

**UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE QUÍMICA
DEPARTAMENTO DE QUÍMICA ANALÍTICA
LAQQA – LABORATÓRIO DE QUIMIOMETRIA EM QUÍMICA ANALÍTICA**



UNICAMP

**PARÂMETROS DE QUALIDADE DE LUBRIFICANTES E ÓLEO DE
OLIVA ATRAVÉS DE ESPECTROSCOPIA VIBRACIONAL,
CALIBRAÇÃO MULTIVARIADA E SELEÇÃO DE VARIÁVEIS**

TESE DE DOUTORADO

WALDOMIRO BORGES NETO

ORIENTADOR: Prof. Dr. RONEI JESUS POPPI

CAMPINAS – SP, AGOSTO DE 2005

**FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DO INSTITUTO DE QUÍMICA
DA UNICAMP**

B644p Borges Neto, Waldomiro.
 Parâmetros de qualidade de lubrificantes e óleo de
 oliva através de espectroscopia vibracional, calibração
 multivariada e seleção de variáveis / Waldomiro Borges
 Neto. -- Campinas, SP: [s.n], 2005.

 Orientador: Ronei Jesus Poppi

 Tese – Universidade Estadual de Campinas, Instituto
 de Química.

 1. Lubrificante. 2. Óleo de oliva. 3. Espectroscopia
 vibracional. 4. Seleção de variáveis. I. Poppi, Ronei
 Jesus. II. Universidade Estadual de Campinas. Instituto
 de Química. III. Título.

Louco é aquele que se considera sábio
quando mede a extensão de sua ignorância.

O sábio traz a pequenina migalha, apanhada no caminho,
e julga conduzir o próprio Himalaia.

A ciência é uma grande montanha de açúcar;
dessa montanha só conseguimos retirar insignificantes pedacinhos.

(Trechos do livro - O Homem que Calculava – Malba Tahan)

Dedico este trabalho em especial à minha amada esposa

Simone Soares de Oliveira Borges

a meus queridos e amados pais

João Dorival Borges e Ivonice Benedita Borges

a meus queridos e amados sogros

Lincoln Soares de Oliveira e Ercília Marciano de Oliveira

aos meus irmãos

Waléria Fabiane Borges e Wagner Lúcio Borges

e sobrinhas

Renata Fabiane Borges Rocha

Gabrielle de Oliveira Penharbel

Manuelle de Oliveira Penharbel

AGRADECIMENTOS

**Portanto a sabedoria entrará no
teu coração, e o conhecimento
será suave à tua alma.**

Pv. 2:10

- A Deus pela essência da vida, meus familiares, amigos e, em especial, pela possibilidade de realizar este trabalho.
- A minha amada esposa Simone, pelo amor, incentivo, paciência e por acreditar em nossos sonhos.
- Ao Prof. Dr. Ronei Jesus Poppi pela oportunidade de realização desse trabalho, orientação, conhecimentos transmitidos, paciência, confiança, disponibilidade e pela amizade durante todo este trabalho. E dizer-te que, para o semeador, a melhor recompensa são os bons frutos gerados pelo seu trabalho, que lhe fornecerão sementes para novas colheitas. 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 grupo LAQQA: Alessandra Borin, Luiz Carlos Moutinho Pataca, Danilo Althmann Mareto, Jez William Batista Braga, Marcelo Garcia Trevisan, Gilmare Antônia da Silva, Patrícia Valderrama, Rosineide Costa Simas, Paulo Henrique Março, Luciana Viviani, Silvio Luis T. Lima, Joana Guilares de Aguiar, Tiago Pucca Araújo e Rafael Antonio Ozzeti. pelos agradáveis momentos vividos, a verdadeira amizade, as viagens, comemorações, muito trabalho, discussões, conselhos, risos e, histórias que guardarei eternamente na memória. Na tentativa de expressar o significado desta amizade, deixo estas singelas palavras: “A amizade é o sentimento que une as almas umas às outras, equilibrando as mentes, dulcificando os corações. Inspiradora de coragem e de abnegação, a amizade enflorêce as almas, abençoando-as com resistências para as lutas.”

- Agradecimento especial aos colegas Alessandra, Luiz e Danilo por toda ajuda nas análises de lubrificantes e nos programas computacionais desenvolvidos.
- À Unicamp por fornecer toda estrutura física e tecnológica para a realização deste trabalho.
- Ao CNPq pelo financiamento do projeto.
- As empresas Wal Química, AGIP Lubrificantes, Rápido Luxo Campinas e Replan Petrobrás, pelo auxílio, dedicação, atenção, doação dos lubrificantes e combustíveis, coleta das amostras de lubrificantes de ônibus e realização das análises químicas.
- À Universidade de Franca e a Faculdade de Engenharia de Alimentos da Unicamp pelo uso dos espectrofotômetros Raman e NIR.
- Aos meus amados pais, João e Ivonice, irmãos, Waléria e Wagner, sobrinhas, Renata, Gabrielle e Manuelle, sogros, Lincoln e Ercília, minhas cunhadas, Silvia e Ana Paula e, toda minha família, por todo amor, apoio, incentivo e compreensão durante toda minha vida.
- Ao amigo e orientador do mestrado Prof. Dr. Paulo Sérgio de Souza por tudo aquilo que representa como uma pessoa muito especial, sendo um exemplo a ser seguido por mim e um conselheiro para todos os momentos.
- Aos Profs. do Instituto de Química da UFG, Ms. Celso Machado, Ms. Maria Gizelda de O. Tavares, Dr. Edésio F. de C. Alcântara e Dr. Lourival C. de Faria, pelo exemplo, incentivo, dedicação e acima de tudo pela amizade.
- A Prof^a Dra. Maria Izabel Maretti Silveira Bueno pela orientação da minha esposa, incentivo e acima de tudo pela amizade.
- Aos meus eternos amigos: Emerson (pesão), Nivanil, Biatriz, Marcelo (bumbeiro), Jaime (macarrão), Renato (lobó), João Jr., Alexandre (primo), Sandra (cabelo de mola), Mário, Sílvia, Divina, Maurílio (veio), Paulinho, Júlio (estrela), Lílian, Olacir, Ivone, Paulo (goiano), Victor (marmelo), Luíza Reino, Lílian Braga e Daleth.
- Aos amigos de Unicamp: Rafael Arromba, Ernesto (tchê), Aloísio (Virgú), Mauro (TC), Divino (TC), Ígor (TC), Gustavo (TC), José Carlos (CG),

Manoel (mané), André (CPG), Prof. Dr. Luiz Carlos Dias, Bel (CPG), Guilherme (rato), Odair, Mário, Eliane, Alaécio, Antônio (jequié), Arnaldo, Teófilo, Emerson, Marcelo Lanza, Prof. Dr. Wilson de Figueiredo Jardim, Gesley, Fernando Barboza, César Alexandre de Mello, Eduardo Osório de Cerqueira, Marcelo Sena, Leoberto Balbinot, Cleidiane Gorete Zampronio, Paulo Henrique Fidêncio e tantos outros.

- A todos aqueles que direta ou indiretamente tiveram sua parcela de participação durante a execução e conclusão deste trabalho.

CURRICULUM VITAE

Dados Pessoais

Nome – Waldomiro Borges Neto
Nascimento – 03/09/1970
Naturalidade – Trindade-GO
E-mail – tago@iqm.unicamp.br

Formação

Graduação – Bacharelado em Química – 1991-1996.
UFG – Universidade Federal de Goiás – Goiânia-GO.
Mestrado em Química – 1999-2001.
UFG – Universidade Federal de Goiás – Goiânia-GO.

Experiência Profissional

Universidade Estadual de Goiás – UEG - 2000–2001.
Função - Professor substituto de Química Analítica e Instrumental.
Escola Senai Roberto Mange – Anápolis-GO - 1999–2001.
Função – professor de ensino técnico.
Centro de Educação Alfa Beta – Goiânia – GO - 1996–1999.
Função – Professor de ensino médio.
Colégio Solução – Goiânia-GO - 1994–1999.
Função – Professor de ensino médio.
Secretaria de Estado da Educação – GO - 1989–1995.
Função – Professor de ensino médio.

Atividades Acadêmicas

Monitoria Acadêmica – Programa de Estágio Docente II – 2/2004 - 7/2004.
Disciplina – QA-581 – Métodos em Química Analítica Instrumental.
Instituto de Química – UNICAMP.
Monitoria Acadêmica – Programa de Estágio Docente I – 8/2003 - 11/2003.
Disciplina – QA-313 – Química III (Engenharia Química).
Instituto de Química – UNICAMP.
Monitoria Acadêmica – Programa de Estágio Docente I – 2/2003 - 7/2003.
Disciplina – QA-581 – Métodos em Química Analítica Instrumental.
Instituto de Química – UNICAMP.

Trabalhos em Eventos Nacionais

- BORGES NETO, Waldomiro; MELLO, Cesar; POPPI, Ronei Jesus. Adulteração de óleo de oliva por espectroscopia raman e calibração multivariada. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.
- PATACA, Luiz Carlos Moutinho; BORGES NETO, Waldomiro; MARCUCCI, Maria Cristina; POPPI, Ronei Jesus. Determinação da cor de mel de abelhas utilizando análise multivariada de imagens. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.
- BORGES NETO, Waldomiro; MARETTO, Danilo Athmann; POPPI, Ronei Jesus. Determinação de etileno glicol em óleo lubrificante por espectroscopia no infravermelho e PLS. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.
- PATACA, Luiz Carlos Moutinho; BORGES NETO, Waldomiro; MARCUCCI, Maria Cristina; POPPI, Ronei Jesus. Determinação do teor de açúcares redutores em mel de

- abelhas por Espectroscopia no Infravermelho Médio e iPLS. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.
- BORGES NETO, Waldomiro; POPPI, Ronei Jesus. Espectroscopia no infravermelho e seleção de variáveis para determinar adulteração de óleo de oliva. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.
 - MELLO, Cesar; RIBEIRO, Diórginis; NOVAES, Fábio; BORGES NETO, Waldomiro; POPPI, Ronei Jesus. Identificação de bactérias utilizando espectroscopia Raman e PLS-discriminante. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.
 - MELLO, Cesar; RIBEIRO, Diórginis; NOVAES, Fábio; BORGES NETO, Waldomiro; POPPI, Ronei Jesus. Rápida determinação de bactérias utilizando espectroscopia de correlação em 2D. In: 28ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2005, Poços de Caldas - MG.
 - BORGES NETO, Waldomiro; POPPI, Ronei Jesus. Estudo de adulteração de Óleo de Oliva por Espectroscopia no Infravermelho e PLS por Intervalos. In: XXVI CONGRESSO LATINOAMERICANO DE QUÍMICA E 27ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2004, Salvador-BA.
 - BORGES NETO, Waldomiro; TREVISAN, Marcello G; POPPI, Ronei Jesus. Avaliação de figuras analíticas de mérito para Calibração com dados Multidimensionais: Um estudo de caso. In: 12º ENCONTRO NACIONAL DE QUÍMICA ANALÍTICA, 2003, São Luís - MA.
 - BORGES NETO, Waldomiro; SOUSA, Rafael A de; POPPI, Ronei Jesus; CADORE, Solange; BACCAN, Nivaldo. Classificação de Água de Coco através de HCA, PCA e Teores de metais determinados por ICP-OES. In: 12º ENCONTRO NACIONAL DE QUÍMICA ANALÍTICA, 2003, São Luís - MA.
 - BORGES NETO, Waldomiro; SOUZA, Paulo Sérgio de; ANTONIOSI FILHO, Nelson Roberto. Aplicação de Métodos de Reconhecimento de Padrões para classificação de óleos vegetais de Oliva. In: 11º ENCONTRO NACIONAL DE QUÍMICA ANALÍTICA, 2001, Campinas-SP.
 - BORGES NETO, Waldomiro; TEIXEIRA, André Luiz; SARTORATTO, Patrícia C; BARBOSA FILHO, Morel Pereira. Análise de propriedades químicas de solos por Espectroscopia no Infravermelho Médio e Calibração Multivariada. In: XL CONGRESSO BRASILEIRO DE QUÍMICA, 2000, Recife-PE.
 - BORGES NETO, Waldomiro; SOUZA, Paulo Sérgio de; ANTONIOSI FILHO, Nelson Roberto. Aplicação de métodos de Reconhecimento de Padrões para a classificação de Óleos Vegetais, utilizando análise Triacilglicéridica. In: XL CONGRESSO BRASILEIRO DE QUÍMICA, 2000, Recife-PE.
 - BORGES NETO, Waldomiro; SOUZA, Paulo Sérgio de; ANTONIOSI FILHO, Nelson Roberto. Classificação de Óleos Vegetais através das Composições de Ácidos Graxos, aplicando métodos de Reconhecimento de Padrões. In: 23ª REUNIÃO ANUAL DA SOCIEDADE BRASILEIRA DE QUÍMICA, 2000, Poços de Caldas.

RESUMO

PARÂMETROS DE QUALIDADE DE LUBRIFICANTES E ÓLEO DE OLIVA ATRAVÉS DE ESPECTROSCOPIA VIBRACIONAL, CALIBRAÇÃO MULTIVARIADA E SELEÇÃO DE VARIÁVEIS

Autor: Waldomiro Borges Neto

Orientador: Ronei Jesus Poppi

Os óleos lubrificantes, geralmente obtidos do petróleo, consistem de uma mistura complexa de hidrocarbonetos de massas molares entre 250-1.000 g mol⁻¹. Sendo a qualidade do óleo lubrificante essencial na preservação do desempenho e longevidade de motores e máquinas, parâmetros de qualidade de lubrificantes como viscosidade a 100 °C, número total de base (TBN) e contaminantes (diesel, gasolina, água e etileno glicol), foram determinados por Espectroscopia no Infravermelho Médio, calibração multivariada por Mínimos Quadrados Parciais (PLS) e métodos de seleção de variáveis como PLS por Intervalos (iPLS), Eliminação de Variáveis Não Informativas por PLS (UVE-PLS) e Algoritmo Genético (GA). Métodos de Espectroscopia Vibracional, aliados aos mesmos métodos de calibração e seleção de variáveis foram utilizados na determinação de contaminantes em óleos de Oliva extra virgem e puro, com óleos vegetais de Algodão, Canola, Girassol, Milho e Soja, com concentrações variando entre 1,00 a 20,00% (m/m). Os modelos obtidos foram avaliados e comparados através do erro quadrático médio de previsão (RMSEP) e média do erro relativo (EM). Os modelos para os parâmetros de qualidade dos lubrificantes apresentaram RMSEPs entre 8,70x10⁻³ e 3,54x10⁻¹ e EMs entre 0,60 e 11,43%, enquanto que nas contaminações do óleo de Oliva RMSEPs entre 6,30x10⁻³ e 5,58x10⁻¹% e EMs entre 0,04 e 5,63% foram obtidos. Os resultados foram satisfatórios, com perspectivas de que as metodologias desenvolvidas neste trabalho possam ser utilizadas como uma alternativa viável no controle de qualidade de lubrificantes e óleos de Oliva.

ABSTRACT

QUALITY PARAMETERS OF LUBRICANTS AND OLIVE OIL THROUGH VIBRATIONAL SPECTROSCOPY, MULTIVARIATE CALIBRATION AND VARIABLE SELECTION

Author: Waldomiro Borges Neto

Adviser: Ronei Jesus Poppi

Lubricating oils usually obtained from petroleum, consist of complex mixtures of hydrocarbons of molar masses among 250-1000 g mol⁻¹. Being the quality of lubricating oil essential for preservation of the performance and longevity of motors and machines, lubricant quality parameters as viscosity at 100°C, total base number (TBN) and contaminants (diesel, gasoline, water and ethylene glycol), were determined by Mid infrared spectroscopy, partial least-squares (PLS) for multivariate calibration and variable selection methods as interval PLS (iPLS), uninformative variable elimination by PLS (UVE-PLS) and Genetic Algorithm (GA). Vibrational spectroscopy methods allies to the same calibration and variable selection methods were used by determination of contaminants in extra virgin and refined olive oils, with vegetable oils of cotton, canola, sunflower, corn and soybean, with concentration in the range 1.00-20.00% (w/w). The obtained models were appraised and compared through the root mean square error of prediction (RMSEP) and relative standard error (EM). The models for the lubricant quality parameters presented RMSEPs between 8.70×10^{-3} and 3.54×10^{-1} and EMs between 0.60 and 11.43%, while in the olive oil contamination RMSEPs between 6.30×10^{-3} and 5.58×10^{-1} % and EMs between 0.04 and 5.63% were obtained. The results were suitable with perspectives in the methodologies developed in this work be used as viable alternative in the quality control of lubricants and olive oils.

LISTA DE ABREVIATURAS OU SIGLAS

ATR - Reflectância total atenuada;
CV – Validação cruzada;
EM - média do erro relativo;
FIR – Infravermelho distante;
FT – transformada de Fourier;
FTIR – Infravermelho com transformada de Fourier;
FTRaman – Raman com transformada de Fourier;
GA – algoritmo genético;
GA-PLS – algoritmo genético com PLS;
GA-iPLS - algoritmo genético usando como entrada as variáveis do iPLS;
GA-UVE-PLS - algoritmo genético usando como entrada as variáveis do UVE-PLS;
iPLS – mínimos quadrados parciais por intervalos;
MIR – Infravermelho médio;
MLR – Regressão Linear Múltipla;
NIR – Infravermelho próximo;
PC - componente principal;
PCA – análise de componentes principais;
PCR – Regressão por componentes principais;
PLS – mínimos quadrados parciais;
PLSR - regressão por PLS;
RMSECV - erro quadrático médio da validação cruzada;
RMSEP - erro quadrático médio de previsão;
savgol – filtro Savitsky-Golay;
TAN - número total de acidez;
TBN - número total de basicidade;
UV/Vis – ultravioleta/visível;
UVE-PLS – eliminação de variáveis não informativas por PLS;
VL- variável latente.

LISTA DE TABELAS

Tabela 1. Regiões espectrais do infravermelho.....	14
Tabela 2. Resultados em ordem decrescente das análises de TBN (mg KOH/g) das 45 amostras de lubrificantes de ônibus.	61
Tabela 3. Resultados para os modelos de calibração multivariada de lubrificante - TBN.....	62
Tabela 4. Resultados em ordem crescente das análises de Viscosidade a 100°C (cSt) das 47 amostras de lubrificantes de ônibus.....	64
Tabela 5. Resultados para os modelos de calibração multivariada de lubrificante - Viscosidade.....	65
Tabela 6. Resultados para os modelos de calibração multivariada de lubrificante - diesel.....	68
Tabela 7. Resultados para os modelos de calibração multivariada de lubrificante - gasolina.....	70
Tabela 8. Resultados para os modelos de calibração multivariada de lubrificante - água.....	73
Tabela 9. Resultados para os modelos de calibração multivariada de lubrificante - etileno.....	75
Tabela 10. Resultados para os modelos de calibração multivariada de Oliva – Algodão no infravermelho médio.....	87
Tabela 11. Resultados para os modelos de calibração multivariada de Oliva – Algodão no infravermelho próximo.....	89
Tabela 12. Resultados para os modelos de calibração multivariada de Oliva – Algodão no Raman.....	91
Tabela 13. Resultados para os modelos de calibração multivariada de Oliva – Canola no infravermelho médio.....	94
Tabela 14. Resultados para os modelos de calibração multivariada de Oliva – Canola no infravermelho próximo.....	96
Tabela 15. Resultados para os modelos de calibração multivariada de Oliva – Canola no Raman.....	98

Tabela 16. Resultados para os modelos de calibração multivariada de Oliva - Girassol.....	100
Tabela 17. Resultados para os modelos de calibração multivariada de Oliva – Girassol no infravermelho próximo.....	102
Tabela 18. Resultados para os modelos de calibração multivariada de Oliva – Girassol no Raman.....	104
Tabela 19. Resultados para os modelos de calibração multivariada de Oliva – Milho no infravermelho médio.....	106
Tabela 20. Resultados para os modelos de calibração multivariada de Oliva – Milho no infravermelho próximo.....	108
Tabela 21. Resultados para os modelos de calibração multivariada de Oliva – Milho no Raman.....	110
Tabela 22. Resultados para os modelos de calibração multivariada de Oliva – Soja no infravermelho médio.....	113
Tabela 23. Resultados para os modelos de calibração multivariada de Oliva – Soja no infravermelho próximo.....	115
Tabela 24. Resultados para os modelos de calibração multivariada de Oliva – todos contaminantes no mesmo modelo.....	118

LISTA DE FIGURAS

Figura 1. Diagrama de energia potencial. (1) oscilador harmônico, (2) oscilador anarmônico.....	13
Figura 2. Representação da ATR.....	19
Figura 3. Representação da construção da matriz X para calibração multivariada.....	31
Figura 4. Decomposição em componentes principais por PCA.....	33
Figura 5. Decomposição em variáveis latentes das matrizes X e Y para modelos PLS.....	36
Figura 6. Gráfico do modelo iPLS do número de intervalos versus RMSECV.....	39
Figura 7. Gráficos do modelo iPLS: (a) número de intervalos versus RMSECV e (b) faixa escolhida para a construção do modelo PLS.....	39
Figura 8. Gráfico do modelo UVE-PLS: valores de <i>t</i> para as variáveis experimentais (1-1.300, em verde) e randômicas (1.301-2.600, em vermelho). O limite de corte é indicado pela linha azul tracejada.....	41
Figura 9. Gráfico das variáveis selecionadas pelo modelo UVE-PLS, em vermelho.....	42
Figura 10. Representação da codificação binária de um espectro na forma de um cromossomo. Cada número de onda corresponde a um gene do cromossomo. As variáveis selecionadas (em vermelho) recebem código 1 e as não selecionadas (em azul) recebem código 0.....	43
Figura 11. Representação do cruzamento entre dois cromossomos pais originando dois cromossomos filhos. Havendo rompimento e troca de material genético do cromossomo pai 1 (em azul) com o material genético do cromossomo pai 2 (em verde).....	44
Figura 12. Representação da mutação do gene 2 do cromossomo filho 2, gerando um novo cromossomo.....	45

Figura 13. Gráficos do modelo GA-PLS: (a) aptidão versus número de variáveis após as gerações, (b) evolução da média e melhor aptidão, (c) evolução do número de variáveis selecionadas e (d) freqüência das variáveis selecionadas após as gerações.....	46
Figura 14. Gráfico das variáveis selecionadas pelo modelo GA-PLS, em vermelho.....	46
Figura 15. Esquema básico da lubrificação em motores automotivos.....	54
Figura 16. Motor automotivo danificado pelo uso de lubrificante degradado.....	54
Figura 17. Representação de bandas espectrais características para espectros de lubrificantes na região do infravermelho médio.....	60
Figura 18. Espectros das 45 amostras de lubrificante de ônibus de 1.936 a 45.896 Km.....	61
Figura 19. Espectros das 45 amostras de lubrificante de ônibus com as variáveis selecionadas pelo GA-PLS para calibração relacionada ao TBN, em vermelho.....	63
Figura 20. Espectros das 47 amostras de lubrificante de ônibus de 522 a 45.896 Km.....	65
Figura 21. Espectro de lubrificante de ônibus com as variáveis selecionadas pelo GA-PLS para calibração relacionada à viscosidade, em vermelho.....	66
Figura 22. Espectros das 40 amostras de lubrificante contaminadas com diesel.	67
Figura 23. Espectro de lubrificante contaminado com diesel com as variáveis selecionadas pelo GA-iPLS em vermelho.....	69
Figura 24. Espectros das 25 amostras de lubrificante contaminadas com gasolina.....	70
Figura 25. Espectro de lubrificante contaminado com gasolina com as variáveis selecionadas pelo GA-iPLS em vermelho.....	71
Figura 26. Espectros das 21 amostras de lubrificante contaminadas com água...72	
Figura 27. Espectros das 40 amostras de lubrificante contaminadas com etileno glicol.....	74

Figura 28. Representação de bandas espectrais características para espectros de óleo de Oliva na região do infravermelho médio.....	83
Figura 29. Representação de bandas espectrais características para espectros de óleo de Oliva na região do infravermelho próximo.....	84
Figura 30. (a) Espectros Raman originais das amostras de óleo de Oliva. (b) Espectros Raman corrigidos pelo método de Gram-Schmidt e a representação de suas bandas espectrais.....	85
Figura 31. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Algodão.....	86
Figura 32. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Algodão com as variáveis selecionadas pelo GA-iPLS em vermelho.....	88
Figura 33. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Algodão.....	88
Figura 34. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Algodão com as variáveis selecionadas pelo GA-iPLS em vermelho.....	90
Figura 35. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Algodão.....	91
Figura 36. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Algodão com as variáveis selecionadas pelo GA-iPLS em vermelho.....	92
Figura 37. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Canola.....	93
Figura 38. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Canola com as variáveis selecionadas pelo GA-iPLS em vermelho.....	94
Figura 39. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Canola.....	95

Figura 40. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Canola com as variáveis selecionadas pelo GA-PLS em vermelho.....	96
Figura 41. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Canola.....	97
Figura 42. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Canola com as variáveis selecionadas pelo GA-iPLS em vermelho.....	98
Figura 43. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Girassol.....	99
Figura 44. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Girassol com as variáveis selecionadas pelo GA-PLS em vermelho.....	101
Figura 45. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Girassol.....	101
Figura 46. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Girassol com as variáveis selecionadas pelo GA-iPLS em vermelho.....	103
Figura 47. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Girassol.....	103
Figura 48. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Girassol com as variáveis selecionadas pelo GA-iPLS em vermelho.....	105
Figura 49. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Milho.....	106
Figura 50. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Milho com as variáveis selecionadas pelo GA-iPLS em vermelho.....	107
Figura 51. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Milho.....	108
Figura 52. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Milho com as variáveis selecionadas pelo GA-PLS em vermelho.....	109

Figura 53. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Milho.....	110
Figura 54. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Milho com as variáveis selecionadas pelo GA-iPLS em vermelho.....	111
Figura 55. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Soja.....	112
Figura 56. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Soja com as variáveis selecionadas pelo GA-iPLS em vermelho.....	113
Figura 57. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Soja.....	114
Figura 58. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Soja com as variáveis selecionadas pelo GA-iPLS em vermelho.....	115
Figura 59. Espectros no infravermelho de todas as amostras de óleo de Oliva contaminadas. (a) no médio e (b) no próximo.....	117

SUMÁRIO

Prefácio.....	1
Capítulo 1 – Espectroscopia Vibracional.....	7
1. Espectroscopia Vibracional.....	9
1.1 Espectroscopia no infravermelho.....	10
1.1.1. A Absorção Molecular.....	10
1.1.2. Princípios da espectroscopia no infravermelho.....	11
1.1.3. Aplicações da espectroscopia no infravermelho.....	14
1.1.4. Espectrômetros com transformada de Fourier.....	18
1.1.5. Espectroscopia de Reflectância Total Atenuada.....	19
1.2. Espectroscopia Raman.....	20
1.2.1. Princípios da espectroscopia Raman.....	20
1.2.2. Aplicações da espectroscopia Raman.....	21
1.2.3. Espectrômetros Raman.....	23
1.3. O uso da espectroscopia vibracional.....	24
Capítulo 2 – Métodos Quimiométricos.....	27
2. Quimiometria.....	29
2.1. Calibração Multivariada.....	30
2.1.1. Análise de Componentes Principais - PCA.....	32
2.1.2. Mínimos Quadrados Parciais - PLS.....	34
2.2. Métodos de Seleção de Variáveis.....	37
2.2.1. PLS por intervalos (iPLS)	38
2.2.2. Eliminação de variáveis não informativas por PLS (UVE-PLS)	40
2.2.3. Algoritmo Genético (GA)	42
2.3. Cálculo dos erros.....	48
2.4. Programas desenvolvidos.....	48

Capítulo 3 – Lubrificantes.....51

3. Lubrificantes.....	53
3.1. Objetivo.....	56
3.2. Parte Experimental.....	56
3.2.1. Coleta das amostras de lubrificante de ônibus.....	56
3.2.2. Medida do Número Total de Basicidade (TBN)	57
3.2.3. Medida de Viscosidade a 100°C.....	58
3.2.4. Contaminação em lubrificantes automotivos.....	58
3.2.5. Espectroscopia no infravermelho médio de lubrificantes.....	59
3.3. Resultados e discussão.....	59
3.3.1. Os espectros de lubrificantes na região do infravermelho médio.....	59
3.3.2. Calibração Multivariada de lubrificante de motor de ônibus relacionado ao parâmetro de qualidade TBN.....	60
3.3.3. Calibração Multivariada de lubrificante de motor de ônibus relacionado ao parâmetro de qualidade Viscosidade a 100°C.....	64
3.3.4. Calibração Multivariada de lubrificante contaminado com Diesel.....	67
3.3.5. Calibração Multivariada de lubrificante contaminado com Gasolina.....	69
3.3.6. Calibração Multivariada de lubrificante contaminado com Água.....	72
3.3.7. Calibração Multivariada de lubrificante contaminado com Etileno Glicol.....	74
3.4. Conclusões da calibração multivariada de lubrificantes.....	76

Capítulo 4 – Óleos de Oliva.....77

4. Óleos de Oliva.....	79
4.1. Objetivo.....	81
4.2. Parte Experimental.....	82
4.2.1. Contaminação de óleo de Oliva.....	82
4.2.2. Espectroscopia no infravermelho médio de óleo de Oliva.....	82
4.2.3. Espectroscopia no infravermelho próximo de óleo de Oliva.....	83
4.2.4. Espectroscopia Raman de óleo de Oliva.....	84

4.3. Resultados e discussão.....	86
4.3.1. Calibração Multivariada de óleo de Oliva contaminado com óleo de Algodão.....	86
4.3.2. Calibração Multivariada de óleo de Oliva contaminado com óleo de Canola.....	93
4.3.3. Calibração Multivariada de óleo de Oliva contaminado com óleo de Girassol.....	99
4.3.4. Calibração Multivariada de óleo de Oliva contaminado com óleo de Milho.....	105
4.3.5. Calibração Multivariada de óleo de Oliva contaminado com óleo de Soja.....	112
4.3.6. Calibração Multivariada de todos os contaminantes em um só modelo.....	116
4.4. Conclusões da calibração multivariada de óleos de Oliva.....	119
Conclusão.....	121
Referências Bibliográficas.....	123

PREFÁCIO

Prefácio

O produto principal de um laboratório químico analítico é a informação sobre a composição química de materiais, normalmente em termos da identidade e/ou quantidade de um ou mais componentes relevantes nas amostras destes materiais.

Assim, pode-se definir qualidade como sendo uma noção relativa, nunca "alta" nem "baixa" em um sentido absoluto, mas adequada ou inadequada o bastante, em termos de que até que ponto um produto, processo ou serviço satisfaz as exigências especificadas anteriormente por um objetivo ou um cliente.

Uma etapa fundamental para melhorar o controle de qualidade é a modelagem da resposta do instrumento, como por exemplo: espectros de absorção, cromatogramas, interferogramas, etc., como função das variações das concentrações de espécies químicas, ou ainda como função de uma outra propriedade qualquer. Esse tipo de modelagem é chamado calibração e, após ser estabelecida, a mesma pode ser usada para a previsão da concentração (ou outra propriedade qualquer) de espécies químicas em amostras desconhecidas.

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 principalmente pela regressão por Mínimos Quadrados Parciais (PLS – do inglês, Partial Least Squares).

Atualmente, é crescente o interesse por parte de muitos grupos de pesquisa, na aplicação da espectroscopia vibracional aliada ao método PLS devido à potencialidade quase que infinita, dessas técnicas em extrair informações com qualidade de análises qualitativas e quantitativas, sendo possível a realização de análises não destrutivas, mais rápidas, sem gasto de reagentes e sem gerar resíduos. Porém, pode ser discutível se o método PLS é capaz de determinar claramente, somente as características espectrais cruciais para a construção de modelos de previsão.

Alguns métodos foram descritos recentemente na literatura para implementar seleção de região espectral para melhorar significativamente o desempenho dos métodos de calibração de espectros inteiros. Esses métodos são chamados métodos de seleção de variáveis, que selecionam regiões específicas do espectro que não são importantes, enquanto geram modelos mais estáveis, robustos e mais simples de interpretar. Na prática, a metodologia está baseada na identificação de um subconjunto dos dados inteiros que produzirão menores erros de previsão. Assim, em espectros vibracionais por exemplo, podem ser eliminados os comprimentos de onda que apenas representam ruídos, informações irrelevantes ou não-linearidades.

Nesse contexto, esta Tese teve como objetivo a aplicação de métodos de calibração multivariada e de seleção de variáveis para determinar parâmetros considerados no controle de qualidade de produtos de grande interesse comercial como lubrificantes e óleos de Oliva, avaliando os modelos obtidos na tentativa de determinar uma melhor relação existente entre medidas espectroscópicas e a propriedade de interesse.

Os temas abordados durante a pesquisa estão distribuídos em quatro capítulos além da conclusão final e referências bibliográficas. O primeiro capítulo, intitulado Espectroscopia Vibracional apresenta um breve histórico da espectroscopia vibracional; uma descrição sobre o fenômeno da absorção molecular; os princípios e aplicações da espectroscopia no infravermelho (médio e próximo) e no Raman; os fundamentos teóricos; aspectos instrumentais dos espectrômetros com transformada de Fourier e dispersivo; a espectroscopia de reflectância total atenuada e o uso da espectroscopia vibracional.

O segundo capítulo, intitulado Quimiometria apresenta o conceito e um breve histórico da quimiometria; os fundamentos teóricos do método de calibração multivariada por PLS e de seleção de variáveis; as aplicações; a descrição do cálculo de erros e dos programas desenvolvidos para o Matlab.

O terceiro capítulo, intitulado Lubrificantes apresenta os conceitos básicos e as funções de lubrificantes; uma descrição das análises mais comuns no controle de qualidade de lubrificantes; a coleta de amostras e a obtenção dos espectros

dos lubrificantes no MIR; medidas do TBN e viscosidade a 100°C; contaminações de lubrificantes por combustíveis, água e etileno glicol; discussão dos resultados dos modelos de calibração multivariada de lubrificantes e as conclusões do capítulo.

O quarto capítulo, intitulado óleos de Oliva apresenta as características e propriedades de óleos de Oliva; uma descrição dos tipos de adulteração; o uso da espectroscopia vibracional na determinação de adulteração de óleo de Oliva; descrição das regiões espectrais do óleo de Oliva; contaminações do óleo de Oliva pela adição de óleos de Algodão, Canola, Girassol, Milho e Soja; discussão dos resultados dos modelos de calibração multivariada do óleo de Oliva e as conclusões do capítulo.

Vislumbrando perspectivas promissoras no uso da espectroscopia vibracional aliada a métodos de calibração multivariada e seleção de variáveis a partir dos resultados obtidos, e principalmente da interpretação química, é possível acreditar que as metodologias são aplicáveis a outros parâmetros e métodos quimiométricos não abordados neste trabalho. Como por exemplo, o estudo de parâmetros de qualidade de lubrificantes industriais e de outros tipos de óleos vegetais.

Tais capítulos têm como finalidade estender ao leitor o conhecimento preliminar à compreensão da potencialidade das análises espectroscópicas em conjunto a métodos de calibração multivariada e de seleção de variáveis para resolver cada vez melhor, problemas em química.

As Conclusões da Tese, avaliam os resultados para a determinação de parâmetros de qualidade de lubrificantes e óleos de Oliva através de espectroscopia vibracional, calibração multivariada e seleção de variáveis. E com base nos resultados e vantagens apontadas, justificar que as metodologias desenvolvidas sejam recomendadas para substituir às metodologias empregadas atualmente. Por fim segue uma lista de Referências Bibliográficas em que são apresentados os trabalhos que contribuíram para a elaboração desta Tese.

CAPÍTULO 1

Espectroscopia Vibracional

1. 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. [1, 2]

Em 1800, o músico e astrônomo alemão Frederick William Herschel ao realizar um experimento utilizando um prisma para separação das faixas espectrais associadas à região do visível, observou que as cores decompostas apresentavam uma quantidade de energia distinta e para monitorar a quantidade de energia associada a cada cor usou um termômetro, verificando que abaixo do vermelho, onde não havia mais luz visível, era a região que apresentava maior temperatura. Assim, Herschel além de descobrir o infravermelho, demonstrou que existem formas de luz que não podem ser observadas pelo olho humano. [2, 3]

Em 1928, o físico indiano Chandrasenkara Venkata Raman descobriu que o comprimento de onda visível de uma pequena fração da radiação espalhada por certas moléculas difere daquele do feixe incidente e, ainda mais, que os deslocamentos de comprimento de onda dependem da estrutura química das moléculas responsáveis pelo espalhamento. Ele recebeu o Prêmio Nobel de Física em 1931 pela descoberta do efeito no espalhamento de luz, que recebeu em sua homenagem o nome de Espalhamento Raman. [1]

Este capítulo enfoca a descrição de conceitos relevantes à compreensão da espectroscopia vibracional, bem como das técnicas e instrumentação empregadas no desenvolvimento de nosso trabalho, limitando sempre que possível, ao uso específico na análise de parâmetros de qualidade de óleos Lubrificantes e de Oliva.

1.1. Espectroscopia no infravermelho

1.1.1. A Absorção Molecular

Quando a radiação atravessa uma camada de um sólido, líquido ou gás, algumas frequências são seletivamente removidas pela absorção, um processo no qual a energia eletromagnética é transferida para átomos, íons ou moléculas que compõem a amostra. A absorção promove essas partículas de seu estado normal à temperatura ambiente, ou estado fundamental, para um ou mais estados excitados de maior energia. [1, 2]

Para que a absorção da radiação ocorra, a energia do fóton de excitação deve ser exatamente igual à diferença de energia entre o estado fundamental e um estado excitado da amostra absorvedora. Uma vez que essa diferença de energia é única para cada espécie, um estudo das frequências da radiação absorvida fornece um meio de caracterizar os constituintes de uma amostra. Para isso, determina-se experimentalmente um gráfico da absorbância em função do comprimento de onda, ou da frequência. [1]

A energia, associada às bandas de uma molécula, é a soma das energias eletrônicas, vibracionais e rotacionais. Para cada estado de energia eletrônica de uma molécula, normalmente existem muitos estados vibracionais possíveis; e, para cada um dos estados vibracionais, são possíveis numerosos estados rotacionais. Como consequência, o número de níveis de energia possível para uma molécula é, normalmente, várias ordens de grandeza maior que o número de níveis de energia possível para uma partícula atômica. Assim, o espectro para uma molécula consiste de uma série de linhas de absorção muito próximas que constituem uma banda de absorção. As energias eletrônicas são observadas nas regiões ultravioleta e visível, as vibracionais na região do infravermelho e as rotacionais na região de microondas. [1, 2]

1.1.2. Princípios da espectroscopia no infravermelho

A radiação infravermelha geralmente não é energética o suficiente para causar transições eletrônicas, a absorção desta radiação está muito restrita a espécies moleculares que têm diferenças de energia pequenas entre vários estados vibracionais e rotacionais. Para absorver radiação infravermelha, uma molécula precisa sofrer uma variação no momento de dipolo como consequência do movimento vibracional ou rotacional. Apenas nessas circunstâncias o campo elétrico alternado da radiação pode interagir com a molécula e causar variações na amplitude de um de seus movimentos. O momento dipolar é determinado pela magnitude da diferença de carga e a distância entre dois centros de carga. Quando uma molécula que possui essa variação do momento dipolar vibra, uma variação regular do momento dipolar ocorre e surge um campo que pode interagir com o campo elétrico associado à radiação. Se a frequência da radiação coincidir exatamente com a frequência vibracional natural da molécula, ocorre uma transferência de energia efetiva e resulta em uma variação da amplitude da vibração molecular; a consequência é a absorção de radiação. Do mesmo modo, a rotação de moléculas assimétricas em torno dos seus centros de massa resulta em uma variação periódica do dipolo que pode interagir com a radiação. Nenhuma variação efetiva no momento de dipolo ocorre durante a vibração ou rotação de uma molécula homonuclear, como O_2 , N_2 ou Cl_2 , conseqüentemente, essas substâncias não podem absorver no infravermelho. [1, 4]

As vibrações moleculares são divididas em duas categorias: estiramentos e deformações angulares. Uma vibração de estiramento envolve uma variação contínua na distância interatômica ao longo do eixo da ligação entre dois átomos podendo acontecer de forma simétrica ou assimétrica. As vibrações de deformação angular são caracterizadas pela variação do ângulo entre duas ligações e podem acontecer no plano ou fora do plano da molécula, de quatro tipos: tesoura (do inglês *scissoring*), balanço (*rocking*), sacudida (*wagging*) e torção (*twisting*). Além desses tipos de vibração, interações ou acoplamentos de vibrações podem ocorrer se as vibrações envolvem ligações de um mesmo átomo

central, tendo como resultado uma variação nas características das vibrações envolvidas. [1, 2]

É possível deduzir o número e a natureza das vibrações em moléculas diatômicas ou triatômicas simples e se essas vibrações vão conduzir a absorções. Entretanto, para moléculas constituídas de muitos átomos possuindo um grande número de centros de vibração, como também para moléculas apresentando interações entre vários centros, a variedade de vibrações possíveis dá origem a espectros na região do infravermelho que são difíceis, senão impossíveis, de se analisar. A utilização das propriedades de simetria molecular e teoria de grupo, em conjunto com o método da mecânica clássica, torna o tratamento vibracional de moléculas poliatômicas bem mais fácil. O número de vibrações possíveis em uma molécula poliatômica pode ser calculado como se segue. Cada átomo possui 3 graus de liberdade (movimentos nas direções x, y e z) e uma molécula com N átomos irá ter $3N$ graus de liberdade. Já que nos interessa somente os movimentos vibracionais, podemos eliminar 3 graus de liberdade translacional, mais 3 graus de liberdade rotacional, para moléculas não-lineares. Os $3N - 6$ graus restantes representam o número de vibrações possíveis na molécula. No caso de moléculas lineares, a rotação em torno do eixo de ligação não é possível e 2 graus de liberdade bastam para descrever o movimento rotacional, restando $3N - 5$ graus de liberdade vibracional. [1]

Considerando que as vibrações acontecem de forma isolada em uma molécula, as características da vibração de estiramento podem se aproximar às de um modelo mecânico consistindo de duas massas ligadas por uma mola, e as frequências fundamentais de quaisquer dois átomos ligados podem ser calculadas assumindo que a energia segue o comportamento de um oscilador harmônico que obedece a Lei de Hooke. Porém, este modelo não descreve completamente o comportamento de partículas de dimensões atômicas, mas pode ser usado para desenvolver as equações de onda da mecânica quântica para a vibração. Assim, em contraste com a mecânica clássica, na qual os osciladores podem ter qualquer energia potencial, os osciladores mecânico-quânticos podem ter apenas determinadas energias discretas, ou seja, as únicas transições que poderiam

ocorrer seriam aquelas em que o número quântico vibracional muda de uma unidade (regra de seleção: $\Delta v = \pm 1$). Uma vez que os níveis vibracionais são igualmente espaçados, apenas uma única banda de absorção deveria ser observada para uma certa vibração molecular. [1, 2]

Embora permitindo obter curvas de energia potencial mais corretas para vibrações moleculares, a complexidade matemática das equações de onda da mecânica quântica impede a sua aplicação quantitativa, exceto para sistemas muito simples. Qualitativamente, as curvas têm que tomar a forma anarmônica diferenciando do comportamento harmônico em vários aspectos, dependendo da natureza da ligação e dos átomos envolvidos. No entanto, a curva harmônica e a anarmônica são muito parecidas para pequenas variações em torno do centro de massa. [1, 2]

As anarmônidades elétrica e mecânica conduzem a desvios de duas espécies. O primeiro deles é que os níveis de energia não são igualmente espaçados, como pode ser visto na representação da curva 2 na Figura 1.

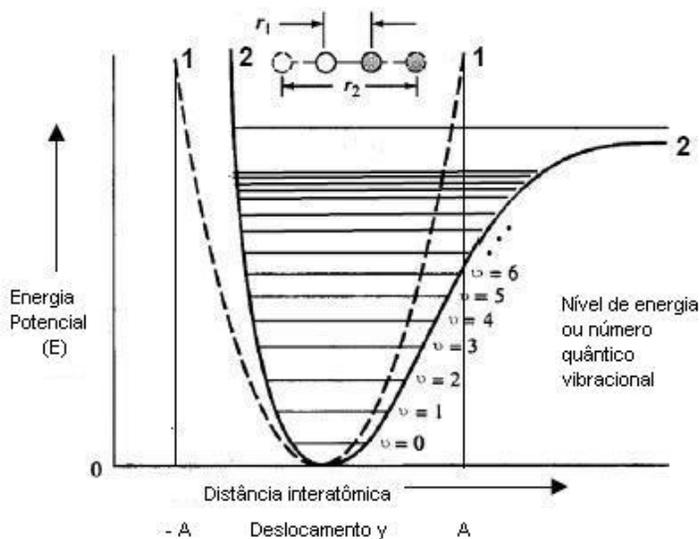


Figura 1. Diagrama de energia potencial. (1) oscilador harmônico, (2) oscilador anarmônico.

A diferença entre os níveis de energia diminui à medida que a energia aumenta. A segunda espécie de desvio é que em números quânticos altos, ΔE se torna menor e a regra de seleção não é seguida rigorosamente e como resultado,

harmônicos que ocorrem em frequências com aproximadamente duas ou três vezes a de uma transição fundamental são observadas, isto é $\Delta v = \pm 2$ ou ± 3 , sendo esse tipo de transição conhecida como sobretom (do inglês, “overtone”). A maior parte dos sinais do tipo sobretom ocorre na região do infravermelho próximo e a intensidade dessas transições é cerca de 10 a 1.000 vezes menor que as observadas para transições fundamentais. [1, 2]

Além dos dois desvios descritos, às vezes são encontradas bandas de combinação quando um fóton excita simultaneamente dois modos vibracionais ocorrendo principalmente entre 5.500 e 4.000 cm^{-1} . A frequência da banda de combinação é aproximadamente a soma ou a diferença das duas frequências fundamentais. Esse fenômeno ocorre quando um quantum de energia é absorvido por duas ligações em vez de uma. [1, 2]

1.1.3. Aplicações 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 cm^{-1} ou comprimento de onda de 780 a 100.000 nm. [1, 2] Do ponto de vista da aplicação e instrumentação, o espectro infravermelho é dividido em radiação no infravermelho próximo (NIR – do inglês, Near Infrared), médio (MIR – do inglês, Middle Infrared) e distante (FIR – Far Infrared). A Tabela 1 apresenta os limites aproximados para cada região.

Tabela 1. Regiões espectrais do infravermelho.

Região	Intervalo de número de onda ($\bar{\nu}$), 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*. [1]

O uso inicial da espectroscopia em química restringiu basicamente a identificação e quantificação de algumas poucas espécies químicas em casos bem específicos devido às limitações instrumentais, superadas posteriormente com o desenvolvimento de equipamentos mecânicos e ópticos mais precisos ampliando essas aplicações. [4, 5]

No fim dos anos 1950 com o aparecimento de espectrofotômetros dispersivos de feixe duplo, que reduziram o tempo necessário para se fazer uma determinação estrutural por um fator de dez, cem ou mesmo mil, as aplicações qualitativas concentraram-se principalmente na região do MIR, uma vez que nessa região é possível observar bandas de absorção de grupos orgânicos específicos como N-O, N-H, C-H, O-H, C-C, C-O, C=C, S-H entre outros. Assim, esta técnica foi largamente empregada pelos químicos orgânicos para auxiliar na caracterização de diversas substâncias químicas. [1-3]

As primeiras aplicações industriais de espectroscopia no infravermelho eram para qualidade e controle de produção nas indústrias petroquímicas, principalmente para combustíveis, lubrificantes e polímeros, como parte dos esforços industriais durante a segunda guerra mundial. Os instrumentos foram projetados pelas maiores companhias químicas da época (Dow, Shell e Cyanamid) e configurados somente para medidas de absorção na região do MIR. [2, 4]

A partir dos anos 1970 com o desenvolvimento dos espectrofotômetros com transformada de Fourier, da informática, do interfaceamento de instrumentos eletrônicos com computadores e a introdução de recursos matemáticos mais sofisticados, teve início uma nova fase de estudos espectroscópicos, em que as análises quantitativas expandiram-se para as regiões espectrais no NIR e MIR. Houve, então, um surpreendente interesse de vários grupos de pesquisas acadêmicos, governamentais e industriais, na tentativa de desenvolver metodologias de análise com essa faixa espectral, tendo como consequência o impacto direto na produtividade, redução de gastos e a minimização de resíduos. [1-5]

A espectroscopia no infravermelho próximo não foi considerada inicialmente como uma técnica analítica com algum valor prático, sendo originalmente, uma extensão da região do visível que não foi explorada até por volta de 1970. Os primeiros trabalhos que proporcionaram interesse pelo estudo da espectroscopia no infravermelho próximo como ferramenta de análise na indústria, foram desenvolvidos na década de setenta pelo grupo de pesquisa do professor Karl Norris, quando este era responsável por um grupo de pesquisa do Departamento de Agricultura dos Estados Unidos. Entretanto, as limitações tecnológicas da época não permitiram o uso mais abrangente do infravermelho próximo como atualmente é observado em diversos setores industriais. [6, 7]

Para a região do NIR, os sinais observados são devidos essencialmente a sobretons e bandas de combinação de estiramentos e deformações angulares de transições fundamentais de ligações X-H, em que X representa átomos de oxigênio, nitrogênio, carbono, grupos aromáticos e também grupos funcionais importantes como os C-O, grupos carbonila, C-N, C-C, entre outros que sofrem estiramento. Diferentemente da espectroscopia no infravermelho médio, os espectros NIR são menos úteis para identificação e mais úteis para análises quantitativas, sendo que a maior desvantagem da técnica, é provavelmente, a baixa sensibilidade a constituintes menores. [1-5]

A princípio, as medidas eram somente realizadas em fotômetros e espectrofotômetros dispersivos baseados em filtros e redes de difração, respectivamente. A configuração destes equipamentos era semelhante à de equipamentos que operavam na região do ultravioleta/visível (UV/Vis), sendo que, em muitos casos tratavam-se de equipamentos que compreendiam a região UV/Vis/NIR. Devido ao reconhecimento do potencial da aplicação do infravermelho próximo principalmente nas análises quantitativas, equipamentos foram desenvolvidos para análises específicas nesta região. [1, 2]

Até o início dos anos 1980, a maioria dos instrumentos para a região do infravermelho era do tipo dispersivo, baseada em redes de difração. A partir daí, no entanto, houve uma mudança drástica nesses instrumentos, de modo que agora a maior parte deles são do tipo de transformada de Fourier. O aparecimento

na última década de espectrômetros com transformada de Fourier (FTIR – do inglês, Fourier transform infrared) relativamente baratos aumentou notavelmente o número e o tipo de aplicações da radiação no infravermelho. Esse aumento surgiu como conseqüência do aumento de uma ordem de magnitude ou mais nas relações sinal-ruído e limites de detecção em comparação aos que podem ser obtidos com instrumentos interferométricos. A partir daí, a espectroscopia no infravermelho médio passa a ser usada também para análise quantitativa de amostras complexas tanto por absorção quanto por emissão. [1, 2]

Na região do NIR as principais aplicações encontram-se na análise quantitativa de materiais industriais e agrícolas e no controle de processos, destacando também as aplicações farmacêuticas, alimentícias e petroquímicas, sendo também uma ferramenta valiosa para a identificação e determinação de aminas primárias e secundárias na presença de aminas terciárias em misturas. [4]

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 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 conseqüência, uma semelhança estreita entre dois espectros nesta região, bem como nas outras, constitui forte evidência da identidade dos compostos que produziram os espectros. [1, 4]

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 semi-sólidas, pastas, géis e outras. [1-3]

Nos últimos anos, a maioria dos fabricantes de instrumentos oferece adaptadores que cabem nos compartimentos de amostra dos instrumentos de espectroscopia no infravermelho e que tornam possível a obtenção imediata de espectros de reflexão. [1-4]

1.1.4. Espectrômetros com transformada de Fourier

Nos anos 1960, foram publicadas as primeiras aplicações químicas da espectroscopia com transformada de Fourier; no final dos anos 1960, já estavam disponíveis comercialmente instrumentos tanto na região FIR quanto na região do MIR. Essa transformação é um filtro digital efetuada matematicamente em um computador por um algoritmo muito rápido e eficiente. A maioria dos instrumentos FTIR utiliza um interferômetro de Michelson, que modula a radiação na região óptica, ao invés de redes de difração. Esse interferômetro divide o feixe de radiação em dois feixes de potência aproximadamente iguais e então recombina-os de forma que as variações na intensidade do feixe recombinao possam ser medidas em função das diferenças do caminho percorrido pelos dois feixes. [1, 2]

Espectrofotômetros FTIR apresentam algumas vantagens, como a eficiência no transporte da radiação até o detector ou vantagem de Jaquinot, pois possuem poucos elementos ópticos e nenhuma fenda para atenuar a radiação, o que aumenta a potência da radiação que incide no detector e melhora a relação sinal/ruído; apresenta um alto poder de resolução ($< 0,1 \text{ cm}^{-1}$) e reprodutibilidade do comprimento de onda, que torna possível a análise de espectros complexos nos quais o número de linhas estreitas e a superposição espectral tornam difícil a interpretação dos espectros individuais; e todos os elementos de resolução para um espectro são medidos simultaneamente, vantagem de Fellgett ou multiplex, o que possibilita a aquisição de dados de um espectro inteiro em cerca de um segundo ou menos. [1]

Os espectrômetros com transformada de Fourier são menos comuns para as regiões UV, Vis e NIR, porque as limitações da relação sinal-ruído nas medidas nestas regiões raramente se devem ao ruído do detector, mas sim ao ruído “shot” ou “flicker” associados com a fonte de radiação e que aumentam conforme aumenta a potência do sinal. Além disso, o ruído total para todos os elementos de resolução em uma medida por FT tende a ser promediado e distribui-se uniformemente sobre todo o espectro. Então, na presença de bandas fracas, a

relação sinal-ruído para bandas intensas é aumentada pela promediação, mas degradada para bandas mais fracas. [1]

1.1.5. Espectroscopia de Reflectância Total Atenuada

A reflexão de radiação pode ser de quatro tipos: especular, difusa, interna e total atenuada (ATR - do inglês, attenuated total reflection). Em nosso trabalho vamos focar 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. [1, 8]

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 um 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 reflectância total atenuada (ATR), mostrado na Figura 2. [1, 8]

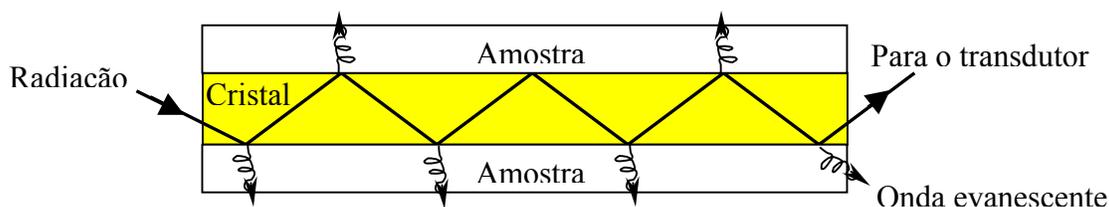


Figura 2. Representação da ATR.

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 absorbâncias, embora dependam do ângulo de incidência, são independentes da espessura da amostra, uma vez que a radiação penetra apenas alguns micrômetros na mesma. [1]

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. [1]

1.2. Espectroscopia Raman

1.2.1. Princípios da espectroscopia Raman

O espalhamento Raman resulta de alterações na polarizabilidade de ligações moleculares. A diferença de comprimento de onda entre a radiação visível incidente e a espalhada corresponde a comprimentos de onda na região do MIR. De fato, o espectro do espalhamento Raman e o espectro de absorção no infravermelho de uma determinada espécie se assemelham muito. Há, no entanto, diferenças suficientes entre os tipos de grupos que são ativos no infravermelho e no Raman para tornar essas técnicas complementares ao invés de competitivas. [1, 2]

Os espectros Raman são obtidos irradiando-se uma amostra com uma fonte laser (do inglês, light amplification by stimulated emission of radiation) potente de radiação monocromática no UV, Vis ou no NIR. A radiação espalhada é de três tipos: espalhamento Stokes, espalhamento anti-Stokes e espalhamento Rayleigh. O último, elástico (sem troca de energia com a molécula), ocorre na mesma frequência da radiação incidente e, portanto, não apresenta nenhuma informação sobre os níveis vibracionais, é significativamente mais intenso que os

outros tipos. Os espalhamentos Stokes e anti-Stokes por sua vez, são inelásticos e provêm a base para o espectro Raman. [1, 4]

No espectro Raman existe uma simetria em relação à linha Rayleigh, do lado de freqüências mais baixas as linhas Stokes, e do lado de freqüências mais altas as linhas anti-Stokes. Geralmente, as linhas anti-Stokes são notadamente menos intensas que as Stokes correspondentes. Por essa razão, apenas a parte Stokes de um espectro é normalmente usada. [1, 4]

As diferenças entre um espectro Raman e um no infravermelho não são surpreendentes quando consideramos que os mecanismos básicos, embora dependentes dos mesmos modos vibracionais, se originam de processos que são mecanisticamente diferentes. A absorção infravermelha exige que um modo vibracional da molécula tenha uma variação no momento de dipolo ou na distribuição de carga associada a ele. Então somente uma radiação de mesma freqüência pode interagir com a molécula e promovê-la a um estado vibracional excitado. Em contraste, o espalhamento envolve uma distorção momentânea dos elétrons distribuídos em torno de uma ligação da molécula, seguida por reemissão da radiação enquanto a ligação volta ao seu estado normal. Na sua forma distorcida, a molécula fica temporariamente polarizada, isto é, ela cria momentaneamente um momento dipolar induzido que desaparece por relaxação e reemissão. Devido a essa diferença fundamental no mecanismo, a atividade Raman de um dado modo vibracional pode diferir significativamente da sua atividade no infravermelho. Por exemplo, uma molécula homonuclear, como N_2 , Cl_2 e H_2 , inativa no infravermelho por não apresentar variação efetiva no momento de dipolo, é ativa no Raman devido à polarizabilidade da ligação entre os átomos.

1.2.2. Aplicações da espectroscopia Raman

Como nas regiões do infravermelho, na espectroscopia Raman também é possível realizar medidas de amostras em todos os estados e formas como, gases, líquidos, sólidos, sistemas binários e terciários como as amostras semi-sólidas, pastas, géis e outras. Assim, tem sido aplicada a análises qualitativas e

quantitativas, em controle de qualidade e controle de processos industriais, agrícolas, alimentícias, petroquímicas, entre outras. [1, 4]

Uma vantagem importante dos espectros Raman sobre os de infravermelho está no fato de que a água não causa interferência; de fato, os espectros Raman podem ser obtidos de soluções aquosas. [1]

A espectroscopia Raman não foi amplamente usada pelos químicos para estudos estruturais até que os lasers se tornaram disponíveis nos anos 1960, o que tornou os espectros muito mais fáceis de serem obtidos. Entretanto, foram considerados de uso muito limitado devido a desvantagens como a interferência da fluorescência da amostra ou de impurezas na mesma, a fotodecomposição da amostra, problemas na calibração de comprimento de onda e dificuldade em atingir espectros de alta resolução. Avanços instrumentais como a tecnologia de interferômetros e fontes no NIR, superaram esses problemas, ampliando o uso da espectroscopia Raman. [2, 9]

A técnica Raman é freqüentemente superior ao infravermelho para investigação de sistemas inorgânicos porque podem ser empregadas soluções aquosas. Além disso, as energias vibracionais de ligações metal-ligante estão geralmente na região de 100 a 700 cm^{-1} ; uma região do infravermelho que é difícil de ser estudada experimentalmente, mas facilmente observadas no Raman. Estudos de Raman são fontes de informação potencialmente úteis no tocante à composição, a estrutura e a estabilidade de compostos de coordenação. As constantes de dissociação de ácidos fortes, como H_2SO_4 , HNO_3 , H_2SeO_4 e H_5IO_6 foram calculadas a partir de medidas Raman. [1]

Os espectros Raman dão mais informação sobre certos tipos de compostos orgânicos que os correspondentes no infravermelho. Por exemplo, a possibilidade de fornecer informações úteis a respeito do grupo funcional olefínico ou então, estimar o tamanho do anel em derivados de cicloparafinas, que podem não ser reveladas por espectros no infravermelho. A espectroscopia Raman tem sido amplamente aplicada ao estudo de sistemas biológicos, com vantagens de exigir amostras pequenas, sensibilidade mínima no tocante à interferência da água, o detalhamento espectral e a sensibilidade conformacional e ambiental. [1]

Os espectros Raman tendem a ter menos bandas do que os espectros no infravermelho. Como consequência, a superposição de bandas em misturas é menos provável e medidas quantitativas são mais simples. Adicionalmente, os dispositivos de amostragem para Raman não estão sujeitos a ataque por umidade e pequenas quantidades de água na amostra não interferem. Apesar dessas vantagens, a espectroscopia Raman ainda não tem sido explorada amplamente para análises quantitativas, principalmente devido ao custo relativamente alto dos espectrômetros Raman em relação ao da instrumentação de absorção. Essa barreira provavelmente tornar-se-á menos importante com a proliferação de lasers de diodo de baixo custo. [1, 2]

Como feixes de laser podem ser focalizados precisamente, é possível fazer análises quantitativas em amostras muito pequenas. Para esse trabalho, instrumentos chamados microsondas de laser são empregados. Essas microsondas têm sido usadas para determinar analitos em células de bactérias, componentes em partículas de fumaça e fuligem e espécies em inclusões microscópicas em minerais. [1]

1.2.3. Espectrômetros Raman

Até o início dos anos 1980, os espectrômetros Raman eram semelhantes em projeto e usados do mesmo modo que os instrumentos dispersivos no UV/Vis. Muitos usavam sistemas de dupla rede de difração, para minimizar o acesso da radiação espúria ao transdutor. Fotomultiplicadoras serviam como transdutores. Agora, no entanto, a maioria dos espectrômetros Raman comercializados são instrumentos com FT equipados com transdutores de Ge resfriados com nitrogênio líquido ou instrumentos multicanal (dispersivos) com detector de carga acoplada. Esses transdutores, em contraste com as válvulas fotomultiplicadoras, são sensíveis à radiação em 783 nm produzida por lasers de diodo, que promovem excitação Raman de muitos compostos sem fluorescência significativa. Os dispositivos de carga acoplada não são, infelizmente, sensíveis à radiação em

1.064 nm do laser Nd:YAG (íons Nd em um cristal hospedeiro contendo granada de Y e Al). [1, 2]

Um espectrômetro dispersivo típico com detector de carga acoplada usa como fonte um laser de diodo e um sistema de filtros que fornece radiação de 783 nm. Esse feixe é focalizado na extremidade da fibra de excitação por meio de uma lente e transmitido a uma sonda de fibra, que consiste da fibra de excitação rodeada por fibras coletoras. Por sua vez, as fibras coletoras conduzem a emissão Raman à fenda de um monocromador, no qual passa por um filtro que elimina a radiação do espalhamento Rayleigh. Uma rede de difração dispersa a radiação e a reflete sobre um dispositivo de carga acoplada (arranjo de diodos). [1]

1.3. O uso da espectroscopia vibracional

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. [1]

O uso comum de tecnologia existe nas três técnicas, MIR, NIR e Raman, e freqüente 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. Uma dúvida persiste no debate em cima de que instrumento é preferido ou até mesmo necessário. Na realidade não há uma resposta de consenso, e a escolha depende da aplicação e a demanda e/ou a complexidade da medida. Dependendo do ambiente operacional, instrumentos

podem ser adaptados com acessórios de manipulação de amostra e programas computacionais para aplicações qualitativas e quantitativas. Nos dias atuais, os fabricantes de instrumentos estão desenvolvendo instrumentos cada vez mais compactos e de custo menor, sendo razoável prever dentro do futuro próximo a miniaturização dos espectrômetros vibracionais.

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.

CAPÍTULO 2

Métodos Quimiométricos

2. 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. A partir dos anos 1970, o desenvolvimento tecnológico relativo ao surgimento de métodos instrumentais computadorizados para a análise química, possibilitou a geração de uma gama de informações, de forma rápida e com boa qualidade, que além de grandes dimensões, são às vezes, complexas e variadas, não podendo dessa maneira ser explicadas por modelos teóricos da química clássica. Foi então que um certo número de químicos, como S. Wold (Umea University – Suécia), se intitularam como quimiométricos dando origem a uma nova área dentro da química analítica. [10-12]

A quimiometria enfoca o estudo de medidas analíticas baseando-se na idéia 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, monitoramento e modelagem de processos multivariados, métodos de inteligência artificial, 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. [11-17]

2.1. 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. [18]

Um modelo de calibração, na verdade, é uma função matemática (f) que relaciona dois grupos de variáveis, uma delas denominada dependente (Y) e a outra denominada independente (X):

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

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. [19, 20]

O passo seguinte à calibração é a validação. Nesta etapa, as variáveis independentes obtidas, para um 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 validação 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. [20]

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. [21]

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. [22-23]

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. Neste tipo de calibração a resposta instrumental é representada na forma de matriz, enquanto a propriedade de interesse, determinada por uma metodologia padrão, é representada por um vetor. [20] A Figura 3 ilustra como uma matriz de dados X de dimensão $n \times m$, ou seja, n objetos (espectros) e m variáveis (números de onda), pode ser construída a partir de um vetor de respostas instrumental.

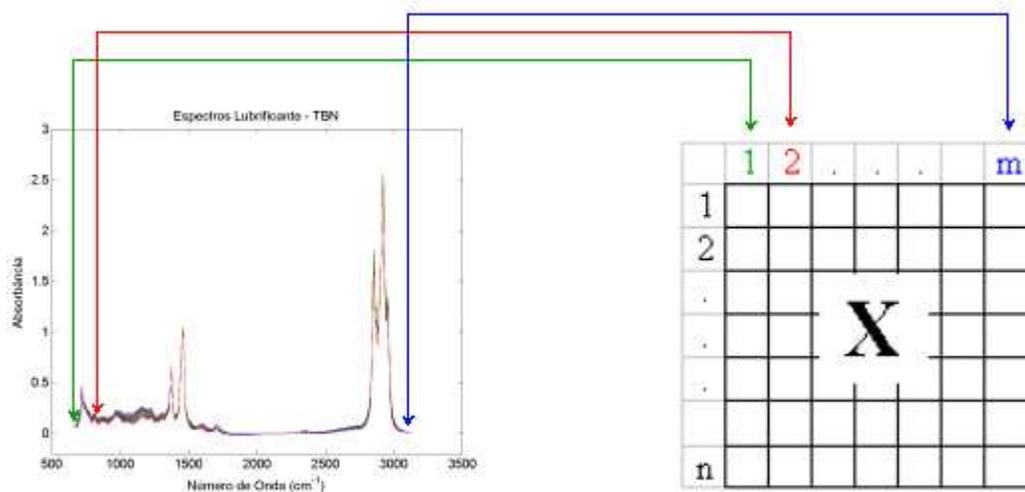


Figura 3. Representação da construção da matriz X para calibração multivariada.

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 – do inglês, Multiple Linear Regression) [16, 24], que utiliza a absorvância em comprimentos de onda selecionados, ou métodos de modelagem global, tais como a Regressão por Componentes Principais (PCR – do inglês, Principal Component Regression) ou Mínimos Quadrados Parciais (PLS – do inglês, Partial Least Squares). [16, 25]

Antes de tratarmos da regressão por PLS vamos introduzir o conceito da análise de componentes principais (PCA – do inglês, Principal Component Analysis) uma vez que nessa regressão, utiliza-se a PCA como a técnica de ortogonalização baseada em uma mudança da base vetorial.

2.1.1. Análise de Componentes Principais - PCA

A PCA foi introduzida na química por Malinowski no final dos anos 1960, com o nome de Análise de Fatores, e a partir da década seguinte uma série de aplicações foi desenvolvida, o que a tornou muito conhecida e explorada. [16, 17]

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 temos:

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

em que Λ é uma matriz diagonal cujos elementos são autovalores de Z , 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 X e a projeção dos dados nessa base é denominada scores (escores), (T). Desse modo, os dados são decompostos por um conjunto de vetores pesos e escores denominado componente principal (PC):

$$X = TP^T \quad (4)$$

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

$$PX = TP^T P \quad (5)$$

$$T = XP \quad (6)$$

A Figura 4 ilustra a decomposição da matriz X de dimensão ($n \times m$) pela análise de componentes principais até A componentes principais.

$$\begin{array}{c} m \\ \square \\ n \end{array} X = \begin{array}{c} 1 \\ \square \\ n \end{array} t_1 \begin{array}{c} 1 \\ \square \\ 1 \end{array} p_1^T + \begin{array}{c} 1 \\ \square \\ n \end{array} t_2 \begin{array}{c} 1 \\ \square \\ 1 \end{array} p_2^T + \dots + \begin{array}{c} 1 \\ \square \\ n \end{array} t_A \begin{array}{c} 1 \\ \square \\ 1 \end{array} p_A^T$$

Figura 4. Decomposição em componentes principais por PCA.

O número máximo de componentes principais obtidos (PCs) 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 freqüentemente aumenta a relação sinal/ruído. [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. [18, 25]

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. [18, 25]

2.1.2. Mínimos Quadrados 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 de que estes também estejam presentes por ocasião da construção do modelo (vantagem de primeira ordem). [17]

Para o método de regressão PCR, a decomposição da matriz \mathbf{X} realizada pela PCA é feita de forma independente da matriz \mathbf{Y} . Enquanto que para o método de regressão PLS a informação de \mathbf{Y} é incorporada, de forma que cada PC do modelo sofre uma pequena modificação para buscar a máxima covariância entre \mathbf{X} e \mathbf{Y} e passa a receber a terminologia de Variável Latente (VL). [11]

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). [16, 24]

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 \mathbf{X} de dimensão $(n \times m)$ e \mathbf{Y} em suas matrizes de escores e pesos:

$$\mathbf{X} = \mathbf{TP}^T + \mathbf{E}_X = \sum \mathbf{t}_h \mathbf{p}_h^T + \mathbf{E}_X \quad (7)$$

$$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{E}_Y = \sum \mathbf{u}_h \mathbf{q}_h^T + \mathbf{E}_Y \quad (8)$$

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. [15-17]

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

$$\mathbf{u}_h = \mathbf{b}_h \mathbf{t}_h \quad (9)$$

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

$$\mathbf{b}_h = \frac{\mathbf{u}_h^T \mathbf{t}_h}{\mathbf{t}_h^T \mathbf{t}_h} \quad (10)$$

A Figura 5 ilustra a decomposição das matrizes \mathbf{X} e \mathbf{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 Iterative Partial Least Squares). [25]

$$\begin{matrix} \begin{matrix} m \\ \mathbf{X} \\ n \end{matrix} = \begin{matrix} 1 \\ \mathbf{t}_1 \\ n \end{matrix} \begin{matrix} m \\ \mathbf{p}_1^T \\ 1 \end{matrix} + \dots + \begin{matrix} 1 \\ \mathbf{t}_h \\ n \end{matrix} \begin{matrix} m \\ \mathbf{p}_h^T \\ 1 \end{matrix} + \begin{matrix} m \\ \mathbf{E}_X \\ n \end{matrix} \\ \\ \begin{matrix} z \\ \mathbf{Y} \\ n \end{matrix} = \begin{matrix} 1 \\ \mathbf{u}_1 \\ n \end{matrix} \begin{matrix} 1 \\ \mathbf{q}_1^T \\ 1 \end{matrix} + \dots + \begin{matrix} 1 \\ \mathbf{u}_h \\ n \end{matrix} \begin{matrix} 1 \\ \mathbf{q}_h^T \\ 1 \end{matrix} + \begin{matrix} z \\ \mathbf{E}_Y \\ n \end{matrix} \end{matrix}$$

Figura 5. Decomposição em variáveis latentes das matrizes \mathbf{X} e \mathbf{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. [16, 25]

Neste trabalho, os modelos PLS foram construídos usando o PLS_Toolbox versão 2.0 (Eigenvector Research) para Matlab 6.1 (MathWorks).

2.2. Métodos de Seleção de Variáveis

Os métodos de calibração multivariada podem ser aplicados para determinações seletivas e seguras do analito de interesse usando dados espectrais, em que o analito e a composição da matriz amostral estão variando. O método PLS consiste num método flexível de espectros totais, sendo que a informação significativa contida nos espectros está concentrada em algumas VLs que são aperfeiçoadas para produzir a melhor correlação com a informação de interesse. Porém, pode ser discutível se o método é capaz de determinar claramente somente as características espectrais cruciais para a construção de modelos de previsão. [16]

Alguns métodos têm sido descritos recentemente na literatura para implementar seleção de região espectral para melhorar significativamente o desempenho dos métodos de calibração de espectros totais. Esses métodos são chamados métodos de seleção de variáveis, que selecionam regiões específicas do espectro (um comprimento de onda ou um conjunto de comprimentos de onda) em que a colinearidade não é tão importante, enquanto gera modelos mais estáveis, robustos e mais simples de interpretar. Na prática, a filosofia está baseada na identificação de um subconjunto dos dados inteiros que produzirão erros de previsão mais baixos. Assim, em espectros vibracionais por exemplo, os comprimentos de onda que apenas induzem a ruídos, informações irrelevantes ou não-linearidades, podem ser eliminados. [26-31]

Os métodos de seleção de variáveis existentes se diferem com relação ao procedimento realizado na seleção da região espectral. Dentre os métodos utilizados atualmente podemos destacar o método de mínimos quadrados parciais por intervalos (iPLS – do inglês, Interval Partial Least Square), [27, 28] o método de eliminação de variáveis não informativas por mínimos quadrados parciais (UVE-PLS – do inglês, Elimination of Uninformative Variables in Partial Least Square) [26, 30] e algoritmo genético (GA – do inglês, Genetic Algorithm). [26, 31] que serão apresentados a seguir pois foram utilizados neste trabalho.

2.2.1. PLS por intervalos (iPLS)

O método iPLS é uma extensão iterativa para o PLS, que desenvolve modelos locais PLS em subintervalos equidistantes de toda a região do espectro.

Seu principal objetivo é prever informação relevante do espectro global nas diferentes subdivisões, de modo que as regiões espectrais cujas variáveis se apresentam como supostamente de menor relevância e/ou interferentes, podem ser removidas, e um novo modelo PLS é construído a partir das variáveis selecionadas. [27]

A sensibilidade do algoritmo PLS para variáveis não informativas é salientada pelos gráficos do iPLS. Os modelos do iPLS são desenvolvidos em subintervalos do espectro com igual peso, e a previsão destes modelos locais e o modelo global é comparado. A comparação é baseada no parâmetro de validação RMSECV (erro quadrático médio da validação cruzada), mas outros parâmetros são também utilizados como r^2 (coeficiente de correlação) e inclinação da reta do gráfico dos valores reais e previstos pelo modelo.

As regiões espectrais com valores de parâmetro de validação menor que o modelo global são selecionadas para a construção de modelos PLS que são comparados com a eficiência do modelo PLS global. Amostras e/ou medidas anômalas detectadas pelo PLS devem ser geralmente removidas antes da aplicação do iPLS.

As Figuras 6 e 7 exemplificam para o modelo de calibração de amostras de lubrificante contaminadas com gasolina, os gráficos do iPLS. Nessa seleção de variáveis foi utilizado o programa iPLS (Lars Nørgaard, Chemometrics Group - KVL, Dinamarca, 2004) do iToolbox para Matlab 6.1 (MathWorks). [27]

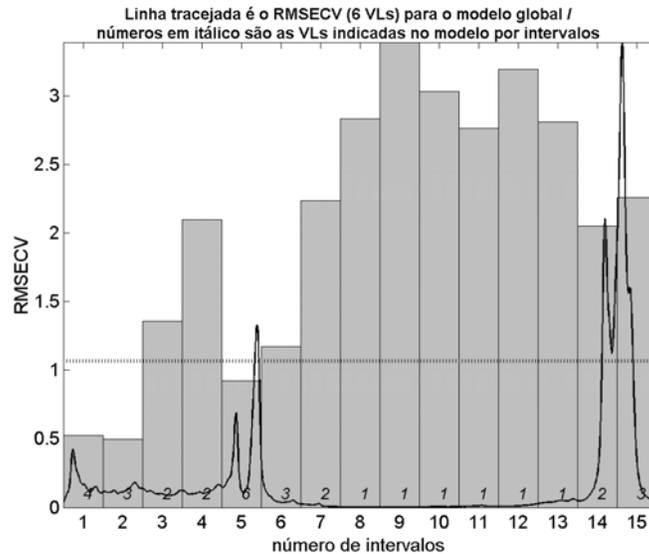


Figura 6. Gráfico do modelo iPLS do número de intervalos versus RMSECV.

Os números no interior das faixas no gráfico da Figura 6, são os números de VLs indicados pelo modelo iPLS para serem usados na construção dos modelos PLS com as variáveis selecionadas em cada intervalo. Nota-se que o 2º intervalo é o que possui menor valor de RMSECV e por isso selecionado como mostrado na Figura 7.

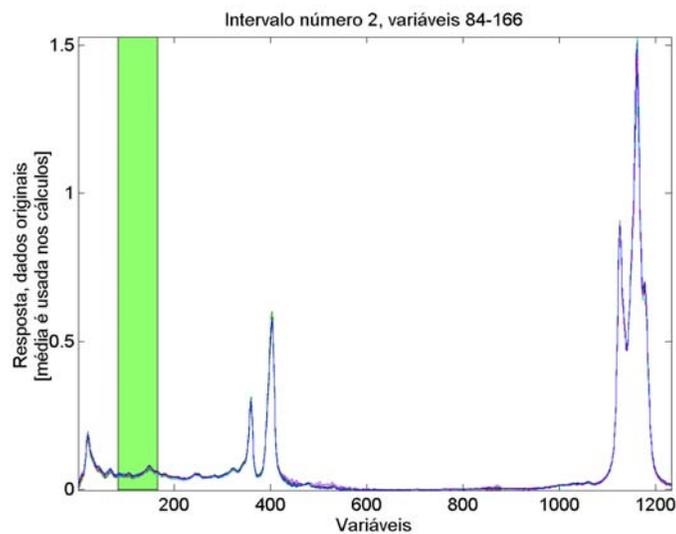


Figura 7. Gráficos do modelo iPLS: (a) número de intervalos versus RMSECV e (b) faixa escolhida para a construção do modelo PLS.

Assim, no exemplo do modelo de calibração de amostras de lubrificante contaminadas com gasolina, o modelo iPLS deve ser construído a partir de um

modelo PLS usando 3 VLs e ao invés das 1.232 variáveis dos espectros totais as 83 variáveis selecionadas no 2º intervalo, ou seja, as variáveis de 84 a 166 correspondentes aos números de onda de 845 a 1.003 cm^{-1} .

2.2.2. Eliminação de variáveis não informativas por PLS (UVE-PLS)

No método UVE-PLS, o algoritmo descobre e elimina de um modelo PLS as variáveis não informativas. O critério usado para distinguir as variáveis informativas e não informativas, é a confiança (estabilidade) dos coeficientes de regressão b , obtidos por validação cruzada do tipo “leave-one-out” (deixe um fora por vez) jackknifing ($i = 1, \dots, n$). Em que um coeficiente de regressão b é deixado de fora na estimativa da média dos coeficientes de regressão b_j ($j = 1, \dots, n$) e do desvio padrão do vetor de n coeficientes b_{ij} , $s(b_j)$ e a seguir o coeficiente deixado de fora é previsto pelo modelo de regressão. Um nível de corte t_j , abaixo do qual esses coeficientes são considerados muito pequenos indicando que a variável correspondente pode ser removida, é calculado por: [30]

$$t_j = b_j/s(b_j) \quad \text{para } j = 1, \dots, n \quad (11)$$

Porém, os desvios padrão $s(b_j)$ para os coeficientes de regressão b para o modelo PLS não podem ser estimados diretamente, sendo necessário usar uma matriz de variáveis randômicas (que simula o ruído dos dados) anexada artificialmente aos dados experimentais.

A partir deste ponto, descreveremos passo a passo o processamento do algoritmo. [26, 30]

Primeiro o algoritmo cria uma matriz de números randômicos (0,0-1,0) com mesma dimensão da matriz de dados \mathbf{X} . Os números são multiplicados por uma constante pequena (por exemplo, 10^{-10}) dando-lhes pelo menos uma ordem de magnitude menor que a imprecisão do instrumento. A multiplicação retém a variação das variáveis, mas torna a influência delas no modelo desprezível.

A nova matriz é acrescentada à matriz de dados X original formando uma matriz estendida com o dobro do número de variáveis da matriz original.

Modelos PLS são construídos para cada amostra usando validação cruzada do tipo “deixe uma fora por vez”, em que cada modelo contém todas amostras menos uma. Isto conduz a uma matriz de coeficientes de regressão b com n linhas (amostras) e uma coluna para cada variável, original e randômica.

Os valores de t são calculados conforme a eq. 11 como sendo a média dos coeficientes b de cada coluna (variáveis) dividida pelo desvio padrão daquela coluna.

Por fim, um limite de corte é fixado por uma faixa com os maiores valores positivos e negativos de t calculados para as variáveis randômicas. Todas variáveis com valor de t igual ou mais baixo, ou seja, dentro desta faixa, são eliminadas do modelo final. Isto significa que todas as variáveis randômicas e originais assumindo conter nada mais que ruído são eliminadas.

A Figura 8 exemplifica para o modelo de calibração de amostras de óleo de Oliva contaminadas com óleo de Milho, o gráfico do UVE-PLS. Nessa seleção de variáveis foi utilizado o programa PLSUVE versão 1.1 (Vita Centner, ChemoAC, Bélgica, 1998) para Matlab 6.1 (MathWorks). [30]

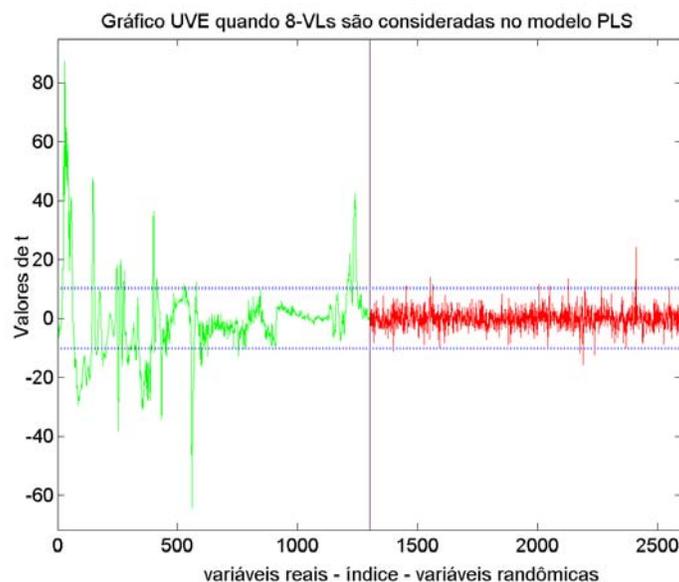


Figura 8. Gráfico do modelo UVE-PLS: valores de t para as variáveis experimentais (1-1.300, em verde) e randômicas (1.301-2.600, em vermelho). O limite de corte é indicado pela linha azul tracejada.

Um novo modelo PLS é construído a partir das variáveis selecionadas pelo UVE-PLS e sua eficiência comparada ao modelo PLS global. A Figura 9 indica as variáveis selecionadas pelo modelo UVE-PLS para o exemplo do modelo de calibração de amostras de óleo de Oliva contaminadas com óleo de Milho.

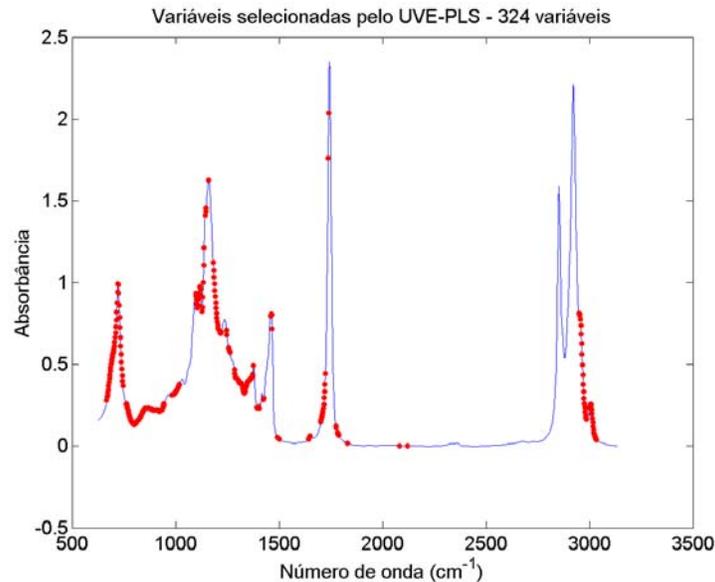


Figura 9. Gráfico das variáveis selecionadas pelo modelo UVE-PLS, em vermelho.

2.2.3. Algoritmo Genético (GA)

O método GA é baseado na teoria da evolução das espécies de Darwin e é utilizado em processos de otimização. No caso da espectroscopia vibracional, cada espectro é tratado como sendo um cromossomo (indivíduo) composto por um conjunto de genes (números de onda por exemplo) que são dispostos em código binário. Cada variável ou número de onda pode receber a codificação binária “1” ou “0” (selecionada ou não selecionada), como representado na Figura 10. [26, 31]

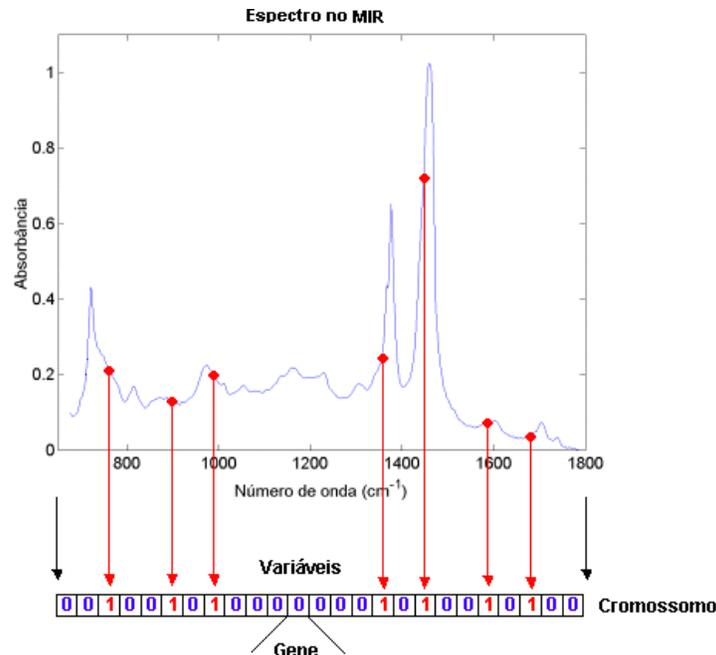


Figura 10. Representação da codificação binária de um espectro na forma de um cromossomo. Cada número de onda corresponde a um gene do cromossomo. As variáveis selecionadas (em vermelho) recebem código 1 e as não selecionadas (em azul) recebem código 0.

O cromossomo original é perturbado aleatoriamente criando vários cromossomos que formam a população inicial. Para cada cromossomo, é avaliada a resposta associada com as condições experimentais correspondentes. Isto é feito construindo um modelo PLS para cada cromossomo (também pode ser usado a MLR), levando em consideração apenas as variáveis que foram codificadas como “1”. O modelo é então avaliado através de validação cruzada para adquirir um valor de aptidão que descreve a qualidade do modelo. [26, 32]

A partir da população inicial, uma nova população a qual pode ser considerada como próxima geração, é obtida pelo cruzamento aleatório entre material genético de cromossomos diferentes. No cruzamento, dois cromossomos pai são divididos geralmente em duas ou três partes, cada uma escolhida aleatoriamente, que são cruzadas e combinadas para formar dois cromossomos filhos que substituirão os cromossomos pai dentro de uma nova geração. Uma nova avaliação é realizada, e os cromossomos com valores de aptidão maiores têm uma probabilidade de reprodução maior que os cromossomos com valores menores, tudo para melhorar a aptidão global da população. [26, 31, 32]

Para uma melhor compreensão da etapa de cruzamento, vamos considerar o exemplo para dois cromossomos pai com dez genes (variáveis), representado na Figura 11, em que demonstraremos um cruzamento entre três partes dos cromossomos pai dando origem a dois cromossomos filhos.

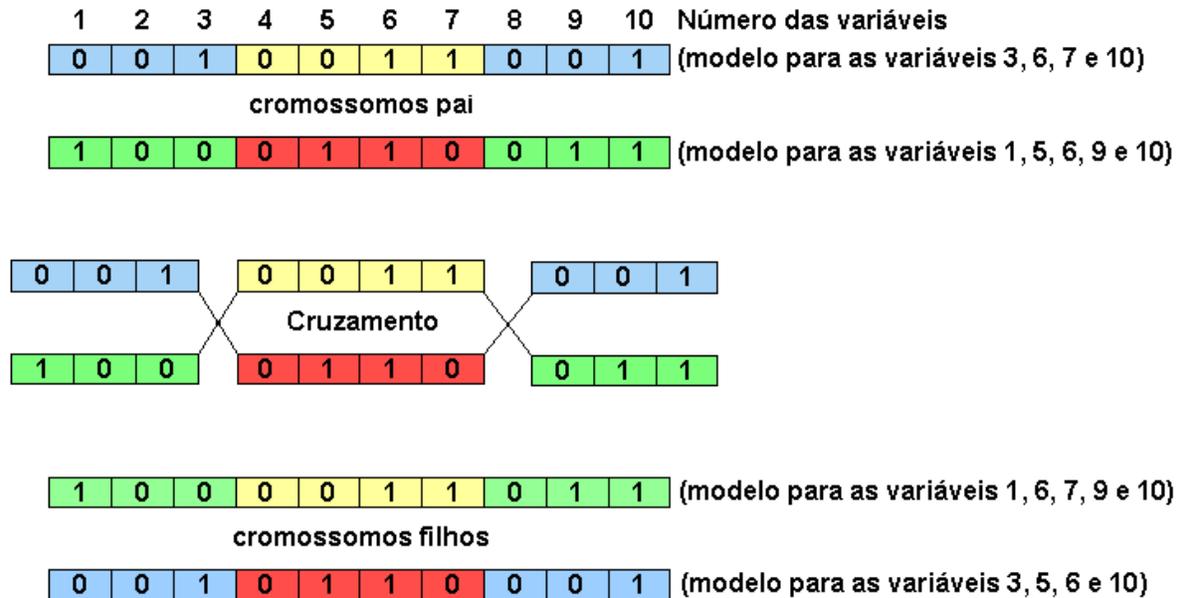


Figura 11. Representação do cruzamento entre dois cromossomos pais originando dois cromossomos filhos. Havendo rompimento e troca de material genético do cromossomo pai 1 (em azul) com o material genético do cromossomo pai 2 (em verde).

Mutações podem ser incorporadas ao modelo e são, às vezes, necessárias para superar alguns problemas na população, sendo utilizadas para:

- dar nova informação genética à população, ou seja, uma variável não selecionada em quaisquer dos cromossomos originais, nunca seria selecionada na próxima geração se mutações não tivessem presentes;
- prevenir que a população se sature com cromossomos semelhantes (convergência prematura).

Uma mutação nada mais é que a inversão de um gene no cromossomo. A taxa de mutação é usualmente definida e fixada de 0,001-0,01. Ainda usando o exemplo para dois cromossomos pai com dez genes (variáveis), só que agora a partir dos cromossomos filhos, podemos representar a mutação do gene 2 do cromossomo filho 2, conforme mostrado na Figura 12.

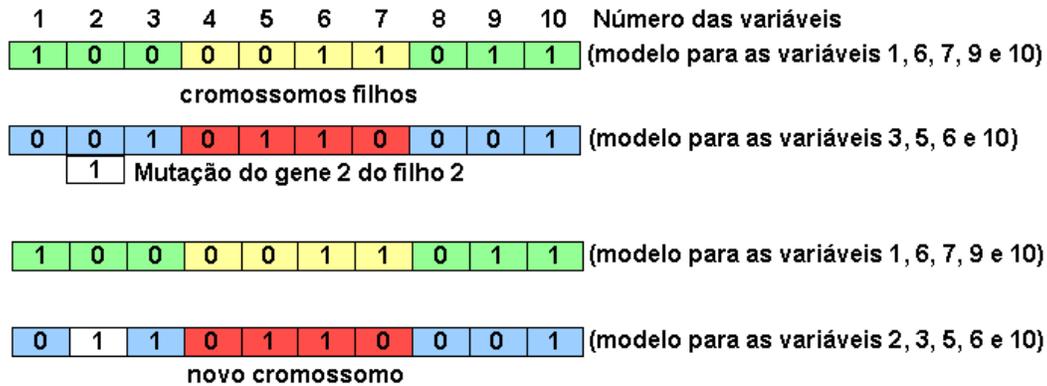


Figura 12. Representação da mutação do gene 2 do cromossomo filho 2, gerando um novo cromossomo.

O algoritmo é repetido até que a condição de término é cumprida. A condição de término é baseada no critério de convergência, em que o algoritmo é encerrado quando uma certa porcentagem dos cromossomos for idêntica ou quando um determinado número de gerações é atingido. [26, 32]

As variáveis do cromossomo mais apto são selecionadas para a construção do modelo de calibração (GA-PLS) e sua eficiência comparada ao modelo PLS global.

A Figura 13 exemplifica para o modelo de calibração de amostras de lubrificante contaminadas com diesel, os gráficos do GA-PLS.

Na avaliação da resposta, ou seja aptidão, deve-se encontrar o valor associado à eficiência de cada cromossomo relacionado ao sistema de interesse, sendo o resultado mais importante no procedimento do algoritmo genético. A aptidão é uma característica intrínseca ao indivíduo, que representa sua habilidade de produzir a melhor resposta. O objetivo é encontrar o menor erro possível, e este será o responsável direto pela vida ou morte dos indivíduos. [31]

Após as gerações, o número de vezes em que as variáveis selecionadas aparecem nos indivíduos mais aptos é representado pela frequência das variáveis incluídas nos modelos.

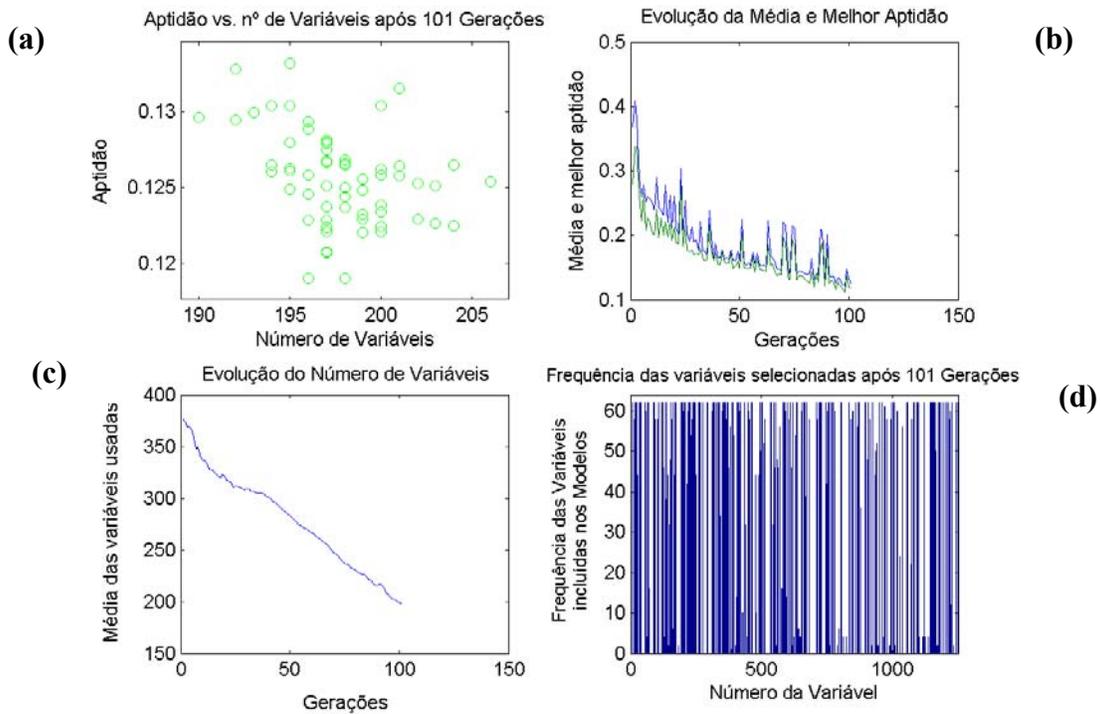


Figura 13. Gráficos do modelo GA-PLS: (a) aptidão versus número de variáveis após as gerações, (b) evolução da média e melhor aptidão, (c) evolução do número de variáveis selecionadas e (d) frequência das variáveis selecionadas após as gerações.

A Figura 14 indica as variáveis selecionadas pelo modelo GA-PLS para o exemplo do modelo de calibração de amostras de lubrificante contaminadas com diesel.

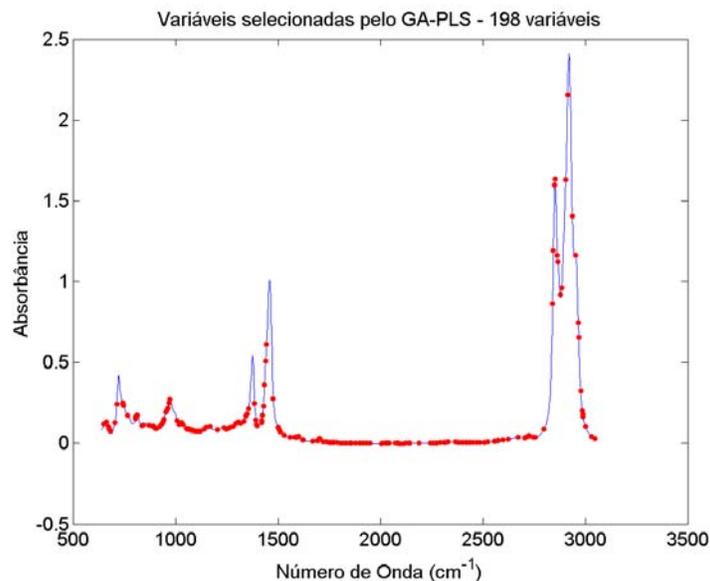


Figura 14. Gráfico das variáveis selecionadas pelo modelo GA-PLS, em vermelho.

A seleção de variáveis por GA tem como desvantagens a possibilidade de apresentar sobre-ajuste no modelo de calibração (devido ao número de gerações ou dimensão da matriz de dados) e o tempo de processamento menos relevante com a evolução dos novos computadores. [26, 31, 32]

Apesar do GA ter início a partir de populações diferentes, devido à natureza probabilística, a falta de reprodutibilidade não representa uma desvantagem pois, mesmo selecionando variáveis diferentes, ele é capaz de manter a informação.

Na tentativa de evitar o sobre-ajuste no modelo de calibração ou até mesmo obter um modelo de calibração com melhor habilidade de previsão, uma alternativa é usar como dados de entrada no algoritmo genético as variáveis selecionadas pelos métodos iPLS (GA-iPLS) ou UVE-PLS (GA-UVE-PLS). Para a seleção de variáveis com algoritmo genético foi utilizado o programa GASELCTR do PLS_Toolbox versão 2.0 (Eigenvector Research) para Matlab 6.1 (MathWorks) nas seguintes condições:

- Tamanho da população: 64 cromossomos;
- Número máximo de gerações: foram avaliados de 20 a 150;
- Convergência: foram avaliados 50 e 80%;
- Taxa de mutação: 0,005 (0,5%);
- Número de variáveis por gene: 1;
- Número de indivíduos na população inicial: 30%;
- Número de partes em que os cromossomos são divididos: 2;
- Algoritmo: PLS;
- Número máximo de VLs para o PLS: 10;
- Tipo de Validação Cruzada: divisão em blocos contínuos;
- Número de blocos para a Validação Cruzada: 5;
- Pré-processamento: nenhum;
- Penalidade: 20% das variáveis incluídas no modelo.

2.3. Cálculo dos erros

A eficiência dos modelos de calibração multivariada pode ser avaliada pelo cálculo dos valores de erro quadrático médio (RMSE – do inglês, root mean square error) e erro relativo (E). Esses valores expressam a exatidão do modelo, ou seja, a proximidade entre o valor calculado pelo modelo (y_{prev}) e o valor verdadeiro ou obtido por um método de referência (y_{real}). Os erros são definidos como: [25, 33]

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (y_{prev} - y_{real})^2}{N}} \quad (12)$$

em que N é o número de amostras.

$$E = \sqrt{\frac{(y_{prev} - y_{real})^2}{(y_{real})^2}} \times 100\% \quad (13)$$

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.

2.4. Programas desenvolvidos

Às vezes, os programas desenvolvidos para o Matlab por serem generalizados, não apresentam todas as respostas na forma pretendida pelo usuário ou não se encontram disponíveis no mesmo programa. Entretanto, uma das maiores vantagens do Matlab é permitir que com o conhecimento adequado, seja possível desenvolver e/ou adaptar novos programas a partir dos algoritmos e programas existentes. Isto é feito com o objetivo de atender ou otimizar as necessidades do usuário.

Neste trabalho, em parceria com o doutorando Luiz Carlos Moutinho Pataca, também do grupo LAQQA (Laboratório de Quimiometria em Química Analítica), foram desenvolvidos os programas tp (tago&passoca) para Matlab:

1) **tp_dados** – para a aquisição dos dados multivariados (espectros), calcular os espectros médios das replicatas das amostras, construção das matrizes de dados dos conjuntos de calibração e previsão, tratamento dos dados e graficar os espectros.

2) **tp_er_pls** – calcular ao mesmo tempo as previsões para seis modelos PLS sendo, dois anteriores e três posteriores ao número de VLs escolhido; calcular os valores de RMSEP, E e a média do erro relativo (EM), para cada modelo PLS; e graficar os valores reais versus previstos para o modelo PLS com menor EM.

3) **tp_grafico_ipls** – graficar as variáveis selecionadas pelo modelo iPLS.

4) **tp_uve_pls** – selecionar as variáveis informativas para o modelo UVE-PLS, construção das matrizes de dados dos conjuntos de calibração e previsão com as variáveis selecionadas, executar o programa **tp_er_pls** e graficar as variáveis selecionadas pelo modelo UVE-PLS.

5) **tp_ga_pls** – selecionar as variáveis informativas para o modelo GA, construção das matrizes de dados dos conjuntos de calibração e previsão com as variáveis selecionadas, executar o programa **tp_er_pls** e graficar as variáveis selecionadas pelo modelo GA.

Esses programas encontram-se disponíveis no LAQQA para os interessados na sua utilização.

CAPÍTULO 3

Lubrificantes

3. Lubrificantes

Os óleos lubrificantes geralmente obtidos do petróleo consistem de uma mistura complexa de hidrocarbonetos de massas molares entre 250-1.000 g mol⁻¹. A formulação dos lubrificantes é obtida misturando até a homogeneidade, uma combinação de óleos básicos e aditivos, estudada e definida para garantir em exercício um nível de exigência adequado no campo de utilização ao qual o lubrificante é destinado.

Os óleos básicos representam o componente ainda preponderante na grande maioria dos lubrificantes. Ao lado das bases minerais, obtidas através do refino do petróleo, adquirem sempre maior importância que as bases sintéticas, assim definidas porque não estão presentes no petróleo ou porque não são obtidas pelo simples tratamento físico-químico das suas frações. [34, 35]

Na formulação, além dos óleos básicos, também estão presentes as bases rerrefinadas, que são os produtos do tratamento químico de lubrificantes usados. Somados a estes, estão os aditivos, que tem por objetivo melhorar as características físicas das bases, como: modificadores de viscosidade, demulsificantes, anti-desgaste, detergentes, dispersantes, anti-corrosivos, anti-oxidantes, entre outros. [34-36]

Na produção de lubrificantes, o petróleo é submetido inicialmente à destilação primária inicial ou topeamento (do inglês, topping), que vem a ser a remoção, por destilação, das frações mais leves. A seguir é feita a destilação a vácuo, separando as diversas frações. [35]

A fração de óleos lubrificantes é submetida a tratamentos subseqüentes, tais como remoção de parafina, remoção de asfalto, refinação ácida, refinação por solventes, etc. Os óleos aromáticos não são adequados para fins de lubrificação. Os óleos lubrificantes minerais podem, então, ser classificados, de acordo com sua origem, em naftênicos e parafínicos. [34]

As principais funções dos lubrificantes são: reduzir a fricção entre superfícies, diminuindo assim o desgaste das mesmas, prevenir a corrosão e a ferrugem, atuar como refrigerante retirando o calor produzido pelo contato das

peças, a limpeza do motor, entre outras. Dessa forma, a qualidade do óleo lubrificante é essencial para a preservação do desempenho e longevidade de motores e máquinas. [34, 35]

Em modelos comerciais de automóveis há diversos pontos de lubrificação: motor, pedais do freio e da embreagem, rolamentos das rodas, eixo traseiro, juntas esféricas, articulação da suspensão, barra da direção entre outros. A Figura 15 apresenta um esquema básico da lubrificação em motores automotivos e a Figura 16 um motor danificado pelo uso de lubrificante degradado.

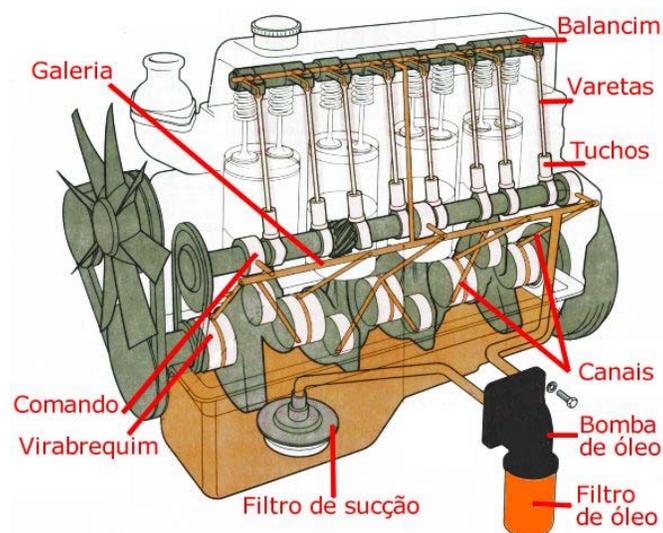


Figura 15. Esquema básico da lubrificação em motores automotivos.



Figura 16. Motor automotivo danificado pelo uso de lubrificante degradado.

Ao longo do processo de uso de um lubrificante podem ocorrer: a contaminação por certos agentes como combustível (gasolina, diesel ou álcool) e refrigerantes (água e etileno glicol); a perda da atividade dos aditivos e o surgimento de produtos da oxidação. [37, 38] Tais contaminações interferem diretamente na viscosidade do lubrificante diminuindo sua eficiência. Uma vez degradado, o lubrificante pode causar desgaste mecânico comprometendo as funções da máquina, podendo gerar prejuízos irreversíveis e de alto custo.

Na presença de oxigênio, que é presente na maioria dos ambientes de lubrificação, e a altas temperaturas, os hidrocarbonetos que compõem os óleos básicos, podem reagir formando produtos contendo carbonilas (produtos de oxidação primária) que em seguida sofrem mais oxidação para produzir ácidos carboxílicos, aldeídos, álcoois, (produtos de oxidação secundária), resultando no aumento dos valores do número total de acidez (TAN). Além disso, os produtos de oxidação formados podem polimerizar aumentando a viscosidade do óleo. A reação de oxidação simplificada é apresentada a seguir: [39-41]



As análises mais comuns no controle de qualidade de lubrificantes são: viscosidade, número total de acidez (TAN) e número total de basicidade (TBN). [35] Outras análises importantes são: determinação de gasolina em óleo por Cromatografia Gasosa [42], determinação de água em produtos de petróleo por um método baseado na reação de Karl-Fisher [43] e determinação de etileno glicol em lubrificante usado por detecção colorimétrica após formação de complexo. [44] Tais métodos consomem tempo de análise, reagentes, são destrutivos e geram resíduos tóxicos.

A viscosidade é a característica mais importante no controle do uso do lubrificante. É pouco provável que um aumento da viscosidade atinja o limite máximo permissível. Como regra orientadora, pode-se dizer que: [35]

1) um aumento na viscosidade pode ser devido a:

- presença de insolúveis;
- produtos oxidados;
- reposição com óleo mais viscoso;
- presença de água.

2) uma diminuição na viscosidade pode ser devido a:

- presença de combustível;
- reposição com óleo menos viscoso;
- cisalhamento do aditivo melhorador do índice de viscosidade.

Nos últimos anos a espectroscopia vibracional tem sido usada para investigar vários parâmetros de qualidade de produtos vindos do petróleo. As técnicas mais efetivas têm sido a espectroscopia vibracional e fluorescência no UV. A espectroscopia no infravermelho tem sido usada para monitorar a presença de contaminantes como a água, combustível, etileno glicol, fluidos hidráulicos, bem como a formação de produtos de oxidação, nitração, as mudanças no TBN e na viscosidade em lubrificantes. [6, 45-48]

3.1. Objetivo

Esta etapa do trabalho tem como objetivo, a construção de modelos de calibração multivariada por PLS e seleção de variáveis, comparando-os quanto à habilidade de quantificar parâmetros utilizados no controle de qualidade de lubrificantes através de espectroscopia no infravermelho médio.

3.2. Parte Experimental

3.2.1. Coleta das amostras de lubrificante de ônibus

As coletas foram realizadas em motores de ônibus da Empresa Rápido Luxo Campinas – Valinhos. A partir do monitoramento de sete ônibus que

transitam percursos em perímetro urbano, dos quais foram coletadas 48 amostras em motores com percursos de 522 Km a 45.896 Km.

O lubrificante utilizado nas medidas de TBN e viscosidade para todos os ônibus foi SHELL® - Rímula Plus, que possui quando novo, viscosidade a 100°C de 14,06 cSt (centiStokes) e TBN de 10,1 mg KOH/g de amostra.

3.2.2. Medida do Número Total de Basicidade (TBN)

A medida do TBN indica a quantidade de ácido, expressa em mg equivalente de KOH, necessária para neutralizar todos os constituintes básicos (alcalinos) em 1,0 g de óleo. Qualquer óleo cujo pH esteja compreendido entre valores 4,0 e 11,0 apresentará um TBN, expresso em mg KOH/g, equivalente à quantidade de ácido necessária para levar o pH do valor em que está até o pH 4,0 e um TAN expresso em mg KOH/g necessário para elevar o pH até o limite superior, pH 11,0. [49]

Para as 48 amostras de óleos de ônibus foram realizadas medidas em duplicatas do TBN, utilizando o procedimento descrito na norma ASTM D4739 [50], na empresa Wal Química. Este método visa a determinação de constituintes básicos contidos em: óleos lubrificantes automotivos, sulfonados de alta basicidade e aditivos. Produtos de petróleo, novos e usados, podem conter constituintes básicos que estão presentes como aditivos ou como produtos de degradação formados durante o uso.

A quantidade relativa destes materiais foi determinada a partir de uma quantidade de amostra calculada em que se adiciona um solvente de titulação, pela titulação potenciométrica semi-automática, com solução alcoólica de HCl 0,1 mol L⁻¹, iniciando com a adição de 0,1 mL a cada 90 segundos até atingir pH 5,0. Em seguida adiciona-se 0,01 mL até pH 4,0. A reação do HCl com os componentes básicos é muito lenta na maior parte das titulações.

A basicidade é usada como dado importante no controle de qualidade das formulações de óleos lubrificantes e na medida da degradação do lubrificante em uso.

3.2.3. Medida de Viscosidade a 100°C

Foram realizadas, pela empresa Wal Química, medidas de viscosidade a 100°C para as mesmas 48 amostras de lubrificante coletadas de ônibus, baseadas na norma ASTM D445. [51]

Pelo procedimento ASTM D445, na medida de viscosidade determina-se o tempo, em segundos (s), para um volume fixo de líquido escoar por gravidade, através do capilar de um viscosímetro calibrado, a uma temperatura rigorosamente controlada, no caso de 100°C. A viscosidade cinemática é o produto da medida do tempo de escoamento (em s) e da constante de calibração do viscosímetro (em cSt/s). Essas medidas foram realizadas em duplicatas.

3.2.4. Contaminação em lubrificantes automotivos

Foram realizadas as contaminações artificiais dos óleos para motor Agip SJ 50 com água e gasolina e Agip SMO SJ 20W-50 com diesel e etileno glicol. As concentrações de água deionizada adicionadas ao lubrificante foram de 0,043 a 0,694% (m/m), e de etileno glicol de 0,05% a 2,05% (m/m), visto que os níveis limites desses contaminantes encontrados em motor automotivo são de 0,3% do volume. Também foram adicionados 8,0% (m/m) de aditivo detergente Infineum P5296, já presente na formulação do óleo fornecido pela Wal Química, fornecedora da Agip. O objetivo do detergente é a homogeneização das amostras durante a contaminação do lubrificante com água e com etileno glicol.

Para a contaminação com gasolina e diesel foram utilizados: gasolina tipo A e diesel tipo D, fornecidos pela refinaria Replan-Petrobras. Como a porcentagem limite em motores é de 2,0% do volume, a faixa de contaminação variou de 0,32 a 8,03% (m/m) para a gasolina e de 0,20 a 8,15% (m/m) para o diesel. Nesse procedimento, foi tomado muito cuidado, fechando-se bem os frascos, para evitar a evaporação destes contaminantes.

3.2.5. Espectroscopia no infravermelho médio de lubrificantes

Os espectros das amostras de lubrificante contaminadas com água e gasolina foram obtidos, em triplicata, no espectrofotômetro Nicolet - 520 FTIR, na região do MIR, abrangendo a região de 650 a 4.000 cm^{-1} . Os espectros foram obtidos com uma resolução de 4 cm^{-1} e 128 varreduras. Foi utilizado acessório de ATR do tipo circle-cell de cristal de ZnSe. Este acessório, além de facilitar o manuseio da amostra, possibilita a radiação incidir em alguns pontos do óleo sem chegar atravessá-lo assim, a presença de fuligem no óleo não interferirá na análise. Caso contrário, a radiação em contato com o óleo espalharia. [1, 8, 38] Os espectros das amostras de lubrificante contaminadas com água e gasolina, usados na construção dos modelos de calibração multivariada e seleção de variáveis foram cedidos pela aluna Alessandra Borin do grupo LAQQA.

Já os espectros das amostras de lubrificante de motores de ônibus e das contaminadas com diesel e etileno glicol foram obtidos também em triplicata, na região de 650 a 4.000 cm^{-1} e resolução de 4 cm^{-1} , porém, usando um espectrofotômetro Bomem – MB100 FTIR com acessório de ATR horizontal de cristal de ZnSe e 64 varreduras.

Todos os espectros obtidos foram posteriormente tratados usando o programa baseline do PLS_Toolbox versão 2.0 (Eigenvector Research) para Matlab 6.1 (MathWorks), para eliminar problemas com a linha base dos espectros.

3.3. Resultados e discussão

3.3.1. Os espectros de lubrificantes na região do infravermelho médio

Os espectros de lubrificantes na região do MIR apresentam bandas características, relacionadas a compostos aromáticos na região de 670 a 1.225 cm^{-1} ; deformação angular (δ) de ligações C-H de grupamentos CH_3 e CH_2 em

1.377 e 1.460 cm^{-1} e estiramento (ν) de ligações C-H de grupamentos CH_2 em 2.853 e 2.922 cm^{-1} [52, 53], como mostrado na Figura 17.

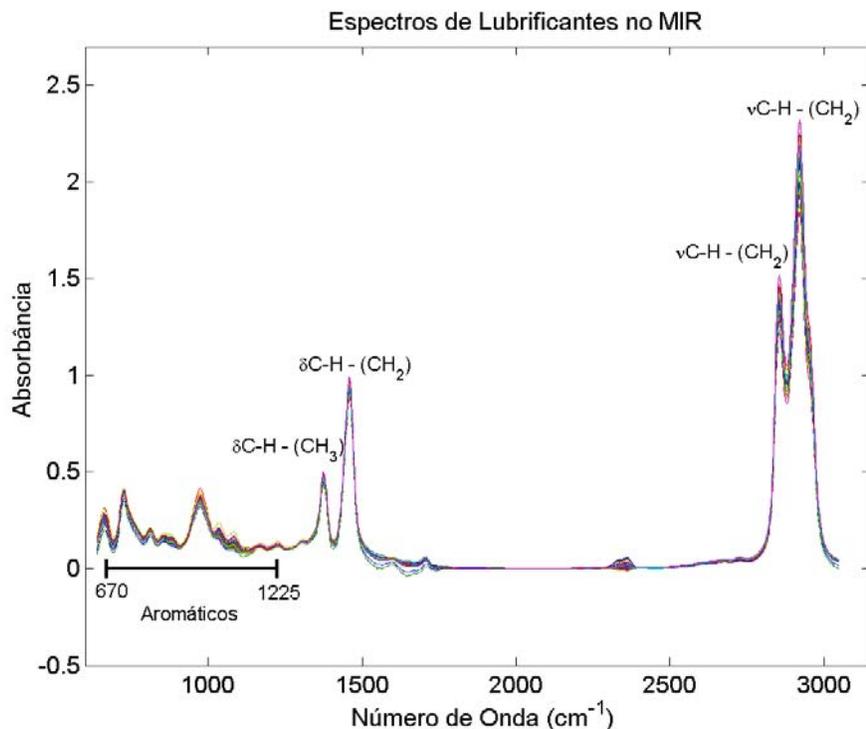


Figura 17. Representação de bandas espectrais características para espectros de lubrificantes na região do infravermelho médio.

3.3.2. Calibração Multivariada de lubrificante de motor de ônibus relacionado ao parâmetro de qualidade TBN

Na Tabela 2, estão apresentados os valores das médias do TBN das 45 amostras de lubrificantes de ônibus nas quilometragens de 1.936 a 45.896 Km usadas na construção dos modelos PLS e com seleção de variáveis. Em destaque cinza encontram-se as amostras selecionadas para o conjunto de previsão.

Os espectros na região do MIR das 45 amostras de lubrificante, relacionados com as medidas de TBN variando entre 3,55 e 8,44 mg KOH/g de amostra, após correção da linha base e selecionado a região de 677 a 3.124 cm^{-1} (1.270 variáveis), são apresentados na Figura 18.

Tabela 2. Resultados em ordem decrescente das análises de TBN (mg KOH/g) das 45 amostras de lubrificantes de ônibus.

Km rodada	Média TBN	Km rodada	Média TBN	Km rodada	Média TBN
32.003	8,44	6.424	6,20	10.162	5,49
16.593	8,16	10.728	6,17	33.565	5,27
1.936	7,96	28.206	6,11	20.781	5,20
17.460	7,88	30.833	6,09	29.545	5,20
2.790	7,52	14.276	6,00	15.948	5,18
25.000	7,21	25.798	5,93	31.219	5,17
20.739	7,16	24.599	5,86	15.296	4,78
9.526	7,14	26.490	5,86	27.909	4,70
5.583	7,12	16.312	5,85	34.000	4,59
7.402	7,03	20.097	5,82	14.380	4,50
2.427	6,96	26.722	5,81	17.914	4,49
14.627	6,57	18.419	5,80	30.564	4,32
28.179	6,53	4.476	5,79	7.748	4,06
9.282	6,27	24.674	5,69	19.298	3,87
5.445	6,25	22.780	5,50	45.896	3,55

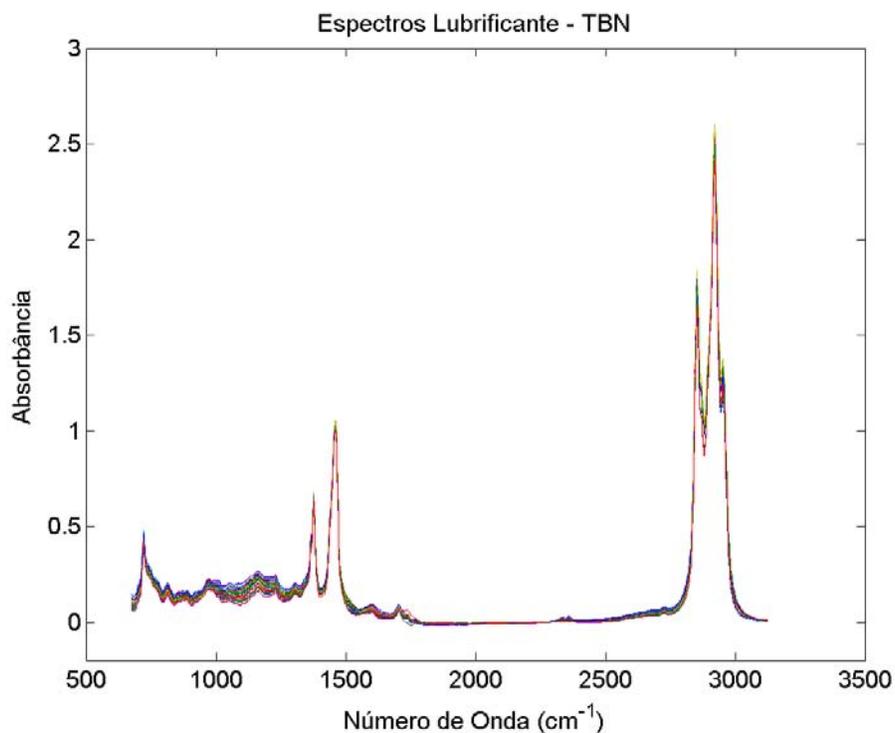


Figura 18. Espectros das 45 amostras de lubrificante de ônibus de 1.936 a 45.896 Km.

Para a construção dos modelos de calibração utilizaram-se 38 amostras e 7 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.270 variáveis do espectro), iPLS (53 variáveis – 779 a 880 cm^{-1}), UVE-PLS (82 variáveis), GA-PLS (187 variáveis), GA-iPLS (11 variáveis) e GA-UVE-PLS (10 variáveis). Os resultados dos modelos avaliados quanto ao número de variáveis latentes (VL), erro quadrático médio de previsão (RMSEP) e média do erro relativo (EM) são apresentados na Tabela 3.

Tabela 3. Resultados para os modelos de calibração multivariada de lubrificante - TBN.

Parâmetro	PLS (1.270 var.)			iPLS (53 var.)			UVE-PLS (82 var.)		
	VL	RMSEP (mgKOH/g)	EM (%)	VL	RMSEP (mgKOH/g)	EM (%)	VL	RMSEP (mgKOH/g)	EM (%)
TBN	7	$8,70 \times 10^{-2}$	1,25	7	$3,54 \times 10^{-1}$	4,67	7	$2,68 \times 10^{-1}$	3,86
	677 – 3.124 cm^{-1}			779 - 880 cm^{-1}			----		
	GA-PLS (187 var.)			GA-iPLS (11 var.)			GA-UVE-PLS (10 var.)		
	VL	RMSEP (mgKOH/g)	EM (%)	VL	RMSEP (mgKOH/g)	EM (%)	VL	RMSEP (mgKOH/g)	EM (%)
	7	$1,00 \times 10^{-1}$	1,28	7	$3,22 \times 10^{-1}$	4,06	7	$3,07 \times 10^{-1}$	4,16
	50 gerações	convergência 50%		25 gerações	convergência 80%		20 gerações	convergência 80%	

Para a calibração multivariada de lubrificante de motor de ônibus relacionado ao parâmetro TBN o melhor modelo foi o PLS (destaque cinza), pois apresentou valores de erros RMSEP e EM menores que os outros modelos, mostrando que nesse caso os métodos de seleção de variáveis não melhoraram a calibração, sendo necessário então usar a informação contida em toda a região do espectro (1.270 variáveis).

O modelo GA-PLS apresentou resultados semelhantes ao PLS, mas a única vantagem foi a redução do número de variáveis selecionadas para o modelo de calibração. Apenas este fato não viabiliza sua aplicação, uma vez que as variáveis selecionadas estão distribuídas ao longo de toda região espectral como

mostra a Figura 19, não melhorando a qualidade da informação química, ou seja, não foi encontrada uma região específica do espectro com informação mais importante que as demais. Como o método de GA demanda um tempo maior de processamento que o PLS e devido à falta de reprodutibilidade pela natureza probabilística, a melhor opção continua sendo o modelo PLS.

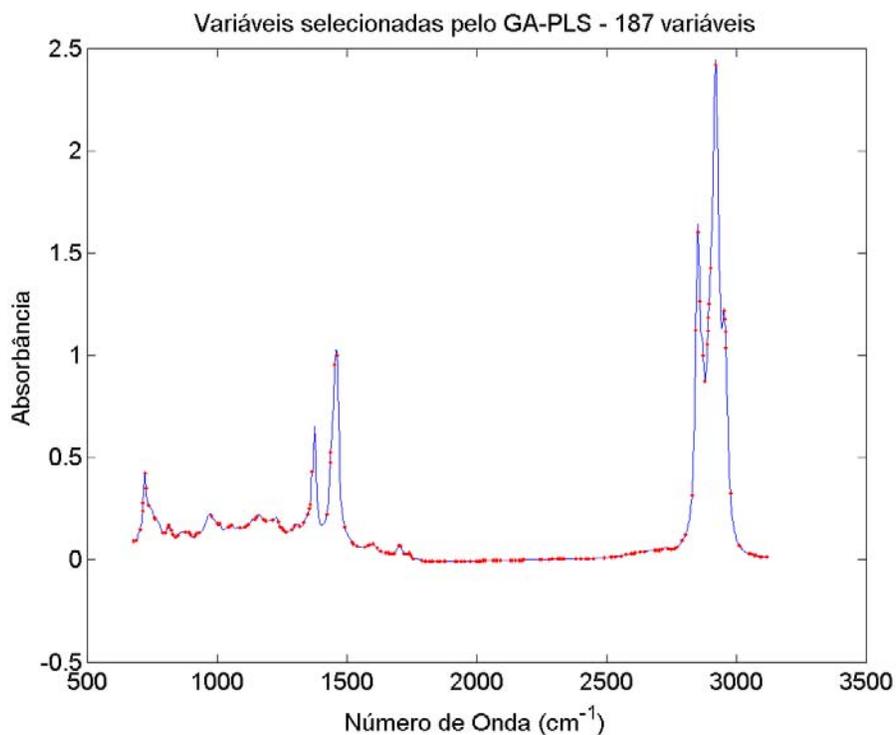


Figura 19. Espectros das 45 amostras de lubrificante de ônibus com as variáveis selecionadas pelo GA-PLS para calibração relacionada ao TBN, em vermelho.

Embora todos os modelos obtidos relacionando os espectros do lubrificante com as medidas de TBN foram eficientes, apresentando valores de RMSEP e EM relativamente baixos, de no máximo $3,54 \times 10^{-1}$ mg KOH/g e 4,67% na previsão das amostras de validação para o modelo iPLS, não são alternativas viáveis ao modelo PLS com 7 VLs, RMSEP de $8,70 \times 10^{-2}$ mg KOH/g e EM de 1,25%.

No decorrer deste e no próximo capítulo, serão discutidos os resultados de cada modelo para os parâmetros estudados, mas optaremos por apresentar os gráficos e figuras apenas dos resultados dos melhores modelos de calibração obtidos.

3.3.3. Calibração Multivariada de lubrificante de motor de ônibus relacionado ao parâmetro de qualidade Viscosidade a 100°C

Na Tabela 4, estão apresentados os valores das médias de viscosidade das 47 amostras de lubrificantes de ônibus nas quilometragens de 522 a 45.896 Km, usadas na construção dos modelos PLS e com seleção de variáveis. Em destaque cinza encontram-se as amostras selecionadas para o conjunto de previsão.

Tabela 4. Resultados em ordem crescente das análises de Viscosidade a 100°C (cSt) das 47 amostras de lubrificantes de ônibus.

Km rodada	Média Viscosidade	Km rodada	Média Viscosidade	Km rodada	Média Viscosidade
6.780	13,28	14.276	13,82	33.565	14,94
522	13,30	21.391	13,85	22.780	15,16
6.424	13,31	26.490	13,88	34.000	15,25
4.476	13,33	9.282	13,92	31.219	15,26
10.162	13,34	17.914	13,95	20.097	15,29
18.419	13,52	27.909	14,02	15.948	15,32
20.739	13,53	29.545	14,02	10.728	15,43
14.380	13,54	24.674	14,05	7.748	15,48
28.206	13,54	1.936	14,09	20.781	15,55
14.627	13,58	45.896	14,26	17.460	15,55
15.296	13,64	24.599	14,29	16.593	15,71
25.798	13,65	5.583	14,30	26.722	15,76
30.564	13,67	7.402	14,35	16.312	16,27
28.179	13,69	30.833	14,54	32.003	16,38
9.526	13,77	2.790	14,54	25.000	17,03
19.298	13,81	5.445	14,63		

Os espectros na região do MIR das 47 amostras de lubrificante, relacionados com as medidas de Viscosidade variando entre 13,28 e 17,03 cSt, após correção da linha base e selecionado a região de 677 a 3.124 cm^{-1} (1.270 variáveis), são apresentados na Figura 20.

Na construção dos modelos de calibração utilizaram-se 38 amostras e 9 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.270 variáveis do espectro), iPLS (53 variáveis - 779 a 880 cm^{-1}), UVE-PLS

(189 variáveis), GA-PLS (138 variáveis), GA-iPLS (15 variáveis) e GA-UVE-PLS (17 variáveis). A Tabela 5 apresenta os resultados dos modelos avaliados.

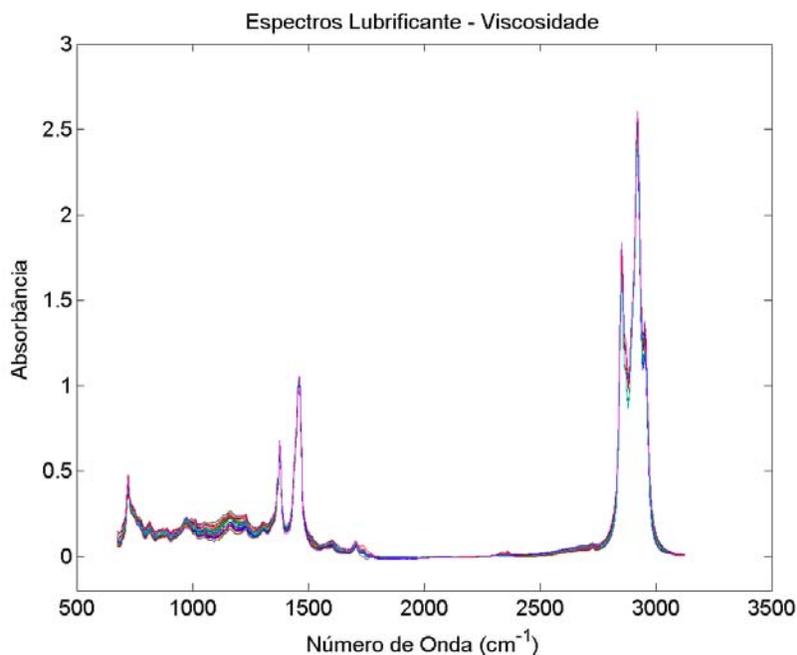


Figura 20. Espectros das 47 amostras de lubrificante de ônibus de 522 a 45.896 Km.

Tabela 5. Resultados para os modelos de calibração multivariada de lubrificante - Viscosidade.

Parâmetro	PLS (1.270 var.)			iPLS (53 var.)			UVE-PLS (189 var.)		
	VL	RMSEP (cSt)	EM (%)	VL	RMSEP (cSt)	EM (%)	VL	RMSEP (cSt)	EM (%)
Viscosidade 100°C	3	0,1479	0,92	5	0,1414	0,78	5	0,1559	0,87
	677 – 3.124 cm ⁻¹			779 - 880 cm ⁻¹			----		
	GA-PLS (138 var.)			GA-iPLS (15 var.)			GA-UVE-PLS (17 var.)		
	VL	RMSEP (cSt)	EM (%)	VL	RMSEP (cSt)	EM (%)	VL	RMSEP (cSt)	EM (%)
	3	0,0964	0,60	5	0,1318	0,72	2	0,1340	0,79
	100 gerações	converg. 80%		25 gerações	converg. 80%		50 gerações	convergência 80%	

Todos os modelos obtidos relacionando os espectros do lubrificante com as medidas de Viscosidade apresentaram valores de erros bem satisfatórios e com pouca diferença na eficiência da previsão das amostras de validação. Entretanto, o modelo GA-PLS (destaque cinza) com 3 VLs, RMSEP de 0,0964 cSt e EM de 0,60%, apresentou resultados melhores, sendo mais robusto (VL menor) que os modelos iPLS, GA-iPLS e UVE-PLS e prevendo melhor (RMSEP e EM menores) que o PLS e GA-UVE-PLS.

Na Figura 21 é apresentado o espectro de uma amostra de lubrificante de ônibus, destacando em vermelho as 138 variáveis selecionadas pelo modelo GA-PLS, distribuídas por todo espectro sem priorizar nenhuma região específica.

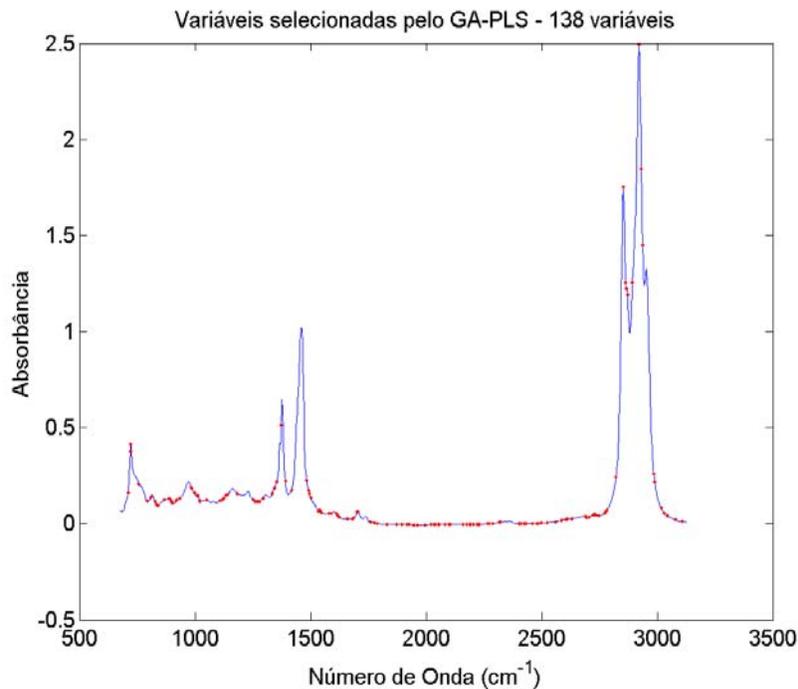


Figura 21. Espectro de lubrificante de ônibus com as variáveis selecionadas pelo GA-PLS para calibração relacionada à viscosidade, em vermelho.

Semelhante ao modelo GA-PLS do TBN, no modelo da Viscosidade não temos um melhora na informação química, mas nesse caso, a eficiência do modelo GA-PLS frente os demais viabiliza o seu uso.

Assim, para a calibração multivariada de lubrificante de motor de ônibus relacionado ao parâmetro Viscosidade, o modelo GA-PLS usando apenas 138

variáveis, selecionadas após 100 gerações e convergência de 80%, ao invés das 1.270 variáveis do modelo PLS mostrou que a seleção de variáveis por GA melhorou a calibração, sendo necessário então usar a informação contida em apenas 138 variáveis.

3.3.4. Calibração Multivariada de lubrificante contaminado com Diesel

Os espectros na região do MIR das 40 amostras de lubrificante contaminadas com diesel nas concentrações entre 0,20 e 8,15% (m/m), após correção da linha base e selecionado a região de 638 a 3.047 cm^{-1} (1.250 variáveis), são apresentados na Figura 22.

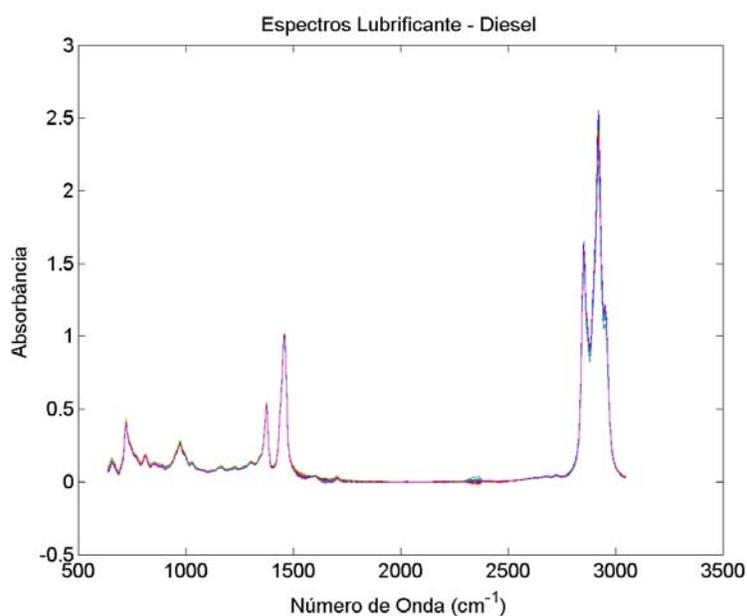


Figura 22. Espectros das 40 amostras de lubrificante contaminadas com diesel.

Na construção dos modelos de calibração utilizaram-se 30 amostras e 10 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.250 variáveis do espectro), iPLS (139 variáveis - 638 a 905 cm^{-1}), UVE-PLS (511 variáveis), GA-PLS (198 variáveis), GA-iPLS (18 variáveis) e GA-UVE-PLS (83 variáveis). A Tabela 6 apresenta os resultados dos modelos avaliados.

Tabela 6. Resultados para os modelos de calibração multivariada de lubrificante - diesel.

Parâmetro	PLS (1.250 var.)			iPLS (139 var.)			UVE-PLS (511 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Diesel</i>	10	$1,7 \times 10^{-1}$	2,60	11	$1,3 \times 10^{-1}$	3,17	10	$2,2 \times 10^{-1}$	4,44
	638 – 3.047 cm^{-1}			638 - 905 cm^{-1}			----		
	GA-PLS (198 var.)			GA-iPLS (18 var.)			GA-UVE-PLS (83 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	12	$1,5 \times 10^{-1}$	3,40	5	$1,0 \times 10^{-1}$	1,13	11	$1,2 \times 10^{-1}$	2,74
	100 gerações	convergência 50%		25 gerações	converg. 80%		50 gerações	convergência 50%	

Todos os modelos obtidos apresentaram valores de erros bem satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação. Nesse sentido, o modelo GA-iPLS (destaque cinza) com 5 VLs, RMSEP de $1,0 \times 10^{-1}\%$ e EM de 1,13%, apresentou os melhores resultados, sendo mais robusto e eficiente que os demais.

Na Figura 23 é apresentado o espectro de uma amostra de lubrificante contaminado com diesel, destacando em vermelho as 18 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a compostos aromáticos presentes no diesel.

Para a calibração multivariada de lubrificante contaminado com diesel o modelo GA-iPLS, mostrou que a seleção de variáveis por GA, usando como entrada as variáveis selecionadas pelo iPLS, 25 gerações e convergência de 80%, melhorou a calibração sendo necessário então usar a informação contida em apenas 18 variáveis na região relacionada a compostos aromáticos presentes no diesel. Isso torna possível uma análise realizada com o monitoramento dessas 18 variáveis ao invés das 1.250 variáveis do espectro global, pois as demais variáveis não são necessárias para o desenvolvimento do modelo de calibração.

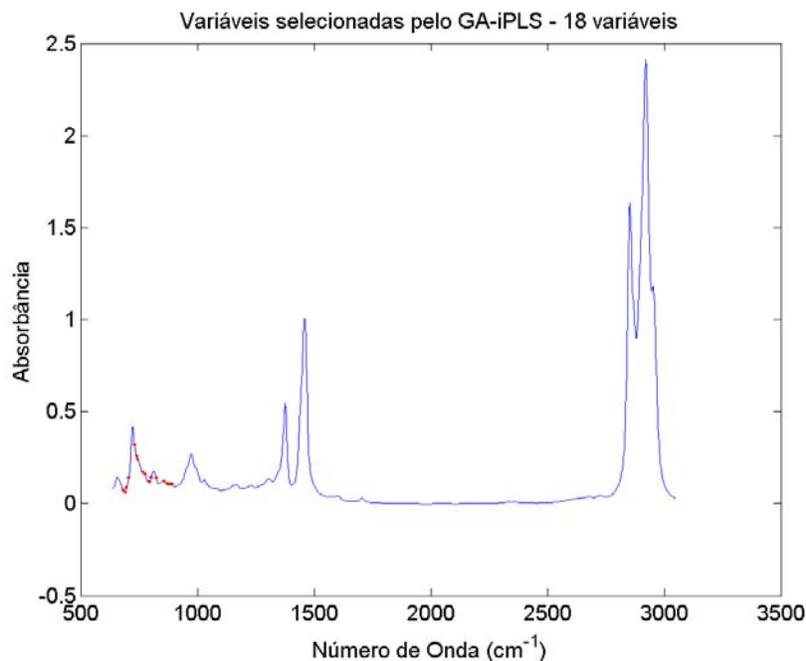


Figura 23. Espectro de lubrificante contaminado com diesel com as variáveis selecionadas pelo GA-iPLS em vermelho.

Neste caso, é importante destacar que além da vantagem de reduzir o número de variáveis, o modelo GA-iPLS possibilita uma melhor interpretação química da relação região espectral e a contaminação. Isso evidencia um dos principais objetivos da seleção de variáveis.

3.3.5. Calibração Multivariada de lubrificante contaminado com Gasolina

Os espectros na região do MIR das 25 amostras de lubrificante contaminadas com gasolina nas concentrações entre 0,32 e 8,03% (m/m), após correção da linha base e selecionado a região de 685 a 3.059 cm^{-1} (1.232 variáveis), são apresentados na Figura 24.

Para a construção dos modelos de calibração utilizaram-se 20 amostras e 5 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.232 variáveis do espectro), iPLS (83 variáveis - 845 a 1.003 cm^{-1}), UVE-PLS

(112 variáveis), GA-PLS (216 variáveis), GA-iPLS (14 variáveis) e GA-UVE-PLS (17 variáveis). A Tabela 7 apresenta os resultados dos modelos avaliados.

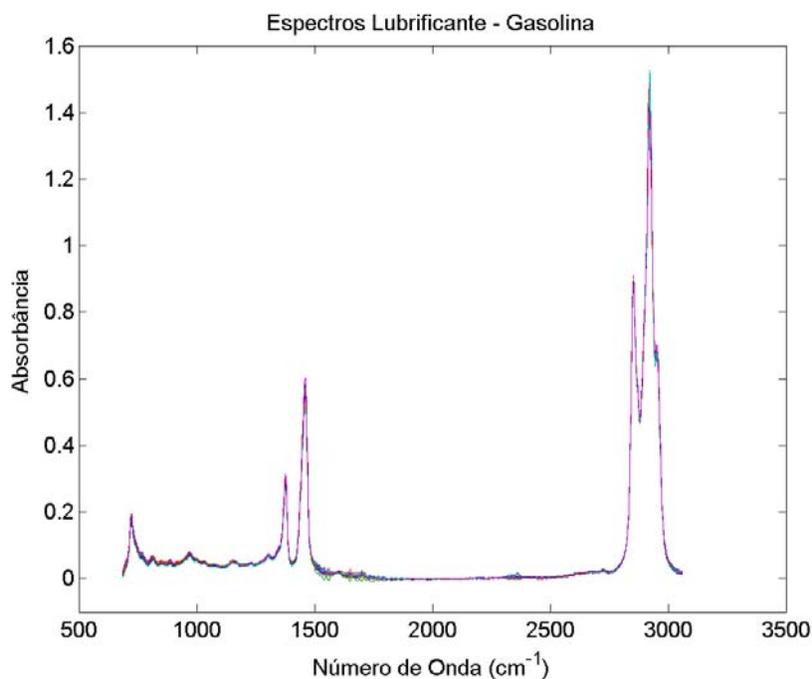


Figura 24. Espectros das 25 amostras de lubrificante contaminadas com gasolina.

Tabela 7. Resultados para os modelos de calibração multivariada de lubrificante - gasolina.

Parâmetro	PLS (1.232 var.)			iPLS (83 var.)			UVE-PLS (112 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
Gasolina	6	$1,2 \times 10^{-1}$	2,91	3	$1,5 \times 10^{-1}$	2,58	4	$2,6 \times 10^{-1}$	7,26
	685 – 3.059 cm^{-1}			845 – 1.003 cm^{-1}			----		
	GA-PLS (216 var.)			GA-iPLS (14 var.)			GA-UVE-PLS (17 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	4	$1,1 \times 10^{-1}$	2,88	3	$8,7 \times 10^{-2}$	2,03	4	$2,1 \times 10^{-1}$	4,33
	100 gerações	convergência 50%		20 gerações	converg. 80%		25 gerações	convergência 80%	

Todos os modelos obtidos apresentaram valores de erros satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação, principalmente para o modelo UVE-PLS. O modelo GA-iPLS (destaque cinza) com 3 VLs, RMSEP de $8,7 \times 10^{-2}\%$ e EM de 2,03%, apresentou os melhores resultados, sendo mais robusto e eficiente que os demais.

Na Figura 25 é apresentado o espectro de uma amostra de lubrificante contaminado com gasolina, destacando em vermelho as 14 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a compostos aromáticos presentes na gasolina.

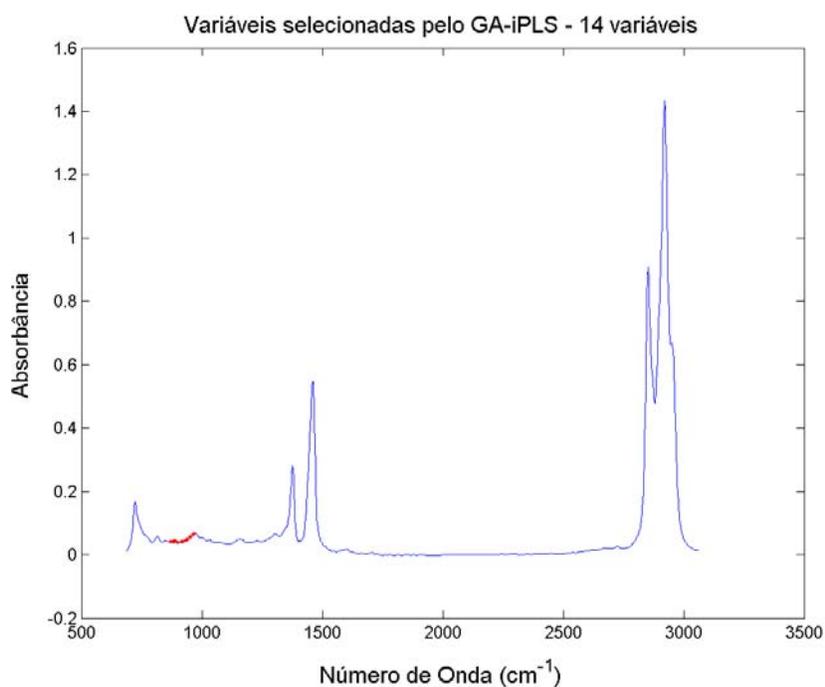


Figura 25. Espectro de lubrificante contaminado com gasolina com as variáveis selecionadas pelo GA-iPLS em vermelho.

Baseando nos critérios de robustez e eficiência, como no caso do diesel, para a calibração multivariada de lubrificante contaminado com gasolina o melhor modelo foi o GA-iPLS, com 20 gerações e convergência de 80%. Sendo necessário então usar a informação contida em apenas 14 variáveis na região relacionada a compostos aromáticos presentes na gasolina. Tornando possível a

análise realizada com o monitoramento dessas 14 variáveis ao invés das 1.232 variáveis do espectro global, visto que as demais variáveis não são fundamentais para o desenvolvimento do modelo de calibração.

Além da vantagem da redução de variáveis, a maior qualidade na informação química se fez presente no modelo GA-iPLS.

3.3.6. Calibração Multivariada de lubrificante contaminado com Água

Os espectros na região do MIR das 21 amostras de lubrificante contaminadas com água nas concentrações entre 0,04 e 0,69% (m/m), após correção da linha base e selecionado a região de 650 a 3.888 cm^{-1} (1.680 variáveis), são apresentados na Figura 26. Além das bandas características dos espectros de lubrificante citadas no item 3.3.1, também tem uma banda larga na região de 3.200 a 3.700 cm^{-1} relacionada a estiramento de ligações O-H.

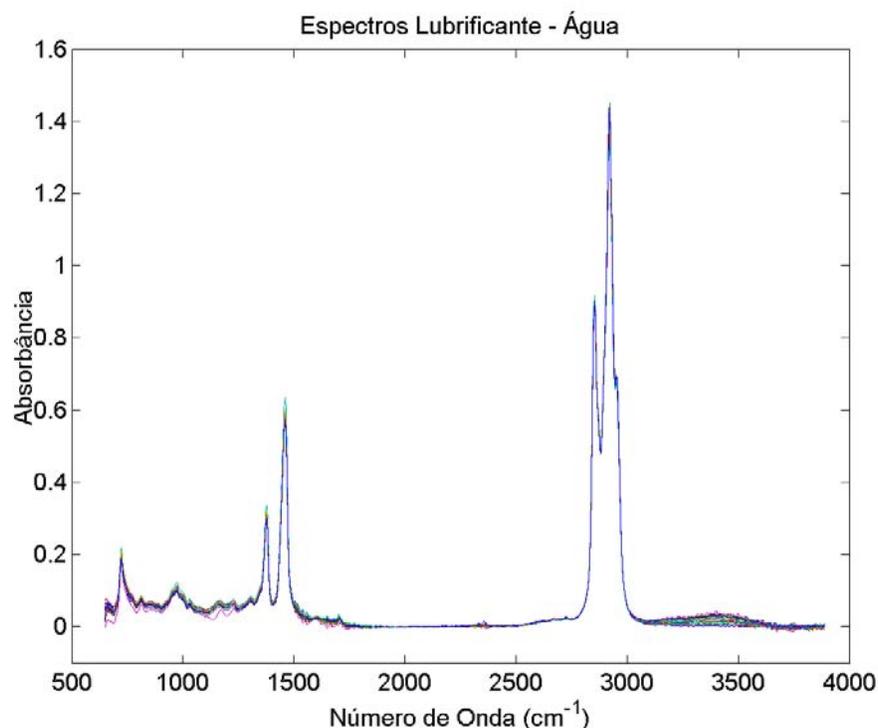


Figura 26. Espectros das 21 amostras de lubrificante contaminadas com água.

Para a construção dos modelos de calibração utilizaram-se 16 amostras e 5 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.680 variáveis do espectro), iPLS (210 variáveis – 3.485 a 3.888 cm^{-1}), UVE-PLS (78 variáveis), GA-PLS (244 variáveis), GA-iPLS (21 variáveis) e GA-UVE-PLS (16 variáveis). A Tabela 8 apresenta os resultados dos modelos avaliados.

Tabela 8. Resultados para os modelos de calibração multivariada de lubrificante - água.

Parâmetro	PLS (1.680 var.)			iPLS (210 var.)			UVE-PLS (78 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
Água	6	$8,7 \times 10^{-3}$	1,74	3	$2,5 \times 10^{-2}$	5,11	3	$3,3 \times 10^{-2}$	8,46
	650 – 3.888 cm^{-1}			3.485 – 3.888 cm^{-1}			----		
	GA-PLS (244 var.)			GA-iPLS (21 var.)			GA-UVE-PLS (16 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	5	$1,9 \times 10^{-2}$	3,77	3	$2,1 \times 10^{-2}$	3,21	2	$3,0 \times 10^{-2}$	6,12
	100 gerações	convergência 50%		25 gerações	convergência 50%		25 gerações	convergência 80%	

Os modelos PLS, GA-PLS, iPLS e GA-iPLS obtidos para a calibração do lubrificante contaminado com água apresentaram valores de erros satisfatórios, enquanto os modelos UVE-PLS e GA-UVE-PLS apresentaram problemas quanto à previsão das amostras de validação. Com maior robustez e eficiência, o modelo PLS (destaque cinza) com 6 VLs, RMSEP de $8,7 \times 10^{-3}\%$ e EM de 1,74% mostrou que neste caso a seleção de variáveis é inviável, sendo necessário utilizar a informação contida em toda a região do espectro (1.680 variáveis).

Geralmente, seria esperado que os métodos de seleção de variáveis selecionassem a região relacionada ao grupo O-H, o que de fato não ocorreu. Como na homogeneização foi necessário adicionar um detergente, é possível que aconteça uma interferência causada por uma interação entre o detergente e a água. Por se tratar de um reagente cedido pela empresa sem a informação de sua composição, não podemos explicar de forma adequada este fato.

O detalhe em questão não afeta a eficiência ou escolha do modelo PLS para a quantificação da contaminação de lubrificante por água.

3.3.7. Calibração Multivariada de lubrificante contaminado com Etileno Glicol

Os espectros na região do MIR das 40 amostras de lubrificante contaminadas com etileno glicol nas concentrações entre 0,05 e 2,05% (m/m), após correção da linha base e selecionado a região de 638 a 3.047 cm^{-1} (1.250 variáveis), são apresentados na Figura 27.

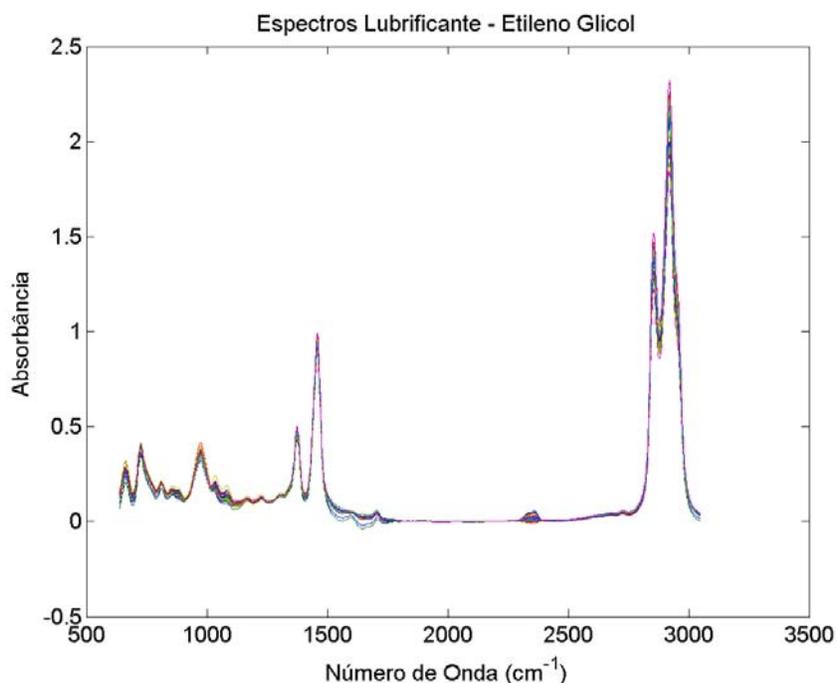


Figura 27. Espectros das 40 amostras de lubrificante contaminadas com etileno glicol.

Na construção dos modelos de calibração utilizaram-se 30 amostras e 10 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.250 variáveis do espectro), iPLS (139 variáveis – 1.175 a 1.441 cm^{-1}), UVE-PLS (451 variáveis), GA-PLS (129 variáveis), GA-iPLS (17 variáveis) e GA-UVE-PLS (62 variáveis). A Tabela 9 apresenta os resultados dos modelos avaliados.

Tabela 9. Resultados para os modelos de calibração multivariada de lubrificante - etileno.

Parâmetro	PLS (1.250 var.)			iPLS (139 var.)			UVE-PLS (451 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Etileno</i> <i>Glicol</i>	7	$4,9 \times 10^{-2}$	3,78	12	$1,1 \times 10^{-1}$	11,43	6	$5,5 \times 10^{-2}$	6,41
	638 – 3.047 cm^{-1}			1.175 – 1.441 cm^{-1}			----		
	GA-PLS (129 var.)			GA-iPLS (17 var.)			GA-UVE-PLS (62 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	5	$6,5 \times 10^{-2}$	5,84	8	$1,1 \times 10^{-1}$	8,24	6	$5,1 \times 10^{-2}$	5,28
	100 gerações	convergência 50%		25 gerações	converg. 80%		25 gerações	convergência 80%	

Os modelos obtidos para a calibração do lubrificante contaminado com etileno glicol apresentaram bons resultados para o PLS, GA-PLS, UVE-PLS e GA-UVE-PLS, sendo que o PLS (destaque cinza) com 7 VLs, RMSEP de $4,9 \times 10^{-2}\%$ e EM de 3,78%, apresentou os melhores resultados. Como no caso da contaminação com água não foi viável a seleção de variáveis.

3.4. Conclusões da calibração multivariada de lubrificantes

A maioria dos modelos de calibração multivariada por PLS e seleção de variáveis relacionando os espectros obtidos na região do infravermelho médio com parâmetros de qualidade de lubrificantes, apresentaram valores de erros satisfatórios, robustez e eficiência na previsão das amostras de validação.

De acordo com as comparações realizadas e resultados apresentados, constatou-se que o método de calibração por PLS foi mais robusto e eficiente na quantificação do parâmetro TBN e dos contaminantes água e etileno glicol. Nesses casos, a utilização dos métodos de seleção de variáveis não foi viável para selecionar regiões nos espectros com informações mais importantes que em outras, sendo necessário utilizar a informação contida em toda a região do espectro.

O método GA-PLS foi o mais eficiente na quantificação do parâmetro Viscosidade a 100°C, mostrando que o GA selecionou 138 dentre as 1.270 variáveis dos espectros, eliminando assim variáveis que não são necessárias para o desenvolvimento do modelo de calibração.

Na quantificação dos contaminantes diesel e gasolina o método GA-iPLS foi o mais eficiente. O GA utilizando como entrada as variáveis selecionadas pelo modelo iPLS selecionou 18 e 14 variáveis respectivamente, na região relacionada a compostos aromáticos presentes no diesel e gasolina como benzeno, tolueno, etilbenzeno, xileno e propilbenzeno, representando uma importante melhora na qualidade da informação química.

As metodologias de análise de lubrificantes novos ou usados por espectroscopia no infravermelho médio, calibração multivariada e seleção de variáveis para alguns casos de determinação dos parâmetros de qualidade, tanto de degradação quanto contaminações, consiste em uma alternativa viável as análises clássicas. Podendo ser aplicadas em laboratórios de controle de qualidade reduzindo tempo, custo de reagentes e quantidade de resíduos tóxicos, além de melhorar a qualidade da informação química.

CAPÍTULO 4

Óleos de Oliva

4. Óleos de Oliva

Os óleos de Oliva são obtidos a partir das frutas de diferentes variedades de oliveiras (*Olea europaea* L.), extensamente cultivadas nos países à margem do Mar Mediterrâneo, tais como, Grécia, Itália e Espanha, onde tiveram importante participação na dieta dos indivíduos nesta região, assim como na sua economia e cultura. Também são cultivadas em outros países como Tunísia, Turquia, Portugal, Síria, Marrocos, Argélia, Argentina e outros. A produção mundial de azeite de Oliva representa 4% da produção mundial de óleos vegetais e 2,5% de óleos comestíveis e gorduras [9].

Na produção de azeite as azeitonas são descaroçadas e amassadas até tornarem-se uma polpa espessa, em seguida uma prensa extrai o suco, que é centrifugado para separar a água do óleo. O azeite de melhor qualidade e preço é conhecido como **extra virgem** e é obtido da primeira prensagem a frio da azeitona, sem sofrer influência de calor, aditivo ou solvente, apresentando um grau de pureza maior e menor acidez. Seguindo a hierarquia, vem o azeite **virgem**, que é extraído da segunda ou terceira prensagem do fruto, por isso, perde um pouco no aroma e sabor, costuma ser mais ácido e levemente adocicado. O azeite **refinado** passa por um processo de neutralização, descoloração e desodorização, após ser extraído das outras prensas da azeitona. O azeite **puro** é a mistura do refinado com o virgem. Esses dois últimos têm preços mais baixos e sabor menos concentrado, sendo mais comercializados em termos gastronômicos, pois embora os quatro tipos sejam distintos têm valores nutritivos semelhantes. O azeite **lampante** é um azeite virgem, impróprio para o consumo humano por apresentar sabor e aroma indesejáveis, originário de azeitonas de má qualidade. O azeite **orujo** é obtido pela extração com solventes dos resíduos de prensagem das azeitonas. Tanto o azeite lampante como o orujo, devem ser refinados para tornarem-se próprios ao consumo humano [53].

Na composição química de óleos e gorduras encontra-se uma faixa de 85 a 95% de triacilglicerídeos. Desta forma, as propriedades dos óleos e gorduras vegetais estão diretamente relacionadas com a composição triacilglicerídica,

baseando-se nos ácidos graxos presentes e sua posição na cadeia. Considerada como a principal vantagem do óleo de Oliva, a predominância de gordura monoinsaturada (cerca de 83%), menos nociva a saúde, auxilia na diminuição dos níveis de colesterol LDL (o mau colesterol) sem afetar os níveis de HDL (o bom colesterol) ou de triacilglicerídeos. O HDL ajuda a desintoxicar as células do organismo e reduz a ocorrência de doenças coronárias. Além disso, o azeite é de fácil digestão e apresenta propriedades medicinais como: estimular o metabolismo, lubrifica as mucosas, hidrata e trata a pele, auxilia na redução da obesidade e conseqüentemente na incidência de diabetes, desobstrui ouvido entupido e atualmente está sendo investigado possível eficiência no tratamento de alguns tipos de câncer [54].

Nos anos 1980 cerca de 400 mortes e 20.000 casos de intoxicações ocorreram devido ao consumo de óleo adulterado (“síndrome da toxicidade de óleo Espanhólico”) [55-57].

Os óleos vegetais que competem com o Oliva, como os de Algodão, Soja, Milho e Girassol, tem predominância de gordura poliinsaturada, que em termos de efeitos à saúde, está numa posição intermediária.

No Brasil não há produção comercial de azeite sendo que a demanda interna é atendida por azeites importados de países da Europa e da Argentina. Infelizmente, por outro lado é um dos países com maior índice de contaminação de azeite [53].

Devido as suas características sensoriais excepcionais e a uma produção limitada, o azeite apresenta elevado valor de mercado, sendo alvo freqüente de adulterações. A adulteração mais comum é a adição de outros óleos vegetais de menor valor comercial. Entretanto, outras fraudes relativas à qualidade do azeite são praticadas devido as diferentes categorias de azeites existentes. Como por exemplo, a adição de Oliva puro ao extra virgem [53, 57, 58].

Nesse sentido, vários métodos físicos e químicos têm sido usados na determinação da autenticidade e detecção do nível de contaminação de óleo de Oliva. Na literatura, encontram-se trabalhos descrevendo métodos qualitativos e quantitativos para determinar a adulteração em óleo de Oliva, dentre eles:

Espectroscopia no UV-Vis [53, 59], Espectrometria de Massas (MS) [60], Cromatografia Gasosa (GC) [53, 61], Cromatografia Líquida de Alta Eficiência (HPLC) [55] e Espectroscopia de Ressonância Magnética Nuclear (NMR) [55]. A maioria desses métodos consome tempo de análise, reagentes, são destrutivos e geram resíduos.

A espectroscopia vibracional em combinação com métodos quimiométricos vem se mostrando nos últimos anos ser uma técnica capaz de extrair informação química de matrizes complexas como os óleos vegetais.

Até o início dos anos 1990, a espectroscopia vibracional era considerada muito limitada para análises de composição de óleos de Oliva, principalmente nas análises quantitativas, sendo utilizada apenas em poucas análises qualitativas. Após o desenvolvimento dos espectrômetros com transformada de Fourier teve início o interesse em análises quantitativas voltadas a detecção de contaminações e quantificação de ácidos graxos e foram retomadas as análises qualitativas. Em ambos os casos utilizando métodos quimiométricos de análises exploratórias e calibração multivariada. Desse período até hoje, podemos destacar dentro dos trabalhos de análise exploratória os trabalhos [9, 62-65], na determinação de adulteração em óleo de Oliva [66-71] e o primeiro trabalho da aplicação de espectroscopia FT-Raman para detecção de adulteração em óleo de Oliva (Baeten et al. – 1996) [8].

4.1. Objetivo

Esta etapa do trabalho tem como objetivo, a construção de modelos de calibração multivariada por PLS e seleção de variáveis, comparando-os quanto à habilidade de quantificar adulterações em óleos de Oliva através de espectroscopia vibracional.

4.2. Parte Experimental

4.2.1. Contaminação de óleo de Oliva

Foram realizadas as contaminações artificiais de 100 amostras de óleos de Oliva, sendo 50 de Oliva Puro de origem italiana e 50 de Extra Virgem de origem portuguesa. Nessa etapa, 10 amostras de óleo de Oliva Puro e 10 de Extra Virgem foram contaminadas com diferentes concentrações variando de 1,00 até 20,00% (m/m), de óleos vegetais de Algodão, Canola, Girassol, Milho e Soja, produzidos no Brasil, o que resultou 20 amostras para cada contaminação. Todos adquiridos no comércio da cidade de Campinas-SP.

4.2.2. Espectroscopia no infravermelho médio de óleo de Oliva

Os espectros das amostras de óleos de Oliva contaminadas foram obtidos, em triplicata, no espectrofotômetro Bomem – MB100 FTIR com acessório de ATR horizontal de cristal de ZnSe, na região do infravermelho médio abrangendo a região de 650 a 4.000 cm^{-1} , com uma resolução de 4 cm^{-1} e 128 varreduras.

Posteriormente, esses espectros foram tratados usando o programa baseline do PLS_Toolbox versão 2.0, como no caso dos espectros de lubrificantes no infravermelho médio.

Os espectros de óleos comestíveis na região do infravermelho médio apresentam bandas características, relacionadas à deformação de ligações C-H fora do plano (*cis* de dupla ligação) em 722 cm^{-1} ; deformação de ligações C-H fora do plano (*trans* de dupla ligação) em 966 cm^{-1} ; região de impressão digital de 900 a 1.500 cm^{-1} ; estiramento de ligações C=O de carbonila de aldeído em 1.730 cm^{-1} (produto de auto-oxidação); estiramento de ligações C=O de grupo éster de triacilglicerídeos em 1.743 cm^{-1} ; estiramento de ligações C-H de grupos metil terminal e metileno de triacilglicerídeos em 2.853 e 2.922 cm^{-1} e estiramento de ligações C-H (*cis* de dupla ligação) em 3.005 cm^{-1} [61]. Aqui representadas na Figura 28.

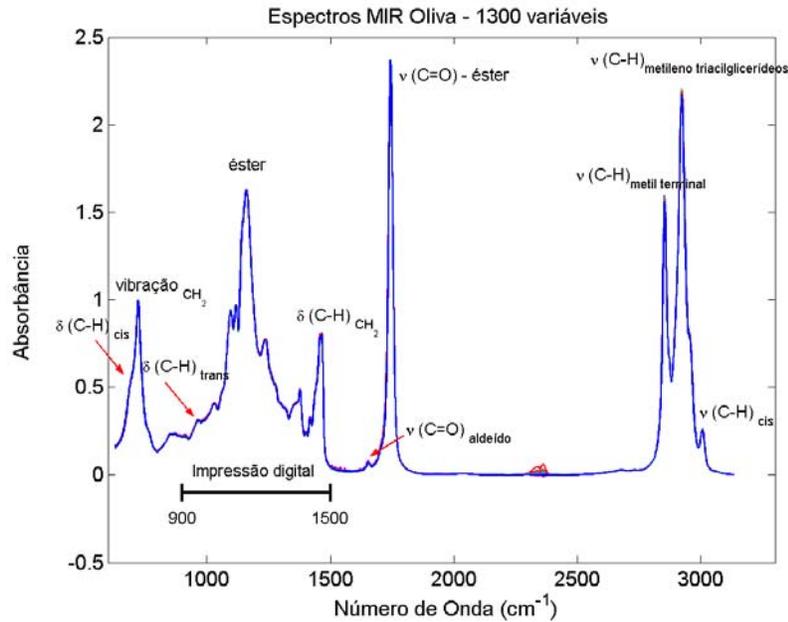


Figura 28. Representação de bandas espectrais características para espectros de óleo de Oliva na região do infravermelho médio.

4.2.3. Espectroscopia no infravermelho próximo de óleo de Oliva

Os espectros das amostras de óleo de Oliva contaminadas foram obtidos, em triplicata, no espectrofotômetro Thermo Nicolet - Nexus 670 FTIR com purga de gás (ar comprimido, 30 psi) da Parker, na região do infravermelho próximo, que abrange a região de 4.400 a 11.000 cm^{-1} . Os espectros foram obtidos com uma resolução de 4 cm^{-1} , 64 varreduras e usando uma cubeta de quartzo com 1 mm de caminho ótico.

Os espectros obtidos na região do infravermelho próximo foram tratados usando os programas baseline e savgol (15 pontos, ordem 2 e sem derivada) do PLS_Toolbox versão 2.0, para filtrar ruídos nos espectros e realizar correção na sua linha base.

Os espectros de óleos comestíveis na região do infravermelho próximo apresentam bandas características, relacionadas a estiramento de ligações C-H (combinação) em 4.661 cm^{-1} ; estiramento de ligações C-H de grupos CH_2 (1ª overtone) em 5.678 e 5.791 cm^{-1} ; estiramento de ligações C-H de grupos CH_2

(combinação) em 7.076 e 7.186 cm^{-1} e estiramento de ligações C-H de grupos CH_2 (2^{a} overtone) em 8.260 e 8.555 cm^{-1} [61]. Aqui representadas na Figura 29.

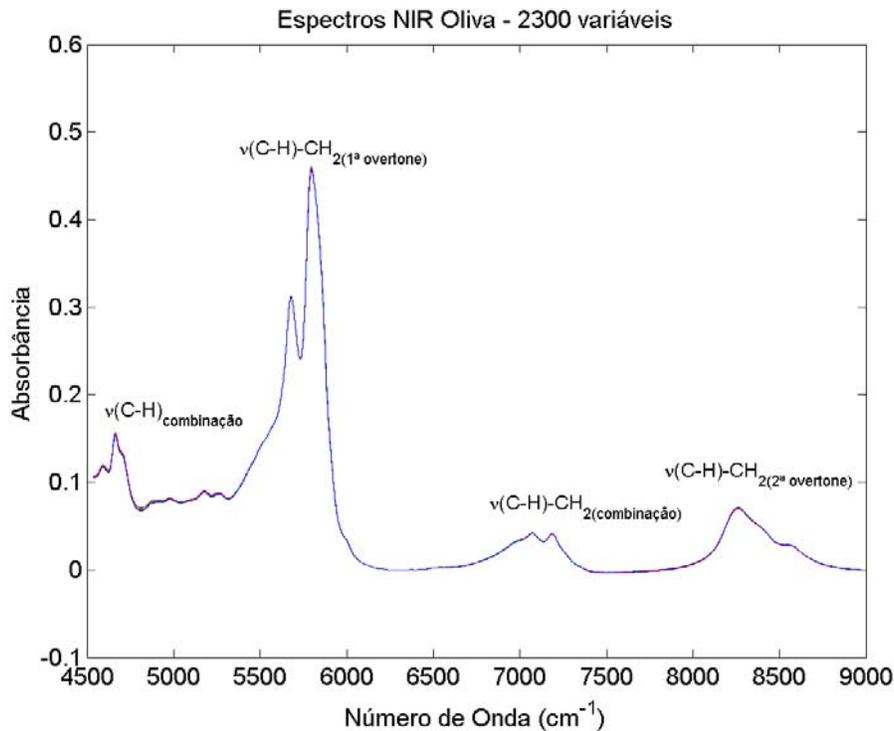


Figura 29. Representação de bandas espectrais características para espectros de óleo de Oliva na região do infravermelho próximo.

4.2.4. Espectroscopia Raman de óleo de Oliva

Os espectros das amostras de óleo de Oliva contaminadas foram obtidos, em triplicata, num espectrofotômetro Raman de baixa resolução – R2001, com LASER de diodo operando na região do infravermelho próximo (785 nm) e detector do tipo CCD de 2.048 elementos da marca Raman System. O espalhamento Raman abrangendo a região de 200 a 2.800 cm^{-1} . Os espectros foram obtidos com uma resolução de 15 cm^{-1} e usando uma cubeta de quartzo com 10 mm de caminho ótico.

Todos os espectros obtidos foram corrigidos usando um programa para Matlab 6.1 de ortogonalização baseado numa modificação do método de Gram-Schmidt. Os espectros originais são apresentados na Figura 30a e os corrigidos e selecionado a região de 683 a 1.572 cm^{-1} na Figura 30b.

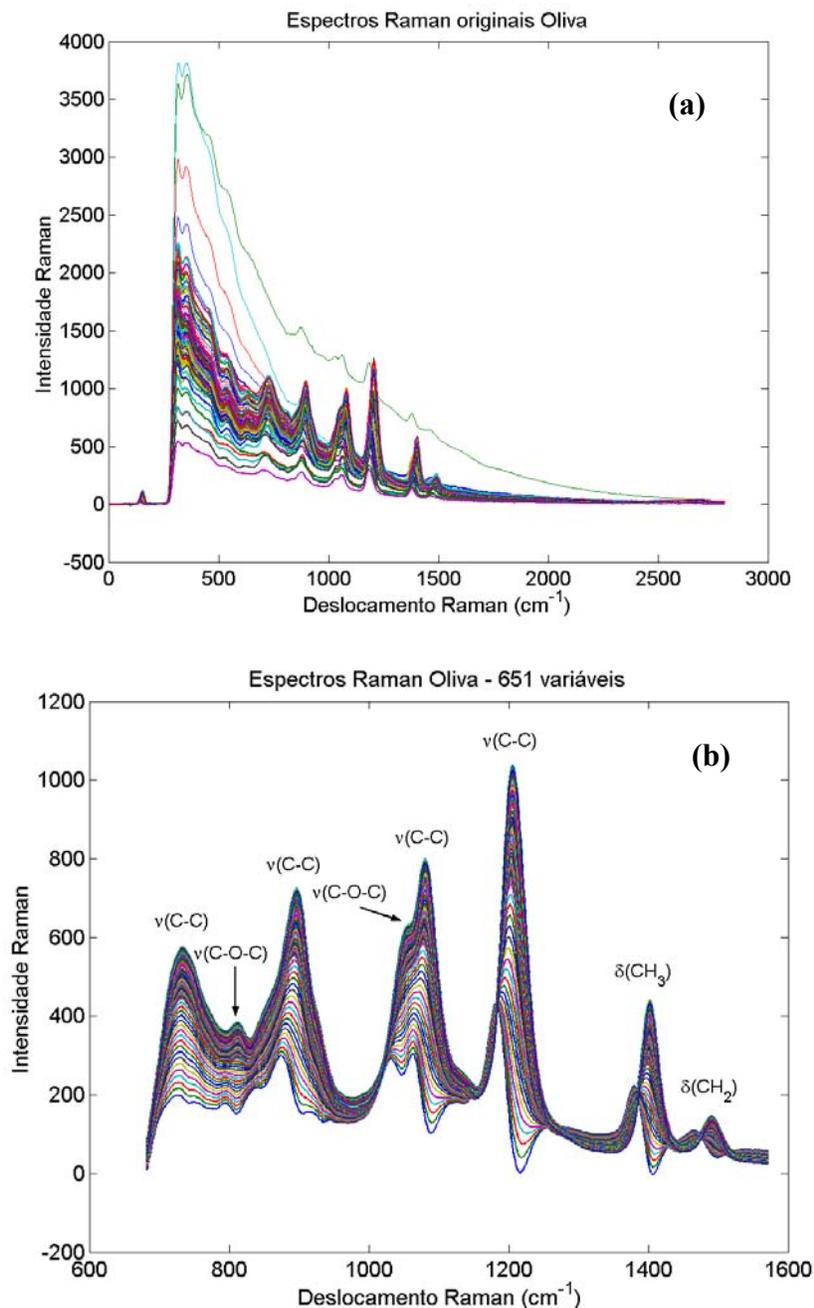


Figura 30. (a) Espectros Raman originais das amostras de óleo de Oliva. (b) Espectros Raman corrigidos pelo método de Gram-Schmidt e a representação de suas bandas espectrais.

Os espectros Raman das amostras de Oliva apresentaram bandas relacionadas a estiramento de ligações C-C em 732, 893, 1.079 e 1.202 cm⁻¹; estiramento de ligações C-O-C em 814 e 1.056 cm⁻¹ e deformação de ligações C-H de grupos CH₃ e CH₂ em 1.399 e 1.486 cm⁻¹.

4.3. Resultados e discussão

4.3.1. Calibração Multivariada de óleo de Oliva contaminado com óleo de Algodão

Os espectros na região do infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Algodão nas concentrações entre 1,25 e 15,00% (m/m), após correção da linha base e selecionado a região de 629 a 3.134 cm^{-1} (1.300 variáveis), são apresentados na Figura 31.

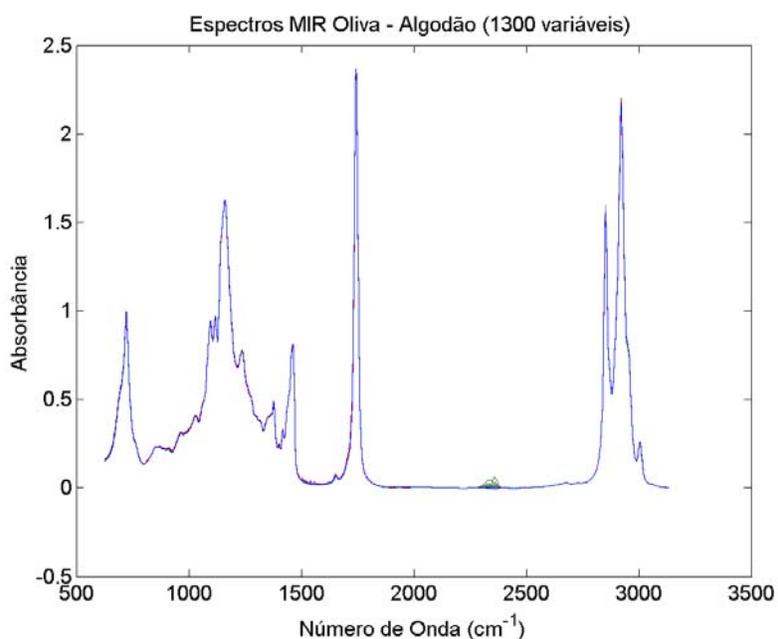


Figura 31. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Algodão.

Na construção dos modelos de calibração utilizaram-se 16 amostras e 4 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.300 variáveis do espectro), iPLS (109 variáveis - 839 a 1.047 cm^{-1}), UVE-PLS (287 variáveis), GA-PLS (270 variáveis), GA-iPLS (17 variáveis) e GA-UVE-PLS (48 variáveis). A Tabela 10 apresenta os resultados dos modelos avaliados.

Tabela 10. Resultados para os modelos de calibração multivariada de Oliva – Algodão no infravermelho médio.

Parâmetro	PLS (1.300 var.)			iPLS (109 var.)			UVE-PLS (287 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
Algodão	10	$1,12 \times 10^{-1}$	1,74	5	$6,36 \times 10^{-2}$	0,52	10	$6,71 \times 10^{-2}$	0,58
	629 – 3.134 cm^{-1}			839 – 1.047 cm^{-1}			----		
MIR	GA-PLS (270 var.)			GA-iPLS (17 var.)			GA-UVE-PLS (48 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	8	$4,00 \times 10^{-2}$	0,40	5	$1,24 \times 10^{-2}$	0,11	7	$4,24 \times 10^{-2}$	0,35
	100 gerações	convergência 80%		25 gerações	convergência 80%		50 gerações	convergência 80%	

Todos os modelos obtidos apresentaram valores de erros bem satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação. Nesse sentido, o modelo GA-iPLS (destaque cinza) com 5 VLs, RMSEP de $1,24 \times 10^{-2}\%$ e EM de 0,11%, apresentou os melhores resultados, sendo mais robusto e eficiente que os demais.

Na Figura 32 é apresentado o espectro no infravermelho médio de uma amostra, destacando em vermelho as 17 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada à deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados).

Para a calibração multivariada de Oliva contaminado com Algodão no MIR, o modelo GA-iPLS mostrou que a seleção de variáveis por GA, usando como entrada as variáveis selecionadas pelo iPLS, 25 gerações e convergência de 80%, melhorou a calibração sendo necessário então usar a informação contida em apenas 17 variáveis na região relacionada à deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados). Isso torna possível uma análise realizada com o monitoramento dessas 17 variáveis ao invés das 1.300 variáveis do espectro global, pois as demais variáveis não são necessárias para o desenvolvimento do modelo de calibração. O modelo GA-iPLS possibilitou uma melhor interpretação química da relação região espectral e contaminação.

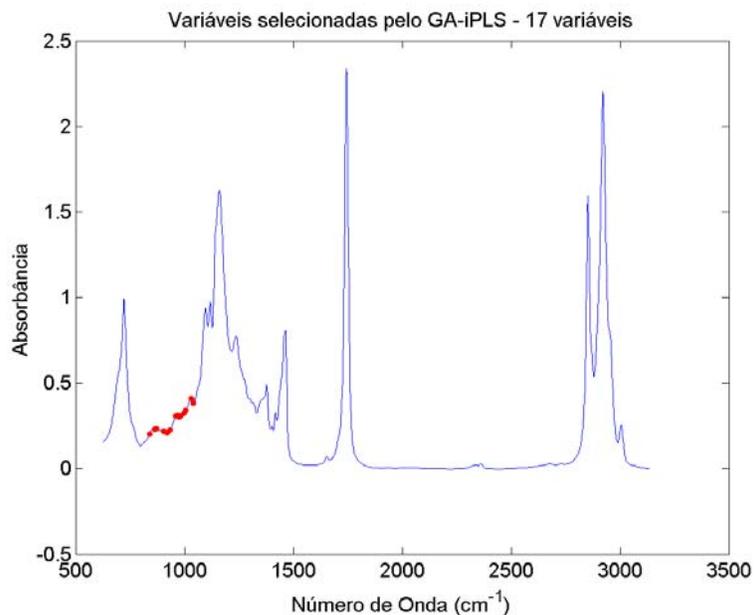


Figura 32. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Algodão com as variáveis selecionadas pelo GA-iPLS em vermelho.

Os espectros na região do infravermelho próximo das mesmas amostras de Oliva contaminadas, após filtrar ruídos nos espectros, corrigir a linha base e selecionado a região de 4.534 a 8.968 cm^{-1} (2.300 variáveis), são apresentados na Figura 33.

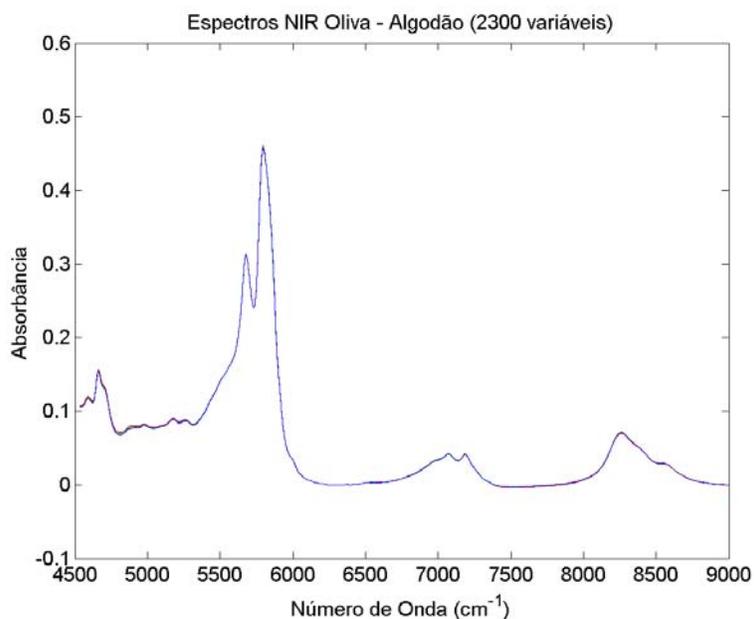


Figura 33. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Algodão.

Na construção dos modelos de calibração e conjunto de previsão foram utilizadas as mesmas amostras dos modelos no infravermelho médio. Em seguida, construídos os modelos: PLS (todas as 2.300 variáveis do espectro), iPLS (154 variáveis – 5.722 a 6.017 cm^{-1}), UVE-PLS (621 variáveis), GA-PLS (349 variáveis), GA-iPLS (16 variáveis) e GA-UVE-PLS (27 variáveis). A Tabela 11 apresenta os resultados dos modelos avaliados.

Tabela 11. Resultados para os modelos de calibração multivariada de Oliva – Algodão no infravermelho próximo.

Parâmetro	PLS (2.300 var.)			iPLS (154 var.)			UVE-PLS (621 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
Algodão	6	$7,41 \times 10^{-2}$	1,17	8	$9,79 \times 10^{-2}$	0,88	8	$1,80 \times 10^{-1}$	1,68
	4.534 – 8.968 cm^{-1}			5.722 – 6.017 cm^{-1}			----		
NIR	GA-PLS (349 var.)			GA-iPLS (16 var.)			GA-UVE-PLS (27 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	5	$4,43 \times 10^{-2}$	0,64	6	$3,65 \times 10^{-2}$	0,35	4	$8,53 \times 10^{-2}$	1,37
	150 gerações	convergência 80%		50 gerações	convergência 80%		100 gerações	convergência 80%	

Todos os modelos obtidos apresentaram valores de erros bem satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação. O modelo GA-iPLS (destaque cinza) com 6 VLs, RMSEP de $3,65 \times 10^{-2}\%$ e EM de 0,35%, apresentou os melhores resultados, sendo mais robusto e eficiente que os demais.

Na Figura 34 é apresentado o espectro no infravermelho próximo de uma amostra, destacando em vermelho as 16 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a estiramento de ligações C-H de grupos CH_2 (1^a overtone).

A seleção de variáveis por GA-iPLS usando como entrada as variáveis selecionadas pelo iPLS, 50 gerações e convergência de 80%, melhorou a calibração sendo necessário usar somente a informação contida em apenas 16 variáveis na região relacionada ao estiramento de ligações C-H de grupos CH_2 (1^a

overtone). Isso torna possível uma análise realizada com o monitoramento dessas 16 variáveis ao invés das 2.300 variáveis do espectro global, pois as demais variáveis não são necessárias para o desenvolvimento do modelo de calibração. Também possibilitou uma melhor interpretação química da relação região espectral e a contaminação.

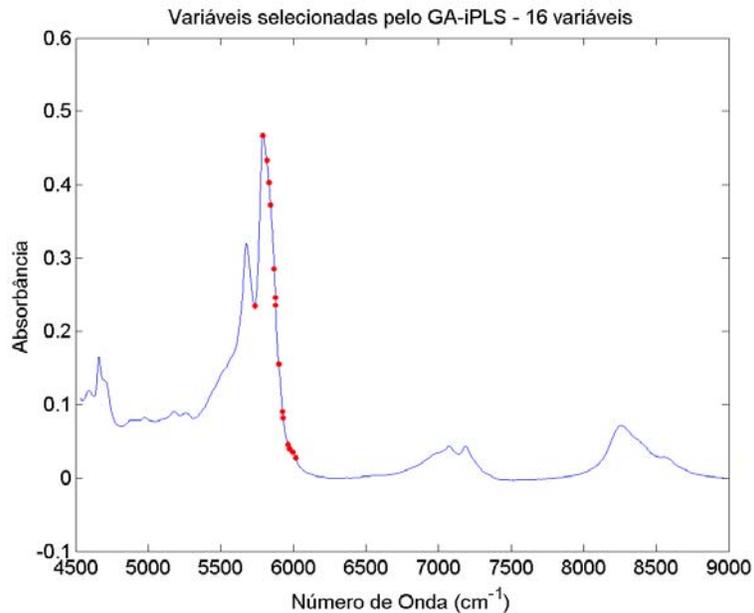


Figura 34. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Algodão com as variáveis selecionadas pelo GA-iPLS em vermelho.

A partir das mesmas amostras utilizadas para a calibração no MIR e no NIR, foram obtidos espectros Raman das amostras de Oliva contaminadas com Algodão, após corrigir o espalhamento e selecionado a região de 683 a 1.572 cm⁻¹ (651 variáveis), são apresentados na Figura 35.

Foram construídos os modelos: PLS (todas as 651 variáveis do espectro), iPLS (108 variáveis – 1.130 a 1.276 cm⁻¹), UVE-PLS (636 variáveis), GA-PLS (17 variáveis), GA-iPLS (10 variáveis) e GA-UVE-PLS (29 variáveis). A Tabela 12 apresenta os resultados dos modelos avaliados.

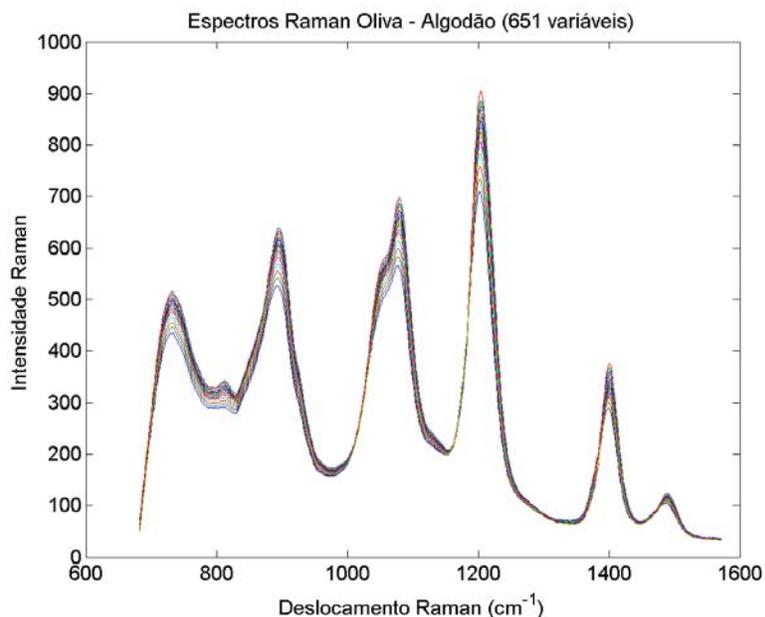


Figura 35. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Algodão.

Tabela 12. Resultados para os modelos de calibração multivariada de Oliva – Algodão no Raman.

Parâmetro	PLS (651 var.)			iPLS (108 var.)			UVE-PLS (636 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Algodão</i>	10	$5,83 \times 10^{-2}$	0,82	10	$5,81 \times 10^{-2}$	0,79	10	$5,83 \times 10^{-2}$	0,82
	683 – 1.572 cm ⁻¹			1.130 – 1.276 cm ⁻¹			----		
<i>Raman</i>	GA-PLS (17 var.)			GA-iPLS (10 var.)			GA-UVE-PLS (29 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	9	$5,84 \times 10^{-2}$	0,80	8	$5,70 \times 10^{-2}$	0,78	10	$5,83 \times 10^{-2}$	0,82
	100 gerações	convergência 50%		25 gerações	convergência 50%		100 gerações	convergência 50%	

Os modelos obtidos, apresentaram valores de erros semelhantes e um número de VLs superior aos modelos no MIR e NIR, representando possíveis problemas de robustez. Porém, este fato não interferiu nas previsões dos modelos de calibração. Considerando este fato e uma diferença muito pequena nos erros, escolheremos o modelo GA-iPLS (destaque cinza) com 8 VLs, RMSEP de $5,70 \times 10^{-2}\%$ e EM de 0,78% como o melhor modelo pela robustez e uma melhor interpretação química da relação região espectral e contaminação.

Na Figura 36 é apresentado o espectro Raman de uma amostra, destacando em vermelho as 10 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a estiramento de ligações C-C.

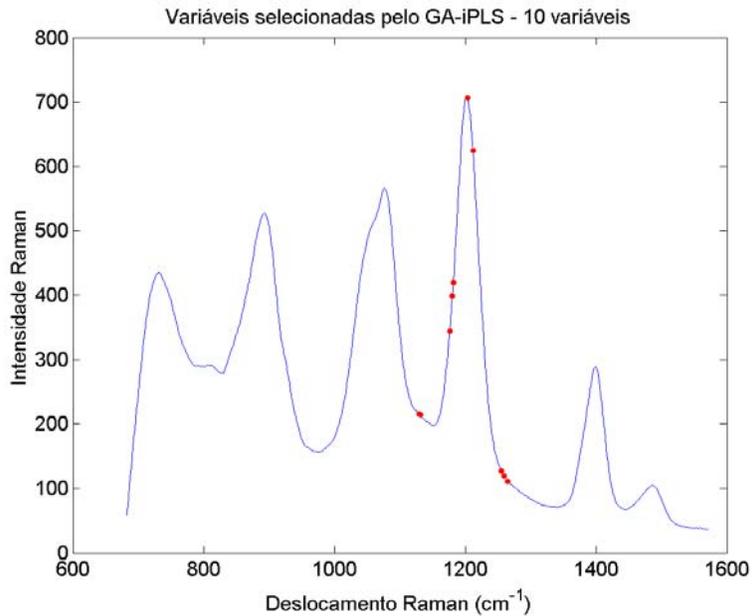


Figura 36. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Algodão com as variáveis selecionadas pelo GA-iPLS em vermelho.

O modelo GA-iPLS usando como entrada as variáveis selecionadas pelo iPLS, 25 gerações e convergência de 50%, melhorou um pouco a calibração sendo necessário então usar a informação contida em apenas 10 variáveis na região relacionada ao estiramento de ligações C-C, no lugar das 651 variáveis do espectro global.

Como os modelos obtidos a partir dos espectros Raman são equivalentes, fica a critério do analista a escolha do método a ser utilizado.

Comparando os diferentes métodos de seleção de variáveis aplicados as três técnicas vibracionais para o caso da contaminação do óleo de Oliva com óleo de Algodão, o método GA-iPLS mostrou-se mais eficiente que os demais, apresentando além da redução no número de variáveis uma melhor qualidade química na resposta.

4.3.2. Calibração Multivariada de óleo de Oliva contaminado com óleo de Canola

Os espectros na região do infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Canola nas concentrações entre 1,25 e 20,00% (m/m), após correção da linha base e selecionado a região de 629 a 3.134 cm^{-1} (1.300 variáveis), são apresentados na Figura 37.

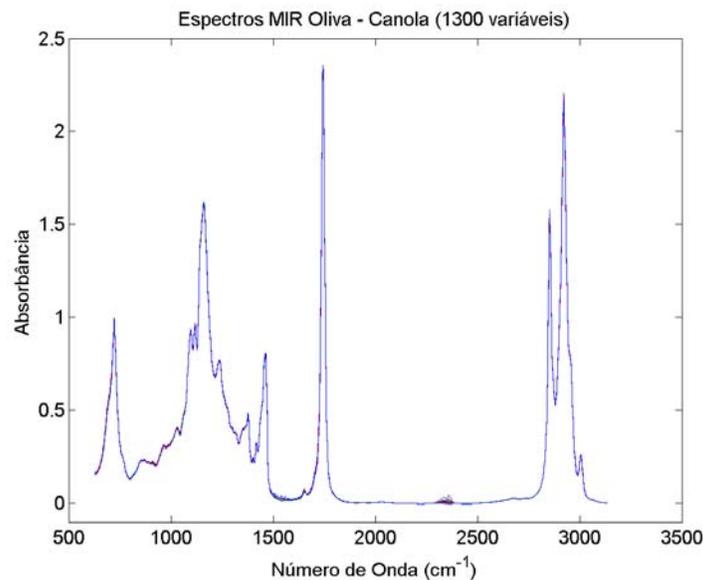


Figura 37. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Canola.

Na construção dos modelos de calibração utilizaram-se 16 amostras e 4 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.300 variáveis do espectro), iPLS (109 variáveis - 839 a 1.047 cm^{-1}), UVE-PLS (252 variáveis), GA-PLS (257 variáveis) e GA-iPLS (17 variáveis). A Tabela 13 apresenta os resultados dos modelos avaliados.

Todos os modelos obtidos apresentaram valores de erros bem satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação. Os modelos GA-iPLS e GA-PLS se equivalem e diferenciam dos demais, levando em conta o número de variáveis selecionadas e a interpretação química, consideramos o modelo GA-iPLS (destaque cinza) com 9 VLs, RMSEP de $1,02 \times 10^{-1}\%$ e EM de 0,64% o melhor.

Tabela 13. Resultados para os modelos de calibração multivariada de Oliva – Canola no infravermelho médio.

Parâmetro	PLS (1.300 var.)			iPLS (109 var.)			UVE-PLS (252 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
Canola	10	$1,89 \times 10^{-1}$	1,52	9	$1,74 \times 10^{-1}$	1,29	8	$1,94 \times 10^{-1}$	1,79
	629 – 3.134 cm^{-1}			839 – 1.047 cm^{-1}			----		
MIR	GA-PLS (257 var.)			GA-iPLS (17 var.)			GA-UVE-PLS (27 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	8	$1,36 \times 10^{-1}$	0,67	9	$1,02 \times 10^{-1}$	0,64	7	$1,64 \times 10^{-1}$	1,39
	100 gerações	convergência 80%		25 gerações	convergência 80%		50 gerações	convergência 80%	

Na Figura 38 é apresentado o espectro no infravermelho médio de uma amostra destacando em vermelho as 17 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados).

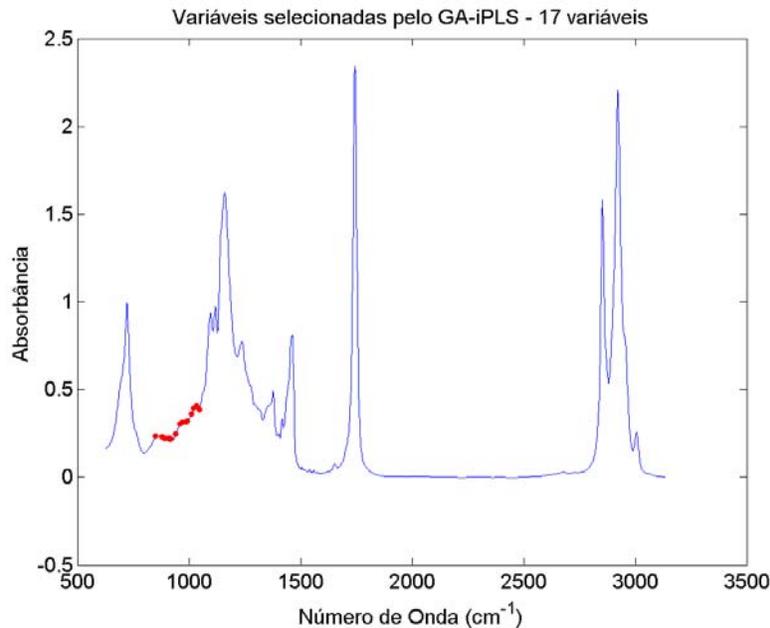


Figura 38. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Canola com as variáveis selecionadas pelo GA-iPLS em vermelho.

Os espectros na região do infravermelho próximo das mesmas amostras de Oliva contaminadas, após filtrar ruídos nos espectros, corrigir a linha base e selecionado a região de 4.534 a 8.968 cm^{-1} (2.300 variáveis), são apresentados na Figura 39.

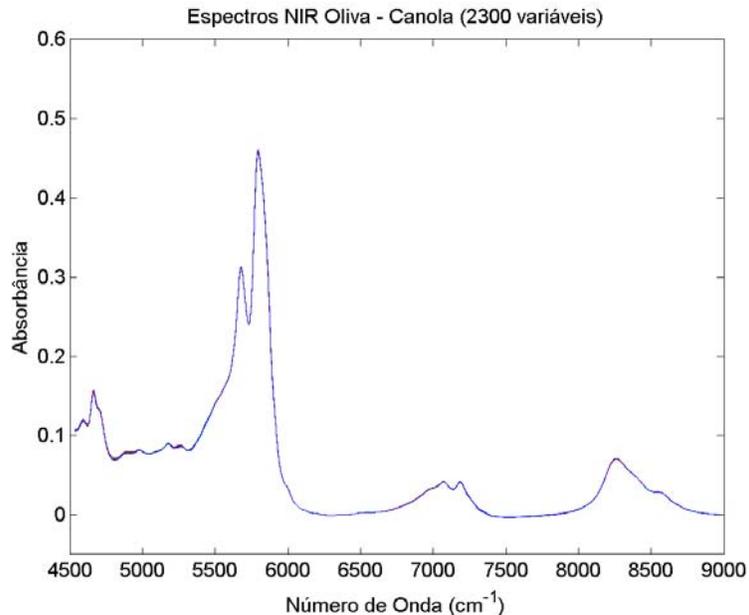


Figura 39. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Canola.

Na construção dos modelos de calibração e conjunto de previsão foram utilizadas as mesmas amostras dos modelos no infravermelho médio. Foram construídos os modelos: PLS (todas as 2.300 variáveis do espectro), iPLS (154 variáveis – 5.722 a 6.017 cm^{-1}), UVE-PLS (589 variáveis), GA-PLS (475 variáveis) e GA-iPLS (16 variáveis). A Tabela 14 apresenta os resultados dos modelos avaliados.

Todos os modelos obtidos apresentaram valores de erros bem satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação a favor do modelo GA-PLS (destaque cinza) com 7 VLs, RMSEP de $2,21 \times 10^{-2}\%$ e EM de 0,20%. O modelo foi semelhante ao GA-UVE-PLS mas com a vantagem de ser um pouco mais robusto.

Tabela 14. Resultados para os modelos de calibração multivariada de Oliva – Canola no infravermelho próximo.

Parâmetro	PLS (2.300 var.)			iPLS (154 var.)			UVE-PLS (589 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Canola</i>	8	$6,47 \times 10^{-2}$	0,55	7	$8,76 \times 10^{-2}$	0,68	9	$3,38 \times 10^{-2}$	0,32
	4.534 – 8.968 cm^{-1}			5.722 – 6.017 cm^{-1}			----		
	GA-PLS (475 var.)			GA-iPLS (16 var.)			GA-UVE-PLS (86 var.)		
<i>NIR</i>	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	7	$2,21 \times 10^{-2}$	0,20	5	$6,49 \times 10^{-2}$	0,56	10	$2,02 \times 10^{-2}$	0,24
	100 gerações	convergência 80%		25 gerações	convergência 50%		50 gerações	convergência 80%	

Na Figura 40 é apresentado o espectro no infravermelho próximo de uma amostra, destacando em vermelho as 475 variáveis selecionadas pelo modelo GA-PLS, distribuídas por todo espectro sem priorizar nenhuma região específica.

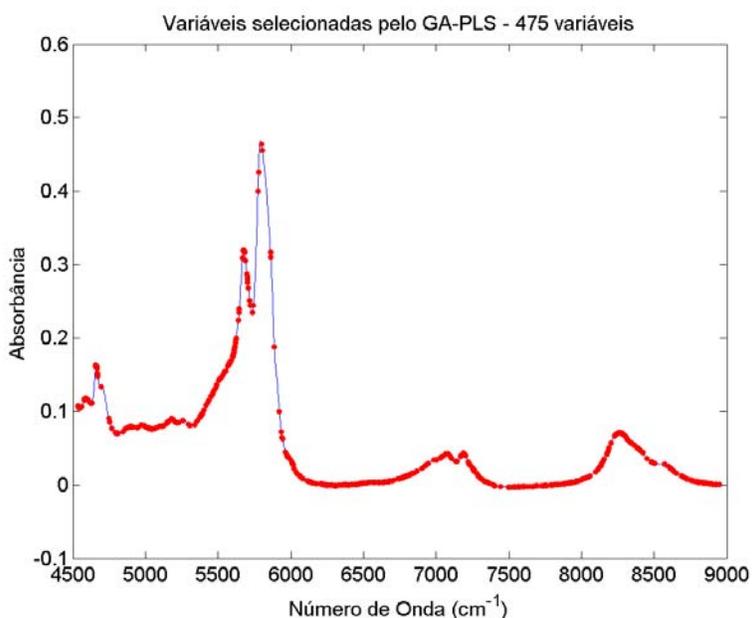


Figura 40. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Canola com as variáveis selecionadas pelo GA-PLS em vermelho.

O modelo GA-PLS, mostrou que a seleção de variáveis por GA, usando 100 gerações e convergência de 80%, melhorou a calibração sendo necessário então

usar a informação contida em apenas 475 variáveis ao invés das 2.300 variáveis do espectro global, pois as demais variáveis não são necessárias para o desenvolvimento do modelo de calibração.

Os espectros Raman das mesmas amostras de Oliva contaminadas, após corrigir o espalhamento e, selecionado a região de 683 a 1.572 cm^{-1} (651 variáveis), são apresentados na Figura 41.

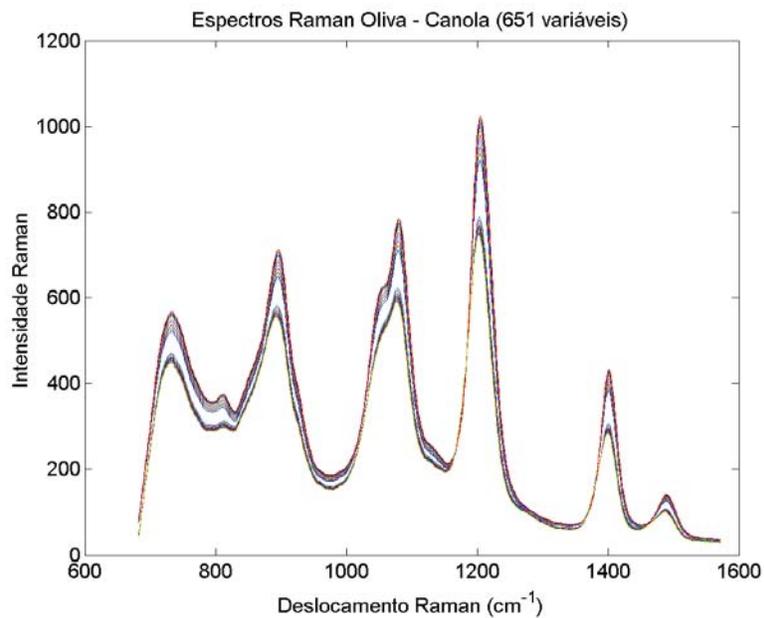


Figura 41. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Canola.

Foram construídos os modelos: PLS (todas as 651 variáveis do espectro), iPLS (108 variáveis – 1.130 a 1.276 cm^{-1}), UVE-PLS (424 variáveis), GA-PLS (81 variáveis) e GA-iPLS (16 variáveis). A Tabela 15 apresenta os resultados dos modelos avaliados.

Os resultados obtidos foram bem satisfatórios, com destaque para os modelos que usam GA. Porém, baseado novamente na qualidade da informação química escolheremos o modelo GA-iPLS (destaque cinza) com 8 VLs, RMSEP de $6,30 \times 10^{-3}\%$ e EM de 0,04% como o melhor.

Tabela 15. Resultados para os modelos de calibração multivariada de Oliva – Canola no Raman.

Parâmetro	PLS (651 var.)			iPLS (108 var.)			UVE-PLS (424 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Canola</i>	9	$1,15 \times 10^{-2}$	0,10	7	$1,35 \times 10^{-2}$	0,11	8	$9,40 \times 10^{-3}$	0,06
	683 – 1.572 cm^{-1}			1.130 – 1.276 cm^{-1}			----		
	GA-PLS (81 var.)			GA-iPLS (16 var.)			GA-UVE-PLS (18 var.)		
<i>Raman</i>	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	8	$6,90 \times 10^{-3}$	0,06	8	$6,30 \times 10^{-3}$	0,04	8	$9,40 \times 10^{-3}$	0,06
	100 gerações	convergência 80%		25 gerações	convergência 80%		100 gerações	convergência 80%	

Na Figura 42 é apresentado o espectro Raman de uma amostra destacando em vermelho as 16 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a estiramento de ligações C-C.

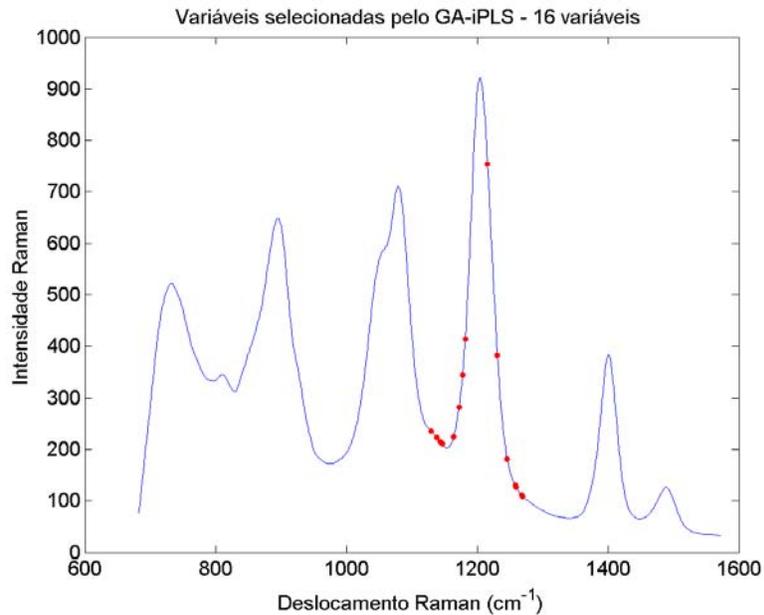


Figura 42. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Canola com as variáveis selecionadas pelo GA-iPLS em vermelho.

O modelo GA-iPLS usando 25 gerações e convergência de 80%, melhorou a calibração sendo necessário então usar a informação contida em apenas 16

variáveis na região relacionada ao estiramento de ligações C-C, no lugar das 651 variáveis do espectro global, pois as demais variáveis não são necessárias para o desenvolvimento do modelo de calibração.

Comparando os diferentes métodos de seleção de variáveis aplicados as três técnicas vibracionais para o caso da contaminação do óleo de Oliva com óleo de Canola, o método GA-iPLS mostrou-se mais eficiente para os modelos no MIR e Raman, apresentando além da redução no número de variáveis uma melhor qualidade química na resposta. Já na região do NIR o melhor modelo foi o GA-PLS que não melhora a qualidade química na resposta.

4.3.3. Calibração Multivariada de óleo de Oliva contaminado com óleo de Girassol

Os espectros na região do infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Girassol nas concentrações entre 1,00 e 17,50% (m/m), após correção da linha base e selecionado a região de 629 a 3.134 cm^{-1} (1.300 variáveis), são apresentados na Figura 43.

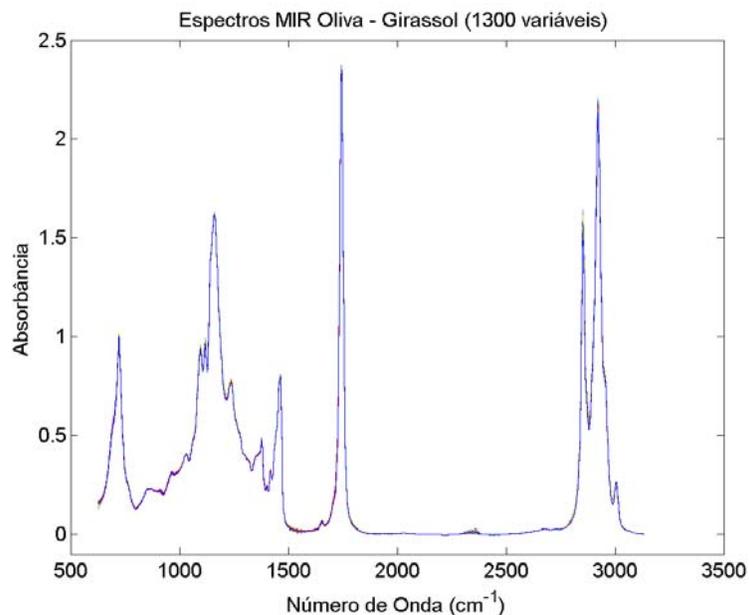


Figura 43. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Girassol.

Na construção dos modelos de calibração utilizaram-se 16 amostras e 4 amostras para o conjunto de previsão. Foram construídos os modelos: PLS (todas as 1.300 variáveis do espectro), iPLS (109 variáveis - 839 a 1.047 cm^{-1}), UVE-PLS (520 variáveis), GA-PLS (241 variáveis) e GA-iPLS (13 variáveis). A Tabela 16 apresenta os resultados dos modelos avaliados.

Tabela 16. Resultados para os modelos de calibração multivariada de Oliva - Girassol.

Parâmetro	PLS (1.300 var.)			iPLS (109 var.)			UVE-PLS (520 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Girassol</i>	7	$6,62 \times 10^{-2}$	0,62	7	$8,08 \times 10^{-2}$	0,94	6	$1,16 \times 10^{-1}$	0,59
	629 – 3.134 cm^{-1}			839 – 1.047 cm^{-1}			----		
<i>MIR</i>	GA-PLS (241 var.)			GA-iPLS (13 var.)			GA-UVE-PLS (53 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	10	$4,34 \times 10^{-2}$	0,35	7	$6,24 \times 10^{-2}$	0,52	6	$7,40 \times 10^{-2}$	0,55
	100 gerações	convergência 80%		25 gerações	convergência 50%		100 gerações	convergência 50%	

Todos os modelos obtidos apresentaram valores de erros bem satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação. Nesse sentido, o modelo GA-PLS (destaque cinza) com 10 VLs, RMSEP de $4,34 \times 10^{-2}\%$ e EM de 0,35%, apresentou os melhores resultados, sendo mais eficiente na previsão que os demais.

Na Figura 44 é apresentado o espectro no infravermelho médio de uma amostra, destacando em vermelho as 241 variáveis selecionadas pelo modelo GA-PLS, distribuídas por todo espectro sem priorizar nenhuma região específica.

O modelo GA-PLS usando 100 gerações e convergência de 80%, melhorou a calibração sendo necessário então usar a informação contida em apenas 241 variáveis ao invés das 1.300 variáveis do espectro global, pois as demais variáveis não são necessárias para o desenvolvimento do modelo de calibração.

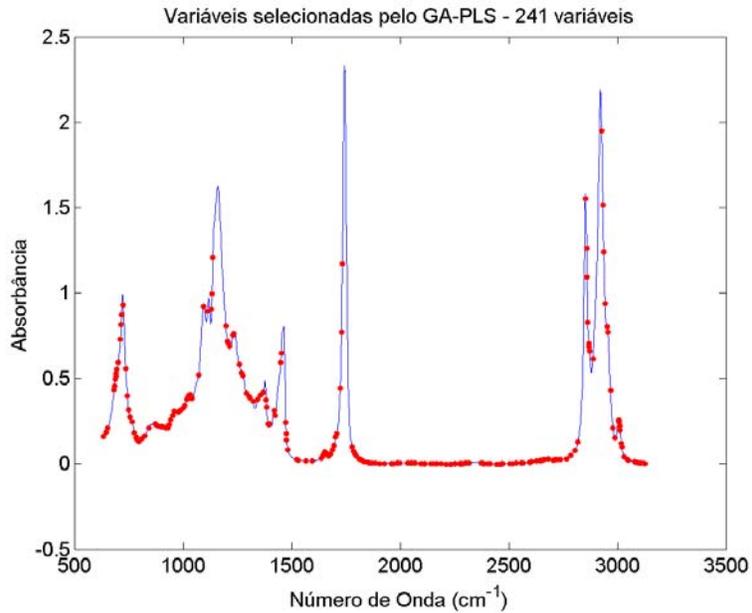


Figura 44. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Girassol com as variáveis selecionadas pelo GA-PLS em vermelho.

Os espectros na região do infravermelho próximo das amostras de Oliva contaminadas com Girassol, após filtrar ruídos nos espectros, corrigir a linha base e selecionado a região de 4.534 a 8.968 cm^{-1} (2.300 variáveis), são apresentados na Figura 45.

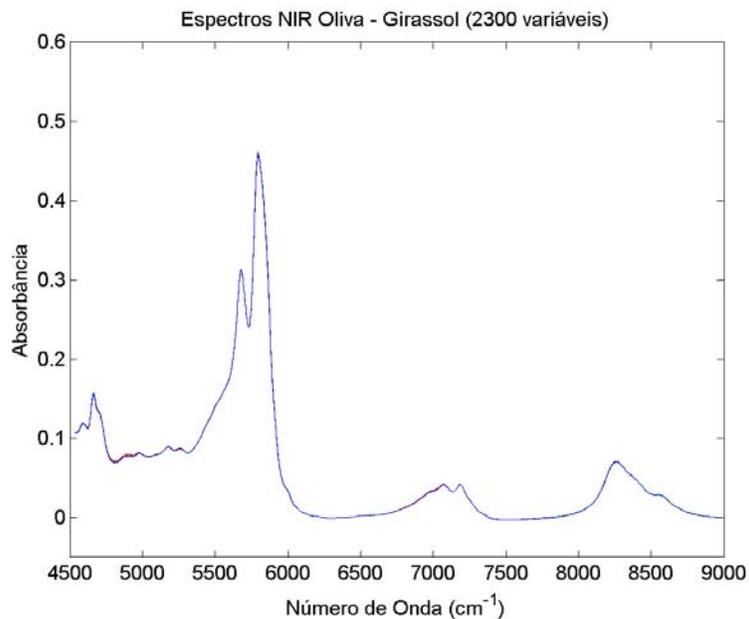


Figura 45. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Girassol.

Em seguida mantendo as mesmas amostras dos modelos no infravermelho médio foram construídos os modelos: PLS (todas as 2.300 variáveis do espectro), iPLS (230 variáveis – 5.421 a 5.863 cm^{-1}), UVE-PLS (925 variáveis), GA-PLS (320 variáveis) e GA-iPLS (28 variáveis). A Tabela 17 apresenta os resultados dos modelos avaliados.

Tabela 17. Resultados para os modelos de calibração multivariada de Oliva – Girassol no infravermelho próximo.

Parâmetro	PLS (2.300 var.)			iPLS (230 var.)			UVE-PLS (925 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Girassol</i>	4	$1,55 \times 10^{-1}$	1,79	3	$1,32 \times 10^{-1}$	0,75	4	$1,54 \times 10^{-1}$	1,68
	4.534 – 8.968 cm^{-1}			5.421 – 5.863 cm^{-1}			----		
<i>NIR</i>	GA-PLS (320 var.)			GA-iPLS (28 var.)			GA-UVE-PLS (80 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	2	$1,02 \times 10^{-1}$	0,72	3	$1,35 \times 10^{-1}$	0,54	2	$1,77 \times 10^{-1}$	1,21
	150 gerações	convergência 80%		25 gerações	convergência 80%		100 gerações	convergência 80%	

Os modelos obtidos apresentaram valores de erros bem satisfatórios mas com diferença quanto à robustez e eficiência da previsão das amostras de validação a favor do modelo GA-iPLS (destaque cinza) com 3 VLs, RMSEP de $1,35 \times 10^{-1}\%$ e EM de 0,54% que apresentou os melhores resultados.

Na Figura 46 é apresentado o espectro no infravermelho próximo de uma amostra, destacando em vermelho as 28 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a estiramento de ligações C-H de grupos CH_2 (1ª overtone).

O modelo GA-iPLS obtido após 25 gerações e convergência de 80%, melhorou a calibração usando no lugar das 2.300 variáveis do modelo global apenas 28 variáveis da região relacionada ao estiramento de ligações C-H de grupos CH_2 (1ª overtone).

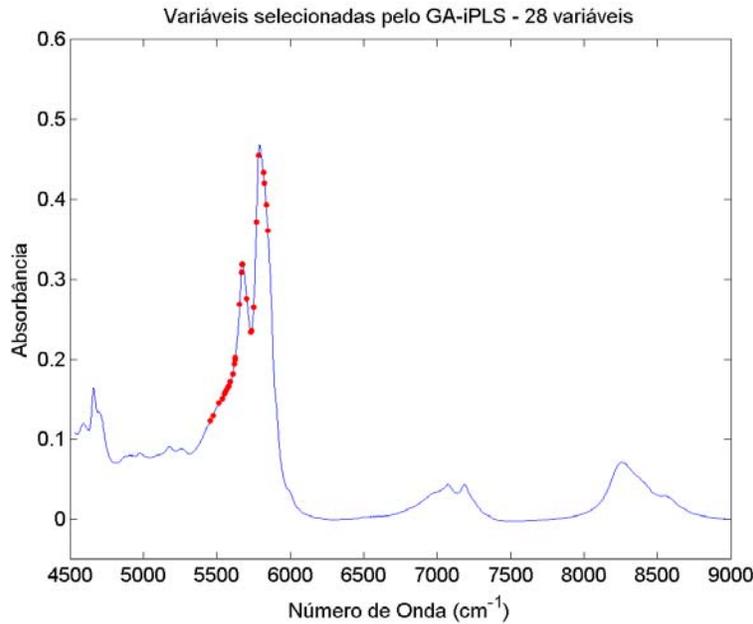


Figura 46. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Girassol com as variáveis selecionadas pelo GA-iPLS em vermelho.

Na Figura 47 são apresentados, os espectros Raman das mesmas amostras de Oliva contaminadas para a calibração, no MIR e NIR após corrigir o espalhamento e selecionado a região de 683 a 1.572 cm^{-1} (651 variáveis).

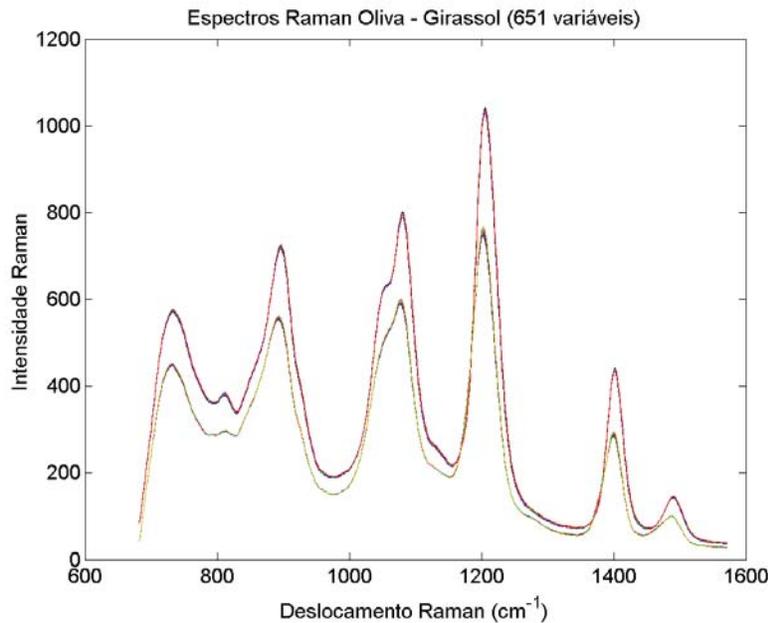


Figura 47. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Girassol.

Foram construídos os modelos: PLS (todas as 651 variáveis do espectro), iPLS (108 variáveis – 1.130 a 1.276 cm^{-1}), UVE-PLS (610 variáveis), GA-PLS (19 variáveis), GA-iPLS (20 variáveis) e GA-UVE-PLS (17 variáveis). A Tabela 18 apresenta os resultados dos modelos avaliados.

Tabela 18. Resultados para os modelos de calibração multivariada de Oliva – Girassol no Raman.

Parâmetro	PLS (651 var.)			iPLS (108 var.)			UVE-PLS (610 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Girassol</i>	11	$1,09 \times 10^{-1}$	1,36	5	$8,49 \times 10^{-2}$	1,26	11	$1,13 \times 10^{-1}$	1,30
	683 – 1.572 cm^{-1}			1.130 – 1.276 cm^{-1}			----		
<i>Raman</i>	GA-PLS (19 var.)			GA-iPLS (20 var.)			GA-UVE-PLS (17 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	11	$6,95 \times 10^{-2}$	1,05	5	$6,62 \times 10^{-2}$	0,98	11	$1,13 \times 10^{-1}$	1,30
	100 gerações	convergência 50%		25 gerações	convergência 80%		100 gerações	convergência 80%	

Os modelos obtidos apresentaram valores de erros satisfatórios mas diferença quanto à robustez e eficiência da previsão das amostras de validação. O modelo GA-iPLS (destaque cinza) com 5 VLs, RMSEP de $6,62 \times 10^{-2}\%$ e EM de 0,98%, apresentou os melhores resultados, sendo mais robusto e eficiente. Embora o modelo GA-PLS apresentou resultados semelhantes ao GA-iPLS em relação a previsão, o problema se encontra no maior número de VLs. Portanto não consiste em uma boa opção para a calibração.

Na Figura 48 é apresentado o espectro Raman de uma amostra destacando em vermelho as 20 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a estiramento de ligações C-C.

O modelo GA-iPLS obtido após 25 gerações e convergência de 80%, melhorou a calibração usando no lugar das 651 variáveis do modelo global apenas 20 variáveis da região relacionada a estiramento de ligações C-C.

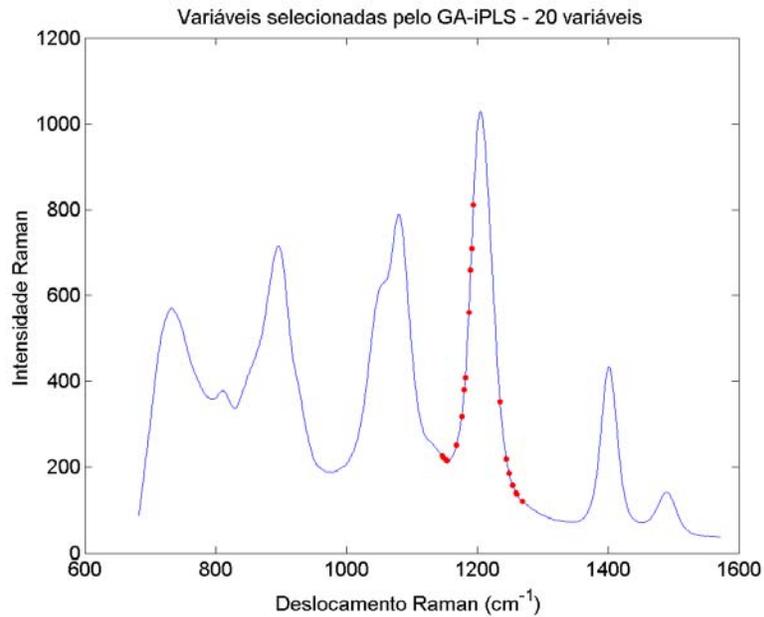


Figura 48. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Girassol com as variáveis selecionadas pelo GA-iPLS em vermelho.

Comparando os diferentes métodos de seleção de variáveis aplicados as três técnicas vibracionais para o caso da contaminação do óleo de Oliva com óleo de Girassol, o método GA-iPLS mostrou-se mais eficiente para os modelos no NIR e Raman, apresentando além da redução no número de variáveis uma melhor qualidade química na resposta. Já na região do MIR o melhor modelo foi o GA-PLS que não melhora a qualidade química na resposta.

4.3.4. Calibração Multivariada de óleo de Oliva contaminado com óleo de Milho

Os espectros na região do infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Milho nas concentrações entre 1,75 e 20,00% (m/m), após correção da linha base e selecionado a região de 629 a 3.134 cm^{-1} (1.300 variáveis), são apresentados na Figura 49.

Foram construídos os modelos: PLS (todas as 1.300 variáveis do espectro), iPLS (109 variáveis - 839 a 1.047 cm^{-1}), UVE-PLS (324 variáveis), GA-PLS (272

variáveis), GA-iPLS (15 variáveis) e GA-UVE-PLS (19 variáveis). A Tabela 19 apresenta os resultados dos modelos avaliados.

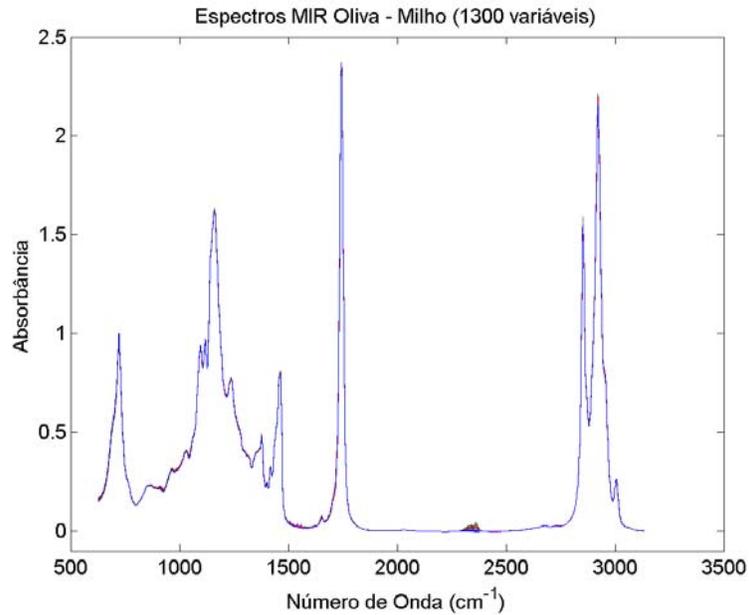


Figura 49. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Milho.

Tabela 19. Resultados para os modelos de calibração multivariada de Oliva – Milho no infravermelho médio.

Parâmetro	PLS (1.300 var.)			iPLS (109 var.)			UVE-PLS (324 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Milho</i>	9	$2,26 \times 10^{-1}$	0,93	8	$8,14 \times 10^{-2}$	0,57	8	$7,38 \times 10^{-2}$	0,44
	629 – 3.134 cm^{-1}			839 – 1.047 cm^{-1}			----		
	GA-PLS (272 var.)			GA-iPLS (15 var.)			GA-UVE-PLS (19 var.)		
<i>MIR</i>	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	9	$9,88 \times 10^{-2}$	0,78	6	$4,82 \times 10^{-2}$	0,26	8	$1,18 \times 10^{-1}$	0,59
	100 gerações	convergência 80%		25 gerações	convergência 80%		100 gerações	convergência 80%	

Os resultados dos modelos foram satisfatórios com baixos valores de erros. O modelo GA-iPLS (destaque cinza) com 6 VLs, RMSEP de $4,82 \times 10^{-2}\%$ e EM de 0,26%, apresentou os melhores resultados. Esse modelo foi obtido após 25 gerações e convergência de 80%.

Na Figura 50 é apresentado o espectro de uma amostra destacando em vermelho as 15 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados).

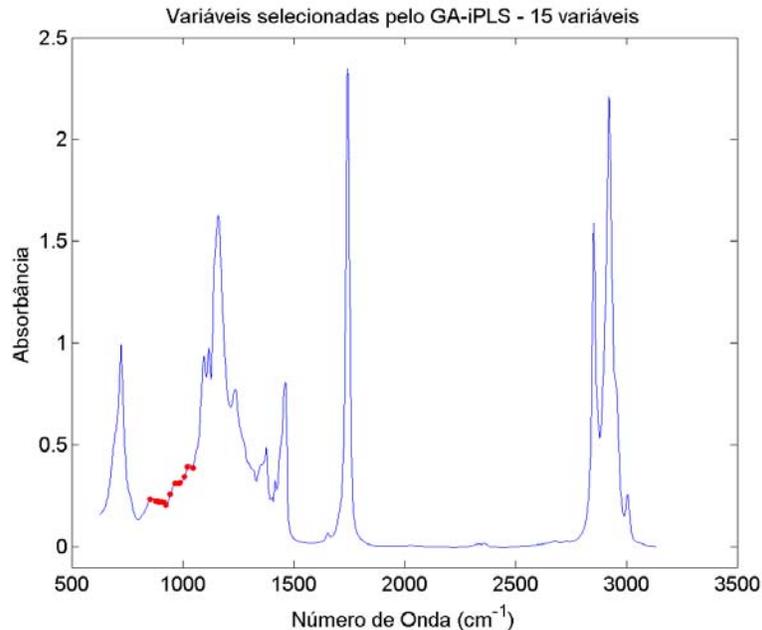


Figura 50. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Milho com as variáveis selecionadas pelo GA-iPLS em vermelho.

Para a calibração multivariada de Oliva contaminado com Milho no MIR, o modelo GA-iPLS melhorou a calibração sendo necessário usar a informação contida em apenas 15 variáveis na região relacionada à deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados) ao invés das 1.300 variáveis do espectro global, pois as demais variáveis não são necessárias para o desenvolvimento do modelo de calibração. O modelo GA-iPLS possibilitou uma melhor interpretação química da relação região espectral e contaminação.

Os espectros na região do infravermelho próximo das mesmas amostras de Oliva contaminadas, após filtrar ruídos nos espectros, corrigir a linha base e selecionado a região de 4.534 a 8.968 cm^{-1} (2.300 variáveis), são apresentados na Figura 51.

Foram construídos os modelos: PLS (todas as 2.300 variáveis do espectro), iPLS (230 variáveis – 5.421 a 5.863 cm^{-1}), UVE-PLS (356 variáveis), GA-PLS (443 variáveis), GA-iPLS (32 variáveis) e GA-UVE-PLS (23 variáveis). A Tabela 20 apresenta os resultados dos modelos avaliados.

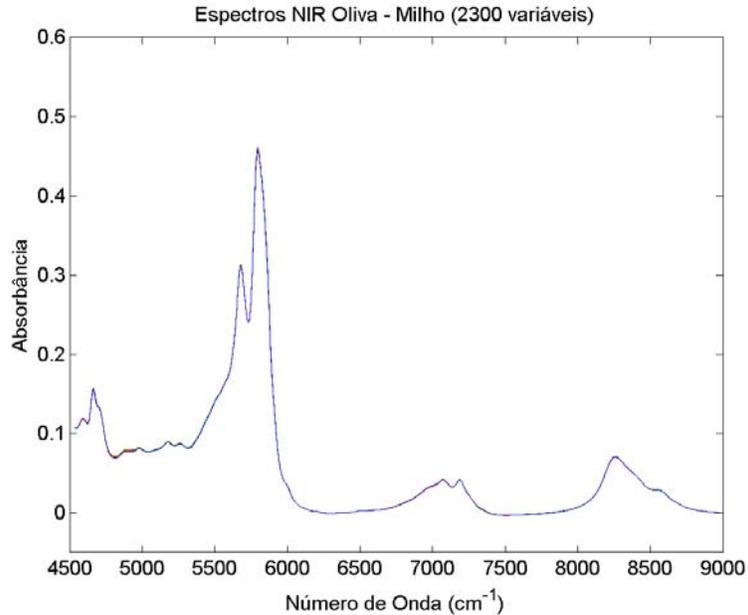


Figura 51. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Milho.

Tabela 20. Resultados para os modelos de calibração multivariada de Oliva – Milho no infravermelho próximo.

Parâmetro	PLS (2.300 var.)			iPLS (230 var.)			UVE-PLS (356 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Milho</i>	8	$2,26 \times 10^{-1}$	2,13	8	$2,11 \times 10^{-1}$	1,89	8	$1,97 \times 10^{-1}$	1,76
	4.534 – 8.968 cm^{-1}			5.421 – 5.863 cm^{-1}			----		
<i>NIR</i>	GA-PLS (443 var.)			GA-iPLS (32 var.)			GA-UVE-PLS (23 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	10	$2,06 \times 10^{-1}$	1,42	6	$1,74 \times 10^{-1}$	1,71	8	$2,21 \times 10^{-1}$	1,74
	100 gerações	convergência 80%		25 gerações	convergência 80%		80 gerações	convergência 80%	

Todos os modelos obtidos apresentaram valores de erros satisfatórios com valores equivalentes na maioria deles. Devido a uma pequena diferença a favor das previsões, podemos considerar o modelo GA-PLS (destaque cinza) com 10 VLs, RMSEP de $2,06 \times 10^{-1} \%$ e EM de 0,35% melhor que os demais.

Na Figura 52 é apresentado o espectro no infravermelho próximo de uma amostra, destacando em vermelho as 443 variáveis selecionadas pelo modelo GA-PLS distribuídas por todo espectro sem priorizar nenhuma região específica.

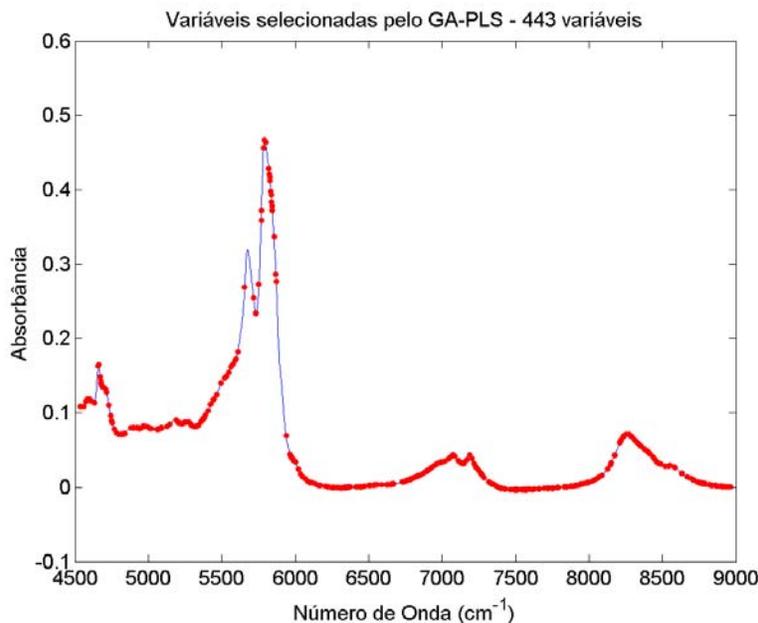


Figura 52. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Milho com as variáveis selecionadas pelo GA-PLS em vermelho.

A partir das amostras utilizadas para a calibração no MIR e NIR, foram obtidos espectros Raman das amostras de Oliva contaminadas com Milho, após corrigir o espalhamento e selecionado a região de 683 a 1.572 cm^{-1} (651 variáveis), são apresentados na Figura 53.

Foram construídos os modelos: PLS (todas as 651 variáveis do espectro), iPLS (108 variáveis – 1.130 a 1.276 cm^{-1}), UVE-PLS (605 variáveis), GA-PLS (122 variáveis), GA-iPLS (12 variáveis) e GA-UVE-PLS (91 variáveis). A Tabela 21 apresenta os resultados dos modelos avaliados.

Todos os modelos obtidos apresentaram valores de erros satisfatórios mas com diferença na eficiência da previsão das amostras de validação. O modelo GA-PLS (destaque cinza) com 7 VLs, RMSEP de $2,84 \times 10^{-2}\%$ e EM de 0,22% foi melhor que os outros modelos.

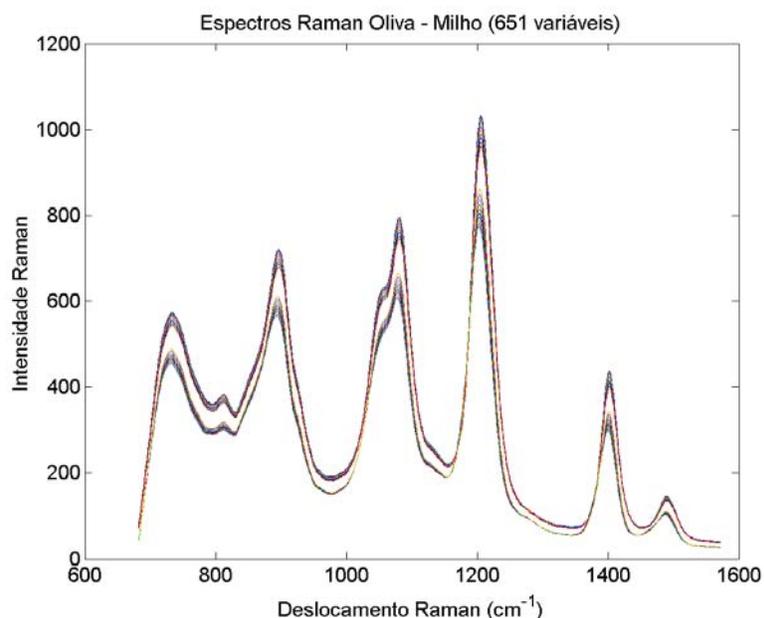


Figura 53. Espectros Raman das 20 amostras de óleo de Oliva contaminadas com óleo de Milho.

Tabela 21. Resultados para os modelos de calibração multivariada de Oliva – Milho no Raman.

Parâmetro	PLS (651 var.)			iPLS (108 var.)			UVE-PLS (605 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
<i>Milho</i>	6	$1,11 \times 10^{-1}$	1,02	8	$6,16 \times 10^{-2}$	0,74	6	$1,14 \times 10^{-1}$	1,03
	683 – 1.572 cm^{-1}			1.130 – 1.276 cm^{-1}			----		
	GA-PLS (122 var.)			GA-iPLS (12 var.)			GA-UVE-PLS (91 var.)		
<i>Raman</i>	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	6	$1,04 \times 10^{-1}$	0,92	7	$2,84 \times 10^{-2}$	0,22	6	$1,14 \times 10^{-1}$	1,03
	50 gerações	convergência 50%		25 gerações	convergência 80%		100 gerações	convergência 80%	

Na Figura 54 é apresentado o espectro Raman de uma amostra destacando em vermelho as 12 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a estiramento de ligações C-C.

O modelo GA-iPLS usando 25 gerações e convergência de 80%, melhorou a calibração sendo necessário usar a informação contida em apenas 12 variáveis na região relacionada ao estiramento de ligações C-C, no lugar das 651 variáveis do espectro global, pois as demais variáveis não são necessárias para o modelo de calibração.

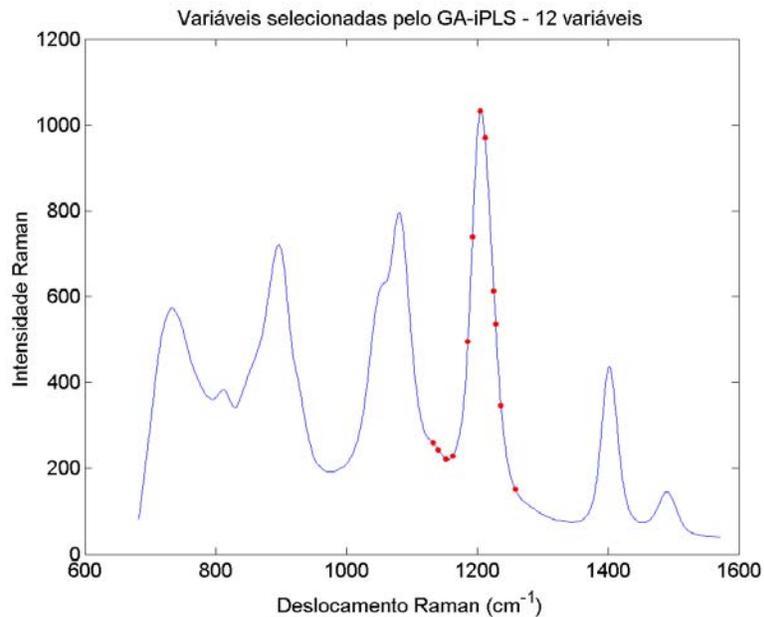


Figura 54. Espectro Raman de uma amostra de óleo de Oliva contaminada com óleo de Milho com as variáveis selecionadas pelo GA-iPLS em vermelho.

Comparando os diferentes métodos de seleção de variáveis aplicados as três técnicas vibracionais para o caso da contaminação do óleo de Oliva com óleo de Milho, o método GA-iPLS mostrou-se mais eficiente para os modelos no MIR e Raman, apresentando além da redução no número de variáveis uma melhor qualidade química na resposta. Já na região do NIR o melhor modelo foi o GA-PLS que não melhora a qualidade química na resposta.

4.3.5. Calibração Multivariada de óleo de Oliva contaminado com óleo de Soja

Os espectros na região do infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Soja nas concentrações entre 1,00 e 15,00% (m/m), após correção da linha base e selecionado a região de 629 a 3.134 cm^{-1} (1.300 variáveis), são apresentados na Figura 55.

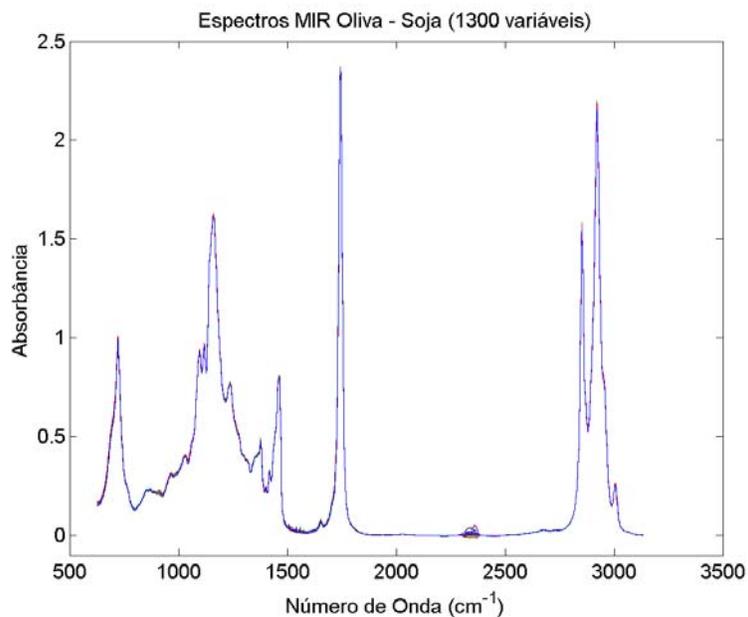


Figura 55. Espectros no infravermelho médio das 20 amostras de óleo de Oliva contaminadas com óleo de Soja.

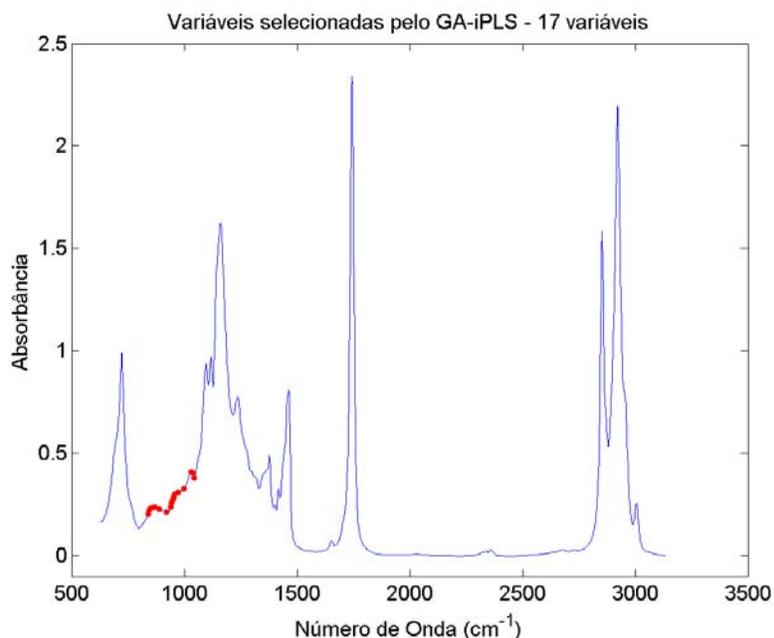
Foram construídos os modelos: PLS (todas as 1.300 variáveis do espectro), iPLS (109 variáveis - 839 a 1.047 cm^{-1}), UVE-PLS (552 variáveis), GA-PLS (257 variáveis), GA-iPLS (17 variáveis) e GA-UVE-PLS (60 variáveis). A Tabela 22 apresenta os resultados dos modelos avaliados.

Os modelos obtidos apresentaram valores de erros satisfatórios mas diferentes, tendo o modelo GA-iPLS com 7 VLs, RMSEP de $8,71 \times 10^{-2}\%$ e EM de 0,89%, apresentados os melhores resultados.

Tabela 22. Resultados para os modelos de calibração multivariada de Oliva – Soja no infravermelho médio.

Parâmetro	PLS (1.300 var.)			iPLS (109 var.)			UVE-PLS (552 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
Soja	10	$1,40 \times 10^{-1}$	1,45	6	$9,80 \times 10^{-2}$	1,85	10	$1,22 \times 10^{-1}$	2,32
	629 – 3.134 cm^{-1}			839 – 1.047 cm^{-1}			----		
MIR	GA-PLS (257 var.)			GA-iPLS (17 var.)			GA-UVE-PLS (60 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	8	$1,16 \times 10^{-1}$	1,76	7	$8,71 \times 10^{-2}$	0,89	6	$1,54 \times 10^{-1}$	1,93
	100 gerações	convergência 80%		25 gerações	convergência 80%		100 gerações	convergência 80%	

Na Figura 56 é apresentado o espectro no infravermelho médio de uma amostra destacando em vermelho as 17 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados).


Figura 56. Espectro no infravermelho médio de uma amostra de óleo de Oliva contaminada com óleo de Soja com as variáveis selecionadas pelo GA-iPLS em vermelho.

Para a calibração multivariada de Oliva contaminado com Soja no MIR, o modelo GA-iPLS melhorou a calibração sendo necessário usar a informação contida em apenas 17 variáveis na região relacionada à deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados) ao invés das 1.300 variáveis do espectro global.

Os espectros na região do infravermelho próximo das mesmas amostras de Oliva contaminadas, após filtrar ruídos nos espectros, corrigir a linha base e selecionado a região de 4.534 a 8.968 cm^{-1} (2.300 variáveis), são apresentados na Figura 57.

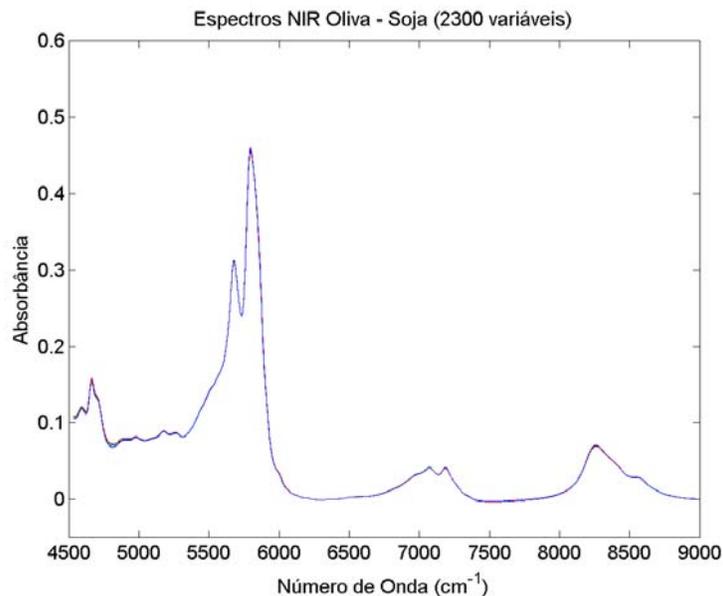


Figura 57. Espectros no infravermelho próximo das 20 amostras de óleo de Oliva contaminadas com óleo de Soja

Foram construídos os modelos: PLS (todas as 2.300 variáveis do espectro), iPLS (154 variáveis – 5.722 a 6.017 cm^{-1}), UVE-PLS (483 variáveis), GA-PLS (441 variáveis), GA-iPLS (20 variáveis) e GA-UVE-PLS (21 variáveis). A Tabela 23 apresenta os resultados dos modelos avaliados.

Todos os modelos obtidos apresentaram valores de erros satisfatórios mas o modelo GA-iPLS (destaque cinza) com 3 VLs, RMSEP de $4,49 \times 10^{-2}\%$ e EM de 0,58%, apresentou os melhores resultados, sendo mais robusto e eficiente que os demais.

Tabela 23. Resultados para os modelos de calibração multivariada de Oliva – Soja no infravermelho próximo.

Parâmetro	PLS (2.300 var.)			iPLS (154 var.)			UVE-PLS (483 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
Soja	3	$8,35 \times 10^{-2}$	1,40	3	$5,43 \times 10^{-2}$	0,96	3	$7,50 \times 10^{-2}$	1,46
	4.534 – 8.968 cm^{-1}			5.722 – 6.017 cm^{-1}			----		
NIR	GA-PLS (441 var.)			GA-iPLS (20 var.)			GA-UVE-PLS (21 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	2	$7,33 \times 10^{-2}$	0,94	3	$4,49 \times 10^{-2}$	0,58	3	$6,81 \times 10^{-2}$	1,09
	100 gerações	convergência 80%		25 gerações	converg. 80%		100 gerações	convergência 80%	

Na Figura 58 é apresentado o espectro no infravermelho próximo de uma amostra, destacando em vermelho as 20 variáveis selecionadas pelo modelo GA-iPLS, na região relacionada a estiramento de ligações C-H de grupos CH_2 (1ª overtone).

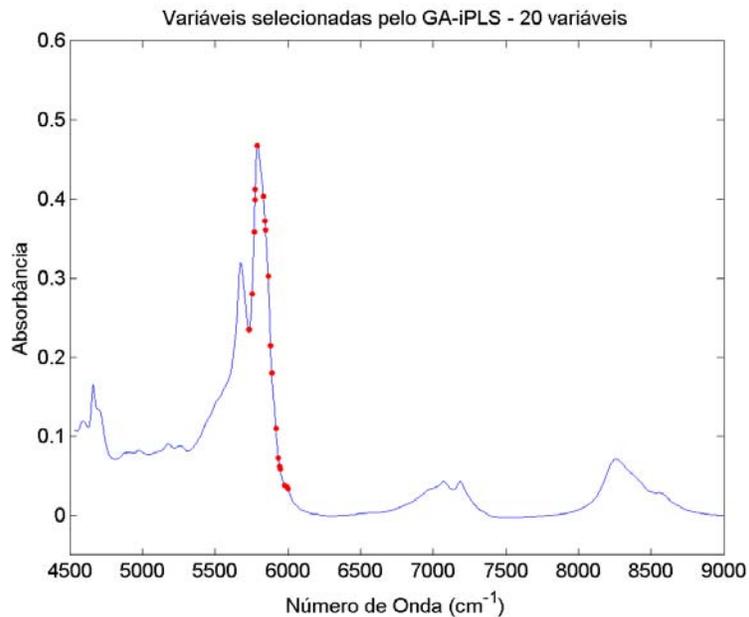


Figura 58. Espectro no infravermelho próximo de uma amostra de óleo de Oliva contaminada com óleo de Soja com as variáveis selecionadas pelo GA-iPLS em vermelho.

O modelo GA-iPLS usando como entrada as variáveis selecionadas pelo iPLS, 25 gerações e convergência de 80%, melhorou a calibração sendo necessário usar a informação contida em apenas 20 variáveis na região relacionada ao estiramento de ligações C-H de grupos CH₂ (1^a overtone).

Comparando os diferentes métodos de seleção de variáveis aplicados as duas técnicas vibracionais para o caso da contaminação do óleo de Oliva com óleo de Soja, o método GA-iPLS mostrou-se mais eficiente que os demais, apresentando além da redução no número de variáveis uma melhor qualidade química na resposta.

4.3.6. Calibração Multivariada de todos os contaminantes em um só modelo

Nesta etapa do trabalho, para cada uma das técnicas de espectroscopia no infravermelho, todas as matrizes dos espectros das cinco contaminações do óleo de Oliva foram agrupadas em uma única matriz de dados. O mesmo foi feito com os respectivos vetores de concentrações de cada um deles.

Essas novas matrizes de dados (uma de espectros no MIR e outra no NIR) foram utilizadas na construção de modelos de calibração por PLS e seleção de variáveis, com o objetivo de verificar a eficiência da previsão das amostras de Oliva adulteradas, utilizando um único modelo de calibração para todas as adulterações. Não se trata de mistura de adulterantes e sim usar todos espectros dentro de um único modelo.

Os espectros na região do infravermelho das 100 amostras de óleo de Oliva contaminadas nas concentrações entre 1,00 e 20,00% (m/m), após correção da linha base e selecionado a região de 629 a 3.134 cm⁻¹ (1.300 variáveis), são apresentados na Figura 59.

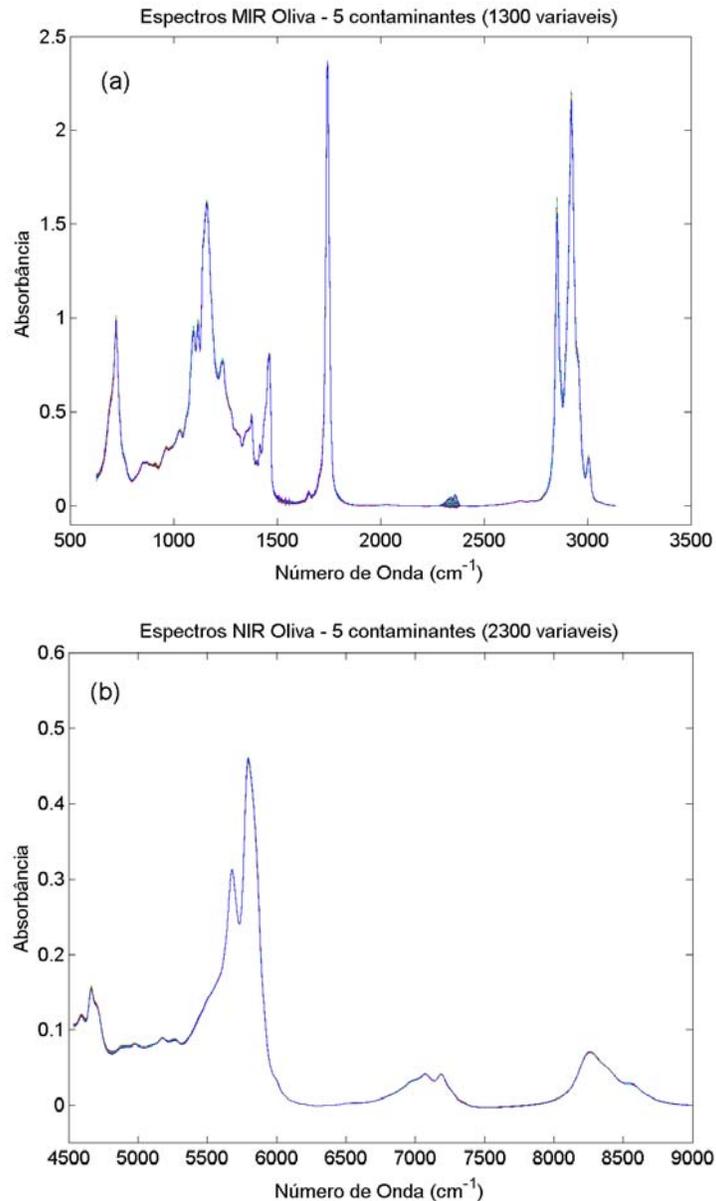


Figura 59. Espectros no infravermelho de todas as amostras de óleo de Oliva contaminadas. (a) no médio e (b) no próximo.

Foram construídos os modelos: PLS, iPLS, UVE-PLS, GA-PLS, GA-iPLS e GA-UVE-PLS. A Tabela 24 apresenta os resultados dos modelos avaliados.

Todos os modelos obtidos apresentaram valores de erros maiores que os erros dos modelos individuais, ou seja, com espectros só de um adulterante. Além do aumento dos erros de previsão, ocorre também um aumento significativo no número de VLs usadas pelo modelo, diminuindo a eficiência de previsão.

Tabela 24. Resultados para os modelos de calibração multivariada de Oliva – todos contaminantes no mesmo modelo.

Parâmetro	PLS (1.300 var.)			iPLS (109 var.)			UVE-PLS (212 var.)		
<i>Todos</i> <i>contaminantes</i>	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	19	$5,19 \times 10^{-1}$	5,55	14	$5,06 \times 10^{-1}$	5,53	16	$4,71 \times 10^{-1}$	5,63
	629 – 3.134 cm^{-1}			839 – 1.047 cm^{-1}			----		
<i>MIR</i>	GA-PLS (166 var.)			GA-iPLS (15 var.)			GA-UVE-PLS (17 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	14	$4,45 \times 10^{-1}$	4,89	9	$5,58 \times 10^{-1}$	5,35	9	$5,42 \times 10^{-1}$	4,83
	100 gerações	convergência 80%		20 gerações	convergência 80%		50 gerações	convergência 80%	
Parâmetro	PLS (2.300 var.)			iPLS (460 var.)			UVE-PLS (332 var.)		
<i>Todos</i> <i>contaminantes</i>	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	19	$2,99 \times 10^{-1}$	2,47	13	$4,90 \times 10^{-1}$	5,11	11	$4,15 \times 10^{-1}$	3,37
	4.534 – 8.968 cm^{-1}			4.534 – 5.419 cm^{-1}			----		
<i>NIR</i>	GA-PLS (470 var.)			GA-iPLS (80 var.)			GA-UVE-PLS (50 var.)		
	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)	VL	RMSEP (%)	EM (%)
	11	$3,31 \times 10^{-1}$	2,49	13	$4,78 \times 10^{-1}$	5,24	10	$3,44 \times 10^{-1}$	3,39
	50 gerações	convergência 80%		20 gerações	convergência 80%		25 gerações	convergência 80%	

Os modelos GA-UVE-PLS (destaque cinza) com 9 VLs, RMSEP de $5,42 \times 10^{-1}\%$ e EM de 4,83% e o GA-PLS (destaque cinza) com 11 VLs, RMSEP de $3,31 \times 10^{-1}\%$ e EM de 2,49% apresentaram os melhores resultados para os modelos no MIR e NIR respectivamente, mostrando que o GA melhorou os modelos de calibração.

Como os modelos de calibração multivariada por PLS e seleção de variáveis obtidos nesta etapa, detectam a presença da contaminação mas não distingue o contaminante, a melhor opção para a calibração foi os modelos individuais conforme mostraram os resultados apresentados anteriormente.

4.4. Conclusões da calibração multivariada de óleos de Oliva

Na quantificação de contaminações de óleos de Oliva por óleos de Algodão, Canola, Girassol, Milho e Soja, os modelos de calibração multivariada por PLS e seleção de variáveis relacionando os espectros vibracionais com as quantidades desses contaminantes, apresentaram resultados satisfatórios com baixos erros de previsão.

O método GA-PLS foi o mais eficiente na quantificação dos contaminantes Canola e Milho por espectroscopia no infravermelho próximo e Girassol por espectroscopia no infravermelho médio, com média de erro relativo de 0,20, 1,42 e 0,35% respectivamente, mostrando que o algoritmo genético eliminou as variáveis que não são necessárias para o desenvolvimento dos modelos de calibração.

Em todos os outros modelos de calibração multivariada, qualquer que seja o contaminante ou técnica vibracional, o método GA-iPLS foi o mais eficiente com média de erro relativo de 0,04 a 0,98%.

Na maioria dos casos estudados, os métodos de seleção de variáveis foram eficientes na eliminação de variáveis não informativas melhorando a previsão de amostras de validação, comparado ao desempenho do modelo de calibração multivariada por PLS.

Foi observado que a região no infravermelho médio relacionada à deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados), contém as variáveis mais relevantes para a construção dos modelos de calibração. Na literatura, é relatado o uso de toda a região de impressão digital para a quantificação desses contaminantes, o que nesse trabalho verificou ser menos eficiente pois nem todas essas variáveis contém informação importante.

Também foi observado, que as variáveis mais relevantes para a construção dos modelos de calibração, no infravermelho próximo estão na região relacionada a estiramento de ligações C-H de grupos CH₂ (1^a overtone) e no Raman estão na região relacionada a estiramento de ligações C-C.

A espectroscopia vibracional aliada à quimiometria, principalmente aos métodos de calibração multivariada e seleção de variáveis, mostraram promissoras em relação às análises de determinação de contaminações em óleos de Oliva, podendo ser uma metodologia viável às análises clássicas.

Conclusão

De acordo com as comparações realizadas e resultados apresentados, constatou-se que os modelos de calibração multivariada por PLS e seleção de variáveis relacionando os espectros vibracionais com parâmetros de qualidade de Lubrificantes e óleos de Oliva, apresentaram valores de erros satisfatórios, robustez e eficiência na previsão das amostras de validação.

Em relação aos parâmetros de qualidade de lubrificantes constatou-se que o método de calibração por PLS foi mais robusto e eficiente na quantificação do parâmetro TBN e dos contaminantes água e etileno glicol. O método GA-PLS foi o mais eficiente na quantificação do parâmetro Viscosidade a 100°C e o método GA-iPLS na quantificação dos contaminantes diesel e gasolina. Em que foram selecionadas variáveis na região relacionada a compostos aromáticos presentes no diesel e gasolina como benzeno, tolueno, etilbenzeno, xileno e propilbenzeno, representando uma importante melhora na qualidade da informação química.

Para a quantificação de contaminações de óleos de Oliva por óleos de Algodão, Canola, Girassol, Milho e Soja constatou-se que o método GA-PLS foi o mais eficiente na quantificação dos contaminantes Canola e Milho por espectroscopia no infravermelho próximo e Girassol por espectroscopia no infravermelho médio. Em todos os outros modelos de calibração multivariada, qualquer que seja o contaminante ou técnica vibracional, o método GA-iPLS foi o mais eficiente.

Foi observado que a região no infravermelho médio relacionada à deformação de ligações C-H fora do plano (*trans* de dupla ligação) e o início da região de impressão digital (*trans/trans* e *cis/trans* de dienos conjugados), contém as variáveis mais relevantes para a construção dos modelos de calibração. Recordando que os óleos adulterantes exceto o Canola são predominantemente poliinsaturados e o óleo de Oliva monoinsaturado, as transições vibracionais nessa região só podem advir da presença de contaminação no óleo de Oliva. Também foi observado, que as variáveis mais relevantes para a construção dos modelos de calibração, no infravermelho próximo estão na região relacionada a

estiramento de ligações C-H de grupos CH_2 (1^a overtone) e no Raman estão na região relacionada a estiramento de ligações C-C, o que pode levar a estudos sobre as cadeias carbônicas de triacilglicerídeos presentes nos óleos e gorduras.

Lembrando que modelos que relacionam medida instrumental com medida de análises clássicas podem, conforme a faixa de erros admitida, ser utilizados em laboratórios de controle de qualidade reduzindo tempo, custo de reagentes, quantidade de resíduos, mão de obra, entre outros. E mediante a melhoria na qualidade da informação química, as metodologias desenvolvidas neste trabalho, apresentam como uma alternativa viável no controle de qualidade de lubrificantes e óleos de Oliva.

Finalmente, acreditamos estar contribuindo para a consolidação da Quimiometria no país, disponibilizando o conhecimento adquirido a todos que se interessarem no uso de métodos quimiométricos para solucionar problemas de natureza química e/ou de áreas afins, principalmente no que tange as aplicações de métodos de calibração multivariada e seleção de variáveis.

REFERÊNCIAS BIBLIOGRÁFICAS

Referências Bibliográficas

- [1] Skoog, D. A.; Holler, F. J.; Nieman, T. A. *Princípios de análise instrumental*. 5. ed. São Paulo: Bookman, 2002.
- [2] Schrader, B. *Infrared and Raman Spectroscopy: methods and applications*. Weinheim: VCH, 1995.
- [3] Burns, D. A.; Ciurczak, E. W. *Handbook of Near-Infrared Analysis*. New York: Marcel Dekker, 2001.
- [4] Coates, J. A review of current new technology: Used in instrumentation for industrial vibrational spectroscopy, *Spectroscopy*, 14, 10, 20-34, 1999.
- [5] Workman Jr, J. J. Interpretative spectroscopy for near infrared, *Appl. Spectrosc. Rev.*, 31, 251, 1996.
- [6] Stuart, A. D.; Trotman, S. M.; Doolan, K. J.; Fredericks, P. M. Spectroscopy measurement of used lubricating oil quality; *Appl. Spectrosc.*, 43, 55-60, 1989.
- [7] Williams, P.; Norris, K. *Near-Infrared technology in the agricultural and food industries*. Minnesota: AACC, 1998.
- [8] Harrick, N.J. Multiple Reflection Cells for Internal Reflection Spectrometry, *Anal. Chem.*, 36, 188-191, 1964.
- [9] Baeten, V.; Meurens, M.; Morales, M. T.; Aparicio, R. Detection of Virgin Olive Oil Adulteration by Fourier Transform Raman Spectroscopy, *J. Agric. Food Chem.*, 44, 2225-2230, 1996.
- [10] Borges Neto, W. Aplicação de métodos de Reconhecimento de Padrões para classificação de Óleos Vegetais. Goiânia, 2001. 74p. Dissertação – Instituto de Química, Universidade Federal de Goiás.
- [11] Otto, M. *Chemometrics – Statistics and Computer application in Analytical Chemistry*, Weinheim: Wiley, 1999.
- [12] Bruns, R. E.; Faigle, J. F. G. Quimiometria, *Quim. Nova*, 8, 84-99, 1995.
- [13] Wold, S.; Sjöström, M. Chemometrics, present and future sucess, *Chemom. Intell. Lab Syst.*, 44, 3-14, 1998.
- [14] Seasholtz, M. B. Making money with chemometrics, *Chemom. Intell. Lab. Syst.*, 45, 55-63, 1999.

- [15] Miller, J. N.; Miller, J. C. *Statistics and Chemometrics for Analytical Chemistry*. London: Prentice Hall, 2000.
- [16] Vandeginste, B. G. M.; Massart, D. L.; Buydens, L. M. C.; Jing, S.; Lewi, P. J.; Smeyers-Verbeke, J. *Handbook of Chemometrics and Qualimetrics: Part B*. Amsterdam: Elsevier, 1998.
- [17] Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A practical guide*. New York: Wiley, 1998.
- [18] Brereton, R. G. Introduction to multivariate calibration in analytical chemistry, *Analyst*, 125, 2125, 2000.
- [19] Helland, I. S. On the structure of partial least square regression, *Communications in statistics – simulation and computation*. 17, 2, 1998.
- [20] Martens, H.; Naes, T. *Multivariate calibration*. New York: Wiley, 1996.
- [21] Sekulic, S.; Seasholtz, M. B.; Wang, Z.; Kowalski, B. R. Nonlinear multivariate calibration methods in analytical chemistry, *Anal. Chem.*, 65, 19, A835-A845, 1993.
- [22] Charnet, R.; De Luna Freire, C. A.; Charnet, E. M. R.; Bovino, H. *Análise de Modelos de Regressão Linear com Aplicações*. Campinas: Unicamp, 1999.
- [23] Chui, Q. S. H.; Zucchini, R. R.; Lichtig, J. Qualidade de medições em química analítica. Estudo de caso: determinação de cádmio por espectrofotometria de absorção atômica com chama, *Quim. Nova*, 24, 2001.
- [24] Beebe, K. R.; Kowalski, B. R. An Introduction to Multivariate Calibration and Analysis, *Anal. Chem.*, 59, 17, 1007A-1017A, 1987.
- [25] Geladi, P.; Kowalski, B. R. Partial Least Squares Regression: A Tutorial, *Anal. Chim. Acta*, 1-17, 1986.
- [26] Abrahamsson, C. et al. Comparison of different variable selection methods conducted on NIR transmission measurements on intact tablets, *Chemom. Intell. Lab. Syst.*, 69, 3-12, 2003.
- [27] Norgaard, L.; Saudland, A.; Wagner, J.; Nielsen, J. P.; Munck, L.; Engelsen, S. B. Interval partial least-square regression (iPLS): A comparative chemometric study with an example from near-infrared spectroscopy, *Appl. Spectrosc.*, 54, 3, 413-419, 2000.

- [28] Oliveira, F. C.; Souza, A. T. P. C.; Dias, J. A.; Dias, S. C. L.; Rubim, J. C. A escolha da faixa espectral no uso combinado de métodos espectroscópicos e quimiométricos, *Quim. Nova*, 27, 2, 2004.
- [29] Osborne, S. D.; Jordan, R. B.; Künnemeyer, R. Method of wavelength selection for partial least square, *Analyst*, 122, 1997.
- [30] Centner, V.; Massart, D. Elimination of uninformative variables for multivariate calibration, *Anal. Chem.*, 68, 21, 3851-3858, 1996.
- [31] Costa Filho, P. A.; Poppi, R. J. Algoritmo genético em química, *Quim. Nova*, 22, 3, 405-411, 1999.
- [32] Zupan, J.; Gasteiger, J. *Neural Networks for Chemistry: an introduction*. Weinheim: VCH, 1993.
- [33] Rohleder, D.; Kiefer, W.; Petrich, W. Quantitative analysis of serum ultrafiltrate by means of Raman spectroscopy, *Analyst*, 129, 906-911, 2004.
- [34] AGIP PETROLI – *Apostila Lubrificante e Lubrificação*, ATEL, 2000.
- [35] Carreteiro, R. P.; Moura, C. R. S. *Lubrificantes e Lubrificação*. São Paulo: Makron Books, 1998.
- [36] Rizvi, S. Q. A. Additives for automotive fuels and lubricants, *Lubric. Eng.*, 33-39, 1999.
- [37] Hilsdorf, J. W.; Reitano, M. F.; Barros, N. D. *Química Tecnológica*, São Bernardo do Campo: Faculdade de Engenharia Industrial, 1993.
- [38] Wiseman M.; Ah-Sue, A. Monitoring oil degradation using FTIR analysis, *Lubric. Eng.*, 236-241, 1992.
- [39] Powell, J. R.; Compton, D. A. C. Automated FTIR spectrometry for monitoring hydrocarbon-based engine oils, *Lubric. Eng.*, 233-239, 1993.
- [40] Dong, J.; Voort et al. Rapid determination of the carboxylic acid contribution to the total acid number of lubricants by Fourier transform infrared spectroscopy, *Lubric. Eng.*, 12-20, 2000.
- [41] Nakanishi, H.; Onodera, K. oxidation stability of synthetic lubricants, *Lub. Eng.*, 29-37, 1997.
- [42] Standard Test Method for Gasoline Diluent in Used Gasoline Engine Oils by Gas Chromatography, ASTM D3525-90, Philadelphia: ASTM, 1990.

- [43] Standard Test Method for Determination of Water in Liquid Petroleum Products by Karl Fisher Reagent, ASTM D1744-99, Philadelphia: ASTM, 1999.
- [44] Standard Test Method for Detecting Glycol-Base Antifreeze in Used Lubricating Oils, ASTM D2982-98, Philadelphia: ASTM, 1998.
- [45] Blanco, M.; Coelho, J.; Iturriaga, H.; MasPOCH, S.; González, R. Determination of Water in Lubricating Oils by Mid- and Near-Infrared Spectroscopy, *Mikrochim. Acta*, 128, 235-239, 1998.
- [46] Paschoal, J.; Barboza, F. D.; Poppi, R. J. Analysis of contaminants in lubricant oil by near infrared spectroscopy and interval partial least-squares, *J. Near Infrared Spectrosc.*, 11, 211-218, 2003.
- [47] Borin, A.; Poppi, R. J. Multivariate Quality Control of Lubricating Oils Using Fourier Transform Infrared Spectroscopy, *J. Braz. Chem. Soc.*, 15, 4, 570-576, 2004.
- [48] Borin, A.; Poppi, R. J. Application of mid infrared spectroscopy and iPLS for the quantification of contaminants in lubricating oil, *Vib. Spectrosc.*, 37, 27-32, 2005.
- [49] Dong, J.; Voort, F. R. V.; Ismail, A. A.; Koblé, E. A.; Pinchuk, D. Rapid determination of the carboxylic acid contribution to the total acid number of lubricants by Fourier transform infrared spectroscopy, *Lubric. Eng.*, 12-20, 2000.
- [50] Standard Test Method for Base Number Determination by Potentiometric Titration, ASTM D4739, Philadelphia: ASTM, 1995.
- [51] Standard Test Method for Kinematic Viscosity of Transparent and Opaque Líquids, ASTM D445, Philadelphia: ASTM, 2001.
- [52] Coates, J. Interpretation of Infrared Spectra, A Practical Approach, *Encyclopedia of Anal. Chem.*, Chichester: Wiley, 10815-10837, 2000.
- [53] Luinge, H.J.; Van der Maas, J.H.; Visser, T. Partial least squares regression as a multivariate tool for the interpretation of infrared spectra, *Chemom. Intell. Lab. Syst.*, 28, 129-138, 1995.
- [54] Aued-Pimentel, S. et al. Azeite de oliva: incidência de adulterações entre os anos de 1993 a 2000, *Rev. Inst. Adolfo Lutz*, 61, 2, 69-75, 2002.

- [55] Harwood, J. L.; Yaqoob, P. Nutritional and health aspects of olive oil, *Eur. J. Lipid Sci. Technol.*, 104, 685-697, 2002.
- [56] Ulberth, F.; Buchgraber, M. Authenticity of fats and oils, *Eur. J. Lipid Sci. Technol.*, 102, 687-694, 2000.
- [57] Firestone, D. Assuring the Integrity of Olive Oil Products, *J. AOAC Int.*, 84, 1, 176-180, 2001.
- [58] Tay, A. et al. Authentication of Olive Oil Adulterated with Vegetable Oils Using Fourier Transform Infrared Spectroscopy, *Lebensm.-Wiss. u.-Technol.*, 35, 99-103, 2002.
- [59] Lai, Y. W.; Kemsley, E. K.; Wilson, R. H. Quantitative analysis of potential adulterants of extra virgin olive oil using infrared spectroscopy, *Food Chem.*, 95-98, 1995.
- [60] Downey, G.; McIntyre, P.; Davies, A. N. Detecting and Quantifying Sunflower Oil Adulteration in Extra Virgin Olive Oils from the Eastern Mediterranean by Visible and Near-Infrared Spectroscopy, *J. Agric. Food Chem.*, 50, 5520-5525, 2002.
- [61] Peña, F.; Cárdenas, S.; Gallego, M.; Valcárcel, M. Direct olive oil authentication: Detection of adulteration of olive oil with hazelnut oil by direct coupling of headspace and mass spectrometry, and multivariate regression techniques, *J. Chromatogr., A*, 1074, 215-221, 2005.
- [62] Dobson, G. Spectroscopy and spectrometry of lipids – Part 1, *Eur. J. Lipid Sci. Technol.*, 103, 815-840, 2001.
- [63] Dahlberg, D. B.; Lee, S. M.; Wenger, S. J.; Vargo, J. A. Classification of Vegetable Oils by FT-IR, *Appl. Spectrosc.*, 51, 8, 1118-1124, 1997.
- [64] Baeten, V.; Aparicio, R. Edible oils and fats authentication by Fourier transform Raman spectrometry, *Biotechnol. Agron. Soc. Environ.*, 4, 196-203, 2000.
- [65] Muik, B. et al. Direct, reagent-free determination of free fatty acid content in olive oil and olives by Fourier transform Raman spectrometry, *Anal. Chim. Acta*, 211-220, 2003.
- [66] Tapp, H. S.; Defernez, M.; Kemsley, E. K. FTIR Spectroscopy and Multivariate Analysis Can Distinguish the Geographic Origin of Extra Virgin Olive Oils, *J. Agric. Food Chem.*, 51, 6110-6115, 2003.

- [67] Barthus,R.C.; Poppi,R.J. Determination of the total unsaturation in vegetable oils by Fourier transform Raman spectroscopy and multivariate calibration, *Vib. Spectrosc.*, 26, 99-105, 2001.
- [68] Barthus,R.C.; Poppi,R.J. Multivariate quality control applied to detect the soybean oil oxidation using Fourier transform Infrared spectroscopy, *Spectrosc. Lett.*, 35, 5, 729-739, 2002.
- [69] Ozen, B. F.; Mauer, L. J. Detection of Hazelnut Oil Adulteration Using FT-IR Spectroscopy, *J. Agric. Food Chem.*, 50, 3898-3901, 2002.
- [70] López-Díez, E. C.; Bianchi, G.; Goodacre, R. Rapid Quantitative Assessment of the Adulteration of Virgin Olive Oils with Hazelnut Oils Using Raman Spectroscopy and Chemometrics, *J. Agric. Food Chem.*, 51, 6145-6150, 2003.
- [71] Ozen, B. F.; Weiss, I.; Mauer, L. J. Dietary Supplement Oil Classification and Detection of Adulteration Using Fourier Transform Infrared Spectroscopy, *J. Agric. Food Chem.*, 51, 5871-5876, 2003.
- [72] Christy, A. A.; Kasemsumran, S.; Du, T.; Ozaki, Y. The Detection and Quantification of Adulteration in Olive Oil by Near-Infrared Spectroscopy and Chemometrics, *Anal. Sci.*, 20, 935-940, 2004.