

MODELAGEM DE EQUAÇÕES ESTRUTURAIS APLICADA NA AVALIAÇÃO DOS FATORES RELACIONADOS COM A VARIABILIDADE DOS DADOS DE MONITORAMENTO DE UMA BARRAGEM DE CONCRETO

Sheila R. Oro^{1*}, Anselmo Chaves Neto², Suellen R. P. Garcia², Tereza R. Mafioleti²

1: Universidade Tecnológica Federal do Paraná
Campus Francisco Beltrão
Linha Santa Bárbara s/n CEP 85601-970 - Caixa Postal 135 - Francisco Beltrão, PR, Brasil
e-mail: sheilaro@utfpr.edu.br

2: Programa de Pós-Graduação em Métodos Numéricos em Engenharia
Universidade Federal do Paraná
Centro Politécnico, CEP 81531-990, Caixa Postal 19081, Curitiba, Paraná, Brasil
e-mail: anselmo@ufpr.br; suellenrpardo@yahoo.com.br; mafioleti@utfpr.edu.br
web: <http://www.ppgmne.ufpr.br/>

Palavras-chave: Modelagem de Equações Estruturais; Análise Fatorial; Monitoramento de Barragens

Resumo. *O monitoramento do comportamento estrutural de barragens de concreto é realizado principalmente através da análise dos dados da instrumentação. Neste estudo buscou-se identificar, analisar e medir as relações existentes entre os instrumentos instalados nos blocos chaves de uma barragem de contrafortes e os principais fatores que influenciam a variabilidade dos dados. A aplicação da técnica multivariada denominada análise fatorial permitiu verificar que aproximadamente 85,5% da variabilidade total observada no conjunto de dados é devida a três fatores. As cargas fatoriais permitiram medir as correlações entre os sensores e os fatores, e com base nessa medida, agrupar os sensores dominantes em cada fator. Com os resultados obtidos na análise fatorial foi criado um diagrama de caminhos no qual foram estabelecidas as relações entre os sensores e os fatores. O modelo de mensuração correspondente a esse diagrama foi testado e validado via análise fatorial confirmatória. Por fim, este modelo foi traduzido num modelo de equações estruturais, o que permitiu a estimação das várias relações de dependência entre os sensores e os fatores, incluindo os casos em que uma variável apresentava-se dependente em uma equação e independente em outra. O modelo identificado foi avaliado como plausível e de ajuste razoável.*

1. INTRODUÇÃO

O monitoramento de barragens, realizado com intuito de avaliar as condições de segurança de suas estruturas, consiste na identificação, análise e acompanhamento das alterações causadas pela incidência de deslocamentos, deformações, tensões, pressões, entre outros fenômenos. Essa avaliação é contínua e realizada por uma equipe especializada, através de inspeções visuais, auscultação geodésica de deslocamentos verticais e/ou horizontais, levantamentos batimétricos e instrumentação de auscultação. [1–4]

No caso de barragem de concreto, conforme [5], para monitorar as subpressões, recalques e deslocamentos dos blocos e de suas fundações são utilizados piezômetros e extensômetros de hastes, instalados na fundação. Enquanto que pêndulos diretos e invertidos, deformímetros, termômetros superficiais, tensômetros múltiplos e medidores mecânicos de junta, instalados na estrutura dos blocos, de acordo com [6], possibilitam avaliar o desempenho estrutural dos blocos da barragem.

A respeito do monitoramento estrutural da barragem de Itaipu, os relatórios técnicos, elaborados periodicamente pela equipe técnica de Itaipu, apresentam os resultados da instrumentação, sugestões de investigações adicionais, critérios a serem empregados no estabelecimento de valores de atenção e de alerta, entre outros.

É importante ressaltar que, dado o enorme volume de informações, originado pela frequência de medições e a quantidade de instrumentos em uso, o estabelecimento de relações entre a instrumentação, seguido da concentração dos dados, sem perda substancial de informação, pode agilizar o processo de identificação de anomalias.

A análise detalhada dos dados da instrumentação de auscultação de barragens, de acordo com [7], requer uma combinação de conhecimentos, especialmente de Engenharia, Matemática e Estatística, e experiência anterior da equipe técnica responsável pela análise destes dados. Isto pode consumir muito tempo, e muitas vezes faz com que seja impossível realizar essa tarefa de forma eficiente. Por isso, técnicas computacionais são indicadas como ferramentas auxiliares na importante tarefa de tomada de decisões.

Nesse contexto, as técnicas estatísticas multivariadas podem ser muito úteis no desenvolvimento de modelos, baseados em métodos numéricos, que captem o comportamento dos dados e representem adequadamente as relações existentes entre os instrumentos instalados na estrutura e fundação da barragem.

As técnicas de análise de agrupamentos e análise fatorial foram utilizadas por [7] para agrupar os instrumentos com comportamentos semelhantes e realizar a classificação hierárquica dos atributos e das hastes de extensômetros instaladas num determinado trecho da barragem de Itaipu.

Os estudos conduzidos por [8], utilizando algoritmos baseados em redes neurais autoassociativas (AANN), análise fatorial exploratória (EFA), quadrado da distância de Mahalanobis (MSD) e decomposição em valores singulares (SVD), mostraram que variações diárias e sazonais de temperatura podem produzir mudanças significativas nas características das estruturas.

Não foram encontrados na literatura relatos de estudos da aplicação da modelagem de equações estruturais no monitoramento de barragens.

No presente estudo, utilizou-se a análise fatorial e a modelagem de equações estruturais para identificar, analisar e medir as relações existentes entre os instrumentos e os principais fatores que influenciam a variabilidade dos dados. A complexidade dos modelos, o grande número de variáveis envolvidas e o desconhecimento da relação entre estas variáveis, podem ter dificultado a análise dos resultados.

2. DADOS

Os dados considerados nesta pesquisa foram cedidos pelo Centro de Estudos Avançados em Segurança de Barragens – CEASB. As leituras manuais realizadas no período compreendido entre janeiro de 1990 e dezembro de 2013 foram obtidas junto ao Sistema de Gestão de Medições - GSM. As 63 variáveis observáveis escolhidas são sensores de pêndulos, bases de alongâmetro, termômetros superficiais para concreto, piezômetros, hastes de extensômetros, nível de água do reservatório e temperatura ambiente.

3. TÉCNICAS ESTATÍSTICAS

Em pesquisas científicas, a inferência a partir de dados coletados pelo pesquisador, sobre determinados processos ou fenômenos, que não se pode diretamente observar, constitui-se num processo iterativo de aprendizado. De tal forma que para explicar um determinado fenômeno, o pesquisador coleta dados de diferentes variáveis e os analisa de acordo com uma hipótese. Com base nos resultados dessa análise, realiza modificações na explicação do fenômeno e procede a uma nova análise, até que os resultados sejam considerados satisfatórios.

Nesse contexto, as técnicas estatísticas podem ser utilizadas com o propósito de obter previsões de valores, simplificação estrutural, formação de grupos, investigação da dependência entre variáveis, construção e testes de hipóteses.

3.1. Séries Temporais

Série temporal, de acordo com [9], é um conjunto de observações acumuladas sequencialmente ao longo do tempo, com a característica de dependência serial. A modelagem de uma série temporal é um procedimento computacional que resulta em equações que permitem a previsão de valores futuros, com base nos valores passados, sem o uso de uma teoria subjacente.

Os principais modelos, de acordo com [9] [10], são: Autoregressivos de ordem (p) – AR(p); Médias Móveis de ordem (q) – MA(q); Mistos de ordem (p,q) – ARMA(p,q); Autoregressivo Integrado Médias Móveis de ordem (p, d, q) – ARIMA (p, d, q); modelos sazonais multiplicativos do tipo ARIMA [(p,d,q)×(P,D,Q)_s]; tal como apresentado nas Equações (1 a 5), respectivamente.

$$\phi(B)Z_t = a_t \quad (1)$$

$$Z_t = \delta + \theta(B)a_t \quad (2)$$

$$\phi(B)Z_t = \delta + \theta(B)a_t \quad (3)$$

$$\phi(B)\nabla^d Z_t = \theta(B)a_t \quad (4)$$

$$\Phi(B^S)\phi(B)\nabla_S^D \nabla^d Z_t = \Theta(B^S)\theta(B)a_t \quad (5)$$

Onde:

Z_t é a série temporal não estacionária;

B é o operador de retardo;

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p;$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q;$$

a_t é o ruído aleatório;

$$\nabla^d = (1 - B)^d;$$

$$\Phi(B^S) = 1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \Phi_p B^{pS};$$

$$\Theta(B) = 1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS};$$

$$\nabla_S^D = (1 - B^S)^D.$$

3.2. Análise Fatorial Exploratória

Conforme [11], dependendo dos objetivos do estudo, o pesquisador pode utilizar a análise fatorial exploratória (EFA) e/ou a análise fatorial confirmatória (CFA). A primeira deve ser utilizada quando não são conhecidas as relações subjacentes entre os conjuntos de dados. Enquanto que a segunda permite testar hipóteses a respeito da estrutura de um conjunto de dados e, dessa forma, construir teorias.

A seleção das variáveis explicativas de um processo ou fenômeno, de acordo com [10][12], pode ser feita a partir de p variáveis correlacionadas e, mediante a aplicação da análise fatorial exploratória, obtém-se um conjunto de m variáveis aleatórias não observáveis não correlacionadas ($m < p$), denominadas fatores, que possuem propriedades especiais em termos de variâncias.

O processo de obtenção dos fatores por componentes principais pode ser resumido na sequência de passos apresentados a seguir, utilizando os conceitos e notações descritos por [10], [13].

Passo 1: Coleta de n observações sobre p variáveis aleatórias distintas e organização na forma matricial (Eq. 6).

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1j} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2j} & \cdots & x_{2p} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{i1} & x_{i2} & \cdots & x_{ij} & \cdots & x_{ip} \\ \vdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nj} & \cdots & x_{np} \end{bmatrix} = [\underline{X}_1 \quad \underline{X}_2 \quad \cdots \quad \underline{X}_p] \quad (6)$$

Passo 2: Estimação das médias (Eq. 7) e variâncias (Eq. 8) das variáveis aleatórias.

$$\mu_j \approx \bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n} \quad , j = 1, 2, \dots, p \quad (7)$$

$$\sigma_j^2 \approx s_j^2 = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \quad , j = 1, 2, \dots, p \quad (8)$$

Passo 3: Padronização dos dados (Eq. 9).

$$\underline{Z}_j = \frac{\underline{X}_j - \mu_j}{\sigma_j} \quad (9)$$

Passo 4: Composição da matriz de covariâncias (Eq. 10) ou, de correlações (Eq. 11).

$$V(\underline{X}) = \Sigma \approx S = \begin{bmatrix} s_1^2 & s_{12} & \cdots & s_{1p} \\ s_{21} & s_2^2 & \cdots & s_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ s_{p1} & s_{p2} & \cdots & s_p^2 \end{bmatrix} \quad (10)$$

Onde

$$s_{jk} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k) \quad , j, k = 1, 2, \dots, p$$

$$\rho \approx R = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1p} \\ r_{21} & 1 & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & 1 \end{bmatrix} \quad (11)$$

Onde

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}\sqrt{s_{kk}}} \quad , j, k = 1, 2, \dots, p$$

Passo 5: Cálculo dos autovalores (λ_j) da matriz de covariâncias (ou de correlações) e os correspondentes autovetores (e_j).

Passo 6: Verificação da viabilidade do uso do modelo fatorial pelo Teste de Bartlett (Eq. 12) e da condição do ajuste do modelo ao conjunto de dados pelo Critério Kaiser-Meyer-Olkin (KMO). O coeficiente KMO (Eq. 13) varia entre 0 e 1. Quanto mais próximo de 1, melhor o ajuste do modelo fatorial aos dados.

$$H_0: \rho = I_{p \times p} \quad \times \quad H_A: \rho \neq I_{p \times p}$$

$$T = - \left[n - \frac{1}{6} (2p + 11) \right] \sum_{j=1}^p \ln \hat{\lambda}_j, \text{ com } T \sim \chi_v^2, v = \frac{p}{2}(p-1) \quad (12)$$

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} q_{ij}^2} \quad (13)$$

Onde

q_{ij} é o elemento pertencente à i -ésima linha e j -ésima coluna da matriz $Q = DR^{-1}D$, com $D = (\sqrt{\text{diagonal}(R^{-1})})^{-1}$.

Passo 7: Escolha do número (m) de fatores, pelo Critério de Kaiser ($\lambda > 1$) ou pelo grau de explicação. Composição do modelo fatorial ortogonal (Eq. 14 e Eq. 15), mediante a obtenção dos fatores (F_k) e suas respectivas cargas fatoriais (Eq. 16), juntamente com as communalidades (Eq. 17) e variâncias específicas (Eq. 18) das variáveis aleatórias originais.

$$\begin{cases} Z_1 = l_{11}F_1 + l_{12}F_2 + \dots + l_{1m}F_m + \varepsilon_1 \\ Z_2 = l_{21}F_1 + l_{22}F_2 + \dots + l_{2m}F_m + \varepsilon_2 \\ \vdots \\ Z_j = l_{j1}F_1 + l_{j2}F_2 + \dots + l_{jm}F_m + \varepsilon_j \\ \vdots \\ Z_p = l_{p1}F_1 + l_{p2}F_2 + \dots + l_{pm}F_m + \varepsilon_p \end{cases} \quad (14)$$

$$Z = L F + \varepsilon \quad (15)$$

Onde:

Z é a matriz de medidas amostrais padronizadas;

L é a matriz de cargas fatoriais, composta pelos elementos l_{ij} ;

F é a matriz de fatores;

ε é a matriz de resíduos.

$$l_{jk} = cov(Z_j, F_k) \quad (16)$$

$$h_j^2 = l_{j1}^2 + l_{j2}^2 + \dots + l_{jk}^2 + \dots + l_{jm}^2 \quad , j = 1, \dots, p ; k = 1, \dots, m \quad (17)$$

$$\psi_j = s_j^2 - \sum_k^m l_{jk}^2 \quad , j = 1, \dots, p \quad (18)$$

Passo 8: Se $m > 1$, realização da rotação *Varimax*, via utilização da matriz de transformação T, e atualização da matriz de cargas fatoriais (L^*) e dos fatores (F^*) para a composição do modelo rotacionado (Eq. 19);

$$Z = L^* F^* + \varepsilon \quad (19)$$

Onde

$$L^* = LT \quad \text{e} \quad F^* = T' F$$

Passo 9: Estimação dos escores fatoriais (Eq. 20).

$$\underline{\hat{f}}_i = (\hat{L}'_z \hat{L}_z)^{-1} \hat{L}'_z \underline{Z}_i \quad , i = 1, 2, \dots, n \quad (20)$$

Passo 10: Interpretação dos fatores.

3.3. Modelagem de Equações Estruturais

Modelagem de equações estruturais (SEM), de acordo com [14], consiste num conjunto de técnicas e procedimentos utilizados para examinar as relações simultâneas de dependência e independência entre variáveis observáveis e latentes. É um processo iterativo baseado na construção e teste de modelos hipotéticos, até que se encontre um modelo consistente com o modelo real da população. Porém, esse modelo real não é conhecido. Então, busca-se o modelo que mais se aproxima da matriz de covariância amostral, ou seja, aquele que se encaixa melhor na estrutura de covariância das variáveis.

Tal como ocorre com outras técnicas multivariadas, o uso da SEM supõe independência das observações, amostras aleatórias, linearidade de todas as relações e normalidade multivariada. Além disso, a ocorrência de valores atípicos (*outliers*) afeta o resultado da SEM, porque interfere na estrutura de variância-covariância das variáveis.

De acordo com [11] os estágios da modelagem de equações estruturais são: definição dos fatores; desenvolvimento do modelo de mensuração; planejamento de estudo para produção dos resultados; avaliação e validação do modelo de mensuração; especificação do modelo estrutural; avaliação e validação do modelo estrutural.

O modelo de mensuração é aquele que define as relações entre as variáveis observáveis e as latentes. Pode ser teórico ou construído com base nos resultados da EFA. Sua validação é feita

através da análise fatorial confirmatória (CFA). Enquanto que o modelo estrutural apresenta as relações entre as variáveis latentes.

As relações causais estabelecidas nos modelos, conforme [15], podem ser diretas, indiretas, espúrias ou associação sem correlação. Uma relação é dita direta quando uma variável causa efeito em outra diretamente, e indireta quando o faz através de uma terceira variável. Relações espúrias ocorrem quando duas variáveis tem uma variável em comum, que exerce efeito em ambas. Por fim, quando duas variáveis possuem uma variável comum, porém não é possível determinar se esta contribui para a covariância daquelas relações espúrias ou indiretas, tem-se a associação sem correlação.

Diagramas de caminhos são comumente usados para a representação visual dos modelos hipotéticos testados. Os caminhos dos fatores para as variáveis observáveis são indicados por setas. Conforme explica [11], cada caminho representa uma relação ou carga que deve existir com base no modelo de mensuração. Usando CFA, apenas as cargas que teoricamente conectam um uma variável observável ao seu correspondente fator são calculadas. Todas as demais são consideradas nulas. Essa é a principal diferença entre CFA e EFA.

De acordo com [10], os modelos são compostos por um conjunto de equações lineares (Eqs. 20-22) usadas para especificar os fenômenos em termos das relações de causa-e-efeito entre suas variáveis.

$$\eta_{(m \times 1)} = B_{(m \times m)}\eta_{(m \times 1)} + \Gamma_{(m \times n)}\xi_{(n \times 1)} + \zeta_{(m \times 1)} \quad (20)$$

$$Y_{(p \times 1)} = \Lambda_Y_{(p \times m)}\eta_{(m \times 1)} + \varepsilon_{(p \times 1)} \quad (21)$$

$$X_{(q \times 1)} = \Lambda_X_{(q \times n)}\xi_{(n \times 1)} + \delta_{(q \times 1)} \quad (22)$$

$$E(\zeta) = 0; \quad Cov(\zeta) = \Psi \quad (23)$$

$$E(\varepsilon) = 0; \quad Cov(\varepsilon) = \Theta_\varepsilon \quad (24)$$

$$E(\delta) = 0; \quad Cov(\delta) = \Theta_\delta \quad (25)$$

Onde $\zeta, \varepsilon, \delta$ são mutuamente não correlacionadas; $Cov(\xi) = \Phi$; ζ, δ são não correlacionadas com ξ ; ε é não correlacionada com η ; B tem zeros na diagonal; e $I - B$ não é singular. Assumindo as condições adicionais (Eqs. 23-25), tem-se $E(\xi) = 0$ e $E(\eta) = 0$.

As quantidades ξ e η , na equação (Eq. 20), são os valores das variáveis latentes, exógenas e endógenas, respectivamente, responsáveis pelas afirmações de causa-e-efeito. As quantidades Y e X , nas Eq. 21 e Eq. 22, são variáveis observáveis linearmente relacionadas com η e ξ através das matrizes de coeficientes de regressão Λ_Y e Λ_X . Os valores de ε e δ são os erros de mensuração de Y e X , respectivamente.

Parâmetros, em SEM, são representações numéricas das relações entre uma variável latente e

uma variável medida. Em CFA, são estimados pelas cargas fatoriais e testados quanto à sua significância. Parâmetros não significativos, com exceção das variâncias, conforme [15], podem ser interpretados como de pouca importância para o modelo e, adicionalmente, indicam que o tamanho da amostra é inadequado.

De acordo com [15], o objetivo inicial do pesquisador em SEM é confirmar um bom ajuste entre o modelo proposto e os dados amostrais. Há diversos índices de avaliação que permitem verificar a adequação geral do modelo. A Tabela 1 apresenta índices de ajuste utilizados neste estudo, com os respectivos valores de referência propostos por [11] e [14]. Uma regra prática, sugerida por [11], é considerar, juntamente com a plausibilidade teórica, pelo menos um índice de ajuste absoluto e um incremental, além do resultado de χ^2 . O teste χ^2 é utilizado para verificar se a matriz de covariância observada é significativamente diferente da matriz de covariância populacional. Conforme aponta [15], essa estatística é sensível ao tamanho da amostra e, na maioria das modelagens que utilizam SEM, dificilmente consegue-se valores baixos, desejáveis para sustentar a hipótese de modelos bem ajustados. Uma alternativa é considerar a razão entre a estatística χ^2 e o número de graus de liberdade (gl).

Estatística	Descrição	Valor de referência
$\frac{\chi^2}{gl}$		>5 – Ajuste ruim 2 a 5 – Ajuste razoável 1 a 2 – Ajuste bom ~1 – Ajuste muito bom
RMSEA	Índice de ajuste absoluto. $RMSEA = \sqrt{\frac{\chi^2 - gl_k}{n - 1}}$	>0,10 – Ajuste ruim 0,5 a 0,10 – Ajuste bom <=0,5 – Ajuste muito bom
CFI	Índice de ajuste incremental $CFI = 1 - \frac{\chi^2_k - gl_k}{\chi^2_n - gl_n}$	>0,9

Tabela1. Índices de ajuste e valores de referência para verificação da adequação do modelo

4. IMPLEMENTAÇÃO DAS TÉCNICAS

Os dados de entrada referem-se às observações registradas por 63 sensores instalados nos blocos D07 e D08 da Barragem de Contrafortes de Itaipu. O fluxograma (Fig. 1) apresenta a sequência de etapas de aplicação das técnicas após a coleta dos dados.

As observações correspondem às leituras realizadas manualmente por auxiliares técnicos, com periodicidade mensal. Porém, no início do período considerado neste estudo, a frequência de leituras era maior. Por isso, resolveu-se trabalhar com as médias mensais das observações. Com o auxílio do *software* Matlab [16], foi implementado um algoritmo para eliminar dados redundantes, gerar as médias mensais das leituras manuais dos instrumentos e identificar lacunas no conjunto de dados, ocasionadas por ausência de leituras em determinados períodos.

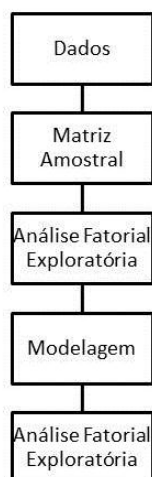


Figura1. Fluxograma das etapas de implementação das técnicas

Em seguida, com o auxílio do *software* Statgraphics [17], cada sensor teve suas medições modeladas por séries temporais e, com base no modelo obtido, foram feitas as previsões para os dados faltantes, utilizando o procedimento *forecasting/backforecasting*.

Ao final desta etapa, obteve-se a matriz de dados amostrais, de ordem 288x63, cujas linhas armazenam as médias mensais das observações e as colunas correspondem aos sensores.

Os sensores X44 e X50 apresentavam gráficos praticamente constantes e foram desconsiderados.

4.1. Identificação dos Fatores

A aplicação da análise fatorial exploratória, realizada com o auxílio do *software* Statgraphics Centurion XVI, possibilitou a avaliação estatística das características dos dados amostrais.

Os sensores X1, X30, X39, X40, X42, X43, X46, X51, X53, X54 e X58, por apresentarem comunalidades abaixo de 0,60, foram descartados.

O critério de seleção adotado foi: nº de fatores que explicam, pelo menos, 80% da variância. Assim, foram identificados três fatores (Tab. 2) que explicam aproximadamente 85,5% da variabilidade total observada no conjunto de dados dos 50 sensores restantes.

Fator	Autovalor	Variância (%)	Variância Acumulada (%)
1	22,1137	44,227	44,227
2	15,3596	30,719	74,947
3	5,27118	10,542	85,489
4	1,70076	3,402	88,89
5	1,09229	2,185	91,075
6	0,533952	1,068	92,143

Tabela2. Seleção do número de fatores com base na variância acumulada

A importância individual dos sensores para cada fator foi medida pela carga fatorial. Os carregamentos (Tab. 3) foram obtidos pelo método das Componentes Principais e a Rotação *Varimax* permitiu o refinamento do processo. Dessa forma, foi possível verificar que cada fator é dominado por um grupo de sensores. A explicação para esta ocorrência é a incidência de um fenômeno externo, não mensurável, cuja interpretação nem sempre é de fácil realização.

Sensor	F1	Sensor	F2	Sensor	F3
X56	0,894187	X26	0,925628	X29	0,936965
X59	0,888431	X24	0,872214	X35	0,93329
X48	0,882771	X10	0,862935	X33	0,9157
X47	0,876733	X9	0,861962	X37	0,915316
X45	0,874272	X25	0,81847	X27	0,912719
X7	0,863159	X21	0,793844	X31	0,885039
X55	0,849457	X5	0,770571	X36	0,853084
X15	0,841515	X23	0,753309	X6	0,823452
X57	0,838069	X3	0,750221	X4	0,7968
X11	0,821506	X22	0,722697	X2	0,750382
X16	0,80819	X13	0,718208	X52	0,711591
X8	0,80755	X18	0,716568	X32	-0,74948
X12	0,761874	X17	0,710969	X39	-0,82742
X41	0,759741	X63	-0,76388		
X34	0,697353	X62	-0,86509		
X28	0,688956	X61	-0,89089		
X19	0,686834				
X14	0,681328				
X20	0,672018				
X38	0,637881				
X60	-0,85742				

Tabela3. Carregamentos dos fatores em cada sensor

4.2. Modelo de Mensuração

Com base nos resultados da análise fatorial exploratória foi composto o modelo fatorial apresentado a seguir.

$$F1 = \sim X56 + X59 + X48 + X47 + X55 + X15 + X57 + X11 + X16 + X8 + X12 + X41 + X34 + X28 + X19 + X14 + X20 + X38 + X60$$

$$F2 = \sim X26 + X24 + X10 + X9 + X25 + X21 + X5 + X23 + X3 + X22 + X13 + X18 + X17 + X63 + X62 + X61$$

$$F3 = \sim X29 + X35 + X33 + X37 + X27 + X31 + X36 + X6 + X4 + X2 + X52 + X32 + X49$$

As equações estruturais que definem as relações entre as variáveis observáveis (sensores) e não observáveis (fatores) são as seguintes:

$X56 = l_{X56}F_1 + e_{X56}$	$X26 = l_{X26}F_2 + e_{X26}$	$X29 = l_{X29}F_3 + e_{X29}$
$X59 = l_{X59}F_1 + e_{X59}$	$X24 = l_{X24}F_2 + e_{X24}$	$X29 = l_{X29}F_3 + e_{X29}$
$X48 = l_{X48}F_1 + e_{X48}$	$X10 = l_{X10}F_2 + e_{X10}$	$X35 = l_{X35}F_3 + e_{X35}$
$X47 = l_{X47}F_1 + e_{X47}$	$X9 = l_{X9}F_2 + e_{X9}$	$X33 = l_{X33}F_3 + e_{X33}$
$X45 = l_{X45}F_1 + e_{X45}$	$X25 = l_{X25}F_2 + e_{X25}$	$X37 = l_{X37}F_3 + e_{X37}$
$X7 = l_{X7}F_1 + e_{X7}$	$X21 = l_{X21}F_2 + e_{X21}$	$X27 = l_{X27}F_3 + e_{X27}$
$X55 = l_{X55}F_1 + e_{X55}$	$X5 = l_{X5}F_2 + e_{X5}$	$X31 = l_{X31}F_3 + e_{X31}$
$X15 = l_{X15}F_1 + e_{X15}$	$X23 = l_{X23}F_2 + e_{X23}$	$X36 = l_{X36}F_3 + e_{X36}$
$X57 = l_{X57}F_1 + e_{X57}$	$X3 = l_{X3}F_2 + e_{X3}$	$X6 = l_{X6}F_3 + e_{X6}$
$X11 = l_{X11}F_1 + e_{X11}$	$X22 = l_{X22}F_2 + e_{X22}$	$X4 = l_{X4}F_3 + e_{X4}$
$X16 = l_{X16}F_1 + e_{X16}$	$X13 = l_{X13}F_2 + e_{X13}$	$X2 = l_{X2}F_3 + e_{X2}$
$X8 = l_{X8}F_1 + e_{X8}$	$X18 = l_{X18}F_2 + e_{X18}$	$X52 = l_{X52}F_3 + e_{X52}$
$X12 = l_{X12}F_1 + e_{X12}$	$X17 = l_{X17}F_2 + e_{X17}$	$X32 = l_{X32}F_3 + e_{X32}$
$X41 = l_{X41}F_1 + e_{X41}$	$X63 = l_{X63}F_2 + e_{X63}$	$X49 = l_{X49}F_3 + e_{X49}$
$X34 = l_{X34}F_1 + e_{X34}$	$X26 = l_{X26}F_2 + e_{X26}$	
$X28 = l_{X28}F_1 + e_{X28}$	$X62 = l_{X62}F_2 + e_{X62}$	
$X19 = l_{X56}F_1 + e_{X56}$	$X61 = l_{X61}F_2 + e_{X61}$	
$X14 = l_{X56}F_1 + e_{X56}$		
$X20 = l_{X56}F_1 + e_{X56}$		
$X38 = l_{X56}F_1 + e_{X56}$		
$X60 = l_{X56}F_1 + e_{X56}$		

O modelo padrão de medida correspondente, representado pelo diagrama de caminhos de relações causais (Fig. 2), apresenta os sensores ligados aos seus respectivos fatores. A cada sensor está associado um termo de erro de medida, que pode ter origem aleatória ou ser derivado de uma determinada característica única da variável observada.

A análise fatorial confirmatória, operacionalizada com o auxílio do *software* R, permitiu a estimação dos 97 parâmetros livres desse modelo, referentes a 47 cargas fatoriais e 50 variâncias residuais. Adotou-se, neste caso, que os fatores são não correlacionados e de variância unitária. O método de estimação utilizado foi Mínimos Quadrados Generalizados (GLS), porque a hipótese de normalidade multivariada dos dados foi negada (valor-p = 0, no teste Shapiro-Francia).

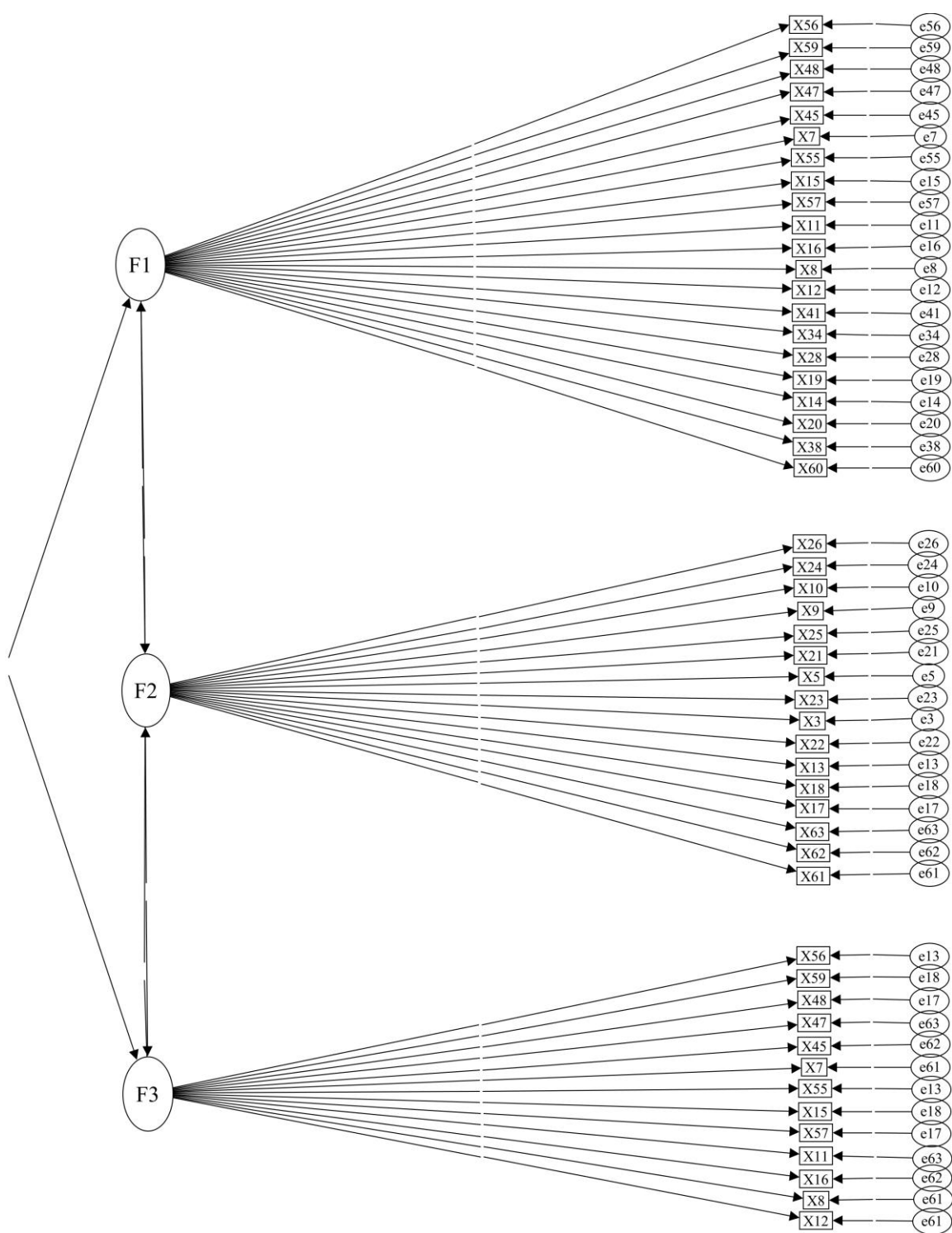


Figura2. Diagrama de caminhos para o modelo padrão de medidas

O tamanho da amostra, composta por 288 observações por sensor, pode ser considerado pequeno. Pois, de acordo com [11], deve-se trabalhar com 5 a 10 casos por parâmetro estimado. Isso significa que para este estudo seria necessário mais de 500 observações. Porém, para obter esse valor seria necessário utilizar dados a partir 1971, o que é impossível, por tratar-se de período anterior à época da construção/operação da barragem. Mesmo assim a análise foi realizada, visto que alguns autores sugerem o uso de amostras maiores que 200.

A análise das estimativas dos parâmetros foi feita com base no valor-p, adotando-se o nível de significância de 0,05, para verificar se a estimativa era estatisticamente diferente de zero, o que indica que o parâmetro era importante para o modelo, exceto quando se tratava de variância residual. Uma parte dos resultados encontra-se apresentada na Figura 3. As colunas informam, respectivamente, o tipo de parâmetro estimado, a estimativa da carga fatorial, o erro padrão, a estatística z, o valor-p, a estimativa da carga fatorial, a estimativa padronizada da carga fatorial.

	Estimate	Std.err	Z-value	P(> z)	Std.lv	Std.all
Latent variables:						
F1 =~						
X56	1.000				1.000	0.084
X59	3.221	1.120	2.877	0.004	3.221	0.266
X48	-6.375	2.044	-3.119	0.002	-6.375	-0.228
X47	-28.828	1.760	-16.375	0.000	-28.828	-0.977
X45	-30.123	2.020	-14.911	0.000	-30.123	-0.899
X7	-0.246	0.481	-0.510	0.610	-0.246	-0.053
X55	-1.502	1.885	-0.797	0.426	-1.502	-0.074
X15	-0.356	0.395	-0.900	0.368	-0.356	-0.100
X57	4.262	3.106	1.372	0.170	4.262	0.134
X11	-0.755	0.766	-0.987	0.324	-0.755	-0.087
X16	1.597	0.485	3.292	0.001	1.597	0.341
X8	-1.232	0.594	-2.076	0.038	-1.232	-0.176
X12	0.432	0.560	0.771	0.441	0.432	0.069
X41	-1.507	1.520	-0.991	0.321	-1.507	-0.100
X34	-2.837	1.604	-1.769	0.077	-2.837	-0.143
X28	-2.379	1.617	-1.471	0.141	-2.379	-0.111
X19	-2.302	0.642	-3.588	0.000	-2.302	-0.402
X14	0.497	0.652	0.763	0.445	0.497	0.065

Figura 3. Primeiras estimativas obtidas para os parâmetros

Dentre todas estimativas, 34% não foram significativas, indicando que os respectivos parâmetros não são importantes para o modelo e poderiam ser excluídos. No entanto, a inexistência de variâncias negativas e de dados discrepantes garante a plausibilidade do modelo.

O total de 1178 graus de liberdade – resultante de 1275 momentos amostrais distintos (variâncias e covariâncias dos sensores), subtraído de 97 parâmetros estimados – indicou um modelo superidentificado, o que permitiu que o modelo fosse testado.

Os fatores não são observáveis diretamente, logo não possuem uma escala métrica definida. Essa dificuldade foi contornada atribuindo-se o valor 1 à carga fatorial do primeiro sensor de cada fator.

Para avaliação do ajuste do modelo foram utilizados os índices apresentados na Tabela 4. O

critério baseado na estatística χ^2 rejeita a hipótese de que o modelo esteja bem ajustado. Por este motivo, é melhor tomar como referência a razão entre χ^2 e os graus de liberdade (χ^2/gl). Os valores dos índices χ^2/gl e RMSEA mostram um ajuste razoável aos dados amostrais. Enquanto que CFI indica pode ser diferentemente especificado. Isso se justifica porque, em geral, esse teste é sensível ao tamanho da amostra, que neste caso, é pequeno.

Estatística	Valor de referência	Resultado
χ^2/gl	>5 – Ajuste ruim 2 a 5 – Ajuste razoável 1 a 2 – Ajuste bom ~1 – Ajuste muito bom	1,935
RMSEA	>0,10 – Ajuste ruim 0,5 a 0,10 – Ajuste bom <=0,5 – Ajuste muito bom	0,057
CFI	>0,9	0,163

Tabela4. Índices de ajuste do modelo

Para identificar as mudanças que melhorariam a especificação do modelo, foram analisados os índices de modificação (IM). Os valores desses índices indicam o quanto seria reduzido o valor de χ^2 caso fosse atribuída uma covariância entre os erros de duas variáveis. A Tabela 5 apresenta os maiores valores de IM obtidos para o modelo testado. A coluna EPC informa o valor padronizado da estimativa do parâmetro acrescentado ao modelo.

Covariâncias	MI	EPC
X18 ~~ X17	57,778	3,241
X56 ~~ X59	43,239	4,14
X19 ~~ X20	34,465	1,407
X23 ~~ X22	24,442	0,679
X11 ~~ X12	22,574	0,509
Total	182,498	

Tabela5. Índices de modificação para o modelo

As mudanças sugeridas são as mais vantajosas e correspondem à existência de correlação entre os erros de medição dos pares de variáveis apresentados. O maior índice de modificação foi observado entre as variáveis X18 e X17, que representam hastes de extensômetros instaladas no mesmo furo de sondagem da fundação do bloco D08. X56 e X59 (piezômetros instalados na fundação do bloco D08, na junta E); X19 e X20 (hastes de extensômetros instaladas no mesmo furo de sondagem da fundação do bloco D08); X23 e X22 (hastes de extensômetros instaladas no mesmo furo de sondagem da fundação do bloco D08); X11 e X12 (hastes de extensômetros instaladas no mesmo furo de sondagem do bloco D07).

Qualquer modificação só pode ser efetuada se tiver significado prático ou base teórica subjacente. Neste caso, a realização das mudanças sugeridas pelos MI, acrescentaria cinco parâmetros ao modelo e reduziria 182 unidades o valor de χ^2 , no mínimo, resultando num modelo melhor especificado.

5. CONCLUSÕES

- No presente estudo, o conjunto de observações relativas às séries temporais de 63 sensores instalados nos blocos chaves D07 e D08, foi submetido à análise fatorial exploratória, o que possibilitou a redução e a representação das variáveis medidas por um subconjunto, composto por três fatores, que concentrou as informações sobre as respostas da estrutura de concreto e da fundação aos diversos fenômenos que atuam sobre ela. Assim, foi possível verificar que cerca de 85% da variabilidade total observada no conjunto de dados é explicada pelos fatores identificados.
- Os carregamentos dos fatores em cada variável medida, que correspondem às estimativas das correlações, permitiram o agrupamento dos sensores. Assim, garantiu-se que elementos de um mesmo grupo são altamente correlacionados com o respectivo fator, devido à influência de um fenômeno externo, e não correlacionados com os demais.
- Os resultados obtidos na EFA serviram de base para a elaboração do modelo de mensuração, plausível e de ajuste razoável.
- A aplicação da análise fatorial confirmatória permitiu verificar que a maioria das relações estabelecidas entre os sensores e os fatores foi significativa. O que representa que uma variação em um fator provoca uma variação correspondente nestes sensores.
- É importante salientar que o aumento do tamanho da amostra e a inclusão dos parâmetros sugeridos pelos índices de modificação (IM) poderiam melhorar o ajustamento do modelo de mensuração. Também podem ser testados, a título de comparação outros modelos hipotéticos. Sugere-se a realização dessas alterações e nova análise do modelo para, só então, proceder a um estudo aprofundado das relações estruturais entre os fatores.

AGRADECIMENTOS

Deixo expressos meus sinceros agradecimentos às seguintes instituições, sem as quais o presente trabalho teria sido impossível:

- ao CEASB – Itaipu Binacional, pelo apoio financeiro, suporte técnico e disponibilização dos dados utilizados neste estudo;
- a PROPPG – UTFPR, pela viabilização deste estudo via Portaria nº. 0398/2014;
- ao PPGMNE – UFPR, pelo incentivo ao desenvolvimento deste trabalho.

REFERÊNCIAS

- [1] P. T. Cruz, *100 barragens brasileiras: casos históricos, materiais de construção, projeto*. São Paulo: Oficina dos Textos, 2006.
- [2] R. F. Leme, T. M. de Carvalho, L. N. de Souza, F. A. de Souza Filho, R. M. Luna, “Segurança de barragens: a Lei 12.334/2012 e a experiência do Ceará,” apresentado no XI Simpósio de recursos Hídricos do Nordeste, João Pessoa - PB, 2012.
- [3] J. V. Carvalho, C. Romanel, “Redes neurais temporais aplicadas ao monitoramento de barragens,” *Revista Eletrônica de Sistemas de Informação*, no. 10, 2007.

- [4] S. C. Kuperman, M. R. Moretti, S. Cifu, T. B. Celestino, G. Re, K. Zoellner, “Criteria to establish limit values of instrumentation readings for old embankment and concrete dams,” 2005.
- [5] S. F. Matos, “Avaliação de instrumentos para auscultação de barragem de concreto. Estudo de caso: deformímetros e tensômetros para concreto na Barragem de Itaipu,” Dissertação, UFPR, Curitiba, PR, 2002.
- [6] Itaipu, *Usina Hidrelétrica de Itaipu - Aspectos de Engenharia*. Foz do Iguaçu: Itaipu Binacional, 2009.
- [7] R. Villwock, M. T. A. Steiner, A. S. Dyminski, A. Chaves Neto, “Itaipu Hydroelectric Power Plant Structural Geotechnical Instrumentation Temporal Data Under the Application of Multivariate Analysis - Grouping and Ranking Techniques,” in *Multivariate Analysis in Management, Engineering and the Sciences*, InTech, 2013, pp. 81–102.
- [8] E. Figueiredo, G. Park, C. R. Farrar, K. Worden, J. Figueiras, “Machine learning algorithms for damage detection under operational and environmental variability,” *Struct. Health Monit.*, vol. 10, no. 6, pp. 559 – 572, Nov. 2011.
- [9] E. L. K. Box and G. M. Jenkins, *Time Series Analysis: forecasting and control*. Holden-Day, 1976.
- [10] R. A. Johnson, D. W. Wichern, *Applied Multivariate Statistical Analysis*, 6th ed. Pearson, 2007.
- [11] J. F. Hair, B. Black, B. Babin, R. E. Anderson, R. L. Tatham, *Análise Multivariada de Dados*. Porto Alegre: Bookman, 2009.
- [12] J. M. Lattin, J. D. Carrol, P. E. Green, *Análise de Dados Multivariados*. Cengage Learning, 2011.
- [13] S. A. Mingoti, *Análise de dados através de métodos de estatística multivariada: uma abordagem aplicada*. Editora UFMG, 2005.
- [14] R. E. Schumacker and L. G. Lomax, *A Beginner’s Guide to Structural Equation Modeling*, 3^a ed. Routledge, 2010.
- [15] J. S. F. Silva, “Modelagem de Equações Estruturais: apresentação de uma metodologia,” Dissertação (Mestrado em Engenharia de Produção), UFRGS, Porto Alegre, 2006.
- [16] *Matlab R2013*. MathWorks, 2013.
- [17] *Statgraphics Centurion XVI*. StatPoint Technologies, 2010.