

PREVISÃO DE SÉRIES TEMPORAIS DE SUBPRESSÃO DE BARRAGENS COM FILTRAGEM SSA E REGRESSÃO MÚLTIPLA COM MODELAGEM ARIMA DOS RESÍDUOS

Júlio C. Royer¹, Volmir E. Wilhelm² e Josiele Patias³

1: Campus Foz do Iguaçu
Instituto Federal do Paraná
Avenida Araucária, 780, Vila A, CEP 85856-000, Foz do Iguaçu, PR, Brasil
e-mail: julio.royer@ifpr.edu.br, web: <http://www.ifpr.edu.br>

2: Departamento de Engenharia de Produção
Centro Politécnico
Universidade Federal do Paraná
Av. Francisco H dos Santos, 210, Jardim das Américas, CEP 81531-980, Curitiba, PR, Brasil
e-mail: volmirw@gmail.com web: <http://www.ufpr.br>

3: Departamento de Engenharia Civil
Itaipu Binacional
Av. Tancredo Neves, 6.731, CEP 85856-970, Foz do Iguaçu, PR, Brasil
e-mail: jpatias@itaipu.gov.br web: <http://www.itaipu.gov.br>

Palavras-chave: Subpressão em Barragens, Regressão Dinâmica, Séries Temporais, ARIMA, SSA

Resumo. *Uma questão de segurança de barragens é monitorar a subpressão nas fundações e mantê-la dentro dos limites de projeto ao longo de todo o seu tempo de operação. Usualmente as decisões de manutenção levam muito em consideração o comportamento histórico da estrutura, e para isso são definidos limites de alerta baseados no máximo histórico, por exemplo. Mas, em função da sazonalidade, a definição de limites de alertas fixos com base no máximo histórico, por exemplo, não ajuda a identificar aumentos inesperados quando deveria estar próximo do mínimo. Este trabalho apresenta uma comparação entre as técnicas de regressão dinâmica e regressão dinâmica com modelagem de série temporal dos resíduos (ARIMA), com e sem a aplicação de SSA (Análise Espectral Singular), com base em séries temporais de temperatura ambiente, temperatura da água e nível do reservatório, utilizadas para definir limites variáveis para alertas de subpressões, compondo uma metodologia em desenvolvimento para a Usina Hidrelétrica de Itaipu para aproximar a previsão da série temporal de subpressão, identificando precocemente desvios do seu comportamento histórico, permitindo a antecipação de procedimentos de verificação local e manutenção, otimizando o trabalho da equipe de segurança de barragens.*

1. INTRODUÇÃO

As fundações rochosas sobre as quais são assentadas as barragens de concreto em geral permitem a infiltração de água, seja pela permeabilidade da rocha, seja pelas fraturas existentes devido à sua formação geológica. Essa água infiltrada gera uma pressão ascendente (chamada de subpressão) que alivia o peso da estrutura e diminui o coeficiente de atrito das descontinuidades, trazendo riscos à segurança da barragem se ultrapassar seus limites de projeto [1,2]. Essa subpressão depende de vários fatores, entre os quais cabe destacar a condutividade hidráulica do maciço rochoso, que é estimado durante a construção por sondagens geológicas e ensaios hidráulicos, e fatores de projeto da barragem, como nível do lago e geometria da barragem, especialmente no que diz respeito ao projeto da cortina de injeção (para diminuir a condutividade hidráulica do maciço), e cortina de drenagem (para aliviar a subpressão). Outro detalhe que precisa ser levado em consideração é que as condições de condutividade hidráulica do maciço sofrem alterações ao longo do tempo, em função de efeitos químicos ou mecânicos, a exemplo de erosões ou depósito de materiais nos caminhos da água ou alterações de PH e de viscosidade da água [3,4,5]. Em função disso a subpressão precisa ser monitorada durante toda a vida útil da barragem, o que, em caso de grandes barragens gera uma grande quantidade de dados, o que, mesmo dispondo de gráficos torna a análise complexa por parte dos engenheiros de segurança. Para facilitar esse trabalho, os sistemas atuais contam com níveis de alerta para cada instrumento.

No caso da barragem de Itaipu há níveis de alerta estáticos. Por exemplo, alerta amarelo se a subpressão ultrapassou o limite histórico e alerta vermelho se alcançou o limite definido em projeto para aquele piezômetro. Mas boa parte dos piezômetros, especialmente os de contato entre o concreto e a rocha sofrem uma influência sazonal importante, e por isso os alertas estáticos não ajudam a identificar precocemente um aumento no nível de subpressão se está próximo ao mínimo sazonal, e pode soar um falso alerta se ultrapassou levemente o limite histórico se for acompanhado por recordes de temperatura, por exemplo.

Se, por um lado, os alertas estáticos não permitem esse tipo de identificação precoce, por outro, só foi encontrado um trabalho na literatura que define uma metodologia para a definição de alertas dinâmicos para subpressão em fundações de barragens [6]. Nele os autores usam regressão múltipla para especificar limites dinâmicos de alerta, desde que as variáveis de entrada consigam explicar pelo menos 70% da variação da subpressão (R^2 ajustado $\geq 0,7$). Caso contrário são utilizados a média e o desvio padrão, especificando os limites do intervalo de confiança de 95%. Mas essa técnica não funcionou em Itaipu, onde os valores da regressão múltipla para esses piezômetros explica algo próximo a 50% da variação da subpressão.

Este trabalho apresenta o estágio atual da metodologia que está em desenvolvimento para Itaipu Binacional, com o uso das técnicas de Análise Espectral Singular (SSA – Singular Spectrum Analysis) [7], para reduzir o nível de ruído da série temporal de subpressão, e Regressão Múltipla com modelagem ARIMA dos resíduos [8,9].

O restante deste artigo está organizado da seguinte forma: a sessão 2 apresenta a modelagem ARIMA de séries temporais, a sessão 3 aborda a técnica SSA aplicada para a redução dos ruídos da série, a sessão 4 descreve a modelagem de Função de Transferência Linear usando Regressão Dinâmica e Modelagem ARIMA da série temporal dos resíduos. Os resultados

obtidos com a aplicação da técnica descrita são apresentados na sessão 5, seguida das conclusões e referências.

2. MODELAGEM DE SÉRIES TEMPORAIS ARIMA

Uma série temporal é uma sequência de dados observados a intervalos regulares de tempo, e com algum tipo de dependência em relação aos valores anteriores. A análise dessa dependência é o ponto central da análise de séries temporais [8,10,11]. Uma técnica clássica para a modelagem de séries temporais é o modelo ARIMA, proposto por Box & Jenkins. Seja uma série temporal da variável Y , representada pelos valores $\{Y_1, Y_2, \dots, Y_t, \dots, Y_{n-1}, Y_n\}$, ou como uma série $\{Y_t, t=1, 2, \dots, n\}$. Essa série pode conter componentes autorregressivas (AR), componentes de médias móveis (MA), e/ou integração (I). A série também pode conter componentes sazonais autorregressivas, de médias móveis ou integração. Para conseguir uma forma de expressão mais compacta, usa-se frequentemente os operadores de Retardo B , definido por $B^m Y_t = Y_{t-m}$, e de diferenciação ∇ (nabla), definido por $\nabla^d = (1-B)^d Y_t$.

Por exemplo, em uma série autorregressiva de ordem 2, AR(2), cada valor de Y_t sofre influência dos dois valores anteriores, Y_{t-1} e Y_{t-2} . Essa série pode ser escrita como $Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + a_t$, onde C é uma constante, φ_1 e φ_2 são coeficientes que indicam o quanto Y_{t-1} e Y_{t-2} influenciam Y_t , e a_t é um choque estocástico aleatório. Usando operador de retardo, essa mesma série pode ser expressa como $\phi(B)Y_t = C + a_t$, onde $\phi(B)$ é o polinômio característico, que para um AR(2) é $\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2$.

Podemos expressar um modelo autorregressivo de ordem p , AR(p), como [10]:

$$Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + a_t,$$

ou

$$\phi(B)Y_t = C + a_t$$

onde $\phi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p$.

Da mesma forma, em um modelo de médias móveis de grau 2, MA(2), Y_t é influenciado pelos dois impactos estocásticos a_t anteriores, e podemos escrevê-lo como:

$Y_t = C + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2}$, ou, usando o operador de retardo: $Y_t = C + \theta(B)a_t$, onde $\theta(B)$ é o polinômio característico, que neste caso é $\theta(B) = 1 - \theta_1 B - \theta_2 B^2$. Igualmente para generalizar para o MA(q) basta fazer $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$.

Um modelo de série temporal pode conter p componentes autorregressivas e q componentes de média móvel, formando um modelo ARMA(p,q), expresso na forma

$$Y_t = C + \varphi_1 Y_{t-1} + \varphi_2 Y_{t-2} + \dots + \varphi_p Y_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q},$$

ou

$$\phi(B)Y_t = C + \theta(B)a_t$$

Para uma série temporal ser representada por um modelo ARMA, ela precisa ser estacionária. Caso a série original não seja estacionária, é possível aplicar transformações na série, e uma

transformação frequentemente usada é diferenciar a série. Assim o modelo que representa a série é o ARIMA(p,d,q), onde p é o grau do polinômio $\phi(B)$, q é o grau do polinômio $\theta(B)$, e d é o grau de diferenciação ∇^d . A representação da série usando o operador de retardo fica $\phi(B)\nabla^d Y_t = \theta(B)a_t$ [10]. É possível que em uma série temporal o valor de Y_t pode ser influenciado significativamente pelo valor de Y_{t-s} , ou a_{t-s} , apresentando comportamento periódico, ou sazonal. Também pode ser necessário diferenciar a série sazonalmente. A representação de um modelo ARIMA(p,d,q)(P,D,Q)_s, em notação de operador de retardo, é:

$$\Phi(B^s)\phi(B)\nabla_s^D\nabla^d Y_t = \Theta(B^s)\theta(B)a_t$$

onde s é a sazonalidade (número de amostras em um ciclo),

$$\begin{aligned}\phi(B) &= 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p, \\ \theta(B) &= 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q, \\ \Phi(B) &= 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_q B^{Ps}, \\ \Theta(B) &= 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \dots - \Theta_q B^{Qs}, \\ \nabla^d &= (1 - B)^d \quad \text{e} \\ \nabla_s^D &= (1 - B^s)^D.\end{aligned}$$

3. ANÁLISE ESPECTRAL SINGULAR

Outra técnica usada em análise e previsão de séries temporais é a análise espectral singular (SSA – Singular Spectrum Analysis). Ela combina elementos da análise clássica de séries temporais, estatística multivariada, sistemas dinâmicos e processamento de sinais. A SSA tem por base a decomposição da série original em um conjunto de componentes interpretáveis, como uma tendência, componentes oscilatórias e um ruído sem estrutura. Ela funciona fazendo uma decomposição de valor singular (SVD – Singular Value Decomposition) de uma matriz trajetória, construída a partir da série temporal. Não é necessário que a série seja estacionária ou que tenha parâmetros conhecidos, o que torna a técnica aplicável a muitas áreas distintas. SSA pode ser aplicada para encontrar tendências, suavizar séries temporais, encontrar componentes sazonais e ciclos, e remoção de ruídos [12,13]. A aplicação SSA é dividida em suas etapas: a decomposição e a reconstrução.

3.1. Decomposição

A etapa da decomposição de uma série temporal $\{Y_1, Y_2, \dots, Y_N\}$ pode ser subdividida em duas subetapas: a incorporação e a decomposição em valores singulares (SVD – Singular Value Decomposition).

Na subetapa de incorporação a matriz trajetória é construída, usando uma janela de tamanho L que percorre a série temporal. A cada passo uma coluna da matriz trajetória, com L linhas é composta. O total de colunas é $K = N - L + 1$.

$$X = [X_1, \dots, X_K] = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & \dots & y_K \\ y_2 & y_3 & \dots & y_{K+1} \\ \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & \dots & y_N \end{pmatrix}$$

O tamanho da janela L fica no intervalo $2 \leq L \leq N-1$. A escolha de L é tópico de pesquisa, sendo que [7] sugere que L seja próximo a $N/2$. Já [14] mostra que o tamanho da janela pode ser significativamente menor, com resultados melhores, e sugere uma faixa de valores, dado por $L = (\ln N)^c$, com $1,5 \leq c \leq 3$, e $L \leq N/2$.

Na subetapa SVD, a matriz trajetória X é decomposta em uma soma de matrizes E_i , com $i=1, \dots, d$, onde d é o número de autovalores λ não nulos:

$$X = \sum_{i=1}^d E_i = \sum_{i=1}^d \sqrt{\lambda_i} U_i V_i^T \quad (1)$$

Onde λ_i e U_i são os autovalores e autovetores da matriz $S = XX^T$, enquanto os vetores $V_i = \frac{X^T U_i}{\sqrt{\lambda_i}}$, para $i = 1, \dots, d$ correspondem às componentes principais da matriz S . Os conjuntos λ_i , U_i e V_i são chamados de autotriplas, e são ordenados em ordem decrescente de autovalor [15].

3.2. Reconstrução

A etapa da reconstrução também pode ser dividida em duas subetapas: agrupamento e média diagonal. Na subetapa do agrupamento, as d componentes E_i são agrupadas de acordo com suas características, em $m < d$ grupos. Por exemplo, podem ser formados três grupos, com componentes de tendência, sazonalidade e de ruído. Assim, para um grupo formado por p componentes, cujos índices são $I = \{i_1, \dots, i_p\}$, a matriz resultante W_I é dada por: $W_I = E_{i_1} + \dots + E_{i_p}$

São computadas m matrizes resultantes para os grupos I_1, \dots, I_m . Logo, a expressão (1) pode ser reescrita como:

$$X = W_{I_1} + \dots + W_{I_m}$$

Na subetapa da média diagonal, cada matriz W_{I_i} , $i = 1, \dots, m$, da expressão acima é convertida em uma série temporal de tamanho N . O procedimento é realizado com o cálculo da média dos elementos w_{jk} onde a soma $j+k$ produz o mesmo valor.

Seja a série $\{y_t^{(i)}, t = 1, \dots, N\}$ a componente da série temporal $\{y_t, t = 1, \dots, N\}$ gerada a partir da matriz elementar W_{I_i} , e $w_{jk}^{(i)}$ o elemento da linha j e coluna k de W_{I_i} . Considere $L^* = \min(L, K)$ e $K^* = \max(L, K)$. O cálculo de $y_t^{(i)}$ é definido pela expressão (2) [16]:

$$y_t^{(i)} = \begin{cases} \frac{\sum_{j=1}^t w_{j,t-j+1}^{(i)}}{t}, \text{ para } 1 \leq t \leq L^* \\ \frac{\sum_{j=1}^{L^*} w_{j,t-j+1}^{(i)}}{L^*}, \text{ para } L^* \leq t \leq K^* \\ \frac{\sum_{j=t-K^*+1}^{N-K^*+1} w_{j,t-j+1}^{(i)}}{N-K^*+1}, \text{ para } K^* \leq t \leq N \end{cases} \quad (2)$$

A contribuição da componente W_{ii} na composição da série temporal final pode ser avaliada pela razão $\frac{\sum_{j=1}^{p_i} (\lambda_{I_{ij}})^{1/2}}{\sum_{l=1}^d (\lambda_l)^{1/2}}$. As componentes $y_t^{(i)}$ geradas também podem ser testadas para verificar se são ruído branco, e neste caso, não serão consideradas na composição da série temporal final. Assim a série temporal final terá menos ruídos do que a série temporal original.

4. FUNÇÃO DE TRANSFERÊNCIA LINEAR

Em um conjunto de observações de uma população em que se observa duas variáveis X e Y, é possível traçar uma regressão linear entre as variáveis para se saber se a variação de X influencia a variação de Y. Por exemplo, se X é a altura do pai, e Y a altura do filho em idade adulta, pode-se querer identificar se (e o quanto) a altura do pai influencia na altura do filho. Isso pode ser feito por meio da regressão linear simples entre X e Y [17].

$$Y_i = C + bX_i + a_i \quad (3)$$

Tem-se assim uma função de transferência linear que é a equação de uma reta, onde C é uma constante, b indica o fator de impacto de X_i em Y_i , e a_i é um resíduo aleatório.

Se o objetivo é saber se (e de que modo) a altura do filho adulto é influenciada pela altura do pai (X_1), e da mãe (X_2), pode-se usar a regressão múltipla.

$$Y_i = C + b_1X_{1,i} + b_2X_{2,i} + a_i \quad (4)$$

No caso de dados de séries temporais, as equações (3) e (4) só são válidas se não houver autocorrelação entre os resíduos, ou seja, se eles forem independentes entre si. Caso contrário, os parâmetros estimados por mínimos quadrados serão inválidos, os coeficientes estimados não serão os melhores, e as previsões não serão ótimas, uma vez que há informações presentes nos resíduos que não estão sendo consideradas [8]. Assim, se houver autocorrelação nos resíduos, pode-se identificar o modelo ARIMA ou SARIMA que mais se ajusta ao comportamento dos resíduos, reescrevendo a equação (3) como [17]:

$$Y_t = C + bX_t + N_t \quad (5)$$

onde N_t é a série de resíduos com autocorrelação. Se, por exemplo, o comportamento de N_t seguir o modelo ARIMA(1,0,1), então podemos escrever N_t como:

$$(1 - \varphi_1 B)N_t = (1 - \theta_1 B)a_t$$

ou

$$N_t = \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)} a_t$$

que pode ser substituído em (5), formando a expressão

$$Y_t = C + bX_t + \frac{(1 - \theta_1 B)}{(1 - \varphi_1 B)} a_t$$

Para iniciar o processo é necessário partir de um modelo ARIMA inicial arbitrário, preferencialmente de baixa ordem, para os resíduos N_t para então estimar os parâmetros C e b , além dos parâmetros φ_i e θ_j . Só depois pode-se calcular os resíduos, e verificar se é ruído branco. Caso não seja ruído branco, verifica-se qual o modelo que melhor o representa, e o processo é repetido até que os resíduos a_t sejam ruído branco.

4.1. Regressão dinâmica

Tanto a regressão linear simples quanto a regressão linear múltipla não permitem a representação do impacto de valores anteriores das variáveis de entrada nos valores subsequentes da variável de saída, pois seus coeficientes são fixos. Mas é frequente encontrar situações reais em que a variável dependente Y_t recebe influência tanto dos valores das variáveis X no instante t ($X_{1,t}, X_{2,t}, \dots, X_{n,t}$), quanto das k_i leituras anteriores ($X_{1,t}, X_{1,t-1}, \dots, X_{1,t-k_1}, X_{2,t}, X_{2,t-1}, \dots, X_{2,t-k_2}, \dots, X_{n,t}, X_{n,t-1}, \dots, X_{n,t-k_n}$).

Esses tipos de situações podem ser modeladas por regressão dinâmica, seguindo a expressão (6), considerando k leituras anteriores, no caso monovariado [8,11,17]:

$$Y_t = C + v(B)X_t + N_t \quad (6)$$

onde $v(B)X_t = v_0X_t + v_1X_{t-1} + \dots + v_kX_{t-k}$ é a função de transferência linear,

C é uma constante, e

N_t é a série de distúrbios, possivelmente autocorrelacionados.

Já para o caso multivariado, com n variáveis, a expressão fica:

$$Y_t = C + v_1(B)X_{1,t} + v_2(B)X_{2,t} + \dots + v_n(B)X_{n,t} + N_t \quad (7)$$

5. PREVISÃO DE SÉRIES TEMPORAIS DE SUBPRESSÃO

Nesta sessão são apresentados passos da metodologia em desenvolvimento e os resultados preliminares obtidos com a aplicação das técnicas de SSA e regressão dinâmica com modelagem ARIMA dos resíduos para a obtenção da função de transferência linear da subpressão em fundações de barragens de concreto. Para a realização dos experimentos foram utilizados dados do piezômetro PS-F-73, localizado no contato concreto-rocha no bloco F19/20, de gravidade aliviada, da barragem principal de Itaipu, o programa estatístico R [18] com o pacote *forecast* [19] e o programa *Caterpillar 1.00* [20].

5.1. Estudo dos fatores de influência

Esta etapa incluiu visitas técnicas às fundações da barragem, entrevistas com os engenheiros, coleta de dados das leituras dos instrumentos, pesquisa em documentos técnicos e estudo de obras técnicas sobre hidrogeotecnia de maciços rochosos fraturados aplicada a fundações de barragens de concreto [1,10,21]. Com isso foram identificadas possíveis influências de:

- a) temperatura ambiente, atribuída à dilatação térmica diferente a que estão sujeitas a face de jusante da barragem (seca) e a face de montante (em contato com o lago), ocasionando diferenças de tensão imposta à fundação nas proximidades do pé de montante. No caso específico desta barragem, a influência não é imediata, devido à propagação mais lenta da temperatura na estrutura de concreto de vários metros de espessura [22];
- b) temperatura da água que passa pelos piezômetros que monitoram a pressão, nos casos em que essa informação está disponível, em função da abertura e obstrução das descontinuidades na fundação ocasionadas pelo efeito da dilatação e contração térmica provocada pela passagem de água em diferentes temperaturas. Uma variação de 20°C para 19°C na água que passa por uma descontinuidade da fundação pode provocar uma alteração de 110% na vazão que passa por ela, alterando significativamente a subpressão na região [1];
- c) nível do reservatório, que influencia diretamente a pressão hidrostática na fundação e a pressão exercida sobre a face de montante da barragem, possivelmente alterando as tensões no pé de montante [2];
- d) intervenções pontuais sobre o sistema, tais como limpeza de drenos ou sismos, no caso de alguns piezômetros. Para o caso do piezômetro estudado neste trabalho, não foi detectada intervenção significativa, em função de sua localização.

5.2. Análise gráfica das séries

Nesta fase foram traçados os gráficos das séries, identificando possíveis correlações entre as variáveis independentes e a subpressão, suas prováveis defasagens de tempo, e observando os *outliers*, que podem ser devidos a erros de leitura ou influência de eventos históricos pontuais (intervenções).

O período selecionado foi do início de 2006 ao início de 2014, em que todas as variáveis (subpressão, nível do lago, temperatura ambiental e temperatura da água que passa pelo piezômetro) estão disponíveis. As figuras 1, 2, 3 e 4 mostram os gráficos das séries originais de leituras nesse período, para a subpressão no piezômetro escolhido, temperatura ambiente, temperatura da água e nível do reservatório.

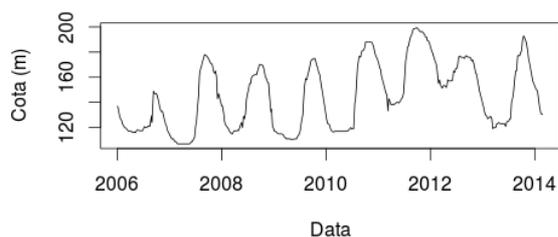


Figura 1: Leituras do piezômetro PS-F-73, em metros sobre o nível do mar, de jan/2006 a fev/2014.

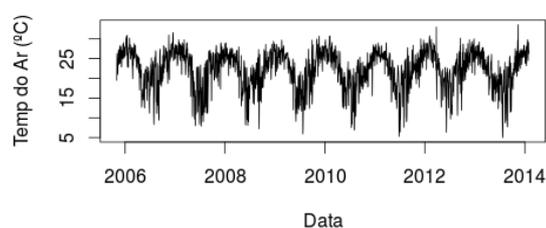


Figura 2: Temperatura ambiente média diária, em °C, de nov/2005 a jan/2014.

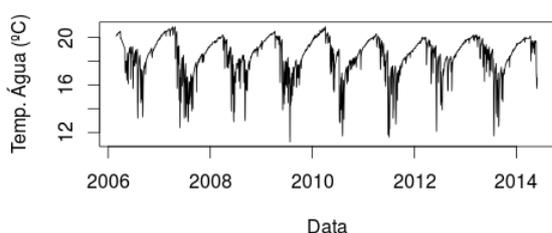


Figura 3: Temperatura média diária da água no PS-F-73, em °C, de mar/2006 a mai/2014.

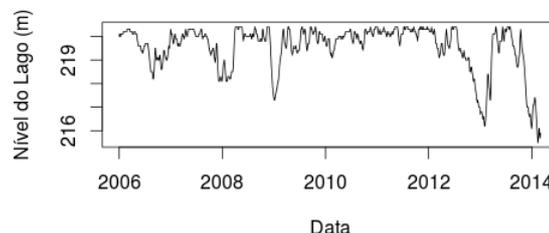


Figura 4: Nível do lago, média diária em metros sobre o nível do mar, de jan/2006 a fev/2014.

Nesta fase foram identificadas:

- uma correlação forte entre a temperatura ambiente e a subpressão, com uma defasagem de aproximadamente 3 meses, com a temperatura máxima no verão levando a uma subpressão mínima no outono, e a temperatura mínima no inverno induzindo a uma subpressão máxima na primavera;
- possível influência da temperatura da água que passa pelo piezômetro na subpressão quase que imediatamente à variação de temperatura;
- resfriamento rápido e aquecimento lento da temperatura da água que passa pelo piezômetro, devido à convecção da água e à condução do calor, respectivamente;
- nenhuma influência visível do nível do reservatório no nível de subpressão. Essa variável permanece na análise para testes estatísticos de significância.

5.3. Regularização das séries temporais

Desde sua construção, a barragem de Itaipu passou por diversas fases até a acomodação da estrutura, e por esse motivo as leituras de subpressão foram realizadas a intervalos diferentes em cada fase. Conforme as estruturas se acomodaram, as leituras tornaram-se mais estáveis e a frequência de leitura foi diminuída. Também há diferenças no comportamento de piezômetros instalados em diferentes feições, sendo lidos mais frequentemente os que

apresentam maior variação. Para o piezômetro PS-F-73 há períodos com leitura diária, duas vezes por semana e semanal. Desta forma, usou-se a frequência semanal, pois está disponível em todo o período. Também foram eliminadas leituras extras realizadas por ocasião de procedimentos de manutenção, leituras duplicadas, e as poucas leituras não disponíveis foram estimadas pela média entre os vizinhos. Essa estimativa de valores mostrou-se consistente com as leituras automáticas, que também se mantiveram próximas à média das leituras das semanas anterior e seguinte. Em função do cálculo de sazonalidade, também foi definido um número fixo de 52 leituras por ano, pois o número apresentava pequenas variações dependendo do dia da semana em que o ano se iniciou.

5.4. Escolha da periodicidade

Para cada uma das variáveis independentes (temperatura ambiente, temperatura da água na fundação e nível do reservatório) foram estabelecidas a média diária, a média dos últimos 7 dias, a média dos últimos 30 dias e a média dos últimos 90 dias, e a periodicidade que apresentou melhor explicação foi mensal. Foram acrescentadas também as leituras das médias mensais de temperatura ambiente com defasagem de 2, 3 e 4 meses, uma vez que há o indício visual a partir do gráfico das séries de que a subpressão é afetada pela temperatura ambiente ocorrida 3 meses antes. As séries geradas ficaram com 96 valores de médias mensais.

5.5. Filtragem de ruídos com SSA

Partindo da série de 96 médias mensais de subpressão, separou-se os últimos 6 valores para a realização de testes, deixando os primeiros 90 valores para a amostra sobre a qual foi feito o ajuste dos modelos. Em seguida procedeu-se à filtragem usando SSA sobre a amostra, com o objetivo de reduzir os ruídos da série. A escolha do tamanho da janela L foi balizada pela fórmula $L = (\ln 90)^c$, com $1,5 \leq c \leq 3$, e $L \leq N/2$ [14]. Sendo a série curta, optou-se por $L=10$, próximo ao limite inferior, e pelas 4 primeiras componentes, cujos autovalores somam 99,93% do total, desprezando as outras 6 componentes. Aplicando os testes Box-Pierce e Ljung-Box sobre os resíduos, ambos aceitam a hipótese dos resíduos serem ruído branco, apresentando resultados de valor-p iguais a 0,5068 e 0,4997 respectivamente. A partir de então, a série reconstruída sem as componentes de ruído passa a ser usada para a modelagem da função de transferência. A seleção de quais componentes usar para reconstruir a série foi feita com o auxílio da ferramenta de plotagem 2D do programa Caterpillar, considerando os padrões regulares como válidos, e os irregulares como ruído. A figura 5 mostra a série original e a série filtrada.



Figura 5: Série PS-F-73 original (vermelha) e filtrada com SSA, L=10, K=4 (preta).

5.6. Regressão múltipla e testes de significância

Foram realizados testes de regressão múltipla para identificar a subpressão em função das três variáveis independentes, com cada uma das periodicidades, chegando-se à conclusão de que para o piezômetro em questão, a periodicidade mensal apresentou maior valor para o teste R^2 ajustado. As tabelas 1 e 2 apresentam o resultado de alguns testes, com e sem defasagem no tempo para a variável temperatura ambiente. Da série total de 96 meses, os primeiros 90 meses foram considerados a amostra para o ajuste do modelo e os últimos 6 para testes:

<i>Variáveis Independentes</i>	<i>Valor Estimado</i>	<i>Erro Padrão</i>	<i>Signif. a 95%</i>
Constante	315,3	26,30	Sim
Temperatura do Ar (-3 meses)	-4,629	0,5785	Sim
Nível do Reservatório	$-7,114 \times 10^{-4}$	0,009363	Não
Temperatura da Água	-3,671	1,525	Sim

<i>Variável Dependente</i>	<i>Subpressão no PS-F-73</i>
Número de leituras na amostra	90
R ² ajustado	0,5359
RMSE	17,62m

Tabela 1: Regressão múltipla para o PS-F-73 com valores médios mensais. Temperatura ambiente com defasagem de 3 meses, Nível do lago e Temperatura da água no PS-F-73 atuais.

A tabela 1 apresenta a primeira tentativa de regressão múltipla com médias mensais, baseada nas três variáveis independentes identificadas, sendo a temperatura ambiente deslocada em 3 meses, de acordo com o comportamento gráfico identificado. Observa-se que o valor do RMSE (raiz do erro quadrático médio) é de 17,62m, e o valor do R^2 ajustado indica que essas variáveis são capazes de explicar 53,59% da variação de subpressão no PS-F-73 dentro da amostra, o que não é considerado satisfatório. Usando o modelo ajustado para prever os próximos 6 meses, o RMSE foi de 9,34m, e o R^2 foi de 0,7821153. O valor de R^2 ajustado ficou muito baixo devido ao pequeno tamanho de amostra (6), e ao número significativo de parâmetros (4). O resultado revela ainda que as variáveis temperatura ambiente e temperatura

da água são significativas ao nível de significância de 95%.

Como o nível do lago não foi significativo foi executada também a regressão múltipla sem essa variável, o que produziu uma pequena melhora no R^2 ajustado dentro da amostra, que passou para 0,5412, e mantendo o RMSE dentro da amostra estável em 17,62m. Na previsão fora da amostra, o RMSE aumentou um pouco, passando para 9,49m, e o R^2 ajustado ficou em 0,4475394, também influenciado pela pequena amostra.

O próximo teste foi incluir as defasagens de 2 e 4 meses para a temperatura ambiente, o que não melhorou significativamente o R^2 ajustado e o RMSE, deixando significativos somente os parâmetros da constante e da temperatura do ar defasada de 4 meses.

Fazendo mais um teste, desta vez retirando a defasagem de 3 meses e o nível do lago, as variáveis melhoraram seu nível de significância, o RMSE permanece praticamente inalterado, em 17,09m e o R^2 ajustado alcança 0,5638, o que ainda não é satisfatório. O resultado é ilustrado na tabela 2.

<i>Variáveis Independentes</i>	<i>Valor Estimado</i>	<i>Erro Padrão</i>	<i>Signif. a 95%</i>
Constante	346,59	33,8905	Sim
Temperatura do Ar (-2 meses)	-2,3343	0,9393	Sim
Temperatura do Ar (-4 meses)	-3,2229	0,6349	Sim
Temperatura da Água	-4,2572	2,2226	Não

<i>Variável Dependente</i>	<i>Subpressão no PS-F-73</i>
Número de leituras na amostra	90
R2 ajustado	0,5638
RMSE	17,09m

Tabela 2: Regressão múltipla para o PS-F-73 com valores médios mensais. Temperatura ambiente com defasagem de 2 e 4 meses e Temperatura da água atual no PS-F-73.

Novamente o ajuste nos parâmetros trouxe um pequeno ganho, mas não o suficiente para considerar aceitável a regressão, pois explica somente 56,38% da variação do PS-F-73.



Figura 6: Gráfico das séries da estimativa pelo modelo de regressão múltipla (vermelho) e da leitura média mensal do PS-F-73 (preto), em metros, dentro da amostra com 90 leituras.

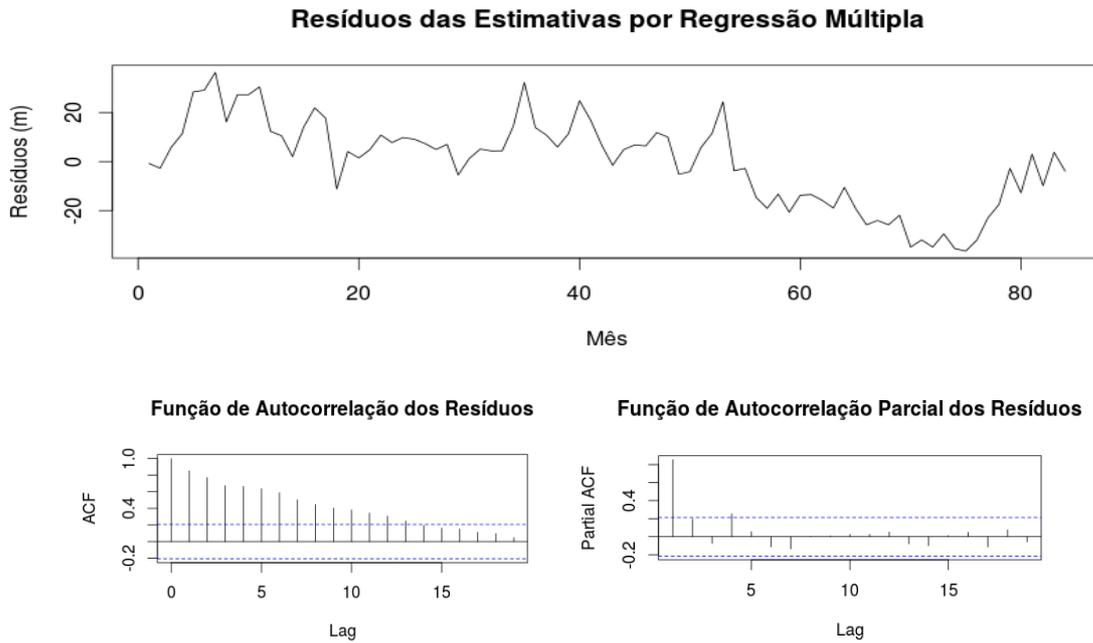


Figura 7: Gráfico, Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) dos resíduos da regressão múltipla dentro da amostra com 90 leituras.

Pode-se observar na figura 6 os gráficos da série de leituras de subpressão (em preto), da estimativa pelo modelo estimado na tabela 2 (em vermelho), dentro da amostra. Pode-se ver que a estimativa tem alguma aproximação à série de leituras, mas não próxima o suficiente. A série de erros, ilustrada na figura 7 mostra erros de -32m a +36m. A série de resíduos também não aparenta ser ruído branco, o que é confirmado pelos gráficos das funções de autocorrelação residual e autocorrelação parcial residual.

Usando este modelo para prever a subpressão nos próximos 6 meses, foi obtido um RMSE de 8,10m, com erros de -0,63m a 16,34m. O gráfico com as leituras e estimativas é mostrado na figura 8.

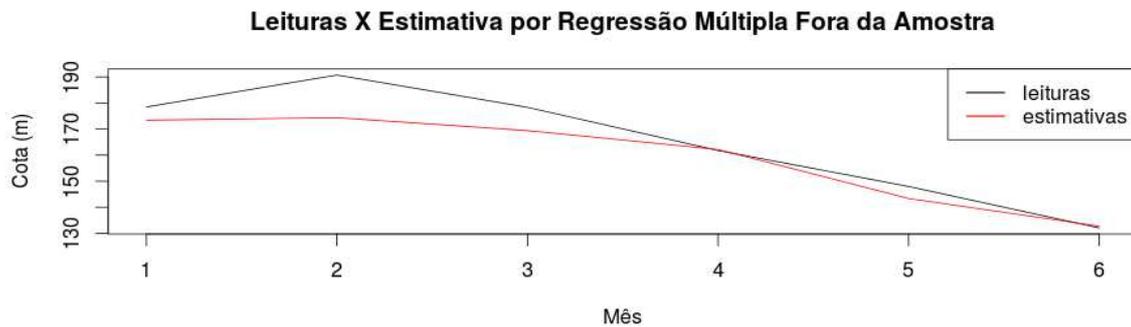


Figura 8: Comparação entre previsão pelo modelo de regressão múltipla (vermelho) e média mensal das leituras do PS-F-73 (preto), em m, fora da amostra com 6 meses.

Na figura 7 nota-se que a FAC e a FACP apontam que os resíduos não são ruído branco, e apresentam um comportamento típico de uma série AR(1). Isso mostra que há mais informações que o modelo não captou, e que os coeficientes estimados estão incorretos. Assim há indicações para refazer a regressão dinâmica incluindo a modelagem ARIMA da série de resíduos.

5.7. Regressão múltipla com modelagem ARIMA dos resíduos

De acordo com a técnica explicada na sessão 4, foi executada a rotina de seleção automática do modelo ARIMA da série temporal das leituras médias mensais de subpressão no piezômetro PS-F-73, considerando a regressão da subpressão a partir das variáveis independentes temperatura ambiental média mensal, com defasagens de 2, 3 e 4 meses, nível médio do lago dos últimos 30 dias e temperatura média dos últimos 30 dias da água que passa pelo piezômetro. Para estimar os parâmetros do modelo foi selecionada uma amostra com as primeiras 90 leituras, deixando as últimas 6 para validação.

<i>Variáveis Independentes e Coeficientes ARIMA</i>	<i>Valor Estimado</i>	<i>Erro Padrão</i>	<i>Signif. a 95%</i>
AR1 (φ_1)	0,2850	0,0986	Sim
AR2 (φ_2)	-0,4647	0,1038	Sim
Temperatura do Ar (-2 meses)	-1,5897	0,3185	Sim
Temperatura do Ar (-3 meses)	-2,5259	0,2952	Sim
Temperatura do Ar (-4 meses)	-1,1322	0,2922	Sim
Temperatura da Água	-1,6296	0,7863	Sim
Nível do Reservatório	0,0021	0,0020	Não

<i>Variável Dependente</i>	<i>Subpressão no PS-F-73</i>
Número de leituras na amostra	90
R2 ajustado	0,9368
RMSE	6,35m
Modelo ARIMA selecionado	ARIMA(2,1,0)

Tabela 3: Modelagem ARIMA dos resíduos N_t , da regressão dinâmica do PS-F-73, dentro da amostra, com 90 leituras.



Figura 9: Comparação entre estimativa pelo modelo escolhido (vermelho) e leitura média mensal (preto) do PS-F-73, em metros, com 90 leituras.

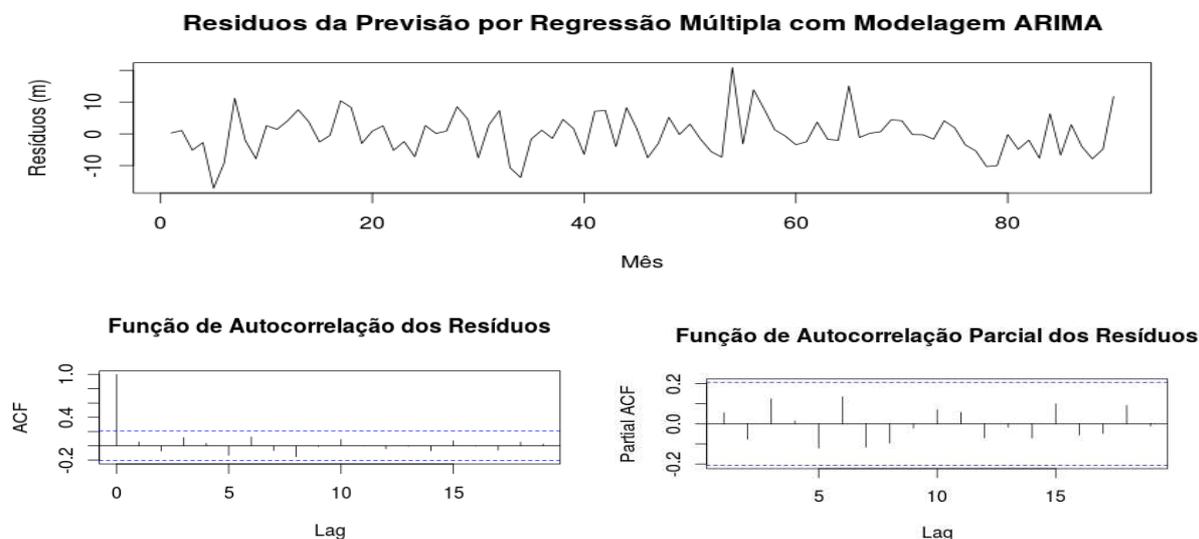


Figura 10: Função de Autocorrelação (FAC) e Função de Autocorrelação Parcial (FACP) dos resíduos a_t da regressão múltipla com modelagem ARIMA dos resíduos N_t .

Pode-se observar na tabela 3 uma melhora significativa utilizando a modelagem ARIMA dos resíduos N_t . O valor da raiz do erro quadrático médio (RMSE) caiu de 17,09m para 6.35m. O valor do R^2 ajustado subiu de 0.5638 para 0.9368, ou seja, sua capacidade explicativa chegou a 93,68%. O modelo selecionado para os resíduos foi ARIMA(2,1,0), e entre as variáveis da regressão, apenas o nível do Lago não é significativo ao nível de confiança de 95%. Retirando o nível do lago e executando novamente a regressão o R^2 ajustado não se altera e o RMSE apresenta um pequeno aumento (6,388212). Já a figura 9 mostra nos gráficos das séries de estimativa pelo modelo da tabela 3 e das leituras uma melhora também significativa em comparação com a figura 6, e a figura 10 exibe a o gráfico da série de resíduos, e as funções de autocorrelação residual e autocorrelação residual parcial, mostrando que os resíduos a_t já podem ser considerados ruído branco, o que é confirmado pelo teste de Box-Pierce.

5.8. SSA e Regressão múltipla com modelagem ARIMA dos resíduos

Com o objetivo de verificar se a aplicação da regressão múltipla com modelagem ARIMA dos resíduos produz um resultado melhor se aplicada à série de subpressão filtrada com SSA, menos ruidosa, em comparação com a série original, foi executado mais um teste, ilustrado na tabela 4. O teste foi aplicado considerando na regressão a temperatura da água e a temperatura ambiente com defasagens de 2, 3 e 4 meses.

Nota-se na tabela 4 que os parâmetros são todos significativos, e o modelo apresenta melhoria significativa na previsão, se comparado ao resultado apresentado na tabela 3, com o RMSE caindo de 6,35m para 2,83m, e o R^2 ajustado subindo de 0,9368 para 0,9871. Os parâmetros para a filtragem SSA que produziram melhores resultados foram: tamanho da janela igual a 10 e número de componentes consideradas igual a 4.

<i>Variáveis Independentes e Coeficientes ARIMA</i>	<i>Valor Estimado</i>	<i>Erro Padrão</i>	<i>Signif. a 95%</i>
AR1 (φ_1)	2,5726	0,0596	Sim
AR2 (φ_2)	-3,2946	0,1336	Sim
AR3 (φ_3)	2,2896	0,1318	Sim
AR4 (φ_4)	-0,8185	0,0572	Sim
Temperatura do Ar (-2 meses)	0,0814	0,0377	Sim
Temperatura do Ar (-3 meses)	0,1064	0,0577	Sim
Temperatura do Ar (-4 meses)	0,0873	0,0340	Sim
Temperatura da Água	-0,1348	0,0403	Sim

<i>Variável Dependente</i>	<i>Subpressão no PS-F-73</i>
Número de leituras na amostra	90
R ² ajustado	0,9871
RMSE	2,83m
Modelo ARIMA selecionado	ARIMA(4,1,0)
Tamanho da janela SSA (L)	10
Número de componentes consideradas na reconstrução	4

Tabela 4: Modelagem ARIMA dos resíduos N_t , da regressão dinâmica sobre a série filtrada por SSA das leituras do PS-F-73, dentro da amostra de 90 leituras.

6. CONCLUSÕES

Este artigo buscou explicitar a metodologia em desenvolvimento para a previsão de séries temporais de subpressão em fundações de barragens de concreto com base em variáveis ambientais, de modo a manter algum grau de explicação sobre a variação da subpressão. Uma composição de métodos foi testada, e foram obtidas melhorias substanciais nos resultados.

Os testes realizados mostraram uma melhoria na obtenção do RMSE de algo próximo a 16m com regressão múltipla para algo próximo a 6m, e uma melhoria no R^2 ajustado da ordem de 55% para 93%, obtendo um modelo de função de transferência linear com uma regressão múltipla e modelagem ARIMA da série temporal dos resíduos.

Outro resultado interessante alcançado foi a melhoria do desempenho da função de transferência linear se a série original for previamente filtrada com a técnica de Análise Espectral Singular (SSA – *Singular Spectral Analysis*). Neste caso, o R^2 ajustado subiu para 98% e o RMSE caiu para abaixo de 3m.

Entre os trabalhos futuros cabe destacar um estudo sobre a otimização dos parâmetros da filtragem SSA.

Agradecimentos

Os autores agradecem aos engenheiros de Itaipu Binacional Dimilson Pinto Coelho e Silvia Frazão Matos, pelas contribuições técnicas ao comportamento de barragens, à Fundação Araucária pelo suporte financeiro parcial, projeto número 44.721, ao IFPR pela possibilidade de dedicação exclusiva à pesquisa, e à UFPR, Unioeste, Ceasb e Fundação PTI por promover a turma de doutorado em métodos numéricos em engenharia, que motivou esta pesquisa.

REFERÊNCIAS

- [1] R. M. Andrade, *Mecânica do escoamento em maciços rochosos fraturados aplicada a barragens*, Engevix, Rio de Janeiro, (1988).
- [2] A. Mascarenhas, *Qualidade hidráulica e deteriorações em fundações de barragens de betão*, Laboratório Nacional de Engenharia Civil, Lisboa, (2005).
- [3] ICOLD, International Commission on Large Dams, *Dam safety Guidelines*, Bulletin 59, Paris, (1987).
- [4] CDBD, Comitê Brasileiro de Barragens, *Guia básico de segurança de barragens*, XXIII Seminário Nacional de Grandes Barragens, Belo Horizonte, (1999).
- [5] C. I. Osako, *A manutenção dos drenos das fundações de barragens: O caso da usina hidrelétrica de Itaipu*, Dissertação de mestrado em construção civil, UFPR, Curitiba, (2002).
- [6] S. C. Kuperman, et al, *Critérios para a fixação de valores limite na instrumentação civil de barragens de concreto e terra*, XXV Seminário Nacional de Grandes Barragens, Salvador, (2003).
- [7] N. Golyandina, V. Nekrutkin, A. A. Zhigljavsky, *Analysis of time series structure: SSA and Related Techniques*, Chapman & Hall, (2001).
- [8] R. Yaffee, M. McGee, *Introduction to time series analysis and forecasting*, Academic Press, San Diego, (2000).
- [9] J. C. Royer, V. E. Wilhelm, J. Patias, *Dam uplift pressure forecast using dynamic regression*, Applied Mathematical Sciences, Vol 9, (2015).
- [10] G. Box, G. Jenkins, G. Reinsel, *Time series analysis: forecasting and control*, John Wiley & Sons, Hoboken, New Jersey, (2008).
- [11] R. Hyndmann, G. Athanasopoulos, *Forecasting: principles and practice*, (2014). Disponível em <<http://otexts.org/fpp>>.
- [12] N. Golyandina e A. Zhigljavsky, *Singular Spectrum Analysis for Time Series*. Springer, (2013).
- [13] H. Hassani, *Singular Spectrum Analysis: Methodology and Comparison*. Journal of Data Science, Vol. 5, pp. 239-257, (2007).
- [14] M. A. R. Khan e D. S. Poskritt, *A note on window length selection in singular spectrum analysis*, Australian & New Zealand Journal of Statistic, 55(2), (2013).
- [15] J. F. M. Pessanha et al, *Usando a análise espectral singular na previsão da produção mensal de um parque eólico*, XVII Simpósio de Pesquisa Operacional e Logística da Marinha, Rio de Janeiro, (2014).
- [16] K. M. Cassiano, M. L. Menezes e J. F. M. Pessanha, *Algumas abordagens para identificação de ruídos na análise espectral de séries temporais*, XLVI Simpósio Brasileiro de Pesquisa Operacional, Salvador, (2014).]
- [17] A. PANKRATZ, *Forecasting with Dynamic Regression Models*, John Wiley & Sons, New York, (1991).
- [18] R Core Team, *R: A language and environment for statistical computing*, R Foundation for Statistical Computing, Vienna, (2014).
- [19] R. Hyndmann, et al, *Forecasting functions for time series and linear models*, R Package

- Version 5.6, (2014). Disponível em <<http://CRAN.R-project.org/package=forecast>>.
- [20] GistaT Group, *Caterpillar 1.00*, programa de modelagem SSA. Disponível em <<http://www.gistatgroup.com/cat/programs.html>>
- [21] Itaipu Binacional, *Usina Hidrelétrica de Itaipu: aspectos técnicos das estruturas civis*, Foz do Iguaçu, 2008.
- [22] A. S. Fiorini et al, *Barragem de Itaipu: comparação dos diferentes tipos de barragens de concreto*, XXIV Seminário Nacional de Grandes Barragens, Fortaleza, (2001).