

PREVISÕES DE VAZÕES MENSAIS VIA COMBINAÇÃO HÍBRIDA ARIMA_NEURAL COM ENCOLHIMENTO E DECOMPOSIÇÃO WAVELET

Levi Lopes Teixeira (UTFPR) prof.levilopes@gmail.com>
Paulo Henrique Siqueira (UFPR) paulohscwb@gmail.com>
Luiz Albino Teixeira Jr (UNILA) luiz.a.t.junior@gmail.com>

Resumo: Neste artigo foram realizadas previsões da série de vazões mensais no posto 266 (Itaipu) localizado no rio Paraná, Brasil. Para tanto, foram usados os métodos de Box & Jenkins, Redes Neurais Artificiais (RNA), Wavelet_RNA e Wavelet_ARIMA. As previsões obtidas com esses métodos foram comparadas às fornecidas pelo método proposto neste artigo, formado por uma combinação híbrida ARIMA_Neural com encolhimento e decomposição wavelet. Na amostra de teste, o erro médio percentual absoluto (MAPE) obtido com o método proposto foi de 1,22%. Com os métodos individuais de Box & Jenkins e RNA ele ficou em 18,01% e 20,12%, respectivamente. Os métodos compostos Wavelet_RNA e Wavelet_ARIMA forneceram previsões melhores em relação aos individuais ARIMA e RNA, sendo 2,8% o MAPE das previsões provenientes do Wavelet_RNA e 3,63% o erro das previsões obtidas através do Wavelet_ARIMA.

Palavras-chave: Wavelet. ARIMA. Redes Neurais Artificiais. Séries Temporais.

MONTHLY FORECASTS OF FLOW VIA COMBINATION WITH HYBRID ARIMA_NEURAL SHRINKING AND WAVELET DECOMPOSITION

Abstract: In this paper were conducted forecasts of series of monthly flows in 266 post (Itaipu) located in Paraná river, Brazil. Therefore, the methods of Box & Jenkins, Artificial Neural Networks (ANN), and Wavelet_ANN Wavelet_ARIMA were used. The estimates obtained using these methods were compared with those provided by the method proposed in this article, formed by a hybrid combination with ARIMA_Neural shrinkage and wavelet decomposition. In the test sample, the mean absolute percentage error (MAPE) obtained with the proposed method was 1.22%. With individual methods of Box & Jenkins RNA and it was in 18.01% and 20.12%, respectively. Compounds methods Wavelet_ANN and Wavelet_ARIMA provided better predictions compared to individual ARIMA and ANN, being 2.8% the MAPE from Wavelet_ANN forecasts and 3.63% the error of the predictions obtained by Wavelet_ARIMA.

Keywords: Wavelet: ARIMA. Artificial Neural Networks. Time Series.

1. INTRODUÇÃO

Estimativas da Empresa de Pesquisa Energética (EPE) apontam que o consumo de energia elétrica no Brasil em 2020 será 659.092 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 afluente) disponível para a movimentação das turbinas geradoras e longos períodos de seca podem levar a uma diminuição da capacidade geradora de energia da usina. O Operador Nacional do Sistema Elétrico (ONS) faz previsão de vazões naturais para os locais de aproveitamento hidrelétricos que constituem o Sistema Interligado Nacional (SIN), sendo denominado PREVIVAZ o modelo de previsão mais usado.

Entre os métodos usados pelo ONS, ressalta-se a frequência de uso dos métodos de Box e Jenkins (Box *et al.*, 1994), eficientes para séries oriundas de processos lineares e com estruturas matemáticas relativamente simples. Segundo WANG *et al.*,(2006a), os processos de geração de vazões são sazonal e não linear. Os mecanismos não lineares são mais fortes em

escalas menores de tempo, sendo as vazões diárias e mensais não lineares (Wang, *et al.*, 2006b). Neste caso, os métodos de Box e Jenkins não são recomendados, sendo indicado outro tipo de modelador, como as redes neurais artificiais (RNA), por exemplo.

Ao longo dos anos, diversos métodos de previsão (ou preditivos) de séries de tempo têm sido propostos com o objetivo de se fazê-la, com o maior acurácia possível. Em termos gerais, podem-se agrupá-los em duas classes básicas: os métodos de previsão base (os de abordagem estatística e os de abordagem de inteligência artificial) e os métodos de combinação de métodos preditivos base (BATES & GRANGER, 1969; NEWBOLD & GRANGER, 1974; MAKRIDAKIS & WINKLER, 1983; GUPTA & WILTON, 1987; MAKRIDAKIS, 1989; WINKLER, 1989). O termo “combinação de métodos preditivos (base)” pode ser utilizado, em sentido amplo, para se referir tanto à combinação de previsões (MAKRIDAKIS & WINKLER, 1983) e quanto à combinação de densidades preditivas bayesianas (FARIA & MUBWANDARIKWA, 2008). Neste artigo, porém, é tratado apenas em sentido estrito de combinação de previsões.

Na literatura, existe um grande número de métodos de previsão base consagrados como, por exemplo, os modelos de Box & Jenkins (HAMILTON, 1994) (que é uma abordagem estatística) e os modelos de redes neurais artificiais (HAYKIN, 2001) (que é uma abordagem de inteligência artificial). De acordo com Zhang (2003), os modelos de Box & Jenkins são eficientes, em termos de modelagem, no mapeamento de estruturas de autodependência linear das séries temporais; enquanto que as redes neurais artificiais os são no mapeamento de estruturas de autodependência não linear. Assim o sendo, uma alternativa plausível é a combinação das previsões geradas por ambos os métodos preditivos, gerando previsões híbridas. Ainda de acordo com o autor, isso é plausível também pelo fato de que as séries temporais, em sua grande maioria, apresentam, simultaneamente, estruturas de autodependência linear e não linear. Com efeito, a utilização de um método de previsão base pode acarretar a perda de informações relevantes, para fins de previsibilidade temporal. Outro problema na utilização de apenas um método preditivo base reside ainda no fato de que, muitas vezes, os processos estocásticos associados às séries temporais são não estacionários (MORETTIN & TOLOI, 2006) ou excessivamente ruidosos (TEIXEIRA JR, 2013), de maneira que uma abordagem mais segura e, de acordo com Wallis (2011), mais eficiente, se implementada de forma adequada, é a combinação de previsões oriundas dos métodos de previsão base.

Por sua vez, da Teoria Wavelet, advém importantes métodos auxiliares de pré-processamento de uma série de tempo (como, por exemplo, o encolhimento e decomposição wavelet da série), antes de sua efetiva modelagem. Desde que sejam utilizados de forma adequada, é possível se obter ganhos preditivos substanciais. Pode-se verificar tal fato, dentre muitas outras referências, em: Donoho & Jonhstone (1994), Donoho *et al* (1995), Lei & Ran (2008), Teixeira Jr *et al* (2012).

Dado o exposto, é proposto, neste artigo, um método de combinação para a geração de previsões híbridas para séries temporais de vazão de afluentes, o qual utiliza, de forma integrada: o encolhimento e a decomposição wavelet, os modelos de Box & Jenkins e as redes neurais artificiais. De forma geral, o método híbrido proposto pode ser descrito em cinco passos gerais: **passo 1:** escolha de base e nível de decomposição para então submeter a série ao encolhimento wavelet; **passo 2:** decomposição wavelet de nível p da série de tempo obtida no passo 1 com o encolhimento wavelet, de forma a gerar $p+1$ componentes ortogonais wavelet; **passo 3:** modelagem de cada componente ortogonal wavelet, do passo 1, por meio de um modelo Box e Jenkins (para mapear informações de autodependência linear) e redes neurais artificiais (para mapear informações de autodependência não linear); **passo 4:** geração

das previsões híbridas linearmente combinadas para as componentes ortogonais wavelet; **passo 5:** a previsão da série observada é obtida com a soma das previsões obtidas no passo 3 com a previsão do resíduo proveniente do encolhimento wavelet.

2. DECOMPOSIÇÃO WAVELET

Enquanto na análise de Fourier o sinal é decomposto em ondas senoidais de várias frequências, na análise wavelet o sinal é decomposto a partir de informações de escala (compressão e dilatação) e translação de uma função wavelet $\omega_{m,n}(\cdot) = 2^{\frac{m}{2}} \omega(2^m(\cdot) - n)$, onde $m, n \in \mathbb{Z}$. Segundo Levan e Kubrusly (2003), o parâmetro m é chamado de parâmetro de escala e n , de parâmetro de translação. Ao contrário das funções trigonométricas seno e cosseno que possuem energia infinita e são usadas na decomposição de Fourier, as funções wavelet são de suporte compacto, isto é, possuem domínio diferente de zero em uma extensão finita e igual a zero em todo o resto, são funções com energia finita e quadraticamente somáveis. A função $\omega_{m,n}$ faz o mapeamento das componentes de detalhes (alta frequência) do sinal. A dita componente de aproximação (baixa frequência) é obtida a partir da função escala $\phi_{m,n}(\cdot) = 2^{\frac{m}{2}} \phi(2^m(\cdot) - n)$.

Entre as famílias de funções wavelet a mais antiga é a de Haar, definida em (1).

$$\omega(x) = \begin{cases} 1, & \text{se } x \in [0, 1/2) \\ -1, & \text{se } x \in [1/2, 1) \\ 0, & \text{caso contrário} \end{cases} \quad (1)$$

Em (2) encontra-se a definição da função escala de Haar.

$$\phi(x) = \begin{cases} 1, & \text{se } x \in [0, 1) \\ 0, & \text{caso contrário} \end{cases} \quad (2)$$

Na Figura 1 estão representados os gráficos das funções escala e wavelet de Haar, respectivamente.

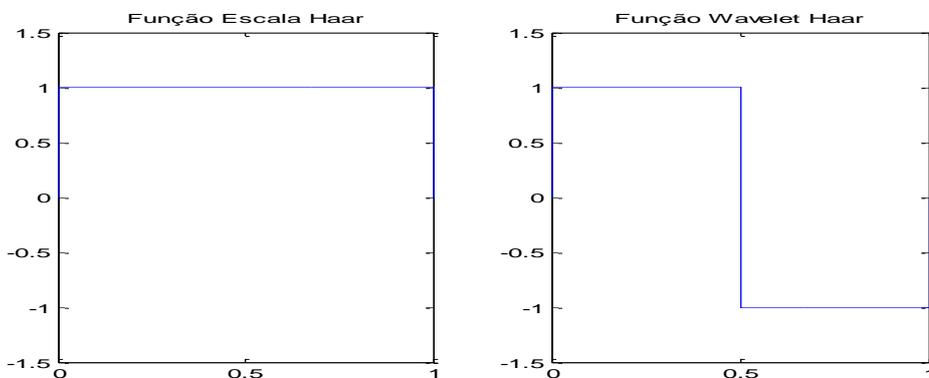


Figura 1 – Gráficos das funções escala e wavelet de Haar.

Baseado em Kubrusly e Levan (2006) e Mallat (1998), o conjunto formado pela união de $\{\phi_{m_0,n}\}_{n \in \mathbb{Z}}$ e $\{\omega_{m,n}\}_{m \geq m_0}$ é uma base ortonormal wavelet do espaço de Hilbert l^2 . Segue que, a decomposição wavelet da função $f(.) \in l^2$ pode ser escrita segundo a expansão (3).

$$f(.) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n} + \sum_{m=m_0}^{\infty} \left(\sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n} \right) \tag{3}$$

Onde $a_{m_0,n}$ é chamado de coeficiente de aproximação e é calculado a partir do produto interno entre $f(.)$ e $\phi_{m_0,n}$. Os coeficientes de detalhes são representados por $d_{m,n}$ e obtidos através do produto interno entre $f(.)$ e $\omega_{m,n}$. Em outras palavras, $f(.)$ é uma combinação linear das funções $\phi_{m_0,n}$ e $\omega_{m,n}$.

3. ENCOLHIMENTO WAVELET

No caso da série temporal ser fortemente ruidosa, recomenda-se a filtragem da mesma, para então modelar a parte determinística obtida com a filtragem. Seja $(y_t)_{t=1}^T$ uma série temporal, onde para todo t : $y_t = f_t + e_t$. Sendo e_t um estado da variável aleatória $\varepsilon_t \sim (0, \sigma^2)$, onde f_t é a componente determinística de y_t . Como não é possível modelar e_t e sendo a série excessivamente ruidosa, indica-se o tratamento dos dados formadores da série temporal a fim de se obter a componente determinística. O encolhimento (*shrinkage*) wavelet é uma das técnicas destinada para este fim.

O Princípio do encolhimento wavelet tem como objetivo a redução (ou mesmo remoção) do ruído presente num sinal, diminuindo (ou zerando) a magnitude dos coeficientes wavelet (MORETTIN, 1999). Outras referências nesta área são os trabalhos de Donoho e Johnstone e Danoho *et al.*

As três etapas do encolhimento wavelet são: 1) decomposição wavelet do sinal ruidoso; 2) selecionar limiares (*thresholds*) que eliminam (ou reduzem) os ruídos; 3) aplicar a transformada wavelet inversa a fim de se obter o sinal menos ruidoso.

Das várias regras de limiarização encontradas na literatura, as mais tradicionais são o limiar duro (*hard threshold*) e limiar suave (*soft threshold*), representadas em (4) e (5), respectivamente.

$$d_{m,n}^H = \begin{cases} d_{m,n}, & \text{se } |d_{m,n}| > \lambda \\ 0, & \text{se } |d_{m,n}| \leq \lambda \end{cases} \tag{4}$$

$$d_{m,n}^S = \begin{cases} \text{sign}(d_{m,n})(|d_{m,n}| - \lambda), & \text{se } |d_{m,n}| > \lambda \\ 0, & \text{se } |d_{m,n}| \leq \lambda \end{cases} \tag{5}$$

Onde $d_{m,n}$ é o coeficiente de detalhe com ruído, $d_{m,n}^H$ e $d_{m,n}^S$ são coeficientes de detalhes sem ruído, $\text{sign}(d_{m,n})$ é o sinal de $d_{m,n}$ e λ o limiar. O *hard threshold* gera uma descontinuidade, um “salto” em λ , eliminando ou mantendo o coeficiente de detalhe. Enquanto o limiar suave

atenua por $(|d_{m,n}| - \lambda)$ os valores de $d_{m,n}$ (se $|d_{m,n}| > \lambda$), eliminando-os em caso contrário. Nos gráficos das Figuras 3 e 4 podem ser observadas essas características.

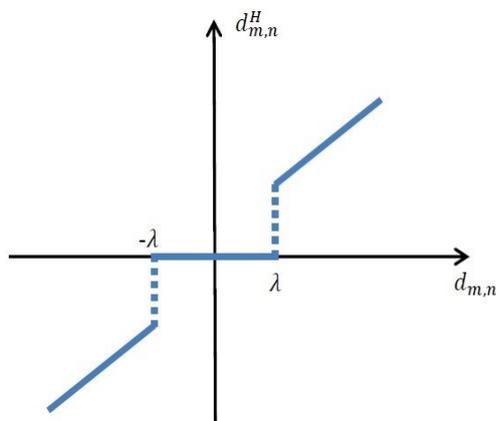


Figura 3 – Gráfico da função limiar duro.

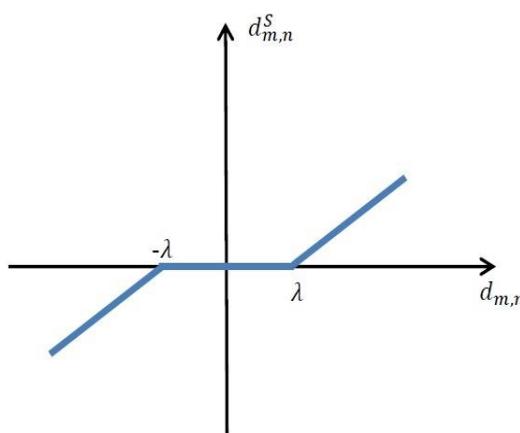


Figura 4 – Gráfico da função limiar suave.

Neste artigo o cálculo do limiar λ foi realizado a partir de duas propostas: universal e SURE (*Estimation Stein's Unbiased Risk Estimate*). Maiores detalhes dessas técnicas podem ser encontradas em Donoho e Johnstone (1995), Johnstone e Silverman (1996) e Morettin (1999).

De acordo com Donoho e Johnstone (1998) o cálculo do limiar universal é dado por (6).

$$\lambda = \sigma \sqrt{2 \log(T)} \tag{6}$$

Sendo σ o nível do ruído e T a cardinalidade da série. O limiar universal é independente da escala m , assumindo um valor global.

Donoho e Johnstone (1995) propuseram selecionar um limiar, pela minimização do estimador não-viesado do risco Stein (SURE), em cada escala m . Seja N_m o número de coeficientes na escala m , então o limiar de SURE é calculado a partir de (7).

$$\lambda_m = \operatorname{argmin}_{0 \leq \lambda \leq \sqrt{2 \log(N_m)}} \operatorname{SURE}(y_m, \lambda) \quad (7)$$

4. REDES NEURAIIS ARTIFICIAIS

O mais eficiente sistema de aprendizado existente na natureza é o cérebro humano, este é constituído por neurônios, que são células capazes de armazenar e processar informações. Os neurônios estão ligados uns aos outros formando uma rede neural e o fluxo das informações (sinais) por esta rede dependerá de uma fenda existente entre os neurônios denominada de fenda sináptica, ou sinapse. Segundo Castro (2010), depois que todos os sinais de entrada são recebidos e ponderados pelas sinapses eles são somados no corpo celular, gerando um sinal de entrada líquida do neurônio, e que será posteriormente propagado para a saída do neurônio.

As Redes Neurais Artificiais (RNA) foram idealizadas com base nos processos de aprendizagem do cérebro humano, de forma que a RNA é capaz de simular as conexões sinápticas. De acordo com Haykin (2001) a RNA é um sistema de processamento paralelo formado por unidades capazes de armazenar conhecimento e disponibilizá-lo para o futuro, sendo formada basicamente por neurônios artificiais distribuídos em camadas. Entre as RNAs, um dos tipos mais usados é a multilayer feedforward que é composta por várias camadas e os sinais são propagados apenas da entrada para a saída, nunca ao contrário.

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. 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) (TEIXEIRA Jr *et al.*, 2012).

O problema de previsão de séries temporais com T observações $\{y_t\}_{t=1}^T$ através de RNA consiste na utilização de dados observados para se prever $y(t+k)$, onde k é o passo de tempo de previsão para o futuro. Define-se o tamanho da janela (n), o valor de k e montam-se as sequências $\{y(t), y(t-1), \dots, y(t-n+1), y(t+1), \dots, y(t+k)\}$ que formaram um conjunto de padrões de treinamento onde $\{y(t), y(t-1), \dots, y(t-n+1)\}$ e $\{y(t+1), \dots, y(t+k)\}$ representam as entradas e saídas, respectivamente.

5. MÉTODO BOX & JENKINS

Box e Jenkins (1970) apresentaram metodologia que fez a integração de conhecimentos existentes na época. A metodologia consiste em ajustar modelos autorregressivos integrados de médias móveis (ARIMA) a um conjunto de dados a partir de três estágios: identificação do modelo, estimação dos parâmetros e verificação do modelo ajustado. Segundo Box e Jenkins (1970) um modelo ARMA plausível para a série temporal $(y_t)_{t=1}^T$, T= cardinalidade, de ordens p (autorregressivo) e q (médias móveis) é descrito pela equação (8).

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q} + a_t \quad (8)$$

O modelo (8) combina valores passados das entradas y_t e choques aleatórios (a_t) descorrelacionados, de média zero e variância constante. Neste modelo $\phi_i, \theta_j \in \mathbb{R}$, com $i = 1, \dots, p$ e $j = 1, \dots, q$, denotam os parâmetros do modelo e δ uma constante. Na hipótese da série temporal ser não estacionária, esta deve ser diferenciada e o modelo ARMA(p,q) substituído pelo ARIMA(p,d,q), sendo d a ordem de diferenciação da série. Para a identificação do modelo, as ordens p e q podem ser determinadas por meio da análise do perfil dos gráficos das funções de autocorrelação (FAC) e autocorrelação parcial (FACP). Identificado o modelo, passa-se ao estágio seguinte que é a estimação dos parâmetros. Para tanto, é necessário utilizar métodos iterativos não lineares de mínimos quadrados, maiores detalhes podem ser encontrados em Box e Jenkins (1970) e Morettin e Toloi (2006). Para a validação do modelo já com os parâmetros estimados pode-se usar testes estatísticos, tais como: teste de Box-Pierce, teste do periodograma acumulado, teste da autocorrelação cruzada, entre outros. No caso da série temporal $(y_t)_{t=1}^T$ apresentar componente sazonal, o modelo de Box & Jenkins plausível é dado genericamente, em (9).

$$\phi(B)(1 - \phi_1 B - \dots - \phi_p B^p) \nabla^d (1 - B^S)^D y_t = \theta(B)(1 - \theta_1 B - \dots - \theta_q B^q) a_t \quad (9)$$

Onde: $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$, $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$, d é a ordem das diferenças simples, D é a ordem das diferenças sazonais, S é o período sazonal, $\phi_k \in \mathbb{R}$ e $\theta_j \in \mathbb{R}$ são os coeficientes dos polinômios não sazonais e $\phi_m \in \mathbb{R}$ e $\theta_n \in \mathbb{R}$ são os coeficientes dos polinômios sazonais.

6. COMBINAÇÃO DE PREVISÕES

Os precursores na área das combinações de previsões são Bates e Granger (1969), que sugeriram uma combinação, expressa em (10), para a previsão da série temporal $\{y_t\}_{t=1}^T$.

$$\hat{y}_t = \rho_1 \times \hat{y}_{1,t} + \rho_2 \times \hat{y}_{2,t} \quad (10)$$

Em (10), a previsão combinada linearmente é representada por \hat{y}_t , enquanto as previsões obtidas através dos métodos 1 e 2 são $\hat{y}_{1,t}$ e $\hat{y}_{2,t}$, respectivamente. Sendo os coeficientes ρ_1 e ρ_2 calculados a partir da matriz de covariância de erros de previsão.

Gupta e Wilton (1977) usaram em seus trabalhos a média aritmética das previsões individuais. Granger e Ramanathan (1984) propuseram, entre outros métodos, uma combinação linear sem restrição para os coeficientes e adição de uma constante. Terui e Dijk (2002) sugerem o uso de filtros de Kalman para o cálculo dos coeficientes e que estes sejam variáveis ao longo do tempo.

Teixeira Jr (2013) utilizou programação matemática para calcular os coeficientes da combinação linear. O autor otimizou um problema de programação não linear, onde a função objetivo era constituída pela soma dos erros de previsão ao quadrado.

Faria e Mubwandarikwa (2008) salientam que a previsão combinada é uma agregadora de informações oriundas de diferentes fontes (no caso, os métodos preditivos base) sobre a flutuação estocástica da série temporal a ser modelada. Em Wallis (2011), é feita uma revisão histórica, na qual são mencionados muitos dos principais artigos sobre combinação de métodos preditivos individuais, publicados em quarenta e dois anos, os quais atestam a sua utilização e a sua eficiência, além de mostrarem uma enorme diversidade de aplicações e abordagens.

7. MATERIAIS E MÉTODOS

7.1 Materiais

A série de vazões mensais $(y_t)_{t=1}^{492}$ usada nesta seção envolve quarenta e um anos de observações no período 1970-2010 e pode ser obtida no site do ONS, com endereço eletrônico: http://www.ons.org.br/operacao/vazoes_naturais.aspx. Os dados que compõem a série são provenientes das medições de vazões no posto 266, nele são obtidas as vazões afluentes à usina hidrelétrica de Itaipu, que está localizada 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 Guaira (Brasil) e Salto del Guairá (Paraguai), ao norte.

7.2 Método Proposto

Seja a série temporal $\{y_t\}_{t=1}^T = \{f_t\}_{t=1}^T + \{\varepsilon_t\}_{t=1}^T$, onde $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. Inicialmente, submete-se a série $\{y_t\}_{t=1}^T$ ao encolhimento wavelet, obtendo-se a série $\{y_{t,EW}\}_{t=1}^T \cong \{f_t\}_{t=1}^T$. Os testes de independência estatística BDS e de estacionariedade *Dickey-Fuller* são aplicados à componente total de resíduo $(r = \{y_t\}_{t=1}^T - \{y_{t,EW}\}_{t=1}^T)$ a fim de se obter indicação que a mesma pode ser classificada estatisticamente como ruído branco. No caso de r não ser classificado como ruído, a sua previsão (obtida por RNA) irá compor a previsão da série observada.

Considere a sequência infinita $\{\tilde{y}_t\}_{t \in \mathbb{Z}}$ obtida da série temporal $\{y_{t,EW}\}_{t=1}^T$ a partir da igualdade $\tilde{y}_t = \begin{cases} y_{t,EW} & t \in \{1, \dots, T\} \\ 0 & t \in \mathbb{Z} \setminus \{1, \dots, T\} \end{cases}$. Segundo Teixeira Jr *et al.*, (2012) a decomposição wavelet de nível p de uma série temporal \tilde{y} em relação a uma determinada base ortonormal wavelet é representada por (11).

$$\tilde{y} \cong \tilde{\tilde{y}} = \sum_{n \in \Gamma_{V_{m_0}}} a_{m_0,n} \phi_{m_0,n} + \sum_{m=m_0}^{m_0+(p-1)} \left(\sum_{n \in \Gamma_{W_m}} d_{m,n} \omega_{m,n} \right) \quad (11)$$

Onde $\Gamma_{V_{m_0}}$ e Γ_{W_m} são subconjuntos de índices em \mathbb{Z} . A determinação de tais conjuntos ocorre através da minimização do erro quadrático médio (MSE) entre \tilde{y} e $\tilde{\tilde{y}}$.

O encolhimento wavelet, as componentes de aproximação $\sum_{n \in \Gamma_{V_{m_0}}} a_{m_0,n} \phi_{m_0,n}$ e detalhe $\sum_{n \in \Gamma_{W_m}} d_{m,n} \omega_{m,n}$ foram obtidas a partir do aplicativo computacional Matlab 8.0, que oferece várias opções de bases wavelets, tais como: as famílias daubechies (dbN), coiflets (coifN) e symlets (symN), desenvolvidas por Ingrid Daubechies (DAUBECHIES, 1992).

Neste trabalho, a série $y_{t,EW}$ foi decomposta ortogonalmente via wavelet de nível $p = 2$, que resulta em duas componentes de detalhe (D_1 e D_2) e uma de aproximação (A_2), utilizando base da família Daubechies com momento igual a 10 (db 10).

Uma vez obtidas as componentes wavelet de aproximação e detalhe, estas foram modeladas via três redes neurais artificiais *feedforward multilayer perceptron* (RNA-MLP) (abordagem não linear) e três modelos ARIMA (abordagem linear). As Redes Neurais foram construídas e treinadas no aplicativo computacional Matlab 8.0 e os parâmetros dos modelos ARIMA obtidos a partir *software EViews 8*.

Da modelagem por RNA e ARIMA resultaram as previsões das componentes wavelet de detalhe ($\hat{D}_{1,RNA}, \hat{D}_{1,ARIMA}, \hat{D}_{2,RNA}$ e $\hat{D}_{2,ARIMA}$) e aproximação ($\hat{A}_{2,RNA}$ e $\hat{A}_{2,ARIMA}$).

Tendo as previsões das componentes wavelet por ARIMA e RNA, fizeram-se três combinações lineares, cada uma envolvendo os pares de previsões ($\hat{A}_{2,RNA}$ e $\hat{A}_{2,ARIMA}$), ($\hat{D}_{1,RNA}$ e $\hat{D}_{1,ARIMA}$) e ($\hat{D}_{2,RNA}$ e $\hat{D}_{2,ARIMA}$). A combinação linear para a componente da aproximação ($\hat{A}_{2,CL}$) está descrita em (12).

$$\hat{A}_{2,CL} = \alpha_{A2,ARIMA} \times \hat{A}_{2,ARIMA} + \alpha_{A2,RNA} \times \hat{A}_{2,RNA} \tag{12}$$

onde $\alpha_{A2,ARIMA}$ e $\alpha_{A2,RNA}$ são parâmetros a serem determinados com a resolução do problema de programação não linear (13), cuja função objetivo é a raiz do erro quadrático médio (RMSE) entre a componente de aproximação $A_2(t)$ e a sua previsão combinada $\hat{A}_{2,CL}(t)$. Da mesma forma procede-se para obter as previsões $\hat{D}_{1,CL}$ e $\hat{D}_{2,CL}$,

$$\text{Min } RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (\hat{A}_{2,CL}(t) - A_2(t))^2} \tag{13}$$

s. a. $\alpha_{A2,ARIMA}$ e $\alpha_{A2,RNA}$ são irrestritas

A previsão da série observada é dada pela soma das previsões das componentes wavelet e previsão do resíduo (\hat{r}), como mostra (14).

$$\{\hat{y}_t\}_{t=1}^T = \{\hat{A}_{2,CL}(t)\}_{t=1}^T + \{\hat{D}_{1,CL}(t)\}_{t=1}^T + \{\hat{D}_{2,CL}(t)\}_{t=1}^T + \hat{r} \tag{14}$$

Para a avaliação dos métodos foram utilizados os Erros MAPE (Erro Médio Percentual Absoluto) e MAE (Erro Médio absoluto), representados em (15) e (16), respectivamente.

$$MAPE = \frac{1}{T} \sum_{t=t_0}^T \left| \frac{y_t - \hat{y}_t}{y_t} \right| \tag{15}$$

$$MAE = \frac{1}{T} \sum_{t=t_0}^T |y_t - \hat{y}_t| \tag{16}$$

Na Figura 5 está representado o fluxograma com as etapas do método proposto para o caso de uma decomposição wavelet de nível 2.

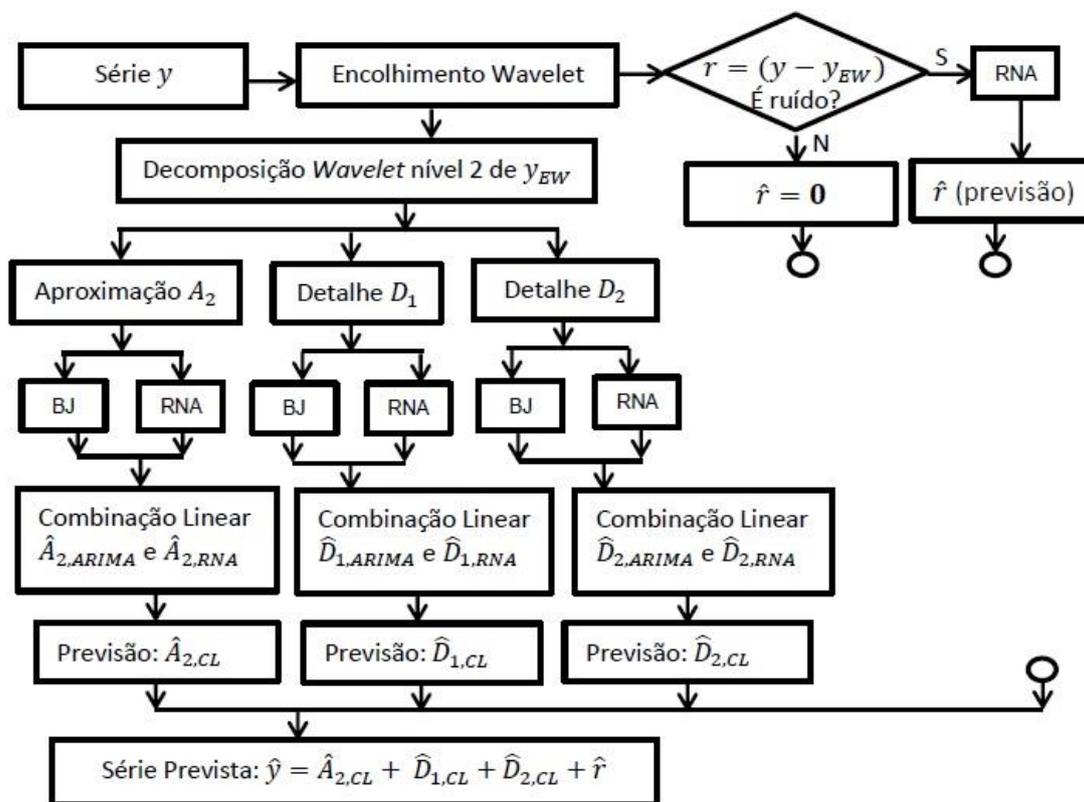


Figura 5- Fluxograma com as etapas do método proposto.

8. RESULTADOS OBTIDOS

8.1 Métodos Individuais Box & Jenkins e RNA

Aplicando-se a metodologia Box e Jenkins na análise da série de vazões mensais, chegou-se a conclusão que o modelo $ARIMA(6,0,6) \times (1,1,2)_{12}$, entre os analisados, é o que melhor se ajusta aos dados da série em estudo. O modelo foi validado a partir da análise de seus resíduos mediante a construção dos correlogramas da FAC e FACP.

A RNA usada na modelagem da série de vazões é do tipo *feedforward multilayer perceptron* com tamanho da janela de entrada igual a cinco, uma camada escondida composta por cinco neurônios artificiais com função de ativação tangente hiperbólica sigmoideal (*tansig*) e um neurônio na camada de saída com função de ativação linear (*purelin*).

A amostra de teste, formada pelas últimas 49 observações, apresentou MAPE igual a 18,01% com o modelo ARIMA e 20,12% com a modelagem por RNA. Nas Figuras 6a, 6b estão os gráficos de dispersão da amostra de teste versus previsão por RNA e ARIMA, respectivamente. Os gráficos em questão são um indicativo da baixa qualidade das previsões obtidas com os métodos individuais ARIMA e RNA. Na Figura 6a o coeficiente de correlação encontrado foi de 0,84 e na Figura 6b obteve-se coeficiente igual a 0,87.

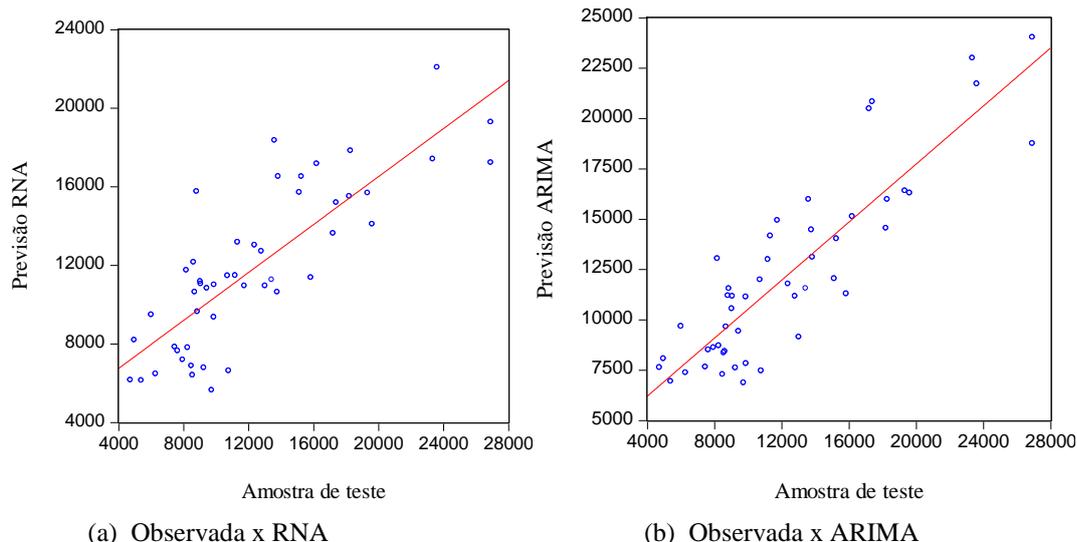


Figura 6 - Diagramas de dispersão entre a amostra de teste e previsões por RNA e ARIMA.

8.2 Métodos Compostos: Wavelet_RNA e Wavelet_ARIMA

Inicialmente, a série de vazões foi decomposta ortogonalmente via wavelet de nível $p = 2$, utilizando a base daubechies com momento 10 (db 10). As três componentes (duas de detalhes e uma de aproximação) obtidas com a decomposição foram ajustadas por três RNA e três modelos ARIMA e as previsões combinadas linearmente a fim de se obter a série observada.

Na Figura 7 podem ser observadas as etapas para a execução do método Wavelet_ARIMA. O fluxograma do método Wavelet_RNA pode ser obtido pela substituição, na Figura 7, do ARIMA pela RNA.

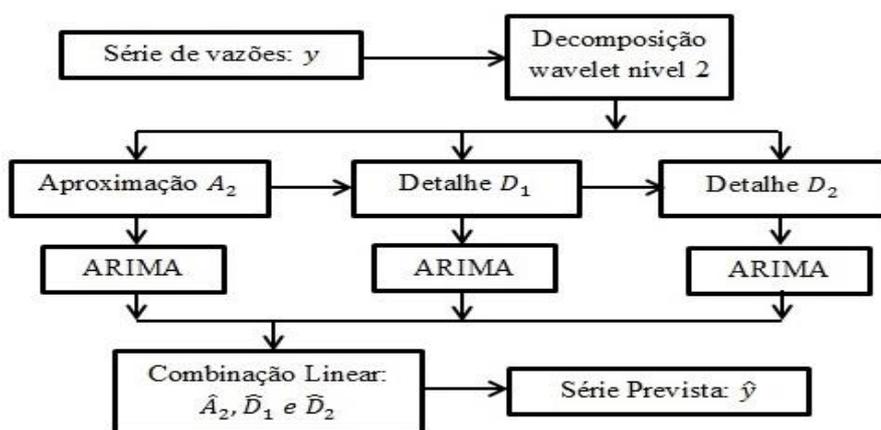


Figura 7 – Fluxograma com as etapas do método Wavelet_ARIMA.

Na figura 8 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).

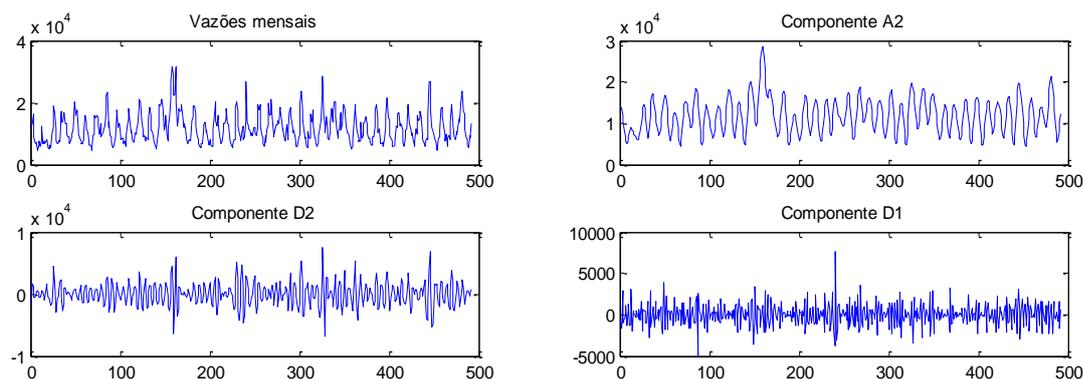


Figura 8 – Gráfico da série de vazões e componentes wavelet.

O Quadro 1 traz 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*).

Quadro 1 – Configurações RNA e ordens dos modelos AR e MA para A2, D1 e D2.

Componente de aproximação de nível 2 (A2):
RNA_1: <i>feedforward</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):
RNA_2: <i>feedforward</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):
RNA_3: <i>feedforward</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)

As previsões $\hat{A}_{2,RNA}$, $\hat{D}_{1,RNA}$ e $\hat{D}_{2,RNA}$ foram somadas a fim de se obter a previsão da série de vazões, conforme (16).

$$(\hat{y}_t)_{t=1}^T = \hat{A}_{2,RNA} + \hat{D}_{1,RNA} + \hat{D}_{2,RNA} \tag{16}$$

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

O MAPE para a amostra de teste obtido com a aplicação do método Wavelet_ARIMA foi de 3,63% e 2,80% com a utilização do Wavelet_RNA, mostrando a eficiência desses métodos em relação aos individuais RNA e ARIMA.

8.3 Método Proposto e Comparativo Entre os Métodos

Para o encolhimento wavelet foram usadas as bases de Daubechies com momentos 1, 5 e 10 (db1, db5 e db10) em níveis $p=2$ e 3, com regra de limiarização suave e limiar universal e SURE. Entre as várias bases e níveis de decomposição avaliados, o encolhimento wavelet determinado a partir da decomposição com a base db1 (haar), em nível $p=2$, forneceu as melhores previsões, sendo então a configuração escolhida na obtenção dos resultados apresentados neste artigo.

Com o encolhimento wavelet da série y obteve-se a série y_{EW} , supostamente menos ruidosa que a observada. Esta última, por sua vez, foi decomposta via wavelet (db10 e $p=2$) e as componentes obtidas ajustadas por RNA e ARIMA. A soma das previsões das componentes wavelet e resíduo determinou a previsão de y , conforme descrito na seção 7.2.

A melhor RNA identificada na modelagem das componentes wavelet tem a seguinte estrutura topológica: tamanho da janela de entrada igual a 12, uma camada escondida composta por 10 neurônios artificiais com função de ativação *tansig* e um neurônio na camada de saída com função de ativação *purelin*. O quadro 2 traz as ordens dos modelos autorregressivos (AR) e média móveis (MA).

Quadro 2 – Ordens dos modelos AR e MA para A2, D1 e D2.

Componente de aproximação de nível 2 (A2):
AR(1), AR(9), AR(10), MA(1) MA(2) MA(4 até 10), MA(12), MA(13) MA(23), MA(24) e MA(25). Com transformação logarítmica aplicada sobre A2.
Componente de detalhe de nível 1 (D1):
AR(1 até 10), MA(1), MA(3), MA(4).
Componente de detalhe de nível 2 (D2):
AR(1 até 10), MA(1 até 5)

No diagrama de dispersão apresentado na Figura 9 se observa alta correlação (coeficiente de correlação = 0,99) entre a amostra de teste e as previsões alcançadas com o método proposto, sendo um indicativo da qualidade das previsões obtidas.

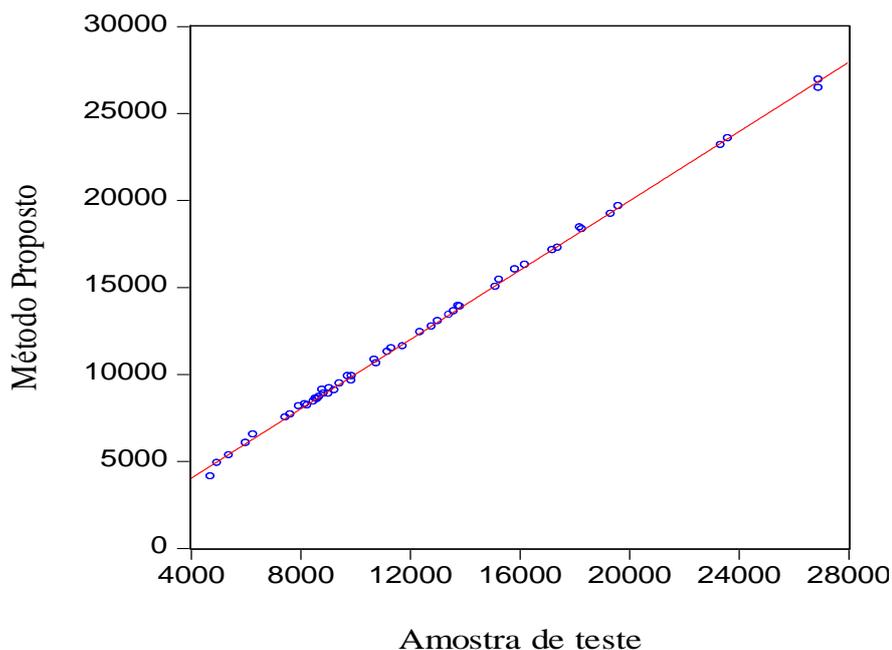


Figura 9 - Diagrama de dispersão entre amostra de teste e previsão pelo método proposto.

A tabela 1 informa o MAPE e o MAE correspondentes às previsões feitas um passo à frente da amostra de teste (49 observações) retirada da série de vazões, onde se pode observar uma superioridade do método proposto em relação aos outros métodos aplicados neste trabalho.

Tabela 1 – MAPE e MAE para amostra de teste retirada da série de vazões posto 266-Itaipu.

Método	Medidas de acurácia	
	MAPE	MAE
Método proposto	1,22%	119,75
Wavelet_RNA	2,80%	290,79
Wavelet_ARIMA	3,63%	396,26
ARIMA	18,01%	2104,39
RNA	20,12%	2417,88

Os melhores resultados foram obtidos pelos métodos Wavelet_RNA, Wavelet_ARIMA e proposto, sendo o último o mais eficiente com MAPE 66,39% menor em relação ao Wavelet_ARIMA e 56,43% em relação ao Wavelet_RNA, como mostra o gráfico comparativo da figura 10.

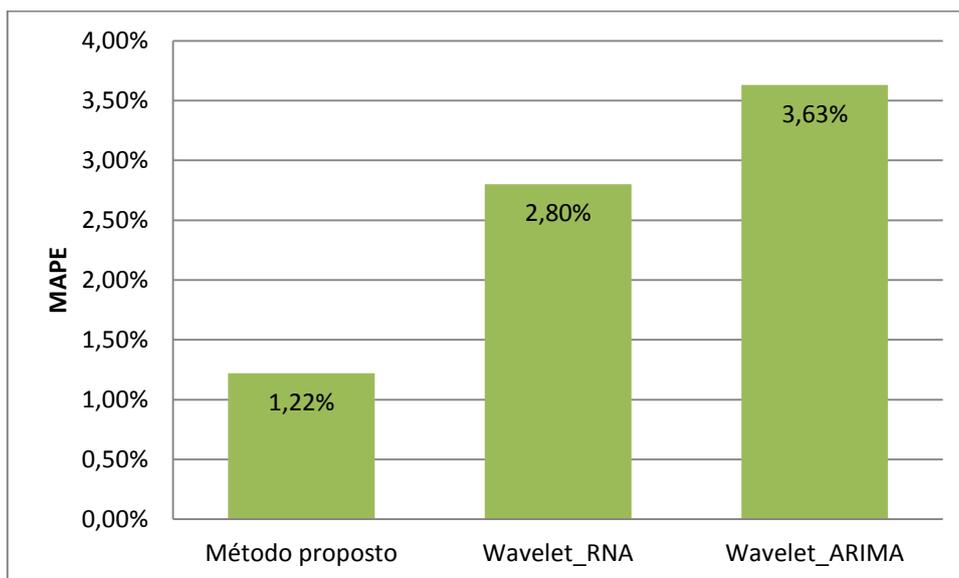


Figura 10 – MAPE amostra de teste para os métodos Wavelet_RNA, Wavelet_ARIMA e proposto.

Nas Figuras 11 e 12, tem-se a comparação entre o método proposto e as abordagens Wavelet-RNA e Wavelet-ARIMA. Nota-se que, em todos os 49 instantes, na amostra de teste, o método proposto apresenta evolução temporal dos valores dos erros percentuais absolutos (APE) mais estáveis que os dos métodos Wavelet-RNA e Wavelet-ARIMA. Nesses métodos, os valores do APE violaram em vários instantes a faixa dos 5% (Figura 11), enquanto que em nenhum instante isso ocorre no método proposto. Além disso, na maioria dos instantes, o método proposto apresenta valores de APE inferiores aos outros dois. A Figura 12 traz os gráficos dos resíduos relativos aos três melhores métodos aplicados neste trabalho. Nesta figura pode-se observar que os resíduos oriundos do método proposto mantêm-se, com exceção de um instante, no intervalo [-400, 400]. Por outro lado, esse intervalo é violado em vários instantes pelos outros dois métodos (Wavelet_RNA e Wavelet_ARIMA), inclusive com picos na ordem de 1200m³/s.

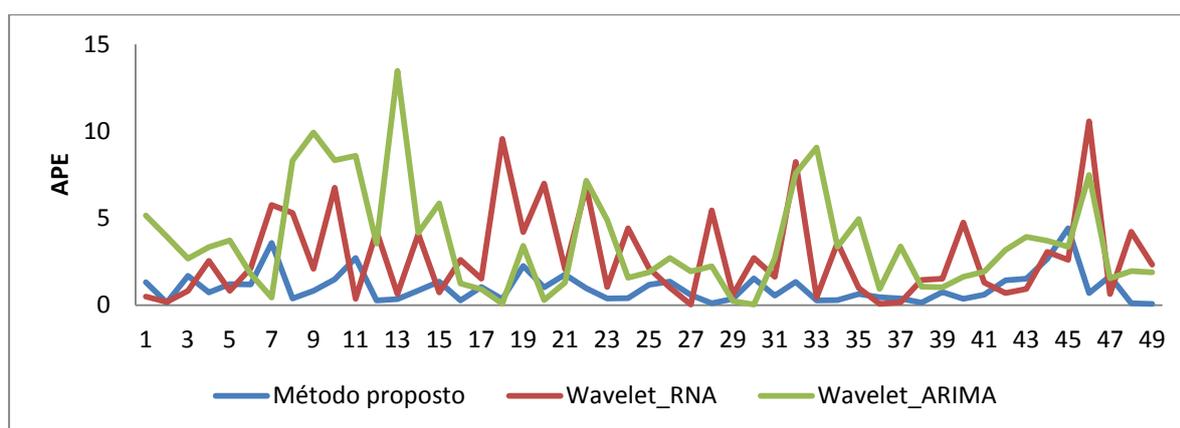


Figura 11- APE na amostra de teste para os métodos Wavelet_RNA, Wavelet_ARIMA e proposto.

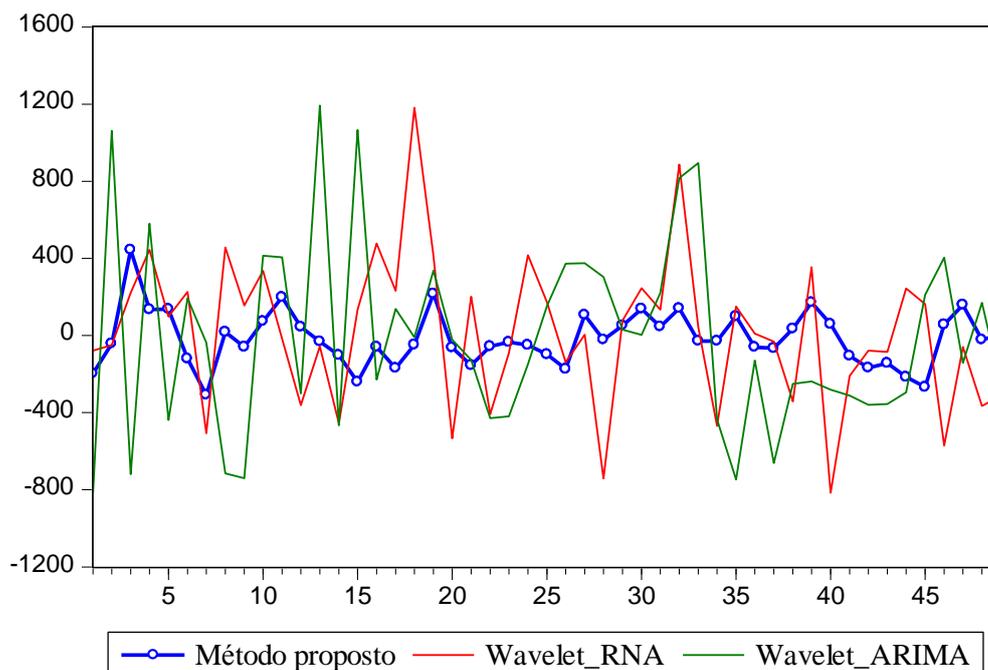


Figura 12 – Resíduos na amostra de teste para os métodos Wavelet_RNA, Wavelet_ARIMA e proposto.

9. CONSIDERAÇÕES FINAIS

Neste artigo foi proposto um método para previsões de séries que envolvem o encolhimento e decomposição wavelet, a modelagem das componentes wavelet via ARIMA e RNA e combinação linear das previsões das componentes wavelet. Comparações com resultados de previsões obtidos através de modelos de Box & Jenkins e RNA, quando usados individualmente, constata consideráveis ganhos preditivos auferidos com o uso do método proposto. Reduzindo o erro preditivo em 90%, aproximadamente. Na comparação com os métodos compostos Wavelet_ARIMA e Wavelet_RNA a redução do MAPE foi da ordem de 60%.

O ponto a ser melhorado no método proposto está da modelagem das componentes wavelet, visto que os modelos ajustados às componentes A2, D1 e D2 não seguem o princípio da parcimônia. De acordo com Tan (2010) as componentes Wavelet apresentam erro heterocedástico e poderiam ser mais facilmente ajustadas com o uso dos modelos da família ARCH.

O comportamento “nervoso” das componentes de detalhes não pode ser um empecilho para o seu uso, haja vista os ganhos preditivos auferidos na aplicação dessa técnica. Kisi, (2010), combinou decomposição wavelet e regressão linear (WR) e comparou os seus resultados aos obtidos com modelagem ARIMA. Constatou que o primeiro método (WR) apresentou melhores resultados de previsão que o segundo. Teixeira Jr *et al.*, (2012), apresentaram um método em que as componentes *wavelets* de uma série temporal constituem os padrões de entrada de uma RNA *feedforward* MLP, cuja saída fornece a previsão da série temporal. Com esta metodologia, chegaram a resultados melhores aos obtidos por RNA convencional (sem considerar componentes wavelet como padrões de entrada).

Resultados alcançados neste trabalho e corroborados por outros encontrados na literatura, já citados anteriormente, fortalecem a conclusão que a combinação de métodos preditivos com incorporação do encolhimento e decomposição wavelet pode auferir ganhos consideráveis em relação a outros que não o fazem.

REFERÊNCIAS

- BATES, J. M. and GRANGER, C. W. J.** *The Combining of Forecasts. Operational Research Quarterly*, v.20, n.4, p. 451-468, 1969.
- BOX, G.; JENKINS, G.** *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day, 1970.
- BOX, G.; JENKINS, G.; REINSEL, G. C.** *Time Series Analysis, Forecasting and Control*. 3.ed. Englewood Cliffs:Prentice-Hall, 1994. 598p.
- CASTRO, L. N.** *Computação Natural*. São Paulo: Editora Livraria da Física, 2010.
- DAUBECHIES, I.** *Ten Lectures on Wavelets (C B M S N S F Regional Conference Series in Applied Mathematics)*, v. 61, Soc for Industrial & Applied Math, Philadelphia, 1992.
- DONOHO, D. L.; JONHSTONE, I. M.** *Ideal Spatial Adaptation by Wavelet Shrinkage. Biometric*, 81,425-455, 1994.
- DONOHO, D.L.; JOHNSTONE, I. M.** *Adapting Unknow Smoothness via Wavelet Shrinkage. Division of Biostatistics/Stanford University Technical Report*, n.161,1995.
- DOHONO, D.L.; JOHNSTONE, I.M.** *Minimax Estimation via Wavelet Shrinkage. The Annals of Statistics*, v.26, n.3, p.879-921. 1998.
- DONOHO, D. L.; JONHSTONE, I. M.; KERKYACHARIAN, G.; and PICARD, D.** *Wavelet Shrinkage: Asymptopia? (with discussion) Journal Royal Statistical Society. B*, 57:301-369, 1995.
- FARIA, A. E.; MUBWANDARIKWA, E.** *Multimodality on the Geometric Combination of Bayesian Forecasting Models. International Journal of Statistics and Management System*, p. 1-25, 2008.
- GRANGER, C.W.J. & RAMANATHAN, R.** *Improved Methods of Forecasting. Journal of Forecasting*, v.3, p.197-204, 1984.
- GUPTA, S. and WILTON, P. C.** *Combination of Forecasts: An Extension. Management Science*. v.33, n.3, p.356-372, 1987.
- HAMILTON, J.** *Time Series Analysis*. Princeton University Press, 1994.
- HAYKIN, S.** *Redes Neurais Princípios e Aplicações*, 2. Ed. Porto Alegre: Bookman, 2001.
- JOHNSTONE, I.M.; SILVERMAN, B.** *Wavelet Threshold Estimators for data with Correlated Noise. Stanford University Technical Report*, 1996.
- KISI, O.** *Wavelet regression model for short-term streamflow forecasting. Journal of hidrology*, p. 344-353, 2010.
- KUBRUSLY, C. S. and LEVAN, N.** *Abstract Wavelets Generated by Hilbert Space Shift Operators. Adv. Math. Sci. Appl.* 14, 643-660, 2006.

LEI, C. and RAN, L. *Short-term Wind Speed Forecasting Model for Wind Farm Based on wavelet Decomposition DRPT, Nanjing, China, pp 2525-2529, 2008.*

LEVAN N. and KUBRUSLY, C. S. *A Wavelet Time-Shift Detail Decomposition. Mathematics and Computers in Simulation. 63, 73-78, 2003.*

MAKRIDAKIS, S. *Why Combining Works? International Journal of Forecasting, v.5, p.601-603, 1989.*

MAKRIDAKIS, S. and WINKLER, R. L. *Averages of Forecasts: Some Empirical Results. Management Science, v.29, n.9, p.987-996, 1983.*

MALLAT, S. *A Wavelet Tour of Signal Processing. Academic Press, San Diego, 1998.*

MORETTIN, P. A. *Ondas e Ondasletas: Da análise de Fourier à Análise. São Paulo: EDUSP, 1999.*

MORETTIN, P. A. e TOLOI, L. M. C. *Análise Séries Temporais. 2ª Ed. ABE Projeto Fisher, Ed. Edgard Blucher, 2006.*

NEWBOLD, P. and GRANGER, C.W. J. *Experience with Forecasting Univariate Times Series and the Combination of Forecasts. Journal of the Royal Statistical Society. V.137, pp.131-165, 1974.*

TAN, Z. et al. *Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models. Applied Energy, n. 87, p. 3606-3610, 2010.*

TEIXEIRA Jr., L. A. et al. *Redes neurais artificiais wavelet causal na previsão da radiação solar direta. Anais... Rio de Janeiro: Simpósio Brasileiro de Pesquisa Operacional 2012.*

TEIXEIRA Jr., L. A. *Combinação SSA-Wavelet de métodos preditivos com ajuste numérico MINIMAX, na geração de previsões e de cenários. Rio de Janeiro, 114 p. Tese (Doutorado em Engenharia Elétrica). Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2013.*

TERUI, N. & DIJK, H. K. *Combined forecasts from linear and nonlinear time series models. International Journal of Forecasting, p. 421-438, 2002.*

WALLIS, K. F. *Combining forecasts - forty years later. Applied Financial Economics, p. 33-41, 2011.*

WANG, A.B.W. et al. *Forecasting daily streamflow using hybrid ANN models. Amsterdam: Journal of Hydrology, v. 324, p. 383-399, 2006a.*

WANG, A.B.W. et al. *Testing for nonlinearity of streamflow processes at different time scales. Amsterdam: Journal of Hydrology, v. 322, p. 247-268, 2006b.*

WINKLER, R. L. *Combining Forecasts: A Philosophical Basis and Some Current Issues. International Journal of Forecasting, v.5, p. 605-609, 1989.*

ZANG G. P. *Time Series Forecasting Using a Hybrid ARIMA and Neural Network Model. Neurocomputing, 50, p.159-175, 2003.*