



Ricardo Vela de Britto Pereira

**Modelagem Híbrida Wavelet Integrada com Bootstrap
na Projeção de Séries Temporais**

Tese de Doutorado

Tese apresentada ao programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio como parte dos requisitos parciais para obtenção do título de Doutor em Engenharia Elétrica.

Orientador: Prof. Reinaldo Castro Souza

Rio de Janeiro
Agosto de 2015



Ricardo Vela de Britto Pereira

Modelagem Híbrida Wavelet Integrada com Bootstrap na Projeção de Séries Temporais

Tese apresentada como requisito parcial para obtenção do grau de Doutor pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Reinaldo Castro Souza
Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Carlos Kubrusly

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Fernando Luiz Cyrino Oliveira

Departamento de Engenharia Industrial – PUC-Rio

Prof. José Francisco Moreira Pessanha
UERJ

Prof. Nival Nunes de Almeida
PUC-Rio

Prof. José Eugenio Leal
Coordenador Setorial do Centro
Técnico Científico

Rio de Janeiro, 14 de agosto de 2015

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Ricardo Vela de Britto Pereira

Possui graduação em Estatística pela Escola Nacional de Ciências Estatísticas (2006) e mestrado em Física pela Pontifícia Universidade Católica do Rio de Janeiro (2011). Atuou pesquisando os seguintes temas: Processos Estocásticos, Séries Temporais, Volatilidade, Equação de Ito-Langevin, Equação de Fokker Planck e Econofísica

Ficha Catalográfica

Pereira, Ricardo Vela de Britto

Modelagem Híbrida Wavelet Integrada com Bootstrap na Projeção de Séries Temporais / Ricardo Vela de Britto Pereira; orientador: Reinaldo Castro Souza – 2015.

104 f. ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, 2015.

Inclui bibliografia

1. Engenharia elétrica – Teses. 2. Wavelet. 3. Redes Neurais Artificiais. 4. Combinação de Previsões. 5. Otimização. 6. Bootstrap. I. Souza, Reinaldo Castro. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 621.3

Resumo

Britto Pereira, Ricardo Vela; Souza, Reinaldo Castro (Orientador). **Modelagem Híbrida Wavelet Integrada com Bootstrap na Projeção de Séries Temporais**. Rio de Janeiro, 2015. 104p. Tese de Doutorado - Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

Na previsão de séries temporais, alguns autores supõem que um método de previsão individual (por exemplo, um modelo ARIMA) produz resíduos (ou erros de previsão) semelhantes a um processo de ruído branco (imprevisível). No entanto, principalmente devido às estruturas de autodependência não mapeadas por um método preditivo individual, tal suposição pode ser facilmente violada na prática. Esta tese propõe um Previsor Híbrido Wavelet (PHW) que integra as seguintes técnicas: decomposição wavelet; modelos ARIMA; redes neurais artificiais (RNAs); combinação de previsões; programação matemática não linear e amostrador Bootstrap. Em termos gerais, o PHW proposto aqui é capaz de capturar, ao mesmo tempo, estruturas com autodependência linear por meio de uma combinação linear wavelet (CLW) de modelos ARIMA, (cujo ajuste numérico ótimo ocorre por programação matemática não linear) e não linear (usando uma RNA wavelet automática) exibidas pela série de tempo a ser predita. Diferentemente de outras abordagens híbridas existentes na literatura, as previsões híbridas produzidas pela PHW proposto levam em conta implicitamente, através da abordagem de decomposição wavelet, as informações oriundas da frequência espectral presentes na série temporal subjacente. Os resultados estatísticos mostram que a metodologia híbrida supracitada alcançou ganhos de precisão relevantes no processo preditivo de quatro séries de tempo diferentes bem conhecidas, quando se compara com outras meteorologistas competitivas.

Palavras-chave

Wavelet; Redes Neurais Artificiais; Box & Jenkins; Combinação de Previsões; Otimização; Bootstrap; Previsões.

Abstract

Britto Pereira, Ricardo Vela; Souza, Reinaldo Castro (Advisor). **Modeling Hybrid Wavelet Integrated with Bootstrap in Projection Temporal Series**. Rio de Janeiro, 2015. 104p. DSc. Thesis – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

In time series analysis some authors presume that a single model (an ARIMA for instance) may yield white noise errors. However that assumption can be easily violated, especially in scenarios where unmapped auto dependency structures are present inside the series. With that being said, this thesis proposes a new approach called Hybrid Wavelet Predictor (HWP) which integrates the following techniques: Wavelet Decomposition, ARIMA models, Neural Networks (NN), Combined Prediction, Non-linear mathematical programming and Bootstrap Sampling. In a broad sense, the proposed HWP is able to capture not only the linear auto-dependent structures from ARIMA using linear wavelet combination (where its optimal numerical adjustment is made through non-linear mathematical programming), but also the non-linear structures by using Neural Network. Differently from others hybrid approaches known to date, the hybrid predictions given by HWP model take into account. Statistical tests show that the hybrid approach stated above increased the prediction's effectiveness by a significant amount when compared with four well known processes.

Keywords

Wavelet; Artificial Neural Network; Box and Jenkins; Combination of Forecasts; optimization; Bootstrap; Forecasts.

Sumário

1. INTRODUÇÃO	11
1.1. Problema	13
1.2. Objetivos	14
1.2.1. Objetivo Geral	14
1.2.2. Objetivos Específicos	14
1.3. Validação da Proposta	15
1.4. Estrutura da Tese	15
2. REVISÃO DA LITERATURA	16
2.1. Análise Wavelet	16
2.1.1. Preliminares	16
2.1.2. Funções wavelet	17
2.1.3. Transformadas wavelet	18
2.1.4. Decomposição wavelet	19
2.2. Modelos de Box e Jenkins	22
2.2.1. Modelos autorregressivos	25
2.2.2. Modelos médias móveis	25
2.2.3. Modelos Autorregressivos e de Médias Móveis	26
2.2.4. Modelos ARIMA	27
2.2.5. Modelos SARIMA	27
2.2.6. Identificação de modelos	28
2.3. Redes Neurais Artificiais	29
2.3.1. Componentes Básicas de uma RNA	32
2.3.1.1. Função de Ativação	32
2.3.1.2. Arquitetura	33
2.3.2. Treinamento neural	34
2.3.3. Normalização dos padrões	36
2.3.4. RNA's multicamadas feedforward, com uma camada escondida	38
2.3.5. Geração de previsões via redes neurais artificiais	39
2.4. Combinação de métodos preditivos individuais	40
2.4.1. Combinação Linear de Métodos Preditivos	41
2.4.1.1. Combinação linear de previsões	41
2.4.1.2. Constantes e pesos adaptativos	43
2.5. Amostrador Bootstrap	45
2.5.1. Intervalos de confiança e de previsão	46
3. METODOLOGIA PROPOSTA	50
3.1. RNA-MLP Wavelet Automática	51
4. RESULTADOS EMPÍRICOS	56
4.1. Conjuntos de dados	56
4.2. Modelagem	59
4.2.1. Modelagem da série temporal Sunspot	60
4.2.2. Modelagem da série temporal <i>Canadian lynx</i>	67
4.2.3. Modelagem da série temporal Exchange Rate	73
4.3. Previsões pontuais e intervalares para os dados de Itaipu	75

4.3.1. O problema da previsão de vazões	75
4.3.2. Caracterização do conjunto de dados	78
4.3.3. Métodos preditivos individuais	80
4.3.4. Métodos preditivos individuais com wavelet	83
4.3.5. Método híbrido proposto: resultados empíricos	85
5. CONCLUSÕES	94
6. SUGESTÕES PARA TRABALHOS FUTUROS	96
7. REFERÊNCIAS BIBLIOGRÁFICAS	97

Lista de figuras

Figura 1 - Funções Wavelet da Família Symelet 20.	20
Figura 2 - Decomposição wavelet de nível 1, com a base ortogonal wavelet sym 20, da série temporal mensal de vazão de afluentes da região sul do Brasil.	21
Figura 3 - Esquematização simplificada de um neurônio biológico.	30
Figura 4 - Arquitetura básica de um neurônio artificial.	31
Figura 5 - RNA MLP <i>feedforward</i> .	38
Figura 6 - Ilustração da dinâmica de previsão, 1 passo à frente, de uma série temporal, com estrutura de autodependência igual a 4, através de uma RNA.	39
Figura 7 - Combinação da RNA com Decomposição <i>Wavelet</i> .	52
Figura 8 - Fluxograma.	55
Figura 9 - Série temporal Manchas Solares (1700-1987).	56
Figura 10 - Annual time series of Canadian lynx (1821-1934).	57
Figura 11 - Série temporal semanal Exchange Rate (1980-1993).	58
Figura 12 - CWs da amostra de treinamento da série temporal (anual) Sunspot.	61
Figura 13 - CW dos erros de previsão da CLW ótima do passo 3.	65
Figura 14 - Previsões híbridas e observações, na amostra de teste.	66
Figura 15 - CW dos erros de previsão da CLW ótima do passo 3.	69
Figura 16 - Previsões e realizações do logaritmo na base de Euler da série temporal Canadian lynx.	71
Figura 17 - Gráfico de dispersão entre as previsões da metodologia proposta e das respectivas observações, na amostra de teste.	72
Figura 18 - Série de vazões mensais obtidas no posto 266-Itaipu no período 1970-2010.	79

Figura 19 - Correlograma da série de vazões mensais.	79
Figura 20 - Autocorrelações dos resíduos do modelo $ARIMA(1,0,2)X(0,1,2)_{12}$.	81
Figura 21 - Diagramas de dispersão entre a amostra de teste e previsões por RNA e ARIMA.	82
Figura 22 - Fluxograma com as etapas do método ARIMA-wavelet.	83
Figura 23 - Gráfico da série temporal de vazões e das CWs A2, D2 e D1.	84
Figura 24 - Gráfico de dispersão entre as CWs A2, D2 e D1 e as suas respectivas previsões.	87
Figura 25 - Resíduos da CLW do passo 3 e suas respectivas previsões.	89
Figura 26 - Realizações e previsões da série temporal de vazão, na amostra de teste (out of sample).	89
Figura 27 – Valores de APE dos previsores CL2 e do método proposto, na amostra de teste.	91
Figura 28 - Diagrama de dispersão entre a amostra de teste e previsões pelo método proposto.	92
Figura 29 - Intervalo de previsão para a série de vazões – amostra de teste.	93

Lista de tabelas

Tabela 1 - Composição das amostras de treino e de teste, nos três conjuntos de dados.	58
Tabela 2 - Estimativas das autocorrelações simples e parciais dos erros de previsão.	63
Tabela 3 - Gráficos da FAC e PACP e p-valores associados aos resíduos ordinários, na amostra de treino.	64
Tabela 4 - Configuração ótima da RNA wavelet automática.	66
Tabela 5 - Estatísticas MSE e MAE, na amostra de teste.	67
Tabela 6 - Configuração ótima da RNA wavelet automática associada à série temporal <i>Canadian lynx</i> .	70
Tabela 7 - Comparativo MSE e MAE, fora da amostra, entre métodos de previsão aplicados à série temporal <i>Canadian Lynx</i> .	72
Tabela 8 - Comparativo entre as estatísticas MSE e MAE entre métodos de previsão aplicados à série temporal log <i>Exchange Rate</i> .	74
Tabela 9 - APE mínimo, máximo e MAPE para seis estruturas distintas de RNA-MLP.	82
Tabela 10 - Configurações RNA e ordens dos modelos AR e MA para A2, D1 e D2.	84
Tabela 11 - Configuração ótima da RNA wavelet automática para projetar os resíduos da CLW ótima, do passo 3.	88
Tabela 12 - Estatísticas de aderência MAPE e MAE.	90

INTRODUÇÃO

Ao longo dos anos, inúmeros *métodos preditivos* (ou *previsores*) vêm sendo propostos com a finalidade de se prever séries temporais. Em linhas gerais, podem-se agrupá-los em três categorias: os estatísticos (como, por exemplo, os *modelos de Box & Jenkins* (2001)¹, os de inteligência artificial (como, por exemplo, as *redes neurais artificiais* (RNAs) (Haykin, 2001)² e os previsores de combinação (como, por exemplo, a combinação linear de previsões oriundas de um modelo de Box & Jenkins e de uma RNA) (Teixeira Jr., 2013)³.

É usual assumir que os resíduos (ou erros preditivos ou de previsão) produzidos por métodos de previsão individuais (lineares ou não lineares) são uma realização de processo ruído branco (caracterizando-os, assim, como não previsíveis). Não obstante, devido principalmente às estruturas de autodependência não mapeadas (explicadas) pelo mesmo, tal suposição, segundo Zhang (2003)⁴, pode ser facilmente violada. Nesta perspectiva, DAGUM, (1980)⁵ ainda destaca que modelos lineares ARIMA (que podem ser interpretados como filtros lineares) são capazes fazer somente o mapeamento de estruturas de autodependência linear; de maneira que podem garantir nível de significância, ausência de estrutura de autocorrelação nos erros de previsão, mas não de estruturas de autodependência não lineares, como, frequentemente, é assumido. Em seus experimentos numéricos, Zhang (2003)⁴ mostra que os resíduos produzidos modelos ARIMA são modeláveis por RNAs (previsor não linear) e que suas previsões “corrigem” as predições (lineares) da abordagem ARIMA, agregando informações e melhorando a acurácia do processo preditivo. Estas previsões “corrigidas” são, frequentemente, referidas como “previsões híbridas”. Formalmente, assuma que \mathbf{y}_t ($t = 1, \dots, T$) seja uma série temporal. De acordo com Zhang (2003)⁴, cada estado \mathbf{y}_t pode ser decomposto tal como: $\mathbf{y}_t = \mathbf{L}_t + \mathbf{N}_t$, onde \mathbf{L}_t e \mathbf{N}_t são, respectivamente, as suas componentes linear e não linear. Consequentemente, a previsão de y_t , denotada por \hat{y}_t , pode ser visualizada como: $\hat{y}_t = \hat{L}_t + \hat{N}_t + \varepsilon_t$, onde, por exemplo, \hat{L}_t (previsão linear) pode ser gerada por um

modelo linear de Box & Jenkins e \hat{N}_t (previsão não linear), por um modelo não linear neural.

De qualquer forma, independente de se utilizar um método preditivo individual linear ou não linear, a maior parte da literatura, implicitamente, assume que existe um modelo “real” que explica a estrutura de autodependência exibida pelos dados temporais, inclusive antes da estimação paramétrica (Chatfield, 2000)⁶. Em outras palavras, esta perspectiva de modelagem somente lida com incertezas associadas aos parâmetros e negligencia incerteza associadas aos previsores. Alguns autores, como, por exemplo, Neuman (2003)⁷, têm salientado que a adoção de apenas um previsor individual pode incorrer em viés estatístico, bem como subestimação da incerteza concernente ao modelo de previsão - o que corrobora o que a afirmativa que é muito difícil encontrar um método preditivo individual único e verdadeiro para modelar a fim de analisar e prever séries de tempo. Recentemente, pesquisas sobre incertezas associadas a previsores têm estado no escopo sobre projeções de séries temporais. Taskaya-Temisel and Casey (2005)⁸ por exemplo, trataram sobre a performance preditiva uma RNA híbrida, chamando à atenção para o risco de subestimação do processo subjacente. Por sua vez, alguns autores (como Dell’Aquila & Ronchitti, 2006⁹; Amendola & Storti, 2008¹⁰ e Lux & Morales-Arias, 2010¹¹) rejeitam a hipótese de que um único e verdadeiro previsor individual (por exemplo, uma RNA ou um modelo de Box e Jenkins) seja atingível (ou, sequer, exista). Ao invés disso, tais autores têm trabalhado no sentido de propor abordagens mais eficientes para se combinar métodos preditivos individuais, levando-se em conta a incerteza associada a cada um deles.

Paralelamente à modelagem híbrida e a combinada (que, basicamente, consistem na integração de métodos preditivos individuais), figuram as abordagens wavelet que realizam, em geral, processamento dos sinais temporais antes de sua efetiva modelagem. Comumente, os métodos wavelet consistem em decompor uma série temporal \mathbf{y}_t ($t = 1, \dots, T$) em termos de componentes espectrais (Donoho & Jonhston, 1994)¹², ou em filtrá-la (Donoho et al, 1995)¹³. Em particular, por meio de uma decomposição wavelet de nível r (Mallat, 2000)¹⁴, uma série temporal \mathbf{y}_t ($t = 1, \dots, T$) é decomposta em termos de $r+1$ componentes wavelet (CWs) –uma CW de aproximação de nível m_0 , denotada

por $A_{m_0,t} (t = 1, \dots, T)$, e r CWs de detalhe de níveis $m_0, \dots, m_0 + (r - 1)$, denotadas por $D_{m_0,t}, \dots, D_{m_0+(r-1),t} (t = 1, \dots, T)$, respectivamente, onde m_0 e r tomam valores inteiros. Devido ao fato de as referidas $r+1$ CWs serem dotadas de frequências espectrais constantes, como destaca Mallat (1998)¹⁵, tendem a possuir um padrão de comportamento mais regular que $y_t (t = 1, \dots, T)$. Com efeito, a modelagem individual, com um nível de acurácia satisfatório, pode incorrer, no processo de reconstrução (isto é, soma das previsões das CW), em ganhos preditivos relevantes, como em Lei; (Ran, 2008¹⁶; Tiwari & Chatterjee, 2010¹⁷; Lima 2011¹⁸; Kisi & Cimen, 2011¹⁹; Nalley, 2012²⁰; Karthikeyan & Kumar, 2013)²¹.

Diante do fato de que não existe uma forma de se combinar previsões, ou de se gerar previsões híbridas, que seja considerada a melhor e do fato de que o método de decomposição wavelet de nível r , quando implementado adequadamente, tende a incorrer em ganhos preditivos, propõem-se aqui uma metodologia híbrida wavelet de integração de modelos de Box & Jenkins e de RNAs para geração de previsões integrada com o amostrador Bootstrap (para geração dos intervalos de previsão). A escolha de tal método de amostragem se deu pelo fato de que ele não trabalha com a distribuição empírica apresentada pelos erros de previsão, evitando vieses decorrentes, por exemplo, da suposição de normalidade dos mesmos.

1.1

Problema

A principal motivação desta tese de doutorado é o desenvolvimento de método híbrido combinado que forneça vantagens preditivas muito superiores aos outros presentes na literatura especializada. Para tanto, o mesmo foi construído com o intuito de mapear informações atinentes às estruturas de autodependência linear (via CLW de modelos de Box e Jenkins) e não linear (via RNA automática) e frequência espectral (via geração das CW).

Em síntese, a metodologia híbrida proposta aqui pode ser executada em sete passo:

- ✓ *Passo 1:* decomposição wavelet de nível r da série temporal de interesse;
- ✓ *Passo 2:* modelagem individual de cada uma das $r+1$ CW via modelo SARIMA;
- ✓ *Passo 3:* Combinação linear wavelet (CLW) das previsões geradas no passo 2;
- ✓ *Passo 4:* Decomposição wavelet de nível k dos resíduos produzidos no passo 3;
- ✓ *Passo 5:* Modelagem simultânea das $k+1$ CWs geradas no passo 4 via redes neurais artificiais multilayer perceptron (RNA-MLP) wavelet automática;
- ✓ *Passo 6:* Geração das previsões (pontuais) híbridas da série temporal original, somando as previsões obtidas nos passos 3 e 6; e
- ✓ *Passo 7:* Geração das previsões intervalares (intervalos de previsão) via bootstrap.

1.2

Objetivos

1.2.1

Objetivo Geral

Projetar séries temporais por meio de um previsor híbrido combinado construído a partir da integração de previsores lineares e não lineares, da decomposição wavelet e da abordagem (não paramétrica) Bootstrap.

1.2.2

Objetivos Específicos

Os objetivos específicos são:

- ✓ Avaliar a importância da decomposição wavelet da série temporal para a qualidade das previsões;
- ✓ Demonstrar empiricamente a eficiência da decomposição wavelet na previsão de séries temporais;

- ✓ Verificar os ganhos preditivos obtidos com a combinação de previsores lineares e não lineares;
- ✓ Utilizar combinação linear e não linear de previsões;
- ✓ Comparar os resultados alcançados com o método proposto aos obtidos por métodos de previsão tradicionais.

1.3

Validação da Proposta

A fim de avaliar o método wavelet híbrido proposto, além de uma série temporal de vazão da usina hidrelétrica de Itaipu, são utilizados três conjuntos de dados oriundos de diferentes áreas e bem conhecidas na literatura (a saber, Sunspot, Canadian lynx e Exchange Rate , Zhang, 2003)⁴.

Tais conjuntos de dados foram escolhidos por serem dotados de diferentes características estatísticas e por serem amplamente conhecidos na literatura. Em especial, vários métodos preditivos lineares e não lineares têm sido aplicados a estes dados, com o objetivo de se gerar previsões pontuais. Assim, foi possível realizar comparações entre os resultados obtidos nesta tese com outros publicados em periódicos de notória relevância internacional. No caso da série temporal de vazão da usina de Itaipu, os resultados do método híbrido proposto foram comparados com abordagens tradicionais (elucidadas mais adiante).

1.4

Estrutura da Tese

Apresente tese está dividida em seis capítulos: (1) Introdução; (2) Revisão Bibliográfica; (3) Descrição do Método Proposto; (4) Experimentos Numéricos; (5) Conclusão; (6) Trabalhos futuros.

2

REVISÃO DA LITERATURA

2.1

Análise Wavelet

2.1.1

Preliminares

Inicialmente, três conceitos fundamentais devem ser definidos: espaço de Hilbert, base ortonormal e série de Fourier. Por espaço de Hilbert, entende-se como sendo um espaço produto interno *completo*. Em particular, a coleção \mathbf{l}^2 de todas as sequências infinitas de números complexos quadraticamente somáveis

(isto é, $\mathbf{l}^2 := \left\{ \mathbf{f}: \mathbf{Z} \rightarrow \mathbf{C}: \sum_{t \in \mathbf{Z}} |\mathbf{f}(t)|^2 < \infty \right\}$ -em que \mathbf{Z} e \mathbf{C} denotam, respectivamente, o conjunto dos números inteiros e o dos números complexos) -equipada de um produto interno $\langle : \rangle$ - ou seja, $(\mathbf{l}^2, \langle : \rangle)$ -, consiste em um caso particular de espaços de Hilbert.

De acordo com Kubrusly & Levan (2006)²³, um subconjunto $\{\mathbf{h}_n\}_{n \in \mathbf{Z}}$ de \mathbf{l}^2 é uma base ortonormal, se os axiomas seguintes são assegurados:

- (i) ortogonalidade (ou seja, $\langle \mathbf{h}_n, \mathbf{h}_m \rangle = 0$, sempre que $n \neq m$);
- (ii) normalização (ou seja, $\|\mathbf{h}_n\| = 1$, para todo $n \in \mathbf{Z}$); e
- (iii) Completamento (ou seja, $\langle \mathbf{x}, \mathbf{h}_m \rangle = 0$, se $\mathbf{x} = \mathbf{0}$). Com efeito, baseado no *Teorema da série de Fourier*, enunciado em Kubrusly (2001),²² se $\{\mathbf{h}_n\}_{n \in \mathbf{Z}}$ é uma base ortonormal para o espaço de Hilbert $(\mathbf{l}^2, \langle : \rangle)$, então, para todo vetor \mathbf{x} em \mathbf{l}^2 , existe uma expansão (única) tal como em (1).

$$\mathbf{x} = \sum_{n \in \mathbf{Z}} \langle \mathbf{x}, \mathbf{h}_n \rangle \mathbf{h}_n \quad (1)$$

A expansão em (1) é referida por *série de Fourier* do vetor \mathbf{x} sobre o espaço de Hilbert $(\mathbf{l}^2, \langle \cdot, \cdot \rangle)$. Daqui em diante, $(\mathbf{l}^2, \langle \cdot, \cdot \rangle)$ é representado, por simplicidade, por \mathbf{l}^2 .

2.1.2

Funções wavelet

Tome o espaço de Hilbert \mathbf{l}^2 . Um elemento $\omega(\cdot) \in \mathbf{l}^2$ - com um produto interno $\langle \cdot, \cdot \rangle: \mathbf{l}^2 \rightarrow \mathbb{C}$ - é uma *função wavelet*, se a coleção de funções $\omega_{m,n}(\cdot) := 2^{\frac{m}{2}} \omega(2^m(\cdot) - n)$, onde $n, m \in \mathbb{Z}$, forma uma base ortonormal. De acordo com Levan e Kubrusly (2003), uma função $f(\cdot)$ em \mathbf{l}^2 admite a sua decomposição através de uma expansão tal como em (1), definida em termos de uma base ortonormal wavelet $\{\omega_{m,n}(\cdot)\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$, conforme em (2).

$$f(\cdot) = \sum_{m \in \mathbb{Z}} \sum_{n \in \mathbb{Z}} \langle f(\cdot), \omega_{m,n}(\cdot) \rangle \omega_{m,n}(\cdot) \quad (2)$$

O índice m , em (2), é chamado de parâmetro de escala (ou de resolução) e o n , de parâmetro de translação [15]. De acordo com Levan e Kubrusly (2003)[26], a projeção de $f(\cdot)$ sobre $\omega_{m,n}(\cdot)$ pode ser interpretada como uma variação de detalhes de $f(\cdot)$, na escala m e translação n . Baseado em Mallat (1998)[15], tem-se que o subespaço *fechado* $\mathbf{W}_m(\omega) := \left(\text{span}\{\omega_{m,n}(\cdot)\}_{n \in \mathbb{Z}} \right)^-$ de \mathbf{l}^2 é chamado de subespaço de detalhes na escala m . Por sua vez, a projeção ortogonal de $f(\cdot)$ sobre $\mathbf{W}_m(\omega)$, denotada por $f_{\mathbf{W}_m(\omega)}(\cdot)$, é definida pela soma parcial descrita em (3).

$$f_{\mathbf{W}_m(\omega)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle f(\cdot), \omega_{m,n}(\cdot) \rangle \omega_{m,n}(\cdot) \quad (3)$$

A projeção ortogonal $f_{\mathbf{W}_m(\omega)}(\cdot)$, em (3), pode ser interpretada como uma componente de detalhes de $f(\cdot)$, na escala m sobre $\mathbf{W}_m(\omega)$. Com efeito, dada a identidade, em (1), segue que $f(\cdot)$ pode ser interpretada como a soma de todas

suas componentes de detalhes $f_{W_m(\omega)}(\cdot)$, em todas as escalas inteiras m , sobre o

subespaço fechado $\left(\left(\sum_{n \in \mathbb{Z}} W_m(\omega) \right)^{\perp}, \langle \cdot, \cdot \rangle \right)$ de l^2 . De acordo com Kubrusly e Levan

(2006) [23], tem-se que $\left(\sum_{n \in \mathbb{Z}} W_m(\omega) \right)^{\perp} = l^2$.

Por sua vez, um elemento $\phi(\cdot) \in l^2$ - com um produto interno $\langle \cdot, \cdot \rangle: l^2 \rightarrow \mathbb{C}$ - é chamado de função (wavelet) escala, se a família de funções $\phi_{m,n}(\cdot) := 2^{\frac{m}{2}} \phi(2^m(\cdot) - n)$, onde $n, m \in \mathbb{Z}$, é tal que: $\langle \phi_{m',n'}(\cdot), \phi_{j,k}(\cdot) \rangle = 0$, sempre que $m' = j$ e $n' \neq k$; e $\langle \phi_{m',n'}(\cdot), \phi_{j,k}(\cdot) \rangle \neq 0$, caso contrário. Baseado em Mallat (1998) [15], o subespaço fechado $V_m(\phi) := \left(\text{span}\{\phi_{m,n}(\cdot)\}_{n \in \mathbb{Z}} \right)^{\perp}$ de l^2 é chamado de subespaço de aproximação na escala m ; e a projeção ortogonal de $f(\cdot)$ sobre $V_m(\phi)$, denotada por $f_{V_m(\phi)}(\cdot)$, é definida pela soma parcial em (4).

$$f_{V_m(\phi)}(\cdot) := \sum_{n \in \mathbb{Z}} \langle f(\cdot), \phi_{m,n}(\cdot) \rangle \phi_{m,n}(\cdot) \quad (4)$$

De acordo com Teixeira Jr. (2013) [24], $f_{V_m(\phi)}(\cdot)$ pode ser interpretada como uma componente de aproximação de $f(\cdot)$, na escala m , sobre $V_m(\phi)$.

2.1.3

Transformadas wavelet

Por transformada wavelet sobre l^2 , entende-se como sendo um produto interno *usual* $\langle \cdot, \cdot \rangle: l^2 \rightarrow \mathbb{C}$ entre uma função $f(\cdot) \in l^2$ e uma função wavelet $\omega_{m,n}(\cdot) \in W_m(\omega)$ ou uma função escala $\phi_{m,n}(\cdot) \in V_m(\phi)$, onde $(m, n) \in \mathbb{Z} \times \mathbb{Z}$. De acordo com Mallat (1998) [15], as transformadas wavelet podem ser classificadas em dois conjuntos disjuntos: o de coeficientes de detalhes, denotado por $\{d_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$, e o de coeficientes de aproximação, denotado por $\{a_{m,n}\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$. Para cada par ordenado $(m, n) \in \mathbb{Z} \times \mathbb{Z}$, as transformadas

wavelet $\mathbf{d}_{m,n}$ e $\mathbf{a}_{m,n}$ são calculadas, respectivamente, por:

$$\mathbf{d}_{m,n} := \langle f(\cdot), \omega_{m,n}(\cdot) \rangle = \sum_{t \in \mathbb{Z}} f(t) \omega_{m,n}(t); \quad \text{e}$$

$$\mathbf{a}_{m,n} := \langle f(\cdot), \phi_{m,n}(\cdot) \rangle = \sum_{t \in \mathbb{Z}} f(t) \phi_{m,n}(t).$$

2.1.4

Decomposição wavelet

Baseado em Levan & Kubrusly (2003)²⁵, uma cadeia de subespaços de aproximação $\{\mathbf{V}_m(\Phi)\}_{m \in \mathbb{Z}}$ em \mathbf{I}^2 é uma Análise Multirresolução Wavelet (ou, simplesmente, AMR wavelet), com uma função escala $\phi(\cdot) \in \mathbf{I}^2$, se os axiomas (a), (b), (c), (d), e (e), abaixo, são satisfeitos.

$$(a) \mathbf{V}_m(\Phi) \subset \mathbf{V}_{m+1}(\Phi), \forall m \in \mathbb{Z};$$

$$(b) \bigcap_{m \in \mathbb{Z}} \mathbf{V}_m(\Phi) = \{0\};$$

$$(c) \left(\bigcup_{m \in \mathbb{Z}} \mathbf{V}_m(\Phi) \right)^- = \mathbf{H};$$

$$(d) v \in \mathbf{V}_m(\Phi) \Leftrightarrow Dv \in \mathbf{V}_{m+1}(\Phi), m \in \mathbb{Z}; \text{ e}$$

$$(e) \left\{ 2^{\frac{m}{2}} \phi(2^m(\cdot) - n) \right\}_{n \in \mathbb{Z}} \text{ é uma base ortonormal de } \mathbf{V}_m(\Phi), m \in \mathbb{Z}.$$

Em Kubrusly e Levan (2006)²³, mostra-se que

$$\mathbf{V}_{m_0}(\Phi) = \left(\sum_{-\infty}^{m_0-1} \mathbf{W}_m(\omega) \right)^-$$

onde $\mathbf{m}_0 \in \mathbb{Z}$. Dado que

$$\mathbf{I}^2 = \left(\sum_{m \in \mathbb{Z}} \mathbf{W}_m(\omega) \right)^- \quad \text{e} \quad \mathbf{V}_{m_0}(\Phi) = \left(\sum_{-\infty}^{m_0-1} \mathbf{W}_m(\omega) \right)^-,$$

segue que, com base no *Teorema da Estrutura Ortogonal*, enunciado em Kubrusly (2001)²², o espaço \mathbf{I}^2 pode ser ortogonalmente expandido conforme em (5).

$$\mathbf{I}^2 = \mathbf{V}_{m_0}(\Phi) + \left(\sum_{m=m_0}^{+\infty} \mathbf{W}_m(\omega) \right)^- \quad (5)$$

Ou, equivalentemente, $f(\cdot)$ pode ser decomposto ortogonalmente sobre \mathbf{I}^2 , conforme descrito em (6).

$$f(\cdot) = f_{V_{m_0}(\phi)}(\cdot) + \sum_{m=m_0}^{+\infty} f_{W_m(\omega)}(\cdot) \quad (6)$$

Dadas as definições das componentes wavelet $f_{V_{m_0}(\phi)}(\cdot)$ e $f_{W_m(\omega)}(\cdot)$ e as identidades (1) e (6), segue que a Série de Fourier da função $f(\cdot)$, sobre \mathbf{I}^2 , em termos da base ortonormal wavelet $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{m,n}(\cdot)\}_{(m,n) \in \{m\}_{m_0}^{+\infty} \times \mathbb{Z}}$, é dada por:

$$f(\cdot) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(\cdot) + \sum_{m=m_0}^{+\infty} \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(\cdot). \quad (7)$$

Onde: $a_{m,n} := \sum_{t \in \mathbb{Z}} f(t) \phi_{m,n}(t)$, $d_{m,n} := \sum_{t \in \mathbb{Z}} f(t) \omega_{m,n}(t)$, sendo que $m_0 \leq m < +\infty$ e $m_0 \in \mathbb{Z}$.

Na Figura 1, têm-se as curvas das funções wavelet da família de Symlet, com momento nulo igual a 20 (notação: “sym 20”) (Moretin, 1999)²⁶. Na parte superior, tem-se o gráfico da função escala; e, na inferior, o gráfico da a função wavelet.

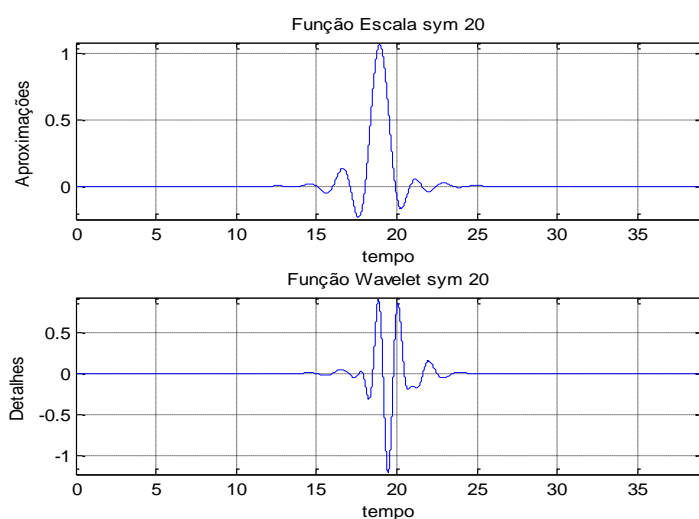


Figura 1 - Funções Wavelet da Família Symelet 20.

Fonte: O autor (2015).

Na Figura 2, têm-se as componentes de aproximação e de detalhe da série temporal mensal de *vazão de afluentes da região Sul do Brasil* (Teixeira Jr., 2013)²⁴ de uma decomposição ortogonal wavelet de nível 1, em termos de uma base ortonormal sym 20 (Daubechies, 1988)²⁷

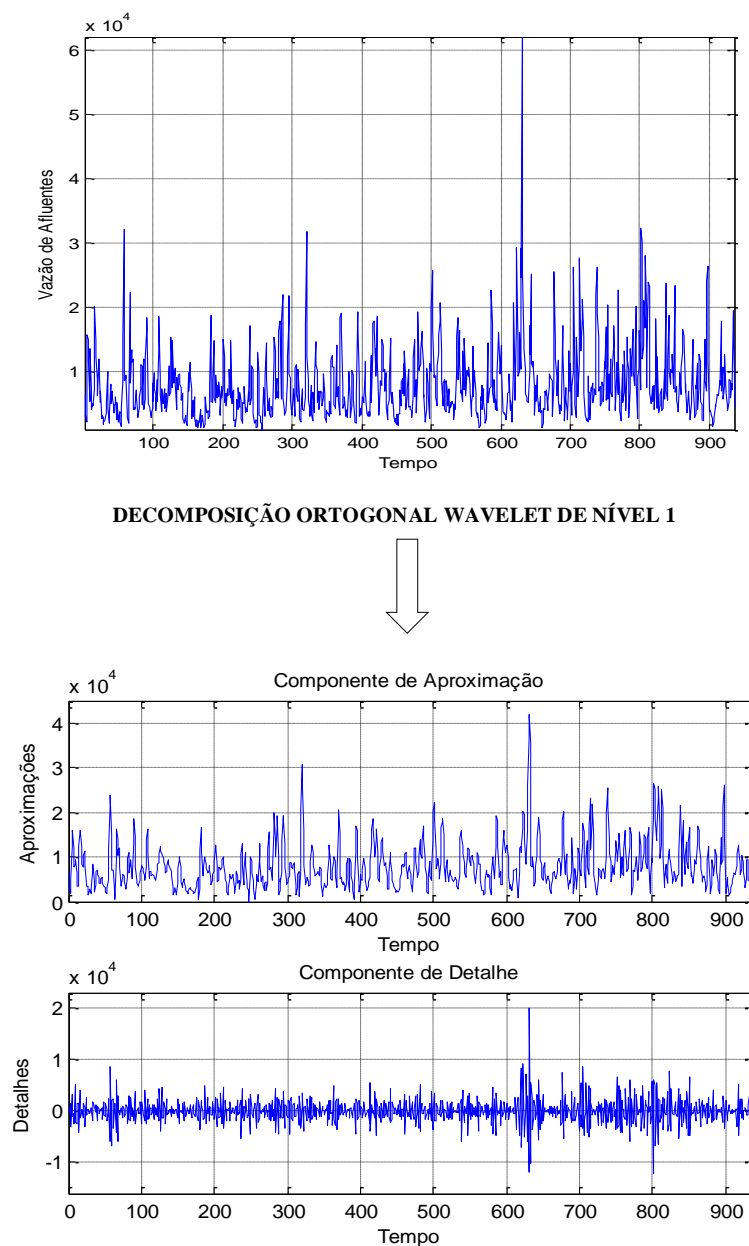


Figura 2 - Decomposição wavelet de nível 1, com a base ortogonal wavelet sym 20, da série temporal mensal de vazão de afluentes da região sul do Brasil.

Fonte: O autor (2015).

2.2

Modelos de Box e Jenkins

Os modelos foram inicialmente formulados pelos pesquisadores Box & Jenkins na década de 1970. Com argumentos de estacionariedade (de segunda ordem ou fraca) e ergodicidade do processo estocástico subjacente, procura-se detectar o sistema probabilístico plausível gerador de uma série temporal, através das informações nela contidas. Adicionalmente, baseia-se na também premissa de que uma série temporal não estacionária pode ser modelada a partir de d diferenciações e da inclusão de uma componente autorregressiva integrada de médias móveis. De acordo com Hill, et al (1999)²⁸, quanto aos modelos de Box e Jenkins: “relacionam os valores correntes de uma variável somente com os seus valores passados e erros passados e correntes”.

O modelo trabalha com dois princípios básicos ainda são assumidos: o da parcimônia, que estabelece um modelo com o menor número de parâmetros; e, o do ciclo iterativo, que consiste em estratégia de “tentativa e erro” de seleção de modelos (Morettin & Toloi, 2006)²⁹. A referida metodologia tem como base a Teoria Geral de Sistemas Lineares, a qual supõe que a passagem de um ruído branco por um filtro de memória infinita gera um processo estacionário de segunda ordem (Souza & Neto,)¹ Assim, assumamos a Equação (8).

$$y_t = \mu + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + \dots, \quad k = 0, 1, \dots \quad (8)$$

Onde: $\psi(B) = (1 + \psi_1 B^1 + \psi_2 B^2 + \dots)$. Desse modo, em (8), tem-se um, de fato, filtro linear com entrada a_t (ruído branco), uma saída, y_t (PE linear discreto) e uma função de transferência, $\psi_k(B)$. Definindo $\tilde{y}_t = y_t - \mu$, segue que:

$$\tilde{y}_t = \psi(B) a_t. \quad (9)$$

Conforme destacado, caso a sequência de pesos $\{\psi_k, k \geq 1\}$ seja finita ou infinita e convergente, o filtro é dito estável, implicando estacionariedade do

processo y_t . Neste caso, μ é a média do processo. Por outro lado, se y_t é não-estacionário, então μ não tem significado específico (Souza & Neto)¹

Considerando a Equação (9), segue que $E[y_t] = \mu + E\left(a_t + \sum_{j=1}^{\infty} \psi_j a_{t-j}\right)$, e

como $E[a_t] = 0$. Com efeito, $E[y_t] = \mu$, se a série $\sum_{j=1}^{\infty} \psi_j$ convergir. Pode-se

escrever \tilde{y}_t de forma alternativa, como a soma ponderada de valores passados

$\tilde{y}_{t-1}, \tilde{y}_{t-2}, \dots$, mais um ruído branco; isto é:

$$\tilde{y}_t = \pi_1 \tilde{y}_{t-1} + \pi_2 \tilde{y}_{t-2} + \dots + a_t \quad (10)$$

Assim, segue-se que:

$$\left(1 - \sum_{j=1}^{\infty} \pi_j B^j\right) \tilde{y}_t = a_t \quad (11)$$

$$\text{Ou: } \pi(B) \tilde{y}_t = a_t.$$

Onde:

$$\pi(B) = (1 - \pi_1 B^1 - \pi_2 B^2 - \dots) \quad (12)$$

Dada a Equação (12), segue que:

$$\pi(B) \psi(B) a_t = a_t \quad (13)$$

De modo que:

$$\pi(B) = \psi^{-1}(B) \quad (14)$$

Esta relação pode ser usada para obter os pesos π_j em função dos pesos ψ_j e vice-versa. Um processo linear é estacionário se a série $\psi(B)$ convergir para $|B| \leq 1$ e invertível, se $|B| \leq 1$. Box et al (1994)³⁰ provam tal fato. No concerne à função de transferência $\psi(B)$, tem-se:

$$\psi_k(B) = \theta(B) / \phi(B) \quad (15)$$

Pois um polinômio infinito equivale a uma divisão de polinômios.

$$\phi(B) y_t = \theta(B) a_t ;$$

$$y_t = \phi(B)^{-1} \theta(B) a_t . \quad (16)$$

Onde:

- (i) $\phi(B) = (1 - \phi_1 B^1 - \dots - \phi_p B^p)$: polinômio característico de ordem p da parte autorregressiva; e
- (ii) $\theta(B) = (1 - \theta_1 B^1 - \dots - \theta_q B^q)$: polinômio característico de ordem q da parte média móvel.

Sendo $\phi(.)$ e $\theta(.)$ polinômios cujos graus são, respectivamente, os índices p e q. Consoante a notação usual, o modelo (17) é denominado ARMA (p,q), um caso particular do método.

Se os pressupostos de normalidade e homocedasticidade da série temporal não são não satisfeitos, é feita sua transformação com o intuito de obter tais propriedades nos dados transformados. Box & Jenkins (1970)³¹ sugerem a classe de transformações de Box & Cox Souza & Neto (2004)¹ destacam que nem sempre é possível estabilizar a variância e obter normalidade, simultaneamente. De acordo com Souza & Neto (2004)¹: “A normalidade é o objetivo principal, podendo levar à variância constante. O inverso também é possível, isto é, estabilizar a variância, sendo que a normalidade pode ser obtida aproximadamente”.

2.2.1

Modelos autorregressivos

Se em (12) $\pi_j = 0, j > p$, obtém-se uma modelo autorregressivo de ordem p , geralmente denotado por AR (p) ou ARMA ($p,0$).

$$\text{AR}(p): \tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \dots + \phi_p \tilde{y}_{t-p} + a_t \quad (17)$$

Renomeando os pesos π_j para ϕ_j , conforme a notação usual, o polinômio característico é dado por $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$, chamado de operador não sazonal autorregressivo (AR). Ou, equivalentemente:

$$\phi(B) \tilde{y}_t = a_t. \quad (18)$$

A representação (18) é equivalente a uma média móvel infinita de ruídos descorrelatados, denotada por MA (∞), que é trivialmente invertível. Porém, para garantir a estacionariedade, requer-se, de acordo com [32], que $|\phi| < 1$; ou, de forma equivalente, que o polinômio $\phi(B) = [\psi(B)]^{-1}$ tenha suas raízes reais ou complexas (a norma, no caso) fora do círculo unitário. Ou seja, que a série $\psi(B)$ convirja quando $|B| \leq 1$.

2.2.2

Modelos médias móveis

Considere o processo linear (2.52) e suponha que $\psi_j = 0, j > q$; obtém-se um processo de médias móveis (moving average) de ordem q , que é denotado por MA (q) ou ARMA ($0,q$).

$$y_t = \mu + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q} \quad (19)$$

Sendo, $\tilde{y}_t = y_t - \mu$, segue que:

$$\tilde{y}_t = \theta(B) a_t. (20)$$

Onde: $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ consiste no operador não sazonal de médias móveis. A condição de inversibilidade do processo MA (q) é similar àquela de estacionariedade do processo AR (p), ou seja, $|\theta| < 1$. Com efeito, a série $\pi(B) = [\theta(B)]^{-1}$ converge para $|B| \leq 1$. Isso é equivalente a afirmar que os zeros do polinômio $\theta(B)$ tomam valores fora do círculo unitário. Assim, o processo MA (q) pode ser escrito como AR (∞). Estes modelos possuem uma representação trivialmente estacionária[1].

2.2.3

Modelos Autorregressivos e de Médias Móveis

O modelo ARMA (p, q) é escrito como na equação 21, onde $\phi(.)$ e $\theta(.)$ são polinômios de ordem p e q , respectivamente. São compostos por duas partes: uma autorregressiva e de outra de médias móveis.

$$\tilde{y}_t = \phi_1 \tilde{y}_{t-1} + \phi_2 \tilde{y}_{t-2} + \dots + \phi_p \tilde{y}_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} \quad (21)$$

Sendo $\phi(B)$ e $\theta(B)$ os polinômios característicos autorregressivos e de médias móveis, respectivamente, podem ser escritos de forma compacta:

$$\phi(B) \tilde{y}_t = \theta(B) a_t. \quad (22)$$

Onde: $\theta(B) = (1 - \theta_1 B^1 - \dots - \theta_q B^q)$: operador não sazonal de médias móveis; e $\phi(B) = (1 - \phi_1 B^1 - \dots - \phi_p B^p)$: operador não sazonal autorregressivo. A fim de se verificar a propriedades de invertibilidade e de estacionariedade, analisam-se os

polinômios $\theta(B)$ e $\phi(B)$, respectivamente. Caso o correlograma da função de autocorrelação (ACF) e da função de autocorrelação parcial (PACF) apresente decaimento senoidal amortecido ou o exponencial, simultaneamente, uma sugestão plausível é o ARMA (1,1).

2.2.4

Modelos ARIMA

O modelo ARIMA (p,d,q) é uma extensão do ARMA (p,q). O índice d representa o número de vezes que a série original foi diferenciada. O resto é similar. Portanto, o Modelo é dado por ARIMA (p,d,q):

$$\phi(B) \nabla^d \tilde{y}_t = \theta(B) a_t \quad (23)$$

Onde: $\nabla^d = (1-B)^d$ consiste no operador diferença não sazonal de ordem d. Se $w_t = \nabla^d z_t$ for estacionária, é possível representar por um modelo ARMA, ou seja, $\phi(B) w_t = \theta(B) a_t$. Logo, w_t é uma diferença de z_t , então esta última é uma integral de w_t .

Por isso é dito que z_t segue um modelo autorregressivo, integrado, de médias móveis. Portanto, o modelo supõe que a d-ésima diferença da série z_t pode ser representada por um modelo ARMA, estacionário e invertível. O ARIMA, na verdade, é um caso especial de um processo integrado (Souza & Neto, 2004)¹.

2.2.5

Modelos SARIMA

Os modelos SARIMA (p,d,q) \times (P,D,Q)_s, também conhecidos como ARIMA multiplicativo, possuem duas partes: a simples e a sazonal. De modo que a sazonalidade da série temporal passa a ser considerada na modelagem. O modelo SARIMA (p,d,q) * (P,D,Q) é dado por:

$$\Phi(B^S) \phi(B) \nabla^d \nabla_s^D \tilde{y}_t = \Theta(B^S) \theta(B) a_t. \quad (24)$$

Onde:

$\nabla^d = (1-B)^d$: operador diferença não sazonal de ordem d;

$\nabla_s^D = (1-B^S)^D$: operador diferença sazonal de ordem D;

$\theta(B) = (1-\theta_1 B^1 - \dots - \theta_q B^q)$: operador não sazonal de médias móveis;

$\phi(B) = (1-\phi_1 B^1 - \dots - \phi_q B^q)$: operador não sazonal autorregressivo;

$\Phi(B^S) = (1-\Phi_1 B^{1S} - \dots - \Phi_P B^{PS})$: operador sazonal autorregressivo; e

$\Theta(B^S) = (1-\Theta_1 B^{1S} - \dots - \Theta_P B^{PS})$: operador sazonal de médias móveis.

Uma vez definido o tamanho do período sazonal, analisam-se os correlogramas da ACF e PACF a fim de estimar os índices p, d e q, como também P, D e Q, atinentes à parte sazonal. O raciocínio para estimação do modelo é similar à parte simples, mas é feita sob os períodos S, 2S, 3S,... [1].

2.2.6

Identificação de modelos

É possível realizar a identificação, a priori, de um modelo plausível através de análises de dependência linear nos correlogramas da ACF e PACF. Algumas características destes gráficos caracterizam determinados processos estocásticos:

- MA (q) – O correlograma da ACF tem picos do lag 1 até o q, caindo de forma brusca a zero no lag q+1. O gráfico da PACF apresenta um decaimento exponencial ou senoidal amortecida. Por exemplo, assumamos um MA (1), o pico será no lag 1, declinando abruptamente a zero a partir do lag 2. Caso $\theta_1 < 0$, o pico será positivo e se $\theta_1 > 0$, negativo. O correlograma da PACF mostra um decaimento exponencial negativo para $\theta_1 > 0$ e senoidal amortecida para $\theta_1 < 0$;

- AR (p) – O correlograma da ACF tem um decaimento exponencial ou senoidal amortecido. O gráfico da PACF tem pico do lag 1 até o p, caindo bruscamente a zero no lag p+1. Por exemplo, assumo um AR (1), o pico será no lag 1, declinando abruptamente a zero a partir do lag 2. Caso $\phi_1 > 0$, o pico será positivo e se $\phi_1 < 0$, negativo. O correlograma da ACF mostra um decaimento exponencial positivo para $\phi_1 > 0$ e senoidal amortecida para $\phi_1 < 0$;
- SAR (P) e SMA (Q) – de igual modo às partes simples ar (p) e ma (q), a análise é realizada, respectivamente, considerando, porém, os ciclos sazonais. Assim, para S = 12, suponha que haja uma queda brusca no lag 36 da FACP e que o correlograma da ACF apresente um decaimento exponencial ou senoidal amortecido nos lag's: 12, 24, 36,... A parte sazonal poderia ser representada pelo polinômio SAR (1).

Para mais detalhes práticos e teóricos acerca da modelagem de Box & Jenkins (1970)³¹, veja: Hill et al (1999)²⁸.

2.3

Redes Neurais Artificiais

De acordo com Haykin (2001)², as redes neurais artificiais (RNA's) com uma, e somente uma, camada escondida consistem em sistemas constituídos por unidades de processamento simples, chamadas de neurônios artificiais (Abalém, 1994)³³ e podem ser utilizadas para aproximar classes de funções contínuas de suporte compacto. Os neurônios artificiais de uma RNA são dispostos em camadas neurais (Tafner, 1996)³⁴ interligadas por um grande número de conexões - denominados pesos sinápticos, os quais ponderam os padrões de entrada [35] recebidos. De acordo com Haykin (2001)², o funcionamento das RNA's é inspirado na estrutura massivamente paralela do cérebro humano, de forma a realizar operações similares, através de conhecimento experimental obtido

previamente a partir dos padrões (ou sinais), tais como: aprendizado, associação, generalização e abstração. Na Figura 3, tem-se uma esquematização simplificada de um neurônio biológico em cuja estrutura um neurônio artificial de uma RNA é inspirado.

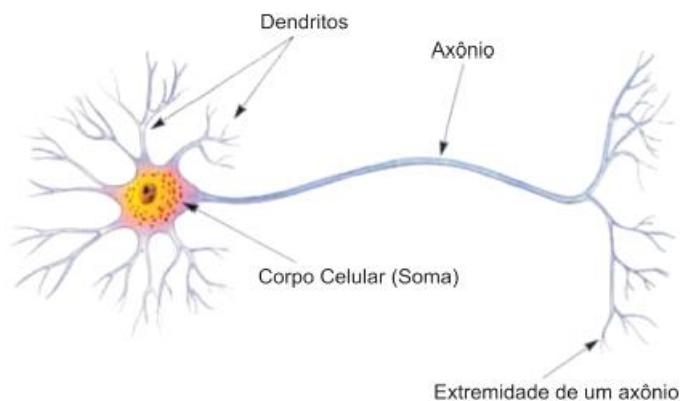


Figura 3 - Esquematização simplificada de um neurônio biológico.

Segundo Tafner (1996)³⁴, um neurônio biológico é constituído, basicamente, pelos seguintes elementos: um *corpo celular* (ou *soma*), que contém o núcleo da célula; diversos *dendritos*, através dos quais impulsos elétricos são recebidos; e um *axônio*, por meio do qual são enviados impulsos elétricos. As interligações entre neurônios são efetuadas através de *sinapses* (isto é, pontos de contato que são controlados por impulsos elétricos e por reações químicas) entre os dendritos e os axônios, formando, assim, uma rede neural biológica de transmissão de informações. Segundo Haykin (2001)², um neurônio artificial é constituído pelos elementos descritos, em (1), (2), (3) e ilustrados, na Figura 4.

- (1) Um *conjunto de pesos sinápticos*, denotado por $\left\{ \left\{ W_{kt} \right\}_{k=1}^K \right\}_{t=1}^T$, e um *conjunto de padrões de entrada*, denotado por $\left\{ x_t \right\}_{t=1}^T$, onde W_{kt} denota o peso sináptico associado ao padrão de entrada x_t (observado no instante t) e ao k -ésimo neurônio artificial. Cada padrão de entrada x_t é multiplicado pelo peso respectivo sináptico W_{kt} , ou seja, $x_t W_{kt}$, para todo $t \in \{1, \dots, T\}$ e para todo $k \in \{1, \dots, K\}$;

- (2) Um *somador* - que consiste em um operador denotado por Σ - para realizar a operação soma dos padrões de entrada em $\{x_t\}_{t=1}^T$ ponderados (por meio da operação de multiplicação) pelos respectivos pesos sinápticos no conjunto $\left\{\{W_{kt}\}_{k=1}^K\right\}_{t=1}^T$. Isto é, $\sum_{t=1}^n x_t W_{kt} + b_k$, onde: x_t é o padrão de entrada, em t ; W_{kt} é o peso sináptico associado ao padrão de entrada x_t e ao k -ésimo neurônio artificial; e b_k é o intercepto (ou nível, ou *bias*), que tem a função de aumentar ou diminuir a entrada líquida da função de ativação, podendo, desse modo, ser positivo ou negativo. A soma ponderada $\sum_{t=1}^n x_t W_{kt} + b_k$, de acordo com Haykin (2001)², pode ser chamada de *regra de propagação*; e

- (3) Uma *função de ativação*, denotada por φ , que é utilizada para restringir a amplitude da saída y_k de um neurônio artificial, dada a soma ponderada $\sum_{t=1}^n x_t W_{kt} + b_k = v_k$. Isto é, $y_k = \varphi(v_k)$, onde $k \in \{1, \dots, K\}$. Além disso, segundo Herrera & Lozano (1998)³⁶ são utilizadas na identificação de relações lineares ou não lineares entre os padrões de entrada e os de saída em modelos neuronais.

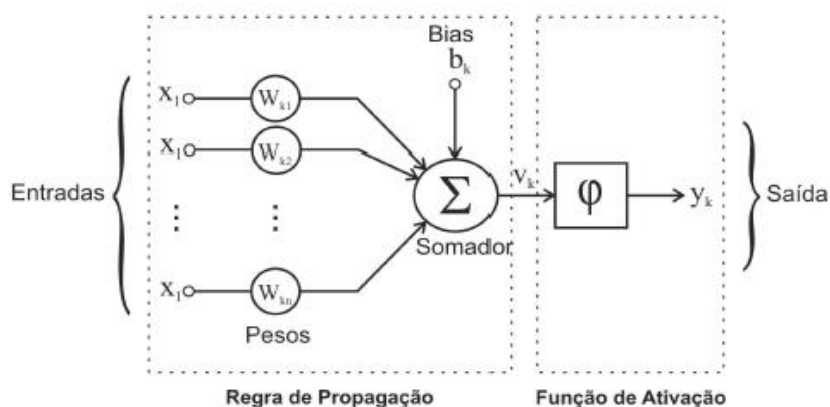


Figura 4 - Arquitetura básica de um neurônio artificial.

2.3.1

Componentes Básicas de uma RNA

De acordo Zsolt (2006)³⁷, as RNA's (básicas) podem ser classificadas de acordo com as características nominadas, a seguir: (a) *função de ativação* dos neurônios artificiais [Seção 2.3.1.1]; (b) a *arquitetura*, que consiste na forma como os neurônios estão interligados na estrutura neuronal [Seção 2.3.1.2]; e (c) O *treinamento neural* [Seção 3.1.3].

2.3.1.1

Função de Ativação

De acordo com Campos (2010)³⁸, os quatro tipos de funções de ativação mais utilizados, em aplicações envolvendo a projeção de séries temporais, são: *linear, degrau, logística, tangente hiperbólica*.

(a) Função Linear:

$$y_k = \sum_{t=1}^n x_t W_{kt} + b_k. \quad (25)$$

A *função linear*, geralmente, é utilizada nos neurônios da camada de saída e é indicada quando as sequências de padrões de saída associados a estes não são conjuntos limitados.

(b) Função Degrau (ou Limiar):

$$y_k = \begin{cases} 1, & \text{se } v_k \geq 0 \\ 0, & \text{se } v_k < 0 \end{cases}. \quad (26)$$

O conjunto imagem da *função degrau* (ou *limiar*) é constituído pelos valores 0 pelos valores -1 ou +1, ou 1. No entanto, por vezes, é necessário que a

sua imagem seja constituída de modo que, neste caso, a função de limiar pode ser particularmente chamada de *função de sinal* (Haykin, 2001)².

(c) **Função Sigmoidal Logística:**

$$y_k = \frac{1}{1 + \exp\{-\alpha v_k\}}, \text{ onde } \alpha \in \mathbb{R}. \quad (27)$$

A função sigmoide logística toma valores no intervalo limitado e fechado $[0,1]$ e consiste em uma transformação monotonicamente crescente (Lima, 2006)³⁹, onde $\sum_{t=1}^n x_t W_{kt} + b_k = v_k$ é a regra de propagação do k-ésimo neurônio artificial.

(d) **Tangente Hiperbólica:**

$$y_k = \tanh(\alpha v_k), \text{ onde } \alpha \in \mathbb{R}. \quad (28)$$

A *função tangente hiperbólica*, denotada por $\tanh(\cdot)$, é utilizada quando se deseja que a saída de um neurônio artificial tome valores no intervalo limitado e fechado $[-1,1]$.

2.3.1.2

Arquitetura

A *arquitetura* de uma RNA se refere à maneira como os seus neurônios artificiais estão interconectados. De acordo com Russell & Norvig (1995)³⁵, os dois tipos básicos de arquitetura de RNA's são a *feedforward* (não recorrente) [letra (a)] e a *recorrente* [letra (b)].

- (a) **RNA's, com arquitetura *feedforward*:** as RNA's com arquitetura *feedforward* são aquelas que não possuem conexões entre neurônios da mesma camada ou entre um neurônio de uma camada com outro neurônio de uma camada anterior. Na Figura 32, é ilustrada uma RNA

feedforward, com uma camada escondida (intermediária ou oculta) e um neurônio na camada de saída. Uma RNA *feedforward* possui, basicamente: um conjunto de nós de entradas, que simplesmente distribuem os padrões de entrada para os seus neurônios artificiais; uma camada de saída, cujos neurônios fornecem as saídas finais da referida RNA; e camadas intermediárias, cujas saídas dos neurônios são as entradas dos neurônios da camada seguinte.

- (b) **RNA's, com arquitetura recorrente:** as RNA's com arquitetura recorrente são aquelas que possuem realimentação das saídas para as entradas, de modo que as saídas destas são determinadas pelas entradas atuais e pelas saídas anteriores (Haykin, 2001)². As estruturas das RNA's recorrentes, portanto, podem apresentar interligações entre neurônios da mesma camada ou entre neurônios de camadas anteriores.

De acordo Russell & Norvig (1995)³⁵, a arquitetura de uma RNA (básica) envolve, em linhas gerais, os seguintes parâmetros: padrões de entrada, número de camadas intermediárias, número de neurônios por camada (entrada, intermediária e de saída), tipo de função de ativação dos neurônios e os padrões de saída.

2.3.2

Treinamento neural

Segundo Abelém (1994)³³, o *treinamento neural* de uma RNA consiste em um processo, no qual, com o uso de um *algoritmo de otimização* (Vanderplaats, 1990)⁴⁰ plausível, é realizado o ajuste numérico dos pesos sinápticos, de forma a minimizar (ou maximizar) uma *função-objetivo* (Ragsdale, 2004)⁴¹. Para tal, de acordo com Haykin (2001)², é comum se utilizar um único conjunto de sinais como padrões de entrada e de saída, simultaneamente, o qual é denominado *amostra de treino* (Campos, 2010)³⁸. Geralmente, o treinamento de uma RNA utiliza como *objetivo* (Ragsdale, 2004)⁴¹ a minimização da função *erro quadrático médio* (EQM) entre os padrões de saída da RNA e os desejados.

Herrera& Lozano (1998)⁴² destacam que é de extrema importância a utilização de padrões (de entrada e de saída) que sejam representativos, quanto a uma *população* (Morettin, & Toloí, 2006)⁴³ de interesse, e que contenham um número de sinais suficiente grande, para que se obtenha uma modelagem adequada. Segundo Haykin (2001)², os procedimentos de treinamento neural ocorrem em dois tipos básicos:

- (a) **Treinamento supervisionado:** os padrões de treino são constituídos pelos conjuntos de entrada e de saída desejadas. Durante o processo de treinamento, as entradas são apresentadas as RNA's e, após isso, são calculadas as saídas neuronais. Em seguida, cada saída da RNA é comparada com a respectiva saída desejada, gerando-se, assim, um sinal de erro (que é a diferença entre as duas saídas). O algoritmo de treinamento, a cada ciclo, atualiza os pesos sinápticos, baseados nos valores de erros nos padrões de treino, a fim de que as saídas da RNA sejam as mais próximas possíveis das saídas desejadas; e
- (b) **Treinamento não supervisionado:** os padrões de treinamento não contêm padrões de saída desejados, de forma que não existem comparações que acarreta geração de sinais de erro. De acordo com Palit & Popovic (2005)⁴⁴, o processo de treinamento não supervisionado, em geral, extrai as propriedades estatísticas do conjunto de padrões de entrada, de maneira a formar agrupamentos disjuntos constituídos de padrões similares, com base em uma *topologia* (Kubrusly, 2001)²².

Após o treinamento neural, é apresentado à RNA treinada um conjunto de padrões de *validação* e de *teste* (Abalém, 1994³³; Campos, 2010)³⁸, que são constituídos de padrões não apresentados antes à RNA, mas que pertençam à mesma população dos padrões de treino, a fim de validá-la e testá-la. Existem vários critérios para a escolha de uma RNA (Russel & Norvig, 1996)³⁵, mas, no caso das RNA's com treinamento supervisionado (que são as de interesse nesta

pesquisa), o erro quadrático médio (nas amostras de treino, validação e teste) é o mais utilizado.

2.3.3

Normalização dos padrões

De acordo Tafner (1996)³⁴, a *normalização dos padrões* de entrada e de saída de RNA's, com treinamento supervisionado, é fundamental para que suas magnitudes fiquem dentro de uma faixa de valores, a fim de evitar que os padrões com maiores magnitudes distorçamos valores dos pesos sinápticos no processo de treinamento neural e de suprimir problemas associados à *paralisia neural* (Haykin, 2001)². Neste caso, os padrões devem ser normalizados em um intervalo limitado compatível com o domínio das funções de ativação dos respectivos neurônios da estrutura neuronal. De acordo com Campos (2010)³⁸, as normalizações de sinais mais comuns, em aplicações envolvendo projeções de séries temporais, são as que os transformam em pontos nos intervalos limitados e fechados $[0,1]$ ou $[-1,1]$; ou, ainda, em pontos *padronizados* (ou seja, elementos de um conjunto de sinais cuja média é igual a 0 e desvio-padrão, igual a 1).

Nesta perspectiva, as normalizações de padrões *MINIMAX*, *método da faixa*, *z-score*, *sigmoidal* são, respectivamente, definidas nos itens (a), (b), (c) e (d), abaixo. Para tanto, assuma previamente que $\{x_t\}_{t=1}^n$ seja, simultaneamente, o conjunto dos padrões de entrada e de saída de uma RNA, com treinamento supervisionado, e que $\{x_t^m\}_{t=1}^n$ seja a sua versão normalizada.

- (a) **Método MINIMAX:** utiliza o valor máximo do conjunto $\{abs(x_t)\}_{t=1}^n$

de sinais, o qual é denotado por $Max\{abs(x_t)\}_{t=1}^n$, onde $abs(x_t)$ denota o valor absoluto de x_t . A normalização dos pontos da sequência $\{x_t\}_{t=1}^n$ ocorre conforme a transformação definida em (29), que toma valores no intervalo $[-1,1]$.

$$x_t^m = x_t / Max\{abs(x_t)\}_{t=1}^n \quad (29)$$

- (b) **Método *premnmx***: utiliza os valores máximo e mínimo do conjunto de padrões $\{x_t\}_{t=1}^n$, os quais são denotados, respectivamente, por $Max\{x_t\}_{t=1}^n$ e $Min\{x_t\}_{t=1}^n$. A normalização de cada ponto x_t ocorre de acordo com a transformação definida em (30), a qual toma valores no intervalo $[-1,1]$.

$$x_t^m = 2 \times \left[\left(x_t - Min\{x_t\}_{t=1}^n \right) / \left(Max\{x_t\}_{t=1}^n - Min\{x_t\}_{t=1}^n \right) \right] - 1 \quad (30)$$

- (c) **Método do z-escore**: realiza a normalização do conjunto $\{x_t\}_{t=1}^n$ em torno de sua média, denotada por \bar{x}_n , utilizando o seu desvio padrão, denotado por S_x , de acordo com a transformação definida em (31). A média de conjunto $\{x_t^m\}_{t=1}^n$ de padrões normalizados é igual a 0 e o desvio-padrão, igual a 1.

$$x_t^m = (x_t - \bar{x}_n) / S_x \quad (31)$$

- (d) **Método Sigmoidal**: utiliza a média e o desvio-padrão do conjunto $\{x_t\}_{t=1}^n$ de padrões, os quais são, respectivamente, denotados por \bar{x}_n e S_x . A normalização de cada ponto x_t em $\{x_t\}_{t=1}^n$ ocorre de acordo com a transformação definida em (32), que toma valores no intervalo $[0,1]$.

$$x_t^m = 1 / (1 + \exp\{(x_t - \bar{x}_n) / S_x\}) \quad (32)$$

2.3.4

RNA' s multicamadas feedforward, com uma camada escondida

Na Figura 5, tem-se a ilustração de uma arquitetura de uma RNA multicamada perceptron (MLP), com duas camadas escondidas com N neurônios; K neurônios na camada de saída e janela igual a m (Haykin, 2001²; Russel & Norvig, 1996)³⁵: a fase *forward* e a fase *backward*.

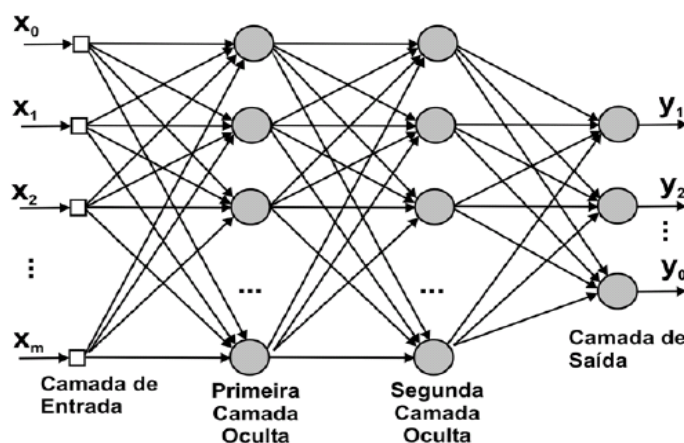


Figura 5 - RNA MLP *feedforward*.

O *backpropagation* é um dos principais algoritmos de treinamento neural. O ajuste numérico dos pesos sinápticos, com o uso deste algoritmo, ocorre por meio de um processo de otimização de duas fases básicas: a *forward* e a *backward*. Na fase *forward*, calcula-se a resposta fornecida pela RNA a partir de um conjunto de padrões de entrada, que depende do tamanho da janela neural. Na fase *backward*, o erro entre a resposta da RNA e a resposta desejada é utilizado no processo de atualização dos pesos sinápticos. Ao longo do treinamento, os padrões de entrada e as respectivas respostas desejadas são apresentados à RNA, de forma a minimizar o erro quadrático médio, da amostra de treino, entre os sinais de saída da RNA e os sinais de saída desejados.

2.3.5

Geração de previsões via redes neurais artificiais

A previsão de valores futuros de uma série temporal, com o uso de uma RNA *feedforward* (com treinamento supervisionado), com uma camada escondida e uma camada de saída com um neurônio artificial, é possível porque as suas observações (sinais) podem ser interpretadas como pontos de alguma função de suporte compacto e, de acordo com a Seção 3.3, tais RNA's são aproximadores deste tipo de mapeamento, na topologia $\| \cdot \|_{\infty}$. Assim sendo, previamente, define-se o tamanho da janela neural L e do horizonte de previsão h desejado. Em uma série temporal com estrutura de dependência temporal *autoregressiva* (Hamilton, 1994)³² (linear ou não linear), tem-se que o conjunto de padrões de entrada é constituído pelos valores passados da própria série temporal que se deseja projetar. O padrão de saída desejado, neste caso, consiste no valor a ser previsto.

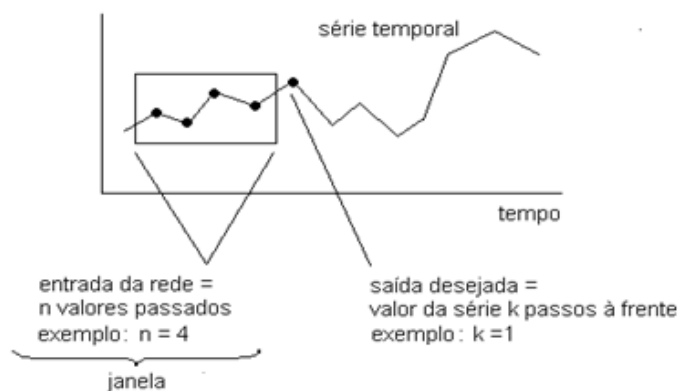


Figura 6 - Ilustração da dinâmica de previsão, 1 passo à frente, de uma série temporal, com estrutura de autodependência igual a 4, através de uma RNA.

Na Figura 6, tem-se a ilustração da dinâmica da produção de uma previsão, 1 passo à frente, para uma série temporal com estrutura de autodependência com defasagem igual a 4 (ou seja, L é igual a 4), através de uma RNA *feedforward*. Note que o vetor de padrões de entrada é constituído pelos últimos 4 sinais temporais, antes do padrão de saída a ser previsto. Para prever o sinal subsequente, a janela da RNA é deslocada um passo à frente (Campos, 2010)³⁸.

2.4

Combinação de métodos preditivos individuais

Tome uma série temporal $\{y_t\}_{t=1}^T$ e considere que se tenha o interesse em modelá-la, a fim de se gerar previsões, dentro e fora da amostra. Assuma que, para tal, se tenha disponível um conjunto com M métodos preditivos plausíveis, o qual é denotado por $\{\mu_m\}_{m=1}^M$, onde $M \in \mathbb{Z} \setminus \{0,1\}$, sendo \mathbb{Z} o conjunto dos números inteiros positivos. De acordo com Faria & Mubwandarikwa (2008)⁴⁵, existem, basicamente, duas possibilidades em se determinar um método preditivo gerador de previsões para a série temporal $\{y_t\}_{t=1}^T$:

1. Escolher um método preditivo em $\{\mu_m\}_{m=1}^M$, baseado em algum de seleção; ou
2. Escolher K métodos preditivos em $\{\mu_m\}_{m=1}^M$ e, em seguida, combiná-los (de forma linear ou não linear).

De acordo com Flores & White (1989)⁴⁶, no processo de combinação de métodos preditivos, é importante se considerar dois aspectos: o primeiro concerne à escolha dos K métodos preditivos base em $\{\mu_m\}_{m=1}^M$; e o segundo, quanto à forma de se combiná-los. Em relação ao primeiro aspecto, os autores destacam que as previsões oriundas dos métodos preditivos base podem ser classificadas em três categorias: *objetivas*, *subjetivas* e *mistas* (isto é, as resultantes da combinação de previsões objetivas e subjetivas). As *previsões objetivas* são aquelas oriundas de previsores com base teórica estatística ou matemática; enquanto que as *previsões subjetivas* são as decorrentes de julgamento humano - geralmente, grupo focado ou opinião de especialistas (Werner & Ribeiro, 2006)⁴⁷. Salienta-se que, nesta pesquisa, todos os esforços se concentram somente na combinação de previsões objetivas. Quanto ao segundo aspecto, embora haja um consenso entre os pesquisadores de que a combinação de previsores individuais, em regra, incorre em ganhos preditivos na projeção de séries temporais, não existe um senso comum em se definir a melhor forma de se fazê-la (Wallis, 2011)⁴⁸.

Para a Definição 1 (a seguir), considere uma série temporal $\{y_t\}_{t=1}^T$ em $(\Omega, \mathcal{X}_\Omega, \rho)$ e assuma que exista o interesse em se proceder à sua modelagem, a fim de se gerar previsões, dentro e fora da amostra, por meio de uma função de combinação de métodos preditivos base. Considere ∇^K seja o conjunto de todas as previsões (dentro e fora da amostra) oriundas dos K métodos preditivos individuais escolhidos, pelo tomador de decisão, em $\{\mu_m\}_{m=1}^M$, onde $K \leq M$; e

tome um mapa $\hat{y}_C : \begin{cases} \nabla^K \rightarrow \mathbf{z} \\ [\hat{y}_{t,k}]_{k=1}^K \in \nabla^K \rightarrow \hat{y}_{t,C} \in \mathbf{z} \end{cases}$ de combinação de previsões (genérico) que transforma, de acordo com sua definição, um vetor $[\hat{y}_{t,k}]_{k=1}^K \in \nabla^K$ de previsões em uma previsão combinada $\hat{y}_{t,C} \in \mathbf{z}$, para todo $t = 1, \dots, T+h$, onde $h \in \mathbf{z}$ é o horizonte de previsão.

2.4.1

Combinação Linear de Métodos Preditivos

As *combinações lineares de métodos preditivos* (abordagem clássica e bayesiana) são vastamente citadas na literatura de projeção de séries temporais. Em Clemen (1989)⁴⁹, por exemplo, é feito um detalhamento sobre as combinações lineares de previsões (abordagem clássica); enquanto que as combinações lineares de densidades preditivas (abordagem bayesiana) são detalhadas em Geneste Zidek (1986)⁵⁰, Mubwandarikwa (2007)⁵¹.

2.4.1.1

Combinação linear de previsões

De acordo com Clemen (1989)⁴⁹, a combinação de previsões consiste em uma abordagem clássica que utiliza algum mecanismo ponderador (calculado de acordo com algum objetivo) associado às previsoras a serem combinadas. Geralmente, adota-se, como objetivo, a minimização da estatística MSE (*mean square error*).

Makridakise Winkler (1983)⁵² propuseram a combinação linear de previsões oriundas de K métodos em $\{\mu_m\}_{m=1}^M$, onde $K \leq M$, utilizando pesos adaptativos variantes no tempo. Os autores verificaram que as previsões combinadas foram mais acuradas que as dos previsores individuais. Em Teixeira Jr (2009)⁵³, foi proposto o uso de uma constante aditiva na combinação linear de previsões. O autor mostrou, em seus experimentos, que a constante adaptativa acarretou ganhos preditivos.

Em Teixeira Jr (2009)⁵³ foram linearmente combinadas previsões de três métodos preditivos base: um método de *amortecimento exponencial* (Morettin & Tolo, 2006)⁴³, um modelo de *Box & Jenkins* e uma rede neural artificial. Os resultados estatísticos associados às previsões linearmente combinadas foram superiores aos das previsões dos previsores individuais. No que tange aos pesos adaptativos lineares, foram assumidos como sendo invariantes no tempo, normalizados, não negativos e com soma igual a uma unidade; e, para obtenção de seus valores ótimos, foi utilizado, como objetivo, o mínimo da estatística MAPE (*mean absolute percentual error*).

Definição 1 – Uma função $\hat{y}_{CL} : \left\{ \begin{array}{l} \nabla^K \rightarrow \mathbf{z} \\ \left[\hat{y}_{t,k} \right]_{k=1}^K \in \nabla^K \rightarrow \hat{y}_{t,CL} \in \mathbf{z} \end{array} \right.$ de combinação linear de previsões para a série temporal $\{y_t\}_{t=1}^T$ é um mapeamento definido por:

$$\hat{y}_{CL,t} = \sum_{k=1}^K \rho_{k,t} \times \hat{y}_{t,k} + \alpha. \quad (33)$$

Onde, em t : $\hat{y}_{CL,t}$ é a previsão linearmente combinada; $\hat{y}_{t,k}$, a previsão do k -ésimo método base; $\rho_{k,t}$, o peso adaptativo associado linearmente à previsão $\hat{y}_{t,k}$; α e, constante adaptativa aditiva.

2.4.1.2

Constantes e pesos adaptativos

No que concerne aos valores - ótimos (abordagem via programação matemática) ou estimados (abordagem estatística) - dos parâmetros adaptativos (ou seja, das constantes e dos pesos adaptativos) das combinações lineares de previsores individuais, existem, na literatura de projeções de séries temporais, muitas propostas de se obtê-los. Não obstante, não existe aquela que seja considerada a mais eficiente.

Segundo Faria & Mubwandarikwa (2008)⁴⁵, é comum a imposição das restrições $\sum_{k=1}^K \rho_{k,t} = 1$ e $\rho_{k,t} \geq 0$ sobre os valores que os pesos adaptativos em $\{\rho_{k,t}\}_{k=1}^K$ podem assumir, onde $t = 1, \dots, T+h$, a fim de preservar a sua interpretabilidade probabilística. Note que, dada a Definição 4.1, se $\sum_{k=1}^K \rho_{k,t} = 1$ e $\rho_{k,t} \geq 0$, então, caso um peso adaptativo assuma valor (ótimo ou estimado) igual a 1, todos os outros pesos adaptativos assumem, por consequência, valores iguais a zero. Tal situação é um caso especial de combinação linear de previsões, visto que, em t , a previsão linearmente combinada é idêntica à previsão do previsor base associado ao peso que assume valor igual a 1.

Em seguida, têm-se algumas abordagens para a obtenção de valores para os parâmetros adaptativos de combinação linear de métodos preditivos individuais:

- (a) Proposto por Granger & Bates (1969)⁵⁴, o chamado *método combinação ideal* é uma abordagem estatística que minimiza a variância dos resíduos da combinação de previsões, utilizando a método de condição de primeira ordem;
- (b) Em Smith & Makov (1978)⁵⁵, Mubwandarikwa (2007)⁵¹, o *método de Quase-Bayes* estima os pesos adaptativos por meio de um problema de classificação utilizando a distribuição de *Dirichlet*;

- (c) Conhecido como *método de pesos de Akaike* e proposto por Teixeira Jr (2009)⁵³, utiliza o critério de informação de *Akaike* (AIC) (Mubwandarikwa, 2007)⁵¹, para obtenção de uma fórmula analítica para o cálculo dos pesos adaptativos;
- (d) Em Mubwandarikwa (2007)⁵¹, é proposto o denominado *método de pesos por priori informativa*, que estabelece, inicialmente, aos pesos adaptativos (interpretados como variáveis aleatórias) uma distribuição de probabilidade *a priori*. À medida que os dados são observados, a distribuição de probabilidade é corrigida, isto é, é gerada uma distribuição de probabilidade *a posteriori*, através da conjugação da distribuição de probabilidade *a priori* e a *função de verossimilhança* (Casella & Berger)⁵⁶ condicionada às observações;
- (e) Em Teixeira Jr (2009)⁵³, os valores para os parâmetros adaptativos da combinação linear de previsões tradicional são obtidos numericamente, por meio de um problema de otimização, cujo objetivo é o mínimo da estatística MAPE; e
- (f) Em Teixeira Jr (2011a)⁵⁷ e em Teixeira Jr (2013)²⁴, os parâmetros adaptativos de uma combinação linear tradicional são interpretados como variáveis de decisão e os valores ótimos são obtidos meio de um PPM multi objetivo MINIMAX, com abordagem de programação por metas (Ragsdale, 2004)⁴¹.

Em síntese, pode-se verificar que os parâmetros adaptativos das combinações lineares de métodos preditivos base podem ser considerados: variantes ou invariantes no tempo; restritos ou irrestritos; e ser interpretados como variáveis aleatórias ou variáveis de decisão. Para visualizar outros exemplos, as seguintes referências: Makridakis (1989)⁵⁸; Winkler(1989)⁵⁹; Gupta & Wilton (1987)⁶⁰; Batchelor & Dua (1995)⁶¹; Flores & White(1998)⁶²; Armstrong, (2001)⁶³; Zou & Yang(2004)⁶⁴; Faria & Mubwandarikwa (2008)⁴⁵; Teixeira Jr (2012b)⁶⁵.

2.5

Amostrador Bootstrap

O método *bootstrap* (ou amostrador Bootstrap) é um procedimento estatístico não paramétrico introduzido por Efron (1979)⁶⁶. A técnica constitui na reamostragem de mesmo tamanho e com reposição dos dados da amostra original, com o objetivo de se estimar uma medida de interesse. O princípio básico do método *bootstrap* estabelece que na ausência de informações sobre a verdadeira distribuição de uma variável aleatória, pode-se obter uma distribuição empírica a partir de reamostragens de uma amostra observada. Considerando-se que esta amostra possui informações da distribuição base, **B** amostras daquela permitirão estimar parâmetros da população da qual foi retirada.

Considere $X = (X_1, X_2, \dots, X_n)$ uma amostra aleatória de uma distribuição F e $\hat{\theta} = s(X)$ estimador de um parâmetro θ desta distribuição. Uma amostra bootstrap $X^* = (X_1^*, X_2^*, \dots, X_n^*)$ é obtida da amostra original X rearranjando os seus n elementos. Assim, uma das possibilidades para X^* é $X^* = (X_1^* = X_2, X_2^* = X_n, \dots, X_n^* = X_1)$, fixando a probabilidade de escolha em X de $1/n$ para cada uma das suas observações. A partir deste procedimento determinam-se B amostras bootstrap independentes da distribuição empírica \hat{F} , obtendo-se então o estimador bootstrap $\hat{\theta}^{*(b)} = s(X^{*(b)})$, com $b = 1, 2, \dots, B$, do parâmetro θ . Sendo o erro padrão do estimador bootstrap representado por (34).

$$\widehat{se}_{boot}(\hat{\theta}) = \left[\frac{1}{B-1} \sum_{b=1}^B (\hat{\theta}^{*(b)} - \hat{\theta}^*)^2 \right]^{\frac{1}{2}} \quad (34)$$

$$\hat{\theta}^* = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{*(b)}.$$

Onde: De acordo com Montgomery & Runger (2003)⁶⁷, a definição do número **B** de amostras bootstrap pode ser realizada com a observação da variação do desvio padrão do estimador bootstrap. A baixa variabilidade ou estabilidade desse valor indica o valor de **B** mais adequado. Com relação à previsão de séries temporais, segundo Souza & Camargo (2004)²,

existem duas maneiras para se fazer bootstrap em uma série de tempo: (a) bootstrap nos resíduos; (b) método moving blocks. Sendo o primeiro mais usado que o segundo em aplicações de bootstrap em séries temporais. Mais detalhes sobre a técnica bootstrap podem ser encontrados em Freedman & Peters (1984)⁶⁸, Efron & Tibshirani (1986)⁶⁹, Souza & Neto (1996)⁷⁰, dentre outros.

2.5.1

Intervalos de confiança e de previsão

A aplicação de modelos de regressão implica na avaliação das incertezas que cercam estimadores de parâmetros e variáveis aleatórias. Os intervalos de confiança e previsão são duas medidas estatísticas valiosas que desempenham esta importante função. O primeiro mede as variações concernentes aos parâmetros do modelo de regressão, enquanto o segundo é mais amplo e considera não só as variações relativas ao modelo de regressão, mas também as variações consequentes dos erros preditivos.

Considere os padrões $\{(y_t, x_t)\}_{t=1}^T$ e o modelo de regressão neural $y(x) = f(x, w) + e(x)$, onde $y(x)$ são os alvos, w são os parâmetros verdadeiros do modelo desconhecido $f(x, w)$ e $e(x)$ o ruído (variável aleatória) com média zero. Denotando por $g(x, \hat{w})$, onde \hat{w} são os estimadores de w , a aproximação de $f(x, w)$, sendo esta aproximação interpretada como a média dos alvos (y) dada a entrada x .

De acordo com Heskes (1997)⁷¹, para problemas de regressão podem-se distinguir dois diferentes aspectos: a acurácia da estimativa da regressão verdadeira e a acurácia das estimativas relativas às saídas observadas. O primeiro aspecto conduz à construção do intervalo de confiança, resultado da distribuição da quantidade $f(x, w) - g(x, \hat{w})$, o segundo aspecto trata da variação $y(x) - g(x, \hat{w})$ e leva ao intervalo de previsão, conforme (35).

$$y(x) - g(x, \hat{w}) = f(x, w) - g(x, \hat{w}) + e(x) \quad (35)$$

Dado (35), segundo Khosravi (2014)⁷², a variância total associada com o resultado da modelagem pode ser expresso como em (36).

$$\sigma_y^2(\mathbf{x}) = \sigma_g^2(\mathbf{x}) + \sigma_e^2(\mathbf{x}) \quad (36)$$

Onde $\sigma_g^2(\mathbf{x})$ é a incerteza do modelo de regressão ajustado e $\sigma_e^2(\mathbf{x})$ a medida da variância do ruído $\mathbf{e}(\mathbf{x})$. Tibshirani (1996)⁷³ sugere, no contexto das regressões, dois tipos de amostragens bootstrap que permitem determinar o intervalo de confiança: bootstrap pairs e bootstrap residual. A seguir, um resumo dos passos que constituem o algoritmo da amostragem *bootstrap pairs*:

- (1) Gerar \mathbf{B} amostras ($20 \leq \mathbf{B} \leq 200$) cada uma com tamanho \mathbf{n} , obtidas com reamostragem das \mathbf{n} observações de treinamento, representada por $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$. Denotando as \mathbf{B} amostras bootstrap por $\{(\mathbf{x}_1^{*b}, y_1^{*b}), (\mathbf{x}_2^{*b}, y_2^{*b}), \dots, (\mathbf{x}_n^{*b}, y_n^{*b})\}$, com $\mathbf{b} = 1, \dots, \mathbf{B}$;

- (2) Para cada amostra bootstrap, minimize $\sum_{i=1}^n [y_i^{*b} - g(\mathbf{x}_i^*, \mathbf{w})]^2$ para obter $\hat{\mathbf{w}}^{*b}$; e

- (3) Estimar o erro padrão do i -ésimo valor predito usando o seguinte estimador: $\left\{ \frac{1}{\mathbf{B}-1} \sum_{b=1}^{\mathbf{B}} [g(\mathbf{x}_i, \hat{\mathbf{w}}^{*b}) - m(\mathbf{x}_i)]^2 \right\}^{1/2}$, onde $m(\mathbf{x}_i) = \frac{1}{\mathbf{B}} \sum_{b=1}^{\mathbf{B}} g(\mathbf{x}_i, \hat{\mathbf{w}}^{*b})$.

O algoritmo do bootstrap residual apresenta alguns passos em comum com o bootstrap pairs. A principal diferença reside na construção das séries bootstrapadas, segue o procedimento para execução do bootstrap residual, tal como:

- (1) Utilizar os dados observados para determinar $\hat{\mathbf{w}}$, fazendo $r_i = y_i - g(\mathbf{x}_i, \hat{\mathbf{w}})$, $i = 1, \dots, n$;
- (2) Gerar \mathbf{B} amostras bootstrap, cada uma com tamanho \mathbf{n} e obtidas com a reamostragem de $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n$. Representando por $\mathbf{r}_1^{*b}, \mathbf{r}_2^{*b}, \dots, \mathbf{r}_n^{*b}$ as \mathbf{B} amostras bootstrap, com $\mathbf{b} = 1, 2, \dots, \mathbf{B}$. Faça $y_i^{*b} = g(\mathbf{x}_i, \hat{\mathbf{w}}) + \mathbf{r}_i^{*b}$;

- (3) Para cada amostra bootstrap, minimize $\sum_{i=1}^n [y_i^{*b} - g(x_i^*, w)]^2$, determinando \hat{w}^{*b} ; e
- (4) Estimar o erro padrão do i-ésimo valor predito usando o seguinte $\left\{ \frac{1}{B-1} \sum_{b=1}^B [g(x_i, \hat{w}^{*b}) - m(x_i)]^2 \right\}^{1/2}$ onde $m(x_i) = \frac{1}{B} \sum_{b=1}^B g(x_i, \hat{w}^{*b})$.

De acordo com Khosravi (2014)⁷², o intervalo de predição pode ser expresso por (37).

$$m(x) - t_{1-\frac{\alpha}{2}, g^l} \sqrt{\sigma_y^2(x)} \leq y(x) \leq m(x) + t_{1-\frac{\alpha}{2}, g^l} \sqrt{\sigma_y^2(x)} \quad (37)$$

Onde g^l são graus de liberdade e α o nível de significância em uma distribuição t de Student. Os autores definiram g^l como sendo a diferença entre o número de amostras de treinamento e o número de parâmetros do modelo RNA. Seguindo nesta linha, o intervalo de confiança pode ser escrito na forma representada em (38).

$$m(x) - t_{1-\frac{\alpha}{2}, g^l} \sqrt{\sigma_g^2(x)} \leq f(x, w) \leq m(x) + t_{1-\frac{\alpha}{2}, g^l} \sqrt{\sigma_g^2(x)} \quad (38)$$

Carney et al (1999)⁷⁴ obtiveram intervalo de confiança e previsão a partir de um *ensemble* (combinação de diversos resultados de previsão) de RNA com *bootstrap*. Os autores sugerem dividir o conjunto com B redes bootstrap em M subconjuntos de tamanho β , executando os seguintes passos:

- (1) Determinar $m^{*i}(x)$ em cada um dos M subconjuntos formados por β redes e montar o conjunto $E = \{m^{*i}(x)\}_{i=1}^M$;
- (2) Obter B^* reamostragens bootstrap de E , formando o conjunto $G = \{E_j^*\}_j^{B^*}$, onde $E_j^* = \{m^{*j1}(x), \dots, m^{*jM}(x)\}$;
- (3) Calcular a variância em cada um dos B^* conjuntos E_j^* :

$$\sigma_j^{2*} = \frac{1}{M} \sum_{k=1}^M (m^{*jk}(x) - m^{*j}(x))^2 ; e$$

(4) Fazer a média das variâncias:

$$\sigma_g^2 = \frac{1}{B^*} \sum_{j=1}^{B^*} \sigma_j^2.$$

Tendo agora uma melhor estimativa para variância da distribuição $P(f(\mathbf{x}, \mathbf{w})|\mathbf{m}^*(\mathbf{x}))$, constrói-se o IC para $f(\mathbf{x}, \mathbf{w})$, conforme (39).

$$\mathbf{m}^*(\mathbf{x}) - z_{1-\alpha} \sigma_g \leq f(\mathbf{x}, \mathbf{w}) \leq \mathbf{m}^*(\mathbf{x}) + z_{1-\alpha} \sigma_g \quad (39)$$

Onde $z_{1-\alpha}$ provêm da distribuição normal padrão com nível de significância α e $\mathbf{m}^*(\mathbf{x})$ é obtido a partir das B amostragens bootstrap. Nix e Weigend (1995) [75] sugeriram uma técnica para estimar a média e a variância de uma distribuição de probabilidades de alvos em um problema de regressão por redes neurais artificiais. Os autores utilizaram para este fim a máxima verossimilhança para formar a função custo, descrita em (40), para treinar uma rede neural.

$$c = \sum_i \frac{1}{2} \left[\frac{(d_i - y(\mathbf{x}_i))^2}{\sigma_e^2(\mathbf{x}_i)} + \ln(\sigma_e^2(\mathbf{x}_i)) \right] \quad (40)$$

Onde $\mathbf{d}(\mathbf{x})$ são os valores observados (alvos) e $\mathbf{y}(\mathbf{x}) \approx \mathbf{f}(\mathbf{x})$ no modelo $\mathbf{d}(\mathbf{x}) = \mathbf{f}(\mathbf{x}) + \mathbf{e}(\mathbf{x})$. Esta proposta é frequentemente adaptada para calcular a variância do ruído ($\sigma_e^2(\mathbf{x})$), sendo utilizada pelos autores Heskes (1997)⁷¹, Carney et al (1999)⁷⁴, Khosravi, (2014)⁷² dentre outros. Bartkiewicz (2012)⁷⁶ realizou previsões através de redes neuro-fuzzy e estimou a variância do ruído ($\sigma_e^2(\mathbf{x})$) treinando uma rede com padrões $\{\mathbf{x}_i, |\mathbf{e}_i|\}$, $i = 1, \dots, n$, sendo $|\mathbf{e}_i| = |\mathbf{y}_i - \mathbf{g}(\mathbf{x}_i, \hat{\mathbf{w}})|$. O autor comparou o intervalo de previsão determinado através da sua proposta com intervalos oriundos de outras técnicas, e concluiu que os resultados foram satisfatórios.

Metodologia Proposta

Tome uma função característica $\chi_{[0,T]}$ [23], a qual é definida por $\chi_{[0,T]}(t) = \begin{cases} 1, & \text{set } t \in [0, T] \\ 0, & \text{set } t \notin [0, T] \end{cases}$ e uma série temporal de cardinalidade igual a T , denotada por $(y_t)_{t=1}^T$. Com efeito, de acordo com Teixeira Jr (2013) [24], a composição $\chi_{[0,T]}y = \tilde{y}: \mathbb{Z} \rightarrow \mathbb{R}$ é pode ser visualizada como um vetor em l^2 . Com efeito, tomando uma base ortonormal wavelet $\{\phi_{m_0,n}\}_{n \in \mathbb{Z}} \cup \{\omega_{m,n}\}_{(m,n) \in \{m\}_{m=m_0}^{+\infty} \times \mathbb{Z}}$ de l^2 , baseado na Seção 2.1, tem-se que série temporal \tilde{y} pode ser expandida, de forma aproximada, tal como segue, em (41).

$$\tilde{y} \simeq \tilde{y} = \sum_{n=1}^T a_{m_0,n} \phi_{m_0,n} + \sum_{m=m_0}^{m_0+(r-1)} \sum_{n=1}^T d_{m,n} \omega_{m,n} \quad (41)$$

O valor ótimo para o parâmetro de escala m_0 usualmente assume valor igual a r . Reescrevendo (41), de forma equivalente, segue que:

$$y \simeq y'_{A_{m_0}} + \sum_{m=m_0}^{m_0+(r-1)} y'_{D_m} \quad (42)$$

Onde: $y'_{V_{m_0}(\phi)} := \sum_{n=1}^T a_{m_0,n} \phi_{m_0,n}$ é uma “estimativa” da projeção ortogonal $y_{A_{m_0}}$ (Seção 2.1); e $y'_{W_m(\omega)} := \sum_{n=1}^{n_m} d_{m,n} \omega_{m,n}$ é uma “estimativa” da projeção ortogonal y_{D_m} (Seção 2.1).

Uma vez obtidas as componentes wavelet (CWs), em (42), realiza-se a sua modelagem individual por meio de um modelo de Box & Jenkins (abordagem linear), a fim de ser mapear estruturas de dependência linear de cada uma. Ou seja, são requeridos $r+1$ modelos de Box & Jenkins distintos. Assim, são obtidos os seguintes $r+1$ conjuntos de previsões, para cada inteiro m no intervalo $m_0 \leq m \leq m_0 + (r - 1)$, sendo h o horizonte de previsão e r , o nível de decomposição wavelet:

$$\hat{y}_{A_{m_0,BJ}} = \left\{ \hat{y}_{A_{m_0,BJ},t} \right\}_{t=t'+1}^{T+h}; \text{ e } \hat{y}_{D_{m,BJ}} = \left\{ \hat{y}_{D_{m,BJ},t} \right\}_{t=t'+1}^{T+h}.$$

Onde t' representa os graus de liberdade perdidos na modelagem de Box e Jenkins. Em seguida, é realizada a combinação linear wavelet (CLW) das referidas previsões, conforme definido em (43)

$$\hat{L}_t = \sum_{k=0}^{t-1} \hat{y}_{A_{m_0,BJ},t-k} \times \hat{\rho}_{A_{m_0,BJ},t-k} + \sum_{m=m_0}^{m_0+(r-1)} \sum_{j=0}^{t-1} \hat{y}_{D_{m,BJ},t} \times \hat{\rho}_{D_{m,BJ},t} + \text{SARIMA}(p,d,q) \times (P,D,Q)_S \quad (43)$$

Onde: SARIMA $(p,d,q) \times (P,D,Q)_S$ é uma estrutura da modelagem de Box e Jenkins; $\hat{y}_{A_{m_0,BJ},t-k}$ é a previsão de um modelo de Box & Jenkins para $y_{A_{m_0},t-k}$; $\hat{\rho}_{A_{m_0,BJ},t-k}$ é peso adaptativo ótimo associado linearmente à previsão $\hat{y}_{A_{m_0,BJ},t-k}$; $\hat{y}_{D_{m,BJ},t}$ é a previsão é a previsão do modelo de Box & Jenkins do estado $y_{D_{m},t-k}$; e $\hat{\rho}_{D_{m,BJ},t}$ é o peso adaptativo ótimo linear associado a $\hat{y}_{D_{m,BJ},t}$.

O ajuste ótimo dos pesos adaptativos em (43) ocorre por meio de um *problema de programação matemática* (PPM) o qual é descrito abaixo $\hat{\rho}_{D_{m,BJ},t}$.

$$\begin{aligned} \text{FO: } & \underset{\rho_{A_{m_0,BJ},t-k}; \rho_{D_{m,BJ},t}}{\text{minimizar}} \quad \text{MSE}_{\text{CLW}} \\ & \text{s.a.:} \end{aligned}$$

$$(1) \hat{L}_t = \sum_{k=0}^{t-1} \hat{y}_{A_{m_0,BJ},t-k} \times \rho_{A_{m_0,BJ},t-k} + \sum_{m=m_0}^{m_0+(r-1)} \sum_{j=0}^{t-1} \hat{y}_{D_{m,BJ},t} \times \rho_{D_{m,BJ},t} + \text{SARIMA}(p,d,q) (P,D,Q)_S;$$

$$(2) \text{MSE}_{\text{CLW}} = \sum_{t=t'+1}^T \frac{(y_t - \hat{L}_t)^2}{T - t'};$$

(3) testes residuais (Análise da função de autocorrelação parcial e simples, Intervalo de confiança, estatística Durbin Watson);

(4) significância dos pesos adaptativos.

3.1

RNA-MLP Wavelet Automática

A integração de uma RNA-MLP e uma decomposição wavelet (DW) de nível r' pode ser realizada de diferentes maneiras. Por exemplo, pode-se aplicar a decomposição wavelet a uma série temporal a ser projetadas e se

utilizar uma RNA-MLP para prever cada componente wavelet, separadamente, e, ao final, somar suas previsões, produzindo as previsões da série temporal subjacente (Teixeira Jr et al, 2011a)⁵⁷. Outra opção consiste em usar as funções wavelet (normalizadas no intervalo $[0,1]$) como funções de ativação dos neurônios de uma RNA tradicional e utilizar os padrões de entrada decompostos através de DW.

Aqui, todavia, optou-se por uma abordagem que combina ambas (denotada por RNA-DW), na qual as componentes wavelet de uma série temporal de interesse constituem os padrões de entrada de uma RNA-MLP cuja saída fornece a previsão da série temporal de interesse (conforme o esquema da Figura 7).

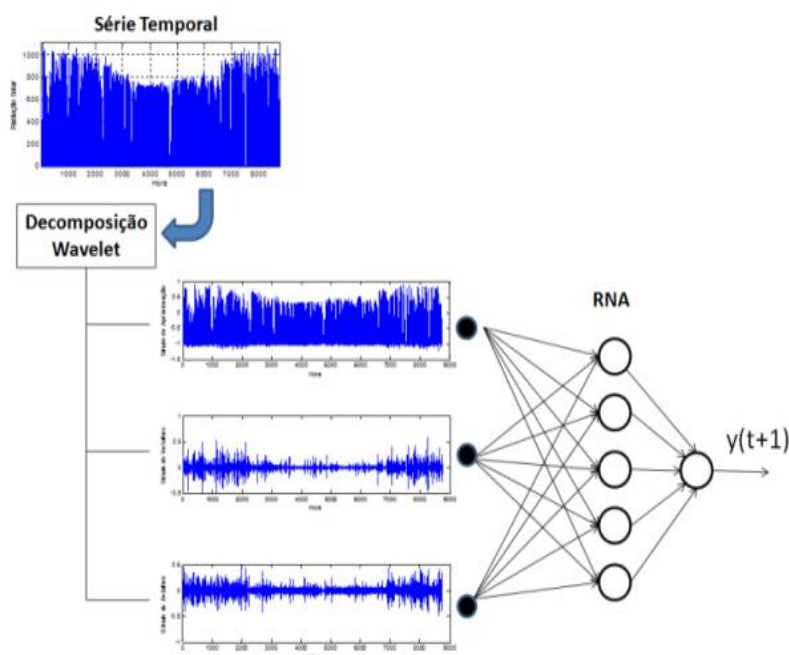


Figura 7 - Combinação da RNA com Decomposição *Wavelet*.

Fonte: O autor (2015).

Na Figura 7, tem-se que são tomados os resíduos produzidos na equação 43, denotados por \mathbf{a}_t ($t = t' + 1, \dots, T$) (onde $\mathbf{a}_t = \mathbf{y}_t - \hat{\mathbf{y}}_t$ é dotado de estrutura de autodependência não linear). Os referidos são decompostos via decomposição wavelet de nível r' e, em seguida as suas $r'+1$ CWs geradas são imputadas, simultaneamente, em RNA-MLP com função de ativação na

camada escondida não linear. O objetivo é gerar previsões para \mathbf{a}_t ($t = t' + 1, \dots, T$).

Tal procedimento é realizado de forma automática por meio de um “laço” computacional. Assim, o algoritmo de busca da configuração ótima na qual considera os seguintes parâmetros:

- (1) O nível r' de decomposição wavelet de \mathbf{a}_t ($t = t' + 1, \dots, T$);
- (2) A base ortonormal wavelet a ser utilizada;
- (3) A normalização dos dados;
- (4) O tamanho de janela j , onde $j=1, \dots, J$ (J é definido pelo tomador de decisão);
- (5) A função de ativação;
- (6) O número de vezes que cada configuração é treinada; e
- (7) Algoritmo de treinamento.

Uma vez definido os parâmetros em (1), (2) e (3), o conjunto das componentes wavelet de \mathbf{a}_t ($t = t' + 1, \dots, T$),

$$\left\{ \left\{ \mathbf{a}_{A_{m_0}}(t) \right\}_{t=t'+1}^T \right\} \cup \left\{ \left\{ \mathbf{a}_{D_m}(t) \right\}_{t=t'+1}^T \right\}_{m=m_0}^{m_0+(r'-1)},$$

são tais que constituem o conjunto de padrões de entrada de uma RNA-MLP para realização do processo de treinamento. Dado uma janela de tamanho igual a j valores passados, a previsão da série temporal (saída de RNA) para cada instante t_0 (nas amostras de treino, validação e de teste) é obtida a partir do conjunto de padrões de entrada dado em (44).

$$\left\{ \left\{ \mathbf{a}_{A_{m_0}}(t) \right\}_{t=t_0-j}^{t_0-1} \right\} \cup \left\{ \left\{ \mathbf{a}_{D_m}(t) \right\}_{t=t_0-L}^{t_0-1} \right\}_{m=m_0}^{m_0+(p-1)} \quad (44)$$

Os parâmetros da RNA-MLP ótimos são tais que minimizam a soma dos erros quadráticos da amostra de treinamento. Adotou-se a mesma metodologia encontrada na literatura para fins de comparação. Portanto, a figura da amostra de validação não foi utilizada nesta tese, somente a amostra de teste, conforme propõe Zhang (2003)⁴.

Em síntese, a metodologia híbrida proposta aqui pode ser executada em sete passo:

- ✓ *Passo 1:* decomposição wavelet de nível r da série temporal de interesse;
- ✓ *Passo 2:* modelagem individual de cada uma das $r+1$ CW via modelo SARIMA;
- ✓ *Passo 3:* Combinação linear wavelet (CLW) das previsões geradas no passo 2;
- ✓ *Passo 4:* Decomposição wavelet de nível k dos resíduos produzidos no passo 3;
- ✓ *Passo 5:* Modelagem simultânea das $k+1$ CWs geradas no passo 4 via RNA-MLP wavelet automática;
- ✓ *Passo 6:* Geração das previsões (pontuais) híbridas da série temporal original, somando as previsões obtidas nos passos 3 e 6; e
- ✓ *Passo 7:* Geração das previsões intervalares (intervalos de previsão) via bootstrap.

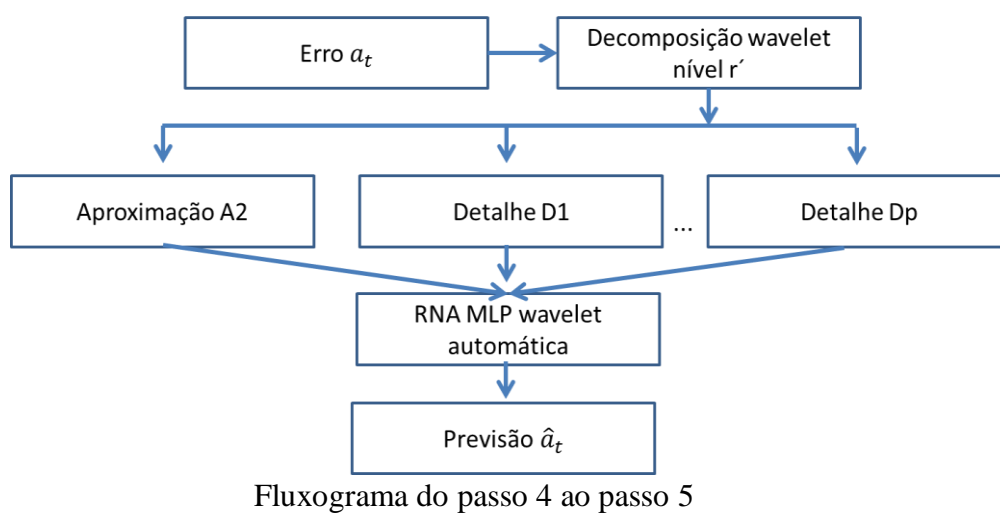
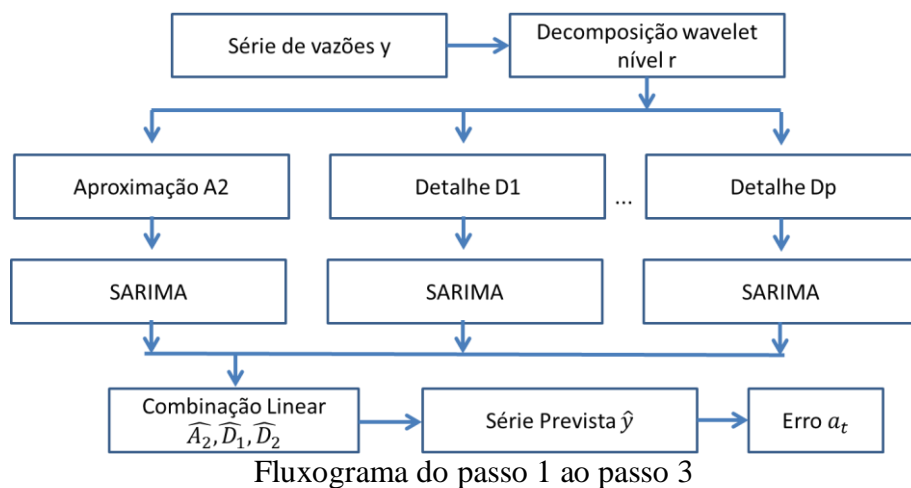


Figura 8 - Fluxograma.

Fonte: O autor (2015).

O passo 6 é dado pelo somatório da série prevista \hat{y} e da previsão \hat{a}_t .

4

Resultados empíricos

4.1

Conjuntos de dados

Na presente seção, além de uma série temporal mensal da usina de Itaipu (descrita mais adiante) são utilizados três conjuntos de dados oriundos de diferentes áreas e bem conhecidas na literatura (a saber, Sunspot, Canadian lynx e Exchange Rate [4]), para demonstrar a eficiência da metodologia híbrida wavelet proposta.

Tais conjuntos de dados são dotados de diferentes características estatísticas e têm sido bastante estudadas no campo da Estatística e no de métodos de aprendizagem máquina. Em especial, vários métodos preditivos lineares e não lineares têm sido aplicados a estes dados, com o objetivo de se gerar previsões pontuais.

A série temporal Sunspot (manchas solares) contém os registros anuais de 1700 até 1987, resultando em 288 observações. Zhang (2003)⁴ destaca que o estudo da atividade das manchas solares possui importância prática, por exemplo, para geofísicos, cientistas do meio-ambiente e climatologistas. Ainda de acordo com o autor, é estatisticamente não linear e não-Gaussiana.

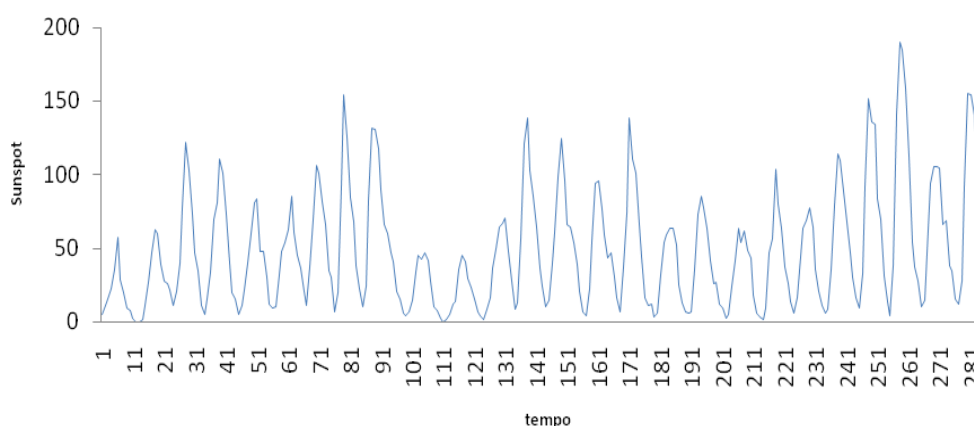


Figura 9 - Série temporal Manchas Solares (1700-1987).

Fonte: O autor (2015)

Por sua vez, a série temporal Canadian Lynx consiste no número de lincos presos por ano no Rio Mackenzie, no distrito do norte do Canadá. Os referidos dados são representados graficamente na Figura 10, na qual se pode perceber uma periodicidade de cerca de 10 anos. Tal conjunto de dados possui 114 pontos, e corresponde ao período de 1821-1934. Também tem sido frequentemente analisada na literatura de séries temporais, com foco na modelagem preditiva, principalmente não linear. Seguindo outros estudos, como por exemplo, Zhang (2003)⁴, são considerados os logaritmos (na base 10) dos dados para serem utilizados para fins de previsão, e não os seus valores originais.

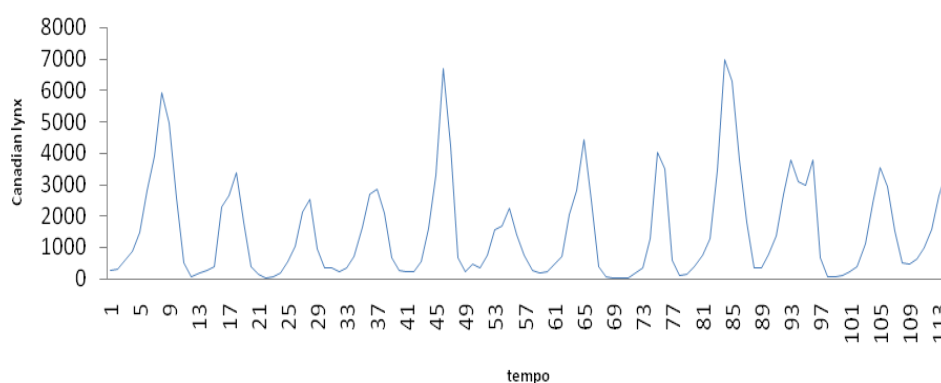


Figura 10 - Annual time series of Canadian lynx (1821-1934).

Fonte: O autor (2015).

O último conjunto de dados temporais é a taxa de câmbio libra esterlina/dólar americano (ou Exchange Rate pound/dolar). Prever a taxa de câmbio é uma tarefa importante ainda e de altíssima complexidade, de maneira que vários modelos preditivos lineares e não lineares foram com este objetivo (Zhang, 2003)⁴. Tal série temporal contém observações com frequência semanal, no período 1980-1993, resultando em 731 pontos. O seu gráfico é exibido na Figura 11. Assim, seguindo os autores considerados para fins de comparação, foi usada a sua versão transformada por meio de uma função logarítmica com base natural, na sua predição temporal. Note que a mesma é, claramente, não estacionária.

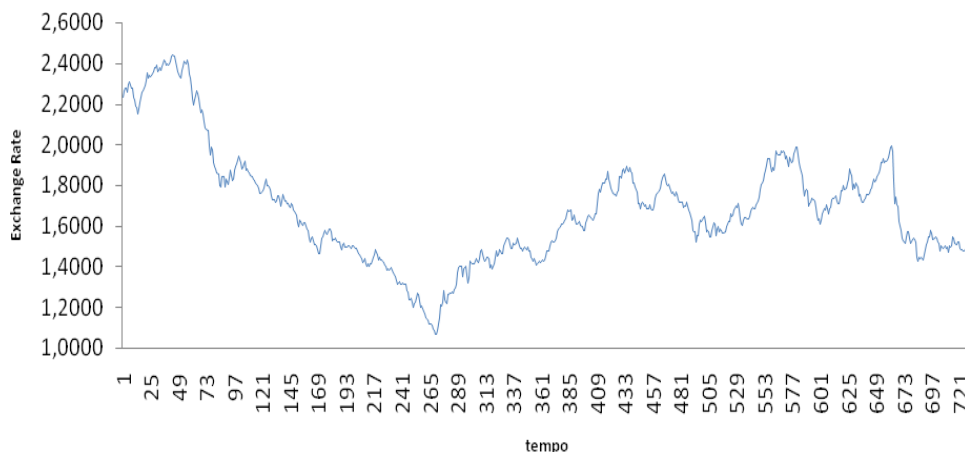


Figura 11 - Série temporal semanal Exchange Rate (1980-1993).

Fonte: O autor (2015).

Segundo Zhang (2003)⁴, para avaliar o desempenho de modelos de previsão aqui considerados, na projeção destas três séries temporais, os conjuntos de dados foram divididos em duas amostras, uma de treinamento e uma de teste. O conjunto de dados de treinamento foi usado, exclusivamente, para o ajuste dos parâmetros dos métodos preditivos; enquanto que a amostra de teste, para avaliar o desempenho preditivo do modelo estabelecido. A divisão de dados (que é a mesma utilizada pelos outros autores considerados aqui) para os três conjuntos de dados são detalhados na Tabela 1.

Tabela 1 - Composição das amostras de treino e de teste, nos três conjuntos de dados.

<i>Dados</i>	<i>Tamanho da amostra</i>	<i>Amostra de treino (tamanho)</i>	<i>Amostra de teste (tamanho)</i>
Sunspot	288	1700–1920 (221)	1921–1987 (67)
Canadian lynx	114	1821–1920 (100)	1921–1934 (14)
Exchange rate	731	1980–1992 (679)	1993 (52)

Fonte: O autor (2015)

Os tamanhos das amostras de treino e de teste foram utilizados seguindo a metodologia utilizada na literatura para fins de comparação. A série e Itaipu será apresentada mais adiante afim de oferecer mais uma contribuição na literatura.

4.2

Modelagem

Para a etapa de modelagens de Box e Jenkins, implementadas no software EViews 8, foram considerados as seguintes análises residuais para a validação: análise de perfil dos gráficos das funções de autocorrelação parcial e simples, teste Ljung-Box (até o Lag 30), estatística de Durbin-Watson, teste de raiz unitária de Dickey-Fuller (Hamilton, 1994)³²

Na modelagem dos resíduos oriundos da CLW de modelos de Box e Jenkins, quanto aos parâmetros wavelet, considerou-se, somente o nível de decomposição wavelet igual a 2. Para essas séries a utilização de um nível de decomposição maior não oferece grande ganho preditivo.

No que tange às bases ortonormais wavelet, consideram-se as seguintes famílias: db1, db2, ..., db45 (ou seja, as famílias de Daubechies com momentos nulos tomando valores entre 1 e 45); sym1,... sym20 (ou seja, as famílias de Daubechies com momentos nulos tomando valores entre 1 e 20) (Mallat, 1998)¹⁵. Quanto à RNA-MLP, foi utilizada somente uma camada escondida e, como algoritmo de treinamento, Levenberg-Marquardt (Haykin, 2001)². Assumiu-se ainda para a RNA-MLP tamanho de janela varando de 1 até 12, número de treinamentos por configuração igual a 25. A modelagem neuronal e a abordagem wavelet foram implementadas no MATLAB 2013a. Finalmente, os parâmetros adaptativos da CLW foram obtidos com auxílio do pacote solver, do Excel 2009, mantendo como inalterados os valores default.

As estatísticas de aderência utilizadas foram: o MSE (*mean square error*) e o MAE (*mean absolute error*), as quais são definidas, respectivamente, por:

$$\text{MSE} = \frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2 \quad \text{e} \quad \text{MAE} = \frac{1}{T} \sum_{t=1}^T |y_t - \hat{y}_t|.$$

Onde: y_t e \hat{y}_t são, respectivamente, o estado e a previsão no instante t . Considerou-se ainda a estatística de aderência MAPE, definida por:

$$\text{MAPE} = \frac{1}{T} \sum_{t=1}^T \frac{|y_t - \hat{y}_t|}{y_t}.$$

4.2.1

Modelagem da série temporal Sunspot

No *passo 1*, uma decomposição wavelet de nível 2, com base ortonormal de Haar [27], da amostra de treino da série temporal Sunspot foi realizada, gerando-se uma CW de aproximação e duas de detalhe. Para modelá-las, no *passo 2*, foi utilizado:

- (a) Um modelo ARIMA (28,1,32), com transformação logarítmica, para prever a CW de aproximação de nível 2. Assuma que $\log(A_{2,t}) - \log(A_{2,t-1}) = z_t$ seja, para cada instante t , a primeira diferença de $A_{2,t}$ (que é CW de aproximação de nível 2 para o estado y_t). Assim, o modelo linear supracitado é algebricamente dado por:

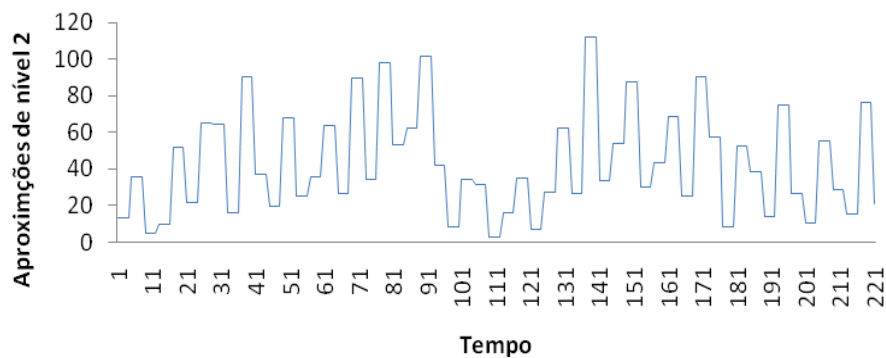
$$z_t = \phi_4 z_{t-4} + \phi_8 z_{t-8} + \phi_{20} z_{t-20} + \phi_{28} z_{t-28} + \theta_{20} a_{20} + \theta_{32} a_{32} + a_t.$$

- (b) Um modelo ARIMA (20,1,28), para projetar a CW de detalhe de nível 2, gerada no passo 1. Assumindo que $z'_t = D_{2,t} - D_{2,t-1}$, para cada instante t , segue que o modelo mencionado é (algebricamente) definido por:

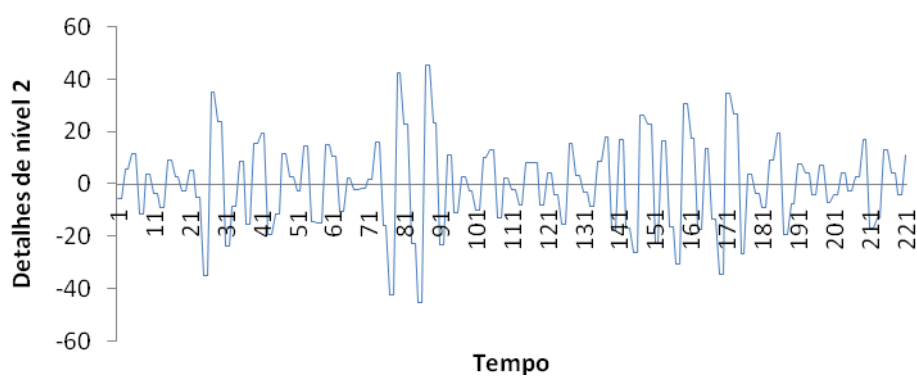
$$(c) \quad z'_t = \phi_2 z'_{t-2} + \phi_4 z'_{t-4} + \phi_{16} z'_{t-16} + \phi_{18} z'_{t-18} + \phi_{20} z'_{t-20} + \theta_4 a_{t-4} + \theta_8 a_{t-8} + \theta_{12} a_{t-12} + \theta_{22} a_{t-22} + \theta_{28} a_{t-28} + a_t.$$

- (d) Um modelo ARIMA (20,0,1) para prever a CW de detalhe de nível 3 (produzida no passo 1) cuja formulação matemática é dada por:

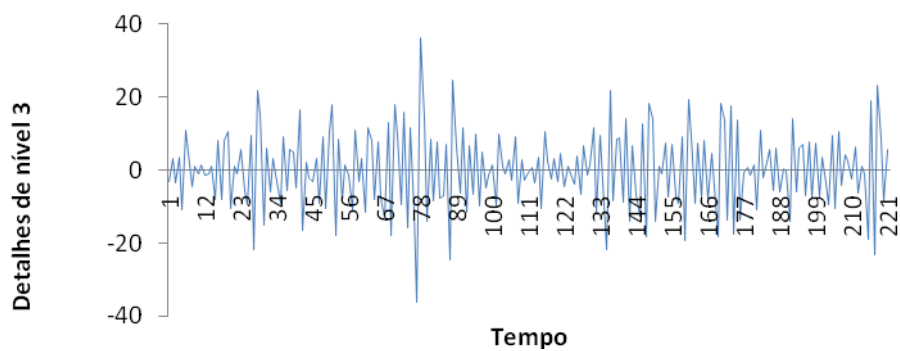
$$D_{3,t} = \phi_4 D_{3,t-4} + \phi_6 D_{3,t-6} + \phi_{10} D_{3,t-10} + \phi_{20} D_{3,t-20} + \theta_1 a_{t-1} + a_t.$$



(a) CW de aproximação de nível 2, $A_{2,t}(t = 1, \dots, 221)$.



(b) CW de detalhe de nível 2, $D_{2,t}(t = 1, \dots, 221)$.



(c) CW de detalhe de nível 3, $D_{3,t}(t = 1, \dots, 221)$.

Figura 12 - CWs da amostra de treinamento da série temporal (anual) Sunspot.

Fonte: O autor (2015).

A CLW ótima, no *passo 3*, é dada na Equação (45).

$$\begin{aligned}\hat{z}_t = & (\hat{y}_{A_2,BJ,t} - \hat{y}_{A_2,BJ,t-1}) \times 0,499067 + (\hat{y}_{A_2,BJ,t-1} - \hat{y}_{A_2,BJ,t-2}) \times \\ & 0,305268 + \hat{y}_{D_2,BJ,t} \times 1,370289 + \hat{y}_{D_2,BJ,t-1} \times 0,411365 - \hat{y}_{D_2,BJ,t-2} \times \\ & 0,675322 - \hat{y}_{D_2,BJ,t-3} \times 0,978313 + \hat{y}_{D_3,BJ,t} \times 0,579351 - \hat{y}_{D_3,BJ,t-1} \times \\ & 0,433425 - \hat{y}_{D_3,BJ,t-3} \times 0,316267 - \hat{z}_{t-1} \times 0,304979 - \hat{z}_{t-2} \times \\ & 0,388336 - \hat{z}_{t-3} \times 0,272105 - \hat{z}_{t-16} \times 0,226281 \quad (45)\end{aligned}$$

Onde: $\hat{z}_t = \hat{L}_t - \hat{L}_{t-1}$, para todo instante t ; $\hat{z}_{t-16} \times 0,226281$ é uma componente sazonal da estrutura Box e Jenkins.

Na saída obtida no EViews, abaixo, pode-se verificar os valores dos parâmetros adaptativos ótimos da CLW, exibidos na Equação (45). Pode-se ainda visualizar três aspectos relevantes: (1) que foi utilizada a primeira diferença na variável de saída (em destaque no topo) - o que é absolutamente pertinente, visto que a série temporal Sunspot é não estacionária de segunda ordem; (2) que o método de estimação foi o de mínimos quadrados ordinários (conforme descrito no PNL na Seção 3); e (3) que a convergência ocorreu na nova interação (que incorreu em baixíssimo esforço computacional). Ademais, a 5% de significância, note que todos os valores ótimos da CLW são significantes (o que implicou a rejeição das hipóteses de eles serem não diferentes de zero, ao nível 5%) e que satisfazem às condições de estacionariedade e invertibilidade (que são restrições necessárias no caso de modelos estatísticos lineares) (Hamilton, 2009)⁷⁷. Quanto ao nível de explicação (comumente, referido como “estatística R^2 ”), a CLW ótima explicou (ou mapeou), aproximadamente, 82,7072% da variabilidade total dos dados, na amostra de treino. Por fim, uma vez que a estatística de Durbin-Watson toma um valor muito próximo ao teórico 2, tem-se forte evidência empírica de inexistência de autocorrelação de lag (defasagem) 1.

Tabela 2 - Estimativas das autocorrelações simples e parciais dos erros de previsão.

Variable	Coefficient t	Std. Error	t-Statistic	Prob.
D(A2,1)	0.499067	0.049996	9.982.111	0.0000
D(A2(-1),1)	0.305268	0.044116	6.919.628	0.0000
D2	1.370.289	0.104430	1.312.157	0.0000
D2(-1)	0.411365	0.109344	3.762.137	0.0002
D2(-2)	-0.675322	0.101110	-6.679.065	0.0000
D2(-3)	-0.978313	0.105535	-9.270.000	0.0000
D3	0.579351	0.073460	7.886.670	0.0000
D3(-1)	-0.433425	0.073789	-5.873.871	0.0000
D3(-3)	-0.316267	0.056216	-5.625.947	0.0000
AR(1)	-0.304979	0.084715	-3.600.057	0.0004
AR(2)	-0.388336	0.079361	-4.893.256	0.0000
AR(3)	-0.272105	0.084616	-3.215.783	0.0016
SAR(16)	-0.226281	0.085553	-2.644.938	0.0090
R-squared	0.827072	Mean dependent var		0.167073
Adjusted R-squared	0.813329	S.D. dependent var		2.172.652
S.E. of regression	9.387.028	Akaike info criterion		7.392.485
Sum squared resid	13305.56	Schwarz criterion		7.638.206
Log likelihood	-5.931.837	Hannan-Quinn criter.		7.492.238
Durbin-Watson stat	1.987.899			

Fonte: O autor (2015).

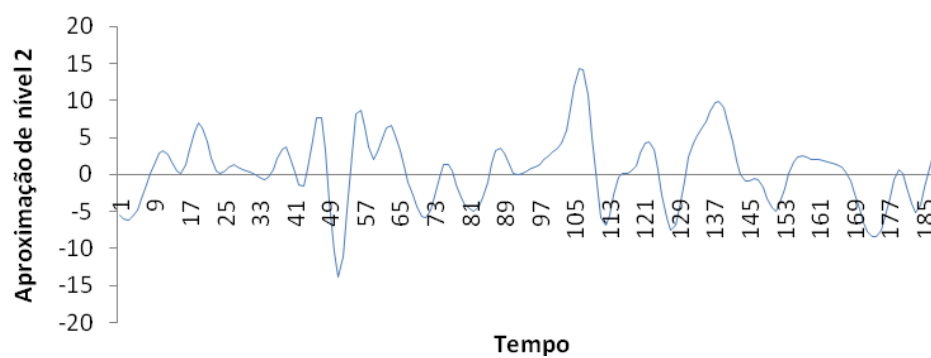
A Tabela 3 exibe as estimativas das autocorrelações simples e parciais dos erros de previsão, dentro da amostra, bem como os seus respectivos intervalos de confiança, além do teste de Ljung-Box, até o lag 30. Analisando-a, note que os valores das estimativas das autocorrelações simples e parciais não violam os limites dos intervalos de confiança (construídos ao nível de credibilidade de 95%), de maneira que não se rejeitou a hipótese nula; além disso, os p-valores do teste Ljung-Box tomaram valores superiores a 5% (nível de significância), até o lag 30, de modo que não se rejeitou a hipótese de inexistência de dependência linear (simples) nos erros de previsão de treino. Inferiu-se, assim, que os resíduos da CLW ótima, em (45) são desprovidos de estrutura de autodependência linear, para qualquer defasagem.

Tabela 3 - Gráficos da FAC e PACP e p-valores associados aos resíduos ordinários, na amostra de treino.

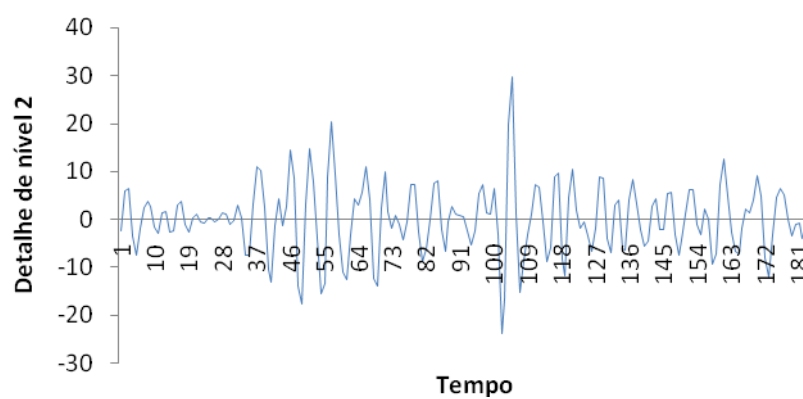
Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
		1	0.004	0.004	0.0023
		2	-0.018	-0.018	0.0586
		3	-0.050	-0.050	0.4831
		4	-0.053	-0.053	0.9654
		5	-0.114	-0.116	3.1762 0.075
		6	-0.049	-0.056	3.5967 0.166
		7	-0.067	-0.080	4.3664 0.225
		8	0.023	0.004	4.4612 0.347
		9	-0.011	-0.034	4.4826 0.482
		10	-0.065	-0.095	5.2226 0.516
		11	0.018	-0.005	5.2830 0.625
		12	0.083	0.058	6.5134 0.590
		13	-0.105	-0.125	8.5108 0.484
		14	-0.082	-0.105	9.7454 0.463
		15	0.007	-0.014	9.7549 0.553
		16	-0.008	-0.034	9.7674 0.636
		17	-0.004	-0.030	9.7703 0.713
		18	0.071	0.040	10.709 0.709
		19	0.015	-0.017	10.753 0.770
		20	0.034	-0.004	10.977 0.811
		21	0.030	0.025	11.148 0.849
		22	0.054	0.069	11.700 0.862
		23	0.039	0.033	11.991 0.886
		24	-0.075	-0.082	13.092 0.873
		25	-0.035	0.007	13.336 0.896
		26	-0.073	-0.063	14.399 0.887
		27	-0.035	-0.047	14.640 0.907
		28	-0.021	-0.015	14.726 0.929
		29	0.032	0.014	14.931 0.943
		30	-0.048	-0.088	15.394 0.950

Fonte: O autor (2015).

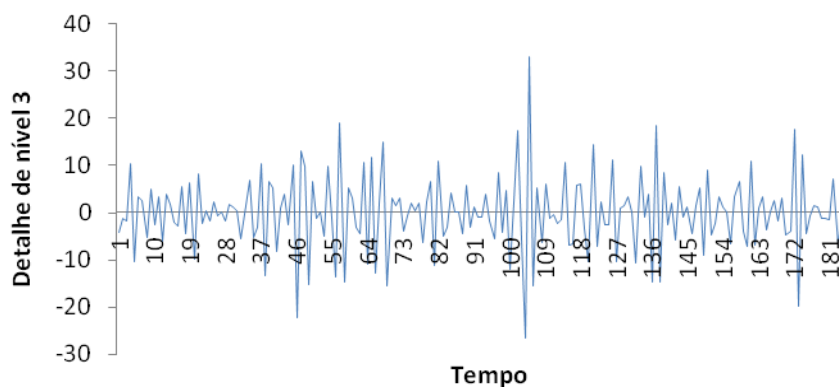
Por sua vez, no *passo 4*, os erros de previsão da CLW ótima do passo 3, denotados por $e_t(t = 35, \dots, 221)$, sendo que $t' = 34$ (graus de liberdade perdidos na modelagem até o passo 3), foram decompostos por meio de uma decomposição wavelet de nível 2, com base ortonormal wavelet de Daubechies com momento nulo igual a 8, cuja notação é dada por “db8”. Na Figura (40), podem-se visualizar as três CWs produzidas aqui.



(a) CW de aproximação de nível 2, $A'_{2,t}(t = 35, \dots, 221)$.



(b) WC of detail at level 2, $D'_{2,t}(t = 35, \dots, 221)$.



(c) WC of detail at level 3, $D'_{3,t}(t = 35, \dots, 221)$.

Figura 13 - CW dos erros de previsão da CLW ótima do passo 3.

Fonte: O autor (2015).

A configuração ótima da RNA wavelet automática do *passo 5* é exibida na Tabela 4.

Tabela 4 - Configuração ótima da RNA wavelet automática.

Neurônios	Normalização	Tamanho da janela	Função de ativação (camada escondida)	Função de ativação (saída)
5	premnmx	7	Tangente hiperbólica	linear

Fonte: O autor (2015).

Finalmente, no *passo 6*, as previsões produzidas nos passos 3 e 5 são simplesmente somadas, gerando previsões híbridas para a série temporal Sunspot, denotadas por $(t=36, \dots, 288)$, onde $t' = 35$. O gráfico com os valores observados e as suas respectivas previsões podem ser contemplados na Figura 14.

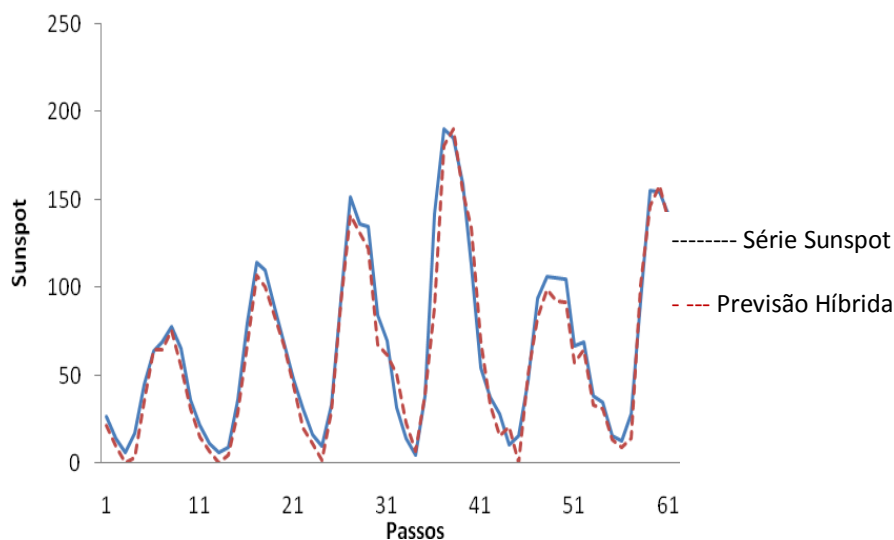


Figura 14 - Previsões híbridas e observações, na amostra de teste.

Fonte: O autor (2015).

A Tabela 5 abaixo mostra os valores das estatísticas de aderência MSE e MAE na amostra de teste (out of sample) do método híbrido proposto, como também de outros dez previsores competitivos existentes na literatura.

Tabela 5 - Estatísticas MSE e MAE, na amostra de teste.

REFERÊNCIAS	Previsores	35 passos à frente		67 passos à Frente	
		MSE	MAE	MSE	MAE
Zhang, G. P. (2003)	ARIMA model	217,0	11,32	306,0	13,03
	ANN	205,3	10,24	351,1	13,54
	Hybrid method	186,8	10,83	280,1	12,78
Adhikari, R. and Agrawal, R. K (2012)	ARIMA model	-	-	483,5	17,63
	ANN	-	-	494,9	15,58
	EANN	-	-	492,7	14,71
	Ensemble	-	-	274,7	12,50
Khashei, M. and Mehdi, B. (2011)	Artificial Neural Network (p,d,q)	125,8	8,94	234,2	12,18
	ANNs and ARIMA Generalized hybrid model	129,4	8,85	218,6	11,45
Khashei, M. and Bijari, M. (2012)	ARIMA and PNN hybrid model	148,8	9,29	234,77	11,55
Britto, R. V. P. (2015)	HWCF	66,5	6,92	137,40	8,71

Fonte: O autor (2015).

4.2.2

Modelagem da série temporal *Canadian lynx*

O conjunto de dados temporais *Canadian lynx* se refere ao registro anual do número de lincas canadenses (espécie de felino natural da América do Norte) capturados no distrito de Rio Mackenzie no Norte do Canadá. Existem, na literatura, vários trabalhos que fazem análise e previsão da série de tempo *Canadian lynx*. Credita-se, porém, a Moran (1953)⁷⁸, como sendo o primeiro artigo a realizar a análise desta. Outras abordagens, utilizadas para a sua previsão, podem ser encontradas, por exemplo, nos seguintes artigos: Zhang (2003)⁴; Subba & Sabr (1984)⁷⁹; Tang & Ghosal (2007)⁸⁰; e Wong & Li (2000)⁸¹. No que tange à predição da série temporal subjacente por meio da metodologia wavelet híbrida proposta, tem-se que, *no passo 1*, foi realizada

uma decomposição wavelet de nível 2, com base ortonormal de Haar, de sua amostra de treino, de modo a gerar uma CW de aproximação de nível 2; e duas de detalhe, de níveis 2 e 3. No passo 2, por sua vez, utilizaram-se os seguintes previsores para prevê-las:

- (a) Um modelo ARIMA (12,1,16) para prever a CW de aproximação de nível 2, com representação dada por:

$$z_t = \phi_4 z_{t-4} + \phi_{12} z_{t-12} + \theta_{12} a_{t-12} + \theta_{16} a_{t-16} + a_t.$$

Onde: $z_t = y_t - y_{t-1}$, para todo instante t .

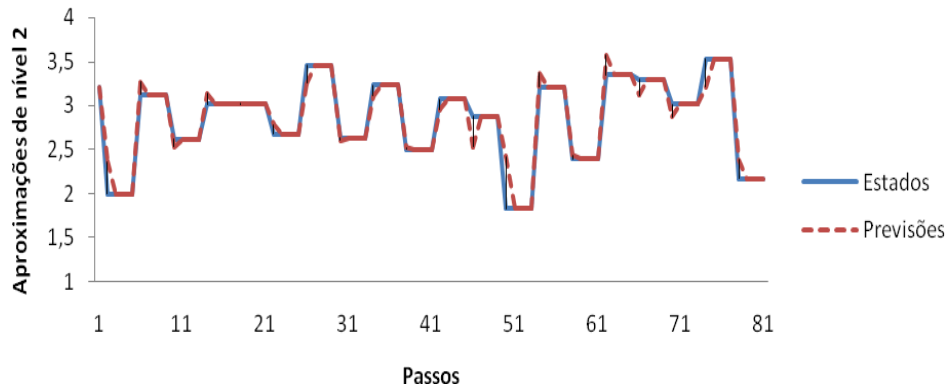
- (b) Um modelo ARIMA (12,1,17), para a CW de detalhe de nível 2.

Assumindo que $z'_t = D_{2,t} - D_{2,t-1}$, para cada instante t , segue que o referido modelo ótimo é dado por:

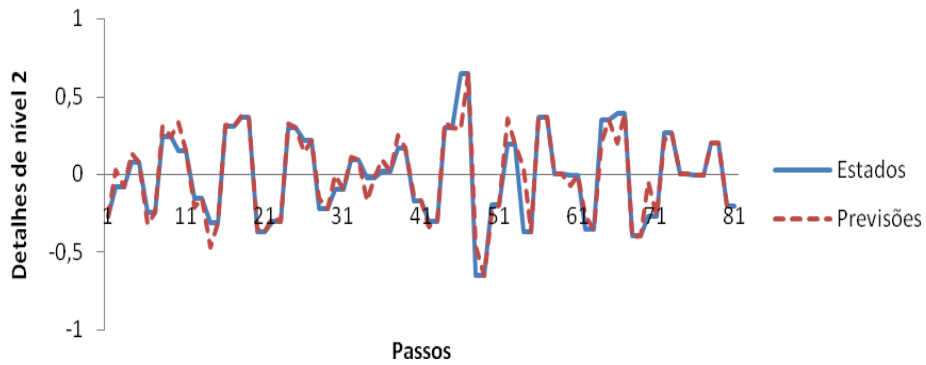
$$z'_t = \phi_4 z'_{t-4} + \phi_{12} z'_{t-12} + \theta_4 a_{t-4} + \theta_{12} a_{t-12} + \theta_{16} a_{t-16} + a_t.$$

- (c) Um modelo ARIMA (9,0,8) para a CW de detalhe de nível 3 com a seguinte formulação matemática:

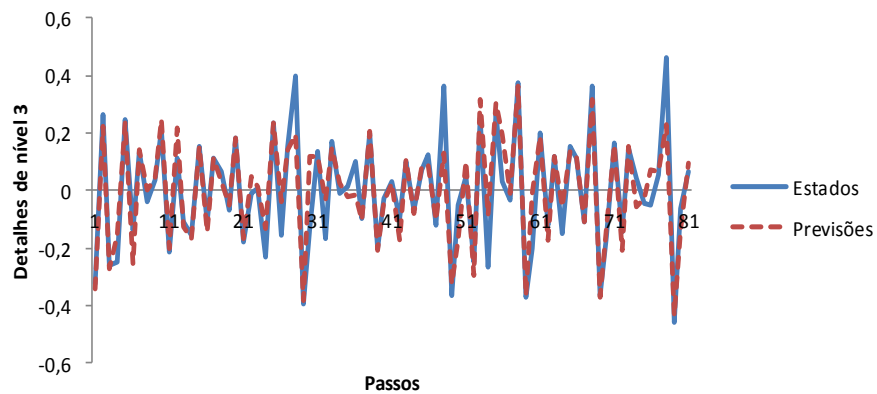
$$\begin{aligned} D_{3,t} = & \phi_1 D_{3,t-1} + \phi_2 D_{3,t-2} + \phi_3 D_{3,t-3} + \phi_4 D_{3,t-4} + \phi_5 D_{3,t-5} + \\ & + \phi_6 D_{3,t-6} + \phi_7 D_{3,t-7} + \phi_8 D_{3,t-8} + \phi_9 D_{3,t-9} + \theta_2 a_{t-2} + \\ & \theta_2 a_{t-2} + a_t. \end{aligned}$$



(a) CW de aproximação de nível 2 $A_{2,t}(t = 19, \dots, 100)$ e suas previsões, $\hat{A}_{2,t}(t = 19, \dots, 100)$.



(b) CW de detalhe de nível 2 $D_{2,t}(t = 19, \dots, 100)$ e suas previsões, $\hat{D}_{2,t}(t = 19, \dots, 100)$.



(c) CW de detalhe de nível 3 $D_{3,t}(t = 19, \dots, 100)$ e suas previsões, $\hat{D}_{3,t}(t = 19, \dots, 100)$.

Figura 15 - CW dos erros de previsão da CLW ótima do passo 3.

Fonte: O autor (2015).

Quanto à combinação CLW ótima do *passo 3*, tem-se que se utilizou, como variável de saída, a série temporal subjacente diferenciada uma vez. Em que os pesos adaptativos ótimos (calculados por Problema de Programação Matemática), são tais que: $\hat{\rho}_{A_2,BJ,t} = 1,039151$, $\hat{\rho}_{D_2,BJ,t} = 1,0152498$, $\hat{\rho}_{D_2,BJ,t-1} = 1,0177965$, $\hat{\alpha} = -0,128429$ e $\phi_1 = -0,98356$, $\theta_2 = -0,283016$. Como variável de saída da CLW, se utilizou a primeira diferença da série temporal subjacente. Quanto à estatística R^2 , a mesma obteve valor próximo a 82,1035%.

Por sua vez, no *passo 4*, os erros de predição da CLW ótima no passo 3, representados por e_t ($t = 19, \dots, 100$), sendo que $t' = 18$ (graus de liberdade perdidos até o passo 3), foram decompostos através de uma decomposição wavelet de nível 2, com base ortonormal wavelet de Daubechies com momento nulo igual a 1 cuja notação é dada por “db1”. Por sua vez, no passo 5, a configuração ótima da RNA wavelet automática (utilizada na projeção dos resíduos produzidos no passo 3) pode ser visualizada na Tabela 4.

Tabela 6 - Configuração ótima da RNA wavelet automática associada à série temporal *Canadian lynx*.

Neurônios	Normalização	Tamanho da janela	Função de ativação (camada escondida)	Função de ativação (saída)
8	premnmx	5	Sigmoide logística	linear

Fonte: O autor (2015).

Finalmente, no passo 6, as previsões h produzidas nos passos 3 e 5 foram simplesmente somadas, gerando as previsões híbridas da série temporal *Canadian lynx*, denotadas por $(t=36, \dots, 288)$, onde $t' = 35$. O gráfico com os valores observados, assim como as suas respectivas previsões, podem ser contemplados na Figura 16.

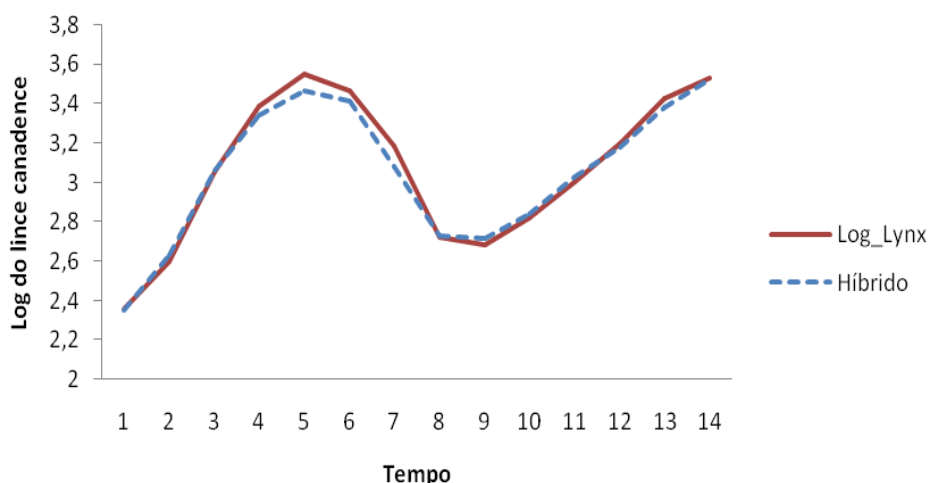


Figura 16 - Previsões e realizações do logaritmo na base de Euler da série temporal Canadian lynx.

Como salientado previamente, os resultados obtidos por meio da metodologia híbrida proposto na predição da série temporal *Canadian Lynx* foram comparados os alcançados por outros trabalhos; como: Zhang (2003)⁴, que comparou os resultados obtidos por três previsores (a saber, ARIMA, RNA e híbrido ARIMA-RNA); Kajitani, Keith & Mcleod (2005)⁸², que utilizaram na projeção da *Canadian Lynx* um modelo RNA *Feed-Forward* (FANN) e modelo SETAR (*Self-Exciting Threshold Autoregression*); Aladag et al (2009)⁸³, que combinaram RNAs recorrentes Elman e um modelo ARIMA; Khashei & Bijari, (2010)⁸⁴, que propuseram um modelo híbrido RNA-ARIMA; Ismail & Shabri (2014)⁸⁵, que utilizaram support vector regression (SVR) e least square SVR (LSSVR). Os resultados, considerando somente as estatísticas de aderência MSE e MAE, destas referências constam na Tabela 7. Na última linha da mesma, têm-se os resultados obtidos com o método híbrido proposto nesta tese de doutorado.

Tabela 7 - Comparativo MSE e MAE, fora da amostra, entre métodos de previsão aplicados à série temporal *Canadian Lynx*.

<i>Método</i>	<i>MSE</i>	<i>MAE</i>
ARIMA (Zhang, 2003)	0,0205	0,1123
RNA (Zhang, 2003)	0,0205	0,1121
Híbrido (Zhang, 2003)	0,0172	0,1040
SETAR (Kajitani, 2005)	0,0140	-
RNA (Khashei, 2010)	0,0136	0,0896
FANN (Kajitani, 2005)	0,0090	-
Híbrido (Aladag, 2009)	0,0090	-
SVR (Ismail, 2014)	0,0085	0,0746
LSSVR (Ismail, 2014)	0,0030	0,0418
Metodologia Proposta	0,0016	0,0211

Fonte: Ismail e Shabri, (2014) – adaptado.

Na tabela 7, verifica-se que o método proposto obteve redução nas estatísticas de aderência MSE e MAE, em relação a todos os outros métodos preditivos listados. Na Figura 17, pode-se verificar a alta correlação dos dados e as respectivas previsões – o que significa bom nível de ajustamento do preditor híbrido proposto aos dados subjacentes.

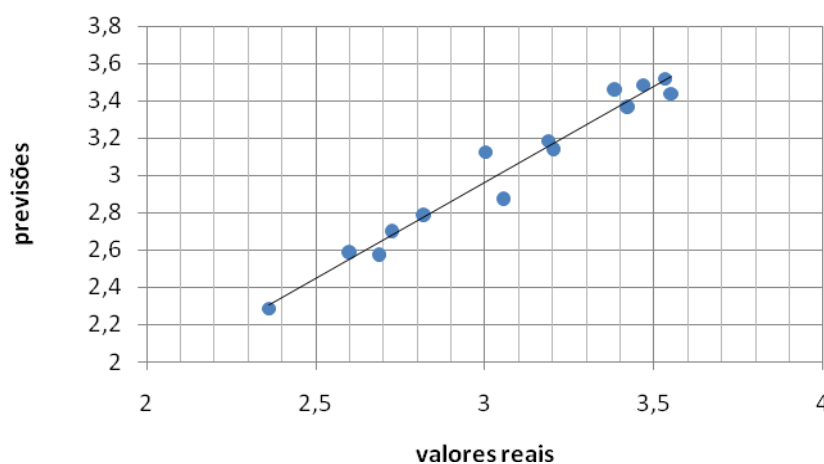


Figura 17 - Gráfico de dispersão entre as previsões da metodologia proposta e das respectivas observações, +na amostra de teste.

Fonte: O autor (2015).

4.2.3

Modelagem da série temporal Exchange Rate

Uma taxa de câmbio consiste no preço de uma moeda em termos de outra moeda. Aqui, a taxa de câmbio considerada envolve as moedas libra esterlina e dólar americano, referida como *Exchange Rate*. No mercado financeiro internacional, fazer a previsão da taxa de câmbio é muito importante, porém de difícil execução. Neste sentido, vários modelos de previsão lineares e não lineares foram desenvolvidos para a realização de tal tarefa, com algumas aplicações apresentando resultados contraditórios, veja, por exemplo, Khashei & Bijari (2011)⁸⁶. A série temporal supracitada possui 731 pontos semanais, compreendendo o período de 1980 a 1993. Seguindo Zhang (2003)⁴, Khashei & Bijari (2010, 2011 e 2012)^{84, 86, 87}, dentre outros, foi utilizada a função logaritmo (na base decimal) para transformar os dados originais e, então, fez-se a sua projeção um passo a frente, para um horizonte de 52 semanas.

Quanto aos parâmetros ótimos da metodologia híbrida proposta, os mesmos se encontram abaixo, em relação a cada passo (até a geração das previsões pontuais):

- ✓ *Passo 1:* decomposição wavelet de nível 2, com base ortonormal wavelet db 40;
- ✓ *Passo 2:* para a predição da CW de aproximação de nível 2, utilizou-se o modelo ARIMA (4,1,24), com as estimativas (de mínimos quadrados [88]) significativas, ao nível 1%, dos parâmetros ϕ_4 , θ_8 , θ_{12} and θ_{24} ; por sua vez, para a projeção das CWs de detalhe de níveis 2 e 3, foram utilizados, respectivamente, os modelos ARIMA (10,1,12), com estimativas significativas para os parâmetros ϕ_2 , ϕ_4 , ϕ_6 , ϕ_8 , θ_2 e θ_{12} , e ARIMA (3,0,4), para os parâmetros ϕ_2 , ϕ_3 , θ_2 , θ_3 e θ_4 (em ambos os casos, há 5% de significância).
- ✓ *Passo 3:* os parâmetros adaptativos da CLW ótima são dados por:
 $\hat{\rho}_{A_2,BJ,t} = 0,984136$, $\hat{\rho}_{D_2,BJ,t} = 0,740954743$, $\hat{\rho}_{D_2,BJ,t-1} = 1,017796$, $\hat{\rho}_{D_3,BJ,t} = 1,0177965$, $\hat{\alpha} = -0,128429$ e $\phi_3 = 1,99367$, $\theta_1 = -0,384025$. Como variável de saída da CLW, se utilizou a

segunda diferença da série temporal subjacente. Quanto à estatística R^2 , esta tomou valor próximo a 81,4651%.

- ✓ *Passos 4e 5:* Quanto aos parâmetros da RNA wavelet automática (para projetar os resíduos produzidos pela CLW ótima, no passo 3), têm-se os seguintes parâmetros ótimos: decomposição de nível 2, com base ortonormal wavelet db10; janela igual a 5; funções de ativação tangente hiperbólica e linear nas camadas oculta (com 8 neurônios) e de saída, respectivamente; e transformação Z-score.
- ✓ *Passo 6:* são somadas as previsões dos passos 3 e 5, gerando as previsões híbridas para a série temporal log Exchange Rate.

Na Tabela 8, encontra-se um comparativo envolvendo os resultados obtidos por Zhang (2003)⁴ e Khashei & Bijari (2010, 2011, 2012)^{84, 86, 87}, que são os apresentaram os melhores resultados estatísticos. Novamente, o método híbrido proposto se apresenta com os menores erros de previsão, em termos das estatísticas de aderência MSE e MAE.

Tabela 8 - Comparativo entre as estatísticas MSE e MAE entre métodos de previsão aplicados à série temporal log *Exchange Rate*.

<i>Método</i>	<i>MSE</i>	<i>MAE</i>
ARIMA (Zhang, 2003)	4,52977	0,005360
RNA (Zhang, 2003)	4,52657	0,005251
Híbrido (Zhang, 2003)	4,35907	0,005121
RNA (Khashei, 2010)	3,76399	0,005107
RNA (Khashei, 2011)	3,64724	0,004969
ARIMA/PNN (Khashei, 2012)	3,35623	0,004717
RNA/PNN (Khashei, 2012)	2,77122	0,004271
Método Proposto	1,01214	0,002269

Fonte: O autor (2015)

Nota: multiplicar todos os valores de MSE por 10^{-5} .

4.3

Previsões pontuais e intervalares para os dados de Itaipu

Nesta seção, são exibidos os resultados empíricos da aplicação do método híbrido proposto em toda a sua extensão, incluindo a estimativa dos intervalos de previsão. Para tanto, foi escolhida uma série de vazões médias mensais, formada a partir de informações coletadas no posto 266-Itaipu, na usina de Itaipu, em Foz do Iguaçu, estado do Paraná. Os resultados alcançados pelo método proposto foram comparados aos resultados obtidos por previsores tradicionais.

4.3.1

O problema da previsão de vazões

Existem vários fatores que podem influenciar a vazão em um determinado ponto de um curso de água, entre eles: evaporação, precipitação, infiltração e percolação. No entanto, a precipitação talvez seja a principal componente na formação da vazão e responsável pela característica aleatória do processo hidrológico. Para o estudo de vazões em usina hidrelétricas, podem-se destacar as seguintes categorias de vazões: natural, incremental, turbinada e vertida. A vazão natural é a observada em um posto de medição, considerando o rio na sua condição natural. A vazão incremental, por sua vez, consiste de toda vazão lateral captada entre um posto de medição e todos os postos de medição imediatamente a montante deste. Por sua vez, a vazão turbinada, é aquela que passa pelas turbinas e gera energia; enquanto que a vazão vertida é a que passa pelos vertedouros da usina, mas não gera energia.

A fim de se modelar a realidade hídrica, são formadas séries temporais de vazões médias diárias, semanais e mensais, o Operador Nacional do Sistema Elétrico (ONS) disponibiliza um banco de dados contendo o registro histórico de vazões naturais de todas as usinas hidrelétricas que compõem o Sistema Interligado Nacional (SIN). O primeiro registro de vazão natural média mensal é referente ao mês de janeiro de 1931, sendo o acesso aos registros associados

ao código denominado posto. No caso da Itaipu, o posto artificial tem o número 66 e o natural, 266.

Estimativas da Empresa de Pesquisa Energética (EPE) apontam que o consumo de energia elétrica no Brasil em 2020 será 650.092 GWh, sendo predominante a produção hidráulica, em virtude da situação privilegiada do país em relação aos recursos hídricos. A produção de energia a partir das usinas hidrelétricas tem ligação direta com a quantidade de água (vazão afluyente) disponível para a movimentação das turbinas geradoras. De forma que longos períodos de seca podem levar a uma diminuição da capacidade geradora de energia da usina. A relação entre vazões afluentes e geração de energia implica um cuidado especial com os dados que cercam o problema.

No Brasil, o NOS abdica das informações sobre as vazões na otimização da geração de energia e na elaboração de estratégias que minimizem efeitos negativos de eventuais enchentes ou secas. Para tanto, o ONS faz previsão de vazões naturais para locais de aproveitamento hidrelétrico que constituem o Sistema Interligado Nacional (SIN), sendo o PREVIVAZ o modelo de previsão mais utilizado.

O PREVIVAZ, usado na previsão de vazões e desenvolvido pelo Centro de Pesquisas de Energia Elétrica (CEPEL), é um modelo estocástico univariado [89], e este modelo é usado para prever as vazões semanais da maioria das bacias do SIN. O PREVIVAZ usa modelos lineares do tipo PARMA (p,q), além de transformações de séries históricas, tipo Box-Cox (Box e Cox, 1964) ou Logarítmicas. Um relatório emitido pelo ONS para previsões de vazões, em dezembro de 2011, menciona outros modelos além do PREVIVAZ. Seguem alguns exemplos de outros modelos e as bacias em que foram aplicados:

- ✓ Para a bacia do rio Uruguai, foi utilizado o modelo MPCV – Modelo de Previsão de Classes de Vazões;
- ✓ Para a bacia do rio Iguaçu, foi utilizado o modelo de Redes Neurais Fuzzy Recorrentes;
- ✓ Para o trecho da bacia do rio Paraná entre as UHEs Rosana, Porto Primavera e Itaipu, foi utilizado o modelo SMAP-MEL.

Entre os métodos usados pelo ONS, já citados, ressalta-se a frequência de uso dos métodos de Box & Jenkins (Box, 1994)³⁰, devido às suas estruturas matemáticas relativamente simples. Os modelos Box-Jenkins para previsão de séries temporais pertencem à família dos autorregressivos integrados de médias móveis (ARIMA).

Na Itaipu a previsão das vazões diárias é realizada em duas fases: uma determinística e outra estocástica. Na fase determinística, usa-se o método de Muskingum para a propagação das vazões. Na fase estocástica, utiliza-se a metodologia de Box e Jenkins. A série temporal supracitada é analisada a fim de se determinar os parâmetros do modelo ARIMA que será utilizado na previsão de vazões médias diárias dez dias à frente. A calibração dos parâmetros do modelo ARIMA escolhido foi realizada no início das operações da Itaipu na década de 80. Entre os principais formadores do lago de Itaipu, estão os rios Ivai e Piquiri, localizados na margem esquerda do rio Paraná. Para estes rios, o departamento especializado da Itaipu faz estimativas da vazão mensal.

Segundo Wang (2006a)⁹⁰, os processos de geração de vazões são sazonais e não lineares. Os mecanismos não lineares são mais fortes em escalas menores de tempo, sendo as vazões diárias e mensais não lineares (Wang, 2006b)⁹¹. São vários os trabalhos existentes na literatura que discutem a previsão de séries oriundas de processos estocásticos não lineares, sendo predominante para este caso o uso de Redes Neurais Artificiais (RNA). (Ballini Wang, 2006a⁹⁰; Valenzuela, 2008⁹²; Ferreira, 2012⁹³; Abreu, 2012⁹⁴; Teixeira Jr, 2012a)⁹⁵, entre outros, utilizaram Redes Neurais Artificiais (RNA) na previsão de séries temporais de vazão.

Embora as RNAs possam ser aplicadas em problemas lineares e não lineares e como ferramenta única em previsões de séries temporais, uma breve inspeção na literatura especializada detecta o uso rotineiro de combinações de previsores e de previsões. Na associação de previsores, pode-se destacar a combinação dos previsores ARIMA e RNA, ARIMA e Support Vector Machine (SVM), e ARIMA e GARCH, nestes casos os autores partem do pressuposto que as séries são geradas por processos lineares e não lineares, simultaneamente. Veja, por exemplo, Zhang (2003)⁴; Hwarng & Ang (2001)⁹⁶;

Valenzuela, (2008)⁹²; Abreu (2012)⁹⁴; Sujjavirivasup (2013)⁹⁷; Yaziz (2013)⁹⁸. Outros usam fazer um pré-processamento da série a partir de *Wavelet*, Filtros de Kalman, Filtros de Partículas e *Singular Spectrum Analysis* (SSA), para então alimentar o previsor, como é o caso de Lima (2011)¹⁸; Menezes (2012)⁹⁹; Hirpa (2013)¹⁰⁰ e Souza (2013)¹.

A previsão de aflúências nos reservatórios das usinas hidrelétricas é de grande importância no planejamento da produção de energia, propicia ao gestor do sistema elementos para a otimização do total da energia produzida, explorando toda a sua potencialidade. Também faz parte das ações da Itaipu o monitoramento do nível do rio Paraná à jusante da barragem, por conta de tratados firmados com Argentina e a preocupação com o fenômeno enchente sobre as populações ribeirinhas. Desta forma, as previsões de aflúências realizadas com certa antecedência e precisão permitem aos tomadores de decisões elaborarem estratégias que minimizem os efeitos negativos de cheias e secas extraordinárias. As previsões de aflúências ainda fornecem informações importantes para outros setores envolvidos com o reservatório, tais como: abastecimento urbano e agrícola, piscicultura, aquicultura, recreação e lazer.

4.3.2

Caracterização do conjunto de dados

Neste experimento, foi utilizada uma série temporal de vazão de afluentes mensal média da usina de Itaipu cujo período é o de janeiro de 1970 até dezembro de 2010 (totalizando 492 observações). Em http://www.ons.org.br/operacao/vazoes_naturais.aspx, tal conjunto pode ser acessado. Mais especificamente, os dados são provenientes das medições no posto 266, da Usina Hidrelétrica de Itaipu, localizado no Rio Paraná, no trecho de fronteira entre o Brasil e o Paraguai, a 14 km ao norte da Ponte da Amizade. A área da usina se estende, desde Foz do Iguaçu, no Brasil, e Ciudad del Este, no Paraguai, ao sul, até Guaíra (Brasil) e Salto del Guairá (Paraguai), ao norte. A escolha dessa série se deu por conta de sua relevância e por ser de difícil modelagem. Na Figura (18), tem-se o gráfico da série temporal supracitada, com todas as 492 observações.

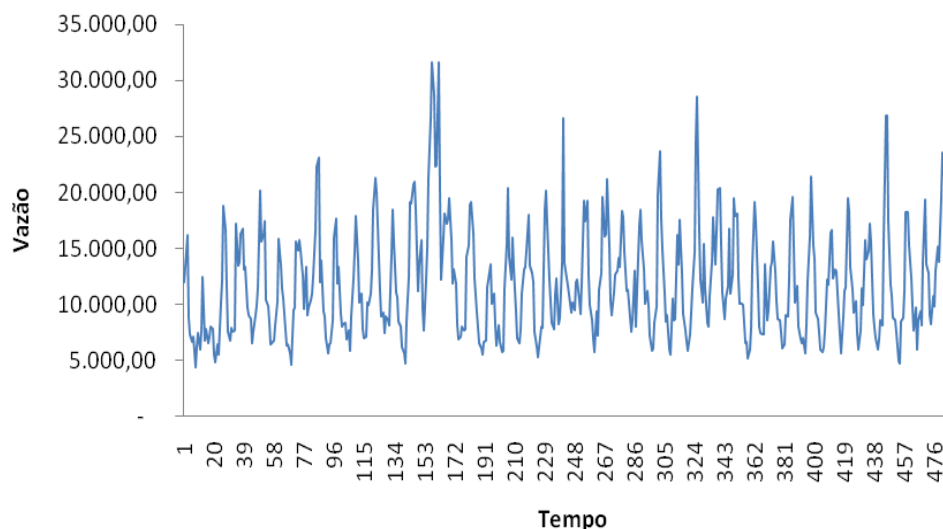


Figura 18 - Série de vazões mensais obtidas no posto 266-Itaipu no período 1970-2010.

Fonte: O autor (2015).

O correlograma (Figura 19) da série temporal observada mostra a existência de componente de sazonalidade com período igual a doze; além de autocorrelações não decrescentes nos lags múltiplos de doze, indicando não-estacionariedade de segunda ordem. A indicação de não-estacionariedade foi confirmada pelo teste de Dickey Fuller aumentado, para um nível de significância de 5%.

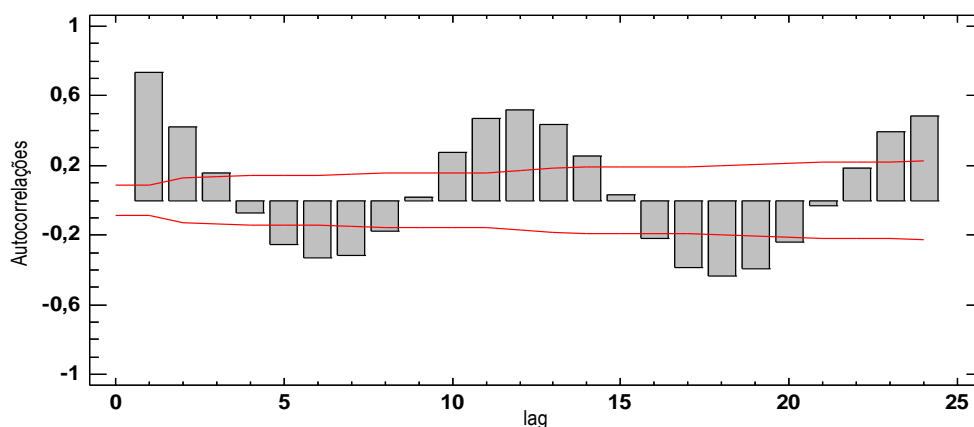


Figura 19 - Correlograma da série de vazões mensais.

Fonte: O autor (2015).

Os resultados obtidos com a modelagem da série de vazões pelo método proposto foram comparados com outras abordagens consagradas na literatura, tais como os previsores individuais de Box & Jenkins e de redes neurais artificiais; além dos compostos, ARIMA-wavelet e RNA-wavelet. Para fins de experimentação, 70 % dos primeiros dados foram utilizados no treinamento dos modelos preditivos individuais (de Box & Jenkins e de redes neurais artificiais), separando 20% para a validação e 10% dos remanescentes (49 pontos), para a amostra de teste. Para tal, foram feitas previsões de um passo à frente, em um horizonte de previsão de 49 meses. Para a avaliação de acurácia dos métodos preditivos, foram utilizados as estatísticas de aderência MAPE e MAE.

4.3.3

Métodos preditivos individuais

Para testar a hipótese de que todos os coeficientes de autocorrelação ρ_k dos erros de previsão eram estatisticamente não diferentes de zero, foi utilizada

a estatística $Q^* = n \sum_{k=1}^m \hat{\rho}_k^2$, desenvolvida por Box e Pierce, em que n é o tamanho da amostra e m a duração da defasagem. A estatística Q^* tem distribuição qui-quadrado com m graus de liberdade. Quando Q^* excede o valor crítico, rejeita-se a hipótese nula de que todos os ρ_k são iguais a zero.

A fim de eliminar o efeito de não-estacionariedade na componente de sazonalidade, foi aplicada uma diferenciação de ordem $D = 12$. Com a série diferenciada, fez-se a análise dos gráficos da FAC e FACP, podendo assim definir possíveis modelos para o ajuste dos dados. Testes com os resíduos indicaram que, entre os modelos candidatos, o ARIMA(1,0,2)x(0,1,2)₁₂ era aquele que melhor se ajustava aos dados da série de vazões em estudo. Para a validação do modelo ARIMA ajustado à série observada, foram analisados os resíduos com a construção de gráficos de autocorrelações, buscando mostrar que os mesmos são não autocorrelacionados.

A Figura 20 mostra o correlograma dos resíduos provenientes da modelagem. Considerando espúria a violação da autocorrelação no lag 15, considerou-se o referido modelo ajustado adequado.

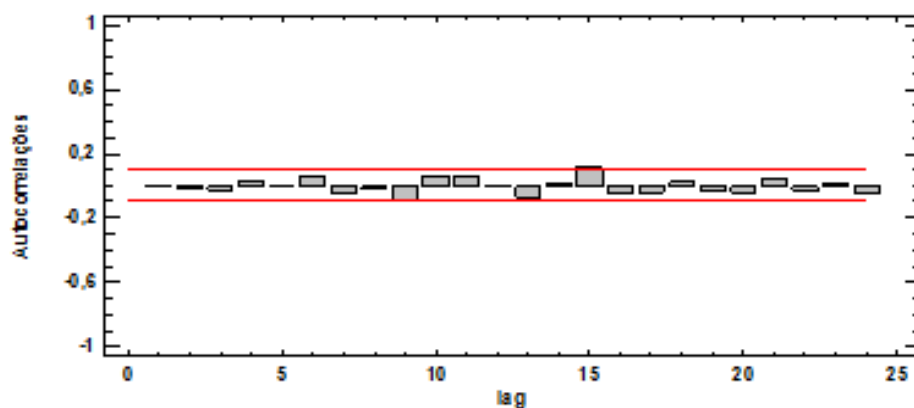


Figura 20 - Autocorrelações dos resíduos do modelo $ARIMA(1,0,2)X(0,1,2)_{12}$

Fonte: O autor (2015).

O teste de Ljung-Box corrobora a conclusão obtida na FAC residual. Como o p-valor (0,203) obtido com o teste de Box-Pierce é maior que 0,05, não se rejeitou a hipótese de os resíduos eram descorrelatados, ao nível de 5%. A RNA usada na modelagem da série subjacente foi *feed-forward multilayer perceptron* (RNA-MLP). Foram realizadas vinte combinações envolvendo tamanho da janela de entrada (J) e quantidade de neurônios na camada escondida (CE), fazendo-se dez simulações para cada combinação. A rede utilizada era constituída por uma camada escondida com função de ativação tangente hiperbólica sigmoideal (*tansig*) e um neurônio na camada de saída com função de linear (*purelin*) e *Levenberg e Marquardt* o algoritmo de treinamento. A busca pela RNA que melhor ajustasse os dados identificou a RNA-MLP com tamanho de janela de entrada igual a quatro (J=4) e camada escondida composta por cinco neurônios artificiais (CE=5), conforme a Tabela 9, onde estão listadas seis das vinte estruturas testadas. Nesta tabela podem ser observados o Erro Médio Percentual Absoluto (MAPE), Erro Percentual Absoluto (APE) mínimo e o APE máximo para as seis configurações listadas.

Tabela 9 - APE mínimo, máximo e MAPE para seis estruturas distintas de RNA-MLP.

J	CE	APE Mínimo			APE Máximo			MAPE		
		Treino	Validação	Teste	Treino	Validação	Teste	Treino	Validação	Teste
5	5	0,060	0,20	0,23	112,3	76,7	79,4	17,8	18,4	21,3
4	4	0,040	0,67	0,67	113,5	70,3	86,4	19,3	20,1	21,6
5	6	0,020	0,44	0,41	132,9	92,5	159,5	16,8	18,6	26,3
5	7	0,005	0,07	0,49	109,3	72,1	146,7	17,2	18,8	27,2
4	5	0,020	0,04	0,68	122,1	87,4	49,4	18,2	18,3	18,9
4	6	0,004	0,07	1,22	126,8	88,3	171,2	18,3	18,9	24,1

Fonte: O autor (2015).

A amostra de teste, formada pelas últimas 49 observações, apresentou MAPE igual a 18,01% com o modelo ARIMA e 18,9% com a modelagem por RNA. Nas Figuras 21a e 21b estão os gráficos de dispersão da amostra de teste versus previsão por RNA e ARIMA, respectivamente. Os gráficos em questão são um indicativo que se pode melhorar a qualidade das previsões obtidas com os métodos individuais ARIMA e RNA.

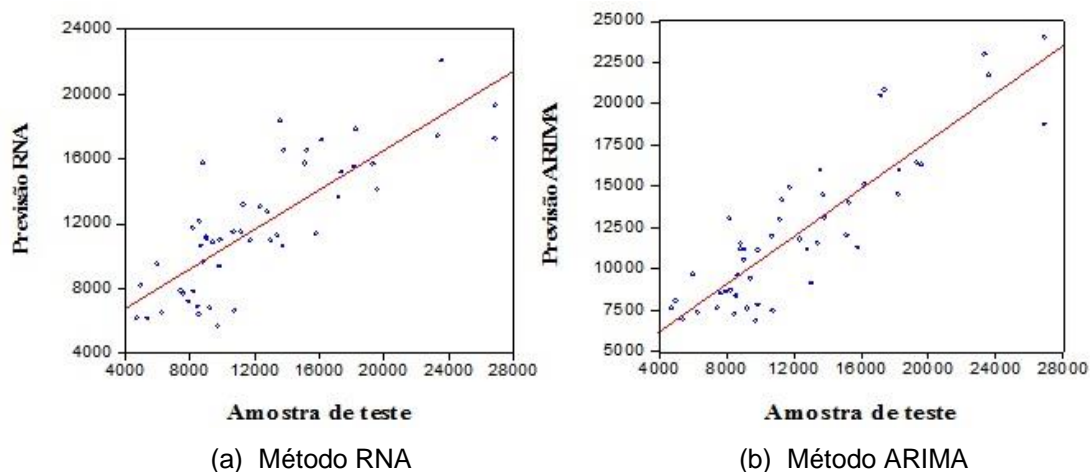


Figura 21 - Diagramas de dispersão entre a amostra de teste e previsões por RNA e ARIMA.

Fonte: O autor (2015).

4.3.4

Métodos preditivos individuais com wavelet

Inicialmente, a série de vazões y foi decomposta ortogonalmente via expansão wavelet de nível 2, utilizando a base Daubechies com momento 10 (“db 10”). As três CWs geradas (sendo duas de detalhes (D1 e D2) e uma de aproximação (A2)) foram ajustadas por três RNA e três modelos ARIMA diferentes e, em seguida, as previsões combinadas linearmente, a fim de se obter as previsões da série temporal original. Na Figura 49, podem ser observadas as etapas para a execução do método ARIMA-wavelet. O fluxograma do método RNA-wavelet pode ser obtido simplesmente pela substituição, na Figura 22, do ARIMA pela RNA.

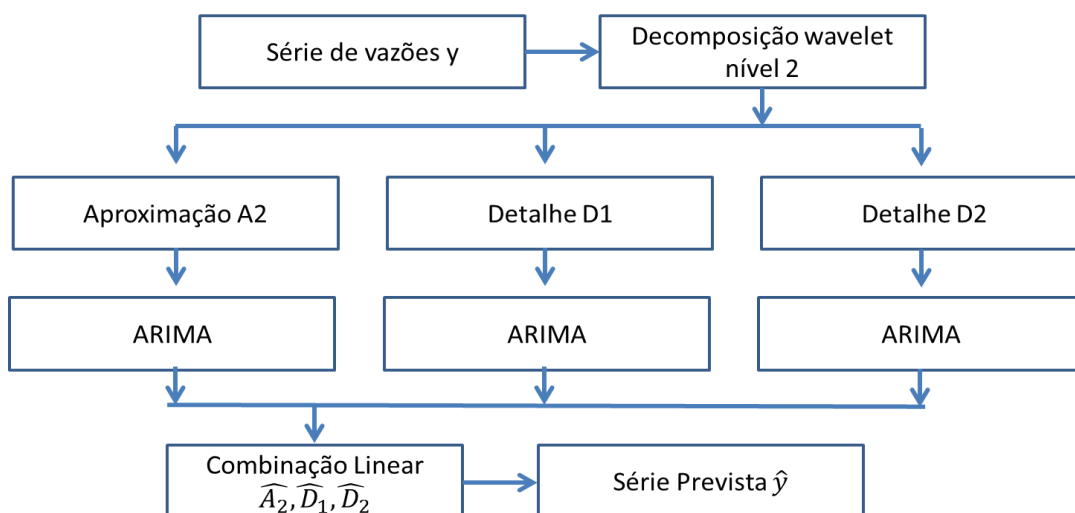


Figura 22 - Fluxograma com as etapas do método ARIMA-wavelet.

Fonte: O autor (2015).

Na Figura 23 podem-se observar os gráficos da série de vazões mensais e as componentes wavelets de aproximação (A2) e detalhes (D1 e D2) obtidas com a base ortonormal wavelet de Daubechies com momento 10 (db 10).

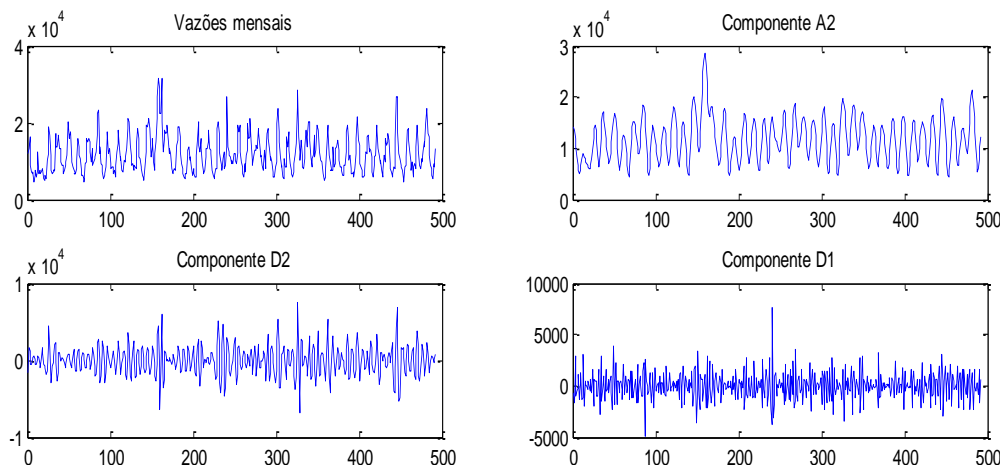


Figura 23 - Gráfico da série temporal de vazões e das CWs A2, D2 e D1.

Fonte: O autor (2015).

A Tabela 10 fornece um resumo dos modelos AR, MA e configurações das RNA utilizadas na modelagem das componentes wavelet A2, D1 e D2, sendo CE a quantidade de neurônios da camada escondida com função de ativação tangente hiperbólica (*tansig*) e J o tamanho da janela de entrada de dados. A camada de saída possui um único neurônio e a função de ativação é linear (*purelin*).

Tabela 10 - Configurações RNA e ordens dos modelos AR e MA para A2, D1 e D2.

Componente de aproximação de nível 2 (A2):	
RNA1: <i>feed-forward</i> , J=20, CE=9, <i>tansig</i> (camada escondida), <i>purelin</i> (camada de saída).	
AR(1), AR(2), AR(4), AR(12), AR(13), AR(24), MA(1 até 3), MA(13) e MA(24)	
Componente de detalhe de nível 1 (D1):	
RNA2: <i>feed-forward</i> , J=20, CE=5, <i>tansig</i> (camada escondida), <i>purelin</i> (camada de saída)	
AR(1 até 8), MA(1 até 8)	
Componente de detalhe de nível 2 (D2):	
RNA3: <i>feed-forward</i> , J=21, CE=9, <i>tansig</i> (camada escondida), <i>purelin</i> (camada de saída)	
AR(1 até 8), AR(10), AR(11), AR(24), MA(1 até 4), MA(6), MA(7) e MA(10)	

Fonte: O autor (2015).

As previsões $\hat{A}_{2,RNA}$, $\hat{D}_{1,RNA}$ e $\hat{D}_{2,RNA}$ foram combinadas linearmente a fim de se obter a previsão da série de vazões, conforme (5.2).

$$\hat{y}_t = \hat{\alpha}_{A2,RNA} \hat{A}_{2,RNA,t} + \hat{\alpha}_{D1,RNA} \hat{D}_{1,RNA,t} + \hat{\alpha}_{D2,RNA} \hat{D}_{2,RNA,t} \quad (5.2)$$

Onde: $\hat{\alpha}_{A2,RNA}$, $\hat{\alpha}_{D1,RNA}$ e $\hat{\alpha}_{D2,RNA}$ são os parâmetros adaptativos ótimos determinados com a solução do problema de programação não linear (5.3), cuja função objetivo é a raiz do erro quadrático médio (RMSE).

Minimizar: RMSE

s. a.

$$\hat{y}_t = \alpha_{A2,RNA} \hat{A}_{2,RNA,t} + \alpha_{D1,RNA} \hat{D}_{1,RNA,t} + \alpha_{D2,RNA} \hat{D}_{2,RNA,t}$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{y}_t - y_t)^2}$$

Onde: $\alpha_{A2,RNA}$, $\alpha_{D1,RNA}$ e $\alpha_{D2,RNA}$ são variáveis de decisão irrestritas.

De forma análoga, obtém-se \hat{y}_t a partir das previsões $\hat{A}_{2,ARIMA,t}$, $\hat{D}_{1,ARIMA,t}$ e $\hat{D}_{2,ARIMA,t}$.

O MAPE para a amostra de teste obtido com a aplicação do método ARIMA-wavelet foi de 3,63%; e de 2,80%, com a utilização do RNA-wavelet – o que mostra a eficiência de tais abordagens em métodos em relação aos individuais RNA e ARIMA.

4.3.5

Método híbrido proposto: resultados empíricos

No *passo 1*, foi realizada uma decomposição wavelet de nível 2, com base ortonormal de Haar, de sua amostra de treino, de modo a gerar: uma CW de aproximação (A2); e duas de detalhe (D2 e D1). No *passo 2*, por sua vez, utilizaram-se os seguintes previsores:

- (a) Um modelo SARIMA (4,1,5) x (0,0,2)₁₂, para prever A2. Assuma que

$$z'_t = D_{2,t} - D_{2,t-1}, \text{ para todo instante } t. \text{ O referido modelo \u00f3timo \u00e9 dado}$$

por:

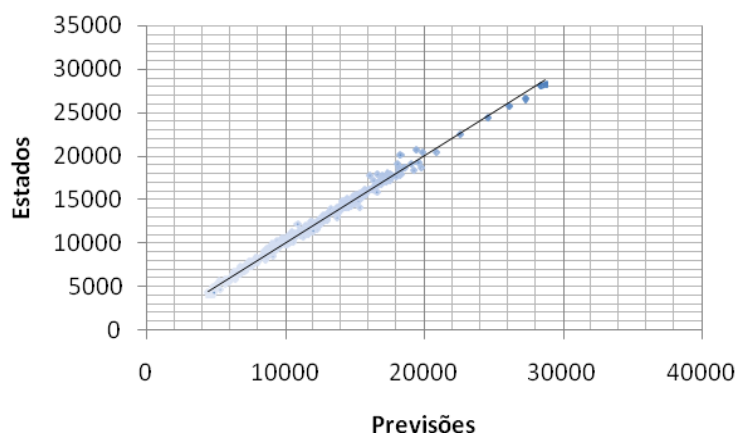
$$(b) \quad z'_t = \phi_1 z'_{t-1} + \phi_2 z'_{t-2} + \phi_3 z'_{t-3} + \phi_4 z'_{t-4} + \Phi_{12} z'_{t-12} + \\ \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_3 a_{t-3} + \theta_{12} a_{t-12} + \theta_{24} a_{t-24} + a_t.$$

- (c) Um modelo SARIMA (2,0,1) x (1,0,2)₁₂ para projetar D2, cuja representa\u00e7\u00e3o dada por:

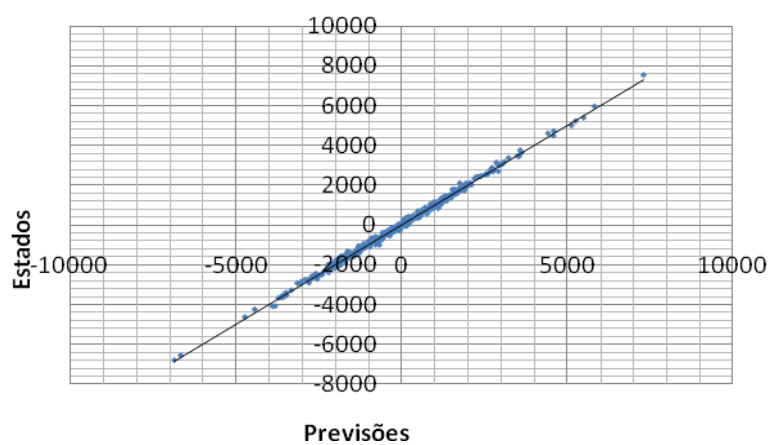
$$D_{2,t} = \phi_2 D_{2,t-2} + \theta_1 a_{t-1} + \Phi_{12} D_{2,t-12} + \theta_{24} a_{t-24} + a_t.$$

- (d) Um modelo ARIMA (1,0,2), para prever D1, o qual \u00e9 dado por:

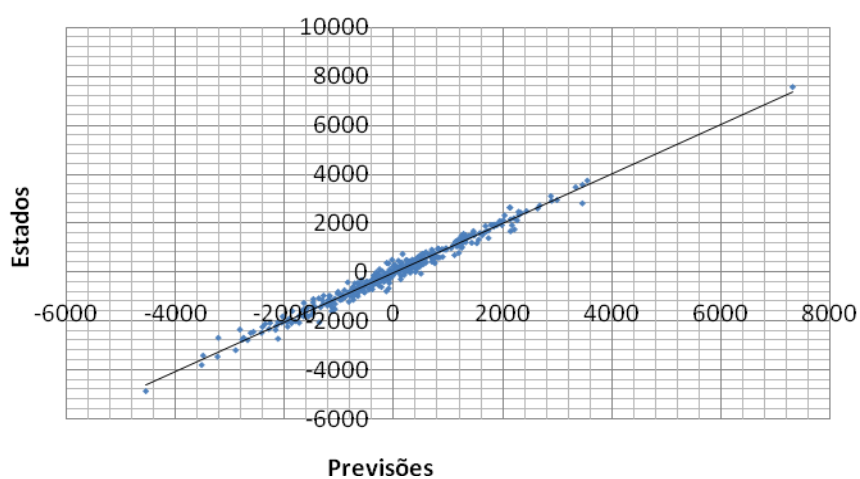
$$D_{1,t} = \phi_1 D_{1,t-1} + \theta_2 a_{t-2} + a_t.$$



(a) CW A2 versus previsões



(b) CW D2 versus previsões



(c) CW D1 versus previsões

Figura 24 - Gráfico de dispersão entre as CWs A2, D2 e D1 e as suas respectivas previsões.

Fonte: O autor (2015).

Na Figura 24, pode-se verificar que os dados são altamente correlacionados com as previsões, o que evidencia nível de acurácia satisfatório. A CLW ótima, no *passo 3*, é dada na Equação (46).

$$\begin{aligned} \hat{z}_t = & (\hat{y}_{A_2,BJ,t} - \hat{y}_{A_2,BJ,t-1}) \times 0,590068 + \hat{y}_{D_2,BJ,t} \times 0,371291 - \\ & \hat{y}_{D_2,BJ,t-1} \times 0,433461 - \hat{y}_{D_2,BJ,t-2} \times 0,766333 - \hat{y}_{D_2,BJ,t-5} \times 0,879413 + \\ & \hat{y}_{D_3,BJ,t} \times 0,680354 - \hat{y}_{D_3,BJ,t-1} \times 0,333435 - \hat{z}_{t-1} \times 0,305078 - \hat{z}_{t-2} \times \\ & 0,489447 - \hat{z}_{t-3} \times 0,383218 - \hat{z}_{t-12} \times 0,927083 \end{aligned}$$

Onde: $\hat{z}_t = \hat{L}_t - \hat{L}_{t-1}$, para todo instante t ; $\hat{z}_{t-12} \times 0,927083$ é uma componente sazonal da estrutura Box e Jenkins.

Por sua vez, no *passo 4*, os erros de predição da CLW ótima no passo 3, representados algebricamente por $e_t(t = 25, \dots, 394)$, sendo que $t' = 24$, foram decompostos via decomposição wavelet de nível 2, utilizando a base ortonormal wavelet de Daubechies com momento nulo igual a 10 cuja notação é dada por “db10”[15]. Por sua vez, a configuração ótima da RNA wavelet automática para projetar os resíduos produzidos no passo 6 pode ser visualizada na Tabela 11. Na Figura 25, têm-se os gráficos das previsões e dos resíduos, na amostra de treinamento.

Tabela 11 - Configuração ótima da RNA wavelet automática para projetar os resíduos da CLW ótima, do passo 3.

Neurônios	Normalização	Tamanho da janela	Função de ativação (camada escondida)	Função de ativação (saída)
db 40	Premnmx	4	Sigmoide logística	linear

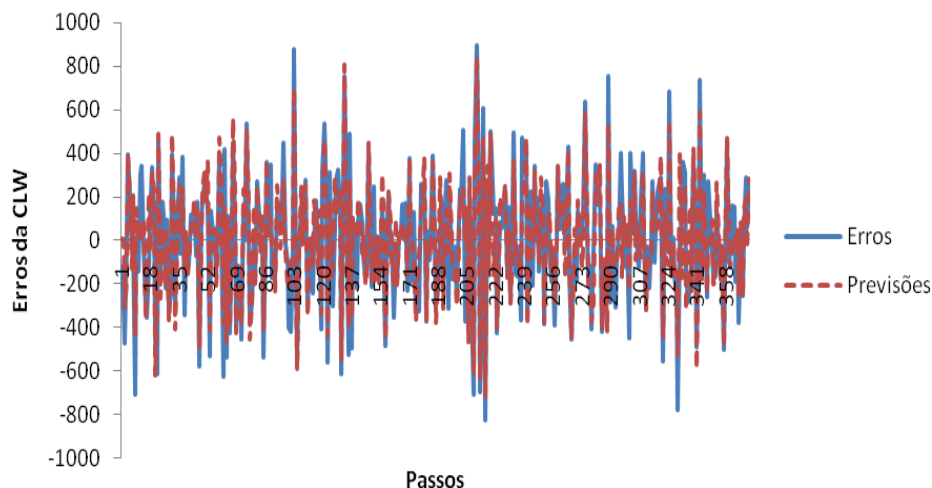


Figura 25 - Resíduos da CLW do passo 3 e suas respectivas previsões.

Fonte: O autor (2015).

Como já salientado, reservou-se 10% do total de dados observados para a amostra de teste (as últimas 49 observações), utilizando-a na avaliação do poder preditivo da metodologia proposta. Todas as previsões são um passo à frente. Com efeito, no passo 7, somando-se as previsões fora da amostra (amostra de teste), dos passos 3 e 6, são geradas as híbridas para a série temporal de vazão. Note que, na escala do gráfico, praticamente não é possível perceber, de forma visual, diferença entre as previsões e os estados da série temporal de vazão, na amostra de teste.

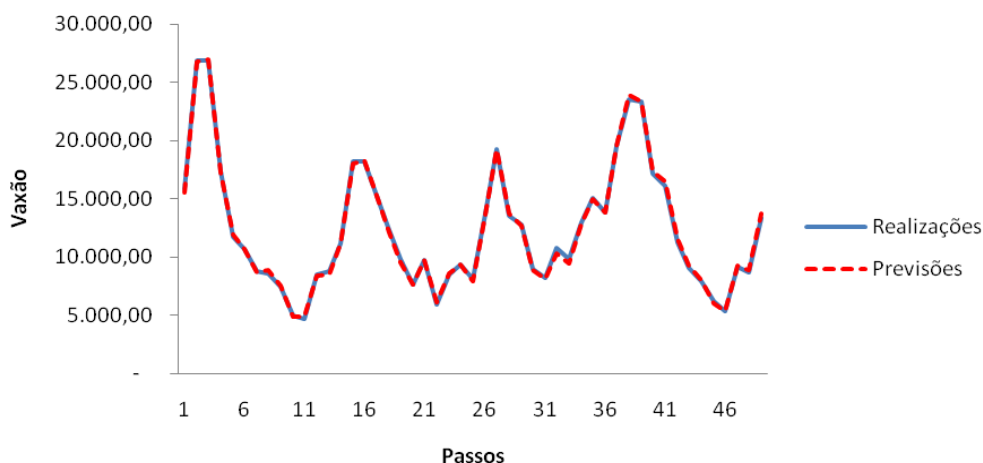


Figura 26 - Realizações e previsões da série temporal de vazão, na amostra de teste (out of sample).

Fonte: O autor (2015).

Na Figura 26, acima, podem-se visualizar os gráficos das realizações da vazão e suas respectivas previsões. Na Tabela 12, abaixo, são expostos os valores das estatísticas residuais associados a todos os previsores aqui implementados, quanto às estatísticas MAPE e MAE, nas amostras de treino, validação e de teste. Por CL1 e CL2, entenda como sendo, respectivamente, as combinações lineares das abordagens ARIMA e RNA; e ARIMA-wavelet e RNA-wavelet.

Tabela 12 - Estatísticas de aderência MAPE e MAE.

Métodos	MAPE			MAE		
	<i>Treinamento</i>	<i>Validação</i>	<i>Teste</i>	<i>Treinamento</i>	<i>Validação</i>	<i>Teste</i>
RNA	16,35%	15,19%	21,78%	1.886,87	1.682,38	2.389,97
ARIMA	16,60%	19,46%	21,25%	2.032,22	2.185,50	2.560,19
CL 1	14,95%	14,82%	19,13%	1.877,10	1.752,25	2.274,33
RNA-wavelet	2,11%	2,47%	2,80%	224,00	232,73	290,79
ARIMA-wavelet	2,74%	3,13%	3,63%	291,92	278,34	396,26
CL 2	2,06%	2,38%	2,64%	218,29	222,45	275,75
Método Proposto	1,27%	1,27%	1,36%	131,83	121,71	141,95

Fonte: O autor (2015).

Na tabela 12, tem-se que CL1 foi superior aos métodos preditivos individuais RNA e BJ, onde os pesos adaptativos lineares ótimos associados às previsões RNA e BJ foram, respectivamente, iguais a 0,618388907 e 0,323249491. O mesmo pode ser verificado para o método de combinação CL2, quando comparado com os métodos base RNA-WAVELET e BJ-WAVELET. Os pesos adaptativos lineares ótimos associados às previsões oriundas de RNA-WAVELET e BJ-WAVELET, respectivamente, iguais a 0,910473771 e 0,089013323. Para tanto, foi utilizado o solver do software EXCEL 2007. O algoritmo de otimização utilizado, em ambos os casos, foi o evolucionário e as configurações padrão foram mantidas inalteradas. É possível verificar que a decomposição wavelet de nível 2, com a base ortonormal de Haar, acarretou ganhos preditivos relevantes. Finalmente, tem-se que a combinação linear híbrida wavelet em estágios múltiplos foi superior todos os outros nas três amostras.

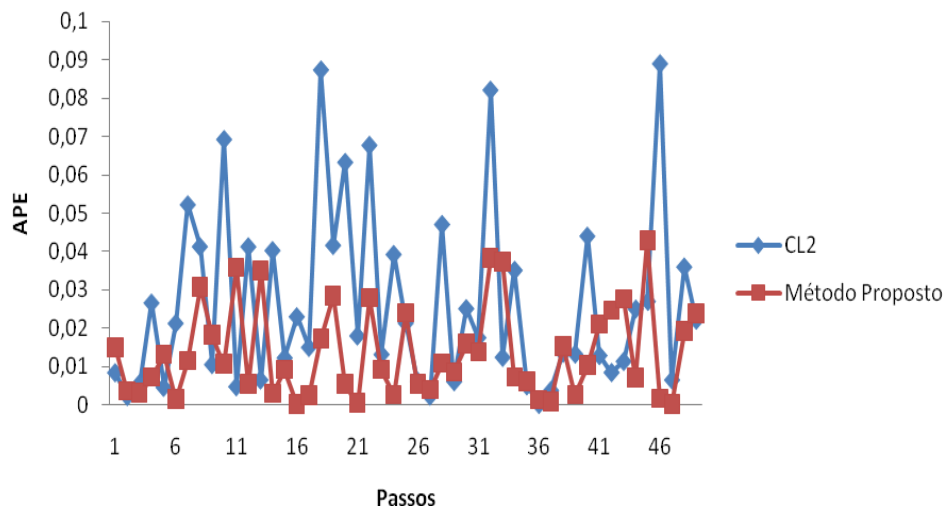


Figura 27 - Valores de APE dos previsores CL2 e do método proposto, na amostra de teste.

Fonte: O autor (2015).

Na Figura 27, tem-se a comparação entre o método proposto e a abordagem CL2 (que foi a melhor na Tabela 12 entre os métodos benchmark). Note que, em todos os 49 instantes, na amostra de teste, o método proposto apresenta evolução temporal dos valores de APE (*absolute percentage error*) mais estáveis que os da CL2 (por exemplo, em 6 instantes os valores de APE da CL2 violam a faixa dos 6%, enquanto que em nenhum instante isso ocorre no método proposto). Além disso, na maioria dos instantes, a combinação proposta apresenta valores de APE inferiores aos da CL2.

Abaixo, na Figura 28, é apresentado o diagrama de dispersão entre as previsões e as realizações da série temporal de vazões, na amostra de teste. Note a presença de forte correlação entre as realizações e as respectivas previsões, corroborando que, de fato, as previsões são acuradas.

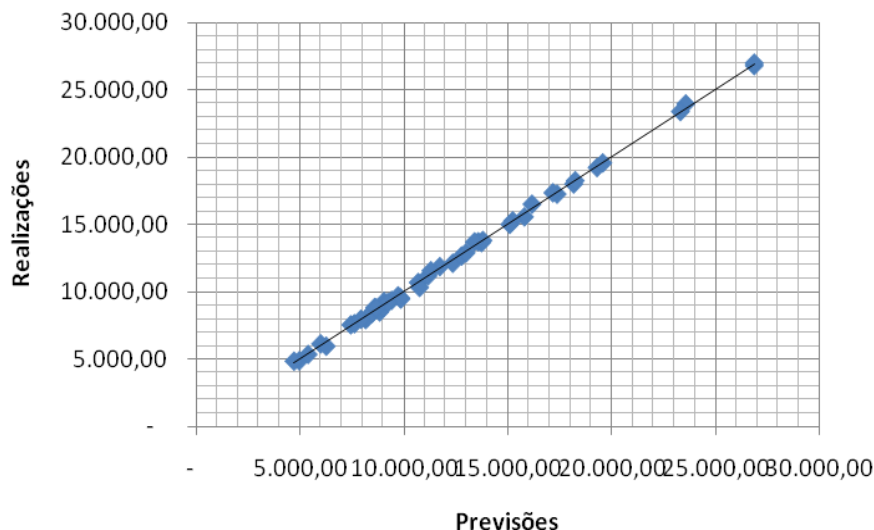


Figura 28 - Diagrama de dispersão entre a amostra de teste e previsões pelo método proposto.

Fonte: O autor (2015).

O passo 7 consiste na geração dos intervalos de previsão via Bootstrap. Neste caso, o processo *bootstrap* foi utilizado segundo a versão reamostragem residual. A quantidade de reamostragens obtidas foi igual a 20 ($B = 20$), que é número mínimo sugerido por Tibshirani (1996) [73]. As séries temporais *bootstrapadas* foram modeladas segundo o método proposto. É importante salientar que, para cada configuração da RNA automática, realizadas 25 simulações, escolhendo-se a que apresentava o menor erro quadrático médio (MSE), na amostra de validação.

Na Figura 29, têm-se as realizações e os limites inferiores e superiores do intervalo de previsão, ao nível de 95 de credibilidade, na amostra de teste, para uma série temporal de vazão mensal da usina de Itaipu. Visualmente, pode-se verificar que não há violações em qualquer instante.

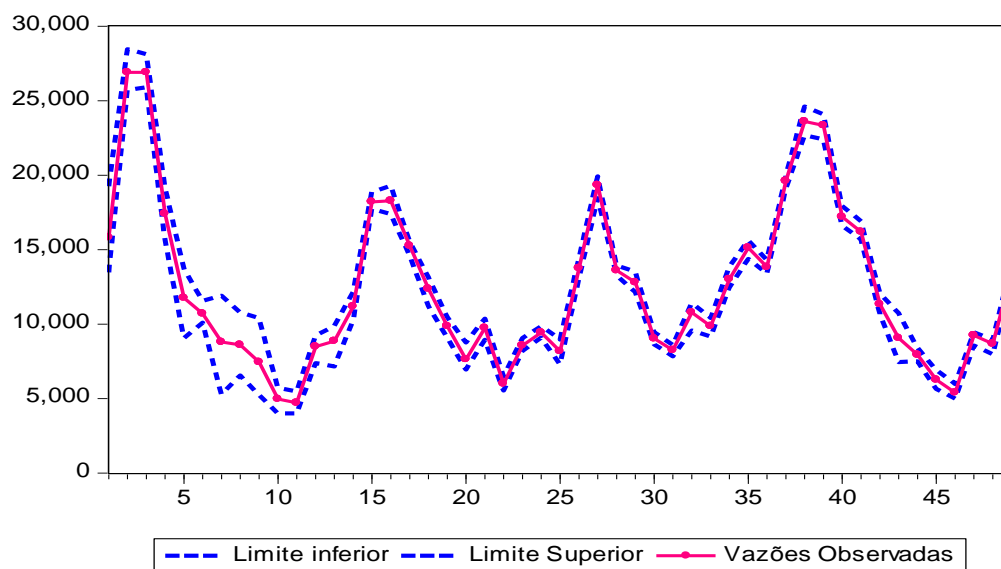


Figura 29 - Intervalo de previsão para a série de vazões – amostra de teste.

Fonte: O autor (2015).

CONCLUSÕES

Nesta tese de doutorado, foi proposta uma metodologia híbrida para projetar pontualmente e de forma intervalar séries temporais, que utiliza, de forma integrada, alguns dos principais métodos e abordagens presentes na Literatura: *decomposição wavelet*; *modelos de Box e Jenkins*; *redes neurais artificiais*; *programação matemática*; e *bootstrap*. Para mostrar a eficiência da metodologia proposta, bem como detalhá-la em um caso aplicado, foi realizada a modelagem da série temporal mensal de vazão da usina hidrelétrica de Itaipu, e de outras três séries temporais vastamente utilizadas para validação de previsores. As projeções consideradas aqui foram de um passo à frente, dentro e fora da amostra. No caso da série temporal de vazão, como *benchmark*, a referida foi modelada, para a mesma finalidade, por meio de outros 6 métodos preditivos consagrados na literatura, sendo 4 individuais (RNAs e modelos ARIMA, com abordagens simples e integradas com uma decomposição wavelet de nível r) e 2 combinados (duas combinações lineares de previsões (CL1 e CL2), conforme a abordagem em Granger & Newbold (1974)¹⁰¹. Para a geração dos intervalos de previsão, utilizou-se o método de Bootstrap.

Em relação à metodologia proposta, é importante destacar alguns pontos. Em relação à abordagem wavelet, considerado um horizonte de previsão igual a k passos à frente, foram calculadas k diferentes séries de *Fourier wavelet*, de modo a se obter k diferentes decomposições wavelet (de nível 2). Para a escolha das melhores bases ortonormais wavelet, no passo 3, considerou-se aquelas que, dado o nível de decomposição igual a 2 (que foi o que apresentou os melhores resultados em termos de modelagem, durante a pesquisa), forneciam o menor EQM (*erro médio quadrático*) entre os sinais da amostra de treino e os respectivos sinais wavelet gerados em decorrência da soma das componentes de aproximação e de detalhe. Na RNA wavelet automática, considerou-se a que gerou as melhores previsões, na amostra de treinamento. Além disso, decomposição wavelet possibilitaram a geração de componentes espectrais wavelet com um padrão de comportamento temporal

menos ruidoso (ou menos estocástico) e mais regular. Em particular, no tocante à regularidade de comportamento, tal fato se ocorreu, porque cada componente ortogonal *wavelet* está associada ao *parâmetro de escalam* [Seção 2.1] das funções *wavelet* que a compõe e este, à frequência da respectiva componente *wavelet* (Box et al, 1994)²⁹. Assim, cada CW forneceu diferentes acerca da flutuação estocástica-temporal da série temporal original, no passo 3. Dessa maneira, ao se combinar linearmente a previsão para as componentes *wavelet*, obteve-se, para cada instante t , uma previsão combinada que, segundo Faria & Mubwandarikwa (2008)⁴⁵, pode ser interpretada como uma agregadora de informações oriundas de diferentes fontes (abordagens preditivas). Destaca-se ainda que foi utilizada a decomposição ortogonal *wavelet*, ao longo da pesquisa, mostraram que as componentes ortogonais *wavelet* possuem melhores padrões de comportamento, para fins de modelagem, que as séries temporais originais.

A combinação linear wavelet (CLW), por se tratar, na verdade, de um modelo linear estatístico, produziu resíduos que, de fato, eram descorrelatados. Em abordagens tradicionais de combinação de previsões, geralmente não é feito qualquer teste estatístico para garantir tal propriedade estatística. Por outro lado, a RNA wavelet automática propiciou fazer uma busca exaustiva por milhares de possíveis previsores para projetar os resíduos descorrelatados (mas não independentes) da CLW, do passo 3, fornecendo previsões acuradas que possibilitam agregar informações não lineares na previsão final (previsão híbrida). No tocante aos intervalos de previsão para a série temporal de vazão, verificou-se que a sua amplitude era relativamente pequena (o que evidencia menor incerteza acerca do futuro) e que não houve nenhuma violação nos mesmos.

Por último, no que tange à utilização dos modelos de Box & Jenkins e das RNAs, isso ocorreu pelo fato de serem reconhecidamente eficientes à modelagem de estruturas de autodependência linear ou não linear, respectivamente, além da facilidade de manipulação operacional devido à gama de software e pacotes existentes.

SUGESTÕES PARA TRABALHOS FUTUROS

Diante dos resultados obtidos nesta tese de doutorado, podem ser estudadas outras metodologias dela decorrentes ou similares, com a finalidade de métodos preditivos cada vez mais eficientes e mais abrangentes, na modelagem de séries temporais que apresentam estruturas temporais complexas de serem modeladas. Dentre tantas possibilidades, eis algumas:

- ✓ Construir um algoritmo computacional que escolha a base ortonormal wavelet e o nível de decomposição ótimos;
- ✓ Utilizar algoritmos genéticos e PSO para obtenção automática dos modelos ARIMA, em um algoritmo computacional;
- ✓ Gerar N densidades preditivas via amostrador QMC híbrido e combiná-las geometricamente;
- ✓ Decompor os sinais via teorema espectral (para operadores normais e compactos) definido em termos de uma base ortonormal wavelet.
- ✓ Utilizar modelos ARIMA-GARCH;
- ✓ Utilizar redes neuro-fuzzy com decomposição wavelet na predição dos resíduos;
- ✓ Utilizar variáveis exógenas (por exemplo, modelos ARIMAX-GARCH e/ou RNAXs).
- ✓ Interpretar os parâmetros adaptativos da CLWH variantes no tempo e estimados por abordagem bayesiana.
- ✓ CLWH multivariada.

REFERÊNCIAS BIBLIOGRÁFICAS

- 1 SOUZA, R.C.; NETO, A.C. **A bootstrap simulation study in ARMA(p,q) structures.** [S.l]: J. of Forecasting.v. 15, p. 343-353, 2004.
- 2 HAYKIN, S. **Redes Neurais: Princípios e Prática.** 2. ed. Porto Alegre: Bookman, 2001.
- 3 TEIXEIRA JR, L.A.; MENESES, M.L.; PESSANHA, J.F.M. SOUZA, R.C.; CASSIANO, K.M. Residential Electricity Consumption Forecasting Using a Geometric Combination Approach. **International Journal of Energy and Statistics**, v. 1, n. 2, 1-13, 2013.
- 4 ZHANG, G.P. **Time series forecasting using a hybrid ARIMA and neural network model.** Atlanta: Neurocomputing, v.50, p. 159-175, 2003.
- 5 DAGUM, E.B. **The X11ARIMA seasonal adjustment method Ottawa: Time Series Research and Analysis Division**, Statistics Canada, 1980.
- 6 CHATFIELD, C. **Time-series forecasting.** New York: Chapman & Hall/CRC, 2000.
- 7 NEUMAN. Maximum likelihood Bayesian averaging of uncertain model predictions. **Stochastic Environmental Research and Risk Assessment**, 17, p. 291-305, 2003.
- 8 TASKAYA-TEMIZEL T.;CASEY, M.C. A comparative study of autoregressive neural network hybrids. **Neural Networks**, [S.l.], 18, p. 781-789, 2005.
- 9 DELL'AQUILA; RONCHETTI. Stock and bond return predictability: the discrimination power of model selection criteria. **Computational Statistics & Data Analysis**, [S.l.], v. 50, n. 6, , p.1478–1495, 10 mar 2006
- 10 AMENDOLA; STORTI; A GMM procedure for combining volatility forecasts. **Computational Statistics & Data Analysis**, [S.l.], v. 52, n. 6, p. 3047-3060, 2008.
- 11 LUX, T.; MORALES-ARIAS, L. Forecasting volatility under fractality, regime-switching, long memory and student-t innovations. **Computational Statistics and Data Analysis**, [S.l.], v. 54, p. 2676-269, 2014.
- 12 DONOHO, D.L.; JONHSTONE, I.M. Ideal Spatial Adaptation by Wavelet Shrinkage. **Biometric**, [S.l.], v. 81, p. 425-455, 1994.

- 13 DONOHO, D.L.; JONHSTONE, I.M.; KERKYACHARIAN, G.; PICARD, D. Wavelet Shrinkage: Asymptopia? (with discussion) **Journal Royal Statistical Society**, [S.l.], v. 57, p 301-369, 1995.
- 14 MALLAT, S. A Wavelet Tour of Signal Processing. **The Sparse Way**, p. 27-39.
- 15 MALLAT, S. **A Wavelet Tour of Signal Processing**. San Diego: Academic Press, 1998.
- 16 LEI, C.; RAN, L. Short-term wind speed forecasting model for Wind Farm Based on Wavelet Decomposition DRPT. Nanjing. China, p. 2525-2529, 2008.
- 17 TIWARI, M.K.; CHATTERJEE, C. Development of an accurate and reliable hourly flood forecasting model using *wavelet-bootstrap*-ANN (WBANN) hybrid approach. **Journal of Hydrology**, [S.l.], p. 458-470, 2010.
- 18 LIMA, F.G. **Modelos de previsões de séries temporais financeiras com combinação de filtros de kalman e wavelets**. 2011, 151p. Tese (Livre Docência em Economia) - Universidade Estadual de São Paulo, Ribeirão Preto, 2011.
- 19 KISI, O.; CIMEN, M. A *wavelet*-support vector machine conjunction model for monthly streamflow forecasting. **Journal of hidrology**, [S.l.], p. 132-140, 2011.
- 20 NALLEY, D. *et al.* Using discrete *wavelet* transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954-2008). **Journal of hydrology**, [S.l.], p. 204-228, 2012.
- 21 KARTHIKEYAN, L; KUMAR, D.N. Predictability of nonstationary time series using wavelet and EMD based ARMA models. **Journal of hydrology**, [S.l.], p. 103-119, 2013.
- 22 KUBRUSLY, C.S. **Elements of Operator Theory**. Boston: Birkhäuser, 2001.
- 23 KUBRUSLY C.S.; LEVAN, N. Abstract Wavelet Generated by Hilbert Space Shift Operators. **Adv. Math. Sci. Appl.**, [S.l.], 14, 643-660, 2006.
- 24 TEIXEIRA JR, L.A.; MENESES, M.L.; PESSANHA, J.F.M.. SOUZA, R.C.; CASSIANO, K.M.. Residential Electricity Consumption Forecasting Using a Geometric Combination Approach. **International Journal of Energy and Statistics**, v. 1, n. 2, 1-13, 2013.
- 25 LEVAN, N.; KUBRUSLY, C.S. A Wavelet Time-Shift Detail Decomposition. **Mathematics and Computers in Simulation**, [S.l.], v. 63, p. 73-78, 2003.

- 26 MORETTIN, P.A. **Ondas e Ondaletas**: Da análise de Fourier à Análise. São Paulo: EDUSP, 1999.
- 27 DAUBECHIES, I. Orthonormal Bases of Compactly Supported. **Wavelet. Comm. Pure and Applied Math.**, 41, p. 909-996, 1988.
- 28 HILL, C.; GRIFFITHS, W.; JUDGE, G.; FARIAS, A. de. **Econometria**. São Paulo: Saraiva, 1999.
- 29 MORETTIN, P.A. e TOLOI, L.M.C. **Análise Séries Temporais**. 2. ed. São Paulo: ABE Projeto Fisher, Ed. Edgard Blucher, 2006.
- 30 BOX, G.; JENKINS, G.; REINSEL, G.C. **Time Series Analysis, Forecasting and Control**. 3.ed. Englewood Cliffs: Prentice-Hall, 598p, 1994.
- 31 BOX, G.; JENKINS, G. **Time Series Analysis: Forecasting and Control**. San Francisco: Holden-Day, 1970.
- 32 HAMILTON, J. **Time Series Analysis**, [S.l]: Princeton University Press, 1994.
- 33 ABELÉM, A. **Redes Neurais Artificiais na Previsão de Séries Temporais**, 1994. Dissertação (Mestrado). Departamento de Engenharia Elétrica Pontifícia Universidade Católica do Rio de Janeiro, 1994.
- 34 TAFNER, M.A. **Redes Neurais Artificiais: Introdução e Princípios de Neurocomputação**. Blumenau: EKO, 1996.
- 35 RUSSELL, S.J.; NORVIG, P. **Artificial Intelligence: a Modern Approach**. New Jersey: Prentice-Hall Inc, 1996.
- 36 HERRERA F.; LOZANO M.V.J.L. **Artificial Intelligence Review**. Tackling Real-coded Genetic Algorithms: Operators and Tools for Behavioral **Analysis Journal**, [S.l.], p. 265-319, 1998.
- 37 ZSOLT, L. **Redes Neurais Artificiais: Fundamentos e Aplicações**. 4. ed. [S.l]: Livraria de Física, 2006.
- 38 CAMPOS, L.C.D. **Modelo Estocástico Periódico baseado em Redes Neurais**. 2010 Tese (Doutorado). Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio).
- 39 LIMA, E.L. **Análise Real. Funções de uma Variável.**, 8. ed., Rio de Janeiro: IMPA, v.1, 2006.
- 40 VANDERPLAATS, G. **Numerical Optimization Techniques for Engineering Design**. 3.ed., Vanderplaats Research e Development, Inc, Colorado Springs, 1990.

- 41 RAGSDALE, C. **Spreadsheet Modeling e Decision Analysis: A Practical Introduction to Management Science**. 4.ed. [S.l.]: South-Western, 2004.
- 42 HERRERA F.; LOZANO M.V.J.L. Artificial Intelligence Review. Tackling Real-coded Genetic Algorithms: Operators and Tools for **Behavioral Analysis, Journal**, [S.l.], p. 265-319, 1998.
- 43 MORETTIN, P.A.; TOLOI, L.M.C. **Análise Séries Temporais**. 2. Ed. [S.l.]: ABE Projeto Fisher; Ed. Edgard Blucher, 2006.
- 44 PALIT, A.K.; POPOVIC, D. **Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications**. London: Springer-Verlag, 2005.
- 45 FARIA, A.E.; MUBWANDARIKWA, E. Multimodality on the Geometric Combination of Bayesian Forecasting Models. **International Journal of Statistics and Management System**, [S.l.], n. 3, p. 1-25, 2008.
- 46 FLORES, B.E.; WHITE, E.M. Subjective versus Objective Combining of Forecasts: An Experiment. **Journal of Forecasting**, [S.l.], v. 8, p. 331-341, 1989.
- 47 WERNER, L.; RIBEIRO, J.L.D. Modelo Composto para Prever Demanda através da Integração de Previsões. **Produção**, [S.l.], v. 16, n. 3, 2006.
- 48 WALLIS, K.F. Combining Forecasts - Forty Years Later. **Applied Financial Economics**, [S.l.], v.21, p. 33-41, 2011.
- 49 CLEMEN, R.T. Combining Forecasts: A Review and Annotated Bibliography. **International Journal of Forecasting**, [S.l.], v. 5, p.559-583, 1989.
- 50 GENEST, C.; ZIDEKJ, V. Combining probability distributions: A critique and annotated bibliography. **Statistical Science**, [S.l.], n.1, p. 114-135, 1986.
- 51 MUBWANDARIKWA, E. **The Geometric Combination of Bayesian Forecasting Models**. Thesis. Department of Statistics. Open University, London (UK), 2007.
- 52 MAKRIDAKIS, S.; WINKLER, R.L. Averages of Forecasts: Some Empirical Results. **Management Science**, [S.l.], v. 29, n. 9, p. 987-996, 1983.
- 53 TEIXEIRA JR, L.A. **Combinação Geométrica de Métodos Preditivos: aplicação à Série de Consumo Residencial de Energia Elétrica**. Dissertação (Mestrado), PUC-Rio, 2009.
- 54 BATES, J.M.; GRANGER, C.W.J. The Combining of Forecasts. **Operational Research Quarterly**, [S.l.], v. 20, n.4, p. 451-468, 1969.

- 55 SMITH, A.F.M.; MAKOV, U.E. A Quasi-Bayes Sequential Procedure for Mixtures. **Journal of the Royal Statistical Society**, [S.l.], Series B (Methodological) . v. 40, n. 1, p. 106-112, 1978.
- 56 CASELLA, G.; BERGER, R.L. **Statistical Inference**, [S.l.:s.n], p. 271-307.
- 57 TEIXEIRA JR, L.A.; PESSANHA, J.F.M; SOUZA, R.C. Combinação Linear de Previsões com Ajuste. **Anais...** In: MINIMAX, 14; Simpósio de Pesquisa Operacional e Logística da Marinha, Rio de Janeiro, Brasil, 2011^a.
- 58 MAKRIDAKIS, S. Why Combining Works? **International Journal of Forecasting**, [S.l.], v. 5, p. 601-603, 1989.
- 59 WINKLER, R.L. Combining Forecasts: A Philosophical Basis and Some Current Issues. **International Journal of Forecasting**, [S.l.], v. 5, p. 605-609, 1989.
- 60 GUPTA, S.; WILTON, P.C. Combination of Forecasts: An Extension. **Management Science**. v. 33, n. 3, p. 356-372, 1987.
- 61 BATCHELOR, R.; DUA, P. Forecaster Diversity and the Benefits of Combining Forecasts. **Management Science**, [S.l.], v. 41, n. 151, p. 68-75, 1995.
- 62 FLORES, B.E.; WHITE, E.M. A Framework for the Combination of Forecasts. **Journal Academic Marketing Science**, [S.l.], v. 16, n. 3-4, p. 95-103, 1998.
- 63 ARMSTRONG, J.S. Combining Forecasting. In: ARMSTRONG, J.S. **Principles of Forecasting: a Handbook for Researchers and Practitioners**, [S.l.]: Kluwer Academic Publishers, 2001.
- 64 ZOU, H.; YANG, Y. Combining Time Series Models for Forecasting. **International Journal of Forecasting**, [S.l.], v. 20, n. 1, p. 69-84, 2004.
- 65 SOUZA, R.C.; TEIXEIRA JR, L.A.; CASSIANO, K.M.; MENEZES L.M.; PESSANHA, J.F.M. Artificial Neural Networks Integrated with Singular Spectrum Analysis and Wavelet Theory in the Short-Term Wind Speed Forecasting. **Proceedings...** In: International Workshop on Energy Efficiency for a More Sustainable World. Portugal: [s.n], 2012b.
- 66 EFRON, B. *Bootstrap* methods: another look at the Jackknife. **The Annals of Statistics**, [S.l.], v.7, n. 1, p. 1-26, 1979.
- 67 MONTGOMERY, D.C.; RUNGER, G.C. **Estatística aplicada e probabilidade para engenheiros**, 2. ed . Rio de Janeiro: LTC, 2003.
- 68 FREEDMAN, D.A.; PETERS, S.C. *Bootstrapping* regression equation: some empirical results. **JASA**, [S.l.], v.79, p. 97-106, 1984.

- 69 EFRON, B.; TIBSHIRANI, R.J. *Bootstrap* methods for standart errors, confidence intervals, and other measures of statistical accuracy. **Statistical Science**, [S.l.], v.1, p. 54-77, 1986.
- 70 SOUZA, R.C.; CAMARGO, M.E. **Análise e Previsão de Séries Temporais**: os modelos ARIMA. Ijuí: SEDIGRAF, 1996.
- 71 HESKES, T. Pratical confidence and predction intervals. **Advances in neural information processing systems**, [S.l.], v. 9, p. 466-472, 1997.
- 72 KHOSRAVI, A. *et al.* Constructing Optimal Prediction Intervals by Using Neural Networks and *Bootstrap* Method. **Neural networks and learning systems**, [S.l.], p. 1-6, 2014.
- 73 TIBSHIRANI, R. A comparison of some error estimates for neural network models. **Neural Computation**, [S.l.], v. 8, p. 152-163, 1996.
- 74 CARNEY, J.G.; CUNNINGHAM, P.; BHAGWAN, U. Confidence and prediction intervals for neural network ensembles. **Anais...** In: THE International Joint Conference Of Neural Networks (IJCNN'99), Washington DC, USA: [s.n.], 1999
- 75 NIX, D.A.; WEIGEND, A.S. Learning local error bars for non – linear regression. **Proceedings...** In: NIPS, 7, p. 489-496, 1995.
- 76 BARTKIEWICZ, W. Prediction Intervals for Short-Term Load Forecasting Neuro-Fuzzy Models. **Przegląd elektro technic zny**, [S.l.], p. 284-287, 2012.
- 77 HAMILTON, J. 2009. **Causes and Consequences of the Oil Shock of 2007-08**, Brookings Papers on Economic Activity, Economic Studies Program, The Brookings Institution, v. 40 (1 (Spring), p. 215-283, 2007.
- 78 MORAN, The statistical analysis of the Canadian Lynx cycle. **Australian Journal of Zoology**, [S.l.], v. 1,n.3, p. 291-298, 1953.
- 79 SUBBA, R.T.; SABR, M.M. An Introduction to Bispectral Analysis and Bilinear Time Series Models. **Lecture Notes in Statistics**, [S.l.], v. 24, 1984.
- 80 TANG, Y.; GHOSAL, S. A consistent nonparametric Bayesian procedure for estimating autoregressive conditional densities. **Computational Statistics & Data Analysis**, [S.l.], v. 51, n. 9, p. 4424-4437, 2007.
- 81 WONG, C. S.; LI, W. K. On a mixture autoregressive model. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, [S.l.], v. 62, n. 1, p. 91-115, 2000.
- 82 KAJITANI, Y.; KEITH, W.H.; McLEOD, A.I. Forecasting Nonlinear Time Series With Feed-Forward Neural Networks: A case Study of Canadian Lynx Data. **Journal of Forecasting**, [S.l.], v. 24, p. 105-117, 2005.

- 83 ALADAG, C.H.; EGRIOGLU, E.; KADILAR, C. Forecasting Nonlinear Time Series with a Hybrid Methodology. **Applied Mathematics Letters**, [S.l.], v. 22, p. 1467-1470, 2009.
- 84 KASHEI, M.; BIJARI, M. An Artificial Neural Network (p,d,q) Model for Times Forecasting. **Expert Systems With Applications**, [S.l.], v. 37, p. 479-489, 2010.
- 85 ISMAIL, S.; SHABRI, A. Time Series Forecasting using Least Square Support Vector Machine for Canadian Lynx Data. **Journal Teknology**, [S.l.], p. 11-15, 2014.
- 86 KASHEI, M.; BIJARI, M. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. **Applied Soft Computing**, [S.l.], v. 11, p. 2664-2675, 2011.
- 87 KASHEI, M.; BIJARI, M. A new class of hybrid models for time series forecasting. **Expert Systems with Applications**, [S.l.], v. 39, p. 4344-4357, 2012.
- 88 CASELLA, C.; BERGER, R.L. **Statistical Inference**, [Sl.:s.n.], p.330-367
- 89 MACEIRA, M.E.P. **Operação Ótima de Reservatórios com Previsão de Afluências**, 1989. Dissertação (Mestrado) Coordenação dos Programas de Pós-Graduação de Engenharia da Universidade Federal do Rio de Janeiro, Brasil, 1989.
- 90 WANG, A.B.W. *et al.* Forecasting daily streamflow using hybrid ANN models. Amsterdam: **Journal of Hydrology**, [S.l.], v. 324, p. 383-399, 2006a.
- 91 WANG, A.B.W. *et al.* Testing for nonlinearity of streamflow processes at different time scales. Amsterdam: **Journal of Hydrology**, [S.l.], v. 322, p. 247-268, 2006b.
- 92 VALENZUELA, O. Hybridization of intelligent techniques and ARIMA models for time series prediction. **Fuzzy Sets and Systems**, [S.l.], v.159, p. 821-845, 2008.
- 93 FERREIRA, C.C. **Previsão de vazões naturais diárias afluentes ao reservatório UHE Tucuruí utilizada a técnica de redes neurais artificiais**. 2012, 117f. Dissertação (Mestrado Engenharia da Computação) – Universidade Federal de Goiás, Goiânia, 2012.
- 94 ABREU, T. *et al.* Metodologia híbrida utilizando os modelos ARIMA e redes neurais artificiais para previsão de cargas elétricas. **Anais...** In: Congresso Brasileiro de Automática, 19., Campina Grande: [s.n.], 2012.
- 95 TEIXEIRA JR, L.A.; MENESES, M.L.; PESSANHA, J.F.M. SOUZA, R.C.; CASSIANO, K.M. Redes Neurais Artificiais *Wavelet* Causal na

Previsão da Radiação Solar Direta **Anais...** In: Simpósio Brasileiro de Pesquisa Operacional, [S. l.:s.n.]: 2012a

- 96 HWARNG, H.B.; ANG, H.T. A simple neural network for ARMA(p; q) time series. Singapore: **Omega**, [S.l.], v. 29, p. 319-333, 2001.
- 97 SUJJAVIRIYASUP, T. Hybrid ARIMA-Support Vector Machine Model for Agricultural Production Planning. **Applied Mathematical Sciences**, [S.l.], v. 7, n. 57, p. 2833-2840, 2013.
- 98 YAZIZ, S.R. The performance of hybrid ARIMA GARCH modeling in forecasting gold price. **Anais...** In: Adelaide: International Congress on Modelling and Simulation, 20, Austrália, 2013.
- 99 MENEZES L.M.; TEIXEIRA JR, L.A.T; CASSIANO K.M.; SOUZA, R.C.; and PESSANHA, J.F.M. SSA, Hierarchical Clustering and ARIMA Model in Forecast of Affluent Flow. **Proceedings...**In: Annual International Symposium on Forecasting, 32, Boston, USA, 2012.
- 100 HIRPA, F.A. **Hydrologic data assimilation for operational streamflow Forecasting**. 2013, 153 f. Doctoral Dissertations, University of Connecticut, 2013.
- 101 GRANGER, C.W.J.; NEWBOLD, P. Spurious regressions in econometrics. **Journal of Econometrics**, [S.l.], v. 2, p. 111-120, Jul 1974.