃO MULTIVARIADA .....	34
2.4.1	<b>Método de seleção de amostras .....</b>	<b>34</b>
2.4.2	<b>Rregressão por mínimos quadrados parciais .....</b>	<b>35</b>
2.4.3	<b>Avaliação dos modelos de calibração.....</b>	<b>38</b>
2.5	MONITORAMENTO ESTATÍSTICO DE PROCESSOS .....	40
2.5.1	<b>Gráficos de controle univariados .....</b>	<b>44</b>
2.5.2	<b>Gráficos de controle para processos autocorrelacionados .....</b>	<b>48</b>
3	<b>MONITORAMENTO DA PRODUÇÃO DE BIODIESEL POR ROTA</b>	
	<b>ALTERNATIVA UTILIZANDO UM ESPECTRÔMETRO PORTÁTIL NO INFRAVERMELHO PRÓXIMO.....</b>	<b>50</b>
3.1	METODOLOGIA .....	53
3.1.1	<b>Produção de biodiesel pelo processo de destilação reativa .....</b>	<b>53</b>
3.1.2	<b>Análises cromatográficas .....</b>	<b>54</b>
3.1.3	<b>Aquisição <i>off-line</i> de dados espectroscópicos com o MicroNIR e seleção do caminho óptico.....</b>	<b>55</b>
3.1.4	<b>Misturas sintéticas.....</b>	<b>56</b>
3.1.5	<b>Monitoramento <i>on-line</i> da produção de biodiesel com o MicroNIR .....</b>	<b>57</b>
3.1.6	<b>Pré-processamento dos dados e desenvolvimento dos modelos de calibração....</b>	<b>58</b>

3.2	RESULTADOS E DISCUSSÃO .....	60
3.2.1	Seleção do caminho óptico e estimativa da composição da mistura reacional no ponto monitorado com o MicroNIR .....	60
3.2.2	Análises cromatográficas das amostras da coluna .....	62
3.2.3	Misturas sintéticas .....	62
3.2.4	Pré-processamento dos dados e desenvolvimento dos modelos de calibração....	65
3.2.5	Avaliação dos modelos quanto à predição das amostras da coluna .....	72
3.2.6	Estimativa do rendimento de biodiesel das amostras da coluna.....	74
3.3	CONCLUSÃO.....	75
4	<b>MONITORAMENTO ESTATÍSTICO DO PROCESSO INDUSTRIAL DE METALIZAÇÃO A VACUO DE FILMES DE POLIÉSTER .....</b>	<b>77</b>
4.1	METODOLOGIA .....	79
4.1.1	Processo de metalização a vácuo e aquisição dos dados de densidade óptica....	79
4.1.2	Análise exploratória dos dados de densidade óptica.....	81
4.1.3	Estratégia de monitoramento estatístico .....	83
4.1.4	Avaliação da estratégia de monitoramento estatístico proposta.....	84
4.1.5	Análise da capacidade do processo .....	84
4.2	RESULTADOS E DISCUSSÃO .....	85
4.2.1	Análise exploratória dos dados de densidade óptica.....	85
4.2.2	Estratégia de monitoramento estatístico .....	90
4.2.3	Avaliação da estratégia de monitoramento proposta.....	94
4.2.4	Análise da capacidade do processo .....	97
4.3	CONCLUSÃO.....	102
5	<b>PERSPECTIVAS.....</b>	<b>104</b>
	<b>REFERÊNCIAS .....</b>	<b>105</b>
	<b>APÊNDICE A – ARTIGO PUBLICADO EM REVISTA CIENTÍFICA.....</b>	<b>113</b>
	<b>APÊNDICE B – DIAGRAMA TERNÁRIO PROVENIENTE DO PLANEJAMENTO DE MISTURAS COM PONTOS ADICIONAIS .....</b>	<b>114</b>
	<b>APÊNDICE C – MISTURAS SINTÉTICAS PREPARADAS PARA O DESENVOLVIMENTO DOS MODELOS DE CALIBRAÇÃO.....</b>	<b>115</b>
	<b>APÊNDICE D – DESCRIÇÃO DOS CÁLCULOS DOS VALORES DE REFERÊNCIA PARA O ÉSTER ALQUÍLICO E O GLICEROL.....</b>	<b>118</b>

<b>APÊNDICE E – MONITORAMENTO DA TENDÊNCIA DOS DADOS DE DENSIDADE ÓPTICA PARA A CORRIDA A2.....</b>	<b>120</b>
<b>APÊNDICE F – COMPARAÇÃO ENTRE OS DADOS ORIGINAIS E DIFERENCIADOS .....</b>	<b>121</b>
<b>APÊNDICE G – MONITORAMENTO DA VARIABILIDADE PARA A CORRIDA A2.....</b>	<b>122</b>

## 1 INTRODUÇÃO

Nas últimas décadas, o rápido avanço tecnológico tem impulsionado o desenvolvimento de sensores analíticos e sistemas computacionais. A associação desses elementos empregados na produção industrial é um dos fatores que tem conduzido à nova geração dos processos industriais de manufatura, amplamente denominada Indústria 4.0. Esta é caracterizada, principalmente, por processos complexos, cujo monitoramento gera uma enorme quantidade de dados, adquiridos em uma elevada frequência e provenientes de fontes diversas (BAKDI; KOUADRI, 2017; HE; WANG, 2018; SZYMAŃSKA, 2018).

Esse cenário, além de proporcionar oportunidades de crescimento e desenvolvimento, também envolve alguns aspectos desafiadores, que estão presentes desde a etapa de projeto até o controle e monitoramento do processo. Adicionalmente, existem ainda as crescentes exigências para a manutenção da segurança, sustentabilidade, qualidade do produto e capacidade de operação. Dessa forma, esse contexto atual tem requerido a utilização de ferramentas poderosas de análise de dados em tempo real, com o objetivo de dar agilidade ao processo de tomada de decisões e ao desenvolvimento de soluções eficientes diante de emergências no processo produtivo (BAKDI; KOUADRI, 2017; SHANG; YOU, 2019).

É nesse contexto que está inserida a ciência de dados, que pode ser entendida como um campo interdisciplinar da ciência em que métodos estatísticos e matemáticos são associados a sistemas computacionais e da informação com o objetivo de evidenciar informações relevantes e novos entendimentos sobre os dados. Diversas ferramentas podem ser aplicadas com esse objetivo. A quimiometria, por exemplo, é considerada uma parte importante dessa ciência (SZYMAŃSKA, 2018).

Usualmente, a quimiometria é definida como uma parte da química utilizada para extrair informações relevantes sobre um sistema químico, através da aplicação de métodos matemáticos e estatísticos. Entretanto, já faz um tempo que as técnicas quimiométricas têm ampliado o seu domínio de aplicação. Dessa forma, principalmente devido às inúmeras aplicações a processos industriais, a quimiometria se tornou também assunto de interesse da engenharia química e de processos (WISE; KOWALSKI, 1995).

A aplicação das ferramentas quimiométricas a processos industriais é bastante abrangente, incluindo, por exemplo, os métodos de planejamento experimental (*Design of Experiment* - DOE), as técnicas de calibração e classificação multivariada e o monitoramento estatístico do processo (FERREIRA; TOBYN, 2014; GRASSI; ALAMPRESE, 2018;

SHANG; YOU, 2019; VIGNADUZZO; MAGGIO; OLIVIERI, 2020).

Considerando-se a importância dessas ferramentas de análise de dados para evidenciar informações relevantes que auxiliem na compreensão e no monitoramento de processos, no presente trabalho, técnicas quimiométricas foram empregadas no tratamento de dados coletados durante dois processos diferentes. Na primeira abordagem, um espectrômetro portátil na região do infravermelho próximo foi utilizado para o monitoramento *on-line* da produção de biodiesel em uma coluna de destilação reativa. Com base nos dados espectroscópicos coletados durante o processo, modelos de regressão por mínimos quadrados parciais (*Partial Least-squares* - PLS) foram desenvolvidos para estimar os teores de etanol, glicerol e éster alquílico (biodiesel) na base da coluna. Já na segunda abordagem, técnicas de monitoramento estatístico de processos foram utilizadas com o objetivo de sugerir uma estratégia de monitoramento para o processo industrial de metalização a vácuo de filmes de poliéster. Nessa abordagem, gráficos de controle foram desenvolvidos com base na densidade óptica do filme metalizado adquirido durante a produção, a qual representa uma medida indireta da espessura do revestimento aplicado ao filme polimérico. Por tratarem de processos distintos, cada abordagem e os seus respectivos objetivos específicos estão contemplados em um item específico desta tese.

## 2 FUNDAMENTAÇÃO TEÓRICA

A seguir, estão inseridos alguns tópicos que fornecem embasamento teórico aos temas abordados nesta tese.

### 2.1 TECNOLOGIA ANALÍTICA DE PROCESSO

De modo geral, a compreensão sobre determinado processo produtivo é essencial para a implementação de um controle eficiente e, por conseguinte, para a redução de sua variabilidade. Nesse contexto, tem-se buscado desenvolver novas metodologias capazes de ampliar o entendimento sobre os processos industriais modernos, os quais são, em geral, complexos e multivariados. A partir dessa necessidade, surgiu a Tecnologia Analítica de Processo (*Process Analytical Technology - PAT*), um tema que tem sido amplamente estudado e aplicado, principalmente, nas indústrias farmacêuticas e de química fina, que exigem um maior rigor em relação ao controle da produção (SIMON et al., 2015; GROBHANS et al., 2018; THAKUR; HEBBI; RATHORE, 2020; VIGNADUZZO; MAGGIO; OLIVIERI, 2020).

A agência *Food and Drug Administration* (FDA), do Departamento de Saúde e Serviços Humanos dos Estados Unidos, em um guia direcionado à indústria farmacêutica, definiu PAT como um sistema voltado para o projeto, a análise e o controle do processo, por meio da medição de parâmetros de processo e atributos de qualidade em diversas etapas da produção, durante o processamento (FDA, 2004). Pode-se identificar quatro aspectos essenciais que constituem esse sistema: a compreensão sobre os principais fatores que afetam a dinâmica do processo e a qualidade do produto final; a análise do processo; a análise dos dados multivariados coletados e o controle do processo (GRASSI; ALAMPRESE, 2018).

Existem diversas ferramentas que podem ser utilizadas a fim de satisfazer aos principais objetivos de estratégias PAT, dentre as quais se destacam os dispositivos para aquisição e análise de dados multivariados, as ferramentas de controle e monitoramento de processos, os modernos analisadores de processos e as técnicas aplicadas à gestão do conhecimento e à melhoria contínua. Essas ferramentas podem ser empregadas sozinhas ou em conjunto, em uma etapa do processo produtivo ou em toda a produção, visando à obtenção de informações que permitam ampliar a compreensão sobre o processo, o gerenciamento dos riscos e a melhoria contínua (FDA, 2004).

Por fim, faz-se necessário ressaltar que o tema PAT é bastante abrangente e tem natureza multidisciplinar, incluindo não apenas as análises químicas, como também as análises físicas, microbiológicas, matemáticas e de risco. Assim, especialistas na área de PAT podem vir de diversas formações, principalmente das áreas de engenharia e instrumentação. De modo geral, algumas aplicações e desenvolvimentos recentes na área de PAT incluem: monitoramento e controle de processos (cristalização, fermentação, destilação, dentre outros), desenvolvimento de sensores analíticos, *scale-up* de laboratório para planta piloto, utilização de técnicas espectroscópicas como ferramentas de PAT (Raman, infravermelho, terahertz) e emprego de métodos quimiométricos (DE BEER et al., 2011; SIMON et al., 2015; GRASSI; ALAMPRESE, 2018).

## 2.2 ESPECTROSCOPIA NO INFRAVERMELHO PRÓXIMO

A espectroscopia no infravermelho é um tipo de espectroscopia vibracional considerada uma importante tecnologia para as determinações quantitativas e qualitativas de compostos moleculares nos estados sólido, líquido e gasoso. O infravermelho está associado à região do espectro eletromagnético em que o comprimento de onda varia de 780 a 1.000.000 nm, o que corresponde a uma variação de número de onda de 12.800 a 10  $\text{cm}^{-1}$ . Considerando-se aspectos relacionados à aplicação e à instrumentação, essa região é geralmente subdividida em infravermelho próximo (NIR - *Near Infrared*), médio (MIR - *Mid Infrared*), e distante (FIR - *Far Infrared*). A faixa do NIR corresponde à região espectral mais energética, em que o comprimento de onda varia entre 780 e 2.500 nm. Já a faixa do FIR, corresponde à região entre 50.000 e 1.000.000 nm. A faixa do MIR, por sua vez, é a região intermediária do infravermelho, que vai de 2.500 a 50.000 nm (SKOOG; HOLLER; CROUCH, 2009).

De um modo geral, quando uma radiação incide sobre um composto molecular, este absorve apenas determinadas frequências dessa radiação. No processo de absorção de energia correspondente ao infravermelho médio, apenas as frequências da radiação incidente que coincidem com as frequências de vibração naturais da molécula são absorvidas. A absorção está condicionada, no entanto, à variação do momento de dipolo das ligações da molécula durante os movimentos de vibração ou de rotação. Dessa forma, apenas as ligações que apresentam um dipolo elétrico que varia na mesma frequência da radiação infravermelha

incidente são capazes de absorvê-la. Moléculas simétricas, por exemplo, são inativas à radiação em toda a região do espectro infravermelho (PAVIA et al., 2009).

Como consequência da absorção da energia na região do infravermelho, a molécula passa do estado fundamental para o estado excitado, sendo essa transição energética capaz de proporcionar um aumento na amplitude dos seus movimentos vibracionais. Os principais modos vibracionais das moléculas que absorvem radiação infravermelha são a deformação axial e a deformação angular. A deformação axial, também denominada estiramento, é caracterizada por alterações na distância entre dois átomos ao longo do eixo da ligação. Já a deformação angular, envolve variações no ângulo formado por duas ligações. Existem diversos modos de deformação angular, os quais podem ocorrer dentro de um mesmo plano ou para fora dele (PAVIA et al., 2009; SKOOG; HOLLER; CROUCH, 2009).

Em particular, as absorções na região do infravermelho próximo (NIR) estão associadas, principalmente, aos sobretons e às bandas de combinação das vibrações fundamentais das ligações C—H, N—H, O—H e S—H. Assim, as principais aplicações dessa técnica estão relacionadas às determinações quantitativas e qualitativas de espécies que apresentam grupos funcionais contendo essas ligações (SKOOG; HOLLER; CROUCH, 2009; SIMPSON, 2010).

O crescente uso da espectroscopia NIR, incluindo em aplicações industriais, está associado, principalmente, às suas características fundamentais já bem estabelecidas, dentre as quais, pode-se citar: rapidez na aquisição espectral, elevada precisão e, na maioria dos casos, ausência de preparo de amostra e manutenção de sua integridade. Adicionalmente, os avanços na área da instrumentação e da quimiometria são fatores que têm impulsionado o rápido desenvolvimento da tecnologia NIR. Em relação à instrumentação, o progresso nessa área tem proporcionado melhorias na aquisição espectral e no tratamento de dados espectroscópicos. Esses avanços são orientados, principalmente, pela miniaturização dos equipamentos e pelos sistemas de aquisição de imagens espetrais. Já no que diz respeito aos métodos quimiométricos, estes são essenciais para a obtenção de informações relevantes a partir dos dados espectroscópicos, principalmente devido à complexa natureza dessa região espectral, caracterizada normalmente por bandas alargadas e sobrepostas, o que dificulta a sua interpretação e a atribuição das bandas de absorção (PASQUINI, 2018).

### **2.2.1 Espectroscopia no infravermelho próximo como uma ferramenta da tecnologia analítica de processo**

Em comparação com outras técnicas analíticas, na última década, a espectroscopia NIR tem se destacado para aplicações na área de PAT, tendo sido empregada em diversos setores produtivos, especialmente, nas indústrias farmacêuticas (PASQUINI, 2018). Além das vantagens anteriormente citadas relacionadas ao uso dessa técnica para aplicações diversas, existem alguns fatores que a tornam particularmente atrativa para utilização como uma ferramenta de PAT. Um desses fatores, por exemplo, está associado à variedade da instrumentação na área do NIR, que permite que essa tecnologia atenda a diversas demandas industriais. Assim, o instrumento de NIR pode ser selecionado de acordo com um critério considerado crítico para a aplicação, o qual pode ser, por exemplo, elevada precisão, rapidez na aquisição dos dados ou exatidão do comprimento de onda (SIMPSON, 2010).

Adicionalmente, ainda do ponto de vista da instrumentação, outro fator que tem impulsionado a utilização do NIR para aplicações industriais consiste no fato de os instrumentos NIR serem, em geral, robustos e compactos. Além disso, possibilitam que a aquisição espectral da amostra seja obtida por meio de sondas conectadas ao equipamento através de cabos de fibra óptica ou por meio de instrumentos ultracompactos e portáteis, o que simplifica bastante a utilização dessa tecnologia em linhas de produção complexas (GRASSI; ALAMPRESE, 2018; MIRSCHEL et al., 2018; PUIG-BERTOTTO; COELLO; MASPOCH, 2019).

Outro aspecto do emprego da tecnologia NIR na área de PAT está relacionado à baixa absorvidade molar dos sobretons e das bandas de combinação, quando comparada a das vibrações fundamentais. Apesar de torná-la uma técnica não muito sensível, apresentando geralmente um limite de detecção em torno de 0,1%, essa baixa absorvidade é muitas vezes considerada vantajosa do ponto de vista de PAT, uma vez que permite que as amostras sejam analisadas sem a necessidade de diluição prévia ou utilização de técnicas para atenuar a radiação incidente. Como resultado, a espectroscopia NIR pode ser empregada para medição direta das amostras, atendendo a um dos principais requisitos da maioria das aplicações na área de PAT (SIMPSON, 2010).

Dessa forma, a possibilidade da medição direta das amostras tem impulsionado o uso do NIR para o monitoramento *in-line* e *on-line* de diversos processos produtivos, visando aplicações, por exemplo, em indústrias farmacêuticas (DE BEER et al., 2011; HOHL et al.,

2017), alimentícias (GRASSI; ALAMPRESE, 2018) e de biocombustíveis (DE LIMA et al., 2014; GELINSKI et al., 2018). Além de ser empregado, inclusive, para determinações quantitativas e qualitativas dos produtos intermediários e finais da produção (GRASSI; ALAMPRESE, 2018; PUIG-BERTOTTO; COELLO; MASPOCH, 2019).

Outro benefício do emprego da espectroscopia NIR em aplicações na área de PAT está relacionado à capacidade de penetração da radiação devido aos comprimentos de onda mais curtos dessa faixa espectral, quando comparados aos de outras técnicas espectroscópicas vibracionais. Ao utilizar essa técnica para a análise de alguns sólidos por reflectância difusa, o feixe da radiação é capaz de penetrá-los com profundidade na ordem de alguns milímetros. Essa característica do NIR permite utilizá-lo para análises não destrutivas, o que pode ser útil, por exemplo, para a avaliação da qualidade de diversos materiais comercializados e utilizados como matéria-prima na produção industrial, evitando-se, assim, problemas de contaminação e heterogeneidade das amostras (SIMPSON, 2010). Por exemplo, essa característica do NIR tem sido avaliada para o monitoramento da qualidade de frutas e verduras ( MAGWAZA et al., 2012; MARQUES et al., 2016; EISENSTECKEN et al., 2019).

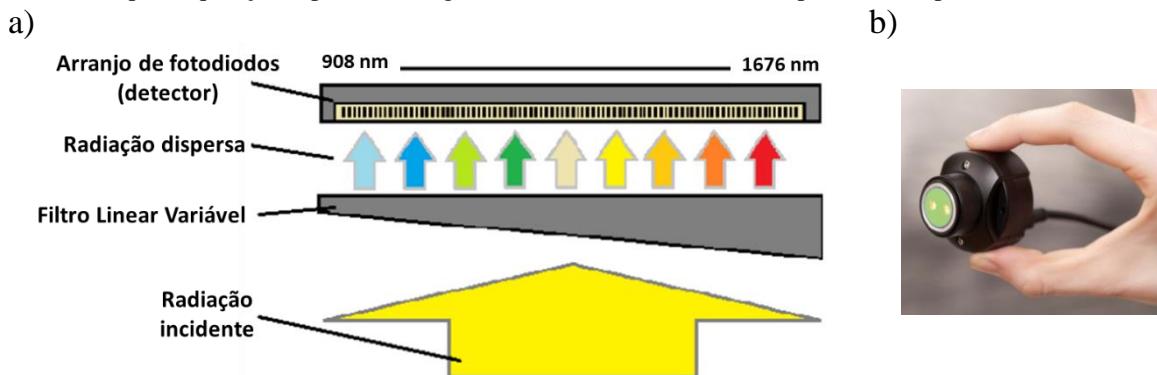
### **2.2.2 Miniaturização dos equipamentos de infravermelho próximo**

Conforme mencionado, um dos principais aspectos relacionados à ascensão da tecnologia NIR está associado aos avanços na área de instrumentação, sendo a tendência à miniaturização um dos fatores mais relevantes. Nesse contexto, na última década, um número crescente de equipamentos portáteis na região do NIR tem sido desenvolvido e comercializado. Pasquini (2018), por exemplo, listou em torno de vinte instrumentos comerciais compactos, alguns dos quais são extremamente leves, pequenos e baratos. Segundo o autor, essa tendência à miniaturização tem transformado a espectroscopia NIR por permitir análises *in loco*.

Um dos equipamentos portáteis disponíveis no mercado é o MicroNIR, produzido pela VIAVI. O MicroNIR abrange uma família de espetrômetros ultracompactos que incluem desde os instrumentos portáteis para análise em campo, como o MicroNIR Pro, utilizado no desenvolvimento desse trabalho, até os instrumentos diretamente voltados para as aplicações na área de PAT, os quais podem ser adaptados para uso em reatores ou tubulações (VIAVI, 2020). O tamanho reduzido desses instrumentos é devido, principalmente, à utilização de um filtro linear variável (*Linear Variable Filter - LVF*) como elemento dispersor. Trata-se de um

filme fino de espessura variável responsável por dispersar a radiação incidente. Dessa forma, cada pixel do arranjo de fotodiodos, utilizado como detector, responde a um comprimento de onda específico da radiação dispersa, permitindo que os espectros sejam adquiridos na região entre 908 e 1676 nm, conforme ilustrado na Figura 1a. Por fim, esse mecanismo de aquisição espectral está instalado em um equipamento ultracompacto (Figura 1b) sem partes móveis, o que confere robustez ao MicroNIR (O'BRIEN et al., 2012; PEDERSON et al., 2014).

Figura 1 – Esquema simplificado do funcionamento do MicroNIR, apresentando a dispersão da radiação incidente por parte do Filtro Linear Variável e a detecção dessa radiação pelo arranjo de fotodiodos, para aquisição espectral na região entre 908 a 1676 nm (a) e espectrômetro portátil MicroNIR Pro (b)



Fonte: Pederson *et al.* (2014)

As principais vantagens associadas à utilização desse tipo de sensor analítico estão relacionadas ao tamanho, ao peso, à robustez e ao reduzido custo de manufatura desses equipamentos. Além disso, por permitir análises em tempo real e *in loco*, tem-se avaliado o potencial dos espectrômetros comerciais portáteis para utilização em diversos campos de atuação (incluindo aplicações visando ao emprego em processos industriais) obtendo-se, em geral, bons resultados (MODROÑO et al., 2017; GALAVERNA et al., 2018; PASQUINI, 2018; PUIG-BERTOTTO; COELLO; MASPOCH, 2019; DA SILVA et al., 2019).

Apesar dos benefícios proporcionados pelos instrumentos miniaturizados na região do NIR, é importante ressaltar alguns desafios associados ao emprego desses equipamentos. Como exemplo, pode-se citar a sua resolução mais baixa e sua faixa espectral reduzida, quando comparadas as de um equipamento de bancada. Além disso, conforme Pasquini (2018), a área de medição reduzida desses instrumentos pode representar uma desvantagem por dificultar a obtenção de espectros representativos, especialmente para amostras não homogêneas. Esse problema, no entanto, pode ser contornado com o emprego de um sistema de amostragem e de análise apropriados. Por fim, de acordo com o mesmo autor, os modelos desenvolvidos utilizando-se instrumentos portáteis podem apresentar erros maiores quando

comparados aos métodos de referência e aos equipamentos de bancada. A escolha em utilizá-los, então, deve ser decidida não apenas nos resultados dos modelos, mas também com base na adequação ao uso. Assim, apesar de proporcionarem uma menor precisão, dependendo da aplicação, a utilização desses equipamentos pode ser bastante vantajosa, dentre outros aspectos, por permitem a obtenção de medidas em tempo real por um custo relativamente baixo para os usuários.

## 2.3 QUIMIOMETRIA

Conforme mencionado anteriormente, assim como os avanços na área da instrumentação, o progresso da quimiometria é um dos fatores que tem impulsionado o uso da espectroscopia NIR. O termo quimiometria é frequentemente definido como a utilização de métodos matemáticos e estatísticos para o tratamento de dados multivariados obtidos a partir de um sistema químico, com o objetivo de extrair o máximo de informação significativa sobre esse sistema, auxiliando na sua compreensão. Geralmente, a disciplina que trata da quimiometria está associada à química, mais especificamente à química analítica (WISE; KOWALSKI, 1995).

Há algumas décadas, no entanto, a quimiometria tem expandido o seu campo de aplicação, não se limitando apenas ao domínio da química analítica. Dessa forma, as ferramentas quimiométricas têm sido empregadas em diversas áreas, tanto do ponto de vista da academia quanto da indústria (SHANG; YOU, 2019; VIGNADUZZO; MAGGIO; OLIVIERI, 2020). Por exemplo, conforme revisado por Ferreira e Tobyn (2014), métodos quimiométricos têm sido amplamente utilizados em aplicações nas áreas ambiental, biomédica e genômica, no monitoramento de biorreatores, e em aplicações nas indústrias farmacêuticas, petroquímicas, alimentícias e poliméricas.

Devido às aplicações aos processos industriais, de acordo com Wise e Kowalski (1995), a quimiometria também se tornou matéria de interesse da engenharia. Dentre as principais ferramentas quimiométricas utilizadas nesse tipo de aplicação, empregadas geralmente no contexto de PAT, destacam-se as técnicas de calibração multivariada (FERREIRA; TOBYN, 2014; SIMON et al., 2015), e os métodos de monitoramento estatístico de processos (SPM – *Statistical Process Control*), com ênfase para os gráficos de controle estatístico (SEVERSON; CHAIWATANODOM; BRAATZ, 2016; TIDRIRI et al., 2016; REIS; GINS, 2017; HE; WANG, 2018).

Nesse contexto, o presente tópico irá abordar alguns temas fundamentais da quimiometria, incluindo a Análise de Componentes Principais (PCA – *Principal Component Analysis*), que é um dos métodos quimiométricos mais importantes, e algumas técnicas de pré-processamento de dados espectroscópicos, as quais foram utilizadas no desenvolvimento deste trabalho. Adicionalmente, será apresentada também uma visão geral das ferramentas quimiométricas comumente aplicadas a processos, incluindo algumas aplicações importantes e recentes. Por fim, com o objetivo de se aprofundar nos métodos quimiométricos aplicados neste trabalho, os dois tópicos subsequentes descreverão dois dos principais temas da quimiometria aplicada a processos: a calibração multivariada, com ênfase no método PLS; e o SPM, com a descrição de algumas ferramentas aplicadas nessa área.

### **2.3.1 Análise de Componentes Principais**

A Análise de Componentes Principais (PCA - *Principal Component Analysis*) é um dos mais importantes e versáteis métodos quimiométricos utilizados para o tratamento de dados multivariados e complexos. De modo geral, essa técnica facilita a avaliação desses dados por meio da redução de sua dimensionalidade, destacando as informações relevantes para a sua interpretação e compreensão. Geralmente, a PCA é utilizada no primeiro contato com os dados, através de uma análise exploratória. No entanto, também pode ser empregada para evidenciar as relações entre as variáveis e entre as amostras, para auxiliar na identificação de padrões, bem como para detectar e interpretar os *outliers*, amostras anômalas presentes nos dados (ESBENSEN; GELADI, 2009; BRO; SMILDE, 2014).

A PCA promove a redução da dimensionalidade dos dados multivariados representando-os através de um novo conjunto de fatores, denominados componentes principais (PC - *Principal Component*), os quais são obtidos por meio de combinações lineares das variáveis originais do conjunto de dados. As componentes principais são ortogonais entre si, o que implica que cada uma contribui de forma independente para explicar a variabilidade presente nos dados. Além disso, elas representam de forma decrescente as direções de maior variabilidade. Assim, a primeira PC (denominada PC1) representa a direção de máxima variabilidade dos dados, a segunda PC (PC2) explica a segunda maior variabilidade, e assim sucessivamente (BEEBE, 1998; ESBENSEN; GELADI, 2009; BRO; SMILDE, 2014).

Os dados a serem submetidos à PCA (e a outros métodos quimiométricos de primeira ordem) devem ser organizados em uma matriz ( $\mathbf{X}$ ), em que as  $i$  linhas contêm as informações de cada amostra (ou objeto) e as  $j$  colunas apresentam as informações referentes às variáveis originais do conjunto de dados, de acordo com a Equação 1.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & \cdots & x_{1j} \\ x_{21} & x_{22} & \cdots & \cdots & x_{2j} \\ x_{31} & x_{32} & \cdots & \cdots & x_{3j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{i1} & x_{i2} & \cdots & \cdots & x_{ij} \end{bmatrix} \quad (1)$$

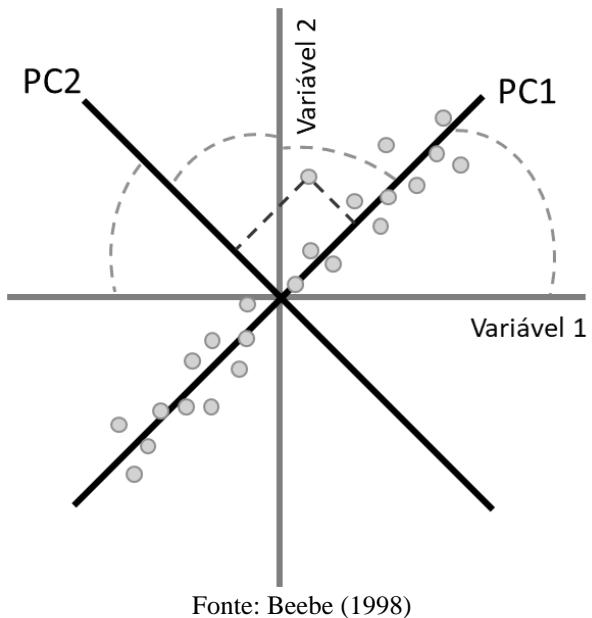
Conforme mencionado, a redução da dimensionalidade dos dados multivariados com a PCA (ou seja, a redução da informação correlacionada e redundante) é possível através da representação desses dados por um conjunto de  $k$  fatores não correlacionados, em que  $k \leq j$ . Matematicamente, essa representação é feita pela decomposição da matriz  $\mathbf{X}$  em dois subconjuntos, de acordo com a Equação 2.

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

Nessa equação, a matriz de escores ( $\mathbf{T}$ ) corresponde às coordenadas das amostras em relação aos eixos das componentes principais e possui dimensão  $i \times k$ . Já a matriz de *loadings* ( $\mathbf{P}$ ), de dimensão  $k \times j$ , representa a relação entre o eixo da PC e o eixo da variável original, sendo calculado pelo cosseno do ângulo entre esses eixos. Por fim, a variabilidade dos dados não explicada pelo modelo é expressa por uma matriz de resíduos ( $\mathbf{E}$ ), que apresenta a mesma dimensão da matriz  $\mathbf{X}$ ,  $i \times j$  (ESBENSEN; GELADI, 2009; BRO; SMILDE, 2014).

Com o objetivo de exemplificar a representação de um conjunto de dados aleatórios utilizando os eixos das componentes principais, um esquema que ilustra a decomposição desses dados através da PCA está apresentado na Figura 2, considerando-se apenas duas variáveis.

Figura 2 - Representação de um conjunto de dados aleatórios descritos originalmente pelas variáveis 1 e 2, utilizando as componentes principais PC1 e PC2. As retas tracejadas representam a projeção das amostras nos eixos das PC's (escores), enquanto as curvas tracejadas representam os ângulos entre a PC e a variável original (loadings)



Fonte: Beebe (1998)

Na Figura 2, os pontos do gráfico representam as amostras, as quais estão inicialmente descritas pelo conjunto de variáveis originais (Variável 1 e 2). Aplicando-se a PCA, esses dados passam a ser descritos pelo novo conjunto de fatores (PC1 e PC2), considerando-se as direções de maior variabilidade nos dados. Adicionalmente, o esquema apresenta também a interpretação geométrica dada aos escores e aos *loadings*. Assim, os escores são representados pela projeção das amostras nos eixos das componentes principais, identificadas pelas retas tracejadas. Enquanto os pesos são obtidos a partir dos ângulos entre o eixo das variáveis originais e o eixo das componentes principais, representados por linhas curvas tracejadas (BEEBE, 1998).

### 2.3.2 Técnicas de pré-processamento aplicadas a dados espectroscópicos

Uma importante etapa da análise quimiométrica consiste na utilização de técnicas de pré-processamento, que são transformações matemáticas aplicadas aos dados antes do desenvolvimento dos modelos, com o intuito de reduzir as fontes de variabilidade irrelevantes para a etapa de modelagem, sejam elas aleatórias ou sistemáticas (BEEBE, 1998). Particularmente, na modelagem de dados espectroscópicos na região do NIR, essas técnicas são frequentemente empregadas para evidenciar a informação química relevante e eliminar, ou ao menos reduzir, os efeitos indesejados do espalhamento da radiação, dentre eles, as

variações na linha de base e a não linearidade. A remoção desses fenômenos físicos é geralmente requerida para melhorar o desempenho dos modelos de calibração, de classificação ou de análise exploratória (RINNAN; VAN DEN BERG; ENGELSEN, 2009; AGELET; HURBURGH, 2010).

Existem diversos pré-processamentos que podem ser aplicados aos dados espectroscópicos. A escolha da técnica mais apropriada depende do tipo do sinal (por exemplo, se são medidas de transmitância ou reflectância), das características das amostras e do objetivo final da etapa de modelagem. Usualmente, os resultados da aplicação dessas técnicas são avaliados de acordo com o aspecto dos dados pré-processados, bem como com os resultados dos modelos desenvolvidos. Dessa forma, apesar de a escolha da técnica mais apropriada se basear também no conhecimento dos dados, frequentemente, mais de um pré-processamento deve ser avaliado (BEEBE, 1998; AGELET; HURBURGH, 2010). A seguir, tem-se uma breve explicação das principais técnicas de pré-processamento de amostras e de variáveis aplicadas aos dados espectroscópicos, as quais foram empregadas no desenvolvimento deste trabalho. As transformações matemáticas apresentadas nesta seção consideram a matriz de dados  $\mathbf{X}$  (resposta instrumental) representada na Equação 1, em que cada linha (i) corresponde a um espectro e cada coluna (j) representa um comprimento de onda ou número de onda.

### 2.3.2.1 Pré-processamento de variáveis

Uma técnica de pré-processamento de variáveis quase sempre empregada aos dados espectroscópicos é a centralização na média. Essa técnica consiste em subtrair de cada elemento da j-ésima coluna da matriz  $\mathbf{X}$  ( $x_{ij}$ ), a média dessa coluna ( $\bar{x}_j$ ), obtendo-se assim um valor corrigido ( $x_{ij}(\text{corr})$ ), conforme Equação 3 (BEEBE, 1998).

$$x_{ij}(\text{corr}) = x_{ij} - \bar{x}_j; i = 1, 2, \dots, I, j = 1, 2, \dots, J \quad (3)$$

A aplicação dessa técnica aos dados espectroscópicos promove a remoção do espectro médio, fazendo com que os eixos das coordenadas sejam movidos para o centro dos dados. Numa PCA, por exemplo, essa transformação permite que as componentes principais sejam relacionadas, de fato, com as direções de maior variabilidade nos dados. Adicionalmente, a centralização na média promove também uma redução da complexidade do modelo,

diminuindo, frequentemente, o número de componentes requeridos na etapa de modelagem (BEEBE, 1998; AGELET; HURBURGH, 2010).

### 2.3.2.2 Pré-processamento de amostras

Uma técnica de pré-processamento de amostras comumente aplicada a dados espectroscópicos é a variação normal padrão (SNV – *Standard Normal Variate*), a qual é empregada com o intuito de reduzir a variabilidade entre as amostras devido ao espalhamento da radiação. Uma vantagem da aplicação desse pré-processamento é que o formato dos espectros corrigidos é semelhante ao dos espectros originais, o que facilita a interpretação dos resultados nas análises posteriores. Na aplicação dessa técnica, o  $i$ -ésimo espectro da matriz  $\mathbf{X}$  ( $\mathbf{x}_i$ ) é subtraído do valor médio do espectro ( $\bar{x}_i$ ), e o resultado é dividido por seu desvio-padrão ( $s_i$ ), obtendo-se assim o espectro corrigido ( $\mathbf{x}_{i(\text{corr})}$ ), conforme apresentado na Equação 4 (BARNES; DHANOA; LISTER, 1989; RINNAN; VAN DEN BERG; ENGELSEN, 2009).

$$\mathbf{x}_{i(\text{corr})} = \frac{\mathbf{x}_i - \bar{\mathbf{x}}_i}{s_i} \quad (4)$$

Outro pré-processamento frequentemente aplicado aos dados espectroscópicos é a correção de espalhamento multiplicativo (MSC – *Multiplicative Scatter Correction*). Assim como a SNV, essa técnica é utilizada para a correção dos efeitos de espalhamento da radiação. Em geral, ambos os pré-processamentos fornecem resultados muito semelhantes, já que a MSC, assim como a SNV, também mantém o formato original dos espectros. A aplicação dessa técnica envolve basicamente duas etapas. Na primeira delas, ajusta-se uma regressão entre os espectros da matriz  $\mathbf{X}$  ( $\mathbf{x}_i$ ) e um espectro de referência ( $\mathbf{x}_{\text{ref}}$ ), conforme apresentado na Equação 5 (RINNAN; VAN DEN BERG; ENGELSEN, 2009).

$$\mathbf{x}_i = b_{0,i} + b_{1,i} \cdot \mathbf{x}_{\text{ref}} + \mathbf{e}_i \quad (5)$$

Geralmente, o espectro médio das amostras de calibração é utilizado como referência, e este é empregado na correção de todos os espectros. Como resultado, tem-se uma estimativa dos coeficientes de regressão linear ( $b_{0,i}$ ) e angular ( $b_{1,i}$ ), denominados coeficientes de correção. Por fim, a parte não modelada dos espectros é computada em um vetor de erros ( $\mathbf{e}_i$ ).

Na segunda etapa da aplicação dessa técnica, os coeficientes de correção são utilizados para a obtenção do espectro corrigido ( $\mathbf{x}_{i(\text{corr})}$ ), de acordo com a Equação 6 (RINNAN; VAN DEN BERG; ENGELSEN, 2009).

$$\mathbf{x}_{i(\text{corr})} = \frac{\mathbf{x}_i - b_{0,i}}{b_{1,i}} = \mathbf{x}_{\text{ref}} + \frac{\mathbf{e}_i}{b_{1,i}} \quad (6)$$

Uma desvantagem da MSC quando comparado à SNV é a necessidade de se ter um espectro de referência, diferentemente do que ocorre na SNV, em que cada espectro é corrigido de forma individualizada (AGELET; HURBURGH, 2010). Assim, no caso de se utilizar o espectro médio como referência, pode haver a necessidade de repetir o procedimento quando uma amostra for excluída do conjunto de dados.

Uma extensão da técnica MSC é a correção de espalhamento multiplicativo estendido (EMSC – *Extended Multiplicative Scatter Correction*). Essa técnica funciona de forma semelhante à MSC, mas realiza um ajuste polinomial de segunda ordem em relação ao espectro de referência. Além disso, se disponível, permite a utilização de informações prévias sobre o espectro de interesse ou sobre os interferentes presentes (MARTENS; STARK, 1991; RINNAN; VAN DEN BERG; ENGELSEN, 2009).

Outra técnica de pré-processamento bastante utilizada para a correção dos efeitos de dispersão da radiação é a derivação espectral, a qual é capaz de remover os efeitos aditivos e multiplicativos dos espectros. Enquanto a primeira derivada remove efeitos aditivos, a segunda derivada é capaz de remover também a tendência linear presente na linha de base (efeitos multiplicativos). Como essa técnica, no entanto, diminui a relação sinal/ruído dos espectros, ela quase nunca é empregada sozinha, sendo associada constantemente a uma técnica de suavização com o intuito de amenizar o aumento indesejado do ruído (BEEBE, 1998; RINNAN; VAN DEN BERG; ENGELSEN, 2009).

Nesse sentido, um método de suavização comumente empregado em conjunto com a derivação espectral é o filtro de Savitzky Golay. Esse filtro faz ajustes polinomiais subsequentes ao longo de todo o espectro, sendo capaz, portanto, de suavizá-lo. Para a aplicação da técnica derivativa com filtro Savitzky Golay, deve-se selecionar o grau da derivada, a ordem do polinômio que será ajustado e o tamanho da janela de pontos utilizados em cada ajuste. Assim, em cada ajuste realizado, o ponto central da janela é substituído pela derivada do polinômio ajustado para a janela. Esta, então, se move para realizar um novo ajuste, de modo que o ponto central do polinômio nesse novo ajuste seja adjacente ao ponto

central do polinômio ajustado anteriormente. Esse procedimento é repetido até que todo o espectro seja corrigido (BEEBE, 1998; RINNAN; VAN DEN BERG; ENGELSEN, 2009; AGELET; HURBURGH, 2010).

Por fim, existe também a técnica de correção de sinal ortogonal (OSC – *Orthogonal Signal Correction*). Diferente das outras técnicas apresentadas, esse pré-processamento utiliza não apenas a informação contida na matriz **X**, como também a propriedade de interesse do analito, que pode ser física ou química. Essa propriedade, geralmente a concentração do analito obtida por um método de referência, é representada por um vetor (**y**), que contém o mesmo número de linhas da matriz **X** analisada. De modo geral, esse pré-processamento remove a variação em **X** que não está relacionada com a propriedade de interesse **y**. Assim, essa técnica é capaz de melhorar o desempenho do modelo de calibração e, frequentemente, requer menos componentes para o desenvolvimento do modelo, facilitando a sua interpretação (WOLD et al., 1998).

### 2.3.3 Aplicação das ferramentas quimiométricas a processos industriais

Há algumas décadas, os métodos quimiométricos têm sido utilizados para o tratamento de dados provenientes de processos industriais, geralmente caracterizados por uma grande quantidade de variáveis correlacionadas (WISE; KOWALSKI, 1995). Esses dados podem compreender tanto as informações químicas obtidas por sensores analíticos instalados na linha de produção, como as medidas espectroscópicas e cromatográficas (GRASSI; ALAMPRESE, 2018; SZYMAŃSKA, 2018), quanto as variáveis de processo, como temperatura e pressão (VITALE; NOORD; FERRER, 2014; BAKDI; KOUADRI, 2017; DU; DU, 2018).

Devido à capacidade dos métodos de análise multivariada para lidar com a complexidade dos dados de processo, tem-se destacado a importância da quimiometria para as aplicações na área de PAT (GRASSI; ALAMPRESE, 2018). Principalmente nas indústrias farmacêuticas, os métodos quimiométricos têm sido cada vez mais empregados como ferramentas de PAT em diversas etapas da produção (FERREIRA; TOBYN, 2014; VIGNADUZZO; MAGGIO; OLIVIERI, 2020). Conforme revisado por Vignaduzzo, Maggio e Olivieri (2020), inclusive, diversas agências reguladoras têm incluído os métodos quimiométricos em guias e documentos direcionados a essas indústrias.

Recentemente, o emprego de modernos analisadores de processos e ferramentas de controle na linha de produção, bem como o desenvolvimento de sistemas computacionais e da

informação, promoveram um aumento considerável no volume de dados coletados durante os processos industriais, tornando-os, portanto, ainda mais complexos. A possibilidade de utilização desses dados para monitorar e controlar os processos representa uma oportunidade de crescimento e desenvolvimento para as indústrias, podendo proporcionar diversas melhorias no processo produtivo (HE; WANG, 2018).

É importante ressaltar, no entanto, os aspectos desafiadores desse cenário, os quais estão frequentemente associados ao gerenciamento, à integração e à extração de informações significativas desses dados, de modo que permitam o controle e o monitoramento do processo de forma eficiente. Nesse contexto, a quimiometria é vista como uma importante parte da ciência de dados que pode ser aplicada ao tratamento de dados multivariados e complexos coletados durante os processos industriais (GE, 2017; SZYMAŃSKA, 2018).

A aplicação dos métodos quimiométricos a dados de processo é bastante abrangente. Por exemplo, ferramentas quimiométricas têm sido amplamente utilizadas no monitoramento estatístico de processos (TIDRIRI et al., 2016; GE, 2017; HE; WANG, 2018), no desenvolvimento de modelos de calibração multivariada para a predição de propriedades de interesse (KADLEC; GABRYS; STRANDT, 2009; LIU; XIE, 2020; VIGNADUZZO; MAGGIO; OLIVIERI, 2020) e na análise multivariada de imagens para monitoramento da qualidade dos produtos e das condições do processo (DUCHESNE; LIU; MACGREGOR, 2012).

Considerando o desenvolvimento de modelos de calibração multivariada, uma abordagem importante dessa aplicação aos processos industriais consiste no desenvolvimento de sensores virtuais. Esses sensores são frequentemente utilizados para estimar variáveis difíceis de medir no processo, através da construção de modelos preditivos com base em variáveis de processo mais facilmente mensuráveis (KADLEC; GABRYS; STRANDT, 2009; LIU; XIE, 2020).

Outra aplicação dos modelos de calibração multivariada, que está relacionada com uma das abordagens descrita no presente trabalho, consiste no desenvolvimento de modelos de regressão para estimar propriedades de interesse, geralmente concentração, de produtos ou compostos intermediários do processo ou da mistura reacional durante a reação. Para esse tipo de aplicação, os métodos espectroscópicos como o de infravermelho próximo são frequentemente utilizados em indústrias farmacêutica, alimentícia e petroquímica (DE BEER et al., 2011; DA SILVA et al., 2015; GRASSI; ALAMPRESE, 2018; DA SILVA et al., 2019). Conforme discutido anteriormente, essa técnica analítica apresenta características

favoráveis para aplicações na área de PAT, incluindo o monitoramento do processo em tempo real.

No que diz respeito ao monitoramento estatístico de processos, os métodos quimiométricos têm sido utilizados para a detecção, a identificação e o diagnóstico de falhas em processos industriais (TIDRIRI et al., 2016; GE, 2017; SHANG; YOU, 2019). De modo geral, a detecção de falhas consiste em avaliar se o processo está ou não em controle estatístico, já a identificação consiste em especificar o tipo de falha e, por fim, o diagnóstico envolve a determinação da causa raiz do problema (GE, 2017).

Nesse contexto, existem diversos *reviews* na literatura que abordam o uso de diferentes técnicas quimiométricas no monitoramento estatístico de processos através de gráficos de controle (TIDRIRI et al., 2016; GE, 2017; REIS; GINS, 2017; HE; WANG, 2018; SHANG; YOU, 2019). Esses trabalhos abordam desde as técnicas mais simples, como PCA e PLS, até técnicas não-lineares, como os métodos baseados em funções *Kernel* e *Support Vector Machine*, para o monitoramento de processos contínuos e em batelada.

## 2.4 CALIBRAÇÃO MULTIVARIADA

A calibração multivariada é uma das vertentes da quimiometria bastante aplicada a processos no contexto de PAT (FERREIRA; TOBYN, 2014; SIMON et al., 2015). Neste tópico, serão abordados alguns temas relacionados à calibração multivariada aplicados no desenvolvimento do presente trabalho, incluindo: o método SPXY (*Sample set Partitioning based on joint x–y distances*), utilizado para a separação de um conjunto de amostras em calibração e predição; o método de calibração multivariada de PLS; e as métricas de desempenho utilizadas para avaliação dos modelos desenvolvidos.

### 2.4.1 Método de seleção de amostras

Com o objetivo de construir e validar um modelo de calibração multivariada a partir de  $N$  amostras conhecidas, existem algumas ferramentas que podem ser utilizadas para separá-las em dois conjuntos: um contendo as amostras de calibração e outro contendo as amostras de predição. O método utilizado no presente trabalho para o desenvolvimento de modelos de regressão foi o algoritmo SPXY. Basicamente, esse algoritmo calcula as distâncias Euclidianas entre todos os pares de amostras ( $p, q$ ), com base nos dados da matriz

**X** representada na Equação 1 ( $d_x(p, q)$ ), e nos valores da propriedade de interesse contidos no vetor **y** ( $d_y(p, q)$ ). Essas distâncias são calculadas, respectivamente, de acordo com as Equações 7 e 8:

$$d_x(p, q) = \sqrt{\sum_{j=1}^J [x_{p,j} - x_{q,j}]^2}; p, q \in [1, I] \quad (7)$$

$$d_y(p, q) = \sqrt{(y_p - y_q)^2}; p, q \in [1, I] \quad (8)$$

Em que  $x_{p,j}$  e  $x_{q,j}$  são as respostas instrumentais, referentes à  $j$  – ésima variável, e  $y_p$  e  $y_q$  são os valores da propriedade de interesse, para as amostras  $p$  e  $q$ , respectivamente.

Em seguida, de acordo com a Equação 9, a distância normalizada ( $d_{xy}(p, q)$ ) é calculada pela soma das distâncias  $d_x(p, q)$  e  $d_y(p, q)$ , divididas pelos seus respectivos valores máximos, com o intuito de atribuir a mesma importância à distribuição das amostras em **X** e em **y** (GALVÃO et al., 2005).

$$d_{xy}(p, q) = \frac{d_x(p, q)}{\max_{p,q \in [1, I]} d_x(p, q)} + \frac{d_y(p, q)}{\max_{p,q \in [1, I]} d_y(p, q)}; p, q \in [1, I] \quad (9)$$

No algoritmo, a seleção das amostras é feita de forma iterativa. Assim, as primeiras amostras selecionadas são aquelas em que a distância  $d_{xy}(p, q)$  é a maior. Nas próximas iterações, o algoritmo seleciona a próxima amostra que apresenta a maior distância em relação a alguma amostra já escolhida. Esse procedimento é repetido até que todas as amostras de calibração sejam selecionadas (GALVÃO et al., 2005).

#### 2.4.2 Regressão por mínimos quadrados parciais

A Regressão por Mínimos Quadrados Parciais (*Partial Least-squares* - PLS) é um dos métodos quimiométricos mais utilizados no desenvolvimento de modelos de calibração multivariada. Assim como na calibração univariada, o objetivo principal dessa técnica é estabelecer uma relação matemática entre os dados adquiridos de uma amostra, contidos na matriz **X**, e uma propriedade de interesse dessa amostra, a qual pode ser de natureza química ou física, e é representada pelo vetor **y**. De forma simplificada, a aplicação desse método

envolve essencialmente duas fases. Na primeira etapa (calibração), os dados da matriz  $\mathbf{X}$  bem como os valores do vetor  $\mathbf{y}$ , obtidos por um método de referência, são utilizados no desenvolvimento do modelo de calibração. Já na segunda etapa (predição), esse modelo é usado para a predição da propriedade de interesse em amostras novas, as quais não foram empregadas na construção do modelo.

A modelagem matemática por PLS é feita através de cálculos iterativos. Para isso, existem diversos algoritmos que podem ser utilizados, os quais são encontrados em pacotes de *softwares* comerciais (ANDERSSON, 2009). A seguir, tem-se uma descrição simplificada das equações utilizadas nessa modelagem, com base no que foi apresentado por Olivieri (2018).

A equação geral para um modelo linear entre a matriz de resposta instrumental ( $\mathbf{X}$ ) e um vetor com a propriedade de interesse do analito ( $\mathbf{y}_n$ ), para  $n$  amostras utilizadas na etapa de calibração, é dada pela Equação 10, em que  $\mathbf{e}$  é o vetor de erros e  $\mathbf{b}_n$ , o vetor dos coeficientes de regressão.

$$\mathbf{y}_n = \mathbf{X}\mathbf{b}_n + \mathbf{e} \quad (10)$$

No caso de as variáveis em  $\mathbf{X}$  serem correlacionadas, essa matriz de dados pode ser substituída por uma versão comprimida, calculada com base nas combinações lineares das variáveis originais, semelhante ao que ocorre na PCA. Essa versão comprimida de  $\mathbf{X}$  corresponde à matriz de escores ( $\mathbf{T}_A$ ), truncada nas  $A$  primeiras colunas, sendo  $A$  o número de fatores utilizados na modelagem. No PLS, esses fatores são denominados variáveis latentes (VLs) e devem reter a principal parte das variabilidades em  $\mathbf{X}$  e em  $\mathbf{y}_n$ . A Equação 10, após a substituição de  $\mathbf{X}$  por  $\mathbf{T}_A$ , pode ser reescrita de acordo com a Equação 11.

$$\mathbf{y}_n = \mathbf{T}_A \mathbf{v}_n + \mathbf{e} \quad (11)$$

É possível perceber que o vetor  $\mathbf{b}_n$  também foi substituído por  $\mathbf{v}_n$ , que é o vetor dos coeficientes de regressão do PLS, definido no espaço das  $A$  variáveis latentes. Para concluir a etapa de calibração,  $\mathbf{v}_n$  pode ser calculado a partir da Equação 12.

$$\mathbf{v}_n = \mathbf{T}_A^+ \mathbf{y}_n \quad (12)$$

Em que  $\mathbf{T}_A^+$  é a inversa generalizada de  $\mathbf{T}_A$ , dada por  $\mathbf{T}_A^+ = (\mathbf{T}_A^T \mathbf{T}_A)^{-1} \mathbf{T}_A$ , a qual foi obtida após algumas transformações matemáticas da Equação 11: multiplicação de ambos os lados da equação por  $\mathbf{T}_A^T$  (matriz transposta de  $\mathbf{T}_A$ ) e posterior multiplicação de ambos os lados da equação por  $(\mathbf{T}_A^T \mathbf{T}_A)^{-1}$ .

Após a estimativa do vetor dos coeficientes de regressão ( $\mathbf{v}_n$ ), pode-se prosseguir para a etapa de predição. No entanto, antes de seguir para essa etapa, é importante compreender como  $\mathbf{T}_A$  foi obtida. De forma simplificada, a matriz de escores  $\mathbf{T}_A$  é calculada de acordo com a Equação 13

$$\mathbf{T}_A = \mathbf{X}^T \mathbf{W}_A (\mathbf{P}_A^T \mathbf{W}_A)^{-1} \quad (13)$$

Na Equação 13,  $\mathbf{W}_A$  é matriz de pesos (*weight loadings*) e  $\mathbf{P}_A$  é a matriz de *loadings*. As A colunas da matriz  $\mathbf{W}_A$  são formadas pelos vetores  $\mathbf{w}_A$  normalizados, obtidos de acordo com a Equação 14.

$$\mathbf{w}_A = \frac{\mathbf{X}^T \mathbf{y}_n}{\|\mathbf{X}^T \mathbf{y}_n\|} \quad (14)$$

Já as A colunas da matriz de *loadings* ( $\mathbf{P}_A$ ), são formadas pelos vetores  $\mathbf{p}_A$ , calculados com base nos vetores da matriz de escores  $\mathbf{T}_A$  ( $\mathbf{t}_A$ ), de acordo com a Equação 15.

$$\mathbf{p}_A = \frac{\mathbf{X}^T \mathbf{t}_A}{\|\mathbf{t}_A^T \mathbf{t}_A\|} \quad (15)$$

A utilização das duas matrizes de pesos e *loadings* na estimativa de  $\mathbf{T}_A$  representa um diferencial do PLS, uma vez que permite que as variáveis latentes do modelo sejam estimadas levando-se em consideração não apenas a matriz  $\mathbf{X}$ , mas também os valores da propriedade de interesse do analito. Assim, a utilização dessas duas matrizes permite que as variáveis latentes

expliquem de forma simultânea a máxima variância espectral dos dados em  $\mathbf{X}$  e a máxima covariância entre  $\mathbf{X}$  e  $\mathbf{y}_n$ . Isso faz com que o modelo PLS seja adaptado para cada analito, sendo considerado, portanto, uma ferramenta poderosa para a construção de modelos de calibração.

Na etapa de predição, o vetor dos coeficientes de regressão do modelo PLS desenvolvido na etapa de calibração ( $\mathbf{v}_n$ ) é utilizado para estimar a propriedade de interesse do analito das amostras do conjunto de predição. Antes disso, no entanto, uma etapa preliminar é requerida, a qual envolve o cálculo dos escores ( $\mathbf{t}_A$ ) para a amostra de predição, a partir dos dados dessa amostra ( $\mathbf{x}$ ), de acordo com a Equação 16:

$$\mathbf{t}_A = (\mathbf{W}_A^T \mathbf{P}_A)^{-1} \mathbf{W}_A^T \mathbf{x} \quad (16)$$

Por fim, calculado o vetor  $\mathbf{t}_A$ , a predição da propriedade de interesse ( $\hat{y}$ ) da amostra pode ser estimada pela Equação 17:

$$\hat{y} = \mathbf{v}_n^T \mathbf{t}_A \quad (17)$$

#### 2.4.3 Avaliação dos modelos de calibração

A avaliação dos modelos de calibração pode ser realizada através da estimativa de alguns parâmetros estatísticos, denominados métricas de desempenho. A raiz do erro quadrático médio (*Root Mean Square Error* - RMSE), por exemplo, é um dos principais parâmetros utilizados para avaliar um modelo de calibração multivariada. O mesmo pode ser calculado para a etapa de calibração (RMSEC), de validação cruzada (RMSECV) e de predição (RMSEP). O RMSEP, em particular, é o parâmetro mais utilizado para expressar a exatidão de um modelo baseado em dados do NIR. Ele pode ser calculado a partir da Equação 18, em que  $y_i$  é o valor medido para a  $i$ -ésima amostra, obtido pelo método de referência,  $\hat{y}_i$  é o valor predito pelo modelo para essa amostra, e  $n_{pred}$  é o número de amostras do conjunto de predição (PASQUINI, 2018).

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n_{pred}} (y_i - \hat{y}_i)^2}{n_{pred}}} \quad (18)$$

Com o objetivo de avaliar diversos modelos (desenvolvidos com diferentes técnicas de pré-processamento, por exemplo), os valores de RMSEP desses modelos podem ser comparados através do teste estatístico de Fischer (teste F), o qual é frequentemente utilizado para testar hipóteses sobre a igualdade entre duas variâncias populacionais.

Outra métrica de desempenho importante utilizada para a avaliação dos modelos de calibração é o *bias*, que avalia a presença de erro sistemático no modelo. Da mesma forma que ocorre para o RMSE, o *bias* também pode ser calculado para os conjuntos de calibração, validação cruzada e predição. Para a etapa de predição, pode ser estimado a partir da Equação 19.

$$bias_{pred} = \frac{\sum_{i=1}^{n_{pred}} (y_i - \hat{y}_i)}{n_{pred}} \quad (19)$$

Uma prática comum na análise dos modelos de calibração consiste em realizar um teste estatístico de *student* (teste t) com o objetivo de avaliar se a predição de um modelo apresenta um *bias* estatisticamente significativo, ou seja, se há, de fato, erros sistemáticos no modelo. Essa é uma recomendação, inclusive, da norma *American Society for Testing and Materials* (ASTM) E1655-05 (*Standard Practices for Infrared Multivariate Quantitative Analysis*), a qual estabelece algumas diretrizes para o desenvolvimento de modelos de calibração multivariada com base em dados de infravermelho (ASTM E1655-05, 2005).

Por fim, o coeficiente de determinação ( $R^2$ ) é comumente utilizado para avaliar os modelos de calibração. Também pode ser calculado para os conjuntos de calibração, validação cruzada e predição. Esse parâmetro permite estimar a porcentagem da variação nos dados que é modelada adequadamente, sendo obtido a partir da Equação 20, em que  $\bar{y}$  é o valor médio de  $y$  considerando todas as amostras (ASTM E1655-05, 2005).

$$R^2_{pred} = \frac{\sum_{i=1}^{n_{pred}} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n_{pred}} (y_i - \bar{y})^2} \quad (20)$$

O  $R^2$ , no entanto, deve ser utilizado com cuidado na avaliação de um modelo de calibração. Pelo fato de o coeficiente de determinação depender da variabilidade dos valores de referência da propriedade de interesse, ele não deve ser utilizado sozinho como um critério de qualidade para os modelos de calibração. Assim, ele deve ser analisado juntamente com

outros parâmetros, como o RMSEP, por exemplo, que é considerado mais eficaz do que o  $R^2$  (AGELET; HURBURGH, 2010; PASQUINI, 2018).

Além da análise dos parâmetros descritos, a avaliação de um modelo de calibração exige também um olhar criterioso sobre as variáveis consideradas críticas para o desenvolvimento do modelo. De acordo com Pasquini (2018), os resultados dos modelos de calibração devem ser acompanhados de uma interpretação química, com o intuito de que correlações internas não sejam negligenciadas. Dessa forma, evita-se que a correlação estabelecida pelo modelo não seja referente ao analito de interesse.

Para a interpretação química dos modelos desenvolvidos, pode-se utilizar os gráficos dos coeficientes de regressão e os gráficos de importância das variáveis na projeção (VIP - *Variable Importance in the Projection*). De modo geral, a análise desses gráficos permite identificar as regiões consideradas mais relevantes para o estabelecimento da relação entre a matriz  $\mathbf{X}$  e o vetor  $\mathbf{y}$ . Tais gráficos estão disponíveis para análise em muitos softwares comerciais. O gráfico de importância das variáveis, inclusive, é bastante empregado como método de seleção de variáveis (ANDERSEN; BRO, 2010), com o intuito de melhorar o desempenho do modelo de calibração, construindo-o com apenas algumas regiões selecionadas do espetro, ao invés do espectro completo.

## 2.5 MONITORAMENTO ESTATÍSTICO DE PROCESSOS

Conforme mencionado, além da calibração multivariada, o monitoramento estatístico de processos (SPM - *Statistical Process Monitoring*) é outro tema abrangido pela quimiometria e amplamente aplicado a processos industriais. O SPM corresponde a um conjunto de ferramentas utilizadas para avaliar e alcançar a estabilidade de um determinado processo produtivo e a melhoria do seu desempenho, através da redução de sua variabilidade (FERRER-RIQUELME, 2009; MONTGOMERY, 2009). Essas ferramentas, quando adequadamente aplicadas a processos industriais, podem acarretar benefícios do ponto de vista da qualidade, da segurança, da eficiência e da lucratividade (REIS; GINS, 2017).

Para a redução da variabilidade de um processo, é necessário identificar e compreender as principais fontes causadoras dessa variabilidade. De modo geral, essas causas podem ser de dois tipos: aleatórias e especiais. As causas aleatórias ou comuns compreendem as diversas fontes de variação da qualidade que atuam de forma aleatória sobre o processo, gerando uma variabilidade natural. Usualmente, diz-se que os processos que operam apenas

com causas desse tipo estão sob controle estatístico, sendo considerados estáveis e previsíveis. Já as causas especiais ou atribuíveis, são as fontes de variação que surgem esporadicamente e não seguem um padrão aleatório. São consideradas falhas no processo, ocasionadas, principalmente, pelo uso de matéria-prima defeituosa, por erros do operador ou pelo controle inadequado das máquinas. Diz-se que um processo que opera também com causas especiais está fora de controle e, portanto, é um processo instável e imprevisível (MONTGOMERY, 2009).

Nesse sentido, com o objetivo de distinguir entre as causas aleatórias e especiais que atuam sobre um determinado processo, gráficos de controle estatístico (também denominados gráficos de monitoramento), podem ser utilizados para monitorá-lo. Esses gráficos são as ferramentas de maior destaque do SPM e permitem identificar o tipo e a magnitude da causa da variabilidade. Assim, o emprego dessas ferramentas de forma adequada pode promover uma melhoria significativa no desempenho do processo, através da remoção das causas especiais detectadas, ou da sua implementação, no caso de serem benéficas (FERRER-RIQUELME, 2009). Adicionalmente, quando aplicadas em tempo real, permitem a detecção precoce de falhas, proporcionando celeridade na tomada de decisão, bem como o desenvolvimento de soluções eficientes diante de urgências no processo produtivo.

Os gráficos de controle surgiram por volta de 1924, por meio do trabalho pioneiro de Walter Shewhart (MONTGOMERY, 2009). Esse marco deu início ao que normalmente se denomina Controle Estatístico de Processos (CEP), que representou um avanço significativo para o monitoramento da qualidade do processo e do produto. De acordo com He e Wang (2018), o CEP corresponde à primeira geração do SPM, que envolve a utilização de gráficos de controle univariados. Esses gráficos são geralmente empregados no monitoramento de uma característica da qualidade do produto, mas também podem ser usados para o monitoramento de uma variável de processo considerada crítica para o controle de qualidade do produto (FERRER-RIQUELME, 2009; MONTGOMERY, 2009).

À medida que os processos industriais se tornam mais complexos, com a implementação de modernos sistemas de controle e automação e a aquisição de uma enorme quantidade de dados provenientes de múltiplos sensores, é necessário o desenvolvimento de metodologias para monitorá-los e controlá-los de forma eficiente. Nesse contexto, de modo a atender à evolução dos sistemas industriais, surgiu o monitoramento estatístico multivariado de processos (MSPM - *Multivariate Statistical Process Monitoring*), que corresponde à segunda geração do SPM. De modo geral, o MSPM abrange diversas metodologias utilizadas

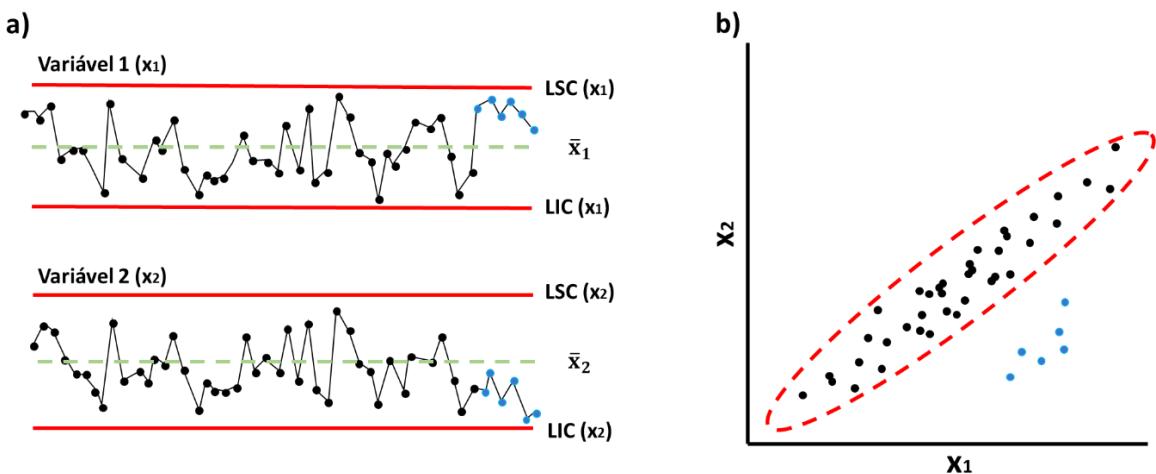
para monitorar processos a partir de dados multivariados, em que as variáveis são analisadas simultaneamente. Por fim, existe ainda a terceira geração do SPM, que engloba os métodos multivariados que têm sido desenvolvidos para lidar com uma complexidade ainda maior, contemplando, por exemplo, características como dinâmica e não linearidade (BAKDI; KOUADRI, 2017; HE; WANG, 2018).

De modo geral, o conceito associado aos gráficos de controle uni e multivariados é semelhante. A aplicação desses gráficos envolve duas fases. Na primeira (denominada Fase I), dados de processo são coletados e empregados na determinação dos limites de controle tentativos, os quais são utilizados para identificar se o processo estava ou não sob controle estatístico durante a coleta desses dados. Nessa etapa, dados considerados fora de controle são removidos e os limites são recalculados. Esse procedimento é repetido até que os limites de controle do gráfico sejam estabelecidos, com base em dados coletados durante condições estáveis e representativas do desempenho do processo. Já na segunda fase (Fase II), o gráfico de controle desenvolvido é utilizado para monitorar o processo. Assim, novos dados são coletados e projetados no gráfico, de preferência em tempo real, permitindo determinar se o processo se mantém ou não em um estado de controle estatístico (MONTGOMERY, 2009).

A principal diferença entre os gráficos de controle uni e multivariados, no entanto, não se resume apenas ao fato de que as cartas de controle multivariadas são utilizadas para o monitoramento simultâneo de diversas variáveis. Uma das principais vantagens da abordagem multivariada, quando comparada à univariada, consiste na possibilidade de explorar informações relacionadas à covariância entre as variáveis. Como resultado, os gráficos multivariados, em geral, são capazes de detectar situações anômalas que não seriam detectadas caso cartas de controle univariadas fossem utilizadas para o monitoramento individual de cada variável (FERRER-RIQUELME, 2009; TIDRIRI et al., 2016; HE; WANG, 2018).

Uma forma de ilustrar a comparação entre as abordagens uni e multivariadas está apresentada na Figura 3.

Figura 3 – Comparação entre as abordagens de monitoramento estatístico univariada (limites de controle em vermelho – linha contínua) (a) e multivariada (elipse de controle em vermelho – linha tracejada) (b) para o monitoramento de duas variáveis aleatórias dependentes



Fonte: He; Wang (2018)

É possível perceber através da Figura 3 que, se cada variável fosse monitorada utilizando-se gráficos de controle univariados (Figura 3a), estes falhariam na detecção de amostras anômalas (em azul), uma vez que, para ambas as variáveis, todos os pontos estão dentro dos limites de controle (linhas contínuas em vermelho). O comportamento incomum dessas amostras apenas fica evidente quando ambas as variáveis são monitoradas de forma simultânea (Figura 3b). Assim, se uma estatística multivariada fosse utilizada no cálculo do limite de controle (representado pela elipse tracejada em vermelho), as amostras anômalas seriam detectadas, uma vez que não seguem o padrão de correlação positiva entre as variáveis  $x_1$  e  $x_2$ , observada para as outras amostras. No exemplo apresentado na Figura 3, foram analisadas apenas duas variáveis aleatórias e dependentes. No entanto, à medida que o número de variáveis dependentes aumenta, há também um aumento da distorção no procedimento de monitoramento por parte das cartas univariadas (MONTGOMERY, 2009; HE; WANG, 2018).

Dessa forma, os benefícios associados ao uso de cartas de controle multivariadas são especialmente importantes quando se lida com dados de processos complexos, de alta dimensionalidade e colinearidade, como a maioria dos processos industriais modernos que usam sistemas avançados de aquisição de dados. Isso se observa, principalmente, em processos químicos, em que as variáveis do processo são altamente correlacionadas devido aos princípios físicos e químicos que os conduzem, como a termodinâmica, a cinética química

e os balanços de massa e energia (FERRER, 2014; SEVERSON; CHAIWATANODOM; BRAATZ, 2016; TIDRIRI et al., 2016; REIS; GINS, 2017; HE; WANG, 2018).

No entanto, apesar de os métodos multivariados serem em geral mais eficientes para o monitoramento dos processos modernos, em algumas situações, a utilização de técnicas univariadas pode ser necessária (ou mais adequada). De modo geral, a seleção da abordagem mais apropriada, seja ela uni ou multivariada, depende das particularidades do processo, dentre as quais: das variáveis de processo ou de qualidade consideradas críticas, da estrutura de correlação dos dados e da dinâmica do processo. Além disso, depende também do objetivo final que se deseja alcançar com a aplicação da ferramenta do SPM.

É importante ressaltar também que, em alguns estudos recentes, ambas as abordagens uni e multivariadas foram utilizadas em conjunto com o objetivo de melhorar o monitoramento do processo (BIN SHAMS; BUDMAN; DUEVER, 2011; HARROU et al., 2015; BAKDI; KOUADRI, 2017; DU; DU, 2018). Nesses estudos, a habilidade de cartas univariadas em detectar pequenos desvios no processo foi explorada com o objetivo de melhorar o desempenho de cartas de controle multivariadas baseadas em fatores/componentes principais e variáveis latentes, como PCA e PLS, no que diz respeito à detecção de falhas em processos industriais.

No contexto dos gráficos de controle aplicados a processos, considerando-se as abordagens empregadas no desenvolvimento do presente trabalho, a seguir, tem-se uma breve apresentação dos principais gráficos de controle univariados. Adicionalmente, serão introduzidas também algumas particularidades relacionadas ao monitoramento de processos autocorrelacionados, tema também abordado na realização deste trabalho.

### 2.5.1 Gráficos de controle univariados

Os gráficos de controle univariados que são convencionalmente empregados no monitoramento de uma característica da qualidade do processo são as cartas de controle de Shewhart, da média móvel exponencialmente ponderada (*Exponentially Weighted Moving Average* - EWMA) e da soma cumulativa (*Cumulative Sum* - CUSUM) (FERRER-RIQUELME, 2009; MONTGOMERY, 2009).

Os gráficos de controle de Shewhart compreendem os pares de gráficos para monitoramento da média e da variabilidade de uma variável crítica de qualidade ou de processo. Para o monitoramento da média, utiliza-se o gráfico de controle  $\bar{x}$ . Já para o

monitoramento da variabilidade, pode-se empregar o gráfico de controle  $s$  (gráfico de controle para o desvio-padrão) ou o gráfico de controle  $R$  (gráfico de controle para a amplitude). Considerando-se o par de gráficos  $\bar{x}$  e  $R$ , os limites inferior (LIC) e superior (LSC) de controle para o gráfico  $\bar{x}$  são obtidos, respectivamente, pelas Equações 21 e 22 (MONTGOMERY, 2009):

$$\text{LIC} = \bar{\bar{x}} - A_2 \bar{R} \quad (21)$$

$$\text{LSC} = \bar{\bar{x}} + A_2 \bar{R} \quad (22)$$

Em que  $\bar{\bar{x}}$  é a média geral das amostras utilizadas na determinação dos limites de controle,  $A_2$  é uma constante tabelada dependente do tamanho das amostras e  $\bar{R}$  é a amplitude média dessas amostras. Já para o gráfico  $R$ , os LIC e LSC são calculados, respectivamente, a partir das Equações 23 e 24 (MONTGOMERY, 2009):

$$\text{LIC} = D_3 \bar{R} \quad (23)$$

$$\text{LSC} = D_4 \bar{R} \quad (24)$$

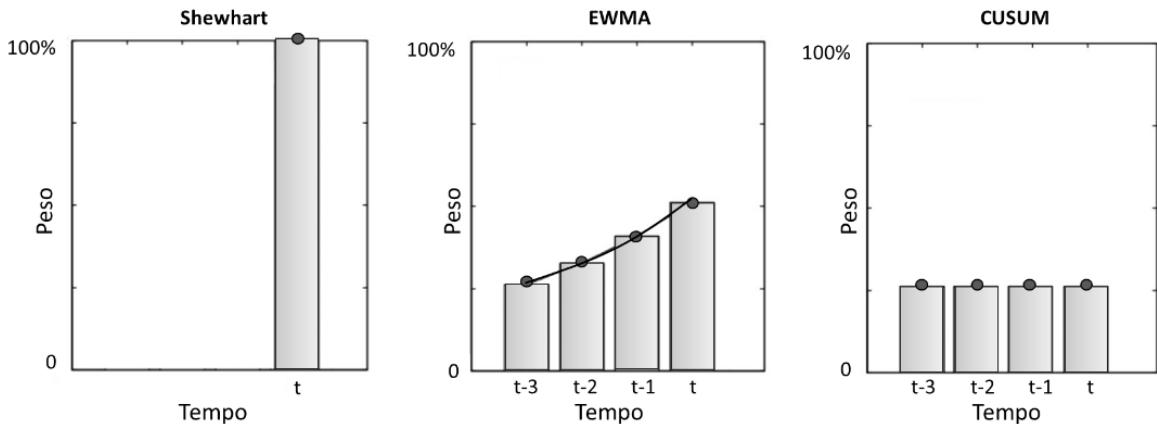
Em que  $D_3$  e  $D_4$  também são constantes tabeladas dependentes do tamanho das amostras.

De modo geral, os gráficos de controle de Shewhart permitem a captura de desvios no processo de grande magnitude. Dessa forma, são mais apropriados para utilização na Fase I de implementação do SPM. Nessa fase, em que são definidos os limites de controle, a possibilidade de ocorrência de grandes desvios no processo é maior, uma vez que o mesmo está sendo ajustado para dentro de controle. Uma desvantagem desses gráficos, no entanto, é que apenas a informação contida na observação mais recente é levada em consideração. Como resultado, os gráficos de Shewhart são relativamente insensíveis a pequenos e médios desvios, sendo menos úteis para a Fase II do monitoramento (MONTGOMERY, 2009; FERRER, 2014).

Por outro lado, os gráficos de controle EWMA e CUSUM consideram não apenas a informação da observação presente, mas também das observações passadas e, portanto, são mais apropriadas para a detecção de pequenos desvios no processo (MONTGOMERY, 2009;

FERRER, 2014). Em geral, essas estatísticas apresentam desempenho semelhante, mas a EWMA é considerada mais fácil de operar e implementar. A principal diferença entre essas cartas é que a CUSUM é calculada considerando pesos iguais às observações anteriores, enquanto que a EWMA considera uma média exponencialmente ponderada dessas observações, em que o peso diminui com a idade da amostra (MONTGOMERY, 2009; KADRI et al., 2016). A comparação entre o cálculo das cartas de controle de Shewhart, EWMA e CUSUM, do ponto de vista do peso aplicado às amostras presente e passadas, está esquematizada na Figura 4.

Figura 4 – Comparação entre o cálculo das cartas de controle de Shewhart, EWMA e CUSUM, do ponto de vista do peso aplicado às amostras presente e passadas



Fonte: Kadri et al. (2016)

A estatística EWMA para a  $i$ -ésima observação ( $x_i$ ) pode ser calculada de acordo com a Equação 25:

$$z_i = \lambda x_i + (1 - \lambda)z_{i-1} \quad (25)$$

Em que  $z_i$  é a estatística EWMA no tempo  $i$ ,  $\lambda$  é o parâmetro de suavização e  $z_{i-1}$  é a estatística EWMA referente ao tempo anterior ( $i - 1$ ). O parâmetro de suavização pode variar de zero a um ( $0 < \lambda \leq 1$ ), e determina o quanto da memória dos dados é retida no cálculo da EWMA. Assim, quanto maior o valor de  $\lambda$ , menores os pesos aplicados às observações mais antigas e, consequentemente, maiores os pesos aplicados às observações mais recentes. Por outro lado, quanto menor o valor de  $\lambda$ , maiores os pesos aplicados às observações mais antigas. Dessa forma, esse parâmetro deve ser selecionado de acordo com as características do processo monitorado (KADRI et al., 2016). O valor de partida ( $z_0$ ), requerido quando  $i = 1$ ,

pode ser considerado como o valor alvo ou a média de valores preliminares do processo que estejam dentro de controle ( $\mu_0$ ) (MONTGOMERY, 2009).

Os limites inferior e superior de controle para esse gráfico são calculados de acordo com as Equações 26 e 27, respectivamente, em que  $L$  representa a largura dos limites de controle (MONTGOMERY, 2009).

$$LIC = \mu_0 - L\sigma \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2i}]} \quad (26)$$

$$LSC = \mu_0 + L\sigma \sqrt{\frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2i}]} \quad (27)$$

Já a carta de controle CUSUM, é baseada no cálculo das somas cumulativas dos desvios dos valores da amostra em relação a um valor alvo ( $\mu_0$ ). Na construção desse gráfico, calcula-se as somas cumulativas dos desvios de  $\mu_0$  que estão acima do alvo, denominadas CUSUM unilateral superior ( $C_i^+$ ), e as somas cumulativas dos desvios de  $\mu_0$  que estão abaixo do alvo, denominadas CUSUM unilateral inferior ( $C_i^-$ ). Essas estatísticas são estimadas, respectivamente, pelas Equações 28 e 29 (FERRER-RIQUELME, 2009; MONTGOMERY, 2009).

$$C_i^+ = \max[0, x_i - (\mu_0 + K) + C_{i-1}^+] \quad (28)$$

$$C_i^- = \max[0, (\mu_0 - K) - x_i + C_{i-1}^-] \quad (29)$$

Em que  $K$  é o valor de referência. Além disso, na construção dessa carta de controle, é definido também um intervalo de decisão ( $H$ ), que age como um limite superior de controle para esse gráfico, o qual é geralmente definido como cinco vezes o desvio-padrão do processo ( $H = 5\sigma$ ).

Por fim, as cartas de controle de CUSUM e EWMA são tradicionalmente empregadas para o monitoramento da média do processo (MONTGOMERY, 2009). No entanto, essas cartas e suas variações também têm sido utilizadas para monitorar outras características do processo, como a dispersão (ZWETSLOOT; AJADI, 2019), os resíduos da modelagem de séries temporais (KADRI et al., 2016; YANG et al., 2018) e o monitoramento simultâneo da média e da variabilidade (SANUSI; MUKHERJEE; XIE, 2019). Adicionalmente, a habilidade

dessas cartas em detectar pequenos desvios no processo tem sido aplicada, inclusive, para melhorar o desempenho de cartas de controle multivariadas utilizadas no monitoramento de processos industriais (BIN SHAMS; BUDMAN; DUEVER, 2011; HARROU et al., 2015; BAKDI; KOUADRI, 2017; DU; DU, 2018).

### 2.5.2 Gráficos de controle para processos autocorrelacionados

De modo geral, a aplicação das cartas de controle univariadas convencionais parte do pressuposto que os dados são independentes e normalmente distribuídos quando o processo está em controle estatístico (MONTGOMERY, 2009). Em diversas situações, no entanto, desvios da normalidade podem não afetar o desempenho dessas cartas (FERRER-RIQUELME, 2009). O gráfico EWMA, por exemplo, quando desenvolvido de forma apropriada, apresenta desempenho satisfatório mesmo em condição de não normalidade (BORROR; MONTGOMERY; RUNGER, 1999; HUMAN; KRITZINGER; CHAKRABORTI, 2011).

A violação da suposição de independência entre as observações, entretanto, pode ocasionar um impacto maior no desempenho dos gráficos de controle. Essa violação é geralmente expressa em termos da função de autocorrelação, que estabelece a correlação entre as amostras ao longo do tempo para uma mesma variável ou série temporal. Em processos modernos com sistemas de aquisição de dados que permitem elevadas taxas de amostragem, a presença de autocorrelação nos dados é algo muito comum, principalmente, devido ao pequeno intervalo de aquisição das medidas em comparação com a dinâmica do processo (FERRER-RIQUELME, 2009; KADRI et al., 2016; REIS; GINS, 2017). Sob essas circunstâncias, o uso de gráficos de controle convencionais pode levar a um número excessivo de alarmes falsos, a uma estimativa inadequada dos parâmetros do processo e a problemas na detecção de desvios nas condições normais de operação (REYNOLDS; LU, 1997).

Dessa forma, devido à ineficiência das cartas de controle tradicionais para lidar com processos autocorrelacionados, duas estratégias são frequentemente empregadas para esse fim (REYNOLDS; LU, 1997). Uma delas consiste no ajuste cuidadoso dos limites de controle das cartas tradicionais com o objetivo de considerar essa autocorrelação (LU; REYNOLDS JR., 2001; REYNOLDS; LU, 1997). A outra abordagem consiste em utilizar modelos de séries temporais para ajustar os dados, de forma a capturar a autocorrelação, e, posteriormente, aplicar as cartas de controle tradicionais para monitorar os resíduos provenientes desse ajuste.

Para essa abordagem, constantemente são empregados os modelos de médias móveis, os modelos autorregressivos e os autorregressivos de médias móveis. De modo geral, essa abordagem se baseia na suposição de que, se um modelo de série temporal for adequadamente utilizado para descrever os dados autocorrelacionados, os resíduos gerados nessa aplicação (que corresponde à diferença entre as medições do processo e os resultados do modelo de séries temporais) podem ser considerados independentes, podendo ser monitorados por cartas de controle convencionais (KADRI et al., 2016).

Adicionalmente, outras abordagens podem ser utilizadas no monitoramento de dados de processo autocorrelacionados (MA et al., 2018). Uma dessas abordagens consiste em reduzir ou remover a tendência dos dados através da utilização de métodos de *detrending* e de diferenciação, os quais são comumente empregados para transformar séries temporais não estacionárias em estacionárias (SALLES et al., 2019).

### **3 MONITORAMENTO DA PRODUÇÃO DE BIODIESEL POR ROTA ALTERNATIVA UTILIZANDO UM ESPECTRÔMETRO PORTÁTIL NO INFRAVERMELHO PRÓXIMO**

O biodiesel consiste em uma mistura de ésteres alquílicos derivados de ácidos graxos de cadeia longa. Usualmente, é produzido a partir da reação de transesterificação entre triacilgliceróis, principais constituintes dos óleos vegetais e gorduras animais, e um álcool de baixa massa molar. Os ésteres, produtos da reação, possuem uma viscosidade mais baixa do que o óleo vegetal, o que permite que o biodiesel produzido possa ser utilizado diretamente nos motores a diesel (KNOTHE; RAZON, 2017). Por conseguinte, o biodiesel tem sido considerado uma alternativa promissora aos combustíveis de origem fóssil, especialmente para o setor de transporte. Este é utilizado com maior frequência misturado ao diesel de petróleo, mas também pode ser empregado como um combustível puro (CHANG; HWANG; WU, 2017). Além disso, por ser renovável e facilmente biodegradável, é visto como uma opção sustentável, cujo uso pode auxiliar na redução da dependência energética e econômica em relação ao petróleo e na minimização de emissão de alguns poluentes gasosos, especialmente, gases do efeito estufa (CHANG; HWANG; WU, 2017; SINGH et al., 2019).

Devido aos benefícios citados, o biodiesel tem sido utilizado em diversos países (CHANG; HWANG; WU, 2017; KNOTHE; RAZON, 2017). No Brasil, por exemplo, o governo tem estimulado o uso do biodiesel através de legislações que determinam o percentual mínimo desse biocombustível que deve ser acrescentado ao diesel vendido ao consumidor final. Na Resolução nº 16, de 29 de outubro de 2018, da Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), ficou estabelecido um cronograma prevendo um aumento gradual desse percentual de 11% v/v, adotado desde junho de 2019, até 15% v/v, previsto para março de 2023 (CNPE, 2018). Recentemente, em 1º de março de 2020, entrou em vigor o percentual mínimo de 12% v/v, conforme estabelecido pela Resolução (ANP, 2020). Dessa forma, os incentivos governamentais e as vantagens proporcionadas pelo uso do biodiesel tornam necessários os estudos que visem a busca por rotas de produção alternativas e métodos de monitoramento da qualidade do produto e do processo, com o objetivo de torná-lo técnica e economicamente competitivo em comparação aos combustíveis de origem fóssil. Essa necessidade é corroborada, inclusive, pela variabilidade de matéria-prima utilizada na produção do biodiesel, a qual, conforme discutido por Singh et al. (2019), ocasiona uma grande variabilidade na qualidade do produto.

No que tange à busca por rotas de produção alternativas, o processo contínuo de destilação reativa tem se mostrado uma opção promissora para a produção de biodiesel. Esse processo, por integrar as etapas de reação e de separação dos produtos em um único equipamento, tende a proporcionar a obtenção de maiores rendimentos, sem a necessidade de utilizar um grande excesso de álcool. Tal excesso é geralmente requerido nos processos convencionais em batelada com o intuito de deslocar o equilíbrio da reação para favorecer a formação de biodiesel. Quando o processo de destilação reativa é utilizado, no entanto, o equilíbrio é deslocado devido à constante retirada do produto, superando-se, assim, as limitações impostas pelo equilíbrio termodinâmico da reação. Há diversos trabalhos na literatura que descrevem o estudo da destilação reativa para a produção de biodiesel, sendo uma grande parte baseada em simulações do processo (BOON-ANUWAT et al., 2015; PÉREZ-CISNEROS et al., 2016; PETCHSOONGSAKUL et al., 2017; PODDAR; JAGANNATH; ALMANSOORI, 2017; JODA; AHMADI, 2019).

No que diz respeito ao controle de qualidade do biodiesel e do seu processo de produção, os métodos tradicionalmente empregados na análise do produto e monitoramento *off-line* da reação são a cromatografia gasosa (CG) e a cromatografia líquida de alta eficiência (HPLC) ( FREEDMAN; BUTTERFIELD; PRYDE, 1986; FELIZARDO et al., 2006; STAMENKOVIC et al., 2008). A CG, inclusive, é o método de referência indicado pela norma europeia para quantificação do teor de éster em biodiesel (EN14103, 2003). Esses métodos, no entanto, são em geral demorados, caros e não podem ser facilmente empregados no monitoramento *on-line* de reações rápidas como a de transesterificação (TREVISAN et al., 2008). Dessa forma, com o objetivo de melhorar o acompanhamento do processo e a qualidade do biodiesel produzido, vários trabalhos têm demonstrado o potencial da espectroscopia NIR para o monitoramento *in-* e *on-line* dessa reação (KILLNER; ROHWEDDER; PASQUINI, 2011; RICHARD et al., 2011; DE LIMA et al., 2014; SALES et al., 2017; GELINSKI et al., 2018; ROUCHI et al., 2019).

Exceto pelo trabalho de Richard et al. (2013), os trabalhos usando NIR citados acima utilizaram processos em batelada. Richard et al. (2013) descreveram o uso de sondas de reflexão e de transreflectância para o monitoramento *on-line* da produção de biodiesel em microrreatores, que operavam no modo contínuo. Modelos PLS foram utilizados para quantificar o teor de oleato de etila nas misturas reacionais. O objetivo principal do trabalho, no entanto, consistiu na transposição de uma reação de transesterificação de um processo em batelada para o sistema de microrreatores, objetivando, dentre outros aspectos, a coleta de

uma quantidade maior de dados nos primeiros instantes da reação (RICHARD et al., 2013).

Além disso, todos os trabalhos mencionados, incluindo o de Richard et al. (2013), utilizaram espectrofotômetros de bancada para monitorar a produção de biodiesel. Em alguns trabalhos, espectros foram adquiridos na faixa de número de onda de 10.000 a 4.000 cm<sup>-1</sup>, que corresponde à faixa de comprimento de onda de 1.000 a 2.500 nm (RICHARD et al., 2011, 2013; GELINSKI et al., 2018; ROUCHI et al., 2019). Em outros trabalhos, os espectros foram adquiridos na faixa de número de onda de aproximadamente 14.000 a 4.000 cm<sup>-1</sup> (714 – 2.500 nm) ( KILLNER; ROHWEDDER; PASQUINI, 2011; DE LIMA et al., 2014; SALES et al., 2017).

A recente tendência de miniaturização de equipamentos NIR, entretanto, tem motivado a utilização de espectrômetros portáteis para análises envolvendo o biodiesel, ao invés do uso de equipamentos de bancada. Paiva et al. (2015), por exemplo, descreveram o uso de um espectrômetro comercial portátil na região do NIR (MicroNIR™ Pro 1700, VIAVI Solutions) para estimar o teor de biodiesel e de óleo vegetal em misturas de biodiesel e diesel. Os autores realizaram um estudo comparativo utilizando um espectrofotômetro de Infravermelho Próximo por Transformada de Fourier (FT-NIR) e obtiveram desempenhos comparáveis para ambos os equipamentos. Em outro trabalho descrito por da Silva et al. (2017), avaliou-se o potencial do MicroNIR para o desenvolvimento de modelos de regressão PLS com o intuito de estimar parâmetros de qualidade da gasolina e misturas de diesel/biodiesel. Os autores também reportaram resultados satisfatórios para o uso do MicroNIR. Por fim, mais recentemente, Correia et al. (2018) utilizaram o MicroNIR para o controle de qualidade de combustíveis. O potencial do espectrômetro foi avaliado em relação à identificação de adulterações em diesel, gasolina do tipo C e etanol hidratado combustível. Dentre outros parâmetros, avaliou-se o teor de biodiesel em amostras de diesel. Em geral, os autores obtiveram resultados promissores, demonstrando o potencial do MicroNIR para ser utilizado no controle de qualidade de combustíveis.

Apesar de já empregado em análises de biodiesel, espectrômetros portáteis na região do NIR ainda não foram utilizados para o monitoramento *on-line* da produção de biodiesel. Esse é, portanto, um aspecto de ineditismo do presente trabalho, o qual descreve o uso do MicroNIR para o monitoramento *on-line* da reação de transesterificação entre etanol e óleo de algodão em um processo contínuo de destilação reativa.

O principal objetivo do trabalho apresentado nesse capítulo (APÊNDICE A) foi avaliar a viabilidade de se utilizar um espectrômetro portátil na região do infravermelho

próximo (MicroNIR), como uma ferramenta de PAT, para monitorar a produção de biodiesel em um processo contínuo de destilação reativa. Para isso, os seguintes objetivos específicos foram definidos:

- a) Estimar a composição da mistura reacional na base da coluna de destilação reativa utilizada para produzir biodiesel a partir de óleo de algodão e etanol;
- b) Definir o caminho óptico mais adequado e as condições experimentais apropriadas para a aquisição de espectros com o MicroNIR;
- c) Coletar e analisar amostras da base da coluna com o MicroNIR e pelos métodos cromatográficos para quantificação de etanol e de éster alquílico;
- d) Avaliar a possibilidade de utilizar misturas sintéticas para compor o conjunto de calibração;
- e) Desenvolver modelos de regressão PLS para quantificar os teores de etanol, glicerol e ésteres alquílicos na base da coluna e avaliar o potencial desses modelos para o monitoramento da produção de biodiesel na coluna de destilação reativa.

### 3.1 METODOLOGIA

A metodologia aplicada para o desenvolvimento deste trabalho está detalhada a seguir.

#### 3.1.1 Produção de biodiesel pelo processo de destilação reativa

As reações de transesterificação para a produção de biodiesel foram conduzidas em uma coluna de destilação reativa em escala laboratorial (coluna de vidro com 150 cm de altura e 3 cm de diâmetro), empacotada com anéis de cobre (diâmetro de 0,7 cm). Ela é composta por pontos de alimentação de óleo e de catalisador no topo e um ponto de alimentação de álcool no fundo. O sistema reacional contém também um refervedor na base, utilizado para a geração de um fluxo de vapor de álcool, o qual era alimentado no fundo da coluna e transferia calor à medida que a percorria em sentido ascendente. Adicionalmente, na parte superior da coluna, há um condensador do tipo total, utilizado para recuperar o álcool não reagido que saía pelo topo.

A coluna podia operar em diversas razões molares álcool/óleo, uma vez que os sistemas de bombeamento de ambos os reagentes permitiam o ajuste de suas vazões de

alimentação. Essas vazões, inclusive, podiam apresentar pequenas variações durante a operação devido à baixa precisão de seus medidores. Nesse trabalho, entretanto, foi adotada uma condição operacional que resultou em um excesso de álcool na base da coluna. No total, foram realizados três experimentos nessa condição. No primeiro deles, alíquotas foram coletadas da base da coluna e analisadas pelo MicroNIR no modo *off-line*, com o intuito de selecionar o caminho óptico mais adequado para a aquisição das medidas e definir as condições de análise. Nos outros dois experimentos, as aquisições espectrais com o MicroNIR foram realizadas no modo *on-line*. Adicionalmente, as alíquotas retiradas da base da coluna foram analisadas também por cromatografia gasosa para estimar os teores de etanol e de éster alquílico. Do total de amostras retiradas, 12 foram empregadas no desenvolvimento dos modelos de calibração.

Nas reações, etanol p.a. (Neon, pureza de 99,85%) e óleo de algodão comercial (Flor de Algodão, adquirido em mercado local) foram utilizados como reagentes. Em algumas situações, ainda, utilizou-se óleo de algodão fornecido pela planta experimental de biodiesel do Centro de Tecnologias Estratégicas do Nordeste (CETENE, Caetés, Pernambuco). Hidróxido de sódio (Dinâmica, pureza mínima de 98%) dissolvido em etanol foi utilizado como catalisador em uma concentração de 1% (m/m) em relação à quantidade de óleo vegetal, aproximadamente.

### **3.1.2 Análises cromatográficas**

Os teores de éster alquílico e de etanol das amostras retiradas da coluna (ponto monitorado pelo MicroNIR) foram estimados utilizando-se um cromatógrafo gasoso, modelo Shimadzu 17-A, com injeção automática e detector de ionização de chama (GC-FID). Do total do volume de cada alíquota, uma parte era enviada para a determinação do teor de etanol, cuja análise era realizada, preferencialmente, no mesmo dia de coleta. À outra parte era adicionado, aproximadamente, 10% v/v de ácido acético glacial, com o intuito de interromper a reação. A mistura era, então, submetida a uma etapa de purificação e mantida sob refrigeração para posterior determinação do teor de éster alquílico por cromatografia.

As análises cromatográficas para estimar a concentração de éster foram realizadas de acordo com a norma EN14103 (EN14103, 2003), utilizando o padrão interno C17 (heptadecanoato de metila) e a coluna capilar *Carbowax*. Um injetor do tipo split/splitless e um detector de ionização de chama foram empregados. A etapa de purificação anterior à

análise consistiu em: lavagens sucessivas com água destilada seguidas de centrifugação e retirada da fase aquosa; secagem com adição de sulfato de sódio anidro (10% (m/m)); e filtração utilizando um papel de filtro qualitativo, um Kitasato e uma bomba a vácuo, conforme descrito por de Lima et al. (2014). Ao menos três lavagens foram realizadas para cada alíquota.

Por falta de uma norma para determinar o teor de etanol em misturas com biodiesel, contendo um elevado teor de etanol, alguns testes foram realizados para otimizar um método cromatográfico que pudesse ser utilizado para esse fim. As análises foram conduzidas em uma coluna capilar ValcoBond VB-1 (fase estacionária 100% dimetilpolisiloxano). A temperatura da coluna para o método otimizado foi programada com o objetivo de se obter uma separação clara para o pico do etanol: a temperatura inicial foi de 50°C, mantida por 6 min, seguida de uma rampa a 30°C/min até 300°C, permanecendo nessa temperatura por 20 min. A curva analítica foi construída com base em soluções padrões contendo biodiesel e etanol p.a., variando-se a concentração de etanol de 50 a 90% (m/m). Cada solução foi analisada em triplicata e um valor de  $R^2$  igual a 99,32% foi obtido para o ajuste linear. O desvio padrão estimado sob condições de repetibilidade ( $N=12$ ) para esse método foi de 1,4% (m/m).

### **3.1.3 Aquisição *off-line* de dados espectroscópicos com o MicroNIR e seleção do caminho óptico**

Neste trabalho, um espetrômetro comercial portátil (MicroNIR<sup>TM</sup> Pro 1700, Viavi Solutions), acoplado a um acessório de transmitância, foi utilizado para a aquisição dos dados espectroscópicos. Esse equipamento contém um Filtro Linear Variável e um arranjo de detector InGaAs de 128 pixel, que permite que os valores de absorbância sejam determinados de forma simultânea para todos os comprimentos de onda. A faixa espectral de operação do MicroNIR comprehende a região entre 908 e 1676 nm e, segundo o fabricante, sua resolução óptica é menor do que 1,25% (geralmente, 1%) do comprimento de onda central (por exemplo, para 1000 nm, a resolução seria menor do que 12,5 nm).

Para a aquisição dos dados no modo *off-line*, utilizou-se um acessório de transmitância desenvolvido no Departamento de Química Analítica da UNICAMP, sob a supervisão do Prof. Dr. Jarbas José Rodrigues Rohwedder, semelhante ao descrito por Paiva et al. (2015). O acessório permite a utilização de cubetas de diferentes tamanhos e, portanto, foi utilizado para a seleção do caminho óptico mais adequado para a análise das amostras retiradas da base da

coluna. Nesse estudo, cubetas de quartzo de 5, 10 e 20 mm foram testadas. Os espectros foram adquiridos em toda faixa espectral do MicroNIR, registrados como uma média de 50 varreduras e com tempo de integração de 50 ms. Para todas as medidas, o 100% de transmitância foi adquirido utilizando-se uma cubeta limpa e vazia e o 0% foi obtido bloqueando-se a passagem da luz entre a fonte externa do acessório (mini lâmpada de tungstênio) e o espectrômetro. Tanto o acessório quanto o MicroNIR foram alimentados por um cabo USB conectado a um laptop, em que os dados foram registrados pelo *software* do instrumento.

O caminho óptico selecionado para a análise das amostras da coluna foi também utilizado na aquisição de espectros das misturas sintéticas. Para cada mistura, as análises foram realizadas em triplicata, no modo *off-line*, seguindo-se o mesmo procedimento descrito. Os espectros médios de cada mistura foram, então, utilizados na construção e validação dos modelos de calibração.

### 3.1.4 Misturas sintéticas

Como a composição da mistura reacional na base de uma coluna de destilação reativa praticamente não varia depois de atingido o estado estacionário, avaliou-se a possibilidade de utilizar amostras preparadas em laboratório (chamadas de misturas sintéticas) para a construção dos modelos de calibração. A utilização de misturas sintéticas também torna o processo de calibração muito mais prático. Com essa finalidade, um planejamento de misturas foi inicialmente adotado para a preparação de misturas ternárias compostas de etanol, óleo de algodão e biodiesel. A faixa de concentração de cada componente foi estabelecida de acordo com a composição da mistura reacional na base da coluna de destilação reativa e de modo a garantir variabilidade suficiente para os modelos de calibração. A partir do planejamento com pontos adicionais, 70 misturas ternárias foram inicialmente preparadas e analisadas com o MicroNIR em temperatura ambiente ( $22 \pm 2^\circ\text{C}$ ). O diagrama ternário com as concentrações dessas misturas está apresentado no APÊNDICE B.

Em seguida, com o objetivo de minimizar as diferenças entre os espectros das misturas sintéticas e das amostras retiradas da coluna, 38 novas misturas foram preparadas, analisadas com o MicroNIR e inseridas nos modelos de calibração. Algumas delas foram preparadas incluindo o glicerol (pureza de 99,5%, Synth), além dos outros três componentes já utilizados. Para algumas dessas misturas quaternárias, respeitou-se a relação estequiométrica entre o

glicerol e o éster alquílico da reação global de transesterificação (1 mol de glicerol para cada 3 mol de éster alquílico). Para outras, no entanto, essa relação não foi levada em consideração, com o objetivo de evitar possíveis correlações indiretas na etapa de modelagem. Por fim, algumas misturas ternárias e quaternárias foram ainda analisadas na mesma temperatura operacional da base da coluna, a qual pode variar em entre 50 a 60°C. As amostras restantes foram analisadas a temperatura ambiente ( $22 \pm 2^\circ\text{C}$ ).

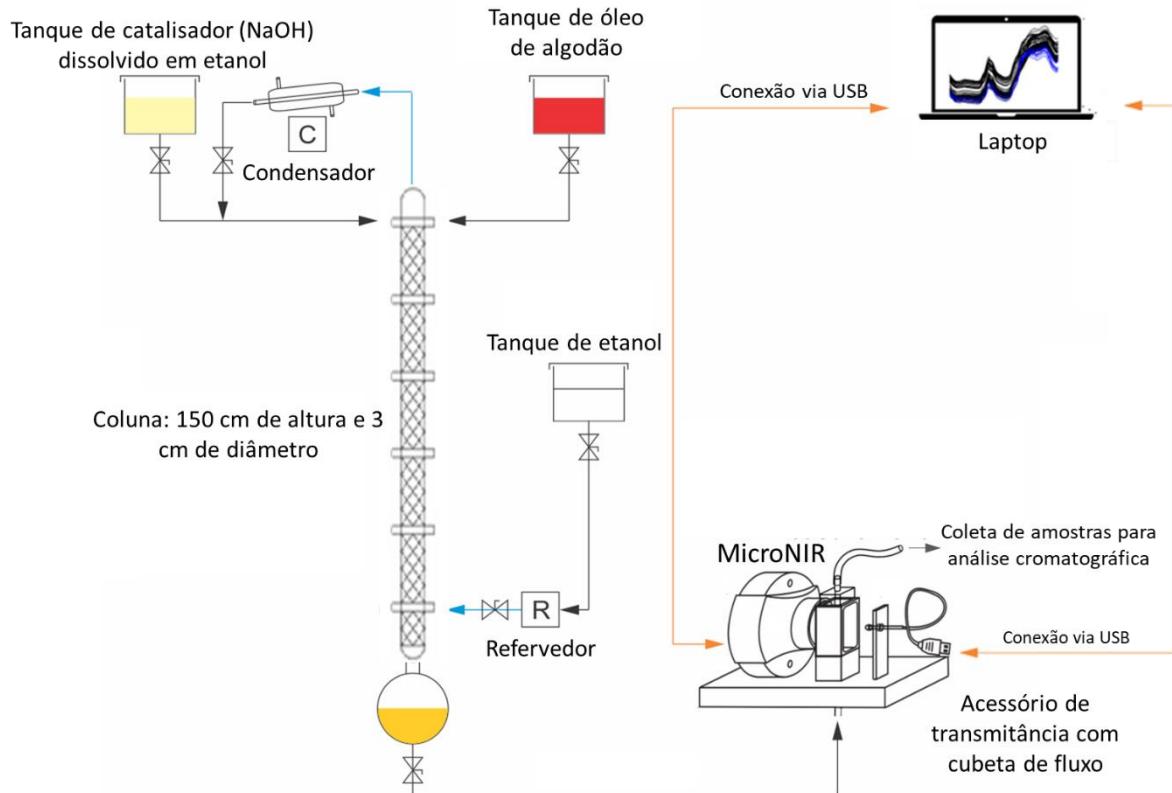
Dessa forma, um total de 108 misturas sintéticas foram preparadas e analisadas com o MicroNIR. As composições dessas misturas, em termos de % (m/m), bem como a temperatura de análise de cada uma, estão apresentadas no APÊNDICE C. Na preparação de cada mistura, a quantidade previamente calculada de cada componente foi pesada em um frasco âmbar, utilizando uma balança analítica Sartorius (modelo BL210S), com precisão de 0,0001 g. Foram pesados em sequência, do menos volátil (glicerina ou óleo) até o mais volátil (etanol). O mesmo etanol e óleo de algodão utilizados na reação de transesterificação, descrita anteriormente, foram usados nessa etapa. Além disso, com o intuito de evitar a ocorrência de erros sistemáticos no preparo das misturas, quatro amostras de biodiesel de algodão com diferentes teores de éster alquílico foram utilizadas (77,8% (m/m), 81,5% (m/m), 90,5% (m/m) e 95,0% (m/m)). Essas amostras foram provenientes da produção de biodiesel em bateladas realizadas no laboratório, empregando-se diferentes condições experimentais. O seguinte procedimento geral foi adotado para cada batelada: inicialmente, o óleo de algodão foi introduzido ao reator e submetido à agitação e aquecimento até alcançar a temperatura de interesse; posteriormente, a solução contendo catalisador (NaOH) previamente dissolvido em etanol foi adicionada ao reator, mantendo-se a reação por aproximadamente 60 min. No fim, o biodiesel foi purificado e analisado por cromatografia gasosa para estimar o teor de éster alquílico, seguindo o mesmo procedimento descrito para as amostras da coluna.

### 3.1.5 Monitoramento *on-line* da produção de biodiesel com o MicroNIR

Uma vez que o caminho óptico foi selecionado e as condições de análise com o MicroNIR no modo *off-line* foram estabelecidas, uma pequena adaptação foi realizada ao acessório de transmitância com o objetivo de utilizar uma cubeta de quartzo de fluxo, de modo a permitir o monitoramento *on-line* da produção de biodiesel. Para essas análises, foram aplicadas as mesmas configurações estabelecidas na aquisição de medidas *off-line*, descritas anteriormente. A diferença é que, no modo *on-line*, a mistura reacional da base da coluna era

continuamente enviada à célula de fluxo, onde os espectros eram coletados automaticamente a cada segundo, utilizando o *software* do MicroNIR. Um esquema da coluna de destilação reativa e do monitoramento *on-line* com o MicroNIR está ilustrado na Figura 5.

Figura 5 – Esquema da coluna de destilação reativa e do monitoramento *on-line* com o MicroNIR



Fonte: A Autora (2020)

### 3.1.6 Pré-processamento dos dados e desenvolvimento dos modelos de calibração

Na etapa de desenvolvimento dos modelos de regressão PLS, alguns pré-processamentos foram avaliados: SNV, MSC, EMSC, 1<sup>a</sup> derivada com filtro Savitzky-Golay e SNV, seguido de OSC. Os dados pré-processados foram também centrados na média para a construção dos modelos, com o objetivo de quantificar os teores de etanol, éster alquílico e glicerol das amostras da coluna.

Na construção dos modelos para quantificação de etanol e éster, 104 misturas sintéticas foram separadas em um conjunto de calibração e predição, aproximadamente 70% para a calibração e 30% para a predição. Esses conjuntos foram selecionados utilizando-se o algoritmo SPXY (GALVÃO et al., 2005). Já na construção do modelo PLS para a

quantificação de glicerol, apenas 38 misturas sintéticas foram utilizadas. Os conjuntos de calibração e de validação também foram selecionados utilizando-se o SPXY, de modo que 30 amostras foram separadas para a calibração e 8 para a previsão. Após a seleção, no entanto, 2 misturas separadas para a calibração foram enviadas para o conjunto de validação, com o intuito de obter uma melhor distribuição dos dados ao longo da faixa de concentração do glicerol.

Adicionalmente, além das misturas sintéticas, para cada modelo, 3 amostras coletadas da coluna foram adicionadas ao conjunto de calibração. As nove amostras da coluna remanescentes foram incluídas no conjunto de predição desses modelos. Com o objetivo de sintetizar essas informações, a quantidade de amostras da coluna e de misturas sintéticas utilizadas no desenvolvimento de cada modelo, antes da exclusão de *outliers*, está apresentada na Tabela 1.

Tabela 1 - Número de misturas sintéticas e amostras da coluna aplicadas às etapas de calibração e predição para o desenvolvimento de modelos PLS para quantificar etanol, éster alquílico e glicerol, antes da exclusão de *outliers*

Componente	Etapa de calibração		Etapa de predição	
	Misturas sintéticas	Amostras da coluna	Misturas sintéticas	Amostras da coluna
Etanol	73	3	31	9
Ester alquílico	73	3	31	9
Glicerol	28	3	10	9

Fonte: A Autora (2020)

Na construção dos modelos PLS, as composições conhecidas das misturas sintéticas (APÊNDICE C) foram empregadas como valores de referência (vetor  $\mathbf{y}$ ). Para as amostras da coluna, os valores de referência para o etanol foram obtidos diretamente por análise cromatográfica, conforme descrito no tópico 3.3.2. Já os valores de referência do éster alquílico e do glicerol, para cada amostra, foram calculados considerando-se os seguintes dados: teor de etanol obtido por cromatografia, teor de éster alquílico obtido por cromatografia após purificação (que corresponde ao teor de éster da mistura de biodiesel e óleo de algodão não reagido), e razão estequiométrica entre o glicerol e o éster alquílico da reação global de transesterificação, conforme descrito no APÊNDICE D.

Na construção de todos os modelos, a detecção e eliminação de amostras anômalas (*outliers*) foi conduzida utilizando-se os gráficos dos resíduos. A seleção do número ótimo de

variáveis latentes baseou-se na técnica de validação cruzada *venetian blinds*, aplicada na etapa de validação interna dos modelos. Por fim, todas as análises quimiométricas e tratamento dos dados espectroscópicos foram realizados utilizando o *software* MATLAB® (R2010a 7.10.0.499, MathWorks) e os modelos foram desenvolvidos utilizando o PLS toolbox (Eigenvector Research Inc., EUA). O cálculo do limite de detecção dos modelos foi realizado através da interface MVC1\_GUI, que funciona em ambiente MATLAB (CHIAPPINI; GOICOECHEA; OLIVIERI, 2020).

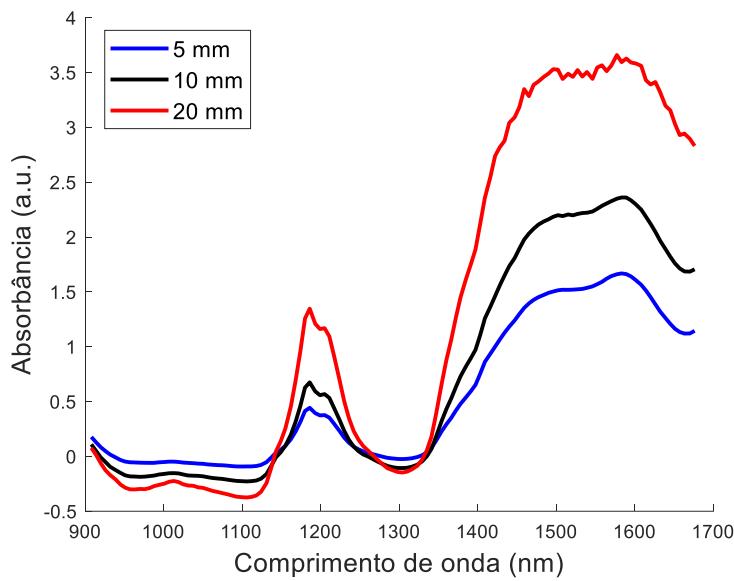
### 3.2 RESULTADOS E DISCUSSÃO

Os principais resultados obtidos estão apresentados e discutidos a seguir.

#### **3.2.1 Seleção do caminho óptico e estimativa da composição da mistura reacional no ponto monitorado com o MicroNIR**

Inicialmente, algumas alíquotas foram retiradas da base da coluna de destilação reativa, após estabilização da coluna, com o intuito de definir as condições adequadas para a aquisição espectral, incluindo a seleção do caminho óptico, e de estimar a composição da mistura reacional nesse ponto. Essas alíquotas foram analisadas com o MicroNIR, no modo *off-line*, utilizando cubetas de quartzo com diferentes caminhos ópticos (5, 10 e 20 mm). Os espectros adquiridos de uma dessas alíquotas utilizando as três cubetas estão mostrados na Figura 6.

Figura 6 - Espectros adquiridos utilizando cubetas de quartzo de diferentes caminhos ópticos



Fonte: A Autora (2020)

É possível perceber através da Figura 6 que o espectro obtido com a cubeta de 20 mm (em vermelho) apresentou sinais de absorção saturados na região entre 1400 e 1700 nm, aproximadamente. Já que a banda de formato alargado nessa região é devida ao primeiro sobretono do grupamento O-H (WORKMAN; WEYER, 2012), a saturação do sinal foi atribuída, principalmente, ao excesso de etanol na mistura reacional. Já o espectro adquirido com a cubeta de 5 mm (em azul), apresentou valores mais baixos de absorbância. Particularmente, para os comprimentos de onda em torno de 1160 e 1450 nm, esses baixos valores poderiam dificultar a quantificação do teor de éster alquílico, já que esses comprimentos de onda são atribuídos, respectivamente, ao quarto e terceiro sobretons do grupamento C=O (WORKMAN; WEYER, 2012). Dessa forma, com base nessas observações, o caminho óptico de 10 mm foi selecionado para a aquisição espectral utilizando o MicroNIR, tanto no modo *off-line*, quanto no modo *on-line*.

Posteriormente, com o intuito de estimar a composição da mistura reacional na base da coluna, as alíquotas coletadas nessa primeira etapa foram tratadas, conforme procedimento descrito no tópico 3.3.2, e analisadas pelos métodos cromatográficos para a quantificação dos teores de éster alquílico (biodiesel) e etanol. O teor de éster para essas amostras ficou em torno de 85% (m/m), considerando apenas a fase contendo biodiesel e óleo não reagido, e o teor de etanol ficou em torno de 80 e 90% (m/m). Esse estudo preliminar foi essencial para definir as faixas de concentração dos modelos de calibração e orientar na preparação das misturas sintéticas.

### 3.2.2 Análises cromatográficas das amostras da coluna

Após a seleção do caminho óptico e estimativa da composição da mistura reacional na base da coluna, prosseguiu-se com a preparação das misturas sintéticas e, em paralelo, com o monitoramento *on-line* da produção de biodiesel. Das amostras da coluna coletadas durante os experimentos e analisadas por cromatografia, 12 foram utilizadas na construção dos modelos de calibração. Outras foram excluídas por estarem fora das faixas de concentração dos modelos. Os teores de etanol e de éster alquílico (considerando apenas a fase contendo biodiesel e óleo de algodão não reagido), obtidos por cromatografia, para essas 12 amostras estão apresentados na Tabela 2.

Tabela 2 – Teores de etanol e éster alquílico, em % (m/m), das amostras retiradas da coluna, obtidos por cromatografia gasosa

Amostra	Teor de etanol obtido por CG (% (m/m))	Teor de éster alquílico obtido por CG após purificação (% (m/m))
C01	80,0	86,0
C02	80,2	82,6
C03	81,6	91,0
C04	83,2	90,4
C05	81,2	85,9
C06	84,9	88,9
C07	89,4	89,2
C08	72,4	92,6
C09	85,9	91,7
C10	86,0	89,9
C11	82,9	91,9
C12	88,3	87,2

Fonte: A Autora (2020)

Os teores de etanol e de éster alquílico das amostras da coluna, apresentados na Tabela 2, estão de acordo com as faixas de concentração dos modelos de calibração desenvolvidos.

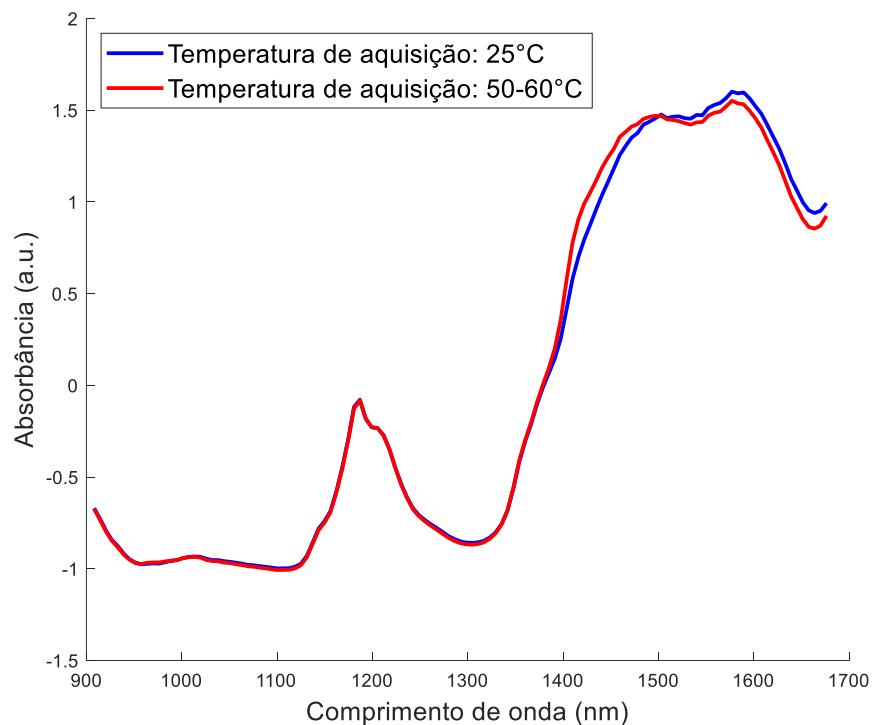
### 3.2.3 Misturas sintéticas

Como a composição da mistura reacional na base da coluna de destilação reativa praticamente não varia após atingido o estado estacionário, as misturas sintéticas foram utilizadas com o objetivo de ampliar as faixas de concentração dos modelos. Embora alguns pequenos ajustes tenham sido realizados nas condições operacionais da coluna com o intuito

de ampliar a faixa de concentração das amostras retiradas da base (Tabela 2), essa variação, por si só, não fornece variabilidade suficiente para a etapa de modelagem, requerendo, portanto, o uso das misturas sintéticas. Além disso, é importante ressaltar que esse tipo de flexibilidade operacional, em geral, não é possível em ambientes industriais. Por fim, o emprego dessas misturas trouxe também simplicidade do ponto de vista experimental, uma vez que reduziu a necessidade de fazer análises cromatográficas, as quais são demoradas, caras e requerem uma etapa de purificação das amostras.

A princípio, por questão de simplicidade, optou-se por preparar misturas ternárias contendo etanol, biodiesel e óleo de algodão. O glicerol foi inicialmente desconsiderado por ser produzido em uma quantidade menor quando comparado ao biodiesel e, principalmente, ao etanol presente em excesso. No entanto, depois de desenvolvidos e analisados os primeiros modelos, sentiu-se a necessidade de melhorar a representatividade das misturas, minimizando as diferenças entre os espectros das amostras coletadas na base da coluna e os espectros das misturas sintéticas. Com esse objetivo, 38 novas misturas foram preparadas, incluindo amostras quaternárias contendo glicerol e amostras analisadas na mesma temperatura de operação da base da coluna (em torno de 50 e 60°C). A princípio, o efeito da temperatura tinha sido considerado insignificante devido à perda de calor que ocorre entre a parte inferior da coluna, de onde sai a amostra, e a posição do MicroNIR no sistema de medição *on-line*. No entanto, os resultados obtidos ao longo do desenvolvimento desse trabalho indicaram a importância de considerar o efeito da temperatura. De forma ilustrativa, estão apresentados na Figura 7 os espectros médios, pré-processados com SNV, das misturas sintéticas 97 e 98 (APÊNDICE C), de mesma composição, mas cujos os espectros foram adquiridos, respectivamente, na temperatura ambiente ( $22 \pm 2^\circ\text{C}$ ) (em azul) e na faixa de temperatura entre 50 e 60°C (em vermelho).

Figura 7 – Espectros médios, pré-processados com SNV, de misturas sintéticas de mesma composição (97 e 98 – APÊNDICE C), adquiridos, respectivamente, na temperatura ambiente ( $22 \pm 2^{\circ}\text{C}$ ) (em azul) e na faixa de temperatura entre 50 e  $60^{\circ}\text{C}$  (em vermelho)

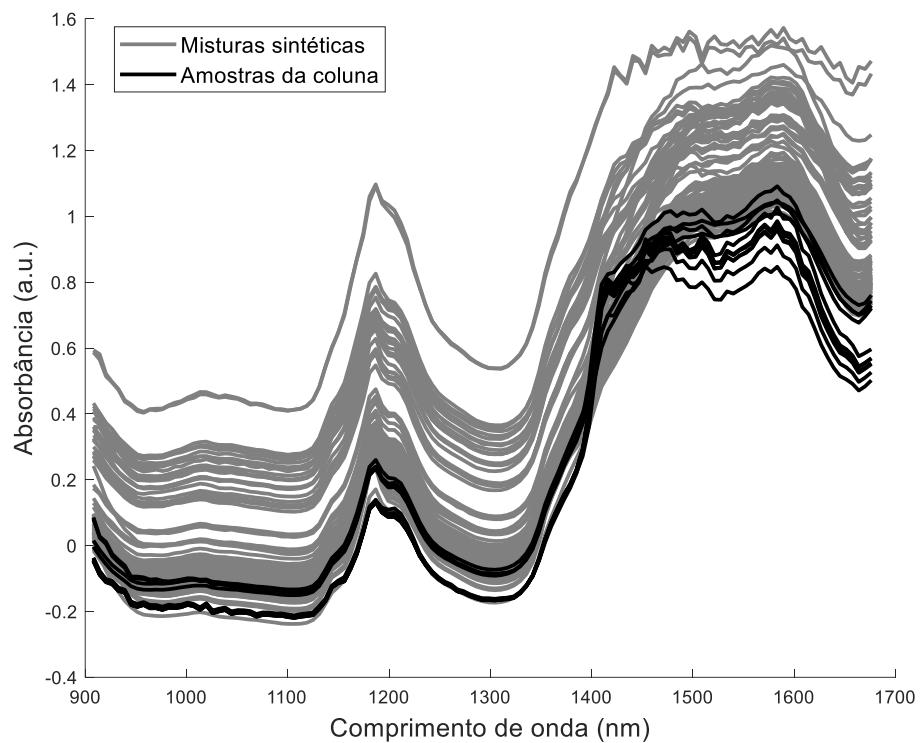


Fonte: A Autora (2020)

Conforme observado na Figura 7, a faixa espectral que foi mais afetada pelo aumento da temperatura foi a região correspondente ao primeiro sobretom do grupamento O-H, presente no etanol e no glicerol.

Para fins de comparação, os espectros médios das 108 misturas sintéticas e das 12 amostras coletadas na base da coluna de destilação reativa, antes da exclusão de *outliers*, estão apresentados na Figura 8.

Figura 8 – Espectros médios das 108 misturas sintéticas preparadas (em cinza) e das 12 amostras retiradas da coluna de destilação reativa (em preto), antes da exclusão dos outliers



Fonte: A Autora (2020).

É possível perceber pela Figura 8 que, apesar da inclusão das novas amostras, ainda existem diferenças entre os espectros dos dois conjuntos de dados, especialmente em torno de 1400 nm, as quais devem ser minimizadas com a aplicação de técnicas de pré-processamento adequadas. Essas técnicas devem ser utilizadas também para corrigir os efeitos de espalhamento da radiação, os quais são atribuídos, principalmente, à emulsão formada durante a reação e à heterogeneidade de algumas misturas sintéticas, sobretudo aquelas com alta concentração de glicerol e de óleo de algodão. Além disso, a passagem de pequenas bolhas de ar na cubeta de fluxo durante a aquisição espectral *on-line* também pode ser uma das causas do espalhamento da radiação e do aspecto ruidoso dos espectros.

### 3.2.4 Pré-processamento dos dados e desenvolvimento dos modelos de calibração

Na etapa de construção dos modelos, algumas técnicas de pré-processamento foram testadas com o intuito de corrigir os efeitos de espalhamento da radiação e minimizar as diferenças entre os espectros das misturas sintéticas e os espectros das amostras da coluna, conforme discutido anteriormente. Os seguintes métodos de pré-processamento foram

testados com essa finalidade: SNV, MSC, EMSC, primeira derivada com filtro Savitzky-Golay (com janela de 7 pontos e polinômio de 2<sup>a</sup> ordem) e OSC.

Antes da aplicação das técnicas citadas, as regiões entre 908-1112 nm e 1400-1676 nm dos espectros utilizados na construção dos modelos para quantificação de etanol e glicerol foram suavizados com filtro Savitzky-Golay (janela de 9 pontos e polinômio de 2<sup>a</sup> ordem), com o intuito de amenizar o aspecto ruidoso dos espectros nessas regiões. O filtro não foi aplicado à região central para não haver perda de informação no pico em torno de 1200 nm, que é atribuído ao segundo sobretom de estiramento C-H.

Após a aplicação das técnicas de pré-processamento, os espectros foram centrados na média para desenvolvimento de modelos de regressão PLS para quantificação de etanol, éster alquílico e glicerol. Os resultados das etapas de calibração e de predição dos modelos desenvolvidos estão apresentados na Tabela 3. Os modelos para etanol e éster alquílico foram construídos levando-se em consideração toda a faixa espectral do MicroNIR, que vai de 908 a 1676 nm. Já para a quantificação do glicerol, os modelos foram desenvolvidos considerando-se apenas a região entre 1348 e 1676 nm, a qual abrange, principalmente, a faixa associada ao glicerol e ao etanol (grupamento O-H).

Tabela 3 – Resultados para as etapas de calibração e de predição dos modelos de regressão PLS para quantificação dos teores de etanol, glicerol e éster alquílico

Composto	Faixa de concentração (% (m/m))	Pré-processamento	VL	RMSEC (% (m/m))	RMSECV (% (m/m))	R <sup>2</sup> <sub>cv</sub>	RMSEP (% (m/m))	R <sup>2</sup> <sub>p</sub>
Etanol	70 – 90	EMSC	5	0,95	1,25	0,93	1,66	0,75
		SNV	6	1,00	1,36	0,91	1,48	0,80
		MSC	6	1,07	1,44	0,90	1,34	0,84
		1 <sup>a</sup> derivada	5	1,05	1,33	0,92	1,81	0,71
		SNV e OSC	2	0,70	1,27	0,93	1,79	0,70
Glicerol	0 – 7,5	EMSC	2	0,76	0,86	0,90	1,08	0,73
		SNV	5	0,50	0,78	0,93	1,29	0,73
		MSC	4	0,56	0,80	0,92	1,25	0,73
		1 <sup>a</sup> derivada	3	0,76	0,89	0,90	0,94	0,76
		SNV e OSC	2	0,36	0,51	0,96	0,98	0,76
Éster alquílico	0 – 23	EMSC	7	2,06	2,65	0,79	2,15	0,79
		SNV	6	2,45	2,95	0,74	2,79	0,64
		MSC	6	2,49	3,04	0,72	2,83	0,63
		1 <sup>a</sup> derivada	6	2,71	3,13	0,71	3,09	0,60
		SNV e OSC	2	2,03	3,42	0,66	3,05	0,58

Fonte: A Autora (2020).

A partir das métricas de desempenho apresentadas na Tabela 3, foi possível selecionar a estratégia de pré-processamento mais adequada para a construção dos modelos PLS e discussão dos resultados. Isso foi feito de modo a priorizar, se possível, a escolha de um único pré-processamento para os três modelos apresentados neste trabalho (para quantificação de etanol, éster alquílico e glicerol).

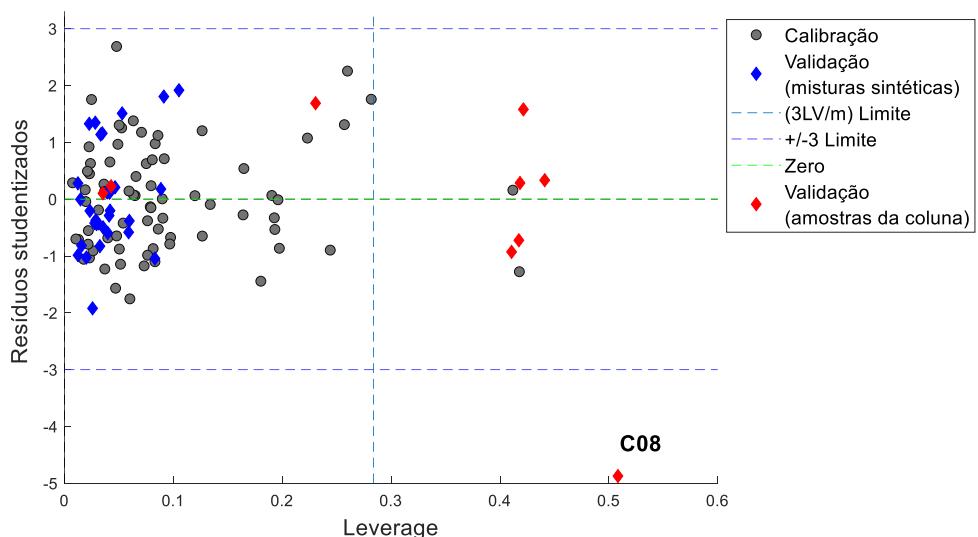
No que diz respeito à quantificação de etanol, os valores de RMSEP dos modelos desenvolvidos com SNV, MSC, 1<sup>a</sup> derivada e SNV seguido de OSC, não apresentaram diferença estatística em relação ao modelo obtido com EMSC, com base em um teste F a um nível de 95% de confiança. O modelo obtido com SNV e OSC foi desenvolvido com um número menor variáveis latentes (VL). No entanto, essa estratégia de pré-processamento não foi selecionada pois, para a quantificação dos três compostos, os modelos desenvolvidos não apresentaram uma predição satisfatória para as amostras da coluna: os valores de RMSEP para a predição das amostras da coluna foram superiores aos valores de RMSEP para a predição das misturas sintéticas (teste F, 95% de confiança). Dessa forma, EMSC e 1<sup>a</sup> derivada poderiam ser utilizadas, já que nesses casos, 5 variáveis latentes foram requeridas para a construção dos modelos. No entanto, a técnica EMSC foi selecionada por ser considerada mais simples do ponto de vista de implementação e interpretação dos resultados, uma vez que mantém o formato dos espectros brutos. Na construção do modelo com EMSC, após a eliminação de *outliers*, 71 misturas sintéticas e 3 amostras da coluna (C10 - C12), totalizando 74 amostras, foram utilizadas na etapa de calibração. Já na etapa de predição, 31 amostras sintéticas e 9 amostras da coluna (C01 - C09) foram consideradas. Os resultados obtidos para esse modelo foram considerados satisfatórios tanto para a calibração (RMSEC de 0,95% (m/m)) quanto para a predição (RMSEP de 1,66% (m/m)).

Para a quantificação do teor de glicerol, o valor de RMSEP do modelo desenvolvido com EMSC não apresentou diferença estatisticamente significativa, com base em um teste F a um nível de 95% de confiança, dos valores de RMSEP dos modelos desenvolvidos com as outras técnicas. O modelo com EMSC foi escolhido, no entanto, pelo fato de ter sido desenvolvido com um número menor de variáveis latentes (VL). Após a investigação e exclusão de *outliers*, esse modelo foi construído utilizando-se 28 amostras no conjunto de calibração: 25 misturas sintéticas e 3 amostras da coluna (C10 à C12). Além disso, 18 amostras foram utilizadas para a etapa de predição, incluindo 9 amostras da coluna (C01 à C09) e 9 misturas sintéticas. Com base nas métricas de desempenho apresentadas, resultados

satisfatórios foram obtidos tanto para a etapa de calibração ( $RMSEC = 0,76\% (m/m)$ ) quanto para a de predição ( $RMSEP = 1,08\% (m/m)$ ).

Quanto aos modelos para quantificação de éster alquílico, exceto para a técnica SNV, os valores de  $RMSEP$  dos modelos desenvolvidos com os outros pré-processamentos não são estatisticamente semelhantes aos obtido com EMSC, com base em um teste F de 95% de confiança. Nesse caso, embora a construção do modelo com SNV tenha requerido menos variáveis latentes, com o objetivo de unificar o pré-processamento utilizado nos três modelos PLS apresentados nesse trabalho, a técnica EMSC foi selecionada. Dessa forma, o modelo com EMSC, construído com 7 variáveis latentes, apresentou resultados aceitáveis para as etapas de calibração ( $RMSEC$  de 2,06% (m/m)) e de predição ( $RMSEP$  de 2,15% (m/m)). Na construção desse modelo, após a retirada de *outliers*, 74 amostras foram utilizadas para a calibração, incluindo 3 amostras da coluna (C10-C12). Já na predição, 31 misturas sintéticas e 8 amostras da coluna foram utilizadas. Das 9 amostras da coluna (C01 à C09), uma foi excluída (C08) por apresentar um valor de resíduo studentizado muito negativo quando comparado ao das outras amostras, conforme ilustrado no gráfico de *leverage* versus resíduos studentizados da Figura 9 para o modelo PLS com os dados pré-processados com EMSC. De acordo com o apresentado em Burns e Ciurczak (2008), os resíduos studentizados indicam quão bem o modelo consegue prever a propriedade de interesse para uma determinada amostra.

Figura 9 – Resíduos studentizados vs. *leverage* para as amostras de calibração (em cinza) e de predição (misturas sintéticas em azul e amostras da coluna em vermelho) para o modelo para quantificação de éster alquílico construído com os dados pré-processados com EMSC

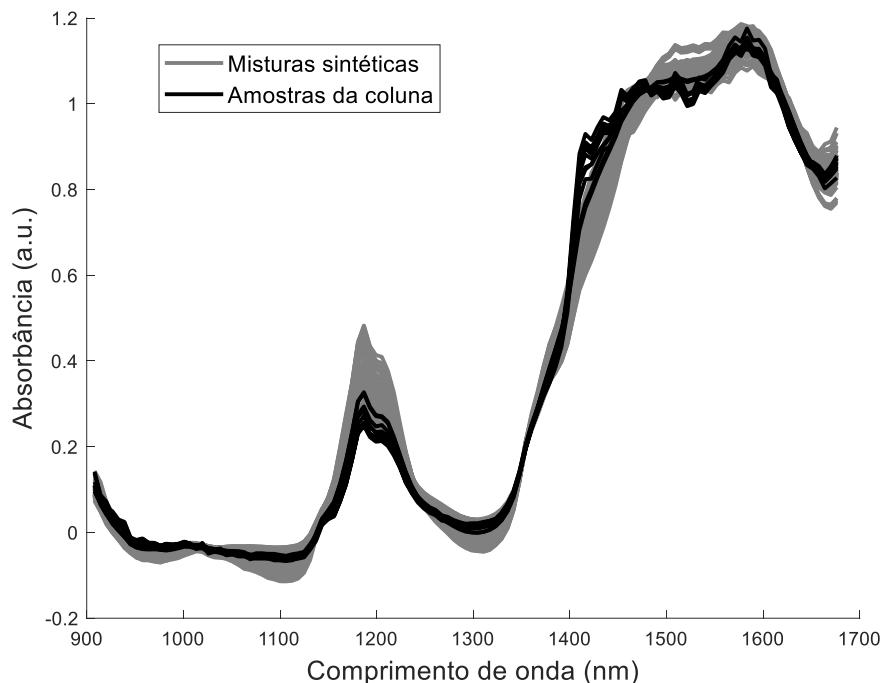


Fonte: A Autora (2020)

Para comparação, Richard et al. (2013) obtiveram valores de RMSEP de 4,10% (m/m) e 3,52% (m/m) quando utilizaram sondas de reflexão e de transflectância, respectivamente, para estimar o teor de oleato de etila na mistura reacional durante a produção de biodiesel. A faixa de concentração adotada para o oleato de etila variou entre zero e 100% (m/m). Para o monitoramento com a sonda de reflexão, 25 amostras foram utilizadas na etapa de predição do modelo, o qual foi desenvolvido considerando uma razão molar etanol/óleo variando de 6 a 45,4. Já para o monitoramento com a sonda de transflectância, 21 amostras constituíram o conjunto de predição e o modelo de calibração foi construído considerando uma razão molar etanol/óleo de 45,4. Dessa forma, o modelo apresentado no presente trabalho para a quantificação de éster alquílico apresentou valor de RMSEP mais baixo quando comparado aos valores obtidos por Richard et al. (2013). No entanto, como o número de amostras utilizadas e as faixas de concentração foram diferentes nos dois trabalhos, não se pode fazer uma comparação direta entre esses resultados.

Diante do exposto, a técnica EMSC foi selecionada para a construção dos três modelos PLS desenvolvidos nesse trabalho. É importante ressaltar que esses modelos não apresentaram *bias* estatisticamente significativo, de acordo com um teste t, em um nível de confiança de 95% (*bias* de predição para os modelos de etanol, éster alquílico e glicerol, respectivamente: -0,22, 0,17 e -0,09). Assim, com o objetivo de avaliar a aplicação dessa técnica na correção espectral, os espectros pré-processados com EMSC para as misturas sintéticas (em cinza) e as amostras da coluna (em preto) que foram utilizadas no desenvolvimento dos modelos, após exclusão de *outliers*, estão apresentados na Figura 10.

Figura 10 – Espectros médios das misturas sintéticas (em cinza) e das amostras da coluna (em preto) pré-processados com EMSC, após a exclusão de outliers

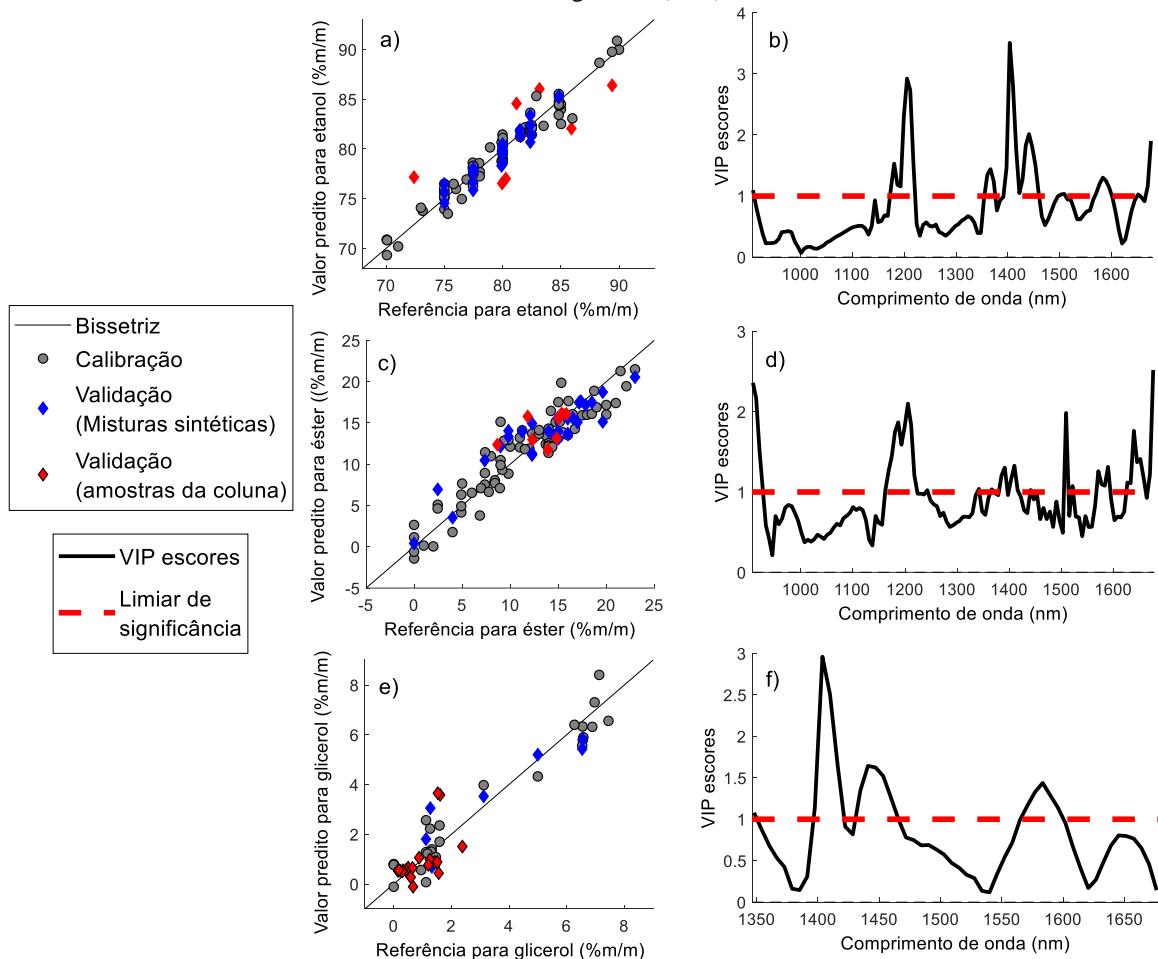


Fonte: A Autora (2020)

Como pode ser observado na Figura 10, de modo geral, a aplicação de EMSC permitiu a correção das regiões ruidosas e dos efeitos resultantes do espalhamento da radiação, apesar de que algumas diferenças entre os espectros das misturas sintéticas e os espectros das amostras da coluna ainda persistiram.

Em relação aos modelos PLS construídos com a técnica EMSC, os gráficos de valores de referência *versus* preditos para as amostras de calibração (em cinza) e as amostras do conjunto de predição (misturas sintéticas em azul e amostras da coluna em vermelho), estão apresentados na Figura 11 (a, c, e). Além disso, os gráficos de importância das variáveis (VIP) para esses modelos estão apresentados na Figura 11 (b, d, f). De acordo com Andersen e Bro (2010), esse tipo de gráfico examina a importância de cada variável para a construção do modelo. A linha tracejada em vermelho representa o limiar de significância. De modo geral, variáveis com valores de VIPs acima do limiar são consideradas importantes para a construção do modelo. Embora esses gráficos sejam bastante utilizados para a seleção de variáveis, os modelos apresentaram resultados melhores quando toda a faixa espectral do MicroNIR foi utilizada para a quantificação de éster e etanol.

Figura 11 – Gráficos dos valores preditos vs. referência (a, c e e) para o conjunto de calibração (em cinza) e o conjunto de predição, composto de misturas sintéticas (em azul) e amostras da coluna (em vermelho); e os VIPs escores (b, d e f) dos modelos PLS para quantificar etanol (a e b), éster alquílico (c e d) e glicerol (e e f)



Fonte: A Autora (2020)

Como observado na Figura 11 (a, c, e), de modo geral, pode-se perceber que as amostras de calibração e de predição estão, de certa forma, bem distribuídas ao longo da bissetriz. Apenas as amostras da coluna (em vermelho), ficaram agrupadas em uma faixa pequena de concentração de éster (Figura 11b) e de glicerol (Figura 11c). Conforme mencionado, embora alguns ajustes tenham sido intencionalmente impostos ao processo, com o intuito de produzir amostras com uma variabilidade maior na composição, essa variabilidade não foi suficiente para abranger toda a faixa dos modelos.

Na Figura 11b, é possível perceber que as variáveis mais importantes para a quantificação do teor de etanol estão localizadas principalmente nas regiões em torno de 1200 nm, atribuída ao 2<sup>a</sup> sobretom da ligação C-H, de 1400 nm, relacionada ao primeiro sobretom do grupamento O-H, e de 1670 nm, associada ao 1º sobretom de estiramento da ligação C-H. Conforme observado, essas regiões são, de fato, importantes para caracterizar o etanol. Ao

analisar a Figura 11d, vê-se que a região em torno de 1200 nm também foi importante na quantificação do éster alquílico. Adicionalmente, as regiões das combinações do grupo C-H, por volta de 1500 nm, e do primeiro sobretom de estiramento da ligação C-H, em torno de 1670 nm, também foram consideradas importantes na predição do teor de éster. Essa última região foi também identificada como importante em outros trabalhos descritos por de Lima et al. (2014) e Killner, Rohwedder e Pasquini (2011), que utilizaram um espectrofotômetro de bancada FT-NIR para monitoramento da produção de biodiesel em batelada. No que diz respeito à quantificação do glicerol, a região em torno de 1400 nm também foi identificada como mais importante.

Ainda com relação aos modelos PLS construídos com a técnica EMSC, os limites de detecção para esses modelos foram estimados empregando a interface MVC1\_GUI (CHIAPPINI; GOICOECHEA; OLIVIERI, 2020). Para a quantificação de etanol, a faixa para o limite de detecção é de 7,9 % (m/m) a 8,3% (m/m). Já para o glicerol, essa faixa é, aproximadamente, em torno de 1,0 % (m/m) e 1,8 % (m/m). Por fim, para o biodiesel, a faixa para o limite de detecção encontra-se entre 3,4% (m/m) e 5,9% (m/m). Conforme observado na Figura 11, as amostras da coluna apresentaram concentrações de etanol e de biodiesel bem acima dos respectivos limites de detecção. Já para o glicerol, algumas amostras da coluna apresentaram teores desse componente abaixo do limite.

### 3.2.5 Avaliação dos modelos quanto à predição das amostras da coluna

A aplicação apresentada no presente trabalho comprehende basicamente dois aspectos principais. O primeiro deles está relacionado à avaliação do desempenho do espectrômetro portátil MicroNIR para quantificar o teor de éster alquílico em misturas sintéticas contendo um elevado teor de etanol em diferentes concentrações (variando entre 70 e 90% (m/m)). Conforme mencionado, essa faixa de concentração foi empregada com o intuito de simular as condições operacionais da coluna no ponto monitorado com o MicroNIR e de modo a fornecer a variabilidade necessária para a etapa de modelagem. De acordo com o que foi discutido, os modelos de calibração apresentaram resultados considerados satisfatórios, conforme as métricas de desempenho apresentadas.

O segundo aspecto abordado nesse trabalho está relacionado à avaliação da capacidade dos modelos em quantificar a composição da mistura reacional na base da coluna. Esse ponto representou um dos grandes desafios desse trabalho, uma vez que os modelos foram

desenvolvidos com base em espectros adquiridos com um espectrômetro portátil, que apresenta uma faixa espectral reduzida, e utilizando, em sua maior parte, misturas sintéticas. As vantagens associadas ao uso das misturas sintéticas já foram discutidas anteriormente. No entanto, existem também alguns desafios relacionados a essa aplicação. O primeiro deles consiste na dificuldade de se obter misturas representativas das amostras da coluna, as quais são mais complexas do que as misturas sintéticas preparadas em laboratório de forma controlada. Adicionalmente, o monitoramento *on-line* das amostras da coluna foi outro fator que trouxe mais complexidade as medidas dessas amostras, quando comparadas às das misturas sintéticas adquiridas no modo *off-line*.

Com o intuito de minimizar as implicações associadas ao uso das misturas sintéticas na etapa de modelagem, algumas estratégias foram utilizadas durante o desenvolvimento desse trabalho, com o objetivo de diminuir as diferenças entre os espectros dessas misturas e das amostras da coluna. Conforme discutido, essas estratégias envolveram basicamente a inclusão de novas misturas aos modelos, passando a considerar a presença do glicerol e o efeito da temperatura. Como consequência dessa inclusão, os modelos PLS apresentaram resultados melhores, considerados satisfatórios. No entanto, é importante ressaltar que esses modelos poderiam apresentar resultados ainda melhores caso o glicerol estivesse presente em todas as misturas e se os espectros dessas misturas fossem adquiridos na mesma temperatura em que as amostras da coluna foram monitoradas.

Com o objetivo de avaliar os modelos desenvolvidos quanto à predição das amostras da coluna, bem como compará-la com a das misturas sintéticas, realizou-se a predição desses dois conjuntos de dados de forma separada. Os parâmetros resultantes dessa análise estão apresentados na Tabela 4.

Tabela 4 – Resultados para a predição das misturas sintéticas e amostras da coluna para os modelos PLS desenvolvidos com EMSC, para quantificação de etanol, éster alquílico e glicerol

Modelo PLS	Amostras sintéticas		Amostras da coluna	
	RMSEP (% (m/m))	Bias	RMSEP (% (m/m))	Bias
Etanol	0,85	-0,19	3,14	-0,32
Éster alquílico	2,13	0,02	2,19	0,75
Glicerol	1,03	-0,23	1,12	0,05

Fonte: A Autora (2020)

Conforme observado na Tabela 4, em todos os casos, o *bias* de predição não foi estatisticamente diferente de zero, de acordo com um teste t a um nível de confiança de 95%. Adicionalmente, os valores de RMSEP para as predições das misturas sintéticas e das amostras da coluna não apresentaram diferença estatisticamente significativa para os modelos de éster alquílico e glicerol, com base em um teste F com 95% de confiança. Para a quantificação de etanol, entretanto, o RMSEP para a predição das amostras da coluna (RMSEP = 3,14) foi estatisticamente superior ao das misturas sintéticas (RMSEP = 0,85), de acordo com um teste F (95% de confiança). Isso pode ser observado também através da análise da Figura 11<sup>a</sup> do item 3.4.4, em que é possível perceber que as amostras da coluna (em vermelho) estão mais afastadas da bissetriz quando comparadas às misturas sintéticas (em azul). Tal fato pode estar associado, dentre outros aspectos, ao problema de reprodutibilidade do método cromatográfico desenvolvido para quantificação do teor de etanol das amostras da coluna. Por exemplo, o desvio-padrão sob condições de repetibilidade para esse método foi estimado em torno de 1,4% (m/m) (com base em 12 amostras). Além disso, o fato de a região atribuída à ligação O-H ser mais afetada pelas diferenças entre os modos *on-line* e *off-line* de aquisição espectral, poderia ser uma possível explicação para essa ocorrência.

Por fim, apesar da dificuldade de reproduzir os espectros das amostras da coluna obtidos *on-line* utilizando as misturas sintéticas analisadas no modo *off-line*, os resultados dos modelos desenvolvidos foram satisfatórios no que diz respeito à predição das amostras da coluna. Considerando a complexidade do sistema analisado, especialmente devido ao excesso de etanol e a faixa de variação de sua concentração, esses resultados foram considerados satisfatórios até mesmo para o modelo do etanol, principalmente quando se considera a aplicação da metodologia proposta para o monitoramento do processo em tempo real.

### 3.2.6 Estimativa do rendimento de biodiesel das amostras da coluna

A partir dos valores preditos dos três modelos desenvolvidos com EMSC para a quantificação de etanol, éster alquílico e glicerol, foi possível estimar o rendimento de éster alquílico (conversão de óleo de algodão em biodiesel) para as amostras coletadas da coluna. Das 9 amostras utilizadas na etapa de predição dos três modelos, a amostra C08 foi excluída na etapa de modelagem do teor de éster, por apresentar um alto valor residual. Dessa forma, foi possível estimar o rendimento de 8 amostras da coluna e compará-los aos valores de referência obtidos diretamente por cromatografia, através do cálculo do erro relativo entre as

medidas. Os valores para cada uma das 8 amostras estão apresentados na Tabela 5.

Tabela 5 – Comparação entre os resultados dos rendimentos de éster alquílico das amostras da coluna estimados a parti dos modelos PLS e adquiridos por cromatografia

Amostra	Rendimento de éster alquílico estimado pelos modelos PLS (%(m/m))	Rendimento de éster alquílico obtido por cromatografia (%(m/m))	Erro relativo (%)
C01	81,0	86,0	5,8
C02	80,5	82,6	2,5
C03	88,3	91,0	3,0
C04	89,9	90,4	0,6
C05	90,5	85,9	5,4
C06	91,5	88,9	2,9
C07	98,6	89,2	10,5
C09	91,9	91,7	0,2

Fonte: A Autora (2020)

Como pode ser visto na Tabela 5, a maioria dos erros relativos ficou abaixo de 5%. Apenas a amostra 7 apresentou um erro relativo considerado bastante alto, em torno de 10%. Apesar disso, um teste t, com 95% de confiança, demonstrou que não há diferença estatisticamente significativa entre os resultados fornecidos pelos dois métodos. De modo geral, os resultados foram considerados satisfatórios do ponto de vista de monitoramento do processo, principalmente quando se leva em consideração o excesso de etanol nas misturas analisadas.

### 3.3 CONCLUSÃO

O presente capítulo avaliou o desempenho de um espectrômetro portátil na região do NIR (MicroNIR) para o monitoramento *on-line* de um processo contínuo de destilação reativa empregado na produção de biodiesel. Apesar do elevado teor de etanol no ponto monitorado com o MicroNIR, os modelos construídos aplicando-se o pré-processamento EMSC apresentaram resultados satisfatórios para a quantificação de etanol (RMSEP = 1,66%(m/m)), éster alquílico (RMSEP = 2,15%(m/m)) e glicerol (RMSEP = 1,08%(m/m)). Os resultados foram considerados satisfatórios, inclusive, quando os modelos foram utilizados na predição de amostras da coluna, apenas.

Os valores preditos a partir dos modelos de calibração foram então utilizados para estimar o rendimento de éster alquílico das amostras da coluna. Os resultados foram comparáveis às medidas de referência obtidas por cromatografia gasosa, com erros relativos abaixo de 5% para a maioria das amostras. É importante ressaltar que, para fins de monitoramento do processo em tempo real, a informação do teor de éster alquílico na mistura reacional seria suficiente. No entanto, o rendimento do biodiesel foi calculado como uma forma de validar a metodologia proposta.

Dessa forma, o presente trabalho demonstrou a viabilidade de utilização do espectrômetro portátil na região do NIR para monitorar a produção de biodiesel em um processo contínuo, mesmo trabalhando-se com uma concentração de etanol elevada, a qual representou uma condição desafiadora para a etapa de modelagem. Além disso, o trabalho demonstrou a possibilidade de se utilizar misturas sintéticas na construção dos modelos de calibração. Conforme discutido, o uso dessas misturas, além de necessário para ampliar a variabilidade dos dados, trouxe benefícios do ponto de vista experimental. Adicionalmente, reduz a dependência em relação aos métodos cromatográficos, utilizados como referência.

#### **4 MONITORAMENTO ESTATÍSTICO DO PROCESSO INDUSTRIAL DE METALIZAÇÃO A VÁCUO DE FILMES DE POLIÉSTER**

A deposição a vácuo é o processo tradicionalmente empregado para a aplicação de revestimentos metálicos, geralmente alumínio, sobre substratos poliméricos, como o politereftalato de etileno e o polipropileno. Os materiais resultantes desse processo são utilizados principalmente na produção de embalagens (BISHOP, 2015). Em geral, os revestimentos são aplicados para atribuir novas propriedades aos filmes poliméricos, dentre as quais, destacam-se as aplicações de barreira, decorativas, de segurança e funcionais. Os revestimentos de barreira, em particular, representam uma categoria de materiais que conferem à embalagem um bloqueio à passagem de luz, vapor de água, oxigênio e outro gases. Eles são aplicados, principalmente, para prolongar o prazo de validade dos produtos, especialmente no ramo alimentício (STRULLER; KELLY; COPELAND, 2014; BISHOP; MOUNT, 2016).

Considerando os processos industriais de deposição a vácuo aplicados na produção de revestimentos de barreira, a deposição de alumínio utilizando barcos de evaporação aquecidos por resistência (também conhecidos como evaporadores) é o método mais empregado (BISHOP, 2015). De um modo geral, os evaporadores são materiais compactos formados por misturas intermetálicas bi ou trifásicas. Usualmente, as misturas bifásicas são compostas por diboreto de titânio, que apresenta baixa resistividade elétrica e alta molhabilidade para o alumínio, e nitreto de boro, que possui alta resistividade elétrica e resistência ao choque térmico. Já às misturas trifásicas, é adicionado também o nitreto de alumínio, que confere condutividade térmica e resistência elétrica ao evaporador (BISHOP, 2011; HERRMANN et al., 2011).

A deposição a vácuo é geralmente conduzida em um sistema industrial conhecido como metalizadora. Em cada corrida desse processo, um rolo de filme polimérico é desenrolado para ser revestido com uma fina camada de alumínio, e é então enrolado na forma de um novo rolo de filme metalizado. Nesse sistema, vários barcos de evaporação são comumente utilizados com o intuito de revestir o filme em toda sua extensão. Eles ficam situados abaixo do filme e ao longo de sua largura. Durante o processo, fios de alumínio são constantemente alimentados à superfície aquecida de cada barco. Quando o alumínio se funde, uma nuvem do metal evaporado é formada e uma fina camada desse metal é depositada sobre uma das superfícies do filme (BISHOP, 2015).

Diversos fatores podem influenciar o desempenho de barreira de filmes revestidos a vácuo. Um desses fatores consiste na qualidade dos materiais utilizados no processo, tais quais do fio de alumínio, dos barcos de evaporação e dos substratos (BISHOP, 2015; BISHOP; MOUNT, 2016). Outro aspecto crítico está relacionado às condições do processo de deposição. Conforme discutido por Mount (2008), as variáveis de processo apresentam um importante e complexo efeito sobre as propriedades de barreira dos filmes metalizados. Nesse contexto, dentre as principais variáveis de processo que influenciam na propriedade de barreira do filme, pode-se destacar a sua densidade óptica, a velocidade com que o filme é desenrolado/enrolado, a taxa de evaporação, a pressão da câmara de evaporação e a temperatura do rolo frio por onde passa o filme que está sendo revestido.

Como a camada de revestimento de um filme metalizado a vácuo é geralmente muito fina, o que dificulta a medição direta da sua espessura, é comum se utilizar outra propriedade para o controle da espessura do revestimento. A densidade óptica (DO) do filme metalizado, por exemplo, pode ser monitorada para esse fim. Trata-se de uma medida da opacidade do filme, a qual está relacionada à espessura da camada metálica depositada sobre o substrato polimérico e, consequentemente, ao seu desempenho de barreira. Geralmente, durante um processo de deposição conduzido em metalizadoras comerciais, medidas de DO são constantemente adquiridas do filme metalizado e utilizadas no controle da alimentação do fio de alumínio e da temperatura do barco de evaporação, com o intuito de controlar o tamanho da poça de fusão no evaporador e a taxa de evaporação do alumínio. Como a taxa de evaporação é facilmente afetada por variações da temperatura, esse controle por meio da DO é essencial para a obtenção de um revestimento metálico uniforme (BISHOP, 2015; BISHOP; MOUNT, 2016). Dessa forma, o monitoramento e controle da DO representam uma maneira eficaz de garantir a consistência da deposição metálica, o que é particularmente importante para evitar defeitos na superfície do revestimento e, portanto, garantir a qualidade do desempenho de barreira do filme.

Nesse contexto, o monitoramento estatístico do processo pelo uso de cartas de controle pode ser considerado uma ferramenta adequada para o processo de metalização a vácuo, com base nas medidas de DO. Por permitir a detecção de situações anômalas, os gráficos de controle são ferramentas da qualidade eficientes para reduzir a variabilidade do processo e melhorar o seu desempenho (MONTGOMERY, 2009). Dessa forma, o presente trabalho descreve uma estratégia de monitoramento estatístico para um processo industrial de metalização a vácuo. Mais especificamente, a abordagem de monitoramento foi baseada nas

medidas de DO de filmes de poliéster revestidos com alumínio, adquiridas continuamente durante a metalização. A literatura relacionada à aplicação de cartas de monitoramento estatístico a esse tipo de processo é bastante escassa. Existe, por exemplo, um trabalho com abordagem diferente da apresentada nesta tese, o qual descreve a aplicação de um controlador preditivo generalizado, associada ao uso de uma carta de monitoramento estatístico CUSUM, para controlar a espessura do filme em um processo de deposição de uma fina camada de revestimento, utilizando a temperatura como variável de entrada do modelo (JIN; GUO; ZHOU, 2006).

O principal objetivo do trabalho descrito neste capítulo consistiu em propor uma estratégia de monitoramento estatístico ao processo industrial de metalização a vácuo de filmes de poliéster, com base nas medidas de DO coletadas ao longo do processo. Para isso, foram estabelecidos os objetivos específicos detalhados a seguir:

- a) Investigar, em um estudo preliminar, as correlações existentes nos dados de DO adquiridos durante a produção sequencial de 8 filmes de poliéster metalizados;
- b) Sugerir uma estratégia de monitoramento estatístico para o processo de metalização com base no estudo preliminar realizado e nas características do processo;
- c) Avaliar a estratégia de monitoramento estatístico proposta utilizando dados de DO coletados durante a produção sequencial de 12 novos rolos de filme metalizado, não utilizados anteriormente;
- d) Complementar a estratégia de monitoramento sugerida com um estudo da capacidade do processo.

#### 4.1 METODOLOGIA

A metodologia empregada no desenvolvimento deste trabalho está descrita a seguir.

##### 4.1.1 Processo de metalização a vácuo e aquisição dos dados de densidade óptica

Os dados de processo utilizados no desenvolvimento do presente trabalho foram disponibilizados por uma empresa com planta industrial no estado de Pernambuco, que atua na fabricação de filmes especiais de poliéster. Esses dados consistem em medidas de densidade óptica adquiridas durante a produção industrial de rolos de filmes metalizados, através do processo de deposição metálica por evaporação a vácuo, conduzido em uma

metalizadora TOPMET™ 2450 (Applied Materials). Esse equipamento utiliza barcos de evaporação aquecidos por resistência com composição padrão, dispostos de forma alternada a cada 0,1 m ao longo da largura do filme, semelhante ao ilustrado na Figura 12.

Figura 12 - Evaporadores dispostos de forma alternada dentro da metalizadora



Fonte: Bishop (2015)

Durante a deposição, valores de DO do filme revestido foram obtidas por sensores alinhados a cada evaporador na direção do comprimento do filme. Nessa metalizadora, até 26 barcos de evaporação podem ser instalados. Entretanto, o número de evaporadores utilizados em cada corrida, bem como o número de sensores ativos, depende da largura do rolo a ser revestido.

De um modo geral, as produções seguiram a operação básica desse tipo de processo, de acordo com o procedimento descrito por Bishop (2015). Durante cada corrida, carretéis de fios de alumínio de alta pureza (mínimo de 99%) são continuamente desenrolados utilizando bombas com motores a passos, as quais permitem o controle da alimentação do fio de alumínio em cada evaporador. Os fios passam, então, por tubos rígidos e são conduzidos até os barcos de evaporação para serem alimentados no local desejado, geralmente a região central dos barcos. Uma vez que o fio entra em contato com a superfície aquecida do evaporador, uma poça de fusão é formada, a partir da qual o alumínio evapora, criando uma nuvem de vapor dentro da metalizadora. À medida que o filme polimérico passa por essa nuvem, o revestimento metálico é depositado sobre uma de suas superfícies.

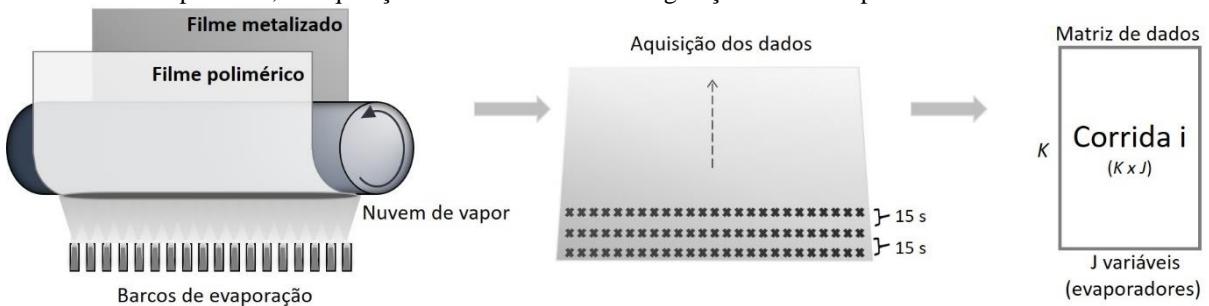
Inicialmente, dados de DO adquiridos durante a produção sequencial de oito rolos de filmes de poliéster de 12  $\mu\text{m}$  de espessura foram avaliados. Essas corridas foram identificadas como A1 a A8, numeradas de acordo com a ordem de produção. Os dados, para essas oito corridas, foram obtidos por meio de 18 sensores de DO alinhados a 18 evaporadores, a cada

15 s aproximadamente. Antes do início dessa produção em sequência, um conjunto de evaporadores novos foi instalado na metalizadora.

Ao fim da metalização de cada rolo de filme polimérico, um relatório contendo os perfis de DO ao longo dos sentidos transversal e longitudinal do filme revestido foi extraído da metalizadora. Cada relatório foi organizado no formato de uma matriz de dados e, para o desenvolvimento do presente trabalho, cada barco de evaporação foi considerado uma variável. Por conseguinte, para cada rolo de filme produzido, foi gerada uma matriz de dados contendo  $J = 18$  colunas e  $K$  linhas, em que  $K$  representa a quantidade de medidas de DO adquiridas ao longo do tempo, podendo ter um valor diferente para cada corrida.

Uma representação esquemática simplificada do processo de metalização de um rolo de filme de poliéster, da aquisição dos dados de DO e da geração da matriz para cada corrida, está apresentada na Figura 13.

Figura 13 - Representação esquemática simplificada do processo de metalização de um rolo de filme de poliéster, da aquisição dos dados de DO e da geração da matriz para cada corrida



Fonte: A Autora (2020)

A matriz de dados da Figura 13, com os valores de DO coletados ao longo de cada corrida do processo, foram coletados como arquivos em Excel. Esses dados foram importados e analisados no MATLAB (MathWorks, MA, USA), utilizando-se rotinas desenvolvidas durante este trabalho.

#### 4.1.2 Análise exploratória dos dados de densidade óptica

Com o objetivo de compreender melhor os dados e, assim, propor uma estratégia de monitoramento mais adequada para o processo, uma análise exploratória foi realizada. Inicialmente, avaliou-se a velocidade dos filmes e as matrizes de dados coletadas para as corridas A1 a A8. Os perfis de DO foram investigados com o objetivo de detectar tendências e correlações nos dados. Dessa forma, os mapas de correlação de cada corrida foram avaliados

com o intuito de identificar padrões de correlação entre as 18 variáveis (barcos de evaporação), assim como para comparar o comportamento de diferentes corridas. Esses mapas foram construídos com base no cálculo dos coeficientes de correlação ( $r$ ) entre todos os possíveis pares de variáveis. Trata-se de uma medida da dependência linear entre duas variáveis e pode ser calculado de acordo com a Equação 30.

$$r = \frac{1}{N-1} \sum_{t=1}^N \left( \frac{x_t - \bar{x}}{s_x} \right) \left( \frac{y_t - \bar{y}}{s_y} \right) \quad (30)$$

Em que:  $\bar{x}$  e  $s_x$  são, respectivamente, a média e o desvio-padrão da variável  $x$ ;  $\bar{y}$  e  $s_y$  são, respectivamente, a média e o desvio-padrão da variável  $y$ ;  $N$  é o número de observações de cada variável; e o subíndice  $t$  indica a posição da medida no tempo.

Adicionalmente, no estudo preliminar, também foram avaliadas a autocorrelação e a correlação cruzada para todas as variáveis. A função de autocorrelação (ACF), calculada de acordo com a Equação 31, é uma medida da correlação entre duas observações (medidas em tempos diferentes) para uma mesma variável (série temporal). Já a função de correlação cruzada (CCF), calculada a partir das Equações 32 e 33, é a medida da correlação entre duas observações em tempos diferentes, para duas variáveis diferentes (MONTGOMERY, 2009; MILLS, 2011).

$$ACF = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2}, \text{ para } k=0, 1, 2, \dots \quad (31)$$

$$CCF = \frac{1}{\tau} \sum_{t=1}^{\tau-k} \frac{(x_t - \bar{x})(y_{t+k} - \bar{y})}{s_x s_y}, \text{ para } k = 0, 1, 2, \dots \quad (32)$$

$$CCF = \frac{1}{\tau} \sum_{t=1}^{\tau-k} \frac{(x_{t-k} - \bar{x})(y_t - \bar{y})}{s_x s_y}, \text{ para } k = 0, -1, -2, \dots \quad (33)$$

Em que  $k$  é o número de períodos de tempo que separam as observações e  $\tau$  o número de pares de observações.

#### 4.1.3 Estratégia de monitoramento estatístico

Com o intuito de estimar a linha de tendência da trajetória de cada variável, a estatística EWMA foi empregada com o objetivo de suavizar o perfil de DO, permitindo assim, uma melhor visualização da tendência dos dados. Para essa abordagem específica, os limites de controle inferior e superior da carta EWMA não foram considerados. Na construção dessa carta, um parâmetro de suavização ( $\lambda$ ) de 0,2 foi empregado e a média das cinco primeiras observações foi considerada como valor inicial ( $z_0$ ), requerido no primeiro cálculo da estatística EWMA.

Outra estratégia empregada para o monitoramento estatístico do processo envolveu a construção de cartas de controle EWMA e da amplitude (R), que foram usadas, respectivamente, para o monitoramento da média e da variabilidade do processo. Essas cartas foram construídas após a remoção da tendência dos dados. Para isso, as técnicas de diferenciação de primeira, segunda e terceira ordens foram avaliadas. Após a diferenciação aplicando-se a ordem mais apropriada, as cartas de controle EWMA foram, então, desenvolvidas para cada variável. Para a construção dessa carta, um parâmetro de suavização ( $\lambda$ ) de 0,5 foi empregado e a média das cinco primeiras observações da variável foi considerada como valor inicial ( $z_0$ ). Além disso, o fator largura dos limites de controle ( $L$ ) foi estabelecido em 2,7. Na definição dos limites de controle e da linha central dessa carta, o desvio-padrão e a média do processo foram calculados com base nos dados diferenciados das 8 corridas analisadas (A1 a A8), para aquela variável específica. O objetivo dessa estratégia foi adotar a média representativa dos dados históricos com o intuito de usar a carta de controle EWMA para o monitoramento do processo em tempo real.

Por fim, a carta de controle R foi também desenvolvida com base nos dados diferenciados. Na construção dessa carta, a amplitude móvel foi calculada para cada 5 observações. A linha central do gráfico foi estimada pelo valor médio da amplitude móvel ( $\bar{R}$ ) calculado considerando os dados diferenciados das 8 corridas analisadas (A1 a A8), para uma posição específica de evaporador. Os limites inferior e superior de controle dessa carta foram calculados com base nas Equações 23 e 24, respectivamente, considerando  $D_3 = 0$  e  $D_4 = 2,114$ . Conforme mencionado, essas constantes são tabeladas e dependem do número de pontos utilizados no cálculo da média móvel (MONTGOMERY, 2009).

#### 4.1.4 Avaliação da estratégia de monitoramento estatístico proposta

Após a proposição de uma estratégia de monitoramento estatístico para o processo estudado, um novo conjunto de dados de DO foi analisado seguindo-se o mesmo procedimento descrito para as corridas A1 a A8. Essa análise foi realizada com o intuito de avaliar a estratégia de monitoramento proposta e examinar se a mesma poderia ser utilizada no monitoramento de outras corridas do processo. Adicionalmente, o comportamento das novas corridas pôde ser comparado ao observado para as oito corridas analisadas inicialmente (A1 a A8).

O novo conjunto de dados de DO foi coletado durante a produção sequencial de 12 rolos de filme metalizado e as corridas de processo foram identificadas por B1 a B12, de acordo com a ordem de produção. Essas corridas foram produzidas na mesma metalizadora utilizada na produção das corridas A1 a A8, sob condições semelhantes, mas variando-se o filme de poliéster utilizado como substrato e o número de sensores de DO ativos. Para as primeiras 5 corridas (identificadas como B1 a B5), filmes de poliéster de 10  $\mu\text{m}$  de espessura foram revestidos e as medidas de DO provenientes de 20 sensores foram consideradas. Já para as corridas B6 a B12, filmes de poliéster de 12  $\mu\text{m}$  de espessura foram metalizados e medidas de DO de 22 sensores foram levadas em consideração. Apesar de terem sido utilizados três tipos de substratos diferentes nesse trabalho, o valor especificado para a densidade óptica do filme metalizado foi igual a 2,15 para todos os casos, valor comumente adotado pelas indústrias. Por fim, vale ressaltar que, antes das corridas B1 e B8, um conjunto de novos evaporadores foi instalado na metalizadora.

#### 4.1.5 Análise da capacidade do processo

Outra estratégia que pode ser utilizada para o monitoramento de um processo em controle estatístico consiste na análise de sua capacidade. No presente trabalho, essa análise foi realizada pelo cálculo da razão de capacidade de um processo descentrado ( $C_{pk}$ ). No cálculo da  $C_{pk}$  para cada variável, a média do processo ( $\bar{\mu}$ ) foi estimada com base na média dos valores de DO para essa variável específica. Já o desvio-padrão ( $\bar{\sigma}$ ) foi calculado pela expressão  $\bar{\sigma} = \bar{R}/d_2$ , em que  $\bar{R}$  é a média da amplitude móvel (calculada para cada 5 pontos) dos valores de DO para aquela variável, e  $d_2 = 2,326$  é uma constante tabelada, dependente do número de observações utilizados no cálculo da amplitude móvel (MONTGOMERY,

2009). Enquanto as cartas de controle podem ser aplicadas para o monitoramento do processo em tempo real, a análise de capacidade sugerida deve ser realizada após o término da corrida.

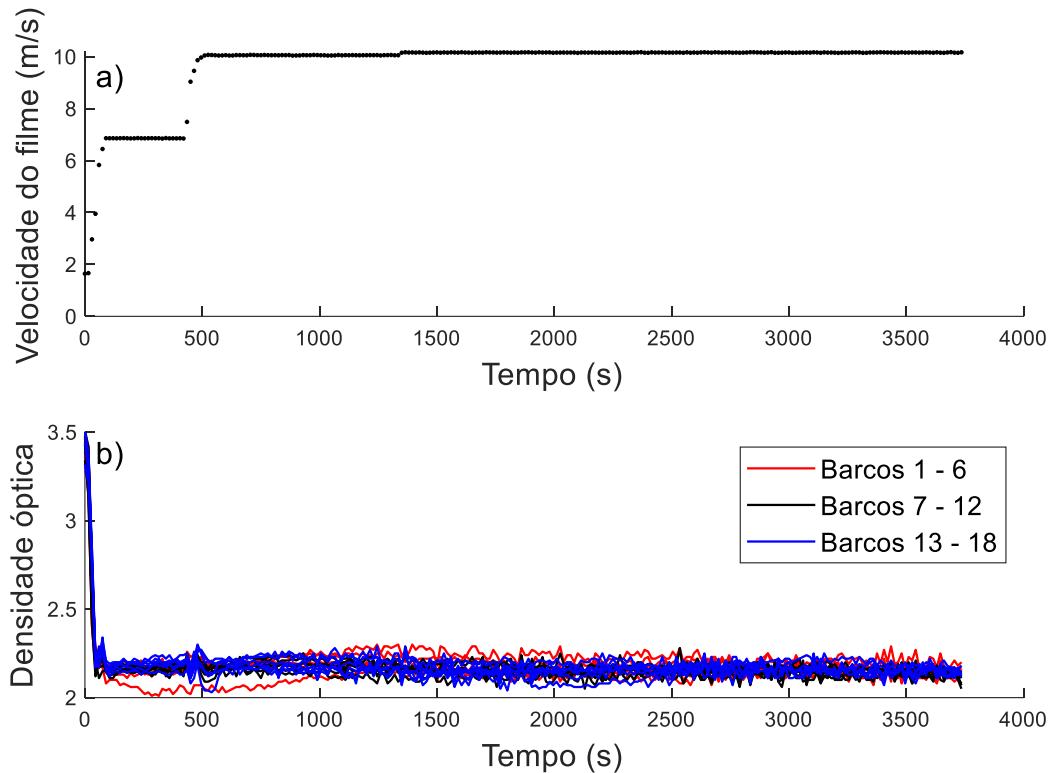
## 4.2 RESULTADOS E DISCUSSÃO

Os resultados e discussão deste trabalho estão apresentados a seguir.

### 4.2.1 Análise exploratória dos dados de densidade óptica

A velocidade com que o filme de poliéster é desenrolado para ser revestido com uma fina camada metálica e posteriormente enrolado na forma de um novo rolo de filme metalizado é uma das variáveis de processo controladas durante a metalização. Associada com outras variáveis, como a taxa de alimentação do fio de alumínio e a temperatura do barco de evaporação, a velocidade do filme tem um efeito importante sobre os perfis de DO do filme metalizado. No período inicial do processo de deposição, a alimentação do fio ao evaporador deve ser realizada de forma gradual, com o objetivo de formar uma poça de fusão adequada, já que o tamanho e o formato da poça exercem influência sobre o perfil da deposição, de acordo com Bishop (2011). Nesse período, a velocidade do filme também é ajustada até alcançar as condições de operação, mantendo-se aproximadamente constante até o término da deposição. No caso de ocorrer uma perturbação no processo, essas variáveis são reguladas com o intuito de manter constante, à medida do possível, a relação entre a velocidade do filme e a taxa de evaporação. Por essa razão, a velocidade do filme e a taxa de alimentação do fio de alumínio apresentam perfis semelhantes. Com o objetivo de ilustrar o comportamento da velocidade do filme estabelecida para o processo e dos perfis de DO, estão apresentados na Figura 14 o perfil da velocidade do filme (em m/s) (a) e os perfis de DO para os 18 evaporadores (b) para a corrida A2.

Figura 14 – Perfil da velocidade do filme (a) e perfis de DO adquiridos ao longo dos 18 evaporadores (b) para a corrida A2

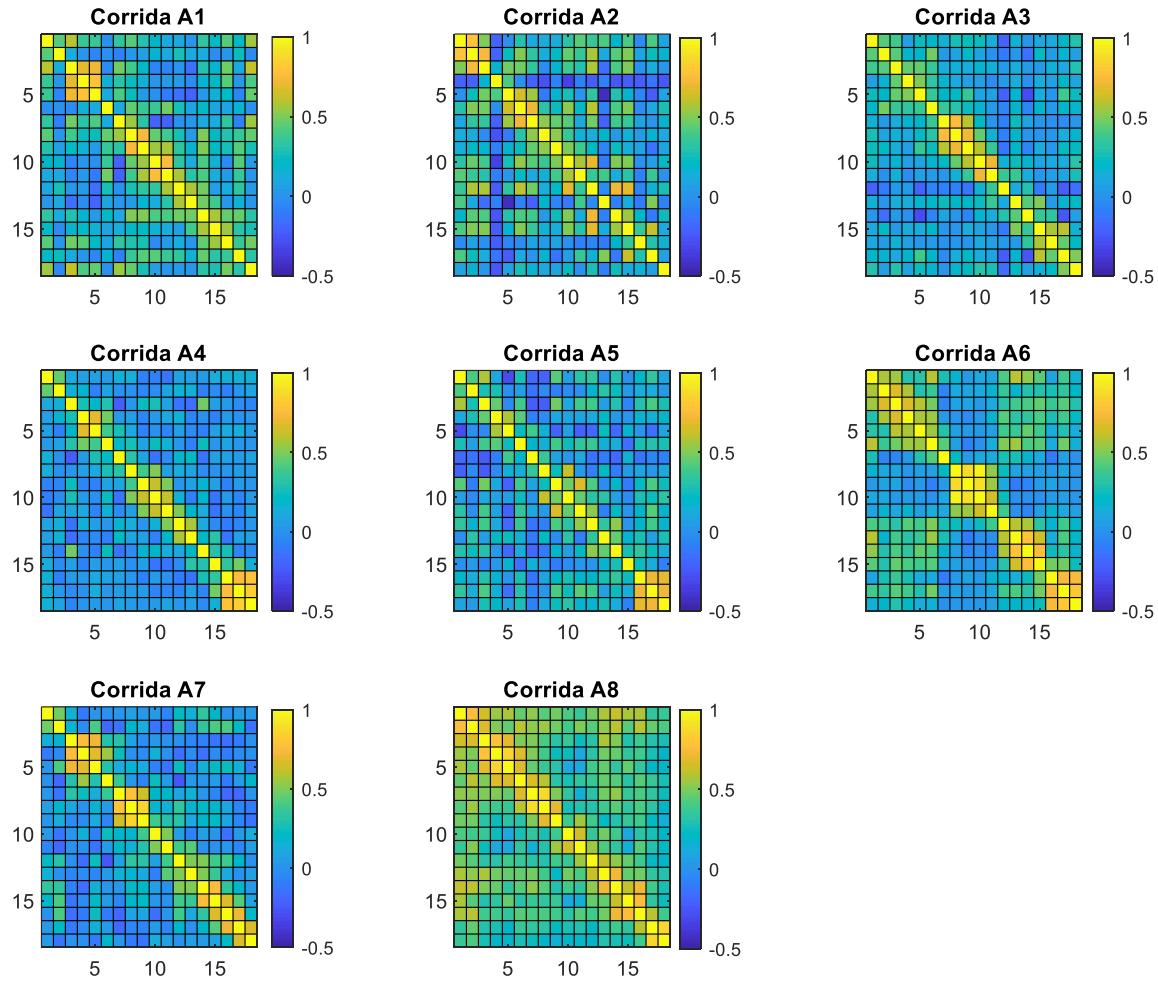


Fonte: A Autora (2020)

Devido à influência da velocidade do filme nos perfis de DO, assim como aos ajustes iniciais nas condições de operação do processo (por exemplo, ajuste para alcançar a velocidade de 10 m/s), todas as análises apresentadas nesse trabalho foram realizadas a partir do momento em que a velocidade do filme foi mantida aproximadamente constante. Um critério de convergência envolvendo o cálculo consecutivo do desvio-padrão da velocidade do filme foi utilizado para definir esse momento para cada corrida. Para a corrida A2, por exemplo, a velocidade do filme foi considerada constante a partir de 600 s, aproximadamente (Figura 14).

Com o objetivo de auxiliar na seleção de uma estratégia de monitoramento adequada para o processo, foi realizada uma análise exploratória nos dados das corridas A1 a A8. De modo geral, esse estudo preliminar envolveu a investigação dos perfis de DO para a detecção de tendências e correlações nos dados. Inicialmente, foram avaliados os mapas de correlação para cada corrida, apresentados na Figura 15. Nesses mapas, a escala de cores está relacionada com os valores do coeficiente de correlação entre todos os possíveis pares de variáveis (evaporadores).

Figura 15 – Mapas de correlação para as corridas A1 a A8. As escalas de cores estão relacionadas com os coeficientes de correlação entre as variáveis



Fonte: A Autora (2020)

Os mapas de correlação apresentados na Figura 15 demonstraram que os dados coletados ao longo dos 18 evaporadores são pouco correlacionados. A fraca correlação entre os dados está provavelmente relacionada com a dificuldade de se ter evaporadores operando com temperaturas iguais. Ainda que o mesmo valor de corrente seja aplicado a todos os barcos de evaporação, a temperatura não necessariamente é a mesma para todos eles, uma vez que as resistividades dos barcos podem apresentar pequenas diferenças. Além disso, a taxa de alimentação do fio de alumínio também pode variar um pouco para cada barco. Como resultado, o tamanho da poça de fusão e, consequentemente, da nuvem de vapor, pode ser diferentemente formada para cada evaporador, ocasionando uma diferença no perfil de deposição. Dessa forma, a soma desses fatores podem influenciar na uniformidade da espessura da camada metálica depositada sobre o filme ao longo de sua largura e, portanto,

justificam a importância de se utilizar as medidas de DO no controle da alimentação do fio de alumínio e da temperatura do barco de evaporação, conforme discutido por Bishop (2015).

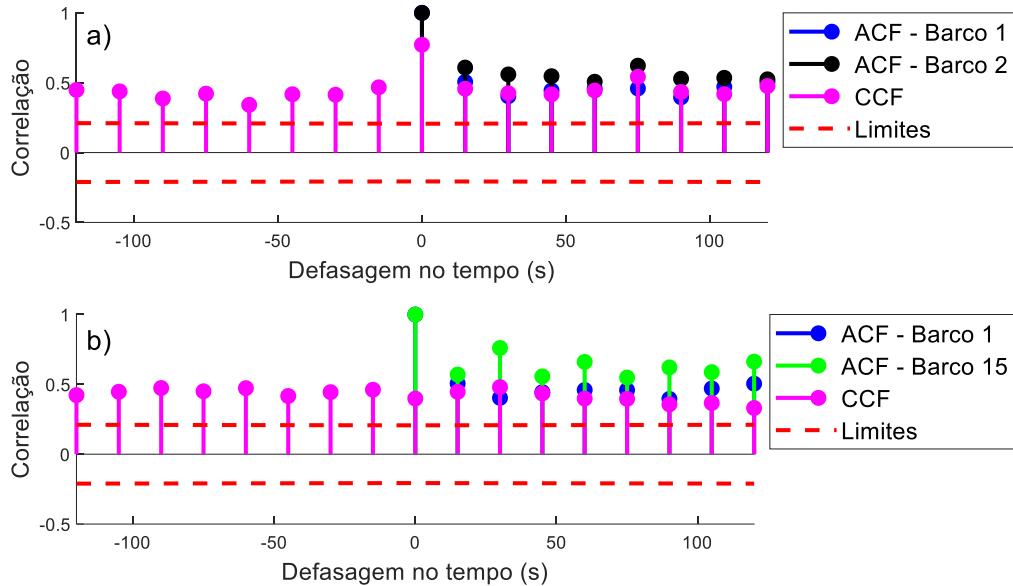
Por outro lado, é possível perceber a partir dos mapas de correlação (Figura 15) que, em alguns casos, os dados de DO adquiridos ao longo de evaporadores vizinhos apresentam uma correlação um pouco maior. Esse fato está provavelmente relacionado com o funcionamento do sistema de controle da metalizadora. Basicamente, nesse sistema, quando uma ação corretiva é necessária, a mesma é realizada não apenas no evaporador que a requereu, mas também nos dois evaporadores vizinhos a esse.

Por fim, os mapas de correlação apresentados na Figura 15 indicaram a ausência de um padrão de correlação que se repita para as 8 corridas analisadas. Isso pode estar relacionado às diferenças no processo entre uma corrida e outra. Dentre os quais, pode-se citar: as variações nos materiais utilizados (barcos de evaporação, fio de alumínio e substratos); a limpeza do equipamento; a utilização de evaporadores mais novos ou mais antigos (desgastados); e as mudanças no sistema de alimentação do alumínio (por exemplo, utilização de carretéis de fios de alumínio mais cheios ou mais vazios e posição da alimentação do alumínio no evaporador).

Além da análise dos mapas de correlação, realizou-se também a avaliação das funções de autocorrelação (ACF) e de correlação cruzada (CCF) para todas as variáveis das corridas A1 a A8. De um modo geral, o autocorrelograma (gráfico para apresentação da ACF) de cada variável permitiu a detecção de uma alta correlação entre observações adquiridas em tempos diferentes, o que pode ser um indicativo de tendência nos dados. Por outro lado, a CCF para cada par de variáveis indicou uma fraca correlação cruzada para as medidas adquiridas ao longo de evaporadores vizinhos. Medidas coletadas ao longo de evaporadores mais distantes não apresentaram, de modo geral, correlação cruzada.

A quantidade de gráficos geradas nesse estudo impossibilitam a apresentação de todos eles, dessa forma, apenas alguns resultados são mostrados a fim de se exemplificar o que foi discutido. Estão apresentadas na Figura 16, por exemplo, as funções ACF e CCF para as variáveis 1 e 2 (Figura 16a) e 1 e 15 (Figura 16b), da corrida A2.

Figura 16 - ACF das variáveis 1 (em azul) e 2 (em preto) e CCF dessas variáveis (em magenta), para a corrida A2 (a); e ACF das variáveis 1 (em azul) e 15 (em verde) e correlação cruzada dessas variáveis (em magenta), para a mesma corrida (b). Em vermelho, limites com 99% de confiança para ACF e CCF

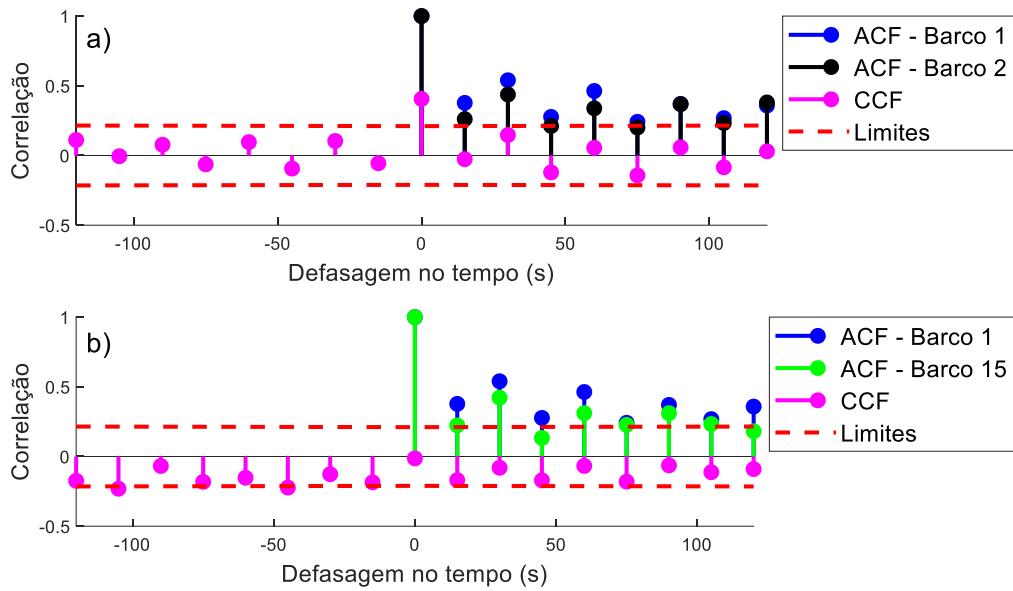


Fonte: A Autora (2020)

É possível perceber através da Figura 16, que todas as três variáveis analisadas (1, 2 e 15) apresentaram autocorrelação (valores acima dos limites com 99% de confiança). A análise da CCF indicou correlação cruzada entre as variáveis 1 e 2 (Figura 16a) e entre as variáveis 1 e 15 (Figura 16b), com valores de correlação acima dos limites. Nesse exemplo, as medidas adquiridas ao longo de evaporadores mais distantes também apresentaram uma alta correlação cruzada.

Com o intuito de dar mais um exemplo sobre o estudo das correlações, estão apresentadas na Figura 17, as funções ACF e CCF para as variáveis 1 e 2 (Figura 17a) e 1 e 15 (Figura 17b), da corrida A5.

Figura 17 - ACF das variáveis 1 (em azul) e 2 (em preto) e CCF dessas variáveis (em magenta), para a corrida A5 (a); e ACF das variáveis 1 (em azul) e 15 (em verde) e correlação cruzada dessas variáveis (em magenta), para a mesma corrida (b). Em vermelho, limites com 99% de confiança para ACF e CCF



Fonte: A Autora (2020)

Conforme apresentado na Figura 17, nesse caso, as variáveis 2 e 15 apresentaram valores mais baixos de autocorrelação. Além disso, tanto os evaporadores vizinhos (1 e 2) quanto os mais distantes (1 e 15) da corrida A5 não apresentaram correlação cruzada.

#### 4.2.2 Estratégia de monitoramento estatístico

As características do processo de metalização a vácuo abordadas anteriormente, bem como a fraca correlação dos dados de DO adquiridos ao longo de diferentes evaporadores, motivaram a aplicação de cartas de controle univariadas para o monitoramento do processo. Novamente, a quantidade de gráficos gerados nesse estudo impossibilita a apresentação de todos eles. Dessa forma, apenas alguns resultados selecionados serão mostrados com o objetivo de ilustrar as principais conclusões.

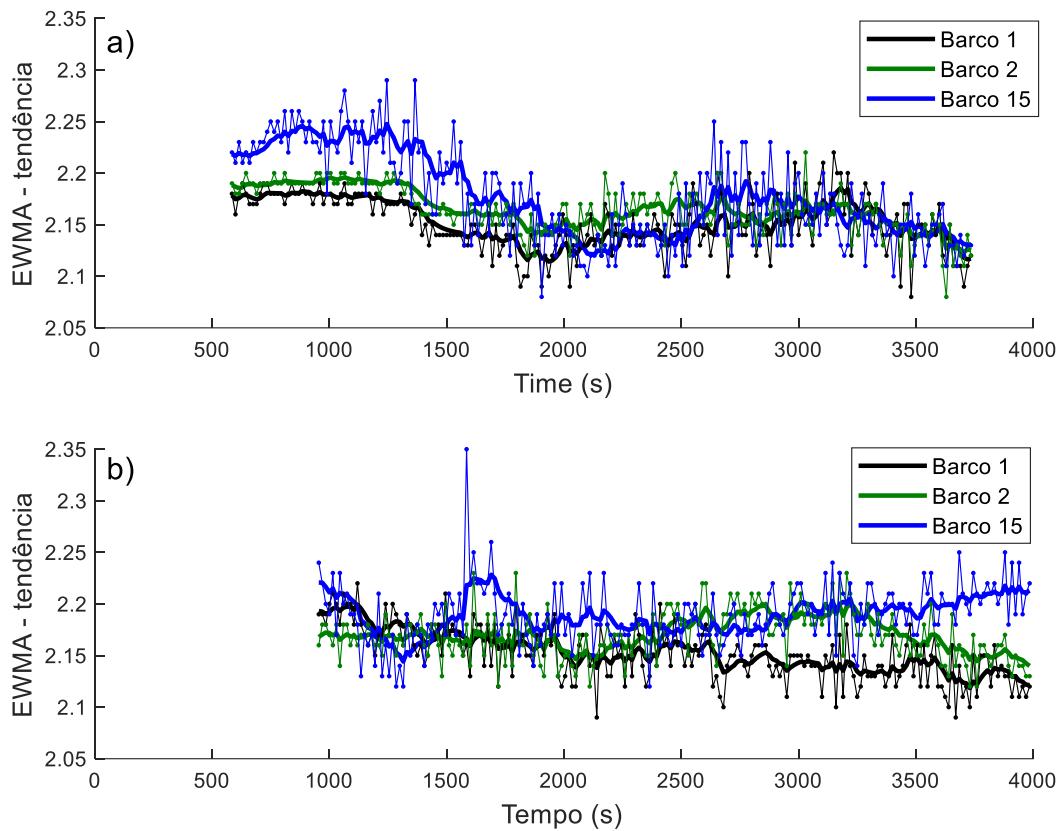
A princípio, levando-se em consideração a importância de monitorar a tendência dos dados, a estatística EWMA foi utilizada como uma ferramenta para suavização dos dados de modo a facilitar a detecção da linha de tendência. O objetivo é que essa detecção, feita de forma mais rápida e clara, possa auxiliar os operadores nas tomadas de decisões e ações corretivas requeridas ao longo do processo. A comparação entre as cartas EWMA para todas as variáveis de cada corrida confirmou que as medidas de DO adquiridas ao longo de evaporadores diferentes, de fato, apresentam um comportamento independente. Assim,

embora evaporadores vizinhos possam apresentar tendências similares, as medidas de DO adquiridas ao longo de evaporadores distintos, em geral, apresentam trajetórias diferentes. Para ilustrar essa observação, as cartas EWMA para as 18 variáveis da corrida A2 estão apresentadas no APÊNDICE E.

Adicionalmente, uma análise comparativa entre as corridas A1 a A8, permitiu verificar que um mesmo evaporador apresentou comportamento distinto para diferentes corridas. Isso pode ser explicado pelos possíveis ajustes no processo realizados entre as corridas, conforme discutido anteriormente.

Com o objetivo de ilustrar essa discussão, as cartas de controle EWMA para o monitoramento da linha de tendência de três variáveis diferentes (barcos 1, 2 e 15) das corridas A2 e A5 estão apresentados, respectivamente, nas Figura 18a e Figura 18b.

Figura 18 – Cartas de controle EWMA para o monitoramento da tendência dos dados (os símbolos conectados por linhas finas representam os dados originais enquanto as linhas grossas mostram a tendência) para os barcos de evaporação 1, 2 e 15 das corridas A2 (a) e A5 (b)



Fonte: A Autora (2020)

Principalmente na Figura 18a, é possível perceber a similaridade entre as medidas de DO adquiridas ao longo dos barcos 1 e 2, e a dissimilaridade entre essas medidas em relação aos valores coletados para o barco 15. Adicionalmente, ao comparar ambas as figuras, é possível identificar que evaporadores na mesma posição (por vezes, o mesmo evaporador) apresentam um comportamento diferente para corridas distintas.

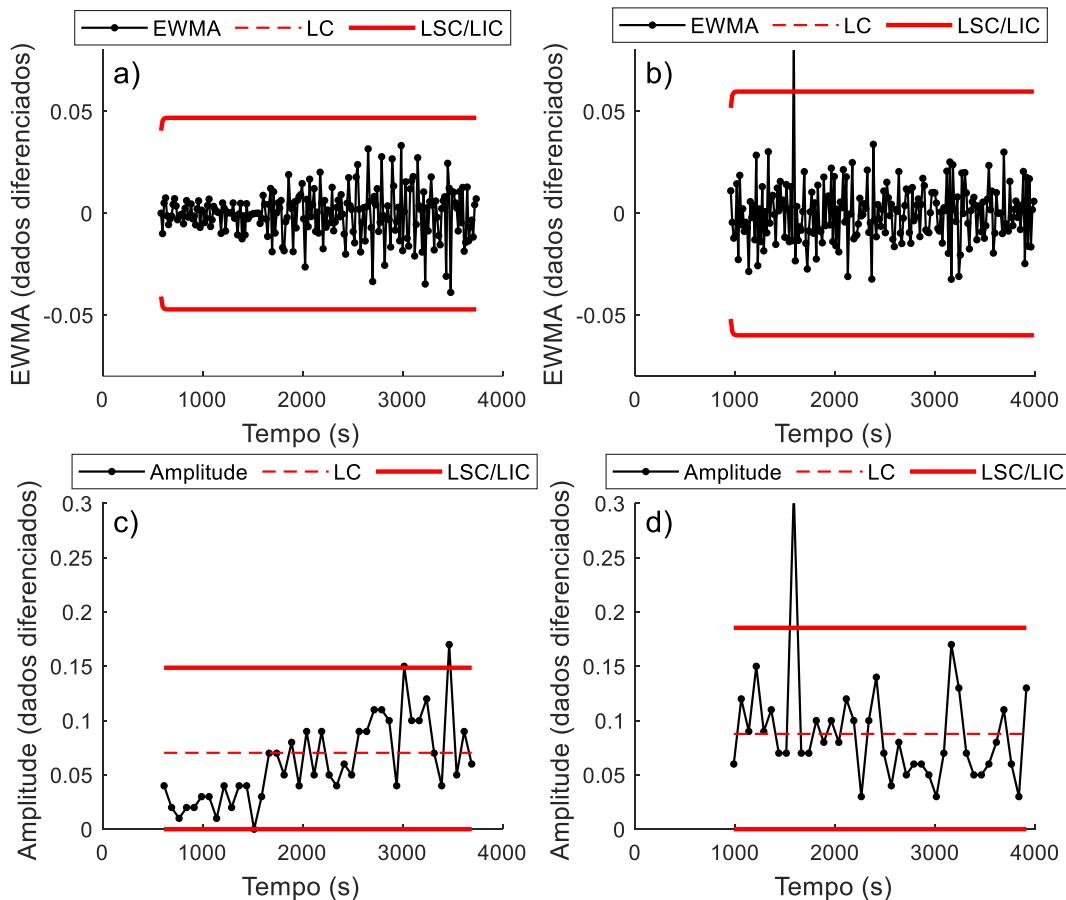
Outra estratégia que pode ser utilizada para o monitoramento do processo envolve o uso de cartas de controle univariadas para o monitoramento da média e da variabilidade do processo. A maioria das cartas de controle convencionais utilizadas para esse fim, entretanto, assumem que os dados são distribuídos de forma independente ao longo do tempo (MONTGOMERY, 2009). Conforme discutido anteriormente, o processo estudado apresenta dados autocorrelacionados, violando a suposição de independência temporal. Existem diversas técnicas que podem ser utilizadas para lidar com esse tipo de dado (REYNOLDS; LU, 1997; LU; REYNOLDS JR., 2001). No presente trabalho, a seguinte estratégia foi adotada: a tendência foi previamente removida pelo processo de diferenciação e, então, cartas de controle convencionais foram desenvolvidas com base nos dados diferenciados. Essa estratégia foi selecionada por ser considerada de fácil interpretação e implementação no processo estudado, quando comparado com outras abordagens como a de modelagem de séries temporais e posterior monitoramento dos resíduos.

Embora a tendência possa ser importante para conhecimento dos dados estudados, ela frequentemente pode prejudicar a interpretação dos dados do processo e, portanto, deve ser removida (SHUMWAY; STOFFER, 2017). No presente trabalho, a primeira, segunda e terceira ordens de diferenciação foram aplicadas para a remoção da tendência. Observou-se que, em geral, a segunda e terceira ordens de diferenciação introduziram ruído aos dados, o que motivou a escolha da primeira ordem, que permitiu, de um modo geral, a efetiva remoção da tendência, conforme ilustrado na Figura 5-1 do APÊNDICE F.

Após a seleção da ordem de diferenciação adequada, cartas de controle foram desenvolvidas com base nos dados diferenciados, para monitoramento da média e da variabilidade. Para monitoramento da média, a carta de controle EWMA foi selecionada pois, segundo Montgomery (2009), ela permite a detecção de desvios de pequena magnitude no processo. A análise da carta EWMA para os dados diferenciados, em geral, resultou em valores aleatoriamente distribuídos, sem uma tendência evidente. No que diz respeito ao monitoramento da variabilidade do processo, cartas de controle da amplitude (R) foram também construídas a partir dos dados diferenciados. Para ilustrar a aplicação dessas

estratégias, cartas EWMA e da amplitude (R) estão apresentadas na Figura 19, para o monitoramento do barco de evaporação 1 da corrida A2 (Figura 19a e c) e do barco 15 da corrida A5 (Figura 19b e d).

Figura 19 – Carta de controle EWMA nos dados diferenciados para o barco de evaporação 1 da corrida A2 (a) e o barco 15 da corrida A5 (b); e carta de controle da amplitude (R) nos dados diferenciados para o barco de evaporação 1 da corrida A2 (c) e o barco 15 da corrida A5 (d). As linhas contínuas em vermelho representam os limites superior e inferior (LSC / LIC) de controle e a linha tracejada em vermelho o limite central (LC) para os gráficos de controle



Fonte: A Autora (2020)

Conforme observado na Figura 19a (carta EWMA para o barco 1 da corrida A2), a média do processo se encontra dentro dos limites de controle. Já para o barco 15 da corrida A5, (Figura 19b), é possível identificar uma situação fora de controle evidenciada por um pico em torno de 1600 s. Esse tipo de distúrbio é comumente observado em caso de falhas na medição do sensor ou no sistema de aquisição de dados (JIN; GUO; ZHOU, 2006). Se o mesmo ocorrer com muita frequência, no entanto, uma investigação mais elaborada do processo deve ser realizada com o intuito de corrigir esse tipo de perturbação.

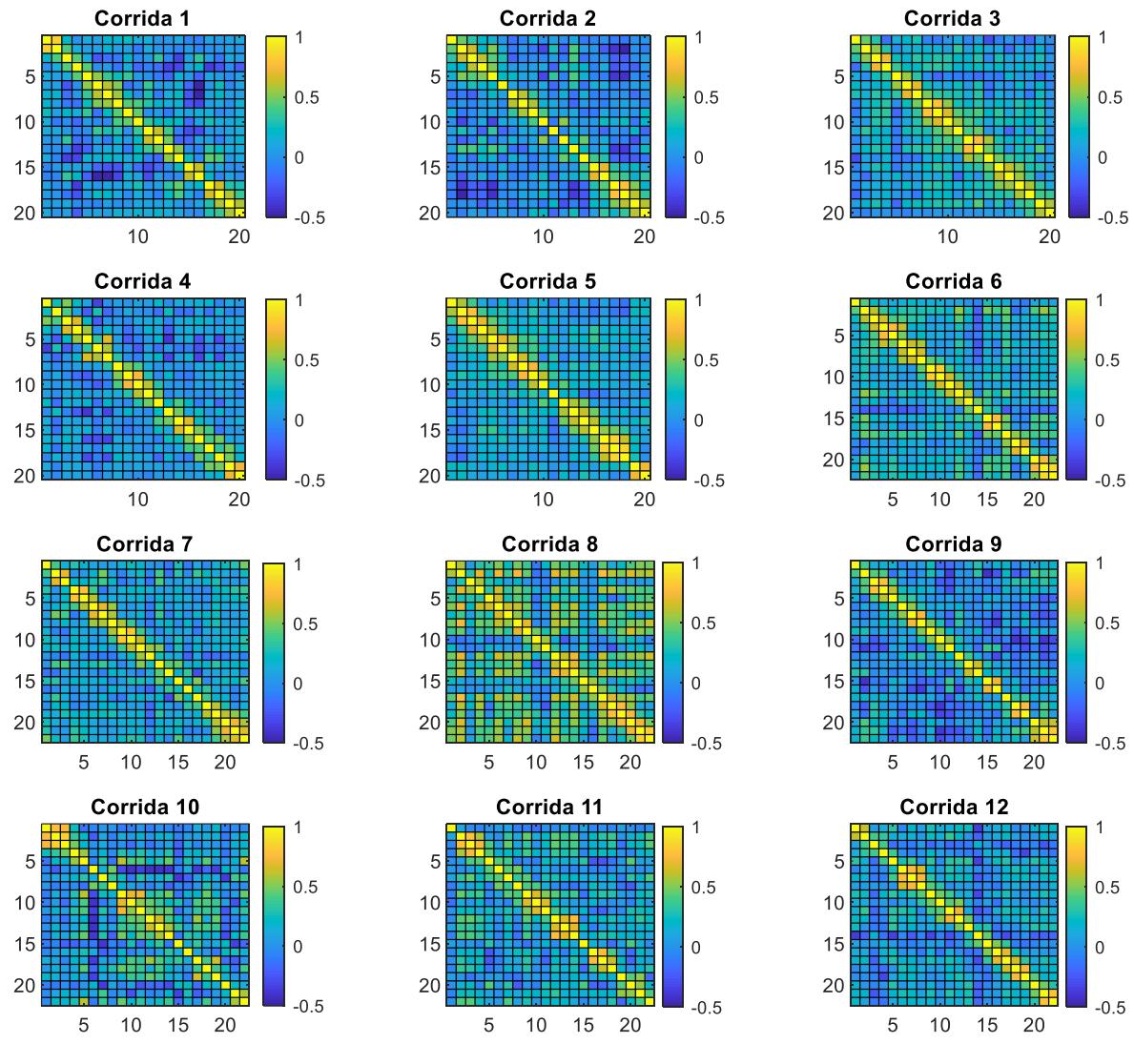
No que diz respeito ao monitoramento da variabilidade do processo para o barco 1 da corrida A2 (Figura 19c), está indicada no gráfico da amplitude uma tendência ascendente na variabilidade dos dados ao longo do processo. De um modo geral, para as duas primeiras corridas (A1 e A2), esse comportamento foi identificado para alguns barcos de evaporação (ver APÊNDICE G para a corrida A2). Apenas com a análise realizada, não é possível garantir a causa para esse tipo de ocorrência. No entanto, uma possível explicação pode estar relacionada às particularidades associadas ao primeiro uso de um barco de evaporação. Basicamente, conforme discutido por Bishop (2011) e (2015), barcos de evaporação novos podem conter compostos voláteis e umidade. Sendo assim, é essencial que se respeite a rampa de aquecimento indicada pelo fabricante com o intuito de eliminar esses compostos e, dessa forma, estabilizar o evaporador para que seja utilizado na corrida seguinte. Para as corridas produzidas na sequência, de A3 em diante, essa tendência de variabilidade crescente ao longo do processo não foi evidenciada (de forma ilustrativa, ver Figura 19d para o barco 15 corrida A5).

Nas últimas corridas (mais especificamente, A6, A7 e A8), as cartas de controle EWMA e R apresentaram, de um modo geral, mais sinais fora de controle, sugerindo um processo mais instável. Durante as oito corridas (A1 a A8), não se obteve informação de que algum evaporador tenha sido substituído. Portanto, a instabilidade das medições de DO para as últimas corridas dessa produção em sequência pode estar relacionada à vida útil dos barcos de evaporação. Conforme discutido por Bishop (2011) e (2015), no final da vida útil do barco, seu desempenho geralmente é reduzido. Essa redução é comumente atribuída às reações químicas que ocorrem no processo, as quais podem ocasionar a erosão e a redução da resistividade do barco. Além disso, pode estar associada também ao processo de corrosão e desgaste do barco, devido à natureza altamente corrosiva do alumínio fundido.

#### **4.2.3 Avaliação da estratégia de monitoramento proposta**

Com o intuito de avaliar se a estratégia de monitoramento empregada poderia ser utilizada para monitoramento de outras corridas do processo e se o comportamento observado para as corridas A1 a A8 se repetiria, os dados de DO adquiridos ao longo da produção sequencial de 12 novas corridas (B1 a B12) foram analisadas seguindo o mesmo procedimento. Assim, inicialmente, foram avaliados os mapas de correlação para essas novas corridas, apresentados na Figura 20.

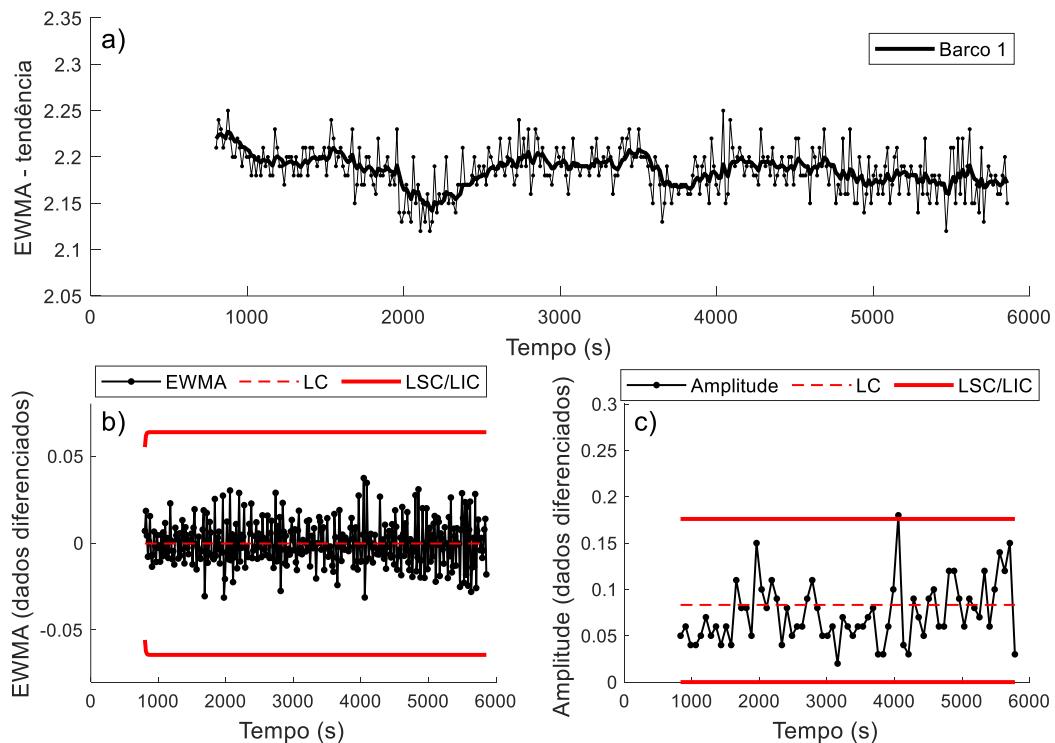
Figura 20 – Mapas de correlação para as corridas B1 a B12. As escalas de cores estão relacionadas com os coeficientes de correlação entre as variáveis



Fonte: A Autora (2020)

Em geral, conforme observado na Figura 20, um comportamento semelhante foi observado para as corridas B1 a B12, principalmente no que se refere à baixa correlação entre as variáveis e à ausência de um padrão de correlação quando diferentes corridas são comparadas. A abordagem de monitoramento proposta poderia, portanto, ser usada no monitoramento desses novos dados. Para ilustrar essa utilização, a estratégia de monitoramento para a variável 1 da corrida B2 está apresentada na Figura 21.

Figura 21 - Estratégia de monitoramento proposta para a corrida B2: carta EWMA para monitoramento da tendência dos dados (símbolos conectados por uma linha fina representam os dados originais enquanto a linha grossa mostra a tendência) (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 1



Fonte: A Autora (2020)

Conforme apresentado na Figura 21, é possível perceber que, a variável 1 da corrida B2, o processo está em controle estatístico. Apenas um pico fora de controle foi detectado em torno de 4000 s. Em comparação com as cartas de controle para as corridas A1 a A8, em geral, as corridas B1 a B12 apresentaram um comportamento similar em relação ao desempenho do evaporador ao longo de sua vida útil. Corridas produzidas usando evaporadores com um tempo de vida mais longo, geralmente, apresentaram um comportamento mais instável, sinalizando mais valores fora de controle nas cartas EWMA e R.

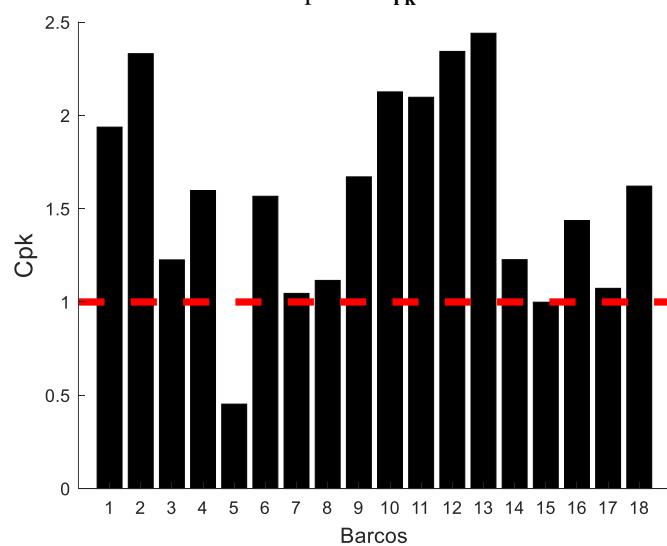
Em relação às particularidades associadas ao primeiro uso do evaporador, com o intuito de reduzir os efeitos disso no processo, as corridas que utilizaram evaporadores novos (B1 e B8) foram conduzidas com a aplicação de uma velocidade de filme mais baixa. Como possível resultado dessa estratégia adotada, a análise das cartas de controle para essas corridas (B1 e B8) não evidenciou um aumento na variabilidade do processo ao longo da produção, como foi observado para as corridas A1 e A2 (resultados não apresentados devido à grande quantidade de gráficos gerados).

#### 4.2.4 Análise da capacidade do processo

Além das cartas de monitoramento estatístico apresentadas, o presente trabalho sugere também a avaliação da capacidade do processo, com o objetivo de estimar a sua variabilidade em relação aos limites de especificação do produto e de ajudar na eliminação ou redução dessa variabilidade. Uma maneira bem estabelecida de expressar a capacidade de um processo descentralizado consiste em estimar a sua razão de capacidade ( $C_{pk}$ ). De modo geral, para um  $C_{pk}$  de valor unitário, o processo é geralmente considerado razoavelmente capaz e quanto maior o valor de  $C_{pk}$ , maior a capacidade do processo de atender às especificações do produto. Essa análise deve ser realizada, no entanto, em conjunto com a avaliação das cartas de monitoramento, uma vez que é válida para um processo em controle estatístico (MONTGOMERY, 2009).

Para ilustrar como a análise de capacidade pode ser aplicada ao processo em estudo, estão apresentados na Figura 22 os valores de  $C_{pk}$  para os 18 barcos de evaporação da corrida A2.

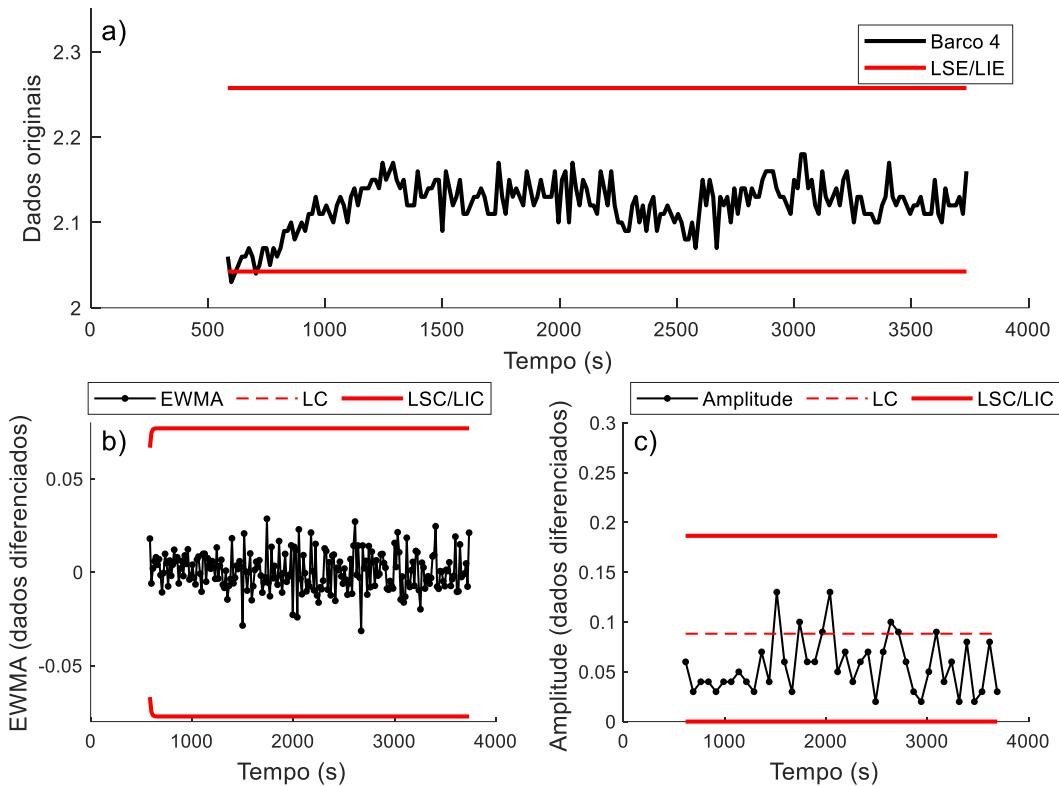
Figura 22 – Valores de  $C_{pk}$  para cada barco de evaporação da corrida A2. A linha tracejada é o limite unitário para o  $C_{pk}$



Fonte: A Autora (2020)

Como pode ser visto na Figura 22, o valor de  $C_{pk}$  foi superior a um para todos os barcos, exceto o 5. Tomando-se como exemplo o barco 4, este apresentou um valor de  $C_{pk}$  maior do que 1,5. O perfil de DO e as cartas EWMA e R para esse barco estão apresentados na Figura 23.

Figura 23 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 4 da corrida A2

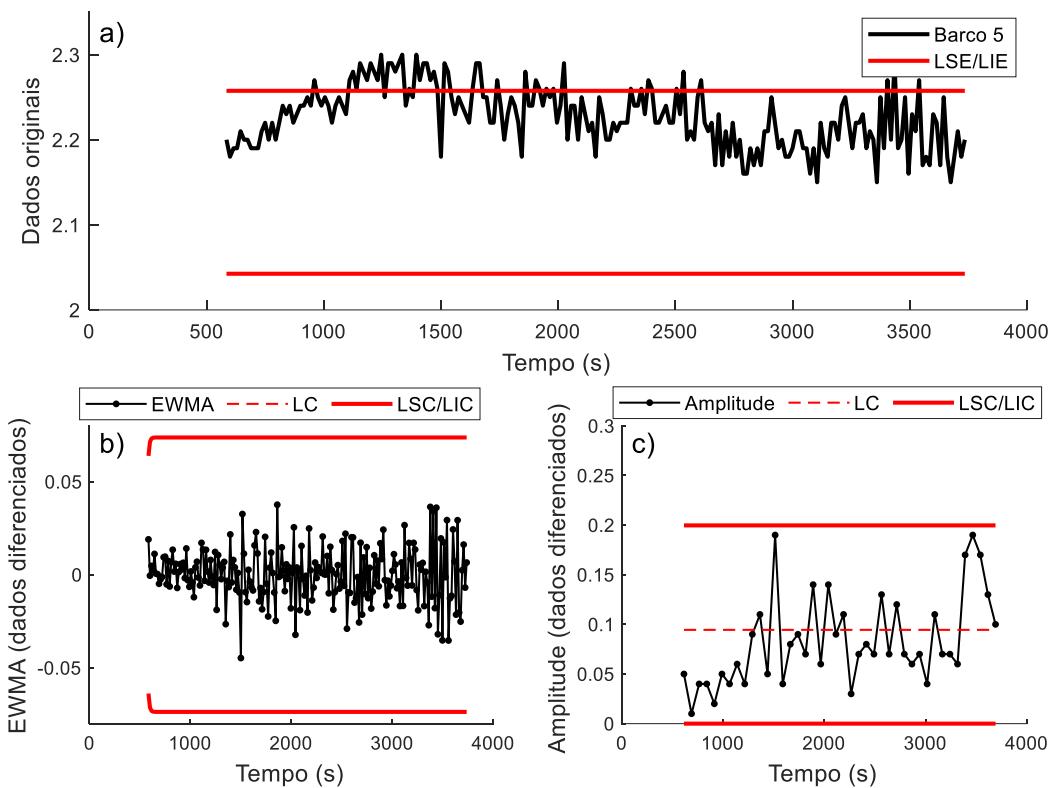


Fonte: A Autora (2020)

A partir da análise da Figura 23, é possível perceber que o barco 4 apresentou apenas alguns poucos pontos fora dos limites de especificação (Figura 23a), o que sugere que o evaporador estava em boas condições. Adicionalmente, as cartas EWMA (Figura 23b) e R (Figura 23c) para essa variável não sinalizaram pontos fora dos limites de controle, apenas a carta da amplitude apresentou algumas sequências de pontos abaixo da linha central do gráfico. Sendo assim, pode-se concluir que o processo é, de fato, capaz de atender às especificações para essa variável.

Tomando-se como exemplo agora o barco de evaporação 5 da corrida A2, vê-se que este apresentou um valor de  $C_{pk}$  bem menor do que 1 (Figura 22). O perfil de DO e as cartas EWMA e R para esse barco estão apresentados na Figura 24.

Figura 24 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 5 da corrida A2

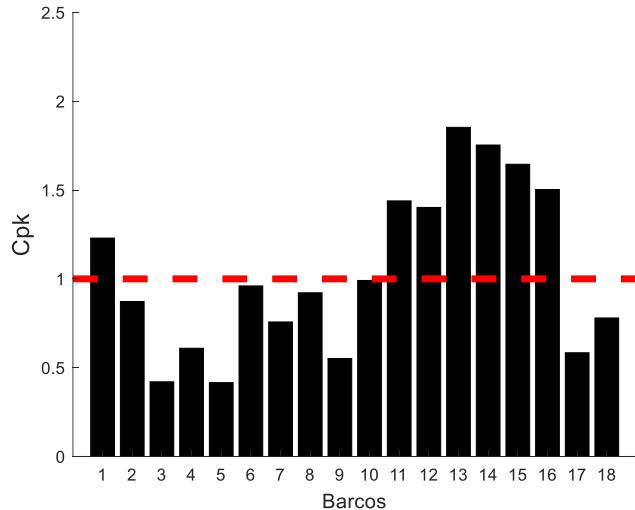


Fonte: A Autora (2020)

Quando o perfil de DO do barco 5 é analisado (Figura 24a), percebe-se que as medidas para esse barco, de fato, apresentaram diversos valores acima do limite superior de especificação (LSE). No entanto, o resultado da análise de capacidade para esse barco deve ser avaliado com cautela, uma vez que a carta de monitoramento da amplitude, apresentada Figura 24c, indica uma tendência crescente da variabilidade ao longo do processo, além de algumas sequências de pontos abaixo da linha central. Essa tendência está associada, provavelmente, à estabilização do barco durante as primeiras vezes que é utilizado, conforme discutido anteriormente. Dessa forma, esse comportamento pode indicar uma situação fora de controle estatístico, o que prejudica a análise do  $C_{pk}$ .

Outro exemplo da análise da capacidade do processo pode ser visto na Figura 25, em que estão apresentados os valores de  $C_{pk}$  para os barcos de evaporação da corrida A7. Essa foi a penúltima corrida da produção sequencial e, provavelmente, barcos de evaporação com um tempo de serviço maior foram utilizados, o que pode levar a algumas situações fora de controle.

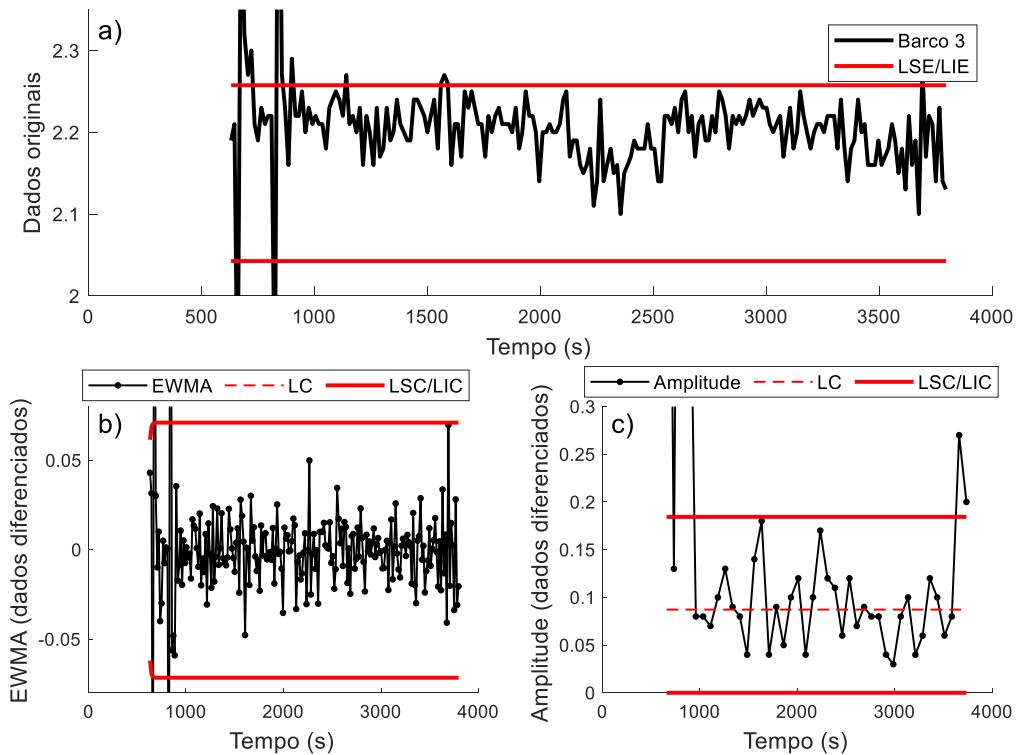
Figura 25 - Valores de  $C_{pk}$  para cada barco de evaporação da corrida A7. A linha tracejada é o limite unitário para o  $C_{pk}$



Fonte: A Autora (2020).

Como pode ser visto na Figura 25, para a corrida A7, os valores de  $C_{pk}$  ficaram abaixo do limite unitário para diversos evaporadores. De forma a ilustrar a análise, pode-se tomar como exemplo o barco 3, que apresentou um valor de  $C_{pk}$  bem abaixo do limite. O perfil de DO e as cartas EWMA e R para esse barco estão apresentados na Figura 26.

Figura 26 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 3 da corrida A7

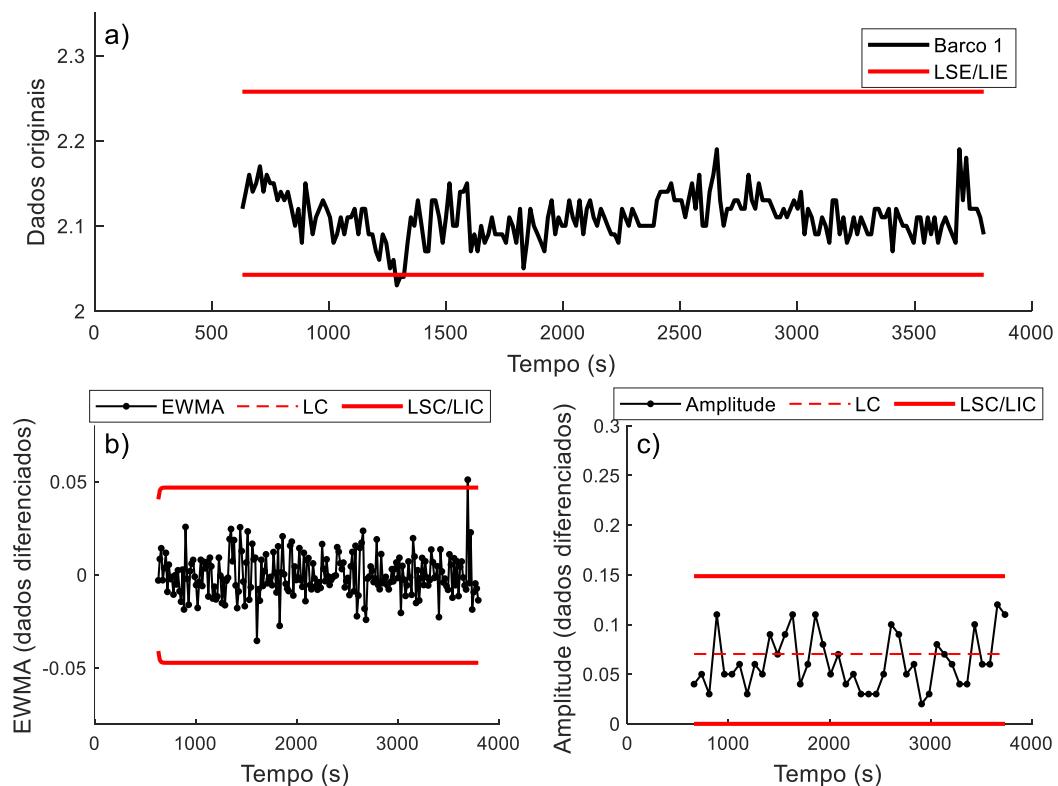


Fonte: A Autora (2020)

A análise do perfil de DO para o barco 3 (Figura 26a) indicou a ocorrência de alguns pontos fora de especificação, principalmente, entre 600 e 900 s de produção. Adicionalmente, a avaliação das cartas EWMA e R para essa variável, (Figura 26b e c, respectivamente), detectaram alguns pontos fora de controle no início e no fim da corrida. Como consequência, a análise de capacidade deve ser realizada com parcimônia, já que o processo não estava em controle estatístico para essa variável.

Considerando ainda a corrida A7, pode-se pegar outro exemplo para ilustrar a estratégia de monitoramento proposta. O barco de evaporação 1, por exemplo, que foi um dos barcos que apresentou um valor de  $C_{PK}$  acima do limite unitário, conforme ilustrado na Figura 25. O perfil de DO e as cartas EWMA e R para essa variável estão apresentados na Figura 27.

Figura 27 – Perfis de DO e limites superior (LSE) e inferior (LIE) de especificação (a), carta EWMA nos dados diferenciados (b) e carta de controle R nos dados diferenciados (c), para a variável 1 da corrida A7



Fonte: A Autora (2020)

Conforme observado na Figura 27, de modo geral, quase todas as medidas de DO para essa variável ficaram dentro dos limites de especificação (Figura 27a). As cartas de controle EWMA e R, ilustradas, respectivamente, na Figura 27b e c, também indicaram que o processo

estava em controle estatístico para essa variável, possibilitando a análise da capacidade do processo de forma coerente.

Enquanto a estratégia de monitoramento estatístico proposta pode ser utilizada para o monitoramento do processo em tempo real, a análise de capacidade sugerida deve ser usada após a conclusão de cada corrida. Embora ela não auxilie nas ações corretivas ao longo do processo, são úteis para indicar, de forma mais rápida e clara, as posições dos evaporadores em que mais valores de DO foram obtidos fora de especificação. Isso permite, portanto, guiar os ajustes realizados entre a metalização de um filme e outro.

#### 4.3 CONCLUSÃO

O presente capítulo apresentou uma estratégia de monitoramento estatístico para as medidas de DO de filmes de poliéster metalizados, adquiridas ao longo de um processo industrial de metalização a vácuo utilizando barcos de evaporação aquecidos por resistência. Um estudo preliminar para detectar tendência e correlação nos dados foi apresentado e a baixa correlação entre as medidas coletadas ao longo de diferentes barcos motivou a utilização de técnicas de monitoramento univariadas. A estratégia proposta consistiu no uso de uma carta EWMA para o monitoramento da linha de tendência dos dados, e cartas EWMA e da amplitude (R), construídas com base nos dados diferenciados, para o monitoramento da média e da variabilidade do processo, respectivamente. A ideia aplicada no desenvolvimento dessas cartas visou ao monitoramento do processo em tempo real, de forma que elas possam ser usadas para auxiliar os operadores em relação às ações corretivas requeridas ao longo do processo. Além disso, o uso dessas cartas apresenta um potencial para redução da variabilidade do processo, buscando-se uma deposição metálica mais consistente e uniforme ao longo de todo o filme.

Adicionalmente, o presente capítulo também apresentou uma análise de capacidade para o processo. No entanto, conforme discutido, essa análise precisa ser utilizada com parcimônia, já que a mesma deve ser aplicada quando o processo estiver em controle estatístico. Ao contrário das cartas de monitoramento estatístico, a análise de capacidade do processo proposta visa ao monitoramento do processo no modo *off-line*, após a sua conclusão. É uma forma mais simples e rápida de guiar os ajustes que são realizados entre corridas, mostrando a posição do evaporador que apresentou mais valores fora de especificação.

Por fim, a ideia principal é que a utilização das ferramentas de monitoramento estatístico apresentadas, através de uma análise mais elaborada dos dados de DO já coletados e disponíveis para uso, possam auxiliar no monitoramento do processo, com o intuito de reduzir a sua variabilidade e, consequentemente, melhorar o desempenho de barreira dos filmes metalizados produzidos.

## 5 PERSPECTIVAS

Em relação ao monitoramento da produção de biodiesel, uma dificuldade encontrada no desenvolvimento do presente trabalho consistiu na coleta de amostras da mistura reacional em outros pontos da coluna de destilação reativa utilizada. Dessa forma, trabalhos futuros poderiam otimizar essas coletas de modo que permitissem a aquisição de espectros em diversos pontos da coluna de destilação reativa. A ideia seria obter um perfil da concentração de ésteres alquílicos (biodiesel) ao longo da coluna, o que poderia auxiliar na definição de suas condições operacionais.

Além disso, os modelos de regressão PLS desenvolvidos no presente trabalho, para estimar as concentrações de éster alquílico, etanol e glicerol, poderiam ser alimentados com mais espectros provenientes de amostras da coluna e de amostras quaternárias, adquiridos na faixa de temperatura de 50 – 60°C. Conforme discutido, a inclusão de algumas amostras nessas condições nos modelos promoveu uma melhoria dos resultados. Dessa forma, se mais amostras fossem incluídas, os resultados apresentados poderiam ser ainda melhores.

No que diz respeito ao monitoramento da metalização de filmes poliméricos, trabalhos futuros poderiam viabilizar o uso de métodos multivariados para o monitoramento estatístico do processo. Conforme discutido, a baixa correlação entre as variáveis e as características do processo motivaram a aplicação de técnicas univariadas no presente trabalho. No entanto, uma desvantagem associada ao uso dessas técnicas está relacionada à quantidade de gráficos que precisam ser analisados. Dessa forma, métodos multivariados não convencionais poderiam ser aplicados a esse processo com o intuito de reduzir a quantidade de gráficos e, assim, facilitar o monitoramento. É importante ressaltar que o uso de métodos multivariados tradicionais tem sido avaliado, no entanto, resultados satisfatórios ainda não foram obtidos, o que indica a necessidade de um maior aprofundamento no tema.

## REFERÊNCIAS

- AGELET, L. E.; HURBURGH, C. R. A tutorial on near infrared spectroscopy and its calibration. **Critical Reviews in Analytical Chemistry**, v. 40, p. 246–260, 2010.
- ANDERSEN, C. M.; BRO, R. Variable selection in regression - a tutorial. **Journal of Chemometrics**, v. 24, p. 728–737, 2010.
- ANDERSSON, M. A comparison of nine PLS1 algorithms. **Journal of Chemometrics**, v. 23, p. 518–529, 2009.
- ANP. **Óleo diesel brasileiro passa a conter mínimo de 12% de biodiesel**. 2020.
- ASTM E1655-05. **Standard Practices for Infrared Multivariate Quantitative Analysis**. American Standard Methods. 2012.
- BAKDI, A.; KOUADRI, A. A new adaptive PCA based thresholding scheme for fault detection in complex systems. **Chemometrics and Intelligent Laboratory Systems**, v. 162, p. 83–93, 2017.
- 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, p. 772–777, 1989.
- BEEBE, K. R. **Chemometrics: a practical guide**. : John Wiley & Sons, 1998.
- BIN SHAMS, M. A.; BUDMAN, H. M.; DUEVER, T. A. Fault detection, identification and diagnosis using CUSUM based PCA. **Chemical Engineering Science**, v. 66, p. 4488–4498, 2011.
- BISHOP, C. A. **Vacuum deposition onto webs, films and foils**. 2. ed. : Elsevier, 2011.
- BISHOP, C. A. **Roll-to-Roll Vacuum Deposition of Barrier Coatings**. 2nd. ed. Beverly: Scrivener Publishing LLC, 2015.
- BISHOP, C. A.; MOUNT, E. M. Vacuum Metallizing for Flexible Packaging. In: WAGNER, J. R. (Ed.). **Multilayer Flexible Packaging**. 2nd. ed. : Elsevier, 2016. p. 235–255.
- BOON-ANUWAT, N.; KIATKITTIPONG, W.; AIOUACHE, F.; ASSABUMRUNGRAT, S. Process design of continuous biodiesel production by reactive distillation: Comparison between homogeneous and heterogeneous catalysts. **Chemical Engineering and Processing: process intensification**, v. 92, p. 33–44, 2015.
- BORROR, C. M.; MONTGOMERY, D. C.; RUNGER, G. C. Robustness of the EWMA Control Chart to Non-Normality. **Journal of Quality Technology**, v. 31, p. 309–316, 1999.
- BRO, R.; SMILDE, A. K. Principal component analysis. **Analytical Methods**, v. 6, p. 2812–2831, 2014.
- BURNS, D. A.; CIURCZAK, E. W. **Handbook of Near Infrared Analysis**. 3. ed. Boca Raton: CRC Press, 2008.

CHANG, W.R.; HWANG, J.J.; WU, W. Environmental impact and sustainability study on biofuels for transportation applications. **Renewable and Sustainable Energy Reviews**, v. 67, p. 277–288, 2017.

CHIAPPINI , F.A.; GOICOECHEA, H.C.; OLIVIERI, A.C. MVC1\_GUI: A MATLAB An upgrade including artificial neural networks modelling graphical user interface for first-order multivariate calibration. **Chemometrics and Intelligent Laboratory Systems**, v. 206, p. 104162-104172, 2020.

CNPE - Conselho Nacional de Política Energética. **Resolução nº16, de 29 de outubro de 2018.** 2018.

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

DA SILVA, N. C.; CVALCANTI, C. J.; HONORATO, F. A.; AMIGO, J. M.; PIMENTEL, M. F. Standardization from a benchtop to a handheld NIR spectrometer using mathematically mixed NIR spectra to determine fuel quality. **Analytica Chimica Acta**, v. 954, p. 32–42, 2017.

DA SILVA, N. C.; MASSA, A. R. C. de G.; DOMINGOS, D.; AMIGO, J. M.; REBOUÇAS, M. das V.; PASQUINI, C.; PIMENTEL, M. F. NIR-based octane rating simulator for use in gasoline compounding processes. **Fuel**, v. 243, p. 381–389, 2019.

DE BEER, T.; BURGGRAEVE, A.; FONTEYNNE, M.; SAERENS, L.; REMON, J. P.; VERVAET, C. Near infrared and Raman spectroscopy for the in-process monitoring of pharmaceutical production processes. **International Journal of Pharmaceutics**, v. 417, p. 32–47, 2011.

DE 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.

DU, Y.; DU, D. Fault detection and diagnosis using empirical mode decomposition based principal component analysis. **Computers and Chemical Engineering**, v. 115, p. 1–21, 2018.

DUCHESNE, C.; LIU, J. J.; MACGREGOR, J. F. Multivariate image analysis in the process industries: A review. **Chemometrics and Intelligent Laboratory Systems**, v. 117, p. 116–128, 2012.

EISENSTECKEN, D.; STÜRZ, B.; ROBATSCHER, P.; LOZANO, L.; ZANELLA, A.; OBERHUBER, M. Postharvest Biology and The potential of near infrared spectroscopy (NIRS) to trace apple origin: Study on different cultivars and orchard elevations. **Postharvest Biology and Technology**, v. 147, p. 123–131, 2019.

EN14103. **Fat and oil derivatives – Fatty Acid Methyl Esters (FAME) – Determination of ester and linolenic acid methyl ester contents.** 2003.

ESBENSEN, K. H.; GELADI, P. Principal Component Analysis: Concept, Geometrical

Interpretation, Mathematical Background, Algorithms, History, Practice. In: BROWN, S. D.; TAUER, R.; WALCZAK, B. (Eds.). **Comprehensive Chemometrics: Chemical and Biochemical Data Analysis**. 2. ed. : Elsevier B.V., 2009.

FDA. **Guidance for Industry PAT — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance**. 2004. Disponível em: <<https://www.fda.gov/media/71012/download>>. Acesso em: 13 fev. 2020.

FELIZARDO, P.; CORREIA, M. J. N.; RAPOSO, I.; MENDES, J. F.; BORDADO, J. M.; BERKEMEIER, R. Production of biodiesel from waste frying oils. **Waste management**, v. 26, p. 487–494, 2006.

FERREIRA, A. P.; TOBYN, M. Multivariate analysis in the pharmaceutical industry: enabling process understanding and improvement in the PAT and QbD era. **Pharmaceutical Development and Technology**, v. 7450, p. 1–15, 2014.

FERRER-RIQUELME, A. J. Statistical Control of Measures and Processes. In: WALCZAK, B.; BROWN, S. D.; TAUER, R. (Eds.). **Comprehensive Chemometrics: Chemical and Biochemical Data Analysis**. 2. ed. : Elsevier, 2009. p. 97–126.

FERRER, A. Latent Structures-Based Multivariate Statistical Process Control : A Paradigm Shift. **Quality Engineering**, v. 26, p. 72–91, 2014.

FREEDMAN, B.; BUTTERFIELD, R. O.; PRYDE, E. Transesterification Kinetics of SSoybean Oil. **Journal of the American Oil Chemists' Society (JAOCS)**, v. 63, p. 1375–1380, 1986.

GALAVERNA, R.; RIBESSI, R. L.; ROHWEDDER, J. J. R.; PASTRE, J. C. Coupling Continuous Flow Microreactors to MicroNIR Spectroscopy: Ultra-compact Device for Facile In-Line Reaction Monitoring. **Organic Process Research and Development**, v. 22, p. 780–788, 2018.

GALVÃO, R. K. H.; ARAUJO, M. C. U.; JOSÉ, 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.

GE, Z. Review on data-driven modeling and monitoring for plant-wide industrial processes. **Chemometrics and Intelligent Laboratory Systems**, v. 171, p. 16–25, 2017.

GELINSKI, E. K.; HAMERSKI, F.; CORAZZA, M. L.; SANTOS, A. F. Biodiesel synthesis monitoring using Near Infrared Spectroscopy. **The Open Chemical Engineering Journal**, v. 12, p. 95–110, 2018.

GRASSI, S.; ALAMPRESE, C. Advances in NIR spectroscopy applied to Process Analytical Technology in food industries. **Current Opinion in Food Science**, v. 22, p. 17–21, 2018.

GROBHANS, S.; RÜDT, M.; SANDEN, A.; BRESTRICH, N.; MORGESTERN, J.; HEISSLER, S.; HUBBUCH, J. In-line Fourier-transform infrared spectroscopy as a versatile process analytical technology for preparative protein chromatography. **Journal of Chromatography A**, v. 1547, p. 37–44, 2018.

HARROU, F.; NOUNOU, M. N.; NOUNOU, H. N.; MADAKYARU, M. PLS-based EWMA fault detection strategy for process monitoring. **Journal of Loss Prevention in the Process Industries**, v. 36, p. 108–119, 2015.

HE, Q. P.; WANG, J. Statistical process monitoring as a big data analytics tool for smart manufacturing. **Journal of Process Control**, v. 67, p. 35–43, 2018.

HERRMANN, M.; RÄTHEL, J.; HÖHN, S.; EICHLER, J.; MICHAELIS, A. Interaction of titanium diboride/boron nitride evaporation boats with aluminium. **Journal of the European Ceramic Society**, v. 31, p. 2401–2406, 2011.

HOHL, R.; SCHEIBELHOFER, O.; STOCKER, E.; BEHZADI, S. S.; HAACK, D.; KOCH, K.; KERSCHHAGGL, P.; LOCHMANN, D.; SACHER, S.; ZIMMER, A. Monitoring of a Hot Melt Coating Process via a Novel Multipoint Near-Infrared Spectrometer. **AAPS PharmSciTech**, v. 18, p. 182–193, 2017.

HUMAN, S. W.; KRITZINGER, P.; CHAKRABORTI, S. Robustness of the EWMA control chart for individual observations. **Journal of Applied Statistics**, v. 38, p. 2071–2087, 2011.

JIN, J.; GUO, H.; ZHOU, S. Statistical process control based supervisory generalized predictive control of thin film deposition processes. **Journal of Manufacturing Science and Engineering**, v. 128, p. 315–325, 2006.

JODA, F.; AHMADI, F. Exergoeconomic analysis of conventional and using reactive distillation biodiesel production scenarios thermally integrated with a combined power plant. **Renewable Energy**, v. 132, p. 898–910, 2019.

KADLEC, P.; GABRYS, B.; STRANDT, S. Data-driven Soft Sensors in the process industry. **Computers and Chemical Engineering**, v. 33, p. 795–814, 2009.

KADRI, F.; HARROU, F.; CHAABANE, S.; SUN, Y.; TAHON, C. Seasonal ARMA-based SPC charts for anomaly detection: Application to emergency department systems. **Neurocomputing**, v. 173, p. 2102–2114, 2016.

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.; RAZON, L. F. Biodiesel fuels. **Progress in Energy and Combustion Science**, v. 58, p. 36–59, 2017.

LIU, Y.; XIE, M. Rebooting data-driven soft-sensors in process industries : A review of kernel methods. **Journal of Process Control**, v. 89, p. 58–73, 2020.

LU, C.; REYNOLDS JR., M. R. Cusum Charts for Monitoring an Autocorrelated Process Cusum. **Journal of Quality Technology**, v. 33, p. 316–334, 2001.

MA, X.; ZHANG, L.; HU, J.; PALAZOGLU, A. A model - free approach to reduce the effect of autocorrelation on statistical process control charts. **Journal of Chemometrics**, v. 32, p. 1–24, 2018.

MAGWAZA, L. S.; OPARA, U. L.; NIEUWOUDT, H.; CRONJE, P. J. R.; SAEYS, W. NIR

Spectroscopy Applications NIR Spectroscopy Applications for Internal and External Quality Analysis of Citrus Fruit — A Review. **Food Bioprocess Technology**, v. 5, p. 425–444, 2012.

MARQUES, E. J. N.; DE FREITAS, S. T.; PIMENTEL, M. F.; PASQUINI, C. Rapid and non-destructive determination of quality parameters in the “Tommy Atkins” mango using a novel handheld near infrared spectrometer. **Food Chemistry**, v. 197, p. 1207–1214, 2016.

MARTENS, H.; STARK, E. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. **Journal of Pharmaceutical & Biomedical Analysis**, v. 9, p. 625–635, 1991.

MILLS, T. C. **The Foundations of Modern Time Series Analysis**. 1. ed. : Palgrave Advanced Texts in Econometrics, 2011.

MIRSCHEL, G.; DAIKOS, O.; SCHERZER, T.; STECKERT, C. Near-infrared chemical imaging used for in-line analysis of functional finishes on textiles. **Talanta**, v. 188, p. 91–98, 2018. Disponível em: <<https://doi.org/10.1016/j.talanta.2018.05.050>>

MODROÑO, S.; SOLDADO, A.; MARTÍNEZ-FERNÁNDEZ, A.; DE LA ROZA-DELGADO, B. Handheld NIRS sensors for routine compound feed quality control: real time analysis and fiefs monitoring. **Talanta**, v. 162, p. 597–603, 2017.

MONTGOMERY, D. C. **Introduction to Statistical Quality Control**. 6th. ed. : John Wiley & Sons, 2009.

MOUNT, E. M. Impact of metallizer process controls on optical and gas barrier uniformity. **Journal of Plastic Film and Sheeting**, v. 24, n. 3–4, p. 203–211, 2008.

O'BRIEN, N. A.; HULSE, C. A.; FRIEDRICH, D. M.; VAN MILLIGEN, F. J.; VON GUNTEN, M. K.; PFEIFER, F.; SIESLER, H. W. Miniature Near-Infrared (NIR) Spectrometer Engine For Handheld Applications. **Next-Generation Spectroscopic Technologies V**, v. 8374, p. 3–9, 2012.

OLIVIERI, A. C. **Chemometrics and Multivariate Calibration**: a pratical approach. 1. ed. : Springer, 2018.

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: A mature analytical technique with new perspectives – A review. **Analytica Chimica Acta**, v. 1026, p. 8–36, 2018.

PAVIA, D. L.; LAMPMAN, G. M.; KRIZ, G. S.; VYVYAN, J. R. **Introduction to spectroscopy**. 4. ed. : Brooks/Cole, 2009.

PEDERSON, C. G.; FRIEDRICH, D. M.; HSIUNG, C.; VON GUNTEN, M.; O'BRIEN, N. A.; RAMAKER, H.-J.; VAN SPRANG, E.; DREISCHOR, M. Pocket-size near-infrared spectrometer for narcotic materials identification. In: PROCEEDINGS OF THE SPIE 2014.

PÉREZ-CISNEROS, E. S.; MENA-ESPINO, X.; RODRÍGUEZ-LÓPEZ, V.; SALES-CRUZ, M.; VIVEROS-GARCÍA, T.; LOBO-OEHMICHEN, R. An integrated reactive distillation

process for biodiesel production. **Computers and Chemical Engineering**, v. 91, p. 233–246, 2016.

PETCHSOONGSAKUL, N.; NGAOSUWAN, K.; KIATKITTIPONG, W.; AIOUACHE, F.; ASSABUMRUNGRAT, S. Process design of biodiesel production: Hybridization of ester- and transesterification in a single reactive distillation. **Energy Conversion and Management**, v. 153, p. 493–503, 2017.

PODDAR, T.; JAGANNATH, A.; ALMANSOORI, A. Use of reactive distillation in biodiesel production: A simulation-based comparison of energy requirements and profitability indicators. **Applied Energy**, v. 185, p. 985–997, 2017.

PUIG-BERTOTTO, J.; COELLO, J.; MASPOCH, S. Evaluation of a handheld near-infrared spectrophotometer for quantitative determination of two APIs in a solid pharmaceutical preparation. **Analytical Methods**, v. 11, p. 327–335, 2019.

REIS, M.; GINS, G. Industrial Process Monitoring in the Big Data/Industry 4.0 Era: From Detection, to Diagnosis, to Prognosis. **Processes**, v. 5, p. 35, 2017.

REYNOLDS, M. R.; LU, C.-W. Control charts for monitoring processes with autocorrelated data. **Nonlinear Analysis, Theory, Methods and Applications**, v. 30, n. 7, p. 4059–4067, 1997.

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.

RICHARD, R.; LI, Y.; DUBREUIL, B.; THIEBAUD-ROUX, S.; PRAT, L. On-line monitoring of the transesterification reaction between triglycerides and ethanol using near infrared spectroscopy combined with gas chromatography. **Bioresource Technology**, v. 102, p. 6702–6709, 2011.

RINNAN, A.; VAN DEN BERG, F.; 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.; DE LA 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.; DE LIMA, S. M.; STRAGEVITCH, L.; PIMENTEL, M. F.; DE JUAN, A. NIR Monitoring and Modelling of Soybean Oil Methanolysis with Multivariate Curve Resolution-Alternating Least Squares with Correlation Constraint. **Journal of the Brazilian Chemical Society**, v. 28, p. 695–706, 2017.

SALLES, R.; BELLOZE, K.; PORTO, F.; GONZALEZ, P. H.; OGASAWARA, E. Nonstationary time series transformation methods: An experimental review. **Knowledge-Based Systems**, v. 164, p. 274–291, 2019.

SANUSI, R. A.; MUKHERJEE, A.; XIE, M. A comparative study of some EWMA schemes for simultaneous monitoring of mean and variance of a Gaussian process. **Computers and Industrial Engineering**, v. 135, p. 426–439, 2019.

SEVERSON, K.; CHAIWATANODOM, P.; BRAATZ, R. D. Perspectives on process monitoring of industrial systems. **Annual Reviews in Control**, v. 42, p. 190–200, 2016.

SHANG, C.; YOU, F. Data Analytics and Machine Learning for Smart Process Manufacturing: Recent Advances and Perspectives in the Big Data Era. **Engineering**, v. 5, p. 1010–1016, 2019.

SHUMWAY, R. H.; STOFFER, D. S. **Time series analysis and its applications with r examples**. 4. ed. New York: Springer, 2017.

SILVA, V. H.; REBOUCAS, M. V; SALLES, A. R.; PIMENTEL, M. F.; PONTES, M. J. C.; PASQUINI, C. Determination of naphtha composition by near infrared spectroscopy and multivariate regression to control steam cracker processes. **Fuel Processing Technology**, v. 131, p. 230–237, 2015.

SIMON, L. L.; PATAKI, H.; MAROSI, G.; MEEMKEN, F.; HUNGERBUHLER, K.; BAIKER, A.; TUMMALA, S.; GLENNON, B.; KUENTZ, M.; STEELE, G.; KRAMER, H. J. M.; RYDZAK, J. W.; CHEN, Z.; MORRIS, J.; KJELL, F.; SINGH, R.; GANI, R.; GERNAEY, K. V; LOUHI-KULTANEN, M.; O'REILLY, J.; SANDLER, N.; ANTIKAINEN, O.; YLIRUUSI, J.; FROHBERG, P.; ULRICH, J.; BRAATZ, R. D.; LEYSSENS, T.; STOSCH, M. Von; OLIVEIRA, R.; TAN, R. B. H.; WU, H.; KHAN, M.; O'GRADY, D.; PANDEY, A.; WESTRA, R.; DELLE-CASE, E.; PAPE, D.; ANGELOSANTE, D.; MARET, Y.; STEIGER, O.; LENNER, M.; ABBOU-OUCHERIF, K.; NAGY, Z. K.; LITSTER, J. D.; KAMARAJU, V. K.; CHIUQ, M.-S. Assessment of Recent Process Analytical Technology (PAT) Trends: A Multiauthor Review. **Organic Process Research & Development**, v. 19, p. 3–62, 2015.

SIMPSON, M. B. Process analytical technology. In: BAKEEV, K. A. (Ed.). **Process analytical technology - spectroscopic tool and implementation strategies for the chemical and pharmaceutical industries**. 2. ed. : Wiley, 2010.

SINGH, D.; SHARMA, D.; SONI, S. L.; SHARMA, S.; KUMARI, 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. **Princípios de análise instrumental**. 6. ed. Porto Alegre: Bookman, 2009.

STAMENKOVIC, O. S.; TODOROVIC, Z. B.; LAZIC, M. L.; SKALA, D. U.; VELJKOVIC, V. B. Kinetics of sunflower oil methanolysis at low temperatures. **Bioresource Technology**, v. 99, p. 1131–1140, 2008.

STRULLER, C. F.; KELLY, P. J.; COPELAND, N. J. Aluminum oxide barrier coatings on polymer films for food packaging applications. **Surface and coatings technology**, v. 241, p. 130–137, 2014.

SZYMANSKA, E. Modern data science for analytical chemical data – a comprehensive

review. **Analytica Chimica Acta**, v. 1028, p. 1–10, 2018.

THAKUR, G.; HEBBI, V.; RATHORE, A. S. An NIR-based PAT approach for real-time control of loading in Protein A chromatography in continuous manufacturing of monoclonal antibodies. **Biotechnology and Bioengineering**, v. 117, n. 3, p. 673–686, 2020.

TIDIRI, K.; CHATTI, N.; VERRON, S.; TIPLICA, T. Bridging data-driven and model-based approaches for process fault diagnosis and health monitoring: A review of researches and future challenges. **Annual Reviews in Control**, v. 42, p. 63–81, 2016.

TREVISAN, M. G.; GARCIA, C. M.; SCHUCHARDT, U.; POPPI, R. J. Evolving factor analysis-based method for correcting monitoring delay in different batch runs for use with PLS: On-line monitoring of a transesterification reaction by ATR-FTIR. **Talanta**, v. 74, p. 971–976, 2008.

VIAVI. **MicroNIR Spectrometers**. 2020. Disponível em: <https://www.viavisolutions.com/en-us/osp/products/micronir-spectrometers>. Acesso em: 11 mar. 2020.

VIGNADUZZO, S. E.; MAGGIO, R. M.; OLIVIERI, A. C. Why should the pharmaceutical industry claim for the implementation of second-order chemometric models - a critical review. **Journal of Pharmaceutical and Biomedical Analysis**, v. 179, p. 112965–112981, 2020.

VITALE, R.; NOORD, O. E. De; FERRER, A. A kernel-based approach for fault diagnosis in batch processes. **Journal of Chemometrics**, v. 28, p. S697–S707, 2014.

WISE, B. M.; KOWALSKI, B. R. Process chemometrics. In: MCLENNAN, F.; KOWALSKI, B. (Eds.). **Process analytical chemistry**. 1. ed. : Springer Science+Business Media Dordrecht, 1995.

WOLD, S.; ANTTI, H.; LINDGREN, F.; OHMAN, J. Orthogonal signal correction of near-infrared spectra. **Chemometrics and Intelligent Laboratory Systems**, v. 44, p. 175–185, 1998.

WORKMAN, J.; WEYER, L. **Practical Guide and Spectral Atlas for Interpretive Near-Infrared Spectroscopy**. 2nd. ed. Boca Raton. 2012.

YANG, H.; HUANG, M.; LAI, C.; JIN, J. An approach combining data mining and control charts-based model for fault detection in wind turbines. **Renewable Energy**, v. 115, p. 808–816, 2018.

ZWETSLOOT, I. M.; AJADI, J. O. A comparison of EWMA control charts for dispersion based on estimated parameters. **Computers & Industrial Engineering**, v. 127, p. 436–450, 2019.

## APÊNDICE A – Artigo publicado em revista científica

Fuel 254 (2019) 115680



Contents lists available at ScienceDirect

Fuel  
journal homepage: [www.elsevier.com/locate/fuel](http://www.elsevier.com/locate/fuel)



### Full Length Article

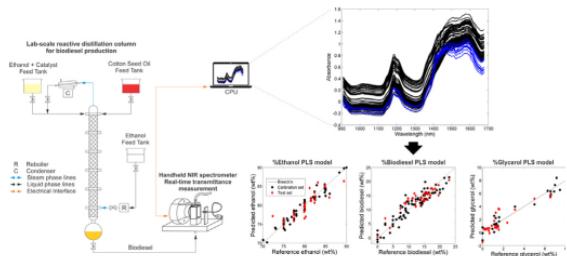
## Handheld near-infrared spectrometer for on-line monitoring of biodiesel production in a continuous process



Rafaella Sales, Neirivaldo C. da Silva, Josivan P. da Silva, Higor H. França,  
Maria Fernanda Pimentel\*, Luiz Stragevitch

Departamento de Engenharia Química, Universidade Federal de Pernambuco, Recife, PE 50.740-521, Brazil

### GRAPHICAL ABSTRACT



### ARTICLE INFO

**Keywords:**  
Near infrared  
Handheld spectrometer  
Process monitoring  
Biodiesel production  
Reactive distillation

### ABSTRACT

In this work, the use of a handheld near-infrared spectrometer (MicroNIR) for on-line monitoring of biodiesel production in a continuous process is described. The process consists of a laboratory scale reactive distillation column where transesterification reactions between cottonseed oil and ethanol take place. This column can operate with different ethanol to oil molar ratios. In the present study, a severe condition using an excess of ethanol at the point monitored with NIRS was applied. This condition caused problems in modelling the content of the alkyl ester, particularly because of the broad and intense band related to the ethanol in the spectral region covered by MicroNIR. Therefore, for process monitoring and control, both an estimate of the alkyl ester as well as of ethanol contents were carried out. To deal with this, PLS calibration models were developed to estimate concentrations of ethanol and alkyl ester at the bottom of the column. In addition, in order to estimate the yield of the alkyl ester, a PLS model to assess glycerol content was also developed. In spite of the problematic conditions under which the column operated, mainly related to the excess of ethanol, the models presented satisfactory results for ethanol (RMSEP = 1.66 wt%), alkyl ester (RMSEP = 2.15 wt%) and glycerol (RMSEP = 1.08 wt%) quantifications and have demonstrated the feasibility of the handheld instrument to monitor biodiesel production in a continuous process.

### 1. Introduction

Implementation of process monitoring and control methods in

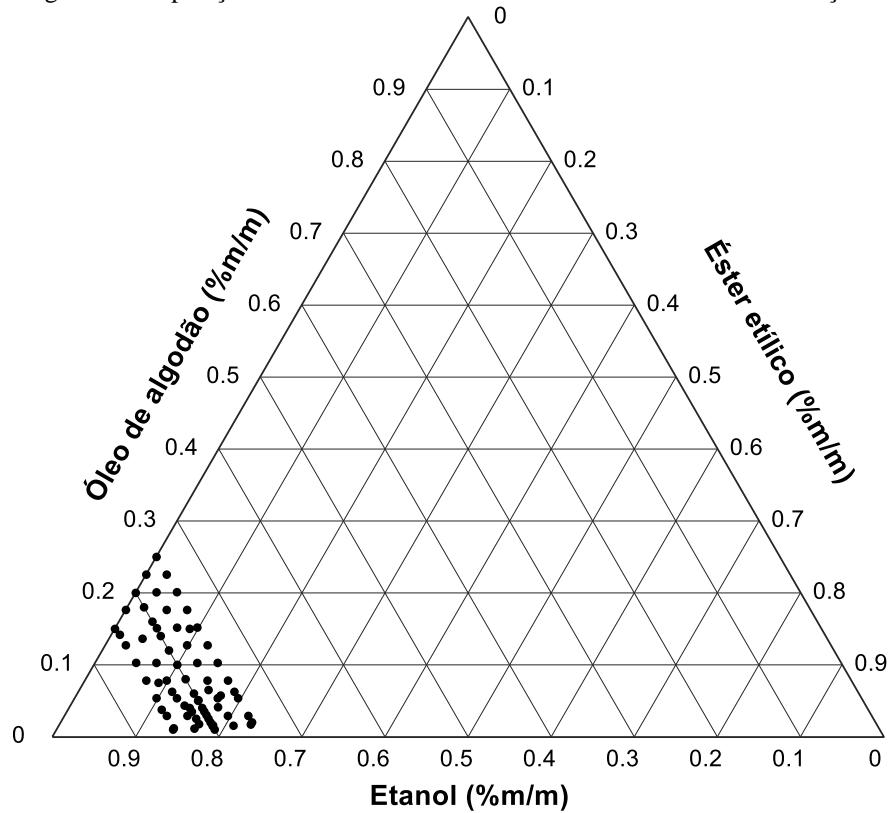
industrial manufacturing is of the utmost importance to reduce process variability, ensuring that product specifications and quality requirements are met. Given the difficulty of handling complex and

\* Corresponding author.

E-mail address: [mfp@ufpe.br](mailto:mfp@ufpe.br) (M.F. Pimentel).

**APÊNDICE B – Diagrama ternário proveniente do planejamento de misturas com pontos adicionais**

Figura – Composição das misturas sintéticas usadas nos modelos de calibração.



Fonte: A Autora (2020).

**APÊNDICE C – Misturas sintéticas preparadas para o desenvolvimento dos modelos de calibração**

Tabela – Frações mássicas de óleo, etanol, éster alquílico e glicerol para cada mistura sintética e a respectiva temperatura de análise com o MicroNIR

(continua)

Mistura sintética	Temperatura de análise (°C)	Teor de óleo (%) (m/m)	Teor de etanol (%) (m/m))	Teor de éster alquílico (%) (m/m))	Teor de glicerol (%) (m/m))
1	25	6,25	82,49	11,25	0,00
2	25	6,25	75,00	18,75	0,00
3	25	3,75	85,00	11,25	0,00
4	25	4,11	78,04	17,85	0,00
5	25	5,73	76,91	17,36	0,00
6	25	4,33	81,96	13,71	0,00
7	25	3,48	81,51	15,01	0,00
8	25	1,76	81,50	16,74	0,00
9	25	4,00	81,50	14,50	0,00
10	25	2,50	81,48	16,02	0,00
11	25	5,00	80,00	15,01	0,00
12	25	4,00	80,00	16,00	0,00
13	25	3,50	80,00	16,50	0,00
14	25	3,00	79,99	17,01	0,00
15	25	2,50	79,99	17,50	0,00
16	25	1,99	80,01	18,00	0,00
17	25	20,00	80,00	0,00	0,00
18	25	18,00	80,01	2,00	0,00
19	25	16,00	79,99	4,01	0,00
20	25	14,00	80,00	6,01	0,00
21	25	12,00	80,01	8,00	0,00
22	25	10,00	80,00	10,00	0,00
23	25	8,00	80,00	12,00	0,00
24	25	6,00	79,99	14,00	0,00
25	25	1,50	80,00	18,50	0,00
26	25	1,00	80,00	19,00	0,00
27	25	5,35	82,37	12,28	0,00
28	25	15,17	77,44	7,39	0,00
29	25	17,63	82,37	0,00	0,00
30	25	2,90	82,33	14,77	0,00
31	25	20,08	77,46	2,46	0,00
32	25	5,35	77,45	17,19	0,00
33	25	7,81	82,37	9,82	0,00
34	25	15,09	79,91	5,00	0,00
35	25	17,63	77,46	4,91	0,00
36	25	7,81	77,46	14,73	0,00
37	25	2,90	77,46	19,64	0,00

**Tabela** – Frações mássicas de óleo, etanol, éster alquílico e glicerol para cada mistura sintética e a respectiva temperatura de análise com o MicroNIR.

(continuação)

Mistura sintética	Temperatura de análise (°C)	Teor de óleo (%) (m/m))	Teor de etanol (%) (m/m))	Teor de éster alquílico (%) (m/m))	Teor de glicerol (%) (m/m))
38	25	5,09	79,91	15,00	0,00
39	25	12,72	77,46	9,82	0,00
40	25	10,26	77,46	12,28	0,00
41	25	10,26	82,37	7,36	0,00
42	25	13,63	82,37	4,00	0,00
43	25	1,54	77,47	21,00	0,00
44	25	22,54	77,46	0,00	0,00
45	25	5,36	84,82	9,82	0,00
46	25	10,27	75,00	14,73	0,00
47	25	20,10	74,99	4,91	0,00
48	25	25,00	75,00	0,00	0,00
49	25	12,72	74,99	12,29	0,00
50	25	2,90	84,81	12,29	0,00
51	25	15,00	85,00	0,00	0,00
52	25	5,36	75,00	19,64	0,00
53	25	7,81	74,99	17,20	0,00
54	25	6,50	77,99	15,51	0,00
55	25	22,54	75,00	2,46	0,00
56	25	12,72	84,83	2,45	0,00
57	25	10,27	84,81	4,91	0,00
58	25	15,18	75,00	9,82	0,00
59	25	7,81	84,83	7,37	0,00
60	25	17,63	75,00	7,37	0,00
61	25	7,50	83,51	8,99	0,00
62	25	1,18	84,83	13,99	0,00
63	25	14,18	84,82	1,00	0,00
64	25	1,00	85,01	14,00	0,00
65	25	2,90	75,00	22,10	0,00
66	25	2,00	75,00	23,00	0,00
67	25	1,70	75,30	23,00	0,00
68	25	1,59	79,90	18,51	0,00
69	25	1,13	82,37	16,50	0,00
70	25	15,00	76,00	9,00	0,00
71	25	7,26	78,92	6,85	6,97
72	25	21,06	70,08	2,30	6,55
73	25	1,87	70,08	21,49	6,56
74	25	9,59	70,07	13,76	6,58
75	25	7,30	70,07	16,07	6,56
76	25	14,19	70,09	9,18	6,54
77	25	1,05	76,50	15,32	7,13
78	25	7,01	78,04	8,40	6,55

**Tabela** – Frações mássicas de óleo, etanol, éster alquílico e glicerol para cada mistura sintética e a respectiva temperatura de análise com o MicroNIR.

(continuação)

Mistura sintética	Temperatura de análise (°C)	Teor de óleo (%) (m/m))	Teor de etanol (%) (m/m))	Teor de éster alquílico (%) (m/m))	Teor de glicerol (%) (m/m))
79	25	14,02	71,04	8,41	6,54
80	25	13,25	79,23	0,93	6,59
81	25	2,50	82,47	15,02	0,00
82	50-60	2,50	82,47	15,02	0,00
83	25	5,38	82,49	11,01	1,12
84	50-60	5,38	82,49	11,01	1,12
85	25	5,38	82,50	9,00	3,12
86	50-60	5,38	82,50	9,00	3,12
87	25	4,00	80,01	16,00	0,00
88	50-60	4,00	80,01	16,00	0,00
89	25	3,47	79,95	14,99	1,59
90	50-60	3,47	79,95	14,99	1,59
91	25	3,37	80,00	15,50	1,12
92	50-60	3,37	80,00	15,50	1,12
93	25	2,50	77,49	20,01	0,00
94	50-60	2,50	77,49	20,01	0,00
95	25	8,17	77,50	13,01	1,32
96	50-60	8,17	77,50	13,01	1,32
97	25	3,50	77,48	14,02	5,00
98	50-60	3,50	77,48	14,02	5,00
99	25	1,75	89,98	6,97	1,30
100	25	0,73	95,23	2,78	1,27
101	25	1,40	89,38	7,78	1,44
102	25	0,64	94,55	3,55	1,26
103	25	1,09	89,81	8,93	0,18
104	25	0,58	94,57	4,75	0,10
105	25	0,00	100,00	0,00	0,00
106	25	2,15	75,80	15,16	6,89
107	25	5,17	73,14	14,24	7,45
108	25	6,74	72,99	14,00	6,27

## APÊNDICE D – Descrição dos cálculos dos valores de referência para o éster alquílico e o glicerol

A descrição dos cálculos dos teores de biodiesel ( $f_{\text{Bio}}$ ) e de glicerol ( $f_{\text{Gli}}$ ), em %(m/m), utilizados como valores de referência para as amostras da coluna, está apresentada a seguir.

Os teores de etanol das amostras da coluna ( $f_{\text{EtOH}}$ ), em %(m/m), utilizados como valores de referência para a construção do modelo, foram obtidos diretamente por cromatografia gasosa. Dessa forma, considerando uma amostra de massa 100 g, tem-se que a massa de etanol da amostra ( $m_{\text{EtOH}}$ ) é dada por  $m_{\text{EtOH}} = \frac{f_{\text{EtOH}} \cdot 100}{100} = f_{\text{EtOH}}$ . Assim, a massa restante da amostra ( $100 - f_{\text{EtOH}}$ ) foi atribuída ao glicerol, ao biodiesel e ao óleo não reagido.

Considerando apenas essa quantidade restante da amostra ( $100 - f_{\text{EtOH}}$ ), pode-se considerar uma massa  $m$  correspondente à mistura de biodiesel e óleo não reagido. Sabendo-se a pureza do biodiesel ( $P_{\text{Bio}}$ ), obtida por cromatografia gasosa após purificação da amostra, pode-se calcular a massa de biodiesel ( $m_{\text{Bio}}$ ) e de óleo ( $m_{\text{Óleo}}$ ) através das Equação 3-1 e Equação 3-2, respectivamente:

$$m_{\text{Bio}} = m \cdot P_{\text{Bio}} \quad (3-1)$$

$$m_{\text{Óleo}} = m \cdot (1 - P_{\text{Bio}}) \quad (3-2)$$

A massa de glicerol ( $m_{\text{Gli}}$ ) pode ser calculada com base na relação estequiométrica entre o biodiesel e o glicerol (3 mols de biodiesel são formados para cada mol de glicerol), a partir da Equação 3-3:

$$m_{\text{Gli}} = m \cdot P_{\text{Bio}} \left( \frac{MM_{\text{Gli}}}{3MM_{\text{Bio}}} \right) \quad (3-3)$$

Em que  $MM_{\text{Gli}}$  é a massa molar do glicerol e  $MM_{\text{Bio}}$  é a massa molar do biodiesel.

Dessa forma, considerando a mistura de biodiesel, glicerol e óleo, o teor de biodiesel  $f'_{\text{Bio}}$ , em m/m, pode ser calculado a partir da Equação 3-4:

$$f'_{\text{Bio}} = \frac{m_{\text{Bio}}}{m_{\text{Bio}} + m_{\text{Óleo}} + m_{\text{Gli}}} \cdot 100\% \quad (3-4)$$

Substituindo as Equações 3-1, 3-2 e 3-3 na Equação 3-4, tem-se a Equação 3-5:

$$f_{\text{Bio}'} = \frac{m \cdot P_{\text{Bio}}}{m \cdot P_{\text{Bio}} + m \cdot (1 - P_{\text{Bio}}) + m \cdot P_{\text{Bio}} \left( \frac{MM_{\text{Gli}}}{3MM_{\text{Bio}}} \right)} \cdot 100\% = \frac{P_{\text{Bio}}}{1 + P_{\text{Bio}} \left( \frac{MM_{\text{Gli}}}{3MM_{\text{Bio}}} \right)} \cdot 100\% \quad (3-5)$$

Esse teor ( $f_{\text{Bio}'}^{'}$ ), no entanto, considera apenas o biodiesel, o óleo e o glicerol. Para encontrar o teor de biodiesel na mistura total ( $f_{\text{Bio}}$ ), basta multiplicar  $f_{\text{Bio}'}^{'}$  pela fração mássica da mistura ternária na mistura total ( $\frac{100 - f_{\text{EtOH}}}{100}$ ), que inclui o etanol, obtendo-se, assim, a Equação 3-6:

$$\frac{f_{\text{Bio}}}{f_{\text{Bio}'}^{'}} = \frac{\frac{m_{\text{Bio}}}{100}}{\frac{m_{\text{Bio}}}{100 - f_{\text{EtOH}}}} \Rightarrow f_{\text{Bio}} = \frac{100 - f_{\text{EtOH}}}{100} f_{\text{Bio}'}^{'}$$

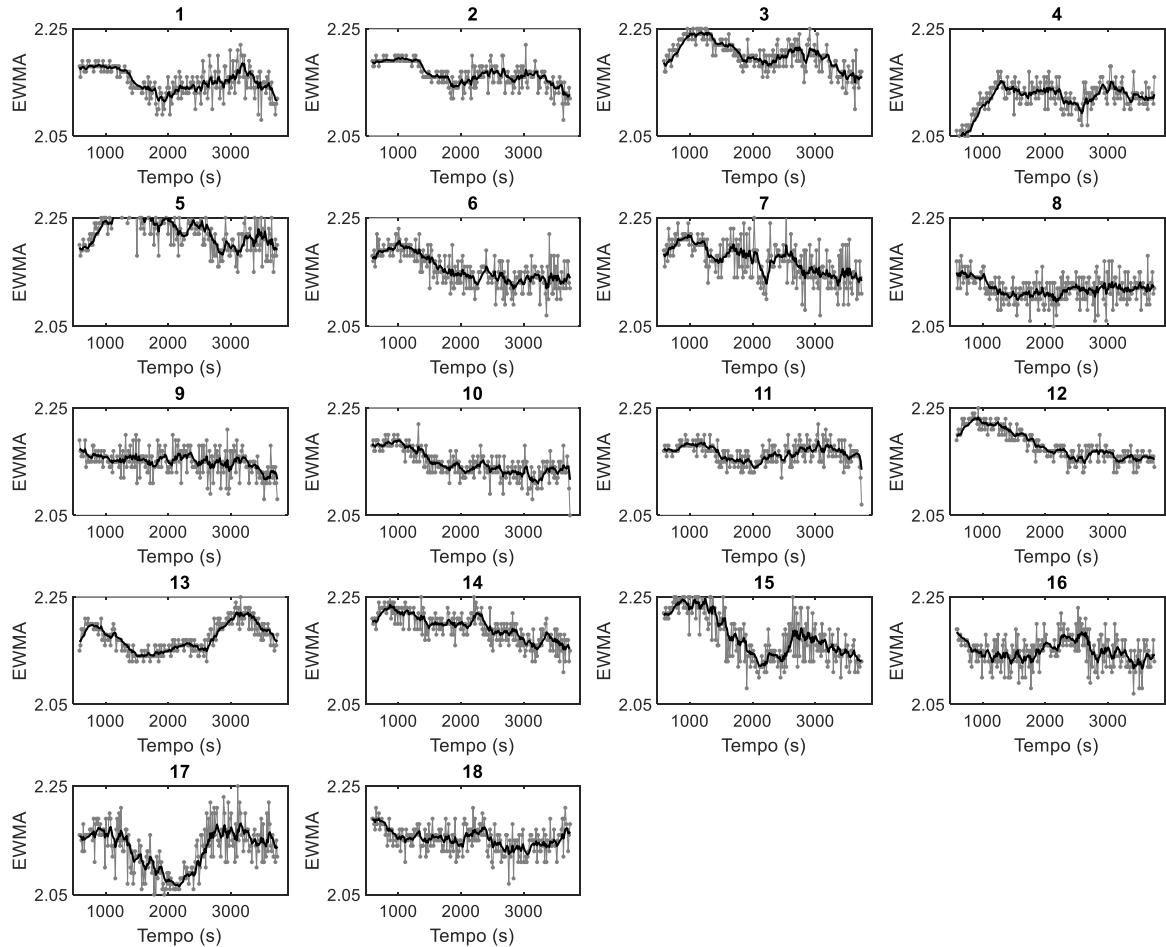
$$f_{\text{Bio}} = \frac{P_{\text{Bio}}}{\left( 1 + P_{\text{Bio}} \left( \frac{MM_{\text{Gli}}}{3MM_{\text{Bio}}} \right) \right)} \left( \frac{100 - f_{\text{EtOH}}}{100} \right) \cdot 100\% \quad (3-6)$$

A partir da relação estequiométrica entre biodiesel e glicerol, pode-se obter o teor de glicerol na mistura total ( $f_{\text{Gli}}$ ), a partir da Equação 3-7:

$$f_{\text{Gli}} = f_{\text{Bio}} \left( \frac{MM_{\text{Gli}}}{3MM_{\text{Bio}}} \right) \quad (3-7)$$

**APÊNDICE E – Monitoramento da tendência dos dados de densidade óptica para a corrida A2**

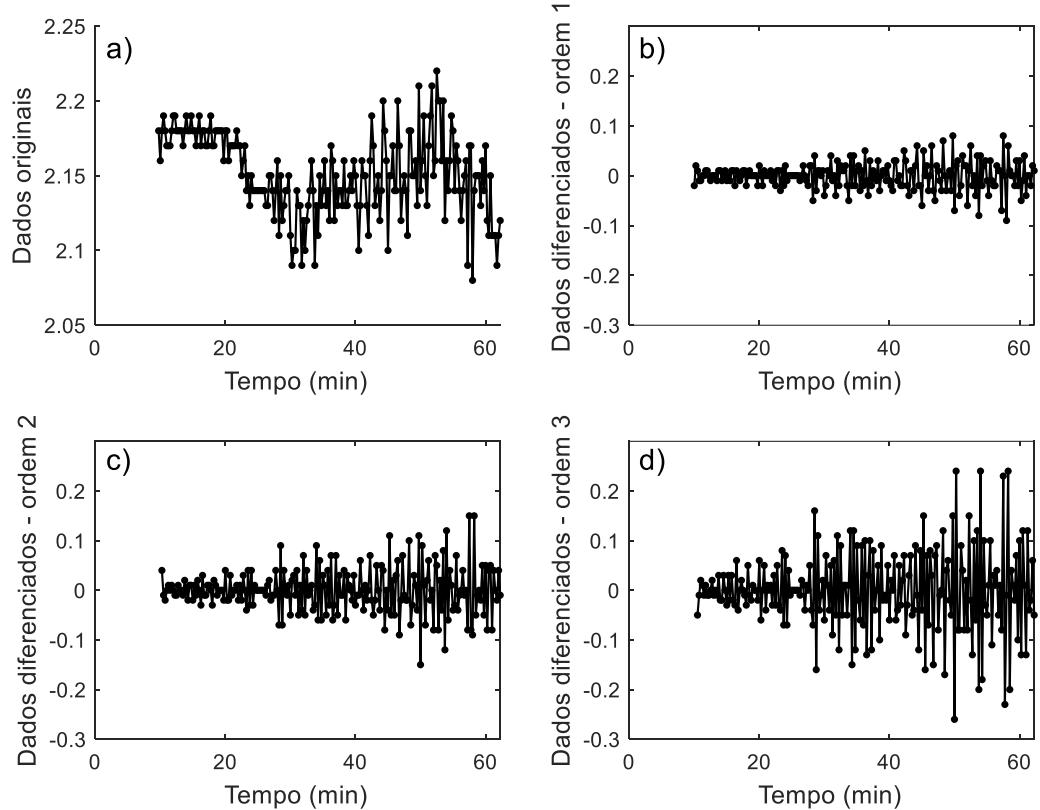
Figura – Cartas EWMA para monitoramento da tendência para as 18 variáveis da corrida A2



Fonte: A Autora (2020).

## APÊNDICE F – Comparação entre os dados originais e diferenciados

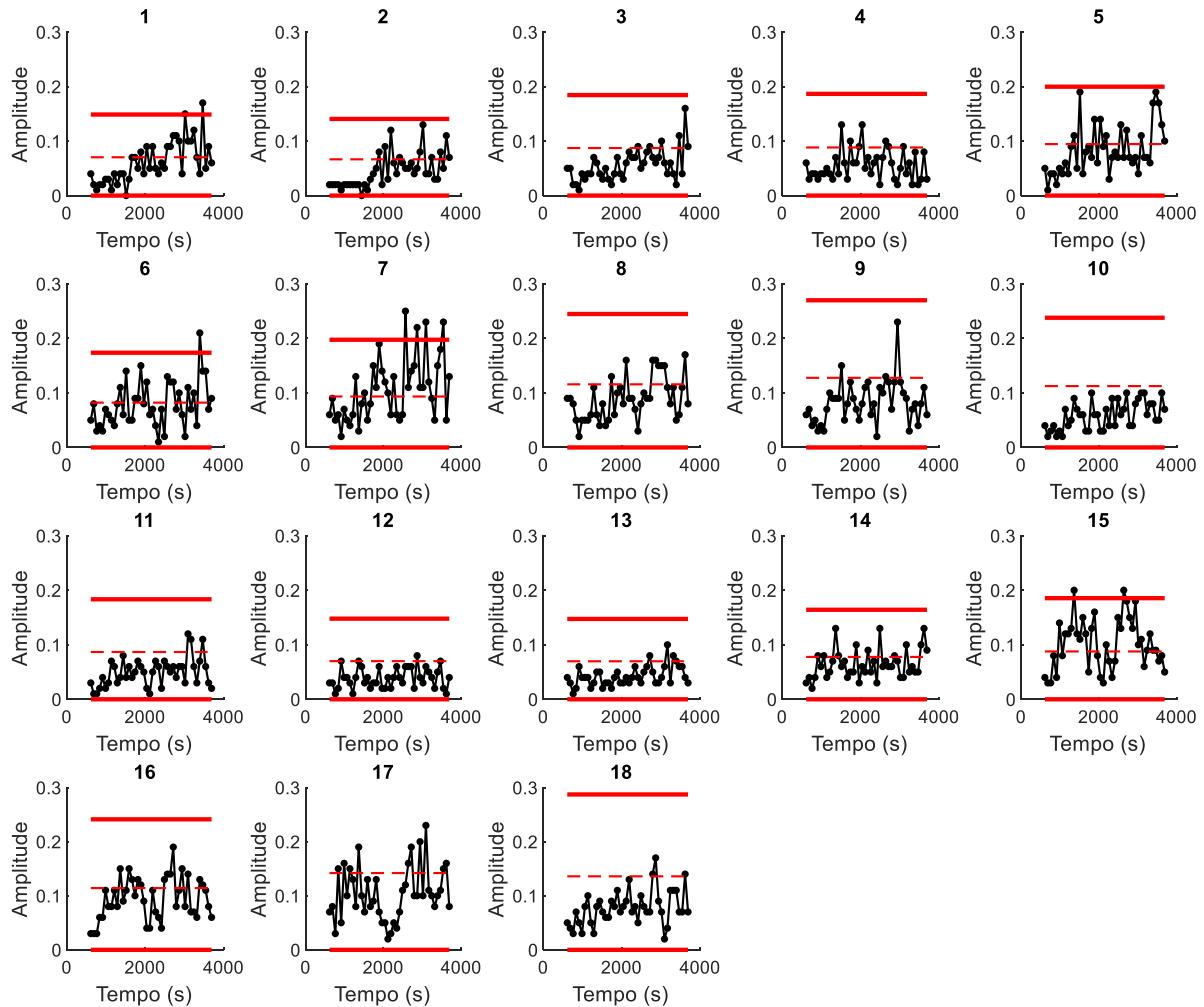
Figura – Dados originais (a) e dados diferenciados com 1<sup>a</sup> (b), 2<sup>a</sup> (c) e 3<sup>a</sup> (d) ordens para a variável 1 da corrida A2.



Fonte: A Autora (2020).

## APÊNDICE G – Monitoramento da variabilidade para a corrida A2

**Figura** – Cartas da amplitude para as 18 variáveis da corrida A2.



Fonte: A Autora (2020).