

Universidade Federal de Minas Gerais
Instituto de Ciências Exatas
Programa de Pós-Graduação em Química

Débora Ribeiro da Silva

**AVALIAÇÃO DE ESPECTRÔMETROS RAMAN E MODELOS DE
CLASSIFICAÇÃO MULTIVARIADA PARA MONITORAMENTO DA
ESTABILIDADE OXIDATIVA DO BIODIESEL**

Belo Horizonte

2025

UFMG/ICEX/DQ. 1.640

D. 893

Débora Ribeiro da Silva

**AVALIAÇÃO DE ESPECTRÔMETROS RAMAN E MODELOS DE
CLASSIFICAÇÃO MULTIVARIADA PARA MONITORAMENTO DA
ESTABILIDADE OXIDATIVA DO BIODIESEL**

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Mestre(a) em Química.

Orientador(a) Profa. Dra. Mariana Ramos de Almeida

Coorientador(a): Profa. Dra. Vânia Marcia Duarte Pasa

Belo Horizonte

2025

Ficha Catalográfica

S586a Silva, Débora Ribeiro da.
2025 Avaliação de espectrômetros Raman e modelos de classificação multivariada para
D monitoramento da estabilidade oxidativa do biodiesel [manuscrito] / Débora Ribeiro da
 Silva. 2025.

113 f. : il., gráfs., tabs.

Orientadora: Mariana Ramos de Almeida.
Coorientadora: Vânia Marcia Duarte Pasa.

Dissertação (mestrado) – Universidade Federal de Minas Gerais – Departamento de Química.

Bibliografia: f. 105-111.

Apêndices: f. 112-113.

1. Química analítica – Teses. 2. Quimiometria – Teses. 3. Raman, Espectroscopia de – Teses. 4. Biodiesel – Teses. 5. Oxidação – Teses. 6. Aprendizagem supervisionada (Aprendizado do computador) – Teses. I. Almeida, Mariana Ramos de, Orientadora. II. Pasa, Vânia Marcia Duarte, Coorientadora. III. Título.

CDU 043



UNIVERSIDADE FEDERAL DE MINAS GERAIS

UFMG

Programa de Pós-Graduação em Química
Departamento de Química - ICEx



"Avaliação de Espectrômetros Raman e Modelos de Classificação Multivariada Para Monitoramento da Estabilidade Oxidativa do Biodiesel"

Débora Ribeiro da Silva

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Profa. Mariana Ramos de
Almeida - Orientadora UFMG

Prof. Monica Benicia Mamian
Lopez Universidade Federal do
ABC

Prof. João Paulo
Ataide Martins
UFMG

Belo Horizonte, 14 de fevereiro de 2025.



Documento assinado eletronicamente por **Mariana Ramos de Almeida, Professora do Magistério Superior**, em 14/02/2025, às 16:28, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **João Paulo Ataíde Martins, Professor do Magistério Superior**, em 14/02/2025, às 16:29, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Mônica Benícia Mamián López, Usuário Externo**, em 14/02/2025, às 16:30, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3961967** e o código CRC **792D3760**.

Referência: Processo nº 23072.209410/2025-32

SEI nº 3961967

AGRADECIMENTOS

Primeiramente, agradeço a minha orientadora a Prof.^a Dr.^a Mariana Ramos de Almeida e ao colega Philipe Martins, pelo apoio e tempo dedicados a construção desse trabalho.

Ao Departamento de Química da UFMG e ao Laboratório de Ensaio de Combustíveis (LEC), por cederem o espaço, os equipamentos e os insumos que foram utilizados nas análises.

A minha orientadora, a Prof.^a Dr.^a Vânia Marcia Duarte Pasa, e a todos os membros do Programa de Recursos Humanos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – PRH-ANP, pelo apoio e parceria.

A todos os órgãos de fomento, incluindo a Fapemig, o CNPq e a CAPES, por apoiarem continuamente o desenvolvimento e a pesquisa na universidade.

Não poderia deixar de agradecer também a minha família e amigos, pelo apoio e estímulo nos momentos mais difíceis. Ao meu namorado pela paciência e carinho, principalmente nos meses dedicados a escrita dessa dissertação.

Aos meus colegas de trabalho, que não só me incentivaram a continuar, como também me deram todo o aporte necessário para que a conciliação entre trabalho e mestrado não fosse tão dispendiosa.

Por fim, gostaria de agradecer a Agilent® e a Luciana Assis Terra por disponibilizarem um dia para que pudéssemos testar o seu equipamento Raman portátil e fazer as medidas utilizadas nessa dissertação.

“Quanto mais aumenta nosso conhecimento, mais evidente fica nossa ignorância.”

John Fitzgerald Kennedy

RESUMO

As reações de oxidação do biodiesel resultam na formação de produtos indesejáveis que podem causar o entupimento e a deposição de sólidos nos motores a diesel. O controle de qualidade do biodiesel é regulamentado no Brasil pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP, de acordo com a Resolução nº 920/2023, que estabelece que a estabilidade oxidativa do biodiesel deve ser igual ou superior a 13 horas, sendo determinada por meio do método Rancimat com aproximadamente 12 horas de duração. Este trabalho propõe o uso da espectroscopia Raman associada a ferramentas quimiométricas para classificação rápida de amostras de biodiesel B100 em relação à sua estabilidade oxidativa. Na primeira parte do trabalho, dois espectrômetros Raman foram avaliados: um espectrômetro Raman portátil com tecnologia SORS e laser de excitação em 830 nm e um espectrômetro compacto de bancada com laser em 1064 nm. Os modelos de classificação multivariada (SIMCA, PLS-DA e floresta aleatória) construídos com os dados obtidos foram avaliados em termos de sensibilidade, especificidade, eficiência e coeficiente de Matthews. Os resultados demonstraram que ambos os espectrômetros são adequados para serem utilizados na análise da estabilidade oxidativa do biodiesel, sendo que o modelo de classificação empregando floresta aleatória apresentou melhor desempenho. Na segunda parte do trabalho, dois métodos de balanceamento de classes foram comparados, o ADASYN e o SMOTE. O método ADASYN apresentou melhor desempenho no balanceamento da classe minoritária, resultando em um modelo com maior eficiência. Com a proposta de um modelo hierárquico, as amostras classificadas como conformes foram novamente classificadas, alcançando eficiência de 94,7% na classificação das amostras conformes com estabilidade oxidativa maior e menor que 17 horas. O emprego da espectroscopia Raman associada a métodos de classificação multivariada mostrou-se adequada para a análise de amostras de biodiesel, reduzindo o tempo de análise em comparação ao Rancimat e viabilizando sua aplicação em diferentes etapas do processo.

Palavras-chave: SORS; balanceamento de classes; amostras sintéticas; PLS-DA; floresta aleatória.

ABSTRACT

Biodiesel oxidation reactions result in the formation of undesirable products that can cause clogging and solids deposition in diesel engines. Biodiesel quality control is regulated in Brazil by the National Agency of Petroleum, Natural Gas and Biofuels (ANP), according to Resolution No. 920/2023, which establishes that the oxidative stability of biodiesel must be equal to or greater than 13 hours, being determined by the Rancimat method with approximately 12 hours of duration. This work proposes the use of Raman spectroscopy associated with chemometric tools for rapid classification of B100 biodiesel samples in relation to their oxidative stability. In the first part of the work, two Raman spectrometers were evaluated: a portable Raman spectrometer with SORS technology and excitation laser at 830 nm and a compact benchtop spectrometer with laser at 1064 nm. The multivariate classification models (SIMCA, PLS-DA and random forest) constructed with the obtained data were evaluated in terms of sensitivity, specificity, efficiency and Matthews coefficient. The results demonstrated that both spectrometers are suitable for use in the analysis of the oxidative stability of biodiesel, with the classification model using random forest presenting better performance. In the second part of the work, two class balancing methods were compared, ADASYN and SMOTE. The ADASYN method presented better performance in balancing the minority class, resulting in a model with greater efficiency. With the proposal of a hierarchical model, the samples classified as compliant were classified again, reaching an efficiency of 94.7% in the classification of compliant samples with oxidative stability greater than and less than 17 hours. The use of Raman spectroscopy associated with multivariate classification methods proved to be suitable for the analysis of biodiesel samples, reducing the analysis time compared to Rancimat and enabling its application in different stages of the process.

Keywords: SORS; class balancing; synthetic samples; PLS-DA; random forest.

LISTA DE FIGURAS

Figura 1. Evolução da produção do biodiesel no Brasil dos anos de 2014 a 2023. (ANP, 2024).....	20
Figura 2. Reação de síntese do biodiesel. (adaptado de RINALDI, et al., 2007).....	22
Figura 3. Mecanismo de transesterificação alcalina em meio homogêneo. (adaptado de RAMOS, et al., 2011).....	23
Figura 4. Mecanismo de transesterificação ácida em meio homogêneo. (adaptado de RAMOS, et al., 2011).....	24
Figura 5. Mecanismo da interação entre o catalisador sólido básico e o álcool. (adaptado de RAMOS, et al., 2011).	25
Figura 6. Mecanismos: a. catalisador heterogêneo ácido de Brønsted-Lowry; b. catalisador heterogêneo ácido de Lewis. (adaptado de RAMOS, et al., 2011).....	25
Figura 7. Matéria-prima utilizada na produção do biodiesel no Brasil no ano de 2023 (Ministério de Minas e Energia, 2023).....	31
Figura 8. Mecanismo radicalar de oxidação do biodiesel (RAMALHO et al., 2006).	34
Figura 9. Estrutura molecular dos antioxidantes BHA, BHT e TBHQ.	35
Figura 10. Esquema da montagem do método Rancimat. (adaptado de MOURA, et al., 2016)	36
Figura 11. Esquema do princípio do espalhamento RAMAN. a. Espalhamento Stokes; b. Espalhamento Anti-Stokes. (adaptado de GONÇALVES, 2022)	38
Figura 12. Espectro Raman obtido para uma amostra de biodiesel.	39
Figura 13. Esquema dos componentes básicos de um espectrômetro Raman. (adaptado de SOUZA et al., 2003)	40
Figura 14. Configuração Raman, a. convencional, b. SORS. (AGILENT, 2024).....	43
Figura 15. Áreas de estudo da quimiometria.	44
Figura 16. Modelo de uma PCA representado por matrizes. (BRO, et al., 2014)	47
Figura 17. Interpretação geométrica de uma PCA. a. Sistema cartesiano com os eixos iguais as variáveis e componentes principais PC1 e PC2 traçadas segundo a direção de maior variabilidade dos dados; b. Representação do score como sendo a distância entre a projeção da amostra no eixo até a origem; c. Ângulo α entre a componente principal e o eixo da variável para determinação do peso.....	48
Figura 18. Limites das hipercaixas definidas para duas classes de amostras, descritas pelo modelo SIMCA (adaptado de FERREIRA, 2015).	51

Figura 19. Estrutura genérica de uma árvore de decisão (adaptado de SANTANA, 2020).	55
Figura 20. Análise das amostras no Agilent Resolve Raman® (Equipamento A).	66
Figura 21. Análise das amostras no Cora 5001 da Anton Paar® (Equipamento B).	66
Figura 22. Espectro Raman de uma amostra de biodiesel conforme (verde) e não conforme (vermelho).	69
Figura 23. Espectros Raman brutos obtidos pelo equipamento A.	70
Figura 24. Gráfico de Escores: PC1 versus PC2.	71
Figura 25. Gráfico de Pesos da PC1.	71
Figura 26. Gráfico de Pesos da PC2.	72
Figura 27. Gráfico de Escores: PC1 versus PC3.	73
Figura 28. Gráfico de Pesos da PC3.	74
Figura 29. Espectros Raman brutos obtidos pelo equipamento B.	75
Figura 30. Gráfico de Escores: PC1 versus PC2.	76
Figura 31. Gráfico de Pesos da PC1.	76
Figura 32. Gráfico de Pesos da PC2.	77
Figura 33. Gráfico de Escores: PC1 versus PC3.	78
Figura 34. Gráfico de Pesos da PC3.	78
Figura 35. Gráfico de T^2 de Hotelling versus Q Resíduos. a. Equipamento A e b. Equipamento B.	80
Figura 36. Gráfico de classificação das amostras segundo a classe modelada (conforme). a. Equipamento A e b. Equipamento B.	81
Figura 37. Espectros Raman brutos balanceados pelo método ADASYN. a. Equipamento A e b. Equipamento B.	83
Figura 38. PLS-DA do conjunto de amostras balanceado pelo ADASYN. a. Equipamento A e b. Equipamento B.	84
Figura 39. Contribuição das variáveis para a construção do modelo PLS-DA, calculado a partir dos dados obtidos pelo equipamento A.	85
Figura 40. Contribuição das variáveis para a construção do modelo PLS-DA, calculado a partir dos dados obtidos pelo equipamento B.	86
Figura 41. Matrizes de confusão dos modelos floresta aleatória construídas a partir dos conjuntos teste balanceados pelo ADASYN, a. Equipamento A e b. Equipamento B.	88

Figura 42. Contribuição normalizada das variáveis para a construção dos modelos floresta aleatória. a. Equipamento A e b. Equipamento B.	89
Figura 43. Análise Raman utilizando o Equipamento A, sobre recipientes de diferentes materiais (AGILENT®, 2024).....	91
Figura 44. Uso do Equipamento B dentro de um laboratório. (ANTON PAAR®, 2024).	92
Figura 45. Espectros Raman brutos obtidos na análise do novo conjunto de dados pelo equipamento B, balanceados pelo a. ADASYN e b. SMOTE.	93
Figura 46. Matrizes de confusão do modelo floresta aleatória construído a partir dos dados balanceados pelo ADASYN, a. conjunto treinamento e b. conjunto teste.....	94
Figura 47. Matrizes de confusão do modelo floresta aleatória construído a partir dos dados balanceados pelo SMOTE, a. conjunto treinamento e b. conjunto teste.	95
Figura 48. Contribuição normalizada das variáveis para a construção dos modelos floresta aleatória, a partir dos dados balanceados pelo a. ADASYN e b. SMOTE. ...	97
Figura 49. Matrizes de confusão do modelo hierárquico construído, a. conjunto treinamento e b. conjunto teste.	100
Figura 50. Contribuição normalizada das variáveis para a construção do modelo hierárquico.....	101

LISTA DE TABELAS

Tabela 1. Especificações do biodiesel definidas pela resolução ANP n° 920/2023. .33	
Tabela 2. Especificações técnicas dos equipamentos Raman utilizados.67	
Tabela 3. Resultados de sensibilidade, especificidade e eficiência obtidos para os conjuntos de treinamento e teste.82	
Tabela 4. Figuras de mérito dos modelos PLS-DA construídos.87	
Tabela 5. Parâmetros utilizados na construção dos modelos de floresta aleatória...87	
Tabela 6. Figuras de mérito dos modelos floresta aleatória construídos.90	
Tabela 7. Figuras de mérito dos conjuntos teste dos modelos floresta aleatória construídos.....96	
Tabela 8. Valores de acordância e concordância calculados para amostras de biodiesel classificadas pelos modelos floresta aleatória.98	

LISTA DE QUADROS

Quadro 1. Vantagens e desvantagens dos principais processos de síntese do biodiesel.....	28
Quadro 2. Equações usadas no cálculo das figuras de mérito utilizadas na validação dos modelos supervisionados.	60

LISTA DE ABREVIATURAS

ACO – Acordância

ADASYN - Abordagem Adaptativa de Sobreamostragem Sintética para Aprendizado Desbalanceado (*Adaptive Synthetic Sampling Approach for Imbalanced Learning*)

ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis

BHA - Butil-Hidroxianisol

BHT - Butil-Hidroxitolueno

CARS - Espectroscopia Raman Anti-Stokes Coerente (*Coherent Anti-stokes Raman Spectroscopy*)

CCD - Dispositivo de Carga Acoplada (*Charge-Coupled Device*)

CNPE - Conselho Nacional de Política Energética

CON - Concordância

CV - Validação Cruzada (*Cross Validation*)

DD-SIMCA – SIMCA orientado aos dados (*Data Driven SIMCA*)

EFC – Eficiência

EN – Norma Europeia (*European Standards*)

ESP - Especificidade

FAME - Ésteres metílicos de ácidos graxos (*Fatty Acid Methyl Esters*)

FN - Falso Negativo

FP - Falso Positivo

FTIR - Infravermelho com Transformada de Fourier (*Fourier-transform infrared spectroscopy*)

HCA- Agrupamentos por Métodos Hierárquicos (*Hierarchical Clusters Analysis*)

Kg - kilogramas

LDA - Análise Discriminante Linear (*Linear Discriminant Analysis*)

LEC - Laboratório de Ensaio de Combustíveis

MCC - Coeficiente de Correlação de Matthews (*Matthews Correlation Coefficient*)

MIR – Infravermelho Médio (*Mid-Infrared*)

mm – milímetros

MSC - Correção de espalhamento multiplicativo (*Multiplicative Scatter Correction*)

mW - miliWatts

NIR - Infravermelho Próximo (*Near-Infrared*)

nm – Nanômetros

OCPLS – Mínimos Quadrados Parciais de Uma Classe (*One Class Partial Least Squares*)

PC - Componente Principal (*Principal Component*)

PCA - Análise de Componentes Principais (*Principal Component Analysis*)

PLS - Mínimos Quadrados Parciais (*Partial Least Squares*)

PLS-DA - Análise Discriminante por Mínimos Quadrados Parciais (*Partial Least Square Discriminant Analysis*)

R² – Coeficiente de determinação

RMSEC - Raiz quadrada do Erro Médio Quadrático da Calibração (*Root Mean Square Errors of Calibration*)

RMSEP - Raiz quadrada do Erro Médio Quadrático da Previsão (*Root Mean Square Errors of Validation*)

RPD – Relação de Desempenho do Desvio (*Residual Prediction Deviation*)

SEN – Sensibilidade

SERS – Espectroscopia Raman Amplificada por Superfície (*Surface Enhanced Raman Spectroscopy*)

SIMCA – Modelagem Independente Suave por Analogia de Classes (*Soft Independent Modeling of Class Analogy*)

SMOTE – Técnica de Sobreamostragem Sintética para Minoria (*Synthetic Minority Over-sampling Technique*)

SNV - Variação normal padrão (*Standard Normal Variate*)

SORS - Espectroscopia Raman com Deslocamento Espacial (*Spatially Offset Raman Spectroscopy*)

TBHQ - Terc-Butil-Hidroquinona

TERS - Espectroscopia Raman Aprimorada por Ponta (*Tip-Enhanced Raman Spectroscopy*)

UHPLC-HRMS - Cromatografia Líquida de Ultra-Alta Performance acoplada à Espectrometria de Massas de Alta Resolução (*Ultra-High Performance Liquid Chromatography-High-Resolution Mass Spectrometry*)

VIP escores - Importância dos Vetores na Projeção dos escores (*Variable Importance on Projection*)

VLs - Variáveis Latentes

VN - Verdadeiro Negativo

VP - Verdadeiro Positivo

SUMÁRIO

1. INTRODUÇÃO	19
2. OBJETIVOS	21
2.1 Objetivos específicos	21
3. REVISÃO DA LITERATURA	22
3.1 Biodiesel.....	22
3.2 Legislação	31
3.3 Estabilidade oxidativa do biodiesel.....	33
3.4 Espectroscopia Raman	37
3.4.1 Conceitos fundamentais	37
3.4.2 Instrumentação.....	39
3.4.3 Espectroscopia Raman com deslocamento espacial (SORS).....	42
3.5 Quimiometria	43
3.5.1 Organização e visualização dos dados experimentais	44
3.5.2 Pré-processamento de dados espectrais	45
3.5.3 Análise de Componentes Principais (PCA)	46
3.5.4 Modelagem de dados – Métodos Supervisionados.....	49
3.5.5 Soft Independent Modeling of Class Analogy – SIMCA	50
3.5.6 Análise Discriminante por Mínimos Quadrados Parciais - <i>Partial Least Squares Discriminant Analysis</i> (PLS-DA).....	52
3.5.7 Árvore de Decisão.....	54
3.5.8 Floresta Aleatória – <i>Random Forest</i>	56
3.5.9 Métodos de balanceamento de classes	57
3.5.10 Validação de modelos de classificação multivariada.....	59
3.6 Estudos da estabilidade oxidativa do biodiesel empregando métodos espectroscópicos e ferramentas quimiométricas.....	61
4. METODOLOGIA	65

4.1 Amostras	65
4.2 Análise espectroscópica.....	65
4.3 Tratamento dos dados obtidos	67
5. RESULTADOS E DISCUSSÕES	69
5.1 Desempenho dos espectrômetros Raman na avaliação da estabilidade oxidativa do biodiesel B100.....	69
5.1.1 Análise exploratória do conjunto de dados obtido pelo Equipamento A..	69
5.1.2 Análise exploratória do conjunto de dados obtido pelo Equipamento B..	74
5.1.3 Avaliação dos Equipamentos A e B por meio do método de classificação SIMCA	79
5.1.4 Avaliação dos Equipamentos A e B por meio do método PLS-DA.....	82
5.1.5 Avaliação dos Equipamentos A e B por meio do método de floresta aleatória	87
5.1.6 Comparação entre os equipamentos A e B.....	90
5.2 Construção de um modelo de classificação hierárquico para avaliação da estabilidade oxidativa do biodiesel com geração de amostras sintéticas	92
REFERÊNCIAS.....	105
APÊNDICE A – Espectros Raman complementares com destaque da banda em 1100 cm⁻¹	112
APÊNDICE B – Espectro Raman complementar com destaque da banda em 800 cm⁻¹	113
APÊNDICE C – Matrizes de confusão dos modelos floresta aleatória dos conjuntos de treinamento.....	113

1. INTRODUÇÃO

Atualmente as principais fontes de energia mundial provêm dos combustíveis fósseis derivados do petróleo, como a gasolina, o óleo diesel e o gás natural. Os combustíveis fósseis são formados ao longo de milhares de anos por meio da decomposição da matéria orgânica. A combustão desses materiais é altamente energética e supre a demanda requerida por indústrias, veículos e residências. Entretanto, o processo de queima dos combustíveis fósseis gera a emissão de gases poluentes causadores do efeito estufa, como o dióxido de carbono (CO₂), além de se tratar de fontes de energia escassas. Os esforços das últimas décadas foram destinados a viabilizar a transição energética, substituindo os combustíveis fósseis por fontes sustentáveis com baixa emissão de carbono.

Hoje em dia o biodiesel, o etanol e o óleo vegetal hidrotratado são os biocombustíveis de maior interesse comercial em nível mundial. No Brasil, o biodiesel puro (B100) é misturado com o diesel para formar as misturas BX, onde X corresponde a porcentagem de biodiesel no diesel. Dados disponibilizados no Anuário Estatístico Brasileiro do Petróleo, Gás Natural e Biocombustíveis (ANP, 2024) mostraram que no período de 2014 a 2023 a produção do biodiesel no Brasil aumentou cerca de 120%, sendo gerados 7,53 milhões de m³ de B100 em 2023 (Figura 1). O óleo de soja foi a principal matéria-prima utilizada na produção do biodiesel nesse ano, correspondendo a 69% da produção total, seguido pelo óleo de fritura e materiais graxos (18,1%), pela gordura animal (8,5%) e pelos óleos vegetais de algodão, canola, girassol, macaúba, milho, palma e palmiste (4,4%).

A qualidade do biodiesel produzido interfere diretamente na aplicação das misturas BX em motores a diesel. Alguns parâmetros importantes, como a estabilidade oxidativa e a contaminação total, podem interferir no funcionamento e desempenho desses motores. Nesse contexto, o estudo e a avaliação da qualidade do biodiesel são muito importantes, evidenciando a necessidade do desenvolvimento de metodologias rápidas e eficientes para esse propósito (CORONADO *et al.*, 2009).

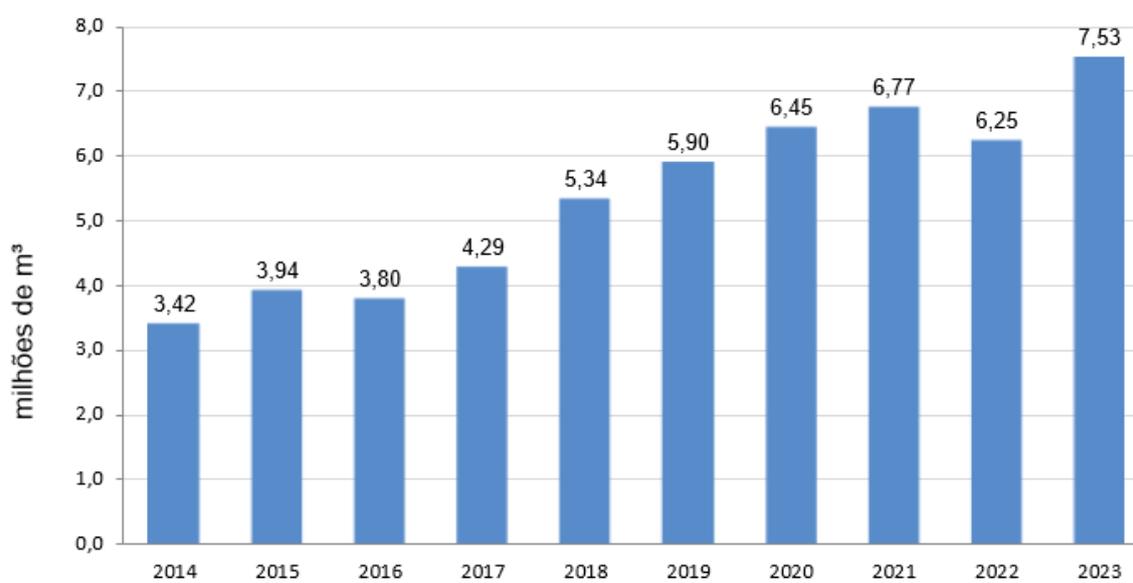


Figura 1. Evolução da produção do biodiesel no Brasil dos anos de 2014 a 2023. (ANP, 2024).

2. OBJETIVOS

Este trabalho tem como objetivo avaliar o uso da espectroscopia Raman associada a métodos quimiométricos para avaliar a estabilidade oxidativa de amostras de biodiesel B100.

2.1 Objetivos específicos

- Obter os espectros Raman de amostras de biodiesel classificadas previamente como “conforme” ou “não conforme”, utilizando espectrômetros Raman portátil e compacto;
- Caracterizar os espectros Raman das amostras de biodiesel;
- Avaliar e comparar o desempenho dos espectrômetros Raman;
- Fazer a análise exploratória do conjunto de dados usando a PCA;
- Fazer o balanceamento das classes de amostras empregando os métodos de reamostragem: *Adaptive Synthetic Sampling Approach for Imbalanced Learning* (ADASYN) e *Synthetic Minority Over-sampling Technique* (SMOTE);
- Construir modelos de classificação supervisionados usando os métodos PLS-DA, floresta aleatória e SIMCA;
- Avaliar o desempenho dos modelos construídos utilizando as figuras de mérito apropriadas;
- Construir um modelo hierárquico para avaliação de amostras de biodiesel com alta e média estabilidade oxidativa.

3. REVISÃO DA LITERATURA

3.1 Biodiesel

O biodiesel é um biocombustível de origem renovável que pode ser produzido a partir de gorduras animais e vegetais. Ele pode ser sintetizado através da reação de transesterificação dos triglicerídeos presentes na gordura com um mono álcool de cadeia curta, na presença de um catalisador homogêneo, heterogêneo ou enzimático. A reação de transesterificação resulta em uma mistura de ésteres alquílicos de ácidos graxos mais glicerol, conforme mostrado na Figura 2 (RINALDI, *et al.*, 2007).



Figura 2. Reação de síntese do biodiesel. (adaptado de RINALDI, *et al.*, 2007).

As reações de transesterificação catalisadas em meio alcalino homogêneo costumam ser mais utilizadas na indústria, já que podem ser realizadas com menor tempo e em temperatura ambiente. Os alcóxidos metálicos e as bases inorgânicas são os catalisadores básicos mais usuais, podendo ser adicionados diretamente no meio reacional ou ser dissolvidos no álcool. Os íons hidróxidos reagem com o mono álcool gerando um alcóxido nucleofílico e água, diminuindo o rendimento da reação de alcoólise. Por sua vez, o agente nucleofílico ataca o carbono da carbonila do triglicerídeo, formando um intermediário tetraédrico que elimina um éster metílico e forma outro alcóxido. O alcóxido formado na etapa anterior reage com a água e forma um diglicerídeo, conforme mostrado no mecanismo da Figura 3. A repetição desse processo por mais dois ciclos produz uma molécula de glicerol e uma mistura de três moléculas de éster metílico, dando origem ao biodiesel. A reação de transesterificação é realizada com uma proporção de 3:1, três mols de álcool para 1 mol de triglicerídeo. Entretanto, o excesso de álcool pode deslocar o equilíbrio químico e aumentar a produção de ésteres graxos. Os ésteres graxos reagem com a base e formam

tensoativos que suprimem a reação de alcoólise, tornando o meio alcalino sensível a presença de tais componentes (RAMOS, *et al.*, 2011).

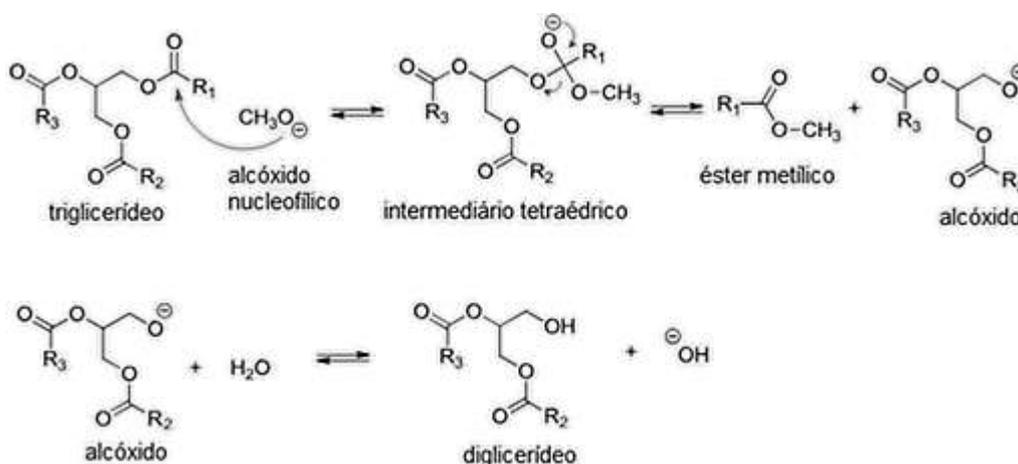


Figura 3. Mecanismo de transesterificação alcalina em meio homogêneo. (adaptado de RAMOS, *et al.*, 2011).

A reação de transesterificação realizada em meio ácido homogêneo ocorre em maior tempo e sob temperaturas mais elevadas, o que aumentam os custos do processo. Os catalisadores ácidos mais comuns são os ácidos de Brønsted-Lowry, como os ácidos sulfúrico e sulfônico. Nessa reação, a carbonila do triglicerídeo é protonada com um íon H⁺ provindo do ácido, dando origem a um carbocátion. O carbocátion sofre um ataque nucleofílico do álcool formando um intermediário tetraédrico. A molécula tetraédrica sofre rearranjo com transferência de próton, formando o diglicerídeo e uma molécula de éster graxo. A repetição desse processo por mais dois ciclos leva a produção de mais duas moléculas de ésteres graxos e uma molécula de glicerol, conforme mostrado na Figura 4. Na reação de transesterificação ácida o controle da produção de ácidos graxos não é necessário, já que eles podem ser esterificados no mesmo meio reacional. Essa característica é vantajosa frente ao processo alcalino, possibilitando o uso de gorduras com menor valor agregado (RAMOS, *et al.*, 2011).

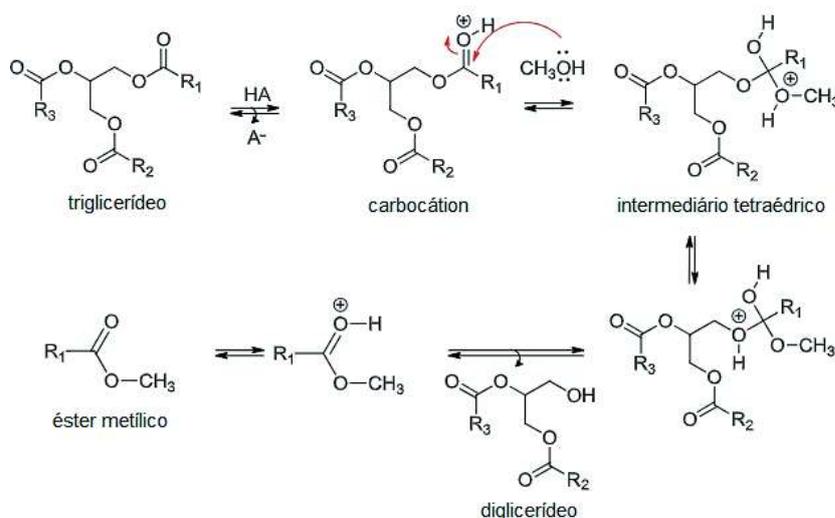


Figura 4. Mecanismo de transesterificação ácida em meio homogêneo. (adaptado de RAMOS, *et al.*, 2011).

O uso de catalisadores homogêneos adiciona etapas de purificação no processo de produção do biodiesel, além de possibilitar a formação de misturas instáveis. Por esse motivo, os catalisadores heterogêneos também podem ser utilizados na síntese do biodiesel, a fim de reduzir as etapas de purificação e garantir o reuso do catalisador. Os catalisadores heterogêneos podem ser classificados como ácidos ou básicos, dependendo da característica do seu sítio ativo. As zeólitas, os óxidos de metais alcalinos terrosos e os aluminossilicatos são alguns dos catalisadores heterogêneos mais utilizados.

Os catalisadores heterogêneos básicos atuam como uma base de Lewis, doando um par de elétrons e se ligando ao hidrogênio da hidroxila do álcool, conforme mostrado no mecanismo da Figura 5. Essa interação favorece o ataque nucleofílico do álcool sobre a carbonila do triglicerídeo, uma vez que ocorre um aumento na densidade eletrônica do oxigênio. A reação segue através do mesmo mecanismo da reação de transesterificação alcalina em meio homogêneo, com a ressalva de que o catalisador não se mistura com o biodiesel formado. As guanidinas são exemplos de catalisadores heterogêneos básicos de Lewis.

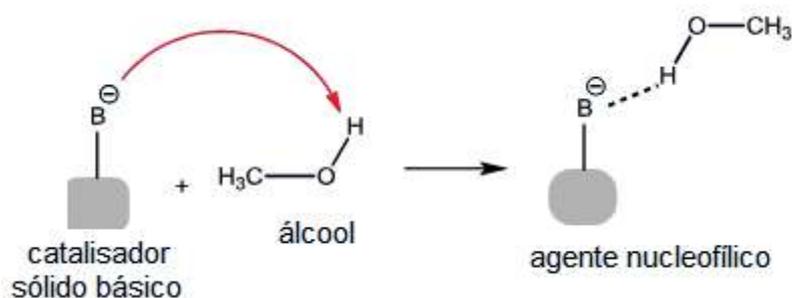


Figura 5. Mecanismo da interação entre o catalisador sólido básico e o álcool. (adaptado de RAMOS, *et al.*, 2011).

Os catalisadores heterogêneos ácidos são derivados de substâncias do tipo ácidos de Brønsted-Lowry ou ácidos de Lewis. Os ácidos de Brønsted-Lowry doam um próton para a carbonila do triglicerídeo, levando a formação do carbocátion (Figura 6.a) e, posteriormente, seguindo as mesmas etapas que o mecanismo de transesterificação ácida em meio homogêneo para formação do biodiesel. Dentre os grupos de catalisadores heterogêneos ácidos de Brønsted-Lowry pode-se citar as resinas trocadoras de íons e as zeólitas. Já os catalisadores ácidos de Lewis atuam através da interação ácido-base entre o par de elétrons do oxigênio da carbonila do triglicerídeo e o metal do catalisador (Figura 6.b). A interação ácido-base aumenta a carga positiva no carbono da carbonila, potencializando o ataque nucleofílico do álcool. Os carboxilatos de zinco e o óxido de nióbio são exemplos de catalisadores heterogêneos ácidos de Lewis.

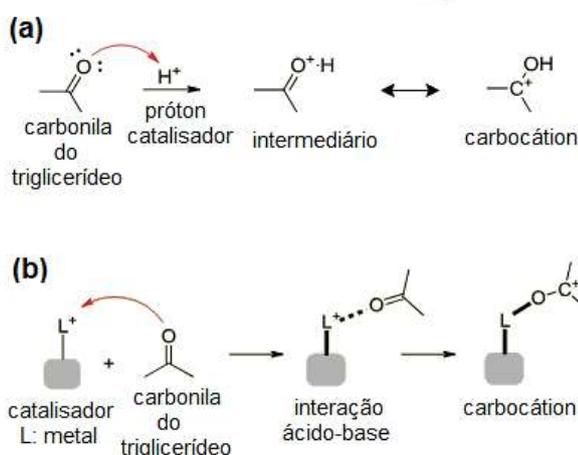


Figura 6. Mecanismos: a. catalisador heterogêneo ácido de Brønsted-Lowry; b. catalisador heterogêneo ácido de Lewis. (adaptado de RAMOS, *et al.*, 2011).

Os catalisadores enzimáticos estão sendo introduzidos na indústria do biodiesel como uma alternativa aos catalisadores químicos. A especificidade, a regioseletividade e a enantioseletividade das enzimas catalisadoras permitem que seu uso gere menor quantidade de subprodutos durante a síntese do biodiesel. Além disso, as enzimas atuam em ampla faixa de pH, variando de 4 a 9, além de serem ativas em condições brandas de temperatura e pressão. As lipases (triacilglicerol hidrolase) são as enzimas empregadas nos processos de alcoólise de gorduras. Elas atuam no meio reacional quebrando as moléculas de gordura e liberando ácidos graxos livres e gliceróis e podem ser obtidas através de fontes animais, vegetais ou microbianas. As lipases obtidas através de fonte animal possuem alto valor agregado, associado as etapas adicionais de isolamento e purificação. As lipases microbianas podem ser obtidas através de processos de fermentação em meio aquoso ou sólido. Já as lipases vegetais são as mais utilizadas na indústria, uma vez que seus processos de extração e purificação são mais econômicos. Apesar das lipases poderem ser utilizadas em sua forma livre, o modo mais comum do seu uso em reações de transesterificação é através de métodos de imobilização. A imobilização melhora a estabilidade e a recuperação das enzimas, além de evitar sua inativação por interação com solventes. Vários métodos são utilizados para a imobilização enzimática, como a adsorção, a ligação covalente ou o encapsulamento. A adsorção é o método mais comum de imobilização aplicado nas lipases. Nesse método, a enzima é adsorvida sobre a superfície de um material suporte, que pode ser composto por vidro de porosidade controlada, sílica, poliamida ou resina. A imobilização por adsorção permite que a lipase possua maior atividade catalítica, já que o procedimento envolve interações fracas do tipo dipolo-dipolo. Entretanto, a adsorção pode acarretar na perda de material catalítico, que pode ficar retido no suporte por meio de interações irreversíveis (BABADI, *et al.*, 2022).

Além da transesterificação catalítica em meio homogêneo ou heterogêneo, existem outras reações menos usuais que também podem ser utilizadas na síntese do biodiesel. Reações de transesterificação assistidas magneticamente, por ultrassom ou por microondas também são citadas na literatura (REIS, *et al.*, 2015). Tratam-se de métodos com alto rendimento, porém que adicionam etapas e insumos de alto valor ao procedimento. Existem dois métodos alternativos de transesterificação que não

fazem uso de catalisadores, são eles a transesterificação em fluido supercrítico e a transesterificação com uso de cossolventes (PAZ *et al.*, 2023).

Na transesterificação em fluido supercrítico o álcool e o triglicerídeo formam uma mistura monofásica no estado supercrítico, sob condições de alta temperatura e pressão. As condições do meio reacional modificam as propriedades termofísicas da mistura, como a viscosidade, a constante dielétrica, o produto iônico e a polaridade. A diminuição da constante dielétrica e o aumento da viscosidade fazem com que o álcool e a gordura formem uma mistura homogênea com maior cinética de reação. Além disso, a alteração do produto iônico permite que o álcool atue como uma espécie de catalisador ácido, favorecendo a síntese do biodiesel. A síntese do biodiesel por fluido supercrítico é vantajosa frente as reações catalíticas, principalmente, por que não sofre influência pela presença de ésteres graxos ou de água no meio reacional. Entretanto, a técnica exige condições extremas de temperatura (250-450 °C) e pressão (350-600 atm) que a deixam mais cara (PELISSON, 2013).

A síntese do biodiesel com o uso de co-solventes foi desenvolvida pelo professor David Boocock na Universidade de Toronto e foi patenteada pela BIOX®. Na transesterificação BIOX® um co-solvente inerte é adicionado no processo de síntese do biodiesel, a fim de formar um sistema monofásico sob condições normais de temperatura e pressão. Nessa reação o ácido graxo livre sofre esterificação ácida, seguida pela etapa de transesterificação do triglicerídeo com o uso do co-solvente. O cossolvente utilizado possui ponto de ebulição semelhante ao do álcool e tem a função de solubilizar o álcool na fase lipídica. Ele pode ser recuperado e reutilizado continuamente ao longo de todo o processo, sem gerar resíduos no biodiesel ou no glicerol formado. O tetrahidrofurano e o éter metil terc-butílico foram os cossolventes testados para essa reação. Entretanto, a toxicidade e a dificuldade em separar esses solventes do álcool são características limitantes desse procedimento (MA, F., *et al.*, 1999).

O Quadro 1 apresenta um comparativo com as principais vantagens e desvantagens dos processos de síntese do biodiesel que foram citados acima. As condições reacionais de cada processo, como temperatura, pressão, qualidade dos reagentes e uso de catalisadores, interferem diretamente nas propriedades do produto final. O uso de temperaturas elevadas, como ocorre na transesterificação ácida em meio homogêneo e na transesterificação em fluido supercrítico, pode acarretar na

formação de compostos oxidantes secundários, como aldeídos e cetonas. O tipo e a concentração do catalisador utilizado nas reações de transesterificação catalítica também podem influenciar na estabilidade oxidativa do biodiesel produzido, já que os íons metálicos atuam como catalisadores da reação de oxidação. Outros fatores que podem influenciar na estabilidade oxidativa do biodiesel durante a síntese são: exposição ao ar e a luz, material componente do tanque reacional, contaminação cruzada, dentre outros.

Quadro 1. Vantagens e desvantagens dos principais processos de síntese do biodiesel.

Tipo	Processo	Vantagens	Desvantagens
Transesterificação catalítica	Alcalina em meio homogêneo	<ul style="list-style-type: none"> - Temperatura e pressão ambiente; - Rendimento acima de 97%; - Rápido (20-120 min). 	<ul style="list-style-type: none"> - Sensível a presença de água e ésteres graxos; - Alto consumo de energia.
	Ácida em meio homogêneo	<ul style="list-style-type: none"> - Não é sensível a presença de ésteres graxos; - Econômico. 	<ul style="list-style-type: none"> - Sensível a presença de água; - Maior tempo de reação; - Altas temperaturas; - Dificuldade na recuperação do catalisador; - Maior razão molar triglicerídeo:álcool.
	Alcalina em meio heterogêneo	<ul style="list-style-type: none"> - Pouca quantidade de resíduos; - Recuperação e reuso do catalisador; - Baixa temperatura; - Econômico. 	<ul style="list-style-type: none"> - Dificuldade na recuperação do glicerol; - Alto consumo de energia.
			<ul style="list-style-type: none"> - Baixa quantidade de catalisador;

	Ácida em meio heterogêneo	<ul style="list-style-type: none"> - Recuperação e reuso do catalisador; - Não conta com processos adicionais de lavagem. 	<ul style="list-style-type: none"> - Alta temperatura (200 °C).
	Enzimática	<ul style="list-style-type: none"> - Altamente seletivo; - Recuperação enzimática; - Não é sensível a presença de ésteres graxos; - Condições brandas de temperatura e pressão; - Baixa formação de resíduos. 	<ul style="list-style-type: none"> - As enzimas podem ser desnaturadas por temperatura ou interação com solventes; - Taxa de reação lenta.
Transesterificação não catalítica	Fluido supercrítico	<ul style="list-style-type: none"> - Não é sensível a presença de água e ésteres graxos; - Alta taxa de reação. 	<ul style="list-style-type: none"> - Requer temperatura (250-400 °C) e pressão (350-600 atm) elevadas; - Maior razão molar triglicerídeo:álcool.
	Cossolvente	<ul style="list-style-type: none"> - Temperatura e pressão ambiente; - Tempo de reação diminuído; - Reuso do cossolvente. 	<ul style="list-style-type: none"> - Alta toxicidade do cossolvente; - Dificuldade na separação do cossolvente e do álcool.

Os triglicerídeos utilizados nos processos de síntese do biodiesel podem ser obtidos através de gorduras animais, óleos vegetais ou óleos residuais. Essas fontes de gordura são majoritariamente compostas por moléculas de triacilgliceróis, ácidos graxos livres, água, esteróis e fosfolípidos. Elas podem ser classificadas com base na sua concentração de ácidos graxos livres. As gorduras de primeira geração contêm

cerca de 1,5% de ácidos graxos livres, como por exemplo os óleos de soja e palma. Já as gorduras de segunda geração possuem menos de 4% de ácidos graxos em sua composição, como é o caso do sebo bovino e dos óleos de cozinha. Por fim, as gorduras de terceira geração contêm mais de 20% de ácidos graxos livres, como é o caso das graxas animais, o que pode ser prejudicial para a síntese do biodiesel a depender do método de transesterificação escolhido (BABADI, *et al.*, 2022).

De modo geral, a fonte de triglicerídeos utilizada na produção do biodiesel varia de acordo com as condições climáticas e a disponibilidade de cada país. No Brasil, segundo o Ministério de Minas e Energia (2023), mais da metade da produção nacional de biodiesel utiliza o óleo de soja como matéria prima, conforme mostrado na Figura 7. A gordura de sebo bovino também tem ganhado espaço nesse mercado, participando de cerca de 8% da produção total de biodiesel no Brasil no ano de 2023. O uso da gordura animal configura menor emissão de carbono para o meio ambiente, além de apresentar menor custo e se tratar de um produto abundante no mercado. Entretanto, as gorduras animais são compostas majoritariamente por ácidos graxos saturados, que são susceptíveis a reações de solidificação quando expostos a baixas temperaturas. A solidificação de tais componentes pode ser prejudicial para os sistemas de injeção e motores a diesel, ocasionando a formação de resíduos e potenciais entupimentos dos sistemas. O grau de insaturação de óleos e gorduras, medido pelo índice de iodo, é uma propriedade importante que deve ser avaliada na matéria prima do biodiesel, já que ele afeta diretamente o seu potencial de oxidação. (RAMOS, *et al.*, 2011).

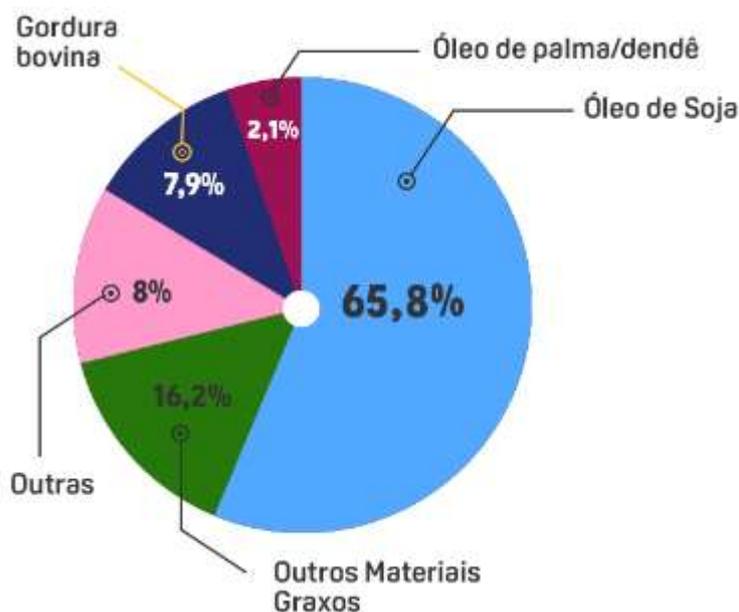


Figura 7. Matéria-prima utilizada na produção do biodiesel no Brasil no ano de 2023 (Ministério de Minas e Energia, 2023).

3.2 Legislação

O biodiesel é totalmente miscível no óleo diesel e possui desempenho semelhante no motor, principalmente no que diz respeito a potência e ao torque. Apesar da combustão do biodiesel fornecer uma quantidade de energia menor, o biocombustível apresenta maior viscosidade em comparação ao derivado do petróleo, o que garante maior lubrificidade e menor ocorrência de desgastes no sistema. Outra vantagem importante é que o biodiesel é obtido através de fontes renováveis com baixa emissão de CO₂, além de possuir ponto de fulgor superior que permite sua aplicação em mecanismos que demandam temperaturas mais elevadas (LÔBO, *et al.*, 2009). No Brasil o biodiesel é utilizado comercialmente apenas como um produto complementar ao óleo diesel. Ele é usado em uma mistura com o óleo diesel, a fim de diminuir gradativamente a demanda pelo combustível derivado do petróleo. A mistura do biodiesel com o diesel é denominada BX, na qual X representa o percentual em volume de biodiesel presente no diesel mineral.

A lei nº 11.097 (BRASIL, 2005), publicada em janeiro de 2005, dispõe sobre a introdução do biodiesel na matriz energética brasileira. Inicialmente foi definido o percentual mínimo obrigatório da adição de 2% v/v de biodiesel no diesel, denominado comercialmente como B2. Ao longo dos anos, novas resoluções foram estabelecidas

com o aumento desse percentual. Em outubro do ano de 2024 foi sancionada a “Lei do Combustível do Futuro” nº 14.993/2024 (BRASIL, 2024) que delibera sobre a promoção e o uso de combustíveis sustentáveis com baixa emissão de carbono, como o diesel verde, o gás natural e o biodiesel. A lei estabelece metas de percentuais obrigatórios na mistura de biodiesel e diesel, iniciando com 15% v/v a partir de março de 2025, com o acréscimo de 1% ao ano até alcançar os 20% v/v no ano de 2030. O Conselho Nacional de Política Energética (CNPE) avaliará a viabilidade dessas metas e definirá o percentual obrigatório da adição de biodiesel no diesel em todo território nacional, respeitando os limites entre 13% e 25% v/v. Também foi acordado pela Lei nº 14.993 que é facultativa a adição de biodiesel no óleo diesel que seja superior ao percentual mínimo obrigatório, desde que as condições de qualidade sejam mantidas. Atualmente, o percentual mínimo obrigatório de biodiesel no diesel é de 14% v/v, segundo definições do CNPE.

As especificações e as obrigações quanto ao controle de qualidade do biodiesel puro (B100) são regulamentadas pela Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – ANP, segundo a resolução ANP nº 920/2023. A Tabela 1 mostra as características e os limites estabelecidos para o biodiesel comercializado no Brasil.

Tabela 1. Especificações do biodiesel definidas pela resolução ANP n° 920/2023.

CARACTERÍSTICAS	UNIDADE	LIMITES
Massa específica a 20 °C	kg/m ³	850 a 900
Viscosidade cinemática a 40 °C	mm ² /s	3,0 a 5,0
Teor de água (máximo)	mg/kg	200,0
Ponto de fulgor (mínimo)	°C	100,0
Teor de éster (mínimo)	% massa	96,5
Cinzas sulfatadas (máximo)	% massa	0,020
Enxofre total (máximo)	mg/kg	10
Sódio + potássio (máximo)	mg/kg	2,5
Cálcio + magnésio (máximo)	mg/kg	2,5
Fósforo (máximo)	mg/kg	4,0
Corrosividade ao cobre, 3h a 50 °C (máximo)	-	1
Índice de acidez (máximo)	mg KOH/g	0,50
Glicerol livre (máximo)	% massa	0,02
Glicerol total (máximo)	% massa	0,23
Monoacilglicerol (máximo)	% massa	0,60
Diacilglicerol (máximo)	% massa	0,20
Triacilglicerol (máximo)	% massa	0,20
Contaminação total (máximo)	mg/kg	24
Metanol e/ou Etanol (máximo)	% massa	0,20
Estabilidade oxidativa a 110 °C (mínimo)	horas	13

3.3 Estabilidade oxidativa do biodiesel

A estabilidade oxidativa do biodiesel demonstra a sua resistência a sofrer processos oxidativos, quando submetido a determinadas condições de temperatura e umidade. Os grupos carboxilatos presentes no biodiesel, assim como a presença de insaturações ou contaminantes, tornam suas moléculas mais susceptíveis a reações de oxidação. Os biodieseis produzidos a partir de óleos vegetais são mais susceptíveis a oxidação, já que o óleo vegetal é formado majoritariamente por ácidos graxos insaturados, como os ácidos oléico e linoléico. Por sua vez, os biodieseis produzidos a partir da gordura animal apresentam maior resistência a oxidação, já que são formados majoritariamente por ácidos graxos saturados, como os ácidos margárico e esteárico. No entanto, além da menor disponibilidade das fontes de

gordura animal, o biodiesel produzido apresenta baixa viscosidade, o que limita sua aplicação em ambientes de baixa temperatura.

A degradação oxidativa do biodiesel ocorre por um mecanismo radicalar composto por três etapas, conforme é mostrado na Figura 8. Na etapa de iniciação do mecanismo radicalar de oxidação do biodiesel os ésteres de ácidos graxos são atacados por espécies altamente reativas, extraindo um átomo de hidrogênio e dando origem a um radical alquila ($R\cdot$). O radical alquila formado na primeira etapa reage com o oxigênio para formar o radical peroxila ($ROO\cdot$) que, por sua vez, é capaz de extrair um átomo de hidrogênio de outro éster dando origem a etapa de propagação. Na etapa final (terminação) os radicais formados nas duas primeiras etapas combinam-se para formar moléculas de peróxido, produto primário da reação de oxidação do biodiesel que, normalmente, se decompõe em álcool, aldeído, cetona ou ácido carboxílico. Os produtos formados na reação de oxidação do biodiesel, normalmente, possuem alto peso molecular e podem causar o entupimento, a deposição de sólidos ou a corrosão dos maquinários que fazem uso do biocombustível. Além disso, o biodiesel com baixa estabilidade oxidativa reduz a qualidade da mistura BX e pode acarretar na formação de água e sólidos no produto final, o que torna esse parâmetro tão importante para seu controle de qualidade. (PULLEN *et al.*, 2012). De acordo com a legislação vigente, o biodiesel comercializado no Brasil deve ter estabilidade oxidativa de no mínimo 13 horas. Esse valor representa o tempo necessário para o início da degradação oxidativa da amostra sob condições padronizadas.

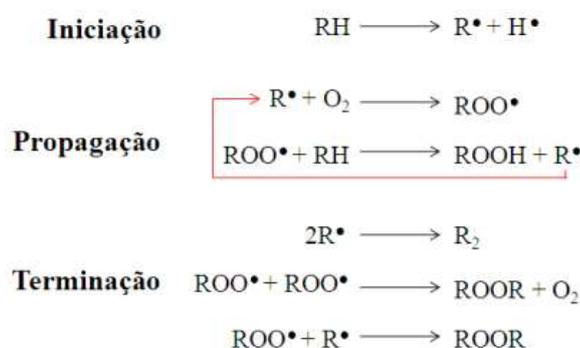


Figura 8. Mecanismo radicalar de oxidação do biodiesel (RAMALHO *et al.*, 2006).

Substâncias antioxidantes podem ser adicionadas na composição do biodiesel para retardar e inibir as reações de oxidação. Alguns antioxidantes já fazem parte da

composição de determinados óleos vegetais, como é o caso dos tocoferóis. Entretanto, as etapas de purificação envolvidas na síntese desses óleos reduzem drasticamente a concentração dos antioxidantes naturais presentes no produto final. Por esse motivo, os antioxidantes mais utilizados no biodiesel são compostos fenólicos de origem sintética, como o butil-hidroxianisol (BHA), o butil-hidroxitolueno (BHT) e o terc-butil-hidroquinona (TBHQ). A estrutura fenólica dos antioxidantes sintéticos doa um próton livre para o radical alquila ($R\cdot$) formado na etapa de iniciação, regenerando a molécula do acilglicerol e interrompendo as reações em cadeia do mecanismo radicalar. Geralmente, as formulações antioxidantes contêm uma mistura das substâncias fenólicas com agentes quelantes. Para isso, é necessário que os antioxidantes combinados possuam efeito sinérgico e propriedades complementares, além de serem estáveis nas mesmas condições de temperatura e pressão. O BHA é o antioxidante mais utilizado no biodiesel, principalmente nos que são sintetizados a partir de gorduras animais. Esse antioxidante é pouco estável em temperaturas elevadas, mas é efetivo para inibir a oxidação de ácidos graxos de cadeia curta, como é o caso dos ácidos utilizados na composição dos óleos de coco e palma (GALVAN, *et al.*, 2014). Normalmente, o BHA é utilizado junto com o BHT devido aos seus efeitos sinérgicos, nos quais o BHA atua como doador de prótons para o radical alquila e o BHT age como regenerador da molécula do BHA. O BHT também é mais estável a temperaturas elevadas, o que amplia a faixa de atuação do BHA quando ambos estão presentes na mistura. Já o TBHQ suporta temperaturas superiores a 200 °C, sendo utilizado principalmente na composição do biodiesel sintetizado a partir de óleo vegetais (DOMINGOS *et al.*, 2007).

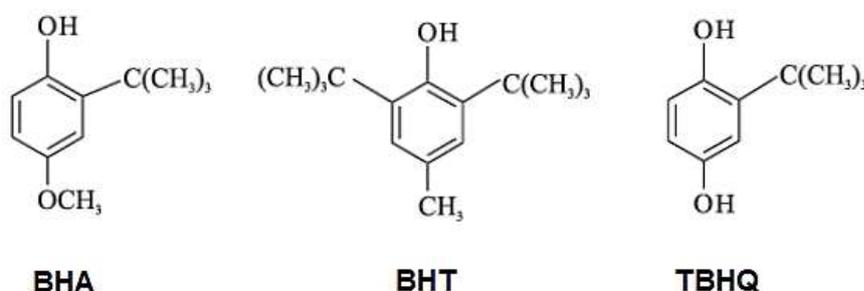


Figura 9. Estrutura molecular dos antioxidantes BHA, BHT e TBHQ.

A resolução ANP n° 920/2023 estabelece que os agentes econômicos autorizados a realizar a mistura do biodiesel com o óleo diesel devem coletar, pelo menos uma vez ao mês, um volume de amostra representativa do tanque de

expedição do B100. Essa amostra deverá ser destinada para a análise para mensurar sua estabilidade oxidativa, de acordo com as normas EN 14112 (*Fat and oil derivatives - Fatty Acid Methyl Esters (FAME) - Determination of oxidation stability (accelerated oxidation test)*) ou EN 15751 (*Automotive fuels - Fatty acid methyl ester (FAME) fuel and blends with diesel fuel - Determination of oxidation stability by accelerated oxidation method*). Os resultados da análise devem ser encaminhados para a ANP, conforme estabelecido pela resolução ANP nº 828/2020.

A norma europeia EN 14112 mede a estabilidade oxidativa do biodiesel pelo método Rancimat, que detecta a variação de condutividade causada pelos vapores liberados durante a reação de oxidação. Nesse método, a amostra é aquecida a 110 °C em um vaso reacional, sob fluxo constante de ar purificado de 10 L/h. Conforme mostrado na Figura 10, o fluxo de ar, representado pelas setas, arrasta os vapores produzidos no aquecimento até uma célula com água deionizada, onde a condutividade é monitorada por um eletrodo que registra variações entre 0 a 300 $\mu\text{S}/\text{cm}$. Um aumento brusco na condutividade marca o fim do período de indução do biodiesel, sendo registrada a sua estabilidade oxidativa em horas. Esse aumento acelerado da condutividade da água está diretamente associado aos produtos formados na reação de oxidação, como os ácidos carboxílicos, que se dissociam e são absorvidos pela água, aumentando sua condutividade (MOURA, *et al.*, 2016).

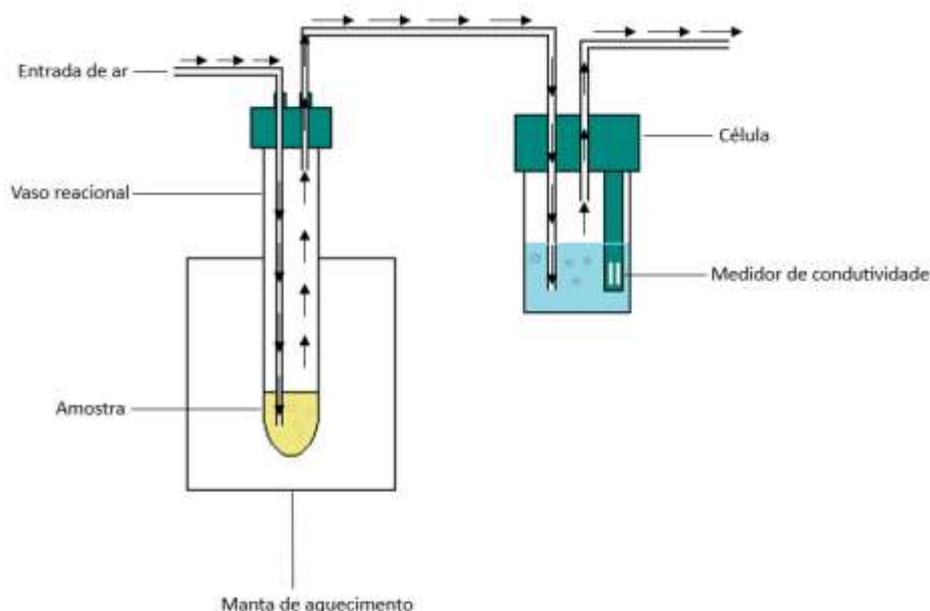


Figura 10. Esquema da montagem do método Rancimat. (adaptado de MOURA, *et al.*, 2016)

Inicialmente, o método proposto na EN 14112 era aplicado apenas na análise do biodiesel puro (B100). Entretanto, com o aumento do uso das misturas do biodiesel com o óleo diesel esse método foi adaptado para suprir as especificidades do novo combustível, originando o método proposto na EN 15751. Na EN 15751 também é utilizado o método Rancimat, porém com o aumento da quantidade de amostra e do comprimento do vaso reacional. O aumento do comprimento do vaso reacional tem por objetivo reduzir a perda de componentes altamente voláteis, uma vez que o comprimento que excede a manta de aquecimento atua como uma espécie de condensador. Já o aumento da quantidade de amostra utilizada acarreta no aumento da sensibilidade da análise, possibilitando a detecção de compostos voláteis formados no início do período de aquecimento (MÉNDEZ *et al.*, 1996).

O método Rancimat possui um tempo médio de 12 horas de análise, além de necessitar de um ambiente controlado com temperatura e fluxo de ar constantes. Por esse motivo, urge a necessidade do desenvolvimento de metodologias que sejam mais rápidas e adaptáveis para a determinação da estabilidade oxidativa do biodiesel B100. Os métodos espectroscópicos associados as ferramentas quimiométricas já tem sido amplamente estudados para esse fim, já que apresentam resultados rápidos e são mais baratos em comparação com outros tipos de análise. Além disso, a possibilidade do uso de espectrômetros portáteis permite que a análise possa ser realizada *in situ*, viabilizando o monitoramento da estabilidade oxidativa do biodiesel durante todas as etapas de produção, transporte e comercialização.

3.4 Espectroscopia Raman

3.4.1 Conceitos fundamentais

A espectroscopia Raman foi descoberta na década de 1920 pelo físico Chandrasekhara Venkata Raman e baseia-se no fenômeno de espalhamento inelástico da radiação. Quando a radiação eletromagnética incide sobre uma substância, diversos fenômenos podem ocorrer, incluindo absorção, emissão e espalhamento da radiação. O espalhamento pode ser classificado em duas categorias principais: o espalhamento elástico e o espalhamento inelástico (RODRIGUES, *et al.* 2012). O espalhamento elástico, chamado de espalhamento *Rayleigh*, ocorre quando a energia espalhada é igual a energia incidida na substância. Quando a radiação

espalhada é diferente da radiação incidida tem-se o que é chamado de espalhamento Raman, um tipo de espalhamento inelástico que permite a identificação de grupos funcionais presentes na substância, por meio da informação da frequência de vibração das suas moléculas.

O espalhamento Raman é dividido em duas classes, o espalhamento *Stokes*, quando a frequência da radiação espalhada é menor que a frequência da radiação incidente, e o *Anti-Stokes*, quando a radiação espalhada tem frequência maior que a radiação incidida. Conforme mostrado na Figura 11.a, o espalhamento *Stokes* ocorre quando a radiação incidente excita uma molécula do seu estado vibracional fundamental v_0 para um estado excitado virtual, de modo que a molécula absorve parte dessa energia e espalha radiação com uma frequência menor ao retornar para o estado vibracional v_1 de mais alta energia. Já no espalhamento *Anti-Stokes*, ao receber a radiação eletromagnética a molécula já se encontra em um estado vibracional excitado v_1 , de modo que ao retornar para o estado vibracional fundamental v_0 ela emite uma radiação com frequência maior que a incidida, conforme mostrado na Figura 11.b (JONES *et al.*, 2019).

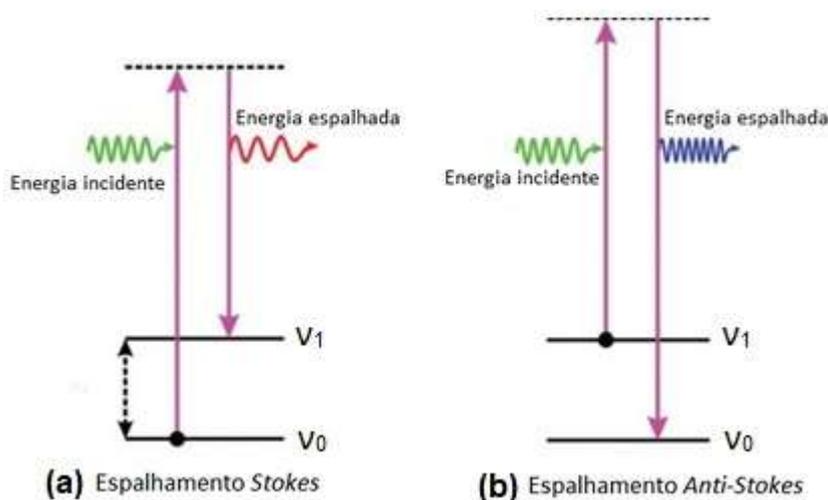


Figura 11. Esquema do princípio do espalhamento RAMAN. a. Espalhamento Stokes; b. Espalhamento Anti-Stokes. (adaptado de GONÇALVES, 2022)

O espectro Raman fornece informações sobre as transições vibracionais permitidas na molécula, caracterizando sua composição molecular e as suas características estruturais. O espectro Raman apresenta informações importantes

sobre os modos vibracionais presentes na amostra como, por exemplo, o estiramento C=O identificado em bandas localizadas na faixa de 1665 a 1750 cm^{-1} ; o estiramento C=C típico de moléculas insaturadas com bandas na faixa de 1600 a 1680 cm^{-1} ; as deformações de CH₂ e CH₃ que podem ser atribuídas as bandas na faixa de 1400 a 1500 cm^{-1} ; a deformação C-H do tipo tesoura na faixa de 1245 a 1280 cm^{-1} ; e o estiramento C-O na região de 800 a 900 cm^{-1} . A Figura 12 apresenta um espectro Raman obtido na análise de uma amostra de biodiesel, com as informações químicas principais destacadas.

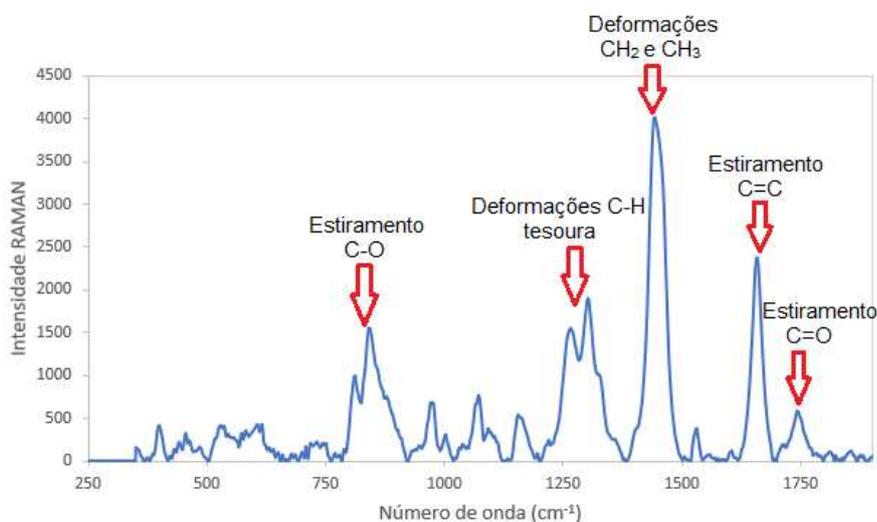


Figura 12. Espectro Raman obtido para uma amostra de biodiesel.

3.4.2 Instrumentação

A configuração geral de um espectrômetro Raman consiste basicamente em uma fonte de radiação monocromática, neste caso, um laser; um filtro de comprimento de onda; um conjunto de espelhos utilizados na focalização do feixe de luz; um conjunto de tubos direcionadores da radiação; o porta amostra; um filtro *notch* que impede a passagem dos feixes espalhados por *Rayleigh*; o espectrômetro; um detector capaz de identificar a intensidade da radiação em cada comprimento de onda e, por fim, um computador e um sistema de processamento de dados que convertem essas informações nos espectros Raman (Figura 13).

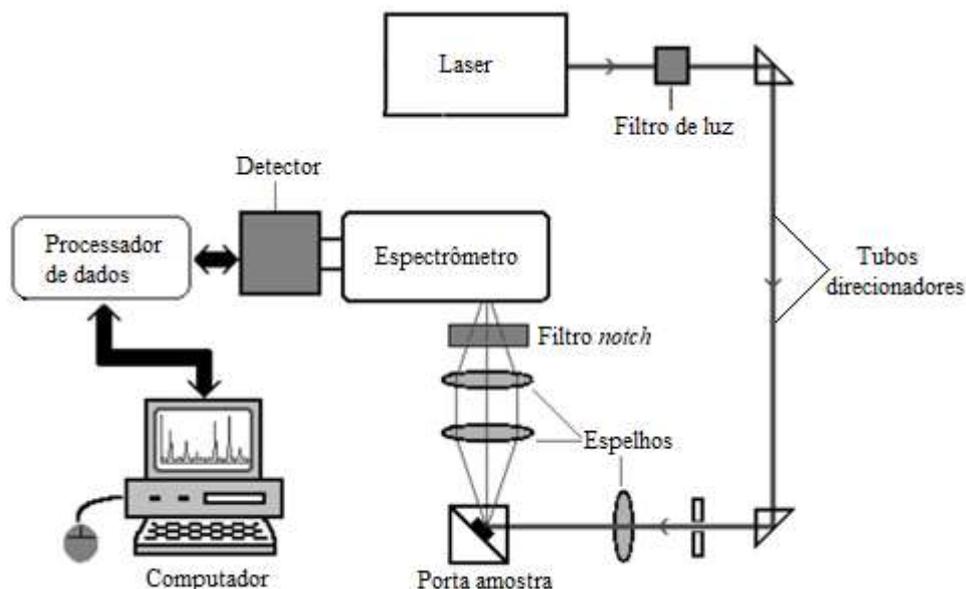


Figura 13. Esquema dos componentes básicos de um espectrômetro Raman. (adaptado de SOUZA et al., 2003)

Existem dois tipos de sistemas Raman, o espectrômetro Raman dispersivo e o espectrômetro Raman por transformada de Fourier. No espectrômetro Raman dispersivo a radiação espalhada inelasticamente é dispersa em uma grade de dispersão e direcionada ao detector. Já no espectrômetro por transformada de Fourier a radiação espalhada é direcionada para um sistema de espelhos, composto por um espelho móvel e um espelho fixo (SCHRADER, 2017). Os espectros são gerados a partir das informações obtidas pelo detector, tratadas com operações matemáticas aplicadas por transformada de Fourier. Nesse caso, a resolução dos espectros é afetada pela posição e distância dos espelhos.

Os lasers utilizados na espectroscopia Raman possuem comprimento de onda na faixa de 200 a 1064 nm, compreendendo a região do ultravioleta até o infravermelho próximo. Normalmente, os lasers com comprimento de onda inferior a 750 nm são utilizados em espectrômetros Raman dispersivos, enquanto que espectrômetros Raman com Transformada de Fourier utilizam de lasers com comprimento de onda superior. Dentre os tipos de laser mais utilizados pode-se citar o laser de HeNe (632 nm), o laser de diodo (660-880 nm) e o laser de Nd:YAG (1064 e 532 nm). O laser escolhido para compor um espectrômetro Raman precisa satisfazer algumas condições como, ter frequência estável, sem que haja mudanças significativas entre replicatas; possuir uma banda de emissão estreita,

preferencialmente menor que 1 cm^{-1} ; ter baixa luminescência, que também possa ser filtrada por componentes do espectrômetro; e emitir um feixe de luz com baixa divergência. Tais características são importantes para reduzir o espalhamento *Rayleigh* durante a análise.

Os filtros utilizados nos espectrômetros Raman são divididos em categorias de acordo com sua função. Os filtros de densidade neutra são utilizados para absorver a luz com um valor de densidade constante, durante toda a faixa de comprimento de onda. Os filtros de passagem longa refletem a luz abaixo de um determinado comprimento de onda e transmitem toda a radiação que não se enquadra nesse limite. Os filtros de passagem curta atuam de forma inversa, transmitindo a luz com menor comprimento de onda e refletindo a luz com maior comprimento de onda. Os filtros passa-banda selecionam uma faixa de frequência de luz que pode ser transmitida, enquanto bloqueia todas as outras frequências. Os filtros bloqueadores de banda bloqueiam uma determinada banda do espectro de emissão. A combinação e o uso desses filtros são importantes, principalmente, para suprimir o sinal provindo do espalhamento *Rayleigh*. Esses filtros são formados por duas placas de vidro preenchidas com um material gelatinoso, denominados como filtros holográficos. A espessura e as propriedades do material gelatinoso irão definir a aplicabilidade do filtro. A principal desvantagem dos filtros holográficos é seu tempo de vida útil, que acaba sendo diminuído a partir da degradação da gelatina em decorrência da absorção de radiação na forma de calor. Como alternativa, os filtros com filmes dielétricos são os mais empregados atualmente. Eles são formados a partir de camadas micrométricas de material transparente dispersas sob um substrato de vidro (GRIFFITHS *et al.*, 1986).

Os detectores utilizados em um espectrômetro Raman precisam ter alta sensibilidade para identificar os fótons emitidos em diferentes comprimentos de onda, de modo a diferenciar o sinal de interesse do ruído de fundo. Nos espectrômetros Raman dispersivos podem ser utilizados detectores do tipo monocanal ou multicanal. No detector monocanal a intensidade da radiação emitida é monitorada de forma sequencial, varrendo toda a faixa espectral de interesse. Já o detector multicanal monitora simultaneamente vários comprimentos de onda, a partir da combinação de detectores monocanal posicionados linearmente. O detector CCD (Dispositivo de Carga Acoplada) é um tipo de detector multicanal com sensibilidade superior aos

demais. O CCD é dividido em várias áreas sensíveis aos fótons, denominadas como pixels, que são dispostos em fileiras intercaladas com os capacitores em um formato de matriz. Os pixels interagem com os fótons emitidos pela amostra, que são posteriormente convertidos em carga elétrica e acumulados nos capacitores vizinhos. O sinal elétrico é transferido para um amplificador e depois para um conversor analógico-digital para originar a resposta. Essa conformação permite que o CCD seja altamente sensível e eficiente para diferenciar sinais em ampla faixa de comprimento de onda. No espectrômetro Raman com transformada de Fourier normalmente é utilizado um detector do tipo semiconductor. Nos detectores semicondutores a radiação é identificada a partir das cargas elétricas geradas no semiconductor ao ser ionizado pela radiação emitida. Os semicondutores mais utilizados nesse tipo de detector são o germânio e o silício.

Os tubos e os componentes transportadores de luz no espectrômetro devem ser produzidos a partir de materiais de fibra de vidro que transmitam ao máximo a radiação incidente. Por isso, tais materiais devem ser deficientes de grupos -OH em sua estrutura, que podem absorver a luz, ou devem ser revestidos por materiais refletivos. O diâmetro, o ângulo e a geometria da fibra de vidro também são importantes, já que essas características determinam o quanto de radiação será transmitida através desses tubos. Já o porta amostra deve ser feito de um material permissivo capaz de transmitir toda a radiação incidente, sem alterar as propriedades físico-químicas da amostra. É importante que os espectros Raman não sejam gerados considerando o material componente do porta amostra, o que pode causar erros de interpretação dos espectros (VANDENABEELE, 2013).

3.4.3 Espectroscopia Raman com deslocamento espacial (SORS)

A espectroscopia Raman convencional baseia-se exclusivamente no espalhamento inelástico da radiação eletromagnética incidida sobre a amostra. Para ampliar suas aplicações e otimizar as imagens espectrais, variações da técnica foram desenvolvidas, como o TERS (*Tip-Enhanced Raman Spectroscopy*), o SORS (*Spatially Offset Raman Spectroscopy*), o SERS (*Surface Enhanced Raman Spectroscopy*) e o CARS (*Coherent Anti-stokes Raman Spectroscopy*). Essas

variações envolvem tanto avanços na instrumentação analítica quanto modificações na superfície da amostra.

Na espectroscopia Raman convencional o laser e o detector são posicionados juntos no equipamento, de forma que a radiação incidida e a radiação espalhada possuem a mesma direção (Figura 14.a). Essa técnica é limitada a detecção de substâncias revestidas por materiais transparentes e claros, já que os recipientes opacos e escuros podem atenuar o feixe de luz antes mesmo de ele entrar em contato com a amostra. Para mitigar tal problema, a espectroscopia Raman com deslocamento espacial (SORS, do inglês *Spatially Offset Raman Spectroscopy*) foi desenvolvida com uma configuração diferente, na qual o laser e o detector do equipamento são separados por um deslocamento espacial ΔS , de modo que os fótons se difundem mais profundamente na amostra, como mostrado na Figura 14.b (MOSCA, *et al.*, 2018). O SORS se baseia em múltiplas medições, considerando diferentes valores de ΔS . Com a aquisição de todas essas medidas, o espectro final da amostra é produzido desconsiderando a contribuição do recipiente. Dessa forma, o SORS pode ser utilizado em amostras contidas em diferentes tipos de recipientes, admitindo maior segurança para o operador e a preservação do conteúdo do frasco (MATOUSEK, 2018).

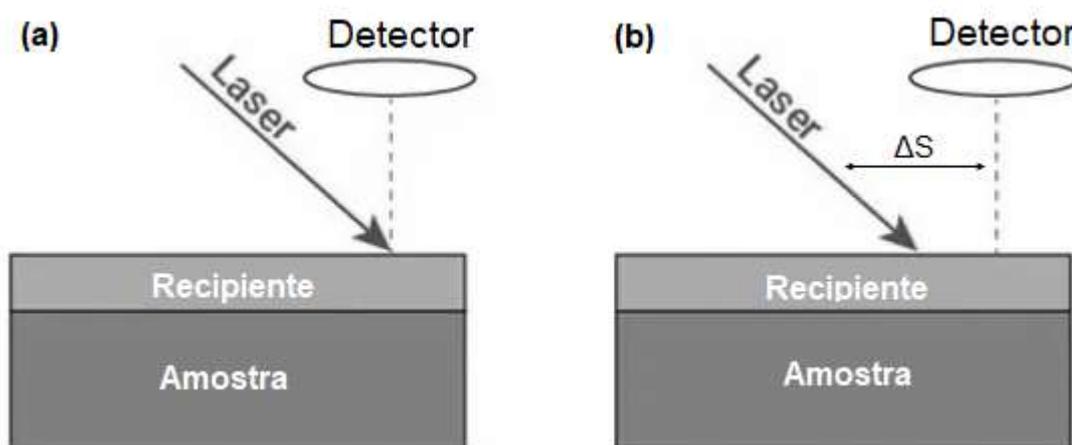


Figura 14. Configuração Raman, a. convencional, b. SORS. (AGILENT, 2024).

3.5 Quimiometria

A quimiometria é o ramo da Química que utiliza de ferramentas matemáticas e computacionais para otimizar experimentos e extrair o máximo de informações do seu conjunto de dados (FERREIRA, 2015). O campo de estudo da quimiometria pode ser dividido em duas grandes áreas, a parte de planejamento de experimentos, voltada para a otimização das condições experimentais, e a parte da análise de dados, que busca entender e extrair todas as informações contidas em dados multivariados. A análise de dados ainda pode ser subdividida em métodos supervisionados (ex. SIMCA, PLS-DA, LDA, PLS), que dependem de alguma informação prévia sobre o conjunto de dados, e métodos não supervisionados (ex. PCA, HCA), que dependem apenas dos resultados experimentais (GONÇALVES, 2022).

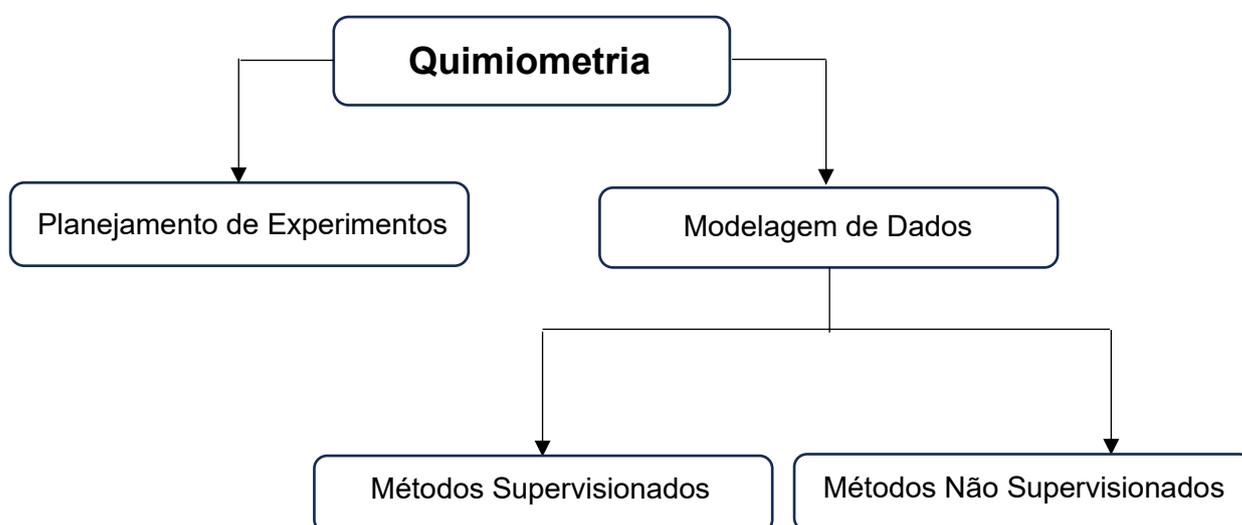


Figura 15. Áreas de estudo da quimiometria.

Os dados experimentais comumente empregados na análise quimiométrica são obtidos a partir da resposta instrumental fornecida pela técnica de análise escolhida. Quando uma amostra é submetida a uma análise instrumental (ex. cromatografia e espectroscopia) o equipamento utilizado emite sinais de resposta em função das variáveis de interesse. Na espectroscopia Raman os picos do espectro de uma amostra são os sinais de resposta que o equipamento fornece em função da faixa do número de onda, determinada como a variável do sistema.

3.5.1 Organização e visualização dos dados experimentais

A primeira etapa para analisar um conjunto de dados instrumentais consiste em organizar, visualizar e pré-processar os valores brutos fornecidos pelo equipamento.

Os dados instrumentais podem ser organizados como uma matriz, denotada por \mathbf{X} ($I \times J$), em um arranjo ordenado de linhas (I) e colunas (J), na qual cada elemento x_{ij} é o sinal de resposta do equipamento para a variável J referente a amostra I . Para dados espectrais as variáveis representadas pela letra J assumem valores numéricos contínuos, sendo compostas pela faixa espectral avaliada (FERREIRA, 2015).

$$\mathbf{X} (I \times J) = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} & \cdots & x_{1j} \\ x_{21} & x_{22} & x_{23} & x_{24} & \cdots & x_{2j} \\ x_{31} & x_{32} & x_{33} & x_{34} & \cdots & x_{3j} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{i1} & x_{i2} & x_{i3} & x_{i4} & \cdots & x_{ij} \end{bmatrix} \quad (1)$$

Durante a organização dos dados é importante se atentar aos elementos x_{ij} ausentes ou discrepantes. Nesses casos, os elementos ausentes ou discrepantes podem ser apenas retirados do conjunto de dados, caso ocorram aleatoriamente no sistema, ou substituídos por valores estimados, quando é observado um padrão sistemático dos dados faltantes. A substituição desses elementos pode ser feita utilizando diversos métodos estatísticos para definir um novo valor, seja: tomando o valor médio dos elementos restantes; utilizando uma média ponderada dos valores mais próximos; ou aplicando exaustivamente a análise de componentes principais (PCA) até obter um novo valor que seja coerente com o conjunto de dados e possua resíduos iguais a zero. No caso da espectroscopia Raman, os espectrômetros atuais possuem *softwares* que fazem a correção de dados discrepantes, como é o caso da identificação e remoção de sinais de raios cósmicos, comumente chamados de *spikes*.

3.5.2 Pré-processamento de dados espectrais

Para reduzir as variações aleatórias ou as informações sistemáticas indesejáveis que ocorrem no conjunto de dados espectrais, seja por erros na experimentação ou por condições de variabilidade na análise, as etapas de pré-tratamento são aplicadas antes da análise exploratória. Na espectroscopia Raman os principais métodos utilizados no pré-processamento dos dados são: alisamento/suavização do espectro, aplicado para redução de ruídos; cálculo das derivadas, utilizado para corrigir os deslocamentos ou a inclinação da linha de base do espectro; e a centragem dos dados na média, utilizado para centrar os espectros em torno do zero (BAHRI *et al.*, 2019).

Em relação às ferramentas de alisamento, o método de Savitzky-Golay é o mais empregado. O algoritmo ajusta um polinômio de determinada ordem aos dados dentro de uma janela com um número específico de pontos, realizando uma média ponderada dos pontos. Nesse caso, a resposta alisada é dada pelo valor do polinômio ajustado e é inserida no ponto central da janela. O método de Savitzky-Golay depende integralmente do tamanho da janela escolhida e do grau do polinômio a ser utilizado, de modo que a mudança desses valores altera significativamente o esboço do gráfico alisado. Alisamentos feitos com polinômio de maior grau e janela pequena são indicados quando há a intenção de preservar o formato do espectro. Já os alisamentos feitos com polinômio de menor grau mas com janelas maiores são recomendáveis quando a diminuição do ruído é mais interessante do que se manter a forma do espectro (FERREIRA, 2015).

Os cálculos das derivadas também são etapas de pré-tratamento que são usualmente utilizadas na transformação de dados de espectroscopia. Elas são aplicadas para corrigir os deslocamentos ou a inclinação da linha de base de um espectro. Quando observa-se o deslocamento da linha de base do espectro, seja por variabilidades instrumentais ou por problemas na amostragem, a primeira derivada pode ser aplicada e calculada a partir da diferença de dois valores consecutivos de intensidade. Para problemas relacionados à inclinação da linha de base, a segunda derivada é comumente utilizada e é calculada com base nos valores obtidos na primeira derivada (MARIA, 2019). Alguns *toolbox* de tratamento de dados (ex. *PLS Toolbox*) já aplicam a transformação da derivada juntamente com o alisamento pelo método de Savitzky-Golay, de modo que os ruídos e a linha de base do espectro são corrigidos simultaneamente.

No método de centragem dos dados na média (*mean center*) o valor médio é calculado para cada variável e subtraído de todas as respostas x_{ij} obtidas. Essa aplicação implica na translação dos eixos para o valor médio de cada um deles, o que não acarreta na deformação da estrutura do espectro e o posiciona em torno do zero.

3.5.3 Análise de Componentes Principais (PCA)

Após escolher e aplicar o pré-processamento adequado ao conjunto de dados, as técnicas de modelagem podem finalmente ser utilizadas para extrair o máximo de

informações relevantes dos resultados. A análise de componentes principais (PCA) é um dos métodos não supervisionados mais utilizado para análise exploratória de dados multivariados. Neste método o conjunto de dados é decomposto em outro sistema de coordenadas, com o intuito de reduzir a dimensão espacial dos dados e ressaltar aqueles que possuem maior variância (RINGNÉR, 2008). Conforme é mostrado na Figura 16, o conjunto de dados representado pela matriz \mathbf{X} é decomposto no produto da matriz \mathbf{T} , que contém os escores, com a matriz \mathbf{P} , que contém os pesos, adicionado a matriz \mathbf{E} , que representa os resíduos não modelados pela PCA.

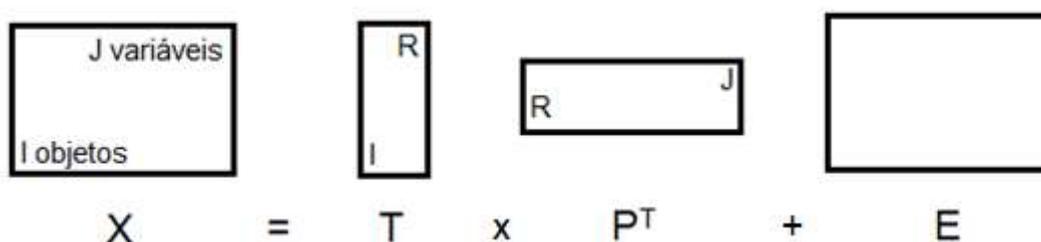


Figura 16. Modelo de uma PCA representado por matrizes. (BRO, *et al.*, 2014)

A decomposição da matriz \mathbf{X} gera as componentes principais (PCs) do modelo PCA, constituindo novas variáveis direcionadas ao longo do eixo de maior dispersão dos dados e que são geradas a partir da combinação linear das variáveis originais. A principal vantagem da PCA é transformar um conjunto de variáveis intercorrelacionadas em um conjunto menor de variáveis ortogonais que explicam a variação dos dados de forma decrescente e independente (HONGYU, 2016). O vetor peso representa a contribuição de cada variável original para as componentes principais e o vetor escore mostra a projeção da variável original na nova componente principal. A combinação dos vetores escore e peso formam as matrizes \mathbf{T} e \mathbf{P} (DUNN, 2014).

Do ponto de vista geométrico, a PCA consiste em representar o conjunto de dados em um novo espaço, no qual os novos eixos (PCs) correspondem às direções de maior variância dos dados, como mostrado na Figura 17.a. Os eixos originais correspondem as variáveis do sistema antes da decomposição dos dados. O novo conjunto de dados é alterado para que a origem dos resultados seja dada pelas PCs traçadas. No sistema de coordenadas da PCA, o escore é dado pela distância do ponto que representa a amostra até a origem no novo espaço das PCs (Figura 17.b),

já o peso é descrito pelo cosseno do ângulo entre cada variável e cada componente principal (Figura 17.c) (DAFFERTSHOFER *et al*, 2004).

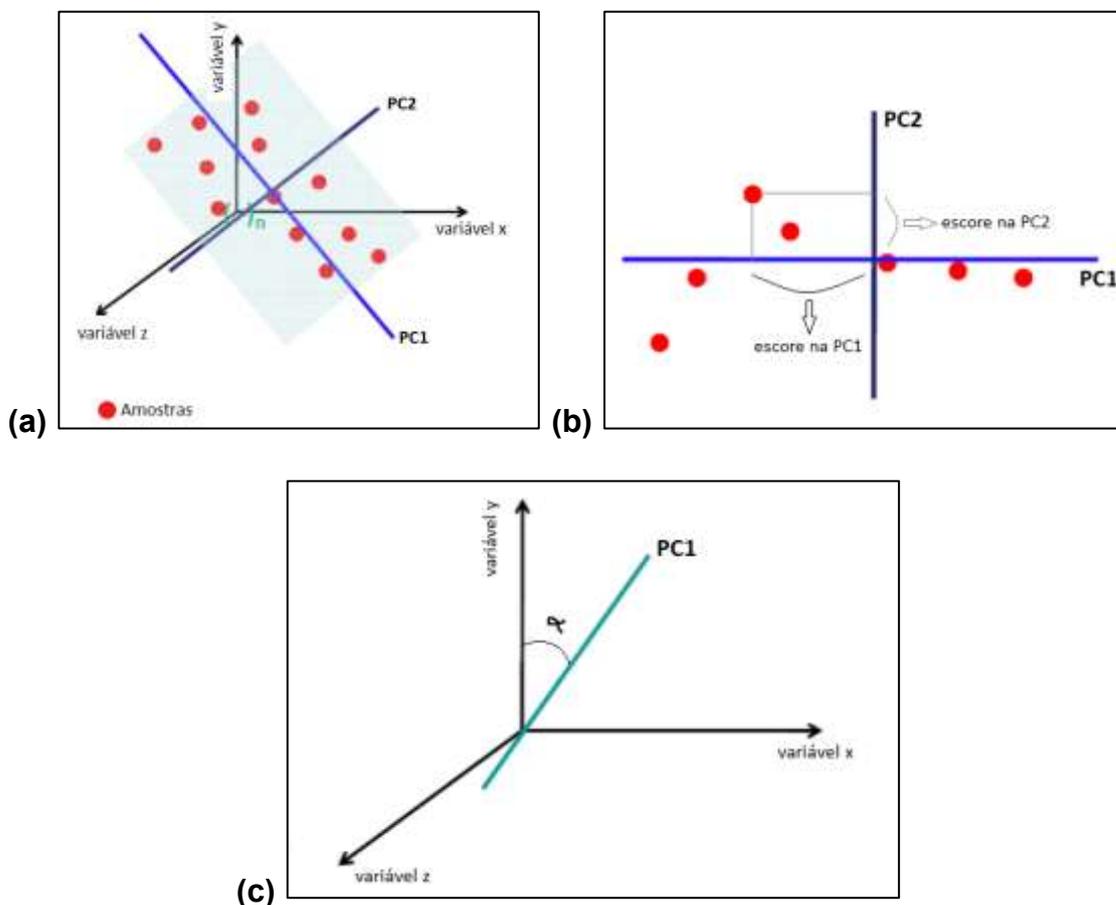


Figura 17. Interpretação geométrica de uma PCA. a. Sistema cartesiano com os eixos iguais as variáveis e componentes principais PC1 e PC2 traçadas segundo a direção de maior variabilidade dos dados; b. Representação do escore como sendo a distância entre a projeção da amostra no eixo até a origem; c. Ângulo α entre a componente principal e o eixo da variável para determinação do peso.

Com o avanço da tecnologia computacional, vários *softwares* são utilizados como ferramentas no tratamento de dados multivariados e já inferem diretamente os resultados de uma PCA sobre um conjunto de dados expressados como uma matriz. Nesse trabalho foi utilizado o MATLAB com o pacote computacional *PLS_Toolbox* (Eigenvector Res. Inc.) para a análise exploratória dos dados. A PCA retorna o gráfico de autovalores das PC's e os gráficos de escores e peso. O gráfico de autovalores revela as variâncias parciais e acumuladas explicadas por cada componente principal, sendo usado para definir o número de PCs que será utilizado na construção da PCA. Normalmente é escolhido o número de componentes principais que descreve a maior parte da variação total do conjunto de dados. Segundo o critério de Kaiser, no gráfico

de autovalores as PCs com autovalores maiores que um representam as componentes principais que explicam a maior variância dos dados (KAISER, 1958). O gráfico de escores das PCs é utilizado para visualizar a distribuição das amostras e sua dispersão em relação às componentes principais, enquanto o gráfico de pesos é utilizado para identificar quais variáveis originais são mais significativas para a formação de cada componente principal (MOITA NETO, *et al.*, 1998).

3.5.4 Modelagem de dados – Métodos Supervisionados

Os métodos supervisionados trabalham com conjuntos de dados dos quais já se conhece alguma informação. Diferentemente da PCA, eles atuam com uma matriz Y ou vetor y , contendo informações prévias sobre a amostra, que podem conter dados numéricos, para métodos quantitativos, ou dados categóricos, para métodos qualitativos. No caso de métodos qualitativos, foco desse trabalho, são atribuídas classes para as amostras do tipo “positivo ou negativo”, “falso ou verdadeiro”, “conforme ou não conforme”, de acordo com as medidas experimentais tomadas como referência.

Para fazer a classificação de uma amostra segundo um método supervisionado é preciso obter um conjunto de dados que contenha amostras representativas de pelo menos uma das classes de interesse. As amostras avaliadas experimentalmente serão divididas em dois conjuntos, o conjunto de treinamento e o conjunto de teste. Os dados experimentais obtidos na análise das amostras do conjunto de treinamento serão usados para construir o modelo empírico com a regra de classificação. O modelo empírico construído é então testado por meio das amostras que constituem o conjunto de teste/validação. Se o teste for satisfatório, o modelo construído poderá ser utilizado para classificar amostras externas ao conjunto de dados previamente selecionado.

Os métodos supervisionados de classificação podem ser divididos em várias categorias, como lineares ou não lineares, paramétricos ou não paramétricos e discriminantes ou de modelagem de classe. Os métodos paramétricos (ex.: SIMCA e LDA) consideram que as variáveis avaliadas seguem uma distribuição normal e usam a homogeneidade da matriz de variância/covariância para delimitar a região de classificação. Os métodos não paramétricos (ex.: k-NN) são usados quando a

distribuição dos dados de cada classe não segue uma distribuição normal ou não é conhecida, baseando-se apenas na matriz inicial e nas informações prévias de cada classe. Nos métodos discriminantes o hiperespaço das variáveis é dividido em regiões distintas, delimitando as áreas correspondentes a cada classe. Já nos métodos de modelagem de classes cada classe é modelada individualmente, sem considerar as demais regiões.

Neste trabalho, foram utilizados os métodos supervisionados de classificação, SIMCA (*Soft Independent Modeling of Class Analogy*), floresta aleatória (*random forest*) e PLS-DA (*Partial Least Squares Discriminant Analysis*), os quais serão desenvolvidos e explicados nos itens abaixo.

3.5.5 Soft Independent Modeling of Class Analogy – SIMCA

O método SIMCA, do inglês *Soft Independent Modeling of Class Analogy*, foi um dos primeiros métodos de classificação supervisionada desenvolvidos na área da Quimiometria. Nesse método assume-se que os valores experimentais obtidos para um conjunto de amostras seguem uma distribuição normal e são modeláveis, considerando que trata-se de um modelo paramétrico e preditivo. Cada uma das classes, definidas de antemão para o conjunto de amostras avaliado, pode ser descrita por um sistema de componentes principais. As classes e suas PCs correspondentes são representadas em um novo espaço tridimensional, com dimensões definidas pelo número de componentes principais em cada classe. Os limites de cada classe são traçados com base no valor absoluto do escore máximo de cada PC, delimitando um hiperplano e, posteriormente, uma hipercaixa que contém os resíduos com as informações que não foram modeladas. Essas classes podem ou não se sobrepor no espaço em que ocupam, de modo que o espaço entre cada classe determina o poder de classificação do modelo SIMCA construído. A Figura 18 mostra a representação de duas classes de amostras e suas respectivas delimitações, seguindo o modelo de classificação SIMCA. Nessa figura é possível observar que duas componentes principais descrevem a classe 1, enquanto que apenas uma componente principal é suficiente para descrever as características das amostras da classe 2.

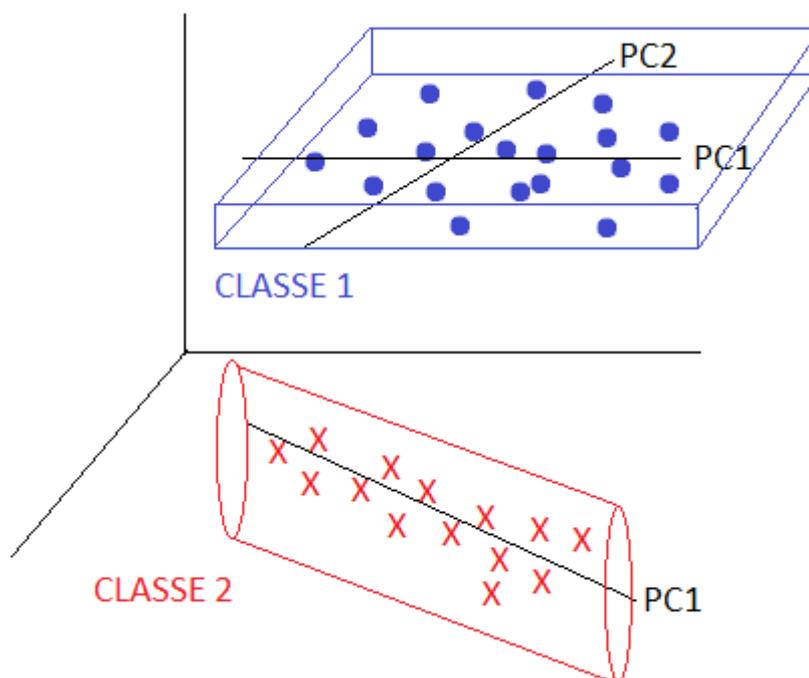


Figura 18. Limites das hipercaixas definidas para duas classes de amostras, descritas pelo modelo SIMCA (adaptado de FERREIRA, 2015).

Os limites da classe de interesse são estabelecidos através dos valores de T^2 de Hotelling e de Q resíduos, calculados a partir das Equações 2 e 3, nas quais: \mathbf{x}^T ($1 \times J$) representa o vetor de medidas de uma amostra, a matriz $\mathbf{\Lambda} = \mathbf{T}^T\mathbf{T} = \text{diag}(\lambda_1, \dots, \lambda_A)$ representa a diagonal com elementos λ_A , que, por sua vez, correspondem aos autovalores da matriz $\mathbf{X}^T\mathbf{X}$ dispostos em ordem decrescente, e \mathbf{P}^T representa a transposta da matriz de pesos \mathbf{P} , que contém os autovetores associados aos autovalores λ_A da matriz $\mathbf{X}^T\mathbf{X}$. Os valores de T^2 de Hotelling fornecem informações sobre a adequação das características da amostra em relação ao modelo construído. Já os valores de Q resíduos indicam o ajuste do modelo em relação ao conjunto de dados avaliado. Espera-se que, para um modelo SIMCA bem ajustado as amostras possuam baixos valores de T^2 de Hotelling e de Q resíduos (MÁRQUEZ *et al.*, 2016).

$$T^2 = xP\Lambda^{-1}P^T x^T \quad (2)$$

$$Q = \|x^T(I - PP^T)\|^2 \quad (3)$$

Após a projeção das classes no novo espaço tridimensional, a amostra teste é atribuída a uma dessas classes, com base na distância da amostra até as fronteiras da classe. Quando o modelo não possui poder de classificação ou quando a amostra

teste é anômala em relação ao conjunto de treinamento, pode ocorrer da amostra teste ocupar duas classes simultaneamente ou nenhuma classe.

Uma amostra teste é representativa da sua classe quando a sua variância residual é menor ou igual à variância residual média da classe. Se sua variância residual for significativamente maior que a variância residual média da classe, a amostra estará distante da classe avaliada ou poderá ser considerada anômala para todo o modelo construído. Para avaliar a significância da variância residual da amostra teste aplica-se um teste F. A estatística F também pode ser utilizada para determinar um valor limite para a variância residual de uma amostra, para que ela se enquadre dentro da classe construída pelo modelo SIMCA (FERREIRA, 2015). Com o valor F_{critico} tabelado com determinado grau de confiança e o valor da variância residual média da classe, também é possível determinar a distância limite no espaço tridimensional para que uma amostra seja identificada naquela classe.

3.5.6 Análise Discriminante por Mínimos Quadrados Parciais - *Partial Least Squares Discriminant Analysis* (PLS-DA)

A análise discriminante por mínimos quadrados parciais (PLS-DA) descreve a relação entre uma matriz multivariada \mathbf{X} e uma matriz \mathbf{Y} (ou vetor \mathbf{y}) contendo os vetores de resposta \mathbf{y} que, por sua vez, especifica as classes de interesse codificadas. Quanto trata-se de um problema de classificação binária, admite-se que a classe 1 é formada pelas amostras da classe de interesse e a classe 0 seja formada pelas amostras que não pertencem à classe de interesse. Assim como é feito na PCA, no PLS-DA os dados de \mathbf{X} e \mathbf{Y} são projetados em um novo espaço, formado pelas variáveis latentes. Cada variável latente relaciona os dados da matriz \mathbf{X} com a predição da variável dependente \mathbf{y} , sendo obtida maximizando a covariância entre os escores \mathbf{t} da matriz \mathbf{X} e da variável dependente \mathbf{y} (BERRUETA *et al.*, 2007).

Inicialmente, as matrizes \mathbf{X} e \mathbf{Y} são submetidas a uma etapa de redução da dimensionalidade, sendo decompostas nas matrizes de escores (\mathbf{T} e \mathbf{U}) e peso (\mathbf{P} e \mathbf{Q}), conforme as Equações 4 e 5.

$$X = T \times P^t + E \quad (4)$$

$$Y = U \times Q^t + F \quad (5)$$

Em seguida, os escores da matriz \mathbf{X} (T) e da matriz \mathbf{Y} (U) são correlacionados através de um vetor de regressão \mathbf{b} , dado por:

$$U = b \times T \quad (6)$$

O vetor de regressão que melhor representa a correlação desejada é calculado considerando o número de variáveis latentes a serem mantidos e a retirada de amostras anômalas do conjunto de dados. Para tal, o PLS-DA utiliza a validação cruzada para repetir a calibração n vezes e determinar o número de variáveis latentes que origina o menor erro de previsão (FERREIRA, 2015).

Os resultados previstos pelo vetor de resposta representam os valores preditos pelo modelo para classificar as amostras na classe 1 ou 0. Para definir os limites entre as classes é aplicada a estatística *Bayesiana*, na qual é construído um histograma ajustado com distribuição normal, que prevê a probabilidade da amostra pertencer a classe 1 ou 0. O cruzamento entre as duas curvas de distribuição de cada classe representa o limiar ótimo para a classificação.

Para avaliar quais variáveis são importantes para a separação das classes é feito o cálculo das variáveis importantes na projeção (*Variable importance in projection - VIP*), determinadas através da seguinte equação:

$$VIP = \sqrt{p \sum_{r=1}^R \left[SS_r \left(\frac{w_{rj}}{\|w_r\|} \right)^2 \right] / \sum_{r=1}^R SS_r} \quad (7)$$

na qual R é o número de variáveis latentes, SS_r é a soma dos quadrados explicados pela r -ésima variável latente e $\left(\frac{w_{rj}}{\|w_r\|} \right)^2$ representa o peso da j -ésima variável em relação ao peso total das variáveis. O valor de SS_r é calculado como descrito na Equação 8, onde t_r é o escore da r -ésima componente e c_r é o coeficiente de regressão que relaciona essa componente com a variável dependente. A predição da variável dependente (y_r) usando apenas essa componente é dada pela Equação 9 e SS_r pode ser interpretado como a variância explicada por y_r . Já o vetor de pesos w_r , que determina a contribuição de cada variável preditora na formação da r -ésima componente latente, é calculado como descrito na Equação 10, onde \mathbf{X} é a matriz das

variáveis preditoras e t_r é o escore da componente. A normalização por $\|X^T t_r\|$ garante que w_r mantenha a variância unitária (WOLD *et al.*, 1998).

$$SS_r = t_r^T t_r c_r^T c_r \quad (8)$$

$$y_r = t_r c_r \quad (9)$$

$$w_r = \frac{X^T t_r}{\|X^T t_r\|} \quad (10)$$

3.5.7 Árvore de Decisão

A árvore de decisão é um método supervisionado que pode ser utilizado para modelos de regressão ou classificação, e usa regras pré-determinadas para dividir um conjunto de dados em subconjuntos que minimizam erros e maximizam as funções de ganho. Em modelos de árvore de decisão binários, cada regra é definida como um nó e possui duas respostas possíveis, visualizadas como as ramificações da árvore. A regra mais importante para o conjunto de dados é a primeira a se apresentar no fluxograma, sendo definida como raiz. A partir da raiz, as próximas regras vão sendo inseridas na árvore, formando suas ramificações até convergir para um conjunto de dados final, definido como folha ou nó terminal. A Figura 19 mostra a estrutura de uma árvore de decisão genérica, cada nó interno representa uma regra de decisão, e cada nó terminal representa o conjunto final de amostras obtido (QUINLAN, 2014).

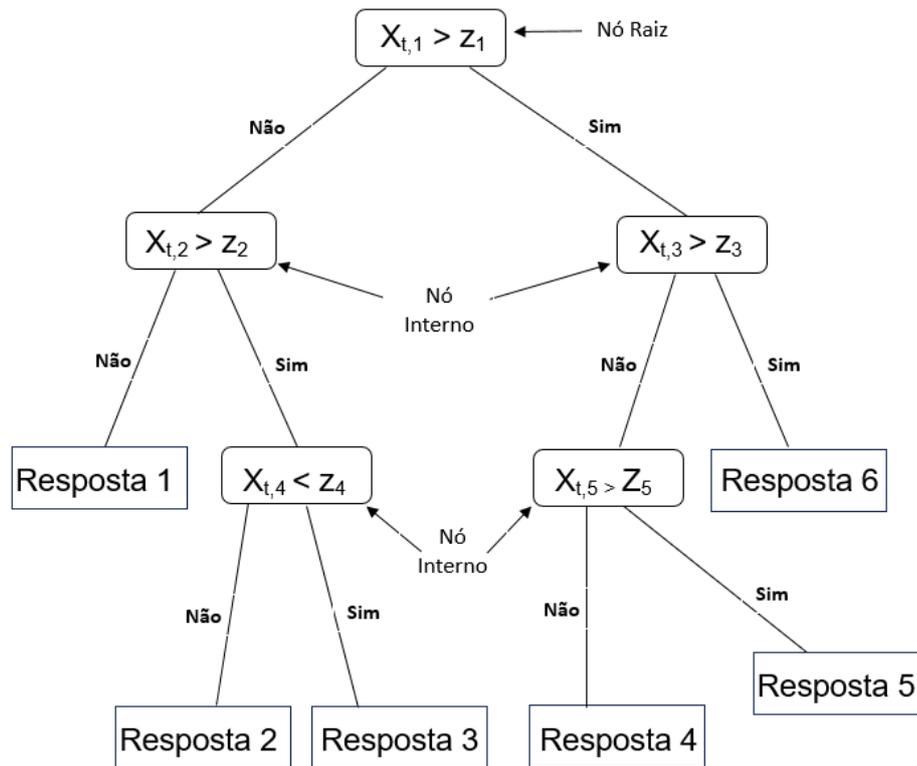


Figura 19. Estrutura genérica de uma árvore de decisão (adaptado de SANTANA, 2020).

Em problemas de classificação binária o índice de Gini é utilizado como regra padrão para definir as divisões da matriz na árvore de decisão. O índice de Gini mede a dispersão dos dados considerando a classe de interesse (a) e a proporção dessa mesma classe (p_a). Quanto menor o valor indicado pelo índice de Gini (G), maior as informações da classe “ a ” contidas naquele nó, resultando na escolha da variável para criar a partição (CAO *et al.*, 2012).

$$G = 1 - \sum_{a=1}^n (p_a^2) \quad (11)$$

A árvore de decisão pode conter múltiplos nós e se tornar muito complexa, apresentando erros elevados devido ao sobreajuste dos conjuntos de treinamento ou calibração. Para evitar esse problema, pode ser realizada a técnica de “poda da árvore”, que consiste em reduzir o número de nós no fluxograma. A poda pode ser feita através da redução do número de nós com menos dissimilaridades, da redução do número mínimo de amostras em cada subconjunto ou da redução do número de variáveis antes de cada divisão. Outra estratégia utilizada para evitar o sobreajuste da árvore de decisão é combinar a predição de várias árvores, como ocorre no modelo floresta aleatória. Essa combinação faz com que cada árvore seja treinada com um

subconjunto aleatório dos dados, reduzindo a variância dos dados e aumentando a robustez do modelo.

3.5.8 Floresta Aleatória – *Random Forest*

A floresta aleatória é um método que combina várias árvores de decisão em um único modelo, a fim de eliminar o sobreajuste e a complexidade encontrada em uma única árvore com múltiplos nós. Existem diversas formas de seleção e aleatorização das árvores de decisão escolhidas para compor a floresta aleatória. O *Bagging* é uma ferramenta utilizada para separar a árvore de decisão em conjuntos de treinamento artificiais, conjuntos *Bootstrap*, e treinar esses subconjuntos, de modo a reduzir os erros de previsão. Cerca de dois terços das amostras do conjunto *Bootstrap* é utilizado para construir a nova árvore de decisão, sendo que o um terço restante é direcionado para testar e avaliar o desempenho do modelo. Outro processo de aleatorização das árvores de decisão é a seleção aleatória das variáveis, que consiste em selecionar um número de variáveis específico para cada nó, durante a construção do modelo (WANG *et al.*, 2020).

O método de classificação por floresta aleatória apresenta diversas vantagens em relação a outros métodos, como: aplicação em sistemas não lineares e em variáveis que não apresentam distribuição normal, diminuição da influência de ruídos e amostras anômalas, e compreensão sobre a importância das variáveis relevantes para o modelo. Dentre os métodos comumente utilizados para avaliar a importância das variáveis em um modelo de floresta aleatória, podem-se citar a permutação binária de uma variável com as demais e a avaliação do decréscimo médio da impureza de Gini (CARVALHO *et al.*, 2016).

A permutação binária de uma variável com as demais permite que a relação original desta variável com a propriedade de interesse seja extinta, sendo construído um modelo com nova eficiência e exatidão para cada permuta realizada. Quando a variável permutada está associada a propriedade de interesse a permutação afeta negativamente a eficiência do modelo, caso contrário a eficácia do modelo não é afetada.

Já o decréscimo médio da impureza de Gini mede o quanto a impureza de Gini diminui quando a variável em questão é selecionada, considerando que o cálculo é ponderado em relação ao número de vezes que essa mesma variável é selecionada. A importância de uma variável β é dada por:

$$Imp(\beta) = \frac{1}{k} \sum \alpha \sum \theta \Delta i \quad (12),$$

na qual α está associada ao número total de árvores da floresta, θ indica somente as árvores que contêm a variável β e Δi é a diminuição da impureza de Gini (GONÇALVES, 2022).

3.5.9 Métodos de balanceamento de classes

Os métodos de classificação preveem a classificação das amostras com base no conjunto de dados e nas classes de amostras previamente estabelecidas. Quando o conjunto de dados contêm classes desbalanceadas, ou seja, uma classe majoritária com um número de amostras superior a classe minoritária, podem ocorrer erros de classificação enviesados pelas características da classe majoritária.

Para minimizar os erros por desbalanceamento de classes, duas estratégias podem ser adotadas: a sobreamostragem, que adiciona amostras na classe minoritária, e a subamostragem, que reduz o número de amostras da classe majoritária, a fim de equilibrar o número de amostras nas duas classes. No entanto, a subamostragem pode resultar na diminuição expressiva do número de amostras ou na perda de dados significativos para a análise classificatória. Por outro lado, a sobreamostragem pode levar ao sobreajuste do modelo, tornando-o menos eficiente para classificar novas amostras (DEY *et al.*, 2023).

Os métodos de sobreamostragem são comumente mais utilizados do que os de subamostragem, já que eles preservam as informações e o número de amostras do conjunto de dados original. O SMOTE (do inglês *Synthetic Minority Over-sampling Technique*) e o ADASYN (do inglês *Adaptive Synthetic Sampling Approach for Imbalanced Learning*) são exemplos de métodos de sobreamostragem utilizados no balanceamento de classes em problemas binários.

O ADASYN é um método de sobreamostragem que utiliza a distribuição das amostras da classe minoritária para gerar novas amostras sintéticas que serão incorporadas a essa classe. As amostras sintéticas são criadas em torno das amostras da classe minoritária com menor grau de aprendizagem, de modo que as novas amostras aumentem as informações contidas na região duvidosa (HAIBO HE *et al.*, 2008).

Para aplicar o ADASYN, primeiramente, é definido um parâmetro β , que expressa o grau de desbalanceamento aceitável para as classes. Para um conjunto perfeitamente balanceado, ou seja, com número igual de amostras na classe majoritária e minoritária, admite-se β igual a 1. Em seguida, é calculado o número de amostras sintéticas G a serem criadas, de acordo com a seguinte equação:

$$G = (m_{majoritárias} - m_{minoritárias}) \times \beta \quad (13),$$

na qual $m_{majoritárias}$ e $m_{minoritárias}$ representam o número de amostras contidas na classe majoritária e minoritária, respectivamente. Para cada amostra pertencente a classe minoritária calcula-se sua dificuldade de classificação r_i , através da observação do número de vizinhos pertencentes a classe majoritária (Δ) e do número de vizinhos mais próximos (k).

$$r_i = \Delta/k \quad (14)$$

Os valores de r_i são normalizados para obter a função de distribuição de densidade \hat{r}_i .

$$\hat{r}_i = \frac{r_i}{\sum_{i=1}^{m_{minoritárias}} r_i} \quad (15)$$

As amostras sintéticas são geradas com base na região de menor densidade da classe minoritária. A amostra sintética s_i gerada é definida a partir da amostra original da classe minoritária x_i e de uma amostra vizinha x_{iz} , considerando um fator λ associado a variação introduzida na amostra sintética.

$$s_i = x_i + (x_i - x_{iz}) \times \lambda \quad (16)$$

O SMOTE também é um método de sobreamostragem que se baseia na interpolação linear para gerar amostras sintéticas para compor a classe minoritária. Para isso, um grupo de amostras da classe minoritária é escolhido aleatoriamente, a

fim de identificar o k-vizinho mais próximo de cada amostra desse grupo (CHAWLA *et al.*, 2002).

Em seguida, a distância entre a amostra x_i e seu k-vizinho x_{iz} é multiplicada por um fator escolhido no intervalo entre 0 e 1. O valor resultante é somado à amostra original para obter a amostra sintética s_i , de forma semelhante à descrita pela Equação 16 utilizada pelo ADASYN (DEY *et al.*, 2023).

A principal diferença entre o ADASYN e o SMOTE é que o ADASYN gera amostras sintéticas a partir de amostras contidas em ambas as classes, enquanto o SMOTE considera apenas amostras da classe minoritária. Além disso, o ADASYN atribui pesos às amostras selecionadas, dando maior ênfase àquelas que são mais difíceis de classificar. Isso faz com que o ADASYN crie amostras sintéticas que irão compor a fronteira entre as classes, enquanto que o SMOTE não realiza essa distinção.

3.5.10 Validação de modelos de classificação multivariada

Para avaliar o desempenho dos modelos de classificação multivariados, são consideradas suas taxas de acerto e de erro. Em um modelo binário, a classe alvo é chamada de positiva, e a classe não alvo, de negativa. Os valores de verdadeiro positivo (VP) e verdadeiro negativo (VN) representam, respectivamente, o número de amostras classificadas corretamente para a classe alvo e a classe não alvo. Da mesma forma, os valores de falsos positivos (FP) e falsos negativos (FN) correspondem às amostras classificadas erroneamente em ambas as classes.

A partir desses valores, são calculadas a sensibilidade (SEN), a especificidade (ESP) e a eficiência (EFC) do modelo. A sensibilidade descreve a capacidade do modelo de classificar corretamente as amostras da classe positiva (classe alvo), enquanto a especificidade mensura a correta classificação das amostras da classe negativa (não alvo). A eficiência corresponde à média entre os valores de sensibilidade e especificidade, avaliando a taxa de acerto total do modelo.

A precisão dos modelos pode ser avaliada por meio dos termos de acordância (ACO) e concordância (CON). A acordância avalia a probabilidade de que amostras idênticas sejam classificadas da mesma forma quando analisadas sob as mesmas

condições experimentais. Já a concordância mede a probabilidade de que duas amostras iguais sejam classificadas na mesma classe, mesmo com variação nas condições experimentais, como dia, analista ou laboratório. Valores de ACO e CON próximos a 1 indicam modelos com alta taxa de repetibilidade e reprodutibilidade. Nas equações utilizadas para os cálculos de acordância e concordância, n representa o número de replicatas analisadas, k o número de classificações corretas para cada lote analítico, e b o número de lotes analisados (GONDIM *et al.*, 2014).

O coeficiente de Matthews (MCC) é comumente utilizado em estudos de aprendizagem de máquina para avaliar o desempenho de métodos de classificação binária. Valores de MCC iguais a -1 indicam modelos com classificação inversa, ou seja, todas as amostras previstas erroneamente. Já valores de MCC iguais a 1 indicam modelos com desempenho perfeito, enquanto que MCC igual a 0 corresponde a modelos com classificação aleatória (CHICCO *et al.*, 2023). O MCC, assim como a eficiência, é uma figura de mérito global. No entanto, apresenta como vantagem a consideração do desbalanceamento de classe na avaliação do desempenho do modelo.

O Quadro 2 apresenta as equações utilizadas para o cálculo das figuras de mérito empregadas neste trabalho.

Quadro 2. Equações usadas no cálculo das figuras de mérito utilizadas na validação dos modelos supervisionados.

Figuras de mérito	Equação
Sensibilidade	$SEN (\%) = \left(\frac{VP}{VP + FN} \right) \times 100$
Especificidade	$ESP (\%) = \left(\frac{VN}{VN + FP} \right) \times 100$
Eficiência	$EFC (\%) = \left(\frac{SEN + ESP}{2} \right) \times 100$
Acordância	$ACO (\%) = \frac{\{k(k - 1) + (n - k)(n - k - 1)\}}{n(n - 1)} \times 100$
Concordância	$CON (\%) = \frac{2[k(k - nb)] + nb(nb - 1) - ACO[nb(n - 1)]}{n^2b(b - 1)} \times 100$

Coeficiente de Matthews	$MCC (\%) = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \times 100$

3.6 Estudos da estabilidade oxidativa do biodiesel empregando métodos espectroscópicos e ferramentas quimiométricas

A estabilidade oxidativa do biodiesel é uma característica crucial para garantir a sua qualidade e desempenho nas misturas BX, além de ser um fator determinante para sua vida útil. Nos últimos anos, diversos estudos têm explorado o uso de métodos espectroscópicos associados a ferramentas quimiométricas para monitorar e prever a estabilidade oxidativa do biodiesel de forma rápida e eficiente. Esses estudos abrangem desde métodos quimiométricos de classificação, que empregam métodos como a Análise Discriminante Linear (LDA) ou Floresta Aleatória para separar amostras de biodiesel em diferentes categorias, até métodos de regressão, como os Mínimos Quadrados Parciais (PLS), que são empregados para prever o tempo de estabilidade oxidativa.

Além disso, abordagens como a análise de componentes principais (PCA), aplicada para reduzir a dimensionalidade dos dados, e a resolução multivariada de curvas (MCR-ALS), utilizada para modelar processos complexos, como a cinética de oxidação, também têm sido empregadas para melhorar a precisão e a interpretabilidade dos modelos. Essas técnicas permitem uma análise rápida e não destrutiva, facilitando o monitoramento da qualidade do biodiesel de maneira eficiente e acessível.

Velvarská e colaboradores (2019) usaram a espectroscopia no infravermelho próximo (NIR, do inglês *Near-Infrared*), associada a ferramentas quimiométricas, para determinar a estabilidade oxidativa das misturas do biodiesel com o óleo diesel identificadas como B7, B20, B30 e B100. Primeiramente, foi construído um modelo de calibração utilizando os espectros NIR de 133 amostras, na faixa de espectral de 10000 a 6450 cm^{-1} , com estabilidade oxidativa conhecida. Para isso, foi utilizado o método de calibração multivariada dos Mínimos Quadrados Parciais (PLS, do inglês *Partial Least Squares*) com sete variáveis latentes. O modelo construído foi validado por um conjunto teste de 30 amostras, obtendo um coeficiente de correlação (R^2) de

99,08%, desvio padrão relativo (RSD) de 8,59% e raiz quadrada do erro médio quadrático da calibração (RMSEC) de 8,73 min (VELVARSKÁ *et al.*, 2019).

Sousa e colaboradores (2019) investigaram o uso da espectroscopia no infravermelho médio com transformada de Fourier (FTIR, do inglês *Fourier-transform infrared spectroscopy*), associada a ferramentas quimiométricas, para a determinação da estabilidade oxidativa do óleo de babaçu. Foram analisadas 30 amostras previamente oxidadas a 100 °C durante 5 dias, e os índices de acidez e peróxido foram determinados. Os espectros obtidos na faixa de 4000 a 650 cm^{-1} foram analisados por análise de componentes principais (PCA) para identificação das variáveis mais significativas. A análise dos gráficos de pesos da PCA indicou que as bandas observadas em 1742, 1163 e 1117 cm^{-1} são as mais importantes na discriminação dos grupos observados, correspondendo às regiões de estiramento C=O, estiramento C-O de ésteres e estiramento C-O de éter, respectivamente. Um modelo de calibração multivariada empregando o PLS foi construído para previsão dos índices de peróxido e de acidez das amostras, apresentando valores de raiz quadrada do erro médio quadrático da previsão (RMSEP) de 7,85 e 0,17 (meq O_2/kg), desvio residual da previsão (RPD) de 2,16 e 3,60, e coeficiente de determinação (R^2) de 0,94 e 0,95, para os índices de peróxido e acidez, respectivamente (SOUSA, *et al.*, 2019).

Meira e colaboradores (2012) estudaram a aplicação da espectrofluorimetria combinada à análise multivariada para a determinação da viscosidade, estabilidade oxidativa e massa específica do biodiesel puro, suas misturas com o óleo diesel e amostras de óleo comercial. Os espectros de emissão foram obtidos na faixa de 230 a 800 nm. Os modelos de calibração foram construídos empregando o PLS, correlacionando os espectros de fluorescência com os valores da estabilidade oxidativa, determinados pelo método Rancimat. Foram obtidos os valores de coeficiente de correlação (R^2) de 0,98557 para o óleo de soja e 0,98423 para o biodiesel de soja. As amostras do conjunto teste apresentaram erro de previsão máximo de 0,6 horas (MEIRA *et al.*, 2012).

Canha e colaboradores (2012) utilizaram o NIR combinado a modelos PLS para avaliar a estabilidade oxidativa de 199 amostras de biodiesel, com e sem a adição de antioxidantes. O modelo de calibração PLS obtido na avaliação da estabilidade oxidativa do biodiesel sem antioxidantes apresentou erro de previsão de 0,6 horas e R^2 de 95%, para uma faixa de calibração de 0,66 a 17,75 horas. O modelo PLS obtido

para as amostras de biodiesel contendo antioxidantes apresentou erro de previsão de 1,28 horas e R^2 de 91%, para uma faixa de calibração de 0,43 a 22,10 horas (CANHA *et al.*, 2012).

Souza e colaboradores (2016) propuseram o uso da espectroscopia Raman e no infravermelho médio (FTIR) para avaliar a hidrólise e a estabilidade oxidativa do biodiesel, do diesel e das suas misturas. Nos espectros FTIR a banda próxima a 1746 cm^{-1} (estiramento C=O) foi utilizada para caracterizar o estado de degradação do biocombustível. Já no Raman, devido à baixa intensidade da banda relativa a carbonila, foi utilizada as bandas de estiramento C=C (1659 cm^{-1}) e C-H de alcenos (3014 cm^{-1}) para caracterizar o nível de oxidação da molécula. Os modelos de regressão construídos apresentaram coeficiente de correlação R^2 de 0,9901 para o FTIR e 0,9944 e 0,9868 para as bandas 1659 cm^{-1} e 3014 cm^{-1} observadas nos espectros Raman (SOUZA *et al.*, 2016).

Gonçalves (2022) avaliou o uso das diferentes técnicas de espectroscopia vibracional (infravermelho médio, infravermelho próximo e Raman) e métodos de classificação para o estudo da estabilidade oxidativa de amostras de biodiesel puro, conformes e não conformes. Para a construção dos modelos de classificação, foram empregados os métodos de análise discriminante linear (LDA) e floresta aleatória. Os resultados obtidos mostraram 97% de eficiência e coeficiente de correlação de Matthews (MCC) de 0,920 para o modelo floresta aleatória construído a partir dos espectros NIR, 98,6% de eficiência e MCC de 0,972 para o modelo de floresta aleatória construído a partir dos espectros Raman e 90,2% de eficiência e MCC de 0,446 para o modelo floresta aleatória construído a partir dos espectros do infravermelho médio (MIR, do inglês *Mid-Infrared*). Nos modelos LDA, a eficiência foi de 93,3% e MCC de 0,875 para o NIR, 92,7% de eficiência e MCC de 0,894 para o Raman, e 92,7% de eficiência e MCC de 0,539 para o MIR. (GONÇALVES, 2022). A espectroscopia Raman e o método de classificação floresta aleatória forneceram os melhores resultados.

Valeriano e colaboradores (2024) estudaram a cinética de oxidação do biodiesel por meio de medidas de espectroscopia Raman associadas ao método de Resolução Multivariada de Curvas com Mínimos Quadrados Alternantes (MCR-ALS) e a Análise de Fatores Evolutivos (EFA). O modelo construído demonstrou que os modos

vibracionais de estiramento C=C e deformação CH₂ e CH₃ são os principais envolvidos no processo de degradação do biodiesel (VALERIANO *et al.*, 2024).

A revisão bibliográfica realizada mostrou que o uso de técnicas espectroscópicas associadas à análise multivariada pode ser muito interessante para otimizar a metodologia de controle de qualidade do biodiesel. A maior parte dos estudos encontrados utiliza amostras preparadas em laboratório, o que limita a aplicabilidade dos resultados, uma vez que essas amostras podem não refletir as variações encontradas em amostras comerciais. Em contraste, neste estudo, foram utilizadas amostras comerciais de biodiesel B100, provenientes de diferentes produtores, a fim de refletir uma variedade de características e condições reais de mercado. Tendo em vista os resultados obtidos por Gonçalves (2022), o presente trabalho utilizou a espectroscopia Raman, combinada com ferramentas quimiométricas, para construção de modelos de classificação que avaliem a estabilidade oxidativa do biodiesel quanto à sua conformidade com a legislação vigente. A ideia principal é possibilitar que tal parâmetro seja medido mais rapidamente e no próprio local de coleta, dando prioridade ao uso de equipamentos portáteis ou compactos.

4. METODOLOGIA

4.1 Amostras

As amostras de biodiesel analisadas nesse trabalho foram disponibilizadas pelo Laboratório de Ensaio de Combustíveis da UFMG (LEC-UFMG). Trata-se de amostras comerciais de B100 que foram enviadas ao LEC no período entre 2022 e 2024. A estabilidade oxidativa dessas amostras foi determinada pelo método Rancimat (Resolução ANP nº 798/2019). As amostras que foram aprovadas no método Rancimat foram definidas como “conforme” (tempo de indução maior ou igual a 13h) e as reprovadas como “não conforme” (tempo de indução menor que 13h). Foram avaliados dois conjuntos de amostras, um primeiro conjunto contendo um total de 90 amostras, sendo 79 amostras da classe “conforme” e 11 amostras da classe “não conforme”, e um segundo conjunto contendo um total de 162 amostras, sendo 147 amostras da classe “conforme” e 14 amostras da classe “não conforme”.

4.2 Análise espectroscópica

As amostras de biodiesel foram analisadas por espectroscopia Raman. Foram utilizados dois tipos de equipamentos Raman, um espectrômetro portátil com tecnologia SORS (Equipamento A) e um espectrômetro de bancada compacto (Equipamento B).

O equipamento portátil utilizado foi o Agilent Resolve Raman (Agilent®, Califórnia, EUA), com laser de 830 nm, potência ajustável de até 475 mW, resolução de 10 a 12 cm^{-1} , faixa espectral de 200 a 2000 cm^{-1} e tecnologia SORS com identificação através de barreiras. Nesse equipamento, o biodiesel foi analisado diretamente no microtubo e sem passar por nenhum tipo de preparo de amostras, conforme mostrado na Figura 20.



Figura 20. Análise das amostras no Agilent Resolve Raman® (Equipamento A).

O equipamento Raman de bancada compacto utilizado foi o modelo Cora 5001 (Anton Paar®, Estíria, Áustria), mostrado na Figura 21, com laser de 1064 nm, potência de 450 mW, resolução de 12 a 17 cm^{-1} , faixa espectral de 100 a 2300 cm^{-1} . Além disso, conta com a tecnologia Autofocus, que permite o ajuste automático do foco da amostra. O equipamento também oferece a possibilidade de análise empregando o laser em 532 nm. No entanto, após testes, foi observada a presença de fluorescência ao utilizar este laser, o que levou à escolha do laser de 1064 nm para o estudo. Nesse equipamento as amostras de biodiesel também foram analisadas diretamente no microtubo.



Figura 21. Análise das amostras no Cora 5001 da Anton Paar® (Equipamento B).

A Tabela 2 mostra as especificações técnicas dos dois equipamentos Raman utilizados.

Tabela 2. Especificações técnicas dos equipamentos Raman utilizados.

	Equipamento A	Equipamento B
Laser (nm)	830	1064 e 532
Potência (mW)	máximo de 475	450
Faixa espectral (cm⁻¹)	200-2000	100-2300
Dimensões (mm)	155x290x73	384x168x355
Peso (kg)	2,2	9,8
Suporte de amostras	Bandeja de <i>vials</i>	Bandeja de <i>vials</i> e lâminas
Duração da bateria (h)	4	-
Exportação de dados	USB (.csv)	USB (.csv, .txt, .png)
Tecnologia	SORS	Autofocus

4.3 Tratamento dos dados obtidos

O desenvolvimento do trabalho aconteceu em duas etapas. Na primeira etapa, foi analisado o conjunto de 90 amostras usando os equipamentos A e B. O desempenho dos espectrômetros foi comparado por meio dos modelos de classificação SIMCA, PLS-DA e floresta aleatória. Na segunda etapa, o conjunto de 162 amostras foi analisado apenas no equipamento B e empregou o método da floresta aleatória. Nesta etapa, foram comparados dois métodos de reamostragem, ADASYN e SMOTE, e a precisão dos modelos foi avaliada por meio das figuras de mérito acordância e concordância. As amostras classificadas como “conformes” (147 amostras) foram, então, reavaliadas por meio de um novo modelo de floresta aleatória, com o objetivo de verificar a classificação das amostras de biodiesel conformes em dois grupos: amostras com estabilidade menor que 17 horas (< 17 horas) e amostras com estabilidade maior que 17 horas (> 17 horas).

Em ambas as etapas, os dados obtidos foram exportados em arquivos txt, organizados em planilhas Excel e limitados na faixa de 350 a 1800 cm⁻¹. Os espectros Raman foram pré-processados empregando primeira derivada com alisamento de *Savitzky-Golay* (janela de 7 pontos e polinômio de segunda ordem), seguido dos dados centrados na média. Após o pré-processamento, foi feita a análise exploratória

dos dados através da PCA utilizando a interface gráfica do *PLS Toolbox* (Eigenvector Res. Inc.) no MATLAB® versão R2010b.

Para os modelos de classificação, PLS-DA e floresta aleatória, os conjuntos de dados foram previamente sobreamostrados pelos métodos ADASYN e SMOTE, para o balanceamento da classe “não conforme” (classe minoritária). As amostras foram divididas em um conjunto de treinamento contendo 70% das amostras e um conjunto de teste com 30% das amostras, utilizando o algoritmo de *Kennard-Stone* (KENNARD, 1969). Os modelos foram construídos no MATLAB® (versões R2010b e R2024a) utilizando os pacotes *PLS-Toolbox* para o SIMCA e o PLS-DA e *Statistics and Machine Learning Toolbox* para a floresta aleatória. No PLS-DA, foi utilizado o método de validação cruzada das venezianas (*venetion blinds*), com o número de 10 splits e a seleção de variáveis latentes foi feita com base no menor erro de classificação da validação cruzada. A otimização dos parâmetros do modelo de floresta aleatória foi feita de acordo com Gonçalves (2022)

A avaliação do desempenho dos modelos construídos foi feita por meio dos cálculos das figuras de mérito discutidas na seção 3.5.10. Para os cálculos de acordância e concordância dos dados da segunda etapa do trabalho, foram realizadas medidas no Equipamento B com seis replicatas de amostras de biodiesel com diferentes níveis de estabilidade oxidativa: 8, 11, 12, 13, 16 e 23 horas. Devido à quantidade de biodiesel disponível, a leitura dessas amostras foi realizada em lâminas de vidro, diferentemente da amostragem utilizada na construção dos modelos de classificação. As medidas de ACO e CON foram realizadas no mesmo equipamento e com as mesmas amostras, variando o dia e o analista responsável para as medidas de CON.

5. RESULTADOS E DISCUSSÕES

5.1 Desempenho dos espectrômetros Raman na avaliação da estabilidade oxidativa do biodiesel B100

Um conjunto contendo 90 amostras de biodiesel B100 (79 amostras “conforme” e 11 amostras “não conforme”) foi analisado utilizando dois espectrômetros Raman, um espectrômetro Raman portátil (Equipamento A) e um espectrômetro Raman compacto (Equipamento B).

A Figura 22 apresenta os espectros Raman de uma amostra de biodiesel B100 conforme (verde) e não conforme (vermelho). A análise visual dos espectros Raman não foi suficiente para diferenciar essas duas amostras, já que o perfil dos dois espectros é muito semelhante. Nota-se que as bandas principais que elucidam o estiramento C=O (1665 - 1750 cm^{-1}), o estiramento C=C (1600 - 1680 cm^{-1}), as deformações de CH₂ e CH₃ (1400 - 1500 cm^{-1}), a deformação C-H do tipo tesoura (1245 - 1280 cm^{-1}) e o estiramento C-O (800 - 900 cm^{-1}), ocorrem tanto para a amostra conforme (verde) quanto para a amostra não conforme (vermelho).

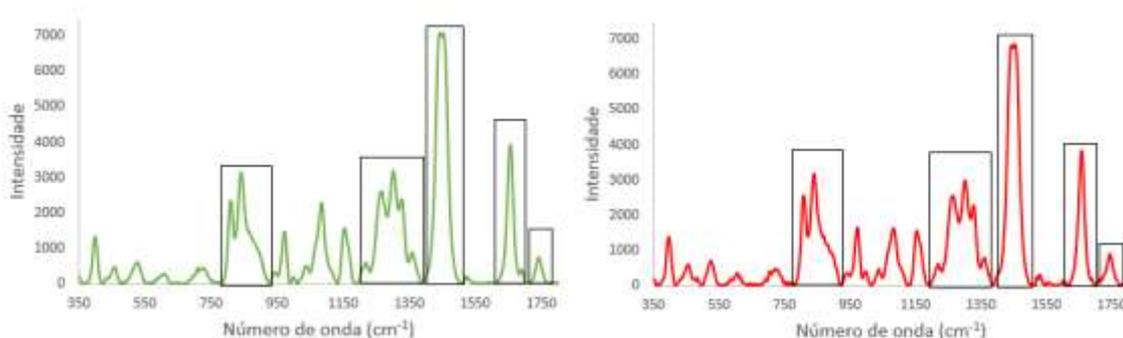


Figura 22. Espectro Raman de uma amostra de biodiesel conforme (verde) e não conforme (vermelho).

Dessa forma, os conjuntos de dados obtidos nos dois equipamentos foram avaliados por ferramentas quimiométricas, PCA, SIMCA, PLS-DA e floresta aleatória, por meio da análise exploratória e da construção de modelos de classificação para previsão da conformidade do biodiesel B100 quanto a estabilidade oxidativa.

5.1.1 Análise exploratória do conjunto de dados obtido pelo Equipamento A

A Figura 23 mostra os espectros brutos, limitados na faixa de 350 a 1800 cm^{-1} , das 90 amostras analisadas no equipamento A.

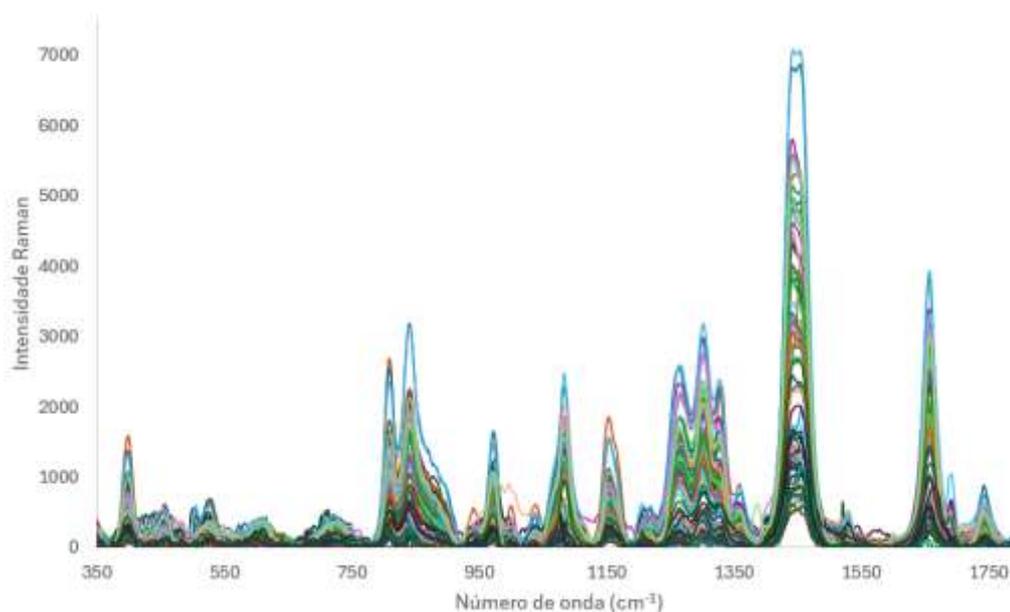


Figura 23. Espectros Raman brutos obtidos pelo equipamento A.

A análise exploratória do conjunto de dados espectrais Raman foi feita através da PCA. Para isso, o conjunto de dados foi pré-processado empregando primeira derivada com alisamento de *Savitzky-Golay* (janela de 7 pontos e polinômio de segunda ordem) seguido dos dados centrados na média. Foram selecionadas três componentes principais que explicam 84,64% da variância total dos dados.

A Figura 24 apresenta o gráfico de escores da PC1 *versus* a PC2 no qual não foi observada uma separação das amostras de biodiesel B100 quanto à sua estabilidade oxidativa. Nesse gráfico é possível observar o agrupamento de algumas amostras no lado negativo e no lado positivo da PC1. Esses dois grupos são compostos por amostras conformes e não conformes, que possuem características semelhantes em relação às variáveis que descrevem a PC1. Ao observar o gráfico de pesos da Figura 25 é possível inferir que o lado positivo da PC1 é caracterizado, principalmente, pela presença de uma banda próxima a 1490 cm^{-1} . Essa banda pode ser associada às deformações CH_2 e CH_3 da molécula e reflete o seu grau de saturação, o que pode estar relacionado à fonte de gordura utilizada na produção do biodiesel. Gorduras animais, como o sebo bovino, apresentam maior quantidade de ácidos graxos saturados, enquanto que os óleos vegetais são majoritariamente formados por ácidos graxos insaturados. Logo, é possível inferir que as amostras de biodiesel que

compõem o grupo localizado no lado positivo de PC1 podem ter origem na reação de transesterificação com gordura animal. Além disso, observa-se na Figura 24 que as amostras localizadas no lado negativo da PC1 estão mais próximas umas das outras, evidenciando a homogeneidade dessas amostras quanto à fonte de gordura utilizada na sua produção.

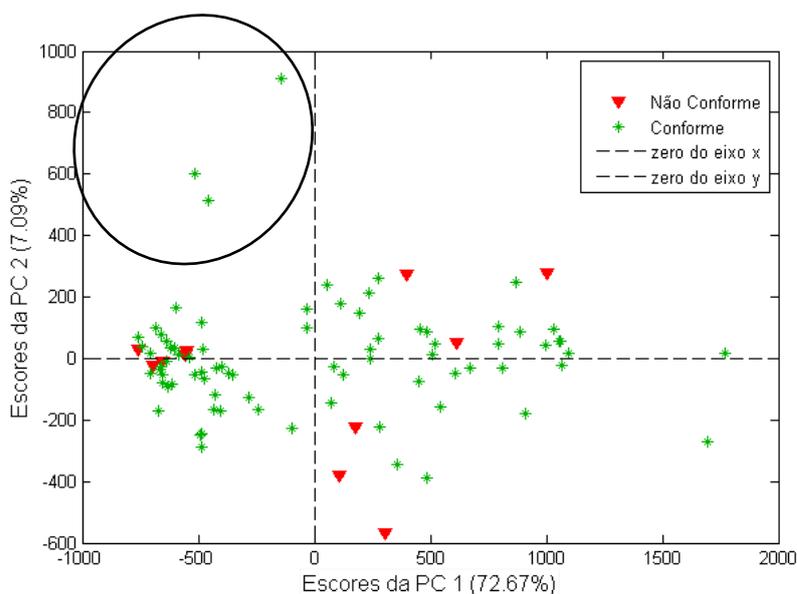


Figura 24. Gráfico de Escores: PC1 versus PC2.

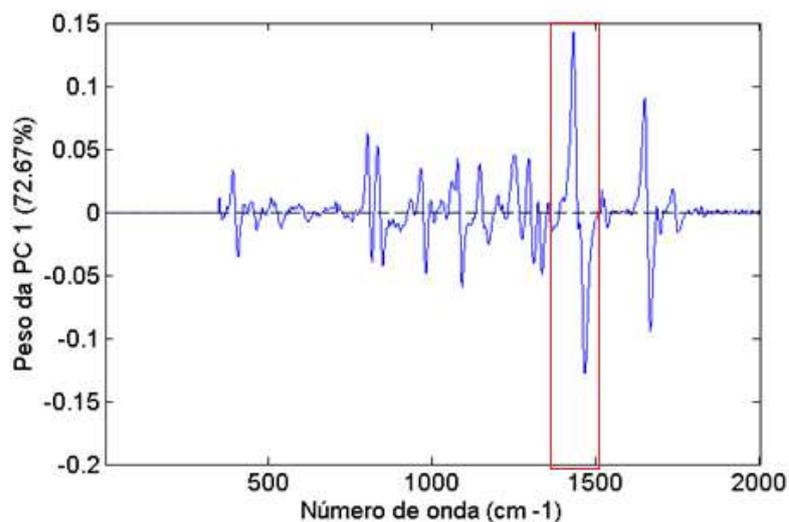


Figura 25. Gráfico de Pesos da PC1.

A Figura 24 também mostra três amostras conformes localizadas no extremo positivo da PC2. Ao avaliar o gráfico de pesos da PC2 (Figura 26) é possível inferir que essas três amostras são caracterizadas, principalmente, pela presença de uma

banda próxima a 1100 cm^{-1} . Avaliando os espectros Raman individuais dessas amostras (APÊNDICE 1A), é possível visualizar a presença de uma banda intensa e estreita próxima a 1100 cm^{-1} . Essa banda pode estar associada à presença de antioxidantes sintéticos tipicamente adicionados ao biodiesel, como o butilhidroxitolueno (BTH) e o terc-butil-hidroquinona (TBHQ).

O BTH e o TBHQ são compostos fenólicos que apresentam uma banda no espectro Raman próxima a 1100 cm^{-1} relacionada a deformação C=C do anel benzênico. A introdução de substituintes no anel benzênico leva a uma maior complexidade do espectro vibracional da molécula, resultando no deslocamento dessa banda para números de onda maiores (PAVIA *et al.*, 2001). Dessa forma, acredita-se que a PC2 separou as amostras de B100 de acordo com a presença de antioxidantes, indicando que as três amostras da classe conforme localizadas no extremo positivo da PC2 contenham maior concentração desses componentes.

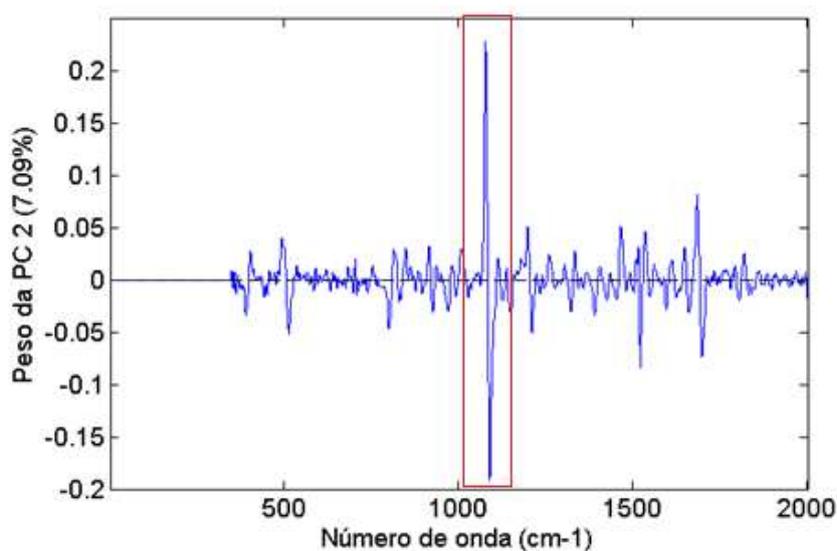


Figura 26. Gráfico de Pesos da PC2.

O gráfico de escores da PC1 *versus* PC3 (Figura 27) apresenta uma amostra não conforme localizada no extremo positivo da PC3. Analisando o gráfico de pesos da PC3, apresentado na Figura 28, é possível inferir que essa amostra é caracterizada, principalmente, pela presença de uma banda próxima a 800 cm^{-1} , conforme é destacado no seu espectro Raman mostrado na Figura 2A (APÊNDICE B). As bandas localizadas na faixa de $(800 - 900\text{ cm}^{-1})$ representam o estiramento C-O da estrutura. Logo, é possível concluir que a amostra da classe não conforme

destacada na PC3, possui uma banda de maior intensidade na faixa de 800 a 900 cm^{-1} , a qual pode estar associada ao estiramento C-O. Essa banda é típica da presença de álcoois, ésteres ou ácidos carboxílicos, compostos que podem ser formados durante o processo de oxidação do biodiesel.

Apesar dessa observação, as outras amostras da classe “não conforme” não tiveram o mesmo comportamento. Isso sugere que outros fatores podem influenciar a intensidade dessa banda, tornando apenas sua presença insuficiente para caracterizar a oxidação do biodiesel.

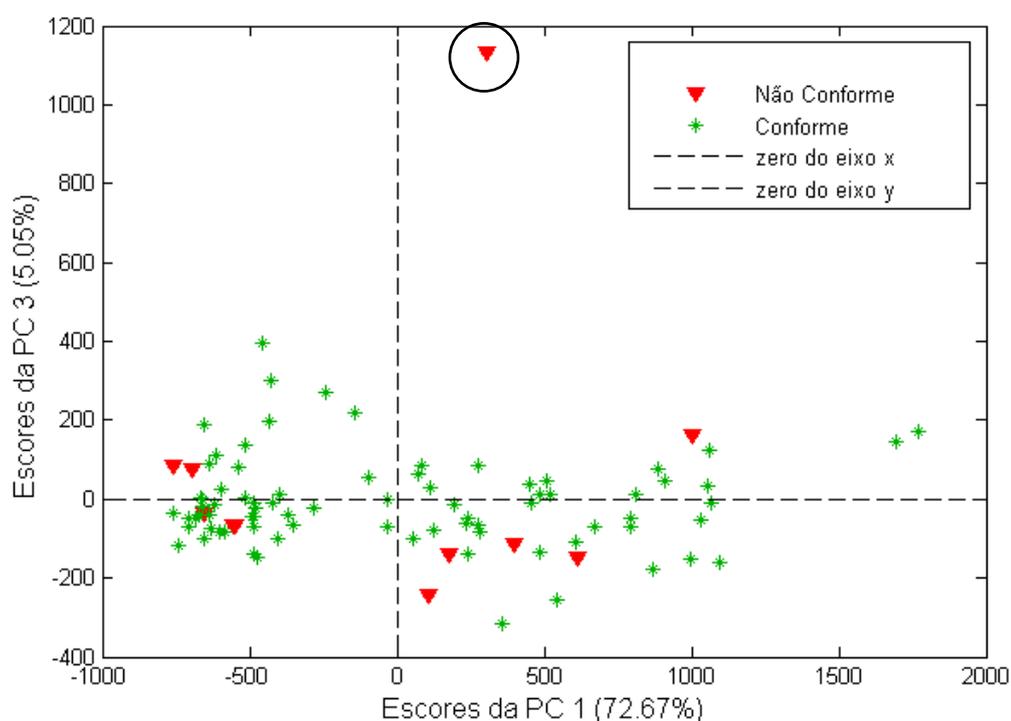


Figura 27. Gráfico de Escores: PC1 versus PC3.

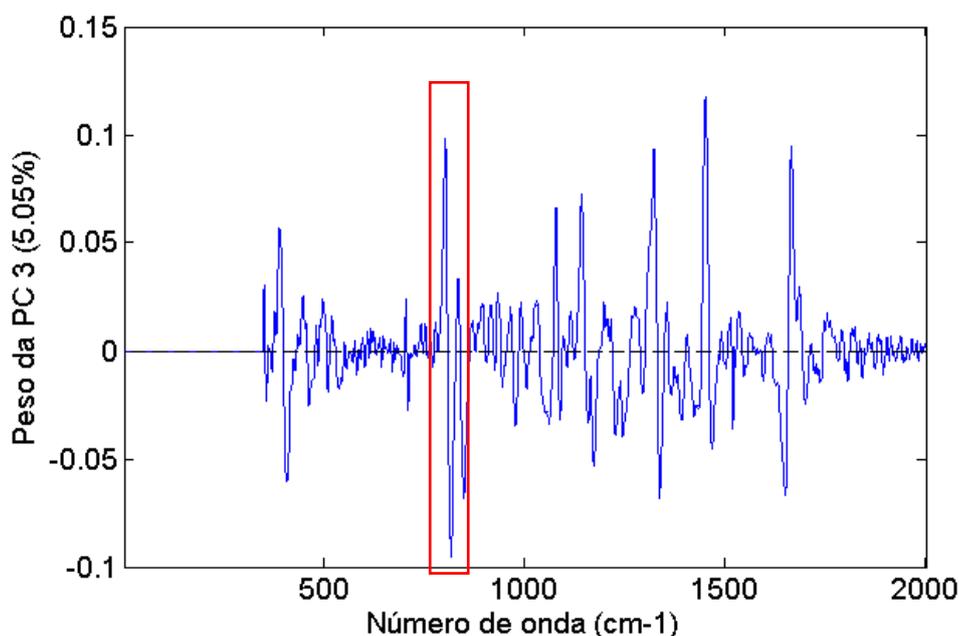


Figura 28. Gráfico de Pesos da PC3.

Os gráficos de escores obtidos para as três componentes principais mostraram que na PCA não houve uma tendência de agrupamento das amostras em relação à sua conformidade ou não conformidade no teste de estabilidade oxidativa. A PC1 separou as amostras em dois grupos, de acordo com a fonte de gordura utilizada na síntese do biodiesel. A PC2 fez a separação das amostras quanto à presença de antioxidantes, evidenciando três amostras no extremo positivo da PC2 com indicativo de maior concentração de tais compostos. A PC3 evidenciou a separação de apenas uma amostra não conforme, localizada no extremo positivo do gráfico de escores. Essa amostra foi caracterizada pelo modo vibracional de estiramento da ligação C-O (800 a 900 cm^{-1}), o que pode estar associado à presença de produtos da oxidação do biodiesel, como peróxidos e álcoois.

5.1.2 Análise exploratória do conjunto de dados obtido pelo Equipamento B

A Figura 29 mostra os espectros brutos, limitados na faixa de 350 a 1800 cm^{-1} , das 90 amostras analisadas no equipamento B.

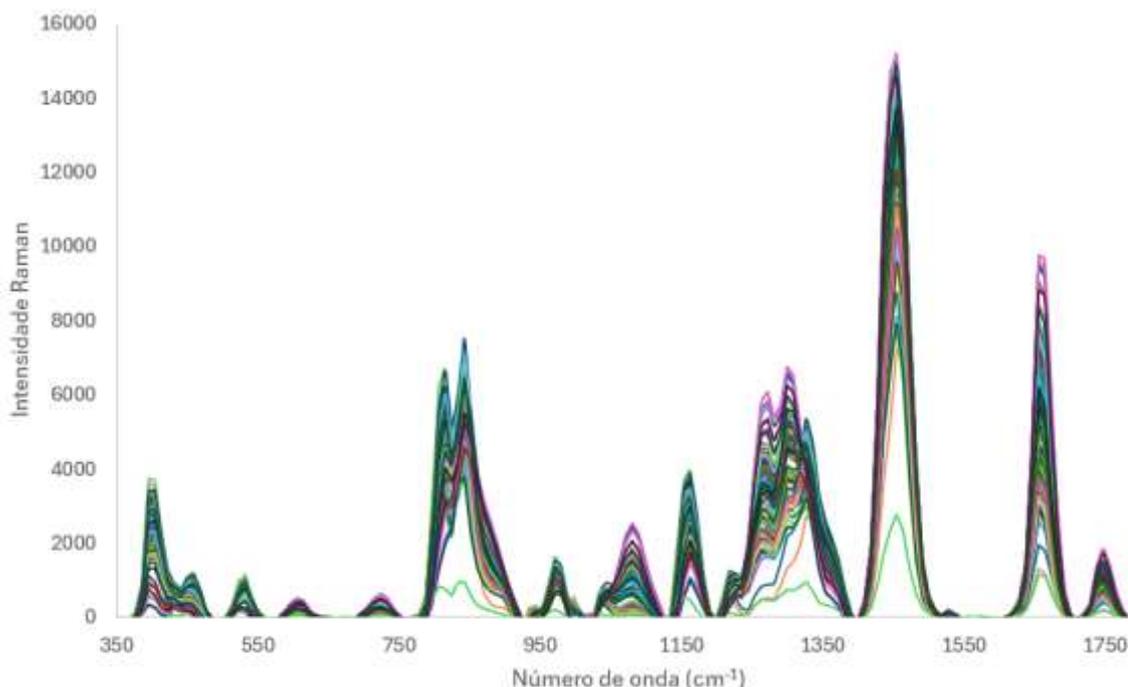


Figura 29. Espectros Raman brutos obtidos pelo equipamento B.

O conjunto de dados espectrais das amostras de biodiesel analisadas pelo equipamento B foi pré-processado empregando primeira derivada com alisamento de *Savitzky-Golay* (janela de 7 pontos e polinômio de segunda ordem) seguido dos dados centrados na média. Foram selecionadas três componentes principais que explicam 97,63% da variância total dos dados.

A Figura 30 apresenta o gráfico de escores da PC1 *versus* a PC2 no qual não foi observada uma separação das amostras de biodiesel B100 em relação à sua estabilidade oxidativa. Nesse gráfico é possível observar o agrupamento de algumas amostras no lado negativo e no lado positivo da PC1. Ao observar o gráfico de pesos da Figura 31, pode-se concluir que o lado positivo da PC1 é caracterizado, principalmente, pela presença de uma banda próxima a 1600 cm^{-1} , associada ao estiramento da ligação C=C.

Já o lado negativo da PC1 é caracterizado, principalmente, pela presença de uma banda próxima de 1300 cm^{-1} , relacionada às deformações C-H do tipo tesoura. As ligações C=C e C-H estão relacionadas ao grau de saturação da molécula de biodiesel que, por sua vez, pode indicar a fonte de triglicerídeos utilizada na síntese do B100. Logo, a PC1 separa as amostras de biodiesel B100 quanto ao seu grau de

saturação, indicando que as amostras do lado positivo e negativo podem ser originárias de fontes de gordura distintos.

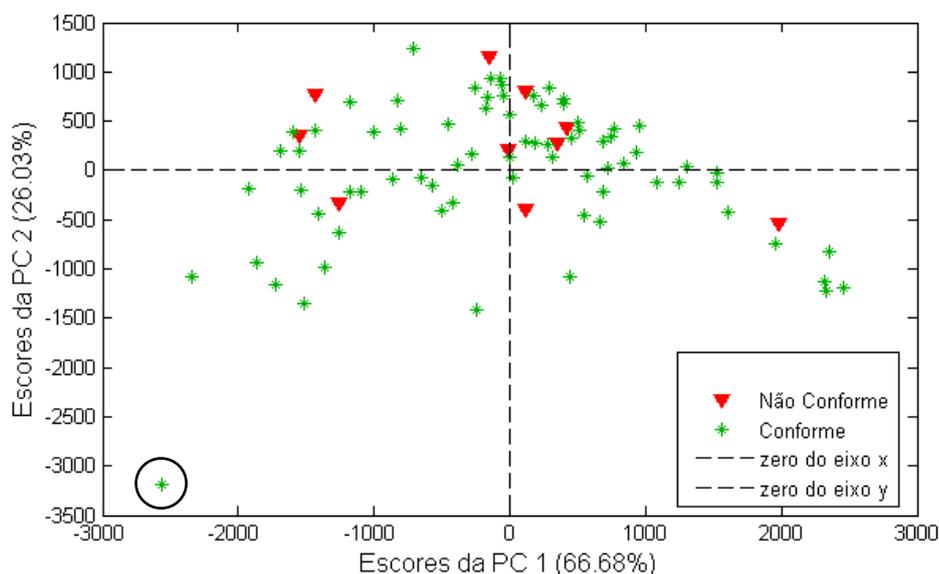


Figura 30. Gráfico de Escores: PC1 versus PC2.

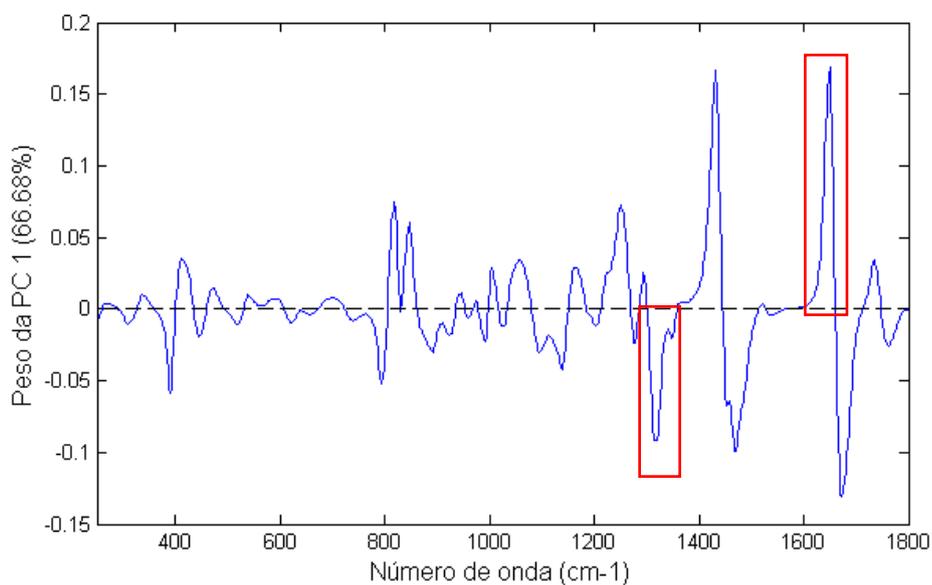


Figura 31. Gráfico de Pesos da PC1.

O gráfico de escores da PC1 *versus* PC2 apresenta uma amostra da classe conforme localizada no extremo negativo da PC2. Observando o gráfico de pesos da PC2, mostrado na Figura 32, observa-se que essa amostra é caracterizada, principalmente, pela presença de uma banda intensa próxima a 1450 cm^{-1} . Essa banda está associada à deformação dos grupos CH_2 e CH_3 , o que pode indicar o grau

de saturação da molécula de biodiesel. Essa observação sugere que a amostra em questão foi sintetizada a partir de uma gordura diferente das demais amostras avaliadas.

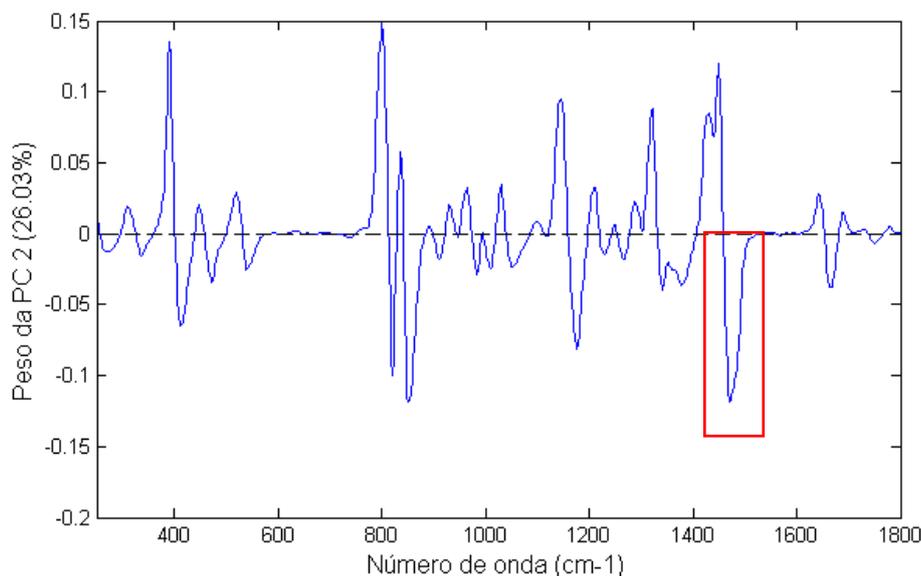


Figura 32. Gráfico de Pesos da PC2.

A Figura 33 mostra o gráfico de escores da PC1 *versus* PC3, onde é observado o mesmo comportamento relatado anteriormente, com a divisão de dois grupos dispersos localizados no lado positivo e negativo da PC3. Esses grupos são formados pela combinação de amostras conformes e não conformes, não evidenciando uma separação referente a estabilidade oxidativa do biodiesel. Avaliando o gráfico de pesos da PC3, Figura 34, observa-se que essa separação em dois grupos na PC3 está relacionada, principalmente à presença de uma banda próxima a 800 cm^{-1} . Essa banda se refere ao estiramento da ligação C-O, o que pode indicar a presença de álcoois, ésteres ou ácidos carboxílicos que foram formados durante a oxidação do biodiesel.

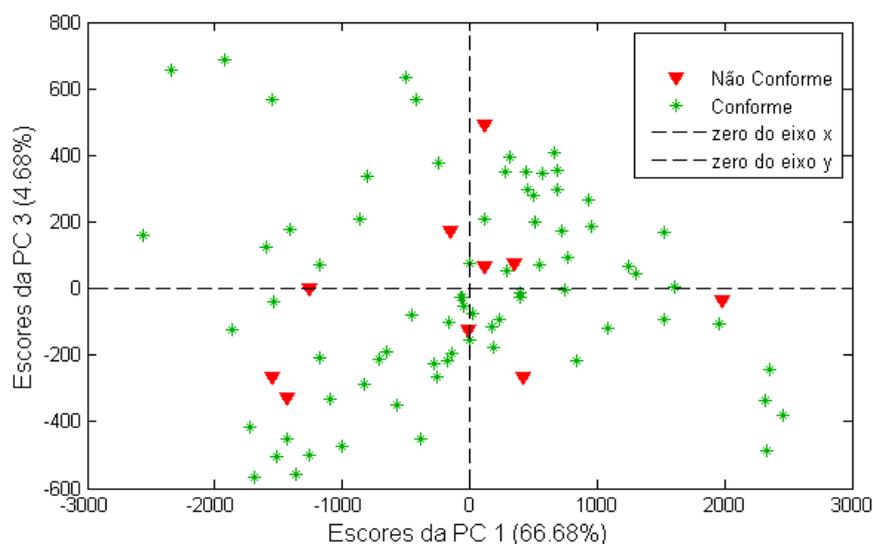


Figura 33. Gráfico de Escores: PC1 versus PC3.

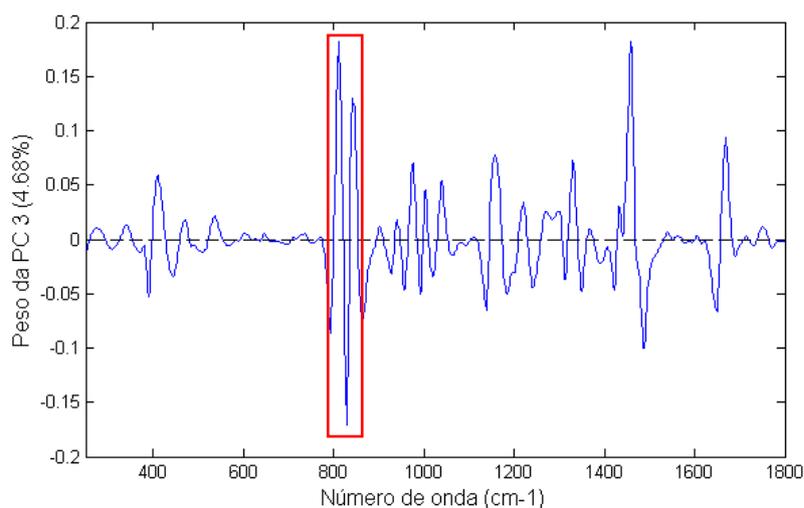


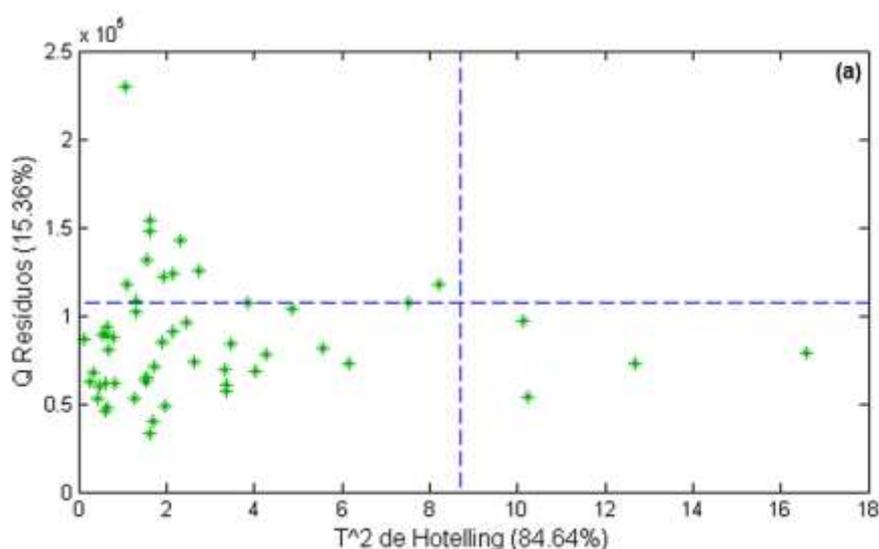
Figura 34. Gráfico de Pesos da PC3.

A partir dos dados obtidos pela PCA não foi possível obter a separação das amostras de biodiesel B100 quanto à sua estabilidade oxidativa. Foi observado que, na PCA dos dados do Equipamento A, as amostras foram separadas com base nas suas características relacionadas à fonte de gordura utilizada na síntese do biodiesel e a presença de antioxidantes. Na PCA dos dados do Equipamento B, resultados semelhantes foram obtidos, uma vez que, as amostras foram separadas levando em consideração o grau de saturação das moléculas e a presença de álcoois, ésteres e ácidos carboxílicos.

5.1.3 Avaliação dos Equipamentos A e B por meio do método de classificação SIMCA

Após realizar a análise exploratória do conjunto de dados, foi construído um modelo de classificação baseado na modelagem de classe SIMCA. O objetivo foi construir um modelo *one class*, considerando que o conjunto de dados era constituído majoritariamente por amostras da classe conforme. A matriz de dados contendo as 79 amostras conformes foi dividida em dois conjuntos: o conjunto de treinamento com 53 amostras (dois terços do número total de amostras conformes) e o conjunto teste com 37 amostras (um terço do número total de amostras conformes e as amostras não conformes). Para fazer a seleção das amostras da classe conforme nos conjuntos de treinamento e de teste, foi utilizado o algoritmo de *Kennard-Stone* (KENNARD, 1969).

A classe de amostras “conforme” foi modelada a partir do conjunto de treinamento. Os espectros foram pré-processados conforme descrito na seção 4.3. Os modelos foram construídos utilizando três componentes principais (PCs). A Figura 35 mostra os gráficos de T^2 de Hotelling *versus* Q resíduos obtidos na análise dos conjuntos de treinamento, utilizando o equipamento A (Figura 35.a) e o equipamento B (Figura 35.b), com os limites de 95% de significância. É possível observar no gráfico que a maior parte das amostras conformes (*) estão dentro dos limites de aceitação estabelecidos para essa classe. Essas amostras apresentaram valores baixos de resíduo e de T^2 . Nos dados obtidos pelo Equipamento A, quatro amostras apresentaram valores de T^2 elevados, o que pode indicar que elas possuem maior influência sobre o modelo construído.



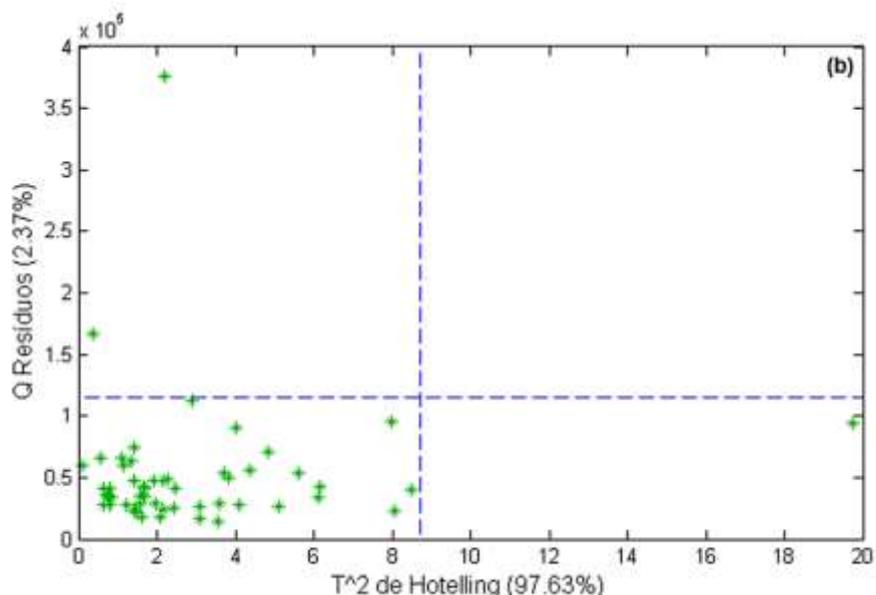


Figura 35. Gráfico de T^2 de Hotelling versus Q Resíduos. a. Equipamento A e b. Equipamento B.

Após avaliar os modelos pelo gráfico de Q resíduos e T^2 de Hotelling, os conjuntos de teste foram utilizados na validação do modelo SIMCA construído. A Figura 36 mostra os gráficos contendo a classificação das amostras de acordo com a classe modelada (conforme). Nota-se que no conjunto de treinamento do Equipamento A (Figura 36.a), apenas seis amostras foram classificadas erroneamente, representando resultados falso negativo. Já no conjunto teste uma amostra conforme não foi classificada corretamente e dez amostras da classe não alvo tiveram resultados falsos positivos (amostras não conformes classificadas como conformes), resultando em um modelo com eficiência de 52,5%. Para os dados obtidos pelo Equipamento B (Figura 36.b), no conjunto de treinamento foi observado que três amostras conformes foram erroneamente classificadas como não conforme (falso negativo). Já no conjunto teste, todas as amostras da classe conforme foram classificadas corretamente e oito amostras da classe não conforme foram classificadas como conforme (resultados falsos positivos), resultando em um modelo com eficiência de 63,5%.

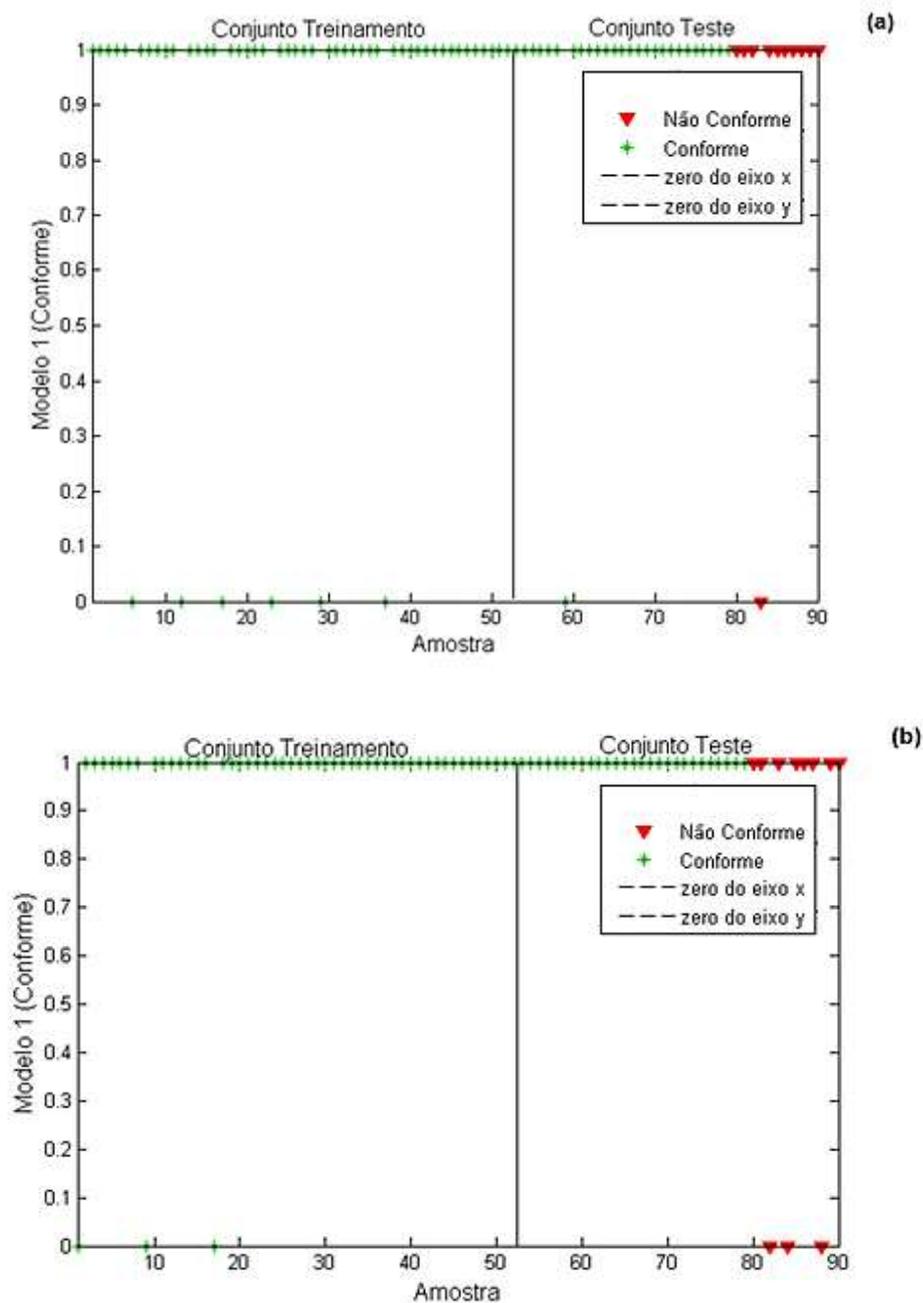


Figura 36. Gráfico de classificação das amostras segundo a classe modelada (conforme). a. Equipamento A e b. Equipamento B.

Os valores de sensibilidade, especificidade e eficiência calculados para os conjuntos de treinamento e teste estão apresentados na Tabela 3. Como o conjunto de treinamento foi formado apenas por amostras alvo (conformes), não houve o cálculo de especificidade e eficiência para o conjunto de treinamento.

Tabela 3. Resultados de sensibilidade, especificidade e eficiência obtidos para os conjuntos de treinamento e teste.

	Conjunto Treinamento	Conjunto Teste		
	SEN	SEN	ESP	EFC
Equipamento A	88%	96%	9%	52,5%
Equipamento B	96%	100%	27%	63,5%

Os resultados obtidos mostraram que a modelagem *one class* SIMCA não é adequada para esse conjunto de dados. Apesar da sensibilidade dos modelos ser alta, a especificidade é muito baixa, reduzindo drasticamente a eficiência dos modelos. Conforme relatado na literatura, conjuntos de amostras com alta variância, nesse caso temos amostras de biodiesel com origem distintas, costumam apresentar modelos *one class* com tendência a alta sensibilidade ao custo de baixa especificidade (BIANCOLILLO *et al.*, 2019). Nunes e colaboradores (2019) utilizaram dados de espectroscopia Raman associados a modelos *one class* para detecção de fraudes em carne bovina e os modelos obtidos apresentaram valores de 94% de sensibilidade e 40% de especificidade (NUNES *et al.*, 2019). Nos casos de conjunto de amostras com alta variância os métodos discriminantes costumam apresentar melhores modelos de predição.

5.1.4 Avaliação dos Equipamentos A e B por meio do método PLS-DA

A Figura 37 mostra os espectros brutos obtidos para as amostras analisadas pelos equipamentos A e B, após o balanceamento de classes realizado pelo método ADASYN. O conjunto de dados final ficou com 158 amostras, 79 amostras conformes e 79 amostras não conformes.

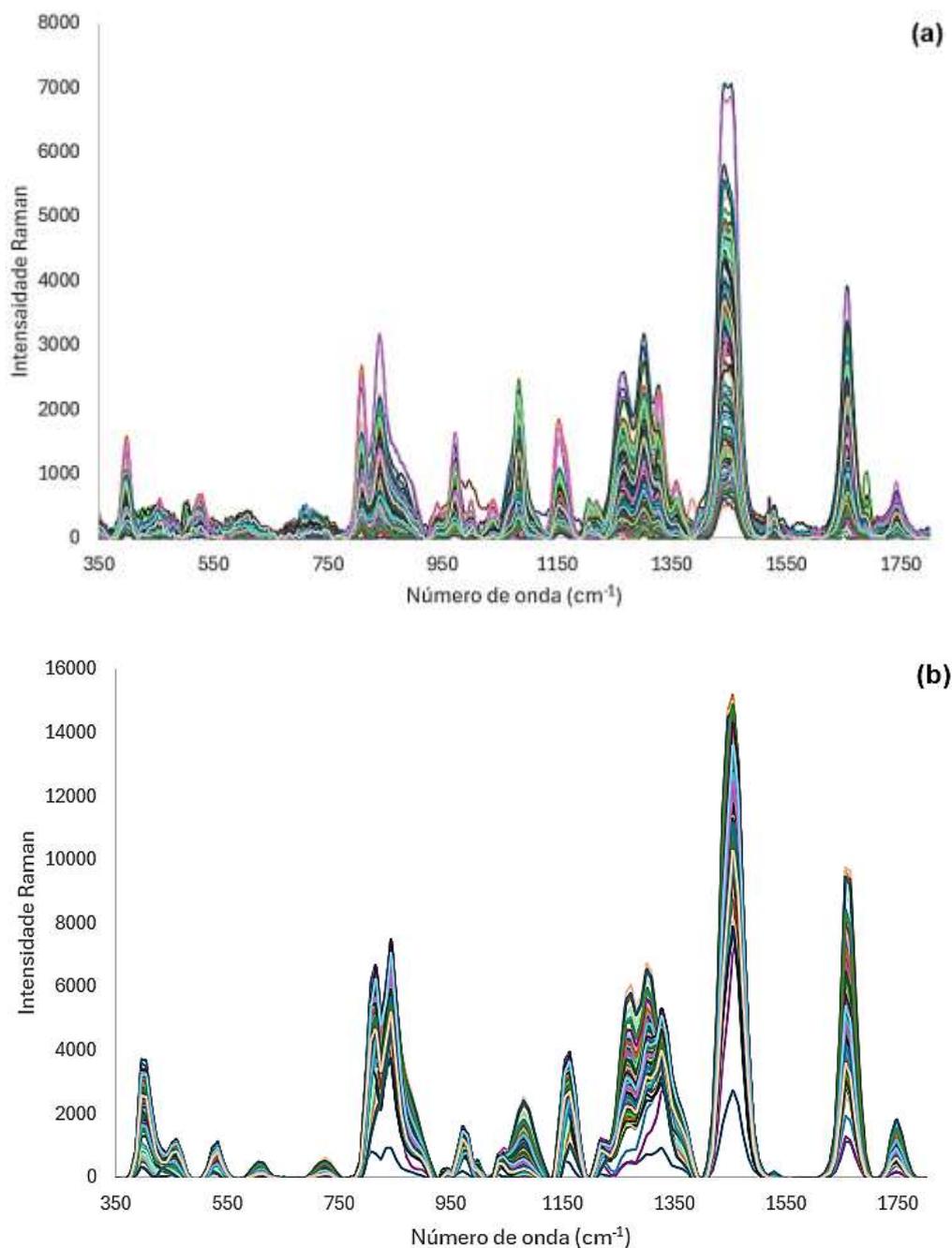


Figura 37. Espectros Raman brutos balanceados pelo método ADASYN. a. Equipamento A e b. Equipamento B.

Os espectros foram pré-processados conforme descrito na seção 4.3. A matriz de dados foi dividida em 106 amostras de treinamento (53 amostras conformes e 53 amostras não conformes) e 52 amostras de teste (26 amostras de cada classe) (KENNARD, 1969). A validação cruzada foi realizada pelo método das venezianas e os modelos foram construídos utilizando quatro variáveis latentes.

A Figura 38 mostra os resultados dos modelos PLS-DA obtidos para as amostras analisadas pelos equipamentos A e B. A classe 1 corresponde as amostras conformes e a classe 0 as amostras não conformes. O limiar entre as classes foi 0,54 para os dados do Equipamento A (Figura 38.a) e 0,50 para os dados do Equipamento B (Figura 38.b).

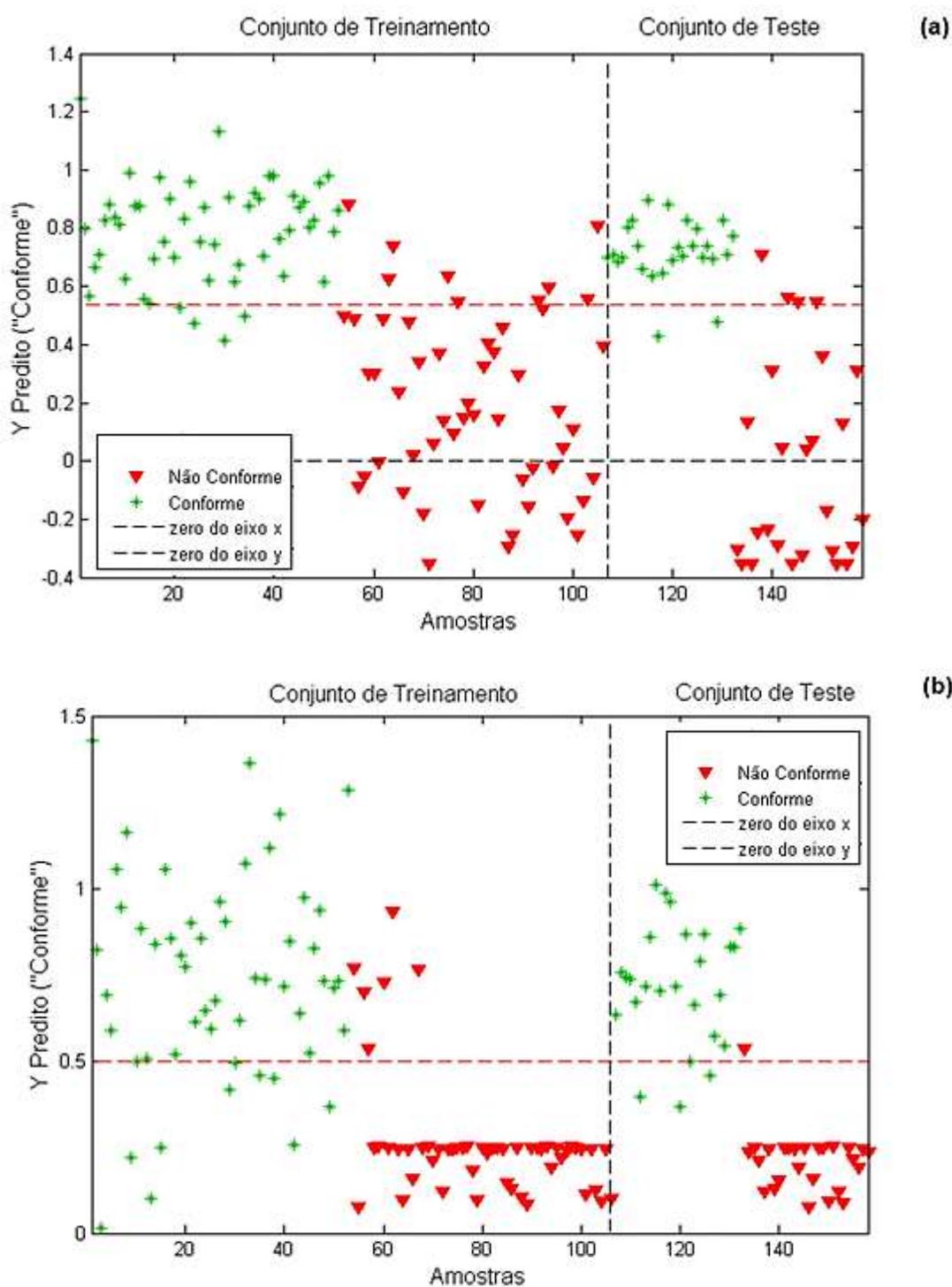


Figura 38. PLS-DA do conjunto de amostras balanceado pelo ADASYN. a. Equipamento A e b. Equipamento B.

Nos resultados obtidos a partir do equipamento A (Figura 38.a), as amostras sintéticas criadas pelo ADASYN apresentaram boa dispersão no conjunto de dados, com variância de 86,9% em **X** e 58,13% em **y**, permitindo a construção de um modelo balanceado sem viés para a classe majoritária. O gráfico de contribuição das variáveis para esse modelo, apresentado na Figura 39, mostra que a banda próxima a 1500 cm^{-1} é a que mais contribuiu para a classificação das amostras. Essa banda está associada às deformações CH_2 e CH_3 da molécula de biodiesel e pode ser relacionada com seu grau de saturação, que é um dos indicativos da oxidação de ácidos graxos. A banda próxima a 1100 cm^{-1} também é significativa para o modelo construído e pode estar associada à presença de antioxidantes na composição do B100, conforme discutido na seção 5.1.1.

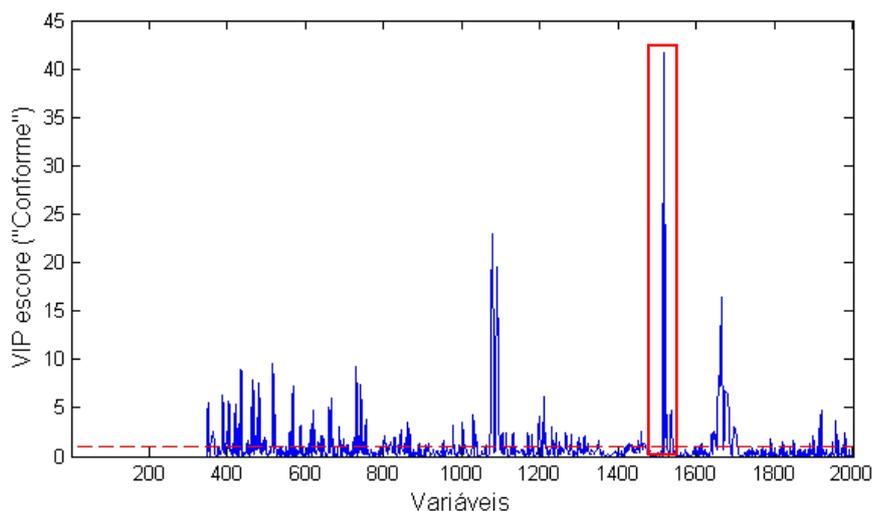


Figura 39. Contribuição das variáveis para a construção do modelo PLS-DA, calculado a partir dos dados obtidos pelo equipamento A.

Já no conjunto de dados obtidos pelo equipamento B (Figura 38.b), é observada uma região com um grupo de amostras não conformes com pouca dispersão. A presença desse grupo indica que o ADASYN criou amostras sintéticas com características muito similares, já que tal método se concentra em criar novas amostras que ocupem a região com alta densidade de amostras da classe minoritária. A baixa dispersão do conjunto de dados pode prejudicar o desempenho do modelo para classificar novas amostras.

A Figura 40 mostra o gráfico de contribuição das variáveis utilizadas na construção do modelo. Nesse gráfico é possível observar que a banda próxima a

região de 1600 cm^{-1} (estiramento C=C) é a que mais contribuiu para a classificação das amostras, demonstrando novamente como o grau de saturação da molécula de biodiesel influencia na sua estabilidade oxidativa.

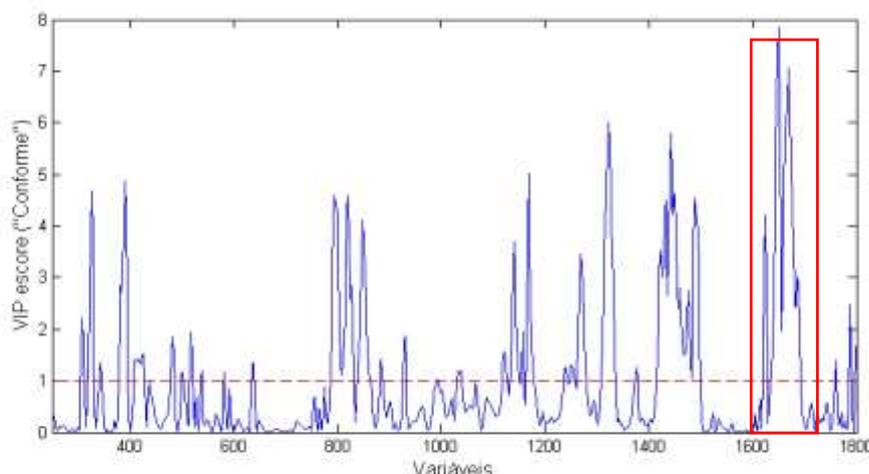


Figura 40. Contribuição das variáveis para a construção do modelo PLS-DA, calculado a partir dos dados obtidos pelo equipamento B.

A Tabela 4 apresenta as figuras de mérito dos modelos PLS-DA construídos com os dados dos equipamentos A e B. O modelo obtido a partir do conjunto de dados do equipamento B foi o que alcançou os melhores valores de eficiência (EFC) e coeficiente de correlação de Matthews (MCC). A eficiência de classificação desse modelo foi de 90,4%, com um MCC de 0,813, demonstrando sua capacidade na discriminação das classes das amostras do biodiesel. Já o modelo PLS-DA obtido para os dados do equipamento A apresentou eficiência de 88,5% e MCC de 0,77.

O modelo PLS-DA do equipamento A apresentou maior sensibilidade em comparação ao modelo do equipamento B. No entanto, sua especificidade foi inferior à do modelo do equipamento B. Conforme observado na Figura 38, o conjunto de amostras não conformes do equipamento A exibiu maior dispersão, o que pode ter resultado em uma menor especificidade do modelo.

No modelo PLS-DA do Equipamento A as bandas próximas a 1500 cm^{-1} (deformação CH_2 e CH_3) e 1100 cm^{-1} (deformação C=C de anel benzênico) foram as que mais contribuíram na classificação das amostras, evidenciando que o modelo diferencia as classes do biodiesel a partir do grau de saturação da molécula e da presença de antioxidantes. No modelo PLS-DA do Equipamento B a banda próxima a

região de 1600 cm^{-1} (estiramento C=C) foi a que mais contribuiu para a classificação das amostras, demonstrando novamente como as diferentes informações do grau de saturação da molécula de biodiesel são importantes na análise discriminante.

Tabela 4. Figuras de mérito dos modelos PLS-DA construídos.

	SEN (%)	ESP (%)	EFC (%)	MCC
Equipamento A	92,3	84,6	88,5	0,77
Equipamento B	84,6	96,2	90,4	0,81

5.1.5 Avaliação dos Equipamentos A e B por meio do método de floresta aleatória

Os modelos floresta aleatória foram construídos com base nos parâmetros ótimos descritos na Tabela 5, definidos a partir de trabalho prévio do grupo (GONÇALVES, 2022). O conjunto de dados contendo as 90 amostras (79 amostras “conforme” e 11 amostras “não conforme”) foi balanceado pelo método ADASYN, resultando em uma matriz de dados de 158 amostras (79 amostras de cada classe). Os dados balanceados foram pré-processados como descrito anteriormente e divididos em 106 amostras de treinamento e 52 amostras de teste (KENNARD, 1969).

Tabela 5. Parâmetros utilizados na construção dos modelos de floresta aleatória.

Parâmetros	Valores
Número de árvores	60
Profundidade máxima da árvore	5
Variáveis disponíveis a cada construção	$\sqrt[3]{942}$
Número mínimo de amostras em um nó	2

As matrizes de confusão dos modelos floresta aleatória, construídos a partir dos conjuntos de treinamento balanceados pelo ADASYN, são apresentadas na Figura 3A (APÊNDICE C), com 100% de eficiência na classificação das amostras. A Figura 41 apresenta as matrizes de confusão dos modelos floresta aleatória obtidos para o conjunto teste dos dados balanceados. Nota-se que em ambos os equipamentos, o modelo da floresta aleatória apresentou 100% de eficiência e MCC igual a 1, com

valores de verdadeiros positivos e verdadeiros negativos iguais a 26 e falsos positivos e falsos negativos iguais a 0.

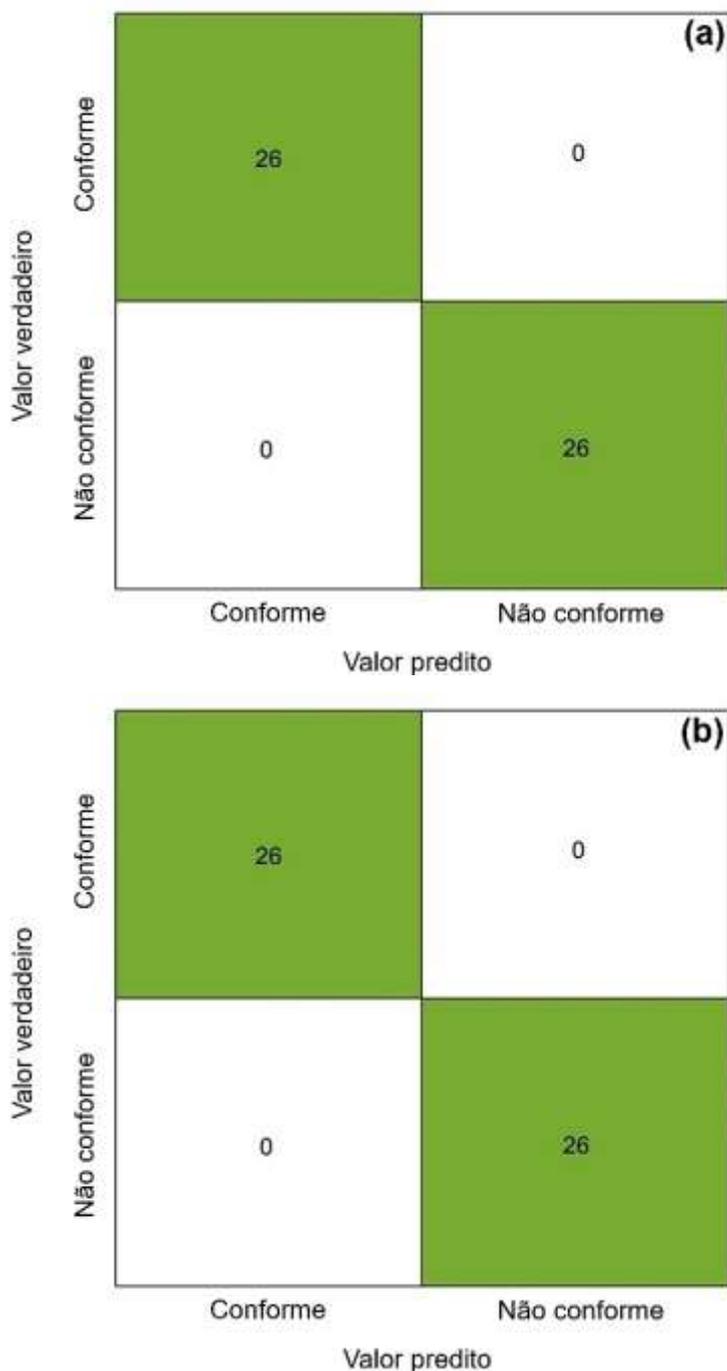


Figura 41. Matrizes de confusão dos modelos floresta aleatória construídas a partir dos conjuntos teste balanceados pelo ADASYN, a. Equipamento A e b. Equipamento B.

Os gráficos da importância das variáveis para a construção desses modelos são mostrados na Figura 42. Nota-se que, para os resultados obtidos a partir do equipamento A (Figura 42.a), as bandas localizadas na região entre 1550 e 1650 cm-

¹ (estiramento C=C) foram as mais relevantes para a classificação das amostras, demonstrando mais uma vez que o grau de insaturação das moléculas de biodiesel está diretamente ligado à sua estabilidade oxidativa. Já para os resultados obtidos a partir do equipamento B (Figura 42.b), a banda próxima a 1400 cm⁻¹ teve mais importância para a construção do modelo, sendo a região que elucida as deformações da ligação C-H. Além disso, é importante observar que, em ambos os casos, toda a faixa espectral avaliada (350 - 1800 cm⁻¹) contribuiu, em algum grau, para a construção do modelo, evidenciando que os modelos de floresta aleatória conseguiram capturar de forma abrangente as informações químicas descritas pelo espectro Raman.

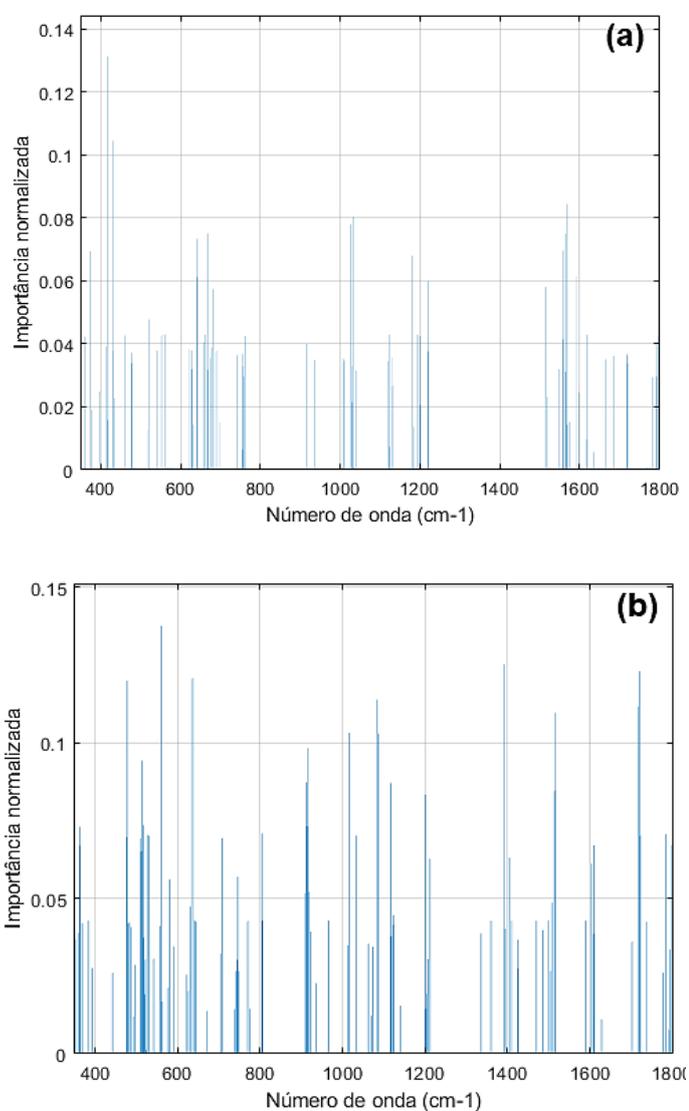


Figura 42. Contribuição normalizada das variáveis para a construção dos modelos floresta aleatória. a. Equipamento A e b. Equipamento B.

A Tabela 6 apresenta as figuras de mérito calculadas para os modelos de floresta aleatória. Os dois modelos construídos tiveram valores máximos de eficiência (100%) e MCC (1), demonstrando que independente do equipamento Raman utilizado na coleta de dados, a modelagem por floresta aleatória foi eficiente na classificação do biodiesel quanto à sua estabilidade oxidativa. A avaliação dos modelos construídos demonstrou que a presença de insaturações, representadas no espectro pelo estiramento C=C (1650 cm^{-1}) e pelas deformações CH₂ e CH₃ (1500 cm^{-1}), foi a característica principal utilizada na construção do modelo floresta aleatória, corroborando com os resultados apresentados em trabalhos anteriores (VALERIANO *et al.*, 2024). Essa observação concorda com a literatura e com o que foi elucidado na revisão bibliográfica, já que as reações de oxidação das moléculas de biodiesel ocorrem nas regiões de insaturação.

Tabela 6. Figuras de mérito dos modelos floresta aleatória construídos.

	SEN (%)	ESP (%)	EFC (%)	MCC
Equipamento A	100	100	100	1
Equipamento B	100	100	100	1

5.1.6 Comparação entre os equipamentos A e B

Os modelos de floresta aleatória construídos a partir dos dados espectrais das 90 amostras analisadas tiveram desempenho semelhante para ambos os equipamentos utilizados, alcançando 100% de eficiência na classificação das amostras de biodiesel B100. Cabe ressaltar que a floresta aleatória lida melhor com desvios de linearidade, o que pode ter levado a modelos com melhores desempenhos em relação ao PLS-DA e SIMCA. A partir desses resultados, torna-se mais relevante comparar as características técnicas dos dois equipamentos e avaliar suas condições de aplicabilidade.

O equipamento A é portátil, leve, com bateria duradoura (~4 horas) e possui a tecnologia SORS, que permite que a amostra de biodiesel seja analisada dentro de diferentes tipos de recipiente, sem a etapa de preparo de amostras, conforme mostrado na Figura 43. A praticidade desse equipamento possibilita sua aplicação em campo, onde as medições Raman poderiam ser realizadas em qualquer etapa da rota

sintética do biodiesel, para posterior classificação da sua estabilidade oxidativa através dos modelos adequados. Isso reduziria as etapas envolvidas no controle de qualidade do biocombustível, além de proporcionar análises mais rápidas, eficientes e que não necessitam de ambiente controlado para serem realizadas. O Equipamento A permite a exportação de dados apenas no formato de texto .csv, o que acrescenta etapas durante o processo de tratamento de dados.



Figura 43. Análise Raman utilizando o Equipamento A, sobre recipientes de diferentes materiais (AGILENT®, 2024).

Por sua vez, o equipamento B é um equipamento de bancada compacto, mais pesado e que necessita de conexão à rede elétrica. Essa característica limita a sua aplicação em campo, exigindo que a análise seja realizada em locais com acesso à rede elétrica e em recipientes que sejam compatíveis com o porta-amostras do equipamento. Ainda assim, a avaliação da estabilidade oxidativa do biodiesel por espectroscopia Raman continua sendo mais rápida que a do método Rancimat, mesmo com o uso de equipamentos de bancada.

O equipamento B possui dois modos para suporte de amostras, o suporte de *vial* para amostras líquidas e a mesa para lâminas de vidro, o que tornam as medidas mais dinâmicas e automatizadas. Sua tecnologia Autofocus ajusta automaticamente o foco da amostra, resultando em espectros mais intensos e com maior razão sinal/ruído. Vale ressaltar que a exportação dos dados pelo equipamento B é feita de forma automática via USB, gerando arquivos nos formatos .pdf, .csv e .txt e arquivos de imagem .png.



Figura 44. Uso do Equipamento B dentro de um laboratório. (ANTON PAAR®, 2024).

Tendo em vista o desempenho e as características observadas em cada equipamento, conclui-se que ambos são adequados para a análise da estabilidade oxidativa do biodiesel, diferenciando-se apenas nas formas de coleta dos espectros e nas propostas de aplicações. O equipamento A poderia ser utilizado no controle de qualidade do biodiesel nas etapas de transporte e armazenamento, enquanto o equipamento B se mostra mais apropriado para a avaliação da conformidade do biodiesel nos centros de distribuição, antes da sua mistura com o diesel.

5.2 Construção de um modelo de classificação hierárquico para avaliação da estabilidade oxidativa do biodiesel com geração de amostras sintéticas

Com base nos resultados mostrados na seção 5.1, o método de classificação floresta aleatória apresentou melhor desempenho na classificação das amostras de biodiesel B100 em conforme e não conforme em relação à sua estabilidade oxidativa. Baseado nesses resultados, um novo conjunto de dados constituído por um total de 162 amostras, sendo 147 amostras da classe “conforme” e 14 amostras da classe “não conforme”, foi analisada pelo equipamento B. Dois métodos de balanceamento de classes, ADASYN e SMOTE, foram utilizados para balancear as amostras da classe minoritária (não conforme).

A Figura 45 mostra os espectros Raman obtidos pelo equipamento B e balanceados pelos métodos ADASYN (Figura 45.a) e SMOTE (Figura 45.b). O conjunto de dados balanceado é constituído por um total de 294 amostras, sendo 147 amostras de cada classe. Visualmente, não é possível observar mudanças nos perfis

espectrais quanto ao tipo de balanceamento empregado para geração das amostras sintéticas, continuando sendo visualizadas as bandas características do biodiesel nas faixas de 800 a 900 cm^{-1} , 1245 a 1350 cm^{-1} , 1400 a 1500 cm^{-1} e 1665 a 1750 cm^{-1} .

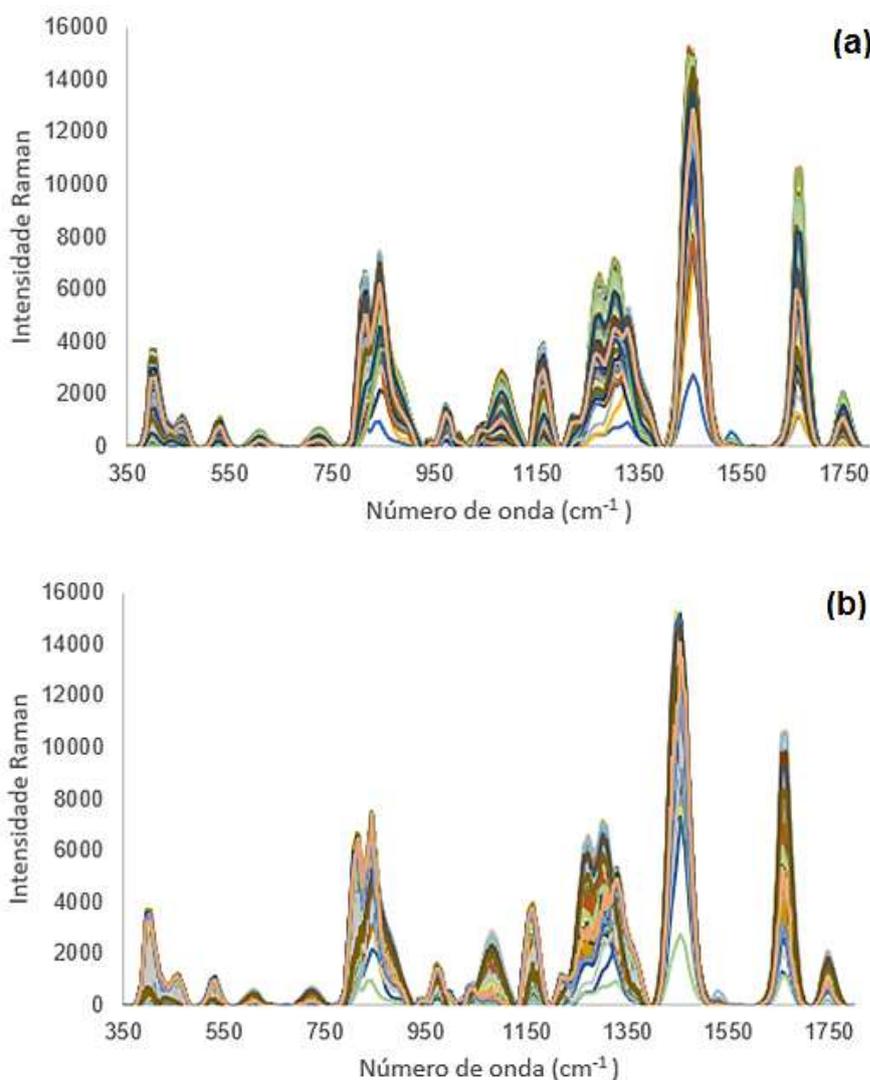


Figura 45. Espectros Raman brutos obtidos na análise do novo conjunto de dados pelo equipamento B, balanceados pelo a. ADASYN e b. SMOTE.

Os conjuntos de dados obtidos foram pré-processados empregando primeira derivada com alisamento de *Savitzky-Golay* (janela de 7 pontos e polinômio de segunda ordem) seguido dos dados centrados na média. A matriz de dados pré-processada foi dividida em 206 amostras de treinamento (103 amostras conformes e 103 amostras não conformes) e 88 amostras de teste (44 amostras de cada classe), para a construção dos modelos de classificação (KENNARD, 1969). Os modelos de classificação com o método floresta aleatória foram construídos a partir dos conjuntos de treinamento e com base nos parâmetros ótimos descritos na Tabela 5 (seção 5.1.5).

A Figura 46 apresenta as matrizes de confusão do modelo floresta aleatória para os conjuntos de treinamento (Figura 46.a) e teste (Figura 46.b), com os dados balanceados pelo ADASYN. As amostras do conjunto de treinamento foram classificadas com 103 verdadeiros positivos, 103 verdadeiros negativos, 0 falsos positivos e 0 falsos negativos. Já as amostras do conjunto teste foram classificadas com 44 verdadeiros positivos, 44 verdadeiros negativos, 0 falsos positivos e 0 falsos negativos, resultando em um modelo com 100% de eficiência para classificação das amostras de biodiesel quanto à estabilidade oxidativa.

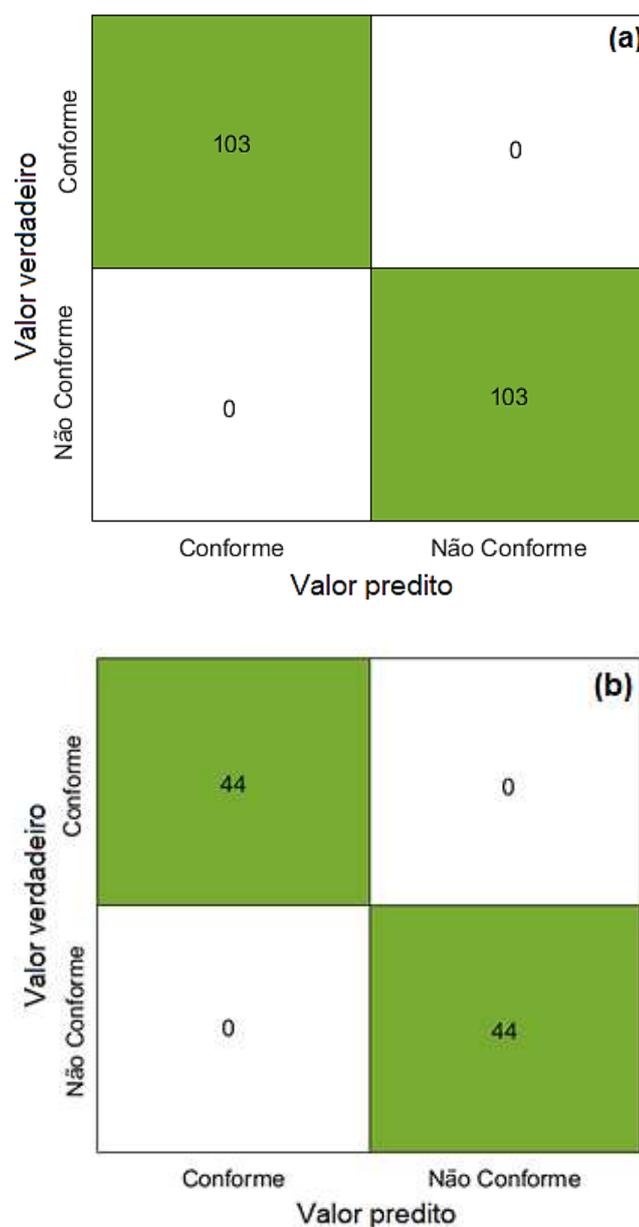


Figura 46. Matrizes de confusão do modelo floresta aleatória construído a partir dos dados balanceados pelo ADASYN, a. conjunto treinamento e b. conjunto teste.

A Figura 47 apresenta as matrizes de confusão do modelo floresta aleatória para os conjuntos de treinamento (Figura 47.a) e teste (Figura 47.b), com os dados balanceados pelo SMOTE. As amostras do conjunto de treinamento foram classificadas com 103 verdadeiros positivos, 103 verdadeiros negativos, 0 falsos positivos e 0 falsos negativos. Já as amostras do conjunto teste, foram classificadas com 44 verdadeiros positivos, 40 verdadeiros negativos, 4 falsos positivos e 0 falsos negativos, resultando em um modelo com 95,5% de eficiência para classificação das amostras de biodiesel quanto à estabilidade oxidativa.

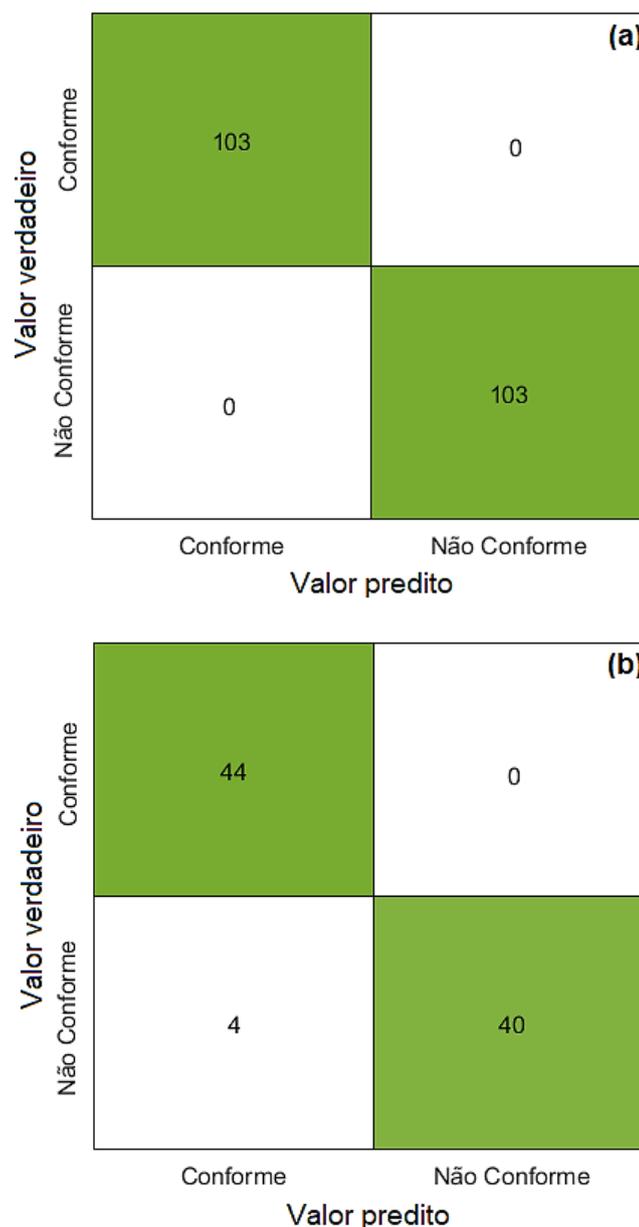


Figura 47. Matrizes de confusão do modelo floresta aleatória construído a partir dos dados balanceados pelo SMOTE, a. conjunto treinamento e b. conjunto teste.

A Tabela 7 apresenta as figuras de mérito calculadas para os conjuntos teste dos modelos floresta aleatória. Os modelos floresta aleatória apresentaram eficiência superior a 95% para classificação das amostras de biodiesel como “conformes” ou “não conformes”. O modelo construído a partir do conjunto de dados balanceado pelo ADASYN obteve melhores resultados, com eficiência, sensibilidade e especificidade de 100% e MCC igual a 1. Já o modelo construído a partir do conjunto de dados balanceado pelo SMOTE, obteve resultados ligeiramente inferiores, com eficiência de 95,5%, especificidade de 90,9%, sensibilidade de 100% e MCC igual a 0,91. O método de balanceamento ADASYN utiliza o peso das amostras do conjunto não-alvo para criar as amostras sintéticas, distribuindo as novas amostras na região duvidosa e aumentando as informações contidas na fronteira entre as classes, o que pode ter corroborado para a sensibilidade e especificidade do modelo. Por sua vez o SMOTE não considera o peso das amostras para criar as amostras sintéticas, distribuindo-as homogeneamente na matriz de dados balanceada e não aumentando as informações no limiar entre as classes, o que pode explicar o menor valor de especificidade. Entretanto, vale ressaltar que ainda não há um consenso na literatura de qual método de balanceamento seja mais eficiente, sendo sua escolha pautada pelas características do conjunto de dados e pelos resultados obtidos.

Tabela 7. Figuras de mérito dos conjuntos teste dos modelos floresta aleatória construídos.

	SEN (%)	ESP (%)	EFC (%)	MCC
ADASYN	100	100	100	1
SMOTE	100	90,9	95,5	0,91

Os gráficos da importância das variáveis para a construção dos modelos de floresta aleatória, balanceados pelos ADASYN e SMOTE, são mostrados na Figura 48.a e Figura 48.b, respectivamente. Observa-se que para o modelo construído a partir dos dados balanceados pelo ADASYN (Figura 48.a) as bandas nas faixas de 600 a 800 cm^{-1} e 1550 a 1650 cm^{-1} foram as que mais contribuíram na discriminação das classes. A banda localizada na faixa de 600 a 800 cm^{-1} está relacionada ao estiramento C-O e pode indicar a presença de produtos de oxidação do biodiesel, como os peróxidos. Já a banda localizada na faixa de 1550 a 1650 cm^{-1} representa o estiramento da C=C, que pode ser usado como indicativo de oxidação da molécula de

biodiesel, já que a reação oxidativa ocorre nas insaturações da sua estrutura. Para o modelo construído a partir dos dados balanceados pelo SMOTE (Figura 48.b) a distribuição do gráfico da contribuição das variáveis é similar ao ADASYN, destacando-se a faixa de 1550 a 1650 cm^{-1} (estiramento C=C), reiterando a importância das insaturações para a classificação da estabilidade oxidativa das amostras de biodiesel por meio da espectroscopia Raman.

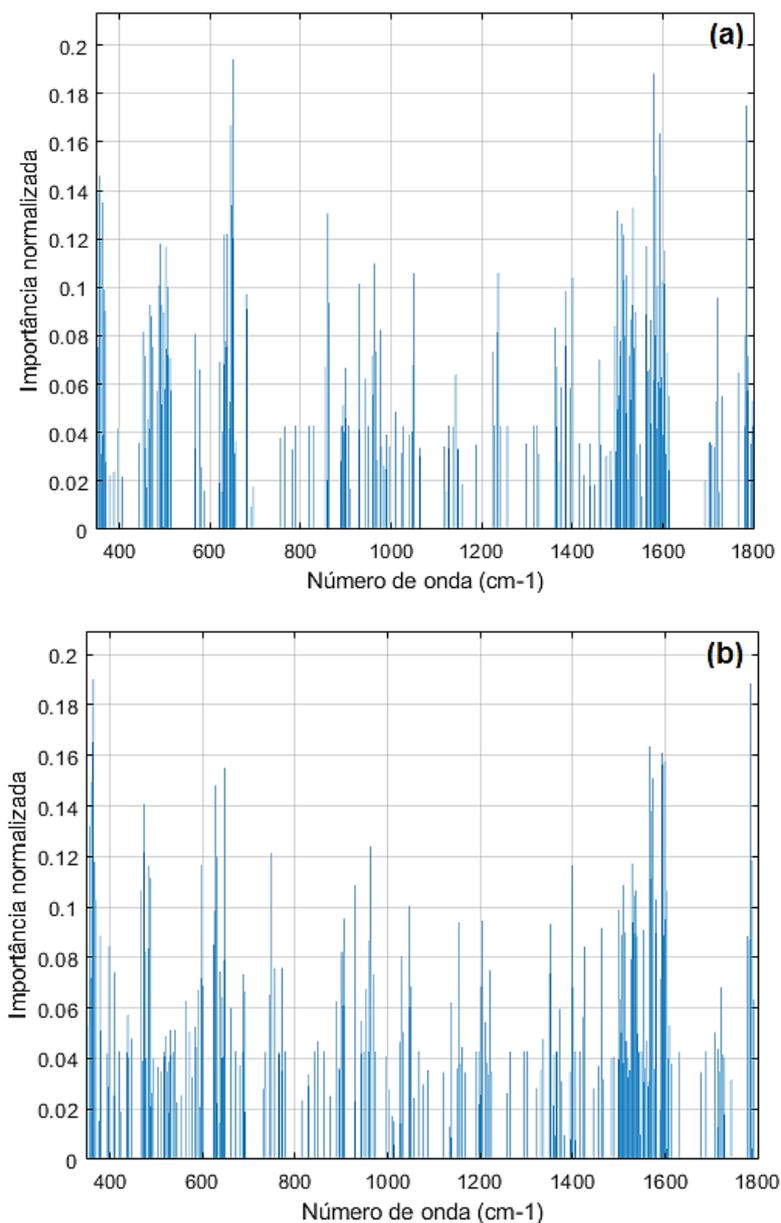


Figura 48. Contribuição normalizada das variáveis para a construção dos modelos floresta aleatória, a partir dos dados balanceados pelo a. ADASYN e b. SMOTE.

Figuras de mérito relacionadas a precisão foram estimadas em seis níveis de estabilidade oxidativa (8, 11, 12, 13, 16, e 23 horas). A acordância (ACO) foi avaliada

em condições de repetibilidade e a concordância (CON) em condição de precisão intermediária, considerando a análise em dois dias, realizada por dois analistas diferentes. A Tabela 8 apresenta os resultados de ACO e CON obtidos para as amostras de biodiesel classificadas pelos modelos floresta aleatória, balanceados pelo ADASYN e pelo SMOTE. Por meio dos valores de ACO e CON calculados, foi observada a falta de precisão e repetibilidade do modelo para valores próximos ao limiar de classificação do biodiesel como conforme ou não conforme. Na faixa de 12 a 13 horas, os valores de ACO e CON obtidos foram inferiores a 50%, indicando que, para esses valores de estabilidade oxidativa, as mudanças nas moléculas do biodiesel ainda são muito sutis, não sendo identificadas de forma eficaz pela técnica de espectroscopia Raman. Em contraste, para os valores extremos de estabilidade oxidativa avaliados (8 e 23 horas) os resultados de ACO e CON obtidos são maiores ($\geq 50\%$), indicando que amostras muito oxidadas (8 horas) ou não oxidadas (23 horas) apresentam alterações mais significativas que podem ser detectadas de forma mais clara pela espectroscopia Raman e identificadas pelo modelo.

Apesar das discussões anteriores, é importante ressaltar que a amostragem utilizada nas medidas para o cálculo da ACO e CON foi diferente da empregada para a construção do modelo. A primeira medida foi feita em lâminas de vidro, enquanto a segunda foi feita diretamente no microtubo da amostra. Acredita-se que essa diferença de amostragem tenha influenciado os resultados de ACO e CON, contribuindo para a redução desses valores. Isso evidencia que a forma de apresentação da amostra impacta diretamente na qualidade dos espectros obtidos e, conseqüentemente, na robustez do modelo construído.

Tabela 8. Valores de acórdância e concordância calculados para amostras de biodiesel classificadas pelos modelos floresta aleatória.

Estabilidade oxidativa (h)	8	11	12	13	16	23
ACO (%) - ADASYN	66,7	100	33,3	16,7	50,0	66,7
ACO (%) - SMOTE	66,7	16,7	16,7	16,7	33,3	50,0
CON (%) - ADASYN	50,0	66,7	16,7	33,3	50,0	50,0
CON (%) - SMOTE	66,7	16,7	33,3	33,3	33,3	83,3

Com o objetivo de construir um modelo hierárquico, em que em um primeiro momento as amostras são classificadas em conforme ou não conforme, e em seguida, as amostras conformes são classificadas em alta estabilidade ou média estabilidade oxidativa, um novo modelo de classificação foi construído. Para a construção deste novo modelo, o conjunto de amostras composto por 147 amostras de biodiesel conformes, sendo 127 amostras conformes com estabilidade oxidativa média (< 17 horas) e 20 amostras conformes com estabilidade oxidativa alta (> 17 horas) foi empregado. O conjunto de dados contendo as amostras de biodiesel com alta estabilidade foi balanceado utilizando o ADASYN, resultando em uma matriz contendo 127 amostras de cada grupo (< 17 horas e > 17 horas).

A matriz de dados foi pré-processada empregando primeira derivada com alisamento de Savitzky-Golay (janela de 7 pontos e polinômio de segunda ordem) seguido dos dados centrados na média, e foi dividida em 178 amostras de treinamento (89 amostras com estabilidade oxidativa alta e 89 amostras com estabilidade oxidativa média) e 76 amostras de teste (38 amostras de cada grupo) (KENNARD, 1969). O modelo floresta aleatória foi construído a partir do conjunto de treinamento, com base nos parâmetros ótimos descritos na Tabela 5 (seção 5.1.5), e validado usando um conjunto teste.

A Figura 49 apresenta as matrizes de confusão para os conjuntos de treinamento (Figura 49.a) e teste (Figura 49.b). As amostras do conjunto de treinamento foram classificadas com 89 verdadeiros positivos, 89 verdadeiros negativos, 0 falsos positivos e 0 falsos negativos. Já as amostras do conjunto teste foram classificadas com 38 verdadeiros positivos, 34 verdadeiros negativos, 4 falsos positivos e 0 falsos negativos. Resultando em um modelo com 100% de sensibilidade, 89,5% de especificidade, 94,7% de eficiência e MCC igual a 0,903, para classificação das amostras de biodiesel com alta (> 17h) e média (<17h) estabilidade oxidativa.

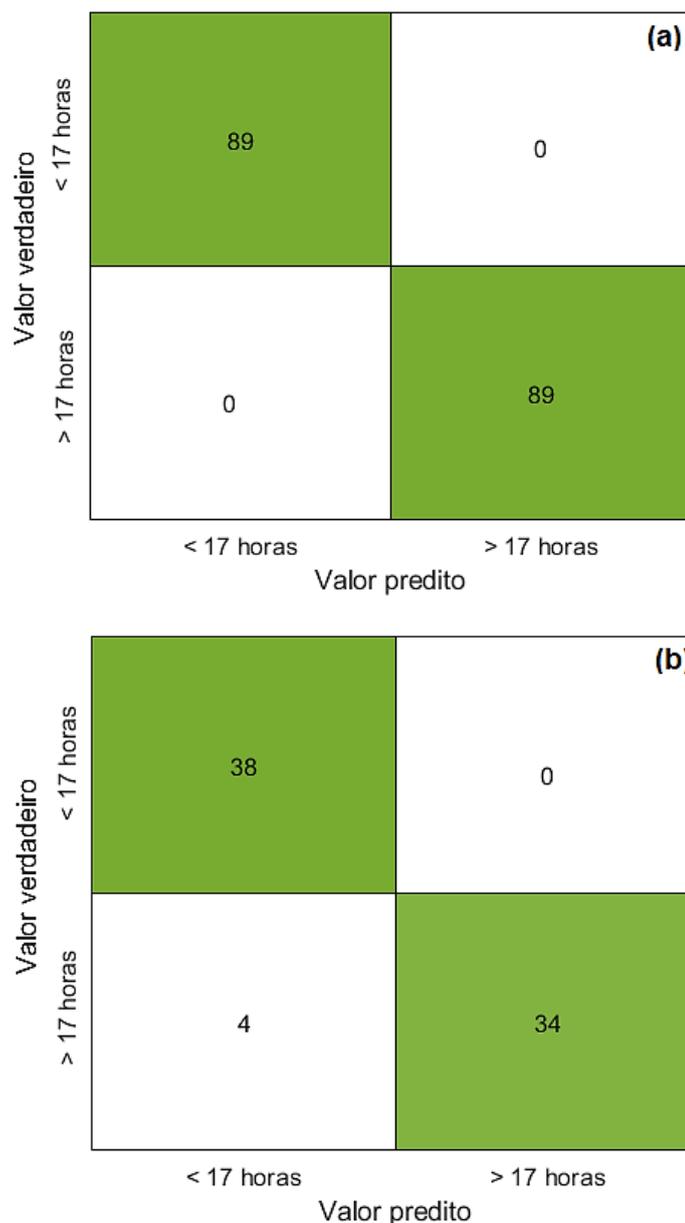


Figura 49. Matrizes de confusão do modelo hierárquico construído, a. conjunto treinamento e b. conjunto teste.

O gráfico da importância das variáveis para a construção do modelo é mostrado na Figura 50. Observa-se que a banda localizada próxima a 1500 cm^{-1} foi a que mais contribuiu na discriminação das amostras conformes com estabilidade oxidativa média e alta, estando associada às deformações dos grupos CH_2 e CH_3 . Como discutido na seção 3.3, a etapa de iniciação do mecanismo radicalar de oxidação do biodiesel ocorre por meio da extração de um átomo de hidrogênio dos ésteres de ácidos graxos. A velocidade da reação de auto-oxidação das substâncias graxas varia de acordo com o número e a posição de insaturações presentes nas suas moléculas. O hidrogênio da posição CH_2 -alílica é o mais susceptível à oxidação,

devido à sua localização favorável ao ataque por radicais livres. As deformações CH_2 no espectro Raman refletem as etapas iniciais do mecanismo radicalar, quando a reação de oxidação ainda não é completa. Essa característica torna as deformações CH_2 uma indicação importante das etapas iniciais da oxidação, especialmente quando as amostras de biodiesel estão próximas ao limiar de 13 horas (KNOTHE *et al.*, 2006). O fato de essas deformações estarem destacadas como importantes para o modelo, mostra que o modelo foi eficiente em capturar mudanças sutis nos espectros, que podem indicar um estágio inicial de oxidação, mas com potenciais alterações nas propriedades do biodiesel a curto prazo.

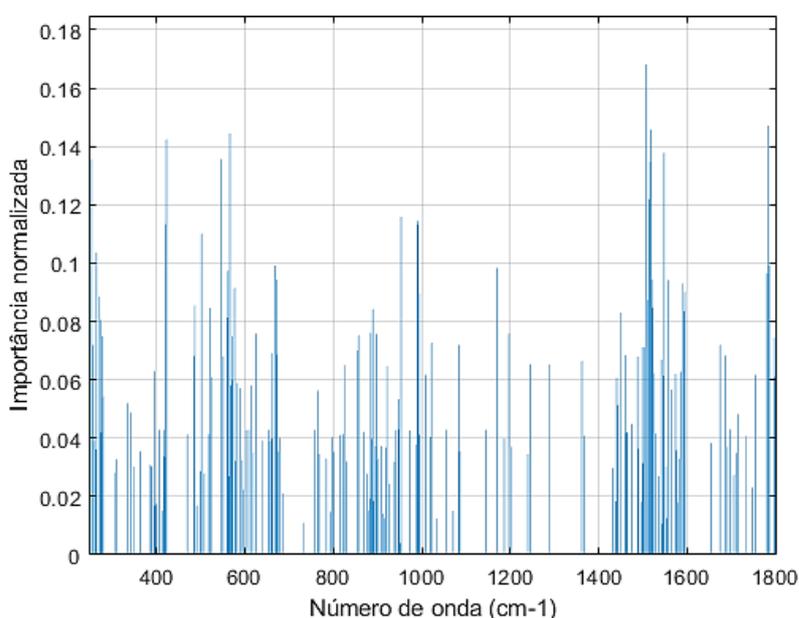


Figura 50. Contribuição normalizada das variáveis para a construção do modelo hierárquico.

Os modelos floresta aleatória construídos obtiveram eficiência maior que 94% e MCC maior que 0,9 na classificação das amostras de biodiesel quanto à sua estabilidade oxidativa. O conjunto de dados balanceado pelo ADASYN resultou no modelo floresta aleatória mais eficiente (100%), indicando que o aumento das informações no limiar entre as classes “conforme” e “não conforme” (13 horas) pode ter contribuído para diferenciação das amostras. As bandas representativas dos estiramentos C-O e C=C contribuíram mais para a modelagem do conjunto de dados balanceado pelo ADASYN, indicando que a presença dos produtos de oxidação do biodiesel e a ocorrência de insaturações, respectivamente, representaram a divisão das classes nesse modelo. Já no modelo construído a partir dos dados balanceados pelo SMOTE a banda do estiramento C=C foi mais significativa, evidenciando como a

presença de insaturações favorece as reações de oxidação da molécula. As figuras ACO e CON calculadas demonstraram que a precisão do método é maior ($\geq 50\%$) para valores extremos de estabilidade oxidativa (8 e 23 horas), nos quais são identificadas mudanças mais significativas no espectro Raman. No modelo construído para verificar se a amostra conforme tinha alta ou média estabilidade, a eficiência de classificação foi de 94,7% e MCC igual a 0,903. Além disso, para esse modelo a região espectral que mais contribuiu foi referente a deformação CH₂, o que pode indicar etapas iniciais do mecanismo radicalar, quando a reação de oxidação ainda não é completa.

6. CONSIDERAÇÕES FINAIS

No presente trabalho foram construídos modelos de classificação para avaliar a estabilidade oxidativa de amostras de biodiesel B100 a partir de dados de espectroscopia Raman. Na primeira parte do trabalho, as amostras de B100 foram submetidas à análise utilizando dois equipamentos: um espectrômetro Raman portátil com tecnologia SORS (Equipamento A) e um espectrômetro Raman compacto (Equipamento B). A partir dos dados dos espectrômetros foram construídos modelos de classificação SIMCA, PLS-DA e floresta aleatória, estes foram avaliados em termos de sensibilidade, especificidade, eficiência e coeficiente de Matthews. Os resultados obtidos demonstraram que ambos os equipamentos possuem desempenho semelhante. Os modelos construídos com a floresta aleatório apresentaram melhores resultados, com 100% de eficiência na classificação das amostras de biodiesel.

Entretanto, as características individuais de cada equipamento refletem no tipo de aplicabilidade de cada um. O equipamento Raman portátil é mais indicado para a avaliação do biodiesel em campo, como nos caminhões de transporte ou nos estoques, já que a tecnologia SORS possibilita que a leitura seja feita sobre diferentes tipos de recipiente. Já o equipamento Raman de bancada precisa ser utilizado com ligação direta a rede elétrica, sendo adequado para as análises de controle de qualidade dentro da fábrica ou nos centros de distribuição.

Na segunda parte do trabalho, foram construídos modelos de floresta aleatória comparando dois métodos de balanceamento de classes: o ADASYN e o SMOTE. O modelo de floresta aleatória construído a partir dos dados balanceados pelo ADASYN alcançou os valores máximos de eficiência (100%) e MCC (1). Os testes de acordância e concordância demonstraram que a metodologia é mais precisa na classificação de amostras de biodiesel com valores extremos de estabilidade oxidativa (8 e 23 horas). Isso ocorre devido a sensibilidade da espectroscopia Raman, que consegue identificar mudanças mais significativas nas amostras de biodiesel muito oxidadas (8 horas) ou muito estáveis (23 horas). As amostras de biodiesel com estabilidade oxidativa próxima ao limiar de referência (13 horas) resultaram em menores valores de ACO e CON, o que pode estar associado ao fato dessas amostras terem sofrido mudanças muito sutis que ainda não são identificadas pela espectroscopia Raman.

Os gráficos de contribuição das variáveis para os modelos de classificação revelam que o estiramento C=C (1600 cm^{-1}), o estiramento C-O (800 cm^{-1}), as deformações C-H (1400 cm^{-1}) e as deformações CH₂ e CH₃ (1500 cm^{-1}) são as características que mais contribuíram na discriminação das classes das amostras de biodiesel. Essa observação corrobora com a interpretação química do espectro Raman, já que esses modos vibracionais indicam o grau de insaturação das moléculas de biodiesel e a presença de produtos de oxidação como peróxidos e álcoois, o que está diretamente relacionado a estabilidade oxidativa do biodiesel.

As amostras classificadas como “conformes” no modelo floresta aleatória foram reavaliadas em um novo modelo para discriminação das amostras com estabilidade oxidativa maior e menor que 17 horas. O modelo construído apresentou 94,7% de eficiência na classificação das amostras conformes com estabilidade oxidativa >17 horas e <17 horas, sendo que a banda localizada próxima a 1500 cm^{-1} (deformações CH₂ e CH₃) foi a que mais contribuiu para a construção do modelo. Os hidrogênios da posição CH₂-alílica são os mais susceptíveis a oxidação, participando da etapa de iniciação do mecanismo radicalar de oxidação do biodiesel.

A proposta de construção de modelos hierárquicos utilizando dados de espectroscopia Raman permitiu extrair mais informações das amostras. Inicialmente, as amostras foram classificadas como 'conformes' ou 'não conformes'. Em um segundo momento, as amostras classificadas como 'conformes' são submetidas a um modelo adicional, que as categoriza em 'alta' ou 'média' estabilidade.

A abordagem apresentada neste trabalho pode ser aplicada em processos de controle da qualidade do biodiesel nas etapas de produção, transporte e armazenamento. O uso de equipamentos portáteis possibilita que a análise da estabilidade oxidativa do biodiesel seja feita de forma rápida, diretamente nos tanques de armazenamento ou caminhões de transporte, acelerando o processo e evitando custos adicionais associados ao envio de amostras para o laboratório.

As perspectivas futuras do trabalho incluem: a análise de um conjunto maior de amostras da classe “não conforme”; o cálculo das figuras de mérito ACO e CON usando a mesma amostragem empregada para a construção dos modelos; e a avaliação de outros parâmetros de qualidade do biodiesel, como a contaminação total e o teor de água.

REFERÊNCIAS

- ANP (2023). Resolução ANP n° 920, de 4 de abril de 2023. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Disponível em: <<https://atosoficiais.com.br/anp/resolucao-n-920-2023-estabelece-a-especificacao-do-biodiesel-e-as-obrigacoes-quanto-ao-controle-da-qualidade-a-serem-atendidas-pelos-agentes-economicos-que-comercializem-o-produto-em-territorio-nacional?origin=instituicao>>.
- ANP (2020). Resolução ANP n° 828, de 1 de setembro de 2020. Agência Nacional do Petróleo, Gás Natural e Biocombustíveis. Disponível em: <<https://atosoficiais.com.br/anp/resolucao-n-828-2020-dispoe-sobre-as-informacoes-constantas-dos-documentos-da-qualidade-e-o-envio-dos-dados-da-qualidade-dos-combustiveis-produzidos-no-territorio-nacional-ou-importados-e-da-outras-providencias?origin=instituicao>>.
- BABADI, A. A. et al. Emerging technologies for biodiesel production: Processes, challenges, and opportunities. *Biomass and Bioenergy*, v. 163, p. 106521, ago. 2022.
- BAHRI, A. et al. Application of visible and near-infrared spectroscopy for evaluation of ewes milk with different feeds. *Animal Production Science*, v. 59, n. 6, p. 1190–1190, 1 jan. 2019.
- BERRUETA, L. A.; ALONSO-SALCES, R. M.; HÉBERGER, K. Supervised pattern recognition in food analysis. *Journal of Chromatography A*, v. 1158, n. 1-2, p. 196–214, jul. 2007.
- BIANCOLILLO, A. et al. Determination of insect infestation on stored rice by near infrared (NIR) spectroscopy. *Microchemical Journal*, v. 145, p. 252–258, mar. 2019.
- BRASIL. Lei n° 11.097, de 13 de janeiro de 2005. Dispõe sobre a introdução do biodiesel na matriz energética brasileira; altera as Leis n°s 9.478, de 6 de agosto de 1997, 9.847, de 26 de outubro de 1999 e 10.636, de 30 de dezembro de 2002; e dá outras providências. *Diário Oficial da União*, Brasília, DF, 13 jan. 2005.
- BRASIL. Lei n° 14.993, de 8 de outubro de 2024. Dispõe sobre a promoção da mobilidade sustentável de baixo carbono e a captura e a estocagem geológica de dióxido de carbono; institui o Programa Nacional de Combustível Sustentável de

Aviação (ProBioQAV), o Programa Nacional de Diesel Verde (PNDV) e o Programa Nacional de Descarbonização do Produtor e Importador de Gás Natural e de Incentivo ao Biometano; altera as Leis nºs 9.478, de 6 de agosto de 1997, 9.847, de 26 de outubro de 1999, 8.723, de 28 de outubro de 1993, e 13.033, de 24 de setembro de 2014; e revoga dispositivo da Lei nº 10.438, de 26 de abril de 2002. Diário Oficial da União, Brasília, DF, 8 out. 2024.

BRO, R.; SMILDE, A. K. Principal component analysis. *Anal. Methods*, v. 6, n. 9, p. 2812–2831, 2014.

CANHA, N. et al. Multivariate near infrared spectroscopy models for predicting the oxidative stability of biodiesel: Effect of antioxidants addition. *Fuel*, v. 97, p. 352–357, 20 fev. 2012.

CAO, D.-S. et al. Tree-based ensemble methods and their applications in analytical chemistry. *TrAC Trends in Analytical Chemistry*, v. 40, p. 158–167, nov. 2012.

CARVALHO JUNIOR, W. DE et al. Regressão linear múltipla e modelo Random Forest para estimar a densidade do solo em áreas montanhosas. *Pesquisa Agropecuária Brasileira*, v. 51, n. 9, p. 1428–1437, set. 2016.

CHAWLA, N. V. et al. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, v. 16, n. 16, p. 321–357, 1 jun. 2002.

CHICCO, D.; JURMAN, G. A statistical comparison between Matthews correlation coefficient (MCC), prevalence threshold, and Fowlkes–Mallows index. *Journal of Biomedical Informatics*, v. 144, p. 104426–104426, 1 ago. 2023.

COMITÉ EUROPÉEN DE NORMALISATION. EN 14112: Fat and oil derivatives - Fatty Acid Methyl Esters (FAME) - Determination of oxidation stability (accelerated oxidation test). Brussels: CEN, 2003.

CORONADO, C. R.; DE CARVALHO, J. A.; SILVEIRA, J. L. Biodiesel CO₂ emissions: A comparison with the main fuels in the Brazilian market. *Fuel Processing Technology*, v. 90, n. 2, p. 204–211, 2009.

DAFFERTSHOFER, A. et al. PCA in studying coordination and variability: a tutorial. *Clinical Biomechanics*, v. 19, n. 4, p. 415–428, maio 2004.

DEY, I.; PRATAP, V. A Comparative Study of SMOTE, Borderline-SMOTE, and ADASYN Oversampling Techniques using Different Classifiers. 1 mar. 2023.

DOMINGOS, A. K. et al. The influence of BHA, BHT and TBHQ on the oxidation stability of soybean oil ethyl esters (biodiesel). *Journal of the Brazilian Chemical Society*, v. 18, n. 2, p. 416–423, abr. 2007.

DUNN, Kevin G. *Process Improvement using Data*. [s. l.], n. 294-34b8, p. 381, 2014. Disponível em: learnche.mcmaster.ca/pid.

FERREIRA, Márcia M C. *Quimiometria: conceitos, métodos e aplicações*. [S. l.]: Editora Unicamp, 2015.

GALVAN, D. et al. Study of oxidation kinetics of B100 biodiesel from soybean and pig fat: activation energy determination. *Química Nova*, v. 37, n. 2, 2014.

GONÇALVES, Vinicius Pires. *Monitoramento da Estabilidade Oxidativa de Biodiesel Empregando Espectroscopia V* MONITORAMENTO DA ESTABILIDADE OXIDATIVA DE BODIESEL EMPREGANDO ESPECTROSCOPIA VIBRACIONAL ASSOCIADA A FERRAMENTAS QUIMIOMÉTRICAS. 2022. 79 f. Tese (Doutorado) - Curso de Química, Departamento de Química, Universidade Federal de Minas Gerais, Belo Horizonte, 2022.

GONDIM, C. et al. An appropriate and systematized procedure for validating qualitative methods: Its application in the detection of sulfonamide residues in raw milk. *Analytica Chimica Acta*, v. 830, p. 11–22, 1 jun. 2014.

GRIFFITHS, Peter R.; DE HASETH, James A. *Practical Raman Spectroscopy: An Introduction*. New York: Springer, 1986.

HAIBO HE et al. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. Disponível em:

<https://ieeexplore.ieee.org/abstract/document/4633969?casa_token=J_CENnbbg04AAAAA:H66LkaQgQseWdiBmYNY3Puy0nrHpFfZ7OA3lo7ZXVSCE-0_bXw-pmbIGkrE7HoIISMjkQqG7Ng>.

HONGYU, K.; SANDANIELO, V. L. M.; JUNIOR, G. J. DE O. Análise de Componentes Principais: Resumo Teórico, Aplicação e Interpretação. *E&S Engineering and Science*, v. 5, n. 1, p. 83–90, 29 jun. 2016.

- JONES, R. R. et al. Raman Techniques: Fundamentals and Frontiers. *Nanoscale Research Letters*, v. 14, n. 1, 12 jul. 2019.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, v. 23, n. 3, p. 187–200, set. 1958.
- KENNARD, R. W.; STONE, L. Computer Aided Design of Experiments. *Technometrics*, v. 11, n. 1, p. 137–148, 1 fev. 1969.
- KNOTHE, G.; GERPEN, J. V.; KRAHL, J.; RAMOS, L.P. Manual do Biodiesel. Traduzido do original “The Biodiesel Handbook” por Luiz Pereira Ramos, São Paulo: Edgard Blücher, 2006.
- LÔBO, I. P.; FERREIRA, S. L. C.; CRUZ, R. S. DA. Biodiesel: parâmetros de qualidade e métodos analíticos. *Química Nova*, v. 32, n. 6, p. 1596–1608, 2009.
- MA, F.; HANNA, M. A. Biodiesel production: a review. *Bioresource Technology*, v. 70, n. 1, p. 1–15, out. 1999.
- MÁRQUEZ, C. et al. FT-Raman and NIR spectroscopy data fusion strategy for multivariate qualitative analysis of food fraud. *Talanta*, v. 161, p. 80–86, 1 dez. 2016.
- MATOUSEK, P. Spatially offset Raman spectroscopy for non-invasive analysis of turbid samples. *TrAC Trends in Analytical Chemistry*, v. 103, p. 209–214, jun. 2018.
- MEIRA, M. et al. Prediction of properties of biodiesel-diesel blends using spectrofluorimetry and multivariate calibration. *Open Chemistry*, v. 10, n. 4, p. 1328–1337, 29 maio 2012.
- MÉNDEZ, E. et al. Validation of the rancimat test for the assessment of the relative stability of fish oils. *Journal of the American Oil Chemists’ Society*, v. 73, n. 8, p. 1033–1037, ago. 1996.
- MOSCA, S. et al. Spatially offset Raman spectroscopy. *Nature Reviews Methods Primers*, v. 1, n. 1, p. 1–16, 11 mar. 2021.
- MOITA NETO, J. M.; MOITA, G. C. Uma introdução à análise exploratória de dados multivariados. *Química Nova*, v. 21, n. 4, p. 467–469, jul. 1998.
- MOURA, N. N., Dutra, R. B. C., SOARES, I., MOURA, N. N., DUTRA, R. B. C., SOARES, I. P., et al. (2016). Avaliação da estabilidade oxidativa do b100 com o uso

de aditivos comerciais e extrativos. III Encontro de Pesquisa e Inovação da Embrapa Agroenergia: Anais.

NUNES, K. et al. Raman spectroscopy and discriminant analysis applied to the detection of frauds in bovine meat by the addition of salts and carrageenan. v. 147, p. 582–589, 1 jun. 2019.

PAVIA, Donald. L.; LAMPMAN, Gary. M.; KRIZ, George. S.; VYVYAN, James R.; Introduction to spectroscopy. 3 ed. São Paulo: Editora Thomson Learning, 2001.

PAZ, M. et al. Evolution of the supercritical extraction of oilseeds matrix cultivated in brasil: a review. International Seven Journal of Health Research, v. 2, n. 4, 14 jul. 2023.

PELISSON, Leidimara. Produção de biodiesel por meio de fluidos supercríticos e sua caracterização utilizando Cromatografia Gasosa de Alta Resolução. 2013. 102 f. Tese (Doutorado) - Curso de Química, Universidade de São Paulo, São Carlos, 2013.

PULLEN, J.; SAEED, K. An overview of biodiesel oxidation stability. Renewable and Sustainable Energy Reviews, v. 16, n. 8, p. 5924–5950, 1 out. 2012.

QUINLAN, J. R. (2014). C4. 5: programs for machine learning. Elsevier.

RAMALHO, V. C.; JORGE, N. Antioxidantes utilizados em óleos, gorduras e alimentos gordurosos. Química Nova, v. 29, n. 4, jul. 2006.

RAMOS, L. P. et al. Biodiesel Production Technologies. Revista Virtual de Química, v. 3, n. 5, 2011.

REIS, M. C. et al. BIODIESEL PRODUCTION FROM FATTY ACIDS OF REFINED VEGETABLE OILS BY HETEROGENEOUS ACID CATALYSIS AND MICROWAVE IRRADIATION. Química Nova, 2015.

RINALDI, R. et al. Síntese de biodiesel: uma proposta contextualizada de experimento para laboratório de química geral. Química Nova, v. 30, n. 5, p. 1374–1380, out. 2007.

RINGNÉR, M. What is principal component analysis? Nature Biotechnology, v. 26, n. 3, p. 303–304, mar. 2008.

RODRIGUES, A. D. G.; GALZERANI, J. C. Espectroscopias de infravermelho, Raman e de fotoluminescência: potencialidades e complementaridades. *Revista Brasileira de Ensino de Física*, v. 34, n. 4, dez. 2012.

SANTANA, Felipe Bachion de. FLORESTA ALEATÓRIA PARA DESENVOLVIMENTO DE MODELOS MULTIVARIADOS DE CLASSIFICAÇÃO E REGRESSÃO EM QUÍMICA ANALÍTICA. 2020. 139 f. Tese (Doutorado) - Curso de Química, Instituto de Química, Universidade Estadual de Campinas, Campinas, 2020.

SCHRADER, B. Raman Spectrometers. *Encyclopedia of Spectroscopy and Spectrometry*, p. 900–905, 2017.

SOUSA, E R. et al. Evaluation of Oxidation and Quality Parameters of Babassu Oil by Medium Infrared Spectroscopy with Fourier Transform (FTIR) and Multivariate Calibration. *Revista Virtual de Química*, v. 11, n. 3, p. 849-865, 2019.

SOUZA, F. DE B. et al. Avaliação do ácido láctico intramuscular através da espectroscopia Raman: novas perspectivas em medicina do esporte. *Revista Brasileira de Medicina do Esporte*, v. 9, n. 6, p. 388–395, nov. 2003.

SOUZA, V. et al. Physicochemical Characterization of Commercial Biodiesel/Diesel Blends and Evaluation of Unconventional Spectroscopic Vibrational Techniques in the Monitoring of Their Oxidation and Hydrolysis during Storage. *Energy & Fuels*, v. 30, n. 10, p. 8399–8409, 8 set. 2016.

VALERIANO, M. C. et al. Raman spectra soft modeling of the biodiesel oxidation through evolving factor analysis & multivariate curve resolution. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, v. 318, p. 124498, out. 2024.

VANDENABEELE, P. *Practical Raman Spectroscopy – An Introduction*. Wiley, v. 1, n. 6, p. 60-99, jul. 2013.

VELVARSKÁ, R. et al. Near-infrared spectroscopy for determining the oxidation stability of diesel, biodiesel and their mixtures. *Chemical Papers*, v. 73, n. 12, p. 2987–2993, 25 jun. 2019.

WANG, Q. et al. Random Forest with Self-Paced Bootstrap Learning in Lung Cancer Prognosis. *ACM Transactions on Multimedia Computing, Communications, and Applications*, v. 16, n. 1s, p. 1–12, 31 jan. 2020.

WOLD, S.; SJÖSTRÖM, M.; ERIKSSON, L. Partial Least Squares Projections to Latent Structures (PLS) in Chemistry. Encyclopedia of Computational Chemistry, 25 set. 1998.

APÊNDICE A – Espectros Raman complementares com destaque da banda em 1100 cm^{-1}

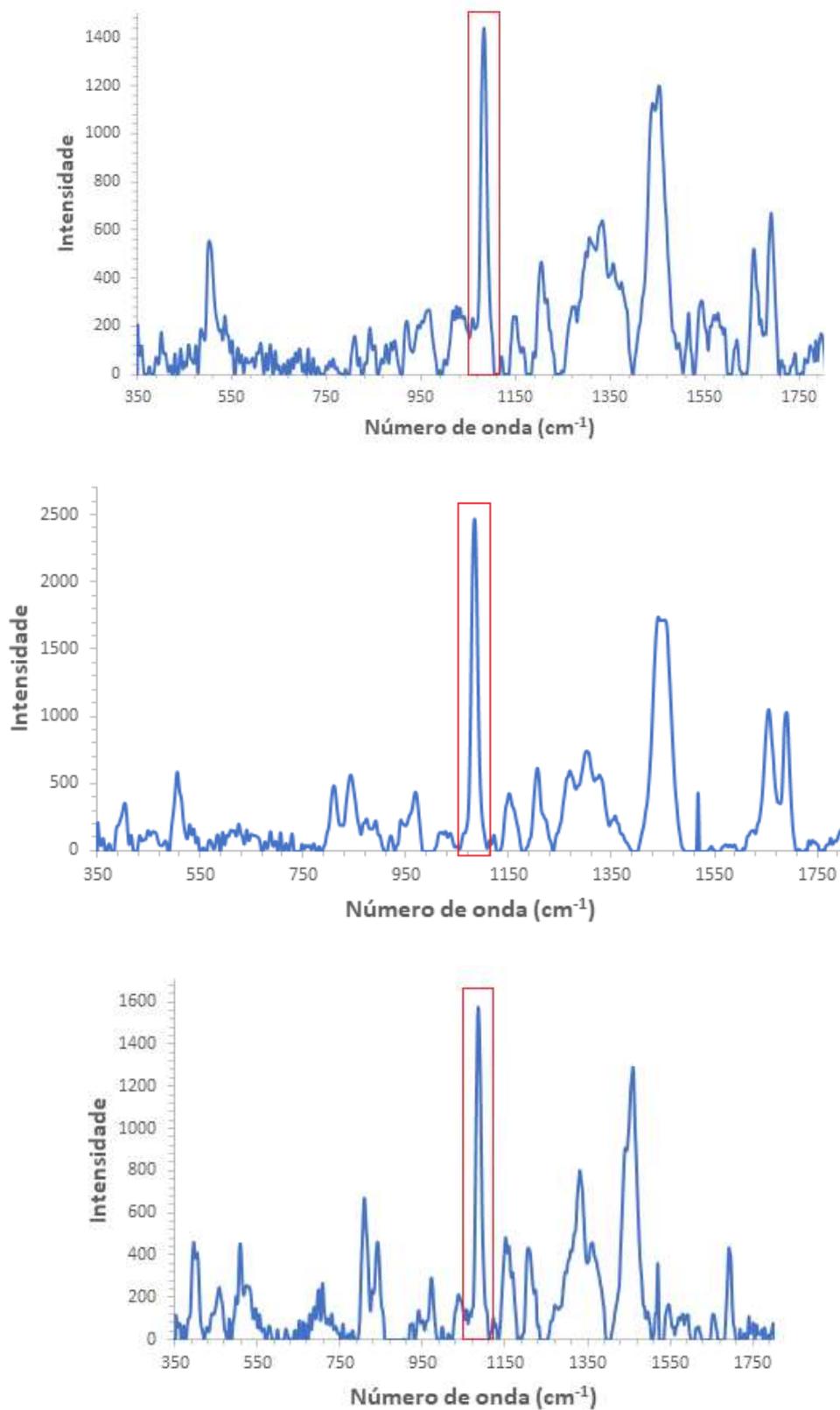


Figura 1A. Espectros Raman das amostras localizadas no extremo positivo da PC2, para o conjunto analisado no Equipamento A.

APÊNDICE B – Espectro Raman complementar com destaque da banda em 800 cm^{-1}

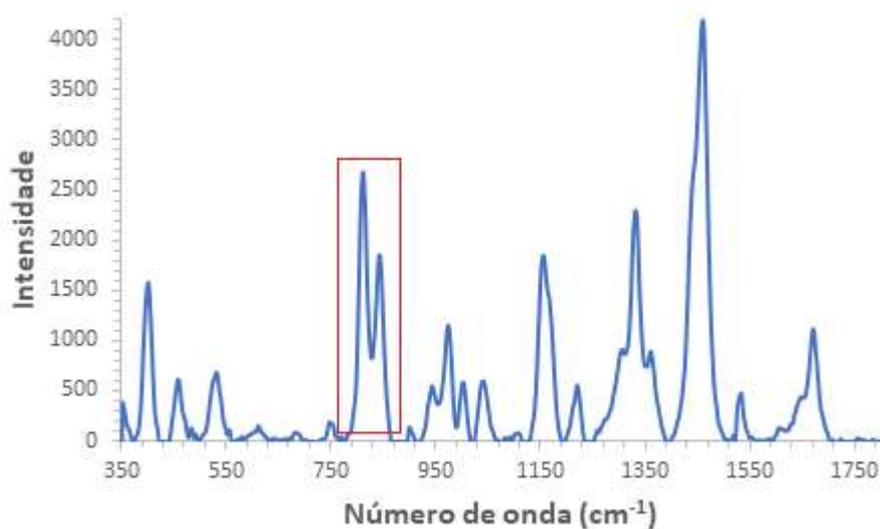


Figura 2A. Espectro Raman da amostra localizada no extremo positivo da PC3, para o conjunto analisado no Equipamento A.

APÊNDICE C – Matrizes de confusão dos modelos floresta aleatória dos conjuntos de treinamento

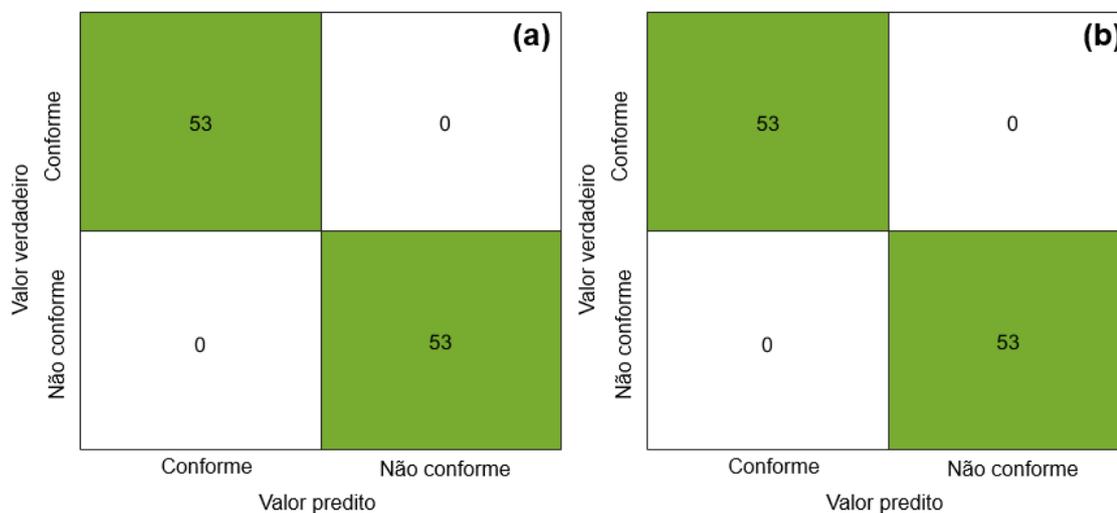


Figura 3A. Matriz de confusão do conjunto de treinamento da floresta aleatória, dados balanceados pelo ADASYN. a. Equipamento A e b. Equipamento B.