

# COMBINAÇÃO LINEAR WAVELET SARIMA-RNA COM ESTÁGIOS MÚLTIPLOS NA PREVISÃO DE SÉRIES TEMPORAIS

**Luiz Albino Teixeira Júnior**

Universidade Federal da Integração Latino-Americana (UNILA), Brasil.

E-mail: luiz.junior@unila.edu.br

**Álvaro Eduardo Faria Júnior**

The Open University (OU), Inglaterra

E-mail: alvaro.faria@open.ac.uk

**Ricardo Vela de Britto Pereira**

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brasil.

E-mail: ricardoestatístico@yahoo.com.br

**Reinaldo Castro Souza**

Pontifícia Universidade Católica do Rio de Janeiro (PUC-Rio), Brasil.

E-mail: reinaldo@ele.puc-rio.br

**Edgar Manuel Carreño Franco**

Universidade Estadual do Oeste do Paraná (UNIOESTE), Brasil

E-mail: emfra.unioeste@gmail.com

## Resumo

Neste artigo, é proposta uma metodologia híbrida para se combinar previsões de séries temporais (estocásticas) chamada de Combinação Linear Wavelet (CLW) SARIMA-RNA com Estágios Múltiplos. Primeiramente, é feita a decomposição wavelet de nível  $p$ , gerando-se  $p+1$  (aproximações das) componentes wavelet (CWs). Em seguida, as CWs são individualmente modeladas por meio de um modelo de Box e Jenkins e de uma rede neural artificial - a fim de se capturar, respectivamente, estruturas lineares e não-lineares plausíveis de autodependência - para, em seguida, serem linearmente combinadas, fornecendo previsões híbridas para cada uma. Por fim, as referidas são linearmente combinadas pela CLW de previsões (a ser definida). Para avaliá-la, utilizaram-se modelos de Box e Jenkins (BJ), redes neurais artificiais (RNA) e sua tradicional Combinação Linear (CL1) de previsões; e RNA integrado com a decomposição wavelet (RNA-WAVELET), modelo BJ integrado com decomposição wavelet (BJ-WAVELET) e sua tradicional Combinação Linear (CL2) de previsões. Todos os métodos preditivos aplicados à série temporal mensal de vazão média de afluentes da barragem da Usina de Itaipu, localizada em Foz do Iguaçu, Brasil. Em todas as análises, a metodologia híbrida proposta obteve desempenho preditivo superior que as outras.

**Palavras-chaves:** Séries Temporais. Decomposição Wavelet. Modelos de Box e Jenkins. Redes Neurais Artificiais. Combinação Linear de Previsões.

## Abstract

In this paper, we put forward a hybrid methodology for combining forecasts to (stochastic) time series referred to as Wavelet Linear Combination (WLC) SARIMA-RNA with Multiple Stages. Firstly, the wavelet decomposition of level  $p$  is performed, generating (approximations of the)  $p+1$  wavelet components (WCs). Then, the WCs are individually modeled by means of a Box and Jenkins' model and an artificial neural network - in order to

capture, respectively, plausible linear and non-linear structures of autodependence - for, then, being linearly combined, providing hybrid forecasts for each one. Finally, all of them are linearly combined by the WLC of forecasts (to be defined). For evaluating it, we used the Box and Jenkins' (BJ) models, artificial neural networks (ANN), and its traditional Linear Combination (LC1) of forecasts; and ANN integrated with the wavelet decomposition (ANN-WAVELET), BJ model integrated with the wavelet decomposition (BJ-WAVELET), and its conventional Linear Combination (LC2) of forecasts. All predictive methods applied to the monthly time series of average flow of tributaries of the Itaipu Dam dam, located in Foz do Iguaçu, Brazil. In all analysis, the proposed hybrid methodology has provided higher predictive performance than the other ones.

**Keywords:** Time Series. Wavelet Decomposition. Box and Jenkins' Models. Artificial Neuron Networks. Linear Combination of Forecasts.

## 1. INTRODUÇÃO

Ao longo dos anos, diversos métodos de previsão (ou preditivos) vêm sendo propostos com o objetivo de se projetar, de forma cada vez mais acurada, séries temporais (estocásticas). Em linhas gerais, podem-se agrupá-los em duas classes disjuntas: os métodos preditivos individuais (que se subdividem em os de abordagem estatística e os de inteligência artificial), e os de combinação de métodos preditivos individuais [Bates & Granger, 1969; Newbold & Granger, 1974; Makridakis & Winkler, 1983; Gupta & Wilton, 1987; Makridakis, 1989; Flores & White, 1989; Flores & White, 1998; Zou & Yang, 2004; Faria & Mubwandarikwa, 2008]. É importante salientar que termo “combinação de métodos preditivos individuais” pode ser utilizado em sentido amplo, referindo-se tanto à abordagem de combinação de previsões [Makridakis & Winkler, 1983], quanto à de combinação de densidades preditivas bayesianas [Faria & Mubwandarikwa, 2008]. Neste artigo, porém, é utilizado apenas para se referir à combinação de previsões.

Na literatura, existe uma gama de métodos preditivos individuais que são amplamente utilizados na projeção de séries temporais, como, por exemplo, os de Box & Jenkins [Hamilton, 1994] e os de redes neurais artificiais (RNAs) [Haykin, 2001] - os quais capturam, respectivamente, estruturas lineares e não-lineares de autodependência. Assim, uma vez que as séries temporais, em geral, exibem ambas, segue a utilização de um modelo de Box & Jenkins ou de uma RNA, de acordo com Zhang (2003) e Wallis (2011), pode implicar perda de informações (lineares, pela RNA; e não-lineares, pelo modelo de Box & Jenkins) relevantes à obtenção de maior poder preditiva. Portanto, previsões geradas da combinação de modelos individuais plausíveis, não viesados e consideravelmente distintos tendem a ser mais acuradas, pois, de acordo com Faria & Mubwandarikwa (2008), são, de fato, agregadoras de informações oriundas de diferentes fontes.

Da Análise Wavelet (ou Teoria Wavelet), por sua vez, advém importantes métodos auxiliares de pré-processamento que consistem, basicamente, em fazer a decomposição, filtragem ou alisamento dos dados temporais, antes de sua efetiva modelagem [Aquino et al., 2009]. Diversas abordagens utilizam, de forma integrada, métodos preditivos individuais e métodos de pré-processamento wavelet. Algumas delas podem ser visualizadas em: Donoho & Johnstone, 1994; Donoho et al., 1995; Lei & Ran (2008), Orgen & Parzen (1996).

Dado o exposto e o fato de haver um número relativamente pequeno de método de combinação de modelos preditivos individuais que fazem o tratamento dos dados, propõe-se a: “combinação linear wavelet SARIMA-RNA com estágios múltiplos”. Basicamente, o método proposto pode ser descrito nos seguintes passos básicos: (1) obtêm-se as mensurações de uma série temporal a ser modelada; (2) realiza-se a sua decomposição de nível  $p$ , gerando-se 1 componente de aproximação e  $p$  componentes de detalhes; (3) modela-se cada uma das  $p+1$  componentes wavelet (CW) do passo (2) através de um modelo de Box & Jenkins (para mapear uma estrutura linear plausível) e de uma rede neural artificial (para mapear uma estrutura não-linear plausível); (4) combinam-se as previsões de cada CW oriundas de ambos os previsores individuais mencionados através de uma linear de previsões tradicional, onde o ajuste dos parâmetros adaptativos é feito por meio de programação não-linear (desde modo, dado que são  $p+1$  componentes wavelet, são necessárias combinações lineares e, conseqüentemente,  $p+1$  problemas de programação não-linear, neste passo; e (5) combinam-se linearmente as previsões híbridas de cada CW geradas no passo (4), gerando as previsões para a série temporal desejada. Note que, no passo (5), é necessário resolver mais um programa não-linear, totalizando  $p+2$  programas não-lineares (que é o que justifica a expressão “estágios múltiplos” no nome do método..

Além da introdução, é dividido em mais sete seções. Na seção 2, são apresentados conceitos básicos de Teoria Wavelet. Nas seções 3 e 4, são introduzidos, respectivamente, os modelos de Box & Jenkins e de redes neurais artificiais. A definição de linear combinação é feita na seção 5. Na seção 6, descreve-se detalhadamente a metodologia proposta. Finalmente, nas seções 7 e 8, são explicitados, respectivamente, os principais resultados e as conclusões.

## 2. ANÁLISE WAVELET

Tome o par ordenado  $(l^2, \langle ; \rangle)$ , onde  $l^2$  é definida pela coleção de todas as sequências infinitas de números complexos quadraticamente somáveis (isto é,  $l^2 := \{f: \mathbb{Z} \rightarrow \mathbb{C}: \sum_{t \in \mathbb{Z}} |f(t)|^2 < \infty\}$ , onde  $\mathbb{Z}$  e  $\mathbb{C}$  são, respectivamente, o conjunto dos números inteiros e complexos) e o mapa  $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$  é um produto interno. Um elemento  $\omega(\cdot)$  em  $l^2$  é um vetor  $l^2$  –wavelet (ou, simplesmente, vetor wavelet), com produto interno  $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$ , se e somente se, a sequência duplamente indexada  $\{\omega_{m,n}(\cdot)\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$  consiste em uma base ortonormal para  $l^2$ , onde o parâmetro  $m$  é chamado de parâmetro de escala diádica e o  $n$ , chamado de parâmetro de translação. Por sua vez, um elemento  $\phi(\cdot) \in l^2$  é um vetor  $l^2$  –escala (ou, simplesmente, vetor escala), com produto interno  $\langle ; \rangle: l^2 \rightarrow \mathbb{C}$ , se, e somente se, a coleção  $\{\phi_{m,n}(\cdot)\}_{(m,n) \in \mathbb{Z} \times \mathbb{Z}}$ , onde  $\phi_{m,n}(\cdot) := 2^{\frac{m}{2}} \phi(2^m(\cdot) - n)$ , para todo  $m, n \in \mathbb{Z}$ , é tal que  $\langle \phi_{l,i}(\cdot); \phi_{j,k}(\cdot) \rangle = 0$ , sempre que  $l = j$  e  $i \neq k$ , e  $\langle \phi_{l,i}(\cdot); \phi_{j,k}(\cdot) \rangle \neq 0$ , se caso contrário.

Baseado em Kubrusly & Leva (2006) e Kubrusly (2011), um vetor  $f(\cdot)$  em  $l^2$  pode ser decomposto ortogonalmente, de forma única, em termos de uma base ortonormal wavelet, representada por  $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{m,n}(\cdot)\}_{(m,n) \in \{m\}_{m=m_0}^{+\infty} \times \mathbb{Z}}$ , tal como em (1).

$$f(\cdot) = f_{V_{m_0}(\phi)}(\cdot) + \sum_{m=m_0}^{+\infty} f_{W_m(\omega)}(\cdot) \quad (1)$$

Onde:  $f_{V_{m_0}(\phi)}(\cdot) := \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(\cdot)$  é a componente de aproximação de nível  $m_0$ , sendo que  $a_{m_0,n} := \langle f(\cdot), \phi_{m_0,n}(\cdot) \rangle$  (isto é, o produto interno usual entre o vetor de sinais  $f(\cdot)$  e o vetor escala de nível  $m_0$  e  $n$ ); e  $f_{W_m(\omega)}(\cdot) := \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(\cdot)$  é a componente de detalhe de nível  $m$ , sendo que  $d_{m,n} := \langle f(\cdot), \omega_{m,n}(\cdot) \rangle$  (ou seja, o produto interno usual entre  $f(\cdot)$  e o vetor wavelet de nível  $m$  e  $n$ ).  $a_{m_0,n}$  e  $d_{m,n} := \langle f(\cdot), \omega_{m,n}(\cdot) \rangle$  são, respectivamente, os coeficientes

wavelet de aproximação e de detalhe. Em (1), tem-se a decomposição wavelet, em termos de  $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{n,m}(\cdot)\}_{(m,n) \in \{m\}_{m=m_0}^{+\infty} \times \mathbb{Z}}$ , de  $f(\cdot)$  sobre  $L^2$ .

### 3. MODELAGEM BOX & JENKINS

De acordo com Liu (2006), se uma série temporal  $(y_t)_{t=1}^T$  é um processo auto-regressivo integrado com médias móveis de ordem  $p$  e  $q$  (ARMA( $p,q$ )), se e somente se, pode ser representado tal como em (2).

$$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (2)$$

Onde:  $(\varphi_i)_{i=1}^p$  e  $(\theta_j)_{j=1}^q$  consistem em listas de parâmetros complexos que satisfazem às condições de invertibilidade e de estacionariedade [Box & Jenkins, 1976];  $(\varepsilon_{t-j})_{j=0}^q$  é uma realização de um processo estocástico ruído branco centrado em zero [Morettin & Toloi, 2006]. Em (7), tem-se uma representação alternativa do modelo, em (3), o qual é dado em termos de dois polinômios.

$$(1 - \varphi_1 B - \dots - \varphi_p B^p) y_t - e_t = (1 - \theta_1 B - \dots - \theta_q B^q) \varepsilon_t \quad (3)$$

Onde:  $B$  é um shift unilateral [Kubrusly (2011)] tal que:  $B^k y_t := y_{t-k}$ , sendo  $k$  um inteiro estritamente positivo. Na classe de modelos ARMA ( $p,q$ ), tem-se que: AR( $p$ ) representa o polinômio  $(1 - \varphi_1 B - \dots - \varphi_p B^p)$  de ordem  $p$  da parte auto-regressiva; e MA( $q$ ), o polinômio  $(1 - \theta_1 B - \dots - \theta_q B^q)$  de ordem  $q$  da parte de médias móveis. Por sua vez, se uma série temporal  $(y_t)_{t=1}^T$  apresentar efeitos de sazonalidade (estacionária ou não estacionária homogênea) ou não estacionariedade homogênea na média, então uma classe de modelos plausíveis é dada genericamente em (4).

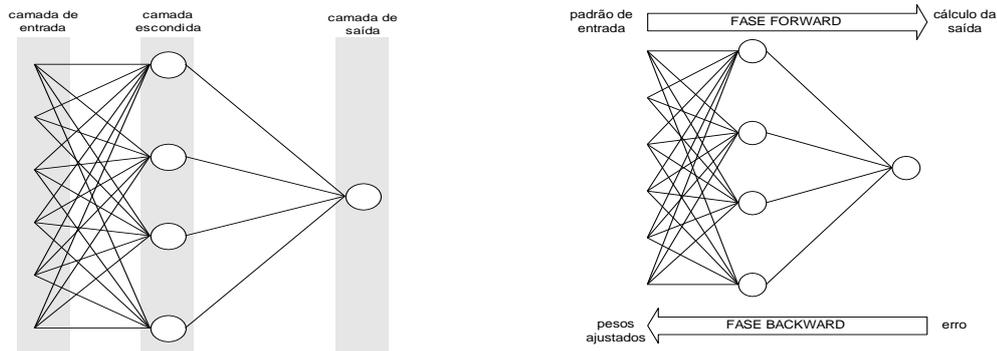
$$\varphi(B)(1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}) \nabla^d \nabla_S^D y_t - e_t = \theta(B)(1 - \Theta_1 B^S - \dots - \Theta_Q B^{QS}) \varepsilon_t \quad (4)$$

Onde:  $\Phi_m \in \mathbb{C}$  e  $\Theta_n \in \mathbb{C}$ , onde  $n = 1, \dots, P$  e  $m = 1, \dots, Q$ , são os parâmetros complexos dos polinômios sazonais referentes, respectivamente, à parte auto-regressiva e de médias móveis sazonais, sendo que satisfazem às condições de invertibilidade e de estacionariedade [Box & Jenkins (1976)];  $D$  é a ordem do operador  $\nabla_S^D$  de diferenças sazonais (necessário em casos de não estacionariedade homogênea da componente sazonal de  $(y_t)_{t=1}^T$ ), que é definido por  $\nabla_S^D := (1 - B^S)^D$ ;  $S$  é o período sazonal (se anual, então  $S = 12$ );  $d$  é a ordem do operador  $\nabla := (1 - B)$  das diferenças simples (necessário em casos de não estacionariedade homogênea na componente média de  $(y_t)_{t=1}^T$ ); e os polinômios  $\varphi(B) := (1 - \varphi_1 B - \dots - \varphi_p B^p)$  e  $\theta(B) := (1 - \theta_1 B - \dots - \theta_q B^q)$  referem às partes simples auto-regressiva e de médias móveis (da equação 3), respectivamente. A classe de modelos de Box e Jenkins, em (4), é denotada por SARIMA ( $P,D,Q$ )  $\times$  ( $p,d,q$ ).

### 4. REDES NEURAIS ARTIFICIAIS

De acordo com Haykin (2001), as redes neurais artificiais são sistemas paralelos distribuídos e compostos por unidades de processamento simples, denominadas neurônios artificiais, e dispostas em uma ou mais camadas interligadas por um grande número de conexões (sinapses), geralmente, unidirecionais (feedforward) e com pesos sinápticos para ponderação das entradas recebidas por cada neurônio. Na Figura 2(a), tem-se uma ilustração da arquitetura mais comum de uma RNA (feedforward) multicamadas perceptron (MLP) com três camadas: camada de entrada (input), camada escondida (ou intermediária) e camada de saída (output).

A primeira camada da rede neural artificial é a camada de entrada, sendo a única exposta aos padrões de entrada. A camada de entrada transmite os valores dos padrões de entrada para os neurônios da camada intermediária para que estes extraiam padrões e transmitam os resultados para a camada de saída (última camada da RNA). A definição do número de neurônios na camada intermediária é realizada de forma empírica.

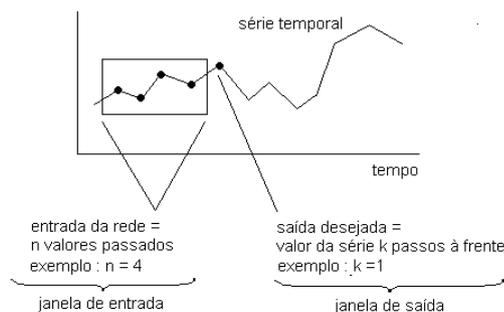


(a) RNA Feedforward de 3 Camadas. (b) Fases do Algoritmo Backpropagation.

**Figura 2 -** Rede Neural Artificial Feedforward MLP e Algoritmo Backpropagation.

O principal algoritmo de treinamento de redes neurais artificiais é o backpropagation, cujo ajuste dos pesos sinápticos ocorre por meio de um processo de otimização realizado em duas fases: forward e backward, conforme ilustrado na Figura 2(b). Na fase forward, é calculada a resposta fornecida pela RNA para determinado padrão de entrada. Na fase backward, o desvio (ou erro) entre a resposta da RNA e a resposta desejada é utilizado no processo de ajuste dos pesos sinápticos. Ao longo do treinamento da RNA, os vários padrões de entrada e as respectivas respostas desejadas são apresentados à RNA, de forma que os pesos sinápticos sejam tais que minimizem a soma dos erros quadráticos (MSE). A previsão de valores futuros de uma série temporal, por meio de uma RNA, inicia-se com a montagem do conjunto de padrões de treinamento (pares entrada/saída), o qual depende da definição do tamanho da janela L de tempo (para os valores passados da própria série temporal que se deseja prever e para as variáveis explicativas) e do horizonte de previsão h. Em um processo auto-regressivo (linear ou não linear), o padrão de entrada é formado pelos valores passados da própria série que se deseja prever.

Por sua vez, o padrão de saída desejada é o valor da observação da série temporal no horizonte de previsão. Na Figura 3, ilustra-se como geralmente é construído o conjunto de treinamento no caso da previsão basear-se nos 4 últimos valores passados. Note que a construção dos padrões de treinamento da rede consiste em mover as janelas de entrada e saída ao longo de toda série temporal, de modo que cada par de janelas (entrada/saída) funciona como um padrão de treinamento e deve ser apresentado repetidas vezes até que o algoritmo de aprendizado convirja.



**Figura 3 -** Montagem do Conjunto de Treinamento.

## 5. COMBINAÇÃO DE PREVISÕES

Assuma que  $(y_t)_{t=1}^T$  seja uma série temporal e que  $\{M_j\}_{j=1}^m$ ,  $m \geq 2$  seja uma coleção de métodos preditivos plausíveis à sua modelagem com o fim de se gerar previsões. Neste processo, pode-se escolher algum método em  $\{M_j\}_{j=1}^m$ , baseado em algum critério de seleção, ou tomar  $k$  métodos em  $\{M_j\}_{j=1}^m$ , onde  $k \leq m$ , e combiná-los. Tome  $k$  métodos na coleção  $\{M_j\}_{j=1}^m$ , onde  $k \leq m$ . Assuma que  $\nabla^k$  seja o conjunto com todas as previsões (dentro e fora da amostra) provenientes dos  $k$  métodos escolhidos em  $\{M_j\}_{j=1}^m$ , considerando um horizonte de previsão igual a  $h$  passos à frente. Considere ainda que  $\nabla_C$  seja o conjunto com todas as previsões oriundas de determinada combinação das previsões pertencentes ao conjunto  $\nabla^k$ .

Por combinação de previsões (genérica), entende-se como sendo uma forma  $\hat{y}_C: \nabla^k \rightarrow \nabla_C$  tal que faz o mapeamento de um vetor de previsões  $(\hat{y}_{t,k})_{k=1}^K \in \nabla^k$  em uma previsão combinada  $\hat{y}_{t,C} \in \nabla_C$  - onde  $t$  é um inteiro em  $1 \leq t \leq T + h$ , sendo  $h$  o horizonte de previsão e  $T$ , a cardinalidade da amostra de treino. Em particular, a combinação linear de previsões consiste de uma forma  $\hat{y}_{CL}: \nabla^k \rightarrow \nabla_{CL}$  que faz o mapeamento de um vetor de previsões  $(\hat{y}_{t,k})_{k=1}^K \in \nabla^k$  em uma previsão linearmente combinada  $\hat{y}_{t,CL} \in \nabla_{CL}$ , a qual é definida em (5).

$$\hat{y}_{CL,t} = \left[ \sum_{i=1}^k (\rho_i) \times (\hat{y}_{t,i}) \right] + \beta \quad (5)$$

Onde:  $\rho_i$  é o peso adaptativo associado à previsão  $\hat{y}_{t,i}$ , sendo  $t$  um instante inteiro no intervalo  $1 \leq t \leq T + h$ , a qual é obtida através do método  $M_i$ ; e  $\beta$  são, respectivamente, as constantes adaptativas multiplicativa e aditiva.

## 6. METODOLOGIA PROPOSTA

Seja  $\chi_{[0,T]}$  um mapa tal que  $\chi_{[0,T]}(t) = \begin{cases} 1, & \text{se } t \in \{1, \dots, T\} \\ 0, & \text{se } t \in \mathbb{Z} - \{1, \dots, T\} \end{cases}$  e  $y(\cdot) := (y_t)_{t=1}^T$ , uma série temporal de tamanho  $T$ , onde  $T > 1$ . Assim, a composição  $\chi_{[0,T]} \circ y(\cdot) = \tilde{y}(\cdot): \mathbb{Z} \rightarrow \mathbb{R}$  pode ser visualizada como uma sequência  $\tilde{y}(\cdot) := (\dots, 0, 0, y_1, y_2, \dots, y_T, 0, 0, \dots)$  no espaço  $l^2$ . Assumindo que  $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}} \cup \{\omega_{m,n}(\cdot)\}_{(m,n) \in \{m\}_{m=m_0}^{+\infty} \times \mathbb{Z}}$  seja uma base ortonormal wavelet para  $l^2$ , segue que, baseado na Seção 2, a série temporal  $\tilde{y}(\cdot)$  admite ser decomposta, de forma aproximada, tal como em (6).

$$\tilde{y}(\cdot) \simeq \tilde{\tilde{y}}(\cdot) = \sum_{n=1}^{n_{m_0}} a_{m_0,n} \phi_{m_0,n}(\cdot) + \sum_{m=m_0}^{m_0+(p-1)} \sum_{n=1}^{n_m} d_{m,n} \omega_{m,n}(\cdot) \quad (6)$$

A expansão, em (6), é chamada de *decomposição wavelet de nível  $p$* . O valor adotado para o parâmetro de nível  $m_0$  é, usualmente, o mesmo de  $p$ . Reescrevendo (6), tem-se que:

$$\tilde{y}(\cdot) \simeq \tilde{\tilde{y}}(\cdot) = y'_{V_{m_0}(\phi)}(\cdot) + \sum_{m=m_0}^{m_0+(p-1)} y'_{W_m(\omega)}(\cdot) \quad (7)$$

Onde, em  $l^2$ :  $y'_{V_{m_0}(\phi)}(\cdot) := \sum_{n=1}^{n_{m_0}} a_{m_0,n} \phi_{m_0,n}(\cdot)$  é uma aproximação da CW de aproximação  $y_{V_{m_0}(\phi)}(\cdot)$ ; e  $y'_{W_m(\omega)}(\cdot) := \sum_{n=1}^{n_m} d_{m,n} \omega_{m,n}(\cdot)$ , da CW de detalhe  $y_{W_m(\omega)}(\cdot)$ .

Uma vez obtidas as  $p+1$  aproximações das CWs, em (7), realiza-se a modelagem de cada uma delas por meio de um modelo de Box & Jenkins (previsor linear) e uma rede neural artificial (previsor não-linear), mapeando, respectivamente, estruturas de autodependência linear e não-linear. De modo a produzir as seguintes sequências de previsões, dentro e fora da amostra, sendo que  $h$  denota o horizonte de previsão e  $T''$  e  $T'$  são os graus de liberdade perdidos pelos modelos RNA e de Box & Jenkins, respectivamente:

- i.  $\hat{y}_{V_{m_0}(\phi),BJ}(\cdot) := \left( \hat{y}_{V_{m_0}(\phi),BJ,t} \right)_{t=T'}^{T+h}$ ; e  $\hat{y}_{V_{m_0}(\phi),RNA}(\cdot) := \left( \hat{y}_{V_{m_0}(\phi),RNA,t} \right)_{t=T''}^{T+h}$ ; e
- ii.  $\hat{y}_{W_m(\omega),BJ}(\cdot) := \left( \hat{y}_{W_m(\omega),BJ,t} \right)_{t=T'}^{T+h}$ ; e  $\hat{y}_{W_m(\omega),RNA}(\cdot) := \left( \hat{y}_{W_m(\omega),RNA,t} \right)_{t=T''}^{T+h}$ , para cada valor inteiro  $m$  no intervalo  $m_0 \leq m \leq m_0 + (p - 1)$ .

Tomando-se as previsões da amostra de treinamento oriundas dos métodos preditivos individuais Box & Jenkins e redes neurais artificial para cada componente de aproximação, é feita a sua combinação linear das previsões, tal como em (8).

$$\hat{y}_{V_{m_0}(\phi),CL,t} := \hat{y}_{V_{m_0}(\phi),BJ,t} \times \rho_{V_{m_0}(\phi),BJ} + \hat{y}_{V_{m_0}(\phi),RNA,t} \times \rho_{V_{m_0}(\phi),RNA} + \beta_{V_{m_0}(\phi),CL} \quad (8)$$

Onde:  $\hat{y}_{V_{m_0}(\phi),BJ,t}$  é a previsão do modelo de Box & Jenkins para  $y_{V_{m_0}(\phi),t}$ ;  $\hat{y}_{V_{m_0}(\phi),RNA,t}$  é a previsão do modelo de redes neurais artificiais para  $y_{V_{m_0}(\phi),t}$ ;  $\rho_{V_{m_0}(\phi),BJ}$  é peso adaptativo associado a  $\hat{y}_{V_{m_0}(\phi),BJ,t}$ ;  $\rho_{V_{m_0}(\phi),RNA}$  é peso adaptativo associado a  $\hat{y}_{V_{m_0}(\phi),RNA,t}$ ;  $\beta_{V_{m_0}(\phi),CL}$  é a constante adaptativa aditiva; e, finalmente,  $\hat{y}_{V_{m_0}(\phi),CL,t}$  é a previsão linearmente combinada para a aproximação  $y_{V_{m_0}(\phi),t}$ .

O ajuste dos parâmetros adaptativos, em (13) ocorre por meio de um problema de programação matemática (PPM) [Ragsdale, 2004], o qual é descrito a seguir.

$$\text{FO: } \min_{\rho_{V_{m_0}(\phi),BJ}; \rho_{V_{m_0}(\phi),RNA}; \beta_{V_{m_0}(\phi),CL}} \left\{ \text{MSE}_{V_{m_0}(\phi),CL} \right\}$$

$$\text{s. a: } \left\{ \begin{array}{l} \hat{y}_{V_{m_0}(\phi),CL,t} := \left( \hat{y}_{V_{m_0}(\phi),BJ,t} \times \rho_{V_{m_0}(\phi),BJ} \right) + \left( \hat{y}_{V_{m_0}(\phi),RNA,t} \times \rho_{V_{m_0}(\phi),RNA} \right) + \beta_{V_{m_0}(\phi),CL} \\ \text{MSE}_{V_{m_0}(\phi),CL} := \sum_{t=1}^T \frac{\left( y_{V_{m_0}(\phi),t} - \hat{y}_{V_{m_0}(\phi),CL,t} \right)^2}{T} \\ \rho_{V_{m_0}(\phi),BJ} + \rho_{V_{m_0}(\phi),RNA} = 1 \\ \rho_{V_{m_0}(\phi),BJ} \geq 0 \\ \rho_{V_{m_0}(\phi),RNA} \geq 0 \end{array} \right.$$

Em seguida, é feita a combinação linear das previsões para cada componente de detalhe de nível  $m$ , tal como em (9).

$$\hat{y}_{W_m(\omega),CL,t} := \hat{y}_{W_m(\omega),BJ,t} \times \rho_{W_m(\omega),BJ} + \hat{y}_{W_m(\omega),RNA,t} \times \rho_{W_m(\omega),RNA} + \beta_{W_m(\omega),CL} \quad (9)$$

Onde:  $\hat{y}_{W_m(\omega),BJ,t}$  é a previsão do modelo de Box & Jenkins para  $y_{W_m(\omega),t}$ ;  $\hat{y}_{W_m(\omega),RNA,t}$  é a previsão do modelo de redes neurais artificiais para  $y_{W_m(\omega),t}$ ;  $\rho_{W_m(\omega),BJ}$  é peso adaptativos associado à previsão  $\hat{y}_{W_m(\omega),BJ,t}$ ;  $\rho_{W_m(\omega),RNA}$  é peso adaptativos associado à previsão

$\hat{y}_{V_{m_0}(\phi),RNA,t}$ ;  $\beta_{W_{m(\omega),CL}}$  é a constante adaptativa aditiva; e  $\hat{y}_{W_{m(\omega),CL,t}}$  é a previsão linearmente combinada para a aproximação  $y_{V_{m_0}(\phi),t}$ .

O PPM utilizado, no ajuste da combinação, em (9), é dado a seguir.

$$\begin{aligned} \text{FO:} & \min_{\rho_{W_{m(\omega),BJ}}, \rho_{W_{m(\omega),RNA}}, \beta_{W_{m(\omega),CL}}} \{MSE_{W_{m(\omega),CL}}\} \\ \text{s. a:} & \left\{ \begin{aligned} \hat{y}_{W_{m(\omega),CL,t}} &= \left( \hat{y}_{W_{m(\omega),BJ,t}} \times \rho_{W_{m(\omega),BJ}} \right) + \left( \hat{y}_{W_{m(\omega),RNA,t}} \times \rho_{W_{m(\omega),RNA}} \right) + \beta_{W_{m(\omega),CL}} \\ MSE_{W_{m(\omega),CL}} &= \sum_{t=1}^T \frac{(y_{W_{m(\omega),t}} - \hat{y}_{W_{m(\omega),CL,t}})^2}{T} \\ \rho_{W_{m(\omega),BJ}} + \rho_{W_{m(\omega),RNA}} &= 1 \\ \rho_{W_{m(\omega),BJ}} &\geq 0 \\ \rho_{W_{m(\omega),RNA}} &\geq 0 \end{aligned} \right. \end{aligned}$$

Tal procedimento é repetido na modelagem individual de cada uma das  $p$  componentes wavelet de detalhe. Ou seja, até aqui são necessários  $p+1$  estágios de ajustes. Por fim, as previsões híbridas linearmente combinadas para as componentes wavelet de aproximação e de detalhe são combinadas, também de forma linear, tal como em (10).

$$\hat{y}_{CL,wave,t} := \left[ \left( \rho_{V_{m_0}(\phi)} \times \hat{y}_{V_{m_0}(\phi),CL,t} \right) + \sum_{m=m_0}^{m_0+(p-1)} \left( \rho_{W_{m(\omega)}} \times \hat{y}_{W_{m(\omega),CL,t}} \right) \right] + \beta_{CL,wave} \quad (10)$$

Onde:  $\hat{y}_{V_{m_0}(\phi),CL,t}$  é a previsão linearmente para  $y_{V_{m_0}(\phi),t}$ ;  $\hat{y}_{W_{m(\omega),CL,t}$  é a previsão linearmente para  $y_{W_{m(\omega),t}$ ;  $\rho_{V_{m_0}(\phi)}$  é peso adaptativos associado à previsão  $\hat{y}_{V_{m_0}(\phi),CL,t}$ ;  $\rho_{W_{m(\omega)}}$  é peso adaptativos associado à previsão  $\hat{y}_{W_{m(\omega),CL,t}$ ;  $\beta_{CL,wave}$  é a constante adaptativa aditiva; e  $\hat{y}_{CL,wave,t}$  é a previsão linearmente combinada para o ponto  $y_t$ .

A formulação do PPM utilizado no ajuste da combinação linear das componentes wavelet, em (10), é descrito a seguir.

$$\begin{aligned} \text{FO:} & \min_{\rho_{V_{m_0}(\phi)}, \rho_{W_{m(\omega)}}, \beta_{CL,wave}} \{MSE_{CL,wave}\} \\ \text{s. a:} & \left\{ \begin{aligned} \hat{y}_{CL,wave,t} &= \rho_{V_{m_0}(\phi)} \times \hat{y}_{V_{m_0}(\phi),CL,t} + \sum_{m=m_0}^{m_0+(p-1)} \rho_{W_{m(\omega)}} \times \hat{y}_{W_{m(\omega),CL,t} + \beta_{CL,wave} \\ MSE_{CL,wave} &= \sum_{t=1}^T \frac{(y_t - \hat{y}_{CL,wave,t})^2}{T} \end{aligned} \right. \end{aligned}$$

A previsão híbrida linearmente combinada para série temporal  $\tilde{y}$  (e, conseguintemente, para a série temporal original  $y$ , nos instantes de 1 até  $T+h$ ) é obtida pela substituição dos valores ótimos dos pesos adaptativos e da constante adaptativa aditiva, em (10).

## 7. EXPERIMENTO COMPUTACIONAL

Para ilustrar a metodologia híbrida proposta, foi modelada uma série temporal de vazão mensal média da usina de Itaipu cujo período é de janeiro de 1970 até dezembro de 2010 (ou seja, 492 observações). A sua escolha se deu por conta de sua relevância e por ser de difícil modelagem. Para fins de experimentação, 80 % dos primeiros dados foram utilizados no treinamento dos modelos individuais (de Box & Jenkins e de redes neurais artificiais), bem como para a obtenção dos parâmetros adaptativos ótimos das combinações de previsões; 10 %

dos dados subsequentes foram utilizados na amostra de validação; e 10 % dos remanescentes, na amostra de teste. Para tal, foram feitas previsões de curto prazo de um passo à frente, em um horizonte de previsão de 49 passos à frente. As estatísticas residuais utilizadas para avaliar o desempenho dos métodos preditivos foram o MAE (mean absolute error) e MAPE (mean absolute percentual error). Na Figura 1, tem-se o gráfico da série temporal supracitada, com todas as 492 observações.

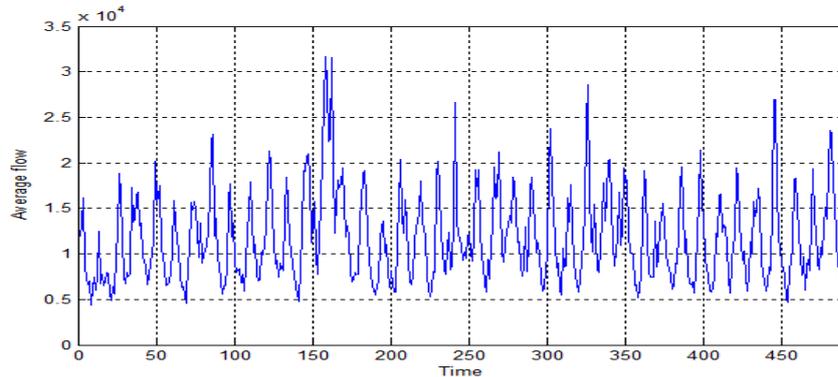


Figura 1 - Série temporal total de vazão média mensal da usina de Itaipu.

### 7.1 DECOMPOSIÇÃO WAVELET

Foi implementada a decomposição wavelet nível 2 (no software MATLAB 2013a). A escolha do nível 2 se deu para preservar a parcimônia no número de modelos individuais a serem estimados. Em relação à base ortonormal wavelet de Haar, a mesma foi utilizada por ter fornecido as melhores previsões.

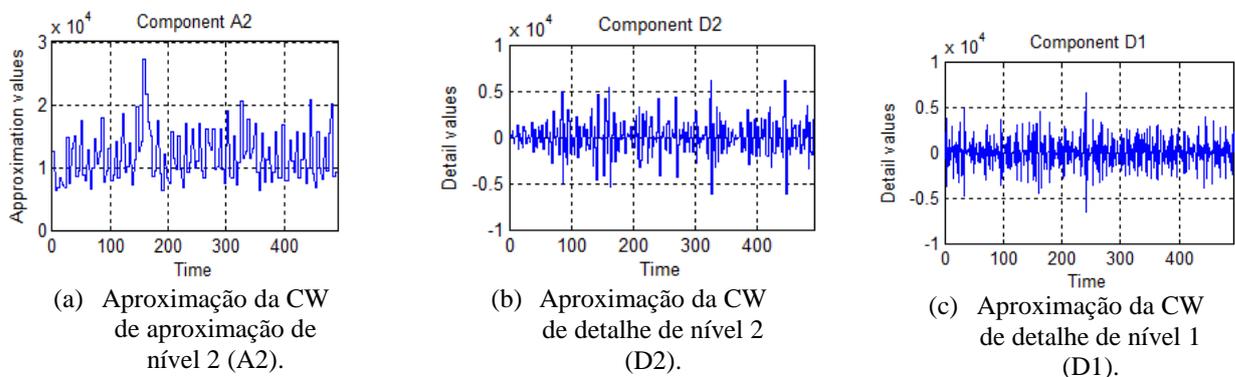


Figura 2 – Aproximação das CWs da série temporal total de vazão média mensal da Usina de Itaipu.

### 7.2 MODELAGEM

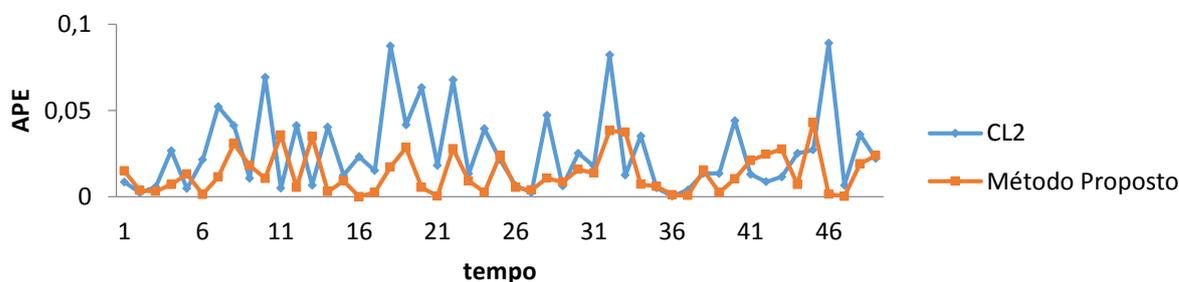
Para efeito de comparação, foi realizada a modelagem da série temporal supracitada utilizando os métodos de redes neurais artificiais (RNA) e de Box & Jenkins (BJ), com abordagem tradicional. Além disso, os métodos preditivos RNA e BJ foram utilizados de forma integrada com a decomposição wavelet de nível 2 (respectivamente, denotadas por RNA-WAVELET e BJ-WAVELET), [Donoho & Jonhstone (1994)]. Por fim, foram utilizadas as combinações lineares das previsões das abordagens BJ e RNA (CL1) e as dos RNA-WAVELET e BJ-WAVELET (CL2), conforme a abordagem de Flores & White (1998). O processo de ajuste dos pesos adaptativos de CL1 e CL2 ocorreu por meio de um PPM cuja função objetivo foi o mínimo do erro quadrado médio dos resíduos da amostra de treinamento (como é usual) e com os pesos convexos (isto é, normalizados e não negativos) [Faria & Mubwandarikwa, 2008].

Métodos	MAPE		Teste	MAE		Teste
	Treinamento	Validação		Treinamento	Validação	

RNA	16,35%	15,19%	21,78%	1.886,87	1.682,38	2.389,97
BJ	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
BJ-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

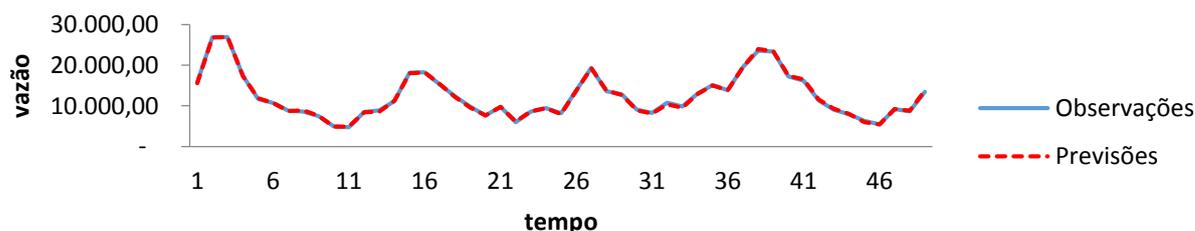
**Tabela 1** - Valores de MAPE e de MAE, nas amostras de treinamento, validação e de teste.

Na tabela 1, 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 3** – Evolução temporal dos valores de APE, na amostra de teste, da CL2 e do método proposto.

Na Figura 3, tem-se a comparação entre o método proposto e a abordagem CL2 (que foi a melhor na Tabela 1 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 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. Na Figura 4, por sua vez, tem-se sobrepostos a curva das observações e das previsões, na amostra de teste. Também é possível verificar a acurácia das previsões geradas pela combinação linear wavelet SARIMA-RNA com estágios múltiplos.



**Figura 4** - Comparação entre as observações e as previsões do método proposto, na amostra de teste.

## 8. CONCLUSÕES

Neste artigo, foi proposta uma nova abordagem híbrida de combinação de previsões, a qual utiliza, de forma integrada, os modelos de Box & Jenkins e de redes neurais artificiais, a Análise Wavelet e a Programação não-linear. De forma geral, o mesmo pode ser descrito em três fases básicas: (1) expande-se, via decomposição wavelet de nível  $p$ , uma série de tempo (a saber, vazão média mensal da Usina de Itaipu), gerando-se  $p+1$  aproximações das CWs (sendo uma CW de aproximação e  $p$  CWs de detalhe); (2) modelam-se, individualmente, as CWs, obtidas em (1), por meio de um modelo Box e Jenkins (para capturar estruturas lineares) e de uma rede neural artificial (para capturar estruturas não-lineares); e (3) combinam-se linearmente as previsões geradas, em (2), utilizando-se as  $p+2$  combinações lineares wavelet de previsões com ajuste numérico via programação não-linear, produzindo previsões combinadas híbridas para a série temporal a ser prevista. Para compará-lo, foram considerados os seguintes métodos de previsão: Box e Jenkins (BJ), redes neurais artificiais (RNA) e sua combinação linear tradicional (CL1); e rede neuronal artificial integrada com a decomposição wavelet (RNA-WAVELET), Box e Jenkins integrado com a decomposição wavelet (BJ-WAVELET) e sua combinação linear tradicional (CL2).

Analisando os resultados, tem-se que, na Tabela 1, o método proposto, em termos das estatísticas MAPE e MAE, foi superior a todos os outros, nas amostras de treino, validação e de teste - o que mostra mais poder de aprendizado (na amostra de treino) e de generalização (nas amostras de validação e de teste). Na Figura 3, por sua vez, obteve, em relação ao segundo melhor método da Tabela 1 (a saber, CL2), maior estabilidade temporal na evolução dos valores de APE (absolute percentual error). Além disso, dadas as curvas da Figura 4, tem-se que o método proposto gerou previsões que se correlacionam fortemente com os dados (evidenciando bom desempenho), ratificando a sua acurácia no experimento computacional. Em última análise, salienta-se que, embora o embasamento teórico da metodologia proposta, pelo menos em parte, seja fundamentado em conteúdos matemáticos de alta complexidade (como, por exemplo, a Teoria Wavelet), os softwares mencionados, no texto, viabilizam a sua utilização em aplicações reais, de maneira a ser operacionalizado de forma, em termos relativos, simples.

## REFERÊNCIAS

- [1]. AQUINO, R. R. B.; LIRA, M. M. S.; OLIVEIRA, J. B.; CARVALHO J.; M. A.; NETO, O. N.; ALMEIDA, G. J. (2009). Application of Wavelet and Neural Network Models for Wind Speed and Power Generation Forecasting in a Brazilian Experimental Wind Park, International Joint Conference on Neural Networks. Atlanta, USA.
- [2]. BATES, J. M. and GRANGER, C. W. J. (1969). Combination Forecasts. Operations Research Quarterly.
- [3]. BOX, G. E. P. and JENKINS, G. M. (1976). Time Series Analysis Forecasting and Control. Ed. San Francisco: Holden-Day.
- [4]. DAUBECHIES, I. (1988). Orthonormal Bases of Compactly Supported *wavelets*. Comm. Pure and Applied Math., 41, pp.909-996.
- [5]. DENG, F.; SU, G.; LIU, C.; WANG, Z. (2010). Global Solar Radiation Modeling Using The Artificial Neural Network Technique. Power and Energy Engineering Conference, Chengdu, March.
- [6]. DONOHO, D. L. and JONHSTONE, I. M. (1994). Ideal Spatial Adaptation by Wavelet Shrinkage. Biometric, 81,425-455.
- [7]. DONOHO, D. L.; JONHSTONE, I. M.; KERKYACHARIAN, G.; and PICARD, D. (1995). Wavelet Shrinkage: Asymptopia? (with discussion) Journal Royal Statistical Society. B, 57:301-369.

- [8]. FARIA, A. E. and MUBWANDARIKWA, E (2008). Multimodality on the Geometric Combination of Bayesian Forecasting Models. *International Journal of Statistics and Management System*, 3, 1-25
- [9]. FLORES, B. E. and WHITE, E. M. (1989). Subjective versus Objective Combining of Forecasts: An Experiment. *Journal of Forecasting*, v.8, p.331-341.
- [10]. FLORES, B. E.; and WHITE, E. M. (1998). A Framework for the Combination of Forecasts. *Journal Academic Marketing Science*, v.16 (3-4), p.95-103.
- [11]. GOLYANDINA N.; NETRUTKIN V.; ZHIGLJAVSKY A. (2001). *Analysis of Time Series Structure: SSA and Related Techniques*. Chapman & Hall / CRC. London.
- [12]. GRANGER, C. W. J. (1989). Invited Review: Combining Forecasts - Twenty Years Later. *Journal of Forecasting*, 8, pp. 167-173.
- [13]. GUPTA, S. and WILTON, P. C. (1987). Combination of Forecasts: An Extension. *Management Science*. v.33, n.3, p.356-372.
- [14]. HAMILTON, J. (1994). *Time Series Analysis*. Princeton University Press.
- [15]. HAYKIN, S. (2001). *Redes Neurais: Princípios e Prática*. 2º ed. Porto Alegre, Bookman.
- [16]. KUBRUSLY, C. S. (2011). *Elements of Operator Theory*. Second Edition. Birkhäuser. Boston.
- [17]. KUBRUSLY, C. S. and LEVAN, N. (2006). Abstract *Wavelets* Generated by Hilbert Space Shift Operators. *Adv. Math. Sci. Appl.* 14, 643-660.
- [18]. LEI, C. and RAN, L. (2008). Short-term Wind Speed Forecasting Model for Wind Farm Based on *wavelet* Decomposition DRPT, Nanjing, China, pp 2525-2529
- [19]. LEVAN N., KUBRUSLY, C. S, (2003). A *Wavelet* “Time-Shift-Detail” Decomposition.
- [20]. LUTKEPOHL, H. (2006). *New Introduction to Multiple Time Series Analysis*. Springer.
- [21]. MALLAT S. (2009). *A Wavelet Tour of Signal Processing*. Academic Press. San Diego.
- [22]. MAKRIDAKIS, S. (1989). Why Combining Works? *International Journal of Forecasting*, v.5, p.601-603.
- [23]. MAKRIDAKIS, S. and WINKLER, R. L. (1983). Averages of Forecasts: Some Empirical Results. *Management Science*, v.29, n.9, p.987-996.
- [24]. MORETTIN, P. A. and TOLOI, L. M. C. (2006). *Análise Séries Temporais*. 2ª Ed. ABE. Projeto Fisher. Editora: Edgard Blucher.
- [25]. NEWBOLD, P. and GRANGER, C.W. J. (1974). Experience with Forecasting Univariate Times Series and the Combination of Forecasts. *Journal of the Royal Statistical Society*. V.137, pp.131-165.
- [26]. OGDEN, R. T. (1997). *Essential Wavelets for Statistical Applications and Data Analysis*. Birkhäuser, Boston.
- [27]. PALIT A. K., POPOVIC D. (2005). *Computational Intelligence in Time Series Forecasting: Theory and Engineering Applications*. Springer-Verlag, London.
- [28]. RAGSDALE, C. (2004). *Spreadsheet Modeling & Decision Analysis: A Practical Introduction to Management Science*, fourth edition, South-Western.
- [29]. RUSSELL, S.J. and NORVIG, P. (1995). *Artificial Intelligence: A Modern Approach*. New Jersey: Prentice-Hall Inc.
- [30]. TAFNER, M. A. (1996). *Redes Neurais Artificiais: Introdução e Princípios de Neurocomputação*. - Blumenau: EKO.
- [31]. WALLIS, K. F. (2011). Combining Forecasts - Forty Years Later. *Applied Financial Economics*, 21, pp. 33-41.
- [32]. WINKLER, R. L. (1989). Combining Forecasts: A Philosophical Basis and Some Current Issues. *International Journal of Forecasting*, v.5, p. 605-609.
- [33]. WINKLER, R. L. and MAKRIDAKIS, S. (1983). The Combination of Forecasting. *Journal of the Royal Statistical Society, Series A*, v.146, 1983, p.150-157.

- [34]. ZERVAS, P.L.; SARIMVEIS, H.; PALYVOS, J.A.; MARKATOS, N.C.G. (2008). Prediction of Daily Global Solar Irradiance on Horizontal Surfaces Based on Neural-Network Techniques. *Renewable Energy*, 33, pp 1796-1803.
- [35]. ZOU, H. & YANG, Y. (2004). Combining Time Series Models for Forecasting. *International Journal of Forecasting*, v.20, n.1, p.69-84.
- [36]. ZSOLT, L. (2006). *Redes Neurais Artificiais: Fundamentos e Aplicações*. 4º Ed, Livraria de Física.