

PREVISÕES DE VAZÕES MENS AIS VIA COMBINAÇÃO HÍBRIDA ARIMA-NEURAL COM ENCOLHIMENTO E DECOMPOSIÇÃO WAVELET

Levi Lopes Teixeira

Universidade Tecnológica Federal do Paraná - UTFPR
Av. Brasil, 4232 -Medianeira-PR
prof.levilopes@gmail.com

Paulo Henrique Siqueira

Universidade Federal do Paraná - UFPR
Centro Politécnico – Jardim das Américas -Curitiba-PR
paulohscwb@gmail.com

Luiz Albino Teixeira Jr

Universidade Federal da Integração Latina-Americana - UNILA
Av. Tancredo Neves, 6731 – PTI - Foz do Iguaçu-PR
luiz.a.t.junior@gmail.com

Edgar Manuel Carreno Franco

Universidade Estadual do Oeste do Paraná - UNIOESTE
Av. Tancredo Neves, 6731 – PTI- Foz do Iguaçu-PR
emfra.unioeste@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. ARIMA, Redes neurais artificiais, Séries temporais.

Área principal: EST.

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. ARIMA, Artificial neural networks, Time series.

Main area: EST

1. Introdução

Existem vários fatores que podem influenciar a vazão em um determinado ponto de um curso de água, entre eles: evaporação, precipitação, infiltração e percolação. Sendo a precipitação a principal componente na formação da vazão e responsável pela característica aleatória do processo hidrológico. Para o estudo de vazões em usina hidrelétricas pode-se destacar as vazões: natural, incremental, turbinada e vertida. A vazão natural é a vazão que seria observada em um posto de medição considerando o rio na sua condição natural. A vazão incremental, por sua vez, é toda vazão lateral captada entre um posto de medição e todos os postos de medição imediatamente a montante deste. A vazão turbinada é aquela que passa pelas turbinas e gera energia, enquanto a vazão vertida é aquela que passa pelos vertedouros da usina e 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.

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), 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 2, 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 4 com a previsão do resíduo proveniente do encolhimento wavelet.

2. Decomposição Wavelet

Na Análise de Fourier (canônica), uma série temporal (unidimensional) com T observações, denotada por $\mathbb{X}(\cdot) = (\mathbb{X}(1), \mathbb{X}(2), \dots, \mathbb{X}(T))$, é ortogonalmente decomposta em termos de senos e cossenos em diversas bandas de frequências (MORETTIN & TOLOI, 2006); enquanto que, na Análise Wavelet, $\mathbb{X}(\cdot)$ é, de forma ortogonal, decomposta, simultaneamente, a partir de informações de escala e de translação capturadas a partir de um conjunto de funções wavelet $\{\mathbb{X}_{\mathbb{M}, \mathbb{N}}(\cdot)\}_{(\mathbb{M}, \mathbb{N}) \in \mathbb{Z}^2}$, onde cada mapa $\mathbb{X}_{\mathbb{M}, \mathbb{N}}(\cdot) = 2^{\frac{\mathbb{M}}{2}} \mathbb{X}(2^{\mathbb{M}}(\cdot) - \mathbb{N})$ é gerado por meio da dilatação (que está associada ao parâmetro m) e translação (que está associada ao parâmetro n) de uma função wavelet $\mathbb{X}(\cdot)$ original (MALLAT, 2009). Segundo Levan e Kubrusly (2003), o parâmetro m é chamado de parâmetro de escala diádica e n , de parâmetro de translação unitária.

Em Teixeira JR *et al* (2013), verifica-se, empiricamente, que cada função wavelet $\mathbb{X}_{\mathbb{M}, \mathbb{N}}(\cdot)$ está associada ao mapeamento das componentes de detalhe (que são as componentes de alta frequência) de $\mathbb{X}(\cdot)$; e, teoricamente, que a componente de aproximação (que consiste na componente de baixa frequência) de $\mathbb{X}(\cdot)$ está associada ao conjunto de funções (wavelet) escala $\{\mathbb{X}_{\mathbb{M}_0, \mathbb{N}}(\cdot)\}_{\mathbb{N} \in \mathbb{Z}}$. Cada mapa $\mathbb{X}_{\mathbb{M}_0, \mathbb{N}}(\cdot) = 2^{\frac{\mathbb{M}_0}{2}} \mathbb{X}(2^{\mathbb{M}_0}(\cdot) - \mathbb{N})$, onde \mathbb{M}_0 é um número inteiro fixo, é gerado por meio da translação unitária n de uma função escala original $\mathbb{X}(\cdot)$ (MALLAT, 2009). De acordo com Daubechies (1988), por definição, as funções wavelet e as funções escala consistem em ondas de curta duração, geralmente assimétricas, que permitem a sua dilatação diádica n (positiva ou negativa) e translação unitária m (positiva ou negativa) de seu domínio. Não obstante, definições mais formais e gerais, do ponto de vista matemático, podem ser verificadas em Teixeira JR *et al* (2013). Segundo Mallat (2009), as famílias wavelet, em geral, são apresentadas em pares constituídos de uma função wavelet $\mathbb{X}(\cdot)$ e uma função escala $\mathbb{X}(\cdot)$ originais. As famílias wavelet de Haar, Daubechies, Coifelet e Symelet são exemplos de famílias que, comumente, são utilizadas na modelagem de séries temporais, para fins de geração de previsões.

De acordo com Kubrusly & Levan (2006) e Mallat (2009), tem-se que a decomposição wavelet de uma série temporal $\mathbb{X}(\cdot)$ é dada, genericamente, em (1).

$$\mathbb{X}(\cdot) = \mathbb{X}_{\mathbb{M}_0, \mathbb{N}}(\cdot) + \sum_{\mathbb{M}=\mathbb{M}_0}^{\infty} \mathbb{X}_{\mathbb{M}, \mathbb{N}}(\cdot) \quad (1)$$

Onde: $\mathbb{X}_{\mathbb{M}_0, \mathbb{N}}(\cdot) = \sum_{\mathbb{N} \in \mathbb{Z}} \mathbb{X}_{\mathbb{M}_0, \mathbb{N}} \mathbb{X}_{\mathbb{M}_0, \mathbb{N}}(\cdot)$ é a componente de aproximação de nível \mathbb{M}_0 de $\mathbb{X}(\cdot)$; $\mathbb{X}_{\mathbb{M}, \mathbb{N}}(\cdot) = \sum_{\mathbb{N} \in \mathbb{Z}} \mathbb{X}_{\mathbb{M}, \mathbb{N}} \mathbb{X}_{\mathbb{M}, \mathbb{N}}(\cdot)$ é a componente de detalhe de nível \mathbb{M} de $\mathbb{X}(\cdot)$; $\mathbb{X}_{\mathbb{M}, \mathbb{N}}$ e $\mathbb{X}_{\mathbb{M}_0, \mathbb{N}}$ são, respectivamente, os coeficientes de detalhe e de aproximação e são calculados, ainda nesta

ordem, a partir do produto interno usual entre $\tilde{w}(\cdot)$ e $\tilde{w}_{j,k}(\cdot)$ (isto é, $\tilde{w}_{j,k} = \sum_{m=1}^M \tilde{w}(m) \tilde{w}_{j,k}(m)$) e entre $\tilde{w}(\cdot)$ e $\tilde{w}_{j_0,k}(\cdot)$ (isto é, $\tilde{w}_{j_0,k} = \sum_{m=1}^M \tilde{w}(m) \tilde{w}_{j_0,k}(m)$), para todo $j_0, j, k \in \mathbb{Z}$, onde \mathbb{Z} é o conjunto dos números inteiros.

Na prática, é usual que o parâmetro j_0 assuma valor igual ao nível de decomposição. Em particular, em uma decomposição wavelet de nível $p=2$, onde o truncamento, em (1), ocorre na segunda componente de detalhe, se adota $j_0 = 2$. Com efeito, tem-se que: $\tilde{w}(\cdot) = \tilde{w}_{2,2}(\cdot) + \tilde{w}_{2,3}(\cdot) + \tilde{w}$, onde \tilde{w} é o vetor de erro de aproximação. Independente do nível p de decomposição wavelet, observa-se empiricamente que \tilde{w} é, em geral, constituído de valores muito próximos a zero, de modo que pode ser desconsiderado para fins de geração de previsões (pontuais).

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 $(\tilde{w}_t)_{t=1}^T$ uma série temporal, onde para todo t : $\tilde{w}_t = \tilde{w}_t + \tilde{w}_t$. Sendo \tilde{w}_t um estado da variável aleatória $\tilde{w}_t \sim (0, \sigma^2)$, onde \tilde{w}_t é a componente determinística de \tilde{w}_t . Como não é possível modelar \tilde{w}_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 (1994) e Danoho *et al* (1995).

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 (2) e (3), respectivamente.

$$\tilde{w}_{j,k}^{\text{hard}} = \begin{cases} \tilde{w}_{j,k}, & |\tilde{w}_{j,k}| > \lambda \\ 0, & |\tilde{w}_{j,k}| \leq \lambda \end{cases} \quad (2)$$

$$\tilde{w}_{j,k}^{\text{soft}} = \begin{cases} \tilde{w}_{j,k} - \lambda \frac{\tilde{w}_{j,k}}{|\tilde{w}_{j,k}|}, & |\tilde{w}_{j,k}| > \lambda \\ 0, & |\tilde{w}_{j,k}| \leq \lambda \end{cases} \quad (3)$$

Onde $\tilde{w}_{j,k}$ é o coeficiente de detalhe com ruído, $\tilde{w}_{j,k}^{\text{hard}}$ e $\tilde{w}_{j,k}^{\text{soft}}$ são coeficientes de detalhes sem ruído, $\tilde{w}_{j,k} - \lambda \frac{\tilde{w}_{j,k}}{|\tilde{w}_{j,k}|}$ é o sinal de $\tilde{w}_{j,k}$ 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 $(|\tilde{w}_{j,k}| - \lambda)$ os valores de $\tilde{w}_{j,k}$ ($|\tilde{w}_{j,k}| > \lambda$), eliminando-os em caso contrário.

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 Johnstone e Silverman (1996) e Morettin (1999).

De acordo com Donoho e Johnstone (1998) o cálculo do limiar universal é dado por (4).

$$\lambda = \sigma \sqrt{2 \ln T} \quad (4)$$

Sendo σ o nível do ruído e T a cardinalidade da série. O limiar universal é independente da escala j , 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 j . Seja n_j o número de coeficientes na escala j , então o limiar de SURE é calculado a partir de (5).

$$\hat{y}_t = \sum_{0 \leq \tau \leq \sqrt{2\pi n}(\hat{\sigma}_t)} \hat{y}_t(\tau, \hat{\sigma}_t) \quad (5)$$

4. Redes Neurais 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).

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 (6).

$$\hat{y}_t = \hat{\mu} + \hat{\alpha}_1 \hat{y}_{t-1} + \hat{\alpha}_2 \hat{y}_{t-2} + \dots + \hat{\alpha}_p \hat{y}_{t-p} - \hat{\beta}_1 \hat{y}_{t-1} - \hat{\beta}_2 \hat{y}_{t-2} - \dots - \hat{\beta}_q \hat{y}_{t-q} + \hat{a}_t \quad (6)$$

O modelo (6) combina valores passados das entradas \hat{y}_t e choques aleatórios (a_t) descorrelacionados, de média zero e variância constante. Neste modelo $\hat{\alpha}_p, \hat{\beta}_q \in \mathbb{R}$, com $\hat{\alpha}_p = 1, \dots, \hat{\alpha}_1 = 1, \dots, \hat{\beta}_q$, denotam os parâmetros do modelo e $\hat{\mu}$ 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 Toloí (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 (7).

$$\Phi(B)(1 - \alpha_1 B - \dots - \alpha_p B^p) \Theta(B) (1 - \beta_1 B^S - \dots - \beta_q B^{qS}) = \Phi(B)(1 - \alpha_1 B - \dots - \alpha_p B^p) \Theta(B) \quad (7)$$

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, $\alpha_i \in \mathbb{R}$ e $\beta_j \in \mathbb{R}$ são os coeficientes dos polinômios não sazonais e $\alpha_i \in \mathbb{R}$ e $\beta_j \in \mathbb{R}$ 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 (8), para a previsão da série temporal $\{y_t\}_{t=1}^T$.

$$\hat{y}_t = \rho_1 \times \hat{y}_{1,t} + \rho_2 \times \hat{y}_{2,t} \quad (8)$$

Em (8), 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 Guaíra (Brasil) e Salto del Guairá (Paraguai), ao norte.

7.2 Método Proposto

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

Considere a sequência infinita $\{\tilde{x}_t\}_{t \in \mathbb{Z}}$ obtida da série temporal $\{x_{j,k}\}_{j,k=1}^n$ a partir da igualdade $\tilde{x}_t = \{x_{j,k}\}_{j,k=1}^n$, $t \in \{1, \dots, n\}$, $t \in \mathbb{Z} \setminus \{1, \dots, n\}$. Segundo Teixeira Jr *et al.*, (2012) a decomposição wavelet de nível j de uma série temporal \tilde{x} em relação a uma determinada base ortonormal wavelet é representada por (9).

$$\tilde{x} \cong \tilde{x} = \sum_{j \in \mathbb{Z}_{j=0}} x_{j,k} x_{j,k} + \sum_{j=0}^{j_0+(j-1)} \left(\sum_{k \in \mathbb{Z}_{k=0}} x_{j,k} x_{j,k} \right) \quad (9)$$

Onde $\mathbb{Z}_{j=0}$ e $\mathbb{Z}_{k=0}$ são subconjuntos de índices em \mathbb{Z} . A determinação de tais conjuntos ocorre através da minimização do erro quadrático médio (MSE) entre \tilde{x} e \tilde{x} .

O encolhimento wavelet, as componentes de aproximação $\sum_{j \in \mathbb{Z}_{j=0}} x_{j,k} x_{j,k}$ e detalhe $\sum_{j \in \mathbb{Z}_{j=0}} x_{j,k} x_{j,k}$ 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 $\tilde{x}_{j,k}$ foi decomposta ortogonalmente via wavelet de nível $j = 2$, que resulta em duas componentes de detalhe (\tilde{x}_1 e \tilde{x}_2) e uma de aproximação (\tilde{x}_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 ($\hat{\tilde{x}}_{1,10000}$, $\hat{\tilde{x}}_{1,100000}$, $\hat{\tilde{x}}_{2,1000}$ e $\hat{\tilde{x}}_{2,100000}$) e aproximação ($\hat{\tilde{x}}_{2,1000}$ e $\hat{\tilde{x}}_{2,100000}$).

Tendo as previsões das componentes wavelet por ARIMA e RNA, fizeram-se três combinações lineares, cada uma envolvendo os pares de previsões ($\hat{\tilde{x}}_{2,1000}$ e $\hat{\tilde{x}}_{2,100000}$), ($\hat{\tilde{x}}_{1,1000}$ e $\hat{\tilde{x}}_{1,100000}$) e ($\hat{\tilde{x}}_{2,1000}$ e $\hat{\tilde{x}}_{2,100000}$). A combinação linear para a componente da aproximação ($\hat{\tilde{x}}_{2,1000}$) está descrita em (10).

$$\hat{\tilde{x}}_{2,1000} = \alpha_{2,100000} \times \hat{\tilde{x}}_{2,100000} + \beta_{2,1000} \times \hat{\tilde{x}}_{2,1000} \quad (10)$$

onde $\alpha_{2,100000}$ e $\beta_{2,1000}$ são parâmetros a serem determinados com a resolução do problema de programação não linear (11), cuja função objetivo é a raiz do erro quadrático médio (RMSE) entre a componente de aproximação $\tilde{x}_2(\tilde{x})$ e a sua previsão combinada $\hat{\tilde{x}}_{2,1000}(\tilde{x})$. Da mesma forma procede-se para obter as previsões combinadas $\hat{\tilde{x}}_{1,1000}$ e $\hat{\tilde{x}}_{2,1000}$.

$$\hat{\sigma}_2^2 = \sqrt{\frac{1}{n} \sum_{t=1}^n (\hat{\psi}_{2,t}(z) - \psi_2(z))^2} \quad (11)$$

A previsão da série observada é dada pela soma das previsões das componentes wavelet e previsão do resíduo ($\hat{\psi}$), como mostra (12).

$$\{\hat{\psi}_t\}_{t=1}^n = \{\hat{\psi}_{2,t}(z)\}_{t=1}^n + \{\hat{\psi}_{1,t}(z)\}_{t=1}^n + \{\hat{\psi}_{2,t}(z)\}_{t=1}^n + \hat{\psi} \quad (12)$$

Para a avaliação dos métodos foram utilizados os Erros MAPE (Erro Médio Percentual Absoluto) e MAE (Erro Médio absoluto), representados em (13).

$$\text{MAPE} = \frac{1}{n} \sum_{t=1}^n \left| \frac{y_t - \hat{y}_t}{y_t} \right| \quad \text{MAE} = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (13)$$

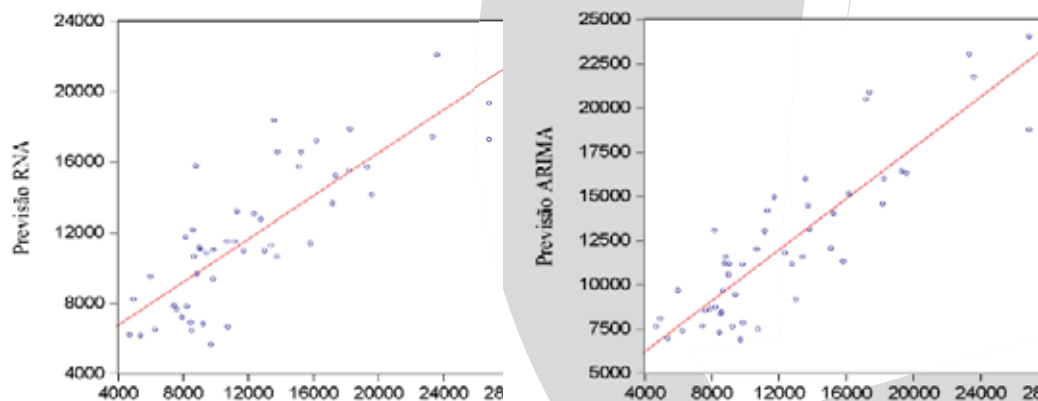
8. Resultados Obtidos

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 sigmoidal (*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 1a, 1b 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 1a o coeficiente de correlação encontrado foi de 0,84 e na Figura 1b obteve-se coeficiente igual a 0,87.



(a) Amostra de teste x RNA

(b) Amostra de teste x ARIMA

Figura 1 - 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). Os gráficos das componentes wavelet obtidas com a decomposição estão representados na figura 2, onde se podem observar os gráficos da série de vazões mensais e as componentes wavelet de aproximação (A2) e detalhes (D1 e D2). As três componentes foram ajustadas por três RNA e três modelos ARIMA.

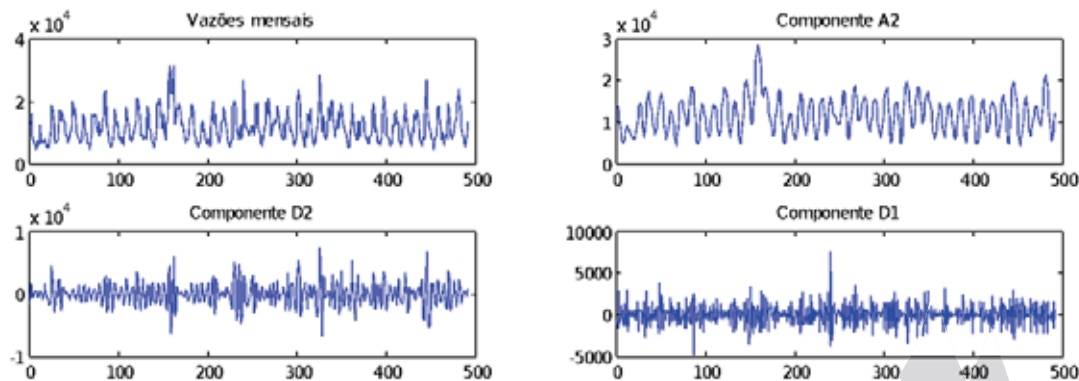


Figura 2 – Gráfico da série de vazões e componentes wavelet.

Com a soma das previsões $\hat{Q}_{2,0000}$, $\hat{Q}_{1,0000}$ e $\hat{Q}_{2,0000}$, obtidas por RNA, conclui-se o método Wavelet_RNA, conforme 14.

$$(\hat{Q}_Q)_{Q=1}^Q = \hat{Q}_{2,0000} + \hat{Q}_{1,0000} + \hat{Q}_{2,0000} \quad (14)$$

Onde $(\hat{Q}_Q)_{Q=1}^Q$ é a previsão da série observada. De forma análoga obtém-se $(\hat{Q}_Q)_{Q=1}^Q$ a partir das previsões $\hat{Q}_{2,000000}$, $\hat{Q}_{1,000000}$ e $\hat{Q}_{2,000000}$, oriundas dos modelos 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 $Q=2$ 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 $Q=2$, forneceu as melhores previsões, sendo então a configuração escolhida na obtenção dos resultados apresentados neste artigo.

Com o encolhimento wavelet da série Q obteve-se a série $Q_{Q,Q}$, supostamente menos ruidosa que a observada. A série $Q_{Q,Q}$, 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 ($Q = Q - Q_{Q,Q}$) determinou a previsão de Q , 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 tangente hiperbólica sigmoidal (*tansig*) e um

neurônio na camada de saída com função de ativação linear (*purelin*). O quadro 1 traz as ordens dos modelos autorregressivos (AR) e média móveis (MA).

Quadro 1 – Ordens dos modelos AR e MA para as componentes wavelet 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)

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.

Na Figura 3, 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%, 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.

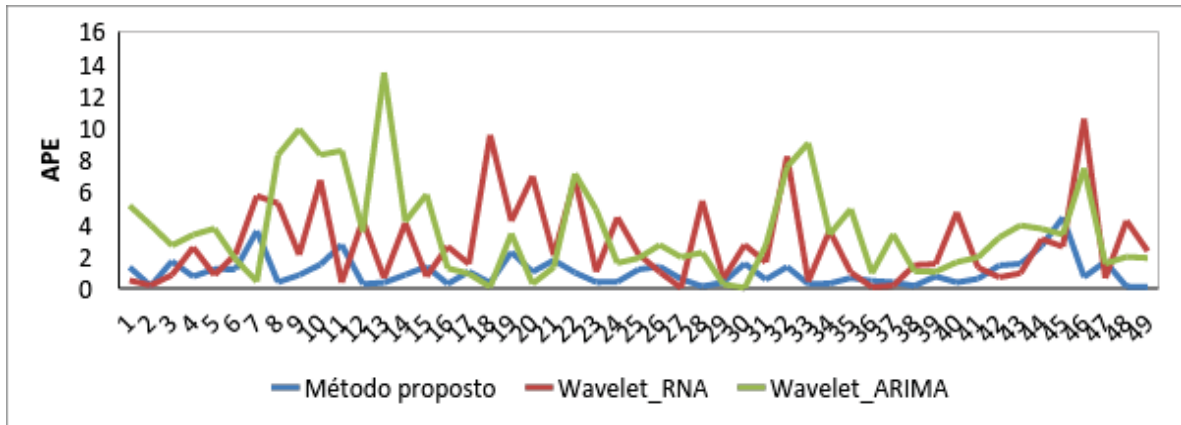


Figura 3- APE na amostra de teste para os métodos Wavelet_RNA, Wavelet_ARIMA e proposto.

Na Figura 4 está representado o gráfico de dispersão da amostra de teste versus previsão pelo método proposto, sendo o coeficiente de correlação entre esses dois conjuntos de dados igual a 0,99.

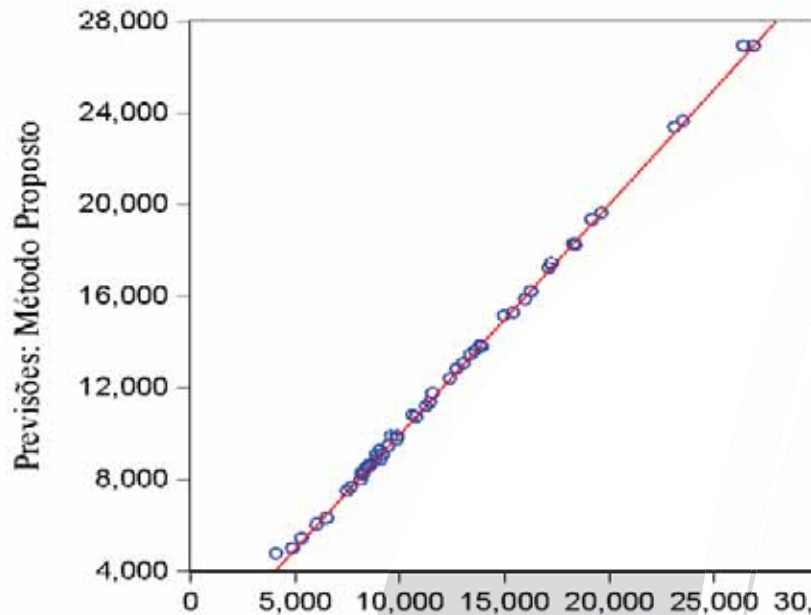


Figura 4 - Diagrama de dispersão entre a amostra de teste e previsões pelo método proposto.

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%.

Dos vários métodos de análise e previsão de séries temporais existentes, os modelos ARIMA e RNA são os mais populares, entretanto, a decomposição wavelet e métodos de filtragem se apresentam como boas opções na busca por previsões acuradas. A combinação de

métodos proposta neste artigo valida esta ideia, haja vista a estabilidade temporal obtida na evolução dos valores de APE (figura 3) e previsões fortemente correlacionadas com os dados (figura (4)), de forma a ratificar a sua acurácia no experimento computacional realizado com o intuito de se fazer as previsões das vazões mensais no posto 266-Itaipu. Em última análise, salienta-se que, embora o embasamento teórico do método proposto seja matematicamente complexo, a utilização dos softwares referenciados no texto viabiliza a sua utilização, ao passo poder ser operacionalizado de forma relativamente simples

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.*
- 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.
- DAUBECHIES, I.** Orthonormal Bases of Compactly Supported wavelets. *Comm. PureandAppliedMath.*, 41, pp.909-996, 1988.
- DONOHO, D. L.; JOHNSTONE, 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. et al.** *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.*
- KUBRUSLY, C. S. and LEVAN, N.** *Abstract Wavelets Generated by Hilbert Space Shift Operators. Adv. Math. Sci. Appl.* 14, 643-660, 2006.
- LEVAN N. and KUBRUSLY, C. S.** *A Wavelet Time-Shift Detail Decomposition. Mathematics and Computers in Simulation.* 63, 73-78, 2003.
- MALLAT, S.** *A Wavelet Tour of Signal Processing. Academic Press, San Diego, 2009.*
- MORETTIN, P. A.** *Ondas e Ondaletas: 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.*
- 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.*

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

