



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE QUÍMICA
DISSERTAÇÃO DE MESTRADO

**“Espectrometria no Infravermelho Médio e Calibração
Multivariada por PLS para quantificação de biodiesel
metílico e etílico de soja, milho e girassol em misturas
com diesel”**

Mestrando: Fábio dos Santos Fonseca

Orientador: Prof. Dr. Waldomiro Borges Neto

Co-orientador: Prof. Dr. Manuel Gonzalo Hernández-Terrones

Uberlândia
Novembro 2011



UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE QUÍMICA
DISSERTAÇÃO DE MESTRADO

“Espectrometria no Infravermelho Médio e Calibração Multivariada por PLS para quantificação de biodiesel metílico e etílico de soja, milho e girassol em misturas com diesel”

Dissertação apresentada ao Programa de Pós-Graduação em Química da Universidade Federal de Uberlândia, como requisito para obtenção de título de Mestre em Química.

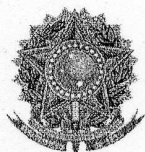
Mestrando: Fábio dos Santos Fonseca

Orientador: Prof. Dr. Waldomiro Borges Neto

Co-orientador: Prof. Dr. Manuel Gonzalo Hernández-Terrones

Área de Concentração: Química Analítica

Uberlândia
Novembro 2011



SERVIÇO PÚBLICO FEDERAL
MINISTÉRIO DA EDUCAÇÃO
UNIVERSIDADE FEDERAL DE UBERLÂNDIA
INSTITUTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA



Ata da defesa de DISSERTAÇÃO DE MESTRADO junto ao Programa de Pós-Graduação em Química, do Instituto de Química da Universidade Federal de Uberlândia
DEFESA DE DISSERTAÇÃO DE MESTRADO ACADÊMICO, NÚMERO 133/COPGQ.

DATA: 25/11/2011

DISCENTE: Fábio dos Santos Fonseca

MATRÍCULA: 10922QMI005

TÍTULO DO TRABALHO: Espectrometria no infravermelho médio e calibração multivariada por PLS para quantificação de biodiesel de soja, milho e girassol; produzidos pelas rotas metílica e etílica, em diesel”.

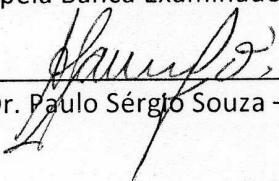
ÁREA DE CONCENTRAÇÃO: Química

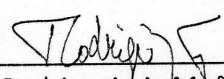
LINHA DE PESQUISA: Físico-Química de Macromoléculas e Colóides


PROJETO DE PESQUISA DE VINCULAÇÃO: Controle de Qualidade de Biocombustíveis.

Às quinze horas do dia vinte e cinco de novembro do ano dois mil e onze, na sala 102 do Bloco 1B, Campus Santa Mônica, reuniu-se a Banca Examinadora composta pelos Professores Doutores: Paulo Sérgio Souza, da Universidade Federal de Goiás, de Goiânia/GO; Rodrigo Alejandro Abarza Muñoz, do IQUFU e Waldomiro Borges Neto, professor orientador e presidente da mesa. Iniciando os trabalhos, o presidente da mesa apresentou o candidato e a Banca Examinadora, agradeceu a presença do público e discorreu sobre as normas e critérios para a realização desta sessão, baseadas na Norma Interna nº 01/2010/COPGQ. Em seguida, o presidente da mesa concedeu a palavra ao candidato para a exposição do seu trabalho e, em seqüência, aos examinadores, em ordem sucessiva, para arguir o apresentador. A duração da apresentação e o tempo de arguição e resposta deram-se conforme as normas do Programa. Ultimada a arguição, desenvolvida dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu os conceitos finais e A provou o candidato. Por sugestão da Banca Examinadora, o título do trabalho será Espectro metílico no Infravermelho médio e Calibração Multivariada por PLS para quantificação de biodiesel metílico e etílico de soja, milho e girassol, em misturas com diesel

Esta defesa de Dissertação de Mestrado Acadêmico é parte dos requisitos necessários à obtenção do título de Mestre. O competente diploma será expedido após cumprimento do estabelecido nas normas do Programa, legislação e regulamentação internas da UFU, em especial do artigo 55 da resolução 12/2008 do Conselho de Pós-graduação e Pesquisa da Universidade Federal de Uberlândia. Nada mais havendo a tratar, deu-se por encerrada a sessão às 17 horas e 45 minutos e lavrada a presente ata, que após lida e aprovada, foi assinada pela Banca Examinadora.


Prof. Dr. Paulo Sérgio Souza – UFG


Prof. Dr. Rodrigo A. A. Muñoz – IQUFU


Prof. Dr. Waldomiro Borges Neto – IQUFU
Orientador

“There is no knowledge that is not power”

(Frase de introdução do jogo Mortal Kombat II, Midway, 1993)

Dedico este trabalho em especial à minha amada esposa

Dinamar Carneiro Fonseca

aos meus queridos e amados filhos

Vitor Carneiro Fonseca e Sabrina Carneiro Fonseca

à meus queridos e amados pais

Ernesto Augusto Fonseca e Helena de São José Rocha

aos meus irmãos

Cristina dos Santos Fonseca

Priscila dos Santos Fonseca

Alexandre Gomes Fonseca

AGRADECIMENTOS

"Ciência é conhecimento organizado.

Sabedoria é vida organizada".

Immanuel Kant (filósofo alemão, fundador da filosofia crítica)

- A Deus pela essência da vida, meus familiares, amigos e, em especial, pela possibilidade de realizar este trabalho;
- À minha amada esposa Dinamar, pelo amor, incentivo, paciência e por acreditar em nossos sonhos;
- Ao Prof. Dr. Waldomiro Borges Neto pela oportunidade de realização desse trabalho, orientação, conhecimentos transmitidos, paciência, confiança, disponibilidade, amizade, diversão e pela teimosia durante todo este trabalho. Agradecerei, portanto, não somente com palavras que ecoam e se perdem com o vento, mas principalmente com ações que permanecerão, reproduzindo aquilo que outrora me passaste: o conhecimento;
- Ao Prof. Dr. Manuel Gonzalo Hernández-Terrones pela co-orientação, parceira e ajuda.
- Ao grupo LQT pelos agradáveis momentos vividos, os congressos, comemorações, muito trabalho, discussões, risos e histórias;
- Ao grupo LABIQ pela parceria na produção dos biodiesel;
- À Universidade Federal de Uberlândia pela oportunidade de aprimorar os meus conhecimentos e pela minha formação;
- Ao Instituto de Química da UFU por fornecer toda estrutura física e tecnológica para a realização deste trabalho;
- À Transpetro SA pelo fornecimento do diesel;
- À FAPEMIG pelo auxílio durante os eventos onde foram apresentados os resultados;
- Aos meus amados pais, Ernesto e Helena; filhos, Vitor e Sabrina; irmãos, Cristina, Priscila e Alexandre, sogros, Odimar e Lúcia; e toda minha família, por todo amor, apoio, incentivo e compreensão durante toda minha vida;
- A todos aqueles que direta ou indiretamente tiveram sua parcela de participação durante a execução e conclusão deste trabalho.

RESUMO

A maior parte da energia consumida no mundo provém do petróleo, do carvão e do gás natural. Essas fontes são limitadas, portanto, fontes alternativas são necessárias e a busca de tais fontes é de suma importância. Os óleos vegetais têm sido apontados como prováveis substituintes ao óleo diesel desde o início do século passado por possuírem características físico-químicas próximas ou passíveis de adequação ao motor diesel. Eles podem ser utilizados nos motores diesel com ou sem processamento. A partir 01 de Janeiro de 2010 ficou estabelecida a adição de 5% de Biodiesel ao Diesel brasileiro. A transformação do óleo vegetal em biodiesel visa apenas uma melhor adaptação ao motor. A cadeia produtiva brasileira de biodiesel atualmente não possui regulamentação quanto à matriz a ser utilizada, sendo regulada atualmente pela disponibilidade de matérias-primas. Ainda não há padronização de método de análise de biodiesel em diesel. O biodiesel puro é analisado primeiro sendo posteriormente adicionado ao diesel.

Neste trabalho, oito modelos de calibração multivariada baseados em espectrometria de infravermelho médio foram desenvolvidos no intuito de determinar o conteúdo de biodiesel em misturas binárias com diesel, na faixa de 1 a 20% v/v, genericamente designados BX. Ésteres (metílicos e etílicos) feitos de óleo de soja, milho, girassol e óleo de soja de fritura foram usados para preparar as misturas com diesel comum. Resultados indicaram que modelos de quadrados mínimos parciais (PLS) baseados em espectros de infravermelho médio foram comprovadamente adequados como métodos analíticos práticos para previsão do conteúdo de biodiesel em misturas com diesel comum na faixa de volume de 1 a 20%. Os modelos PLS foram validados com conjunto de amostras independentes, sendo os valores de RMSEP (erro quadrático médio de previsão) inferiores a 0,2266; RMSECs (erros quadráticos médios de previsão) inferiores a 0,1532. Correlações lineares observadas no gráfico de valores previstos vs. Reais com coeficientes de correlação (R^2) maiores que 0,999. Adicionalmente, os erros médios de previsão encontrados foram inferiores a 2,21%.

ABSTRACT

Most of the energy consumed worldwide comes from oil, coal and natural gas. These sources are limited, so alternative sources are needed and the search for such sources is very important. Vegetable oils have been mentioned as likely replacements for diesel fuel since the beginning of last century by having physical and chemical characteristics that could be close or suitable for a diesel engine. They can be used in diesel engines with or without processing. Since January 1st of 2010 was established the addition of 5% biodiesel to diesel in Brazil. The transformation of vegetable oil into biodiesel refers only to better adapt to the engine. The Brazilian Biodiesel Production Chain currently has no regulations regarding the matrix to be used, and is currently regulated by the availability of raw materials. There is still no standardized method for analyzing biodiesel in diesel. Pure biodiesel is analyzed first and then added to diesel.

In this work, eight multivariable calibration models based on middle-infrared spectroscopy were developed in order to determine the content of biodiesel in diesel fuel (binary blends), in the range of 1 to 20% v/v, generically called BX. Esters (methyl and ethyl) made with soybean, corn, sunflower seed oils and used soybean frying oil were used to prepare the blends with conventional diesel. Results indicated that partial least squares (PLS) models based on MID infrared spectra were proven suitable as practical analytical methods for predicting biodiesel content in conventional diesel blends in the volume fraction range from 1% to 20%. PLS models were validated by independent prediction set and the RMSEPs (Root Mean Square Errors of Prediction) were estimated as low as 0,2266; RMSECs (Root Mean Square Errors of Calibration) were estimated as low as 0,1532. Linear correlations were observed for predicted vs. observed values plots with correlation coefficient (R) greater than 0,999. Additionally, the average preview errors found were less than 2,21%.

LISTA DE ABREVIATURAS OU SIGLAS

^1H NMR – (do inglês, Nuclear Magnetic Resonance) Ressonância Magnética Nuclear de Prótons
ABIOVE - Associação Brasileira das Indústrias de Óleos Vegetais
ABNT - Associação Brasileira de Normas Técnicas
AGL - Ácidos Graxos Livres
ANP - Agência Nacional do Petróleo, Gás Natural e Biocombustíveis
ANVISA – Agência Nacional de Vigilância Sanitária
ASTM - Sociedade Americana para Testes e Materiais
ATR – (do inglês, Attenuated Total Reflectance) Refletância Total Atenuada;
B100 - Biodiesel puro
CNPE - Conselho Nacional de Política Energética;
CV – (do inglês, Cross Validation) Validação cruzada;
 \bar{E} - média do erro relativo;
FTIR - (do inglês, Fourier Transform Infrared) Infravermelho com transformada de Fourier;
MIR - (do inglês, Middle Infrared) Infravermelho Médio
MLR - (do inglês, Multiple Linear Regression) Regressão Linear Múltipla
NIR - (do inglês, Near Infrared) Infravermelho Próximo
PCA – (do inglês, Principal Component Analysis) Análise de Componentes Principais
PCR – (do inglês, Principal Component Regression) Regressão por Componentes Principais
PC – (do inglês, Principal Component) Componente Principal
PLS – (do inglês, Partial Least Squares) Quadrados Mínimos Parciais
RMSE – (do inglês, Root Mean Square Error) Raiz do Erro Quadrático Médio
RMSEC – (do inglês, Root Mean Square Error of Calibration) Erro Quadrático Médio da Calibração
RMSEP - (do inglês, Root Mean Square Error of Prediction) Erro Quadrático Médio da Previsão
UV – Ultravioleta
Vis – Visível
VL - Variável Latente

LISTA DE TABELAS

Tabela 1 - Composição média de ácidos graxos dos óleos de soja, girassol e milho.....	6
Tabela 2 - Especificações do óleo diesel e do biodiesel de óleo residual de soja.....	8
Tabela 3 - Regiões do Infravermelho.	12
Tabela 5 - Classificação dos valores de linearidade [54].....	40
Tabela 5 - Concentração real e prevista para mistura de biodiesel etílico de soja em diesel (%v/v).	42
Tabela 6 - Concentração real e prevista para mistura de biodiesel metílico de soja em diesel (% v/v).	45
Tabela 7 - Concentração real e prevista de mistura de biodiesel etílico de soja residual em diesel.....	46
Tabela 8 - Concentração real e prevista para mistura de biodiesel metílico de soja residual em diesel.	48
Tabela 9 - Valores de concentração real e prevista para mistura de biodiesel etílico de milho em diesel.	50
Tabela 10 - Concentração real e prevista para mistura de biodiesel metílico de milho em diesel.	52
Tabela 11 - Concentração real e prevista para mistura de biodiesel etílico de girassol em diesel.	54
Tabela 12 - Concentração real e prevista para mistura de biodiesel metílico de girassol em diesel.	56
Tabela 13 - Resumo dos modelos PLS.....	57
Tabela 14 - Comparação dos resultados obtidos com resultados da literatura científica.....	58
Tabela 15 - Resultados obtidos e da literatura para quantificação por PLS.	58

LISTA DE FIGURAS

Figura 1 - Reação de transesterificação de triglicerídeos.	3
Figura 2 - Gráfico de matérias-primas utilizadas na produção do biodiesel brasileiro (Fonte: ANP).	5
Figura 3 - Demanda obrigatória de biodiesel no Brasil.	5
Figura 4 - Vendas e importações de diesel no Brasil.	9
Figura 5 - Representação do ATR.	14
Figura 6 - Bandas de absorção de estiramento da carbonila.	15
Figura 7 - Espectro do diesel puro.	17
Figura 8 - Espectro do biodiesel puro.	17
Figura 9 - Espectros de misturas binárias biodiesel/diesel de concentrações de 1 a 20% v/v.	17
Figura 10 - Representação da construção da matriz de dados X para calibração multivariada.	20
Figura 11 - Espectro MIR com flutuação na linha de base (a) Espectro MIR corrigido e cortado (b).	21
Figura 12 - Formação da Componente Principal Zero.	24
Figura 13 - Formação da Componente Principal Um.	25
Figura 14 - Formação da Componente Principal Dois.	25
Figura 15 - Decomposição em componentes principais por PCA.	26
Figura 16 - Formação da Variável Latente 1.	27
Figura 17 - Decomposição em variáveis latentes das matrizes X e Y para modelos PLS.	29
Figura 18 - Espectros originais das misturas binárias biodiesel/diesel (1 a 20% v/v)	36
Figura 19 - Espectros de misturas binárias biodiesel/diesel (1 a 20% v/v) após correção da linha de base.	36
Figura 20 - Espectros de misturas binárias biodiesel/diesel (1 a 20% v/v) após correção da linha de base e cortes.	36
Figura 21 - Resultado final dos espectros de misturas binárias biodiesel/diesel (1 a 20% v/v).	37
Figura 22 - Áreas de grande variação de absorvâncias entre os espectros.	37
Figura 23 - Tela de escolha do número de variáveis latentes.	39
Figura 24 - Gráfico dos valores de RMSEC VS Número de Variáveis latentes.	39
Figura 25 - Exemplo de desvios equidistantes da linha de tendência.	41
Figura 26 - Espectros MIR para as amostras de misturas de biodiesel etílico de soja em diesel.	41
Figura 27 - Gráfico dos valores de RMSEC VS VL (biodiesel etílico de soja em diesel).	42
Figura 28 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel etílico de soja).	43
Figura 29 - Espectros MIR para biodiesel metílico de soja em diesel.	43
Figura 30 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel metílico de soja em diesel.	44
Figura 31 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de soja).	44
Figura 32 - Espectros MIR para biodiesel etílico de soja residual em diesel.	45
Figura 33 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de soja residual em diesel.	46
Figura 34 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel etílico de soja residual).	46
Figura 35 - Espectros MIR para biodiesel metílico de soja residual em diesel.	47
Figura 36 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel metílico de soja residual.	48
Figura 37 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de soja residual).	48
Figura 38 - Espectros MIR para biodiesel etílico de milho em diesel.	49
Figura 39 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de milho.	50
Figura 40 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel etílico de milho).	50
Figura 41 - Espectros MIR para biodiesel metílico de milho em diesel.	51
Figura 42 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel metílico de milho.	52
Figura 43 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de milho).	52
Figura 44 - Espectros MIR para biodiesel etílico de girassol em diesel.	53
Figura 45 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de girassol.	54
Figura 46 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de milho).	54
Figura 47 - Espectros MIR para biodiesel metílico de girassol em diesel.	55
Figura 48 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de girassol.	55
Figura 49 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de milho).	56

SUMÁRIO

CAPÍTULO 1: Biodiesel	1
1. Biodiesel.....	2
1.1. Definições de Biodiesel.....	2
1.2. Obrigatoriedade do uso do Biodiesel no Brasil	2
1.3. Obtenção do Biodiesel	2
1.3.1. Diferença entre etanol e metanol na produção de Biodiesel.....	4
1.3.2. Matérias-primas utilizadas na produção de Biodiesel	4
1.3.3. Degradação Térmica de óleos comestíveis	7
1.4. Substituição do óleo diesel por Biodiesel.....	8
1.5. Parâmetros de Qualidade do Biodiesel.....	10
CAPÍTULO 2: Espectroscopia Vibracional	11
2. Espectroscopia Vibracional.....	12
2.1. Aplicações e usos da espectroscopia no Infravermelho.....	12
2.1.1. Espectroscopia de Refletância Total Atenuada (ATR).....	13
2.2. Interpretação de Espectros	14
2.2.1. Absorção característica de grupos de moléculas orgânicas	15
2.2.2. Caracterização de grupos funcionais nos espectros	16
CAPÍTULO 3: Métodos Quimiométricos	19
3. Quimiometria	20
3.1. Organização dos Dados para o programa.....	20
3.2. Pré-Tratamento de Dados	21
3.3. Análise Multivariada	21
3.4. Calibração Multivariada	22
3.4.1. Análise de Componentes Principais - PCA.....	23
3.4.2. Quadrados Mínimos Parciais - PLS	26
3.5. Cálculo dos erros	29
3.6. Linearidade dos Modelos.....	30
CAPÍTULO 4: Procedimento Experimental.....	31
4. Procedimento Experimental.....	32
4.1. Preparação das amostras	32
4.2. Obtenção dos espectros no MIR	32
4.3. Importação dos dados para Matlab para geração da matriz de dados	32
4.4. Tratamento dos espectros das misturas binárias biodiesel/diesel.....	32
4.5. Construção dos Modelos.....	33
CAPÍTULO 5: Resultados e Discussões.....	34
5. Resultados e Discussões	35

5.1. Preparação das amostras	35
5.2. Obtenção dos espectros de Infravermelho Médio	35
5.4. Tratamento dos espectros das misturas binárias biodiesel/diesel	35
5.5. Construção dos Modelos	37
5.5.1. Correlação dos espectros com propriedade de interesse	37
5.5.2. Dados Centrados na Média	38
5.5.3. Determinação no número de variáveis latentes para construção dos modelos	38
5.5.4. Validação Cruzada.....	39
5.5.5. Linearidade dos Modelos	40
5.5.6. Gráficos de Valores Reais versus Valores Previstos	40
5.5.7. Erros de previsão	41
5.6. Modelos de quantificação de biodiesel em diesel.....	41
5.6.1. Biodiesel etílico de soja.....	41
5.6.2. Biodiesel metílico de soja.....	43
5.6.3. Biodiesel etílico de soja residual	45
5.6.4. Biodiesel metílico de soja residual.....	47
5.6.5. Biodiesel etílico de milho.....	49
5.6.6. Biodiesel metílico de milho	51
5.6.7. Biodiesel etílico de girassol	53
5.6.8. Biodiesel metílico de girassol	55
5.6.9. Resumo dos modelos	56
5.6.10. Comparação dos modelos com outros trabalhos publicados.....	57
CAPÍTULO 6: Conclusão.....	60
6. Conclusão.....	61
CAPÍTULO 7: Bibliografia.....	62

CAPÍTULO 1: Biodiesel

1. Biodiesel

1.1. Definições de Biodiesel

Biodiesel foi definido pela *National Biodiesel Board* dos Estados Unidos como o derivado mono-alquil éster de ácidos graxos de cadeia longa, proveniente de fontes renováveis como óleos vegetais ou gordura animal, cuja utilização está associada à substituição de combustíveis fósseis em motores de ignição por compressão (motores do ciclo diesel). [1]

O Art. 6º, inciso XXIV da Lei nº 11.097/2005 o define como “biocombustível derivado de biomassa renovável para uso em motores a combustão interna com ignição por compressão ou, conforme regulamento, para geração de outro tipo de energia, que possa substituir parcial ou totalmente combustíveis de origem fóssil.” [2]

Já a Resolução ANP nº 7, de 19 de março de 2008, coloca da seguinte forma em seu Art. 2º, inciso I: “biodiesel – B100 – combustível composto de alquil-ésteres de ácidos graxos de cadeia longa, derivados de óleos vegetais ou de gorduras animais conforme a especificação contida no Regulamento Técnico, parte integrante desta Resolução.” [3]

De modo sucinto, biodiesel é o combustível produzido a partir de óleos vegetais ou gorduras animais que visa substituir total ou parcialmente o combustível diesel de petróleo.

1.2. Obrigatoriedade do uso do Biodiesel no Brasil

A Lei Nº 11.097, de 13 de Janeiro de 2005 dispõe sobre a introdução do biodiesel na matriz energética brasileira, introduzindo um percentual mínimo de adição ao diesel de 2% (v/v). [2]

A Resolução nº 2, de 13 de março de 2008, do Conselho Nacional de Política Energética – CNPE aumentou a mistura obrigatória de biodiesel no diesel de 2% (v/v) para 3% (v/v). [4]

A Resolução nº 2, de 27 de abril de 2009, do CNPE, determinou que o percentual de biodiesel em diesel fosse de 4% (v/v) a partir de julho de 2009, permitindo percentuais entre 4 e 5% (v/v), dependendo da viabilidade econômica. Sendo que a partir 01 de Janeiro de 2010 ficou estabelecida a adição de 5% (v/v) de biodiesel ao diesel. Esta regra foi estabelecida pela Resolução nº 6/2009 do Conselho Nacional de Política Energética (CNPE), publicada no Diário Oficial da União (DOU) em 26 de outubro de 2009. [5]

1.3. Obtenção do Biodiesel

Os óleos vegetais têm sido apontados como prováveis substituintes ao óleo diesel desde o início do século passado por possuírem características físico-químicas próximas ou passíveis de adequação ao motor diesel. Eles podem ser utilizados nos motores diesel com ou sem processamento. O craqueamento (térmico ou catalítico) e a transesterificação são processos que visam à adequação do combustível ao motor.

Para a obtenção de biodiesel, a reação de transesterificação de óleos vegetais com álcoois primários pode ser realizada tanto em meio ácido quanto em meio básico, conforme demonstrado na Figura 1.

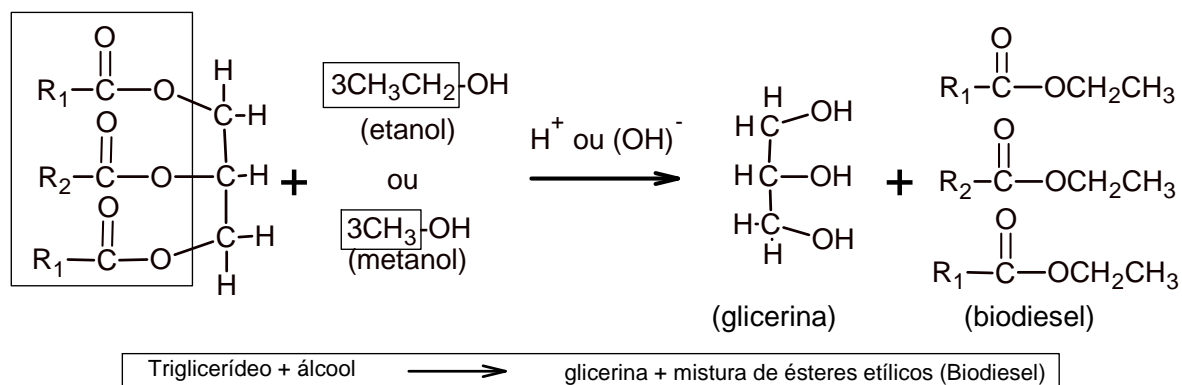


Figura 1 - Reação de transesterificação de triglicerídeos.

Transesterificação é processo mais utilizado atualmente para a produção de biodiesel. A reação de síntese, geralmente empregada em nível industrial, utiliza uma razão molar óleo/álcool de 1:6 na presença de 0,4% de hidróxido de sódio ou de potássio, porque o meio básico apresenta melhor rendimento e menor tempo de reação do que o meio ácido. Por outro lado, o excesso de agente transesterificante (álcool primário) faz-se necessário para deslocar o equilíbrio estabelecido na reação em direção à formação de biodiesel. [6]

O produto da reação, basicamente constituído de uma mistura de ésteres, glicerina, catalisador e sabão; apresenta-se em duas fases distintas que são separadas no próprio reator através de uma válvula de descarga. A fase superior desta mistura contém os ésteres, enquanto a camada inferior é formada essencialmente de glicerina, podendo ser separada por decantação simples. No entanto, antes da filtração, a mistura de ésteres deve ser lavada com água quente de modo a eliminar traços de catalisador, de sabão ou de glicerina residual.

Os óleos vegetais, quando transformados em ésteres etílicos ou metílicos através da reação de transesterificação, possuem características físico-químicas mais próximas às do óleo diesel do que as do óleo "in natura". Além disso, a mistura de ésteres formada, também denominada biodiesel, possui características mais independentes da espécie vegetal utilizada. Os ésteres podem ser utilizados em motores diesel de injeção direta ou indireta. O processo de transesterificação de óleos vegetais via rota etílica ou rota metílica resulta em diversos benefícios em relação ao respectivo óleo vegetal: reduz a densidade e a viscosidade, enquadrando-se dentro da especificação do óleo diesel; diminui o ponto de fulgor e o ponto de névoa; reduz o resíduo de carbono a valores menores que os do diesel; dá origem a produtos com menor peso molecular, embora seja mantida a estrutura original dos ácidos graxos e aumenta o número de cetano (tornando-se maior ou igual ao diesel). [7] A transformação do óleo vegetal em biodiesel visa apenas uma melhor adaptação ao motor.

1.3.1. Diferença entre etanol e metanol na produção de Biodiesel

Existem algumas diferenças entre esses dois compostos, como descrito a seguir.

O metanol tem como característica principal ser 45% mais eficiente que o etanol (álcool etílico), pois é quimicamente mais reativo. Assim, o tempo e a temperatura no processo são menores, tornando-o mais atrativo economicamente. No entanto, por ser um combustível produzido a partir de petróleo, o biodiesel gerado acaba não sendo de fonte totalmente renovável. Além de ser muito tóxico, apresenta maior risco de incêndio, pois seu ponto de ignição é baixo. [8]

O etanol possui produção nacional instalada, devido a incentivos brasileiros na utilização do etanol na mistura com a gasolina, ou sozinho. Outra particularidade importante na utilização do etanol, para o processamento do biodiesel, é a presença de maior índice de cetanos e maior lubrificidade que o metanol, aumentando assim, o rendimento do biocombustível no veículo. Pelo fato do etanol ser produzido a partir de biomassa, podendo ser a cana-de-açúcar, o milho, a beterraba; produz-se biodiesel a partir de uma fonte 100% renovável, e com menor risco de explosão. Contudo, o etanol apresenta maior afinidade química com a glicerina, dificultando a sua separação e aumentando consequentemente o gasto de energia para esse processo. [8]

A alcoólise (quebra da molécula utilizando álcool) com metanol é tecnicamente mais viável do que a alcoólise com etanol, particularmente se esse corresponde ao etanol hidratado, cujo teor em água (4-6%) retarda a reação. O uso de etanol anidro na reação efetivamente minimiza este inconveniente, embora não implique em solução para o problema inerente à separação da glicerina do meio de reação que, no caso da síntese do éster metílico, pode ser facilmente obtida por simples decantação. [8]

1.3.2. Matérias-primas utilizadas na produção de Biodiesel

Os óleos vegetais são produtos naturais constituídos por uma mistura de ésteres derivados do glicerol (triacilgliceróis ou triglicerídeos), cujos ácidos graxos contêm cadeias de 8 a 24 átomos de carbono com diferentes graus de insaturação.

Conforme a espécie de oleaginosa, variações na composição química do óleo vegetal são expressas por variações na relação molar entre os diferentes ácidos graxos presentes na estrutura. Portanto, a análise da composição de ácidos graxos constitui o primeiro procedimento para a avaliação preliminar da qualidade do óleo bruto e/ou de seus produtos de transformação.

A maior parte do biodiesel brasileiro tem como matéria-prima a soja (Figura 2).

O óleo de soja atualmente ocupa a 2ª posição na oferta mundial de óleos e gorduras. Em 1990, a produção desse óleo situou-se em torno de 16,1 milhões de toneladas, seguida do óleo de palma (conhecido no Brasil como óleo de Dendê) com 10,8 milhões de toneladas. Outros óleos vegetais com produção mundial significativa foram os de canola e girassol, ambos com aproximadamente 8 milhões de toneladas, e os de algodão e amendoim, com aproximadamente 4 milhões de toneladas cada um.

Mesmo sendo substituíveis entre si, cada um desses óleos possui especificidades que os tornam mais ou menos adequados dependendo de seus usos finais. [9]

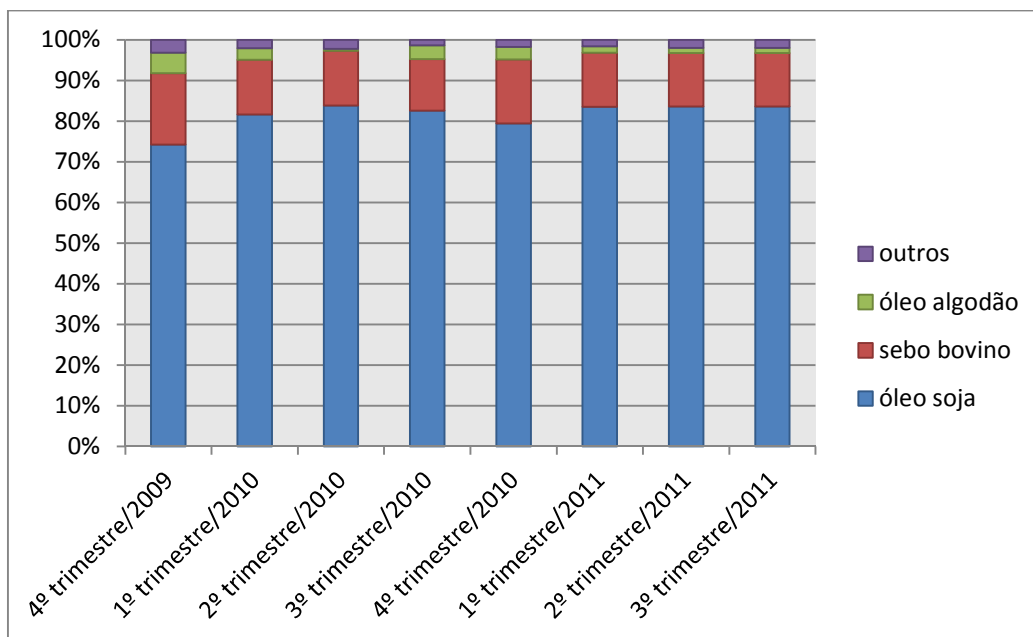


Figura 2 - Gráfico de matérias-primas utilizadas na produção do biodiesel brasileiro (Fonte: ANP).

Muitos deles, tal como ocorre com a soja, tem a sua produção vinculada ao crescimento da demanda de outros produtos da oleaginosa. É o caso do algodão, cuja principal demanda é a pluma para a indústria têxtil e de vestuário.

A soja no Brasil é produzida por mais de 243 000 agricultores dos mais variados portes (a soja ocupa em torno de 57% das culturas temporárias no Brasil), atingindo em 2005 a produção de 53 milhões de toneladas de grãos das quais 22 milhões foram exportados e 29 milhões destinados à indústria esmagadora que trabalha atualmente com 72% da capacidade instalada e é composta por um parque de mais de 50 empresas na qual apenas 12 representam 70%. [10] Em 2005 foram produzidos 57 milhões de toneladas cúbicas de óleo de soja no Brasil (mesma produção obtida em 2009, sendo que deste total, 30 milhões de toneladas cúbicas foram exportadas). [11] Com a previsão de demanda de biodiesel crescente (Figura 3), outras fontes além da soja deverão ter suas posições consolidadas e expandidas para abastecer este mercado em expansão.

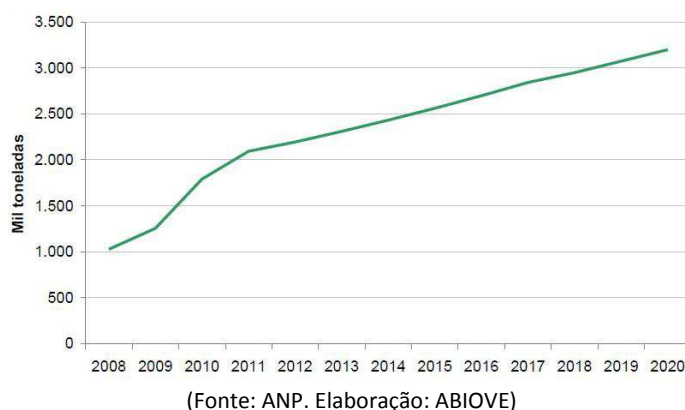


Figura 3 - Demanda obrigatória de biodiesel no Brasil.

Como mostrado na Tabela 1, os dois biodieseis derivados do óleo de soja, girassol e milho formados em maior quantidade derivam dos ácidos linoleico e oleico; como acontece também no óleo de milho e girassol.

Tabela 1 - Composição média de ácidos graxos dos óleos de soja, girassol e milho.

Nº de carbonos	Ácido graxo	Concentração (%)		
		Óleo de soja	Óleo de girassol	Óleo de milho
12	Láurico	0,1 (máx.)	-	-
14	Mirístico	0,2 (máx.)	0,5 (máx.)	0,3 (máx.)
16	Palmítico	9,9 – 12,2	3,0 – 10,0	9,0 – 14,0
16 (1 dupla)	Palmitoleico	Traços – 0,2	< 1,0	< 0,5
18	Esteárico	3 – 5,4	1,0 – 10,0	0,5 – 4,0
18 (1 dupla)	Oleico	17,7 – 26	14,0 – 35,0	24,0 – 42,0
18 (2 duplas)	Linoleico	49,7 – 56,9	55,0 – 75,0	34,0 – 62,0
18 (3 duplas)	Linolênico	5,5 – 9,5	< 0,3	< 2,0
20	Araquídico	0,2 – 0,5	< 1,5	< 1,0
20 (1 dupla)	Gadolênico	0,1 – 0,3	< 0,5	< 0,5
22	Behênico	0,3 – 0,7	< 1,0	< 0,5
22 (1 dupla)	Erúnico	0,3 (máx.)	0,3 (máx.)	-
24	Lignocérico	0,4 (máx.)	< 0,5	0,5 (máx.)
24 (1 dupla)	Nervônico	-	< 0,5	-

(Fonte: ANVISA)

O Brasil possui expressiva produção de biodiesel a partir de sebo bovino (Fig. 2), no entanto não é correto estimar a produção de sebo bovino, gordura de frango ou banha de porco com base na demanda por biodiesel, pois se sabe que esses produtos são derivados da produção e demanda por carnes. Adicionalmente, a utilização de sebo bovino para produção de biodiesel permite maiores ganhos, pois há legislação que limita a quantidade de sebo a ser vendida com a carne.

Tendo em vista as restrições na oferta de óleo em cada mercado, a demanda mundial crescente por óleo para uso alimentar impulsionou, muito antes do surgimento da demanda da indústria de biodiesel, o crescimento da produção de oleaginosas com teor de óleo mais elevado que a soja. Ou seja, a demanda mundial por óleos vegetais incentivou o crescimento da produção de variedades cujo rendimento em óleo por hectare era mais vantajoso. [10]

O caso mais proeminente dessa mudança no padrão de demanda foi a produção de óleo de palma. Neste caso, o crescimento foi tão acelerado que a palma desbancou a soja como a maior oferta de óleos vegetais do mundo. Outro caso importante foi do óleo de canola, que aumentou a produção para 19,1 milhões de toneladas, uma variação de 145% em menos de duas décadas. [9]

Atualmente, o óleo de soja responde por 73,3% da disponibilidade de óleos e gorduras no Brasil, seguida pelo sebo bovino, gordura de frango e banha de porco, que juntos somam 11,6%. Observa-se que os demais óleos têm menor participação na oferta nacional, tal como o óleo de palma e o óleo de algodão. [10]

De acordo com as projeções da ABIOVE, em 2020 o Brasil deverá colher em torno de 105 milhões de toneladas de soja. Processada localmente, o país poderá contar com 20 milhões de toneladas de óleo de soja. Isso sem contar com o desenvolvimento de

outras oleaginosas, tais como girassol e canola, as quais deverão passar por um momento de grande crescimento nos próximos anos, incentivadas pela demanda por óleo do biodiesel. [10]

O programa de biodiesel, ao contrário de prejudicar a produção de alimentos e a agricultura familiar, favorece a produção de farelos essenciais para a produção de carnes, especialmente as de frango e suína. Se, portanto, uma tonelada de biodiesel de soja exige a produção da mesma quantidade de óleo vegetal, pode-se dizer que ela torna disponíveis outras quatro toneladas de farelo de soja para a produção de rações. No caso da produção de frangos, sabe-se, pela composição média das rações utilizada pelos criadores, que essas quatro toneladas serão adicionadas ao milho e outros ingredientes e resultarão em 12 toneladas de carne. Não se quer dizer com isso que o biodiesel leva ao aumento da produção de soja por si só, mas sim que ele pode viabilizar o processamento local da soja atualmente exportada in natura. [12]

No longo prazo, as consequências para o aumento da produção de oleaginosas com maior teor de óleo, tais como a canola e o girassol, serão diversas. Em termos quantitativos, a produção de farelo será menor que a que seria obtida caso o óleo utilizado fosse de soja. No caso do girassol, uma tonelada de óleo é produzida juntamente com 1,1 toneladas de farelo. Na canola, esse valor é de 1,5 toneladas. [12]

Porém, em termos qualitativos, esses farelos serão utilizados em composições diferentes da soja nas rações, pois possuem propriedades com efeitos variados sobre a produção de carnes, sendo mais ou menos adequados a diferentes criações e fases de crescimento. [12]

Outro benefício que será obtido com a diversificação de oleaginosas será o melhor aproveitamento do solo, pois serão plantadas como culturas de inverno, tal como a canola, e em sistema de rotação com outras lavouras, como é o caso do girassol. [12]

A cadeia produtiva de biodiesel atualmente não possui regulamentação quanto à matriz a ser utilizada, sendo regulada atualmente pela disponibilidade de matérias-primas.

1.3.3. Degradação Térmica de óleos comestíveis

A fritura por imersão é um processo que utiliza óleos ou gorduras vegetais como meio de transferência de calor, cuja importância é indiscutível para a produção de alimentos em lanchonetes e restaurantes comerciais ou industriais em nível mundial, bem como no preparo de alimentos nas residências. Em estabelecimentos comerciais, utilizam-se fritadeiras elétricas descontínuas com capacidades que variam de 15 a 350 litros, cuja operação normalmente atinge temperaturas entre 180-200°C. Já em indústrias de produção de empanados, salgadinhos e congêneres, o processo de fritura é normalmente contínuo e a capacidade das fritadeiras pode ultrapassar 1000 litros. O tempo de utilização do óleo varia de um estabelecimento para outro, principalmente pela falta de legislação que determine a troca do óleo usado. [13] Por essa razão, considerando a grande diversidade de estabelecimentos que utilizam esses óleos, é difícil fazer um levantamento preciso da disponibilidade desse resíduo em grandes centros urbanos.

Os destinos mais comuns do óleo usado são produção de sabão, de massa de vidraceiro e de ração animal, mas que também têm parte de seu volume descartado

diretamente no esgoto doméstico. A produção de biodiesel a partir de óleo residual pode dar destino melhor a este óleo.

Vários projetos de reciclagem têm sido bem sucedidos no Brasil; dentre eles destacam-se o aproveitamento de papel, plásticos, metais, óleos lubrificantes automotivos e industriais, soro de leite e bagaço de cana.

A utilização de biodiesel de óleos de fritura em motores do ciclo diesel apresentou bons resultados. Os testes foram realizados em bancada dinâmométrica e em veículo de carga média com motor turbinado a diesel. Por outro lado, a avaliação da emissão de gases demonstrou que houve um aumento relativo na liberação de gases nitrogenados, particularmente quando o biocombustível foi comparado ao diesel convencional. A Tabela 4 apresenta algumas das características físicas e químicas do biodiesel usado nesses testes. Apesar de não atender a uma especificação definida, o biodiesel de óleos de fritura apresentou características bastante semelhantes aos ésteres de óleos "novos". Por outro lado, mesmo sendo um biodiesel de óleo parcialmente oxidado, suas características foram próximas às do óleo diesel convencional (Tabela 2), apresentando inclusive uma boa homogeneidade mediante análise de sua curva de destilação. [14]

Tabela 2 - Especificações do óleo diesel e do biodiesel de óleo residual de soja.

Características	Óleo diesel*	Biodiesel
Densidade 15°C (kg/m ³)	0,849	0,888
Ponto inicial de destilação (°C)	189	307
10%	220	319
20%	234	328
50%	263	333
70%	286	335
80%	299	337
90%	317	340
Ponto final de destilação (°C)	349	342
Aromáticos (% v/v)	31,5	-
Carbono (%)	86,0	77,4
Hidrogênio (%)	13,4	12,0
Oxigênio (%)	0,0	11,2
Enxofre (%)	0,3	0,03
Índice de cetano	46,1	44,6
Número de cetano	46,2	50,8
Valor calórico (MJ/kg)	42,30	37,50

* Óleo diesel com especificação para combustíveis do tipo US-2D (EUA).

Apesar dos excelentes resultados obtidos, é inevitável admitir que o óleo de fritura trouxesse consigo muitas impurezas, oriundas do próprio processo de cocção de alimentos. Portanto, para minimizar esse problema, é sempre aconselhável proceder à pré-purificação e secagem dos óleos antes da reação de transesterificação.

1.4. Substituição do óleo diesel por Biodiesel

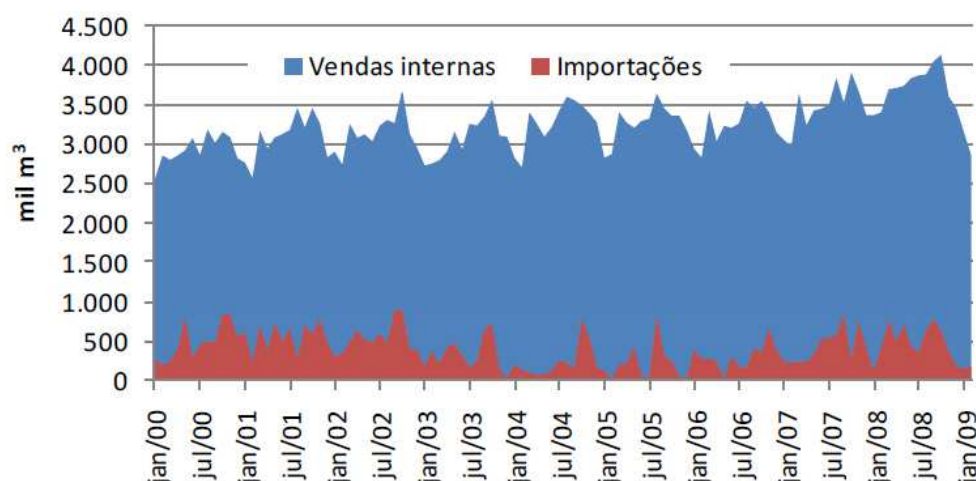
A maior parte da energia consumida no mundo provém do petróleo, do carvão e do gás natural. Essas fontes são limitadas, portanto, fontes alternativas são necessárias e a busca de tais fontes é de suma importância. Neste contexto, os óleos vegetais e gordura

animal aparecem como uma alternativa para substituição ao óleo diesel em motores de ignição por compressão. A possibilidade de emprego de combustíveis de origem agrícola em motores diesel é bastante atrativa tendo em vista o aspecto ambiental, por serem fontes renováveis de energia.

Diversas razões justificam um programa que visa substituir o diesel de petróleo.

O Brasil não tem autossuficiência em diesel. Desde 2000 são importados, em média, 12% do volume consumido. Essa dependência tem custos econômicos e estratégicos, já que o diesel é o combustível que abastece toda a logística rodoviária. [10]

Desta forma, constata-se que o Brasil vem, de fato, substituindo diesel importado, caro e sempre sujeito às intempéries políticas das principais regiões produtoras mundiais. Portanto, o biodiesel diminui a nossa necessidade de importações de diesel. Desde 2005, ano de início do programa brasileiro, mais de 1,5 bilhões de dólares deixaram de serem gastos com a importação de diesel mineral. [10] A Figura 4 apresenta graficamente esta relação entre as vendas/importação de diesel.



(Fonte: ANP. Elaboração: ABIOVE)

Figura 4 - Vendas e importações de diesel no Brasil.

O biodiesel puro de soja reduz as emissões de hidrocarbonetos (HC) em 67%, monóxido de carbono (CO) em 48%, dióxido de carbono (CO₂) em 78%, material particulado em 47% e óxidos de enxofre (SO_x) em 100% (o biodiesel é naturalmente isento de enxofre). [61]

Assim, o biodiesel se encaixa perfeitamente no contexto atual de procura de combustíveis alternativos que sejam mais limpos, reduzindo as fontes de gases do Efeito Estufa.

A utilização de biodiesel no transporte rodoviário pesado oferece grandes vantagens para o meio ambiente, principalmente em grandes centros urbanos, tendo em vista que a emissão de poluentes é menor que a do óleo diesel. [7] As emissões de monóxido e dióxido de carbono, enxofre e material particulado foram inferiores às do diesel convencional. No entanto, os níveis de emissões de gases nitrogenados foram maiores para diferentes tipos de biodiesel. [23]

Foi testada na frota de transporte coletivo da cidade de Curitiba a utilização de biodiesel de óleo de soja, doado pela *American Soybean Association*. O biodiesel foi misturado ao diesel convencional na proporção de 20% (v/v), com o propósito de verificar a eficiência desse combustível na redução da poluição ambiental. Os testes foram realizados em 20 ônibus de diferentes marcas durante três meses consecutivos e, ao final dos trabalhos, apresentaram redução média de fumaça em torno de 35%. [15]

1.5. Parâmetros de Qualidade do Biodiesel

A Resolução ANP Nº 04, de 02/02/2010 (DOU 03/02/2010) estabelece a especificação do biodiesel a ser comercializado pelos diversos agentes econômicos autorizados em todo o território nacional. [16]

O Regulamento Técnico ANP anexo à Resolução ANP Nº 04 aplica-se ao biodiesel, de origem nacional ou importada, a ser comercializado em território nacional adicionado na proporção prevista na legislação aplicável ao óleo diesel conforme a especificação em vigor, e em misturas específicas autorizadas pela ANP; dispendo as seguintes normas aplicáveis:

- A determinação das características do biodiesel será feita mediante o emprego das normas da Associação Brasileira de Normas Técnicas (ABNT), das normas internacionais *American Society for Testing and Materials* (ASTM), da *International Organization for Standardization* (ISO) e do *Comité Européen de Normalisation* (CEN);
- Os dados de incerteza, repetitividade e reprodutibilidade fornecidos nos métodos relacionados neste Regulamento devem ser usados somente como guia para aceitação das determinações em duplicata do ensaio e não devem ser considerados como tolerância aplicada aos limites especificados neste Regulamento.
- A análise do produto deverá ser realizada em uma amostra representativa do mesmo, obtida segundo métodos ABNT NBR 14883 – Petróleo e produtos de petróleo – Amostragem manual ou ASTM D 4057 – Prática para Amostragem de Petróleo e Produtos Líquidos de Petróleo (Practice for Manual Sampling of Petroleum and Petroleum Products) ou ISO 5555 (Animal and vegetable fats and oils – Sampling).

A quantificação do éster no biodiesel puro é feita de acordo com as normas ABNT 15342 (1) (2) e EN ISO 14103. [3]

CATÍTULO 2: Espectroscopia Vibracional

2. Espectroscopia Vibracional

A espectroscopia compreende o estudo da interação da radiação eletromagnética com a matéria, tendo como um de seus principais objetivos a determinação dos níveis de energia e transições de espécies atômicas e moleculares.

A espectroscopia vibracional divide-se em duas técnicas distintas, a espectroscopia de absorção no Infravermelho e a espectroscopia Raman, que geram informações na região do Infravermelho. [17-18]

2.1. Aplicações e usos da espectroscopia no Infravermelho

A região espectral do Infravermelho está situada antes da região do visível no espectro eletromagnético e abrange a radiação com números de onda no intervalo de aproximadamente $12\ 800$ a $10\ \text{cm}^{-1}$ ou comprimento de onda de 780 a $100\ 000\ \text{nm}$. Do ponto de vista da aplicação e instrumentação, o espectro Infravermelho é dividido em radiação no Infravermelho próximo (NIR), médio (MIR) e distante (FIR). [17-18] A Tabela 3 apresenta os limites aproximados para cada região.

Tabela 3 - Regiões do Infravermelho.

Região	Intervalo de número de onda (ν), cm^{-1}	Região em comprimento de onda (λ), nm	Região de frequência (ν), Hz
Próximo (NIR)	12 800 a 4 000	780 a 2 500	$3,8 \times 10^{14}$ a $1,2 \times 10^{14}$
Médio (MIR)	4 000 a 200	2 500 a 5 000	$1,2 \times 10^{14}$ a $6,0 \times 10^{12}$
Distante (FIR)	200 a 10	5 000 a 100 000	$6,0 \times 10^{12}$ a $3,0 \times 10^{11}$

Fonte: Skoog, D. A.; Holler, F. J.; Nieman, T. A. Princípios de análise instrumental. [17]

A região do MIR é provavelmente onde se encontra a maioria das pesquisas desenvolvidas e o maior número de aplicações. Ainda hoje, a maioria das aplicações consiste na identificação de compostos orgânicos, pois nessa região ocorrem essencialmente transições fundamentais e existe uma faixa espectral conhecida como região de impressão digital ($1\ 200$ a $700\ \text{cm}^{-1}$). Nessa região, pequenas diferenças na estrutura e na constituição de uma molécula resultam em mudanças significativas na distribuição das bandas de absorção. Em consequência, uma semelhança estreita entre dois espectros nesta região, bem como nas outras, constitui forte evidência da identidade dos compostos que produziram os espectros. [17, 21]

Para as regiões do Infravermelho, em geral, é possível realizar medidas de amostras em todos os estados e formas como, gases, líquidos, sólidos, sistemas binários e terciários como as amostras semissólidas, pastas, géis e outras. [17-19, 28]

Nos últimos anos, muitos fabricantes de instrumentos ofereceram adaptadores para os compartimentos de amostra dos instrumentos de espectroscopia no Infravermelho e que tornam possível a obtenção imediata de espectros de reflexão. [17-19, 21-22]

É cada vez maior o interesse e a necessidade de vários grupos de pesquisas acadêmicos, governamentais e industriais, de desenvolver métodos sensíveis, rápidos e altamente específicos para uma variedade de compostos químicos. A espectroscopia vibracional parece ser apropriada a essa necessidade mais que qualquer outra ferramenta analítica. As técnicas vibracionais tanto no Infravermelho quanto no Raman, além de

serem técnicas não destrutivas, não geram subprodutos tóxicos, em certos casos não requerem preparação de amostras, não usam reagentes e são adequadas para processos “on-line”. Dessa forma, a espectroscopia vibracional oferece oportunidades analíticas quase que ilimitadas para muitas áreas de produção e controle de qualidade. [17]

O uso comum de tecnologia existe nas três técnicas, MIR, NIR e Raman, e frequente sobreposição significativa acontece nos conceitos de projeto. Devido a diferenças em exigências de mercado e uso, os conceitos de instrumentos são vistos diferentemente e, em alguns casos, duas ou mais das técnicas podem competir de fato para a mesma aplicação.

Em geral, na obtenção de informações qualitativas e quantitativas a partir dos complexos espectros vibracionais, por tratar de dados multivariados, é imprescindível a utilização de Métodos Quimiométricos. Tais métodos são desenvolvidos e disponibilizados em programas computacionais, e são, juntamente com os avanços tecnológicos dos instrumentos, os responsáveis pela popularização do uso da espectroscopia vibracional.

Diversos trabalhos têm sido feitos utilizando Infravermelho, tais como: Adulteração de misturas biodiesel/diesel por óleos vegetais determinadas por FTNIR [23], Aplicação de espectroscopias Raman e Infravermelho na identificação e quantificação de plastificantes em filmes comerciais de PVC estirável [24], Determinação de ácido oleico em sementes de girassol por espectroscopia no Infravermelho e calibração multivariada [25], Determinação com FTIR de ácidos graxos livres em óleos de peixes destinados à produção de biodiesel [26], Aplicação de MIR e iPLS para a quantificação de contaminantes em óleos lubrificantes [27].

2.1.1. Espectroscopia de Refletância Total Atenuada (ATR)

A reflexão de radiação pode ser de quatro tipos: especular, difusa, interna e total atenuada (ATR). Neste trabalho o foco será apenas a ATR, que é uma técnica não destrutiva que permite obter espectros no Infravermelho de amostras como líquidos, sólidos insolúveis, filmes, adesivos, pós, pastas, entre outros. [17, 29]

A espectroscopia de ATR baseia-se no fato de que quando um feixe de radiação passa de um meio mais denso (cristal de ATR) para um menos denso (amostra) ocorre uma reflexão. A fração do feixe incidente que é refletida aumenta com o ângulo de incidência; além de certo ângulo crítico, a reflexão é total.

No processo de reflexão, o feixe se comporta como se penetrasse um pouco no meio menos denso antes que a reflexão ocorra. A profundidade de penetração, que varia de uma fração até vários comprimentos de onda, depende do comprimento de onda da radiação incidente, dos índices de refração dos dois materiais e do ângulo do feixe incidente em relação à interface.

A radiação penetrante é chamada onda evanescente. Se o meio menos denso absorve essa onda evanescente, ocorre atenuação do feixe nos comprimentos de onda das bandas de absorção. Esse fenômeno é conhecido como refletância total atenuada (ATR), mostrado na Figura 5. [17, 29]

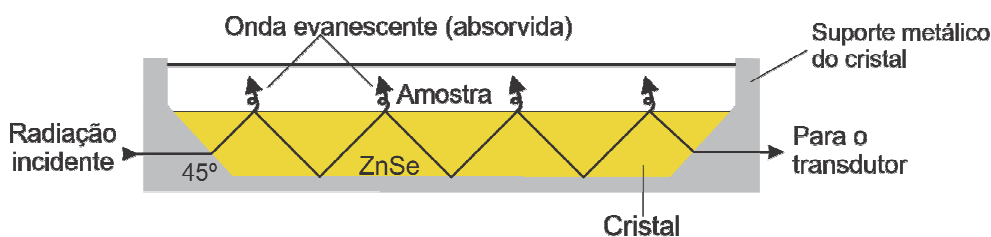


Figura 5 - Representação do ATR.

Os espectros de ATR são semelhantes, mas não iguais aos espectros comuns de absorção. Em geral, enquanto as mesmas bandas são observadas, suas intensidades relativas diferem. As absorvâncias, embora dependam do ângulo de incidência, são independentes da espessura da amostra, uma vez que a radiação penetra apenas alguns micrômetros na mesma. [17]

Uma das maiores vantagens da espectroscopia de ATR é que os espectros de absorção são obtidos rapidamente em uma grande variedade de tipos de amostras com um mínimo de preparação. Assim, linhas, fios, tecidos, fibras, pós, pastas, suspensões, polímeros, borrachas e outros materiais podem ser analisados pressionando as amostras sobre o cristal ATR. Para amostras líquidas, o cristal ATR pode ser mergulhado no líquido. Soluções aquosas podem também ser usadas, desde que o cristal não seja solúvel em água. [17]

2.2. Interpretação de Espectros

Não há regras rígidas para interpretar um espectro de Infravermelho. Certos requerimentos, entretanto, devem existir antes que se tente interpretar um espectro: [28]

- a) O espectro deve possuir resolução e intensidade adequados;
- b) Tratamento preciso das vibrações de moléculas complexas é inviável;
- c) Espectros de Infravermelho devem ser interpretados por comparações empíricas de espectros e extrapolações de estudos de moléculas mais simples;
- d) Muitas das absorções de grupos variam em uma grande faixa porque as bandas surgem de complexas interações vibracionais da molécula.

Duas áreas importantes para exames preliminares de espectros são as regiões de $4\ 000 - 1\ 300$ e $900 - 650\ \text{cm}^{-1}$. A parte de alta frequência do espectro é chamada de região de grupo funcional. Os estiramentos característicos de importantes grupos como O-H, N-H e C=O ocorrem nesta parte do espectro. A ausência de absorção característica destes grupos pode ser comumente usada como evidência de ausência destes grupos na molécula. A interpretação deve ser cuidadosa, pois bandas muito largas podem encobrir bandas menores, como é o caso da banda de OH em um composto que possua equilíbrio ceto-enólico que pode encobrir a banda da carbonila da cetona na faixa de $1\ 850$ a $1\ 540\ \text{cm}^{-1}$, levando à falsa evidência de ausência de grupo carbonila. [28]

Bandas fracas na região de alta frequência são resultantes da absorção fundamental de grupos funcionais como S-H e $\text{C}\equiv\text{C}$; muito valiosos na determinação da estrutura. Tais bandas fracas têm pouco valor em regiões mais complicadas do espectro. Sobreposição de bandas e combinações de bandas de bandas de baixa frequência comumente aparecem em regiões de alta frequência do espectro. Sobreposição de

bandas e combinações de bandas são caracteristicamente fracas, exceto quando ocorre ressonância de Fermi. [28]

A falta de fortes bandas de absorção na região de $900 - 650 \text{ cm}^{-1}$ geralmente indica estrutura não aromática. A parte intermediária do espectro $1300 - 900 \text{ cm}^{-1}$ é conhecida como região de “impressão digital”. O padrão de absorção nesta região é normalmente complexo, com bandas originadas de interações entre os modos vibracionais. Esta parte do espectro é extremamente valiosa quando examinada em comparação com outras regiões. A absorção nesta região intermediária é provavelmente única para cada molécula. [28]

Qualquer conclusão obtida após o exame de uma banda particular deve ser confirmada quando possível pelo exame de outras partes do espectro. Por exemplo, a conclusão da existência de grupo carbonila de éster deve ser confirmada pela observação de uma forte banda de estiramento da ligação C-O na região de $1300 - 1100 \text{ cm}^{-1}$. [28]

Compostos similares podem gerar espectros virtualmente idênticos em condições normais. [28] Diferenças muito sutis ocorrem, sendo sua visualização muito difícil.

O uso do espectro Infravermelho para caracterização de compostos/grupos orgânicos constitui um passo importante em determinações de estruturas, como apresentado nos trabalhos: Investigações espectroscópicas de oxalatos de ferro (II) e ferro (III) [30], Análise térmica com FTIR de sílica gel organofuncionalizada [31], Síntese e caracterização de complexos de EDTA úteis como elementos-traços de suplementação [32].

2.2.1. Absorção característica de grupos de moléculas orgânicas

A absorção de estiramento da ligação do grupo carbonila, C=O, ocorre em uma ampla faixa de número de ondas, que vai de aproximadamente 1850 a 1610 cm^{-1} , dependendo da classe e da estrutura do composto em particular, como pode ser visto na Figura 6 (em destaque a faixa da carbonila de éster, presente no biodiesel). Todo efeito que resulte em aumento da força da ligação C=O terá como consequência incremento na frequência de estiramento desta ligação, enquanto efeitos que diminuam a força da ligação resultarão em decréscimo no número de ondas de absorção. [33]

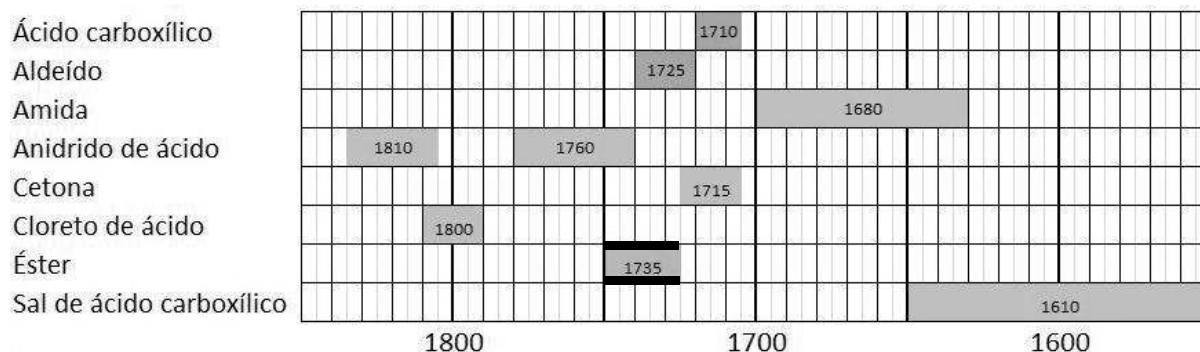


Figura 6 - Bandas de absorção de estiramento da carbonila.

Ésteres possuem duas bandas de absorção caracteristicamente fortes do estiramento das ligações C=O e C-O. A intensa vibração de estiramento da ligação C=O ocorre em altas frequências (baixos comprimentos de onda) do que em cetonas normais.

A constante de força do grupo carbonila é aumentada pela natureza de atração de elétrons do oxigênio adjacente (efeito indutivo). A absorção ocorre na região de $1750 - 1735 \text{ cm}^{-1}$. Conjugação tem pouco ou nenhum efeito na frequência de absorção da carbonila.

A vibração de estiramento da ligação C-O dos ésteres consiste em duas vibrações assimétricas acopladas: C-C(=O)-O e O-C-C, sendo a primeira mais importante. Estas bandas ocorrem na região de $1300 - 1100 \text{ cm}^{-1}$. A banda de C-C(=O)-O mostra-se com maior intensidade em $1210 - 1163 \text{ cm}^{-1}$; comumente mais larga e mais intensa do que a banda de estiramento de C=O. Se o éster for derivado de ácidos α,β -insaturados, ocorrem várias bandas na região de $1300 - 1160 \text{ cm}^{-1}$. Banda O-C-C de ésteres de álcoois primários ocorre em $1164 - 1031 \text{ cm}^{-1}$, e em 1100 cm^{-1} para ésteres formados de álcoois secundários. Ésteres metílicos de ácidos graxos de longas cadeias apresentam um padrão de três bandas próximas a 1250 , 1205 e 1175 cm^{-1} (sendo mais forte a banda em 1175 cm^{-1}). [28]

A longa cadeia carbônica derivada dos ácidos graxos gera bandas de absorção relativas aos grupos metil, metileno, C=C, C=C-C=C. As ligações C-C geram banda de absorção de torção na região de 500 cm^{-1} (fora do espectro); geram, também, bandas de estiramento fracas na faixa de $1200 - 800 \text{ cm}^{-1}$.

As ligações C-H geram bandas de estiramento na região de $3000 - 2840 \text{ cm}^{-1}$. Para o grupo metil (CH_3), temos a banda de estiramento assimétrico em 2962 cm^{-1} , estiramento simétrico em 2872 cm^{-1} , torção simétrica (muito estável) em 1375 cm^{-1} , torção assimétrica em 1450 cm^{-1} , tesoura em $1439 - 1399 \text{ cm}^{-1}$. Para o grupo metileno (CH_2) temos banda de estiramento assimétrico em 2926 cm^{-1} , de estiramento simétrico em 2853 cm^{-1} , de tesoura em 1465 cm^{-1} , de balanço em 720 cm^{-1} e bandas fracas de torção e balanço em $1350 - 1150 \text{ cm}^{-1}$. Temos ainda, banda em $1667 - 1640$ devido à ligação C=C sem conjugação e duas bandas em 1650 e 1600 (ou somente em 1600) para o grupo C=C-C=C. [28]

2.2.2. Caracterização de grupos funcionais nos espectros

O diesel brasileiro é constituído por alcanos, assim, possui grandes cadeias carbônicas com ligações carbono-hidrogênio; são estas mesmas ligações C-H presentes em grande quantidade nos biodieseis devido a grande cadeia carbônica derivada dos ácidos graxos que constituem os óleos utilizados na fabricação de biodiesel. Esta semelhança estrutural faz com que o espectro do diesel puro (Figura 7) tenha absorvâncias semelhantes em algumas regiões ao ser comparado com espectro do biodiesel puro (Figura 8).

No entanto, quando se observam espectros de misturas binárias de biodiesel e diesel de concentrações variando de 1 a 20%; tem-se que os espectros tornam-se mais semelhantes ao espectro do biodiesel puro; como pode ser visto na Figura 9.

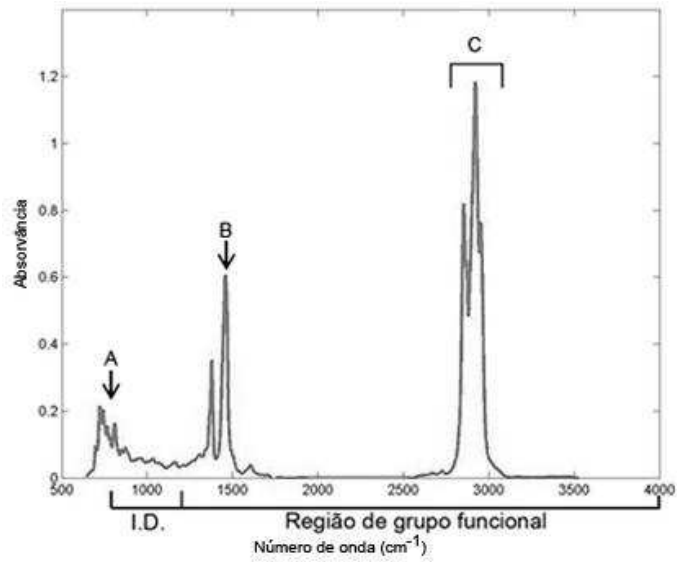


Figura 7 - Espectro do diesel puro.

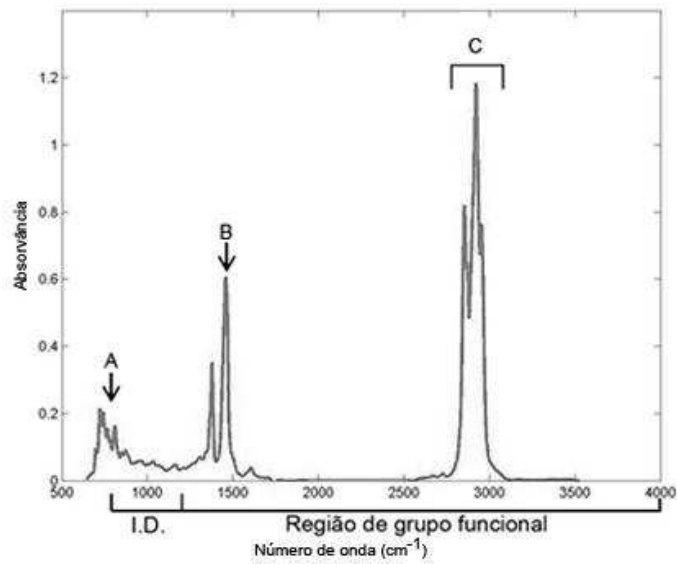


Figura 8 - Espectro do biodiesel puro.

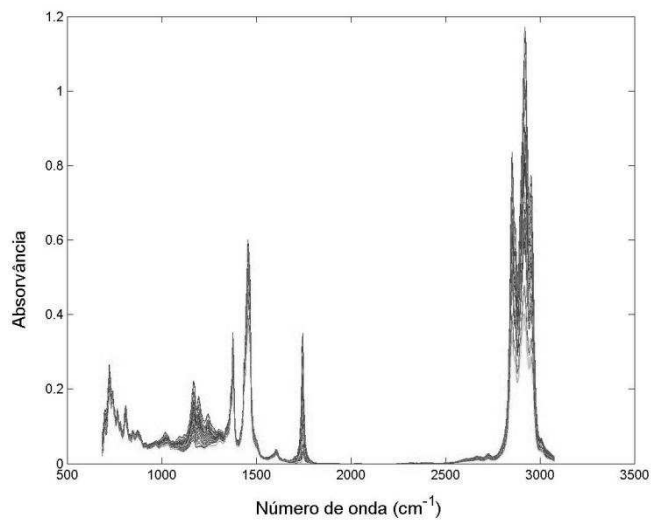


Figura 9 - Espectros de misturas binárias biodiesel/diesel de concentrações de 1 a 20% v/v.

Dentre as regiões do espectro na faixa de varredura, algumas são mais importantes, dando indícios das estruturas dos componentes que geraram os espectros. Estas regiões estão assinaladas nas Figuras 7 e 8; são elas:

I.D.: região de Impressão Digital. A parte intermediária do espectro $1\ 300 - 900\ \text{cm}^{-1}$ é conhecida como região de “impressão digital”. O padrão de absorção nesta região é normalmente complexo, com bandas originadas de interações entre os modos vibracionais. Esta parte do espectro é extremamente valiosa quando examinada em comparação com outras regiões. A absorção nesta região intermediária é provavelmente única para cada molécula.

A: ausência de fortes bandas de absorção na região de 900 a $650\ \text{cm}^{-1}$ geralmente indica estrutura não aromática, evidenciando que o diesel brasileiro é composto por hidrocarbonetos não aromáticos;

B: bandas de torção simétrica (muito estável) em $1\ 375\ \text{cm}^{-1}$, torção assimétrica em $1\ 450\ \text{cm}^{-1}$, tesoura em $1\ 439 - 1\ 399\ \text{cm}^{-1}$ para grupo metil ($-\text{CH}_3$); banda de tesoura para $-\text{CH}_2$.

C: ligações C-H geram bandas de estiramento na região de $3\ 000 - 2\ 840\ \text{cm}^{-1}$. Banda de estiramento assimétrico em $2\ 962\ \text{cm}^{-1}$, estiramento simétrico em $2\ 872\ \text{cm}^{-1}$ para grupo metil (CH_3). Banda de estiramento assimétrico em $2\ 926\ \text{cm}^{-1}$ para grupo metileno (CH_2).

D: Bandas fracas de torção e balanço em $1\ 350 - 1\ 150\ \text{cm}^{-1}$ para $-\text{CH}_2$. Forte banda de estiramento da ligação C-O na região de $1\ 300 - 1\ 100\ \text{cm}^{-1}$ (*). Banda C-C(=O)-O em $1\ 210 - 1\ 163\ \text{cm}^{-1}$. Banda O-C-C de ésteres ocorre em $1\ 164 - 1\ 031\ \text{cm}^{-1}$. Ésteres metílicos de ácidos graxos de longas cadeias apresentam um padrão de três bandas próximas a $1\ 250$, $1\ 205$ e $1\ 175\ \text{cm}^{-1}$ (sendo mais forte a banda em $1\ 175\ \text{cm}^{-1}$).

E: Banda carbonila de éster em $1\ 735\ \text{cm}^{-1}$.

F: mesma característica da região assinalada com letra **C** na Figura 6, pois o biodiesel possui grandes cadeias carbônicas com várias ligações carbono-hidrogênio; as quais geram absorvância nesta região.

CAPÍTULO 3: Métodos Quimiométricos

3. Quimiometria

A Quimiometria pode ser definida como a aplicação de métodos matemáticos, estatísticos e de lógica formal para o tratamento de dados químicos, de forma a extrair uma maior quantidade de informações e melhores resultados analíticos. A princípio, esses métodos foram desenvolvidos e empregados em outras áreas do conhecimento, enquanto que as soluções dos problemas em Química Analítica eram obtidas através de métodos tradicionais de análises por via úmida como, titulação, precipitação, reações específicas, extração, etc. Esses métodos tradicionais, ainda muito importantes, são em alguns casos, considerados lentos e dispendiosos. [34-36]

A Quimiometria enfoca o estudo de medidas analíticas baseando-se na ideia da observação indireta, relacionando essas medidas à composição química de uma substância e deduzindo o valor de uma propriedade de interesse através de alguma relação matemática. Devido à sofisticação das técnicas analíticas, novos métodos Quimiométricos são desenvolvidos até os dias de hoje, com o objetivo de resolver problemas de análise de dados multivariados. Conforme o objetivo do estudo, a Quimiometria pode ser dividida em algumas áreas muito aplicadas atualmente como: processamento de sinais analíticos, planejamento e otimização de experimentos, reconhecimento de padrões e classificação de dados, calibração multivariada, seleção de variáveis, monitoramento e modelagem de processos multivariados, métodos de inteligência artificial, tratamento de imagens, dentre outras. A principal linha de pesquisa da Quimiometria aplicada à Química Analítica tem sido a construção de modelos de regressão a partir de dados de primeira ordem, ou seja, dados que podem ser representados através de um vetor para cada amostra, sendo a construção desses modelos denominada de calibração multivariada. [34-35, 37-40]

3.1. Organização dos Dados para o programa

A resposta instrumental, para a construção dos modelos, é representada na forma de matriz; enquanto a propriedade de interesse, determinada por uma metodologia padrão, é representada por um vetor. [41] A Figura 10 ilustra como uma matriz de dados **X** de dimensão $n \times m$, ou seja, n objetos (espectros) e m variáveis (números de onda) que pode ser construída a partir de um vetor de respostas instrumental.

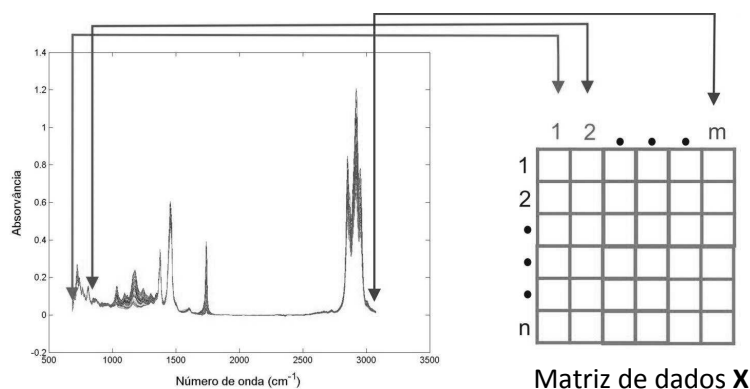


Figura 10 - Representação da construção da matriz de dados **X** para calibração multivariada.

É importante salientar que, para a montagem da matriz \mathbf{X} , o programa não lê o espectro, mas sim um arquivo com números que formam o espectro; ou seja, ao serem obtidos os espectros, estes devem ser exportados num formato numérico para que possa ser lido pelo programa (no caso deste trabalho, arquivos texto padrão ASCII).

A utilização dos espectros pelo programa de quantificação é semelhante à construção da curva de calibração em métodos univariados para relacionamento com a propriedade de interesse.

3.2. Pré-Tratamento de Dados

O objetivo é remover matematicamente fontes de variação indesejáveis que não serão removidas naturalmente durante a análise dos dados e que podem influenciar os resultados finais, pois os sinais medidos consistem de: sinal verdadeiro + ruído aleatório.

Em geral, as matrizes de dados de natureza multivariada necessitam de manipulações matemáticas que possibilitem contornar ou reduzir fontes de variação não informativas para futuras modelagens. Estas manipulações são conhecidas como pré-processamento (nas variáveis – matriz \mathbf{X}) ou transformação (nas amostras – matriz com propriedade de interesse para PLS).

Em medidas baseadas em espectros MIR, geralmente é possível observar ruído aleatório e variação da linha de base (Figura 11a). O método mais simples para superar tais variações não informativas é a Correção da Linha de Base (*Baseline*). Sendo o conjunto de espectros com a linha de base corrigida mostrados na Figura 11b.

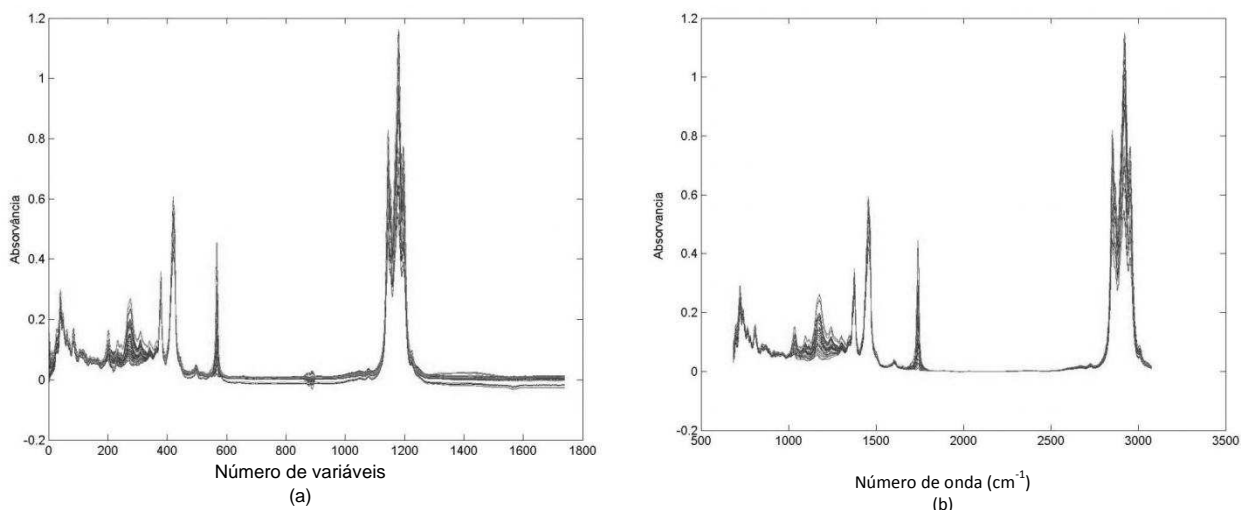


Figura 11 - Espectro MIR com flutuação na linha de base (a) Espectro MIR corrigido e cortado (b).

3.3. Análise Multivariada

Análise de dados obtidos por experimentação ou observação irá, comumente, sugerir uma explicação modificada do fenômeno. De forma geral, neste processo iterativo de aprendizado, variáveis são comumente adicionadas ou retiradas do conjunto em estudo. Assim, as complexidades da maioria dos fenômenos requerem que se façam observações em muitas variáveis diferentes. [42]

A necessidade de entender as relações entre muitas variáveis faz da análise multivariada um assunto inerentemente difícil. Comumente, a mente humana é oprimida pela absoluta importância dos dados. Além disso, a matemática necessária para fazer inferências de técnicas multivariadas é mais complexa do que em técnicas univariadas. [42]

Os objetivos de investigações científicas que naturalmente levam a métodos multivariados são: [42]

1. Redução de dados ou Simplificação estrutural;
2. Triagem e Agrupamento;
3. Investigação da dependência entre variáveis;
4. Previsão;
5. Construção e teste de hipóteses.

3.4. Calibração Multivariada

A calibração pode ser definida como uma série de operações que estabelecem, sob condições específicas, uma relação entre medidas instrumentais e valores para uma propriedade de interesse correspondente. [43]

Tal método faz uso de Regressão Linear Múltipla para resolver o problema de prever uma variável Y de n variáveis X_1, X_2, \dots, X_n ; $n > 1$. Tradicionalmente, a variável Y é chamada *variável dependente*, enquanto as variáveis X_1, \dots, X_n são chamadas *variáveis independentes*. No processo de calibração multivariada baseado em dados espectrais temos a construção de um modelo matemático que relaciona as medidas de absorvância ao longo dos espectros com os parâmetros de interesse (concentrações de analitos, acidez, viscosidade, índice de refração, etc.) nas amostras; sendo que, no caso destes modelos, relacionam-se os valores de absorvância com concentração do biodiesel em misturas com diesel. [44]

Um modelo de calibração, na verdade, é uma função matemática (f) que relaciona dois grupos de variáveis; isto é, relacionam-se os valores de Y por uma função de regressão $f()$:

$$Y = f(X) = Xb \quad (1)$$

Esta etapa representa a calibração e por isso o conjunto de dados empregado para essa finalidade é chamado de conjunto de calibração. Os parâmetros do modelo são denominados de coeficientes de regressão (b) determinados matematicamente a partir dos dados experimentais. [41, 45]

O passo seguinte à calibração é a previsão. Nesta etapa, as variáveis independentes obtidas, para outro conjunto de amostras, são utilizadas em conjunto com o coeficiente de regressão, para calcular os valores previstos para a variável dependente. No conjunto de previsão utilizam-se amostras cujas variáveis dependentes sejam conhecidas para que seja possível estabelecer uma comparação entre os valores previstos pelo modelo e os valores conhecidos previamente através de uma metodologia padrão, o que permitirá a avaliação sobre o desempenho do modelo de calibração proposto. [41]

Existem diversos métodos para a construção de modelos de calibração, sendo que a função que ajusta as variáveis dependentes e independentes pode ser linear ou não, dependendo da complexidade do sistema em estudo. [46]

Dentre os métodos de calibração existentes, sem dúvida, os mais difundidos são ainda os métodos de calibração univariada que também são conhecidos como calibração de ordem zero, ou seja, tem-se apenas uma medida instrumental para cada uma das amostras de calibração, isto é, para cada amostra tem-se apenas um escalar. No entanto, a aplicação da calibração univariada é restrita, visto que, quando a amostra não é livre de interferentes e a medida é realizada diretamente por uma metodologia instrumental, isso provavelmente provocará desvios na determinação da propriedade de interesse e a aplicação deste método de calibração torna-se inviável. [47]

Em calibração multivariada, mais de uma resposta instrumental é relacionada com a propriedade de interesse. Esses métodos de calibração possibilitam a análise mesmo na presença de interferentes, desde que esses interferentes estejam presentes nas amostras utilizadas para a construção do modelo de calibração. Outras possibilidades apresentadas por esse tipo de calibração são determinações simultâneas e análises mesmo sem resolução. Isso faz com que os modelos de calibração multivariada sejam uma alternativa quando os métodos univariados não são aplicáveis.

Devido à presença de mais de uma variável utilizada para previsão, usa-se o termo multivariado para o processo.

Tem sido possível através de Métodos Quimiométricos, mais especificamente, métodos de calibração multivariada, a determinação de propriedades químicas e físicas de amostras ativas na região do Infravermelho e Raman. Isso tem sido feito por métodos com modelagem local, como a Regressão Linear Múltipla (MLR) [38-39], que utiliza a absorvância em comprimentos de onda selecionados, ou métodos de modelagem global, tais como a Regressão por Componentes Principais (PCR) ou Quadrados Mínimos Parciais (PLS). [38, 48]

A Calibração multivariada tem sido usada em diversas áreas da Química, sendo citados os trabalhos: Alternativas quimiométricas para a resolução de problemas analíticos clássicos: Determinação espectrofotométrica de misturas de zircônio e háfnio [49], Comparação de algoritmos PLS no monitoramento de parâmetros de gasolina e gasóleo com MIR e NIR [50], Determinação dos níveis de biodiesel misturados em diferentes amostras de diesel com RMN [51], Determinação de açúcar total, glicose, frutose e sacarose em soluções aquosas de sucos de frutas por PLS-NIR [52].

Como o Método de Calibração por Quadrados Mínimos Parciais faz uso da Análise de Componentes Principais (PCA), segue abaixo descrição do PCA.

3.4.1. Análise de Componentes Principais - PCA

Introduzida na Química por Malinowski no final dos anos 1960, com o nome de Análise de Fatores, tendo a partir da década seguinte uma série de aplicações, foi desenvolvida, tornando-a muito conhecida e explorada. [38-39]

O objetivo da Análise de Componentes Principais é a redução do número de variáveis, pela troca das variáveis originais por outras chamadas Componentes Principais.

O primeiro passo para a Análise de Componentes Principais é a formação de uma matriz de variância/covariância dos dados (\mathbf{Z}) que irá isolar a fonte de variação dos dados:

$$\mathbf{Z} = \mathbf{X}^T \mathbf{X} \quad (2)$$

A matriz \mathbf{Z} deve ser simétrica, ou seja, $\mathbf{Z} = \mathbf{Z}^T$, caso contrário não há esperança de diagonalizar essa matriz. Como qualquer matriz simétrica é ortogonalmente diagonalizável, então por uma transformação unitária tem-se:

$$\mathbf{\Lambda} = \mathbf{P}^{-1} \mathbf{Z} \mathbf{P} = \mathbf{P}^T \mathbf{Z} \mathbf{P} \quad (3)$$

em que $\mathbf{\Lambda}$ é uma matriz diagonal cujos elementos são autovalores de \mathbf{Z} , \mathbf{P} é a matriz de autovetores, denominada loadings (pesos). Para realizar a ortogonalização necessitamos de um resultado essencial sobre matrizes simétricas:

- 1) se a matriz é simétrica os autovalores são reais;
- 2) se a matriz é simétrica os autovalores de autoespaços diferentes são ortogonais.

Basicamente, os loadings formam uma nova base ortonormal que explica dos dados de \mathbf{X} e a projeção dos dados nessa base é denominada scores (escores), (\mathbf{T}). Desse modo, os dados são decompostos por um conjunto de vetores pesos e escores denominado componente principal (PC):

$$\mathbf{X} = \mathbf{T} \mathbf{P}^T \quad (4)$$

Como a matriz \mathbf{P}^T é ortonormal, $\mathbf{P}^T \mathbf{P} = \mathbf{I}$ (matriz identidade), portanto:

$$\mathbf{P} \mathbf{X} = \mathbf{T} \mathbf{P}^T \mathbf{P} \quad (5)$$

$$\mathbf{T} = \mathbf{X} \mathbf{P} \quad (6)$$

Geometricamente, pode-se representar a decomposição da matriz de dados de alta dimensão, com o cálculo de novas variáveis a partir das variáveis originais por:

a) Componente Principal Zero:

O modelo mais simples para representar o conjunto de dados é utilizar seu ponto médio, x_m (Figura 12), aonde as coordenadas são definidas pelas médias das variáveis.

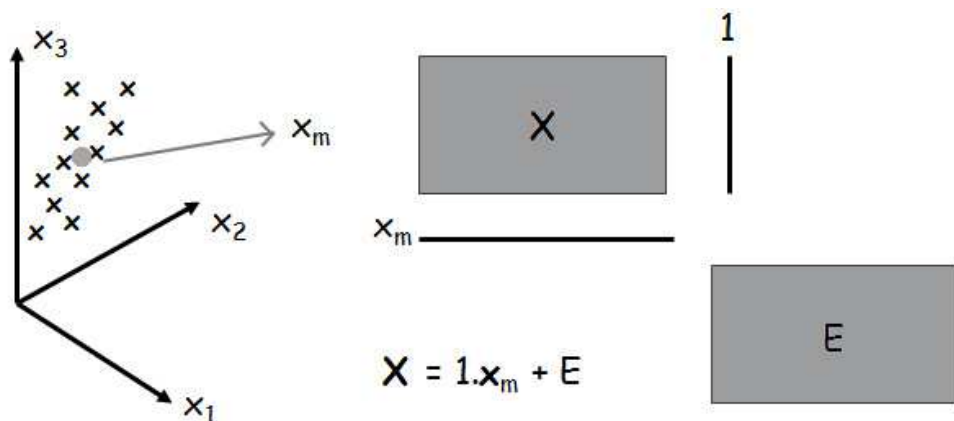


Figura 12 - Formação da Componente Principal Zero.

b) Componente Principal Um (PC1)

Ajusta-se a posição de uma reta (utilizando mínimos quadrados) passando por x_m (Figura 13); de forma a capturar a maior quantidade possível de informação dos dados originais, ou seja, ajusta-se a reta de forma que se obtenha o menor erro possível para a representação. A projeção de cada ponto do conjunto de dados originais nesta reta formam os escores. Essa reta é a Primeira Componente Principal (PC1).

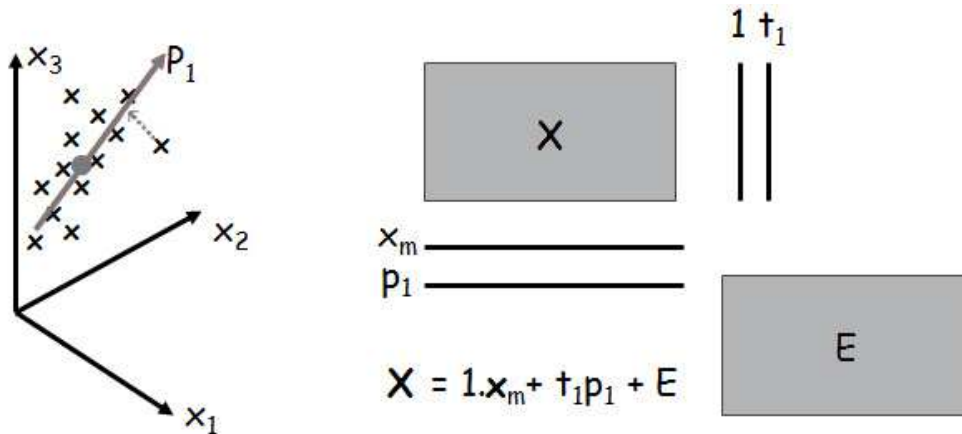


Figura 13 - Formação da Componente Principal Um.

c) Componente Principal Dois (PC2):

É a reta que passa por x_m e é perpendicular à PC1 (Figura14). Desta forma, ela pode capturar informações do conjunto de dados pela projeção dos dados em outra posição diferente de PC1.

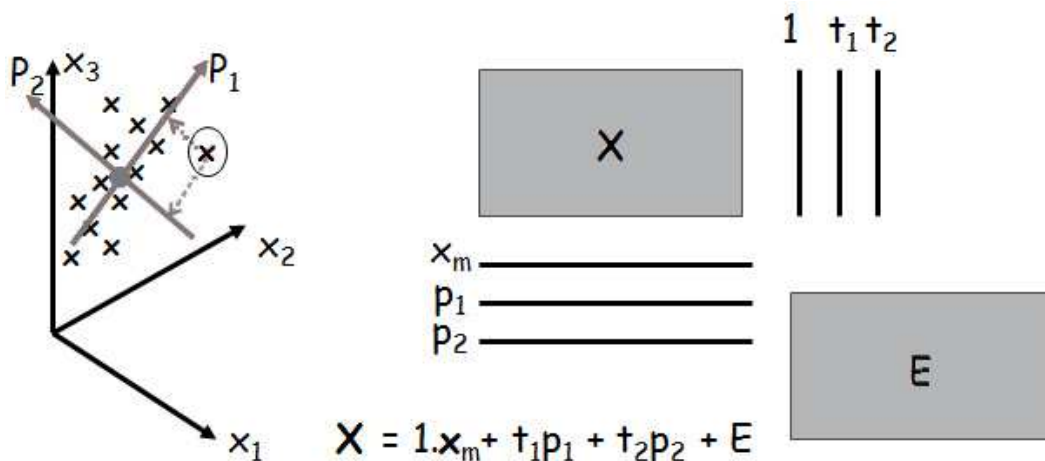


Figura 14 - Formação da Componente Principal Dois.

d) Demais Componentes Principais:

As demais retas são ajustadas passando por x_m e sempre perpendiculares às outras componentes, como estão em planos diferentes, dizemos que são ortogonais entre si (o termo perpendicular aplica-se somente ao mesmo plano de projeção).

A Figura 15 ilustra a decomposição da matriz X de dimensão $(n \times m)$ pela Análise de Componentes Principais até A componentes principais.

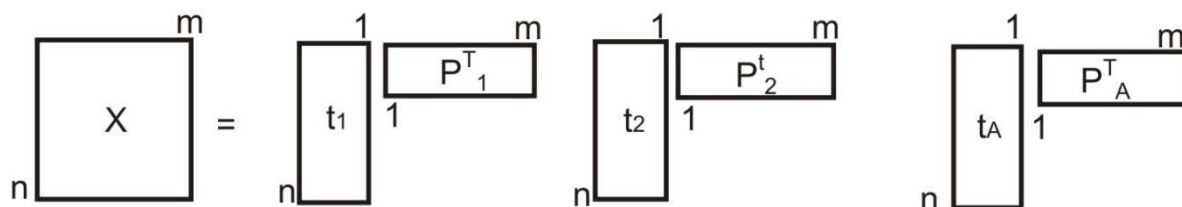


Figura 15 - Decomposição em componentes principais por PCA.

O número máximo de componentes principais (PCs) obtidos é igual ao número de vetores de dados utilizados (posto da matriz X de dados independentes), sendo que nem todas as PCs possuem informações úteis.

Normalmente, as últimas PCs modelam ruído inerente aos dados, sendo assim, a eliminação dessas PCs frequentemente aumenta a relação sinal/ruído; além do ruído aleatório nunca está correlacionado, de tal forma que ele fica na matriz de resíduos. [16, 17]

Para a determinação do número correto de PC o método mais utilizado consiste no método de Validação Cruzada (CV – do inglês, Cross Validation), o qual se baseia na habilidade de previsão de um modelo construído por parte de um conjunto de dados, seguido pela previsão do restante do conjunto de dados, que é realizada pelo modelo construído. A validação cruzada pode ser realizada em blocos, ou seja, um número determinado de amostras é deixado de fora no processo de construção do modelo e a seguir essas amostras são previstas pelo modelo construído, ou ainda por um caso conhecido como “leave-one-out” (deixe uma fora por vez), em que uma amostra é deixada de fora no processo de construção do modelo e a seguir essa amostra é prevista pelo modelo construído. Em ambos os casos, o processo é repetido até que todas as amostras tenham sido previstas e o erro quadrático médio da validação cruzada (RMSECV) é calculado. [43, 48]

O cálculo é realizado para o número de componentes de 1 até A , e os resultados de RMSECV são apresentados em um gráfico em função do número de PCs. O comportamento típico para esses gráficos é a observação de um mínimo ou um patamar, que indica a melhor dimensionalidade do modelo de regressão, ou seja, o melhor número de PCs que produziu o menor erro de previsão sem perda significativa da variância dos dados. [43, 48]

O objetivo comum do uso do PCA é classificar, separar, amostras apresentando o resultado visualmente na forma de gráfico de duas ou três dimensões. No entanto ele pode ser utilizado, também, para melhorar o resultado de outras ferramentas como pode ser visto no trabalho: Simulação de tratamento de água de descarte industrial usando rede artificial neural a análise de componentes principais [53], já que o PCA reduz as variáveis, ao trocar o sistema original por novas variáveis chamadas PCs.

3.4.2. Quadrados Mínimos Parciais - PLS

O PLS ou a regressão por PLS (PLSR) é considerado o método de regressão mais utilizado para a construção de modelos de calibração multivariada a partir de dados de primeira ordem. Este método, assim como a PCR, não requer um conhecimento exato de todos os componentes presentes nas amostras podendo realizar a previsão de amostras

mesmo na presença de interferentes, desde que estes também estejam presentes por ocasião da construção do modelo (vantagem de primeira ordem). [39]

Para o método de regressão PCR, a decomposição da matriz **X** realizada pela PCA é feita de forma independente da matriz **Y**. Enquanto que para o método de regressão PLS a informação de **Y** é incorporada, de forma que cada PC do modelo sofre uma pequena modificação para buscar a máxima covariância entre **X** e **Y** e passa a receber a terminologia de Variável Latente (VL). [34]

Geometricamente, representa-se esta incorporação da matriz de respostas **Y** à matriz de dados **X** como mostrado na Figura 16:

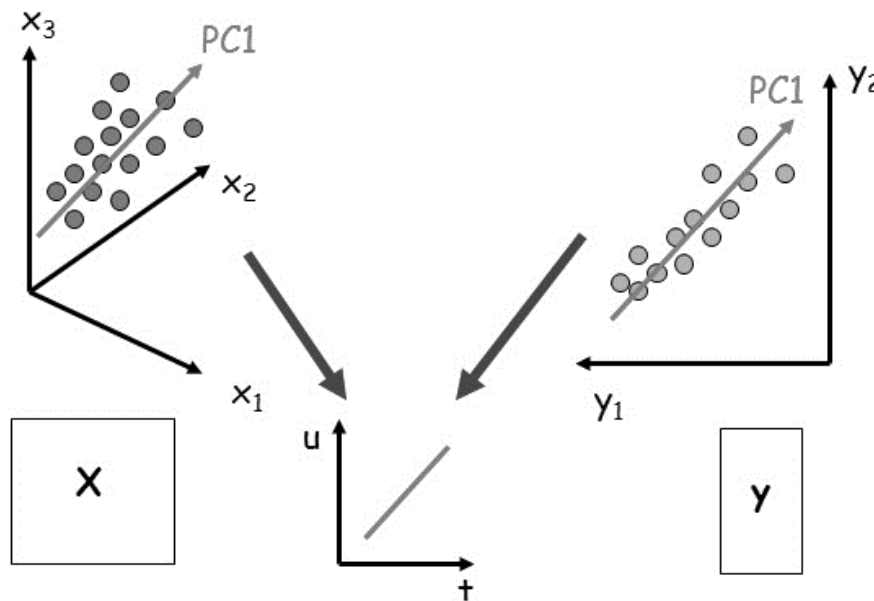


Figura 16 - Formação da Variável Latente 1.

Como representado na Figura 15, no PLS é feita a Análise de Componentes Principais tanto na matriz de dados **X** quanto na matriz de respostas **Y** e, para melhorar a correlação linear entre as duas matrizes, as Componentes Principais calculadas podem sofrer pequenas modificações em seus ângulos. Devido a estas modificações no ângulo das Componentes Principais, elas deixam de ser ortogonais e passam a serem chamadas Variáveis Latentes (VL).

O termo *Parcial*, de Quadrados Mínimos Parciais, refere-se à característica da construção do modelo que se faz pela associação linear entre as variáveis Y, X_1, \dots, X_n após retirar o “efeito” de qualquer subgrupo das $n-1$ variáveis restantes; ou seja, mede-se a dependência entre Y e X_m após remover a dependência linear de Y com um subgrupo de k de $n-1$ variáveis independentes $X_i, i=1, \dots, p; i \neq m$. Esta dependência linear de Y com o subgrupo de k variáveis é então chamado de “efeito” do subgrupo. Como a correlação normal faz uso da matriz de covariância de erros, usa-se correlação parcial. [42, 44]

A equação 1 pode ser reescrita na forma:

$$Y_i = \alpha + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + e_i \quad i = 1, \dots, n. \quad (7)$$

Aonde x representa os valores das variáveis independentes, Y é a resposta correspondente, e é o componente do erro, α e β são parâmetros.

Esta equação sugere que a resposta Y varia linearmente com cada variável independente quando as outras permanecem constantes, deixando de lado o componente do erro. [54]

Para estimar os parâmetros, usam-se *Quadrados Parciais*. Para este modelo, é necessário variar α , β_1 , β_2 e β_3 simultaneamente para minimizar a soma dos desvios quadrados: [54]

$$\sum_{i=1}^n (y_i - \alpha - \beta_1 x_{i1} - \beta_2 x_{i2} - \beta_3 x_{i3})^2 \quad (8)$$

O modelo PLSR é obtido através de um processo iterativo, no qual se otimiza ao mesmo tempo a projeção das amostras sobre o(s) peso(s), para a determinação dos escores, e o ajuste por uma função linear dos escores da matriz \mathbf{X} aos escores da matriz \mathbf{Y} de modo a minimizar os desvios. Essa otimização simultânea ocasiona pequenas distorções nas direções dos pesos, de modo que, rigorosamente eles perdem a ortogonalidade, levando a pequenas redundâncias de informação. Porém são essas pequenas redundâncias que otimizam a relação linear entre os escores, e estas distorções da ortogonalidade entre os componentes principais na PLSR fazem com que não sejam mais componentes principais (que são ortogonais) e sim variáveis latentes (VLs). [38-39]

A PLSR estende o conceito do modelo inverso (propriedade como função da medida instrumental) trocando as variáveis originais por um subconjunto truncado das variáveis latentes dos dados originais. Considerando um caso geral para a determinação de mais de uma espécie de interesse, logo \mathbf{Y} é uma matriz de dimensão $(n \times z)$, em que z é o número de colunas de \mathbf{Y} , tem-se a decomposição de ambas as matrizes de dimensão $(n \times m)$ e \mathbf{Y} em suas matrizes de escores e pesos:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E}_X \quad (9)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{E}_Y \quad (10)$$

onde \mathbf{X} é a matriz de dados (medida instrumental), \mathbf{Y} é a matriz de resposta (concentração, por exemplo), \mathbf{T} e \mathbf{U} são os escores para as duas matrizes de dados, \mathbf{P} e \mathbf{Q} são os respectivos pesos, h é número de VLs, \mathbf{E}_X e \mathbf{E}_Y são os respectivos resíduos compostos pelas VLs descartadas, ou seja, as matrizes que contem a parte não modelada. [38-40]

Uma relação linear é, então, estabelecida entre os escores de \mathbf{X} e os escores de \mathbf{Y} para cada VL: [11]

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h \quad (11)$$

em que, \mathbf{b}_h é o vetor de coeficientes de regressão do modelo linear para cada VL, obtido através de:

$$\mathbf{b}_h = \frac{\mathbf{u}_h^T \mathbf{t}_h}{\mathbf{t}_h^T \mathbf{t}_h} \quad (12)$$

A Figura 16 ilustra a decomposição das matrizes X e Y no produto das matrizes de escores e pesos. A decomposição pode ser realizada através de diversos algoritmos que procedem a referida decomposição por passos diferentes chegando ao final em resultados praticamente iguais. Um exemplo desses algoritmos é o NIPALS (do inglês, Nonlinear Interactive Partial Least Squares – Quadrados Mínimos Parciais Interativos Não lineares). [48]

$$\begin{array}{c}
 \begin{array}{c} m \\ \square \\ n \end{array} X = \begin{array}{c} 1 \\ \square \\ n \end{array} t_1 \begin{array}{c} m \\ \square \\ 1 \end{array} P_1^T + \dots + \begin{array}{c} 1 \\ \square \\ n \end{array} t_h \begin{array}{c} m \\ \square \\ 1 \end{array} P_h^T + \begin{array}{c} m \\ \square \\ n \end{array} E_X \\
 \\
 \begin{array}{c} z \\ \square \\ n \end{array} Y = \begin{array}{c} 1 \\ \square \\ n \end{array} u_1 \begin{array}{c} z \\ \square \\ 1 \end{array} q_1^T + \dots + \begin{array}{c} 1 \\ \square \\ n \end{array} u_h \begin{array}{c} z \\ \square \\ 1 \end{array} q_h^T + \begin{array}{c} z \\ \square \\ n \end{array} E_Y
 \end{array}$$

Figura 17 - Decomposição em variáveis latentes das matrizes X e Y para modelos PLS.

No final do processo a variância explicada pela primeira VL será maior que a variância explicada pela segunda VL e a terceira VL explicará uma variância menor que a segunda VL, e assim sucessivamente até o número de VLs definido e o algoritmo, geralmente, converge rapidamente. [38, 48]

Para determinar o número ótimo de VLs, duas técnicas de validação são comumente utilizadas: série de teste e validação cruzada. O método de validação cruzada, utilizado neste trabalho, consiste em usar uma matriz contendo as amostras de calibração, cujo procedimento se faz com a retirada de uma ou mais amostras para a predição e as amostras remanescentes são utilizadas para modelagem. Repete-se este procedimento até que se tenha feito a predição com todas as amostras. O número ótimo de VLs será aquele capaz de produzir o menor erro médio quadrático de previsão (RMSEP), sendo seu método de cálculo citado adiante.

Os modelos construídos com PLS têm sido utilizados na correlação de propriedades de substratos diversos, como: Aplicação de métodos de calibração multivariada para a determinação simultânea de riboflavina (VB2), tiamina (VB1), piridoxina (VB6) e nicotinamida (VPP) [55], Determinação de ácido oleico em sementes de girassol por espectroscopia no Infravermelho e calibração multivariada [25].

3.5. Cálculo dos erros

A eficiência dos modelos de calibração multivariada pode ser avaliada pelo cálculo dos valores de erro quadrático médio (RMSE) e erro relativo (E). Esses valores expressam a exatidão do modelo, ou seja, a proximidade entre o valor calculado pelo modelo ($y_{previsão}$) e o valor verdadeiro ou obtido por um método de referência (y_{real}). Os erros são definidos como: [48]

$$RMSE = \sqrt{\sum_{i=1}^n (y_{previsão} - y_{real})^2} \quad (13)$$

em que n é o número de amostras.

$$E = \sqrt{\frac{(y_{previs\tilde{a}o} - y_{real})^2}{(y_{real})^2}} \times 100 \quad (14)$$

RMSEC é o RMSE calculado a partir das amostras de calibração, isto é, uma medida do erro na modelagem; RMSECV a partir das amostras da validação cruzada e RMSEP das amostras do conjunto de previsão.

3.6. Linearidade dos Modelos

Uma simples associação entre as variáveis X e Y produz um par de valores ou, graficamente, aponta esta dispersão em uma linha reta. Uma pequena dispersão muito próxima à linha indica forte associação (correlação). A medida numérica desta relação é chamada Coeficiente de Correlação de Amostra; [54] também chamado de Coeficiente de Correlação de Pearson, em homenagem ao seu criador, sendo simbolizada por R. Observando a equação 15, infere-se que o valor de R varia no intervalo: $-1,0 \leq R \leq 1,0$.

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{[\sum_{i=1}^n (x_i - \bar{x})^2][\sum_{i=1}^n (y_i - \bar{y})^2]}} \quad (15)$$

aonde $(x_1, y_1), \dots, (x_n, y_n)$ são os n pares de observações.

Valores próximos a 0 indicam falta de linearidade e valores próximos a 1 e -1 indicam alta linearidade. [42]

CAPÍTULO 4: Procedimiento Experimental

4. Procedimento Experimental

Os biodieseis utilizados no presente trabalho foram produzidos por transesterificação em via metílica (usando metanol) e via etílica (usando etanol) no Laboratório de Biocombustíveis (LABIQ) da Universidade Federal de Uberlândia - MG.

O diesel foi cedido pela Transpetro S/A.

4.1. Preparação das amostras

Para cada modelo foram preparadas 40 amostras, sendo estas amostras divididas em dois grupos: 30 amostras para calibração e 10 amostras para previsão.

Utilizando os valores de densidade do diesel e dos biodieseis determinados no laboratório de análises da Transpetro S/A; os valores de concentração para as misturas (1 a 20% v/v) foram convertidos para massa. As massas foram aferidas utilizando balança analítica. Assim, para preparar uma mistura B5 utilizando biodiesel etílico de milho, de volume final 10 mL, são misturados 0,5 mL de biodiesel e 9,5 mL de diesel. Estes volumes foram convertidos em massa utilizando a densidade, o que corresponde a 0,4387g de biodiesel e 8,1102g de diesel.

Os valores das massas de biodiesel e diesel aferidos na balança foram anotados para posterior cálculo de concentração de cada mistura.

4.2. Obtenção dos espectros no MIR

Os espectros de absorvância foram obtidos no equipamento Shimadzu IRPrestige-21, utilizando resolução de 4 cm^{-1} , 32 varreduras, apodização do tipo Happ Genzel na faixa de comprimento de onda de $650\text{ a }4000\text{ cm}^{-1}$, com o uso do acessório ATR com cristal de seleneto de zinco (ZnSe). A mistura de biodiesel/diesel, após agitada, é adicionada diretamente sobre o cristal de seleneto de zinco. Utilizam-se em torno de 0,5 mL, o necessário para cobrir todo o cristal. Foram obtidos cinco espectros para cada amostra. Após a obtenção de cada espectro, o ATR foi limpo: primeiro com algodão seco para retirar a mistura de biodiesel e diesel, depois foi utilizado algodão embebido em álcool etílico 96% e secado com algodão. Depois de limpo e seco é realizada a leitura do branco antes de cada leitura de amostra.

4.3. Importação dos dados para Matlab para geração da matriz de dados

Os espectros obtidos foram exportados em arquivo texto padrão ASCII para serem importados para o Matlab para criação da matriz de dados. Foi criado um arquivo de texto com os valores reais de concentração para ser importado para o Matlab.

4.4. Tratamento dos espectros das misturas binárias biodiesel/diesel

Todos os espectros foram corrigidos pelo Método de Correção de Linha de Base (*Baseline*), utilizando *Matlab* 6.1 (*MathWorks Inc.*). Os espectros foram cortados, usando como faixa de trabalho de $685\text{ a }3077\text{ cm}^{-1}$.

4.5. Construção dos Modelos

Os modelos foram construídos utilizando *PLS_Toolbox 3.0* para *Matlab 6.1 (MathWorks Inc.)*. Os dados foram centrados na média, utilizada regressão com SIMPLS e Validação Cruzada do tipo *Leave one out*. O número de Variáveis Latentes (VL) foi determinado utilizando o gráfico de Variância capturada (%) versus Número de Variáveis Latentes. A eficiência dos modelos foi avaliada pelos valores de RMSEC, RMSEP, Erro médio de previsão (%), gráfico de Valores Reais versus Previstos e valores de linearidade (R^2).

CAPÍTULO 5: Resultados e Discussões

5. Resultados e Discussões

5.1. Preparação das amostras

Para fazer a leitura dos espectros no Infravermelho Médio utilizando o acessório ATR; são gastos 0,5 mL de amostra. Considerando cinco espectros para cada mistura, são necessários 2,5 mL de amostra para obtenção dos espectros, mesmo assim, decidiu-se preparar misturas de volume final 10 mL. Não foram preparadas misturas de volume final 5 mL devido a grande dificuldade de manipulação de quantidades pequenas de biodiesel para preparar misturas na faixa de concentração de 1 a 20% v/v. É esta dificuldade na manipulação das quantidades pequenas de biodiesel que justificam os valores de concentração diferentes do planejado. Inicialmente foram feitos os cálculos (a faixa de concentração de 1 a 20% v/v foi igualmente dividida entre as amostras) e, ao executar a pesagem do biodiesel e diesel, buscou-se ficar o mais próximo possível das massas teóricas calculadas para cada concentração. Como a massa de biodiesel pesada é sempre menor do que a massa de diesel, as misturas foram iniciadas com a pesagem do biodiesel, pois como a massa é muito pequena, pequenas variações levavam a grandes desvios de concentração (principalmente nas amostras de menor concentração) e, no caso de erros de adição, não se perde o diesel.

As amostras utilizadas para calibração do modelo e as amostras utilizadas para previsão foram preparadas de forma independente no intervalo de concentração de 1 a 20% v/v cada, tomando o cuidado de não repetir na previsão valores de concentração contidos na calibração.

Os valores de massa de biodiesel e diesel adicionados com o auxílio da balança analítica foram utilizados para o cálculo da concentração real das amostras, sendo este o método de referência para comparação com os valores previstos pelos modelos.

5.2. Obtenção dos espectros de Infravermelho Médio

A vantagem do uso de álcool etílico hidratado para a limpeza é que qualquer resíduo é facilmente identificado pela presença da larga banda de absorção do grupo OH. Os espectros foram obtidos com 32 varreduras, ou seja, o espectro obtido da amostra é o resultado da média de 32 varreduras. A apodização, Happ Genzel, utilizada é a que apresenta resolução adequada para líquidos, segundo manual do equipamento. A resolução utilizada foi de 4 cm^{-1} , ou seja, esta é a diferença entre os números de onda lidos pelo equipamento.

5.4. Tratamento dos espectros das misturas binárias biodiesel/diesel

Ao sobrepor os espectros das 40 misturas binárias de biodiesel/diesel com concentrações variando na faixa de 1 a 20% v/v; vê-se a variação da linha de base dos espectros (Figura 18). Assim, é necessário fazer a remoção matemática das fontes de variação indesejáveis que não serão removidas naturalmente durante a análise dos dados e que podem influenciar os resultados finais, já que os sinais medidos consistem de sinal verdadeiro e ruído aleatório. Para ajustar estas diferenças, eliminando as flutuações da linha de base (Figura 19), usou-se a correção da linha de base (*Baseline*) dos espectros.

Todos os espectros foram obtidos pela varredura do intervalo de 650 a 4000 cm^{-1} ; no entanto as duas extremidades dos espectros continham dados desnecessários, assim foram os espectros cortados, ficando a faixa de trabalho de 685 a 3077 cm^{-1} (Figura 20); evitando uso de informações desnecessárias na matriz de dados. Após estes tratamentos, ainda foram ajustadas as escalas dos eixos e suas legendas (Figura 21).

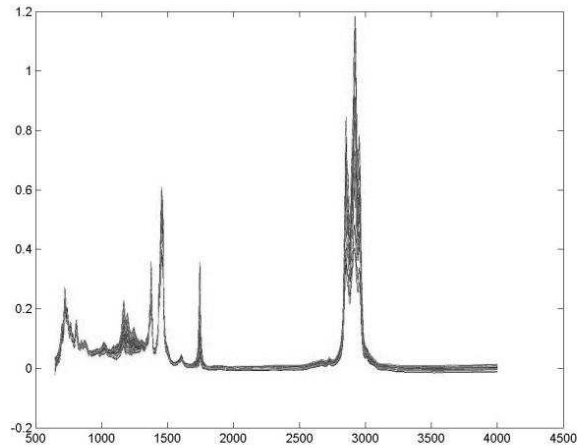


Figura 18 - Espectros originais das misturas binárias biodiesel/diesel (1 a 20% v/v)

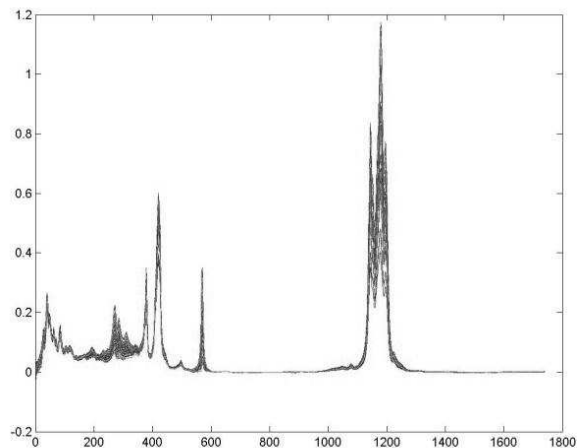


Figura 19 - Espectros de misturas binárias biodiesel/diesel (1 a 20% v/v) após correção da linha de base.

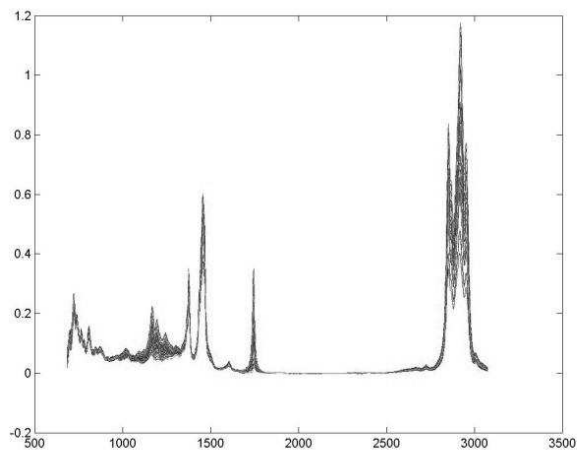


Figura 20 - Espectros de misturas binárias biodiesel/diesel (1 a 20% v/v) após correção da linha de base e cortes.

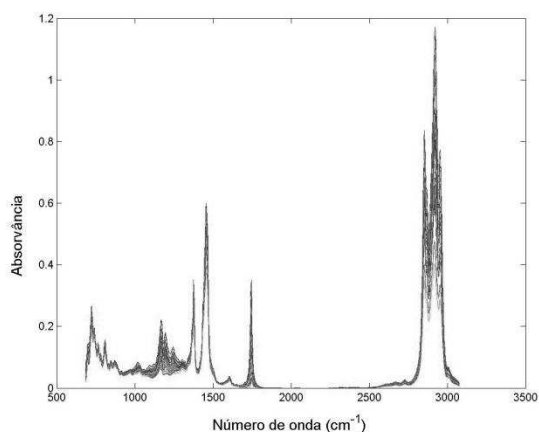


Figura 21 - Resultado final dos espectros de misturas binárias biodiesel/diesel (1 a 20% v/v).

Estes procedimentos foram feitos para cada conjunto de 40 espectros que constituem as misturas de um dos biodieseis usados com diesel, na faixa de concentração de 1 a 20% v/v.

5.5. Construção dos Modelos

5.5.1. Correlação dos espectros com propriedade de interesse

Terminada a etapa de ajustes dos espectros, constrói-se a matriz de dados **X** a partir dos dados dos espectros corrigidos e cortados.

A diferença entre os espectros é muito sutil, impossibilitando a identificação/correlação visual dos espectros com a propriedade de interesse: a concentração de biodiesel em misturas binárias com diesel. Mas são justamente estas regiões aonde ocorrem variações entre os espectros (Figura 22) que são necessárias à modelagem utilizando PLS. Poder-se-ia pensar em utilizar somente estas regiões para a construção dos modelos, diminuindo a quantidade de dados; no entanto, utilizando apenas o PLS, não se pode garantir que não haja informações relevantes em regiões diferentes das três cuja variação é visivelmente mais notável.

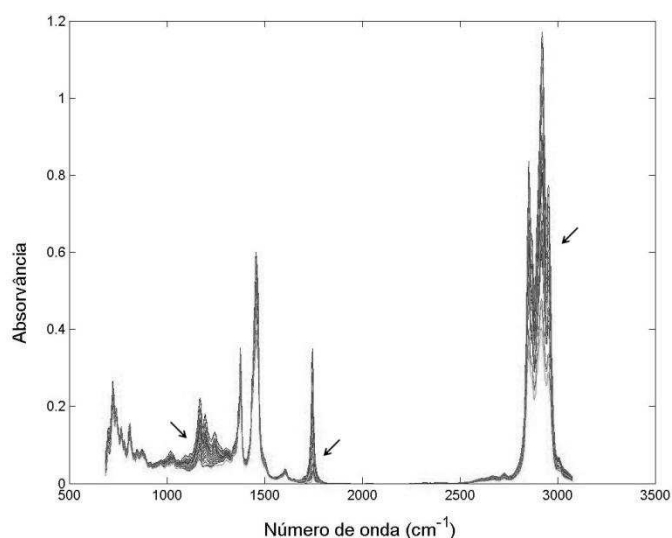


Figura 22 - Áreas de grande variação de absorvâncias entre os espectros.

5.5.2. Dados Centrados na Média

Os dados experimentais originais podem não ter uma distribuição adequada para a análise, dificultando a extração de informações úteis e interpretação dos mesmos. Nestes casos, um pré-processamento nos dados originais pode ser de grande valia. Para centrar os dados na média, calcula-se a média das intensidades para cada comprimento de onda e subtrai-se cada intensidade do respectivo valor médio; desta forma o eixo cartesiano tem sua origem centrada no valor médio das medidas.

5.5.3. Determinação no número de variáveis latentes para construção dos modelos

As Variáveis Latentes são as novas variáveis que serão utilizadas para representar os dados. Como as variáveis originais estão sendo trocadas por novas variáveis, deve-se avaliar até que ponto a troca é representativa do conjunto de dados, avaliar quantas variáveis latentes devem ser utilizadas para representar o conjunto de dados.

Na construção do modelo, é apresentada uma tela do programa em que informações importantes para decidir qual o número de variáveis latentes são apresentadas (Figura 23): a porcentagem de variância capturada, ou seja, a quantidade de dados originais que está sendo explicada em função do número de variáveis latentes. A primeira coluna apresenta o número de variáveis latentes; a segunda coluna apresenta a porcentagem de variância explicada da matriz de dados \mathbf{X} (variáveis); a terceira coluna apresenta a variância acumulada da matriz de dados \mathbf{X} ; a quarta coluna apresenta a porcentagem de variância explicada da matriz de respostas \mathbf{Y} ; a quinta coluna apresenta a variância acumulada da matriz de respostas \mathbf{Y} .

Aliado a Figura 23, tem-se o gráfico do Erro Quadrático Médio de Calibração (RMSEC) em função no número de variáveis latentes (Figura 24). Na Figura 24 constata-se que quanto maior a quantidade de variáveis latentes, menor é o erro, pois maior a quantidade de informação capturada. No entanto, ao fazer uso de grande número de variáveis latentes, pode-se estar incorporando ao modelo informações desnecessárias e ruídos.

Os modelos de regressão linear começam a perder a linearidade em torno de 15 variáveis latentes, assim, deve-se escolher a menor quantidade de variáveis latentes possível para manter o modelo linear. Por conseguinte, usam-se as informações das Figuras 23 e 24 em conjunto; para determinar o número de variáveis latentes em que se tenha boa representatividade dos dados originais e menor Erro Médio Quadrático de Calibração.

Para este caso específico, foram escolhidas cinco variáveis latentes, pois o ganho a partir da sexta variável latente é muito pequeno, não justificando o uso de mais variáveis latentes.

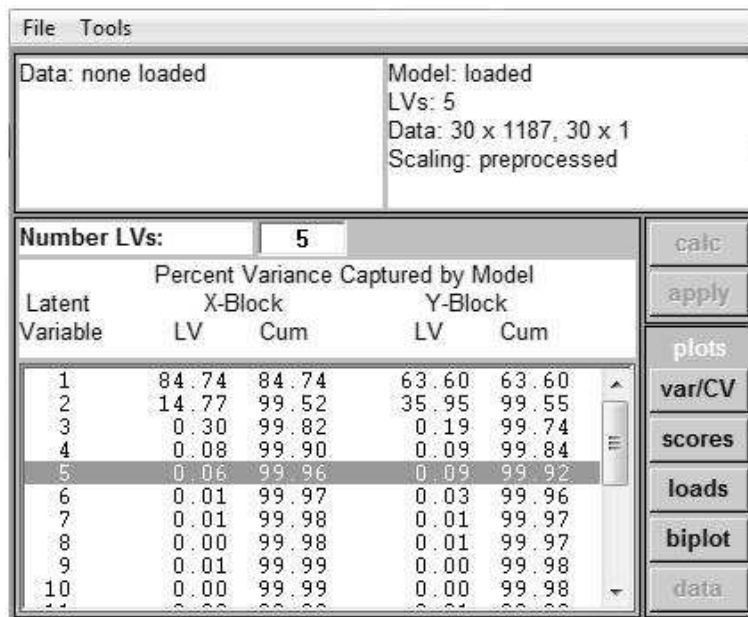


Figura 23 - Tela de escolha do número de variáveis latentes.

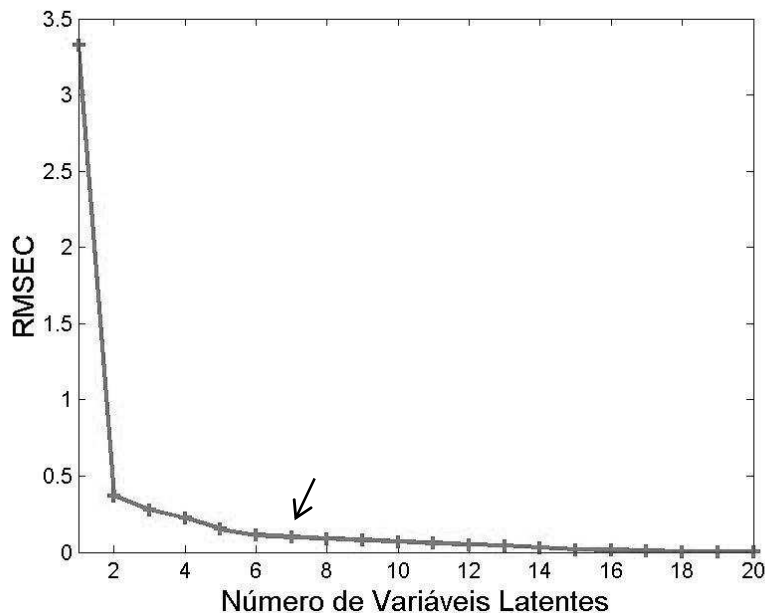


Figura 24 – Gráfico dos valores de RMSEC VS Número de Variáveis latentes.

5.5.4. Validação Cruzada

A Validação Cruzada do tipo *leave-one-out* (“deixar um de fora”) é baseada na avaliação da magnitude dos erros de previsão do modelo de calibração. Esta avaliação é feita pela comparação das concentrações previamente conhecidas da seguinte forma: primeiro, retira-se uma amostra do conjunto de calibração e constrói-se o modelo com as demais; depois o modelo construído é usado para prever a concentração da amostra que foi retirada inicialmente e calcula-se o erro desta previsão; este procedimento de retirar amostra e construir o modelo com as demais é repetido para todas as demais amostras; então se calcula a soma dos erros quadráticos médios para o conjunto de calibração (RMSEC). Todo esse processo é repetido para modelos construídos com uma, duas, três,

e assim por diante, variáveis latentes; desta forma, a quantidade mais adequada de variáveis latentes será correspondente ao menor valor de RMSEC. Graficamente, a relação entre o número de variáveis latentes e RMSEC é mostrada na Figura 24.

5.5.5. Linearidade dos Modelos

A linearidade mede o grau e o tipo do relacionamento entre as variáveis estudadas, ou seja, refere-se à capacidade do método de gerar resultados linearmente proporcionais à concentração do biodiesel, enquadrados em faixa de 1 a 20% v/v.

No entanto, a linearidade é comumente apresentada nos gráficos em função de R^2 , tendo seus valores possíveis variando no intervalo $0 \leq R^2 \leq 1$; podendo também vir expresso em porcentagem. No presente trabalho, são identificados os valores de R^2 nos gráficos de valores reais versus valores previstos.

Como na criação dos modelos de quantificação por PLS usa-se regressão linear, este coeficiente é muito importante para avaliação dos modelos.

Comparando-se os valores do coeficiente de linearidade, temos a classificação estatística do tipo de correlação linear de acordo com a faixa de valores:

Tabela 4 - Classificação dos valores de linearidade [54].

Valores de R^2	$0,9990 \leq R^2 \leq 0,9995$	$0,9995 < R^2 < 1$	1
Classificação da linearidade	forte	fortíssima	perfeita

5.5.6. Gráficos de Valores Reais versus Valores Previstos

Estes gráficos possuem grande importância na avaliação da qualidade do modelo; pois, através deles, podem ser avaliadas as distribuições das amostras no intervalo de previsão de concentração e a distribuição das amostras em torno da linha de tendência.

A boa distribuição das amostras no intervalo indica que todo o intervalo foi coberto pelas mostras de previsão, assim, o erro médio associado às medidas de previsão é representativo de todo o intervalo proposto. Os erros das amostras de previsão não costumam ser iguais, assim a concentração de valores apenas numa faixa determinada do intervalo deixa o modelo tendencioso.

A distribuição em torno da linha de tendência é um parâmetro valioso na avaliação da linearidade do modelo, pois; como se usam valores de R^2 para medir a linearidade, desvios equidistantes da linha de tendência, um acima e outro abaixo da linha, levam a altos valores de R^2 (Figura 25), sendo então a visualização da distribuição dos pontos em torno da linha de tendência essencial para confirmação dos valores de linearidade (O coeficiente de correlação é fortemente influenciado pela média da distribuição). A simples visualização de linearidade sem o cálculo de R^2 é chamada de Relação Linear Aparente, assim, o modelo com boa linearidade deve apresentar valores de R^2 próximos a um e linearidade aparente.

Do ponto de vista da relação entre os valores reais de concentração e os valores calculados de concentração no gráfico, a relação linear será direta quando apresentarem-se distribuídos ao longo de uma linha com inclinação positiva (coeficiente angular da reta maior que zero); a relação linear será indireta quando apresentarem-se distribuídos ao

longo de uma linha com inclinação negativa (coeficiente angular da reta menor que zero) e não haverá relação linear quando a distribuição ao longo da reta tiver inclinação zero.

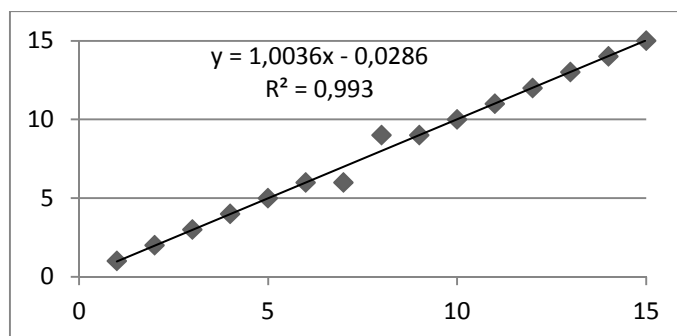


Figura 25 - Exemplo de desvios equidistantes da linha de tendência.

5.5.7. Erros de previsão

Para todos os modelos, foram calculados os erros individuais da previsão de cada amostra e o erro médio de previsão considerando todas as amostras utilizadas na previsão. O valor de concentração real foi obtido pelas massas dos componentes das misturas binárias, aferidas com a balança, assim a pesagem utilizando a balança analítica é o método de referência para cálculo dos erros percentuais.

5.6. Modelos de quantificação de biodiesel em diesel

5.6.1. Biodiesel etílico de soja

Os espectros de Infravermelho Médio das 40 misturas de biodiesel etílico de soja e diesel; depois de terem seus espectros agrupados, tratados e cortados; são apresentados na Figura 26.

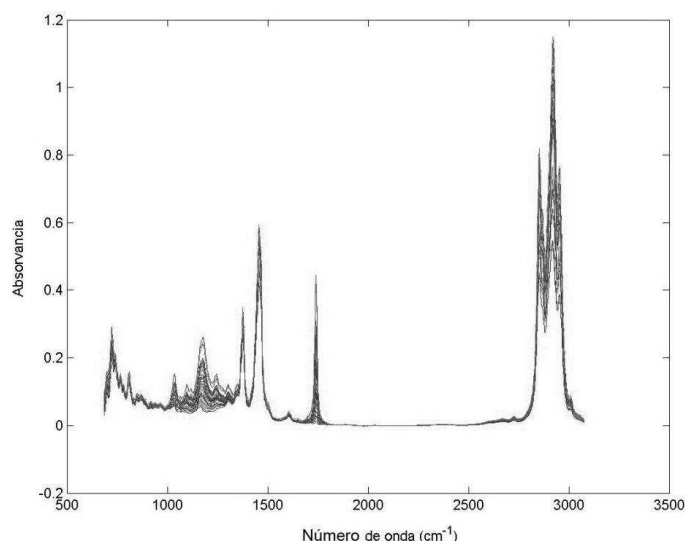


Figura 26 - Espectros MIR para as amostras de misturas de biodiesel etílico de soja em diesel.

O melhor modelo de calibração por PLS foi construído com 5 variáveis latentes (Figura 27), apresentando variância explicada de 99,96% para a matriz de dados e 99,92% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,1515 e 0,2266. Comparando-se estes dois valores de erro

observa-se que o modelo é coerente, pois apresenta erro de calibração (construção do modelo) menor que o erro de previsão. A média dos erros percentuais dos valores de concentração das misturas de biodiesel/diesel previstas foi de apenas 1,39%; estando este valor muito abaixo do limite de erro permitido pela metodologia de quantificação padrão ASTM (American Society for Testing and Materials, organização norte-americana de desenvolvimento de normas), a qual permite erros de até 10%.

Neste modelo, foram preparadas 40 amostras no intervalo de 1 a 20% v/v e destas foram retiradas aleatoriamente, pelo programa, 10 amostras deste conjunto que foram utilizadas para a previsão. A separação de 30 amostras para calibração e 10 amostras para previsão foi feita pelo programa e não preparadas separadamente como nos demais modelos. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 5.

O modelo apresentou fortíssima correlação linear ($R^2 = 0,9996$), boa distribuição das amostras ao longo do intervalo de concentração (1 a 20%) e boa distribuição das amostras ao longo da linha de tendência (Figura 28).

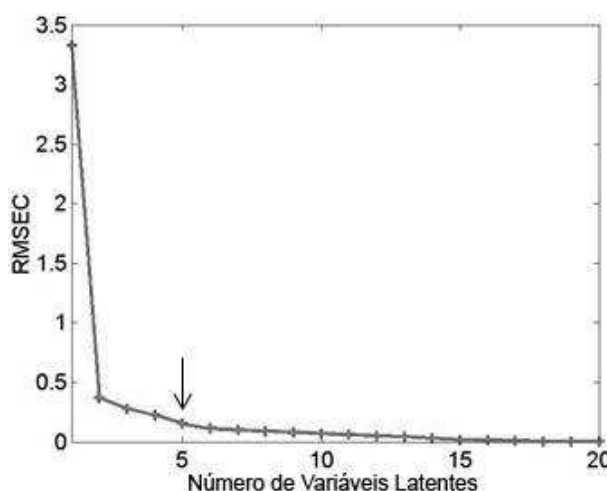


Figura 27 - Gráfico dos valores de RMSEC VS VL (biodiesel etílico de soja em diesel).

Tabela 5 - Concentração real e prevista para mistura de biodiesel etílico de soja em diesel (%v/v).

Valor real	Valor Previsto	Erro (%)
1,9896	1,9883	0,07
2,9014	3,0217	4,15
6,8090	6,8784	1,02
9,7503	9,8071	0,58
10,9014	11,0193	1,08
13,1797	13,0853	0,72
13,7348	13,4889	1,79
17,6153	17,6472	0,18
18,0682	17,8675	1,11
19,0720	18,4668	3,17

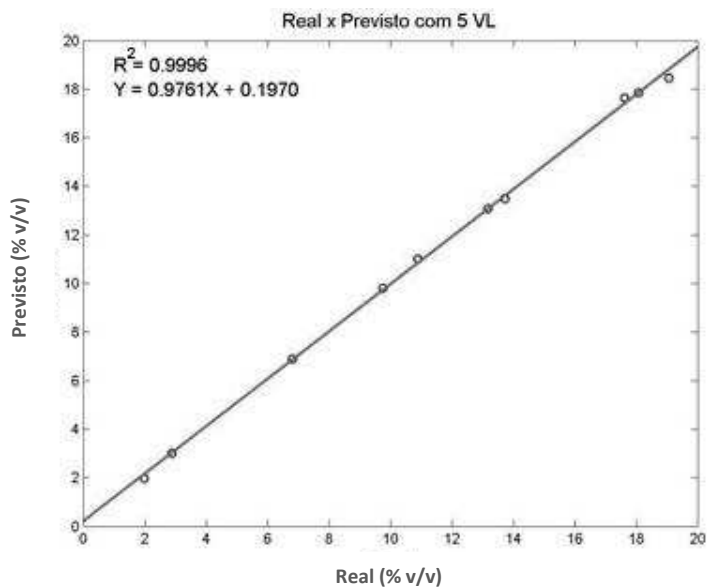


Figura 28 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel etílico de soja).

5.6.2. Biodiesel metílico de soja

Os espectros de Infravermelho Médio das 40 misturas de biodiesel metílico de soja e diesel; depois de terem seus espectros agrupados, tratados e cortados; são apresentados na Figura 29.

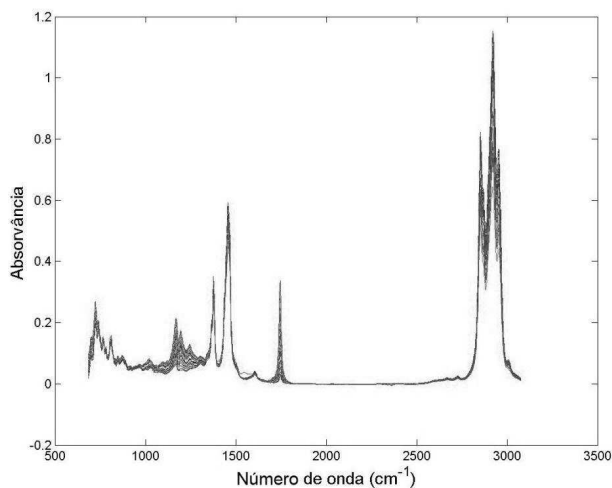


Figura 29 - Espectros MIR para biodiesel metílico de soja em diesel.

O melhor modelo de calibração por PLS foi construído com 6 variáveis latentes (Figura 30), apresentando variância explicada de 99,97% para a matriz de dados e 99,98% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,0916 e 0,1609. Comparando-se estes dois valores de erro observa-se que o modelo é coerente, pois apresenta erro de calibração (construção do modelo) menor que o erro de previsão. A média dos erros percentuais dos valores de concentração das misturas de biodiesel/diesel usadas na previsão é muito pequena,

apenas 1,26%. A amostra de concentração em torno de 5% v/v apresentou erro quase dez vezes menor do que o limite estabelecido pela ASTM.

Neste modelo, foram preparadas 40 amostras no intervalo de 1 a 20% v/v e destas foram retiradas aleatoriamente, pelo programa, 10 amostras deste conjunto que foram utilizadas para a previsão. A separação de 30 amostras para calibração e 10 amostras para previsão foi feita pelo programa. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 6.

O modelo apresentou forte correlação linear ($R^2 = 0,9992$), boa distribuição das amostras ao longo do intervalo de concentração (1 a 20%) e boa distribuição das amostras ao longo da linha de tendência (Figura 31). Os intervalos entre as amostras não foram semelhantes devido à escolha aleatória das amostras de previsão.

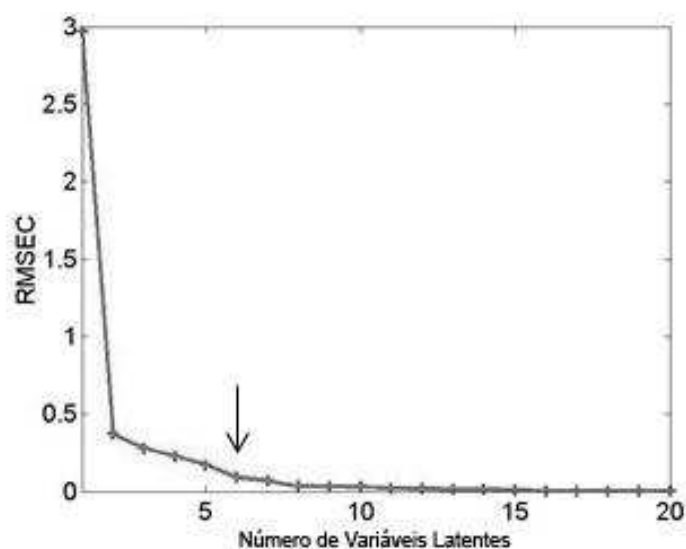


Figura 30 – Gráfico dos valores de RMSEC VS VL para modelo com biodiesel metílico de soja em diesel.

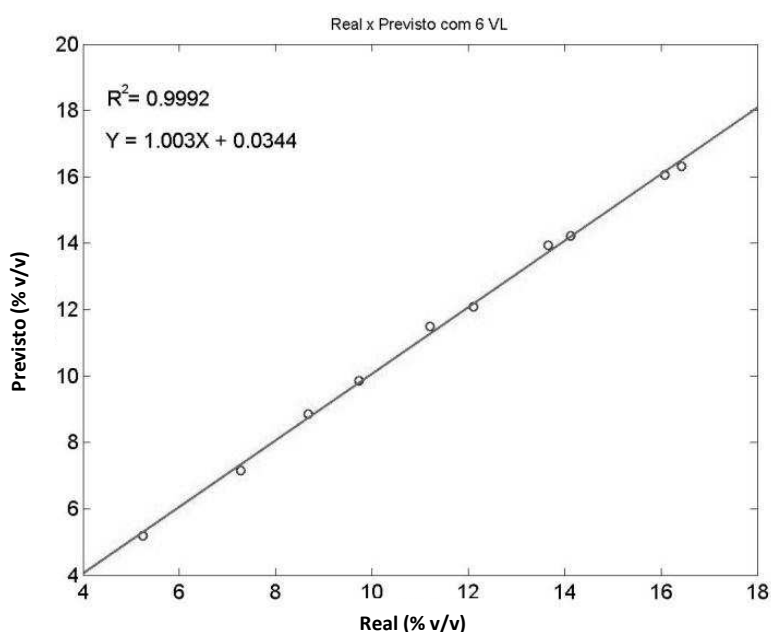


Figura 31 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de soja).

Tabela 6 - Concentração real e prevista para mistura de biodiesel metílico de soja em diesel (% v/v).

Valor real	Valor Previsto	Erro (%)
5,2406	5,1836	1,09
7,2731	7,1479	1,72
8,6678	8,8580	2,19
9,7310	9,8531	1,25
11,2015	11,4941	2,61
12,1054	12,0835	0,18
13,6534	13,9409	2,11
14,1211	14,2191	0,69
16,0783	16,0590	0,12
16,4267	16,3253	0,62

5.6.3. Biodiesel etílico de soja residual

Os espectros de Infravermelho Médio das 40 misturas de biodiesel etílico de soja residual e diesel; depois de terem seus espectros agrupados, tratados e cortados; são apresentados na Figura 32.

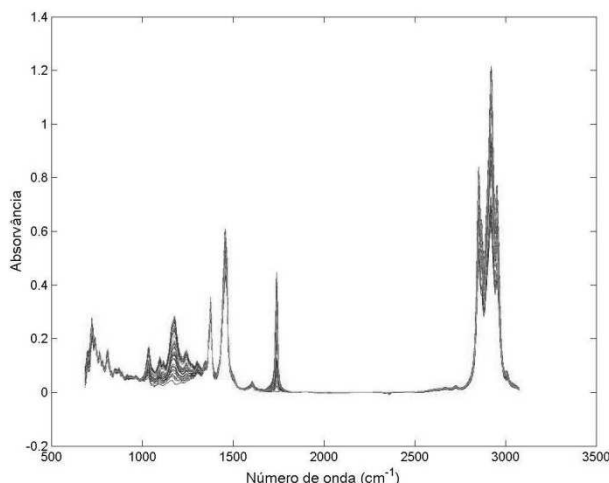


Figura 32 - Espectros MIR para biodiesel etílico de soja residual em diesel.

O melhor modelo de calibração por PLS foi construído com 11 variáveis latentes (Figura 33), apresentando variância explicada de 99,99% para a matriz de dados e 100% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,0417 e 0,0661. De acordo com estes dois valores de erro, conclui-se que o modelo possui coerência, visto que apresenta erro de calibração (construção do modelo) menor que o erro de previsão. A média dos erros percentuais dos valores de concentração previstos das misturas de biodiesel/diesel foi de apenas 0,53%; sendo este erro muito pequeno, quase vinte vezes menor do que o limite estabelecido pela ASTM.

Neste modelo, foram preparadas 30 amostras no intervalo de 1 a 20% v/v e 10 amostras no intervalo de 1 a 20% v/v. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 7.

O modelo apresentou fortíssima correlação linear ($R^2 = 0,9999$), boa distribuição das amostras ao longo do intervalo de concentração (1 a 20%) e boa distribuição das amostras ao longo da linha de tendência (Figura 34).

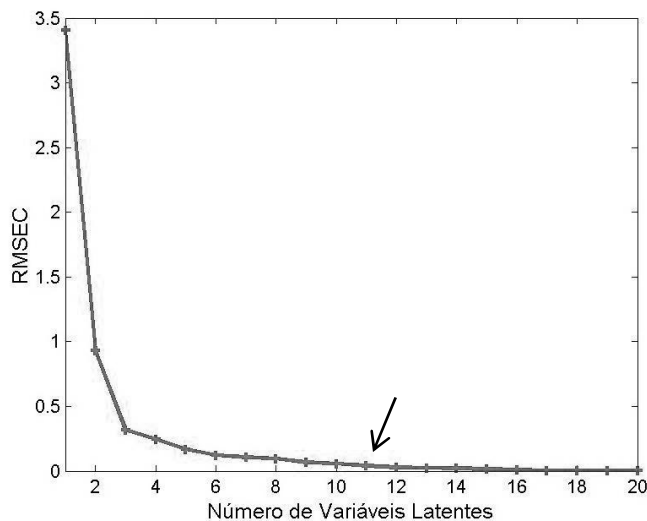


Figura 33 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de soja residual em diesel.

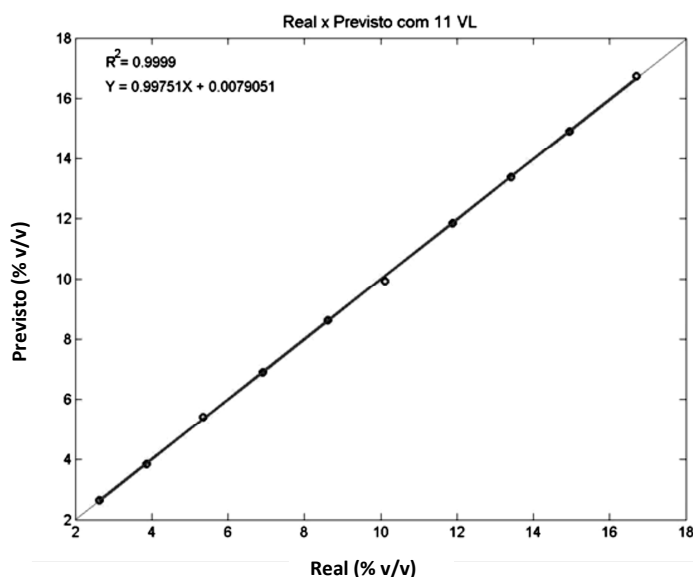


Figura 34 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel etílico de soja residual).

Tabela 7 - Concentração real e prevista de mistura de biodiesel etílico de soja residual em diesel.

Valor real	Valor Previsto	Erro (%)
2,3200	2,6441	0,5364
3,8600	3,8490	0,2855
5,3500	5,4248	1,3974
6,9100	6,9008	0,1337
8,6100	8,6186	0,0999
10,1200	9,9418	1,7606
11,8800	11,8511	0,2429
13,4200	13,3808	0,2923
14,9500	14,9155	0,2311
16,7000	16,7479	0,2871

5.6.4. Biodiesel metílico de soja residual

Os espectros de Infravermelho Médio das 40 misturas de biodiesel metílico de soja residual em diesel; depois de serem agrupados, tratados e cortados; estão representados na Figura 35.

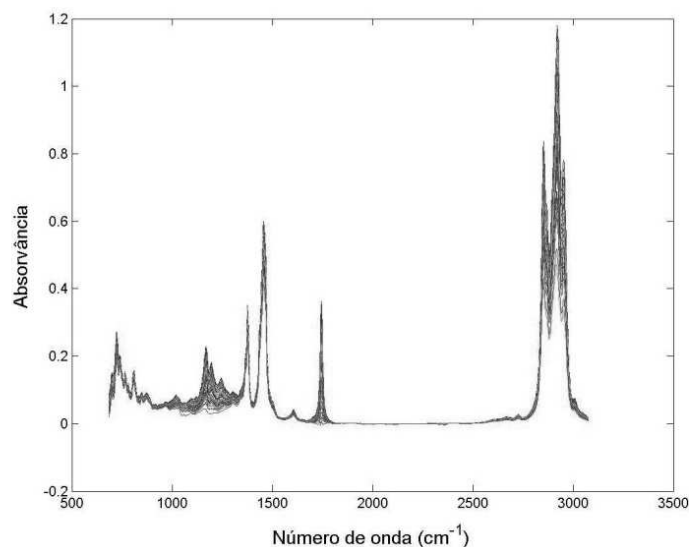


Figura 35 - Espectros MIR para biodiesel metílico de soja residual em diesel.

O melhor modelo de calibração por PLS foi construído com 5 variáveis latentes (Figura 36), apresentando variância explicada de 99,94% para a matriz de dados e 99,94% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,1346 e 0,1487. Comparando-se os valores de erro de calibração e previsão, conclui-se que o modelo possui coerência, já que apresenta erro de calibração (construção do modelo) menor que o erro de previsão.

A média dos erros percentuais dos valores de concentração das misturas de previsão de biodiesel/diesel foi de apenas 2,05%; sendo este erro muito pequeno, quase cinco vezes menor do que o limite estabelecido pela ASTM.

Neste modelo, o conjunto de 30 mostras do conjunto de calibração foi preparado independente do conjunto de 10 amostras de previsão, ambos tendo suas concentrações variando no intervalo de 1 a 20% v/v. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 8.

O modelo apresentou fortíssima correlação linear ($R^2 = 0,9998$), boa distribuição das amostras no intervalo de concentração (1 a 20%) e boa distribuição das amostras ao longo da linha de tendência (Figura 37).

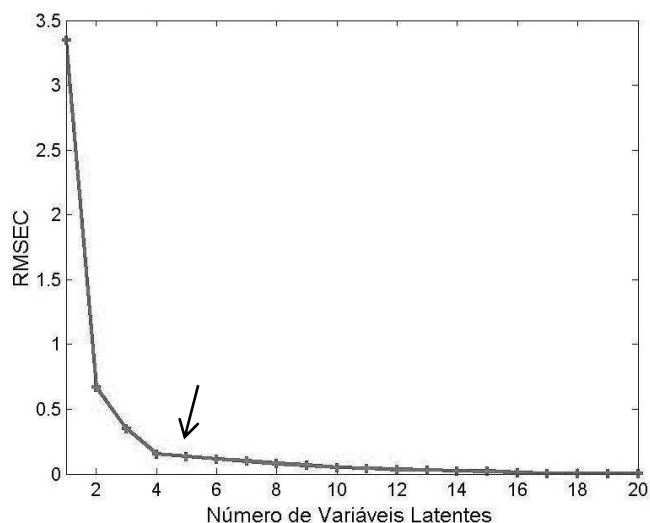


Figura 36 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel metílico de soja residual.

Tabela 8 - Concentração real e prevista para mistura de biodiesel metílico de soja residual em diesel.

Valor real	Valor Previsto	Erro (%)
2,5038	2,6462	5,69
3,6650	3,5157	4,07
5,2054	4,9845	4,24
6,7650	6,7280	0,55
8,3127	8,2508	0,74
9,8733	9,7987	0,76
11,4119	11,2436	1,48
12,9623	12,9264	0,28
14,4951	14,2443	1,73
16,1086	15,9463	1,01

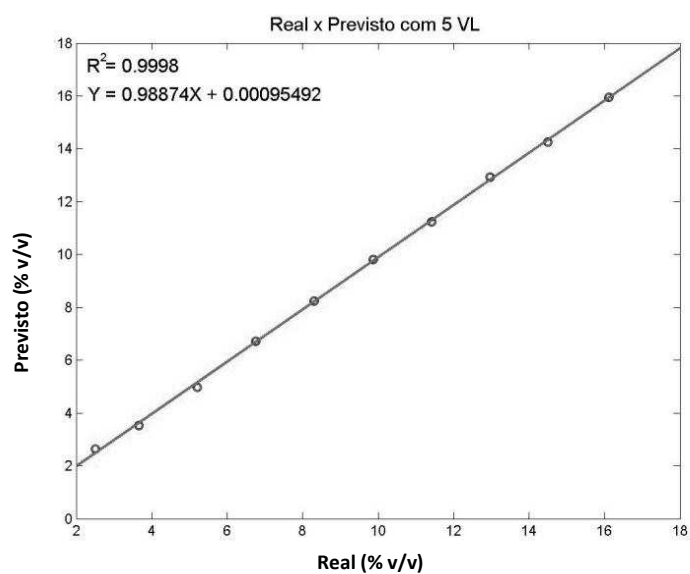


Figura 37 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de soja residual).

5.6.5. Biodiesel etílico de milho

Os espectros de Infravermelho Médio para as 40 misturas de biodiesel etílico de milho em diesel; depois de agrupados, tratados e cortados; estão representados na Figura 38.

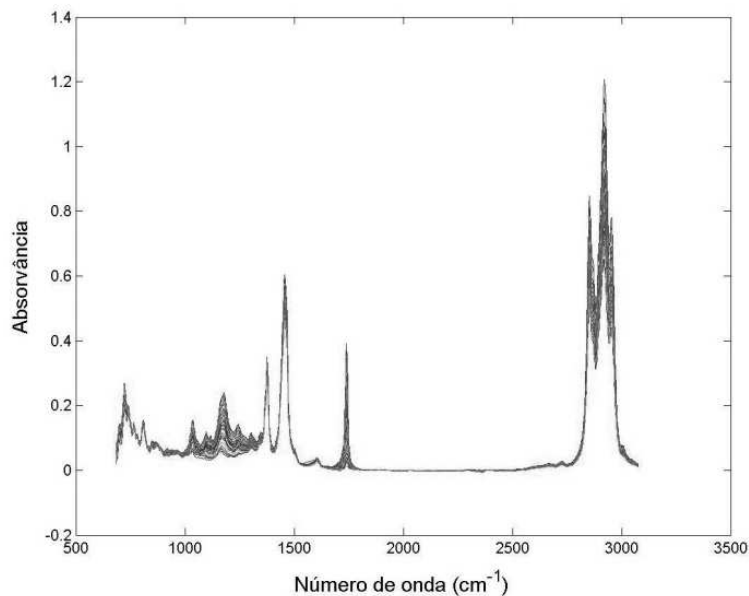


Figura 38 - Espectros MIR para biodiesel etílico de milho em diesel.

O melhor modelo de calibração por PLS foi construído com 9 variáveis latentes (Figura 39), apresentando variância explicada de 99,49% para a matriz de dados e 99,97% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,0330 e 0,1681. Comparando-se os valores de erro de calibração e previsão, conclui-se que o modelo também possui coerência, por conseguinte apresenta erro de calibração (construção do modelo) menor que o erro de previsão.

A média dos erros percentuais dos valores de concentração das misturas de previsão de biodiesel/diesel foi de apenas 1,75%; sendo este erro muito pequeno, mais de cinco vezes menor do que o limite estabelecido pela ASTM.

Neste modelo, o conjunto de 30 mostras do conjunto de calibração foi preparado independente do conjunto de 10 amostras de previsão, ambos tendo suas concentrações variando no intervalo de 1 a 20% v/v. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 9.

O modelo apresentou forte correlação linear ($R^2 = 0,9995$), boa distribuição das amostras no intervalo de concentração (1 a 20%) e boa distribuição das amostras ao longo da linha de tendência (Figura 40).

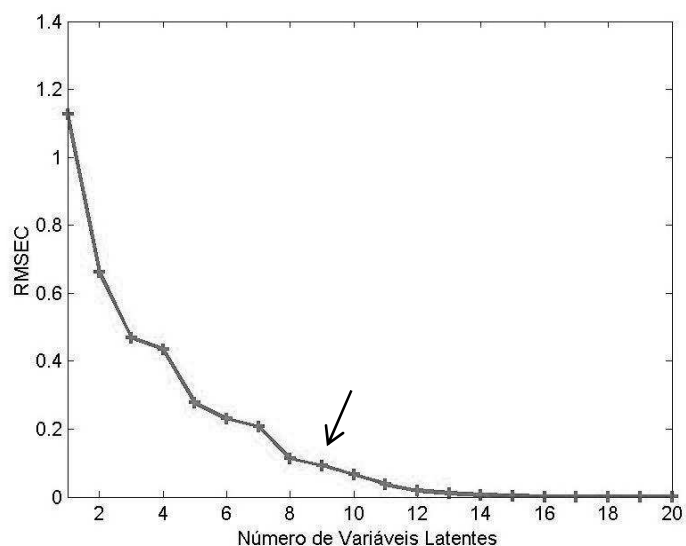


Figura 39 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de milho.

Tabela 9 - Valores de concentração real e prevista para mistura de biodiesel etílico de milho em diesel.

Valor real	Valor Previsto	Erro (%)
3,0826	3,2386	5,06
4,9702	5,0680	1,97
6,6625	6,7042	0,63
8,4640	8,6358	2,03
9,7238	9,8520	1,32
11,2156	10,9543	2,33
12,9377	12,6295	2,38
14,6213	14,4382	1,25
16,2022	16,2277	0,16
17,9065	17,8475	0,33

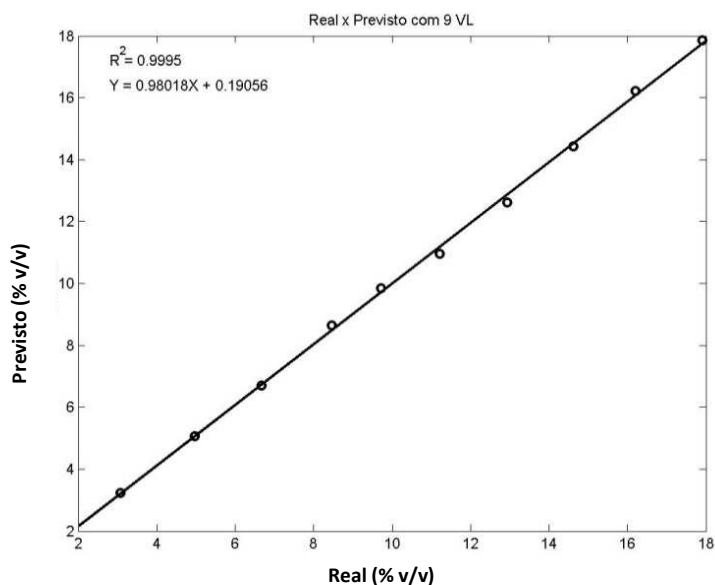


Figura 40 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel etílico de milho).

5.6.6. Biodiesel metílico de milho

Os espectros de Infravermelho Médio para as 40 misturas de biodiesel metílico de milho em diesel; depois de agrupados, tratados e cortados; estão representados na Figura 41.

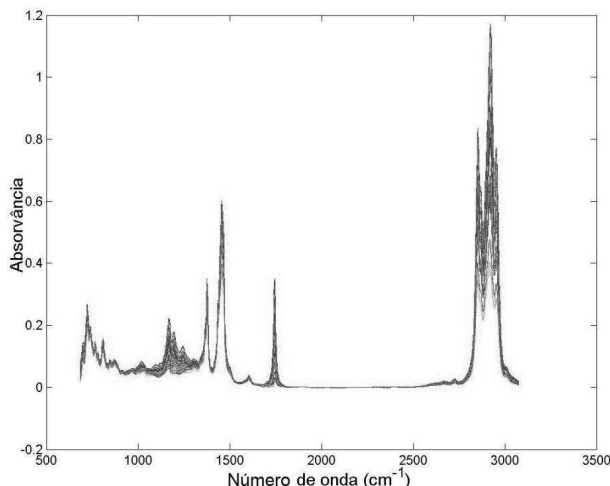


Figura 41 - Espectros MIR para biodiesel metílico de milho em diesel.

O melhor modelo de calibração por PLS foi construído com 4 variáveis latentes (Figura 42), apresentando variância explicada de 99,96% para a matriz de dados e 99,91% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,1532 e 0,1525. Comparando-se os valores de erro de calibração e previsão, conclui-se que o modelo também possui coerência, sendo estes dois valores estatisticamente iguais, o modelo não possui erro de previsão menor do que erro de calibração. A média dos erros percentuais dos valores de concentração das misturas de previsão de biodiesel/diesel foi de apenas 1,40%; sendo este erro muito pequeno, mais de sete vezes menor do que o limite estabelecido pela ASTM.

Neste modelo, o conjunto de mostras do conjunto de calibração foi preparado independente do conjunto de amostras de previsão, ambos tendo suas concentrações variando no intervalo de 1 a 20% v/v.

Algumas amostras de previsão apresentaram valores muito próximos de concentração, devido à dificuldade de manipulação dos valores de massa na preparação das amostras; por isso foi reparada mais uma amostra de previsão, ficando o conjunto de previsão com 11 amostras. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 10.

O modelo apresentou forte correlação linear ($R^2 = 0,9994$), tendo sua distribuição de amostras no intervalo de concentração (1 a 20%) e distribuição das amostras ao longo da linha de tendência mostradas no Gráfico 15.

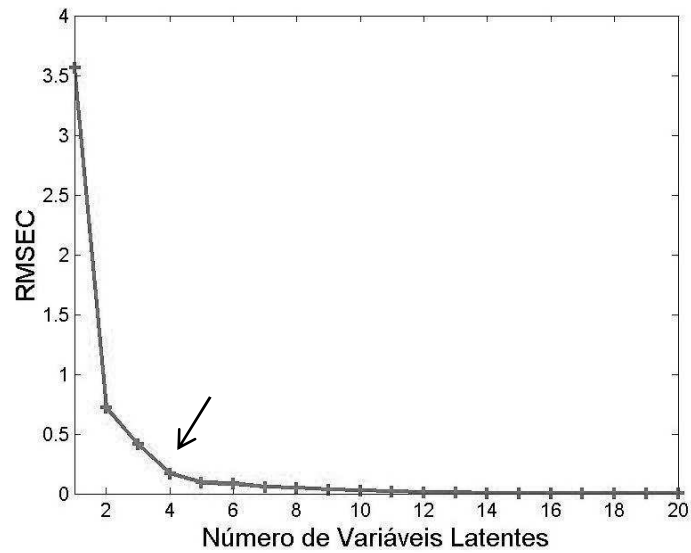


Figura 42 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel metílico de milho.

Tabela 10 - Concentração real e prevista para mistura de biodiesel metílico de milho em diesel.

Valor real	Valor Previsto	Erro (%)
8,1318	7,8699	3,2207
10,0590	9,7746	2,8271
4,9198	4,7393	3,6697
6,4302	6,5098	1,2381
8,2827	8,3419	0,7150
9,9824	9,9350	0,4752
11,4818	11,5020	0,1757
12,7572	12,8800	0,9629
14,6659	14,7324	0,4537
16,1080	15,9983	0,6808
18,0789	18,2517	0,9556

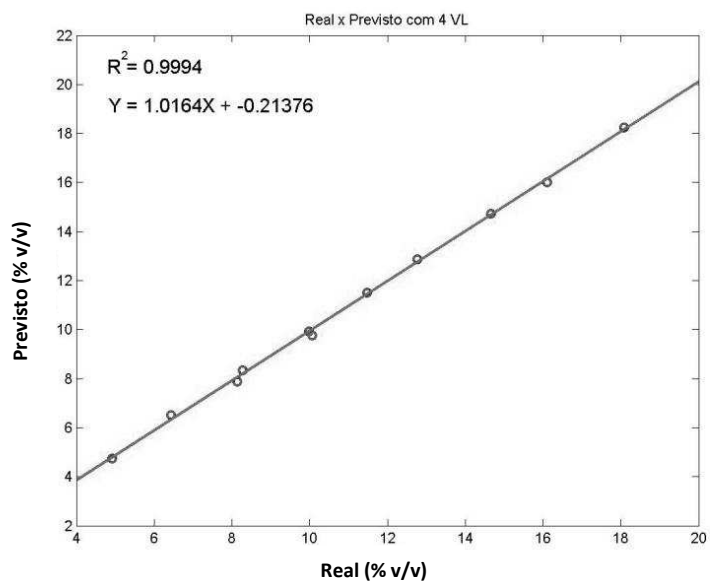


Figura 43 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de milho).

5.6.7. Biodiesel etílico de girassol

Os espectros de Infravermelho Médio para as 40 misturas de biodiesel etílico de girassol em diesel; depois de agrupados, tratados e cortados; estão representados na Figura 44.

O melhor modelo de calibração por PLS foi construído com 12 variáveis latentes (Figura 45), apresentando variância explicada de 99,99% para a matriz de dados e 100% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,0314 e 0,1563. Comparando-se os valores de erro de calibração e previsão, conclui-se que o modelo também possui coerência, logo apresenta erro de calibração (construção do modelo) menor que o erro de previsão.

A média dos erros percentuais dos valores de concentração das misturas de previsão de biodiesel/diesel foi de apenas 1,35%; sendo este erro muito pequeno, mais de sete vezes menor do que o limite estabelecido pela ASTM. A amostra de concentração 5% v/v apresentou erro de apenas 1,37%.

Neste modelo, o conjunto de 30 mostras do conjunto de calibração foi preparado independente do conjunto de 10 amostras de previsão, ambos tendo suas concentrações variando no intervalo de 1 a 20% v/v. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 11.

O modelo apresentou fortíssima correlação linear ($R^2 = 0,9998$), boa distribuição das amostras no intervalo de concentração (1 a 20%) e boa distribuição das amostras ao longo da linha de tendência (Figura 46).

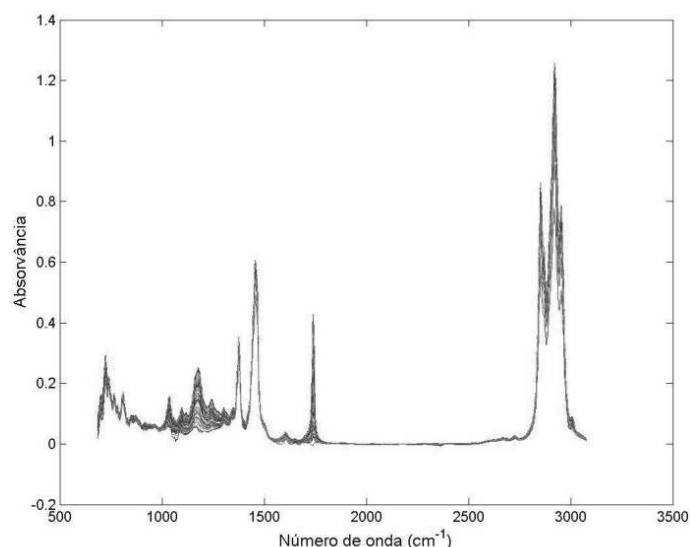


Figura 44 - Espectros MIR para biodiesel etílico de girassol em diesel.

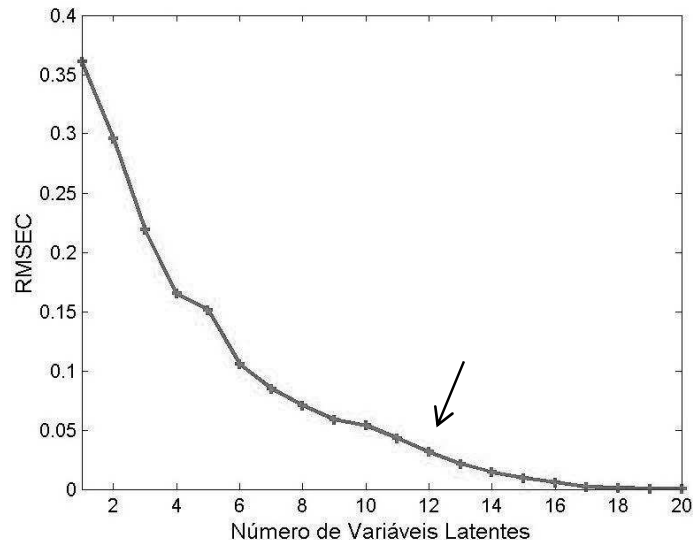


Figura 45 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de girassol.

Tabela 11 - Concentração real e prevista para mistura de biodiesel etílico de girassol em diesel.

Valor real	Valor Previsto	Erro (%)
1,0800	1,0655	1,34
3,1100	3,0174	2,98
5,0100	4,9412	1,37
7,3000	7,2114	1,21
9,4600	9,4332	0,28
11,5500	11,8227	2,36
13,6700	14,0208	2,57
15,7800	15,9137	0,85
17,8500	17,9332	0,47
19,9600	19,9479	0,06

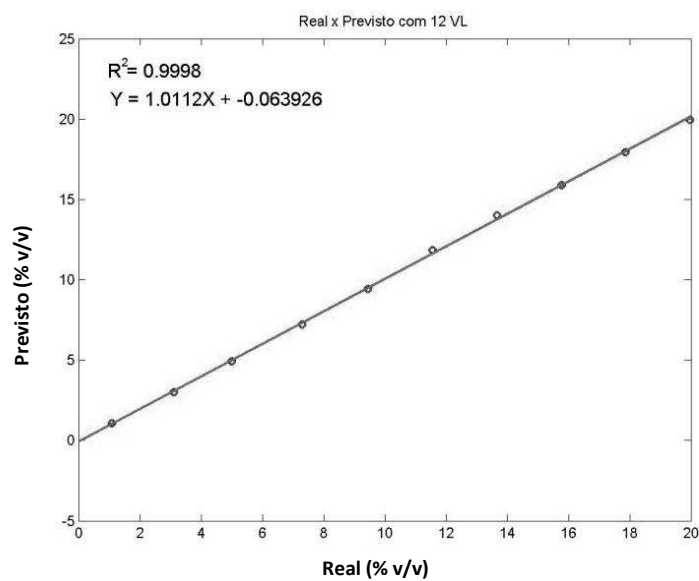


Figura 46 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de milho).

5.6.8. Biodiesel metílico de girassol

Os espectros de Infravermelho Médio para as 40 misturas de biodiesel etílico de girassol em diesel; depois de agrupados numa mesma figura, tratados e cortados; estão representados na Figura 30.

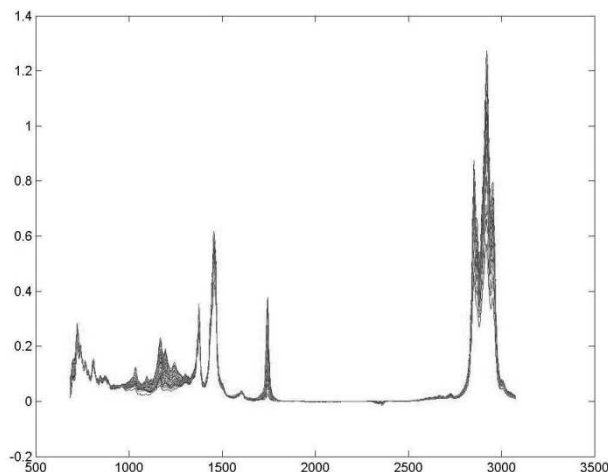


Figura 47 - Espectros MIR para biodiesel metílico de girassol em diesel

O melhor modelo de calibração por PLS foi construído com 7 variáveis latentes (Figura 48), apresentando variância explicada de 99,95% para a matriz de dados e 100% para a matriz de respostas. Os valores de RMSEC e RMSEP para este modelo são, respectivamente; 0,0251 e 0,0521. Comparando-se os valores de erro de calibração e previsão, tem-se que por apresentar erro de calibração (construção do modelo) menor que o erro de previsão, o modelo é coerente. A média dos erros percentuais dos valores de concentração das misturas de previsão foi de apenas 1,12%; sendo este erro muito pequeno, quase nove vezes menor do que o limite estabelecido pela ASTM.

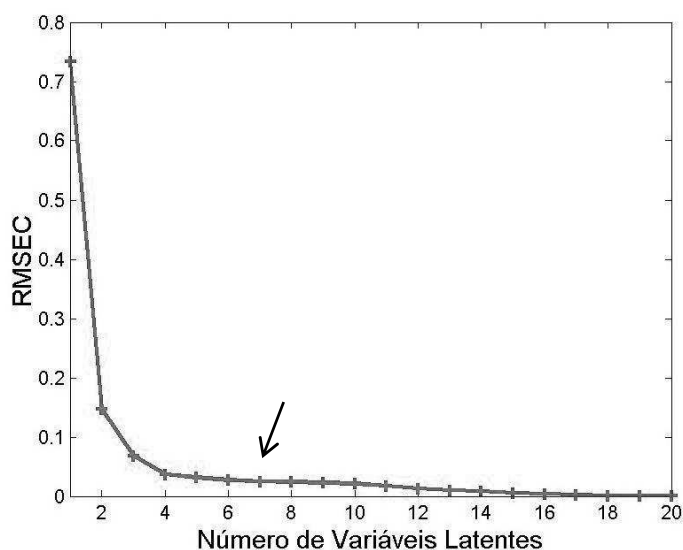


Figura 48 - Gráfico dos valores de RMSEC VS VL para modelo com biodiesel etílico de girassol.

Também neste modelo, o conjunto de 30 mostras do conjunto de calibração foi preparado independente do conjunto de 10 amostras de previsão, ambos tendo suas

concentrações variando no intervalo de 1 a 20% v/v. Os valores de concentração (% v/v) real e previsto e erros percentuais individuais das amostras utilizadas para previsão estão mostrados na Tabela 12.

O modelo apresentou fortíssima correlação linear ($R^2 = 0,9999$), boa distribuição das amostras no intervalo de concentração (1 a 20%) e boa distribuição das amostras ao longo da linha de tendência (Figura 49).

Tabela 12 - Concentração real e prevista para mistura de biodiesel metílico de girassol em diesel.

Valor real	Valor Previsto	Erro (%)
2,3000	2,3470	2,04
2,8900	2,9658	2,62
3,1100	3,1916	2,62
3,3900	3,4135	0,69
4,2000	4,2420	1,00
5,9700	5,9630	0,12
6,5100	6,5567	0,72
7,8000	7,8281	0,36
10,3200	10,3874	0,65
11,9600	12,0113	0,43

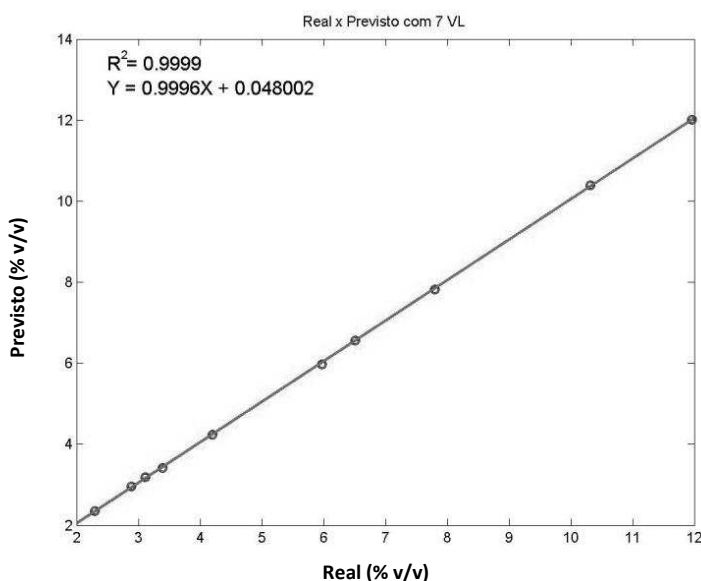


Figura 49 - Gráfico dos valores de concentração (% v/v) reais versus previstos (biodiesel metílico de milho).

5.6.9. Resumo dos modelos

Apesar de utilizar espectros MIR bastante semelhantes, sendo praticamente impossível extrair informações visuais, tanto na correlação do espectro com o tipo de biodiesel adicionado quanto na relação dos perfis com os respectivos valores de concentração de biodiesel na mistura com diesel; todos os modelos de calibração, independente do tipo de biodiesel, foram capazes de estabelecer uma boa relação linear entre os espectros e suas respectivas concentrações (de acordo com os valores de R e

VL). Mostrando que a regressão por PLS foi eficiente na redução das dimensões das variáveis originais sem perder informação relevante.

Todos os modelos de calibração obtidos apresentaram valores de erros muito satisfatórios, com a média de erros relativos entre 0,53% e 2,21%, RMSEC entre 0,0251 e 0,1532% (v/v) e RMSEP entre 0,0521 e 0,2266% (v/v), uma vez que para tal metodologia são aceitos erros de até 10% (padrão ASTM). Vale destacar que em todos os modelos os valores de RMSEC foram sempre menores que os de RMSEP concordante com o critério de que a calibração deve ser mais eficiente do que a etapa de previsão. Lembrando que a ANP regulamenta o teor de biodiesel em diesel em 5% (v/v), nas tabelas destacamos em negrito as concentrações previstas para as amostras com concentração próximas a este limite, onde se verifica que os erros destas previsões ficaram entre 0,12 e 4,24%, bem abaixo do erro aceito para a metodologia. O resumo dos modelos é apresentado na Tabela 13.

Tabela 13 - Resumo dos modelos PLS.

Biodiesel	Rota	RMSEC	RMSEP	Erro médio (%)	VL	R
Milho	etílica	0,0330	0,1681	1,75	9	0,9995
Milho	metílica	0,1532	0,1525	1,40	4	0,9994
Soja	etílica	0,1515	0,2266	1,39	5	0,9996
Soja	metílica	0,0916	0,1609	1,26	6	0,9992
Soja residual	etílica	0,0417	0,0661	0,53	11	0,9999
Soja Residual	metílica	0,1346	0,1487	2,05	5	0,9998
Girassol	etílica	0,0314	0,1563	1,35	12	0,9998
Girassol	metílica	0,0251	0,0521	1,12	7	0,9999

Outro aspecto muito importante na avaliação de um modelo de calibração consiste na análise dos gráficos de valores reais versus os valores previstos pelo modelo para o conjunto de amostras de previsão. A partir dos gráficos apresentados verifica-se que as amostras estão distribuídas uniformemente na faixa de concentração analisada, sendo representativa e não tendenciosa. Também se observa que todos os modelos foram altamente lineares com coeficientes de correlação maiores que 0,999; o que corrobora com o uso de regressão linear para criação dos modelos.

5.6.10. Comparação dos modelos com outros trabalhos publicados

Foram pesquisados, na literatura científica, trabalhos com o objetivo de quantificar biodiesel em diesel, o resumo dos resultados obtidos pelos autores foi agrupado na Tabela 16 para melhor comparação com os resultados obtidos neste trabalho.

Pela análise da Tabela 14 vemos que os modelos obtidos tiveram valores de erros muito bons, mostrando a eficiência da metodologia proposta. O valor de erro entre parênteses [59] representa o erro para a concentração atual da legislação brasileira (o alto erro médio deve-se a previsão de biodiesel em amostras que não o continham – segundo autor do estudo).

Tabela 14 - Comparação dos resultados obtidos com resultados da literatura científica.

Pontos "curva calibração"	Biodiesel	BX analisado	Análise de amostras por	Quantificação por	Erros (%)	Referência
6	Soja (m)	B0,7 B1,9 B4,3	CGAR-EM-MSI ^b	Pico do éster linoleato de metila	4 a 18	[56]
< 200	Soja (m), mamona (m), soja r (m)	B0 a B5	NIR, MID	PLS	10,2	[58]
10	Soja (m)	B0,5 a B5	MID	Área pico carbonila	3,0 a 6,7	[57]
64000	Soja (m)	B0,5 a B20	¹ H NMR ^c	PLS	9,82 (0,4)	[59]
5		B1 e B15	GC-FID ^d		0,47 a 4,00	[60]
1187	Milho (e)	B1 a B20	MID	PLS	1,75	a
1187	Milho (m)	B1 a B20	MID	PLS	1,40	a
1187	Soja (e)	B1 a B20	MID	PLS	1,39	a
1187	Soja (m)	B1 a B20	MID	PLS	1,26	a
1187	Soja r (e)	B1 a B20	MID	PLS	0,53	a
1187	Soja r (m)	B1 a B20	MID	PLS	2,06	a
1187	Girassol (e)	B1 a B20	MID	PLS	1,35	a
1187	Girassol (m)	B1 a B20	MID	PLS	1,12	a

a: presente trabalho

b: cromatografia em fase gasosa acoplada à espectrometria de massas com monitoramento seletivo de íons

c: ressonância magnética nuclear de prótons

d: cromatografia em fase gasosa com detector de ionização por chama

Utilizando apenas as referências que citam os parâmetros apresentados neste trabalho, foi montada a Tabela 15, com resultados de quantificação por PLS. Desta forma podem ser comparados também o Número de Variáveis Latentes, RMSEC, RMSEP, R² e tipo de Pré-Processamento.

Tabela 15 - Resultados obtidos e da literatura para quantificação por PLS.

Pontos de dados	Biodiesel	BX analisado	Análise de amostras por	PLS					Referência
				VL	RMSEC (*)	RMSEP (*)	E (%)	R ²	
< 200	Soja (m), mamona (m), soja residual (m)	B0 a B5	NIR, MID	3	0,28	0,25	10,2	0,986	[58]
64000	Soja (m)	B0,5 a B20	¹ H NMR	3	0,55	0,82	9,82	0,9882	[59]
1187	Milho (e)	B1 a B20	MID	9	0,03	0,17	1,75	0,9995	a
1187	Milho (m)	B1 a B20	MID	4	0,15	0,15	1,40	0,9994	a
1187	Soja (e)	B1 a B20	MID	5	0,15	0,23	1,39	0,9996	a
1187	Soja (m)	B1 a B20	MID	6	0,09	0,16	1,26	0,9992	a
1187	Soja residual (e)	B1 a B20	MID	8	0,09	0,14	2,21	0,9999	a
1187	Soja residual (m)	B1 a B20	MID	5	0,13	0,15	2,06	0,9998	a
1187	Girassol (e)	B1 a B20	MID	12	0,03	0,16	1,35	0,9997	a
1187	Girassol (m)	B1 a B20	MID	7	0,03	0,05	1,12	0,9999	a

a: presente trabalho e: formado pela rota etílica m: formado pela rota metílica

(*) os valores de RMSEC e REMSEP obtidos no presente trabalho foram arredondados a fim de ficarem no mesmo padrão dos apresentados nos artigos.

De acordo com os resultados apresentados na Tabela 17 observa-se que o uso do Infravermelho Médio com PLS leva a modelos mais eficientes: com menor erro médio e maior linearidade (muito desejável, já que o PLS faz uso de regressão linear). Sendo tal metodologia mais simples (análise direta não destrutiva) e apresenta menor custo (em relação à compra de equipamento e o uso de ^1H NMR).

CAPÍTULO 6: Conclusão

6. Conclusão

No presente trabalho, foram desenvolvidas metodologias baseadas na utilização da técnica de espectrometria na região do Infravermelho Médio (MIR) aliada ao método de calibração por Quadrados Mínimos Parciais (PLS) visando à quantificação de biodiesel de soja, milho e girassol, oriundos das rotas etílica e metílica, em diesel.

De acordo com os resultados apresentados, constatou-se que os modelos de calibração por PLS apresentaram valores de erros satisfatórios (bem abaixo do limite de 10% estabelecido na ASTM), alta linearidade (com valores de R^2 superiores a 0,999) e eficiência na previsão das concentrações das amostras, mesmo quando comparados com técnica mais moderna ($^1\text{H NMR}$).

Como os modelos que relacionam medida instrumental com medida de análises clássicas podem, conforme a faixa de erros admitida, serem utilizados em laboratórios de controle de qualidade reduzindo tempo de análise, custo com reagentes, quantidade de resíduos, mão de obra, entre outros. As metodologias desenvolvidas neste trabalho apresentam-se como alternativa viável no controle de qualidade do diesel comercializado quanto ao teor de biodiesel na mistura estabelecido pela ANP.

É importante ressaltar que, de acordo com os resultados apresentados, vislumbra-se a perspectiva de estudos futuros dentro do grupo de pesquisa, aplicando outras matérias-primas para produção de biodiesel e métodos de calibração multivariada com seleção de variáveis, tentando encontrar regiões espectrais que contenham informações mais relevantes para o modelo e conseqüentemente eliminação das variáveis supostamente de menor relevância e/ou interferência, ganhando maior eficiência. Este conhecimento aprimorado permitirá o desenvolvimento de equipamentos com detectores mais simples, portáteis e de menor custo; passíveis de serem utilizados em monitoramento "on line".

CAPÍTULO 7: Bibliografía

1. NATIONAL BIODIESEL BOARD. Definição de biodiesel. Disponível em: <<http://www.biodiesel.org>>. Acesso em: 12 abr. 2010.
2. BRASIL. *Lei nº 11.097/2005*. Introdução do biodiesel na matriz energética brasileira. Disponível em: <<http://www.planalto.gov.br/leg.asp>>. Acesso em: 10 abr. 2010.
3. AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP, Resolução nº 7, De 19 de março de 2008. Regulamenta inserção do biodiesel na matriz energética brasileira. Disponível em: <<http://www.anp.gov.br>>. Acesso em: 16 ago. 2009.
4. Conselho Nacional de Política Energética – CNPE, Resolução nº 2, de 13 de março de 2008. Reduz o prazo de que trata o § 1º do art. 2º da Lei nº 11.097, de 13 de janeiro de 2005, e dá outras providências. Disponível em: <<http://www.biodiesel.gov.br/legislacao.html>>. Acesso em 15 fev. 2010.
5. Conselho Nacional de Política Energética – CNPE, Resolução nº 2, de 27 de abril de 2009. Altera o percentual de adição de biodiesel ao diesel para a faixa de 4 a 5% e estipula a obrigatoriedade da adição de 5% a partir de 01 de janeiro de 2010. Disponível em: <<http://www.biodiesel.gov.br/legislacao.html>>. Acesso em 55 mar. 2010.
6. Comércio, M.d.l.e.d., *Produção de Combustíveis Líquidos a Partir de Óleos Vegetais*, Industrial, S.d.T., Editor. 1985, Coordenadoria de Informações Tecnológicas: Brasília.
7. BOARD, N.B., ed. *Anais do Congresso Internacional de Biocombustíveis Líquidos*. 1998, Secretaria de Estado da Ciência: Curitiba.
8. FREEDMAN, B., BUTTERFIELD, R.O., e Pryde, E.H., J. Am. Oil Chem. Soc, 1986. **63**(1375).
9. OIL WORLD, *Demand and Prices from 1976 through 2020*. Disponível em: <<http://www.oilworld.biz>>. Acesso em: 17 out. 2009.
10. ASSOCIAÇÃO BRASILEIRA DAS INDÚSTRIAS DE ÓLEOS VEGETAIS. *Produção Nacional de Soja*. Disponível em: <<http://www.abiove.com.br>>. Acesso em: 23 fev. 2010.
11. AMERICAN SOYBEAN ASSOCIATION. A reference guide to important soybean facts and figures. Disponível em: <<http://soystats.com>>. Acesso em: 10 set. 2010.
12. CIRAD, *Guide technique pour une utilisation énergétique des huiles végétales*. 2008, Patrick Rousset: Brasília. p. 288.
13. COSTA NETO, P.R., et al., *Produção de biocombustível alternativo ao óleo diesel através da transesterificação de óleo de soja usado em frituras*. Química Nova, 2000. **23**: p. 531-537.
14. MITTELBACH, M. e TITTHART, P., J. Am. Oil Chem. Soc., 1988. **65**(1185).
15. LAURINDO, J.C., *Anais do Congresso Internacional de Biocombustíveis Líquidos*. 1998, Secretaria de Estado da Ciência: Curitiba.
16. AGÊNCIA NACIONAL DO PETRÓLEO, GÁS NATURAL E BIOCOMBUSTÍVEIS – ANP, Resolução nº 4, De 02 de fevereiro de 2010. estabelece a especificação do biodiesel a ser comercializado pelos diversos agentes econômicos autorizados em todo o território nacional. Disponível em: <<http://www.anp.gov.br>>. Acesso em: 06 fev. 2010.
17. SKOOG, D.A., HOLLER, F.J., NIEMAN, T.A., *Princípios de análise instrumental*, ed. 5. 2002, São Paulo: Bookman.
18. SCHRADER, B., *Infrared and Raman Spectroscopy: methods and applications*. 1995, Weinheim: VCH.
19. BURNS, D.A. e CIURCZAK, E.W., *Handbook of Near-Infrared Analysis*. 2001, New York: Marcel Dekke.
20. LEVINE, I.N., *Molecular Spectroscopy*. 1975, New York: John Wiley & Sons Inc. 491.
21. COATES, J., *Interpretation of Infrared Spectra, A Practical Approach, Encyclopedia of Anal. Chem.* 2000, Chichester: John Wiley & Sons Inc.
22. WORKMAN Jr, J.J., *Interpretative spectroscopy for near infrared*. Appl. Spectrosc., 1996. 31: p. 251.
23. OLIVEIRA, C. C. F., BRANDÃO, R. R. C., RAMALHO, F. H., DA COSTA, A. F. L., SUAREZ, A. Z. P., RUBIM, C. J., *Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy*. Analytica Chimica Acta, 2007. 587(2): p. 194-199.
24. DE SOUZA, M .L., CORIO, P., TEMPERINI, M. L. A., TEMPERINI, J. A., *Aplicação de espectroscopias raman e Infravermelho na identificação e quantificação de plastificantes em filmes comerciais de PVC esticável*. Química Nova, 2009. 32: p. 1452-1456.

25. CANTARELLI, M. A.; FUNES I. G.; MARCHEVSKY, E. J. CAMIÑA, J. M., *Determination of oleic acid in sunflower seeds by infrared spectroscopy and multivariate calibration method*. Talanta, 2009. **80**(2): p. 489-492.
26. ARYEE, A.N.A., VAN DE VOORT, F.R., SIMPSON, B.K., *FTIR determination of free fatty acids in fish oils intended for biodiesel production*. Process Biochemistry (Amsterdam, Netherlands), 2009. 44(4): p. 401-405.
27. BORIN, A.; POPPI, R. J., *Application of mid infrared spectroscopy and iPLS for the quantification of contaminants in lubricating oil*. Vibrational Spectroscopy, 2005. **37**(1): p. 27-32.
28. SILVERSTEIN, R.M., WEBSTER, F.X., KIEMLE, D.J., *Spectrometric Identification of Organic Compounds*, ed. 7. 2005, New Jersey: John Wiley and Sons Inc. 512.
29. HARRICK, N.J., *Multiple Reflection Cells for Internal Reflection Spectrometry*. Anal. Chem., 1964. 36: p. 188-191.
30. D'ANTONIO, M.C., *Spectroscopic investigations of iron(II) and iron(III) oxalates*. Journal of the Brazilian Chemical Society, 2009. 20: p. 445-450.
31. FOSCHIERA, J.L., PIZZOLATO, T.M., BENVENUTTI, E.V., *FTIR thermal analysis on organofunctionalized silica gel*. Journal of the Brazilian Chemical Society, 2001. 12: p. 159-164.
32. BARAN, E.J., WAGNER, C.C., TORRE, M.H., *Synthesis and Characterization of EDTA Complexes Useful for Trace Elements Supplementation*. Journal of the Brazilian Chemical Society, 2002. 13: p. 576-582.
33. BARBOSA, L.C.A., *Espectroscopia no Infravermelho na caracterização de compostos orgânicos*. 2007, Viçosa: UFV.
34. BRUNS, R.E.; FAIGLE, J.F.G., *Quimiometria*. Química Nova, 1995. 8: p. 84-99.
35. BORGES NETO, W., *Aplicação de métodos de Reconhecimento de Padrões para classificação de Óleos Vegetais*, in Instituto de Química. 2001, Universidade Federal de Goiás: Goiânia. p. 74.
36. OTTO, M., *Chemometrics – Statistics and Computer application in Analytical Chemistry*. 1999, Weinheim: John Wiley & Sons Inc.
37. BARROS NETO, B.D., SCARMINIO, I.S., BRUNS, R.E., *25 anos de quimiometria no Brasil*. Química Nova, 2006. 29: p. 1401-1406.
38. VANDEGINSTE, B.G.M., MASSART, D. L. , BUYDENS, L. M. C. *Buydens, Handbook of Chemometrics and Qualimetrics: Part A, Part B*. 1998, Amsterdam: Elsevier.
39. BEEBE, K.R., PELL, R.J., SEASHOLTZ, M.B., *Chemometrics: A practical guide*. 1998, New York: John Wiley & Sons Inc.
40. MILLER, J.N., MILLER, J.C., *Statistics and Chemometrics for Analytical Chemistry*. 2000, London: Prentice Hall.
41. MARTENS, H.; NAES, T., *Multivariate calibration*. 1996, New York: John Wiley & Sons Inc.
42. JOHNSON, R.A.; WICHERN, D.W., *Applied Multivariate Statistical Analysis*, ed. 5. 2002, New Jersey: Prentice Hall. 767.
43. BRERETON, R.G., *Introduction to multivariate calibration in analytical chemistry*. Analyst, 2000. 125: p. 2125.
44. AFIFI, A.A., AZEN, S.P., *Statistical Analysis - A computer oriented approach*, ed. 2. 1979, San Diego: Academic Press INC.
45. HELLAND, I.S., *On the structure of partial least square regression*. Communications in statistics – simulation and computation, 1998. 17.
46. SEKULIC, S., SEASHOLTZ, M. B., WANG, Z.Y., KOWALSKI, B. R., *Nonlinear multivariate calibration methods in analytical chemistry*. Anal. Chem., 1993. 65(19): p. A835-A845.
47. CHARNET, R., *et al.*, *Análise de Modelos de Regressão Linear com Aplicações*. 1999, Campinas: Unicamp.
48. GELADI, P., KOWALSKI, B.R., *Partial Least Squares Regression: A Tutorial*. Anal. Chim. Acta, 1986. 1: p. 17.
49. PERALTA-ZAMORA, P.; CORNEJO-PONCE, L.; NAGATA, N.; POPPI, R. J., *Alternativas quimiométricas para a resolução de problemas analíticos clássicos: determinação espectrofotométrica de misturas de zircônio e háfnio*. Química Nova, 1997. 20: p. 469-474.

50. FELÍCIO, C. C.; BRÁS, L. P.; LOPES, J. A.; CABRITA L.; MENEZES, J. C., *Comparison of PLS algorithms in gasoline and gas oil parameter monitoring with MIR and NIR*. Chemometrics and Intelligent Laboratory Systems, 2005. 78(1-2): p. 74-80.
51. MONTEIRO, M.R., *et al.*, *Determination of biodiesel blend levels in different diesel samples by ¹H NMR*. Fuel, 2009. 88(4): p. 691-696.
52. RAMBLA, F.J., GARRIGUES, S., DE LA GUARDIA, M., *PLS-NIR determination of total sugar, glucose, fructose and sucrose in aqueous solutions of fruit juices*. Analytica Chimica Acta, 1997. 344(1-2): p. 41-53.
53. OLIVEIRA-ESQUERRE, K.P., MORI, M., BRUNS, R.E., *Simulation of an industrial wastewater treatment plant using artificial neural networks and principal components analysis*. Brazilian Journal of Chemical Engineering, 2002. 19: p. 365-370.
54. BHATTACHARYA, G.K., JOHNSON, R.A., *Statistical Concepts and Methods*. 1977, Singapura: John Wiley & Sons Inc. 639.
55. BARTHUS, R.C., MAZO, L.H., POPPI, R.J., *Aplicação de métodos de calibração multivariada para a determinação simultânea de riboflavina (VB2), tiamina (VB1), piridoxina (VB6) e nicotinamida (VPP)*. Química Nova, 2007. 30: p. 1638-1643.
56. FARIA, R.C.M.; REZENDE, M. J. C. ; REZENDE, C. M.; PINTO, A. C., *Desenvolvimento e validação de metodologia de análise de misturas biodiesel: diesel utilizando cromatografia gasosa-espectrometria de massas*. Química Nova, 2007. 30: p. 1900-1905.
57. RIBEIRO, N. M.; GUARIEIRO, L. L. N.; PINTO, A. C.; AGUIAR, P. F., *Metodologia analítica para quantificar o teor de biodiesel na mistura biodiesel: diesel utilizando espectroscopia na região do Infravermelho*. Química Nova, 2008. 31: p. 421-426.
58. FERNANDA PIMENTEL, M.; RIBEIRO, G.M.G.S.; DA CRUZ, R.S.; STRAGEVITCH, L.; PACHECO FILHO, J.G.A.; TEIXEIRA, L.S.G.; *Determination of biodiesel content when blended with mineral diesel fuel using infrared spectroscopy and multivariate calibration*. Microchemical Journal, 2006. 82(2): p. 201-206.
59. MONTEIRO, M. R., AMBROZIN, A. R. P., SANTOS, M. S., BOFFO, E. F., PEREIRA-FILHO, E. R., LIÃO, L. M., FERREIRA, A. G., *Evaluation of biodiesel-diesel blends quality using ¹H NMR and chemometrics*. Talanta, 2009. 78(3): p. 660-664.
60. RAGONESE, C.; TREANCHIDA, P.Q.; SCIARRONE, D.; MONDELLO, L., *Conventional and fast gas chromatography analysis of biodiesel blends using an ionic liquid stationary phase*. Journal of Chromatography A, 2009. 1216(51): p. 8992-8997.
61. CAIXETA, G., *Estudo de Cenário de Produção de Biodiesel*, Dissertação de Mestrado - Instituto de Engenharia Química, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2009.