

Forecasting of the relative displacements in block I11 of the Itaipu hydroelectric plant dam through a ARIMA-SVR hybrid approach

Previsão dos deslocamentos relativos no bloco I11 da barragem da usina hidrelétrica de Itaipu por meio de uma abordagem híbrida ARIMA-SVR

Samuel Bellido Rodrigues¹

Levi Lopes Teixeira¹

Luiz Albino Teixeira Júnior²

Arinei Carlos Lindbeck da Silva³

Paulo Henrique Siqueira³

Edgar Manuel Carreño Franco⁴

¹Universidade Tecnológica Federal do Paraná

samuelb@utfpr.edu.br

levilopes@utfpr.edu.br

²Universidade Federal da Integração Latino-Americana - UNILA

Luiz.a.t.junior@gmail.com

³Universidade Federal do Paraná

arineicls@gmail.com

paulohs@ufpr.br

⁴Universidade Estadual do Oeste do Paraná

emfra.uinoeste@gmail.com

Abstract: Last four decades, the well-known ARIMA models have been widely used for time series forecasting. However,

in recent years, the support vector regression (SVR) has also appeared as a promising alternative for predicting them. In several papers, the performance of the ARIMA models and the SVRs are confronted and different conclusions are drawn such that they cannot be considered as conclusive. Thus, this paper puts forward a hybrid approach for predicting real time series, referred to as the ARIMA-SVR hybrid approach. It combines both the ARIMA models and the SVRs in order to take advantage simultaneously from the linear auto-dependence (by using the ARIMA models) and from the non-linear auto-dependence (by using the SVRs models) exhibited by the time series. In addition, the proposed methodology uses the wavelet decomposition and linear combination of forecasts (CL). The numerical results have indicated that the ARIMA-SVR hybrid approach may be effective to obtain accuracy gains in the predictive process.

Keywords: Time series forecasting, ARIMA models, decomposition wavelet, support vector regression, linear combination of forecasts.

Resumo: Nas últimas quatro décadas, os modelos ARIMA têm sido amplamente utilizados na previsão de séries temporais. No entanto, em recentes anos, a regressão por vetores suporte (SVR) tem também aparecido como uma alternativa promissora para a predizê-las. Em vários artigos, os desempenhos dos modelos ARIMA e SVR são confrontados e diferentes conclusões são obtidas, de modo que não podem ser consideradas como conclusivas. Assim, este artigo propõe uma abordagem híbrida para prever séries temporais reais, referida como abordagem híbrida ARIMA-SVR. Ela combina ambos os modelos ARIMA e SVR para tirar proveito, simultaneamente, da auto-dependência linear (usando os modelos ARIMA) e da auto-dependência não-linear (usando o modelo SVR) exibidas pela série temporal. Além disso, a metodologia proposta utiliza a decomposição wavelet e combinação linear de previsões (CL). Os resultados numéricos indicaram que a abordagem híbrida de ARIMA-SVR pode ser eficaz para obter ganhos de precisão no processo preditivo.

Palavras-Chave: previsão de séries temporais, modelos ARIMA, decomposição wavelet, regressão por vetores suporte, combinação linear de previsões.

1 Introdução

A previsão de séries temporais é uma importante área de pesquisa. Para prever uma série de tempo são utilizadas observações passadas da mesma variável que são coletadas e analisadas com o intuito de desenvolver um modelo que apresenta a relação implícita entre passado e futuro. Várias pesquisas têm sido realizadas ao longo dos últimos anos para o desenvolvimento e aprimoramento da previsão de séries temporais, como por exemplo, Aladag, Egrioglu e Kadilar [1], Babu e Reddy [2] e Pai e Lin [3].

O objetivo central consiste em modelar a série temporal de dados existente com o intuito de se prever os valores futuros de dados desconhecidos, com a melhor acurácia possível. Tal precisão nas previsões de séries temporais é fundamental para muitos processos de decisão. Com efeito, a pesquisa pela busca de modelos mais completos para

serem utilizados para prever séries temporais nunca foi interrompida [2] [4] [5].

Dentre os modelos e algoritmos existentes para prever uma série temporal, o método ARIMA [6], que consiste em ajustar modelos lineares auto-regressivos integrados de médias móveis, é sem dúvida um dos mais populares. Diversas aplicações são encontradas na literatura, como, por exemplo, em [2] [7] [8].

A regressão por vetores suporte (SVR) [9], que é uma máquina de aprendizagem baseada na teoria da aprendizagem estatística, e adere ao princípio da minimização do risco estrutural, é uma técnica que vem sendo usada frequentemente nos últimos anos na modelagem de séries temporais não lineares com extraordinário sucesso [10] [11] [12].

Segundo Granger e Remanathan [13], a combinação de previsões possibilita a geração de melhores previsões. De acordo com Hollauer, Issler e Notini [14], a diversificação de previsões leva à diminuição do erro, e Faria e Mubwandarikwa [15] 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.

Dado o exposto, é proposta neste artigo uma nova abordagem híbrida para previsão de séries temporais, a qual utiliza, de forma integrada, as modelagens ARIMA e SVR, decomposição Wavelet e combinação de previsões.

2 Metodologia Box-Jenkins

A metodologia proposta por Box e Jenkins [6] 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 [6], um modelo ARMA plausível para a série temporal $(y_t)_{t=1}^T$, T sendo a cardinalidade, de ordens p (autorregressivo) e q (médias móveis) é descrito pela equação 1.

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

O modelo da equação 1 combina valores passados das entradas Z_t e choques aleatórios (a_t) descorrelacionados, de média zero e variância constante. Nesse 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 [6] e Morettin e Tolo [16]. 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 e teste da autocorrelação cruzada. No caso da série temporal $(y_t)_{t=1}^T$ apresentar componente sazonal, o modelo Box-Jenkins plausível é dado genericamente, na equação 2:

$$(1 - \Phi_1 B - \dots - \Phi_p B^p)(1 - \Phi B^S - \dots - \Phi_p B^{pS})(1 - B)^d (1 - B^S)^D y_t = (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \theta_1 B^S - \dots - \theta_p B^{qS}) \varepsilon_t \quad (2)$$

onde 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.

3 Wavelet

A análise Wavelet tem se mostrado nos últimos anos uma ferramenta eficiente em vários ramos de pesquisa, tais como: engenharia, física, matemática e estatística.

Função Wavelet é uma função capaz de decompor ou representar uma série temporal originalmente descrita num domínio de tempo de tal forma que a série possa ser analisada em diferentes escalas de frequência de tempo. A decomposição de uma função com o uso de Wavelet é conhecida como transformada Wavelet.

Na Análise de Fourier (canônica), uma série temporal (unidimensional) com T observações, denotada por $f(\cdot) = (f(1), f(2), \dots, f(T))$, é ortogonalmente decomposta em termos de senos e cossenos em diversas bandas de frequências [16]; enquanto que, na Análise Wavelet, $f(\cdot)$ é, de forma ortogonal, decomposta, simultaneamente, a partir de informações de escala e de translação capturadas a partir de um conjunto de funções Wavelet $\{\omega_{m,n}(\cdot)\}_{(m,n) \in \mathbb{Z}^2}$, onde cada mapa $\omega_{m,n}(\cdot) = 2^{\frac{m}{2}} \omega(2^m(\cdot) - n)$ é gerado por meio da dilatação (que está associada ao parâmetro m) e translação (que está associada ao parâmetro n) de uma função Wavelet $\omega(\cdot)$ original conforme Mallat [17]. Segundo Levan e Kubrusly [18], o parâmetro m é chamado de parâmetro de escala diádica e n , de parâmetro de translação unitária.

Em Teixeira Jr et al [19], verifica-se, empiricamente, que cada função Wavelet $\omega_{m,n}(\cdot)$ está associada ao mapeamento das componentes de detalhe (que são as componentes de alta frequência) de $f(\cdot)$; e, teoricamente, que a componente de aproximação (que consiste na componente de baixa frequência) de $f(\cdot)$ está associada ao conjunto de funções (Wavelet) escala $\{\phi_{m_0,n}(\cdot)\}_{n \in \mathbb{Z}}$. Cada mapa $\phi_{m_0,n}(\cdot) = 2^{\frac{m_0}{2}} \phi(2^{m_0}(\cdot) - n)$, onde m_0 é um número inteiro fixo, é gerado por meio da translação unitária n de uma função escala original $\phi(\cdot)$ [17]. De acordo com Daubechies [20], 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 [19]. Segundo Mallat [17], as famílias Wavelet, em geral, são apresentadas em pares constituídos de uma função Wavelet $\omega(\cdot)$ e uma função escala $\phi(\cdot)$ 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 e Levan [21] e Mallat [17], tem-se que a decomposição Wavelet de uma série temporal $f(\cdot)$ é: dada, genericamente, pela equação 3.

$$f(\cdot) = f_{A_{m_0}}(\cdot) + \sum_{m=m_0}^{\infty} f_{D_m}(\cdot) \quad (3)$$

onde: $f_{A_{m_0}}(\cdot) = \sum_{n \in \mathbb{Z}} a_{m_0,n} \phi_{m_0,n}(\cdot)$ é a componente de aproximação de nível m_0 de $f(\cdot)$; $f_{D_m}(\cdot) = \sum_{n \in \mathbb{Z}} d_{m,n} \omega_{m,n}(\cdot)$ é a componente de detalhe de nível m de $f(\cdot)$; $d_{m,n}$ e $a_{m_0,n}$ são, respectivamente, os coeficientes de detalhe e de aproximação e são calculados, ainda nesta ordem, a partir do produto interno usual entre $f(\cdot)$ e $\omega_{m,n}(\cdot)$ (isto é, $d_{m,n} = \sum_{t=1}^T f(t) \omega_{m,n}(t)$) e entre $f(\cdot)$ e $\phi_{m_0,n}(\cdot)$ (isto é, $a_{m_0,n} = \sum_{t=1}^T f(t) \phi_{m_0,n}(t)$), para todo $m_0, m, n \in \mathbb{Z}$, onde \mathbb{Z} é o conjunto dos números inteiros.

Na prática, é usual que o parâmetro m_0 assuma valor igual ao nível de decomposição. Em particular, em uma decomposição Wavelet de nível $p=2$, onde o truncamento, na equação 3, ocorre na segunda componente de detalhe, adota-se $m_0 = 2$. Com efeito, tem-se que: $f(\cdot) = f_{A_2}(\cdot) + f_{D_2}(\cdot) + f_{D_3}(\cdot) + \varepsilon$, onde ε é o vetor de erro de aproximação. Independente do nível p de decomposição Wavelet, observa-se empiricamente que ε é, em geral, constituído de valores muito próximos a zero, de modo que pode ser desconsiderado para fins de geração de previsões (pontuais).

4 Support Vector Regression

A regressão de vetores de suporte (SVR, *Support Vector Regression*) é fundamentada em uma máquina de vetor de suporte (SVM, *Support Vector Machine*) [9]. O modelo produzido por SVR depende apenas de um subconjunto dos dados de formação e por sua função de custo que é utilizada para a construção do modelo, ignorando os dados de formação perto do modelo de previsão.

As descrições a seguir sobre o SVR baseiam-se no tutorial de Smola e Schölkopf [22]. Mais detalhes, podem ser encontrados nesse guia.

Seja uma amostra de dados de treinamento $\{(x_1, y_1), \dots, (x_l, y_l)\} \subset X \times \mathbb{R}$, onde X indica o espaço de padrões de entrada, o objetivo é encontrar uma função $f(x)$ que apresente no máximo um desvio ε dos alvos y_i , obtido para todos os dados de treinamento. Logo, descreve-se a função linear de aproximação por $f(x) = \langle w, x \rangle + b$ com $w \in X$, $b \in \mathbb{R}$, onde $\langle \cdot, \cdot \rangle$ denota o produto interno em X . Contudo, nem sempre é possível garantir a viabilidade do problema, já que existem pontos que violam as restrições. Assim, chega-se à formulação

mencionada por Vapnik [9], com variáveis de folga, de acordo com a expressão 4:

$$\begin{aligned} & \text{minimize } \frac{1}{2} \|w\|^2 + C \left(\sum_{i=1}^l \xi_i + \xi_i^* \right) \\ & \text{sujeito a: } \begin{cases} y_i - \langle w, x_i \rangle - b \leq \varepsilon + \xi_i \\ \langle w, x_i \rangle + b - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, \quad \forall i = 1, \dots, l \end{cases} \end{aligned} \quad (4)$$

Para a formulação dual, serão introduzidos alguns multiplicadores de Lagrange (α_i, α_i^*) não negativos conforme relatam Vapnik [9] e Smola e Schölkopf [22]. Logo, o problema, agora de maximização, na sua forma dual é dado pela expressão 5:

$$\begin{aligned} & \text{Maximizar } -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) \\ & \quad + y_i \sum_{i=1}^l (\alpha_i - \alpha_i^*) \\ & \text{Sujeito a } \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ & \quad \alpha_i, \alpha_i^* \in [0, C], \quad \forall i = 1, \dots, l \end{aligned} \quad (5)$$

e a função linear de aproximação passa a ser representada pela equação 6:

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \quad (6)$$

A formulação dual do problema SVR fornece como alternativa trabalhar em um espaço de alta dimensionalidade. Assim, pode-se realizar um mapeamento não linear dos dados de entrada para um espaço de dimensão maior, onde a regressão linear torna-se possível. Para isso, utiliza-se a abordagem baseada em funções Kernel $K(x, x') := \langle \varphi(x), \varphi(x') \rangle$, cuja introdução no problema de otimização faz com que este passe a ser descrito pela expressão 7:

$$\begin{aligned} & \text{Maximizar } -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) K(x_i, x_j) - \varepsilon \sum_{i=1}^l (\alpha_i + \alpha_i^*) \\ & \quad + y_i \sum_{i=1}^l (\alpha_i - \alpha_i^*) \\ & \text{Sujeito a } \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ & \quad \alpha_i, \alpha_i^* \in [0, C], \quad \forall i = 1, \dots, l \end{aligned} \quad (7)$$

e a função de aproximação da SVR não linear passa a ser representada pela equação 8.

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x_i, x) + b \quad (8)$$

Dentre as funções Kernel mais utilizadas no algoritmo SVR destacam-se o Kernel: linear, polinomial homogêneo, polinomial não homogêneo, sigmoideal e gaussiano.

5 Combinação de Previsões

A combinação, de acordo com Clemen [23], é uma metodologia atraente na obtenção de previsões, pois ao invés de escolher a melhor técnica, o problema passa a ser de quais técnicas podem ajudar na melhoria da acurácia.

Conforme Teixeira Jr [24], as previsões pontuais são combinadas linearmente, utilizando algum mecanismo ponderador os quais visam à minimização da variância dos resíduos combinados ou outra função residual (a ser definida pelo tomador de decisão). Os pesos podem ser fixos ou variáveis (não necessariamente positivos) ou somar uma unidade ou não. Destaca-se ainda que as médias, simples, ponderada ou harmônica também poderiam ser utilizadas na combinação de previsões.

Já Bates e Granger [25] propuseram o método de combinar os resultados dos modelos de previsão mais eficientes, com a finalidade de agregar as características mais relevantes de cada um dos modelos utilizados em uma determinada situação. Os autores sugeriram uma combinação, para a previsão da série temporal $\{y_t\}_{t=1}^T$, onde a previsão é combinada linearmente.

Foi utilizada no trabalho de Gupta e Wilton [26] a média aritmética das previsões individuais. Já no trabalho de Granger e Ramanathan [13] foi proposta, entre outros métodos, uma combinação linear sem restrição para os coeficientes e adição de uma constante. No trabalho de Terui e Dijk [27] é proposto o uso de filtros de Kalman para o cálculo dos coeficientes e que estes sejam variáveis ao longo do tempo.

No trabalho de Teixeira Jr [24] utilizou-se 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.

6 A metodologia híbrida

A comparação entre a modelagem linear ARIMA e da modelagem não linear SVR nos últimos anos vem despertando a atenção de cientistas em volta do mundo. Existem na literatura científica estudos comparativos entre o desempenho desses métodos, como por exemplo, [5], [11], [28], [29], [30], [31], [32], [33] e [34]. No entanto, não se pode afirmar que algum deles é um modelo universal e o mais apropriado para todas as situações.

Do fato de ser de interesse extrair completamente características lineares e não lineares de um conjunto de dados em um problema real a metodologia híbrida que tem ambas as capacidades demonstra ser uma boa estratégia de utilização, já que, combinando diferentes modelos, características distintas podem ser capturadas. Na literatura podem-se encontrar alguns trabalhos que visam a capturar estruturas lineares e não lineares por meio de metodologias híbridas, tais como: [1], [2], [3], [4], [12], [27], [35], [36] e [37].

Seja a série temporal $y_t = L_t + N_t + \varepsilon_t$, onde L_t denota a componente linear, N_t a componente não linear e $\varepsilon_t \sim i.i.d.N(0, \sigma^2)$. Inicialmente, ajusta-se um modelo ARIMA à série y_t , obtendo-se a sua previsão \hat{y}_t e a componente total de resíduo $e_t = y_t - \hat{y}_t$. Características lineares do ARIMA permitem representar a previsão de L_t por \hat{y}_t . Assim, a série de resíduos pode ser escrita na forma $e_t = y_t - \hat{L}_t$. Os testes de autocorrelação residual, independência estatística BDS e de estacionariedade Dickey-Fuller são aplicados à componente e_t a fim de se obter indicação que ela não pode ser classificada estatisticamente como ruído branco, de forma que os resíduos devem possuir estruturas lineares e/ou não lineares que podem ser modeladas. Nesse caso, a previsão da série observada resultará da soma das previsões dos resíduos e \hat{L}_t ($\hat{y}_t = \hat{L}_t + \hat{e}_t$). Os resíduos são decompostos ortogonalmente via Wavelet e suas componentes modeladas por SVR, aproveitando a sua eficiência na modelagem de séries oriundas de processos não lineares. Obtidas as previsões das componentes Wavelet, estas serão combinadas linearmente a fim de se obter a previsão \hat{e}_t .

Considere-se a sequência infinita $(\tilde{e})_{t \in \mathbb{Z}}$ obtida da série temporal $(e_t)_{t=1}^T$, onde T é a cardinalidade da série, a partir da igualdade $\tilde{e}_t = \begin{cases} e_t, & t \in \{1, \dots, T\} \\ 0, & t \in \mathbb{Z} \setminus \{1, \dots, T\} \end{cases}$. Segundo Teixeira Jr et al [19], a decomposição Wavelet de nível p de uma série temporal \tilde{e}_t em relação a uma determinada base ortonormal Wavelet é representada pela equação 9.

$$\tilde{e}_t \cong \tilde{\tilde{e}}_t = \sum_{n \in \Gamma_{V_{m_0}}} a_{m_0, n} \phi_{m_0, n} + \sum_{m=m_0}^{m_0+(p-1)} \left(\sum_{n \in \Gamma_{W_m}} d_{m, n} \omega_{m, n} \right) \quad (9)$$

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

As componentes de aproximação $\sum_{n \in \Gamma_{V_{m_0}}} a_{m_0, n} \phi_{m_0, n}$ e detalhe $\sum_{n \in \Gamma_{W_{m_0}}} a_{m_0, n} \phi_{m_0, n}$ provenientes da decomposição da série de resíduos foram obtidos a partir do aplicativo computacional Matlab 8.0, que oferece várias opções de bases Wavelet, tais como: as famílias daubechies (db), coiflets (coif) e symlets (sym), desenvolvidas por Ingrid Daubechies [20].

Neste trabalho, a série e_t foi decomposta em nível $p = 2$, que resulta em duas componentes de detalhe (D_1 e D_2) e uma de aproximação (A_2). Uma vez obtidas as componentes Wavelet de aproximação e detalhe para a série de resíduos, estas foram modeladas por SVR, resultando as previsões das componentes Wavelet de detalhe (\hat{D}_1 e \hat{D}_2) e aproximação (\hat{A}_2). No passo seguinte obteve-se a previsão do resíduo a partir da combinação linear descrita pela equação 10:

$$\hat{e}_t = \rho_1 \hat{A}_2 + \rho_2 \hat{D}_1 + \rho_3 \hat{D}_2 \quad (10)$$

onde ρ_1, ρ_2 e ρ_3 são parâmetros a serem determinados com a resolução do problema de programação não linear (expressão 11), cuja função objetivo é a raiz do erro

quadrático médio (RMSE) entre o resíduo e_t e a sua previsão combinada \hat{e}_t .

$$\text{Min } RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (e_t - \hat{e}_t)^2} \quad (11)$$

s. a. ρ_1, ρ_2 e ρ_3 são irrestritas

Na hipótese do modelo ARIMA ter capturado as estruturas lineares da série e o SVR as não lineares, então $\hat{N}_t = \hat{e}_t$ e a previsão da série temporal pela modelagem híbrida proposta, a qual denominar-se-á aqui ARIMA-SVR, é representada por $\hat{y}_t = \hat{L}_t + \hat{N}_t$.

Em resumo, a metodologia híbrida proposta consiste nos passos representados no fluxograma da figura 1.

Para a avaliação do método foram utilizados os erros RMSE (Raiz do Erro Quadrático Médio) e MAE (Erro Absoluto Médio), representados nas equações 12.

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=t_0}^T (y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{T} \sum_{t=t_0}^T |y_t - \hat{y}_t| \quad (12)$$

7 Materiais e métodos

A Itaipu é uma empresa que se preocupa com a segurança de suas operações, mantendo um monitoramento do desempenho de sua estrutura, dentre outras formas, por meio da análise de dados coletados pela instrumentação civil da barragem. Dentre os diversos fatores que se pode considerar em uma barragem de concreto, os movimentos relativos entre blocos, e entre estes e suas fundações, surgem como um assunto de extremo interesse e análise.

Os blocos estão sujeitos principalmente à ação do nível do lago e das variações de temperatura. Por exemplo: no verão ocorrem dilatações no concreto, que provocam uma tendência de deformação do bloco para a direção montante, que por sua vez pode fazer com que aumentem as tensões de compressão no pé de montante dos blocos. No inverno o concreto se contrai, provoca uma tendência de deformação do bloco a jusante. Pode-se então identificar um comportamento cíclico da estrutura, intimamente sujeito a condições ambientais da região [38].

Por conseguinte, fazer a previsão de deslocamentos, mais precisamente dos deslocamentos relativos entre blocos, é de grande relevância para ajudar a compreender e prever o comportamento futuro da barragem de concreto de Itaipu.

O comportamento dos deslocamentos diferenciais entre blocos se dá segundo três direções ortogonais entre si, recalques diferenciais entre blocos, deslocamentos cisalhantes horizontais e movimentos de abertura-fechamento das juntas. A base de alongâmetro é utilizada para medir deslocamentos horizontais e verticais entre juntas, aberturas de fissuras de barragens, galerias, túneis, maciços rochosos e edificações. Do instrumento base de alongâmetro se obtêm as seguintes medidas: duas no piso,

abertura entre juntas e deslizamento no sentido montante-jusante, e duas na parede, recalque entre os blocos e abertura entre os blocos. Um vídeo ilustrativo do movimento e de como a base de alongâmetro funciona pode ser visto em: <https://www.youtube.com/user/CeasbPTI/videos>.

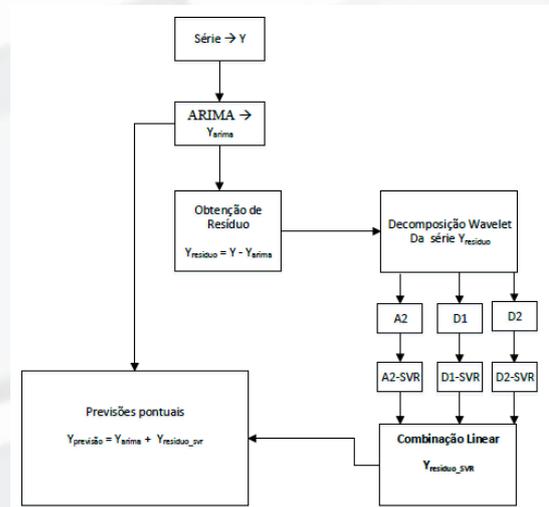


Figura 1: Fluxograma ARIMA-SVR.

Os dados que formam a série temporal (y_t) são provenientes da base de alongâmetro que mede a abertura e o fechamento, o recalque e o deslizamento entre os blocos.

A figura 2 mostra o comportamento dos deslocamentos relativos do bloco I11, nas direções de abertura entre juntas e deslizamento (que são os deslocamentos horizontais), da base de alongâmetro JS-I-35:

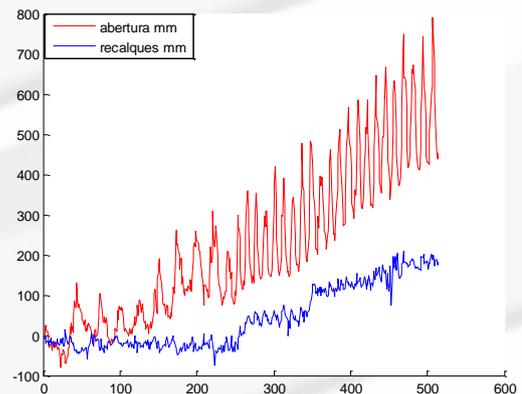


Figura 2: deslocamentos relativos do bloco I11.

A série escolhida para aplicação da metodologia é bivariada e contém os dados referentes ao período de 1982-2014 do instrumento base de alongâmetro (JS-I-35), localizado no bloco I-11/10 na Barragem de Ligação Esquerda, entre a Estrutura de Desvio e a Barragem de Enrocamento, e que medem a abertura (JS-I-35-X) e os recalques (JS-I-35-Y) entre

os blocos 10 e 11. Os valores são dados em milésimos de milímetros.

Na primeira etapa desse trabalho foi feito o tratamento das informações, para serem estruturadas de acordo com a metodologia de previsão. Optou-se por trabalhar com o período de 1992-2013, com valores mensais.

Para ambas as séries foram deixadas para treino o período de 1992-2012 totalizando 252 valores e os 12 últimos valores o ano de 2013 para teste. A metodologia usada é de um passo à frente.

Primeiramente foi realizada a modelagem ARIMA para as séries JS-I-35-X e JS-I-35-Y. Em seguida foi obtido o resíduo para cada série, denotado por r-JS-I-35-X e r-JS-I-35-Y. Para cada série de resíduo foi aplicada a decomposição Wavelet, obtendo-se assim as componentes que foram modeladas via SVR e combinadas para obtenção da previsão do resíduo para posterior obtenção das previsões pontuais.

Para modelagem ARIMA foi utilizado o software *Eviews* sendo feita a análise gráfica da série temporal e identificado o modelo apropriado a partir das funções de autocorrelação e autocorrelações parciais. Os valores encontrados se encontram na Tabela 1.

Na decomposição Wavelet e na modelagem SVR foi utilizada uma caixa de ferramentas em *Matlab* que é uma linguagem de alto nível e oferece um ambiente interativo que permite realizar tarefas de computação intensiva. Na Tabela 2 encontram-se as bases utilizadas para decomposição Wavelet, escolhidas após alguns testes, e as respectivas componentes.

Tabela 1: Modelo ARIMA.

SÉRIE	MODELO	Equação
JS-I-35-X	ARIMA (36,0,11)	$W_t = 0.418W_{t-1} - 0.040W_{t-4} +$ $+ 0.398W_{t-11} + 0.285W_{t-36} +$ $+ 0.141a_{t-2} + 0.272a_{t-5} +$ $- 0.377a_{t-11}$
JS-I-35-Y	ARIMA (9,0,16)	$W_t = 0.894W_{t-1} - 0.163W_{t-6} +$ $+ 0.279W_{t-9} - 0.362a_{t-1} +$ $+ 0.324a_{t-6} - 0.154a_{t-16}$

Para o uso do SVR foi utilizada uma biblioteca para máquinas de vetores suporte para *Matlab* denominada LIBSVM [39], sendo que a biblioteca LIBSVM é atualmente um dos aplicativos de software mais utilizados para regressão em vetores suporte. Ao treinar o modelo SVR, existem alguns parâmetros para serem escolhidos que influenciam no desempenho do modelo. Portanto, a fim de obter um modelo ajustado, tais parâmetros têm que ser selecionados adequadamente. Foram realizadas várias simulações para as séries supracitadas. Estão apresentados na Tabela 3 os parâmetros que melhor se ajustaram, baseado nos dados de treinamento e que foram escolhidos de acordo com as medidas de acurácia *mean square error* (MSE) e *Squared correlation coefficient* (SCC).

Tabela 2: Decomposição Wavelet.

SÉRIES	Base	Aproximação	Detalhes
r-JS-I-35-X	Daubechies 8	r-A2-X	r-D1-X r-D2-X
r-JS-I-35-Y	Daubechies 8	r-A2-Y	r-D1-Y r-D2-X

Em seguida foi feita a combinação das séries dos resíduos previstas via SVR. Das componentes: r-A2-X, r-D1-X e r-D2-X obtiveram-se os parâmetros, $\rho_{r_{A_2X}} = 1.015$, $\rho_{r_{D_1X}} = 1.069$ e $\rho_{r_{D_2X}} = 0.5616$, e, para a série dos resíduos r-A2-Y, r-D1-Y e r-D2-Y foram obtidos os parâmetros $\rho_{r_{A_2Y}} = 1.0099$, $\rho_{r_{D_1Y}} = 1.0360$ e $\rho_{r_{D_2Y}} = 1.0501$. Tais parâmetros foram ajustados por meio da otimização de um problema de programação não linear, cuja função objetivo é a raiz do erro quadrático médio (RMSE), a partir do Solver do *Excel*.

8 Resultados

Neste trabalho foi considerada para análise da previsão a metodologia um passo à frente. As medidas de acurácia utilizadas para avaliação foram: raiz quadrada do erro quadrático médio (RMSE) e a média dos erros absolutos (MAE).

A Tabela 4 apresenta as medidas de acurácia MAE e RMSE correspondentes às previsões da série JS-I-35-X, para modelagem ARIMA, SVR e para a metodologia híbrida. Observa-se que a metodologia híbrida ARIMA-SVR obteve um ganho preditivo de 18,76% quando comparada com o método ARIMA e 27,41% quando comparada com o método SVR para a medida de acurácia MAE, e de 29,15% quando comparada com o método ARIMA e 56,14% quando comparada com o método SVR para a medida de acurácia RMSE.

A comparação entre o valor real e o valor previsto via ARIMA, SVR e ARIMA-SVR para os doze valores fora da amostra de dados da série JS-I-35-X são apresentados na Figura 3.

São mostradas na Tabela 5 as medidas de acurácia MAE e RMSE para as previsões da série JS-I-35-Y dos modelos ARIMA, SVR e para a metodologia híbrida ARIMA-SVR. Nesta série a metodologia híbrida ARIMA-SVR obteve um ganho preditivo de 91,25% quando comparada com o método ARIMA e 70,14% quando comparada com o método SVR para a medida de acurácia MAE, e de 102,56% quando comparada com o método ARIMA e 88,56% quando comparada com o método SVR para a medida de acurácia RMSE.

A comparação entre o valor real e o valor previsto via ARIMA, SVR e ARIMA-SVR para os doze valores fora da amostra de dados da série JS-I-35-Y é apresentada na Figura 4.

Tabela 3: Parâmetros SVR.

SÉRIES	Kernel	Gamma	Degree	Coef0	Cost	Epsilon	Janela
r-A2_X	RBF	0.5	-	-	100	0.02	4
r-D1_X	RBF	0.8	-	-	100	0.02	4
r-D2_X	Polinomial	-	3	800	1	0.402	2
r-A2_Y	RBF	1	-	-	100	0.0002	4
r-D1_Y	RBF	2	-	-	10	0.02	4
r-D2_Y	RBF	0.9	-	-	5	0.002	2

Tabela 4 : Comparação da previsão para a série JS-I-35-X.

Série	Método	Medidas de acurácia	
		MAE	RMSE
JS-I-35-X	ARIMA	33,865	45,648
JS-I-35-X	SVR	37,608	55,189
JS-I-35-X	ARIMA-SVR	29,515	35,346

Tabela 5: Comparação da previsão para a série JS-I-35-Y.

Série	Método	Medidas de acurácia	
		MAE	RMSE
JS-I-35-Y	ARIMA	10,094	12,344
JS-I-35-Y	SVR	8,980	11,490
JS-I-35-Y	ARIMA-SVR	5,278	6,094

9 Conclusão

Com o intuito de diminuir erros de previsão, o estudo de séries temporais vem crescendo bastante nos últimos anos, pois é sensato que a maximização da eficiência do método preditivo ou da metodologia que leva às previsões de menor erro propiciará a técnicos e engenheiros uma ferramenta de apoio nos processos de tomadas de decisões.

Dos vários métodos de análise e previsão de séries temporais existentes, o modelo ARIMA é sem dúvida o mais popular deles, entretanto nos últimos anos o método de Regressão por Vetores Suporte (SVR) surge como uma novidade promissora nesse campo, demonstrando sua boa capacidade de generalização por meio da obtenção de bons resultados. Embora estas duas técnicas mostrem seu

potencial, não se pode dizer que uma técnica é melhor que a outra. Portanto, a metodologia híbrida aqui apresentada serve como uma alternativa na obtenção de melhores resultados.

Neste trabalho, foi proposta uma modelagem híbrida para previsão de séries temporais usando os modelos ARIMA, decomposição Wavelet, a regressão de vetores suporte (SVR) e a combinação de previsões. A eficácia do método foi verificada por meio da aplicação em um problema real que mede os deslocamentos relativos entre blocos.

Os dados que formam a série temporal que é bivariada são provenientes do instrumento denominado base de alongâmetro e que mede a abertura (JS-I-35-X) e os recalques (JS-I-35-Y) entre os blocos 10 e 11 da barragem da Usina Hidrelétrica de Itaipu. Os resultados obtidos mostram que a metodologia híbrida ARIMA-SVR obteve ganhos preditivos para ambas as séries testadas, comprovando que a metodologia proposta para previsões de séries temporais contribui para a melhora da acurácia nas previsões pontuais, nesse caso comparadas com as técnicas ARIMA e SVR.

Agradecimentos

Às instituições: Itaipu Binacional, Programa de Pós Graduação em Métodos Numéricos em Engenharia (PPGMNE) da Universidade Federal do Paraná (UFPR), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Centro de Estudos Avançados em Segurança de Barragens (CEASB), Fundação Parque Tecnológico Itaipu (FPTI), e, em especial, aos engenheiros: Cláudio Neumann Júnior e Dimilson Coelho.

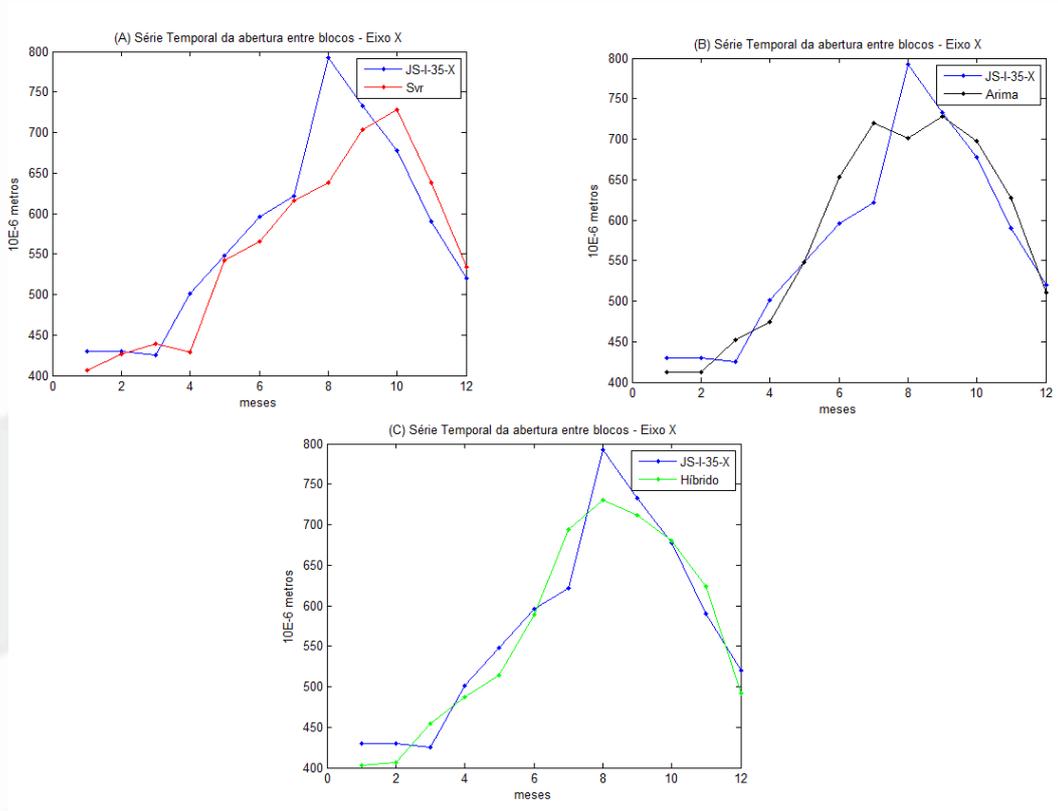


Figura 3: Gráfico da amostra de teste e previsões para a série JS-I-35-X.

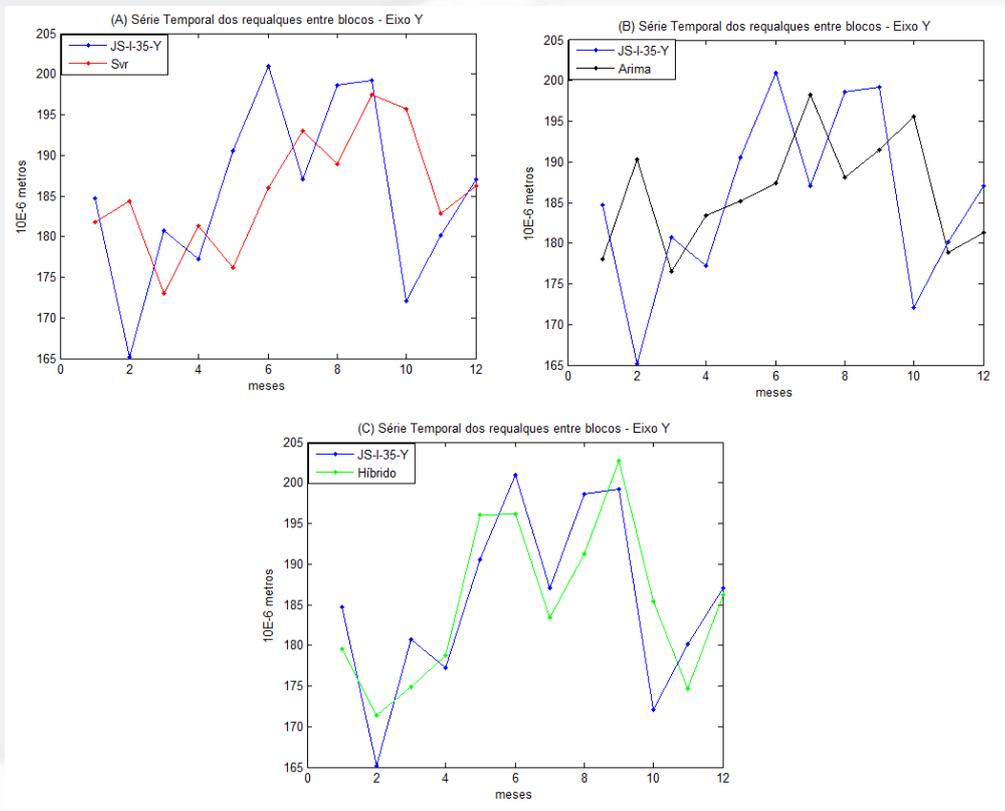


Figura 4: Gráfico da amostra de teste e previsões para a série JS-I-35-Y.

Referências

- [1] C. H. Aladag, E. Egrioglu e C. Kadilar. Forecasting nonlinear time series with a hybrid methodology, **Applied Mathematics Letters**, pp. 1467-1470, 2009.
- [2] C. N. Babu e B. E. Reddy. A moving-average filter based hybrid ARIMA-ANN model for forecasting time series data, **Applied Soft Computing**, pp. 27-38, 2014.
- [3] P. F. Pai e C. S. Lin. A hybrid ARIMA and support vector machines model in stock price forecasting, **The International Journal of Management Science**, pp. 497-205, 2005.
- [4] R. Alwee, S. M. H. Shamsuddin e R. Sallehuddin. Hybrid Support Vector Regression and Autoregressive Integrated Moving Average Models Improved by Particle Swarm Optimization for Property Crime Rates Forecasting with Economic Indicators, **The ScientificWorld Journal**, pp. 1-11, 2013.
- [5] O. Kosanan e N. Kantanantha. Thailand's Rubber Production Forecasting Comparison, **Proceedings of the International MultiConference of Engineers and Computer Scientists 2014 Vol II**, Hong Kong, 2014.
- [6] G. E. Box e G. M. Jenkins. **Time Series Analysis: Forecasting and Control**, Oakland, Califórnia: Holden-Day, 1970.
- [7] K. Suresh e S. K. Priya. Forecasting sugarcane yield of Tamilnadu using ARIMA models, **Sugar Tech**, vol. 13, nº 1, pp. 23-26, 2011.
- [8] C. Babu e B. Reddy. Predictive data mining on average global temperature using variants of ARIMA models, em **2012 International Conference on Advances in Engineering, Science and Management**, ICAESM, 2012.
- [9] V. N. Vapnik. **The Nature of Statistical Learning Theory**, New York: Springer - Verlag, 1995.
- [10] V. Vapnik, S. E. Golowich e A. Smola. Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing, **Advances in Neural Information Processing Systems 9**, pp. 281-287, 1996.
- [11] V. Anandhi e R. M. Chezian. Support Vector Regression to Forecast the Demand and Supply of Pulpwood, **International Journal of Future Computer and Communication**, vol. 2, 2013.
- [12] W. C. Hong. Electric load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm, **Energy**, pp. 5568-5578, 2011.
- [13] C. W. Granger, R. Ramanathan. Improved Methods of Forecasting, **Journal of Forecasting**, vol. 3, pp. 197-204, 1984.
- [14] G. Hollauer, J. V. Issler e H. H. Notini. Prevendo o crescimento da produção industrial usando um número limitado de combinações de previsões, **Economia Aplicada**, vol. 12, pp. 177-198, 2008.
- [15] A. E. Faria, E. Mubwandarikwa. Multimodality on the Geometric Combination of Bayesian Forecasting Models, **International Journal of Statistics and Management System**, pp. 1-25, 2008.
- [16] P. P. Morettin e C. M. Tolo. **Análise de Séries Temporais**, São Paulo: Blucher, 2006.
- [17] S. Mallat. **A Wavelet Tour of Signal Processing**, San Diego: Academic Press, 2009.
- [18] N. Levan e C. S. Kubrusly. A Wavelet Time-Shift Detail Decomposition, **Mathematics and Computers in Simulation**, vol. 63, pp. 73-78, 2003.
- [19] L. A. Teixeira Jr, J. F. M. Pessanha, M. L. Menezes, K. M. Cassiano e R. C. Souza. Redes Neurais Artificiais E Decomposição Wavelet Na Previsão Daradiação Solar Direta, **Simpósio Brasileiro de Pesquisa Operacional**, Rio de Janeiro, 2012.
- [20] I. Daubechies. Orthonormal basis of compactly supported wavelets, **Comm. Appl. Math.**, pp. 909-996, 1988.
- [21] C. S. Kubrusly, N. Levan. Abstract Wavelets Generated by Hilbert Space Shift Operators, **Adv. Math. Sci. Appl.**, vol. 14, pp. 643-660, 2006.
- [22] A. J. Smola e B. Schölkopf. A Tutorial on Support Vector Regression, **NeuroCOLT Technical Report**, 1998.
- [23] R. T. Clemen. Combining forecasts: A review and annotated bibliography, **International Journal of Forecasting**, pp. 559-583.
- [24] L. A. Teixeira Jr. Combinação SSA-Wavelet de métodos preditivos com ajuste numérico MINIMAX, na geração de previsões e de cenários, **Puc**. Rio de Janeiro, 2013.
- [25] J. M. Bates, C. W. J. Granger. The Combining of Forecasts., **Operational Research Quarterly**, vol. 20, nº 4, pp. 451-468, 1969.
- [26] S. Gupta, P. C. Wilton. Combination of Forecasts: An Extension, **Management Science**, vol. 33, nº 3, pp. 356-372, 1987.
- [27] N. Terui, H. K. Dijk. Combined forecasts from linear and nonlinear time series models, **International Journal of Forecasting**, p. 421-438, 2002.
- [28] M. K. Okasha. Using Support Vector Machines in Financial Time Series Forecasting, **International Journal of Statistics and Applications**, pp. 28-39, 2014.
- [29] K. Kandananond. A Comparison of Various Forecasting Methods for Autocorrelated Time Series, **International Journal of Engineering Business Management**, pp. 1-6, 2012.
- [30] X. Zhang, T. Zhang, A. A. Young e X. Li. Applications and Comparisons of Four Time Series Models in Epidemiological Surveillance Data, **PLoS ONE**, 2014.
- [31] M. Moharrampour, A. Mehrabi, H. Hajikandi, S. Sohrabi e J. Vakili. Comparison of Support Vector Machines (SVM) and Autoregressive integrated moving average (ARIMA) in daily flow forecasting, **Journal of River Engineering**, 2013.
- [32] L. L. Teixeira, S. B. Rodrigues, P. H. Siqueira, A. C. L. d. Silva e L. A. Teixeira Jr. Comparativo entre ARIMA, redes neurais artificiais, máquinas de vetores suporte para regressão e combinação linear de previsões de vazões mensais no posto 266-itaipu, **XXI SIMPEP**, 2014.

- [33] E. K. Lafdani, A. M. N. Nia e A. Ahmadi. Daily suspended sediment load prediction using artificial neural networks and support vector machines, **Journal of Hydrology**, vol. 478, pp. 50-62, 2013.
- [34] Z. He, X. Wen, H. Liu , J. Du. A comparative study of artificial neural network, **Journal Hydrology**, vol. 509, pp. 379-386, 2014.
- [35] F. M. Tseng, H. C. Yu, G. H. Tzeng. Combining neural network model with seasonal time series ARIMA model, **Technological Forecasting & Social Change**, pp. 71-87, 2002.
- [36] P. G. Zhang. Time series forecasting using a hybrid ARIMA and neural network model, **Neurocomputing**, pp. 159-175, 2003.
- [37] B. Zhu e Y. Wei. Carbon price forecasting with a novel hybrid ARIMA and least Squares support vector machines methodology, **Omega**, pp. 517-524, 2013.
- [38] C. I. Osako. A manutenção dos drenos nas Fundações de Barragens o caso da Usina Hidrelétrica de Itaipu, Dissertação Curso de Pós-Graduação em Construção Civil, Setor de Tecnologia, **Universidade Federal do Paraná**, 2002.
- [39] C. C. Chang, C. J. Lin. LIBSVM: a library for support vector machines, [Online]. Available: <http://www.csie.ntu.edu.tw/~cjlin/libsvm>. [Acesso em 5 Janeiro 2014].