



## AVALIAÇÃO DO DESEMPENHO DE ONDALETAS PARA A CONSTRUÇÃO DE MODELOS DE REGRESSÃO MULTIVARIADA EMPREGANDO DADOS DE ESPECTROSCOPIA NO INFRAVERMELHO

M. Werlang<sup>1</sup>, M. Rizzi<sup>2</sup>, E. M. K. Theisen<sup>2</sup>, G. Fenner<sup>2</sup>, M. F. Ferrão<sup>3,\*</sup>, A. E. Gerbase<sup>4</sup>, R. E. P. Pazos<sup>3</sup>

<sup>1</sup> Acadêmica do Curso de Matemática - Universidade de Santa Cruz do Sul - UNISC - RS

<sup>2</sup> Mestre em Sistemas e Processos Industriais - Universidade de Santa Cruz do Sul - UNISC - RS

<sup>3</sup> Professor do Programa de Pós-graduação em Sistemas e Processos Industriais - Universidade de Santa Cruz do Sul - UNISC - RS

<sup>4</sup> Professora do Programa de Pós-graduação em Química - Universidade Federal do Rio Grande do Sul - UFRGS - RS

\*E-mail: [ferrao@unisc.br](mailto:ferrao@unisc.br)

### RESUMO

Neste trabalho utilizou-se a Transformada Discreta Ondaleta (TDO) Daubechies, com 4 coeficientes (Daub4), para compactar a dimensão da matriz de dados espectrais, obtidos por espectroscopia no infravermelho com transformada de Fourier no intervalo 650-4.000  $\text{cm}^{-1}$ , através da reflectância total atenuada (FTIR-ATR) de amostras de polióis de óleo de soja, visando a determinação do valor de hidroxilas (VOH) que para as amostras estudadas corresponde a faixa 23,66-195,04 mg de KOH/g. Através dos modelos de regressão por mínimos quadrados parciais (PLS), comparou-se o desempenho de cada um dos 5 conjuntos de dados compactados, com o original e/ou entre eles. O conjunto de dados dos espectros de polióis, compactado a 1/4 da sua dimensão original, foi considerado como o mais robusto apresentando redução do RMSEP em 20,28% em relação ao modelo com os espectros não compactados e coeficientes de determinação semelhantes. Esse resultado implicou em um modelo de dimensões menores com a mesma capacidade preditiva, assim a TDO (Daub4), mostra ser um método robusto para a redução da dimensão da matriz de dados espectrais, quando pretende-se construir modelos de regressão multivariados.

**Palavras-chave:** PLS, otimização, ondaletas, redução da dimensionalidade, infravermelho, polióis de soja.

### 1 Introdução

As ferramentas de análise multivariada, aplicadas à dados de espectroscopia no infravermelho têm resultado em metodologias de grande utilidade à química analítica. A evolução tecnológica possibilitou o processamento de grandes quantidades de informação, obtidas através dos espectros das amostras estudadas [1].

Porém, trabalhar com um conjunto muito grande de dados, gera um maior custo operacional para alcançar os objetivos propostos e resultados adequados para cada problema. Os espectros no infravermelho são exemplos de conjuntos de dados, onde os comprimentos de onda correspondem às variáveis, e as intensidades encontram-se relacionadas com as estruturas moleculares, dos compostos presentes nas amostras que serão analisadas. Os espectros no infravermelho médio obtidos com resolução de 4  $\text{cm}^{-1}$ , podem apresentar de 1.000 até 2.000 variáveis para uma única amostra, que para determinadas modelagens empregam matrizes com um número muito elevado de dados, resultando em custo computacional significativo, ou até mesmo inexecutável.

Uma das formas de minimizar esse custo de processamento é diminuir a dimensionalidade da matriz de dados, primando por manter o mais fielmente possível as informações pertinentes ao conjunto de dados original. A Transformada Discreta Ondaleta (*Wavelet*) [2-3] apresenta-se como uma excelente solução para a compressão dos dados espectrais, sendo que sua teoria já vem formalizada no processamento de diversos tipos de sinais. Uma das aplicações, que vêm obtendo êxito, consiste em gerar novos conjuntos de dados com dimensões menores e com capacidade de manter, aproximadamente, todas as informações químicas contidas no conjunto de dados inicial. A Transformada Discreta Ondaleta (TDO) tem demonstrado, desde 1989, ser eficiente na área de tratamento de diversos tipos de sinais obtidos na análise instrumental, com aplicações em distintos estudos na química analítica [2]. No caso dos sinais espectrais, uma parte considerável do conjunto de variáveis contém informações irrelevantes ou ruídos, que resultam em maior dificuldade na construção de modelos multivariados. A contribuição das TDOs, na compressão dos dados, se torna ainda mais importante, quando se deseja empregar algum algoritmo de seleção de

variáveis [4], uma vez que reduz a complexidade do sistema, favorecendo o desempenho desse algoritmo. Também, se considerarmos a eliminação de ruídos instrumentais, as TDOs podem auxiliar a construção dos modelos multivariados de regressão, como por exemplo o PLS (*Partial Least Squares*) [4-11], melhorando a capacidade preditiva destes.

Este trabalho apresenta a utilização das TDOs para compactação dos dados espectrais, obtidos por espectroscopia no infravermelho com transformada de Fourier através da reflectância total atenuada horizontal (FTIR-HATR) de amostras de polióis de óleo de soja, visando a determinação do valor de hidroxilas (VOH).

### 1.1 Transformada Discreta Ondaleta (Wavelet)

O avanço nas pesquisas com Transformada Ondaleta permitiu perceber algumas limitações da Transformada de Fourier (TF), a exemplo do estudo de sinais não estacionários e na variação da largura das janelas espectrais, tanto na frequência como no tempo, aplicadas aos dados de interesse [2-3]. Além disso, a TDO é uma técnica computacionalmente mais rápida do que a TF [2].

A TDO realiza o tratamento do sinal, com todas as suas características, independente do tempo e da frequência. Atuando diretamente na frequência do sinal analítico, que contribui para aumentar a diferenciação entre sinal e ruído; e a conservação de picos estreitos com a respectiva informação [3].

Do ponto de vista matemático, a função “ondaleta mãe” (“*mother wavelet*”), gera as funções ondaleta filhas que irão se decompor hierarquicamente [12] conforme esquema apresentado na figura 1. Para cada decomposição são gerados dois subsinais de comprimentos  $n/2$ , onde  $n$  representa o sinal inteiro. O primeiro subsinal corresponde aos sinais escalas (*scalings*, A1), enquanto que o segundo, aos sinais ondaletas (*wavelets*, D1).

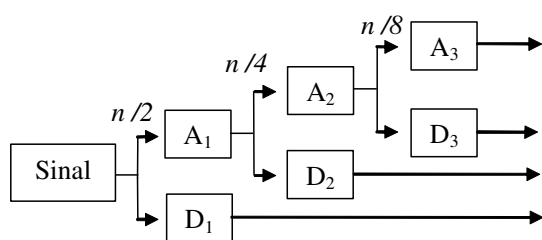


Figura 1 – Esquema do processo de decomposição da Transformada Discreta Ondaleta.

Para criar as funções ondaletas-filhas, além de escolher o tipo de ondaletas é necessário utilizar um banco de filtros, compostos por coeficientes especiais [12]. Toda TDO decompõe um sinal discreto em dois subsinais com a metade do seu tamanho [12-13]. Para tal são empregados filtros digitais, de passa-baixa, os quais produzem os coeficientes de aproximação do sinal (A1), e de passa-alta que produzem os coeficientes de

detalhamento do sinal (D1). O ruído do sinal estará mais concentrado nessa segunda metade [3], e dependendo da aplicação esta parte pode ser descartada ou não [14]. Processando os coeficientes de aproximação (A1) novamente como dados de entrada com o mesmo banco de filtros, gera-se assim outro conjunto de coeficientes de aproximação e de detalhe com um nível de resolução inferior [12-14]. Porém, é desejável que o sinal, que se pretende compactar, esteja com dimensão  $2^n$ , com  $n$  pertencendo aos números naturais ou de valor par, já que este sinal, é sempre reduzido à metade em cada nível de compactação [12].

Para ilustrar este processo, na figura 2a é apresentado um sinal de uma função, e na figura 2b, o sinal da função num primeiro nível de transformada [12]. Com a junção adequada dos sinais escalas e ondaletas da primeira transformada, pode-se recuperar o sinal original, isso implica na preservação da informação dessa função [3].

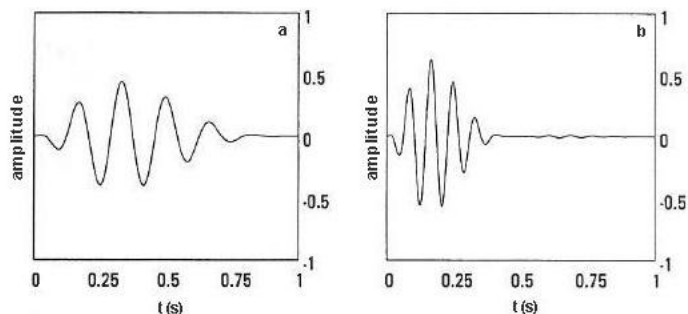


Figura 2 – Representação do sinal de uma função  $g(s)$ , onde (a) é a representação do sinal dessa equação e (b) a representação do sinal dessa equação com um nível de transformada.

Do ponto de vista algorítmico, a TDO, harmoniza as técnicas de decomposição e alisamento (*smoothing*) [2]. A seleção do nível adequado de decomposições e do tipo de ondaletas dependerá do sinal e da experiência envolvidos.

Uma família de ondaletas bastante comum é a TDO Daubechies com quatro coeficientes (Daub4). As funções escalas e ondaletas e seus respectivos sinais são gerados [12] por estes coeficientes:

$$h_0 = \frac{1+\sqrt{3}}{4\sqrt{2}} \quad h_1 = \frac{3+\sqrt{3}}{4\sqrt{2}} \quad h_2 = \frac{3-\sqrt{3}}{4\sqrt{2}} \quad h_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

e

$$g_0 = h_3 \quad g_1 = -h_2 \quad g_2 = h_1 \quad g_3 = -h_0$$

Para obter os sinais escalas da Daub4 é necessário multiplicar os coeficientes escalas pelo sinal ( $s$ ) [12,14], conforme a equação 1.

$$a_i = h_0 s_{2i} + h_1 s_{2i+1} + h_2 s_{2i+2} + h_3 s_{2i+3}, \text{ com } i = [0, \infty) \quad (1)$$

E para obtenção dos sinais ondaletas, procede-se da mesma forma, resultando na equação 2.

$$c_i = g_0 s_{2i} + g_1 s_{2i+1} + g_2 s_{2i+2} + g_3 s_{2i+3}, \text{ com } i = [0, \infty) \quad (2)$$

### 1.2 Regressão por Mínimos Quadrados Parciais (PLS – Partial Least Squares)

O método de regressão multivariada por mínimos quadrados parciais – PLS (*Partial Least Squares*) é uma ferramenta robusta para modelar e analisar qualquer conjunto de dados, permitindo um estudo simultâneo de várias espécies presentes, sendo indiferente à ausência ou não de diferenças espectrais fortes nessas espécies, assim como, uma possível alta correlação dos dados [9]. Esse método de regressão, quando aplicado a problemas da química, propõem uma relação entre dois conjuntos de dados, quando existir dependência entre as propriedades dos mesmos [6]. Assim, quando se tratar de dados espectrais, esse modelo utiliza simultaneamente informações dos espectros e as concentrações (ou propriedades) dessas amostras, para realizar a calibração, onde as informações dos espectros são representadas numa matriz **X** e por sua vez, as concentrações, ou outra variável de interesse, em uma matriz **Y** ou vetor **y** [7]. As linhas da matriz **X** representam as amostras, e as colunas as variáveis, sendo que matriz **X** contém os dados independentes, enquanto que **Y**, os dados dependentes [6].

Segundo Pimentel *et al.* [11], o processo de calibração normalmente consiste em duas etapas. Na primeira, a etapa **descritiva**, medidas feitas numa série de padrões analíticos de concentrações conhecidas são usadas para construir um modelo que relacione a grandeza medida com a concentração da espécie de interesse. Na segunda etapa, conhecida como **preditiva**, usa-se esse modelo para prever concentrações de novas amostras, a partir dos sinais analíticos medidos para elas. A base matemática do PLS é a decomposição da matriz **X**, entendida como a soma de várias matrizes com dimensões unitárias **M<sub>i</sub>** para  $i = 1, 2, \dots, n$ , acrescida da matriz de resíduos (correspondendo ao erro), como está representado na equação 3:

$$\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \dots + \mathbf{M}_n + \mathbf{E} \quad (3)$$

em que  $n$  representa o número de fatores utilizados para truncar a igualdade (as variáveis latentes) e **E** a matriz de resíduos relacionada com esse número de fatores. A decomposição da matriz **X** é realizada conforme as equações 4 e 5 [8,11,15]:

$$\mathbf{X} = \mathbf{t}_1 \mathbf{p}_1^t + \mathbf{t}_2 \mathbf{p}_2^t + \dots + \mathbf{t}_n \mathbf{p}_n^t + \mathbf{E} \quad (4)$$

$$\mathbf{X} = \mathbf{TP}^t + \mathbf{E} \quad (5)$$

Dessa forma a matriz **X**, que contém todas as informações espectrais, passa a ser representada por uma nova matriz **X**, com dimensão menor, descrita pelas matrizes **M<sub>n</sub>** sem a perda de informações relevantes das variáveis originais [6].

E de forma análoga para a matriz **Y**, resultando na equação 6:

$$\mathbf{Y} = \mathbf{UQ}^t + \mathbf{F} \quad (6)$$

onde **T** e **U** são os escores (*scores*) de **X** e **Y**, e **P** e **Q** são os pesos (*loadings*) de **X** e **Y**, respectivamente. O relacionamento de **X** e **Y** resulta na equação linear 7 representando a curva de calibração:

$$\mathbf{c} = \mathbf{Xb} + \mathbf{e} \quad (7)$$

onde **X** representa uma matriz de resposta para as amostras da calibração, **c** é um vetor com os valores das concentrações das amostras, **e** indica os erros do modelo e os ruídos dos espectros e **b** é um vetor representando os parâmetros do modelo [8].

Neste trabalho foi empregado o PLS1 no qual a matriz **Y** representa uma matriz coluna, isto é um vetor **y**.

## 2 Metodologia

### 2.1 Aquisição dos Dados

As amostras de polióis foram preparadas no Instituto de Química da Universidade Federal do Rio Grande do Sul (UFRGS), a partir de óleos de soja refinado. A *American Oil Chemists' Society* (AOCS) [17] recomenda o método de titulação como método de referência, cujo valor de hidroxilas é expresso em miligramas de hidróxido de potássio (KOH) por grama de amostra. Para a obtenção dos espectros desse conjunto de amostras foi utilizado um espectrofotômetro Nicolet Magma 550 equipado com o acessório de HATR (*Horizontal Attenuated Total Reflection*).

### 2.2 Compactação dos Sinais com Transformada Discreta Ondaleta

Para tratar os dados dos espectros dos polióis foi utilizada a TDO Daubechies, com 4 coeficientes, provenientes de um banco de filtros (Daub4), conforme descrito anteriormente.

Na figura 3 é apresentada a matriz-exemplo desses coeficientes, empregada para o cálculo da TDO Daub4 [14], para as compactações realizadas.

$$\begin{bmatrix} h_0 & h_1 & h_2 & h_3 & 0 & 0 & 0 & 0 \\ g_0 & g_1 & g_2 & g_3 & 0 & 0 & 0 & 0 \\ 0 & 0 & h_0 & h_1 & h_2 & h_3 & 0 & 0 \\ 0 & 0 & g_0 & g_1 & g_2 & g_3 & 0 & 0 \\ 0 & 0 & 0 & 0 & h_1 & h_2 & h_3 & h_4 \\ 0 & 0 & 0 & 0 & g_0 & g_1 & g_2 & g_3 \\ h_2 & h_3 & 0 & 0 & 0 & 0 & h_0 & h_1 \\ g_2 & g_3 & 0 & 0 & 0 & 0 & g_0 & g_1 \end{bmatrix} \begin{bmatrix} s_0 \\ s_1 \\ s_2 \\ s_3 \\ s_4 \\ s_5 \\ s_6 \\ s_7 \end{bmatrix}$$

Figura 3 – Esquema da multiplicação da matriz de coeficientes Daub4 com o sinal.

O sinal é representado por um vetor e os coeficientes da TDO Daub4 por uma matriz quadrada, com coeficientes escalas e

ondaletas em linhas intercaladas e transladadas a cada duas colunas, de dimensão igual ao número de linhas do vetor. Procedese com a operação de multiplicação entre matrizes, obtendo como resposta outro vetor, no qual os sinais escalas e sinais ondaletas estarão intercalados, que representará a compactação pela metade do sinal original. Quando os coeficientes chegam ao final da linha, eles são automaticamente recolocados no início da mesma linha (tal como ocorre, nas duas últimas linhas da matriz, com  $h_2, h_3, g_2, g_3$ ). Na seqüência são separados os sinais escalas para a primeira metade do intervalo e os sinais ondaletas para a segunda parte do intervalo estudado [14].

Para os espectros dos polióis foram realizados cinco níveis de compactações: 1/2, 1/4, 1/8, 1/16, 1/32 do sinal original, no ambiente MATLAB [18].

### 2.3 Construção e Avaliação dos Modelos Multivariados

O *iToolbox* para MATLAB foi usado para construir os modelos multivariados por PLS [4], utilizando somente os sinais escalas. Sendo que o conjunto de dados espectrais foi dividindo em 3 grupos, como segue:

**Calibração:** 76 espectros de diferentes polióis obtidos por FT-IR/ATR foram empregados para a construção dos modelos de calibração. As amostras com valores extremos de VOH foram incluídas no conjunto de calibração correspondendo a faixa de 23,66-195,04 mg de KOH/g.

**Validação:** 35 espectros de obtidos por FT-IR/ATR, que não foram utilizados na calibração, foram empregados para a validação dos modelos correspondendo a faixa de 23,77-194,84 mg de KOH/g.

**Previsão:** outros 35 espectros, correspondendo a faixa de 25,90-194,84 mg de KOH/g, de polióis obtidos por FT-IR/ATR foram empregados para avaliar o desempenho dos modelos de calibração.

Todos os modelos de calibração desenvolvidos foram avaliados comparando os valores de hidroxila (VOH) obtidos pelo método proposto e pelo método oficial, com base no coeficiente de determinação ( $R^2$ ) e dos erros de validação [8] do conjunto de amostras usadas para a validação externa [16]. Esse erro é calculado conforme equação 8, sendo denominado de RMSEV (*Root Mean Squares Error of Validation*):

$$RMSEV = \sqrt{\frac{\sum_{i=1}^n (y_{v_i} - \hat{y}_{v_i})^2}{n}} \quad (8)$$

onde  $y_{v_i}$  é o valor de referência e  $\hat{y}_{v_i}$  é o valor estimado e  $n$  o número total de amostras de validação [8].

De forma análoga é calculado o RMSEP (*Root Mean Squares Error of Prediction*), sendo neste caso utilizados os valores das amostras de previsão.

### 3 Discussão dos Resultados

Para ilustrar o comportamento dos espectros de polióis mediante os procedimentos da Transformada Discreta Ondaleta (Daub4), utilizou-se o espectro da amostra 01, antes e depois da aplicação de alguns níveis de compactação. Na figura 4 são apresentados esses resultados, sendo que a figura 4a representa o espectro na sua forma original, com os números de onda que identificam as estruturas moleculares presentes na amostra; já nas figuras 4b, 4c e 4d são representados os espectros com o primeiro, segundo e terceiro níveis de compactação, respectivamente, juntamente com as variáveis resultantes. Pela comparação dessas figuras observa-se que aumentando o número de compactação, aumenta a amplitude do sinal.

Os resultados obtidos para os modelos multivariados PLS, com o emprego da TDO, para os 5 níveis de compactação, são apresentados na tabela 1. Verifica-se que existem, de maneira geral, semelhanças nos resultados desses modelos, para todos os níveis de compactação estudados.

Tabela 1 – Resultados dos modelos de regressão com os valores de RMSEC, RMSEV e RMSEP obtidos.

Modelos PLS						
C <sup>a</sup>	NV <sup>b</sup>	R <sup>2</sup>	RMSEC <sup>c</sup>	RMSEV <sup>c</sup>	RMSEP <sup>c</sup>	VL <sup>d</sup>
1	1736	0,9509	11,39	10,65	12,87	10
1/2	868	0,9740	14,20	10,70	14,85	10
1/4	434	0,9503	11,52	10,64	10,26	10
1/8	216	0,9483	11,68	10,65	12,87	10
1/16	108	0,9469	11,84	10,78	13,26	10
1/32	54	0,9411	12,47	10,99	12,75	9

<sup>a</sup> Nível de compactação

<sup>b</sup> Número de variáveis

<sup>c</sup> Valores expressos em mg de KOH/g

<sup>d</sup> Variáveis latentes

Dentre os modelos PLS gerados, um que pode ser considerado como robusto (menores valores de RMSEC, RMSEV e RMSEP) foi construído com a compactação em 1/4 das informações espectrais originais, isto é, com 2 níveis de transformada, obtendo-se um  $R^2$  equivalente ao do conjunto original dos dados, e com redução do valor RMSEP em 20,28% quando comparado com o modelo dos dados originais, empregando o mesmo número de variáveis latentes. Essa matriz de espectros compactados conseguiu descrever, praticamente, todas as informações contidas no conjunto de dados não compactado, sem causar perda significativa de informações. Nas figuras 5 e 6 são apresentados os gráficos dos erros entre os valores previstos e os valores do método de referência para as amostras de previsão dos modelos com os dados iniciais, e com a aplicação de 2 níveis de transformada, respectivamente.

Verifica-se nos modelos, que o erro tende a aumentar para VOH maiores, porém para o modelo de calibração que emprega todo o conjunto espectros essa medida é mais expressiva. Pode-se atribuir esse comportamento a uma menor linearidade do método



de referência para estas amostras, ou ainda ao fato destas apresentarem distintos sinais do estiramento O-H, podendo ser necessário mais variáveis latentes para sua modelagem. Outro

aspecto interessante refere-se ao fato de que modelo que utilizou os dados compactados apresenta as amostras mais ajustadas na reta de calibração, conforme pode ser observado na Figura 7.

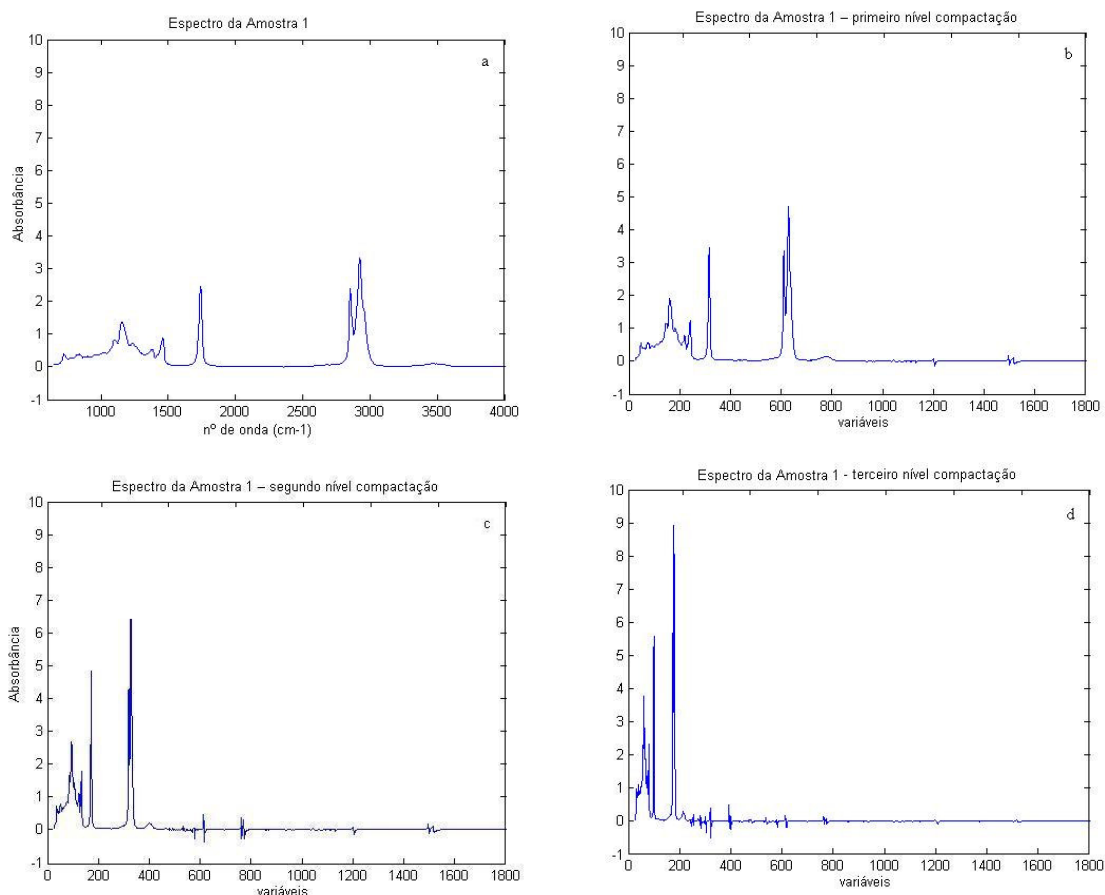


Figura 4 – Sinais obtidos antes e após a aplicação da TDO para o espectro da amostra 01.

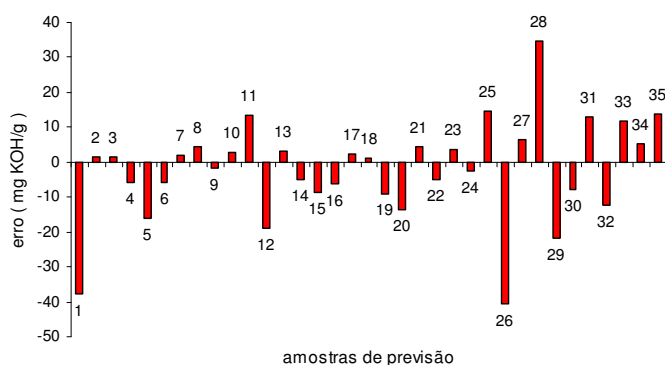


Figura 5 – Gráfico dos erros absolutos entre os valores determinados pelo método AOCS e previstos por FTIR/ATR para o modelo PLS construído com as variáveis originais, empregando 10 variáveis latentes.

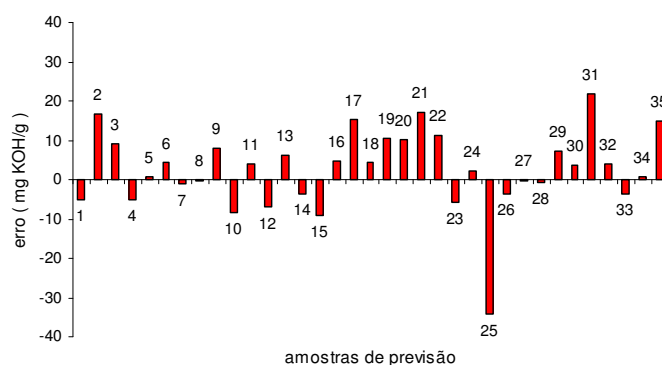


Figura 6 – Gráfico dos erros absolutos entre os valores determinados pelo método AOCS e previstos por FTIR/ATR para o modelo PLS construído com 1/4 das variáveis, empregando 10 variáveis latentes.

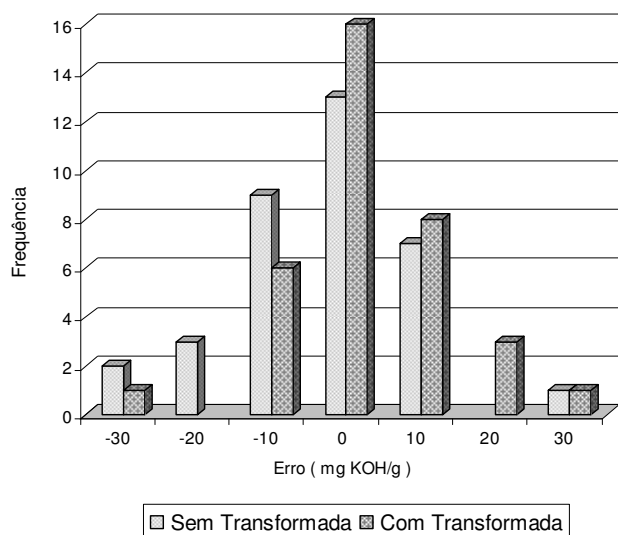


Figura 7 – Histograma dos erros de previsão (mg de KOH/g) para os modelos de calibração PLS.

### Considerações Finais

Este trabalho mostrou a eficiente contribuição da TDO Daub4 para compactar a dimensão do conjunto de dados espectrais, visando à construção de modelos de calibração PLS mais robustos, para determinação do VOH em polióis de óleos de soja empregando os sinais de FT-IR/ATR.

Nos modelos de regressão por mínimos quadrados parciais (PLS), foi constatado, que a compactação do conjunto espectros até 1/8 da sua dimensão original através da TDO Daub4 não inviabiliza a sua capacidade de determinar e explicar, praticamente, as informações presentes no conjunto de dados iniciais.

Desta forma, metodologias para o controle de insumos e/ou produtos industriais podem ser racionalmente desenvolvidas, utilizando a quantidade necessária de dados de espectroscopia por reflexão total atenuada no infravermelho (FT-IR/ATR), obtendo-se métodos mais rápidos e não destrutivos.

### Agradecimentos

Os autores agradecem a CAPES e ao CNPQ pelo suporte recebido.

### EVALUATION OF WAVELET PERFORMANCE TO BUILD MULTIVARIATE REGRESSION MODELS APPLYING INFRARED SPECTROSCOPY DATA

**ABSTRACT:** In this work Discrete Transform Wavelet (DWT) Daubechies with 4 coefficients was used (Daub4) to compact the

matrix dimension of spectral data obtained by Fourier Transform Infrared Spectroscopy in the 650 - 4000  $\text{cm}^{-1}$  range with Attenuated Total Reflectance (FTIR/ATR) of soybean polyol samples, aiming at the determination of hydroxyl value (OHV), which for the studied samples corresponds to the KOH/g 23.66-195.04 mg range. By Partial Least-Squares (PLS) regression models, the performance of each of the 5 different compacted data sets was compared with the original and/or each other. The polyol spectrum data set compacted to 1/4 of its original dimension was considered the better, presenting a 20.28% RMSEP reduction in relation to the model with non-compacted spectra. This result brought about a model with smaller dimensions with the same predictive capacity, thus DWT (Daub4) is a good model for the reduction of the spectral data matrix when multivariate regression models are aimed at being built.

**Keywords:** PLS, optimization, wavelets, dimensional reduction, infrared, soybean polyols.

### REFERÊNCIAS

- COSTA FILHO, P. A.; POPPI R. J.; *Quím. Nova*, 25, 46-52p. 2002.
- CHAU, Y.; LIANG, J. G.; XUE-GUANG, S. *Chemometrics: from basics to wavelet transform*. New Jersey – Canada: John Wiley & Sons Inc., . v. 164. 2004.
- DANTAS FILHO, H. A. *Desenvolvimentos de técnicas quimiométricas de compressão de dados e redução de ruído aplicadas a óleo diesel e madeira de eucalipto empregando espectroscopia NIR*. Tese, UNICAMP, Campinas, 2007.
- NORGAARD, L.; SAUDLAND, A.; WAGNER, J.; NIELSEN J. P.; MUNCK, L. & ENGELSEN, S. B.; *Appl. Spectrosc.*, 54, 413-419p. 2000.
- BRERETON, G. Richard. *Chemometrics: data analysis for the laboratory and chemical plant*. West Sussex: Willey, 2003.
- CORDEIRO, G. A. *Desenvolvimento de metodologias espectroscópicas multivariadas para quantificação de fármacos em formas farmacêuticas*. Dissertação, UFPR, Curitiba, 2006.
- GARCIA, M. H. F., FARIAS, S. B., FERREIRA, B. G. *Polímeros*, 14, 322-325p, 2004
- FURTADO, J. C., FERRÃO, M. F., KONZEN, P. H. A., MOLZ, R. F., BASSANI, I. A.; *Tecno-Log.*, 6, 41-71p, 2002
- FERREIRA, M. M. C., ANTUNES, A. M., MELGO, M. S., VOLPE, P. L. O.; *Quím. Nova*, 22, 724-731p. 1999
- RIBEIRO, F. A. L. *Aplicação de métodos de análise multivariada no estudo de hidrocarbonetos policíclicos aromáticos*. Dissertação, UNICAMP, Campinas, 2001.
- PIMENTEL, M. F.; BARROS NETO, B. *Quím. Nova*, 19, 268-277p. 1996.
- DAUBECHIES, I. *Ten lectures on wavelets*. SIAM, Philadelphia, PA, 1992.

13. MALLAT, S. *A wavelet tour of signal processing*. New York: Academic Press, 1998.
14. JENSEN, A., La COUR-HARBO, A. *Ripples in mathematics*. Springer, 2001.
15. FERRÃO, M. F. *Aplicação de técnicas espectroscópicas de reflexão no infravermelho no controle de qualidade de farinha de trigo*. Tese, UNICAMP, Campinas, 2000.
16. SANTOS M. C. S. *Utilização da espectroscopia na região do infravermelho (FTIR) e quimiometria na identificação do café torrado e moído adulterado*. Dissertação, UFLA, Lavras, 2005.
17. AMERICAN OIL CHEMISTS' SOCIETY. *Official Methods and Recommended Practices of de American Oil Chemists' Society*, 4th edn, Champaign, 1993, revised 1997.
18. WISE, B. M.; GALLAGHER, N. B.; *PLS Toolbox for use with Matlab*, v. 2.1, Eigenvector Technologies: Manson, 1998.