

Controle estatístico multivariado de instrumentos de monitoramento de barragens

Emerson Lazzarotto¹

Centro de Engenharias e Ciências Exatas, UNIOESTE, Foz do Iguaçu, PR

Resumo. Barragens de usinas hidrelétricas são monitoradas por instrumentos que avaliam leituras de diferentes tipos de movimentos dos blocos que a compõe. No monitoramento é necessário prever os valores futuros de leituras e construir cartas estatísticas de controle para indicar eventuais anormalidades. Monitorar muitos instrumentos é operacionalmente inviável. Assim, este trabalho propõe uma metodologia multivariada para monitorar instrumentos de barragens de concreto, que integra: Análise de Componentes Principais (ACP), modelos ARIMAX-GARCH, gráfico de controle multivariado T^2 e amostragem não paramétrica *Kernel Density Estimation* (KDE). A partir dos dados de um conjunto de instrumentos da barragem, são calculadas, via método ACP, as Componentes Principais (CP's) e são produzidas previsões através de modelos ARIMAX-GARCH apenas para CP's representativas. A análise é realizada sob os resíduos de CP's via gráfico de controle multivariado T^2 , sem assumir a normalidade, em que são geradas densidades preditivas associadas aos valores da estatística T^2 , de modo não paramétrico, via KDE. Finalmente, o limite superior da carta de controle das CP's é obtido e são encontrados os pontos fora do limite de controle do gráfico T^2 . A metodologia foi aplicada utilizando dados de instrumentos da barragem de concreto da usina hidrelétrica de Itaipu. Os resultados mostraram que o método proposto contribui para a redução da análise de um alto número de instrumentos a poucas CP's que reúnem instrumentos correlacionados e com comportamento similar e para que a taxa de falsos alarmes se mantenha dentro dos valores esperados dando credibilidade a carta de controle multivariada.

Palavras-chave. Controle estatístico de qualidade, instrumentos de monitoramento de barragem, análise multivariada, gráfico de controle T^2 .

1 Introdução

As barragens de usinas hidrelétricas devem garantir que a produção, a geração e a segurança sejam preservadas durante o seu período operacional. Esta tarefa é executada, em grandes barragens, por meio de inspeções visuais e instrumentos de monitoramento. Como as barragens são obras muito particulares, mesmo com a determinação de limites de variação de projeto, estes podem não ser mais reais durante a fase de operação da barragem. Assim, é necessário estudar o comportamento dos dados de leituras dos instrumentos

¹emerson.lazzarotto@gmail.com

de modo que seja possível determinar limites estatísticos de controle confiáveis, prever com máxima acurácia possível o seu futuro comportamento e colaborar na indicação prévia de algum tipo de tendência que possa recomendar obras de reparo.

Gráficos de controle são mecanismos simples e eficazes de monitorar os dados da instrumentação. Entretanto, quando há necessidade de monitorar muitas variáveis, sobretudo correlacionadas, é sugerida a aplicação de gráficos de controle multivariados [7]. A estatística T^2 de Hotelling monitora a média e a covariância de várias variáveis. Porém, presume a hipótese de normalidade multivariada e a ausência desta hipótese pode aumentar sensivelmente a emissão de falsos alarmes e desacreditar o sistema de monitoramento.

A análise de componentes principais tem sido empregada para reduzir a análise em dados de instrumentos de barragens em gráficos de controle multivariados, criar variáveis não correlacionadas e construir distribuições não paramétricas para os resíduos do gráfico T^2 por [2], [6] e [10]. No Brasil, [5] limitou-se ao caso normal univariado.

Este trabalho propõe uma metodologia multivariada em instrumentos que monitoram barragens que objetiva reduzir o custo operacional das análises e diminuir a taxa de falsos alarmes, sem perda da capacidade de detecção de mudanças estatisticamente significativas. Uma aplicação na usina hidrelétrica de Itaipu é realizada para testá-la em um caso real.

2 Fundamentos teóricos

2.1 Análise de Componentes Principais (ACP)

A ACP é uma técnica multivariada em que variáveis correlacionadas são transformadas em um conjunto de variáveis não correlacionadas, capaz, em muitas ocasiões, de representar um problema multivariado em um espaço de dimensão muito mais reduzida [3].

Dado um vetor de dados de variáveis aleatórias $X' = (X_1, X_2, \dots, X_p)$, as CP's $Z_i = a'_i X = a_{i1}X_1 + a_{i2}X_2 + \dots + a_{ip}X_p$, onde $i = 1, \dots, p$ e $a'_i = (a_{i1}, \dots, a_{ip})$, segundo [4], são dadas por $Z_i = e'_i X = e_{i1}X_1 + e_{i2}X_2 + \dots + e_{ip}X_p$ e $Var(Z_i) = \mathbf{e}_i' \Sigma \mathbf{e}_i = \lambda_i$, onde $(\lambda_i, \mathbf{e}_i)$ é o i -ésimo par ordenado de autovalor/autovetor da matriz de covariância e $Var(Z_i)$ é a variância da CP Z_i . A proporção total de variância explicada pela k -ésima componente principal é dada por $V_e^k = \lambda_k / \sum_{i=1}^p \lambda_i$. Em um processo estável, os valores de algumas 'poucas' CP's também deverão estar estáveis e, se as CP's apresentam estabilidade, a variação aleatória das variáveis originais também deverá estar estável [4].

2.2 Modelos ARIMAX-GARCH

A abordagem de [1] é uma metodologia para análise de séries temporais e consiste do ajuste de modelos autorregressivos integrados de médias móveis (ARIMA(p, d, q)) que são modelos lineares aplicáveis a séries temporais que são consagrados na literatura pelo bom desempenho no ajuste e previsão.

Se y_t uma série temporal e $(x_{i,t})$ representam realizações de r variáveis exógenas de y_t , então y_t pode ser representada por um modelo ARIMAX (p, d, q), ARIMA com variáveis exógenas, dado por:

$$\nabla^d y_t = \phi_0 + \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j e_{t-j} + \sum_{i=1}^r \sum_{k=0}^l \alpha_{ik} x_{i,t-k} + e_t. \quad (1)$$

Modelos autorregressivos com heteroscedasticidade condicional generalizada -GARCH- são não lineares no que se refere à variância. No modelo GARCH(r, s) o choque aleatório e_t , em (1), é representado por $e_t = \sigma_t \epsilon_t$ de modo que a variância é dada por

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^r \alpha_i e_{t-i}^2 + \sum_{i=1}^s \beta_i \sigma_{t-i}^2. \quad (2)$$

Um modelo composto por (1) e (2) é denominado ARIMAX-GARCH. A análise de adequabilidade de um modelo se dá através de testes sobre a autocorrelação nos resíduos.

2.3 Kernel Density Estimation

Gráficos de controle não paramétricos são úteis quando não se tem conhecimento da verdadeira distribuição de probabilidades dos dados. Um núcleo estimador da densidade (KDE) é uma maneira não paramétrica de estimar a função densidade de probabilidade de uma variável aleatória. A técnica KDE oferece uma maneira não paramétrica de estimar a função densidade de probabilidade sem a especificação de um modelo paramétrico [9]. A núcleo-estimativa da densidade de uma série $X = (X_i), i = 1, \dots, N$ no ponto x dependendo número de observações, de um parâmetro de suavização e da núcleo-função K . Em particular, o núcleo de Epanechnikov K_e e da distribuição normal K_n , são dados por

$$K_e(x) = (3(1 - x^2))/4, -1 \leq x \leq 1 \text{ e } K_n(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, x \in \mathbb{R}. \quad (3)$$

2.4 Gráficos de controle multivariados

Na prática, há necessidade de monitorar o controle de diversas variáveis e, embora múltiplos gráficos univariados possam ser aplicados, isto pode conduzir a interpretações enganosas sobre o estado de um processo, sobretudo quando existir correlação entre as variáveis, deve-se examiná-las conjuntamente e não separadamente [7].

Dadas m observações de p variáveis $X_i = (X_{i1}, X_{i2}, \dots, X_{ip})$, a estatística T^2 , para dados que seguem uma distribuição normal multivariada, conforme [8], é dada por

$$T^2 = (X_i - \bar{X})' [S]^{-1} (X_i - \bar{X}) \quad (4)$$

onde \bar{X} é a média amostral e S é a matriz de covariância. Se $F_{\alpha,p,m-p}$ é o $1 - \alpha$ percentil da distribuição F com m e $m - p$ graus de liberdade e se $\beta_{\alpha,p/2,(m-p-1)/2}$ é o $1 - \alpha$ percentil da distribuição β com $p/2$ e $(m - p - 1)/2$ graus de liberdade, segundo [7], o limite superior de controle das fase II e I, são dados por:

$$LSC_{II} = \frac{p(m+1)(m-1)}{m(m-p)} F_{\alpha,p,m-p} \text{ e } LSC_I = \frac{(m-1)^2}{m} \beta_{\alpha,p/2,(m-p-1)/2}. \quad (5)$$

3 Metodologia Proposta

As principais etapas do método proposto são as seguintes.

- Análise e retenção de k componentes principais (ACP) das n variáveis.
- Modelagem ARIMAX-GARCH das componentes principais (CP's) retidas.
- Seleção dos melhores modelos ARIMAX-GARCH para cada CP.
- Obtenção dos resíduos das CP's retidas e dos valores da estatística T^2 dos resíduos.
- Construção de uma estimativa empírica da densidade via KDE.
- Avaliação e comparação da taxa de falsos alarmes obtido via suposição de normalidade multivariada e via KDE para o gráfico T^2 dos resíduos das CP's.

4 Resultados da aplicação

As séries temporais utilizadas são do Bloco E-6 da barragem da usina hidrelétrica de Itaipu Binacional, indicado na Figura 1. A amostra foi separada, conforme a Tabela 1, em amostra de treinamento, para construção e estimativa dos parâmetros dos modelos e amostra de validação, para avaliar a capacidade preditiva do modelo.



Figura 1: Barragem da usina de Itaipu e o trecho E em destaque.

Tabela 1: Separação da amostra.

Amostra	Fase	Qtd. dados	Período
Treinamento	I	360	10/82 a 09/12
Validação	II	23	10/12 a 08/14
Total		383	

Integram a análise realizadas médias mensais das séries de 4 tipos de instrumentos: 8 piezômetros (código 1A a 1H), 2 pêndulos (código 2X e 2Y), 12 bases de alongômetro (códigos 3A1, 3A2 até 3F1 a 3F2) e 10 extensômetros (código 4A, 4i1 a 4i3, i=B,C,D), totalizando 32 séries, além da temperatura ambiente e nível do lago com *lags* de retardo que são utilizados como variáveis exógenas (explicativas) nos modelos ARIMAX-GARCH.

A ACP levou a seleção (retenção) de cinco CP's (z_1, z_2, z_3, z_4, z_5) para posterior construção de modelos ARIMAX-GARCH e do gráfico T^2 devido ao fato que elas explicam quase 90% da variabilidade e porque estas 5 CP's incluem todas as variáveis significativamente em uma, e somente uma, das CP's. Assim, o monitoramento é simplificado e não perde nenhuma variável.

A CP z_1 é principalmente determinada pelo instrumento $2X$, além de todos do tipo 3, 4A, do grupo 4B e 1A. Variáveis com coeficiente negativo em z_1 têm tendência de decrescimento e as demais de crescimento, conforme a Figura 2, onde o eixo x é o tempo e o eixo y contém a leitura padronizada do instrumento. Estes devem ser considerados os instrumentos mais importantes da análise, porque apresentam maior variabilidade.

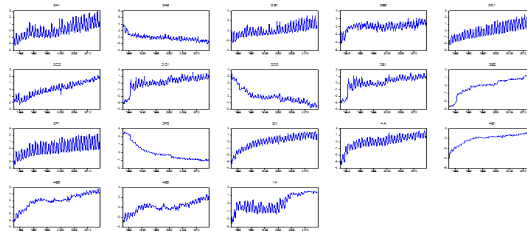


Figura 2: Instrumentos mais importantes da CP z_1 .

A CP z_2 é determinada pelos instrumentos: 1B,1C,1F,1G,1H e 2Y. z_2 é 'aproximadamente' a soma destas variáveis, inclusive todos os coeficientes delas são positivos em z_2 . Os instrumentos mais importantes na determinação da CP z_3 são 4C1, 4C2 e 4C3. A CP z_3 se aproxima da soma das contribuições do instrumento 4C nas hastes 1,2 e 3 e inclusive agrupa as 3 hastes do mesmo instrumento. A CP z_4 é principalmente determinada pelas 3 hastes do instrumento 4D e 1D. A CP z_5 é determinada por 1E.

A Tabela 2 apresenta características dos melhores modelos selecionadas para cada CP: o grau de explicação R^2 , os termos significativos da temperatura ambiente e dos modelos ARIMA. Para z_1 a ordem GARCH é (0, 2). Para as demais CP's os modelos são ARIMA. Nenhum coeficiente para o nível do lago mostrou-se significativo.

Tabela 2: Resultados dos melhores modelos encontrados.

CP	z_1	z_2	z_3	z_4	z_5
R^2	99,38	96,03	94,62	94,54	59,06
Tempamb	0 a 10	0 a 3	4	0 a 3	Nenhum
Arima	ar(1),sar(12) ma(1),sma(24)	ar(1),ar(3) sar(12),ma(12)	ar(1)	Ar(1),ar(2),sar(12) ma(1),ma(7),ma(8)	Ar(1),ar(10) ma(12)

Os valores da probabilidade do teste de Jarque-Bera obtidos levaram a rejeição da normalidade univariada dos resíduos das CP's e, conseqüentemente, da normalidade multivariada (MVN). Portanto, rejeita-se a MVN dos resíduos das CP's.

Para a construção do gráfico T^2 dos resíduos, foram incluídas 338 observações na fase

I (23 foram perdidas no processo de estimativa) e 361 na fase II. Valores do limite superior de controle (LSC) das fases I e II para taxas de falsos alarmes fixadas em $\alpha = 1\%$ e 5% do gráfico T^2 e o número de observações fora dos limites de controle (FLC) para o gráfico T^2 dos resíduos e das CP's foram obtidos. Todos os processos podem ser declarados fora do limite de controle, presumindo normalidade multivariada.

Para cada cenário da construção do KDE, se considerou os núcleos da equação (3), número de observações no núcleo $N = 200$ ou $N = 300$ e foi obtido o percentil de 95 e 99% que é o LSC exibido na Tabela 3 para a série temporal t_{res}^2 bem como a quantidade de observações acima do LSC para a fase I (FLC I) e da fase II (FLC II). A Tabela 3 apresenta ainda o LSC e FLC presumindo MVN. O LSC determinado via KDE faz com que o número de FLC do gráfico T^2 esteja próximo do número de falsos alarmes esperado.

Tabela 3: Resultados dos valores do LSC e FLC obtidos via KDE para a série temporal t_{res}^2 .

Série				t_{res}^2		95%		99%
Núcleo	N	LSC ₉₅	LSC ₉₈	LSC ₉₉	FLC _I	FLC _{II}	FLC _I	FLC _{II}
N	200	13,17	16,46	23,70	19	20	5	5
N	300	13,33	17,25	25,98	17	18	5	5
E	200	13,18	16,46	23,04	19	20	5	5
E	300	13,33	16,82	25,55	17	18	5	5
MVN fase I		10,97	13,22	14,85	28		12	
MVN fase II		11,07	13,39	15,09		29		12
Esperado					16,9	18,1	3,38	3,61

Para exemplificar estes cenários, a Figura 3 exibe o gráfico T^2 da série t_{res}^2 , $\alpha = 5\%$, incluindo dados das fases I e II, com o LSC determinado via normalidade multivariada (em vermelho), gerando 29 FLC e determinado via KDE (em rosa), gerando 18 FLC.

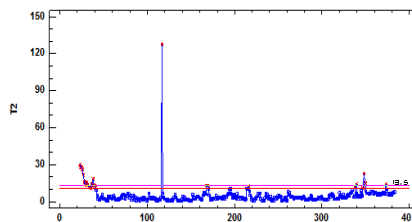


Figura 3: Gráfico T^2 da série t_{res}^2 , LSC MVN x KDE.

Quando o LSC do gráfico T^2 é determinado sem presumir normalidade multivariada, os falsos alarmes ficam dentro dos valores esperados e tornam viáveis os gráficos de controle multivariados aplicados ao monitoramento de instrumentos em Itaipu.

5 Conclusões

Este trabalho elaborou uma metodologia para monitorar dados de instrumentos de auscultação de barragem, estabelecer limites estatísticos de controle e reduzir o monitoramento a algumas CP's representativas e monitorar o gráfico T^2 dos resíduos onde o limite superior independe de normalidade. A aplicação em Itaipu mostrou que é possível reduzir a análise de 32 instrumentos à análise de 5 CP's e o LSC determinado via KDE reduz a taxa de falsos alarmes e dá credibilidade ao sistema de monitoramento (Tabela 3).

O modelo de CP's conseguiu fazer uma clusterização do conjunto de instrumentos em 5 conjuntos disjuntos. A hipótese de MVN não é satisfeita, gerando uma quantidade inaceitável de falsos alarmes no gráfico T^2 . Esta restrição do problema foi superada pelo uso do LSC obtido via KDE. Conforme se observa na Tabela 2, a temperatura ambiente e seus *lags* comprovou ser uma variável importante nos modelos ARIMAX-GARCH das CP's e, por consequência, das variáveis originais, os instrumentos. A variável do nível do reservatório não se mostrou significativa nos modelos utilizados. Uma justificativa para isso é a baixa variabilidade do nível do reservatório em Itaipu.

Referências

- [1] G. Box, G. Jenkins e G. Reinsel. *Time Series Analysis Forecasting and Control*. Hoboken-NJ: John Wiley and Sons, Inc., 2008.
- [2] L. Cheng e D. Zheng. Two online dam safety monitoring models based on the process of extracting environmental effect, *Adv. Engen. Software*, vol. 57, pp. 48-56, 2013.
- [3] J. Jackson. *A user's guide to principal components*. New York: John Wiley and Sons, Inc., 1991.
- [4] R. Johnson e D. W. Wichern. *Applied Multivariate Statistical Analysis - 6th. ed.*, Upper Saddle River - NJ: Pearson Prentice Hall, 2007.
- [5] S. Kuperman, *et all.*, Criteria to establish limit values of instrumentation readings for old embankment and concrete dams, *ASDSO Annual Conference*, 2004.
- [6] C. Loh, *et all.*, Applic. adv. statist. methods for extracting long-term trend in static monitoring data from an arch dam, *Struct. Health Monit.*, vol. 10, pp. 587-601, 2011.
- [7] R. Mason e J. Young. *Multivariate statistical process control with industrial applications*, Philadelphia: ASA - Society for Industrial and Applied Mathematics, 2002.
- [8] D. Montgomery. *Introdução ao Controle Estatístico de Qualidade*, 4^a ed., Rio de Janeiro-RJ: LTC, 2013.
- [9] M. Wand e M. Jones. *Kernel smoothing*, New York: Chapman and Hall, 1995.
- [10] H. Yu, Z. Wu, T. Bao e L. Zhang, Multivariate analysis in dam monitoring data with PCA, *Science China Technological Sciences*, v. 53, n.4, pp. 1088-1097, April 2010.