

<https://eventos.utfpr.edu.br//sicite/sicite2019>

Modelo de regressão para dados de instrumentação de barragem

Regression model for dam instrumentation data

RESUMO

Giovanna de Oliveira Moreira
giovannadomoreira@gmail.com
Universidade Tecnológica Federal
do Paraná, Toledo, Paraná, Brasil

Suellen Ribeiro Pardo Garcia
suellenpardo@utfpr.edu.br
Universidade Tecnológica Federal
do Paraná, Toledo, Paraná, Brasil

Os riscos atrelados ao rompimento de uma barragem são de grandes proporções como degradação ambiental, perdas de vida humana e animal, além de prejuízos materiais. O monitoramento dessas grandes estruturas consiste na inspeção visual dos engenheiros e técnicos realizada periodicamente e também do acompanhamento das leituras realizadas por instrumentos instalados na estrutura. Os instrumentos instalados geram uma massa de dados que permite avaliar a estabilidade da estrutura, verificação e confirmação de hipóteses que asseguram ou não a segurança na obra. É de interesse dos engenheiros e técnicos conhecer a previsão das leituras desses instrumentos, para em tempo hábil fazer tomadas de decisões, evitando acidentes. Os modelos estatísticos são geralmente utilizados para fazer essas previsões, particularmente, os modelos de regressão. Apresenta-se neste artigo um modelo de regressão linear múltipla para os movimentos relativos de um bloco de concreto. Utiliza-se a técnica Análise de Componente Principais (ACP) reduzindo 14 sensores em 3 novas variáveis, que são as componentes principais. O modelo é construído para a primeira componente principal que explica 47% da variabilidade dos dados originais e considera-se como variáveis que influenciam essa componente, o modelo HST (Hydrostatic Season Time), amplamente utilizado no contexto de segurança de barragens. Após vários ajustes, o modelo final apresentou um coeficiente de determinação de 0,89, confirmando o quanto o modelo consegue explicar os valores observados.

PALAVRAS-CHAVE: Deslocamentos, Pêndulos, HST, ACP.

Recebido: 19 ago. 2019.

Aprovado: 01 out. 2019.

Direito autoral: Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



ABSTRACT

The risks associated with the rupture of a dam are of great proportions such as environmental degradation, loss of human and animal life, and material damage. Monitoring of these large structures consists of periodically inspecting the engineers and technicians and also monitoring the readings performed by instruments installed on the structure. The installed instruments generate a mass of data that allows to evaluate the stability of the structure, verification and confirmation of hypotheses that assure or not the safety in the work. It is in the interest of engineers and technicians to know the forecast of the readings of these instruments, in order to make timely decisions, avoiding accidents. Statistical models are generally used to make these predictions, particularly regression models. This paper presents a multiple linear regression model for the relative movements of a concrete block. The Principal Component Analysis (PCA) technique is used by reducing 14 sensors into 3 new variables, which are the main components. The model is built for the first major component that explains 47% of the variability of the original data and is considered as variables that influence this component, the Hydrostatic Season Time (HST) model, widely used in the dam safety context. After several adjustments, the final model

presented a coefficient of determination of 0.89, confirming how much the model can explain the observed values.

KEYWORDS: Displacements, Pendulums, HST, PCA.

INTRODUÇÃO

O monitoramento de barragens é a atividade que pode evitar ou ao menos diminuir os danos de uma possível ruptura na estrutura. Caracteriza-se por inspeções visuais e regulares na estrutura e a coleta e análise dos dados da instrumentação instalada nos blocos com o objetivo de identificar eventuais alterações no comportamento observado, gerando alertas para as equipes responsáveis pela gestão da segurança (PENNA, 2015; ARAGÃO, 2015; FUSARO, 2015).

Com relação a análise dos dados da instrumentação, os modelos estatísticos utilizados, mais especificadamente, os modelos de regressão visam prever as leituras dos instrumentos e assim ser possível construir intervalos de confiança para essas leituras. Esses intervalos de confiança são utilizados como referência para as leituras mais recentes observadas.

Entre os modelos de regressão encontrados na literatura está o modelo de regressão linear múltipla HST (Hydrostatic Season Time) proposto inicialmente em 1958 por Ferry, Will e Beaujoint (GARCIA; et al., 2015). Esse modelo é geralmente utilizado para monitorar deslocamentos, considerando funções que descrevem o efeito da pressão hidrostática na barragem, o efeito das variações de temperatura ambiente e os eventuais efeitos irreversíveis devido a deformações permanentes, pois as barragens estão sob ação destes três efeitos.

Além dos modelos estatísticos, técnicas da análise multivariada dos dados, também são aplicadas aos dados de monitoramento de barragem em busca de redução de variáveis, agrupamento de instrumentos, interpretação entre instrumentos relacionados, entre outros (AHMADI-NEDUSHAN). A técnica de análise de componentes principais consiste em extrair poucas variáveis de um grupo grande de variáveis originais, reduzindo o número de análises individuais, por exemplo, um conjunto de sensores de determinado instrumento pode ser reduzido a poucas componentes, desde que essas componentes representem os dados de forma satisfatória.

Neste contexto o presente artigo utilizará da teoria da técnica de análise de componentes principais em sensores dos pêndulos direto e invertido extraindo sua primeira componente principal como nova variável independente e da modelagem do HST, extrair as variáveis dependentes do modelo, que influenciam no comportamento dessa componente.

MATERIAIS E MÉTODOS

O conjunto de dados é composto pelos valores mensurados pelos sensores dos pêndulos direto e invertido instalados ao longo do bloco f19/20 da barragem da Usina Hidrelétrica de Itaipu e esses valores foram aferidos manualmente pelos técnicos e engenheiros.

As leituras consideradas foram entre janeiro de 2000 até junho de 2015 com periodicidade mensal, ou seja, dispõe-se de uma amostra de 186 observações ($n = 186$). Existem 7 sensores dos pêndulos, 6 sensores do pêndulo direto (cof17, cof18, cof19, cof20, cof21 e cof22) dispostos ao longo da estrutura do bloco e, 1 sensor do pêndulo invertido (cof23) localizado na base do bloco. Assim, tem-se 14 variáveis, pois os deslocamentos são na direção X, ou seja, no sentido do fluxo (direção montante-jusante) e, na direção Y, perpendicular ao fluxo (margem direita-esquerda). Portanto, a amostra é uma matriz de dimensão 186×14 .

A amostra foi dividida em conjunto de treinamento e conjunto de teste. O conjunto de treinamento são os dados de janeiro de 2000 a dezembro de 2012 e foram utilizados na construção do modelo apresentado neste artigo. O conjunto de teste, ainda não utilizado, pois será feita a previsão pelo modelo, fora da amostra, e depois o erro será calculado.

A metodologia do trabalho consistiu em estudar a teoria do método de análise de componentes principais, pois as componentes são as variáveis dependentes do modelo de regressão linear que será apresentado.

A análise de componentes principais é a técnica que utiliza princípios da álgebra linear para transformar variáveis correlacionadas, em um número menor de variáveis chamadas de componentes principais. As novas variáveis são combinações lineares das variáveis originais e poucas delas explicam a maior parte da variabilidade dos dados (MINGOTI).

Na sequência, um estudo sobre a teoria dos modelos de regressão linear múltipla foi realizado, para auxiliar no estudo do modelo HST. O HST é um modelo de regressão linear múltipla que descreve a relação de efeitos reversíveis (efeitos da pressão hidrostática e da temperatura) e efeitos irreversíveis (efeitos do tempo) no deslocamento da estrutura. Esse relacionamento das ações das variáveis ambientais no bloco de uma barragem e sua resposta pode ser modelado matematicamente e correspondem a um período em que a configuração da estrutura permanece a mesma (AHMADI-NEDUSHAN).

No HST, a função do nível do reservatório (carga hidrostática) é dada por um polinômio de quarto grau, a função da temperatura dada por uma soma de funções trigonométricas e, a função dos efeitos irreversíveis por uma função polinomial de terceiro grau (AHMADI-NEDUSHAN, 2002). Uma alternativa comum na literatura é substituir a função periódica da temperatura pela temperatura no concreto da barragem (LÉGER, 2007; LECLERC, 2007), e ainda existem variações da função do efeito irreversível na literatura, considera-se aqui a proposta de Xi et (2011) e o modelo é dado pela seguinte forma:

$$Y = a_0 + \sum_{i=1}^4 a_i z^i + \sum_{i=1}^k b_i T_i + c_0 \theta + c_1 \ln \theta \quad (1)$$

Em que a_0 é o termo independente, $z = \frac{H - H_{min}}{H_{max} - H_{min}}$, h é a altura do nível do reservatório, T_i são as leituras de temperatura dos termômetros instalados no concreto, $\theta = \frac{t - t_0}{100}$ em que t_0 é o tempo que se iniciou a leitura do sensor.

E por fim, após escolhido o modelo, com as variáveis importantes para descrever o comportamento da componente principal, apresenta-se os valores previstos pelo modelo.

RESULTADOS E DISCUSSÕES

A técnica de análise de componentes principais foi aplicada aos 14 sensores dos pêndulos direto e invertido e a escolha do número de componentes se deu pelo critério de variabilidade explicada pelas mesmas. A primeira componente principal explicou 47% da variabilidade dos dados, a segunda componente explicou 28% e a terceira 10%, ou seja, as três primeiras componentes explicaram 85% da variabilidade dos dados, o que foi considerado satisfatório. Assim, três novas variáveis representariam um conjunto de 14 sensores. O modelo será construído apenas para a primeira componente principal (CP1).

A Tabela 1 apresenta os resultados da primeira simulação, considerando o modelo dado em (1). Nota-se que a coluna 5 apresenta valor p, em geral, maior do que 0,05 (nível de significância), ou seja, nestes testes, aceita-se a hipótese nula de que o coeficiente referente a determinada variável no modelo é nulo. Apenas as variáveis T_1 , T_5 e T_6 apresentaram coeficientes significativos.

Tabela 1 – Modelo 1

Variável	Coefficiente	Erro Padrão	Estatística t	Valor p
C	-6,931	6,523	-1,062	0,289
Z	-34,355	31,595	-1,087	0,278
Z²	123,302	100,841	1,222	0,223
Z³	-142,912	120,714	-1,183	0,238
Z⁴	55,504	49,272	1,126	0,261
T₁	0,997	0,231	4,311	0,000
T₂	-0,445	0,313	-1,423	0,156
T₃	0,083	0,298	0,280	0,779
T₄	0,013	0,055	0,234	0,815
T₅	-0,345	0,045	-7,595	0,000
T₆	-0,106	0,024	-4,383	0,000
θ	2,090	1,842	1,135	0,258
ln θ	-0,165	3,564	-0,046	0,963

Fonte: Autoria própria (2019).

O coeficiente de determinação ajustado foi de 0,89.

A partir disso, foram retiradas, uma a uma, as variáveis que apresentaram coeficiente estatisticamente nulos. No modelo 1, foi retirada a “ln θ ” que apresentou o maior valor p (0,96). As tabelas dos novos modelos gerados não serão apresentadas, apenas a discussão dos resultados.

Após retirada, no modelo 2, as variáveis T_1 , T_5 , T_6 e T têm coeficientes significativos e retirou-se a variável T_4 (modelo 3) que apresentou maior valor p (0,82). No modelo 3, as mesmas variáveis apresentaram coeficientes significativos, então retirou-se a variável T_3 que apresentou maior valor p (0,75). No modelo 4, obteve-se as mesmas variáveis com coeficientes significativos, então Z foi retirado, pois apresentou maior valor p (0,26). No modelo 5, a variável constante (termo independente) passou a ser importante e Z^4 foi retirado, pois apresentou maior valor p (0,79). No modelo 6, retirou-se a variável Z^3 com maior valor p (0,16). E por fim, no modelo 7, retirou-se a constante com valor de 0,11.

Apresenta-se na Tabela 2, os resultados do modelo final (modelo 8), mais parcimonioso, com todos os coeficientes significativos.

Tabela 1 – Modelo Final

Variable	Coefficiente	Erro Padrão	Estatística t	Valor p
Z^2	2,237	0,411	5,433	0,000
T_1	0,665	0,086	7,684	0,000
T_2	-0,444	0,117	-3,779	0,000
T_5	-0,323	0,044	-7,273	0,000
T_6	-0,116	0,021	-5,492	0,000
θ	2,119	0,193	10,965	0,000

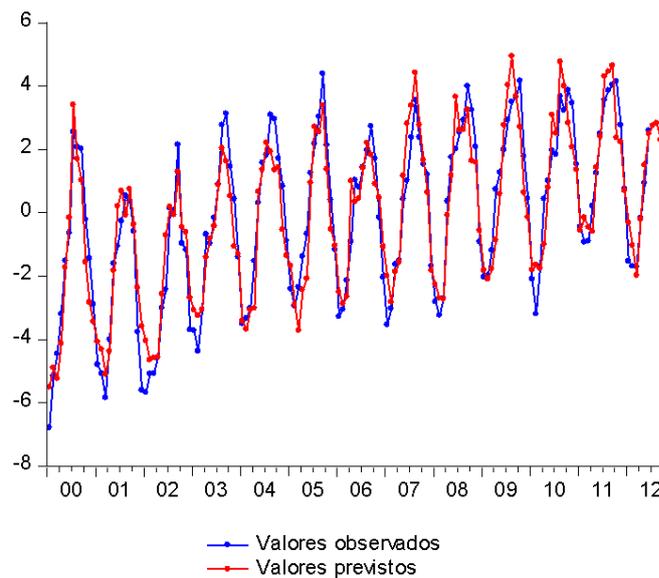
Fonte: Autoria própria (2019).

O valor do coeficiente de determinação ajustado foi de 0,89 e equação do modelo de regressão linear múltipla é dada por:

$$CP1 = 2,24 Z^2 + 0,67 T_1 + 0,44 T_2 - 0,32 T_5 - 0,12 T_6 + 2,12 \theta \quad (2)$$

A Figura 1 apresenta os valores reais da primeira componente principal dos sensores e os valores ajustados pelo modelo em (2).

Figura 1 – Comparação entre valores previstos e observados



Fonte: Autoria própria (2019).

CONCLUSÃO

O artigo apresenta a aplicabilidade da técnica de análise de componentes principais em dados de monitoramento de barragem visto que 14 variáveis podem ser reduzidas a novas 3 variáveis e ainda assim explicar 85% da variabilidade dos dados. O modelo construído para a primeira componente principal, que explica 47% da variabilidade dos dados originais dos sensores, apresentou um coeficiente de determinação de 0,89, confirmando o quanto o modelo consegue explicar os valores observados. Assim, pode-se prosseguir para outra etapa da modelagem que seria a avaliação do modelo para fins de previsão dos movimentos relativos do bloco da barragem de concreto.

REFERÊNCIAS

- AHMADI-NEDUSHAN, B. **Multivariate Statistical Analysis of monitoring data for concrete dams**. Tese de Doutorado do Departamento de Engenharia Civil e Mecânica Aplicada, McGill University. Montreal, p. 211. 2002.
- GARCIA, Suellen et al. **Modelo estatístico para o monitoramento de movimentos relativos na barragem da uhe itaipu**. Foz do Iguaçu, 2015. Disponível em: http://www.cbdb.org.br/xxsngb/download/trabalhos_tecnicos/tema112/MODELO%20ESTAT%3%8DSTICO%20PARA%20O%20MONITORAMENTO%20DE%20MOVIMENTOS%20RELATIVOS%20NA%20BARRAGEM%20DA%20UHE%20ITAIPU.pdf. Acesso em: 3 ago. 2019.
- LÉGER, P.; LECLERC, M. **Hydrostatic, temperature, time-displacement model for concrete dams**. Journal of engineering mechanics, v. 133, n. 3, p. 267-277, Março 2007. Disponível em: <https://ascelibrary.org/doi/full/10.1061/%28ASCE%290733-9399%282007%29133%3A3%28267%29>. Acesso em: 8 ago. 2019.
- LOMBARDI, G.; AMBERG, F.; DARBRE, G. R. **Algorithm for the prediction of functional delays in the behaviour of concrete dams**. The international journal on hydropower & dams, v. 15, n. 3, p. 111-116, 2008. Disponível em: <https://www.lombardi.ch/fr-fr/SiteAssets/Publications/1280/Pubb-0460-L-Algorithm%20for%20the%20prediction%20of%20functional%20delays%20in%20the%20behaviour%20of%20concrete%20dams.pdf>. Acesso em: 3 ago. 2019.
- MACHADO, LUCAS V. T., AZEEZ, DOLPAO G. **Incidentes e acidentes em barragens**. Rio de Janeiro 2018. Disponível em: https://www.itr.ufrj.br/sigabi/wp-content/uploads/7_sigabi/VASCONCELLOS_LUCAS_355a358.pdf. Acesso em: 3 ago. 2019.
- MINGOTI, S. A.; **Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada**. Belo Horizonte: Editora UFMG, 1 ed. 2005.
- PENNA, Daniel C. R.; ARAGÃO, Giani A. S.; FUSARO, Teresa C. **Complementariedade entre o monitoramento e análise de risco na gestão da segurança de barragens**. Foz do Iguaçu, 2015. Disponível em: http://www.cbdb.org.br/xxsngb/download/trabalhos_tecnicos/tema113/COMPLEMENTARIEDADE%20ENTRE%20O%20MONITORAMENTO%20E%20AN%3%81LISES%20DE%20RISCO%20NA%20GEST%3%83O%20DA%20SEGURAN%3%87A%20DE%20BARRAGENS.pdf. Acesso em: 3 ago. 2019.
- XI, G. Y.; YUE, J. P.; ZHOU, B. X.; TANG, P. **Application of an artificial immune algorithm on a statistical model of dam displacement**. Computer & Mathematics with Applications, v. 62, n. 10, p. 3980-3986, Novembro 2011.