



**UnB**

Instituto de Química

Programa de Pós-Graduação em Química

Tese de Doutorado

Avaliação e comparação de ensaios de estabilidade de óleo diesel utilizando FTIR, GC-MS e ferramentas quimiométricas

Vinicius Leandro Skrobot

Prof. Dr. Jez Willian Batista Braga

Brasília, DF

2020



**UnB**

Instituto de Química

Programa de Pós-Graduação em Química

Tese de Doutorado

Avaliação e comparação de ensaios de estabilidade de óleo diesel utilizando FTIR, GC-MS e ferramentas quimiométricas

Tese apresentada no Programa de Pós-Graduação em Química do Instituto de Química da Universidade de Brasília como parte dos requisitos para obtenção do título de Doutor em Química.

Vinicius Leandro Skrobot

Orientador: Prof. Dr. Jez Willian Batista Braga

Brasília, DF

2020

*Dedico este trabalho aos meus pais, Antônio Vitor Skrobot e Denilza Pires Skrobot por terem se esforçado tanto para me educar em todos os sentidos e que me fazem sentir orgulho por minhas conquistas. Aos meus irmãos, Victor Hugo Skrobot e Helton Valmor Skrobot, pelo amor e companheirismo por todos esses anos. À minha esposa, Débora Cristina Araújo Miguel, pelo amor, apoio, paciência e por me estimular a persistir neste projeto. Por fim, dedico este trabalho aos meus filhos, Elisa Araújo Skrobot e Lucas Araújo Skrobot, que, de diferentes formas, me estimulam a tentar ser alguém melhor.*

## Agradecimentos

- Ao Prof. Dr. Jez Willian Batista Braga, por ser um profissional exemplar, pela confiança depositada em mim, pelas orientações pacientes e pela amizade.
- À minha esposa, pelo apoio, pela paciência e por me estimular a buscar o sonho de concluir o doutorado.
- Ao Caio de Souza Santos da Unb, pela realização de ensaios ao longo de um ano de projeto de iniciação científica.
- Ao Luiz Filipe Brandão Filho da ANP, por compartilhar sua experiência em cromatografia gasosa e espectrometria de massas.
- Ao Instituto de Química da Unb, pela oportunidade de participar do Programa de Pós-graduação em Química.
- A todos os membros da equipe do Centro de Pesquisas e Análises Tecnológicas da ANP que me auxiliaram na realização de ensaios, troca de experiências ou auxílio no uso de equipamentos analíticos. Em especial, agradeço ao Coordenador do CPT, Fábio da Silva Vinhado por todo o apoio.
- À Agência Nacional do Petróleo, Gás Natural e Biocombustíveis, pela oportunidade de realizar este projeto e pelos conhecimentos adquiridos na área de combustíveis e química analítica.
- À equipe do Grupo de Automação, Quimiometria e Química Ambiental do IQ pela recepção e pelo aprendizado.
- Ao prof. Eustáquio Vinicius Ribeiro de Castro da UFES por ter me apresentado ao mundo da quimiometria durante o mestrado.

## SUMÁRIO

1. INTRODUÇÃO E OBJETIVOS .....	1
2. FUNDAMENTAÇÃO TEÓRICA .....	3
2.1. Óleo diesel .....	3
2.1.1. Produção de óleo diesel .....	4
2.2. Biodiesel .....	6
2.2.1. Produção de biodiesel .....	7
2.3. Misturas de óleo diesel e biodiesel .....	8
2.4. Estabilidade do óleo diesel, do biodiesel e de suas misturas .....	9
2.4.1. Ensaios para avaliação de estabilidade de óleo diesel e biodiesel .....	12
2.5. Espectrometria no infravermelho .....	13
2.6. Cromatografia gasosa e espectrometria de massas .....	16
2.7. Ferramentas Quimiométricas .....	18
2.7.1. Ferramentas quimiométricas multidimensionais .....	18
2.7.1.1. Histórico .....	18
2.7.1.2. Conceitos e notação .....	19
2.7.1.3. Análise de Fatores Paralelos (PARAFAC) .....	22
2.7.1.4. Resolução de Curvas Multivariadas (MCR) .....	27
2.7.2. Calibração multivariada .....	34
2.7.2.1. Regressão linear em notação matricial e geométrica .....	34
2.7.2.2. Seleção do conjunto de amostras de calibração .....	36
2.7.2.3. Pré-tratamento de dados .....	37
2.7.2.4. Identificação de amostras anômalas .....	39
2.7.2.5. Métodos de regressão multivariada .....	41
2.7.2.6. Avaliação e otimização do modelo de calibração .....	49
2.7.2.7. Validação e estimativa de figuras de mérito do modelo .....	50

3.	REVISÃO BIBLIOGRÁFICA.....	52
3.1.	Quimiometria e espectrometria no infravermelho .....	52
3.2.	Cromatografia Gasosa.....	55
3.3.	PARAFAC .....	57
3.4.	MCR-ALS.....	58
SEÇÃO 1 – APLICAÇÃO DE FERRAMENTAS MULTIDIMENSIONAIS PARA AVALIAÇÃO DE ENSAIOS DE ESTABILIDADE DE ÓLEO DIESEL .....		60
4.	MATERIAIS E MÉTODOS DA SEÇÃO 1.....	60
4.1.	Obtenção das amostras de óleo diesel.....	60
4.2.	Avaliação prévia das amostras coletadas.....	63
4.3.	Ensaio de Estabilidade .....	64
4.3.1.	Ensaio de estabilidade ao armazenamento a 43 °C (ASTM D4625).....	64
4.3.2.	Ensaio rápido de oxidação em pequena escala (ASTM D7545).....	65
4.3.3.	Ensaio de estabilidade em alta temperatura (ASTM D6468).....	66
4.3.4.	Ensaio de estabilidade à oxidação por oxidação acelerada (EN 15751) ..	67
4.4.	Espectrometria no infravermelho.....	68
4.5.	Cromatografia gasosa acoplada à Espectrometria de massas .....	70
4.6.	Ferramentas quimiométricas.....	72
5.	RESULTADOS E DISCUSSÃO DA SEÇÃO 1.....	72
5.1.	Avaliação preliminar.....	72
5.2.	Ensaio de estabilidade .....	74
5.3.	Espectrometria no infravermelho médio.....	76
5.4.	Aplicação de PARAFAC nos dados de ATR-FTIR .....	82
5.5.	GC-MS.....	94
5.6.	Aplicação de MCR-ALS nos dados de GC-MS .....	101
6.	CONCLUSÕES DA SEÇÃO 1 .....	125

SEÇÃO 2 – USO DE CALIBRAÇÃO MULTIVARIADA PARA PREDIÇÃO DE RESULTADOS DE ENSAIO DE ESTABILIDADE UTILIZANDO ESPECTROMETRIA NO INFRAVERMELHO MÉDIO.....	129
7. MATERIAIS E MÉTODOS DA SEÇÃO 2.....	129
7.1. Obtenção das amostras de óleo diesel.....	129
7.2. Avaliação prévia das amostras de óleo diesel.....	130
7.3. Ensaio de estabilidade oxidativa pelo método da oxidação acelerada.....	130
7.4. Espectrometria no infravermelho médio.....	131
7.5. Calibração multivariada.....	131
8. RESULTADOS E DISCUSSÃO DA SEÇÃO 2.....	131
8.1. Avaliação prévia das amostras.....	131
8.2. Ensaio de estabilidade oxidativa pelo método da oxidação acelerada.....	134
8.3. Espectrometria no infravermelho.....	135
8.4. Calibração multivariada.....	137
8.4.1. Avaliação inicial por Análise por Componentes Principais - PCA.....	137
8.4.2. Regressão por Mínimos Quadrados Parciais - PLS.....	138
8.4.3. Regressão por Vetores de Suporte – SVR.....	141
9. CONCLUSÕES DA SEÇÃO 2.....	146
10. CONCLUSÕES GERAIS.....	148
11. BIBLIOGRAFIA.....	149
APÊNDICE 1 – Diagnóstico de consistência de núcleo.....	1
APÊNDICE 2 – Comparação dos ensaios para avaliação de estabilidade de óleo diesel utilizados neste estudo.....	1
APÊNDICE 3 – Avaliação prévia das amostras de óleo diesel - SEÇÃO 1.....	1
APÊNDICE 4 – Espectros de referência do banco de dados Massbank®.....	1
APÊNDICE 5 – Avaliação prévia das amostras de óleo diesel - SEÇÃO 2.....	1

## Índice de Figuras

Figura 1: Volume de combustíveis líquidos ( $m^3$ ) comercializados no Brasil em 2019 (Fonte: Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – disponível em <a href="http://www.anp.gov.br/dados-estatisticos">http://www.anp.gov.br/dados-estatisticos</a> - acessado em 23/03/2020).....	3
Figura 2: Cadeia de distribuição da mistura óleo diesel/biodiesel no Brasil. [adaptado de Silva, Alexandre, D. Distribuição e comercialização de biodiesel e suas misturas. 6º Congresso Internacional de bioenergia. 2011.] .....	4
Figura 3: Esquema de refino típico para produção de óleo diesel <sup>2</sup> . .....	4
Figura 4: Reação de transesterificação na qual R representa diferentes cadeias de ácidos graxos e R', usualmente, é igual a $CH_3$ <sup>6</sup> .....	8
Figura 5: Esquema típico de obtenção de biodiesel a partir de óleo de soja. O ácido fosfórico é adicionado na segunda etapa para transformar lipídeos não hidrolisáveis em hidrolisáveis. <sup>41</sup> .....	8
Figura 6: Representação esquemática das reações <b>A.</b> de oxidação e auto-oxidação do óleo diesel e <b>B.</b> da formação de peróxidos e hidroperóxidos em biodiesel. <sup>49,52</sup> .....	11
Figura 7: Representação esquemática das principais regiões do espectro eletromagnético. ....	13
Figura 8: Representação de uma célula ATR [adaptado de Catálogo Pike Technologies] <sup>63</sup> .....	14
Figura 9: Representação de um conjunto de dados com 3 dimensões.....	20
Figura 10: Desdobramento de um conjunto de dados de 3 dimensões.....	21
Figura 11: Em PARAFAC, um conjunto de dados de 3 dimensões é decomposto na soma de produtos de 3 vetores. Os vetores de pesos são chamados de A, B e C ou pelo nome da dimensão (e.g.: “pesos da dimensão das amostras”, “pesos da dimensão dos comprimentos de ondas”).[adaptado de Smilde et al. <sup>72</sup> ]. .....	22
Figura 12: Deconvolução de uma matriz <b>D</b> nas matrizes de concentrações <b>C</b> e de espectros <b>S</b> <sup>86</sup> . .....	27
Figura 13: Aplicação de MCR-ALS em conjuntos de dados de três dimensões mostrando o agrupamento de matrizes em colunas (alto) e linhas (abaixo) <sup>86</sup> .....	28
Figura 14: Obtenção de um modelo MCR-ALS <sup>86</sup> .....	29
Figura 15: Representação da ferramenta SIMPLISMA. O gráfico à esquerda representa a composição de 13 misturas dos componentes A, B e C. À direita, o gráfico representa a	

projeção de vetores representando, por exemplo, uma relação m/z de um espectro de massas <sup>92</sup> .....	32
Figura 16: Regressão linear representada como projeção do vetor y no espaço delimitado pela matriz C de duas colunas <sup>77</sup> .....	35
Figura 17: Separação de amostras de duas classes por uma linha. As formas vermelhas são os <i>vetores de suporte</i> conforme explicado no texto.....	44
Figura 18: Projeção de observações no espaço de características por uma função kernel linear e definição do hiperplano de separação das classes. [Extraído de Filgueiras <sup>108</sup> ].	45
Figura 19: Utilização de vetores de suporte para regressão. [Extraído de Filgueiras <sup>108</sup> ]	47
Figura 20: Regressão por vetores de suporte. As figuras em vermelho são os vetores de suporte. As linhas tracejadas correspondem à banda $\epsilon$ . [adaptado de Guenther e Schonlau <sup>107</sup> ].....	48
Figura 21: Locais de coleta das amostras de óleos diesel utilizadas no estudo.....	62
Figura 22: Esquema do frasco e foto da estufa utilizados para realização do ensaio conforme norma ASTM D4625.....	65
Figura 23: Coleta de amostra oxidada após realização de ensaio conforme norma ASTM D7545.....	66
Figura 24: Forno utilizado para realização do ensaio de estabilidade em alta temperatura com tubo de ensaio e amostra.....	67
Figura 25: Esquema e foto do equipamento Rancimat <sup>®</sup> utilizado para realização do ensaio conforme norma CEN EN 15751.....	68
Figura 26: Exemplo de curva de condutividade versus tempo destacando uma inflexão (ou período de indução).....	68
Figura 27: Espectrômetro Bruker <sup>®</sup> Tensor II <sup>®</sup> com célula ATR de ZnSe.....	69
Figura 28: Sistema GC-MS Agilent <sup>®</sup> modelo 6890N/5973N com amostrador automático.....	70
Figura 29: Resultados dos ensaios de rotina das 18 amostras deste estudo (— limite superior; — limite inferior). A amostra do posto 2 apresentou teor de biodiesel acima do limite da especificação ANP, entretanto, abaixo do limite de tolerância adotado no laboratório.....	74
Figura 30: Resultados dos ensaios de estabilidade conforme norma CEN EN15751....	75
Figura 31: Resultados dos ensaios de estabilidade conforme norma ASTM D7545....	76
Figura 32: Correlação entre os resultados dos ensaios conforme normas D7545 e EN15751.....	76

Figura 33: Espectros das 18 amostras (em duplicata) antes dos ensaios de estabilidade. Em <b>A</b> , toda a região estudada e, em <b>B</b> e <b>C</b> , aproximação de regiões com variações comentadas no texto. ....	78
Figura 34: Espectros de uma das amostras antes e após todos os ensaios de estabilidade. ....	79
Figura 35: Comparação da média dos 36 espectros de óleo diesel antes dos ensaios (linha azul) com a média dos 36 espectros após ensaios de degradação (linha vermelha). <b>A</b> : comparação com o ensaio conforme norma ASTM D4625. <b>B</b> : comparação com o ensaio conforme norma ASTM D7545. <b>C</b> : comparação com o ensaio conforme norma ASTM D6468. <b>D</b> : comparação com o ensaio conforme norma CEN EN15751. ....	80
Figura 36: Organização dos dados para aplicação da ferramenta PARAFAC aos dados de ATR-FTIR. ....	83
Figura 37: Modelo PARAFAC com 7 componentes – pesos da dimensão C (ensaios). 85	
Figura 38: Modelo PARAFAC com 7 componentes - pesos das 3 primeiras componentes da dimensão B (espectros). ....	86
Figura 39: Modelo PARAFAC com 7 componentes – pesos da 4ª componente da dimensão dos espectros. ....	88
Figura 40: Modelo PARAFAC com 7 componentes – pesos da 5ª componente da dimensão dos espectros. ....	89
Figura 41: Modelo PARAFAC com 7 componentes – pesos da 6ª componente da dimensão dos espectros. ....	90
Figura 42: Modelo PARAFAC com 7 componentes – pesos da 7ª componente da dimensão dos espectros. ....	91
Figura 43: Pesos das componentes da dimensão das amostras para o modelo PARAFAC com 7 componentes. A numeração das amostras no eixo das abscissas considera duas replicatas por amostra. ....	92
Figura 44: Pesos das amostras na 5ª componente no modelo PARAFAC com 7 componentes. ....	93
Figura 45: Pesos das amostras nas componentes 1 versus 2 no modelo PARAFAC com 7 componentes. ....	94
Figura 46: <b>A</b> . Cromatograma de Íons Totais, TIC, de amostra de óleo diesel antes de qualquer ensaio de degradação. A atribuição dos picos segue o procedimento descrito no subitem 4.5. <b>B</b> . TICs das três amostras (antes dos ensaios) com deslocamento de linha-base para facilitar a comparação. ....	95

Figura 47: Comparação da parte inicial dos TICs da amostra 8 antes e após o ensaio conforme norma D4625.....	98
Figura 48: Alterações observadas na comparação da amostra 8 antes e após o ensaio de estabilidade de acordo com a norma ASTM D7545. Note a maior supressão do sinal para o linoleato de metila em 78,3 min. O mesmo comportamento foi observado nas demais amostras.....	98
Figura 49: Alterações observadas após o ensaio de acordo com a norma D6468 em: <b>A.</b> amostra 8 e, em <b>B.</b> e <b>C.</b> amostra 2. Note a diminuição geral dos picos em <b>A.</b> .....	99
Figura 50: Alterações observadas para o ensaio conforme norma CEN EN15751: Em <b>A.</b> , observa-se a supressão quase completa de compostos voláteis e o pico surgido em 8,7 min. As alterações em <b>C</b> só foram observadas para a amostra 2 e as destacadas em <b>B</b> e <b>D</b> são comuns às três amostras. Os picos destacados são comentados no texto. ....	100
Figura 51: Organização dos dados de GC-MS para aplicação de MCR-ALS. Os números nos retângulos identificam a amostra ensaiada.....	101
Figura 52: Autovalores (em escala logarítmica) obtidos por SVD para o conjunto de dados de GC-MS.....	104
Figura 53: Pesos da primeira componente da matriz <b>C</b> (perfil cromatográfico recuperado). <b>A:</b> todas as amostras. <b>B:</b> Apenas a amostra 2 sem degradação destacando o tempo de retenção que corresponde ao do palmitato de metila.....	105
Figura 54: Perfil espectral recuperado da primeira componente do modelo MCR destacando os picos típicos de ésteres alifáticos. ....	106
Figura 55: Espectro do palmitato de metila obtido no banco de dados Massbank <sup>®155</sup> . 106	
Figura 56: Somatório das áreas do perfil de concentrações da 1 <sup>a</sup> . componente.....	107
Figura 57: Pesos da segunda componente da matriz <b>C</b> (perfil de concentrações). <b>A:</b> todas as amostras. <b>B:</b> Apenas amostra 2 sem degradação. ....	109
Figura 58: Perfil espectral recuperado da segunda componente do modelo MCR. ....	110
Figura 59: Somatório das áreas do perfil de concentrações da 2 <sup>a</sup> . componente.....	111
Figura 60: Perfis de cromatográficos recuperados na segunda componente da amostra 2 sem degradação e após os quatro ensaios de estabilidade. Os perfis foram deslocados verticalmente para facilitar a comparação. ....	112
Figura 61: Pesos da terceira componente da matriz <b>C</b> (perfil de concentrações) da amostra 2 sem degradação.....	113
Figura 62: Pesos da quarta componente da matriz <b>C</b> (perfil de concentrações).....	114
Figura 63: Perfil espectral recuperado da quarta componente do modelo MCR.....	114

Figura 64: Somatório das áreas do perfil de concentrações da 4 <sup>a</sup> . componente.....	115
Figura 65: Perfil de concentrações recuperado na 5 <sup>a</sup> . componente do modelo MCR. .	116
Figura 66: Espectro recuperado pela 5 <sup>a</sup> . componente do modelo MCR.....	116
Figura 67: Somatório das áreas do perfil de concentrações da 5 <sup>a</sup> . componente.....	117
Figura 68: Pesos da sexta componente da matriz <i>C</i> (perfil de concentrações).....	118
Figura 69: Perfil espectral recuperado da sexta componente do modelo MCR. ....	119
Figura 70: Somatório das áreas do perfil de concentrações da 6 <sup>a</sup> . componente.....	119
Figura 71: Pesos da décima componente da matriz <i>C</i> (perfil de concentrações).....	120
Figura 72: Perfil espectral recuperado da décima componente do modelo MCR. ....	121
Figura 73: Somatório das áreas do perfil de concentrações da 10 <sup>a</sup> . componente.....	122
Figura 74: Resultados da avaliação prévia das 50 amostras discutidas na Seção 2. A variabilidade dos resultados de Teor de biodiesel é discutida no texto. O resultado de Teor de enxofre para uma das amostras não estava disponível, assim como um dos resultados de Teor de Biodiesel. ....	132
Figura 75: Resultado dos ensaios de Estabilidade à Oxidação conforme norma EN 15751 em 50 amostras de óleo diesel. ....	134
Figura 76: Histograma dos resultados de estabilidade conforme norma EN 15751 em 50 amostras de óleo diesel. ....	135
Figura 77: Espectros médios (triplicata) das 50 amostras de óleo diesel utilizadas nesta seção. ....	136
Figura 78: Três espectros de uma mesma amostra com aproximação de região correspondente ao estiramento C-H. ....	137
Figura 79: Escores das amostras nas duas primeiras componentes de PCA. A linha tracejada corresponde a um nível de confiança de 95%. ....	138
Figura 80: RMSEC e RMSECV para modelo PLS com diferentes números de variáveis latentes (com 49 amostras). ....	139
Figura 81: Valores estimados pelo modelo PLS versus erros de predição (validação cruzada) para modelo PLS com 49 amostras e 6 variáveis latentes. ....	139
Figura 82: RMSEC e RMSECV para modelo PLS com diferentes números de variáveis latentes (com 48 amostras). ....	140
Figura 83: Valores de Hotelling $T^2$ e Parâmetro <i>Q</i> para amostras do modelo PLS com 2 variáveis latentes após exclusão de 2 amostras anômalas. ....	141
Figura 84: Otimização dos parâmetros de $\nu$ -SVR por Grade de Pesquisa. O X indica a combinação de menor erro quadrático médio.....	142

Figura 85: Valores estimados pelo modelo $\nu$ -SVR versus erro de predição de validação cruzada.....	144
Figura 86: Espectros de massas de referência para <b>A</b> : Linoleato de metila, <b>B</b> : Oleato de Metila, <b>C</b> : Estearato de metila. ....	1

## Índice de Tabelas

Tabela 1: Composição em termos de ácidos graxos (% m/m) de algumas matérias-primas para biodiesel <sup>6,38</sup> . O código sob o nome do ácido graxo corresponde ao número de átomos de carbono na cadeia alquílica e o número de insaturações na mesma. ....	7
Tabela 2: Evolução do teor de biodiesel para uso rodoviário no Brasil. ....	9
Tabela 3: Absorções importantes no infravermelho médio características de substâncias majoritárias no óleo diesel, no biodiesel e nos seus produtos de degradação. <sup>62,65</sup> .....	15
Tabela 4: Locais e datas de coletas das amostras de óleo diesel utilizadas neste estudo. A Bandeira se refere ao nome da distribuidora ostentado no posto de revenda. Postos de “Bandeira Branca” podem revender combustíveis de qualquer distribuidora.....	62
Tabela 5: Resultados dos ensaios de estabilidade pelos métodos CEN EN15751 e ASTM D7545. ....	75
Tabela 6: Modelos PARAFAC com diferentes números de componentes. ....	83
Tabela 7: Variância capturada por componente do modelo PARAFAC.....	84
Tabela 8: Parâmetros de modelos MCR-ALS obtidos com diferentes números de componentes. ....	103
Tabela 9: ANOVA de um fator das áreas recuperadas para a primeira componente na matriz C. ....	108
Tabela 10: Avaliação das componentes cuja interpretação é subjetiva.....	123
Tabela 11: Valores de RMSECV para os modelos de calibração multivariados com diferentes pré-tratamentos. Os parâmetros dos modelos estão descritos em 8.4.2 e 8.4.3. ....	145

## Lista de abreviaturas e acrônimos

- ABNT - Associação Brasileira de Normas Técnicas
- ALS - Alternating Least Squares ou Mínimos Quadrados Alternantes
- ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis.
- ATR - Attenuated Total Reflectance ou Reflectância Total Atenuada.
- BXX - Mistura de óleo diesel e biodiesel no qual o biocombustível está a XX % v/v.
- CEN - Committee European Normalization ou Comitê Europeu de Normalização.
- CNPE - Conselho Nacional de Política Energética.
- CORCONDIA - Core Consistency Diagnostic ou Diagnóstico de Consistência de Núcleo ou Consistência trilinear.
- DLaTGS - Deuterated Lanthanum  $\alpha$  Alanine doped TriGlycine Sulphate ou Sulfato de triglicina dopado com  $\alpha$  alanina lantânio deuterado
- DTLD - Direct Trilinear Decomposition ou Decomposição Trilinear Direta.
- EFA - Evolving Factor Analysis ou Análise Evolucionária de Fatores.
- EI - Electron Ionization ou Ionização eletrônica.
- FID - Fire Ionization Detector ou Detector por ionização de chama.
- FNNLS - Fast non-negativity-constrained linear least squares ou Restrição rápida de não negatividade por mínimos quadrados.
- FTIR - Fourier Transform Infrared ou Infravermelho com Transformada de Fourier.
- GC-GC - na prática: Comprehensive Chromatography ou Cromatografia abrangente.
- GC-MS – Gas Chromatography – Mass spectrometry ou Cromatografia Gasosa acoplada à Espectrometria de Massas.
- GRAM -Generalized Rank Annihilation ou Método Generalizado de Aniquilação de Posto)
- HCC - Hidrocraqueamento.
- HPLC - High Performance Liquid Chromatography ou Cromatografia líquida de alto desempenho.
- m/z - Razão massa/carga
- MCR-ALS - Multivariate Curve Resolution - Alternating Least Squares ou Resolução de Curvas Multivariadas - Mínimos Quadrados Alternantes
- MIR - Mid-infrared ou Infravermelho Médio
- MSTFA - n-metiltrimetilsililfluoroacetamida.
- NNLS - Non-negativity-constrained linear least squares ou Restrição de não negatividade por mínimos quadrados
- PARAFAC - Parallel Factor Analysis ou Análise de Fatores Paralelos.
- PCA - Principal Component Analysis ou Análise por Componentes Principais.
- PFTBA – Perfluorotributíla
- PMQC - Programa de Monitoramento da Qualidade de Combustíveis
- PNPB - Programa Nacional de Produção e Uso do Biodiesel.
- RBF - Radial Basis Function ou Função de base radial.
- REMSEC - Root Mean Squared Error of Calibration ou Raiz da média quadrática dos erros de calibração.
- RMN - Ressonância Magnética Nuclear

RMSECV - Root Mean Squared Error of Cross Validation ou Raiz da média quadrática dos erros de validação cruzada.

RMSEP - Root Mean Squared Error of Prediction ou Raiz da média quadrática de predição.

SEV - Standard Error of Validation ou Erro padrão de validação.

SIMPLISMA - Simple-to-use interactive self-modeling mixture analysis ou Resolução de curvas por auto-modelagem interativa.

SNV - Standard Normal Variate ou Padronização normal de sinal.

SQE - Soma dos Quadrados dos Erros

SVD - Singular Value Decomposition ou Decomposição em Valores Singulares.

SVM - Support Vector Machine ou Máquina de vetores de suporte.

SVR - Support Vector Regression ou Regressão por vetores de suporte

TIC - Total Ion Chromatogram ou Cromatograma de Íons Totais

$t_R$  - Tempo de retenção

TRR - Transportador Revendedor Retalhista.

u.m.a. - unidades de massa atômica.

VIP - Variable importance in projection ou Importância de variável na projeção.

WCOT - Wall-Coated Open Tubular ou Tubo (cromatográfico) aberto com parede revestida.

## Resumo

A estabilidade ao armazenamento do óleo diesel automotivo é um tema de grande interesse econômico, pois sua degradação pode gerar produtos que podem danificar motores e deteriorar a qualidade das emissões gasosas. A mistura de biodiesel ao óleo diesel trouxe novos desafios. Os métodos de avaliação da estabilidade do óleo diesel se baseiam em diferentes formas de estressá-lo e mensurar sua degradação, sendo de grande importância o entendimento dos fenômenos envolvidos nesses testes e a correlação dos resultados. Em um primeiro estudo, quatro métodos normalizados de ensaios de estabilidade foram aplicados a amostras de óleo diesel coletadas em postos de revenda nas cinco regiões do Brasil. As alterações nas amostras foram avaliadas por espectrometria no infravermelho médio (FTIR) e, para algumas amostras, por cromatografia gasosa acoplada à espectrometria de massas (GC-MS). Os dados foram analisados com os modelos de Análise de Fatores Paralelos, PARAFAC, e Resolução de Curvas Multivariadas, MCR. O modelo PARAFAC dos dados FTIR permitiu observar alterações sutis e específicas nas amostras para cada tipo de ensaio a despeito das diferenças significativas entre os combustíveis. De forma complementar, a aplicação de MCR nos dados de GC-MS permitiu “separar matematicamente” classes de compostos ou espécies químicas específicas, facilitando a avaliação dos fenômenos envolvidos nos ensaios de estabilidade. As ferramentas quimiométricas permitiram observar que os ensaios promovem perda de compostos leves e a oxidação de alguns componentes, especialmente os ésteres mais insaturados do biodiesel. A forma e a extensão desses fenômenos, entretanto, variou muito de um ensaio para o outro. Os resultados mostraram que nenhum dos ensaios de curta duração provoca fenômenos semelhantes aos causados pelo teste de longa duração, que simula o comportamento do combustível em campo. No segundo estudo, cinquenta amostras de óleo diesel coletadas no Distrito Federal foram analisadas por FTIR e com o ensaio que segue a norma CEN EN 15751, com o objetivo de criar um modelo para prever o resultado desse ensaio a partir de FTIR. Foram testadas as ferramentas de Regressão nos Mínimos Quadrados Parciais, PLS e Regressão por Vetores de Suporte, SVR. Infelizmente os resultados não foram satisfatórios, mostrando que a espectrometria no FTIR, nas condições deste estudo, foi incapaz de detectar as substâncias relevantes que afetam os resultados do ensaio de estabilidade selecionado.

**Palavras-chave:** Óleo diesel, estabilidade de combustíveis, PARAFAC, MCR, quimiometria, calibração multivariada.

## Abstract

Storage stability of automotive diesel fuel is an issue of strong economic interest since its degradation generates products that may harm engines and deteriorate the quality of gaseous emissions. The mixture of biodiesel in diesel fuel brought new concerns. The test methods to evaluate diesel fuel stability are based on different ways of stressing the fuel and measuring its degradation, being of great value the understanding of the phenomena involved on these tests and the correlation among the results. In a first study, four standard test methods for stability tests were applied to diesel fuel samples collected in gas stations at the five Brazil regions. The modifications were evaluated with Fourier transform infrared spectroscopy (FTIR) and, for some samples, with gas chromatography-mass spectrometry (GC-MS). The data were evaluated with Parallel Factor Analysis models, PARAFAC, and Multivariate Curve Resolution, MCR. The PARAFAC model for FTIR data allowed to observe subtle and specific transformations for each type of assay despite the significant differences among fuel samples. In addition, the use of MCR in GC-MS data allowed to “mathematically separate” classes of compounds or specific chemical species making it easier the interpretation of the involved phenomena. The chemometric tools allowed to observe that the assays promoted the loss of light species and the oxidation of others, specially the more unsaturated biodiesel compounds. The form and extension of those phenomena, however, varied greatly from one assay to another. The results showed that none of the short terms tests provoked similar phenomena to the long term one, which simulates the behavior of the fuel in real conditions. In the second study, fifty diesel fuel samples collected in Federal District were analyzed with FTIR and with the stability test according to CEN standard EN 15751, aiming to create a model to estimate its results from infrared data. Partial least squares regression, PLS, and support vector regression, SVR were tested. Unfortunately, these results were not satisfactory, showing that FTIR spectrometry, in the conditions of this study, was not capable of detecting the relevant substances that affect the results of chosen stability test.

**Keywords:** Diesel fuel, fuel stability, PARAFAC, MCR, chemometrics, multivariate calibration.

## 1. INTRODUÇÃO E OBJETIVOS

Com um consumo anual de mais de 55 bilhões de litros por ano, o óleo diesel é o combustível automotivo mais importante no Brasil<sup>1</sup>. O óleo diesel é definido como um derivado do petróleo constituído por hidrocarbonetos de 10 a 25 átomos de carbono com faixa de destilação situada, comumente, entre 150 °C e 400 °C, e que apresenta um conjunto de propriedades que permitem sua utilização em motores que funcionam segundo o ciclo Diesel<sup>2</sup>. Por força de lei<sup>3</sup>, todo óleo diesel comercializado no Brasil para uso automotivo sofre adição de biodiesel em percentual estabelecido pelo Conselho Nacional de Política Energética, CNPE. O teor de biodiesel vem crescendo desde a implementação do Programa Nacional de Uso e Produção do Biodiesel em 2004, quando era, opcionalmente, adicionado a 2 % (V/V), sendo este percentual obrigatório em todo o território nacional a partir de janeiro de 2008. O teor atual é de 12 % (V/V)<sup>4</sup>.

A estabilidade do óleo diesel é objeto de pesquisas acadêmicas<sup>5</sup> e da indústria há muitos anos, mas as novas tecnologias e rotas de produção, bem como crescentes exigências ambientais e da indústria<sup>6</sup>, alteraram a composição e a suscetibilidade desse combustível às degradações químicas e biológicas. Isso gerou novas demandas por métodos de avaliação de estabilidade e de investigação dos processos físicos e químicos envolvidos. Apesar dos benefícios ambientais e econômicos<sup>7</sup>, a mistura do biodiesel ao óleo diesel promovida em diversos países do mundo<sup>6</sup> trouxe novos elementos complicadores a esse cenário tanto por conta da própria biodegradabilidade e suscetibilidade à oxidação do biodiesel<sup>8,9</sup>, quanto por contaminantes presentes no biocombustível<sup>10</sup>.

O termo “estabilidade de combustível” refere-se à resistência desse produto a alterações<sup>5</sup> que podem estar relacionadas à *estabilidade ao armazenamento* ou à *estabilidade térmica*. O primeiro termo se refere ao comportamento do combustível em condições normais de armazenamento e o segundo à sua resistência à exposição a altas temperaturas por curtos períodos de tempo, como aqueles observados em sistemas de recirculação de combustíveis<sup>11,12</sup>. O simples acompanhamento das propriedades de um combustível ao longo do tempo não é uma prática exequível para avaliação da estabilidade ao armazenamento; tanto porque existem muitos fatores que afetam qualitativa e quantitativamente a degradação do combustível<sup>13</sup>, quanto pelo tempo necessário para essa avaliação. O impacto de alterações – dos parâmetros físicos e químicos – no desempenho do óleo diesel é outro fator complicador na avaliação da

estabilidade do óleo diesel, sendo necessário determinar quais alterações são aceitáveis e não acarretam perda de desempenho do combustível.

Nesse contexto, a indústria desenvolveu diversas normas técnicas que estabelecem métodos de avaliação da estabilidade de combustíveis para fornecer métodos reproduzíveis em condições laboratoriais. Essas normas variam entre aquelas que aceleram brandamente a degradação do produto<sup>14</sup>, até métodos que buscam estressar ao máximo o combustível no menor intervalo de tempo<sup>11,15</sup>. Os ensaios variam grandemente em relação ao custo de equipamentos, expertise requerida do operador e demanda de tempo. A própria variedade de métodos disponíveis para avaliação de estabilidade do óleo diesel demonstra a importância e a complexidade do assunto. A título de exemplo, somente a norma ASTM D975<sup>16</sup>, que estabelece a especificação do óleo diesel nos EUA, cita 7 métodos para avaliação de estabilidade. Cada método está baseado em uma forma diferente de estressar o combustível até que ele sofra alterações mensuráveis<sup>11,14,17</sup>.

Tendo em vista a importância de se ter conhecimento dos fenômenos e transformações químicas ligadas à utilização desses testes de estabilidade do óleo diesel, e considerando as constantes mudanças que vêm ocorrendo na sua composição (e suas misturas com biodiesel), a presente tese de doutorado tem os seguintes objetivos:

- Entender e comparar as alterações químicas e físicas promovidas pelos ensaios de estabilidade utilizando a espectrometria no infravermelho por transformada de Fourier (FTIR, do inglês *Fourier Transform Infrared Spectroscopy*) e a cromatografia gasosa com detecção por espectrometria de massas (GC-MS do inglês, *Gas Chromatography – Mass Spectrometry*);
- Avaliar a aplicabilidade de ferramentas quimiométricas multidimensionais para extrair informações subjacentes dos resultados obtidos com as técnicas FTIR e GC-MS. Dessa forma, espera-se entender os tipos de alterações químicas e físicas promovidas para cada tipo de ensaio considerando os diferentes óleos diesel (e biodieseis) disponíveis no mercado.
- Criar um modelo de calibração multivariada empregando medidas de reflectância total atenuada e FTIR (ATR-FTIR, do inglês, *Attenuated Total Reflectance – Fourier Transform Infrared Spectroscopy*) que permita, de forma direta e rápida, estimar os resultados do ensaio de estabilidade à oxidação descrito pela norma EN15751.

## 2. FUNDAMENTAÇÃO TEÓRICA

### 2.1. Óleo diesel

A Figura 1 mostra a preponderância da participação do óleo diesel dentre os combustíveis líquidos no Brasil. Isso se deve, principalmente à forte participação do modal rodoviário na matriz de transportes brasileira, que, em 2019, respondia por 61 % do total das cargas transportadas<sup>18</sup>.

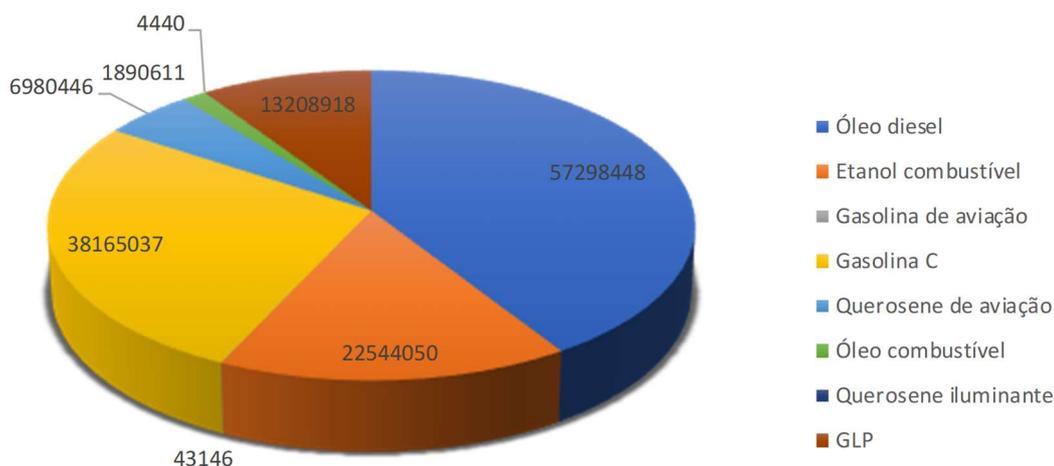


Figura 1: Volume de combustíveis líquidos (m<sup>3</sup>) comercializados no Brasil em 2019 (Fonte: Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – disponível em <http://www.anp.gov.br/dados-estatisticos> - acessado em 23/03/2020).

A Figura 2 mostra a cadeia de distribuição de combustíveis no Brasil com ênfase para o óleo diesel. Nota-se que a mistura óleo diesel/biodiesel é a composição de produtos de diversas origens e que pode estar sujeita a diferentes condições de transporte e armazenamento. O TRR, Transportador Revendedor Retalhista, é o agente que entrega o produto para grandes consumidores, como transportadoras, fazendas e locais que utilizam geradores de energia de grande porte.

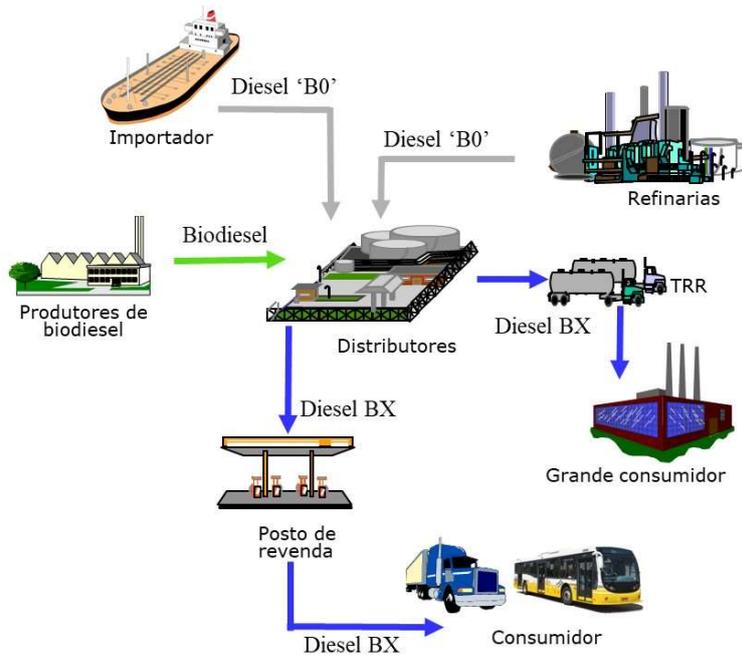


Figura 2: Cadeia de distribuição da mistura óleo diesel/biodiesel no Brasil. [adaptado de Silva, Alexandre, D. Distribuição e comercialização de biodiesel e suas misturas. 6º Congresso Internacional de bioenergia. 2011.]

### 2.1.1. Produção de óleo diesel

A produção de óleo diesel, no Brasil e no mundo, é baseada principalmente no processo de destilação a partir de cortes de gasóleos atmosféricos, leve e pesado, seguido de hidrodessulfurização<sup>2</sup>. A Figura 3 apresenta um esquema de refino que representa, de forma geral, as refinarias brasileiras.

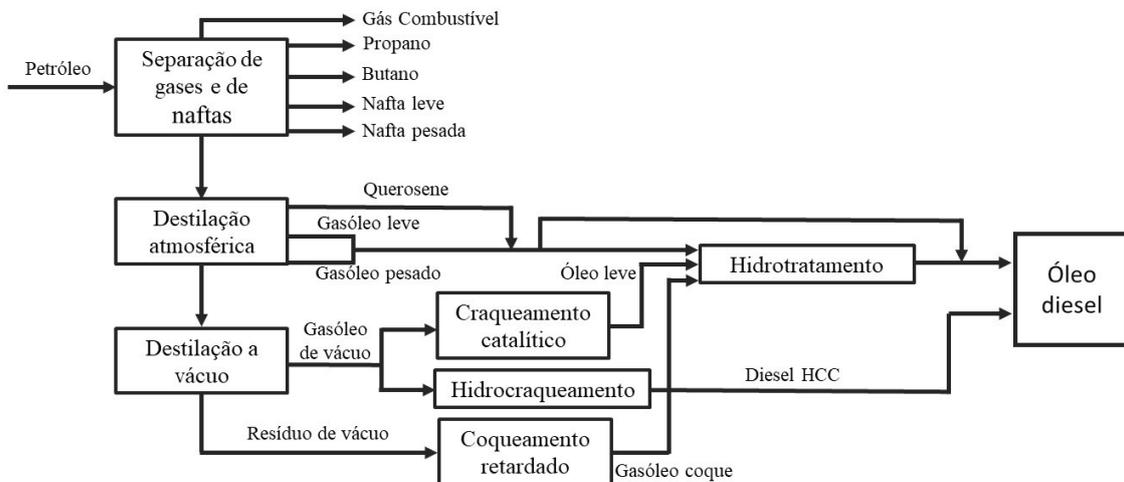


Figura 3: Esquema de refino típico para produção de óleo diesel<sup>2</sup>.

A seguir, são descritas, de forma sucinta, as etapas de refino mais importantes para produção de óleo diesel<sup>2,19</sup>:

- Destilação atmosférica – Por meio do aquecimento do petróleo (previamente dessalgado) a 300 – 400 °C, várias frações são separadas com base nas temperaturas de ebulição. As frações pesadas, que não podem ser vaporizadas, seguem para a destilação a vácuo. Nesta etapa, obtém-se produtos na faixa da nafta, gasolina, querosene, gasóleo leve, gasóleo pesado, resíduo de fundo e gás de refinaria.
- Destilação a vácuo – Frações pesadas da destilação atmosférica são aquecidas a 400 °C e pressões de 40 a 100 mbar. Nessas condições, é possível vaporizar substâncias que sofreriam decomposição térmica antes de entrarem em ebulição. Os produtos usuais desta etapa são: gasóleo leve, gasóleo pesado e resíduo de vácuo.
- Coqueamento retardado – É um processo térmico não catalítico de craqueamento. Nesta etapa, a carga (geralmente, resíduo de vácuo) é aquecida em uma fornalha antes de ser conduzida ao circuito de alimentação das câmaras de coqueamento, onde se forma e acumula o coque. Formam-se, principalmente, gasóleo, coque e gás etano.
- Hidrotratamento – Consiste na reação de contaminantes presentes no petróleo com hidrogênio a pressões e temperaturas elevadas. As reações ocorrem em leitos catalíticos que promovem a redução de compostos sulfurados, nitrogenados e oxigenados, além da saturação de olefinas. Dessa forma, obtém-se um combustível com menor teor de enxofre e maior estabilidade química e térmica.
- Hidrocraqueamento (HCC) – utiliza um reator no qual o fracionamento ocorre a pressões elevadas (35 kg/cm<sup>2</sup> a 100 kg/cm<sup>2</sup>), na presença de hidrogênio a temperaturas de 280 a 475 °C.

Seja para manutenção de propriedades ao longo da cadeia de distribuição do combustível, seja para adequar suas características à legislação vigente, utiliza-se aditivos na produção de óleo diesel, sendo os mais comuns<sup>2</sup> os biocidas, os antiespumantes e os melhoradores de cetano, lubricidade e escoamento a frio.

## 2.2. Biodiesel

O biodiesel é definido como um combustível composto de alquil ésteres de ácidos carboxílicos de cadeia longa, produzido a partir da transesterificação e/ou esterificação de matérias graxas, de gorduras de origem vegetal ou animal, e que atenda a especificação contida na Resolução ANP n° 45 de 25 de agosto de 2014<sup>20</sup>. Apesar de diferenças entre especificações adotadas em cada país<sup>21-23</sup>, o biodiesel pode ser considerado o mesmo produto no mundo inteiro, até mesmo porque, as características construtivas e de desempenho dos motores seguem padrões mundiais. Parâmetros (e limites) específicos de cada país apresentam diferenças que refletem características locais de clima, matéria-prima e forma de introdução do biocombustível na cadeia<sup>16,21,24</sup>.

Dentre as várias opções tecnológicas para uso de óleos vegetais como combustíveis, a razão para adoção do biodiesel como sucedâneo do óleo diesel – ou como produto a ser misturado a ele – reside, essencialmente, na viscosidade desse produto, já que óleos vegetais apresentam viscosidades com uma ordem de grandeza superior ao óleo diesel, o que tem impacto negativo no sistema de injeção dos motores diesel<sup>6</sup>. Outras estratégias para uso de óleos vegetais nesse tipo de motor foram e continuam sendo objeto de estudo e vão do uso de óleos vegetais *in natura*<sup>25-28</sup>, até o uso de produtos de craqueamento dessa biomassa<sup>29,30</sup>, passando por misturas binárias de óleo diesel e óleo vegetal<sup>31-33</sup> e misturas ternárias contendo óleo diesel, óleo vegetal e surfactantes (que estabilizam a mistura)<sup>25,34,35</sup>.

A matéria-prima utilizada para produção de biodiesel afeta significativamente suas propriedades físicas e químicas. Assim, por exemplo, biodiesel produzido exclusivamente a partir de sebo bovino causa entupimento de filtros por cristalização se a temperatura desce abaixo de 9°C<sup>6</sup>, mas apresenta boa estabilidade à oxidação por ser composto majoritariamente de ésteres de ácidos graxos saturados<sup>6,36</sup>. Já o óleo de soja fornece um biodiesel com fluidez adequada até -2°C, mas requer o uso intensivo de antioxidantes em função da presença de cadeias insaturadas em sua composição<sup>37</sup>. A Tabela 1 mostra a composição aproximada em termos de ácidos graxos de matérias-primas típicas para produção de biodiesel.

Tabela 1: Composição em termos de ácidos graxos (% m/m) de algumas matérias-primas para biodiesel<sup>6,38</sup>. O código sob o nome do ácido graxo corresponde ao número de átomos de carbono na cadeia alquílica e o número de insaturações na mesma.

<b>Óleo ou gordura</b>	<b>Mirístico C 14:0</b>	<b>Palmitico C 16:0</b>	<b>Estearico C 18:0</b>	<b>Oleico C18:1</b>	<b>Linoleico C 18:2</b>	<b>Linolênico C 18:3</b>	<b>Erúxico C 22:1</b>
<b>Canola</b>		1,5-6	1-2,5	52-66,9	16,1-31	6,4-14,1	1-2
<b>Algodão</b>	0,6-1,5	21,4-26,4	2,1-5	14,7- 21,7	46,7-58,2		
<b>Soja</b>		2,3-13,3	2,4-6	17,7- 30,8	49-57,1	2-10,5	0-0,3
<b>Sebo bovino</b>	2,1-6,9	25-37	9,5-34,2	14-50	26-50		

### **2.2.1. Produção de biodiesel**

O Brasil contava, em 2019, com 51 usinas autorizadas a produzir biodiesel, totalizando uma capacidade nominal de 9,2 milhões de metros cúbicos anual<sup>39</sup>. Em 2019, foram comercializados 5,9 milhões de metros cúbicos de biodiesel sendo a matéria-prima mais utilizada o óleo de soja (68,3 % em vol.) seguido de gordura animal (14,1 % em vol.).

A principal rota para produção do biodiesel no Brasil utiliza transesterificação com catalisador básico e metanol. Essa escolha se justifica pela eficiência da reação aliada ao baixo custo dos insumos<sup>40</sup>. A Figura 4 representa a principal reação envolvida na produção de biodiesel.

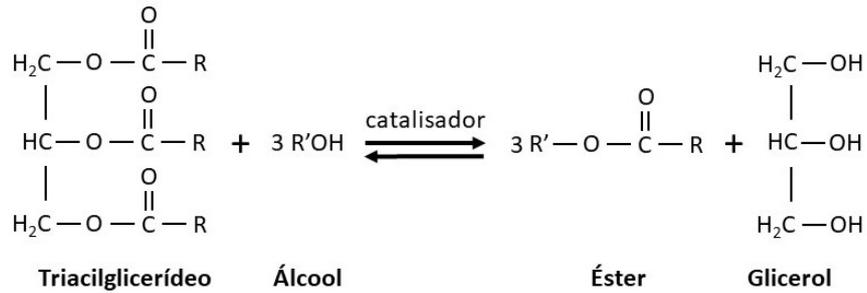


Figura 4: Reação de transesterificação na qual R representa diferentes cadeias de ácidos graxos e R', usualmente, é igual a CH<sub>3</sub> <sup>6</sup>.

A Figura 5 apresenta um esquema típico de produção de biodiesel a partir de óleo de soja e metanol<sup>41</sup>. Detalhes do processo podem ser diferentes em cada planta em função da tecnologia utilizada ou particularidades da matéria-prima.

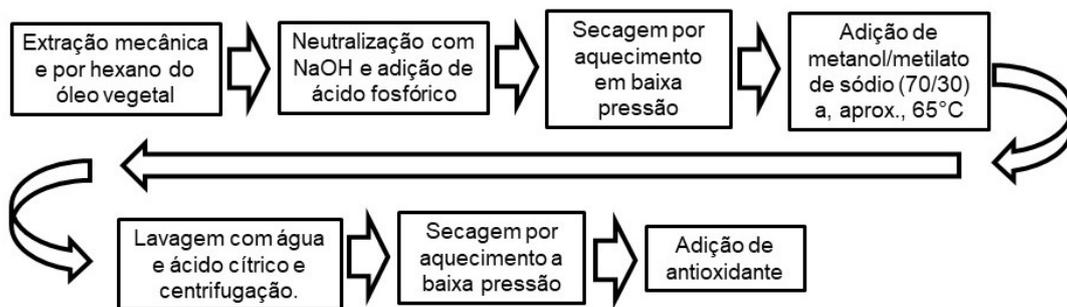


Figura 5: Esquema típico de obtenção de biodiesel a partir de óleo de soja. O ácido fosfórico é adicionado na segunda etapa para transformar lipídeos não hidrolisáveis em hidrolisáveis.<sup>41</sup>

### 2.3. Misturas de óleo diesel e biodiesel

O uso de biodiesel em mistura ao óleo diesel foi adotado em várias partes do mundo como uma medida para reduzir o impacto ambiental das emissões veiculares, em especial, o aumento do efeito estufa<sup>6,7</sup>. Entretanto, outros benefícios também são esperados ou atribuídos a essa estratégia energética, como redução da dependência de combustíveis fósseis, fixação de trabalhadores no campo e redução de desigualdades sociais<sup>7,42</sup>.

O biodiesel foi introduzido na matriz energética brasileira por meio do Programa Nacional de Produção e Uso do Biodiesel (PNPB) conforme Lei n° 11.097 de 13 de janeiro de 2005<sup>3</sup> que, incluindo legislações subsequentes<sup>43,44,4,45</sup>, estabeleceu um

cronograma de implementação e aumento gradual de concentração do biocombustível no óleo diesel descrito na Tabela 2.

Tabela 2: Evolução do teor de biodiesel para uso rodoviário no Brasil.

<b>Prazo</b>	<b>Teor de biodiesel / %/v/v</b>
Janeiro de 2003	Facultativo até 2
Janeiro de 2008	2
Julho de 2008	3
Julho de 2009	4
Janeiro de 2010	5
Julho de 2014	6
Novembro de 2014	7
Março de 2017	8
Março de 2018	10
Setembro de 2019	11
Março de 2020	12

#### **2.4. Estabilidade do óleo diesel, do biodiesel e de suas misturas**

Diferentes óleos diesel produzidos dentro das especificações de qualidade podem apresentar diferenças significativas em função de métodos de refino, petróleo utilizado, manuseio, contaminações e mistura com outros combustíveis<sup>2,46</sup>. Esses aspectos afetam a forma como o combustível se degrada no armazenamento. Além disso, muito frequentemente, os processos químicos de degradação são devidos à presença de substâncias em concentrações traço<sup>46,47</sup>.

*Nota: A presença de água no óleo diesel pode promover degradações decorrentes de fenômenos microbiológicos que, apesar de grande importância prática, não fazem parte do escopo deste estudo<sup>2</sup>.*

A perda da estabilidade do óleo diesel está relacionada à presença de compostos nitrogenados básicos, tiofenóis, olefinas e ácidos orgânicos, que são precursores de degradação, que levam à formação de gomas e sedimentos e à mudança de cor do produto. Essas reações podem ser aceleradas pela catálise por metais e pela alta temperatura do sistema de injeção dos motores. Também podem ocorrer reações de esterificação de

hidrocarbonetos aromáticos com compostos nitrogenados e sulfurados gerando sedimentos<sup>2,5</sup>.

Não há um modelo unificado que permita explicar todos os fenômenos associados à degradação do óleo diesel<sup>5</sup>. De forma geral, entretanto, a instabilidade do óleo diesel pode ser atribuída a reações de oxidação e auto-oxidação (ver Figura 6). Os hidrocarbonetos insaturados e compostos com heteroátomos reagem com oxigênio, formando hidroperóxidos que se decompõem em produtos de oxidação solúveis e insolúveis. Estes, concentram os átomos de oxigênio, enxofre e nitrogênio em compostos de alta massa molar (200 a 600 Da). Mecanismos de nucleação ou aglomeração levam à formação de gomas e depósitos aderentes<sup>5</sup>.

Em função da sua composição, o biodiesel é mais suscetível à degradação por oxidação que combustíveis fósseis<sup>21,9</sup>. Isso é especialmente válido quando há alta concentração de ésteres com duas ou mais insaturações, já que grupos metílenos adjacentes a ligações duplas são particularmente suscetíveis a ataques radicalares que constituem o primeiro passo no processo de oxidação<sup>48</sup>. Os hidroperóxidos assim formados podem polimerizar para formar sedimentos insolúveis e gomas, que são associados com entupimento de filtros e formação de depósitos no sistema de injeção e câmara de combustão<sup>21,49</sup>. A preocupação com compostos insaturados é tão grande que a concentração de linolenato de metila é restrita na norma EN 14214 (especificação europeia do biodiesel) em função da propensão dessa substância a sofrer oxidação<sup>22,50</sup>. Apesar disso, o valor limite foi estabelecido em 12 % m/m com vistas a não excluir o óleo de canola, importante matéria-prima para biodiesel na Europa que é rica em éster de ácido linolênico<sup>6,50</sup>.

A estabilidade do biodiesel está ligada à sua composição em termos de ácidos graxos, mas, também, a substâncias presentes em menores quantidades<sup>51</sup>. Essas advêm da matéria-prima, do processo de produção ou da degradação do biodiesel ao longo do tempo.

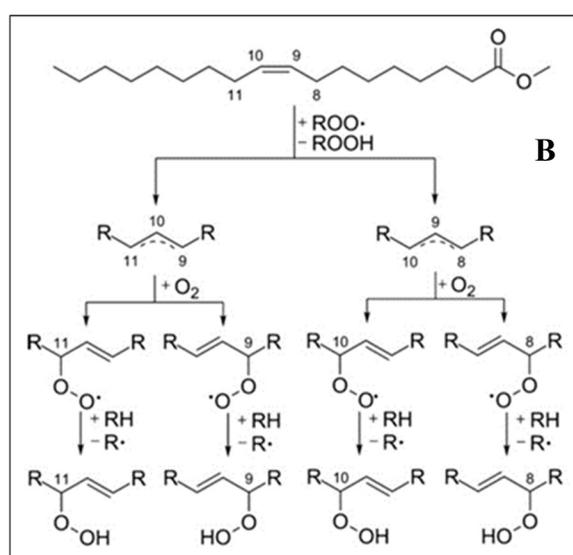
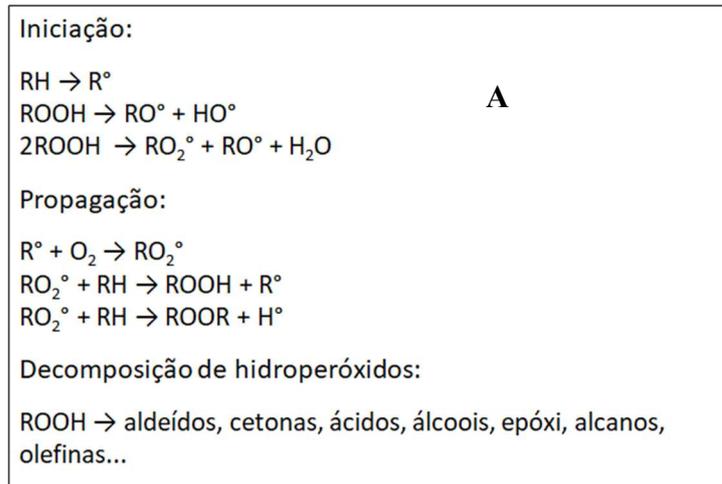


Figura 6: Representação esquemática das reações **A.** de oxidação e auto-oxidação do óleo diesel e **B.** da formação de peróxidos e hidroperóxidos em biodiesel.<sup>49,52</sup>

Esteróis, antioxidantes naturais e hidrocarbonetos presentes nos óleos vegetais podem contaminar o biodiesel, mesmo após as etapas usuais de purificação<sup>10,38,53,54</sup>. Também podem estar presentes glicerídeos intermediários formados na transesterificação (mono e diglicerídeos), além de triglicerídeos que não reagiram, água e sabões. Também podem restar pequenas quantidades de glicerol, ácidos graxos livres, álcool e catalisador. Todos esses contaminantes podem, dependendo da concentração e das condições de uso, causar problemas em função do acúmulo de depósitos nos motores e entupimento ou mau funcionamento de filtros e bicos injetores. Esses produtos podem, também, contribuir para acelerar o processo de deterioração do próprio biodiesel e causar corrosão em algumas partes dos motores.

Glicosídeos de esteróis são uma categoria especialmente importante de substâncias presentes no biodiesel no que se refere à estabilidade<sup>53,54</sup>. Essa classe de substâncias ocorre em baixa concentração nos óleos vegetais, mas pode contribuir significativamente na formação de depósitos em biodieseis e, conseqüentemente, na alteração de aspecto e, possivelmente, no entupimento de filtros de motores<sup>54,55</sup>. Vários autores relatam preocupação com a formação de peróxidos em biodiesel que poderia afetar adversamente elastômeros presentes nos motores<sup>5,12,16,36</sup>.

Em congresso recente, com a participação de representantes de fabricantes de motores e peças automotivas, foram relatadas ocorrências de campo envolvendo o uso de misturas diesel/biodiesel que incluem problemas de partida, baixa pressão de combustível, oscilação na rotação, presença de micro-organismos e borra, evidência de oxidação e corrosão e entupimento de filtros<sup>56</sup>. De acordo com os participantes, esses problemas são acentuados quando os veículos, geradores e equipamentos ficam parados por alguns meses e ocorrem mesmo quando o combustível está – inicialmente – dentro das especificações.

#### ***2.4.1. Ensaios para avaliação de estabilidade de óleo diesel e biodiesel***

Há duas categorias de ensaios de avaliação de estabilidade de combustíveis líquidos: os testes de longa duração e os testes acelerados. Na primeira categoria, simulam-se condições de estocagem reais e as alterações são monitoradas por testes convenientes. Na segunda categoria, estão os testes que aceleram a degradação do combustível por exposição a: **i.** temperaturas elevadas, **ii.** grandes quantidades de ar ou oxigênio, **iii.** catalisadores; ou **iv.** vapor. Esses ensaios são acompanhados de técnicas que permitam mensurar direta ou indiretamente alterações no produto. Alterações de cor, por exemplo, são monitoradas pelo ensaio de Cor ASTM<sup>57,58</sup>. Transmitância e espalhamento de luz também são utilizados, mas o parâmetro mais utilizado é a medida de material insolúvel formado, o que é feito, usualmente, por gravimetria<sup>5,8,11,59</sup>.

Alguns dos métodos mais tradicionais de acompanhamento de estabilidade de óleo diesel baseiam-se na quantificação de material insolúvel<sup>11,14,16,60</sup>. Esses ensaios focam no principal impacto de uso de combustíveis degradados: entupimento de filtros e outros elementos nos motores. Esses métodos, invariavelmente, exigem a presença e a atenção do analista por um período relativamente longo e, muitas vezes, apresentam baixa reprodutibilidade. Esses fatores levam ao constante desenvolvimento de novos métodos.

A taxa de degradação para vários tipos de combustíveis não se altera uniformemente com o aumento de temperatura. A relação entre temperatura e taxa de reação depende das energias de ativação das etapas de controle nas reações químicas. Isso varia de combustível para combustível e os produtos de degradação podem ser diferentes em condições de oxidação acelerada<sup>14</sup>. Assim sendo, a correlação dos resultados obtidos com ensaios acelerados e o comportamento do combustível em condições reais de armazenamento são questionáveis.

## 2.5. Espectrometria no infravermelho

A região espectral do infravermelho compreende radiação com números de ondas no intervalo de aproximadamente  $12.800$  a  $10\text{ cm}^{-1}$  (ou comprimentos de onda de  $0,78$  a  $1000\text{ }\mu\text{m}$ ) conforme representado na Figura 7. Do ponto de vista tanto da aplicação como da instrumentação, o espectro infravermelho é convenientemente dividido em infravermelho próximo, médio e distante<sup>61</sup>.

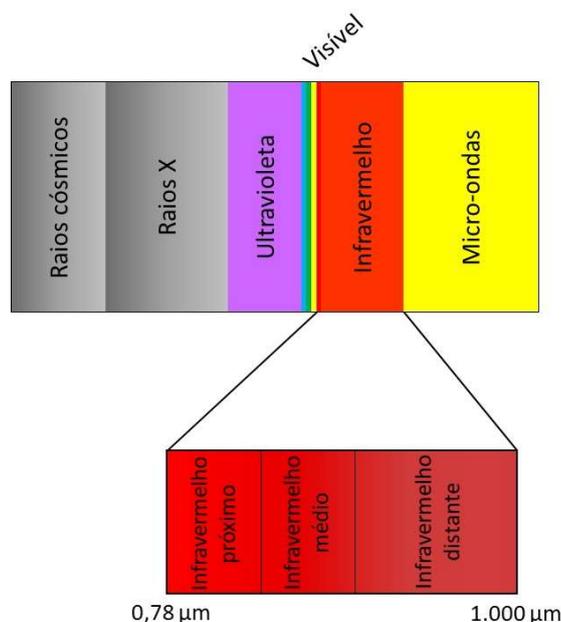


Figura 7: Representação esquemática das principais regiões do espectro eletromagnético.

Para absorver radiação infravermelha, uma molécula precisa sofrer uma variação no momento de dipolo como consequência do movimento vibracional ou rotacional. Apenas nessas circunstâncias o campo elétrico alternado da radiação pode interagir com a molécula e causar variações na amplitude de seus movimentos. As transições vibracionais (estiramento e deformação angular) são as mais importantes na região do infravermelho médio e, para sólidos e líquidos, fornecem bandas vibracionais mais ou menos alargados

(em função de interações entre as vibrações dos átomos e o resto do composto). Como cada ligação química corresponde a um nível de energia (quantizado), o espectro de uma molécula pode se tornar extremamente complexo.

A instrumentação padrão para espectrometria no infravermelho envolve o uso de interferômetros e a transformada de Fourier<sup>62</sup>. Nesses equipamentos, a radiação contendo todos os comprimentos de onda de interesse é separada em dois feixes. Um deles percorre uma distância fixa, e o outro, uma distância variável determinada por um espelho móvel. Fazendo-se variar as distâncias percorridas pelos dois feixes, obtém-se uma sequência de interferências construtivas e destrutivas e, conseqüentemente, variações na intensidade e frequência da radiação recebida pelo detector, formando o chamado interferograma. Uma operação matemática chamada de transformada de Fourier converte o interferograma, que está no domínio do tempo, para a forma mais familiar de um interferograma no domínio de frequências<sup>62</sup>.

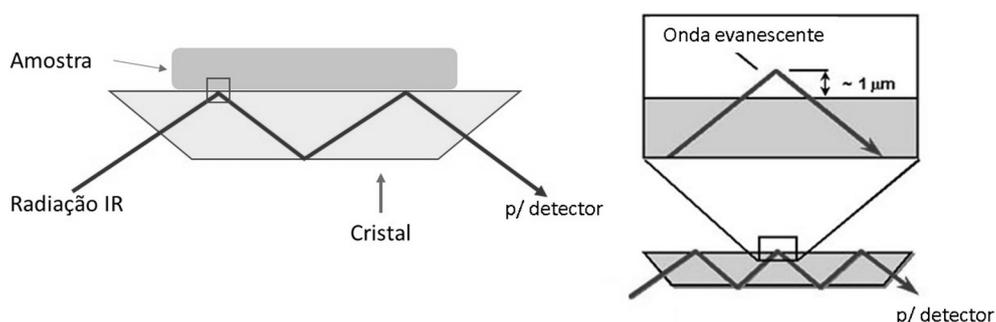


Figura 8: Representação de uma célula ATR [adaptado de Catálogo Pike Technologies]

63

Há vários recursos para manuseio de amostras para permitir que o material seja exposto à radiação infravermelha. Especificamente para amostras líquidas, pode-se comprimir uma gota do produto entre duas placas de sal ou injetá-lo numa célula que mantém as duas placas em distância fixa (usualmente, entre 0,01 e 1 mm). Outra opção, também aplicável em sólidos, é o uso de espectrometria de reflexão cujas variantes são intituladas Especular, Interna, Total Atenuada e Difusa, sendo as duas últimas mais comuns em aplicações analíticas<sup>62</sup>. Nas células de Reflectância Total Atenuada (do inglês, *Attenuated Total Reflectance* ou ATR), a radiação penetra no cristal em ângulo adequado e sofre reflexão total nas suas paredes internas. A cada reflexão, a radiação interage com o material que está em contato com o cristal sofrendo absorção (ver Figura 8). Dependendo do ângulo de incidência, do cristal e do comprimento de onda, a

penetração da radiação na amostra (onda evanescente) vai de 0,5 a 10  $\mu\text{m}$ <sup>63</sup>. O espectro no infravermelho obtido por ATR é muito similar ao obtido com célula de transmissão, exceto pelo fato de a magnitude da onda evanescente se alterar com o comprimento de onda. Usualmente, essa diferença é compensada (matematicamente) pelo *software* do equipamento<sup>64</sup>.

Tabela 3: Absorções importantes no infravermelho médio características de substâncias majoritárias no óleo diesel, no biodiesel e nos seus produtos de degradação.<sup>62,65</sup>

Número de ondas / $\text{cm}^{-1}$	Grupo	Tipo de vibração	Observação
3650-3584	OH	Estiramento	Álcoois (lig. Intermolecular)
3450	OH	Estiramento	Lig. Intermolecular
3008	CH	Estiramento	-CH=CH- (cis-olefina)
2953	CH	Estiramento assimétrico	Alifático (-CH <sub>3</sub> )
2925	CH	Estiramento assimétrico	Alifático (-CH <sub>2</sub> )
2854	CH	Estiramento simétrico	Alifático (-CH <sub>2</sub> )
1746	C=O	Estiramento	$\nu(\text{C}=\text{O})$ éster
1717	C=O	Estiramento	$\nu(\text{C}=\text{O})$ cetona
1463	CH	Deformação tesoura no plano	Alifático (-CH <sub>2</sub> )
1376	CH	Deformação angular simétrica	Alifático (-CH <sub>3</sub> )
1164	C-O	Estiramento	$\nu(\text{C}-\text{O})$ éster
1025	C-O-C	Estiramento	$\nu(\text{C}-\text{O}-\text{C})$ éster
966	CH	Torção fora do plano	Trans (-CH=CH-)
722	CH	Deformação balanço no plano	Alifático (-CH <sub>2</sub> )
675	CH	Torção fora do plano	Aromático ( $\bar{\nu}\text{C}-\text{H}$ )

O espectro no infravermelho médio inclui sinais específicos de substâncias presentes no óleo diesel e biodiesel, assim como dos produtos de degradação dessa mistura. A

Tabela 3 resume as absorções típicas desses produtos (e seus prováveis produtos de degradação).

## 2.6. Cromatografia gasosa e espectrometria de massas

Na cromatografia gasosa (GC), a amostra é vaporizada e injetada no topo de uma coluna cromatográfica. A eluição é feita por fluxo de um gás inerte que atua como fase móvel. A cromatografia gasosa (ou, mais apropriadamente, gás-líquido) baseia-se na partição do analito entre uma fase móvel gasosa e uma fase líquida imobilizada na superfície de um sólido inerte<sup>61</sup>. A eficiência da coluna cromatográfica requer que a amostra tenha tamanho adequado e seja introduzida como um plugue de vapor, pois a injeção lenta de amostras muito grandes causa espalhamento da banda e reduz a resolução. O método mais comum de injeção de amostra envolve o uso de microseringas para injetar uma alíquota através de um septo autosselante de borracha de silicone, em um vaporizador instantâneo localizado no topo da coluna (a porta de amostragem está geralmente cerca de 50 °C acima do ponto de ebulição do componente menos volátil da amostra). Para colunas analíticas comuns, os volumes injetados das amostras variam de alguns décimos de microlitro até 20 µL.

As colunas cromatográficas podem variar de 2 a 100 m de comprimento e serem construídas de aço inoxidável, vidro, sílica fundida ou Teflon<sup>61</sup>. Para permitir a separação dos componentes injetados na coluna, o cromatógrafo dispõe de sistemas que permitem variar o fluxo de gás de arraste e da temperatura do forno que contém a coluna.

As colunas mais populares em laboratórios que analisam combustíveis são as do tipo Tubo aberto com parede revestida, WCOT (do inglês, Wall-Coated Open Tubular) de sílica fundida. Elas consistem em um tubo de sílica com diâmetro da ordem de ¼ de milímetro revestido com a fase estacionária. O polidimetil-siloxano é um exemplo muito comum de fase estacionária utilizada para separação de hidrocarbonetos e outros compostos apolares. O poliglicol, por outro lado, é adequado para separação de compostos polares<sup>66</sup>.

Na espectrometria de massas, a amostra é vaporizada ou atomizada e, a seguir, uma fração substancial das espécies é ionizada para, então, serem separadas em função da razão massa/carga. Na técnica de ionização eletrônica (EI), uma fonte de elétrons bombardeia moléculas na fase de vapor (possivelmente oriunda de um cromatógrafo) com um feixe de elétrons de alta energia (tipicamente, 70 eV) e registra o resultado do impacto

dos elétrons como um espectro de íons separados na base da razão massa/carga ( $m/z$ ). Em função da energia do feixe de elétrons, moléculas orgânicas sofrem fragmentação o que, em condições adequadas, fornece uma “impressão digital” da molécula. O padrão de fragmentação pode ser comparado com bancos de dados de espectros de massas obtidos nas mesmas condições. A maior parte dos íons formados tem carga unitária<sup>62</sup>.

Há muitas opções de instrumentação para seleção dos íons gerados. O mais comum para acoplamento às técnicas cromatográficas é o filtro de massas de quadrupolo. Esse tipo de filtro usa quatro barras sujeitas a uma diferença de potencial (o “quadrupolo”). Íons atravessam o sistema em velocidade constante na direção paralela aos polos (direção  $z$ ), porém oscilam de modo complexo nas direções  $x$  e  $y$ , devido à aplicação simultânea de uma voltagem de corrente contínua ( $V_{dc}$ ) e de uma voltagem de radiofrequência ( $V_{rf}$ ) aos polos. Existe sempre uma determinada “oscilação estável” que faz com que um íon particular chegue à extremidade do quadrupolo sem tocar os polos. Essa oscilação depende da razão  $m/z$  do íon. Assim, apenas íons com uma determinada  $m/z$  atravessarão o filtro quadrupolo em um determinado conjunto de condições. Os demais íons atingirão os polos e serão retidos. A varredura de massas é feita variando as frequências  $rf$  e  $dc$ , porém mantendo-se suas razões.

O uso de cromatografia gasosa acoplada à espectrometria de massas data do final dos anos 1950<sup>67</sup> e permite unir a capacidade de separação da cromatografia e de identificação de compostos da espectrometria de massas. O cromatograma de íons totais (ou, TIC, do inglês, *Total Ion Chromatogram*) fornecido por um sistema GC-MS pode ser utilizado em aplicações qualitativas e quantitativas.

No primeiro caso (uso qualitativo), o usuário pode, simplesmente, examinar os espectros correspondentes de cada pico do TIC e compará-los com bibliotecas de compostos. Outra abordagem é reconstruir o TIC utilizando um ou mais íons característicos do composto de interesse<sup>67</sup>. Assim, por exemplo, o íon de razão massa/carga 74 é característico de compostos contendo carbonila, o que permite obter um cromatograma somente com picos correspondentes ao biodiesel a partir de um TIC de uma mistura óleo diesel-biodiesel.

Compostos altamente polares podem não ser suficientemente voláteis para análise por GC em temperaturas razoáveis. As técnicas de derivatização permitem alterar esses compostos tornando-os mais voláteis<sup>68</sup>. A silanização é a técnica mais prevalente de

derivatização<sup>68</sup> e consiste em introduzir um grupo silil na molécula, usualmente na forma de dimetilsilil, trimetilsilil, t-butildimetilsilil ou clorometildimetilsilil. O mecanismo envolve a substituição de um hidrogênio ativo (nos grupos -OH, -COOH, -NH, -NH<sub>2</sub> e -SH) por um grupo trimetilsilil. Os derivados silanizados são mais voláteis e estáveis produzindo picos cromatográficos estreitos e simétricos<sup>68</sup>.

A derivatização pode ser particularmente útil em GC-MS, já que é possível identificar facilmente os compostos silanizados monitorando a razão massa/carga do grupo silil<sup>69</sup>. Numa derivatização com MSTFA (n-metiltrimetilsililtrifluoroacetamida), por exemplo, a razão m/z 73 indicará o composto que sofreu silanização e a massa desse composto será dada pela massa do íon molecular – 15 u.m.a<sup>69</sup>.

## 2.7. Ferramentas Quimiométricas

A quimiometria é definida como a aplicação da estatística para a análise de dados químicos (da química orgânica, analítica ou medicinal) e o planejamento de experimentos químicos e simulações.<sup>70</sup> De forma geral, a quimiometria se utiliza de estatística multivariada, que consiste em um conjunto de métodos estatísticos/matemáticos utilizados em situações nas quais várias variáveis são medidas simultaneamente, em cada elemento amostral. Comumente, as variáveis são correlacionadas entre si e, quanto maior o número de variáveis, mais complexa torna-se a análise por métodos comuns de estatística univariada<sup>71</sup>.

### 2.7.1. Ferramentas quimiométricas multidimensionais

#### 2.7.1.1. Histórico

Algumas das primeiras ideias em análise multidimensional foram publicadas por Raymond Cattell<sup>72</sup>. Em 1944, ele propôs o uso do princípio de perfis paralelos proporcionais para a análise simultânea de várias matrizes. Esse princípio estabelece que um conjunto de fatores comuns poderia ser ajustado com diferentes pesos por dimensão para várias matrizes de dados simultaneamente.

Num relatório técnico de 1963 (publicado em 1964), Ledyard Tucker introduziu a decomposição de um conjunto de dados hipotético de 3 dimensões em matrizes de pesos **A**, **B** e **C**. Neste documento, utilizava-se o termo *modo* ao invés de *dimensão*, o que ainda é seguido por muitos autores atualmente.

Outro modelo de três dimensões foi introduzido independentemente por Carroll & Chang e por Harshman em 1970. Os primeiros autores chamaram seu modelo de CANDECOMP (do inglês *Canonical Decomposition*) enquanto Harshman propôs o nome PARAFAC (do inglês *Parallel Factor Analysis*). Como os dois modelos são equivalentes, alguns autores o chamam de PARAFAC-CANDECOMP. Neste trabalho, optou-se por utilizar apenas PARAFAC, por ser o mais comum nos textos envolvendo quimiometria<sup>73</sup>.

Merece destaque neste breve histórico, o desenvolvimento das ferramentas para solução *algébrica* de dados trilineares como o Método Generalizado de Aniquilação de Posto (GRAM - *Generalized Rank Annihilation Method*) proposto por Sanchez e Kowalski em 1986<sup>74</sup> e, dos mesmos autores, a Decomposição trilinear direta (DTD – *Direct Trilinear Decomposition*)<sup>73</sup>. O primeiro método utiliza a decomposição de dados em autovetores, mas é limitado a dados nos quais ao menos uma das dimensões seja igual a dois. O segundo método superou essa limitação, mas ambos apresentam limitações quanto aos critérios de otimização dos parâmetros, o que limita sua utilização.

Mais recentemente, Tauler et al.<sup>75</sup> propuseram a extensão da ferramenta de Resolução de Curvas Multivariadas (MCR, *Multivariate Curve Resolution*) para dados em três dimensões. No artigo de 1993, os autores aplicaram MCR a dados de titulação espectroscópica de um sistema em multiequilíbrio para obter os perfis de concentração das espécies presentes. Já em 1995, Tauler<sup>76</sup> utilizou MCR em dados de cromatografia líquida com detector de arranjo de diodos para demonstrar sua capacidade de resolução de sinais mesmo em situações complexas, como quando há desvios de trilinearidade.

#### 2.7.1.2. *Conceitos e notação*

Se um instrumento gera uma tabela de dados para cada amostra, como é o caso dos dados de excitação/emissão gerados num espectrofluorímetro, este resultado é uma matriz  $\mathbf{X}$  ( $J \times K$ ), na qual  $J$  comprimentos de onda de excitação são usados para medir a emissão em  $K$  comprimentos de onda. Tomando  $I$  medidas como essa, por exemplo, em diferentes ocasiões/amostras, obtém-se um conjunto de dados tridimensional de dimensões  $\underline{\mathbf{X}}$  ( $I \times J \times K$ )<sup>72</sup>. Obviamente, é possível imaginar conjuntos de dados com qualquer número de dimensões, entretanto, nesse trabalho, apenas conjuntos com 3 ou menos dimensões serão utilizados (ver Figura 9).

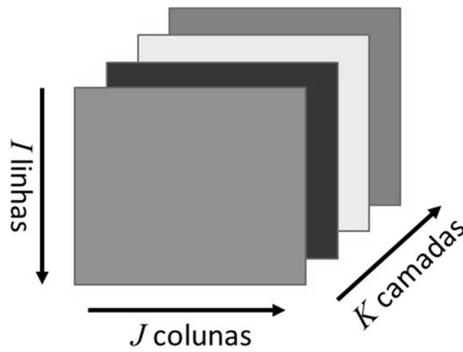


Figura 9: Representação de um conjunto de dados com 3 dimensões.

Para analisar conjuntos de dados de 3 dimensões, algumas operações elementares precisam ser relembradas<sup>72</sup>:

➤ Produto de Kronecker:

O produto de Kronecker de duas matrizes  $\mathbf{A}$  ( $I \times J$ ) e  $\mathbf{B}$  ( $K \times M$ ) é definido como:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \cdots & a_{1J}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{I1}\mathbf{B} & \cdots & a_{IJ}\mathbf{B} \end{bmatrix}$$

➤ Produto Hadamard

O produto Hadamard é definido para duas matrizes  $\mathbf{A}$  e  $\mathbf{B}$  de tamanhos iguais ( $I \times J$ ), como:

$$\mathbf{A} * \mathbf{B} = \begin{bmatrix} a_{11}b_{11} & \cdots & a_{1J}b_{1J} \\ \vdots & \ddots & \vdots \\ a_{I1}b_{I1} & \cdots & a_{IJ}b_{IJ} \end{bmatrix}$$

➤ Produto Khatri-Rao

O produto Khatri-Rao de duas matrizes  $\mathbf{A}$  e  $\mathbf{B}$  requer que elas sejam divididas em igual número de partições. Para as aplicações deste estudo, é útil o caso especial no qual as matrizes são particionadas em suas  $K$  colunas:

$$\mathbf{A} = [\mathbf{A}_1 \dots \mathbf{A}_K] \text{ e } \mathbf{B} = [\mathbf{B}_1 \dots \mathbf{B}_K]$$

$$\mathbf{A} \mid \otimes \mid \mathbf{B} = [\mathbf{A}_1 \otimes \mathbf{B}_1 \dots \mathbf{A}_K \otimes \mathbf{B}_K]$$

➤ Multiplicação de vetores

O produto externo dos vetores  $\mathbf{a}$  e  $\mathbf{b}$  é igual a uma matriz, já o produto tensorial,  $\Delta$ , de  $\mathbf{a}$ ,  $\mathbf{b}$  e  $\mathbf{c}$ , é um conjunto de dados de 3 dimensões.

$$\mathbf{a}\mathbf{b}' = \text{matriz}$$

$$\mathbf{a}\Delta\mathbf{b}\Delta\mathbf{c} = \text{conjunto de dados de 3 dimensões}$$



➤ Desdobramento (em inglês, *unfolding*)

Consiste, neste contexto, em transformar um conjunto de dados de 3 dimensões  $\underline{\mathbf{X}}$  ( $I \times J \times K$ ) em uma matriz. Uma das formas de fazê-lo está ilustrada na Figura 10 e fornece uma matriz de dimensões ( $I \times JK$ ):

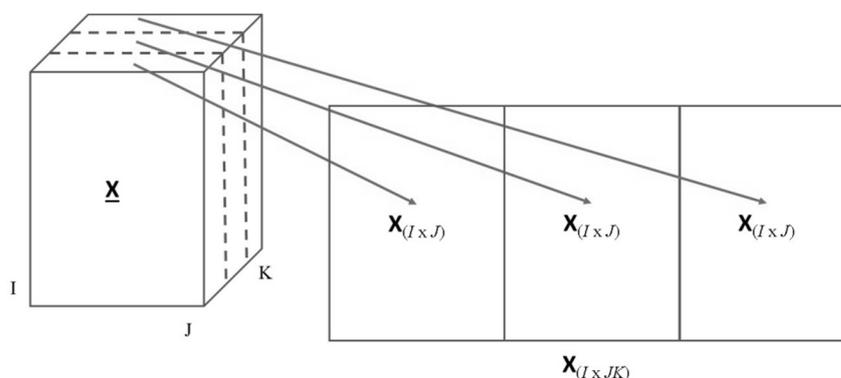


Figura 10: Desdobramento de um conjunto de dados de 3 dimensões.

Note que o desdobramento pode ser realizado de 5 outras formas: ( $I \times KJ$ ), ( $J \times IK$ ), ( $J \times KI$ ), ( $K \times IJ$ ), ( $K \times JI$ ).

Um conceito importante em quimiometria é o de posto ou dimensionalidade<sup>77</sup>. Para uma matriz  $\mathbf{A}$  ( $I \times J$ ) o número de colunas independentes é chamado de posto-coluna da  $\mathbf{A}$ <sup>72</sup>. Este é o posto-matemático, em oposição ao pseudo-posto ou posto-químico, que está relacionado às fontes *significativas* de variabilidade da matriz. Assim, por exemplo, os espectros no ultravioleta de 10 amostras *contendo a mesma espécie absorvente* em diferentes concentrações, obtidos com 100 comprimentos de onda, fornecem uma matriz ( $10 \times 100$ ) que, em princípio, deveria ter posto unitário. As linhas seriam apenas combinações lineares umas das outras. Entretanto, em função do ruído experimental, o posto (matemático) será igual a 10. Há várias abordagens para estimar o posto-químico de um conjunto de dados incluindo o conhecimento químico do sistema, o uso de valores singulares (de decomposição em valores singulares, SVD) e validação cruzada, na qual

uma parte dos dados é separada para testar a capacidade do modelo em prever os valores dos dados excluídos<sup>77</sup>.

O conceito de posto se torna mais complicado para dados de dimensionalidade maior que dois. O posto de um conjunto de dados de 3 dimensões  $\underline{X}$  é definido como o menor número de componentes trilineares necessários para ajustar  $\underline{X}$  exatamente<sup>72</sup>. Assim, para um conjunto de dados com dimensões 2 x 2 x 2, o posto máximo é 3, o que quer dizer que existe algum conjunto de dados com essas dimensões que não pode ser descrito utilizando apenas duas componentes<sup>78</sup>. Para um conjunto de dados de 3 x 3 x 3, o posto máximo é 5. Não há regras explícitas para determinação do posto máximo de qualquer conjunto de dados, exceto para duas dimensões e alguns conjuntos simples de 3 dimensões<sup>72</sup>.

### 2.7.1.3. Análise de Fatores Paralelos (PARAFAC)

Para dados em três dimensões, o modelo de Análise de Fatores Paralelos, PARAFAC, é composto pelas matrizes de pesos **A**, **B** e **C** que contêm os elementos  $a_{if}$ ,  $b_{jf}$  e  $c_{kf}$ , respectivamente, relativos às três dimensões dos dados<sup>73</sup>. O modelo trilinear (Equação 1) é ajustado para minimizar a soma dos quadrados dos resíduos  $e_{ijk}$ .

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk} \quad (1)$$

onde F é o número de fatores selecionados. A Figura 11 é a representação gráfica da Equação 1, na qual cada tríade de vetores corresponde a uma *componente PARAFAC*. Os vetores são, usualmente, chamados de pesos, mas em alguns textos, é possível encontrar a primeira dimensão (quando ela se refere às amostras) sendo chamada de escores se ela for associada às amostras<sup>79</sup>.

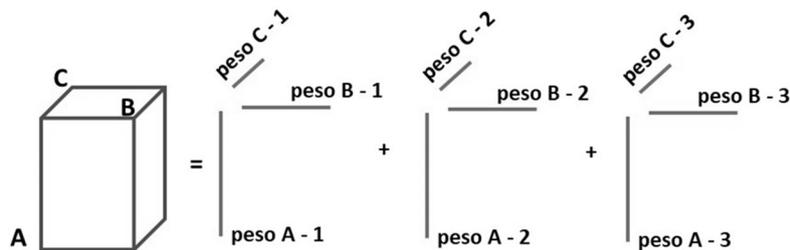


Figura 11: Em PARAFAC, um conjunto de dados de 3 dimensões é decomposto na soma de produtos de 3 vetores. Os vetores de pesos são chamados de A, B e C ou pelo nome da dimensão (e.g.: “pesos da dimensão das amostras”, “pesos da dimensão dos comprimentos de ondas”).[adaptado de Smilde et al.<sup>72</sup>].

Uma forma conveniente de representar o modelo PARAFAC utiliza o produto Khatri-Rao e o conjunto de dados  $\underline{\mathbf{X}}$  desdobrado na dimensão das linhas (Equação 2).

$$\mathbf{X}_{(I \times JK)} = \mathbf{A} (\mathbf{C} \mid \otimes \mid \mathbf{B})^T \quad (2)$$

onde  $T$  indica a transposta da matriz, ou seja, a troca das linhas por colunas.

### 2.7.1.3.1. Algoritmos para PARAFAC

Há diferentes formas de estimar os parâmetros de um modelo PARAFAC. Encontrar esses parâmetros é um problema de otimização e, em algumas situações, uma solução diretamente computável é possível; em outras, um algoritmo iterativo precisa ser utilizado<sup>72</sup>. As duas ferramentas mais importantes para ajuste de modelos em análise multidimensional são chamadas de *mínimos quadrados alternantes* e soluções baseadas em *autovalores*<sup>72</sup>.

O algoritmo a seguir estima os valores de  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  para um conjunto de dados  $\underline{\mathbf{X}}$  de três dimensões<sup>72</sup>  $I \times J \times K$  utilizando mínimos quadrados. O objetivo é minimizar a diferença entre o modelo dado na Equação 2 e  $\underline{\mathbf{X}}$  desdobrado como  $(I \times JK)$ .

$$\text{Objetivo: } \min \|\mathbf{X} - \mathbf{A} (\mathbf{C} \mid \otimes \mid \mathbf{B})^T\|^2$$

1. Inicialize  $\mathbf{B}$  e  $\mathbf{C}$  com um determinado valor de  $F$

$$2. \mathbf{Z} = (\mathbf{C} \mid \otimes \mid \mathbf{B})$$

$$\mathbf{A} = \mathbf{X}_{(I \times JK)} \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}$$

$$3. \mathbf{Z} = (\mathbf{C} \mid \otimes \mid \mathbf{A})$$

$$\mathbf{B} = \mathbf{X}_{(J \times IK)} \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}$$

$$4. \mathbf{Z} = (\mathbf{B} \mid \otimes \mid \mathbf{A})$$

$$\mathbf{C} = \mathbf{X}_{(K \times JI)} \mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}$$

5. Retorne ao passo 2 até que a mudança relativa no ajuste seja pequena.

onde  $\mathbf{Z} (\mathbf{Z}^T \mathbf{Z})^{-1}$  é a pseudo-inversa de  $\mathbf{Z}$  (inversa de Moore-Penrose).

Há várias opções na literatura para inicialização dos parâmetros<sup>80</sup>, desde a inicialização aleatória, que evita mínimos locais<sup>80</sup>, até ferramentas analíticas (em oposição às numéricas) como DTLD (do inglês, Direct Trilinear Decomposition ou Decomposição Trilinear Direta), GRAM (do inglês, Generalized Rank Annihilation ou Método Generalizado de Aniquilação de Postos) e SVD (do inglês, Singular Value Decomposition ou Decomposição em Valores Singulares). A seguir, algumas características de cada ferramenta:

### **Inicialização aleatória**

- > Pode não levar à melhor inicialização, pois a estimativa inicial pode estar distante dos parâmetros do modelo ótimo e consumir muito recurso computacional.
- > Útil quando existe o risco de “mínimos locais”. O modelo pode ser estimado repetidamente e os resultados comparados.

### **Inicialização analítica**

- > Busca-se o cálculo direto dos parâmetros. A estimativa inicial tende a estar muito mais próxima do valor ideal.
- > Risco de a solução ficar “presa” a “mínimos locais”, pois repetidas execuções do algoritmo levarão à mesma solução.

O método GRAM utiliza decomposição em valores singulares para estimar diretamente os parâmetros de um modelo PARAFAC que será utilizado na inicialização. O método é limitado a conjuntos de dados que tenham, no máximo, duas camadas em uma das dimensões. Quando não é este o caso, se utiliza o algoritmo DTLTD. A “estimativa” dessas camadas é feita por combinação linear de todas as camadas originais. Os detalhes de GRAM e DTLTD estão fora do escopo deste trabalho e podem ser encontrados na literatura<sup>72,74</sup>.

A definição de  $F$  passa pelo conceito de dimensionalidade discutido no subitem 2.7.1.2 e será apresentada adiante.

O critério de parada do algoritmo (passo 5) pode, por exemplo, ser uma alteração de ajuste menor que 0,0001 %, entre uma iteração e a próxima. Teoricamente, a convergência do ajuste não implica convergência dos parâmetros, mas, em situações práticas, este é o caso.

O algoritmo mostrado acima é computacionalmente ineficiente devido ao consumo excessivo de memória. Na prática, são utilizados outros algoritmos que se beneficiam de propriedades específicas de álgebra matricial, utilizando menos memória e levando aos mesmos resultados<sup>72</sup>. A discussão desses algoritmos está fora do escopo deste trabalho.

#### 2.7.1.3.2. *Estimativa do número de componentes*

Dois parâmetros são comumente utilizados para definir o número de componentes num modelo PARAFAC: o Diagnóstico de Consistência de Núcleo ou CORCONDIA (do inglês, *Core Consistency Diagnostic*) e a soma de quadrados de erros do modelo<sup>81</sup>. O parâmetro CORCONDIA é uma medida do quão próximo o modelo PARAFAC está de um modelo trilinear perfeito e está detalhado no Apêndice 1. Também deve-se considerar o conhecimento químico do sistema, por exemplo, fazendo-se a comparação entre espectros de referência e recuperados.

Na prática, modelos PARAFAC são construídos com números crescentes de componentes e os parâmetros acima são determinados para cada um<sup>81-83</sup>. O número de componentes é, então, escolhido como aquele que fornece o maior valor de CORCONDIA com o máximo de variação explicada utilizando o menor número de fatores possível (princípio da parcimônia<sup>73</sup>).

Como uma regra geral, valores de CORCONDIA acima de 90% indicam a adequação do modelo PARAFAC e valores abaixo de 50% indicam falta de trilinearidade ou excesso de componentes<sup>73</sup>.

#### 2.7.1.3.3. *Uso de restrições nos modelos PARAFAC*

Restrições são condições de contorno impostas ao modelo que buscam adequá-lo ao sentido físico dos dados ou ao conhecimento a priori<sup>78</sup> do sistema. As restrições mais comuns para modelos PARAFAC são<sup>81</sup>:

- Não negatividade: faz com que os perfis calculados (pesos) não possam ser negativos.
- Ortogonalidade: impõe que as componentes sejam ortogonais, ou seja, não correlacionadas entre si.
- Unimodalidade: impõe que os perfis calculados tenham apenas um máximo.

É importante observar que essas restrições podem ser impostas a todas as dimensões simultaneamente ou apenas às dimensões especificadas pelo usuário.

#### 2.7.1.3.4. *Pré-processamento de dados*

Apesar de serem procedimentos triviais em dados de duas dimensões, as operações de centralização na média e o escalamento são mais complexos em dados

multidimensionais porque eles podem ser aplicados e combinados de diversas formas<sup>84</sup>. Por exemplo, ao centralizar na média os dados em uma das dimensões, o escalamento feito em qualquer outra dimensão é afetado<sup>73</sup>.

Outra categoria de pré-processamento consiste na correção de desvios de tempo de retenção de dados cromatográficos, o que é especialmente necessário para aplicação de PARAFAC, já que essa ferramenta não permite desvios de trilinearidade<sup>81</sup>.

Neste estudo, nenhum pré-processamento de dados foi aplicado.

#### 2.7.1.3.5. *Algumas propriedades de PARAFAC*

O modelo matemático de PARAFAC pressupõe que exista uma relação trilinear entre os pesos e os dados experimentais. Isso pode não ser verdadeiro em todos os sistemas e acaba por dificultar a aplicação dessa ferramenta em dados cromatográficos, por exemplo, pois, como já foi citado, os deslocamentos de tempos de retenção constituem desvios significativos do modelo trilinear.

Outro aspecto importante de PARAFAC é que se trata de um modelo não cumulativo<sup>73</sup>. Isso implica que os parâmetros estimados para um modelo com  $F + 1$  componentes não são iguais aos de um modelo com  $F$  componentes mais um fator adicional. Além disso, as componentes não são ortogonais entre si, exceto que isso seja imposto com o uso de restrições<sup>72</sup>.

Um aspecto importante da ferramenta PARAFAC é o fato de apresentar *solução única* para o conjunto de dados<sup>80</sup>. Utilizando-se ferramentas tradicionais de decomposição de dados, os pesos obtidos numa decomposição bilinear podem não refletir os espectros puros, exceto que seja inserida alguma informação externa. Este problema é chamado de ambiguidade rotacional. Essa situação não ocorre com PARAFAC, pois, se os dados forem realmente trilineares, os espectros intrínsecos (ou qualquer que seja a variável) serão encontrados se o número de componentes correto for selecionado e a relação sinal/ruído for apropriada. O significado matemático da solução única é que o modelo PARAFAC não pode sofrer rotação sem perda de ajuste.

Por fim, cabe mencionar uma importante propriedade de PARAFAC (mas que não é exclusiva dele): a vantagem de segunda ordem, que consiste na possibilidade de quantificar analitos na presença de interferentes não modelados<sup>85</sup>.

#### 2.7.1.4. Resolução de Curvas Multivariadas (MCR)

A ferramenta de Resolução de Curvas Multivariadas tem como objetivo a obtenção de estimativas dos perfis puros dos componentes presentes em uma matriz de dados<sup>86</sup>. Usualmente, MCR é implementado por mínimos quadrados alternantes (em inglês, Alternating Least Squares, ou ALS); por isso, o acrônimo mais utilizado é MCR-ALS.

MCR permite recuperar informações sobrepostas não seletivas provenientes de uma matriz **D** em contribuições com sentido físico dos componentes puros do sistema. A Equação 3 mostra o modelo MCR e a notação adotada neste trabalho.

$$\mathbf{D} = \mathbf{C} \mathbf{S}^T + \mathbf{E} \quad (3)$$

onde **D** é uma matriz com a resposta instrumental;

**C** é a matriz com as intensidades relativas;

**S** é a matriz de perfis espectrais;

**E** é a matriz com a parte não modelada de **D**.

A Figura 12 mostra a representação dessa decomposição para uma matriz **D** obtida de uma corrida cromatográfica com detecção em vários comprimentos de onda.

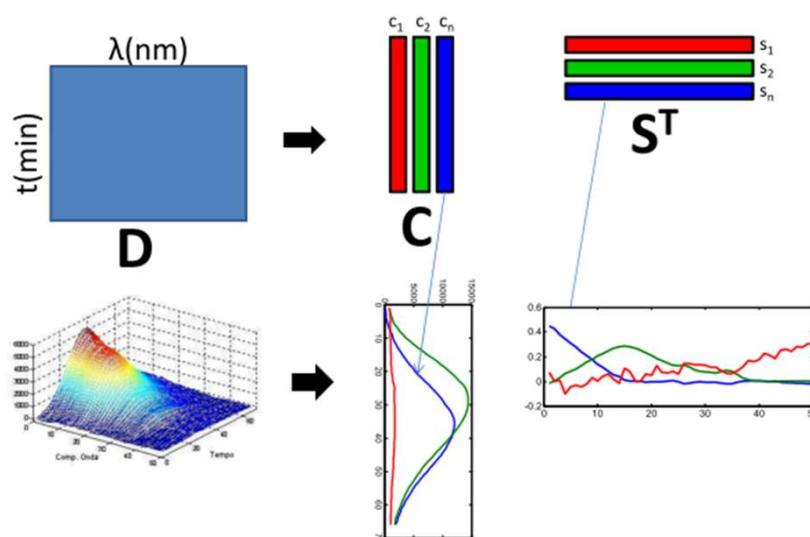


Figura 12: Deconvolução de uma matriz **D** nas matrizes de concentrações **C** e de espectros **S<sup>86</sup>**.

Ao contrário do que ocorre com PARAFAC, MCR apresenta liberdade (ou ambiguidade) rotacional<sup>81</sup>. Isso quer dizer que é possível promover a rotação das matrizes **C** e **S** sem alteração na matriz **E** (ou seja, sem alteração na qualidade do ajuste). Esse

problema pode ser minimizado pela utilização de restrições, conforme será explicado adiante.

Ferramentas de resolução de curvas “model-free”, ou seja, que não dependem de modelo matemático ou dados de referência, existem há muitos anos<sup>87</sup>, mas foram estendidas para aplicação em dados de 3 dimensões somente no início dos anos 90<sup>75,76</sup>. Essa extensão permite contornar o problema de ambiguidades rotacionais e de intensidade nas soluções encontradas por ferramentas tradicionais<sup>75</sup>. Para tanto, as matrizes precisam ser agrupadas de forma conveniente em uma *matriz aumentada*. A Figura 13 mostra, esquematicamente, como é a deconvolução em um modelo MCR-ALS utilizando duas formas de agrupamento de matrizes. O agrupamento em coluna (“column-wise”) é indicado quando são realizados diferentes experimentos (ou experimentos em condições diferentes) utilizando o mesmo instrumento analítico. Já o agrupamento em linha (“row-wise”) é utilizado para realização ou acompanhamento do mesmo experimento/fenômeno em instrumentos diferentes. Há outras formas de agrupamento que fogem do escopo deste trabalho.

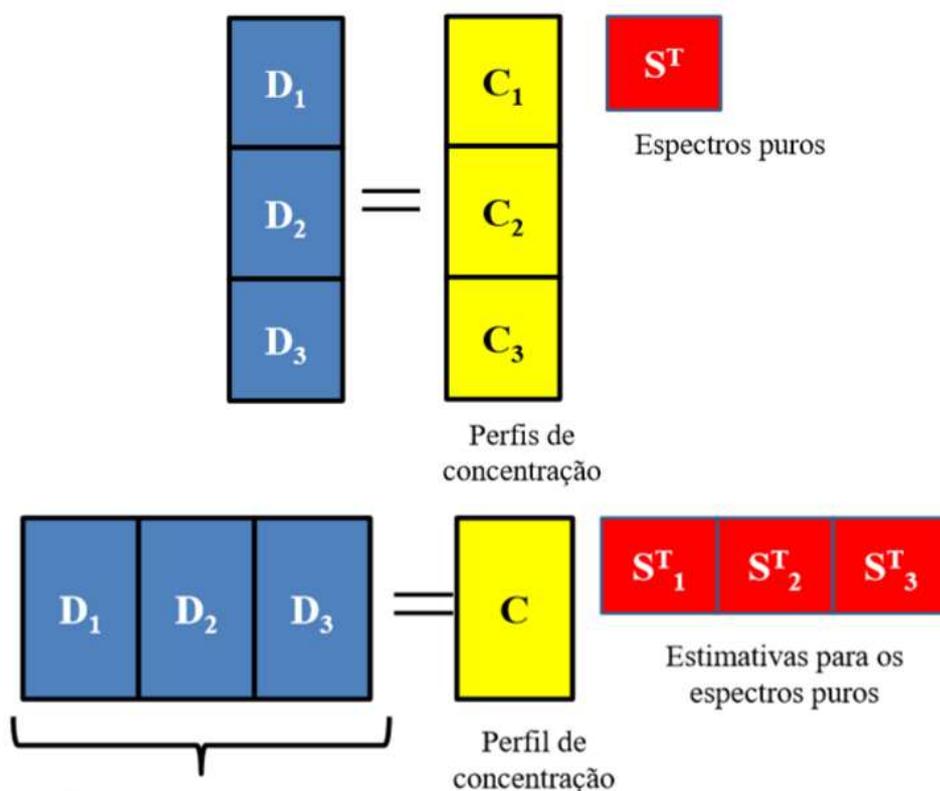


Figura 13: Aplicação de MCR-ALS em conjuntos de dados de três dimensões mostrando o agrupamento de matrizes em colunas (alto) e linhas (abaixo)<sup>86</sup>.

Nos casos em que não existe uma estrutura trilinear completa, como quando apenas a dimensão espectral é comum às camadas, a recuperação dos perfis é mais difícil e depende do grau de complexidade, da seletividade dos dados e das restrições aplicadas durante a otimização<sup>76</sup>. Tauler, entretanto, demonstrou que o uso de MCR-ALS em dados cromatográficos consegue recuperar espectros e perfis puros de forma adequada mesmo quando há desvios de trilinearidade por deslocamento de tempos de retenção ou alteração da forma dos picos entre corridas<sup>76</sup>.

Vale mencionar que MCR, assim como PARAFAC, oferece a vantagem de segunda ordem quando utilizado em análises quantitativas<sup>88</sup>.

#### 2.7.1.4.1. Ajuste do modelo MCR-ALS

A Figura 14 mostra as etapas para o ajuste de um modelo MCR-ALS. A matriz de dados **D** – que pode ser uma matriz aumentada – é decomposta nas matrizes **C** e **S** a partir de um número de componentes **F**. O uso de restrições, como mencionado anteriormente, busca minimizar o problema de liberdade rotacional de MCR-ALS impondo condições de contorno que forneçam um modelo com sentido físico para o sistema em estudo.

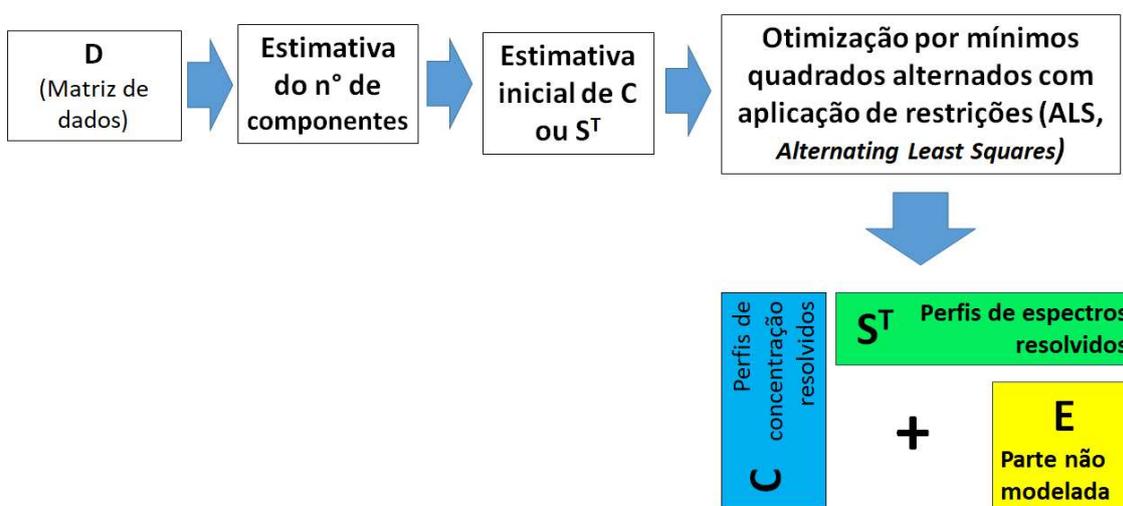


Figura 14: Obtenção de um modelo MCR-ALS<sup>86</sup>.

#### 2.7.1.4.2. Estimativa do número de componentes

O número de componentes **F** do modelo pode ser obtido com base no conhecimento do sistema. Assim, o número de espécies químicas cujos perfis estão misturados na matriz **D** pode ser utilizado como estimativa (talvez inicial) de **F**.

Um parâmetro importante na definição do número de componentes é a falta de ajuste definida, neste contexto, como a diferença entre a matriz **D** e os dados reproduzidos por  $\mathbf{CS}^T$ , o que é calculado conforme Equação 4<sup>89</sup>:

$$\text{Falta de ajuste (\%)} = 100 \sqrt{\frac{\sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2}} \quad (4)$$

onde

$e_{ij}$  são os elementos da matriz de erros, e

$d_{ij}$  são os elementos da matriz **D** com os dados experimentais.

Outro parâmetro relevante é a variância explicada calculada conforme equação 5:

$$R^2 = \frac{\sum_{i,j} d_{ij}^2 - \sum_{i,j} e_{ij}^2}{\sum_{i,j} d_{ij}^2} \quad (5)$$

Outra abordagem para estimativa de  $F^{89}$  é o uso de Decomposição por Valores Singulares (usualmente, utiliza-se a sigla em inglês SVD) que permite estimar quantas fontes independentes de variabilidade existem em **D**. Nesse método de decomposição, a matriz de **X** é reescrita como o produto da equação 6.

$$\mathbf{X} = \mathbf{U} \mathbf{S} \mathbf{V}^T \quad (6)$$

As colunas de **U** geram uma base ortonormal do espaço-linha de **X** e as colunas de **V** geram uma base ortonormal do espaço-coluna de **X**. Os elementos da diagonal da matriz **S** são chamados de autovalores e podem ser usados para estimar a dimensionalidade intrínseca de **X**<sup>90</sup>. Como esses elementos indicam a proporção de variabilidade capturada pelo modelo da equação 6, e como eles são obtidos em ordem decrescente de importância, pode-se utilizar os valores  $s_{ij}$  para decidir a real dimensionalidade dos dados em **X**.

Por fim, não pode ser descartada a opção da seleção de número de componentes por tentativa e erro. Nesse caso, para cada valor de F testado, o modelo obtido precisa ser avaliado criticamente.

#### 2.7.1.4.3. Obtenção das estimativas iniciais de **C** e **S**

Para que o algoritmo iterativo possa ser iniciado, é necessário fornecer uma matriz inicial de **C** ou **S**. É possível inserir perfis (de concentração ou espectral) de uma ou mais

espécies quando houver conhecimento a priori de que essa(s) espécie(s) está(ão) presente(s).

Outra abordagem é a estimativa numérica de **C** ou **S**. Para modelos que buscam modelar dados evolutivos<sup>91</sup>, pode ser aplicada a ferramenta de Análise Evolucionária de Fatores (ou EFA, do inglês Evolving Factor Analysis)<sup>87</sup>. Neste trabalho, essa opção não foi considerada, sendo utilizado o método de *pureza de variáveis* baseado em SIMPLISMA (do inglês, *Simple-to-Use Interactive Self-Modeling Mixture Analysis* ou Resolução de curvas por auto-modelagem interativa)<sup>92</sup>.

O método SIMPLISMA será explicado com o exemplo utilizado no artigo original de Windig e Guilment<sup>92</sup>: considere um conjunto de 13 misturas de 3 componentes A, B e C. Como o somatório das concentrações é igual a 1, a representação de cada mistura no gráfico da Figura 15 (esquerda) fica em um plano. Cada espectro dessas misturas é composto por centenas de variáveis (e.g. *m/z* para MS). Se as três componentes contribuem para os espectros, cada variável será uma combinação (possivelmente) linear das componentes. Como consequência, cada variável estará contida pelos três eixos como representado na Figura 15 (esquerda). Se as variáveis são representadas por vetores cujos comprimentos são limitados pelo plano triangular, a projeção dessas variáveis resultará na projeção (ou componente) da Figura 15. Essa projeção, por sinal, é feita com a ferramenta de Análise de Componentes Principais<sup>77</sup>. Uma variável que tem contribuição de apenas umas das componentes, chamada de variável pura, vai coincidir com o eixo da componente. A primeira componente pura pode ser encontrada determinando o vetor com o maior comprimento na projeção. Como os comprimentos dos vetores de projeção são determinados pela pureza das componentes, é possível plotar um espectro que representa a pureza usando os comprimentos desses vetores. Uma variável com uma intensidade relativamente alta será relativamente pura; uma variável com uma baixa intensidade terá contribuições de várias componentes. A representação dos valores de pureza na forma de um espectro é a base para da abordagem SIMPLISMA.

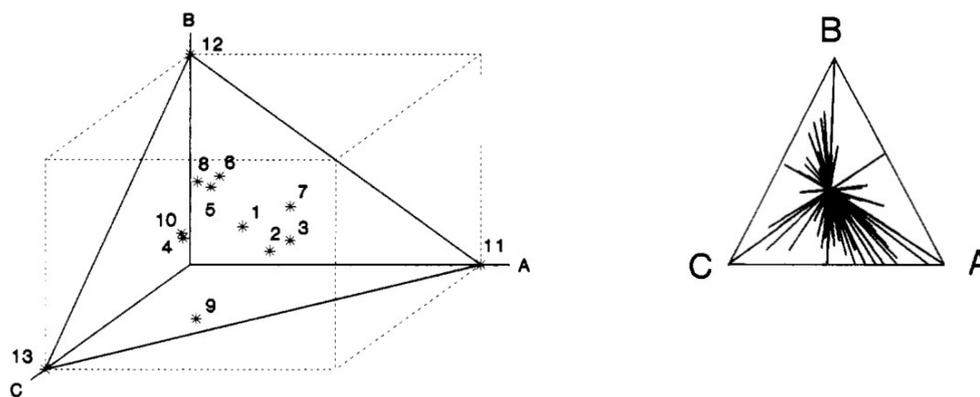


Figura 15: Representação da ferramenta SIMPLISMA. O gráfico à esquerda representa a composição de 13 misturas dos componentes A, B e C. À direita, o gráfico representa a projeção de vetores representando, por exemplo, uma relação m/z de um espectro de massas<sup>92</sup>.

#### 2.7.1.4.4. ALS e uso de restrições

Um algoritmo para estimativa de **C** e **S** a partir de **D** será esquematizado para a situação na qual a estimativa inicial é feita para **S**.

1- Estime **S** com um determinado número de componentes

2- Obtenha **C**, por mínimos quadrados

$$\mathbf{C} = \mathbf{D} \mathbf{S} (\mathbf{S}^T \mathbf{S})^{-1}$$

3- Obtenha **S** por mínimos quadrados

$$\mathbf{S}^T = \mathbf{D} \mathbf{C} (\mathbf{C}^T \mathbf{C})^{-1}$$

4- Repita os passos 2 e 3 até convergência (ou critério de parada).

A convergência é atingida quando duas iterações consecutivas apresentam diferenças relativas de desvio padrão dos resíduos (entre o modelo calculado e os dados experimentais) menor que um valor pré-selecionado; usualmente, 0,1%<sup>89</sup>.

As restrições são condições de contorno impostas ao algoritmo ALS que podem ser aplicadas na matriz **C** e/ou **S**. Algumas restrições aqui descritas são comuns à ferramenta PARAFAC, mas outros aspectos serão abordados:

➤ Não negatividade:

Consiste em ‘forçar’ as variáveis a assumir apenas valores iguais ou maiores que zero. Isso é mandatário, por exemplo, quando o algoritmo é

aplicado para resolver dados espectrais, já que sinais negativos não têm sentido físico.

Essa restrição pode ser aplicada simplesmente impondo que valores negativos sejam substituídos por zero (“*force-to-zero*”) ou aplicando abordagens baseadas em mínimos quadrados<sup>89</sup>. As ferramentas NNLS (do inglês, *Non-negativity-constrained linear least squares*, ou Restrição de não negatividade por mínimos quadrados) e sua extensão, FNNLS (onde a letra F acrescenta a palavra *Fast* ao acrônimo) são exemplos de algoritmos que substituem valores negativos nas componentes por valores estimados com base em mínimos quadrados. O detalhamento destes algoritmos está fora do escopo deste trabalho e pode ser encontrado na literatura<sup>93</sup>, mas cabe mencionar que, de acordo com os proponentes dessas ferramentas, simplesmente substituir valores negativos por zeros pode gerar erros de estimativas de vetores de regressão e levar algoritmos (de modelos multidimensionais) a divergir, i.e. iterações sucessivas levam a modelos que descrevem os dados de forma progressivamente pior.

➤ Unimodalidade<sup>94</sup>:

Essa restrição não permite que **C** ou **S** apresente mais de um máximo (ou mínimo). Essa restrição é aplicável, por exemplo, quando o modelo é usado para resolver dados cromatográficos: uma determinada componente do modelo deveria apresentar apenas um máximo, já que o cromatograma de uma única substância não deveria apresentar mais que um máximo. Há diferentes algoritmos para aplicação da unimodalidade. O leitor é convidado a buscar na literatura citada a descrição e a comparação do desempenho desses algoritmos.

➤ Balanço de massas:

Nessa restrição, a soma dos componentes (esperados) é igual a uma constante (100%, por exemplo) ou a um valor conhecido<sup>89</sup>. Essa abordagem se aplica quando, por exemplo, a matriz **D** contém perfis de *todos* os componentes de um sistema reacional.

#### 2.7.1.4.5. *Identificação de compostos com base em banco de dados de espectros de massas*

Neste trabalho, especificamente no subitem 5.6, os espectros recuperados por MCR-ALS foram comparados com o banco de dados Massbank<sup>®95</sup>, um site que agrega mais de onze mil espectros de massas obtidos por ionização eletrônica e adota um algoritmo que calcula a similaridade entre os espectros obtidos e aqueles pertencentes ao banco de dados com base em uma correlação de cossenos conforme metodologia proposta por Stein e Scott<sup>96</sup> e cujos parâmetros foram aprimorados por Horai et al<sup>95</sup>.

Os espectros recuperados também foram avaliados visualmente comparando-se os picos característicos das principais classes de compostos orgânicos dos espectros recuperados com os da literatura. Essas comparações foram feitas com base, principalmente, em Silverstein e Webster<sup>62</sup>, Kitson, Larsen e McEwen<sup>69</sup> e Settle<sup>67</sup>.

#### 2.7.2. *Calibração multivariada*

Apenas os principais aspectos teóricos e práticos da calibração multivariada serão apresentados aqui. Caso o leitor deseje se aprofundar no assunto, recomenda-se a leitura das publicações que tratam em detalhes esse tema<sup>77, 81, 90, 97, 98, 99</sup>.

##### 2.7.2.1. *Regressão linear em notação matricial e geométrica*

Considere que a Equação 7 é um modelo de regressão linear ordinária de um sistema analítico<sup>77,81</sup>.

$$y = b_0 + b_1x + e \quad (7)$$

onde

y = concentração

x = absorvância

b<sub>0</sub> = intercepto

b<sub>1</sub> = coeficiente de regressão

e = erro aleatório

A Equação 7 corresponde ao chamado modelo *inverso* de calibração, no qual as concentrações estão no eixo das ordenadas e as respostas instrumentais estão no eixo das abscissas. Isso é o oposto do método *clássico* de calibração que assume o formato da lei

de Beer-Lambert. Ocorre que, no método de regressão por mínimos quadrados, assume-se que não existe erro experimental (ou ele é insignificante) nos dados do eixo das abscissas. Na maioria dos casos, entretanto, os maiores erros estão na concentração ou propriedade de interesse e não na medida do sinal analítico.

É possível obter a solução matricial dos coeficientes em notação matricial, nesse caso, o modelo passa a ser representado conforme Equação 8.

$$y = \mathbf{C}b + e \quad (8)$$

onde  $y$  e  $b$  são os vetores de concentrações e de coeficientes de regressão, respectivamente e  $\mathbf{C}$  é a matriz de absorbâncias. Nesse caso, uma das colunas de  $\mathbf{C}$  é um vetor composto exclusivamente de números um (para um modelo com termo constante).

A solução para  $b$  pode ser obtida a partir da pseudo-inversa de Moore-Penrose conforme Equação 9.

*Nota: como será visto adiante, essa solução é válida tanto em sistemas univariados quanto multivariados.*

$$\hat{b} = (\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T y \quad (9)$$

A mesma solução pode ser entendida do ponto de vista geométrico. Considere que a matriz  $\mathbf{C}$  representa um espaço-coluna cujas dimensões correspondem ao número de colunas. Assim, se  $\mathbf{C}$  tem duas colunas, esse espaço pode ser representado por um plano conforme Figura 16. O que se busca na regressão linear é projetar o vetor  $y$  do espaço gerado por  $\mathbf{C}$ , o que é feito por meio de uma projeção ortogonal. O vetor de coeficientes de regressão corresponde às coordenadas no espaço-coluna dessa projeção (ver Figura 16).  $e$ , por outro lado, é o vetor de resíduos cujo valor é a distância ortogonal de  $y$  até sua projeção no espaço  $\mathbf{C}$ .

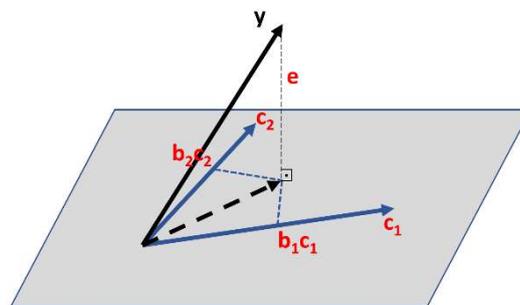


Figura 16: Regressão linear representada como projeção do vetor  $y$  no espaço delimitado pela matriz  $\mathbf{C}$  de duas colunas<sup>77</sup>.

Apesar de muito utilizados, os métodos univariados de calibração utilizados em química são limitados, pois assumem que a resposta medida é influenciada unicamente pelo constituinte de interesse ou que, no caso de haver algum interferente presente, que sua contribuição seja um acréscimo constante na resposta<sup>77</sup>. Na calibração multivariada, a quantidade física ou química (resposta) é obtida como uma função de muitas grandezas medidas (preditores)<sup>99</sup>. Calibrações multivariadas utilizam preditores não específicos, geralmente informação física de espectros, especialmente no infravermelho. Mas podem ser utilizados espectros UV, visível, Raman, de fluorescência, de RMN e de massas, além de preditores de químicos e eletroquímicos<sup>99</sup>.

#### 2.7.2.2. Seleção do conjunto de amostras de calibração

A despeito das diferenças entre as ferramentas de calibração, algumas diretrizes são válidas para qualquer calibração multivariada no que se refere à seleção de amostras para compor o *conjunto de calibração*<sup>97,99</sup>:

- a. As amostras devem incluir todos os compostos esperados nas amostras a serem analisadas, o que permite modelar efeitos de interferência – esse aspecto não é essencial em todos os modelos de calibração de segunda ordem que podem se beneficiar da chamada *vantagem de segunda ordem*<sup>81</sup>.
- b. A faixa de concentração (ou propriedade de interesse) deve ser suficientemente abrangente para cobrir todas as amostras esperadas no uso do modelo de calibração.
- c. As concentrações (ou propriedade de interesse) devem estar preferencialmente uniformemente distribuídas dentre as amostras de calibração.
- d. O número de amostras deve ser suficiente para, estatisticamente, definir a relação entre os espectros e a concentração (ou propriedade de interesse).

A norma ASTM E1655<sup>97</sup> apresenta procedimentos específicos para aplicações de calibração multivariada com espectrometria no infravermelho, dentre eles:

- a. A faixa de concentração deve ser, preferencialmente, cinco vezes, mas não menos que três vezes o desvio padrão de reprodutibilidade ( $\text{Reprodutibilidade}/2,77$ ) do método de referência.
- b. Na obtenção dos espectros para calibração, variações experimentais ou de preparo de amostras esperadas no uso do método devem ser reproduzidas;

- c. Se um modelo multivariado é desenvolvido com três ou menos variáveis latentes, então o conjunto de calibração deve conter, no mínimo, 24 amostras após a eliminação de outliers.
- d. Se o modelo utiliza mais do que três variáveis, então o conjunto de calibração deve conter, no mínimo, seis vezes o número de variáveis.

Para novas aplicações, a norma orienta que seja feito um estudo de exequibilidade de calibração, que deveria contar com aproximadamente, 30 a 50 amostras cobrindo toda a faixa de propriedade/concentração.

### 2.7.2.3. *Pré-tratamento de dados*

Antes da aplicação de ferramentas quimiométricas, é usualmente necessário pré-tratar os dados. O objetivo do pré-tratamento é reduzir variações indesejáveis que não foram removidas durante a aquisição dos dados e que não serão eliminadas naturalmente durante a análise, mas que podem influenciar os resultados<sup>77</sup>. Quando bem aplicadas, as ferramentas de pré-tratamento de dados garantem que as variações observadas nos dados são de origem química, e não de ruído instrumental<sup>100</sup>. Os pré-tratamentos podem ser divididos em duas classes<sup>77</sup>: a *transformação* se refere aos pré-tratamentos aplicados às amostras (o que, usualmente, corresponde às linhas da matriz de dados); e o *pré-processamento* se refere ao pré-tratamento aplicado às variáveis (colunas da matriz de dados).

A *seleção de variáveis* engloba uma série de estratégias de transformação de dados e tem por objetivo selecionar apenas aquelas variáveis (e.g.: comprimentos de ondas) relevantes para o problema, eliminando ruído e/ou sinais relacionados a interferentes. A seleção de variáveis pode consistir em algo tão simples quanto selecionar e descartar regiões do espectro com base em inspeção visual ou no conhecimento do sistema<sup>90</sup>.

As abordagens para seleção de variáveis são tão abundantes que Mehmood, et al.<sup>101</sup>, dividiram os métodos de seleção de variáveis para aplicação de PLS em diferentes categorias: métodos de filtragem, empacotados e embutidos. Cada abordagem para seleção de variáveis tem suas vantagens e desvantagens e, muitas vezes, a escolha é feita com base em tentativa e erro.

Neste trabalho, utilizou-se o algoritmo de Importância de variável por projeção, VIP (do inglês, Variable importance in projection)<sup>101</sup>. VIP busca medir a importância

acumulada de cada variável original  $j$  a partir de cada vetor  $w$  de variáveis do modelo. A importância de cada variável é dada por  $v_j$  definida na Equação 10:

$$v_j = \sqrt{p \sum_{a=1}^A [SS_a \left( \frac{w_{aj}}{\|w_a\|^2} \right)] / \sum_{a=1}^A (SS_a)} \quad (10)$$

onde  $SS_a$  é a soma dos quadrados explicados pela  $a$ -ésima componente. Portanto,  $v_j$  mede a contribuição de cada variável a partir da sua variância explicada e  $\frac{w_{aj}}{\|w_a\|^2}$  representa a importância da  $j$ -ésima variável. O usuário deve decidir o critério de corte para as variáveis (e.g.: percentual de variáveis a serem descartadas) mantendo-se, é claro, apenas aquelas com os maiores valores de  $v_j$ .

Uma transformação muito utilizada em espectroscopia é o uso de *derivadas*. Deslocamentos da linha de base (causados por problemas instrumentais, por exemplo) são eliminados se as absorbâncias são substituídas pela primeira ou segunda derivada dos espectros. Uma das formas de calcular as derivadas utiliza o método Savitzky-Golay que consiste em ajustar, por mínimos quadrados, um polinômio de grau  $n$  a uma “janela” (sequência) de pontos do espectro<sup>77</sup>. Essa janela “avança” ao longo de todo o espectro substituindo os valores de absorbância da janela pelos valores dados pelo polinômio, o que *suaviza* ou *alisa* o espectro. Como o espectro passa a ser descrito por polinômios, as derivadas podem ser obtidas diretamente desses polinômios.

Outra transformação importante em espectroscopia é a *Padronização normal de sinal* (a sigla usual é SNV; do inglês *Standard Normal Variate*). Ela permite corrigir efeitos aditivos e multiplicativos causados por fenômenos físicos, especialmente, aqueles oriundos de interferências de espalhamento de luz causado pelo tamanho da partícula sólida. SNV corresponde a autoescalar cada linha da matriz original dos dados (ver Equação 11)<sup>77</sup>.

$$\begin{bmatrix} \vdots \\ x_{i\ SNV} \\ \vdots \end{bmatrix} = \frac{1}{s_i} \left( \begin{bmatrix} \vdots \\ x_i \\ \vdots \end{bmatrix} - \begin{bmatrix} \bar{x}_i \\ \vdots \\ \bar{x}_i \end{bmatrix} \right) \quad (11)$$

$$\bar{x}_i = \frac{1}{J} \sum_{j=1}^J x_{ij} \text{ e } s_i = \sqrt{\sum_{j=1}^J (x_{ij} - \bar{x}_i)^2}$$

A *centralização na média* é um procedimento muito comum de pré-processamento de dados e consiste em subtrair cada valor da matriz de dados pela média dos valores da variável (coluna) correspondente, conforme Equação 12.

$$x_{ij(\text{cm})} = x_{ij} - \bar{x}_j \quad (12)$$

onde  $\bar{x}_j = \frac{1}{I} \sum_{i=1}^I x_{ij}$

Esse procedimento apenas translada os eixos das variáveis sem afetar a estrutura dos dados<sup>77</sup>. Apesar de simples, a centralização pode ser indispensável na construção do modelo de calibração, já que remove um termo constante no conjunto de dados que poderia “ocupar” a primeira componente (ou variável latente) do modelo ou pode até influenciar na direção segunda componente<sup>78</sup>.

É importante alertar que, quando uma amostra desconhecida é analisada com um modelo multivariado, a média dos espectros de calibração deve ser subtraída do espectro da amostra desconhecida e a média dos valores de referência deve ser adicionada ao resultado obtido por interpolação no modelo.

#### 2.7.2.4. *Identificação de amostras anômalas*

Há dois tipos de amostras anômalas que podem ser identificadas durante a calibração<sup>97</sup>. O primeiro tipo é o de amostras que apresentam composição ou propriedade extrema em relação ao restante do conjunto de calibração. Essa amostra tem grande influência em, pelo menos, um dos coeficientes de regressão e sua inclusão no conjunto de calibração pode levar a um modelo inadequado se seu valor de referência for (mesmo que levemente) inexato. O segundo tipo de amostra anômala é aquele para o qual o valor estimado pelo modelo é significativamente diferente do valor de referência. Essa diferença pode advir de erro na obtenção do valor de referência, do espectro dessa amostra ou uma falha do modelo (sem mencionar erros de transcrição de dados).

Uma ferramenta poderosa para identificação de amostras anômalas é o parâmetro chamado de alavancagem (*leverage*) que será explicado a seguir.

Retomando a equação 8, podemos obter o vetor  $y$  projetado no espaço coluna  $C$  com a Equação 13.

$$\hat{\mathbf{y}} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T\mathbf{y} \quad (13)$$

A matriz  $\mathbf{H}$  é definida exatamente como a parte da equação 13 que faz essa projeção (Equação 14).

$$\mathbf{H} = \mathbf{C}(\mathbf{C}^T\mathbf{C})^{-1}\mathbf{C}^T \quad (14)$$

Os elementos da diagonal de  $\mathbf{H}$  ( $h_{ii}$ ) são chamados de alavancagem (em inglês, leverage) e medem a influência da amostra  $i$  no modelo de regressão linear, ou, mais especificamente, na estimativa dos coeficientes de regressão. A matriz  $\mathbf{C}$  utilizada, é claro, dependerá do modelo de calibração aplicado.

Do ponto de vista geométrico, a alavancagem é uma medida da distância da amostra ao centroide do conjunto de dados, no subespaço definido pelos  $A$  fatores do modelo. Amostras distantes do centro têm maior influência no modelo do que amostras próximas dele.

O valor médio de alavancagem,  $h_{med}$ , é dado pela Equação 15. Esse parâmetro pode ser utilizado para estabelecer um teste de significância para distinguir amostras anômalas: amostras com alavancagem maior que duas ou três vezes a alavancagem média (dependendo do nível de confiança desejado) são consideradas suspeitas e devem ser avaliadas<sup>77</sup>.

$$h_{med} = \frac{\sum_{i=1}^I h_{ii}}{I} = \frac{A}{I} \quad (15)$$

Já a estatística Hotelling,  $T^2$ , representa uma medida da variabilidade de cada amostra dentro do modelo<sup>102</sup>. Ela também indica quão longe uma amostra está do centro do modelo. Para um modelo PCA (o raciocínio é análogo para outros modelos), a estatística Hotelling  $T^2$  é a soma quadrática dos escores normalizados (Equação 16):

$$T_i^2 = \mathbf{t}_i\mathbf{\Lambda}^{-1}\mathbf{t}_i^T \quad (16)$$

onde  $t_i$  se refere à  $i$ -ésima linha da matriz de escores e  $\mathbf{\Lambda}$  é a matriz diagonal contendo os autovalores das componentes do modelo PCA.

Como  $T^2$  segue uma distribuição de Fischer, é possível definir um limite de confiança com base num valor de  $F$  de Fischer tabelado com um nível de significância escolhido (usualmente, 95%).

*Nota: para modelos nos quais os pesos não são ortogonais,  $\lambda$  precisa ser estimada, o que faz com que  $T^2$  não seja tão exata e útil<sup>102</sup>.*

O parâmetro  $Q$  (ou  $Q$  residuals) é uma estatística de falta de ajuste calculada a partir da soma dos quadrados de cada linha da matriz de erros  $E^{102}$  (Equação 17):

$$Q = \mathbf{e}_i^T \mathbf{e}_i \quad (17)$$

onde  $\mathbf{e}_i$  é o vetor de erros da amostra  $i$ .

O parâmetro  $Q$  indica o quão bem cada amostra se adequa ao modelo. É uma medida da diferença, ou resíduo, entre a amostra e sua projeção nos  $k$  fatores utilizados no modelo<sup>102</sup>.

#### 2.7.2.5. Métodos de regressão multivariada

A seguir, será feita uma breve explanação sobre as ferramentas quimiométricas de calibração multivariada utilizadas neste trabalho: Regressão por Mínimos Quadrados Parciais, PLS e Regressão por Vetores de Regressão, SVR.

É importante frisar que, qualquer que seja a ferramenta de calibração escolhida, é de suma importância a confiabilidade dos dados de referência (dados utilizados para gerar o modelo de calibração), já que a exatidão do modelo é altamente dependente da exatidão e precisão dos valores de referência<sup>97</sup>.

##### 2.7.2.5.1. Regressão por Mínimos Quadrados Parciais – PLS

A regressão em mínimos quadrados parciais, PLS ou PLSR, busca aproximar o espaço das medidas originais por um de dimensão reduzida, com a imposição de restrição na decomposição da matriz  $\mathbf{X}$  direcionando-a para uma solução cujo alvo é a propriedade de interesse (expressa em  $\mathbf{y}$ ). PLS gera um sistema de variáveis com um compromisso de ‘explicação’ da variância em  $\mathbf{X}$  e  $\mathbf{y}$  simultaneamente; essas variáveis são intituladas *latentes*<sup>77</sup>.

A abordagem PLS foi proposta por Herman Wold em 1975 para modelagem de conjuntos de dados complexos em termos de cadeias de matrizes<sup>103</sup>. O estudo incluiu um algoritmo NIPALS para estimativa dos parâmetros do modelo. O acrônimo PLS se refere à parte central do algoritmo. Mais especificamente, cada parâmetro do modelo PLS é estimado iterativamente como a inclinação de uma regressão (por mínimos quadrados) entre a matriz coluna da variável  $y$  e outro parâmetro vetor como variável  $x$ .

Há duas variantes de PLS, denominadas PLS1 e PLS2. Na primeira, calculam-se as variáveis latentes para um único vetor  $\mathbf{y}$  (e.g.: a concentração de *um* analito). Em PLS2, o conjunto de variáveis latentes é modelado para todas as colunas de  $\mathbf{Y}$  simultaneamente (e.g.: uma matriz com diferentes propriedades de interesse em cada coluna). Na maioria das situações, obtém-se modelos melhores se as variáveis latentes são otimizadas para cada vetor  $\mathbf{y}$ <sup>101</sup>. Entretanto, quando há correlação entre as variáveis (colunas) de  $\mathbf{Y}$ , PLS2 pode fornecer um modelo superior. Neste trabalho, apenas PLS1 será utilizado e, como é usual na literatura, será referenciado, daqui em diante, simplesmente por PLS.

A seguir, apresenta-se um algoritmo NIPALS para construção de modelo de regressão PLS<sup>77</sup>. O pré-processamento da matriz  $\mathbf{X}$  e do vetor  $\mathbf{y}$  foi desconsiderado para facilitar o entendimento.

Cálculo da primeira variável latente:

Bloco X (variáveis preditoras)

$$\mathbf{X} = \mathbf{y}\mathbf{w}_1^T + \mathbf{E}_x \Rightarrow \hat{\mathbf{w}}_1 = \mathbf{X}^T\mathbf{y}(\mathbf{y}^T\mathbf{y})^{-1} \quad (\text{cálculo do fator-peso } \mathbf{w})$$

$$\hat{\mathbf{w}}_1 = \hat{\mathbf{w}}_1 / \sqrt{\hat{\mathbf{w}}_1^T \hat{\mathbf{w}}_1} \quad (\text{normalização de } \mathbf{w}_1)$$

$$\mathbf{t}_1 = \mathbf{X}\hat{\mathbf{w}}_1 \quad (\text{cálculo dos escores})$$

$$\mathbf{X}^T = \mathbf{I}_1\mathbf{t}_1^T \Rightarrow \hat{\mathbf{I}}_1 = \mathbf{X}^T\mathbf{t}_1(\mathbf{t}_1^T\mathbf{t}_1)^{-1} \quad (\text{cálculo dos loadings})$$

Bloco y (variáveis dependentes)

$$\mathbf{y} = \mathbf{t}_1\mathbf{q}_1 + \mathbf{f} \Rightarrow \hat{\mathbf{q}}_1 = (\mathbf{t}_1^T\mathbf{t}_1)^{-1}\mathbf{t}_1^T\mathbf{y} \quad (\text{regressão do vetor } \mathbf{y} \text{ no vetor } \mathbf{t}_1)$$

Atualização de  $\mathbf{X}$  e  $\mathbf{y}$

$$\mathbf{E} = \mathbf{X} - \mathbf{t}_1\hat{\mathbf{I}}_1^T$$

$$\mathbf{f} = \mathbf{y} - \mathbf{t}_1\hat{\mathbf{q}}_1$$

$$\mathbf{X} = \mathbf{E}$$

$$\mathbf{y} = \mathbf{f}$$

[As próximas variáveis latentes são calculadas com o mesmo algoritmo, utilizando  $\mathbf{X}$  e  $\mathbf{y}$  atualizados.]

Após o cálculo das variáveis latentes (quantas forem necessárias), calcula-se o vetor de regressão  $\mathbf{b}$ :

$$\hat{\mathbf{b}} = \mathbf{W}(\mathbf{L}^T\mathbf{W})^{-1}\mathbf{q}$$

Do ponto de vista geométrico, PLS pode ser interpretado como uma projeção da matriz de dados  $\mathbf{X}$  num hiperplano de  $A$ -dimensões de tal forma que as coordenadas dessa projeção ( $\mathbf{t}_a$ ,  $a = 1, 2, \dots, A$ ) são boas preditoras de  $\mathbf{y}$ <sup>103</sup>. A direção desse plano é determinada pelas inclinações,  $\mathbf{l}_{aj}$ , de cada variável latente em relação a cada eixo de coordenada  $\mathbf{x}_j$ . Essa inclinação, por sinal, é o cosseno do ângulo entre as variáveis latentes e os eixos  $\mathbf{x}_j$ .

Em tese, PLS deveria produzir os melhores resultados dentre as ferramentas de regressão linear multivariada, pois as variáveis latentes são definidas para incorporar a informação da propriedade de interesse na construção do modelo. Isso, entretanto, pode depender da escolha adequada das variáveis inseridas modelo, da escolha do número de variáveis latentes e da seleção de um conjunto de amostras de calibração representativo, para citar apenas alguns fatores<sup>77</sup>.

#### 2.7.2.5.2. Regressão por Vetores de Suporte – SVR

A Máquina de Vetores de Suporte, SVM (do inglês, *Support Vector Machine*) é um método de aprendizagem de máquina (*machine learning*) desenvolvido por Cortes e Vapnik<sup>104</sup> em 1995 para problemas de classificação binária. O algoritmo de Vetores de Suporte é uma generalização não linear do algoritmo *Generalized Portrait* desenvolvido na Rússia nos anos 60<sup>105</sup>. Apesar da SVM ter sido desenvolvida para resolver problemas de classificação, a ferramenta pode ser utilizada para problemas de regressão conforme será explicado adiante.

SVM permite modelar sistemas com comportamento não linear fornecendo modelos robustos<sup>106</sup>. Enquanto modelos lineares tradicionais buscam explicar como uma variável independente afeta a variável dependente, o propósito de máquinas de aprendizado é a predição<sup>107</sup>.

##### a. Classificação com vetores de suporte

A Figura 17 mostra um exemplo de um conjunto de dados num espaço de duas variáveis e que pertencem a duas classes (classificadas como -1 e +1). A separação pode ser feita por uma linha (um hiperplano de uma dimensão) e uma das formas de encontrá-

la é dada pelo problema de otimização abaixo<sup>107</sup> no qual busca-se minimizar a *margem*  $M$  – a distância mínima entre a linha e o ponto mais próximo de cada classe classificada como  $-1$  e  $+1$ .

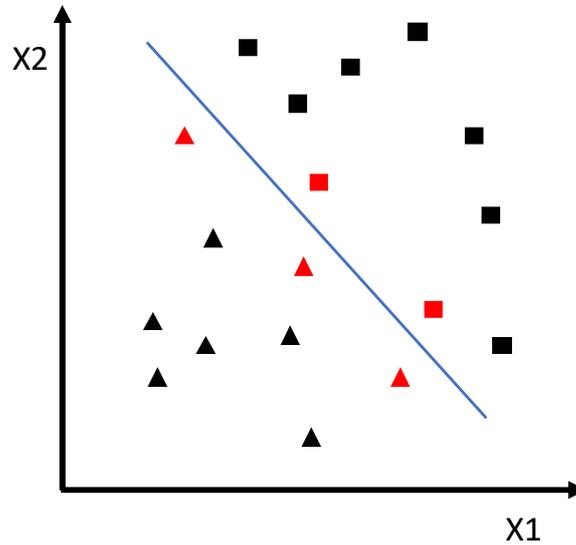


Figura 17: Separação de amostras de duas classes por uma linha. As formas vermelhas são os *vetores de suporte* conforme explicado no texto.

$$\begin{aligned} &\text{maximize } M && (18) \\ &\beta_j, j = 1, \dots, p \end{aligned}$$

$$\text{sujeito a } \begin{cases} \sum_{j=1}^p \beta_j^2 \\ y_i(\beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}) \geq M \end{cases} \quad (19)$$

onde  $M$  é a *margem*,  $p$  é o número de variáveis  $x$ , e  $y_i \in \{-1, 1\}$ ,  $i = 1, \dots, n$ . Note que a segunda linha da Equação 19 representa um hiperplano que, em um sistema de duas dimensões, corresponde à uma linha. A restrição de  $\beta_j$  na Equação 19 não é essencial, mas torna o cálculo das distâncias de qualquer observação para o hiperplano conveniente.

É importante reforçar que os vetores de suporte do modelo acima são as amostras que demarcam a fronteira entre as classes. Neste modelo, há cinco vetores de suporte (formas em vermelho na Figura 17) e a otimização das Equações 18 e 19 considera apenas as observações às margens da linha de divisão das classes, o que reduz drasticamente o tempo computacional.

Na maioria das situações, não é possível separar completamente as observações, o que torna o problema de otimização acima insolúvel. Isso é resolvido inserindo-se a *variável de folga*  $\zeta$  no algoritmo que permite que algumas observações estejam “do lado

errado” da linha de separação das classes. E, para evitar que o número de erros seja grande demais, insere-se a variável *erro total de orçamento* ou *custo*  $C$ . Os detalhes de como isso é implementado estão além dos objetivos deste texto, sendo suficiente entender que valores altos de  $C$  aumentam a penalidade por classificações erradas na elaboração (treinamento) do modelo, o que pode levar ao sobreajuste do modelo. Isso, por sua vez, pode levar a um desempenho ruim do modelo para novas observações. Por isso, a escolha de  $C$  é uma etapa importante na otimização do modelo.

Quando não é possível realizar uma separação linear das classes (mesmo com a aplicação de uma variável de folga), como ilustrado na Figura 18, é possível aplicar uma transformação dimensional nos dados por meio de uma *função kernel* não linear  $K$ . Essa função projeta os dados no *espaço de características* onde, espera-se, as classes possam ser divididas por uma função linear. A separação linear no espaço das características equivale à aplicação não linear no espaço original do dados<sup>108</sup>.

No exemplo da Figura 18, a função kernel é dada pela Equação 20:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i \cdot \mathbf{x}_j \quad (20)$$

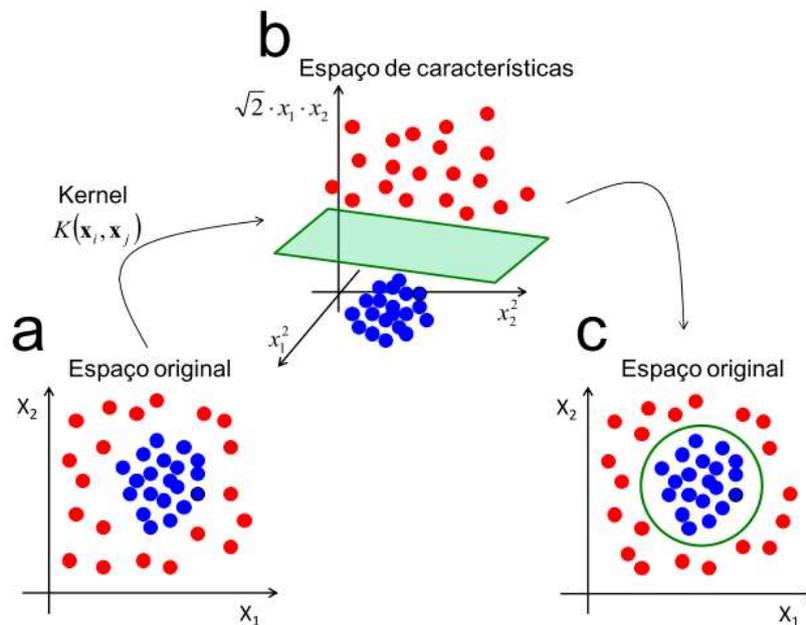


Figura 18: Projeção de observações no espaço de características por uma função kernel linear e definição do hiperplano de separação das classes. [Extraído de Filgueiras<sup>108</sup>]

O aumento de dimensão promovido pela função kernel dada na Equação 20 *pode* tornar as observações linearmente separáveis. Há, entretanto, outras funções mais eficientes. As funções kernel mais conhecidas são: linear, sigmoide e a função de base radial (RBF, do inglês, *Radial Basis Function*), que será utilizada neste estudo por ser considerada a mais flexível<sup>107</sup>, ou seja, ela se adequa a conjuntos de dados com diferentes comportamentos. A função RBF é dada por:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \cdot \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (21)$$

onde  $\gamma > 0$  é um parâmetro adicional de ajuste que regula o quão próximas as observações devem estar para contribuir para a decisão de classificação e os índices  $i$  e  $j$  indicam as distâncias entre todos os pares de observações. A função RBF é considerada uma função de propósito geral, quando não há conhecimento da estrutura dos dados.

*b. Regressão por vetores de suporte*

A utilização de vetores de suporte para regressão pode ser realizada transformando-se os dados de treinamento em dois conjuntos de dados e determinando-se o hiperplano de separação desses dados conforme ilustrado na Figura 19. Para cada amostra  $x_i$  (Figura 19-A), um número positivo  $d$  é adicionado e subtraído do correspondente valor de interesse  $y_i$  (Figura 19-B) gerando duas classes (Figura 19-C). O hiperplano de máxima separação obtido para separação dessas duas classes corresponderá à função de regressão do modelo. As diferenças dessa função de regressão para uma regressão por mínimos quadrados serão discutidas adiante.

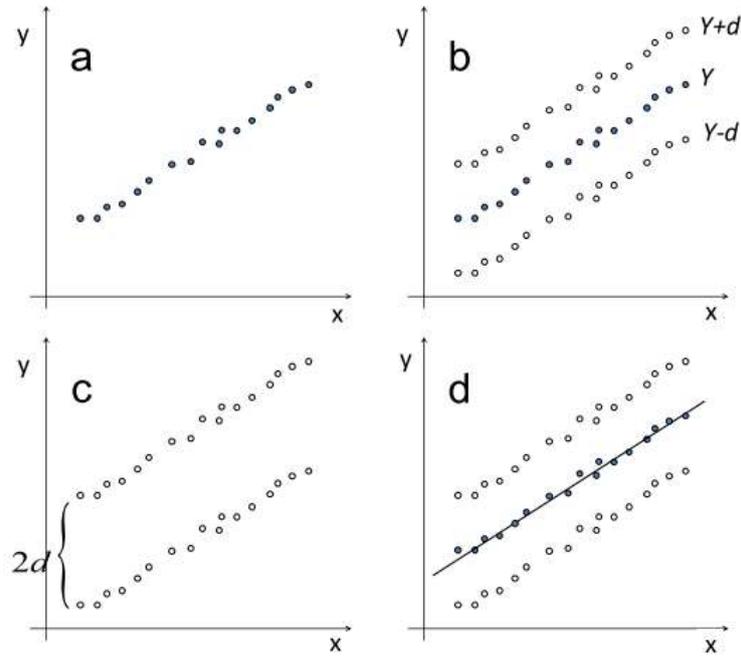


Figura 19: Utilização de vetores de suporte para regressão. [Extraído de Filgueiras<sup>108</sup>]

A seguir<sup>107</sup>, apresenta-se o problema de otimização para regressão por máquina de vetores considerando a situação (não ilustrada acima) na qual não há separação perfeita entre as classes, o que requer os usos das variáveis  $\zeta$  e  $C$  já comentadas anteriormente. A variável  $C$  regula o quão intensamente devem ser evitadas (penalizadas) observações fora da *banda*  $\varepsilon$ . As variáveis  $\xi$  e  $\xi^*$  se referem aos erros positivos e negativos, respectivamente.

$$\text{Minimize } \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \quad (22)$$

$$\text{Sujeito a } \begin{cases} y_i - x_i^t \beta - \beta_0 & \leq \varepsilon + \xi_i \\ -(y_i - x_i^t \beta - \beta_0) & \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* & \geq 0 \end{cases} \quad (23)$$

A Figura 20 ilustra três aspectos importantes da regressão por vetores de suporte<sup>107,108</sup>: *i.* todas as observações sobre e fora da *banda*  $\varepsilon$  são vetores de suporte; *ii.* a *banda*  $\varepsilon$  forma um tubo ao redor da função de regressão no qual são aceitos erros de exatidão. *iii.* enquanto na regressão por mínimos quadrados busca-se minimizar a perda

quadrática para *todas* as observações, em SVR as observações dentro da banda  $\epsilon$  são desconsideradas. A Figura 20 ilustra essa situação para uma observação.

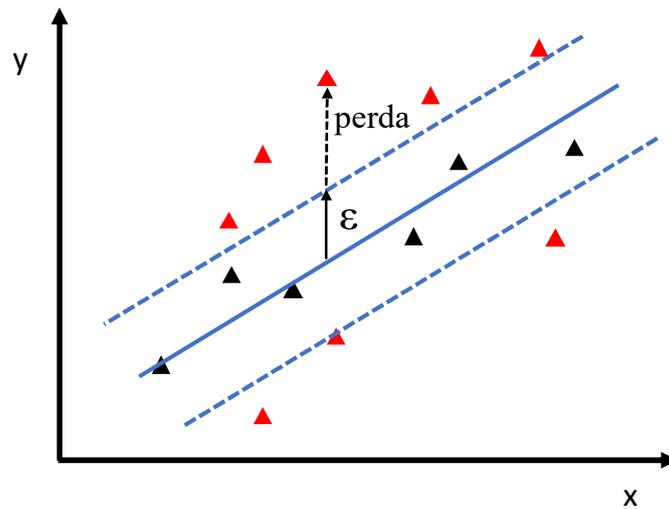


Figura 20: Regressão por vetores de suporte. As figuras em vermelho são os vetores de suporte. As linhas tracejadas correspondem à banda  $\epsilon$ . [adaptado de Guenther e Schonlau<sup>107</sup>]

Assim como discutido no item a (Classificação com vetores de suporte), as mesmas funções kernel podem ser utilizadas em conjuntos de dados complexos ou com comportamento não linear.

As equações 22 e 23 correspondem ao que se denomina atualmente como  $\epsilon$ -SVR - em função do papel de  $\epsilon$  no algoritmo. Esse algoritmo tende a levar a modelos com banda  $\epsilon$  reduzida, mas com número excessivo de vetores de suporte. Entretanto, no algoritmo  $\nu$ -SVR incorpora-se um parâmetro adicional à Equação 22 para controlar o número de vetores de suporte na Equação 24.

$$\text{Minimize } \frac{1}{2} \|\beta\|^2 + C \left[ \nu\epsilon + \frac{1}{N} \sum_{i=1}^n (\xi_i + \xi_i^*) \right] \quad (24)$$

Após a escolha de  $\nu$ , a banda  $\epsilon$  é automaticamente definida pelo algoritmo.  $\nu$  é o limite inferior de proporção entre vetores de suporte e o número total de amostras<sup>108</sup>. O algoritmo  $\nu$ -SVR foi selecionado para este projeto por ser o que tende a dar os melhores resultados de acordo com a literatura.

Os parâmetros ideais para o modelo SVR devem ser otimizados conforme será discutido adiante. Assim, numa regressão por vetores de suporte, utilizando o algoritmo  $\nu$ -SVR e função de base radial, os seguintes parâmetros devem ser otimizados:  $\nu$ ,  $C$  e  $\gamma$ .

### 2.7.2.6. Avaliação e otimização do modelo de calibração

Há várias ferramentas para avaliar e otimizar o modelo de calibração. Usualmente, elas são aplicadas exclusivamente nos dados do conjunto de calibração<sup>97</sup> e não se confundem com a etapa de validação do modelo (a ser discutida adiante). Nem todas as formas de avaliação e otimização discutidas neste item são aplicáveis a todas as ferramentas de calibração sendo necessário avaliação caso a caso.

O coeficiente de determinação é dado pela Equação 25 e tem a mesma interpretação em modelos uni ou multivariados:  $R^2$  determina a quantidade de variação nos dados que foi modelada pelo modelo de calibração. Valores de  $R^2$  próximos de um são desejáveis; valores muito baixos devem ser investigados.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i) / (n - k - 1)}{\sum_{i=1}^n (y_i - \bar{y})^2 / (n - 1)} = \frac{SS_{reg}}{SS_{tot}} \quad (25)$$

O parâmetro RMSEC (Root Mean Squared Error of Calibration, ou Raiz da média quadrática dos erros de calibração) fornece um sumário dos resíduos do modelo enquanto o RMSECV (Root Mean Squared Error of Cross Validation, ou Raiz da média quadrática dos erros de validação cruzada) resume os erros das estimativas para validação cruzada (Equações 26 e 27).

$$RMSEC = \sqrt{\frac{\sum_{i=1}^I (y_i - \hat{y}_i)^2}{I - A}} \quad (26)$$

$$RMSECV = \sqrt{\frac{\sum_{i=1}^I (y_i - \hat{y}_{cv})^2}{I}} \quad (27)$$

Onde  $y_i$  e  $\hat{y}_i$  são os valores do parâmetro de interesse da amostra e seu valor estimado, respectivamente,  $A$  é o número de variáveis latentes e  $\hat{y}_{cv}$  é o valor estimado para amostra segregada para validação cruzada. Para modelos centrados na média o denominador da equação 26 passa a ser  $I - A - 1$ .

Há muitas variantes para realização de validação cruzada. Neste trabalho, adotou-se o *Venetian Blinds* que consiste em dividir as amostras em subgrupos com  $n$  camadas, sendo cada camada formada por  $x$  amostras. Um certo número de camadas é, então, retirado do conjunto de calibração (de todos os subgrupos). Após construção do modelo de calibração e estimativa do valor da propriedade de interesse para as amostras retiradas, elas retornam ao conjunto de calibração e novas camadas são retiradas de forma que uma mesma amostra nunca é retirada duas vezes. Por fim, o parâmetro que resume o

desempenho do modelo, RMSECV, por exemplo, é calculado. Considera-se que essa ferramenta é a mais adequada para o número de amostras e a forma como elas foram selecionadas.

A implementação usual de regressão por vetores de suporte utiliza um sistema particular de otimização denominado *grade de pesquisa* ou de *parâmetros*<sup>108-110</sup>. Como descrito no item b do subitem 2.7.2.5.2, a aplicação do método v-SVR com função de base radial requer a otimização de três parâmetros ( $C$ ,  $\gamma$  e  $\nu$ ). A grade de pesquisa testa o modelo SVR por validação cruzada com todas as combinações desses parâmetros dentro de um intervalo valores e com um determinado número de valores. Maiores detalhes podem ser buscados em Filgueiras<sup>108</sup>.

#### 2.7.2.6.1. Otimização do número de variáveis no modelo

A inclusão de um número excessivo de variáveis latentes (ou fatores, dependendo da literatura utilizada) leva ao sobreajuste do modelo, o que faz com que ele apresente alta variância e seja apenas de aplicação *local*, ou seja, aplicável em amostras muito similares àquelas utilizadas no conjunto de calibração. Por outro lado, um número insuficiente de fatores leva um modelo subajustado com alta *tendência* nos resultados<sup>77</sup>.

Uma primeira avaliação do número de fatores a serem incluídos pode ser baseada nos autovalores dos modelos multivariados<sup>98</sup>. Fatores que retêm grande parte da variabilidade dos dados tendem a ser relevantes para o modelo (essa afirmativa é especialmente verdadeira para PLS).

A seguir, um gráfico de número de fatores *versus* os valores de RMSEC ou RMSECV pode ser utilizado para decidir o modelo adequado. O mínimo nesse gráfico indicará o máximo de fatores a ser empregado. Se não houver um mínimo claro, deve-se buscar um patamar inferior nos valores de RMSECV e optar pelo modelo com o menor número de componentes nesse patamar (novamente, o princípio da parcimônia).

#### 2.7.2.7. Validação e estimativa de figuras de mérito do modelo

A validação consiste na “comprovação, através do fornecimento de evidência objetiva, de que os requisitos para uma aplicação ou uso específicos pretendidos foram atendidos”<sup>111</sup>. Apesar de parecer vaga, essa definição está alinhada ao conceito de *adequação ao propósito*<sup>112</sup> de ensaios, que enfatiza a comparação das características de desempenho de um método analítico contra os requisitos da aplicação do método. Na

prática, a estimativa das figuras de mérito e a sua avaliação crítica constituem o processo de validação de um método analítico.

A validação requer um conjunto de amostras com a propriedade de interesse conhecida e que não foram utilizadas na etapa de modelagem. O modelo será aplicado nessas amostras e os resultados serão estatisticamente avaliados. Muitas vezes, dispõe-se de um conjunto grande de amostras, que será dividido entre aquelas que comporão o conjunto de calibração e aquelas que serão utilizadas para validação. Essa divisão deve ser feita de forma criteriosa, sendo muito comum o uso do algoritmo Kennard-Stone<sup>113</sup> que busca separar as amostras de modo uniforme.

Um conjunto de validação ideal deve ter as seguintes características (adaptado da norma ASTM E1655<sup>97</sup>):

- As amostras devem conter todos os compostos interferentes que se espera para as amostras nas quais o modelo será utilizado.
- A faixa analítica de propriedade ou concentração deve ser abrangente e uniformemente distribuída.
- O número de amostras deve ser suficiente para avaliar estatisticamente o desempenho do modelo. De forma geral, se cinco ou menos fatores são utilizados no modelo, o conjunto de validação deve conter, no mínimo, 20 amostras. Se o modelo utiliza mais que cinco fatores, o conjunto de validação deve conter, no mínimo, cinco vezes o número de fatores.

Dentre os vários possíveis critérios que podem ser utilizados para avaliação dos resultados da validação, serão apresentados o Erro médio quadrático de predição (RMSEP, do inglês, Root Mean Squared Error of Prediction) e a Tendência de validação:

**a.** Erro médio quadrático de predição, RMSEP:

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (v_i - \hat{v}_i)^2}{n}} \quad (28)$$

onde  $v_i$  e  $\hat{v}_i$  é o valor da amostra de validação  $i$  e sua estimativa, respectivamente.  $n$  é o número de amostras de validação.

**b.** Tendência de validação (*bias*):

$$\bar{e}_v = \frac{\sum_{i=1}^n (v_i - \hat{v}_i)}{n} \quad (29)$$

É possível calcular outros parâmetros estatísticos e – estatisticamente – determinar a significância da tendência nos resultados (utilizando, por exemplo, a distribuição *t* de *Student*). Entretanto, o autor entende que a avaliação dos parâmetros de validação deve considerar, primordialmente, a adequação ao uso do modelo. Assim, por exemplo, um modelo com erros significativos pode ser útil para apontar amostras que estejam muito acima de um valor limite ou pode ser utilizado como ensaio de triagem para ensaios adicionais. Dessa forma, apenas a análise crítica dos resultados de validação permite decidir se um modelo é útil ou não.

Infelizmente, conforme será visto adiante, os modelos de calibração avaliados neste projeto não forneceram resultados razoáveis na etapa de validação e, por essa razão as demais figuras de mérito geralmente descritas e estimadas no processo de validação não serão detalhadas nesse texto. Assim sendo, caso o leitor deseje se aprofundar na área de figuras de mérito, recomenda-se a leitura de publicações específicas<sup>97,114–116</sup>.

### **3. REVISÃO BIBLIOGRÁFICA**

#### **3.1. Quimiometria e espectrometria no infravermelho**

Há um número enorme de aplicações de espectrometria no infravermelho em óleo diesel, biodiesel e suas misturas com e sem uso de ferramentas quimiométricas. As aplicações vão de monitoramento de processos de produção<sup>50</sup> à diferenciação de matérias-primas<sup>117</sup>, passando pela identificação de adulterações<sup>118</sup> e o monitoramento da qualidade de produto<sup>119</sup>, apenas para citar algumas.

Também há diversas aplicações que tratam do uso de FTIR para avaliar ou acompanhar a estabilidade do óleo diesel e/ou biodiesel. Dentre os estudos já realizados, pode-se destacar o de Farias et al.<sup>51</sup>, que utilizaram FTIR e termogravimetria para demonstrar a importância das insaturações no processo de oxidação de biodieseis de mamona e óleo de maracujá. A espectrometria no infravermelho médio permitiu destacar as diferenças entre as cadeias carbônicas dos ésteres de cada matéria-prima. Li et al.<sup>120</sup> também utilizaram termogravimetria, mas com atmosfera inerte, e acompanhamento por FTIR para identificar as energias envolvidas e os produtos formados na degradação térmica de biodiesel de óleo de amendoim (com foco neste óleo como resíduo de fritura).

Ogbu e Ajiwe<sup>121</sup> utilizaram FTIR para avaliar a estabilidade térmica de óleo diesel e suas misturas com biodieseis de duas matérias-primas locais (Nigéria/África) concluindo que os picos correspondentes a OH (de hidroperóxidos), C=O (de compostos carbonílicos) e C=C (insaturações) são úteis para acompanhar os processos de formação de óxidos e de degradação dos combustíveis. Neste estudo, os produtos foram aquecidos a diferentes temperaturas por 10 minutos e, então, analisados no espectro infravermelho.

Em função da complexidade e do grande tempo requerido para execução da maioria dos ensaios de estabilidade, uma variedade de métodos alternativos foi proposta. Alguns com o objetivo de prever os resultados dos ensaios de estabilidade tradicionais, outros simplesmente relacionando-os com resultados de métodos analíticos. Muitos se baseiam em análise termogravimétrica (TGA/DTA métodos)<sup>122,120,123,124</sup>, mas há usos de espectroscopia de ressonância paramagnética eletrônica<sup>125</sup>, além de métodos laboratoriais mais básicos como viscosidade e massa específica<sup>122</sup>. O uso de espectrometria no infravermelho, entretanto, é predominante.

Lira et al.<sup>126</sup> utilizaram espectrometria no infravermelho médio e próximo e calibração multivariada para prever propriedades ligadas à estabilidade do biodiesel. Cinco amostras de biodiesel etílico (obtidos de soja, rabanete e sebo) e metílico (obtido de soja com e sem aditivo antioxidante) foram submetidas a envelhecimento acelerado. Ao longo do processo de envelhecimento, foram coletadas alíquotas que foram analisadas quanto à estabilidade oxidativa, índice de acidez e teor de água por métodos tradicionais. As ferramentas PLS e MLR, com diferentes opções de pré-processamento de dados e algoritmos de seleção de regiões espectrais, foram utilizadas para predição desses parâmetros. Constatou-se que os espectros no infravermelho médio e próximo forneceram valores de RMSEP comparáveis e podem ser utilizados para predição dos parâmetros de estabilidade. Os autores advertem, entretanto, que a presença de catalisadores e antioxidantes podem alterar as propriedades estudadas significativamente sem que o espectro infravermelho sofra alterações. Observa-se, também, que a variedade de amostras utilizadas no estudo é muito pequena para representar os processos e as matérias-primas existentes no mercado, o que limita seu uso prático.

Vieira e Pasquini<sup>127,128</sup> utilizaram Espectroscopia de Emissão no Infravermelho Próximo, NIRES, para monitorar a formação de hidroperóxidos em amostras de biodiesel sujeitas a temperaturas elevadas (140 – 180 °C). As curvas de NIRES vs tempo apresentaram boa correlação,  $R^2 = 0,88$ , com os resultados obtidos com o método CEN

EN 14112<sup>129</sup>; um ensaio muito similar ao método EN 15751 empregado neste estudo. O método proposto requer menor volume de amostra e, mais importante, menos tempo: cerca de um décimo, nas condições do estudo. Apesar do estudo utilizar poucas amostras, os autores mostraram que o monitoramento de um composto de degradação intermediário pode fornecer, nas condições do estudo, um parâmetro quantitativo útil para avaliação da estabilidade do biodiesel.

Ni, Norgard e Morup<sup>106</sup> demonstraram a relevância da Regressão por vetores de suporte, SVR para aplicações em espectroscopia no infravermelho e sistemas não lineares. Os autores compararam o desempenho de sete ferramentas de regressão com dados de infravermelho próximo (disponíveis publicamente) que apresentam comportamentos lineares e não lineares. As ferramentas são três variantes de máquina de vetores de suporte, PLS e três variantes de redes neurais. O estudo concluiu que, se os dados passam por pré-processamento adequado, todas as ferramentas avaliadas apresentam desempenhos similares com dados lineares ou com pouca não linearidade (causada por efeitos suaves de espalhamento de luz). Para dados sem pré-processamento ou com não linearidade severa, os autores afirmam que somente as ferramentas de aprendizado de máquina e redes neurais têm desempenho aceitável. Apesar de interessante, o estudo explorou poucas opções de pré-processamento de dados, o que poderia favorecer os métodos de aprendizado de máquinas (menos dependentes de pré-tratamentos).

Filgueiras<sup>108</sup> construiu modelo SVR com espectros no infravermelho médio para determinação de propriedades de petróleo (grau API, viscosidade cinemática e teor de água). O autor concluiu que SVR apresentou resultados mais exatos em relação a PLS para grau API e viscosidade e resultados equivalentes para teor de água. O autor apresentou abordagem que permite avaliar a importância das variáveis originais no modelo de forma análoga ao que se obtém com os loadings de PLS.

Alves, Henrique e Poppi<sup>130</sup> utilizaram SVR e infravermelho próximo para predição dos resultados de Ponto de fulgor e Número de cetano de óleo diesel. Os autores enfatizaram que algumas propriedades de produtos de petróleo apresentam correlações não lineares com dados espectroscópicos. Afirmaram, também, que, apesar de serem adequados à calibração de sistemas não lineares, as redes neurais artificiais tendem a fornecer resultados com sobreajuste.

Filgueiras, Alves e Poppi<sup>131</sup> utilizaram SVR e NIR para quantificar biodiesel de sebo misturado ao biodiesel de soja em óleo diesel B20 (20% v/v biodiesel em óleo diesel). Nesse estudo, o modelo de calibração elaborado com SVR forneceu erros de predição significativamente menores que modelos PLS apesar de ambos apresentarem erros sistemáticos.

### 3.2. Cromatografia Gasosa

A análise cromatográfica de óleo diesel é um grande desafio em função do enorme número de substâncias presentes, muitas delas similares, o que praticamente inviabiliza sua separação efetiva, conforme será demonstrado na seção de Resultados e Discussão da Seção 1. O acoplamento de um espectrômetro de massas e/ou o uso de ferramentas quimiométricas de resolução de sinais acaba se mostrando essencial.

Besser et al.<sup>9</sup> utilizaram um método de degradação acelerada estabelecido em norma (para aplicação em lubrificantes e alterado para óleo diesel) em óleo diesel com 20% v/v de biodiesel metílico de canola, obtendo os seguintes produtos de degradação: mono-metil-éster de ácido nonadióico, metil éster de ácido 9-oxo-nonadioico e ácidos carboxílicos de 1 a 9 átomos de carbono, assim como ácido hidroxicarboxílicos, ácidos dicarboxílicos e ácidos carboxílicos com um grupo epóxido. Para o óleo diesel misturado a biodiesel de soja, os mesmos produtos de oxidação foram encontrados, mas com uma variedade maior de ésteres metílicos de C8 a C18, ácidos carboxílicos insaturados até C8, álcoois e aldeídos. Os ensaios foram realizados na parte líquida e nos precipitados formados por meio de GC-MS com silanização.

Bacha et al.<sup>13</sup> utilizaram GC-MS para identificar os produtos de degradação de ensaio de estabilidade oxidativa de biodieseis de soja e canola e mistura óleo diesel/biodiesel de canola (10 % v/v). O ensaio foi realizado conforme norma ASTM D7545 (também utilizada neste estudo), identificando compostos epóxi com 19 átomos de carbono, derivados de ésteres metílicos, produtos de oxidação de cadeia curta (entre C<sub>2</sub> e C<sub>11</sub>) com o alcanos, alquenos, ácidos, álcoois, aldeídos e cetonas. Os autores admitem que os produtos de oxidação são tão complexos que muitos dos resultados devem ser considerados apenas como qualitativos.

O papel de contaminantes em baixas concentrações na estabilidade do óleo diesel também tem sido objeto de estudo. Pieber et al.<sup>55</sup> apresentaram um novo método para identificação e quantificação de esteril-glicosídeos em biodiesel. Este contaminante é

associado a precipitações e entupimentos de filtros, especialmente, para biodiesel de soja. O método se baseia no monitoramento singular de íons específicos da fragmentação dos esteril-glicosídeos silanizados e, de acordo com os autores, atinge níveis de detecção mais baixos que outros métodos publicados.

A aplicação de GC-MS para caracterização de óleo diesel mostra-se útil também quando esse combustível é obtido a partir de craqueamento de óleos vegetais. Lima et al.<sup>132</sup> utilizaram GC-MS para caracterizar óleo diesel obtido da pirólise de óleos de soja, palma e mamona. A identificação dos compostos representativos e contaminantes formados foi realizada com base na comparação dos espectros obtidos com a biblioteca de espectros do equipamento utilizado. A identificação das classes de compostos foi corroborada por espectrometria no infravermelho e GC-FID.

Flitsch et al.<sup>52</sup> avaliaram os produtos de degradação de biodiesel de canola submetido ao ensaio Rancimat (conforme norma europeia EN 14112) por meio de cromatografia gasosa acoplada à espectrometria de massas e cromatografia de troca iônica, concluindo que o ensaio promove a formação de ácidos fórmico e acético, ácidos graxos de 5 a 18 átomos de carbono e epóxidos. Também foram encontrados aldeídos e álcoois.

Pedley et al.<sup>133</sup> estudaram o papel dos fenalenos e indoís na formação de depósitos em óleo diesel do Reino Unido, analisando amostras armazenadas em condições normais e, também, utilizando o método ASTM D4625<sup>14</sup> para forçar o envelhecimento em 13 semanas. Utilizou-se cromatografia em camada delgada para análise do diesel. O estudo concluiu que as variedades de óleo diesel que contêm essas espécies são as mais suscetíveis à formação de depósitos.

Em função da complexidade da matriz óleo diesel-biodiesel, a cromatografia bidimensional abrangente, GC-GC, é uma opção para resolver compostos que não seriam identificáveis com cromatografia convencional. A associação de GC-GC-FID e Resolução de Curvas Multivariadas permitiu identificar a matéria-prima do biodiesel misturado ao óleo diesel e, adicionalmente, quantificar o biocombustível<sup>134</sup>. A Análise de Componentes Principais, PCA, foi utilizada como etapa preliminar para selecionar regiões relevantes do cromatograma. GC-GC, por sinal, parece ser uma técnica promissora na área de combustíveis em função da incomparável capacidade de resolução<sup>135</sup>.

### 3.3. PARAFAC

Há um enorme número de aplicações de PARAFAC para deconvolução espectral, especialmente para dados de espectrofotometria de fluorescência<sup>81,136-139</sup>. Esse tipo de aplicação de PARAFAC está, conceitualmente, bastante distante da utilizada neste trabalho e não será discutida nesta seção, que focará em aplicações menos usuais. Conforme será detalhado adiante, aplicações de PARAFAC em dados de Cromatografia (gasosa ou líquida) acoplada à espectrometria de massas não são raros, apesar da limitação já comentada de PARAFAC para dados cromatográficos (risco de perda de trilinearidade com o deslocamento dos tempos de retenção). Aplicações de PARAFAC para acompanhar a degradação de produtos ou sistemas também serão discutidas a seguir.

Johnson, Rose-Pehrsson e Morris<sup>46</sup> avaliaram um óleo diesel naval (equivalente a um óleo diesel marítimo brasileiro) após aplicação de condições de estresse utilizando GC-MS e PARAFAC. As alíquotas foram analisadas por GC-MS e os cromatogramas de íons totais submetidos a Análise de Componentes Principais Multidimensional e, em regiões pré-selecionadas com base em Análise de Variância, utilizando PARAFAC. O espectro de massas recuperado nessa região foi, de modo tentativo, associado a um composto aromático que parece estar relacionado ao processo de degradação. Como PARAFAC permitiu identificar apenas um composto gerado nas condições de estresse, os autores concluíram que os resultados permitiram apenas demonstrar o *potencial* da abordagem escolhida.

Toraman et al.<sup>140</sup> aplicaram PARAFAC para decompor e extrair informações relevante de dados Py-GC-MS (pirólise acoplada à GC-MS) obtidos de biomassa geneticamente modificada. O estudo comparou os bio-óleos obtidos a partir de pirólise de álamo geneticamente modificado e selvagem indicando alterações sutis na composição da lignina. O estudo também utilizou PLS-DA (ou *PLS-Discriminant Analysis*) para separar as amostras oriundas de material geneticamente modificado daquele de origem natural e indicar as substâncias responsáveis por essa discriminação.

Morales, Ortiz e Sarabia<sup>141</sup> aplicaram PARAFAC em dados de GC-MS para otimizar os parâmetros de microextração por fase sólida (MSPE) de misturas de clorofenóis em água. PARAFAC permitiu resolver as quatro variedades de clorofenóis presentes na mistura e testar diferentes condições de microextração graças à já comentada vantagem de segunda ordem.

Ortiz et al.<sup>83</sup> utilizaram PARAFAC para avaliar os impactos da temperatura e do tempo de armazenamento na degradação de peixes. A degradação foi mensurada por meio da quantificação de aminas biogênicas que caracterizam a putrefação desse produto (putrecina, cadaverina, tiramina, entre outras). O estudo gerou um conjunto de dados de 8 x 7 x 3 dimensões, sendo 8 momentos, 7 aminas e 3 temperaturas de armazenamento diferentes (-18°C, 4°C e 25°C). PARAFAC permitiu identificar e diferenciar os compostos relacionados com a degradação em cada tipo de armazenamento.

Rocha et al.<sup>142</sup> utilizaram PARAFAC em dados de GC-MS de Materiais de Referência Certificados, MRCs, buscando classificar os produtos e extrair os perfis cromatográficos e espectrais das amostras. Os autores afirmam que o modelo gerado neste estudo poderia ser utilizado para classificar, por similaridade, novos combustíveis indicando possíveis aplicações. Apesar de interessante, essa aplicação ainda carece de confirmação. Os 20 MRCs utilizados incluíram combustíveis (rodoviários e aeronáuticos), lubrificantes automotivos e biocombustíveis (biodieseis de soja e sebo bovino).

Smilde, Bro e Geladi<sup>72</sup> avaliaram parâmetros de qualidade da água ao redor da cidade de Veneza ao longo de 44 meses. Esses parâmetros incluem turbidez, NH<sub>4</sub><sup>+</sup>, demandas química e biológica de oxigênio, material suspenso, fósforo, entre outros. PARAFAC permitiu observar aspectos sazonais na presença de poluentes, bem como separar poluentes de origem industriais de outras. É importante observar que a dimensão temporal dos fenômenos estudados poderia ser perdida (ou “misturada”) à dimensão da localidade com ferramentas quimiométricas convencionais.

Bernal, Díaz e Barrera<sup>143</sup> analisaram o extrato de cálices de *Hibiscus sabdariffa* por HPLC-UV-DAD. PARAFAC permitiu identificar compostos representativos da variabilidade dos dados entre as diversas amostras de flores bem como agrupar as amostras por similaridades. Os pesos da primeira componente da dimensão cromatográfica foram decisivos para diferenciação das amostras.

### 3.4. MCR-ALS

De acordo com Março et al.<sup>88</sup> as aplicações mais comuns de MCR-ALS incluem monitoramento de reação em escala de laboratório e industrial com diferentes tipos de equipamentos hifenados, como por cromatografia e análise por injeção em fluxo. Outra área de aplicação de MCR-ALS de grande importância é na resolução de imagens

hiperespectrais. Nessa técnica, coleta-se o espectro de cada pixel de uma imagem ou objeto. Ruckebusch e Blanchet<sup>144</sup> afirmam que MCR-ALS é o algoritmo mais popular e flexível nessa área. MCR-ALS é especialmente adequada para espectroscopia Raman sendo possível encontrar aplicações na literatura que vão da identificação de minerais em meteoritos lunares<sup>145</sup> a adulteração de medicamentos<sup>146</sup>.

Marques et al.<sup>147</sup> aplicaram MCR-ALS em dados de LC-MS para avaliar o efeito de arsênio (III) no perfil lipídico de cianobactérias. Os autores analisaram o extrato lipídico de duas espécies de bactérias tratadas com o metal pesado por 7 dias. Os espectros de massas resolvidos por MCR-ALS indicaram redução de lipídeos associados ao processo de fotossíntese explicando como esses organismos são afetados pelo As(III).

Câmara et al.<sup>118</sup> aplicaram MCR-ALS em dados de infravermelho médio e próximo para identificar a presença de óleo vegetal em mistura diesel-biodiesel. Os desempenhos de MCR-ALS e PLSR para quantificação dos contaminantes foram comparados, mostrando que a primeira ferramenta fornece erros de predição bastante inferiores. Os autores enfatizaram que MCR-ALS pode ser utilizado para recuperar os espectros de outros contaminantes mesmo na presença de espécies interferentes.

Kwok e Taylor<sup>146</sup> utilizaram MCR-ALS em dados de microscopia Raman para identificar falsificação de medicamento em forma de comprimidos. Os pesquisadores compararam comprimidos legítimos e falsificados do medicamento Cialis<sup>®</sup> e aplicaram MCR-ALS nos complexos espectros obtidos. Por comparação dos espectros recuperados com os espectros de componentes puros, os autores concluíram que as versões falsificadas contêm compostos não ativos diferentes qualitativa e quantitativamente, além de apresentar heterogeneidade no componente ativo.

Le Dréau et al.<sup>65</sup> utilizaram FTIR para acompanhar o envelhecimento de óleos vegetais comestíveis numa célula de envelhecimento fabricada pelo grupo. O uso de MCR-ALS permitiu extrair o espectro das principais espécies formadas e degradadas em óleos de amendoim, canola, nozes, uva, soja e girassol. As componentes MCR-ALS permitiram determinar quais espécies se alteram no envelhecimento de cada tipo de óleo.

## **SEÇÃO 1 – APLICAÇÃO DE FERRAMENTAS MULTIDIMENSIONAIS PARA AVALIAÇÃO DE ENSAIOS DE ESTABILIDADE DE ÓLEO DIESEL**

Nesta seção, 4 métodos normalizados de avaliação de estabilidade foram aplicados em 18 amostras de óleo diesel (contendo biodiesel a, aproximadamente, 8 %v/v) coletadas em postos de revenda nas 5 regiões do Brasil. Os métodos selecionados seguem as normas americanas ASTM D4625, D6468 e D7545 e a norma europeia CEN EN15751. As alterações físicas e químicas foram avaliadas por espectrometria no infravermelho médio e, para algumas amostras, por cromatografia gasosa acoplada à espectrometria de massas. Os dados obtidos foram avaliados com as ferramentas quimiométricas multidimensionais Análise de Fatores Paralelos, PARAFAC, e MCR-ALS com o objetivo de avaliar e comparar as transformações promovidas pelos ensaios.

### **4. MATERIAIS E MÉTODOS DA SEÇÃO 1**

#### **4.1. Obtenção das amostras de óleo diesel**

As amostras de óleo diesel utilizadas neste estudo foram coletadas em postos de revenda entre dezembro de 2017 e fevereiro de 2018, período no qual vigorava o teor obrigatório de 8 % v/v de biodiesel. As amostras foram transportadas em recipientes de 5 litros de PEAD (Polietileno de alta densidade) próprios para armazenamento de combustíveis e acondicionadas em caixas de isopor<sup>®</sup> contendo, quando necessário, gelo. As caixas dispunham de termômetro com registro de temperatura mínima e máxima, o que permite afirmar que nenhuma das amostras foi exposta a temperaturas menores que 4 °C ou maiores que 20 °C. Assim que possível, os recipientes foram transferidos para geladeiras dos laboratórios da ANP onde foram armazenadas em temperaturas de, aproximadamente, 2 °C.

Apenas amostras de óleo diesel S10 foram coletadas por ser um produto mais sensível à degradação<sup>36</sup> e porque, num futuro breve, este se tornará o tipo principal comercializado no Brasil, já que os veículos pesados novos comercializados desde 2012 só podem utilizar esse combustível<sup>148</sup>.

A escolha do número de amostras e dos locais de coleta foi feita com base em diversos fatores simultaneamente. A primeira premissa era coletar o maior número de amostras possível, já que se pretende que o estudo não seja limitado a produtos vindos de uma única combinação de refinaria e usina de biodiesel. Por outro lado, afirmar que o estudo

seria aplicável a qualquer mistura diesel/biodiesel nacional exigiria um esforço inexecutável, pois, além dos fatores enumerados abaixo, as características desse produto alteram-se constantemente por mudanças nos processos de produção e pela entrada de produto importado, em princípio, de qualquer local do mundo que produza dentro das especificações brasileiras. Outros fatores considerados (em ordem de prioridade):

- 1- O ensaio de estabilidade conforme norma ASTM D 4625 implicou um limite de 18 amostras, já que as estufa disponível comportava, no máximo, 36 frascos (todos os ensaios foram realizados em duplicata).
- 2- Com base nos dados de venda de combustíveis para distribuidoras, buscou-se localidades abastecidas por diferentes refinarias e por distribuidoras que importam óleo diesel regularmente;
- 3- Somente postos de revenda que receberam o produto da distribuidora recentemente (menos que 3 dias da coleta) foram selecionados. Essa informação era obtida junto ao gerente do posto ou em consulta às notas fiscais de recebimento do produto.
- 4- Respeitando-se as premissas acima, buscou-se localidades o mais próximo possível de Brasília para minimizar custos e tempo de transporte – o que minimiza períodos de agitação e variações de temperatura durante o transporte.

Dessa forma, as coletas foram realizadas em 10 cidades distribuídas em 6 estados e no Distrito Federal, englobando todas as 5 regiões do País, conforme Tabela 4. A localização das cidades é ilustrada na Figura 21.

Tabela 4: Locais e datas de coletas das amostras de óleo diesel utilizadas neste estudo. A Bandeira se refere ao nome da distribuidora ostentado no posto de revenda. Postos de “Bandeira Branca” podem revender combustíveis de qualquer distribuidora.

Posto de revenda	Data da coleta	Local	Nome do posto*	Bandeira	Distribuidora de origem
1	25/01/2018	Brasília/DF	109 Norte	Shell	Shell
2	13/01/2018	Londrina/PR	Araguaia	Branca	Biostratum
3	21/01/2018	Belo Horizonte/MG	Bel	Ipiranga	Ipiranga
4	14/01/2018	Itápolis/SP	Cazulão	Branca	Redepetro
5	13/01/2018	Londrina/PR	Duim	BR	BR
6	21/01/2018	Belo Horizonte/MG	Formula BR	BR	BR
7	21/01/2018	Goiânia/GO	Goianão	Ipiranga	Ipiranga
8	13/01/2018	São J. do Rio Preto/SP	HM	Branca	Ciapetro
9	16/12/2017	Porto Nacional/TO	Imperial	BR	BR
10	18/01/2018	Brasília/DF	Jarjour	Branca	Total
11	14/01/2018	Assis/SP	Panema	Ipiranga	Ipiranga
12	16/02/2018	Brasília/DF	Petronorte	Ipiranga	Ipiranga
13	17/12/2017	Barreiras/BA	Pit Stop Dourado	Branca	Larco
14	13/01/2018	Monte Alegre/MG	Rancho Alegre	Branca	Federal
15	16/12/2017	Porto Nacional/TO	Real	Branca	Total
16	17/12/2017	Barreiras/BA	Sabba	Shell	Shell
17	15/01/2018	Brasília/DF	Serv Car	BR	BR
18	12/01/2018	Goiânia/GO	Xodó	Branca	Max Dist. de Petróleo

\* Nome “fantasia” que é precedido de “Posto” ou “Auto posto”.

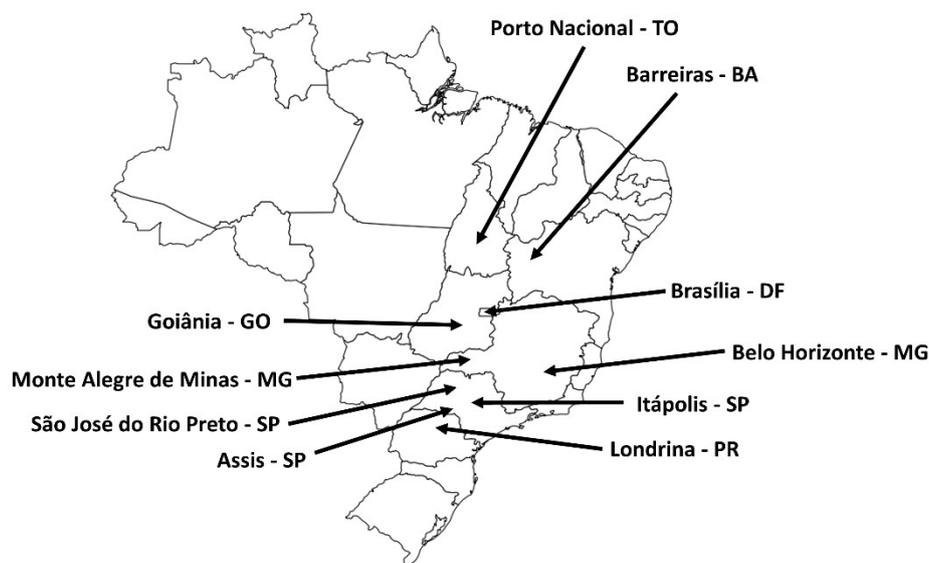


Figura 21: Locais de coleta das amostras de óleos diesel utilizadas no estudo.

## 4.2. Avaliação prévia das amostras coletadas

Este estudo incluiu apenas óleos diesel que atendem as especificações da ANP. Para garantir esse requisito, as 18 amostras coletadas foram analisadas nos ensaios utilizados rotineiramente para monitoramento da qualidade dos combustíveis pela ANP. Os ensaios são aqueles que indicam alterações por contaminação (acidental ou intencional) e degradação de maneira mais eficiente. A seguir, esses ensaios são descritos bem como as normas técnicas aplicadas. Todos os ensaios foram realizados por analistas treinados, com equipamentos calibrados e verificados conforme normas técnicas de ensaios e conforme previsto na norma de gestão laboratorial ABNT NBR ISO/IEC 17025<sup>149</sup>:

- Cor e Aspecto: Apesar de serem características superficiais e, até certo ponto, subjetivas, a cor e o aspecto do óleo diesel são parâmetros de qualidade muito importantes na avaliação da qualidade de combustíveis. O parâmetro cor permite diferenciar os óleos diesel S10 e S500, já que o segundo contém, obrigatoriamente, um corante vermelho que serve para alertar o consumidor sobre o teor de enxofre no produto. Já o parâmetro Aspecto contempla quaisquer contaminantes insolúveis em concentração significativa, o que inclui contaminantes do processo de transporte e armazenamento e substâncias formadas no processo de degradação do combustível, como água e materiais de massa molar alta oriundas da degradação do óleo diesel e/ou biodiesel.
- Massa específica (conforme norma ABNT NBR 14065): Parâmetro relacionado ao projeto do motor. No ciclo de combustão, a bomba injetora controla o *volume* de produto injetado na câmara de combustão; a massa específica garante que a *massa* correta de combustível será injetada. Além disso, é um parâmetro que pode indicar contaminações grosseiras com outros líquidos de densidades diferentes.
- Ponto de Fulgor (conforme norma ASTM D93): primordialmente, está relacionado à segurança no manuseio do óleo diesel, mas é um importante indicativo de cuidado no manuseio, já que pequenas contaminações com combustíveis mais voláteis (gasolina e etanol, por exemplo) provocam grandes alterações nesse parâmetro. A presença desses contaminantes pode reduzir o número de cetano (com queda de eficiência do motor) e causar danos em polímeros presentes no motor.
- Destilação (conforme norma ASTM D86): é um parâmetro muito importante de qualidade, pois permite avaliar a proporção dos compostos presentes no óleo

diesel. O balanço correto de hidrocarbonetos leves, médios e pesados é considerado no projeto dos motores e necessário para partida do motor, operação adequada em diferentes regimes de rotação e autonomia sem formação de resíduos ricos em carbono<sup>2</sup>.

- Teor de biodiesel (conforme norma CEN EN14078): além de garantir a proporção prevista em lei do biocombustível, garante que a mistura tenha a lubricidade adequada e as características de combustão esperadas.

### **4.3. Ensaios de Estabilidade**

A seguir, são descritos os ensaios de estabilidade avaliados neste estudo. Todos os ensaios buscam realizar a degradação do óleo diesel de modo a fornecer parâmetros objetivos que mensurem essa degradação após a aplicação de condições de estresse. Em dois dos ensaios utilizados (conforme D4625 e D6468), esses parâmetros não foram medidos, pois isso envolveria filtrar todo o combustível e lavar o filtro com solvente. Esse procedimento alteraria as características do combustível seja pela sua aeração ao longo da filtração, seja pela adição do solvente utilizado para lavagem dos filtros. Como o resultado, propriamente dito, de cada ensaio está fora dos objetivos deste estudo, para estes ensaios, as amostras passaram apenas pela etapa de estresse, sendo armazenadas em refrigerador em seguida.

Considera-se que o método ASTM D4625 é capaz de simular as condições de armazenamento reais, mas de forma acelerada. Ou seja, esse ensaio não levaria à formação de espécies químicas ou outras alterações significativamente diferentes daquelas observadas em campo. Outro ensaio que merece destaque é o que segue a norma EN15751, pois ele é considerado o ensaio padrão para biodiesel e óleo diesel em muitas normas e legislações<sup>16,20,22,150</sup>. Esse entendimento é bastante consolidado na indústria e academia<sup>48,14,8,151</sup>. A seguir, são descritos os ensaios escolhidos para este estudo. O Apêndice 2 apresenta uma tabela de comparação entre os ensaios aqui descritos.

#### ***4.3.1. Ensaio de estabilidade ao armazenamento a 43 °C (ASTM D4625)***

O ensaio consiste em aquecer 400 mL de óleo diesel em frascos de vidro borossilicato âmbar (Figura 22) por períodos de 4 a 24 semanas. O frasco dispõe de um respiro que permite a equalização de pressão com o ambiente. Após o período selecionado, a amostra é resfriada até a temperatura ambiente e o teor de materiais insolúveis é determinado por filtração e subsequente lavagem com iso-octano.

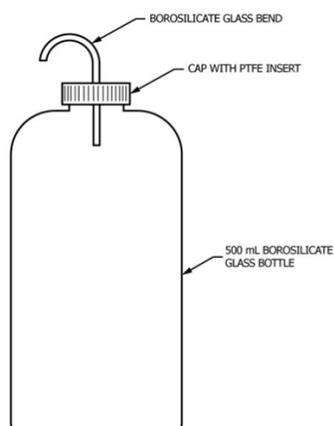


Figura 22: Esquema do frasco e foto da estufa utilizados para realização do ensaio conforme norma ASTM D4625.

Estudos demonstraram que, grosso modo, os teores de insolúveis obtidos neste ensaio correspondem aos valores obtidos em condições normais de armazenamento multiplicadas por  $4^{14}$ , ou seja, cada semana do ensaio corresponde a, aproximadamente, um mês de armazenamento em condições reais. Assim, se o estudo é conduzido por 13 semanas (*período escolhido neste projeto*), espera-se obter o mesmo teor de insolúveis que seria observado para o mesmo produto armazenado ao longo de um ano em condições reais.

Todos os frascos, tampas e respiros utilizados neste estudo eram novos e o aquecimento foi realizado numa estufa de secagem da marca Selecta, modelo Digitronic. A temperatura foi monitorada em diferentes momentos com termo higrômetro calibrado obtendo-se temperaturas sempre no intervalo  $(43,0 \pm 0,5) \text{ }^\circ\text{C}$  (enquanto a norma estabelece um limite de  $\pm 1 \text{ }^\circ\text{C}$ ). A posição de cada amostra na estufa foi definida aleatoriamente.

Após o período de 13 semanas, os frascos foram retirados da estufa e deixados resfriar naturalmente. Os respiros foram retirados, os recipientes foram lacrados com Parafilm<sup>®</sup> e armazenados em refrigerador a, aproximadamente,  $2 \text{ }^\circ\text{C}$  até a realização dos ensaios subsequentes.

#### ***4.3.2. Ensaio rápido de oxidação em pequena escala (ASTM D7545)***

O ensaio consiste em introduzir 5 mL do combustível em um vaso de pressão, folheado a ouro, que é pressurizado com oxigênio a 700 kPa e, subsequentemente,

aquecido a 140 °C<sup>15</sup>. A pressão atingida após esse aquecimento (aproximadamente, 1000 kPa nas condições deste estudo) é monitorada até que uma queda de 10 % ocorra (breakpoint). Essas etapas são realizadas automaticamente pelo equipamento da marca AntonPaar®, modelo PetroOxy®. Após a finalização do ensaio, resfriamento do vaso de reação e equalização da pressão, o óleo diesel oxidado foi coletado com pipeta de Pasteur (Figura 23) e armazenada em frasco âmbar, em refrigerador, até realização dos ensaios subsequentes.

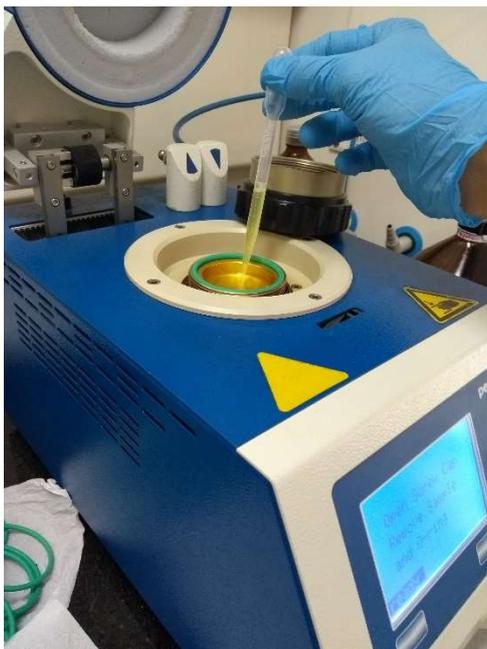


Figura 23: Coleta de amostra oxidada após realização de ensaio conforme norma ASTM D7545.

#### ***4.3.3. Ensaio de estabilidade em alta temperatura (ASTM D6468)***

Cinquenta mililitros de combustível são aquecidos por 90 ou 180 minutos a 150 °C em tubos de ensaio (abertos)<sup>11</sup>. Neste estudo, selecionou-se, por conveniência, o período de 90 minutos. Após aquecimento e resfriamento, o combustível é filtrado e a quantidade de material insolúvel é estimada pela reflectância dos elementos filtrantes utilizando equipamento específico e por comparação com padrões (de reflectância).

Neste estudo, utilizou-se um forno de calibração de medidores de temperatura que utiliza convecção de ar quente (mostrado na Figura 24) ao invés de um banho líquido conforme previsto na norma técnica. Não houve um estudo de comparação entre o desempenho do forno utilizado e aquele previsto em norma, entretanto, entende-se que a exatidão da medição de temperatura e o princípio do forno utilizado (convecção)

garantem resultados comparáveis com os que seriam obtidos com o equipamento estabelecido na norma.



Figura 24: Forno utilizado para realização do ensaio de estabilidade em alta temperatura com tubo de ensaio e amostra.

Após o período de 90 minutos, as alíquotas de óleo diesel foram vertidas diretamente para frascos âmbar de 50 mL que, após resfriamento natural, foram armazenados em refrigerador a, aproximadamente, 2 °C até a realização de ensaios subsequentes.

#### ***4.3.4. Ensaio de estabilidade à oxidação por oxidação acelerada (EN 15751)***

Neste ensaio, um fluxo de ar purificado é borbulhado através de 7,5 gramas de amostra do combustível aquecido a 110 °C em tubo de ensaio<sup>152</sup>. Esse fluxo de ar (10 L/min) é direcionado a uma célula de medição adjacente que contém água destilada e um condutímetro (ver Figura 25). Este registra o período de indução que corresponde a um rápido crescimento na condutividade ou, mais precisamente, à inflexão no gráfico de condutividade versus tempo conforme Figura 26. Essa inflexão – que corresponde ao período de indução – indica a formação de ácidos carboxílicos de baixa massa molecular que são carregados pelo fluxo de ar. O software do equipamento determina o ponto de inflexão pelo cálculo da segunda derivada da curva de condutividade; o ponto de máximo corresponde ao período de indução.

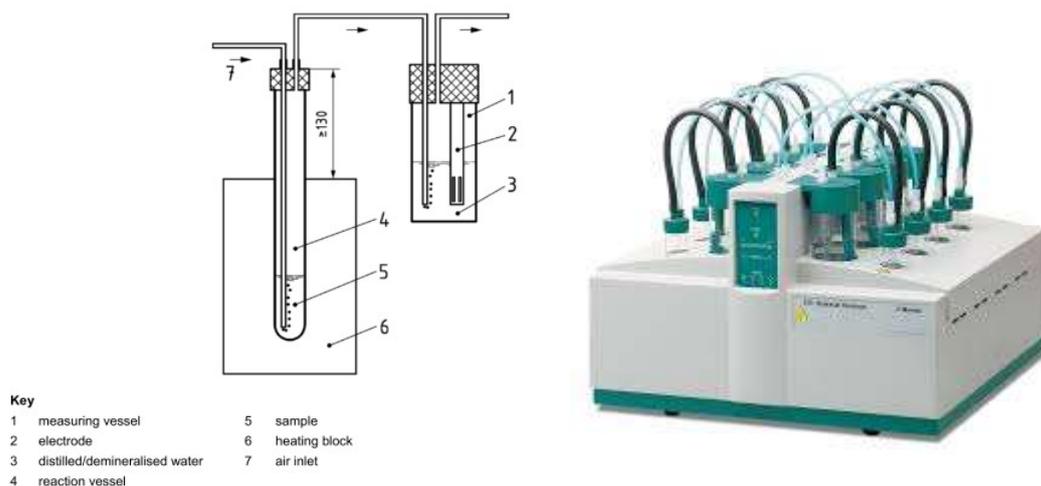


Figura 25: Esquema e foto do equipamento Rancimat<sup>®</sup> utilizado para realização do ensaio conforme norma CEN EN 15751.

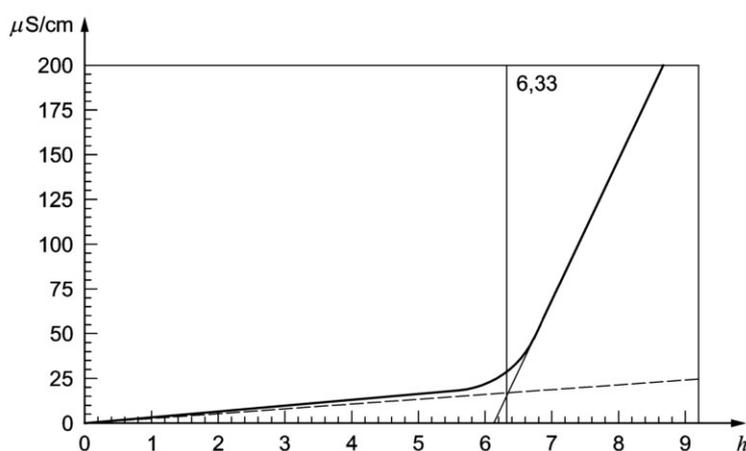


Figura 26: Exemplo de curva de condutividade versus tempo destacando uma inflexão (ou período de indução).

Após detecção do período de indução, o equipamento automaticamente interrompe o fluxo de ar e o aquecimento. O produto (ou o que restou no tubo de ensaio) foi, então, vertido em frasco âmbar de 50 mL e armazenado em refrigerador a, aproximadamente, 2 °C até a realização dos ensaios subsequentes.

#### 4.4. Espectrometria no infravermelho

Todas as amostras de óleo diesel deste projeto, antes e após a realização dos ensaios de estabilidade, foram analisadas por espectrometria no infravermelho médio em equipamento da marca Bruker<sup>®</sup>, modelo Tensor II<sup>®</sup> utilizando célula de reflectância total atenuada (ATR) – também denominado HATR, destacando a orientação da célula – de Seleneto de Zinco com 10 reflexões (marca Pike<sup>®</sup>) (Figura 27).

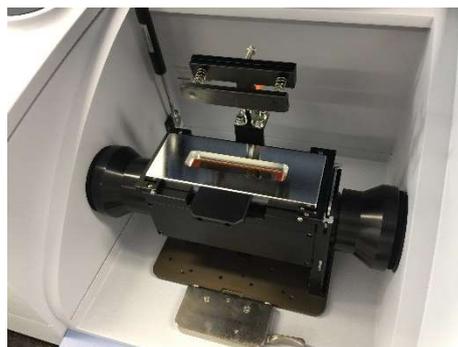


Figura 27: Espectrômetro Bruker® Tensor II® com célula ATR de ZnSe.

Para cada espectro obtido, foi realizada a aquisição de espectro de *background* com o acessório de ATR vazio, o que elimina/anula o efeito de espécies absorventes atmosféricas como CO<sub>2</sub> e umidade, assim como contaminantes presentes na célula de ATR que, eventualmente, não tenham sido retirados pelas etapas de limpeza.

As seguintes configurações foram adotadas após otimização:

Faixa espectral: 4.000 a 400 cm<sup>-1</sup> (os espectros foram, posteriormente, “cortados” em 650 cm<sup>-1</sup> em função da faixa de transparência do ZnSe)

Resolução: 4 cm<sup>-1</sup>

Número de varreduras por amostra: 16

Número de varreduras para aquisição de *background*: 16

Tipo de espectro: Absorbância corrigida para ATR

Abertura de feixe: 6 mm

Detector: DLaTGS (do inglês, Deuterated Lanthanum  $\alpha$  Alanine doped TriGlycine Sulphate)

A adição do óleo diesel à célula de ATR foi realizada com pipetas de Pasteur descartáveis, utilizando-se de 1 a 1,5 mL de amostra. A limpeza era realizada com tolueno, seguido de ciclohexano e secagem com lenço de papel e jato de ar seco.

#### 4.5. Cromatografia gasosa acoplada à Espectrometria de massas

*Nota: Infelizmente, em julho de 2019, o refrigerador no qual estavam guardadas as alíquotas que seriam analisadas por GC-MS parou de funcionar momentaneamente. O gelo acumulado no equipamento derreteu e molhou os recipientes, fazendo as etiquetas se descolarem e/ou apagarem. Dessa forma, foi impossível identificar a maioria das alíquotas. Apenas três puderam ser identificadas quanto aos quatro ensaios de estabilidade. Apesar desse contratempo, entendeu-se que o estudo com três amostras agregaria informação relevante para o projeto, principalmente porque elas tinham origens bastante diversas (Paraná, São Paulo e Tocantins).*

As amostras de óleo diesel 2, 8 e 15 (vide Tabela 4) foram analisadas por GC-MS antes e após a realização dos ensaios de estabilidade. Os ensaios foram realizados de forma preliminar em agosto e setembro de 2018 e, após otimização das condições, repetidos em agosto de 2019.

Utilizou-se um sistema da marca Agilent® modelo 6890N/5973N com amostrador automático (Figura 28). O cromatógrafo está configurado para injeção do tipo *split* e o espectrômetro de massas é do tipo quadrupolo com detector fotomultiplicador.



Figura 28: Sistema GC-MS Agilent® modelo 6890N/5973N com amostrador automático.

Inicialmente, as amostras foram silanizadas e, a seguir, diluídas em n-heptano na proporção de 1:10. Para tanto, 100 µL de cada amostra de óleo diesel foram transferidos

diretamente para vials cromatográficos de 2 mL. A seguir, foram adicionados 100 µL de piridina e 100 µL de MSTFA (*n*-metiltrimetilsililtrifluoroacetamida). Os vials foram, então, fechados, agitados e deixados em repouso por 5 minutos. Por fim, foram adicionados 700 µL de *n*-heptano. Utilizou-se micropipeta de 100 a 1000 µL Eppendorf calibrada com erro inferior a 3% e ponteiros descartáveis, além de piridina (Anidra – marca Vetec<sup>®</sup>), MSTFA (> 98,5% mín. - marca Sigma Aldrich<sup>®</sup>) e *n*-heptano (PA - 99 % mín. - marca Neon<sup>®</sup>).

Para garantir a comparabilidade de resultados e, principalmente, a eficiência da identificação dos compostos por busca na biblioteca de dados, realizou-se o procedimento de *tuning* antes de cada batelada de ensaios. Este procedimento calibra e ajusta as intensidades e as razões massa-carga do analisador quadrupolo com base na análise de um padrão com valores conhecidos (neste caso, PFTBA - perfluorotributilamina). Os parâmetros de ajuste foram comparados com os valores informados no manual fabricante. Também foram avaliados os parâmetros que indicam a estanqueidade do analisador de massas como a presença de picos característicos de oxigênio, nitrogênio e água, conforme instruções do fabricante.

Com o objetivo de maximizar a resolução dos componentes da mistura diesel/biodiesel e partindo-se da experiência com análise de outras amostras de óleos diesel, vários parâmetros foram testados chegando-se às condições cromatográficas e espectrométricas abaixo. Cada corrida cromatográfica consome 117 minutos:

Injeção:

Volume: 1 µL (microseringa de 10 µL)

Modo de injeção: split com taxa de 50:1

Condições cromatográficas:

Coluna: HP-5MS, temp. máx: 350 °C, 30 m x 250 µm x 0,25 µm (comprimento x diâmetro x espessura de filme)

Fluxo do gás de arraste: fluxo constante de 1 mL/min de He

Programação de temperatura para forno do cromatógrafo:

Taxa de aquecimento / °C/min	Temperatura / °C	Tempo de permanência na temperatura / min.
-	40	1,0
2,0	200	15,0
10,0	310	10,0

Programação do espectrômetro de massas:

*Solvent delay*: 6 min

Intervalo de varredura: 30,0 a 600,0 Da

Taxa de varredura: 2,60 s/scan

A identificação dos compostos foi realizada com base na comparação dos espectros obtidos com a biblioteca disponível no *software* Agilent Chemstation: NIST 02 (com 174.948 espectros). A comparação é feita automaticamente pelo programa. Somente compostos com índice de similaridade acima de 95% foram considerados. Além disso, a avaliação visual dos espectros e a coerência das espécies apresentadas pelo software também foram consideradas.

#### 4.6. Ferramentas quimiométricas

Os dados foram exportados dos equipamentos analíticos para um computador Dell® com processador Intel Core i3® e 8 GB de memória RAM. Os cálculos multivariados foram realizados com o programa MATLAB® versão R2016b (Mathworks Inc.®) utilizando os seguintes pacotes (“Toolboxes”):

- PARAFAC: N-way Toolbox para Matlab, versão 3.3 (Rasmus Bro e Claus Anderson).
- MCR-ALS: MCR-ALS GUI, versão 2.0 (Joaquim Jaumot, Anna Juan, Romá Tauler).

## 5. RESULTADOS E DISCUSSÃO DA SEÇÃO 1

### 5.1. Avaliação preliminar

Os ensaios de rotina não apresentaram sinais de contaminação ou degradação das amostras de óleo diesel. Uma das amostras apresentou o teor de biodiesel fora do limite

vigente à época dos ensaios ( $8 \pm 1$  % v/v), entretanto, dentro da tolerância analítica adotada pelo laboratório.

Os detalhes dos resultados dos ensaios são apresentados no Apêndice 3 e a Figura 29 mostra os mesmos resultados na forma de gráficos juntamente com os limites da especificação adotada pela ANP. Os dados também foram comparados na ordem de coleta (não mostrado), o que não indicou tendências. Infere-se, portanto, que as condições de transporte e armazenamento foram suficientes para evitar alterações nas propriedades dos óleos diesel.

Observa-se que os resultados apresentaram grande variabilidade dentro dos limites de especificação, o que demonstra que a estratégia de coletas foi bem-sucedida, pois obteve-se amostras com grande diversidade de composição e propriedades.

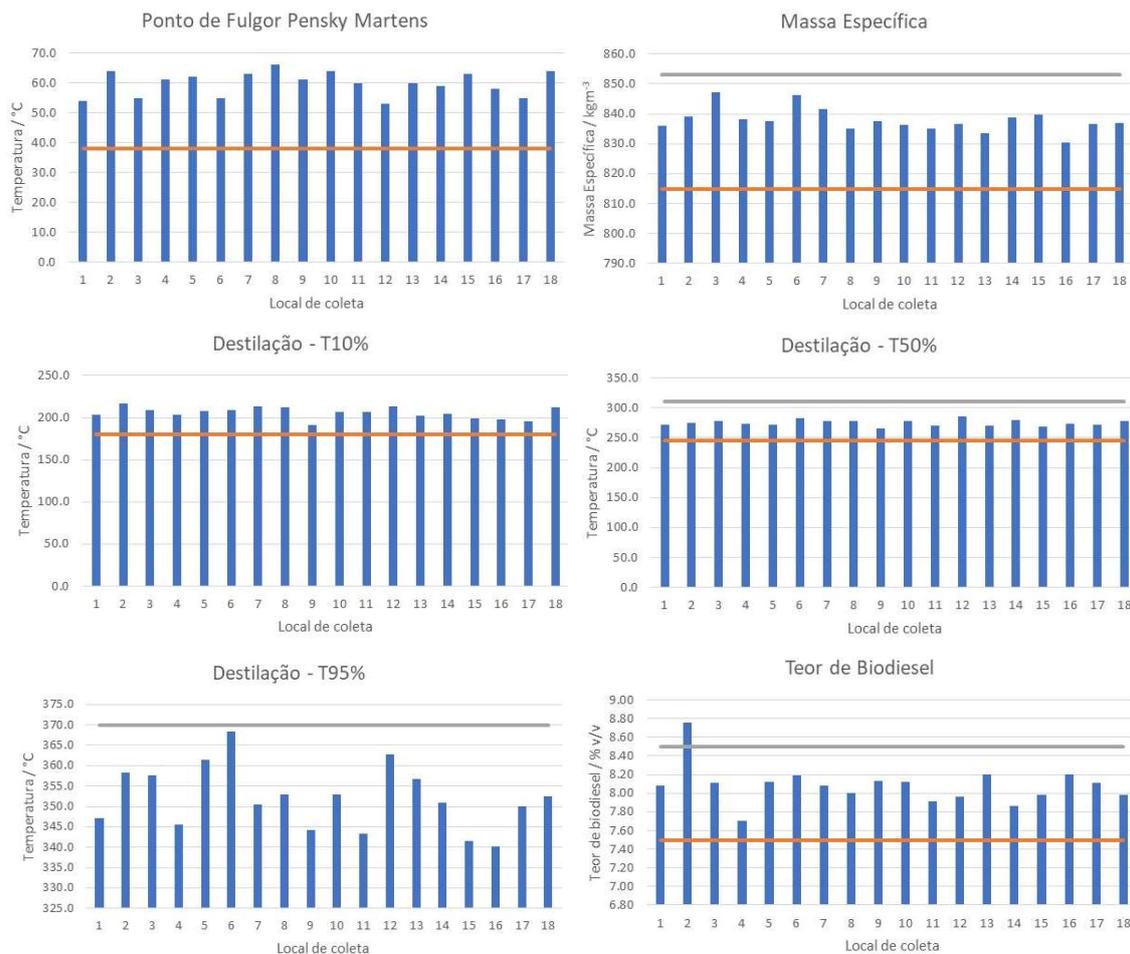


Figura 29: Resultados dos ensaios de rotina das 18 amostras deste estudo (— limite superior; — limite inferior). A amostra do posto 2 apresentou teor de biodiesel acima do limite da especificação ANP, entretanto, abaixo do limite de tolerância adotado no laboratório.

## 5.2. Ensaio de estabilidade

Conforme descrito no subitem 4.3, os ensaios realizados conforme normas ASTM D4625 e D6468 não foram executados até obtenção do resultado final, pois isso requereria a realização de filtração (e mistura do combustível ao solvente de filtração).

Já para os ensaios conforme normas ASTM D7545 e CEN EN15751, os resultados são obtidos sem manipulação da amostra e, por isso, estão detalhados na Tabela 5.

Tabela 5: Resultados dos ensaios de estabilidade pelos métodos CEN EN15751 e ASTM D7545.

Identificação	CEN 15751 / horas			ASTM D7545 / minutos		
	Resultado 1	Resultado 2	Média	Resultado 1	Resultado 2	Média
1	23,6	23,8	23,7	82,8	81,9	82,4
2	20,3	19,4	19,8	86,5	87,3	86,9
3	30,1	30,5	30,3	87,1	85,0	86,0
4	36,0	35,1	35,5	89,1	90,1	89,6
5	18,6	18,4	18,5	76,0	76,6	76,3
6	31,5	32,8	32,2	85,6	83,6	84,6
7	25,4	25,0	25,2	77,7	75,4	76,5
8	35,0	33,6	34,3	93,5	94,3	93,9
9	36,6	34,8	35,7	79,1	79,1	79,1
10	31,2	30,6	30,9	86,1	87,5	86,8
11	24,6	25,7	25,2	77,6	76,6	77,1
12	31,9	31,0	31,4	93,0	90,3	91,6
13	31,4	31,9	31,7	79,8	80,3	80,0
14	29,1	28,1	28,6	92,2	90,9	91,5
15	23,5	23,2	23,3	72,3	79,9	76,1
16	*	*	*	109,1	106,8	108,0
17	12,8	12,9	12,9	80,3	79,7	80,0
18	23,4	22,0	22,7	76,0	76,7	76,3

\* Resultado não pôde ser obtido.

Assim como foi observado nos ensaios de rotina, há uma grande variabilidade de resultados nos ensaios estabilidade (ver Figura 30 e Figura 31). A amostra do posto 16 mostrou comportamento anômalo no ensaio de estabilidade pela norma CEN EN15751, não sendo possível estabelecer a inflexão da curva de condutividade versus tempo. O resultado para o ensaio conforme ASTM D7545 para a mesma amostra sugere que ela apresenta uma estabilidade excepcionalmente alta, o que poderia explicar porque a inflexão não foi observada no ensaio anterior mesmo após 48 horas.

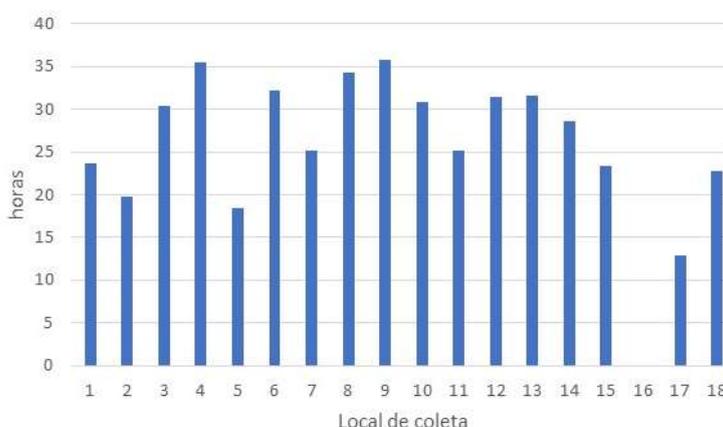


Figura 30: Resultados dos ensaios de estabilidade conforme norma CEN EN15751.

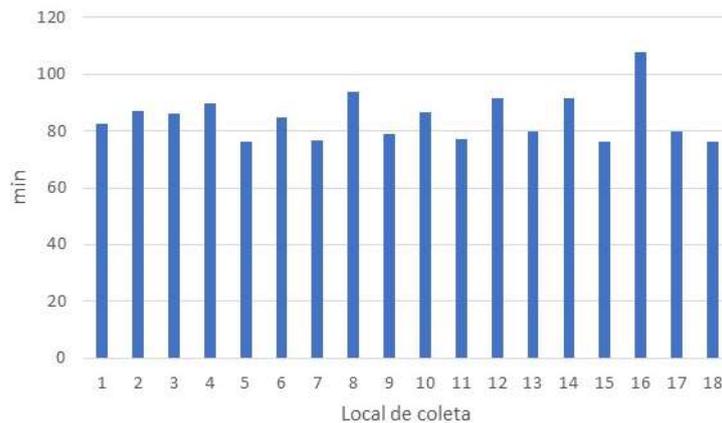


Figura 31: Resultados dos ensaios de estabilidade conforme norma ASTM D7545.

A Figura 32 mostra a correlação entre os dois ensaios de estabilidade. Apesar de estar fora do escopo deste projeto, a correlação entre ensaios pode ser de grande interesse se um deles é mais rápido ou barato que o outro. O coeficiente de correlação de Pearson é igual a 0,259, o que é significativo para um nível de confiança de 95 % ( $r_{\text{tabelado}, n=17} = 0,482$ )<sup>153</sup>. Entretanto, a utilização dos resultados de um ensaio para previsão do resultado do outro, neste caso, teria alta incerteza.

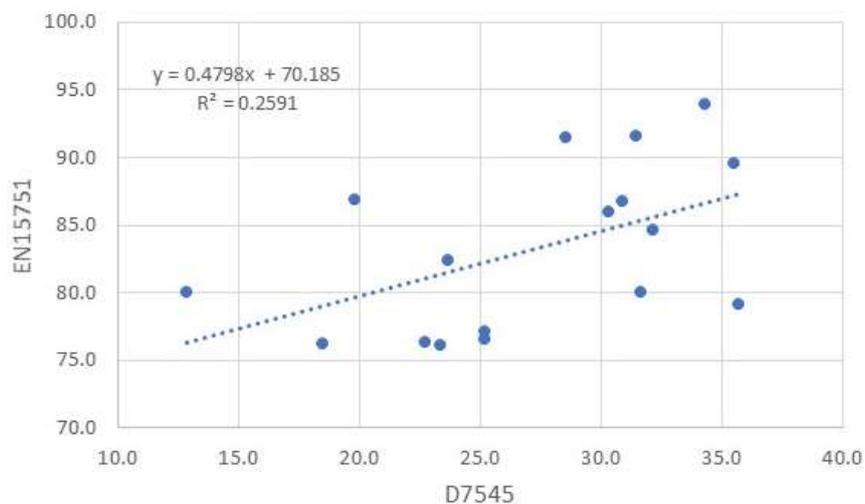


Figura 32: Correlação entre os resultados dos ensaios conforme normas D7545 e EN15751.

### 5.3. Espectrometria no infravermelho médio

A Figura 33-A mostra 36 espectros das 18 amostras (em duplicata) antes dos ensaios de estabilidade. Nota-se variação entre os espectros nas seguintes regiões:

- Picos em, aproximadamente,  $3008\text{ cm}^{-1}$  que (Figura 33-C), provavelmente, são devidos ao estiramento CH de olefinas cis. Essa variação pode ser atribuída às diferenças de cadeias carbônicas dos biodieseis presentes nas misturas. O biodiesel de gordura animal, por exemplo, apresenta menos insaturações que o biodiesel de soja.
- Entre  $2950$  e  $2850\text{ cm}^{-1}$  (Figura 33-C), o que corresponde aos estiramentos simétricos e assimétricos de  $\text{CH}_3$  e  $\text{CH}_2$  alifáticos de atribuição incerta.
- Aproximadamente,  $1746\text{ cm}^{-1}$  (Figura 33-B) que corresponde ao estiramento C=O da carbonila provavelmente em função de pequenas variações no teor de biodiesel (também há variações menores na região de ca.  $1164\text{ cm}^{-1}$  relacionadas ao estiramento C-O de éster).
- Entre, aproximadamente,  $1463$  e  $1376\text{ cm}^{-1}$  (Figura 33-B), correspondentes às deformações angulares CH de hidrocarbonetos alifáticos (pouco específicas).

Além disso, há variações na região da “impressão digital” que correspondem a uma combinação de várias bandas de absorção e que são de difícil atribuição.

A Figura 34 mostra os espectros no infravermelho de uma das amostras (escolhida aleatoriamente) antes e após os ensaios de estabilidade. As regiões nas quais foram observadas alterações visíveis foram marcadas em função do tipo de vibração molecular associada. É importante mencionar que a Figura 34 seria ligeiramente diferente para outras amostras, mas, em todos os casos, há variações nas mesmas regiões do espectro.

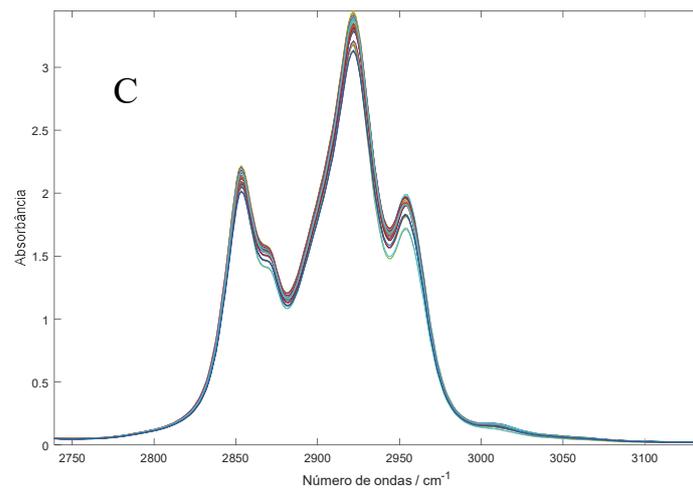
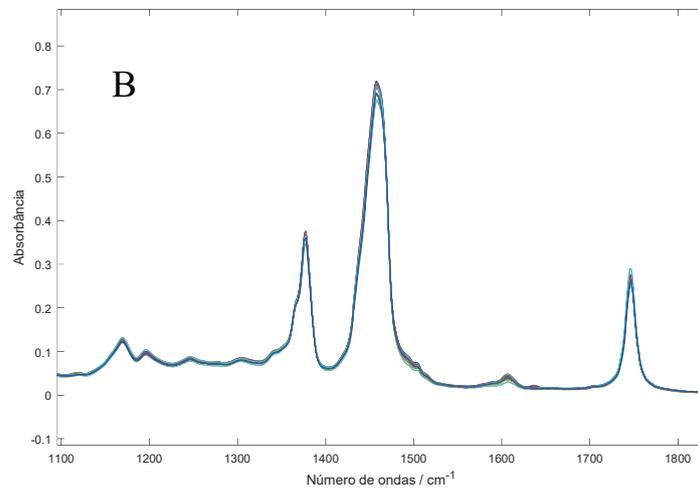
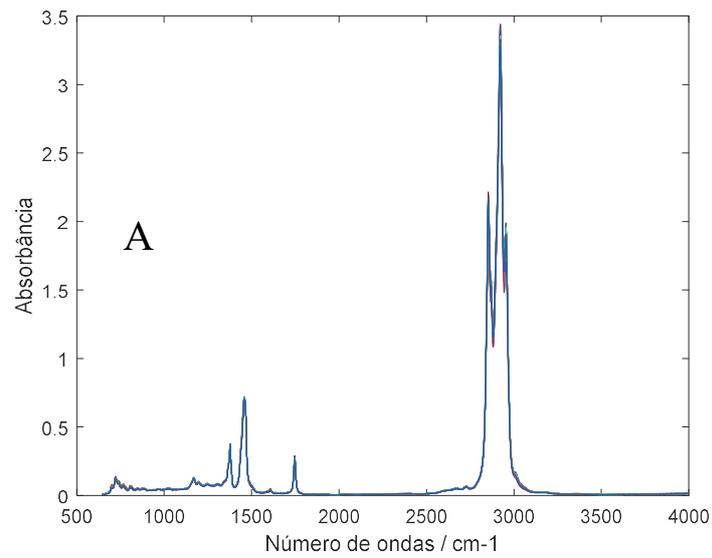


Figura 33: Espectros das 18 amostras (em duplicata) antes dos ensaios de estabilidade. Em **A**, toda a região estudada e, em **B** e **C**, aproximação de regiões com variações comentadas no texto.

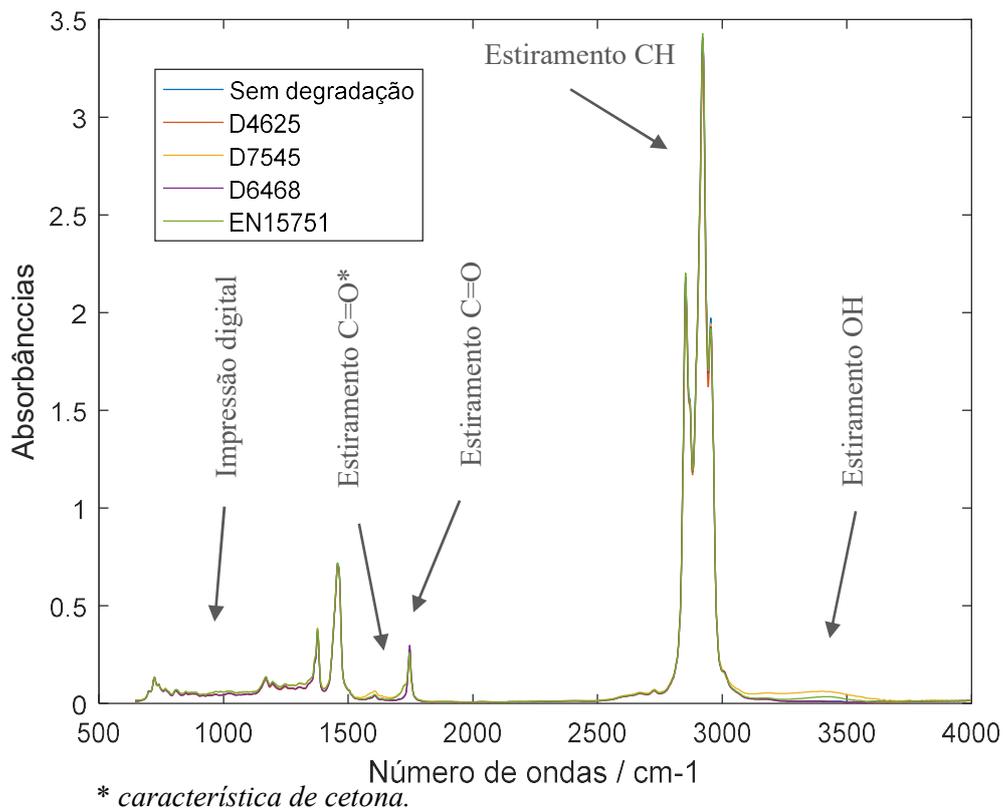


Figura 34: Espectros de uma das amostras antes e após todos os ensaios de estabilidade.

A Figura 35 mostra as médias dos espectros das amostras sem degradação e após cada ensaio de estabilidade.

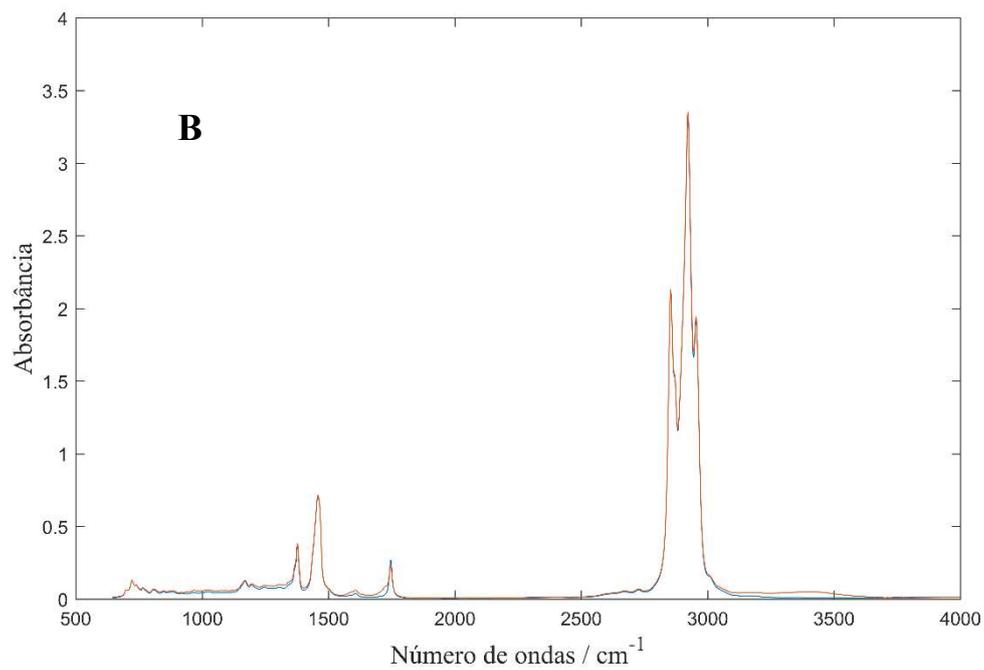
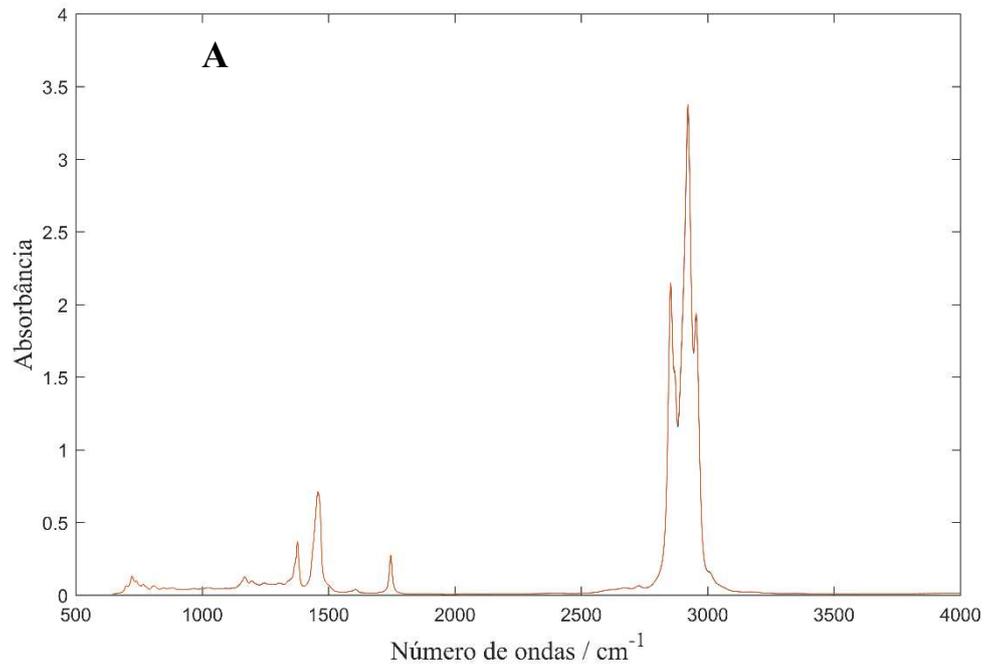
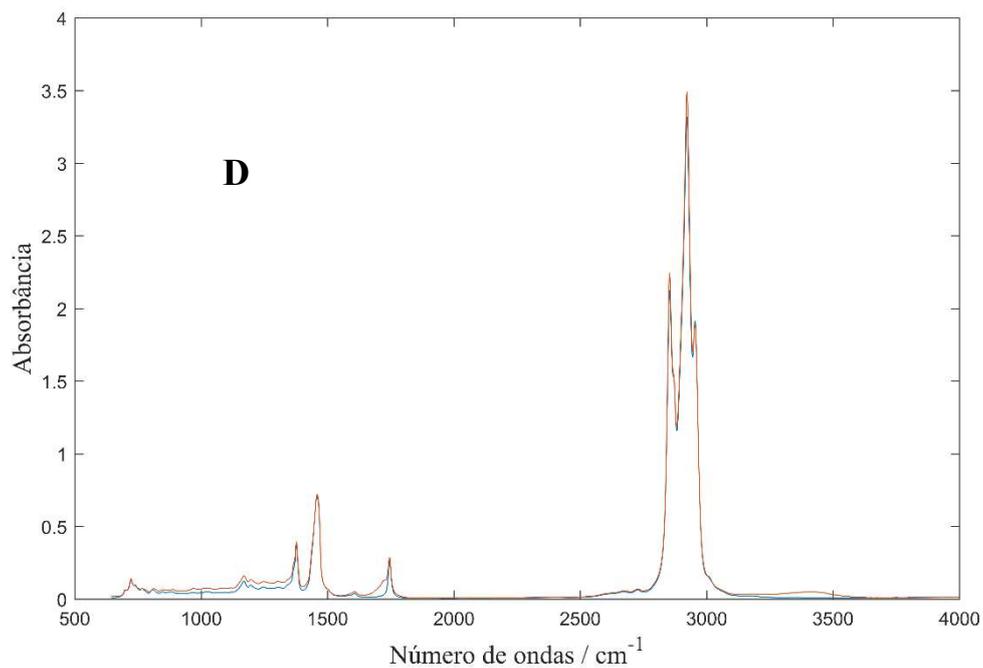
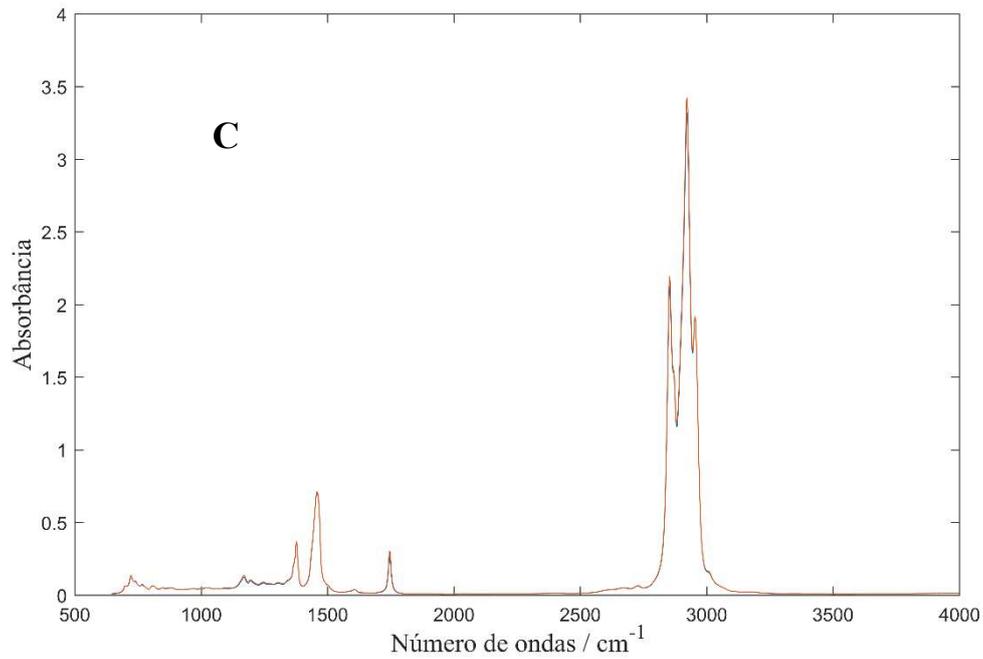


Figura 35: Comparação da média dos 36 espectros de óleo diesel antes dos ensaios (linha azul) com a média dos 36 espectros após ensaios de degradação (linha vermelha). **A**: comparação com o ensaio conforme norma ASTM D4625. **B**: comparação com o ensaio conforme norma ASTM D7545. **C**: comparação com o ensaio conforme norma ASTM D6468. **D**: comparação com o ensaio conforme norma CEN EN15751.

Continuação da Figura 35:



Com base na Figura 35, algumas observações podem ser feitas:

- O método D4625 parece não promover qualquer alteração significativa nos espectros. Ou as alterações são mascaradas pela variabilidade entre as amostras.
- O método D7545 promove alterações significativas, principalmente, na região de ca. 3450 cm<sup>-1</sup> que corresponde ao estiramento OH, em ca. 1717 cm<sup>-1</sup>, que pode ser indicativo de presença de cetonas e na região de impressão digital.

- O método D6468 parece não promover qualquer alteração significativa nos espectros.
- O método EN15751 parece promover as mesmas alterações que o método D7545 com menor intensidade na região de OH.

Vale acrescentar que a comparação visual dos espectros em duplicatas (verdadeiras) demonstraram boa reprodutibilidade de resultados. Ou seja, os espectros da mesma amostra que foram submetidos aos mesmos ensaios em duplicata são muito similares. Outra observação importante é que a não observação de mudanças nítidas no espectro de infravermelho pelos métodos D4625 e D6468 não implica a conclusão de que não ocorreram mudanças na composição da amostra, mas somente que as mudanças não puderam ser detectadas por inspeção visual dos espectros.

#### **5.4. Aplicação de PARAFAC nos dados de ATR-FTIR**

Após a organização dos dados de ATR-FTIR conforme Figura 36, a otimização do modelo PARAFAC foi realizada através da construção de modelos com diferentes números de componentes, modos de inicialização, critérios de convergência e restrições. Especialmente para o último parâmetro, é possível utilizar conhecimento a priori para tomada de decisão. Assim, decidiu-se por utilizar apenas a restrição de não negatividade para a dimensão dos espectros, já que não há sentido físico em absorbâncias negativas. Os modos de inicialização tiveram pouco ou nenhum impacto na qualidade dos modelos obtidos, o que indica que o conjunto de dados parece não apresentar mínimos locais ou degenerescência significativa das componentes calculadas. Todas as combinações dos outros tipos de restrições comuns (ortogonalidade e unimodalidade) foram testados nas três dimensões com diferentes números de componentes, fornecendo sempre valores muito baixos para CORCONDIA e muito altos para a Soma de Quadrados de Erros.

*Nota: Os resultados desta seção geraram um artigo na Revista Energy & Fuels de 2019<sup>154</sup>.*

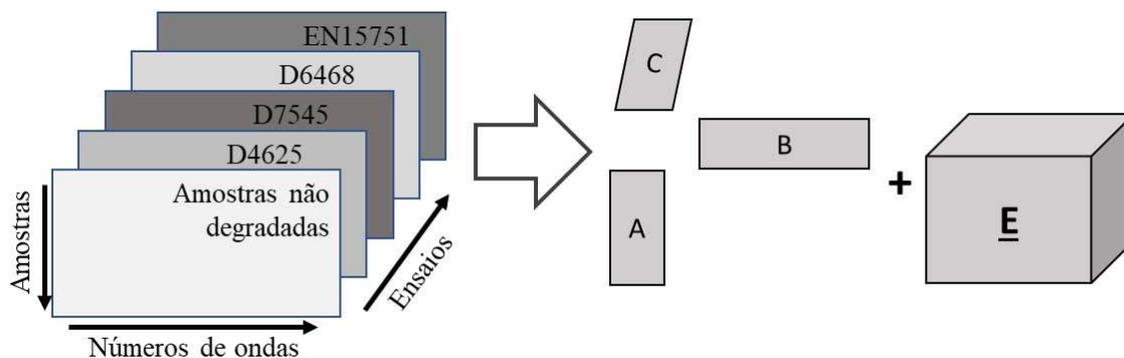


Figura 36: Organização dos dados para aplicação da ferramenta PARAFAC aos dados de ATR-FTIR.

A Tabela 6 mostra os valores de CORCONDIA e Soma de Quadrados de Erros, SQE, para modelos PARAFAC com diferentes números de componentes e fixando-se os seguintes parâmetros:

- Restrições: apenas não negatividade na dimensão espectral.
- Inicialização: DTLD (*Direct Trilinear Decomposition*).
- Critério de convergência: mudança relativa na soma de quadrados de erros entre iterações sucessivas:  $< 10^{-6}$ .

Tabela 6: Modelos PARAFAC com diferentes números de componentes.

Nº de componentes	Iterações	SQE (%)	CORCONDIA (%)
2	28	44,1	-0,9
3	195	31,5	0,3
4	80	19,1	38,0
5	69	14,6	32,2
6	308	7,4	71,0
7	263	5,0	89,7
8	805	3,9	46,7
9	351	3,3	0,1

Nota-se que o valor de SQE melhora continuamente, mas os valores de CORCONDIA pioram com a adição de mais que 7 componentes ao modelo. Um número de iterações excessivamente alto poderia ser um indicativo de um modelo instável, o que

não é o caso para este conjunto de dados. Se houvesse diferentes números de componentes fornecendo valores de similares de CORCONDIA, a SQE seria decisiva para escolha do melhor modelo. Contudo, neste conjunto de dados não foi necessário usar esse critério, pois o CORCONDIA indica 7 componentes como ideal. Foram testados modelos com números maiores de componentes, mas sempre se observou o surgimento de componentes degeneradas que não acrescentam informação relevante ao modelo.

A Tabela 7 mostra a variância capturada por componente para o modelo PARAFAC de 7 componentes. Note que, em função de como o algoritmo do programa funciona, a primeira componente captura o máximo de variância e as componentes subsequentes acumulam, gradativamente, menos variância. Chama a atenção o fato de que as últimas componentes capturam muito pouca variabilidade apesar de que, como será demonstrado, modelam informações relevantes.

Tabela 7: Variância capturada por componente do modelo PARAFAC

<b>Componente</b>	<b>% variância do modelo</b>
1	56,82
2	2,42
3	1,68
4	0,05
5	0,03
6	0,02
7	0,01

A Figura 37 mostra os pesos das 7 componentes da dimensão C, ou seja, na dimensão dos ensaios. Nota-se que as primeiras 3 componentes são similares para os cinco ensaios. Uma possível interpretação para esse fato é de que o modelo não estaria capturando variabilidade que está relacionada aos diferentes ensaios de estabilidade, mas às amostras. Em outras palavras, as três maiores fontes de variabilidade no conjunto de dados advêm das diferenças entre amostras, o que é concordante com a grande variação existente nas mesmas devido à origem em diferentes refinarias, cidades e estados.

Já a 4ª componente só é preponderante para o ensaio EN15751. Essa componente, portanto, captura prioritariamente a variabilidade especificamente gerada por este ensaio. De forma análoga, a 5ª componente está relacionada ao ensaio D7545, a 6ª componente às amostras não degradadas (possivelmente, modelando variabilidade que ainda não havia

sendo capturada nas 3 primeiras componentes) e a 7ª componente está relacionada ao ensaio D4625. O ensaio D6468 não tem uma componente específica, o que pode indicar que as transformações que esse ensaio promove não foram detectadas pelo espectro de infravermelho ou estão mascaradas pela variabilidade entre amostras.

Aqui fica clara a importância de incluir uma amostragem significativa no conjunto de dados e de realizar ensaios com replicatas. Somente nessas condições é possível diferenciar as transformações causadas pelos ensaios das variações entre amostras.

Vale mencionar que o sinal dos pesos não tem um sentido físico no modelo PARAFAC, mas o fato das componentes apresentarem sinais opostos aponta variações em sentidos opostos.

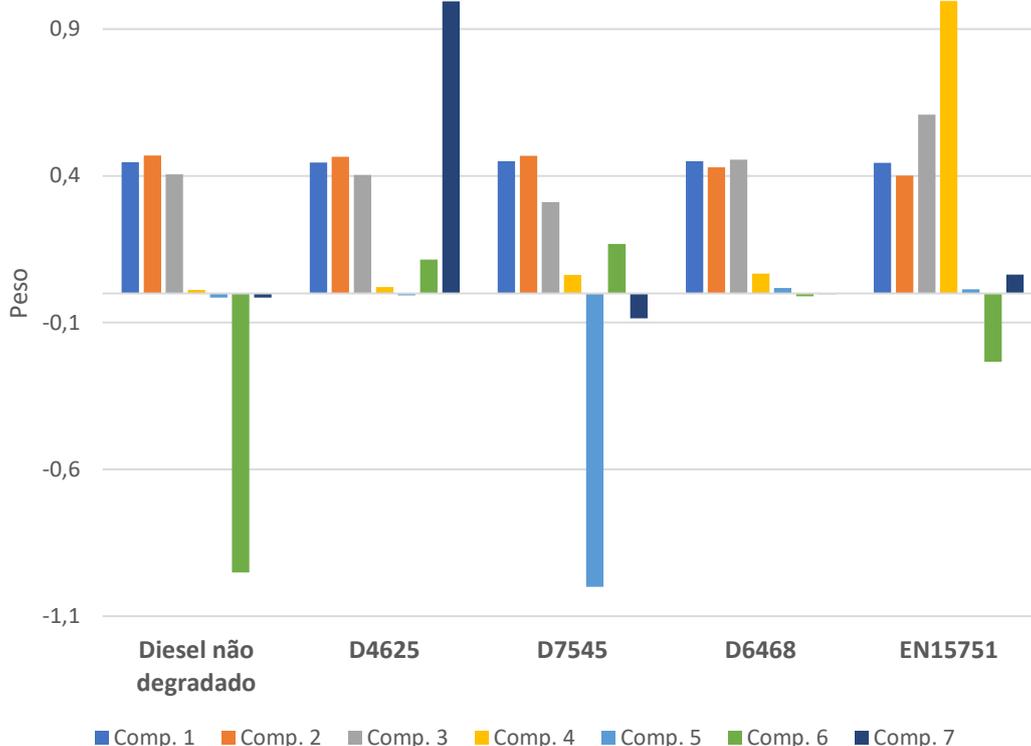


Figura 37: Modelo PARAFAC com 7 componentes – pesos da dimensão C (ensaios).

A Figura 38 mostra os pesos na dimensão espectral (matriz **B**) das três primeiras componentes. Os perfis lembram os espectros das amostras originais (especialmente na 1ª componente). Para a 2ª componente, o sinal correspondente ao estiramento C=O do biodiesel não aparece. Essa avaliação, em partes, corrobora com a interpretação feita para as 3 primeiras componentes na dimensão C: os espectros parecem estar relacionados às

*amostras* e não a uma característica particular impingida pelos ensaios de estabilidade. Além disso, pode-se inferir que as componentes 1 e 3 estão relacionadas ao biodiesel e sua variação nas amostras antes da degradação e a componente 2 ao óleo diesel antes da degradação.

A Figura 39 mostra os pesos da 4ª componente da dimensão espectral. Essa componente, está associada ao ensaio de estabilidade pela norma EN15751 (conforme avaliação da dimensão C). Assim, é possível avaliar quais transformações são preponderantes para esse ensaio. Nota-se que a região dos estiramentos C=O e OH (e interações intermoleculares) são bastante importantes nessa componente, assim como a região da “impressão digital”.

Essa observação é compatível com as transformações esperadas para o óleo diesel nas condições do ensaio, uma vez que o borbulhamento de ar e a temperatura de 110°C podem promover a oxidação dos ésteres insaturados a ácidos carboxílicos e álcoois. Eles também podem oxidar a ácidos os hidrocarbonetos que compõem o óleo diesel. O sinal correspondente ao estiramento CH pode estar associado à perda de frações mais voláteis (ricas em hidrocarbonetos).

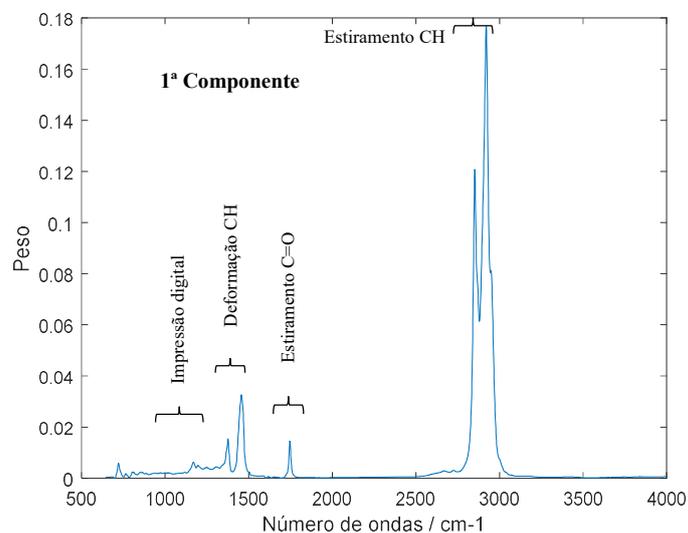
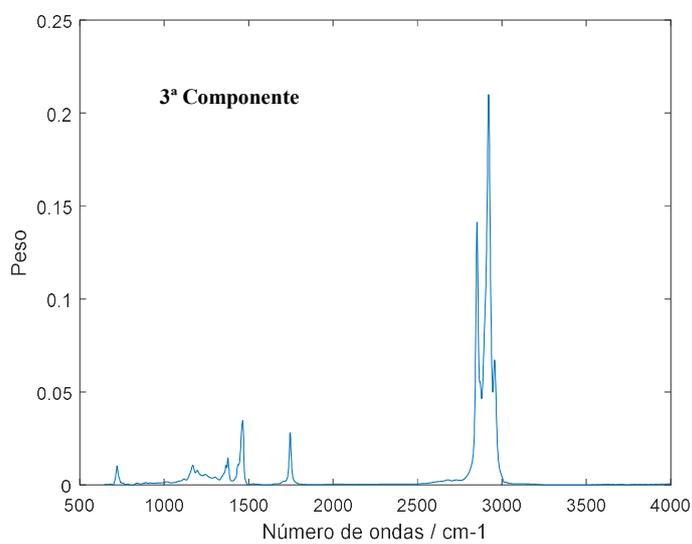
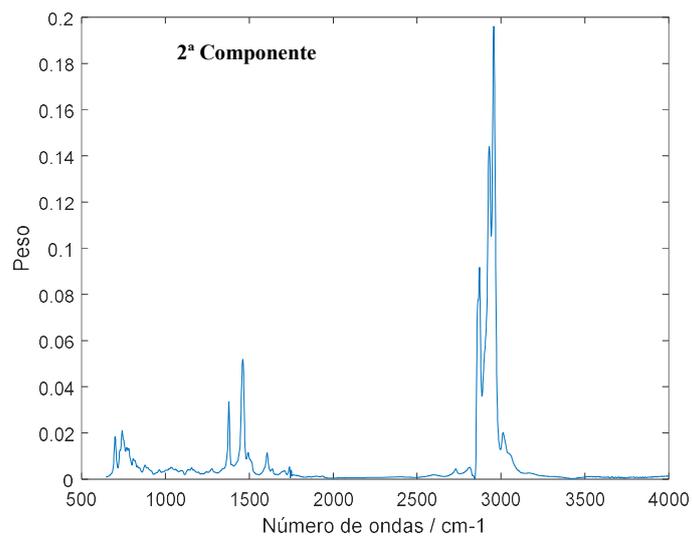


Figura 38: Modelo PARAFAC com 7 componentes - pesos das 3 primeiras componentes da dimensão B (espectros).

Continuação da Figura 38:



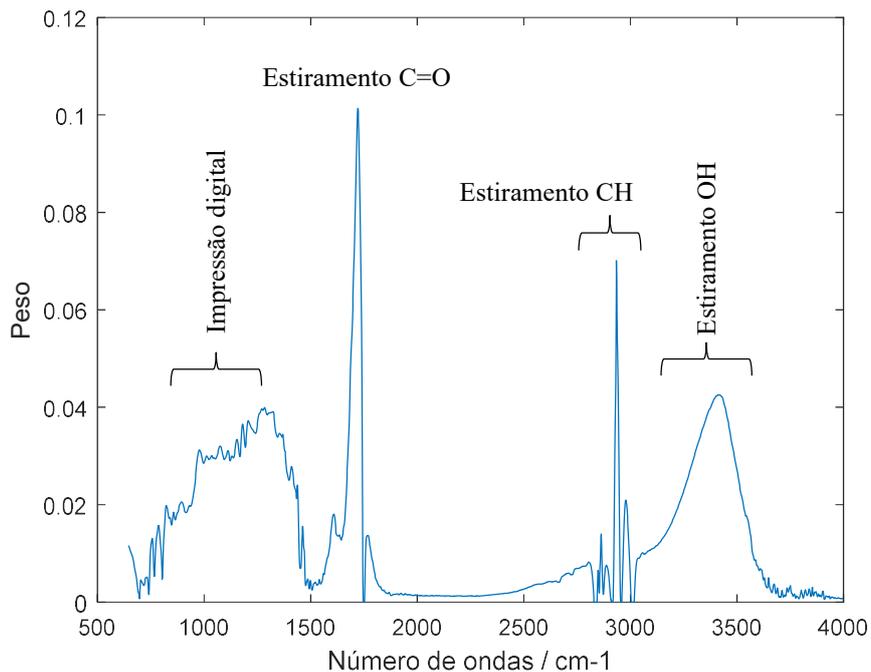


Figura 39: Modelo PARAFAC com 7 componentes – pesos da 4ª componente da dimensão dos espectros.

A 5ª componente (Figura 40) é associada ao ensaio de acordo com a norma D7545 e também tem características de sinal de carbonila, ainda mais importante para a região do estiramento OH, mas nenhuma participação de grupos CH. Neste ensaio, há grande propensão para oxidação devido à pressurização com O<sub>2</sub> e não há perda de material por volatilização como ocorre com o ensaio anterior, o que pode explicar a menor importância dos sinais de CH.

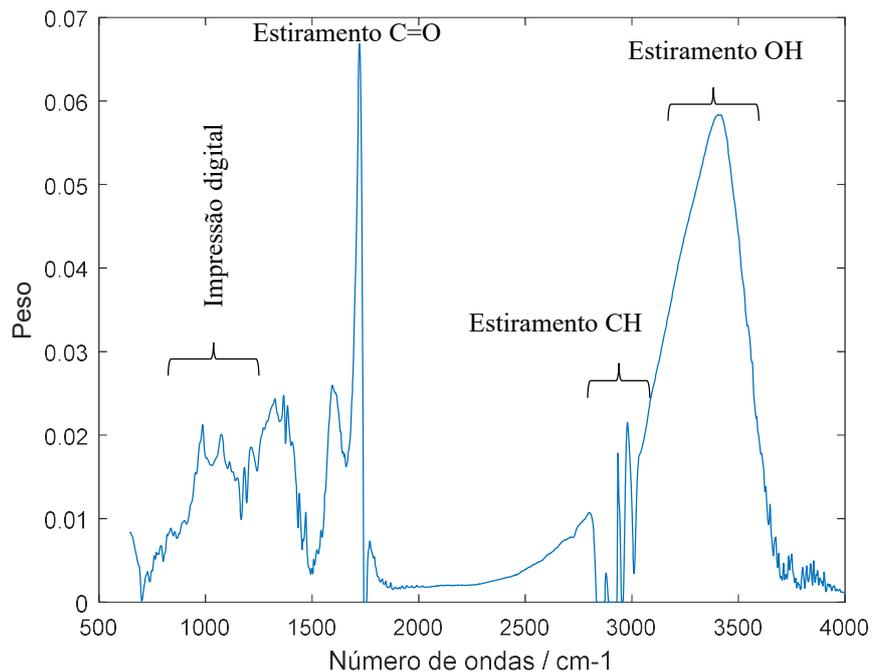


Figura 40: Modelo PARAFAC com 7 componentes – pesos da 5ª componente da dimensão dos espectros.

A 6ª componente (Figura 41) foi associada novamente à condição dos espectros não degradados, o que indicaria que se trata de parte da modelagem das variações entre amostras que não foi capturada com as três primeiras componentes. Mais uma vez, trata-se de um perfil muito similar ao espectro usual do óleo diesel.

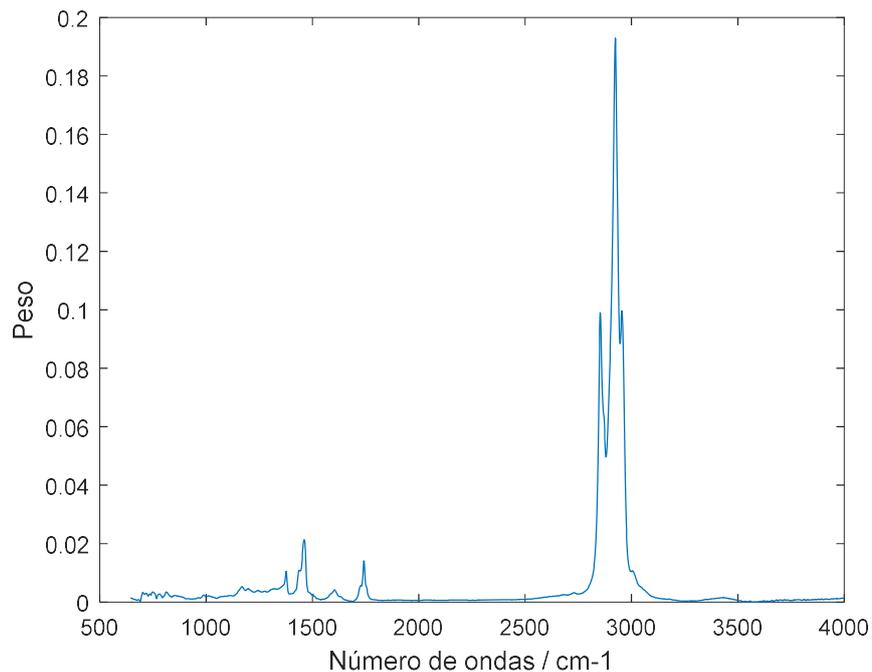


Figura 41: Modelo PARAFAC com 7 componentes – pesos da 6ª componente da dimensão dos espectros.

Por fim, a 7ª componente, mostrada na Figura 42, foi associada ao método de ensaio mais brando, o D4625, para o qual se espera uma preponderância de evaporação de hidrocarbonetos leves em relação às outras transformações. Essa informação é compatível com o perfil da 7ª componente, que é dominada pelos sinais de hidrocarbonetos parafínicos (estiramento e deformações CH).

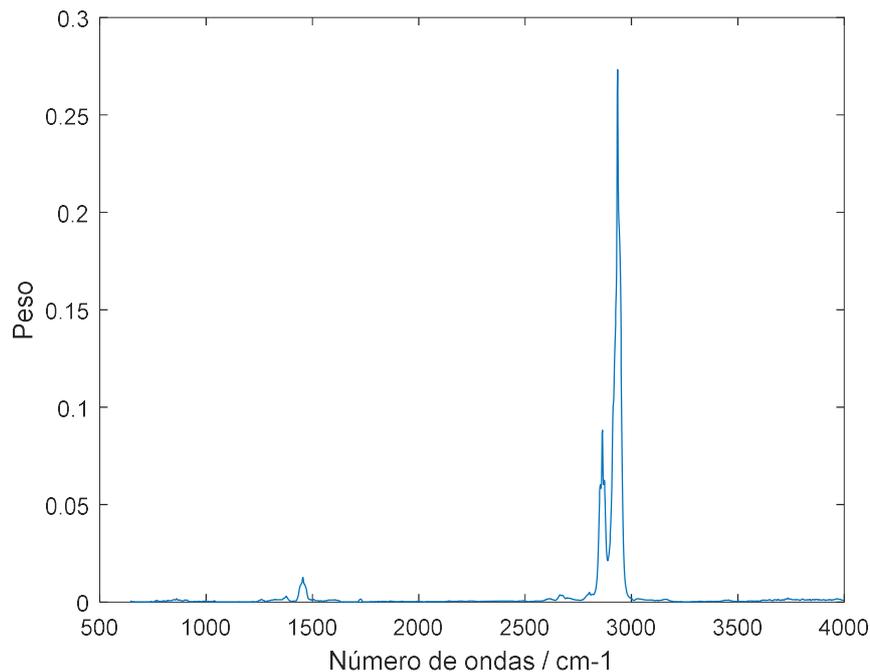


Figura 42: Modelo PARAFAC com 7 componentes – pesos da 7ª componente da dimensão dos espectros.

A Figura 43 mostra os pesos da dimensão das amostras (matriz A) que modela a variabilidade relacionada a cada amostra. Nota-se que a primeira componente tem os maiores pesos (e as demais têm pesos decrescentes), o que está relacionado com a forma como o algoritmo foi implementado: as componentes são ordenadas de acordo com a contribuição para o modelo, ou seja, de acordo com a proporção de variabilidade capturada pela componente.

Como os dados estão organizados com as duplicatas lado a lado, nota-se a formação de pares na Figura 43. Entretanto, há exceções: a amostra 10 na componente 1, por exemplo, está descolada de sua replicata. Por outro lado, esse mesmo par de resultados aparece agrupado quando observado em outra componente, mesmo observando as componentes uma a uma e com aproximação. Analisando outras componentes em separado, foi possível encontrar outros resultados discordantes entre replicatas. Esse comportamento demonstra que as transformações promovidas pelos ensaios de estabilidade são complexas e não tiveram uma elevada reprodutibilidade entre as replicatas para algumas amostras.

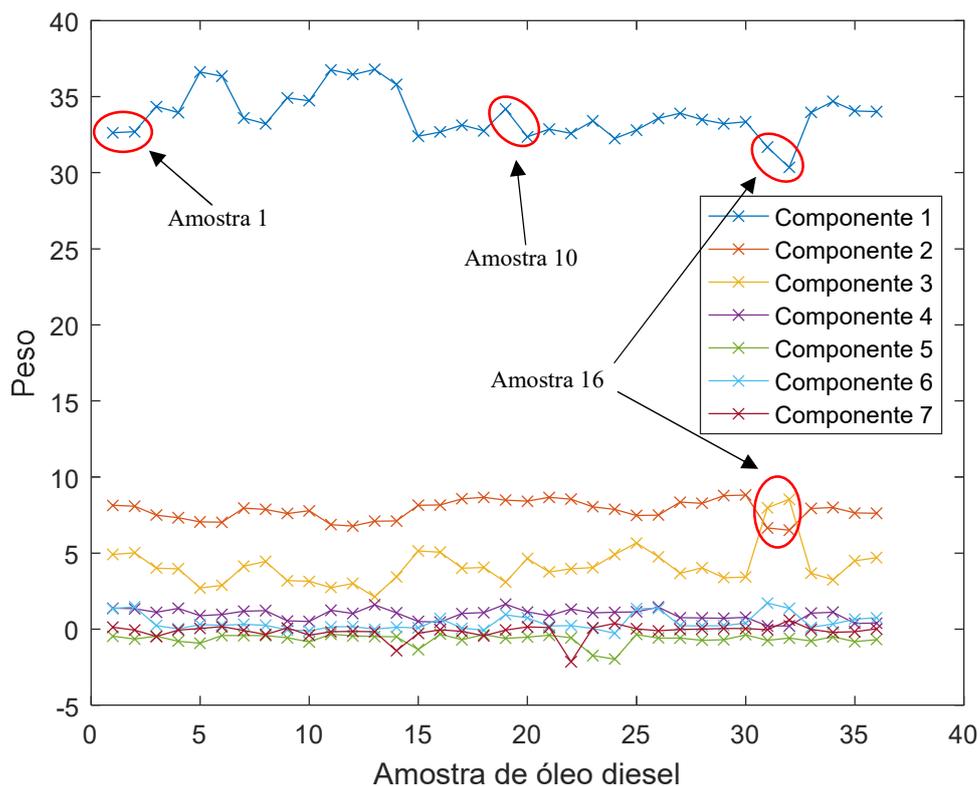


Figura 43: Pesos das componentes da dimensão das amostras para o modelo PARAFAC com 7 componentes. A numeração das amostras no eixo das abscissas considera duas replicatas por amostra.

Comparando as mesmas componentes nas três dimensões, é possível relacionar o comportamento das amostras (matriz A) nas dimensões espectral (matriz B) e de ensaios (matriz C). Assim, por exemplo, investigou-se a amostra 16 (numerações 31 e 32 na Figura 44) que se destaca nas componentes 1, 2, 3 e, com menos intensidade, na 6. Essa amostra já foi comentada no subitem 5.2 por apresentar comportamento anômalo para o ensaio de estabilidade conforme norma CEN EN15751 (não foi possível obter o resultado deste ensaio devido à elevada estabilidade da amostra). Além disso, essa amostra forneceu o maior resultado para o ensaio conforme D7545. Como as quatro componentes comentadas estão associadas às características do diesel (e não de qualquer ensaio específico), entende-se que seu comportamento é oriundo da sua composição e não de aditivos ou uma substância específica. Mesmo não sendo possível estabelecer uma relação de causa e efeito, chama a atenção o fato de que a amostra 16 é a de maior massa específica entre todas ( $830,5 \text{ kg/m}^3$ ).

A amostra 12 (e sua duplicata) destaca-se na 5ª componente conforme Figura 44. Essa componente está associada ao ensaio conforme D7545, o que poderia indicar um comportamento diferenciado neste (e somente neste) ensaio. A avaliação dos resultados dos ensaios de estabilidade não indicou qualquer característica ou comportamento excepcional dessa amostra. A princípio, o modelo PARAFAC indica que essa amostra foi a que mais sofreu modificações no ensaio D7545, provavelmente devido à sua oxidação.

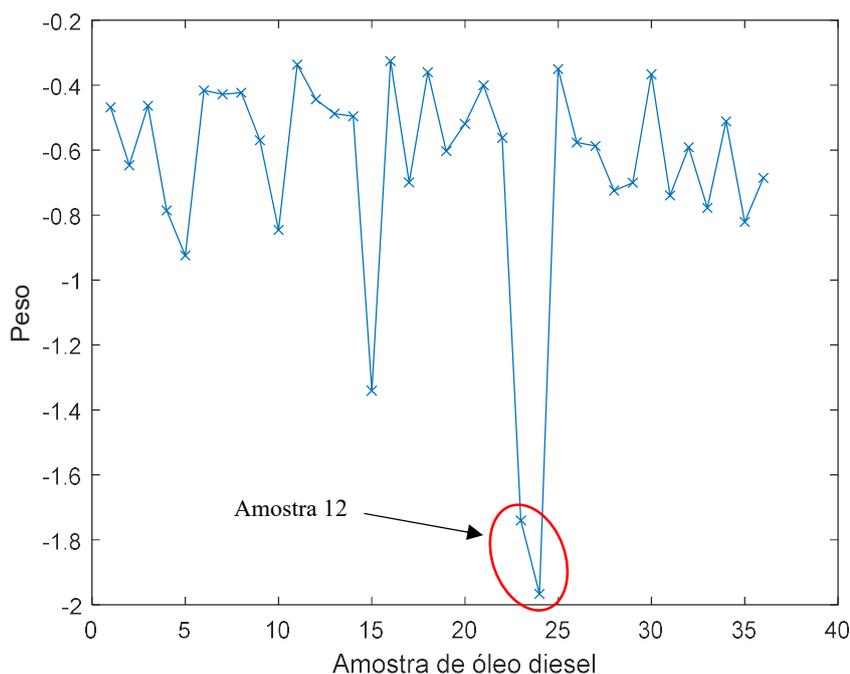


Figura 44: Pesos das amostras na 5ª componente no modelo PARAFAC com 7 componentes.

O gráfico com os pesos das componentes 1 e 2 da dimensão das amostras é mostrado na Figura 45. A amostra 16, como esperado, se destaca das demais, mas também é possível visualizar um aglomerado com as amostras 3, 6 e 7. Assim como a amostra 16 tem a menor massa específica, as três amostras mencionadas têm os maiores valores para essa característica (847,1, 846,1 e 841,5 kg/m<sup>3</sup>, respectivamente). Nota-se, portanto, que a 1ª componente é dominada por compostos de alta densidade, o que reforça a interpretação de que as primeiras componentes são dominadas pelas características das amostras (e não dos ensaios de estabilidade).

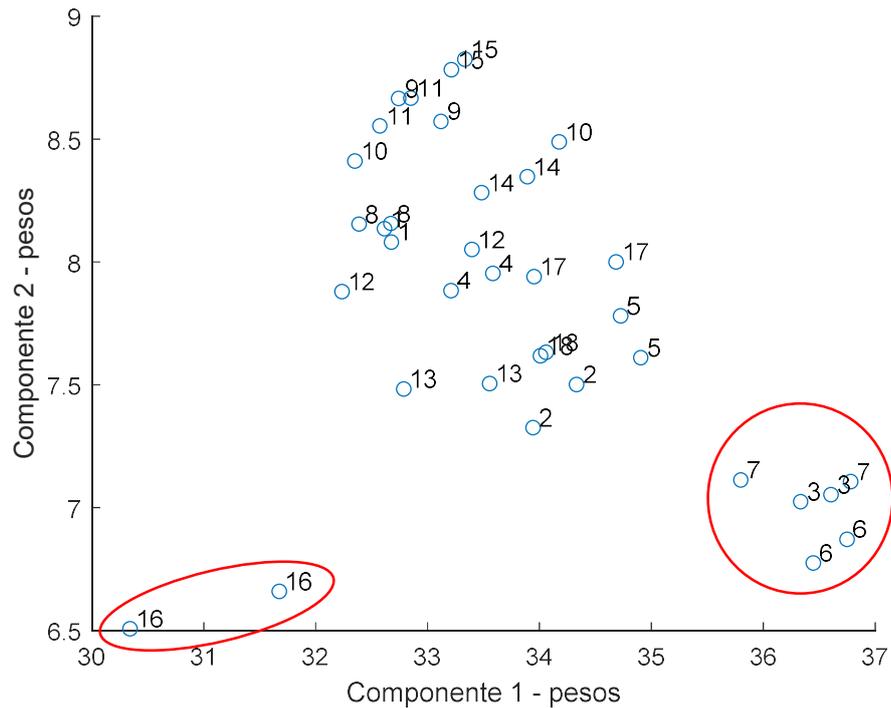


Figura 45: Pesos das amostras nas componentes 1 versus 2 no modelo PARAFAC com 7 componentes.

### 5.5. GC-MS

Conforme descrito no subitem 4.5, três amostras – antes e após os ensaios de estabilidade – foram analisadas utilizando-se GC-MS. O Cromatograma de Íons Totais, TIC, de uma das amostras antes dos ensaios de degradação é mostrado na Figura 46-A com a identificação dos principais picos realizada conforme descrito no subitem 4.5. Na Figura 46-B, os TICs das três amostras antes dos ensaios são mostrados com deslocamento de linha-base para facilitar a comparação. Observa-se que as amostras, apesar de terem sido coletadas em regiões muito diferentes, apresentam, basicamente, a mesma composição. As alterações mais perceptíveis se referem à altura dos picos correspondentes ao biodiesel - o que pode estar ligado à matéria prima utilizada ou ao processos de refino: palmitato, linoleato, oleato e estearato de metila.

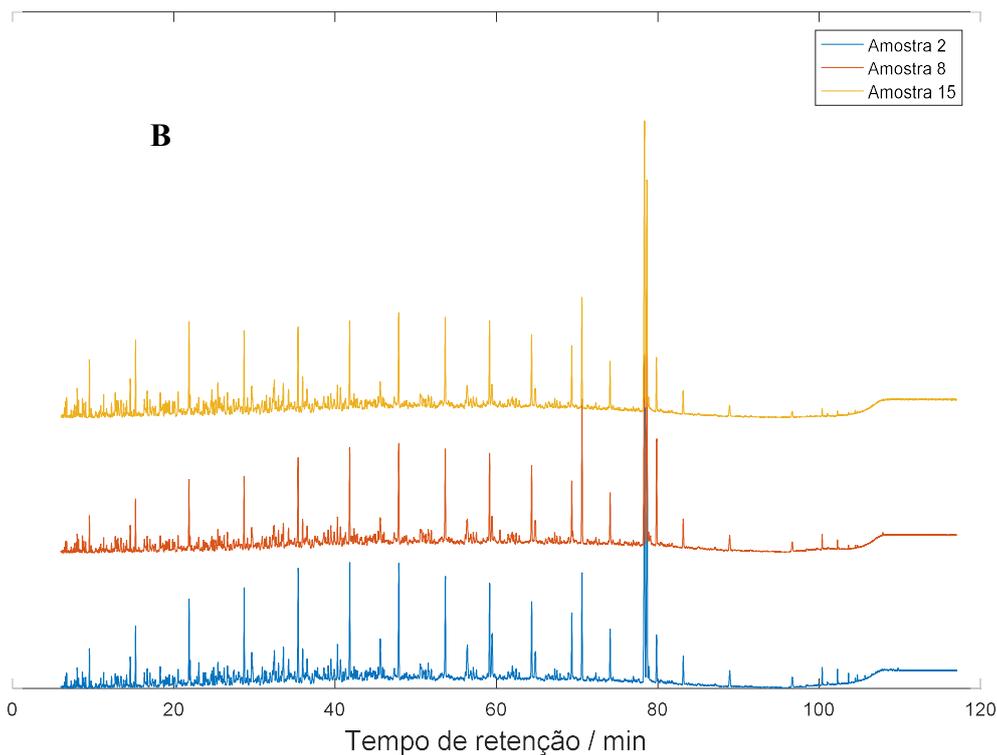
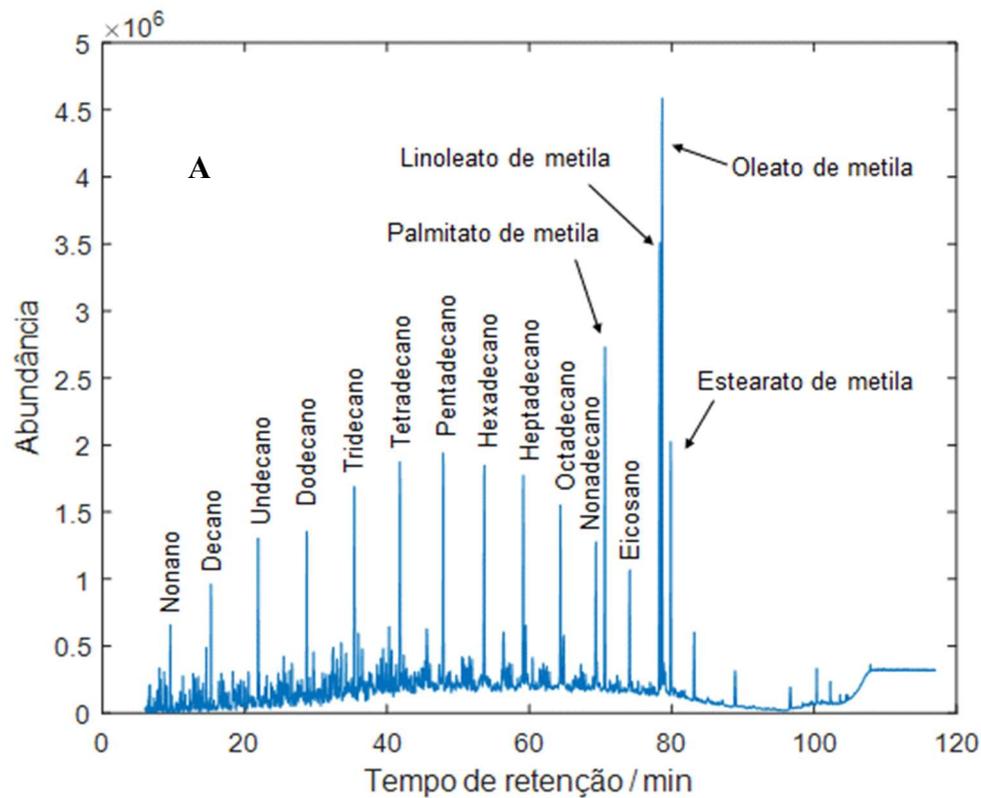


Figura 46: **A.** Cromatograma de Íons Totais, TIC, de amostra de óleo diesel antes de qualquer ensaio de degradação. A atribuição dos picos segue o procedimento descrito no subitem 4.5. **B.** TICs das três amostras (antes dos ensaios) com deslocamento de linha-base para facilitar a comparação.

Os TICs foram avaliados visualmente buscando-se picos que tenham surgido, diminuído ou aumentado na comparação dos TICs antes e após os ensaios de estabilidade. Os espectros correspondentes a esses picos foram, então, comparados com os espectros do banco de dados NIST 02. Vale mencionar que as avaliações comparativas podem ser prejudicadas por pequenas diferenças de diluição durante a silanização das amostras antes dos ensaios cromatográficos. Adicionalmente, nos ensaios em que há evaporação de frações do combustível, pode ocorrer aumento das áreas dos picos das substâncias que não foram volatilizadas.

Essa avaliação levou às seguintes constatações:

- O ensaio de estabilidade pela norma ASTM D4625 causou apenas leve diminuição dos picos de baixo tempo de retenção (ver Figura 47). Isso leva a crer que as alterações significativas, se existirem, envolvem compostos que não podem ser detectados nas condições do ensaio, seja por envolver compostos que não podem ser analisados por GC – mesmo com derivatização – ou porque eles estão em baixa concentração e/ou mal resolvidos.
- O ensaio de estabilidade pela norma ASTM D7545 promoveu o surgimento de picos em 20,4 e 58,1 minutos. Esses picos são destacados na Figura 48, mas apenas o primeiro composto pôde ser identificado com base na biblioteca de espectros do equipamento: o composto foi identificado como sendo o trimetilsilil éster do ácido hexanóico. Entende-se que é o produto da silanização do ácido hexanóico, composto que pode ser formado na degradação do biodiesel, especificamente, por hidrólise de hexanoato de metila. Já os picos em 78,3 e 78,6 min., atribuídos ao linoleato de metila e ao oleato de metila, respectivamente, apresentaram diminuição de área. O primeiro composto, entretanto, apresentou uma diminuição de altura de pico muito maior (36 % contra 7 %, respectivamente, para a amostra selecionada) o que é coerente com os dados da literatura que indicam que os compostos insaturados do biodiesel são suscetíveis à oxidação e, que quanto mais insaturados, mais suscetíveis o são. Apesar de pequenas variações quantitativas, essas observações são as mesmas para as três amostras. É importante observar que não houve alterações significativas dentre os compostos mais leves, o que é coerente com o fato de que o ensaio é realizado num vaso fechado que não permite perda de material.

- O ensaio conforme norma D6468 promoveu a perda de compostos mais voláteis. Nota-se que, aproximadamente, até 20 minutos de corrida cromatográfica, todos os picos do TIC da amostra submetida a esse ensaio têm significativa diminuição de abundância. Quando possível, esses compostos foram identificados, na amostra não estressada, como hidrocarbonetos leves. Entretanto, em 8,7 min, para as três amostras, há um pico que não pôde ser identificado com base na biblioteca de espectros (ver Figura 49-A), que pode ser oriundo dos reagentes de silanização. Em 34,2 e 42,2 min, e apenas para a amostra 2, ocorreu a diminuição das intensidades dos picos identificados, respectivamente, como 1,2,3,4-tetrahidro-6-metil-naftaleno e 2,7-dimetilnaftaleno (ver Figura 49-B e Figura 49-C). Esse comportamento particular da amostra 2 pode indicar falta de reprodutibilidade do método.

Não houve alteração significativa nos picos dos ésteres, ao contrário do observado no ensaio D7545, indicando que este ensaio não tem características tão oxidantes quanto aquele.

- O ensaio conforme EN15751 promoveu a quase completa supressão dos compostos até, aproximadamente, 25 min de corrida. E, mesmo até, aproximadamente, 35 min, os picos estão significativamente diminuídos. Quando possível, esses compostos foram identificados, na amostra não estressada, como hidrocarbonetos leves. Na Figura 50-A, observa-se a supressão dos compostos voláteis e o pico surgido em 8,7 min (que, conforme discutido acima, não pôde ser identificado com base na biblioteca de espectros). Em **B**, observa-se que apenas o pico correspondente ao linoleato de metila sofreu redução significativa (ficando o pico do oleato de metila, em 78,6 min, visualmente inalterado). Essa redução foi de 55 %, ainda mais intensa, portanto, que a observada para o ensaio conforme D7545. Na Figura **50-C** destacam-se alterações ocorridas apenas para a amostra 2: em 34,2 min houve diminuição do pico identificado como sendo 2,3-dihidro-1,1,3-trimetil-1-H-indeno e, em 34,6 min, houve redução de um pico cuja identidade não pôde ser determinada. Na Figura **50-D** observou-se o surgimento de uma série de pequenos picos que não puderam ser identificados, exceto aquele destacado em 88,2 min, que corresponde ao trans-9,10-epoxiestearato de metila,<sup>48</sup> um composto intermediário no processo de autooxidação do oleato de metila<sup>48</sup>.

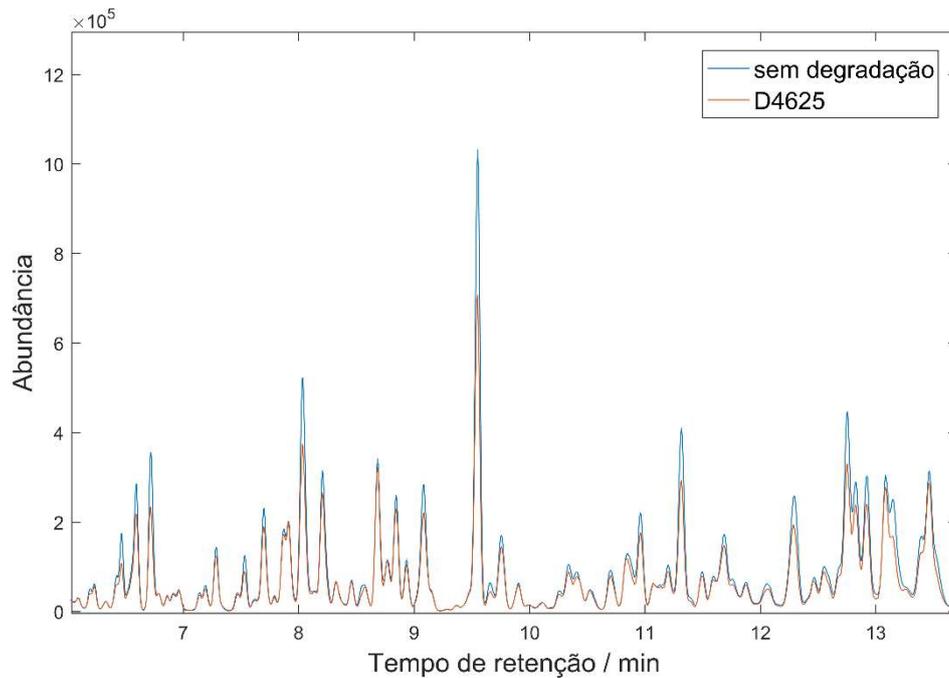


Figura 47: Comparação da parte inicial dos TICs da amostra 8 antes e após o ensaio conforme norma D4625.

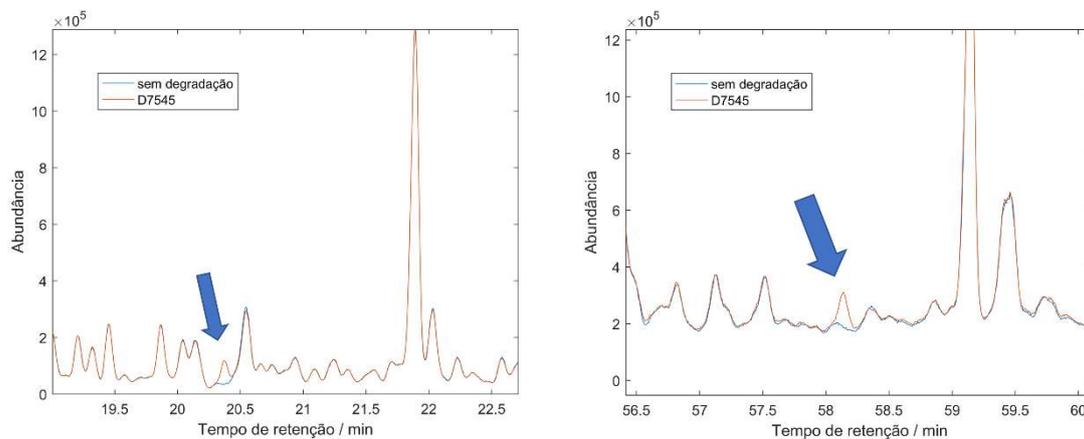


Figura 48: Alterações observadas na comparação da amostra 8 antes e após o ensaio de estabilidade de acordo com a norma ASTM D7545. Note a maior supressão do sinal para o linoleato de metila em 78,3 min. O mesmo comportamento foi observado nas demais amostras.

Continuação da Figura 48:

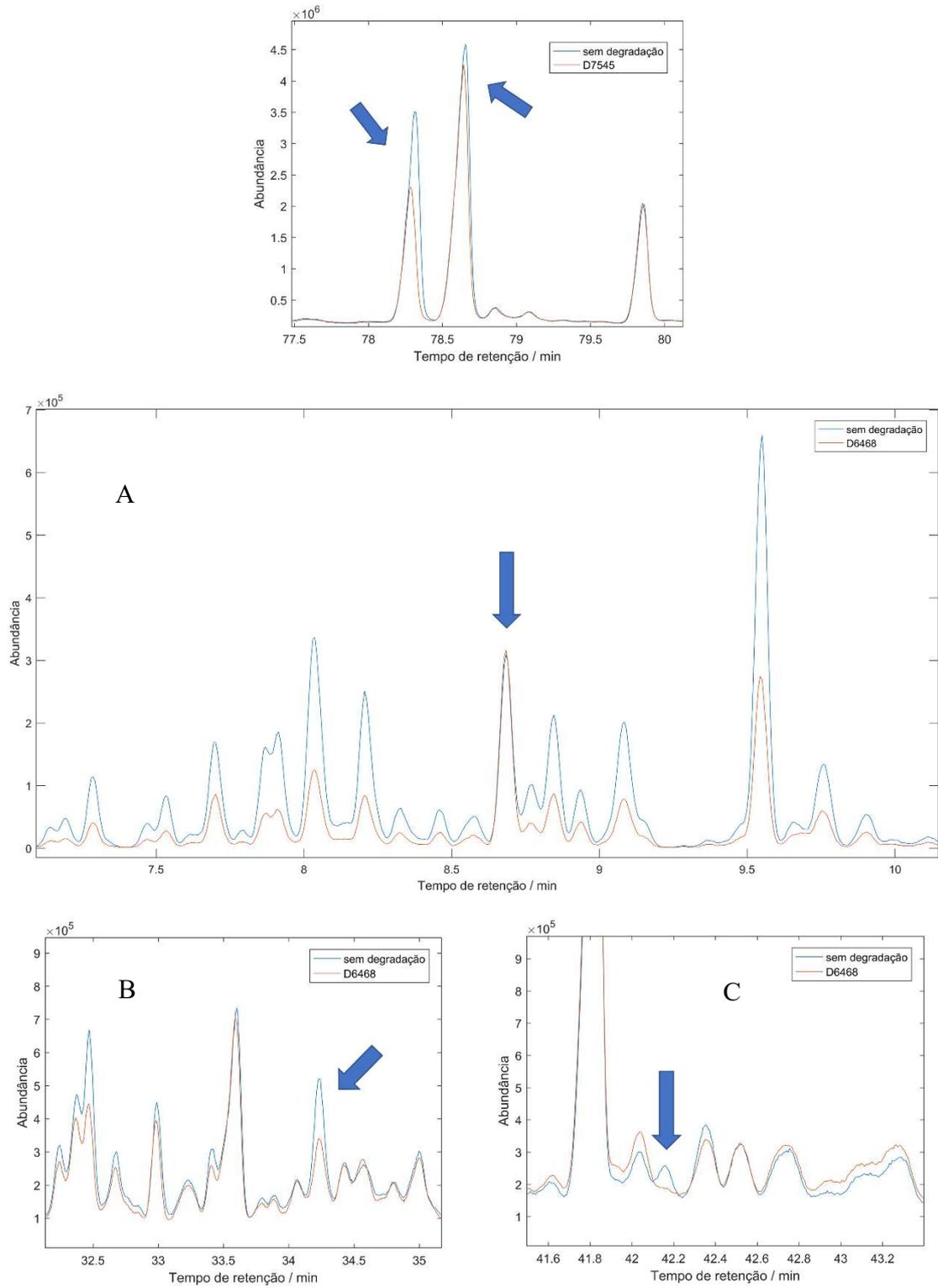


Figura 49: Alterações observadas após o ensaio de acordo com a norma D6468 em: A. amostra 8 e, em B. e C. amostra 2. Note a diminuição geral dos picos em A.

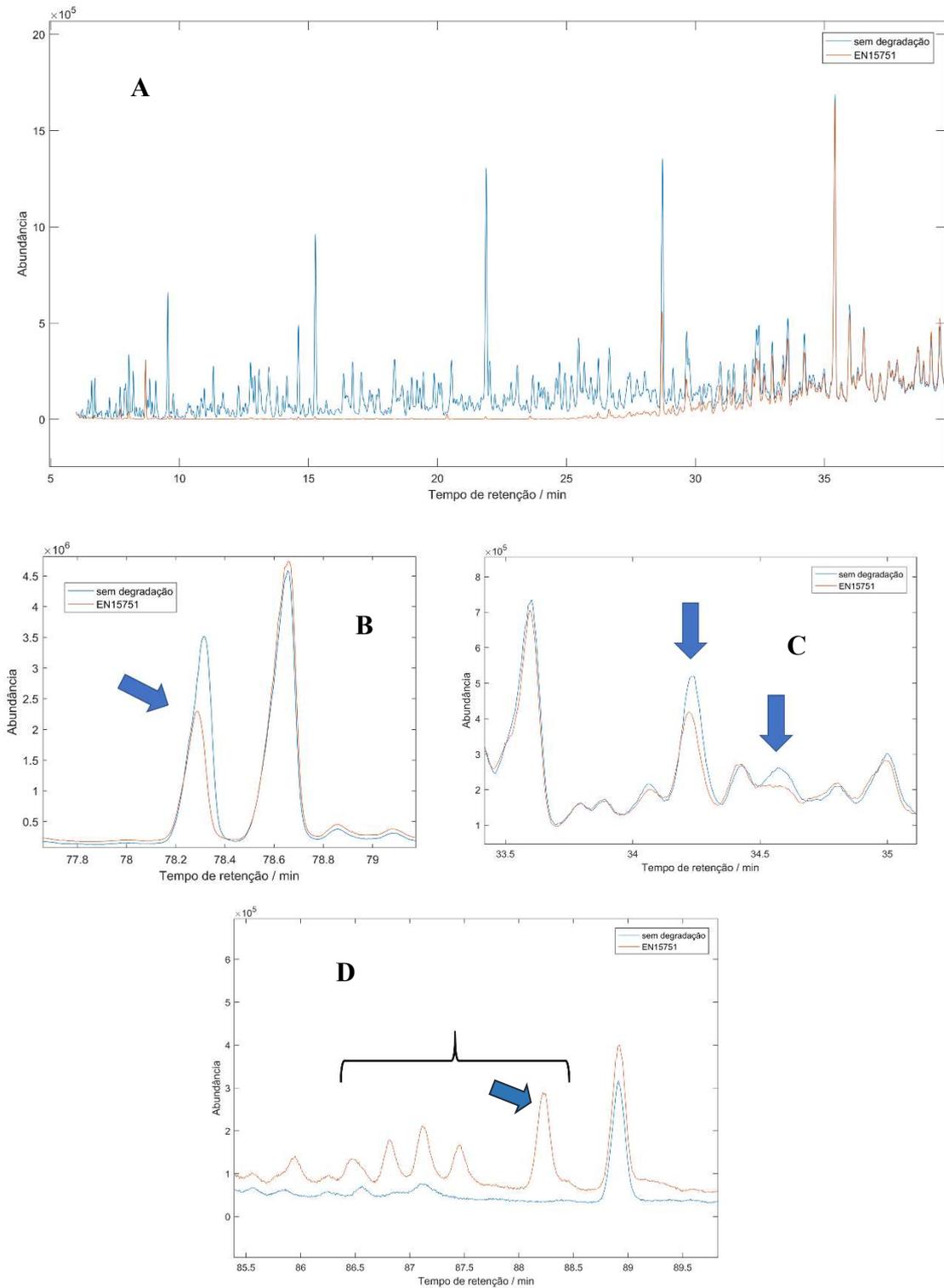


Figura 50: Alterações observadas para o ensaio conforme norma CEN EN15751: Em A, observa-se a supressão quase completa de compostos voláteis e o pico surgido em 8,7 min. As alterações em C só foram observadas para a amostra 2 e as destacadas em B e D são comuns às três amostras. Os picos destacados são comentados no texto.

## 5.6. Aplicação de MCR-ALS nos dados de GC-MS

Os dados de GC-MS foram agrupados na direção das colunas (“column-wise”) conforme Figura 51, fornecendo uma matriz aumentada de 257.460 linhas por 571 colunas, sendo que cada linha corresponde a um espectro de massas medido em um determinado tempo de retenção e cada coluna a uma razão massa-carga (entre 30 e 600 Da).

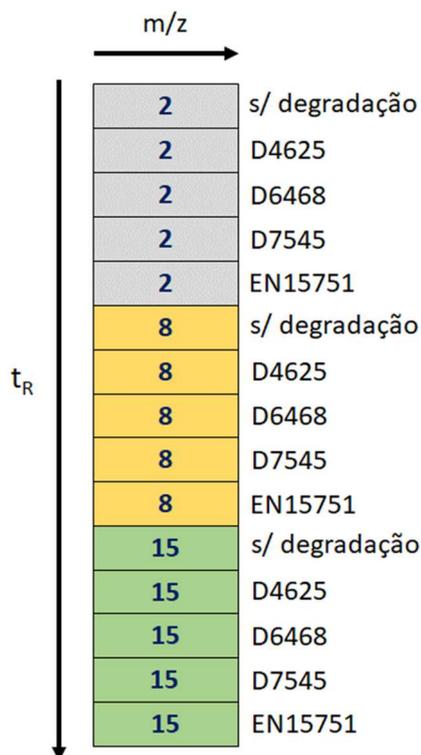


Figura 51: Organização dos dados de GC-MS para aplicação de MCR-ALS. Os números nos retângulos identificam a amostra ensaiada.

A aplicação de MCR-ALS foi realizada com os parâmetros a seguir:

➤ Estimativa inicial:

A estimativa inicial foi feita na matriz **C** (que contém as intensidades relativas) por meio do método de pureza de variáveis puras SIMPLISMA (descrito em 2.7.1.4.3).

➤ Restrições:

O uso de não negatividade em **C** e **S** é praticamente mandatório pela natureza das grandezas utilizadas. Utilizou-se a opção “force to zero” para implementação dessa restrição. O método FNNLS foi tentado, mas foi abandonado por apresentar

tempos computacionais proibitivos (o algoritmo requereu *dias* de processamento para cada modelo). Além disso, testes preliminares indicaram que FNNLS não gerava modelos significativamente melhores.

A unimodalidade foi testada em **C** com o objetivo de obter perfis de intensidades relativas (perfis cromatográficos) melhores. Dentre os algoritmos *vertical*, *horizontal* e *médio*, o que forneceu os melhores valores de variância capturada e falta de ajuste foi o último. Entretanto, os espectros de massas recuperados apresentaram picos deslocados ou ausentes em relação aos espectros de referência. Sem o uso de unimodalidade, os espectros ficaram bastante próximos dos espectros de referência como será detalhado adiante. Portanto, foi decidido não empregar a restrição de unimodalidade no modelo final.

➤ Normalização:

Decidiu-se pela normalização da matriz **S** (espectros puros) utilizando-se norma euclidiana que permite que as variações devido às concentrações dos componentes puros sejam contempladas na matriz **C**.

Para definição do número de componentes, o modelo MCR foi calculado – de forma exploratória – utilizando-se um grande número de componentes: 50. Apesar desse número de componentes ser elevado, deve-se observar que o número de espécies químicas presentes e detectadas na amostra de diesel é consideravelmente superior a esse número de componentes. Cada componente obtida foi avaliada visualmente nas matrizes **C** e **S**, concluindo-se que algumas componentes apresentaram perfis de concentrações recuperados com picos únicos em tempos de retenção correspondentes aos ésteres que compõem o biodiesel (conforme identificado por GC-MS). Os perfis espectrais dessas componentes foram coerentes com as identificações feitas com base em tempos de retenção conforme será detalhado adiante. Por outro lado, dentre os 50 perfis de concentrações recuperados, a grande maioria apresentou vários picos e perfis espectrais recuperados que não apresentaram nenhuma correspondência com o banco de dados espectrais utilizado<sup>95</sup> ou por comparação com dados da literatura<sup>62,69</sup>. Além disso, houve várias componentes que pareciam modelar hidrocarbonetos (essa afirmação se baseou nos tempos de retenção e perfis espectrais recuperados).

Dessa forma, decidiu-se testar novos modelos com números menores de componentes, avaliando-se as componentes obtidas e a qualidade geral dos seus ajustes.

A Tabela 8 mostra os valores obtidos de falta de ajuste, coeficiente de determinação,  $R^2$ , e número de iterações para esses modelos.

Tabela 8: Parâmetros de modelos MCR-ALS obtidos com diferentes números de componentes.

<b>Nº de componentes</b>	<b>Falta de ajuste %</b>	<b><math>R^2</math></b>	<b>Iterações</b>
4	31,0	90,4	9
5	27,4	92,5	10
6	24,8	93,9	19
7	22,8	94,8	15
8	21,7	95,3	14
9	20,5	95,8	12
10	19,6	96,1	15
15	16,4	97,3	18
18	15,7	97,5	10
19	14,4	97,9	17
20	14,1	98,0	26
50	13,7	98,1	21

Nota-se que, como esperado, a falta de ajuste diminui e o  $R^2$  aumenta com o aumento do número de componentes, entretanto, este comportamento não é abrupto, o que dificulta a decisão sobre o número de componentes ideal.

Outra ferramenta que pode auxiliar nessa decisão é a Decomposição de Valores Singulares descrita no item 2.7.1.4.2. O gráfico dos autovalores versus autovetores mostrado na Figura 52 indica que, aproximadamente, a partir da quinta componente, os incrementos de variabilidade capturada são pequenos. Entretanto, modelos obtidos com poucas componentes não conseguiram fornecer perfis de concentração recuperados com picos únicos. Vale lembrar que os autovalores não se referem ao modelo MCR, mas aos autovetores obtidos por decomposição em valores singulares.

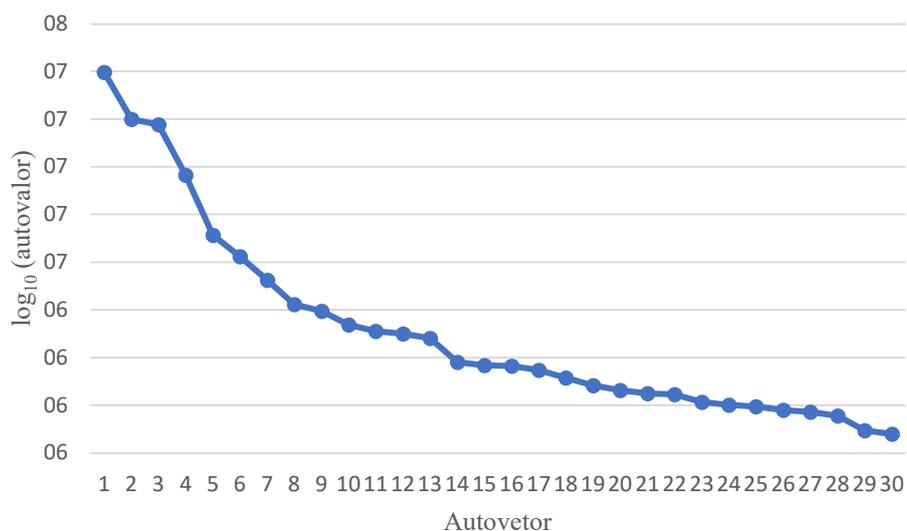


Figura 52: Autovalores (em escala logarítmica) obtidos por SVD para o conjunto de dados de GC-MS.

A decisão sobre o número de componentes do modelo final recaiu na análise qualitativa das componentes uma a uma de modelos com diferentes números de componentes. O modelo com 19 componentes forneceu picos praticamente unitários nos perfis de concentrações correspondentes aos ésteres que compõem o biodiesel, além de outras componentes com características interessantes como será detalhado adiante. Modelos com mais componentes não forneceram perfis de concentração ou espectrais interpretáveis. Por outro lado, modelos com 18 ou menos componentes não conseguem gerar componentes com picos unitários para os principais ésteres do biodiesel. Este acabou sendo o critério de decisão para escolha do modelo com *19 componentes*.

Antes de iniciar a avaliação do modelo obtido, é importante enfatizar que toda a discussão deste tópico foca em compostos que podem ser analisados por GC (com silanização), o que pode excluir muitos compostos de alto peso molecular.

A Figura 53-A mostra o perfil de intensidades relativas, que é a estimativa do perfil cromatográfico, obtido para a primeira componente (primeira coluna da matriz **C**) do modelo MCR-ALS. Essa figura mostra de forma conjunta o perfil cromatográfico estimado para as 15 corridas cromatográficas. Por outro lado, a Figura 53-B apresenta um destaque da Figura 53-A mostrando apenas o perfil recuperado da amostra 2 sem degradação. Comparando-se o tempo de retenção do TIC da Figura 46-A, deduz-se que essa componente captura/modela o composto palmitato de metila.

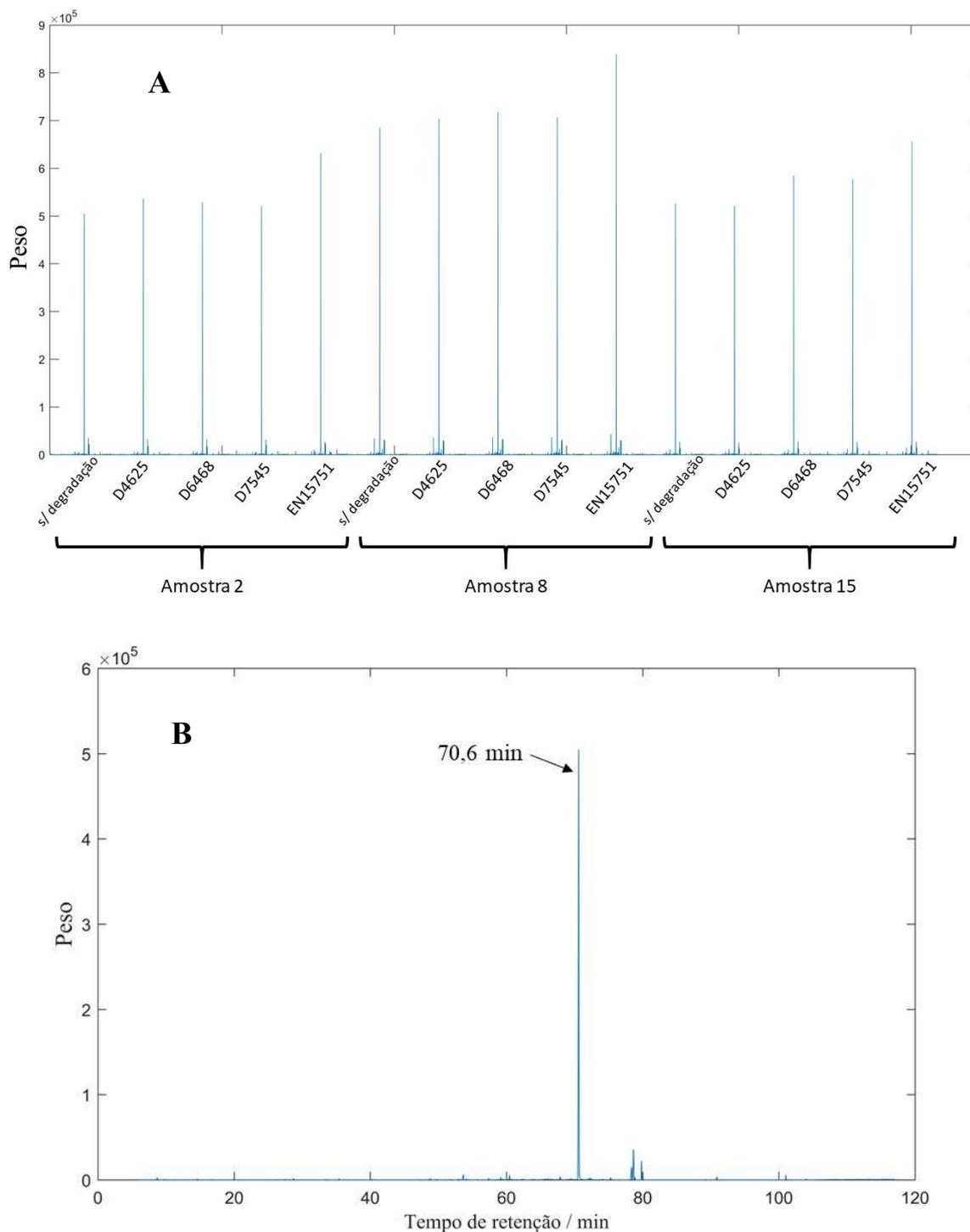


Figura 53: Pesos da primeira componente da matriz **C** (perfil cromatográfico recuperado). **A**: todas as amostras. **B**: Apenas a amostra 2 sem degradação destacando o tempo de retenção que corresponde ao do palmitato de metila.

Analisando-se a mesma componente na matriz **S** (perfil espectral), obtém-se um espectro de massas recuperado/estimado (Figura 54) com as características típicas de ésteres alifáticos<sup>62</sup>: pico base em  $m/z$  74 característico do rearranjo de McLafferty; pico

em  $m/z$  43 de íon alquila; pico em  $m/z$  87 característico de íon alquila contendo oxigênio; pico em  $m/z$  239 que corresponde ao íon M-31 ( $\text{OCH}_3$ ); e pico do íon molecular M em  $m/z$  270 e M+1  $m/z$  271.

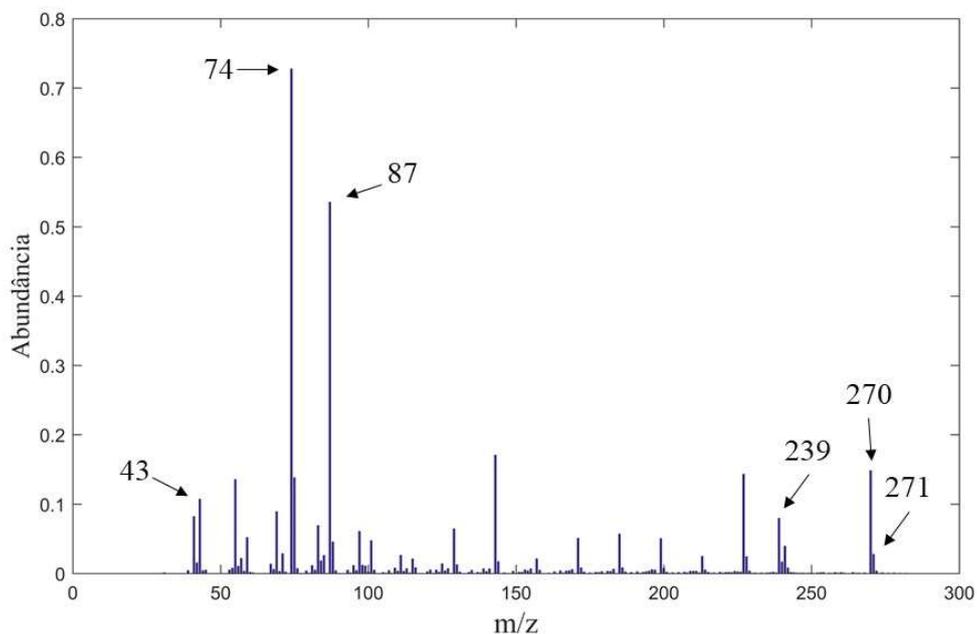


Figura 54: Perfil espectral recuperado da primeira componente do modelo MCR destacando os picos típicos de ésteres alifáticos.

O banco de dados Massbank<sup>®155</sup> permitiu comparar o espectro de massas recuperado com o banco de dados do programa. Neste caso, o espectro de referência mais similar ao recuperado por MCR foi, como esperado, o do palmitato de metila, com 91,7 % de similaridade. Visualmente, também é possível perceber a similaridade com o espectro de referência da Figura 55.

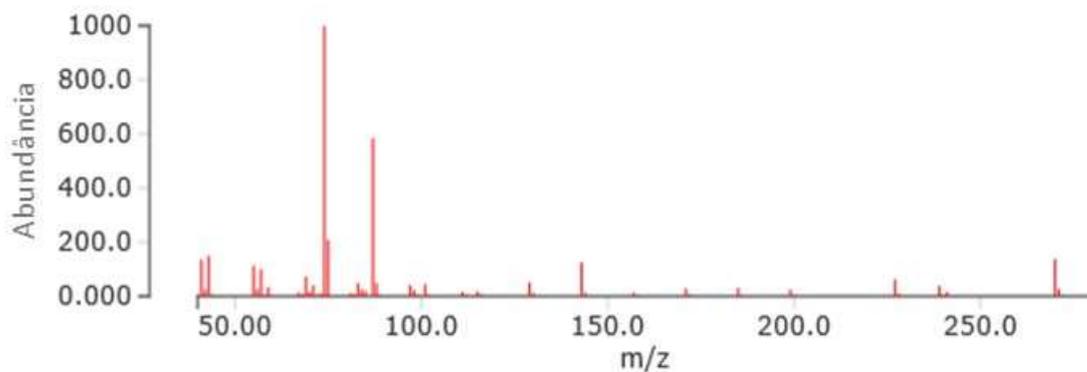


Figura 55: Espectro do palmitato de metila obtido no banco de dados Massbank<sup>®155</sup>.

Observando novamente a Figura 53-A, pode-se notar diferenças significativas entre as três amostras analisadas. Comparando de forma geral as três amostras nessa componente, percebe-se que as amostras 2 e 15 são mais similares entre si e apresentam intensidades sistematicamente menores que a amostra 8. Como destacado anteriormente, essas diferenças são fruto das origens do óleo diesel e do biodiesel que compõem as amostras. Para permitir uma melhor comparação dos resultados dos ensaios de estabilidade, a soma das intensidades dos cromatogramas estimados para as cinco corridas de uma mesma amostra (antes e após os quatro ensaios de degradação) foram normalizadas para 100%. Essa normalização visa minimizar a diferença entre as amostras e possíveis variações na silanização e na diluição. Assim, por exemplo, o somatório das áreas da amostra 2 para as cinco condições (amostra original e após ensaios) é igual 100% e cada teste de estabilidade fornece um percentual dessa área. Adicionalmente, decidiu-se desprezar picos recuperados com valores muito pequenos, pois observou-se que estes representavam basicamente falhas na modelagem dos dados ou ruído e poderiam impactar negativamente no cálculo de área total por amostra. Decidiu-se excluir todos os picos recuperados com pesos/alturas inferiores a 1% do valor de intensidade do pico máximo da amostra não degradada, pois esses picos são considerados ruído. Na prática, esses picos foram substituídos pelo valor zero. A Figura 56 mostra o resultado dessa abordagem. Vale mencionar que, para critérios de comparação, após essa normalização, as três amostras foram consideradas como replicatas.

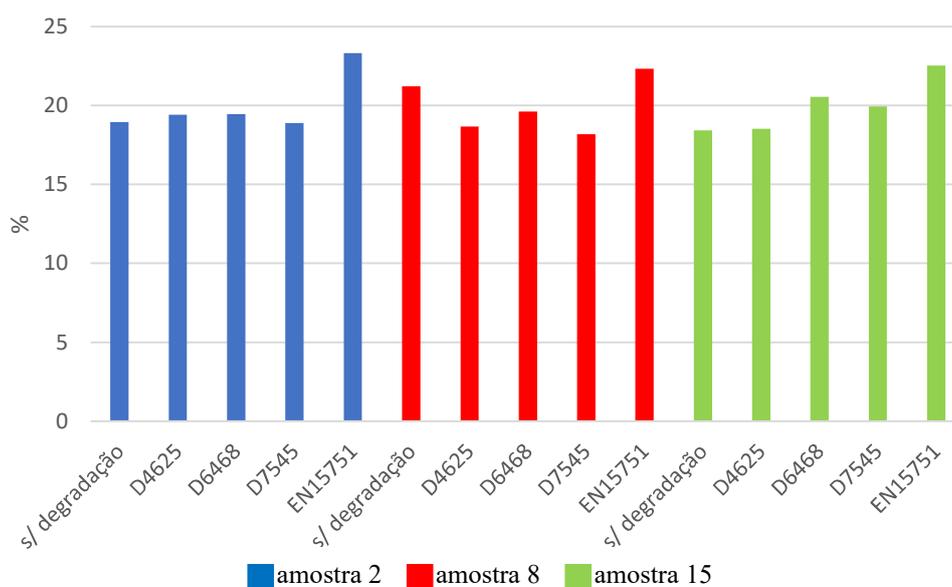


Figura 56: Somatório das áreas do perfil de concentrações da 1ª. componente.

A seguir, é realizada a análise de variância, ANOVA, dos dados normalizados, Tabela 9. Como o valor  $F$  calculado (9,70) é superior ao valor  $F$  crítico (3,48), concluiu-se que há diferença significativa (com nível de significância de 5%) entre os ensaios e as amostras sem degradação. O teste Tukey (com o mesmo nível de significância) indicou que apenas os ensaios conforme norma EN15751 forneceram áreas totais significativamente maiores (ou, maiores que a *diferença mínima significativa*) que as das amostras sem degradação.

Tabela 9: ANOVA de um fator das áreas recuperadas para a primeira componente na matriz C.

RESUMO						
<i>Grupo</i>	<i>Contagem</i>	<i>Soma</i>	<i>Média</i>	<i>Variância</i>		
s/ degradação	3	58,58	19,53	2,20		
D4625	3	56,61	18,87	0,23		
D6468	3	59,62	19,87	0,36		
D7545	3	57,01	19,00	0,79		
EN15751	3	68,18	22,73	0,27		

ANOVA						
<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>
Entre grupos	29,81	4	7,45	<b>9,70</b>	0,00	<b>3,48</b>
Dentro dos grupos	7,68	10	0,77			
Total	37,49	14				

Uma explicação para esse fenômeno seria que o palmitato de etila não é consumido/degradado nos ensaios, mas várias outras substâncias são, o que acaba por concentrar esse éster (lembrando que a quantidade de óleo diesel injetada no cromatógrafo foi sempre a mesma). A avaliação das próximas componentes mostrará, justamente, que o ensaio conforme EN15751 promove perda e degradação significativa de vários compostos, ao contrário dos demais ensaios.

A segunda componente apresentou vários picos na matriz C (Figura 57). Com base no padrão dos picos recuperados e de seus tempos de retenção, conclui-se que, para essa componente, o modelo não conseguiu isolar um composto, mas uma *classe* de compostos: os hidrocarbonetos do óleo diesel (e, como será exposto adiante, possivelmente, os hidrocarbonetos *alifáticos*). Interessante notar, especialmente na Figura 57-B, que MCR parece ter “limpado” o cromatograma dos picos do biodiesel e da linha

de base em forma de morro - bastante visível na Figura 46-A. Além disso, a Figura 45-B mostra que desprezar os sinais que apresentaram intensidades abaixo de 1% em relação à amostra não degradada reduziu consideravelmente o ruído, mas não suprimiu os picos relevantes para a análise.

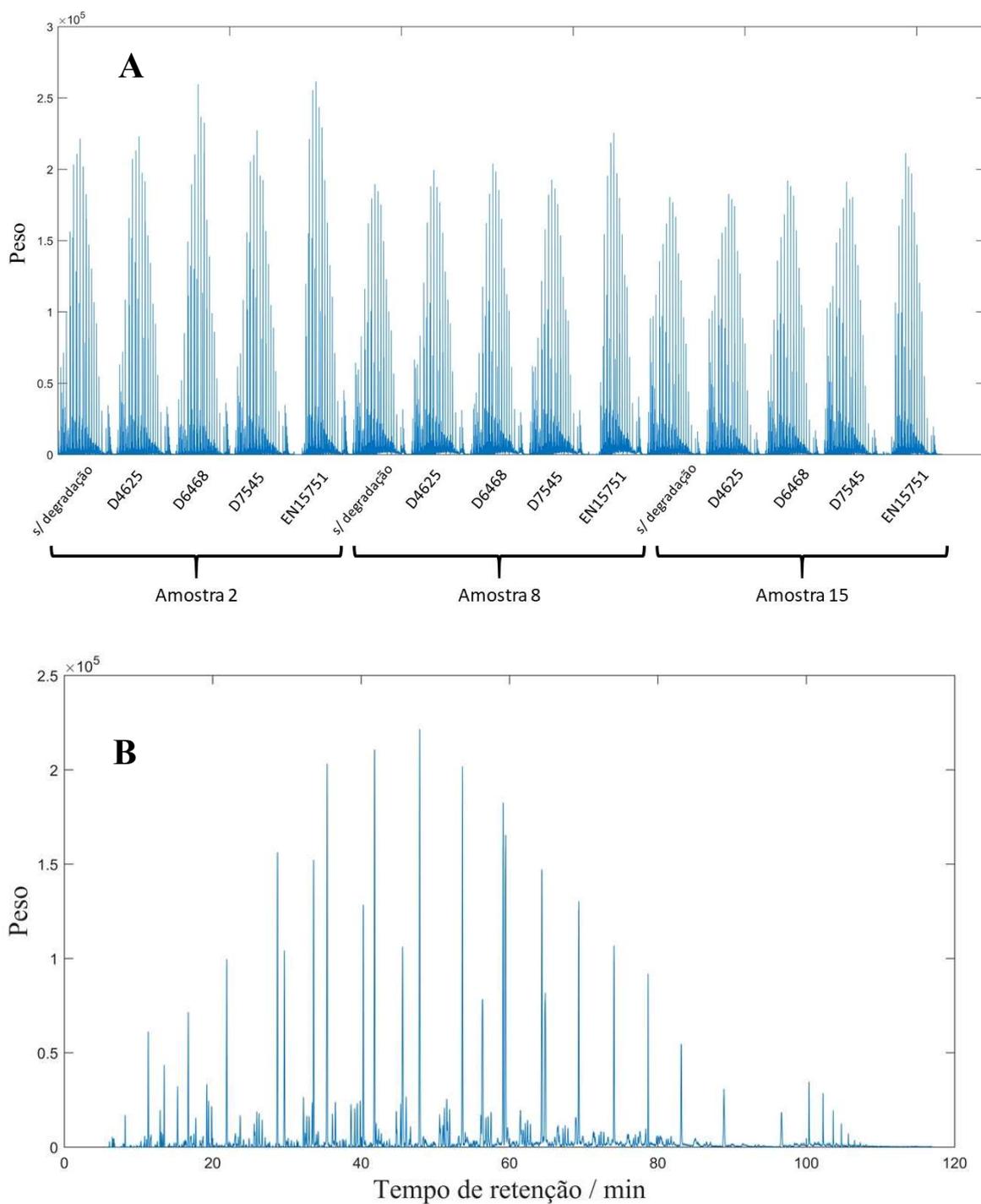


Figura 57: Pesos da segunda componente da matriz C (perfil de concentrações). **A:** todas as amostras. **B:** Apenas amostra 2 sem degradação.

Analisando-se a mesma componente na matriz **S** (perfil espectral), obtém-se um espectro de massas recuperado (Figura 58) com características de hidrocarbonetos saturados<sup>62</sup>: picos pronunciados com massas iguais a  $C_nH_{2n+1}$ . A comparação do espectro recuperado com o banco de dados Massbank<sup>®</sup> indicou 20 hidrocarbonetos *alifáticos* com índices de similaridades entre 70 e 80 %: o primeiro seria o 3-etil-2,6,10-trimetilundecano com 79,9 % de similaridade. Essa “dificuldade” do algoritmo em indicar um composto com alta similaridade não surpreende, já que a componente não está modelando uma única substância como se espera usualmente com MCR. Por outro lado, todos os compostos indicados pelo algoritmo são hidrocarbonetos, o que é coerente com a análise inicial de que a componente forneceu um cromatograma do óleo diesel livre de biodiesel.

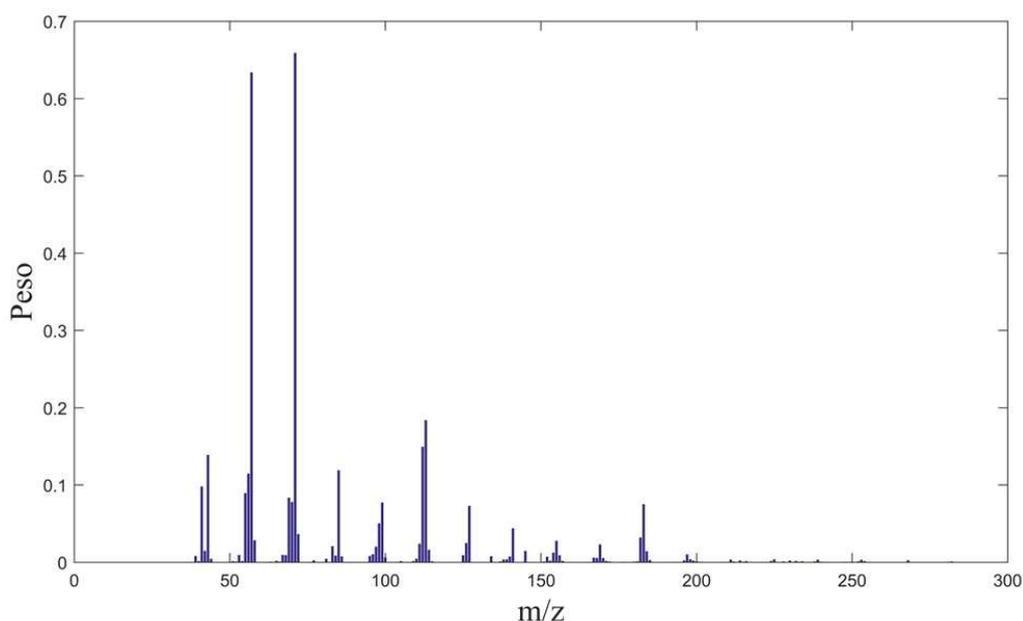


Figura 58: Perfil espectral recuperado da segunda componente do modelo MCR.

Novamente, os cromatogramas recuperados foram normalizados, os picos de intensidade inferior a 1% em relação à amostra não degradada foram excluídos e a área total foi calculada para facilitar a comparação dos resultados. A Figura 59 mostra o resultado dessa abordagem. Apesar de alguns resultados se destacarem, como há grande variação entre amostras, a ANOVA indicou que não há diferenças significativas (nível de significância de 5%) entre os ensaios. O valor de  $F$  calculado (1,2) ficou abaixo do valor  $F$  crítico (3,5).

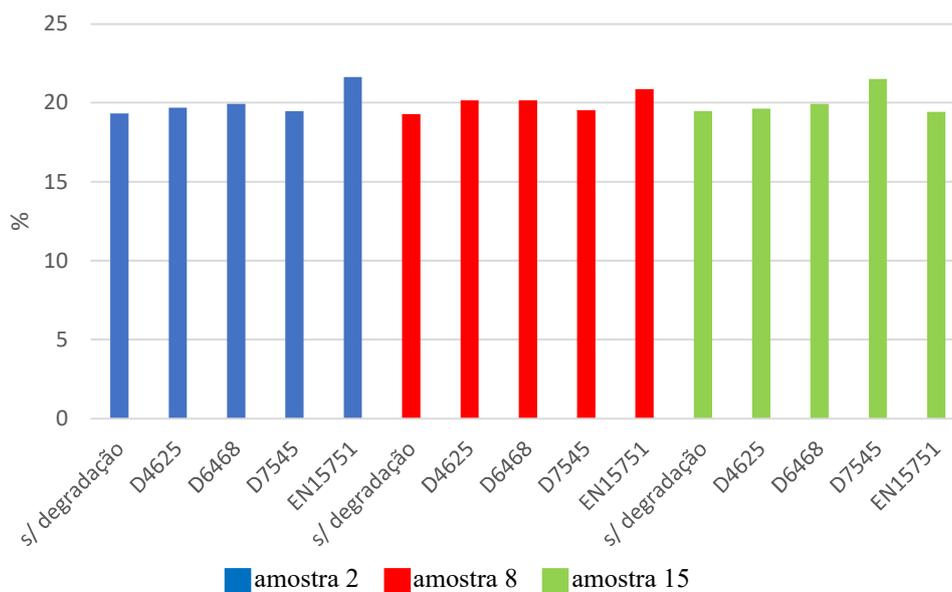


Figura 59: Somatório das áreas do perfil de concentrações da 2ª. componente.

Apesar do resultado da ANOVA, sobrepondo-se uma das amostras (de fato, qualquer uma das três amostras) antes e após os quatro ensaios, percebe-se claramente a diferença na parte inicial do perfil de concentração recuperado para a amostra submetida ao ensaio de acordo com a norma EN15751. Contudo, como as diferenças ocorrem em picos relativamente pequenos em relação ao cromatograma como um todo, representam apenas uma pequena fração da área total do cromatograma. O que justifica a não detecção dessa mudança pela ANOVA. Esse comportamento já havia sido observado na comparação dos TICs (ver Figura 50-A) sem o uso de MCR. Uma das claras vantagens do MCR é que, supostamente, ele permite comparar todo o perfil de hidrocarbonetos sem “interferência” do biodiesel, o que indica que o único ensaio que causa alterações significativas nessa classe de compostos é aquele de acordo com a norma EN15751. As amostras submetidas ao ensaio conforme D6468 também apresentaram diminuição de picos recuperados em baixos tempos de retenção, mas de forma muito menos pronunciada. Reforça-se aqui a premissa de que a avaliação é feita com base nos compostos passíveis de detecção nas condições cromatográficas utilizadas.

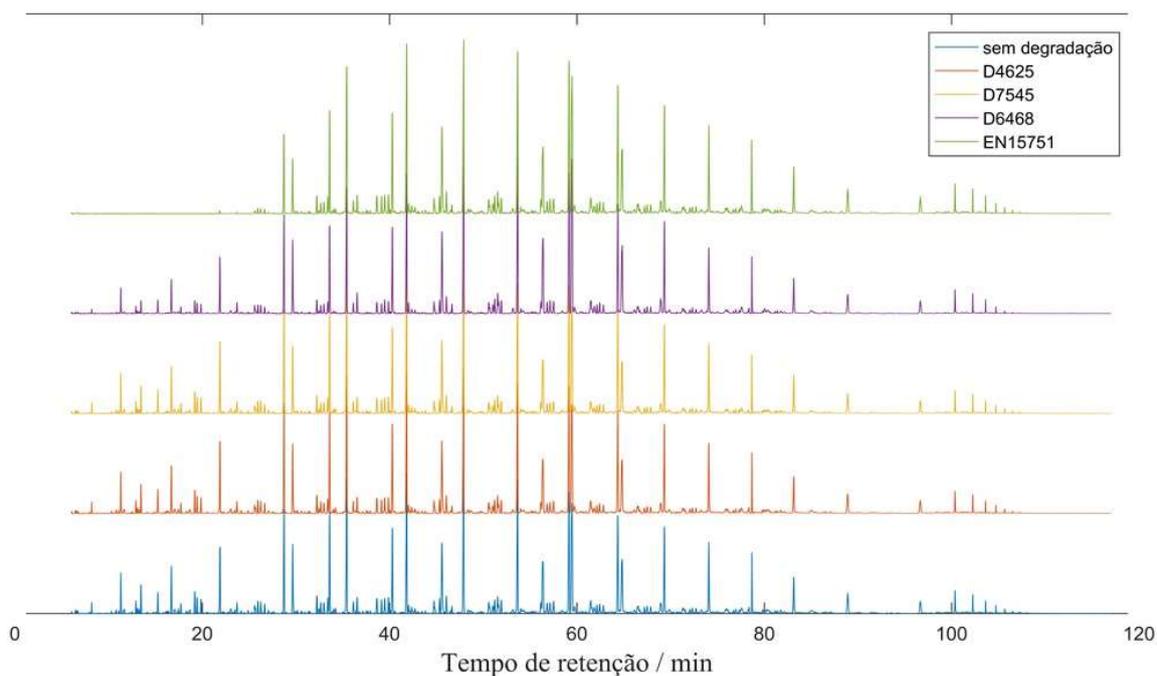


Figura 60: Perfis de cromatográficos recuperados na segunda componente da amostra 2 sem degradação e após os quatro ensaios de estabilidade. Os perfis foram deslocados verticalmente para facilitar a comparação.

A terceira componente apresentou o mesmo comportamento para todas as amostras e testes de estabilidade. Por essa razão, apenas os pesos da amostra 2 sem degradação é exibida na Figura 61. O perfil espectral da terceira componente (não mostrado) indicou picos intensos apenas em  $m/z$  207 e 281. Apesar de não ter sido possível identificar um composto com esse perfil espectral, atribui-se essa componente ao fenômeno de “sangramento” da coluna cromatográfica porque o perfil de concentração acompanha a rampa de aquecimento da coluna e porque o sinal em 207  $m/z$  é conhecido como indicador de sangramento da coluna cromatográfica<sup>67</sup>. Claramente, essa componente capturou o sinal da linha de base, o que “limpou” as demais componentes.

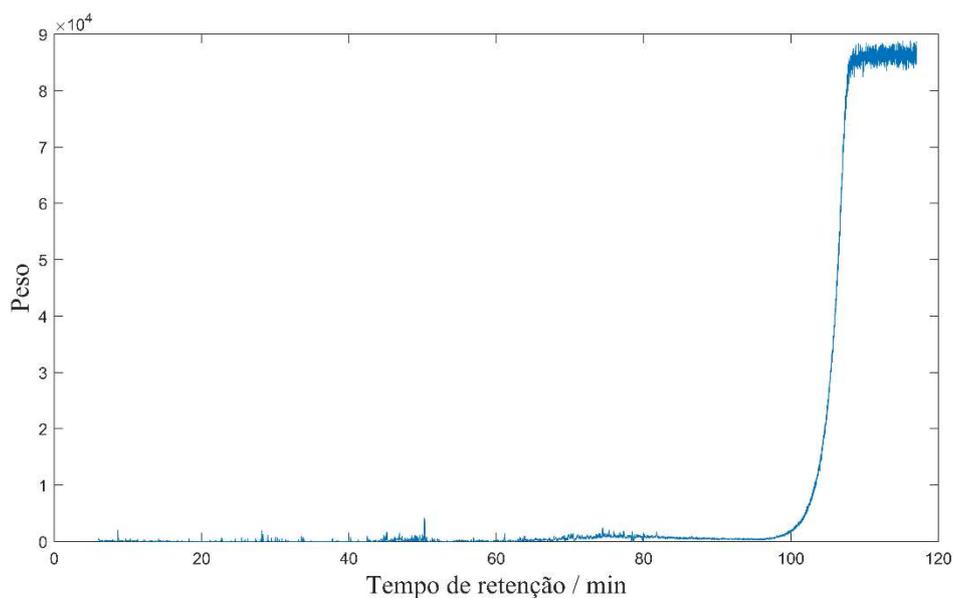


Figura 61: Pesos da terceira componente da matriz *C* (perfil de concentrações) da amostra 2 sem degradação.

A quarta componente da matriz de concentrações mostrada na Figura 62 capturou praticamente uma única espécie que, ao contrário do observado na primeira componente, apresenta comportamento nitidamente diferenciado entre os ensaios de estabilidade. O tempo de retenção desses picos recuperados é de 78,4 minutos, o que corresponde ao linoleato de metila, de acordo com os dados de GC-MS.

O espectro recuperado na quarta componente é mostrado na Figura 63 e, assim como o perfil da primeira componente, apresenta os picos típicos de ésteres, mas, nesse caso, com os padrões de fragmentação de cadeias insaturadas. A comparação com o banco de dados Massbank<sup>®</sup> indicou como composto mais provável o linoleato de metila com 91,2 % de similaridade, confirmando a interpretação feita com base no tempo de retenção.

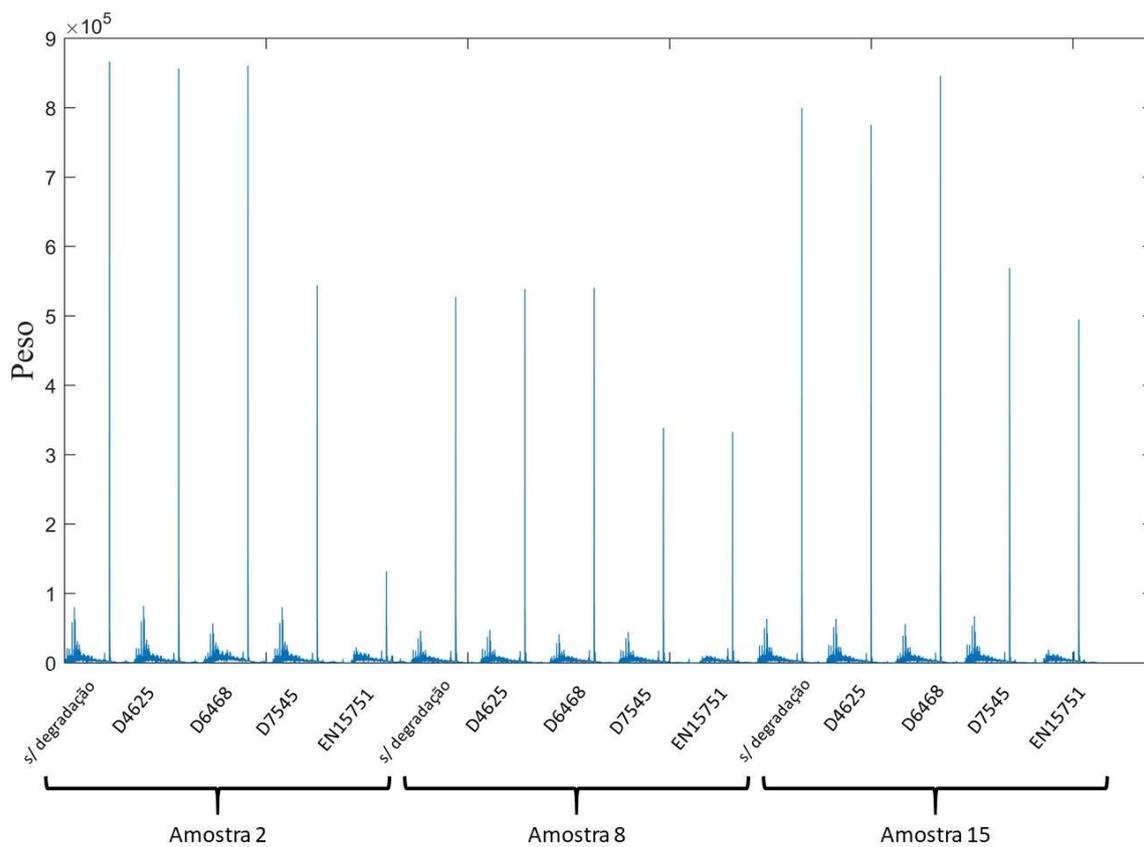


Figura 62: Pesos da quarta componente da matriz  $C$  (perfil de concentrações).

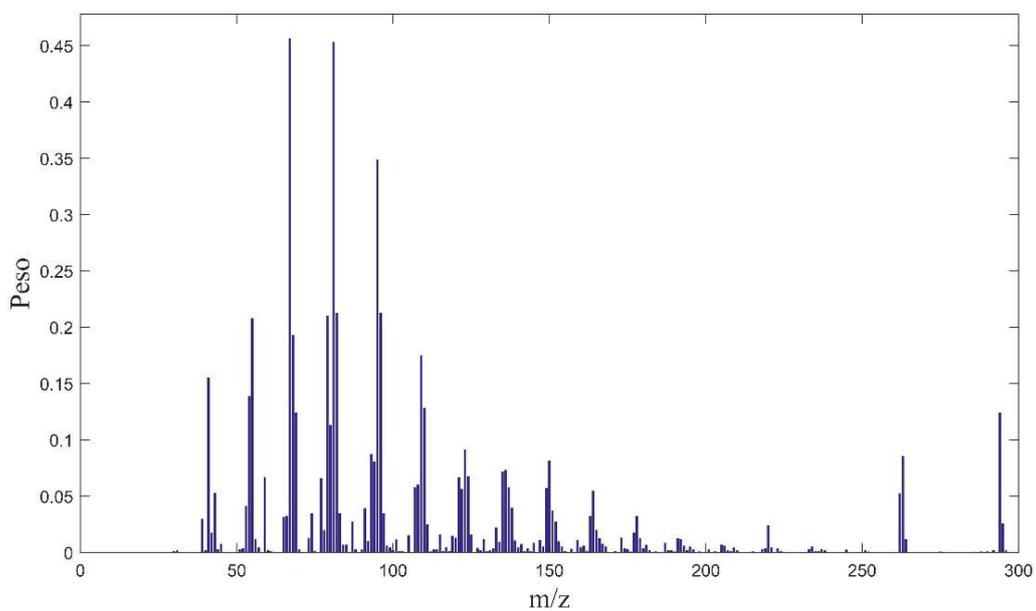


Figura 63: Perfil espectral recuperado da quarta componente do modelo MCR.

A Figura 62 mostrou grande variação entre amostras, o que torna a abordagem de normalização de áreas por amostra ainda mais importante. A Figura 64 deixa mais visível que as amostras que sofreram os ensaios conforme normas D7545 e EN15751 são as únicas que forneceram áreas consistentemente menores. Isso foi confirmado com ANOVA que forneceu um valor  $F_{\text{calculado}}$  de 38,2 para um  $F_{\text{crítico}}$  de 3,5 e com o teste Tukey que apontou valores de diferença significativa para os ensaios conforme normas D7545 e EN15751. Essa componente, portanto, modela basicamente o comportamento do linoleato de metila e mostra que apenas os ensaios com características mais oxidantes causam redução significativa em sua concentração.

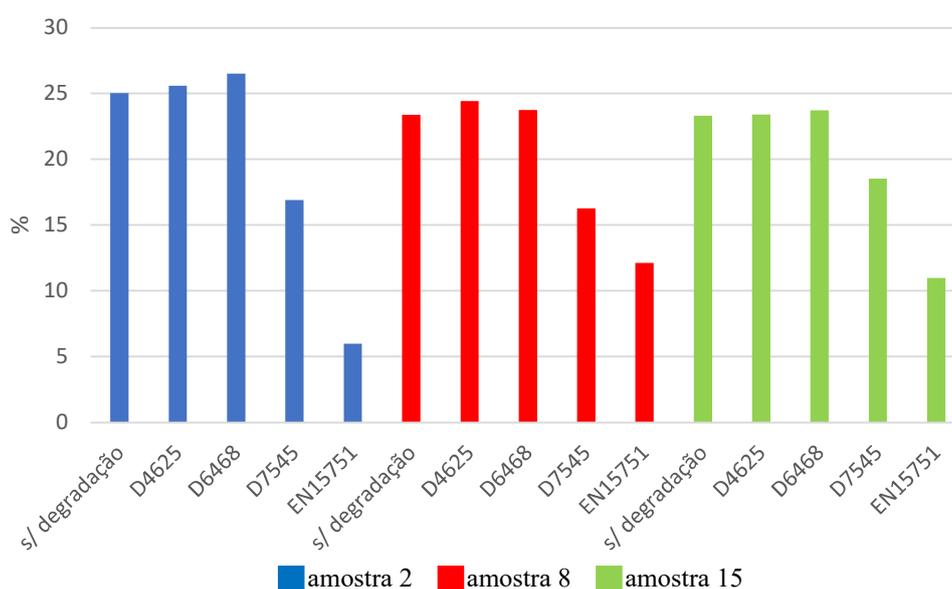


Figura 64: Somatório das áreas do perfil de concentrações da 4ª. componente.

A quinta componente forneceu uma estimativa do perfil cromatográfico (perfil de intensidade relativas) com vários picos até, aproximadamente, 20 minutos de corrida. Conforme Figura 65, o comportamento observado foi nitidamente diferenciado para as amostras submetidas aos ensaios conforme normas D6468 e, principalmente, EN15751, nos quais o pico majoritário é consideravelmente menor após o ensaio D6468 e praticamente suprimido pelo ensaio EN15751.

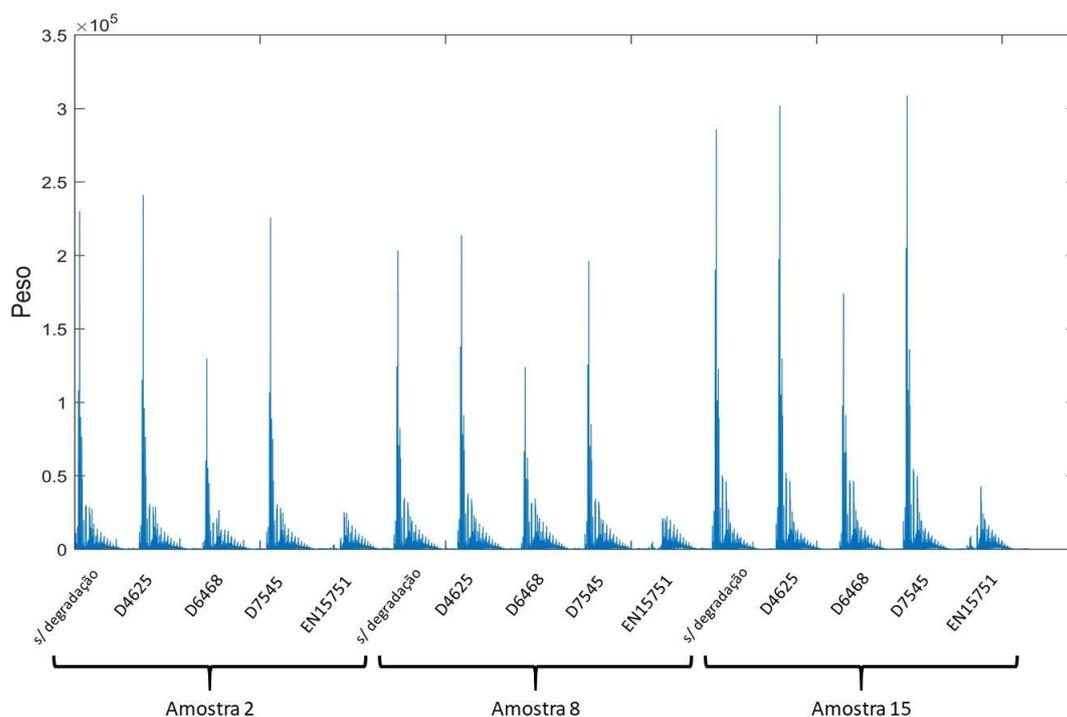


Figura 65: Perfil de concentrações recuperado na 5<sup>a</sup>. componente do modelo MCR.

O perfil da componente espectral, mostrado na Figura 66, é dominado por um sinal em  $m/z$  105, que é característico de compostos que contêm um anel aromático. A busca na base de dados espectrais Massbank<sup>®</sup> indicou vários compostos monoaromáticos com alta similaridade ao espectro recuperado. O espectro de referência é mostrado no Apêndice 4.

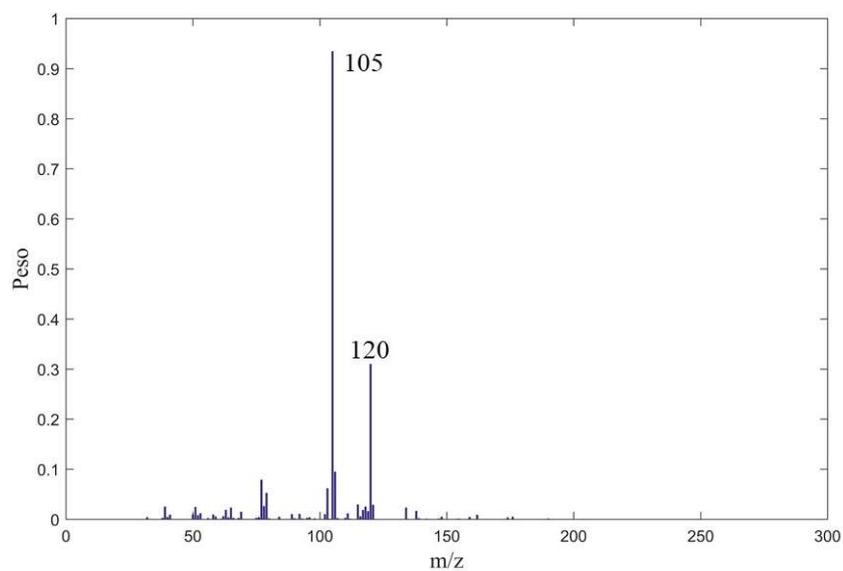


Figura 66: Espectro recuperado pela 5<sup>a</sup>. componente do modelo MCR.

As áreas normalizadas (como descrito anteriormente) para cada amostra são mostradas na Figura 67. A análise de variância apontou (com um nível de significância de 5%) que há diferença entre os ensaios ( $F_{\text{calculado}}: 101,1; F_{\text{crítico}}: 3,5$ ) e o teste Tukey (com o mesmo nível de significância) confirmou que os ensaios conforme D6468 e EN15751 apresentam áreas menores que as das amostras originais. Esses ensaios são justamente os que mais tendem a promover perda de material volátil, como discutido anteriormente. Conclui-se que a quinta componente foi capaz de modelar o comportamento de uma classe específica de hidrocarbonetos (monoaromáticos), revelando um padrão que não seria perceptível sem o uso de MCR-ALS. Esses compostos são particularmente importantes na avaliação da estabilidade da mistura óleo diesel-biodiesel por contribuírem para a dissolução de produtos de oxidação do biocombustível<sup>36</sup>.

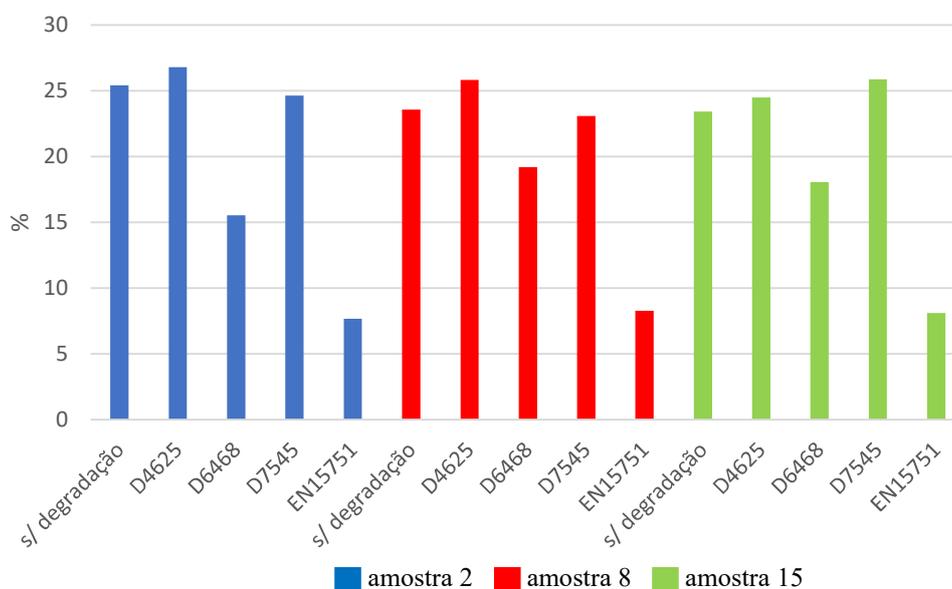


Figura 67: Somatório das áreas do perfil de concentrações da 5ª. componente.

A sexta componente capturou um pico praticamente unitário em 78,7 minutos conforme Figura 68. Por comparação com os tempos de retenção observados na análise qualitativa dos dados de GC-MS, esse pico corresponde ao oleato de metila.

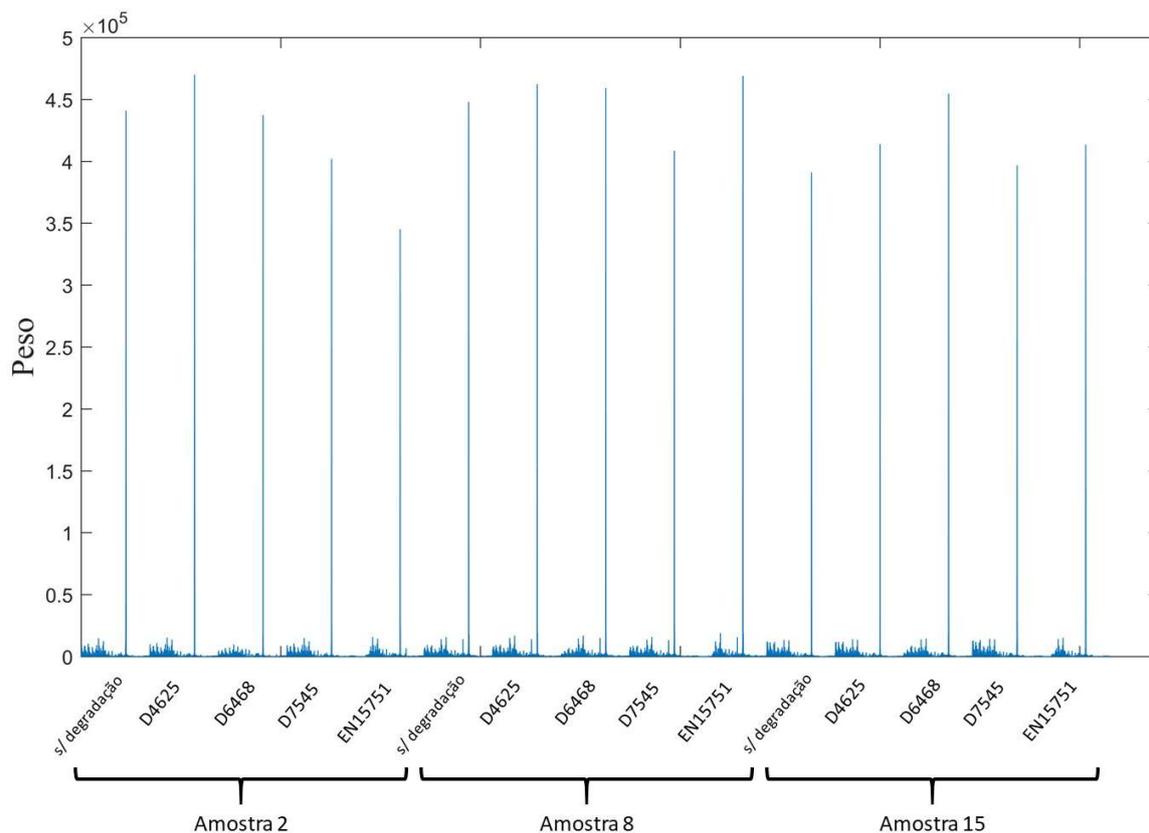


Figura 68: Pesos da sexta componente da matriz  $C$  (perfil de concentrações).

A mesma componente na matriz  $S$  fornece um espectro recuperado típico de ésteres (ver Figura 69): picos intensos em  $m/z$  74 e  $m/z$  87 (característicos de rearranjos de McLafferty), em  $m/z$  265 de  $M - 31$  ( $OCH_3$ ) e o pico do íon molecular  $M$  em  $m/z$  296 e  $M+1$   $m/z$  297 (ésteres insaturados tendem a apresentar  $M$  mais intensos que os saturados). A comparação com o banco de dados Massbank<sup>®</sup> fornece, como primeira, segunda e terceira opções os compostos *cis*-12-octadecanoato de metila, (*z*)-15-octadecanoato de metila e oleato de metila, apresentando 88,3 %, 86,8 % e 84,5 % de similaridade, respectivamente. Os três compostos apresentam similaridades estruturais, mas a presença dos dois primeiros no biodiesel não é esperada de acordo com a literatura<sup>6</sup>. Portanto, apesar de um índice de similaridade menor em relação às componentes discutidas anteriormente, essa componente foi identificada como oleato de metila (o espectro de referência do banco de dados é mostrado no Apêndice 4). Outro aspecto que deve ser considerado em relação ao índice de similaridade é que não se pode excluir, também, a possibilidade de que as condições de obtenção do espectro de referência sejam ligeiramente diferentes das utilizadas neste trabalho, o que poderia justificar um pequeno percentual de discordância nos espectros de massa.

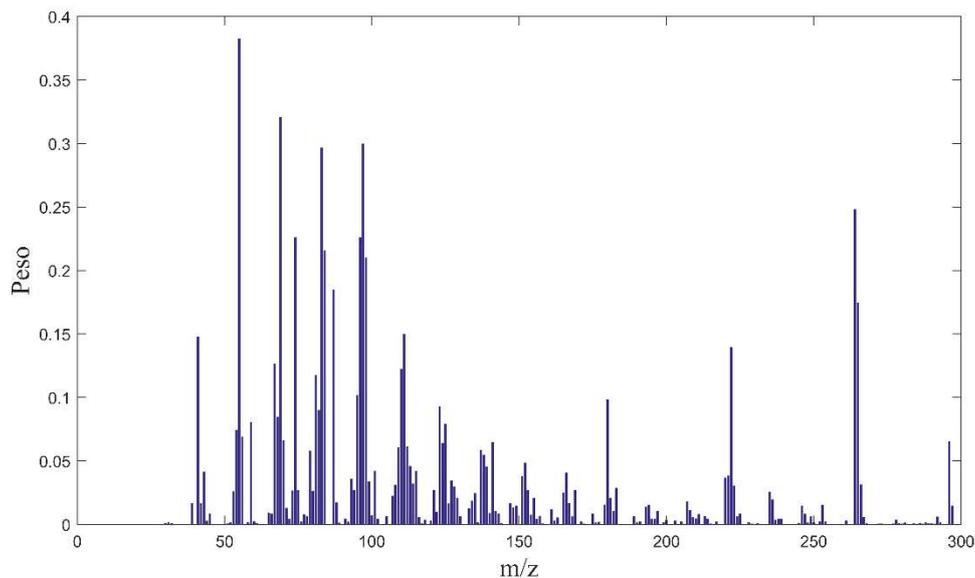


Figura 69: Perfil espectral recuperado da sexta componente do modelo MCR.

As áreas normalizadas (como descrito anteriormente) do perfil de concentrações para cada amostra são mostradas na Figura 70. A análise de variância apontou (com um nível de significância de 5%) que não existe diferença significativa entre os ensaios ( $F_{\text{calculado}}: 1,5; F_{\text{crítico}}: 3,5$ ), o que demonstra que o oleato de metila não foi afetado significativamente, mesmo nos ensaios mais severos. De forma estrita, a ANOVA permite concluir que a variação entre as amostras e a variação entre os ensaios, que nessa componente é devido ao comportamento do oleato de metila, são equivalentes, com 95% de confiança.

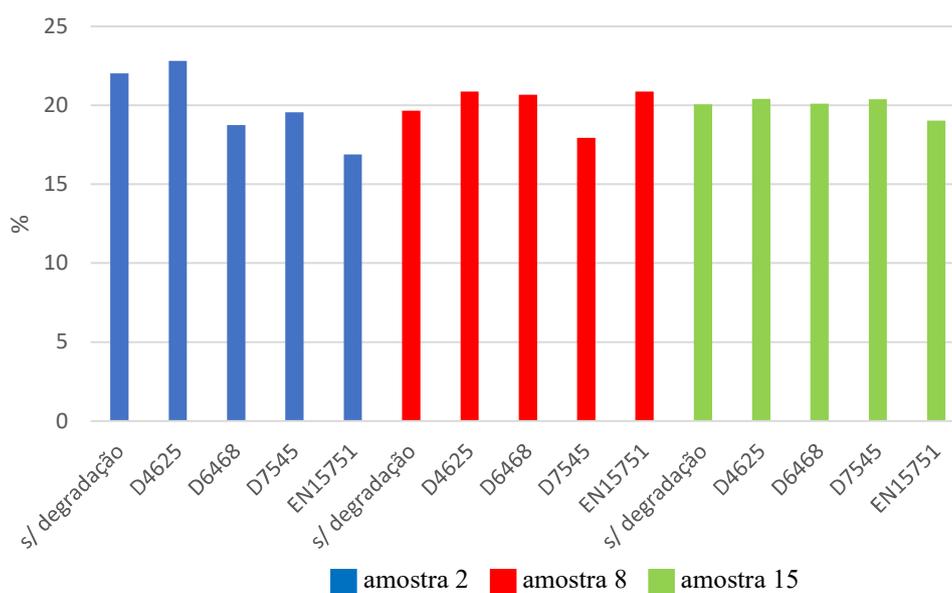


Figura 70: Somatório das áreas do perfil de concentrações da 6ª. componente.

A sétima componente forneceu perfil de concentrações recuperado com vários picos sem um padrão claro. O perfil espectral recuperado não apresenta similaridade razoável com nenhum composto do banco de dados Massbank<sup>®</sup>. A interpretação dessa componente, portanto, não pôde ser feita de forma objetiva e será apresentada de forma resumida na Tabela 10, juntamente com outras componentes de interpretação subjetiva.

A componente 10 apresentou picos praticamente unitários na matriz C conforme Figura 71. O tempo de retenção desses picos recuperados foi de 79,8 minutos, o que corresponde ao tempo de retenção do estearato de metila conforme dados de GC-MS.

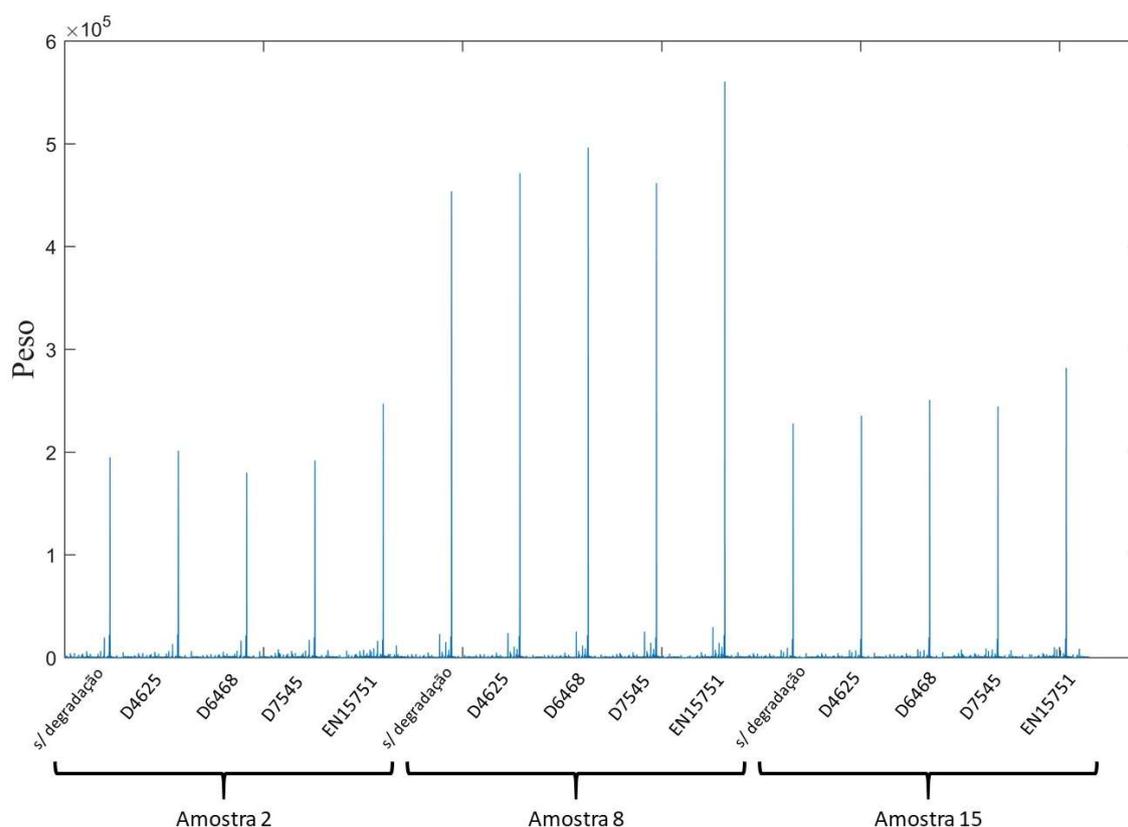


Figura 71: Pesos da décima componente da matriz C (perfil de concentrações).

A avaliação da décima componente na matriz de espectros recuperados (S) (Figura 72) confirma essa interpretação já que os pesos da 10<sup>a</sup> componente são aqueles típicos de ésteres saturados: pico base em m/z 74 característico do rearranjo de McLafferty; pico em m/z 43 de íon alquila; pico em m/z 87 característico de íon alquila contendo oxigênio; pico em m/z 267 que corresponde ao íon M-31 (OCH<sub>3</sub>); e pico do íon molecular M em m/z 298 e M+1 m/z 299.

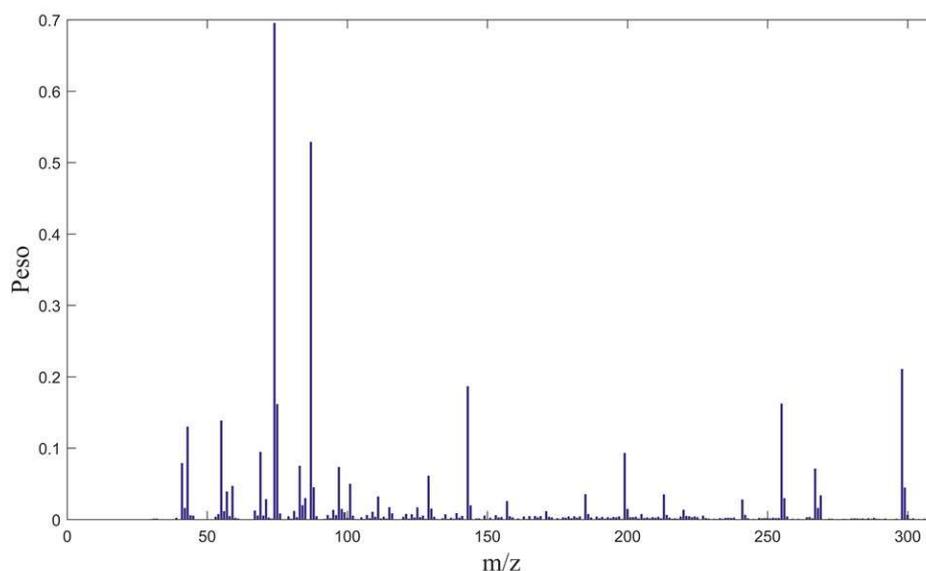


Figura 72: Perfil espectral recuperado da décima componente do modelo MCR.

A comparação desse espectro (recuperado) com o banco de dados Massbank<sup>®</sup> indicou, como esperado, o estearato de metila como primeiro composto com 86,5 % de similaridade (o espectro de referência é mostrado no Apêndice 4).

As áreas normalizadas (como descrito anteriormente) do perfil de concentrações para cada amostra são mostradas na Figura 73. A análise de variância apontou (com um nível de significância de 5%) que há diferença entre os ensaios ( $F_{\text{calculado}}$ : 16,3;  $F_{\text{crítico}}$ : 3,48) e o teste Tukey (com o mesmo nível de significância) apontou que apenas as amostras submetidas ao ensaio conforme EN15751 apresentam áreas significativamente maiores que as das amostras originais.

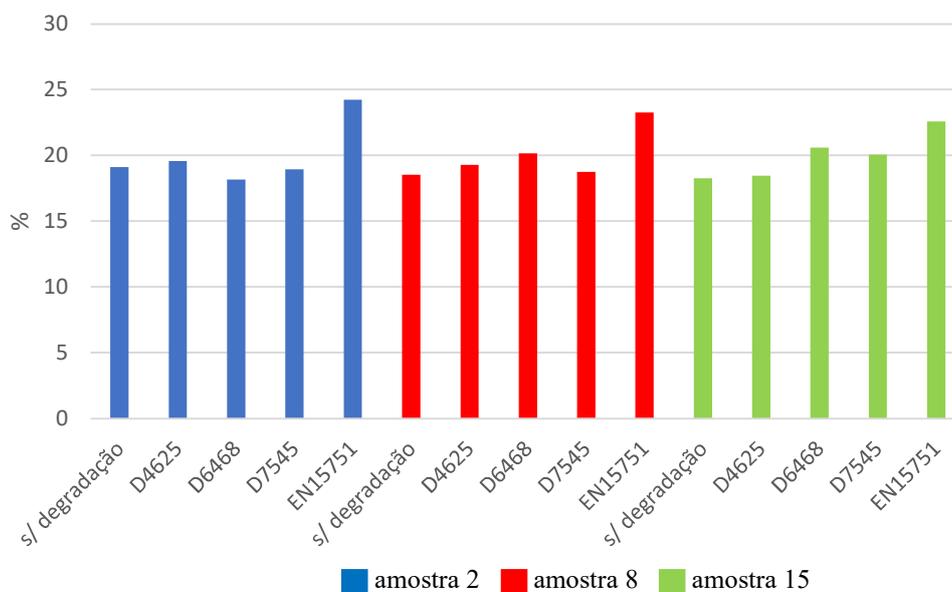


Figura 73: Somatório das áreas do perfil de concentrações da 10ª. componente.

Mais uma vez, assim como na primeira componente, interpreta-se que o aumento da concentração desse éster é causado pela diminuição significativa de outros compostos (e.g.: linoleato de metila e hidrocarbonetos leves), o que acaba por concentrar o estearato de metila.

Conforme descrito anteriormente, muitas componentes modelaram a variabilidade dos dados de formas que não puderam ser interpretadas ou identificadas de forma tão direta quanto aquelas discutidas até aqui. A Tabela 10 resume a avaliação dessas componentes.

Tabela 10: Avaliação das componentes cuja interpretação é subjetiva.

Componente	Avaliação das componentes*
7	Apresentou centenas de picos no perfil cromatográfico recuperado na matriz C. Não foi possível atribuir nenhum composto ou classe de compostos com similaridade razoável ao espectro de massas recuperado.
8	Foram observados vários picos em tempos de retenção entre 20 e 80 min. no perfil cromatográfico recuperado. Não foi possível atribuir nenhum composto ou classe de compostos com similaridade razoável ao espectro de massas recuperado.
9	Vários picos em tempos de retenção distribuídos ao longo de todo o cromatograma, mas, especialmente, no início do cromatograma recuperado. Os picos foram similares em todas as condições, exceto para as amostras do ensaio de acordo com EN15751, que apresentou picos significativamente menores. O espectro recuperado apresenta similaridade acima de 80 % para vários compostos aromáticos. Esta componente parece ser um caso de degeneração da componente em relação à 5ª. componente.
11	Esta componente apresentou alguns picos no início do cromatograma recuperado, principalmente, em, aprox., 35 min. Não foi possível atribuir nenhum composto ou classe de compostos com similaridade razoável ao espectro de massas recuperado.
12	Foram observados dezenas de picos no perfil cromatográfico recuperado, principalmente, em tempos de retenção baixos, mas se estendendo até ca. 80 min. Os picos são significativamente menores para as amostras submetidas aos ensaios conforme D6468 e EN15751. O perfil espectral recuperado tem 95 % ou mais de semelhança com todos os xilenos, mas, como o perfil de concentrações tem muitos picos, não foi possível associar a 12ª. componente a essa classe de compostos.
13	Observou-se centenas de picos ao redor de cerca de 40 min. Não foi possível atribuir nenhum composto ou classe de compostos com similaridade razoável com base no espectro de massas de recuperado.
14	Observou-se centenas de picos com perfil similar ao da segunda componente, ou seja, picos que parecem agregar os hidrocarbonetos do óleo diesel. O espectro de massas recuperado apresentou 80 % de similaridade com o pentadecano. O algoritmo indica similaridade com vários outros hidrocarbonetos alifáticos/saturados. Essa componente parece ser degenerada em relação à segunda componente.

...Continuação da Tabela 10.

15	Observou-se a presença de centenas de picos que parecem corresponder ao perfil de hidrocarbonetos até ca. 80 min. Não foi possível atribuir nenhum composto com similaridade razoável ao espectro recuperado.
16	Foram observados centenas de picos que parecem corresponder ao perfil de hidrocarbonetos a partir de cerca de 35 min. O perfil espectral recuperado apresenta centenas de picos sem padrões claros. Não foi possível atribuir nenhum composto com similaridade razoável.
17	Foram observados centenas de picos recuperados que parecem corresponder ao perfil de hidrocarbonetos de baixo tempo de retenção até ca. 45 min. Não foi possível atribuir nenhum composto ou classe com similaridade razoável ao espectro de massas recuperado.
18	Essa componente apresentou um perfil de concentrações recuperado que lembra o perfil da linha-base (3ª componente), mas com alguns picos sobressaindo nos tempos de retenção baixos. Não foi possível atribuir nenhum composto ou classe de compostos com similaridade razoável ao espectro recuperado.
19	O perfil cromatográfico recuperado para essa componente contém picos ao longo de todos os tempos de retenção com perfil que lembra o TIC de uma amostra. O espectro de massas recuperado apresentou poucos picos e não pôde ser associado a nenhum composto com similaridade alta.

\* As comparações de espectros foram feitas com base de dados Massbank<sup>®155</sup> e literatura<sup>62,67,69</sup>.

Os principais resultados descritos nesta seção são resumidos a seguir:

- As componentes de números 1, 4, 6 e 10 modelaram as componentes do biodiesel (palmitato, linoleato, oleato e estearato de metila, respectivamente).
- O linoleato de metila foi consumido em quantidade significativa apenas nos ensaios conforme normas D7545 e EN15751. Isso demonstra a suscetibilidade desse composto às condições oxidantes dos ensaios.
- O oleato de metila não apresentou variação significativa em relação à amostra original para nenhum dos ensaios, apesar de apresentar uma insaturação na cadeia alquílica.
- A 2ª componente modelou os principais hidrocarbonetos do óleo diesel fornecendo um cromatograma (recuperado) livre de linha base e da presença do biodiesel.

- A linha de base foi modelada na terceira componente, o que contribuiu para que as demais componentes fossem obtidas livres da linha de base em forma de sino e da sua elevação final causada pelo aumento de temperatura da programação cromatográfica.
- Várias componentes forneceram resultados que não permitiram uma interpretação clara. Esse fato, em parte, é justificado pela grande complexidade dos dados de GC-MS de amostras de combustível constituído pela mistura de óleo diesel e biodiesel. A 9ª. componente, entretanto, parece capturar a “classe” de compostos “aromáticos”, que apresentaram comportamento homogêneo para todas as condições exceto para aquelas submetidas ao ensaio conforme EN15751 (que apresentaram diminuição de sinal). A reação de aromáticos com hidrocarbonetos insaturados, compostos de enxofre e nitrogênio já foi relatada na literatura como um mecanismo de formação de goma e sedimentos<sup>2,5</sup> em combustíveis. Interessante notar que essa é a primeira componente na qual o biodiesel não aparece como composto quimicamente ativo em um dos ensaios de estabilidade (especificamente, o EN15751).

Como considerações finais desta seção, é importante observar que o uso de três amostras por ensaio, apesar de estar aquém do planejado inicialmente e daquilo que seria desejável para fornecer inferências estatísticas conclusivas, foi importante para demonstrar que há grandes variações na composição e no comportamento dos óleos diesel frente aos ensaios de estabilidade, o que requer atenção em qualquer comparação de resultados. Além disso, observou-se que o uso de MCR-ALS em dados de GC-MS confirmou e expandiu as observações obtidas a partir da espectrometria no infravermelho.

## **6. CONCLUSÕES DA SEÇÃO 1**

A abordagem escolhida para seleção e coleta de amostras de óleo diesel foi bem-sucedida. Apesar de não ser possível estender os resultados deste estudo para outros óleos diesel comercializados no Brasil, os ensaios de rotina, de estabilidade, infravermelho e GC-MS mostraram que as amostras incluem uma grande variabilidade em termos de composição e propriedades, o que garante que os resultados não sejam restritos a um tipo ou origem de combustível.

Os ensaios de estabilidade mostram grande dispersão de resultados, mesmo entre amostras coletadas nas mesmas localidades, o que confirma a complexidade do problema em estudo.

Este estudo confirmou a capacidade da espectrometria no infravermelho médio para avaliar a estabilidade do óleo diesel. É possível identificar a formação de compostos oxigenados e a degradação de compostos insaturados com relativa facilidade. Isso, em parte, decorre dos excelentes níveis de reprodutibilidade alcançados com os equipamentos FTIR modernos.

Após a otimização do modelo PARAFAC, foi possível demonstrar que o conjunto de dados estudados é adequadamente ajustado por um modelo trilinear. É importante frisar este aspecto, já que a aplicação de PARAFAC no tipo de dados deste projeto não é usual e não seria surpreendente se não fosse possível ajustar um modelo razoável. Entretanto, os dados de CORCONDIA, a Soma de Quadrados de Erros e, mais importante, a coerência das componentes obtidas em todas as dimensões, demonstraram a possibilidade de uso de PARAFAC neste conjunto de dados. A coerência citada se manifesta, por exemplo, na correlação entre as características do ensaio de estabilidade e as regiões espectrais apontadas como relevantes para modelagem.

A ferramenta PARAFAC permitiu identificar combinações de regiões espectrais que melhor manifestam as alterações promovidas pelos ensaios conforme normas D4625, D7545 e EN15751. Para o ensaio conforme D6468, não foi possível identificar alterações específicas no infravermelho possivelmente porque elas foram mascaradas por variações entre amostras. Esse fato reforça a importância de uma amostragem representativa (ou, a mais representativa possível). A variação entre amostras, por sinal, mostrou-se um fator de grande relevância, tanto que parece ter “consumido” três das sete componentes utilizadas para modelagem dos dados. Em relação a cada ensaio, infere-se que:

- O ensaio de acordo com a norma D4625 tem, como característica, a componente 7 que é a de menor importância. Isso sugere que esse ensaio promove as alterações menos drásticas. As alterações se resumem à região de estiramento CH, o que estaria relacionado a perdas de hidrocarbonetos por evaporação.
- O método D7545 está relacionado à 5ª componente que tem importância intermediária. As transformações estão relacionadas à região do estiramento

OH e C=O, sugerindo que esse ensaio promove formação de compostos ácidos e oxidação das carbonilas do biodiesel.

- Já o método conforme norma EN15751 é o que tem maior importância dentre os três (4ª componente), promovendo transformações na região de estiramento C=O, CH e OH, o que sugere, em conjunto com a avaliação das condições de ensaio, que o método promove, além de oxidações, perdas por ebulição/evaporação.

A cromatografia gasosa acoplada à espectrometria de massas permitiu identificar os principais compostos afetados nos ensaios: hidrocarbonetos leves e ésteres, especialmente os mais insaturados. A identificação de produtos de degradação é extremamente difícil por eles estarem presentes em pequenas concentrações e pela dificuldade em resolver (do ponto de vista cromatográfico) compostos numa matriz tão complexa.

A aplicação de MCR-ALS aos dados de GC-MS permitiu separar cada componente do biodiesel em diferentes componentes, o que pode ser útil na quantificação desses constituintes e na identificação da matéria-prima utilizada na produção de biodiesel. Mesmo aquelas componentes que modelam compostos que já são bem resolvidos cromatograficamente, são úteis, pois as componentes estão livres das influências das linhas de base, e por estarem isoladas, facilitam a implementação de rotinas automatizadas de quantificação.

A ferramenta MCR-ALS aplicada aos dados de GC-MS permitiu inferir que:

- Os ésteres insaturados são os compostos mais afetados por alterações químicas e os hidrocarbonetos leves os mais afetados por evaporação/ebulição.
- Os ensaios de acordo com as normas D6468 e EN15751 são similares por promover evaporação/ebulição de hidrocarbonetos leves. O segundo ensaio, entretanto, promove muito mais perda de material. A espectrometria IR não conseguiu identificar a perda de compostos leves promovida pelo método D6468, o que não é surpreendente, já que GC-MS é uma técnica analítica muito mais sensível.
- O linoleato de metila tem um comportamento diferenciado dentre os ésteres, sendo consumido nos ensaios de acordo com a norma D7545 e EN15751.

- O ensaio conforme norma EN15751 promove alterações mais severas nas amostras, tanto por ebulição/evaporação quanto por processos químicos.
- É possível visualizar alterações promovidas pelo ensaio conforme D6468 que não são percebidas com a combinação FTIR-PARAFAC.
- MCR-ALS permitiu indicar o consumo de compostos aromáticos nos ensaios de acordo com as normas EN15751 e D6468, o que é bastante relevante, já que isso é apontado como um mecanismo importante de formação de gomas e sedimentos<sup>2,5</sup>.

De forma geral, o estudo confirmou que o biodiesel é a fração mais sensível aos fenômenos químicos de degradação, enquanto o óleo diesel é mais sujeito a alterações por volatilização.

Adicionalmente, o estudo demonstrou que cada ensaio avaliado promove alterações diferentes nas amostras. Dessa forma, não é possível obter os resultados de um ensaio a partir do outro, o que seria especialmente desejável no caso do método previsto na norma D4625, que consome várias semanas e que é considerado como referência para simulação de alterações em campo.

Por fim, observa-se que os resultados das técnicas analíticas “tradicionais” FTIR e GC-MS foram complementados com as ferramentas quimiométricas. PARAFAC e MCR-ALS permitiram destacar padrões e simplificar os dados obtidos, principalmente, excluindo sinais analíticos irrelevantes, o que destaca as informações relevantes para o estudo.

## **SEÇÃO 2 – USO DE CALIBRAÇÃO MULTIVARIADA PARA PREDIÇÃO DE RESULTADOS DE ENSAIO DE ESTABILIDADE UTILIZANDO ESPECTROMETRIA NO INFRAVERMELHO MÉDIO**

Conforme discutido anteriormente, o ensaio de estabilidade que atende à norma EN15751 é considerado como referência para biodiesel e suas misturas com óleo diesel em vários países. Entretanto, conforme descrito no item 4.3.4., trata-se de ensaio que demanda muito tempo e mão de obra. O estudo de Lira et al.<sup>126</sup> mostrou a possibilidade de previsão dos resultados do mesmo ensaio de estabilidade com espectrometria IR e ferramentas quimiométricas. Entretanto, como comentado no item 3.1, o estudo foi limitado a poucas amostras de biodiesel puro obtido em condições controladas. Portanto, os resultados desta seção procuram confirmar a possibilidade de prever o resultado do ensaio de estabilidade descrito pela norma EN15751 empregando um conjunto de amostras de diesel comercial que representem, ao menos de forma aproximada, parte da variabilidade presente nas amostras do mercado atual do diesel.

No presente estudo, espectros no infravermelho médio foram obtidos para um conjunto amplo de amostras coletadas em postos de revenda com o objetivo de criar modelo de calibração multivariada que permitiria prever o resultado do ensaio de estabilidade à oxidação. Dessa forma, seria possível estimar o resultado de um ensaio que requer mais de 20 horas – sem contar as etapas demoradas de limpeza e preparação – a partir de uma análise rápida e bastante reprodutível.

### **7. MATERIAIS E MÉTODOS DA SEÇÃO 2**

#### **7.1. Obtenção das amostras de óleo diesel**

Cinquenta amostras dentre aquelas usualmente coletadas como parte do Programa de Monitoramento de Combustíveis, PMQC, da ANP foram selecionadas para este estudo. As amostras foram coletadas no Distrito Federal entre os meses de novembro de 2019 e janeiro de 2020. Para garantir que o conjunto de amostras fosse representativo de todas as regiões do DF, elas foram gradativamente separadas a cada rodada de coletas. No período, a legislação previa que o óleo diesel comercializado em postos de revenda contivesse 11 % v/v de biodiesel<sup>4</sup>.

As amostras foram coletadas diretamente dos bicos de abastecimento para garrafas de alumínio com capacidade de 1 litro. Elas foram transportadas e deixadas nos

laboratórios da ANP no mesmo dia da coleta, permanecendo em ambiente climatizado até a realização dos ensaios de rotina do PMQC.

## **7.2. Avaliação prévia das amostras de óleo diesel**

As amostras de óleo diesel foram analisadas entre 4 e 16 dias após a coleta nos principais parâmetros estabelecidos na Resolução ANP n° 50 de 23/12/2013, que estabelece os requisitos de qualidade do óleo diesel<sup>156</sup>. Os parâmetros selecionados são aqueles que melhor caracterizam adulterações ou contaminações do produto e já foram detalhados na seção 4.2:

- Ponto de fulgor conforme norma ASTM D93
- Massa Específica conforme norma ABNT NBR 14065
- Enxofre total conforme norma ASTM D5453
- Destilação conforme norma ASTM D86
- Teor de biodiesel conforme norma CEN EN14078
- Aspecto conforme normas ABNT NBR 14954

Logo após a conclusão desses ensaios, os recipientes originais de coleta foram colocados em refrigerador até a realização dos ensaios de estabilidade e espectrometria no infravermelho. Idealmente, o ensaio de estabilidade seria realizado antes dos demais para evitar a exposição ao ar e à temperatura ambiente, mas isso prejudicaria a dinâmica dos ensaios de rotina. Entretanto, entende-se que, para os objetivos deste estudo, uma eventual perda de estabilidade seria pouco relevante, já que o que se busca é a relação matemática entre os resultados de estabilidade e os espectros.

## **7.3. Ensaio de estabilidade oxidativa pelo método da oxidação acelerada**

O ensaio de estabilidade oxidativa pelo método da oxidação acelerada está estabelecido na norma CEN EN 15751 e foi detalhado no item 4.3.4. Os resultados discutidos nessa seção não foram realizados em duplicata como ocorreu na Seção 1. Tendo em vista o maior número de amostras para calibração não foi possível realizar replicatas para esses ensaios. Além disso, os resultados anteriores mostraram que o equipamento é bastante preciso. Observou-se, por exemplo, que os resultados em

duplicata sempre apresentaram variação menor que a repetibilidade estabelecida na norma, que é de 1,3 horas<sup>152</sup> para a média de valores obtidos neste estudo.

#### **7.4. Espectrometria no infravermelho médio**

O equipamento e os acessórios utilizados para obtenção dos espectros no infravermelho utilizados nessa seção, bem como suas configurações, são os mesmos descritos no item 4.4. Uma importante diferença dessa seção é que os espectros foram obtidos em triplicata (três leituras sucessivas), com o objetivo inicial de estimar a estrutura de seus ruídos.

*Nota: Como será discutido adiante, não foi possível criar um modelo adequado para previsão do resultado do ensaio de estabilidade com base nos espectros no infravermelho. Assim sendo, a investigação do ruído e seus efeitos na precisão do modelo de calibração tornaram-se irrelevantes. Entretanto, as triplicatas foram úteis para eliminar a possibilidade de erros grosseiros na obtenção dos espectros. Os modelos desta seção foram obtidos com os espectros médios, ou seja, cada observação é a média de três espectros obtidos em condições de repetibilidade.*

#### **7.5. Calibração multivariada**

Os arquivos eletrônicos com os resultados dos ensaios de espectrometria IR foram exportados no formato de dados para MATLAB<sup>®</sup> e os resultados dos ensaios de estabilidade foram manualmente inseridos. Os cálculos foram realizados com o programa MATLAB<sup>®</sup> versão R2016b (Mathworks<sup>®</sup> Inc.) utilizando o suplemento PLS Toolbox<sup>®</sup> Versão 8.7 (Eigenvector Research<sup>®</sup> Inc.). Para os modelos SVR, PLS Toolbox<sup>®</sup> utiliza o algoritmo LIBSVM versão 2.9<sup>109</sup>.

## **8. RESULTADOS E DISCUSSÃO DA SEÇÃO 2**

### **8.1. Avaliação prévia das amostras**

Os resultados obtidos na avaliação prévia das amostras de óleo diesel são apresentados na Tabela do Apêndice 5 e nos gráficos da Figura 74. Os resultados mostram boa variabilidade e, exceto para o Teor de biodiesel, valores dentro das especificações vigentes. Em função do aumento gradual do teor de biodiesel, e da mistura de diferentes lotes de produtos (com diferentes concentrações de biodiesel) nas distribuidoras e revendedores de combustíveis, são aceitas variações no teor.

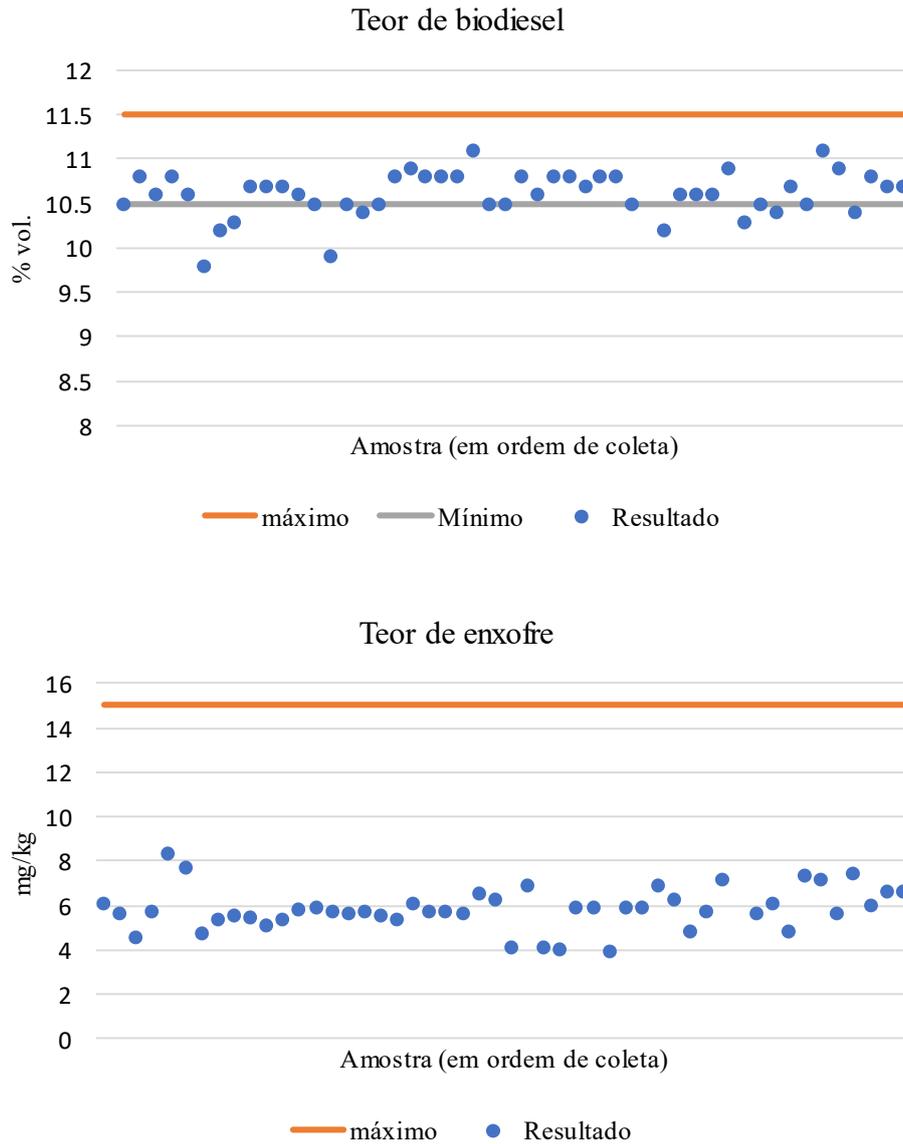
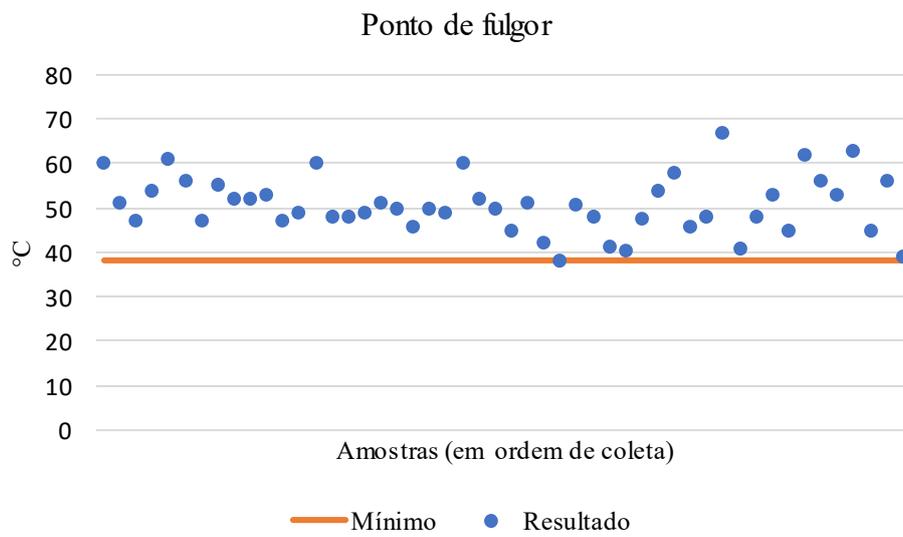
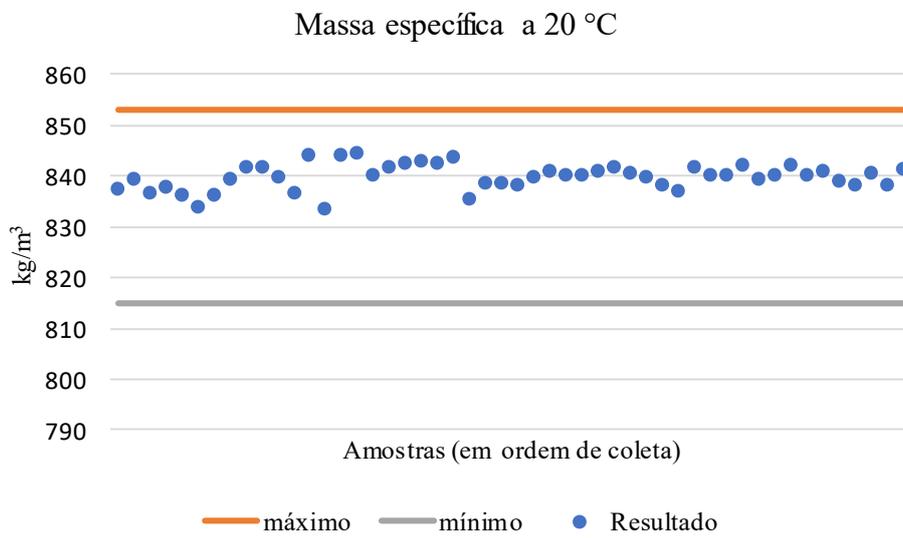
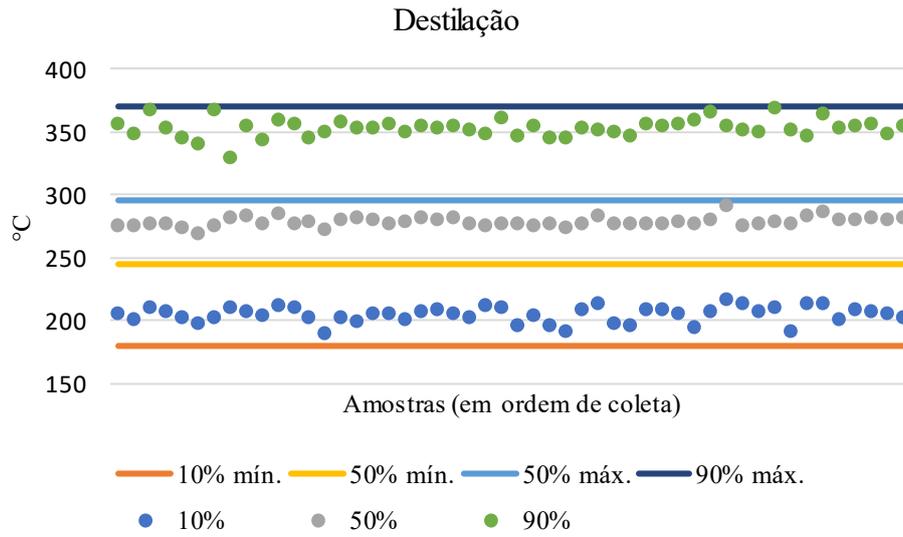


Figura 74: Resultados da avaliação prévia das 50 amostras discutidas na Seção 2. A variabilidade dos resultados de Teor de biodiesel é discutida no texto. O resultado de Teor de enxofre para uma das amostras não estava disponível, assim como um dos resultados de Teor de Biodiesel.

Continuação da Figura 74:



## 8.2. Ensaio de estabilidade oxidativa pelo método da oxidação acelerada

Os resultados para o ensaio de Estabilidade Oxidativa conforme norma EN 15751 são apresentados na Tabela do Apêndice 5 e na Figura 75. Nota-se na figura que não há qualquer tendência clara ao longo do tempo. Por outro lado, o histograma da Figura 76 mostra uma distribuição não homogênea dos resultados, o que é indesejável num conjunto de calibração, conforme discutido em 2.7.2.2. Não surpreende que essa distribuição seja aproximadamente normal, já que reflete o comportamento real de produtos coletados no mercado. As variações observadas podem ser explicadas, principalmente, pela quantidade de biodiesel presente, pelo tipo de matéria prima utilizada na produção do biocombustível e pelos aditivos utilizados na sua produção.

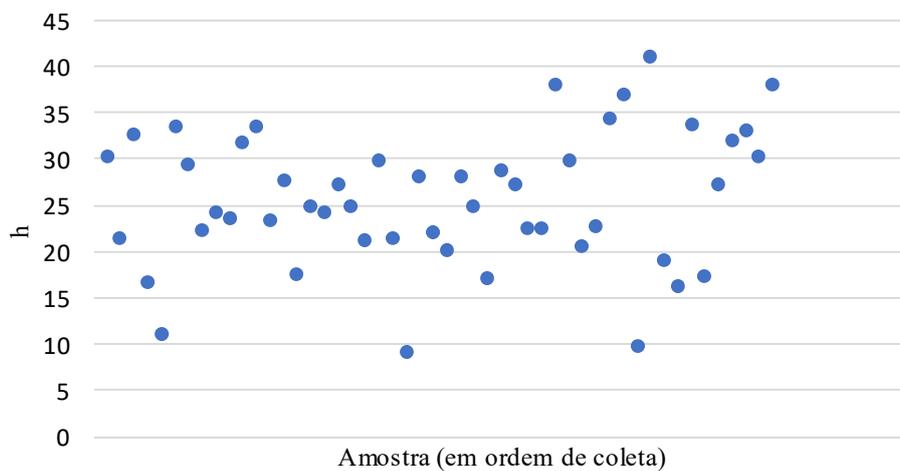


Figura 75: Resultado dos ensaios de Estabilidade à Oxidação conforme norma EN 15751 em 50 amostras de óleo diesel.

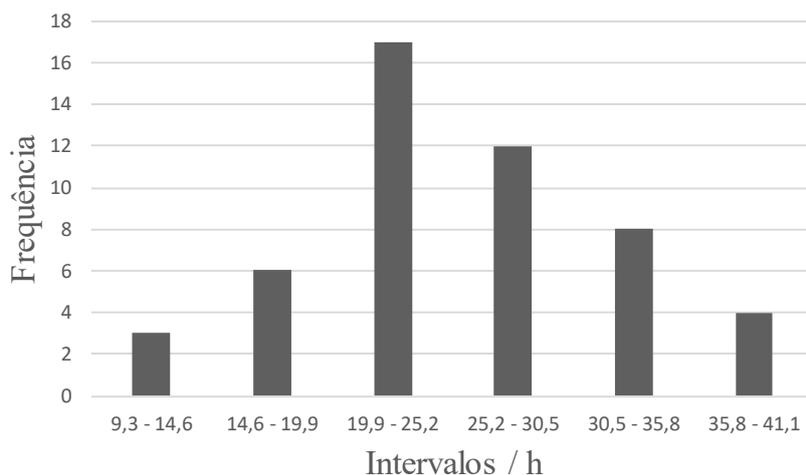


Figura 76: Histograma dos resultados de estabilidade conforme norma EN 15751 em 50 amostras de óleo diesel.

### 8.3. Espectrometria no infravermelho

A Figura 77 mostra os 50 espectros sobrepostos, sendo, cada um, a média de três espectros da mesma amostra. Nota-se que há pequenas variações nos sinais em todas as regiões de absorção. Na região do estiramento C-H e C=O, essas variações são mais pronunciadas, sendo difícil atribuir uma causa para a primeira região, já para a região de C=O, a causa provável é a variação nos teores de biodiesel discutida no item 5.3.

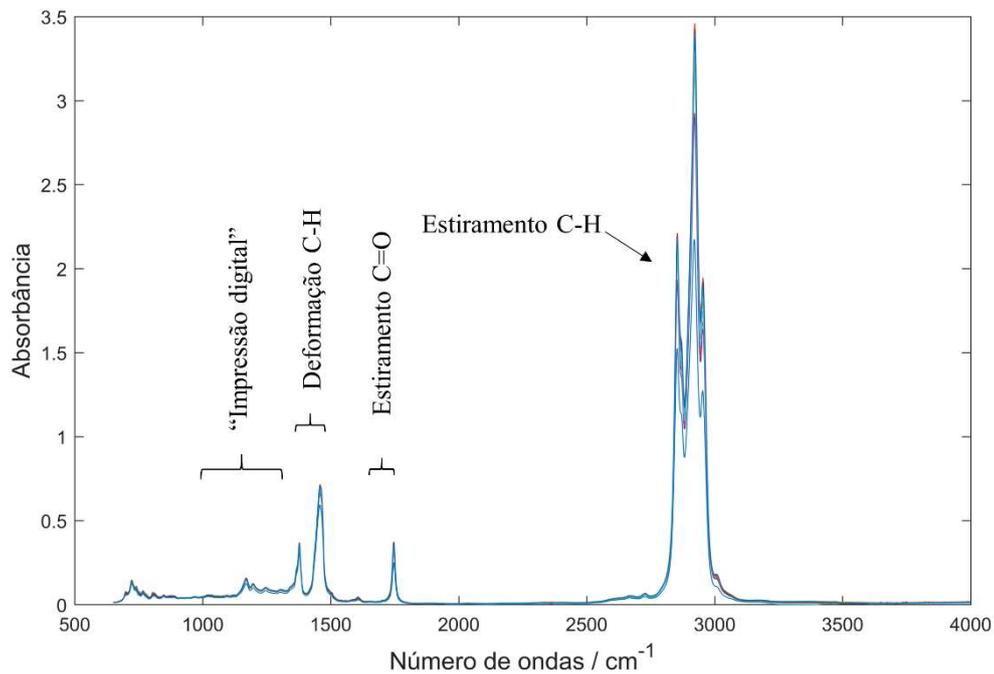


Figura 77: Espectros médios (triplicata) das 50 amostras de óleo diesel utilizadas nesta seção.

Com o objetivo de ilustrar a repetibilidade da obtenção dos espectros, a Figura 78 mostra os três espectros na região de estiramento C-H de uma amostra. Observou-se esse mesmo grau de similaridade em todas as triplicatas. Como discutido anteriormente, os modelos quimiométricos foram elaborados com os espectros médios.

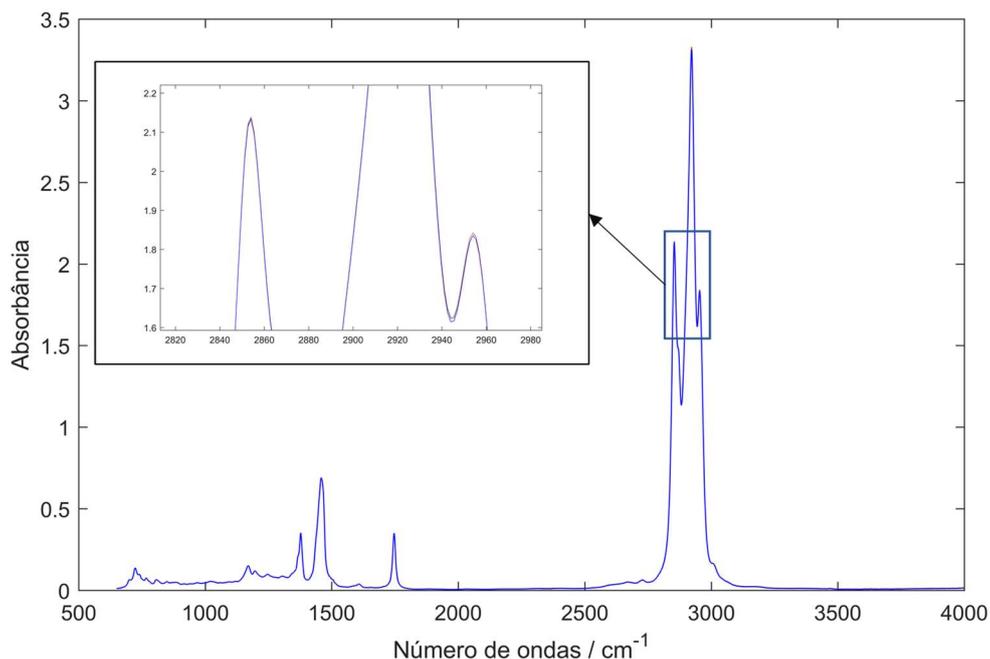


Figura 78: Três espectros de uma mesma amostra com aproximação de região correspondente ao estiramento C-H.

## 8.4. Calibração multivariada

### 8.4.1. Avaliação inicial por Análise por Componentes Principais - PCA

Independentemente da ferramenta de calibração a ser utilizada, é uma boa prática avaliar os dados preliminarmente por meio de Análise por Componentes Principais, PCA. Essa ferramenta multivariada consiste numa transformação linear dos dados originais em um conjunto menor variáveis ortogonais – as componentes principais – que retêm o máximo de variabilidade possível dos dados originais<sup>77</sup>. Isso permite identificar eventuais amostras anômalas, grupos heterogêneos de amostras e regiões espectrais relevantes.

O modelo PCA foi calculado a partir dos 50 espectros (cada espectro sendo a média da triplicata) e obtido por meio do algoritmo de Decomposição de Valores Singulares com centralização dos espectros na média. Os escores das duas primeiras componentes são mostrados na forma de gráfico na Figura 79. A amostra com identificação 29 se destaca na primeira componente – que acumula mais de 94% da variabilidade dos dados – e fica bastante afastada do círculo tracejado que corresponde a um intervalo de confiança de 95%.

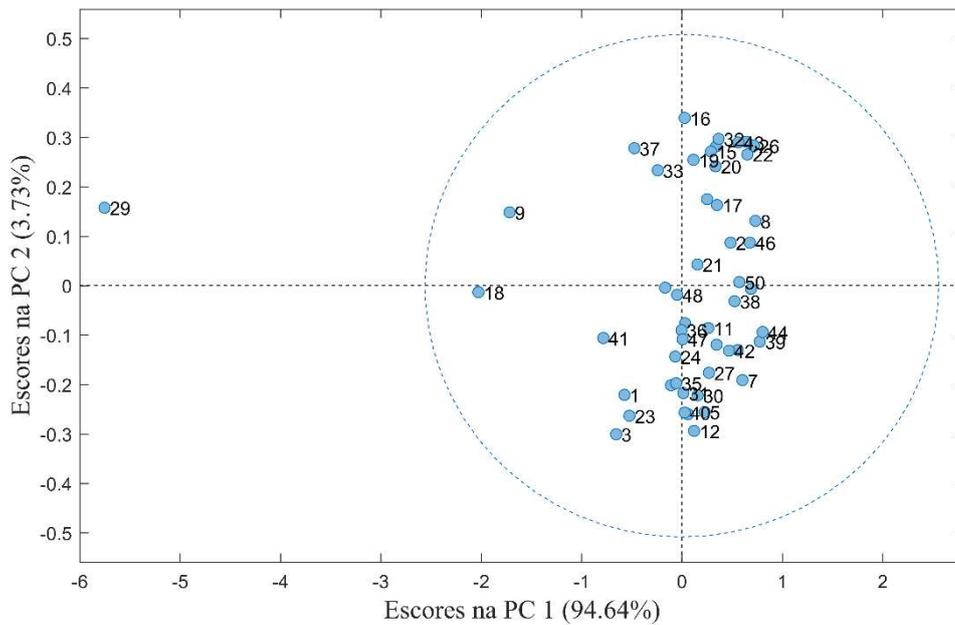


Figura 79: Escores das amostras nas duas primeiras componentes de PCA. A linha tracejada corresponde a um nível de confiança de 95%.

Devido ao seu comportamento anômalo, a amostra 29 foi excluída do conjunto de dados e o restante das amostras foi mantido para o desenvolvimento do modelo de calibração com centralização de dados na média.

#### 8.4.2. Regressão por Mínimos Quadrados Parciais - PLS

Como uma tentativa inicial, todas as amostras foram utilizadas para o desenvolvimento da calibração, não sendo realizada uma divisão entre amostras de calibração e validação. A Figura 80 mostra a variação dos valores de RMSEC e RMSECV para diferentes números de variáveis latentes. Os valores de RMSECV foram calculados com a ferramenta Venetian Blinds com 10 camadas e 1 amostra por camada. O gráfico não apresentou um mínimo evidente para nenhum dos parâmetros, sendo observada apenas a tendência decrescente natural dos valores de RMSEC. Além disso, os valores de RMSECV revelam que o PLS não consegue reduzir os erros com a adição de mais variáveis latentes. De forma tentativa, decidiu-se testar um modelo com 6 variáveis latentes.

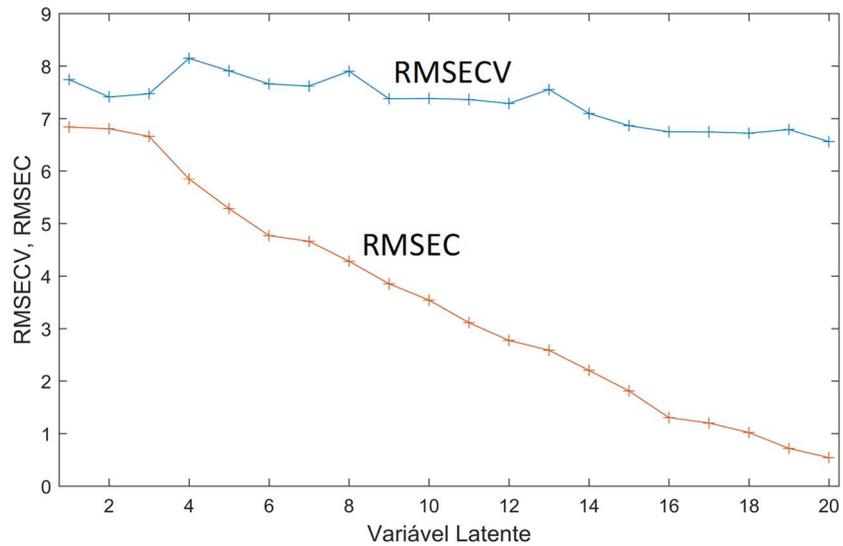


Figura 80: RMSEC e RMSECV para modelo PLS com diferentes números de variáveis latentes (com 49 amostras).

Esse modelo apresentou um RMSECV de 7,66 h e um coeficiente de determinação,  $R^2$  de calibração, de 0,55. Ambos os valores já indicam modelos de baixa qualidade. Adicionalmente, a Figura 81 mostra a distribuição dos erros de predição em função do valor estimado, o que enfatiza a amplitude dos erros.

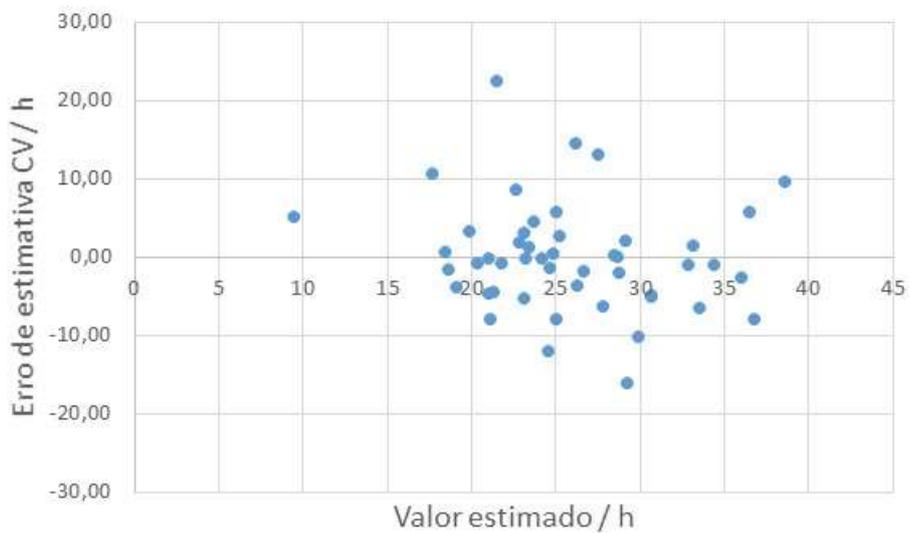


Figura 81: Valores estimados pelo modelo PLS versus erros de predição (validação cruzada) para modelo PLS com 49 amostras e 6 variáveis latentes.

Decidiu-se, então, excluir a amostra 27 em função do elevado valor de resíduos Q e, com base nos valores de RMSEC e RMSECV (Figura 82), optou-se por um modelo com duas variáveis latentes.

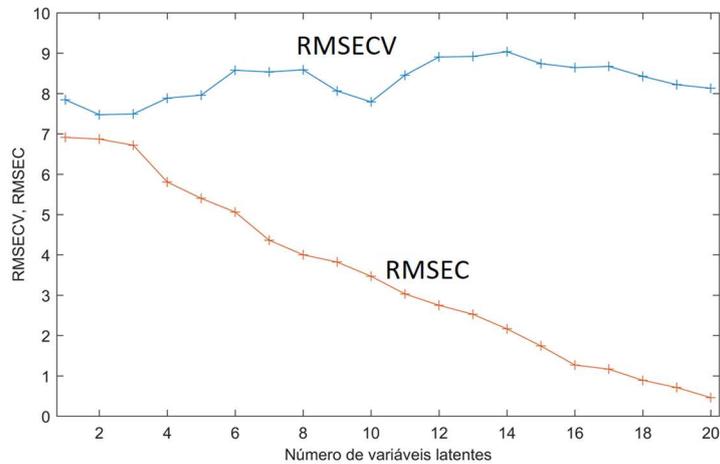


Figura 82: RMSEC e RMSECV para modelo PLS com diferentes números de variáveis latentes (com 48 amostras).

Esse modelo apresentou um RMSECV de 7,47, ligeiramente melhor que o modelo anterior, mas a Figura 83 mostra que outras amostras foram indicadas como anômalas nos parâmetros Q e  $T^2$  Hotelling. Também não houve melhora na tendência dos resultados mostrada na Figura 81. Decidiu-se excluir as seis amostras indicadas como anômalas (3, 5, 6, 9, 14, 18) e reconstruir o modelo.

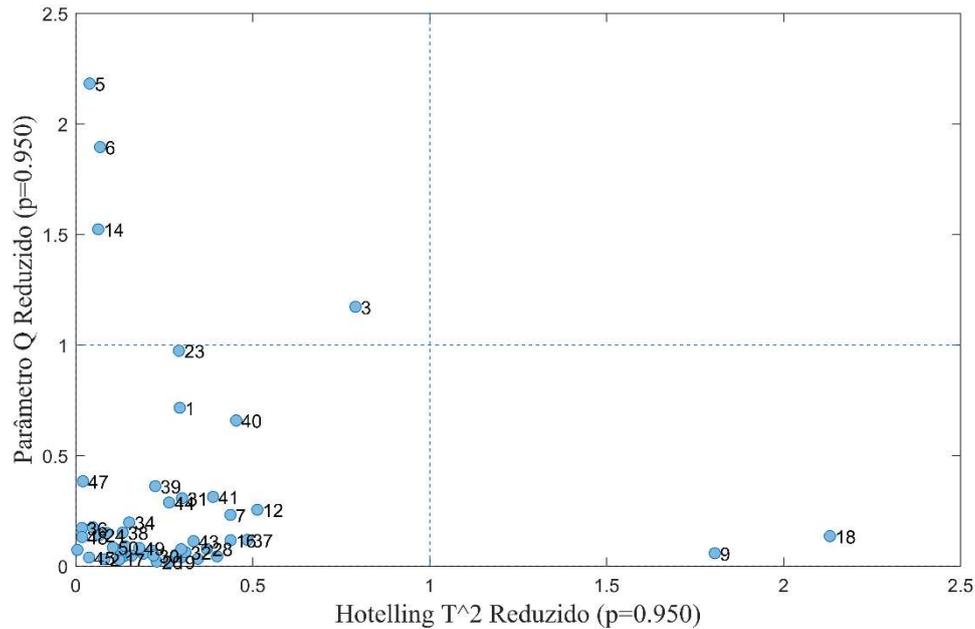


Figura 83: Valores de Hotelling  $T^2$  e Parâmetro Q para amostras do modelo PLS com 2 variáveis latentes após exclusão de 2 amostras anômalas.

Mais uma vez, com base nos valores de RMSEC e RMSECV, decidiu-se por um modelo com 2 variáveis latentes. Este forneceu um valor de RMSECV de 7,67 h; pior do que o obtido com o modelo anterior e com o mesmo tipo de tendência para valores altos e baixos de estabilidade. Além disso, os valores de Q e Hotelling  $T^2$  indicaram mais quatro amostras como anômalas. Esses resultados mostram claramente a ocorrência do fenômeno de “bola de neve”<sup>97</sup>, no qual, a cada retirada de amostras supostamente anômalas, novas amostras são classificadas como anômalas indefinidamente.

Assim sendo, optou-se por explorar outras opções de pré-tratamento de dados após a retirada das amostras 27 e 29, que permitiram obter os melhores valores de RMSECV (utilizando-se apenas centralização de dados pela média).

Os resultados obtidos com diferentes pré-tratamentos serão apresentados adiante em conjunto com aqueles obtidos com Regressão por Vetores de Suporte na Tabela 11. Mesmo nas melhores condições, o valor de RMSECV foi alto (6,88).

#### 8.4.3. Regressão por Vetores de Suporte – SVR

O primeiro modelo de calibração com Vetores de Suporte foi elaborado com os parâmetros abaixo. A função RBF foi escolhido por ser a mais flexível (útil quando o

comportamento do sistema não é conhecido a priori) e o tipo de vetor de suporte ( $v$ -SVR) evita sobreajuste do modelo:

- Exclusão da amostra 29 (com base nos resultados de PCA).
- Kernel: RBF (Radial Basis Function)
- Tipo de vetor de suporte:  $v$ -SVR.
- Pré-tratamento: somente centralização dos dados na média.
- Validação cruzada: Venetian blinds, 10 camadas e uma amostra por subcamada.
- Grade de pesquisa:

$\gamma$ :  $1 \times 10^{-6}$ ;  $3,1 \times 10^{-6}$ ;  $1 \times 10^{-5}$ ;  $1 \times 10^{-4}$ ;  $3 \times 10^{-4}$ ;  $1 \times 10^{-3}$ ;  $1 \times 10^{-2}$ ;  $3 \times 10^{-2}$ ; 0,1; 0,3.

$C$ :  $1 \times 10^{-3}$ ;  $3 \times 10^{-3}$ ;  $1 \times 10^{-2}$ ;  $3 \times 10^{-2}$ ; 0,1; 0,3; 1; 3,2; 10; 31,6; 100.

$\nu$ : 0,2; 0,5; 0,8.

A Figura 84 apresenta os resultados da otimização por grade de pesquisa. O ponto indicado com a letra X marca as combinações de menores erros quadráticos médios para todas as combinações de pares de parâmetros.

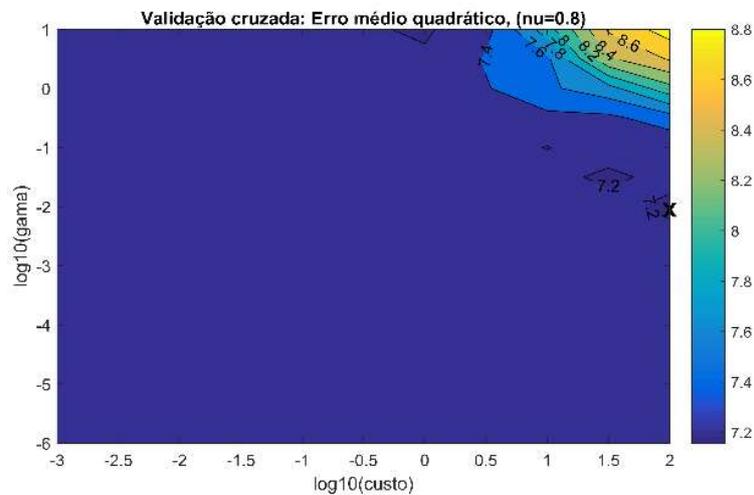
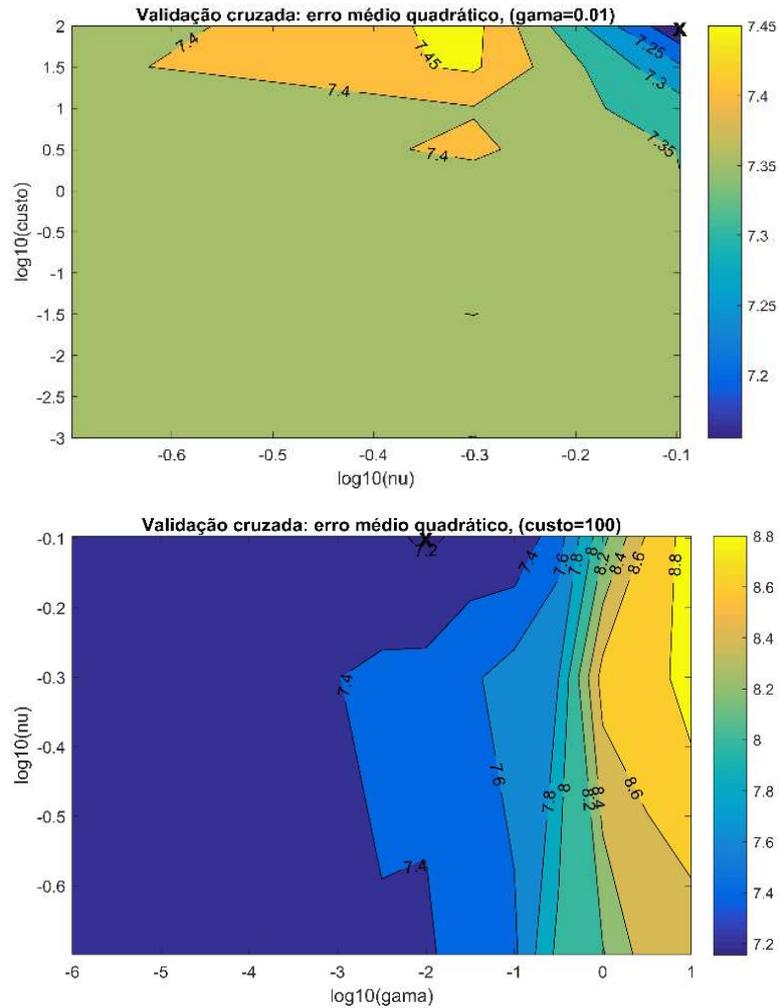


Figura 84: Otimização dos parâmetros de  $v$ -SVR por Grade de Pesquisa. O X indica a combinação de menor erro quadrático médio.

Continuação Figura 84:



O modelo otimizado apresentou as seguintes características:

- Custo: 100
- $\nu$ : 0,8
- $\gamma$ : 0,01
- número de vetores de suporte: 42

Esse modelo apresentou um valor de RMSECV alto: 7,15 h e parece prever resultados muito similares para todas as amostras conforme ilustra a Figura 85. A seguir, outras opções de pré-tratamento de dados foram testadas. Os resultados são apresentados e comparados com os obtidos por PLS na Tabela 11.

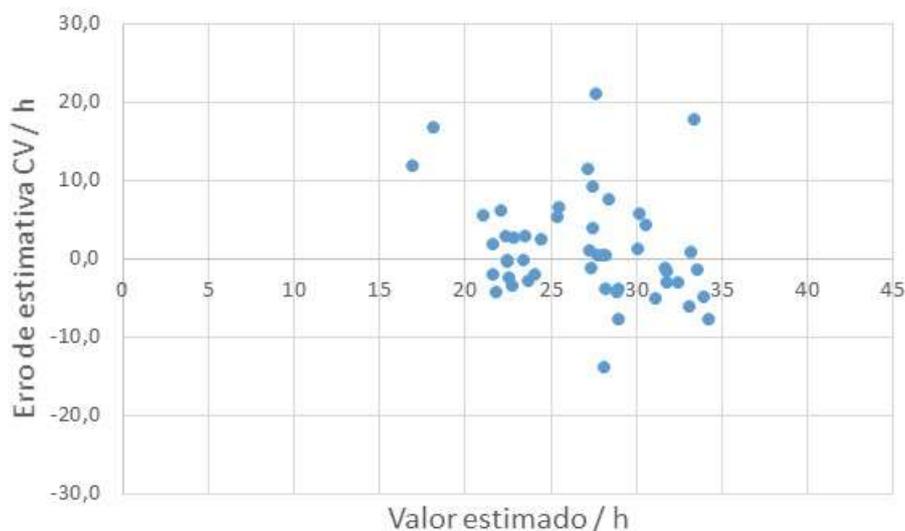


Figura 85: Valores estimados pelo modelo  $\nu$ -SVR versus erro de predição de validação cruzada.

Na Tabela 11 são mostrados os resultados de RMSECV para os modelos PLS e SVR com diferentes pré-tratamento dos dados. Nota-se que nenhum dos modelos apresentou valores satisfatórios de erros de validação cruzada. Chama a atenção o alto número de vetores de suportes otimizados para os modelos SVR. Isso demonstra a dificuldade do modelo em se ajustar aos dados levando a sobreajuste.

Os valores de RMSECV para SVR foram sempre inferiores, mas, para se afirmar que os modelos com vetores de suporte são *estatisticamente* superiores, seria necessária uma avaliação estatística dos resultados, o que não foi possível e não se justifica, já que nenhum dos modelos apresentou-se útil, mesmo como teste semi-quantitativo da estabilidade de uma amostra de óleo diesel. Ademais, não foi possível afirmar que o sistema em estudo apresenta comportamento não linear simplesmente porque os valores dos modelos SVR são numericamente inferiores. De fato, nenhum dos modelos fornece previsões razoáveis do ensaio de estabilidade com base nos espectros das amostras antes do ensaio.

Tabela 11: Valores de RMSECV para os modelos de calibração multivariados com diferentes pré-tratamentos. Os parâmetros dos modelos estão descritos em 8.4.2 e 8.4.3.

Pré-tratamento dos dados	RMSECV / h	
	PLS (n° de variáveis latentes)	v-SVR (C; v; $\gamma$ ; n° de vetores de suporte)
Centralização na média	7,47 (2)	7,15 (100; 0,8; 0,01; 42)
Centralização na média e 1ª. derivada <sup>1</sup>	7,40 (1)	7,07 (10; 0,8; 10; 41)
Centralização na média e 2ª. derivada <sup>2</sup>	7,33 (1)	7,12 (100; 0,8; 10; 41)
Centralização na média e SNV <sup>3</sup>	7,32 (1)	7,00 (1; 0,8; 3,16; 42)
Centralização na média, 1ª. derivada <sup>1</sup> e SNV <sup>3</sup>	7,28 (1)	6,96 (0,31; 0,8; 0,31; 41)
Centralização na média, 2ª. derivada <sup>2</sup> e SNV <sup>3</sup>	6,89 (6)	6,83 (100; 0,8; 0,03; 45)
Centralização na média, 2ª. derivada <sup>2</sup> , SNV <sup>3</sup> e seleção de variáveis <sup>4</sup>	6,88 (6)	6,71 (100; 0,8; 0,32; 49)

<sup>1</sup> Savitzky-Golay, polinômio de 1° grau, janela de 15 pontos.

<sup>2</sup> Savitzky-Golay, polinômio de 2° grau, janela de 15 pontos.

<sup>3</sup> Standard Normal Variate Scaling (Padronização Normal de Sinal)

<sup>4</sup> Utilizou-se VIP (*Variable Importance in Projection* ou Importância da Variável na Projeção)<sup>101</sup>

Do ponto de vista matemático, uma das explicações para o baixo desempenho dos modelos é a baixa variância capturada de  $y$  (vetor contendo os valores da propriedade de interesse), ou seja, o quanto os modelos conseguem capturar da variação dos valores de estabilidade à oxidação. Para o modelo PLS, esse parâmetro é facilmente obtido sendo de apenas 63,7 % para o modelo com o menor valor de RMSECV.

Considerando que as ferramentas de calibração utilizadas são robustas e consolidadas, e que a aquisição dos espectros foi feita de forma adequada, resta concluir que no conjunto de amostras estudado o espectro infravermelho médio não apresentou mudanças espectrais devido a substâncias relevantes para os processos de oxidação/degradação promovidos pelo ensaio conforme norma EN 15751. Outra hipótese, talvez até mais provável, seria atribuir os elevados erros das calibrações à variabilidade existente entre as amostras e aos antioxidantes que são adicionados ao biodiesel e ao óleo diesel. Como esses produtos são adicionados em baixas concentrações, a espectrometria no infravermelho médio seria incapaz de detectá-los e modelá-los adequadamente, mas influenciam de forma significativa na estabilidade oxidativa das amostras. O artigo de Ammawath et al.<sup>157</sup> demonstra o quão difícil é detectar

antioxidantes tipicamente utilizados em biodiesel, mesmo em matrizes de óleo vegetal puro.

Mais uma vez, é interessante comparar os resultados obtidos neste estudo com os de Lira et al.<sup>126</sup>, que utilizaram espectrometria no infravermelho (próximo e médio) e calibração multivariada para prever o resultado do mesmo ensaio de estabilidade utilizado neste estudo em amostras de *biodiesel*. Os modelos em PLS e MLR forneceram resultados bastante exatos, mas é importante observar que, ao contrário do presente trabalho, aquele focou em apenas cinco amostras de biodiesel (com diferentes graus de degradação induzida), sendo que algumas receberam aditivo antioxidante (em quantidade controlada). Em suas conclusões, os próprios autores admitem que a aplicabilidade do modelo pode ser afetada por substâncias não detectáveis por espectrometria no infravermelho.

## 9. CONCLUSÕES DA SEÇÃO 2

Os resultados demonstraram que não foi possível prever o resultado do ensaio de estabilidade de acordo com a norma EN15751 com base nos espectros no infravermelho médio nas condições deste estudo. Essa afirmativa é possível em função de uma série de cuidados tomados no planejamento e execução do estudo:

- Os espectros foram obtidos em triplicata em equipamento operando nas condições especificadas pelo fabricante.
- Foram utilizadas ferramentas multivariadas consagradas, inclusive SVR, que não depende de relação linear entre variáveis preditoras e preditas.
- Foi obtido um número significativo de amostras com propriedades bastante distintas e uma faixa de resultados ampla para o parâmetro de interesse.
- As amostras incluídas no modelo foram coletadas e transportadas em condições previstas em normas e foram aprovadas nos principais ensaios previstos na legislação, o que evita que produtos com características anômalas afetem o modelo.

A explicação mais direta para os resultados obtidos é a de que a espectrometria IR não consegue detectar a(s) substância(s) que são realmente importantes para prever o comportamento do produto nas condições do ensaio. De forma mais especulativa, pode-se inferir que aditivos antioxidantes (especialmente do biodiesel) sejam importantes para os resultados, mas sua baixa concentração não permite que sejam detectados. Destaca-se

que a constituição e teor desses aditivos pode variar entre as produtoras do óleo diesel, o que também contribui para dificultar a obtenção de uma relação entre o espectro e a estabilidade das amostras. Por fim, outros compostos que afetam a estabilidade do óleo diesel podem agir de forma isolada ou não como antioxidantes do óleo diesel, antioxidantes naturais remanescentes da matéria prima do biodiesel ou, ainda, traços de metais (especialmente de transição).

## 10. CONCLUSÕES GERAIS

Dentre os resultados deste estudo, ficou confirmado que o óleo diesel é uma matriz complexa assim como os fenômenos envolvidos nos ensaios de estabilidade. Apesar disso, o uso de espectrometria no infravermelho médio associada à ferramenta quimiométrica PARAFAC conseguiu, de maneira sistemática, apontar as principais alterações promovidas pelos ensaios conforme normas D4625, D7545 e EN15751. Para o ensaio conforme D6468, não foi possível identificar alterações específicas no infravermelho, possivelmente, porque elas foram mascaradas por variações entre amostras. Além disso, PARAFAC permitiu destacar as diferenças entre amostras, o que destaca a importância da variabilidade amostral existente nas amostras de diesel comercial. A aplicação de MCR-ALS nos ensaios de GC-MS não permitiu identificar em detalhes muitos produtos de degradação minoritários, o que pode decorrer das limitações da cromatografia gasosa em detectar compostos de alto peso molecular ou da intensa coeluição dos compostos. Entretanto, a decomposição dos dados GC-MS promovida por MCR facilitou a interpretação dos principais fenômenos observados nos ensaios de estabilidade: perda de hidrocarbonetos mais voláteis nos ensaios conforme normas D6468 e EN15751 e oxidação dos ésteres mais insaturados de forma significativa no ensaio conforme norma EN15751. A técnica de GC-MS foi mais sensível que o FTIR e pode fornecer com maior riqueza de detalhes informações sobre a diminuição da concentração de alguns compostos específicos, tais como os relacionados ao biodiesel. Contudo, infelizmente, a restrição do estudo a três amostras impossibilitou a obtenção de maior detalhamento nas conclusões.

Apesar do sucesso na identificação de alterações nos ensaios de estabilidade com FTIR, prever os resultados do ensaio de estabilidade mais utilizado pelo mercado (EN15751) a partir dos espectros IR do óleo diesel mostrou-se inviável. Mesmo com diferentes pré-tratamentos dos dados espectrais, PLS (a ferramenta de calibração multivariada mais utilizada) não foi capaz de prever os resultados do ensaio de estabilidade com exatidão razoável. A calibração com vetores de suporte, SVR, que permite criar modelos não lineares também não apresentou resultados significativamente melhores, o que confirmou que a espectrometria no infravermelho médio, nas condições deste estudo, não consegue “enxergar” as substâncias que, de fato, são responsáveis pelo comportamento óleo diesel diante das condições de estresse do ensaio de estabilidade.

## 11. BIBLIOGRAFIA

- (1) Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Dados Estatísticos <http://www.anp.gov.br/dados-estatisticos> (acessado em 8 de fev. de 2020).
- (2) Farah, M. A. *Petróleo e seus Derivados*, 1a. ed.; LTC: Rio de Janeiro, 2012.
- (3) Brasil. *Lei N° 11.097, de 13 de Janeiro de 2005. Dispõe sobre a Introdução do Biodiesel na Matriz Energética Brasileira*; 2005.
- (4) *Conselho Nacional de Política Energética - Resolução N° 16, de 29 de Outubro de 2018.*; Conselho Nacional de Política Energética: Brasil, 2018.
- (5) Batts, B. D.; Fathoni, A. Z. A Literature Review of Fuel Stability Studies with Particular Emphasis on Diesel Oil. *Energy and Fuels* **1991**, No. 5, 2–21.
- (6) Knothe, G.; Gerpen, J. Van; Krahl, J.; Ramos, L. P. *Manual do Biodiesel*, 1a. ed.; Edgard Blücher: São Paulo, 2006.
- (7) Holanda, A. *Biodiesel e Inclusão Social*, 1a.edição.; Câmara dos Deputados, Coordenação de Publicações: Brasília, 2004.
- (8) Biostab. *Stability of Biodiesel - Summary Report*; BLT Wieselburg, Áustria, 2003.
- (9) Besser, C.; Pisarova, L.; Frauscher, M.; Hunger, H.; Litzow, U.; Orfaniotis, A.; Dörr, N. Oxidation Products of Biodiesel in Diesel Fuel Generated by Artificial Alteration and Identified by Mass Spectrometry. *Fuel* **2017**, 206, 524–533.
- (10) Haupt, J.; Brankatschk, G.; Wilharm, T. *Sterol Glucoside Content in Vegetable Oils as a Risk for the Production of Biodiesel – Study of the Technological Chain Impact Sterol Glucoside Content in Vegetable Oils as a Risk for the Production of Biodiesel - Final Report. AGQM.*; Berlin, Germany, 2009.
- (11) ASTM - International. *D6468 - Standard Test Method for High Temperature Stability of Middle Distillate Fuels 1*; 2008; pp 8–13.
- (12) Østerstrøm, F. F.; Anderson, J. E.; Mueller, S. A.; Collings, T.; Ball, J. C.; Wallington, T. J. Oxidation Stability of Rapeseed Biodiesel/Petroleum Diesel Blends. *Energy and Fuels* **2016**, 30 (1), 344–351.
- (13) Bacha, K.; Ben-Amara, A.; Vannier, A.; Alves-Fortunato, M.; Nardin, M. Oxidation Stability of Diesel/biodiesel Fuels Measured by a Petrooxy Device and Characterization of Oxidation Products. *Energy and Fuels* **2015**, 29 (7), 4345–4355.
- (14) ASTM - International. *D4625 - Standard Test Method for Middle Distillate Fuel Storage Stability at 43 °C ( 110 °F )*; 2016; pp 1–5.
- (15) ASTM - International. *D7545 - Oxidation Stability of Middle Distillate Fuels - Rapid Small Scale Oxidation Test (RSSOT)*; 2014; p 7.
- (16) ASTM - International. *D975 - Standard Specification for Diesel Fuel Oils*; 2018; pp 1–25.
- (17) ASTM - International. *D7462 - Standard Test Method for Oxidation Stability of Biodiesel ( B100 ) and Blends of Biodiesel with Middle Distillate Petroleum Fuel (Accelerated Method)*; 2011; Vol. i, pp 1–6.

- (18) Confederação Nacional do Transporte. Transporte em números. Dezembro de 2019. <https://www.cnt.org.br/analises-transporte> (acessado 23 de mar. de 2020).
- (19) Szklo, A. S. *Fundamentos do Refino de Petróleo*, 1a.ed.; Interciência: Rio de Janeiro, 2005.
- (20) *Agência Nacional do Petróleo, Gás Natural e Biocombustíveis - Resolução ANP N° 45 de 25.8.2014*; Brasil, 2014; pp 1–10.
- (21) Bonazza, B.; Herman, M.; Saunders, B.; Bennett, J.; Kloss, E.; da Silva Junior, J. F. *White Paper on Internationally Compatible Biofuel Standards - Tripartite Task Force Brazil, European Union & United States of America*; 2007.
- (22) European Committee for Standardization. *BS EN 14214:2012+A1:2014 Liquid Petroleum Products - Fatty Acid Methyl Esters (FAME) for Use in Diesel Engines and Heating Applications - Requirements and Test Methods*; 2013; p 28.
- (23) ASTM - International. *D6751 - Standard Specification for Biodiesel Fuel Blend Stock*; Estados Unidos, 2015; pp 1–10.
- (24) European Committee for Standardization. *EN 590 - Automotive Fuels — Diesel — Requirements and Test Methods*; 2013.
- (25) Rakopoulos, D. C.; Rakopoulos, C. D.; Giakoumis, E. G. Impact of Properties of Vegetable Oil, Bio-Diesel, Ethanol and N-Butanol on the Combustion and Emissions of Turbocharged HDDI Diesel Engine Operating under Steady and Transient Conditions. *Fuel* **2015**, *156*, 1–19.
- (26) Guerra, E.; Fuchs, W. Biocombustível Renovável: Uso de Óleo Vegetal em Motores. *Rev. Acadêmica Ciências Agrár. Ambient.* **2010**, *8* (I), 103–112.
- (27) Ramadhas, A. S.; Jayaraj, S.; Muraleedharan, C. Use of Vegetable Oils as I.C. Engine Fuels - A Review. *Renew. Energy* **2004**, *29* (5), 727–742.
- (28) Capuano, D.; Costa, M.; Di Fraia, S.; Massarotti, N.; Vanoli, L. Direct Use of Waste Vegetable Oil in Internal Combustion Engines. *Renew. Sustain. Energy Rev.* **2017**, *69* (October 2016), 759–770.
- (29) Barreto, C. C. K.; Oliveira, C. C.; Souza, G. G.; Suarez, P. A. Z.; Rubim, J. C. Evaluation of the Stability during Storage of a Diesel-like Fuel Obtained by the Pyrolysis of Soybean Oil. *Biomass and Bioenergy* **2012**, *37*, 42–48.
- (30) Silva, A. A. L.; Santos, A. G. D.; Di Souza, L.; Caldeira, V. P. da S.; Júnior, G. E. L.; Araújo, A. S. Síntese e Caracterização de Biodiesel de Sebo Bovino e de sua Mistura B10. *Orbital - Electron. J. Chem.* **2015**, *7* (1), 21–27.
- (31) Franco, Z.; Nguyen, Q. D. Flow Properties of Vegetable Oil-Diesel Fuel Blends. *Fuel* **2011**, *90* (2), 838–843.
- (32) Senthur Prabu, S.; Asokan, M. A.; Prathiba, S.; Ahmed, S.; Puthean, G. Effect of Additives on Performance, Combustion and Emission Behavior of Preheated Palm Oil/diesel Blends in DI Diesel Engine. *Renew. Energy* **2018**, *122*, 196–205.
- (33) De, B.; Panua, R. S. An Experimental Study on Performance and Emission Characteristics of Vegetable Oil Blends with Diesel in a Direct Injection Variable Compression Ignition Engine. *Procedia Eng.* **2014**, *90*, 431–438.

- (34) Atmanli, A.; Ileri, E.; Yilmaz, N. Optimization of Diesel-Butanol-Vegetable Oil Blend Ratios Based on Engine Operating Parameters. *Energy* **2016**, *96*, 569–580.
- (35) Pinzi, S.; López, I.; Leiva-Candia, D. E.; Redel-Macías, M. D.; Herreros, J. M.; Cubero-Atienza, A.; Dorado, M. P. Castor Oil Enhanced Effect on Fuel Ethanol-Diesel Fuel Blend Properties. *Appl. Energy* **2018**, *224* (January), 409–416.
- (36) McCormick, R. L.; Westbrook, S. R. *Empirical Study of the Stability of Biodiesel and Biodiesel Blends Milestone Report Empirical Study of the Stability of Biodiesel and Biodiesel Blends Milestone Report*; 2007.
- (37) Obadiah, A.; Kannan, R.; Ramasubbu, A.; Kumar, S. V. Studies on the Effect of Antioxidants on the Long-Term Storage and Oxidation Stability of Pongamia Pinnata (L.) Pierre Biodiesel. *Fuel Process. Technol.* **2012**, *99*, 56–63.
- (38) Moratto, E.; Alves, R. F. *Óleos E Gorduras Vegetais (Processamento E Análises)*, 1a. ed.; Universidade Federal de Santa Catarina: Florianópolis, 1986.
- (39) Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Anuário estatístico 2020. <http://www.anp.gov.br/arquivos/central-conteudos/anuario-estatistico/2020/texto-secao-4.pdf> (acessado em 25 de julho de 2020).
- (40) Revista Biodiesebr. Etanol X Metanol: Disputa Inflmável. 2008. <https://www.biodieselbr.com/revista/003/etanol-metanol-disputa> (acessado em 10 de setembro de 2020).
- (41) Lacerda, J., Entrevista concedida a Vinicius Skrobot em visita técnica à usina de biodiesel da Caramuru. Ipameri-GO. 07 de Junho de 2017.
- (42) Ma, F.; Hanna, M. A. Biodiesel Production: A Review. *Bioresour. Technol.* **1999**, *70* (1), 1–15.
- (43) *Lei Nº 13.263, de 23 de Março de 2016. Altera a Lei Nº 13.033, de 24 de Setembro de 2014, Para dispor sobre os percentuais de adição de biodiesel ao óleo diesel comercializado no território nacional.* 2016.
- (44) *Brasil. Lei Nº. 13.263, de 23 de Março de 2016. Altera a Lei Nº 13.033, de 24 de setembro de 2014, Para dispor sobre os percentuais de adição de biodiesel ao óleo diesel comercializado no território nacional.* 2016; Vol. 153, pp 93–94.
- (45) *Agência Nacional Do Petróleo, Gás Natural E Biocombustíveis - Despacho Nº 621 de 6 de Agosto de 2019; Diário Oficial da União: Brasil, 2019; p 2019.*
- (46) Johnson, K. J.; Rose-Pehrsson, S. L.; Morris, R. E. Monitoring Diesel Fuel Degradation by Gas Chromatography - Mass Spectroscopy and Chemometric Analysis. *Energy and Fuels* **2004**, *18* (3), 844–850.
- (47) Goede, S.; Wilken, C.; Ajam, M.; Roets, P.; Engelbrecht, P.; Woolard, C. A Comparison of the Stability Performance of Blends of Paraffinic Diesel and Petroleum-Derived Diesel with RME Biodiesel Using Laboratory Stability Measurement Techniques. *J. Fuels* **2015**, *2015*, 15.
- (48) Christensen, E.; McCormick, R. L. Long-Term Storage Stability of Biodiesel and Biodiesel Blends. *Fuel Process. Technol.* **2014**, *128*, 339–348.
- (49) Saluja, R. K.; Kumar, V.; Sham, R. Stability of Biodiesel – A Review. *Renew. Sustain. Energy Rev.* **2016**, *62*, 866–881.

- (50) Knothe, G. Analyzing Biodiesel: Standards and Other Methods. *JAOCs, J. Am. Oil Chem. Soc.* **2006**, 83 (10), 823–833.
- (51) Farias, R. M. C.; Conceição, M. M.; Candeia, R. A.; Silva, M. C. D.; Fernandes, V. J.; Souza, A. G. Evaluation of the Thermal Stability of Biodiesel Blends of Castor Oil and Passion Fruit. *J. Therm. Anal. Calorim.* **2011**, 106 (3), 651–655.
- (52) Flitsch, S.; Neu, P. M.; Schober, S.; Kienzl, N.; Ullmann, J.; Mittelbach, M. Quantitation of Aging Products Formed in Biodiesel during the Rancimat Accelerated Oxidation Test. *Energy and Fuels* **2014**, 28 (9), 5849–5856.
- (53) Wang, H.; Tang, H.; Salley, S.; Simon Ng, K. Y. Analysis of Sterol Glycosides in Biodiesel and Biodiesel Precipitates. *J. Am. Oil Chem. Soc.* **2010**, 87 (2), 215–221.
- (54) Bezerra, K. da S.; Antoniosi Filho, N. R. Comparative Analysis of Content in Free Steroids of Different Oils, Fats and Biodiesel By Gas Chromatography. *Quim. Nova* **2015**, 38 (4), 498–505.
- (55) Pieber, B.; Schober, S.; Goebel, C.; Mittelbach, M. Novel Sensitive Determination of Steryl Glycosides in Biodiesel by Gas Chromatography-Mass Spectroscopy. *J. Chromatogr. A* **2010**, 1217 (42), 6555–6561.
- (56) AEA - Associação Brasileira de Engenharia Automotiva. Workshop “Combustíveis : Características , Qualidade e Ocorrências No Campo.” In *AEA - Associação Brasileira de Engenharia Automotiva. Resumo Consolidado do Workshop “Combustíveis : Características , Qualidade e Ocorrências no Campo”*; Rio de Janeiro, 2018.
- (57) Nayyar, A.; Sharma, D.; Soni, S. L.; Mathur, A. Characterization of N-Butanol Diesel Blends on a Small Size Variable Compression Ratio Diesel Engine: Modeling and Experimental Investigation. *Energy Convers. Manag.* **2017**, 150 (May), 242–258.
- (58) ASTM. *Standard Test Method for ASTM Color of Petroleum Products (ASTM Color Scale) D1500*; United States of America, 2017.
- (59) *ASTM D4625 - Standard Test Method for Middle Distillate Fuel Storage Stability at 43°C (110 °F)*; United States of America. 2014.
- (60) *ASTM D2274. Standard Test Method for Oxidation Stability of Distillate Fuel Oil (Accelerated Method)*; 2014; Vol. i, pp 1–6.
- (61) Skoog, D. A.; Holler, F. J.; Nieman, T. A. *Princípios de Análise Instrumental*, 5a. edição. Bookman: Porto Alegre, 2002.
- (62) Silverstein, R. M.; Webster, Francis, X. *Identificação Espectrométrica de Compostos Orgânicos*, 6a. edição.; Editora, LTC: Rio de Janeiro, 2000.
- (63) *PIKE - Technologies - Catálogo Pike Technologies - FTIR, NIR and UV-Vis Accessories and Supplies*; Madison, 2017; Vol. Vol. 1.
- (64) Bruker Optik GmbH. *Tensor II - Manual de Usuário*; Ettlingen, 2016.
- (65) Le Dréau, Y.; Dupuy, N.; Artaud, J.; Ollivier, D.; Kister, J. Infrared Study of Aging of Edible Oils by Oxidative Spectroscopic Index and MCR-ALS Chemometric Method. *Talanta* **2009**, 77 (5), 1748–1756.

- (66) Agilent Technologies. Catálogo de consumíveis para cromatografia e espectroscopia Agilent [www.agilent.com/cs/library/catalogs/Public/5991-5213PTBR](http://www.agilent.com/cs/library/catalogs/Public/5991-5213PTBR) (acessado em 11 de julho de 2020).
- (67) Settle, F. A. *Handbook of Instrumental Techniques for Analytical Chemistry*, 1a. ed.; Prentice Hall, 1997.
- (68) Mohd, M. A. Derivatization Reactions and Reagents for Gas Chromatography Analysis. In *Advanced Gas Chromatography - Progress in Agricultural, Biomedical and Industrial Applications*; 2007.
- (69) Kitson, F. G.; Larsen, B. S.; McEwen, C. N. *Gas Chromatography and Mass Spectrometry - A Practical Guide*, 1a. ed.; Academic Press: London, 1996.
- (70) IUPAC. Compendium of Chemical Terminology - Gold Book <http://goldbook.iupac.org/PDF/goldbook.pdf> <http://goldbook.iupac.org/I03352.html> (acessado em 4 de julho de 2020).
- (71) Mingoti, S. A. *Análise de Dados Através de Métodos de Estatística Multivariada - Uma Abordagem Aplicada*, 1a.ed.; UFMG, Editora: Belo Horizonte, 2005.
- (72) Smilde, A. K.; Rasmus, B.; Geladi, P. *Multi-Way Analysis Applications in the Chemical Sciences*; John Wiley & Sons, Ltd, 2004.
- (73) Sena, M. M.; Trevisan, M. G.; Poppi, R. J. PARAFAC: Uma Ferramenta Quimiométrica para Tratamento de Dados Multidimensionais. Aplicações na Determinação Direta de Fármacos em Plasma Humano por Espectrofluorimetria. *Quim. Nova* **2005**, 28 (5), 910–920.
- (74) Sánchez, E.; Kowalski, B. R. Generalized Rank Annihilation Factor Analysis. *Anal. Chem.* **1986**, 58 (2), 496–499.
- (75) Tauler, R.; Izquierdo-Ridorsa, A.; Casassas, E. Simultaneous Analysis of Several Spectroscopic Titrations with Self-Modeling Curve Resolution. *Chemom. Intell. Lab. Syst.* **1993**, 18, 293–300.
- (76) Tauler, R. R. Multivariate Curve Resolution Applied to Second Order Data. *Chemom. Intell. Lab. Syst.* **1995**, 30 (1), 133–146.
- (77) Ferreira, M. M. C. *Quimiometria - Conceitos, Métodos e Aplicações*, 1a. ed.; Unicamp: Campinas, 2015.
- (78) Bro, R. Tese: Multi-Way Analysis in the Food Industry. Models, Algorithms, and Applications., Royal Veterinary and Agricultural University - Denmark, 1998.
- (79) Gardlo, A.; Smilde, A. K.; Hron, K.; Hrdá, M.; Karlíková, R.; Friedecký, D.; Adam, T. Normalization Techniques for PARAFAC Modeling of Urine Metabolomic Data. *Metabolomics* **2016**, 12 (117).
- (80) Bro, R. PARAFAC. Tutorial and Applications. *Chemom. Intell. Lab. Syst.* **1997**, 38 (2), 149–171.
- (81) Braga, J. W. B. Tese: Aplicação e Validação de Modelos de Calibração de Segunda Ordem em Química Analítica, Universidade Estadual de Campinas, 2008.

- (82) Morais, E. C.; Esmerino, E. A.; Monteiro, R. A.; Pinheiro, C. M.; Nunes, C. A.; Cruz, A. G.; Bolini, H. M. A. Prebiotic Low Sugar Chocolate Dairy Desserts: Physical and Optical Characteristics and Performance of PARAFAC and PCA Preference Map. *J. Food Sci.* **2016**, *81* (1), S156–S164.
- (83) Ortiz, M. C.; Sanllorente, S.; Herrero, A.; Reguera, C.; Sarabia, L. A. Impact of Time and Temperature of Storage on the Spoilage of Swordfish and the Evolution of Biogenic Amines through a Multiway Model. *J. Chemom.* **2018**, *32* (3), e2965.
- (84) Smilde, A. K. Three-Way Analyses Problems and Prospects. *Chemom. Intell. Lab. Syst.* **1992**, *15* (2–3), 143–157.
- (85) Bro, R. Review on Multiway Analysis in Chemistry - 2000-2005. *Crit. Rev. Anal. Chem.* **2006**, *36* (3–4), 279–293.
- (86) Tauler, R. Multivariate Curve Resolution: theory and applications <http://www.mcrals.info/> (acessado em 9 de setembro de 2020).
- (87) Maeder, M.; Zuberbuehler, A. D. The Resolution of Overlapping Chromatographic Peaks by Evolving Factor Analysis. *Anal. Chim. Acta* **1986**, *181* (C), 287–291.
- (88) Março, P. H.; Valderrama, P.; Alexandrino, G. L.; Poppi, R. J.; Tauler, R. Resolução Multivariada de Curvas com Mínimos Quadrados Alternantes: Descrição, Funcionamento e Aplicações. *Quim. Nova* **2014**, *37* (9), 1525–1532.
- (89) Jaumot, J.; Gargallo, R.; De Juan, A.; Tauler, R. A Graphical User-Friendly Interface for MCR-ALS: A New Tool for Multivariate Curve Resolution in MATLAB. *Chemom. Intell. Lab. Syst.* **2005**, *76* (1), 101–110.
- (90) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; 2003; Vol. 8.
- (91) Jaumot, J.; de Juan, A.; Tauler, R. MCR-ALS GUI 2.0: New Features and Applications. *Chemom. Intell. Lab. Syst.* **2015**, *140*, 1–12.
- (92) Windig, W.; Guilment, J. Interactive Self-Modeling Mixture Analysis. *Anal. Chem.* **1991**, *63* (14), 1425–1432.
- (93) Bro, R.; de Jong, S. A Fast Non-Negative Constrained Linear Least Squares Algorithm. *J. Chemom.* **1997**, *11* (Outubro de 1996), 393–401.
- (94) De Juan, A.; Vander Heyden, Y.; Tauler, R.; Massart, D. L. Assessment of New Constraints Applied to the Alternating Least Squares Method. *Anal. Chim. Acta* **1997**, *346* (3), 307–318.
- (95) Horai, H.; Arita, M.; Kanaya, S.; Nihei, Y.; Ikeda, T.; Suwa, K.; Ojima, Y.; Tanaka, K.; Tanaka, S.; Aoshima, K.; et al. MassBank: A Public Repository for Sharing Mass Spectral Data for Life Sciences. *J. Mass Spectrom.* **2010**, *45* (7), 703–714.
- (96) Stein, S. E.; Scott, D. R. Optimization and Testing of Mass Spectral Library Search Algorithms for Compound Identification. *J. Am. Soc. Mass Spectrom.* **1994**, *5* (9), 859–866.
- (97) ASTM - International. *ASTM - E1655 - Standard Practices for Infrared Multivariate Quantitative Analysis*; United States of America, 2019.
- (98) Bayne, C. K.; Kramer, R. *Chemometric Techniques for Quantitative Analysis*;

- 1999; Vol. 41.
- (99) Forina, M.; Lanteri, S.; Casale, M. Multivariate Calibration. *J. Chromatogr. A* **2007**, *1158*, 61–93.
- (100) Flood, M. E.; Connolly, M. P.; Comiskey, M. C.; Hupp, A. M. Evaluation of Single and Multi-Feedstock Biodiesel - Diesel Blends Using GCMS and Chemometric Methods. *Fuel* **2016**, *186*, 58–67.
- (101) Mehmood, T.; Liland, K. H.; Snipen, L.; Sæbø, S. A Review of Variable Selection Methods in Partial Least Squares Regression. *Chemom. Intell. Lab. Syst.* **2012**, *118*, 62–69.
- (102) Inc., E. Eigenvector Research Inc. PLS Toolbox Documentation. T-Squared Q Residuals and Contributions. 2016. Eigenvector Research Inc.
- (103) Wold, S.; Sjostrom, M.; Eriksson, L. PLS-Regression: A Basic Tool of Chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130.
- (104) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20*, 273–297.
- (105) Smola, A. J.; Scholkopf, B. A Tutorial on Support Vector Regression. *Stat. Comput.* **2004**, 199–222.
- (106) Ni, W.; Nørgaard, L.; Mørup, M. Non-Linear Calibration Models for near Infrared Spectroscopy. *Anal. Chim. Acta* **2014**, *813*, 1–14.
- (107) Guenther, N.; Schonlau, M. Support Vector Machines. *Stata J.* **2016**, *16* (4), 917–937.
- (108) Filgueiras, P. R. Tese: Regressão por Vetores de Suporte Aplicado na Determinação de Propriedades Físico-Químicas de Petróleos e Biocombustíveis. Universidade estadual de campinas, 2014.
- (109) Chang, C.; Lin, C. *LIBSVM: A Library for Support Vector Machines*; 2019.
- (110) Eigenvector. Documentation (Manual do software) - Plstoolbox. Suport Vector Machine.
- (111) Associação Brasileira de Normas Técnicas - NBR ISO 9000 - Sistema de Gestão da Qualidade - Fundamentos e Vocabulário; 2015.
- (112) B. Magnusson and U. Örnemark (eds.). *Eurachem Guide: The Fitness for Purpose of Analytical Methods – A Laboratory Guide to Method Validation and Related Topics*; 2014.
- (113) Kennard, A. R. W.; Stone, L. A. Computer Aided Design of Experiments. *Technometrics* **1969**, *11* (1), 137–148.
- (114) Olivieri, A. C. Analytical Figures of Merit: From Univariate to Multiway Calibration. *Chem. Rev.* **2014**, *114* (10), 5358–5378.
- (115) Olivieri, A. C.; Faber, N. M.; Ferré, J.; Boqué, R.; Kalivas, J. H.; Mark, H. Uncertainty Estimation and Figures of Merit for Multivariate Calibration: (IUPAC Technical Report). *Pure Appl. Chem.* **2006**, *78* (3), 633–661.
- (116) Lorber, A.; Fabert, K.; Kowalski, B. R. Net Analyte Signal Calculation in

- Multivariate Calibration. *Anal. Chem.* **1997**, *69* (8), 1620–1626.
- (117) Mazivila, S.; Santana, F. B.; Mitsutake, H.; Gontijo, L. C.; Santos, D. Q.; Neto, W. B. Discrimination of the Type of Biodiesel/diesel Blend (B5) Using Mid-Infrared Spectroscopy and PLS-DA. *Fuel* **2015**, *142*, 222–226.
- (118) Câmara, A. B. F.; de Carvalho, L. S.; de Morais, C. L. M.; de Lima, L. A. S.; de Araújo, H. O. M.; de Oliveira, F. M.; de Lima, K. M. G. MCR-ALS and PLS Coupled to NIR/MIR Spectroscopies for Quantification and Identification of Adulterant in Biodiesel-Diesel Blends. *Fuel* **2017**, *210* (Agosto), 497–506.
- (119) Correia, R. M.; Tosato, F.; Domingos, E.; Rodrigues, R. R. T.; Aquino, L. F. M.; Filgueiras, P. R.; Lacerda, V.; Romão, W. Portable near Infrared Spectroscopy Applied to Fuel Quality Control. *Talanta* **2018**, *176* (Julho 2017), 59–68.
- (120) Li, H.; Niu, S.; Lu, C.; Wang, Y. Comprehensive Investigation of the Thermal Degradation Characteristics of Biodiesel and Its Feedstock Oil through TGA-FTIR. *Energy and Fuels* **2015**, *29* (8), 5145–5153.
- (121) Ogbu, I. M.; Ajiwe, V. I. E. FTIR Studies of Thermal Stability of the Oils and Methyl Esters from Afzelia Africana and Hura Crepitans Seeds. *Renew. Energy* **2016**, *96*, 203–208.
- (122) Jain, S.; Sharma, M. P. Review of Different Test Methods for the Evaluation of Stability of Biodiesel. *Renew. Sustain. Energy Rev.* **2010**, *14* (7), 1937–1947.
- (123) Mohamed Shameer, P.; Ramesh, K. FTIR Assessment and Investigation of Synthetic Antioxidant on the Fuel Stability of Calophyllum Inophyllum Biodiesel. *Fuel* **2017**, *209* (Julho), 411–416.
- (124) Fernandes, V. J.; Araujo, A. de S.; Vinhado, F. da S.; Pivesso, P. R. Caracterização de Resíduo Sólido Formado em Biodiesel de Sebo Bovino. *Quim. Nova* **2012**, *35* (10), 1901–1906.
- (125) Velasco, J.; Andersen, M. L.; Skibsted, L. H. Evaluation of Oxidative Stability of Vegetable Oils by Monitoring the Tendency to Radical Formation. A Comparison of Electron Spin Resonance Spectroscopy with the Rancimat Method and Differential Scanning Calorimetry. *Food Chem.* **2004**, *85* (4), 623–632.
- (126) Lira, L. F. B.; Albuquerque, M. S.; Pacheco, J. G. A.; Fonseca, T. M.; Cavalcanti, E. H. de S.; Stragevitch, L.; Pimentel, M. F. Infrared Spectroscopy and Multivariate Calibration to Monitor Stability Quality Parameters of Biodiesel. *Microchem. J.* **2010**, *96* (1), 126–131.
- (127) Vieira, F. S.; Pasquini, C. Determination of the Oxidative Stability of Biodiesel Using near Infrared Emission Spectroscopy. *Fuel* **2014**, *117*, 1004–1009.
- (128) Vieira, F. S.; Pasquini, C. Near Infrared Emission Photometer for Measuring the Oxidative Stability of Edible Oils. *Anal. Chim. Acta* **2013**, *796*, 101–107.
- (129) *Committee European Normalisation. EN 14112 Fat and Oil Derivatives – Fatty Acid Methyl Esters (FAME) - Determination of Oxidation Stability (Accelerated Oxidation Test)*. 2003.
- (130) Alves, J. C. L.; Henriques, C. B.; Poppi, R. J. Determination of Diesel Quality Parameters Using Support Vector Regression and near Infrared Spectroscopy for

- an in-Line Blending Optimizer System. *Fuel* **2012**, *97*, 710–717.
- (131) Filgueiras, P. R.; Alves, J. C. L.; Poppi, R. J. Quantification of Animal Fat Biodiesel in Soybean Biodiesel and B20 Diesel Blends Using near Infrared Spectroscopy and Synergy Interval Support Vector Regression. *Talanta* **2014**, *119*, 582–589.
- (132) Lima, D. G.; Soares, V. C. D.; Ribeiro, E. B.; Carvalho, D. A.; Cardoso, É. C. V.; Rassi, F. C.; Mundim, K. C.; Rubim, J. C.; Suarez, P. A. Z. Diesel-like Fuel Obtained by Pyrolysis of Vegetable Oils. *J. Anal. Appl. Pyrolysis* **2004**, *71* (2), 987–996.
- (133) Pedley, J. F.; Bersnek, L. A.; O’Connell, M. G.; Solly, R. K. The Chemistry of Sediment Formation in Australian Middle Distillate Fuels. In *HMSO Londres*; 1990; pp 1100–1107.
- (134) Mogollon, N. G. S.; Ribeiro, F. A. de L.; Lopez, M. M.; Hantao, L. W.; Poppi, R. J.; Augusto, F. Quantitative Analysis of Biodiesel in Blends of Biodiesel and Conventional Diesel by Comprehensive Two-Dimensional Gas Chromatography and Multivariate Curve Resolution. *Anal. Chim. Acta* **2013**, *796*, 130–136.
- (135) Wang, F. C. Y. Comprehensive Three-Dimensional Gas Chromatography Mass Spectrometry Separation of Diesel. *J. Chromatogr. A* **2017**, *1489*, 126–133.
- (136) Heibati, M.; Stedmon, C. A.; Stenroth, K.; Rauch, S.; Toljander, J.; Säve-Söderbergh, M.; Murphy, K. R. Assessment of Drinking Water Quality at the Tap Using Fluorescence Spectroscopy. *Water Res.* **2017**, *125*, 1–10.
- (137) Alostaz, M.; Biggar, K.; Donahue, R.; Hall, G. Soil Type Effects on Petroleum Contamination Characterization Using Ultraviolet Induced Fluorescence Excitation-Emission Matrices (EEMs) and Parallel Factor Analysis (PARAFAC). *J. Environ. Eng. Sci.* **2008**, *7* (6), 661–675.
- (138) Divya, O.; Mishra, A. K. Multivariate Methods on the Excitation Emission Matrix Fluorescence Spectroscopic Data of Diesel-Kerosene Mixtures: A Comparative Study. *Anal. Chim. Acta* **2007**, *592* (1), 82–90.
- (139) Lenhardt, L.; Zeković, I.; Dramićanin, T.; Milićević, B.; Burojević, J.; Dramićanin, M. D. Characterization of Cereal Flours by Fluorescence Spectroscopy Coupled with PARAFAC. *Food Chem.* **2017**, *229*, 165–171.
- (140) Toraman, H. E.; Abrahamsson, V.; Vanholme, R.; Van Acker, R.; Ronsse, F.; Pilate, G.; Boerjan, W.; Van Geem, K. M.; Marin, G. B. Application of Py-GC/MS Coupled with PARAFAC2 and PLS-DA to Study Fast Pyrolysis of Genetically Engineered Poplars. *J. Anal. Appl. Pyrolysis* **2018**, *129*, 101–111.
- (141) Morales, R.; Ortiz, M. C.; Sarabia, L. A. Usefulness of a PARAFAC Decomposition in the Fiber Selection Procedure to Determine Chlorophenols by Means SPME-GC-MS. *Anal. Bioanal. Chem.* **2012**, *403* (4), 1095–1107.
- (142) Fortunato de Carvalho Rocha, W.; Schantz, M. M.; Sheen, D. A.; Chu, P. M.; Lippa, K. A. Unsupervised Classification of Petroleum Certified Reference Materials and Other Fuels by Chemometric Analysis of Gas Chromatography-Mass Spectrometry Data. *Fuel* **2017**, *197*, 248–258.
- (143) Bernal, F. A.; Orduz-díaz, L. L.; Coy-barrera, E. Application of PARAFAC and

- OPLS-DA Analyses on HPLC Fingerprints for the Characterization of Hibiscus Sabdariffa Calyxes. *Quim. Nova* **2016**, 39 (2), 160–166.
- (144) Ruckebusch, C.; Blanchet, L. Multivariate Curve Resolution: A Review of Advanced and Tailored Applications and Challenges. *Anal. Chim. Acta* **2013**, 765, 28–36.
- (145) Smith, J. P.; Smith, F. C.; Booksh, K. S. Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS) with Raman Imaging Applied to Lunar Meteorites. *Appl. Spectrosc.* **2017**, 72 (3), 404–419.
- (146) Kwok, K.; Taylor, L. S. Analysis of Counterfeit Cialis® Tablets Using Raman Microscopy and Multivariate Curve Resolution. *J. Pharm. Biomed. Anal.* **2012**, 66, 126–135.
- (147) Marques, A. S.; Bedia, C.; Lima, K. M. G.; Tauler, R. Assessment of the Effects of As ( III ) Treatment on Cyanobacteria Lipidomic Profiles by LC-MS and MCR-ALS. *Anal. Bioanal. Chem.* **2016**, 408 (21), 5829–5841.
- (148) *Brasil. Ministério Do Meio Ambiente. Proconve - Matriz do Transporte de Cargas no Brasil. Brasília.* 2010.
- (149) *Associação Brasileira de Normas Técnicas. NBR ISO/IEC 17025 - Requisitos Gerais Para Competência de Laboratórios de Ensaio e Calibração.* 2017.
- (150) ASTM - International. *D7467 - Standard Specification for Diesel Fuel Oil , Biodiesel Blend (B6 to B20)*; 2017; pp 1–18.
- (151) Yaakob, Z.; Narayanan, B. N.; Padikkaparambil, S.; Unni K., S.; Akbar P., M. A Review on the Oxidation Stability of Biodiesel. *Renew. Sustain. Energy Rev.* **2014**, 35, 136–153.
- (152) European Committee for Standardization. *EN 15751 - Automotive Fuels - Fatty Acid Methyl Ester (FAME) Fuel and Blends with Diesel Fuel - Determination of Oxidation Stability by Accelerated Oxidation Method*; 2014.
- (153) Triola, M. F. *Introdução à Estatística*, 9a edição.; LTC: Rio de Janeiro, 2005.
- (154) Skrobot, V. L.; Santos, C. de S.; Braga, J. W. B.. Exploratory Analysis of Automotive Diesel Fuel Stability Test Methods by Infrared Spectroscopy and Parallel Factor Analysis. *Energy and Fuels* **2019**, 33 (7), 6170–6176.
- (155) Consortium, M. Massbank <https://massbank.eu/MassBank/Index>.
- (156) *Agência Nacional do Petróleo, Gás Natural e Biocombustíveis - Resolução ANP N° 50 DE 23.12.2013*; Brasil, 2013; pp 1–10.
- (157) Ammawath, W.; Man, Y. B. C.; Baharin, B. S.; Rahman, R. B. A. A New Method for Determination of Tert-Butylhydroquinone (TBHQ) in RBD Palm Olein with FTIR Spectroscopy. *J. Food Lipids* **2004**, 11 (4), 266–277.

## APÊNDICE 1 – DIAGNÓSTICO DE CONSISTÊNCIA DE NÚCLEO

A seguir, será descrito o conceito de CORCONDIA (do inglês, Core Consistency Diagnostic). Para tanto, será necessário recorrer a um outro tipo de modelo multidimensional - que não foi aplicado neste estudo. Trata-se do modelo Tucker, mais especificamente o modelo Tucker3<sup>73</sup>. O modelo Tucker3 é definido abaixo:

$$\underline{\mathbf{X}} = \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^T + \underline{\mathbf{E}} \quad (30)$$

onde as matrizes  $\mathbf{A}$  ( $I \times D$ ),  $\mathbf{B}$  ( $J \times E$ ) e  $\mathbf{C}$  ( $K \times F$ ) contêm os pesos relativos às três dimensões; a matriz  $\mathbf{G}$  ( $D \times EF$ ) é a matriz central (ou núcleo) e  $\underline{\mathbf{E}}$  contém os erros do modelo. Ao contrário do que ocorre com os modelos PARAFAC, é possível ter um número diferente de componentes em cada dimensão. Os elementos da matriz  $\mathbf{G}$  definem como os pesos individuais interagem nas diferentes dimensões, ou seja, os seus valores absolutos indicam a importância de cada interação entre componentes para o modelo.

Conforme descrito anteriormente, o modelo PARAFAC pode ser representado pela equação 2 (reproduzida abaixo),

$$\mathbf{X}_{(I \times J \times K)} = \mathbf{A}(\mathbf{C} \odot \mathbf{B})^T + \mathbf{E} \quad (2)$$

que é equivalente ao modelo Tucker3 a seguir<sup>78</sup> (Equação 7):

$$\mathbf{X}_{(I \times J \times K)} = \mathbf{A}\mathbf{T}^{(F \times FF)}(\mathbf{C} \otimes \mathbf{B})^T + \mathbf{E} \quad (31)$$

No qual a matriz  $\mathbf{T}$  (que corresponde a  $\underline{\mathbf{T}}$  transformada em matriz na direção apropriada) contém zeros em todas as coordenadas, exceto na superdiagonal que contém apenas valores iguais a um.

Após ajustado o modelo PARAFAC ( $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$ ), a verificação de que uma estrutura trilinear é apropriada<sup>78</sup> pode ser obtida calculando-se a matriz central de um modelo Tucker3,  $\underline{\mathbf{G}}$ , obtido por mínimos quadrados, dadas as *mesmas* matrizes  $\mathbf{A}$ ,  $\mathbf{B}$  e  $\mathbf{C}$  conforme abaixo<sup>78</sup>:

$$\underset{\mathbf{G}}{\operatorname{argmin}} \|\mathbf{X} - \mathbf{A}\mathbf{G}(\mathbf{C} \otimes \mathbf{B})^T\|_F^2 \quad (32)$$

Se o modelo PARAFAC é válido,  $\underline{\mathbf{G}}$  deve ser semelhante a  $\underline{\mathbf{T}}$ . Se o modelo não é trilinear, ou se muitas componentes foram inseridas, então  $\underline{\mathbf{G}}$  será (significativamente) diferente de  $\underline{\mathbf{T}}$ .

Uma forma simples de avaliar se a estrutura do modelo é adequada consiste em analisar os elementos da superdiagonal (e aqueles fora da superdiagonal) de  $\underline{\mathbf{G}}$ . Se os

elementos da superdiagonal são próximos de um e os elementos fora dela são próximos de zero, pode-se afirmar que não há sobreajuste do modelo. Caso contrário, ou foram utilizadas muitas componentes no modelo ou ele não é válido (questão da trilinearidade). É possível, também, que existam outliers no conjunto de dados<sup>78</sup>. A Equação 32 permite calcular a Superidentidade ou *Core Consistency* ou, ainda, CORCONDIA de “*CORe CONsistency DIAgnostic*” (o último termo foi selecionado neste trabalho):

$$CORCONDIA = 100 \left( \frac{1 - \sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F (g_{def} - t_{def})^2}{\sum_{d=1}^F \sum_{e=1}^F \sum_{f=1}^F t_{def}^2} \right) \quad (33)$$

**APÊNDICE 2 – COMPARAÇÃO DOS ENSAIOS PARA AVALIAÇÃO DE ESTABILIDADE DE ÓLEO DIESEL UTILIZADOS  
NESTE ESTUDO**

<b>Norma Técnica</b>	<b>Escopo</b>	<b>Condições</b>	<b>Parâmetro de medição</b>	<b>Volume de combustível por ensaio / mL</b>
ASTM D4625 -Standard Test Method for Middle Distillate Fuel Storage Stability at 43°C	Estabilidade inerente ao armazenamento de combustíveis destilados	400mL – 43°C – 4 a 24 semanas	Total insolúvel: diesel é filtrado (filtro de nylon 47 mm 0,8um). Lava-se filtro e garrafa com isooctano e pesa-se filtro antes e depois.	400
ASTM D6468 - Standard Test Method for High Temperature Stability of Middle Distillate Fuels	Estabilidade relativa de destilados médios sob condição de envelhecimento de alta temperatura com exposição limitada ao ar	50 mL – 150°C – exposição ao ar – 90 ou 180min	Percentual de reflectância do filtro Whatman n. 1 de 47 mm lavado com isooctano.	50
ASTM D7545 - Standard Test Method for Oxidation Stability of Middle Distillate Fuels—Rapid Small Scale Oxidation Test (RSSOT)	Estabilidade de combustíveis destilados médios como óleo diesel e óleos de aquecimento com até 100% de biodiesel, sob condições de oxidação acelerada	5 mL – 700 kPa de O <sub>2</sub> – 140°C – (teste continua até queda de 10% de pressão)	tempo até 10%	5
CEN EN 15751 Automotive fuels - Blends of Fatty acid methyl ester (FAME) with diesel fuel – Determination of oxidation stability by rapidly accelerated oxidation method at 110 °C	Estabilidade à oxidação de combustíveis para motores diesel, por meio da medição do período de indução de combustíveis até 20 h a 110 °C.	7,5 g – borbulhamento de O <sub>2</sub> – 110°C – (teste até inflexão de curva de condutividade ou 20h)	tempo até inflexão.	7,5

### APÊNDICE 3 – AVALIAÇÃO PRÉVIA DAS AMOSTRAS DE ÓLEO DIESEL - SEÇÃO 1

Identificação	Cor (visual)	Aspecto	Ponto de fulgor/°C	Massa Específica/kg m <sup>-3</sup>	Destilação/°C			Teor de biodiesel/%v/v
					T 10%	T 50%	T 95%	
1	Amarelo	LIALMP	54	835,9	202,8	271,7	347,1	8,1
2	Amarelo	LIALMP	64	839,1	216,4	275,2	358,3	8,8
3	Amarelo	LIALMP	55	847,1	208,6	278,2	357,7	8,1
4	Amarelo	LIALMP	61	838,3	203,2	273,0	345,6	7,7
5	Amarelo	LIALMP	62	837,6	208,2	271,6	361,4	8,1
6	Amarelo	LIALMP	55	846,1	209,0	282,3	368,4	8,2
7	Amarelo	LIALMP	63	841,5	213,4	277,4	350,5	8,1
8	Amarelo	LIALMP	66	834,9	211,8	277,7	352,8	8,0
9	Amarelo	LIALMP	61	837,6	190,9	265,1	344,1	8,1
10	Amarelo	LIALMP	64	836,4	206,9	278,0	352,9	8,1
11	Amarelo	LIALMP	60	835,1	207,1	269,8	343,2	7,9
12	Amarelo	LIALMP	53	836,7	213,0	286,1	362,7	8,0
13	Amarelo	LIALMP	60	833,5	202,0	270,0	356,7	8,2
14	Amarelo	LIALMP	59	838,7	204,4	279,5	350,9	7,9
15	Amarelo	LIALMP	63	839,6	198,8	267,8	341,5	8,0
16	Amarelo	LIALMP	58	830,5	197,9	272,6	340,1	8,2
17	Amarelo	LIALMP	55	836,5	195,8	271,4	349,9	8,1
18	Amarelo	LIALMP	64	836,8	212,1	277,8	352,5	8,0

LIALMP = Límpido e Isento de Água Livre e Material Particulado.

**APÊNDICE 4 – ESPECTROS DE REFERÊNCIA DO BANCO DE DADOS  
MASSBANK®**

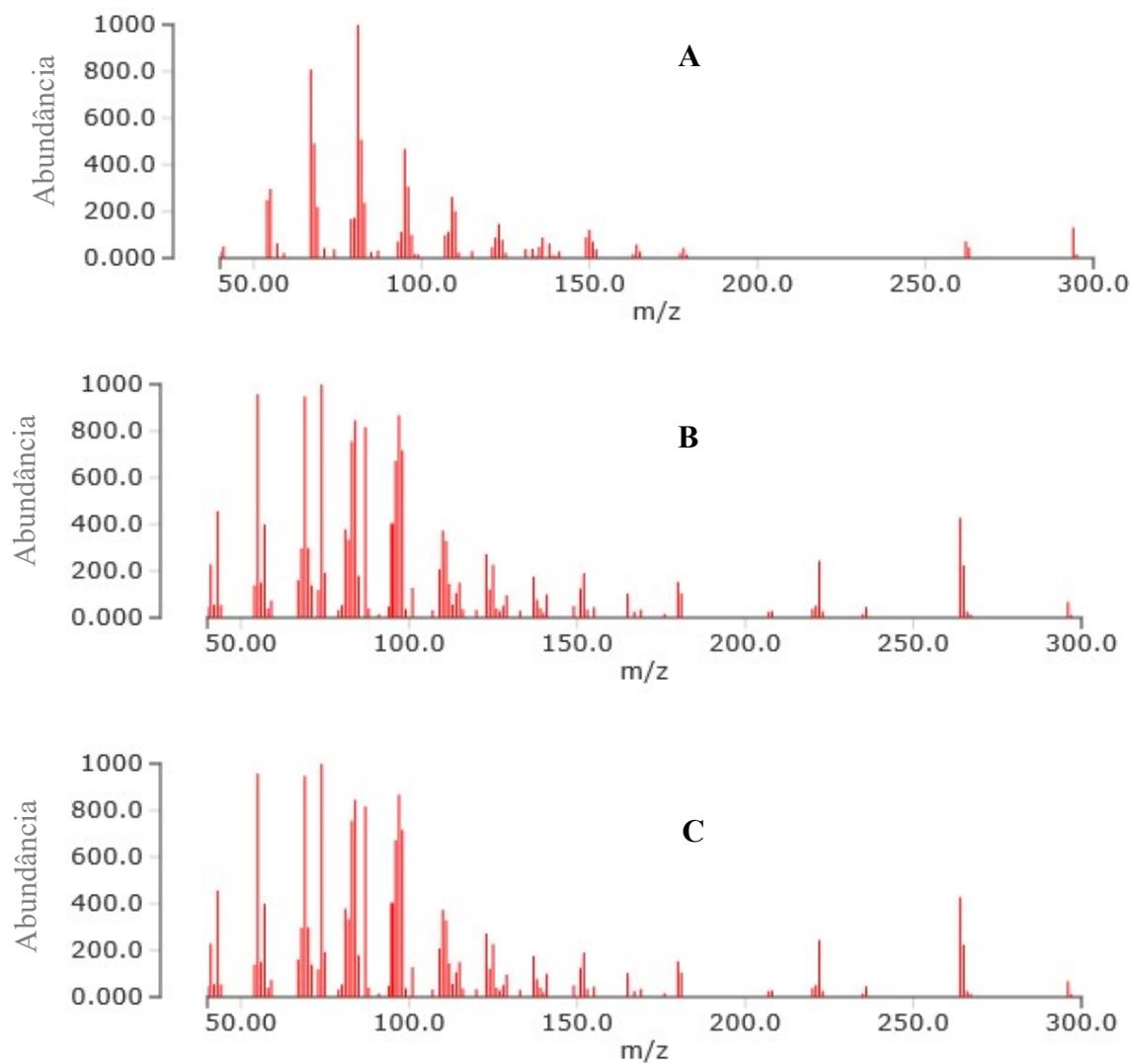


Figura 86: Espectros de massas de referência para **A**: Linoleato de metila, **B**: Oleato de Metila, **C**: Estearato de metila.

**APÊNDICE 5 – AVALIAÇÃO PRÉVIA DAS AMOSTRAS DE ÓLEO  
DIESEL - SEÇÃO 2**

Identificação	Estabilidade EN 15751 / h	Teor de biodiesel / %vol.	Teor de enxofre / mg/kg	Destilação / °C			Massa específica 20°C / kg/m³	Ponto de fulgor / °C	Aspecto *
				10% rec	50% rec	95% rec			
1	30,36	10,5	6,1	205	275,9	355,5	837,2	60	LII
2	21,45	10,8	5,6	201,3	275,4	347,4	839,3	51	LII
3	32,58	10,6	4,5	210,5	276,8	366,7	836,4	47	LII
4	16,63	10,8	5,7	206,4	276,6	353,4	837,9	54	LII
5	11,05	10,6	8,3	201,5	274,2	345,5	836,2	61	LII
6	33,47	9,8	7,7	197,5	268,1	340,4	833,9	56	LII
7	29,54	10,2	4,7	202,7	274,5	367,4	836,1	47	LII
8	22,39	10,3	5,3	209,8	281,9	329,1	839,3	55	LII
9	24,26	10,7	5,5	206,4	283,3	353,9	841,5	52	LII
10	23,53	10,7	5,4	203,8	276,7	343	841,5	52	LII
11	31,83	10,7	5,1	211,4	284,1	359,9	839,8	53	LII
12	33,53	10,6	5,3	210,4	277,3	356	836,7	47	LII
13	23,4	10,5	5,8	202,8	278,1	344,9	844	49	LII
14	27,67	9,9	5,9	190,2	272,4	349,1	833,3	60	LII
15	17,5	10,5	5,7	202,5	280	358,4	844,1	48	LII
16	25,01	10,4	5,6	199,7	280,8	352,3	844,5	48	LII
17	24,25	10,5	5,7	205,4	279,1	352,7	840	49	LII
18	27,39	10,8	5,5	205,5	277,1	355,9	841,6	51	LII
19	25,01	10,9	5,3	201,2	278,9	350	842,4	50	LII
20	21,31	10,8	6,1	206,8	280,7	355	842,8	46	LII
21	29,94	10,8	5,7	208,5	280,4	353	842,4	50	LII
22	21,38	10,8	5,7	204,9	281,5	353,7	843,5	49	LII
23	9,25	11,1	5,6	201,9	277,2	350,9	835,3	60	LII
24	28,15	10,5	6,5	211,7	275,8	347,8	838,5	52	LII
25	22,14	10,5	6,2	210,3	277,4	360,9	838,7	50	LII
26	20,22	10,8	4,1	195,2	276,9	346,6	838,2	45	LII
27	28,17	10,6	6,9	204,4	275,1	354,2	839,6	51	LII
28	24,82	10,8	4,1	196,5	276,8	345,4	840,7	42	LII
29	17,16	10,8	4	191,1	274,3	345,6	840,1	38	LII
30	28,77	10,7	5,9	209	277,4	352,2	839,9	50,5	LII
31	27,26	10,8	5,9	212,8	282,6	351,3	840,8	48	LII
32	22,58	10,8	3,9	197,4	277,3	349,2	841,8	41,5	LII

...Continuação Apêndice 5.

33	22,64	10,5	5,9	195,8	276,9	345,9	840,5	40,5	LII
34	38,04	-	5,9	208,1	276,9	356,1	839,8	47,5	LII
35	29,95	10,2	6,9	208,8	276,1	354,8	838	54	LII
36	20,66	10,6	6,2	205,7	277,7	355,8	837,1	58	LII
37	22,65	10,6	4,8	194,6	276,3	358,5	841,8	46	LII
38	34,29	10,6	5,7	206,2	279,6	365,4	839,9	48	LII
39	36,89	10,9	7,1	216,7	290,6	354,5	840	67	LII
40	9,82	10,3	-	213,2	275,8	351,2	842	41	LII
41	41,14	10,5	5,6	207,2	277,3	350,1	839,2	48	LII
42	19,03	10,4	6,1	210,1	277,5	368,2	840,2	53	LII
43	16,3	10,7	4,8	191,1	276,6	351,4	841,9	45	LII
44	33,74	10,5	7,3	214	282,3	347,1	840	62	LII
45	17,34	11,1	7,1	212,5	285,5	364,5	840,9	56	LII
46	27,38	10,9	5,6	200	279,1	353,4	839	53	LII
47	31,94	10,4	7,4	208,5	279,1	355,1	838	63	LII
48	33,09	10,8	6	206,3	282,1	356	840,6	45	LII
49	30,37	10,7	6,6	205,5	279,4	348,5	838,2	56	LII
50	38,13	10,7	6,6	202,8	280,7	354,8	841,2	39	LII

\* LII = Límpido e Isento de Impurezas