



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
E INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIAS E ENGENHARIA DE PETRÓLEO

BRUNO CÉSAR ZANARDO HONÓRIO

**Aplicação da Transformada *Wavelet* na
Atenuação de Ruídos e Determinação de Ciclos
Sedimentares em Perfis Geofísicos**

Este exemplar corresponde à redação final da dissertação defendida por Bruno César Zanardo Honório aprovada pela Comissão julgadora em **14/09/2011**.

CAMPINAS

2011



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
E INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIAS E ENGENHARIA DE PETRÓLEO

Aplicação da Transformada *Wavelet* na Atenuação de Ruídos e Determinação de Ciclos Sedimentares em Perfis Geofísicos

Autor: Bruno César Zanardo Honório

Orientador: Prof. Dr. Alexandre Campana Vidal

Co-orientador: Prof. Dr. Emilson Pereira Leite

Programa: Ciências e Engenharia de Petróleo

Área de Concentração: Reservatórios e Gestão

Dissertação de mestrado acadêmico apresentada à Comissão de Pós Graduação em Ciências e Engenharia de Petróleo da Faculdade de Engenharia Mecânica e Instituto de Geociências, como requisito para a obtenção do título de Mestre em Ciências e Engenharia de Petróleo.

Campinas, 2011

SP – Brasil

FICHA CATALOGRÁFICA ELABORADA PELA
BIBLIOTECA DA ÁREA DE ENGENHARIA E ARQUITETURA - BAE - UNICAMP

H759a Honório, Bruno César Zanardo
Aplicação da transformada wavelet na atenuação de
ruídos e determinação de ciclos sedimentares em perfis
geofísicos / Bruno César Zanardo Honório. --Campinas,
SP: [s.n.], 2011.

Orientadores: Alexandre Campane Vidal, Emilson
Pereira Leite.

Dissertação de Mestrado - Universidade Estadual de
Campinas, Faculdade de Engenharia Mecânica e
Instituto de Geociências.

1. Geofísica. 2. Processamento de sinais. 3. Análise
espectral . 4. Wavelet (Matemática). I. Vidal,
Alexandre Campane. II. Leite, Emilson Pereira. III.
Universidade Estadual de Campinas. Faculdade de
Engenharia Mecânica e Instituto de Geociências. IV.
Título.

Título em Inglês: Application of wavelet transform to well log denoising and sedimentary
cyclicality

Palavras-chave em Inglês: Geophysics, Signal processing, Spectral analysis , Wavelet
(Mathematical)

Área de concentração: Reservatórios e Gestão

Titulação: Mestre em Ciências e Engenharia de Petróleo

Banca examinadora: Alexandre Campane Vidal, João Marcos Travassos Romano, Sérgio Sacani
Sancevero

Data da defesa: 14-09-2011

Programa de Pós Graduação: Engenharia Mecânica



UNIVERSIDADE ESTADUAL DE CAMPINAS
FACULDADE DE ENGENHARIA MECÂNICA
E INSTITUTO DE GEOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM
CIÊNCIAS E ENGENHARIA DE PETRÓLEO

DISSERTAÇÃO DE MESTRADO ACADÊMICO

Aplicação da Transformada *Wavelet* na Atenuação de Ruídos e Determinação de Ciclos Sedimentares em Perfis Geofísicos

Autor: Bruno César Zanardo Honório
Orientador: Prof. Dr. Alexandre Campanhe Vidal
Co-orientador: Prof. Dr. Emilson Pereira Leite

A Banca Examinadora composta pelos membros abaixo aprovou esta Dissertação:

Prof. Dr. Alexandre Campanhe Vidal, Presidente
DGRN/IG/UNICAMP

Prof. Dr. João Marcos Travassos Romano
DMO/FECC/UNICAMP

Dr. Sérgio Sacani Sancevero
ROXAR

Campinas, 14 de setembro de 2011

DEDICATÓRIA

Aos meus pais, Zé e Kuca, cujo altruísmo guia minha conduta.

Aos meus irmãos, Lucas e Negrão; um de sangue, ambos de coração.

AGRADECIMENTOS

A conclusão deste trabalho representa o fim de uma etapa que teria sido muito mais difícil se não fosse a companhia, física ou não, de algumas pessoas, as quais sou extremamente grato.

Agradeço aos meus pais, pois, além de sempre me apoiarem e acreditarem nos meus anseios e sonhos, me passaram os valores pessoais que serão minha maior herança. Admiro vocês profundamente.

Ao meu orientador, professor Dr. Alexandre Vidal, pelo acolhimento no início desta jornada, discussões, correções e a confiança depositada. Espero ter respondido à altura e continuar lado a lado nessa empreitada.

Aos amigos que passaram por esta estadia campineira, às vezes dividindo uma sala de aula, outras o lar, idéias, viagens ou simplesmente um ombro amigo: Brunão, João Paulo, Tiagones, Meco, Renan, Brazuk, Glaucão, Sups, Letícia, Naka e Polderrrrman. Nossas lembranças jamais passarão.

Aos companheiros que conheci e convivi nesse caminho da geociência: Luis Fernando Cypriano (Chaveiro) pelas conversas que me direcionaram no início dessa caminhada, Alexandre Milagres pelos cafezinhos e prosas no meio da tarde, a Lenita Fioriti pelas aulas de geologia, Rodrigo S. Portugal, Daniel L. Macedo, José Jadson, Evângela P. Alves e Luís d'Afonseca (Akiles) pelo tempo e esforço despendido em nossos seminários de geofísica.

Aos companheiros de ofício, Alê (Champs), Juliana Bueno, Ancilla Carvalho, Michele C. Kuroda e Paola Faccini. Um agradecimento especial ao pesquisador Dr. Rodrigo Drummond pelas discussões, conselhos e revisões em meu exame de qualificação, bem como pela ajuda na elaboração de alguns programas e ao professor Emilson P. Leite pela cooperação na realização deste trabalho.

Agradeço aos professores e demais funcionários da comunidade DEP/DGRN, em especial à Sônia A. G. de Oliveira e Michelle C. Fulaneto, e ao pessoal do CEPETRO, Sr. Amador A. de Souza, José A. T. Junior e Nathassia M. Marques que sempre se mostraram solícitos em ajudar na burocracia necessária.

À PETROBRAS pelo apoio financeiro fornecido através do Programa de Formação de Recursos Humanos junto à rede de Caracterização e Modelagem Geológica de Reservatórios (CARMOD).

Por fim, à Lídia Bertolo e sua família. Uma pelo carinho, confiança e bons momentos vividos nesse período (mesmo longe, você está presente) e à outra por me proporcionar um ambiente familiar mesmo eu estando distante de minha casa. Isso foi de grande valor pra mim.

Os homens são miseráveis porque não sabem ver nem entender os bens que estão ao seu alcance.

Pitágoras

Mudam as caras, mas todas podem ter as mesmas expressões. Mudam as línguas, mas todas têm suas palavras carinhosas e os seus calões.

As orações e os deuses também variam, mas o alívio que eles trazem vem do mesmo lugar. Mudam os olhos e tudo que eles olham, mas quando molham, todos olham com o mesmo olhar.

Seja onde for uma lágrima de dor tem apenas um sabor e uma única aparência. A palavra saudade só existe em português, mas nunca faltam nomes se o assunto é ausência.

A solidão apavora, mas a nova amizade encoraja. E é por isso que a gente viaja procurando um reencontro, uma descoberta que compense a nossa mais recente despedida.

Nosso peito muitas vezes aperta. Nossa rota é incerta.

Mas o que não incerto na vida?

Gabriel o Pensador

RESUMO

HONÓRIO, Bruno César Zanardo, Aplicação da Transformada *Wavelet* na Atenuação de Ruídos e Determinação de Ciclos Sedimentares em Perfis Geofísicos. Campinas, Faculdade de Engenharia Mecânica e Instituto de Geociências, Universidade Estadual de Campinas, 2011. 105 p. Dissertação de Mestrado.

Perfis geofísicos de poços são importantes fontes de informação para o estudo de rochas em subsuperfície, sendo comumente utilizados na caracterização de reservatórios de petróleo. É sabido que os sinais obtidos como respostas do meio geológico contêm ruídos que podem afetar a interpretação do objeto estudado e que a transformada *wavelet* (WT) é melhor adaptada que a transformada de Fourier (FT) para a análise de sinais não-estacionários, como aqueles obtidos da perfilagem geofísica. Por outro lado, existem diversos parâmetros que devem ser considerados quando se trabalha com a WT, tais como a escolha da função *wavelet* base (*wavelet* mãe), o nível de decomposição, assim como a função e as regras que "controlam" como e quais coeficientes serão utilizados para reconstrução do sinal.

Este estudo analisa o processo de atenuação de ruídos em perfis geofísicos de poços através da transformada *wavelet* ortogonal. Numa primeira abordagem, foi estudado o processo de atenuação em sinais sintéticos e então, foi estudado o processo em perfis geofísicos reais. Uma vez que os dados de perfis de poços são geralmente utilizados para a classificação de litologias, foi proposto um método associado com o algoritmo de classificação K-Vizinhos Mais Próximos (KNN) para investigar como as diferentes combinações dos parâmetros afetam os sinais de saída e o seu desempenho na classificação litológica, tornando assim um processo conduzido pelos dados. Foi investigada as funções de *thresholding* tradicionalmente empregadas (*hard e soft threshold*) bem como uma recente abordagem (*customized threshold*), sendo esta a que proporcionou melhores resultados. O potencial da transformada *wavelet* como ferramenta para auxiliar na interpretação geológica é evidenciado pela identificação de importantes características geológicas, como regiões de cimentação carbonática e a determinação de ciclos sedimentares, do Campo de Namorado, Bacia de Campos, Brasil.

Palavras-Chave: Geofísica, Processamento de sinais, Análise espectral, Wavelet

ABSTRACT

HONÓRIO, Bruno César Zanardo, Application of *Wavelet* Transform to Well Log Denoising and Sedimentary Cyclicity. Campinas, Faculdade de Engenharia Mecânica, Universidade Estadual de Campinas, 2011. 105 p. Dissertação de Mestrado.

Geophysical well logs are an important tool for the characterization of subsurface rocks, being commonly used in the study of reservoir geology. It is well known that signals obtained as responses from geological media contain noise that can affect their interpretation, and that wavelet transform (WT) is more suitable than the Fourier transform (FT) to denoise non-stationary signals, as the ones obtained from well logs. On the other hand, there are several parameters that must be considered when working with the WT, such as the wavelet basis function choice (mother wavelet), the decomposition level and also the function and rules that “control” which and how the coefficients will be used for signal reconstruction.

This study analyzes the process of denoising geophysical well log data by orthogonal wavelet transform. In a first approach, it was studied the denoising process in synthetic signals and then, the process in real geophysical well logs. Since the well log data are usually used in lithology classification, we propose a method associated with the K-Nearest Neighbor (KNN) classification algorithm to investigate how different combinations of parameters affect the output signals and its performance in the lithological classification, thus making a data driven process. It was evaluated the thresholding function traditionally used (*hard e soft threshold*) as well a recent approach (*customized threshold*), this being the one that provided better results. The potential of the wavelet transform as a tool to aid geological interpretation is evidenced by the identification of important geological features, such as regions of carbonate cementation and determination of sedimentary cycles of the Namorado Field, Campos Basin, Brazil.

Key Words:

Geophysics, Signal Processing, Spectral analysis, Wavelet

SUMÁRIO

DEDICATÓRIA	iv
AGRADECIMENTOS	v
RESUMO.....	vii
ABSTRACT	ix
SUMÁRIO.....	x
LISTA DE FIGURAS.....	xiii
LISTA DE TABELAS	xv
LISTA DE NOMENCLATURAS	xvii
Capítulo 1	1
INTRODUÇÃO	1
1.1. Motivação e Objetivos	2
Capítulo 2	5
FUNDAMENTAÇÃO TEÓRICA.....	5
2.1. Introdução	5
2.2. Análise Espectral.....	9
2.3. Análise <i>Wavelet</i>	15
2.3.1. Transformada <i>Wavelet</i> Contínua	16
2.3.2. Transforma <i>Wavelet</i> Discreta	23
2.3.3. Transformada <i>Wavelet</i> Ortogonal.....	25

2.3.4. Análise Multirresolução.....	26
Capítulo 3	29
ATENUAÇÃO DE RUÍDOS E DETERMINAÇÃO DE CICLOS SEDIMENTARES EM PERFIS GEOFÍSICOS.....	29
3.1. Introdução	29
3.2. Método	31
3.2.1. Escolha da Função <i>Wavelet</i>	32
3.2.2. Métodos de <i>Thresholding</i>	35
3.2.3. Validação	43
3.2.4. Fluxograma do método	46
Capítulo 4	49
RESULTADOS	49
4.1. Dados sintéticos	49
4.2. Perfis de poços	51
Capítulo 5	67
CONCLUSÕES	67
Capítulo 6	69
REFERÊNCIAS.....	69
ANEXO A.....	73
APÊNDICE A	83
A1 - K-Vizinhos Mais Próximos.....	83

LISTA DE FIGURAS

Figura 2.1: Função original e a decomposição via FT (Misiti et al., 1997)	10
Figura 2.2: a) Sinal composto por duas ondas senoidais; b) Conteúdo de frequência obtido via FT	11
Figura 2.3: Sinais compostos por duas senóides de 10 e 20 (Hz) e seus respectivos espectros obtidos via FT.....	12
Figura 2.4: Sinais compostos por duas senóides de 10 e 20 (Hz) e os respectivos espectros obtidos via STFT.....	14
Figura 2.5: Função original e a decomposição via WT (Misiti et al., 1997).....	15
Figura 2.6: Parte real e imaginária da <i>wavelet</i> Morlet e espectro de frequência para diferentes escalas (Kumar & Foufoula, 1997)	16
Figura 2.7: Plano tempo-frequência da transformada <i>wavelet</i> (modificado de Protázio, 2002)... ..	19
Figura 2.8: Obtenção dos coeficientes <i>wavelet</i> (Misiti et al., 1997).	20
Figura 2.9: Fluxograma da CWT.....	20
Figura 2.10: Sinais compostos por duas senóides de 10 e 20 (Hz) e os respectivos espectros obtidos via CWT (<i>wavelet</i> Morlet).....	21
Figura 2.11: Determinação da frequência central da <i>wavelet</i> Morlet.....	22
Figura 2.12: Sinal no domínio do tempo e análise do conteúdo de frequências obtidos através da FT, STFT e WT, respectivamente	26
Figura 2.13: Ilustração da análise multirresolução. S: Sinal original; cAj e cDj: componentes de Aproximação e Detalhe no nível j (Misiti <i>et al.</i> , 1997).....	27
Figura 3.1: Ilustração de um registro estratigráfico composto por três ondas senoidais de diferentes frequências e amplitudes (Weedon, 2003).....	30
Figura 3.2: a) Sinal semelhante à <i>wavelet</i> Db2 sem e com ruído branco Gaussiano adicionado (SNR =15dB).....	33
Figura 3.3: a) <i>Wavelet</i> Daubechies2; b) <i>Wavelet</i> Symlet2	34
Figura 3.4: Sinal ruidoso (vermelho) e sinal com ruído atenuado (roxo)	35
Figura 3.5: Atenuação de ruídos através de diferentes abordagens (Weedon, 2003).....	36

Figura 3.6: <i>Hard, Soft e Customized Thresholding</i>	39
Figura 3.7: Curva de Lorenz.....	41
Figura 3.8: a) <i>Thresholding</i> baseado no conceito da curva de Lorenz; b) Curva de Lorenz para os coeficientes <i>wavelet</i>	43
Figura 4.1: Sinais sintéticos antes e pós processamento	50
Figura 4.2: Eficiência da função de <i>threshold</i> customizada para o sinal 'Heavy sine'. O ponto de mínimo define o melhor alfa (menor MSE)	51
Figura 4.3: Taxa média de corretas classificações para diferentes valores de alfa no “ <i>customized threshold</i> ”	58
Figura 4.4: Perfis de Raios Gama original e filtrado e os respectivos escalogramas	59
Figura 4.5: Desempenho do KNN utilizando 50% das amostras como treino	60
Figura 4.6: Classificação litológica baseada no perfil original (esquerda) e perfil filtrado (direita) comparados com o testemunho (centro) (50% das amostras como treino)	60
Figura 4.7: Desempenho do KNN utilizando 60% das amostras como treino	61
Figura 4.8: Classificação litológica baseada no perfil original (esquerda) e perfil filtrado (direita) comparados com o testemunho (centro) (60% das amostras como treino)	61
Figura 4.9: Desempenho do KNN utilizando 70% das amostras como treino	62
Figura 4.10: Classificação litológica baseada no perfil original (esquerda) e perfil filtrado (direita) comparados com o testemunho (centro) (70% das amostras como treino)	62
Figura 4.11: Perfis geofísicos e escalograma do perfil GR filtrado (GRf). As setas amarelas e vermelhas indicam regiões de intercalações de arenito/folhelho e cimentação carbonática, respectivamente.	63
Figura 4.12: Arenito médio gradado destacado após eliminação do ruído	64
Figura 4.13: a) Ciclos sedimentares identificados por Souza Jr. (1997) e relacionados com o nível 1 do escalograma filtrado, em b).	66

LISTA DE TABELAS

Tabela 3-1: Entropia de Shannon	34
Tabela 3-2: Regras de escolha para os valores de <i>Thresholding</i>	40
Tabela 3-3: Litologia do poço estudado	45
Tabela 4-1: Erro médio quadrático para diferentes sinais e esquemas de atenuação de ruído.....	49
Tabela 4-2: Corretas classificações utilizando 50% das amostras como treino e diferentes esquemas de atenuação de ruído (Ruído branco Gaussiano).....	52
Tabela 4-3: Corretas classificações utilizando 50% das amostras como treino e diferentes esquemas de atenuação de ruído (<i>Scaled white noise</i>).....	54
Tabela 4-4: Corretas classificações utilizando 50% das amostras como treino e diferentes esquemas de atenuação de ruído (<i>Nonwhite noise</i>)	55
Tabela 4-5: Entropia de Shannon para o perfil GR	57
Tabela A-1: Corretas classificações utilizando 60% das amostras como treino e diferentes esquemas de atenuação de ruído (Ruído branco Gaussiano).....	73
Tabela A-2: Corretas classificações utilizando 60% das amostras como treino e diferentes esquemas de atenuação de ruído (<i>scaled white noise</i>).....	74
Tabela A-3: Corretas classificações utilizando 60% das amostras como treino e diferentes esquemas de atenuação de ruído (<i>nonwhite noise</i>).....	76
Tabela A-4: Corretas classificações utilizando 70% das amostras como treino e diferentes esquemas de atenuação de ruído (Ruído branco Gaussiano).....	77
Tabela A-5: Corretas classificações utilizando 70% das amostras como treino e diferentes esquemas de atenuação de ruído (<i>scaled white noise</i>).....	79
Tabela A-6: Corretas classificações utilizando 70% das amostras como treino e diferentes esquemas de atenuação de ruído (<i>nonwhite noise</i>).....	80

LISTA DE NOMENCLATURAS

Letras gregas:

Ψ – *wavelet* mãe

$\Psi_{a,b}$ – *wavelet* filha ou família de *wavelets* no caso de todo o domínio de a e b

Φ – função de escala

$\delta(\cdot)$ – função de limiarização (*thresholding*)

Δ – período de amostragem

σ – desvio padrão

λ – limiar de corte no processo de atenuação de ruído

α – parâmetro de suavização da função de *thresholding* customizada

γ – limiar de corte da função de *thresholding* customizada

Letras latinas:

c – coeficiente *wavelet*

C_ψ – constante de admissibilidade

F_c – frequência central

H – entropia de Shannon

E – entropia de Shannon padronizada

L^2 – espaço de todas as funções quadraticamente integráveis

W – matriz ortonormal

R – conjunto dos números reais

Z – conjunto dos números inteiros

Subscritos:

a, j – parâmetros de escala

b, k – parâmetros de deslocamento

Abreviações:

Coifn - *Wavelet Coiflet* com n momentos nulos

CWT - Transformada *Wavelet* Contínua (*Continuous Wavelet Transform*)

Dbn – *Wavelet* Daubechies com n momentos nulos

DWT - Transformada *Wavelet* Discreta (*Discrete Wavelet Transform*)

FSF - Filtros de Frequência Seletiva (*Frequency-Selective Filtering*)

FT - Transformada de Fourier (*Fourier Transform*)

GR - Raios-Gama (*Gamma Ray*)

GRf – perfil de Raios-Gama filtrado

KNN - K-Vizinhos Mais Próximos (*K-Nearest Neighbor*)

MRA - Análise Multirresolução (*Multiresolution Analysis*).

QMF - Filtros-espelho em Quadratura (*Quadrature Mirror Filter*)

Sign – Função sinal

Symn – *Wavelet* Symlet com n momentos nulos

STFT - Transformada de Fourier de Curta Duração (*Short Time Fourier Transform*)

TF - Filtros de Limiarização (*Threshold Filtering*)

WF - Filtros de Wiener (*Wiener Filtering*)

WT - Transformada *Wavelet* (*Wavelet Transform*)

CAPÍTULO 1

INTRODUÇÃO

Investigações geofísicas do interior da Terra envolvem a coleta, sobre ou próxima à sua superfície, de medidas que sofrem influência da distribuição interna de algumas propriedades físicas. A análise dessas medidas permite avaliar como essas propriedades variam vertical e lateralmente.

Uma ampla gama de levantamentos geofísicos existe, para cada um dos quais o método é sensível a uma propriedade física particular. Os perfis geofísicos de poços são uma fonte comum de dados para a avaliação de depósitos sedimentares em subsuperfície. Em particular, as técnicas de perfilagem são bastante utilizadas na exploração de hidrocarbonetos, uma vez que fornecem importantes informações *in situ* de possíveis rochas-reservatório. Também são utilizadas em explorações dos recursos hídricos pelas mesmas razões.

A maneira mais simples de apresentar esses dados é traçar um gráfico mostrando a variação da quantidade medida com relação a distância ou tempo. O gráfico apresentará alguma forma de onda mais ou menos complexa, mostrando variações da propriedade física referente à própria geologia, sobrepostas à informações não-geológicas, à imprecisão instrumental e erros na coleta de dados. Somado a isso, a precisão desses sinais pode ser incerta devido à dificuldade de interpolação da curva entre os pontos coletados (Philip, et al., 2002). Portanto, a análise de sinais representa um aspecto essencial do processamento de dados geofísicos, visando a separação do sinal referente à geologia das informações indesejadas e, com isso, a interpretação do sinal em termos da estrutura da rocha.

Geralmente, os sinais obtidos como resposta do meio geológico contêm características não-estacionárias ou transitórias, incluindo ciclicidades, tendências e mudanças abruptas, características essas que muitas vezes são a parte mais importante de um sinal, e que são de difícil visualização através de uma análise direta dos dados.

Para muitas aplicações, portanto, a melhor representação dos sinais é feita no domínio da frequência, o que é obtido através de métodos de análise espectral, sendo a transformada de Fourier o método tradicionalmente empregado. No entanto, essa técnica não apresenta as características apropriadas para a análise de dados de natureza não-estacionária, uma vez que se perde informação sobre a localização temporal do evento na transformação de domínios.

Para suprir essa limitação, uma modificação no método, transformada de Fourier de curta duração (*Short Time Fourier Transform - STFT*), permite a representação do sinal em ambos os domínios, analisando segmentos do sinal através de uma “função janela”. Uma janela pequena possibilita boa localização temporal, sendo utilizada para capturar comportamentos transientes, enquanto uma janela grande fornece precisão do conteúdo de frequência à custa da perda da informação temporal do evento. O comprimento da janela determina a resolução tempo-frequência, e, portanto, uma vez estabelecido o tamanho da função janela, a resolução tempo-frequência é fixada para toda a análise.

Nas últimas três décadas, a transformada *wavelet* se desenvolveu e ganhou aceitação em diversas áreas, como análises sísmicas, processamento de imagens, óptica, estudo de turbulência, mecânica quântica, caos, fractal, pesquisas médicas, dentre outras (Lau & Weng, 1995).

A análise *wavelet* se caracteriza por ser um método de análise multirresolução (Mallat, 1989). Através dessa abordagem, as propriedades locais do sinal podem ser acessadas em ambos os domínios e, diferentemente da STFT que possui resolução fixa, a WT permite a análise em diversos níveis de resolução simultaneamente. Com isso, uma função ou sinal pode ser decomposto em componentes de baixa, média e alta frequência, sendo que cada componente pode ser estudada individualmente.

1.1.Motivação e Objetivos

Este estudo analisa o processo de atenuação de ruídos em perfis geofísicos de poços através da transformada wavelet ortogonal. Muito embora a WT apresente propriedades mais adequadas que a FT para a análise de sinais não-estacionários, existem diversos parâmetros que devem ser considerados quando se a utiliza, tais como a escolha da função wavelet base (wavelet

mãe), o nível de decomposição, as função e as regras que "controlam" como e quais coeficientes serão utilizados para reconstrução do sinal. Foi investigada as funções de thresholding tradicionalmente empregadas (*hard e soft threshold*) bem como uma recente abordagem (*customized thresahold*).

Uma vez que os dados de perfis de poços são geralmente utilizados para a classificação de litologias, foi proposto um método associado com o algoritmo de classificação KNN para investigar como as diferentes combinações dos parâmetros afetam os sinais de saída e o seu desempenho na classificação litológica. Considerando que a maior taxa de corretas classificações será obtida quanto maior a fidelidade do sinal em relação às características do objeto estudado, o método proposto permite avaliar a combinação dos parâmetros de forma quantitativa. Qualitativamente, objetivou-se a interpretação do resultado da filtragem em termos geológicos. Como resultado dessas duas abordagens, visou-se a obtenção de um dado geologicamente mais representativo.

No segundo capítulo, são apresentados os fundamentos teóricos da análise espectral segundo as ópticas da análise de Fourier e da análise wavelet. No terceiro capítulo, são apresentadas as questões que envolvem o problema de atenuação de ruídos utilizando a WT, de modo a considerar os principais parâmetros envolvidos e a estratégia proposta para direcionar tal escolha. O quarto capítulo apresenta os resultados obtidos para os sinais modelados assim como para perfis geofísicos reais, sendo que neste ultimo caso, buscou-se a interpretação do resultado de filtragem em termos geológicos. Por fim, são apresentadas as principais conclusões.

CAPÍTULO 2

FUNDAMENTAÇÃO TEÓRICA

“Simplificando ao extremo, diremos que uma wavelet é uma idealização matemática de uma nota musical. Da mesma forma que se representa uma composição musical sob a forma de uma série de notas colocadas em uma partitura, pode-se imaginar a utilização de “notas matemáticas” na representação de certos objetos matemáticos, tais como funções ou sinais”. (Alex Grossmann, Encyclopédie Universalis)¹

2.1. Introdução

As *wavelets*, que ainda não possuíam esse nome, foram introduzidas em 1910 pelo matemático Alfred Haar, na sua tese de doutorado sobre a teoria do sistema de funções ortogonais (Haar, 1910). A função que ele descreveu é agora chamada de *wavelet* Haar.

No entanto, foi dentro do contexto da geofísica que as *wavelets* se desenvolveram. O francês Jean Morlet, então geofísico da companhia petrolífera *Elf Aquitaine*, estudava a resposta de pulsos sísmicos, os quais são amplamente utilizados na caracterização geológica de reservatórios de petróleo. Devido às limitações da transformada de Fourier na análise de sinais sísmicos, Morlet cria um novo método para decompor esses sinais no que chamou de “*wavelets* de forma constante” (*wavelets of constant shape*), que posteriormente se tornou “*wavelet* Morlet” (Morlet et al., 1982a; Morlet et al., 1982b). Sua formalização foi dada posteriormente por Grossmann & Morlet (1984) e Goupillaud et al. (1984). O termo utilizado originalmente era em

¹ En schématisant à l'extrême, nous dirons qu'une ondelette est l'idéalisation mathématique d'une note de musique. De même que l'on représente une œuvre musicale sous forme de séries de notes portées sur une partition, de même on peut songer à utiliser des “notes mathématiques” pour représenter certains objets mathématiques, tels des fonctions ou des signaux.

francês “*ondelette*”, cujo significado é “pequena onda”. Transferindo-o para o inglês e traduzindo “onde” para “*wave*”, obteve-se o termo amplamente conhecido como “*wavelet*”².

Stephane Mallat incorporou às *wavelets* um salto adicional. Ele descobriu relações entre filtros-espelho em quadratura (*Quadrature Mirror Filter - QMF*), algoritmos piramidais e *wavelets* de base ortonormal (Mallat, 1989). O método de análise multirresolução por ele proposto inseriu as *wavelets* no *mainstream* do processamento de sinais digitais (Guan & Du, 2004). Inspirado em parte por esses resultados, Yves Meyer construiu a primeira *wavelet* “não-trivial”. Ao contrário das *wavelets* Haar, as *wavelets* Meyer são continuamente diferenciáveis, porém não possuem suporte compacto.

Ingrid Daubechies utilizou os trabalhos de Mallat para construir um conjunto de funções *wavelet* de base ortonormal (Daubechies, 1988), que são talvez as mais elegantes e se tornaram a pedra angular das aplicações das *wavelets* hoje (Graps, 1995).

Após esse desenvolvimento, as *wavelets* encontraram diversas aplicações nas mais diferentes áreas, das quais é dado um foco principal dentro do contexto da geofísica.

Lau & Weng (1995) fornecem uma introdução e um tutorial para se estudar séries temporais de registros climáticos segundo a óptica da WT. Nesse trabalho, os autores evidenciam a potencialidade da análise *wavelet* para extrair informações relevantes sobre as oscilações da temperatura da Terra.

Goel & Vidakovic (1995) consideraram algumas questões referentes à WT e propuseram um esquema de filtragem baseado na curva de Lorenz. A curva de Lorenz originalmente foi aplicada dentro do contexto da economia para avaliar a distribuição de renda numa sociedade. Dentro do contexto da WT, o método utiliza a distribuição de energia dos coeficientes *wavelet* e estabelece um balanço entre o ganho de informação sobre um evento particular e a perda no número de coeficientes *wavelet*. Portanto, fornece um limiar ótimo do número de coeficientes que devem ser mantidos para a reconstrução do sinal. Os autores relatam que a vantagem de se

² Este texto mantém o termo ‘wavelet’ já consagrado em toda literatura apesar de concordar com a tradução ‘ondaletas’ proposta por Morettin (www.ime.usp.br/~pam/)

utilizar esse esquema está no fato de não ser necessária nenhuma suposição sobre a distribuição do ruído.

Motivados por esse trabalho, Katul & Vidakovic (1995) consideraram o problema relacionado à turbulência atmosférica sobre duas ópticas diferentes: *thresholding* universal proposto por Donoho & Johnstone (1994) e o *thresholding* baseado na curva de Lorenz. Nesse trabalho, os autores buscam separar turbilhões coerentes anisotrópicos (*attached eddies*) dos turbilhões de escalas bem menores estatisticamente isotrópicos (*detached eddies*). Os autores defendem que esse esquema de *thresholding* é mais adequado para a análise de turbulência do que os baseados nos modelos clássicos.

Ainda numa comparação sobre esquemas de *thresholding*, Goring (2004) estudou ciclos oceânicos de longa duração sob a óptica da WT. Esses ciclos são obtidos eliminando o efeito da maré em registros do nível do mar. O registro resultante contém não somente os ciclos oceânicos, mas também o ruído decorrente do erro do instrumento e do alias das ondas de curto período. O autor então avalia o processo de eliminação de ruído adotando dois critérios para o valor do limiar de corte dos coeficientes *wavelet*: um baseado na curva de Lorenz e outro baseado na regra de *threshold* universal. No conjunto de dados avaliado pelo autor (Marsden Point), a maior parte da energia do ruído é concentrada no primeiro nível de detalhe, e o *threshold* universal se mostra mais apropriado. No entanto, o autor conclui que este resultado não é necessariamente aplicável à outros conjunto de dados, devendo ser avaliado os pormenores de cada conjunto de dados em particular.

Deighan & Watts (1997) demonstraram que as propriedades de localização tempo-frequência obtidas pela WT se mostraram eficazes na supressão do ground roll³. Quando comparado com as técnicas de filtragem convencionais que utilizam a transformada de Fourier, o método baseado na WT proporcionou melhores resultados, sem aumento no tempo

³ Um tipo de ruído coerente gerado pelas ondas de superfície, tipicamente de baixa velocidade, baixa frequência e alta amplitude (Oilfield glossary, Schlumberger).

computacional. Eles sugerem que trabalhos adicionais sejam feitos com o intuito de selecionar a *wavelet* ótima para a filtragem de dados sísmicos, uma vez que existem diversas possibilidades.

Prokoph & Agterberg (2000) analisaram através da WT dados de Raios-Gama (Gamma Ray - GR) oriundos de perfis de poços. Além de localizar descontinuidades e determinar ciclos sedimentares de alta frequência, eles encontraram uma correlação entre os ciclos de predominância de GR e os ciclos de Milankovitch⁴, sugerindo que os ciclos climáticos foram fatores importantes para a deposição marinha do membro Egret, Canadá. Em trabalho semelhante, Vega (2003) constatou essa mesma correlação no campo de gás ao leste do Mar da Irlanda.

Lyrio et al. (2004) propuseram uma técnica baseada na WT 1D para a filtragem de um conjunto de dados gravimétricos. O procedimento é baseado na análise da distribuição da energia em diferentes escalas. Foi demonstrado que, para dados dessa natureza, a energia do sinal é bem separada da energia das componentes de ruído. Eles concluem que abordagens convencionais para estimar o valor do *threshold* são ineficazes nas aplicações de dados gravimétricos, onde o ruído é desconhecido e dependente da localização. A escolha do valor do *threshold* baseada na curva de energia acumulada e a escala a partir da qual se faz a limiarização dos coeficientes *wavelet*, fornece uma forma eficiente para atenuar o ruído de qualquer conjunto de dados gravimétricos.

Yu et al. (2010) demonstraram que a WT é eficiente no processo de supressão de ruído em perfis de poços. Além disso, constataram que perfis acústicos que passaram por esse processo são

⁴ Os ciclos de Milankovitch são ritmos da órbita da Terra e estão diretamente relacionados com ciclos climáticos. Três causas do ciclo de Milankovitch são a precessão, obliquidade e excentricidade na órbita da Terra. A precessão pode ser entendida fazendo-se uma analogia entre a rotação da Terra e o movimento de um pião: o eixo de rotação varre um cone com um determinado período, no caso da Terra, de cerca de 21.000 (21 ky) anos. Obliquidade se refere a mudanças na inclinação do eixo da Terra com relação ao plano das órbitas (41ky anos). A excentricidade é relacionada a mudanças na forma da órbita da Terra ao redor do sol, de mais circular para mais elíptica (100 e 400 ky). Essas variáveis atuando em conjunto controlam a radiação solar e, com isso, afetam o clima global, resultando em alterações do nível do mar (Plint, et al. 1993).

superiores aos dados originais em revelar informações geológicas, tais como camadas contendo gás.

Demirel & Anbarjafari (2011) propuseram um método para aumentar a resolução de imagens de satélite. A técnica por eles proposta, utiliza a DWT para decompor a imagem de entrada em diferentes sub-bandas. Então, as imagens na sub-banda de alta frequência e a imagem de entrada de baixa resolução são interpoladas. Ao combinar essas imagens, uma nova imagem de melhor resolução é obtida. Os resultados quantitativos (pico na razão sinal-ruído e erro médio quadrático) e visuais mostraram a superioridade da técnica proposta sobre as convencionais e técnicas “estado-da-arte” para o melhoramento de resolução em imagens.

2.2. Análise Espectral

As séries temporais são entendidas como qualquer sequência de medições ou observações coletadas em uma ordem específica (Weedon, 2003). Normalmente, as medidas são feitas em intervalos constantes de alguma escala, tais como a espessura de camadas de rocha, a distância geográfica, tempo, número de bandas de crescimento, etc. Alguns autores se referem aos dados coletados em relação a uma escala de profundidade ou espessura como "séries de profundidade". No entanto, por razões históricas, séries temporais é o termo matemático correto (Schwarzacher, 1975), termo este que será adotado para se referir ao conjunto de dados analisados, sendo eles de natureza espacial ou temporal.

Em ciências naturais, é muito comum a idéia de decompor um problema em suas partes mais simples, estudá-las individualmente, e então construir um modelo do problema a partir do entendimento de cada uma dessas componentes. Esse pensamento tem sua origem na obra de Renè Descartes, “Discurso sobre o método”, e molda, ainda hoje, o padrão de investigação de um fenômeno ou processo, dentro e fora do meio científico.

A idéia fundamental por trás da análise *wavelet* é examinar os dados de acordo com a escala. Farge et al. (1993) definem a análise *wavelet* como um microscópio matemático. As *wavelets* são funções que satisfazem certos requisitos matemáticos e são usadas na representação

de dados ou outras funções. Até então, essa idéia não é nova: Joseph Fourier, no início dos anos 1800, descobriu que poderia sobrepor senos e cossenos para representar outras funções. Segundo sua teoria, conhecida como análise de Fourier⁵, qualquer função periódica $f(x)$ pode ser representada como uma soma ponderada de simples formas de onda, no caso, senos e cossenos ($e^{i\omega x}$), onde o peso numa frequência particular ω é dado por $\hat{f}(\omega)$. A expressão matemática é dada por (Papoulis, 1962):

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(\omega) e^{i\omega x} d\omega, \quad (1)$$

O conteúdo $\hat{f}(\omega)$ é obtido através da expressão:

$$\hat{f}(\omega) = \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx. \quad (2)$$

A Figura 2.1 ilustra a maneira como a análise de Fourier “enxerga” uma função.

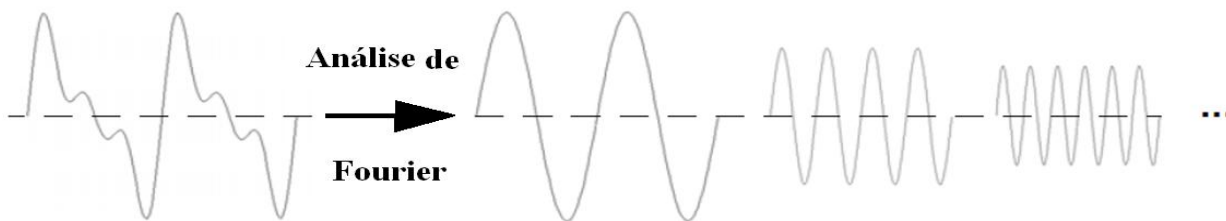


Figura 2.1: Função original e a decomposição via FT (Misiti et al., 1997)

⁵ Ao decompor uma função em partes mais simples, este processo é chamado análise. A operação de reconstruir a função a partir dessas partes é conhecida como síntese. Assim, nesse contexto, o termo síntese de Fourier descreve o processo de reconstruir uma função a partir de componentes mais simples obtidas pela análise de Fourier. No entanto, o termo análise de Fourier usualmente se refere a ambos os processos.

A utilidade da transformada de Fourier reside na capacidade de analisar uma função através de seu conteúdo de frequência, o que geralmente revela características que são de difícil visualização no domínio do tempo (Chakraborty et al., 1995). Graficamente, a transformada de Fourier de uma função é analisada, onde na Figura 3.2a tem-se o sinal original composto por duas senóides e na Figura 3.2b o conteúdo de frequência de 10 e 20 (Hz) obtidos via FT:

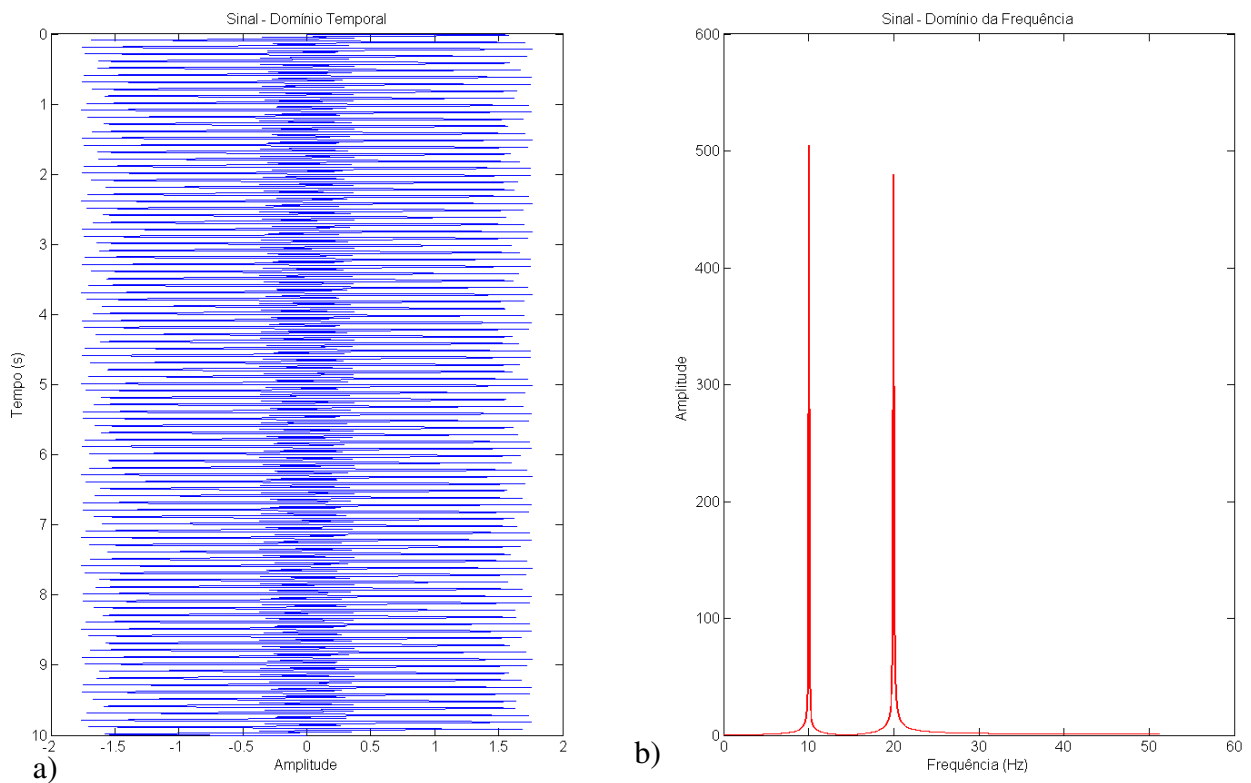


Figura 2.2: a) Sinal composto por duas ondas senoidais; b) Conteúdo de frequência obtido via FT

Para muitas funções, a análise de Fourier é bastante útil, uma vez que o conteúdo de frequência é de grande importância. No entanto, a transformada de Fourier apresenta uma limitação: ao identificar as frequências constituintes do sinal na transformação do domínio do tempo para o domínio da frequência, perde-se totalmente a informação sobre a localização temporal do evento, ou seja, é impossível dizer quando determinado conteúdo ocorre. Como exemplo, considere os sinais compostos pelas duas ondas senoidais de frequências 10 e 20 (Hz),

como ilustrado na Figura 2.3. Para o sinal 1, tem-se a senóide de 10 (Hz) no intervalo de 0 a 5 (s) e a senóide de 20 (Hz) no intervalo de 5 a 10 (s), enquanto o sinal 2 possui ambas as senóides em todo o intervalo.

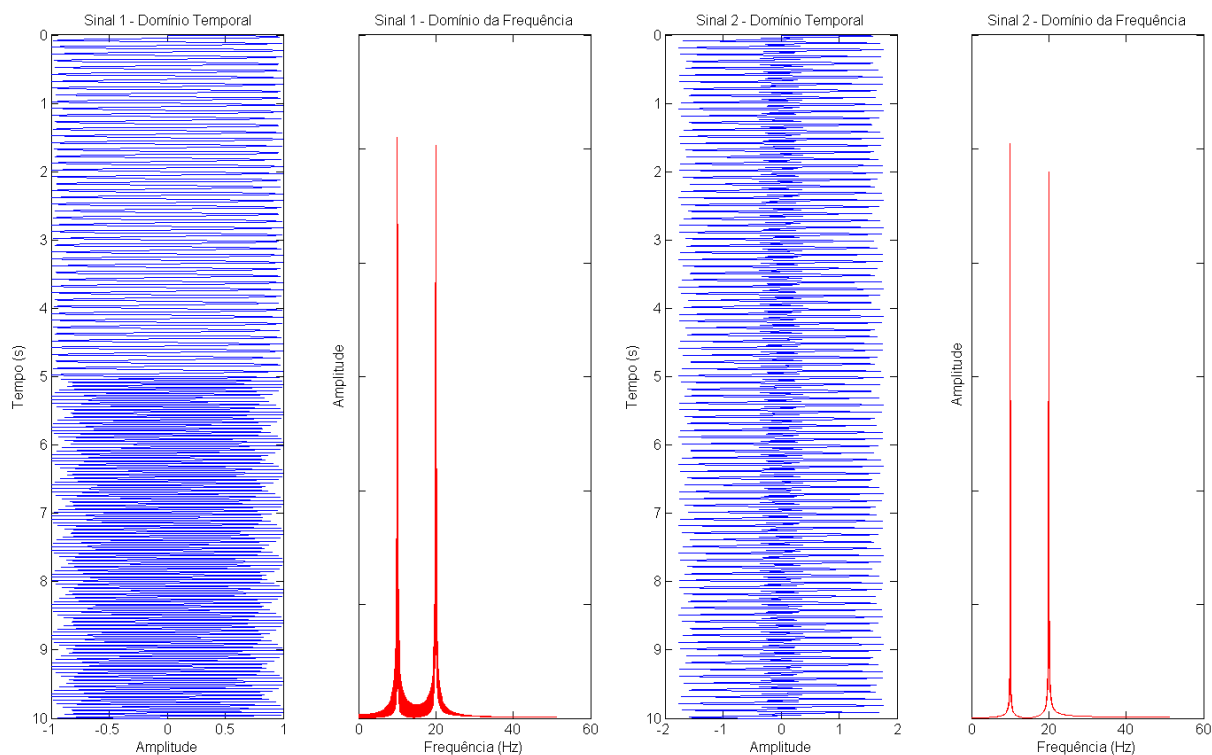


Figura 2.3: Sinais compostos por duas senóides de 10 e 20 (Hz) e seus respectivos espectros obtidos via FT.

Do ponto de vista da Transformada de Fourier, não há distinção entre as duas respostas obtidas para o espectro de frequência, muito embora os sinais de entrada sejam diferentes. Esta é uma das limitações da FT e, portanto, seria interessante uma abordagem que contemplasse tal situação de modo a caracterizar de forma mais fidedigna o sinal analisado.

Por volta de 1945, Léon Brillouin, Dennis Gabor, Claude Shannon e Jean Ville já estavam conscientes das limitações da transformada de Fourier (Meyer, 2006), e buscavam uma técnica mais eficiente para a representação tempo–frequência.

Gabor, em 1946, aprimora a abordagem da transformada de Fourier de modo a estabelecer uma janela temporal e assim, analisar o sinal dentro desse intervalo. Essa técnica é conhecida como Transformada de Fourier de Curta Duração (*Short Time Fourier Transform - STFT*) ou, analogamente, como Transformada de Fourier Janelada (*Windowed Fourier Transform - WFT*). Este conceito permite avaliar um sinal tanto em termos do tempo quanto da frequência, estudando o sinal parte por parte, ou janela por janela.

Matematicamente, a STFT é dada por:

$$Gf(\omega, x) = \int_{-\infty}^{\infty} f(u)g(x - u)e^{-i\omega u} du, \quad (3)$$

onde $g(x)$ representa a função janela.

Com esta abordagem, tornou-se possível uma correlação entre o conteúdo de frequência da função com o instante que tal evento ocorre, ou seja, fornece informação sobre qual e onde ocorre determinada frequência. Analogamente à análise feita via FT para o caso de dois sinais compostos por senóides de 10 e 20 (Hz), através da STFT é obtida uma resposta mais fiel àquilo que de fato o sinal representa (Figura 2.4):

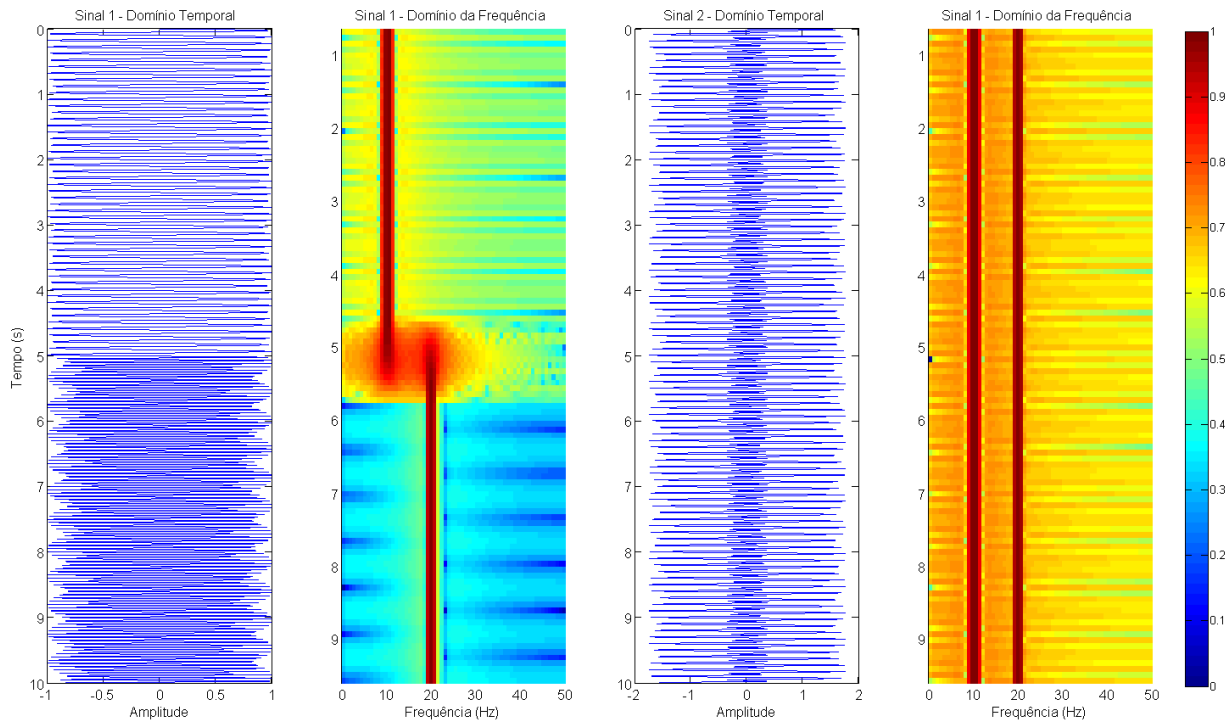


Figura 2.4: Sinais compostos por duas senóides de 10 e 20 (Hz) e os respectivos espectros obtidos via STFT.

Evidentemente, há uma distinção entre os espectros de frequências obtidos para cada um dos sinais. No entanto, apesar da STFT apresentar uma evolução frente à FT, também apresenta sua limitação: uma vez estabelecido o tamanho da janela, este é fixado para todo o sinal de modo a negligenciar as características que possuem tamanhos maiores ou menores que àquele estabelecido pela janela. Isso pode ser notado no intervalo de 4.5 – 5.5 (s) onde não há uma clara definição do conteúdo de frequência presente.

Muitos sinais, no entanto, requerem uma abordagem mais flexível, ou seja, uma técnica na qual se possa variar o tamanho da janela e com isso determinar com maior precisão as características do sinal, janela pequena para analisar altas frequências e janelas grandes para analisar baixas frequências, sendo este o princípio da análise *wavelet*.

2.3. Análise *Wavelet*

Assim como senos e cossenos na análise de Fourier, as *wavelets* são utilizadas como funções base na representação de outras funções. Uma função $\Psi(x) \in L^2(R)$ ⁶ é chamada *wavelet* se possui média zero (eq. 4) e se satisfaz algumas condições específicas, e.g., condição de admissibilidade (eq. 5), (Grossmann & Morlet, 1984):

$$\int_{-\infty}^{\infty} \Psi(x) dx = 0; \quad (4)$$

$$0 < \int_{-\infty}^{\infty} \frac{|\hat{\Psi}(\omega)|^2}{|\omega|} d\omega < \infty, \quad (5)$$

onde $\hat{\Psi}(\omega)$ é a transformada de Fourier de $\Psi(x)$.

A Figura 2.5 ilustra a representação de um sinal segundo a óptica da análise *wavelet*:

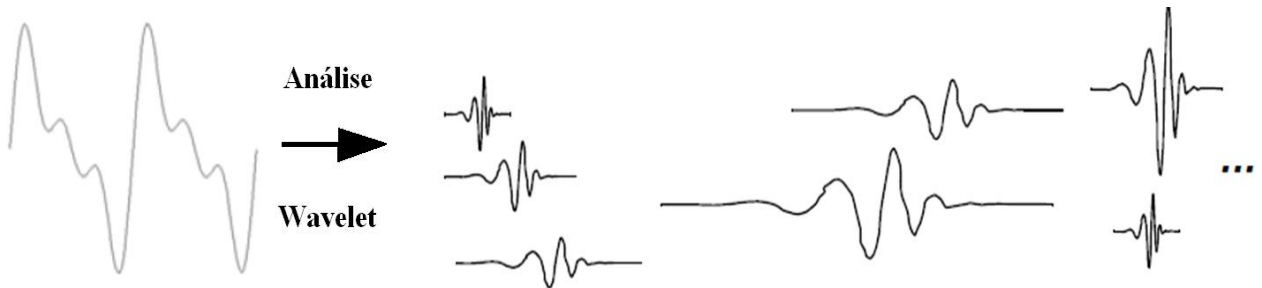


Figura 2.5: Função original e a decomposição via WT (Misiti et al., 1997)

Uma família de *wavelets* pode ser obtida através dos parâmetros de escala \mathbf{a} e deslocamento \mathbf{b} da *wavelet* mãe $\Psi(x)$:

$$\Psi_{a,b}(x) \equiv \frac{1}{\sqrt{a}} \Psi\left(\frac{x-b}{a}\right); \quad a > 0 \text{ e } b \in R. \quad (6)$$

⁶ Espaço de todas as funções quadraticamente integráveis: uma função \mathbf{f} pertence a $L^2(S)$ se $\int_S f^2$ é finita

Alterar o valor de a , dilata ($a > 1$) ou contrai ($a < 1$) a função $\Psi(x)$: se uma *wavelet* é comprimida, seu espectro é espalhado e as componentes de alta frequência são avaliadas; ao contrário, se uma *wavelet* é dilatada, um espectro mais estreito é obtido e as componentes de menor frequência são consideradas (Figura 2.6)

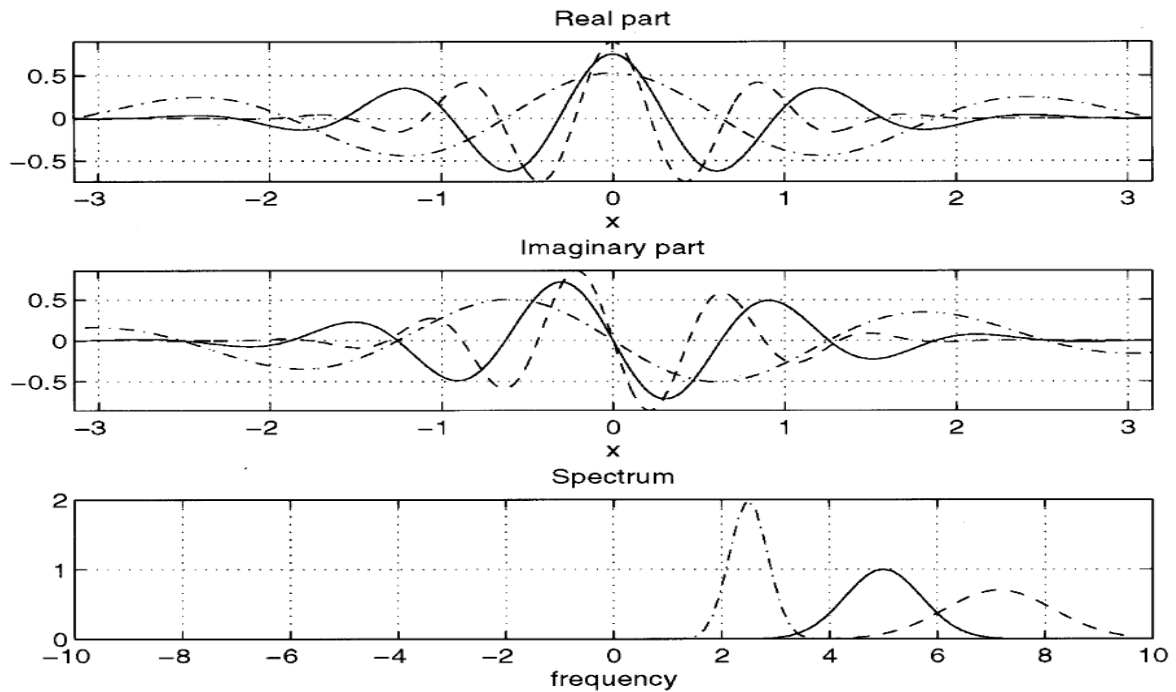


Figura 2.6: Parte real e imaginária da *wavelet* Morlet e espectro de frequência para diferentes escalas (Kumar & Foufoula, 1997)

2.3.1. Transformada *Wavelet* Contínua

Uma vez escolhida a família de *wavelets*, a transformada *wavelet* contínua (*Continuous Wavelet Transform* - CWT) de uma função $f(x) \in L^2(\mathbb{R})$ é definida como:

$$W f(a, b) = \int_{-\infty}^{\infty} f(x) \overline{\Psi_{a,b}}(x) dx, \quad (7)$$

onde $\overline{\Psi_{a,b}}(x)$ denota o complexo conjugado de $\Psi_{a,b}$

A transformada *wavelet* é uma transformação linear, ou seja, a transformada *wavelet* da soma de dois sinais é a soma das transformadas *wavelets* de cada sinal individual. O termo “contínuo” da CWT se refere aos valores assumidos pelos parâmetros de escala e posição (equação 6). A CWT fornece uma representação redundante do sinal, ou seja, a CWT de uma função numa dada escala e localização pode ser obtida a partir da CWT da mesma função em outras escalas e posições.

A transformada *wavelet* inversa é dada por (Daubechies, 1992):

$$f(x) = \frac{1}{C_\Psi} \int_{-\infty}^{\infty} \int_0^{\infty} a^{-2} W f(a,b) \Psi_{a,b}(x) da dx, \quad (8)$$

onde C_Ψ é constante e depende da escolha da função *wavelet*. A equação 8 pode ser vista como uma forma de reconstruir $f(x)$ através da superposição das *wavelets* $\Psi_{a,b}(x)$, uma vez conhecida a transformada $Wf(a,b)$.

Como é possível notar da Figura 2.6, assim como na análise de Fourier, o tamanho da janela estabelece a resolução do conteúdo de frequência: uma função com um suporte mais compacto vai ter uma melhor resolução no domínio do tempo e uma resolução mais pobre no domínio da frequência, enquanto uma função com suporte mais amplo terá uma resolução mais pobre no domínio do tempo e uma melhor resolução no domínio da frequência. No entanto, a análise *wavelet* contempla uma ampla variedade no tamanho da janela e, conseqüentemente, diferentes resoluções são obtidas.

A característica de dependência entre a resolução tempo-frequência é governada pelo princípio de incerteza de Heisenberg, o qual estabelece que certos pares de propriedades físicas não podem ser medidos simultaneamente com precisão arbitrariamente alta. Para entender essa relação entre suporte da função e sua resolução, é útil identificar que a equação 7 pode ser reescrita utilizando o teorema de Parseval, de modo a obter:

$$W f(a, b) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \hat{f}(x) \overline{\hat{\Psi}_{a,b}}(\omega) d\omega. \quad (9)$$

Com isso, as propriedades de localização da transformada *wavelet* podem ser compreendidas analisando o desvio padrão de $|\Psi_{a,b}(x)|^2$ e $|\hat{\Psi}_{a,b}(\omega)|^2$, representados por σ_a e $\hat{\sigma}_a$, respectivamente. Tais parâmetros, que são funções da escala, satisfazem as seguintes relações (Kumar & Foufoula, 1997):

- O desvio padrão σ_a de $|\Psi_{a,b}|^2$ satisfaz a igualdade:

$$\sigma_a = a \sigma_1; \quad (10)$$

- O desvio padrão $\hat{\sigma}_a$ de $|\hat{\Psi}_{a,b}|^2$ satisfaz a igualdade:

$$\hat{\sigma}_a = \hat{\sigma}_1/a; \quad (11)$$

- O centro do passa-banda ω_a^0 , o qual é definido como a moda de $|\hat{\Psi}_{a,b}|^2$, satisfaz:

$$\omega_a^0 = \omega_1^0/a. \quad (12)$$

Através dessas expressões, é possível notar que, quanto maior o parâmetro escala, ou seja, mais dilatada a função, tanto ω_a^0 quanto $\hat{\sigma}_a$ decrescem, indicando que o centro do passa-banda é deslocado para as componentes de baixas frequências e que há uma diminuição da incerteza ($\hat{\sigma}_a$), proporcionando uma melhor resolução no domínio da frequência, e vice-versa (Figura 2.6).

No plano tempo-frequência, a célula de resolução para a transformada *wavelet* no ponto (x_0, ω_a^0) é dado por:

$$[x_0 \pm a \sigma_1 \times (\omega_1^0/a) \pm (\hat{\sigma}_1/a)]. \quad (13)$$

A área dessa célula é expressa por:

$$[\sigma_a \times \hat{\sigma}_a]. \quad (14)$$

A expressão 13 possui dimensões variáveis e depende do parâmetro de escala. No entanto, a área da célula de resolução independe dos parâmetros de escala e posição, ou seja, possui tamanho constante. Logo, ao aumentar a resolução sobre domínio do tempo, diminui a resolução sobre o domínio da frequência e vice-versa (Figura 2.7).

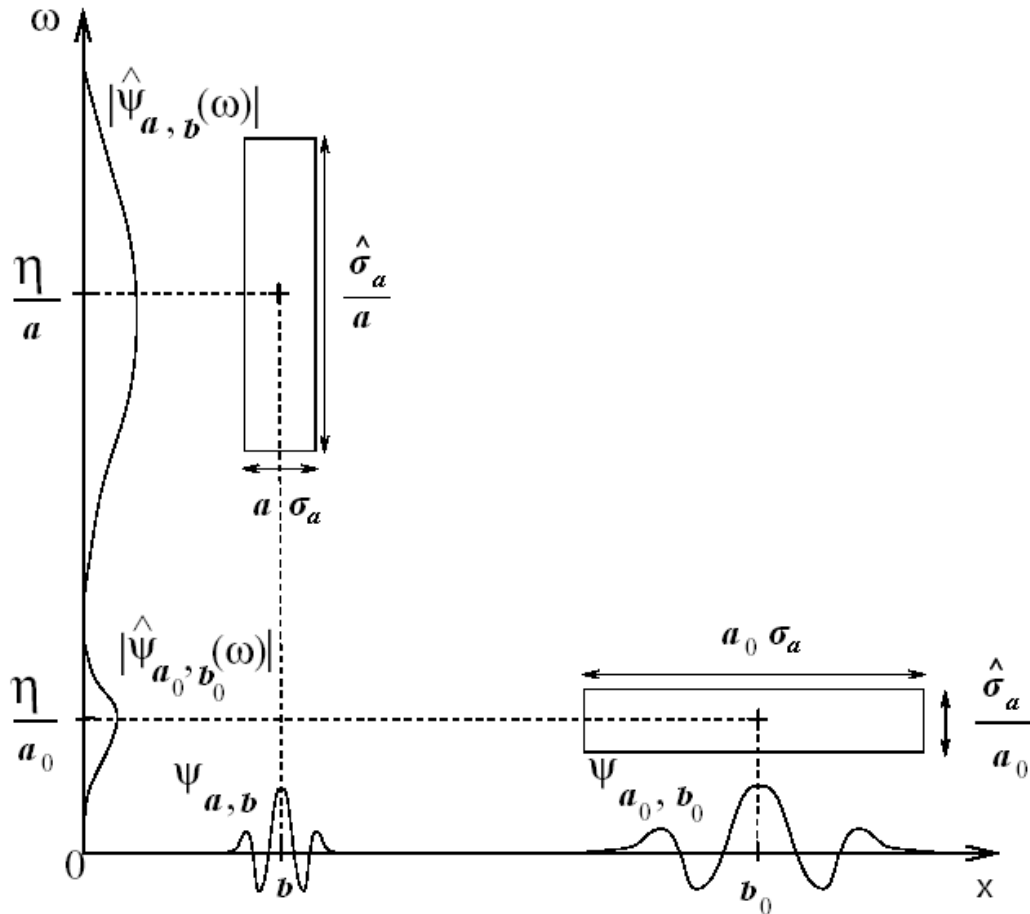


Figura 2.7: Plano tempo-frequência da transformada *wavelet* (modificado de Protázio, 2002).

O sinal analisado através da CWT é representado como uma série de coeficientes que medem a correlação entre o sinal e as *wavelets* utilizadas. Assim, quanto maior a similaridade entre a *wavelet* e a seção do sinal, maior será o coeficiente C . Note que o resultado dependerá da forma da *wavelet* escolhida.

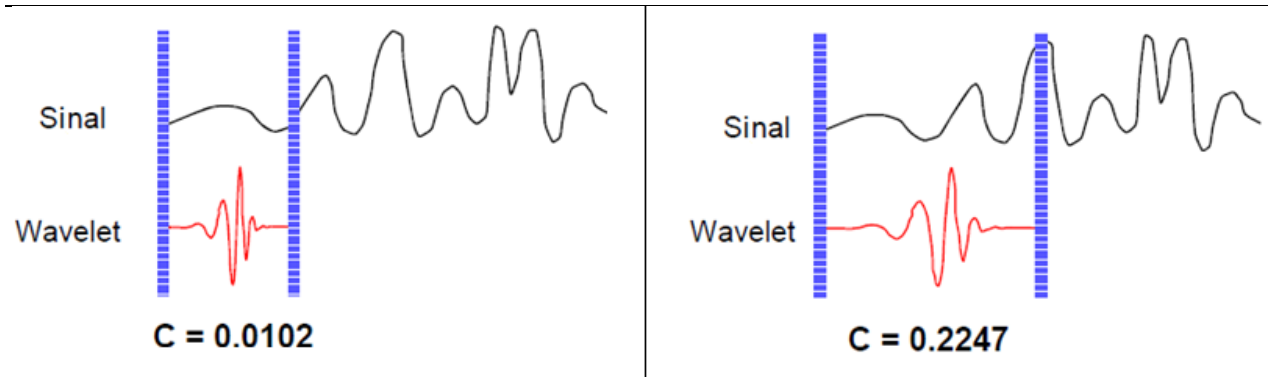


Figura 2.8: Obtenção dos coeficientes *wavelet* (Misiti et al., 1997).

Em suma, o processo de obtenção dos coeficientes *wavelet* pode ser sintetizado da seguinte forma:

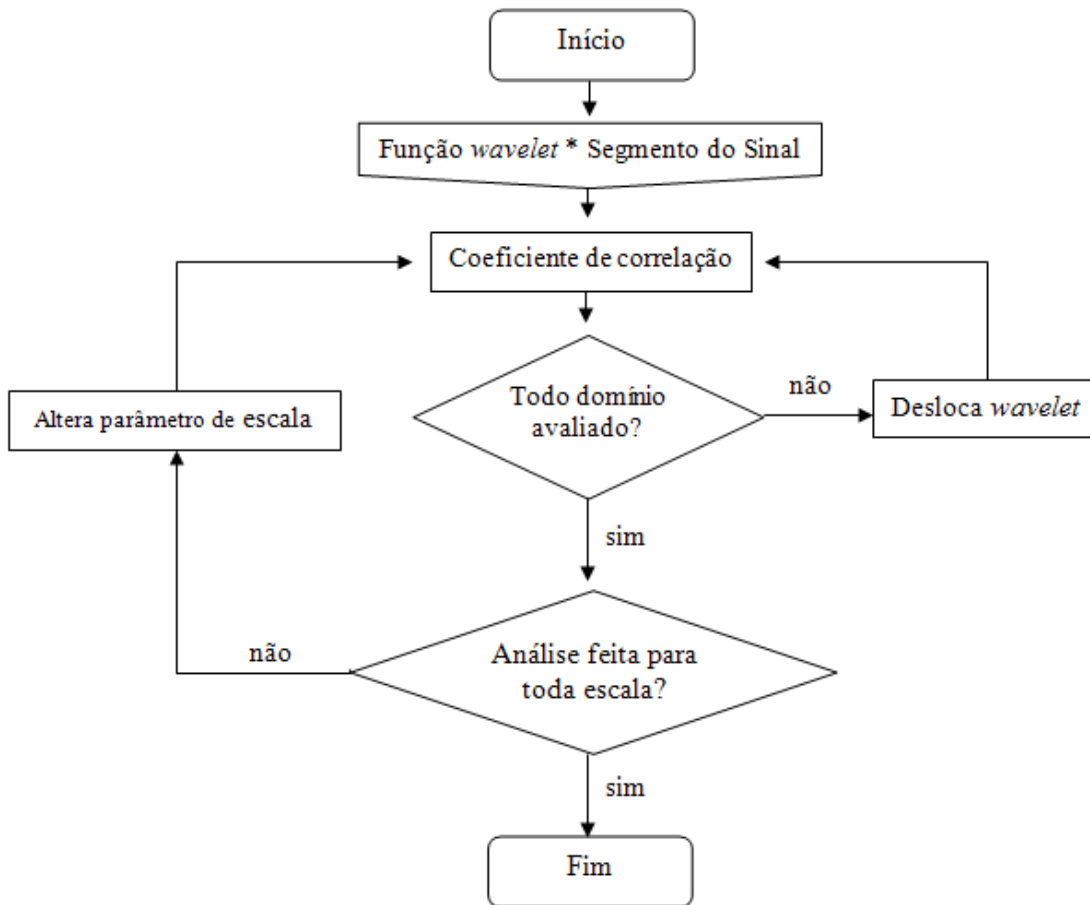


Figura 2.9: Fluxograma da CWT

- Uma seção do sinal é comparada com a função *wavelet* pré-escolhida.
- O coeficiente C , que representa a similaridade entre a *wavelet* e a seção do sinal, é calculado;
- A *wavelet* é deslocada e os passos anteriores são repetidos até que todo o domínio do sinal tenha sido analisado;
- A escala a da *wavelet* é alterada e os passos anteriores são repetidos;
- Os passos anteriores são repetidos para todas as escalas a .

A representação de amplitude desses coeficientes é chamada de escalograma (Figura 2.10). A mesma análise feita para os sinais compostos pelas senóides de 10 e 20 (Hz) é realizada, agora usando a CWT:

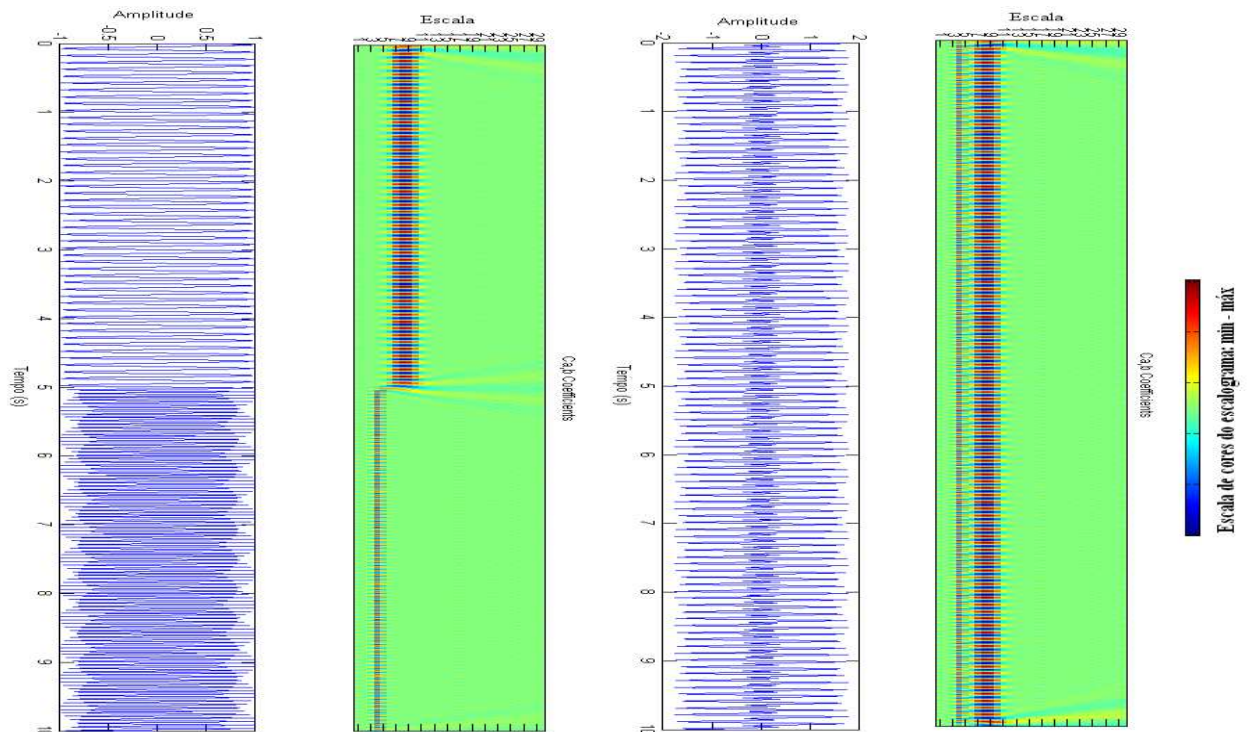


Figura 2.10: Sinais compostos por duas senóides de 10 e 20 (Hz) e os respectivos espectros obtidos via CWT (*wavelet* Morlet).

Como é possível observar da Figura 2.10, a transformada *wavelet* não faz a representação tempo-frequência propriamente dita, e sim, tempo-escala. Para fazer uma correspondência entre escala e frequência, é mais plausível falar numa pseudo-frequência associada a uma determinada escala.

A idéia é associar a uma dada *wavelet* um sinal periódico de frequência F_c . Assim, a frequência F_c é uma caracterização conveniente e simples da frequência dominante da *wavelet*. Portanto, quando uma *wavelet* é dilatada por uma fator a , sua frequência se altera para F_c/a . Por fim, se o período de amostragem é Δ , a frequência associada a escala a fica (Abry, 1997):

$$F_a = \frac{F_c}{a \cdot \Delta}. \quad (15)$$

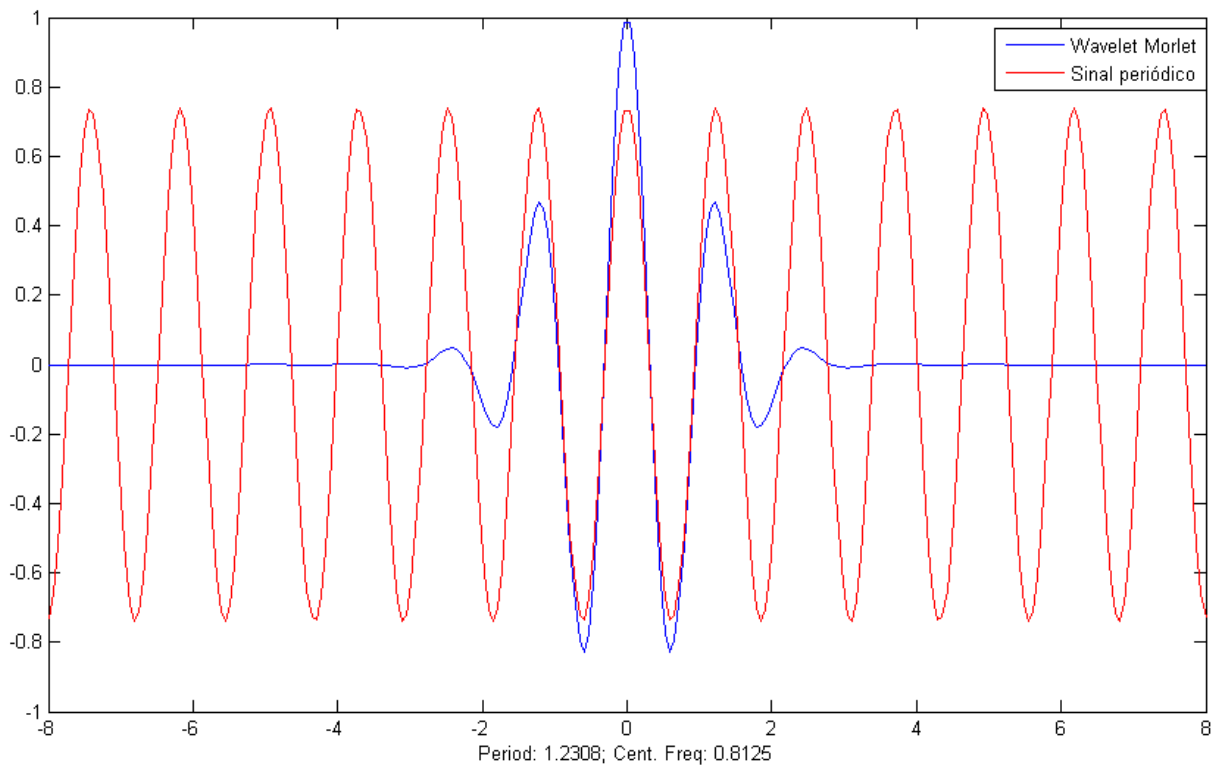


Figura 2.11: Determinação da frequência central da *wavelet* Morlet

No exemplo dos sinais da Figura 2.10, a CWT do sinal analisado obteve as escalas $a=8$ e $a=4$. O sinal possui um período de amostragem ≈ 0.01 . Portanto, $F_8 \approx 10$ (Hz) e $F_4 \approx 20$ (Hz).

2.3.2. Transforma *Wavelet* Discreta

Um inconveniente da CWT é que a representação do sinal é muitas vezes redundante devido aos valores atribuídos aos parâmetros \mathbf{a} e \mathbf{b} (ver equação 6). Uma forma de contornar isso é utilizar valores discretos para tais parâmetros, o que caracteriza a transformada *wavelet* discreta (*Discrete Wavelet Transform* - DWT). Dessa forma, pode-se obter tanto uma representação redundante quanto não-redundante utilizando escolhas apropriadas de funções *wavelets* e esquemas de discretização (Kumar & Foufoula, 1997).

Uma forma para tal é escolher um esquema de discretização dado por:

$$a = a_0^j; \quad b = kx_0 a_0^j; \quad a_0 > 1, \quad j \text{ e } k \in \mathbb{Z} \quad (16)$$

Como mencionado no início desse capítulo, a transformada *wavelet* é entendida como um microscópio matemático. Tal analogia provém do fator de magnificação (a_0^{-j}) no qual um determinado problema é avaliado: pequenas escalas proporcionam uma grande magnificação e vice-versa.

Através dos valores assumidos pelos parâmetros de escala e posição definidos acima, a família de funções *wavelets* é então expressa por:

$$\Psi_{j,k}(x) \equiv \frac{1}{\sqrt{a_0^j}} \Psi\left(\frac{x - kx_0 a_0^j}{a_0^j}\right) = a_0^{-j/2} \Psi(a_0^{-j} x - kx_0), \quad (17)$$

Utilizando a família de *wavelets* obtidas pela equação 17, tem-se para a transformada *wavelet* discreta:

$$W f(j, k) = a_0^{-j/2} \int_{-\infty}^{\infty} f(x) \Psi(a_0^{-j} x - k x_0) dx \quad (18)$$

Através de escolhas apropriadas para os parâmetros a_0 e x_0 e utilizando a família de *wavelets* discretas, o sinal $f(x)$ pode ser completamente caracterizado assim como no caso da CWT. Essas funções *wavelets* que proporcionam a completa representação do $f(x)$ são chamadas de “frames”. A condição necessária e suficiente para isso é que os coeficientes *wavelets* satisfaçam:

$$A \|f\|^2 \leq \sum_j \sum_k |W f(j, k)|^2 \leq B \|f\|^2, \quad (19)$$

onde $\|f\|^2$ denota a energia (ou norma L^2) da função $f(x)$, $A > 0$ e $B < \infty$ são constantes características da *wavelet* e da escolha dos parâmetros a_0 e x_0 (Daubechies, 1992). Satisfeitas essas condições, pode-se obter a expansão em série para $f(x)$ dada por:

$$f(x) = \frac{2}{A+B} \sum_j \sum_k W f(j, k) \Psi_{j,k} + \gamma, \quad (20)$$

onde γ é o erro.

Em geral, um frame não é uma base ortogonal (essa situação ocorre somente quando $A=B=1$). Os frames fornecem uma representação redundante do sinal $f(x)$, sendo a razão A/B denominado fator de redundância. Em outras palavras, os coeficientes *wavelets* obtidos numa vizinhança são correlacionados uns com os outros, resultando na melhor resolução das características do sinal analisado. No entanto, isso gera um aumento da complexidade computacional, resultando em algoritmos mais lentos (Kumar & Foufoula, 1997).

Para muitos sinais que possuem características não-estacionárias, ou seja, cujo conteúdo de frequência varia ao longo do tempo, a base mais conveniente para a decomposição desses sinais é uma base ortogonal, local e universal: local no sentido que não se espalham em todo o

domínio (como senos e cossenos o são na FT), e universais, pois podem ser utilizadas para uma eficiente decomposição de qualquer conjunto de dados.

2.3.3. Transformada *Wavelet* Ortogonal

Um conjunto de bases ortogonais pode ser obtido atribuindo ao esquema de discretização (equação 16) $a_0 = 2$ e $x_0=1$, de tal forma que a família de *wavelets* obtidas (equação 17) fica:

$$\Psi_{j,k}(x) \equiv \frac{1}{\sqrt{2^j}} \Psi\left(\frac{x - k2^j}{2^j}\right) = 2^{-j/2} \Psi(2^{-j}x - k), \quad (21)$$

A propriedade mais notável dessa base é que as funções são ortogonais à suas translações e dilatações. Todas as funções pertencentes a L^2 podem ser aproximadas, com precisão arbitrariamente alta, a partir de combinações lineares das *wavelets* $\Psi_{j,k}$:

$$f(x) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} D_{j,k} \Psi_{j,k}(x), \quad (22)$$

onde $D_{j,k}$ é dado por:

$$D_{j,k} = \int f(x) \Psi_{j,k}(x) dx. \quad (23)$$

O coeficiente $D_{j,k}$ mede a contribuição da escala 2^j na posição $k2^j$. A expansão obtida através da equação (22) é semelhante à obtida para as séries de Fourier, com as seguintes diferenças:

- A série é duplamente indexada, cujos índices indicam a escala e a localização;
- A função base tem a propriedade de localização tempo-frequência.

A Figura 2.7 elucida a representação tempo-frequência obtida por diferentes abordagens:

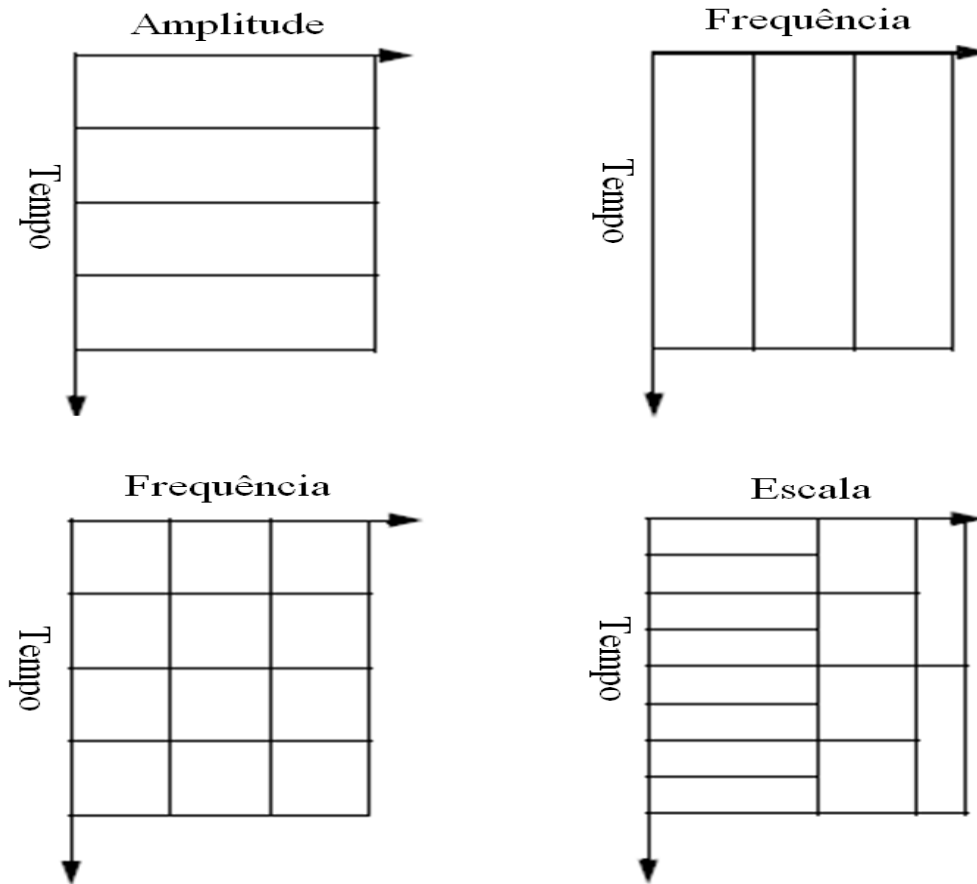


Figura 2.12: Sinal no domínio do tempo e análise do conteúdo de frequências obtidos através da FT, STFT e WT, respectivamente

2.3.4. Análise Multirresolução

A implicação mais importante dessa escolha dos valores dos parâmetros é a conexão com a análise multirresolução (*Multiresolution Analysis* - MRA). A MRA está relacionada com o estudo de sinais ou processos representados por diferentes níveis de resolução, o que permite que um sinal complexo seja decomposto em diversos outros mais simples, e assim, possíveis de serem analisados separadamente.

Em suma, o processo de decomposição através da MRA consiste em separar o sinal original em duas partes: uma contendo as características mais gerais do sinal (grande escala – baixa frequência) e outra contendo as características de detalhe (pequena escala – alta frequência). Essas componentes são denominadas “aproximação” e “detalhe”. Uma vez que o sinal tenha sido decomposto, a componente de aproximação pode ser novamente submetida ao processo de separação em novas componentes de aproximação e detalhe, e assim sucessivamente. Em teoria, esse processo poderia ser mantido indefinidamente até que os detalhes individuais representassem uma única amostra ou pixel (no caso de imagens). Na prática, seleciona-se um número adequado de níveis com base na natureza do sinal. A Figura 2.13 ilustra a decomposição via MRA:

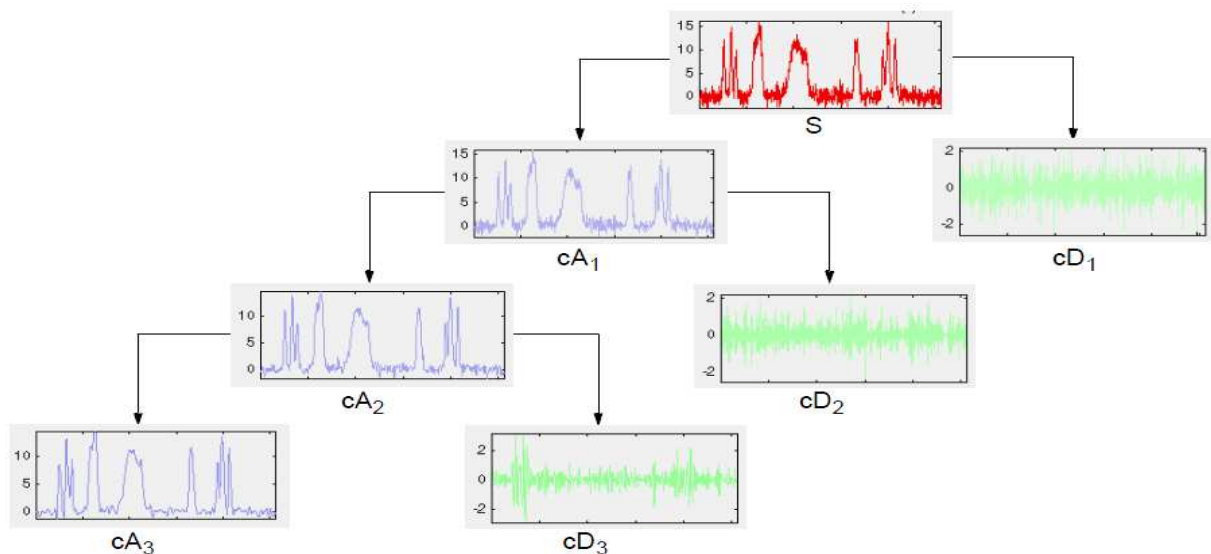


Figura 2.13: Ilustração da análise multirresolução. S: Sinal original; cA_j e cD_j : componentes de Aproximação e Detalhe no nível j (Misiti *et al.*, 1997).

Matematicamente, a MRA é dada da seguinte forma: Seja $f_j(x)$ a função que representa as características maiores que uma determinada escala (todas as características maiores que essa

determinada escala estão presentes). Para obter a próxima função de maior resolução, um detalhe $f'_j(x)$ é adicionado, de modo que:

$$f_{j-1} = f_j(x) + f'_j(x), \quad (24)$$

a qual é a equação recursiva básica da teoria multirresolução. Na Figura 2.13, $cA_{j-1} = cA_j(x) + cD_j(x)$, ou seja, $cA_{j-1} \equiv f_{j-1}$; $cA_j(x) \equiv f_j(x)$ e $cD_j(x) \equiv f'_j(x)$.

No âmbito da MRA através das *wavelets*, a função $f_j(x)$ é aproximada por:

$$f_j(x) = \sum_{k=-\infty}^{\infty} C_{j,k} \Phi_{j,k}(x), \quad (25)$$

sendo a função suave (smooth function) $\Phi_{j,k}$ e os coeficientes $C_{j,k}$ dados por:

$$\Phi_{j,k}(x) = 2^{-j/2} \Phi(2^{-j}x - k); \quad C_{j,k} = \int \Phi_{j,k}(x) dx. \quad (26)$$

A função Φ é chamada de função de escala (*scaling function*) e atua como uma função de amostragem, de modo que os coeficientes $C_{j,k}$ são amostras discretas de $f(x)$. Os detalhes $f'_j(x)$ são aproximados utilizando a família de *wavelets* ortogonais, e é expressa como:

$$f'_j(x) = \sum_{k=-\infty}^{\infty} D_{j,k} \Psi_{j,k}(x), \quad (27)$$

onde os coeficientes $D_{j,k}$ são definidos na equação 23.

CAPÍTULO 3

ATENUAÇÃO DE RUÍDOS E DETERMINAÇÃO DE CICLOS SEDIMENTARES EM PERFIS GEOFÍSICOS

3.1. Introdução

Nos dados aplicados à cicloestratigrafia, os registros ou sinais do ambiente que são “codificados” durante a sedimentação muitas vezes são corrompidos por processos que não fazem parte do sistema normal de deposição. Tais processos incluem as diferentes formas de discordâncias no registro sedimentar e torna o reconhecimento dos fenômenos normais de deposição uma questão mais laboriosa.

Assim como as interrupções, as variações na taxa de acumulação e a diagênese frequentemente modificam as formas finais do conjunto de dados cíclicos. Somado a isso, os próprios processos de gravação podem introduzir distorções que precisam ser levadas em consideração. Portanto, os dados utilizados na cicloestratigrafia contêm informações sobre a variabilidade normal do ambiente, variações ambientais anormais e os processos que produzem os próprios registros que, inerentemente, possuem ruídos. Em outras palavras, as informações estratigráficas que são observadas podem ser consideradas como o produto de muitos processos sobrepostos (Weedon, 2003).

A compreensão de que os dados estratigráficos são compostos por componentes regulares, mais as componentes irregulares, é baseado em uma visão linear dos processos envolvidos. Na realidade, os processos não-lineares são abundantes em sistemas ambientais (e.g., Le Treut & Ghil, 1983). Em sistemas não-lineares, a saída não varia na proporção direta à entrada. Há muitos aspectos em dados cicloestratigráficos que não podem ser facilmente investigados pelos métodos lineares de análise. Apesar da opinião de que as abordagens não-lineares podem explicar mais sobre os dados do que os métodos lineares, estes últimos são atualmente melhor compreendidos matematicamente e são os mais frequentemente usados (Weedon, 2003).

A Figura 3.1 ilustra um registro estratigráfico composto pela sobreposição de três ondas senoidais com diferentes comprimentos de onda e amplitudes. Como é possível observar, é pouco provável que as três componentes de frequência sejam identificadas e caracterizadas somente através de uma simples inspeção visual do sinal resultante.

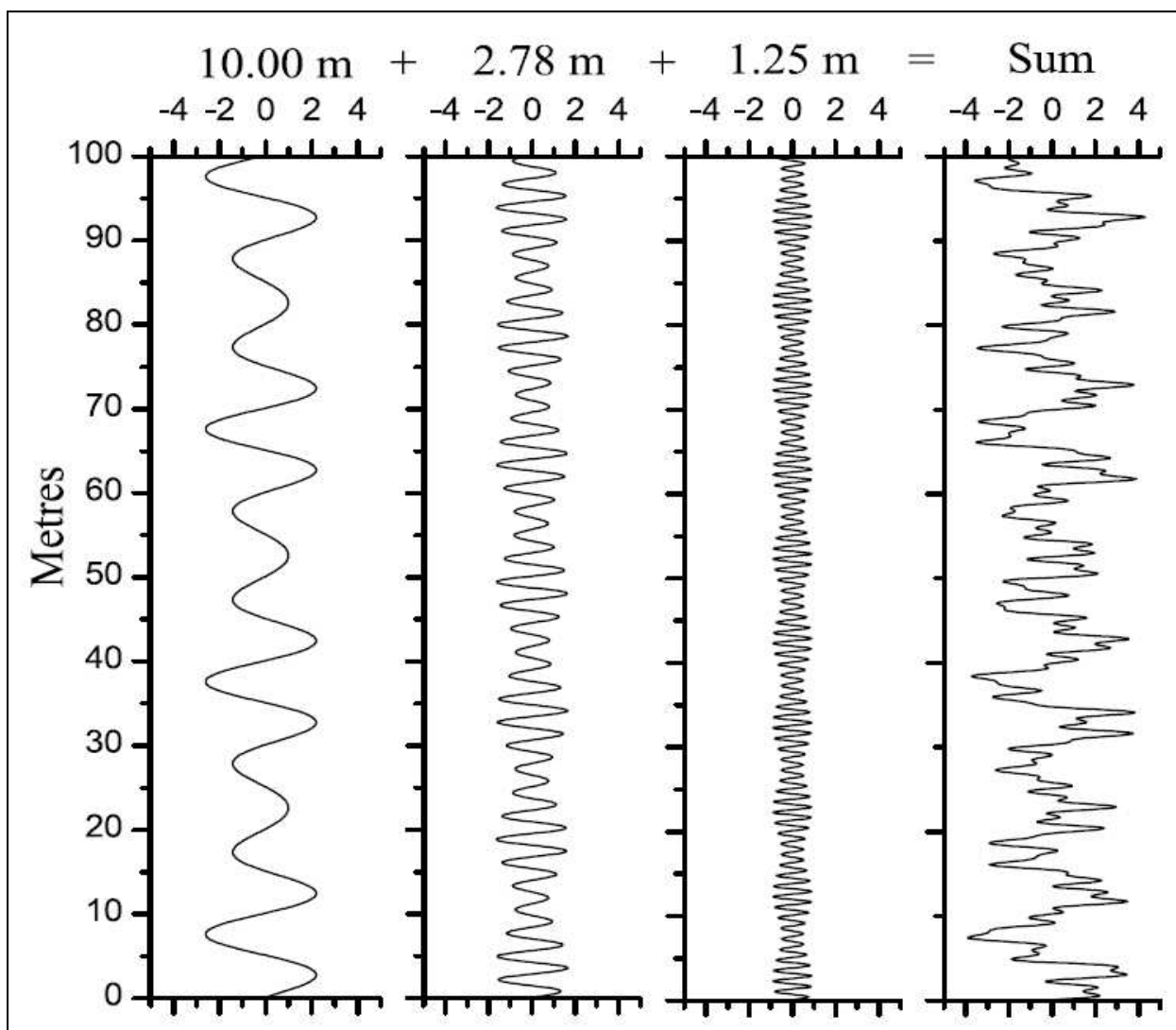


Figura 3.1: Ilustração de um registro estratigráfico composto por três ondas senoidais de diferentes frequências e amplitudes (Weedon, 2003).

Em geral, séries temporais reais consistem da sobreposição de dezenas ou centenas de componentes e, para tanto, a análise espectral se mostra uma ferramenta útil para caracterizar o conteúdo de frequência de tais sinais.

3.2.Método

É prática comum considerar as formas de onda obtidas pelos métodos geofísicos como combinação de sinal e ruído. O sinal é a parte da onda que se relaciona com as estruturas geológicas sob investigação. O ruído, por sua vez, são todas as outras componentes. Portanto, em diversas situações é necessário um pré-processamento do sinal a fim de eliminar ou diminuir a contribuição de componentes indesejadas, e tornar o sinal mais fiel às características do objeto estudado.

Em geral, nos perfis geofísicos, o ruído está associado ao conteúdo de alta frequência. No entanto, mudanças abruptas que refletem características do objeto detectado também são registradas nas altas frequências. Portanto, deve-se haver uma distinção entre essas componentes ao filtrar um sinal, evitando que informações importantes sejam descartadas durante esse processo.

Através da abordagem via MRA, as propriedades locais do sinal podem ser acessadas em ambos os domínios e, uma vez que se pode diferenciar efetivamente o conteúdo do sinal, torna-se uma ferramenta adequada para a supressão e eliminação de ruído.

Ao empregar a WT para atenuação de ruídos, alguns parâmetros devem ser levados em consideração, tais como a escolha da função *wavelet*, a escolha do limiar de corte para as componentes julgadas referentes ao ruído, e como utilizar os coeficientes retidos para a reconstrução do sinal, tópicos esses que serão abordados nas subseções seguintes.

3.2.1. Escolha da Função *Wavelet*

Diante à diversidade de funções *wavelet* base, diversos estudos buscaram um critério objetivo para selecionar a função mais adequada. A base “ótima” não é única e depende do processo investigado (Katul & Vidakovic, 1995)

Ao decompor um sinal através da DWT, a energia do ruído é captada majoritariamente nas componentes de detalhe (alta frequência), as quais possuem menores amplitudes e estão distribuídas em diversos coeficientes. Já a energia das componentes úteis do sinal se restringe a poucos coeficientes, porém com amplitudes maiores (Yu et al., 2010). Esse desequilíbrio entre a energia das componentes pode ser medido através da entropia de Shannon.

Dentro do contexto da WT, Coifman & Wickerhauser (1992) foram os primeiros a utilizar o conceito de entropia em problemas relacionados à construção do algoritmo “*wavelet packets*”, a fim de obter um critério de seleção da melhor base. Vidakovic & Katul (1994) aplicaram esse conceito em problemas relacionados à turbulência atmosférica.

Para uma distribuição discreta de probabilidade, a entropia de Shannon é definida como (Katul & Vidakovic, 1995):

$$H(c) = - \sum_n c_n \log c_n, \quad (28)$$

onde $0 \log 0 \stackrel{\text{def}}{=} 0$. Seja S o sinal normalizado ($\|S\| = 1$) e c_n o conjunto de coeficientes *wavelet* obtidos pela transformada *wavelet* do sinal S . O conjunto c_n é normalizado e pode ser visto como uma distribuição discreta de probabilidade correspondente ao sinal S . Assim, a entropia padronizada para o sinal S dada a WT é:

$$E(S|W) = \frac{H(c^2)}{H\left(\left(\frac{1}{n}, \dots, \frac{1}{n}\right)\right)}. \quad (29)$$

A equação 29 é diretamente proporcional à magnitude do quadrado dos coeficientes *wavelet*. Caso um dado coeficiente seja pequeno (como, por exemplo, os coeficientes

relacionados ao ruído), $c \log c \approx -c$ e sua contribuição para a entropia é insignificante. Assim, quando uma função base produz uma medida de entropia relativamente baixa, a *wavelet* utilizada consegue captar mais eficientemente as componentes do sinal e do ruído e, portanto, esta função base é a mais apropriada e define a melhor escolha, uma vez que a capacidade de separar tais componentes é maximizada. Em outras palavras, quanto maior a semelhança entre a *wavelet* utilizada e o sinal analisado, melhor será o processo de atenuação de ruído.

Para ilustrar essa situação, foram avaliados os valores de entropia de Shannon para um sinal semelhante à *wavelet* Db2 com ruído branco Gaussiano adicionado (Figura 3.2).

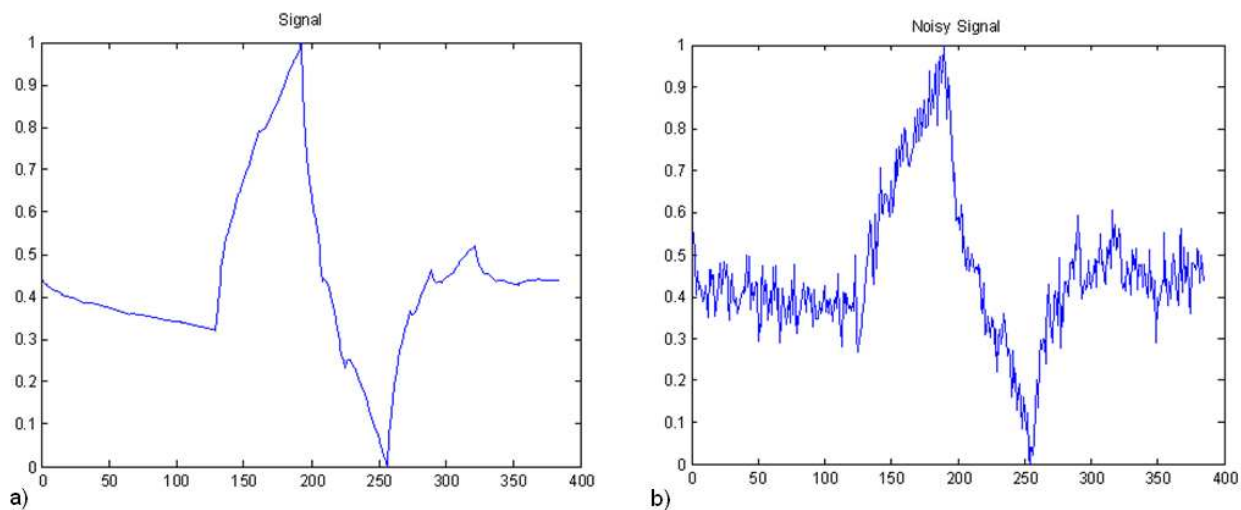


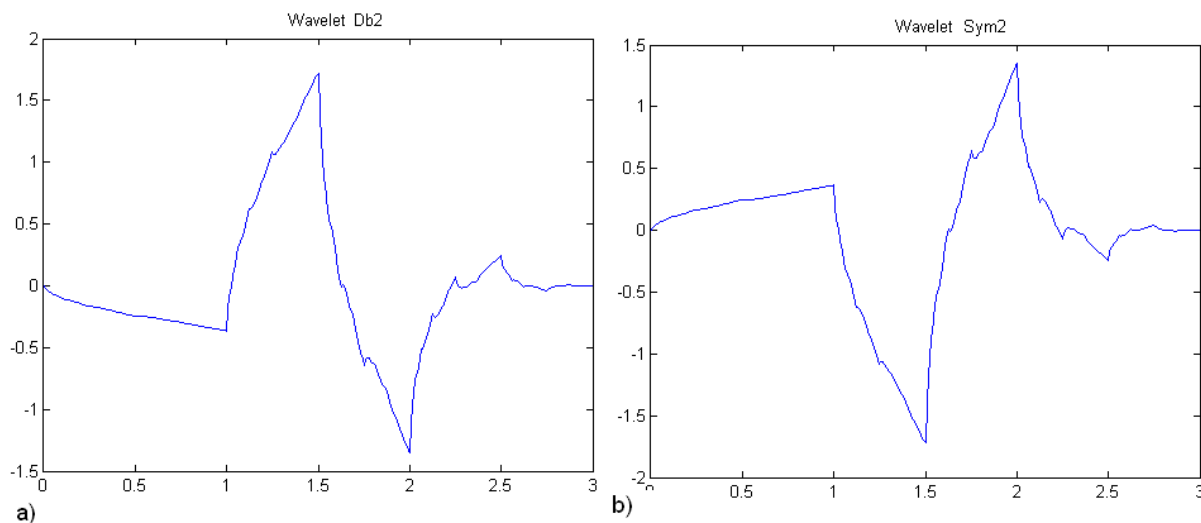
Figura 3.2: a) Sinal semelhante à *wavelet* Db2 sem e com ruído branco Gaussiano adicionado (SNR =15dB)

A entropia de Shannon calculada para diferentes escolhas de *wavelet* base é apresentada na Tabela 3-1.

Tabela 3-1: Entropia de Shannon

<i>Wavelet</i>	Entropia	<i>Wavelet</i>	Entropia
Db1	0,45	Sym3	0,42
Db2	0,34	Sym4	0,42
Db3	0,42	Sym5	0,39
Db4	0,40	Sym6	0,41
Db5	0,37	Sym7	0,38
Db6	0,40	Sym8	0,41
Db7	0,43	Coif1	0,42
Db8	0,44	Coif2	0,42
Db9	0,46	Coif3	0,43
Db10	0,47	Coif4	0,45
Sym2	0,34	Coif5	0,46

Os menores valores da entropia são observados ao fazer a decomposição através das funções Db2 e Sym2, o que definiria a melhor escolha da *wavelet* base para processar o sinal em questão. O fato de ter obtido essas duas funções e não somente a função Db2 é decorrente da similaridade de tais funções, sendo uma o inverso da outra (Figura 3.3).

Figura 3.3: a) *Wavelet* Daubechies2; b) *Wavelet* Symlet2

Como mencionado, outra questão envolvendo a atenuação de ruído através da WT é a escolha do limiar de corte para as componentes julgadas referentes ao ruído, o qual será discutido na seção seguinte. Por hora, basta dizer que os valores abaixo desse limiar são descartados, e o sinal é reconstruído com os coeficientes restantes. Assim, o sinal decomposto com a *wavelet* Db2 e reconstruído com os coeficientes retidos é apresentado na Figura 3.4:

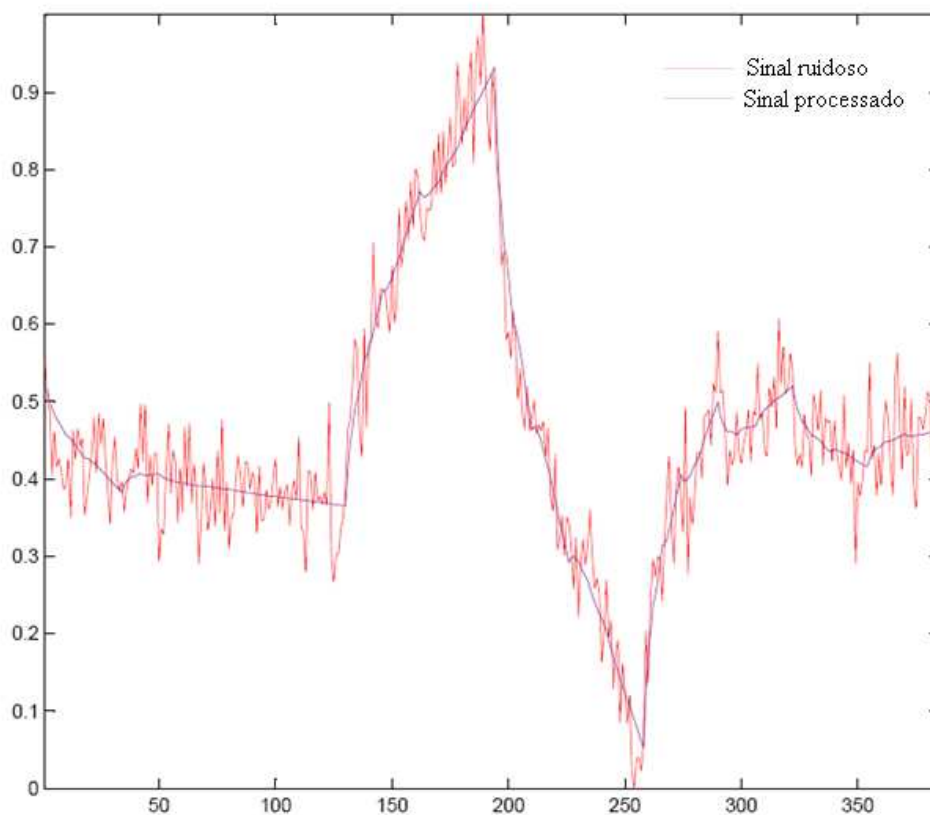


Figura 3.4: Sinal ruidoso (vermelho) e sinal com ruído atenuado (roxo)

3.2.2. Métodos de *Thresholding*

O processo de filtragem envolve a manipulação de uma série temporal de modo a alterar as características espectrais dos dados. Em linhas gerais, os modos de filtragem podem ser divididos em filtros de frequência seletiva (*frequency-selective filtering* - *FSF*), filtros de limiarização (*threshold filtering* - *TF*) e filtros de Wiener (*Wiener filtering* - *WF*). Em cada caso,

um procedimento diferente é utilizado para isolar o sinal do "ruído" nas séries temporais. O método FSFs envolve a remoção de componentes de frequência que não são desejados. A abordagem via TF remove todas as informações relacionadas às variações abaixo de um determinado limite (ou nível de amplitude). A WF pode ser utilizada para modelar o fundo espectral e em seguida, retirá-lo de uma ampla gama de frequências (Figura 3.5).

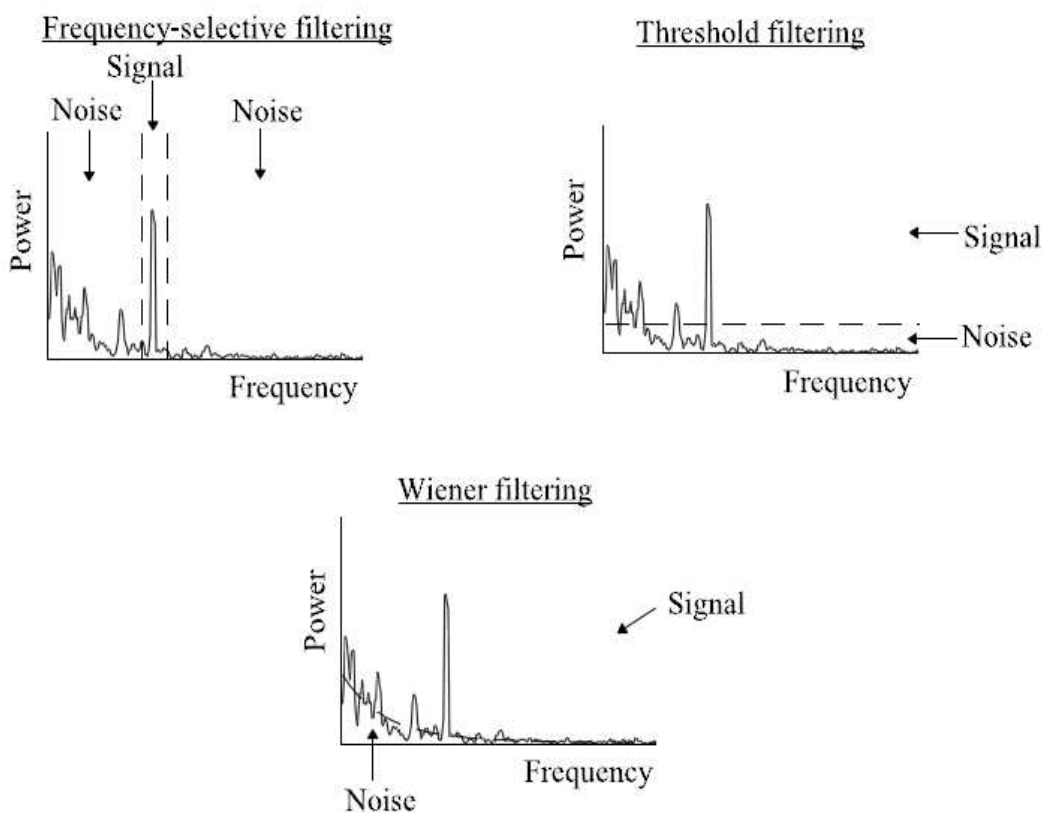


Figura 3.5: Atenuação de ruídos através de diferentes abordagens (Weedon, 2003)

No presente trabalho, a abordagem via *thresholding* foi utilizada.

O conceito de *wavelet thresholding* aplicado ao processo de atenuação de ruído foi introduzido por Donoho & Johnstone (1994) e, basicamente, são classificados como *Hard e Soft Thresholding*. Considere um modelo básico de sinal $s(n)$ que contém ruído expresso por:

$$s(n) = f(n) + \sigma R(n). \quad (30)$$

Numa aproximação simplista, o ruído $R(n)$ assume um modelo de ruído branco Gaussiano $N(0,1)$. A idéia fundamental do processo é reduzir a contribuição desse parâmetro, e com isso, evidenciar o sinal $f(n)$.

Matematicamente, a transformada *wavelet* é equivalente à multiplicação do vetor contendo as amostras discretas de s pela matriz ortonormal W :

$$w = Ws, \quad (31)$$

onde w é a matriz de coeficientes *wavelets*. Então, os coeficientes w são limiarizados pela função de *thresholding* $\delta(\cdot)$. O esquema de atenuação de ruído pode ser expresso por:

$$\tilde{s} = W^{-1} \left(\delta_{\lambda}^{H,S}(Ws) \right), \quad (32)$$

onde \tilde{s} é o sinal reconstruído a partir dos coeficientes obtidos dos operadores do *hard* δ_{λ}^H ou *soft thresholding* δ_{λ}^S , os quais são expressos por:

$$\delta_{\lambda}^H = \begin{cases} w & \forall |w| > \lambda \\ 0 & c. c. \end{cases}, \quad (33)$$

$$\delta_{\lambda}^S = \begin{cases} \text{sign}(w)(|w| - \lambda) & \forall |w| > \lambda \\ 0 & c. c. \end{cases}. \quad (34)$$

Em ambos os casos, o limiar de corte λ é escolhido baseado na energia do sinal e na variância do ruído e será discutido com mais detalhe posteriormente. Se o coeficiente *wavelet* é maior que λ , assume-se que sua contribuição para o sinal é significativa e, portanto, utilizado na reconstrução. Caso contrário, é considerado oriundo do ruído e então descartado. A função para o *hard-thresholding* opera de modo um pouco diferente da função do *soft-thresholding*: enquanto o primeiro mantém ou descarta o valor do coeficiente (*keep or kill*), o último “manipula” os valores

mantidos (*shrinkage*) de modo a deixar a função de *threshold* contínua (Figura 3.6). Uma análise teórica mais detalhada sobre o *soft-thresholding* pode ser encontrada em (Donoho, 1995).

Devido à descontinuidade do *hard thresholding* (Figura 3.6), esta função é conhecida por produzir artefatos no sinal atenuado, especialmente quando o nível de ruído é significativo (Chang et al., 2000). Neste mesmo trabalho, os autores concluem através de seus resultados de simulação que o *soft thresholding* produz um menor erro de estimativa do que o estimador do *hard thresholding* e, por isso, geralmente escolhido. No entanto, em alguns casos relata-se que o *hard thresholding*, apesar de sua “desvantagem”, resulta numa estimativa superior que o *soft thresholding* para algumas classes de sinais (e.g., Yoon & Vaidyanathan, 2004). Esse fato motivou esses últimos autores a desenvolver um método de *thresholding* diferente, denominado “*Customized Threshold*”, que matematicamente é expressa por:

$$\delta_{\lambda}^c = \begin{cases} w - \text{sign}(w)(1 - \alpha)\lambda & \text{se } |w| \geq \lambda \\ 0 & \text{se } |w| \leq \gamma \\ \alpha\lambda \left(\frac{|w| - \gamma}{\lambda - \gamma}\right)^2 \left\{ (\alpha - 3) \left(\frac{|w| - \gamma}{\lambda - \gamma}\right) + 4 - \alpha \right\} c.c. & \end{cases}, \quad (35)$$

onde $0 < \gamma < \lambda$ e $0 \leq \alpha \leq 1$.

O valor de γ , neste caso, define o valor abaixo no qual os valores são descartados e o valor de α define a “forma” da função de *thresholding* δ_{λ}^c . A função do “*Customized Threshold*” pode ser vista como uma combinação linear do *soft e hard threshold*, por exemplo, quando $\lambda=1$ e $\gamma=\lambda/2$, ou seja:

$$\alpha \delta_{\lambda}^H + (1 - \alpha)\delta_{\lambda}^S. \quad (36)$$

Note que:

$$\lim_{\alpha \rightarrow 0} \delta_\lambda^C = \delta_\lambda^S \quad e \quad \lim_{\alpha \rightarrow 1, \gamma \rightarrow \lambda} \delta_\lambda^C = \delta_\lambda^H \quad (37)$$

Uma questão importante referente ao *Customized threshold*, é que os parâmetros γ e α são escolhidos baseados na minimização do erro médio quadrático, ou seja, em sinais reais e, em particular, os perfis geofísicos onde não se conhece a forma de onda livre do ruído, isso fica inviável. O valor de γ foi escolhido como sendo $\gamma = \lambda/2$, o qual define o intervalo onde ocorrerá a suavização da função de *threshold*. Já para o valor de gama, foi elaborado um método baseado num critério de classificação litológica: a maior taxa de corretas classificações (amostras classificadas iguais às amostras do testemunho) define α . Para a classificação litológica, utilizou-se o algoritmo “K-Vizinhos Mais Próximos” (*K-Nearest Neighbor* - KNN) (Cover & Hart, 1967). No subitem “4.1.3-Validação” é discutido os pormonores desse procedimento e o apêndice A1 apresenta os conceitos do KNN. Graficamente, as funções de *thresholding* citadas são dadas como na Figura 3.6

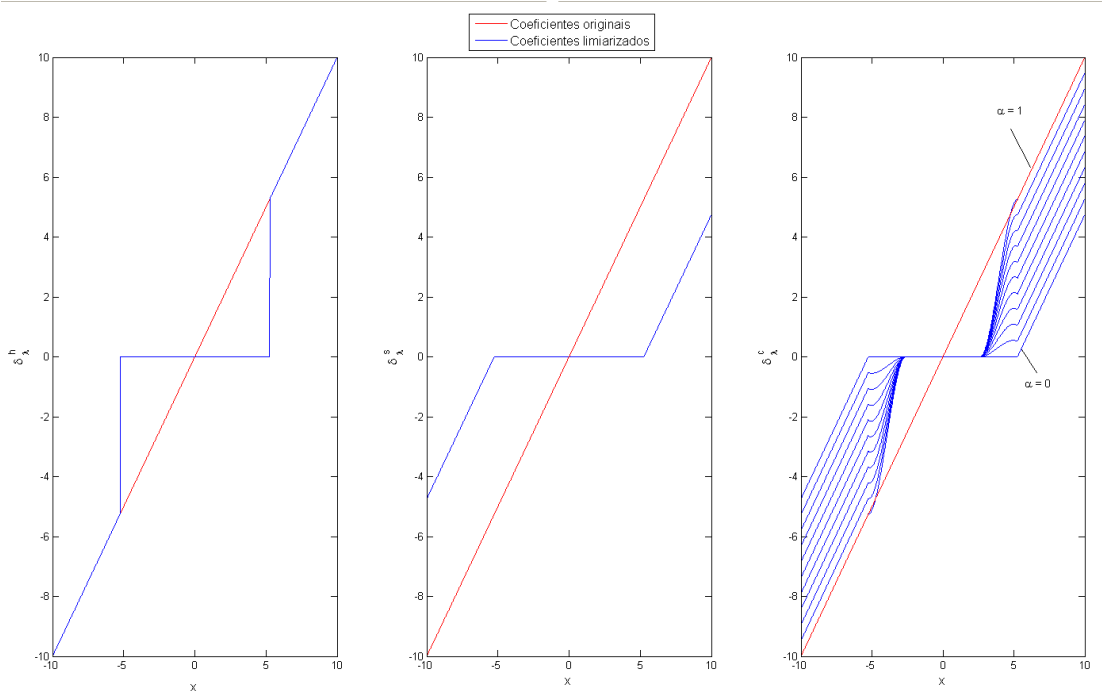


Figura 3.6: *Hard, Soft e Customized*⁷ *Thresholding*

⁷ Para diversos valores de α 's, quando $\lambda = \text{'universal'}$ e $\gamma = \lambda/2$

A determinação do valor de *thresholding* λ é uma questão importante no processo de eliminação de ruído. Um limiar baixo pode resultar num sinal bastante similar ao dado de entrada, no entanto, com a presença de ruído. No outro extremo, os coeficientes que possuem informações relevantes podem ser anulados, tornando o sinal de saída excessivamente “suave”.

As regras que foram utilizadas no presente trabalho estão listadas na Tabela 3-2:

Tabela 3-2: Regras de escolha para os valores de *Thresholding*

Regra do <i>Thresholding</i>	Princípio
'universal'	<i>Threshold</i> fixo proporcional ao logaritmo do número de amostras N
'rigorousSURE'	Baseado no princípio de Estimativa Imparcial de Risco de Stein
'heuristicSURE'	Combinação dos dois métodos acima
'minimax'	Baseado no princípio de minimização do máximo risco
'Lorenz'	Baseado na curva de Lorenz

Donoho & Johnstone (1994) mostraram que, para n variáveis independentes e identicamente distribuídas, o valor máximo esperado é $\sqrt{2 \log_e n}$, o que conduz ao *thresholding* universal, também conhecido como “forma fixa”. Esta é uma das primeiras regras propostas e fornece um *threshold* rápido, automático e fácil (Katul & Vidakovic, 1995). A expressão do *threshold* universal é dada por:

$$\lambda = \hat{\sigma} \sqrt{2 \log_e (n)}, \quad (38)$$

onde $\hat{\sigma}$ é o desvio padrão das amostras (geralmente restrita às componentes de detalhe).

Para a regra 'rigorousSURE', o estimador de risco é baseado no princípio de “estimativa imparcial de risco de Stein” (Stein’s Unbiased Risk Estimate - SURE) (Stein, 1981).

'HeuristicSURE' é uma mistura das regras 'universal' e 'rigorousSURE'. Se a razão sinal-ruído é muito pequena, a estimativa SURE é muito ruidosa. Se essa situação for detectada, o *thresholding* fixo é usado (Misiti et al., 1997). Outra forma de obter o valor de *threshold* é baseada no princípio de minimização do máximo risco, ou ‘minimax’, cuja expressão é dada por:

$$\lambda = 0.3936 + 0.1829 \left(\frac{\log(n)}{\text{Log}(2)} \right), \quad (39)$$

Diante das diferentes técnicas de *thresholding*, Katul & Vidakovic (1995) consideraram o problema relacionado à turbulência atmosférica sobre a óptica da curva de Lorenz (Lorenz, 1905).

A curva de Lorenz foi aplicada inicialmente dentro do contexto da economia, buscando avaliar a desigualdade na distribuição de renda entre as famílias numa determinada sociedade (Figura 3.7).

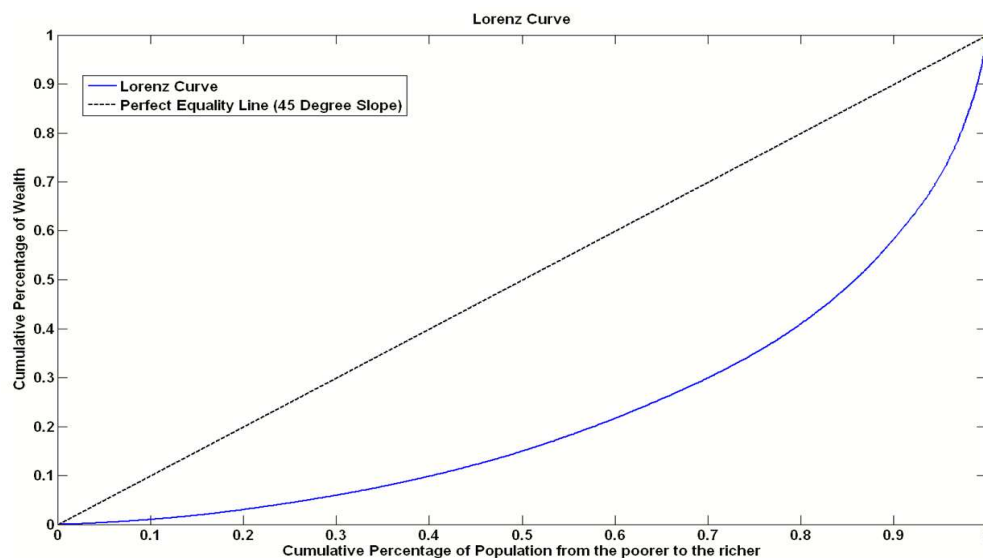


Figura 3.7: Curva de Lorenz

O eixo vertical da curva de Lorenz se refere à porcentagem acumulada de renda e o eixo horizontal à porcentagem acumulada de população, a qual é colocada em ordem crescente de renda individual. Numa situação igualitária de distribuição de renda, a curva de Lorenz coincidiria com a bissetriz do gráfico (*Perfect Equality Line*). No caso de uma desigualdade total, com apenas um indivíduo detentor de toda a renda (última posição no eixo horizontal), a curva de Lorenz seria representada por linha quebrada constituída pelo eixo horizontal e pela linha vertical com início no ponto de 100% da população acumulada. Na prática, a curva de Lorenz situa-se numa posição intermediária entre as duas situações-limite.

Dentro do contexto da atenuação de ruído através da transformada *wavelet*, o conceito da curva de Lorenz é relacionado à energia acumulada contida nos coeficientes *wavelet* (Figura 3.8a). Uma vez que, ao decompor um sinal através da DWT, as energias das componentes de ruído possuem menores amplitudes, o princípio do método é encontrar um limiar onde o acréscimo de coeficientes não é justificado pelo ganho de energia ou, inversamente, a perda de energia não é justificada ao rejeitar uma quantidade maior de coeficientes (Figura 3.8b). Esse limiar corresponde ao ponto onde a inclinação da reta tangente à curva de Lorenz, que coincide com a inclinação da diagonal. A vantagem do *threshold* baseado na curva de Lorenz é que não se requer nenhuma suposição sobre a distribuição do ruído.

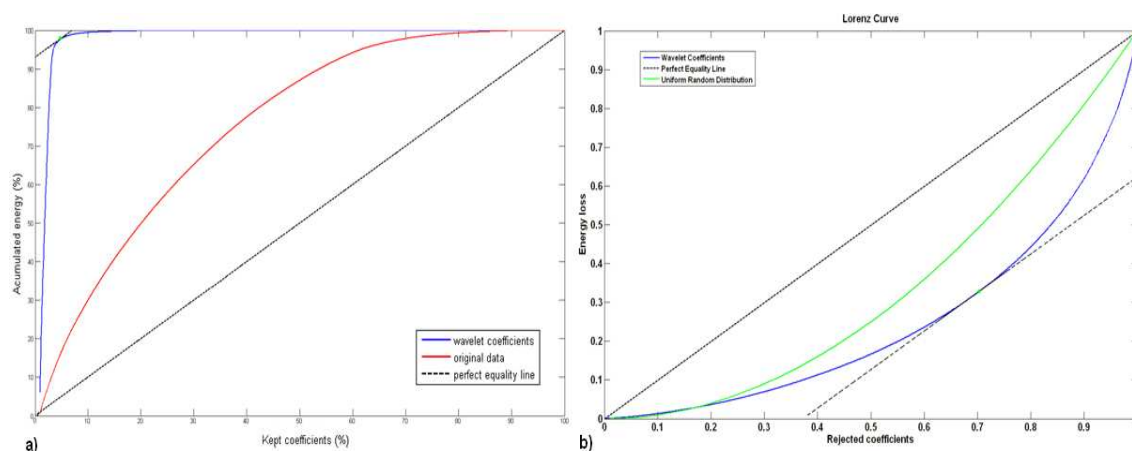


Figura 3.8: a) *Thresholding* baseado no conceito da curva de Lorenz; b) Curva de Lorenz para os coeficientes *wavelet*.

As regras acima possuem suas próprias vantagens e desvantagens nas aplicações de ordem prática. Assim, deve-se escolher a regra adequada para cada problema particular.

3.2.3. Validação

Para avaliar o desempenho dos processos de filtragem, o problema foi analisado segundo duas abordagens: uma para sinais sintéticos e outra para os perfis geofísicos reais.

Os sinais sintéticos serão modelados para diferentes formas de onda: ‘blocks’, ‘bumps’, ‘heavy sine’, ‘doppler’ e ‘quadchirp’. Uma vez que se conhece a função que descreve o sinal, o desempenho pode ser avaliado calculando o erro médio quadrático: quanto menor o erro, melhor a estimativa do sinal e, portanto, mais eficiente o processo de atenuação do ruído.

$$\|f - \tilde{F}\|^2 \triangleq \frac{1}{N} \sum_{i=0}^{N-1} |f_i - \tilde{F}_i|^2. \quad (40)$$

Com o intuito de deixar a análise mais robusta, o processo de atenuação foi repetido dez vezes, sendo que a cada nova iteração, uma nova forma de ruído foi adicionada.

Para os perfis geofísicos, foi adotado um segundo critério: uma vez que o dado analisado se refere às repostas físicas do meio geológico, uma forma quantitativa para avaliar como a escolha dos parâmetros influencia o processo de eliminação de ruído pode ser obtida através de um critério de classificação litológica. A premissa básica é que se um determinado sinal contaminado por ruído é utilizado como dado de entrada num processo de classificação, espera-se que o desempenho (taxa de acerto na classificação) obtido seja inferior quando comparado ao utilizar um sinal mais “limpo”. Inversamente, um sinal mais fiel às características geológicas resultará numa maior taxa de corretas classificações. Assim, foi adotado o algoritmo KNN para a classificação litológica e avaliação do processo.

Para corroborar a utilização da técnica de classificação como critério de avaliação do processo, o perfil geofísico foi avaliado para diversas escolhas de *wavelets* base. Uma vez que a *wavelet* mais apropriada pode ser obtida a partir da entropia de Shannon, foi verificada a correspondência entre a *wavelet* obtida no processo de classificação litológica e a *wavelet* obtida pela entropia de Shannon.

No presente trabalho, foi avaliado o desempenho com valores de k entre 1-25. Como se trata de um algoritmo de classificação supervisionada, é necessário um conjunto de treino (ou aprendizagem). Neste caso, o conjunto de treino se refere aos trechos do poço que são testemunhados. O poço estudado possui 12 classes litológicas (Tabela 3-3). O algoritmo classificador foi treinado com 50%, 60% e 70% das amostras de cada classe, retiradas aleatoriamente do conjunto de dados. Uma vez que as amostras de treino foram escolhidas aleatoriamente, optou-se por repetir o processo de classificação dez vezes. A validação é dada pela porcentagem de acertos entre as eletrofácies classificadas (amostras não escolhidas para treino) através do algoritmo KNN e as respectivas amostras do testemunho. Esse mesmo procedimento foi adotado para a escolha do parâmetro α na função do *customized threshold*.

Tabela 3-3: Litologia do poço estudado

Classe	Descrição
1	Diamictito Arenoso Lamoso
2	Conglomerados Residuais
3	Arenito Grosso, Amalgamado
4	Arenito Médio Fino Laminado
5	Arenito Médio Gradado ou Maciço
6	Arenito Médio Cimentado
7	Arenito/Folhelho Interestratificado
8	Arenito/Folhelho Finamente Interestratificado
9	Siltito Argiloso Estratificado
10	Interlaminado Siltito Argiloso e Marga
11	Interlaminado Arenoso Bioturbado
12	Arenito Cimentado, com Feições de Escorregamento

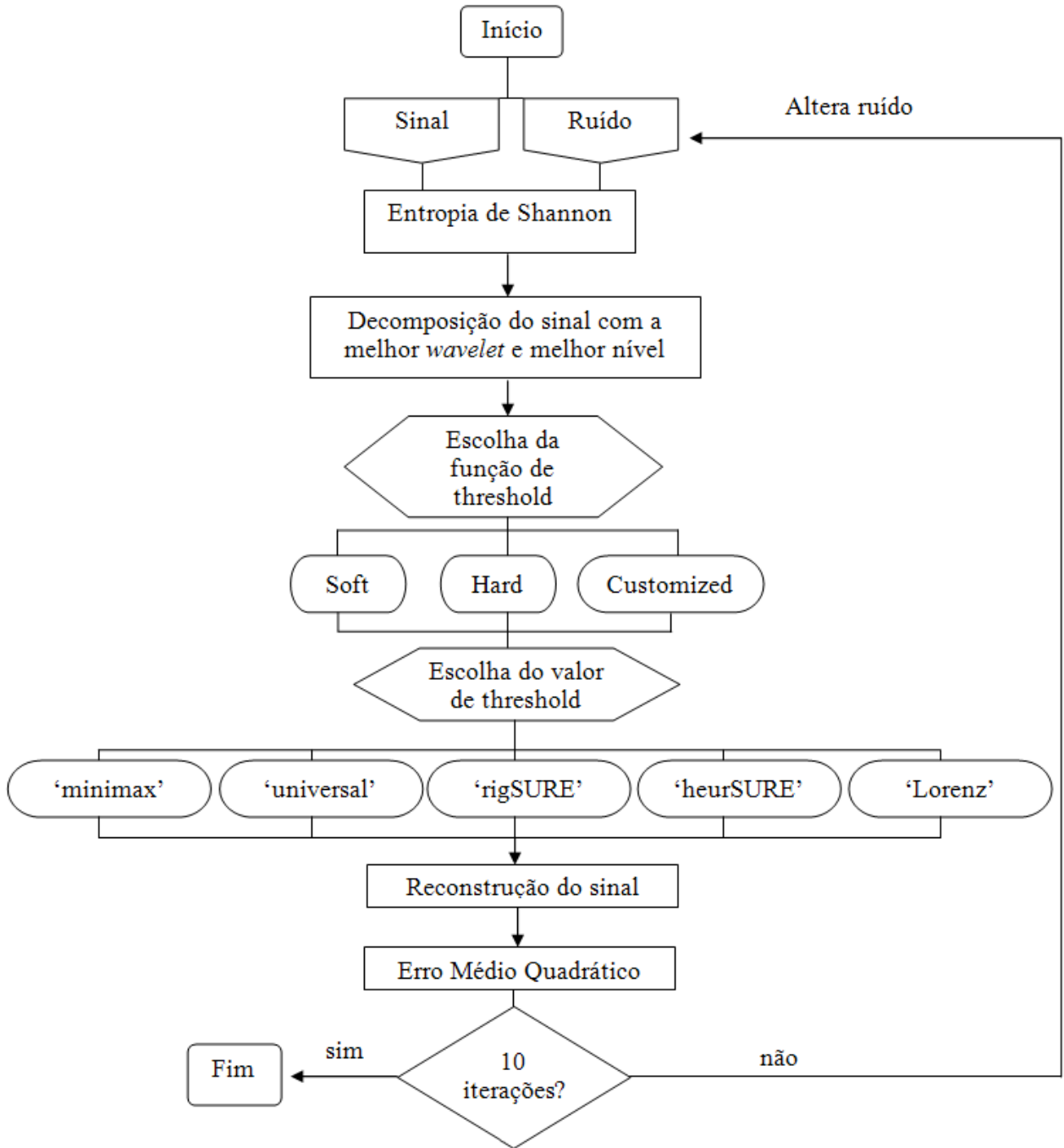
Reservatório

Possível Reservatório

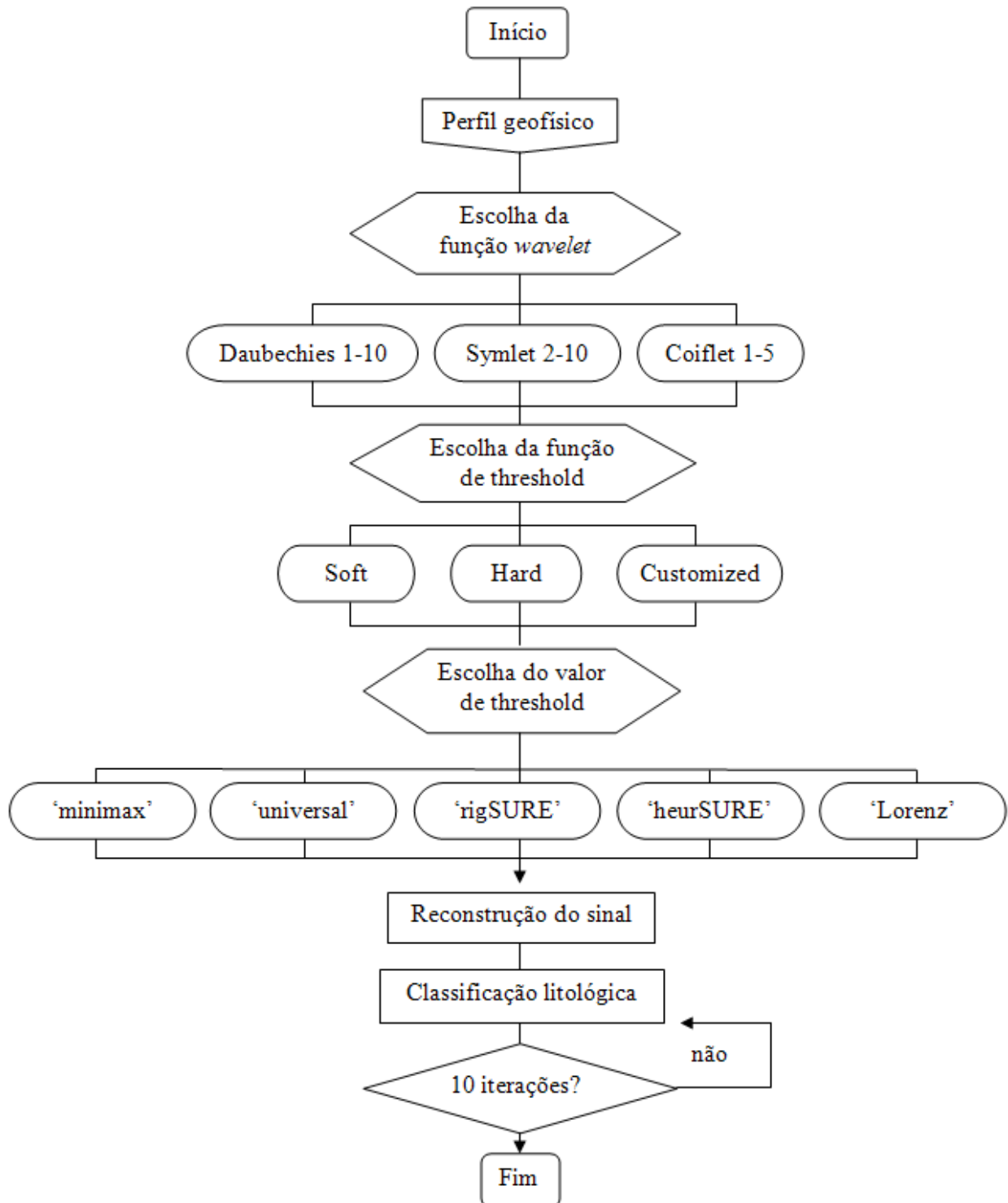
Não Reservatório

3.2.4. Fluxograma do método

Dados sintéticos



Perfis de poços



CAPÍTULO 4

RESULTADOS

4.1.Dados sintéticos

Para cada um dos sinais analisados, ruído branco gaussiano foi adicionado (SNR=7dB). Uma vez que esse ruído é gerado aleatoriamente, o processo foi repetido 10 vezes com o intuito de deixar a análise mais robusta.

O erro médio quadrático para cada conjunto de sinal, regra e função de *threshold* é apresentado nas tabelas abaixo. Os valores são referentes às médias após 10 iterações:

Tabela 4-1: Erro médio quadrático para diferentes sinais e esquemas de atenuação de ruído

SINAL	Regra	Universal	Rigrsure	Heursure	Minimax	Lorenz	Wavelet/nível
	Função						
Blocks	Soft	1,79	12,39	0,49	0,69	13,72	Db1 / 8
	Hard	0,27	11,23	0,72	0,46	4,29	
	Custom	0,26	0,81	0,72	0,44	1,17	
Bumps	Soft	1,91	0,64	0,84	0,95	4,46	Coif4 / 3
	Hard	0,96	0,91	0,78	0,67	2,11	
	Custom	0,64	0,63	0,63	0,58	1,47	
Heavy sine	Soft	0,52	0,24	0,23	0,27	1,32	Db3 / 5
	Hard	0,30	0,47	0,26	0,46	0,78	
	Custom	0,24	0,24	0,21	0,25	0,40	
Doppler	Soft	1,56	0,49	0,69	0,77	4,03	Sym4 / 5
	Hard	0,64	0,82	0,70	0,65	1,83	
	Custom	0,45	0,49	0,56	0,54	0,97	
Quadchirp	Soft	3,52	0,78	0,74	1,49	10,79	Db8 / 4
	Hard	0,87	0,97	0,97	0,75	3,03	
	Custom	0,67	0,77	0,74	0,71	1,25	

Em todas as situações, a função de *threshold* customizada gerou os melhores resultados. Como é possível observar, a regra do *threshold* é um parâmetro que depende do sinal avaliado, de modo que o problema deve ser analisado na sua particularidade. A Figura 4.1 mostra os sinais estudados bem como os respectivos sinais com o ruído atenuado:

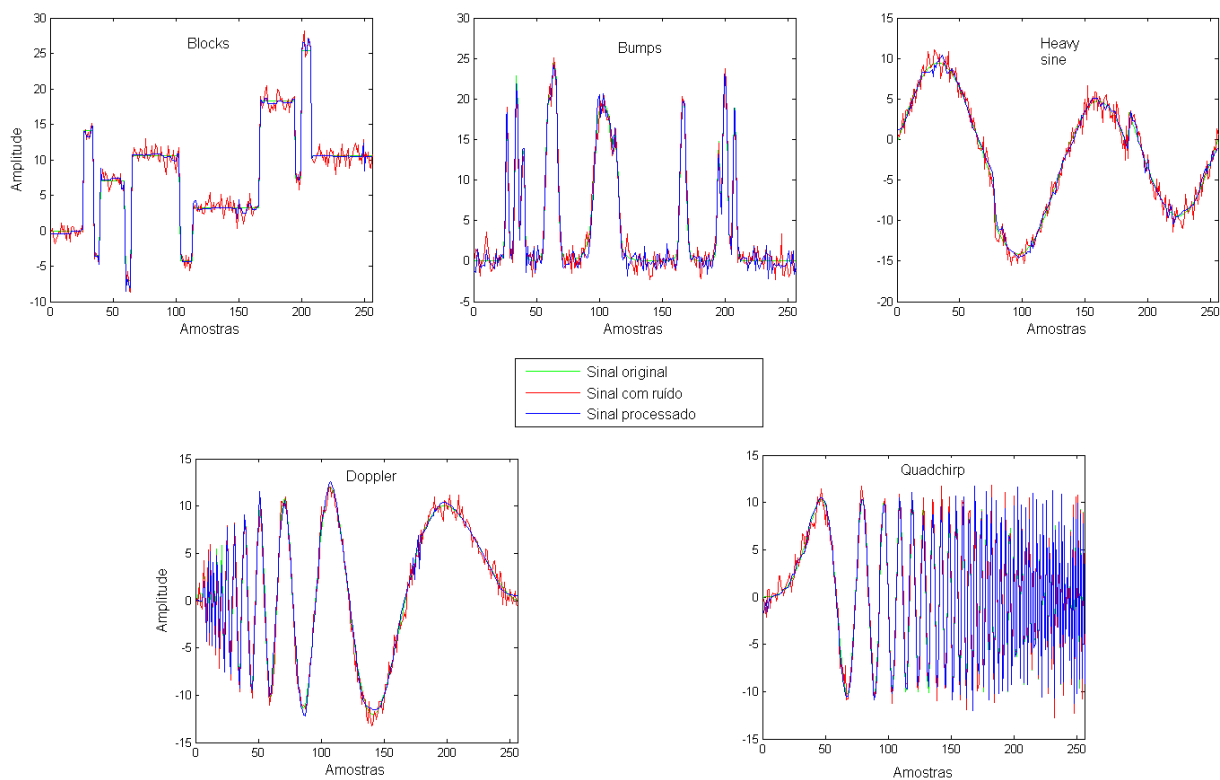


Figura 4.1: Sinais sintéticos antes e pós processamento

Na Figura 4.2, é ilustrada a eficiência da função de *threshold* customizada com diferentes valores de alfa para o sinal ‘Heavy sine’:

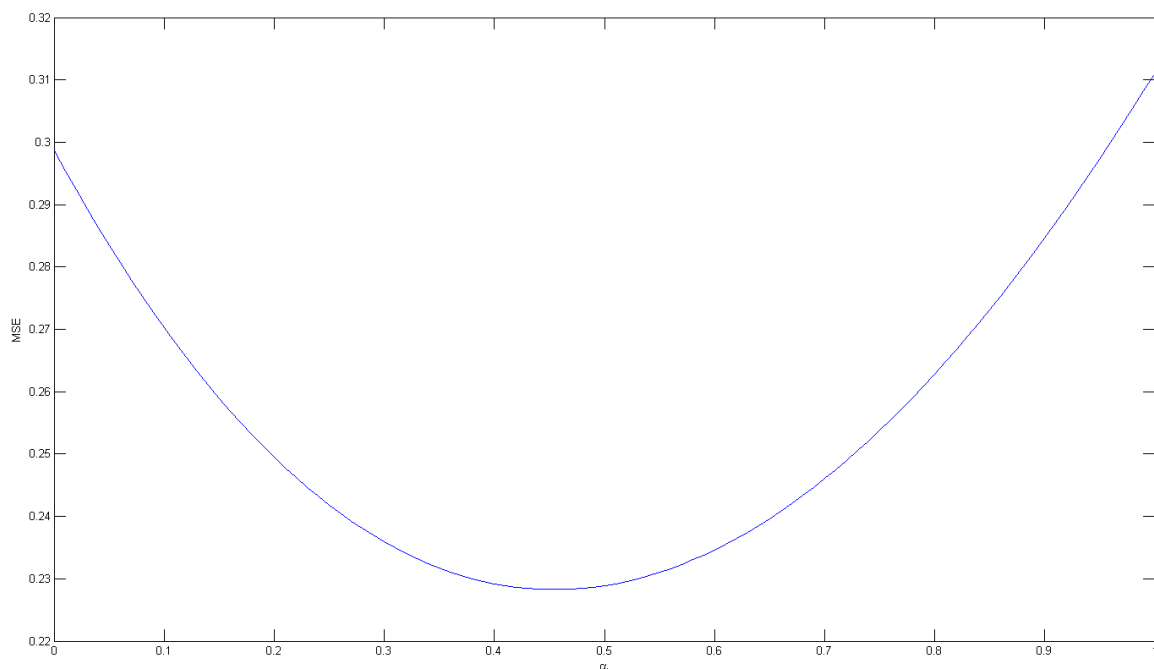


Figura 4.2: Eficiência da função de *threshold* customizada para o sinal 'Heavy sine'. O ponto de mínimo define o melhor alfa (menor MSE)

Como fica evidente a partir da Figura 4.2, entre os dois extremos que definem o *soft* e o *hard thresholding* existem valores que tornam o MSE um tanto quanto menor. A superioridade da função de *threshold* customizada frente às demais impulsionou a adaptação dessa função para a análise de perfis geofísicos, a qual é apresentada a seguir.

4.2. Perfis de poços

Por meio da combinação dos diferentes parâmetros, a melhor escolha foi definida pelo maior acerto na determinação das eletrofácies, validação realizada através dos trechos do testemunho. A comparação é feita analisando somente as maiores taxas de acerto na classificação dentro das 25 possíveis (número de vizinhos) para cada sinal.

O *threshold* pode ser global, no qual um único limiar é aplicado em todas as escalas, pode ser dependente de escala, onde um valor (possivelmente) diferente de *threshold* é aplicado para cada escala, ou mesmo zonal, no qual um segmento do sinal é dividido em várias zonas e, para cada zona, um limiar diferente é aplicado (Mallat, 1989). Essa diferenciação está relacionada ao tipo do ruído. No presente trabalho, foram consideradas três situações distintas: No primeiro caso, aplica-se um esquema de *threshold* global (ruído branco gaussiano $N(0, \sigma=1)$). Nos segundo e terceiro casos, um esquema de *threshold* dependente da escala é aplicado: é avaliado o desvio padrão das amostras de detalhe no primeiro nível (D1) e então, os valores de *threshold* são multiplicados por este fator (*scaled white noise*); esse processo é repetido avaliando o desvio padrão em cada nível (*nonwhite noise*).

Por simplicidade, os resultados aqui apresentados são referentes à classificação utilizando 50% das amostras para treinamento do algoritmo KNN. Os demais resultados encontram-se no Anexo A:

- Ruído branco Gaussiano:

Tabela 4-2: Corretas classificações utilizando 50% das amostras como treino e diferentes esquemas de atenuação de ruído (Ruído branco Gaussiano)

FUNÇÃO	<i>wavelet</i>	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
	regra								
SOFT	Universal	57,40	42,66	42,15	42,88	40,23	42,49	40,45	41,07
	Rigsure	36,95	36,95	37,57	37,01	38,08	36,50	35,76	35,71
	Heursure	37,91	36,67	37,06	37,74	37,46	35,93	36,61	36,72
	Minimaxi	51,02	40,73	40,79	40,00	38,59	40,51	38,76	38,31
	Lorenz	52,15	45,82	48,14	46,38	46,38	47,74	47,51	48,64
HARD	Universal	56,16	39,15	36,10	35,65	36,67	38,08	37,18	38,19
	Rigsure	36,89	35,54	36,38	37,06	37,51	36,84	36,16	36,72
	Heursure	36,38	36,55	37,23	38,25	37,46	36,16	36,67	36,50
	Minimaxi	49,49	37,01	35,93	37,40	35,31	36,16	35,42	34,92
	Lorenz	52,15	45,82	48,14	46,38	46,38	47,74	47,51	48,64
CUSTOM	Universal	59,44	43,95	42,71	43,22	40,85	43,39	41,98	41,30
	Rigsure	38,08	39,10	38,42	38,64	38,53	38,36	37,80	37,85
	Heursure	38,53	37,85	38,64	38,93	38,53	38,14	38,14	37,68

4. RESULTADOS

	Minimaxi	52,60	42,32	41,02	40,62	40,06	41,30	41,19	40,17
	Lorenz	52,15	45,82	48,14	46,38	46,38	47,74	47,51	48,64
FUNÇÃO	<i>wavelet</i> regra	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
SOFT	Universal	39,44	40,73	43,45	42,88	41,13	42,20	42,20	40,00
	Rigsure	36,16	35,88	35,88	36,84	36,44	36,38	36,27	36,44
	Heursure	36,78	35,59	37,34	35,42	37,57	36,67	36,33	35,14
	Minimaxi	38,70	40,00	40,79	39,60	38,70	40,51	39,72	38,25
	Lorenz	44,92	51,07	45,99	48,14	52,03	36,10	49,66	48,53
HARD	Universal	35,03	37,40	39,15	35,20	36,61	39,10	35,93	35,42
	Rigsure	37,12	36,21	36,05	35,48	35,25	37,18	36,16	38,02
	Heursure	36,72	36,50	36,78	35,25	37,06	37,12	36,72	36,44
	Minimaxi	35,14	38,81	37,18	36,05	36,33	36,67	35,93	35,93
	Lorenz	44,92	51,07	45,99	48,14	52,03	36,10	49,66	48,53
CUSTOM	Universal	40,06	42,26	44,01	43,28	42,77	42,71	41,53	41,64
	Rigsure	37,97	38,08	38,25	38,87	37,91	37,51	39,10	38,76
	Heursure	38,02	37,91	37,80	38,25	37,97	37,23	37,85	38,42
	Minimaxi	38,47	40,62	41,53	40,28	41,19	41,92	41,47	39,72
	Lorenz	44,92	51,07	45,99	48,14	52,03	36,10	49,66	48,53
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	41,41	40,79	41,30	41,30	42,20	42,26	42,15	41,53
	Rigsure	36,84	36,78	36,10	36,67	37,34	36,44	36,33	37,91
	Heursure	36,38	36,67	36,10	37,01	36,44	37,68	38,25	37,63
	Minimaxi	39,77	40,06	39,44	39,89	40,56	39,94	40,96	39,72
	Lorenz	47,18	44,63	49,66	45,25	49,49	48,36	49,27	45,37
HARD	Universal	36,55	36,61	37,40	38,02	38,14	35,08	36,89	35,82
	Rigsure	37,74	35,48	35,65	36,10	37,23	36,27	36,44	36,84
	Heursure	35,93	37,18	34,75	36,55	35,82	36,61	36,78	37,34
	Minimaxi	37,29	36,61	36,72	35,54	37,12	37,34	37,57	36,50
	Lorenz	47,18	44,63	49,66	45,25	49,49	48,36	49,27	45,37
CUSTOM	Universal	42,82	43,11	42,15	43,56	42,54	43,67	44,24	43,28
	Rigsure	38,36	38,02	38,25	38,25	38,02	38,25	38,53	37,91
	Heursure	38,36	38,14	38,19	38,14	38,08	38,02	37,91	38,47
	Minimaxi	41,24	40,68	40,73	40,68	41,24	40,56	41,92	40,85
	Lorenz	47,18	44,63	49,66	45,25	49,49	48,36	49,27	45,37

- *Scaled white noise*

Tabela 4-3: Corretas classificações utilizando 50% das amostras como treino e diferentes esquemas de atenuação de ruído (*Scaled white noise*)

FUNÇÃO	<i>wavelet</i>	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
	regra								
SOFT	Universal	60,23	40,85	40,40	38,36	38,93	41,24	38,98	38,36
	Rigsure	38,47	35,48	35,88	35,76	37,23	36,05	35,88	36,38
	Heursure	38,36	36,61	36,67	36,78	36,89	35,82	36,38	36,67
	Minimaxi	53,79	39,32	38,98	38,81	38,59	39,32	37,68	37,85
	Lorenz	54,41	45,93	46,27	49,21	43,50	42,03	41,81	45,71
HARD	Universal	55,99	38,59	37,40	38,87	36,16	35,99	38,08	36,95
	Rigsure	36,95	35,37	35,14	36,27	37,51	36,55	36,38	36,95
	Heursure	37,74	35,14	35,99	36,55	35,37	35,37	34,75	35,59
	Minimaxi	54,46	35,54	35,93	36,27	37,51	36,16	37,46	37,63
	Lorenz	54,41	45,93	46,27	49,21	43,50	42,03	41,81	45,71
CUSTOM	Universal	60,96	43,73	41,24	41,86	39,72	41,07	39,55	39,55
	Rigsure	40,85	37,97	38,98	38,25	38,81	38,42	38,59	38,19
	Heursure	41,02	38,25	37,97	38,93	38,02	38,02	38,14	38,02
	Minimaxi	58,31	41,69	39,60	40,11	39,60	39,60	38,93	39,15
	Lorenz	54,41	45,93	46,27	49,21	43,50	42,03	41,81	45,71
FUNÇÃO	<i>wavelet</i>	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
	regra								
SOFT	Universal	37,68	39,55	43,62	40,00	39,49	41,47	39,49	37,40
	Rigsure	36,16	36,10	35,93	35,93	34,69	35,88	36,05	35,48
	Heursure	37,23	36,44	35,31	37,23	36,10	36,44	35,20	36,61
	Minimaxi	37,57	38,14	40,06	39,27	38,93	38,98	37,12	36,95
	Lorenz	43,45	41,30	46,33	45,37	47,12	38,53	45,37	41,81
HARD	Universal	35,54	37,06	38,14	36,21	37,29	36,44	36,89	35,20
	Rigsure	35,76	36,61	36,10	36,05	36,21	36,95	36,16	36,38
	Heursure	36,61	36,67	36,16	35,37	37,12	35,65	36,50	36,55
	Minimaxi	35,48	37,51	37,34	36,84	36,27	34,52	34,97	35,76
	Lorenz	43,45	41,30	46,33	45,37	47,12	38,53	45,37	41,81
CUSTOM	Universal	38,70	40,00	43,90	41,24	40,45	41,86	40,40	38,87
	Rigsure	38,47	38,14	38,14	37,63	38,08	37,68	38,19	37,91
	Heursure	38,14	38,59	38,19	38,47	38,25	37,97	37,80	38,31
	Minimaxi	38,08	39,60	41,98	39,60	39,27	40,00	39,38	38,81

	Lorenz	43,45	41,30	46,33	45,37	47,12	38,53	45,37	41,81
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	38,87	39,32	37,85	41,58	39,66	39,89	39,72	38,31
	Rigsure	37,06	36,27	36,10	37,06	37,63	35,20	36,50	36,05
	Heursure	36,16	36,44	36,84	35,88	35,99	36,27	35,93	36,21
	Minimaxi	39,60	40,17	37,18	40,51	38,70	38,81	38,14	37,68
	Lorenz	43,33	42,20	44,58	43,79	44,52	45,03	46,16	44,18
HARD	Universal	36,38	36,05	36,05	38,93	37,23	36,05	36,84	35,65
	Rigsure	35,37	36,16	35,76	37,18	36,21	35,48	35,31	36,21
	Heursure	36,21	35,93	36,67	36,38	36,10	34,92	36,61	35,42
	Minimaxi	36,27	36,38	35,93	38,19	35,31	36,50	34,80	36,27
	Lorenz	43,33	42,20	44,58	43,79	44,52	45,03	46,16	44,18
CUSTOM	Universal	40,40	39,27	40,11	42,77	41,02	40,00	41,30	40,00
	Rigsure	38,19	37,91	37,80	38,42	38,19	37,74	37,85	37,68
	Heursure	37,85	38,19	37,68	37,85	38,36	37,74	37,80	37,97
	Minimaxi	39,04	39,21	38,98	40,73	40,28	39,15	39,38	38,64
	Lorenz	43,33	42,20	44,58	43,79	44,52	45,03	46,16	44,18

- *Nonwhite noise*

Tabela 4-4: Corretas classificações utilizando 50% das amostras como treino e diferentes esquemas de atenuação de ruído (*Nonwhite noise*)

FUNÇÃO	<i>wavelet</i> regra	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
SOFT	Universal	50,79	43,45	42,32	45,48	46,61	38,53	42,49	43,39
	Rigsure	44,63	39,49	39,55	43,39	41,92	42,03	38,36	40,73
	Heursure	45,82	39,10	39,10	41,53	41,58	38,93	37,51	39,44
	Minimaxi	52,37	44,80	41,86	43,33	43,62	39,60	36,89	42,20
	Lorenz	54,24	42,60	50,62	46,44	47,46	48,42	45,82	47,74
HARD	Universal	49,21	41,86	40,17	42,60	37,91	43,05	36,78	43,39
	Rigsure	40,17	38,02	33,28	40,51	35,03	35,31	35,37	34,52
	Heursure	40,23	37,40	35,99	36,16	35,37	38,36	34,97	37,57
	Minimaxi	51,81	42,54	41,36	42,37	40,34	39,89	36,89	35,99
	Lorenz	54,24	42,60	50,62	46,44	47,46	48,42	45,82	47,74
CUSTOM	Universal	53,16	48,19	43,39	46,84	47,85	44,46	42,94	43,05
	Rigsure	44,86	40,28	41,13	44,86	43,39	42,71	40,00	42,49
	Heursure	45,76	40,73	39,89	42,94	41,47	39,55	38,93	41,69

4. RESULTADOS

	Minimaxi	55,65	46,27	43,22	45,20	43,90	42,66	40,73	44,92
	Lorenz	54,24	42,60	50,62	46,44	47,46	48,42	45,82	47,74
FUNÇÃO	<i>wavelet</i> regra	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
SOFT	Universal	38,25	42,54	45,20	42,66	49,77	36,50	46,84	44,29
	Rigsure	41,47	42,71	38,53	40,28	39,83	40,11	41,07	43,45
	Heursure	33,33	41,19	39,77	38,25	39,89	38,53	39,15	43,39
	Minimaxi	40,62	42,60	46,55	41,64	45,37	42,54	44,97	38,64
	Lorenz	46,55	47,46	42,82	51,41	51,47	38,08	51,81	47,68
HARD	Universal	39,27	40,73	42,71	40,79	42,88	41,30	40,96	41,41
	Rigsure	35,71	40,68	35,93	32,94	37,74	34,75	37,23	35,48
	Heursure	36,50	45,20	35,42	34,86	35,14	34,52	34,92	37,18
	Minimaxi	36,05	39,60	45,88	41,58	41,47	39,94	41,30	40,68
	Lorenz	46,55	47,46	42,82	51,41	51,47	38,08	51,81	47,68
CUSTOM	Universal	43,33	45,93	48,08	44,58	49,89	40,96	49,44	45,03
	Rigsure	42,43	44,07	40,23	40,73	42,82	41,47	43,22	44,46
	Heursure	38,08	47,68	41,13	39,38	42,88	39,66	39,44	43,84
	Minimaxi	41,47	45,88	46,50	43,95	47,29	42,99	46,55	43,22
	Lorenz	46,55	47,46	42,82	51,41	51,47	38,08	51,81	47,68
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	45,08	40,62	44,52	46,55	49,32	42,26	42,20	43,50
	Rigsure	42,37	36,44	42,49	40,96	41,41	40,28	43,45	43,22
	Heursure	42,49	37,18	36,95	37,06	37,01	40,85	45,59	40,85
	Minimaxi	42,66	39,89	45,48	42,66	47,12	42,37	43,56	42,94
	Lorenz	46,50	45,88	46,27	44,92	52,09	46,78	49,38	47,68
HARD	Universal	43,67	40,06	40,06	43,62	43,67	39,49	44,69	38,53
	Rigsure	37,97	35,48	37,80	37,97	36,50	36,78	37,97	37,51
	Heursure	40,00	35,54	39,27	34,46	33,95	35,93	38,14	33,50
	Minimaxi	40,11	39,04	39,27	36,21	42,32	39,21	41,58	39,66
	Lorenz	46,50	45,88	46,27	44,92	52,09	46,78	49,38	47,68
CUSTOM	Universal	45,88	44,18	46,38	46,67	52,37	44,46	45,88	44,92
	Rigsure	44,92	40,62	43,39	42,03	42,26	40,85	46,33	43,28
	Heursure	43,67	39,55	40,28	40,56	37,97	41,36	46,27	42,09
	Minimaxi	48,19	43,05	46,05	43,90	48,08	44,01	45,76	44,80
	Lorenz	46,50	45,88	46,27	44,92	52,09	46,78	49,38	47,68

Em geral, todos os sinais processados melhoraram a taxa de corretas classificações. No entanto, os maiores valores são encontrados ao utilizar o sinal processado com a *wavelet* Db1, a função de *threshold* customizada, o valor de *threshold* obtido pela regra ‘universal’ com correção a partir da componente de detalhe D1 (*scaled white noise*). Para o sinal processado com esses parâmetros, obteve-se 60,96% para as taxas de corretas classificações.

Quando comparada à classificação com o dado original (35,96%), o sinal processado com tais parâmetros, o método proposto proporcionou uma melhora de 25% quando utilizados 50% das amostras como treino no KNN.

A mesma característica é observada para diferentes valores do conjunto de treino do KNN (Anexo A). Quando utilizados 60% e 70% de amostras como treino e esses mesmos parâmetros, obteve-se 64,17% e 67,52% de corretas classificações, frente a 36,21% e 36,44% obtidos com o dado original, o que representa uma melhora de 27,96% e 31,08%.

Quando avaliada a entropia de Shannon para a seleção da melhor função base, obteve-se o seguinte resultado:

Tabela 4-5: Entropia de Shannon para o perfil GR

Função	Entropia	Função	Entropia
Db1	0,5199	Sym4	0,5254
Db2	0,5203	Sym5	0,5277
Db3	0,5249	Sym6	0,5311
Db4	0,5300	Sym7	0,5351
Db5	0,5329	Sym8	0,5368
Db6	0,5372	Sym9	0,5393
Db7	0,5414	Sym10	0,5418
Db8	0,5449	Sym11	0,5456
Db9	0,5488	Sym12	0,5477
Db10	0,5520	Sym13	0,5508
Db11	0,5554	Sym14	0,5530
Db12	0,5588	Sym15	0,5559
Db13	0,5620	Sym16	0,5570
Db14	0,5647	Sym17	0,5584
Db15	0,5678	Sym18	0,5605
Db16	0,5702	Sym19	0,5649
Db17	0,5728	Sym20	0,5660
Db18	0,5753	Coif1	0,5233
Db19	0,5780	Coif2	0,5340
Db20	0,5807	Coif3	0,5435
Sym2	0,5203	Coif4	0,5521
Sym3	0,5249	Coif5	0,5598

De fato, a função base obtida pelo método proposto (Db1) é corroborada pela minimização da entropia de Shannon, sugerindo que os demais parâmetros obtidos também são os mais apropriados para a análise do dado em questão.

A adaptação proposta para a função de *threshold* customizada se mostrou interessante, pois obteve as maiores taxas de corretas classificações e representou uma melhora frente ao *hard* e *soft threshold*. O comportamento do algoritmo classificador utilizando o sinal processado com esta função pode ser avaliado através da Figura 4.3, onde se pode obter o valor de alfa mais apropriado:

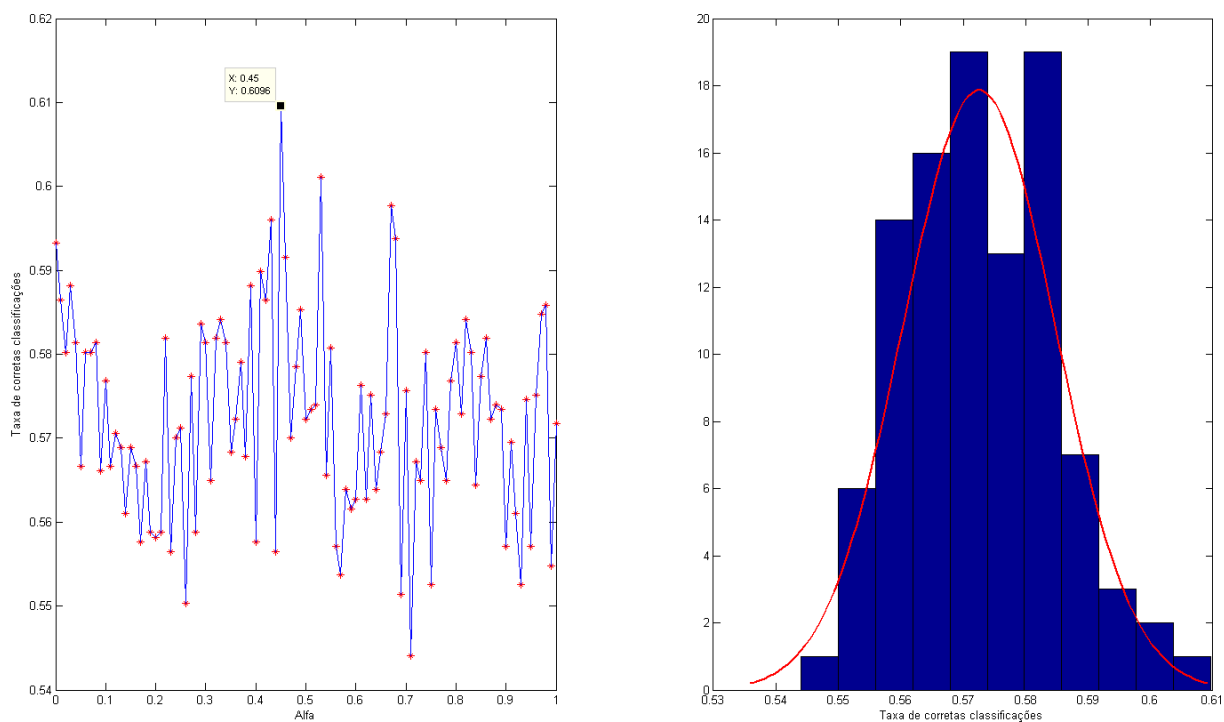


Figura 4.3: Taxa média de corretas classificações para diferentes valores de alfa no “*customized threshold*”

A Figura 4.4 mostra os perfis, original e filtrado pelo método proposto, do poço estudado:

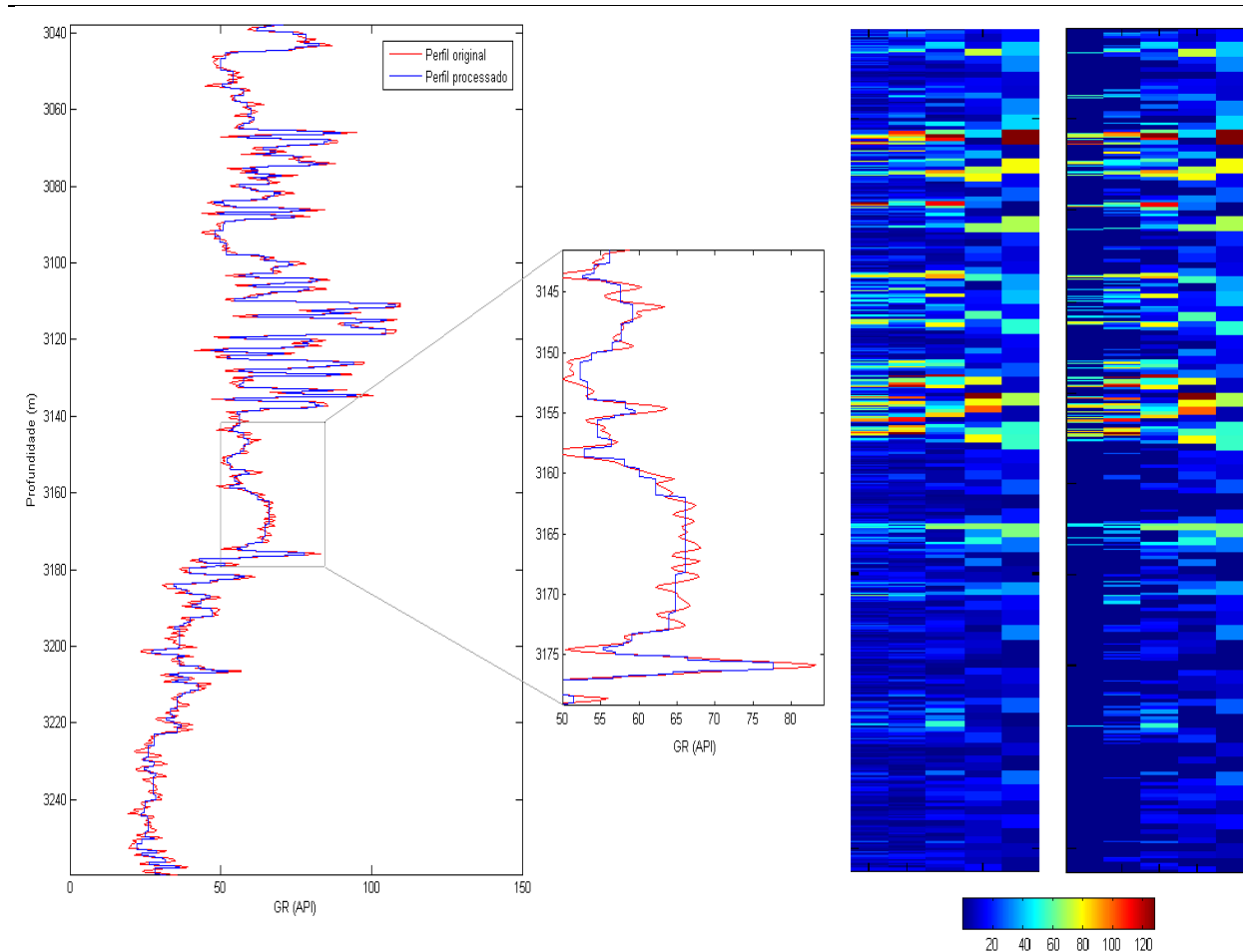


Figura 4.4: Perfis de Raios Gama original e filtrado e os respectivos escalogramas

O desempenho na classificação litológica para diferentes valores de números de vizinhos do algoritmo KNN é comparado para os perfis original e perfil filtrado (Figura 4.5, Figura 4.7 e Figura 4.9). Além disso, a classificação das eletrofácies obtida com esses dois conjuntos de dados pode ser comparada visualmente com o testemunho, onde se evidencia claramente uma melhor correspondência entre a litologia classificada com o dado processado e a litologia presente no testemunho. Para melhor visualização, as 12 eletrofácies foram agrupadas em três grupos principais: reservatório, possível reservatório e não reservatório (Figura 4.6, Figura 4.8 e Figura 4.10).

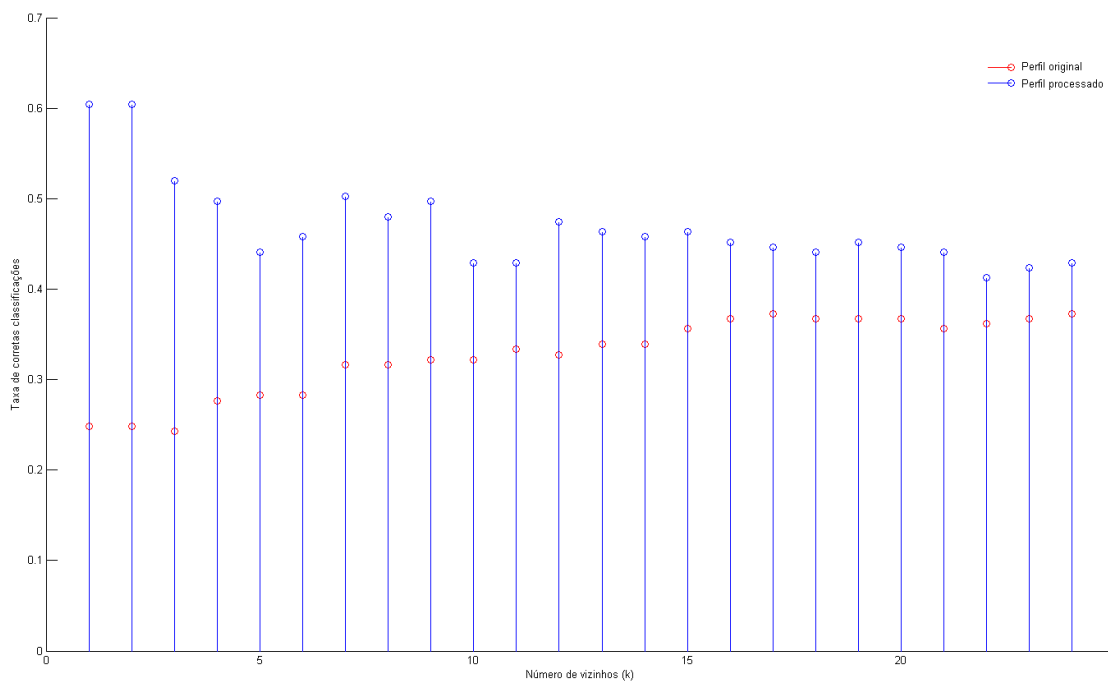


Figura 4.5: Desempenho do KNN utilizando 50% das amostras como treino

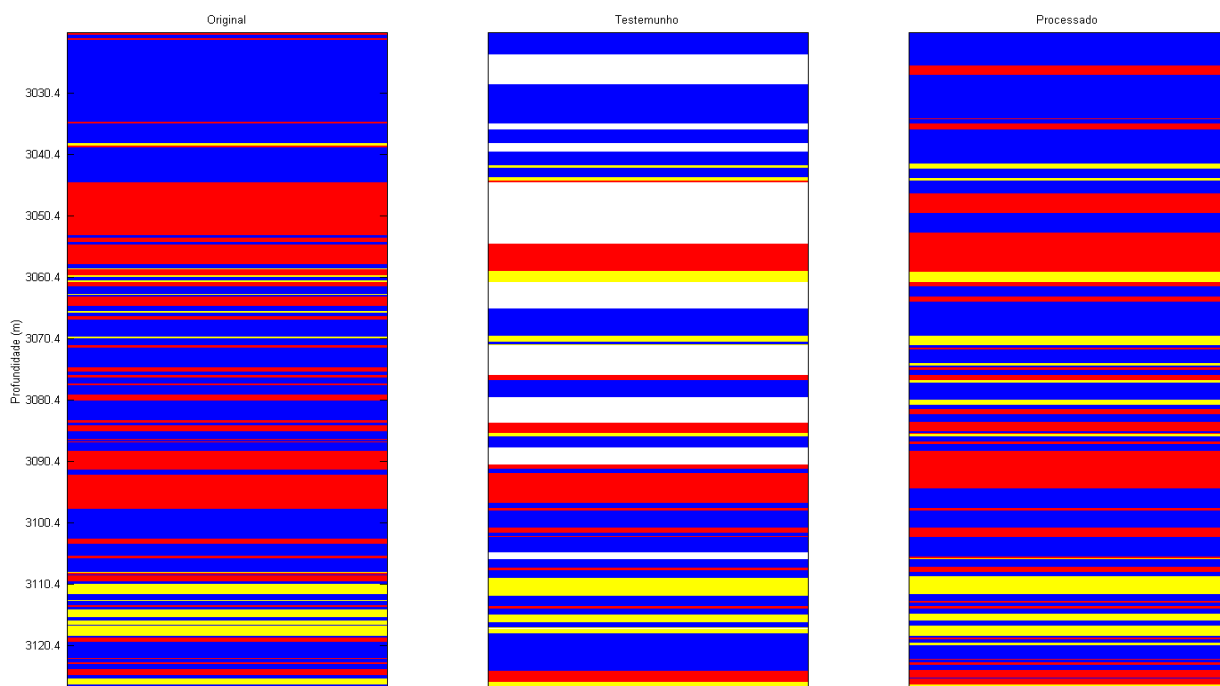


Figura 4.6: Classificação litológica baseada no perfil original (esquerda) e perfil filtrado (direita) comparados com o testemunho (centro) (50% das amostras como treino)

A mesma análise é feita utilizando 60% e 70% das amostras como treino

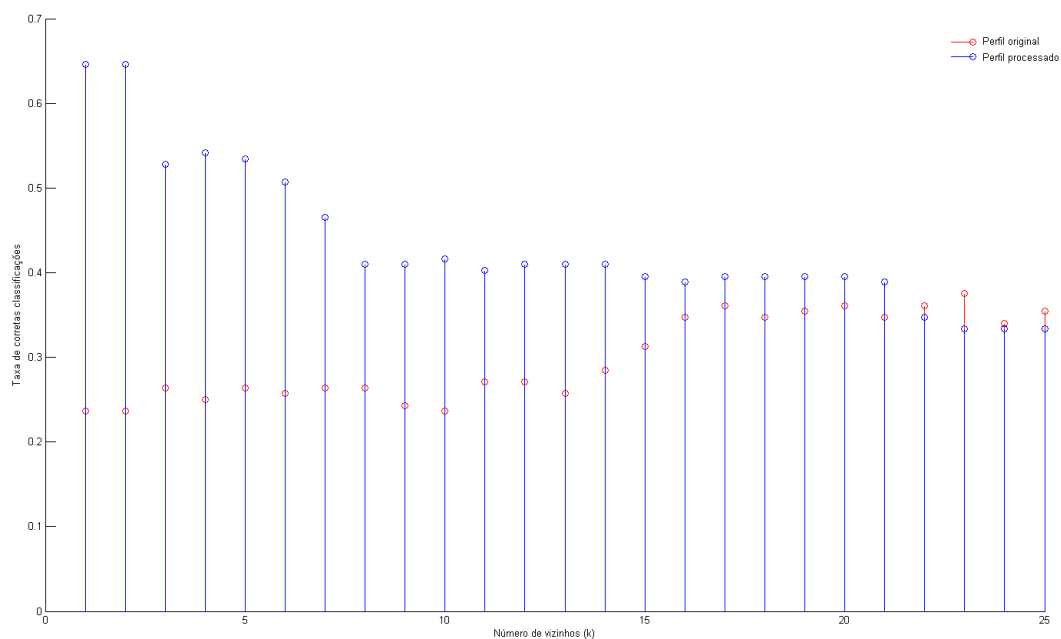


Figura 4.7: Desempenho do KNN utilizando 60% das amostras como treino

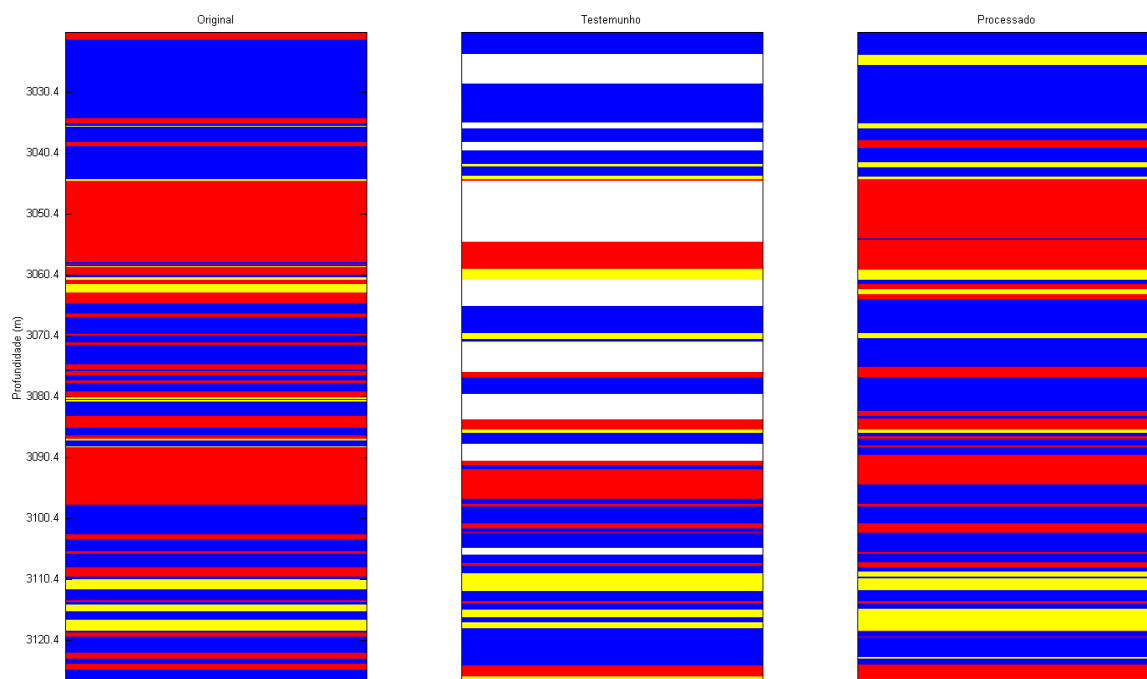


Figura 4.8: Classificação litológica baseada no perfil original (esquerda) e perfil filtrado (direita) comparados com o testemunho (centro) (60% das amostras como treino)

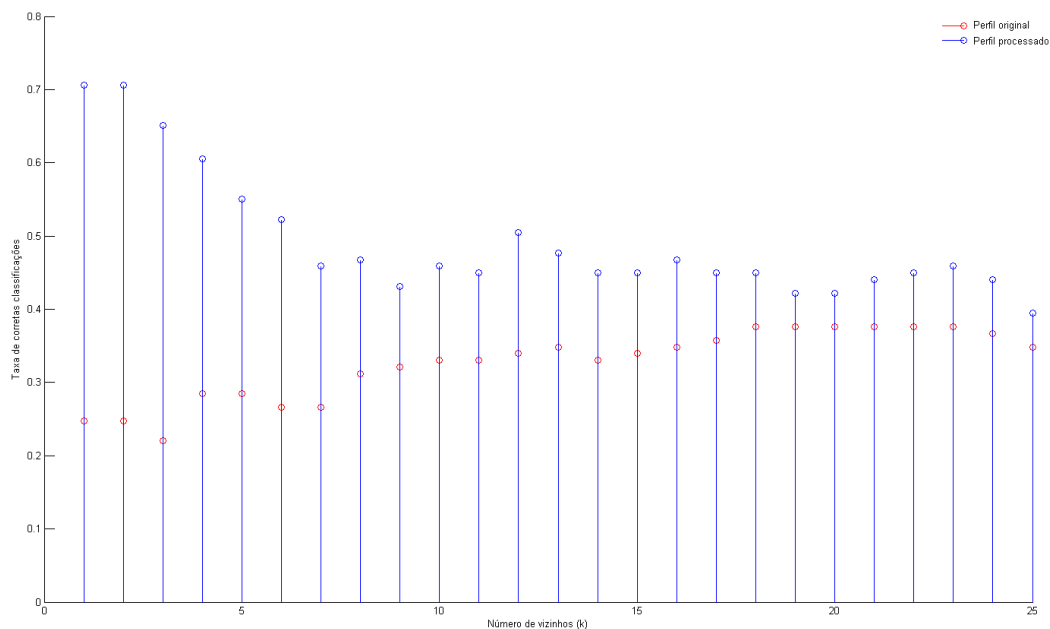


Figura 4.9: Desempenho do KNN utilizando 70% das amostras como treino

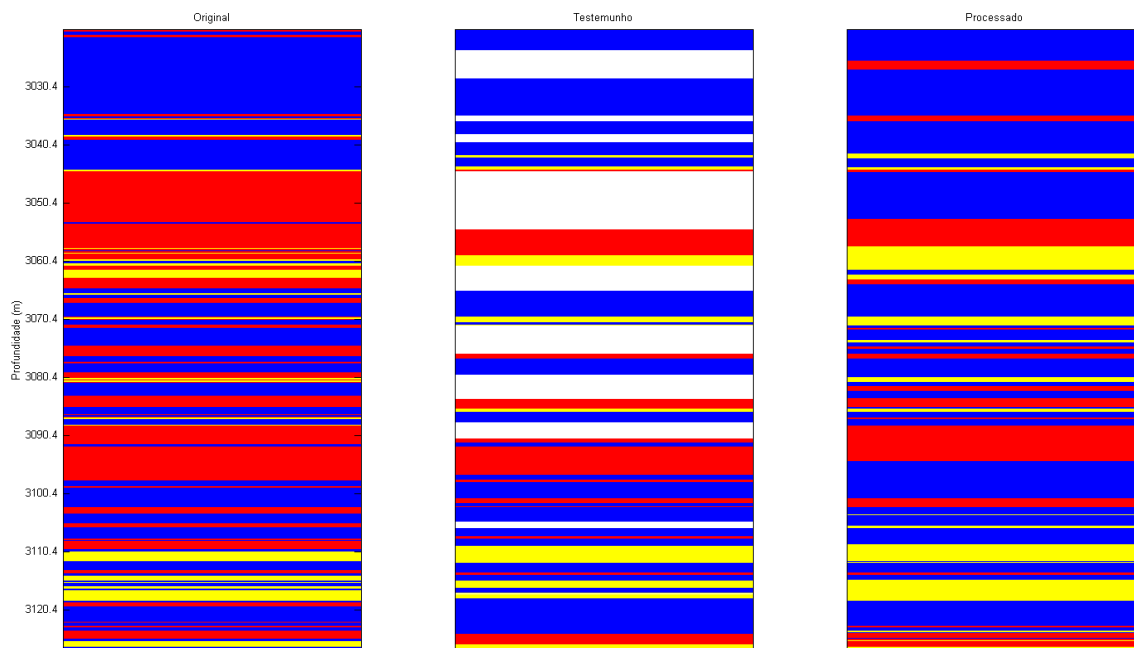


Figura 4.10: Classificação litológica baseada no perfil original (esquerda) e perfil filtrado (direita) comparados com o testemunho (centro) (70% das amostras como treino)

Uma análise mais detalhada do resultado da filtragem é realizada através da Figura 4.11. Nitidamente há uma atenuação mais acentuada na banda de maior frequência (nível 1), como esperado, porém preservando alguns eventos importantes, o que seria eliminado por processos de filtragem baseados, por exemplo, na transformada de Fourier.

Ao analisar os níveis de maior frequência do escalograma de coerência obtido, juntamente com perfis geofísicos e testemunho, evidencia-se que os eventos registrados e mantidos são referentes à intercalações de diferentes litologias, em muitos casos, entre camadas de arenitos e folhelhos bem marcados no perfil raios gama (seta amarela). Outros registros preservados para o nível de maior frequência se referem a picos isolados que são relacionados à cimentação carbonática, marcada pela associação dos perfis densidade e raios gama (seta vermelha).

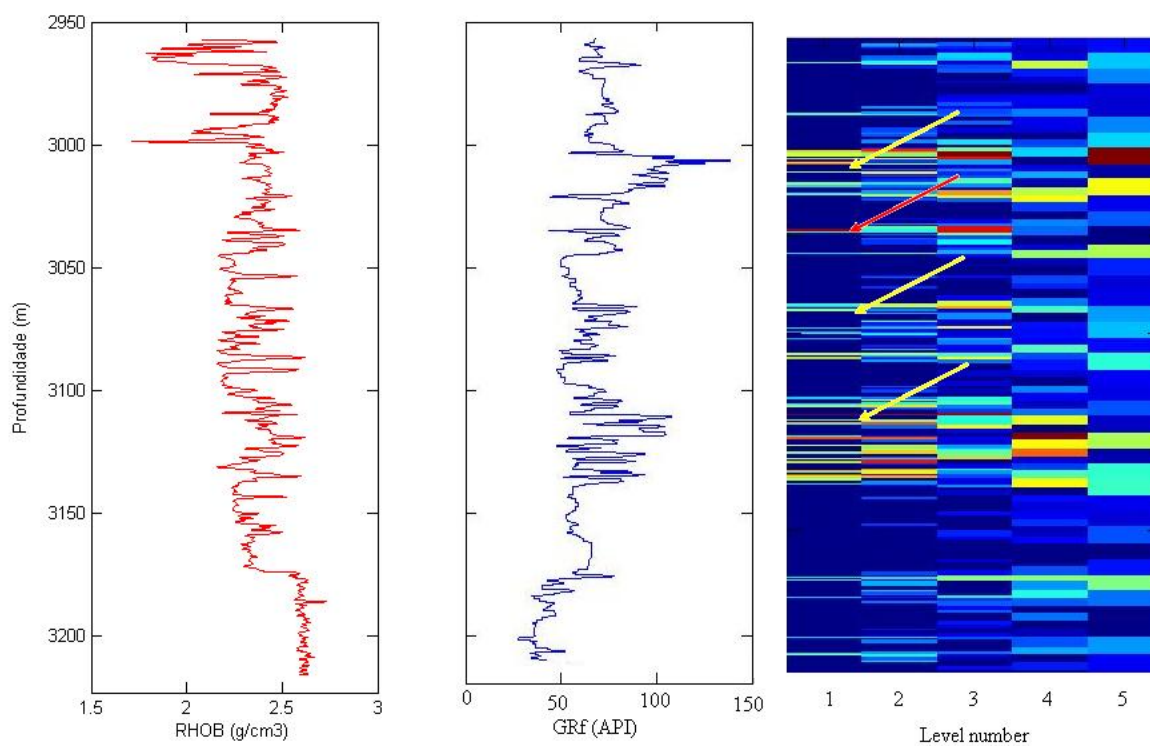


Figura 4.11: Perfis geofísicos e escalograma do perfil GR filtrado (GRf). As setas amarelas e vermelhas indicam regiões de intercalações de arenito/folhelho e cimentação carbonática, respectivamente.

Além disso, regiões que apresentam litologia contínua e homogênea podem ser evidenciadas por essa abordagem, pois nesse caso não se espera variações no sinal. Isso pode ser observado na região indicada na Figura 4.12, na qual a filtragem retirou corretamente valores de coerência referentes ao ruído, fato este evidenciado por ser este trecho descrito em testemunho, que apontou a presença de um intervalo composto unicamente por arenito médio gradado (seta vermelha).

Vale ressaltar que não apenas a fácies arenito médio gradado, mas também a fácies arenito grosso amalgamado, as duas principais fácies reservatório, apresentam-se no campo como pacotes espessos e homogêneos, registrado pelo padrão de forma de caixa nos perfis. Como apontado anteriormente, a correta filtragem relacionada a essas litologias será de grande importância quando utilizado um método matemático de classificação litológica a partir de perfis.

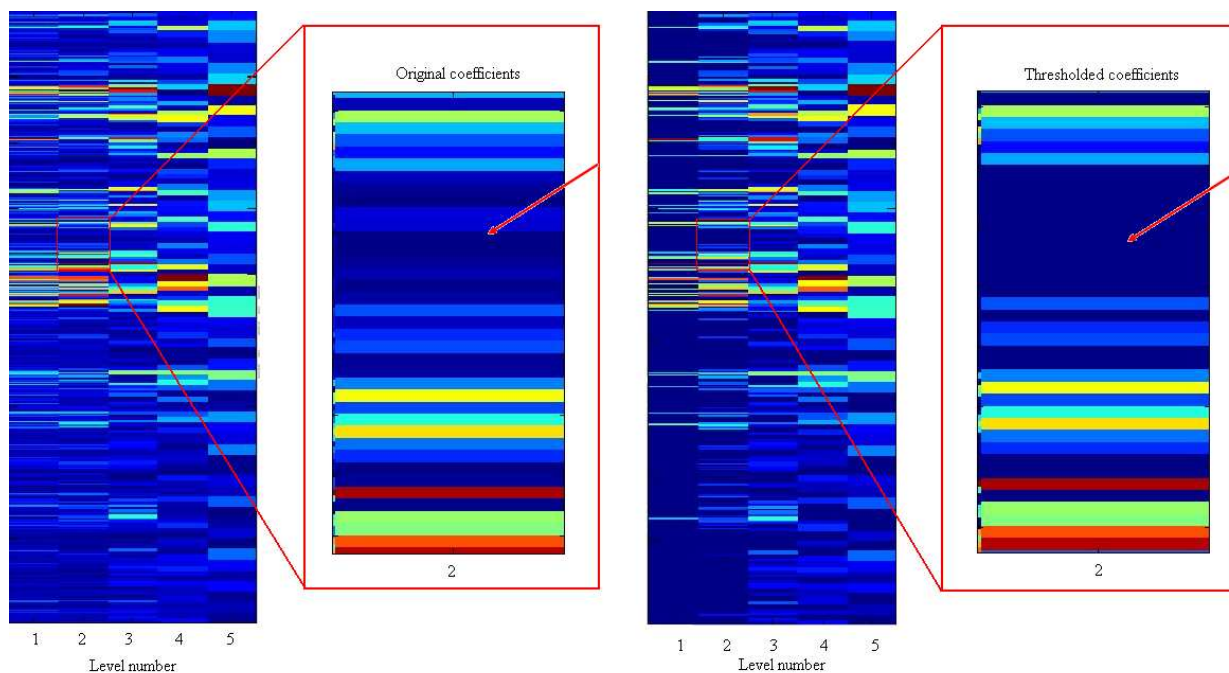


Figura 4.12: Arenito médio gradado destacado após eliminação do ruído

Os resultados apresentados no escalograma de coerência para sequência com variações de fácies como também para litologia contínuas validam esse método como uma ferramenta auxiliar na caracterização geológica. Além disso, foi possível associar o resultado obtido ao estudo de

ciclicidade dos eventos sedimentares quando observado os níveis de coerência e as variações no sinal.

Um exemplo desta situação pode ser ilustrado para os dados do Campo de Namorado, em que Sousa Jr. (1997), através do estudo geoestatístico, aplica curvas de proporções verticais de arenito e folhelho/marga para todo o reservatório. O autor identifica que a evolução vertical das proporções mostra três ciclos de maior sedimentação de arenito, que considera como unidades genéticas e definidas nos poços com auxílio do perfil raios gama. O autor reconhece também três sistemas de deposição: 1) sistemas de canais composto em sua maioria por depósitos turbidíticos ligados à correntes de alta densidade; 2) sistema de canal dique-marginal; 3) sistema argilo-margoso que representa a sedimentação hemipelágica.

A interpretação realizada por Sousa Jr. (1997) para o poço NA07 pode ser comparada com os resultados obtidos no escalograma (Figura 4.13). Nesta comparação foi possível observar a principal associação de fácies reservatório e depósitos de canal que ficam bem definidas nas regiões filtradas, como também os depósitos de dique-marginal e hemipelágicos com padrão serrilhado no perfil e registrado no menor nível do escalograma. A ciclicidade observada na relação entre arenitos e folhelhos/margas é registrada pelo padrão encontrado no nível 1 do escalograma: a região filtrada indica o início do processo deposicional (fácies reservatório), enquanto os “pontos” mantidos são referentes ao fim do ciclo, demarcando as zonas de deposição de folhelhos/margas.

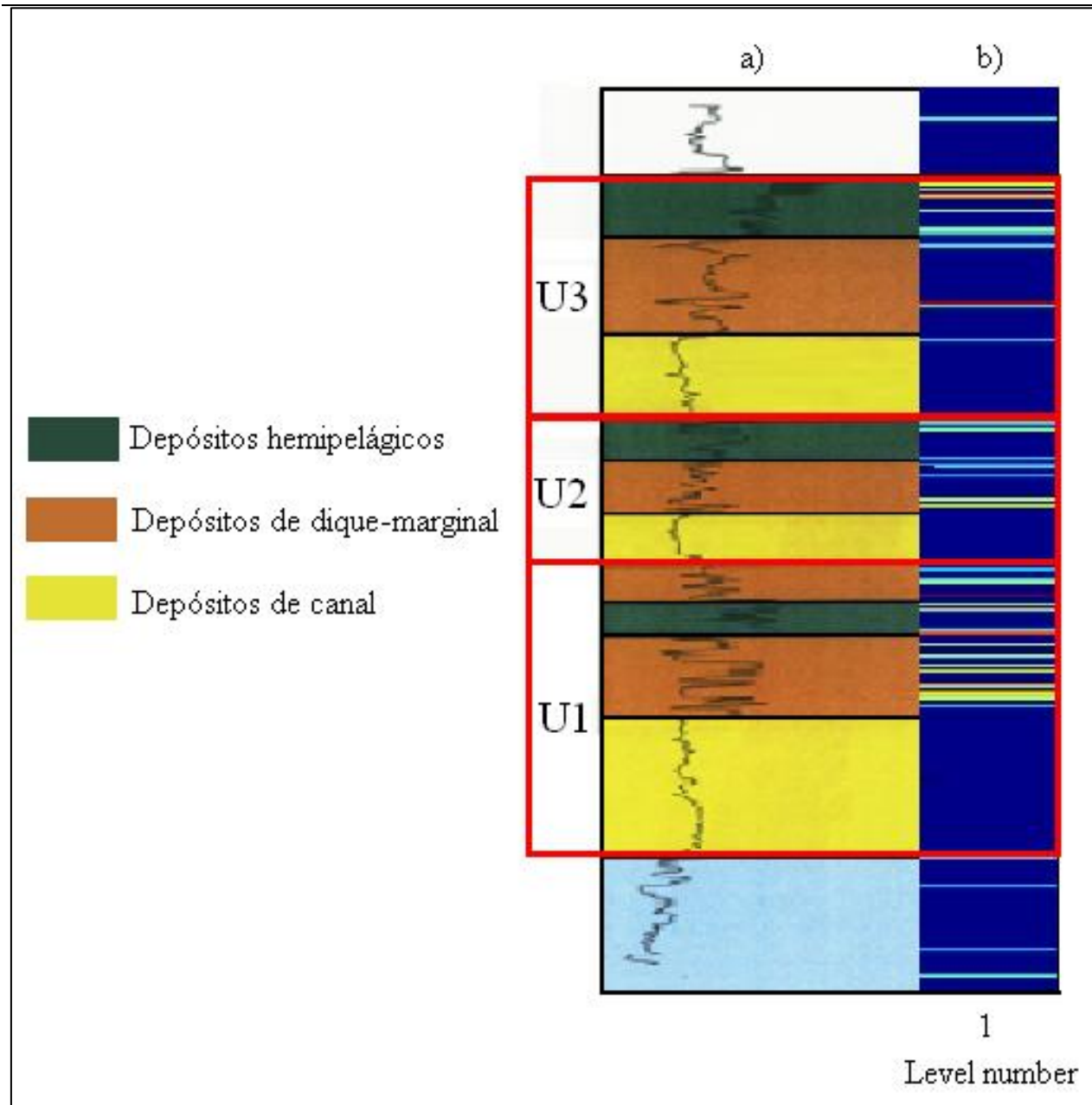


Figura 4.13: a) Ciclos sedimentares identificados por Souza Jr. (1997) e relacionados com o nível 1 do escalograma filtrado, em b).

Os resultados observados demonstram a potencialidade da aplicação do método para extrair mais informações dos dados de poço e, com isso, uma ferramenta útil no trabalho de caracterização geológica de reservatórios de petróleo.

CAPÍTULO 5

CONCLUSÕES

Foi desenvolvido um método para a análise de perfis geofísicos segundo a óptica da transformada *wavelet*. Um enfoque especial foi dado ao problema de atenuação de ruído, visando, com isso, a obtenção de um dado geologicamente mais representativo.

Para o processo de filtragem, foi avaliada a influência da seleção da função base e das regras e funções de *thresholding*. O critério de classificação litológica como método de seleção dos melhores parâmetros foi corroborado a partir da minimização da entropia de Shannon para a seleção da melhor função base. Foi estudado o desempenho das funções de *thresholding* tradicionalmente empregados (*hard* e *soft threshold*), bem como de métodos recém propostos (*customized threshold*). Nesse caso, foi constatado que em sinais sintéticos esta última abordagem proporciona melhores resultados, motivo pelo qual essa função foi adaptada para a análise de perfis geofísicos reais, se mostrando também superior às demais abordagens.

A limitação dessa abordagem é decorrente da necessidade de um critério para a escolha do parâmetro alfa. No caso da modificação proposta para a análise de perfis geofísicos, deve-se haver um conjunto de treinamento e classificação a fim de obter tal parâmetro.

Uma análise conjunta do resultado da filtragem com outros perfis geofísicos permitiu identificar, no próprio escalograma, regiões de intercalações entre camadas de arenitos e folhelhos além de pontos de cimentação carbonática. Quando comparada à caracterização realizada por Sousa Jr. (1997), o método proposto evidenciou as principais associações de fácies reservatório e depósitos de canal que ficaram bem definidas nas regiões filtradas, bem como os depósitos de dique-marginal e hemipelágicos com padrão serrilhado no perfil e registrado nos menores níveis do escalograma. Esta associação na relação de arenitos e folhelhos/margas permitiu relacionar o padrão encontrado no nível 1 do escalograma (zonas de alta frequência seguidas por uma região filtrada) com os três ciclos deposicionais identificados, por Sousa Jr, para o campo de Namorado.

Além de gerar um dado geologicamente mais representativo e obter os parâmetros mais adequados para analisar um determinado sinal segundo a óptica da transformada *wavelet*, o

método proposto permitiu extrair importantes informações geológicas a partir do escalograma obtido.

CAPÍTULO 6

REFERÊNCIAS

- Abry, P. (1997). Ondelettes et Turbulence. Multirésolutions Algorithmes de Décomposition, Invariance d'échelles. *Diderot Editeur, Paris*.
- Chakraborty, A., & Okaya, D. (1995). Frequency-Time Decomposition of Seismic Data Using Wavelet-Based Methods. *Geophysics, Vol. 60, No. 6, 1906-1916* .
- Chang, S. G., Yu, B., & Vetterli, M. (2000). Adaptive Wavelet Thresholding for Image Denoising and Compression. *IEEE Transactions on Image Processing., Vol. 9, 1532-1546, .*
- Coifman, R., & Wickerhauser, M. (1992). Entropy-Based Algorithms for Best Basis Selection. *IEEE Transactions on Information Theory, Vol. 38, No. 2* .
- Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory. Vol. 13, No.1, 21–27*.
- Daubechies, I. (1988). Orthonormal Bases of Compactly Supported Wavelets. *Communications on Pure Applied Mathematics, Vol. 41, No. 7, 909-996* .
- Daubechies, I. (1992). Ten Lectures on Wavelets. *CBMS-NSF Conference Series in Applied Mathematics. SIAM Ed.*
- Deighan, A. J., & Watts, D. R. (1997). Ground-roll Suppression Using the Wavelet Transform. *Geophysics, Vol. 62, No. 6., 1896-1903* .
- Demirel, H., & Anbarjafari, G. (2011). Discrete Wavelet Transform-Based Satellite Image Resolution Enhancement. *IEEE Transactions on Geoscience and Remote Sensing, Vol. 49, No. 6* .
- Donoho, D. (1995). De-Noising by Soft-thresholding. *IEEE - Transactions on Information Theory. Vol. 41, No. 3, 613–627* .
- Donoho, D., & Johnstone, I. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. *Biometrika, Vol.81, No.3, 425–455* .
- Einsele, G., Ricken, W., & Seilacher, A. (1991). Cycles and events in stratigraphy - basic concepts and terms. *Berlin: Springer-Verlag. 1-19*.

- Farge, M., Hunt, J. C., & Vassilicos, J. C. (1993). *Wavelets, Fractals and Fourier Transforms: New Developments and New Applications*. Oxford University Press .
- Fischer, A., de Boer, P., & Premoli Silva, I. (1990). Cyclostratigraphy in Cretaceous. Resources, Events and Rhythms. Background and Plans for Research. *Ginsburg, R.N. & Beaudoin, B., NATO Advanced Science Institutes Series C: Mathematical and Physical Sciences, Kluwer Academic, Dordrecht. 304, 139-172* .
- Goel, P., & Vidakovic, B. (1995). Wavelet Transformations as Diversity Enhancers. *Discussion Paper 95-04,ISDS, Duke University* .
- Goring, D. (2004). Denoising Long-Wave Records.
- Goupillaud, P., Grossmann, A., & Morlet, J. (1984). Cycle-octaves and Related Transforms in Seismic Signal Analysis. *Geoexploration, Vol. 23, 85–102* .
- Graps, A. (1995). An Introduction to Wavelets. *IEEE Computational Science and Engineering* , 2.
- Grossmann, A., & Morlet, J. (1984). Decomposition of Hardy functions into square integrable wavelets of constant shape. *SIAM J. Math. Anal.* 15 , 723–736.
- Guan, L., & Du, Y. (2004). Review of the Application of Wavelet Analysis in the Petroleum Industry. *5th Canadian International Petroleum Conference (55th Annual Technical Meeting)* . Calgary, Alberta, Canada.
- Haar, A. (1910). Zur Theorie der Orthogonalen Funktionensysteme. *Math. Ann.* , 69, 331-371.
- Katul, G., & Vidakovic, B. (1995). The Partitioning of Attached and Detached Eddy Motion in the Atmospheric Surface Layer Using Lorentz Wavelet Filtering. *Boundary-Layer Meteorology, Vol. 77, No. 2, 153-172*.
- Kumar, P., & Foufoula, E. (1997). Wavelet Analysis for Geophysical Applications. *Reviews of Geophysics, Vol. 35, No. 4, 385-412* .
- Lau, K., & Weng, H. (1995). Climate Signal Detection Using Wavelet Transform: How to Make a Time Series Sing. *Bulletin of American Meteorological Society, Vol.76, No.12* .
- Le Treut, H., & Ghil, M. (1983). Orbital Forcing, Climatic Interactions, and Glaciation Cycles. *Journal of Geophysical Research, Vol. 88, 5167-5190*.

- Lorenz, M. O. (1905). Methods of Measuring the Concentration of Wealth. *American Statistical Association*; Vol. 9, No. 70, 209–219.
- Lyrio, J. C., Tenorio, L., & Li, Y. (2004). Efficient Automatic Denoising of Gravity Gradiometry Data. *Geophysics*, Vol. 69, No. 3, 772-782; .
- Mallat, S. G. (1989). A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. Vol. 2, No. 7 .
- Mallat, S. G. (1999). A Wavelet Tour of Signal Processing. *Academic Press*, second edition, 434-465 .
- Meyer, Y. (2006). A Tribute to Morlet. *Fundamental papers in wavelet theory* .
- Misiti, M., Misiti, Y., Oppenheim, G., & Poggi, J. (1997). Wavelet Toolbox™ 4 - User's Guide.
- Morlet, J., Arens, G., Farge, E., & Giard, D. (1982b). Wave Propagation and Sampling Theory, 2, Sampling Theory and Complex Waves. *Geophysics*, Vol. 47, No. 2, 222–236 .
- Morlet, J., Arens, G., Farge, E., & Giard, D. (1982a). Wave Propagation and Sampling Theory, 1. Complex Signal and Scattering in Multilayered Media. *Geophysics*, Vol. 47, No.2, 203–221 .
- Papoulis, A. (1962). The Fourier Integral and Its Applications. *McGraw-Hill*, New York .
- Philip, K., Brooks, M., & Hill, I. (2002). An Introduction To Geophysical Exploration . *Blackwell Science* .
- Plint, A., Eyles, N., Eyles, C., & Walker, R. (1993). Control of Sea Level Change. in Walker, R.G., and James, N.P., eds., *Facies Models – Response to Sea Level Change*, Geological Society of Canada, Toronto, 15-25 .
- Prokoph, A., & Agterberg, F. (2000). Wavelet Analysis of Well-Logging Data From Oil Source Rock, Egret Member, Offshore Eastern Canada. *Bulletin of American Association of Petroleum Geologists*, Vol. 84, No.10, 1617-1632 .
- Protázio, J. M. (2001). Análise Wavelet Aplicada a Sinais Geofísicos. *Dissertação de Mestrado em Ciências e Engenharia de Petróleo - FEM/IG - Unicamp* .
- Schwarzacher, W. (1975). Sedimentation Models and Quantitative Stratigraphy. *Elsevier*, Amsterdam, 161.

- Souza Jr., O. G. (1997). Stratigraphie Séquentielle et Modélisation Probabiliste des Réservoirs d'un Cône Sous Marin Profond (Champ de Namorado Brésil) - Intégration des Données Géologiques et Géophysiques. *Docteur Thèse. Université Paris 6* .
- Stein, C. M. (1981). Estimation of the Mean of a Multivariate Normal Distribution. *The Annals of Statistics. Vol 9, No.6, 1135–1151* .
- Taner, M. (2001). Seismic Attributes . *CSEG Recorder, 49-56* .
- Torrence, C., & Compo, G. (1998). A Practical Guide to Wavelet Analysis. *Bulletin of the American Meteorological Society, Vol. 79, No. 1, 61-78* .
- Vega, N. R. (2003). Reservoir Characterization Using Wavelet Transforms. *Ph.D. dissertation, Texas A&M University, U.S.A.* .
- Vidakovic, B. a. (1994). Wavelet Shrinkage of Turbulence Signals. *Discussion Paper 94-25, ISDS, Duke University* .
- Weedon, G. (2003). Time-Series Analysis and Cyclostratigraphy. Examining stratigraphic records of environmental cycles. *Cambridge University Press* .
- Yoon, B., & Vaidyanathan, P. P. (2004). Wavelet-Based Denoising by Customized Thresholding. *International Conference on Acoustics, Speech, and Signal Processing - ICASSP* .
- Yu, J., Guo, K., Yuan, X., Fu, W., & Xun, Z. (2010). Wavelet Denoising of Well Logs and its Geological Performance. *Energy Exploration & Exploitation, Vol. 28, No. 2,87-95* .

ANEXO A

Tabelas contendo as porcentagens de corretas classificações utilizando o método proposto:

- 60% das amostras para treinamento do algoritmo KNN e ruído branco Gaussiano:

Tabela A-1: Corretas classificações utilizando 60% das amostras como treino e diferentes esquemas de atenuação de ruído (Ruído branco Gaussiano)

FUNÇÃO	<i>wavelet</i>	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
	regra								
SOFT	Universal	60,35	42,85	42,71	41,39	40,42	41,60	40,56	40,63
	Rigsure	37,64	36,60	38,26	39,10	37,57	36,60	35,90	34,93
	Heursure	38,26	38,33	38,40	35,49	38,13	37,71	37,22	35,07
	Minimaxi	52,92	42,85	39,86	41,11	39,86	40,63	41,67	38,96
	Lorenz	52,36	47,85	51,32	50,28	44,10	47,92	48,47	47,85
HARD	Universal	60,21	39,17	36,18	37,64	39,17	38,33	36,67	37,78
	Rigsure	37,99	36,18	36,81	36,88	38,13	35,49	36,88	36,11
	Heursure	37,29	37,92	35,42	37,29	36,67	37,78	35,63	35,69
	Minimaxi	51,60	37,15	37,99	36,60	37,50	37,64	36,04	35,00
	Lorenz	52,36	47,85	51,32	50,28	44,10	47,92	48,47	47,85
CUSTOM	Universal	62,57	44,72	43,40	43,13	41,94	43,61	42,92	42,15
	Rigsure	39,10	38,54	38,47	38,89	39,10	38,75	38,47	38,13
	Heursure	39,51	38,26	39,38	39,31	39,31	38,68	38,89	38,68
	Minimaxi	54,86	42,78	41,39	41,46	41,53	42,36	41,60	40,56
	Lorenz	52,36	47,85	51,32	50,28	44,10	47,92	48,47	47,85
FUNÇÃO	<i>wavelet</i>	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
	regra								
SOFT	Universal	39,03	41,81	41,94	40,49	41,25	41,04	41,39	40,42
	Rigsure	37,29	37,99	37,01	37,15	37,71	36,74	37,36	37,64
	Heursure	37,22	37,36	37,57	38,40	37,78	34,79	35,97	37,36
	Minimaxi	37,57	39,79	40,56	40,76	40,14	40,69	38,68	39,38
	Lorenz	43,47	50,21	46,60	49,38	51,53	36,94	50,35	50,07
HARD	Universal	35,35	37,64	38,54	37,01	37,29	39,10	38,61	35,56
	Rigsure	35,69	37,57	36,53	36,46	35,69	36,46	36,67	36,39

	Heursure	35,90	38,06	36,94	35,83	37,36	35,21	35,76	37,78
	Minimaxi	35,49	37,57	37,78	38,33	36,94	37,92	36,04	36,32
	Lorenz	43,47	50,21	46,60	49,38	51,53	36,94	50,35	50,07
CUSTOM	Universal	40,90	42,92	44,93	43,26	43,54	44,17	43,40	41,74
	Rigsure	38,96	38,40	38,75	38,68	38,68	38,26	38,75	38,82
	Heursure	38,75	39,03	38,47	38,89	37,92	37,99	39,03	38,96
	Minimaxi	39,51	42,08	42,50	40,90	41,39	42,50	41,18	39,58
	Lorenz	43,47	50,21	46,60	49,38	51,53	36,94	50,35	50,07
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	41,04	42,50	40,69	40,90	40,76	41,81	42,99	42,36
	Rigsure	37,43	37,01	37,57	36,39	36,88	37,01	38,61	38,26
	Heursure	37,15	37,57	37,29	38,33	37,29	37,64	36,04	37,43
	Minimaxi	39,38	41,32	40,97	41,53	41,67	40,97	41,74	39,86
	Lorenz	48,75	45,21	50,35	44,79	50,76	48,33	49,79	45,76
HARD	Universal	36,81	38,33	37,43	37,78	36,81	38,06	38,26	35,90
	Rigsure	36,46	36,39	37,43	36,81	36,67	35,90	35,90	36,18
	Heursure	35,42	38,40	37,57	37,78	37,99	36,53	36,32	36,32
	Minimaxi	38,75	36,67	36,88	37,22	37,15	36,94	35,69	37,15
	Lorenz	48,75	45,21	50,35	44,79	50,76	48,33	49,79	45,76
CUSTOM	Universal	42,36	42,43	42,36	43,82	43,13	44,38	44,10	43,26
	Rigsure	38,68	38,47	38,68	39,44	38,75	38,40	38,96	38,82
	Heursure	38,89	38,06	38,13	39,65	38,47	39,10	39,10	39,24
	Minimaxi	42,08	41,81	40,42	41,11	40,90	41,04	42,99	41,88
	Lorenz	48,75	45,21	50,35	44,79	50,76	48,33	49,79	45,76

- 60% das amostras para treinamento do algoritmo KNN e *scaled white noise*:

Tabela A-2: Corretas classificações utilizando 60% das amostras como treino e diferentes esquemas de atenuação de ruído (*scaled white noise*)

FUNÇÃO	<i>wavelet</i> regra	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
SOFT	Universal	60,69	42,64	40,83	42,01	36,67	40,83	38,40	39,44
	Rigsure	40,83	36,81	35,63	36,81	37,43	37,43	35,90	36,74
	Heursure	41,88	37,57	37,92	36,04	36,74	37,85	35,56	36,18
	Minimaxi	57,08	40,35	38,75	37,92	40,21	37,71	38,33	37,99
	Lorenz	55,56	47,64	46,88	50,35	45,42	41,60	42,22	44,86

HARD	Universal	60,28	38,47	37,08	37,57	37,78	36,81	37,43	37,29
	Rigsure	38,96	36,39	36,32	37,22	35,83	36,60	37,57	37,22
	Heursure	41,32	36,60	36,81	35,83	36,60	35,49	37,01	36,39
	Minimaxi	57,99	36,60	37,08	38,61	36,67	37,22	37,50	36,88
	Lorenz	55,56	47,64	46,88	50,35	45,42	41,60	42,22	44,86
CUSTOM	Universal	64,17	44,72	42,08	41,53	40,07	41,32	40,28	39,72
	Rigsure	43,06	38,54	38,61	38,82	38,96	38,40	38,68	37,99
	Heursure	43,68	38,47	38,68	39,03	38,33	38,13	38,61	38,40
	Minimaxi	62,36	41,74	40,14	40,28	40,00	39,93	39,31	38,96
	Lorenz	55,56	47,64	46,88	50,35	45,42	41,60	42,22	44,86
FUNÇÃO	<i>wavelet</i> regra	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
SOFT	Universal	37,36	39,93	43,89	39,58	39,31	41,32	39,65	37,78
	Rigsure	37,78	37,01	37,50	36,18	36,60	37,29	37,01	37,43
	Heursure	37,15	36,53	37,22	37,78	37,15	35,90	36,81	37,64
	Minimaxi	37,57	37,85	40,69	38,47	38,89	40,35	38,47	36,53
	Lorenz	43,96	42,71	48,13	46,53	46,94	40,76	45,07	42,36
HARD	Universal	36,81	37,78	38,82	37,29	38,13	37,36	37,78	36,25
	Rigsure	36,04	36,04	36,04	36,18	36,67	36,74	35,56	36,39
	Heursure	38,47	36,81	35,76	35,69	37,08	36,60	36,60	37,15
	Minimaxi	36,67	37,71	36,04	37,22	37,01	36,53	36,60	37,22
	Lorenz	43,96	42,71	48,13	46,53	46,94	40,76	45,07	42,36
CUSTOM	Universal	38,54	40,63	44,51	42,36	40,63	42,92	41,18	39,86
	Rigsure	38,26	38,40	38,40	38,13	38,68	38,54	38,96	38,47
	Heursure	38,68	38,33	38,96	37,99	38,75	39,31	38,19	39,03
	Minimaxi	38,82	40,42	42,50	40,21	39,79	40,63	40,21	39,03
	Lorenz	43,96	42,71	48,13	46,53	46,94	40,76	45,07	42,36
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	39,79	41,46	38,40	40,35	40,00	39,31	41,94	38,26
	Rigsure	35,56	36,25	36,04	36,46	36,18	36,11	35,83	36,11
	Heursure	35,69	36,88	36,32	37,92	35,14	36,74	36,18	36,39
	Minimaxi	37,50	38,19	38,75	39,79	38,68	38,96	38,40	38,54
	Lorenz	44,17	44,24	45,28	45,21	47,01	45,49	46,18	46,74
HARD	Universal	37,15	35,63	36,67	38,33	37,36	37,71	37,99	36,74
	Rigsure	36,25	36,94	38,13	35,35	37,43	36,88	37,01	35,97
	Heursure	35,97	36,81	37,57	35,56	36,67	36,67	36,81	35,63
	Minimaxi	35,83	36,46	35,28	38,19	35,56	36,46	36,88	36,81

	Lorenz	44,17	44,24	45,28	45,21	47,01	45,49	46,18	46,74
CUSTOM	Universal	41,60	41,11	40,49	42,50	41,74	40,56	41,39	40,97
	Rigsure	38,89	38,40	38,96	38,54	38,54	37,85	38,89	38,47
	Heursure	38,33	38,75	38,75	38,61	38,82	38,33	38,61	38,26
	Minimaxi	40,63	39,72	40,00	41,18	40,21	39,93	39,72	39,17
	Lorenz	44,17	44,24	45,28	45,21	47,01	45,49	46,18	46,74

- 60% das amostras para treinamento do algoritmo KNN e *nonwhite noise*:

Tabela A-3: Corretas classificações utilizando 60% das amostras como treino e diferentes esquemas de atenuação de ruído (*nonwhite noise*)

FUNÇÃO	<i>wavelet</i> regra	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
	SOFT	Universal	51,74	44,72	42,57	46,11	47,29	40,35	42,57
Rigsure		43,54	38,54	41,67	43,89	43,26	41,46	38,96	40,35
Heursure		46,25	40,42	38,47	41,67	40,42	38,13	38,75	39,58
Minimaxi		52,29	45,83	44,24	42,15	43,54	41,81	37,64	43,13
Lorenz		54,51	42,08	51,46	48,75	49,31	48,96	46,39	49,86
HARD	Universal	50,00	43,26	42,36	41,18	40,35	42,43	38,40	43,26
	Rigsure	43,47	36,60	33,19	38,33	35,90	36,39	35,63	34,58
	Heursure	41,88	37,85	35,00	34,79	38,06	36,39	35,14	37,36
	Minimaxi	54,17	43,19	41,39	40,21	42,85	39,24	39,31	37,43
	Lorenz	54,51	42,08	51,46	48,75	49,31	48,96	46,39	49,86
CUSTOM	Universal	54,44	48,96	43,47	47,85	50,42	44,03	42,64	44,03
	Rigsure	46,25	40,56	40,76	45,35	44,58	43,19	41,53	42,15
	Heursure	47,43	41,25	39,44	43,96	42,36	40,35	39,03	41,60
	Minimaxi	58,26	47,85	44,10	46,11	44,44	43,61	41,04	45,35
	Lorenz	54,51	42,08	51,46	48,75	49,31	48,96	46,39	49,86
FUNÇÃO	<i>wavelet</i> regra	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
SOFT	Universal	38,54	43,06	44,31	41,74	49,31	36,94	48,68	43,47
	Rigsure	40,63	43,26	39,93	41,11	41,39	41,81	42,57	43,33
	Heursure	34,44	40,63	41,04	38,26	41,39	38,82	36,74	43,26
	Minimaxi	40,90	41,60	45,42	42,15	46,04	41,32	47,29	39,31
	Lorenz	43,75	47,08	45,35	49,51	51,67	38,33	51,39	49,58
HARD	Universal	38,96	42,71	41,81	43,06	42,64	40,28	40,14	43,26
	Rigsure	36,60	40,42	36,46	34,79	38,68	34,58	37,36	37,43
	Heursure	37,50	46,11	36,32	36,46	35,35	34,79	37,92	37,71

	Minimaxi	37,78	39,79	45,28	42,29	42,15	39,51	41,53	39,93
	Lorenz	43,75	47,08	45,35	49,51	51,67	38,33	51,39	49,58
CUSTOM	Universal	43,68	45,49	49,10	43,75	50,49	42,50	49,38	45,14
	Rigsure	43,40	44,58	40,35	40,76	43,75	43,89	43,26	44,72
	Heursure	37,99	48,47	41,74	39,44	44,58	41,11	40,28	44,10
	Minimaxi	42,71	46,46	47,57	43,82	48,19	44,31	46,60	43,47
	Lorenz	43,75	47,08	45,35	49,51	51,67	38,33	51,39	49,58
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	45,49	38,61	45,69	45,07	49,31	42,01	43,89	45,00
	Rigsure	42,50	39,44	42,78	40,21	41,25	40,83	45,42	43,82
	Heursure	41,88	37,99	38,75	39,31	37,92	41,18	43,40	40,69
	Minimaxi	41,81	40,97	43,47	42,57	45,14	42,15	42,43	42,22
	Lorenz	47,92	45,76	46,67	44,93	52,22	47,71	50,56	47,78
HARD	Universal	43,54	39,03	40,76	43,68	45,07	39,65	45,28	39,44
	Rigsure	35,14	34,10	36,18	38,68	34,58	35,83	39,44	37,15
	Heursure	38,26	36,04	39,51	34,65	34,65	36,04	37,22	33,89
	Minimaxi	40,63	38,96	39,65	35,76	43,40	40,76	42,36	41,11
	Lorenz	47,92	45,76	46,67	44,93	52,22	47,71	50,56	47,78
CUSTOM	Universal	46,18	44,93	47,29	48,47	51,67	45,49	47,64	46,04
	Rigsure	44,93	40,90	44,17	42,57	43,33	41,32	46,94	44,10
	Heursure	43,68	41,88	39,93	41,81	38,75	41,74	45,56	41,60
	Minimaxi	47,71	43,26	46,67	45,42	48,26	44,51	45,97	43,96
	Lorenz	47,92	45,76	46,67	44,93	52,22	47,71	50,56	47,78

- 70% das amostras para treinamento do algoritmo KNN e ruído branco Gaussiano:

Tabela A-4: Corretas classificações utilizando 70% das amostras como treino e diferentes esquemas de atenuação de ruído (Ruído branco Gaussiano)

FUNÇÃO	<i>wavelet</i> regra	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
SOFT	Universal	62,94	43,76	42,66	42,84	41,74	43,12	41,74	41,10
	Rigsure	39,91	37,89	36,97	37,98	38,07	37,98	34,59	36,15
	Heursure	38,99	37,98	37,06	36,79	39,08	36,51	35,69	37,43
	Minimaxi	55,32	40,55	38,90	41,47	40,92	40,37	40,46	39,36
	Lorenz	53,58	45,78	52,75	51,10	46,15	50,09	49,91	49,91
HARD	Universal	64,68	37,06	37,89	36,79	38,90	38,72	38,81	36,51
	Rigsure	38,53	36,79	35,69	38,07	39,08	37,06	36,24	36,70

	Heursure	37,80	36,88	37,25	37,52	38,53	36,79	37,43	36,88
	Minimaxi	54,50	39,45	37,06	37,34	36,88	37,16	36,42	37,25
	Lorenz	53,58	45,78	52,75	51,10	46,15	50,09	49,91	49,91
CUSTOM	Universal	65,14	46,33	44,31	44,86	43,12	43,49	43,85	42,29
	Rigsure	40,09	39,17	39,82	40,00	40,55	39,82	39,08	38,81
	Heursure	40,00	38,99	40,37	40,37	40,55	39,27	39,36	39,72
	Minimaxi	58,44	42,75	41,93	41,56	42,20	41,47	43,49	41,28
	Lorenz	53,58	45,78	52,75	51,10	46,15	50,09	49,91	49,91
FUNÇÃO	<i>wavelet</i> regra	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
SOFT	Universal	39,17	42,02	43,03	44,86	41,10	42,39	40,73	41,74
	Rigsure	36,70	35,69	36,42	38,26	37,52	36,24	37,52	36,15
	Heursure	36,97	36,70	37,06	39,08	38,07	36,15	39,45	37,16
	Minimaxi	38,72	41,28	42,02	40,55	39,91	42,29	39,82	41,93
	Lorenz	44,86	50,64	47,52	50,83	53,21	35,87	49,82	49,82
HARD	Universal	35,69	38,26	37,16	37,61	35,78	38,72	37,34	36,15
	Rigsure	38,99	38,62	36,51	37,71	36,51	35,96	37,80	38,72
	Heursure	37,06	37,06	37,25	40,28	36,61	36,70	35,96	36,06
	Minimaxi	36,15	38,90	36,97	38,62	36,79	37,61	35,78	35,60
	Lorenz	44,86	50,64	47,52	50,83	53,21	35,87	49,82	49,82
CUSTOM	Universal	40,92	43,12	45,32	43,39	44,13	45,23	43,67	42,48
	Rigsure	39,36	39,72	40,00	39,36	39,63	39,08	40,00	40,73
	Heursure	39,91	39,72	40,73	40,28	39,54	38,26	40,09	39,63
	Minimaxi	40,00	43,03	43,39	41,65	41,56	43,49	42,94	41,19
	Lorenz	44,86	50,64	47,52	50,83	53,21	35,87	49,82	49,82
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	43,30	41,74	42,39	41,83	42,02	43,58	42,20	43,58
	Rigsure	37,89	37,89	37,34	37,16	37,61	36,33	36,06	38,62
	Heursure	37,80	36,70	38,53	38,62	38,17	37,61	38,17	37,98
	Minimaxi	40,92	38,90	41,10	39,54	39,17	40,73	42,48	40,55
	Lorenz	48,90	45,05	50,28	44,04	51,56	49,63	52,39	46,24
HARD	Universal	37,43	38,17	36,97	37,89	36,51	36,42	37,80	37,25
	Rigsure	36,42	37,80	35,69	35,96	38,07	36,70	34,31	37,71
	Heursure	35,23	37,43	34,22	35,87	36,51	36,97	36,88	36,15
	Minimaxi	38,72	37,80	36,06	36,79	37,52	37,34	37,43	37,25
	Lorenz	48,90	45,05	50,28	44,04	51,56	49,63	52,39	46,24
CUSTOM	Universal	43,21	44,86	42,75	44,22	43,49	45,41	44,50	43,49

	Rigsure	41,10	38,81	39,72	41,65	39,45	39,36	39,54	39,82
	Heursure	39,91	38,99	40,46	39,63	39,36	40,55	40,83	41,10
	Minimaxi	42,84	42,39	41,74	42,57	43,03	41,65	44,40	43,03
	Lorenz	48,90	45,05	50,28	44,04	51,56	49,63	52,39	46,24

- 70% das amostras para treinamento do algoritmo KNN e *scaled white noise*:

Tabela A-5: Corretas classificações utilizando 70% das amostras como treino e diferentes esquemas de atenuação de ruído (*scaled white noise*)

FUNÇÃO	<i>wavelet</i>	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
	regra								
SOFT	Universal	64,59	42,75	39,17	41,74	38,81	40,92	39,82	37,71
	Rigsure	44,86	37,61	37,16	37,98	38,35	37,98	35,87	35,96
	Heursure	41,10	35,32	36,24	37,34	38,44	35,87	38,17	38,44
	Minimaxi	61,65	39,45	38,53	38,26	37,16	37,61	38,44	38,35
	Lorenz	56,51	47,16	43,49	54,86	44,22	42,57	42,84	45,41
HARD	Universal	62,75	39,36	37,80	40,00	39,08	35,69	37,43	40,09
	Rigsure	43,03	35,96	37,52	37,25	35,69	35,87	37,43	36,42
	Heursure	41,38	35,87	36,51	35,32	37,89	35,78	35,60	36,06
	Minimaxi	58,17	37,25	38,26	38,72	39,72	35,69	36,33	36,97
	Lorenz	56,51	47,16	43,49	54,86	44,22	42,57	42,84	45,41
CUSTOM	Universal	67,52	44,86	44,22	42,57	42,66	41,65	41,93	40,00
	Rigsure	45,69	39,72	38,72	39,63	39,45	38,90	38,99	39,17
	Heursure	46,24	40,46	40,09	40,28	40,73	39,54	38,81	39,54
	Minimaxi	64,86	42,75	40,92	41,19	41,47	40,46	40,73	39,72
	Lorenz	56,51	47,16	43,49	54,86	44,22	42,57	42,84	45,41
FUNÇÃO	<i>wavelet</i>	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
regra									
SOFT	Universal	36,51	40,18	45,05	42,94	39,45	42,75	40,18	37,71
	Rigsure	35,96	37,89	36,97	36,70	36,33	35,50	37,98	34,86
	Heursure	37,25	37,43	36,79	37,98	36,97	36,70	38,62	37,80
	Minimaxi	38,44	37,25	41,10	39,91	38,81	40,55	39,36	37,16
	Lorenz	43,76	42,20	46,88	47,52	47,52	41,56	48,62	42,57
HARD	Universal	35,41	37,61	38,53	37,06	37,89	37,61	37,43	34,68
	Rigsure	38,26	36,61	36,33	34,22	36,97	34,68	35,87	36,42
	Heursure	36,97	36,06	36,24	35,32	36,88	35,78	36,70	37,52
	Minimaxi	36,97	38,53	38,35	37,80	36,24	35,78	34,86	37,89

	Lorenz	43,76	42,20	46,88	47,52	47,52	41,56	48,62	42,57
CUSTOM	Universal	40,92	42,20	45,14	42,84	41,10	43,21	41,47	40,92
	Rigsure	39,45	39,45	39,82	39,54	38,53	40,37	38,99	40,00
	Heursure	39,82	39,45	39,91	40,28	39,54	39,91	39,63	40,64
	Minimaxi	39,17	40,73	42,84	41,10	40,83	41,38	40,83	40,37
	Lorenz	43,76	42,20	46,88	47,52	47,52	41,56	48,62	42,57
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	40,55	41,38	36,51	41,47	42,48	38,62	40,73	40,00
	Rigsure	36,42	37,16	35,87	36,79	37,16	37,71	36,06	36,88
	Heursure	38,81	36,15	37,43	37,98	35,14	37,25	36,24	36,42
	Minimaxi	37,71	38,17	38,62	41,93	38,53	37,98	37,89	39,82
	Lorenz	43,94	45,41	45,05	45,23	46,51	45,05	47,80	45,69
HARD	Universal	36,88	35,69	38,17	37,52	39,08	36,15	38,35	37,06
	Rigsure	35,96	37,16	36,24	36,79	36,97	37,61	34,95	36,70
	Heursure	36,42	34,86	37,43	37,34	36,42	35,32	37,61	36,70
	Minimaxi	36,15	35,87	37,98	35,32	37,43	37,16	37,98	36,61
	Lorenz	43,94	45,41	45,05	45,23	46,51	45,05	47,80	45,69
CUSTOM	Universal	41,65	42,57	40,73	42,75	41,74	42,66	42,75	41,93
	Rigsure	39,27	39,45	39,17	38,90	39,54	39,72	38,90	39,54
	Heursure	39,82	40,18	40,00	38,90	39,63	38,62	38,53	38,90
	Minimaxi	41,10	41,19	40,73	41,56	41,19	41,01	40,92	40,46
	Lorenz	43,94	45,41	45,05	45,23	46,51	45,05	47,80	45,69

- 70% das amostras para treinamento do algoritmo KNN e *nonwhite noise*:

Tabela A-6: Corretas classificações utilizando 70% das amostras como treino e diferentes esquemas de atenuação de ruído (*nonwhite noise*)

FUNÇÃO	<i>wavelet</i> regra	Db1	Db2	Db3	Db4	Db5	Db6	Db7	Db8
SOFT	Universal	51,93	45,78	43,85	46,88	47,43	37,71	43,76	45,41
	Rigsure	46,15	38,07	41,19	44,59	42,20	43,03	39,82	40,55
	Heursure	47,80	40,09	38,62	39,27	41,38	38,26	38,07	39,54
	Minimaxi	53,94	49,17	41,83	44,68	47,61	42,20	38,35	42,29
	Lorenz	52,84	44,68	55,87	48,44	47,80	50,64	47,98	47,89
HARD	Universal	50,09	45,05	41,28	44,13	40,46	45,23	37,43	41,19
	Rigsure	43,49	37,16	33,67	43,76	38,17	37,71	36,70	35,32
	Heursure	43,67	37,52	37,71	36,70	39,08	36,97	34,95	38,07

	Minimaxi	52,11	44,13	43,58	39,36	42,75	39,63	39,17	37,25
	Lorenz	52,84	44,68	55,87	48,44	47,80	50,64	47,98	47,89
CUSTOM	Universal	55,05	49,72	44,50	49,45	49,45	45,78	44,50	44,86
	Rigsure	46,61	42,29	41,28	46,97	43,85	42,57	41,65	43,12
	Heursure	49,17	42,39	40,83	43,76	43,58	41,74	39,72	42,39
	Minimaxi	56,88	49,08	44,95	48,53	46,51	43,94	41,83	46,79
	Lorenz	52,84	44,68	55,87	48,44	47,80	50,64	47,98	47,89
FUNÇÃO	<i>wavelet</i> regra	Db9	Db10	Sym2	Sym3	Sym4	Sym5	Sym6	Sym7
SOFT	Universal	39,82	42,66	43,85	42,29	49,17	38,62	46,70	43,21
	Rigsure	40,55	41,47	38,99	40,28	39,54	41,19	41,47	44,22
	Heursure	35,69	42,20	40,37	38,17	40,09	37,43	38,81	45,78
	Minimaxi	41,10	42,75	47,34	42,39	46,70	41,19	44,95	41,65
	Lorenz	44,04	48,53	47,34	51,65	52,02	38,99	50,28	51,10
HARD	Universal	38,62	41,38	41,74	43,85	43,21	42,84	42,57	42,57
	Rigsure	35,96	40,00	37,06	33,49	36,88	35,05	38,07	38,07
	Heursure	38,26	46,51	38,17	35,60	35,69	35,05	35,69	37,52
	Minimaxi	35,69	40,09	44,77	41,47	39,17	40,64	40,46	40,73
	Lorenz	44,04	48,53	47,34	51,65	52,02	38,99	50,28	51,10
CUSTOM	Universal	44,95	46,70	48,99	45,05	50,64	42,66	50,18	45,78
	Rigsure	43,03	44,59	40,73	41,83	42,84	43,12	42,66	46,06
	Heursure	39,27	49,82	42,57	40,55	45,32	41,47	40,83	46,06
	Minimaxi	43,49	46,24	49,82	45,50	49,08	45,96	47,71	44,68
	Lorenz	44,04	48,53	47,34	51,65	52,02	38,99	50,28	51,10
FUNÇÃO	<i>wavelet</i> regra	Sym8	Sym9	Sym10	Coif1	Coif2	Coif3	Coif4	Coif5
SOFT	Universal	45,41	43,30	49,27	44,86	50,73	44,86	46,97	46,51
	Rigsure	43,12	39,45	42,94	41,56	40,92	40,55	46,33	43,03
	Heursure	42,39	38,72	40,28	39,72	34,40	42,48	43,76	39,27
	Minimaxi	44,31	44,77	45,05	42,66	47,80	42,11	45,78	42,11
	Lorenz	49,54	46,06	48,35	45,60	54,13	48,90	51,93	47,06
HARD	Universal	42,75	41,10	41,01	45,23	45,23	41,01	47,06	38,72
	Rigsure	34,86	34,22	34,59	40,92	36,88	36,51	37,25	39,17
	Heursure	39,17	37,71	38,72	36,15	35,14	37,71	37,43	32,57
	Minimaxi	40,64	41,01	40,46	35,50	43,03	41,47	41,47	40,46
	Lorenz	49,54	46,06	48,35	45,60	54,13	48,90	51,93	47,06
CUSTOM	Universal	47,25	46,15	48,35	48,44	53,30	47,61	48,99	46,79
	Rigsure	44,31	41,19	44,59	42,75	43,39	41,28	46,97	44,04

Heursure	44,68	40,73	40,55	42,29	38,90	42,29	46,24	42,11
Minimaxi	49,27	44,86	46,24	46,51	50,92	45,14	46,24	45,50
Lorenz	49,54	46,06	48,35	45,60	54,13	48,90	51,93	47,06

APÊNDICE A

A1 - K-Vizinhos Mais Próximos

O método KNN é um algoritmo de classificação baseado na distância estrutural entre um conjunto de testes e conjunto de treino (Cover *et al.*, 1967). Dado um conjunto de testes com diversas amostras, cada amostra será analisada espacialmente e seu rótulo (classificação) será igual ao do maior número de vizinhos contidos na região escolhida. Este método está entre os algoritmos mais simples de aprendizagem de máquinas. Sua utilização é recomendada para casos onde computações probabilísticas ou escolhas iniciais podem complicar a solução. De fato, a única entrada necessária, além dos conjuntos de testes e treinos, é o número de vizinhos k da vizinhança ((Bremner *et al.*, 2005).

Seja $T^n = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ um conjunto de treino com n amostras rotuladas, ou seja, com uma identificação relacionada a cada amostra e \mathbf{x} um ponto que se deseja classificar.

O ponto mais próximo $\mathbf{x}' \in T^n$, matematicamente, é dado por:

$$\mathbf{x}' = \underset{\mathbf{x}_i \in T^n}{\min} d(\mathbf{x}, \mathbf{x}_i), \quad (42)$$

onde $d(\mathbf{x}, \mathbf{x}_i)$ é a distância entre os pontos \mathbf{x} e \mathbf{x}_i .

Se $k = 1$ e $\mathbf{x}' \in T^n$ é o ponto mais próximo de \mathbf{x} , o rótulo atribuído à amostra \mathbf{x} será o mesmo rótulo da amostra \mathbf{x}_i . Analiticamente, se $k = n$ e $X = \{\mathbf{x}', \mathbf{x}'', \dots, \mathbf{x}^{(n)}\}$ são os n pontos mais próximos de \mathbf{x} , o rótulo atribuído ao ponto \mathbf{x} será o rótulo de maior frequência entre os elementos de X (Figura A1).

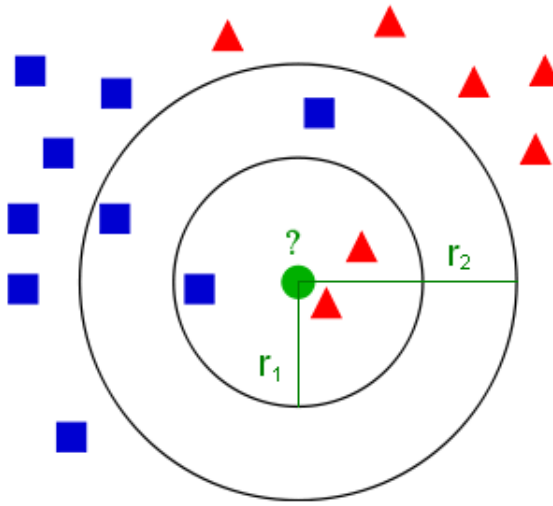


Figura A 1: Elucidação do método KNN

Seja o objetivo classificar a variável representada pelo círculo verde central. Escolhe-se um número de vizinhos, por exemplo, três (aqui elucidado pelo círculo de raio r_1), e assim o algoritmo avalia a representatividade das amostras presentes nesta vizinhança. Uma vez que há duas amostras de triângulos vermelhos e somente uma de quadrado azul, a variável seria classificada como “triângulo vermelho”. Ao considerar cinco vizinhos (denotada pelo círculo de raio r_2), há três quadrados e dois triângulos e, portanto, a variável seria classificada como “quadrado azul”. Este exemplo simples ajuda a entender a idéia central deste método.

Costuma-se escolher o número de vizinhos k como sendo ímpar, evitando possíveis empates entre os números de classes para uma classificação. Na necessidade de um desempate, os pontos de \mathbf{X} que tem maior peso nos rótulos são os mais próximos de \mathbf{x} .

Não existem regras que limitam o número de vizinhos, mas algumas observações devem ser feitas: se k é muito grande, o ponto de teste \mathbf{x} pode ser classificado erroneamente devido ao grande número de outros rótulos que não os corretos. O mesmo ocorre quando se tem poucas amostras $\mathbf{x}_i \in \mathbf{T}^n$ com o rótulo da amostra que se deseja classificar, uma vez que a contribuição das amostras corretas será pequena.

