

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

APLICAÇÃO DE FERRAMENTAS QUIMIOMÉTRICAS E TÉCNICAS ESPECTROSCÓPICAS
NA ANÁLISE DE COMBUSTÍVEIS

CARLA FELIPPI CHIELLA RUSCHEL

TESE DE DOUTORADO

ORIENTADOR: PROF. DR. MARCO FLÔRES FERRÃO

CO-ORIENTADOR: PROF. DR. DIMITRIOS SAMIOS

2017

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE QUÍMICA

PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA

APLICAÇÃO DE FERRAMENTAS QUIMIOMÉTRICAS E TÉCNICAS ESPECTROSCÓPICAS
NA ANÁLISE DE COMBUSTÍVEIS

Tese apresentada como requisito parcial
para a obtenção do grau de Doutor em Química

Orientador: Prof. Dr. Marco Flôres Ferrão

Co-Orientador: Prof. Dr. Dimitrios Samios

2017

DEDICATÓRIA

Dedico este trabalho ao meu esposo Eder Ruschel como uma forma de gratidão pelo incondicional apoio e incentivo, sem o qual teria sido impossível a conclusão de mais esta etapa acadêmica.

AGRADECIMENTOS

Ao meu marido Eder Ruschel, pelo carinho, apoio, incentivo e compreensão neste longo período de aprendizagem.

Ao meu orientador, Professor Doutor Marco Flôres Ferrão, pela oportunidade de trabalharmos juntos, pela sua amizade, confiança, respeito e dedicação, por dividir sua experiência e compartilhar seus ensinamentos ao longo de todo esse tempo de trabalho mútuo.

Ao meu co-orientador Professor Doutor Dimitrios Samios, pela colaboração, pelos ensinamentos e sugestões.

A banca, pela disposição em avaliar e colaborar com sugestões e correções que qualificaram ainda mais o trabalho final.

Aos funcionários do CECOM/LABCOM-UFRGS, pelas análises referentes às especificações padrão da ANP, pelo carinho, respeito, amizade e gentileza, especialmente à Alexandro Helgueira, Ana Maria Nucci e Marcos Oliveira da Rosa.

À Professora Doutora Denise Azambuja pela orientação no estágio docente.

Aos Professores Doutor Aloir Merlo, Doutor João Henrique Zimnoch dos Santos e Doutora Tânia Pizzolato e as doutorandas Luma Fritsch e Rafaela Raupp da Rosa pela valiosa ajuda, empréstimo de equipamentos, computadores e laboratório.

Ao Professor Doutor Francisco Paulo dos Santos pela colaboração, dedicação, confiança, treinamento, ensinamentos, empréstimo de material e imensa ajuda com as análises de RMN.

Aos colegas de estudo, de laboratório, amigos, professores e ex-colegas de laboratório, pela imensa ajuda e colaboração, pela paciência, pela convivência harmoniosa, carinho e respeito, amizade e companheirismo, em especial à Camila Correa, Marcelo Marcelo, Bernardo Vacaro, Eduardo Lopes, Fabio Weiler, Andre Rech, Fernanda Bock, Juliana da Costa, Layane Lenardon, Manuella Schneider, Gilson Helfer, Adilson da Costa, Camila Cerveira, Pâmela Barela, Katherine Kellermann, Jucelaine Poletti, Natália Medeiros, Diogo Pompéu de Moraes e Janaína Barbará.

A todos os amigos e familiares que torceram pelo meu sucesso neste curso.

À Universidade Federal do Rio Grande do Sul e funcionários do Instituto de Química pela grande colaboração, auxílio, empréstimo de material, laboratório e equipamentos, em especial, aos técnicos Marcia Weissheimer, Larissa Rodrigues, Jorge Candido, Regis Machado e ao motorista José Carlos Martins pelo transporte até a REFAP.

À REFAP S.A. pelas amostras de combustíveis.

À CAPES pela bolsa de estudos.

PRODUÇÃO CIENTÍFICA

1. Artigos publicados gerados a partir dos resultados da tese

1.1 Carla F. C. Ruschel, Marco F. Ferrão, Francisco P. Dos Santos, Dimitrios Samios. Otimização do processo de transesterificação em duas etapas para produção de biodiesel através do planejamento experimental Doehlert. *Química Nova*. Vol. 39, Nº 3, Págs. 267-272, 2016.

1.2 Carla F. C. Ruschel, Marcelo C. A. Marcelo, Marco F. Ferrão, Francisco P. Dos Santos. Determination of gasoline adulteration using ^1H NMR spectroscopy and chemometrics techniques. *Analytical Letters*. No prelo.

2. Demais artigos publicados durante o doutorado

2.1 Carla Felippi Chiella Ruschel, Chun Te Huang, Dimitrios Samios, Marco Flôres Ferrão. Análise exploratória aplicada a espectros de reflexão total atenuada no infravermelho com transformada de Fourier (ATR-FTIR) de blendas de biodiesel/diesel. *Química Nova*. Vol. 37, Nº 5, Págs. 810-815, 2014.

2.2 Carla Felippi Chiella Ruschel, Chun Te Huang, Dimitrios Samios, Marco Flôres Ferrão, Carlos Itsuo Yamamoto, Rubia Carla Barato Plochanski. Determinação do Número de Cetano de Blendas de Biodiesel/Diesel Utilizando Espectroscopia no Infravermelho Médio e Regressão Multivariada. *Orbital. The Electronic Journal of Chemistry*. Vol 6, Nº 1, Págs. 39-46, 2014.

2.3 Carla Felippi Chiella Ruschel, Chun Te Huang, Dimitrios Samios, Marco Flôres Ferrão, Carlos Itsuo Yamamoto & Rubia Carla Barato Plochanski. Environmentally Friendly Determination of Quality Parameters of Biodiesel/Diesel Blends Using Fourier Transform Infrared Spectra. *Journal of American Oil Chemistry Society*. Vol. 92, Nº 3, Págs. 309-315, 2015.

2.4 Camila Correa, Carla Felippi Chiella Ruschel, Fátima M. Bento, Marco Flôres Ferrão. Aplicação de Cartas de Controle Multivariadas no Controle e Garantia da Qualidade do Biodiesel (B100). *Revista Virtual de Química*. Vol. 7, Nº 6, Págs. 2273-2289, 2015.

3. Trabalhos apresentados em eventos

3.1 Método Verde Para Determinação Da Densidade De Blendas De Diesel/Biodiesel Através De FTIR-HATR e Regressão Multivariada. Carla. F. C. Ruschel. XX Encontro de Química da Região Sul, Novembro de 2013. Lajeado, RS.

3.2 Carla F. C. Ruschel, Chun T. Huang, Marco F. Ferrão, Dimitrios Samios. Análise Exploratória de Blendas Diesel/Biodiesel a partir de Dados de Espectroscopia por Reflexão Total Atenuada no Infravermelho Médio. I Escola de Inverno de Quimiometria, Agosto de 2013, São Carlos, S.P.

3.3 Carla Felippi Chiella Ruschel, Chun Te Huang, Dimitrios Samios, Marco Flôres Ferrão. Determinação do Número de Cetano de Blendas Biodiesel/Diesel por Infravermelho e Regressão Multivariada com Seleção de Variáveis. 17º ENQA, Outubro, 2013, Belo Horizonte, M.G.

3.4 Carla Felippi Chiella Ruschel, Chun Te Huang, Dimitrios Samios, Marco Flôres Ferrão, Carlos Itsuo Yamamoto, Rubia Carla Barato Plocharski. Determinação da massa específica de blendas de biodiesel/Diesel utilizando FTIR e regressão multivariada. II Escola de Inverno de Quimiometria, Julho de 2015. Campinas, S.P.

3.5 Carla Felippi Chiella Ruschel, Francisco Paulo dos Santos, Dimitrios Samios, Marco Flôres Ferrão. Optimization of Transesterification Double Step Process to the Production of Biodiesel using Doehlert Design. 8º Congresso Argentino de Quimica Analítica. Novembro de 2015, La Plata, Argentina.

3.6 Carla Felippi Chiella Ruschel, Marcelo Caetano Alexandre Marcelo, Francisco Paulo dos Santos, Marco Flôres Ferrão. Determinação das adulterações de gasolina com óleo diesel e de etanol com água empregando ^1H RMN e quimiometria. 18º Encontro Nacional de Química Analítica, Setembro de 2016, Florianópolis, S.C.

SUMÁRIO

LISTA DE FIGURAS.....	X
LISTA DE TABELAS.....	XIII
LISTA DE ABREVIATURAS.....	XIV
RESUMO.....	XVI
ABSTRACT.....	XVIII
1. REVISÃO BIBLIOGRÁFICA.....	1
1.1 Aspectos Econômicos e Conceituais dos Combustíveis e Biocombustíveis.....	1
1.2 Produção de Biodiesel.....	3
1.2.1 A Reação de Transesterificação.....	4
1.2.1.1 Catálise Básica.....	5
1.2.1.2 Catálise Ácida.....	5
1.2.1.3 A metodologia Transesterification Double Step Process (TDSP).....	6
1.3 Adulteração de Combustíveis.....	7
1.3.1 Controle de Qualidade dos Combustíveis.....	8
1.4 Técnicas Espectroscópicas.....	9
1.4.1 Espectroscopia de Ressonância Magnética Nuclear de Hidrogênio (RMN ¹ H).....	9
1.4.2 Espectroscopia de Infravermelho com Transformada de Fourier (FTIR).....	14
1.5 Ferramentas Quimiométricas.....	19
1.5.1 Planejamento Experimental ou Matriz Doehlert.....	20
1.5.2 Análise por Componentes Principais ou Principal Component Analysis (PCA).....	23
1.5.3 Análise de Agrupamentos Hierárquicos ou Hierarchical Cluster Analysis (HCA).....	25
1.5.4 Regressão pelo método de Mínimos Quadrados Parciais ou Partial Least Squares (PLS).....	26
1.5.5 Análise Discriminante por Mínimos Quadrados Parciais ou Partial Least Squares Discriminant Analysis (PLS-DA).....	27
1.5.6 Máquinas de Vetor de Suporte ou Support Vector Machines (SVM).....	28
1.5.7 Cartas de Controle Multivariadas ou Multivariate Control Charts (CCM).....	31
1.5.7.1 Cartas de Controle com base no Sinal Analítico Líquido (NAS).....	31

1.5.7.2 Construção das Cartas de Controle	32
2. OTIMIZAÇÃO DO PROCESSO TDSP PARA PRODUÇÃO DE BIODIESEL EMPREGANDO PLANEJAMENTO EXPERIMENTAL DOEHLERT	38
2.1 Objetivo.....	38
2.2 Materiais e métodos.....	38
2.2.1 Reagentes	38
2.2.2 Softwares	38
2.3 Planejamento Doehlert.....	38
2.3.1 Produção do biodiesel	39
2.3.2 Caracterização do biodiesel por ressonância magnética nuclear de hidrogênio (RMN ¹ H)	40
2.4 Resultados e Discussão	44
2.4.1 Análise da conversão para catálise básica segundo planejamento Doehlert.....	44
2.4.2 Análise da conversão para catálise ácida segundo planejamento Doehlert	45
2.4.3 Análise do gráfico de probabilidade normal e da superfície de resposta	46
3. DETERMINAÇÃO DA ADULTERAÇÃO DA GASOLINA EMPREGANDO RMN ¹ H E MÉTODOS QUIMIOMÉTRICOS	51
3.1 Objetivo.....	51
3.2 Materiais e métodos.....	51
3.2.1 Softwares	52
3.2.2 Obtenção e Alinhamento dos Espectros de RMN ¹ H	52
3.3 Resultados e Discussão	53
3.3.1 PCA e HCA	53
3.3.2 PLS-DA e PLS.....	55
3.3.2.1 Definição dos conjuntos	55
3.3.2.2 Sensibilidade e Especificidade	56
4. AVALIAÇÃO DA ADULTERAÇÃO DE COMBUSTÍVEIS EMPREGANDO FTIR-ATR E MÉTODOS QUIMIOMÉTRICOS.....	60
4.1 Objetivo.....	60

4.2 Materiais e métodos.....	60
4.2.1 Espectroscopia de FTIR-ATR.....	63
4.2.2 Softwares	65
4.2.3 Definição dos conjuntos	65
4.3 Resultados e discussão	66
4.3.1 Análise Exploratória	66
4.3.2 Métodos de Classificação	68
4.3.3 Métodos de Regressão Multivariada	72
4.3.4 Cartas de Controle Multivariadas	75
5. CONCLUSÕES	80
6. REFERÊNCIAS	82

LISTA DE FIGURAS

Figura 1. a) Estados de spin para o próton na ausência do campo magnético externo aplicado e com campo magnético externo aplicado, b) Processo de absorção da ressonância pelo próton.....	10
Figura 2. Representação do fenômeno da ressonância magnética nuclear com absorção da radiação quando $\nu = \omega$	12
Figura 3. a) Exemplo de uma FID (sinal no domínio do tempo) b) Exemplo de um sinal no do espectro no domínio da frequência. Adaptado de Macomber ³³	14
Figura 4. Parte do espectro eletromagnético que relaciona o infravermelho com outros tipos de radiação mais próximos ²⁹	15
Figura 5. Exemplo de modos de estiramento e dobramento das moléculas.....	16
Figura 6. Esquema representativo de um espectrômetro de infravermelho com transformada de Fourier (FTIR) ²⁹	17
Figura 7. Esquema representativo de um acessório de refletância total atenuada (ATR).....	18
Figura 8. Figura geométrica representativa da matriz Doehlert para sistema com duas variáveis ou fatores.....	21
Figura 9. Esquema das transformações realizadas pela PCA numa matriz de dados X.	25
Figura 10. Representação de um conjunto de dados classificados por SVM com hiperplano de separação ótima com margens e vetores de suporte.	29
Figura 11. Representação do hiperplano que define a função de regressão por SVR.	30
Figura 12. Representação da divisão do espectro r nos vetores: r^{\perp} (NAS+resíduo), r_{K*} (vetor NAS), r_{INT} (vetor interferente) e r_{RES} (vetor resíduo).	32
Figura 13. Valores codificados para planejamento Doehlert com ponto central em triplicata.	39
Figura 14. Esquema representativo da reação geral de transesterificação do óleo de soja com álcool primário (metanol) produzindo biodiesel e glicerol como subproduto.....	40
Figura 15. Espectro de RMN ¹ H de óleo de soja.	41

Figura 16. Espectro de RMN ¹ H de biodiesel de óleo de soja produzido pelo método TDSP (experimento 2 da etapa 1).	42
Figura 17. Gráfico de probabilidade normal para a massa de catalisador (g) e volume de metanol (mL) na primeira parte da reação.....	48
Figura 18. Superfície de resposta do planejamento Doehlert para a massa de catalisador (g) e volume de metanol (mL) na primeira parte da reação.	48
Figura 19. Espectros de RMN ¹ H das misturas de gasolina alinhados pelo algoritmo RAFFT.....	53
Figura 20. Dendrograma dos dados espectrais de RMN ¹ H das misturas de gasolina.	54
Figura 21. Análise de Componentes Principais de espectros de RMN ¹ H das misturas de gasolina.	55
Figura 22. Classificação dos espectros de RMN ¹ H das amostras de gasolina por PLS-DA.....	56
Figura 23. Escores VIP obtidos na análise de PLS-DA do conjunto de espectros de RMN ¹ H.....	57
Figura 24. Modelo PLS/RMN ¹ H para previsão do percentual de de óleo diesel como adulterante nas misturas de gasolina e etanol.	58
Figura 25. a) espectros de etanol não adulterado (0,5 a 7,5% (m/m) de água), b) espectros de etanol adulterado com 8 a 25% (m/m) de água, c) espectros de misturas de gasolina com 27% + 1 (v/v) de etanol anidro não adulteradas, d) espectros de gasolina com 27% + 1 (v/v) de etanol anidro adulteradas com 5 a 30% (v/v) de óleo diesel.	63
Figura 26. Espectros das misturas de gasolina com pré-tratamento por Savitzky - Golay utilizando polinômio de primeiro grau com uma janela de 15 pontos.	64
Figura 27. a) Região espectro de infravermelho utilizada para a construção das CCM para as amostras de etanol, b) Mesma região após o pré-processamento: suavização, através do algoritmo de Savitsky-Golay utilizando uma janela de 15 pontos e polinômio de segundo grau, normalização, aplicação de variação normal padrão (SNV) e por fim foram derivada por Savitsky-Golay com uma janela de 15 pontos e polinômio de segundo grau.	65

Figura 28. Análise de agrupamento hierárquico dos espectros de FTIR-ATR a) das misturas de etanol, b) das misturas de gasolina.	67
Figura 29. Análise de por componentes principais dos espectros de FTIR-ATR a) das misturas de etanol, b) das misturas de gasolina.	68
Figura 30. Classificação por PLS-DA dos espectros de FTIR-ATR do conjunto de amostras de etanol.	69
Figura 31. Classificação por PLS-DA dos espectros de FTIR-ATR do conjunto de amostras de gasolina.	70
Figura 32. Classificação por SVM-DA dos espectros de FTIR-ATR do conjunto de amostras de etanol.	71
Figura 33. Classificação por SVM-DA dos espectros de FTIR-ATR do conjunto de amostras de gasolina.	72
Figura 34. Quantificação dos adulterantes: a) água em etanol hidratado combustível, b) óleo diesel na mistura de gasolina e etanol pelo modelo PLS utilizando dados de FTIR-ATR.	73
Figura 35. Quantificação dos adulterantes: a) água em etanol hidratado combustível, b) óleo diesel na mistura de gasolina e etanol pelo modelo SVM utilizando dados de FTIR-ATR.	74
Figura 36. Cartas de controle multivariadas baseadas em dados de FTIR-ATR para determinação de adulteração de misturas de etanol. A primeira carta é do sinal de analítico líquido (NAS) (água), a carta intermediária é do interferente ou matriz da amostra (etanol) e a última carta é dos resíduos.	76
Figura 37. Cartas de controle multivariadas baseadas em dados de FTIR-ATR para determinação de adulteração de misturas de gasolina e etanol. A primeira carta é do sinal de analítico líquido (NAS) (etanol), a carta intermediária é do interferente ou matriz da amostra (gasolina) e a última carta é dos resíduos que é o adulterante (óleo diesel).	78

LISTA DE TABELAS

Tabela I. Matriz de Doehlert com duas variáveis	21
Tabela II. Valores de conversão obtidos utilizando condições reacionais testadas para catálise básica	43
Tabela III. Valores de conversão obtidos utilizando condições reacionais testadas para catálise ácida segundo planejamento Doehlert	43
Tabela IV. Análise de Variância - Modelo Quadrático para catálise básica	47
Tabela V. Análise de Variância - Modelo Linear para catálise ácida	49
Tabela VI. Análise de Variância - Modelo Quadrático para catálise ácida	49
Tabela VII. Percentual dos componentes nas misturas de gasolina	52
Tabela VIII. Percentual de cada componente das misturas de etanol	61
Tabela IX. Percentual de cada componente das misturas de gasolina	62
Tabela X. Parâmetros dos modelos PLS-DA e SVM-DA para misturas de etanol e gasolina	70
Tabela XI. Parâmetros do modelo PLS e SVM para misturas de etanol e gasolina	72

LISTA DE ABREVIATURAS

ABNT	ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS
AEAC	ÁLCOOL ETÍLICO ANIDRO COMBUSTÍVEL
AEHC	ÁLCOOL ETÍLICO HIDRATADO COMBUSTÍVEL
ANOVA	ANÁLISE DE VARIÂNCIA
ANP	AGÊNCIA NACIONAL DE PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS
ASTM	<i>AMERICAN SOCIETY FOR TESTING AND MATERIALS</i>
ATR	<i>ATTENUATED TOTAL REFLECTANCE</i> OU REFLECTÂNCIA TOTAL ATENUADA
B100	BIODIESEL
BDEP	BANCO DE DADOS DE EXPLORAÇÃO E PRODUÇÃO
CCM	CARTAS DE CONTROLE MULTIVARIADAS
CDCl ₃	CLOROFÓRMIO DEUTERADO
C _{ME}	CONVERSÃO EM ÉSTER METÍLICO
CH ₃ OH	ÁLCOOL METÍLICO
FID	DECAIMENTO DA INDUÇÃO LIVRE OU <i>FREE INDUCTION DECAY</i>
FTIR	<i>FOURIER TRANSFORM INFRARED</i>
HCA	<i>HIERARCHICAL CLUSTER ANALYSIS</i>
IV	INFRAVERMELHO
I _{ME}	INTEGRAL DO SINAL DO ÉSTER METÍLICO
I _{α-CH₂}	INTEGRAL DO GRUPO METILENO ADJACENTE AO GRUPO ÉSTER
KOH	HIDRÓXIDO DE POTÁSSIO
KNN	<i>K-NEAREST NEIGHBOUR</i>
ME	ÉSTER METÍLICO
NAS	SINAL ANALÍTICO LÍQUIDO OU <i>NET ANALYTE SIGNAL</i>
PC	<i>PRINCIPAL COMPONENT</i> OU COMPONENTE PRINCIPAL
PCA	<i>PRINCIPAL COMPONENT ANALYSIS</i>
PCR	<i>PRINCIPAL COMPONENT REGRESSION</i>
PLS	<i>PARTIAL LEAST SQUARES</i>
PLS-DA	<i>PARTIAL LEAST SQUARES - DISCRIMINANT ANALYSIS</i>
PMQC	PROGRAMA DE MONITORAMENTO DA QUALIDADE DOS COMBUSTÍVEIS

RMN ¹ H	RESSONÂNCIA MAGNÉTICA NUCLEAR DE HIDROGÊNIO
SIMCA	<i>SOFT INDEPENDENT MODELLING OF CLASS ANALOGY</i>
SVM	<i>SUPPORT VECTOR MACHINE</i>
SVM-DA	<i>SUPPORT VECTOR MACHINE - DISCRIMINANT ANALYSIS</i>
SVR	SUPPORT VECTOR REGRESSION
TDSP	<i>TRANSESTERIFICATION DOUBLE STEP PROCESS</i>
TMS	TETRAMETILSILANO
UNICAMP	UNIVERSIDADE ESTADUAL DE CAMPINAS

RESUMO

A presente tese de doutorado visou primeiramente em otimizar a metodologia denominada *Transesterification Double Step Process* (TDSP) para a produção de biodiesel. O biodiesel foi produzido a partir de óleo de soja, pela rota metílica, empregando diferentes condições reacionais a partir do planejamento experimental Doehlert. O produto foi caracterizado por espectroscopia de ressonância magnética nuclear de hidrogênio (RMN ^1H) o que permitiu também o cálculo da conversão do material de partida em biodiesel. Foram utilizados ainda gráficos de probabilidade normal e de superfícies de resposta para avaliar a influência das variáveis consideradas no planejamento, como a quantidade de catalisador e o volume de metanol, além da temperatura e tempo de reação. Para a etapa da catálise básica da reação a melhor conversão foi obtida utilizando-se 0,40 g de catalisador e 60 mL de metanol enquanto para a catálise ácida a condição inicial com temperatura de 65 °C e tempo de reação de 90 minutos utilizada no método TDSP se manteve como a melhor condição de reação. Em um segundo momento, foram produzidos conjuntos de amostras de gasolina e etanol, com amostras adulteradas e não-adulteradas, que foram caracterizados através de RMN ^1H e de espectroscopia no infravermelho com transformada de Fourier com acessório de reflectância total atenuada (FTIR-ATR), para posterior aplicação de ferramentas quimiométricas aos dados espectroscópicos gerados. Aos dados de RMN ^1H foram aplicados modelos não supervisionados como a análise por agrupamentos hierárquicos (HCA) e a análise por componentes principais (PCA), modelos supervisionados como a análise discriminante empregando mínimos quadrados parciais (PLS-DA) e de regressão multivariada por mínimos quadrados parciais (PLS). HCA e PCA apresentaram uma tendência de separação entre as amostras adulteradas e não adulteradas. Além disso, o modelo PLS-DA foi capaz de classificar corretamente todas as amostras para ambas as classes. Já o algoritmo PLS apresentou um excelente modelo de regressão para quantificar o Diesel como adulterante da gasolina com valor de RMSECV igual a 2,32% (v/v), RMSEP igual a 1,42% (v/v) e R^2 igual a 0,992. Por fim, com os dados de infravermelho foram aplicados além de modelos por HCA, PCA, PLS e PLS-DA, modelos de regressão por máquinas de vetores de suporte (SVM) e de classificação com análise discriminante (SVM-DA) e cartas de controle multivariadas (CCM) também para avaliação de adulteração desses combustíveis. Para os modelos com base nos dados de infravermelho, no geral, todos os resultados mostraram que tais modelos são ótimas alternativas para distinguir as amostras adulteradas das não-adulteradas para os combustíveis estudados. Para as amostras de gasolina ambos modelos de classificação PLS-DA e SVM-DA apresentaram sensibilidade e especificidade máximas, o que não ocorreu para os modelos com as amostras de etanol. O modelo quantitativo de PLS para as amostras de etanol apresentou valor de RMSECV igual

a 1,721 % (m/m), RMSEP igual a 0,859 % (m/m) e R^2 igual a 0,985, enquanto o modelo PLS para as amostras de gasolina apresentou valor de RMSECV igual a 0,832 % (v/v), RMSEP igual a 0,543 % (v/v) e R^2 igual a 0,998. O modelo quantitativo de SVM para as amostras de etanol apresentou valor de RMSECV igual a 1,056 % (m/m), RMSEP igual a 0,379 % (m/m) e R^2 igual a 0,999, enquanto o modelo SVM para as amostras de gasolina apresentou valor de RMSECV igual a 1,084 % (v/v), RMSEP igual a 0,571 % (v/v) e R^2 igual a 0,997.

ABSTRACT

The present doctoral thesis aimed first optimizing the methodology called Transesterification Double Step Process (TDSP) for biodiesel production. The biodiesel was produced from soybean oil by the methyl route, using different reaction conditions from the Doehlert experimental design. The product was characterized by hydrogen nuclear magnetic resonance (^1H NMR) spectroscopy which also allowed the calculation of the conversion of the starting material to biodiesel. We also used normal probability and response surface graphs to evaluate the influence of the variables considered in the planning, such as the catalyst amount and the volume of methanol, besides the temperature and reaction time. For the basic catalysis stage of the reaction the best conversion was obtained using 0.40 g of catalyst and 60 mL of methanol while for the acid catalysis the initial condition with temperature of 65 °C and reaction time of 90 minutes used in the method TDSP remained the best reaction condition. Second, samples of gasoline and ethanol were produced with adulterated and unadulterated samples, which were characterized by ^1H NMR and Fourier transform infrared spectroscopy with attenuated total reflectance (FTIR-ATR), for later application of chemometric tools to the spectroscopic data generated. To the ^1H NMR data, unsupervised models by hierarchical cluster analysis (HCA) and principal component analysis (PCA) were applied. In addition supervised models such as discriminant analysis employing partial least squares (PLS-DA) and partial least squares multivariate regression (PLS) were also studied. HCA and PCA showed a tendency of separation between adulterated and unadulterated samples. In addition, the PLS-DA model was able to correctly classify all samples for both classes. The PLS algorithm presented an excellent regression model to quantify Diesel as a gasoline adulterant with RMSECV value equal to 2.32 % (v/v), RMSEP equal to 1.42 % (v/v) and R^2 equal to 0.992. Finally, with the infrared data, were applied models by HCA, PCA, PLS and PLS-DA, support vector machine regression (SVM) and with discriminant analysis (SVM-DA) and multivariate control charts (CCM) also to evaluate the adulteration of these fuels. At models based on infrared data, in general, all the results proved to be excellent alternatives to distinguish the adulterated samples from the unadulterated samples for the fuels studied. To the gasoline samples both PLS-DA e SVM-DA classification models presented maximum sensibility and specificity, which did not occur for models with ethanol samples. The quantitative PLS model to ethanol samples presented RMSECV value equal to 1.721 % (m/m), RMSEP equal to 0.859 % (m/m) and R^2 equal to 0.985, while the PLS model to gasoline samples presented RMSECV value equal to 0.832 % (v/v), RMSEP equal to 0.543 % (v/v) and R^2 equal to 0.998. The quantitative SVM model to ethanol samples presented RMSECV value equal to 1.056 % (m/m), RMSEP equal to 0.379 %

(m/m) and R^2 equal to 0.999, while the SVM model to gasoline samples presented RMSECV value equal to 1.084 % (v/v), RMSEP equal to 0.571 % (v/v) and R^2 equal to 0.997.

1. REVISÃO BIBLIOGRÁFICA

1.1 Aspectos Econômicos e Conceituais dos Combustíveis e Biocombustíveis

A importância dos combustíveis automotivos pode ser mensurada através dos seus dados de produção e consumo, sendo sabidamente um produto de interesse mundial. O volume de petróleo produzido no mundo em 2015 aumentou em 2,8 milhões de barris/dia (3,2%) em relação a 2014, passando de 88,8 milhões para 91,7 milhões de barris/dia. Já o consumo mundial, no ano de 2015, totalizou 95 milhões de barris/dia, após aumento de 2% (1,9 milhão de barris/dia) em comparação a 2014. No ranking de países que mais consumiram petróleo em 2015, os Estados Unidos estavam em primeiro lugar, com 19,4 milhões de barris/dia (20,4% do total mundial). Em seguida veio a China, com consumo médio de 12 milhões de barris/dia de petróleo (12,6% do total mundial), na terceira colocação ficou a Índia, com 4,2 milhões de barris/dia (4,4% do total mundial) e o Brasil alcançou o quinto lugar, com consumo de cerca de 3,2 milhões de barris/dia correspondendo a 3,3% do total mundial¹.

A gasolina, um dos combustíveis mais utilizados, é constituída basicamente por hidrocarbonetos e, em menor quantidade, por produtos oxigenados. Esses hidrocarbonetos são, em geral, mais leves do que aqueles que compõem o óleo diesel, pois são formados por moléculas de menor cadeia carbônica (4 a 12 átomos de carbono). Além destes, contém compostos de enxofre, de nitrogênio e compostos metálicos, todos em baixas concentrações e a sua faixa de destilação varia de 30 a 220 °C². Como a gasolina é uma mistura e não uma substância pura ela não tem um valor de temperatura de ebulição fixo, e sim uma faixa de temperatura na qual o líquido vaporiza. Os compostos mais leves evaporam preferencialmente a temperaturas mais baixas enquanto os compostos mais pesados necessitam de maior temperatura para destilar.

Este combustível no Brasil ainda se divide em dois tipos: Gasolina A, que é o combustível formulado por meio da mistura de correntes provenientes do refino de petróleo (gasolina pura), destinado aos veículos automotivos dotados de motores ciclo Otto, que se diferenciam dos motores do ciclo Diesel por necessitarem de uma fonte de ignição (no caso uma centelha) e a gasolina tipo C, que é aquela revendida nos postos, sendo obtida a partir da mistura de gasolina A com um percentual obrigatório de álcool anidro, atualmente de 27% (v/v)³.

O óleo Diesel, por sua vez, também é um combustível fóssil constituído basicamente por hidrocarbonetos, porém, estes são mais pesados do que os que compõem a gasolina, pois são

formados por moléculas com maior cadeia carbônica, com 9 a 35 átomos de carbono. Além disso, possui uma faixa de destilação superior à da gasolina, que vai de 140 a 380 °C. Nos motores a Diesel, diferentemente dos motores à gasolina, a ignição se dá por compressão, ou seja, o combustível e ar são comprimidos e sofrem ignição iniciando a sua combustão. O Diesel combustível comercializado nos postos se classifica em “Interior”, que é o Diesel de coloração amarela e pode conter no máximo 500 mg.L⁻¹ (ppm) de enxofre na sua composição. Já o Diesel dito “Metropolitano” recebe um corante vermelho para sua identificação e o limite máximo é de 10 mg.L⁻¹ (ppm) de enxofre na sua composição, pois a quantidade de automóveis por área nas capitais é maior do que no interior dos estados⁴. O controle dessa propriedade visa evitar a formação de compostos corrosivos durante a combustão, além de compostos poluentes como óxidos de enxofre, responsáveis pela chuva ácida além de material particulado. Estas partículas, devido ao seu tamanho muito pequeno, são facilmente inaladas e estão relacionadas com a incidência de problemas respiratórios.

No caso dos biocombustíveis, que incluem tanto o etanol quanto o biodiesel, segundo definição utilizada pela própria Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP), também houve aumento na sua produção e consumo. A produção total de etanol em 2015 cresceu 6,3%, totalizando 29,9 milhões de m³. Para o biodiesel a partir de novembro de 2014 houve um aumento na proporção adicionada ao óleo diesel para 7% (Lei nº 13.033/2014). Em 2015, a capacidade nominal para produção de biodiesel (B100) no Brasil era de cerca de 7,4 milhões de m³ (20,4 mil m³/dia), entretanto a produção nacional foi de 3,9 milhões de m³, o que correspondeu a somente 53,3% da capacidade total. Em comparação a 2014, a produção de biodiesel (B100) foi 15,1% maior¹.

O etanol combustível atualmente é destinado ao uso em motores “flex” que possui como principal componente o álcool etílico, especificado sob as formas de álcool etílico anidro combustível e de álcool etílico hidratado combustível. O álcool etílico anidro combustível (AEAC) ou etanol anidro combustível é destinado para compor a mistura com gasolina A na formulação da gasolina C. Já o álcool etílico hidratado combustível (AEHC) é o combustível destinado à venda no posto revendedor para o consumidor final, conforme especificação da ANP⁵. Historicamente pode-se considerar que o AEHC foi implementado no ano de 1975 durante o governo do presidente Ernesto Geisel com o lançamento do programa chamado de “Pró-álcool” que visava conter gastos com importação de petróleo e incentivar a produção do primeiro combustível brasileiro proveniente de fonte renovável, a cana de açúcar. Os primeiros carros movidos a álcool foram produzidos em 1979 em pequena escala, apenas 0,5% do total da frota em circulação no país. Já em 1986 foi o ápice desta tecnologia naquela época, alcançando 72,6% do total de automóveis produzidos naquele ano. Nos anos seguintes com o aumento da oferta e redução do preço do petróleo, a utilização da cana-de-

açúcar voltou a ser apenas para a produção de açúcar, gerando falta de álcool combustível e o setor sofreu uma crise de abastecimento. Nos anos 2000, com o surgimento dos carros bicombustíveis ou “flex” o consumo de álcool aumentou novamente, mantendo-se até hoje^{2,6}.

Outro biocombustível muito importante atualmente adicionado ao óleo Diesel comercial, o biodiesel, diferentemente da gasolina e do óleo Diesel, provenientes de fontes petroquímicas, é constituído por ésteres alquílicos provenientes de ácidos graxos de cadeia longa, produzidos a partir da reação de óleos ou gorduras de origem vegetal ou animal, com álcoois de cadeia curta como o metanol, que é o mais empregado, ou alternativamente o etanol⁷. O óleo de soja foi a principal matéria-prima utilizada na produção de biodiesel no Brasil compondo 77,7% do total, em 2015. A segunda matéria-prima no ranking de produção das usinas brasileiras foi a gordura animal com 18,8% do total, seguida por óleo de algodão, com 2% do total, e o restante, 1,5%, é produzido a partir de outras fontes. O interesse na utilização de biodiesel como substituinte parcial ou total em misturas com o óleo Diesel petroquímico se deve as suas propriedades semelhantes, porém, o biodiesel se destaca pela biodegradabilidade, menor toxicidade, além de poder ser produzido a partir de diversas fontes renováveis e pela menor geração de poluentes atmosféricos^{1,8,9}.

1.2 Produção de Biodiesel

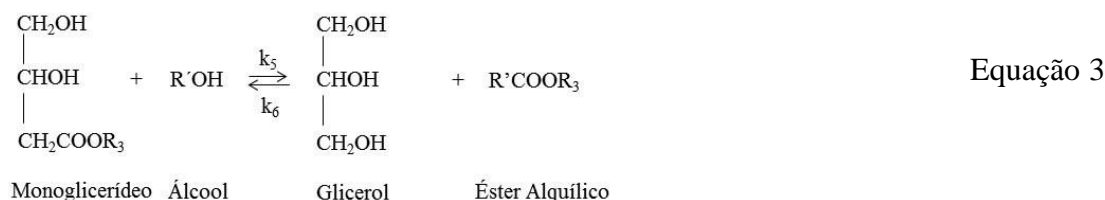
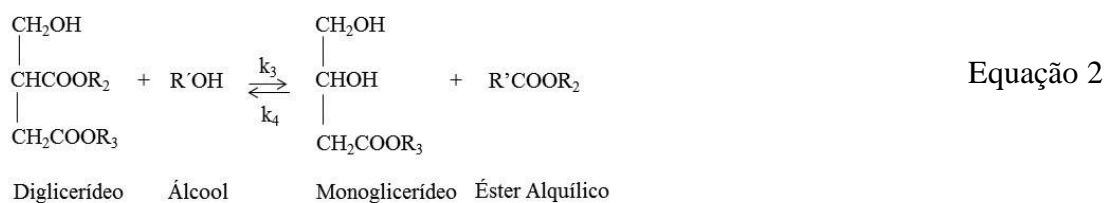
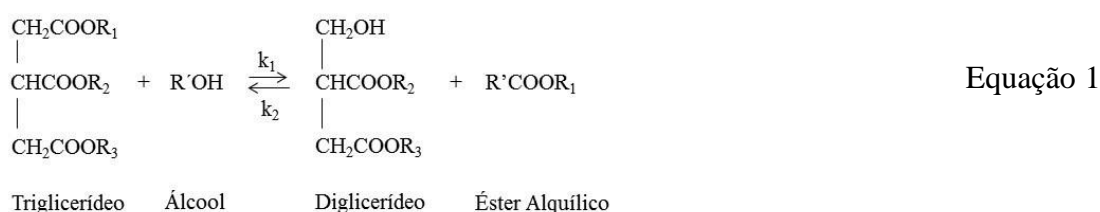
A comercialização e, conseqüentemente, a qualidade dos combustíveis que abastecem os veículos é um assunto que sempre gera interesse de todos os setores da sociedade, como da indústria, consumidores finais, órgãos reguladores e pesquisadores da área. Este tema possui importância para o desenvolvimento de novas tecnologias que produzirão benefícios tanto econômicos quanto sociais¹⁰. Neste sentido, atualmente, há necessidade de pesquisas para diversificar fontes de matéria-prima, incluindo renováveis¹¹ ou óleos residuais¹², diminuir custos de produção, criar novas metodologias de produção¹³, melhorar características físico-químicas¹⁴, qualidade e consequente rendimento e emissões do motor¹⁵, entre outros.

Embora o óleo vegetal possa ser utilizado diretamente como combustível em motores, a sua elevada viscosidade deve ser reduzida antes de ser utilizado em motores a Diesel comuns. Do ponto de vista técnico, a alta viscosidade dos óleos vegetais causa uma atomização do combustível ruim na câmara de combustão do motor e, eventualmente, resulta em problemas operacionais, tais como depósitos de resíduos no motor. A principal solução para o problema tem sido a redução da viscosidade através da reação de transesterificação de óleos e gorduras que produz um combustível mais limpo e ambientalmente seguro. A pirólise e a micro emulsão são métodos que também

poderiam ser utilizados, mas são mais caros e produzem biodiesel de menor qualidade, comparados ao método de transesterificação¹⁶.

1.2.1 A Reação de Transesterificação

A transesterificação de óleos vegetais reflete na qualidade do biodiesel, reduzindo a sua viscosidade e melhorando outras propriedades físico-químicas. O mecanismo geral da reação de transesterificação é apresentado na sequência de equações 1, 2 e 3^{17,18}:



Na indústria de biodiesel, a transesterificação é um método universal e bem estabelecido para a produção de biodiesel. Na presença de um catalisador ácido ou básico, os triglicerídeos de óleos ou gorduras reagem com o álcool utilizado, normalmente metanol, para produzir o biodiesel, o qual é também conhecido como éster metílico de ácidos graxos, além do subproduto glicerol. Como a transesterificação consiste em uma série de reações consecutivas e reversíveis, os triglicerídeos são convertidos na primeira etapa em diglicerídeos, após em monoglicerídeos e, finalmente, em glicerol, de maneira que um mol de éster é formado em cada etapa.

Através da contínua pesquisa e desenvolvimento, muitas empresas oferecem tecnologias inovadoras para produzir biodiesel de alta qualidade de forma consistente a um preço competitivo via processo de transesterificação. No entanto, há várias questões sobre a produção de biodiesel comercial que necessitam ser melhoradas, como a construção de uma indústria de biodiesel mais sustentável e

o desenvolvimento de uma tecnologia robusta para produzir, refinar e recuperar os resíduos finais, além de sempre tentar otimizar a tecnologia de produção de biodiesel, e para isso é realmente importante aprimorar as tecnologias convencionais.

1.2.1.1 Catálise Básica

Um dos processos de produção de biodiesel mais utilizados na indústria é através da catálise básica convencional. Suas principais vantagens são uma reação rápida com baixa razão molar de álcool/triglicerídeo com alto rendimento (mínimo exigido de 96,5%)¹⁹. No entanto, a catálise básica requer condições anídras mais rígidas comparativamente a ácida. A presença de água conduz à hidrólise irreversível dos lipídeos. Pode formar emulsão ou sabão se a concentração de catalisador for maior do que a necessária²⁰. Normalmente as espécies utilizadas neste processo são hidróxido de potássio, hidróxido de sódio, e bases de Lewis. Há a formação de espécies ativas semelhante à um alcóxido que atacam a carbonila dos triglicerídeos, dando origem a um intermediário, a partir do qual o éster de alquila é formado²¹. A regeneração das espécies ativas ocorre após a liberação do mono alquil éster. A formação do glicerol ocorre no final do processo. Normalmente, um pequeno excesso de álcool é necessário para deslocar o equilíbrio para os produtos e para evitar reações paralelas²². O tempo de reação diminui quando empregam temperaturas próximas do ponto de ebulição do álcool utilizado.

1.2.1.2 Catálise Ácida

A transesterificação por catálise ácida, no entanto, inclui a combinação de reações reversíveis. A sua elevada conversão é devido à capacidade de reagir tanto com ácidos graxos quanto com sais de ácidos graxos presentes no sistema. Os ácidos mais empregados são o ácido clorídrico, ácido sulfúrico e o trifluoreto de boro¹⁹. No entanto, sabe-se que o ácido como catalisador atua muito mais lentamente do que a base. A velocidade da reação de produção de biodiesel com catalisador ácido está relacionada com as condições da reação de transesterificação como a proporção molar de álcool/triglicerídeo, a temperatura, concentração do catalisador e pureza dos reagentes. Os álcoois mais utilizados nesse processo são o metanol, etanol, propanol, ente outros. A razão molar de triglicerídeos/álcool é geralmente, cerca de 12 moléculas de álcool para uma de triglicerídeo. O excesso é usado a fim de deslocar o equilíbrio de reação, evitar reações reversíveis e para acelerar o processo²⁰. A temperatura influencia a velocidade do ataque eletrofílico e desta forma afeta os rendimentos e, conseqüentemente, diminui o custo de produção. A concentração de catalisador está diretamente

relacionada com o rendimento: quando é superior a um por cento promove reações reversas, enquanto uma concentração baixa resulta em um processo de evolução limitado. As impurezas podem promover reações paralelas, o que pode ser minimizado com maior temperatura e pressão, mas isso não otimiza o rendimento²³.

1.2.1.3 A metodologia *Transesterification Double Step Process (TDSP)*

O processo de transesterificação em duas etapas ou *Transesterification Double Step Process (TDSP)* utilizado neste estudo foi inicialmente desenvolvido por Samios e colaboradores²⁴ para produzir biodiesel metílico a partir de alguns óleos clássicos no Brasil como girassol e linhaça, por ser um método simples e rápido de síntese, que une as vantagens da catálise básica aos da catálise ácida convencionais facilitando o procedimento de separação dos produtos, e resulta em ótima eficiência de conversão. Ao longo dos anos também vem sendo aprimorado desde a sua criação, resultando na aplicação para produção de biodiesel etílico também²⁵, além de possibilitar o uso de outras fontes de matéria-prima, como banha, sebo e a reutilização de óleo de fritura por exemplo²⁶.

O mecanismo relacionado com este processo foi descrito por Samios e colaboradores como um mecanismo de transesterificação consecutivo ácido-base incorporando as características principais e princípios da catálise básica e ácida. A primeira reação compreende a formação da espécie ativa do catalisador (alcóxido). Uma vez formada, esta espécie reage com o triglicerídeo o que constitui o primeiro passo do processo típico de catálise básica. A segunda etapa ocorre com a adição de ácido sulfúrico além de mais álcool quando a mistura apresenta pH entre 12 e 14, ou seja, ausência total de íons H^+ . A espécie ativa ácida produzida pela adição do ácido seria responsável pela neutralização e, simultaneamente, a produção de uma espécie $R1^+$, que ataca diretamente os grupos carbonílicos. A diferença entre este passo e a catálise ácida clássica, é que o ácido é adicionado numa mistura de reação, que inclui alcóxido e apresenta um elevado valor de pH. Na catálise ácida clássica, tal como descrito anteriormente, a solução de ácido/álcool é adicionada ao triglicerídeo neutro. A espécie muito ativa $R1^+$ reage imediatamente com mono, di e triglicerídeos residuais, bem como, sabão, já presente na mistura de reação, resultando num processo rápido e com grau de conversão muito alto, o que é evidenciado pela pureza do produto final. É importante observar que de acordo com este mecanismo, descrito nos trabalhos de Samios e colaboradores, os ésteres metílicos dos ácidos graxos produzidos pelo processo de catálise básica, não sofrem qualquer degeneração pela catálise ácida²⁴⁻²⁶.

O objetivo desta metodologia é de superar dificuldades observadas na produção de biodiesel de maneira convencional através da catálise básica ou ácida, por um processo mais rápido do que a catálise ácida, de fácil separação de fases e de elevada pureza do biodiesel obtido. As condições

operacionais do método TDSP propostos são de relativamente baixa temperatura e de facilitar a separação de fases. Este método supera a necessidade do emprego de materiais anidros, de óleo vegetal com baixa acidez, elimina qualquer problema de formação de sabão ou emulsão (frequentemente presentes na catálise básica), requer tempos muito mais curtos do que a transesterificação por catálise ácida e resulta em graus elevados de conversão. Por outro lado, o método TDSP em comparação com uma única etapa de catálise básica, necessita do emprego de cerca do dobro da quantidade de metanol e catalisador, o que torna o processo mais caro. Todavia, o método TDSP tem a capacidade de produzir biodiesel com conversões mais elevadas do que o método básico e em tempos operacionais menores do que a catálise ácida, mostrando-se uma boa opção a ser otimizada no intuito de tornar-se uma nova realidade tecnológica e industrial.

1.3 Adulteração de Combustíveis

Outro tema recorrente, tanto para combustíveis petroquímicos quanto biocombustíveis, é o problema da adulteração. A adulteração visa obtenção de maior lucro implicando na diminuição da qualidade do produto, o que gera normalmente problemas ao consumidor como aumento do consumo de combustível pelo motor, além de desgaste e corrosão de peças do veículo.

A Agência Nacional do Petróleo, Gás Natural e Biocombustíveis (ANP), implantada pelo Decreto nº 2.455, de 14 de janeiro de 1998, é o órgão regulador das atividades que integram a indústria do petróleo e gás natural e a dos biocombustíveis no Brasil. Autarquia federal, vinculada ao Ministério de Minas e Energia, a ANP é responsável pela execução da política nacional para o setor energético do petróleo, gás natural e biocombustíveis. A ANP é também um centro de referência em dados sobre a indústria do petróleo e gás natural. A agência mantém o Banco de Dados de Exploração e Produção (BDEP), promove estudos sobre o potencial petrolífero e o desenvolvimento do setor, divulga as estatísticas oficiais sobre reservas e produção no Brasil, realiza pesquisas periódicas sobre qualidade dos combustíveis e lubrificantes e sobre preços desses produtos. Na área dos biocombustíveis, divulga dados sobre autorizações, produção e comercialização de biodiesel e etanol. Algumas das principais atribuições práticas da ANP são: autorizar e fiscalizar as atividades de produção, importação, exportação, transporte, armazenagem, distribuição, revenda e comercialização de derivados de petróleo, gás natural e biocombustíveis, além de estabelecer as especificações técnicas (características físico-químicas) dos derivados de petróleo, gás natural e dos biocombustíveis e realizar o permanente monitoramento da qualidade desses produtos nos postos revendedores dos combustíveis²⁷.

As adulterações mais comuns registradas pela ANP, no ano de 2016, foram: Para o óleo Diesel a adulteração mais recorrente envolveu o teor de biodiesel (que deveria ser, atualmente, de 7% em volume) com 29,9% das não-conformidades. No caso da gasolina C a principal adulteração foi a variação no percentual de etanol adicionado, que por lei, deveria ser de 27% em volume, totalizando 53,1% das não-conformidades verificadas. Além disso, dados provenientes das ações de fiscalização mostraram que outro problema antigo de adulteração ou contaminação muito recorrente, acontece pela adição de óleo Diesel na gasolina, tanto pela adição intencional do Diesel como adulterante²⁸, uma vez que seu valor comercial é menor, quanto pela contaminação que ocorre, muitas vezes, por problemas de vazamento nos tanques de armazenamento nos postos de combustíveis ou até mesmo por erro no abastecimento dos mesmos, pelos caminhões tanque. Para o álcool combustível (hidratado) a principal infração foi relativa à massa específica/teor alcoólico com 58,1% das não-conformidades¹ fator este relacionado ao controle da qualidade e quantidade máxima da água permitida neste combustível. Estes dois últimos tipos de adulteração apresentados foram estudados neste trabalho, através da aplicação de ferramentas quimiométricas a dados de análises por ¹H RMN e FTIR-ATR de amostras adulteradas propositalmente ou não, produzidas de forma a reproduzir este problema, que consiste em duas das adulterações mais antigas e ainda muito recorrentes no etanol hidratado e na gasolina comercializados nos postos de combustíveis em todo país.

1.3.1 Controle de Qualidade dos Combustíveis

O controle de qualidade dos combustíveis automotivos brasileiros é realizado pelo Programa de Monitoramento da Qualidade dos Combustíveis (PMQC) que visa atender ao disposto no artigo 8º da Lei 9.478/1997 e trata da garantia de qualidade e do suprimento de combustíveis no mercado nacional. Os principais objetivos são o levantamento dos indicadores gerais da qualidade dos combustíveis comercializados no Brasil e a identificação de focos de não conformidade, visando orientar e aperfeiçoar a atuação da área de fiscalização da agência. O PMQC conta para sua execução com uma rede de instituições e centros de pesquisa. Nesse Programa o universo de postos revendedores de cada estado é dividido em regiões com um número semelhante de postos. Os postos revendedores são selecionados com base em sorteios aleatórios, realizados pela instituição contratada para efetuar a coleta e análise de amostras de gasolina, etanol e óleo diesel²⁷.

Para verificar e atestar a conformidade da qualidade dos combustíveis, de acordo com as normas, as instituições e centros de pesquisas vinculados à ANP utilizam ensaios físico-químicos bem estabelecidos nas amostras coletadas no PMQC e seguem rigorosamente os métodos da

Associação Brasileira de Normas técnicas (ABNT) ou da *AMERICAN SOCIETY FOR TESTING AND MATERIALS* (ASTM) desenvolvidos para avaliar cada propriedade, com exceção aos ensaios relativos ao aspecto e cor da amostra, no método visual, e o de octanagem, realizado com equipamento portátil de infravermelho. Estes testes, porém, utilizam alguns equipamentos de custo elevado e outros têm um tempo de execução relativamente longo, limitando o número de amostras analisadas em função do tempo, de modo que sempre é interessante o desenvolvimento de novas tecnologias que aprimorem, modernizem e viabilizem processos de controle de diferentes parâmetros de forma mais eficientes quando se trata da análise da qualidade de qualquer produto.

1.4 Técnicas Espectroscópicas

1.4.1 Espectroscopia de Ressonância Magnética Nuclear de Hidrogênio (RMN 1H)

A ressonância magnética nuclear (RMN) juntamente com a espectroscopia na região do infravermelho constituem duas das mais importantes técnicas espectroscópicas. Através da espectroscopia de RMN vários núcleos podem ser estudados tais como o flúor, nitrogênio e oxigênio, porém os dois mais utilizados são indiscutivelmente o hidrogênio seguido do carbono. A ressonância fornece a quantidade de átomos magneticamente distintos ou não equivalentes do isótopo em estudo, além de obter informações sobre a natureza do ambiente químico de cada átomo. Desta forma, em muitos casos, a junção da espectroscopia de infravermelho médio com RMN tem sua grande aplicação na determinação ou confirmação da estrutura química de uma molécula.

Para que ocorra a ressonância os núcleos dos átomos devem possuir a propriedade chamada *spin*. Para tanto, qualquer núcleo que possua massa ou número atômico ímpar, possui consequentemente um momento angular de *spin* e um momento magnético μ . Para cada núcleo com *spin*, o número de estados de *spin* permitidos é determinado pelo seu número quântico de *spin* nuclear I . O número I é uma constante física e existem $2I+1$ estados de *spin* permitidos, que vão de $-I$ a $+I$. No caso do hidrogênio, seu número quântico de *spin* é igual a $1/2$ e consequentemente dois estados de *spin* permitidos: $-1/2$ e $+1/2$.

Em seu estado normal os estados de *spin* de um núcleo têm energia dita degenerada, ou seja, igual e possui dois *spins* opostos: $-1/2$ e $+1/2$ e seus momentos magnéticos nucleares apontam em direções opostas. Porém, quando um campo magnético externo é aplicado, os estados de *spin* se modificam, uma vez que o núcleo possui carga e gera um campo magnético próprio, ou momento magnético μ que se alinha totalmente a favor ou contra esse campo. O estado de *spin* $+1/2$ possui

energia menor uma vez que está alinhado a favor do campo, enquanto o estado de *spin* -1/2 tem energia maior alinhando-se oposto ao campo.

O fenômeno da ressonância magnética nuclear ocorre quando os núcleos, já alinhados ao campo externo aplicado, são induzidos a absorver energia e a mudar sua orientação de *spin*. Essa energia absorvida é quantizada e igual à diferença de energia entre os dois estados de *spin* envolvidos, conforme equação 4 e na prática ela depende da intensidade do campo externo aplicado B_0 , como apresentado na Figura 1.

$$E_{\text{absorvida}} = h\nu = (E_{\text{estado}-1/2} - E_{\text{estado}+1/2}) \quad \text{Equação 4}$$

Através da equação 5 vemos que quanto maior for a força do campo magnético aplicado, maior é a diferença de energia entre os estados de *spin*.

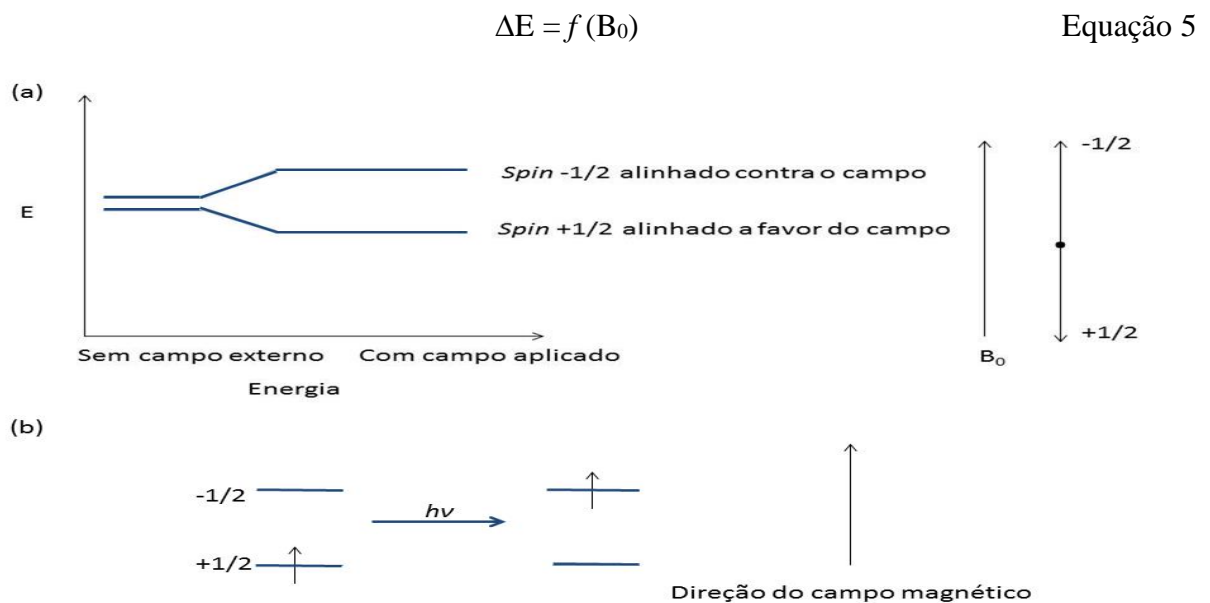


Figura 1. a) Estados de spin para o próton na ausência do campo magnético externo aplicado e com campo magnético externo aplicado, b) Processo de absorção da ressonância pelo próton.

Outro parâmetro importante é a razão magnetogírica γ que é uma constante relativa a cada núcleo e determina a dependência da energia com o campo magnético. A magnitude da separação dos níveis de energia também depende do núcleo em questão, pois cada núcleo possui uma razão diferente entre momento magnético e momento angular, uma vez que cada um possui uma carga e massa específicas. Como o momento angular é quantizado por $h/2\pi$, tem-se:

$$\Delta E = f (\gamma B_0) = h\nu$$

$$\Delta E = \gamma \frac{h}{2\pi} B_0 = h\nu$$

Chegando-se a frequência da energia absorvida:

$$\nu = \frac{(\gamma)}{2\pi} B_0$$

Equação 6

Assim conhecendo o valor da razão magnetogírica, pode-se calcular em que frequência da radiação o próton vai absorver em determinada intensidade de campo.

A natureza da transição nuclear de *spin* ocorre porque os prótons absorvem energia em função do campo magnético aplicado e conseqüentemente ocorre o fenômeno de precessão que é a rotação do núcleo sobre o seu próprio eixo com uma determinada frequência, chamada de frequência de Larmor. Essa frequência é diretamente proporcional à intensidade do campo aplicado, assim quanto maior for o campo, maior será a velocidade ou frequência angular (ω) de precessão. O núcleo possui carga e seu movimento de precessão gera um campo elétrico que oscila na mesma frequência. Quando se aplica a radiofrequência ocorre absorção de energia. Portanto, a ressonância ocorre quando a onda de radiofrequência aplicada ao núcleo for igual a frequência do campo gerado pelo núcleo em precessão, ocorrendo o acoplamento dos dois campos e transferência de energia da radiação para o núcleo o que causa a mudança de *spin*. Deste modo, é dito que há ressonância entre o núcleo e a onda eletromagnética incidente, conforme Figura 2.

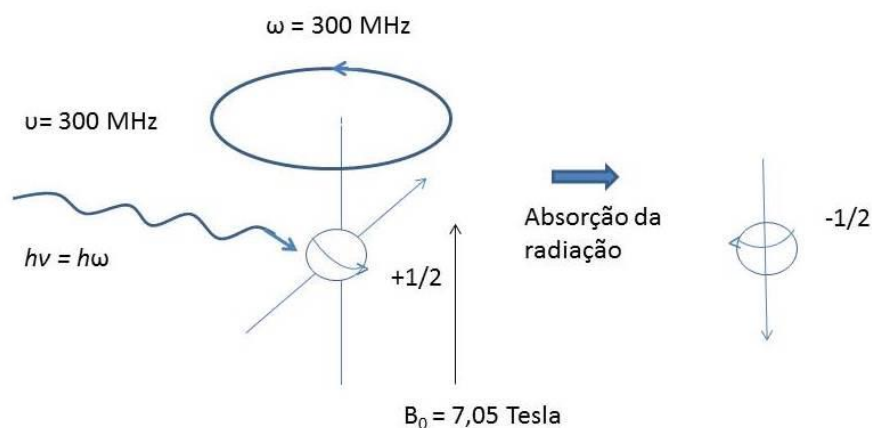


Figura 2. Representação do fenômeno da ressonância magnética nuclear com absorção da radiação quando $\nu = \omega$.

Outro fator importante quando se estuda RMN é o efeito de blindagem e suas consequências sobre o deslocamento químico. A ressonância só possui aplicabilidade porque os prótons de uma molécula absorvem em frequências diferentes. Isso se deve ao fato de que cada próton está em um “ambiente eletrônico” diferente uma vez que cada um possui uma densidade eletrônica diferente ao seu redor. Quando se aplica um campo externo os elétrons de valência se movimentam gerando uma corrente diamagnética local, ou seja, um campo de direção oposta ao campo aplicado. Esse efeito é chamado de blindagem diamagnética ou anisotropia. Quanto maior for a densidade eletrônica, maior é a blindagem exercida. Desta forma, como cada próton tem um ambiente eletrônico particular, conseqüentemente há uma blindagem eletrônica específica, resultando em uma frequência de ressonância distinta. Porém essas diferenças nas frequências não são grandes o suficiente dificultando a sua medição diretamente para cada próton. Para tanto é empregada uma substância de referência ou padrão, que normalmente é o tetrametilsilano (TMS). Assim, os demais deslocamentos químicos de cada próton são medidos em relação aos prótons desse padrão.

O deslocamento de um próton em relação ao padrão depende da intensidade do campo aplicado. Então para não haver diferença no valor do deslocamento para a mesma amostra em diferentes campos ou equipamentos, foi definido um novo parâmetro chamado deslocamento químico (δ) que é obtido dividindo-se o deslocamento medido em hertz de um próton pela frequência em Megahertz do espectrômetro utilizado, conforme equação 7. Assim, a medida torna-se independente do campo e convencionou-se o deslocamento químico em ppm (partes por milhão) em relação à frequência do espectrômetro, tornando o “zero” dessa escala, em ppm, referente ao TMS.

$$\delta = \frac{\text{deslocamento (Hz)}}{\text{frequência espectrômetro(MHz)}} \quad \text{Equação 7}$$

Para ser analisada, a amostra deve ser dissolvida em solvente livre de prótons como, por exemplo, o clorofórmio deuterado (CDCl_3) e o porta amostra é um tubo cilíndrico. A amostra gira sobre seu próprio eixo para que o campo magnético aplicado seja uniforme. No equipamento, há uma bobina com gerador de radiofrequência que altera a orientação do *spin* dos prótons. Quando a amostra absorve energia e há mudança na orientação de *spin* induzindo um sinal de radiofrequência, outra bobina detectora recebe o sinal de ressonância que gera os sinais do espectro. No espectro, o sinal em 0 ppm refere-se ao TMS (padrão) e os prótons mais blindados que precessam numa velocidade menor, relativamente aos prótons mais desblindados, aparecem mais à direita do espectro, ou seja, possuem deslocamentos químicos menores, enquanto os mais desblindados estão mais à esquerda e possuem deslocamentos químicos maiores. Outro aspecto importante é chamado de pulso que é uma descarga de energia dada sobre a amostra, com alta potência, em um tempo curto, que é capaz de excitar simultaneamente todos os núcleos dos átomos que constituem a molécula. Para isso a fonte é ligada e desligada rapidamente, de modo a obedecer o princípio de Heisenberg, pois se a duração do pulso for muito curta, haverá imprecisão na frequência necessária. Então é utilizada uma faixa de frequências em torno daquela desejada. Quando o pulso é interrompido os núcleos perdem energia e voltam ao *spin* original, e emitem radiação, o que se chama de processo de relaxação. Porém como numa molécula há vários núcleos distintos, são emitidas várias frequências distintas ao mesmo tempo e essa emissão é chamada de decaimento da indução livre ou *free induction decay* (FID). O FID é resultado da sobreposição de todas as frequências emitidas e sua intensidade diminui com o passar do tempo conforme os núcleos perdem a excitação. Através de um método matemático chamado transformada de Fourier consegue-se extrair as frequências de cada núcleo. O FID é descrito em relação ao sinal no domínio do tempo, como apresentado na Figura 3, mas com a transformada de Fourier esse sinal é convertido para o domínio da frequência, que é a forma que vimos no espectro final. Esse método agrega maior sensibilidade e menor tempo à análise, permitindo analisar núcleos com baixa abundância ou amostras muito diluídas²⁹⁻³².

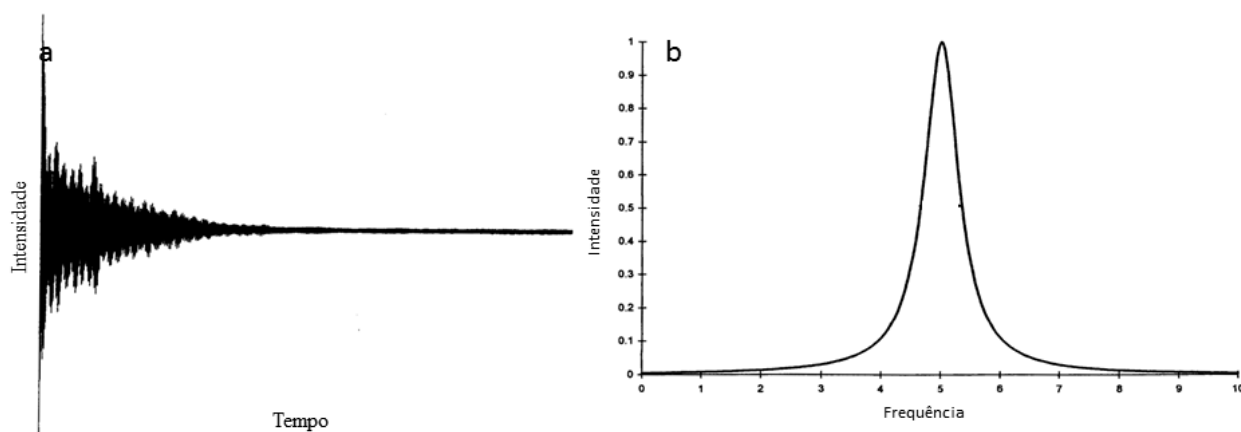


Figura 3. a) Exemplo de uma FID (sinal no domínio do tempo) b) Exemplo de um sinal no do espectro no domínio da frequência. Adaptado de Macomber³³.

1.4.2 Espectroscopia de Infravermelho com Transformada de Fourier (FTIR)

A maioria dos compostos que possuem ligações covalentes, orgânicos ou inorgânicos, tem a capacidade de absorver em várias frequências de radiação eletromagnética na região do infravermelho do espectro eletromagnético. A região vibracional do infravermelho é a que possui mais aplicabilidade na química. Nesta região temos comprimentos de onda maiores do que aqueles da luz visível, de 400 a 800 nm, mas menores do que os das micro-ondas, que são maiores que 1mm. Através do espectro eletromagnético, Figura 4, podemos ver a relação inversamente proporcional do comprimento de onda (λ) em relação à frequência (ν): $\nu = c/\lambda$, onde c = velocidade da luz. A energia, por sua vez, é diretamente proporcional à frequência: $E = h\nu$, onde h = constante de Planck. Por meio desta equação pode-se concluir que a região de raios X possui radiação de alta energia, o suficiente para quebrar ligações químicas. Por outro lado, as energias mais baixas como aquelas em forma de radiofrequência são capazes apenas de causar transições de *spins* nucleares ou eletrônicos, como é o caso daquela utilizada em RMN²⁹⁻³².

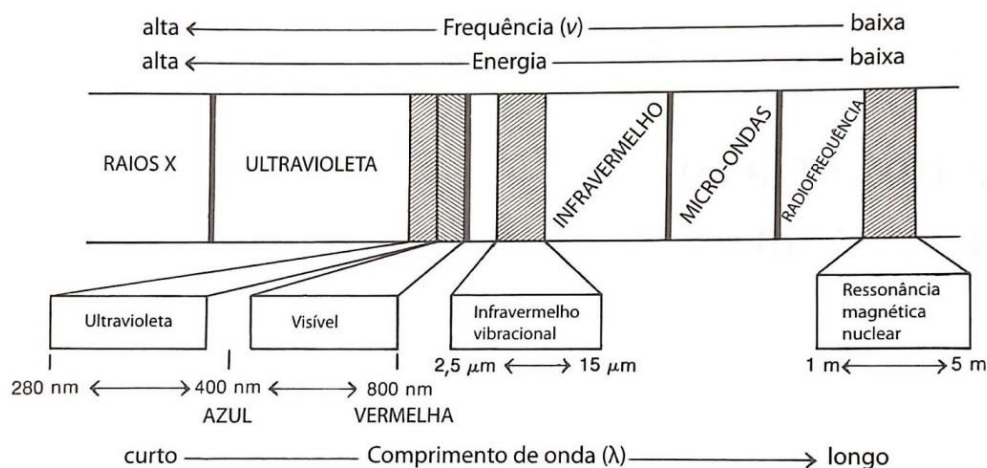


Figura 4. Parte do espectro eletromagnético que relaciona o infravermelho com outros tipos de radiação mais próximos²⁹.

A forma mais usual de nos referirmos ao infravermelho médio é em termos do número de onda, em cm^{-1} , uma vez que esta unidade é diretamente proporcional à energia, ou seja, número de onda maior corresponde à uma energia maior. Desta forma, a escala de número de onda vai de 4000 a 400 cm^{-1} .

Assim como ocorre no RMN, quando há absorção de energia quantizada no infravermelho pelas frequências selecionadas de radiação no infravermelho, ocorre excitação das moléculas que atingem um estado de maior energia. A energia envolvida neste caso é da ordem de 8 a 40 KJ/mol e corresponde a faixa de frequências vibracionais de estiramento e dobramento das ligações das moléculas covalentes. Porém somente ligações que possuem momento de dipolo variável no tempo absorvem no IV. Ligações simétricas como a do H_2 não absorvem radiação no infravermelho. Para ocorrer transferência de energia, a ligação deve ter um dipolo elétrico que varie na mesma frequência da radiação aplicada, assim o dipolo acopla-se com o campo eletromagnético incidente. No infravermelho, o mesmo tipo de ligação em diferentes moléculas não terá frequências de vibração exatamente iguais, assim como na ressonância o ambiente químico em que os hidrogênios estão, influencia no sinal gerado para cada hidrogênio. Apesar de as frequências que as moléculas absorvem serem iguais, os espectros de moléculas diferentes não serão exatamente os mesmos²⁹⁻³².

A utilização tradicional e principal do espectro de IV é para verificar a estrutura das moléculas. Isto se dá através da identificação dos modos vibracionais dos grupos funcionais presentes. As vibrações mais comuns são o estiramento e o dobramento, conforme apresentados na Figura 5. Em geral, vibrações do tipo estiramento assimétrico ocorrem em frequências mais altas do que os simétricos e os estiramentos, por sua vez, ocorrem em frequências mais altas do que os dobramentos.

Ainda há termos muito utilizados como tesoura (scissoring), balanço (wagging), torção (twisting) e sacudida (rocking) para descrever deformações vibracionais. Em grupos funcionais em que há três ou mais átomos, com pelo menos dois deles iguais, originam-se os estiramentos simétricos ou assimétricos, como no CH_3 e NO_2 , por exemplo²⁹⁻³².

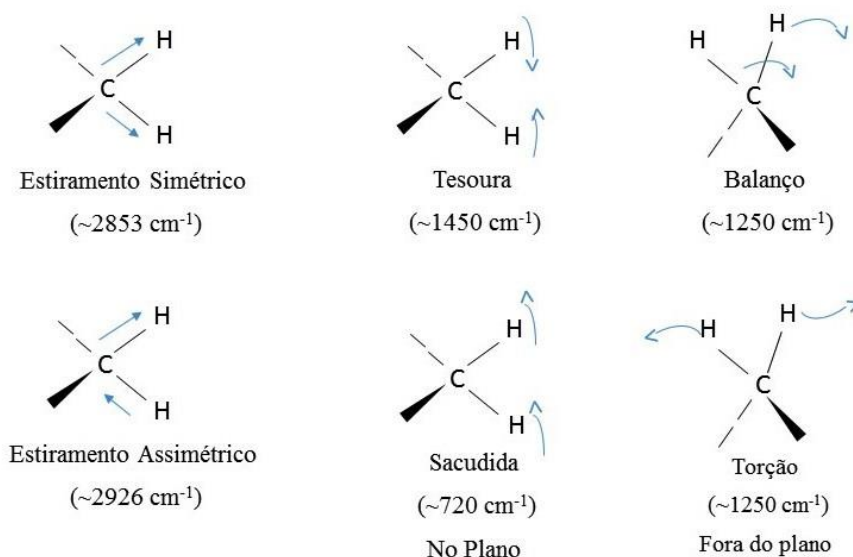


Figura 5. Exemplo de modos de estiramento e dobramento das moléculas.

Estes tipos de vibrações são classificados como fundamentais e tratam da excitação do estado fundamental para o excitado, de maior energia. Outras bandas mais fracas presentes são as harmônicas, de combinação e de diferença. As harmônicas são resultado da excitação de bandas que passam do estado fundamental para outro de maior energia e são múltiplos inteiros da frequência fundamental. Já as bandas de combinação tratam da soma de duas bandas que ocorrem quando duas frequências vibracionais se acoplam numa molécula originando uma vibração com uma nova frequência ativa na região do infravermelho. As bandas de diferença são análogas as de combinação, porém a frequência resultante é a diferença entre duas bandas. Outro aspecto importante é chamado de ressonância de *Fermi* que é uma vibração acoplada a partir de uma vibração fundamental com uma frequência de combinação ou harmônica. Esse fenômeno é bastante frequente nos grupos carbonila. Outro acoplamento possível de ocorrer é quando os estiramentos e dobramentos unem-se às frequências rotacionais, apesar destas últimas não terem a mesma região no espectro. Isto gera uma banda alargada nessas absorções, em função da excitação simultânea de rotações e vibrações²⁹⁻³².

Os espectrômetros utilizados para analisar amostras por infravermelho produzem um interferograma, ou gráfico de intensidade versus tempo, que é um sinal complexo, que contém todas as frequências que formam o espectro no domínio do tempo. Porém, assim como na técnica de RMN,

pode-se tirar conclusões mais úteis se o espectro estiver plotado na forma de intensidade versus frequência, ou seja, espectro no domínio da frequência. A operação matemática complexa chamada de transformada de Fourier aplicada ao infravermelho também serve para separar as frequências do interferograma. O infravermelho com transformada de Fourier ou *Fourier Transform Infrared* (FTIR) é rápido, mais sensível e tem melhor relação sinal/ruído do que os antigos equipamentos dispersivos de infravermelho. A Figura 6 apresenta um esquema representativo de um espectrômetro de FTIR²⁹⁻³².

No equipamento de FTIR o interferômetro controla a energia enviada para a amostra analisada. Essa energia vinda da fonte atravessa o divisor de feixes e um espelho num ângulo de 45 graus em relação a radiação de entrada, a separa em dois feixes perpendiculares: um segue a direção inicial e outro percorre um desvio num ângulo de 90 graus. Este último bate num espelho fixo e é refletido de volta. O outro vai para um espelho móvel e também é refletido. Isto é feito para variar a trajetória do segundo feixe. Quando os dois feixes se reencontram e se recombina, porém, as diferenças nos caminhos causam interferências construtivas e destrutivas. Esse feixe origina o interferograma. Em seguida ele atravessa a amostra, a qual absorve simultaneamente em todas as frequências do espectro. O sinal do interferograma que chega ao detector informa a quantidade de energia absorvida em cada frequência.

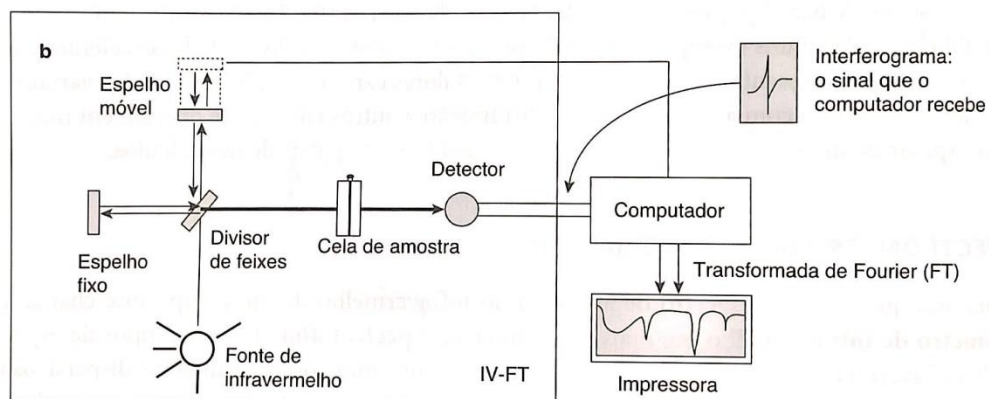


Figura 6. Esquema representativo de um espectrômetro de infravermelho com transformada de Fourier (FTIR)²⁹.

A aplicação da transformada de Fourier ao final desse processo permite extrair as frequências individuais e produzir o espectro. Os programas atuais subtraem automaticamente o espectro de fundo também chamado de branco (*background*) que contém somente informações atmosféricas e do caminho óptico, ou seja, um espectro obtido sem a amostra. O espectro final é o resultado dessa subtração²⁹⁻³².

Já a espectroscopia na região do infravermelho acoplada a um acessório de refletância total atenuada ou *Attenuated Total Reflectance* (ATR) é uma técnica simples que se destaca por não necessitar preparo de amostra, e ser muito utilizada para analisar líquidos tanto quantitativamente quanto qualitativamente. O infravermelho com reflexão total atenuada (FTIR-ATR) emprega cristais com alto índice de refração, como os de seleneto de zinco (ZnSe), muito utilizado, interagindo com a amostra sobre a superfície do cristal, no qual ocorrem múltiplas reflexões da radiação do infravermelho. O ATR mede as variações que ocorrem no feixe de infravermelho que sofre a reflexão quando interage com a amostra. A profundidade de penetração (dp) do feixe é dada pela equação 8³⁴⁻³⁷.

$$d_p = \frac{\lambda}{2\pi n_1 \left(\sin^2 \theta_i - \left(\frac{n_3}{n_2} \right)^2 \right)^{1/2}}$$

Equação 8

Onde: λ = comprimento de onda no vácuo, λ/n_1 = comprimento de onda do elemento de reflexão interna; n_2 = índice de refração do cristal; n_3 = índice de refração da amostra e θ_1 = ângulo de incidência do feixe^{37,38}.

A radiação que penetra na amostra recebe a denominação de onda evanescente e quando a amostra absorve num determinado comprimento de onda evanescente ocorre a atenuação do feixe cujo comprimento de onda corresponde às bandas de absorção apresentadas no espectro de infravermelho³⁴⁻³⁷. A Figura 7 representa o funcionamento do ATR.

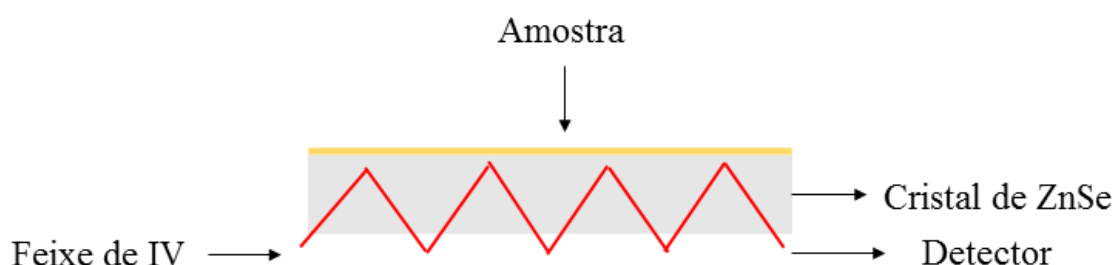


Figura 7. Esquema representativo de um acessório de refletância total atenuada (ATR).

1.5 Ferramentas Quimiométricas

O conceito de quimiometria é bastante amplo, sua aplicação em maior escala é relativamente nova e varia de acordo com a interpretação de cada autor. Para Sergio L. C. Ferreira, autor do livro intitulado “Introdução às Técnicas de Planejamento de Experimentos”³⁹, é a disciplina que emprega fundamentos matemáticos e estatísticos aliados a recursos de informática para planejar experimentos visando otimização e também avaliação de dados de origem química, com o objetivo de obter a máxima quantidade de informações, e se subdivide em: Técnicas de planejamento de experimentos, análise multivariada e calibração multivariada. Já para Edenir R. Pereira Filho, autor do livro “Planejamento Fatorial em Química: Maximizando a obtenção de resultados”, a quimiometria é uma área da química que utiliza conhecimentos de matemática e estatística para a identificação de informações relevantes de um problema em estudo ⁴⁰. E pode ser dividida em quatro vertentes: Análise exploratória de dados químicos com a utilização de HCA (*Hierarchical Cluster Analysis*) e PCA (*Principal Component Analysis*), calibração multivariada com a utilização da PCR (*Principal Component Regression*) e PLS (*Partial Least Squares*), além de modelos de classificação através de ferramentas como KNN (*k-Nearest Neighbour*), SIMCA (*Soft Independent Modelling of Class Analogy*) e PLS-DA (*PLS for Discriminant Analysis*) e planejamento fatorial para otimizar condições experimentais. Segundo Márcia M. C. Ferreira, autora do livro “Quimiometria Conceitos, métodos e Aplicações”, há na literatura tantas definições de quimiometria quanto quimiometristas, mas há um consenso geral que esta ciência é a intersecção de três grandes áreas: Química, Matemática e Estatística. Além disso, a autora destaca que é importante perceber a sua natureza interdisciplinar uma vez que ela possui grande aplicação na química medicinal e computacional como no estudo de novos fármacos, bem como nas indústrias farmacêuticas, química e de alimentos aplicada ao controle de qualidade e no tratamento de imagens, além da área médica de metabolômica auxiliando na identificação de marcadores de doenças ⁴¹. No livro intitulado “Como fazer experimentos” de Bruns e colaboradores, os autores com mais de 20 anos de experiência acadêmica em quimiometria a definem como o emprego de técnicas estatísticas, matemáticas e gráficas para resolver problemas químicos, mas que, na realidade, podem ser aplicadas às mais diversas áreas⁴².

Historicamente um dos primeiros trabalhos de grande importância para esta área foi o de Kowalski e Bender⁴³ além do estudo de Lawton e Sylvestre publicado em 1971 que teve papel importante na fundamentação desta metodologia⁴⁴. Outros trabalhos importantes foram o de Geladi e Esbensen considerados como alguns dos fundadores da área de quimiometria⁴⁵. O primeiro livro relativo a este assunto foi resultado de um simpósio ocorrido em 1976 organizado pela *American*

Chemical Society e já incluía temas como otimização de experimentos, análise fatorial e SIMCA⁴⁶ e somente em 1987 foram publicadas revistas sobre quimiometria como o *Journal of Chemometrics* e *Chemometrics and Intelligent Laboratory Systems*.

1.5.1 Planejamento Experimental ou Matriz Doehlert

O planejamento experimental Doehlert ou matriz Doehlert, por sua vez, foi desenvolvido por David H. Doehlert em 1972⁴⁷, e possui aplicação inclusive na otimização de métodos analíticos. Este tipo de planejamento descreve uma distribuição de pontos de maneira uniforme sobre um espaço esférico, que contempla um número menor de experimentos a serem realizados comparativamente aos planejamentos de superfície de resposta. As suas principais vantagens são: i) permite fazer planejamento com número de níveis diferentes para cada fator em estudo, ii) não contém experimentos com todos os fatores em nível superior ou inferior, iii) possui elevada eficiência com relação ao número de experimentos necessários para a otimização, determinado pela expressão $n = k^2 + k + 1$, onde n é o número de experimentos e k é o número de fatores envolvidos.

Para exemplificar, uma matriz Doehlert para estudo de duas variáveis possui um ponto central e outros seis pontos que formam a figura de um hexágono. Nesse caso, o fator julgado como principal possui cinco níveis diferentes enquanto o outro fator possui três níveis. Geralmente se utiliza o fator que tem maior efeito no processo com o maior número de níveis com o objetivo de se obter mais informações acerca do sistema estudado. A Figura 8 representa a matriz de Doehlert para um sistema com duas variáveis e a Tabela I apresenta os valores codificados para cada experimento.

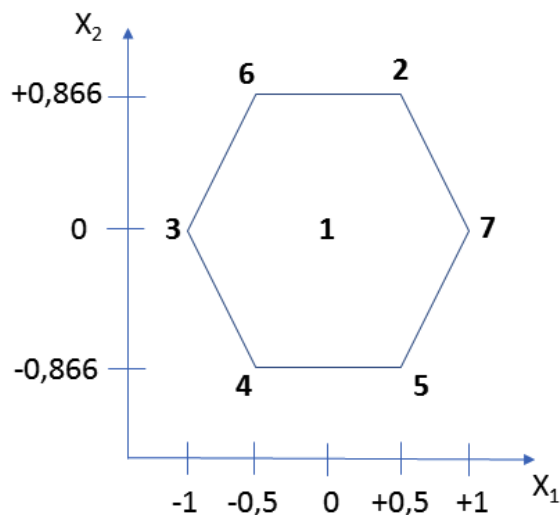


Figura 8. Figura geométrica representativa da matriz Doehlert para sistema com duas variáveis ou fatores.

Tabela I. Matriz de Doehlert com duas variáveis

Experimento	X₁	X₂
1	0	0
2	+0,5	+0,866
3	-1	0
4	-0,5	-0,866
5	+0,5	-0,866
6	-0,5	+0,866
7	1	0

Em seguida os dados experimentais são processados geralmente num programa estatístico gerando um modelo com uma equação representativa e uma superfície de resposta. A validação do modelo é avaliada, por exemplo, através de uma regressão linear ou quadrática, utilizando um determinado nível de confiança, comumente utilizando a análise de variância (ANOVA). A etapa de validação dos modelos é importante para avaliação do estudo de otimização, uma vez que ela aponta os valores de máximo ou de mínimo, no caso, por exemplo, de modelos quadráticos, e estes pontos definem as condições experimentais empregadas no estudo. A validação dos modelos pode ser feita empregando várias ferramentas estatísticas, geralmente estabelecidas utilizando os dados previstos e os dados experimentais. Essas ferramentas de análise de planejamentos experimentais possuem recursos para avaliar o ajuste entre esses dados. A obtenção de um bom modelo necessita ter uma

regressão significativa e uma falta de ajuste não significativa. Desta forma, a maior parte da variação dos dados será descrita pela equação de regressão e o restante é descrito pelos resíduos. Do total da variação descrita pelos resíduos, a maior parte deve estar com o chamado erro puro, que é referente ao erro experimental, e não com a falta de ajuste, que está diretamente relacionada com o modelo proposto. Deve-se observar ainda o valor do coeficiente de determinação, ou R^2 do sistema, que representa o quanto da variação é explicada pela falta de ajuste do modelo, ou seja, quanto mais próximo de um for o valor de R^2 , melhor será o ajuste do modelo. É necessário ainda observar o gráfico dos resíduos que devem estar aleatoriamente distribuídos, apresentando homocedasticidade (distribuição homogênea dos resíduos, sem apresentar tendência nos dados, ou seja possui variância constante dos erros para observações diferentes).

A metodologia de superfície de resposta envolve o planejamento experimental e regressão empregados para analisar o problema em que os fatores influenciam na resposta, com o objetivo de otimizar o processo. A superfície de resposta se caracteriza por um modelo polinomial, e pela sua derivatização se pode obter o valor das variáveis independentes. Para um modelo quadrático a superfície de resposta de um processo no qual as variáveis independentes influenciam na resposta experimental, pode ser representada pela equação 9.

$$R = a + bA + cB + dA^2 + eB^2 + fAB + E \quad \text{Equação 9}$$

Onde: a é o termo independente, b e c são os termos lineares, d e e são os coeficientes quadráticos, f é o coeficiente de interação entre as duas variáveis, e E é o erro aleatório associado ao modelo.

A superfície que representa graficamente esta função é a superfície de resposta, podendo ser representada na forma de um gráfico. A superfície é caracterizada pelo ponto de máximo, ponto de mínimo e ponto de sela. O ponto de sela é a região da superfície de resposta que contém a resposta máxima para os níveis de algumas variáveis e ao mesmo tempo a resposta mínima para os níveis da outra variável do sistema. Os sinais do modelo quadrático indicam o ponto estacionário. Se os valores dos termos são todos negativos o ponto estacionário é de máximo, se forem todos positivos, é um ponto de mínimo e se alguns sinais são positivos e outros negativos é um ponto de sela^{39,40}.

Neste estudo é proposta a inserção do planejamento experimental Doehlert, como um instrumento para otimizar as condições de síntese do biodiesel produzido pelo método TDSP. Modelos multivariados, como é o caso do planejamento Doehlert, são mais completos que os univariados, pois permitem o estudo de várias variáveis simultaneamente. Este modelo reduz o número de experimentos a serem realizados e seu uso em conjunto com a metodologia de superfícies

de resposta é muito utilizado para otimizar condições reacionais^{48,49}. A literatura apresenta, entre inúmeras aplicações, o uso de planejamentos experimentais como outra forma eficiente de avaliar e aprimorar métodos de síntese de biodiesel. Em 2009, Valle et. al. utilizaram planejamento Doehlert aliado ao fatorial fracionado na produção de biodiesel por etanolise de *Raphanus sativus L. var. óleo Oleiferus Stokes* catalisada por etóxido de sódio⁵⁰. Em outro trabalho, Borges e colaboradores utilizaram o planejamento experimental fatorial fracionado para estabelecer as condições experimentais que melhoraram a eficiência de transesterificação de óleo de soja por rotas etílica e metílica⁵¹. Já Hamze e colaboradores estudaram a transesterificação do óleo de fritura na qual a metodologia de superfície de resposta com base no planejamento Box-Behnken foi utilizada para investigar os efeitos dos principais parâmetros de operação, obtendo bom rendimento de reação⁵².

1.5.2 Análise por Componentes Principais ou Principal Component Analysis (PCA)

A análise exploratória é composta de ferramentas específicas para extrair conclusões sobre grandes conjuntos de dados. Juntamente com a *Hierarchical Cluster Analysis* (HCA) ou Análise de agrupamentos por métodos hierárquicos, estas ferramentas compõem dois dos métodos mais importantes, classificados como métodos de reconhecimento de padrões, através dos quais se pode reconhecer tendências de agrupamento em conjuntos de amostras.

Estes métodos quimiométricos são utilizados na identificação de padrões entre grupos de amostras com o objetivo de classificá-las e dividem-se em dois tipos: métodos supervisionados e não-supervisionados, baseando-se em 3 condições:

- i) dados do mesmo tipo são classificados como semelhantes;
- ii) a análise utilizada é capaz de detectar as suas semelhanças ou diferenças;
- iii) tem que existir diferenças significativas entre os dados.

Nos métodos supervisionados, é convencionalizado que os dados já possuem uma classe preestabelecida e que essa informação é utilizada para a análise e construção dos modelos de classificação. Já os métodos não supervisionados não exigem um conhecimento prévio do conjunto de dados, que serão agrupados somente de acordo com a informação dos próprios dados experimentais.

A PCA consiste de um método não supervisionado e por isso este tipo de análise é chamada de análise exploratória e sua ideia básica é projetar os dados num espaço de dimensão menor em relação ao espaço do conjunto de dados inicial sem prejudicar a interação entre os mesmos. Esta técnica

permite evidenciar as informações mais relevantes relacionadas aos dados, facilitando tirar conclusões sobre eles. Visualmente, inclusive, torna-se mais fácil verificar as semelhanças e diferenças, interpretar os dados e ainda constatar se há dados com comportamento atípico que não se enquadram em nenhum grupo.

Um método de projeção consiste em reduzir o número de variáveis envolvidas, assim como ocorre com a sombra de algum objeto que reduz a sua forma do espaço tridimensional para um plano com apenas duas dimensões. Essa redução pode ser denominada como compressão dos dados, sem perda de informação relevante e com a vantagem de facilitar a interpretação dos mesmos. A PCA faz essa compressão, projetando os dados num espaço de menor dimensão com base na correlação entre as variáveis presentes nas amostras. Na química, os dados de espectroscopia e cromatografia, por exemplo, são compostos de muitas variáveis, que conseqüentemente estão correlacionadas e possuem, portanto, aplicabilidade para a PCA. Quando há correlação entre as variáveis dos conjuntos de dados, pode-se reduzir o número de variáveis reais que descrevem quase 100% de toda informação relevante. Essa compressão dos dados é realizada através de combinações lineares das variáveis iniciais, unindo aquelas que possuem informação análoga. Isto resulta num novo conjunto de variáveis com as informações que se deseja^{41,53,54}.

Na PCA essas novas variáveis passam a se chamar componentes principais ou *principal component* (PC). Estas PCs são ditas não correlacionadas e ortogonais uma em relação a outra, ou seja, a informação contida numa PC é exclusivamente sua e, portanto, não está presente nas outras PCs. Outra propriedade das PCs é que a quantidade de informação descrita por cada PC diminui da PC1 em direção a PC2, PC3 e assim por diante, ou seja, a PC1 sempre contém a máxima variância em relação aos dados iniciais. A segunda componente principal (PC2) tem a segunda maior variância dos dados, porém num subespaço que é ortogonal à PC1. As demais componentes, da mesma forma, são ortogonais às duas primeiras e descrevem a máxima variância restante. Portanto, a informação redundante é removida, permitindo descrever toda a informação realmente relevante através de poucas novas variáveis ou PCs^{41,53,54}.

Considerando uma matriz **X**, cada vetor-linha representa uma amostra e cada vetor-coluna representa uma variável. Já a PCA pode ser descrita pela decomposição da matriz em outras duas matrizes: matriz **T** de escores e matriz **P** de pesos, conforme mostrado na Figura 9.

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}$$

Equação 10

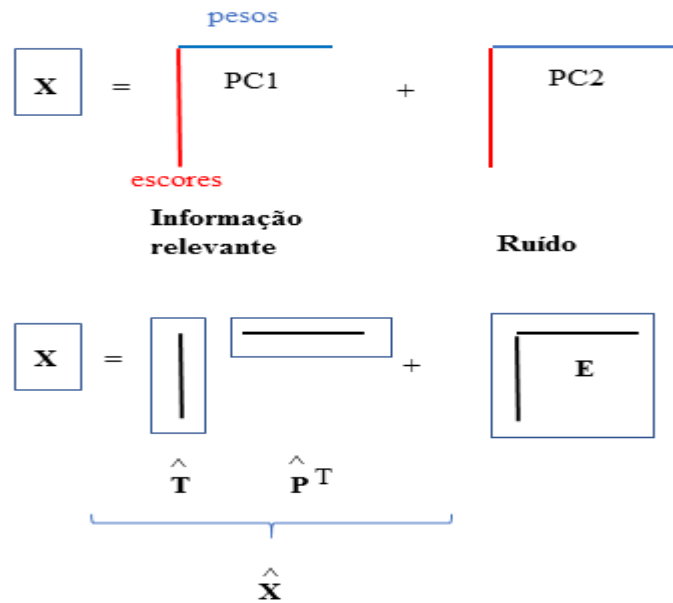


Figura 9. Esquema das transformações realizadas pela PCA numa matriz de dados \mathbf{X} .

Onde \mathbf{E} é a matriz resíduo e os escores estão relacionados às amostras e os pesos às variáveis. Esta alteração modifica o espaço das variáveis originais para outro já que as variações sistemáticas dos dados estão agora contidas nas novas PCs. Com as novas PCs definidas pela projeção dos dados, se houver correlação entre as variáveis originais, poucas PCs são suficientes para representar os dados e as demais contém, assim, a informação aleatória. Assim, a matriz inicial pode ser reescrita pela equação 11:

$$\mathbf{X} = \hat{\mathbf{X}} + \mathbf{E} \quad \text{Equação 11}$$

Onde $\hat{\mathbf{X}}$ é igual a \mathbf{TP}^T e representa uma aproximação dos dados originais e \mathbf{E} é a matriz de erros que contém variações aleatórias, como erros experimentais ou ruído^{41,53,54}.

1.5.3 Análise de Agrupamentos Hierárquicos ou Hierarchical Cluster Analysis (HCA)

A análise de agrupamentos hierárquicos ou HCA consiste basicamente em agrupar dados de acordo com a sua similaridade ou diferença. Junto com PCA, estes métodos de reconhecimento de padrões são importantes para reduzir a dimensão dos dados e detectar amostras que apresentam comportamento anômalo dentro de um conjunto. Estes processos são muito úteis, principalmente, quando temos grande quantidade de informações. Na HCA, a formação de grupos ocorre de acordo com a distância entre as amostras ou objetos. Algumas das formas mais comuns de cálculo da

distância é a distância Euclidiana entre dois pontos e distância de Mahalanobis e existem alguns métodos mais usados para agrupar objetos como K-Nearest Neighbour (KNN) e o método de Ward entre outros⁵⁵.

1.5.4 Regressão pelo método de Mínimos Quadrados Parciais ou Partial Least Squares (PLS)

O modelo de regressão por Mínimos Quadrados Parciais (PLS) consiste em transformar o espaço que contém os valores originais medidos em outro de dimensão reduzida, decompondo a matriz \mathbf{X} com o objetivo de encontrar uma solução, cujo objetivo é a propriedade de interesse. Um fator importante para o modelo PLS é as variáveis latentes que mantêm uma relação entre a explicação da variância em \mathbf{X} e a predição da variável dependente \mathbf{y} . Desta forma, a informação sobre a propriedade de interesse está incluída nos cálculos das variáveis latentes que são eficientes em transportar as informações relevantes para a construção do modelo de calibração. O modelo PLS é construído numa única etapa e tanto a informação da matriz \mathbf{X} quanto da propriedade de interesse são utilizados na decomposição e compressão de dados. Além disso, no método PLS cada variável latente é obtida, a fim de maximizar a covariância entre os escores da matriz \mathbf{X} e variável dependente \mathbf{y} e esta definição de fatores otimiza a decomposição no sentido da predição da propriedade^{41,56}. O modelo PLS pode ser representado pelas equações 12 a 16, que representam a decomposição das matrizes em escores e pesos.

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^t + \mathbf{t}_2 \mathbf{p}_2^t + \dots + \mathbf{t}_h \mathbf{p}_h^t + \mathbf{E} \quad \text{Equação 12}$$

$$\mathbf{Y} = \mathbf{u}_1 \mathbf{q}_1^t + \mathbf{u}_2 \mathbf{q}_2^t + \dots + \mathbf{u}_h \mathbf{q}_h^t + \mathbf{F} \quad \text{Equação 13}$$

Na modelagem PLS, \mathbf{X} e \mathbf{y} são representadas por escores e pelos pesos:

$$\mathbf{X} = \mathbf{T} \cdot \mathbf{P}^t + \mathbf{E} \quad \text{Equação 14}$$

$$\mathbf{Y} = \mathbf{U} \cdot \mathbf{Q}^t + \mathbf{F} \quad \text{Equação 15}$$

A relação entre as duas matrizes de dados \mathbf{X} e \mathbf{y} pode ser obtida, correlacionando-se os escores de cada bloco, afim de obter uma relação linear descrita pela equação 16:

$$\mathbf{U} = \mathbf{b} \mathbf{T} \quad \text{Equação 16}$$

Desta forma o modelo PLS utiliza toda a informação proveniente das matrizes de dados para a construção do modelo e se calcula o número e variáveis latentes, por meio de combinações lineares, a fim de se obter a melhor correlação entre os dados. Escolhe-se o número de variáveis latentes de forma a resultar no menor erro possível, ou seja, a diferença entre os valores de referência e previstos pelo modelo, da propriedade em estudo, deve ser o menor possível. O erro quadrático médio (RMSE) é calculado com base na equação 17:

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_r - y_p)^2}{n}} \quad \text{Equação 17}$$

onde y_r e y_p correspondem, respectivamente, aos valores de referência e previsão, e n é o número de amostras de cada conjunto de dados.

Outro elemento importante para avaliar o modelo é o coeficiente de determinação (R^2), que é calculado por:

$$R^2 = 1 - \frac{\sum_i (y_r - y_p)^2}{\sum_i (y_r - \bar{y})^2} \quad \text{Equação 18}$$

onde y_r e y_p correspondem, respectivamente, aos valores de referência e previstos pelo modelo e \bar{y} é o valor médio.

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

Os modelos de classificação com base no PLS-DA podem ser descritos como a relação entre uma matriz multivariada \mathbf{X} independente e um vetor de respostas que admite os valores 0 ou 1 e indica a que classe uma amostra pertence. A ideia da abordagem PLS-DA é usar uma regressão que opera uma decomposição linear nas matrizes \mathbf{X} e \mathbf{Y} para calcular os parâmetros do modelo, uma vez que existe uma relação entre os dois espaços. Na PLS-DA as amostras modeladas são discriminadas em duas classes diferentes, por exemplo: as amostras são denominadas da classe 1 se amostras pertencem à classe de interesse e são denominados classe 0 se as amostras pertencem a outra classe. Desta forma, se uma amostra pertence à classe de interesse o valor de y correspondente será aproximadamente igual a 1, mas se esta amostra pertence a outra classe o valor de y será aproximadamente igual a zero^{57,58}.

A classificação PLS-DA pode ser descrita por dois parâmetros: sensibilidade e especificidade⁵⁹. Eles são calculados pelas equações 19 e 20:

$$\text{Sensibilidade} = \frac{VP}{VP+FN} \times 100 \quad \text{Equação 19}$$

$$\text{Especificidade} = \frac{VN}{VN+FP} \times 100 \quad \text{Equação 20}$$

Onde VP é o número de verdadeiros positivos, FN representa os falsos negativos, VN representa os verdadeiros negativos e FP representa os falsos positivos.

Desta forma, quando o modelo prevê corretamente todas as amostras, tanto a sensibilidade quanto a especificidade serão iguais a 1.

1.5.6 Máquinas de Vetor de Suporte ou Support Vector Machines (SVM)

Quando se utiliza métodos de calibração multivariada o PLS é o método mais aplicado pela sua maior simplicidade e rapidez computacional. Porém nem sempre o PLS produz os resultados mais adequados, principalmente quando se trata de matrizes complexas, com dados não lineares, onde se torna muito mais apropriado utilizar modelos derivados das máquinas de vetor de suporte (SVM).

As máquinas de vetor de suporte ou Support Vector Machines (SVM) são um conjunto de métodos de aprendizado supervisionado que analisam dados linearmente separáveis e reconhecem padrões, usados tanto para classificação (SVM-DA) quanto para regressão por vetores de suporte (SVR). O método SVM-DA utiliza como entrada um conjunto de dados e prediz, para cada entrada dada, a qual classe faz parte. Dado um conjunto de elementos de treinamento, cada um marcado como pertencente a uma dentre duas classes, o algoritmo de treinamento do SVM constrói um modelo que atribui novos elementos de teste a uma categoria ou outra. Um modelo SVM faz a representação desses elementos como pontos no espaço, que são mapeados de tal maneira, que os elementos de cada classe sejam divididos por um espaço bem definido, que seja tão grande quanto possível. Os novos elementos teste são então mapeados neste mesmo espaço e preditos como pertencentes a uma ou outra classe. Desta forma, o que o modelo SVM faz é calcular uma linha de separação, chamado de hiperplano, representado pela equação 21.

$$f(x) = w \cdot x + b \quad \text{Equação 21}$$

Onde w é o vetor de pesos, x é o vetor de amostras, b é uma constante e a função $f(x)$ é o hiperplano ótimo de separação.

Esse hiperplano procura maximizar a distância entre os pontos mais próximos em relação a cada uma das classes, conforme exemplificado na Figura 10:

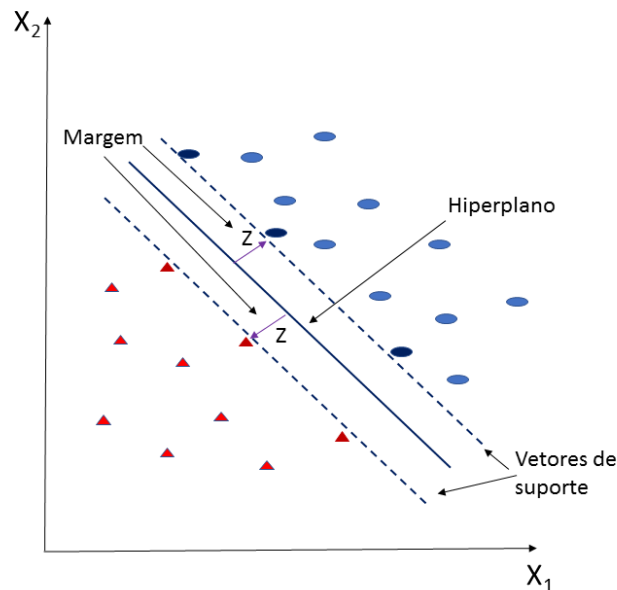


Figura 10. Representação de um conjunto de dados classificados por SVM com hiperplano de separação ótima com margens e vetores de suporte.

Essa distância (z) entre o hiperplano e o ponto mais próximo de cada classe é nomeada de margem. O modelo SVM define primeiramente as classes, por meio de cada ponto pertencente a cada uma das classes, e em seguida ajusta o modelo de maneira a maximizar a margem entre o hiperplano de separação e os vetores de suporte. Desta forma, as amostras nas quais $w \cdot x + b = 1$ pertencem à classe +1, enquanto as amostras que tiverem $w \cdot x + b = -1$ correspondem à classe -1. Os vetores de suporte são, portanto, as amostras de fronteira entre as classes, com as quais o modelo SVM é construído.

A representação anteriormente apresentada é válida quando as amostras são linearmente separáveis, o que na prática pode não ocorrer, necessitando assim, de um ajuste no modelo. Esse ajuste se dá com a introdução de mais uma variável ξ no modelo, chamada variável de folga, que admite um erro proporcional à distância da amostra ao hiperplano definido pelos vetores de suporte da classe à qual pertence, e é desta forma, associado à classificação de algumas amostras. Consequentemente a margem para este novo caso, pode se tornar muito grande, fazendo com que os erros de classificação sejam ponderados por uma nova constante C . Quando a constante possui um valor baixo várias amostras podem ser classificadas erroneamente, enquanto para um alto valor de C ,

classificam-se poucas amostras erroneamente, gerando mais vetores de suporte. Portanto, o valor de C é um parâmetro a ser ajustado no modelo, juntamente com a variável ξ .

Existem ainda muitos problemas que tratam de dados não linearmente separáveis que necessitam de uma transformação na dimensão dos dados de entrada. Nesses casos, aplica-se o SVM por ser um método não linear, mas com o mesmo objetivo de encontrar um hiperplano de separação máxima entre classes. Nos problemas de separação não linear, não é possível aplicar uma função linear que separe as amostras. Assim é necessário mapear as amostras no espaço original para outro de alta dimensão chamado de espaço de características através de uma função *Kernel* não linear. A aplicação dessa função faz com que as amostras tenham uma nova dimensão dada em função das variáveis originais. Esta nova separação linear no espaço de características é equivalente a aplicar uma função não linear no espaço original dos dados. Existem várias funções *Kernel* para vetores de suporte, dentre as quais, as mais utilizadas são a polinomial, sigmoide e a função de base radial ou (RBF *Radial Basis Function*) que possui base exponencial.

Por fim o SVM pode ainda ser aplicado em problemas quantitativos, em forma de regressão, adicionando-se e subtraindo-se uma constante positiva d ao valor da propriedade quantitativa de interesse y_i de cada amostra, originando assim as classes. Desta maneira, o hiperplano de separação máxima passa justamente por y , como é apresentado na Figura 11.

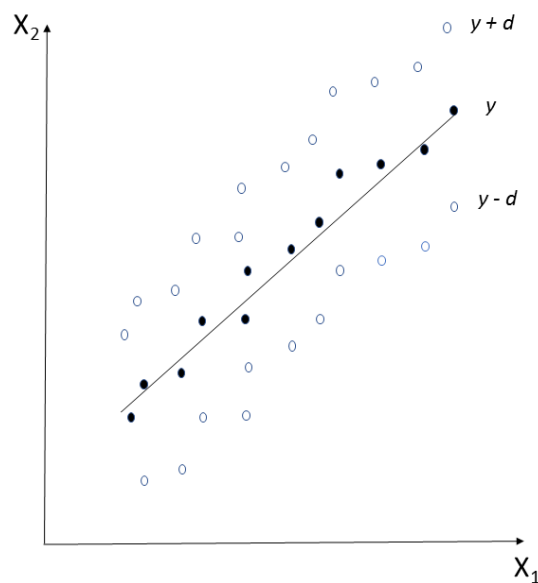


Figura 11. Representação do hiperplano que define a função de regressão por SVR.

Para os cálculos na regressão por SVR, é utilizada a mesma base da função *Kernel* utilizada para a classificação com a inclusão de um limite de tolerância ε em torno da função de regressão que delimita o erro de exatidão. A função é resolvida por multiplicadores de Lagrange e derivadas parciais que resultam na equação 22.

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha^*) K(x_i, x) + b \quad \text{Equação 22}$$

Onde $K(x_i, x)$ é a função *Kernel* aplicada aos dados de entrada e α e α^* representam vetores que ponderam as amostras.

Na regressão também existem parâmetros da função *Kernel* a serem otimizados como C , γ e a margem de ε . Os algoritmos mais utilizados para regressão atualmente são o ν -SVR e o ϵ -SVR⁶⁰⁻⁶².

1.5.7 Cartas de Controle Multivariadas ou Multivariate Control Charts (CCM)

Quando se trata de um grande número de variáveis, as cartas de controle multivariadas baseadas no sinal analítico líquido ou *Net Analyte Signal* (NAS) têm muita aplicabilidade, principalmente no monitoramento do controle de qualidade de processos. Estas cartas tem o objetivo de detectar anormalidades nos parâmetros de controle nos processos de produção, diminuindo a quantidade de produtos fora das especificações e conseqüentemente, custos. Além disso, permitem identificar qualitativamente as amostras dentro ou fora de controle através da divisão, de um dado espectroscópico, por exemplo, em três constituintes: o NAS, os resíduos e os interferentes. A principal relevância desse método é permitir separar a variação sistemática da propriedade que se quer verificar nos dados daquela devido ao restante dos compostos presentes (matriz)^{37,63-65}.

1.5.7.1 Cartas de Controle com base no Sinal Analítico Líquido (NAS)

O vetor NAS é definido como parte do sinal analítico que é ortogonal às contribuições dos outros constituintes presentes na amostra. Além disso, ele é diretamente proporcional à concentração do analito^{37,65-67}.

Com base no NAS, o sinal instrumental r é dividido em três constituintes independentes, como apresentado na Figura 12: vetor NAS (r_k^*) que representa o analito, o vetor interferente (r_{INT}) que representa outro constituinte da amostra (normalmente a matriz) e um vetor de resíduo (r_{RES}) que representa as variações espectrais não explicadas pelo vetor NAS e o vetor interferente, que geralmente contém somente o ruído instrumental. O vetor de regressão NAS (b_k) representa a direção de projeção do vetor NAS (r_k^*)^{37,64,65}.

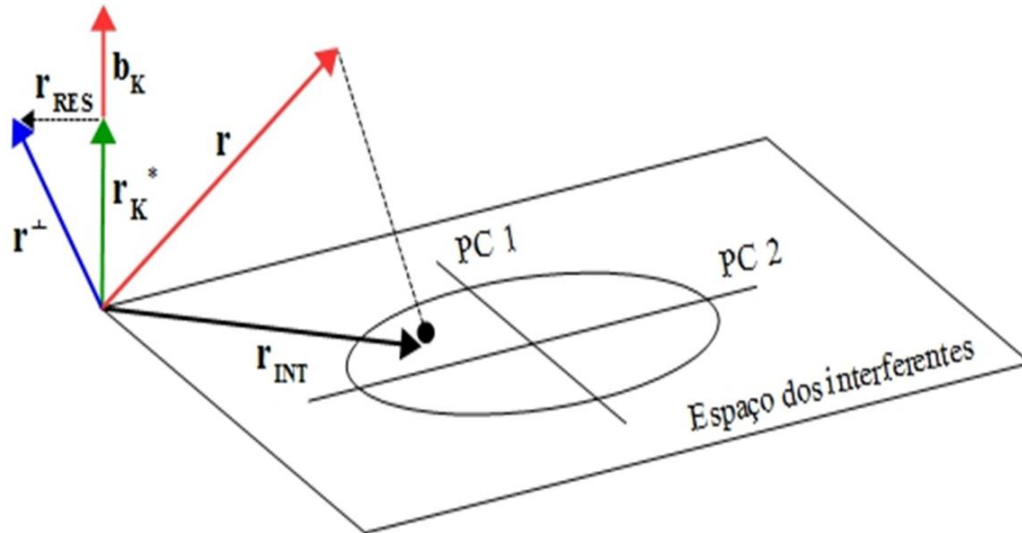


Figura 12. Representação da divisão do espectro r nos vetores: r^\perp (NAS+resíduo), r_{k^*} (vetor NAS), r_{INT} (vetor interferente) e r_{RES} (vetor resíduo).

Conforme a equação 23, qualquer espectro r ($J \times 1$) pode ser dividido em três partes independentes, como apresentado na Figura 12: o vetor NAS (r_{k^*}); um vetor de interferência (r_{INT}); e um vetor resíduo (r_{RES}). Em seguida, três cartas de controle podem ser construídas: carta NAS, carta dos interferentes e carta dos resíduos^{37,64,65}.

$$r = r_{k^*} + r_{INT} + r_{RES} \quad \text{Equação 23}$$

1.5.7.2 Construção das Cartas de Controle

Para a construção das cartas de controle há duas etapas: na primeira define-se o espaço dos interferentes (PP^+), que descreve todas as variações devido a outros componentes da amostra. Isto é definido através de uma PCA. Na segunda é calculado o vetor de regressão NAS (b_k) (para o analito), do vetor interferente e do vetor resíduo. Para tanto, dois conjuntos de amostras são necessários: um com todos os constituintes, exceto o analito, para construção do espaço dos interferentes e outro conjunto incluindo o analito^{37,64,65}.

O espaço dos interferentes é definido através de uma PCA, conforme equação 11. Os pesos definem o espaço dos interferentes e o número de PCs, além disso ilustra o número de fenômenos independentes nos espectros das amostras da matriz X , por exemplo, variação na composição dos componentes majoritários.

O número de componentes principais selecionadas deve conter informação suficiente em relação à amostra. A PCA contém ruído espectral no espaço das PCs, ou seja, não é possível realizar uma separação completa da variação sistemática e de uma não sistemática.

Tendo o espaço do interferente definido, são calculados o vetor interferente, o vetor NAS e o vetor resíduo com amostras consideradas dentro do controle e com todos os constituintes, incluindo o analito.

Já o vetor interferente (\mathbf{r}_{INT}) é determinado pela projeção do espectro \mathbf{r} no espaço dos interferentes, equação 24:

$$\mathbf{r}_{INT} = \mathbf{P}\mathbf{P}^+ \mathbf{r} \quad \text{Equação 24}$$

onde \mathbf{r} é o vetor de dimensão $J \times 1$ com todas as informações espectrais, inclusive do analito e \mathbf{P}^+ é a pseudoinversa da matriz \mathbf{P} .

Por fim, o vetor de regressão NAS é definido como a parte ortogonal ao espaço dos interferentes e é definido pela equação 25:

$$\mathbf{B}_k = (\mathbf{I}_J - \mathbf{P}\mathbf{P}^+) \mathbf{R} \quad \text{Equação 25}$$

onde \mathbf{B}_k é a matriz que contém os vetores, os quais são todos ortogonais ao espaço dos interferentes de dimensão $I \times J$, em que I é o número de amostras dentro do controle utilizadas para calcular o vetor de regressão NAS, \mathbf{I}_J é a matriz identidade de dimensão $J \times J$, \mathbf{R} é a matriz de dimensão $I \times J$ com somente amostras dentro do controle.

No entanto, \mathbf{B}_k contém várias direções ortogonais ao espaço do interferente, assim a média para todos os vetores ortogonais é calculada conforme a equação 26, obtendo-se desta forma uma única direção:

$$\mathbf{b}_k = \frac{\sum_{i=1}^I \mathbf{B}_{k,i}}{I} \quad \text{Equação 26}$$

onde \mathbf{b}_k é o vetor de regressão NAS e I é número de amostras dentro do controle usadas para construir o modelo.

Ainda, segundo a equação 27, o vetor NAS de qualquer espectro \mathbf{r} pode ser calculado através da projeção do espectro para o vetor de regressão NAS (\mathbf{b}_k).

$$\mathbf{r}_k^* = \mathbf{b}_k (\mathbf{b}_k^T \mathbf{b}_k)^{-1} \mathbf{b}_k^T \mathbf{r} \quad \text{Equação 27}$$

É importante notar que existe uma pequena diferença entre o \mathbf{r}^\perp e \mathbf{r}_k^* , conforme Figura 12, que corresponde ao vetor residual (\mathbf{r}_{RES}), calculado como a parte do espectro que não é descrita pelo vetor NAS nem pelo vetor interferente, através de uma subtração vetorial, equação 28:

$$\mathbf{r}_{RES} = \mathbf{r} - \mathbf{r}_{INT} - \mathbf{r}_k^* \quad \text{Equação 28}$$

Após estarem definidos os três vetores, as CCM podem ser construídas com amostras consideradas dentro do controle (\mathbf{R}_{NOC}) onde o NOC indica amostras dentro do controle⁶⁶.

Para construir a carta NAS, os espectros das amostras dentro do controle (\mathbf{R}_{NOC}) são projetados no vetor de regressão NAS, ambas proporcionais à concentração do analito na amostra. O valor da projeção é um escalar chamado de valor de NAS. Para determinar os limites estatísticos para a carta NAS, o valor de NAS para cada um dos sinais instrumentais das amostras dentro do controle é calculado de acordo com a equação 29:

$$\mathbf{nas}_{NOC} = \mathbf{R}_{NOC}^T \mathbf{b}_k \quad \text{Equação 29}$$

onde \mathbf{nas}_{NOC} é um vetor com o valor de NAS para cada espectro das amostras dentro do controle.

Estes limites são calculados através da média dos valores de NAS ($\overline{\mathbf{nas}}_{NOC}$) e o desvio padrão (S_{NOC}) das amostras dentro do controle, de acordo com a equação 30 para um limite de 95% de confiança no caso deste trabalho. Com estes valores, os limites estatísticos são plotados na carta NAS.

$$\overline{\mathbf{nas}}_{NOC} \pm 2 \times S_{NOC} \quad \text{Equação 30}$$

O mesmo processo é realizado para novas amostras, ou conjuntos de amostras utilizadas para validar os limites da carta NAS, conforme equação 31:

$$\mathbf{nas}_{novas} = \mathbf{R}_{novas}^T \mathbf{b}_k \quad \text{Equação 31}$$

A carta de controle do interferente, por sua vez, tem como base a distância da projeção dos espectros das amostras dentro do controle no espaço do interferente (\mathbf{PP}^+), que foi determinado através da PCA. Desta forma, as amostras projetadas na área delimitada pela elipse formada pelas duas componentes principais, apresentada na Figura 13, são consideradas dentro de controle; e aquelas projetadas além desses limites, são consideradas fora de controle.

Inicialmente, os espectros das amostras dentro de controle são projetados sobre o espaço dos interferentes, de acordo com a equação 32:

$$\mathbf{R}_{\text{INT}} = \mathbf{P}\mathbf{P}^+\mathbf{R}_{\text{NOC}} \quad \text{Equação 32}$$

onde \mathbf{R}_{INT} é a matriz dos interferentes composta pelas amostras dentro de controle de dimensão $I \times J$, e I é o número de amostras.

Em seguida, \mathbf{R}_{INT} é projetado no modelo PCA para obterem-se os valores de escores, conforme equação 33:

$$\mathbf{T}_{\text{NOC}} = \mathbf{R}_{\text{INT}}^T \mathbf{P}(\mathbf{P}^T \mathbf{P})^{-1} \quad \text{Equação 33}$$

onde \mathbf{T}_{NOC} é a matriz de escores com dimensão $I \times A$, em que A é o número de componentes principais no modelo PCA.

Portanto, a distância das amostras até o centro da elipse pode ser calculada através da equação 34:

$$\mathbf{d}_{\text{NOC},i} = (\mathbf{t}_{\text{NOC},i} - \bar{\mathbf{t}}_{\text{NOC}})^T \mathbf{S}^{-1} (\mathbf{t}_{\text{NOC},i} - \bar{\mathbf{t}}_{\text{NOC}}) \quad \text{Equação 34}$$

onde $\mathbf{d}_{\text{NOC},i}$ é a distância dos interferentes até o centro da elipse formados pelas duas primeiras PCs, no caso do exemplo da Figura 13, $\mathbf{t}_{\text{NOC},i}$ é o vetor escore com dimensão $1 \times A$, $\bar{\mathbf{t}}_{\text{NOC}}$ é o valor médio dos escores contidos na matriz \mathbf{T}_{NOC} , \mathbf{S} é a matriz de covariância centrada na média dos escores com dimensão $A \times A$.

Já o limite para carta do interferente é determinado através da equação 35:

$$\mathbf{D}_{\text{lim},0,95} = \mathbf{F}_{0,95(A, I_{\text{NOC}}-A)} \left(\frac{A(I_{\text{NOC}}^2 - I_{\text{NOC}})}{I_{\text{NOC}}(I_{\text{NOC}}-A)} \right) \quad \text{Equação 35}$$

onde $\mathbf{F}_{0,95}$ é valor da distribuição F com 95% de confiança em que o índice $(I_{\text{NOC}}-A)$ refere-se ao número de graus de liberdade, I_{NOC} é o número de amostras dentro do controle e A é número de componentes principais.

Um processo semelhante é realizado para as novas amostras, de acordo com as equações 36 a 38.

$$\mathbf{R}_{\text{INT},\text{novas}} = \mathbf{P}\mathbf{P}^+\mathbf{R}_{\text{novas}} \quad \text{Equação 36}$$

$$\mathbf{T}_{\text{novas}} = \mathbf{R}_{\text{INT},\text{novas}}^T \mathbf{P}(\mathbf{P}^T \mathbf{P})^{-1} \quad \text{Equação 37}$$

$$\mathbf{d}_{\text{novas},i} = (\mathbf{t}_{\text{NOC},1} - \bar{\mathbf{t}}_{\text{NOC}})^T \mathbf{S}^{-1} (\mathbf{t}_{\text{NOC},i} - \bar{\mathbf{t}}_{\text{NOC}}) \quad \text{Equação 38}$$

Os espectros das amostras dentro do controle também são utilizados para calcular o limite da carta do resíduo, que se baseia no Q-statistic^{67,69}. Para determinar os vetores residuais dos sinais instrumentais das amostras dentro do controle, inicialmente calcula-se a matriz NAS ($\mathbf{R}_{\text{NOC}}^*$), uma vez que a carta dos resíduos representa a parte do sinal que não foi modelada através da carta NAS e dos interferentes, conforme equação 39:

$$\mathbf{R}_{\text{NOC}}^* = \mathbf{b}_k (\mathbf{b}_k^T \mathbf{b}_k)^{-1} \mathbf{b}_k^T \mathbf{R}_{\text{NOC}} \quad \text{Equação 39}$$

Em seguida a matriz resíduo (\mathbf{R}_{RES}), pela equação 40:

$$\mathbf{R}_{\text{RES}} = \mathbf{R}_{\text{NOC}} - \mathbf{R}_{\text{INT}} - \mathbf{R}_{\text{NOC}}^* \quad \text{Equação 40}$$

E de acordo com Jackson e Mudholkar,⁶⁹ o limite é calculado da seguinte forma:

$$\mathbf{Q}_{\text{NOC}} = \mathbf{R}_{\text{RES}}^T \mathbf{R}_{\text{RES}} \quad \text{Equação 41}$$

onde \mathbf{Q}_{NOC} é a matriz de dimensão $I \times I$.

A diagonal da matriz \mathbf{Q}_{NOC} contém os valores Q-statistic que são plotados na carta de resíduos. O limite de confiança Q é calculado conforme as equações 42 a 44:

$$Q_\alpha = \theta_1 \left[c_\alpha \frac{\sqrt{2\theta_2 h_0^2}}{\theta_1} + 1 + \frac{\theta_2 h_0 (h_0 - 1)}{\theta_1^2} \right]^{\frac{1}{h_0}} \quad \text{Equação 42}$$

onde C_α é o desvio padrão normal.

$$\theta_i = \sum_{j=A+1}^n \lambda_j^i \quad i = 1, 2, 3 \quad \text{Equação 43}$$

onde A corresponde ao número de PCs retidos no modelo, e “n” é o número total de componentes principais (ou número total de amostras).

$$h_0 = 1 - \left(\frac{2\theta_1 \theta_3}{3\theta_2^2} \right) \quad \text{Equação 44}$$

O mesmo processo é realizado para novas amostras, de acordo com as equações 45 a 47.

$$\mathbf{R}_{\text{NOC}}^* = \mathbf{b}_k (\mathbf{b}_k^T \mathbf{b}_k)^{-1} \mathbf{b}_k^T \mathbf{R}_{\text{NOC}} \quad \text{Equação 45}$$

$$\mathbf{R}_{\text{RES}} = \mathbf{R}_{\text{NOC}} - \mathbf{R}_{\text{INT}} - \mathbf{R}_{\text{NOC}}^* \quad \text{Equação 46}$$

$$\mathbf{Q}_{\text{NOC}} = \mathbf{R}_{\text{RES}}^T \mathbf{R}_{\text{RES}} \quad \text{Equação 47}$$

Como exemplo de aplicações das CCM, um dos primeiros trabalhos que envolveram cartas de controle multivariado, em 2005, foi o de Skibsted e colaboradores para monitorar a concentração de um princípio ativo em formulações farmacêuticas⁶⁴.

Outros trabalhos interessantes incluem o de Rocha e colaboradores que utilizaram as cartas de controle multivariadas baseadas na divisão do espectro de infravermelho próximo em analito (NAS), interferente e resíduo para monitorar a concentração de um fármaco (nimesulida) e identificar a adulteração da composição do excipiente em formulações farmacêuticas. A carta NAS foi utilizada para verificar a concentração de nimesulida, e foi capaz de indicar corretamente as amostras de acordo com a concentração do princípio ativo no fármaco. Já a composição do excipiente foi monitorada através da carta dos interferentes, indicando também corretamente quando houve variação da composição⁷⁰.

Já em 2009, Oliveira e colaboradores utilizaram as CCM para monitorar tanto o teor de biodiesel em blendas de biodiesel/diesel (B2), quanto a adulteração com óleo de soja e contaminação com gasolina. Através da carta NAS foi possível determinar o teor de biodiesel na blenda e a composição do diesel através da carta dos interferentes⁷¹.

Em 2011, dos Santos Junior também utilizou as cartas para monitorar os principais parâmetros de qualidade de óleos lubrificantes de motor como viscosidade dinâmica e cinemática, índice de viscosidade e a concentração de elementos como cálcio, magnésio, fósforo e zinco. Neste trabalho foi possível determinar as amostras que estavam fora dos parâmetros de qualidade⁶⁵.

No trabalho de Mitsutake e colaboradores publicado em 2015, as cartas de controle com base no NAS e infravermelho médio, foram utilizadas para controlar a qualidade de misturas de B5 de biodiesel/Diesel. Foram testadas novas amostras-controle e amostras fora de controle em relação ao teor de biodiesel, além de biodiesel adulterado com óleos vegetais e diesel adulterado com óleo lubrificante automotivo residual, querosene e gasolina. Os resultados obtidos apresentam uma excelente distinção entre as amostras dentro e fora das especificações de qualidade, com 91% e 100% respectivamente, que foram corretamente classificados, o que demonstra que a metodologia desenvolvida é uma alternativa viável para o monitoramento deste tipo de combustível⁷².

2. OTIMIZAÇÃO DO PROCESSO TDSP PARA PRODUÇÃO DE BIODIESEL EMPREGANDO PLANEJAMENTO EXPERIMENTAL DOEHLERT

2.1 Objetivo

Este capítulo tem o objetivo de apresentar a otimização do processo de produção de biodiesel intitulado *Transesterification Double Step Process* (TDSP) através da aplicação do planejamento experimental Doehlert.

2.2 Materiais e métodos

2.2.1 Reagentes

Para a produção do biodiesel este estudo utilizou o óleo de soja comercial marca Leve (Cambé, Brasil), além de álcool metílico (MeOH) (P.A.) - Neon (São Paulo, Brasil) e os catalisadores foram o hidróxido de potássio (KOH) (P.A.) - Synth (Diadema, Brasil) e o ácido sulfúrico (H₂SO₄) (P.A.) - Proquímios (Rio de Janeiro, Brasil). Também foram empregados sulfato de sódio - (P.A.) – Dinâmica (Diadema, Brasil) e clorofórmio deuterado (99,8% D com 0,05% (v/v) de TMS) - Cambridge Isotope Laboratories, Inc. (Andover, USA) para as análises de RMN.

2.2.2 Softwares

Os dados do planejamento Doehlert foram processados utilizando planilhas eletrônicas do Microsoft Excel desenvolvidas pelo Laboratório de Quimiometria Teórica e Aplicada da Universidade Estadual de Campinas (UNICAMP)⁷³.

2.3 Planejamento Doehlert

Na Figura 13 são apresentados os níveis para os parâmetros estudados em cada fase da síntese do biodiesel metílico de óleo de soja. Primeiramente foram estudadas as variáveis julgadas como mais importantes, em função do conhecimento prévio da metodologia TDSP, para a primeira etapa da reação que corresponde à catálise básica: a massa de catalisador (g de KOH) e o volume de álcool

(mL de CH_3OH). Dentre essas duas variáveis a quantidade de catalisador foi variada em 5 níveis e a de metanol em 3, uma vez que neste processo o metanol pode ser reaproveitado na segunda etapa da reação, enquanto que o catalisador tem maior custo. Assim a massa de KOH utilizada foi de 0,30 a 0,70 g e o volume de CH_3OH variou entre 20 e 60 mL. Na segunda etapa da reação que corresponde à catálise ácida foram estudadas as variáveis: tempo (min.) e temperatura de reação ($^{\circ}\text{C}$). Para estas variáveis o tempo foi estudado em 5 níveis e a temperatura em 3. Desta forma, o tempo de reação variou de 30 a 150 min. e a temperatura de 55 a 75 $^{\circ}\text{C}$. Foram realizadas ainda reações em triplicata referentes ao ponto central do planejamento Doehlert.

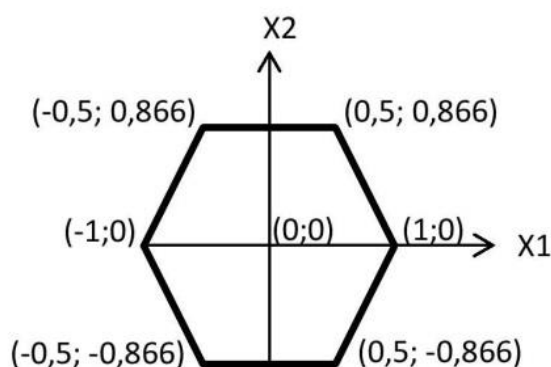


Figura 13. Valores codificados para planejamento Doehlert com ponto central em triplicata.

2.3.1 Produção do biodiesel

Para este estudo foram produzidas, no total, 18 amostras de acordo com o planejamento experimental Doehlert e com base na metodologia TDSP para produção de biodiesel metílico. Para avaliar as condições da primeira parte da reação de transesterificação (catálise básica) e da segunda parte (catálise ácida) foram sintetizadas no total 9 amostras, para cada etapa, estando incluídas as 3 replicatas referentes ao ponto central.

No geral, independentemente das quantidades e condições de reação utilizadas, o processo de produção segue um padrão estabelecido pelo método TDSP. Primeiramente 100 mL de óleo foram aquecidos até a temperatura de refluxo do metanol. Em seguida, foi adicionada a mistura de KOH e metanol que originam a espécie ativa do catalisador (metóxido) e são mantidos, na primeira etapa da reação a 65 $^{\circ}\text{C}$ por 30 min. Após este tempo, foi adicionado o ácido sulfúrico e mais metanol utilizado

em excesso, para garantir o deslocamento da reação no sentido dos produtos, conforme Figura 14. Ao final, o produto foi colocado no evaporador rotativo para retirada do metanol excedente, transferido para um funil de separação para retirada do glicerol formado e foram feitas sucessivas lavagens para neutralizar o pH. O biodiesel foi seco em sulfato de sódio e por fim, passou por um processo de aquecimento para evaporação de resquícios de água, foi filtrado e envasado.

É importante observar que para os experimentos da catálise básica, a segunda parte da reação foi mantida constante, com quantidades de ácido sulfúrico e metanol iguais a 0,30 mL e 20 mL, respectivamente. Já para as variáveis tempo e temperatura na segunda etapa da reação, foram analisados os resultados de conversão obtidos por espectros de ressonância magnética nuclear de Hidrogênio (RMN ^1H) da primeira etapa. Assim as quantidades de catalisador e metanol que resultaram na maior conversão foram fixadas para o segundo estudo (catálise ácida).

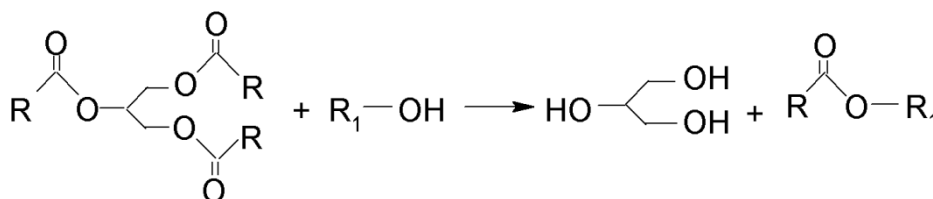


Figura 14. Esquema representativo da reação geral de transesterificação do óleo de soja com álcool primário (metanol) produzindo biodiesel e glicerol como subproduto.

2.3.2 Caracterização do biodiesel por ressonância magnética nuclear de hidrogênio (RMN ^1H)

Os produtos finais da síntese do biodiesel foram caracterizados por Ressonância Magnética Nuclear de Hidrogênio (RMN ^1H). Para a análise foram colocados 15 μL de biodiesel em um tubo de RMN e adicionados 550 μL de CDCl_3 . As amostras foram analisadas num espectrômetro Bruker *AVANCE* operando na frequência de 400 MHz para ^1H e 100 MHz para ^{13}C , equipado com uma sonda de 5 mm com detecção direta. Os espectros foram obtidos selecionando-se alguns parâmetros na aquisição e no processamento, para fornecer uma melhor resolução digital com base nos parâmetros. As condições de análise foram 16 scans e $\text{td} = 256\text{k}$, janela espectral de 5,5 ppm, $\text{lb} = 0,30\text{ Hz}$; $\text{si} = 256\text{ k}$ fornecendo uma resolução digital de 0,017 Hz/ponto. Esta técnica é rápida e simples em comparação com análises cromatográficas que exigem preparação da amostra e/ou purificação. Na análise de RMN ^1H , as informações sobre a amostra são obtidas sem necessitar derivatização ou causar inutilização da amostra, além de ser necessária uma quantidade muito pequena de amostra.

Uma avaliação dos espectros de óleo de soja e do biodiesel metílico de óleo de soja permite inicialmente avaliar o teor residual de tri, di e monoglicerídeos nos ésteres metílicos resultantes,

através do sinal correspondente aos hidrogênios metilênicos da sequência glicérica (-CH₂-CH-CH₂-). Estes sinais localizados entre 4,1 e 4,3 ppm correspondem a quatro átomos de hidrogênio. Já o sinal referente ao grupo CH está localizado a 5,2 ppm, juntamente com sinais de átomos de hidrogênio das duplas ligações. Os sinais de mono e diglicerídeos, formados como intermediários da reação exibem sinais dos hidrogênios glicéricos na mesma região que hidrogênios glicéricos de triglicerídeos da matéria-prima, Figura 15.

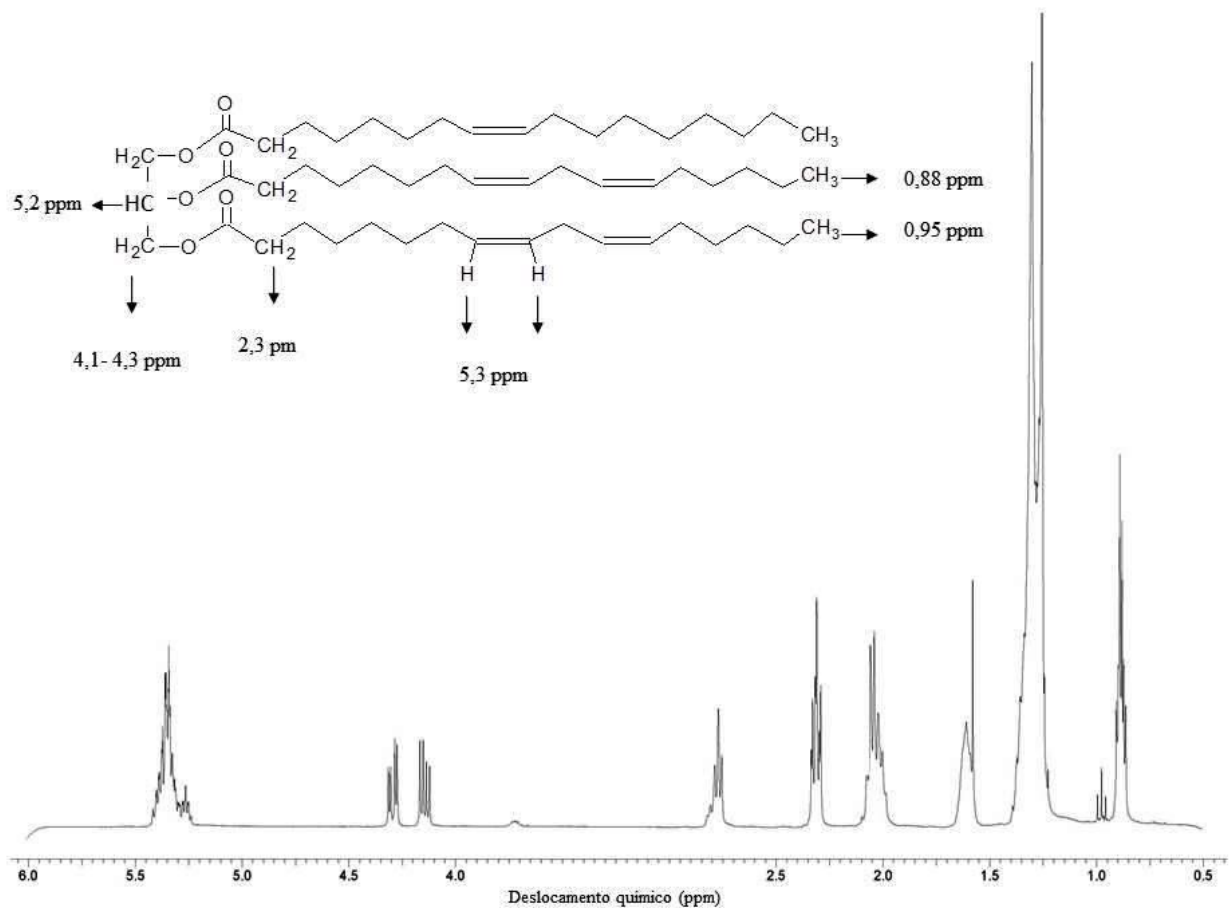


Figura 15. Espectro de RMN ¹H de óleo de soja.

Como pode ser visto no exemplo da Figura 16, não há nenhum resíduo de mono, di ou triglicerídeos, o que fica evidenciado a partir do desaparecimento total de sinais entre 4,1 e 4,3 ppm. Assim, para avaliar a conversão de óleos em ésteres metílicos de ácidos graxos, é observado o aparecimento de um novo sinal a 3,6 ppm. Este sinal indica a formação de biodiesel, porque está na região do espectro correspondente aos átomos de hidrogênio do grupo éster metílico (COOCH₃). A conversão foi analisada pelo surgimento deste sinal além da comparação com os sinais que desaparecem relacionados com o óleo de partida, em virtude de outros produtos secundários poderem

ter se formado nesta mesma região do espectro de RMN. Para avaliar a conversão em éster metílico (C_{ME}), utilizou-se o sinal do hidrogênio do grupo metileno adjacente ao grupo carbonila ($\alpha\text{-CH}_2$) em 2,3 ppm. Este sinal foi escolhido como referência para o cálculo, uma vez que está presente em todas as moléculas de derivados de triglicerídeos, incluindo o óleo que não reagiu, mono e diglicerídeos e todas as outras espécies de cadeia longa que podem ser formados durante o procedimento. A Equação 48 apresenta o cálculo da conversão da reação, pela relação dos valores da integral dos sinais de RMN de hidrogênio a partir dos hidrogênios do éster metílico (ME) e dos átomos de hidrogênio no grupo metileno adjacente ($\alpha\text{-CH}_2$) à carbonila, apresentados na Figura 16.

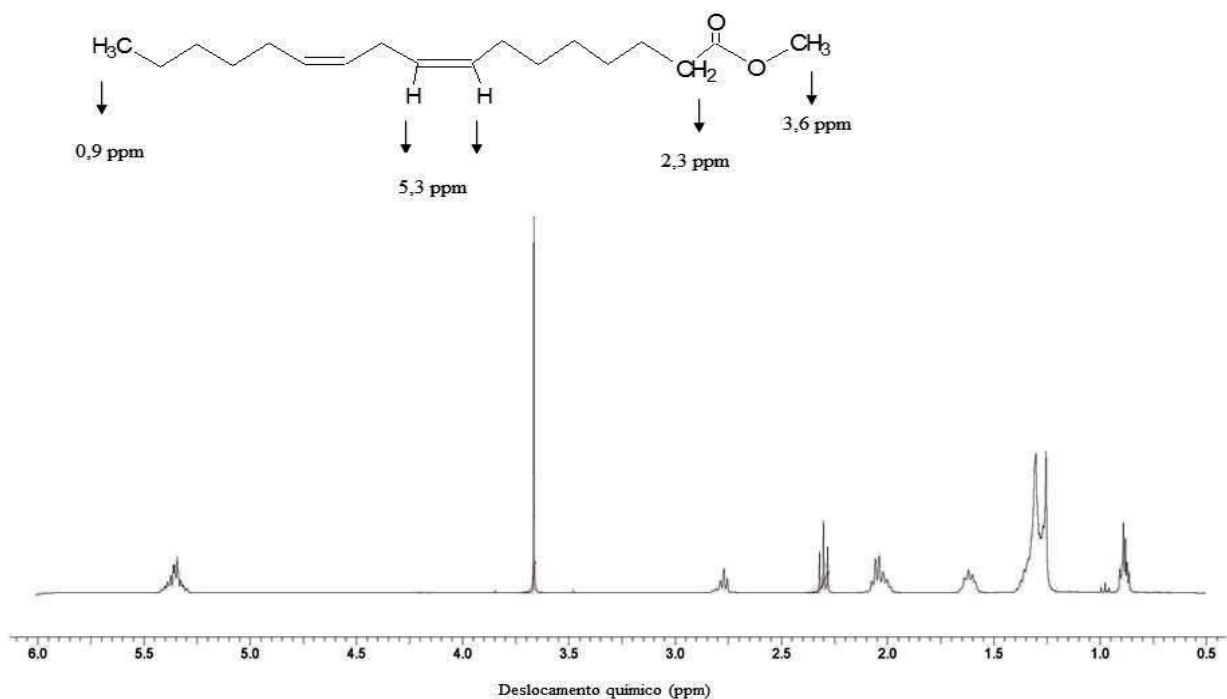


Figura 16. Espectro de RMN ^1H de biodiesel de óleo de soja produzido pelo método TDSP (experimento 2 da etapa 1).

$$C_{ME} = 100 \times \left(\frac{\frac{I_{ME}}{3}}{\frac{I_{\alpha\text{-CH}_2}}{2}} \right) \quad \text{Equação 48}$$

I_{ME} é o valor da integral do sinal do éster metílico e $I_{\alpha\text{-CH}_2}$ é o valor da integral do grupo metileno adjacente ao grupo éster. Assim a integral do éster metílico foi dividida pelo fator 3 e a integral do grupo metileno foi dividida pelo fator 2, uma vez que o éster metílico tem três hidrogênios em cada

sinal -OCH₃ em 3,6 ppm e dois hidrogênios do sinal referente ao grupo metileno adjacente a carbonila (α -CH₂) em 2,3 ppm²⁶.

Utilizando a Equação 41, para o cálculo da conversão do óleo de soja em ésteres metílicos, tem-se como exemplo, para o experimento número 2, o seguinte resultado:

$$C_{ME} = 100 \times \left(\frac{\frac{1}{3}}{\frac{0,67}{2}} \right) = 99,50 \%$$

Tabela II. Valores de conversão obtidos utilizando condições reacionais testadas para catálise básica

Experimento	Catalisador (g de KOH) (X1)	Metanol (mL) (X2)	Conversão (%)
1	0,70 (1)*	40 (0)*	99,50
2	0,60 (0,5)*	60 (0,866)*	99,50
3	0,30 (-1)*	40 (0)*	96,62
4	0,40 (-0,5)*	20 (-0,866)*	91,32
5	0,60 (0,5)*	20 (-0,866)*	93,90
6	0,40 (-0,5)*	60 (0,866)*	100,00
7	0,50 (0)*	40 (0)*	99,50
8	0,50 (0)*	40 (0)*	98,04
9	0,50 (0)*	40 (0)*	99,50

* Os valores entre parênteses são os valores codificados do planejamento Doehlert

Tabela III. Valores de conversão obtidos utilizando condições reacionais testadas para catálise ácida segundo planejamento Doehlert

Experimento	Tempo (min.) (X1)	T (°C) (X2)	Conversão (%)
10	150 (1)*	65 (0)*	99,42
11	120 (0,5)*	75 (0,866)*	99,42
12	30 (-1)*	65 (0)*	90,10
13	60 (-0,5)*	55 (-0,866)*	98,03
14	120 (0,5)*	55 (-0,866)*	95,24
15	60 (-0,5)*	75 (0,866)*	99,50
16	90 (0)*	65 (0)*	99,84
17	90 (0)*	65 (0)*	99,60
18	90 (0)*	65 (0)*	99,99

* Os valores entre parênteses são os valores codificados do planejamento Doehlert

Para os demais experimentos as conversões foram calculadas da mesma forma, gerando os dados apresentados nas Tabelas II e III, com as variáveis estudadas na catálise básica e na catálise ácida, respectivamente.

2.4 Resultados e Discussão

2.4.1 Análise da conversão para catálise básica segundo planejamento Doehlert

Segundo Samios e colaboradores, para a catálise básica, a quantidade estequiométrica ideal dos reagentes é de 0,10 mols de óleo: 1 mol de álcool: 0,001 mols de catalisador, aproximadamente, ou seja, para 100 mL de óleo utilizados, deveriam ser colocados 40 mL de metanol e 0,5 g de catalisador, que foram as condições utilizadas para o ponto central do planejamento²⁴.

Uma boa conversão de uma reação de transesterificação de óleo ou gordura, segundo o que é exigido pela ANP para um biodiesel comercial, é de pelo menos 96,5%. Analisando os dados da Tabela II, iniciando pelos experimentos 7, 8 e 9, referentes ao ponto central do planejamento Doehlert foram obtidos, em média, valores de conversão em torno de 99%. Partindo do ponto central, no qual foi utilizado 40 mL de metanol, pode-se observar também os experimentos 1 e 3, que utilizaram a mesma quantidade de metanol, porém variando a quantidade de catalisador que foi diminuída para 0,30 no experimento 3 e foi aumentada para 0,70 no experimento 1. Esta variação resultou em valores de conversão maiores quando a quantidade de catalisador foi aumentada e em conversões menores quando a quantidade de catalisador foi diminuída mostrando a sua influência na conversão.

Quando o volume de metanol foi fixado em 60 mL, que foi o volume máximo testado, nos experimentos 2 e 6, a conversão aumentou para 100% (maior valor obtido entre todos os experimentos) quando a massa de catalisador utilizada foi de 0,40 g e permaneceu praticamente igual a conversão do ponto central quando a massa de catalisador é de 0,60 g; o que também mostra a maior relevância da quantidade de metanol em relação a quantidade de catalisador.

Da mesma forma, quando foi diminuído o volume de metanol para 20 mL, independentemente da quantidade de catalisador utilizada, o percentual de conversão da reação de transesterificação diminuiu consideravelmente. Além disso, nota-se também que quando se soma a redução do valor da outra variável, isto também contribui para a diminuição da conversão como é observado no experimento 4.

Quando analisamos do ponto de vista da mesma quantidade de catalisador, nos experimentos 4 e 6, confirmamos que o volume de metanol é primordial para uma boa conversão da reação. Assim

como no experimento 3, onde é diminuído ainda mais a quantidade de catalisador, porém a de metanol aumenta, o que aumenta também a conversão. Já analisando os experimentos 2 e 5, que utilizam maiores quantidades de catalisador, verificamos também que a conversão é muito maior quando o volume de metanol utilizado também é alto. Além disso, quando a quantidade de catalisador é a maior, como no experimento 1, porém se diminui a de metanol, como acontece no restante dos experimentos, a conversão fica acima de 99%.

As observações anteriores comprovam que o metanol, como é um reagente que participa da reação de transesterificação, é fundamental para a produção de biodiesel, e que a sua quantidade mínima ideal, para 100 mL de óleo, é de 60 mL para obtenção de uma conversão ótima de 100%. Já a quantidade mínima de catalisador, para 100 mL de óleo, é de 0,40 g.

2.4.2 Análise da conversão para catálise ácida segundo planejamento Doehlert

Segundo Samios e colaboradores, para a catálise ácida, a temperatura ideal de reação que é a temperatura de refluxo do solvente (metanol) é de aproximadamente 65 °C e o tempo de reação necessário é de 90 minutos²⁴.

Analisando os dados da Tabela III, os experimentos 16, 17 e 18 são referentes às condições determinadas para o ponto central do planejamento experimental Doehlert, ou seja, tempo de uma hora e meia e temperatura de 65 °C. Nessas condições, foram obtidos ótimos valores de conversão, em média, de 99,8%, mostrando que a combinação desta temperatura e tempo de reação para a catálise ácida é suficiente para a obtenção de biodiesel. Além disso, foram estudados outros tempos de reação enquanto se manteve constante a temperatura em 65 °C. Nessas condições, nos experimentos 10 e 12, foi observado que quando o tempo foi reduzido 3 vezes houve uma queda na conversão e quando o tempo foi aumentado consideravelmente não houve grande melhora na conversão.

Quando a temperatura de reação foi reduzida em 10 °C, ou seja, passou para 55 °C, nos experimentos 13 e 14, foram estudadas tempos de reação de uma e duas horas. Foi observado que com o aumento do tempo de reação houve diminuição na conversão. Por fim, foi avaliada a condição de temperatura fixada em 75 °C, 10 °C a mais que o ponto central, com tempos de reação de uma e duas horas, nos experimentos 11 e 15. Nestes casos, independentemente do tempo de reação obteve-se valores de conversão muito próximos.

Quando se observam os dados a partir da variação da temperatura, fixando para fins de comparação, o tempo, como nos experimentos 13 e 15, onde o tempo de reação é de uma hora e a temperatura foi aumentada de 55 para 75 °C, verificou-se um aumento pequeno na conversão. Já para

o caso dos experimentos 11 e 14, nos quais o tempo foi fixado em duas horas e a temperatura foi aumentada de 55 para 75 °C, observou-se também uma redução do valor de conversão obtido.

Conclui-se que, em geral, para os experimentos de variação de temperatura e tempo de reação da catálise ácida, há um efeito sobre o equilíbrio dessa reação, no qual o aumento de uma das variáveis avaliadas, combinado com a redução da outra, resultam em melhores valores de conversão. Tempos e temperaturas de reação muito altos não favorecem necessariamente a reação, deslocando o equilíbrio da reação no sentido dos reagentes, como no caso do experimento 11. Assim, se necessitamos, por exemplo, a otimização do estudo para tempos menores, devemos utilizar uma temperatura mais elevada de reação. De maneira geral, uma boa condição é a que foi utilizada no ponto central, com temperatura de 65 °C e tempo de 90 minutos, obtendo valores satisfatórios de conversão. Além disso, houveram resultados não esperado como é o caso dos valores de conversão dos experimentos 13 e 14, e que podem ser explicados pelo fato de que possivelmente a influência da variação de 10 °C na temperatura estudada, não tenha sido suficientemente grande para gerar diferenças significativas nos percentuais de conversão.

2.4.3 Análise do gráfico de probabilidade normal e da superfície de resposta

Após terem sido calculados os percentuais de conversão através dos espectros de RMN ¹H, os dados foram processados utilizando planilhas eletrônicas desenvolvidas pelo Laboratório de Quimiometria Teórica e Aplicada da Universidade Estadual de Campinas (UNICAMP)⁷¹. No caso da catálise básica, como pode ser visto na Figura 17, a variável que mais influencia é a variável 2, ou seja, o volume de metanol utilizado; em seguida a variável 1, ou seja, a quantidade de catalisador (KOH) foi a segunda mais importante para a reação. Na sequência, é mostrado que o efeito de segunda ordem para a variável 2 foi o terceiro mais importante e que a combinação das variáveis 1 e 2, e o efeito de segunda ordem para a variável 1 não apresentaram influência sobre a reação.

Além do gráfico de probabilidade normal, os resultados dos experimentos realizados com base nos planejamentos Doehlert, como a conversão em ésteres metílicos (%) determinada a partir do RMN ¹H, possibilitaram gerar a superfície de resposta, que está apresentada na Figura 18. Foram testados modelos lineares e quadráticos, porém apesar de ambos os modelos apresentarem homocedasticidade para os resíduos e regressão significativa com falta de ajuste não significativa, o modelo linear apresentou, para a catálise básica, um percentual de variância explicada de somente 78,95%. Desta forma, o modelo quadrático foi escolhido por representar melhor o comportamento dos dados e a função matemática ajustada para a catálise básica é apresentada na equação 49 referente

a superfície de resposta apresentada na Figura 18. Para esta equação foi obtida uma variância explicada de aproximadamente 98%. Para ambos os modelos foi considerado um nível de significância de 0,05, sendo os resultados da análise da variância apresentados na Tabela IV.

$$Y_1 = 99,013 + 1,320(X_1) + 4,145(X_2) - 0,953(X_1^2) - 3,434(X_2^2) - 1,732(X_1)(X_2) \quad \text{Equação 49}$$

onde Y_1 é o percentual de conversão, X_1 é a massa de catalisador e X_2 é o volume de metanol.

Tabela IV. Análise de Variância - Modelo Quadrático para catálise básica

FV*	SQ*	nGL*	MQ*	Fcalc.*	p*
Regressão	73,26	5	14,65	29,16 SG	0,0095
Resíduos	1,51	3	0,50		
F. Ajuste	0,086	1	0,086	0,012	0,76
Erro Puro	1,42	2	0,71		
Total	74,77	8			
% variação explicada				97,98	
% máx. de variação explicável				98,10	

*Onde: SG=Valor significativo, FV= Fonte de variação, SQ= soma quadrática, nGL= n° de graus de liberdade, MQ= média quadrática, Fcalc. = valor de F calculado e p= nível de significância.

Como pode ser visto na Tabela IV o modelo quadrático proposto para a catálise básica apresenta uma regressão significativa e uma falta de ajuste não significativa. Desta forma, se observa ainda que a maior parte da variação é descrita pela equação de regressão e o restante pelos resíduos. Da variação contida nos resíduos, a maior parte é devido ao erro puro (erro experimental), e o restante é descrito pela falta de ajuste que está diretamente relacionada com o modelo.

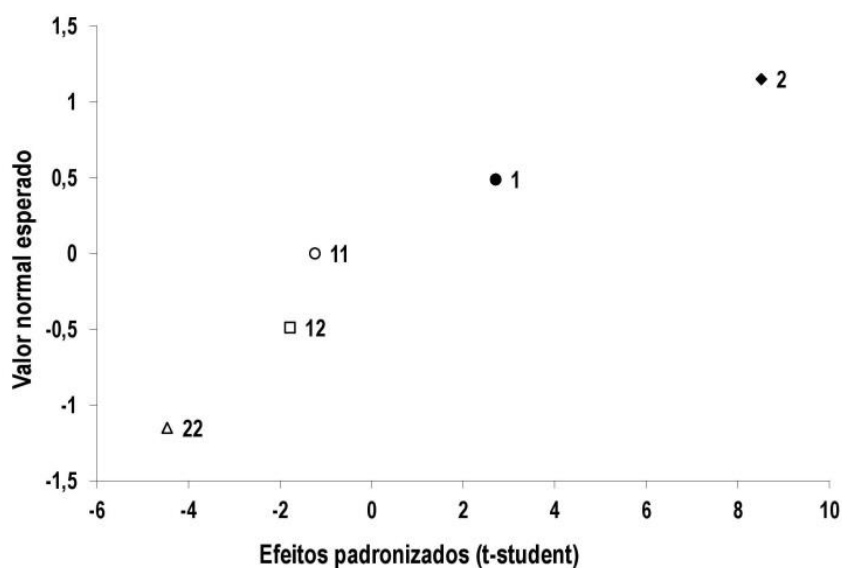


Figura 17. Gráfico de probabilidade normal para a massa de catalisador (g) e volume de metanol (mL) na primeira parte da reação.

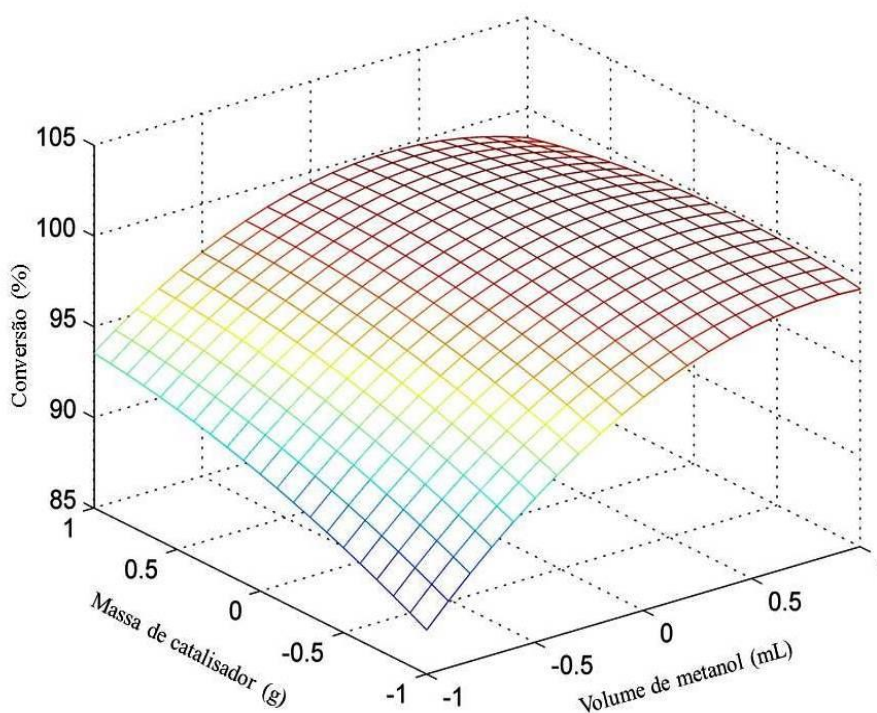


Figura 18. Superfície de resposta do planejamento Doehlert para a massa de catalisador (g) e volume de metanol (mL) na primeira parte da reação.

O gráfico de probabilidade normal, Figura 17, e o resultado obtido para a superfície de resposta, Figura 18, sugerem a maior influência do volume de metanol no aumento da conversão da etapa

básica da reação de transesterificação, em relação à massa de catalisador. Já para a catálise ácida, como pode ser visto nas Tabelas V e VI, tanto o modelo linear quanto o quadrático propostos apresentaram falta de ajuste significativa.

Tabela V. Análise de Variância - Modelo Linear para catálise ácida

FV*	SQ*	nGL*	MQ*	Fcalc.*	p*
Regressão	30,54	3	10,20	0,91	0,50
Resíduos	55,59	5	11,10		
Falta de Ajuste	55,52	3	18,50	478,18 SG	0,021
Erro Puro	0,077	2	0,040		
Total	86,13	8			
% variação explicada				35,46	
% máx. De variação explicável				99,91	

*Onde: FV= Fonte de variação, SQ= soma quadrática, nGL= n° de graus de liberdade, MQ = média quadrática, Fcalc. = valor de F calculado e p= nível de significância, SG = Valor significativo.

Tabela VI. Análise de Variância - Modelo Quadrático para catálise ácida

FV*	SQ*	nGL*	MQ*	Fcalc.*	p*
Regressão	61,29	5	12,30	1,48	0,40
Resíduos	24,84	3	8,28		
Falta de Ajuste	24,77	1	24,80	639,95 SG	0,0016
Erro Puro	0,077	2	0,040		
Total		8			
% variação explicada				71,16	
% máx. de variação explicável				99,91	

*Onde: FV= Fonte de variação, SQ= soma quadrática, nGL= n° de graus de liberdade, MQ = média quadrática, Fcalc. = valor de F calculado e p= nível de significância, SG = Valor significativo.

Com base nos resultados apresentados, se pode concluir que a aplicação do planejamento experimental Doehlert possibilitou a otimização da reação de transesterificação do óleo de soja com álcool metílico através da metodologia *Transesterification Double Step Process* (TDSP) para a produção de biodiesel. Para a catálise básica foi obtido um modelo satisfatório, com intervalo de confiança de 95%, uma vez que, a regressão obtida foi significativa ($p < 0,05$) e o valor de F calculado foi aproximadamente 3 vezes maior que o F teórico. Além disso, não houve falta de ajuste significativo e a maior parte da variação descrita pelos resíduos está relacionada ao erro puro (experimental), o que é o ideal. A melhor conversão (100%) foi obtida utilizando 0,40 g de catalisador

e 60 mL de metanol. A variável metanol apresentou maior influência sobre o sistema seguida da variável catalisador. Para a catálise ácida não foi possível obter um modelo linear ou quadrático adequado para a variação de temperatura (de 10 °C) e de tempo (de 30 minutos) propostos para o estudo, uma vez que ambos apresentaram falta de ajuste, embora as condições descritas pelo ponto central já sejam suficientes para se obter ótimas conversões do ponto de vista da catálise ácida.

3. DETERMINAÇÃO DA ADULTERAÇÃO DA GASOLINA EMPREGANDO RMN ¹H E MÉTODOS QUIMIOMÉTRICOS

3.1 Objetivo

Neste capítulo, dados de ressonância magnética nuclear de hidrogênio aliados a técnicas quimiométricas de análises não-supervisionadas como HCA e PCA, além de regressão multivariada como PLS e análise supervisionada por PLS-DA, foram utilizados para avaliar a adulteração de gasolina.

3.2 Materiais e métodos

Primeiramente foram produzidas 25 amostras de gasolina, conforme apresentado na Tabela VII.

Para os estudos de classificação e quantificação as amostras foram preparadas pela adição de $25\% \pm 1$ (v/v) de etanol anidro na gasolina pura, para simular as amostras de acordo norma reguladora da ANP vigente na época do desenvolvimento desta parte do estudo. Desta forma, dez amostras contendo $75\% \pm 1$ de gasolina (v/v), sem adição de óleo diesel, formam o grupo não adulterado, enquanto que quinze amostras foram propositadamente adulteradas com 5 a 30% de óleo Diesel, substituindo gradativamente o conteúdo de gasolina (com volume de etanol mantido em 25 %) formando o grupo de amostras adulteradas.

Tabela VII. Percentual dos componentes nas misturas de gasolina

Amostra	% v/v gasolina	% v/v etanol	% v/v óleo Diesel
1	76	24	0
2	76	24	0
3	75,5	24,5	0
4	75,5	24,5	0
5	75	25	0
6	75	25	0
7	74,5	25,5	0
8	74,5	25,5	0
9	74	26	0
10	74	26	0
11	70	25	5
12	70	25	5
13	70	25	5
14	65	25	10
15	65	25	10
16	65	25	10
17	60	25	15
18	60	25	15
19	60	25	15
20	55	25	20
21	55	25	20
22	55	25	20
23	45	25	30
24	45	25	30
25	45	25	30

3.2.1 Softwares

A construção de todos os modelos quimiométricos (PCA, HCA, PLS, PLS-DA) utilizando dados de RMN ^1H foram implementados no software MATLAB R2013 (MathWorks, EUA) sendo também empregadas rotinas do PLS-Toolbox 6.2.1 (Eigenvector Research Inc.) e o pacote iToolbox.

3.2.2 Obtenção e Alinhamento dos Espectros de RMN ^1H

Para a aquisição dos espectros de RMN ^1H as amostras foram preparadas em tubos de RMN de 5 milímetros pela dissolução de 15 μL do combustível em 550 μL de clorofórmio deuterado (CDCl_3) e foram homogeneizadas por vórtex. As amostras foram introduzidas num espectrômetro Bruker Avance modelo com 400 MHz de frequência, equipado com uma sonda de 5 mm de detecção direta e um padrão de tetrametilsilano (TMS) foi utilizado como referência interna. Outros parâmetros

instrumentais importantes utilizados foram: janela espectral de 10,0127 ppm, número de pontos na aquisição = 64 k, ganho do receptor = 90,5.

O pré-tratamento dos dados de RMN, para posterior utilização nos modelos quimiométricos, consistiu no alinhamento espectral utilizando o algoritmo de alinhamento recursivo por transformada rápida de Fourier ou *Recursive Alignment by Fast Fourier Transform (RAFFT)*^{74,75}. Os espectros já alinhados são apresentados na Figura 19.

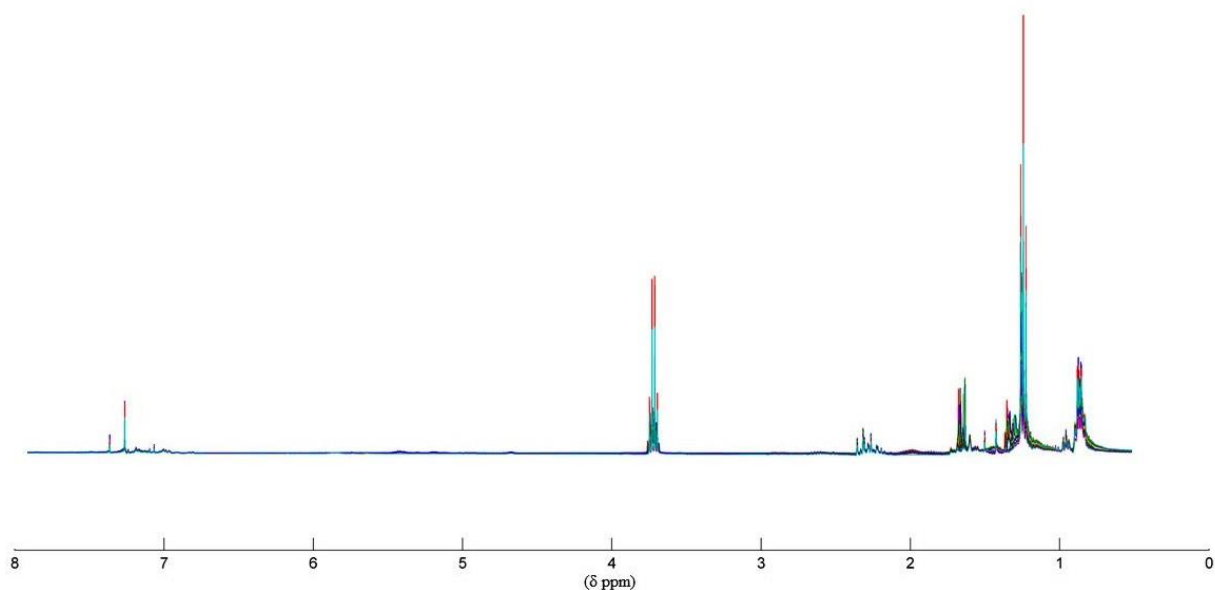


Figura 19. Espectros de RMN ¹H das misturas de gasolina alinhados pelo algoritmo *RAFFT*.

3.3 Resultados e Discussão

3.3.1 PCA e HCA

A análise de agrupamento hierárquico das amostras de gasolina foi realizada utilizando o método de Ward e com os dados previamente centrados na média. A Figura 20 apresenta o dendrograma formado a partir dos espectros de RMN ¹H.

Como pode ser visto na Tabela VII, as amostras foram divididas em não-adulteradas (amostras 1 a 10) e adulteradas (amostras 11 a 25). Para um agrupamento ideal, esperava-se que esta técnica separasse as amostras nestas duas categorias, o que não foi observado. De acordo com a Figura 20, formaram-se dois grupos: o grupo em verde são as amostras com maior adulteração (conteúdo de diesel adicionado igual a 15, 20 e 30% v/v) e em vermelho estão as amostras não adulteradas e

amostras com menor conteúdo de adulteração, são elas: 11, 12, 13, 14, 15, 16 e 19 (com teor de Diesel adicionado igual a 5 a 10% v/v além da amostra 19 com 15% v/v).

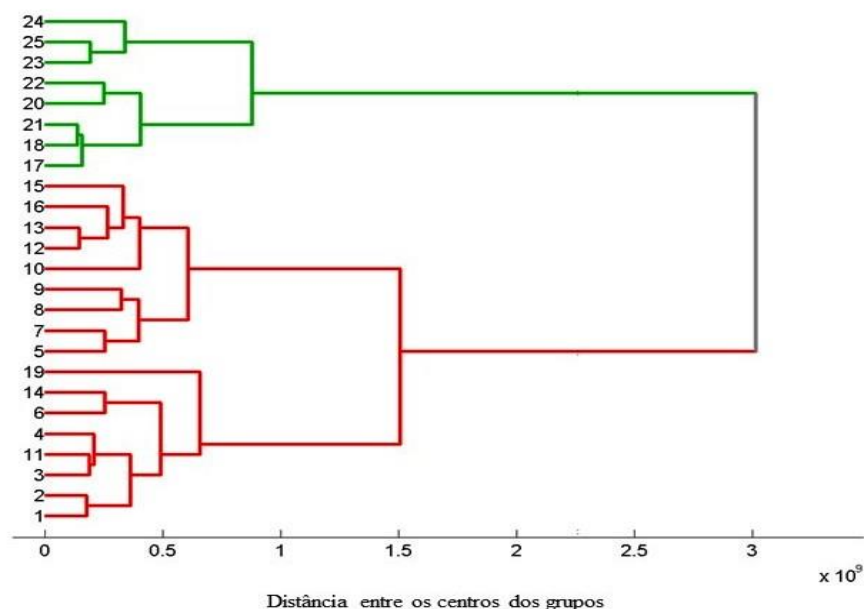


Figura 20. Dendrograma dos dados espectrais de RMN ^1H das misturas de gasolina.

Para a PCA, os dados também foram centrados na média e considerou-se cinco componentes principais que resultaram em uma variância acumulada de 91,11%. Os escores da PCA para a PC1, PC2 e PC3, com base nos espectros de RMN ^1H , estão apresentados na Figura 21.

De acordo com a Figura 21, podemos observar uma tendência de agrupamento de amostras não adulteradas (vermelho) e amostras adulteradas com óleo diesel (verde), especialmente na PC1, em que as amostras não adulteradas apresentam escores negativos enquanto as amostras adulteradas têm uma tendência a se agrupar com valores positivos ao longo da PC1, exceto para as amostras 11, 14 e 19. Além disso, não foi observado um padrão nas componentes principais PC1, PC2 e PC3 relacionadas com a adição de Diesel ou com a diminuição do teor em gasolina nas misturas.

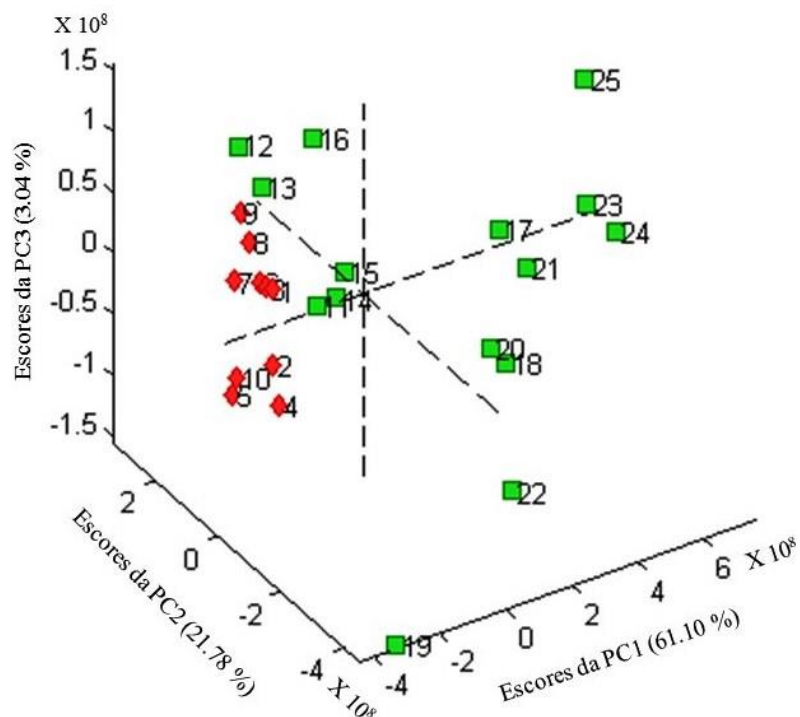


Figura 21. Análise de Componentes Principais de espectros de RMN ^1H das misturas de gasolina.

Portanto, com base nos resultados da HCA e PCA tornou-se apropriado avaliar os dados empregando o método de classificação PLS-DA e de calibração por PLS, uma vez que os métodos não supervisionados apesar de não serem capazes de agrupar as amostras de forma totalmente correta, indicaram que os dados de RMN ^1H podem ser diferenciados com relação à adulteração.

3.3.2 PLS-DA e PLS

Os modelos PLS-DA foram elaborados para classificar as amostras de gasolina adulteradas e não adulteradas enquanto os modelos PLS foram elaborados, a fim de quantificar o adulterante presente nas misturas, onde Y é igual à sua percentagem na mistura.

3.3.2.1 Definição dos conjuntos

Para o modelo PLS-DA empregando as amostras de gasolina adulteradas e não adulteradas, o conjunto de dados foi separado em 64% para o conjunto de treinamento e 36% para o conjunto de teste através do algoritmo de Kennard-Stone⁷⁶. Desta forma, um total de dezesseis amostras foram utilizadas no conjunto de treinamento, sendo dez amostras do conjunto adulterado e seis amostras de conjunto não adulterado. Além disso, para o conjunto teste foram utilizados um total de nove

amostras, sendo cinco amostras do conjunto adulterado e quatro amostras de conjunto não adulterado. Para ambos os modelos PLS e PLS-DA os dados foram centrados na média.

Os modelos PLS foram construídos com base nos dados de RMN e nas concentrações das misturas de gasolina. Primeiramente, foram corrigidas linha de base e a fase dos espectros de RMN. Depois, foi separado um conjunto de dados para calibração (60%) e de previsão (40%) através de algoritmo desenvolvido por Kennard-Stone. O número apropriado de variáveis latentes (VLs) foi determinado utilizando método de validação cruzada *leave-one-out*. O erro quadrático médio de validação cruzada ou *Root Mean Square Error Of Cross-Validation* (RMSECV), a raiz quadrada média do erro de previsão ou *Root Mean Square Error Of Prediction* (RMSEP) e o coeficiente de determinação (R^2) foram calculadas com base nas equações 17 e 18:

3.3.2.2 Sensibilidade e Especificidade

Para avaliar os modelos de classificação devem ser considerados dois parâmetros importantes: Sensibilidade e especificidade que são calculadas de acordo com as equações 19 e 20.

A classificação de gasolina por modelos PLS-DA é apresentada na Figura 22. A classificação com 4 variáveis latentes (VLs) foi 100% correta para todas as amostras de gasolina nos conjuntos de treinamento e teste, onde as amostras em vermelho são as adulteradas e as amostras em verde são as não-adulteradas. A sensibilidade (amostras pertencentes à classe e classificadas corretamente nesta classe) e especificidade (amostras não pertencentes à classe modelada e corretamente classificados como não pertencentes) foram iguais a 100% para todos os modelos.

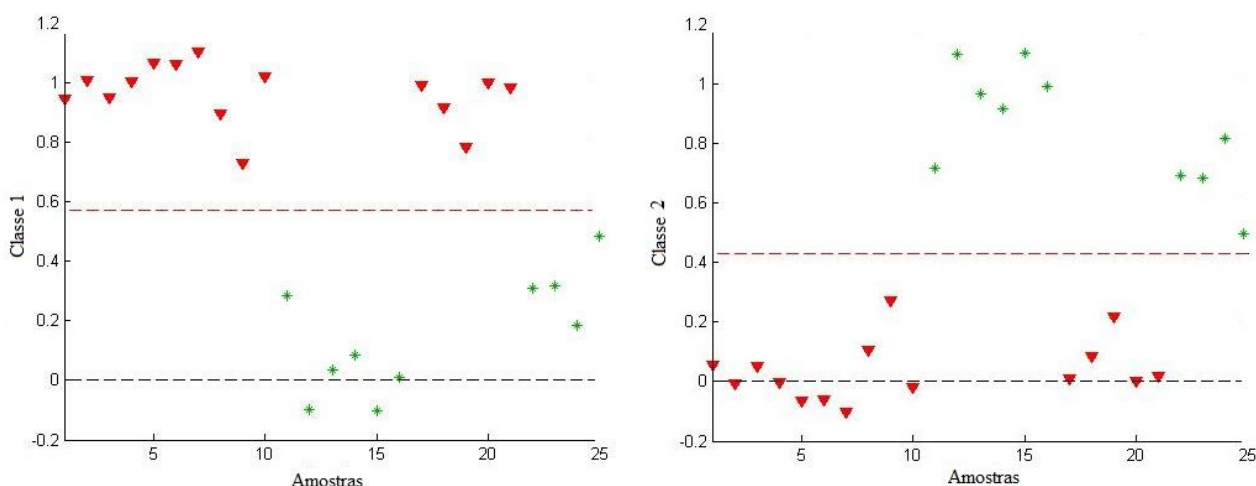


Figura 22. Classificação dos espectros de RMN ^1H das amostras de gasolina por PLS-DA.

A projeção da importância da variância ou *Variance Importance Projection* (VIP), Figura 23, apresenta os sinais mais importantes dos espectros de RMN, que contribuíram para a classificação

das amostras de gasolina por PLS-DA. O sinal A refere-se aos hidrogênios presentes em CH_2 de parafinas, olefinas e aromáticos. O sinal B pode ser relacionado a átomos de hidrogênio de CH de isoparafinas, de CH_2 de naftênicos cíclicos, de Ph-C-CH_2 de aromáticos e de C=C-CH_3 de olefinas. O sinal C corresponde a hidrogênios de CH_2 de etanol (Ferreira e colaboradores 2007). Além disso, a gasolina e óleo diesel têm composições muito complexas, e podem conter até centenas de compostos que variam em termos de comprimento de cadeia, ramificação e de insaturação entre outros fatores, refletindo na formação dos seus componentes, tais como aromáticos, naftênicos e olefinas. Por outro lado, a principal diferença é que o diesel tem cadeias maiores em relação à gasolina e, portanto, tem mais CH_2 , conferindo a ele característica mais apolar. A gasolina, por sua vez, possui cadeias menores com caráter mais polar em comparação ao diesel, misturando-se mais facilmente ao etanol do que ao diesel. Portanto, pode ser sugerido que estas variações estruturais e de composição que incluem diferentes proporções dos componentes, observada nos espectros de RMN, podem contribuir para a classificação das misturas de gasolina adulteradas com óleo diesel e não adulteradas.

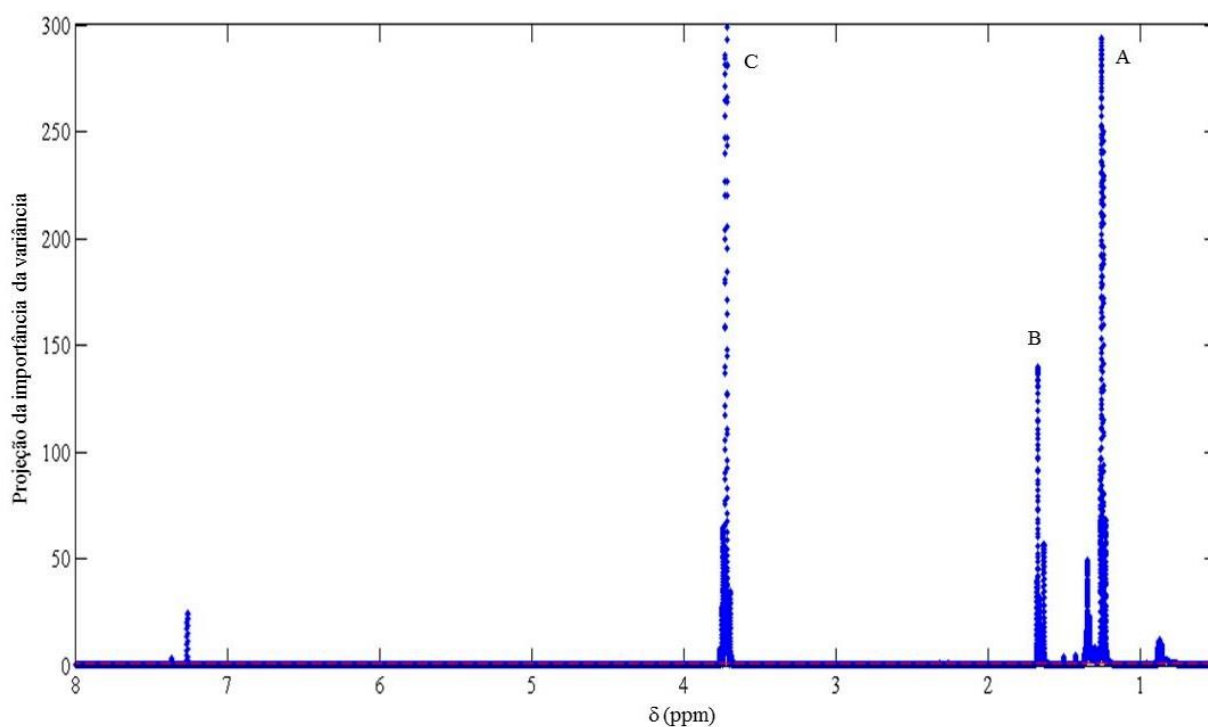


Figura 23. Escores VIP obtidos na análise de PLS-DA do conjunto de espectros de RMN ^1H .

O modelo PLS para predição da quantidade de diesel como adulterante da mistura de gasolina e etanol, variada nas amostras, apresentou excelente resultado, uma vez que o coeficiente de determinação (R^2) foi igual a 0,992 o que significa uma excelente correlação entre os dados de RMN ^1H e a quantidade de diesel adicionada às misturas. Além disso, os valores de RMSECV igual a 2,32%

(v/v) e RMSEP igual a 1,42% (v/v) são baixos e o modelo utilizou apenas 4 variáveis latentes, considerando que a natureza de ambos combustíveis (gasolina e diesel) têm composições químicas bastante complexas. Este resultado foi obtido utilizando-se um número reduzido de amostras preparadas a partir de combustíveis puros. Posteriormente seria interessante aplicar esse mesmo modelo em amostras comerciais para comparar os resultados, considerando-se que neste outro tipo de amostras, normalmente é necessário um número bem maior de amostras durante a modelagem para se representar corretamente a variabilidade das mesmas.

Na Figura 24 observamos o percentual de Diesel adicionado como adulterante da mistura gasolina e etanol predito pelo modelo PLS com dados de RMN ^1H versus a quantidade de diesel adicionadas às misturas.

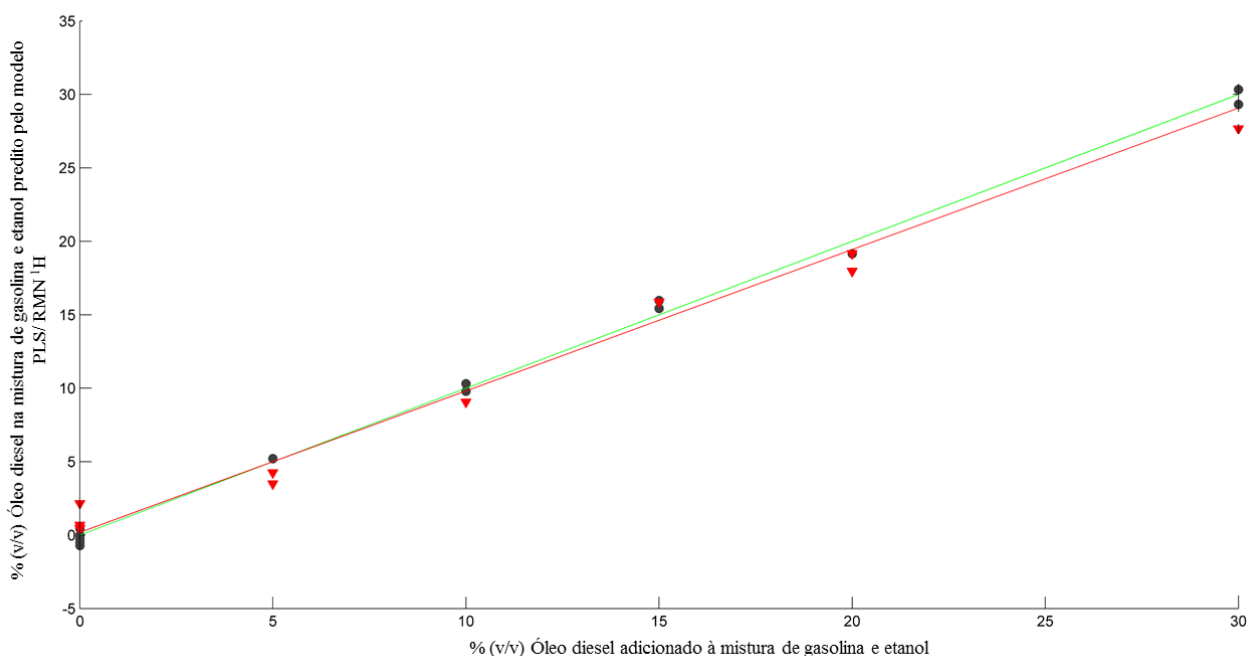


Figura 24. Modelo PLS/RMN ^1H para previsão do percentual de de óleo diesel como adulterante nas misturas de gasolina e etanol.

Como resultado do estudo sobre a adulteração da gasolina empregando dados de RMN ^1H , temos que a HCA e a PCA foram propostas para uma avaliação inicial das amostras com base nos espectros. Estas análises não apresentaram uma tendência de agrupamento de acordo com a adulteração das amostras, com exceção das amostras mais adulteradas. Já os modelos PLS-DA, com base em dados RMN ^1H , apresentaram excelentes resultados em relação à classificação das amostras de gasolina em adulteradas e não adulteradas. No conjunto proposto, a classificação apresentada para ambas as classes apresentou sensibilidade e especificidade igual a 100%. Além disso, a espectroscopia de RMN ^1H associada a regressão multivariada por PLS foi eficiente para

determinação do teor de óleo diesel e, conseqüentemente, para identificar a adulteração da gasolina. Foi possível a quantificação do óleo diesel adicionado como adulterante em misturas de gasolina em concentrações entre 0 a 30% (v /v).

4. AVALIAÇÃO DA ADULTERAÇÃO DE COMBUSTÍVEIS EMPREGANDO FTIR-ATR E MÉTODOS QUIMIOMÉTRICOS

4.1 Objetivo

O capítulo 4 tem o objetivo de avaliar a adulteração dos combustíveis gasolina e etanol empregando diversas ferramentas quimiométricas, desde análise exploratória como HCA e PCA, além de métodos de classificação como PLS-DA e SVM-DA, métodos de regressão multivariada como SVM e PLS e por fim, as CCM aplicadas aos espectros de infravermelho com transformada de Fourier com acessório de refletância total atenuada.

4.2 Materiais e métodos

Para o estudo com aplicação das técnicas quimiométricas aos dados de infravermelho médio com transformada de Fourier com acessório de refletância total atenuada (FTIR-ATR) primeiramente foram preparadas 20 amostras de etanol e 25 amostras de gasolina, conforme apresentado nas Tabelas VIII e IX.

Para tanto, 10 amostras de etanol hidratado foram produzidas com percentual mínimo de 92,5% (v/v) de etanol anidro compondo as amostras conforme exige a norma reguladora vigente da ANP número 19 de 2015 que permite, portanto, um teor máximo de água igual a 7,5% (v/v). As demais amostras deste conjunto compõem as amostras propositalmente adulteradas com teor de água maior que o permitido, variando de 8 até 25 % (v/v), produzidas em duplicata.

O conjunto de amostras de gasolina comum foi produzido com o teor de etanol anidro permitido atualmente, de acordo norma reguladora do MAPA número 75 de 2015, igual a $27\% \pm 1$ (v/v), nas 10 primeiras amostras ditas não adulteradas. As demais amostras foram produzidas, em triplicata, com percentual de óleo diesel, adicionado como adulterante, que variou de 5 até 30 % em volume, substituindo gradativamente o conteúdo de gasolina. Nestas amostras o percentual de etanol foi mantido constante e igual a 27%.

Tabela VIII. Percentual de cada componente das misturas de etanol

Amostras	% m/m água	% m/m etanol
1	0,5	99,5
2	1,0	99
3	1,5	98,5
4	2	98
5	3	97
6	4	96
7	5	95
8	6	94
9	7	93
10	7,5	92,5
11	8	92
12	8	92
13	10	90
14	10	90
15	15	85
16	15	85
17	20	80
18	20	80
19	25	75
20	25	75

Tabela IX. Percentual de cada componente das misturas de gasolina

Amostras	% v/v gasolina	% v/v etanol	% v/v óleo Diesel
1	74	26	0
2	74	26	0
3	73,5	26,5	0
4	73,5	26,5	0
5	73	27	0
6	73	27	0
7	72,5	27,5	0
8	72,5	27,5	0
9	72	28	0
10	72	28	0
11	68	27	5
12	68	27	5
13	68	27	5
14	63	27	10
15	63	27	10
16	63	27	10
17	58	27	15
18	58	27	15
19	58	27	15
20	53	27	20
21	53	27	20
22	53	27	20
23	43	27	30
24	43	27	30
25	43	27	30

4.2.1 Espectroscopia de FTIR-ATR

As amostras utilizadas nesta parte do estudo foram analisadas num espectrômetro Perkin-Elmer modelo Spectrum 400 com acessório de refletância total atenuada horizontal (HATR), utilizando um cristal de seleneto de zinco, com resolução de 4 cm^{-1} , e com 8 e 32 varreduras, na faixa espectral de $4000\text{-}750\text{ cm}^{-1}$ respectivamente, para os conjuntos de amostras de gasolina e de etanol. Para as amostras do conjunto de gasolinas, foi necessário utilizar menos varreduras, pois a mistura elui muito rapidamente sobre o porta amostra, e portanto se for utilizado um número de varreduras maior que 8 e conseqüentemente, um tempo de análise maior, isto ocasiona a evaporação da amostra, prejudicando a análise. Os conjuntos de espectros resultantes para amostras dos dois combustíveis são mostrados na Figura 25.

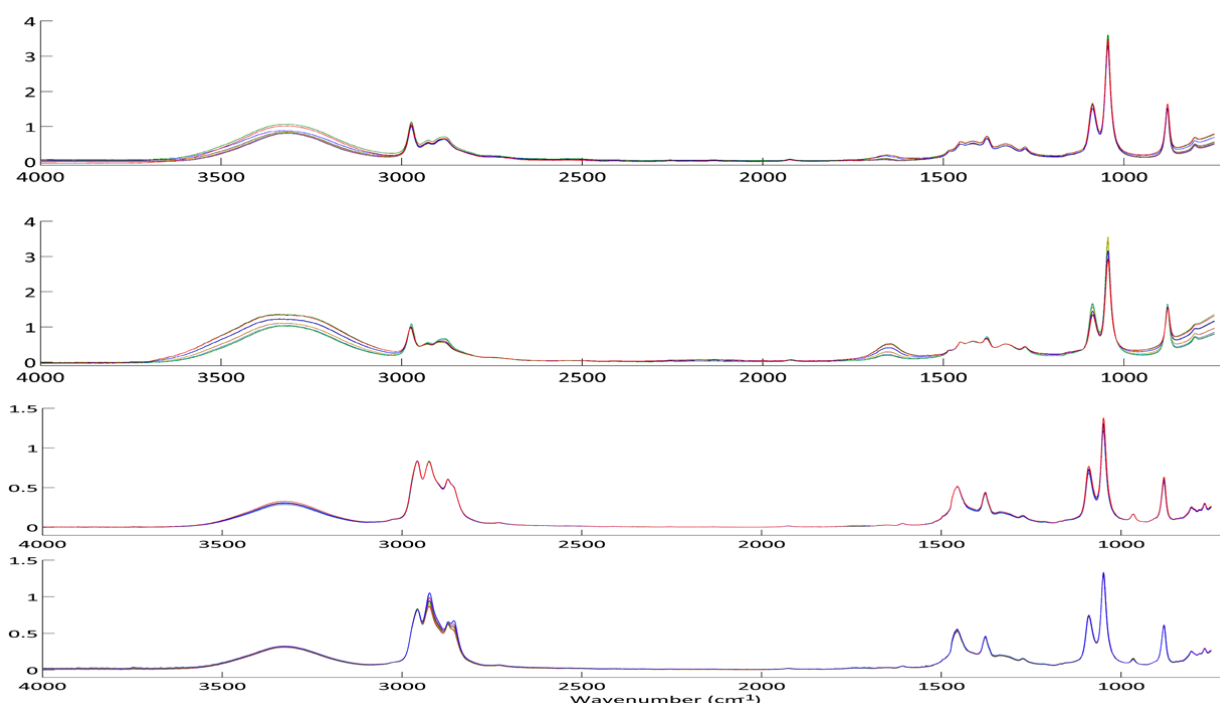


Figura 25. a) espectros de etanol não adulterado (0,5 a 7,5% (m/m) de água), b) espectros de etanol adulterado com 8 a 25% (m/m) de água, c) espectros de misturas de gasolina com 27% \pm 1 (v/v) de etanol anidro não adulteradas, d) espectros de gasolina com 27% \pm 1 (v/v) de etanol anidro adulteradas com 5 a 30% (v/v) de óleo diesel.

Para a análise exploratória (HCA e PCA), análise quantitativa (PLS e SVM), métodos de classificação (PLS-DA e SVM-DA) e para as cartas de controle multivariadas (CCM) de gasolina se utilizou a toda região do espectro ($4000\text{-}750\text{ cm}^{-1}$). Nas misturas de gasolina e etanol adulteradas ou não com óleo diesel, nesta região do espectro, as bandas que visualmente se percebe maior diferença são uma banda de absorção em 3300 cm^{-1} devido ao estiramento da ligação O-H, enquanto que para a gasolina há bandas de maior intensidade na região de $2850\text{-}3000\text{ cm}^{-1}$ referente aos estiramentos

simétricos e assimétricos das ligações C-H de grupos C-H, CH₂ e CH₃ de hidrocarbonetos. Já para o Diesel existem bandas semelhantes as da gasolina porém menos intensas. Isto ocorre porque a gasolina é composta de hidrocarbonetos que são, em geral, mais leves do que aqueles que compõem o Diesel, pois são formados por moléculas de menor cadeia carbônica (4 a 12 átomos de carbono), conseqüentemente possuem menor proporção de CH₂ nas cadeias e maior proporção de CH₃ que o Diesel. No espectro pode-se observar os sinais dos estiramentos simétricos e assimétricos de CH₂ em 2840-2870 cm⁻¹ e 2920-2930 cm⁻¹, respectivamente, além dos sinais dos estiramentos simétrico e assimétricos de CH₃ em 2865-2885 cm⁻¹ e 2950-2975 cm⁻¹ respectivamente. Para o estudo da gasolina, o pré-tratamento realizado na matriz dos espectros de infravermelho, no software Matlab, foi apenas uma suavização pelo algoritmo de Savitzky- Golay utilizando polinômio de primeiro grau com uma janela de 15 pontos, Figura 26.

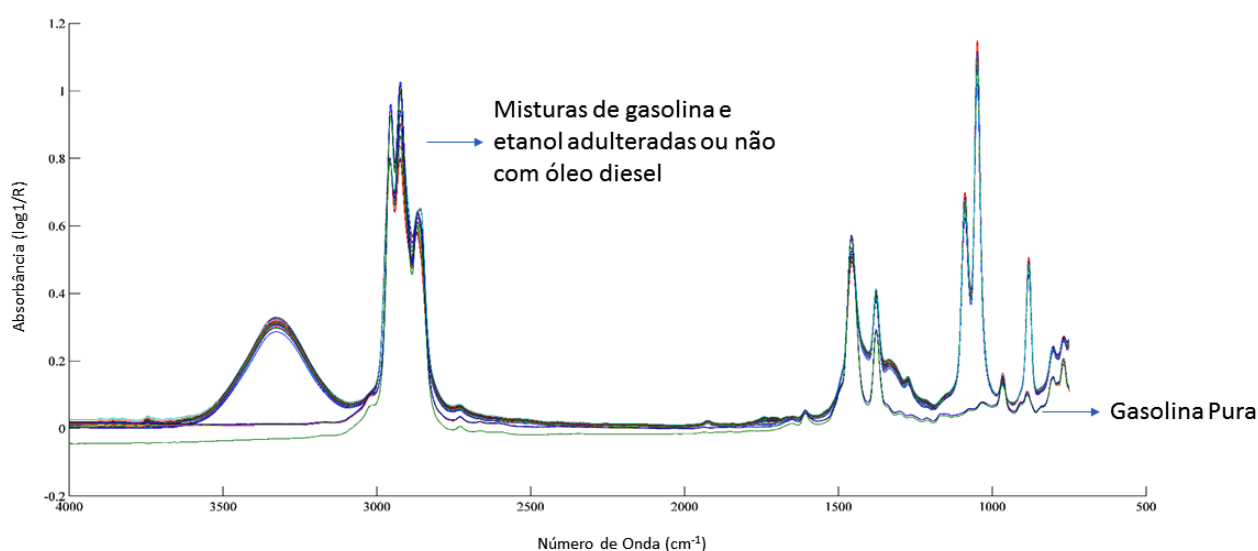


Figura 26. Espectros das misturas de gasolina com pré-tratamento por Savitzky - Golay utilizando polinômio de primeiro grau com uma janela de 15 pontos.

Para as CCM do etanol adulterado ou não com água, o melhor resultado foi obtido usando a região espectral de 3800-2600 cm⁻¹. Nesta região se encontram bandas importantes como o estiramento da ligação O-H presente tanto na molécula da água quanto do etanol na região de 3050-3650 cm⁻¹, estiramentos simétricos e assimétricos de CH₂ em 2840-2870 cm⁻¹ e 2920-2930 cm⁻¹, respectivamente, além dos sinais dos estiramentos simétricos e assimétricos de CH₃ em 2865-2885 cm⁻¹ e 2950-2975 cm⁻¹, respectivamente. Observa-se, porém que, neste caso, os estiramentos simétricos de CH₂ e CH₃ estão sobrepostos. Para a construção destas cartas foi necessário um pré-tratamento dos espectros no software Matlab, no qual os dados foram suavizados, através do algoritmo de Savitsky-Golay utilizando uma janela de 15 pontos e polinômio de segundo grau,

normalizados, tratados com variação normal padrão (SNV) e por fim foram derivados por Savitsky-Golay com uma janela de 15 pontos e polinômio de segundo grau, Figura 27a e b.

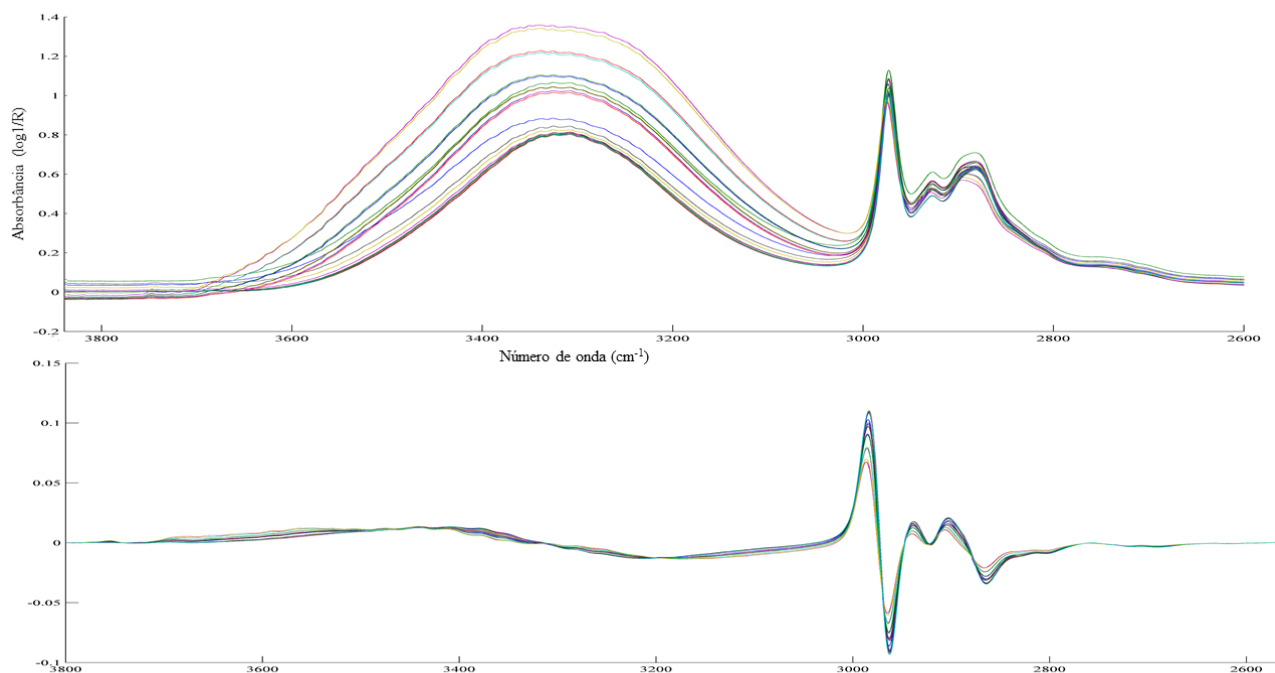


Figura 27. a) Região espectro de infravermelho utilizada para a construção das CCM para as amostras de etanol, b) Mesma região após o pré-processamento: suavização, através do algoritmo de Savitsky-Golay utilizando uma janela de 15 pontos e polinômio de segundo grau, normalização, aplicação de variação normal padrão (SNV) e por fim foram derivada por Savitsky-Golay com uma janela de 15 pontos e polinômio de segundo grau.

4.2.2 Softwares

A construção dos modelos quimiométricos desenvolvidos utilizando dados de FTIR-ATR foram implementados no software MATLAB R2013 (MathWorks, EUA), sendo também empregadas rotinas do PLS-Toolbox 6.2.1 (Eigenvector Research Inc.) e o pacote iToolbox. Além disso, a elaboração das cartas de controle foi realizada em rotina pré-estabelecida em ambiente Matlab (<http://www.models.kvl.dk>, Copenhagen, Dinamarca).

4.2.3 Definição dos conjuntos

Para a aplicação dos métodos de regressão e das cartas de controle, primeiramente ambos os conjuntos de dados foram separados em conjuntos de calibração ou treinamento, contendo 60% das amostras e conjunto de predição ou teste, contendo 40% das amostras, através do algoritmo desenvolvido por Kennard-Stone. Para os modelos de regressão e classificação os dados foram

centrados na média e o número apropriado de variáveis latentes (VLs) foi determinado usando o método de validação cruzada denominado *leave-one-out*. Os valores de RMSECV, RMSEP e R^2 foram calculados conforme as equações 43 e 44.

4.3 Resultados e discussão

4.3.1 Análise Exploratória

A análise de agrupamento hierárquico (HCA) foi o primeiro modelo construído com o objetivo de fazer uma avaliação inicial do conjunto de dados. O dendrograma resultante da HCA foi construído usando o método de Ward sem pré-processamento dos dados de infravermelho para ambos os conjuntos de combustíveis. A análise de componentes principais (PCA) foi o segundo método quimiométrico utilizado para avaliar os dados de gasolina e etanol. Para este método os conjuntos de dados foram centrados na média e foram utilizadas as 3 primeiras componentes principais (PCs) que resultaram em uma variância acumulada de 99,86% para o conjunto do etanol e de 99,56% para o conjunto da gasolina. A principal diferença entre HCA e PCA é que HCA usa toda a informação proveniente do espectro, enquanto a informação utilizada na PCA é composta, como o próprio nome informa, por componentes principais que contém a maior parte das informações relevantes contidas nos dados e pode excluir aquela informação dita aleatória contida na matriz de erros.

Com base no estudo inicial obtido com a aplicação de HCA, Figura 28a, pode-se ver que estão formando dois grupos principais, um menor que contém as amostras não adulteradas e outro superior com amostras adulteradas. Todas as amostras foram agrupadas corretamente, em ordem crescente de quantidade de adulterante, embora as amostras 9 e 10 tenham sido agrupadas às amostras adulteradas, provavelmente pela composição muito próxima as amostras 11 e 12. Quanto às amostras de gasolina na Figura 28b, a HCA mostrou uma tendência correta de agrupamento, em ordem crescente de quantidade de diesel adicionado como adulterante. Houve a formação de um subgrupo com as amostras não adulteradas, outro intermediário com as amostras menos adulteradas e outro superior, com as amostras com maior percentual de adulterante (em verde), que une corretamente inclusive as amostras em triplicata.

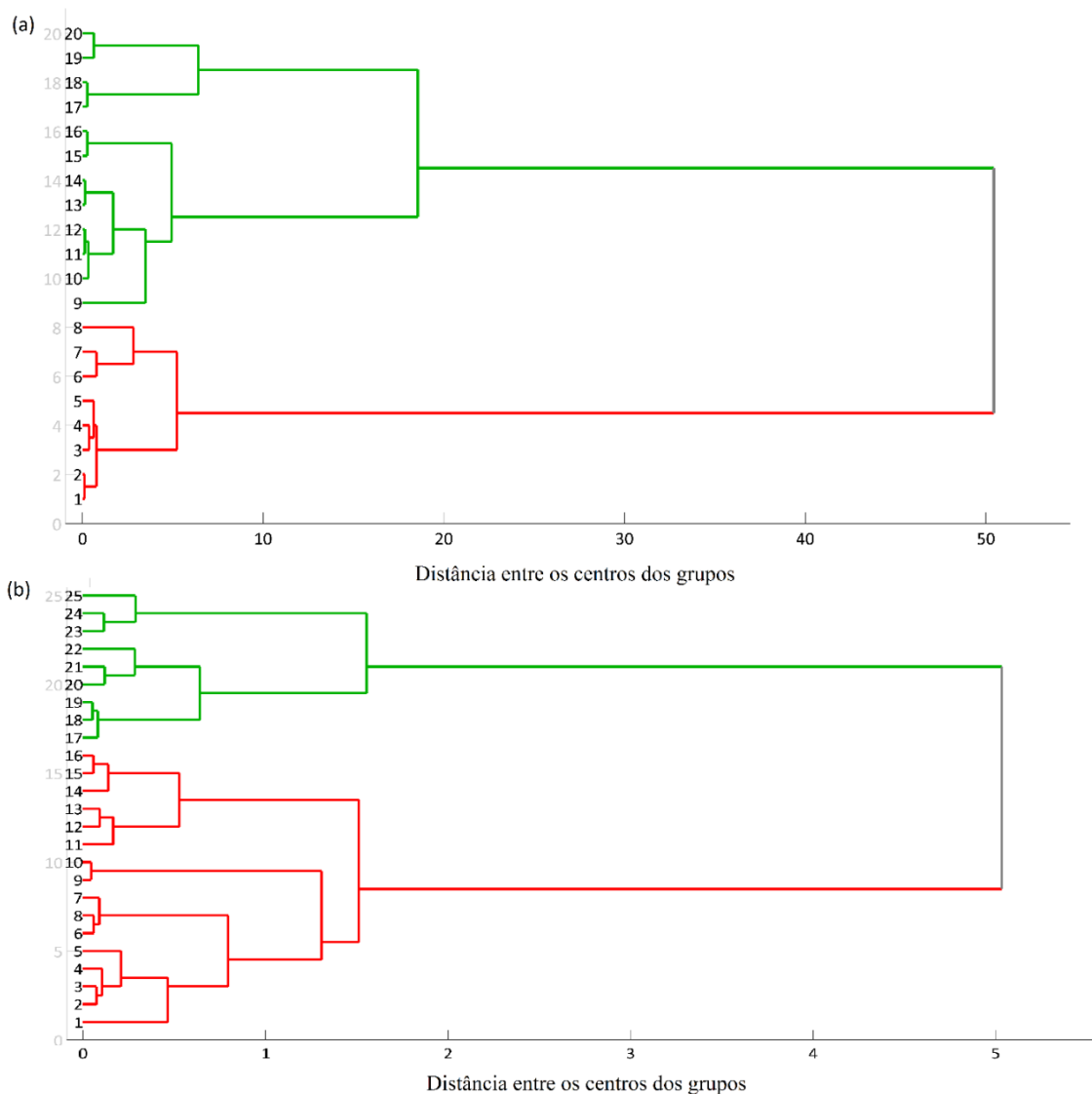


Figura 28. Análise de agrupamento hierárquico dos espectros de FTIR-ATR a) das misturas de etanol, b) das misturas de gasolina.

Observando a Figura 29, que foi plotada em 3 dimensões para melhor visualização, observando apenas a primeira componente principal, nota-se que, para ambos conjuntos, há uma clara tendência de separação entre as amostras, onde as amostras não adulteradas estão ao longo do eixo negativo das PC1 juntamente com algumas amostras com pouca quantidade de adulterante, enquanto as amostras com percentual mais alto dos adulterantes estão dispostas ao longo do eixo positivo da PC1.

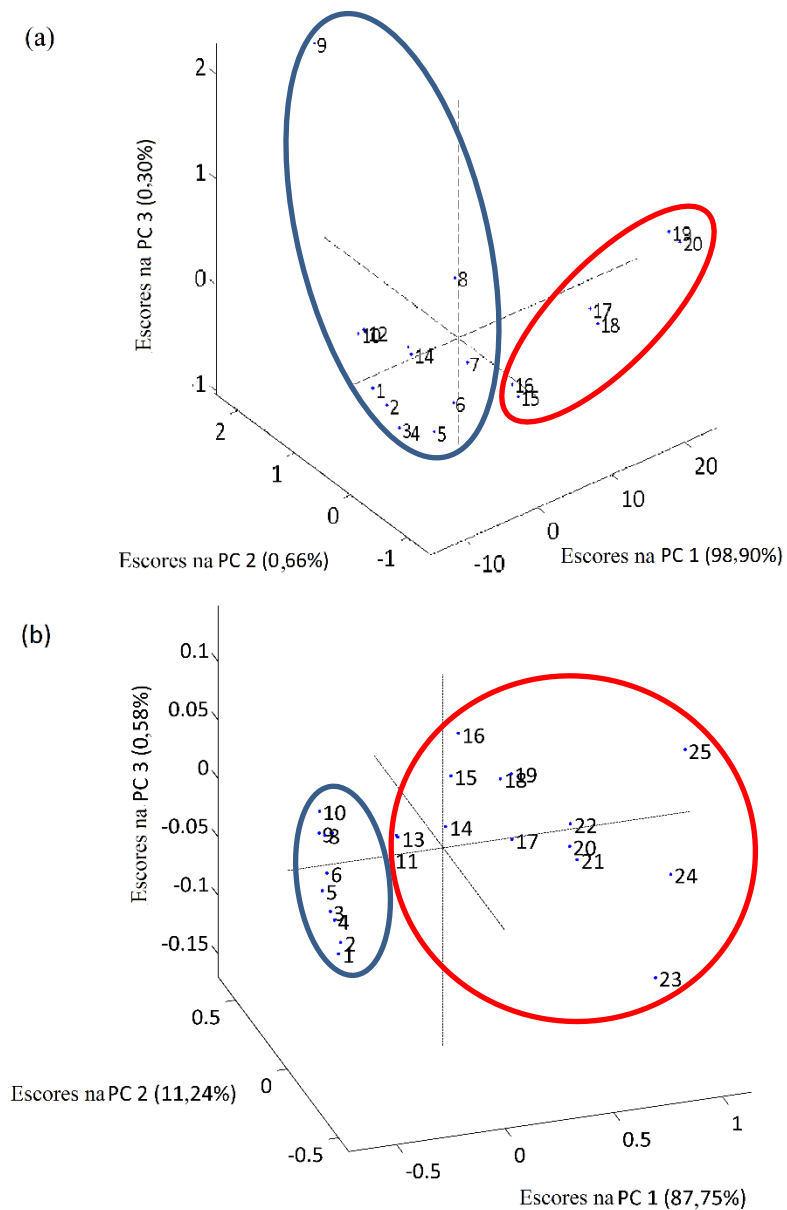


Figura 29. Análise de por componentes principais dos espectros de FTIR-ATR a) das misturas de etanol, b) das misturas de gasolina.

4.3.2 Métodos de Classificação

Os modelos PLS-DA e SVM-DA com caráter qualitativo foram construídos para classificar as amostras de etanol e gasolina em adulteradas e não adulteradas.

A classificação de amostras de etanol e gasolinas por modelos PLS-DA é mostrada nas Figuras 30 e 31.

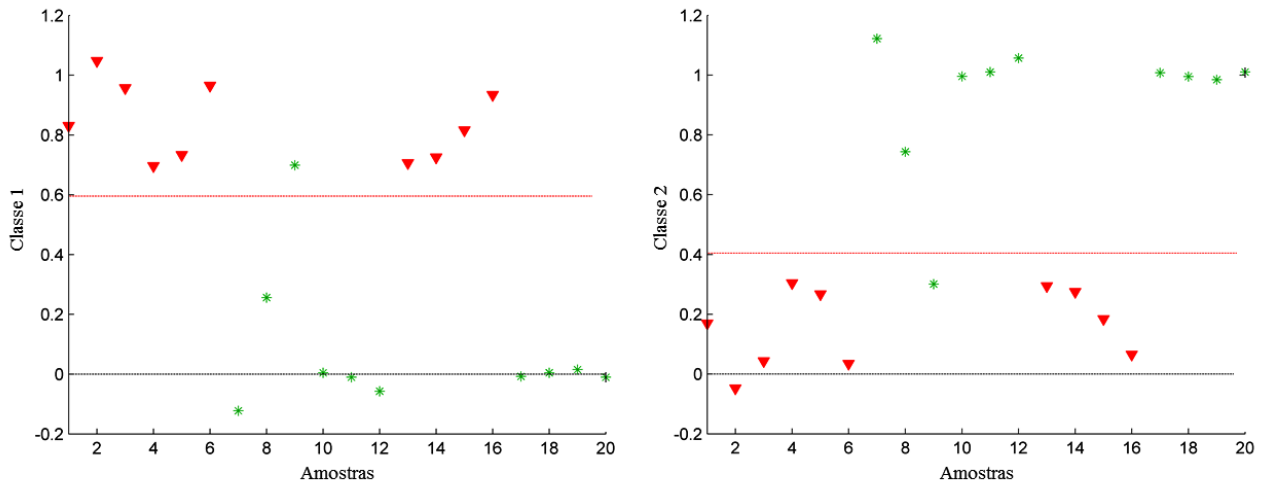


Figura 30. Classificação por PLS-DA dos espectros de FTIR-ATR do conjunto de amostras de etanol.

Os asteriscos verdes, na Figura 30 referem-se ao etanol não adulterado e os triângulos vermelhos referem-se a amostras de etanol adulterado. A linha pontilhada vermelha refere-se ao limiar que separa ambas as classes. A linha pontilhada preta refere-se ao valor zero. A classificação das amostras de etanol utilizou 2 variáveis latentes (VLs) e apresentou, conforme mostrado na Tabela X, sensibilidade igual a 100% e especificidade de 90% para a primeira classe e sensibilidade igual a 90% e especificidade igual a 100% para a segunda classe, considerando sensibilidade igual as amostras pertencentes à classe e classificadas corretamente nesta classe e especificidade as amostras não pertencentes à classe modelada e corretamente classificadas como não pertencentes. Desta forma ocorreu uma vez um falso positivo na classe um e um falso negativo na classe dois.

A classificação das amostras de gasolina utilizando 2 VLs apresentou excelentes resultados, conforme mostrado na Tabela X, uma vez que a sensibilidade e especificidade foram iguais a 100% para ambas as classes, como pode ser visto na Figura 31, ou seja, todas as amostras foram corretamente classificadas pelo modelo proposto.

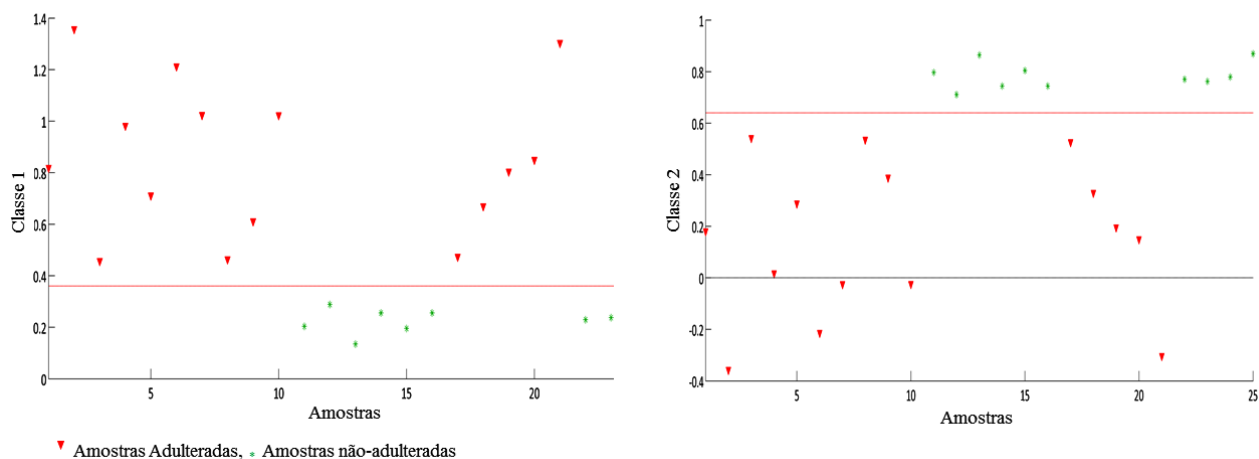


Figura 31. Classificação por PLS-DA dos espectros de FTIR-ATR do conjunto de amostras de gasolina.

Tabela X. Parâmetros dos modelos PLS-DA e SVM-DA para misturas de etanol e gasolina

	Etanol		Gasolina	
	Sensibilidade	Especificidade	Sensibilidade	Especificidade
*Classe 1	100 %	90 %	100 %	100 %
*Classe 2	90 %	100 %	100 %	100 %
**Classe 1	80 %	100 %	100 %	100 %
**Classe 2	100 %	80 %	100 %	100 %

*PLS-DA, **SVM-DA

Os resultados dos modelos SVM-DA construídos para fins de comparação com os resultados do PLS-DA para os dados de etanol e gasolina são mostrados nas Figuras 32 e 33.

Como pode ser visto na Figura 33, o modelo SVM-DA proposto obteve um resultado ligeiramente inferior ao PLS-DA, relacionado aos parâmetros de sensibilidade e especificidade, calculados pelas equações 44 e 45, uma vez que PLS-DA apresentou sensibilidade e especificidade iguais a 100% e 90% para a classe 1 e 90% e 100% para a classe 2 para o conjunto de etanol, enquanto que para a SVM-DA esses parâmetros foram iguais a 100% e 90% para a classe 1 e 90% e 100% para a classe 2, como pode ser observado na Tabela X.

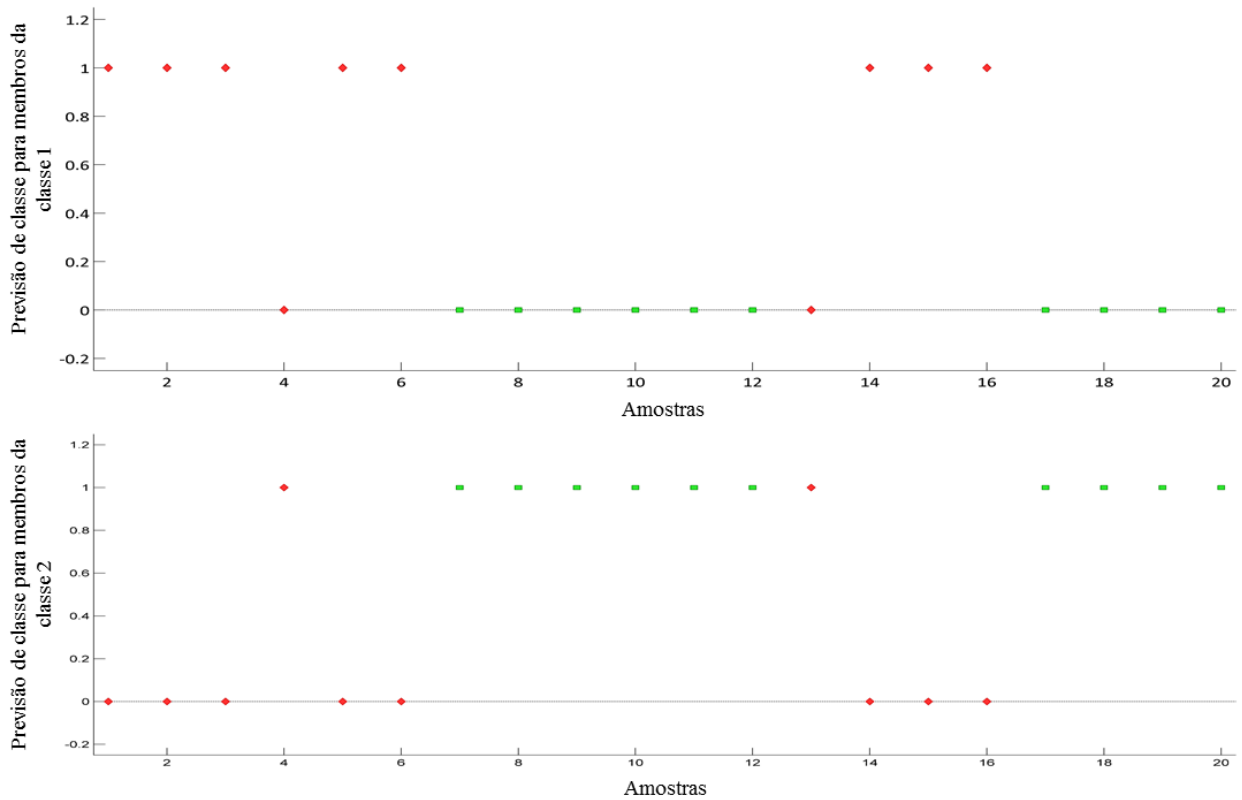


Figura 32. Classificação por SVM-DA dos espectros de FTIR-ATR do conjunto de amostras de etanol.

A mesma comparação pode ser feita para o conjunto de gasolina, observando a Figura 33. Como pode ser visto, o modelo SVM-DA proposto sugere um resultado equivalente ao PLS-DA, relacionado aos parâmetros de sensibilidade e especificidade, uma vez que ambos apresentaram sensibilidade e especificidade iguais a 100% para ambas as classes, como pode se concluir observando a figura e verificando-se que todas as amostras foram corretamente classificadas.

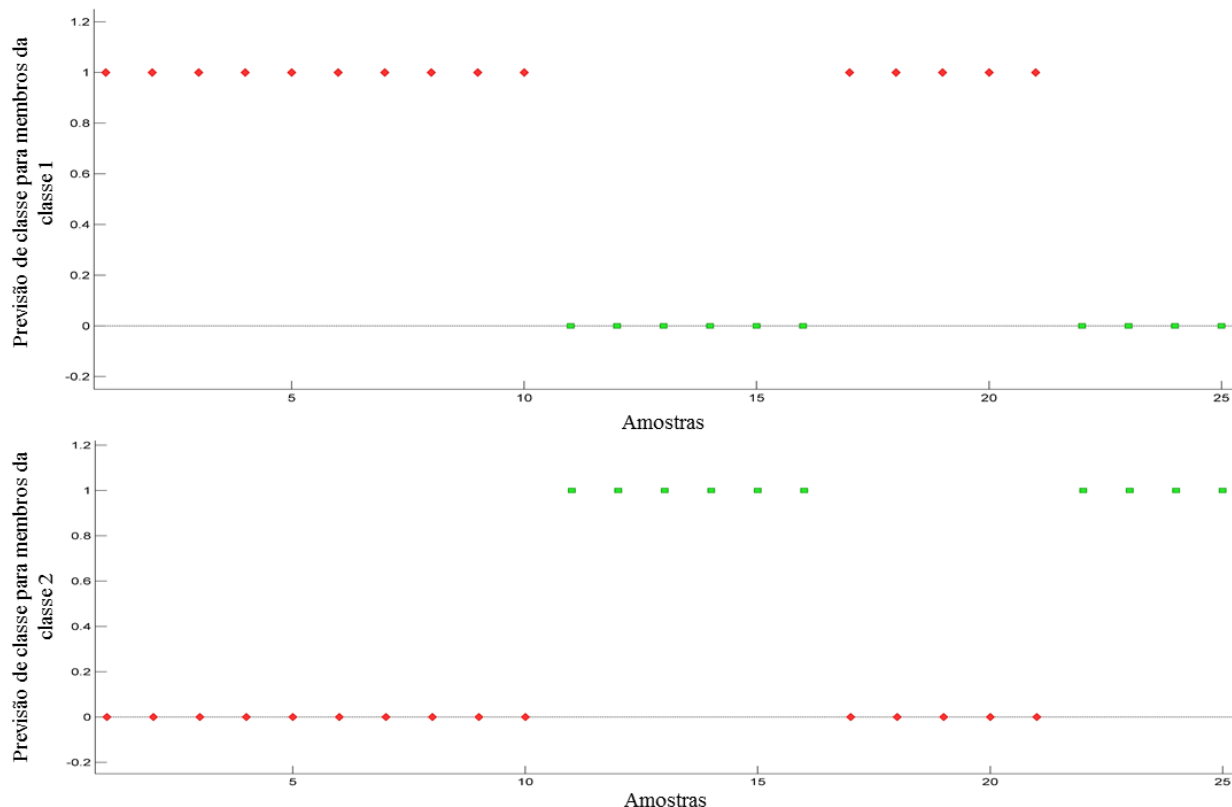


Figura 33. Classificação por SVM-DA dos espectros de FTIR-ATR do conjunto de amostras de gasolina.

4.3.3 Métodos de Regressão Multivariada

Na tabela XI são mostrados os parâmetros do modelo PLS e SVM para misturas de etanol e gasolina.

Tabela XI. Parâmetros do modelo PLS e SVM para misturas de etanol e gasolina

	VLs	R ²	RMSECV(%)***	RMSEP(%)***
*Etanol	2	0,985	1,721	0,859
*Gasolina	3	0,998	0,832	0,543
**Etanol		0,999	1,056	0,379
**Gasolina		0,997	1,084	0,571

*PLS, **SVM, ***% (m/m) para etanol e % (v/v) para gasolina, VLs = variáveis latentes

Ambos os modelos PLS apresentaram excelentes resultados, já que o primeiro utiliza apenas duas variáveis latentes com coeficiente de determinação (R²) igual a 0,985 e baixos valores de RMSECV e RMSEP para o modelo onde y é igual à porcentagem de água no etanol, como pode ser visto na Figura 34a. O modelo PLS para avaliar a quantidade de diesel como adulterante na gasolina

apresentou um valor de R^2 igual a 0,998. Este resultado mostra uma excelente correlação entre as quantidades de adulterantes colocadas nas misturas e as mesmas quantidades previstas pelo método espectroscópico FTIR-ATR juntamente com o método PLS, Figura 34. Além disso, o erro médio quadrático de validação cruzada e o erro médio quadrático de predição são baixos e ambos usaram um número baixo de variáveis latentes que correspondem ao mesmo número de componentes de misturas.

Foram construídos também modelos de regressão pelo método SVM conforme apresentado na Figura 35. Nestes modelos o pré-processamento dos dados foi centrar na média e para a validação cruzada foi utilizado o método *leave-one-out*. Para o etanol, os parâmetros de custo, epsilon e gama foram iguais a 100, 0,01 e 0,031623, respectivamente; já para o conjunto de gasolina foram iguais a 100, 0,01 e 0,01, respectivamente.

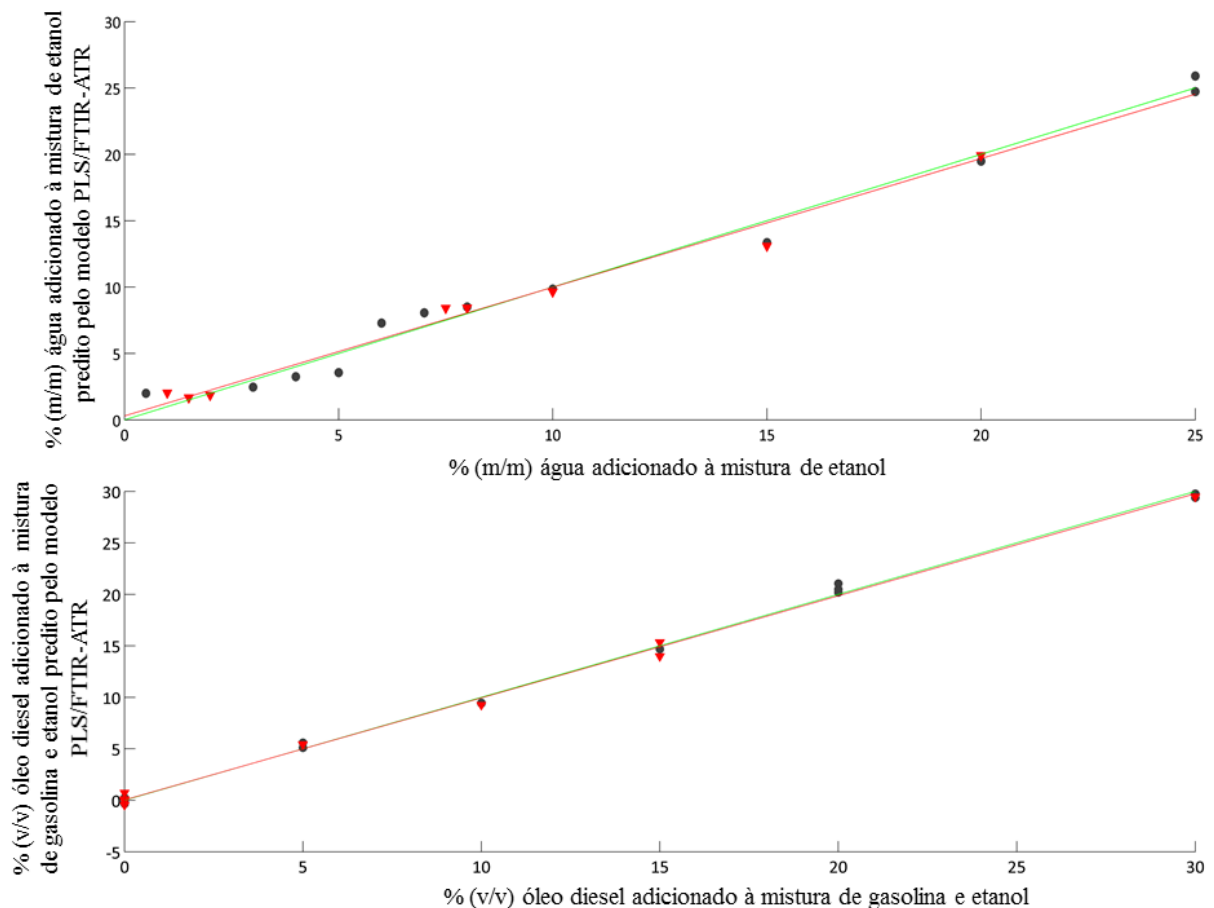


Figura 34. Quantificação dos adulterantes: a) água em etanol hidratado combustível, b) óleo diesel na mistura de gasolina e etanol pelo modelo PLS utilizando dados de FTIR-ATR.

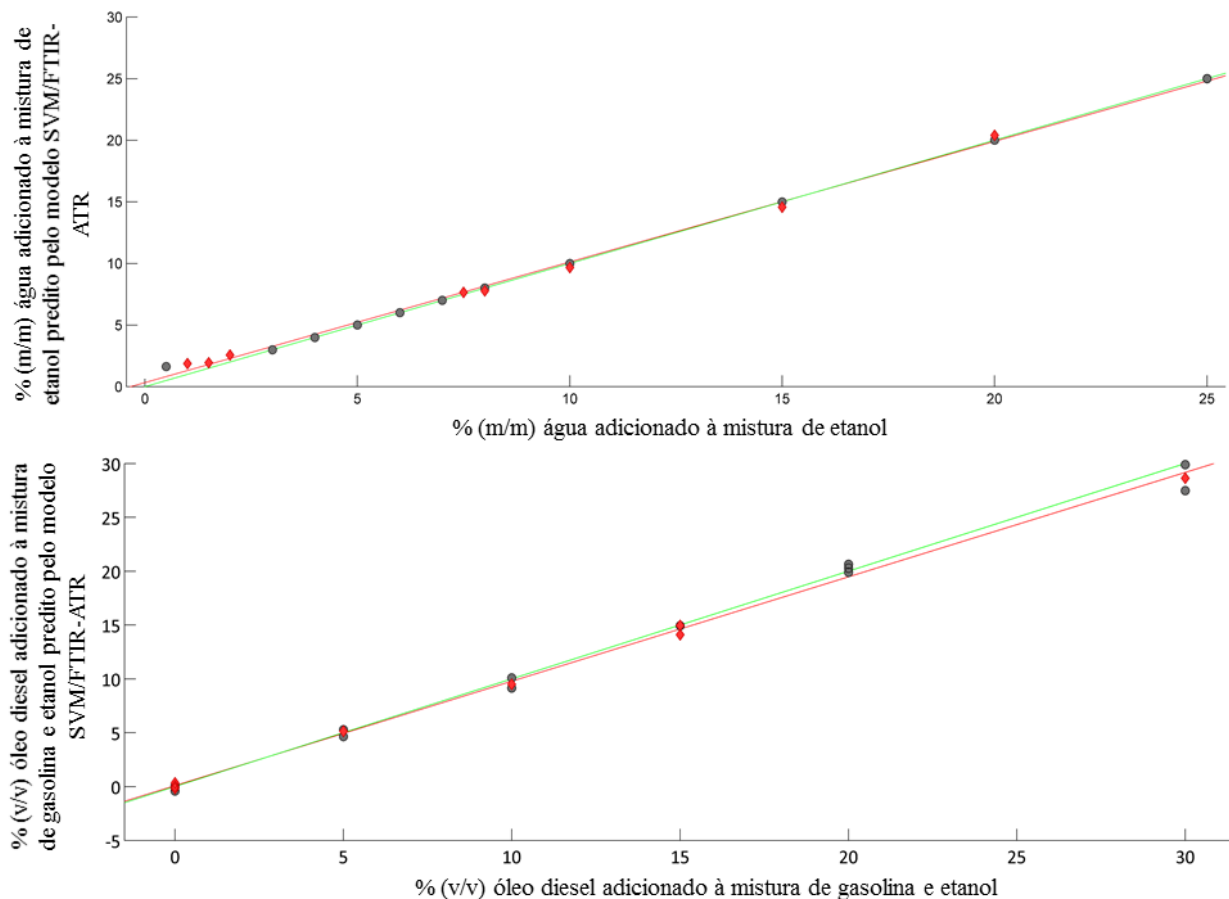


Figura 35. Quantificação dos adulterantes: a) água em etanol hidratado combustível, b) óleo diesel na mistura de gasolina e etanol pelo modelo SVM utilizando dados de FTIR-ATR.

Para o conjunto de amostras de etanol, os valores de RMSECV, RMSEP e R^2 obtidos pelo modelo foram iguais a 1,319%, 0,474% e 0,996, respectivamente. Estes resultados indicam que o modelo SVM é uma excelente opção para representar este tipo de dados. A correlação obtida entre a quantidade de água adicionada em misturas de etanol como adulterante e a mesma quantidade prevista a partir de dados FTIR-ATR foi excelente. Já para o conjunto de gasolina os valores de RMSECV, RMSEP e R^2 foram iguais a 1,084%, 0,571% e 0,997, respectivamente. O modelo SVM mostrou-se também uma excelente opção para as misturas de gasolina, uma vez que foi obtida através do SVM uma excelente correlação entre a quantidade de óleo diesel adicionado em misturas de gasolina como adulterante e esta mesma quantidade de combustível prevista a partir de dados FTIR-ATR.

Comparando-se os resultados obtidos por ambos modelos PLS e SVM podemos observar que para o etanol, o modelo SVM apresentou melhores resultados em comparação com o modelo PLS. O coeficiente de determinação (R^2) obtido no modelo SVM é maior do que no PLS, enquanto que o erro quadrático médio de validação cruzada (RMSECV) e o erro quadrático médio de predição (RMSEP) foram muito menores para o SVM do que para o PLS. Comparando os dados para o conjunto de

amostras de gasolina, os resultados obtidos foram muito semelhantes, uma vez que o modelo PLS apresentou valores muito próximos de R^2 e do erro de predição, enquanto o erro quadrático médio de validação cruzada (RMSECV) foi ligeiramente menor que o modelo SVM .

4.3.4 Cartas de Controle Multivariadas

Os resultados das cartas de controle multivariados (MCC) são mostrados nas Figuras 36 e 37 para conjuntos de etanol e gasolina, respectivamente, onde as amostras em azul pertencem ao conjunto de calibração e as amostras em vermelho ao conjunto de validação.

A carta NAS mostra o sinal analítico líquido, ou seja, o sinal do analito, que para este conjunto de amostras é a água. Como pode ser visto na carta NAS, os limites da carta foram definidos por sete amostras do conjunto de calibração (em azul). Este conjunto consiste de dois terços das amostras não adulteradas que foram selecionadas pelo algoritmo Kennard-Stone e o restante (três amostras) foram unidas às amostras adulteradas para formar o conjunto de validação. Como pode ser visto na Figura 36, as amostras do conjunto de validação (em vermelho) mostraram o seguinte comportamento: Três amostras não adulteradas utilizadas para validação, isto é, contendo menos de 7,5% de água estão dentro dos limites da carta NAS , como era esperado; mas duas amostras contendo 8% de água não ficaram acima do limite superior, como esperado. Para as outras amostras contendo 10%, 15%, 20% e 25% de água na mistura de álcool obteve-se um bom resultado, uma vez que estão acima do limite superior estipulado pela carta NAS, o que era esperado, uma vez que todos contêm mais água do que é permitido pela norma que regula o teor de água em álcool etílico hidratado.

A carta do interferente representa a matriz, que por sua vez, mostra a quantidade de etanol nas amostras, tanto no conjunto de calibração (em azul) quanto no conjunto de validação (em vermelho). Esta carta mostra que as sete amostras do conjunto de calibração e as três amostras de teste do conjunto de validação contendo a quantidade mínima exigida pelo norma da ANP, que é igual a 92,5% de etanol em sua composição estão dentro do limite esperado. As amostras contendo menos de 92,5% de etanol estão fora do limite da carta do interferente. Isto mostra que quanto menor a quantidade de etanol presente nas amostras, mais fora do limite estipulado estão as amostras. Esse comportamento está de acordo com o planejamento experimental, já que o percentual de etanol foi gradualmente substituído por água e é menor que o mínimo estipulado para as amostras de calibração (92,5% em volume de etanol).

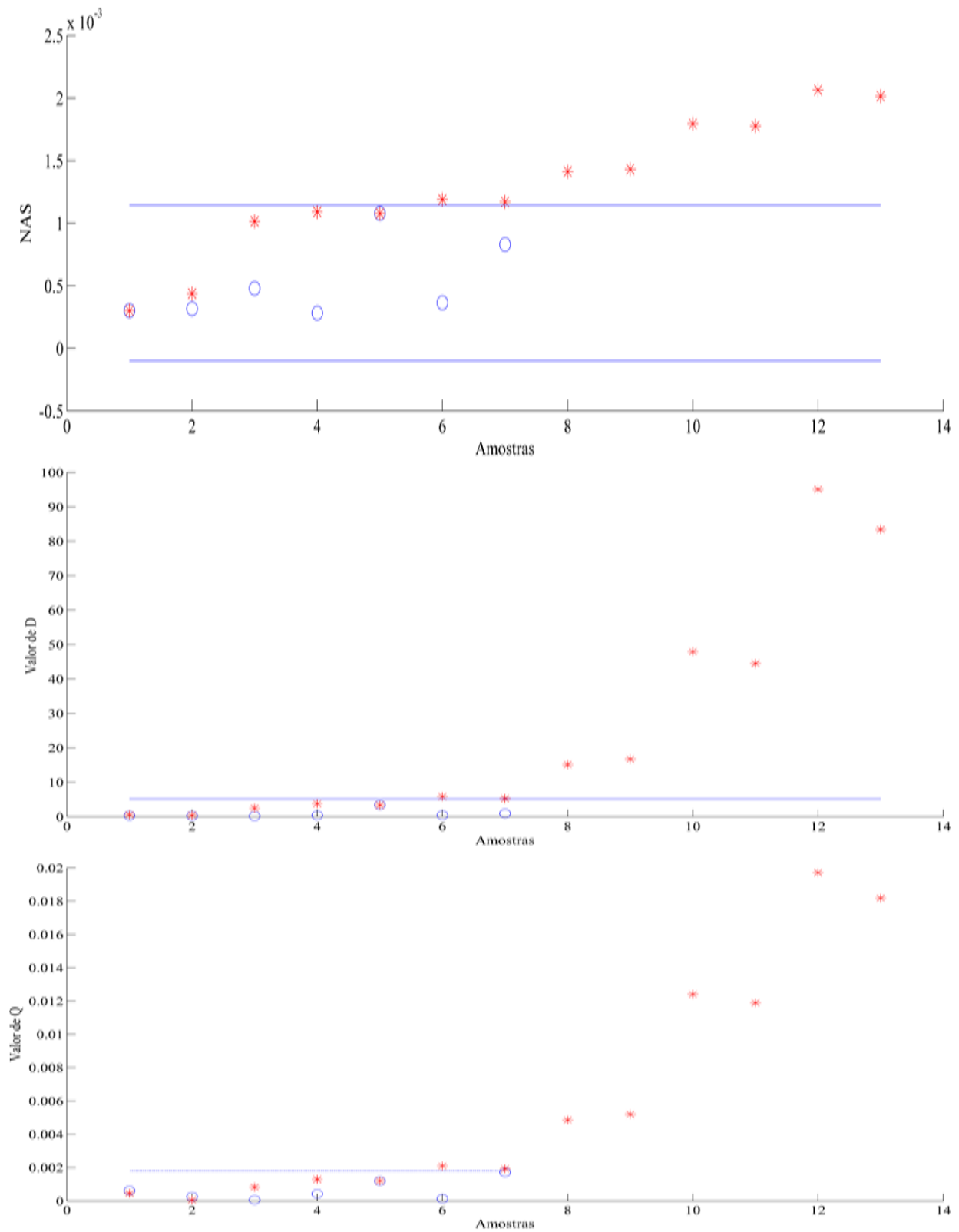


Figura 36. Cartas de controle multivariadas baseadas em dados de FTIR-ATR para determinação de adulteração de misturas de etanol. A primeira carta é do sinal de analítico líquido (NAS) (água), a carta intermediária é do interferente ou matriz da amostra (etanol) e a última carta é dos resíduos.

Dessa forma, quanto mais o teor de etanol foi reduzido nas amostras de validação (de 92% para 75% em volume), maior a distância dessas amostras das amostras de conjunto de calibração. Pode-se também observar que as amostras em duplicata estão próximas, o que mostra uma boa reprodutibilidade da preparação das amostras e obtenção dos espectros.

A carta dos resíduos está em concordância com os resultados apresentados na carta NAS e do interferente, mostrando que o teor de água está dentro do limite máximo de 7,5% em massa nas amostras de calibração e acima desse limite nas amostras de validação, com exceção das duas amostras contendo 8% de água, que neste gráfico de resíduos também estavam abaixo do limite.

Na Figura 37, a carta NAS mostra o sinal líquido do analito, que para este conjunto é o etanol. Como pode ser visto, os limites da carta foram definidos por sete amostras do conjunto de calibração (em azul). Este conjunto consiste de dois terços das amostras não adulteradas que foram selecionadas pelo algoritmo Kennard-Stone e o restante (três amostras) foram unidas às amostras adulteradas para formar o conjunto de validação. De acordo com a Figura 38, todas as amostras do conjunto de calibração e validação estão dentro dos limites da carta NAS, como esperado, uma vez que todas elas têm $27\% \pm 1$ (% v/v) de etanol.

A carta do interferente, que para este conjunto de amostras se refere à matriz (gasolina), mostra que as amostras do conjunto de calibração (em azul) e as três amostras de teste usadas para validação, estão dentro do limite. Além disso, a carta mostra claramente que as outras amostras do conjunto de teste estão fora do limite estipulado, o que está de acordo com o planejamento, uma vez que a porcentagem de gasolina foi gradualmente substituída por Diesel para simular adulteração e, portanto, contém menos que o mínimo estipulado para as amostras de calibração (72% em volume de gasolina). Assim, quanto mais o teor de gasolina foi reduzido nas amostras de validação (68-43% em volume), maior a distância destas amostras das amostras do conjunto de calibração. Pode-se também observar que as amostras em triplicata estão próximas, o que mostra uma boa reprodutibilidade da preparação das amostras e obtenção dos espectros.

A carta dos resíduos, por sua vez, refere-se ao componente da mistura que não foi modelado pela carta NAS nem pela carta do interferente. Assim, pode ser visto que há outro componente na mistura além de gasolina e etanol. Portanto, no caso deste conjunto, composto de gasolina, etanol e Diesel, a carta dos resíduos representa o Diesel adicionado à mistura como adulterante. Nota-se que nas amostras de calibração (em azul) e nas três primeiras amostras do conjunto de validação (em vermelho) que estão dentro do limite da carta dos resíduos, não há, portanto, nenhum adulterante presente. A partir da quarta amostra do conjunto de validação, há presença e aumento do teor de

Diesel, nas amostras 11 a 25 da Tabela IX, (em triplicata) e que pertencem ao conjunto teste, o que está de acordo com o planejamento.

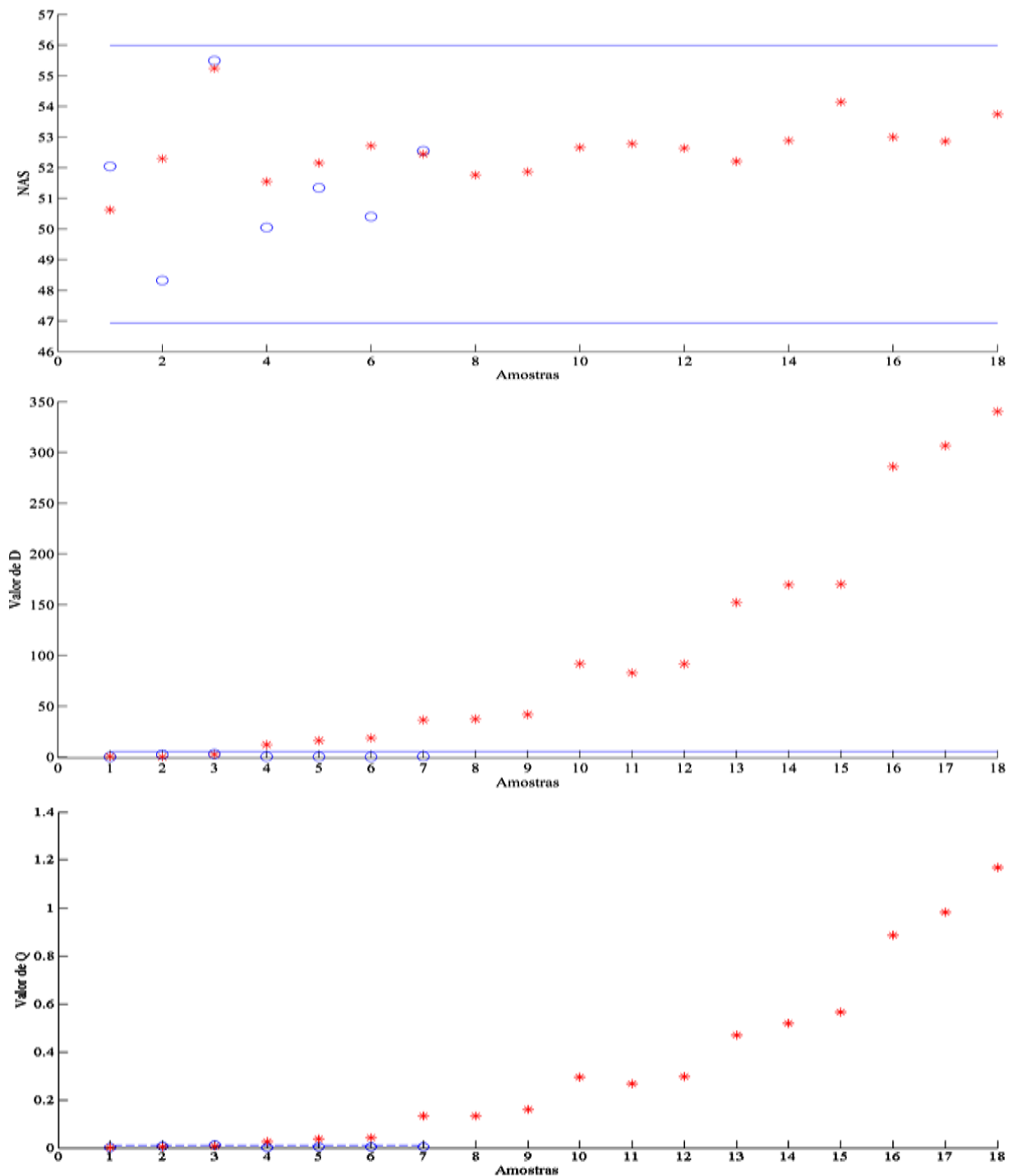


Figura 37. Cartas de controle multivariadas baseadas em dados de FTIR-ATR para determinação de adulteração de misturas de gasolina e etanol. A primeira carta é do sinal de analítico líquido (NAS) (etanol), a carta intermediária é do interferente ou matriz da amostra (gasolina) e a última carta é dos resíduos que é o adulterante (óleo diesel).

Como pode ser visto, ambos os conjuntos das cartas de controle multivariadas aplicadas aos espectros de infravermelho se mostraram como ferramentas quimiométricas úteis para avaliar as amostras de gasolina e etanol propostos no estudo, uma vez que permitiram separar os sinais dos analitos (gasolina e etanol), sinais dos interferentes (água e etanol) e dos resíduos (diesel) de cada conjunto estudado.

5. CONCLUSÕES

A aplicação do planejamento experimental Doehlert no estudo da reação de transesterificação do óleo de soja com álcool metílico, através da metodologia *Transesterification Double Step Process* (TDSP) para a produção de biodiesel, atingiu o percentual de conversão máximo, nas reações em que se estudou a etapa da catálise básica, utilizando-se massa de catalisador igual a 0,40 g e volume de metanol de 60 mL. Realizando-se a mesma reação com quantidades estequiométricas dos reagentes segundo a proposta original da metodologia TDSP, é necessário utilizar 0,50 g de catalisador e 40 mL de metanol. Já para o estudo da etapa da catálise ácida, os resultados obtidos aplicando planejamento Doehlert são equivalentes aqueles alcançados apenas com a metodologia TDSP, ou seja, temperatura igual a 65 °C e tempo de reação igual a 90 minutos.

Na outra etapa do estudo que fez uso de métodos quimiométricos com dados de RMN, a HCA e a PCA foram propostas para avaliar as amostras de gasolina, e embora tenham apresentado uma tendência de agrupamento apenas para as amostras mais adulteradas, indicaram que os dados de RMN poderiam ser utilizados na diferenciação de amostras adulteradas. Já os modelos PLS-DA apresentaram excelentes resultados em relação à classificação de gasolinas adulteradas e não adulteradas, apresentando sensibilidade e especificidade igual a 100%. Além disso, a mesma técnica espectroscópica associada a regressão multivariada por PLS foi eficiente para quantificar óleo diesel em gasolina adulterada.

Finalmente, foram aplicados dois modelos de análise exploratória (HCA e PCA), dois de classificação (PLS-DA e SVM-DA), dois de regressão multivariada (PLS e SVM) e cartas de controle multivariadas a espectros de infravermelho médio. A HCA se mostrou uma ferramenta rápida e simples para uma avaliação inicial de ambos os conjuntos de combustível estudados; no entanto não foi possível diferenciar amostras muito próximas do limite de adulteração. O modelo de PCA construído para o conjunto de etanol apresentou uma tendência de separação de amostras não adulteradas e adulteradas, enquanto que para o conjunto de gasolina foi observada uma clara separação entre amostras de gasolina não adulteradas e adulteradas. Já os modelos PLS-DA mostraram bons resultados para classificar as amostras de combustível estudadas neste trabalho em adulterado e não adulterado. Para os conjuntos propostos, obteve-se uma boa classificação com presença de apenas um falso positivo e um falso negativo para etanol e uma excelente classificação para gasolina com sensibilidade e especificidade igual a 100% uma vez que não apresentou falsos positivos ou falsos negativos. Além disso, através de regressão multivariada pelo método PLS, foi possível uma quantificação eficiente dos adulterantes, a determinação simultânea de quantidades de água e óleo diesel utilizados como adulterantes em misturas de etanol e gasolina, respectivamente,

em concentrações variando entre 8 a 25% de água (m/m) e 0 a 30% de óleo diesel (v/v). Os modelos SVM-DA, por sua vez, apresentaram resultados satisfatórios, principalmente para o conjunto de gasolina onde todas as amostras foram corretamente classificadas. Os modelos SVM para ambos os conjuntos de combustível apresentaram excelentes valores de coeficiente de determinação, erro quadrático médio de validação cruzada e erro quadrático médio de predição mostrando a aplicabilidade deste método para quantificar adulteração nessas misturas. O último método utilizado, as CCM também se mostraram bastante adequadas, uma vez que as cartas detectaram uma percentagem de adulteração acima de 10% em massa de água colocada na mistura com álcool, e de 5% a 30% em volume de óleo diesel em misturas de gasolina e portanto, provavelmente podem detectar quantidades maiores de adulterantes também que são adicionados, como uma prática comum, com o objetivo de obter lucro indevido.

Este trabalho permitiu concluir que se pode utilizar tanto ressonância magnética nuclear de hidrogênio quanto infravermelho médio para identificar adulteração em combustíveis automotivos, juntamente com alguns métodos quimiométricos. Para trabalhos futuros seria interessante aplicar estes mesmos métodos em amostras comerciais, para fins de comparação com os resultados obtidos neste estudo.

6. REFERÊNCIAS

1. Anuário Estatístico Brasileiro Do Petróleo, Gás Natural E Biocombustíveis ANP 2016.
2. Dias, J.A.; Machado, P. A. de O.; Machado, E.H.; Diana, R.A.D. Entendendo a adulteração de combustíveis. 2007. 3ª Ed.
3. Portaria do Ministério da Agricultura, Pecuária e Abastecimento (MAPA) Nº 75, de 05/03/2015.
4. Resolução Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) Nº 65, de 9.12.2011 - DOU 12.12.2011.
5. Resolução Agência Nacional de Petróleo, Gás Natural e Biocombustíveis (ANP) Nº 19, de 15/4/2015 republicada DOU 20/4/2015.
6. de Oliveira, M.F.; Saczk, A.A.; Okumura, L.L.; Stradiotto, N.R. *Energy Fuels* **2009**, *23*, 4852.
7. Chambriard, M.M.R.; *Resolução ANP nº14*. Agência Nacional de Petróleo, Gás Natural e Biocombustíveis. 2012.
8. Demirbas, A. *Energy Convers. Manage.* **2009**, *50*, 14.
9. Guarieiro, L.L.N.; de Souza, A.F.; Torres, E.A.; de Andrade, J.B.; *Atmos. Environ.* **2009**, *43*, 2754.
10. Rico, J.A.P.; Sauer, I.L., *Renewable Sustainable Energy Rev.* **2015**, *45*, 513.
11. Aransiola, E.F.; Ojumu, T.V.; Oyekola, O.O.; Madzimbamuto, T.F.; Ikhu-Omoregbe, D.I.O.; *Biomass Bioenergy* **2014**, *61*, 276.
12. Zhang, H.; Aytun, O.U.; Wang, Q.; Zhao, Z.; *Renewable Sustainable Energy Rev.* **2014**, *38*, 677.
13. Nurfitri, I.; Maniam, G.P.; Hindryawati, N.; Yusoff, M.M.; Ganesan, S.; *Energy Convers. Manage.* **2013**, *74*, 395.
14. Yaakob, Z.; Narayanan, B.N.; Padikkaparambil, S.; Unni K.S.; Akbar P.M.; *Renewable Sustainable Energy Rev.* **2014**, *35*, 136.
15. Shahir, V.K.; Jawahar, C.P.; Suresh, P.R.; *Renewable Sustainable Energy Rev.* **2015**, *45*, 686.
16. Oh, P.P.; Lau, H.L.N.; Chen, J.; Chong, M.F.; Choo, Y.M. *Renewable Sustainable Energy Rev.* **2012**, *16*, 5131.
17. Vicente, G.; Martinez, M.; Aracil, J. *Bioresour. Technol.* **2004**, *92*, 297.
18. Sharma, Y.C.; Singh, B.; Upadhyay, S.N. *Fuel* **2008**, *87*, 2355.
19. Marchetti, J.M.; Miguel, V.U.; Errazu, A.F. *Renewable Sustainable Energy Rev.* **2007**, *11*, 1300.
20. Liu, Y.; Lotero, E.; Goodwin, J.G. *J. Mol. Catal. A: Chem.* **2006**, *245*, 132.
21. Singh, A.; He, B.; Thompson, J.; Van Gerpen, J. *Applied Engineering in Agriculture* **2006**, *22*, 597.
22. Freedman, B., Pryde, E.H.; Mounts, T.L. *J. Am. Oil Chem. Soc.* **1984**, *61*, 1638.
23. Srivastava, A.; Prasad, R. *Renewable Sustainable Energy Rev.* **2000**, *4*, 111.

24. Samios, D.; Pedrotti, F.; Nicolau, A.; Reiznautt, Q.B.; Martini, D.D.; Dalcin, F.M. *Fuel Process. Technol.* **2009**, *90*, 599.
25. Guzatto, R.; Defferrari, D.; Reiznautt, Q.B.; Cadore, Í.R.; Samios, D. *Fuel* **2012**, *92*, 197.
26. Guzatto, R.; de Martini, T.L.; Samios, D. *Fuel Process. Technol.* **2011**, *92*, 2083.
27. <http://www.anp.gov.br>, acessado em janeiro de 2017.
28. Obeidat, S.M. e Alomary, A.A. *Appl. Magnet. Reson.* **2016**, *47*, 1273.
29. Pavia, D.L.; Lampman, G.M.; Kriz, G.S.; VyVyan, J.R. *Introdução à Espectroscopia*. 2010, Tradução da 4ª edição norte-americana, Editora Cengage Learning.
30. Balci, M. *Basic ¹H- and ¹³C-NMR Spectroscopy*. 2005, 1ª Edição, Editora Elsevier.
31. Lambert, J.B.; Mazzola, E.P. *Nuclear Magnetic Resonance Spectroscopy: An Introduction to Principles, Applications, and Experimental Methods*. 2004, 1ª Edição, Editora Pearson.
32. Silverstein, R.M.; Webster, F.X.; Kiemle, D.J.; D.L. Bryce. *Spectrometric Identification of Organic Compounds*, 2014, 8ª Edição, Editora John Wiley & Sons, Inc.
33. Macomber, R.S. *A Complete Introduction to Modern NMR Spectroscopy*, 1998, 1ª Edição, Editora John Wiley & Sons, Inc.
34. Barbosa, L.C. de A. *Espectroscopia no infravermelho na caracterização de compostos orgânicos*, 2007, 1ª Edição, Editora UFV.
35. Ferrão, M.F. Aplicação de técnicas espectroscópicas de reflexão no infravermelho no controle de qualidade de farinha de trigo, 2000, Tese de Doutorado, Universidade Estadual de Campinas.
36. Aliske, M.A. Medidas de espectrometria no infravermelho médio para determinação do teor de biodiesel em óleo Diesel, 2010, Dissertação de Mestrado, Universidade Federal do Paraná.
37. Correa, C. Aplicação de cartas de controle multivariadas no controle e garantia de qualidade do biodiesel e suas blendas (B100), 2014, Dissertação de Mestrado, Universidade Federal do Rio Grande do sul.
38. Lindon, J.C.; Tranter, G.E.; Holmes, J.L. *Encyclopedia of Spectroscopy and Spectrometry Part 1*, 2000, editora Elsevier.
39. Ferreira, S.L.C. *Introdução às Técnicas de Planejamento de Experimentos*. 2015, 1ª Edição, Editora Vento Leste.
40. Pereira Filho, E.R. *Planejamento Fatorial em Química*. 2015, 1ª Edição, Editora EdUFSCar.
41. Ferreira, M.M.C. *Quimiometria Conceitos, Métodos e Aplicações*. 2015, 1ª Edição, Editora UNICAMP.
42. Barros Neto, B.de.; Scarminio, I.S.; Bruns, R.E. *Como Fazer Experimentos*. 2010, 4ª Edição, Editora Bookman.

43. Kowalski, B.R.; Bender, C. *J. Am. Chem. Soc.* **1972**, *94*, 5632.
44. Lawton, W.; Sylvestre, E. *Technometrics* **1971**, *13*, 617.
45. Geladi, P.; Esbensen, K. *J. Chemom.* **1990**, *4*, 337.
46. Kowalski, B.R. *Chemometrics: Theory and Application*. ACS Symp. Ser. 52, American Chemical Society, Washington, D.C., 1977.
47. Doehlert, D.H.; Klee, V.L.; *Discrete Mathematics* **1972**, *2*, 309.
48. Ferreira, S.L.C.; Dos Santos, W.N.L.; Quintella, C.M.; Neto, B.B.; Bosque-Sendra, J.M.; *Talanta* **2004**, *63*, 1061.
49. Depoi, F.S.; Bentlin, F.R.S.; Ferrão, M.F.; Pozebon, D.; *Anal. Methods* **2012**, *4*, 2809.
50. Valle, P.W.P.A.; Rezende, T.F.; Souza, R.A.; Fortes, I.C.P.; Pasa, V.M.D.; *Energy Fuels* **2009**, *23*, 5219.
51. Borges, K.A.; Squissato, A.L.; Santos, D.Q.; Neto, W.B.; Batista, A.C.F.; Silva, T.A.; Vieira, A.T.; de Oliveira, M.F.; Hernández-Terrones, M.G.; *Energy* **2014**, *67*, 569.
52. Hamze, H.; Akia, M.; Yazdani, F.; *Process Saf. Environ. Prot.* **2015**, *94*, 1.
53. Wold, S.; Esbensen, K.; Geladi, P. *Chemom. Intell. Lab. Syst.* **1987**, *2*, 37.
54. Bro, R.; Smilde, A.K. *Anal. Meth.* **2014**, *6*, 2812.
55. dos Santos, V.H.J.M., Bruzza, E. do C. ; de Lima, J.E. ; Lourega, R.V.; Rodrigues, L.F. *Energy Fuels* **2016**, *30*, 4905.
56. Ferrão, M.F.; Viera, M.D.; Pazos, R.E.P.; Fachini, D.; Gerbase, A.E.; Marder, L. *Fuel* **2011**, *90*, 701.
57. da Silva, V.A.G., Talhavini, M.; Zacca, J.J.; Trindade, B.R.; Braga, J.W.B. *J. Braz. Chem. Soc.* **2014**, *25*, 1552.
58. Vitale, R.; Bevilacqua, M.; Bucci, R.; Magrì, A.D.; Magrì, A.L.; Marini, F. *Chemom. Intell. Lab. Syst.* **2012**, *121*, 90.
59. Grasel, F.S; Ferrão, M.F. *Anal. Methods.* **2016**, *8*, 644.
60. Oliveira Junior, H.A., Caldeira, A.M.; Machado, M.A.S.; Souza, R.C.; Tanscheit, R. *Inteligência Computacional aplicada à Administração, Economia e Engenharia em Matlab*. 2007. Editora Thomson.
61. Chen, N.; Lu, W.; Yang, J.; Guozheng, L. *Support Vector Machine in Chemistry*. 2004, Editora World Scientific.
62. Filgueiras, P.R. Regressão por vetores de suporte aplicado na determinação de propriedades físico-químicas de petróleo e biocombustíveis, 2014, Tese de Doutorado, Universidade Estadual de Campinas.

63. Correa, C.; Ruschel, C. F. C.; Bento, F. M.; Ferrão, M. F. *Rev. Virtual Quim.* **2015**, *7*, 2273.
64. Skibsted, E.T.S.; Westerhuis, J.A.; Boelens, H.F.; Smilde, A.K.; Witte, D.T.; *Applied Spectroscopy.* **2004**, *58*, 264.
65. dos Santos Junior, A. A. Determinação de Parâmetros de óleos lubrificantes de motores de ciclo Otto e ciclo Diesel através da Espectroscopia no Infravermelho, métodos multivariados e cartas de controle. **2011**. Dissertação de Mestrado. Universidade de Brasília.
66. Lorber, A. *Anal. Chem.* **1986**, *58*, 1167.
67. Lorber, A.; Faber, N.M.; Kowalski, B.R. *Anal. Chem.* **1997**, *69*, 1620.
68. Skibsted, E.T.S.; Boelens, H.F.M.; Westerhuis, J.A.; Smilde, A.K.; Broad, N.W.; Rees, D.R.; Witte, D.T. *Anal. Chem.* **2005**, *77*, 7103.
69. Jackson, J. E., & Mudholkar, G. S. *Technometrics* **1979**, *21*, 341.
70. Rocha, W.F. de C.; Rosa, A.L.; Martins, J.A.; Poppi, R.J. *J. Mol. Struct.* **2010**, *982*, 73-78.
71. de Oliveira, I. K.; de Carvalho Rocha, W. F.; Poppi, R. J. *Anal. Chim. Acta* **2009**, *642*, 217.
72. Mitsutake, H.; Guimarães, E.; Freitas, H.C.; Gontijo, L.C.; Santos, D.Q.; Borges Neto, W.J. *Chemom.* **2015**, *29*, 411.
73. Teófilo, R.F.; Ferreira, M.M.C.; *Quim. Nova* **2006**, *29*, 338.
74. Jiang, W., Zhang, Z.M.; Yun, Y.H.; Zhan, D.J.; Zheng, Y.B.; Liang, Y.Z.; Yang, Z.Y.; Yu, L. *Chromatographia* **2013**, *76*, 1067.
75. Wong, J.W.H., Durante, C.; Cartwright, H.M. *Anal. Chem.* **2005**, *77*, 5655.
76. Kennard, R.W. e Stone, L.A. *Technometrics*, **1969**, *11*, 137.