

Milton Perceus Santos de Melo

**Modelo híbrido inteligente baseado em filtragem *wavelet*  
e redes neurais com aplicações à previsão de séries  
temporais**

Recife

16 de Maio de 2018



**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**

**Modelo híbrido inteligente baseado em filtragem *wavelet* e redes neurais com aplicações à previsão de séries temporais**

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 16/05/2018 pela comissão examinadora.

**Área de concentração: Biometria e Estatística Aplicada**

**Orientador: Prof. Dr. Tiago Alessandro Espínola Ferreira**

**Coorientadores: 1. Prof. Dr. Carlos Tadeu dos Santos Dias; 2. Prof. Dr. Paulo Salgado Gomes de Mattos Neto**

**Recife**

**16 de Maio de 2018**

Dados Internacionais de Catalogação na Publicação (CIP)  
Sistema Integrado de Bibliotecas da UFRPE  
Biblioteca Central, Recife-PE, Brasil

M528m      Melo, Milton Perceus Santos de  
              Modelo híbrido inteligente baseado em filtragem wavelet  
              e redes neurais com aplicações à previsão de séries temporais /  
Milton Perceus Santos de Melo. – 2018.  
              105 f. : il.

              Orientador: Tiago Alessandro Espínola Ferreira.  
              Coorientadora: Carlos Tadeu dos Santos Dias, Paulo  
Salgado Gomes de Mattos Neto.

              Tese (Doutorado) – Universidade Federal Rural de  
Pernambuco, Programa de Pós-Graduação em Biometria e  
Estatística Aplicada, Recife, BR-PE, 2018.

              Inclui referências, anexo(s) e apêndice(s).

              1. Decomposição Wavelet 2. Seleção da Wavelet-mãe  
3. Sistemas híbridos inteligentes 4. Otimização na previsão  
5. Pré-processamento de séries temporais I. Ferreira, Tiago  
Alessandro Espínola, orient. II. Dias, Carlos Tadeu dos Santos,  
coorient. III. Mattos Neto, Paulo Salgado Gomes de, coorient.  
IV. Título

CDD 519.5

**UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO**  
**PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO**  
**PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA E ESTATÍSTICA APLICADA**  
**Modelo híbrido inteligente baseado em filtragem *wavelet* e redes neurais com**  
**aplicações à previsão de séries temporais**

Milton Perceus Santos de Melo

Tese julgada adequada para obtenção do título de Doutor em Biometria e Estatística Aplicada, defendida e aprovada por unanimidade em 16/05/2018 pela comissão examinadora.

Orientador:

---

**Prof. Dr. Tiago Alessandro Espínola  
Ferreira**  
Orientador

Banca examinadora:

---

**Prof. Dr. Antonio Samuel Alves da  
Silva**  
DEINFO/UFRPE

---

**Prof. Dr. Carlos Tadeu dos Santos  
Dias**  
Coorientador - Esalq/USP

---

**Prof. Dr. Moacyr Cunha Filho**  
DEINFO/UFRPE

---

**Prof. Dr. Paulo Salgado Gomes de  
Mattos Neto**  
Coorientador - CIn/UFPE

*Dedico o presente trabalho a Deus, a minha mãe e ao meu filho  
que foram meus maiores apoios nos momentos de angústia.*

# Agradecimentos

À Deus por todas as graças derramadas em minha vida e a Nossa Senhora pela sua intercessão e proteção como mãe.

Aos meus pais, Milton Lima de Melo e Maria Auxiliadora Santos de Melo, pela vida, exemplo de personalidade, trabalho e amor.

Ao meu filho, Thales Henrique de Melo, por está sempre presente na minha vida, apoiando e incentivando esta jornada acadêmica com sua alegria e amor de maneira à renovar minhas energias.

À minha irmã, Nadjamara Santos de Melo que, com muito amor, está sempre acompanhando meus passos.

Aos meus familiares, bem como aos meus amigos, que me apoiaram, direta ou indiretamente, nessa caminhada.

Aos meus orientadores Prof. Dr. Tiago Alessandro Espínola Ferreira, Prof. Dr. Carlos Tadeu dos Santos Dias e Prof. Dr. Paulo Salgado Gomes de Mattos Neto. Vocês são exemplos de competência, dedicação e amor à profissão docente. Agradeço por todo aprendizado, companheirismo e amizade.

Aos amigos do doutorado e mestrado do Programa de Pós-graduação em Biometria e Estatística Aplicada/UFRPE que se fizeram presentes em minha vida.

Ao programa de Pós-graduação em Biometria e Estatística Aplicada (PPGBEA) da Universidade Federal Rural de Pernambuco representado pelo Prof. Dr. Moacyr Cunha Filho, bem como também ao secretário Marco Antônio Gomes dos Santos por toda ajuda.

Aos professores da banca examinadora pela disponibilidade e sugestões para melhoria deste trabalho e a todos os professores do PPGBEA que contribuíram para minha formação.

Agradeço à CAPES (Comissão de Aperfeiçoamento de Pessoal do Nível Superior) pelo apoio financeiro no início de minha formação e, por fim, a Secretaria de Educação do estado de Pernambuco, a Universidade Estadual de Alagoas pelo incentivo durante a fase de conclusão deste trabalho e a todas as pessoas que, apesar de não ter citado os nomes, contribuíram direta ou indiretamente no decorrer deste trabalho.

*“Não vos amoldeis às estruturas deste mundo,  
mas transformai-vos pela renovação da mente,  
a fim de distinguir qual é a vontade de Deus:  
o que é bom, o que lhe é agradável, o que é perfeito.  
(Bíblia Sagrada, Romanos 12, 2)*

# Resumo

Combinações de métodos aplicados à previsão de séries temporais têm mostrado, quanto à acurácia, resultados superiores quando comparados aos modelos individuais. Uma alternativa bastante utilizada de agregação de modelos é a utilização de métodos de pré-processamento em séries temporais em conjunto com preditores individuais e, neste contexto, a análise *Wavelet* tem importantes contribuições. Contudo, a eficiência da análise *Wavelet* enquanto método de filtragem de séries temporais pode ser otimizada uma vez que em sua abordagem tradicional a ideia é apenas retirar um tipo específico de ruído, o ruído branco. Neste contexto, esta Tese apresenta a proposta de um novo método híbrido de previsão de séries temporais, intitulado *Multiresolution time-delay added evolutionary forecasting* (MTAEF), que tem como objetivo otimizar o processo de filtragem para maximizar o desempenho do processo de previsão. O modelo híbrido inteligente *Time-delay added evolutionary forecasting* (TAEF) foi o preditor escolhido em virtude de sua capacidade não linear de ajuste e generalização de informação a partir dos dados, uma vez que combina Redes neurais artificiais com Algoritmos genéticos modificados em um mecanismo de busca evolutiva pela dimensionalidade mínima necessária para a reconstrução do espaço de fase gerador da série. Para averiguar a eficiência do modelo proposto, foram realizados experimentos computacionais em três séries temporais clássicas da literatura especializada. As referidas séries são as seguintes: *Canadian lynx*, *Wolf's sunspot* e *Exchange rate*. O processo de verificação do desempenho e comportamento do modelo foi realizado em 50 repetições independentes por meio da análise da robustez do modelo e do erro de previsão. Os resultados alcançados pelo método proposto, em relação à outras abordagens, apresentaram maior acurácia e menor variabilidade corroborando assim com a abordagem proposta.

**Palavras-chave:** Decomposição *Wavelet*, Seleção da *Wavelet*-mãe, Sistemas híbridos inteligentes, Otimização na previsão, Pré-processamento de séries temporais.

# Abstract

Some combinations of methods applied to time series forecast have resulted in better outcomes concerning their efficiency when compared to individual time series models. A widely used alternative option for model aggregation is the use of denoise methods in combination with individual forecasting methods. Within this context, the Wavelets analysis has brought great contributions. However, the efficiency of Wavelets analysis as a time series denoise method can be enhanced, since the purpose of traditional denoising is only to remove a specific type of noise, the so called white noise. In such a context, this Thesis has the purpose to present a new hybrid time series forecasting model caled Multiresolution Time-delay Added Evolutionary forecast (TAEF) method. This model aims to optimize the denoising method, maximizing the performance of forecasting process. The Time-delay Added Evolutionary forecast model (TAEF) was the forecaster selected due to its nonlinear adjustment capacity and generalization of information from collected data since it combines Artificial Neural Networks with modified genetic algorithms to search for the minimum dimensionality necessary to reconstruct the space-generating phase of the series. To verify the efficiency of the proposed method, computational experiments were carried out with three classical time series from corresponding specialized literature. These series are the following ones: Canadian Lynx, Wolf's Sunspot and Exchange Rate. The verification process of model performance and efficiency was based on 50 (fifty) independent replicates by means of robustness analysis and forecasting error. The results achieved by the proposed method, compared to other approaches, provided higher accuracy and lower variability corroborating the proposed method.

**Key-words:** Wavelet decomposition, Wavelet-mother selection, Intelligent Hibrid Systems, Forecasting optimization, Time series preprocessing.

# Lista de Figuras

Figura 1 – Algoritmo de decomposição de Mallat. . . . .	18
Figura 2 – Algoritmo de reconstrução de Mallat. . . . .	19
Figura 3 – Rede Neural MLP alimentada com múltiplas camadas. . . . .	25
Figura 4 – Arquitetura de um Rede Neural MLP. . . . .	25
Figura 5 – Esquema da previsão da série temporal pelo procedimento. (a) Horizonte de previsão $h = 1$ , (b) horizonte de previsão $h = 2$ e (c) Horizonte de previsão $h = 3$ . . . . .	27
Figura 6 – Esquema de formatação de dados entrada/saída na RNA. (a) Caso <i>one-step-ahead</i> , (b) caso <i>multi-step-ahead</i> . . . . .	28
Figura 7 – Esquema da previsão da série temporal para três procedimentos <i>multi-step-ahead</i> . (a) Horizonte de previsão $h = 1$ , (b) horizonte de previsão $h = 2$ e (c) Horizonte de previsão $h = 3$ . . . . .	29
Figura 8 – Esquema padrão do Algoritmo genético. . . . .	31
Figura 9 – Esquema do Método TAEF. . . . .	34
Figura 10 – Arquitetura do Sistema de Decomposição <i>Wavelet</i> - WavDS. . . . .	37
Figura 11 – Esquematização do método proposto Multiresolution Time-delay Added Evolutionary Forecasting - MTAEF. . . . .	39
Figura 12 – Gráfico do conjunto de dados <i>Sunspot</i> . . . . .	40
Figura 13 – Gráfico do conjunto de dados <i>Lynx</i> . . . . .	41
Figura 14 – Gráfico do conjunto de dados <i>Exchange rate</i> . . . . .	42
Figura 15 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados <i>Sunspot</i> . . . . .	47
Figura 16 – (a)Gráfico QQ-Plot da gaussiana para os resíduos dos dados <i>Sunspot</i> , (b)Gráfico da FAC para os resíduos dos dados <i>Sunspot</i> , (c)Gráfico da FACP para os resíduos dos dados <i>Sunspot</i> . . . . .	47
Figura 17 – Resultados obtidos após a Decomposição <i>Wavelet</i> no 2º nível para os dados <i>Sunspot</i> . . . . .	48
Figura 18 – Filtros FIR para Wavelet Discreta de Meyer . . . . .	49
Figura 19 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados <i>Lynx</i> . . . . .	50

Figura 20 – (a) Gráfico QQ-Plot da gaussiana para os resíduos dos dados <i>Lynx</i> , (b)Gráfico da FAC para os resíduos dos dados <i>Lynx</i> , (c)Gráfico da FACP para os resíduos dos dados <i>Lynx</i> . . . . .	50
Figura 21 – Resultados obtidos após a Decomposição <i>Wavelet</i> no 2º nível para os dados <i>Lynx</i> . . . . .	51
Figura 22 – Filtros FIR para <i>Wavelet</i> Symlet . . . . .	52
Figura 23 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados <i>Exchange rate</i> nas primeiras 365 semanas. . . . .	52
Figura 24 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados <i>Exchange rate</i> nas últimas 365 semanas. . . . .	53
Figura 25 – (a)Gráfico QQ-Plot da gaussiana para os resíduos dos dados <i>Exchange rate</i> , (b)Gráfico da FAC para os resíduos dos dados <i>Exchange rate</i> , (c)Gráfico da FACP para os resíduos dos dados <i>Exchange rate</i> . . . . .	54
Figura 26 – Resultados obtidos após a Decomposição <i>Wavelet</i> no 2º nível para os dados <i>Exchange rate</i> . . . . .	55
Figura 27 – Filtros FIR para <i>Wavelet</i> Coiflet . . . . .	56
Figura 28 – (a)Resultados da previsão do modelo de RNAs após filtragem via WavDS: (a) <i>Sunspot</i> no 2º caso, (b) <i>Lynx</i> , (c) <i>Exchange rate</i> no 3º caso. . . . .	62
Figura 29 – Gráfico comparativo das previsões do método de MTAEF para os dados sem ajuste de fase - caso (a); e com ajuste de fase - caso (b), para o conjunto de teste dos dados <i>Lynx</i> . . . . .	65
Figura 30 – Gráfico comparativo do ajuste do método MTAEF para os dados sem ajuste de fase (a) e com ajuste de fase (b) para o conjunto de treino dos dados <i>Lynx</i> . . . . .	66
Figura 31 – Gráfico comparativo do ajuste do método MTAEF para os dados sem ajuste de fase (a) e com ajuste de fase (b) para o conjunto de validação dos dados <i>Lynx</i> . . . . .	66
Figura 32 – Gráficos dos resultados da previsão sem ajuste de fase para o conjunto de dados <i>Sunspot</i> - 4º caso. . . . .	68
Figura 33 – Gráfico ilustrativo da Função <i>Wavelet</i> Discreta de Meyer . . . . .	78
Figura 34 – Gráfico ilustrativo da Função <i>Wavelet</i> Symlet 8 . . . . .	79
Figura 35 – Gráfico ilustrativo da Função <i>Wavelet</i> Coiflet 5 . . . . .	79
Figura 36 – Gráfico ilustrativo da Função <i>Wavelet</i> Haar . . . . .	80
Figura 37 – Gráfico ilustrativo da Função <i>Wavelet</i> Daubechies 2 . . . . .	80
Figura 38 – Gráfico ilustrativo da Função Escala Discreta de Meyer . . . . .	81
Figura 39 – Gráfico ilustrativo da Função Escala Symlet 8 . . . . .	82
Figura 40 – Gráfico ilustrativo da Função Escala Coiflet 5 . . . . .	82

Figura 41 – Gráfico ilustrativo da Função Escala Haar . . . . . 83  
Figura 42 – Gráfico ilustrativo da Função Escala Daubechies 2 . . . . . 83

## Lista de tabelas

Tabela 1 – Composição amostral dos conjuntos de dados <i>Sunspot</i> , <i>Lynx</i> e <i>Rate</i> . . .	43
Tabela 2 – Horizontes de previsão utilizados nas situações experimentais para os conjuntos de dados <i>Sunspot</i> , <i>Lynx</i> e <i>Rate</i> . . . . .	43
Tabela 3 – Estatísticas descritivas dos resultados da previsão via RNAs para os conjuntos <i>Sunspot</i> e <i>Lynx</i> antes e depois da utilização da filtragem após 50 repetições do experimento. . . . .	57
Tabela 4 – Estatísticas descritivas dos resultados da previsão via RNAs para os conjuntos <i>Exchange rate</i> antes e depois da utilização da filtragem após 50 repetições do experimento. . . . .	59
Tabela 5 – Resultados das previsões obtidos pelo modelo de RNAs após filtragem via WavDS e por modelos híbridos para os conjuntos de dados <i>Sunspot</i> e <i>Lynx</i> na 1 <sup>a</sup> situação experimental. . . . .	61
Tabela 6 – Resultados das previsões obtidos pelo modelo de RNAs após filtragem via WavDS e por modelos híbridos para os conjuntos de dados <i>Exchange rate</i> na 1 <sup>a</sup> situação experimental. . . . .	61
Tabela 7 – Resultados das previsões obtidos pelos métodos MTAEF e modelos híbridos para os conjuntos de dados <i>Sunspot</i> na 2 <sup>a</sup> situação experimental. . . . .	63
Tabela 8 – Resultados das previsões obtidos pelos métodos WavDS+RNA, MTAEF e modelos híbridos para os conjuntos de dados <i>Sunspot</i> na 2 <sup>a</sup> situação experimental. . . . .	67

# Lista de abreviaturas e siglas

AG	Algoritmo genético
AGM	Algoritmo genético modificado
AMR	Análise de múltipla resolução
AR	Autorregressivos
ARV	Variância relativa média
ARIMA	Autorregressivos integrados de médias móveis
CP1	Primeiro critério de parada
CP2	Segundo critério de parada
DW	Decomposição <i>Wavelet</i>
FAC	Função de autocorrelação
FIR	Filtros de resposta a impulso finita
FACP	Função de autocorrelação parcial
GARCH	Modelo de heterocedasticidade condicional autoregressivo generalizado
iid	Independente e indenticamente distribuído
IA	Inteligência artificial
LM	<i>Levenberg Marquardt</i>
MAD	Desvio mediano absoluto
MAE	Erro absoluto médio
MAPE	Erro percentual absoluto médio
MLP	<i>Multilayer perceptron</i>
MIMO-AR	<i>Multiple input-multiple output autoregressive</i>

MSE	Erro quadrático médio
MTAEF	<i>Multiresolution time-delay added evolutionary forecasting</i>
POCID	<i>Prediction on change in direction</i>
RBF	<i>Radial basis function</i>
RNAs	Redes neurais artificiais
Rprop	<i>Resiliente propagation</i>
SVM	<i>Support vector machines</i>
TAEF	<i>Time-delay added evolutionary forecasting</i>
Theil	Estatística U de Theil
TW	Transformada <i>Wavelet</i>
TWI	Transformada <i>Wavelet</i> inversa
TWD	Transformada <i>Wavelet</i> discreta
WavDS	Sistema de decomposição <i>Wavelet</i>
WNN	<i>Wavelet neural network</i>

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Motivação	1
1.2	Objetivos	3
1.2.1	Geral	3
1.2.2	Específicos	3
1.3	Organização da Tese	4
<b>2</b>	<b>Revisão de Literatura</b>	<b>6</b>
2.1	Análise de séries temporais e previsão	6
2.1.1	Pré-processamento em séries temporais	8
2.1.2	Ruído em séries temporais	8
2.2	Wavelets aplicadas à previsão de séries temporais	9
2.3	Modelos Híbridos que utilizam <i>Wavelets</i> e RNAs	10
2.4	Algoritmos Genéticos aplicados à Previsão de Séries Temporais	12
<b>3</b>	<b>Metodologia e Materiais</b>	<b>13</b>
3.1	Métodos	13
3.1.1	Análise Wavelet	13
3.1.1.1	Função Escala	16
3.1.1.2	Análise de Múltipla Resolução	18
3.1.2	Estrutura da Filtragem de Séries Temporais via Wavelets	20
3.1.3	Redes Neurais Artificiais - RNAs	23
3.1.3.1	Redes Neurais <i>Multilayer Perceptron</i>	24
3.1.3.2	Previsão de séries temporais via RNAs	26
3.1.3.3	O problema da previsão de séries temporais via RNAs	29
3.1.4	Algoritmos Genéticos - AGs	30
3.1.5	Time-delay Added Evolutionary Forecasting - TAEF	32
3.2	Metodologia Proposta	35
3.2.1	O Sistema de Decomposição Wavelet - WavDS	35
3.2.2	Multiresolution Time-delay Added Evolutionary Forecasting - MTAEF	38
3.3	Materiais	39
3.3.1	Dados	39
3.3.2	Métricas de Acurácia	43

<b>4</b>	<b>Resultados e Discussão</b>	<b>46</b>
4.1	Análise dos resultados da filtragem do método WavDS	46
4.1.1	Caso Wolf's sunspot	46
4.1.2	Caso Canadian lynx	49
4.1.3	Caso Exchange rate	52
4.2	Análise do desempenho do filtro WavDS	56
4.3	Análise do método híbrido MTAEF	62
<b>5</b>	<b>Considerações Finais</b>	<b>69</b>
5.1	Conclusões	69
5.2	Atividades Futuras	70
	<b>Referências Bibliográficas</b>	<b>72</b>
	<b>Apêndices</b>	<b>77</b>
	APÊNDICE A Gráficos das Funções Wavelets ( $\psi$ ) utilizadas nesta pesquisa	78
	APÊNDICE B Gráficos das Funções Escala ( $\phi$ ) utilizadas nesta pesquisa	81
	<b>Anexos</b>	<b>84</b>
	ANEXO A Evolução da população dos Algoritmos genéticos	85
	A.1 O Teorema dos Esquemas	85

# 1 Introdução

Neste capítulo, a motivação com justificativas, assim como uma breve evolução histórica a cerca da teoria que envolve análise *Wavelet* são descritas na Seção 1.1. Os objetivos desta Tese, incluindo o geral e os específicos são apresentados na Seção 1.2. No final do capítulo, na Seção 1.3, é apresentada a organização desta Tese.

## 1.1 Motivação

Metodologias de previsão de séries temporais sugerem que observações passadas de uma variável sejam analisadas para aplicar modelos que descrevam suas características fundamentais.

Usualmente, pesquisadores consideram que uma série temporal é um processo determinístico adicionado de ruído. Deste modo, dados com características ruidosas possuem inicialmente a indisponibilidade de informações transparentes do seu comportamento no passado para captar plenamente a sua dependência com o futuro (ABABNEH; WADI; ISMAIL, 2013; WADI; ISMAIL, 2011; CATALÃO; POUSINHO; MENDES, 2011; LIU et al., 2013; KILIÇ; UĞUR, 2016; LIU et al., 2016).

Neste sentido a análise *Wavelet* como metodologia de pré-processamento nos dados, com a proposta de redução ou remoção de ruído, tem atraído a atenção de pesquisadores da área por ser uma metodologia que possibilita obtenção de maior precisão com a utilização de um procedimento de previsão e, conseqüentemente, modelos com resultados mais confiáveis (LIU et al., 2013; CHANDAR; SUMATHI; SIVANANDAM, 2016; LIU; LU; WANG, 2012; JOO; KIM, 2015; KAO et al., 2013; RODRIGUEZ; BARBA, 2016).

A análise *Wavelet* é um método que possui fundamentação matemática obtida a partir da derivação da teoria de análise frequencial de Fourier, que foi inicialmente estabelecida no século XIX, e possui comprovadamente grande importância e influência até os dias atuais (MORETTIN, 2014; SIDNEY, 1998).

A idéia inicial sobre as *Wavelets* foi pela primeira vez mencionadas no apêndice da tese de Haar em 1909 (SIDNEY, 1998). A *Wavelet* de Haar ficou no anonimato por vários anos, até que na década de 30 vários grupos trabalhando independentemente, pesquisaram a representação de funções usando uma base variável com a escala (MORETTIN, 2014).

A fundamentação matemática atual sobre *Wavelets* foi inicialmente desenvolvida por Jean Morlet e por Alex Grossman em 1981. Os métodos que envolvem análise *Wavelet* foram desenvolvidos principalmente por Meyer que assegurou a disseminação dos métodos (SIDNEY, 1998).

A pesquisa que envolve pré-processamento de séries temporais por meio de *Wavelets* (*Wavelet denoise*) teve seu início desde a pesquisa desenvolvida por Mallat (MALLAT, 1989) até sua metodologia tradicional apresentada por Donoho (DONOHO, 1995). Desde então, as pesquisas que envolvem *Wavelets* tem crescido rapidamente nos últimos anos (LIU; LU; WANG, 2012; MORETTIN, 2014; WANG; ZHANG; ZHOU, 2016).

Todavia, em questões referentes à pré-processamento nos dados, é de concordância entre pesquisadores que o procedimento *Wavelet denoise* possui limitações.

A retirada apenas de um tipo específico de ruído, o ruído branco gaussiano e a seleção de alguns parâmetros da filtragem via *Wavelets*, como por exemplo a escolha da função *Wavelet* e os limiares de filtragem, são limitações que interferem diretamente nos resultados da filtragem (KAO et al., 2013; RODRIGUEZ; BARBA, 2016; FLORES et al., 2016; LIU et al., 2016; ALI; EL-DAHSHAN; YAHIA, 2017; SEO et al., 2015; WANG; ZHANG; ZHOU, 2016).

Em sua utilização no âmbito de séries temporais, diversos pesquisadores comprovaram a eficácia com a utilização de *Wavelets* em conjunto com modelos preditivos, denominando assim modelos considerados híbridos, que combinam pré-processamento em séries temporais via Decomposição *Wavelet* (DW) com a utilização de modelos de previsão (ABABNEH; WADI; ISMAIL, 2013; CATALÃO; POUSINHO; MENDES, 2011; LIU et al., 2013; KHANDELWAL; ADHIKARI; VERMA, 2015; MUSTAFA; ALZUBAYDI, 2013; KILIÇ; UĞUR, 2016).

Nesse contexto, Redes neurais artificiais (RNAs) do tipo *Multilayer perceptron* (MLP) são comumente utilizadas em conjunto com métodos de filtragem via *Wavelets* e foram apresentadas à literatura como uma alternativa que possibilitou, em diversos casos, ganho em exatidão quando comparados os seus resultados aos resultados obtidos a partir de modelos tradicionais (CATALÃO; POUSINHO; MENDES, 2011; MINU; LINEESH; JOHN, 2010; CHANDAR; SUMATHI; SIVANANDAM, 2016; ADAMOWSKI; CHAN, 2011).

Contudo, diversos autores afirmam que a seleção dos parâmetros das RNAs é dada de maneira heurística e que este é um problema à ser resolvido (HAYKIN, 2007; FERREIRA; VASCONCELOS; ADEODATO, 2008; DE MATTOS NETO et al., 2014; DE MATTOS NETO; FERREIRA, 2016; DONATE et al., 2013; LIU; LU; WANG, 2012;

DE MATTOS NETO; JUNIOR; FERREIRA, 2010; ASADI et al., 2012).

Donate et al. (2013) enfatiza que a questão fundamental para o sucesso de um modelo MLP para análise e previsão de séries temporais é a adequada seleção de seus parâmetros. Para contornar este problema os autores utilizam um procedimento meta-heurístico que envolve computação evolutiva (HOLLAND, 1992) chamado Algoritmo genético (AG).

Métodos meta-heurísticos inspirados na teoria da evolução, por meio de AGs (HOLLAND, 1992), são realizados de maneira interativa permitindo assim a evolução do processo para encontrar uma solução aceitável de um determinado problema (FERREIRA; VASCONCELOS; ADEODATO, 2008; HOLLAND, 1992; DE MATTOS NETO; JUNIOR; FERREIRA, 2010; DE MATTOS NETO et al., 2014; DE MATTOS NETO; FERREIRA, 2016; DONATE et al., 2013; ASADI et al., 2012; ARAÚJO; FERREIRA, 2009).

Para FERREIRA (2006), AGs são métodos de Inteligência artificial (IA) utilizados para aprendizagem, busca e otimização de resultados. Por meio de um processo iterativo, o AG busca a melhor solução para um determinado problema partindo de um conjunto de estados iniciais que são formados por possíveis soluções deste problema.

Uma metodologia que combina Algoritmo genético modificado (AGM) e RNAs do tipo MLP, intitulada *Time-delay added evolutionary forecasting* (TAEF) (FERREIRA; VASCONCELOS; ADEODATO, 2008), proporcionou resultados superiores em exatidão quando comparados a de diversos modelos híbridos (FERREIRA; VASCONCELOS; ADEODATO, 2008; DE MATTOS NETO et al., 2014; DE MATTOS NETO; FERREIRA, 2016; DE MATTOS NETO; JUNIOR; FERREIRA, 2010; LIU; LU; WANG, 2012).

Contudo, a busca por métodos preditores que possibilitem melhorias na precisão da previsão de séries temporais é incessante e tem se tornado cada vez maior com a utilização de modelos híbridos. Para Liu, Lu e Wang (2012) e Asadi et al. (2012) a combinação que envolve métodos de filtragem e AGM são de suma importância para o sucesso do modelo preditivo.

Em virtude de todo o exposto, é proposto nesta Tese uma novo procedimento híbrido que combina *Wavelets*, RNAs e AGM intitulado *Multiresolution Time-delay Added Evolutionary Forecasting* (TAEF). Este método é a composição de um procedimento inovador de pré-processamento em séries temporais, intitulada Sistema de decomposição via *Wavelet* (WavDS), que é combinado com o método TAEF (FERREIRA; VASCONCELOS; ADEODATO, 2008) em virtude de sua capacidade de ajuste à dados com características não lineares e seu poder de generalização na previsão de séries temporais.

Com a proposta, esta Tese visa deixar como principal contribuição uma nova

metodologia de pré-processamento de dados utilizando *Wavelets* que tem por objetivo eliminar diferentes tipos de ruídos que estão presentes na série temporal e fornecer um indicativo para a seleção da função *Wavelet* superando assim limitações encontradas na literatura da área (ABABNEH; WADI; ISMAIL, 2013; ALI; EL-DAHSHAN; YAHIA, 2017; LIU et al., 2013; KHANDELWAL; ADHIKARI; VERMA, 2015; LIU et al., 2016; MUSTAFA; ALZUBAYDI, 2013; WANG; ZHANG; ZHOU, 2016; KILIÇ; UĞUR, 2016).

## 1.2 Objetivos

A fim de definir os objetivos deste trabalho apresenta-se o objetivo geral e os objetivos específicos.

### 1.2.1 Geral

O objetivo geral deste trabalho é propor uma nova metodologia híbrida que utiliza, concomitantemente, um sistema de decomposição via *Wavelets*, AGMs e RNAs do tipo MLP para previsão de séries temporais.

### 1.2.2 Específicos

A fim de que o objetivo geral seja alcançado, faz-se necessário realizar os seguintes objetivos específicos:

- Avaliar a influência do pré-processamento em séries temporais na qualidade das previsões por meio da análise da variabilidade estatística e da acurácia nos seu resultados;
- Realizar um levantamento do referencial teórico a respeito de modelos híbridos que utilizam em conjunto *Wavelets* com técnicas de previsão de séries temporais;
- Analisar as principais contribuições e limitações dos modelos híbridos citados no item anterior;
- Propor uma nova abordagem para pré-processamento de séries temporais que utilizam a DW superando as limitações encontradas na literatura;
- Analisar as principais contribuições e limitações do modelo de previsão de RNAs do tipo MLP na sua utilização em conjunto com *Wavelets* na composição de modelos híbridos;

- Analisar a variabilidade estatística e a robustez dos resultados da previsão do modelo MLP;
- Comparar os resultados experimentais obtidos à alguns dos resultados disponibilizados na literatura.

### 1.3 Organização da Tese

Com o intuito de apresentar esta metodologia, esta Tese foi estruturada em 5 capítulos. Destes, os próximos 4 capítulos estão descritos a seguir:

**Capítulo 2 - Revisão de Literatura:** Neste capítulo são expostas questões referentes à análise e previsão de séries temporais e sua relação com o ruído. Alguns detalhes sobre o procedimento de filtragem via *Wavelets* aplicado ao problema de previsão de séries temporais são apresentados. A maneira como a qual os AG podem contribuir na previsão de séreis temporais também é mostrada.

**Capítulo 3 - Materiais e Métodos:** Neste capítulo, os métodos que dão todo aporte teórico para a implementação das metodologias propostas são apresentados. As limitações encontradas com os métodos tradicionais de filtragem de dados são expostas e a justificativa para os procedimentos propostos do método são apresentados. São comentadas as contribuições para o estado da arte, a justificativa da escolha do modelo de RNAs como processo preditor e, por fim, o banco de dados com a descrição dos tamanhos amostrais e transformações realizadas são apresentados.

**Capítulo 4 - Resultados e Discussão:** Neste capítulo, os métodos propostos descritos no Capítulo 3 são aplicados em três tradicionais séries temporais em várias janelas de tempo. Os resultados dos experimentos realizados com os métodos propostos são apresentados e comentados.

**Capítulo 5 - Considerações Finais:** Neste capítulo são apresentadas as conclusões do trabalho desenvolvido a partir dos experimentos realizados com os métodos propostos aplicados às séries temporais. Por fim, é realizada uma discussão sobre a nova abordagem proposta, bem como suas limitações e propostas para trabalhos futuros.

## 2 Revisão de Literatura

Neste capítulo, são apresentados na Seção 2.1 alguns conceitos básicos e definições sobre pré-processamento em séries temporais e ruído. Importantes resultados que mostram a melhora em acurácia com a utilização de modelos de previsão em conjunto com *Wavelets*, assim como a utilização de *Wavelets* em conjunto com RNAs, são apresentados nas Seções 2.2 e 2.3. Na Seção 2.4 são apresentados alguns trabalhos que mostram a maneira com a qual os AGs podem contribuir na previsão de séries temporais.

### 2.1 Análise de séries temporais e previsão

Uma série temporal pode ser definida por um conjunto de informações, ou observações, de uma variável de interesse. Geralmente estas informações são capturadas consecutivamente e igualmente espaçadas, resultando assim em uma sequência ordenada no tempo (BOX et al., 2015). Ela pode ser considerada uma série temporal discreta ou contínua, estacionária ou não estacionária, homocedástica ou heterocedástica, tanto no contexto univariado como no multivariado.

A suposição básica que norteia a análise de séries temporais é que há um processo relacionado com o tempo que exerce influência sobre os dados no passado e no futuro. Também observa-se que existe um comportamento aleatório exercendo influência nas informações. Assim, o processo gerador da série costuma atuar criando padrões aleatórios e não aleatórios que podem ser detectados mediante alguns processos estatísticos. Desta forma, a série temporal pode ser considerada a combinação aditiva de dois processos, um determinístico e outro estocástico (BOX et al., 2015; WU; HUANG, 2009; WADI; ISMAIL, 2011; ABABNEH; WADI; ISMAIL, 2013; SIDNEY, 1998; VERMA; VERMA, 2012; DONOHO, 1995; KILIÇ; UĞUR, 2016; FLORES et al., 2016). Em virtude disto, nesta Tese, a série temporal também será considerada a combinação aditiva destes dois processos.

A série temporal é considerada determinística quando pode ser definida por uma função matemática que representa sua dependência temporal com precisão. A maior parte das séries temporais que representam fenômenos reais contêm um elemento aleatório que dificulta sua representação com exatidão. Quando uma série temporal contém uma componente estocástica ou aleatória, ela é denominada um processo estocástico. Em virtude

disto, a série temporal  $\{y_t\}_{t=1,2,\dots,N}$  apresentada na Equação 1 pode ser representada pela soma da componente determinística ( $\mu$ ), que representa o comportamento fundamental da série temporal, com um componente aleatório ( $e_t$ ) que também é conhecido como erro ou ruído.

$$y_t = \mu + e_t \quad (1)$$

de modo que  $t$  é um índice cronológico ou espacial de tamanho  $N$  (considera-se geralmente o tempo).

O objetivo principal da análise de séries temporais é, por meio da compreensão do seu mecanismo gerador, sintetizar a dependência temporal entre as informações para prever observações futuras. De modo geral, pode-se destacar os seguintes objetivos específicos no estudo de séries temporais (BOX et al., 2015; KILIÇ; UĞUR, 2016):

- Descrição: Descrever propriedades da série, tais como tendência, sazonalidade, ciclo e a existência de valores discrepantes ou alterações estruturais decorrentes das mudanças de tendências ou de sazonalidades;
- Explicação: Utilizar a variabilidade de uma série para explicar a variabilidade de outra, no caso multivariado;
- Previsão: Realizar previsões baseadas em valores anteriormente registrados. Neste estudo, assume-se a existência de incerteza e desenvolve-se esforços para reduzir os erros inerentes à previsão.
- Controle: Analisar origens de falhas para melhorar processos e assim obter um maior rigor, quando os valores da série temporal medem a qualidade de uma atividade.

De forma geral, o problema de previsão de séries temporais pode ser analisado sob duas diretrizes: a temporal e a espectral. A análise das séries no domínio do tempo (temporal) e no domínio da frequência (espectral) são dois enfoques alternativos e complementares (WANG et al., 2015). A representação do domínio do tempo dá a amplitude da série no instante de tempo. Entretanto, em muitos casos, importantes informações sobre o comportamento da série só podem ser visualizadas nas componentes de frequência de uma série temporal, em vez de um único valor da amplitude no tempo (KILIÇ; UĞUR, 2016; BULLMORE et al., 2001; WANG; ZHANG; ZHOU, 2016).

Apesar de ser mais utilizada na literatura, a análise no domínio do tempo não é adequada para compreender completamente a estrutura dos dados, especialmente em séries temporais não estacionárias que possuem componentes de altas frequências (LIU et al., 2016; CHANDAR; SUMATHI; SIVANANDAM, 2016; KILIÇ; UĞUR, 2016; NALBAND; SREEKRISHNA; PRINCE, 2016; WANG et al., 2015; WANG; ZHANG; ZHOU, 2016).

Como exemplo, pode-se citar o ruído branco, ele possui difícil visualização quando a série é inspecionada no domínio do tempo. Já quando sua análise é realizada no domínio da frequência pode-se identificar mais facilmente suas frequências harmônicas e amplitudes (BULLMORE et al., 2001; VERMA; VERMA, 2012; FLORES et al., 2016; KAO et al., 2013; LIU et al., 2013; KILIÇ; UĞUR, 2016).

O ruído é uma perturbação indesejada dentro de uma faixa de frequências. Consiste num distúrbio que afeta e distorce o comportamento fundamental da série e possui propriedades estatísticas conhecidas de amplitude, distribuição e densidade espectral (BULLMORE et al., 2001; FLORES et al., 2016).

### 2.1.1 Pré-processamento em séries temporais

Na realização de uma análise estatística envolvendo séries temporais a dificuldade encontrada na grande maioria das vezes é devido a comportamentos irregulares, considerados ruídos, que estão presentes nos dados (BULLMORE et al., 2001; ZHANG; YANG; LIU, 2005; FLORES et al., 2016; WU; HUANG, 2009).

Pesquisadores comprovaram que dados com características pertinentes ao ruído branco possuem inicialmente a indisponibilidade de informações transparentes do seu comportamento no passado para captar plenamente a sua dependência com o futuro (ABABNEH; WADI; ISMAIL, 2013; WADI; ISMAIL, 2011; CATALÃO; POUSINHO; MENDES, 2011; LIU et al., 2013; WANG; ZHANG; ZHOU, 2016).

Contudo, estas anomalias devem ser previamente detectadas e devidamente removidas antes da etapa de análise dos dados. Esta fase é chamada de pré-processamento e é uma etapa crucial na análise de séries temporais principalmente quando a estabilidade de um modelo é requerida (LIU et al., 2013; FLORES et al., 2016; ZHANG; YANG; LIU, 2005; LIU et al., 2016; JOO; KIM, 2015; WANG; ZHANG; ZHOU, 2016). Segundo Zhang, Yang e Liu (2005) o pré-processamento nos dados é a etapa mais importante e corresponde a 80% do sucesso no processo de análise da série temporal.

Os métodos de pré-processamento podem visar objetivos diferentes, como descrição de dados, investigações de correlações e/ou filtragem de ruído para previsão valores futuros, e cada qual com seus diferentes pressupostos. Nesta Tese a ênfase dada ao pré-processamento é na questão que envolve a remoção de diversos tipos de ruído para previsão de séries temporais.

### 2.1.2 Ruído em séries temporais

Segundo Flores et al. (2016) o ruído assumido como presente nas séries temporais era apenas o ruído branco gaussiano e estudos recentes fornecem evidências de que a presença de outros tipos de ruído possuem interferência no comportamento de uma série temporal e na eficácia de modelo de previsão. Deste modo, tendo em vista as diferentes caracterizações e magnitudes pertinentes a um ruído, métodos mais eficazes de pré-processamento nos dados se fazem necessários devido às características únicas e intrínsecas que comumente não são perceptíveis nas séries temporais (BULLMORE et al., 2001).

A densidade espectral é uma distribuição de energia no espectro de frequências e é uma propriedade que permite distinguir os diferentes tipos de ruído. Estes ruídos são associados à cores que identificam suas propriedades. Há ruído rosa, marrom, vermelho, cinza e branco. A classificação do ruído, a partir da densidade espectral, é apresentada por nomes associados às suas respectivas cores (BULLMORE et al., 2001; FLORES et al., 2016; NALBAND; SREEKRISHNA; PRINCE, 2016). Como exemplo podemos citar o ruído branco, o mais popular deles.

O ruído branco gaussiano é uma série de dados aleatórios com uma densidade de potência espectral plana, ou seja, a potência é constante em todas as faixas de frequência. Ele apresenta constantes valores de energia para todo o espectro de frequências. É estacionário e sem dependência temporal, resultando em um processo estocástico onde todas as variáveis aleatórias seguem distribuição normal com média e variância constantes e covariâncias nulas (BULLMORE et al., 2001; GANNET, 1996).

O Modelo de decomposição empírico (EEMD) (NALBAND; SREEKRISHNA; PRINCE, 2016) e o modelo de Decomposição *Wavelet* (DONOHO, 1995) podem ser considerados os métodos de pré-processamento em séries temporais que proporcionam maior êxito (JOO; KIM, 2015; WANG; ZHANG; ZHOU, 2016; ALI; EL-DAHSHAN; YAHIA, 2017). A principal limitação destes métodos é a retirada de apenas um tipo específico de ruído, o ruído branco gaussiano .

Análise de *Wavelets*, como metodologia de filtragem nos dados com a proposta de remoção de ruído, tem atraído atenção de pesquisadores por ser um método que possibilita a obtenção de maior acurácia com a utilização de um procedimento de previsão de séries temporais com baixo custo computacional (KILIÇ; UĞUR, 2016; LIU et al., 2016; CHANDAR; SUMATHI; SIVANANDAM, 2016; JOO; KIM, 2015; SEO et al., 2015; KHANDELWAL; ADHIKARI; VERMA, 2015; RODRIGUEZ; BARBA, 2016).

## 2.2 Wavelets aplicadas à previsão de séries temporais

Diversos pesquisadores obtiveram melhoria significativa nos resultados de previsão com o procedimento anteriormente mencionado. [Soltani \(2002\)](#) utilizou a Decomposição *Wavelet* (DW) como procedimento de filtragem no pré-processamento das clássicas séries temporais *Sunspots* e *Mackey–Glass* e verificou que os resultados obtidos proporcionaram melhora nos resultados da previsão. Na pesquisa apresentada por [Wadi e Ismail \(2011\)](#) foram comparados os resultados da previsão por meio dos modelos autorregressivos integrados de médias móveis (ARIMA) antes e depois do procedimento de filtragem via *Wavelets*, enfatizando a importância da Transformada na retirada de efeitos irregulares presentes nos dados.

Em [Kao et al. \(2013\)](#) a Transformada *Wavelet* é proposta como um método que possibilita a redução nos efeitos irregulares, permitindo melhor visualização do comportamento fundamental da série temporal, preservando padrões, tendências e surpresas nos dados. Joo & Kim ([JOO; KIM, 2015](#)) propuseram um procedimento semelhante ao realizado em Wadi & Ismail ([WADI; ISMAIL, 2011](#)) quando comparam o desempenho com modelos de previsão com e sem a decomposição *Wavelet* em oito cenários, aplicando modelos (ARIMA) nos dados com características não sazonais e modelos ARIMA Sazonais (SARIMA) em dados com características sazonais.

De modo geral, pesquisadores comprovaram a eficácia da filtragem via decomposição *Wavelet*. No primeiro nível da decomposição o conjunto de dados é subdividido em dois subconjuntos caracterizando um *downsampling*, como descrito na Seção 3.1.1.2. Destes subconjuntos, um é composto pelas tendências e o outro por variações da série temporal ([ABABNEH; WADI; ISMAIL, 2013](#); [CATALÃO; POUSINHO; MENDES, 2011](#); [LIU et al., 2016](#); [PANDEY; SINGH; SINHA, 2010](#); [KHANDELWAL; ADHIKARI; VERMA, 2015](#); [SOLTANI, 2002](#); [MUSTAFA; ALZUBAYDI, 2013](#)).

Desta forma, a procura por modelos que melhoram a acurácia na previsão de séries temporais tem-se tornado cada vez maior com a utilização de modelos combinados denominados híbridos. Mesmo com algumas limitações, como por exemplo a escolha da *Wavelet*-mãe, a TW tem importantes contribuições neste sentido.

[Mustafa e ALZUBAYDI \(2013\)](#) comprovaram a eficácia da previsão de séries temporais com a utilização da filtragem via *Wavelet* como pré-processamento dos dados por meio da análise comparativa de quatro *Wavelets*-mãe antes da utilização dos modelos ARIMA, com a *Wavelet Discrete Meyer* apresentando melhores resultados.

[Liu et al. \(2013\)](#) informam que muitos tipos de *wavelets*, incluindo *Daubechies*, *Symmlets*, *Gaussian*, *Mexican hat*, *Morlet* e *Shannon* podem ser utilizadas como *Wavelet*-

mãe em análise de séries temporais, contudo a *Wavelet* escolhida em seu trabalho foi *Daubechies* único e exclusivamente pelo fato desta *Wavelet* ser umas das mais utilizadas na literatura.

Rodriguez e Barba (2016) verificaram que filtragem nos dados via *Wavelet Fejer-Korovkin*, quando aplicado ao modelo MIMO-AR (*Multiple Input-Multiple Output Autoregressive*) proporcionou melhores resultados quando comparado ao pré-processamento realizado via *Wavelet*-mãe *Daubechies* para a aplicação do mesmo modelo.

Para contornar o problema com a seleção da *Wavelet*-mãe apropriada, Khandelwal, Adhikari e Verma (2015) realizaram a média das previsões após o pré-processamento nos dados com três *Wavelets*-mãe.

Contudo, na literatura da área, pode-se perceber que as principais *Wavelets*-mãe escolhidas para a TW são *Haar* e *Daubechies* único e exclusivamente por terem sido as mais utilizadas em outros artigos (ABABNEH; WADI; ISMAIL, 2013; ALI; EL-DAHSHAN; YAHIA, 2017; CHANDAR; SUMATHI; SIVANANDAM, 2016; JOO; KIM, 2015; LIU et al., 2016; MUSTAFA; ALZUBAYDI, 2013; SEO et al., 2015). Desta maneira pode-se evidenciar que a seleção da *Wavelet*-mãe na TW para a decomposição *Wavelet* é um problema à ser resolvido .

## 2.3 Modelos Híbridos que utilizam *Wavelets* e RNAs

Diversos estudos comprovaram que modelos de RNAs e Decomposição *Wavelet* utilizados em conjunto proporcionaram melhorar a eficiência nos resultados da previsão em comparação com modelos tradicionais de previsão. Mesmo sendo utilizados com abordagens e contextos diferentes, eles apresentaram resultados satisfatórios.

Minu, Lineesh e John (2010) utiliza o modelo *Wavelet Neural Network* (WNN) para realizar previsão de uma série temporal do mundo real com características não lineares e não-estacionárias para avaliar a eficiência do método proposto. Analisaram a série temporal de observações mensais de ataques terroristas no mundo no período de fevereiro de 1968 a janeiro de 2007. Comparam os resultados do modelo WNN com os resultados dos modelos Modelo de heterocedasticidade condicional autoregressivo generalizado (GARCH) e RNA do tipo MLP. Obtiveram ganho em acurácia superior a 99% em ambos os casos.

A ideia básica do modelo WNN é utilizar o pré-processamento nos dados por meio da decomposição via *Wavelets* e utilizar RNAs do tipo MLP em cada coeficiente *wavelet* do processo.

Na mesma linha que Minu, Lineesh e John (2010), Chandar, Sumathi e Sivanandam

(2016), Adamowski e Chan (2011) e Seo et al. (2015) utilizam TWD e RNAs. Chandar, Sumathi e Sivanandam (2016) utilizam *waveletes* e RNAs para previsão de séries temporais financeiras e compararam os resultados encontrados com os resultados encontrados pelo modelo de RNAs, já Adamowski e Chan (2011) compararam os resultados encontrados por meio do modelo que utiliza *Waveletes* combinadas à RNAs com os resultados provindos dos modelos ARIMA e de RNAs aplicados separadamente na previsão de níveis de lençóis freáticos e verificaram melhora nos resultados. Seo et al. (2015) aplicaram esta mesma metodologia para realizar previsões de níveis de água com a utilização de diferentes *wavelets*-mãe e encontraram que a performance do modelo, no que diz respeito a arquitetura da rede, é dependente da *wavelets*-mãe escolhida.

Khandelwal, Adhikari e Verma (2015) propõe outra abordagem como proposta de previsão em três conjunto de dados, utilizando como procedimento de previsão o caso *multi-step-ahead*. Utilizam a DW como uma metodologia que divide as séries temporais em estudo em componentes lineares e não lineares. Nas componentes lineares são aplicados os tradicionais modelos ARIMA, e as componentes não lineares em conjunto com os erros encontrados das previsões dos modelos ARIMA são apresentados como entrada para o modelo de RNAs (KHANDELWAL; ADHIKARI; VERMA, 2015).

Liu et al. (2013) fizeram análise em uma série temporal de velocidade de vento com *Waveletes* e RNAs a partir de quatro modelos que possuem formulações diferentes. Encontraram que o modelo *Wavelet Packet*-RNA foi o melhor entre eles. Segundo os autores, este modelo utiliza uma Transformada *Wavelet* especial de modo que, diferente do que é descrito na Seção 3.1.1.2, não são apenas os coeficientes de aproximação  $cA_{j,k}$  que são submetidos ao procedimento de *downsampling*, os coeficientes de detalhes  $cD_{j,k}$  também são. Após a utilização da transformada, cada coeficiente separadamente é apresentado como entrada na rede e as previsões são encontradas (LIU et al., 2013).

Contudo, de maneira geral, pode-se comprovar que a combinação da decomposição via *wavelets* em conjunto com os modelos de RNAs do tipo MLP apresentam bons resultados.

## 2.4 Algoritmos Genéticos aplicados à Previsão de Séries Temporais

AGs são aplicados de maneira indireta na previsão de séries temporais de modo a dar apoio a outros métodos por uma melhor modelagem e, conseqüentemente, proporcionar melhores resultados em exatidão com a utilização de um modelo de previsão (FERREIRA, 2006).

Ferreira, Vasconcelos e Adeodato (2008) apresentaram à literatura um modelo

híbrido inteligente, intitulado *Time-delay added evolutionary forecasting* (TAEF), que combina RNAs do tipo MLP com AGM para previsão de séries temporais. Com a proposta, limitações existentes com a escolha dos parâmetros das MLPs foram contornadas por meio de uma análise comparativa de seus resultados com resultados obtidos por modelos tradicionais em diferentes métricas.

Segundo FERREIRA (2006) a combinação de modelos preditivos de séries temporais com AGM caracterizam sistemas que são chamados Sistemas híbridos inteligentes.

Asadi et al. (2012) propôs uma abordagem que combina pré-processamento nas séries temporais e AGMs para suporte na realização de previsões via MLPs no mercado financeiro de ações. Segundo os autores, esta combinação é uma alternativa promissora. Já Liu et al. (LIU; LU; WANG, 2012) procurou contornar características de complexidade, não linearidade e incerteza que são inerentes às séries temporais por meio de um modelo que combina Análise *Wavelet*, por meio da DW (DONOHO, 1995), RNAs e o AG para viabilizar a dificuldade com a seleção dos parâmetros do modelo de RNAs que serão comentados na Subseção 3.1.3.

De Mattos Neto et al. (2014) utilizaram o procedimento que envolve AGs para o ajuste dos parâmetros de RNAs do tipo MLP e realização de previsões de séries temporais (Séries de partículas inaláveis -  $MP_{10}$  e Série de partículas inaláveis finas  $MP_{2,5}$ ) que são utilizadas como indicadores de qualidade do ar para duas cidades diferentes. Esta análise apresentou desempenho superior quando comparou seus resultados obtidos aos obtidos por outras técnicas em todas as séries analisadas.

Na mesma linha de pesquisa, De Mattos Neto e Ferreira (2016) utilizaram AGM para a busca dos melhores parâmetros das MLPs para realizar previsões no mercado de ações em séries temporais de quatro companhias financeiras internacionais (Amazon, Apple, Google e Intel).

Desta maneira, de acordo tudo que foi exposto nas Seções 2.1, 2.2, 2.3 e 2.4, a modelagem que faz uso de sistemas híbridos inteligentes e métodos de pré-processamento nos dados, compõe um conjunto de técnicas que otimizam os resultados de uma metodologia de previsão.

## 3 Metodologia e Materiais

Neste capítulo, uma nova metodologia para pré-processamento de dados e um novo método híbrido para análise e previsão de séries temporais são propostos, assim como a descrição dos conjuntos de dados utilizados nas análises é apresentada. Primeiramente, na Seção 3.1, as metodologias utilizadas nesta Tese são apresentadas. Na Subseção 3.1.1 o procedimento de análise via *Wavelets* é formulado e na Subseção 3.1.2 é mostrada a fundamentação matemática a cerca da filtragem de séries temporais via *Wavelets*. A metodologia que envolve previsão de séries temporais por meio de RNAs é abordada na Subseção 3.1.3. Na Subseção 3.1.4 e 3.1.5 são abordados, respectivamente, os tópicos referentes à Algoritmos genéticos e ao modelo híbrido que será utilizado nesta Tese. Na Seção 3.2, uma nova abordagem para filtragem de séries temporais, com sua respectiva motivação e justificativa, assim como o método híbrido proposto são apresentados. Na Seção 3.3 os dados, com suas respectivas descrições, assim como as situações que serão analisadas neste estudo experimental são apresentados.

Neste capítulo a descrição dos conjuntos de dados utilizados nos experimentos é apresentada, assim como são analisados e comparados todos resultados obtidos

### 3.1 Métodos

#### 3.1.1 Análise Wavelet

As *Wavelets*, também conhecidas por *Ondelette* ou Ondaletas, são representações matemáticas de pequenas ondas. Em virtude disto o termo em inglês *Wavelets* é utilizado para sua representação (MORETTIN, 2014; TEIXEIRA, 2015; SOUZA et al., 2007).

Pode-se pensar em *Wavelets* como alternativas a outros sistemas de funções usados como base para a representação de funções pertencentes a certos espaços, como os senos, cosseno, polinômios ortogonais e etc. O problema básico apresentado aqui é o de codificar informações que estão apresentadas em forma de dados. Um fato importante sobre as *Wavelets* é que elas são localizadas no tempo e no espaço. Este comportamento as tornam ideais para analisar séries temporais não estacionárias (MORETTIN, 2014; TEIXEIRA, 2015; SOUZA et al., 2007).

As *Wavelets* são representadas em blocos, possuem propriedades de translação e

dilatação em uma única função  $\psi_{a,b}(t)$  que é chamada função *Wavelet* no tempo  $t$ . Na Equação 2 esta relação é apresentada.

$$\psi_{a,b}(t) = |a|^{\frac{-1}{2}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

de modo que  $a, b \in \mathbb{R}$  com  $a > 0$ .

A alteração dos parâmetros  $a$  e  $b$  da Equação 2 geram várias *Wavelets* a partir da função *Wavelet*  $\psi(t)$ , que é comumente chamada *Wavelet*-mãe, e que está definida em  $t \in \mathbb{R}$  (MORETTIN, 2014; SOUZA et al., 2007).

O termo  $a$  representa o parâmetro de contração, caso seu valor aumente, ou de dilatação, no caso contrario, e refere-se a escala utilizada. O termo  $b$  é a representação do parâmetro de translação, isto é, a localização da *Wavelet*-mãe. O termo  $|a|^{\frac{-1}{2}}$  corresponde a um fator de normalização para a energia da *Wavelet*  $\psi_{a,b}$  manter-se localizada em uma dada região do espaço. Isto significa dizer que as amplitudes da *Wavelet*-mãe  $\psi(t)$  são apreciáveis apenas nesta região localizada (SIDNEY, 1998). Algumas funções *Wavelets*-mãe estão ilustradas graficamente no Apêndice A.

Segundo Morettin (MORETTIN, 2014), quando se trabalham com dados de valores discretos, por conveniência matemática, definem-se valores especiais para  $a$  e  $b$ :  $a = 2^{-j}$  e  $b = k \cdot 2^{-j}$ , com  $j, k \in \mathbb{R}$ . Por analogia com a análise de Fourier, considere o espaço  $L^2(\mathbb{R})$  de todas as funções mensuráveis de quadrado integrável <sup>1</sup> sobre  $\mathbb{R}$ . A ideia é considerar dilatações (ou compressões) e translações de uma única função  $\psi(t)$  (*Wavelet*-mãe) de modo a cobrir  $\mathbb{R}$ . Desta forma as *Wavelets*  $\psi_{j,k}(t)$ , que são geradas a partir da *Wavelet*-mãe, possuem representação matemática dada por:

$$\psi_{j,k}(t) = 2^{\frac{j}{2}} \psi(2^j t - k), \quad (3)$$

com  $j, k \in \mathbb{Z}$ , são obtidas a partir de uma dilatação binária  $2^j$  e uma translação diádica  $k \cdot 2^{-j}$ . Desta forma está sendo considerado um caso particular da Equação 2.

Os parâmetros  $j$  e  $k$  representam, respectivamente, escala e localização temporal. A parametrização no tempo e no espaço por  $k$  e a frequência ou escala por  $j$  torna o procedimento computacionalmente eficaz (WADI; ISMAIL, 2011; ABABNEH; WADI; ISMAIL, 2013; CHANDAR; SUMATHI; SIVANANDAM, 2016; MORETTIN, 2014; TEIXEIRA, 2015; JOO; KIM, 2015; KHANDELWAL; ADHIKARI; VERMA, 2015) .

As funções  $\psi_{j,k}(t)$  dadas na Equação 3 podem formar bases que não precisam ser necessariamente ortogonais. Contudo, trabalhar com bases ortogonais possibilita a

<sup>1</sup> A função  $f(t)$  é do tipo quadrado integrável sobre um intervalo  $(a,b)$  se a integral definida  $\int_a^b |f(x)|^2 dx$  existe.

reconstrução perfeita da série temporal original, a partir dos coeficientes da transformada (SIDNEY, 1998; MORETTIN, 2014; TEIXEIRA, 2015; SOUZA et al., 2007). Daí vem o interesse em considerar na análise bases de funções que sejam ortogonais.

A fundamentação matemática da teoria que envolve as *Wavelets* está inserida no espaço de Hilbert ( $L^2(\mathbb{R})$ ).

Considere uma base ortornormal gerada por  $\psi_{j,k}(t)$ , ou seja,

$$\langle \psi_{j,k}, \psi_{l,m} \rangle = \delta_{j,l} \delta_{k,m}^2, \text{ com } j, k, l, m \in \mathbb{Z} \quad (4)$$

de tal modo que, para qualquer função  $f(t)$  de quadrado integrável sobre  $\mathbb{R}$ , tem-se que:

$$f(t) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} c_{j,k} \psi_{j,k}(t).^3 \quad (5)$$

Pode-se dizer que a Equação 5 é uma série de *Wavelets* da função  $f$  e os coeficientes *Wavelets*  $c_{j,k}$  são calculados por meio da seguinte equação:

$$c_{j,k} = \langle f, \psi_{j,k} \rangle = \int_{-\infty}^{+\infty} f(t) \psi_{j,k}(t) dt.^4 \quad (6)$$

Uma transformada ortogonal é concisa e cada coeficiente é calculado como produto interno da série temporal com a função base. Daí o interesse em considerar bases *Wavelets* que sejam ortogonais (MORETTIN, 2014; SOUZA et al., 2007).

Contudo, para que uma função matemática possa ser considerada uma *Wavelet*, ela deve satisfazer as seguintes propriedades (SIDNEY, 1998; MORETTIN, 2014):

P1.  $\int_{-\infty}^{\infty} \psi(t) dt = 0$ ;

Segundo Sidney (1998) esta propriedade, conhecida como condição de admissibilidade, permite que as *Wavelets* apresentem características oscilatórias ao longo do eixo temporal. Também garante a existência da Transformada *Wavelet* inversa (TWI) que possibilita a reconstrução da função em sua formatação original.

<sup>2</sup>  $\delta_{j,l}$  (com  $j, l \in \mathbb{Z}$ ) representa a função Delta de Kronecker.

<sup>3</sup> A convergência da função  $f(t)$  deve ser entendida em média quadrática.

<sup>4</sup> A relação de Parseval vale aqui, a saber:

$$\int_{-\infty}^{+\infty} f^2(t) dt = \sum_j \sum_k c_{j,k}^2.$$

Segundo Morettin (MORETTIN, 2014), a energia por unidade de tempo de  $f(t)$ , é decomposta em um número infinito de termos, sendo cada termo  $c_{j,k}$  a contribuição à potência, do termo da série temporal  $f(t)$ .

A condição de admissibilidade também pode ser expressa no domínio de Fourier como descrito na Equação 7:

$$C = \int_{-\infty}^{\infty} \frac{|\Psi(\omega)|^2}{|\omega|} dt < \infty, \quad (7)$$

de modo que  $\Psi(\omega)$  é o espectro da *Wavelet*-mãe  $\psi(t)$  (transformada de Fourier da função  $\psi(t)$ ) e  $\omega$  é a frequência angular. Deste modo, podemos observar que o parâmetro  $C$  necessita ser finito.

$$P2. \int_{-\infty}^{\infty} |\psi(t)| dt < \infty :$$

Também conhecida como condição de regularidade, esta propriedade garante que função *Wavelet* tenha sua energia localizada em uma certa região (Suporte compacto<sup>5</sup>).

$$P3. \int_{-\infty}^{\infty} |\psi(t)|^2 dt = 1 :$$

Esta propriedade, que por vezes é chamada de energia unitária, está relacionada com a condição de que as funções *Wavelets* devem estar localizadas em uma certa região finita no espaço. As propriedades *P2* e *P3* são as características que diferenciam a análise *Wavelet* da análise de Fourier (SIDNEY, 1998).

$$P4. \text{ Os primeiros } r - 1 \text{ momentos de } \psi \text{ anulam-se, isto é: } \int_{-\infty}^{\infty} t^j \psi(t) dt = 0, \text{ com } j = 0, 1, 2, \dots, r - 1, \text{ para algum } r \leq 1 \text{ e } \int_{-\infty}^{\infty} |t^r \psi(t)| dt < \infty$$

O parâmetro  $r$  (momentos nulos) está relacionado ao grau de suavidade (ou regularidade) da *Wavelet*-mãe  $\psi$ . Quanto maior o valor de  $r$ , mais suave será  $\psi$  (SIDNEY, 1998; MORETTIN, 2014). Desta forma, temos que uma *Wavelet*-mãe  $\psi(t)$  é uma função, tal que, a família de funções  $\psi_{a,b}(t)$  pode gerar uma base ortonormal para  $L^2(\mathbb{R})$ .

A descoberta de bases ortonormais na forma da Equação 3, e de suporte compacto, sintetizou a família de *Wavelets* e possibilitou uma análise mais eficiente do que a obtida com outros sistemas de funções (SIDNEY, 1998).

### 3.1.1.1 Função Escala

Para algumas escolhas especiais de  $\psi$ , as funções  $\psi_{a,b}$  que estão apresentadas na Equação 2 constituem uma base ortonormal para o espaço de funções do  $L^2(\mathbb{R})$ . Uma

<sup>5</sup> Uma *Wavelet* é dita de suporte compacto, se toda a energia dessa *Wavelet* está restrita a um intervalo finito, ou seja, se a função é exatamente zero fora desse intervalo.

maneira atraente de gerar as *Wavelets* com estas características é com a utilização da função escala  $\phi(t)$ , ou *Wavelet*-pai, que é a solução da seguinte equação:

$$\phi(t) = \sqrt{2} \sum_k l_k \phi(2t - k). \quad (8)$$

Segundo [Sidney \(1998\)](#), [Souza et al. \(2007\)](#) e [Morettin \(2014\)](#), esta função gera uma família ortonormal em  $L^2(\mathbb{R})$ ,

$$\phi_{j,k}(t) = 2^{\frac{j}{2}} \phi(2^j(t - k)), \text{ com } j, k \in \mathbb{Z}. \quad (9)$$

Nestas condições, a *Wavelet*-mãe  $\psi$  pode ser obtida a partir da função escala  $\phi$  por meio da seguinte equação:

$$\psi(t) = \sqrt{2} \sum_k h_k \phi(2t - k), \quad (10)$$

de modo que:

$$h_k = (-1)^k l_{k-1}. \quad (11)$$

Os coeficientes  $l_k$  e  $h_k$  são coeficientes de filtros digitais FIR (Filtros de resposta a impulso finita), que também são conhecidos como filtros passa-altas e filtros passa-baixas (frequências), respectivamente. Estes filtros são chamados *Quadrature mirror filters* e são utilizados para calcular a Transformada *Wavelet* discreta (TWD) ([SIDNEY, 1998](#); [VERMA; VERMA, 2012](#); [MORETTIN, 2014](#)).

Estes coeficientes são calculados por meio das seguintes expressões:

$$l_k = \sqrt{2} \int_{-\infty}^{\infty} \phi(t) \phi(2t - k) dt, \quad (12)$$

$$h_k = \sqrt{2} \int_{-\infty}^{\infty} \psi(t) \psi(2t - k) dt, \quad (13)$$

Contudo, considerando o sistema ortonormal  $\{\phi_{j,k}(t), \psi_{j,k}(t)\}$ , qualquer função  $f(t) \in L^2(\mathbb{R})$  pode ser escrita de acordo com a seguinte expressão ([SIDNEY, 1998](#); [MORETTIN, 2014](#); [SOUZA et al., 2007](#)):

$$f(t) = \sum_k cA_{j_0,k} \phi_{j_0,k}(t) + \sum_{j \geq j_0} \sum_k cD_{j,k} \psi_{j,k}(t), \quad (14)$$

de modo que  $cA$  e  $cD$  representam conjuntos de coeficientes *Wavelets* e  $j_0$  representa a escala mais suave.

O primeiro somatório apresentado na Equação 14 representa a tendência da série temporal, as componentes de baixas frequências, enquanto o segundo somatório representa os pormenores da série, os componentes de altas frequências (WADI; ISMAIL, 2011; ABABNEH; WADI; ISMAIL, 2013; TEIXEIRA, 2015). Desta forma, em cada nível de escala  $j$ , dois conjuntos são gerados: um conjunto composto por coeficientes *wavelet* ( $cA$ ), que chamados de coeficientes de aproximação, e outro conjunto composto por coeficientes *Wavelets* ( $cD$ ) que por sua vez são chamados coeficientes de detalhes .

### 3.1.1.2 Análise de Múltipla Resolução

Com a utilização da Função escala apresentada na Equação 8, e a definição de *Wavelet* dada na Equação 9, temos que os coeficientes  $cA_{j_0,k}$  e  $cD_{j,k}$  podem ser calculados por meio da seguinte expressão (SIDNEY, 1998; MORETTIN, 2014; SOUZA et al., 2007):

$$cA_{j_0,k} = \int_{-\infty}^{\infty} f(t)\phi_{j_0,k}(t)dt, \quad (15)$$

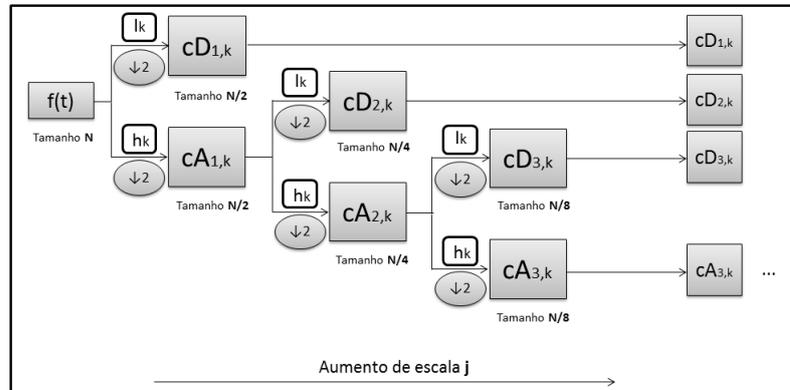
$$cD_{j,k} = \int_{-\infty}^{\infty} f(t)\psi_{j,k}(t)dt. \quad (16)$$

Além disso, as Equações 15 e 16 podem ser interpretadas como um processo de convolução seguido de uma decimação <sup>6</sup> (*downsampling*) por dois (notação:  $\downarrow 2$ ) que, ao ser aplicada em uma série temporal, as amostras pares (ou ímpares) são eliminadas. Logo, no nível  $j$ , tem-se metade dos coeficientes do nível  $j - 1$  (MALLAT, 1989).

Deste modo, com base na Equação 14, a cada nova escala de  $j$ , o coeficiente de aproximação  $cA_{j,k}$  passa por uma nova divisão  $l_k/h_k$ , de maneira a gerar novos coeficientes  $cA_{j+1,k}$  e  $cD_{j+1,k}$ , e assim sucessivamente. Logo, a cada novo nível  $j + 1$ , tem-se metade dos coeficientes do nível anterior  $j$ . Este procedimento é ilustrado no terceiro nível de resolução na Figura 1. Ele é conhecido como Algoritmo de decomposição em árvore, Algoritmo de decomposição de Mallat ou, ainda, Algoritmo piramidal (MALLAT, 1989; SIDNEY, 1998; MORETTIN, 2014; SOUZA et al., 2007).

<sup>6</sup> Representa uma operação de sub-amostragem, que consiste em eliminar todos os coeficientes de índice par de uma sequência. A sub-amostragem garante que o número total de pontos permaneça constante após a TWD (SIDNEY, 1998).

Figura 1 – Algoritmo de decomposição de Mallat.



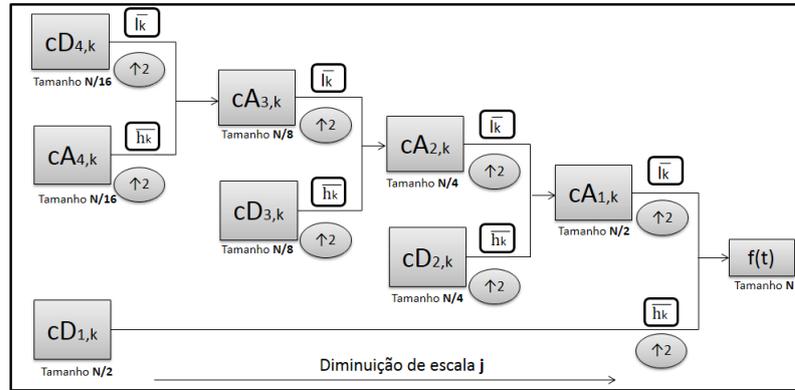
Fonte: Próprio autor.

Os coeficientes  $l_k$  e  $h_k$ , que podem ser observados nas Equações 8 e 9, são utilizados para calcular a TWD e separar a série temporal em componentes de altas frequências (que podem ser considerados detalhes e ruídos) e em componentes de baixas frequências (em vários níveis de resolução). Em virtude disto, é oriundo desta ideia o nome Análise de Multipla Resolução (AMR) (SIDNEY, 1998; SOUZA et al., 2007).

Segundo Morettin (2014) a Análise de Múltipla Resolução (AMR) permite analisar a série temporal em várias escalas de resolução, permitindo assim estudar o comportamento dos dados em várias magnificações.

A Transformada *Wavelet* inversa (TWI) essencialmente realiza as mesmas operações em uma direção contrária. Aos mesmos coeficientes são aplicados a TWI em ordem reversa utilizando o algoritmo piramidal de reconstrução. Desta forma, ao invés de ocorrer um procedimento de decimação, as séries temporais são interpoladas. Desta forma caracteriza-se um *upsampling* por dois (notação:  $\uparrow 2$ ), ou seja, zeros são colocados entre os coeficientes da transformada. (MALLAT, 1989). A representação da reconstrução da série de dados da função  $f$  por meio da estrutura de banco de filtros FIR no 4º nível de resolução é ilustrado na Figura 2.

Figura 2 – Algoritmo de reconstrução de Mallat.



Fonte: Próprio autor.

A relação entre os níveis de resolução pode ser descrita sob a forma de um aninhamento dos espaços de funções, onde cada espaço de maior resolução contém espaços de menor resolução.

De acordo com [Morettin \(2014\)](#), AMR é uma sequência crescente de subespaços fechados  $\{V_j, \text{com } j \in \mathbb{Z}\}$  que aproximam  $L^2(\mathbb{R})$ , e possui as seguintes propriedades:

$$(M1.) \{0\} \dots \subset V_{-1} \subset V_0 \subset V_1 \subset \dots ;$$

$$(M2.) L^2(\mathbb{R}) = \overline{\bigcup_j V_j};$$

$$(M3.) \bigcap_j V_j = \{0\}$$

As propriedades  $M1$ ,  $M2$  e  $M3$  garantem que a sequência de subespaços  $(V_j)_{j \in \mathbb{Z}}$  não é vazia, pois a função nula está contida em cada espaço da sequência.

$$(M4.) f(t) \in V_j \Leftrightarrow f(2t) \in V_{j+1}, \forall j.$$

Na propriedade  $M4$  é descrito que a relação de inclusão entre os subespaços  $V_j$  é uma relação de auto-similaridade toda vez que uma função  $f$  for reescalada por um fator 2.

$$(M5.) V_{j+1} = V_j \oplus W_j .$$

Já nesta propriedade é indicado que as translações da função escala formam uma base ortonormal para o espaço  $V_j$ , com  $W_j$  sendo o complemento ortogonal de  $V_j$  em  $V_{j+1}$ .

Deste modo, considerando uma função  $f \in L^2(\mathbb{R})$ , a ideia aqui é obter aproximações da função  $f$  em vários níveis de resolução. Cada subespaço fechado  $V_j$  será constituído de funções aproximantes, sendo que a melhor aproximação é obtida considerando-se a projeção ortogonal de  $f$ , ( $\hat{f}(t)$ ), sobre cada  $V_j$  (MORETTIN, 2014). Desta forma,  $\forall g \in V_j$ , temos:

$$\|g(t) - f(t)\| \geq \|\hat{f}(t) - f(t)\|. \quad (17)$$

Assim, de acordo com a propriedade (M1), ao passar do nível de resolução  $j$  (ou escala  $2^j$ ) para o nível  $j + 1$ , ganha-se informação no momento em que adiciona-se detalhe. Na medida em que o nível de resolução aumenta, ( $j \rightarrow \infty$ ), de acordo com a propriedade (M2), a função aproximada converge para  $f(t)$ . Por outro lado, se  $j \rightarrow \infty$ , temos por (M3) que  $\hat{f}(t) \rightarrow f(t)$ . Como pode ser visto na propriedade (M4), o espaço  $V_{j+1}$  é obtido de  $V_j$  escalando-se as funções aproximadoras pela razão dos respectivos níveis de resolução. E, de acordo com a propriedade (M5), a informação perdida quando diminuimos o nível de resolução ( $V_{j+1}$  para  $V_j$ ) pode ser representado pelo espaço  $W_j$  (SIDNEY, 1998; MORETTIN, 2014; SOUZA et al., 2007).

### 3.1.2 Estrutura da Filtragem de Séries Temporais via Wavelets

A capacidade de representar uma série temporal no domínio tempo-frequência, possibilitou a TW ser uma técnica eficiente no pré-processamento em séries temporais (VALENCIA et al., 2016).

O trabalho mais referenciado que envolve filtragem via *Wavelets* foi apresentado por Donoho (1995). Ele é baseado na limiarização dos coeficientes *Wavelets* de uma série temporal. O fato da AMR fornecer a decomposição da série temporal em várias escalas é o fator primordial para a adequabilidade do procedimento. A maior parte do ruído é representada pelos coeficientes *Wavelets* nas escalas mais finas e, descartando estes coeficientes, o resultado é uma filtragem natural do ruído (SIDNEY, 1998; MUSTAFA; ALZUBAYDI, 2013; JOY; PETER; JOHN, 2013; MORETTIN, 2014; VALENCIA et al., 2016; AL-QAZZAZ et al., 2014; ALI; EL-DAHSHAN; YAHIA, 2017; VERMA; VERMA, 2012).

De acordo com Morettin (2014), o procedimento de limiarização dos coeficientes *Wavelets* corresponde a uma operação de suavização. A TW preserva a energia, a soma dos quadrados dos dados é igual à soma dos quadrados dos coeficientes das *Wavelets*, mas neste caso, a energia fica concentrada em poucos coeficientes. Desta forma são necessários poucos coeficientes para descrever a função de interesse. Por outro lado, o ruído branco é invariante por uma transformada ortogonal e, passar para o domínio *Wavelet*, não o afeta.

Conforme definição dada na Equação 1, o comportamento fundamental ( $\mu_t$ ) de uma série temporal que foi corrompida por ruído branco pode ser encontrado com a utilização da DW por meio da realização de três principais etapas (ALI; EL-DAHSHAN; YAHIA, 2017; VALENCIA et al., 2016; VERMA; VERMA, 2012; DONOHO, 1995):

1. **Decomposição:** A *Wavelet*-mãe e o nível  $j$  de decomposição são determinados. Em seguida a decomposição da série temporal por meio da TW é realizada;
2. **Limiarização:** É realizada a estimação do limiar e a limiarização dos coeficientes *Wavelets*;
3. **Reconstrução:** A aplicação da TWI no mesmo nível  $j$  para a reconstrução da série temporal filtrada.

A segunda etapa do procedimento é parte fundamental no processo de filtragem devido ao ruído estar representado nos coeficientes *Wavelets* da decomposição. O método é baseado em descartar os coeficientes que representam o ruído na série (MORETTIN, 2014; VALENCIA et al., 2016; AL-QAZZAZ et al., 2014; ALI; EL-DAHSHAN; YAHIA, 2017; VERMA; VERMA, 2012) .

Segundo Donoho (1995), os procedimentos de filtragem *Wavelet* que promovem a limiarização dos coeficientes *Wavelets*  $c_{j,k}$  são o *Hard* e *Soft thresholdings*.

Apesar de atualmente existirem outros métodos de limiarização (truncamento) de coeficientes *Wavelets*, os métodos de limiarização *Hard* e *Soft thresholdings* são os mais utilizados na literatura da área e, em virtude disto, foram analisados nesta pesquisa (JOY; PETER; JOHN, 2013; MORETTIN, 2014; VALENCIA et al., 2016; ALI; EL-DAHSHAN; YAHIA, 2017; VERMA; VERMA, 2012). Suas formulações matemáticas são apresentadas nas Equações 18 e 19, respectivamente.

$$thr_{\lambda}^H(t) = \begin{cases} 0, & \text{se } |t| \leq \lambda, \\ t, & \text{se } |t| > \lambda. \end{cases} \quad (18)$$

$$thr_{\lambda}^S(t) = \begin{cases} 0, & \text{se } |t| \leq \lambda, \\ y_t(t)(|t| - \lambda), & \text{se } |t| > \lambda. \end{cases} \quad (19)$$

De modo que  $y_t(t)$  é a série temporal no tempo  $t$ .

O *Hard Thresholding* ( $thr_{\lambda}^H$ ) trunca os coeficientes que são menores ou iguais a um valor específico de limiar  $\lambda$  ajustando-os para zero, enquanto que o *Soft Thresholding* ( $thr_{\lambda}^S$ ) ajusta para zero os coeficientes com valor igual ou abaixo do limiar  $\lambda$  bem como

suaviza (pondera as observações) os outros coeficientes que estão acima do valor deste limiar.

A seleção do limiar é um procedimento importante afetando diretamente a qualidade da série temporal filtrada. Portanto a escolha entre *hard* ou *soft* deve ser feita caso a caso, dependendo da aplicação e, portanto, dos dados (MORETTIN, 2014; VALENCIA et al., 2016; JOY; PETER; JOHN, 2013).

Outro aspecto imprescindível é a estimação do limiar  $\lambda$ , esta é outra questão fundamental neste processo. Existem quatro critérios padrões de estimação deste limiar: *minimaxi*, *sqtwlog*, *rigrsure* e *heursure*. Eles são definidos da seguinte forma (JOY; PETER; JOHN, 2013; VERMA; VERMA, 2012; VALENCIA et al., 2016; AL-QAZZAZ et al., 2014; ALI; EL-DAHSHAN; YAHIA, 2017; DONOHO, 1995):

- **Critério Minimax:** A regra de estimação do limiar que se baseia no princípio Minimax. Este critério utiliza um limiar  $\lambda$  fixo que é calculado por meio da seguinte equação:

$$\lambda = \begin{cases} \hat{\sigma}(0,3936 + 0,1829 \cdot \log_2 N), & \text{se } N \geq 32, \\ 0, & \text{se } N < 32, \end{cases} \quad (20)$$

com  $\hat{\sigma} = \text{mediana} \left( \frac{|\omega_1|}{0,6745}, \frac{|\omega_2|}{0,6745}, \frac{|\omega_3|}{0,6745}, \dots, \frac{|\omega_j|}{0,6745} \right)$  sendo o estimador do desvio padrão ( $\sigma$ ) do ruído,  $\omega_j$  o vetor de coeficientes *Wavelets* para cada escala  $j$  e  $N$  o tamanho da série temporal (DONOHO, 1995);

- **Critério Sqtwolog:** O valor do limiar é calculado pelo método do limiar universal (logaritmo da raiz quadrada) que é dado por:

$$\lambda_j = \hat{\sigma}_j \sqrt{2 \log(N_j)}, \quad (21)$$

com  $N_j$  o comprimento da série ruidosa na escala  $j$  e  $\hat{\sigma}_j$  é o estimador do desvio padrão do ruído no domínio *Wavelet* em cada escala  $j$  que é desconhecido e pode ser estimado pelo desvio mediano absoluto (*MAD*) na escala  $j$  que está descrito na Equação 22:

$$\hat{\sigma}_j = \text{mediana} \left( \frac{|\omega_1|}{0,6745}, \frac{|\omega_2|}{0,6745}, \frac{|\omega_3|}{0,6745}, \dots, \frac{|\omega_j|}{0,6745} \right); \quad (22)$$

- **Critério Rigrsure:** Também conhecido por *SureShrink*, possibilita que o limiar  $\lambda$  seja estimado de forma a minimizar o risco não viesado de Stein para cada coeficiente *Wavelet*. Para isto considere o vetor de risco  $\Omega = [\Omega_1, \Omega_2, \Omega_3, \dots, \Omega_N]$ , o vetor que

contém o quadrado dos valores dos coeficientes *Wavelet* do menor para o maior nível. Desta maneira, o limiar  $\lambda$  é encontrado para o nível  $j$  por meio da seguinte expressão:

$$\lambda_j = \hat{\sigma}_j \sqrt{\varphi_b}, \quad (23)$$

com  $\varphi_b$  sendo o  $b$ -ésimo coeficiente *Wavelet* associado ao menor valor  $\Omega_b$  (com  $b \in [1, N]$ ) do vetor de risco  $\Omega$ .

- **Critério Heursure:** É um critério de seleção de limiar que utiliza a combinação dos critérios Sqtwlog e Rigrsure. É Dependente da relação existente entre a série temporal e o ruído. Se esta relação é pequena o método de estimação Rigrsure fornece um estimador ruim quando comparado ao estimador fornecido pelo critério Sqtwlog. Sendo os limiares  $\lambda_1$  e  $\lambda_2$  obtidos pelos critérios Sqtwlog e Rigrsure, respectivamente, o limiar pelo critério Heursure é dado por:

$$\lambda = \begin{cases} \lambda_1, & \text{se } A > B, \\ \min(\lambda_1, \lambda_2), & \text{se } A \leq B. \end{cases} \quad (24)$$

De modo que  $A = \frac{s-N}{N}$  e  $B = \log_2 N^{\frac{3}{2}} \sqrt{N}$ , com  $N$  sendo o comprimento do vetor de coeficientes *Wavelets* e  $s = \sum_{i=1}^N \omega_i^2$ .

### 3.1.3 Redes Neurais Artificiais - RNAs

De acordo com [Haykin \(2007\)](#) Redes Neurais Artificiais (RNAs) são um sistema de computação paralelo formado por unidades de processamento idênticas e interligadas que são chamadas neurônios (ou nodos) artificiais. Estes neurônios são distribuídos em uma ou mais camadas e possuem unidades capazes de armazenar conhecimento com a possibilidade de disponibilizá-lo no futuro.

São técnicas computacionais que apresentam um modelo matemático inspirado na estrutura neural de organismos inteligentes que adquirem conhecimento por meio de experiências e possuem amplo alcance na modelagem de problemas com características não lineares ([HAYKIN, 2007](#); [KHASHEI; BIJARI, 2011](#); [FERREIRA, 2006](#)).

Dentre suas principais propriedades, vale-se destacar sua capacidade de aprendizagem, adaptabilidade, robustez, capacidade de armazenamento e conhecimento adquirido e seu poder de generalização ([HAYKIN, 2007](#)):

As RNAs foram idealizadas com base no processo de aprendizagem do cérebro humano, de forma que uma RNA tem a capacidade de simular as conexões neurais que também são conhecidas como conexões sinápticas. Cada neurônio de uma RNA é

condicionado a receber um sinal que é ponderado pelo seu respectivo peso na conexão de entrada e é processado por uma função matemática, chamada função de transferência (ou ativação). Após esta etapa, um novo sinal de saída é produzido e propagado pela rede (HAYKIN, 2007).

A função de transferência  $g(\cdot)$  realiza um mapeamento não linear das observações passadas  $y_{t-k}$  (em que  $k = 0, 1, \dots, L$ ) com o objetivo de construir as previsões  $y_{t+1}$ . Os três tipos básicos de função de transferência são: Limiar, Linear e Sigmóidal. A escolha depende do objetivo da análise. Para fins de previsão a função sigmoidal, como definida na Equação 25, é comumente utilizada na construção de RNAs (FERREIRA, 2006).

$$g(x) = \frac{1}{1 + \exp(-\lambda x)} \quad (25)$$

De modo que:

- $\lambda$  é o parâmetro que determina a sensibilidade na resposta da função;
- $\exp(\cdot)$  é a função exponencial que possui como base o número de Euler;
- $x$  representa os dados de entrada na rede.

A função Sigmóide logística é definida no intervalo de tempo  $[0,1]$ . Em virtude disto todos os dados que alimentarão a rede devem ser escalonados neste intervalo (FERREIRA, 2006). Cada dado escalonado  $x_n$  é encontrado por meio de uma transformação linear que é apresentada na Equação 26.

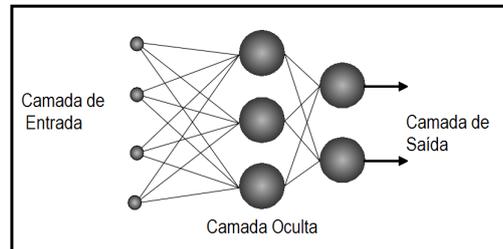
$$x_t = \frac{y_t - y_{\min}}{y_{\max} - y_{\min}} \quad (26)$$

De modo que  $y_{\min}$  representa o menor valor e  $y_{\max}$  representa o maior valor da série temporal  $\{y_t\}_{t=1,2,\dots,N}$ .

### 3.1.3.1 Redes Neurais *Multilayer Perceptron*

As RNAs do tipo *Multilayer Perceptron* (MLP) são formadas por várias camadas (camadas de entrada, camadas intermediárias ou ocultas e camadas de saída) que são interligadas por conexões sinápticas. A primeira camada corresponde às variáveis de entrada do problema. A segunda camada é a camada escondida ela é utilizada para capturar as relações não lineares existentes nos dados. Já a terceira camada é usada para fornecer o valor de previsão (FERREIRA, 2006). Na Figura 3 a seguir é apresentado um exemplo de uma RNA de 2 camadas com 4 entradas e 2 saídas.

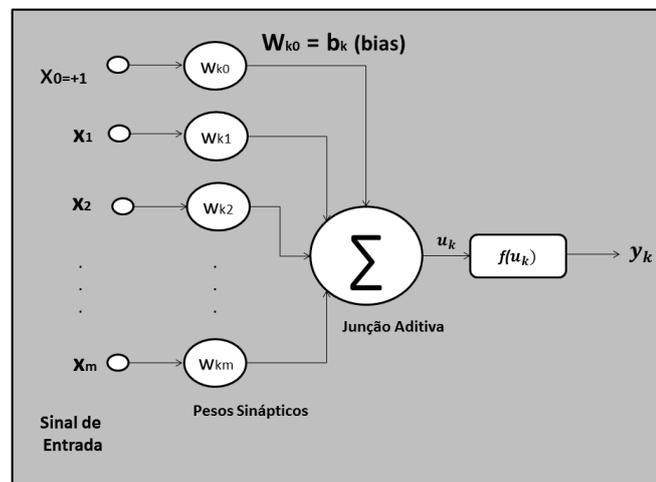
Figura 3 – Rede Neural MLP alimentada com múltiplas camadas.



Fonte: Próprio autor.

Segundo Haykin (2007) a série temporal é propagada da camada de entrada para a camada de saída. A ligação entre os neurônios e suas camadas ocorrem por meio de parâmetros numericamente ajustáveis, chamados pesos sinápticos, que são as ponderações entre as conexões e representam a "força" de cada conexão. A soma das entradas ponderadas são transformadas por meio da função de transferência (HAYKIN, 2007; FERREIRA, 2006). Este processo é ilustrado na Figura 4.

Figura 4 – Arquitetura de um Rede Neural MLP.



Fonte: Próprio autor.

A representação matemática da saída do neurônio  $k$  que foi ilustrada na Figura 4 é descrita na Equação 27.

$$y_k = f(u_k) = f\left(\sum_{j=1}^m w_{kj}x_j + b_k\right) \quad (27)$$

O limiar  $b_k$  tem o papel de aumentar ou diminuir a influência do valor de entrada líquida para função de ativação do neurônio  $k$  (HAYKIN, 2007).

Segundo FERREIRA (2006) uma das propriedades de suma importância nas RNAs é a habilidade de aprender por meio de um conjunto exemplo dado para treinamento que extrai características necessárias para representar a informação e, com isto, melhorar seu desempenho. Isto é feito por meio de um processo iterativo que realiza ajustes em seus pesos sinápticos para melhorar a generalização da informação obtida no conjunto de exemplo comumente chamado de conjunto de treino.

De maneira geral, a maneira pela qual uma RNA do tipo MLP se relaciona com o ambiente possibilita os seguintes paradigmas de aprendizado: O supervisionado e o não supervisionado. Uma rede neural MLP é treinada de forma supervisionada quando um padrão estabelecido na saída da rede é direcionado à sua respectiva entrada, forçando assim a rede a ajustar os pesos sinápticos. O aprendizado neste tipo de rede é geralmente feito por meio de um algoritmo de retro-propagação do erro (KIŞI; UNCUOĞLU, 2005). O processo iterativo é concluído quando a RNA atinge uma solução desejada por meio de um critério de parada (frequentemente um erro mínimo desejável) para uma classe de problemas (HAYKIN, 2007; FERREIRA, 2006). Em cada conexão sináptica um neurônio tem um fator de peso reajustável por um processo iterativo associado a ele por meio do algoritmo de treinamento que conduz todo o procedimento.

Existem muitos tipos de algoritmos de aprendizagem para diferentes modelos de RNAs, estes algoritmos diferem entre si principalmente pelo modo como os pesos sinápticos são modificados. Os algoritmos de treinamento do tipo gradiente descendente *LM-Levenberg Marquard* e *Rprop-Resiliente Propagation*, produzem bons resultados e por isto são os mais populares (HAYKIN, 2007; KIŞI; UNCUOĞLU, 2005; FERREIRA, 2006).

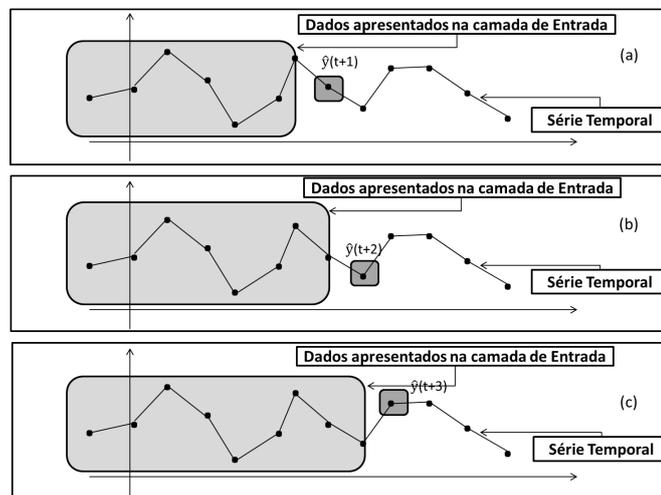
A série temporal que é apresentada à RNA é dividida em três partes caracterizadas por um conjunto de treino, validação e teste. O primeiro conjunto é utilizado na fase de aprendizagem da rede para treinar o modelo, o conjunto de validação é utilizado para validar o treinamento realizado e o terceiro conjunto constitui o teste do modelo. Não existe regra definitiva para dividir a série temporal nestes três conjuntos mais sim um a padronização que segue o relatório técnico *Proben1* (PRECHELT et al., 1994; FERREIRA, 2006). Segundo esta padronização, os dados devem ser divididos em, 50%, 25% e 25% para os conjuntos de treino, validação e teste, respectivamente (PRECHELT et al., 1994). Nesta Tese, além das análises seguindo esta padronização, analisou-se também para a divisão 70%, 15% e 15%, seguindo a mesma ordem nos conjuntos de dados.

### 3.1.3.2 Previsão de séries temporais via RNAs

A previsão de séries temporais MLPs consiste em analisar as  $N$  observações da série temporal  $\{y_t\}_{t=1,2,\dots,N}$  para prever  $y_{N+h}$  observações, sendo  $h$  o horizonte de previsão.

Contudo, pesquisadores utilizam  $n_1$  observações ( $n_1 < N$ ) da séries temporal  $y_t$  para obter  $h$  previsões ( $n_1 + h = N$ ) e realizar a análise da exatidão do modelo de previsão (ZHANG, 2003; KHASHEI; BIJARI, 2011; KHANDELWAL; ADHIKARI; VERMA, 2015; BABU; REDDY, 2014; WADI; ISMAIL, 2011). Este processo iterativo está ilustrado na Figura 5 para  $h = 1, h = 2$  e  $h = 3$ , nos casos (a), (b) e (c), respectivamente. Nelas pode-se observar que é necessário conter alguns dados da série para realizar o próximo passo da previsão ( $\hat{y}_{(n_1+h)}$ ).

Figura 5 – Esquema da previsão da série temporal pelo procedimento. (a) Horizonte de previsão  $h = 1$ , (b) horizonte de previsão  $h = 2$  e (c) Horizonte de previsão  $h = 3$ .



Fonte: Próprio autor.

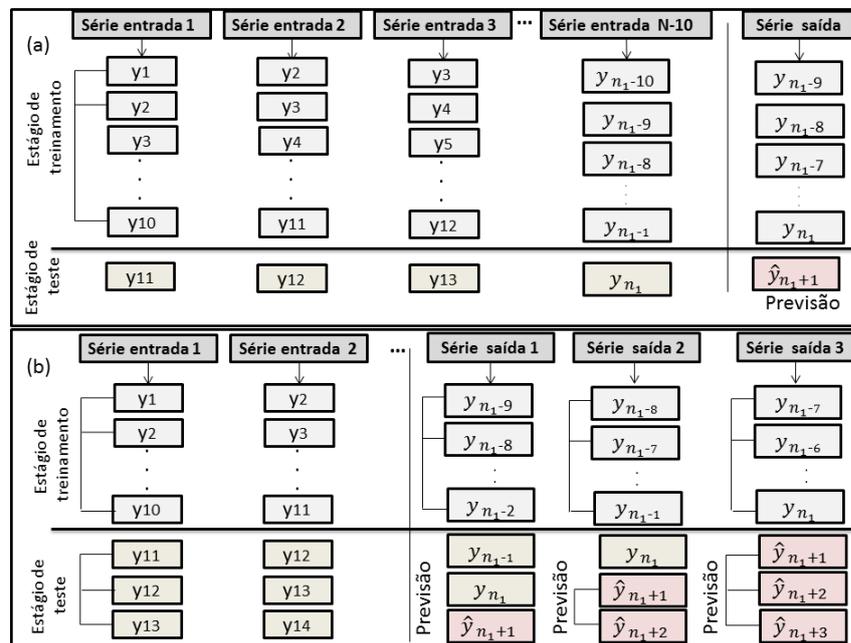
A RNA necessita de uma quantidade  $n$  de dados na camada de entrada da rede correspondente aos retardos temporais relevantes (*lags*) que possibilitam o estabelecimento de  $n_1 - n$  padrões de entrada/saída e, assim, a determinação do próximo valor futuro da série (FERREIRA, 2006; LIU et al., 2013; HAYKIN, 2007).

Nos casos em que o horizonte de previsão é maior que um, a previsão da série temporal pode ser realizadas sob duas diretrizes de previsão: Um passo à frente (*one-step-ahead*) ou Múltiplos passos à frente (*Multi-step-ahead*) (BABU; REDDY, 2014; BABU; SURE, 2016).

No procedimento *one-step-ahead*, deve-se possuir a quantidade de dados  $N = n_1 + h$  da série temporal para alimentar o processo iterativo com os dados reais. Isto é, para fazer a previsão com horizonte  $h$  de uma série temporal de tamanho  $N$  é necesssário inicialmente apresentar como dados de entrada a série com comprimento  $n_1$  para prever cada valor  $\hat{y}_{n_1+h}$ . Este processo está ilustrado na Figura 5 para os horizontes de previsão  $h = 3$ .

Uma vez em que o *lag*  $n$  e o horizonte de previsão  $h$  estão definidos, a RNA constrói internamente uma sequência  $y_{n_1}, y_{n_1-1}, y_{n_1-2}, y_{n_1-n+1}$  para a camada de entrada da rede com o objetivo de se obter  $y_{n_1+1}, \dots, y_{n_1+h}$ . O caso particular onde o número de *lags* é 10 para os procedimentos *one-step-ahead*, está ilustrado na Figura 6(a), e *multi-step-ahead*, na Figura 6(b). Neste exemplo, temos  $n_1 = 10$  padrões de entrada/saída. O horizonte de previsão é  $h = 1$ , no caso (a), e  $h = 3$  no caso (b).

Figura 6 – Esquema de formatação de dados entrada/saída na RNA. (a) Caso *one-step-ahead*, (b) caso *multi-step-ahead*

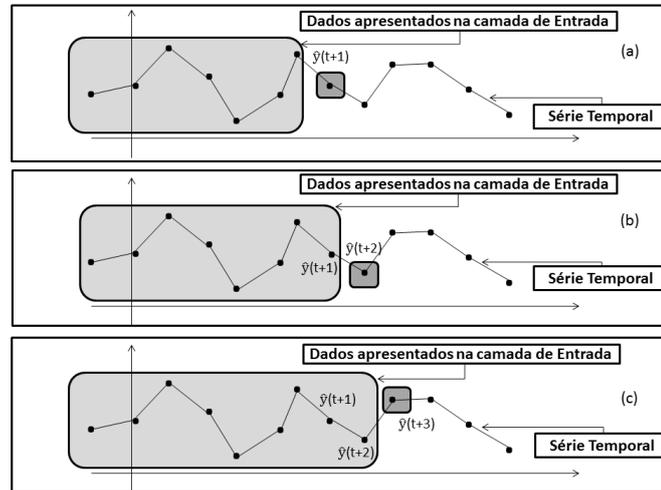


Fonte: Próprio autor.

No caso geral, os dados  $y_{n_1+1}, y_{n_1+2}, \dots, y_{n_1+h-1}$  alimentarão a série de entrada em cada uma etapa do processo iterativo que resultará nos valores previstos  $\hat{y}_{n_1+2}, \hat{y}_{n_1+3}, \dots, \hat{y}_{n_1+h}$  e servirão como parâmetros para a análise da exatidão do modelo por meio do erro da previsão  $e_h = |y_{n_1+h} - \hat{y}_{n_1+h}|$ .

De maneira semelhante ao caso anterior, a previsão *multi-step-ahead* possui o diferencial da substituição dos valores da série temporal pelos valores previstos no modelo para realizar a previsão do próximo valor. Os dados  $\hat{y}_{n_1+1}, \hat{y}_{n_1+2}, \dots, \hat{y}_{n_1+h-1}$  alimentarão a série de entrada em cada uma etapa do processo iterativo que resultará nos valores previstos  $\hat{y}_{n_1+2}, \hat{y}_{n_1+3}, \dots, \hat{y}_{n_1+h}$ . Este processo iterativo está ilustrado na Figura 7 para  $h = 1$  no caso (a),  $h = 2$  no caso (b) e  $h = 3$  no caso (c).

Figura 7 – Esquema da previsão da série temporal para três procedimentos *multi-step-ahead*. (a) Horizonte de previsão  $h = 1$ , (b) horizonte de previsão  $h = 2$  e (c) Horizonte de previsão  $h = 3$ .



Fonte: Próprio autor

O modelo matemático que envolve a saída desejada  $y_{t+1}$  e as entradas  $y_{t-n+1}, \dots, y_{t-2}, y_{t-1}, y_t$  é apresentada na Equação 28.

$$y_{t+1} = \alpha_0 + \sum_{j=1}^q \alpha_j g(\beta_{0j} + \sum_{i=1}^n \beta_{ij} y_{t-i+1}) + e_{t+1} \quad (28)$$

de modo que  $\alpha_0$  e  $\beta_{i,j}$ , com  $j = 0, 1, \dots, q$  e  $i = 0, 1, \dots, n$  são os parâmetros das RNAs (Pesos sinápticos),  $n$  é a quantidade de neurônios na camada de entrada (*lags*),  $q$  é a quantidade de neurônios ocultos na camada intermediária,  $e_{t+1}$  é o erro de aproximação no tempo  $t + 1$ , e  $g(\cdot)$  a função de transferência.

### 3.1.3.3 O problema da previsão de séries temporais via RNAs

O que pode ser o fator primordial para o fracasso do modelo é o fato de que os dados são não evidentes, são frequentemente marcados por ruído e isto dificulta a obtenção da correta seleção dos parâmetro do modelo de MLPs (FERREIRA, 2006).

A correta escolha do número de *lags* é uma tarefa complexa e imprescindível para a fiel representação do comportamento fundamental da série temporal pelo modelo de RNAs. Outra questão não menos importante é a determinação da quantidade de neurônios na camada oculta. É nesta etapa que as características não lineares da série temporal são capturadas, contudo este procedimento deve ser realizado com cautela pois uma quantidade grande de neurônios nesta camada pode proporcionar problemas de superajustamento (HAYKIN, 2007; FERREIRA, 2006; FERREIRA; VASCONCELOS; ADEODATO, 2008;

FIRMINO; DE MATTOS NETO; FERREIRA, 2015; DE MATTOS NETO et al., 2014; DE MATTOS NETO; FERREIRA, 2016).

Segundo Ferreira, Vasconcelos e Adeodato (2008) a abordagem alternativa que combinam Algoritmo Genético com RNAs possibilita uma forma diferenciada para encontrar os parâmetros necessários para representar a informação subjacente que gera as séries temporais. Esta abordagem, intitulada *Time-delay Added Evolutionary Forecasting*(TAEF), está descrita na Subseção 3.1.5.

Existem muitos modelos de RNAs propostos na literatura, os modelos *Support Vector Machines*(SVM) e *Radial Basis Function*(RBF) são alguns exemplos (HAYKIN, 2007). Nesta Tese a RNA utilizada foi a do tipo MLP em virtude de sua praticidade, flexibilidade, adaptabilidade e eficiência em modelar séries temporais com características não lineares para fins de previsão.

As RNAs são uma metodologia que é aplicada fortemente em modelagem de séries temporais não lineares. Foi efetivamente empregada em pesquisas realizadas nos últimos anos, como por exemplo nos artigos de Ferreira, Vasconcelos e Adeodato (2008), Khashei e Bijari (2011), Babu e Reddy (2014), Chandar, Sumathi e Sivanandam (2016), Oliveira e Ludermir (2016) e Rodriguez e Barba (2016).

### 3.1.4 Algoritmos Genéticos - AGs

Algoritmos genéticos (AGs) são métodos de IA que utilizam a metáfora da teoria evolução das espécies, em uma perspectiva de simulação destes processos, com o objetivo de solucionar problemas de busca e otimização (HOLLAND, 1992; FERREIRA, 2006). Heurísticas inspiradas na Teoria da evolução são iterativamente aplicadas, permitindo a população evoluir para uma solução aceitável do problema (HOLLAND, 1992).

A ideia da teoria da evolução, de maneira simplista, pode ser encarada como uma técnica de otimização que é realizada de maneira a transferir informações genéticas dos pais para os filhos (HOLLAND, 1992; FERREIRA, 2006).

De forma geral, uma técnica de otimização apresenta (FERREIRA, 2006):

- i. **Espaço de Busca:** Espaço onde estão localizados todas as soluções do problema;
- ii. **Função Objetivo:** Função que avalia todas as soluções produzidas, atribuindo a cada uma delas um valor numérico que quantifica sua qualidade. A melhor hipótese é determinada por meio de uma medida numérica pré-determinada chamada *fitness*.

Como procedimento de otimização, os AGs realizam uma busca pela melhor solução para o problema partindo de um conjunto de estados iniciais que são formados por possíveis soluções do problema. Cada uma destas possíveis soluções são chamadas de *indivíduos* e o grupo de indivíduos de *população*.

A função que determina o *fitness* gera um índice de ordenamento por aptidão para todos os indivíduos da população. Quando a situação requer a minimização de um parâmetro, como um erro (caso dos problemas de previsão), a representação da função *fitness* pode ser dada (FERREIRA; VASCONCELOS; ADEODATO, 2008):

$$fitness = \frac{1}{1 + erro} \quad . \quad (29)$$

Os indivíduos de uma população são sujeitos a uma série de operações, tais como *seleção*, *cruzamento* e *mutação*. Como resultado deste processo, uma nova geração de indivíduos é criada e este processo é repetido até que um dado objetivo seja atingido (HOLLAND, 1992; FERREIRA, 2006; FERREIRA; VASCONCELOS; ADEODATO, 2008).

Existe um grande número de diferentes implementações dos AGS. De maneira geral, os AGs são implementados por meio do seguinte esquema (FERREIRA, 2006)

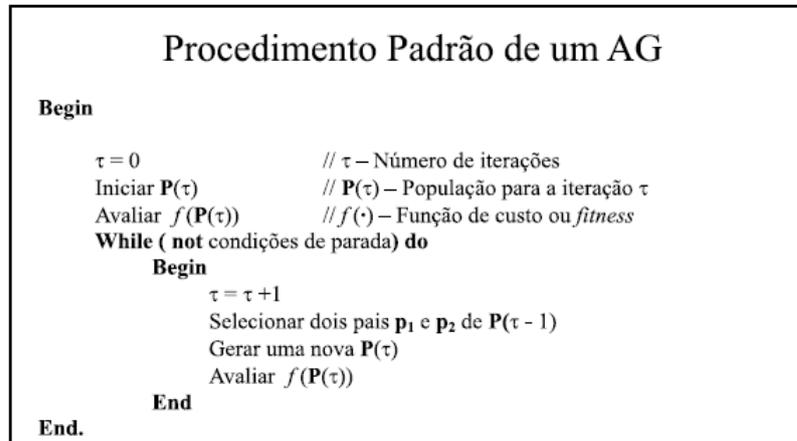
1. Os AGs operam interativamente sobre um conjunto de soluções (indivíduos) que formam a população;
2. Em cada interação do processo, todos os indivíduos são avaliados por meio da função *fitness*;
3. Uma nova população é então gerada por meio de um procedimento probabilístico realizado a partir da população atual.

Dado um AG, a caracterização matemática da evolução temporal de uma população é possível com o uso do Teorema dos Esquemas (HOLLAND, 1992; FERREIRA, 2006). Este teorema é baseado no conceito dos padrões que descrevem um conjunto de *strings* de bits.

Com a utilização da representação dos indivíduos pelos esquemas, o Teorema dos esquemas caracteriza a evolução da população do AG em termos de números de instâncias que representam cada esquema. Tal teorema está comentado no Anexo A.

Um algoritmo que ilustra a representação de um genérico AG é apresentado na Figura 8.

Figura 8 – Esquema padrão do Algoritmo genético.



Fonte: Ferreira (FERREIRA, 2006)

### 3.1.5 Time-delay Added Evolutionary Forecasting - TAEF

O método TAEF (*Time-delay Added Evolutionary Forecasting*) é uma metodologia inspirada no teorema de Takens (TAKENS, 1981) que consiste em um modelo híbrido inteligente composto por uma RNA do tipo MLP combinada com AGM (FERREIRA; VASCONCELOS; ADEODATO, 2008).

Inicialmente, o método TAEF procura encontrar o modelo mais adequado para prever a série temporal e, em seguida, realiza teste de hipóteses, a fim de ajustar as distorções de fase que podem aparecer na representação de algumas séries temporais.

O TAEF é um método para previsão de séries temporais que realiza uma busca evolutiva, por meio de AGMs em duas etapas. Na Primeira etapa os AGMs são utilizados para encontrar os parâmetros de RNA, esta etapa fase é considerada a etapa de otimização de parâmetros. A busca pela dimensionalidade mínima necessária para reprodução do fenômeno gerador da séries e de seus valores subsequentes é realizada na etapa de ajustamento de fase (FERREIRA; VASCONCELOS; ADEODATO, 2008; DE MATTOS NETO et al., 2014).

Seguindo este princípio, os parâmetros da RNA (arquitetura) definidos pelo algoritmo na primeira etapa do TAEF são:

1. O número de *Lags*;
2. O número de neurônios na camada oculta;
3. O algoritmo de treinamento.

Após a melhor arquitetura da RNA encontrada, o TAEF avalia se a RNA obtida possui desempenho melhor na predição do que um modelo *Random walk* (FERREIRA; VASCONCELOS; ADEODATO, 2008; DE MATTOS NETO et al., 2014).

O método TAEF analisa o ajuste do modelo a partir de quatro diferentes algoritmos de treinamento. Além dos algoritmos de treinamento comentados na Subseção 3.1.3 ( *LM-Levenberg Marquard* e *Rprop-Resiliente Propagation* ), outros dois são avaliados: O algoritmo Gradiente conjugado escalado e o algoritmo Um passo secante (FERREIRA, 2006).

Segundo Ferreira (FERREIRA, 2006), nas previsões das séries temporais podem ocorrer um comportamento de defasagem temporal, que pode ser encarado como um ajuste de fase temporal relativa entre dados reais e previstos. Desta forma, existem duas possibilidades quanto à defasagem temporal das séries: a primeira, onde as séries temporais não tendem a um modelo de *Random walk* (Previsões em *fase*), e a segunda, para as séries que tendem a um modelo de *Random walk* (Previsões fora de *fase*).

Esta avaliação é dada aplicando-se dois procedimentos de previsão (Em fase e Fora de fase), ao conjunto de validação, logo após o treinamento da RNA. Nos dois resultados, é aplicado o teste de hipóteses *t-student* com um grau de confiança de 99% ( $\alpha = 0,01$ ) (FERREIRA; VASCONCELOS; ADEODATO, 2008; DE MATTOS NETO et al., 2014; DE MATTOS NETO; FERREIRA, 2016; FERREIRA, 2006). As hipóteses analisadas no teste são as seguintes (sendo testados todos os pares de pontos - Reais *vs.* Previstos):

$H_0$ : O erro MSE da previsão em fase é menor ou igual que o erro MSE da previsão fora de fase ;

$H_1$ : O erro MSE da previsão em fase é maior que o erro MSE da previsão fora de fase.

Uma vez a hipótese nula sendo não rejeitada, o método irá sugerir que a previsão está em fase, desta maneira não será necessária a intervenção do ajuste de fase na previsão. Caso a hipótese nula seja rejeitada, será feita a intervenção de ajuste de fase (FERREIRA, 2006; FERREIRA; VASCONCELOS; ADEODATO, 2008). A análise do resultado do teste *t-student* possibilita duas importantes interpretações:

**1ª: Em fase (Hipótese nula aceita);**

O processo gerador da séries temporal apresenta relação com o modelo preditivo, em outras palavras, a RNA não é treinada como um modelo *Random walk*.

**2ª: Fora de fase (Hipótese nula rejeitada).**

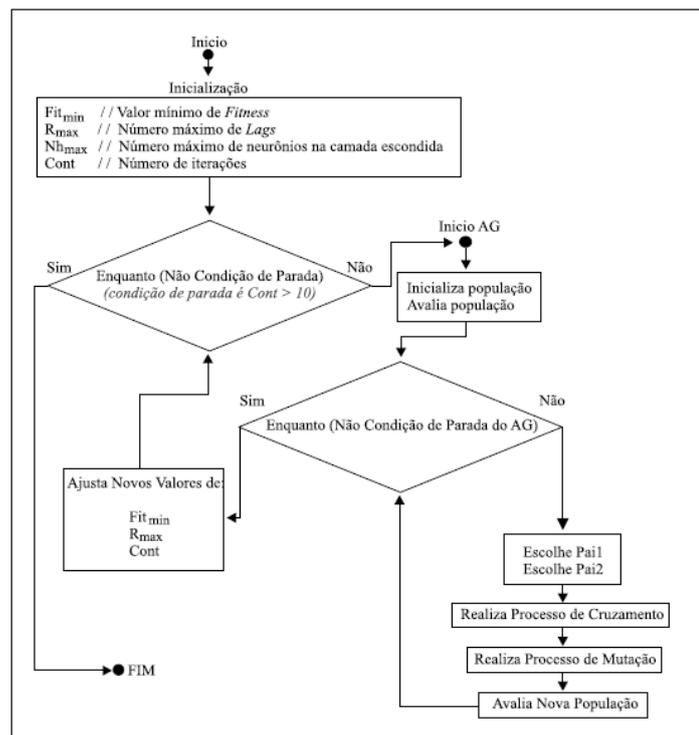
Existe uma incompatibilidade de atraso de tempo entre a série temporal prevista quando comparada com a série temporal real. Em outras palavras, a RNA treinada tende a se comportar como um modelo *Random walk*.

Ao inicializar o TAEF, devem-se determinar os seguintes parâmetros (FERREIRA, 2006):

- i). Valor mínimo aceitável de *fitness* que deve ser alcançado por pelo menos um indivíduo da população (Equação 29).
- ii). Número máximo de *Lags* a serem investigados -  $R_{max}$ ;
- iii). Quantidade de neurônios na camada oculta -  $Nh_{max}$ ;
- iv). Número de iterações do método - *Cont*.

No item i), o valor *erro* considerado pelo método TAEF na Equação 29 é o Erro quadrático médio (MSE). Estes parâmetros constituem apenas valores iniciais que serão ajustados a medida que o método evolui em busca do melhor modelo possível para a previsão da série temporal (FERREIRA, 2006). O método TAEF, com a inclusão do teste comportamental, está esquematizado na Figura 9.

Figura 9 – Esquema do Método TAEF.



Fonte: Ferreira (FERREIRA, 2006).

Segundo FERREIRA (2006), após sua inicialização dada pelo usuário, o TAEF inicializa um AGM para a realização da busca pela melhor configuração dos parâmetros das RNAs bem como o algoritmo de treinamento aplicado. Quando algum critério de parada é satisfeito, o AGM testa se o melhor indivíduo encontrado na população conseguiu obter sucesso quanto à meta estabelecida para seu valor de *fitness*. Caso este sucesso não tenha sido obtido, o TAEF considera que uma possível causa é a baixa dimensionalidade do  $R_{max}$ , e repete o procedimento em uma nova iteração e aumenta a dimensionalidade.

Entretanto, se o sucesso do indivíduo foi alcançado (indivíduo alcançar, ou superar, a meta estabelecida para seu menor valor de *fitness*), o método TAEF encara que uma solução aceitável foi encontrada (segundo critérios previamente estabelecidos pelo usuário) (FERREIRA, 2006). Contudo, se o número máximo de iterações (*Cont*) não tiver sido alcançado, o método TAEF considera a possibilidade de melhorar a própria meta anteriormente estabelecida pelo usuário.

Desta forma, o método ajusta o valor do parâmetro de mínimo valor de *fitness* aceitável ( $Fit_{min}$ ) como sendo o próprio valor de *fitness* alcançado pelo melhor indivíduo, ajustando o valor da dimensionalidade de busca dos *Lags* ( $R_{max}$ ) para o valor do *Lag* de mais alta ordem utilizado pelo indivíduo.

Este procedimento tenta buscar a dimensionalidade mínima necessária para a correta reconstrução do espaço de fase gerador da série temporal. Uma vez que estes valores são ajustados, repetições deste experimento são realizadas toda vez que o melhor indivíduo consiga superar o *fitness* mínimo estabelecido para sua iteração. Desta maneira geram-se novas metas para próxima iteração do método (FERREIRA, 2006; FERREIRA; VASCONCELOS; ADEODATO, 2008; DE MATTOS NETO et al., 2014; DE MATTOS NETO; FERREIRA, 2016).

## 3.2 Metodologia Proposta

### 3.2.1 O Sistema de Decomposição Wavelet - WavDS

Na Seção 2 foi apresentada a maneira com a qual o ruído pode se manifestar no comportamento fundamental do processo gerador da série temporal e sua relação negativa com a utilização de modelos de previsão. Também foi apresentado como pesquisadores buscaram alternativas para realizar um pré-processamento nos dados de maneira que este problema fosse resolvido e foi verificado que a análise de *Wavelets* possibilitou importantes contribuições neste contexto.

Donoho e Johnstone (1994) foram os pesquisadores que inicialmente propuseram

um método de filtragem baseado na retirada do ruído gaussiano branco por meio de *Wavelets*. Utilizaram o método de estimação do limiar  $\lambda$  conhecido atualmente por *Sqtwolog* que foi descrito na Equação 21.

No ano seguinte, Donoho (1995), apresenta a interpretação formal para o conceito de filtragem provando o esquema para tentar retirar o ruído por suavização ou limiarização no domínio *Wavelet* e afirma ainda que os métodos apresentados anteriormente para filtragem obtiveram apenas sucessos parciais.

Entretanto, como descrito na Subseção 2.1.2, séries temporais também são compostas por outros tipos de ruídos que podem afetar o desempenho dos modelos de previsão, o ruído rosa e marron são alguns exemplos (KIŞI; UNCUOĞLU, 2005; FLORES et al., 2016). Flores et al. (2016) apresentam uma análise comparativa em séries temporais de naturezas diferentes e verificam como a presença de diferentes tipos de ruído podem afetar o desempenho de diversos modelos de previsão.

Desta forma o trabalho apresentado por Donoho (1995), principal trabalho que envolve filtragem via *Wavelets*, possui a limitação na retirada dos outros tipos de ruído, os que não possuem as características de normalidade, homocedasticidade e não correlacionariedade que são pertinentes a um ruído branco. Outra questão relevante abordada na literatura da área é no que diz respeito a dificuldade para selecionar a *Wavelet*-mãe que será utilizada no processo de filtragem. Frequentemente, pesquisadores utilizam as *Wavelets Haar* e *Daubechies* por serem as mais utilizadas em trabalhos anteriores (KILIÇ; UĞUR, 2016; CHANDAR; SUMATHI; SIVANANDAM, 2016; JOO; KIM, 2015; KAO et al., 2013; KHANDELWAL; ADHIKARI; VERMA, 2015; RODRIGUEZ; BARBA, 2016; MUSTAFA; ALZUBAYDI, 2013; SANG; WANG; WU, 2010) .

Para superar as limitações anteriormente mencionadas também é proposto nesta Tese uma metodologia de pré-processamento de séries temporais intitulado WavDS (Sistema de Decomposição *Wavelet*), um sistema de decomposição da série temporal via *Wavelets* que é realizado em duas etapas.

Na primeira etapa, é realizada a tradicional filtragem *Wavelet* unidimensional proposta por Donoho em 1995 (DONOHO, 1995). Com a proposta, o ruído, que aqui é considerado gaussiano branco, deve ser eliminado) por meio da filtragem *Wavelet*. Nesta etapa, a série temporal  $y_t$  é dada de entrada no procedimento tradicional de filtragem via DW (DONOHO, 1995). Inicialmente faz-se uma seleção dos parâmetros necessários para o procedimento de filtragem que são enumerados à seguir:

- i. Regra de seleção de limiar;
- ii. Critério de estimação de limiar;

*iii.* Escolha da *Wavelet*-mãe;

*iv.* Nível de decomposição *Wavelet*.

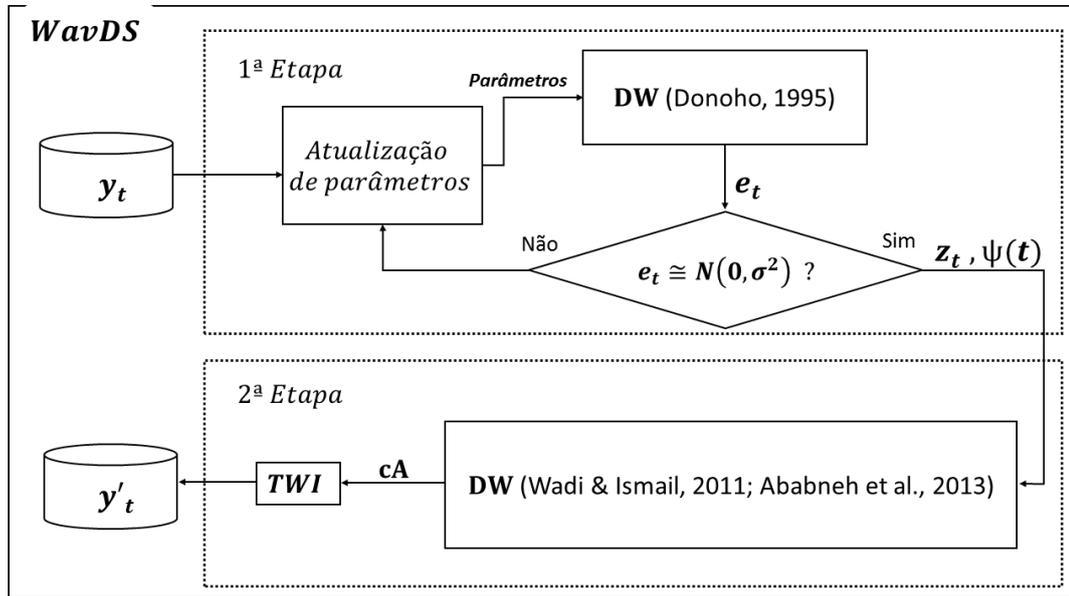
Na Subseção 3.1.2 estão descritas as formulações matemáticas das opções da escolha da regra de seleção de limiar (*Hard Thresholding* e *Soft Thresholding*) e dos critérios de estimação do limiar (*Minimaxi*, *Sqtwolog*, *Rigrsure* e *Heusure*). Após a definição da regra de seleção do limiar  $\lambda$  e do critério de estimação do limiar  $\lambda$  descritos na Subseção 3.1.1, a função *Wavelet* que será utilizada como *Wavelet*-mãe, e o nível da DW possível (1º nível), devem ser definidos para dar início ao procedimento de filtragem. Nesta momento da primeira etapa deve-se escolher (de maneira heurística) a função *Wavelet* que será utilizada como *Wavelet*-mãe.

Após a primeira etapa do procedimento ser realizado, o ruído da filtragem ( $e_t$ ) é gerado e analisado como critério de parada para a primeira etapa. Caso o ruído tenha as características pertinentes a um ruído branco gaussiano ( $e_i \approx N(0, \sigma^2) - iid$ ) a primeira etapa do método WavDS é concluída e são determinados uma série pré-processada ( $z_t$ ) e a *Wavelet*-mãe ( $\psi(t)$ ) (*Wavelet* selecionada que possibilitou  $e_i \approx N(0, \sigma^2) - iid$ ). Este procedimento está ilustrado na Figura 10.

Pela proposta, a análise do ruído deve ser analisada por meio de quatro diferentes testes de normalidade, são eles: Teste Anderson darling (SCHOLZ; STEPHENS, 1987), Teste Jarque-Bera (JARQUE; BERA, 1987), Teste Lilliefors (LILLIEFORS, 1967) e o teste Kolmogorov-Smirnov (JR, 1951).

Caso o ruído provindo da primeira etapa não tenha as características pertinentes a um ruído branco gaussiano, uma nova seleção de parâmetros é realizada a fim de que o critério de parada estabelecido para conclusão da primeira etapa seja encontrado. A modificação dos parâmetros para satisfazer o critério de parada ( $e_i = N(0, \sigma^2)$ ) segue a mesma ordem dada anteriormente para a seleção dos parâmetros. Analisam-se as combinações possíveis envolvendo os parâmetros *i*, *ii* e *iii* antes de modificar o nível de decomposição (parâmetro *iv*).

Por meio do método proposto, a *Wavelet*-mãe ( $\psi(t)$ ) que proporcionar esta eliminação do ruído branco no menor nível de decomposição (na primeira etapa do WavDS) é a *Wavelet* selecionada como mãe para a segunda etapa do procedimento.

Figura 10 – Arquitetura do Sistema de Decomposição *Wavelet* - WavDS.

Fonte: Próprio autor.

A segunda etapa do WavDS consiste na utilização da decomposição *Wavelet* no primeiro nível nos dados obtidos a partir da realização da primeira etapa do procedimento. Utiliza-se a Transformada *Wavelet* e a respectiva Transformada *Wavelet* Inversa nos dados a fim de separar as componentes da série temporal em componentes de altas e baixas frequências (ABABNEH; WADI; ISMAIL, 2013; WADI; ISMAIL, 2011).

Neste momento da segunda etapa do procedimento WavDS, o conjunto de dados é subdividido em dois subconjuntos por meio da Transformada *Wavelet*, um formado pelos coeficientes de aproximações (tendências - componentes de baixas frequências) que é responsável pelo principal comportamento da série e outro pelos coeficientes de detalhes (variações - componentes de altas frequências) que são responsáveis pelos pequenos ajustes (ABABNEH; WADI; ISMAIL, 2013; WADI; ISMAIL, 2011).

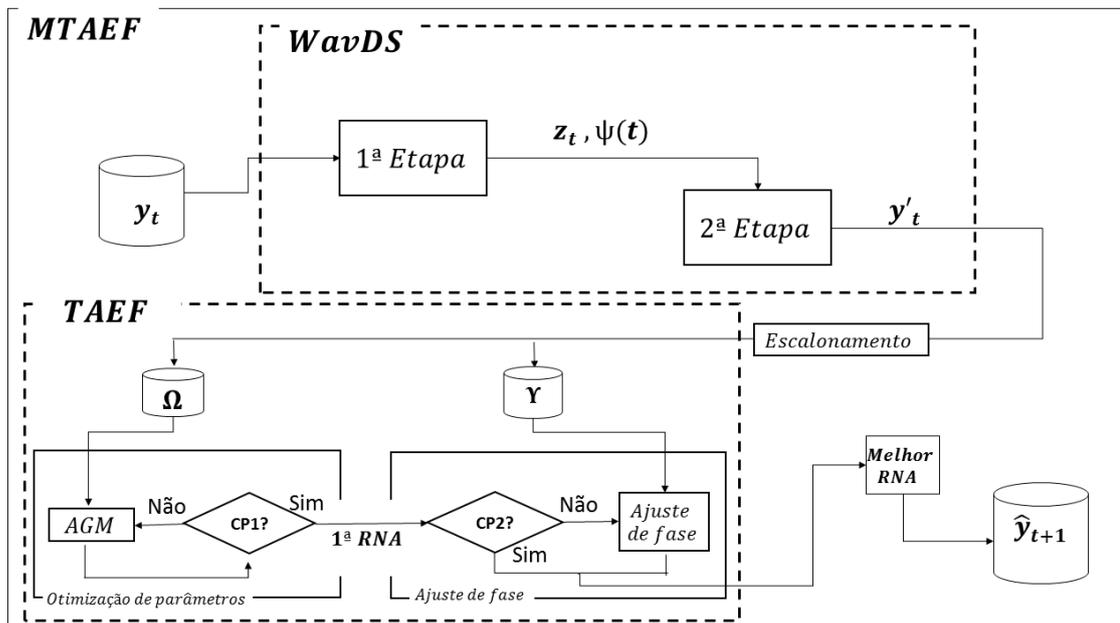
Assim, de maneira similar aos trabalhos apresentados por Wadi e Ismail (2011) e Ababneh, Wadi e Ismail (2013), descartam-se os coeficientes de detalhes ( $cD$ ) que são considerados outros tipos de ruído e utiliza-se a série dos coeficientes de aproximação ( $cA$ ) na Transformada *Wavelet* Inversa para produzir os dados pré-processados ( $y'_t$ ) à serem utilizados no modelo de previsão.

### 3.2.2 Multiresolution Time-delay Added Evolutionary Forecasting - MTAEF

O modelo proposto de previsão de séries temporais, intitulado *Multiresolution Time-delay Added Evolutionary Forecasting* (MTAEF) é inspirado nos trabalhos realizados por Liu, Lu e Wang (2012) e Ferreira, Vasconcelos e Adeodato (2008). Tem por objetivo minimizar, concomitantemente, os problemas relacionados ao pré-processamento da série temporal e a seleção dos parâmetros do modelo de RNAs do tipo MLP.

Combina o método de filtragem WavDS descrito na Subseção 3.2.1 com o método TAEF descrito na Subseção 3.1.5 com o objetivo de minimizar os problemas que envolvem as temáticas à cerca do pré-processamento das séries temporais e seleção dos melhores parâmetros das RNAs que foram descritos na Subseção 3.1.3. Desta forma, como pode ser visualizado na esquematização apresentada na Figura 11, o método MTAEF proposto é um procedimento que pode ser caracterizado em duas etapas distintas.

Figura 11 – Esquematização do método proposto Multiresolution Time-delay Added Evolutionary Forecasting - MTAEF.



Fonte: Próprio autor.

Na primeira etapa, é realizado o pré-processamento na série temporal por meio do procedimento de filtragem proposto WavDS que foi descrito na Subseção 3.2.1.

Na segunda etapa do método MTAEF, assim como realizado por Wadi e Ismail (2011) e por Ababneh, Wadi e Ismail (2013), um modelo de previsão de séries temporais é aplicado aos dados pré-processados ( $y'_t$ ). Foi escolhido o modelo de RNAs do tipo MLPs, por meio do método TAEF, como procedimento de previsão.

O método escolhido para a realização da previsão das séries temporais foi proposto por Ferreira, Vasconcelos e Adeodato (2008) em virtude de sua capacidade de ajuste e generalização de dados com características não lineares que são pertinentes às séries temporais do mundo real.

Na Figura 11, os dados filtrados a partir do pré-processamento via WavDS são utilizados pelo método TAEF. Inicialmente eles são escalonados no intervalo  $[0, 1]$  e separados no conjunto de treino ( $\Omega$ ) e validação ( $\gamma$ ) para dar início às etapas de otimização de parâmetros e ajuste de fase (FERREIRA; VASCONCELOS; ADEODATO, 2008). O critério de parada para a otimização dos parâmetros das RNAS (CP1?) e o critério de parada para o ajuste de fase (CP2?) estão descritos na Subseção 3.1.5.

O método MTAEF é um procedimento que pretende contribuir para que a realização do processo de previsão seja dado com estabilidade estatística possibilitando a redução de diversos tipos de ruído ao mesmo tempo em que garante a manutenção da estrutura de auto dependência da série temporal.

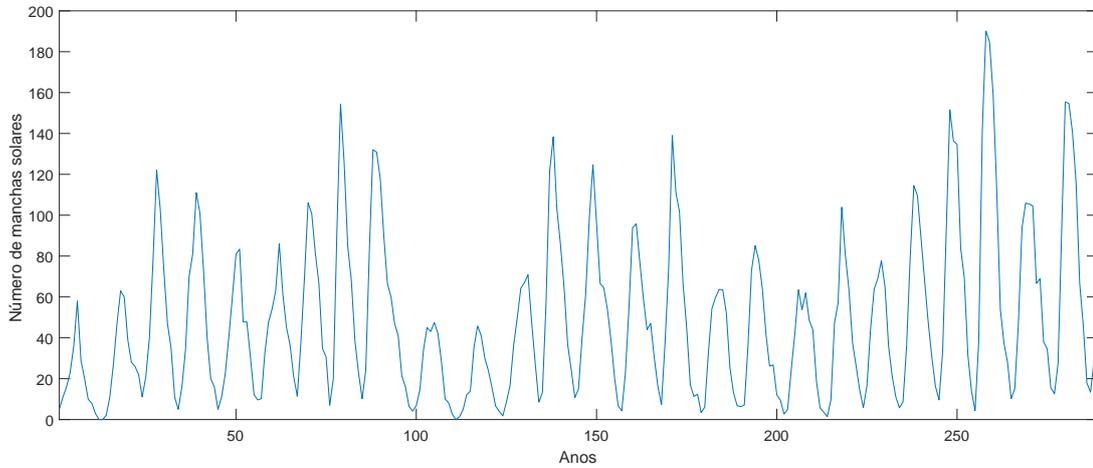
## 3.3 Materiais

### 3.3.1 Dados

As tradicionais séries, *Wolf's sunspot*, *Canadian lynx* e *Exchange rate* foram utilizadas neste estudo experimental. Estas séries temporais são amplamente estudadas por possuírem diferentes características estatísticas e virem de diversas áreas do conhecimento (KHASHEI; BIJARI, 2011; ZHANG, 2003; BABU; REDDY, 2014; BABU; SURE, 2016). As séries temporais *Wolf's sunspot* e *Canadian lynx* podem ser encontradas em <https://datamarket.com> enquanto a série *Exchange rate* pode ser obtida em <https://fred.stlouisfed.org/series/EXUSUK>.

A série *Wolf's sunspot* corresponde ao número anual de manchas solares no período de 1700 a 1987. Esta série possui um total de 288 dados. O seu gráfico é apresentado na Figura 12. Também conhecida como *Sunspot*, ela é considerada uma série temporal não linear e é utilizada frequentemente para a avaliação de modelos não lineares (ZHANG, 2003; KHASHEI; BIJARI, 2011; KHANDELWAL; ADHIKARI; VERMA, 2015; FIRMINO; DE MATTOS NETO; FERREIRA, 2015; BABU; SURE, 2016).

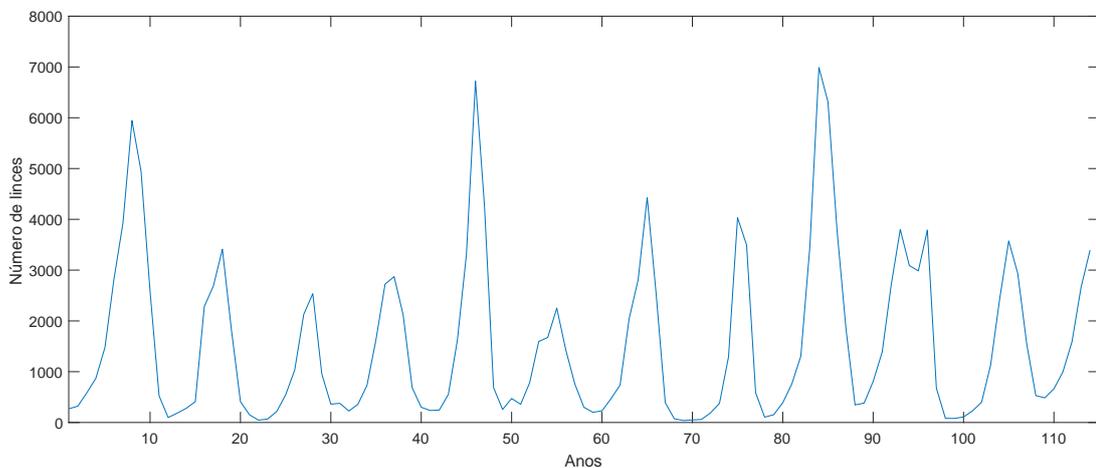
Figura 12 – Gráfico do conjunto de dados *Sunspot*.



Fonte: Próprio autor.

A série *Canadian lynx* apresenta o número anual de linces capturados no rio *Mackenzie* que fica localizado ao norte do Canadá. O gráfico que representa esta série é apresentado na Figura 13.

Figura 13 – Gráfico do conjunto de dados *Lynx*.

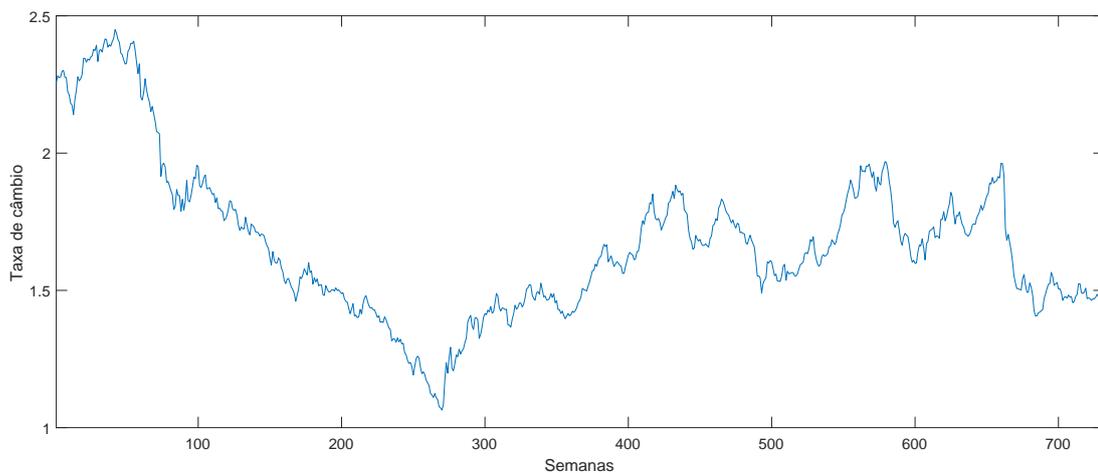


Fonte: Próprio autor.

Este conjunto de dados possui 114 observações no período de 1821 a 1934. Assim como a série *Sunspot*, também existe extensa literatura com foco em modelagem não linear que a utiliza (KHASHEI; BIJARI, 2011; ZHANG, 2003). Assim como realizado por Zhang (2003) e Khashei e Bijari (2011), a transformação nos dados foi realizada com a utilização do logaritmo na base decimal para garantir condições experimentais de comparação.

O próximo conjunto de dados é o *Exchange rate*. Seu gráfico é apresentado na Figura 14. Corresponde à 730 observações semanais da taxa de câmbio entre a libra esterlina e o dólar norte-americano. Esta série é analisada no período de 1980 a 1993. Neste caso, assim como realizado por Zhang (2003) e por Khashei e Bijari (2011), o logaritmo natural foi utilizado para garantir condições experimentais de comparação dos resultados (KHASHEI; BIJARI, 2011; KHANDELWAL; ADHIKARI; VERMA, 2015; ZHANG, 2003).

Figura 14 – Gráfico do conjunto de dados *Exchange rate*.



Fonte: Próprio autor.

A análise dos dados foi conduzida com a proposta de verificar o desempenho do método MTAEF e do procedimento de filtragem WavDS, por meio de uma análise comparativa que foi realizada em dois diferentes contextos caracterizados nesta Tese como situações experimentais. A primeira situação experimental diz respeito a análise do processo de filtragem WavDS e a segunda situação experimental diz respeito a análise do método MTAEF proposto.

Para isto, na primeira situação analisada, o experimento foi conduzido com a proposta de avaliar o desempenho do método de pré-processamento WavDS por meio da análise da variabilidade estatística e da acurácia dos resultados da previsão.

A análise foi conduzida a partir da comparação das previsões encontradas com a utilização do filtro WavDS combinado com modelos de RNAs do tipo MLP, com a combinação de tradicionais métodos de filtragem via *Wavelets* com MLPs e tradicionais modelos de previsão (RNAs do tipo MLPs e Autorregressivos integrados de médias móveis - ARIMA ) sem a utilização de pré-processamento nas séries.

Nesta primeira situação de experimentação, as séries temporais e seus respectivos horizontes de previsão (Amostra teste) que estão descritos na Tabela 1 foram analisados

como sugeridos nos artigos de Kashei & Bijari (KHASHEI; BIJARI, 2011) e Zhang (ZHANG, 2003) de modo a garantir equidade na análise comparativa dos resultados.

Tabela 1 – Composição amostral dos conjuntos de dados *Sunspot*, *Lynx* e *Rate*.

Séries	Amostra total	Amostra treino	Amostra teste
Sunspot	288	221	67
Lynx	114	100	14
Exchange rate	730	678	52

Os conjuntos de dados *Sunspot* e *Lynx* também são utilizados na segunda situação experimental. Nesta situação, os horizontes de previsão utilizados para análise do método preditivo proposto MTAEF foram analisados como sugeridos nos artigos de Firmino, De Mattos Neto e Ferreira (2015) e Babu e Sure (2016).

Desta maneira, podemos dividir os conjuntos de teste para a análise dos modelos em cinco distintos horizontes (casos), de modo que os três primeiros casos serão analisados na primeira situação experimental (Análise do desempenho do método de pré-processamento WavDS - Seção 4.2) e os dois últimos casos serão analisados na segunda situação experimental (Análise do método híbrido MTAEF - Seção 4.3). Estes cinco casos estão descritos na Tabela 2.

Tabela 2 – Horizontes de previsão utilizados nas situações experimentais para os conjuntos de dados *Sunspot*, *Lynx* e *Rate*.

Séries	1º caso	2º caso	3º caso	4º caso	5º caso
Sunspot	35	67	—	72	25
Lynx	14	—	—	28	—
Exchange rate	5	26	52	—	—

### 3.3.2 Métricas de Acurácia

As métricas utilizadas para a análise do desempenho de previsão da série temporal  $\{y_t\}_{t=1,2,\dots,N}$  são definidas em termos do erro de previsão  $e_t$ , o qual é formulado a partir da diferença entre o valor real da série ( $y_t$ ) e o valor previsto pelo modelo preditivo ( $\hat{y}_t$ ), conforme descrito na Equação 30.

$$e_t = y_t - \hat{y}_t \quad (30)$$

A análise dos resultados será norteada por meio de diferentes indicadores de exatidão que são utilizados na literatura da área (ZHANG, 2003; KHASHEI; BIJARI, 2011; WANG

et al., 2013; BABU; REDDY, 2014; FIRMINO; DE MATTOS NETO; FERREIRA, 2015; BABU; SURE, 2016).

Zhang (2003), Khashei e Bijari (2011) e demais pesquisadores (WANG et al., 2013; BABU; REDDY, 2014; BABU; SURE, 2016) quantificaram a acurácia na previsão das séries por meio de dois diferentes indicadores de erro, o Erro Quadrático Médio (*Mean Square Error - MSE*) e o Erro Absoluto Médio (*Mean Absolute error - MAE*), que são calculados por meio das seguintes equações:

$$MSE = \frac{1}{h} \sum_{t=1}^h e_t^2 \quad (31)$$

$$MAE = \frac{1}{h} \sum_{t=1}^h |e_t| \quad (32)$$

Com  $h$  sendo o horizonte de previsão (quantidade de dados previstos).

Na segunda situação experimental, o método MTAEF tem o seu desempenho comparado, por meio da análise realizada em todos conjuntos de dados apresentados sob diferentes métricas de acurácia.

Além das métricas anteriormente comentadas, Firmino, De Mattos Neto e Ferreira (2015) utilizam outras medidas para analisar o desempenho do modelo preditivo proposto, são elas: *POCID* (*Prediction on change in direction*), *Theil* (Estatística U de Theill), *MAPE* (*Mean absolute percentual error*) e *ARV* (Variância relativa média).

A métrica *POCID* mede o percentual de acerto relativo a tendência da série temporal (FIRMINO; DE MATTOS NETO; FERREIRA, 2015). Ela está definida na Equação 33.

$$POCID = 100 \frac{\sum_{t=1}^h D_t}{h} \quad (33)$$

de modo que:

$$D_t = \begin{cases} 1, & \text{se } (y_t - y_{t-1}) \cdot (\hat{y}_t - \hat{y}_{t-1}) > 0 \\ 0 & \text{se caso contrário.} \end{cases}$$

A estatística *U de Theil* se baseia no *MSE* do preditor escalonado pelo erro de previsão de um modelo do tipo *Random walk* (FIRMINO; DE MATTOS NETO; FERREIRA, 2015). Ela está definida na Equação 34.

$$Theil = \frac{\sum_{t=1}^h (y_t - \hat{y}_t)^2}{\sum_{t=1}^h (y_t - y_{t-1})^2} \quad (34)$$

de modo que:

- **Theil=1** Modelo tem desempenho igual a um modelo *Random walk*;
- **Theil>1** Modelo tem desempenho inferior a um modelo *Random walk*;
- **Theil<1** Modelo tem desempenho superior a um modelo *Random walk*.

Em outras palavras, quanto mais próximo de zero a estatística *U de Theil* é, mais bem ajustado está o modelo aos dados.

Dentre as métricas utilizadas nesta pesquisa, na Equação 35 é definido o *MAPE*, uma medida que quantifica o Erro médio percentual.

$$MAPE = \frac{1}{h} \sum_{t=1}^h \left| \frac{e_t}{y_t} \right| \quad (35)$$

de modo que, neste caso específico,  $y_t$  é o ponto da série no instante  $t$ .

Outro quantificador de acurácia adotado é a variância relativa média (*ARV*), que associa o desempenho do modelo com o desempenho de se prever simplesmente a media da série e que possui definição dada por:

$$ARV = \frac{1}{h} \frac{\sum_{t=1}^h (\hat{y}_t - y_t)^2}{\sum_{t=1}^h (y_t - \bar{y}_t)^2}. \quad (36)$$

O desempenho do MTAEF é comparado com o desempenho de diferentes modelos híbridos propostos na literatura (ZHANG, 2003; FERREIRA; VASCONCELOS; ADEODATO, 2008; KHASHEI; BIJARI, 2011; WANG et al., 2013; BABU; REDDY, 2014; FIRMINO; DE MATTOS NETO; FERREIRA, 2015; BABU; SURE, 2016).

## 4 Resultados e Discussão

Neste capítulo as três séries temporais descritas na Seção 3.3 foram analisadas e os resultados desta análise estão comentados nas seções à seguir. Os experimentos foram realizados com o objetivo de avaliar o desempenho do método de filtragem WavDS e do método MTAEF por meio da análise de métricas tradicionais de acurácia na previsão de séries temporais. O procedimento de previsão considerado é o de um passo à frente (*one-step-ahead*).

Este procedimento foi implementado com a utilização do *software* MATLAB versão R2016a com a *Wavelet Toolbox* (MISITI et al., 1996) para a análise de *Wavelet* e a utilização do *NNtool* (DEMUTH; BEALE, 2009) para a análise e previsão de séries temporais via RNAs.

### 4.1 Análise dos resultados da filtragem do método WavDS

Na Subseção 4.1.1 os resultados para o conjunto de dados *Sunspot* são apresentados. Os resultados para os conjuntos de dados *Canadian lynx* e *Exchange rate* são apresentados na Subseção 4.1.2 e na Subseção 4.1.3.

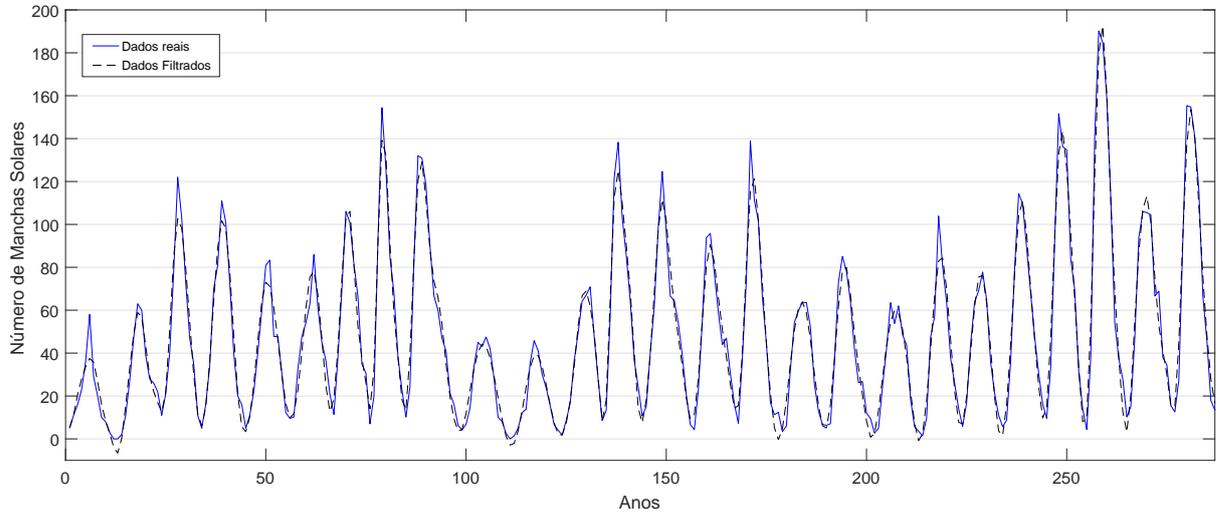
#### 4.1.1 Caso Wolf's sunspot

Os parâmetros descritos na Subseção 3.2.1 utilizados neste conjunto de dados, que possibilitaram que o resíduo removido na primeira etapa do procedimento WavDS apresente características pertinentes ao ruído branco gaussiano, estão apresentados à seguir:

- i.* Regra de seleção de limiar: ***Soft thresholding*** ;
- ii.* Critério de estimação de limiar: ***Minimaxi***;
- iii.* Escolha da *Wavelet*-mãe: **Discreta de Meyer**;
- iv.* Nível de decomposição *Wavelet*: **2º Nível**.

Na Figura 15 é apresentado o gráfico dos dados resultantes da filtragem realizada pelo método WavDS sobreposto ao gráfico dos dados reais para o conjunto *Sunspot*.

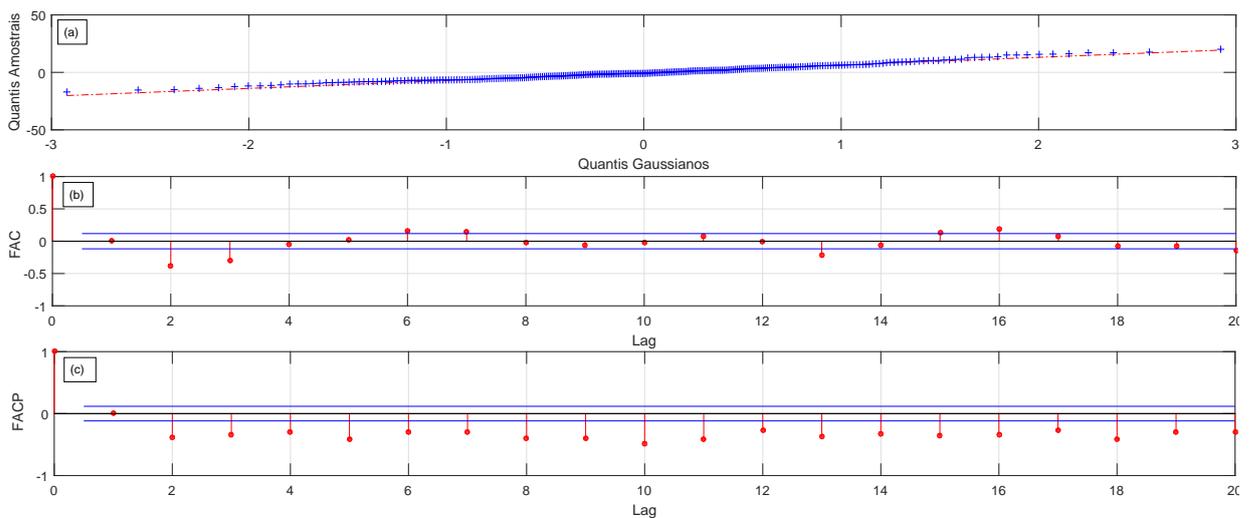
Figura 15 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados *Sunspot*.



Fonte: Próprio autor.

Por meio da análise gráfica realizada nas Figuras 16(a), 16(b) e 16(c), pode-se verificar que os resíduos removidos na 1ª etapa do WavDS possuem características pertinentes a um ruído branco gaussiano.

Figura 16 – (a)Gráfico QQ-Plot da gaussiana para os resíduos dos dados *Sunspot*, (b)Gráfico da FAC para os resíduos dos dados *Sunspot*, (c)Gráfico da FACP para os resíduos dos dados *Sunspot*.



Fonte: Próprio autor.

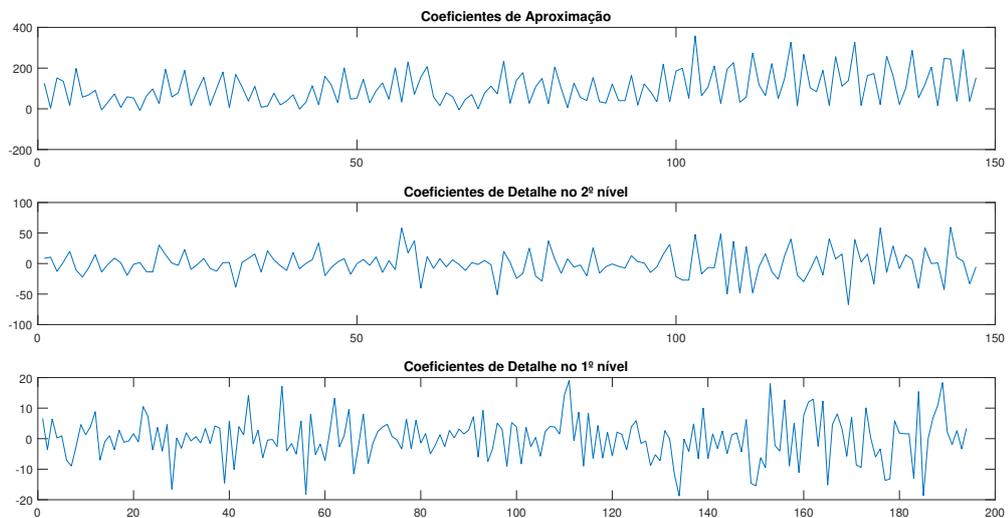
A comparação realizada nos percentis amostrais e teóricos dos gráficos sugerem que

os processos seguem distribuição normal. Tal análise foi verificada com a utilização do teste de normalidade Lilliefors (LILLIEFORS, 1967) considerando 0,01 o nível de significância de probabilidade.

Por meio da análise das Funções de Auto-correlação (FAC) e Auto-correlação Parcial (FACP) ilustradas nas Figuras 16(b) e 16(c) pode-se observar no primeiro retardo temporal (*Lag*) que os processos são estacionários em média e variância. Em virtude das características lineares das FAC e FACP, assim como a verificação da normalidade dos dados dada anteriormente, pode-se assumir que os resíduos encontrados podem ser de fato caracterizados como ruído branco.

Assim como descrito na Subseção 3.1.1.2 a DW no segundo nível possibilita a decomposição da série temporal em três subconjuntos, dois conjuntos formado pelos coeficientes de detalhes  $cD$  (para cada nível) e o outro pelos coeficientes de aproximação  $cA$  no segundo nível. Este conjuntos estão ilustrados graficamente na Figura 17.

Figura 17 – Resultados obtidos após a Decomposição *Wavelet* no 2º nível para os dados *Sunspot*.

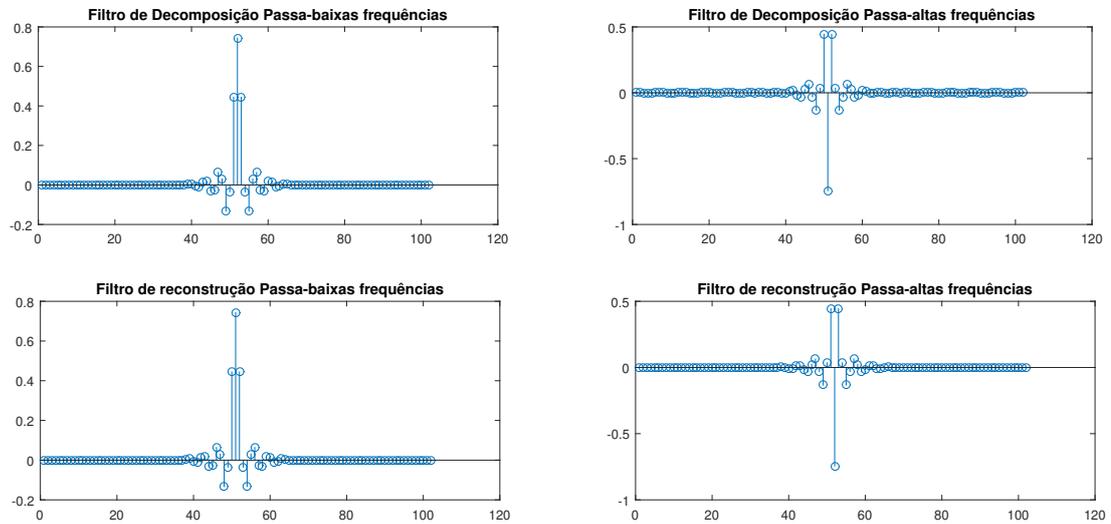


Fonte: Próprio autor.

Conforme descrito na Subseção ??, a função escala  $\phi(t)$  apresentada na Equação 8 e a *Wavelet*-mãe  $\psi(t)$  apresentada na Equação 10 podem ser obtidas por meio de sua representação por Filtros digitais FIR.

A representação gráfica do banco de filtros FIR para decomposição e reconstrução da séries temporal para a *Wavelet* discreta de Meyer está ilustrada na Figura 18.

Figura 18 – Filtros FIR para Wavelet Discreta de Meyer



Fonte: Próprio autor.

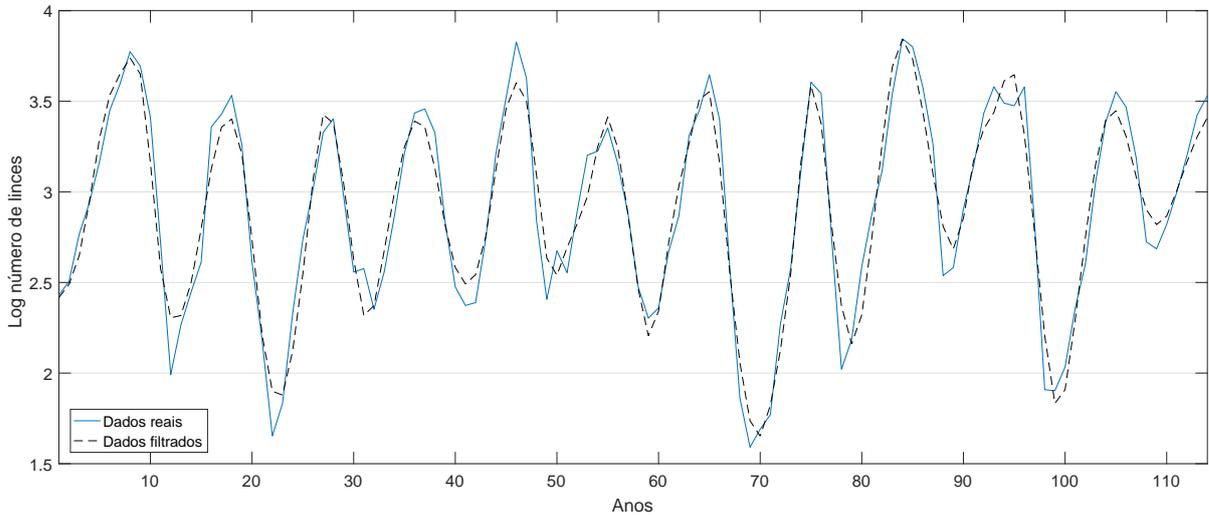
#### 4.1.2 Caso Canadian lynx

Os parâmetros utilizados para conjunto de dados *Canadian lynx* que possibilitaram que o resíduo removido na primeira etapa do procedimento WavDS (Subseção 3.2.1) apresente características pertinentes ao ruído branco gaussiano estão apresentados à seguir:

- i.* Regra de seleção de limiar: ***Soft thresholding*** ;
- ii.* Critério de estimação de limiar: ***Minimaxi***;
- iii.* Escolha da *Wavelet*-mãe: ***Symlet***;
- iv.* Nível de decomposição *Wavelet*: **2º Nível**.

Como pode ser observado, por meio do filtro WavDS a *Wavelet*-mãe selecionada para os dados *Lynx* foi diferente da *Wavelet*-mãe selecionada para *Sunspot*. Os gráficos com os resultados do pré-processamento realizado pelo procedimento WavDS são apresentados na Figura 19 para o caso dos dados *Lynx*.

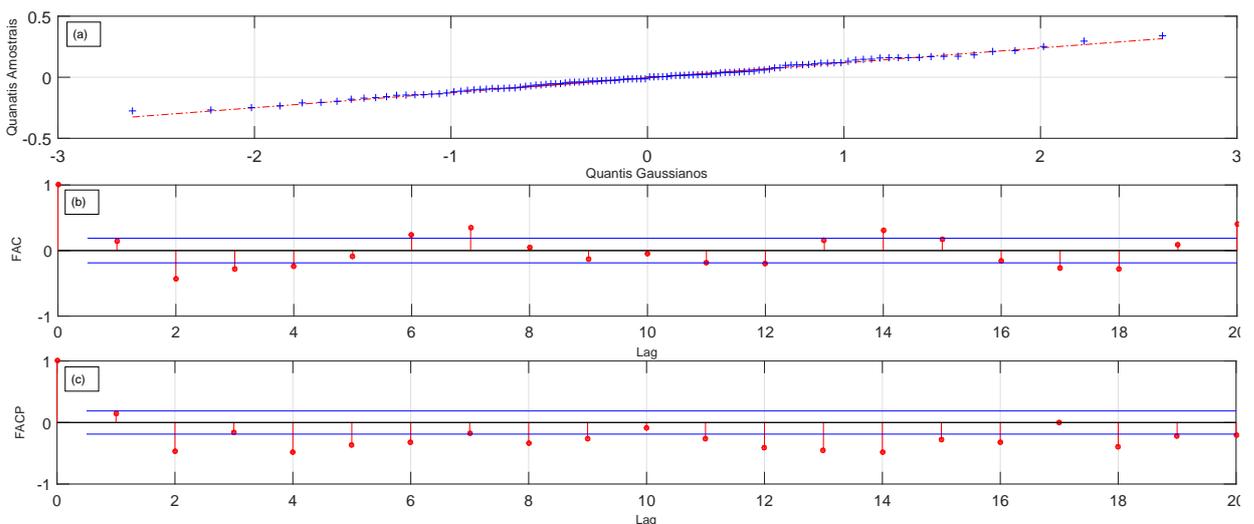
Figura 19 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados *Lynx*.



Fonte: Próprio autor.

Por meio da análise gráfica realizada nas Figuras 20(a), 20(b) e 20(c), de maneira semelhante ao caso sunspot, pode-se verificar que os resíduos removidos na 1ª etapa do WavDS possuem características pertinentes a um ruído branco gaussiano.

Figura 20 – (a) Gráfico QQ-Plot da gaussiana para os resíduos dos dados *Lynx*, (b) Gráfico da FAC para os resíduos dos dados *Lynx*, (c) Gráfico da FACP para os resíduos dos dados *Lynx*.



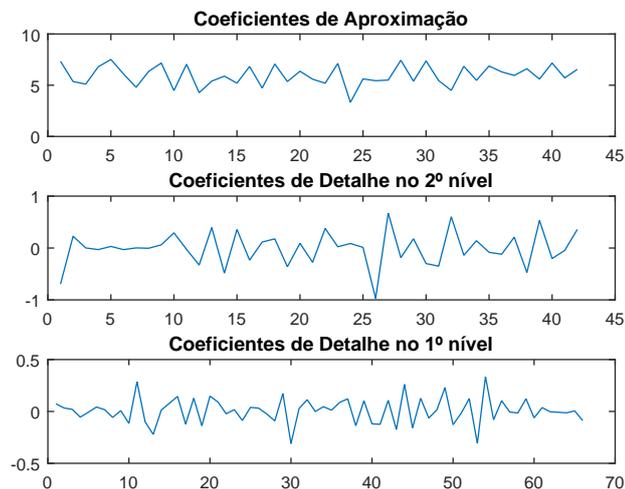
Fonte: Próprio autor.

A comparação realizada nos percentis amostrais e teóricos dos gráficos também

sugere que os resíduos seguem distribuição normal. Tal análise foi verificada tanto com a utilização do teste de normalidade Lilliefors (LILLIEFORS, 1967) quanto com a utilização do teste de normalidade jarque-Bera (JARQUE; BERA, 1987) considerando 0,01 e 0,05 os níveis de significância.

De maneira semelhante ao conjunto de dados *Sunspot*, por meio da análise das Funções de Auto-correlação (FAC) e Auto-correlação Parcial (FACP) ilustradas nas Figuras 20(b) e 20(c) pode-se observar que no primeiro *Lag* os processos são estacionários em média e variância. Desta forma, pode-se concluir que existem evidências que permitem assumir que os resíduos encontrados podem ser considerados como ruído branco.

Figura 21 – Resultados obtidos após a Decomposição *Wavelet* no 2º nível para os dados *Lynx*.



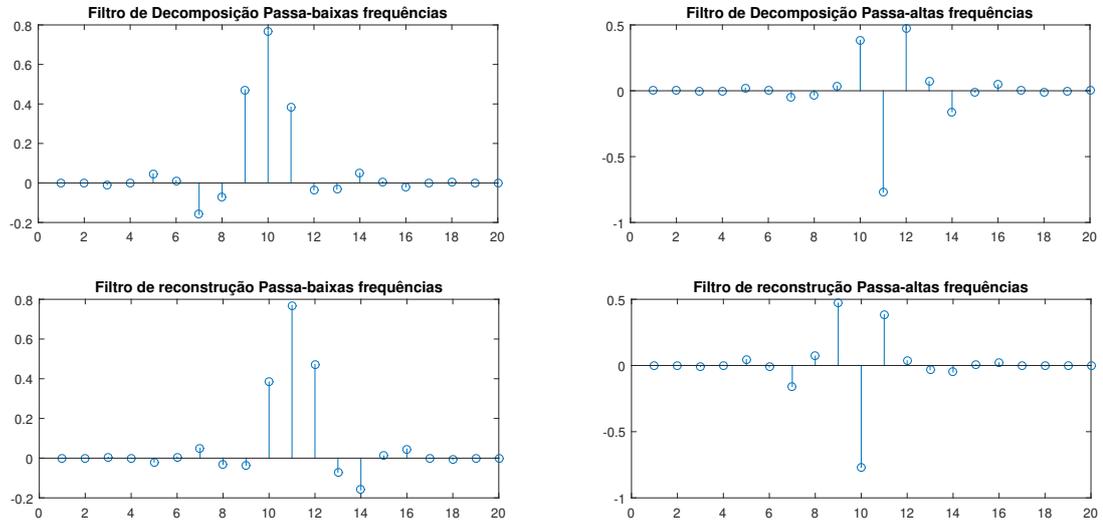
Fonte: Próprio autor.

A ilustração gráfica da função *Wavelet*-mãe  $\psi$  e da função escala  $\phi$  referentes à *Wavelet Symlet* são apresentadas no Apêndice A e no Apêndice B, respectivamente.

Os conjuntos de coeficientes obtidos a partir da DW estão apresentados graficamente na Figura 21 para os coeficientes de aproximação no segundo nível, coeficientes de detalhe no segundo nível e coeficientes de detalhe no primeiro nível.

A representação gráfica da *Wavelet Symlet* por banco de filtros FIR para decomposição e reconstrução da séries temporal, conforme descrito na Subseção 3.1.1, é apresentada na Figura 22.

Figura 22 – Filtros FIR para Wavelet Symlet

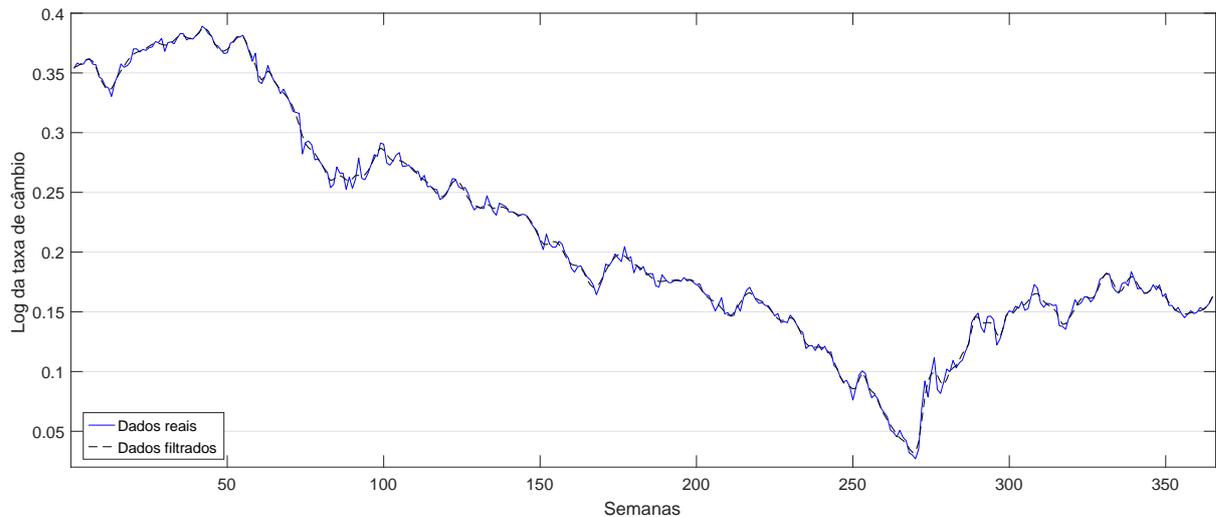


Fonte: Próprio autor.

#### 4.1.3 Caso Exchange rate

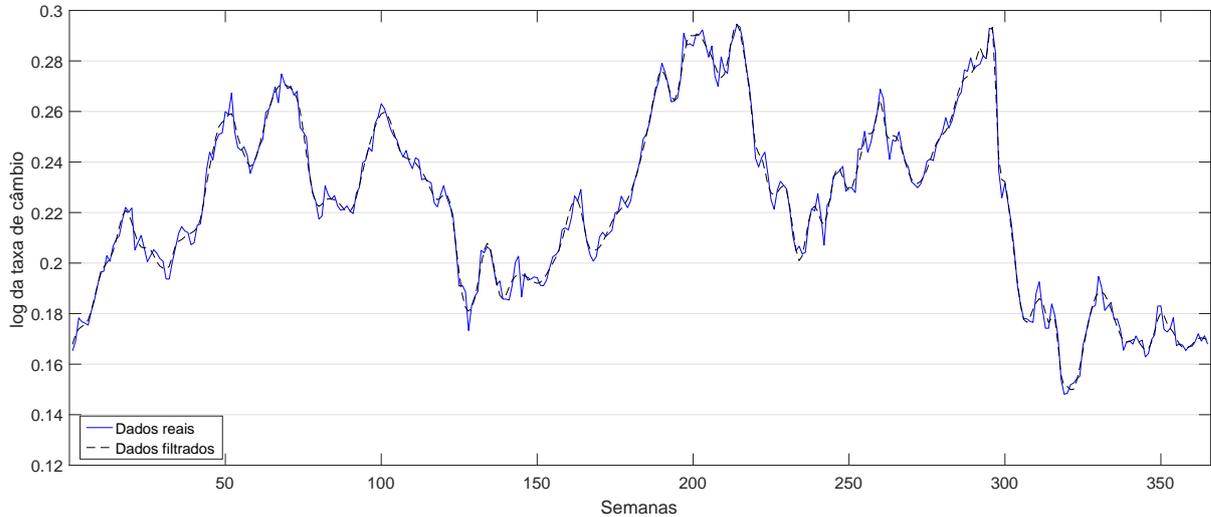
Para os dados *Exchange rate*, para uma melhor visualização, os resultados da filtragem foram divididos em dois gráficos que são apresentados na Figura 23 e na Figura 24.

Figura 23 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados *Exchange rate* nas primeiras 365 semanas.



Fonte: Próprio autor.

Figura 24 – Resultados obtidos após o pré-processamento via WavDS para o conjunto de dados *Exchange rate* nas últimas 365 semanas.



Fonte: Próprio autor.

Na Figura 23 está o gráfico do resultado da filtragem sobreposto aos dados reais para as primeiras 365 semanas e na Figura 24 para o restante dos dados.

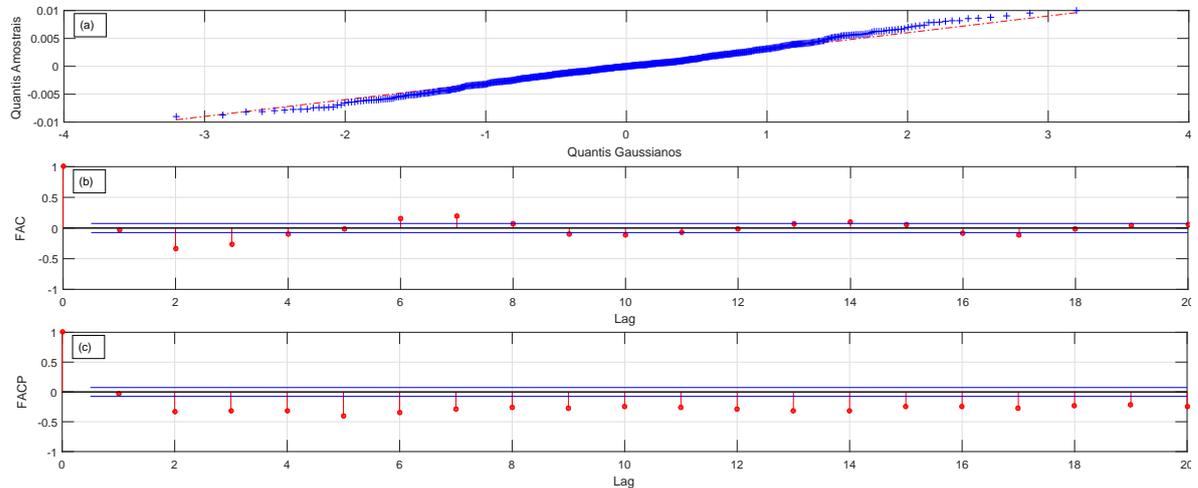
Os parâmetros selecionados na primeira etapa do WavDS estão descritos a seguir:

- i. Regra de seleção de limiar: ***Soft thresholding*** ;
- ii. Critério de estimação de limiar: ***Minimaxi***;
- iii. Escolha da *Wavelet*-mãe: ***Coiflet***;
- iv. Nível de decomposição *Wavelet*: **2º Nível**.

Deste modo, pode-se perceber que para todos os casos analisados, o método de pré-processamento proposto proporcionou que houvesse alteração na seleção da *Wavelet*-mãe superando assim limitações encontradas na literatura (WADI; ISMAIL, 2011; ABABNEH; WADI; ISMAIL, 2013).

Os gráficos das funções *Wavelet* que representam a *Wavelet Coiflet* são ilustrados nos Apêndices A e B. Os resíduos da filtragem do conjunto de dados *Exchange rate* são apresentados na figura a seguir.

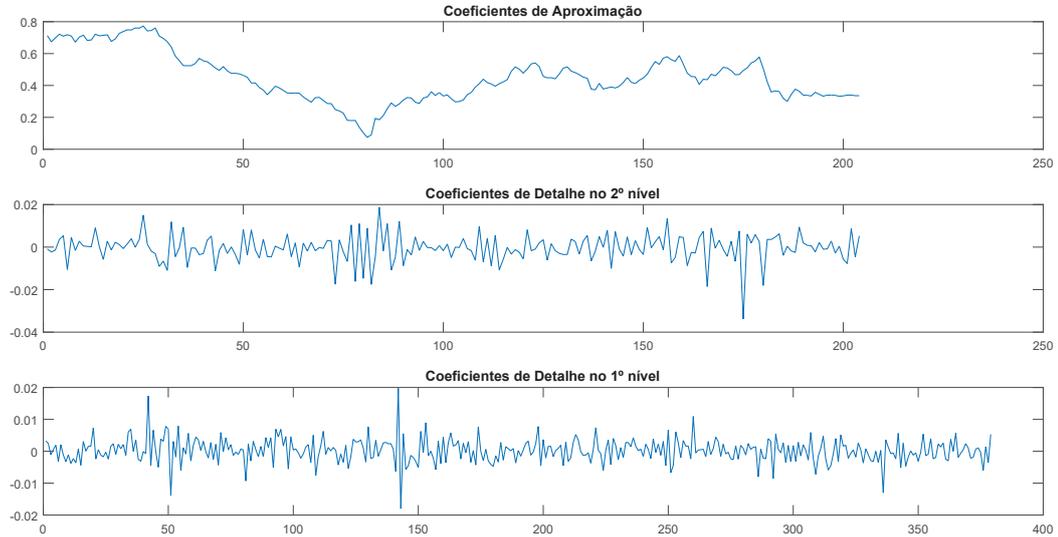
Figura 25 – (a)Gráfico QQ-Plot da gaussiana para os resíduos dos dados *Exchange rate*, (b)Gráfico da FAC para os resíduos dos dados *Exchange rate*, (c)Gráfico da FACP para os resíduos dos dados *Exchange rate*.



Fonte: Próprio autor

De maneira similar aos resultados dos conjuntos de dados *Wolf's sunspot* e *Canadian lynx*, a análise das Funções de Auto-correlação (FAC) e Auto-correlação Parcial (FACP) ilustradas nas Figuras 25(b) e 25(c) sugerem que os resíduos encontrados podem ser caracterizados como ruído branco. A normalidade dos resíduos foi verificada com a utilização do teste de Lilliefors (LILLIEFORS, 1967) considerando como 0,01 o nível de significância.

Figura 26 – Resultados obtidos após a Decomposição *Wavelet* no 2º nível para os dados *Exchange rate*.

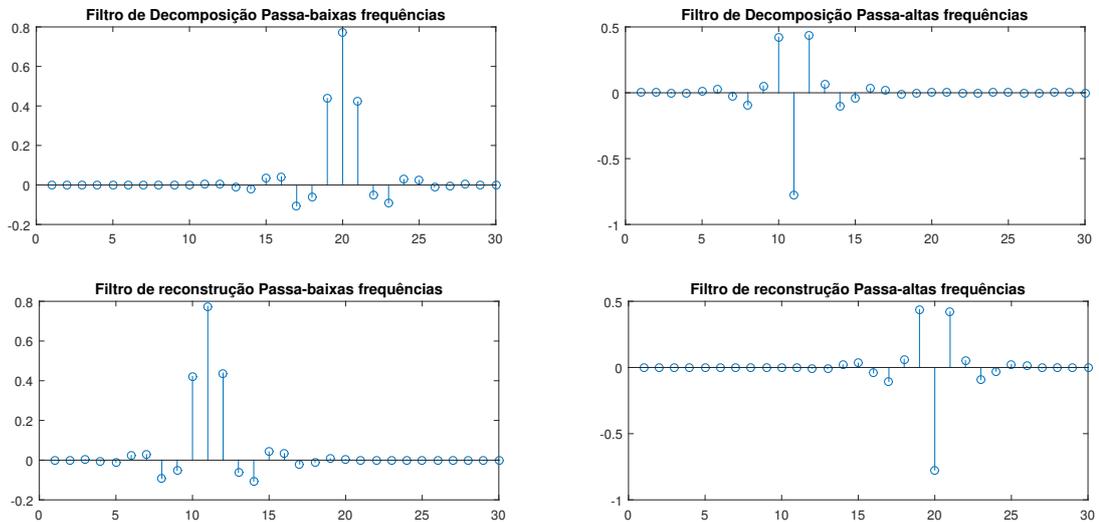


Fonte: Próprio autor.

A ilustração gráfica dos três subconjuntos de dados obtidos por meio da DW ( $cD$  para cada nível de decomposição e  $cA$  no segundo nível) estão ilustrados graficamente na Figura 26.

Já a representação gráfica do banco de filtros FIR para decomposição e reconstrução da séries temporal para a *Wavelet Coiflet* está ilustrada na Figura 27.

Figura 27 – Filtros FIR para Wavelet Coiflet



Fonte: Próprio autor.

## 4.2 Análise do desempenho do filtro WavDS

Nesta seção estão apresentadas a análise da variabilidade estatística e da acurácia dos resultados encontrados na previsão de cada um dos conjuntos de dados.

O desempenho do método de pré-processamento WavDS foi avaliado por meio da análise realizada sob duas diretrizes:

- Exatidão dos resultados da previsão com a utilização de RNAs do tipo MLP;
- Variabilidade estatística destes resultados.

O experimento foi realizado com 50 repetições independentes comparando-se a exatidão dos resultados das previsões com o método de RNAs encontrados a partir das seguintes situações:

- 1º Dados sem a utilização de pré-processamento;
- 2º Dados com filtragem realizada pelo método clássico apresentado por Donoho (1995) com as principais *Wavelets*-mãe utilizadas na literatura (*Haar* e *Daubechies*) (KILIÇ; UĞUR, 2016; CHANDAR; SUMATHI; SIVANANDAM, 2016; JOO; KIM, 2015; KHANDELWAL; ADHIKARI; VERMA, 2015; RODRIGUEZ; BARBA, 2016) ;
- 3º Dados filtrados a partir do método WavDS.

A exatidão destes resultados foi analisada seguindo o critério do menor valor das medidas de erro encontrado pelos valores previstos do modelo. Para isto, o MSE e o MAE foram as métricas adotadas. Estas métricas de acurácia são comumente utilizadas na literatura da área (ZHANG, 2003; KHASHEI; BIJARI, 2011; BABU; REDDY, 2014; BABU; SURE, 2016).

Os resultados numéricos mostram em todos os casos analisados que, após a utilização do método de filtragem WavDS, houve melhora em acurácia e redução na variabilidade estatística dos resultados da previsão via MLPs que é existente em virtude da reinicialização dos pesos sinápticos da RNAs, conforme descrito na Subseção 3.1.3.

As estatísticas descritivas utilizadas para a análise de variabilidade foram o coeficiente de variação (*cv*) e o desvio padrão (*std*). Os valores de mínimo (*min*) e de máximo (*max*) dos resultados da previsão foram utilizados para analisar a exatidão do modelo de previsão. Também é apresentado o valor médio (*med*) dos dados

Os resultados com a utilização das RNAs antes e depois do pré-processamento nos dados *Wolf's sunspot* e *Canadian lynx* são apresentados na Tabela 3.

Tabela 3 – Estatísticas descritivas dos resultados da previsão via RNAs para os conjuntos *Sunspot* e *Lynx* antes e depois da utilização da filtragem após 50 repetições do experimento.

Séries	Estatísticas	Sunspot – 1º caso		Sunspot – 2º caso		Lynx	
		<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>
Sem filtragem	<i>min</i>	148,96	08,66	310,76	12,50	0,0184	0,0953
	<i>med</i>	219,89	11,35	468,74	15,32	0,0411	0,1398
	<i>max</i>	314,85	14,05	794,55	19,08	0,0912	0,2331
	<i>std</i>	40,75	1,18	100,26	1,54	0,0147	0,0292
	<i>cv</i>	18,53%	10,39%	21,38%	10,05%	35,76%	20,88%
Filtragem utilizando	<i>min</i>	145,55	8,09	314,61	14,02	0,0141	0,0937
	<i>med</i>	199	10,31	679,87	18,38	0,0423	0,1476
Wavelet-mãe	<i>max</i>	276,90	12,43	2596,4	27,97	0,0855	0,1781
Haar	<i>std</i>	30,6	1,05	390,00	3,04	0,0112	0,0180
	<i>cv</i>	15,37%	10,18%	57,36%	16,53%	26,47%	12,19%
Filtragem utilizando	<i>min</i>	91,06	6,06	156,71	9,15	0,0121	0,0737
	<i>med</i>	135,21	7,84	342,08	12,84	0,0297	0,1124
Wavelet-mãe	<i>max</i>	202,77	9,85	787,21	18,00	0,0635	0,1760
Daubechies	<i>std</i>	25,63	0,83	142,46	2,15	0,0107	0,0207
	<i>cv</i>	18,95%	10,58%	41,64%	16,74%	36,02%	18,41%
Filtragem utilizando	<i>min</i>	<b>61,28</b>	<b>4,53</b>	<b>109,63</b>	<b>6,99</b>	<b>0,0071</b>	<b>0,0666</b>
	<i>med</i>	74,31	5,26	153,59	8,39	0,0207	0,1129
WavDS	<i>max</i>	97,21	6,08	262,45	10,39	0,0396	0,1584
	<i>std</i>	7,89	0,33	32,77	0,8	0,0055	0,0170
	<i>cv</i>	10,61%	6,27%	21,33%	9,53%	26,57%	15,05%

De acordo com os resultados apresentados na Tabela 3, no primeiro caso do conjunto de dados *Sunspot*, os resultados encontrados com o modelo de previsão via RNAs após a

utilização do pré-processamento nos dados com o método WavDS, em comparação com a utilização do modelo de RNA sem a filtragem, possibilitaram aumento na acurácia em aproximadamente 58,9% no MSE e 47,7% no MAE (de acordo com análise realizada nos valores mínimos).

Neste primeiro caso analisado, pode-se também observar por meio da análise comparativa dos coeficientes de variação que existe redução na variabilidade dos resultados da previsão via RNAs quando utiliza-se o filtro WavDS. A redução foi de 18,53% para 10,61%, resultando assim no aumento em robustez de aproximadamente 42,7% no MSE e 39,6% no MAE.

Na Tabela 3, o pior valor previsto (*max*) observado no MSE encontrado pela RNA com o uso do WavDS é mais próximo do real valor da série quando comparado com o melhor valor previsto (*min*) encontrado com o método de RNAs sem a utilização WavDS em, aproximadamente, 15,5% no conjunto de dados da *Sunspot* para o segundo caso.

Pode-se também observar aumento na acurácia dos resultados encontrado da previsão no conjunto de dados *Canadian Lynx*. Nesta análise, o ganho em exatidão foi de 61,4% para o MSE e 30,1% para o MAE que, por si só, torna o método de filtragem WavDS atrativo.

Os melhores resultados encontrados pelos modelos de RNAs com a utilização do método WavDS estão descritos na Tabela 4 para o conjunto de dados *Exchange rate*. Melhoraram em média o MSE e o MAE em aproximadamente 76,3% e 52,6%, respectivamente.

Tabela 4 – Estatísticas descritivas dos resultados da previsão via RNAs para os conjuntos *Exchange rate* antes e depois da utilização da filtragem após 50 repetições do experimento.

Series		Rate – 1º caso		Rate – 2º caso		Rate – 3º caso	
		<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>
Sem filtragem	<i>min</i>	3,0732	0,0043	4,0526	0,0049	4,0287	0,0048
	<i>med</i>	14,013	0,0080	22,082	0,0107	15,781	0,0080
	<i>max</i>	56,143	0,0174	88,942	0,0262	88,065	0,0180
	<i>std</i>	12,218	0,0031	20,366	0,0050	15,029	0,0030
	<i>cv</i>	87,19%	38,75%	92,22%	46,72%	95,23%	37,5%
Filtragem utilizando	<i>min</i>	1,8541	0,0031	2,5621	0,0037	1,8116	0,0030
	<i>med</i>	2,0298	0,0032	2,9906	0,0041	2,0112	0,0032
Wavelet-mãe	<i>max</i>	2,6605	0,0033	3,3791	0,0043	2,3937	0,0035
Haar	<i>std</i>	0,1319	5,82e-05	0,1734	1,27e-04	0,1131	1,01e-4
	<i>cv</i>	6,49%	1,81%	5,79%	3,09%	5,62%	3,15%
Filtragem utilizando	<i>min</i>	1,1611	0,0024	1,5412	0,0027	1,1344	0,0023
	<i>med</i>	1,2427	0,0025	1,7395	0,0030	1,2566	0,0025
Wavelet-mãe	<i>max</i>	1,3629	0,0027	2,3069	0,0032	1,6613	0,0028
Daubechies	<i>std</i>	0,0427	6,48e-05	0,1137	10,4e-05	0,0841	7,63e-05
	<i>cv</i>	3,43%	2,59%	6,53%	3,46%	6,69%	3,05%
Filtragem utilizando	<i>min</i>	<b>0,9850</b>	<b>0,0023</b>	<b>0,8539</b>	<b>0,0023</b>	<b>0,7204</b>	<b>0,0020</b>
	<i>med</i>	1,0219	0,0024	0,8903	0,0024	0,7344	0,0021
WavDS	<i>max</i>	1,0920	0,0027	1,1805	0,0027	0,8871	0,0022
	<i>std</i>	0,0204	4,8e-5	0,0578	6,2e-5	0,0231	3,01e-05
	<i>cv</i>	1,99%	2%	6,49%	2,58%	3,14%	1,43%

Nota: Todos os valores das estatísticas descritivas para o *MSE* foram multiplicadas por  $10^{-5}$ .

Os melhores resultados da previsão foram observados no 2º e 3º casos, onde pode-se perceber uma melhora em acurácia superior a 78% no caso do MSE e 53% no caso do MAE.

Pode-se observar também alta redução no coeficiente de variação do procedimento de previsão com RNAs. Obtêve-se redução média de aproximadamente 95,5% no MSE e 95,1% MAE nos três casos analisados de acordo com a análise dos coeficientes de variação (*cv*).

Os parâmetros utilizadas pelos modelos de RNAs do tipo MLP são comumente representados da seguinte forma ( $q \times n \times p$ ), de modo que o parâmetro  $q$  representa a quantidade de neurônios na camada de entrada (número de *Lags*),  $n$  representa a quantidade de neurônios na camada escondida e  $p$  a quantidade de neurônios na camada de saída (ZHANG, 2003; KHASHEI; BIJARI, 2011).

Nesta pesquisa os parâmetros utilizados nos modelos de RNAs foram os mesmos utilizados pelos artigos publicados por Zhang (ZHANG, 2003) e Kashei & Bijari (KHASHEI; BIJARI, 2011): ( $4 \times 4 \times 1$ ), ( $7 \times 5 \times 1$ ) e ( $7 \times 6 \times 1$ ) para, respectivamente, *Wolf's sunspot*, *Canadian lynx* e *Exchange rate*.

Por meio da análise realizada a partir dos resultados apresentados na Tabela 3 e na Tabela 4 pode-se perceber que a utilização do método WavDS como procedimento de filtragem, com o indicativo para a escolha da *Wavelet*-mãe e retirada de diferentes tipos de ruído, possibilitou em todas situações analisadas ganho em acurácia, tanto no que diz respeito ao MSE quanto ao MAE, quando comparados aos resultados encontrados a partir da utilização do procedimento de filtragem clássico via *Wavelets* (DONOHO, 1995).

Mediante análise comparativa das metodologias de filtragem que foram realizadas por meio dos resultados encontrados na Tabela 3 e na Tabela 4, vale destacar que o método WavDS possibilitou:

- Superioridade média acima de 60% no MSE nos melhores valores em comparação com os resultados encontrados após a filtragem via *Wavelet Haar* nos dados *Sunspot*;
- Superioridade média acima de 57% no MAE nos melhores valores em comparação com os resultados encontrados após a filtragem via *Wavelet Haar* nos dados *Sunspot*.

Também foi verificado na grande maioria dos casos com a utilização do procedimento de previsão via RNAs que houve redução na variabilidade nos resultados. Dentre demais contribuições, vale-se destacar:

- Redução média na variabilidade dos resultados acima de 30% tanto no MSE quanto no MAE em comparação com os resultados encontrados após a filtragem via *Wavelet Haar* nos dados *Sunspot*;
- Redução na variabilidade dos resultados em, aproximadamente, 44% no MSE e 40% no MAE em comparação com os resultados encontrados após a filtragem via *Wavelet Daubechies* nos dados *Sunspot*.
- Redução na variabilidade dos resultados em, aproximadamente, 69.4% MSE em comparação com os resultados encontrados após a filtragem via *Wavelet Haar* nos dados *Exchange rate* no 1º caso.
- Redução média na variabilidade dos resultados em aproximadamente 53% em comparação com os resultados encontrados após a filtragem via *Wavelet Daubechies* nos dados *Exchange rate* no 3º caso.

De acordo com as situações experimentais descritas na Seção 3.3, o horizonte de previsão para o conjunto de dados *Sunspot* no primeiro caso é 35 dados e no segundo caso

o horizonte é 67 dados. Já o conjunto de dados *Lynx* possui 14 dados que serão utilizados para a análise de previsão.

Os resultados publicados para estas situações na literatura da área (ZHANG, 2003; KHASHEI; BIJARI, 2011) são apresentados na Tabela 5 em conjunto com os melhores resultados obtidos com o modelo de MLP após a filtragem dos dados por meio procedimento WavDS. Para o conjunto de dados *Exchange rate* os resultados publicados (ZHANG, 2003; KHASHEI; BIJARI, 2011) estão apresentados na Tabela 6.

Tabela 5 – Resultados das previsões obtidos pelo modelo de RNAs após filtragem via WavDS e por modelos híbridos para os conjuntos de dados *Sunspot* e *Lynx* na 1ª situação experimental.

Séries	Sunspot – 1º caso		Sunspot – 2º caso		Lynx	
	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>
ARIMA	216,965	11,39	306,08	13,03	0,02048	0,1122
RNA	205,302	10,243	351,19	13,54	0,02046	0,1121
ZHANG (2003)	186,82	10,83	280,15	12,78	0,017	0,1039
KASHEI & BIJARI (2011)	129,42	8,847	218,64	11,44	0,009	0,085
WANG (2013)	95,50	5,678	214,92	7,139	0,018	0,1035
WavDS+RNA	<b>61,28</b>	<b>4,53</b>	<b>109,63</b>	<b>6,99</b>	<b>0,007</b>	<b>0,066</b>

Tabela 6 – Resultados das previsões obtidos pelo modelo de RNAs após filtragem via WavDS e por modelos híbridos para os conjuntos de dados *Exchange rate* na 1ª situação experimental.

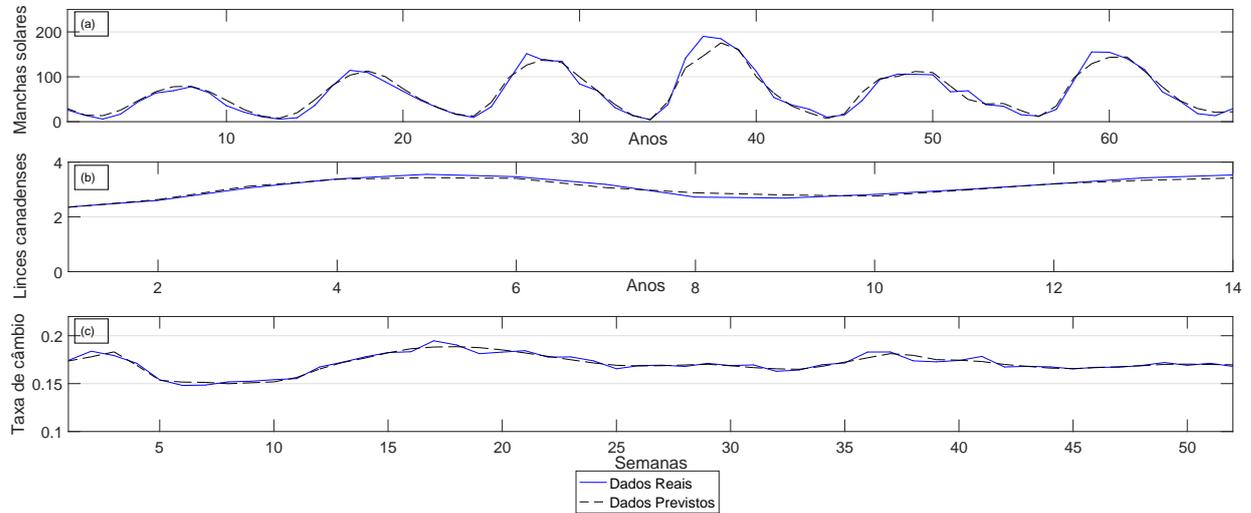
Séries	Rate – 1º caso		Rate – 2º caso		Rate – 3º caso	
	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>	<i>MSE</i>	<i>MAE</i>
ARIMA	3,684	0,005	5,657	0,0060	4,529	0,0053
RNA	2,763	0,004	5,710	0,0059	4,526	0,0052
ZHANG (2003)	2,672	0,0041	5,655	0,0058	4,359	0,0052
KASHEI & BIJARI (2011)	2,399	0,0039	4,278	0,0053	3,647	0,0049
WavDS+RNA	<b>0,985</b>	<b>0,002</b>	<b>0,853</b>	<b>0,0023</b>	<b>0,720</b>	<b>0,0020</b>

Nota: Todos os valores das estatísticas descritivas para o *MSE* foram multiplicadas por  $10^{-5}$ .

De acordo com os resultados apresentados na Tabela 5 e na Tabela 6, pode-se perceber que os modelos de RNAs com a filtragem via WavDS apresentaram resultados superiores em acurácia tanto no MSE quanto no MAE, quando comparados os seus resultados com os resultados divulgados na literatura da área (ZHANG, 2003; KHASHEI; BIJARI, 2011; WANG et al., 2013; BABU; REDDY, 2014; BABU; SURE, 2016) para o caso dos dados *Sunspot* e *Lynx*.

Na Figura 28 a seguir são ilustradas as previsões encontradas pelo método de filtragem WavDS combinado com o modelo de RNAs para os três conjuntos de dados.

Figura 28 – (a) Resultados da previsão do modelo de RNAs após filtragem via WavDS: (a) *Sunspot* no 2º caso, (b) *Lynx*, (c) *Exchange rate* no 3º caso.



Fonte: Próprio autor.

Conforme pôde ser observado, os melhores resultados encontrados no que diz respeito a melhora na acurácia da previsão nas séries temporais com a utilização do método proposto WavDS foram evidenciados no caso da série *Exchange rate* que são de natureza financeiras e são bastante ruidosos (ABABNEH; WADI; ISMAIL, 2013; WADI; ISMAIL, 2011; KILIÇ; UĞUR, 2016).

Contudo, pode-se evidenciar nesta pesquisa que o método de pré-processamento proposto WavDS possibilitou, com a abordagem inovadora para a escolha da *Wavelet*-mãe e a retirada de diversos tipos de ruído, que as previsões realizadas com o modelo de RNA obtivessem ganho em acurácia e robustez, sendo estas as principais contribuições do método.

### 4.3 Análise do método híbrido MTAEF

A análise comparativa da exatidão da previsão dos resultados encontrados por meio do método proposto MTAEF, do método de pré-processamento WavDS em conjunto com a utilização de RNAs e com recentes resultados divulgados na literatura da área (ZHANG, 2003; KHASHEI; BIJARI, 2011; WANG et al., 2013; BABU; REDDY, 2014; FIRMINO; DE MATTOS NETO; FERREIRA, 2015; BABU; SURE, 2016) são apresentados nesta seção.

Entres estes, destacam-se trabalhos quem envolvem a temática de que o erro na previsão de séries temporais possui informação e, em virtude disto, também deve ser

modelado e incluído no modelo preditivo (DE MATTOS NETO et al., 2017; FIRMINO; DE MATTOS NETO; FERREIRA, 2015; FIRMINO; DE MATTOS NETO; FERREIRA, 2014).

Nesta linha de pesquisa, Firmino, De Mattos Neto e Ferreira (2015) apresentam um procedimento que melhora os resultados de modelos preditores mostrando que o resíduo do modelo possui informação e, descartá-lo, pode levar a perda em acurácia na previsão do modelo.

Em sua pesquisa, os autores fizeram uma análise comparativa entre as previsões dos modelos ARIMA, RNA e TAEF sem a correção dos resíduos, modelo com resíduos corrigidos pelo modelo ARIMA recursivo, entre eles o TAEF (TAEFa) e modelos corrigidos por modelo MLPs, entre eles o TAEF (TAEFm)(FIRMINO; DE MATTOS NETO; FERREIRA, 2015).

Estes modelos foram analisados em vários conjuntos de dados, entre eles os conjuntos *Sunspot* e *Lynx* e, em virtude disto, estes resultados estão apresentados na Tabela 7 e foram escolhidos para serem analisados em conjunto com o modelo MTAEF proposto de modo que os melhores resultados estão destacados em negrito.

Tabela 7 – Resultados das previsões obtidos pelos métodos MTAEF e modelos híbridos para os conjuntos de dados *Sunspot* na 2ª situação experimental.

Séries	Métricas	ARIMA	RNA	TAEF	TAEFa	TAEFm	WavDS+RNA	MTAEF
Sunspot	MSE	0,0133	0,0346	0,0118	0,0053	0,0064	0,0073	<b>0,0027</b>
	POCID	73,33%	73,33%	<b>100%</b>	93,33%	93,33%	80,28%	92,62 %
	Theil	0,5002	1,2995	0,4414	0,1971	0,2402	<b>0,0076</b>	0,1443
	MAPE	39,95%	52,67%	14,44%	11,16%	<b>8,38%</b>	40,43%	16,64%
	ARV	0,2268	0,6575	0,3561	0,1111	0,1326	0,1433	<b>0,0487</b>
Lynx	MSE	<b>0,0024</b>	0,0091	0,0124	0,0112	0,0087	0,0061	0,0037
	POCID	<b>100%</b>	83,33%	83,33%	83,33%	83,33%	87,5%	79,16%
	Theil	1,3606	0,9934	1,3607	1,2215	0,9542	2,52	<b>0,2332</b>
	MAPE	24,23%	62,68%	47,34%	54,15%	43,53%	17,01%	<b>10,95%</b>
	ARV	0,3698	0,6724	0,3574	0,3411	0,2665	0,1392	<b>0,0937</b>

Outro motivo para a realização da análise comparativa destes resultados, com os resultados provindos do modelo MTAEF, é o fato dos autores (FIRMINO; DE MATTOS NETO; FERREIRA, 2015) utilizarem o método TAEF (FERREIRA; VASCONCELOS; ADEODATO, 2008) individualmente e em conjunto com outras técnicas que visam otimizar os resultados de um modelo preditivo. Desta forma tem-se uma oportunidade que possibilita a comparação entre modelos híbridos de previsão e métodos otimizadores de preditores.

Por meio dos resultados apresentados na Tabela 7 é possível ter uma comparação direta entre todos os resultados alcançados e pode-se perceber que o modelo MTAEF teve

desempenho global superior aos demais modelos em virtude de uma maior quantidade de métricas com desempenho melhor que os demais modelos.

Em uma análise comparativa entre os modelos MTAEF e TAEF, destacam-se alguns resultados:

- No caso dos dados *sunspot* pode-se observar no MSE, por exemplo, um ganho percentual em acurácia de aproximadamente 77,1%. Com exceção da métrica POCID, o MTAEF superou o modelo TAEF em todas as métricas analisadas.
- No caso dos dados *Lynx* pode-se observar na estatística U de Theil, por exemplo, que o modelo MTAEF é aproximadamente 82,6% mais bem ajustado aos dados.

Quando a comparação é realizada entre os modelos MTAEF e TAEFa (TAEF com correção dos erros pelo modelo ARIMA recursivo), pode-se observar na Tabela 7 que o modelo MTAEF foi melhor em sete dentre dez oportunidades distintas. No caso dos dados *Sunspot* as métricas foram: MSE, estatística U de Theil, e a medida ARV. Já no caso dos dados *Lynx* as métricas com superior resultados foram o MSE, a estatística U de Theil, o erro MAPE e o ARV. Com base na análise destes resultados podemos destacar algumas interpretações:

- A medida ARV, que compara o desempenho do modelo com a previsão média da série, revela que o modelo fornece resultados superiores do que a utilização do valor médio da série;
- A estatística U de Theil, que gera um limiar de comparação com um modelo do tipo *Random walk*, mostra que o modelo aplicado à série tem desempenho melhor.

Já quando a análise comparativa é realizada entre os modelos MTAEF e TAEFm (TAEF com correção dos erros por RNAs do tipo MLP), pode-se perceber por meio da análise realizada nos dados descritos na Tabela 7 que o modelo MTAEF foi melhor do que o modelo TAEFm nas mesmas sete distintas oportunidades em que foi melhor do que o modelo TAEFa e com um menor custo computacional.

Desta maneira, em virtude dos casos analisados na Tabela 7, temos que o método MTAEF apresentou resultados mais robustos e com maior acurácia quando comparados aos resultados provindos de modelo TAEF (FERREIRA; VASCONCELOS; ADEODATO, 2008).

Outra importante conclusão desta análise na Tabela 7 é que o procedimento de filtragem WavDS em conjunto com RNAs (WavDS+RNA) possibilitou melhores resultados

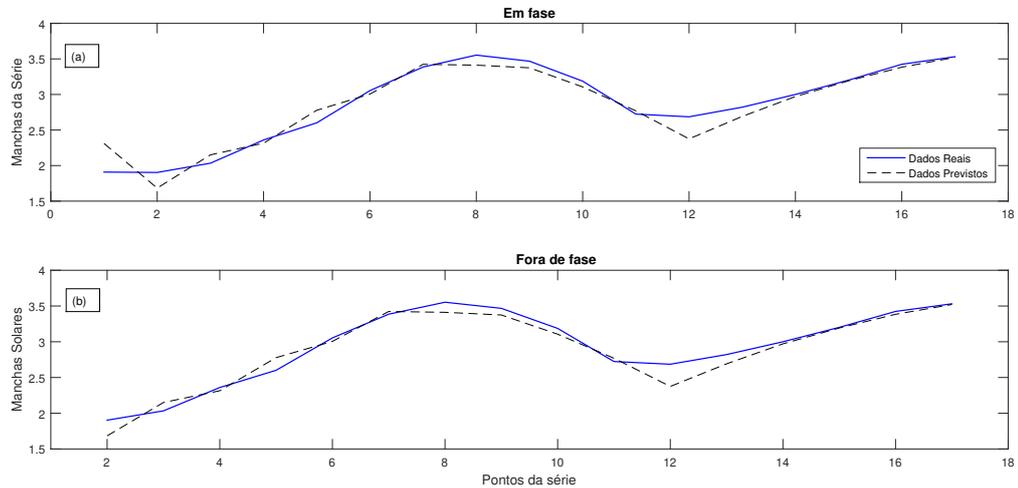
em diversas oportunidades, quando comparados os seus aos resultados publicados por Fimino et al. (FIRMINO; DE MATTOS NETO; FERREIRA, 2015):

- **WavDS+RNA vs. TAEF:** A filtragem nos dados por meio do filtro WavDS possibilitou melhora em acurácia em diversas métricas para os diferentes dados. Nos dados *Lynx*, por exemplo, a superioridade é a mais perceptível. Apenas na métrica U de Theill o modelo não foi superior;
- **WavDS+RNA vs. TAEFa:** No conjunto de dados *Lynx* o WavDS possibilitou, com exceção da métrica U de Theil, que o modelo de RNAs fosse melhor que o modelo TAEFa.
- **WavDS+RNA vs. TAEFm:** Em comparação realizada com os resultados provindos do modelo TAEFm, a análise é semelhante à anterior (TAEFa). O filtro WavDS possibilitou, com exceção da métrica U de Theil, que o modelo de RNAs fosse melhor.
- **WavDS+RNA vs. MTAEF:** Apenas na estatística U de Theil, para o conjunto *Lynx*, os resultados foram melhores.

Assim como descrito na Subseção 3.1.5, o método MTAEF avalia a melhor rede que deve ser utilizada pelo modelo de RNAs e o algoritmo de treinamento. Para os dados *Sunspot* a arquitetura da rede foi  $(9 \times 7 \times 1)$  com o algoritmo de treinamento Gradiente conjugado escalado e, para os dados *Lynx*, a arquitetura da rede foi  $(7 \times 7 \times 1)$  com o algoritmo de treinamento *Levenberg marquardt* selecionado para treinamento.

Outra análise realizada pelo método MTAEF é no que diz respeito ao ajuste de fase, por meio de uma avaliação realizada no conjunto de validação da RNA para saber em qual procedimento (Em fase ou Fora de fase) o modelo é melhor ajustado aos dados. O modelo realiza o teste de hipóteses *t-student* para identificar em quais das situações o modelo é melhor ajustado aos dados. Em ambos os casos analisados (*Sunspot* e *Lynx*) o teste de hipóteses *t-student* comprovou que não foi necessário o ajuste de fase.

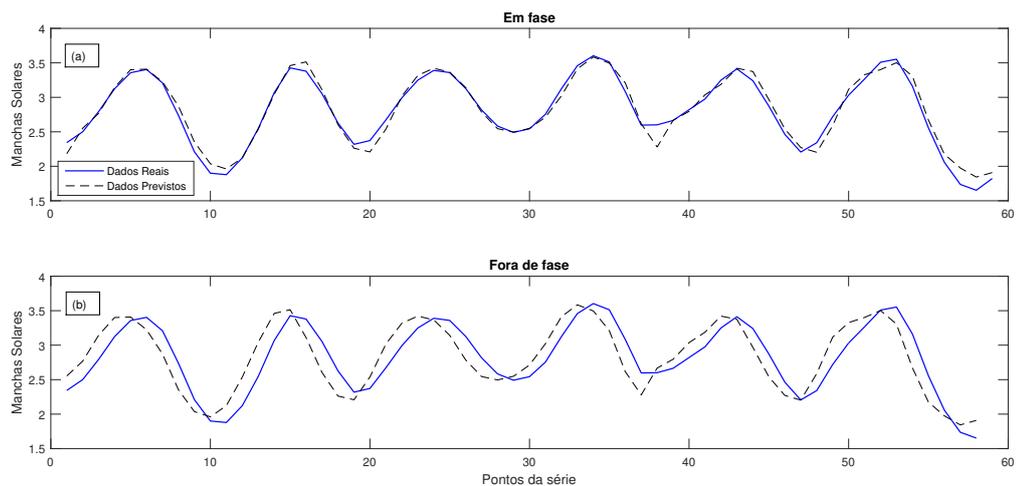
Figura 29 – Gráfico comparativo das previsões do método de MTAEF para os dados sem ajuste de fase - caso (a); e com ajuste de fase - caso (b), para o conjunto de teste dos dados *Lynx*



Fonte: Próprio autor.

Na Figura 29 são ilustrados os gráficos das previsões Em fase - caso (a) e Fora de fase - caso (b), para o conjunto de dados *Lynx*, propostas pelo modelo MTAEF.

Figura 30 – Gráfico comparativo do ajuste do método MTAEF para os dados sem ajuste de fase (a) e com ajuste de fase (b) para o conjunto de treino dos dados *Lynx*.

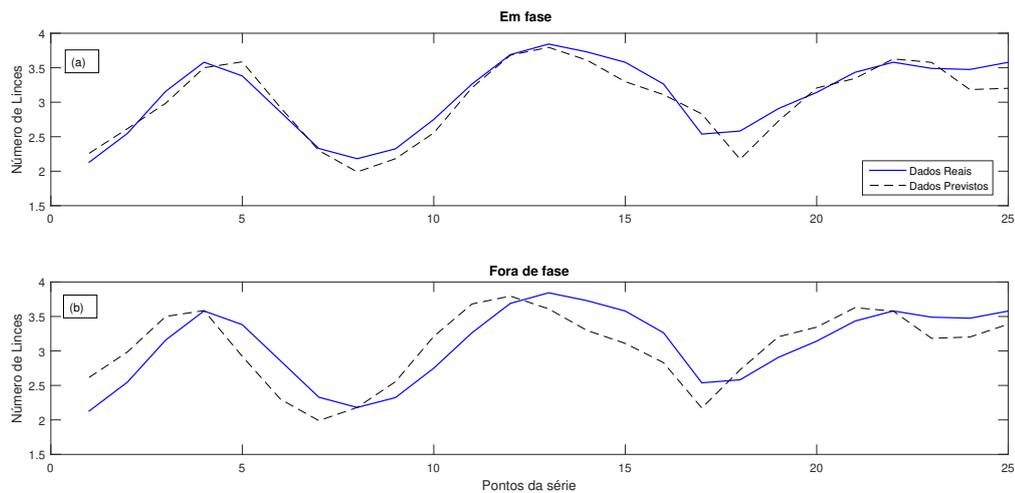


Fonte: Próprio autor.

Por meio da análise da Figura 29 não é possível identificar em qual das opções o modelo gera uma melhor previsão, contudo, como descrito anteriormente o teste de

hipóteses *t-student* comprovou que não foi necessário o ajuste de fase. Tal comprovação pode ser visualizada por meio da Figura 30 e na Figura 31. Elas representam os ajustes do modelo nos conjuntos de validação e teste que são utilizados pelos modelos de RNAs.

Figura 31 – Gráfico comparativo do ajuste do método MTAEF para os dados sem ajuste de fase (a) e com ajuste de fase (b) para o conjunto de validação dos dados *Lynx*.



Fonte: Próprio autor.

A próxima análise da segunda situação experimental foi realizada com base nos artigos publicados por Babu & Reddy (BABU; REDDY, 2014) e Babu & Sure (BABU; SURE, 2016). Os resultados do experimento realizado, em conjunto com os resultados publicados nos artigos, são apresentados na Tabela 8.

Tabela 8 – Resultados das previsões obtidos pelos métodos WavDS+RNA, MTAEF e modelos híbridos para os conjuntos de dados *Sunspot* na 2ª situação experimental.

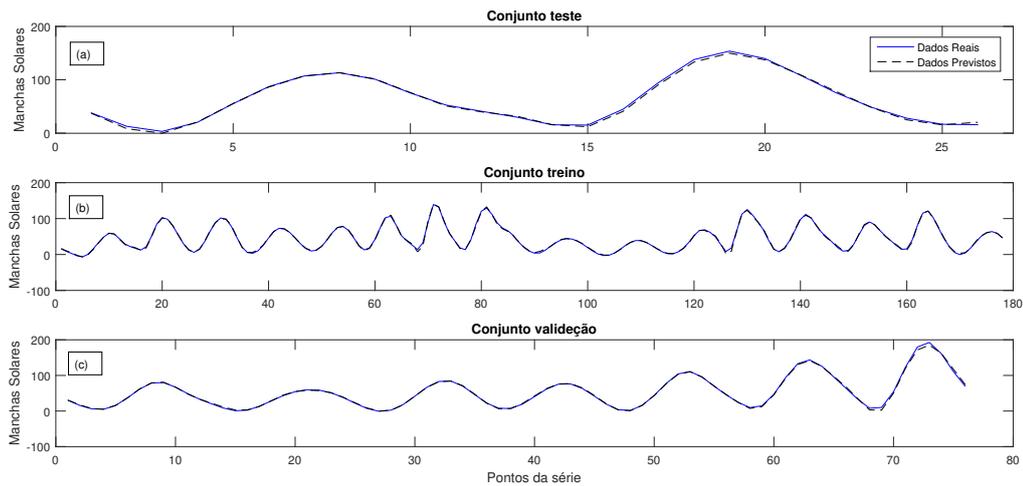
Modelos	Sunspot - 4º caso	
	<i>MSE</i>	<i>MAE</i>
ARIMA	308,1	14,663
RNA	285,1	14,351
ZHANG (2003)	298,5	14,223
KASHEI & BIJARI (2011)	269,5	13,405
WANG (2013)	298,0	14,211
BABU & REDDY (2014)	155,5	9,8718
BABU & SURE (2016)	95,55	8,1604
WavDS+RNA	88,51	7,3113
MTAEF	<b>6,103</b>	<b>1,8546</b>

De acordo com os resultados que estão apresentados na Tabela 8, pode-se observar

que o modelo que combina WavDS com RNAs proporcionou melhores resultados do que os resultados publicados por Babu & Sure (BABU; SURE, 2016). Considerando-se o baixo custo computacional empregado pelo método (WavDS+RNAs), este é um dos principais resultados encontrados nesta pesquisa.

Contudo, os melhores resultados foram, de fato, encontrados pelo método MTAEF. O ganho percentual em acurácia, quando comparados os resultados do método MTAEF com o resultado publicado por Babu & Reddy (BABU; SURE, 2016) foi de, aproximadamente, 93,61% no MSE (redução de 95,5 para 6,103) e 77,27% no MAE (redução de 95,5 para 1,8546). O gráfico que ilustra o desempenho do modelo para este caso está apresentado na Figura 32.

Figura 32 – Gráficos dos resultados da previsão sem ajuste de fase para o conjunto de dados *Sunspot* - 4º caso.



Fonte: Próprio autor.

O modelo MTAEF, assim como nos resultados apresentados na Tabela 6, detectou que não foi necessário o ajuste de fase. Para este caso, a rede selecionada foi  $(9 \times 7 \times 1)$  e o algoritmo de treinamento foi o Gradiente conjugado escalado. Nesta análise, também foi observado que o modelo proposto MTAEF atendeu às expectativas (minimizar o erro na previsão) uma vez que busca concomitantemente minimizar os problemas relativos à previsão de séries temporais via MLPs (filtragem dos dados e seleção de parâmetros).

De acordo com os dados apresentados na Tabela 8, Tabela 6 e Tabela 5, o procedimento WavDS foi eficiente. Proporcionou melhora nos resultados obtidos pelas RNAs e superou a acurácia encontrada em resultados publicados (ZHANG, 2003; KHASHEI; BIJARI, 2011; WANG et al., 2013; BABU; REDDY, 2014; BABU; SURE, 2016).

## 5 Considerações Finais

Neste capítulo as considerações finais da Tese são apresentadas. Primeiramente, na Seção 5.1, um breve resumo da problemática envolvendo séries temporais é descrito. Nesta seção as principais contribuições do modelo são apresentadas. Na Seção 5.2 é apresentada a principal limitação encontrada e o direcionamento para as atividades futuras.

### 5.1 Conclusões

A previsão do comportamento futuro de séries temporais com base na evolução histórica das próprias séries é um problema de previsão clássico, com interessantes aplicações práticas em diversas áreas. A busca por combinações de modelos para aumentar a precisão nos resultados na previsão está se tornando um foco por parte de muitos pesquisadores. Estes modelos combinados, denominados híbridos, buscam promover previsores superiores aos modelos individuais por usufruir especificamente de cada uma das características pertinentes aos dados.

Contudo, de maneira geral, as séries temporais podem ser consideradas um processo determinístico acrescido de ruído. Tendo em vista as diversas caracterizações que podem ocorrer em um ruído, a filtragem nos dados não é uma alternativa trivial. Outra questão de extrema relevância é a variabilidade existente nos resultados obtidos com a utilização dos modelos de RNAs. Esta variabilidade existente pode conduzir à interpretações equivocadas de seus resultados uma vez que na realização de um experimento em uma análise o resultado pode ser satisfatório e em outra não. Em virtude disto, nesta Tese, estes temas foram abordados por serem fundamentais na problemática que envolve previsão de séries temporais.

O MTAEF é um método que, conjuntamente, visa atender as seguintes questões: filtragem e previsão. Busca explorar o potencial de cada técnica retirando o ruído presente na série e possibilitando prever os resultados futuros com estabilidade e alto grau de precisão.

A tese foi conduzida baseada na hipótese de que as séries temporais são influenciadas por ruído aditivo e esta hipótese foi confirmada por meio das análises que foram realizada sob as óticas de obtenção de estabilidade estatística no comportamento fundamental da série. Esta estabilidade foi verificada por meio da análise da variabilidade dos resultados,

e da obtenção de maior precisão na utilização de um procedimento de previsão após a utilização do filtro WavDS também aqui proposto. Com isto a detecção de padrões comportamentais e extração de importantes informações implícitas em virtude do ruído presente nas séries temporais foram verificadas. Dentre as principais contribuições desta Tese, merecem destaque:

- Desenvolvimento de um novo procedimento de filtragem *Wavelet* que possui um indicativo para escolha da *Wavelet*-mãe que será utilizada na DW que possibilita a retirada de diversos tipos de ruídos;
- Inovação na metodologia de filtragem em dados de séries temporais que possibilita que o método de previsão via RNAs possa ser desenvolvido em um processo que possibilite resultados estatisticamente mais estáveis e, conseqüentemente, mais confiáveis;
- Nova abordagem híbrida que combina técnicas de filtragem via *Wavelets* e previsão de séries temporais.

O MTAEF foi comparado com o método individual de RNAs, com a utilização do método individual de RNAs após filtragem tradicional via *Wavelets*, com a utilização do método individual de RNAs após filtragem via o filtro WavDS e com resultados encontrados por meio de tradicionais modelos híbridos que foram publicados na literatura da área em diversos diferentes cenários. Com base nas análises, os resultados encontrados corroboraram a abordagem proposta.

## 5.2 Atividades Futuras

Na primeira etapa do procedimento WavDS, no que diz respeito a escolha da *Wavelet*-mãe e seleção da combinação dos diversos parâmetros, observou-se que não há apenas uma combinação de *Wavelet*-mãe e parâmetros que possibilitou a retirada do ruído gaussiano branco na primeira etapa do procedimento, mais sim algumas combinações deles. Em alguns casos analisados os resultados finais das previsões eram muito próximos e não interferiram na eficácia do modelo. A maneira com a qual deve-se realizar a investigação da influência na escolha das *Wavelets*-mãe e demais parâmetros na 1ª etapa do modelo foi a principal limitação encontrada em virtude da quantidade de opções disponíveis de *Wavelets*-mãe e demais parâmetros que atendem às exigências do método WavDS proposto.

Contudo, dada a confirmação de que a abordagem com o filtro WavDS proporcionou resultados com superior precisão em relação às técnicas tradicionais de filtragem via

*Wavelets*, o trabalho futuro a ser desenvolvido é refinar o procedimento realizado nesta pesquisa. Faz-se necessário uma análise mais sofisticada do comportamento do ruído provindo da 1ª e da 2ª etapa do método WavDS, por meio da caracterização do ruído, para verificar a possibilidade de melhorar a acurácia encontrada com a utilização do método MTAEF como sugerido por [Firmino, De Mattos Neto e Ferreira \(2015\)](#).

## Referências Bibliográficas

- ABABNEH, F.; WADI, S. A.; ISMAIL, M. T. Haar and daubechies wavelet methods in modeling banking sector. v. 8, p. 551–566, 2013.
- ADAMOWSKI, J.; CHAN, H. F. A wavelet neural network conjunction model for groundwater level forecasting. **Journal of Hydrology**, Elsevier, v. 407, n. 1, p. 28–40, 2011.
- AL-QAZZAZ, N. K. et al. Selection of mother wavelets thresholding methods in denoising multi-channel eeg signals during working memory task. In: IEEE. **Biomedical Engineering and Sciences (IECBES), 2014 IEEE Conference on**. [S.l.], 2014. p. 214–219.
- ALI, M. N.; EL-DAHSHAN, E.-S. A.; YAHIA, A. H. Denoising of heart sound signals using discrete wavelet transform. **Circuits, Systems, and Signal Processing**, Springer, p. 1–16, 2017.
- ARAÚJO, R.; FERREIRA, T. A. An intelligent hybrid morphological-rank-linear method for financial time series prediction. **Neurocomputing**, Elsevier, v. 72, n. 10-12, p. 2507–2524, 2009.
- ASADI, S. et al. Hybridization of evolutionary levenberg–marquardt neural networks and data pre-processing for stock market prediction. **Knowledge-Based Systems**, Elsevier, v. 35, p. 245–258, 2012.
- BABU, C. N.; REDDY, B. E. A moving-average filter based hybrid arima–ann model for forecasting time series data. **Applied Soft Computing**, Elsevier, v. 23, p. 27–38, 2014.
- BABU, C. N.; SURE, P. Partitioning and interpolation based hybrid arima–ann model for time series forecasting. **Sādhanā**, Springer, v. 41, n. 7, p. 695–706, 2016.
- BOX, G. E. et al. **Time series analysis: forecasting and control**. [S.l.]: John Wiley & Sons, 2015.
- BULLMORE, E. et al. Colored noise and computational inference in neurophysiological (fmri) time series analysis: resampling methods in time and wavelet domains. **Human brain mapping**, Wiley Online Library, v. 12, n. 2, p. 61–78, 2001.
- CATALÃO, J. P. d. S.; POUSINHO, H. M. I.; MENDES, V. M. F. Short-term wind power forecasting in portugal by neural networks and wavelet transform. **Renewable energy**, Elsevier, v. 36, n. 4, p. 1245–1251, 2011.
- CHANDAR, S. K.; SUMATHI, M.; SIVANANDAM, S. Prediction of stock market price using hybrid of wavelet transform and artificial neural network. **Indian Journal of Science and Technology**, v. 9, n. 8, 2016.

DE MATTOS NETO, P.; FERREIRA, T. A. Applying a general hybrid intelligent system for ultra-high-frequency stock market forecasting. In: IEEE. **Neural Networks (IJCNN), 2016 International Joint Conference on**. [S.l.], 2016. p. 2104–2109.

DE MATTOS NETO, P. et al. A perturbative approach for enhancing the performance of time series forecasting. **Neural Networks**, Elsevier, v. 88, p. 114–124, 2017.

DE MATTOS NETO, P.; JUNIOR, A. R. L.; FERREIRA, T. A. Time series forecasting using a perturbative intelligent system. In: ACM. **Proceedings of the 12th annual conference on Genetic and evolutionary computation**. [S.l.], 2010. p. 1477–1478.

DE MATTOS NETO, P. et al. Hybrid intelligent system for air quality forecasting using phase adjustment. **Engineering Applications of Artificial Intelligence**, Elsevier, v. 32, p. 185–191, 2014.

DEMUTH, H.; BEALE, M. Matlab neural network toolbox user's guide version 6. the mathworks inc. Citeseer, 2009.

DONATE, J. P. et al. Time series forecasting by evolving artificial neural networks with genetic algorithms, differential evolution and estimation of distribution algorithm. **Neural Computing and Applications**, Springer, v. 22, n. 1, p. 11–20, 2013.

DONOHO, D. L. De-noising by soft-thresholding. **IEEE transactions on information theory**, IEEE, v. 41, n. 3, p. 613–627, 1995.

DONOHO, D. L.; JOHNSTONE, I. M. Ideal spatial adaptation by wavelet shrinkage. **biometrika**, JSTOR, p. 425–455, 1994.

FERREIRA, T. A.; VASCONCELOS, G. C.; ADEODATO, P. J. A new intelligent system methodology for time series forecasting with artificial neural networks. **Neural Processing Letters**, Springer, v. 28, n. 2, p. 113–129, 2008.

FERREIRA, T. A. E. **Uma nova metodologia híbrida inteligente para a previsão de séries temporais**. Tese (Doutorado) — Universidade Federal de Pernambuco, 2006.

FIRMINO, P. R. A.; DE MATTOS NETO, P.; FERREIRA, T. A. Correcting and combining time series forecasters. **Neural networks**, Elsevier, v. 50, p. 1–11, 2014.

FIRMINO, P. R. A.; DE MATTOS NETO, P.; FERREIRA, T. A. Error modeling approach to improve time series forecasters. **Neurocomputing**, Elsevier, v. 153, p. 242–254, 2015.

FLORES, J. J. et al. Comparison of time series forecasting techniques with respect to tolerance to noise. In: IEEE. **Power, Electronics and Computing (ROPEC), 2016 IEEE International Autumn Meeting on**. [S.l.], 2016. p. 1–6.

GANNET, E. Glossary of telecommunications terms,” federal standard 1037c. **Information Technology Service (ITS)**, v. 7, 1996.

HAYKIN, S. **Redes neurais: princípios e prática**. [S.l.]: Bookman Editora, 2007.

HOLLAND, J. H. **Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence**. [S.l.]: MIT press, 1992.

JARQUE, C. M.; BERA, A. K. A test for normality of observations and regression residuals. **International Statistical Review/Revue Internationale de Statistique**, JSTOR, p. 163–172, 1987.

JOO, T. W.; KIM, S. B. Time series forecasting based on wavelet filtering. **Expert Systems with Applications**, Elsevier, v. 42, n. 8, p. 3868–3874, 2015.

JOY, J.; PETER, S.; JOHN, N. Denoising using soft thresholding. **International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering**, v. 2, n. 3, p. 1027–1032, 2013.

JR, F. J. M. The kolmogorov-smirnov test for goodness of fit. **Journal of the American statistical Association**, Taylor & Francis Group, v. 46, n. 253, p. 68–78, 1951.

KAO, L.-J. et al. A hybrid approach by integrating wavelet-based feature extraction with mars and svr for stock index forecasting. **Decision Support Systems**, Elsevier, v. 54, n. 3, p. 1228–1244, 2013.

KHANDELWAL, I.; ADHIKARI, R.; VERMA, G. Time series forecasting using hybrid arima and ann models based on dwt decomposition. **Procedia Computer Science**, Elsevier, v. 48, p. 173–179, 2015.

KHASHEI, M.; BIJARI, M. A novel hybridization of artificial neural networks and arima models for time series forecasting. **Applied Soft Computing**, Elsevier, v. 11, n. 2, p. 2664–2675, 2011.

KILIÇ, D. K.; UĞUR, Ö. Multiresolution analysis of s&p500 time series. **Annals of Operations Research**, Springer, p. 1–20, 2016.

KIŞI, Ö.; UNCUOĞLU, E. Comparison of three back-propagation training algorithms for two case studies. CSIR, 2005.

LILLIEFORS, H. W. On the kolmogorov-smirnov test for normality with mean and variance unknown. **Journal of the American statistical Association**, Taylor & Francis Group, v. 62, n. 318, p. 399–402, 1967.

LIU, C.; LU, B.; WANG, J. Traffic flow prediction based on wavelet analysis, improved genetic algorithm, and bp neural network. In: **CICTP 2012: Multimodal Transportation Systems—Convenient, Safe, Cost-Effective, Efficient**. [S.l.: s.n.], 2012. p. 1126–1133.

LIU, H. et al. Forecasting models for wind speed using wavelet, wavelet packet, time series and artificial neural networks. **Applied Energy**, Elsevier, v. 107, p. 191–208, 2013.

LIU, T. et al. Time series forecasting based on wavelet decomposition and feature extraction. **Neural Computing and Applications**, Springer, p. 1–13, 2016.

- MALLAT, S. G. A theory for multiresolution signal decomposition: the wavelet representation. **IEEE transactions on pattern analysis and machine intelligence**, Ieee, v. 11, n. 7, p. 674–693, 1989.
- MINU, K.; LINEESH, M.; JOHN, C. J. Wavelet neural networks for nonlinear time series analysis. **Applied Mathematical Sciences**, v. 4, n. 50, p. 2485–2495, 2010.
- MISITI, M. et al. Wavelet toolbox. **The MathWorks Inc., Natick, MA**, v. 15, p. 21, 1996.
- MORETTIN, P. A. **Ondas e Ondaletas Vol. 24**. [S.l.]: Edusp, 2014.
- MUSTAFA, Q.; ALZUBAYDI, T. H. Comparing the box-jenkins models before and after the wavelet filtering in terms of reducing the orders with application. **Journal of Concrete & Applicable Mathematics**, v. 11, n. 2, 2013.
- NALBAND, S.; SREEKRISHNA, R.; PRINCE, A. A. Analysis of knee joint vibration signals using ensemble empirical mode decomposition. **Procedia Computer Science**, Elsevier, v. 89, p. 820–827, 2016.
- OLIVEIRA, J. F. de; LUDERMIR, T. B. A hybrid evolutionary decomposition system for time series forecasting. **Neurocomputing**, Elsevier, v. 180, p. 27–34, 2016.
- PANDEY, A. S.; SINGH, D.; SINHA, S. K. Intelligent hybrid wavelet models for short-term load forecasting. **IEEE Transactions on Power Systems**, IEEE, v. 25, n. 3, p. 1266–1273, 2010.
- PRECHELT, L. et al. Proben1: A set of neural network benchmark problems and benchmarking rules. Technical Report 21/94, 1994.
- RODRIGUEZ, N.; BARBA, L. Fejer-korovkin wavelet based mimo model for multi-step-ahead forecasting of monthly fishes catches. In: **15th Mexican International Conference on Artificial Intelligence (MICAI), At Mexico**. [S.l.: s.n.], 2016.
- SANG, Y.-F.; WANG, D.; WU, J.-C. Uncertainty analysis of decomposition level choice in wavelet threshold de-noising. **Entropy**, Molecular Diversity Preservation International, v. 12, n. 12, p. 2386–2396, 2010.
- SCHOLZ, F. W.; STEPHENS, M. A. K-sample anderson–darling tests. **Journal of the American Statistical Association**, Taylor & Francis Group, v. 82, n. 399, p. 918–924, 1987.
- SEO, Y. et al. Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. **Journal of Hydrology**, Elsevier, v. 520, p. 224–243, 2015.
- SIDNEY, B. C. **Introduction to wavelets and wavelet transforms: a primer**. [S.l.]: Prentice Hall International, 1998.
- SOLTANI, S. On the use of the wavelet decomposition for time series prediction. **Neurocomputing**, Elsevier, v. 48, n. 1, p. 267–277, 2002.

SOUZA, F. de et al. Comparação das bases de wavelets ortonormais e biortogonais: implementação, vantagens e desvantagens no posicionamento com gps. **Trends in Applied and Computational Mathematics**, v. 8, n. 1, p. 149–158, 2007.

TAKENS, F. Detecting strange attractors in turbulence. In: **Dynamical systems and turbulence, Warwick 1980**. [S.l.]: Springer, 1981. p. 366–381.

TEIXEIRA, L. L. **Projeção de séries temporais por meio de um método híbrido wavelet-neural integrado com bootstrap**. Tese (Doutorado) — Universidade Federal do Paraná, 2015.

VALENCIA, D. et al. Comparison analysis between rigrsure, sqtwolog, heursure and mini-maxi techniques using hard and soft thresholding methods. In: IEEE. **Signal Processing, Images and Artificial Vision (STSIVA), 2016 XXI Symposium on**. [S.l.], 2016. p. 1–5.

VERMA, N.; VERMA, A. Performance analysis of wavelet thresholding methods in denoising of audio signals of some indian musical instruments. **International Journal of Engineering Science and Technology**, v. 4, n. 5, p. 2040–2045, 2012.

WADI, M. A.; ISMAIL, M. T. Selecting wavelet transforms model in forecasting financial time series data based on arima model. **Applied Mathematical Sciences**, v. 5, n. 7, p. 315–326, 2011.

WANG, L. et al. An arima-ann hybrid model for time series forecasting. **Systems Research and Behavioral Science**, Wiley Online Library, v. 30, n. 3, p. 244–259, 2013.

WANG, W.-c. et al. Improving forecasting accuracy of annual runoff time series using arima based on eemd decomposition. **Water Resources Management**, Springer, v. 29, n. 8, p. 2655–2675, 2015.

WANG, X.; ZHANG, X.; ZHOU, G. Wavelet transform used to preprocessing of apple near infrared spectra. **Journal of Residuals Science & Technology**, v. 13, n. 7, 2016.

WU, Z.; HUANG, N. E. Ensemble empirical mode decomposition: a noise-assisted data analysis method. **Advances in adaptive data analysis**, World Scientific, v. 1, n. 01, p. 1–41, 2009.

ZHANG, C.; YANG, Q.; LIU, B. Guest editors' introduction: Special section on intelligent data preparation. **IEEE Transactions on Knowledge and Data Engineering**, IEEE, v. 17, n. 9, p. 1163–1165, 2005.

ZHANG, G. P. Time series forecasting using a hybrid arima and neural network model. **Neurocomputing**, Elsevier, v. 50, p. 159–175, 2003.

## Apêndices

## APÊNDICE A – Gráficos das Funções Wavelets ( $\psi$ ) utilizadas nesta pesquisa

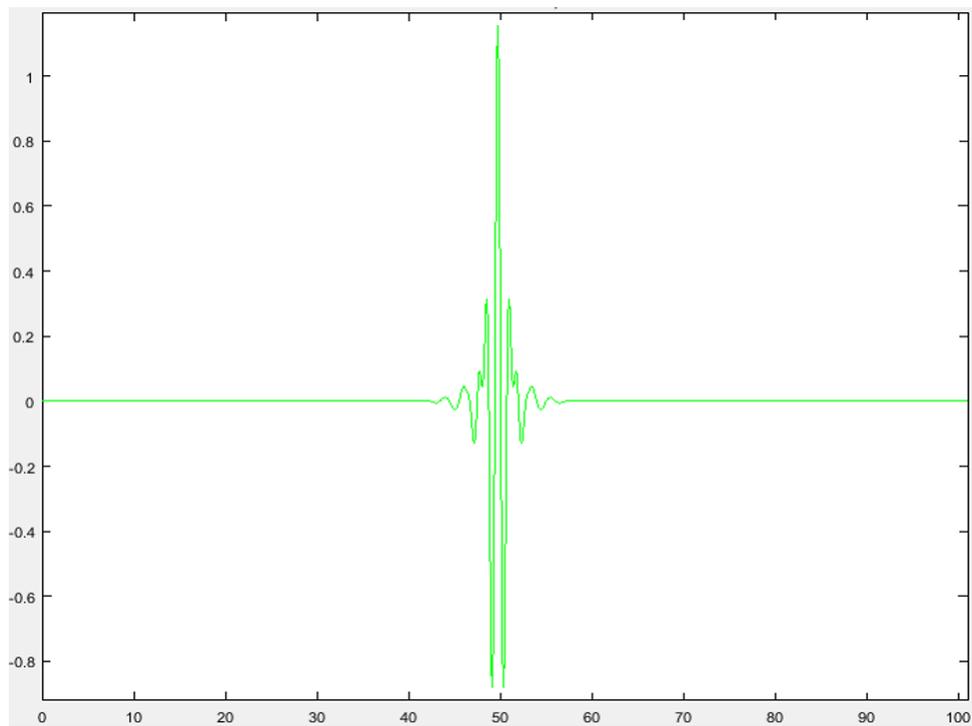


Figura 33 – Gráfico ilustrativo da Função *Wavelet* Discreta de Meyer

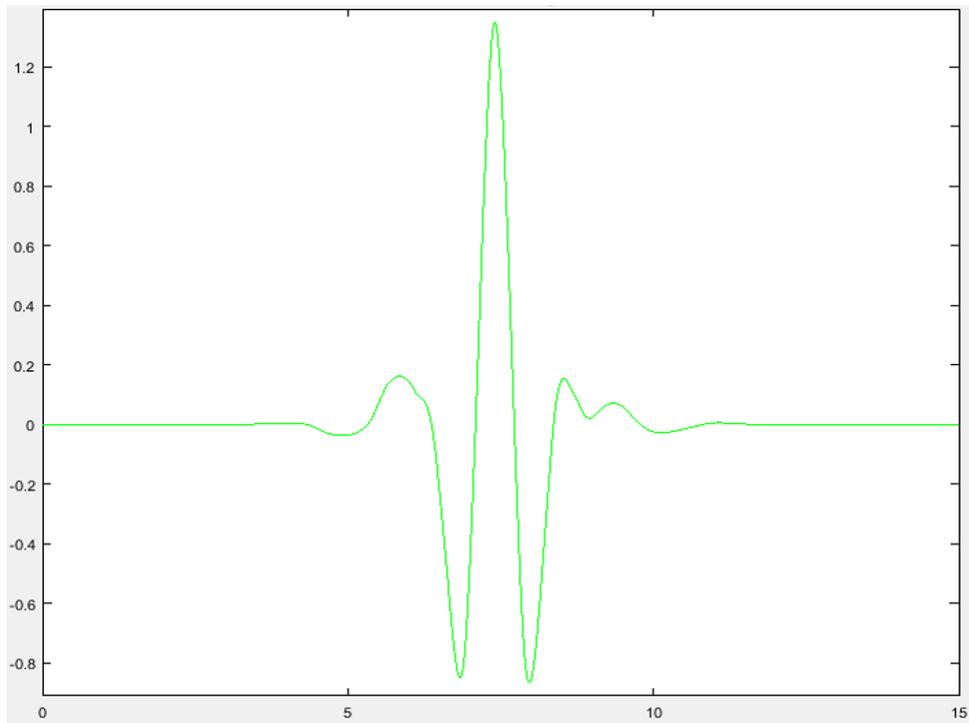


Figura 34 – Gráfico ilustrativo da Função *Wavelet* Symlet 8

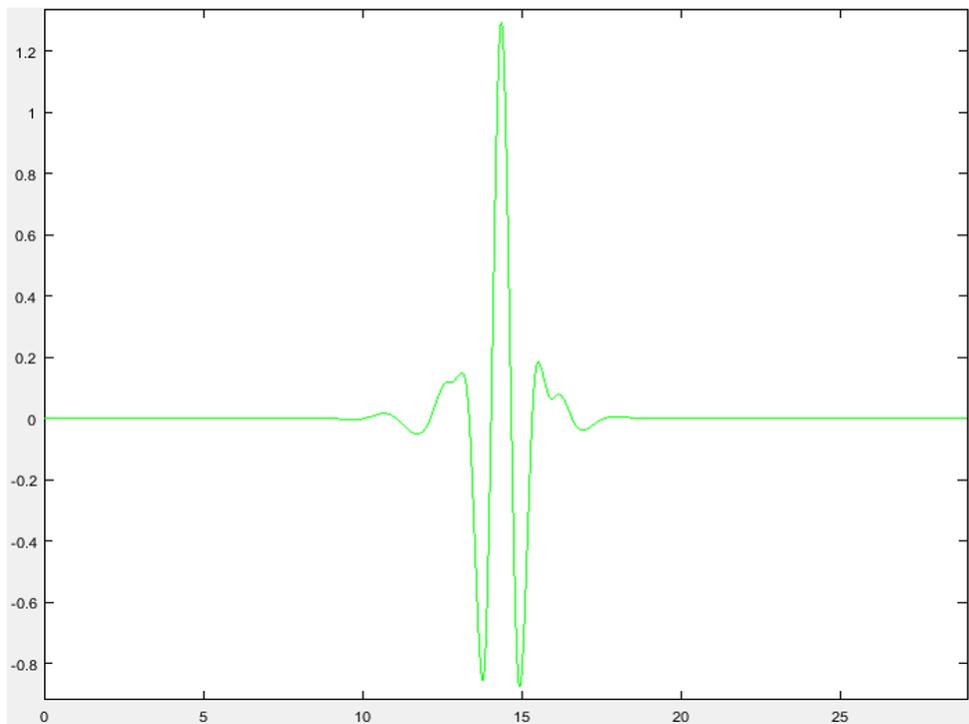


Figura 35 – Gráfico ilustrativo da Função *Wavelet* Coiflet 5

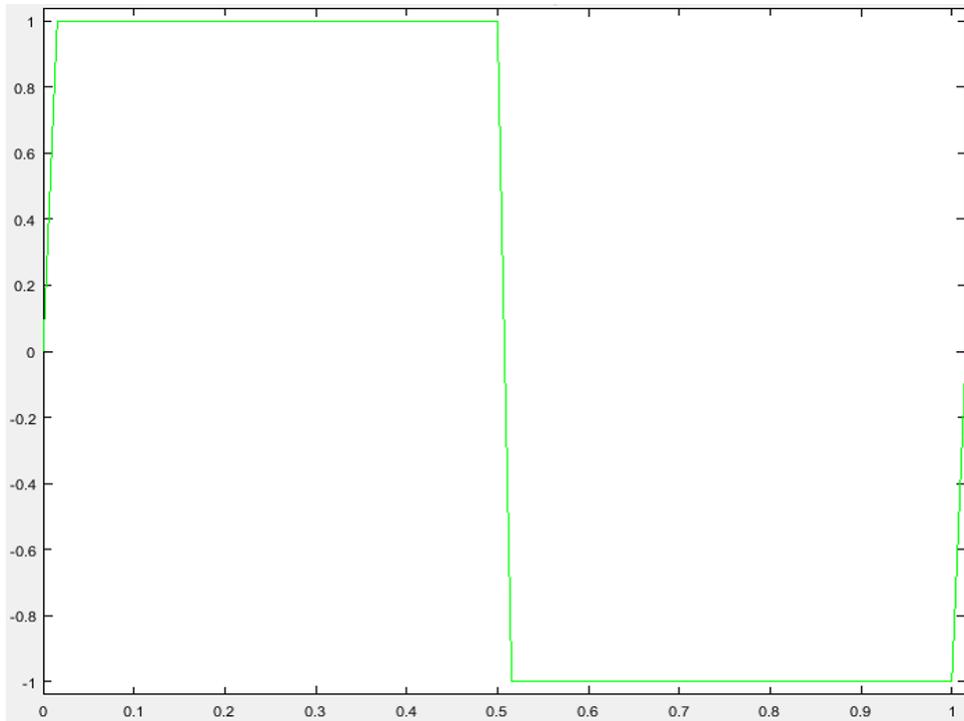


Figura 36 – Gráfico ilustrativo da Função *Wavelet* Haar

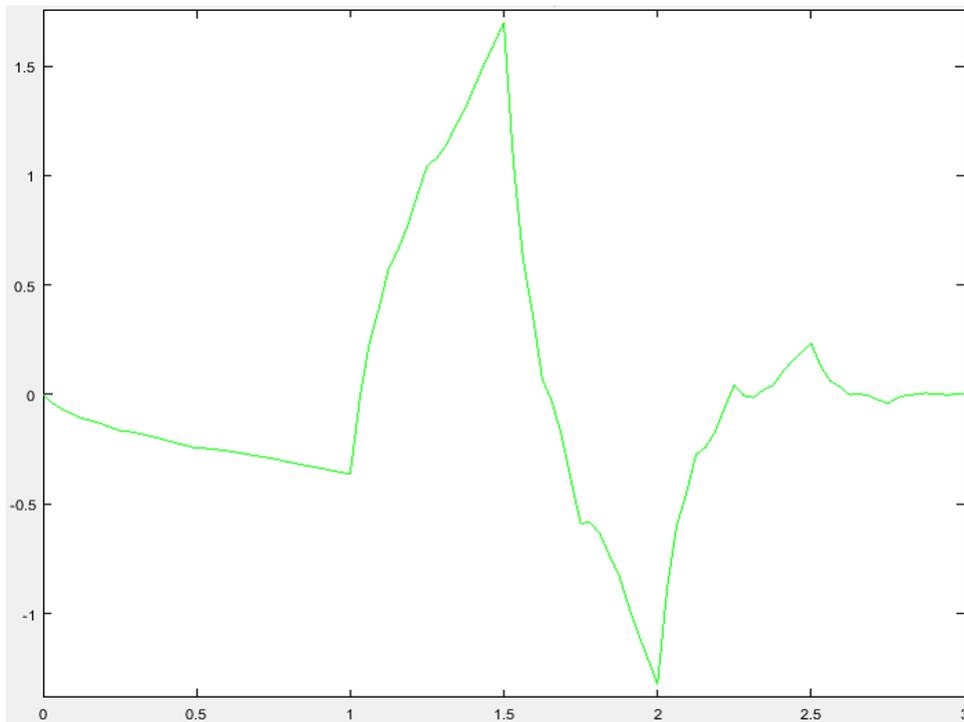


Figura 37 – Gráfico ilustrativo da Função *Wavelet* Daubechies 2

## APÊNDICE B – Gráficos das Funções Escala ( $\phi$ ) utilizadas nesta pesquisa

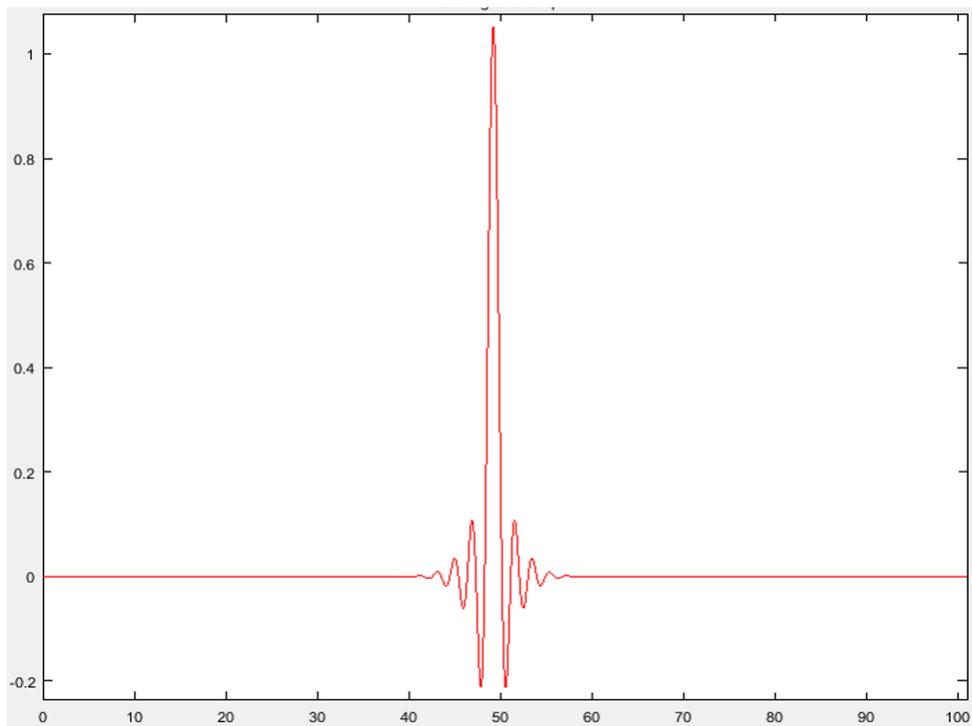


Figura 38 – Gráfico ilustrativo da Função Escala Discreta de Meyer

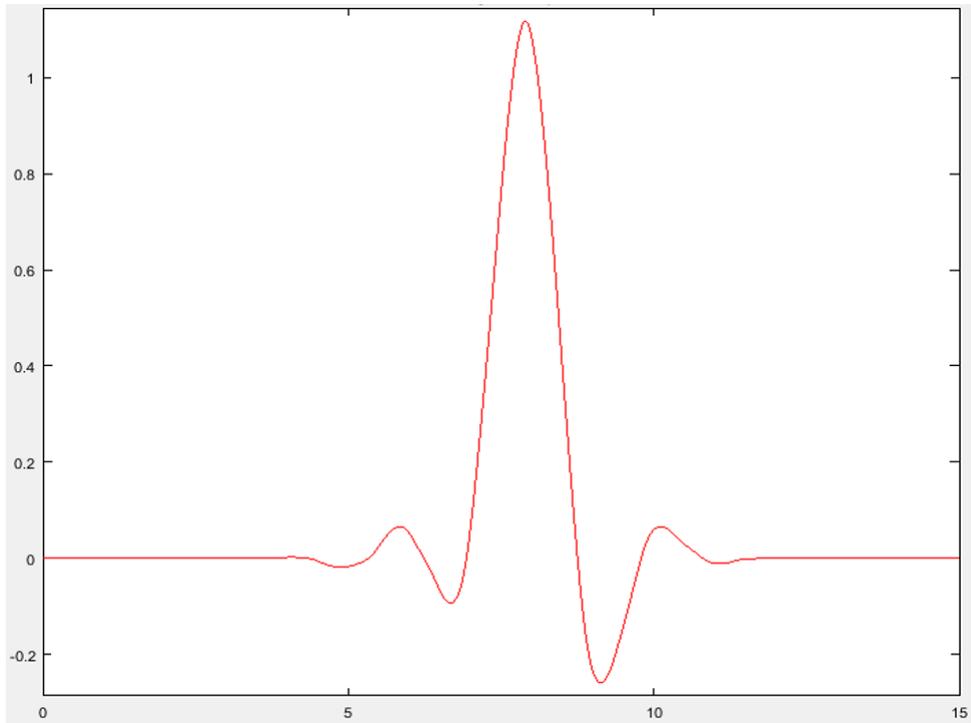


Figura 39 – Gráfico ilustrativo da Função Escala Symlet 8

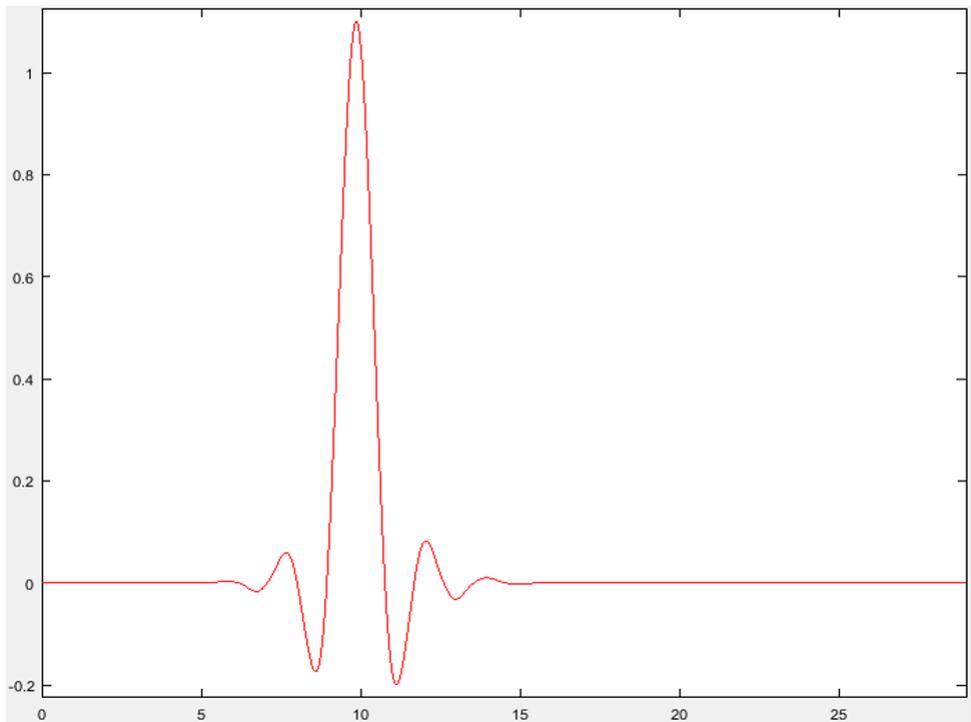


Figura 40 – Gráfico ilustrativo da Função Escala Coiflet 5

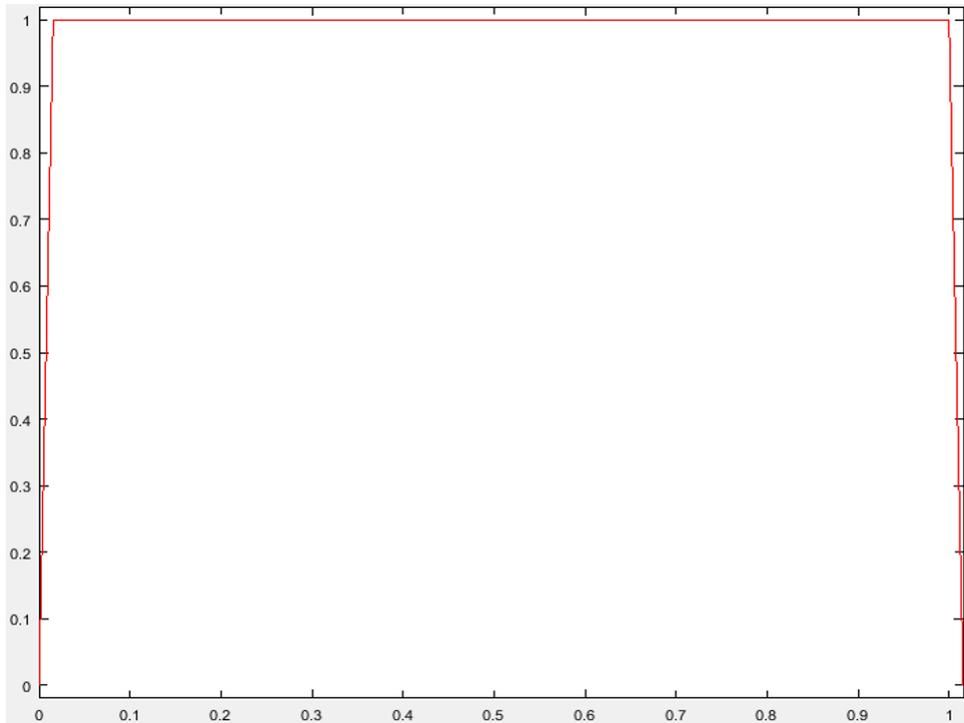


Figura 41 – Gráfico ilustrativo da Função Escala Haar

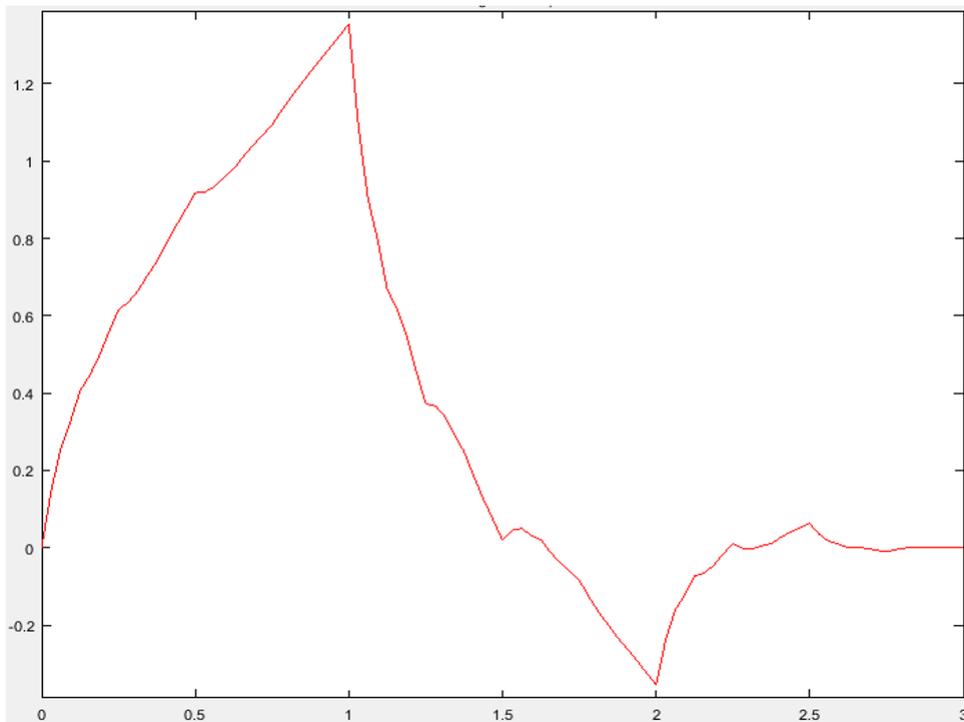


Figura 42 – Gráfico ilustrativo da Função Escala Daubechies 2

# Anexos

# ANEXO A – Evolução da população dos Algoritmos genéticos

## A.1 O Teorema dos Esquemas

Seja  $m(s, t)$  o número de instâncias de um esquema  $s$  em uma população no tempo  $t$ . o Teorema dos esquemas irá descrever o valor esperado de  $m(s, t + 1)$  em termos de  $m(s, t)$ . A evolução da população de um AG depende dos procedimentos de seleção, combinação (cruzamento) e mutação (HOLLAND, 1992).

Considerando inicialmente o processo de seleção, seja  $f(h)$  a função *fitness* (Equação 29) de uma *string*,  $\bar{f}(t)$  a média da função de *fitness* sobre todas as *strings* da população no tempo  $t$  e  $N$  o número total de indivíduos da população. Sendo  $h \in S \cap P_t$  (com  $h$  um indivíduo que seja representado pelo esquema  $S$  e membro da população do AG no tempo  $t$  ( $P_t$ )), seja  $\hat{u}(s, t)$  o *fitness* médio das instâncias do esquema  $S$  no tempo  $t$ . Deseja-se calcular o valor esperado de  $m(s, t + 1)$ , representado na Equação 37 por  $E[m(s, t + 1)]$ .

Assumindo que a probabilidade de seleção seja dada por:

$$Pr(h) = \frac{f(h)}{\sum_{i=1}^N f(h_i)} = \frac{f(h)}{N\bar{f}(t)}, \quad (37)$$

a probabilidade de seleção de um representante do esquema  $S$  é:

$$Pr(h \in S) = \sum_{h \in (S \cap P_t)} \frac{f(h)}{N\bar{f}(t)} = \frac{\hat{u}(s, t)}{N\bar{f}(t)} m(s, t). \quad (38)$$

Contudo, por definição,

$$\hat{u}(s, t) = \frac{\sum_{h \in (S \cap P_t)} f(h)}{m(s, t)} \quad (39)$$

o que leva a interpretar a Equação 37 como a probabilidade de que um indivíduo seja selecionado pelo AG, sendo este a instância do esquema  $S$ . Desta forma, o valor esperado do número de instâncias de  $S$  de  $N$  passos independentes de seleção, que irá criar a nova geração, é exatamente  $N$  vezes a Equação 37:

$$E[m(s, t + 1)] = \frac{\hat{u}(s, t)}{\bar{f}(t)} m(s, t) \quad (40)$$

A Equação 40 mostra que o número de instâncias de  $S$ , sobre a geração  $t + 1$  é proporcional ao fitness médio  $\hat{u}(s, t)$  das instâncias deste mesmo esquema no tempo  $t$ , e inversamente proporcional ao fitness médio  $\bar{f}(t)$  de todos os membros da população no tempo  $t$ .

Entretanto, a Equação 40 não leva em consideração as operações de cruzamento e mutação. Todas as duas operações podem tanto criar novas instâncias de  $S$  como também podem destruir as existentes. Considerando apenas o efeito de destruição de instâncias, o Teorema dos Esquemas terá um limite inferior para o valor esperado da frequência dos esquemas  $S$ , dado por:

$$E[m(s, t + 1)] \geq \frac{\hat{u}(s, t)}{\bar{f}(t)} m(s, t) \left(1 - p_c \frac{d(S)}{l - 1}\right) (1 - p_m)^{o(S)} \quad (41)$$

de modo que  $p_c$  é a probabilidade do operador de cruzamento (um ponto) ser aplicado a um indivíduo arbitrário e  $p_m$  a probabilidade que um bit arbitrário de um indivíduo qualquer seja alterado (mutação). O termo  $o(S)$  é o número de bits definidos no esquema  $S$ ,  $d(S)$  é a distância entre o bit definido mais à esquerda e o mais à direita no esquema  $S$  e, por fim,  $l$  é o comprimento de um indivíduo da população.

A Equação 41 é o próprio teorema dos Esquemas, onde o lado direito é composto por três termos. O primeiro, idêntico ao da Equação 40, reflete o efeito do processo de seleção, o segundo termo (termo central do lado direito) descreve o efeito do operador de cruzamento (probabilidade de um indivíduo representado por um esquema  $S$  ainda ser representado pelo mesmo esquema após a aplicação do operador de cruzamento). Por fim, o último termo descreve a probabilidade de um indivíduo representado pelo esquema  $s$  ainda ser representado pelo mesmo esquema após a aplicação da operação mutação.