

UNIVERSIDADE FEDERAL DA PARAÍBA
CENTRO DE CIÊNCIAS EXATAS E DA NATUREZA
DEPARTAMENTO DE QUÍMICA
PROGRAMA DE PÓS-GRADUAÇÃO EM QUÍMICA



DISSERTAÇÃO DE MESTRADO

**Uso do método MDL para filtragem de ruído instrumental
empregando a Transformada *Wavelet***

por

Gledson Emidio José

JOÃO PESSOA, março de 2008

Gledson Emidio José

**Uso do método MDL para filtragem de ruído instrumental
empregando a Transformada *Wavelet***

**Dissertação apresentada ao programa
de Pós-Graduação em Química, da
Universidade Federal da Paraíba, como
parte dos requisitos à obtenção do
título de mestre em Química, área de
concentração em Química Analítica.**

Orientador: Prof. Dr. Mário César Ugulino de Araújo

Co-orientador: Prof. Dr. Roberto Kawakami Harrop Galvão



JOÃO PESSOA, março de 2008

J 83 u José, Gledson Emidio.

Uso do método MDL para filtragem de ruído instrumental empregando a Transformada *Wavelet* / Gledson Emidio José. – João Pessoa, 2008.

109 p.

Orientador: Mário César Ugulino de Araújo.

Co-orientador: Roberto Kawakami Harrop Galvão.

Dissertação (mestrado) – UFPB/CCEN

1. Química Analítica. 2. Remoção de ruído. 3.

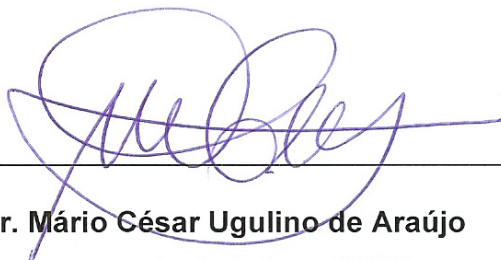
Transformada *Wavelet*.

UFPB/BC

CDU 543 (043)

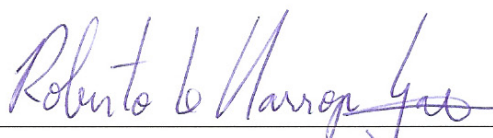
Uso do Método MDL para Filtragem de Ruído Instrumental Empregando a Transformada *Wavelet*

Aprovada pela banca examinadora:



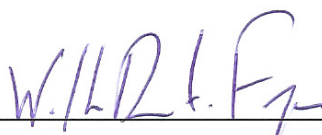
Prof. Dr. Mário César Ugulino de Araújo

Departamento de Química – UFPB
1º orientador



Prof. Dr. Roberto Kawakami Harrop Galvão

Divisão de Engenharia Eletrônica – ITA
2º orientador



Prof. Dr. Wallace Duarte Fragoso

Departamento de Química Fundamental – UFPE
Avaliador



Prof. Dr. Edvan Cirino da Silva

Departamento de Química – UFPB
Avaliador

Para meu pai e minha mãe,
EMIDIO e CIRLEI

Agradecimentos

Ao **Conselho Nacional de Desenvolvimento Científico e Tecnológico, CNPq**, pela bolsa de estudos concedida;

À **Universidade Federal da Paraíba**, pelo apoio institucional;

À **Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES**, através do **Programa Nacional de Cooperação Acadêmica, PROCAD 0081/05-1**, pelo auxílio financeiro durante a missão de estudo;

Ao professor **Mário César Ugulino de Araújo**, pela orientação e confiança nos últimos quatro anos de pesquisa;

Ao professor **Roberto Kawakami Harrop Galvão**, pela orientação e acolhimento cordial e atencioso em todas as minhas visitas ao Instituto Tecnológico da Aeronáutica;

Aos amigos **Francisco Gambarra, Márcio Coelho e Sófacles Carreiro**, pelas valiosas discussões quimiométricas e pelo companheirismo;

Ao professor **Wallace Fragoso**, pelos conselhos e conversas sobre Quimiometria;

Ao professor **Célio Pasquini**, pelo acolhimento no Grupo de Instrumentação e Automação (GIA-UNICAMP);

À **Simone Simões**, pela obtenção dos espectros de Capropril;

À **Heronides Adonias**, pela cessão dos espectros de óleo diesel;

Ao **Núcleo Multi-usuário de Caracterização de Análise da UFPB, NUCAL**, pelas análises em RMN, e à **Josean Fachine**, pela cessão da amostra de extrato vegetal;

À **Juliana Cortez**, pela estada em Campinas durante uma das missões de estudo;

Aos **amigos e familiares** que me apoiaram nessa lida;

À **todos** que me ajudaram direta ou indiretamente nesse processo;

À **Fabiane**, por tudo.

Sumário

LISTA DE FIGURAS.....	x
LISTA DE TABELAS.....	xiv
LISTA DE ABREVIATURAS E SIGLAS.....	xvi
RESUMO.....	xvii
ABSTRACT.....	xviii
1 INTRODUÇÃO	2
1.1 Objetivos	5
2 FUNDAMENTAÇÃO TEÓRICA	6
2.1 A Transformada de Fourier.....	7
2.2 A Transformada <i>Wavelet</i>	8
2.3 Remoção de ruído por Transformada <i>Wavelet</i>	12
2.3.1 Escolha da função <i>wavelet</i>	15
2.3.1.1 Escolha baseada na literatura	18
2.3.1.2 Escolha por inspeção visual dos resultados	18
2.3.1.3 Escolha pela forma da função <i>wavelet</i>	18
2.3.1.4 Escolha pelo MDL	18
2.3.2 Limiar	19
2.3.2.1 Limiar Universal.....	19
2.3.2.2 Limiar por variância.....	20
2.3.2.3 Limiar MDL	20
2.3.3 Escolha da função de limiarização.....	21
3 EXPERIMENTAL	24
3.1 Técnicas	24

3.2 Experimentos	25
3.2.1 Experimento 1 – Dados simulados.....	26
3.2.2 Experimento 2 – Captopril em comprimidos / Espectrometria NIRR.....	26
3.2.3 Experimento 3 – Dipirona sódica em pó / Espectrometria NIRR.....	27
3.2.4 Experimento 4 – Óleo Diesel Combustível / Espectrometria NIR	28
3.2.5 Experimento 5 – Extrato vegetal / Espectrometria ¹ H RMN	29
3.3 Programas e Pacotes Computacionais	30
4 RESULTADOS E DISCUSSÕES	33
4.1 Experimento 1 – Sinais simulados.....	34
4.1.1 Ruído com desvio-padrão de 0,01.....	35
4.1.2 Ruído com desvio-padrão de 0,3.....	40
4.1.3 Ruído com desvio-padrão de 1,0.....	44
4.1.4 Análise dos diferentes níveis de ruído	47
4.2 Experimento 2 – Captopril / Espectrometria NIRR	49
4.3 Experimento 3 – Dipirona Sódica / Espectrometria NIRR.....	57
4.4 Experimento 4 – Óleo Diesel / Espectrometria NIR	62
4.5 Experimento 5 – Extrato vegetal / ¹ H RMN	67
4.6 Estudo dos resultados obtidos nos Experimentos 1-5.....	71
4.6.1 Escolha da <i>Wavelet</i>	72
4.6.2 Limiar e função de limiarização.....	73
4.6.3 Espectros obtidos com poucas varreduras	74
5 CONCLUSÕES	77
5.1 Propostas futuras	78
REFERÊNCIAS BIBLIOGRÁFICAS	80
APÊNDICE A	86
APÊNDICE B	86
APÊNDICE C	93

Lista de figuras

Figura 2.1 Esquema gráfico mostrando o funcionamento da translação e escalamento da função <i>wavelet Symlet 8</i>	9
Figura 2.2 Esquema da decomposição de um sinal pelos filtros passa-baixa e passa-alta.....	10
Figura 2.3 Esquema de um banco de filtros <i>wavelet</i> para dois níveis de decomposição.	11
Figura 2.4 Esquema de um banco de filtros <i>wavelet</i> para três níveis de decomposição, mostrando os coeficientes em cada etapa.	11
Figura 2.5 Esquema geral do Filtro de Quadratura Espelhada, onde é mostrado o processo de decomposição e reconstrução do sinal.....	12
Figura 2.6 Forma de algumas funções escala (ϕ) e <i>wavelet</i> (ψ) e da família Daubechies.	16
Figura 2.7 Forma de algumas funções <i>wavelet</i> (ψ) e escala (ϕ) da família <i>Coiflet</i>	16
Figura 2.8 Forma de algumas funções <i>wavelet</i> (ψ) e escala (ϕ) da família <i>Symlet</i>	17
Figura 2.9. Valor do coeficiente antes da limiarização e valor do coeficiente após a limiarização Suave e Rígida.	22
Figura 3.1 (a) equipamento ABB-BOMEM; (b) suporte para análises de sólidos.....	27

Figura 3.2 Espectrômetro com Transformada de Fourier Perkin Elmer Spectrum GX.	28
Figura 3.3 Espectrômetro FT-NIR BOMEM usado para as análises de óleo diesel. .	29
Figura 3.4 (a) Espectrômetro Bruker DRX-200 utilizado para as análises de ^1H RMN; (b) Detalhe do compartimento de inserção da amostra.....	30
Figura 4.1 Sinal sintético livre de ruído e os sinais gerados pela adição de ruído em três níveis de desvio-padrão.	34
Figura 4.2 Sinais reconstruídos pela TW usando as técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso (SD=0,01).....	35
Figura 4.3 Curva do RMSE médio x função <i>wavelet</i> para as três técnicas de remoção de ruído por TW em um total de cem sinais gerados com ruído (SD=0,01).	36
Figura 4.4 Gráfico de barras apresentando a posição da <i>wavelet</i> selecionada pelo MDL para cada um dos cem sinais gerados, com relação ao RMSE obtido.....	37
Figura 4.5 Curva do RMSE x função <i>wavelet</i> para a técnica MDL dos sinais número 5 e 17.	37
Figura 4.6 Frequência de seleção da mesma função <i>wavelet</i> pelo MDL nos cem sinais sintéticos.	38
Figura 4.7 Gráfico de barras com o número de coeficientes selecionados pelo limiar universal (barras cinza) e MDL (barras pretas) através das 22 <i>wavelets</i> para (a) o espectro 5 e (b) o espectro 17.....	39
Figura 4.8 De cima para baixo, sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso (SD=0,3).	40

Figura 4.9 Curva do RMSE médio x função <i>wavelet</i> para as três técnicas de remoção de ruído por Transformada wavelet em um total de cem sinais gerados com ruído (SD=0,3).....	41
Figura 4.10 Gráfico de barras apresentando a posição da <i>wavelet</i> selecionada pelo MDL para cada um dos cem sinais gerados, com relação ao RMSE obtido.	41
Figura 4.11 Curva do RMSE x função <i>wavelet</i> para a técnica MDL dos sinais número 6 e 60.	42
Figura 4.12 Frequência de seleção da mesma função <i>wavelet</i> pelo MDL nos cem sinais sintéticos.	42
Figura 4.13 Curva do RMSE x função <i>wavelet</i> da técnica MDL para os sinais número 6 e 60.	43
Figura 4.14 Sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso (SD=1,0).....	44
Figura 4.15 Curva do RMSE médio x função <i>wavelet</i> para as três técnicas de remoção de ruído por Transformada wavelet em um total de cem sinais gerados com ruído (SD=1,0).....	45
Figura 4.16 Gráfico de barras apresentando a posição da <i>wavelet</i> selecionada pelo MDL para cada um dos cem sinais gerados, com relação ao RMSE obtido.	45
Figura 4.17 Curva do RMSE x função <i>wavelet</i> para a técnica MDL dos sinais número 11 e 20.	46
Figura 4.18 Frequência de seleção da mesma função <i>wavelet</i> pelo MDL nos cem sinais sintéticos.	46
Figura 4.19 Curva do RMSE x função <i>wavelet</i> da técnica MDL para os sinais número 11 e 20. A <i>wavelet</i> escolhida é marcada por um triângulo.	47

Figura 4.20 Espectro de refletância de um comprimido do princípio-ativo captopril na região de 3.999-10.000 cm^{-1}	49
Figura 4.21 Coeficientes de detalhe do primeiro nível de decomposição para o espectro de captopril.	50
Figura 4.22 Espectro de referência (100 varreduras) e espectro ruidoso (1 varredura) de um comprimido do princípio-ativo captopril na região de 4.610-8.928 cm^{-1}	51
Figura 4.23 Cento e um espectros sobrepostos de um comprimido do princípio-ativo captopril na região de 4610-8928 cm^{-1} . (a) antes e (b) depois da correção da linha de base por EMSC. Em detalhe, corte ampliado da região entre 5300-5700 cm^{-1}	51
Figura 4.24 Sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso para a mesma amostra de captopril.	52
Figura 4.25 Comparação do número de coeficientes <i>wavelet</i> escolhidos através das 22 <i>wavelets</i> para os espectros 5 e 47, respectivamente, usando-se o limiar universal (barras cinza) e o limiar MDL (barras azuis).....	56
Figura 4.26 Gráfico de barras apresentando a extensão da seleção do filtro wavelet pelos nos espectros individuais.....	57
Figura 4.27 Espectro de referência e um espectro de varredura unitária de uma amostra do princípio-ativo dipirona sódica, (a) antes e (b) depois da correção da linha de base pelo EMSC.	58
Figura 4.28 Sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso para a mesma amostra do princípio-ativo dipirona sódica.	59
Figura 4.29 De cima para baixo, sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso para a mesma amostra do princípio-ativo dipirona.	60

Figura 4.30 Curva do RMSE médio x função <i>wavelet</i> para as três técnicas de remoção de ruído por Transformada <i>wavelet</i> para os espectros de dipirona sódica.	61
Figura 4.31 Espectro bruto, espectro de referência e os espectros reconstruídos após remoção de ruído pelas técnicas em estudo.	63
Figura 4.32 Corte realizado na região do <i>short-NIR</i> , do espectro bruto, espectro de referência e dos espectros reconstruídos após remoção de ruído.	63
Figura 4.33 Curva do RMSE médio x função <i>wavelet</i> para as quatro técnicas de remoção de ruído por Transformada <i>Wavelet</i> para os espectros de óleo diesel.	64
Figura 4.34 Curvas do número de coeficientes selecionados pelas quatro técnicas de remoção de ruído, para as 22 funções <i>wavelet</i> .	66
Figura 4.35 Espectro de ^1H RMN obtido pela promediação de (cima) 1024 varreduras, e (baixo) 16 varreduras.	67
Figura 4.36 Espectros de ^1H RMN reconstruídos pela TW através das técnicas MDL, Rígida e Suave, e os espectros obtidos pela promediação de (cima) 1024 varreduras, e (baixo) 16 varreduras.	68
Figura 4.37 Corte nos espectros de ^1H RMN reconstruídos pelas técnicas MDL (vermelho), Rígida (verde) e Suave (azul).	69
Figura 4.38 Corte nos espectros de captopril, reconstruídos pela técnica Suave usando-se as wavelets Daubechies 1-3 e 9-10.	72

Lista de tabelas

Tabela 1. Valores de RMSE médios obtidos com 1 varredura e 7 níveis de decomposição no banco de filtros.....	53
Tabela 2. Comparação entre as técnicas para o melhor (espectro 47) e o pior (espectro 5) resultado na seleção da função <i>wavelet</i> pelo critério MDL.	54
Tabela 3. Valores de RMSE, obtidos pelas quatro técnicas frente aos vinte e dois filtros wavelet, entre o espectro de referência e o obtido pela promediação de 16 varreduras..	65
Tabela 4. Comparação entre as técnicas para o melhor e o pior resultado na seleção da função <i>wavelet</i>	70

Lista de abreviaturas e siglas

a	Variável de escalamento
b	Variável de translação
c	Coeficientes de aproximação
Coif	Coiflet
CDCL ₃	Clorofórmio deuterado
CLAE-DAD	Cromatografia Líquida de Alta Eficiência- Detecção de Arranjo de diodo
d	Coeficientes de detalhe
Db	Daubechies
DTGS	Sulfato de Triglicina Deuterada
EMSC	Correção de Sinal Multiplicativo Estendida
FT-NIR	Transformada de Fourier – Infravermelho Próximo
FT	Transformada de Fourier
H	Filtro passa-baixa
InGaAs	Arseneto de Gálio e Índio
KBr	Brometo de Potássio
L	Filtro passa-alta
LIBS	Espectrometria por Emissão com Plasma Induzido por Laser
MDL	Comprimento Mínimo de Descrição
MRA	Análise Multi-resolucional
NIR	Infravermelho Próximo
NIRR	Refletância Difusa no Infravermelho Próximo
PAT	Tecnologia Analítica de Processos
QMF	Filtros de Quadratura Espelhada
RMN	Ressonância Magnética Nuclear
SD	Desvio-padrão
SNR	Relação sinal-ruído
Sym	<i>Symlet</i>
TFJ	Transformada de Fourier Janelada
TW	Transformada <i>Wavelet</i>
TWC	Transformada <i>Wavelet</i> Contínua

TWD Transformada *Wavelet* Discreta
TWI Transformada *Wavelet* Inversa
WT *Wavelet* Transform

Resumo

Existem várias técnicas que visam à melhora da relação sinal-ruído em sinais espectrométricos. Uma das mais populares consiste na realização de medidas repetidas e subsequente promediação das varreduras individuais, o que, dependendo da aplicação, pode demandar muito tempo, além de promover deslocamentos na intensidade e no eixo de comprimento ou número de onda. Uma alternativa é a remoção de ruído usando técnicas de processamento de sinais baseadas na Transformada *Wavelet* (TW). Nesta dissertação, três técnicas de remoção de ruído por TW, a saber, Suave (limiarização Suave e limiar universal), Rígida (limiarização Rígida e limiar universal) e MDL (limiarização Rígida e limiar MDL) foram comparadas em relação à qualidade dos seus espectros reconstruídos e a capacidade de compressão de dados. Além de sinais simulados, foram utilizados espectros de Ressonância Magnética Nuclear e Espectrometria NIR, obtidos por um número pequeno de varreduras. Entre essas técnicas, o Comprimento Mínimo de Descrição (MDL, do inglês: *Minimum Description Length*) tem a vantagem de selecionar automaticamente tanto o tipo do filtro quanto o limiar. Assim, os resultados obtidos pelo MDL em relação à seleção da função *wavelet* também foram avaliados. Varreduras individuais foram utilizadas a fim de estudar o comportamento das diferentes técnicas frente à ação de ruído instrumental real. Os resultados mostraram que o método MDL pode ser uma boa alternativa como método de remoção de ruído instrumental em dados espectrométricos, provendo espectros pouco ruidosos, além das vantagens quanto à parcimônia e seleção automática da função *wavelet* e do limiar, permitindo uma análise sem inferências subjetivas por parte do usuário. O procedimento MDL aplicado a espectros promediados por um número pequeno de varreduras mostrou ser útil em Ressonância Magnética Nuclear, onde um considerável número de varreduras deve ser adquirido para obter uma adequada relação sinal-ruído, e na Espectrometria NIR como ferramenta em Tecnologia de Processos Analíticos (PAT), onde a rapidez de medidas é imprescindível.

Palavras-chave: Remoção de ruído, Transformada *Wavelet*, MDL, Espectrometria NIR, Espectrometria de RMN, varreduras individuais.

Abstract

There are several approaches that seek to improve the signal-to-noise ratio (SNR) in spectroscopic signals, with the most popular consisting of carrying out repeated measurements and coaveraging the individual scans. However, such process is time-consuming and may promote drifts in the intensity and wavelength axes. An alternative consists of using signal processing techniques for noise reduction, such as the Wavelet Transform (WT). Among the denoising techniques based on the WT, the MDL (Minimum Description Length) method selects both the filter type and threshold in an automatic manner. The Hard, Soft and MDL techniques are compared in terms of denoising and compression, and the MDL results are evaluated with regard to filter selection issues. This dissertation presents an investigation of wavelet denoising procedures for NIR and RMN spectra. In particular, coaveraged spectra are employed in order to study the effect of actual instrumental noise, rather than simulated noise. A simulated example with Gaussian noise in three levels of standard deviation also available. The results showed that the MDL method may be a good alternative to standard wavelet denoising methods concerning the removal of real instrumental noise in NIR spectrometry, due to the automatically selection of filter type and threshold, which does not require subjective choices from the analyst. The MDL denoising procedure applied to coaveraged scans should be particularly useful in nuclear resonance spectroscopy, where a considerable number of scans may be required to achieve an adequate signal-to-noise ratio, and in applications of Near Infrared Spectroscopy, as tool for process analysis, where rapid measurements are vital.

Keywords: Wavelet Denoising, Minimum Description Length, NIR spectrometry, NMR spectrometry, individual scans.

introdução



1 Introdução

A Química Analítica moderna é uma ciência eminentemente instrumental, muito diversa daquela praticada no final do século XIX, quando os procedimentos de precipitação, extração ou destilação eram essenciais. A partir de 1900, novos métodos baseados em fenômenos físicos, como a condutividade, a absorção e a emissão de luz, a relação carga/massa *etc.*, trouxeram uma riqueza de possibilidades para o campo das análises químicas, as quais se concretizaram durante os cem anos subsequentes, devido aos avanços nas áreas da microeletrônica digital, óptica, mecânica e computação. Como resultado, tivemos o aparecimento de técnicas mais rápidas e sensíveis, capazes de produzir uma grande quantidade de dados sobre a amostra em questão, possibilitando o desenvolvimento de métodos de análise mais robustos. Um exemplo destas técnicas é a Cromatografia Líquida de Alta Eficiência acoplada a um sistema de detecção de arranjo de diodos (CLAE-DAD), que gera milhões de dados espectrocromatográficos a cada medida (CHAU FON-TIN *et al.*, 2004).

Devido à enorme quantidade de dados gerados por tais técnicas, a premência por ferramentas capazes de extrair informação química desse emaranhado de números promoveu a união de várias áreas do conhecimento, como a Matemática, a Estatística e a Computação. Surgiu assim a Quimiometria, um conjunto de ferramentas aplicadas a uma análise racional de medidas químicas. Multidisciplinar por natureza, visa criar ou selecionar procedimentos ótimos de medida e experimentos, bem como extrair a máxima informação possível dos dados químicos (MASSART *et al.*, 1997). A Quimiometria permeia todo o processo analítico, desde a parte inicial, no planejamento dos experimentos (BARROS NETO *et al.*, 2007), a coleta dos dados, o tratamento dos sinais obtidos (WENTZELL e BROWN, 2000), a manipulação dos dados através de reconhecimento de padrões (MASSART *et al.*, 1997b) ou quantificação (GELADI, 2002) de um ou mais constituinte, até a etapa final de interpretação dos resultados.

Uma etapa crucial em uma modelagem quimiométrica é o tratamento das medidas analíticas adquiridas. A maioria delas é obtida via instrumentos computadorizados, como em espectrometrias no Infravermelho próximo (NIR, do

inglês: *Near Infrared*) e no Infravermelho Médio (MIR, do inglês: *Medium Infrared*), em Ressonância Magnética Nuclear (RMN) e também em Cromatografia Líquida de Alta Eficiência (WENTZELL e BROWN, 2000). A fidelidade da informação analítica resultante é dependente de como ela foi primeiramente processada. Tal tratamento exerce influência direta na qualidade das etapas subseqüentes da modelagem (BRERETON, 2003), e por isso, deve-se proceder com cautela na escolha das diferentes opções de técnicas de processamento de sinais.

Um dos problemas que afetam a integridade de uma medida analítica é a presença de ruído, que pode ser originado por interferentes físicos no aparato experimental, ou qualquer flutuação do sinal recebido pelo detector. Toda medida apresenta um componente ruidoso, afetando sobremaneira a obtenção de informação útil de um sistema químico.

Para contornar esse problema, os espectrômetros modernos trabalham com um procedimento de promediação, que consiste em registrar um espectro várias vezes e depois calcular o espectro médio. Como o ruído é aleatório, observando-se valores negativos e positivos, este efeito é diminuído pela promediação. Assim, a parte determinística do sinal é amplificada (FELINGER, 1998), aumentando-se a relação sinal-ruído por um fator de \sqrt{N} , com N sendo o número de repetições (WENTZELL e BROWN, 2000), o que permite a remoção de ruído sem distorcer o sinal.

Porém, para se obter um sinal confiável, N precisa ser grande, e quanto maior for N, mais tempo será necessário para a aquisição do sinal, o que é desvantajoso se a rapidez das respostas for fundamental, como em Tecnologia Analítica de Processos (PAT, do inglês: *Process Analytical Technology*). De fato, em aplicações da PAT, o controle de qualidade de um produto (por exemplo, medicamentos) é realizado diretamente na linha de produção, em várias etapas da cadeia produtiva. Um controle apropriado requer que as informações obtidas através dos espectros sejam analisadas imediatamente, de modo a permitir a tomada de decisões sem a interrupção do processo (FDA, 2004).

Uma alternativa consiste em aplicar técnicas de processamento de sinais após a obtenção do espectro promediado com um número pequeno de varreduras. As principais técnicas de processamento de sinais empregadas pelos químicos analíticos são a Savitzky-Golay e a Transformada de Fourier (TF) (CERQUEIRA et al, 2000). No entanto, desde 1990, métodos baseados na Transformada *Wavelet*

(TW) têm aparecido como opção a essas técnicas, recebendo grande atenção da comunidade quimiométrica (CHAU FON-TIN et al., 2004).

O interesse crescente dos químicos pela Transformada *Wavelet* se justifica pela sua capacidade de representar simultaneamente as propriedades do tempo e da frequência de um sinal, o que pode revelar informações químicas importantes acerca da substância ou analito em estudo. Além disso, a natureza multiresolucional da transformada, obtida pelas versões transladadas e escalonadas da função *wavelet*, possibilita a representação do sinal em várias escalas, o que pode ser vantajoso para fins de extração de características.

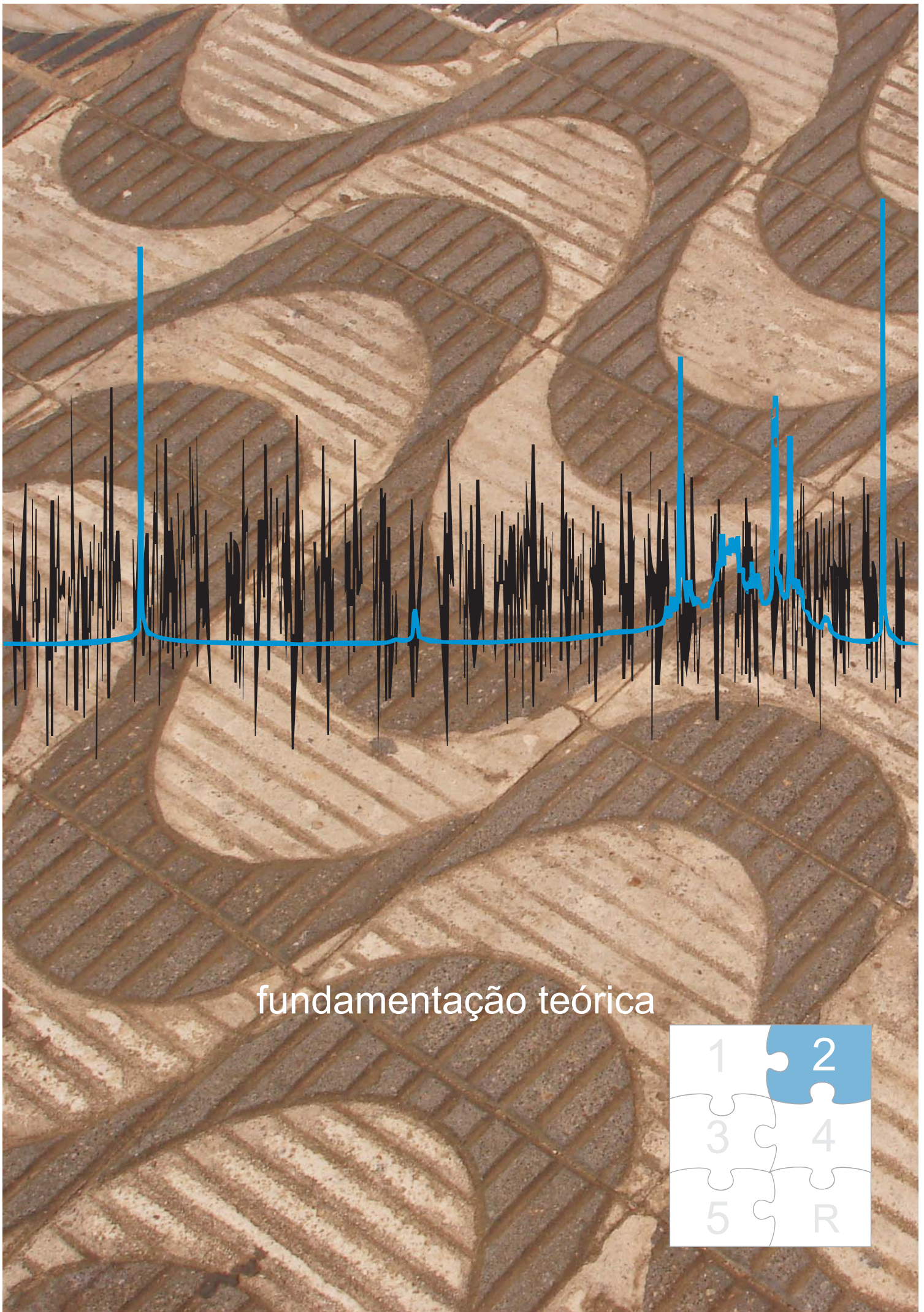
O bom desempenho da TW com relação à compressão de sinais tem tornado possível a utilização eficiente de muitos dos métodos espectrométricos modernos, devido à alta dimensionalidade das respostas analíticas. Na Transformada *Wavelet*, essa compressão pode ser realizada empregando técnicas de limiarização, com duas técnicas sendo mais amplamente utilizadas, a saber, limiarização Rígida (*Hard thresholding*) e limiarização Suave (*Soft thresholding*), que consistem em remover ou atenuar os coeficientes *wavelet* de pequena amplitude (DAUBECHIES, 1992), que possuem baixa relação sinal-ruído. Para isso, há a necessidade de uma distinção entre quais coeficientes deverão ser atenuados ou eliminados, o que é realizado pelas técnicas de limiar, cujo representante mais conhecido e utilizado é o limiar universal (DONOHO, 1995b), apesar de nos últimos anos outras técnicas terem sido desenvolvidas (SAITO, 1996) (GALVÃO et al., 2006).

A escolha da *wavelet*, da função de limiarização e do limiar são as questões que dominam as pesquisas sobre remoção de ruído por Transformada *Wavelet*. Segundo (MANGANIELLO et al, 2002) o único método que permite selecionar automaticamente o limiar e a função *wavelet* com um critério definido é o Comprimento Mínimo de Descrição (MDL, do inglês: *Minimum Description Length*), introduzido por Jorma Rissanem (RISSANEM, 2000) e adaptado por Naoki Saito (SAITO, 1996). Baseado no princípio da parcimônia, o MDL busca o mínimo de coeficientes que expliquem a maior parte da informação.

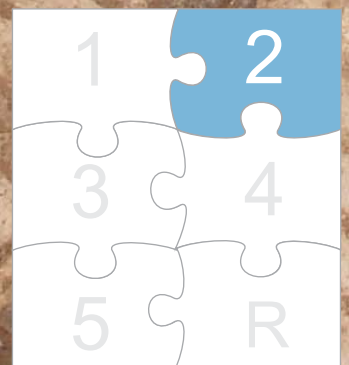
Como o MDL não requer a decisão do analista sobre estes parâmetros, torna-se uma técnica muito mais simples e viável para a manipulação dos espectros obtidos pela promediação de poucas varreduras, trazendo agilidade e economia às análises, com conseqüente obtenção mais rápida da informação de interesse.

1.1 Objetivos

A presente dissertação teve como objetivo estudar a remoção de ruído instrumental por Transformada *Wavelet* através do uso da técnica MDL, além de realizar uma investigação com as abordagens Suave e Rígida, visando comparar as características de cada uma no tocante à melhora da relação sinal-ruído de espectros promediados por um número pequeno de varreduras.



fundamentação teórica



2 Fundamentação teórica

Todos os métodos analíticos instrumentais produzem medidas. Na espectrometria NIR, por exemplo, a medida é um espectro, que é comumente obtido passando a radiação infravermelha através de uma amostra e determinando que fração da radiação incidente é absorvida.

Um espectro é uma seqüência de medidas discretas, constituída por um componente que contém a informação desejada do analito ou substância de estudo, e outro que contém informação espúria, que traz diversas conseqüências à precisão da medida (WENTZELL e BROWN, 2000). Nenhuma medida está livre desse último componente, que pode ter proveniência química ou instrumental. Dessa forma, o espectro (y) pode ser caracterizado como:

$$y = x + \varepsilon \quad (2.1)$$

com x sendo o sinal relacionado ao analito e ε a função ruído.

Em medidas espectrométricas, o ruído pode ser homoscedástico ou heteroscedástico (BARROS NETO et al., 2007). Ambos mascaram a informação química sobre o analito de interesse e podem diminuir a eficácia de modelos quimiométricos.

No ruído homoscedástico, características como média e desvio-padrão permanecem constantes ao longo de todo o espectro. Já a natureza do ruído heteroscedástico depende das características da medida para cada dado (WENTZELL e BROWN, 2000), o que dificulta a eliminação do ruído.

A maioria dos equipamentos produz algum componente heteroscedástico. Além do mais, o ruído homoscedástico em um espectro pode ser convertido em heteroscedástico através de uma transformação não-linear como transmitância para absorbância (ALSBERG et al., 2007) (WOODWARD et al., 2008), tornando a remoção do componente ruidoso uma tarefa não-trivial, por serem desconhecidas as características do ruído. Sua análise e eliminação são realizadas usando-se

estimativas aproximadas de suas características, através do uso de técnicas elaboradas de processamento de sinais, como as transformadas matemáticas.

O objetivo de qualquer transformada é revelar certas características que não são aparentes no sinal original, facilitando o processo de análise e processamento da transformação. As Transformadas de Fourier e *Wavelet* fazem a conversão do sinal no domínio temporal (ou de comprimento de onda) para o domínio das frequências. Isso pode ser uma grande vantagem, visto que, dependendo da característica do sinal analítico, grande parte da informação está inacessível visualmente, sendo mais facilmente extraídas no domínio das frequências, possibilitando a visualização de uma gama muito maior de informações que estariam imperceptíveis no domínio temporal.

2.1 A Transformada de Fourier

A Transformada de Fourier exerce um papel importante na Química, sendo comumente usada na maioria dos instrumentos analíticos (SHAO et al., 2003). Descoberta em 1807 por Joseph Fourier, baseia-se no princípio de que uma forma de onda periódica pode ser expressa como uma infinita soma de senos e co-senos de várias frequências.

A TF opera sobre um sinal contínuo, usando a sobreposição de funções seno e co-seno para a conversão de domínio do sinal. Se o sinal for estacionário, isto é, se as características do sinal não mudarem com o tempo, não haverá problemas, mas como a maioria dos sinais apresenta características não-estacionárias e transitórias, como início e fim de eventos, *drifts*, mudanças abruptas de eventos etc, (MISITI et al., 1996), a TF pode não ser apropriada para a detecção de informação do sinal. Na Transformada de Fourier, apenas a informação das frequências é mantida, e toda a informação referente ao tempo é descartada. Além disso, como o sinal é tratado globalmente, não observando as peculiaridades de cada faixa espectral, pode haver comprometimento do processamento caso tenhamos algum trecho muito ruidoso. (GALVÃO et al., 2001)

Na tentativa de resolver estas deficiências, Gabor introduziu uma versão da TF que é janelada no tempo, técnica que chamou de Transformada de Fourier

Janelada (TFJ), e que dispõe o sinal em ambos os domínios tempo e freqüência. Este melhoramento consiste em usar uma janela de comprimento finito e movê-la ao longo do sinal em questão (ALSBERG, 1997). Assim, para cada deslocamento da janela, a TF é processada na região do sinal compreendida pelo comprimento da janela.

Porém, como o tamanho da janela temporal na TFJ é fixa, estabelecer uma largura ideal que contemple as particularidades de cada porção do sinal é uma tarefa não-trivial. Em regiões de alta freqüência, necessitaremos de uma janela pequena; em baixas freqüências, a janela deverá ser maior. Haverá sempre a perda de informação, seja no tempo ou na freqüência.

2.2 A Transformada *Wavelet*

A transformada *Wavelet* é uma ferramenta que divide o sinal em componentes de diferentes freqüências, e então estuda cada componente com uma resolução combinada com sua escala (DAUBECHIES, 1992). Sua vantagem comparada com a Transformada de Fourier é a boa localização em ambos os domínios de tempo e freqüência. Semelhante à TFJ, a TW é calculada através de uma janela que percorre todo o sinal. A diferença da TW é que essa janela é completamente modulável e escalonável. Assim, o processo é repetido inúmeras vezes com janelas maiores ou menores para cada novo ciclo (MISITI et al., 1996). Como resultado, tem-se uma coleção de representações tempo-freqüência do sinal, em diferentes resoluções.

As funções de base usadas na TW são chamadas *wavelets* $\Psi_{(a,b)}$. Como o próprio nome sugere, *wavelets* são pequenas ondas que diferem do zero apenas em um domínio limitado, chamado suporte. Este suporte pode variar, permitindo um diferente grau de localização.

As *wavelets* são versões escalonadas e transladadas de uma função protótipo chamada *wavelet*-mãe, Ψ . O processo de escalamento da *wavelet* é realizado pelo parâmetro **a**, que permite alongá-la ou comprimi-la. Já a translação, desempenhada pelo parâmetro **b**, consiste em movê-la ao longo do eixo das abscissas, como uma função janelante. Mudando-se o parâmetro **b**, o centro da localização no tempo é

movido, pois cada $\Psi_{a,b}(t)$ é localizado em volta de b , com um espalhamento proporcional ao parâmetro a (MALLAT, 1999), provendo uma descrição tempo-freqüência de $f(t)$ (DAUBECHIES, 1992). Uma descrição do processo de rescalonamento e translação pode ser vista na Figura 2.1.

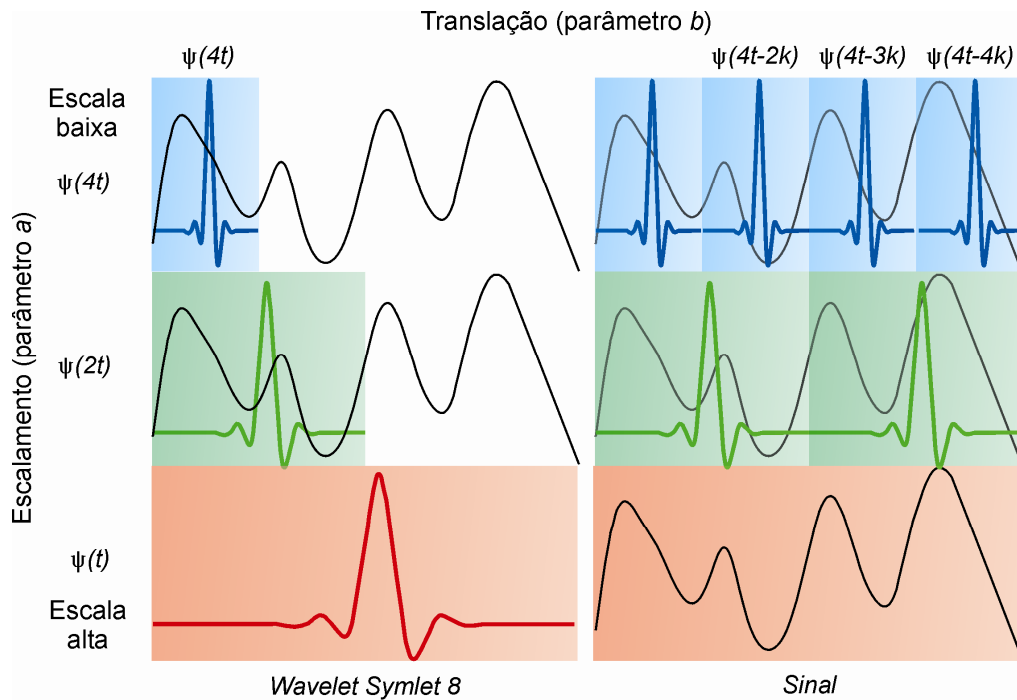


Figura 2.1 Esquema gráfico mostrando o funcionamento da translação e escalamento da função *wavelet Symlet 8*.

Conhecendo as funções e os parâmetros usados na análise *wavelet*, podemos calcular a Transformada *Wavelet* Contínua (TWC) de um sinal:

$$\begin{aligned} \text{TWC}(a,b)[f(t)] &= \int_{-\infty}^{\infty} \psi_{a,b}^* f(t) dt \\ &= \langle \psi_{a,b} | f(t) \rangle \end{aligned} \tag{2.2}$$

Sendo a função $\Psi_{a,b}$ definida como:

$$\Psi_{a,b} = \frac{1}{\sqrt{|a|}} \Psi\left(\frac{t-b}{a}\right), \quad a,b \in \mathfrak{R}, a \neq 0 \tag{2.3}$$

Por razões práticas, estes parâmetros são freqüentemente discretizados, levando a então chamada Transformada *Wavelet* Discreta (TWD), com os valores de

a e **b** respeitando a restrição $a = 2^m$ e $b = n2^m$, onde m e n são números inteiros (GALVÃO et al., 2001). Depois da discretização, a função *wavelet* é definida como:

$$\psi_{a,b}(t) = 2^{-\frac{a}{2}} \psi(2^{-j}t - b), \quad (2.4)$$

Realizar a Transformada *wavelet* em um sinal é equivalente a processá-lo com um banco de filtros, gerando como produto os coeficientes *wavelet*. Uma forma eficiente de se produzir tais filtros foi proposta por Stéphane Mallat, em 1982, e nomeada de Análise Multi-resolucional (MRA, do inglês: *Multiresolucional Analysis*) (STRANG e NGUYEN, 1996).

Na análise MRA, as funções *wavelets* são relacionadas a um banco de filtros, que possui dois filtros, um passa-baixa (L) e outro passa-alta (H), que são aplicados ao sinal de entrada, um espectro NIR, por exemplo, separando-o em bandas de frequência (STRANG e NGUYEN, 1996). A forma destes filtros é determinada pelo tipo de função *wavelet* usada (ALSBERG, 1997b). O filtro passa-baixa, relacionado à chamada função escala Φ , gera os coeficientes de aproximação (c), que são os componentes de baixa frequência. O filtro passa-alta, por sua vez, está relacionado à função *wavelet*, produzindo os coeficientes de detalhe (d), de alta frequência, conforme o esquema presente na Figura 2.2.

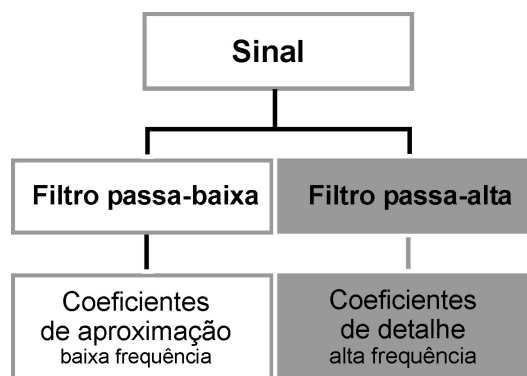


Figura 2.2 Esquema da decomposição de um sinal pelos filtros passa-baixa e passa-alta.

Após a filtragem, os coeficientes passam por um processo de decimação, que consiste em remover um de cada dois coeficientes. O conjunto de coeficientes de aproximação, por sua vez, passa em seguida por uma nova divisão L/H, de maneira

a gerar coeficientes c_{m+1} e d_{m+1} e assim sucessivamente (GALVÃO et al., 2001), conforme podemos ver na Figura 2.3.

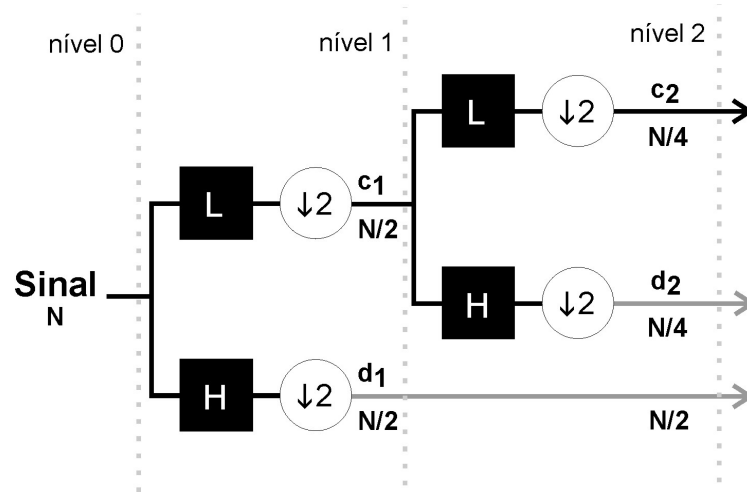


Figura 2.3 Esquema de um banco de filtros *wavelet* para dois níveis de decomposição.

Como resultado, teremos um vetor com o mesmo tamanho do sinal de entrada, contendo os coeficientes de aproximação do último nível de decomposição, que possui baixa resolução, e os coeficientes de detalhe de todos os níveis de decomposição, de baixa e alta resolução. Uma representação gráfica desse processo é encontrada na Figura 2.4.

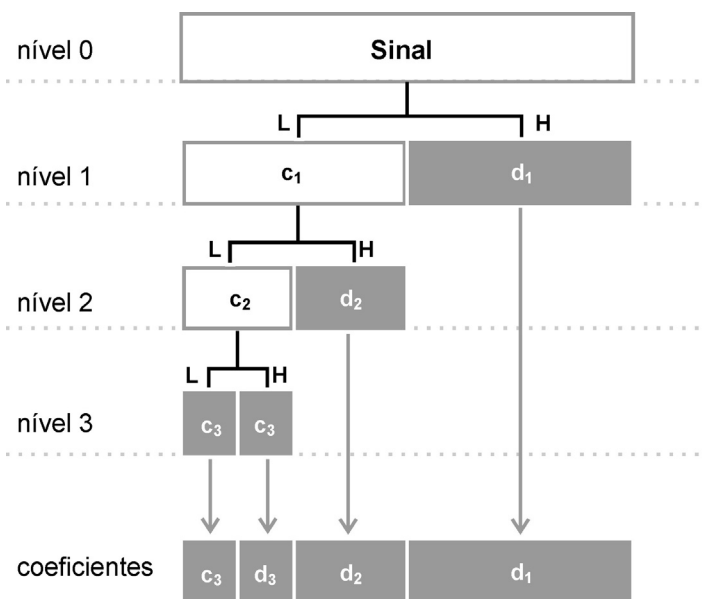


Figura 2.4 Esquema de um banco de filtros *wavelet* para três níveis de decomposição, mostrando os coeficientes em cada etapa.

Para a reconstrução ou síntese do sinal (Figura 2.5), que deve ser realizada sem perda de informação, a Transformada *Wavelet* Inversa é aplicada, seguindo o procedimento inverso à decomposição. Os filtros de decomposição (passa-altas e passa-baixas), juntamente com os respectivos filtros de reconstrução, formam um sistema chamado de filtros de Quadratura Espelhada (QMF) (MISITI et al., 1996).

Como os coeficientes de aproximação e detalhe passaram por decimação, eles não podem ser combinados diretamente para reconstruir o sinal. Eles devem ser reconstruídos separadamente, pela inserção de zeros entre os coeficientes, e depois combinados.

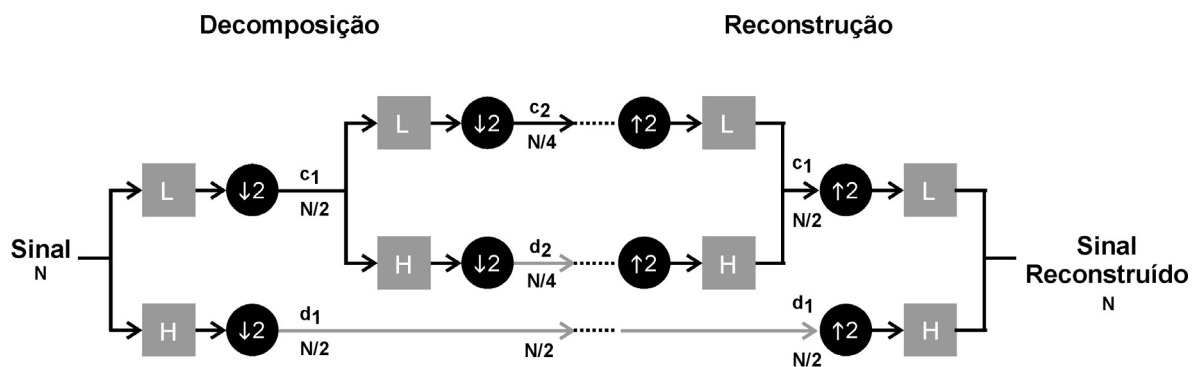


Figura 2.5 Esquema geral do Filtro de Quadratura Espelhada, onde é mostrado o processo de decomposição e reconstrução do sinal.

2.3 Remoção de ruído por Transformada *Wavelet*

Uma medida experimental geralmente é contaminada por algum tipo de ruído, que pode interferir na interpretação dos resultados e nas etapas de modelagem quimiométrica. O ruído é um fenômeno que afeta todas as frequências, como o sinal relacionado ao analito tende a dominar os componentes de baixa frequência, é esperado que a maioria dos componentes de alta frequência, abaixo de certo nível, seja ruído (ALSBERG et al., 1997a). Os métodos de remoção de ruído por Transformada *Wavelet* têm como objetivo separar esse componente ruidoso do sinal.

O uso da Transformada *Wavelet* pelos químicos analíticos na tarefa de remoção de ruído tem crescido vertiginosamente, em substituição a técnicas como a Savitzky-Golay (BEEBE et al., 1998). Na realização da remoção de ruído pela TW, três etapas devem ser seguidas:

- Transformação do sinal em questão para o domínio *wavelet*;
- Seleção dos coeficientes mais informativos, e subsequente tratamento;
- Reconstrução do espectro filtrado.

■ ETAPA 1 — Transformação do sinal para o domínio *wavelet*

Uma das premissas da formulação empregada para a TW é que o sinal deve ter um comprimento diádico, com N sendo o número de níveis de decomposição. Caso isso não seja possível, existem vários métodos para a extrapolação do comprimento do sinal, como extensão simétrica e *zero-padding* (CHAU FON-TIN et al., 2004). Tais técnicas têm como objetivo diminuir a distorção das bordas do espectro, e envolvem a inserção de alguns coeficientes adicionais em cada estágio do processo de decomposição, a fim de obter uma perfeita reconstrução (MISITI et al., 2000). Na técnica *zero-padding*, a adequação do comprimento do sinal é realizada pela inserção de zeros, enquanto na extensão simétrica é feita uma replicação simétrica do sinal. Neste texto, porém, trabalharemos apenas com dados com comprimento múltiplo de 2^N .

Para a realização da decomposição do sinal através do banco de filtros *wavelet*, devemos escolher quantos níveis de decomposição serão utilizados. O número máximo depende do tamanho do sinal e da função *wavelet*. Quanto maior for o nível, maior será a compressão do sinal e a remoção de ruído. No entanto, efeitos de borda podem prejudicar a análise caso o número seja muito alto. Recomenda-se utilizar o número máximo de níveis dado pela função *wmaxlev.m*, do Matlab® *Wavelet Toolbox* (MISITI et al., 2000).

A escolha da *wavelet* que será usada depende da natureza dos dados. A qualidade dos sinais reconstruídos é diretamente dependente da função *wavelet* empregada.

■ ETAPA 2 — Seleção dos coeficientes mais informativos e subsequente tratamento

Uma transformada não remove informação ou ruído, ela apenas move-os de lugar. Para que a transformada *wavelet* seja aplicada na tarefa de remoção de ruído, ou em outras aplicações, deve-se fazer a seleção dos coeficientes mais representativos. Tais coeficientes podem ser separados respeitando um limiar de corte. Valores acima do limiar podem ser mantidos ou atenuados e valores abaixo do corte podem ser excluídos. Para isso, devemos escolher o tipo de função de limiarização. Este procedimento baseia-se na idéia básica de que a energia de uma função freqüentemente será concentrada em poucos coeficientes *wavelet*, enquanto a energia do ruído será espalhada ao longo de todos os coeficientes (POORNACHANDRA e KUMARAVEL, 2006), isto é, somente alguns coeficientes contribuem com o sinal propriamente dito.

■ ETAPA 3 — Reconstrução do espectro filtrado

A reconstrução do espectro é realizada através da Transformada *Wavelet* Inversa (TWI), utilizando somente os coeficientes selecionados pelas funções de limiarização. Após esta etapa, tem-se uma versão menos ruidosa do sinal original.

Ao depararmos com estas etapas, algumas perguntas aparecem imediatamente:

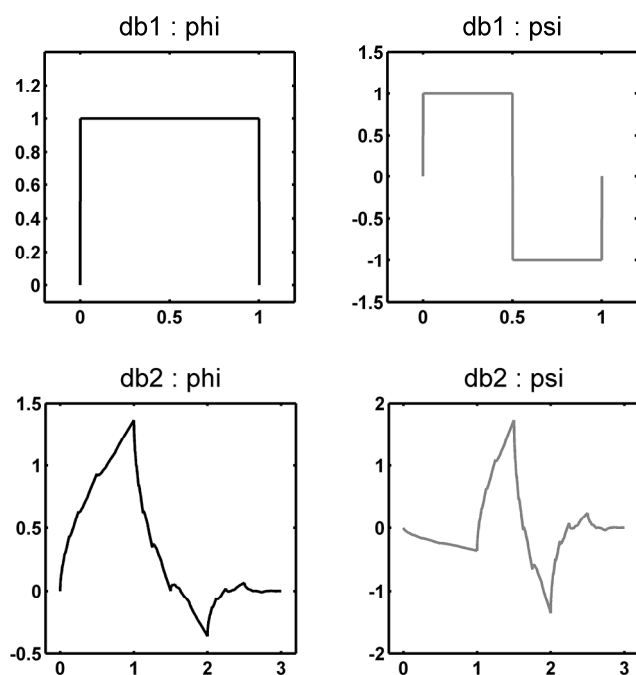
- Qual função *wavelet* devemos usar?
- Quais coeficientes escolheremos?
- Que faremos com os coeficientes escolhidos?

Dependendo da característica do sinal utilizado, as respostas para essas perguntas podem variar completamente. Por isso, devem ser tratadas com muita cautela e experiência sobre o sinal em questão.

2.3.1 Escolha da função *wavelet*

Como existe uma infinidade de diferentes *wavelets* descritas na literatura, escolher uma que se adapte bem à natureza dos dados é uma tarefa desafiadora. Apesar de algumas famílias serem mais difundidas, as funções que as integram possuem propriedades que variam entre si, como ordem dos filtros de decomposição, momentos de decaimento, grau de oscilação, ortogonalidade etc, dificultando ainda mais a escolha (ARGOUD et al., 2004). É necessário entender as características da *wavelet* para determinar se ela é adequada ou não para a natureza do sinal em questão. Entre as famílias mais conhecidas e usadas na literatura de Química Analítica temos as famílias Daubechies, *Symlet* e *Coiflet*.

As Daubechies compreendem *wavelets* que diferem no comprimento do filtro. A primeira é a *wavelet* mais antiga e também a mais simples, conhecida como *wavelet* de Haar. Elas apresentam uma forma muito assimétrica. Na Figura 2.6, são mostradas as formas de algumas funções escala e *wavelet* da família Daubechies.



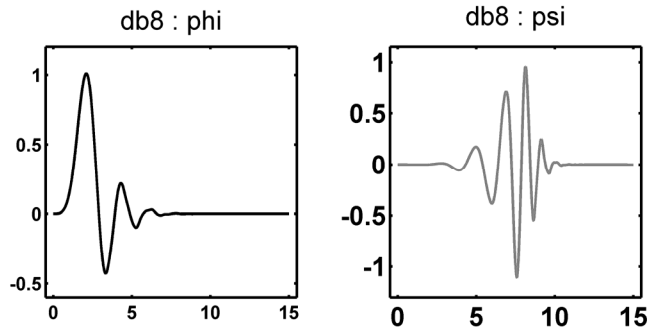


Figura 2.6 Forma de algumas funções escala (ϕ) e *wavelet* (ψ) e da família Daubechies.

As *Coiflets* foram construídas por Daubechies em colaboração com R. Coifman (DAUBECHIES, 1992). Na Figura 2.7 são mostradas algumas funções escala e *wavelet* da família *Coiflet*.

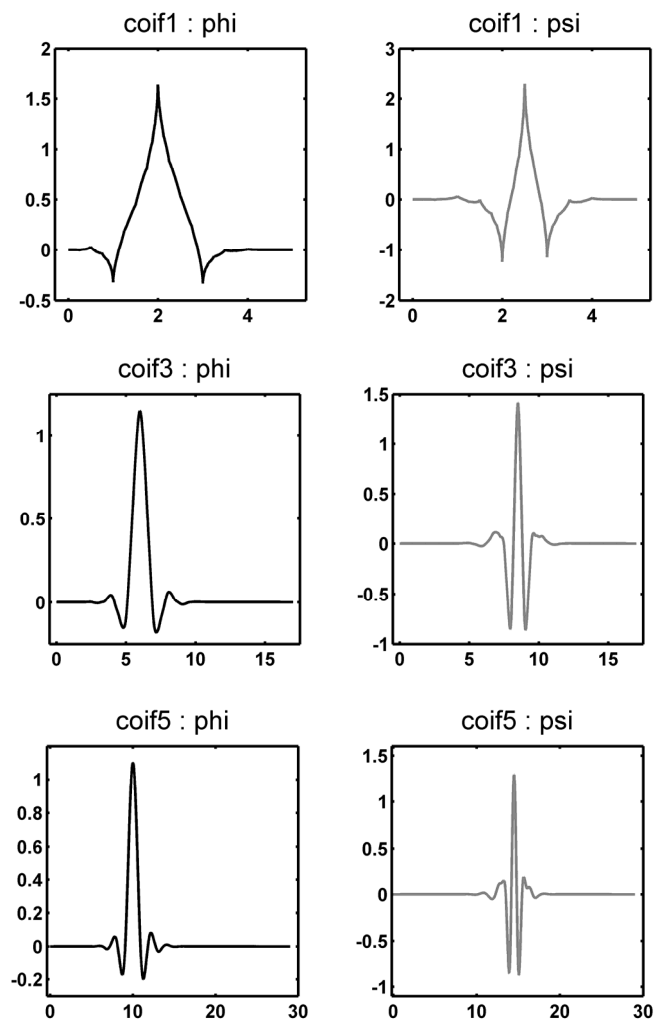


Figura 2.7 Forma de algumas funções *wavelet* (ψ) e escala (ϕ) da família *Coiflet*.

As *Symlets*, também produzidas por Ingrid Daubechies, apresentam maior simetria e são comumente preferidas na análise de espectros vibracionais (ALSBERG, 1997). Na Figura 2.8, podem ser visualizadas algumas das funções escala e *wavelet* da família *Symlet*.

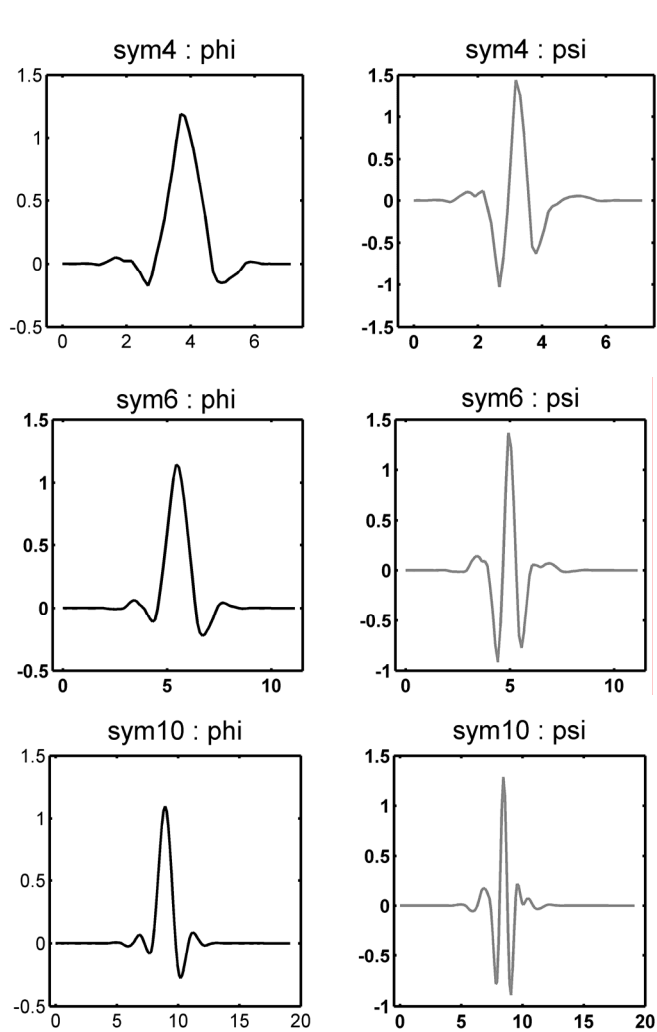


Figura 2.8 Forma de algumas funções *wavelet* (*psi*) e escala (*phi*) da família *Symlet*.

Vale salientar que os filtros correspondentes às *wavelets* *dbN*, *coifN* e *symN* possuem comprimento $2N$, $6N$ e $2N$, respectivamente.

Entre os métodos mais utilizados para a escolha da *wavelet* temos o uso baseado em indicações da literatura, comparação visual, forma da *wavelet* e a técnica MDL.

2.3.1.1 Escolha baseada na literatura

Devido à ausência de técnicas mais robustas para a escolha da *wavelet* que se adapte bem à natureza dos dados, freqüentemente tal escolha se baseia em indicações na literatura que a tenham utilizado em um tipo particular de sinal. Em Espectroscopia NIR, por exemplo, a função *wavelet Symlet 8* tem sido muito usada, talvez por influência do clássico tutorial de TW escrito por [Alsberg e colaboradores \(1998b\)](#).

2.3.1.2 Escolha por inspeção visual dos resultados

Testar o comportamento das diferentes *wavelets* frente a uma determinada aplicação não é uma tarefa fácil, pode consumir muito tempo computacional e a comparação efetiva dos resultados nem sempre é possível.

2.3.1.3 Escolha pela forma da função *wavelet*

A forma da função *wavelet* que pretendemos usar deve ser capaz de traduzir as características do espectro. É um critério muito usado, embora seja difícil de implementar devido à quantidade de *wavelets* disponíveis, além de ser passível de inferências do usuário.

2.3.1.4 Escolha pelo MDL

O MDL é o único método que realiza a escolha sistemática da *wavelet*. Como ele também seleciona automaticamente um limiar, será tratada mais detalhadamente na próxima seção.

2.3.2 Limiar

A escolha do limiar de corte é a mais importante da remoção de ruído por Transformada *Wavelet*, e por isso, requer muita atenção e critério para a escolha da configuração correta dos parâmetros necessários.

Os coeficientes com pequena amplitude podem ser descartados sem perda significativa de informação. Os coeficientes restantes, que portam a maior parte da informação, podem ser aplicados nas tarefas de remoção de ruído, na seleção de variáveis para regressão, análise discriminante, detecção de pontos de inflexão etc. No entanto, a escolha destes coeficientes é uma tarefa trabalhosa, e deve ser realizada criteriosamente, para não perder informação válida, ou remover pouco ruído indesejável.

A maioria das técnicas de remoção de ruídos trabalha zerando os coeficientes abaixo de um limiar pré-estabelecido, e/ou atenuando os coeficientes acima deste limiar. Escolher o limiar correto é a questão-chave.

2.3.2.1 Limiar Universal

O limiar universal é baseado nas propriedades estatísticas do ruído branco, Gaussiano. Sua fórmula é a seguinte:

$$\text{limiar} = \sigma \sqrt{2 \log N} \quad (2.5)$$

onde σ é o desvio padrão do ruído e N é o comprimento do sinal. Como na maioria dos sinais reais o σ é desconhecido, uma estimativa é obtida através da mediana dos coeficientes da mais fina escala de resolução (DONOHO et al., 1995b), isto é, dos coeficientes de detalhe do primeiro nível de decomposição, sendo dividido pelo fator 0,674. A equação da estimativa do ruído pode ser vista abaixo:

$$\sigma = |\text{mediana}(d1)| / 0,674 \quad (2.6)$$

onde d_1 são os coeficientes de detalhe do primeiro nível de decomposição. A estimativa é feita nesses coeficientes, pois se assume que sejam compostos essencialmente por ruído. No entanto, caso haja presença de alguma porção de sinal entre estes coeficientes, pode ocorrer superestimação do ruído (DONOHO e JOHNSTONE, 1994).

2.3.2.2 Limiar por variância

Este método baseia-se na escolha dos coeficientes cuja variância seja maior que um valor estipulado, por exemplo, uma porcentagem de variância de 90%. Dessa forma, os coeficientes que apresentarem variância inferior a 90% são eliminados.

2.3.2.3 Limiar MDL

O critério MDL, proposto por Jorma Rissanen em 1978 (RISSANEM, 2000), estabelece que a melhor representação de um sinal é aquela que minimiza o número de *bits* utilizados para representar o sinal e os parâmetros desta representação. É baseado na idéia de que qualquer regularidade nos dados pode ser usada para comprimi-lo (GRUNWALD, 2005), e assim, descrevê-lo usando um número menor de pontos que o original.

Naoki Saito (1996) aplicou o critério MDL para estimar um sinal discreto de uma medida ruidosa usando uma biblioteca de funções *wavelet*. Seu método objetiva obter simultaneamente uma boa remoção de ruído e compressão dos sinais, selecionando automaticamente o limiar e a função *wavelet*.

O MDL consiste em uma função custo que objetiva expressar um compromisso entre o número de coeficientes retidos e o erro da reconstrução do sinal. É expresso como a soma de dois termos conflitantes, segundo a Equação 2.7,

$$\text{MDL}(k^*, m^*) = \min \left(\frac{3}{2} k \log(N) + \frac{N}{2} \log \left(\sum (x_m^2 - x_{mk}^2) \right) \right) \quad (2.7)$$

onde k é o número de maiores coeficientes que foram retidos, m designa o tipo de filtro, x_m os coeficientes do tipo da transformada m , x_{mk} para os k maiores coeficientes em amplitude, e k^* e m^* para os valores otimizados. O valor correspondente a k^* é o limiar escolhido, e m^* é a função *wavelet* escolhida.

O primeiro termo representa uma função penalizante, que aumenta linearmente com o número de coeficientes *wavelet* retidos. O segundo termo caracteriza a energia residual, que é o erro entre o sinal reconstruído e o sinal original. O número de coeficientes escolhido como o mínimo por esta equação é considerado o limiar. Semelhante operação ocorre com a seleção da melhor *wavelet*.

2.3.3 Escolha da função de limiarização

Após a escolha do limiar de corte, precisamos saber em qual abrangência do universo de coeficientes iremos aplicá-lo, e em que magnitude eles serão modificados. Duas estratégias apresentam vasto uso pela comunidade científica, ambas desenvolvidas por David Donoho e colaboradores ([DONOHO, 1995](#)), sendo a primeira chamada de função de limiarização Suave (*Soft thresholding*), e a segunda chamada de função de limiarização Rígida (*Hard thresholding*).

Nessas duas abordagens clássicas, todos os coeficientes cujas magnitudes sejam inferiores ao valor do limiar são zerados, pois se presume que sejam resultantes de ruído. A diferença entre elas reside no tratamento dado aos coeficientes com magnitudes superiores ao limiar. Enquanto na abordagem Rígida os coeficientes acima do limiar são mantidos inalterados, na abordagem Suave eles são atenuados pelo valor do limiar.

Embora sejam utilizadas intensamente, tais técnicas são passíveis de alguns problemas. Entre eles, a limiarização Suave pode ser tendenciosa devido à atenuação de coeficientes de maior amplitude. Já a limiarização Rígida, por ser uma função descontínua, como podemos ver na [Figura 2.9](#), pode ser sensível a

pequenas mudanças no perfil do sinal. (POORNACHANDRA e KUMARAVEL, 2006). O comportamento das duas funções de limiarização é mostrado em um sinal de 100 pontos, conforme a Figura 2.9.

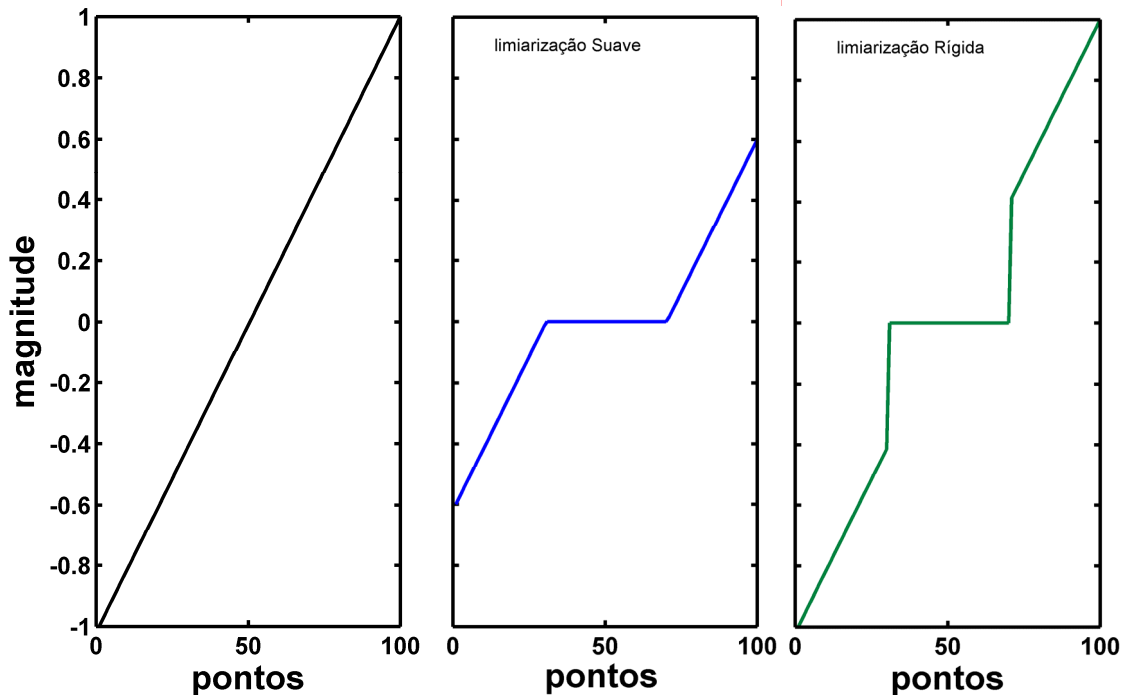
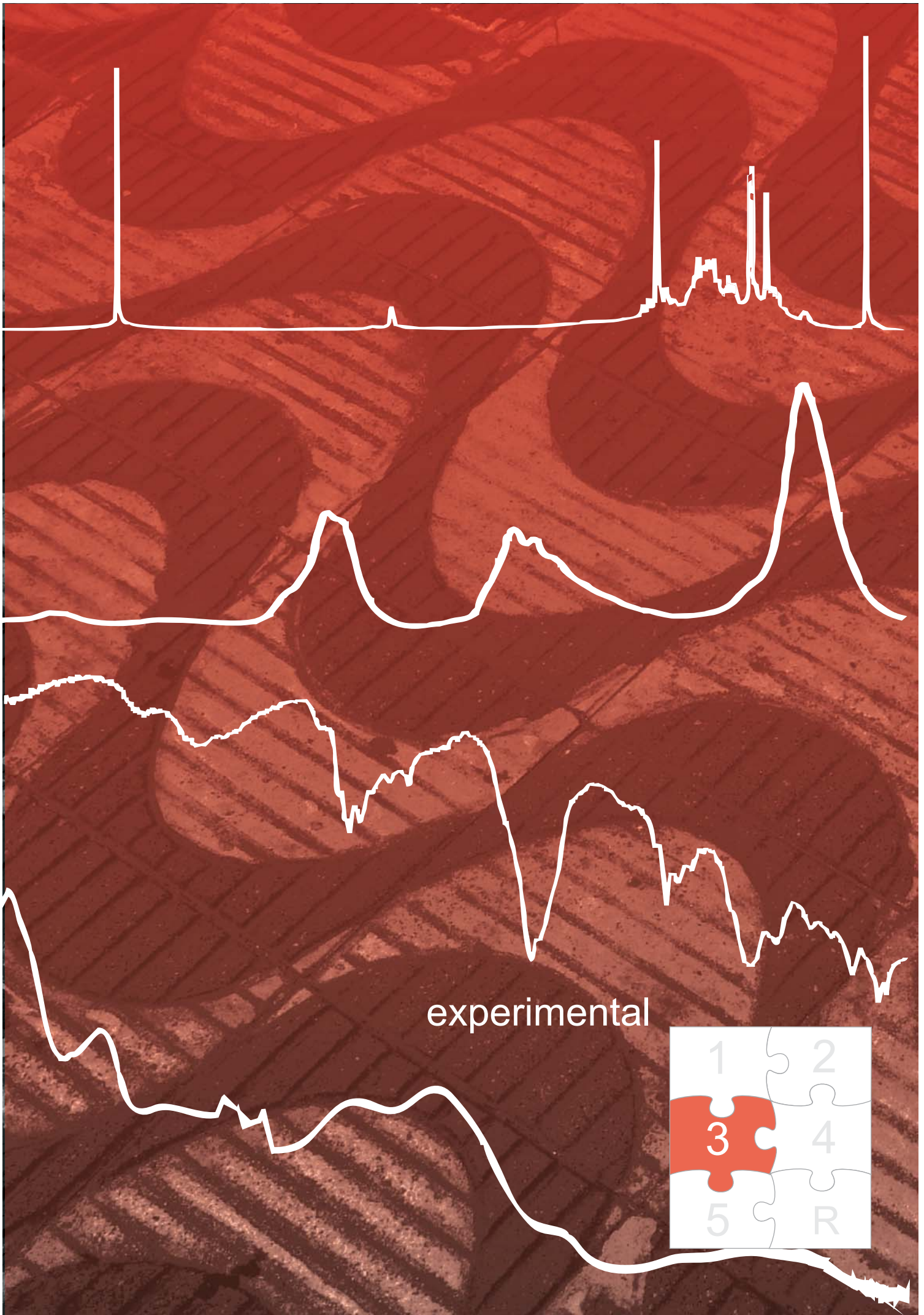


Figura 2.9. Valor do coeficiente antes da limiarização e valor do coeficiente após a limiarização Suave e Rígida.

Vale notar que tanto a limiarização Rígida quanto a Suave atuam somente nos coeficientes de detalhe (DONOHO et al., 1995b), dado que os mesmos contém a maior parte do ruído.

Uma abordagem mais simples chama-se limiarização Bruta (*crude thresholding*), que consiste em zerar todos os coeficientes de detalhe, mantendo apenas os de aproximação. No entanto, tal abordagem pode causar excessiva distorção do sinal.

Outras técnicas são descritas na literatura, sendo a maioria delas derivadas das abordagens propostas por Donoho e colaboradores (1995b). Neste trabalho, restringiremos a investigação às funções de limiarização Rígida e Suave.



experimental



3 Experimental

3.1 Técnicas

Neste experimental é explorado o uso de três técnicas de remoção de ruído instrumental por Transformada *Wavelet*, sendo elas:

- Suave — limiarização Suave e limiar universal;
- Rígida — limiarização Rígida e limiar universal;
- MDL — limiarização Rígida e limiar MDL.

Afim de avaliar a sensibilidade desses métodos de remoção de ruído com respeito à escolha dos filtros *wavelet*, uma investigação foi conduzida com vinte e duas *wavelets*, das famílias Daubechies (1-10), *Symlet* (4-10) e *Coiflet* (1-5). Como mencionado anteriormente, os filtros passa-altas e passa-baixas para DbN, SymN e CoifN têm comprimento 2N, 2N, e 6N, respectivamente. A escolha dessas funções baseou-se no seu uso intensivo na literatura, conforme podemos ver em [WALCZAK e MASSART \(1997\)](#) e [CAI e HARRINGTON \(1998\)](#).

Em todos os experimentos avaliados nesta dissertação, os cálculos foram realizados usando o nível máximo de decomposição que permita um bom funcionamento do banco de filtros *wavelet*, dado pela função *wmaxlev.m*, do Matlab® *Wavelet Toolbox* ([MISITI et al., 1996](#)).

3.2 Experimentos

Para avaliar o desempenho das técnicas aqui abordadas, cinco experimentos foram avaliados:

- **Experimento 1** — Sinais simulados;

- **Experimento 2** — Espectros do medicamento Captopril em comprimidos, obtidos por Espectrometria de Refletância Difusa no Infravermelho Próximo;

- **Experimento 3** — Espectros do princípio ativo Dipirona Sódica em pó, obtidos por Espectrometria de Refletância Difusa no Infravermelho Próximo;

- **Experimento 4** — Espectros de óleo diesel combustível obtidos por Espectrometria no Infravermelho Próximo, medidos em transmitância;

- **Experimento 5** — Espectros de um extrato vegetal medidos em Espectrometria de Ressonância Magnética Nuclear.

3.2.1 Experimento 1 – Dados simulados

Neste exemplo simulado, um sinal sintético $\{x(\lambda)\}$, $\lambda=1, 2, \dots, 1024$, foi gerado pela superposição de seis funções Gaussianas com médias igualmente distribuídas no intervalo de 1-1024. Os pesos do pico e o desvio padrão das Gaussianas foram aleatoriamente gerados na faixa de 2-4 e 0-30, respectivamente. Ao sinal gerado foi adicionado ruído Gaussiano branco de média zero, em três níveis diferentes de desvio-padrão, a saber, 0,01, 0,3 e 1,0. Para cada nível de ruído foram gerados cem espectros ruidosos.

3.2.2 Experimento 2 – Captopril em comprimidos / Espectrometria NIRR

Este experimento envolve a análise do medicamento Captopril na forma farmacêutica comprimido, 25 mg, através da espectrometria de Refletância Difusa no Infravermelho Próximo (NIRR). As análises foram realizadas no Grupo de Instrumentação e Automação (GIA), da Universidade de Campinas (UNICAMP), em um espectrômetro FT-NIR Bomen (ABB-Canadá), modelo MB 160D, equipado com um detector à base de Arseneto de Gálio e Índio (InGaAs). Os espectros foram medidos em um acessório para sólidos, na faixa espectral compreendida entre 3.999 cm^{-1} e 10.000 cm^{-1} , com uma resolução de 2 cm^{-1} . Para a centralização da amostra sob a fonte de radiação foi adaptado ao acessório para sólidos um suporte metálico (Figura 3.1b). Na medida do branco foi utilizado Teflon, 99% *reflective Spectralon* (Labsphere, North Sutton, E.U.A.). A temperatura ambiente foi mantida constante em 24°C e a umidade relativa do ar em 48%.



Figura 3.1 (a) equipamento ABB-BOMEM; (b) suporte para análises de sólidos

Inicialmente, a partir de uma mesma amostra desse medicamento foram registrados cinquenta espectros de varredura individual. Um espectro adicional da amostra, obtido pela promediação de cem varreduras foi registrado para fins de referência nos cálculos de RMSE. Todos os espectros foram obtidos na faixa espectral de 3.999 a 10.000 cm^{-1} e submetidos à técnica EMSC (Correção de Sinal Multiplicativo Estendida, do inglês: *Extended Multiplicative Signal Correction*) (Martens et al, 2003), para redução de possíveis espalhamentos da linha de base.

3.2.3 Experimento 3 – Dipirona sódica em pó / Espectrometria NIRR

O terceiro experimento envolve a análise do princípio ativo Dipirona Sódica, na forma farmacêutica pó, através da espectrometria de Refletância Difusa no Infravermelho Próximo (NIRR). O registro dos espectros foi realizado em um espectrômetro FT-IR Perkin Elmer, modelo Spectrum GX, equipado com um detector a base DTGS (Sulfato de Triglicina Deuterada, do inglês: *Deuterated TriGlicine Sulfate*) e fonte de Halogênio-Tungstênio com envelope de quartzo. O acessório de refletância difusa Perkin Elmer foi acoplado ao equipamento. As análises foram realizadas no Laboratório de Automação e Instrumentação em Química Analítica e Quimiometria (LAQA), da Universidade Federal da Paraíba.



Figura 3.2 Espectrômetro com Transformada de Fourier Perkin Elmer Spectrum GX.

Os espectros foram registrados na faixa espectral de $7710 - 4000 \text{ cm}^{-1}$, em uma resolução de $4,0 \text{ cm}^{-1}$, com temperatura do ambiente constante em $25,5^\circ\text{C}$ e umidade relativa do ar de 42%. O sinal do branco foi registrado com brometo de Potássio (KBr).

A partir da mesma amostra de dipirona, foram registrados cinquenta espectros de varredura individual e um espectro de referência, obtido pela promediação de 256 varreduras, disponibilizando para as etapas subseqüentes duas matrizes, uma matriz de dimensão 50×3712 ; outra de 1×3712 , respectivamente.

3.2.4 Experimento 4 – Óleo Diesel Combustível / Espectrometria NIR

Neste experimento foi reproduzido o experimental realizado em [GALVÃO et al. \(2007\)](#). Espectros NIR de uma única amostra de óleo diesel combustível foram adquiridos usando o mesmo equipamento utilizado no [Experimento 2](#), na faixa espectral de 850 a 1300 nm, utilizando uma cubeta de quartzo de 10 mm. O espectro resultante tem 2560 variáveis espectrais.



Figura 3.3 Espectrômetro FT-NIR BOMEM utilizado para as análises de óleo diesel.

A fim de reduzir possíveis espalhamentos da linha de base que podem comprometer os cálculos de RMSE, os espectros foram submetidos a um procedimento que consiste em remover de cada espectro, a média deles mesmos, conforme realizado em [GALVÃO et al. \(2007\)](#).

Neste experimento, além das três técnicas de remoção de ruído usadas nos outros estudos, foi utilizada a técnica Subótima, proposta na referência supracitada.

3.2.5 Experimento 5 – Extrato vegetal / Espectrometria ^1H RMN

Neste experimento, espectros de ^1H RMN foram obtidos de uma mesma amostra de um diterpeno (o ácido *ent*-7Racetoxitrachilobano-18-óico), do tipo *ent*-trachilobano, isolado das cascas do caule da espécie *Xylopi*a *langsdo*rffiana, conhecida popularmente como “pimenteira da terra”. Detalhes sobre a espécie e a obtenção do extrato vegetal podem ser encontrados em [TAVARES et al., \(2006\)](#). Os espectros RMN foram obtidos em um espectrômetro Bruker DRX-200 (Bruker GmbH, Rheinstetten, Alemanha), operando a 200 MHz para a frequência ^1H , usando CDCl_3 como solvente. As análises foram realizadas no Núcleo Multi-usuário de Caracterização e Análises da UFPB (NUCAL).

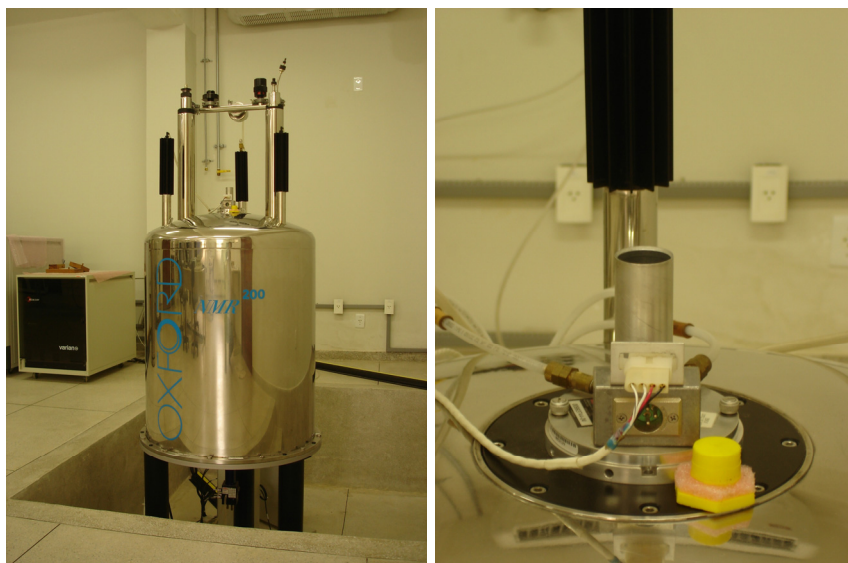


Figura 3.4 (a) Espectrômetro Bruker DRX-200 utilizado para as análises de ^1H RMN; (b) Detalhe do compartimento de inserção da amostra.

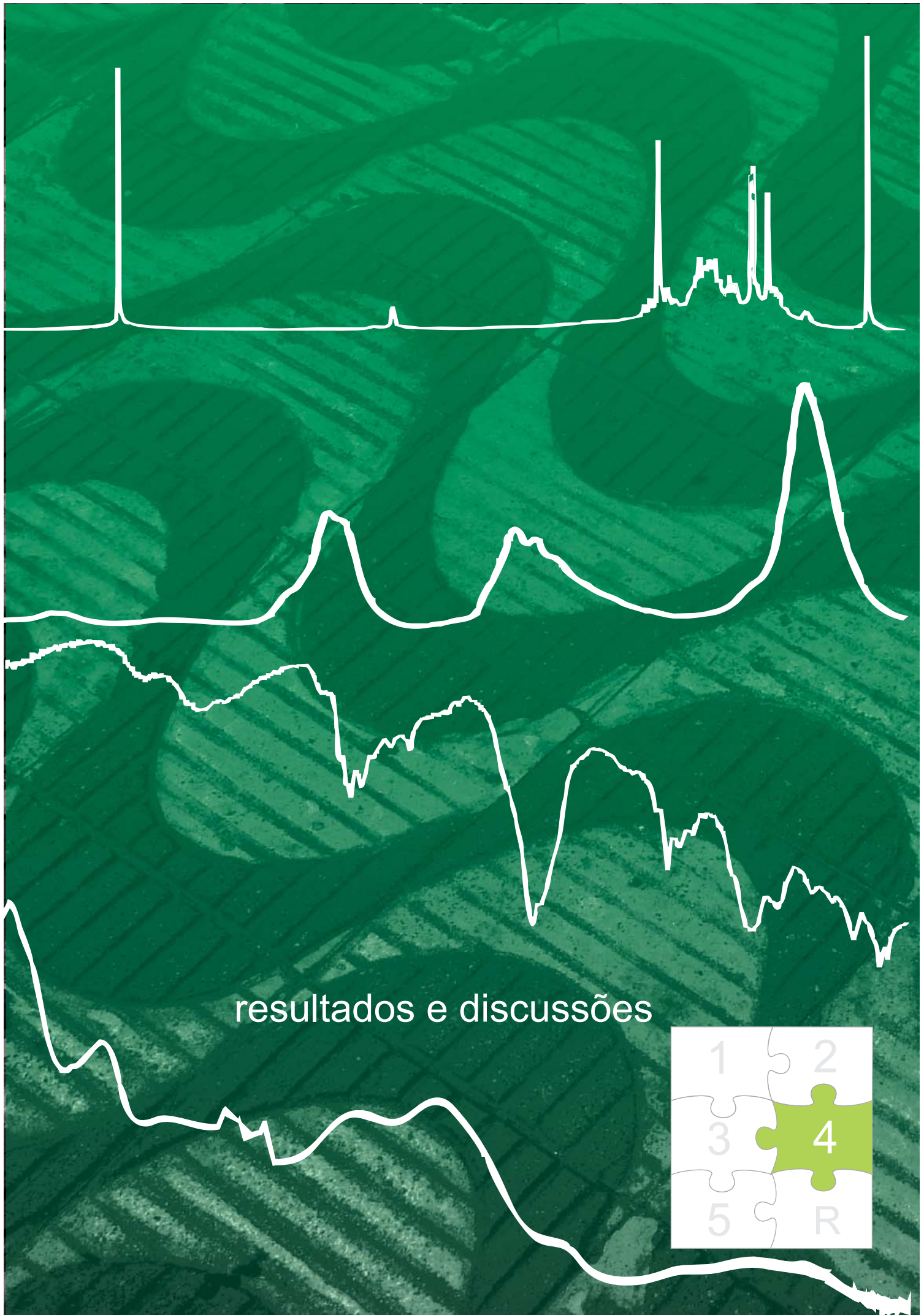
Inicialmente foi obtido um espectro resultante da promediação de dezesseis transientes (ou varreduras). Um espectro promediado por 1024 transientes foi registrado para fins de referência nos cálculos de RMSE. Os dados foram coletados sobre 32 kpts.

3.3 Programas e Pacotes Computacionais

Os algoritmos das três técnicas de remoção de ruído por Transformada *wavelet* foram desenvolvidos usando o pacote *Wavelet Toolbox*, contido no *software* Matlab[®] 6.5 (The Mathworks, Natick, E.U.A.), baseado nos fluxogramas apresentados no [Apêndice](#) da dissertação, bem como os respectivos códigos-fonte.

Todos os cálculos consoantes às técnicas de remoção de ruído foram realizados usando-se o *software* Matlab[®] 6.5, usando os algoritmos acima mencionados. Para a execução da técnica EMSC foi utilizado o *software* Unscrambler 9.7, da CAMO (Trondheim, Noruega).

Um computador portátil Toshiba Satellite[®], equipado com um processador Intel[®] Celeron[®] de 1,33 MHz, com 512 MB de memória, foi utilizado em todo o trabalho.



resultados e discussões



4 Resultados e discussões

Comparar a eficiência das técnicas de remoção de ruído aplicadas a dados reais é uma tarefa complicada, pois o sinal verdadeiro sem o componente ruidoso é desconhecido. Assim, geralmente avaliam-se as técnicas por meio de inferências visuais nos sinais resultantes, muito embora seja uma métrica assaz subjetiva e passível de interferências pessoais. Neste trabalho, visando superar esta dificuldade, adota-se como sinal livre de ruído um espectro obtido pela promediação de um grande número de varreduras, e então, a raiz do erro médio quadrático (RMSE) é calculada entre este espectro e o obtido com um reduzido número de varreduras. O RMSE, então, é a medida da disparidade entre estes espectros, sendo calculado como,

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{\lambda=1}^N [\mathbf{x}f(\lambda) - \mathbf{x}(\lambda)]^2} \quad (4.1)$$

onde \mathbf{N} é o número de pontos do sinal, $\mathbf{x}f$ é o espectro reconstruído e \mathbf{x} o espectro de referência.

4.1 Experimento 1 – Sinais simulados

Para a avaliação do desempenho das técnicas de remoção de ruído aqui estudadas, além do uso de dados reais, é interessante realizar uma comparação usando exemplos simulados, nos quais são conhecidas as propriedades do ruído.

Neste experimento, a partir de um sinal de 1024 pontos, livre de ruído, foram gerados sinais ruidosos pela adição de ruído branco Gaussiano em três níveis de desvio-padrão (SD), cobrindo desde pouco, (SD=0,01), médio (SD=0,3), até elevado grau de ruído (SD=1,0). Para cada um desses níveis, cem sinais foram gerados, e então, submetidos às três técnicas de remoção de ruído por TW, a saber, Rígida, Suave e MDL, usando-se as vinte e duas funções *wavelet* descritas no [Capítulo 3](#).

O erro entre o sinal verdadeiro e os sinais reconstruídos pelas técnicas para cada função *wavelet* foi calculado segundo a [Equação 4.1](#) disposta anteriormente. Os bancos de filtros para todas as técnicas foram construídos com cinco níveis de decomposição, número máximo escolhido pela função *wmaxlev.m*. O comprimento do sinal foi fixado em 1024 pontos a fim de seguir uma das premissas da formulação empregada para a TW, que requer o sinal múltiplo de 2^N , com N sendo o número de níveis de decomposição.

Na [Figura 4.1](#) é mostrado o sinal livre de ruído, bem como os obtidos pela adição de ruído nos três níveis de desvio-padrão.

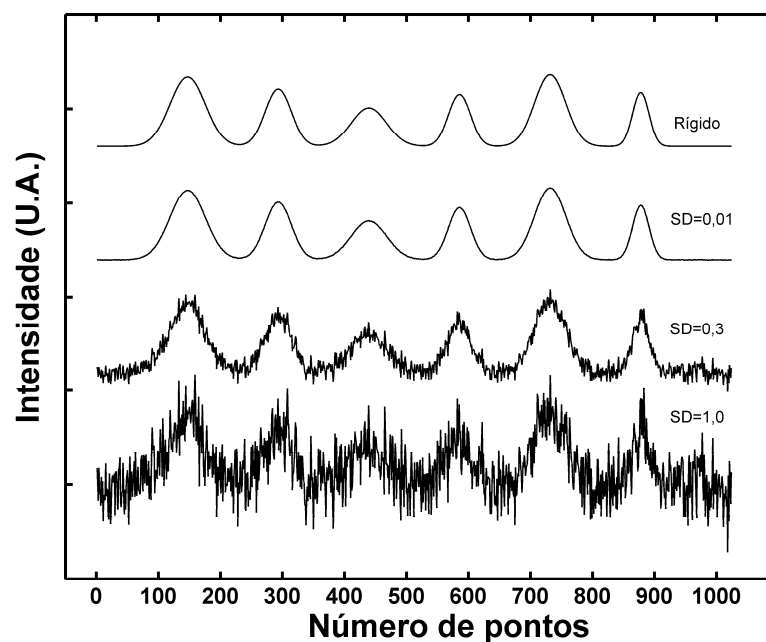


Figura 4.1 Sinal sintético livre de ruído e os sinais gerados pela adição de ruído em três níveis de desvio-padrão.

4.1.1 Ruído com desvio-padrão de 0,01

Para os sinais contaminados com pouco ruído ($SD= 0,01$), todas as técnicas apresentaram boa eficiência, removendo satisfatoriamente o ruído presente no sinal bruto, sem introduzir distorção perceptível. O resultado pode ser visualizado na [Figura 4.2](#), onde são apresentados os sinais reconstruídos pelas técnicas MDL, Rígida e Suave, o sinal ruidoso e o de referência, livre de ruído.

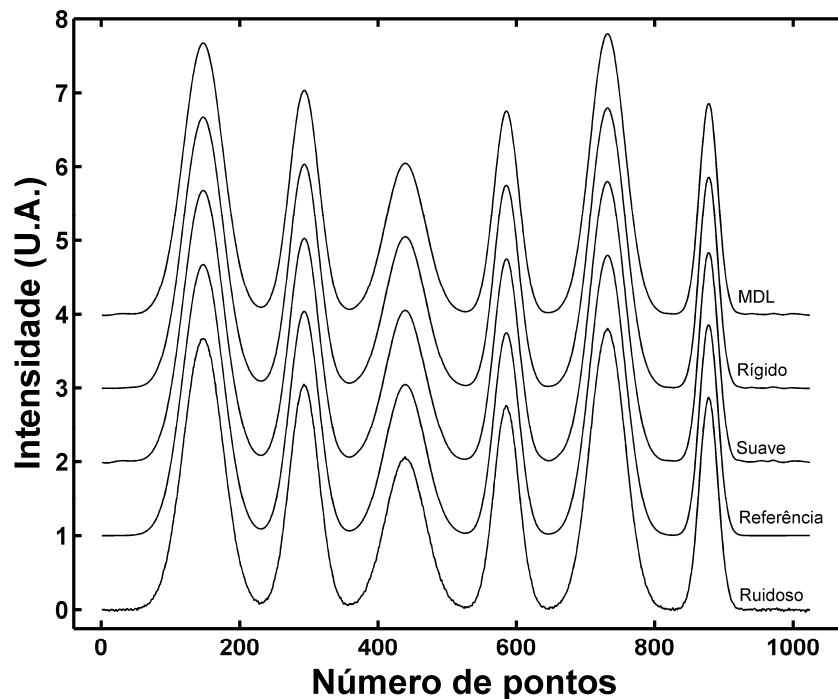


Figura 4.2 Sinais reconstruídos pela TW usando as técnicas MDL, Rígida e Suave, sinal de referência e sinal ruidoso ($SD=0,01$).

Para avaliar o comportamento das três técnicas com relação ao RMSE obtido entre o sinal de referência e o sinal reconstruído para cada uma das *wavelets*, um total de cem sinais ruidosos foram gerados aleatoriamente e submetidos à remoção de ruído, sendo calculado o RMSE médio obtido para cada *wavelet*. Os valores resultantes estão apresentados graficamente na [Figura 4.3](#). Como podemos notar, a técnica Suave ficou aquém das técnicas Rígida e MDL, as quais obtiveram erros muito próximos, quiçá por utilizarem a mesma função de limiarização (limiarização Rígida). De fato, a maior agressividade da função de limiarização Suave trouxe pequenas distorções nas extremidades do sinal, que contribuíram para a elevação dos RMSE's para a técnica Suave.

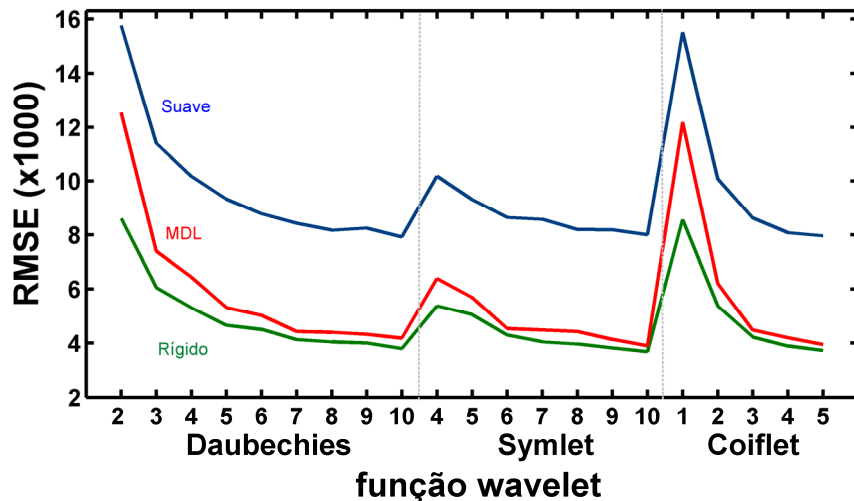


Figura 4.3 Curva do RMSE médio x função *wavelet* para as três técnicas de remoção de ruído por TW em um total de cem sinais gerados com ruído (SD=0,01).

Como discutido no [Capítulo 2](#), o método MDL pode selecionar com um critério definido, tanto o limiar quanto o tipo de filtro. O critério adotado pelo MDL é a minimização da sua função-custo, que representa o interesse conflitante de dois termos, um relacionado à parcimônia do resultado, outro buscando a diminuição do erro da reconstrução do sinal. A minimização dessa função-custo, isto é, a obtenção do menor valor de MDL, representa um compromisso entre estes dois objetivos. Assim, entre a biblioteca de *wavelets*, a que resultar no menor valor de MDL, será a escolhida. A validade dessa seleção foi avaliada pelos RMSE's obtidos entre o sinal de referência e o sinal reconstruído pelo MDL através de uma biblioteca de vinte e duas *wavelets*, para cada um dos cem sinais gerados. Dessa forma, podemos julgar se a minimização da função custo do MDL para cada sinal pode resultar em baixos RMSE's.

Através da [Figura 4.4](#), podemos observar que dos cem sinais gerados, em 14 deles o MDL escolheu o filtro que forneceu o menor RMSE. Em 10 casos a função escolhida resultou no 2º menor erro, e em 17 no 3º menor erro. Em 65% dos casos, o MDL selecionou um filtro cujo RMSE estava entre os cinco menores erros entre os vinte de dois filtros. Em 31% dos casos (barras cinza escuro) o MDL escolheu o filtro que forneceu o menor RMSE ou um que não tinha diferenças significativas com este, segundo um teste-*F* com 95% de confiança.

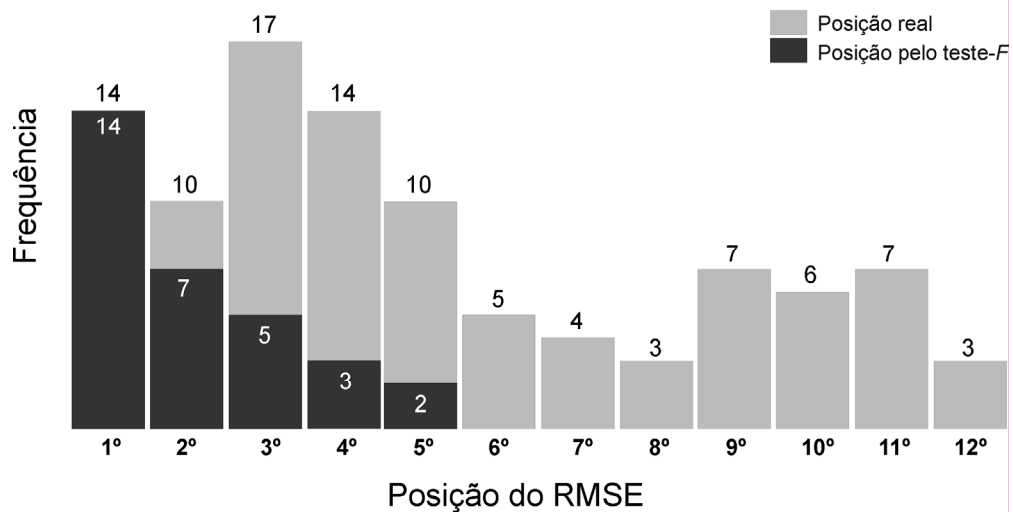


Figura 4.4 Gráfico de barras apresentando a posição da *wavelet* selecionada pelo MDL para cada um dos cem sinais gerados, com relação ao RMSE obtido.

Na Figura 4.5, abaixo, são mostradas as curvas de RMSE x função *wavelet* obtidas nos casos onde o MDL obteve o seu melhor e pior desempenho em relação à escolha adequada da *wavelet*.

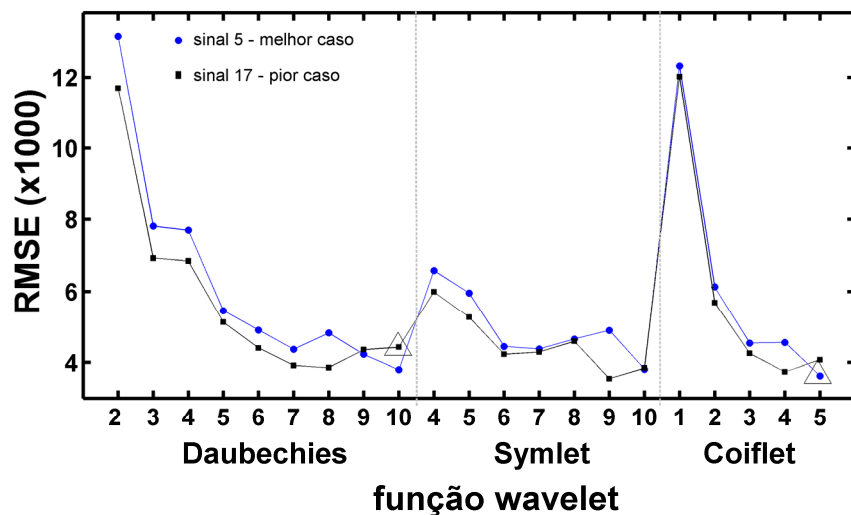


Figura 4.5 Curva do RMSE x função *wavelet* para a técnica MDL dos sinais número 5 e 17.

Para o espectro 5, um caso em que o MDL obteve desempenho ótimo, a *wavelet* selecionada (*Coiflet* 5) obteve o menor RMSE entre as 22 funções, enquanto que para o espectro 17, a *wavelet* Daubechies 10, selecionada como ótima, resultou no 12º menor erro. Apesar da baixa colocação em termos de RMSE

da *wavelet* selecionada pelo MDL para esse sinal, um erro relativamente baixo foi obtido, mostrando a eficiência da técnica. Com efeito, além de remover satisfatoriamente o componente ruidoso com erros comparáveis aos obtidos pela técnica Rígida, o MDL possui a vantagem da escolha automática da *wavelet* que deve ser utilizada, sem perdas significativas no RMSE.

O MDL mostrou considerável robustez na escolha da *wavelet*, selecionando apenas cinco *wavelets* diferentes em todos os cem sinais (Figura 4.6), sendo que em 91 % dos casos foram selecionadas apenas duas *wavelets*, Daubechies 10 (70%) e *Coiflet* 5 (21%).

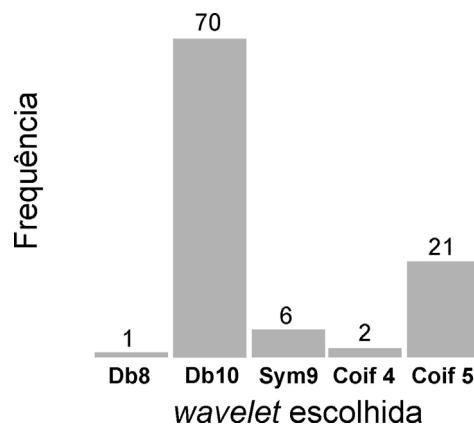


Figura 4.6 Frequência de seleção da mesma função *wavelet* pelo MDL nos cem sinais sintéticos.

Em relação à capacidade de compressão dos dados, as duas estratégias de limiar (MDL e universal) mostraram eficiências comparáveis, selecionando um número muito pequeno de coeficientes *wavelet*, conforme pode ser visto na Figura 4.7, onde dois gráficos de barras apresentam o número de coeficientes obtido pelas vinte e duas funções para os espectros 5 e 17 estudados anteriormente. Apesar do comportamento semelhante das duas estratégias, podemos perceber uma leve vantagem da técnica MDL para a maioria das funções *wavelets*, sendo que em três casos (Db1, Db2 e Coif1) as duas técnicas mostraram diferenças mais expressivas. Uma explicação para tal discrepância pode ser a maior sensibilidade do limiar universal frente a filtros de comprimento curto. De fato, os filtros Db1, Db2 e Coif1 são os menores entre os vinte e dois filtros *wavelet*, possuindo comprimento 2, 4 e 6, respectivamente. Para os dois sinais, conforme o comprimento do filtro aumenta,

podemos notar que as diferenças no número de coeficientes selecionados entre os dois limiares diminui.

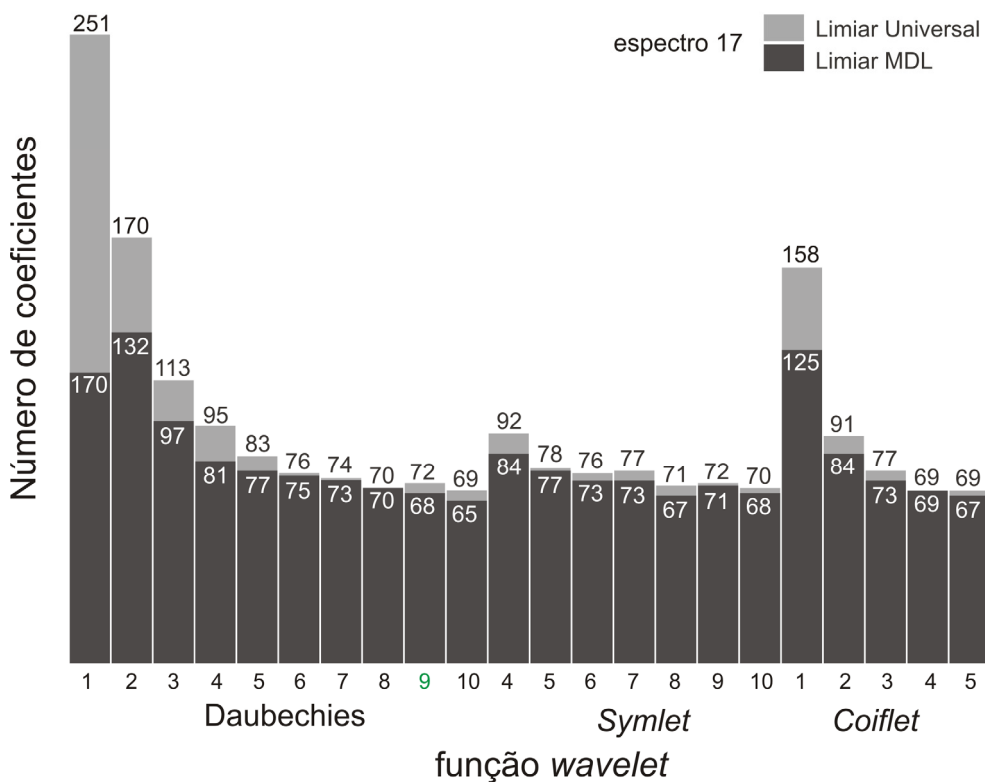
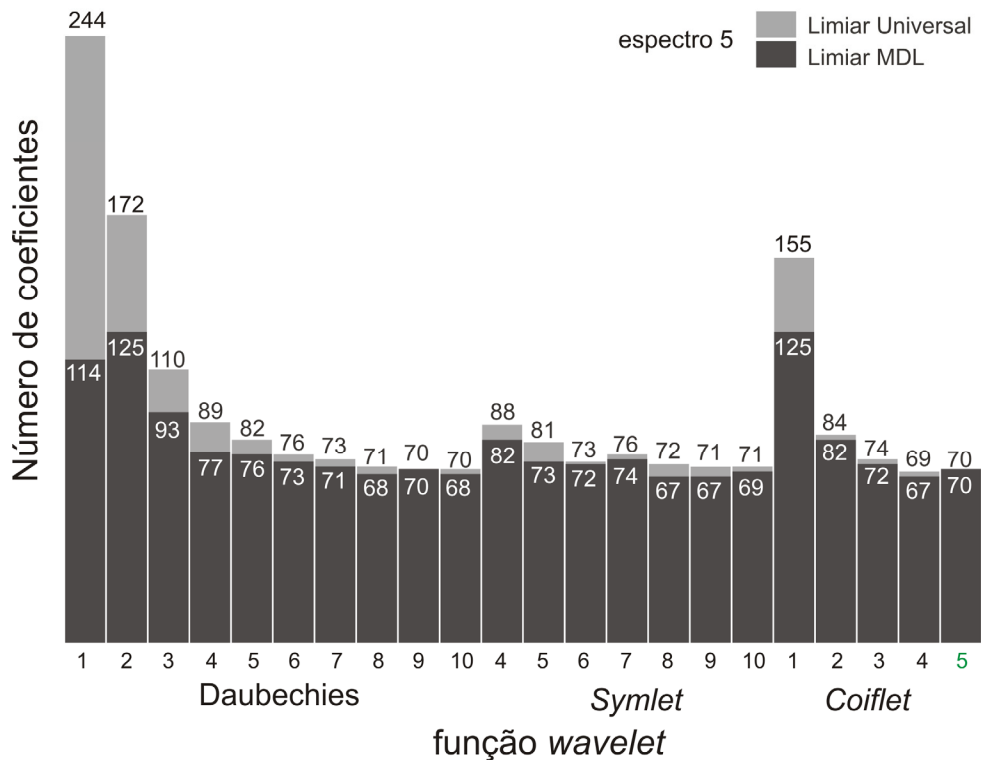


Figura 4.7 Gráfico de barras com o número de coeficientes selecionados pelo limiar universal (barras cinza) e MDL (barras pretas) através das 22 wavelets para (a) o espectro 5 e (b) o espectro 17. A wavelet selecionada pelo MDL é escrita em verde.

4.1.2 Ruído com desvio-padrão de 0,3

Igualmente aos sinais do exemplo anterior, todas as três técnicas mostraram desempenhos semelhantes, removendo todo o componente ruidoso presente no sinal bruto, porém, mantendo distorções em algumas regiões, todas causadas pelo ruído proeminente deste sinal, conforme podemos ver na [Figura 4.8](#), abaixo.

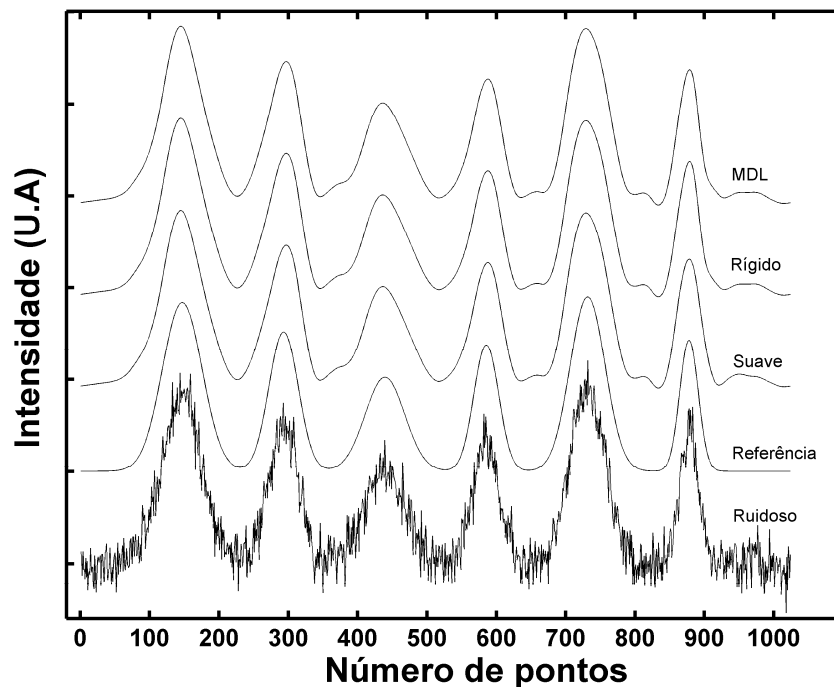


Figura 4.8 De cima para baixo, sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso (SD=0,3).

Em termos de RMSE, conforme pode ser visto na [Figura 4.9](#), o padrão de eficiência das técnicas se manteve, com a técnica Rígida apresentando os menores erros, seguida pelo MDL e pela técnica Suave. Esta, por sua vez apresentou melhora em relação aos resultados anteriores, obtendo erros próximos aos obtidos pelas outras técnicas.

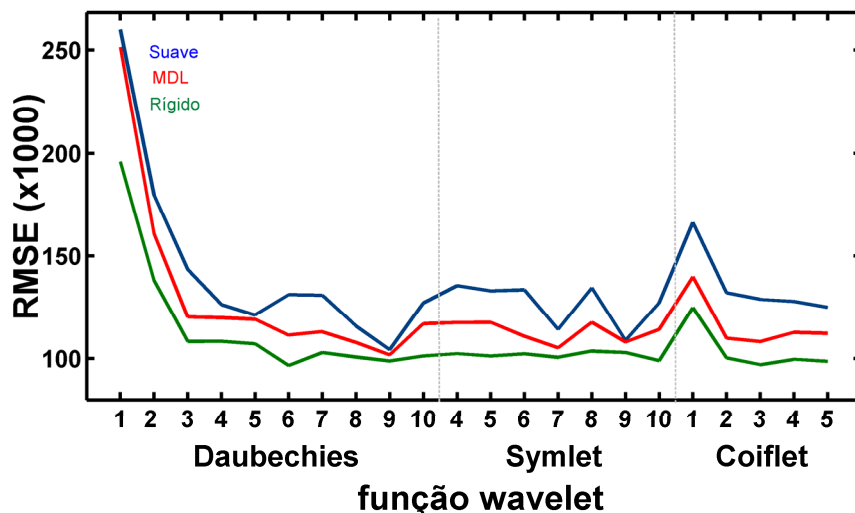


Figura 4.9 Curva do RMSE médio x função *wavelet* para as três técnicas de remoção de ruído por Transformada *wavelet* em um total de cem sinais gerados com ruído (SD=0,3).

A eficiência do MDL na seleção da função ótima foi avaliada com base nos RMSE's obtidos pelas vinte e duas *wavelets* em um total de cem sinais gerados.

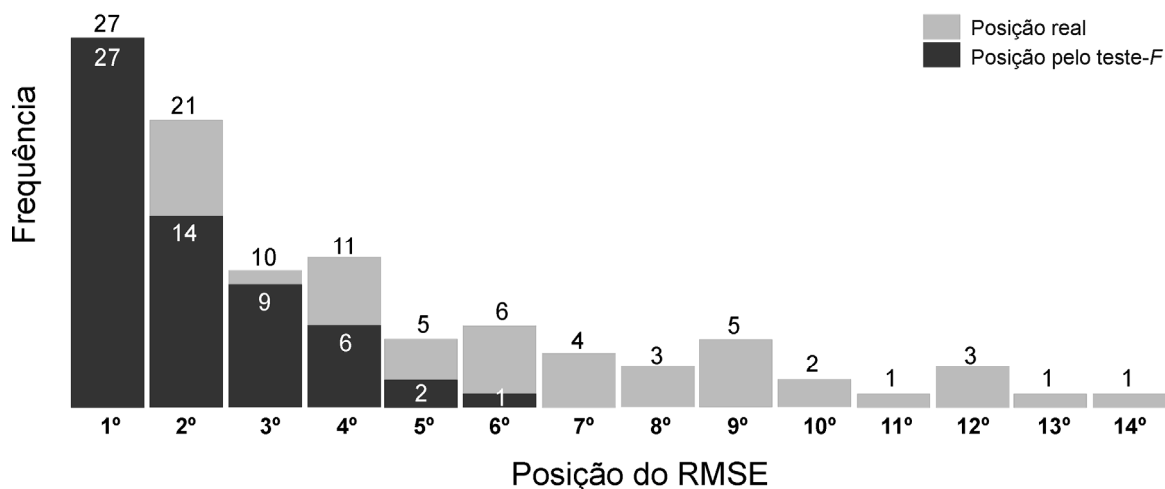


Figura 4.10 Gráfico de barras apresentando a posição da *wavelet* selecionada pelo MDL para cada um dos cem sinais gerados, com relação ao RMSE obtido.

Através do gráfico de barras da [Figura 4.10](#), podemos notar que o MDL obteve bons resultados, selecionando em 27% dos casos a *wavelet* que resultou no menor RMSE. Em 74% dos sinais, o MDL escolheu *wavelets* que obtiveram até o quinto menor erro. Através de um teste-*F* com 95% de confiança, podemos concluir que em 59% dos casos o MDL selecionou a *wavelet* que resultou no menor RMSE ou uma cujo erro não tem diferença significativa com o menor (barras cinza-escuro).

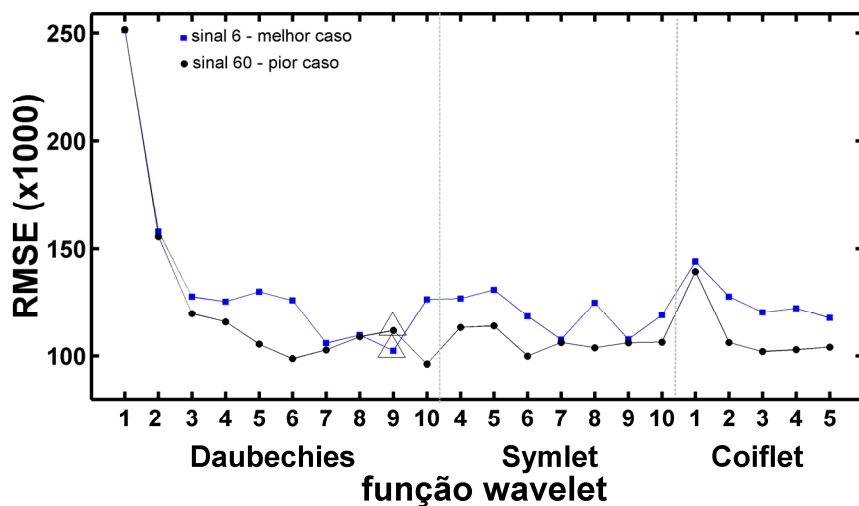


Figura 4.11 Curva do RMSE x função *wavelet* para a técnica MDL dos sinais número 6 e 60.

Entre os cem sinais gerados, o MDL apresentou sua pior eficiência na seleção da *wavelet* do sinal número 60, onde a *wavelet* Daubechies 9 foi escolhida, obtendo o 14º menor erro. Porém, essa mesma função forneceu o menor erro entre as vinte e duas funções quando selecionada pelo MDL para o sinal 6 (Figura 4.11). Vale ressaltar que o MDL mostrou considerável robustez na seleção da *wavelet*, escolhendo em 74% dos casos a mesma função (Db9), também escolhida maciçamente nos sinais menos ruidosos do experimento anterior. O gráfico de barras contendo a frequência de seleção da *wavelet* em cem sinais pode ser visto na Figura 4.12.

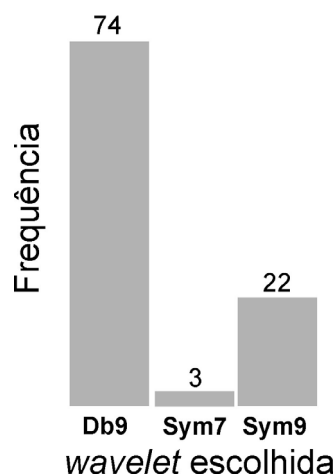


Figura 4.12 Frequência de seleção da mesma função *wavelet* pelo MDL nos cem sinais sintéticos.

Ambos os limiares MDL e Universal apresentaram eficiência na compressão dos sinais, sendo capazes de reconstruir o sinal usando apenas uma pequena porção dos coeficientes (menos de 10%), como pode ser visto na [Figura 4.13](#), onde é apresentado o número de coeficientes selecionados para cada *wavelet* para os sinais 6 e 60, discutidos anteriormente.

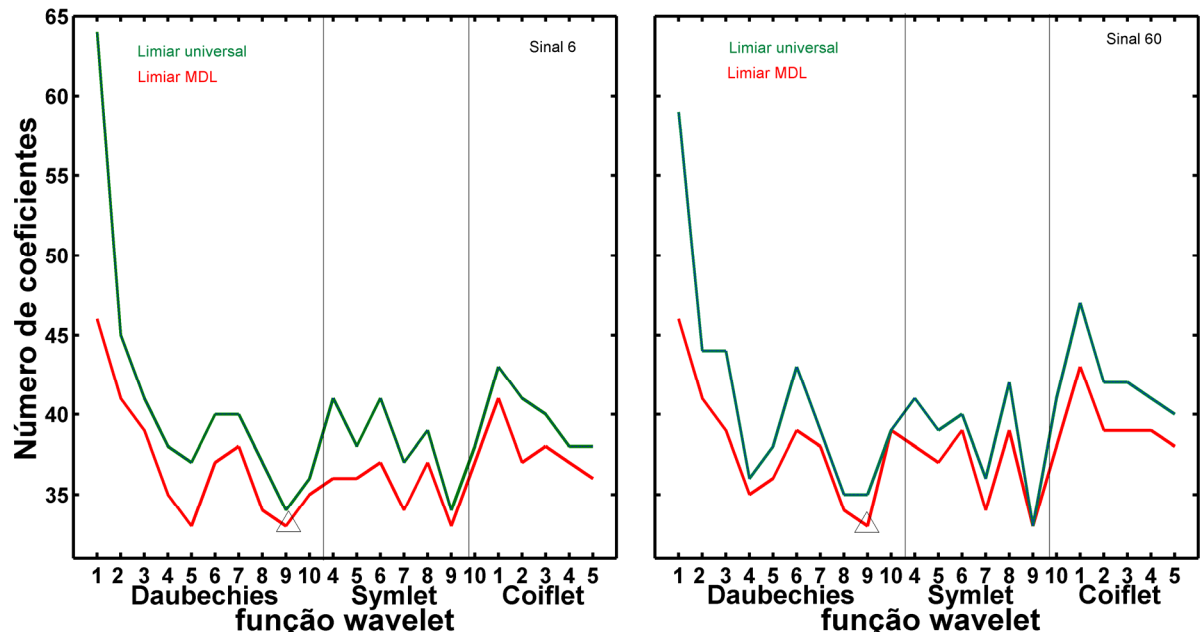


Figura 4.13 Curva do RMSE x função *wavelet* da técnica MDL para os sinais número 6 e 60.

O MDL produziu os melhores resultados, selecionando menos coeficientes para todas as *wavelets* no sinal 6, e empatando em duas no sinal 60. Em ambos os sinais, a *wavelet* escolhida pelo MDL (Daubechies 9) resultou no menor número de coeficientes. Um compromisso entre os dois termos conflitantes da função-custo MDL foi obtido, visto que no sinal 6 foi possível a seleção de um pequeno número de coeficientes com um baixo RMSE. Embora para o sinal 60 a *wavelet* escolhida não tenha obtido o menor RMSE, o baixo número de coeficientes justifica seu uso, principalmente considerando-se que o erro obtido é aceitavelmente baixo.

4.1.3 Ruído com desvio-padrão de 1,0

Na [Figura 4.14](#) são mostrados os sinais reconstruídos para cada técnica, o sinal bruto e o livre de ruído.

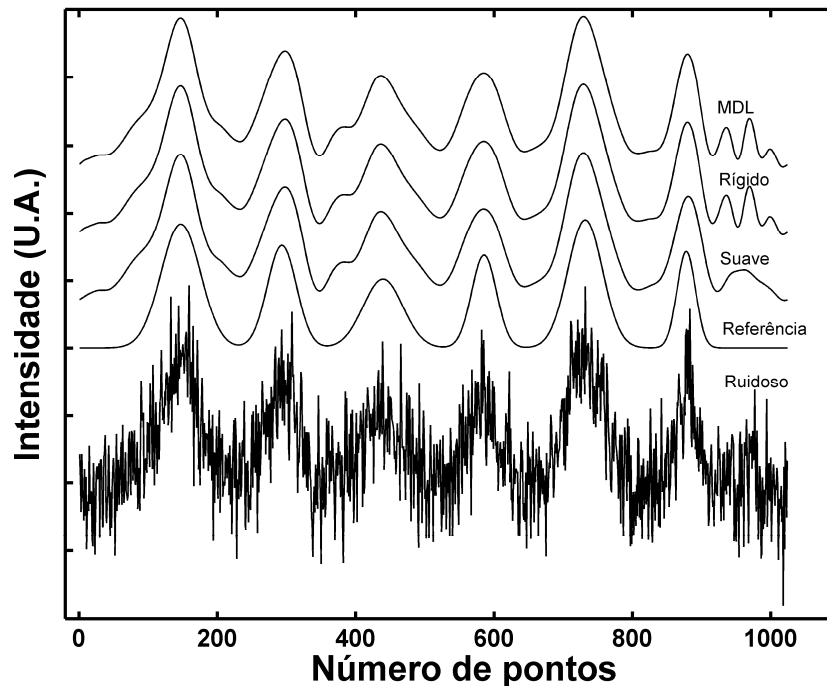


Figura 4.14 Sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso (SD=1,0).

O comportamento observado nos casos anteriores se manteve para os sinais extremamente ruidosos deste exemplo (SD=1,0). As técnicas mostraram resultados similares quando comparadas visualmente e em termos de RMSE. O alto desvio-padrão do ruído causou distorções em várias regiões do sinal bruto, as quais não puderam ser corrigidas, embora o ruído tenha sido satisfatoriamente removido, independentemente da técnica utilizada, conforme observado na [Figura 4.14](#). Novamente, a técnica Suave mostrou maior rigor, o que pode ser constatado na parte final do seu sinal reconstruído, a qual apresenta uma única banda, ao invés das três bandas presentes nos sinais reconstruídos pelas técnicas Rígida e MDL. Isso pode ser explicado pelo rigor na atenuação dos coeficientes realizada pela técnica Suave, cuja diminuição da magnitude dos coeficientes acima do limiar promoveu a união das três bandas em apenas uma. Porém, como essas bandas são provenientes de deformações do sinal bruto, não houve perda de informação útil,

servindo-nos apenas de exemplo sobre a influência da manipulação dos coeficientes.

Em termos de RMSE, as três técnicas apresentaram a maior similaridade observada entre os experimentos simulados ora estudados. A técnica Rígida prevaleceu em boa parte das *wavelets*, porém, foi suplantada em algumas pelas técnicas MDL e Suave (Figura 4.15).

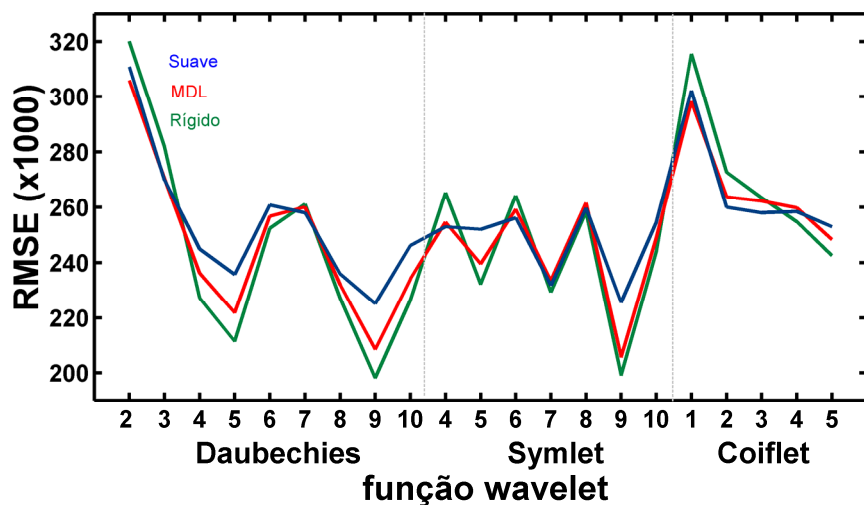


Figura 4.15 Curva do RMSE médio x função *wavelet* para as três técnicas de remoção de ruído por Transformada *wavelet* em um total de cem sinais gerados com ruído (SD=1,0).

Em relação à escolha da *wavelet* pelo MDL, pela análise do gráfico de barras da Figura 4.16, temos que em 75% dos casos o MDL selecionou um filtro cujo RMSE estava entre os cinco menores erros, dos vinte e dois filtros analisados. Em 40% dos casos, a função escolhida pelo MDL forneceu o menor RMSE ou um que não tinha diferenças significativas com o menor, segundo um teste-*F* com 95% de confiança.

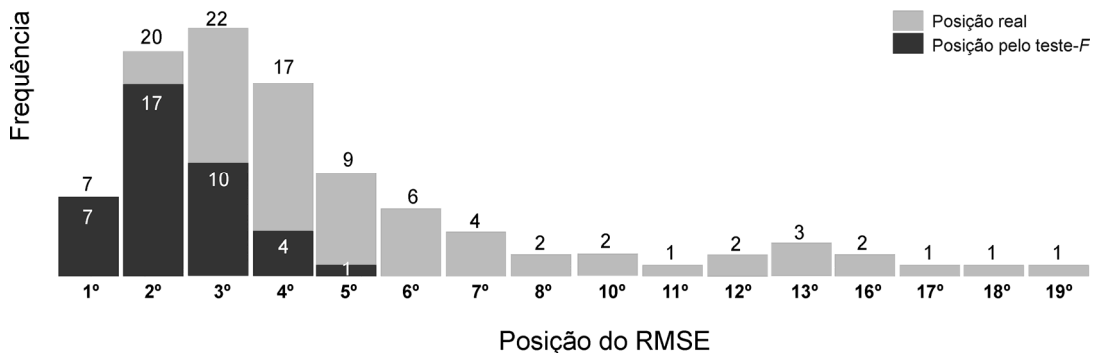


Figura 4.16 Gráfico de barras apresentando a posição da *wavelet* selecionada pelo MDL para cada um dos cem sinais gerados, com relação ao RMSE obtido.

Como podemos ver pelas Figuras 4.16 e 4.17, o caso em que o MDL obteve seu pior desempenho na seleção do filtro *wavelet* foi o sinal 20, com a seleção do filtro Daubechies 5, obtendo o 19º pior erro. O melhor desempenho pôde ser observado em sete casos, onde os filtros selecionados pelo MDL forneceram o mais baixo RMSE entre os vinte e dois filtros. Um exemplo é o sinal 11, no qual o MDL selecionou a *wavelet* Daubechies 9.

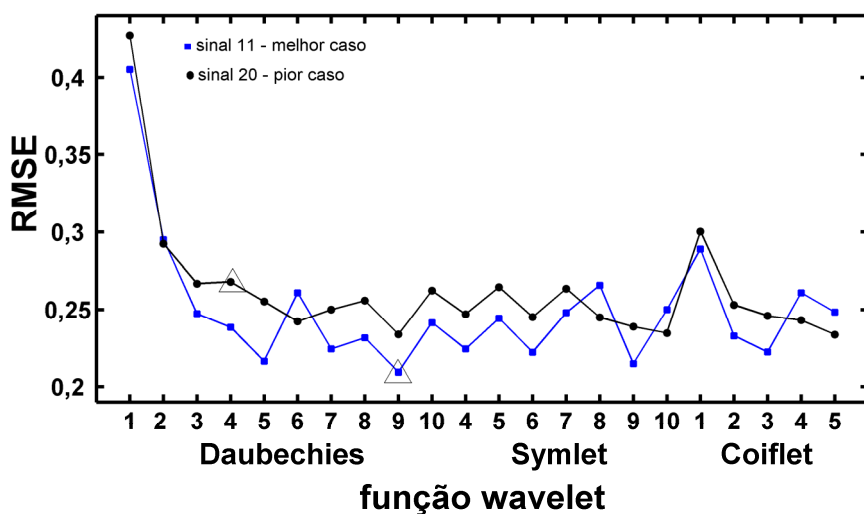


Figura 4.17 Curva do RMSE x função *wavelet* para a técnica MDL dos sinais número 11 e 20.

Apesar de o MDL ter selecionado apenas duas *wavelets* (Daubechies 9 e *Symlet* 9) em 73% dos casos (Figura 4.18), o universo de *wavelets* selecionadas para este exemplo foi 7, um pouco maior que nos casos anteriores (5 e 3, respectivamente). Além disso, as duas *wavelets* selecionadas confirmaram uma tendência de seleção em sinais mais ruidosos, visto terem sido intensamente escolhidas nos dois exemplos simulados mais ruidosos, mostrando a robustez do MDL na tarefa de seleção da *wavelet* mais adequada à natureza dos dados.

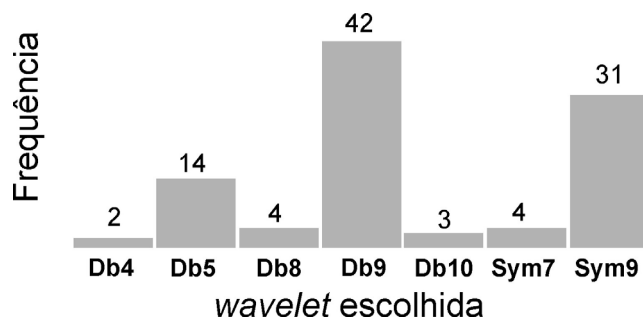


Figura 4.18 Frequência de seleção da mesma função *wavelet* pelo MDL nos cem sinais sintéticos.

O número de coeficientes *wavelet* selecionados pelas duas técnicas de limiar foi o menor observado neste estudo com sinais simulados. O aumento no nível de ruído pode ter diminuído o número de coeficientes realmente relacionados ao componente do sinal. Na [Figura 4.19](#) podem ser observadas as curvas do número de coeficientes selecionados em relação à função *wavelet* para os dois sinais discutidos acima. Com exceção de duas *wavelets* no sinal 11, todas as outras *wavelets*, deste e do sinal 20, selecionaram um número fixo de coeficientes (33), enquanto o limiar universal oscilou entre 32 e 37, obtendo menos coeficientes que o limiar MDL.

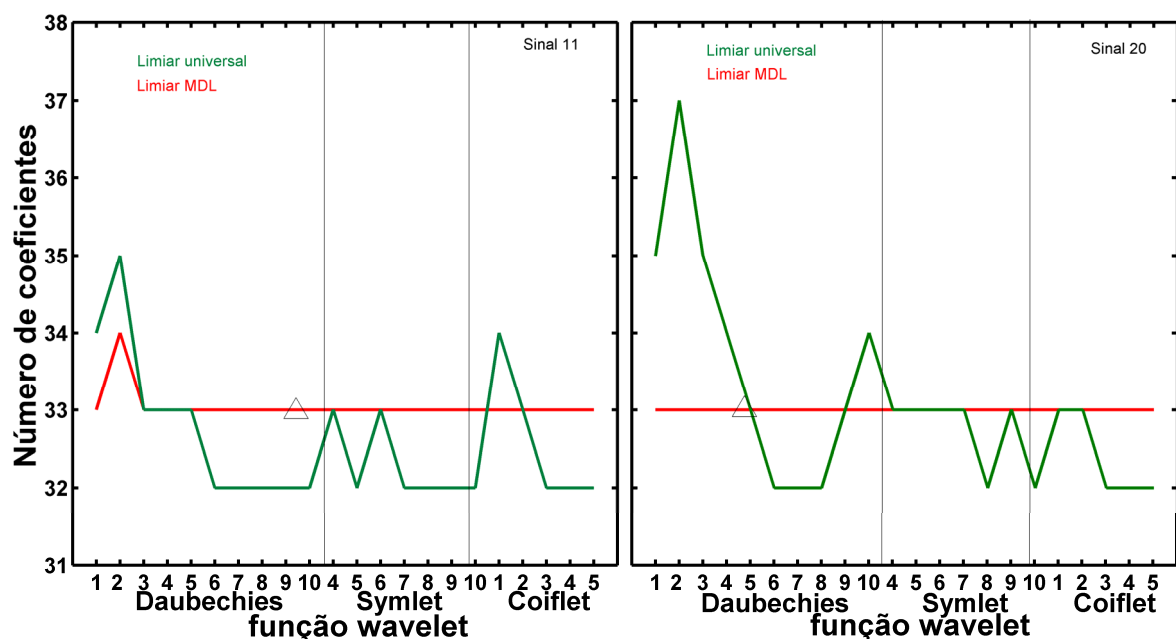


Figura 4.19 Curva do RMSE x função *wavelet* da técnica MDL para os sinais número 11 e 20. A *wavelet* escolhida é marcada por um triângulo.

4.1.4 Análise dos diferentes níveis de ruído

Pela análise do comportamento das três técnicas na remoção de ruído em sinais sintéticos gerados em três níveis de ruído homoscedástico, podemos concluir que todas se mostraram eficientes, sendo capazes de remover satisfatoriamente o componente ruidoso visível. A técnica Rígida apresentou os melhores resultados, tanto em termos das características qualitativas do sinal como em relação ao RMSE, sendo seguida pelo MDL, que obteve desempenho similar. A técnica Suave obteve resultados menos satisfatórios que as outras técnicas, porém, com uma melhora considerável à medida que se aumentou o nível de ruído, chegando a resultados

similares às outras técnicas para algumas *wavelets*. A técnica Suave utiliza a função de limiarização Suave, diferentemente das outras duas técnicas, que fazem uso da limiarização Rígida. O maior rigor da limiarização Suave, devido à atenuação dos coeficientes com amplitudes acima do limiar, foi o responsável pelo fraco desempenho dessa técnica.

Em todos os casos, os maiores RMSE's foram observados nas *wavelets* cujos comprimentos dos filtros são menores. O mesmo comportamento se deu em relação ao número de coeficientes necessários para a reconstrução do sinal, com uma menor compressão dos dados obtida nessas *wavelets*. De fato, as *wavelets* Daubechies 1 a 3, e a *coiflet* 1, de comprimento 2, 4, 6 e 6, respectivamente, obtiveram sempre os maiores erros e um elevado número de coeficientes. Esse comportamento é devido ao fato que em filtros de curto comprimento as “janelas” são menores, por conseguinte, mas janelas são necessárias para analisar todo o sinal, com conseqüente aumento no número de coeficientes selecionados. Além disso, a suavidade da *wavelet* aumenta conforme se aumenta o comprimento do filtro (ROSAS-OREA et al., 2005), permitindo uma melhor aproximação do sinal.

A parcimônia na seleção de coeficientes apresentou uma tendência relacionada ao nível de ruído, com a seleção de um número menor de coeficientes conforme se aumentou o nível do ruído, como conseqüência da diminuição no número de coeficientes relacionados ao componente do sinal propriamente dito. Todas as três técnicas apresentaram uma representação mais compacta dos sinais aqui estudados.

Em relação à escolha da *wavelet*, o MDL manteve uma eficiência constante em todos os casos, selecionando sempre filtros que forneceram baixos RMSE's. Além disso, apresentou considerável robustez, selecionando a *wavelet* Sym9 em todos os níveis de ruído estudados, principalmente nos dois últimos, onde a *wavelet* Db9 também teve ótimo desempenho, sendo majoritariamente escolhida. Vale notar que ambas possuem filtro com comprimento 18, o segundo maior entre as vinte e duas *wavelets*.

4.2 Experimento 2 – Captopril / Espectrometria NIRR

Comumente, espectros de refletância difusa sofrem mais com a presença de ruído que os espectros de absorbância e transmitância, principalmente devido ao forte espalhamento da radiação. Neste experimento, a amostra (um comprimido) apresenta compactação uniforme dos grânulos, possibilitando a aquisição de um espectro menos suscetível aos espalhamentos de luz. Dessa forma, os espalhamentos não comprometeram de maneira acentuada os espectros, conforme podemos ver através da [Figura 4.20](#), que apresenta um espectro bruto de varredura unitária na faixa espectral de 3.999 a 10.000 cm^{-1} .

Apesar de o espectro ser resultado de apenas uma varredura, com exceção de algumas porções, não apresenta ruído considerável. De fato, a intensidade do ruído não é uniforme em toda a faixa espectral, sendo mais intenso na região conhecida como *short-NIR*, o que pode ser explicado pela degradação da sensibilidade do detector à base de InGaAs.

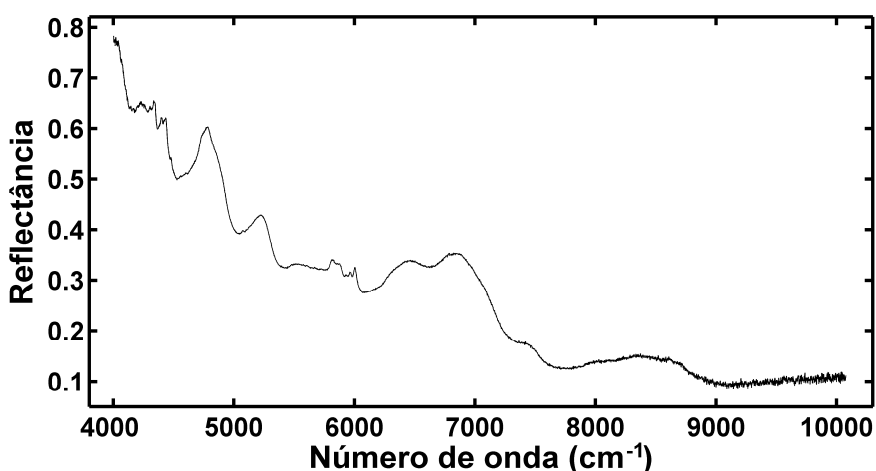


Figura 4.20 Espectro de refletância de um comprimido do princípio-ativo captopril na região de 3.999-10.000 cm^{-1} .

Essa heteroscedasticidade no sinal pode ser evidenciada também pela inspeção dos coeficientes *wavelet* do primeiro nível de detalhe ([WALCZAK e MASSART, 1997b](#)). Conforme podemos ver na [Figura 4.21](#), a frequência nos primeiros quinhentos e últimos mil coeficientes apresenta alta variabilidade na intensidade, o que pode reduzir a eficiência da limiarização *wavelet*, devido à

dificuldade em selecionar um limiar que contemple uniformemente todos os coeficientes *wavelet*.

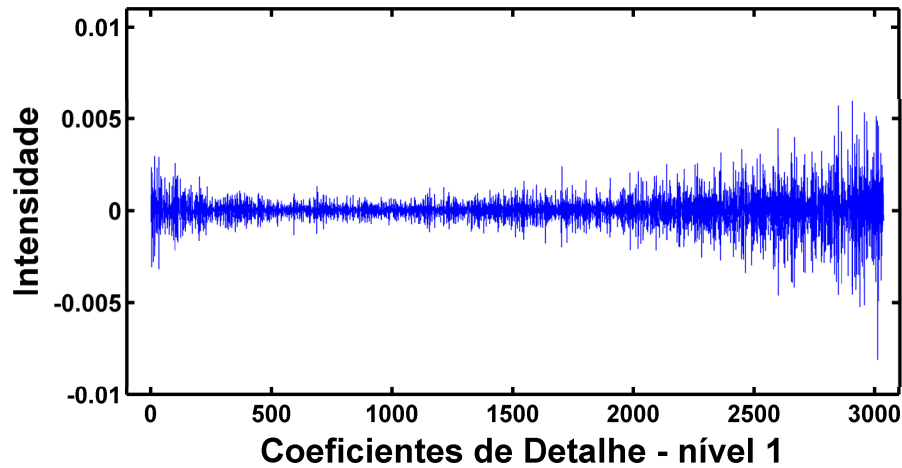


Figura 4.21 Coeficientes de detalhe do primeiro nível de decomposição de um espectro de captopril.

Assim, decidiu-se pela exclusão das regiões acima citadas, de forma também a satisfazer um dos critérios necessários ao adequado funcionamento do banco de filtros *wavelet* na formulação adotada, que requer sinais com comprimento múltiplo de 2^N .

Para o cálculo do nível de decomposição máximo em que o banco de filtros possa funcionar adequadamente, utilizamos a função *wmaxlev.m*, do *Wavelet Toolbox*, com uma estimativa de 4.500 pontos, a qual resultou em sete níveis. Visto que $2^7=128$, o múltiplo desse resultado mais próximo de 4500 é 4480. Assim, a faixa de comprimento de onda foi reduzida pela exclusão das últimas 20 variáveis, resultando em uma região de trabalho entre 4610 cm^{-1} e 8.928 cm^{-1} , disponibilizando para os cálculos subseqüentes uma matriz de tamanho 51×4480 . Na [Figura 4.22](#) é mostrado o espectro de referência e um espectro de varredura individual, utilizados nos cálculos subseqüentes.

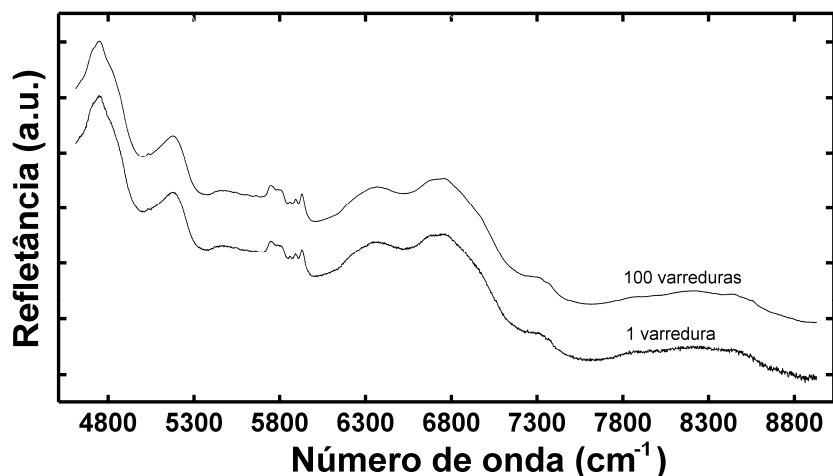


Figura 4.22 Espectro de referência (100 varreduras) e espectro ruidoso (1 varredura) de um comprimido do princípio-ativo captopril na região de 4.610-8.928 cm^{-1} .

Na [Figura 4.23a](#), podemos visualizar o espalhamento da linha de base nos espectros, principalmente entre o espectro de referência (verde) e os demais, o que pode trazer influências negativas sobre o cálculo do RMSE. Visando contornar esse problema, todos os espectros foram submetidos ao emprego da técnica EMSC, com o resultado do tratamento podendo ser visto na [Figura 4.23b](#).

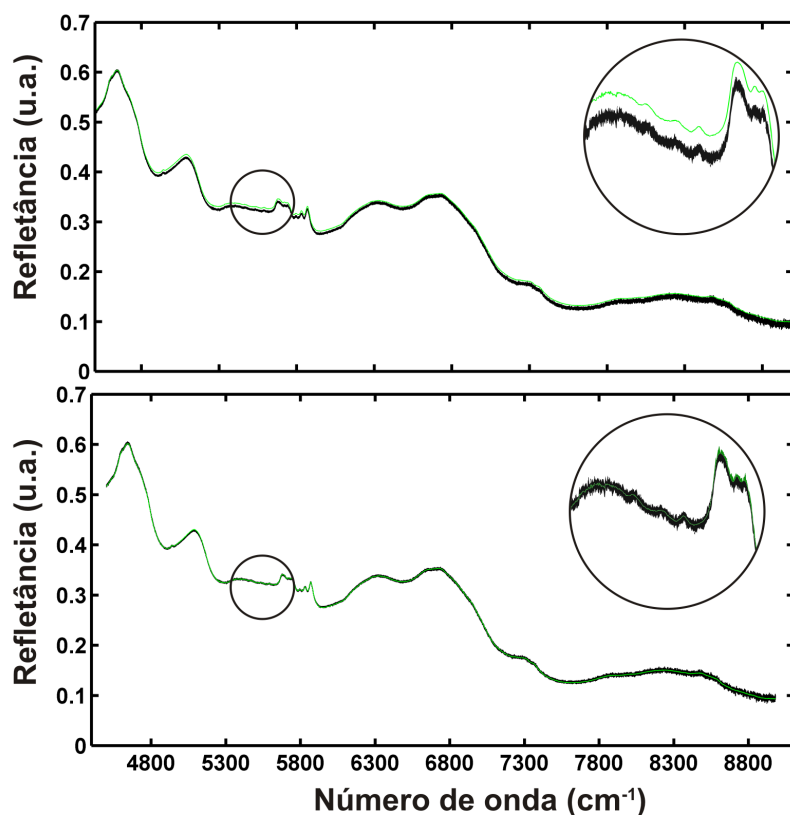


Figura 4.23 Cento e um espectros sobrepostos de um comprimido do princípio-ativo captopril na região de 4610-8928 cm^{-1} . (a) antes e (b) depois da correção da linha de base por EMSC. Em detalhe, corte ampliado da região entre 5300-5700 cm^{-1} .

Cada um dos cinquenta espectros de varredura unitária foi submetido às três técnicas de remoção de ruído por Transformada *Wavelet*, e o RMSE individual e médio foi calculado entre os espectros reconstruídos e o espectro de referência, obtido pela promediação de cem varreduras.

Pela comparação qualitativa entre os espectros reconstruídos, o espectro bruto de varredura unitária e o espectro de referência, podemos constatar que todas as técnicas mostraram desempenhos diferentes na remoção do componente ruidoso (Figura 4.24).

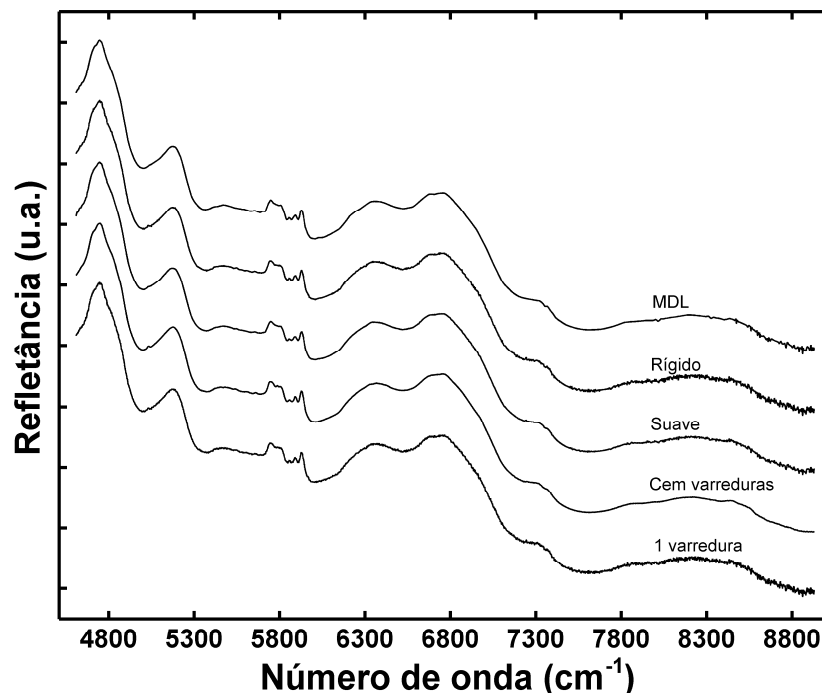


Figura 4.24 Sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso para a mesma amostra de captopril.

Uma porção considerável do ruído presente no espectro bruto foi mantida no espectro reconstruído pela técnica Rígida, principalmente quando comparado com o obtido pela abordagem Suave, a qual apresentou um espectro visualmente limpo abaixo de 7900 cm⁻¹. Por outro lado, a técnica MDL foi mais eficiente, removendo praticamente todo o ruído, persistindo apenas alguns artefatos na região posterior a 8300 cm⁻¹, difíceis de serem eliminados devido à baixa relação sinal-ruído do espectro bruto nesta região. Como podemos ver pela Figura 4.24, o espectro obtido pela técnica MDL, embora seja a reconstrução de uma única varredura, é muito similar ao de referência, obtido pela média de cem varreduras.

As técnicas MDL e Rígida mostraram resultados discrepantes, apesar de usarem a mesma função de limiarização (Rígida). Isso é devido ao limiar adotado,

mostrando que o limiar MDL foi mais eficiente na distinção entre ruído e sinal que o limiar universal, empregado na técnica Rígida.

Na [Tabela 1](#) são mostrados os RMSE's médios para as três técnicas frente às vinte e duas funções *wavelet*. É interessante notar que, na grande maioria das *wavelets* analisadas, as duas abordagens que fazem uso do limiar universal ficaram aquém do método MDL, corroborando os resultados obtidos pela inspeção visual dos espectros.

Tabela 1 Valores de RMSE médios obtidos com 1 varredura e 7 níveis de decomposição no banco de filtros.

RMSE ($\times 10^{-3}$)			
Filtro <i>wavelet</i>	Técnicas		
	Suave	Rígida	MDL
Db1	1,33	1,37	1,57
Db2	1,03	1,33	1,05
Db3	1,01	1,34	1,01
Db4	1,01	1,35	0,99
Db5	1,02	1,35	1,00
Db6	1,03	1,36	1,00
Db7	1,02	1,36	1,01
Db8	1,03	1,36	1,01
Db9	1,04	1,37	1,02
Db10	1,05	1,37	1,02
Sym4	1,02	1,34	1,01
Sym5	1,02	1,35	1,00
Sym6	1,03	1,36	1,01
Sym7	1,03	1,37	1,00
Sym8	1,03	1,36	1,00
Sym9	1,04	1,36	1,01
Sym10	1,03	1,36	1,00
Coif1	1,03	1,32	1,06
Coif2	1,02	1,35	1,00
Coif3	1,03	1,36	1,01
Coif4	1,03	1,36	1,01
Coif5	1,03	1,37	1,00

Pela análise da [Tabela 1](#) e da [Figura 4.24](#), podemos perceber que os resultados para cada técnica foram influenciados apenas pela capacidade de

remoção de ruído de cada uma, já que nenhuma delas causou deformações ao sinal, o que poderia trazer perda de informação química.

Como discutido anteriormente, o método MDL pode ser usado para selecionar uma *wavelet* apropriada para o processo de remoção de ruído. Essa escolha é baseada no menor valor obtido pela minimização da função custo MDL. Através dos RMSE's obtidos pela técnica MDL para os vinte e dois filtros, em cada um dos cinquenta espectros, podemos avaliar a eficiência dessa escolha. Na [Tabela 2](#) são apresentados os erros obtidos pelas três técnicas nos espectros em que o MDL obteve o melhor e o pior desempenho, respectivamente, no que diz respeito à seleção do filtro ótimo.

Tabela 2 Comparação entre as técnicas para o melhor (espectro 47) e o pior (espectro 5) resultado na seleção da função *wavelet* pelo critério MDL.

Filtro wavelet	Espectro 47					Espectro 5				
	Suave	Rígida	k	MDL	k	Suave	Rígida	k	MDL	k
Db1	1,33	1,38	679	1,57	196	1,30	1,35	674	1,55	194
Db2	1,05	1,35	651	1,05	152	1,01	1,30	598	0,99	145
Db3	1,02	1,35	642	1,04	139	0,99	1,32	625	0,98	128
Db4	1,02	1,36	663	1,01	148	0,99	1,32	653	0,93	138
Db5	1,04	1,36	659	1,01	141	1,00	1,33	633	0,94	147
Db6	1,05	1,37	709	1,02	135	1,01	1,33	636	0,93	139
Db7	1,05	1,38	682	1,04	142	1,02	1,35	671	0,96	134
Db8	1,05	1,38	739	1,06	152	1,02	1,36	712	0,98	142
Db9	1,06	1,37	716	1,07	146	1,03	1,35	725	0,97	147
Db10	1,06	1,39	761	1,03	148	1,04	1,35	725	0,97	144
Sym4	1,02	1,35	624	1,02	135	1,00	1,33	627	0,98	133
Sym5	1,03	1,36	667	1,01	142	1,01	1,33	648	0,98	146
Sym6	1,03	1,37	678	1,01	132	1,02	1,33	612	0,98	131
Sym7	1,05	1,38	711	1,07	154	1,02	1,34	677	0,94	145
Sym8	1,04	1,38	697	1,03	137	1,02	1,34	670	0,96	131
Sym9	1,05	1,38	709	1,02	155	1,03	1,35	688	0,98	146
Sym10	1,05	1,38	716	1,01	148	1,02	1,34	660	0,97	130
Coif1	1,04	1,34	616	1,06	138	1,03	1,30	571	1,04	138
Coif2	1,03	1,36	670	1,02	137	1,00	1,33	622	0,94	123
Coif3	1,03	1,37	693	1,03	138	1,02	1,33	609	1,01	125
Coif4	1,05	1,38	717	1,03	141	1,02	1,34	649	0,95	128
Coif5*	1,04	1,38	730	1,00	134	1,02	1,34	640	0,99	129

Entre os cinquenta espectros analisados, o espectro 47 foi o que o MDL apresentou o melhor desempenho na seleção do filtro *wavelet*, obtendo o mínimo valor da função-custo para a *wavelet* Coiflet 5, que obteve o menor RMSE entre os vinte e dois filtros. Já no espectro 5, apesar de o MDL ter selecionado o mesmo filtro *wavelet* (Coiflet 5), este apresentou o 18º menor erro, sendo o pior desempenho do MDL na seleção do filtro ótimo. É interessante notar que, em ambos os espectros, os erros obtidos pela *wavelet* escolhida através do uso da técnica MDL foram menores que os obtidos pelas outras técnicas.

A abordagem MDL tem um compromisso com a parcimônia na seleção dos coeficientes *wavelet*. Sua função-custo busca um limiar que permita uma reconstrução com o mínimo possível de coeficientes e com um erro aceitavelmente baixo, garantido pela função penalizante contida no segundo termo da função custo. Tal característica foi evidenciada experimentalmente através do número de coeficientes selecionados para cada *wavelet* pelo método MDL. Pela inspeção dos gráficos de barras da [Figura 4.25](#), podemos notar que o limiar MDL obteve para todas as funções um número muito menor de coeficientes que as técnicas baseadas no limiar universal.

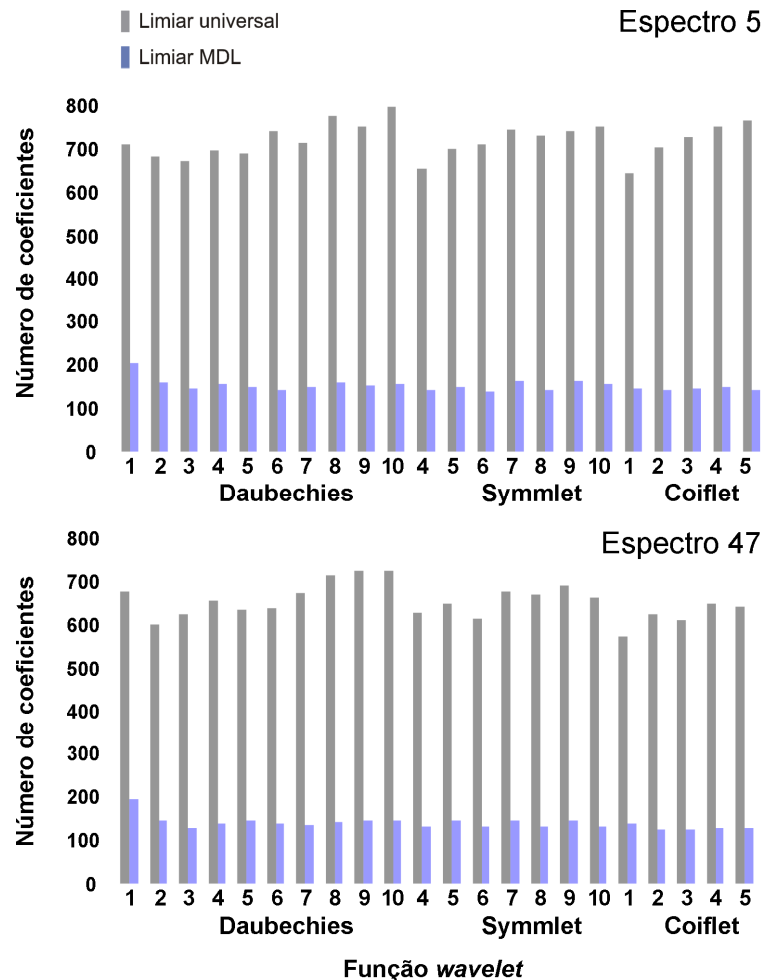


Figura 4.25 Comparação do número de coeficientes *wavelet* escolhidos através das 22 *wavelets* para os espectros 5 e 47, respectivamente, usando-se o limiar universal (barras cinza) e o limiar MDL (barras azuis).

Para o espectro 47, por exemplo, o MDL indicou que apenas 134 coeficientes, de um total de 4480, seriam necessários para a reconstrução do espectro com mínima distorção, enquanto que as técnicas baseadas no limiar universal indicaram 730 coeficientes, o quádruplo dos coeficientes da técnica MDL para o mesmo espectro. Para o espectro 5, o MDL escolheu apenas 128 coeficientes, sendo o 5º resultado mais parcimonioso entre as vinte e duas *wavelets*, o que pode explicar o erro relativamente alto da *wavelet* escolhida pelo MDL para este espectro, quiçá, pela busca de um equilíbrio entre a parcimônia e o erro da reconstrução, inerentes à função-custo do MDL.

O MDL apresentou considerável robustez na seleção da função *wavelet*, selecionando o mesmo filtro (*coiflet 5*) em 64% dos cinquenta espectros individuais. O filtro Sym8 foi selecionado em 20% dos casos, e somente cinco filtros foram selecionados nos restantes 16% dos espectros (Figura 4.26).

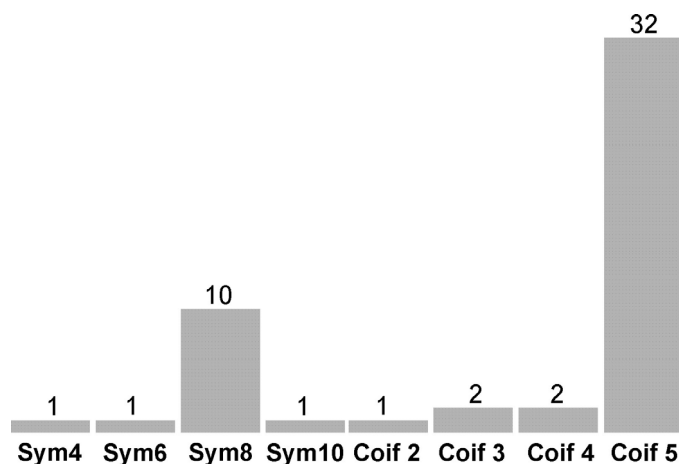


Figura 4.26 Gráfico de barras apresentando a extensão da seleção do filtro wavelet pelos nos espectros individuais.

Vale salientar que para as técnicas Rígida e Dura não há um critério objetivo para a escolha do filtro *wavelet*, dificultando em muito a tarefa do analista.

4.3 Experimento 3 – Dipirona Sódica / Espectrometria NIR

O foco deste experimento é a remoção de ruído em sinais de refletância difusa na região do infravermelho próximo. Porém, diferentemente do experimento anterior, os espectros foram registrados em um equipamento de outro fabricante (Perkin Elmer) e usando um detector diferente (DTGS). Além disso, a amostra em questão, por se tratar de um sólido em pó, apresenta maior suscetibilidade a produzir espectros ruidosos. Tal comportamento pode ser confirmado pela inspeção da [Figura 4.27a](#), que apresenta o espectro de dipirona sódica obtido pela promediação de 256 varreduras (verde) e um espectro de varredura unitária (preto), na região de 7.500 a 4.000 cm^{-1} .

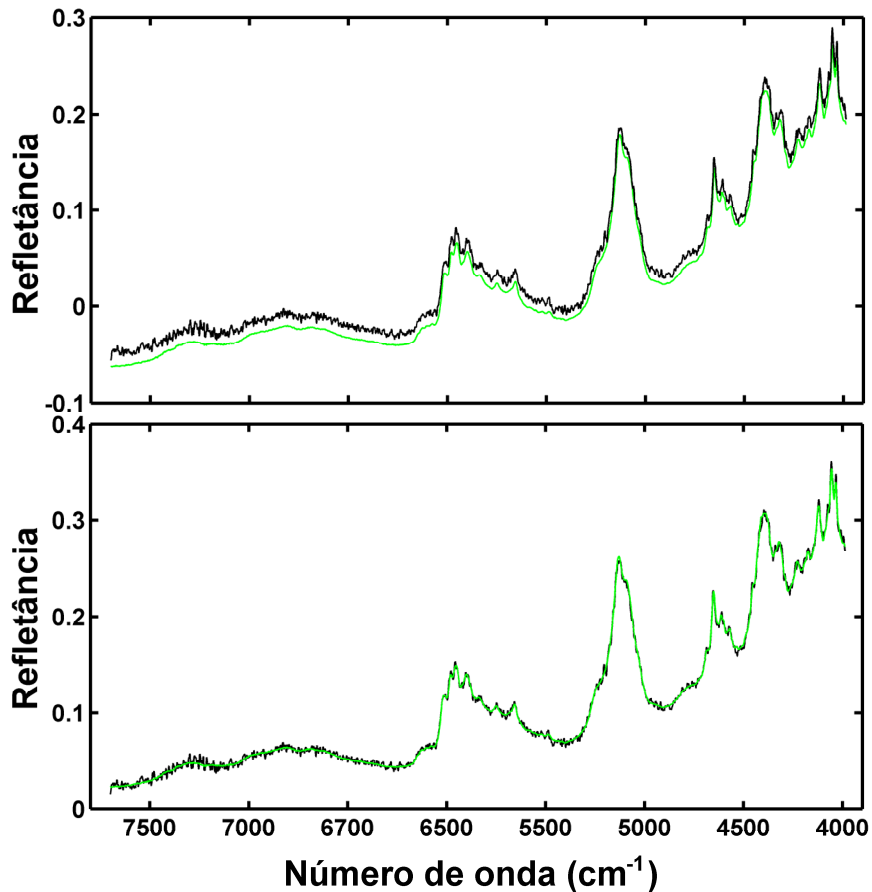


Figura 4.27 Espectro de referência e um espectro de varredura unitária de uma amostra do princípio-ativo dipirona sódica, (a) antes e (b) depois da correção da linha de base pelo EMSC.

Notadamente, o espectro de varredura unitária do princípio-ativo dipirona apresenta-se mais ruidoso que o espectro do experimento anterior, apresentando ruído proeminente em todas as regiões, a ponto de deformar algumas feições presentes no sinal de referência, o que pode comprometer a obtenção de um espectro similar a este pelas técnicas aqui estudadas. Os problemas de deslocamento da linha de base também se mostraram maiores, sendo aplicada a técnica EMCS para a correção dos mesmos (Figura 4.27b).

Todos os espectros resultantes compreendem 3.712 variáveis, um número adequado para a aplicação das técnicas de remoção de ruído por TW. O nível máximo de decomposição do banco de filtros foi escolhido pela função *wmaxlev.m*, que indicou o uso de 7 níveis.

Cada um dos cinquenta espectros de varredura unitária foi submetido às três técnicas de remoção de ruído por Transformada *wavelet*, e o RMSE individual e

médio foi calculado entre os espectros reconstruídos e o espectro de referência, obtido pela promediação de 256 varreduras.

Na [Figura 4.28](#), podemos visualizar os espectros reconstruídos pelas três técnicas, o espectro de referência e o espectro de varredura unitária.

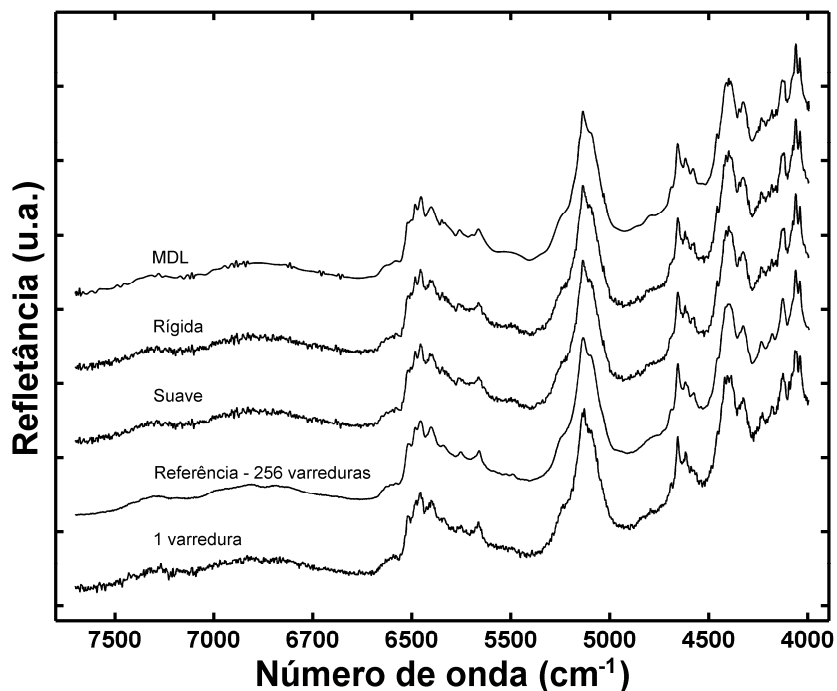


Figura 4.28 Sinais reconstruídos pelas técnicas MDL, Rígida e Suave, sinal de referência e sinal ruidoso para a mesma amostra do princípio-ativo dipirona sódica.

Como podemos notar, o ruído proeminente no espectro de varredura se manteve nos espectros reconstruídos pelas técnicas Suave e Rígida. A técnica MDL removeu considerável porção do ruído, porém, alguns artefatos ainda permaneceram devido ao elevado ruído no espectro bruto. Ao trabalharmos com amostras suscetíveis a produzirem espectros muito ruidosos, talvez uma varredura seja insuficiente para a obtenção de um espectro reconstruído fiel ao espectro de referência. Em tais casos, um estudo com mais varreduras é recomendável. Na [Figura 4.29](#) é mostrado o comportamento das técnicas na remoção de ruído em um espectro obtido pela promediação de quatro varreduras.

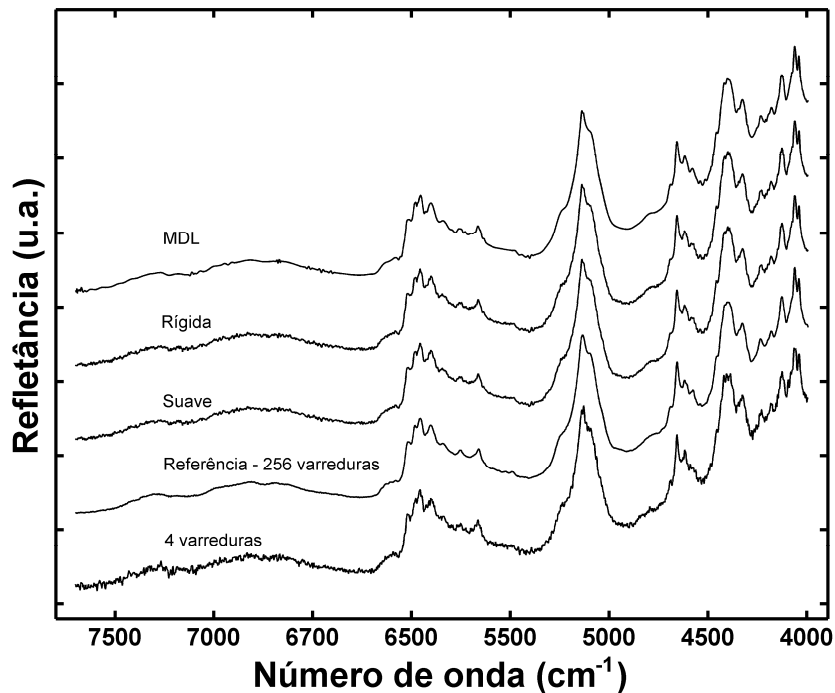


Figura 4.29 De cima para baixo, sinais reconstruídos pelas técnicas MDL, Rígido e Suave, sinal de referência e sinal ruidoso para a mesma amostra do princípio-ativo dipirona.

Igualmente ao experimento com o espectro de uma varredura, as técnicas baseadas no limiar universal também não foram eficientes na extração do componente ruidoso, mantendo considerável porção em seus espectros reconstruídos. Por outro lado, a técnica MDL mostrou-se eficiente, excluindo a maior parte do ruído e conservando apenas pequenas deformações inerentes ao espectro bruto, em menor proporção que as observadas no espectro reconstruído de uma única varredura. Na região compreendida entre 5.400 e 4.000 cm^{-1} , o espectro reconstruído pela técnica MDL é similar ao de referência, apesar da grande diferença no número de varreduras, e conseqüentemente, no tempo despendido para a sua obtenção.

De fato, para o equipamento em questão, um espectro promediado por quatro varreduras foi obtido em 1 minuto, enquanto o espectro obtido por 256 varreduras necessitou de aproximadamente 35 minutos. Em um experimento normal, comumente utiliza-se uma média de 32 a 64 varreduras, mesmo assim, demandando um tempo considerável. Isso pode ser uma desvantagem em aplicações onde a rapidez é fator preponderante, como em Tecnologia Analítica de Processos.

Na Figura 4.30 são mostrados os RMSE's obtidos pelas técnicas Rígida, Suave e MDL, com relação as vinte e duas *wavelets*. Essa avaliação quantitativa corrobora os resultados obtidos pela inspeção visual dos espectros, com a técnica MDL obtendo os menores erros, bem abaixo das técnicas baseadas no limiar universal, com exceção para a *wavelet* Daubechies 1, onde o MDL apresentou um RMSE elevado.

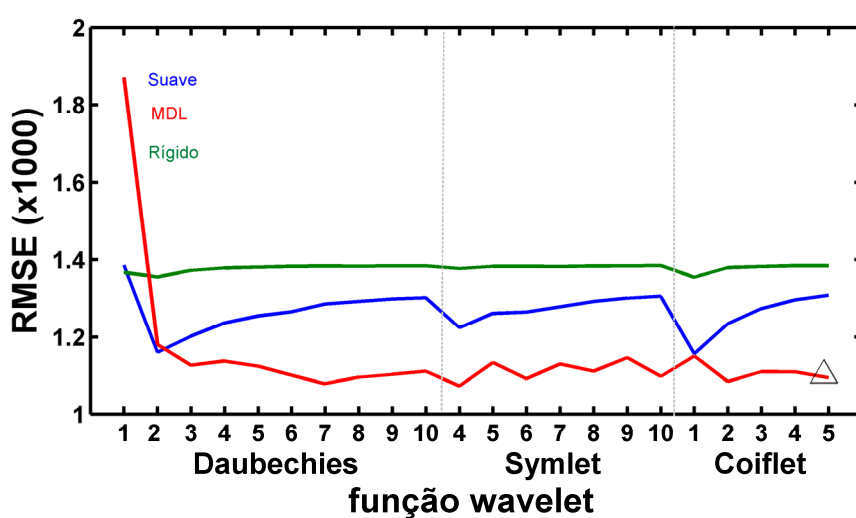


Figura 4.30 Curva do RMSE médio x função *wavelet* para as três técnicas de remoção de ruído por Transformada *wavelet* para os espectros de dipirona sódica.

Entre as vinte e duas *wavelets* contidas na biblioteca de funções, o MDL apontou como ótima a *wavelet coiflet 5*, que forneceu o quinto menor RMSE, cujo valor não difere estatisticamente do menor erro, segundo um teste *F* realizado com 95% de confiança.

Em relação à capacidade de compressão das duas funções de limiar para a *wavelet* escolhida pelo limiar MDL, apenas 194 coeficientes foram escolhidos de um total de 3712, enquanto que o limiar universal escolheu 1565 coeficientes, um número oito vezes maior, reforçando as vantagens de compressão de sinais da abordagem MDL.

Dessa forma, para o problema de remoção de ruído em espectros de refletância difusa de amostras sólidas em pó usando-se um número pequeno de varreduras, podemos concluir que a técnica MDL apresentou resultados mais satisfatórios que as técnicas Rígida e Suave. O MDL, além de resultar em erros mais baixos em relação ao espectro de referência, permitiu a reconstrução com menos

informação espúria do sinal, através da seleção de um número muito menor de coeficientes, possibilitando um espectro muito similar ao espectro de referência. Além disso, a *wavelet* a ser utilizada pode ser escolhida automaticamente, não necessitando de inferências do analista, principalmente por não haver um critério definido usando-se as outras duas técnicas.

4.4 Experimento 4 – Óleo Diesel / Espectrometria NIR

Neste experimento os resultados obtidos por [GALVÃO et al \(2007\)](#) em um artigo publicado na *Analytica Chimica Acta* foram reproduzidos e contrastados com os obtidos pela técnica MDL.

Para a remoção da linha de base, diferentemente dos experimentos anteriores, foi empregado o procedimento usado pelos autores, que consiste em subtrair a média de cada espectro sobre ele mesmo.

Para comparar as técnicas em termos visuais, os autores utilizaram um critério para definir um número de varreduras ideal, acima do qual o tempo consumido não justificaria varreduras adicionais. O critério baseia-se na melhora da qualidade do sinal na medida em que se aumenta o número de varreduras, até um ponto em que a melhora não seja mais significativa. Detalhes sobre o critério podem ser encontrados na referência supracitada. Segundo o critério, foi adotado como espectro bruto, o obtido pela promediação de dezesseis varreduras, sendo, então, calculado o RMSE em relação ao espectro de referência.

Os espectros compreendem 2560 variáveis, portanto, um número adequado para a aplicação das técnicas de remoção de ruído por TW. O número máximo de níveis de decomposição do banco de filtros foi escolhido pela função *wmaxlev.m*, que indicou 7 níveis.

A [Figura 4.31](#) apresenta o espectro bruto, bem como os espectros de referência e os tratados pelas quatro técnicas em estudo. Podemos notar que, com exceção da técnica Rígida, todas as outras técnicas resultaram em espectros mais suavizados. Quando comparados com o espectro de referência, os resultados são semelhantes aos obtidos no Experimental 2, com o MDL apresentando o espectro mais similar.

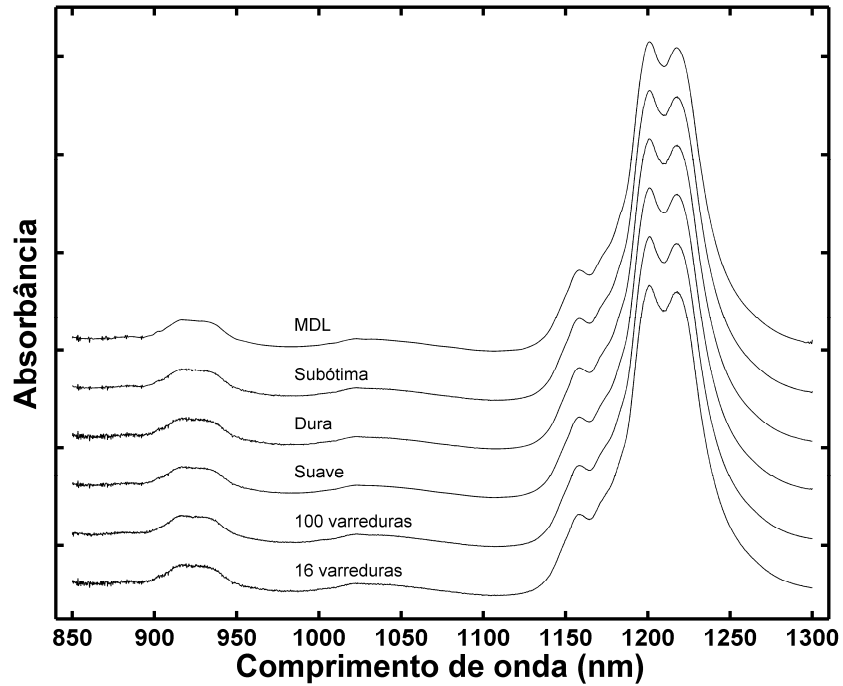


Figura 4.31 Espectro bruto, espectro de referência e os espectros reconstruídos após remoção de ruído pelas técnicas em estudo. Diferentes deslocamentos do eixo y foram adicionados para melhor visualização.

As principais diferenças podem ser vistas na região do *short-NIR*, onde a relação sinal-ruído é relativamente pobre. A [Figura 4.32](#) apresenta essa região (850-900 nm), em maior detalhe.

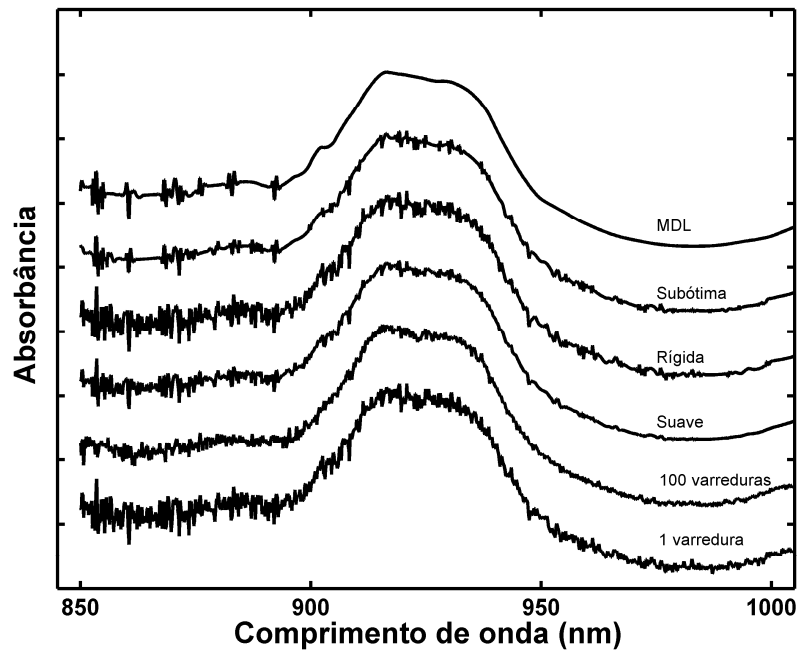


Figura 4.32 Corte realizado na região do *short-NIR*, do espectro bruto, espectro de referência e dos espectros reconstruídos após remoção de ruído.

Pela inspeção da região do *short-NIR*, podemos notar que o MDL apresentou o melhor resultado, removendo todo o ruído presente no espectro bruto na região acima de 900 nm, sem distorcer o sinal, restando apenas alguns artefatos na região inferior a essa, devido ao comprometimento do espectro bruto.

Contudo, os resultados qualitativos contrastam com os obtidos pela comparação através do cálculo do RMSE, como podemos ver na [Figura 4.33](#) e na [Tabela 3](#).

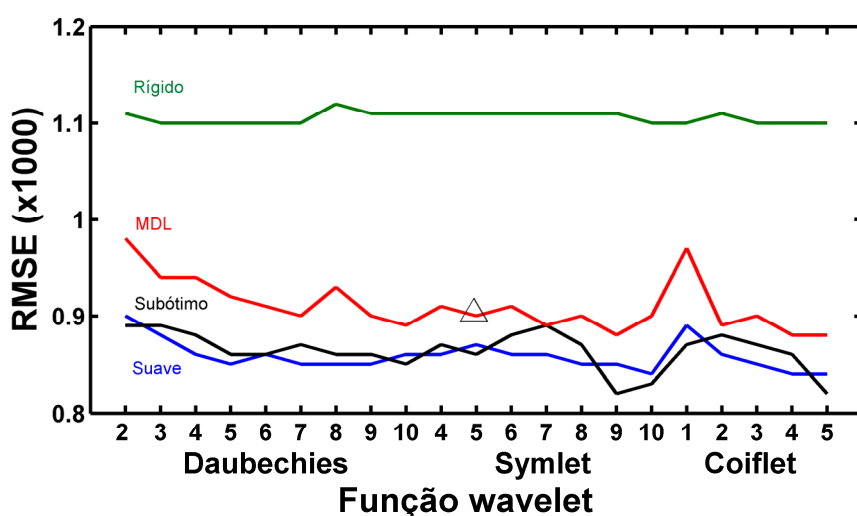


Figura 4.33 Curva do RMSE médio x função *wavelet* para as quatro técnicas de remoção de ruído por Transformada *Wavelet* para os espectros de óleo diesel.

O MDL resultou em erros maiores que as técnicas Suave e Subótima, para todos os filtros *wavelet*. Este comportamento conflitante entre os resultados de RMSE e a inspeção visual dos espectros, que aponta claramente uma superioridade do MDL em relação à eficiência da remoção de ruído, pode ser explicado pela presença de considerável ruído na região do *short-NIR* do espectro de referência, que comprometeu a eficiência da métrica do RMSE. De fato, o MDL resultou em um espectro mais limpo que o próprio espectro de referência.

Tabela 3 Valores de RMSE, obtidos pelas quatro técnicas frente aos vinte e dois filtros wavelet, entre o espectro de referência e o obtido pela promediação de 16 varreduras. A *wavelet* selecionada com base no critério MDL seria a Sym5, indicada com um asterisco*.

Filtro Wavelet	Técnicas de remoção de ruído						
	Suave	Rígida	k	Sub ótima	k	MDL	k
Db1	1,55	1,26	487	0,96	1043	2,18	151
Db2	0,90	1,11	459	0,89	387	0,98	115
Db3	0,88	1,10	435	0,89	333	0,94	91
Db4	0,86	1,10	427	0,88	311	0,94	97
Db5	0,85	1,10	409	0,86	319	0,92	94
Db6	0,86	1,10	371	0,86	302	0,91	80
Db7	0,85	1,10	396	0,87	318	0,90	93
Db8	0,85	1,12	468	0,86	327	0,93	87
Db9	0,85	1,11	443	0,86	331	0,90	91
Db10	0,86	1,11	434	0,85	321	0,89	90
Sym4	0,86	1,11	430	0,87	339	0,91	89
Sym5*	0,87	1,11	433	0,86	314	0,90	76
Sym6	0,86	1,11	427	0,88	345	0,91	86
Sym7	0,86	1,11	421	0,89	322	0,89	91
Sym8	0,85	1,11	426	0,87	345	0,90	88
Sym9	0,85	1,11	440	0,82	322	0,88	83
Sym10	0,84	1,10	409	0,83	344	0,90	93
Coif1	0,89	1,10	432	0,87	405	0,97	111
Coif2	0,86	1,11	424	0,88	329	0,89	83
Coif3	0,85	1,10	387	0,87	331	0,90	86
Coif4	0,84	1,10	419	0,86	320	0,88	87
Coif5	0,84	1,10	406	0,82	327	0,88	88

A função *wavelet* escolhida pelo MDL como ótima foi a *wavelet Symlet 5*, que apresentou o 12º menor erro, porém, através de um teste-*F* com 95% de confiança, constata-se que este valor não é estatisticamente significativo ao menor valor. Vale ressaltar que a *wavelet Symlet 5* permitiu a reconstrução do sinal com apenas 76 coeficientes, o resultado mais parcimonioso entre as 22 *wavelets*, bem menor que o número de coeficientes necessários para a reconstrução dos espectros pelo limiar Universal, que necessitou de até seis vezes mais coeficientes que o limiar MDL. O

limiar Subótimo, por sua vez, necessitou de quatro vezes mais coeficientes que o MDL, porém, superando as técnicas baseadas no limiar universal, com exceção para o filtro *wavelet* Haar, no qual apresentou desempenho assaz fraco. O comportamento de cada uma das vinte e duas *wavelets*, com respeito ao número de coeficientes selecionados em cada técnica, pode ser visto na [Figura 4.34](#) e na [Tabela 3](#), acima.

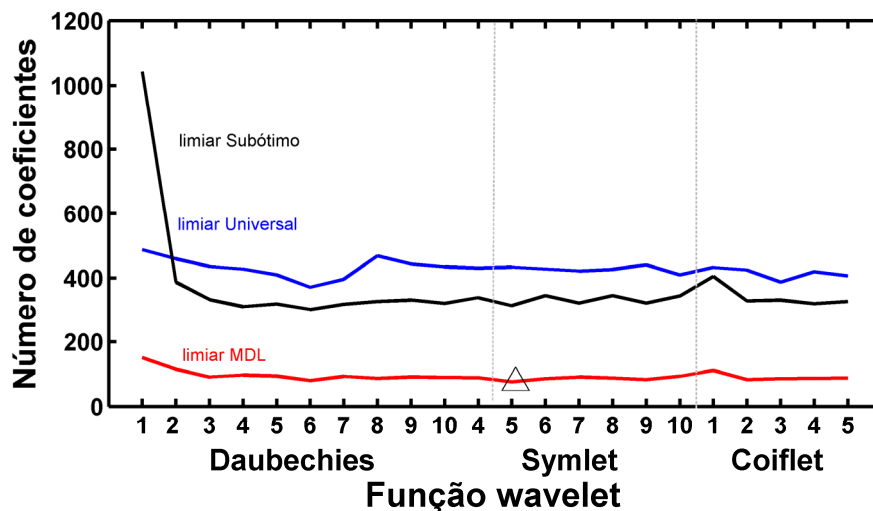


Figura 4.34 Curvas do número de coeficientes selecionados pelas quatro técnicas de remoção de ruído, para as 22 funções *wavelet*.

É interessante notar o comportamento das técnicas com relação ao filtro *wavelet* Daubechies 1, que apresentou resultado diverso aos outros filtros. Devido ao comprimento deste filtro (2), o menor entre todas as 22 *wavelets*, mais coeficientes foram necessários para a reconstrução dos espectros, elevando o RMSE para as técnicas MDL, Rígida e Suave. A técnica Subótima obteve para essa *wavelet*, um RMSE similar aos obtidos para as outras 21, porém, sacrificando a parcimônia, necessitando de 1043 coeficientes *wavelet* para a adequada reconstrução do espectro, um número onze vezes maior que o selecionado pelo MDL para a mesma função.

4.5 Experimento 5 – Extrato vegetal / ^1H RMN

Neste experimento, diferentemente dos outros abordados nesta dissertação, é estudada a remoção de ruído em espectros de Ressonância Magnética Nuclear. Semelhante à espectrometria no infravermelho próximo, a RMN também opera com Transformada de Fourier, sendo um exemplo clássico do emprego da promediação de sinais objetivando o aumento da relação sinal-ruído. Cortes do espectro bruto obtido pela promediação de dezesseis transientes, e do espectro de referência, promediado por mil e vinte e quatro transientes, são mostrados na [Figura 4.35](#).

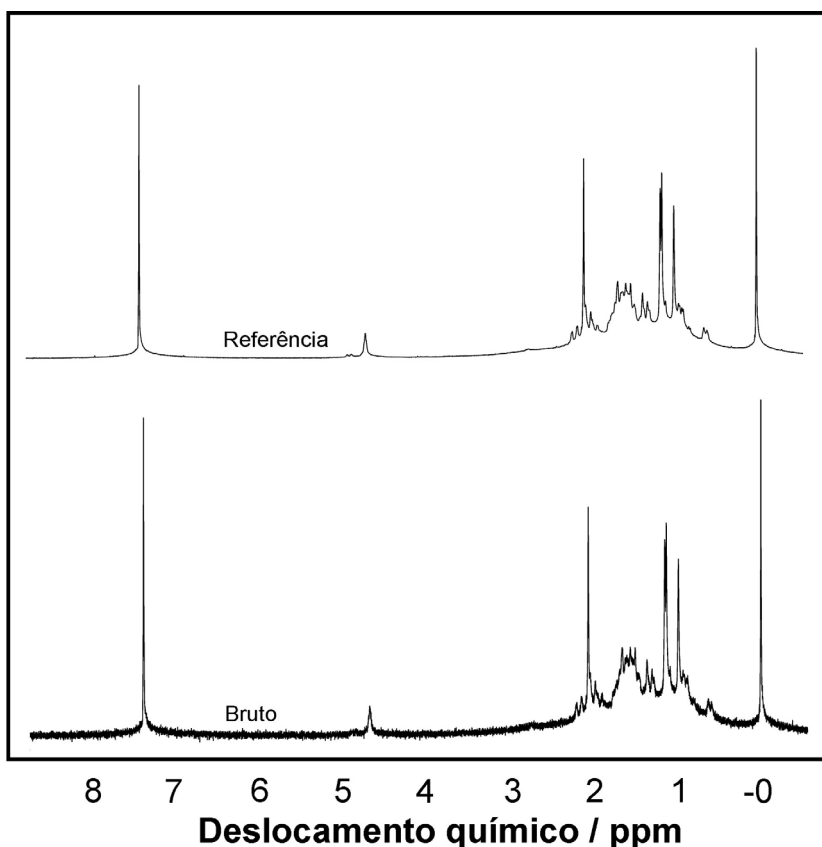


Figura 4.35 Espectro de ^1H RMN obtido pela promediação de (cima) 1024 varreduras, e (baixo) 16 varreduras.

Os dois espectros resultaram em 32.768 pontos cada, ou seja, um comprimento adequado para o bom funcionamento do banco de filtros *wavelet*. Espectros de RMN geralmente são registrados com um tamanho múltiplo de 16.384 (chamado de 16 k) para a adequada execução do algoritmo que calcula a

transformada de Fourier (LAMBERT e MAZZOLA, 2004). Assim, não houve preocupação com a adequação no tamanho do espectro, como também não foi necessário o uso de técnicas de atenuação da linha de base, já que os espectros se apresentaram bem sobrepostos. O nível máximo de decomposição do banco de filtros wavelet foi fixado em 10 através da função *wmaxlev.m*.

Através da Figura 4.36 pode ser observado que as técnicas Suave, Rígida e MDL foram eficazes na remoção do ruído presente no espectro bruto obtido por dezesseis varreduras. Apesar do proeminente ruído contido neste, os espectros reconstruídos pelas três técnicas se mostraram muito similares ao espectro de referência, obtido pela promediação de 1024 transientes. Igualmente ao observado nos sinais sintéticos do Experimento 1, a natureza homoscedástica do ruído nestes espectros contribuiu fortemente para o bom funcionamento das duas funções de limiar.

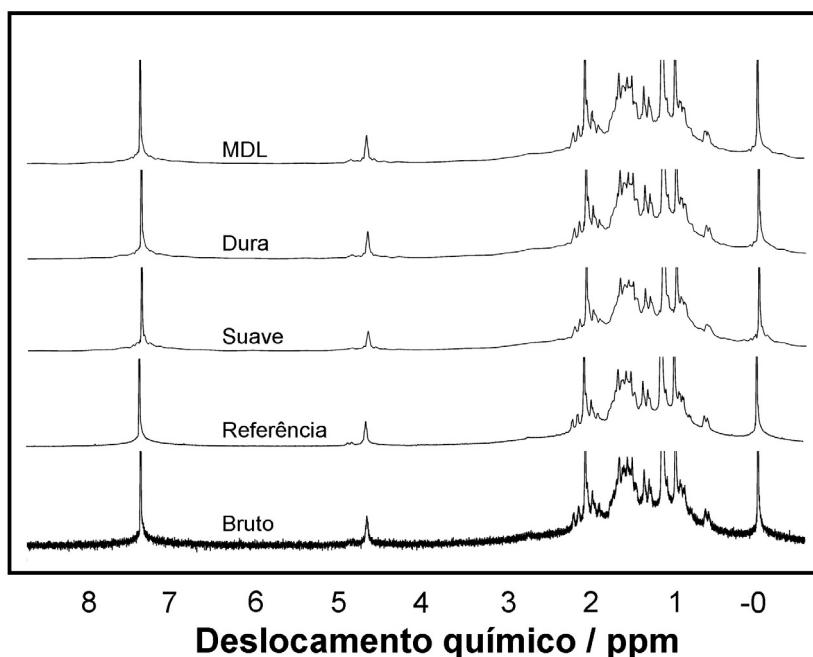


Figura 4.36 Espectros de ^1H RMN reconstruídos pela TW através das técnicas MDL, Rígida e Suave, e os espectros obtidos pela promediação de (cima) 1024 varreduras, e (baixo) 16 varreduras.

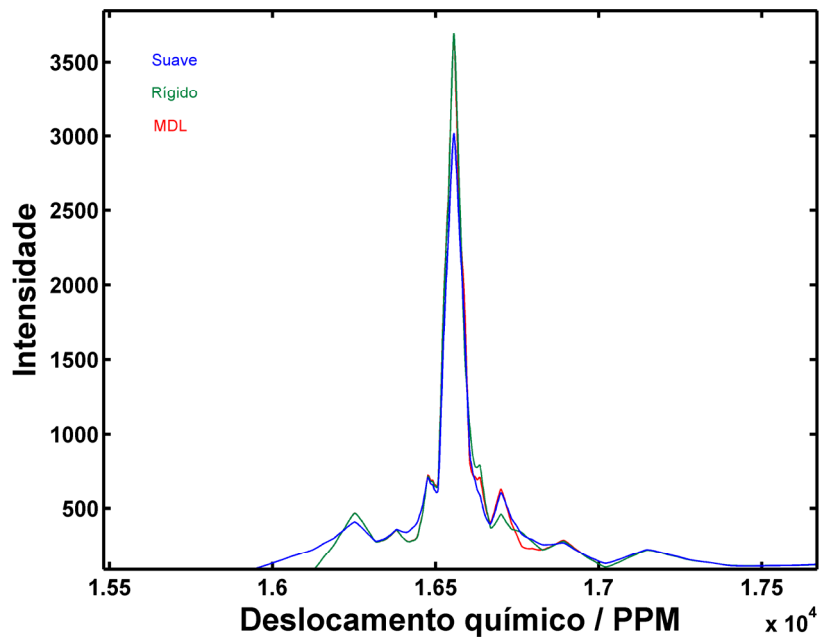


Figura 4.37 Corte nos espectros de ^1H RMN reconstruídos pelas técnicas MDL (vermelho), Rígida (verde) e Suave (azul).

Dessa forma, como ambos os limiares funcionaram bem, a diferença de eficiência entre as três técnicas foi influenciada pelo rigor no tratamento dos coeficientes *wavelet*, isto é, pela função de limiarização, o que pode ser constatado pela análise das Figuras 4.36-37, além da Tabela 4. De fato, a técnica Suave, devido à sua função de limiarização mais severa, atenuando os coeficientes de detalhe cujos valores estão acima do limiar, resultou nos maiores valores de RMSE, justamente por distorcer o sinal em algumas regiões, como por exemplo, a diferença em intensidade no pico localizado em volta de 1,65 ppm (Figura 4.37), em relação aos espectros reconstruídos através das técnicas que fazem uso da função de limiarização Rígida (Rígida e MDL). Entre elas, a técnica Rígida obteve os menores valores de RMSE, também corroborando os resultados obtidos com sinais sintéticos com ruído homoscedástico (Experimento 1).

Tabela 4 Comparação entre as técnicas para o melhor e o pior resultado na seleção da função *wavelet*. A *wavelet* selecionada com base no critério MDL é indicada com um asterisco.

Filtro Wavelet	RMSE				
	Técnicas de remoção de ruído				
	Suave	Rígida	k	MDL	k
Db1	135,81	85,67	358	94,90	310
Db2	111,25	67,65	257	74,12	229
Db3	108,32	64,09	258	69,82	233
Db4	107,61	63,43	260	69,87	232
Db5	112,91	62,22	278	68,40	254
Db6	115,19	63,31	294	67,87	272
Db7	113,33	64,69	287	70,54	268
Db8	116,34	64,05	305	69,99	283
Db9	118,70	66,73	310	70,27	294
Db10	121,58	63,15	338	69,82	309
Sym4	109,03	64,23	248	71,51	222
Sym5	107,96	64,07	260	72,22	232
Sym6	110,33	63,79	260	69,62	239
Sym7	109,58	61,61	267	68,36	238
Sym8	113,04	65,82	270	71,89	248
Sym9	110,71	63,01	286	71,73	256
Sym10	113,68	65,19	275	72,40	249
Coif1	111,96	66,31	265	74,91	233
Coif2*	107,30	63,08	247	69,11	226
Coif3	112,40	64,77	265	71,61	242
Coif4	113,04	64,63	271	71,61	248
Coif5	114,54	65,62	281	71,94	259

A boa eficiência na remoção de ruído desempenhada pelas três técnicas aqui estudadas, além do pequeno número de coeficientes *wavelet* selecionados por ambas as técnicas de limiar (até cem vezes menores), pode ser explicado pela homoscedasticidade do ruído, que é uma das premissas para o bom funcionamento das técnicas de remoção de ruído por Transformada *Wavelet*.

De fato, com o ruído sendo homoscedástico, as funções de limiar podem trabalhar de forma adequada, visto que o limiar escolhido pode abranger toda a faixa

de coeficientes, devido à constância da variância do ruído, o que não ocorre em espectros heteroscedásticos. O número de coeficientes selecionados pelos limiares universal e MDL para os espectros de RMN corroboram este comportamento, visto que ambas selecionaram números aproximados de coeficientes, com uma leve vantagem para a técnica MDL em todas as *wavelets*. Para a *wavelet* selecionada pelo MDL (Coif2), foi necessário menos que o 1% (226 coeficientes) do número de variáveis iniciais (32768 variáveis).

Este experimento mostrou a viabilidade da remoção de ruído por Transformada *Wavelet* em espectros de RMN obtidos por poucas varreduras, possibilitando a economia de tempo de equipamento e operador, visto que para a obtenção de um espectro, por exemplo, de 16 varreduras, é necessário menos de um minuto, enquanto para obter um espectro de 1024 varreduras são requeridos 84 minutos.

Além disso, o uso da técnica MDL permite que a filtragem seja realizada sem inferências do operador, já que a técnica seleciona automaticamente um limiar e a função *wavelet* adequada. Somados o tempo de equipamento e o tempo computacional para o cálculo da TW, o procedimento completo pode ser efetuado em menos de dois minutos, para espectros de tamanho 32 kptos.

4.6 Estudo dos resultados obtidos nos Experimentos 1-5

Neste experimental, cinco experimentos foram realizados acerca do problema da remoção de ruído em sinais analíticos usando técnicas baseadas na Transformada *Wavelet*. Para uma comparação fidedigna entre as técnicas aqui abordadas, a saber, MDL, Rígida e Suave, foram realizados estudos usando dados simulados com diferentes níveis de ruído, dados instrumentais de Espectrometria NIR, nos modos de refletância, usando dois equipamentos diferentes e duas formas farmacêuticas, e em absorbância, usando amostras de óleo diesel. Dados de Ressonância Magnética Nuclear também foram explorados.

O comportamento das diferentes técnicas com relação à escolha da *wavelet*, limiar e função de limiarização, em espectros obtidos usando-se um número baixo de varreduras, é resumido a seguir em forma de tópicos.

4.6.1 Escolha da *Wavelet*

Selecionar uma função *wavelet* que melhor se adapte à natureza dos dados é uma tarefa não-trivial. Das quatro técnicas estudadas aqui, somente o MDL possui um critério definido para efetuar tal escolha. As outras técnicas requerem que as várias funções *wavelet* sejam testadas, e os resultados sejam comparados visualmente, tornando-se uma tarefa ineficiente e dependente da inferência subjetiva do analista (CAI e HARRINGTON, 1998). Uma outra possibilidade é o uso de uma função referendada pela literatura para um tipo particular de sinal, o que pode nos levar a um resultado diverso do ótimo, visto que para um dado sinal, duas funções *wavelet* similares podem dar resultados diferentes. Um exemplo deste comportamento pode ser observado na Figura 4.38, que apresenta um espectro de uma amostra de Captopril (Experimental 2) reconstruído após o emprego da técnica Suave usando-se *wavelets* da família Daubechies.

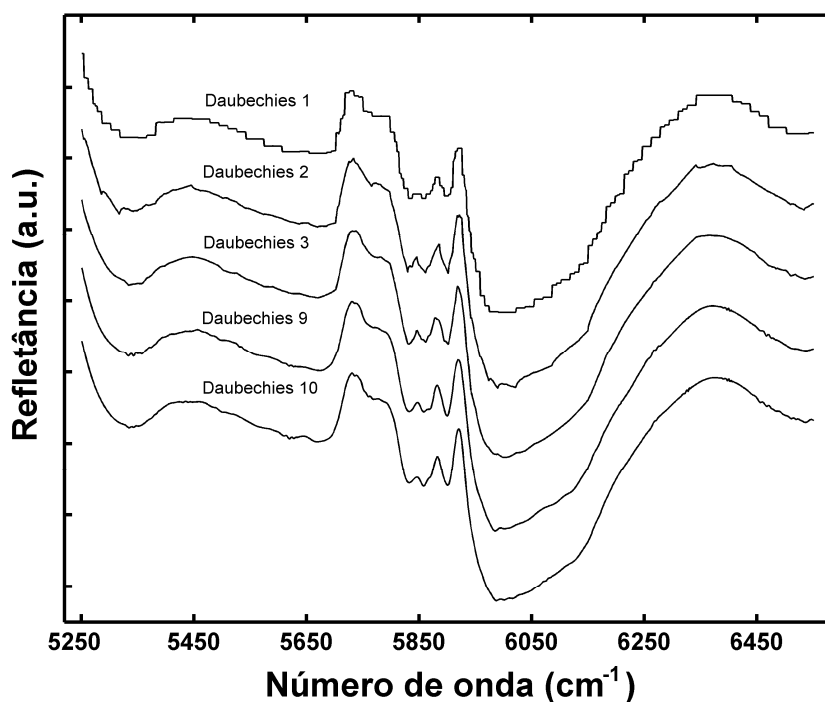


Figura 4.38 Corte nos espectros de captopril, reconstruídos pela técnica Suave usando-se as wavelets Daubechies 1-3 e 9-10.

Como podemos notar, os espectros reconstruídos apresentam feições diferentes, embora as *wavelets* sejam da mesma família e de comprimento de filtro próximos (Daubechies 1-3 e 9-10). O espectro reconstruído através da função Daubechies 1 distorceu completamente as curvas, deixando-as retas. Conforme se aumentou o tamanho do filtro, as feições dos espectros se alteraram, tornando-se mais suaves. Esse comportamento nos indica que a escolha de uma *wavelet* ao acaso pode levar a resultados não satisfatórios.

Nos cinco experimentos realizados nesta dissertação, o MDL desempenhou considerável eficiência na escolha da *wavelet* ótima. Nos experimentos com sinais sintéticos, onde foi realizado um estudo com cem sinais gerados e suas respectivas *wavelets* selecionadas, o MDL sempre escolheu *wavelets* que forneceram baixos valores de RMSE, entre os cinco menores, em aproximadamente 50% dos casos. Nos estudos com sinal instrumental real, a *wavelet* selecionada pelo MDL obteve sempre erros que não diferiam significativamente com relação aos menores obtidos, comprovando a eficácia da técnica na seleção da melhor função.

O MDL mostrou ser robusto na seleção da mesma *wavelet*, visto que apenas algumas funções foram escolhidas em predominância. É interessante notar que na maioria dos casos o MDL selecionou *wavelets* com filtros de comprimento maior, como a Sym9 e Db9, que possuem comprimento de 18.

4.6.2 Limiar e função de limiarização

O desempenho da função de limiar é muito dependente da natureza do ruído. Em ruído homoscedástico, como a variância do ruído permanece constante em todo o sinal, a estimativa do limiar torna-se mais fácil, pois pode abranger igualmente todos os coeficientes. Já em dados heteroscedásticos, a estimativa do limiar pode ser subestimada para alguns coeficientes e superestimada para outros, causando uma distinção errônea entre ruído e sinal.

Nos experimentos com dados sintéticos, cujo ruído é sabidamente homoscedástico, todas as três técnicas obtiveram bom comportamento. O limiar Universal obteve melhores resultados quando associado à limiarização Rígida, que resultou sempre em baixos valores de RMSE, além de espectros mais fidedignos ao

sinal de referência, independente do limiar. A técnica MDL, embora tenha obtido erros um pouco maiores que a técnica Rígida, possibilitou a reconstrução dos espectros filtrados com um número menor de coeficientes, em todos os experimentos, mostrando maior parcimônia que o limiar universal, sem distorcer o sinal. O mesmo comportamento pode ser observado nos sinais de Ressonância Magnética Nuclear, que apresentam um forte caráter homoscedástico.

Em ruídos cuja variância não é uniforme em todo o sinal, como os encontrados nos experimentos com Espectrometria NIR, a estimativa do ruído realizada pelo limiar universal foi subestimada, visto ambas as técnicas Rígida e Suave serem incapazes de remover o ruído de maneira satisfatória. De fato, a técnica MDL, apesar de utilizar a limiarização Rígida, obteve espectros muito similares aos de referência, além de baixos erros, devido, sobretudo, a sua função de limiar.

Vale salientar que o limiar MDL mostrou maior robustez frente aos diferentes regimes do ruído, mostrando menor suscetibilidade à heteroscedasticidade do ruído que o limiar universal, visto que em todos os experimentos com espectrometria NIR resultou nos espectros menos ruidosos e mais semelhantes ao espectro de referência.

Em todos os casos, o limiar MDL selecionou um número menor de coeficientes que o limiar Universal, principalmente nos dados com ruído heteroscedástico, cuja diferença chegou a oito vezes menos coeficientes, indicando que o limiar MDL foi mais eficaz na estimativa do ruído. O limiar Subótimo também apresentou bons resultados, mas devido à utilização em apenas um experimento, não foi possível fazer inferências mais aprofundadas.

4.6.4 Espectros obtidos com poucas varreduras

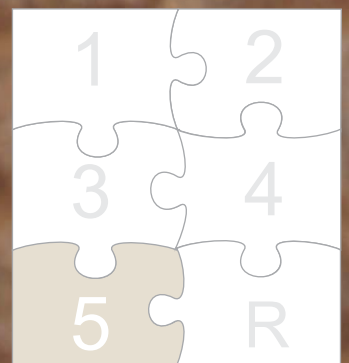
A possibilidade de se obter espectros usando-se um número pequeno de varreduras é muito atraente. Além da economia gerada pela diminuição de tempo, tanto de operador e de equipamento, a agilidade nas análises é uma importante característica, principalmente em aplicações da Tecnologia de Processos Analíticos,

agilizando a análise quimiométrica dos dados, e conseqüentemente, a obtenção de informação.

Entre as técnicas estudadas nessa dissertação, o MDL é a que mais se adapta a esse fim, visto que resultou nos espectros mais similares aos espectros de referência, obtido pela promediação de um número elevado de varreduras. Além disso, sua capacidade de selecionar automaticamente o limiar e a *wavelet* adequada à natureza dos sinais permite que a análise seja feita livre de inferências subjetivas do operador.

Nos experimentos usando a Espectrometria NIR, foi possível a obtenção de espectros fidedignos aos de referência, com o uso de apenas uma varredura. No entanto, dependendo do nível de ruído, algumas varreduras podem ser acrescentadas, de acordo com um estudo pormenorizado. Nos dados de RMN, devido à necessidade de um tempo maior de aquisição, o uso de um número menor de varreduras torna-se fundamental, visto que um espectro obtido por 1024 varreduras pode demandar até sessenta minutos para ser adquirido, ao passo que um espectro obtido pela promediação de dezesseis varreduras pode ser obtido em apenas um minuto.

conclusões



5 Conclusões

Nesta dissertação foi abordado o problema da remoção de ruído instrumental por Transformada *Wavelet* usando o método Mínimo Comprimento de Descrição. O método MDL seleciona automaticamente o limiar e a função *wavelet*, dispensando a tomada de decisões do analista, que requerem conhecimento do assunto, e que geralmente pode ser muito subjetiva.

A técnica MDL se mostrou capaz de selecionar uma *wavelet* adequada aos sinais, escolhendo sempre funções *wavelet* que forneceram baixos RMSE's. Além disso, apresentou resultados satisfatórios na remoção de ruído, obtendo espectros suavizados em sinais com ruído homoscedástico e heteroscedástico, enquanto que as técnicas Rígida e Suave obtiveram desempenhos variáveis, o que dificultaria a escolha de uma delas.

Em relação à capacidade de compressão dos sinais, o MDL mostrou ser parcimonioso, selecionando um número menor de coeficientes *wavelet* em todos os experimentos.

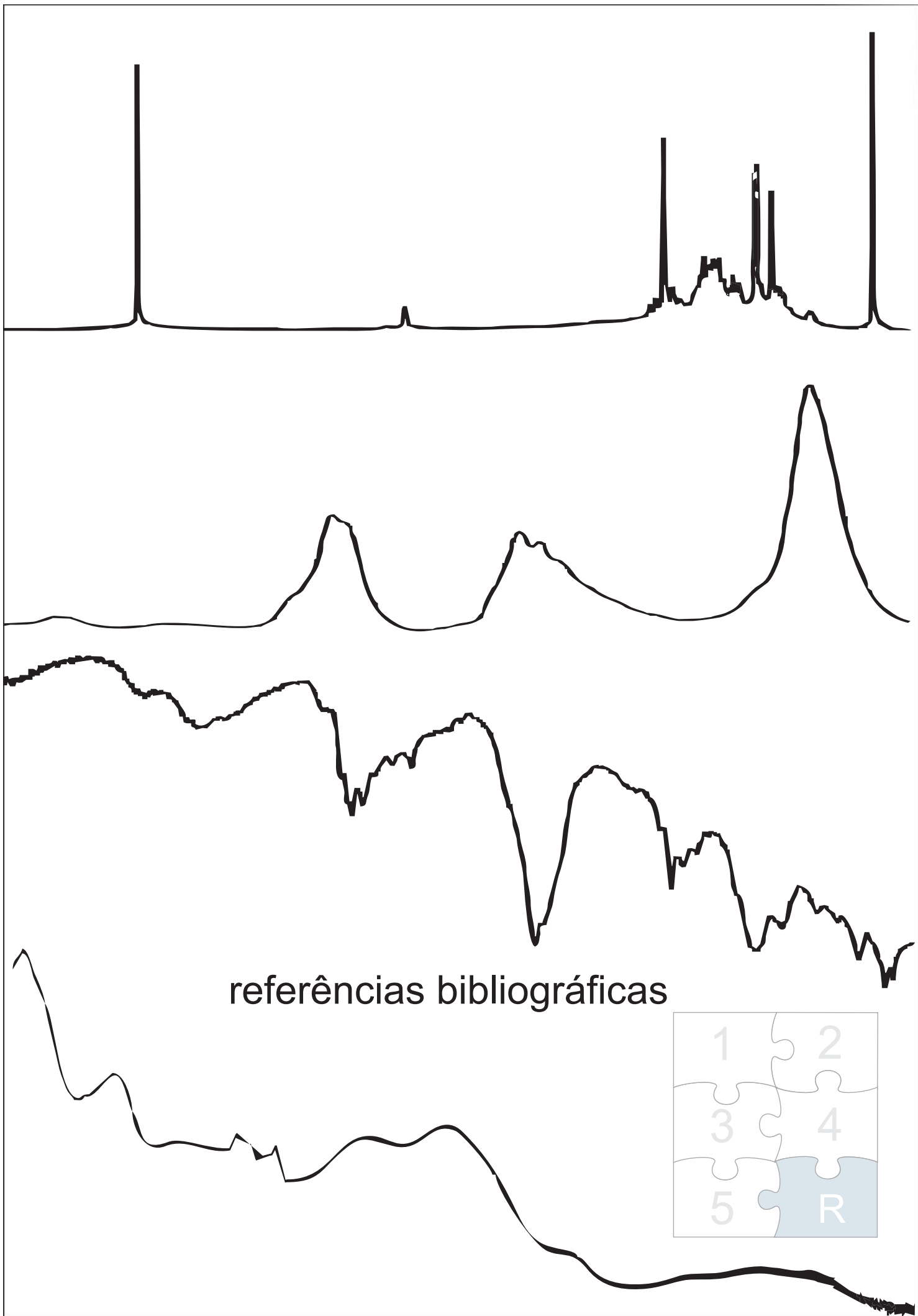
A melhora da relação sinal-ruído em espectros obtidos por um número pequeno de varreduras mostrou ser possível, principalmente usando a técnica MDL, que obteve sempre espectros similares ao espectro de referência, obtido por um número elevado de varreduras.

Enfim, este trabalho mostrou que o método MDL pode ser uma boa alternativa aos métodos padrão de remoção de ruído por Transformada *Wavelet*, em relação à remoção de ruído instrumental real em técnicas espectrométricas.

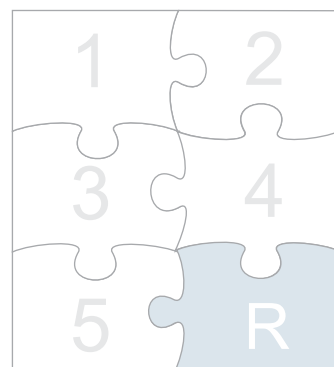
5.1 Propostas futuras

As propostas futuras para a continuidade deste trabalho são:

- construção de modelos quimiométricos de Calibração Multivariada em conjuntos de dados com espectros obtidos pela promediação de um número pequeno de varreduras;
- estudo do comportamento da técnica MDL em medidas analíticas obtidas por outras técnicas, como CLAE e Espectrometria de Emissão Óptica com Plasma Induzido por Laser (LIBS, do inglês: *Laser Induced Breakdown Spectroscopy*).



referências bibliográficas



Referências Bibliográficas

ALSBERG, B. K.; WOODWARD, A. M.; WINSON, M. K.; ROWLAND, J.; KELL, D.B., *Wavelet Denoising of Infrared Spectra*. *Analyst*, **122**: 645, 1997.

ALSBERG, B. K.; WOODWARD, A. M.; KELL, D. B., *An introduction to wavelet transforms for chemometricians: A time-frequency approach*. *Chem. Intell. Lab. Sys.* **37**: 215, 1997.

ARGOULD, F.I.M.; AZEVEDO, F.M.; MARINO NETO, J., *Estudo comparativo entre Funções wavelet e suas diferentes aplicabilidades em reconhecimento de padrões epileptogênicos em eletroencefalograma*. *Revista Brasileira de Engenharia Biomédica*, **20**: 49, 2004.

BEEBE, K. R.; PELL, R. J.; SEASHOLTZ, M. B., *Chemometrics: A Practical Guide*. New York: John Wiley & Sons, 1998.

BRERETON, R.G., *Data Analysis for the Laboratory and Chemical Plant*. Chichester: John Wiley & Sons, 2003.

BARROS-NETO, B.; SCARMINIO, I. S.; BRUNS, R. E. *Como Fazer Experimentos*. Campinas: Editora UNICAMP, 2007.

CAI, C.S., HARRINGTON, P.D., *Different discrete wavelet transforms applied to denoising analytical data*. *J. Chem. Inf. Comput. Sci.* **38**: 1161, 1998.

CERQUEIRA, E. O.; POPPI, R. J.; KUBOTA, L. T.; MELLO, C., *Utilização de filtro de Transformada de Fourier para a minimização de ruídos em Sinais Analíticas*. *Química Nova*, **23**: 690 (2000).

CHAU, F.T.; LIANG, Y.Z.; GAO, J.; SHAO, X.G. *Chemometrics: From basics to Wavelet Transform*. Nova Jersey: Wiley-Interscience, 2004.

DONOHO, D.L.; JOHNSTONE, I.M., *Ideal spatial adaptation via wavelet shrinkage*. *Biometrika*, **81**: 425, 1994.

DONOHO, D.L., *De-noising by soft thresholding*, *IEEE Transactions on Information Theory*, **41**: 613, 1995.

DONOHO, D.L.; JOHNSTONE, I.M.; KERKYACHARIAN, G; PICARD, D., *Wavelet Shrinkage: Asymptopia*. *J. Roy. Statist. Soc.* **57-2**: 301, 1995.

DAUBECHIES, I. *Ten Lectures on Wavelets*. CBMS-NSF Series in Applied Mathematics n. 61, Philadelphia: SIAM, 1992.

FELINGER, A., *Data Analysis and Signal Processing in Chromatography*. Data Handling in Science and Technology, Volume 21, Elsevier, 1998.

FOOD AND DRUG ADMINISTRATION (FDA). *PAT, Guidance for Industry — A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance*. Disponível em: <http://www.fda.gov/cder/OPS/PAT.htm>. Acessado em 23/01/2008.

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**: 874, 2001.

GALVÃO, R.K.H.; DANTAS FILHO, H.A.D.; 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**: 159, 2007.

GELADI, P., *Some recent trends in the calibration literature*. *Chem. Intell. Lab. Sys.* **60**: 211, 2002.

GRUNWALD, P., *A Tutorial Introduction to the Minimum Description Length Principle*. MIT Press, 2004.

LAMBERT, J.B.; MAZZOLA, E.P., *Nuclear Magnetic Resonance Spectroscopy: An Introduction to Principles, Applications, and Experimental Methods*, Nova Jersey: Prentice Hall, 2004.

MALLAT, S., *A wavelet tour of signal processing*. Londres: Academic Press, 1999.

MANGANIELLO, L.; VEGA, C.; RÍOS, A.; VALCÁRCEL, M., *Use of Wavelet Transform to enhance piezoelectric signals for analytical purposes*, *Analytica Chimica Acta*, **456: 93 (2002)**.

MARTENS, H.; NIELSEN, J. P.; ENGELSEN, S. B., *Light Scattering and Light Absorbance Separated by Extended Multiplicative Signal Correction. Application to Near-Infrared Transmission Analysis of Powder Mixtures*, *Analytical Chemistry*, **78:394 (2003)**.

MASSART, D. L.; VANDEGINSTE, B. G. M.; BUYDENS, S. J.; LEWI, P. J.; SMEYERS-VERBEKE, J., *Handbook of Chemometrics and Qualimetrics: Parte A*, Amsterdam: Elsevier, 1997.

MASSART, D. L.; VANDEGINSTE, B. G. M.; BUYDENS, S. J.; LEWI, P. J.; SMEYERS-VERBEKE, J., *Handbook of Chemometrics and Qualimetrics: Parte B*, Amsterdam: Elsevier, 1997.

MISITI, M., MISITI, Y., OPPENHEIM, G., POGGI, J.M. *Wavelet Toolbox User's Guide*. Natick, Mass.: The Mathworks, Inc. 1996.

POORNACHANDRA, S.; KUMARAVEL, N., *Subband-Adaptive Shrinkage for Denoising of ECG Signals*. *Journal on Applied Signal Processing*. **1, 2006**.

RISSANEM, J., *MDL denoising*. *IEEE Transactions on Information Theory*, 46: 7, 2000.

SAITO, N., *Simultaneous noise suppression and signal compression using a library of orthonormal bases and the Minimum Description Length criterion*, in E. Foufoula-Georgiou & P. Kumar, eds., in *Wavelets in Geophysics*, 299-324 New York: Academic Press, 1994

SHAO, X. G.; LEUNG, K. M.; CHAU, F. T., *Wavelet: A new trend in Chemistry*. *Acc. Chem. Res.* **36: 276 (2003)**.

STRANG G., NGUYEN T., *Wavelets and Filter Banks*, Wellesley: Cambridge Press, 1996.

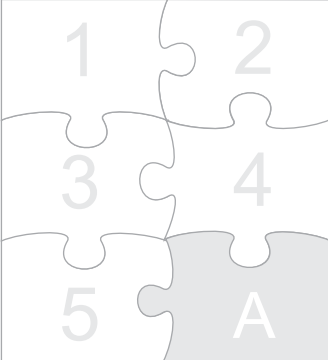
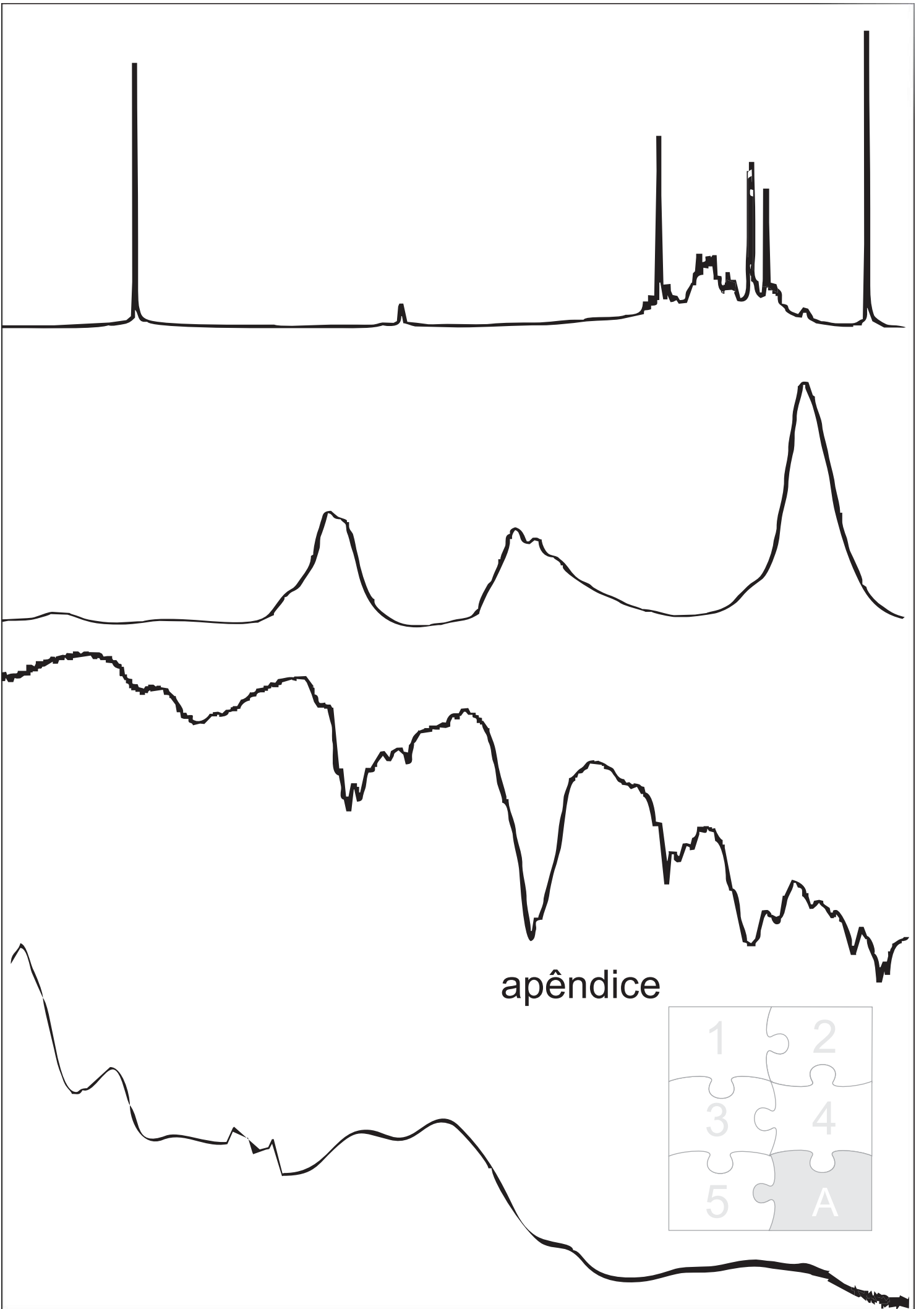
TAVARES, J.F.; QUEIROGA, K.F.; SILVA, M.V.B.; DINIZ, M.F.F.M.; FILHO, J.M.B.; DA-CUNHA, E.V.L.; SIMONE, C.A.; ARAÚJO JÚNIOR, J.X.; MELO, P.S.; HAUN, M.; SILVA, M.S., *ent-Trachylobane Diterpenoids from *Xylopi* *langsdorffiana**. *J. Nat. Prod.* **69: 960, 2007**.

WALCZAK, B.; MASSART, D.L. *Wavelet packet transform applied to a set of signals: A new approach to the best-basis selection*. *Chem. Intell. Lab. Sys.* **38: 39, 1997**.

WALCZAK, B.; MASSART, D.L. *Wavelets: Something for Analytical Chemistry?* *Trends in Analytical Chemistry*. *Chem. Intell. Lab. Sys.* **16: 451, 1997**.

WENTZELL, P.D., BROWN, C.D. *Signal Processing in Analytical Chemistry*. *Encyclopedia of Analytical Chemistry*. Chichester: John Wiley & Sons Ltd, 2000.

WOODWARD, A.M., ALSBERG, B.J., KELL, D.B. *The effect of heteroscedastic noise on the chemometric modelling of frequency domain data*. *Chem. Intell. Lab. Sys.* **40: 101-107, 1998**.



Apêndice

Apêndice A

% Algoritmo para remoção de ruído por Transformada *Wavelet*, através do método MDL e limiarização Rígida.

Escrito por Gledson Emidio José, sob orientação do Prof. Dr. Roberto Kawakami Harrop Galvão.

Programa utilizado: Matlab® 6.5, com o pacote *Wavelet* Toolbox.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
clear
```

```
close all
```

```
Niveis = 7; % número máximo de níveis de decomposição dado pela função wmaxlev.m, do Matlab.
```

% Passo 1: Inicialização dos dados

```
load xtrue % Espectro de referência
```

```
xtrue = Matrix;
```

```
load xruido % Espectros de varredura unitária
```

```
x = Matrix;
```

```
RMSE = []; % Usando apenas a promediação dos espectros
```

```
RMSEmdl = [];
```

```
E1 = sqrt(mean((xtrue - x(1,:)).^2)); % cálculo do RMSE
```

```
RMSE(1) = E1;
```

```
RMSEmdl(1) = E1;
```

```
mdl=[];
```

```
Xmdl=[];
```

```
M=size(x,1);
```

% Passo 2: Geração da biblioteca de wavelets

```
wname1 =
```

```
{'db1','db2','db3','db4','db5','db6','db7','db8','db9','db10','sym4','sym5','sym6','sym7','sym8',  
'sym9','sym10','coif1','coif2','coif3','coif4','coif5'};
```

```
for i = 1:M % loop para as amostras (N=número de amostras)
```

```
    disp(['amostra numero ' num2str(i)]);
```

```
    for w=1:22 % laço para testar todas as wavelets
```

```
        wname=wname1{w};
```

```
        % Calculo do RMSE
```

```
        e = xtrue - x(i,:);
```

```
        RMSE(w,i) = sqrt(mean(e.^2));
```

% Passo 3: Análise wavelet usando a abordagem MDL

```
    % disp('Será feito a escolha da melhor wavelet pelo principio MDL!')
```

% Passo 3.1: Decomposição wavelet

```
[C,L] = wavedec(x(i,:),Niveis,wname);
```

```
C1=C(1:L(2)); % separação dos coeficientes de aproximação
```

```
C2=C(L(2)+1:end); % separação dos coeficientes de detalhe
```

```
[Csort,index] = sort(abs(C2)); % colocando em ordem crescente de valor absoluto
```

```
N=length(C2);
```

% Passo 3.2: Calculo da função custo MDL

```
    for k = 1:ceil((N-1)/2); % k = Número de coeficientes mantidos
```

```
        Cdescartados = Csort(1:N-k);
```

```
        %função custo MDL
```

```
        mdl(k,w) = (3/2)*k*log2(N) + (N/2)*log2(sum(Cdescartados.^2));
```

```
    end % encerra k
```

```
[mdl_min(w),zopt(w)] = min(mdl(:,w));
```

```
Cdenoise = C2;
```

```
Cdenoise(index(1:N-zopt(w))) = 0; % excluindo os espectros abaixo do limiar MDL
```

```
Cdenoise2= [C1 Cdenoise]; % reordenando os espectros
```

```
mdlcoeficientes(w,:)= Cdenoise2;
```

% Passo 3.3: Reconstrução do espectro filtrado

```
[coluna,linha]=min(mdl_min);
```

```
Xmdl(w,:) = waverec(Cdenoise2,L,wname);
```

```
    % Calculo do RMSE
```

```
    RMSEmdl(w,i) = sqrt(mean((xtrue - Xmdl(w,:)).^2));
```

```
end % encerra w
```


Apêndice B

% Algoritmo para remoção de ruído por Transformada *Wavelet*, usando limiar Universal e limiarização Rígida e Suave.

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%  
clear,  
close all
```

```
Niveis = 7; % número máximo de níveis de decomposição dado pela função wmaxlev.m
```

```
wname1 =
```

```
{'db1','db2','db3','db4','db5','db6','db7','db8','db9','db10','sym4','sym5','sym6','sym7','sym8','sym9','sym10','coif1','coif2','coif3','coif4','coif5'};
```

% Passo 1: Carregar os dados

```
load xtrue % Espectro de referência
```

```
xtrue = Matrix;
```

```
load xruido % Espectros de varredura unitária
```

```
x = Matrix;
```

```
RMSE = []; % Usando apenas a promediação dos espectros
```

```
RMSEsoft = []; % Usando a TW e a limiarização Suave com limiar universal
```

```
RMSEhard = []; % Usando a TW e a limiarização Rígida com limiar universal
```

```
E1 = sqrt(mean((xtrue - x(1,:)).^2)); % calculo do RMSE
```

```
RMSE(1) = E1;
```

```
RMSEsoft(1) = E1;
```

```
RMSEhard(1) = E1;
```

```
M=size(x,1);
```

```
for i = 1:M
```

```
    disp(['amostra numero ' num2str(i)]);
```

% Passo 2: Gerar a biblioteca de *wavelets*

```
for w=1:22 % laço para testar todas as wavelets
```

```
    wname=wname1{w};
```

% Calculo do RMSE

e = xtrue - x(i,:);

RMSE(w,i) = sqrt(mean(e.^2));

% Passo 3: Analise wavelet usando o limiar Universal com limiarização Suave e Rígida

% Passo 3.1: Decomposição do espectro

[tmean,L] = wavedec(x(i,:),Niveis,wname);

% Passo 3.2: Estimar desvio-padrão do ruído usando a mediana sob a hipótese de homoscedasticidade.

D1 = detcoef(tmean,L,1); % Primeiro nível de detalhe

sigma_mediana = median(abs(D1))/0.6745;

% Passo 3.3: Calcular o limiar (universal)

n = length(tmean);

lim = sigma_mediana * sqrt(2*log(n));

% Passo 3.4: Aplicar a limiarização Suave e a Rígida

tsoft = tmean;

thard = tmean;

for j=(L(2)+1):n % Os coeficientes de aproximação não são modificados

tj = tmean(j);

if abs(tj) < lim

tsoft(j) = 0;

thard(j) = 0;

else

tsoft(j) = sign(tj)*(abs(tj)-lim);

end

end

% Passo 3.5: Reconstrução do espectro filtrado

[xsoft(w,:)] = waverec(tsoft,L,wname);

[xhard(w,:)] = waverec(thard,L,wname);

% Calculo do RMSE

e = xtrue - xsoft(w,:);

RMSEsoft(w,i) = sqrt(mean(e.^2));

