



**Universidade Estadual de Campinas
Instituto de Química
Programa de Pós-Graduação em Química**

Tese de Doutorado:

“Desenvolvimento de técnicas quimiométricas de compressão de dados e de redução de ruído instrumental aplicadas a óleo diesel e madeira de eucalipto usando espectroscopia NIR”.

Heronides Adonias Dantas Filho

Orientador: Prof. Dr. Celio Pasquini

Co-orientador: Prof. Dr. Mário César Ugulino de Araújo (DQ-UFPB)

Campinas, Março de 2007

FICHA CATALOGRÁFICA ELABORADA PELA BIBLIOTECA DO INSTITUTO DE
QUÍMICA DA UNICAMP

D235d	<p>Dantas Filho, Heronides Adonias. Desenvolvimento de técnicas quimiométricas de compressão de dados e de redução de ruído instrumental aplicadas a óleo diesel e madeira de eucalipto usando espectroscopia NIR / Heronides Adonias Dantas Filho. -- Campinas, SP: [s.n], 2007.</p> <p>Orientador: Celio Pasquini. Co-orientador: Mário César Ugulino de Araújo.</p> <p>Doutorado - Universidade Estadual de Campinas, Instituto de Química.</p> <p>1.Compressão de dados. 2. Ruído instrumental. 3. Madeira de eucalipto. 4. Espectroscopia NIR. I. Pasquini, Celio. II. Araújo, Mário César Ugulino. III. Universidade Estadual de Campinas. Instituto de Química. IV. Título.</p>
-------	--

Título em inglês: Development of chemometric technics for data compression and reduction of instrumental noise applied to diesel oil and eucalyptus wood employing NIR spectroscopy.

Palavras-chaves em inglês: Data compression, Instrumental noise, Eucalyptus wood, NIR spectroscopy

Área de concentração: Química Analítica

Titulação: Doutor em Ciências

Banca examinadora: Prof. Dr. Celio Pasquini (orientador), Prof. Dr. Roberto Kawakami Harrop Galvão (Divisão de Engenharia Eletrônica-ITA), Prof. Dr. Ronei Jesus Poppi (IQ-Unicamp), Prof. Dr. Roy Eduard Bruns (IQ-Unicamp), Prof. Dr. Marco Flores Ferrão (DQF-UNISC)

Data de defesa: 16/03/2007

Dedicatória

Só eles sabem e principalmente sentem a satisfação de ter um filho e irmão conquistando mais um importante degrau da ascensão profissional. Dedico a conclusão desta tese aos meus irmãos: Jackeline, José Paulo e Célia e à outra pessoa fundamental em toda a nossa vida, Maria do Céu Medeiros Dantas: minha mãe.

Agradecimentos

- A Deus, pela origem da minha vida e acompanhamento por toda a minha jornada e por ter fornecido a luz essencial para a superação de mais uma batalha.
- A toda minha família, pelo apoio, incentivo e por me fazer crer que é um dos elementos essenciais para o sucesso de um ser humano, os outros dois são: Deus e os amigos.
- Ao meu pai, Heronides Adonias Dantas (*in memoriam*).
- Ao prof. Dr. Celio Pasquini, pela frutífera oportunidade de trabalho, amizade, paciência, confiança e orientação segura. Também o considero um genial professor-pesquisador, pelas idéias sempre criativas e também como cidadão exemplar.
- Ao prof. Dr. Mário César Ugulino de Araújo da Universidade Federal da Paraíba - UFPB, pelos conselhos, contribuições, amizade e co-orientação, bem como à prof^a Dr^a Teresa Saldanha, do mesmo laboratório da UFPB.
- Ao prof. Dr. Roberto Kawakami do ITA, pela indispensável contribuição e discussões sempre valiosas em todo o andamento deste trabalho. Também outro profissional exemplar.
- À Kelly Fernandes pela compreensão, amor, apoio e companheirismo nos momentos difíceis e nos que proporcionaram felicidade também. Nas circunstâncias em que a razão não foi suficiente, ela teve fundamental importância.
- Aos professores Jarbas e Ivo pela presença amigável e com produtivas discussões dos trabalhos durante o convívio no laboratório I-144 e também por constituir um grupo de pesquisa coeso.
- Aos colegas do GIA: Fabiano Gonzaga, Alexandre Fonseca, Alexandre Sacorague (Xaolin), Claudete Fernandes, Fernando Yonehara, Ismael Chagas, Juliana Cortez, Emerson Aquino, Márcia Breitzkreitz, Lucas Mariano, Pedro Paro, Giovana Povia e todos os demais integrantes do laboratório que de alguma forma contribuíram para o ambiente sempre saudável, produtivo e alegre, condições indispensáveis para a execução de um bom trabalho.
- A Leonardo Chagas, pelas valiosas discussões sobre os dados de madeira de eucalipto.

- Aos funcionários das empresas de Papel e Celulose que disponibilizaram os dados de madeira de eucalipto.
- Aos colegas do Laboratório de Química Ambiental (Lab. I-155) pelo convívio durante o trabalho de doutorado.
- Aos funcionários da CPG, em especial a Bel, Rodrigo e Elias.
- Aos colegas e amigos do Laboratório de Instrumentação e Automação em Química Analítica e Quimiometria (LAQA) da UFPB: Sérgio Ricardo, Edvaldo Gaião, Ricardo Alexandre, Gledson Emídio, Simone Simões, Alessandra Félix, Márcio Coelho (pela força na finalização da tese), Edvan Cirino, Wallace Fragoso, Glauciene Paula, Sueni, Socorro, Osmundo, Amália, Luciano Farias, Elaine, os quais tenho enorme amizade, vínculo profissional e gratidão.
- Aos amigos da república da Rua Jerônimo Pattaro, 219, em especial a Eduardo Marafon, Tiago Germano, Edson Filho, Rodrigo Possari, Mário Killner, Sebastião Milton, Márcio Cataia, Kristerson e Ricardo.
- Aos funcionários e técnicos do laboratório de ensino da época do estágio docente, em especial à Mirian, Divino, Mauro, Érico.
- À Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), por promover as condições necessárias à execução deste trabalho de doutorado e pela concessão do auxílio financeiro dedicado a este projeto. Um modelo de Fundação de amparo à pesquisa que deve ser tomado como referência ao longo de todo o Brasil.
- Àqueles que devem entender perfeitamente que essa etapa da tese não é tão simples como se imagina, por não ter identificado os seus nomes nesses agradecimentos, mas são conscientes do quanto contribuíram, direta ou indiretamente para a conclusão desta obra.

Curriculum Vitae Resumido

1 - Formação Acadêmica

- 1.1- 2001 - Graduação em Química Industrial – Universidade Federal da Paraíba.
- 1.2- 2003 – Mestrado em Química - Universidade Federal da Paraíba.
- 1.3- 2007 – Doutorado em Química – Universidade Estadual de Campinas.

2 – Artigos publicados em revistas internacionais (5 últimos)

2.1 - Breitzkreitz, M. C., Raimundo Jr. I. M., Rohwedder, J. J. R., Pasquini, C., Dantas Filho, H. A., José, G. E., Araújo, M. C. U., *Determination of Total Sulphur in Diesel Fuel Employing NIR Spectroscopy and Multivariate Calibration*, **Analyst**, 128 (2003) 1204 – 1207;

2.2 - Galvão, R. K. H., José, G. E., Dantas Filho, H. A., Araújo, M. C. U., Silva, E. C., Paiva, H. M., Saldanha, T. C. B., *Optimal Wavelet Filter Construction using X and Y data*, **Chemometrics and Intelligent Laboratory Systems**, 70 (2004) 1 – 10.

2.3 - Dantas Filho, H. A., Galvão, R. K. H., Araújo, M. C. U., Silva, E. C., Saldanha, T. C. B., José, G. E., Pasquini, C., Raimundo Jr. I. M., Rohwedder, J. J. R., *A Strategy for Selecting Calibration Samples for Multivariate Modelling*, **Chemometrics and Intelligent Laboratory Systems**, 72 (2004) 83 – 91.

2.4 – Dantas Filho, H. A., Souza, E. S. O. N., Visani, V., Barros, S. R. R. C., Saldanha, T. C. B., Araújo, M. C. U., Galvão., R. K. H., *Simultaneous Spectrometric Determination of Cu^{2+} , Mn^{2+} and Zn^{2+} in Polivitaminic/Polimineral Drug using SPA and GA Algorithms for Variable Selection*, **Journal of the Brazilian Chemical Society**, 16 (2005) 58 – 61.

2.5 – Galvão, R. K. H., Dantas Filho, H. A., Martins, M. N., Araújo, M. C. U., Pasquini, C., *Sub-optimal Wavelet Denoising of Coaveraged Spectra Employing Statistics from Individual Scans*, **Analytica Chimica Acta**, 67 (2007) 159 – 167.

3 – Trabalhos apresentados em Congressos

3.1 - Breitzkreitz, M. C., Raimundo Jr. I. M., Rohwedder, J. J. R., Pasquini, C., Dantas Filho, H. A., José, G. E., Araújo, M. C. U., *Determination of Total Sulphur in Diesel Fuel Employing NIR Spectroscopy and Multivariate Calibration*, **11th International Conference on Near Infrared Spectroscopy**, 2003;

3.2- Dantas Filho, H. A., Galvão, R. K. H., Araújo, M. C. U., Silva, E. C., Saldanha, T. C. B., José, G. E., Pasquini, C., Raimundo Jr. I. M., Rohwedder, J. J. R., *A Strategy for Selecting Calibration Samples for Multivariate Modelling*, **9th International Conference on Chemometrics in Analytical Chemistry**, 2004;

3.3 – Dantas Filho, H. A., Soares, S. F. C., Fragoso, W. D., Araújo, M. C. U. Galvão, R. K. H., Silva, E. C., Pasquini, C. *Comparative Study of NAS-based figures of merit for MLR with variable selection and full-spectrum PLS*, **10th International Conference on Chemometrics in Analytical Chemistry**, 2006;

3.4 - Dantas Filho, H. A., Galvão, Pasquini, C., Araújo, M. C. U., *Optimal Wavelet Filtering of Coaveraged Instrumental Signals Spectra Employing Statistics from Individual Scans*, **10th International Conference on Chemometrics in Analytical Chemistry**, 2006;

4 – Experiência Profissional

4.1 - Monitor da disciplina Química Analítica Clássica (QA-282, 1º semestre de 2005) pelo Programa de Estágio Docência da Unicamp;

4.2 – Monitor da disciplina Química Analítica Clássica (QA-213, 2º semestre de 2005) pelo Programa de Estágio Docência da Unicamp;

4.3 – Monitor da disciplina Química Analítica Clássica (QA-282, 1º semestre de 2006) pelo Programa de Estágio Docência da Unicamp;

RESUMO

Título: DESENVOLVIMENTO DE TÉCNICAS QUIMIOMÉTRICAS DE COMPRESSÃO DE DADOS E DE REDUÇÃO DE RUÍDO INSTRUMENTAL APLICADAS A ÓLEO DIESEL E MADEIRA DE EUCALIPTO USANDO ESPECTROSCOPIA NIR”.

Autor: Heronides Adonias Dantas Filho

Orientador: Prof. Dr. Celio Pasquini

Co-orientador: Prof. Dr. Mário César Ugulino de Araújo.

Neste trabalho foram desenvolvidas e aplicadas técnicas de seleção de amostras e de variáveis espectrais para calibração multivariada a partir do Algoritmo das Projeções Sucessivas (APS). Também foi utilizada a transformada wavelet para resolver problemas de redução de ruído associado a dados de espectroscopia NIR (Infravermelho Próximo), na construção de modelos de calibração multivariada baseados em Regressão Linear Múltipla (MLR) para estimativa de parâmetros de qualidade de óleo diesel combustível e também de madeira de eucalipto. Os espectros NIR de transmitância para óleo diesel e de reflectância para madeira de eucalipto foram registrados empregando-se um equipamento NIR-Bomem com detector de Arseneto de Gálio e Índio. Para a aplicação em óleo diesel, foram estudadas as regiões espectrais: 850 – 1.100 nm, 1.100 – 1.570 nm e 1.570 - 2.500 nm. Para as amostras de madeira de eucalipto foi empregada a região de 1.100 – 2.500 nm. Os resultados do uso de técnicas de seleção de variáveis e amostras por MLR comprovaram sua simplicidade frente os modelos de regressão por mínimos quadrados parciais (PLS) que empregam toda a região espectral e transformação em variáveis latentes e são mais complexos de interpretar. O emprego de seleção de amostras demonstrou ainda a possibilidade de procedimentos de recalibrações e transferência de calibração que utilizam um número reduzido de amostras, sem a perda significativa da capacidade preditiva dos modelos MLR. O uso de filtragem wavelet também teve sua eficiência comprovada tanto no contexto da calibração multivariada quanto na filtragem de espectros NIR a partir de varreduras individuais. Na maioria dos casos de que trata esta tese e por consequência das técnicas quimiométricas empregadas, melhorias quanto à minimização do erro (RMSEP) associado à quantificação dos parâmetros de qualidade, bem como redução do tempo empregado na aquisição de varreduras de espectros NIR foram as principais contribuições fornecidas.

ABSTRACT

Title: “DEVELOPMENT OF CHEMOMETRIC TECHNIQUES FOR DATA COMPRESSION AND REDUCTION OF INSTRUMENTAL NOISE APPLIED TO DIESEL OIL AND EUCALYPTUS WOOD USING NIR SPECTROSCOPY”.

Author: Heronides Adonias Dantas Filho

Advisor: Prof. Dr. Celio Pasquini

Co-advisor: Prof. Dr. Mário César Ugulino de Araújo.

This work describes two techniques for spectral variable and sample selection based on the Successive Projections Algorithm (SPA), aiming the construction of multivariate regression models. Also, the wavelet transform was employed to solve problems related to noise reduction associated with spectroscopic data in the near infrared spectral region (NIR), and employed in the construction of multivariate calibration models based in Linear Multiple Regression (LMR) to estimate the quality parameters of diesel fuel and eucalyptus wood. The NIR transmission spectra for diesel samples and the reflectance spectra obtained for wood samples were acquired by using a NIR-Bomen equipment with AsGaIn detector. For application in diesel, the following spectral regions have been investigated: 850 – 1100 nm, 1100 – 1570 nm and 1570 – 2500 nm. For wood samples the spectral region employed was from 1100 – 2500 nm. The results obtained by using the variable selection techniques and LMR demonstrate their simplicity when compared with its counterpart Partial Least Square (PLS) which employs full spectral region and latent variables, being, therefore, more difficult to be interpreted. The use of wavelet filtering also demonstrates its efficiency both for multivariate calibration and NIR spectral data filtering. In most of the cases approached in this work, and in consequence for the chemometric techniques employed, improvements in the error (RMSEP) associated with the quality parameters as well a decrease in the analysis time were the main achievements of this work.

Índice

Lista de Abreviaturas	xvii
Lista de Tabelas	xix
Lista de Figuras	xxv
1 Introdução	01
1.1 Métodos de Seleção de Variáveis	08
1.1.1 O Algoritmo das Projeções Sucessivas (APS) para seleção de variáveis.....	10
1.1.2 O APS empregado em seleção de amostras	12
1.2 Técnicas de Filtragem de Sinais Instrumentais.	13
1.2.1 Método da Transformada de Fourier (FT)	15
1.2.2 Método da Transformada Wavelet (WT)	17
1.2.3 Filtragem Wavelet de espectros NIR médios a partir de varreduras individuais.....	22
1.3 Espectroscopia NIR	23
1.3.1 Análise NIR de Óleo Diesel	25
1.3.2 Análise NIR de Madeira de Eucalipto	29
2 Objetivos	33
3 Experimental	37
3.1 Materiais e Métodos	39
3.2 Equipamentos e Acessórios	39
3.2.1 Equipamento e Sistema de registro dos espectros	39
3.2.2 Regiões NIR estudadas em função das cubetas empregadas	40
3.3 Software e Hardware	46
3.4 Calibração Multivariada	47
3.4.1 Pré-Tratamento dos Dados	47
3.4.2 Conjunto de amostras	48
4 Resultados e Discussão	51
4.1 Modelos de calibração por Mínimos Quadrados Parciais (PLS-1)	53
4.2 Modelos de calibração por Regressão Linear Múltipla (MLR)	56
4.2.1 Modelos MLR com Seleção de variáveis por APS	56
4.2.2 Modelos MLR com Seleção de variáveis e amostras de Calibração por APS.....	58
4.2.3 Seleção de amostras e corte de árvores de Madeira de eucalipto	60
4.2.4 Filtragem Wavelet e calibração multivariada por MLR	66
4.2.5 Filtragem Wavelet de espectros NIR médios a partir da estatística de varreduras individuais.....	71
5 Conclusões	75
6 Referências Bibliográficas	79
7 Apêndice	89

Lista de Abreviaturas

MLR	Regressão Linear Múltipla.
PCR	Regressão em Componentes Principais
PLS	Regressão por Mínimos Quadrados Parciais.
APS	Algoritmo das Projeções Sucessivas.
NIR	Infravermelho Próximo
GA	Algoritmo Genético
UV-VIS	Ultravioleta – Visível
RMSEP	Erro Quadrático Médio de Previsão
LED	Diodo Emissor de Luz
FT	Transformada de Fourier
FTI	Transformada de Fourier Inversa
WT	Transformada Wavelet
DWT	Transformada Wavelet Discreta
SNR	Razão Sinal/Ruído
RMSE	Erro Quadrático Médio de Reconstrução
GLP	Gás Liquefeito de Petróleo
KS	Kennard - Stone
KSX	Kennard – Stone baseado nas distâncias em X
AEAC	Álcool Etílico Anidro Carburante
ASTM	American Society for Testing and Materials
InGaAs	Arseneto de Gálio e Índio
V.L	Variáveis Latentes
Repro.	Reprodutibilidade dos métodos de referência
S.G.	Savitzky-Golay
selmist	Algoritmo para Seleção de Amostras
rlmqr	Regressão Linear Múltipla com decomposição QR

Lista de Tabelas

Tabela 1	Uso da espectroscopia NIR em análises químicas.....	25
Tabela 2	Composição química aproximada das madeiras.....	29
Tabela 3	Caminhos ópticos das celas de fluxo empregadas e faixa espectral obtida com o registro dos espectros com o uso das respectivas células.....	41
Tabela 4	Unidades de medida e os diversos Parâmetros de Qualidade estudados para as matrizes de Óleo Diesel e Madeira de Eucalipto.....	46
Tabela 5	Número de pontos à esquerda e a direita do ponto médio empregados na suavização dos espectros, em função da região espectral explorada, para as amostras de óleo diesel.....	48
Tabela 6	Número de amostras de óleo diesel em cada um dos conjuntos de calibração, validação e previsão, em função da região espectral e do parâmetro de qualidade empregadas nos modelos PLS e MLR.....	49
Tabela 7	Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade de óleo diesel para o sub-conjunto de amostras de previsão dos modelos PLS.....	55
Tabela 8	Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade, para o sub-conjunto de amostras de previsão dos modelos MLR, empregando seleção de variáveis, nas regiões de 850 - 1.100 nm (R.E. 1), 1.100 - 1.570 nm (R.E. 2) e 1.570 - 2.500 nm (R.E. 3).....	57
Tabela 9:	Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade, para o sub-conjunto de amostras de previsão dos modelos MLR, empregando seleção de variáveis e amostras de calibração, nas regiões de 850 - 1.100 nm (R.E. 1), 1.100 - 1.570 nm (R.E. 2) e 1.570 - 2.500 nm (R.E. 3).....	59
Tabela 10:	Características dos modelos MLR e PLS construídos para a estimativa dos parâmetros de qualidade de madeira de eucalipto.....	63
Tabela 11:	Valores de RMSEP para os modelos PLS construídos para a estimativa dos parâmetros de qualidade de madeira de eucalipto, para as amostras do conjunto de previsão.....	64
Tabela 12.	Valores de RMSEP e R (correlação) para a estimativa dos parâmetros de qualidade de óleo diesel, empregando filtragem Wavelet e modelos MLR, aplicados ao conjunto de amostras da série de previsão, com os espectros derivativos, nas regiões espectrais: 850 - 1.100 nm (R.E.1), 1.100 - 1.570 nm (R.E.2) e 1.570 - 2.500 nm (R.E.3).....	68
Tabela 13.	Valores de RMSEP dos modelos PLS com filtro de Savitzky-Golay (PLS-SG) e dos modelos MLR com filtragem Wavelet (MLR-WT), para o conjunto de amostras da série de previsão, nas regiões espectrais: 850 - 1.100 nm (R.E.1), 1.100 - 1.570 nm (R.E.2) e 1.570 - 2.500 nm (R.E.3).....	69

Tabela 14.	Valores de RMSEP dos modelos MLR com filtro polinomial por Savitzky-Golay e seleção de variáveis por APS (MLR-SG-APS) e regressão MLR com filtragem e seleção de variáveis por transformada Wavelet (MLR-WT), aplicados ao conjunto de amostras da série de previsão, com os espectros derivativos nas regiões espectrais: 850 - 1.100 nm (R.E.1), 1.100 - 1.570 nm (R.E.2) e 1.570 - 2.500 nm (R.E.3).....	69
Tabela 15.	Comparação entre RMSEP obtidos na estimativa dos parâmetros de qualidade, para o conjunto de amostras de previsão, empregando os modelos MLR, filtro polinomial por Savitzky-Golay e seleção de variáveis e amostras de calibração por APS (MLR-APS-AM) e também com emprego de regressão MLR com filtragem e seleção de variáveis por transformada Wavelet (MLR-WT), nas regiões espectrais: 850 - 1.100 nm (C.O. 50), 1.100 - 1.570 nm (C.O. 10) e 1.570 - 2.500 nm (C.O. 1).....	70
Tabela 16	Valores de RMSE ($\times 1000$) para 16 scans e diferentes filtros wavelet.....	73
Tabela B1:	Número de amostras de madeira de eucalipto em cada um dos conjuntos de calibração, validação e previsão, em função do parâmetro de qualidade empregado nos modelos PLS e MLR na região espectral de 1.100 - 2.500 nm.....	92
Tabela B2:	Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade de madeira de eucalipto para o sub-conjunto de amostras de previsão dos modelos PLS na região espectral de 1.100 - 2.500 nm.....	92
Tabela F1	Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano, usando modelos MLR, na região espectral de 850 - 1.100 nm, com caminho óptico de 50 mm.....	114
Tabela F2:	Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano, usando modelos MLR, na região espectral de 1.100 - 1.570 nm, com caminho óptico de 10 mm.....	115
Tabela F3	Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano, usando modelos MLR, na região espectral de 1.570 - 2.500 nm, com caminho óptico de 1 mm.....	116
Tabela F4	Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de madeira de eucalipto usando modelos MLR na região espectral de 1.100 - 2.500 nm.....	117
Tabela G1	Identificação das amostras de calibração com seus respectivos valores dos parâmetros de qualidade de óleo diesel (entre parênteses), selecionadas pelo APS, na região espectral de 850 - 1.100 nm, com caminho óptico de 50 mm.....	118

Tabela G2	Identificação das amostras de calibração com seus respectivos valores dos parâmetros de qualidade de óleo diesel (entre parênteses), selecionadas pelo APS, na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10 mm.....	119
Tabela G3	Identificação das amostras de calibração selecionadas pelo APS na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano, usando modelos MLR, na região espectral de 1.570 – 2.500nm, com caminho óptico de 1 mm.....	120
Tabela G4	Identificação das amostras de calibração com seus respectivos valores dos parâmetros de qualidade de madeira de eucalipto (entre parênteses), selecionadas pelo APS, na região espectral de 1.100 – 2.500 nm.....	121
Tabela J1	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.....	129
Tabela J2	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “extrativos” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.....	130
Tabela J3	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “holocelulose” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.....	131
Tabela J4	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “lignina” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.....	132
Tabela J5	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “rendimento” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.....	133
Tabela J6	Comparação entre modelos PLS-SG, MLR-WT e MLR-APS. Valores de RMSEP para a estimativa dos parâmetros de qualidade de madeira de eucalipto, para o sub-conjunto de amostras de previsão, na região espectral de 1.100 – 2.500 nm.	134
Tabela J7	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Índice de Cetano” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 –1.100 nm.....	135

Tabela J8	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 –1.100 nm.....	136
Tabela J9	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Enxofre” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 –1.100 nm.....	137
Tabela J10	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 50” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 – 1.100 nm.....	138
Tabela J11	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 85” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 – 1.100 nm.....	139
Tabela J12	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Índice de Cetano” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.....	140
Tabela J13	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.....	141
Tabela J14	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Enxofre” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.....	142
Tabela J15	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 50” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 – 1.570 nm.....	143
Tabela J16	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 85” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 – 1.570 nm.....	144

Tabela J17	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Índice de Cetano” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.....	145
Tabela J18	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.....	146
Tabela J19	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Enxofre” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.....	147
Tabela J20	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 50” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.....	148
Tabela J21	Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 85” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.....	149

Lista de Figuras

Figura 1.	Representação da calibração multivariada usada em quimiometria.....	03
Figura 2.	Regressão PCR e transformação em Componentes Principais (PC1 e PC2).....	06
Figura 3.	Regressão PLS e transformação em Variáveis Latentes (PC1 e PC2).....	07
Figura 4.	Exemplo de operações de projeções efetuadas com o uso do APS. para selecionar comprimentos de onda a partir de 5 variáveis espectrais ($J = 5$), 3 amostras de calibração ($M_{cal} = 3$), e comprimento de onda tomado como partida ($k(0) = 3$), resultado da primeira iteração: $k(1)=1$	11
Figura 5.	Fluxograma do uso do APS para seleção de variáveis e amostras para calibração multivariada.....	13
Figura 6.	Diagrama de emprego das transformadas FT e WT aplicados ao sinal original.....	14
Figura 7.	Espectro dividido em 8 regiões com o uso de janelas retangulares não-justapostas.....	16
Figura 8.	Exemplos de Wavelets dos tipos: a) Morlet e b) Chapéu Mexicano.....	18
Figura 9.	Wavelets das famílias Simlet, Daubechie e Coiflet, empregadas na decomposição dos espectros NIR de óleo diesel e madeira de eucalipto.....	18
Figura 10.	Efeito de modificações na escala.....	19
Figura 11.	Algoritmo de Decomposição em Árvore. Os coeficientes wavelet no nível m correspondem a d_m	19
Figura 12.	Três possíveis formas de se decompor o espectro em coeficientes wavelet.....	20
Figura 13.	Árvore de Reconstrução Wavelet com dois níveis.....	21
Figura 14.	a) Localização da radiação NIR no espectro eletromagnético. b) Detalhamento dos comprimentos de onda que delimitam a região NIR e sua localização entre as regiões visível e infravermelho médio.....	23
Figura 15.	Vibrações fundamentais da molécula da água e suas respectivas transmitâncias na região do infravermelho médio.....	24
Figura 16.	Amostra de óleo diesel acondicionada em recipiente de vidro (a) e as estruturas químicas dos hidrocarbonetos presentes nesse combustível (b).....	26
Figura 17.	Diversas formas estruturais da celulose: unitárias e polímero.....	30
Figura 18.	Fórmula estrutural de algumas hemiceluloses.....	30
Figura 19.	Base estrutural da macromolécula da lignina.....	31
Figura 20.	Serragem de madeira de eucalipto acondicionada em recipiente de vidro.....	32

Figura 21.	a) Região espectral de 850-1.100 nm usando cubeta de vidro de 50 mm, b) Região espectral de 1.100-1.570 nm usando cubeta de quartzo de 10 mm, c) Região espectral de 1.570-2.500 nm usando cubeta de vidro de 1 mm.....	40
Figura 22.	Sistema montado para a aquisição dos espectros das amostras de óleo diesel, usando a cubeta de fluxo de 10 mm de caminho óptico. As setas na cor amarela indicam o sentido do fluxo da alíquota de diesel.....	42
Figura 23.	Detalhe da célula de fluxo (Cub.) de 10 mm empregada no registro do espectro. As setas de cor amarela indicam o sentido do feixe de radiação NIR.....	42
Figura 24.	Detalhe do sistema com a cubeta de fluxo (Cub.) de 50 mm. As setas de cor laranja indicam o sentido do feixe de radiação NIR. As setas de cor verde indicam o sentido do fluxo da alíquota de óleo diesel.....	43
Figura 25.	Detalhe do sistema com a cubeta de fluxo (Cub.) de 1 mm. As setas de cor laranja indicam o sentido do feixe de radiação NIR. As setas de cor amarela indicam o sentido do fluxo da alíquota de óleo diesel.....	44
Figura 26.	Sistema semelhante ao existente nas empresas de celulose e papel, montado no laboratório GIA para a aquisição dos espectros de reflectância das amostras de madeira de eucalipto.....	45
Figura 27.	Espectro NIR de reflectância de madeira de eucalipto, na região espectral de 1.100 – 2.500nm.....	45
Figura 28.	Gráficos de variância residual em X (a) e em Y (b) em função das amostras de calibração (azul) e validação (vermelho) para óleo diesel. Modelo PLS construído para o parâmetro densidade na região espectral de 1.100 – 1.570 nm.....	53
Figura 29.	Gráfico dos coeficientes de regressão em função das variáveis espectrais para óleo diesel no modelo PLS aplicado ao parâmetro densidade, na região espectral de 1.100 – 1.570 nm. Variáveis destacadas em círculo foram eliminadas do modelo.....	54
Figura 30.	Gráfico dos valores previstos em função dos valores de referência, para as amostras de validação, sem (a) e com (b) otimizações de ajustes dos modelos PLS, para o parâmetro densidade. Região espectral de 1.100 – 1.570 nm; dados de Óleo Diesel.....	54
Figura 31.	Imagens representativas de uma plantação de eucalipto (a) e de árvores cortadas (b) para o processo de cozimento e branqueamento da polpa da madeira.....	60
Figura 32.	Exemplo de seleção por KS das 3 primeiras amostras de um conjunto de dados contendo 6 amostras. As amostras 1, 2 e 4 foram nessa ordem consideradas como as mais representativas em relação ao espaço amostral pesquisado, considerando as distâncias existentes entre cada uma delas.....	62
Figura 33.	(a) Gráfico dos scores (PC1 x PC2) das 4.636 árvores da matriz de dados, variância explicada de 97% nas 2 primeiras PC, (b) Scores para as 200 amostras selecionadas para possível corte, variância explicada de 84% nas 2 primeiras PC (b).....	62

Figura 34.	Tipos de família wavelet ($\Psi(t)$) empregadas na decomposição dos espectros NIR de óleo diesel e madeira de eucalipto.....	67
Figura 35.	(a) Varredura Individual, (b) Espectro médio de 100 varreduras usado como referência para cálculo do RMSE, na região de 850 – 1.300 nm.....	71
Figura 36.	RMSE em relação ao espectro de referência (100 scans) e como uma função do número de varreduras na obtenção dos espectros médios. Comparação entre os métodos: Sem filtragem, Proposto, rígido e suave.....	72
Figura A1	Principais compostos absorventes na região NIR.....	91
Figura D1	Gráficos dos valores de RMSEP versus número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis, na estimativa dos diversos parâmetros de qualide de óleo desel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 850 – 1.100 nm, com caminho óptico de 50 mm.....	106
Figura D2	Gráficos dos valores de RMSEP versus número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis, na estimativa dos diversos parâmetros de qualide de diesel interior e metropolitano para as amostras do conjunto de previsão, na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10 mm.....	107
Figura D3.	Gráficos dos valores de RMSEP versus número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis, na estimativa dos diversos parâmetros de qualide de diesel interior e metropolitano para as amostras do conjunto de previsão, na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1 mm.....	108
Figura D4.	Gráficos dos valores de RMSEP versus número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis, na estimativa dos diversos parâmetros de madeira de eucalipto para as amostras do conjunto de previsão, na região espectral de 1.100 – 2.500 nm.....	109
Figura E1.	Gráficos dos valores de RMSEP versus número de amostras de calibração empregados nos modelos MLR com seleção de variáveis espectrais e amostras de calibração, na estimativa dos diversos parâmetros de qualide de diesel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 850 – 1.100 nm, com caminho óptico de 50mm.....	110
Figura E2.	Gráficos dos valores de RMSEP versus número de amostras de calibração empregados nos modelos MLR com seleção de variáveis espectrais e amostras de calibração, na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10mm.....	111

Figura E3.	Gráficos dos valores de RMSEP versus número de amostras de calibração empregados nos modelos MLR com seleção de variáveis espectrais e amostras de calibração, na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1mm.....	112
Figura E4.	Gráficos dos valores de RMSEP versus número de amostras de calibração empregados nos modelos MLR com seleção de variáveis espectrais e amostras de calibração na estimativa dos diversos parâmetros de qualide de madeira de eucalipto para as amostras do conjunto de previsão na região espectral de 1.100 – 2.500 nm.....	113
Figura H1.	Gráficos dos valores previstos versus valores de referência para as amostras do conjunto de previsão, empregando modelos MLR com seleção de variáveis e amostras de calibração na estimativa dos diversos parâmetros de qualide de diesel interior, na região espectral de 850 – 1.100 nm, com caminho óptico de 50mm.....	122
Figura H2.	Gráficos dos valores previstos versus valores de referência para as amostras do conjunto de previsão, empregando modelos MLR com seleção de variáveis e amostras de calibração na estimativa dos diversos parâmetros de qualide de diesel interior, na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10mm.....	123
Figura H3.	Gráficos dos valores previstos versus valores de referência para as amostras do conjunto de previsão, empregando modelos MLR com seleção de variáveis e amostras de calibração na estimativa dos diversos parâmetros de qualide de diesel interior, na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1mm.....	124

Capítulo 1

Introdução

“Nunca ande pelo caminho traçado, pois ele conduz
somente até onde os outros foram.”

Alexandre Graham Bell

1 - Introdução

Nos últimos anos, têm-se observado um aumento significativo da importância da quimiometria^[1,2], área da química que emprega métodos matemáticos e estatísticos para a resolução de problemas químicos. Nesse sentido, diversos tutoriais têm sido publicados, tanto sobre aspectos gerais da área^[3] como sobre técnicas mais avançadas^[4,5].

De modo geral, a quimiometria pode ser dividida em várias áreas de estudo, tais como:

- Planejamento e Otimização de experimentos;
- Cartas de controle multivariado;
- Reconhecimento de Padrões;
- Calibração Multivariada;
- Resolução de curvas e misturas;
- Processamento de Sinais;
- Inteligência artificial (sistemas especialistas, redes neurais, algoritmos genéticos).

Dentre essas a calibração multivariada^[6,7] é uma das áreas da quimiometria mais empregada junto a dados espectroscópicos. A figura 1 ilustra, de forma simples, a idéia de calibração multivariada.

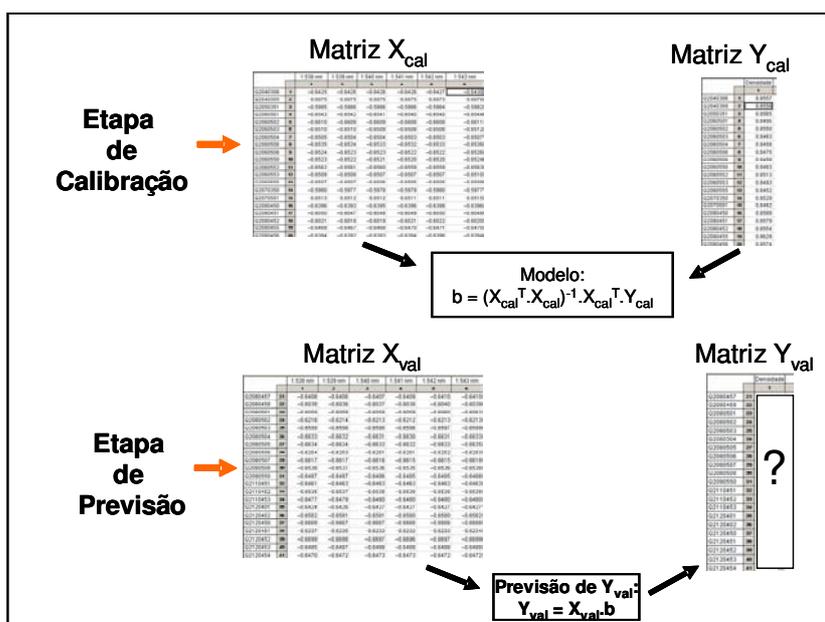


Figura 1: Representação da calibração multivariada usada em quimiometria.

O objetivo básico dessa técnica é procurar estabelecer uma relação quantitativa entre o espectro (matriz X) e a concentração ou propriedades das espécies de interesse (matriz Y), quando houver uma relação de dependência entre ambas.

Na etapa inicial de construção do modelo de calibração necessita-se das informações referentes aos sinais instrumentais (matriz independente - X), bem como o conhecimento da composição química quantitativa para uma grande quantidade de amostras (matriz dependente - Y). Porém, para a quantificação da composição de amostras do conjunto de previsão, os valores da matriz Y são determinados empregando-se apenas as informações da matriz "X" e usando o vetor de regressão "b" obtido na etapa de calibração.

A utilização de calibração multivariada em dados químicos propiciou uma série de benefícios tais como: melhoria na obtenção da informação quantitativa com aumento de seletividade e confiabilidade, resolução do problema de sobreposição de sinais analíticos, otimização de controle de qualidade e determinações quantitativas simultâneas de compostos químicos ou propriedades físico-químicas empregando-se apenas um instrumento analítico (espectrofotômetro, cromatógrafo, etc.).

Um dos primeiros métodos de calibração multivariada inversa empregado foi a Regressão Linear Múltipla (MLR)^[6,7], que procura resolver um sistema de equações lineares. Na regressão MLR assume-se que as concentrações ou propriedades físico-químicas (Matriz Y) é função linear das respostas instrumentais (Matriz X) e pode ser escrita matricialmente como:

$$Y = X\beta + e \quad \text{Eq. 1}$$

Onde: β = vetor dos coeficientes de regressão linear

e = vetor de resíduos ou erro de modelagem.

A representação matricial para a equação 1 pode ser escrita da seguinte forma:

$$\begin{matrix} | & & | \\ Y & = & X \\ | & & | \\ n & & n \end{matrix} \begin{matrix} | & & | \\ & & \beta \\ | & & | \\ & & n \end{matrix} \begin{matrix} | \\ + \\ | \\ n \end{matrix} \begin{matrix} | \\ e \\ | \\ n \end{matrix} \quad \text{Eq. 2}$$

Onde: n = número de amostras da matriz

q = número de variáveis da matriz

O vetor de regressão “b” é estimado, na etapa de calibração, utilizando o critério dos “mínimos quadrados” que minimiza o vetor de resíduos ($e = Y - Xb$) e é calculado por:

$$\beta = (X^T X)^{-1} X^T Y \quad \text{Eq. 3}$$

O vetor de concentração Y_{am} de uma amostra desconhecida é obtida por:

$$Y_{am} = X_{am} \beta \quad \text{Eq. 4}$$

Contudo, a MLR padece de problemas de robustez numérica e sensibilidade a ruído, em especial quando os dados apresentam elevada sobreposição espectral.

Considere o exemplo abaixo de uma matriz X_{cal} e o cálculo de seu determinante:

$$\begin{array}{c} X_1 \quad X_2 \\ X_{cal} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \\ 3 & 6 \end{bmatrix} \Rightarrow X_{cal}^T X_{cal} = \begin{bmatrix} 14 & 28 \\ 28 & 56 \end{bmatrix} \\ \det(X_{cal}^T X_{cal}) = 14 \times 56 - 28 \times 28 = 784 - 784 = 0 \end{array} \quad \text{Eq. 5}$$

Quando as colunas de X_{cal} são linearmente dependentes, as variáveis das colunas x_1 e x_2 são colineares, ou seja, quando x_1 aumenta, x_2 aumenta na mesma proporção. Portanto, a variável x_1 ou a variável x_2 porta informação redundante. Outra limitação da MLR diz respeito a pouca redução de ruído associado aos dados e que prejudica também a eficiência da calibração e previsão de constituintes de uma amostra.

Para resolver as limitações da MLR, surgiram as técnicas de Regressão em Componentes Principais (PCR) e por Mínimos Quadrados Parciais (PLS)^[7,8], que vêm sendo mais amplamente empregadas.

Na regressão PCR a matriz X é decomposta em um produto de duas outras matrizes, denominadas de scores “T” e loadings “P”, conforme a equação 6:

$$X = T P^T + E \quad \text{Eq. 6}$$

O vetor de regressões “ β ” em PCR é encontrado a partir de:

$$\beta = (T^T \cdot T)^{-1} \cdot T^T \cdot Y \quad \text{Eq. 7}$$

No passo seguinte, usa-se regressão por MLR para relacionar a matriz Y com a matriz dos scores T:

$$Y = T \beta + E \quad \text{Eq. 8}$$

A figura 2 representa a regressão PCR e a transformação das variáveis originais em componentes principais.

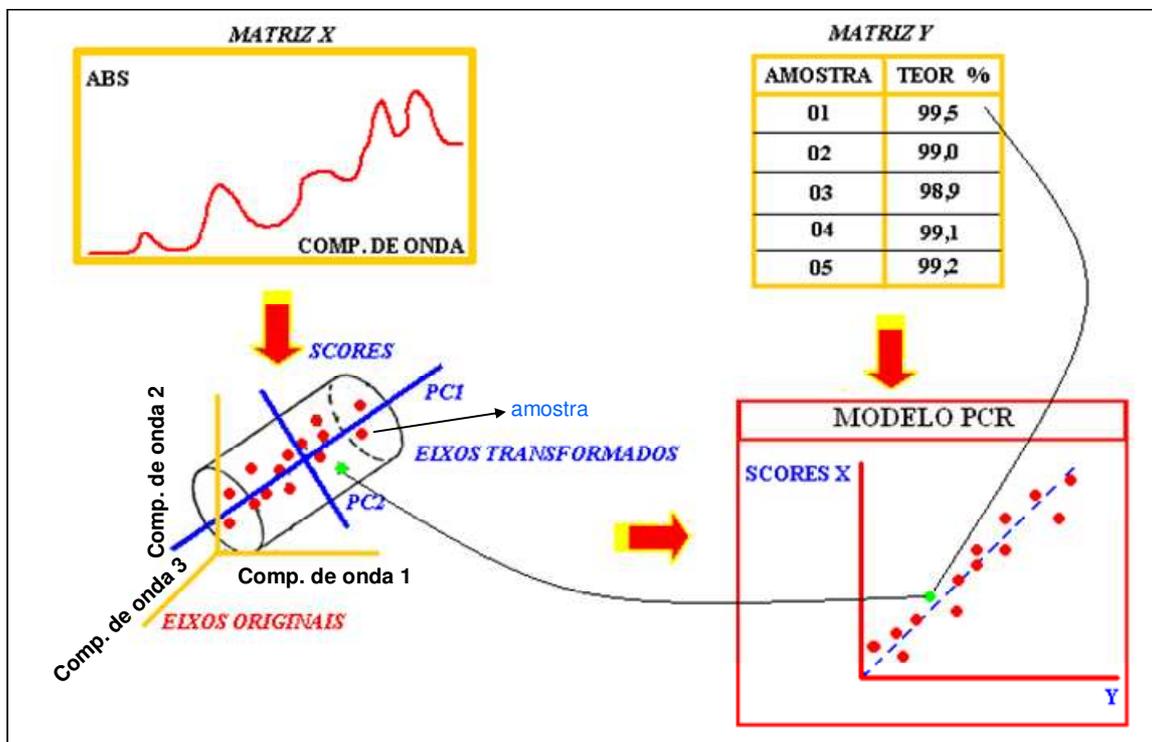


Figura 2: Regressão PCR e transformação em Componentes Principais (PC1 e PC2).

Na regressão por PLS, tanto a matriz X como a matriz Y são transformadas em PC (variáveis latentes), formando dois grupos de matrizes de scores e loadings (T, U e P, Q) e dois blocos de equações:

$$X = T P^T + E \quad e \quad Y = U Q^T + F \quad \text{Eq.9}$$

O modelo final consiste em relacionar linearmente os scores do bloco X (matriz T) com os scores do bloco Y (matriz U), dando:

$$U = \beta T + Ge \quad Y = \beta T Q^T + H \quad \text{Eq.10}$$

Nas equações 9 e 10, os termos E, F, G e H correspondem as matrizes de resíduos, nas determinações dos termos X, Y e U. A figura 3 representa a regressão PLS e a transformação das variáveis originais em variáveis latentes.

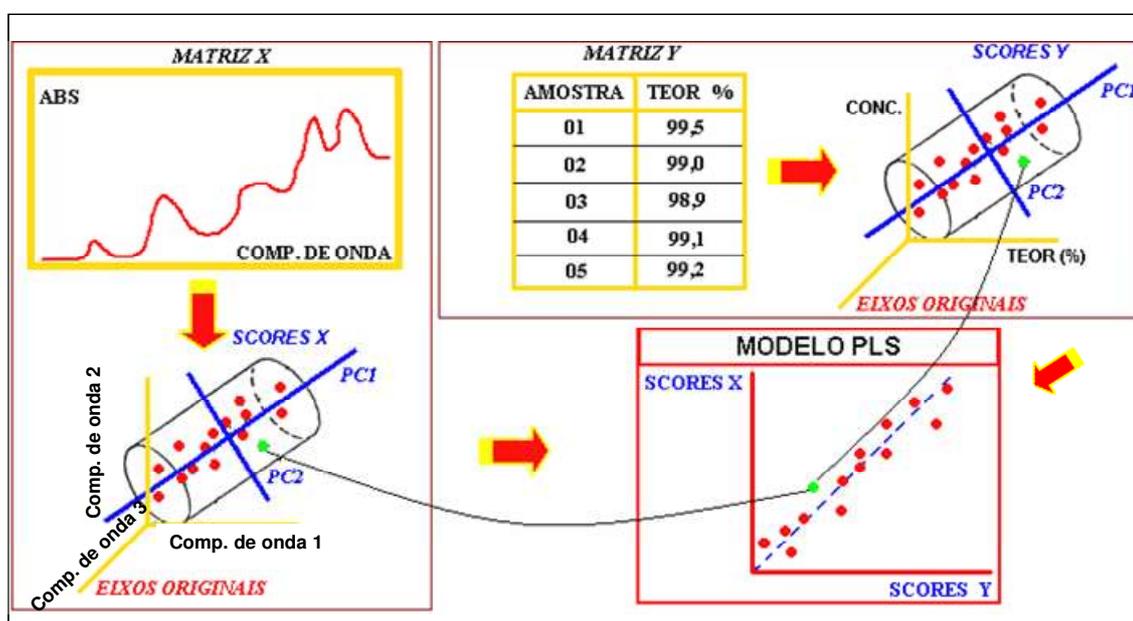


Figura 3: Regressão PLS e transformação em Variáveis Latentes (PC1 e PC2).

Devido à existência de ortogonalidade entre as componentes principais nas regressões PCR e PLS, o problema de colinearidade entre as variáveis originais, que era um dos inconvenientes na regressão MLR, é contornado. A pequena sensibilidade ao ruído instrumental associado aos dados, também tornou as regressões PCR e PLS bem aceitas e amplamente divulgadas.

Uma das críticas feitas às regressões baseadas em componentes principais é a utilização de transformações matemáticas que não conservam o significado físico das variáveis originais, tornando a interpretação e o tratamento dos dados mais complexos que no caso da regressão MLR.

1.1 - Métodos de Seleção de Variáveis.

Devido à complexidade dos espectros para a realização de uma análise quantitativa de mais de um componente, é necessária a utilização de muitas variáveis (comprimentos de onda), exigindo o emprego de ferramentas matemáticas específicas para este fim.

A fim de simplificar tanto os modelos MLR quanto os modelos PCR e PLS de calibração multivariada, grande destaque tem sido dado à aplicação de técnicas de seleção de variáveis^[9-22], que mantêm as variáveis que contém informações relevantes, eliminando aquelas que carregam apenas ruído ou redundância. Com isso torna-se muitas vezes possível a utilização de MLR que é um método de calibração mais simples que os baseados em componentes principais (PCR e PLS), tanto no que diz respeito à calibração propriamente dita, quanto à interpretação dos resultados.

Selecionar a combinação de variáveis que leve à otimização do modelo é uma das maiores dificuldades da análise multivariada. Um dos problemas práticos é saber quantos e quais comprimentos de onda devem ser escolhidos, principalmente quando ocorre alta sobreposição espectral. Para contornar esses problemas, diversos algoritmos matemáticos têm sido desenvolvidos.

Entre as diversas técnicas de seleção de variáveis empregadas em MLR, foram desenvolvidos e empregados, ao longo de muitos anos, os métodos iterativos de seleção, como o “forward selection”^[14], “backward elimination”^[14] e o “stepwise regression”^[15].

O método “Forward Selection” emprega um algoritmo iterativo que inicia com uma única variável “x” e mais variáveis são adicionadas sucessivamente à regressão até que uma condição de parada seja satisfeita.

A variável inicial é aquela que possui maior correlação com a variável dependente “y”. Em cada iteração subsequente, o efeito de incluir cada uma das variáveis remanescentes é avaliado de acordo com um “teste-F” parcial. Cada teste envolve a inclusão da variável considerada, construção do novo modelo, cálculo dos valores correspondentes de $\hat{\sigma}^2$ (estimativa da variância dos resíduos), SS_{reg} (ou *RSS- soma dos quadrados da regressão*) e realizado o teste-F parcial. A variável com o maior valor de F é então incluída no modelo. Este processo de seleção continua até que nenhuma das variáveis remanescentes tenha um valor de F parcial maior que um dado valor crítico.

A limitação do método “Forward Selection” é o fato de que uma vez que a variável foi incluída no modelo, ela não pode ser removida na iteração subsequente. Então, algumas combinações em que esta variável não está presente nunca são testadas durante o processo de seleção.

O método de seleção “Backward Elimination” inicia-se pela construção de um modelo que utiliza todas as variáveis. Posteriormente, o efeito de remover cada uma das variáveis do modelo é avaliado de acordo com um teste F parcial como no algoritmo “Forward Selection”. A variável com o menor valor de F parcial é então removida. Este processo de seleção continua até que todas as variáveis do modelo tenham valores de F parcial maiores que um dado valor crítico.

A limitação encontrada no método “*Backward Elimination*” é o fato de que a variável uma vez removida do modelo, ela não pode ser reintegrada em iterações subsequentes. Então, algumas combinações em que esta variável está presente nunca serão testadas durante o processo de seleção. Como pode ser visto, esta desvantagem é exatamente o inverso da limitação mostrada para o método “*Forward Selection*”. Além disso, se o número total de variáveis (K) é maior ou igual ao número de amostras de calibração (N), não será possível construir um modelo MLR com todas as variáveis selecionadas, como requerido no passo inicial do algoritmo.

O método “Stepwise Regression” pode ser considerado uma combinação dos métodos “*Forward*” e “*Backward*” descritos acima. O algoritmo resultante adiciona progressivamente novas variáveis ao modelo, iniciando com uma variável x com a maior correlação com a variável dependente y , como no método “*Forward Selection*”. Além disso, o algoritmo emprega um mecanismo para eliminar variáveis como no método “*Backward Elimination*”.

O algoritmo “Stepwise Regression” tem duas principais desvantagens. Primeiramente, selecionar os níveis de significância α associados ao teste parcial F pode não ser uma escolha direta. Em segundo lugar, o teste parcial F é baseado em suposições estatísticas (tais como distribuição gaussiana dos dados) que pode não ser verdadeiro para uma série de dados em particular.

Nos últimos anos, o algoritmo genético (GA)^[23-30] tem se destacado como técnica de seleção de variáveis. Trata-se de uma ferramenta matemática não determinística inspirada

na teoria da seleção natural de Darwin e tem sido aplicado a dados espectrais nas regiões ultravioleta-visível (UV-VIS)^[31,32] e infravermelho próximo (NIR)^[33,34]. Porém, devido ao caráter estocástico deste algoritmo, há uma relativa demora em se encontrar a melhor combinação de parâmetros (população inicial, número de gerações, etc.) que leve à solução do problema, que no caso do presente trabalho é encontrar o menor erro quadrático médio na predição dos parâmetros de qualidade de amostras de óleo diesel combustível e madeira de eucalipto.

1.1.1 - O Algoritmo das Projeções Sucessivas (APS) para seleção de variáveis.

Em 2001, Araújo et al desenvolveram o Algoritmo das Projeções Sucessivas (APS)^[35], que é mais simples de utilizar que o GA e foi desenvolvido inicialmente para seleção de variáveis. Trata-se de um algoritmo determinístico cujo principal objetivo é selecionar variáveis minimamente redundantes e com isso solucionar problemas de colinearidade entre elas. Para isso utiliza-se operações de projeção de vetores e a técnica tem sido aplicada em dados de espectroscopia de emissão em plasma^[36] e na região do infravermelho próximo^[37].

Apesar dos fundamentos teóricos do APS já terem sido divulgados e detalhados em trabalhos anteriores^[35,36,38], faz-se necessário a inclusão de um exemplo prático neste momento.

Considere a seleção de comprimentos de onda escolhidos de um conjunto de $j=5$ variáveis espectrais: (400 nm (X1), 450 nm (X2), 500 nm (X3), 550 nm (X4) e 600 nm (X5)) para a determinação multivariada da concentração de $A=2$ analitos (Co e Ni) usando-se $M_{cal} = 3$ misturas de calibração. A seqüência de operações do algoritmo é obedecida de seguinte maneira:

- a) Escolha do primeiro vetor ($k(0)$) tomado como partida, para o cálculo das projeções. Todos os outros vetores são posteriormente testados como vetor inicial, para a formação das cadeias de variáveis candidatas.

- b) Cálculo das projeções dos demais vetores (comprimentos de onda) em um subespaço ortogonal ao do vetor inicial, conforme figura 4.

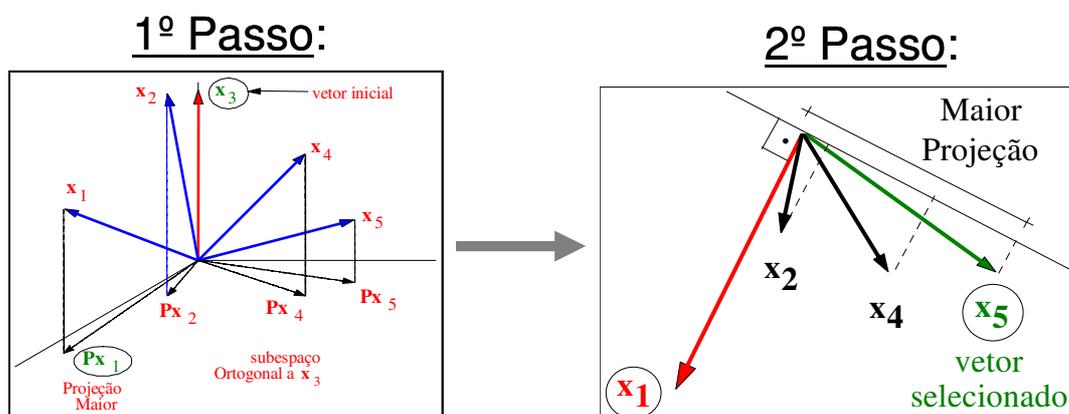


Figura 4: Exemplo de operações de projeções efetuadas com o uso do APS. para selecionar comprimentos de onda a partir de 5 variáveis espectrais ($J=5$), 3 amostras de calibração ($M_{cal} = 3$), e comprimento de onda tomado como partida ($k(0) = 3$, resultado da primeira iteração: $k(1)=1$).

- c) Organização das cadeias de variáveis, onde cada cadeia é composta pelo vetor de partida (neste exemplo X_3) além das variáveis mais ortogonais (X_1 e X_5).

O número de variáveis a selecionar em cada cadeia é limitado pelo número de misturas (amostras) de calibração, como toda regressão MLR. Neste exemplo, como $M_{cal} = 3$, esse também deve ser o número máximo de comprimentos de onda a serem selecionados para cada cadeia de variáveis. No exemplo da figura 2 a primeira cadeia de variáveis é formada pelos comprimentos de onda X_3 , X_1 e X_5 que seriam: 500, 400 e 600 nm.

Para se determinar qual a melhor cadeia de variáveis, faz-se uma regressão por MLR para cada uma das cadeias, empregando-se como matriz X os valores de absorvância nos comprimentos de onda da cadeia de variáveis e como matriz Y as concentrações dos metais: Co e Ni.

Realiza-se o cálculo do erro quadrático médio de previsão (RMSEP) de amostras de um conjunto de teste (diferente do de calibração) associado a cada um dos modelos MLR construídos e dado pela equação 11:

$$RMSEP = \sqrt{\frac{1}{M} \sum_{m=1}^M (y_m - yref_m)^2} \quad Eq. 11$$

Onde:

M = Número de amostras de calibração.

y_m = Valor de Y previsto pelo modelo para a amostra m .

$yref_m$ = Valor de Y obtido pelo método de referência para a amostra m .

A cadeia de variáveis cujo modelo MLR levar ao menor RMSEP será escolhida para este propósito.

Vale notar que tanto o uso do APS quanto a calibração por MLR são procedimentos simples e rápidos de serem implementados em qualquer programa computacional disponível.

1.1.2 - O APS empregado em seleção de amostras.

Um assunto que também tem atraído a atenção de muitos pesquisadores é a busca de métodos quimiométricos de seleção de amostras em calibração multivariada^[39-46]. Isto torna-se importante por economizar trabalho experimental e computacional, se as rotinas de análise multicomponente necessitam utilizar instrumentos que precisam ser periodicamente recalibrados, por causa dos efeitos de envelhecimento, mudanças nas condições ambientais de trabalho ou manutenção regular, por exemplo. Seleção de amostras também é importante no contexto da transferência de modelos de calibração multivariada^[46,47], uma vez que proporciona benefícios tais como a redução do custo envolvido não somente na repetição das análises, mas também no transporte físico das mesmas.

Apesar de o APS ter sido inicialmente desenvolvido para seleção de comprimentos de onda, neste trabalho o algoritmo foi adaptado para selecionar em uma seqüência de operações primeiro as variáveis espectrais e depois as amostras de transferência de calibração, conforme a figura 5.

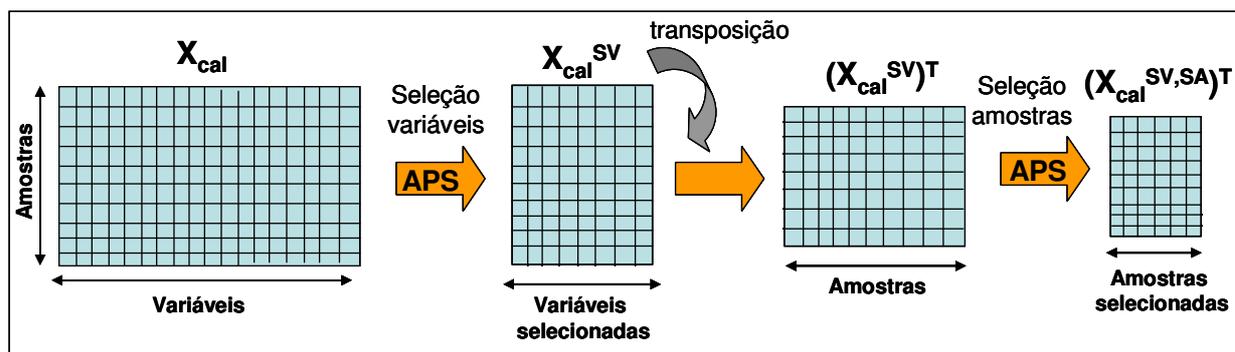


Figura 5: Fluxograma do uso do APS para seleção de variáveis e amostras para calibração multivariada.

Conforme se verifica na figura 5, uma simples transposição da matriz de dados após a seleção inicial de variáveis espectrais é suficiente para promover a operação de seleção de amostras de calibração que agora representam as colunas da matriz de dados.

Vale notar também que apenas um algoritmo empregado para selecionar variáveis e amostras proporciona vários benefícios à calibração multivariada, tais como a construção de modelos MLR mais simples e mais fáceis de interpretar, servir de guia para a escolha de diodos emissores de luz (LED) a serem empregados na construção de fotômetros, bem como a seleção de amostras para transferência de calibração e recalibrações mais simples.

1.2 - Técnicas de Filtragem de Sinais Instrumentais.

Outra área da quimiometria que tem auxiliado bastante os químicos é a utilização de técnicas de processamento de sinais, visando melhorias na interpretação e utilização dos dados^[5].

Entre os ruídos que normalmente contaminam respostas instrumentais, podem ser citados o Johnson, o Shot, o Flicker e o ambiental^[48]. Os métodos mais comumente utilizados em química para remoção de tais interferências são os filtros de média boxcar, média móvel e Savitzky-Golay^[48]. Embora de fácil implementação, tais filtros tendem a distorcer picos estreitos e uma parte da informação útil contida no sinal é geralmente removida junto com o ruído, uma vez que tais filtros atuam diretamente sobre o sinal

analítico e não sobre as frequências que compõem tal sinal. Esta característica torna estes métodos de suavização pouco seletivos, isto é, incapazes de diferenciar o que é sinal do que é ruído. Este problema pode ser minimizado ao se empregar métodos de filtragem, tais como transformada de Fourier e transformada Wavelet, que atuam diretamente no domínio das frequências que compõem o sinal analítico, possibilitando a remoção mais seletiva do ruído associado aos dados.

Para uma melhor compreensão de como os métodos que empregam a transformada de Fourier e Wavelet atuam sobre o sinal analítico, é apresentado na figura 6 um diagrama esquemático do procedimento de implementação desses métodos.

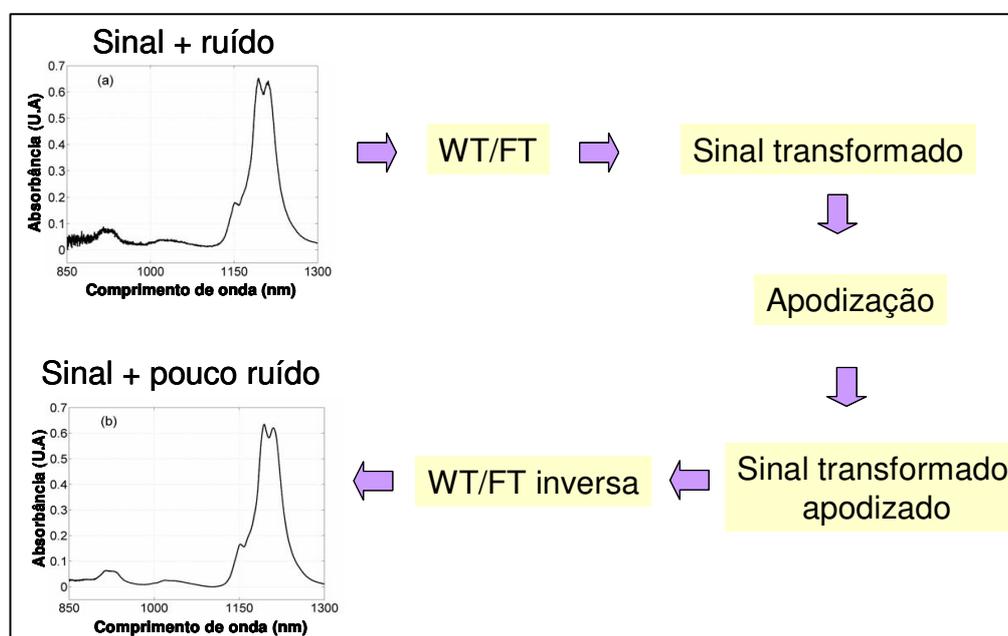


Figura 6. Diagrama do processo de filtragem do espectro NIR de diesel usando FT e WT aplicados ao sinal original.

Como pode ser verificado na figura 6 o sinal analítico é transformado em espectros de frequências através da FT ou WT e, em seguida, é aplicada uma função de apodização (corte de frequências) para remover sinais espúrios, ou seja, sinais não relacionados à informação analítica. Após a apodização, tem-se um sinal transformado com menor ruído, no qual aplica-se a inversa da FT ou WT para restaurar o sinal original com o ruído minimizado.

1.2.1 - Método da Transformada de Fourier (FT)

Em sua versão discreta, a FT de um sinal $f(x)$ com N pontos, é calculada como^[5]:

$$F(\omega) = \sum_{x=0}^{N-1} f(x) \exp(-i\omega x) \quad \text{Eq. 12}$$

onde x é uma variável cuja natureza depende da técnica instrumental empregada na geração do sinal (por exemplo, comprimento de onda no caso de espectroscopia). Vale ressaltar que alguns autores empregam um fator de normalização de $1/N$ na Equação 2, mas a expressão aqui apresentada é a empregada pelo software MATLAB^[49]. O parâmetro ω , denominado frequência, assume os seguintes valores:

$$\omega = 2\pi \frac{k}{N}, \quad k = 0, 1, \dots, N-1 \quad \text{Eq. 13}$$

onde k é chamado “índice de frequência”.

Como resultado da aplicação da FT, obtém-se um conjunto de N coeficientes, indexados por k . Anulando os coeficientes em que a presença do ruído é preponderante (normalmente em frequências elevadas) e aplicando a Transformada de Fourier Inversa (FTI), obtém-se um sinal mais “limpo”. O critério de apodização é um parâmetro muito importante quando se usa a FT na suavização de ruídos instrumentais. Contudo, uma das limitações da FT encontra-se não no critério de apodização, mas sim no fato de que ela não permite analisar em separado diferentes trechos do sinal. Isso traz duas dificuldades:

- 1) O analista não pode identificar quais regiões do espectro são mais relevantes para a tarefa em questão.
- 2) Se uma região do espectro é particularmente ruidosa, ou porta informação irrelevante para a análise, a sua contribuição não pode ser eliminada no cálculo dos coeficientes.

Uma solução para esse problema consiste no uso da Transformada de Fourier Janelada. Nesse caso, o espectro é dividido em regiões e sobre cada região é aplicada a FT (figura 7).

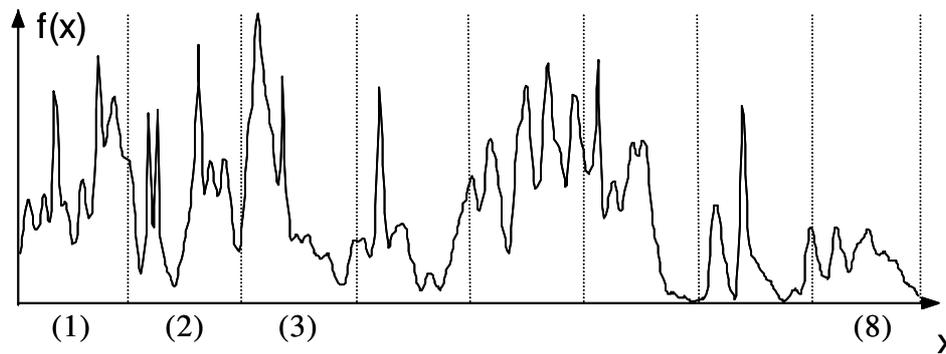


Figura 7. Espectro dividido em 8 regiões com o uso de janelas retangulares não-justapostas.

Matematicamente, a Transformada de Fourier Janelada pode ser formulada da seguinte maneira:

$$F(p, \omega) = \sum_{x=0}^{n-1} f(x)\omega(x-p)\exp(-i\omega x) \tag{Eq.14}$$

onde $w(x)$ é uma função de janelamento, responsável pela delimitação do trecho que está sendo considerado no sinal. A posição da janela dentro do sinal é dada pelo parâmetro p .

No exemplo da figura 7, a função $w(x)$ realiza cortes abruptos, dividindo o espectro em intervalos bem delimitados (janelas). Contudo, tal procedimento pode fazer com que variáveis adjacentes, que normalmente portam informação similar, sejam colocadas em janelas separadas. Para minimizar esse problema, conhecido como “efeito de borda”, costumam-se usar funções-janela com decaimento suave, tais como a gaussiana:

$$\omega(x) = \exp\left(-\frac{x^2}{2}\right) \tag{Eq. 15}$$

Outra questão relevante é a escolha da largura para a função-janela. Suponha, por exemplo, que sejam usadas janelas retangulares com M comprimentos de onda. Cada janela dá então origem a M coeficientes Fourier e, portanto, a preservação do número de variáveis requer o uso de N/M janelas. Note-se que, aumentando M , melhora-se a análise de cada janela (mais coeficientes), mas perde-se resolução espacial (menos janelas). Esse problema da área de processamento de sinais^[50] faz lembrar o Princípio da Incerteza de Heisenberg.

As limitações da FT aplicadas ao processamento de sinais não estacionários, bem como a dificuldade do emprego de janelas espectrais de largura variável aplicados aos dados, promoveu a inspiração para o surgimento de uma outra categoria de transformada matemática: as Wavelets.

1.2.2 - Método da Transformada Wavelet (WT)

Trata-se de uma família de funções matemáticas (do inglês: Wave = onda, let = pequena) que transformam sinais em diferentes componentes de frequência e permitem o estudo de cada componente com uma resolução conhecida como escala.

A WT emprega conversões de natureza similar às da Transformada de Fourier por janelas, isto é, oscilações limitadas dentro de uma janela. Tais oscilações são chamadas “wavelets”. A diferença encontra-se na forma de parametrizar as funções, que agora permite o uso de janelas de largura variável:

$$v_{a,b}(x) = \frac{1}{\sqrt{a}} \psi\left(\frac{x-b}{a}\right) \quad \text{Eq. 16}$$

O parâmetro a é chamado “escala”, e determina a largura da janela de análise. O parâmetro b , chamado “posição”, ou “deslocamento”, dá a posição central da janela dentro do espectro. Há uma ampla gama de possíveis escolhas para a função $\psi(x)$, denominada wavelet-mãe, que pode ou não ter uma forma similar à da equação 12. Duas possíveis alternativas, por exemplo, são:

1) Wavelet originalmente proposta por Morlet^[51]:

$$\psi(x) = \exp\left(-\frac{x^2}{2}\right) \exp(-i5x) \quad \text{Eq. 17}$$

em que $i = \sqrt{-1}$.

2) Segunda derivada da gaussiana, ou “chapéu mexicano”:

$$\psi(x) = (1 - x^2) \exp\left(-\frac{x^2}{2}\right) \quad \text{Eq. 18}$$

Essas wavelets estão apresentadas na figura 8.

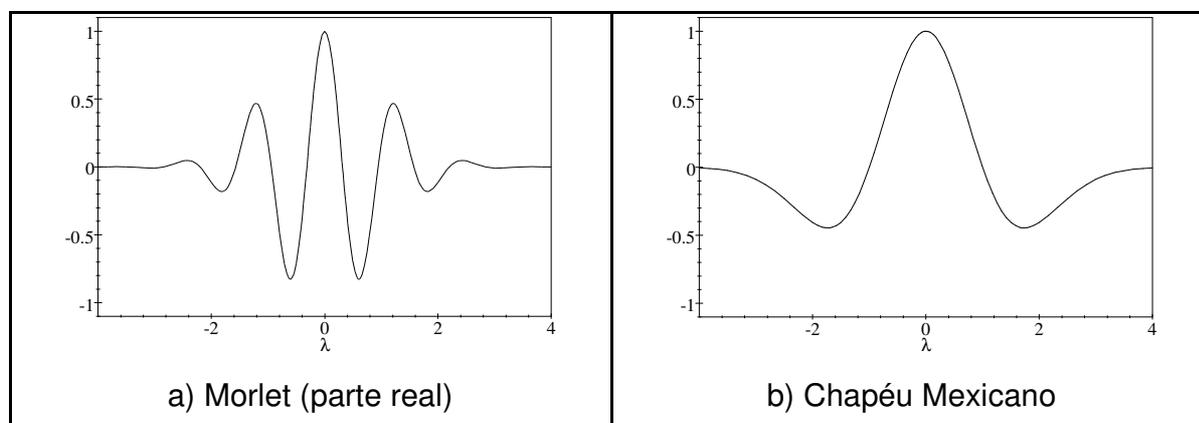


Figura 8. Exemplos de Wavelets dos tipos: a) Morlet e b) Chapéu Mexicano.

Outras famílias de wavelets bastante empregadas em quimiometria e que foram usadas nesta tese, são as do tipo Symlet, Daubechie e Coiflet, representadas na figura 9.

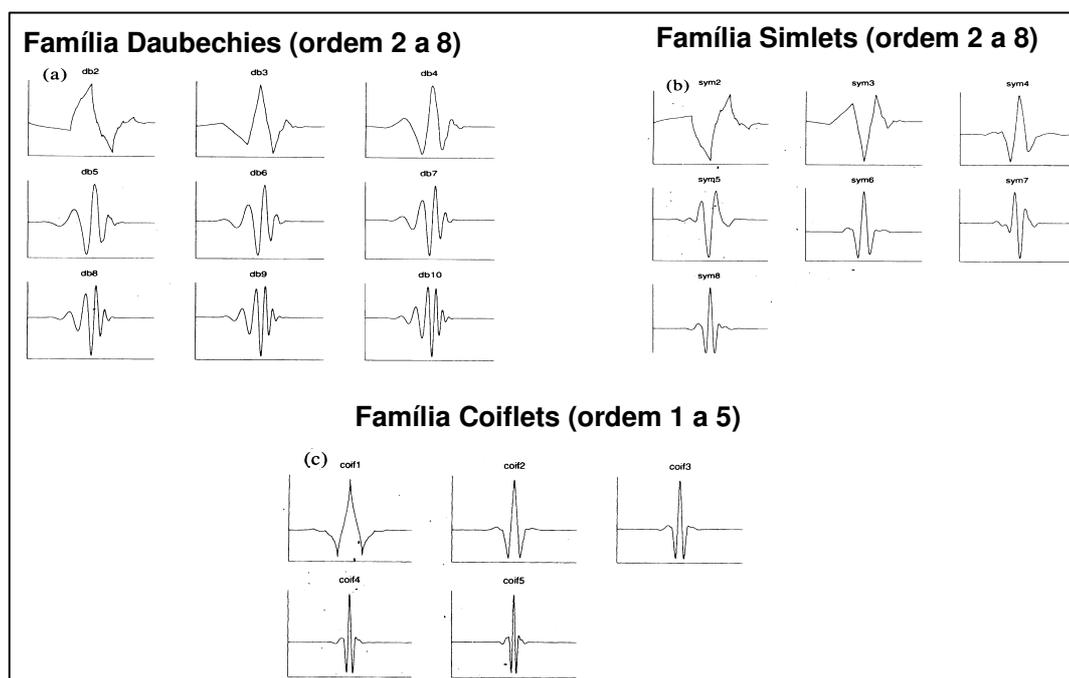


Figura 9. Wavelets das famílias Simlet, Daubechies e Coiflet, empregadas na decomposição dos espectros NIR de óleo diesel e madeira de eucalipto.

A Figura 10 apresenta o efeito do aumento e da redução na escala sobre a wavelet Morlet.

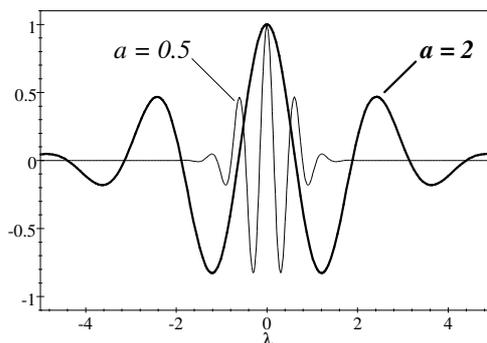


Figura 10. Efeito de modificações na escala

Embora os coeficientes a, b na equação 16 possam, em princípio, assumir qualquer valor real (com $a \neq 0$), é de praxe fazer a seguinte restrição:

$$a = 2^m, \quad b = n2^m \tag{Eq. 18}$$

com (m, n) inteiros. Pode-se mostrar^[51,52] que dessa forma a informação do espectro é preservada utilizando-se um número de coeficientes igual ao número original de variáveis. Além disso, torna-se possível obter os coeficientes de uma forma computacionalmente eficiente, através do algoritmo de decomposição em árvore proposto por Mallat^[52,53] e representado esquematicamente na figura 11.

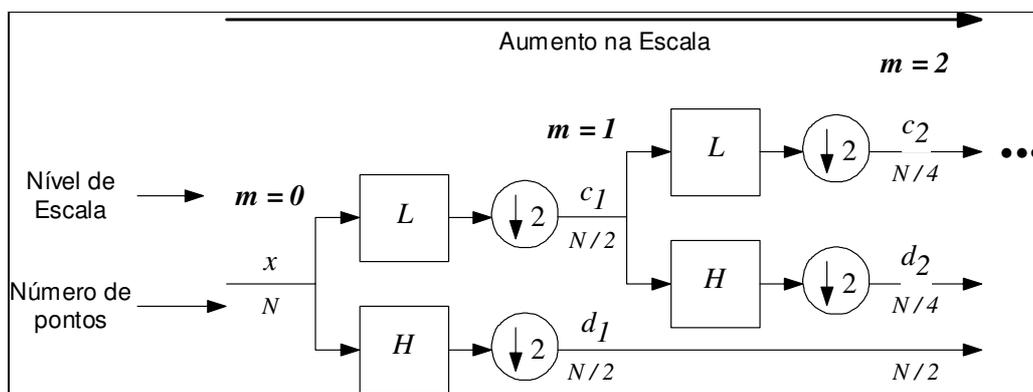


Figura 11. Algoritmo de Decomposição em Árvore. Os coeficientes wavelet no nível m correspondem a d_m .

Nessa figura, os blocos L e H representam filtros digitais passa-baixas e passa-altas, respectivamente, que estão associados à wavelet adotada na análise. Normalmente, tais filtros são implementados na forma de média móvel, de modo que o esforço computacional envolvido é pequeno. O símbolo $(\downarrow 2)$ representa a operação de sub-

amostragem, que consiste em eliminar todos os coeficientes de índice ímpar de uma seqüência. A sub-amostragem garante que o número total de pontos permaneça constante em cada nível da árvore, de modo a não introduzir redundâncias.

O algoritmo de decomposição em árvore pode ser interpretado da seguinte forma: Em cada divisão L / H , uma “fatia” da informação é separada (coeficientes wavelet d_m , também chamados de coeficientes de “detalhe” no nível m) e o restante segue para uma nova divisão (coeficientes c_m , ou coeficientes de “aproximação” no nível m). A saída da árvore de decomposição depende do número de níveis de escala (também chamados “níveis de resolução”) empregados. A figura 12 exemplifica três possíveis decomposições.

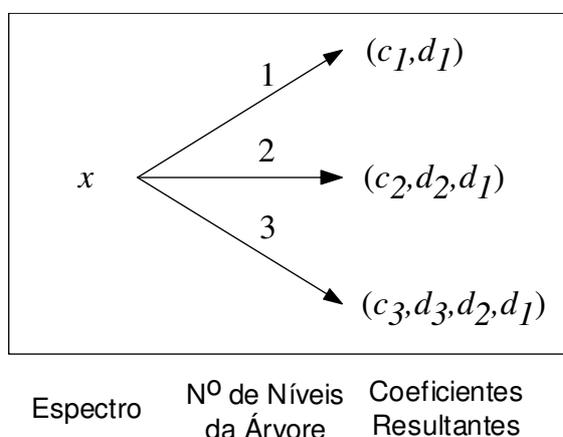


Figura 12. Três possíveis formas de se decompor o espectro em coeficientes wavelet.

Do ponto de vista quimiométrico, os coeficientes wavelet de pequena escala estarão associados a características bem localizadas do espectro, tais como raias estreitas e bem pronunciadas. Os coeficientes de escalas maiores estarão associados a características que se espalham por regiões espectrais mais largas. O “resíduo” c_m corresponderá a uma “linha média” do espectro, removidos os picos mais proeminentes. À semelhança da Transformada de Fourier Discreta, espera-se que informações de relevância estejam concentradas em um número reduzido de coeficientes, de modo a facilitar a análise.

Note-se que um segundo algoritmo em árvore pode ser usado para reconstruir o espectro a partir dos coeficientes (Transformada Wavelet Inversa) como mostra a Figura 13.

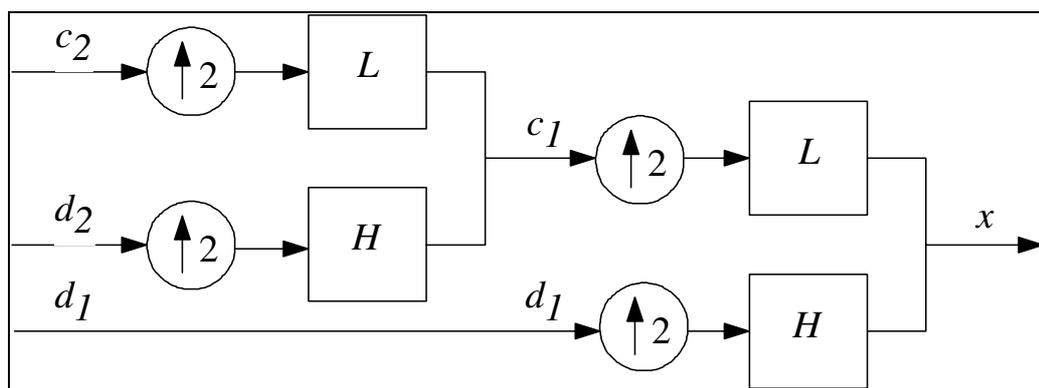


Figura 13. Árvore de Reconstrução Wavelet com dois níveis.

Nessa árvore de reconstrução, o símbolo $(\uparrow 2)$ representa a operação de inserção de zeros entre os pontos de uma seqüência. É possível mostrar que, sob certas condições, os mesmos filtros L, H usados na fase de decomposição também podem ser usados na reconstrução do espectro^[51].

Um interessante paralelo pode ser traçado entre o esquema de reconstrução wavelet e processos de visão, que eram o campo de pesquisa original de Mallat. Ao olharmos uma pessoa à distância, é possível distinguir apenas os contornos de seu corpo (características de larga escala). Contudo, ao nos aproximarmos, passamos a perceber mais detalhes, tais como os traços de seu rosto (características de pequena escala). Ou seja, nesse processo, a formação da imagem passa pela incorporação sucessiva de detalhes em escalas cada vez mais finas.

Em estudos químicos a WT tem sido empregada desde 1989 e já foram publicados mais de uma centena de artigos^[54]. Na química analítica a WT tem sido empregada principalmente para remoção de ruído em calibração multivariada^[55-58].

Galvão e demais pesquisadores^[5] realizaram um estudo comparativo sobre redução de ruídos espectrais usando transformadas de Fourier e Wavelet. Foi demonstrado que em espectros que apresentam poucos pontos notáveis, tais como espectros de absorção molecular UV-VIS, os resultados da filtragem via FT e WT são semelhantes. Entretanto, para sinais com baixo “fator de preenchimento” (filling factor), como espectros de emissão em plasma, o uso da WT conduz a melhores filtragens, principalmente por não distorcer o formato de picos estreitos.

Nas duas últimas décadas, o método de filtragem de sinais baseado na transformada Wavelet vem ganhando popularidade^[54-67]. Técnicas “Wavelet” de filtragem de sinais podem contribuir para a redução do ruído associado aos espectros NIR e por consequência promover aos modelos de calibração multivariada baseada em MLR, melhoria na exatidão da estimativa dos parâmetros de qualidade do óleo diesel e da madeira de eucalipto, que constitui uma das aplicações realizadas neste projeto.

1.2.3 - Filtragem Wavelet de espectros NIR médios a partir de varreduras individuais.

Uma forma usual de melhorar a relação sinal/ruído (SNR) em um sinal instrumental consiste na aquisição de repetidas medidas e cálculo da média das varreduras (scans) individuais^[48,68]. Porém, o número de varreduras não pode ser excessivamente grande devido à restrições de tempo, bem como variações na intensidade (“drift”) dos comprimentos de onda.

Neste contexto, a transformada Wavelet discreta (DWT) pode ser uma ferramenta útil^[69]. Para este propósito, o sinal é convertido para o domínio wavelet, geralmente por um algoritmo baseado em um banco de filtros^[50], e uma transformação é aplicada para atenuar (comprimir) os coeficientes wavelet com baixa razão sinal/ruído. Um procedimento simples consiste em manter os coeficientes de aproximação (componentes de baixa frequência do sinal) da DWT, e substituir todos os coeficientes de detalhe (componentes de alta frequência do sinal) com zeros^[70]. Tal aproximação considera que as características mais importantes do sinal se concentram em baixas frequências, enquanto os elementos de alta frequência são representados pelo ruído.

Métodos mais elaborados geralmente empregam uma operação de apodização (corte), em que os coeficientes cujo valor estiverem abaixo de um dado limiar de apodização são substituídos por zeros. Os coeficientes restantes podem ser mantidos inalterados (“método do limiar rígido”) ou podem ser reduzidos pelo valor do limiar adotado (“limiar suave”)^[71].

Métodos de apodização requerem uma escolha agressiva do limiar de apodização que pode não ser de escolha fácil se as estatísticas do ruído (ou pelo menos a intensidade do ruído) for desconhecida.

Neste trabalho foi desenvolvido e aplicado, a dados simulados e a espectros NIR, um novo método de filtragem wavelet^[72] que explora as estatísticas de scans individuais obtidas durante o processo de aquisição de espectros médios.

O método proposto consiste na compressão dos coeficientes wavelet do sinal ruidoso por um fator que minimiza o erro quadrático médio (RMSE) esperado com relação ao sinal verdadeiro. Uma vez que o sinal verdadeiro não é conhecido, uma estimativa sub-ótima do fator de compressão é calculada usando a estatística das varreduras adquiridas das amostras.

1.3 – Espectroscopia NIR

Uma das técnicas mais empregadas no âmbito das análises não destrutivas é a espectroscopia na região do infravermelho próximo (NIR). Essa técnica abrange os comprimentos de onda no intervalo de 780-2.500 nm do espectro eletromagnético^[48]. A figura 14 representa a localização da radiação NIR no espectro eletromagnético.

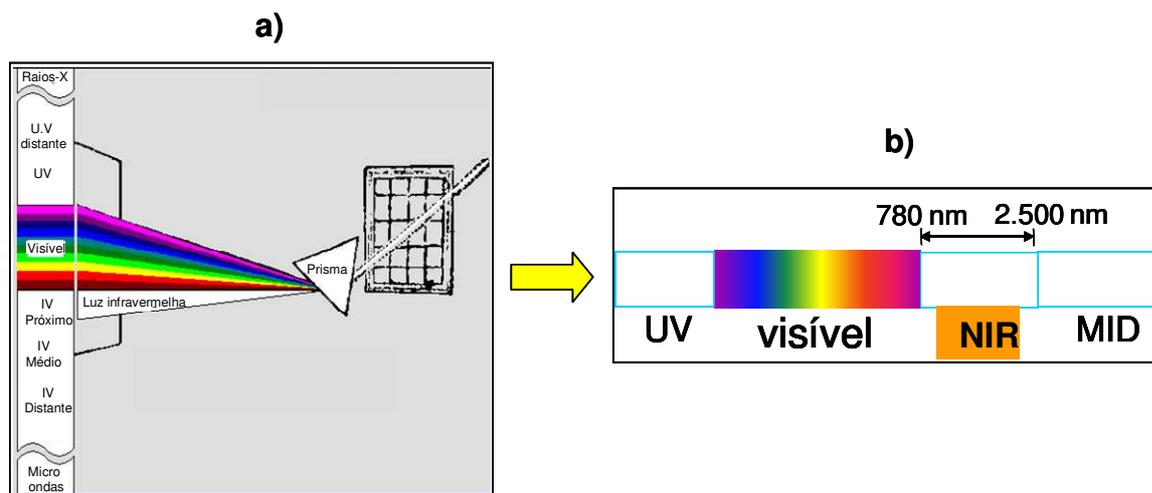


Figura 14. a) Localização da radiação NIR no espectro eletromagnético. b) Detalhamento dos comprimentos de onda que delimitam a região NIR e sua localização entre as regiões visível e infravermelho médio.

As ocorrências espectrais da radiação NIR se devem quase que totalmente às bandas de absorção relacionadas aos sobretons de transições fundamentais que ocorrem no infravermelho médio e às combinações de transições que ocorrem também nessa região.

A Figura 15 representa as bandas de absorção do espectro de transmitância da água, absorvendo na região do infravermelho médio (2,5 – 50 μm).

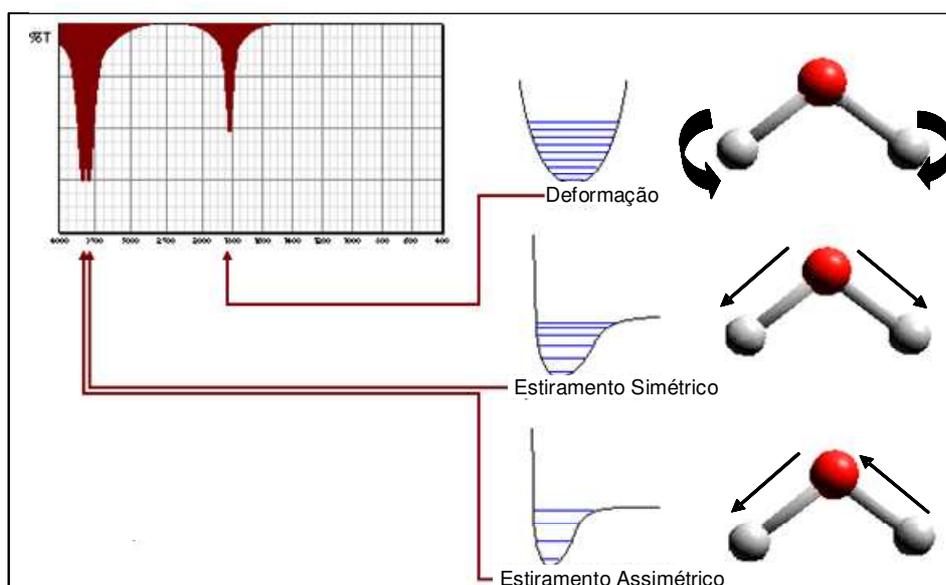


Figura 15. Vibrações fundamentais da molécula da água e suas respectivas transmitâncias na região do infravermelho médio.

As moléculas que mais podem absorver radiação NIR são aquelas constituídas das ligações C-H, N-H, O-H e S-H, que por serem de alta energia e envolverem átomos de massa relativamente baixa, têm transições intensas na região do infravermelho médio, devido a sua alta anarmonicidade. A tabela de absorções NIR, localizada no Apêndice A, resume os principais grupos característicos que podem absorver na região NIR.

A partir disso, pode-se obter um vasto campo de aplicação da radiação NIR ^[73], cuja natureza química das amostras se enquadre nas características citadas acima. A Tabela 1 ilustra as principais aplicações para o setor industrial.

Tabela 1. *Uso da espectroscopia NIR em análises químicas.*

1	Determinação do peso molecular de polímeros como o propileno e etilenoglicol
2	Determinação de constituintes de blendas têxteis contendo algodão-poliéster
3	Acabamento de fibras têxteis
4	Grau de hidrogenação de ácidos esteáricos
5	Misturas de hidrocarbonetos
6	Octanagem de gasolina
7	Teor total de alcalóides, nicotina e açúcares redutores em tabaco
8	Determinação do teor de proteína e umidade em produtos agrícolas
9	Monitoramento de quase todos os parâmetros da indústria petroquímica
10	Determinação de tamanho de partículas
11	Teor de açúcar em beterraba
12	Adulteração de suco de laranja
13	Grau BRIX de um grande número de produtos ou matérias primas
14	Etanol em gasolina
15	Água em solventes
16	Teor de gordura em leite
17	Aditivos em papel
18	Homogeneização de blendas de pós
19	Identificação e quantificação de princípios ativos de fármacos
20	Classificação de polímeros na indústria de reciclagem

1.3.1 - Análise NIR de Óleo Diesel

Para a economia brasileira, um dos campos de aplicação mais importantes da espectroscopia NIR é a indústria de combustíveis^[74]. O Brasil utiliza vários tipos de combustíveis, tais como gasolina, óleo diesel, biodiesel, álcool, gás natural, gás liquefeito de petróleo (GLP) e gás metano veicular, que já vem sendo comercializado em cidades das regiões Sul, Sudeste e do Nordeste. Entre estes, o óleo diesel é economicamente muito importante, devido ao modelo energético brasileiro, no qual predomina o transporte de carga por rodovias em veículos pesados que utilizam este combustível. Por esse motivo, no nosso país, o principal produto do fracionamento do petróleo é o óleo diesel.

A Figura 16 ilustra uma amostra de óleo diesel acondicionada num recipiente de vidro, bem como as estruturas químicas dos hidrocarbonetos presentes.

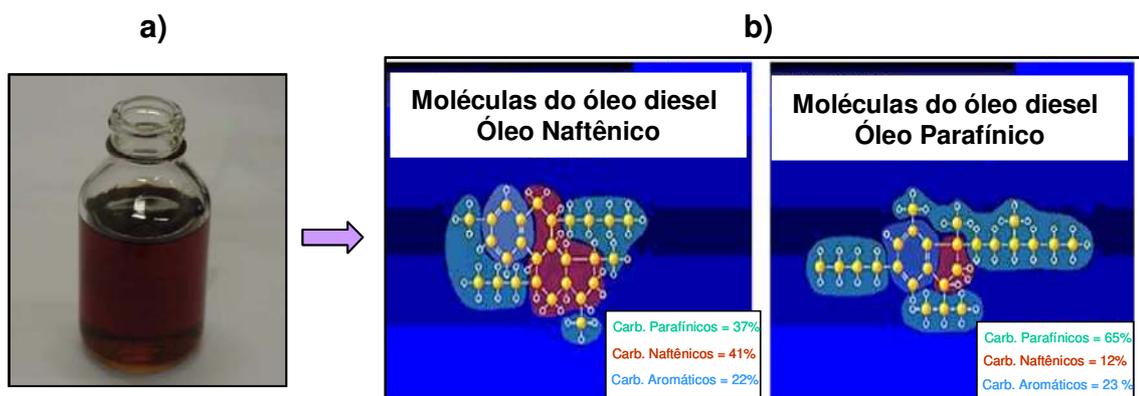


Figura 16. Amostra de óleo diesel acondicionada em recipiente de vidro (a) e as estruturas químicas dos hidrocarbonetos presentes nesse combustível (b).

Os combustíveis normalmente utilizados nos motores diesel brasileiros são constituídos de misturas de hidrocarbonetos, com uma faixa de pontos de ebulição variando aproximadamente entre 150°C e 370°C, o que corresponde aos destilados intermediários do petróleo, que destilam após o querosene e assemelham-se aos gasóleos mais leves [75].

A composição química do óleo diesel é muito variável no que diz respeito à distribuição desses hidrocarbonetos, que podem ser classificados em três tipos: parafínicos, naftênicos e aromáticos para os produtos de destilação direta, aparecendo ainda as olefinas quando o óleo diesel contém, também, produtos de craqueamento. A predominância de um ou outro tipo depende do petróleo que deu origem ao combustível e do processamento e tratamento que ele sofreu na refinaria. A massa molecular média é também variável, mas, geralmente, o óleo diesel apresenta hidrocarbonetos variando de nove a vinte átomos de carbono. A distribuição dos hidrocarbonetos dessas quatro categorias (parafínicos, naftênicos, aromáticos e olefínicos) representadas na figura 16 é muito importante, porque as propriedades do combustível são funções dela. Assim, a relação carbono/hidrogênio, tão importante na combustão, é variável, sendo maior nos hidrocarbonetos aromáticos e menor nos parafínicos. Propriedades como a densidade, faixa de destilação, poder calorífico, viscosidade, ponto de fluidez, número de cetano, etc., variam com a distribuição dos diversos hidrocarbonetos [75].

De acordo com a Portaria Nº 310 da ANP, de 27 de dezembro de 2001 [76], o óleo diesel automotivo classifica-se, conforme sua aplicação, em Óleo Diesel Automotivo

Metropolitano (D) e Óleo Diesel Automotivo Interior (B). O tipo B é o mais usado em todo o país, exceto em regiões metropolitanas, por ter maior teor de enxofre e outras especificações que levam a maior emissão de poluentes que o tipo D. Este último é exigido em quatorze regiões metropolitanas de grande população. Além do óleo diesel automotivo também existe o diesel marítimo (com um tipo especial para navios militares) e o tipo padrão usado em testes de motores.

Na certificação do óleo diesel, os parâmetros de controle de qualidade considerados são: teor de enxofre, temperaturas de destilação, corrosividade ao cobre, água e sedimentos, índice de cetano, teor máximo de cinzas, densidade, resíduos de carbono e viscosidade. Para determinação destes parâmetros de qualidade são empregados os métodos de referência estabelecidos nas normas da ASTM (do inglês, American Society for Testing and Materials), conforme indicado a seguir juntamente com breves comentários sobre a relevância dos mesmos.

◆ **Enxofre:** Método ASTM D 1266/2622/4294

Os compostos de enxofre são indesejáveis, pois conferem ao produto corrosividade e odor desagradável.

Durante a operação intermitente do motor, ou quando ele opera a baixas temperaturas, a condensação de umidade em contato com o dióxido (SO_2) e o trióxido (SO_3) de enxofre formados na combustão, pode conduzir à formação de ácidos sulfuroso (H_2SO_3) e sulfúrico (H_2SO_4), ambos corrosivos.

◆ **Curva de Destilação:** Método ASTM D 86

As características de volatilidade do óleo diesel exercem grande influência sobre o seu desempenho e tais características são determinadas através do ensaio de destilação. O ponto inicial de ebulição e as temperaturas de 50% e 85% destilados são usados para caracterizar a volatilidade.

◆ **Corrosividade ao Cobre:** Método ASTM D 130

A tendência corrosiva de um combustível depende do teor de enxofre total. Na forma elementar e de alguns compostos como sulfetos e mercaptanas, o enxofre ataca o cobre presente no sistema do motor.

◆ Água e Sedimentos: Método ASTM D 1796

A água pode congelar (nas regiões frias) obstruindo filtros e canalizações estreitas ou provocar ferrugem nos motores. Também pode provocar irregularidades na queima quando ela vaporiza nas linhas de alimentação do combustível. Os sedimentos consistem geralmente de material carbonáceo, metais ou outros materiais inorgânicos. Os sedimentos podem ocasionar desgaste e criar depósitos, tanto no sistema de injeção quanto no próprio motor.

◆ Índice de Cetano: Método ASTM D 4737

É um parâmetro calculado para estimar o número de cetano cuja determinação pelo ensaio motor requer equipamento especial de alto custo, demanda tempo e mão de obra especializada. O número de cetano é uma medida da qualidade de ignição do óleo diesel, determinado através de teste com motor padrão de um único cilindro (CFR).

◆ Teor Máximo de Cinzas: Método ASTM D 482

A cinza é constituída de material não combustível sob a forma de sólidos e compostos metálicos solúveis. Pode estar presente como areia, argila, óxidos de ferro e alumínio que aparecem após a combustão. Os componentes do sistema de injeção e motor são fabricados com precisão e tolerância mínima a esses materiais abrasivos.

◆ Densidade: Método ASTM D 1298/ D 4052

É uma propriedade importante do combustível, pois está diretamente ligada ao teor energético total contido em uma determinada massa ou volume do mesmo e pode indicar a presença de contaminação.

◆ Resíduo de Carbono: Método ASTM D 524

O termo “Resíduo de Carbono” designa o resíduo formado durante a evaporação e pirólise de um produto de petróleo. A medida deste parâmetro permite fazer uma previsão grosseira da tendência de um óleo diesel para formar depósitos carbonosos no motor.

◆ **Viscosidade:** Método ASTM D 445

Tem efeito no desempenho do sistema da bomba e injetor. Uma viscosidade muito baixa pode exigir uma manutenção mais freqüente do sistema de injeção enquanto uma viscosidade muito alta poderá provocar pressões excessivas neste mesmo sistema.

Com a introdução da análise NIR de modo on-line durante o processo de produção do óleo diesel em refinarias e devido à rapidez com que os resultados são obtidos, pode-se inclusive usar tubulações de retorno e reaproveitar o produto cujos parâmetros de qualidade não atendam às normas estabelecidas no processo, evitando que o óleo diesel de má qualidade seja entregue nas distribuidoras de combustíveis.

1.3.2 - Análise NIR de Madeira de Eucalipto.

Outra aplicação da espectroscopia NIR, juntamente com as técnicas multivariadas de análise matemática, que tem ganhado destaque no âmbito nacional é na área de controle de qualidade de madeira de eucalipto para indústria de papel e celulose.

Os materiais lignocelulósicos são compostos basicamente por celulose, hemicelulose, lignina e constituintes em menores quantidades^[77]. A Tabela 2 apresenta a composição química aproximada das madeiras.

Tabela 2. Composição química aproximada das madeiras.

Componente	Concentração (%)
1 - Celulose	~50
2 - Hemicelulose	~20
3 – Lignina	15 a 35
4 – Menores Constituintes	≤ 10

A celulose, principal componente da parede celular da fibra da madeira e de fórmula química $(C_6H_{10}O_5)_n$, é um polissacarídeo linear, constituído por um único tipo de unidade de açúcar. A molécula da celulose pertence à função química dos carboidratos ou mais especificamente a dos glicídeos^[77]. As fórmulas estruturais da celulose são apresentadas na figura 17.

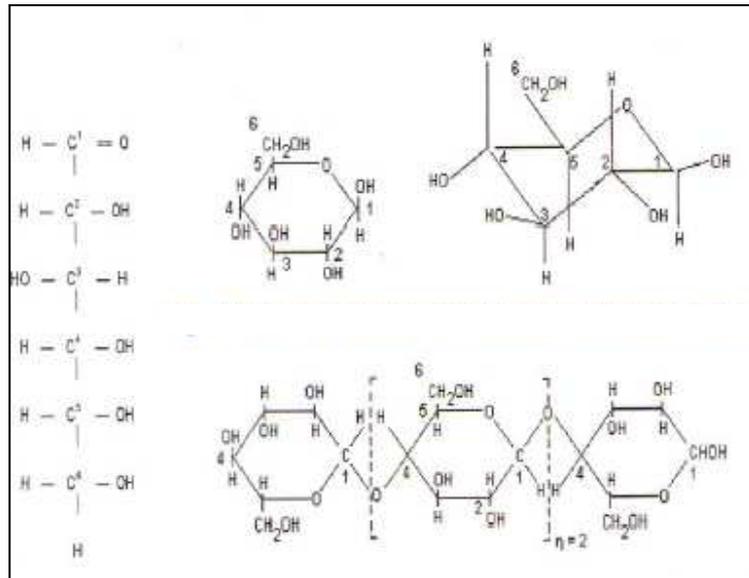


Figura 17. Diversas fórmulas estruturais da celulose: unitárias e polímero.

As hemiceluloses também são polissacarídeos de baixa massa molecular, os quais estão intimamente associados com a celulose nos tecidos das plantas^[77]. Porém, diferem da celulose por serem constituídas de vários tipos de unidades de açúcar, além de serem polímeros ramificados e de cadeia curta. As fórmulas estruturais das hemiceluloses são apresentadas na figura 18.

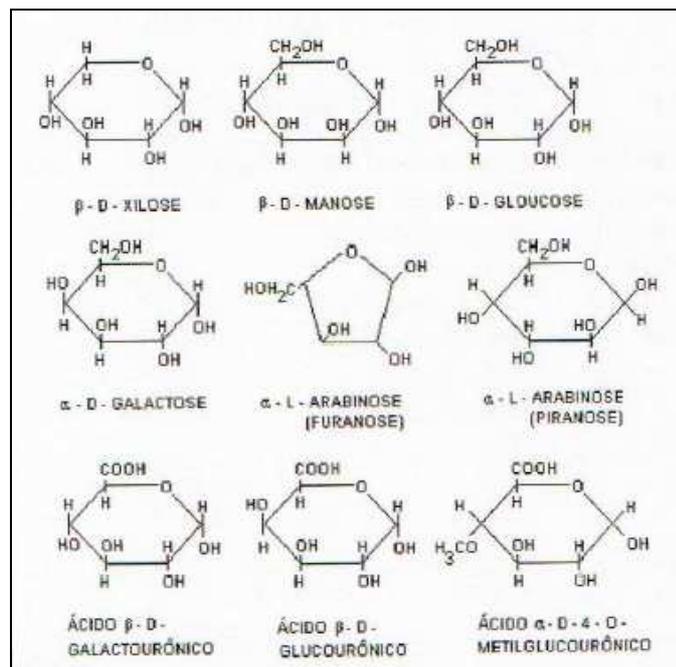


Figura 18. Fórmulas estruturais de algumas hemiceluloses.

A lignina é uma substância química que confere rigidez à parede da célula e, nas partes da madeira, age como um agente permanente de ligação entre as células, gerando uma estrutura resistente ao impacto, compressão e dobra. A base estrutural da macromolécula da lignina é apresentada na figura 19.



Figura 19. Base estrutural da macromolécula da lignina.

Por promover a opacidade e aspecto amarelo ao papel branco de escritório, constituindo assim propriedades indesejáveis, as indústrias de papel e celulose se preocupam também com o desenvolvimento de pesquisas na área de engenharia florestal voltada para a plantação de árvores de eucalipto cuja quantidade de lignina seja a mínima possível. Vale lembrar que a lignina não pode ser totalmente eliminada de uma árvore de eucalipto porque ela é o constituinte responsável por dar firmeza à árvore no solo.

A grande economia de reagentes poluentes empregados na etapa de branqueamento da polpa do eucalipto é atingida porque tais reagentes só serão adicionados no processo em quantidades adequadas e levando-se em consideração a composição química dos lotes de cavacos de eucalipto definida na etapa inicial de registro dos espectros NIR. A figura 20 apresenta uma amostra de serragem de madeira de eucalipto, acondicionada em recipiente de vidro.



Figura 20. Serragem de madeira de eucalipto acondicionada em recipiente de vidro.

Portanto, as grandes contribuições da espectroscopia NIR com calibração multivariada dedicadas aos dados de madeira de eucalipto concentram-se no aspecto econômico e preservação do meio ambiente, uma vez que a rapidez com que os resultados são obtidos ainda na etapa de corte das árvores, permite que o processo de cozimento da polpa do eucalipto seja otimizado em curtos intervalos de tempo e empregando apenas a quantidade necessária de reagentes danosos ao meio ambiente.

Capítulo 2

Objetivos

“O insucesso é apenas uma oportunidade para
recomeçar de novo com mais inteligência.”

Henry Ford

2 - Objetivos

No âmbito geral deste trabalho, foram desenvolvidas e aplicadas ferramentas quimiométricas de seleção de amostras de calibração e de variáveis, a partir do Algoritmo das Projeções Sucessivas (APS), bem como utilizada a transformada wavelet, para resolver problemas de remoção de ruído associado a dados de espectroscopia NIR e possibilitar a utilização de modelos de calibração baseado em MLR para estimativa de parâmetros certificadores de qualidade de óleo diesel combustível e madeira de eucalipto.

De forma mais específica, os objetivos deste projeto de doutorado obedecem, de modo não seqüencial aos seguintes itens:

- a) Utilização de técnicas de seleção de variáveis espectrais e amostras de calibração através do Algoritmo das Projeções Sucessivas (APS), em modelos MLR para a estimativa dos diversos parâmetros de qualidade de óleo diesel e madeira de eucalipto;
- b) Utilização de técnicas de seleção de amostras representativas, a partir do algoritmo de Kennard-Stone (KS) para auxiliar no corte de madeiras;
- c) Emprego de transformada wavelet na construção dos modelos de Regressão Linear Múltipla (MLR) para a estimativa dos parâmetros de qualidade de óleo diesel e madeira de eucalipto;
- d) Uso de transformada wavelet para filtragem de sinais espectrais NIR médios a partir da estatística de varreduras individuais.

Capítulo 3

Experimental

“São fúteis e cheias de erros as ciências que não
nasceram da experimentação, mãe de todo
conhecimento.”

Leonardo da Vinci

3 - Experimental

3.1 - Materiais e Métodos

As amostras de óleo diesel combustível dos tipos B (interior) e D (metropolitano), foram fornecidas pela Central Analítica do Instituto de Química da UNICAMP (IQM-UNICAMP), cujo serviço de monitoramento da qualidade de gasolina, óleo diesel e álcool etílico anidro carburante (AEAC) é realizado em postos da região metropolitana e circunvizinhanças da cidade de Campinas-SP.

Os parâmetros de qualidade das amostras adquiridas, foram determinados e fornecidos pela Central Analítica, segundo as normas ASTM^[78-81] que regem a análise de cada parâmetro, constituindo com isso a matriz de variáveis dependentes (matriz Y).

Elas foram acondicionadas em recipientes de vidro âmbar, com capacidade de 100 mL e conservadas em geladeira sob temperatura de aproximadamente 8°C.

Os espectros (Matriz X) e os parâmetros de qualidade (Matriz Y) das amostras de madeira de eucalipto foram cedidas por empresas de Celulose e Papel instaladas no Brasil.

3.2 - Equipamentos e Acessórios

3.2.1 – Equipamento e Sistema de registro dos espectros

O registro dos espectros de absorção das amostras (Matriz X) de diesel foi realizado utilizando-se o Espectrofotômetro NIR com Transformada de Fourier da marca Bomem, modelo MB 160 D, fonte a base de quartzo-halogênio e detector a base de Arseneto de Gálio e Índio (InGaAs). Tal instrumento opera na região espectral de 714,29 a 2.631,58 nm (nanômetros) e é interfaceado a um microcomputador que foi usado para aquisição dos dados.

Cada espectro arquivado no computador foi obtido como uma média de 100 varreduras, com uma resolução de 2 cm^{-1} , numa temperatura de cerca de 23°C com cerca

de 34% de umidade relativa do ar, registrando-se o espectro da cubeta vazia como branco a cada análise de cerca de 50 amostras de diesel.

3.2.2- Regiões NIR estudadas em função das cubetas empregadas

Para a análise das amostras de óleo diesel foram empregadas três tipos de células (cubetas) de fluxo, uma vez que não foi possível obter espectros de óleo diesel em toda a região NIR (750 – 2.500 nm) com apenas uma cubeta e com a mesma relação sinal-ruído em toda a faixa espectral em estudo. As cubetas de caminho óptico de 50 mm e 1 mm foram construídas no IQ-UNICAMP e são de material a base de vidro. A cubeta com caminho óptico de 10 mm é de material a base de quartzo e foi adquirida da empresa Hellma^[82]. A figura 21 descreve as células de fluxo empregadas no registro dos espectros, bem como as regiões NIR abrangidas com o uso de cada uma das cubetas.

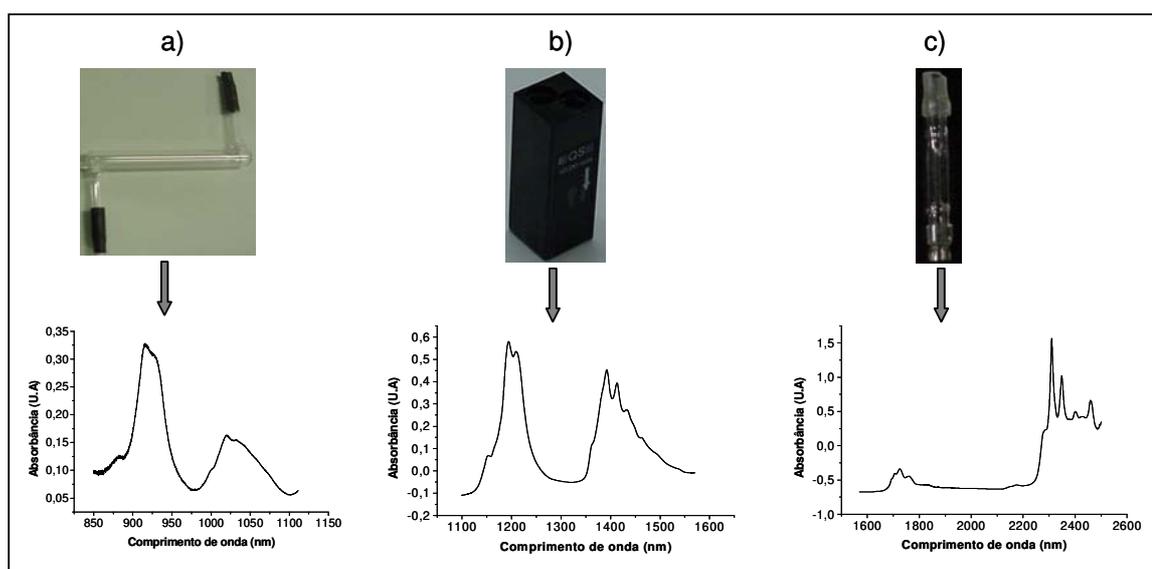


Figura 21. a) Região espectral de 850-1.100 nm usando cubeta de vidro de 50 mm, b) Região espectral de 1.100-1.570 nm usando cubeta de quartzo de 10 mm, c) Região espectral de 1.570-2.500 nm usando cubeta de vidro de 1 mm.

As sub-regiões NIR mostradas pela figura 21 possuem informações de absorção dos diversos tipos de hidrocarbonetos presentes no diesel, tais como: parafínicos, naftênicos, aromáticos e olefínicos, sendo a composição química do diesel relacionada

com os parâmetros de qualidade empregando modelos de calibração multivariada. A tabela de absorções NIR localizada no “Apêndice A” desta tese resume os modos vibracionais observados na região NIR^[83] para as ligações dos diversos tipos de moléculas existentes tanto nas amostras de diesel quanto nas serragens das madeiras de eucalipto.

Cada uma das amostras de óleo diesel analisadas foi submetida ao registro de três espectros médios, um para cada sub-região NIR. Portanto, cada amostra foi analisada três vezes.

As regiões espectrais abrangidas pelo uso de cada uma das cubetas de fluxo na análise das amostras de óleo diesel, encontram-se descritas na tabela 3.

Tabela 3. Caminhos ópticos das celas de fluxo empregadas e faixa espectral obtida com o registro dos espectros com o uso das respectivas células.

Caminho óptico (mm)	Região Espectral (R.E.) em nm
50	R.E.1 = 850 – 1.100
10	R.E.2 = 1.100 – 1.570
1	R.E.3 = 1.570 – 2.500

As matrizes de dados (amostras x comprimentos de onda) foram então organizadas em função da faixa espectral estudada, bem como foram construídos modelos quimiométricos para cada uma das respectivas sub-regiões.

Para tornar rápida a análise das amostras de diesel, bem como possibilitar melhor reprodutibilidade do registro dos espectros, foi montado um sistema simples de análise de amostras em fluxo, para cada uma das celas empregadas, conforme é ilustrado na figura 22.

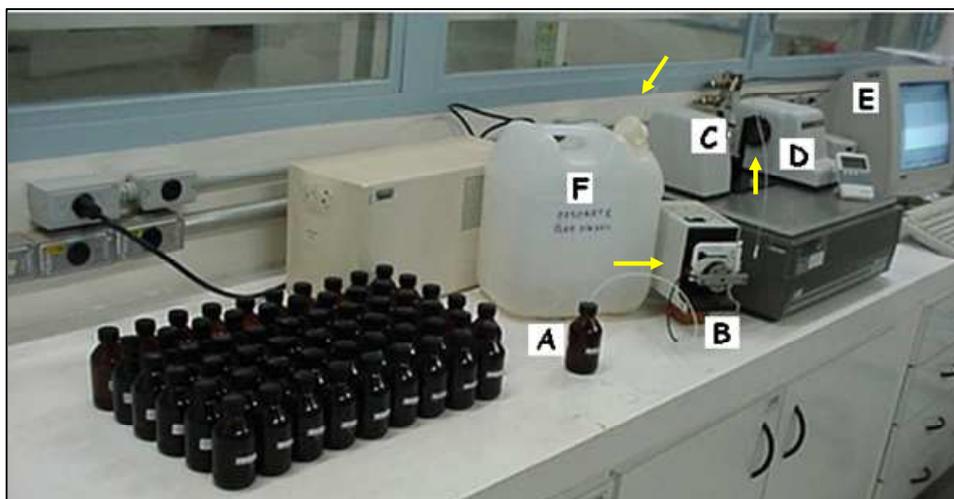


Figura 22 .Sistema montado para a aquisição dos espectros das amostras de óleo diesel, usando a cubeta de fluxo de 10 mm de caminho óptico. As setas na cor amarela indicam o sentido do fluxo da alíquota de diesel.

Conforme verificado na figura 22, a alíquota da amostra a ser analisada é aspirada do recipiente (A) e é transportada, através da bomba peristáltica (B) até a célula (C). O espectrômetro (D) registra o espectro da amostra com a bomba peristáltica parada, enquanto o microcomputador (E) arquiva os dados da análise e em seguida a bomba é acionada novamente para levar a alíquota analisada até o descarte (F).

Para melhor visualização da localização da célula, a figura 23 descreve onde a alíquota da amostra foi introduzida, no momento do registro do espectro.

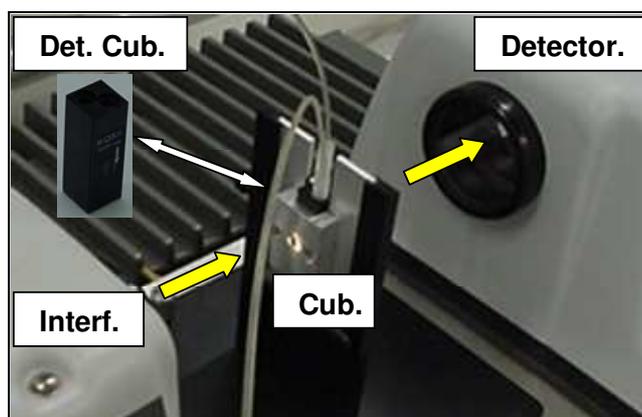


Figura 23 .Detalhe da célula de fluxo (Cub.) de 10 mm empregada no registro do espectro. As setas de cor amarela indicam o sentido do feixe de radiação NIR.

A radiação NIR, que sai do interferômetro, interage com a alíquota da amostra localizada na célula de fluxo (Cub.) e a intensidade de radiação amortecida é registrada no detector. A alíquota da amostra permanece parada dentro da célula (Det. Cub.) até que as 100 varreduras do espectro sejam realizadas.

Com o sistema montado desta forma, melhorias relacionadas à repetitividade na aquisição dos espectros NIR são obtidas, uma vez que a célula de fluxo se mantém na mesma posição, desde a aquisição do espectro do branco até o registro dos espectros das amostras.

Sistemas de fluxo semelhantes aos da figura 23 também foram montados, para aquisição dos espectros com células de 50 mm e 1 mm de caminho óptico, que podem ser visualizados nas figuras 24 e 25.

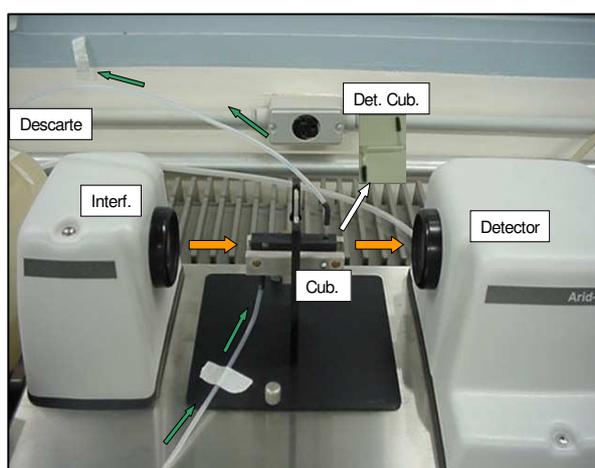


Figura 24. Detalhe do sistema com a cubeta de fluxo (Cub.) de 50 mm. As setas de cor laranja indicam o sentido do feixe de radiação NIR. As setas de cor verde indicam o sentido do fluxo da alíquota de óleo diesel.

Com o sistema da figura 24 montado, consegue-se obter espectros na região de comprimentos de onda de 850 – 1.100 nm. Já na figura 25, com o sistema de fluxo montado para o emprego de célula de 1 mm de caminho óptico, foram obtidos espectros na região de 1.570 – 2.500 nm.

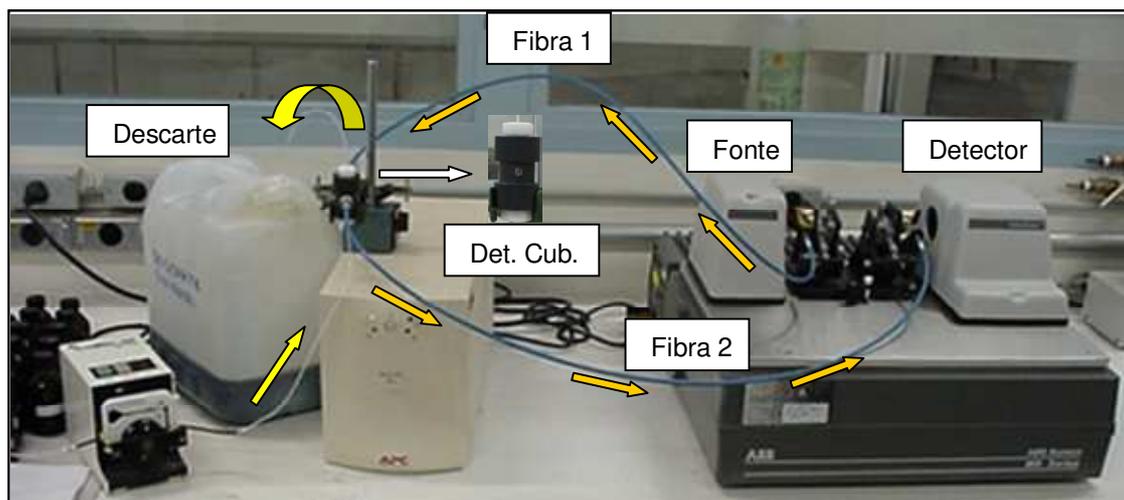


Figura 25. Detalhe do sistema com a cubeta de fluxo (Cub.) de 1 mm. As setas de cor laranja indicam o sentido do feixe de radiação NIR. As setas de cor amarela indicam o sentido do fluxo da alíquota de óleo diesel.

O sistema descrito na figura 25 é provido de espelhos toroidais recobertos com ouro, que são responsáveis por focar o feixe de radiação NIR que sai do interferômetro para a fibra óptica, bem como da fibra óptica para o detector.

A fibra óptica número 1 (Fibra 1) transporta o feixe de radiação NIR advinda da fonte que interage com a alíquota da amostra localizada na célula e a intensidade de radiação amortecida é transportada pela fibra óptica número 2 (Fibra 2), até ser registrada no detector.

As regiões espectrais estudadas por consequência do uso das células de fluxo serão detalhadas no item “Regiões NIR estudadas”.

Para as amostras de madeira de eucalipto, apesar de os dados terem sido fornecidos por diversas empresas, o equipamento empregado é semelhante ao existente no laboratório do Grupo de Instrumentação e Automação (GIA-IQM-UNICAMP) e ilustrado nas figuras 22-25, com a característica de usar um assessorio para reflectância difusa destinada a análise de sólidos, visualizado na figura 26.

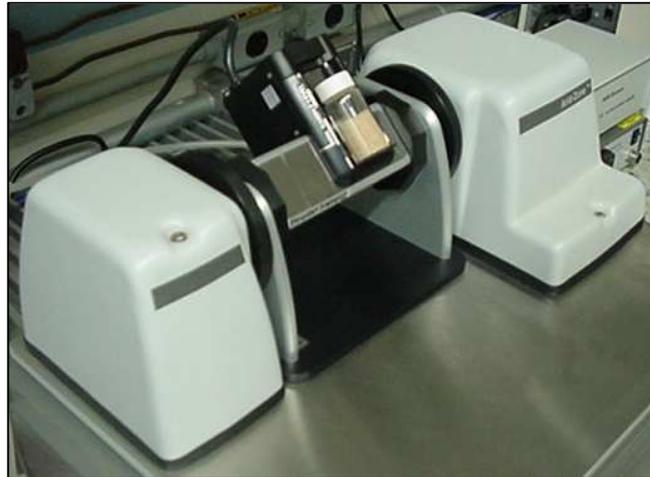


Figura 26: Sistema semelhante ao existente nas empresas de celulose e papel, montado no laboratório GIA para a aquisição dos espectros de reflectância das amostras de madeira de eucalipto.

Os espectros de reflectância difusa das amostras de madeira de eucalipto foram registrados na região espectral de 1.100 – 2.500 nm, com resolução de 2 nm e 50 varreduras para cada espectro. Como branco foi registrado o espectro de um material que reflete cerca de 99% da radiação NIR e que é geralmente feito de um material a base de Poli-Tetra-Flúor-Etileno (PTFE).

Para a aplicação do estudo referente a madeira de eucalipto foi explorada a região de 1.100 – 2.500 nm e um espectro de reflectância difusa representativo de uma amostra de madeira de eucalipto encontra-se na figura 25.

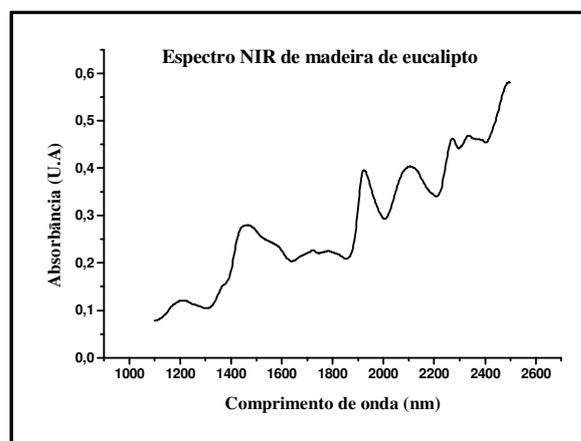


Figura 27. Espectro NIR de reflectância de madeira de eucalipto, na região espectral de 1.100 – 2.500nm.

No âmbito da calibração multivariada, tanto para as amostras de óleo diesel quanto para madeira de eucalipto, a informação química contida nos espectros NIR é relacionada aos parâmetros de qualidade investigados neste estudo, que se encontram na Tabela 4:

Tabela 4. Unidades de medida e os diversos Parâmetros de Qualidade estudados para as matrizes de Óleo Diesel e Madeira de Eucalipto.

Amostra	Parâmetro de Qualidade	Unidade	Norma/Referência
Óleo diesel	^a Índice de Cetano	--	ASTM D-4737 ^[78]
	Densidade	g.L ⁻¹	ASTM D-4052 ^[79]
	Teor de Enxofre	% m/m	ASTM D-4294 ^[80]
	Temperatura de 50 % recuperado	°C	ASTM D-86 ^[81]
	Temperatura de 85% recuperado	°C	ASTM D-86 ^[81]
Madeira de Eucalipto	Densidade	kg/m ³	Tappi T258 cm-94 ^[84]
	Teor de Extrativos	%	Tappi T204 cm-97 ^[85]
	Rendimento de Polpa	%	Lanna et al. ^[86]
	Teor de Holocelulose	%	Browning ^[87]
	Teor de Lignina	%	Tappi T 222 om-98 ^[88]

a - Índice de Cetano é adimensional.

Existem mais parâmetros de qualidade analisados tanto para óleo diesel quanto para madeira de eucalipto, porém para fins de aplicações do uso das técnicas quimiométricas de seleção de variáveis e de redução de ruído para calibração multivariada estudadas nesse trabalho, os parâmetros de qualidade aqui investigados foram considerados suficientes.

3.3 – Software e Hardware

Nas etapas de construção e validação dos modelos quimiométricos baseados em mínimos quadrados parciais (PLS-1) foi utilizado o programa UNSCRAMBLER[®], versão 9.2^[89]. e na etapa de desenvolvimento dos algoritmos para seleção de variáveis espectrais, amostras de calibração e redução de ruído, empregou-se o ambiente MATLAB[®] for Windows versão 6.5^[49]. Para a aquisição dos dados referente aos espectros foi empregado um micro computador celeron, com velocidade de processamento de 450 MHz, 128 Mb de memória RAM. Na etapa de processamento dos cálculos quimiométricos

foi utilizado um microcomputador Dell-Pentium IV com velocidade de processamento de 3.0 GHz e com 1.000 Mb de memória RAM.

3.4– Calibração Multivariada

Para estimar os parâmetros de qualidade de óleo diesel e madeira de eucalipto a partir da espectrometria NIR, fez-se necessário a construção de modelos de calibração multivariada, cuja função básica é relacionar as respostas instrumentais, nesse caso espectros, de cada uma das amostras, denominada de matriz X, com cada um dos parâmetros de qualidade das amostras de óleo diesel e madeira de eucalipto, denominada de matriz Y.

Neste trabalho, foram construídos modelos quimiométricos por PLS-1 e MLR^[6], e este último modelo foi elaborado empregando o APS como técnica de seleção de variáveis e amostras de calibração^[47].

3.4.1 - Pré-Tratamento dos Dados

A fim de tornar os dados melhor condicionados para utilização dos modelos quimiométricos, fez-se necessário o pré-tratamento associado as amostras (linhas) e as variáveis (colunas) das matrizes de dados.

Para os dados de óleo diesel e com o intuito de minimizar o ruído instrumental, uma vez que a resolução empregada foi de 2 cm^{-1} , realizou-se uma suavização pelo método de Savitzky-Golay^[90] aplicada às amostras, utilizando-se polinômio de segunda ordem. A janela de suavização a empregar dependeu da região espectral explorada, conforme a tabela 5.

Tabela 5: Número de pontos a esquerda e a direita do ponto médio empregados na suavização dos espectros, em função da região espectral explorada, para as amostras de óleo diesel.

Região Espectral (nm)	Caminho óptico (mm)	Nº de pontos da janela
850 – 1.100	50	40
1.100 – 1.570	10	40
1.570 – 2.500	1	10

No pré-tratamento associado às amostras, empregou-se também primeira derivada^[7] como forma de corrigir deslocamento sistemático da linha base, já que o espectrômetro empregado é de feixe simples e há uma pequena variação de índice de refração entre as amostras.

O método de pré-processamento empregado nas variáveis (colunas) da matriz consistiu apenas na centralização dos dados na média^[7], uma vez que dados espectrais possuem variáveis contínuas (comprimentos de onda) e não se justifica utilização de auto-escalamento ou qualquer tipo de pré-tratamento.

Para a aplicação as amostras de madeira de eucalipto, não foi realizada nenhum tipo de suavização preliminar dos espectros, uma vez que a resolução de 2 nm empregada no registro desses espectros proporcionou uma significativa redução da intensidade de ruído já na etapa experimental de análise. Porém os procedimentos de centralização na média e primeira derivada foram também aplicados a esses dados.

As operações de pré-processamento descritas acima serviram tanto para os modelos PLS-1 quanto MLR.

3.4.2 - Conjunto de amostras

A fim de verificar a robustez dos modelos de calibração empregados, foram utilizados três conjuntos de amostras na construção e validação dos modelos PLS-1 e MLR. O método de seleção de amostras empregado foi baseado no algoritmo de Kennard-Stone^[45], escolhendo-se cerca de 50% das amostras para calibração (cal.) e os 50% restante divididos igualmente para as amostras de validação (val.) e previsão (prev.)

Todos os modelos empregaram validação por série de teste e a quantidade de amostras a compor cada um dos conjuntos é descrita na tabela 6.

Tabela 6. Número de amostras de óleo diesel em cada um dos conjuntos de calibração, validação e previsão, em função da região espectral e do parâmetro de qualidade empregadas nos modelos PLS e MLR.

Parâmetro de Qualidade	Calibração (cal.)			Validação (val.)			Previsão (prev.)		
	^a R.E. 1	^b R.E. 2	^c R.E. 3	R.E. 1	R.E. 2	R.E. 3	R.E. 1	R.E. 2	R.E. 3
Cetano	137	137	97	68	68	48	68	68	48
Densidade	137	137	97	68	68	48	68	68	48
Enxofre	85	85	63	42	42	32	43	43	31
T - 50%	137	137	97	68	68	48	68	68	48
T - 85%	137	137	97	68	68	48	68	68	48

a- R.E. 1 = Região espectral de 850 – 1.100 nm, b- R.E. 2 = Região espectral de 1.100 – 1.570 nm

c- R.E. 3 = Região espectral de 1.570 – 2.500 nm.

Como pode ser observado na tabela 6, para as amostras de óleo diesel, o número de amostras usadas nos modelos de calibração para o parâmetro enxofre é menor quando comparado aos outros parâmetros investigados. Isso se deve ao fato de que no momento da obtenção das amostras, o equipamento empregado nesta determinação passou por um curto período de manutenção. Ainda assim o número de amostras da matriz de dados é considerado representativo.

Para a região espectral de 1.570 – 2.500 nm (R.E. 3) foram também analisadas menos amostras do que em relação as regiões de 850 – 1.100 nm (R.E. 1) e 1.100 – 1.570 nm (R.E. 2), uma vez que foram retiradas amostras da matriz de dados cujos espectros possuíam alta variabilidade durante as 100 varreduras. Isso se deve ao fato de não haver reprodutibilidade no posicionamento das fibras ópticas ao se analisar amostras em dias diferentes e cuja posição das fibras ópticas não foi reprodutível. Contudo o número de amostras para calibração, validação e previsão é também considerado representativo.

Para as amostras de madeira de eucalipto, não houve problemas relacionados à utilização de fibras ópticas no registro dos espectros NIR e na tabela B1 localizada no apêndice desta tese encontra-se conteúdo semelhante ao da tabela 6.

Capítulo 4

Resultados e Discussão

“Nos campos da observação, o acaso favorece apenas as mentes preparadas.”

Louis Pasteur

4 – Resultados e Discussão

4.1 - Modelos de calibração por Mínimos Quadrados Parciais (PLS-1)

A fim de comparar a eficiência de modelos quimiométricos por Regressão Linear Múltipla (MLR) que empregam o APS como técnica de seleção de variáveis e amostras de calibração, foram construídos modelos quimiométricos baseado no método dos mínimos quadrados parciais (PLS), empregando-se o software Unscrambler, versão 9.1.

A regressão por PLS é bastante discutida na literatura^[91] e sua eficiência em dados de combustíveis já é largamente conhecida^[92]. Foram construídos modelos quimiométricos para cada um dos parâmetros de qualidade do diesel e madeira de eucalipto e foram analisados elementos estatísticos, tais como: correlação, erro médio quadrático de previsão (RMSEP), inclinação da reta (slope), tendência (bias), com o intuito de avaliar o ajuste do modelo ao parâmetro investigado.

Foram utilizados procedimentos de otimização dos modelos, tais como retirada de amostras anômalas e seleção de variáveis espectrais, conforme ilustra a figura 26.

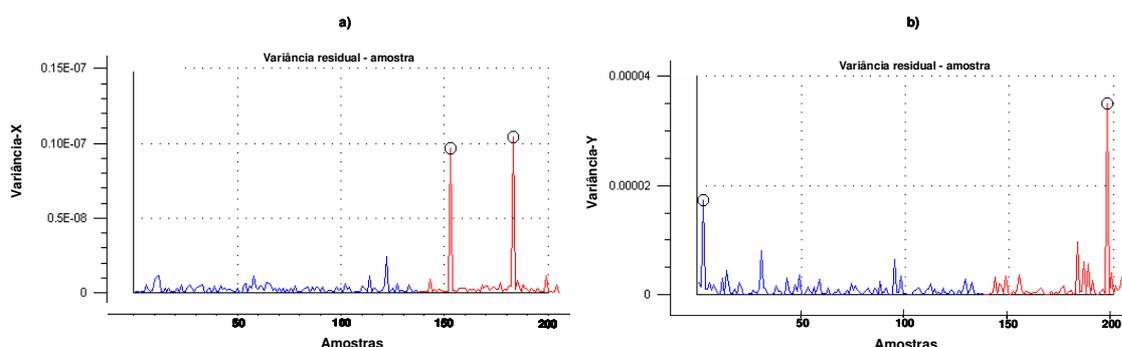


Figura 28. Gráficos de variância residual em X (a) e em Y (b) em função das amostras de calibração (azul) e validação (vermelho) para óleo diesel. Modelo PLS construído para o parâmetro densidade na região espectral de 1.100 – 1.570 nm.

Amostras de calibração (cal.) e validação (val.), como as sinalizadas por círculos na figura 6 e cuja variância residual em X e em Y são muito altas, indica que o modelo não conseguiu explicar a variância contida nelas, sendo necessário a retirada do modelo.

O procedimento de seleção de variáveis espectrais dos modelos PLS baseou-se na eliminação de comprimentos de onda que não apresentam informação em relação ao parâmetro de qualidade em estudo, informativos, conforme ilustra a figura 27 para o parâmetro densidade em óleo diesel.

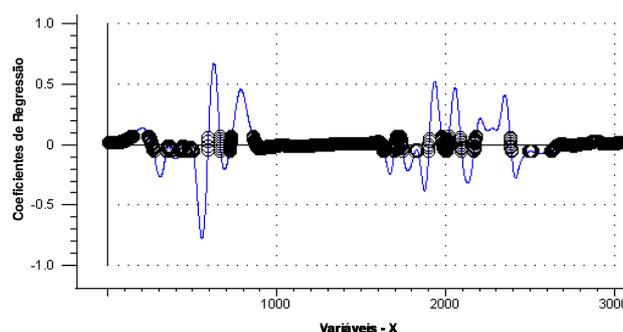


Figura 29. Gráfico dos coeficientes de regressão em função das variáveis espectrais para óleo diesel no modelo PLS aplicado ao parâmetro densidade, na região espectral de 1.100 – 1.570 nm. Variáveis destacadas em círculo foram eliminadas do modelo.

Variáveis espectrais cujos valores dos coeficientes de regressão equivale a 10% ou menos do maior valor de coeficiente encontrado no modelo^[93] foram eliminadas.

Com os procedimentos de ajuste dos modelos descritos acima, obtiveram-se elementos estatísticos otimizados, comparando-se com modelo PLS sem ajustes, conforme descrito na figura 28.

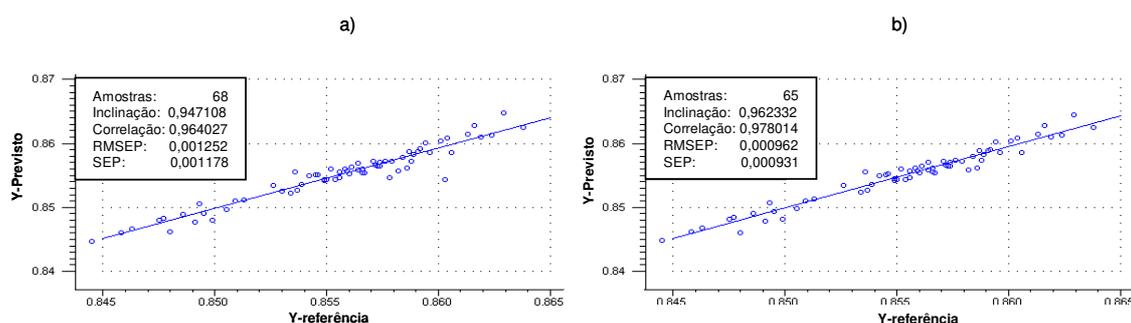


Figura 30. Gráfico dos valores previstos em função dos valores de referência, para as amostras de validação, sem (a) e com (b) otimizações de ajustes dos modelos PLS, para o parâmetro densidade. Região espectral de 1.100 – 1.570 nm; dados de Óleo Diesel.

Os procedimentos de ajuste dos modelos PLS foram efetuados em cada um dos parâmetros de qualidade investigados e para cada uma das regiões espectrais

exploradas. Os valores de RMSEP, associados à estimativa dos parâmetros de qualidade das amostras de óleo diesel obtidos dos modelos PLS, são abordados na tabela 7. Para aplicação à madeira de eucalipto os resultados encontram-se no APÊNDICE B desta tese.

Tabela 7: Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade de óleo diesel para o sub-conjunto de amostras de previsão dos modelos PLS.

Parâmetro	Unidade	ª Faixa do parâmetro	R.E. 1		R.E. 2		R.E. 3		c Repto.
			RMSEP	R	RMSEP	R	RMSEP	R	
Cetano	-	39,9 – 48,3	1,4 (4) ^b	0,5736	1,0 (12)	0,7481	1,0 (7)	0,7767	^d 2,0
Densidade	g.L ⁻¹	0,8443 – 0,8658	0,0018 (7)	0,9209	0,0009 (11)	0,9778	0,0010 (5)	0,9768	^e 0,0012
Enxofre	% (m/m)	0,05 – 0,29	0,03 (6)	0,8054	0,02 (5)	0,8749	0,02 (5)	0,9374	^f 0,05
T-50 %	°C	255,1 – 291,1	6,5 (4)	0,7080	4,8 (7)	0,8452	2,9 (10)	0,9510	^g 9,7
T-85 %	°C	319,4 – 370,1	10,5 (4)	0,4553	6,3 (10)	0,8249	8,4 (8)	0,7165	^h 10,6

^{a)} Faixa do parâmetro = Menor e maior valor para cada parâmetro de qualidade no conjunto de amostras de calibração.

^{b)} Entre parênteses encontra-se o número de V.L (variáveis latentes) empregadas.

^{c)} Repto. = Valores de reprodutibilidade dos métodos de Referência associado a análise de cada parâmetro de qualidade investigado.

Obs: Destacado em vermelho encontram-se os menores RMSEP e número de V.L empregados, bem como a maior correlação empregada na modelagem de cada parâmetro de qualidade.

Considerando os valores de RMSEP obtidos na tabela 7, é constatado que os modelos PLS que empregaram a região espectral de 1.570 – 2.500 nm (R.E. 3) fornece maiores correlações (R) e menores ou semelhantes valores de RMSEP na estimativa dos parâmetros cetano, densidade, enxofre e T-85 %, quando comparado com os modelos das faixas espectrais de 850 - 1.100 nm (R.E 1) e de 1.100 – 1.570 nm (R.E. 2). Tal resultado pode ser atribuído ao fato de que na região de primeiro sobreton e bandas de combinação de absorção de hidrocarbonetos (1.570 – 2.500 nm) os picos possuem uma relativa resolução (finos e separados) que ajudam na elucidação da relação das matrizes X e Y na modelagem por PLS e MLR.

A região espectral de 850 - 1.100 nm (R.E. 1), empregada nos modelos PLS resultou nos maiores RMSEP e menores coeficientes de correlação (R), dentre as regiões espectrais estudadas. A provável explicação para isso é que nessa região espectral há picos de absorção dos hidrocarbonetos que são altamente sobrepostos, pouco definidos e

pouco intensos, cuja razão sinal/ruído é baixa. Esta é considerada a hipótese mais provável e que justifica a obtenção de modelos PLS e MLR pouco ajustados aos parâmetros de qualidade de óleo diesel investigados neste trabalho.

Foi realizado um teste-t emparelhado para comparar as estimativas dos valores previstos em função dos valores de referência para cada um dos parâmetros de qualidade estudados, e em cada região espectral explorada. A um nível de 95 % de confiança, não houve diferença significativa entre as estimativas da técnica NIR com modelos PLS em comparação com os valores obtidos pelos métodos de referência.

4.2 – Modelos de calibração por Regressão Linear Múltipla (MLR)

O modelo de calibração multivariada baseado em MLR carrega a característica de ser mais simples e mais fácil de ser compreendido, que a regressão por PLS, cuja regressão é realizada em variáveis latentes (V.L) e que podem não ter uma interpretação física direta dos dados.

Neste trabalho foram empregados modelos MLR empregando tanto a seleção de variáveis espectrais quanto seleção de amostras de calibração, a partir do APS.

4.2.1 - Modelos MLR com Seleção de variáveis por APS

As técnicas de seleção de variáveis, quando aplicadas a modelos MLR^[33,35] de calibração multivariada, promovem efeitos semelhantes aos encontrados nos modelos PCR e PLS, tais como: remoção de colinearidade, redundância e ruído de toda a série de dados. Porém, como já mencionado em itens anteriores, o ganho no uso de MLR com técnicas de seleção de variáveis, encontra-se na simplicidade do tratamento dos dados.

A construção de fotômetros dedicados e baratos, a base de diodos emissores de luz (LED), foi grandemente favorecida com o emprego das técnicas de seleção de variáveis aplicadas a regressão MLR, pois com a não transformação dos dados de uma matriz em componentes principais, a tarefa de escolha dos comprimentos de onda mais

informativos para o parâmetro de qualidade em estudo se tornou em um procedimento simples e direto.

Para cada um dos parâmetros de qualidade do óleo diesel e de madeira de eucalipto, foram executados modelos de calibração multivariada empregando seleção de variáveis por APS seguido de modelagem por MLR e comparados com modelos PLS.

O pré-tratamento dado as matrizes desta etapa do trabalho, seguiu os mesmos procedimentos do pré-processamento dos dados para modelos PLS. Aliás, uma vez obtidas as matrizes de dados pré-processadas, no software Unscrambler, foram executados modelos PLS ainda neste software, bem como foram exportadas as matrizes pré-processadas para o software Matlab para serem submetidas ao emprego dos algoritmos que empregaram este ambiente de programação.

Os algoritmos: *selecao*, *aps*, *sumsq*, *selvaramost*, *selmist*, *centraliza* e *rlmqr* usados na etapa de seleção de variáveis, bem como de amostras de calibração multivariada, encontram-se descritos no “APÊNDICE C” desta tese, onde também são disponibilizadas tabelas com as variáveis espectrais e amostras de calibração, empregadas na construção dos modelos MLR.

Os resultados das estimativas dos parâmetros de qualidade do óleo diesel para as amostras da série de previsão, empregando seleção de variáveis por APS e regressão por MLR encontram-se na tabela 8. Resultados semelhantes para os dados de madeira de eucalipto encontram-se no “APÊNDICE J”.

Tabela 8: Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade, para o sub-conjunto de amostras de previsão dos modelos MLR, empregando seleção de variáveis, nas regiões de 850 - 1.100 nm (R.E. 1), 1.100 - 1.570 nm (R.E. 2) e 1.570 - 2.500 nm (R.E. 3).

Parâmetro	Unidade	R.E. 1		R.E. 2		R.E. 3	
		RMSEP	R	RMSEP	R	RMSEP	R
Cetano	-	1,4 (28) ^a	0,5853	1,1 (22)	0,6984	1,0 (14)	0,7636
Densidade	g.L ⁻¹	0,0023 (73)	0,8661	0,0010 (54)	0,9737	0,0009 (24)	0,9787
Enxofre	% (m/m)	0,04 (30)	0,7421	0,03 (11)	0,8520	0,02 (9)	0,9064
T-50 %	°C	6,8 (38)	0,6851	3,9 (21)	0,9032	4,8 (72)	0,8691
T-85 %	°C	9,0 (8)	0,5827	8,2 (75)	0,6823	8,3 (34)	0,7099

^{a)} Entre parênteses encontra-se o número de comprimentos de onda empregados nos modelos MLR.

A partir dos elementos estatísticos (RMSEP, R) obtidos dos modelos MLR com seleção de variáveis, indicados na tabela 8, verifica-se por motivos semelhantes aos da modelagem por PLS que a região espectral de 1.570 – 2.500 nm, cujo caminho óptico empregado foi de 1 mm (R.E.3) é mais útil para ser empregada na estimativa dos parâmetros: Índice de cetano, densidade e teor de enxofre, uma vez que resultou em modelos MLR com menor RMSEP e variáveis espectrais, assim como maiores coeficientes de correlação (R) foram obtidos.

O modelo MLR que empregou as variáveis selecionadas, a partir da região espectral de 1.100 – 1.570 nm (R.E. 2) forneceu melhor ajuste para o parâmetro T-50 % e T-85 %. Já os modelos MLR que empregaram seleção de variáveis por APS, na região de 850 – 1.100 nm (R.E. 3) não se mostrou bem ajustado para a estimativa de todos os parâmetros de qualidade investigados.

Ao se comparar os modelos MLR com seleção de variáveis, cujos resultados estão na tabela 7 com os modelos PLS cujos elementos estatísticos de ajuste encontram-se na tabela 6, e que empregaram a quase totalidade das variáveis espectrais existentes na matriz X, verifica-se uma grande redução no número de comprimentos de onda empregados nos modelos MLR. Este fato é bem notado, por exemplo, na estimativa do parâmetro T-85 %, empregando-se a região de 850 – 1.100 nm (R.E. 1), onde foram empregados no modelo PLS, 1.728 comprimentos de onda, transformados em variáveis latentes; já no modelo MLR, este número é reduzido para apenas 8 comprimentos de onda, sem a necessidade de transformação em variáveis latentes.

4.2.2 – Modelos MLR com Seleção de variáveis e amostras de calibração por APS

Técnicas de compressão de dados, cujo objetivo é a seleção de amostras em calibração multivariada, tem atraído a atenção de muitos pesquisadores^[39-46]. Tal procedimento pode salvar um trabalho experimental e computacional se a rotina de análises multicomponentes requerer o uso de um equipamento que necessite ser periodicamente recalibrado.

Seleção de amostras é também importante no contexto da transferência de modelos de calibração multivariada^[46], em que uma sub-série de amostras disponíveis deve ser medida em diferentes espectrômetros (escravo) com respeito a um único espectrômetro (mestre). Neste caso, a seleção de um pequeno número de amostras representativas reduz o custo envolvido não somente na repetição das análises, mas também no transporte físico das amostras.

Foram também construídos modelos MLR que empregaram tanto a seleção de variáveis espectrais como amostras de calibração, para cada parâmetro de qualidade de óleo diesel e de madeira de eucalipto, nas mesmas regiões espectrais dos modelos PLS e MLR já construídos. Inclusive o algoritmo “selvaramost”, que está descrito no “Apêndice C4” desta tese, teve sua rotina desenvolvida para selecionar primeiro as variáveis selecionadas e em seguida escolher as amostras de calibração mais representativas, restringidas às variáveis selecionadas.

Os resultados dos parâmetros de ajuste (RMSEP, R) dos modelos MLR com seleção de variáveis e amostras de calibração para óleo diesel, bem como valores de reprodutibilidade dos métodos de referência, faixa de valores dos parâmetros de qualidade e unidade de medida destas grandezas, encontram-se na tabela 9. Vale salientar que os comprimentos de onda empregados em cada modelo MLR são os mesmos que os empregados nos modelos MLR da tabela 7. Os resultados para a aplicação de madeira de eucalipto encontra-se no “APENDICE J” desta tese.

Tabela 9. Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade, para o sub-conjunto de amostras de previsão dos modelos MLR, empregando seleção de variáveis e amostras de calibração, nas regiões de 850 - 1.100 nm (R.E. 1), 1.100 - 1.570 nm (R.E. 2) e 1.570 – 2.500 nm (R.E. 3).

Parâmetro	Unidade	R.E. 1		R.E. 2		R.E. 3	
		RMSEP	R	RMSEP	R	RMSEP	R
Cetano	-	1,6 (12) ^a	0,3897	1,3 (3)	0,5738	1,0 (3)	0,7363
Densidade	g.L ⁻¹	0,0038 (55)	0,7575	0,0010 (42)	0,9722	0,0012 (22)	0,9639
Enxofre	% (m/m)	0,03 (3)	0,7844	0,02 (11)	0,9056	0,02 (9)	0,8707
T-50 %	°C	7,3 (10)	0,6432	4,8 (19)	0,8662	4,8 (10)	0,8698
T-85 %	°C	9,8 (8)	0,4866	10,7 (42)	0,6032	9,8 (8)	0,6833

^{a)} Entre parênteses encontra-se o número de amostras de calibração empregados nos modelos MLR.

Verifica-se na tabela 9 que em geral não houve perda da capacidade preditiva dos modelos MLR, ao empregar um número reduzido de amostras de calibração. Apenas para o parâmetro densidade e utilizando-se a região espectral de 850 – 1.100 nm (R.E. 1) o valor de RMSEP foi estatisticamente superior ao valor de reprodutibilidade do método padrão.

A região de 1.570 – 2.500 nm provou ser mais útil para os modelos MLR empregados na estimativa de Índice de cetano, enxofre, T-50 % e T-85 %, cujos RMSEP obtidos foram menores ou iguais aos da região de 850 – 1.100 nm e 1.100 – 1.570 nm.

4.2.3 - Seleção de amostras e corte de árvores de madeira de eucalipto

Outro assunto interessante e que motivou a aplicação de técnicas de seleção de amostras para calibração multivariada com uma conotação diferente da empregada no item anterior concentra-se na resolução de um dos problemas enfrentados nas indústrias de papel e celulose do Brasil e certamente do exterior, cuja intenção é selecionar para a etapa inicial de construção dos modelos multivariados, as árvores cuja composição química seja a mais variada possível e com isso podem ser selecionadas amostras mais representativas de uma determinada área plantada.



a)



b)

Figura 31. *Imagens representativas de uma plantação de eucalipto (a) e de árvores cortadas (b) para o processo de cozimento e branqueamento da polpa da madeira.*

Considerando que na etapa da derrubada das árvores só há informação referente ao espectro das amostras, uma vez que muitos parâmetros de qualidade da madeira de eucalipto só são determinados quando os cavacos das amostras de madeira são

submetidos ao processo de cozimento e branqueamento da polpa de eucalipto e determinação dos demais parâmetros no laboratório, torna-se importante o emprego de um método de seleção de amostras que leve em conta apenas a variabilidade dos espectros (Matriz X).

Um algoritmo de seleção de amostras bastante conhecido na literatura e que é empregado até os dias atuais foi desenvolvido pelos pesquisadores R. W. Kennard e L. A. Stone em 1969^[45], denominado aqui na tese de *K.S.* Trata-se de um algoritmo matemático que tem por objetivo selecionar um subconjunto representativo de um conjunto de N amostras.

Para assegurar uma distribuição uniforme do subconjunto de amostras representadas pelo espaço de dados baseado na resposta instrumental X , o KS segue um procedimento orientado no qual novas seleções são realizadas em regiões do espaço distantes das amostras já selecionadas. Para este propósito, o algoritmo emprega a distância euclideana entre os vetores \mathbf{x} de cada par (p, q) de amostras calculada como:

$$d_x(p, q) = \sqrt{\sum_{j=1}^J [x_p(j) - x_q(j)]^2}; \quad p, q \in [1, N] \quad \text{Eq.1}$$

Para dados espectrais, $x_p(j)$ e $x_q(j)$ são as respostas instrumentais nos j -ésimos comprimentos de onda para as amostras p e q , respectivamente. J representa o número de comprimentos de onda no espectro.

A seleção inicia-se escolhendo-se o par $(p1, p2)$ de amostras para as quais a distância $dx(p1, p2)$ seja a maior. Em cada iteração subsequente, o algoritmo seleciona a amostra que apresentar a maior distância em relação a alguma amostra já selecionada. Este procedimento é repetido até o número de amostras especificado pelo analista ser alcançado. A figura 30 representa o critério de escolha das amostras adotado pelo KS.

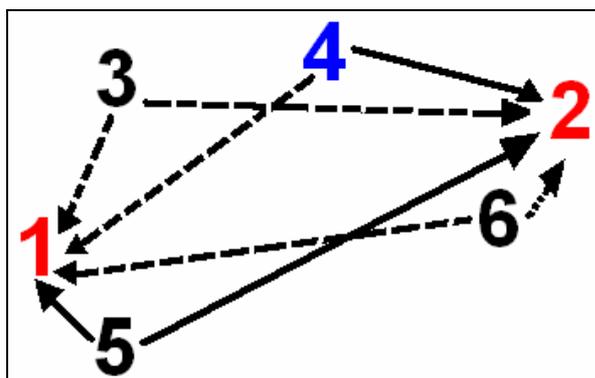
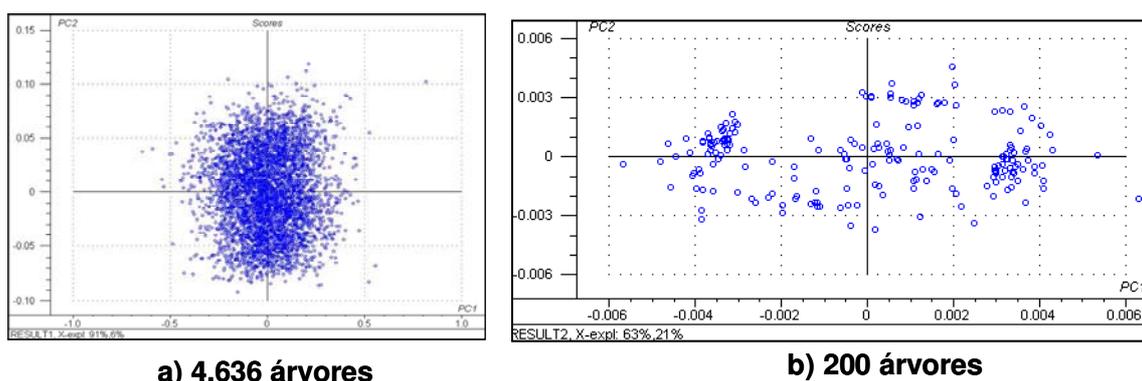


Figura 32. Exemplo de seleção por KS das 3 primeiras amostras de um conjunto de dados contendo 6 amostras. As amostras 1, 2 e 4 foram nessa ordem consideradas como as mais representativas em relação ao espaço amostral pesquisado, considerando as distâncias existentes entre cada uma delas.

O conjunto de dados referente a esta etapa do trabalho consiste de 4.636 amostras, cujos espectros NIR foram registrados na região espectral de 1.100 – 2.500 nm, com resolução de 2 nm e média de 50 varreduras. A serragem das amostras de madeira de eucalipto foi obtida realizando-se furos a uma altura de 1,5 m em cada uma das árvores e sem que as mesmas fossem derrubadas. A matriz de dados é composta por 4.636 amostras e 700 variáveis. A figura 31 ilustra um gráfico de *scores* obtidos por análise por componentes principais (PCA) de todas as amostras da matriz de dados, bem como as 200 amostras selecionadas para o corte usando o algoritmo *K.S*.



a) 4.636 árvores

b) 200 árvores

Figura 33. (a) Gráfico dos scores (PC1 x PC2) das 4.636 árvores da matriz de dados, variância explicada de 97% nas 2 primeiras PC, (b) Scores para as 200 amostras selecionadas para possível corte, variância explicada de 84% nas 2 primeiras PC (b).

Apesar de não se ter empregado análise por Componentes Principais na seleção das amostras, os gráficos dos scores da figura 31 ilustram a distribuição das 200 amostras (figura 30-b) selecionadas pelo KS e sua representatividade em relação as 4.636 amostras da figura 30-a.

A decisão de se cortar 200 árvores leva em consideração o porte do laboratório da planta química, ou seja, a sua capacidade de processar o cozimento, branqueamento e obtenção dos demais parâmetros de qualidade da madeira.

Para avaliar a representatividade das amostras selecionadas com esta intenção, foi realizado um estudo com um banco de dados de 100 amostras da mesma empresa, com os seus respectivos valores dos parâmetros de qualidade (matriz Y). Foram construídos modelos de regressão por PLS empregando o KS e o método aleatório como técnicas de seleção de amostras, para as seguintes situações abordadas na tabela 10.

Tabela 10. Características dos modelos MLR e PLS construídos para a estimativa dos parâmetros de qualidade de madeira de eucalipto.

Modelo de calibração	Método de Validação	Método de seleção de amostras	Amostras de calibração	Amostras de previsão
PLS-KSX-25	Validação cruzada completa	K.S.X.	25	75
PLS-KSX-50	Validação cruzada completa	K.S.X.	50	50
PLS-Aleat-25	Validação cruzada completa	Aleatório	25	75
PLS-Aleat-50	Validação cruzada completa	Aleatório	50	50
PLS-100	Validação cruzada completa	--	100	100

Conforme se verifica na tabela 10, Os modelos PLS empregaram diferentes quantidades de amostras de calibração, de modo a se analisar a eficiência da estimativa dos parâmetros de qualidade da madeira à medida que o número de amostras de calibração aumenta ou diminui.

Para se manter a mesma proporção entre as 200 árvores no conjunto de 4.636 amostras, deveria ter sido criado um modelo PLS com apenas 4 amostras de calibração e 96 de validação externa, porém as 4 amostras de calibração representariam muito pouco das 100 amostras do conjunto e pouca variabilidade seria encontrada neste conjunto de

calibração. Seria interessante obter um banco de dados com um número maior de amostras.

O método aleatório foi também empregado em um dos modelos PLS com o intuito de avaliar se uma amostragem aleatória traria vantagens em relação ao método *K.S.*

Os resultados de RMSEP obtidos dos modelos PLS sob as diversas condições detalhadas na tabela 10 encontram-se na tabela 11.

Tabela 11. Valores de RMSEP para os modelos PLS construídos para a estimativa dos parâmetros de qualidade de madeira de eucalipto, para as amostras do conjunto de previsão.

Parâmetro de Qualidade	Modelos					Repro. ^b	Faixa do parâmetro
	PLS-ksx (25cal)	PLS-ksx (50cal)	PLS-aleat (25cal)	PLS-aleat (50cal)	PLS (100cal)		
	RMSEP	RMSEP	RMSEP	RMSEP	RMSEP		
Lig_total (%) ^[88]	0,91 (4) ^a	0,84 (4)	1,08 (3)	0,98 (5)	0,83 (6)	1,0	24,8 – 30,9
Rend_dep (%) ^[86]	1,26 (4)	0,86 (6)	1,58 (1)	1,05 (4)	0,99 (5)	1,0	48,7 – 55,8
Siring/Guai ^[94]	0,22 (7)	0,14 (5)	0,21 (3)	0,20 (4)	0,18 (6)	0,5	2,2 – 3,2
A_consus (%) ^[95]	0,64 (4)	0,75 (5)	0,59 (3)	0,61 (3)	0,50 (8)	0,5	9,2 – 14,0
A_E (%) ^[95]	1,01 (3)	0,63 (5)	0,92 (4)	0,74 (12)	0,70 (8)	0,5	12,3 – 17,4

a) Entre parênteses encontra-se o número de variáveis latentes usadas nos modelos PLS.

b) Valores de reprodutibilidade dos métodos de referência para os parâmetros de qualidade: Lignina total (*Lig_total*), Rendimento depurado (*Rend_dep*), Siringil/Guaiacil (*Siring/Guai*), Álcali Consumido (*A_Consu*) e Álcali Efetivo (*A_E*).

Verifica-se na tabela 11 que, para os parâmetros de qualidade Álcali Efetivo (*A_E*) e Álcali Consumido (*A_consus*) os valores dos erros de previsão (RMSEP) de quase todos os modelos de regressão são maiores que os valores de reprodutibilidade dos métodos de referência. Uma provável explicação para isso encontra-se no fato de que a repetibilidade desses métodos de referência no momento da execução das análises tenha sofrido variações altas e ter proporcionado reflexos de RMSEP elevados para os modelos PLS.

Apesar de alguns modelos PLS não ter possuído eficiência inicial, deve-se levar em conta que o procedimento de seleção de amostras é apenas uma das etapas na busca por modelos mais robustos e eficientes para a estimativa de parâmetros de qualidade. Outros procedimentos de otimização de modelos de regressão multivariada, tais como: seleção de variáveis espectrais e uso de técnicas de redução de ruído ainda devem ser

empregados nos modelos de regressão multivariada, em etapas subseqüentes ao de seleção de amostras e que deverão contribuir significativamente para a redução dos RMSEP associados aos modelos.

Outro fato a se notar na tabela 11 e que confirma as expectativas é que os modelos com 25 amostras de calibração possuem RMSEP mais altos que os RMSEP dos modelos com 50 amostras de calibração. Quanto maior o número de amostras de calibração, maior a representatividade do modelo e menor deverá ser o RMSEP encontrado.

O modelo mais bem ajustado, segundo a tabela 11 foi para o parâmetro Siringil/Guaiacil (Siring/Guai), que forneceu valores de RMSEP (0,19 em média) bem abaixo do valor de reprodutibilidade (0,50) do método de referência. Tanto o valor de repetibilidade do método deve ter tido pouca variabilidade como também a faixa de variação dos valores encontrados é pequena (2,2 – 3,2).

Observações, ainda a partir da tabela 11, indicam que os métodos PLS cuja seleção de amostras de calibração se deu de forma aleatória possuíram RMSEP hora maiores, hora menores que os modelos PLS baseado no algoritmo KS para seleção de amostras. O inconveniente do método aleatório é a não representatividade de amostras que pode levar a modelos “instáveis”.

Considerando que a idéia desse estudo é a seleção de 200 árvores com propriedades físico-químicas variadas, apoiada na decisão de um algoritmo matemático (KS) juntamente com o conhecimento do analista e sem a necessidade da derrubada da árvore para o registro dos espectros, a simples aplicação de modelos de calibração multivariada possui interessante e prática aplicação ao problema levantado pelas indústrias de papel e celulose.

Lembrando que uma vez selecionadas as 200 amostras representativas e determinados os seus respectivos parâmetros de qualidade para compor o conjunto de calibração, o tempo de alguns segundos destinados à previsão dos parâmetros de qualidade das amostras de previsão que futuramente entrarão no processo de cozimento e branqueamento da planta química e empregando-se a técnica NIR, tem promissora aplicação para as indústrias de papel e celulose, já que os métodos de referência demoram horas e até dias para fornecer os valores dos parâmetros de qualidade.

4.2.4 – Filtragem Wavelet e calibração multivariada por MLR.

Nesta etapa do trabalho, foi investida a maior parte do tempo no desenvolvimento e testes dos algoritmos empregados nos modelos baseados em Regressão Linear Múltipla (MLR) com filtragem Wavelet (ver apêndice), bem como a construção dos modelos PLS utilizados como método quimiométrico comparativo.

Todos os procedimentos relacionados a aquisição, registro dos espectros (matriz X) e obtenção dos valores de referência (matriz Y) dos parâmetros de qualidade, relacionados as amostras de óleo diesel e eucalipto analisadas neste projeto, já foram detalhados no item Materiais e Métodos desta tese.

Para fins de construção dos modelos de calibração multivariada por MLR, empregando-se filtros baseados na transformada Wavelet, foram escritos em ambiente Matlab^[49] os algoritmos: Estudo, decomposição e fastselecaolight. As rotinas dos algoritmos encontram-se no “Apêndice I” desta tese e o significado dos mesmos encontra-se descrito a seguir:

- Algoritmo “decomposicao”: Usado para decompor os espectros para o domínio wavelet, usando a wavelet e o número de níveis de decomposição que são testados em cada iteração.

- Algoritmo “fastselecaolight”: Rotina baseada no algoritmo das projeções sucessivas (APS)^[35] e tem por finalidade selecionar a melhor wavelet/nível de decomposição que conduziu ao menor erro quadrático médio de previsão (RMSEP), bem como é usado também na seleção de coeficientes wavelet dentro da wavelet/nível de decomposição escolhida.

- Algoritmo “estudo”: Rotina principal da série de algoritmos empregados neste trabalho e que reúne as subrotinas: decomposição e fastselecaolight com o intuito de realizar a otimização do tipo de wavelet e nível de decomposição empregado na análise dos sinais NIR, bem como divulgar os resultados de RMSEP, correlação e demais elementos da otimização empregados neste estudo.

As wavelets empregadas no estudo de otimização foram as das famílias Daubechie e Simlet, conforme ilustra a figura 31.

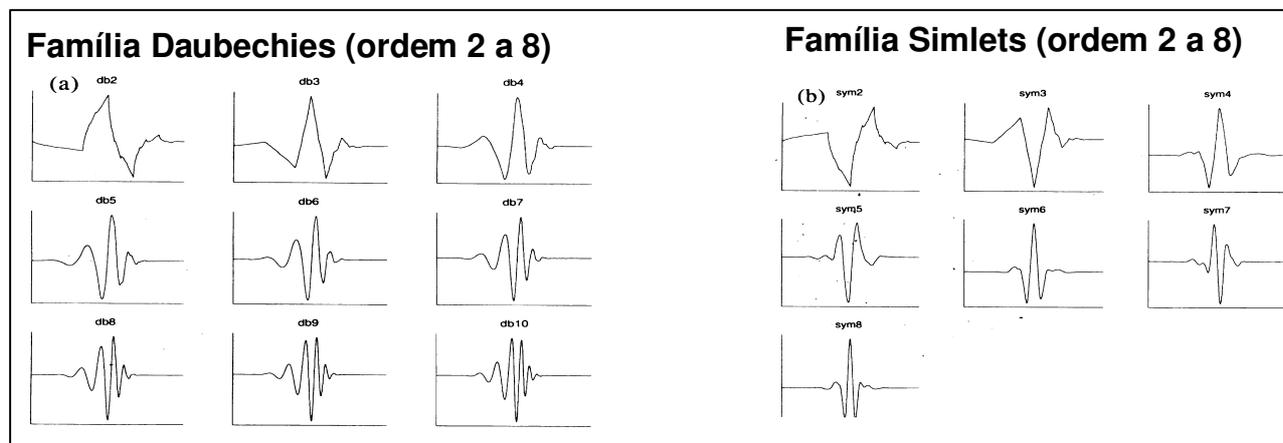


Figura 34. Tipos de família wavelet ($\Psi(t)$) empregadas na decomposição dos espectros NIR de óleo diesel e madeira de eucalipto.

Para cada modelo MLR de calibração multivariada, foram testadas as wavelets do tipo Daubechies de ordem 1 a 10 e as simlets 4 a 10, com 1 a 7 níveis de decomposição.

As regiões NIR estudadas, a descrição da quantidade de amostras empregadas em cada uma das regiões espectrais, os pré-processamentos efetuados e a construção dos modelos PLS descritos nos capítulos anteriores, serviram de base de estudo para os resultados obtidos neste ponto da tese.

Nesta etapa do trabalho foram empregados modelos MLR com filtragem Wavelet e os diversos modelos de calibração multivariada construídos até o momento foram comparados quanto a parsimônia e eficiência na estimativa dos diversos parâmetros de qualidade de óleo diesel e madeira de eucalipto investigados.

Os modelos de regressão MLR com filtragem wavelet (MLR-WT) foram aplicados aos espectros NIR derivativos para as regiões espectrais já consideradas no capítulo Materiais e Métodos. Os resultados das estimativas dos parâmetros de qualidade para as amostras da série de previsão, empregando filtragem Wavelet e regressão por MLR e aplicados a dados de óleo diesel encontram-se na tabela 12. Os resultados para madeira de eucalipto encontram-se no “Apêndice J” desta tese, bem como o detalhamento dos demais resultados para óleo diesel, tais como o tipo de Wavelet e o seu nível de decomposição que levaram à modelos MLR com menores valores de RMSEP.

Tabela 12. Valores de RMSEP e R (correlação) para a estimativa dos parâmetros de qualidade de óleo diesel, empregando filtragem Wavelet e modelos MLR, aplicados ao conjunto de amostras da série de previsão, com os espectros derivativos, nas regiões espectrais: 850 - 1.100 nm (R.E.1), 1.100 - 1.570 nm (R.E.2) e 1.570 – 2.500 nm (R.E.3).

Parâmetro de qualidade	R.E. 1		R.E. 2		R.E. 3		^b Repro.
	RMSEP	R	RMSEP	R	RMSEP	R	
Cetano	1,0	0,7091	0,9	0,8332	0,9	0,7605	2,0
Densidade (g.cm⁻³)	0,0009	0,9679	0,0005	0,9936	0,0005	0,9930	0,0012
Enxofre (% m/m)	0,02	0,8993	0,02	0,9558	0,02	0,9214	0,05
T-50 % (°C)	4,1	0,8706	2,8	0,9505	2,9	0,9456	9,7
T-85 % (°C)	7,4	0,6522	5,1	0,8914	4,9	0,8822	10,6

^{a)} Em vermelho encontram-se os menores RMSEP na comparação entre as regiões espectrais R.E.1, R.E.2 e R.E.3.

^{b)} Repro. = Valores de reprodutibilidade dos métodos de referência associado a análise de cada parâmetro de qualidade investigado.

Analisando-se a tabela 12, pode-se constatar conclusões semelhantes com relação ao uso de modelos PLS e modelos MLR com seleção de variáveis, ou seja, a região espectral 1.570 – 2.500 nm (R.E.3), por possuir picos de absorção dos hidrocarbonetos que são pouco sobrepostos, bem definidos e muito intensos, cuja razão sinal/ruído é alta pode proporcionar bons modelos de calibração multivariada.

Outra constatação deste estudo e que pode ser verificado no “APÊNDICE J” desta tese é que o tipo de wavelet e o número de níveis de decomposição são realmente importantes para a otimização dos modelos MLR com filtragem Wavelet na estimativa dos parâmetros de qualidade de óleo diesel e madeira de eucalipto.

Foi realizada uma comparação entre os modelos de regressão por PLS que empregam filtros polinomiais por Savitzky-Golay com os modelos que empregam MLR com filtragem Wavelet e os resultados encontram-se na tabela 13.

Tabela 13. Valores de RMSEP dos modelos PLS com filtro de Savitzky-Golay (PLS-SG) e dos modelos MLR com filtragem Wavelet (MLR-WT), para o conjunto de amostras da série de previsão, nas regiões espectrais: 850 - 1.100 nm (R.E.1), 1.100 - 1.570 nm (R.E.2) e 1.570 – 2.500 nm (R.E.3).

Parâmetro de qualidade	R.E.1		R.E.2		R.E.3	
	PLS-SG	MLR-WT	PLS-SG	MLR-WT	PLS-SG	MLR-WT
Cetano	^a 1,4 (4)	1,0	1,0 (12)	0,9	1,0 (7)	0,9
Densidade (g.cm⁻³)	^b 0,0018 (7)	0,0009	0,0009 (11)	0,0005	0,0010 (5)	0,0005
Enxofre (% m/m)	0,03 (6)	0,02	0,02 (5)	0,02	0,02 (5)	0,02
T-50 % (°C)	6,5 (4)	4,1	4,8 (7)	2,8	2,9 (10)	2,9
T-85 % (°C)	10,5 (4)	7,4	6,3 (10)	5,1	8,4 (8)	4,9

^{a)} Entre parênteses encontra-se o número de variáveis latentes (V.L) empregado nos modelos PLS.

^{b)} Em vermelho encontram-se os menores RMSEP na comparação PLS com MLR nas respectivas regiões espectrais.

Verifica-se na tabela 13 que para todos os parâmetros de qualidade de óleo diesel os modelos MLR-WT demonstraram ser mais eficientes que os modelos PLS-SG, evidenciando a importância do filtro Wavelet na redução do ruído instrumental dos espectros NIR.

Comparação semelhante entre os modelos MLR que empregam filtragem Wavelet (MLR-WT) com modelos MLR que usam filtro de Savitzky-Golay e seleção de variáveis espectrais pelo Algoritmo das Projeções Sucessivas (MLR-SG-APS), também foi realizada e os resultados encontram-se na tabela 14.

Tabela 14. Valores de RMSEP dos modelos MLR com filtro polinomial por Savitzky-Golay e seleção de variáveis por APS (MLR-SG-APS) e regressão MLR com filtragem e seleção de variáveis por transformada Wavelet (MLR-WT), aplicados ao conjunto de amostras da série de previsão, com os espectros derivativos nas regiões espectrais: 850 - 1.100 nm (R.E.1), 1.100 - 1.570 nm (R.E.2) e 1.570 – 2.500 nm (R.E.3).

Parâmetro de qualidade	R.E.1		R.E.2		R.E.3	
	MLR-SG-APS	MLR-WT	MLR-SG-APS	MLR-WT	MLR-SG-APS	MLR-WT
Cetano	^a 1,4 (28)	1,0	1,1 (22)	0,9	1,0 (14)	0,9
Densidade (g.cm⁻³)	^b 0,0023 (73)	0,0009	0,0010 (54)	0,0005	0,0009 (24)	0,0005
Enxofre (% m/m)	0,04 (30)	0,02	0,03 (11)	0,02	0,02 (9)	0,02
T-50 % (°C)	6,8 (38)	4,1	3,9 (21)	2,8	4,8 (72)	2,9
T-85 % (°C)	9,0 (8)	7,4	8,2 (75)	5,1	8,3 (34)	4,9

^{a)} Entre parênteses encontra-se o número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis por APS.

^{b)} Em vermelho encontram-se os menores RMSEP na comparação MLR-SG-APS com MLR-WT nas respectivas regiões espectrais.

Pela tabela 14, constata-se novamente a eficiência dos modelos MLR-WT, agora comparado com os modelos MLR-SG-APS. Vale notar neste ponto que o filtro Wavelet realmente foi eficaz na redução do ruído dos espectros, uma vez que aqui a comparação se faz entre duas técnicas de filtragem de sinais que empregam a mesma técnica de calibração multivariada.

Outra comparação foi realizada, agora entre os modelos MLR que usam filtragem Wavelet (MLR-WT) com modelos MLR que usam filtro de Savitzky-Golay, seleção de variáveis espectrais e amostras de calibração pelo Algoritmo das Projeções Sucessivas (MLR-APS-AM), e os resultados encontram-se na tabela 15.

Tabela 15. Comparação entre RMSEP obtidos na estimativa dos parâmetros de qualidade, para o conjunto de amostras de previsão, empregando os modelos MLR, filtro polinomial por Savitzky-Golay e seleção de variáveis e amostras de calibração por APS (MLR-APS-AM) e também com emprego de regressão MLR com filtragem e seleção de variáveis por transformada Wavelet (MLR-WT), nas regiões espectrais: 850 - 1.100 nm (C.O. 50), 1.100 - 1.570 nm (C.O. 10) e 1.570 – 2.500 nm (C.O. 1).

Parâmetro de qualidade	850 – 1.100 nm (R.E.1)		1.100 – 1.570 nm (R.E.2)		1.570 – 2.500 nm (R.E.3)	
	MLR-APS-AM	MLR-WT	MLR-APS-AM	MLR-WT	MLR-APS-AM	MLR-WT
Cetano	^a 1,6 (28,12)	1,0	1,3 (22,3)	0,9	1,0 (14)	0,9
Densidade (g.cm⁻³)	^b 0,0038 (73,55)	0,0009	0,0010 (54,42)	0,0005	0,0009 (24)	0,0005
Enxofre (% m/m)	0,03 (30,3)	0,02	0,02 (11,11)	0,02	0,02 (9)	0,02
T-50 % (°C)	7,3 (38,1)	4,1	4,8 (21,19)	2,8	4,8 (72)	2,9
T-85 % (°C)	9,8 (8,8)	7,4	10,7 (75,42)	5,1	8,3 (34)	4,9

^a) Entre parênteses encontra-se o número de comprimentos de onda e amostras de calibração, respectivamente empregadas nos modelos MLR.

^b) Em vermelho encontram-se os menores RMSEP na comparação MLR-APS-AM com MLR-WT, nas respectivas regiões espectrais.

Tomando-se como fator de comparação os RMSEP fornecidos na tabela 15, verifica-se que os modelos MLR-APS-AM demonstraram ser menos eficientes que os modelos MLR-WT, para todos os parâmetros de qualidade investigados.

4.2.5 – Filtragem Wavelet de espectros NIR médios a partir da estatística de varreduras individuais.

Nesta etapa do trabalho, um novo método de filtragem wavelet, proposto por kawakami et al.^[72] e que explora as estatísticas de scans individuais obtidas durante o processo de aquisição de espectros médios foi aplicado a dados NIR de óleo diesel. Os resultados aplicados a dados simulados não foram explorados nesta tese, mas encontram-se no artigo^[72] que trata deste assunto.

Foram registrados espectros NIR de uma única amostra de diesel, empregando-se um espectrofotômetro FT-NIR Bomem, com uma resolução de 2cm^{-1} na região espectral de 850 – 1.300 nm, à temperatura ambiente (24°C) e 48% de umidade relativa do ar. Os espectros resultantes possuem 2.560 comprimentos de onda (pontos). Inicialmente, foram adquiridos 50 scans individuais (um scan individual é descrito na figura. 35-a). Em seguida, um espectro médio (usado como referência para cálculo do RMSE) foi obtido para a mesma amostra usando 100 scans (figura 35-b).

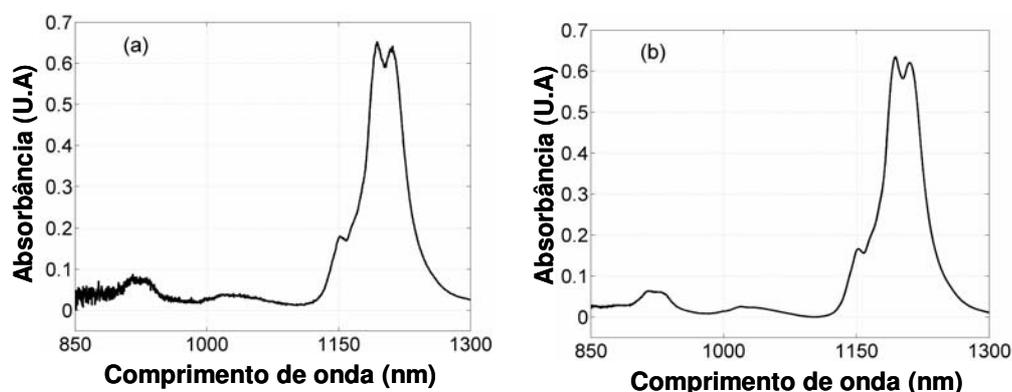


Figura 35. (a) Varredura Individual, (b) Espectro médio de 100 varreduras usado como referência para cálculo do RMSE, na região de 850 – 1.300 nm.

Todos os cálculos envolvidos neste estudo experimental foram executados em ambiente Matlab 6.5^[49] e seu respectivo pacote de Wavelets. O processo de filtragem dos espectros foi realizado utilizando-se inicialmente um banco de filtros da família Symlet 8 com 7 níveis de decomposição.

A figura 36 descreve o estudo comparativo entre os métodos de apodização: proposto, rígido (Hard) e suave^[71] (soft), aplicados a dados NIR de óleo diesel.

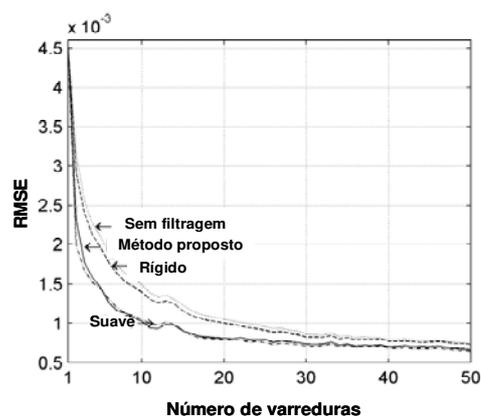


Figura 36. RMSE em relação ao espectro de referência (100 scans) e como uma função do número de varreduras na obtenção dos espectros médios. Comparação entre os métodos: Sem filtragem, Proposto, rígido e suave.

Conforme se verifica na figura 36, o método proposto forneceu valores de RMSE comparáveis aos melhores resultados obtidos pelo método suave (soft). Lembrando que o inconveniente dos métodos rígido (Hard) e suave (Soft) é a escolha do limiar de corte de frequências que o analista deve fornecer e que não é necessário no método proposto.

Uma das conseqüências práticas deste estudo também é mostrar, a partir de um teste-F utilizado, que não há diferença estatística significativa entre adquirir um número de 16 ou 32 scans, para o cálculo dos espectros NIR médios, reduzindo consideravelmente o tempo de aquisição de espectros médios, para atingir uma satisfatória razão sinal/ruído.

Alem disso, a fim de avaliar a sensibilidade do método de filtragem com relação ao banco de filtros wavelet, uma investigação também foi realizada^[72] para o caso de 16 scans e utilizando-se 21 wavelets das famílias Daubechies (db), Symlet (sym), e Coiflet (coif), como em estudos anteriores encontrados na literatura^[71,96] e uma tabela com os valores de RMSE em função dos tipos de wavelet e do método de apodização empregados apresenta-se na tabela 16.

Tabela 16. Valores de RMSE ($\times 1000$) para 16 scans e diferentes filtros wavelet.

Filtros Wavelet	Método de Filtragem		
	Corte Suave	Corte Rígido	Método Proposto
db2	0,90	1,11	0,89
db3	0,88	1,10	0,89
db4	0,86	1,10	0,88
db5	0,85	1,10	0,86
db6	0,86	1,10	0,86
db7	0,85	1,10	0,87
db8	0,85	1,12	0,86
db9	0,85	1,11	0,86
db10	0,86	1,11	0,85
sym4	0,86	1,11	0,87
sym5	0,87	1,11	0,86
sym6	0,86	1,11	0,88
sym7	0,86	1,11	0,89
sym8	0,85	1,11	0,87
sym9	0,85	1,11	0,82
sym10	0,84	1,10	0,83
coif1	0,89	1,10	0,87
coif2	0,86	1,11	0,88
coif3	0,85	1,10	0,87
coif4	0,84	1,10	0,86
coif5	0,84	1,10	0,82

Constata-se, nesta tabela que o método proposto foi sempre melhor que o método de corte rígido. Além disso, com exceção da família sym 7 e da sym 9, o método proposto apresentou valores de RMSE similares aos do método suave em todos os demais casos. Para a família sym 9 o método proposto é significativamente melhor que o método de corte suave, por outro lado a situação contrária é observada para a família sym 7.

Capítulo 5

Conclusões

“Só as vitórias disputadas com esforço
e perseverança são dignificantes.”

Alexandre Magno

5 - Conclusões

Os resultados obtidos com o emprego de modelos MLR com seleção de variáveis espectrais e amostras de calibração aplicados a dados de óleo diesel e madeira de eucalipto demonstraram ser mais simples que os modelos PLS, que empregam toda a região espectral e transformação em variáveis latentes e são mais complexos de interpretar.

Para os resultados de óleo diesel, embora não tenham sido observados modelos MLR com os melhores ajustes em apenas uma região espectral entre as investigadas neste trabalho (850 – 1.100 , 1.100 – 1.570, 1.570 – 2.500) e que justificasse o uso de apenas uma cela de caminho óptico, pode-se eleger a região de 1.100 – 1.570 nm, cuja cela de fluxo empregada foi de 10 mm como uma opção mais prática de registro de espectros de futuras amostras e que empregue equipamento NIR comercial, pois não há necessidade de utilização de fibras ópticas como na cela de 1 mm, bem como a razão sinal/ ruído dos espectros é maior que na região que emprega a cela de 50 mm, sendo que esta necessita de um grande volume de óleo diesel no caminho óptico.

Porém, se a intenção da utilização de técnicas de seleção de variáveis é a construção de fotômetros a base de LED e dedicados a estimativa dos parâmetros de qualidade de óleo diesel investigados neste trabalho, sugere-se que também sejam investigados LEDs que emitam tanto na região do terceiro sobreton (750 – 1.100 nm) que são de custo reduzido, quanto na região das bandas de primeiro sobreton e de combinação (1.570 – 2.500 nm), cuja altura do pico de emissão é alta.

Os resultados obtidos com modelos MLR que empregaram técnicas de seleção de amostras de calibração além de seleção de variáveis espectrais para os dados de madeira de eucalipto e óleo diesel demonstraram a possibilidade de procedimentos de recalibrações e transferência de calibração que utilizam um número reduzido de amostras, sem a perda significativa da capacidade preditiva dos modelos MLR utilizados na estimativa dos diversos parâmetros de qualidade.

Os resultados de seleção de árvores empregando-se o KS demonstraram promissora aplicação para a derrubada das árvores mais representativas a compor os

modelos de calibração para a estimativa dos diversos parâmetros de qualidade das madeiras de eucalipto que são determinadas por empresas de papel e celulose do Brasil.

No estudo relacionado ao uso de modelos MLR-WT, foi demonstrado que a transformada Wavelet é uma técnica eficiente para a redução de ruído instrumental de espectros NIR aplicada a calibração multivariada por MLR. Além disso, o estudo da otimização das famílias wavelet e do nível de decomposição se tornou importante para a eficiência do tratamento dos espectros NIR, resultando em modelos MLR-WT com capacidade preditiva mais eficiente que os modelos PLS.

Para aplicações de que trata o estudo de filtragem Wavelet de espectros NIR médios a partir da estatística de varreduras individuais, o método proposto nesta tese deve ser particularmente útil para dados de ressonância magnética nuclear e também para espectroscopia na região do infravermelho distante, onde um número considerável de varreduras pode ser requerido para se conseguir uma adequada relação sinal/ruído.

Outras demonstrações de eficiência da transformada Wavelet em diversas aplicações^[54-58,62,67,69] tornam esta técnica de processamento de sinais uma ferramenta atrativa para a quimiometria.

Capítulo 6

Referências Bibliográficas

“O bom senso é o que há de mais bem distribuído no mundo, pois cada um pensa estar bem provido dele”

René Descartes

6 - Referências Bibliográficas

1. Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi P. J., Smeyers-Verbeke, J., **Handbook of Chemometrics and Qualimetrics: Part A**, Elsevier, Amsterdam, 1997.
2. Massart, D. L., Vandeginste, B. G. M., Buydens, L. M. C., de Jong, S., Lewi P. J., Smeyers-Verbeke, J., **Handbook of Chemometrics and Qualimetrics: Part B**, Elsevier, Amsterdam, 1997.
3. Ferreira, M. M. C. et al., *Quimiometria I: calibração multivariada, um tutorial*, **Química Nova**, 22 (1999), 724 - 731.
4. Costa, P. A.; Poppi, R. J., *Algoritmo Genético em Química*, **Química Nova**, 22 (1999), 405 - 411.
5. Galvão, R. K. H., Araújo, M. C. U., Saldanha, T. C. B., Visani, V., Pimentel, M. F.; *Estudo comparativo sobre filtragem de sinais instrumentais usando transformadas de Fourier e Wavelet*, **Química Nova**, 24 (2001), 874 - 884.
6. Martens, H., and Naes, T.; **Multivariate Calibration**, Wiley, New York, 1989.
7. Beebe, K. R., Pell, R.J. and Seasholtz, M.B.; **Chemometrics - A Practical Guide**, Wiley-Interscience, New York, 1998.
8. Geladi, P., Kowalski, B. R., Partial Least Squares: A tutorial, **Analytica Chimica Acta**, 185 (1986), 1-17.
9. Jouan-Rimbaud, D., Walczak, B., Poppi, R. J., de Noord, O. E., Massart, D. L., *Application of wavelet transform to extract the relevant component from spectral data for multivariate calibration* **Analytical Chemistry**, 69 (1997), 4317 – 4323.
10. Xu, L., Jiang, J. H., Wu, H. L., Shen, G. L., Yu, R. Q., *Variable-weighted PLS*, **Chemometrics and Intelligent Laboratory Systems**, 85 (2007), 140 – 143.
11. Bogomolov, A., Hachey, M., *Application of SIMPLISMA purity function for variable selection in multivariate regression analysis: A case study of protein secondary structure determination from infrared spectra*, **Chemometrics and Intelligent Laboratory Systems**, 87 (2007), 57-67.
12. Centner, V., Massart, D. L., de Noord, O. E., de Jong, S., Vandeginste, B. M., Sterna, C., *Elimination of uninformative variables for multivariate calibration*, **Analytical Chemistry**, 68 (1996), 3851 – 3858.
13. Gusnanto, A., Pawitan, Y., Huang, J., Lane, B., *Variable selection in random calibration of near-infrared instruments: ridge regression and partial least squares regression settings*, **Journal of Chemometrics**, 17 (2003) 174 – 185.
14. Sutter, J. M., Kalivas, J. H., *Comparison of Forward Selection, Backward Elimination, and Generalized Simulated Annealing for Variable Selection*, **Microchemical Journal**, 47 (1993) 60 -66.

15. Kabe, D. G., *Stepwise Multivariate Linear Regression*, **Journal of the American Statistical Association**, 58 (1963), 770-773.
16. Chong, I. G., Jun, C. H., *Performance of some variable selection methods when multicollinearity is present*, **Chemometrics and Intelligent Laboratory Systems**, 78 (2005) 103 – 112.
17. Todeschini, R., Galvagni, D., Vilchez, J. L., del Olmo, M., Navas, N., *Kohonen artificial neural networks as a tool for wavelength selection in multicomponent spectrofluorimetric PLS modelling: application to phenol, o-cresol, m-cresol and p-cresol mixtures*, **Trends in Analytical Chemistry**, 18 (1999) 93 – 98.
18. Kalivas, J. H., Roberts, N. Sutter, J. M., *Elimination of uninformative variables for multivariate calibration*, **Analytical Chemistry**, 61 (1989), 2024 – 2030.
19. Lima, S. L. T., Mello, C., Poppi, R. J., *PLS pruning: a new approach to variable selection for multivariate calibration based on Hessian matrix of errors*, **Chemometrics and Intelligent Laboratory Systems**, 76 (2005), 2024 – 2030.
20. Hemmateenejad, B., Ghavami, R., Miri, R., Shamsipur, M., *Net analyte signal-based simultaneous determination of antazoline and naphazoline using wavelength region selection by experimental design-neural networks*, **Talanta**, 68 (2006) 1222 – 1229.
21. Rao, K. R., Lakshminarayanan, S., *Partial correlation based variable selection approach for multivariate data classification methods*, **Chemometrics and Intelligent Laboratory Systems**, 86 (2007) 68 – 81.
22. Ferreira, M. M. C., Montanari, C. A., Gaudio, A. C., *Seleção de variáveis em QSAR*, **Química Nova**, 25 (2002) 439 – 448.
23. Lucasius, C. B.; Kateman, G.; *Genetic algorithms for large-scale optimization in chemometrics: An application*, **Trends in Analytical Chemistry**, 10 (1991), 254 - 261.
24. Díez, I. E., Sáiz, J. M. G., Pizarro, C., Forina, M., *Alternating conditional expectations regression with selection of significant predictors by genetic algorithms*, **Analytica Chimica Acta**, 555 (2006) 96 – 106.
25. Cho, S. J., Hermsmeier, M. A., *Genetic Algorithm Guided Selection: Variable Selection and Subset Selection*, **Journal of Chemical Information and Computer Sciences**, 42 (2002) 927-936.
26. Goicoechea, H. C., Olivieri, A., *A new family of genetic algorithms for wavelength interval selection in multivariate analytical spectroscopy*, **Journal of Chemometrics, Journal of Chemometrics**, 17 (2003) 338-345.
27. Franco, V. G., Perin, J. C., Mantovani, V. E., Goicoechea, H. C., *Monitoring substrate and products in a bioprocess with FTIR spectroscopy coupled to artificial neural networks enhanced with a genetic-algorithm-based method for wavelength selection*, **Talanta**, 68 (2006) 1005 – 1012.

28. Leardi, R., Seasholtz, M. B., Pell, R. J., *Variable selection for multivariate calibration using a genetic algorithm: prediction of additive concentrations in polymer films from Fourier transform-infrared spectral data*, ***Analytica Chimica Acta***, 461 (2002) 189 – 200.
29. Leardi, R., *Genetic algorithms in chemometrics and chemistry: a review*, ***Journal of Chemometrics***, 15 (2001) 559 – 569.
30. Gourvéneq, S., Capron, X., Massart, D. L., *Genetic algorithms (GA) applied to the orthogonal projection approach (OPA) for variable selection*, ***Analytica Chimica Acta***, 519 (2004) 11 – 21.
31. Leardi, R., Boggia, R., Terrile, M.; *Genetic algorithms as a strategy for feature selection*, ***Journal of Chemometrics***, 6 (1992), 267 - 281.
32. Jouan-Rimbaud, D., Massart, D. L.; Leardi, R.; Noord, O. E.; *Genetic algorithms as a tool for wavelength selection in multivariate calibration*, ***Analytical Chemistry***, 67 (1995), 4295 - 4301.
33. Guchardi, R., Costa Filho, P. A.; Poppi, R. J.; Pasquini, C.; *Determination of ethanol and methyl tert-butyl ether (MTBE) in gasoline by NIR–AOTF-based spectroscopy and multiple linear regression with variables selected by genetic algorithm*, ***Journal of Near Infrared Spectroscopy***, 6 (1998), 333 - 339.
34. Leardi, R.; *Application of genetic algorithm-PLS for feature selection in spectral data sets*, ***Journal of Chemometrics***, 14 (2000), 643 - 655.
35. Araújo, M. C. U., Saldanha, T. C. B., Galvão, R. K. H., Yoneyama, T., Chame, H. C., Visani, V.; *The successive projections algorithm for variable selection in spectroscopic multicomponent analysis*, ***Chemometrics and Intelligent Laboratory Systems***, 57 (2001), 65-73.
36. Galvão, R. K. H., Pimentel, M. F., Araújo, M. C. U., Yoneyama, T., Visani, V.; *Aspects of the successive projections algorithm for variable selection in multivariate calibration applied to plasma emission spectrometry*, ***Analytica Chimica Acta***, 443 (2001), 107 - 115.
37. Breitzkreitz, M. C., Raimundo, I. M. Jr., Rohwedder, J. J. R., Pasquini, C., Dantas Filho, H. A., José, G. E., Araújo, M. C. U., *Determination of total sulfur in diesel fuel employing NIR spectroscopy and multivariate calibration*, ***Analyst***, 128 (2003), 1204 - 1207.
38. Chame, H. C., “O Algoritmo de Projeções Sucessivas para Seleção de Variáveis em Análise Espectrofotométrica Multicomponente Simultânea de Metais” - Dissertação de Mestrado - ***Universidade Federal da Paraíba*** (2000).
39. Honigs, D. E., Hieftje, G. M., Mark, H. L., Hirschfeld, T. B.; *Unique-sample selection via near-infrared spectral subtraction*, ***Analytical Chemistry***, 57 (1985), 2299 - 2303.

40. Puchwein, G.; *Selection of calibration samples for near-infrared spectrometry by factor analysis of spectra*, **Analytical Chemistry**, 60 (1988), 569 - 573.
41. Navarro-Villoslada, F., Pérez-Arribas, L. V., Leon-González, M. E., Pólo-Diez.; *Selection of calibration mixtures and wavelengths for different multivariate calibration methods*, **Analytica Chimica Acta**, 313 (1995), 93 - 101.
42. Ferré, J., Xavier-Rius, F., *Constructing D-optimal designs from a list of candidate samples*, **Trends in Analytical Chemistry**, 16 (1997), 70 - 73.
43. Wang, Y. D., Veltkamp, D. J., Kowalski, B. R.; *Multivariate instrument standardization*, **Analytical Chemistry**, 63 (1991), 2750 - 2756.
44. Bouveresse, E., Massart, D. L.; *Improvement of the piecewise direct standardisation procedure for the transfer of NIR spectra for multivariate calibration*, **Chemometrics and Intelligent Laboratory Systems**, 32 (1996), 201 - 203.
45. Kennard, R. W., Stone, L. A., *Computer Aided Design of Experiments*, **Technometrics**, 11 (1969), 137 - 148.
46. Feudale, R. N., Woody, N. A., Tan, H. Myles, A. J., Brown, S. D., Ferré, J.; *Transfer of multivariate calibration models: a review*, **Chemometrics and Intelligent Laboratory Systems**, 64 (2002), 181-192.
47. Dantas Filho, H. A., Galvão, R. K. H., Araújo, M. C. U., Silva, E. C., Saldanha, T. C. B., José, G. E., Pasquini, C., Raimundo, I. M. Jr., Rohwedder, J. J. R., *A strategy for selecting calibration samples for multivariate modeling*, **Chemometrics and Intelligent Laboratory Systems**, 72 (2004), 83-91.
48. Skoog, D. A., Holler, F. J., Nieman, T. A.; **Principles of Instrumental Analysis (tradução)**, 5ª ed., Editora Bookman, Porto Alegre-RS, 2002.
49. Matlab, versão 6.5, The Math Works inc, 2003, Natick, Massachusetts, USA, 2003.
50. Vetterli, M., Kovacevic, J.; **Wavelets and Subband Coding**, Prentice-Hall, New Jersey, 1995.
51. Daubechies, I.; **Ten Lectures on Wavelets**, SIAM, Philadelphia, 1992.
52. Strang, G., Nguyen, T., **Wavelets and Filter Banks**. Wellesley-Cambridge Press; Wellesley, 1996.
53. Misiti, M., Misiti, Y., Oppenheim, G., Poggi, J. M.; **Wavelet Toolbox User's Guide**, The Mathworks, Natick, 1996.
54. Leung, A. K.; Chau, F.; Gao, J. C.; *A review on applications of wavelet transform techniques in chemical analysis: 1989 –1997*, **Chemometrics and Intelligent Laboratory Systems**, 43 (2004), 165 - 184.

-
55. Brown, P. J.; Fearn, T. e Vanucci, M.; *Bayesian Wavelet Regression on Curves With Application to a Spectroscopic Calibration Problem*, **Journal of American Statistic Associated**, (2001), 398 – 408.
 56. Jetter, K.; Depezynski, U.; Molt, K. e Niemoller, A.; *Principles and applications of wavelet transformation to chemometrics*, **Analytica Chimica Acta**, 368 (2000), 169 – 180.
 57. Eriksson, L.; Trygg, J.; Johansson, E.; Bro, R. e Wold, S.; *Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated process fluorescence data*, **Analytica Chimica Acta**, 420 (2000), 181 - 195.
 58. Coelho, C. J.; Galvão, R. K. H.; Araújo, M. C. U.; Saldanha, T. C. B.; Visani, V. e Pimentel, M. F. P.; *A solution to the wavelet transform optimization problem in multicomponent analysis*, **Chemometrics and Intelligent Laboratory Systems**, 66 (2003), 205 - 217.
 59. Rioul, O., Vetterli, M.; *Wavelets and signal processing*, **IEEE Signal Processing Magazine**, 8 (1991), 14-38.
 60. Meyer, Y.; **Wavelets-Algorithms and Applications**, SIAM, Philadelphia, 1993.
 61. Alsberg, B. K.; Woodward, A. M., Kell, D. B.; *An introduction to wavelet transforms for chemometricians: A time-frequency approach*, **Chemometrics and Intelligent Laboratory Systems**, 37 (1997), 215-239.
 62. Daubechies, I.; *Where do wavelets come from? A personal point of view*, **Proceedings of the IEEE**, 84 (1996), 510-513.
 63. Coelho, C. J., Galvão, R. K. H., Araújo, M. C. U., Pimentel, M. F., Silva, E. C.; *A Linear Semi-infinite Programming Strategy for Constructing Optimal Wavelet Transforms in Multivariate Calibration Problems*, **Journal of Chemical Information and Computer Sciences**, 43 (2003), 928-933.
 64. Galvão, R. K. H., José, G. E., Dantas Filho, H. A., Araújo, M. C. U., Silva, E. C., Paiva, H. M., Saldanha, T. C. B., Souza, E. S. O. N., *Optimal wavelet filter construction using X and Y data*, **Chemometrics and Intelligent Laboratory Systems**, 70 (2004), 1-10.
 65. Smrcok, L., Durik, M., Jorik, V., *Wavelet denoising of powder diffraction patterns*, **Powder Diffraction**, 14 (1999), 300-304.
 66. Medeiros, V. M.; Araújo, M. C. U.; Galvão, R. K. H.; da Silva, E. C.; Saldanha, T. C. B.; Toscano, I. A. S.; de Oliveira, M. S. R.; Freitas, S. K. B. e Mariano Neto, M.; *Screening analysis of river seston downstream of an effluent discharge point using near-infrared reflectance spectrometry and wavelet-based spectral region selection*, **Water Research**, 39 (2005), 3098 – 3097.
 67. Fearn, T. e Davies, M. C.; *A comparison of Fourier and wavelet transforms in the processing of near infrared spectroscopic data: Part 1. Data compression*, **Journal of Near Infrared Spectroscopy**, 11 (2003), 3–15.

-
68. Galvão, R. K. H., Hadjiloucas, S., Bowen, J. W.; *Use of the statistical properties of the wavelet-transform coefficients for optimization of integration time in Fourier transform spectrometry*, **Optics Letters**, 27 (2002) 643-645.
69. Walczak, B., **Wavelets in Chemistry**, Elsevier Science, New York, 2000.
70. Cai, C.S., Harrington, P.D., *Different Discrete Wavelet Transforms Applied to Denoising Analytical Data*, **Journal of Chemical Information and Computer Sciences**, 38 (1998) 1161-1170.
71. Donoho, D.L., *De-noising by soft-thresholding*, **IEEE Transactions Information Theory**, 41 (1995) 613-627.
72. Galvão, R. K. H., Dantas Filho, H. A., Martins, M. N., Araújo, M. C. U., Pasquini, C., *Sub-optimal wavelet denoising of coaveraged spectra employing statistics from individual scans*, **Analytica Chimica Acta**, 581 (2007) 159 - 167.
73. BURNS, D. R., CIURCZAK, E. W., **Handbook of Near-Infrared Analysis**, Marcel Dekker editor, New York, 1992.
74. **Revista Ciência Hoje**, Combustíveis de Qualidade - uma busca constante, vol 28, nº 165, p. 2-7, outubro de 2000.
75. <http://pt.wikipedia.org/wiki/Diesel>, acessado em 04 de maio de 2007.
76. Agência Nacional do Petróleo - ANP, Portaria Nº 310, DE 27 DE DEZEMBRO DE 2001.
77. <http://www.celuloseonline.com.br/>, acessado em 04 de maio de 2007.
78. Annual Book of ASTM Standards, Standard Test Method for Calculated Cetane Index by Four Variable Equation, **ASTM D-4737**, ASTM International, Philadelphia, USA, 1990.
79. Annual Book of ASTM Standards, Density and Relative Density of Liquids by Digital Density Meter, **ASTM D-4052**, ASTM International, Philadelphia, USA, 1995.
80. Annual Book of ASTM Standards, Sulfur in Petroleum Products by Energy Dispersive X-Ray Fluorescence Spectroscopy, **ASTM D-4294**, ASTM International, Philadelphia, USA, 1990.
81. Annual Book of ASTM Standards, Distillation of Petroleum Products at Atmospheric Pressure, **ASTM D 86**, ASTM International, Philadelphia, USA, 1995.
82. <http://www.hellma.com.br/cubetas.html>, acessado em 04 de maio de 2007.
83. Workman, J. J. JR.; *Interpretative Spectroscopy for Near Infrared*, **Applied Spectroscopy Reviews**, 31 (1996), 251-320.

-
84. Technical Association of Pulp and Paper Industry - **TAPPI T258 cm-94**, Basic Density and Moisture Content of Pulpwood, Atlanta, 1994.
 85. Technical Association of Pulp and Paper Industry - **TAPPI T204 cm-97**, Solvent Extractives of Wood and Pulp, Atlanta, 1997.
 86. Lanna, A. E. et al. Maximizing Pulp Yield Potential of Cenibra's Wood Supply. In: Brazilian Symposium on the Chemistry of Lignins and other Wood Components. 7., 2001. Belo Horizonte. Proceedings... Belo Horizonte: 2001. p 159-167.
 87. Browning, B. L.; The Chemistry of Wood. New York, Interscience, 1963. 689 p.
 88. Technical Association of Pulp and Paper Industry - **TAPPI T222 om-98**, Acid-Insoluble Lignin in Wood and Pulp, Atlanta, 1998/1999.
 89. Unscrambler-Ex User's Guide, versão 9.2, **CAMO A/S**, Oslo, Norway, 2004.
 90. Savitzky, A. and Golay, M. J. E.; *Smoothing and differentiation of data by simplified least squares procedures*, **Analytical Chemistry**, 36 (1964), 1627 – 1639.
 91. Dardenne, P., Sinnaeve, G., and Baeten, V.; *Multivariate calibration and chemometrics for near infrared spectroscopy: which method?* **Journal of Near Infrared Spectroscopy**, 8 (2000), 229-237.
 92. Bohács, G., Ovadi, Z., Salgó, A., *Prediction of gasoline properties with near infrared spectroscopy*, **Journal of Near Infrared Spectroscopy**, 6 (1998), 341-348.
 93. Saldanha, T. C. B, de Araujo, M, C, U, Neto B. B., Chame, H. C.; *Simultaneous Analysis of Co(II), Cu(II), Mn(II), Ni(II) and Zn (II) in the Ultraviolet Region Using PAR and Multivariate Calibration*, **Analytical Letters**, 33 (2000), 1187-1202.
 94. Lim, S. Y.; Dence, C. W.; *Methods in lignin Chemistry*. Berlim: Springer-Verlag, 1992. 578p.
 95. Associação Brasileira Técnica de Celulose e Papel – ABTCP. Norma 5/68: Solubilidade em NaOH a 1%, São Paulo, 1974. Não paginado.
 96. Pasti, L., Walczak, B., Massart, D.L., Reschiglian, P., *Optimization of signal denoising in discrete wavelet transform*, **Chemometrics and Intelligent Laboratory Systems**, 48 (1999) 21-34.
 97. Meyer, Y.; **Wavelets-Algoritms and Applications**, SIAM, Philadelphia, 1993.

Capítulo 7

Apêndices

“Enquanto as leis forem necessárias, os homens
não estarão capacitados para a liberdade.”

Pitágoras

Apêndice A- Tabela de Absorções NIR

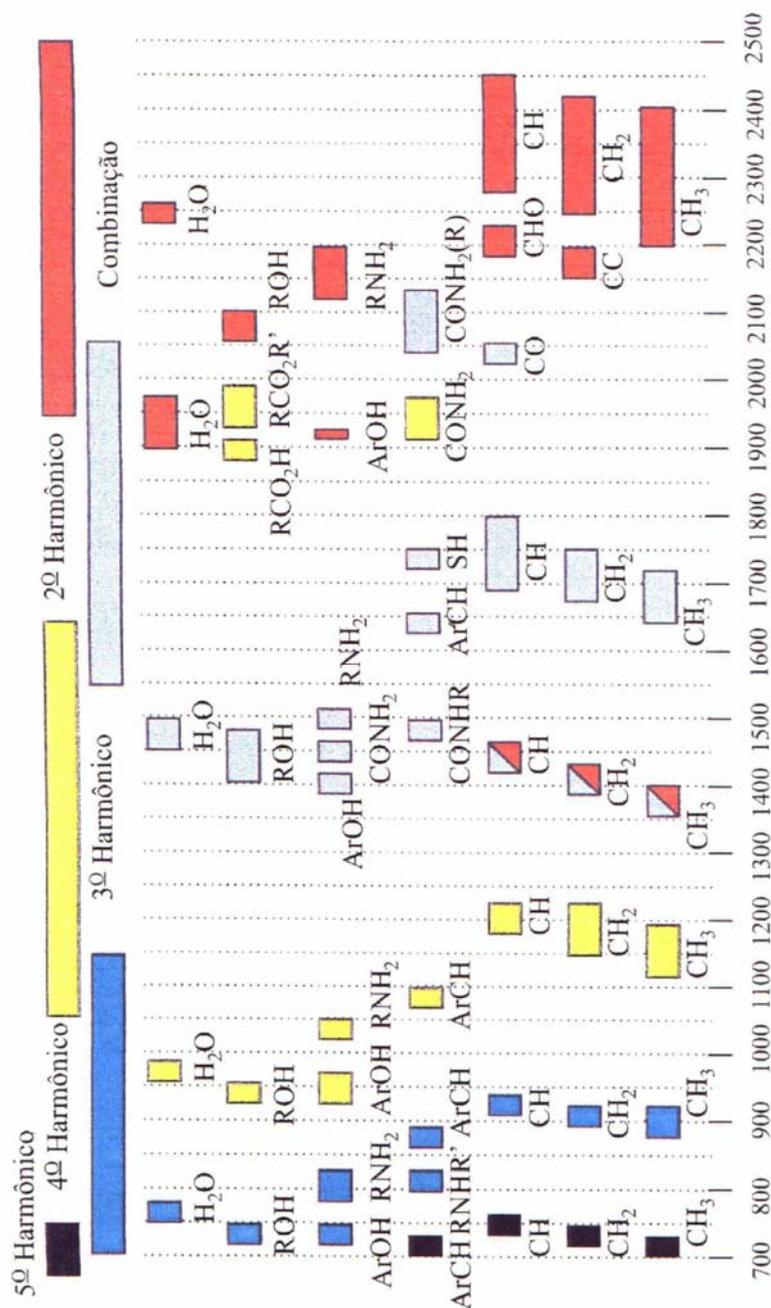


Figura A1. Principais compostos absorventes na região NIR

Apêndice B- Modelos PLS para madeira de eucalipto

Tabela B1. Número de amostras de madeira de eucalipto em cada um dos conjuntos de calibração, validação e previsão, em função do parâmetro de qualidade empregado nos modelos PLS e MLR na região espectral de 1.100 – 2.500 nm.

Parâmetro de Qualidade	Calibração (cal.)	Validação (val.)	Previsão (prev.)
Densidade	57	28	29
Extrativos	68	33	34
Rendimento	70	34	35
Holocelulose	31	15	15
Lignina	70	34	35

Tabela B2. Valores de RMSEP e coeficiente de correlação (R) na estimativa dos parâmetros de qualidade de madeira de eucalipto para o sub-conjunto de amostras de previsão dos modelos PLS na região espectral de 1.100 – 2.500 nm.

Parâmetro	Unidade	Faixa do parâmetro	RMSEP	R	Repro.
Densidade	g.cm ⁻³	0,4170 – 0,5750	0,016 (13)	0,8561	0,025
Extrativos	%	1,05 – 4,77	0,54 (6)	0,6381	0,5
Rendimento	%	44,9 – 53,3	0,60 (7)	0,9208	1,0
Holocelulose	%	64,16 – 72,89	1,19 (5)	0,8337	1,0
Lignina	%	25,44 – 38,00	1,39 (13)	0,7112	1,0

Apêndice C1- Seleção de Variáveis e Amostras por APS

- Algoritmo selecao

Colaborador: prof. Dr. Roberto Kawakami Harrop Galvão (Div. Eng. Eletrônica – ITA)

```
function [I,R,Lopt] = selecao(Xcal,Ycal,Xteste,Yteste,N1,N2,analitos)

% SELECAO - Selecao de Variaveis baseada no Algoritmo das Projecoes Sucessivas (APS)
% Passos do programa:
% 1) Centralizacao na media e auto-escalamento das colunas de Xcal
%   (somente para fins de geracao das cadeias de variaveis)
% 2) Uso do APS para gerar cadeias contendo no minimo 'N1'
%   e no maximo 'Nmist_cal' variaveis a partir de cada coluna de Xcal.
%   Total de cadeias: (Nmist_cal - N1 + 1) x (#variaveis em Xcal)
% 3) Calibracao de um modelo via Regressao Linear Multipla
%   com Decomposicao QR para cada cadeia de variaveis
% 4) Escolha da melhor cadeia de variaveis com base no RMSEP (série de teste)
%
% [I,R,Lopt] = selecao(Xcal,Ycal,Xteste,Yteste,N1,N2,analitos)
%
% Xcal --> (Calibracao) Respostas instrumentais (#objetos x #variaveis)
% Ycal --> (Calibracao) Concentracoes (#objetos x #componentes)
% Xteste --> (Teste) Respostas instrumentais (#objetos x #variaveis)
% Yteste --> (Teste) Concentracoes (#objetos x #componentes)
% N1 --> Limite inferior para o no. de variaveis a selecionar
%       Recomenda-se fazer N1 = Numero total de analitos
% N2 --> Limite superior para o no. de variaveis a selecionar
%       N2 deve ser menor ou igual ao numero de misturas de calibracao
% analitos --> Indices dos analitos a serem considerados nas matrizes Ycal e Yteste.
%       Exemplos: [1,3,5] --> Analitos 1,3,5
%                 2:5 --> Analitos 2,3,4,5
%                 4 --> Analito 4
%       Default: Todos os analitos.
%
% I --> Resultado da seleção (melhor cadeia de variaveis)
% R --> menor RMSEP obtido em funcao do no. de variaveis usadas
% (obs: R(N) vale zero para N<N1)
```

```
% Lopt --> Matriz contendo as cadeias de variaveis associadas a R
%
% Programas Auxiliares: APS.M, SUMSQR.M
%
% Se os analitos foram especificados, as matrizes Y sao
% reduzidas as colunas indicadas.
if nargin == 7
    disp('Considerando apenas os analitos indicados!')
    Ycal = Ycal(:,analitos);
    Yteste = Yteste(:,analitos);
end

Nmist_cal = size(Xcal,1);
Nlambdas = size(Xcal,2);
Nmist_teste = size(Yteste,1);
Ncomp = size(Yteste,2);

clc
disp('- Selecao de variaveis via APS -')
disp(' ')
disp('Escolha uma opcao e tecele <ENTER>')
disp('1 - Gerar cadeias de variaveis via APS')
disp('2 - Carregar cadeias de variaveis previamente geradas')
opcao = input('Opcao: ');

if (opcao == 1) % Gerar

    disp(' ')
    disp('O APS sera utilizado para gerar cadeias de variaveis.')
    disp(' ')
    disp('Entre com o nome de um arquivo para salvar as cadeias')
    filename = input('(entre apostrofos e sem extensao): ');

% No MATLAB, a inicializacao previa de vetores torna a execucao mais rapida
L = zeros(N2,Nlambdas);

% Centralização na Média e Auto-Escalonamento das colunas de Xcal
```

```
opcao = input('Deseja fazer centralizacao e auto-escalonamento para fins de selecao ? (Sim: 1 - Nao:
0)');
if (opcao == 1)
    disp('Sera feita centralizacao e auto-escalonamento!')
    for i=1:Nlambdas
        x = Xcal(:,i);
        Xcaln(:,i)=(x-mean(x))/std(x);
    end
else
    disp('Nao sera feita centralizacao e auto-escalonamento!')
    Xcaln = Xcal;
end
espera = input('Tecla algo para continuar...');

% Aplicacao do APS
clc
t0 = clock; % Comando usado para estimar quanto tempo levará a otimização
for i=1:Nlambdas

    L(:,i) = (aps(Xcaln,i,N2)); % Seleciona N2 comprimentos de onda

    % Apresenta na Tela uma estimativa do tempo restante para o fim da otimização
    if (rem(i,10) == 0)
        tempo = etime(clock,t0);
        home
        disp(['Aplicando APS. Tempo restante = ',num2str(round((Nlambdas-i)*tempo/10)), 's',blanks(10)]);
        t0 = clock;
    end

end

save(filename,'L')

else % Carregar

    filename = input('Entre com o nome do arquivo (entre apostrofos e sem extensao): ');
    load(filename)
end
```

```

% Determinação do RMSEP p/ todas as cadeias de variáveis
disp('Efetuando regressões para determinar a melhor cadeia de variáveis ...')
disp(' ')

R = zeros(1,N2);
Lopt = zeros(N2,N2);
rmsep = zeros(N2,Nlambdas);

for N = N1:N2 % Para a cadeia de comprimento N
  for i = 1:Nlambdas % partindo da variável i
    lambdas = L(1:N,i); % Variáveis da cadeia
    % Respostas instrumentais nas variáveis da cadeia
    Xcal2 = Xcal(:,lambdas);
    Xteste2 = Xteste(:,lambdas);
    % Determina o modelo via RLM + QR nas variáveis da cadeia
    W = Xcal2\Ycal;
    % Previsão das concentrações da série de teste
    Ypred = Xteste2*W;
    % Cálculo do RMSEP associado a cadeia
    rmsep(N,i)=sqrt(sumsqr(Ypred-Yteste)/(Nmist_teste*Ncomp));
  end
  [R(N) imin] = min(rmsep(N,:));
  Lopt(1:N,N)=L(1:N,imin);
end

%rmsep
[Rbest,Nbest] = min(R(N1:N2));
Nbest = Nbest+N1-1;
rmsepopt = rmsep(Nbest,:);
l = (Lopt(1:Nbest,Nbest))';
disp(['Menor RMSEP obtido: ' num2str(Rbest)])
disp(['Número ideal de variáveis: ' num2str(Nbest)])
figure,plot([N1:N2],R(N1:N2)),grid
axis([N1 N2 -inf inf])
xlabel('Número de variáveis')
ylabel('RMSEP')

```

Apêndice C2- Seleção de Variáveis e Amostra por APS - Algoritmo aps

```

function I = aps(Xcal,I1,N)

% APS - Algoritmo das Projeco es Sucessivas
%% I = aps(Xcal,I1,N)
%
% Xcal --> Matriz de respostas instrumentais (#objetos x #variaveis)
% I1 --> Variavel usada como ponto de partida
% N --> No. de variaveis a selecionar (incluindo I1)
%% I --> Variaveis selecionadas (incluindo I1)
%
% Recomenda-se que as colunas de Xcal sejam centradas na media.
% O programa SELECAO.M faz isso automaticamente.
%
% Inicialização do Algoritmo em I1
I = zeros(1,N);
I(1) = I1;

% Algoritmo das Projeções Sucessivas

for n = 2:N

% Em cada iteracao, todas as colunas de Xcal sao projetadas no subespaco
% ortogonal a coluna selecionada na iteracao anterior (vetor de referencia)

x = Xcal(:,I(n-1)); % Vetor de referencia para a iteracao atual
xn = x/(x'*x);
Xcal = Xcal - xn*x'*Xcal;
norms = sum(Xcal.^2); % Quadrado da norma das projeco es (soma dos quadrados de cada
elemento)
[dummy,I(n)] = max(norms); % I(n) : Indice da projecao de maior norma
% O vetor escolhido eh o "mais ortogonal" em relacao ao vetor de referencia.
end

```

Apêndice C3- Seleção de Variáveis e Amostra por APS - Algoritmo sumsqr

```
function s = sumsqr(x)
```

```
s = sum(sum(x.*x));
```

Apêndice C4- Seleção de Variáveis e Amostra por APS - Algoritmo selvaramost

```
function [l,m,rmsev,rmsep,rv,rp] = selvaramost(Xcal,Ycal,Xval,Yval,Xpred,Ypred,N2,analito)
```

```
% SELVARAMOST - Selecao de Variaveis e Amostras baseada no Algoritmo das Projecoes Sucessivas  
(APS)
```

```
%
```

```
% [l,m,rmsev,rmsep,rv,rp] = selvaramost(Xcal,Ycal,Xval,Yval,Xpred,Ypred,N2,analito)
```

```
%
```

```
% Xcal --> (Calibracao) Respostas instrumentais (#objetos x #variaveis)
```

```
% Ycal --> (Calibracao) Concentracoes (#objetos x #componentes)
```

```
% Xval --> (Teste) Respostas instrumentais (#objetos x #variaveis)
```

```
% Yval --> (Teste) Concentracoes (#objetos x #componentes)
```

```
% Xpred --> (Predicao) Respostas instrumentais (#objetos x #variaveis)
```

```
% Ypred --> (Predicao) Concentracoes (#objetos x #componentes)
```

```
% N2 --> Limite superior para o no. de variaveis a selecionar
```

```
% N2 deve ser menor ou igual ao numero de misturas de calibracao
```

```
% analito --> Indice do analito a ser considerado nas matrizes Y
```

```
%
```

```
% l --> Resultado da seleção de variaveis
```

```
% m --> Resultado da seleção de misturas
```

```
% rmsev --> Erro quadratico medio de validacao
```

```
% rmsep --> Erro quadratico medio de predicao
```

```
% rv --> Coeficiente de correlacao obtido na validacao
```

```
% rp --> Coeficiente de correlacao obtido na predicao
```

```
%
```

% Programas Auxiliares: SELECAO.M, SELMIST.M, CENTRALIZA.M, RLMQR.M

if length(analito) > 1

error('Este programa so trabalha com um analito de cada vez')

end

% Passo 0: Isolar o analito desejado

Ycal = Ycal(:,analito);

Yval = Yval(:,analito);

Ypred = Ypred(:,analito);

% Passo 1: Centralização na media

[Xcalc,Ycalc,Xvalc,Yvalc,Xpredc,Ypredc] = centraliza(Xcal,Ycal,Xval,Yval,Xpred,Ypred);

disp('Passo 1: Os dados foram centrados na media.')

espera = input('Tecla algo para continuar...');

% Passo 2: Selecao de variaveis

N1 = 1;

l = selecao(Xcalc,Ycalc,Xvalc,Yvalc,N1,N2);

disp('Passo 2: As variaveis foram selecionadas.')

espera = input('Tecla algo para continuar...');

% Passo 3: Selecao de amostras restritas as variaveis selecionadas no passo 2

[R,m] = selmist(Xcalc(:,l),Ycalc,Xvalc(:,l),Yvalc,1,length(l));

disp('Passo 3: As amostras foram selecionadas.')

espera = input('Tecla algo para continuar...');

% Passo 4: Calculo das estatisticas de ajuste do modelo

% 4.1. RMSEV e RMSEP

[Yvalcp,Eabsv,rmsev] = rlmqr(Xcalc(m,:),Ycalc(m,:),Xvalc,Yvalc,l);

[Ypredcp,Eabsp,rmsep] = rlmqr(Xcalc(m,:),Ycalc(m,:),Xpredc,Ypredc,l);

% 4.2. Correlacao

% Validacao

x = Yvalc;

```
y = Yvalcp;
n = length(y);
num = n*sum(x.*y) - sum(x)*sum(y);
den = sqrt( ( n*sum(x.^2) - (sum(x))^2 ) * ( n*sum(y.^2) - (sum(y))^2 ) );
rv = num/den;

% Predicao
x = Ypredc;
y = Ypredcp;
n = length(y);
num = n*sum(x.*y) - sum(x)*sum(y);
den = sqrt( ( n*sum(x.^2) - (sum(x))^2 ) * ( n*sum(y.^2) - (sum(y))^2 ) );
rp = num/den;

% Apresentacao dos resultados

disp('Variaveis selecionadas: ')
l
disp('Amostras de calibracao selecionadas: ')
m

disp(['RMSEV: ' num2str(rmsev)])
disp(['RMSEP: ' num2str(rmsep)])
disp(['Correlacao na Validacao: ' num2str(rv)])
disp(['Correlacao na Predicao: ' num2str(rp)])

% Graficos de Valor Predito vs Valor de Referencia
% Validacao
close all
figure(1)
plot(Yval,Yvalcp + mean(Ycal),'o')
title(['Validacao: RMSEV = ' num2str(rmsev) ', Correlacao = ' num2str(rv)])
xlabel('Valor de referencia')
ylabel('Valor predito')
% Reta bissetriz
minimo = min([Yval;Yvalcp + mean(Ycal)]);
maximo = max([Yval;Yvalcp + mean(Ycal)]);
line([minimo,maximo],[minimo,maximo])
```

```

% Predicao
figure(2)
plot(Ypred,Ypredcp + mean(Ycal),'o')
title(['Predicao: RMSEP = ' num2str(rmse) ', Correlacao = ' num2str(rp)])
xlabel('Valor de referencia')
ylabel('Valor predito')
% Reta bissetriz
minimo = min([Ypred;Ypredcp + mean(Ycal)]);
maximo = max([Ypred;Ypredcp + mean(Ycal)]);
line([minimo,maximo],[minimo,maximo])

```

Apêndice C5- Seleção de Variáveis e Amostra por APS - Algoritmo selmist

```

function [R,m,rmseopt,Mopt] = selmist(Acal,Ccal,Ateste,Cteste,N1,N2,analitos)

% [R,m,rmseopt,Mopt] = selmist(Acal,Ccal,Ateste,Cteste,N1,N2,analitos)
%
% Este programa determina as melhores misturas-padrão
% para calibracao multivariada, empregando o algoritmo APS.
% Critério Adotado: RMSEP obtido na série de teste
% Método de Calibração: RLM via pseudo-inversa c/ QR
%
% Acal --> (Nmist_cal x Nlambdas)
% Ccal --> (Nmist_cal x Ncomp)
% Ateste --> (Nmist_teste x Nlambdas)
% Cteste --> (Nmist_teste x Ncomp)
% N1,N2 --> menor e maior número de misturas a serem testadas
%      (ex: N1=5,N2=10 fará com que o programa busque a melhor seleção
%      que possua no mínimo 5 e no máximo 10 misturas)
% analitos --> vetor contendo os indices dos analitos que serao
%      levados em consideracao nas matrizes Ccal e Cteste
%      (ex: [1,3,5] ou simplesmente 4)
%      Default: Todos os analitos.
%
% R --> menor rmsep obtido em funcao do n. de misturas usadas

```

```

% (obs: R(N) vale zero para N<N1)
% m --> melhor seleção encontrada
% rmseopt --> rmsep em função do ponto de partida do APS para o melhor
% número de misturas encontrado.
% Mopt --> Matriz (N2 x (N2-N1+1)): Resultado da Seleção
% em função do n. de misturas a serem selecionadas
%
% Programas Auxiliares: APS.M, SUMSQR.M
%
% Observação: Esse programa nao faz centralizacao na media nem auto-escalamento.
% Se necessario, esses procedimentos devem ser efetuados externamente.

if nargin == 7
    Ccal = Ccal(:,analitos);
    Cteste = Cteste(:,analitos);
end

disp('Está sendo adotado o critério de menor RMSEP obtido via RLM com QR')

Nmist_cal = size(Acal,1);
Nmist_teste = size(Cteste,1);
Ncomp = size(Cteste,2);

M = zeros(N2,Nmist_cal); % No MATLAB, a inicializacao previa de vetores
                        % torna a execucao do programa mais rapida
for i=1:Nmist_cal
    M(:,i) = (aps(Acal',i,N2)); % Seleciona N2 misturas
    %if rem(i,10)==0,i,end
end

R = zeros(1,N2);
Mopt = zeros(N2,N2);
rmsep = zeros(N2,Nmist_cal);

% Determinação do RMSEP p/ todas as seleções efetuadas

for N=N1:N2
    for i=1:Nmist_cal

```

```
%if rem(i,100)==0,i,end
mist = M(1:N,i);
Acal2=Acal(mist,:);

Ccal2=Ccal(mist,:);

W = Acal2\Ccal2; % Determina o modelo via RLM + QR
Cpred = Ateste*W;

rmsep(N,i)=sqrt(sumsqr(Cpred-Cteste)/(Nmist_teste*Ncomp));
%if rem(i,10)==0,i,end
end
[R(N) imin] = min(rmse(N,:));
Mopt(1:N,N)=M(1:N,imin);
end

[Rbest,Nbest] = min(R(N1:N2));
Nbest = Nbest+N1-1;
rmsepopt = rmsep(Nbest,:);
m = (Mopt(1:Nbest,Nbest))';

disp('Menor RMSEP obtido:')
disp(Rbest)
disp('Número ideal de misturas de calibração:')
disp(Nbest)

if N2>N1
figure,plot([N1:N2],R(N1:N2)),grid
axis([N1 N2 -inf inf])
xlabel('Número de Misturas de Calibração')
ylabel('RMSEP')
end
```

Apêndice C6- Seleção de Variáveis e Amostra por APS

- Algoritmo centraliza

```

function [Xcalc,Ycalc,Xvalc,Yvalc,Xpredc,Ypredc] = centraliza(Xcal,Ycal,Xval,Yval,Xpred,Ypred);

% Função CENTRALIZA - Centralização na media das colunas
% Cada coluna das matrizes de calibração Xcal e Ycal eh centrada em sua media.
% Esse procedimento elimina a necessidade de se incluir um termo de bias no modelo de regressao
% As colunas das matrizes de validação (Xval,Yval) e predição (Xpred,Ypred) sao centradas na
% media das colunas de Xcal e Ycal.
%
% [Xcalc,Ycalc,Xvalc,Yvalc,Xpredc,Ypredc] = centraliza(Xcal,Ycal,Xval,Yval,Xpred,Ypred);
%
% Os argumentos de saida correspondem as matrizes centralizadas
% Caso Xpred e Ypred nao sejam fornecidos, as matrizes Xpredc e Ypredc devolvidas pelo programa
% serao vazias.
% Centralização do X
m = mean(Xcal);
nmistcal = size(Xcal,1);
nmistval = size(Xval,1);
Xcalc = Xcal - repmat(m,nmistcal,1);
Xvalc = Xval - repmat(m,nmistval,1);
% Centralização do Y
my = mean(Ycal);
Ycalc = Ycal - repmat(my,nmistcal,1);
Yvalc = Yval - repmat(my,nmistval,1);
if nargin > 4
    nmistpred = size(Xpred,1);
    Xpredc = Xpred - repmat(m,nmistpred,1);
    Ypredc = Ypred - repmat(my,nmistpred,1);
else
    Xpredc = [];
    Ypredc = [];
end

```

Apêndice C7- Seleção de Variáveis e Amostra por APS - Algoritmo rlmqr

```
function [Cpred,Eabs,rmsep] = rlmqr(Acal,Ccal,Ateste,Cteste,lambda,analitos)
```

```
% [Cpred,Eabs,rmsep] = rlmqr(Acal,Ccal,Ateste,Cteste,lambda,analitos);
```

```
%
```

```
% Procura determinar as concentracoes das misturas de teste
```

```
% usando Regressão Linear Múltipla com Decomposição QR nos
```

```
% comprimentos de onda indicados no vetor lambda
```

```
%
```

```
% analitos --> Componentes a serem analisados
```

```
% (Default: Todos os Componentes)
```

```
%
```

```
% Cpred --> Concentrações Preditas na Série de Teste
```

```
% Eabs --> Erro Absoluto Médio: mean(abs(Cpred-Cteste))
```

```
% rmsep --> Erro Quadratico Médio: sqrt(mean((Cpred-Cteste2).^2))
```

```
if nargin<6
```

```
    analitos = [1:size(Ccal,2)];
```

```
end
```

```
disp('Usando RLM com QR !')
```

```
Acal2 = Acal(:,lambda);
```

```
Ateste2 = Ateste(:,lambda);
```

```
Ccal2 = Ccal(:,analitos);
```

```
Cteste2 = Cteste(:,analitos);
```

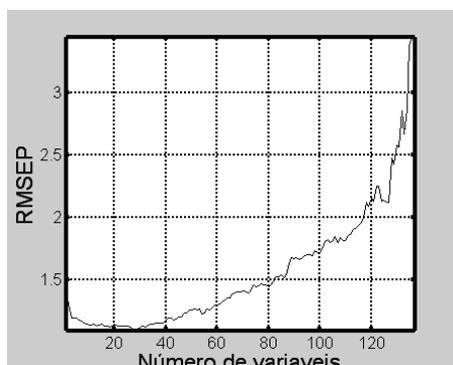
```
W = Acal2\Ccal2; % Determina o modelo via RLM + QR
```

```
Cpred = Ateste2*W;
```

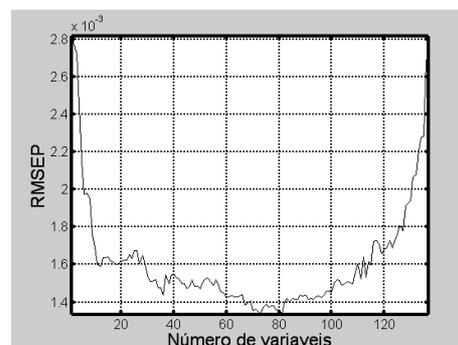
```
Eabs = mean(abs(Cpred-Cteste2));
```

```
rmsep = sqrt(mean((Cpred-Cteste2).^2));
```

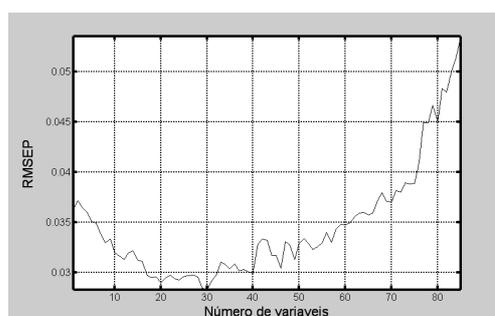
Apêndice D – Gráficos de Seleção de variáveis



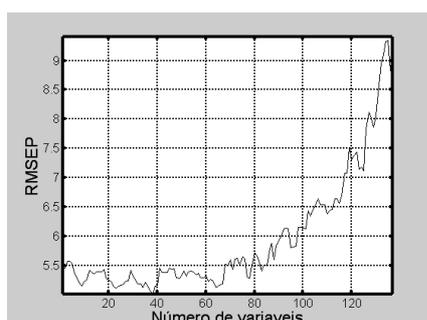
Índice de Cetano



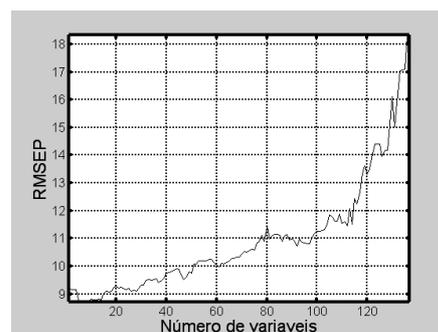
Densidade



Enxofre

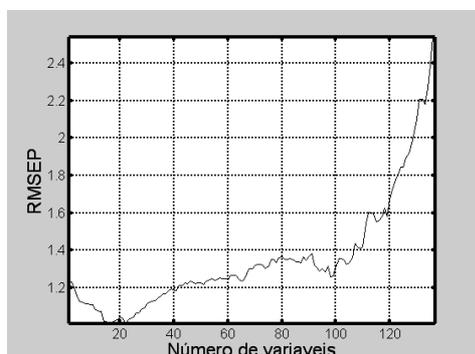


T - 50 %

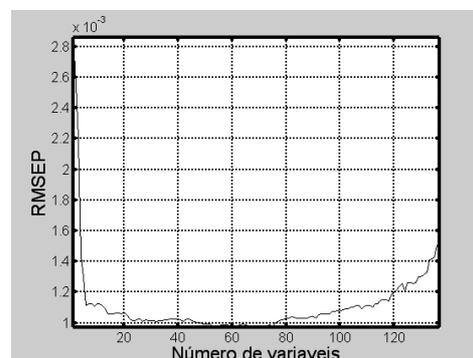


T - 85 %

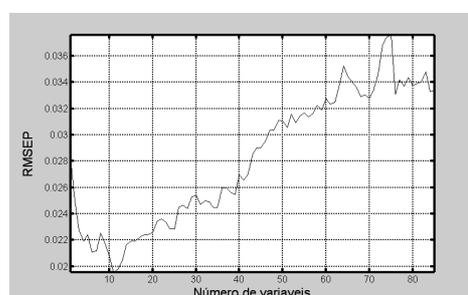
Figura D1. Gráficos dos valores de RMSEP versus número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis, na estimativa dos diversos parâmetros de qualidade de óleo diesel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 850 – 1.100 nm, com caminho óptico de 50 mm.



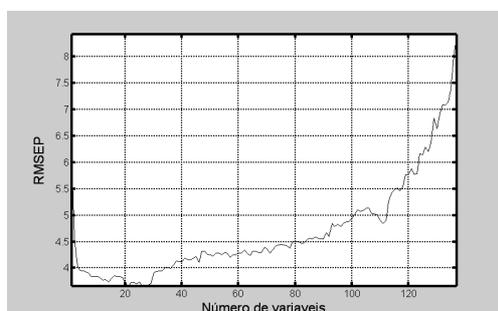
Índice de Cetano



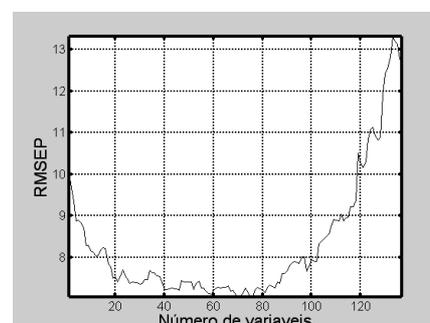
Densidade



Enxofre

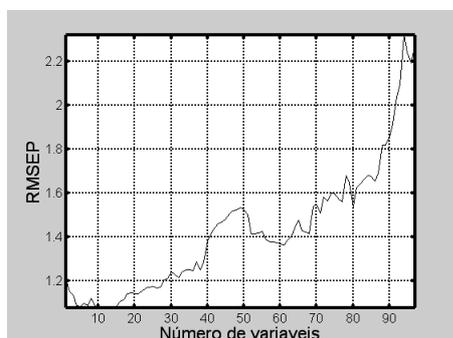


T-50%

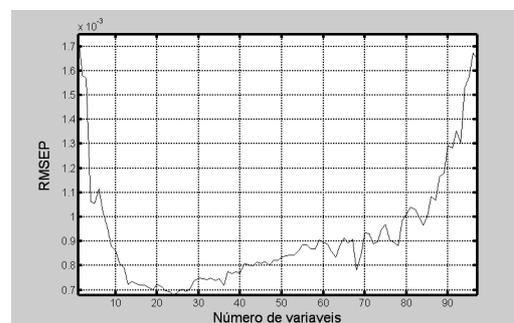


T-85%

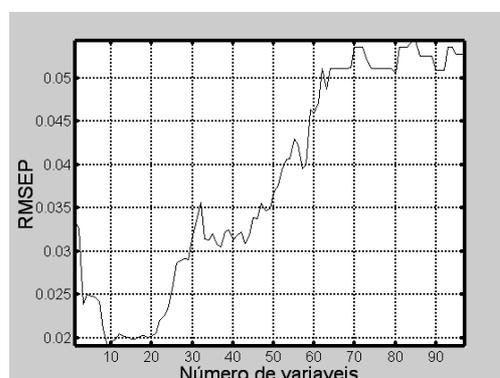
Figura D2. Gráficos dos valores de *RMSEP* versus número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis, na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano para as amostras do conjunto de previsão, na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10 mm.



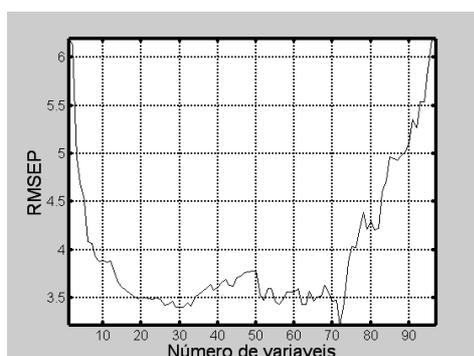
Índice de Cetano



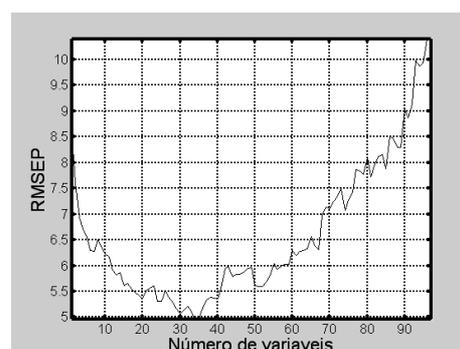
Densidade



Enxofre

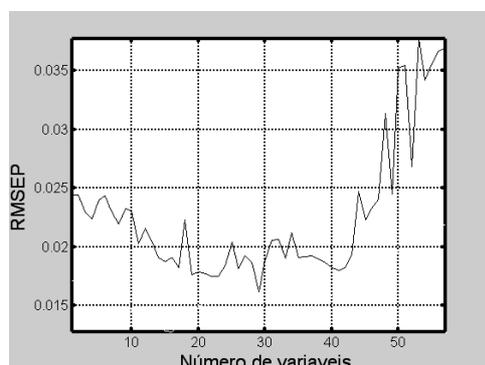


T-50%

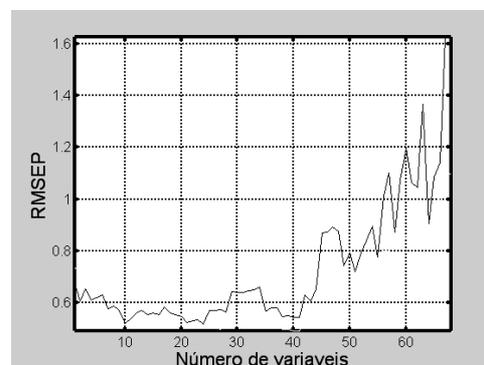


T-85%

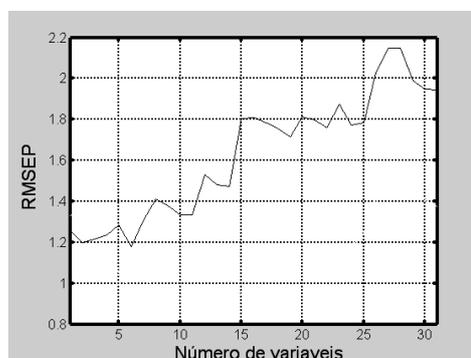
Figura D3. Gráficos dos valores de RMSEP versus número de comprimentos de onda empregados nos modelos MLR com seleção de variáveis, na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano para as amostras do conjunto de previsão, na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1 mm.



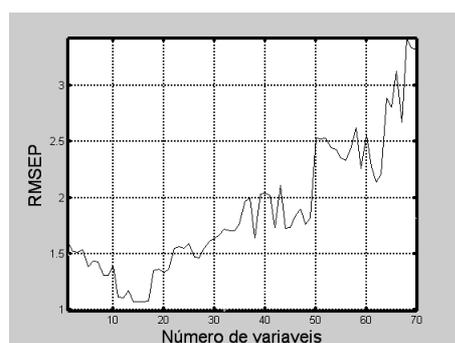
Densidade



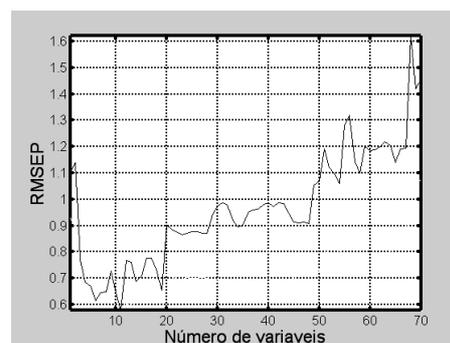
Extrativos



Holocelulose



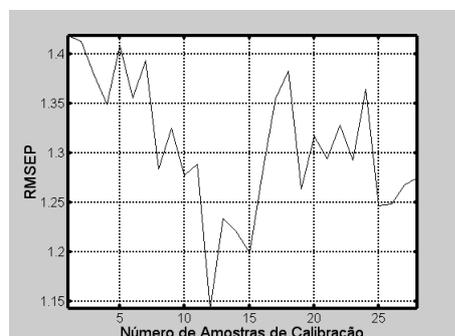
Lignina



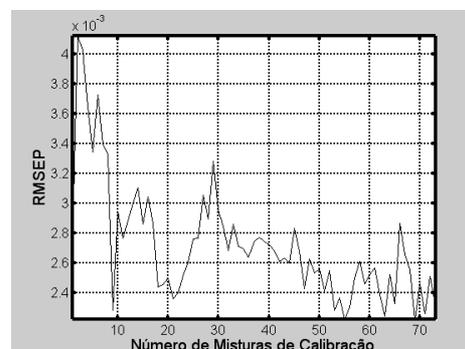
Rendimento

Figura D4. Gráficos dos valores de *RMSEP* versus número de comprimentos de onda empregados nos modelos *MLR* com seleção de variáveis, na estimativa dos diversos parâmetros de madeira de eucalipto para as amostras do conjunto de previsão, na região espectral de 1.100 – 2.500 nm.

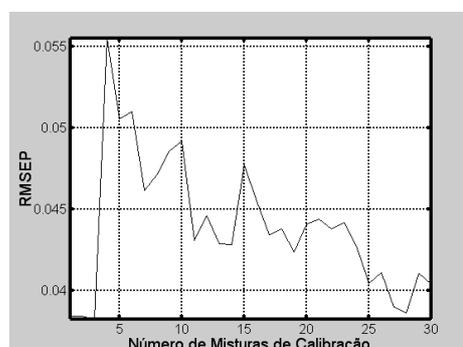
Apêndice E – Gráficos de Seleção de Amostras



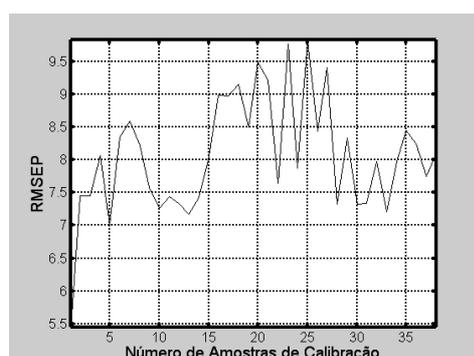
Índice de Cetano



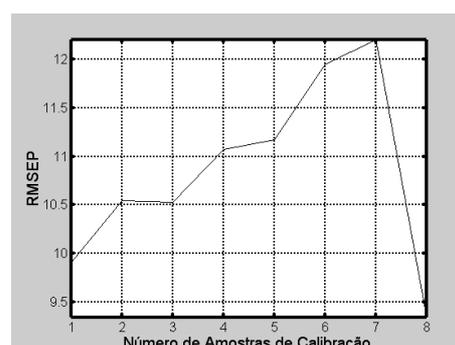
Densidade



Enxofre

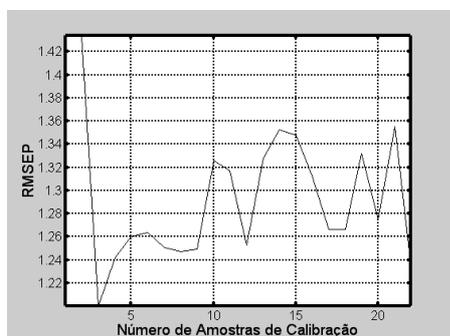


T-50 %

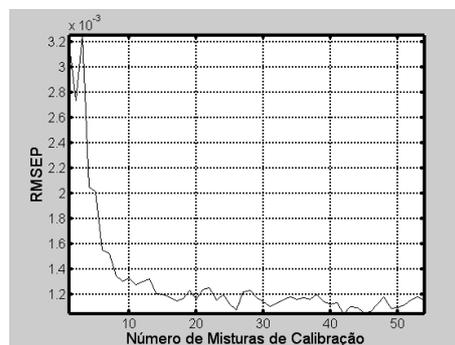


T-85 %

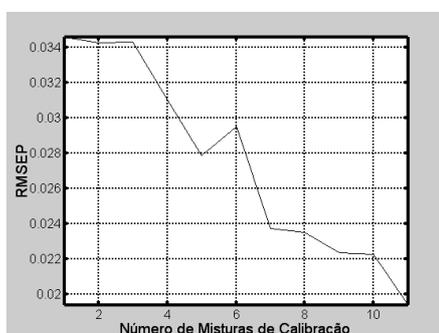
Figura E1. Gráficos dos valores de RMSEP versus número de amostras de calibração empregados nos modelos MLR com seleção de variáveis espectrais e amostras de calibração, na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 850 – 1.100 nm, com caminho óptico de 50mm.



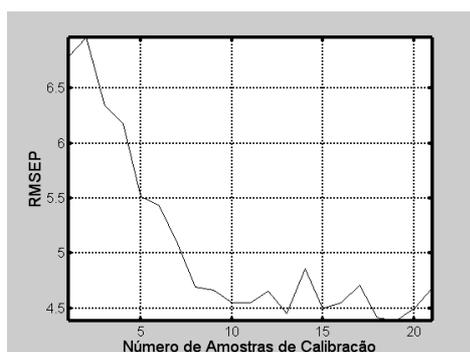
Índice de Cetano



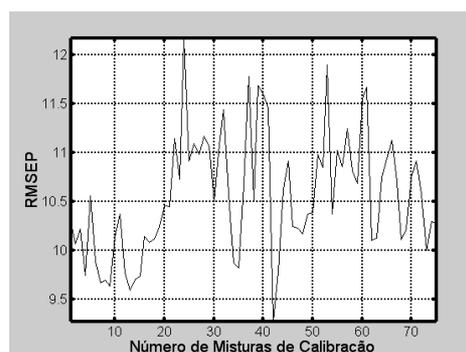
Densidade



Enxofre

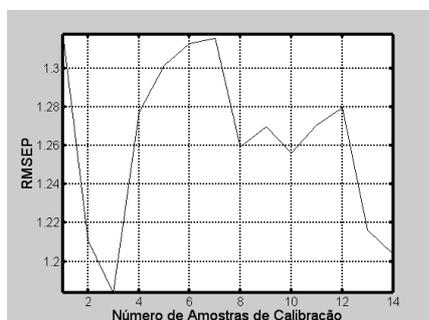


T-50 %

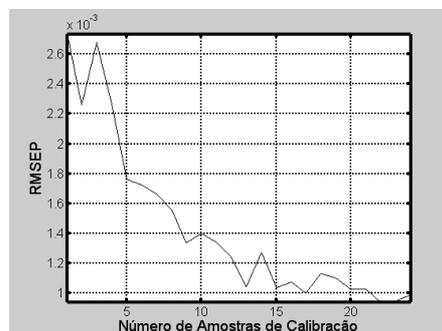


T-85 %

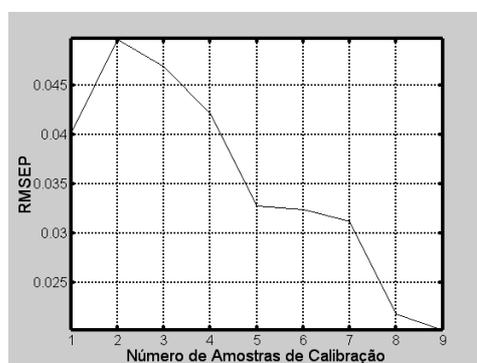
Figura E2. Gráficos dos valores de RMSEP versus número de amostras de calibração empregados nos modelos MLR com seleção de variáveis espectrais e amostras de calibração, na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10mm.



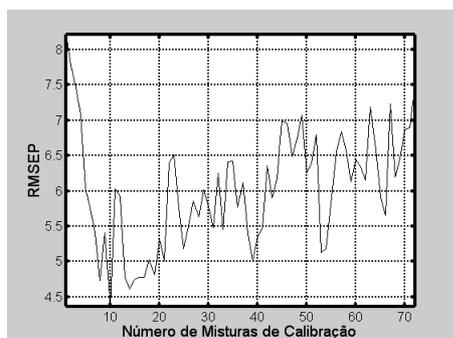
Índice de Cetano



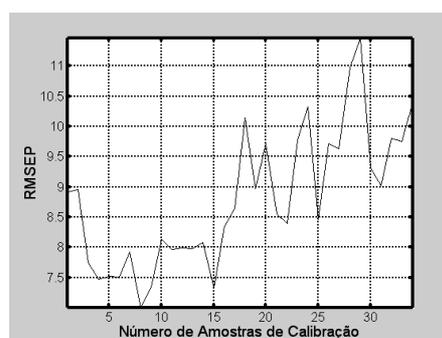
Densidade



Enxofre

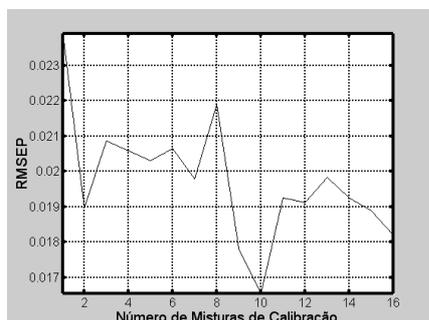


T-50 %

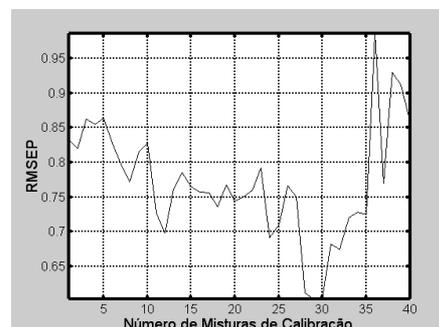


T-85 %

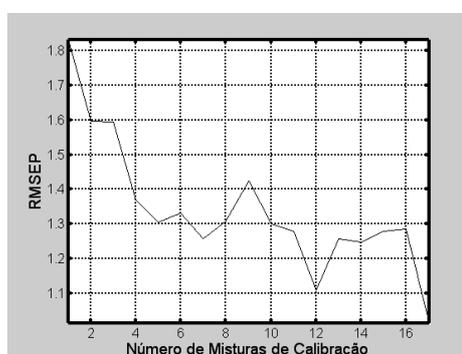
Figura E3. Gráficos dos valores de *RMSEP* versus número de amostras de calibração empregados nos modelos *MLR* com seleção de variáveis espectrais e amostras de calibração, na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano para as amostras do conjunto de previsão na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1mm.



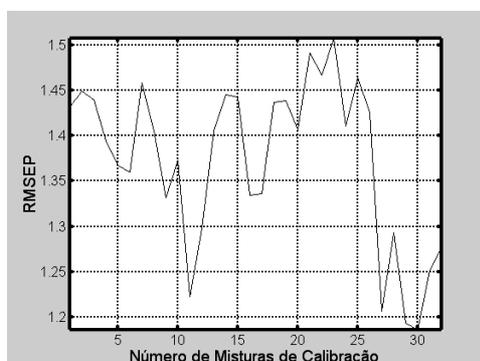
Densidade



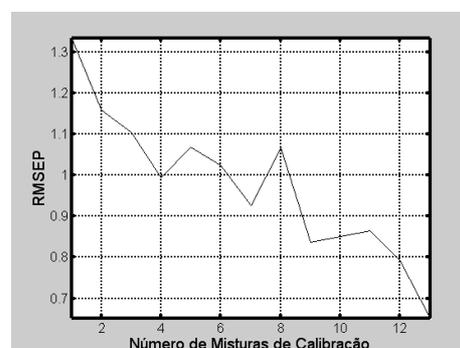
Extrativos



Holocelulose



Lignina



Rendimento

Figura E4. Gráficos dos valores de RMSEP versus número de amostras de calibração empregados nos modelos MLR com seleção de variáveis espectrais e amostras de calibração na estimativa dos diversos parâmetros de qualidade de madeira de eucalipto para as amostras do conjunto de previsão na região espectral de 1.100 – 2.500 nm.

Apêndice F – Tabelas de variáveis selecionadas

Tabela F1. Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano, usando modelos MLR, na região espectral de 850 – 1.100 nm, com caminho óptico de 50 mm.

Parâmetro de qualidade	Variáveis Selecionadas (nm)														
Cetano	856,00 904,31	856,71 908,60	859,12 911,99	861,09 914,40	862,79 916,01	866,27 920,11	869,75 923,06	872,16 926,54	876,63 930,92	879,57 937,97	882,43 941,28	887,61 958,07	890,38 967,17	894,22	898,86
Densidade	856,00 868,06 884,75 909,49 930,12	856,62 868,95 886,36 910,47 931,28	857,43 870,38 888,15 912,52 933,42	858,32 871,45 890,11 913,68 935,03	858,86 872,34 891,72 914,94 936,63	859,84 873,23 892,97 916,01 937,97	860,55 874,22 895,11 917,35 940,92	861,09 875,20 896,09 919,13 944,94	861,71 876,54 897,61 920,47 951,37	862,79 877,34 898,86 921,63 958,33	863,68 877,97 901,18 923,69 963,96	864,48 879,75 902,34 924,58 967,26	865,29 880,73 904,31 925,92 972,62	866,54 882,25 906,18 926,72	867,34 883,15 907,88 928,69
Enxofre	856,62 898,86	858,86 904,31	861,10 906,18	864,04 908,86	865,20 912,17	868,14 915,29	869,84 917,26	872,16 918,95	876,36 923,51	878,32 928,24	881,09 931,19	885,56 937,97	890,11 941,01	892,61 958,42	895,65 1.047,10
T- 50 %	856,00 890,20 931,37	857,87 891,72 934,94	858,86 894,31 937,97	859,93 897,43 940,74	861,09 900,92 944,94	862,61 904,40 958,16	864,48 907,17 966,01	866,89 910,02 972,53	869,48 911,90	872,16 914,58	875,29 917,70	877,61 920,92	880,82 923,69	884,93 925,83	887,70 927,97
T- 85 %	856,62	861,09	909,58	919,31	937,97	940,74	958,24	975,84							

Tabela F2. Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano, usando modelos MLR, na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10 mm.

Parâmetro de qualidade	Variáveis selecionadas (nm)											
Cetano	1.137,09 1.232,00	1.146,12 1.377,58	1.158,06 1.386,92	1.185,62 1.390,29	1.190,97 1.403,15	1.195,57 1.411,11	1.200,01 1.417,08	1.202,91 1.423,20	1.207,51 1.489,02	1.211,95 1.459,63	1.217,46	1.221,44
Densidade	1.109,99 1.179,80 1.202,30 1.230,32 1.418,15	1.119,03 1.181,94 1.203,22 1.235,52 1.423,35	1.126,07 1.183,32 1.204,29 1.356,15 1.432,54	1.135,71 1.186,38 1.205,36 1.366,41 1.438,20	1.141,38 1.188,83 1.207,05 1.373,75 1.459,48	1.146,12 1.190,97 1.208,73 1.378,65 1.480,60	1.151,17 1.194,19 1.211,18 1.386,77	1.158,83 1.195,72 1.214,55 1.393,04	1.162,04 1.197,25 1.216,54 1.398,25	1.168,78 1.199,24 1.218,99 1.403,15	1.172,76 1.200,31 1.221,59 1.406,67	1.175,51 1.201,23 1.225,26 1.412,18
Enxofre	1.153,62	1.186,08	1.195,72	1.202,61	1.356,15	1.386,77	1.402,23	1.411,87	1.417,69	1.459,63	1.479,38	
T- 50 %	1.137,70 1.377,58	1.146,12 1.388,91	1.158,83 1.393,65	1.179,80 1.402,38	1.186,23 1.411,41	1.191,59 1.417,23	1.195,57 1.423,66	1.198,63 1.455,19	1.202,00 1.460,55	1.209,04	1.215,01	1.220,98
T- 85 %	1.100,20 1.163,73 1.191,59 1.206,28 1.229,09 1.393,65 1.453,36	1.106,17 1.167,40 1.192,96 1.207,51 1.235,37 1.395,49 1.459,94	1.109,99 1.170,00 1.194,34 1.209,34 1.239,96 1.398,25 1.485,81	1.119,03 1.172,60 1.195,72 1.209,34 1.356,15 1.401,16	1.125,15 1.175,97 1.197,40 1.211,49 1.362,12 1.403,15	1.129,59 1.179,19 1.198,48 1.212,56 1.366,71 1.406,97	1.137,09 1.18225 1.199,39 1.213,94 1.373,75 1.410,80	1.143,67 1.183,93 1.200,62 1.215,01 1.379,72 1.417,23	1.148,42 1.186,08 1.201,69 1.216,85 1.383,24 1.423,35	1.151,48 1.187,15 1.202,61 1.219,29 1.386,77 1.429,32	1.153,47 1.188,22 1.203,68 1.221,74 1.389,06 1.434,22	1.158,67 1.190,21 1.205,36 1.224,04 1.391,36 1.441,11

Tabela F3. Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de diesel interior e metropolitano, usando modelos MLR, na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1 mm.

Parâmetro de qualidade	Variáveis selecionadas (nm)											
Cetano	1.876,62 2.490,17	2.161,14 2.495,74	2.256,48	2.299,88	2.304,70	2.308,78	2.313,23	2.339,20	2.352,18	2.392,98	2.453,82	2.476,08
Densidade	1.876,62 2.421,18	2.251,65 2.450,11	2.269,83 2.453,82	2.305,07 2.458,27	2.309,15 2.461,98	2.313,23 2.468,29	2.321,76 2.471,25	2.339,57 2.474,96	2.345,87 2.483,50	2.351,07 2.490,17	2.354,78 2.493,14	2.404,48 2.496,11
Enxofre	1.867,35	2.256,85	2.304,70	2.309,15	2.312,86	2.318,79	2.338,83	2.482,01	2.489,80			
T- 50 %	1.876,62 2.313,23 2.355,15 2.404,48 2.443,80 2.475,71	2.252,39 2.315,08 2.359,23 2.407,45 2.447,14 2.477,19	2.257,22 2.318,79 2.361,82 2.411,53 2.450,85 2.479,41	2.266,49 2.325,84 2.365,53 2.414,87 2.454,19 2.482,75	2.272,80 2.329,92 2.368,87 2.419,32 2.456,04 2.486,46	2.288,75 2.334,37 2.381,48 2.423,03 2.458,64 2.489,06	2.292,83 2.338,83 2.385,56 2.425,26 2.460,13 2.491,29	2.301,36 2.341,42 2.389,65 2.427,48 2.461,98 2.492,77	2.305,07 2.344,02 2.392,98 2.431,56 2.465,32 2.495,74	2.307,67 2.345,50 2.397,06 2.434,16 2.468,29 2.496,85	2.309,15 2.348,47 2.399,66 2.438,61 2.472,00	2.311,75 2.351,44 2.402,26 2.441,21 2.473,85
T- 85 %	1.876,62 2.344,39 2.461,98	2.112,18 2.347,36 2.464,95	2.252,39 2.352,55 2.467,54	2.281,33 2.392,98 2.471,25	2.300,25 2.404,48 2.474,59	2.300,25 2.423,03 2.477,56	2.306,92 2.438,24 2.486,09	2.309,15 2.442,69 2.489,80	2.313,23 2.446,77 2.492,77	2.316,94 2.450,48 2.495,37	2.335,12 2.453,82	2.339,20 2.457,90

Tabela F4. Identificação dos comprimentos de onda selecionados pelo APS na estimativa dos diversos parâmetros de qualidade de madeira de eucalipto usando modelos MLR na região espectral de 1.100 – 2.500 nm.

Parâmetro de qualidade	Variáveis selecionadas (nm)									
Densidade	1.124,00 1.938,00 2.314,00	1.202,00 1.952,00 2.336,00	1.342,00 2.028,00 2.364,00	1.398,00 2.060,00 2.402,00	1.420,00 2.096,00 2.448,00	1.654,00 2.234,00 2.462,00	1.882,00 2.258,00 2.468,00	1.896,00 2.276,00 2.480,00	1.906,00 2.288,00 2.488,00	1.916,00 2.302,00
Extrativos	1.126,00	1.418,00	1.882,00	1.896,00	1.918,00	2.236,00	2.260,00	2.350,00	2.466,00	2.488,00
Holocelulose	1.126,00	1.896,00	1.918,00	2.026,00	2.234,00	2.382,00				
Lignina	1150.00 2294.00	1414.00 2316.00	1654.00 2462.00	1892.00 2488.00	1910.00	1926.00	2022.00	2234.00	2260.00	2274.00
Rendimento	1.150,00 2.488,00	1.416,00	1.894,00	1.904,00	1.918,00	2.158,00	2.236,00	2.260,00	2.298,00	2.458,00

Apêndice G – Tabelas de amostras selecionadas

Tabela G1. Identificação das amostras de calibração com seus respectivos valores dos parâmetros de qualidade de óleo diesel (entre parênteses), selecionadas pelo APS, na região espectral de 850 – 1.100 nm, com caminho óptico de 50 mm.

Parâmetro de qualidade	Amostras selecionadas					Variação do parâmetro
Cetano	R2160458 (40,8) R2130350 (46,3) R2020507 (42,9)	R2110451 (41,0) Q2110451 (45,5) R2040353 (48,3)	R2170450 (41,9) R2110305 (44,4)	V2180301 (42,0) Q2060503 (44,0)	R2020502 (42,3) Q2060504 (43,5)	39,9 – 48,3
Densidade	V2060551 (0,8457) R2260362 (0,8476) R2090453 (0,8491) R2110350 (0,8503) Q2210301 (0,8531) R2220407 (0,8546) Q2120453 (0,8556) Q2110453 (0,8561) R2030452 (0,8576) R2260355 (0,8584) R2160454 (0,8606)	Q2060508 (0,8458) R2230402 (0,8477) Q2160402 (0,8491) V2120401 (0,8509) Q2220353 (0,8536) Q2220352 (0,8551) R2090454 (0,8556) Q2210303 (0,8562) R2220401 (0,8576) Q2120450 (0,8586) R2160455 (0,8614)	Q2060550 (0,8463) Q2190351 (0,8479) Q2060501 (0,8495) Q2120402 (0,8526) R2220302 (0,8539) Q2250350 (0,8552) Q2040301 (0,8557) R2030408 (0,8573) R2030458 (0,8579) R2260357 (0,8592) Q2080453 (0,8619)	R2130351 (0,8473) Q2110451 (0,8483) S2030401 (0,8496) R2050301 (0,8526) R2090501 (0,8540) R2020504 (0,8553) R2110305 (0,8558) R2080405 (0,8573) S2080401 (0,8580) R2090401 (0,8600) Q2240401 (0,8634)	R2160457 (0,8475) R2030406 (0,8488) R2020551 (0,8498) R2040353 (0,8529) R2110354 (0,8546) R2020553 (0,8555) R2110302 (0,8559) R2250411 (0,8573) R2260350 (0,8582) R2130304 (0,8602) R2040450 (0,8648)	0,8443 – 0,8658
Enxofre	R2110451 (0,10)	Q2110451 (0,15)	R2020507 (0,19)			0,05 – 0,29
T- 50 %	Q2070501 (260,0)					255,1 – 291,1
T- 85 %	V2210350 (325,0) R2110454 (345,6)	Q2110451 (327,9) R2130350 (347,4)	R2170455 (337,1) R2020507 (362,4)	R2160454 (341,7)	R2180354 (342,4)	319,4 – 370,1

Tabela G2. Identificação das amostras de calibração com seus respectivos valores dos parâmetros de qualidade de óleo diesel (entre parênteses), selecionadas pelo APS, na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10 mm.

Parâmetro de qualidade	Amostras selecionadas					Variação do parâmetro
Cetano	R2080403 (41,1)	R2020502 (42,3)	Q2110451 (45,5)			39,9 – 48,3
Densidade	Q2060507 (0,8443) Q2190351 (0,8479) V2120401 (0,8509) Q2210302 (0,8539) Q2080452 (0,8554) Q2110453 (0,8561) R2250409 (0,8578) R2270350 (0,8591) R2040450 (0,8648)	Q2060508 (0,8458) Q2110451 (0,8483) Q2190353 (0,8511) Q2220350 (0,8539) R2020553 (0,8555) Q2210303 (0,8562) R2110301 (0,8579) R2160454 (0,8606) R2080403 (0,8658)	Q2060550 (0,8463) Q2160402 (0,8491) R2040353 (0,8529) R2110354 (0,8546) Q2120453 (0,8556) R2250406 (0,8568) R2130303 (0,8586) R2110454 (0,8607)	R2130351 (0,8473) S2030401 (0,8496) Q2080506 (0,8536) Q2210305 (0,8549) R2020506 (0,8558) V2170450 (0,8574) R2260358 (0,8587) R2180304 (0,8611)	R2160457 (0,8475) R2110350 (0,8503) R2040355 (0,8537) R2020504 (0,8553) R2110302 (0,8559) R2030452 (0,8576) R2020552 (0,8588) Q2080455 (0,8629)	0,8443 – 0,8658
Enxofre	R2020504 (0,09) Q2060550 (0,17) R2040453 (0,27)	V2170450 (0,10) Q2160402 (0,18)	Q2110451 (0,15) R2170450 (0,23)	Q2060508 (0,16) Q2120452 (0,24)	Q2180401 (0,16) R2250409 (0,24)	0,05 – 0,29
T- 50 %	Q2090550 (255,1) R2020502 (265,6) R2030403 (272,9) R2020507 (282,3)	Q2190351 (260,9) Q2060550 (267,0) R2270351 (273,4) R2160455 (282,9)	Q2210301 (262,1) R2260356 (267,3) R2260359 (276,7) R2070350 (289,6)	R2170450 (262,4) Q2080452 (272,1) R2020554 (280,7) R2080409 (291,1)	R2110451 (263,0) Q2120453 (272,6) R2040453 (281,0)	255,1 – 291,1
T- 85 %	Q2160401 (320,2) Q2210301 (326,1) R2160453 (331,1) Q2060552 (335,9) Q2120453 (341,1) R2020505 (346,8) R2040353 (354,2) R2190302 (358,6) R2040357 (365,1)	Q2160408 (320,9) Q2060555 (327,2) R2090450 (332,4) R2070350 (339,0) Q2120454 (341,6) R2130350 (347,4) R2020552 (354,4) R2040355 (360,8) R2110304 (370,1)	Q2060504 (323,0) Q2110451 (327,9) R2170454 (332,6) R2170450 (339,3) R2160454 (341,7) R2130351 (348,7) R2110302 (354,7) R2130306 (361,3)	V2210350 (325,0) R2230402 (329,7) Q2210304 (333,0) S2080402 (340,7) Q2210306 (344,0) Q2080506 (349,8) R2110354 (358,3) R2020507 (362,4)	Q2220350 (325,6) Q2210311 (330,6) Q2060508 (334,3) R2270351 (340,9) R2080403 (345,3) R2020501 (352,4) R2040356 (358,4) R2110353 (363,5)	319,4 – 370,1

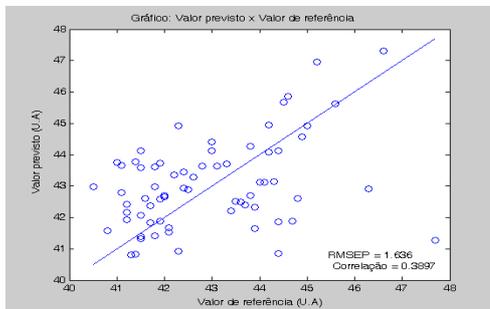
Tabela G3. Identificação das amostras de calibração com seus respectivos valores dos parâmetros de qualidade de óleo diesel (entre parênteses), selecionadas pelo APS, na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1 mm.

Parâmetro de qualidade	Amostras selecionadas					Variação do parâmetro
Cetano	R2040453 (39,9)	R2130306 (43,7)	Q2110451 (45,5)			39,9 – 48,3
Densidade	Q2070501 (0,8462) R2170450 (0,8506) R2050350 (0,8550) R2270351 (0,8574) R2040453 (0,8638)	R2130350 (0,8476) R2220303 (0,8534) S2020550 (0,8555) R2030452 (0,8576) R2080403 (0,8658)	R2030403 (0,8480) R2180353 (0,8537) R2140351 (0,8559) R2110301 (0,8579)	Q2160404 (0,8491) Q2220350 (0,8539) R2260359 (0,8567) R2180356 (0,8587)	S2030401 (0,8496) Q2210358 (0,8549) R2110306 (0,8574) R2130304 (0,8602)	0,8443 – 0,8658
Enxofre	R2170454 (0,10) Q2120452 (0,24)	Q2110451 (0,15) Q2080456 (0,25)	Q2060504 (0,16) S2030401 (0,26)	Q2080452 (0,22) R2030402 (0,27)	S2250401 (0,23)	0,05 – 0,29
T- 50 %	Q2090550 (255,1) R2260359 (276,7)	R2170450 (262,4) R2260358 (277,5)	R2020553 (264,3) R2040453 (281,0)	R2130350 (273,1) R2250406 (283,6)	Q2110451 (275,0) R2180350 (288,2)	255,1 – 291,1
T- 85 %	Q2060504 (323,0) R2260358 (347,2)	Q2110451 (327,9) R2030402 (347,9)	S2030401 (333,8) V2180301 (360,6)	R2020553 (341,8)	Q2080456 (343,2)	319,4 – 370,1

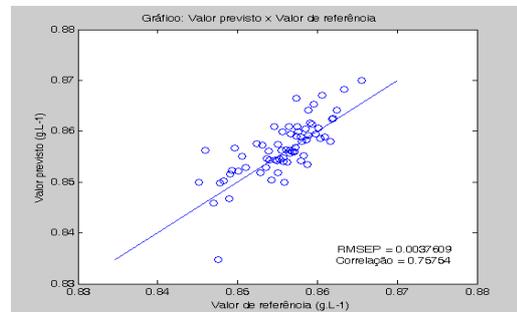
Tabela G4. Identificação das amostras de calibração com seus respectivos valores dos parâmetros de qualidade de madeira de eucalipto (entre parênteses), selecionadas pelo APS, na região espectral de 1.100 – 2.500 nm.

Parâmetro de qualidade	Amostras selecionadas				Varição do parâmetro
Densidade	AE 385 C145 (0,4730) AE-374 VR3715 (0,5640) AE 402 C022 (0,4790)	AE248 TA793 (0,4980) AE 377 VR3709 (0,4950) AE249 C026 (0,5060)	AE261 C189 (0,5270) AE248 C186 (0,5160)	AE 375 VR3715 (0,5460) AE 375 VR3715 (0,5550)	0,4170 – 0,5740
Extrativos	MG3003 TR0203 (1,46) C219 T-5 (1,85) AE 367 VR3806 (1,59) VR3709 T-8 (3,27) AE 375 VR3715 (4,38) AE376 VR3715 (2,56) AE 261 C185 (1,88) C041 T-3 (3,98)	AE367 VR3806 (1,59) AE 402 C022 (0,91) AE261 C081 (2,68) AE303 C072 (3,68) AE248 C106 (4,77) AE 390 C145 (3,03) AE 390 C022 (2,24)	MG3.012 P002 (1,63) AE 303 B261 (2,21) AE249 C026 A-10 (2,70) AE249 C026 (3,89) AE249 C026 A-25 (1,09) AE 377 VR3709 (1,38) MG3.012 S006 (2,57)	AE372 VR3806 (1,72) TC30 T-4 (2,35) AE 303 B261 (2,21) C041 T-3 (4,43) AE 389 C022 (2,09) AE 389 C022 (2,09) AE 373 VR3806 (2,97)	1,05 – 4,77
Holocelulose	AE248 C106 (66,34) AE249 C026 A-10 (69,4) AE 261 C185 (67,42) AE376 VR3796 (70,90) AE 375 VR3796 (68,87)	AE261 C189 (66,60) AE262 G34 (69,82) MG3.012 P002 (69,22) AE248 C186 (65,28)	AE 303 B261 (69,18) MG3.012 S006 (79,19) AE304 B261 (68,29) MG3.012 P002 (69,22)	AE249 C026 (68,75) AE261 C081 (66,51) AE249 C026 A-24 (69,92) AE 375 VR3796 (69,75)	68,75 – 73,51
Lignina	AE249 C026 A-13 (26,64) C219 T-5 (27,57) AE 390 C145 (28,73) AE376 VR3796 (25,45) AE 354 C101 (28,58) AE261 C081 (30,81) AE 366 VR3806 (27,45) VR3709 T-1 (30,92)	MG3.012 S006 (27,74) AE 367 VR3806 (30,55) AE 354 C197 (30,81) AE249 C026 A-12 (27,67) AE 385 C145 (27,66) AE261 C189 (30,96) MG3.012 P002 (29,12) AE 373 VR3806 (30,47)	AE262 G34 (27,76) AE376 VR3715 (29,80) AE249 TA793 (31,63) AE 354 P002 (28,25) AE 373 VR3709 (29,39) AE261 C185 (31,42) AE304 B261 (28,84)	AE249 C026 A-10 (27,90) AE 375 VR3715 (29,94) AE 390 C022 (32,97) MG3003 TR0203 (28,35) AE303 C072 (29,50) TC30 T-4 (32,30) AE 377 VR3709 (28,17)	25,44 – 38,00
Rendimento	AE 261 C185 (44,90) C219 T-5 (54,71) AE 389 C022 (50,46) AE 377 VR3709 (51,91)	AE261 C081 (47,40) AE249 C026 A-24 (48,59) MG3.012 P002 (50,81)	VR3709 T-1 (51,93) AE 303 B261 (51,03) AE372 VR3806 (51,40)	MG3003 TR0203 (53,25) C041 T-3 (50,07) AE 377 VR3709 (52,32)	44,90 – 55,51

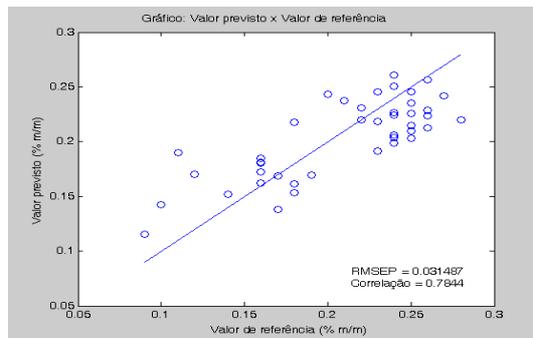
Apêndice H – Figuras de modelos APS-MLR



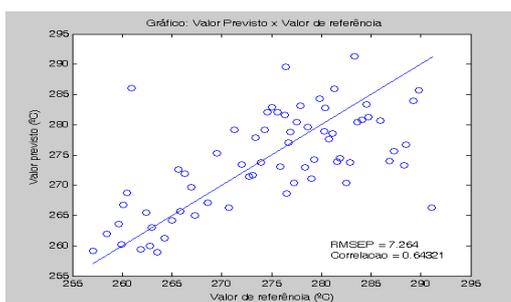
Índice de Cetano



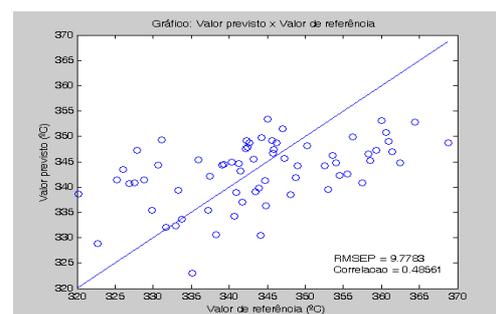
Densidade



Enxofre

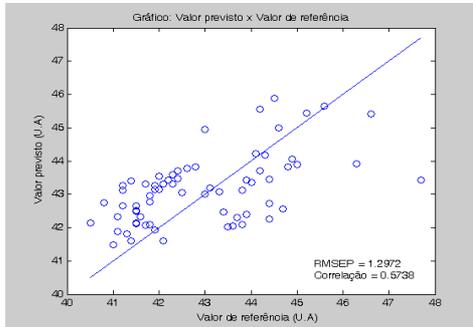


T - 50 %

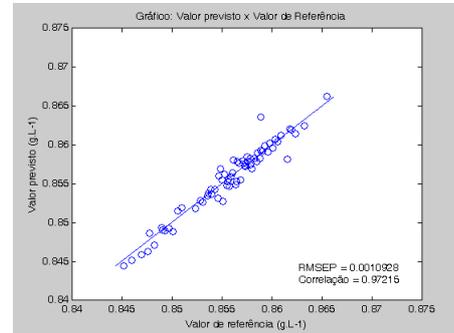


T - 85 %

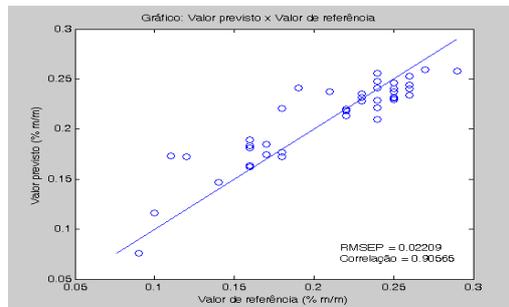
Figura H1. Gráficos dos valores previstos versus valores de referência para as amostras do conjunto de previsão, empregando modelos MLR com seleção de variáveis e amostras de calibração na estimativa dos diversos parâmetros de qualidade de diesel interior, na região espectral de 850 – 1.100 nm, com caminho óptico de 50mm.



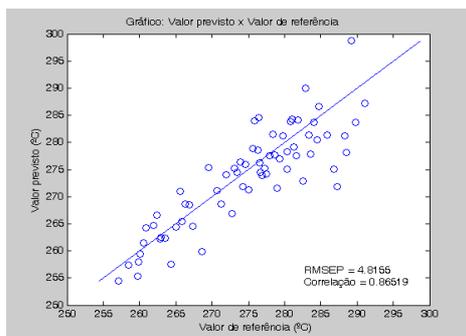
Índice de Cetano



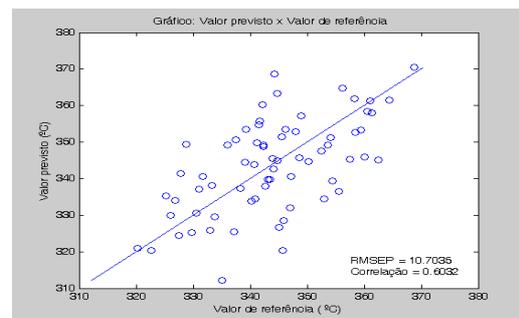
Densidade



Enxofre

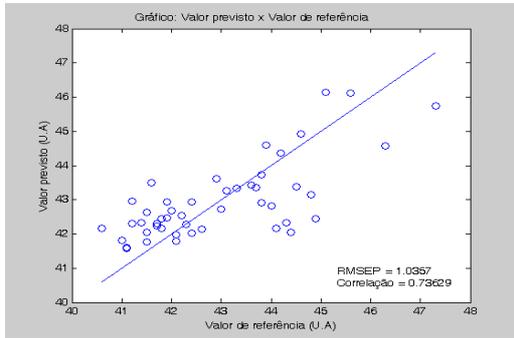


T - 50 %

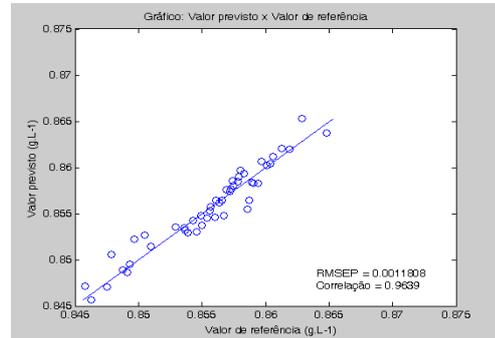


T - 85 %

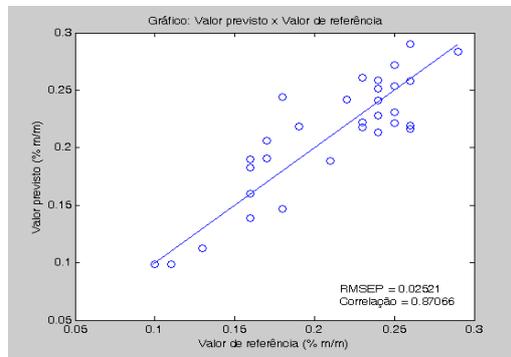
Figura H2. Gráficos dos valores previstos versus valores de referência para as amostras do conjunto de previsão, empregando modelos MLR com seleção de variáveis e amostras de calibração na estimativa dos diversos parâmetros de qualidade de diesel interior, na região espectral de 1.100 – 1.570 nm, com caminho óptico de 10mm.



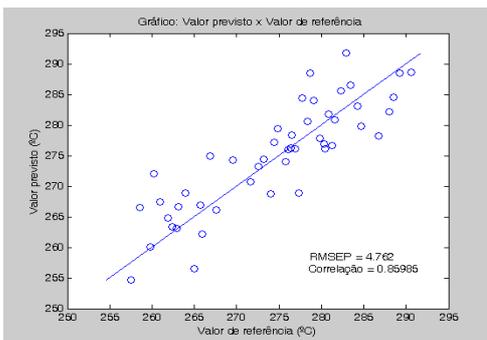
Índice de Cetano



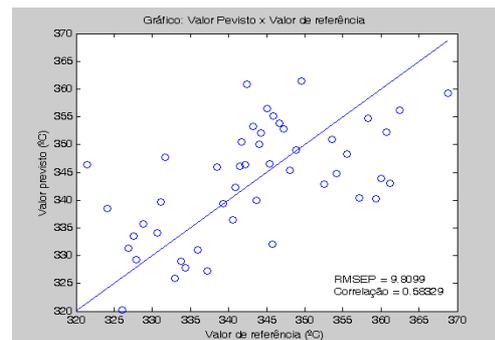
Densidade



Enxofre



T - 50 %



T - 85 %

Figura H3. Gráficos dos valores previstos versus valores de referência para as amostras do conjunto de previsão, empregando modelos MLR com seleção de variáveis e amostras de calibração na estimativa dos diversos parâmetros de qualidade de diesel interior, na região espectral de 1.570 – 2.500 nm, com caminho óptico de 1mm.

Apêndice I – Algoritmos Wavelet

Apêndice I1- Algoritmo estudo

```
clear,close all,clc

% Passo 0: Carregar os dados
% Matriz X --> Respostas Instrumentais
% Matriz Y --> Propriedade

load Xcal
Xcal = Matrix;
load Xval
Xval = Matrix;
load Xpred
Xpred = Matrix;
load Ycal
Ycal = Matrix;
load Yval
Yval = Matrix;
load Ypred
Ypred = Matrix;

% Wavelets consideradas:
% db1 - db10
% sym4 - sym10 --> Obs: sym1 - sym3 sao iguais a db1 - db3
for i=1:10
    nome_wavelet{i} = ['db' num2str(i)];
end
for i=11:17
    j = i - 7;
    nome_wavelet{i} = ['sym' num2str(j)];
end
% for i=38:42
%     j = i - 37;
%     nome_wavelet{i} = ['coif' num2str(j)];
```

```
% end

for wavelet = 1:17
    for niveis = 1:7 % Checar com wmaxlev

        % Passo 1: Decomposicao wavelet com a familia, o comprimento e o numero de
        % niveis de decomposicao especificado

        Tcal = decomposicao(Xcal,niveis,nome_wavelet{wavelet});
        Tval = decomposicao(Xval,niveis,nome_wavelet{wavelet});
        Tpred = decomposicao(Xpred,niveis,nome_wavelet{wavelet});

        % Passo 2: Remover os coeficientes wavelet que respondem por apenas 5% da
        % variancia total explicada na calibracao

        v = var(Tcal); % v contem a variancia de cada coluna de T (isto eh, de cada coef. wavelet)
        vtotal = sum(v);
        v_percentual = 100*v/vtotal;

        % Ordenacao da menor variancia para a maior variancia
        [v_sort,index] = sort(v_percentual);

        v_acumulada = cumsum(v_sort);
        corte = max(find(v_acumulada < 1)); % Maior numero de coeficientes para o qual a variancia
        explicada eh menor que 1%

        index_mantidos = index(corte + 1:end);

        Tcal = Tcal(:,index_mantidos);
        Tval = Tval(:,index_mantidos);
        Tpred = Tpred(:,index_mantidos);

        N2 = min(size(Tcal,1),size(Tcal,2)) - 2;

        [coef_selecionados,RMSEV(wavelet,niveis),CORRV(wavelet,niveis),RMSEP(wavelet,niveis),
        CORRP(wavelet,niveis)] = fastselecaolight(Tcal,Ycal,Tval,Yval,Tpred,Ypred,1,N2,0,1);

        Ncoef(wavelet,niveis) = length(coef_selecionados);
```

close all

% Passo 3: Executar o APS e apresentar o resultado em termos de RMSEP

end % fim do for niveis

end % Fim do for wavelet

save resultados RMSEV CORRV RMSEP CORRP nome_wavelet Ncoef

Apêndice I2- Algoritmo decomposicao

```
function T = decomposicao(X,niveis,wname)
```

```
Nobjetos = size(X,1);
```

```
for i = 1:Nobjetos
```

```
    T(i,:) = wavedec(X(i,:),niveis,wname);
```

```
end
```

Apêndice J – Resultados Modelos MLR-WT

Tabela J1. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.

		Nível de Decomposição							Média	Desvio Padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.021306	0.027183	0.024404	0.024851	0.023822	0.024165	0.023212	0.024135	0.0017715	0.021306	0.027183
	Db 2	0.023384	0.023991	0.029078	0.021815	0.021553	0.021736	0.021521	0.023297	0.0027306	0.021521	0.029078
	Db 3	0.024823	0.021396	0.020678	0.023794	0.022802	0.022518	0.022524	0.022648	0.0013873	0.020678	0.024823
	Db 4	0.025673	0.023181	0.022638	0.021897	0.021499	0.024865	0.024778	0.023504	0.0016143	0.021499	0.025673
	Db 5	0.020952	0.021476	0.020843	0.023737	0.021329	0.021661	0.021726	0.021675	0.00096858	0.020843	0.023737
	Db 6	0.026443	0.021968	0.024759	0.019203	0.024143	0.024018	0.022153	0.023241	0.0023482	0.019203	0.026443
	Db 7	0.026910	0.023151	0.022898	0.019222	0.021068	0.023008	0.021743	0.022571	0.0023672	0.019222	0.02691
	Db 8	0.022320	0.028853	0.023968	0.020325	0.023184	0.021128	0.021906	0.023098	0.0028131	0.020325	0.028853
	Db 9	0.025249	0.020314	0.024693	0.023061	0.022100	0.021823	0.022783	0.02286	0.0016955	0.020314	0.025249
	Db 10	0.021823	0.019923	0.022667	0.024261	0.024353	0.024337	0.024400	0.023109	0.0017344	0.019923	0.024400
	Sim 4	0.025094	0.020982	0.019611	0.023424	0.023517	0.024593	0.020284	0.022501	0.0021817	0.019611	0.025094
	Sim 5	0.021844	0.022651	0.022060	0.024136	0.022294	0.023602	0.022392	0.022711	0.00084452	0.021844	0.024136
	Sim 6	0.025952	0.025832	0.022601	0.018462	0.022760	0.027415	0.018789	0.023116	0.0035279	0.018462	0.027415
	Sim 7	0.023878	0.022550	0.022184	0.025006	0.024268	0.023871	0.023871	0.023661	0.00097712	0.022184	0.025006
	Sim 8	0.025909	0.023218	0.024366	0.022437	0.019648	0.023176	0.023968	0.023246	0.0019353	0.019648	0.025909
	Sim 9	0.023682	0.024416	0.019259	0.020837	0.021724	0.023613	0.022861	0.022342	0.0018318	0.019259	0.024416
Sim 10	0.025291	0.020883	0.021899	0.021643	0.020901	0.021008	0.021124	0.021821	0.0015786	0.020883	0.025291	
Valor mínimo		0.020952	0.019923	0.019259	0.018462	0.019648	0.021008	0.018789				
Valor máximo		0.02691	0.028853	0.029078	0.025006	0.024353	0.027415	0.024778				
Média		0.024149	0.023057	0.022859	0.022242	0.02241	0.023326	0.022355				
Desvio Padrão		0.0019149	0.002428	0.0023189	0.0020585	0.0013543	0.0016182	0.0015082				

Tabela J2. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “extrativos” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.

		Nível de Decomposição							Média	Desvio Padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.63272	0.66684	0.60169	0.57413	0.62502	0.62432	0.63679	0.62307	0.029025	0.57413	0.66684
	Db 2	0.64042	0.65192	0.53334	0.64957	0.55391	0.55372	0.55551	0.59120	0.053125	0.53334	0.65192
	Db 3	0.67440	0.62516	0.49221	0.63477	0.61681	0.55826	0.55997	0.59451	0.061118	0.49221	0.6744
	Db 4	0.67809	0.6609	0.52047	0.74546	0.72864	0.7986	0.83552	0.70967	0.103730	0.52047	0.83552
	Db 5	0.67644	0.67104	0.51441	0.69846	0.60653	0.59168	0.54086	0.6142	0.070837	0.51441	0.69846
	Db 6	0.65494	0.59404	0.59797	0.57400	0.67753	0.56449	0.64309	0.61515	0.043312	0.56449	0.67753
	Db 7	0.61770	0.78055	0.58994	0.75301	0.53682	0.56967	0.5723	0.63143	0.095908	0.53682	0.78055
	Db 8	0.61843	0.58905	0.65944	0.76757	0.66866	0.72006	0.63278	0.66514	0.061335	0.58905	0.76757
	Db 9	0.61112	0.62446	0.57003	0.56019	0.75232	0.68349	0.68349	0.64073	0.069205	0.56019	0.75232
	Db 10	0.58098	0.61553	0.61408	0.57315	0.57600	0.73153	0.71758	0.62984	0.067072	0.57315	0.73153
	Sim 4	0.57305	0.88694	0.50183	0.49997	0.67846	0.49909	0.52901	0.59548	0.143720	0.49909	0.88694
	Sim 5	0.64422	0.60416	0.74338	0.62218	0.58757	0.61711	0.60996	0.63265	0.051809	0.58757	0.74338
	Sim 6	0.61588	0.66592	0.61139	0.53074	0.50793	0.53234	0.51042	0.5678	0.062379	0.50793	0.66592
	Sim 7	0.72388	0.71111	0.59181	0.61993	0.78925	0.56642	0.56729	0.65281	0.088171	0.56642	0.78925
	Sim 8	0.6466	0.90709	0.49006	0.76563	0.75849	0.48075	0.48947	0.64830	0.168930	0.48075	0.90709
	Sim 9	0.64593	0.57062	0.52821	0.65385	0.63208	0.62329	0.6251	0.61130	0.045340	0.52821	0.65385
	Sim 10	0.61623	0.68655	0.55215	0.45605	0.49773	0.46822	0.61486	0.55597	0.086834	0.45605	0.68655
Valor mínimo		0.57305	0.57062	0.49006	0.45605	0.49773	0.46822	0.48947				
Valor máximo		0.72388	0.90709	0.74338	0.76757	0.78925	0.7986	0.83552				
Média		0.6383	0.67717	0.57132	0.62816	0.63493	0.5990	0.60729				
Desvio Padrão		0.037275	0.09702	0.066327	0.094516	0.088243	0.091222	0.084703				

Tabela J3. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “holocelulose” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.

		Nível de Decomposição							Média	Desvio Padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	1.5880	1.4292	1.5263	1.534	1.4962	1.4449	1.5085	1.5039	0.054194	1.4292	1.5880
	Db 2	1.2355	0.98785	1.3007	1.1719	1.1662	1.0200	1.0223	1.1292	0.12058	0.98785	1.3007
	Db 3	1.4798	1.5502	1.5002	1.6379	1.1627	1.1542	1.1472	1.3760	0.21299	1.1472	1.6379
	Db 4	1.9483	1.1534	1.323	1.4122	1.448	1.4525	1.4573	1.4564	0.24277	1.1534	1.9483
	Db 5	1.0316	1.024	1.0844	1.3652	2.1843	1.3703	1.373	1.3475	0.40319	1.0240	2.1843
	Db 6	1.0693	1.3781	1.3168	1.4938	1.5569	1.5276	1.5518	1.4135	0.17712	1.0693	1.5569
	Db 7	1.1601	1.3647	1.3888	1.4888	1.1365	1.1469	1.1853	1.2673	0.14325	1.1365	1.4888
	Db 8	1.1177	1.1727	1.3709	1.312	1.3835	1.4839	1.1475	1.2840	0.13957	1.1177	1.4839
	Db 9	1.1235	1.2796	1.4845	1.5185	1.2415	1.3165	1.3422	1.3295	0.13708	1.1235	1.5185
	Db 10	1.1394	1.114	1.5682	1.4156	1.8308	2.1531	1.7334	1.5649	0.37677	1.1140	2.1531
	Sim 4	0.9988	1.225	0.97411	1.2105	1.3278	1.3479	1.4999	1.2263	0.18954	0.97411	1.4999
	Sim 5	1.4695	1.2536	1.2258	1.1467	1.6702	1.6897	1.4954	1.4216	0.21739	1.1467	1.6897
	Sim 6	0.99403	1.2075	1.3583	1.4817	1.3631	1.4873	1.3862	1.3254	0.17347	0.99403	1.4873
	Sim 7	1.2115	1.0247	1.3194	1.6461	1.6803	1.6113	1.6147	1.4440	0.25794	1.0247	1.6803
	Sim 8	0.99289	1.1698	1.5239	1.2431	1.3419	1.3514	1.3614	1.2835	0.16877	0.99289	1.5239
	Sim 9	1.6212	1.4469	1.4574	1.3089	1.5972	1.674	1.3939	1.4999	0.13343	1.3089	1.6740
	Sim 10	0.99334	1.1352	1.4679	1.2775	1.3765	1.4395	1.4432	1.3047	0.18054	0.99334	1.4679
Valor mínimo		0.99289	0.98785	0.97411	1.1467	1.1365	1.0200	1.0223				
Valor máximo		1.9483	1.5502	1.5682	1.6461	2.1843	2.1531	1.7334				
Média		1.2456	1.2304	1.3642	1.3920	1.4684	1.4512	1.392				
Desvio Padrão		0.27803	0.16034	0.15932	0.15428	0.26973	0.2556	0.18328				

Tabela J4. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “lignina” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.

		Nível de Decomposição							Média	Desvio Padrão	Valor mínimo	Valor máximo	
		1	2	3	4	5	6	7					
Família Wavelet	Db 1	1.4846	1.3425	1.0244	0.9992	1.0662	1.0687	1.0681	1.1505	0.18613	0.9992	1.4846	
	Db 2	1.3563	1.3341	1.2755	1.3812	1.5248	1.4221	1.4215	1.3879	0.079175	1.2755	1.5248	
	Db 3	1.1695	1.2411	1.3594	1.4620	1.3351	1.4815	1.4073	1.3508	0.11392	1.1695	1.4815	
	Db 4	1.7765	1.2492	1.1386	1.2641	1.3172	1.2759	1.3599	1.3402	0.20417	1.1386	1.7765	
	Db 5	1.5437	1.4176	1.2023	1.2750	1.3548	1.3272	1.3228	1.3491	0.10863	1.2023	1.5437	
	Db 6	1.6813	1.2301	1.0418	1.3939	1.7149	1.9588	1.7716	1.5418	0.32791	1.0418	1.9588	
	Db 7	1.5637	1.3971	2.7911	1.3025	1.8278	1.4001	1.4025	1.6693	0.52393	1.3025	2.7911	
	Db 8	1.2408	1.5573	1.0500	1.5634	1.3094	1.0924	1.2539	1.2953	0.20279	1.0500	1.5634	
	Db 9	1.3289	1.4227	1.3042	1.6886	1.3648	1.6447	1.5838	1.4768	0.15897	1.3042	1.6886	
	Db 10	1.1234	1.5690	1.3416	1.0433	1.1630	1.2134	1.1840	1.2340	0.17338	1.0433	1.5690	
	Sim 4	1.6273	1.3125	1.3062	1.3240	1.4057	1.1340	1.2646	1.3392	0.15118	1.1340	1.6273	
	Sim 5	1.3406	1.2717	1.3746	1.2760	1.4630	1.2760	1.2781	1.3257	0.072556	1.2717	1.4630	
	Sim 6	1.2690	1.3512	1.1775	1.1854	1.2980	1.2380	1.1363	1.2365	0.07542	1.1363	1.3512	
	Sim 7	1.5359	1.5283	1.1828	1.3591	1.7982	1.0849	1.2300	1.3885	0.24819	1.0849	1.7982	
	Sim 8	1.2067	1.1217	1.0898	1.8998	1.0160	1.1875	1.1412	1.2375	0.29879	1.0160	1.8998	
	Sim 9	1.3323	1.2556	1.9193	1.2782	1.3175	1.132	1.2553	1.3557	0.25684	1.1320	1.9193	
	Sim 10	1.2053	1.6400	1.2615	1.2189	1.1398	1.3649	0.9394	1.2528	0.21485	0.9394	1.6400	
	Valor mínimo		1.1234	1.1217	1.0244	0.99916	1.016	1.0687	0.93945				
	Valor máximo		1.7765	1.6400	2.7911	1.8998	1.8278	1.9588	1.7716				
Média		1.3992	1.3672	1.3436	1.3479	1.3774	1.3119	1.2953					
Desvio Padrão		0.19495	0.14124	0.42583	0.21879	0.23373	0.22874	0.1942					

Tabela J5. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “rendimento” aplicado ao sub-conjunto de amostras de previsão dos dados de madeira de eucalipto na região espectral de 1.100 – 2.500 nm.

		Nível de Decomposição							Média	Desvio Padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.66612	0.61175	0.61751	0.65406	0.64126	0.64423	0.64342	0.63976	0.019212	0.61175	0.66612
	Db 2	0.58191	0.56645	0.57144	0.62919	0.63763	0.60713	0.66112	0.60784	0.03627	0.56645	0.66112
	Db 3	0.66819	0.60376	0.61825	0.58345	0.59685	0.65687	0.55463	0.61171	0.040027	0.55463	0.66819
	Db 4	0.64073	0.61899	0.57340	0.63545	0.64529	0.61422	0.57435	0.61463	0.02999	0.5734	0.64529
	Db 5	0.63005	0.53147	0.71744	0.53114	0.64282	0.65982	0.79900	0.64453	0.095971	0.53114	0.79900
	Db 6	0.60544	0.59827	0.62997	0.63044	0.78028	0.55888	0.64401	0.63533	0.069821	0.55888	0.78028
	Db 7	0.64126	0.62446	0.55476	0.59554	0.58563	0.53775	0.59436	0.59054	0.03618	0.53775	0.64126
	Db 8	0.69042	0.72228	0.55703	0.54047	0.60132	0.56572	0.54505	0.60318	0.073761	0.54047	0.72228
	Db 9	0.68403	0.59712	0.62612	0.60632	0.57364	0.57545	0.58985	0.6075	0.038293	0.57364	0.68403
	Db 10	0.66177	0.63052	0.56267	0.55691	0.65158	0.62153	0.61351	0.61407	0.040657	0.55691	0.66177
	Sim 4	0.68246	0.59904	0.57268	0.58043	0.50987	0.67903	0.56811	0.5988	0.062324	0.50987	0.68246
	Sim 5	0.63024	0.59961	0.65017	0.58773	0.54500	0.73980	0.56175	0.61633	0.065508	0.54500	0.7398
	Sim 6	0.68599	0.66850	0.53493	0.58777	0.60281	0.62483	0.51462	0.60278	0.063679	0.51462	0.68599
	Sim 7	0.60254	0.62171	0.65770	0.74706	0.51403	0.60710	0.66007	0.63003	0.070912	0.51403	0.74706
	Sim 8	0.68801	0.60037	0.57668	0.62110	0.58816	0.80309	0.49868	0.62516	0.096572	0.49868	0.80309
	Sim 9	0.60595	0.60339	0.56429	0.60540	0.62399	0.64286	0.64036	0.61232	0.026865	0.56429	0.64286
	Sim 10	0.68942	0.69887	0.61067	0.60615	0.58209	0.61223	0.5332	0.61895	0.058227	0.53320	0.69887
	Valor mínimo		0.58191	0.53147	0.53493	0.53114	0.50987	0.53775	0.49868			
Valor máximo		0.69042	0.72228	0.71744	0.74706	0.78028	0.80309	0.79900				
Média		0.65027	0.61745	0.59975	0.60580	0.60719	0.63239	0.59977				
Desvio Padrão		0.035911	0.045143	0.047158	0.049171	0.06214	0.065521	0.07200				

Tabela J6. Comparação entre modelos PLS-SG, MLR-WT e MLR-APS. Valores de RMSEP para a estimativa dos parâmetros de qualidade de madeira de eucalipto, para o sub-conjunto de amostras de previsão, na região espectral de 1.100 – 2.500 nm.

Parâmetro de Qualidade	Modelo de Regressão		
	PLS-SG	MLR-WT	MLR-APS
	<i>RMSEP</i>	<i>RMSEP</i>	<i>RMSEP</i>
Densidade (Kg.m ⁻³)	0,016 (13)	0,019	0,016 (29)
Extrativos (% m/m)	0,54 (6)	0,67	0,52 (10)
Holocelulose (% m/m)	1,19 (5)	2,24	1,18 (6)
Lignina (% m/m)	1,39 (13)	1,56	1,07 (14)
Rendimento (%)	0,60 (7)	0,62	0,58 (11)

Tabela J7. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Índice de Cetano” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 –1.100 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	1.0677	1.0617	1.0294	1.1087	1.043	1.019	1.0532	1.0547	0.0294	1.0190	1.1087
	Db 2	1.0494	1.1013	1.0951	1.0872	1.0015	1.0008	1.0887	1.0606	0.0439	1.0008	1.1013
	Db 3	1.0754	1.0542	1.0628	1.0712	1.0638	1.0231	1.0291	1.0542	0.0204	1.0231	1.0754
	Db 4	1.0811	1.0617	1.0998	1.0858	1.0763	1.0121	1.0706	1.0696	0.0281	1.0121	1.0998
	Db 5	1.0861	1.1082	1.0488	1.1971	1.0622	1.0374	1.0009	1.0772	0.0631	1.0009	1.1971
	Db 6	1.0871	1.0923	1.0996	1.1162	1.0485	1.0886	1.0316	1.0806	0.0297	1.0316	1.1162
	Db 7	1.0405	1.0987	1.0384	1.0611	1.0659	1.0193	1.1016	1.0608	0.0310	1.0193	1.1016
	Db 8	1.0429	1.0614	1.0846	1.1651	1.0909	1.0361	1.0774	1.0798	0.0429	1.0361	1.1651
	Db 9	1.059	1.096	1.0547	1.1067	1.0707	1.0182	1.0531	1.0655	0.0295	1.0182	1.1067
	Db 10	1.0623	1.0574	1.0807	1.0634	1.048	1.0689	1.0341	1.0593	0.0150	1.0341	1.0807
	Sim 4	1.0589	1.0862	1.0749	1.0621	1.0844	1.0097	1.0845	1.0658	0.0271	1.0097	1.0862
	Sim 5	1.0971	1.0898	1.0714	1.2161	0.99121	1.0356	1.044	1.0779	0.0708	0.9912	1.2161
	Sim 6	1.0571	1.0861	1.0485	1.0628	1.0863	1.0092	1.0907	1.0630	0.0288	1.0092	1.0907
	Sim 7	1.1022	1.0538	1.0228	1.0415	1.0739	1.0788	1.0872	1.0657	0.0277	1.0228	1.1022
	Sim 8	1.056	1.0854	1.0984	1.0618	1.0869	1.0089	1.072	1.0671	0.0296	1.0089	1.0984
	Sim 9	1.0764	1.0853	1.0746	1.0447	1.0382	1.033	1.0799	1.0617	0.0221	1.0330	1.0853
Sim 10	1.0568	1.0994	1.0923	1.0717	1.087	1.0082	1.0717	1.0696	0.0307	1.0082	1.0994	
Valor mínimo		1.0405	1.0538	1.0228	1.0415	0.9912	1.0008	1.0009				
Valor máximo		1.1022	1.1082	1.0998	1.2161	1.0909	1.0886	1.1016				
Média		1.0680	1.0811	1.0692	1.0955	1.0599	1.0298	1.0630				
Desvio Padrão		0.0182	0.0185	0.0252	0.0517	0.0291	0.0260	0.0275				

Tabela J8. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 – 1.100 nm.

	Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo	
	1	2	3	4	5	6	7					
Família Wavelet	Db 1	0.0016229	0.0015448	0.0014111	0.0013894	0.0011739	0.0010881	0.0010944	0.0013321	0.00021618	0.0010881	0.0016229
	Db 2	0.0015108	0.0015124	0.0017438	0.0015299	0.0013511	0.0011806	0.0010563	0.0014121	0.00023367	0.0010563	0.0017438
	Db 3	0.0015417	0.0014504	0.0015819	0.0013663	0.0014803	0.0010723	0.0010637	0.0013652	0.00021421	0.0010637	0.0015819
	Db 4	0.0015212	0.0016198	0.0016512	0.0015931	0.0013125	0.0012125	0.0012859	0.0014566	0.00018111	0.0012125	0.0016512
	Db 5	0.0015877	0.0015508	0.001624	0.0017391	0.0013333	0.0013201	0.0010487	0.0014577	0.00023597	0.0010487	0.0017391
	Db 6	0.0015283	0.0016731	0.0013668	0.001558	0.0013246	0.00098892	0.00099343	0.0013476	0.00027014	0.00098892	0.0016731
	Db 7	0.0014804	0.0017503	0.0015234	0.0015718	0.0012566	0.0010038	0.00086594	0.0013503	0.0003213	0.00086594	0.0017503
	Db 8	0.0016129	0.0014262	0.0016652	0.0013787	0.0013054	0.0010449	0.000999	0.0013475	0.00025604	0.0009990	0.0016652
	Db 9	0.0015967	0.0014019	0.001434	0.001281	0.0010497	0.0011301	0.0010667	0.0012800	0.00020812	0.0010497	0.0015967
	Db 10	0.0016399	0.0016933	0.0015641	0.0016884	0.0011037	0.0010834	0.0010627	0.0014051	0.00030422	0.0010627	0.0016933
	Sim 4	0.0016113	0.0014887	0.0014939	0.0015929	0.0013032	0.0012842	0.0012768	0.0014359	0.00014577	0.0012768	0.0016113
	Sim 5	0.0017135	0.0016196	0.0014465	0.0015375	0.0011931	0.0010392	0.0010647	0.0013734	0.00027327	0.0010392	0.0017135
	Sim 6	0.0016135	0.0016484	0.0016926	0.001479	0.001276	0.0011207	0.0012188	0.0014356	0.00022967	0.0011207	0.0016926
	Sim 7	0.0015106	0.0018137	0.0015167	0.0013793	0.0012554	0.0010203	0.0010394	0.0013622	0.00028359	0.0010203	0.0018137
	Sim 8	0.0014931	0.001645	0.0014719	0.0014039	0.0012156	0.0011063	0.0012429	0.0013684	0.00018789	0.0011063	0.001645
	Sim 9	0.001571	0.0018653	0.0016529	0.0015154	0.0012905	0.0011526	0.0010587	0.0014438	0.00028847	0.0010587	0.0018653
Sim 10	0.0014936	0.0015163	0.0015005	0.0016378	0.0011957	0.0010972	0.00097952	0.0013458	0.0002512	0.00097952	0.0016378	
Valor mínimo	0.0014804	0.0014019	0.0013668	0.001281	0.0010497	0.00098892	0.00086594					
Valor máximo	0.0017135	0.0018653	0.0017438	0.0017391	0.0014803	0.0013201	0.0012859					
Média	0.0015676	0.0016012	0.0015494	0.0015083	0.0012600	0.0011144	0.0010834					
Desvio Padrão	0.0000644	0.0001328	0.0001088	0.0001265	0.0001003	0.0000929	0.0001122					

Tabela J9. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Enxofre” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 –1.100 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.031885	0.029448	0.032108	0.030736	0.031725	0.030753	0.02664	0.030471	0.0019203	0.02664	0.032108
	Db 2	0.029741	0.028056	0.030915	0.03154	0.034132	0.027939	0.027091	0.029916	0.0024762	0.027091	0.034132
	Db 3	0.035679	0.030624	0.030114	0.031539	0.026725	0.034284	0.024257	0.03046	0.0039916	0.024257	0.035679
	Db 4	0.031229	0.035201	0.031237	0.031031	0.030043	0.028008	0.024326	0.030154	0.0033455	0.024326	0.035201
	Db 5	0.034204	0.030351	0.029773	0.028256	0.030157	0.030329	0.030788	0.030551	0.0018036	0.028256	0.034204
	Db 6	0.031965	0.035648	0.028294	0.026912	0.027651	0.02617	0.02861	0.029321	0.0033442	0.02617	0.035648
	Db 7	0.031324	0.028024	0.035684	0.032712	0.028416	0.031133	0.024559	0.030265	0.0036108	0.024559	0.035684
	Db 8	0.030774	0.029787	0.032949	0.030956	0.027416	0.027721	0.023966	0.029081	0.0029616	0.023966	0.032949
	Db 9	0.030736	0.029197	0.031089	0.033325	0.029257	0.0308	0.029566	0.030567	0.0014467	0.029197	0.033325
	Db 10	0.034149	0.02894	0.028757	0.029358	0.027417	0.028969	0.032831	0.03006	0.0024501	0.027417	0.034149
	Sim 4	0.033102	0.032727	0.031829	0.027379	0.032562	0.028131	0.027488	0.03046	0.0026508	0.027379	0.033102
	Sim 5	0.035569	0.030668	0.033527	0.028803	0.032344	0.033395	0.025583	0.031413	0.0033638	0.025583	0.035569
	Sim 6	0.030995	0.036103	0.025505	0.033927	0.027257	0.026397	0.023496	0.029097	0.004672	0.023496	0.036103
	Sim 7	0.031751	0.031618	0.034903	0.028807	0.027289	0.031519	0.026673	0.030366	0.0029137	0.026673	0.034903
	Sim 8	0.030508	0.032229	0.031893	0.036438	0.032627	0.028555	0.022763	0.030716	0.0042433	0.022763	0.036438
	Sim 9	0.030298	0.031343	0.028918	0.028555	0.033396	0.033215	0.024789	0.030073	0.0030038	0.024789	0.033396
Sim 10	0.033768	0.034554	0.03463	0.034384	0.027743	0.026394	0.022981	0.030636	0.0048326	0.022981	0.03463	
Valor mínimo		0.029741	0.028024	0.025505	0.026912	0.026725	0.02617	0.022763				
Valor máximo		0.035679	0.036103	0.035684	0.036438	0.034132	0.034284	0.032831				
Média		0.032216	0.031442	0.031302	0.030862	0.029774	0.02963	0.026259				
Desvio Padrão		0.00185	0.002612	0.0026332	0.0026685	0.002529	0.0025434	0.00287				

Tabela J10. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 50” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 –1.100 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	5.6058	5.3145	5.6309	4.936	5.4505	5.8876	4.9398	5.3950	0.35828	4.9360	5.8876
	Db 2	5.5972	5.6515	5.6131	5.5231	5.1921	5.4979	4.3657	5.3487	0.4595	4.3657	5.6515
	Db 3	5.5888	5.554	5.5522	5.4986	5.6782	4.5207	4.2996	5.2417	0.5742	4.2996	5.6782
	Db 4	5.5751	5.5869	5.6757	5.5128	5.257	4.5553	4.1728	5.1908	0.5899	4.1728	5.6757
	Db 5	5.5679	5.6615	5.5218	5.4526	5.31	5.4984	4.2177	5.3186	0.4972	4.2177	5.6615
	Db 6	5.5595	5.5701	5.5275	5.5142	5.4168	5.1723	5.0656	5.4037	0.2031	5.0656	5.5701
	Db 7	5.5661	5.3418	5.548	5.5084	5.629	4.0875	4.5334	5.1735	0.6098	4.0875	5.6290
	Db 8	5.5575	5.5222	5.5829	5.4427	5.1599	4.4647	4.3478	5.1540	0.5309	4.3478	5.5829
	Db 9	5.6379	5.5816	5.6283	5.5073	5.2313	5.3441	4.6947	5.3750	0.3361	4.6947	5.6379
	Db 10	5.6155	5.4788	5.5361	5.5208	5.6318	4.9213	4.966	5.3815	0.3040	4.9213	5.6318
	Sim 4	5.5408	5.5144	5.5302	5.6025	5.3919	4.3243	4.1966	5.1572	0.6169	4.1966	5.6025
	Sim 5	5.5833	5.5057	5.5106	5.6211	5.6358	4.6475	4.4813	5.2836	0.4961	4.4813	5.6358
	Sim 6	5.7191	5.6658	5.655	5.5825	5.4084	4.4606	4.1812	5.2389	0.6400	4.1812	5.7191
	Sim 7	5.3215	5.6649	5.5111	5.3443	5.2952	5.2006	4.6529	5.2844	0.3181	4.6529	5.6649
	Sim 8	5.7226	5.4263	5.6593	5.5539	5.4167	4.4672	4.1752	5.2030	0.6184	4.1752	5.7226
	Sim 9	5.5965	5.5151	5.3482	5.6227	5.4351	4.5597	4.4696	5.2210	0.4921	4.4696	5.6227
Sim 10	5.5527	5.6807	5.3607	5.5267	5.4157	4.4704	4.2605	5.1811	0.5697	4.2605	5.6807	
Valor mínimo		5.3215	5.3145	5.3482	4.936	5.1599	4.0875	4.1728				
Valor máximo		5.7226	5.6807	5.6757	5.6227	5.6782	5.8876	5.0656				
Média		5.5828	5.5433	5.5524	5.4865	5.4091	4.8282	4.4718				
Desvio Padrão		0.085095	0.1098	0.092849	0.15749	0.16033	0.50952	0.29637				

Tabela J11. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 85” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 850 –1.100 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	7.9116	7.7388	8.6558	8.0534	8.0258	9.1982	9.0429	8.3752	0.58416	7.7388	9.1982
	Db 2	7.9001	7.8496	8.3759	7.4487	8.4582	8.1055	8.2003	8.0483	0.34705	7.4487	8.4582
	Db 3	7.9692	7.8339	7.7705	8.3887	8.5061	8.0080	8.5030	8.1399	0.31740	7.7705	8.5061
	Db 4	7.8054	7.7481	8.0788	7.4724	8.5798	7.6913	7.9150	7.8987	0.35420	7.4724	8.5798
	Db 5	7.9394	7.8202	7.8698	7.5521	8.2609	8.9608	7.8680	8.0387	0.45674	7.5521	8.9608
	Db 6	7.8926	7.7535	7.5781	8.2747	7.615	8.2637	7.8464	7.8891	0.28316	7.5781	8.2747
	Db 7	7.8883	7.7521	8.3013	7.5059	8.1977	8.0072	8.5354	8.0268	0.34871	7.5059	8.5354
	Db 8	7.8581	7.8010	7.5116	7.5373	8.7369	7.9728	8.1878	7.9436	0.42213	7.5116	8.7369
	Db 9	7.8313	7.7300	7.8234	8.0137	8.5833	8.2148	7.4847	7.9545	0.35827	7.4847	8.5833
	Db 10	7.8019	8.3657	7.798	7.6846	7.4914	8.1453	7.5034	7.8272	0.32419	7.4914	8.3657
	Sim 4	7.8314	7.8366	7.7485	8.3048	9.0772	7.7262	7.7002	8.0321	0.50485	7.7002	9.0772
	Sim 5	7.6661	7.6921	7.9547	8.4767	7.5344	8.3729	8.3628	8.0085	0.39212	7.5344	8.4767
	Sim 6	7.8429	7.8858	7.6628	7.6946	8.0822	8.1778	8.2774	7.9462	0.23797	7.6628	8.2774
	Sim 7	7.9637	7.6553	7.9003	8.0749	8.7939	8.2856	8.5641	8.1768	0.39725	7.6553	8.7939
	Sim 8	7.8514	7.7590	8.0874	7.5380	8.6179	8.087	8.3328	8.0391	0.36302	7.5380	8.6179
	Sim 9	8.0868	7.7902	7.6097	7.9274	8.1976	7.8638	8.4007	7.9823	0.26618	7.6097	8.4007
	Sim 10	7.8556	7.7963	7.6929	7.4996	8.5300	8.0744	8.3635	7.9732	0.36982	7.4996	8.5300
Valor mínimo	7.6661	7.6553	7.5116	7.4487	7.4914	7.6913	7.4847					
Valor máximo	8.0868	8.3657	8.6558	8.4767	9.0772	9.1982	9.0429					
Média	7.8762	7.8122	7.907	7.8499	8.3111	8.1856	8.1817					
Desvio Padrão	0.089461	0.15418	0.3085	0.36011	0.45216	0.38536	0.41247					

Tabela J12. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Índice de Cetano” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.91394	0.91313	0.94586	0.97496	0.88729	0.95951	0.86588	0.9229	0.0393	0.86588	0.97496
	Db 2	0.91501	0.93102	0.94438	0.91637	0.95239	0.88687	0.92072	0.9238	0.0216	0.88687	0.95239
	Db 3	0.91632	0.95315	0.94774	0.91415	0.93582	1.0016	0.96761	0.9481	0.0305	0.91415	1.00160
	Db 4	0.92877	0.93454	0.88469	0.93141	0.94929	0.92581	0.92872	0.9262	0.0199	0.88469	0.94929
	Db 5	0.92933	0.90932	0.95788	0.92185	0.94187	0.88423	0.92349	0.9240	0.0234	0.88423	0.95788
	Db 6	0.94363	0.96456	0.95047	0.93156	0.96344	0.91448	0.93732	0.9436	0.0179	0.91448	0.96456
	Db 7	0.92425	0.93326	0.92829	0.9361	0.94517	0.99329	0.95500	0.9451	0.0237	0.92425	0.99329
	Db 8	0.92644	0.90972	0.94979	0.96199	0.95082	0.89121	0.9204	0.9301	0.0254	0.89121	0.96199
	Db 9	0.91871	0.92769	0.95132	0.90954	0.96591	0.91709	0.94659	0.9338	0.0210	0.90954	0.96591
	Db 10	0.91914	0.92791	0.9482	0.95265	0.96161	0.89057	0.93032	0.9329	0.0240	0.89057	0.96161
	Sim 4	0.91807	0.95423	0.93854	0.96945	0.9348	0.88776	0.9868	0.9414	0.0329	0.88776	0.9868
	Sim 5	0.92761	0.94035	0.93828	0.95876	0.97094	0.97575	0.96301	0.9535	0.0182	0.92761	0.97575
	Sim 6	0.93539	0.94408	0.94341	0.96665	0.92832	0.90661	0.91868	0.9347	0.0195	0.90661	0.96665
	Sim 7	0.9274	0.91375	0.94639	0.94194	0.9465	0.95033	0.96656	0.9418	0.0169	0.91375	0.96656
	Sim 8	0.91758	0.9404	0.9467	0.98565	0.92965	0.90687	0.95434	0.9402	0.0260	0.90687	0.98565
	Sim 9	0.91518	0.93173	0.94379	0.9334	0.93987	0.9736	0.94077	0.9398	0.0177	0.91518	0.9736
	Sim 10	0.91805	0.95672	0.93068	0.93295	0.94928	1.0008	0.95442	0.9490	0.0268	0.91805	1.0008
	Valor mínimo		0.91394	0.90932	0.88469	0.90954	0.88729	0.88423	0.86588			
Valor máximo		0.94363	0.96456	0.95788	0.98565	0.97094	1.0016	0.9868				
Média		0.9232	0.9344	0.9410	0.9435	0.9443	0.9333	0.9400				
Desvio Padrão		0.0082	0.0168	0.0162	0.0229	0.0192	0.0430	0.0275				

Tabela J13. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.00061313	0.00061309	0.00066293	0.00061941	0.00059759	0.0005411	0.00057565	0.00060327	0.000038021	0.00054110	0.00066293
	Db 2	0.00060795	0.00062357	0.00064508	0.00069253	0.00055925	0.00054086	0.00056623	0.00060507	0.000053746	0.00054086	0.00069253
	Db 3	0.00062436	0.00061731	0.00065758	0.00066000	0.00054899	0.00058539	0.00061892	0.00061608	0.000039099	0.00054899	0.00066000
	Db 4	0.00062392	0.00065498	0.00050615	0.00058786	0.00067829	0.00057907	0.00054132	0.00059594	0.000061163	0.00050615	0.00067829
	Db 5	0.0006459	0.00064922	0.00063214	0.00060308	0.0006159	0.00055631	0.00059895	0.00061450	0.000032308	0.00055631	0.00064922
	Db 6	0.00064203	0.00061559	0.00067760	0.00066301	0.00049113	0.00058856	0.00058645	0.00060920	0.000062685	0.00049113	0.0006776
	Db 7	0.00063791	0.00063784	0.00063851	0.00064774	0.00062042	0.00055789	0.00056426	0.00061494	0.00003772	0.00055789	0.00064774
	Db 8	0.00064057	0.00064559	0.00065792	0.00063822	0.00060675	0.00057587	0.00054323	0.00061545	0.000042268	0.00054323	0.00065792
	Db 9	0.0006434	0.00064825	0.0006598	0.0006694	0.00058414	0.00059451	0.00058654	0.00062658	0.000036793	0.00058414	0.0006694
	Db 10	0.00064762	0.00062199	0.00065383	0.0006182	0.00049792	0.00056506	0.0005333	0.00059113	0.000059888	0.00049792	0.00065383
	Sim 4	0.00064336	0.00064973	0.00066137	0.00065073	0.00058444	0.00052264	0.00051612	0.00060406	0.000063035	0.00051612	0.00066137
	Sim 5	0.00061167	0.00065362	0.0005092	0.00060735	0.0005482	0.00058405	0.00056934	0.00058335	0.000046987	0.00050920	0.00065362
	Sim 6	0.00064135	0.00062151	0.00061254	0.00057912	0.00060626	0.0005767	0.00055474	0.00059889	0.000029966	0.00055474	0.00064135
	Sim 7	0.00064117	0.00063497	0.00071394	0.00060775	0.00060367	0.00055225	0.00053446	0.00061260	0.000059774	0.00053446	0.00071394
	Sim 8	0.00064214	0.00064116	0.0005343	0.00064346	0.00060493	0.0005245	0.00050815	0.00058552	0.000061058	0.00050815	0.00064346
	Sim 9	0.00060609	0.00065152	0.00053503	0.00059702	0.00057472	0.0005455	0.00060135	0.00058732	0.000039595	0.00053503	0.00065152
	Sim 10	0.0006413	0.00067498	0.00064738	0.00062793	0.00060306	0.00052292	0.00050152	0.00060273	0.000065764	0.00050152	0.00067498
	Valor mínimo	0.00060609	0.00061309	0.00050615	0.00057912	0.00049113	0.00052264	0.00050152				
	Valor máximo	0.00064762	0.00067498	0.00071394	0.00069253	0.00067829	0.00059451	0.00061892				
	Média	0.00063258	0.00063852	0.00062384	0.00063017	0.00058386	0.0005596	0.00055885				
Desvio Padrão	0.00001458	0.00001734	0.00006263	0.00003151	0.00004532	0.00002397	0.00003403					

Tabela J14. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Enxofre” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.016914	0.016503	0.01944	0.016137	0.016465	0.016781	0.016082	0.016903	0.0011593	0.016082	0.01944
	Db 2	0.016863	0.017194	0.016592	0.01685	0.016816	0.016899	0.016218	0.016776	0.00030269	0.016218	0.017194
	Db 3	0.017189	0.015925	0.016187	0.016485	0.017197	0.01644	0.016491	0.016559	0.00047759	0.015925	0.017197
	Db 4	0.016873	0.016565	0.016794	0.016165	0.016703	0.015537	0.016236	0.01641	0.00046977	0.015537	0.016873
	Db 5	0.015981	0.015973	0.016081	0.01716	0.016434	0.016796	0.015073	0.016214	0.00067255	0.015073	0.017160
	Db 6	0.015999	0.015891	0.016836	0.016721	0.016729	0.01683	0.015838	0.016406	0.00046935	0.015838	0.016836
	Db 7	0.016009	0.016573	0.016634	0.016195	0.017518	0.016481	0.017018	0.016633	0.00050652	0.016009	0.017518
	Db 8	0.015917	0.015955	0.016799	0.016329	0.016427	0.015604	0.016345	0.016197	0.00039649	0.015604	0.016799
	Db 9	0.016596	0.015864	0.016534	0.016967	0.017484	0.015943	0.017659	0.016721	0.00069676	0.015864	0.017659
	Db 10	0.016595	0.017188	0.017916	0.016367	0.017969	0.017456	0.016285	0.017111	0.00070843	0.016285	0.017969
	Sim 4	0.01688	0.015861	0.016905	0.01638	0.016287	0.017288	0.016229	0.016547	0.00049250	0.015861	0.017288
	Sim 5	0.016832	0.017227	0.017174	0.016727	0.016827	0.016919	0.017743	0.017064	0.00035221	0.016727	0.017743
	Sim 6	0.016878	0.015982	0.015967	0.016352	0.017923	0.016113	0.016523	0.016534	0.00069334	0.015967	0.017923
	Sim 7	0.015962	0.016027	0.017172	0.017116	0.018103	0.017880	0.015969	0.01689	0.00091618	0.015962	0.018103
	Sim 8	0.016848	0.016866	0.017206	0.016149	0.017972	0.016781	0.017005	0.017005	0.00055475	0.016149	0.017972
	Sim 9	0.016843	0.015864	0.016377	0.016442	0.016071	0.017700	0.015508	0.016401	0.00071644	0.015508	0.017700
	Sim 10	0.015924	0.015975	0.01661	0.016191	0.017984	0.016892	0.017336	0.016702	0.00076212	0.015924	0.017984
Valor mínimo		0.015917	0.015861	0.015967	0.016137	0.016071	0.015537	0.015073				
Valor máximo		0.017189	0.017227	0.01944	0.01716	0.018103	0.017880	0.017743				
Média		0.016535	0.01632	0.016896	0.016514	0.017112	0.016726	0.016457				
Desvio Padrão		0.0004524	0.0005204	0.0008087	0.0003442	0.0006973	0.0006637	0.0007308				

Tabela J15. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 50” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	4.0078	3.8165	3.6075	3.9784	3.9211	3.3427	3.4177	3.7274	0.2724	3.3427	4.0078
	Db 2	3.5332	3.4352	3.7618	3.436	3.7472	3.0113	3.3708	3.4708	0.2545	3.0113	3.7618
	Db 3	3.5302	3.4606	3.5076	3.7936	2.912	3.6677	2.8362	3.3868	0.3682	2.8362	3.7936
	Db 4	3.5847	3.5257	4.2521	3.2992	3.4183	3.2809	3.1501	3.5016	0.3631	3.1501	4.2521
	Db 5	3.5863	3.5705	3.6993	3.2706	3.32	3.2345	3.6393	3.4744	0.1926	3.2345	3.6993
	Db 6	3.5464	3.823	3.735	3.2074	2.8687	2.9901	3.454	3.3749	0.3648	2.8687	3.8230
	Db 7	3.6485	3.5872	3.5235	3.5495	2.9173	3.5734	3.1546	3.4220	0.2752	2.9173	3.6485
	Db 8	3.5548	3.551	3.6173	3.2911	3.3356	3.1956	3.1695	3.3878	0.1843	3.1695	3.6173
	Db 9	3.6535	3.5261	3.959	2.8934	3.5561	3.2161	3.313	3.4453	0.3417	2.8934	3.9590
	Db 10	3.5815	3.6535	3.7241	3.5214	2.8563	3.0248	3.1082	3.3528	0.3471	2.8563	3.7241
	Sim 4	3.7331	3.5686	3.6868	3.5994	3.0063	3.0286	3.4664	3.4413	0.3019	3.0063	3.7331
	Sim 5	3.5847	3.4241	3.1602	3.6512	3.052	3.8381	3.1707	3.4116	0.2944	3.0520	3.8381
	Sim 6	3.7513	3.3876	3.5946	3.6565	3.0099	3.0628	3.2167	3.3828	0.2957	3.0099	3.7513
	Sim 7	3.633	3.4142	3.5157	3.167	2.9168	3.6174	3.4374	3.3859	0.2591	2.9168	3.6330
	Sim 8	3.6147	3.808	3.5748	3.5653	3.0073	3.3491	3.0458	3.4236	0.3024	3.0073	3.8080
	Sim 9	3.6085	3.376	3.7001	3.117	3.3551	3.1817	2.861	3.3142	0.2895	2.8610	3.7001
Sim 10	3.5627	3.4777	3.7006	3.2976	3.0004	3.3251	3.2312	3.3708	0.2314	3.0004	3.7006	
Valor mínimo	3.5302	3.3760	3.1602	2.8934	2.8563	2.9901	2.8362					
Valor máximo	4.0078	3.8230	4.2521	3.9784	3.9211	3.8381	3.6393					
Média	3.6303	3.5533	3.6659	3.4291	3.1883	3.2906	3.2378					
Desvio Padrão	0.1158	0.1465	0.2227	0.2723	0.3252	0.2531	0.2146					

Tabela J16. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 85” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.100 –1.570 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	6.7882	6.8825	6.8302	6.0992	6.6176	6.12	6.5511	6.5555	0.3262	6.0992	6.8825
	Db 2	6.7477	6.5241	6.2044	5.7632	5.7481	5.4638	5.5259	5.9967	0.5012	5.4638	6.7477
	Db 3	6.5005	6.2017	6.0933	5.5219	5.8168	6.0854	5.3554	5.9393	0.3998	5.3554	6.5005
	Db 4	6.3424	7.0681	6.2788	6.82	5.5432	5.9223	6.3588	6.3334	0.5121	5.5432	7.0681
	Db 5	6.3854	6.5152	6.4845	6.2975	6.2297	6.3436	6.0798	6.3337	0.1500	6.0798	6.5152
	Db 6	6.3078	6.6703	6.9709	6.0342	5.9031	5.5216	5.9981	6.2009	0.4905	5.5216	6.9709
	Db 7	6.3756	6.393	6.3778	6.4827	5.6984	5.6807	5.754	6.1089	0.3745	5.6807	6.4827
	Db 8	6.7257	6.499	6.0988	6.5151	5.7403	6.3237	6.2176	6.3029	0.3236	5.7403	6.7257
	Db 9	6.1367	6.2445	6.4253	6.3459	6.3355	6.3976	5.9167	6.2575	0.1792	5.9167	6.4253
	Db 10	6.0975	6.1314	6.0471	5.8041	5.8979	5.6339	5.7369	5.9070	0.1917	5.6339	6.1314
	Sim 4	6.3076	6.215	6.8687	6.3848	5.5197	5.6491	6.2891	6.1763	0.4589	5.5197	6.8687
	Sim 5	6.0967	7.1447	7.0319	6.2866	5.5118	5.7209	6.2177	6.2872	0.6128	5.5118	7.1447
	Sim 6	6.2906	6.5094	6.1422	6.2882	6.1602	6.0583	6.2337	6.2404	0.1452	6.0583	6.5094
	Sim 7	6.2478	6.7176	7.0047	5.7454	5.9333	6.4876	5.9841	6.3029	0.4566	5.7454	7.0047
	Sim 8	6.3124	6.324	6.2686	6.4525	6.1634	5.7566	5.3477	6.0893	0.3948	5.3477	6.4525
	Sim 9	6.6874	6.4563	6.2642	6.1787	5.7374	5.986	5.0690	6.0541	0.5320	5.0690	6.6874
Sim 10	6.7589	6.8645	6.4376	6.3267	6.173	5.7939	5.4263	6.2544	0.5115	5.4263	6.8645	
Valor mínimo		6.0967	6.1314	6.0471	5.5219	5.5118	5.4638	5.0690				
Valor máximo		6.7882	7.1447	7.0319	6.8200	6.6176	6.4876	6.5511				
Média		6.4182	6.5507	6.4605	6.1969	5.9253	5.9379	5.8860				
Desvio Padrão		0.2385	0.3028	0.3453	0.3331	0.3123	0.3201	0.4220				

Tabela J17. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Índice de Cetano” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.

		Nível de Decomposição							Média	Desvio Padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	1.1234	1.1551	1.1493	1.0755	0.99973	0.95733	0.95309	1.0591	0.088389	0.95309	1.1551
	Db 2	1.1172	1.1332	1.0874	1.0369	1.021	0.98728	1.0136	1.0567	0.055974	0.98728	1.1332
	Db 3	1.1267	1.1681	1.0831	0.99848	1.0309	1.0324	1.0333	1.0676	0.061065	0.99848	1.1681
	Db 4	1.0721	1.1902	1.2348	0.95649	0.98171	1.0326	1.0579	1.0751	0.103000	0.95649	1.2348
	Db 5	1.1235	1.0764	1.1007	1.1228	1.0655	1.0664	0.99555	1.0787	0.044122	0.99555	1.1235
	Db 6	1.1177	1.0702	1.1454	0.99555	1.0510	0.9969	0.9811	1.0511	0.064050	0.98109	1.1454
	Db 7	1.1455	1.2230	1.1132	0.94863	0.97657	1.0008	0.99351	1.0573	0.10324	0.94863	1.2230
	Db 8	1.1149	1.1717	1.1425	1.0237	1.2980	1.2424	0.97722	1.1386	0.11336	0.97722	1.2980
	Db 9	1.1577	1.0741	1.1827	0.9925	1.0564	0.99161	1.0005	1.0651	0.078958	0.99161	1.1827
	Db 10	1.1315	1.2489	1.0941	0.95214	1.0324	1.0205	0.95693	1.0624	0.10528	0.95214	1.2489
	Sim 4	1.0998	1.1485	1.1055	1.0023	0.99492	0.98104	0.99411	1.0466	0.068762	0.98104	1.1485
	Sim 5	1.1148	1.1506	1.0383	1.0659	1.0975	1.0811	1.0032	1.0788	0.048868	1.00320	1.1506
	Sim 6	1.1129	1.0855	1.0843	0.99565	0.97884	1.0222	0.96814	1.0354	0.058288	0.96814	1.1129
	Sim 7	1.1297	1.1431	1.1152	1.0128	1.0018	0.96216	0.96847	1.0476	0.078847	0.96216	1.1431
Sim 8	1.0935	1.1432	1.0901	1.0696	1.0232	0.97287	0.96794	1.0515	0.065773	0.96794	1.1432	
Sim 9	1.0871	1.1872	1.1274	0.97507	1.2883	1.0387	1.2892	1.1419	0.12033	0.97507	1.2892	
Sim 10	1.1136	1.1932	1.1668	1.0626	1.0194	1.0651	0.98569	1.0866	0.075566	0.98569	1.1932	
Valor mínimo		1.0721	1.0702	1.0383	0.94863	0.97657	0.95733	0.95309				
Valor máximo		1.1577	1.2489	1.2348	1.1228	1.298	1.2424	1.2892				
Média		1.1166	1.1507	1.1212	1.0169	1.0539	1.0266	1.0082				
Desvio Padrão		0.020634	0.051769	0.046052	0.049172	0.095772	0.066659	0.077234				

Tabela J18. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Densidade” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	0.000586	0.0006588	0.0006833	0.0007042	0.0006455	0.0006921	0.0006103	0.0006544	0.00004358	0.0005865	0.0007042
	Db 2	0.0005833	0.0008021	0.0008820	0.0007487	0.0007272	0.0007620	0.0006819	0.0007411	0.00009369	0.00058338	0.0008820
	Db 3	0.0005859	0.0006841	0.0006834	0.0006134	0.0007279	0.0007441	0.0007859	0.0006893	0.00007112	0.00058595	0.0007860
	Db 4	0.0005790	0.0006315	0.0006849	0.0008280	0.0007334	0.0006531	0.0006918	0.0006860	0.00007953	0.00057906	0.0008280
	Db 5	0.0006154	0.0006913	0.0006988	0.0006697	0.0006693	0.0007034	0.0006902	0.0006769	0.00003015	0.00061548	0.0007034
	Db 6	0.0005950	0.0006797	0.0006410	0.0006865	0.0007766	0.0006430	0.0007084	0.0006758	0.00005795	0.00059508	0.0007770
	Db 7	0.0006183	0.0006356	0.0006482	0.0006253	0.0007883	0.0006644	0.0007042	0.0006692	0.00005989	0.00061838	0.0007884
	Db 8	0.0005942	0.0006373	0.0006645	0.0007123	0.0006917	0.0007028	0.0006814	0.0006692	0.00004146	0.00059422	0.0007124
	Db 9	0.0006856	0.0005445	0.0007463	0.0007578	0.0006456	0.0007121	0.0006559	0.0006783	0.00007260	0.00054455	0.0007579
	Db 10	0.0005842	0.0006337	0.0007760	0.0006262	0.0006451	0.0007847	0.0006554	0.0006722	0.00007723	0.00058422	0.0007847
	Sim 4	0.0005721	0.0005506	0.0008400	0.0007712	0.0008545	0.0006851	0.0007187	0.0007132	0.00012011	0.00055063	0.0008540
	Sim 5	0.0006206	0.0007035	0.0007922	0.0007218	0.0006992	0.0006891	0.0007384	0.0007093	0.00005210	0.00062068	0.0007923
	Sim 6	0.0006224	0.0006871	0.0006647	0.0006525	0.0006811	0.0006814	0.0007056	0.0006708	0.00002711	0.00062249	0.0007057
	Sim 7	0.0006000	0.0007198	0.0007899	0.0006750	0.0006630	0.0006936	0.0006328	0.0006821	0.00006160	0.00060004	0.0007899
	Sim 8	0.0006211	0.0006449	0.0006077	0.0006555	0.0007527	0.0006595	0.0006757	0.0006596	0.00004714	0.00060774	0.0007527
	Sim 9	0.0005865	0.0006892	0.0007779	0.0006576	0.0007067	0.0007623	0.0007733	0.0007077	0.00007044	0.00058651	0.0007779
	Sim 10	0.0006004	0.0006817	0.0006378	0.0007006	0.0007281	0.0006788	0.0006928	0.0006743	0.00004244	0.0006004	0.0007281
Valor mínimo		0.0005721	0.0005445	0.0006077	0.0006134	0.0006451	0.0006430	0.0006103				
Valor máximo		0.0006856	0.0008021	0.0008820	0.0008280	0.0008545	0.0007847	0.0007859				
Média		0.0006030	0.0006633	0.0007188	0.0006945	0.0007139	0.0007007	0.0006943				
Desvio Padrão		0.0000268	0.0000600	0.0000787	0.0000578	0.0000571	0.0000407	0.0000448				

Tabela J19. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “Enxofre” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.

	Nível de Decomposição							Média	Desvio Padrão	Valor mínimo	Valor máximo	
	1	2	3	4	5	6	7					
Família Wavelet	Db 1	0.023521	0.022319	0.023619	0.021708	0.023185	0.022694	0.021603	0.022664	0.00082437	0.021603	0.023619
	Db 2	0.024452	0.025142	0.022467	0.026077	0.024269	0.021245	0.022563	0.023745	0.0017051	0.021245	0.026077
	Db 3	0.02164	0.020624	0.026646	0.024044	0.024437	0.023904	0.022931	0.023461	0.0019697	0.020624	0.026646
	Db 4	0.021803	0.020971	0.024595	0.022393	0.023433	0.023676	0.023672	0.022935	0.0012602	0.020971	0.024595
	Db 5	0.022074	0.021086	0.020712	0.025212	0.022955	0.020033	0.018928	0.021571	0.0020716	0.018928	0.025212
	Db 6	0.022233	0.024523	0.025355	0.024681	0.020379	0.021912	0.021694	0.022968	0.0018722	0.020379	0.025355
	Db 7	0.022117	0.021151	0.021713	0.020702	0.023781	0.024926	0.021327	0.022245	0.001542	0.020702	0.024926
	Db 8	0.021596	0.021816	0.024992	0.02346	0.023934	0.021586	0.022120	0.022786	0.0013465	0.021586	0.024992
	Db 9	0.020567	0.024514	0.021185	0.022448	0.020123	0.021472	0.022775	0.021869	0.0015012	0.020123	0.024514
	Db 10	0.021793	0.023579	0.021401	0.025651	0.022569	0.023501	0.022561	0.023008	0.0014146	0.021401	0.025651
	Sim 4	0.020913	0.022249	0.021934	0.024478	0.021684	0.020710	0.021934	0.021986	0.0012353	0.02071	0.024478
	Sim 5	0.02002	0.023992	0.021739	0.021724	0.020853	0.023235	0.022191	0.021965	0.0013479	0.02002	0.023992
	Sim 6	0.020421	0.025365	0.022071	0.022100	0.021799	0.023585	0.023299	0.022663	0.0015809	0.020421	0.025365
	Sim 7	0.021279	0.023383	0.020494	0.020877	0.023124	0.024123	0.025161	0.022634	0.0017751	0.020494	0.025161
	Sim 8	0.021606	0.021221	0.025635	0.020502	0.021980	0.023148	0.023764	0.022551	0.0017571	0.020502	0.025635
	Sim 9	0.02167	0.02532	0.021853	0.024024	0.024492	0.024737	0.023818	0.023702	0.0014129	0.02167	0.02532
	Sim 10	0.021567	0.025207	0.024574	0.022524	0.021938	0.023261	0.023880	0.023279	0.0013574	0.021567	0.025207
	Valor mínimo	0.02002	0.020624	0.020494	0.020502	0.020123	0.020033	0.018928				
	Valor máximo	0.024452	0.025365	0.026646	0.026077	0.024492	0.024926	0.025161				
	Média	0.021722	0.023086	0.022999	0.023094	0.022643	0.022809	0.022601				
Desvio Padrão	0.0010632	0.0017469	0.0019261	0.0017611	0.0013864	0.0014168	0.0013824					

Tabela J20. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 50” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo
		1	2	3	4	5	6	7				
Família Wavelet	Db 1	3.8792	3.9603	3.5422	3.6956	3.9817	3.4356	3.3125	3.6867	0.2655	3.3125	3.9817
	Db 2	3.6729	3.6003	3.4488	3.4011	3.2262	2.9076	3.1275	3.3406	0.2706	2.9076	3.6729
	Db 3	3.5802	3.1878	3.6854	3.4823	3.3267	2.9203	3.2529	3.3479	0.2594	2.9203	3.6854
	Db 4	3.7659	3.9209	3.4771	3.6604	3.34	3.4524	3.3983	3.5736	0.2139	3.3400	3.9209
	Db 5	3.8208	3.4602	3.2323	3.2992	3.0237	3.5097	3.1914	3.3625	0.2602	3.0237	3.8208
	Db 6	3.9247	3.4910	3.5942	3.5933	3.6274	3.5585	3.2337	3.5747	0.2039	3.2337	3.9247
	Db 7	3.8717	3.4307	3.3000	3.4903	3.5493	3.2659	3.2301	3.4483	0.2215	3.2301	3.8717
	Db 8	3.7638	4.0163	3.914	3.1551	3.2097	3.5564	3.4744	3.5842	0.3329	3.1551	4.0163
	Db 9	3.4448	3.6284	3.8223	3.8298	3.9257	3.6188	3.4443	3.6734	0.1915	3.4443	3.9257
	Db 10	3.8496	3.2318	3.8622	3.2837	3.3912	3.4640	3.4912	3.5105	0.2532	3.2318	3.8622
	Sim 4	3.8501	4.0109	3.8202	3.8734	3.493	3.3424	3.4302	3.6886	0.2602	3.3424	4.0109
	Sim 5	3.6668	3.6187	3.6358	3.3419	3.7531	3.3066	3.7565	3.5828	0.1847	3.3066	3.7565
	Sim 6	3.4241	3.5252	3.6818	3.6126	3.2791	3.129	3.5056	3.4511	0.1923	3.1290	3.6818
	Sim 7	3.417	4.0188	3.5358	3.6225	3.3055	3.3428	3.3164	3.5084	0.2543	3.3055	4.0188
	Sim 8	3.3887	3.6594	3.5664	3.7465	3.4660	3.4345	3.5319	3.5419	0.1272	3.3887	3.7465
	Sim 9	3.6354	3.6199	3.5141	3.6413	3.4588	3.1397	3.5248	3.5049	0.1755	3.1397	3.6413
	Sim 10	3.6719	3.5133	3.3836	3.6057	3.355	3.193	3.0881	3.4015	0.2127	3.0881	3.6719
Valor mínimo		3.3887	3.1878	3.2323	3.1551	3.0237	2.9076	3.0881				
Valor máximo		3.9247	4.0188	3.914	3.8734	3.9817	3.6188	3.7565				
Média		3.6840	3.6408	3.5892	3.5491	3.4537	3.3281	3.3712				
Desvio Padrão		0.1794	0.2629	0.1944	0.2001	0.2529	0.2122	0.1735				

Tabela J21. Valores de RMSEP em função do tipo de wavelet e do nível de decomposição empregado no modelo MLR-WT para a estimativa do parâmetro de qualidade “T - 85” aplicado ao sub-conjunto de amostras de previsão dos dados de Óleo Diesel na região espectral de 1.570 – 2.500 nm.

		Nível de Decomposição							Média	Desvio padrão	Valor mínimo	Valor máximo	
		1	2	3	4	5	6	7					
Família Wavelet	Db 1	5.8069	6.0912	5.7103	5.2737	5.7744	4.9976	5.4483	5.5861	0.3692	4.9976	6.0912	
	Db 2	6.4387	5.6243	6.0336	5.601	5.6889	5.2508	4.9483	5.6551	0.4874	4.9483	6.4387	
	Db 3	6.4115	5.7693	5.6917	6.2018	5.6299	5.6322	5.5981	5.8478	0.3240	5.5981	6.4115	
	Db 4	6.0156	5.8406	5.9201	5.5126	5.0555	5.9436	5.5014	5.6842	0.3453	5.0555	6.0156	
	Db 5	5.5841	6.2831	5.8488	5.6276	5.3446	5.677	5.8215	5.7410	0.2917	5.3446	6.2831	
	Db 6	6.1265	6.353	6.004	6.1192	5.7414	5.6214	5.6922	5.9511	0.2717	5.6214	6.3530	
	Db 7	6.1649	6.2442	5.5286	5.4505	5.7767	5.9817	5.588	5.8192	0.3168	5.4505	6.2442	
	Db 8	5.8034	5.9015	6.1031	5.7885	5.618	5.9375	5.9501	5.8717	0.1532	5.6180	6.1031	
	Db 9	6.0552	6.1037	5.8525	6.1666	5.6565	5.6757	6.8373	6.0496	0.4021	5.6565	6.8373	
	Db 10	5.9206	5.7549	5.7853	5.5898	5.4992	5.3473	5.4411	5.6197	0.2073	5.3473	5.9206	
	Sim 4	5.3241	6.0447	6.4119	5.5565	5.7162	5.9375	5.793	5.8263	0.3515	5.3241	6.4119	
	Sim 5	5.508	5.9863	5.9341	5.6816	5.5738	5.7553	5.5802	5.7170	0.1850	5.5080	5.9863	
	Sim 6	5.7495	5.9012	5.7209	5.4672	5.8327	5.5125	5.4337	5.6597	0.1870	5.4337	5.9012	
	Sim 7	5.8858	5.7699	5.8644	5.03	5.3974	5.3346	5.1082	5.4843	0.3571	5.0300	5.8858	
	Sim 8	5.8039	5.9515	5.9302	5.4826	5.7845	5.8111	5.5591	5.7604	0.1771	5.4826	5.9515	
	Sim 9	5.8832	6.0959	5.7967	5.7447	5.9357	5.6057	5.7072	5.8242	0.1624	5.6057	6.0959	
	Sim 10	5.8047	6.0564	5.8552	5.5078	6.1853	5.2853	5.6358	5.7615	0.3120	5.2853	6.1853	
	Valor mínimo		5.3241	5.6243	5.5286	5.0300	5.0555	4.9976	4.9483				
	Valor máximo		6.4387	6.353	6.4119	6.2018	6.1853	5.9817	6.8373				
	Média		5.8992	5.9866	5.8818	5.6354	5.6594	5.6063	5.6261				
Desvio Padrão		0.2911	0.2015	0.1953	0.3061	0.2506	0.2849	0.3956					