

MONITORAMENTO DE MOVIMENTOS RELATIVOS EM BARRAGEM DE CONCRETO VIA MODELO ESTATÍSTICO MULTIVARIADO

Suellen R. P. Garcia^{1*}, Anselmo C. Neto², Sheila R. Oro¹ e Claudio N. Júnior³

1: Universidade Tecnológica Federal do Paraná
Campus Toledo
Rua Cristo Rei, 19 CEP 85902-490, Toledo-PR, Brasil
e-mail: suellenpardo@utfpr.edu.br; 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
web: <http://www.ppgmne.ufpr.br/>

3: Divisão de Engenharia Civil e Arquitetura - Itaipu Binacional
Avenida Tancredo Neves, 6731, CEP 85866-900, Foz do Iguaçu-PR, Brasil
e-mail: neumann@itaipu.gov.br

Palavras-chave: Monitoramento de Barragem, Modelo Multivariado, Mínimos Quadrados Parciais, PLSR, Deslocamentos

Resumo.

Os deslocamentos diferenciais entre blocos representam comportamentos estruturais importantes no monitoramento de uma barragem de concreto. Os pêndulos diretos e as bases de alongâmetro captam os deslocamentos que ocorrem devido à variação do nível do reservatório, variação de temperatura e eventuais deformações permanentes. Modelos estatísticos são geralmente utilizados para previsões, baseados em correlações existentes entre as variáveis ambientais e a resposta da barragem. Neste trabalho apresenta-se uma metodologia para a construção de um modelo de regressão por mínimos quadrados parciais para dados de monitoramento da Barragem de ITAIPU, localizada entre Brasil e Paraguai. O modelo reduz para 4 componentes um conjunto de 33 sensores de instrumentos (pêndulos diretos, bases de alongâmetro e termômetros) e as leituras do nível do reservatório, ou seja, 34 variáveis. As 4 componentes explicam 90% das variáveis preditoras e mais de 70% de cada variável de resposta (deslocamentos). O modelo é validado com o processo leave-on-out de validação cruzada e contribui com a previsão de novas observações.

1. INTRODUÇÃO

O monitoramento da estrutura de uma barragem é realizado por meio de inspeções visuais e instrumentação. A instrumentação instalada gera uma enorme massa de dados, que se analisada devidamente, fornece informações sobre o comportamento da estrutura mediante efeitos externos, como por exemplo, variações do nível do reservatório e da temperatura ambiente.

O objetivo da análise dos dados da instrumentação é propiciar informação que pode ser usada em uma interpretação física das deformações e, na previsão, seja do comportamento futuro da própria estrutura ou para estudo do comportamento de uma estrutura semelhante [1]. Modelos estatísticos são propostos na literatura, para tal objetivo. Estes são muitas vezes baseados em correlações existentes entre fatores tais como: o nível de água do reservatório, temperatura ambiente, desgaste devido ao tempo e a resposta da barragem a estas ações como tensões, deformações e deslocamentos [2].

Dois grandes desafios encontrados ao propor tais modelos para dados de monitoramento de barragens é que primeiro, as variáveis independentes, ou seja, variações do nível do reservatório e temperatura, podem gerar dados multicolineares, de modo que não seja possível utilizar algumas técnicas estatísticas clássicas. Segundo, o número de variáveis dependentes é geralmente alto, pois são muitos os sensores de instrumentação de uma barragem de concreto. A multicolinearidade cria uma variância compartilhada entre as variáveis, diminuindo assim a capacidade de prever a variável dependente, bem como examinar a importância relativa de cada variável independente [3].

O método de regressão por mínimos quadrados parciais (*Partial Least Squares regression* - PLSR) não tem como hipótese que as variáveis sejam não correlacionadas e nem requer que os resíduos sigam uma distribuição normal, como ocorre na regressão por mínimos quadrados ordinários (*Ordinary Least Squares* - OLS). O método PLSR utiliza as componentes PLS obtidas, a fim de maximizar a covariância entre as variáveis independentes e as variáveis dependentes [4].

PLSR generaliza e combina características de regressão multivariada (MR), análise de correlação canônica (CCA) e Análise de Componentes Principais (PCA) sem impor suas restrições [1].

PLS (*Partial Least Squares*) foi desenvolvido na década de 1960 por Herman Wold como uma técnica econométrica, mas os maiores defensores são engenheiros químicos. PLS é aplicada em calibração de espectrometria, no monitoramento e controle de processos industriais onde um processo pode ter centenas de variáveis controláveis e dezenas de saídas [5]. Outras aplicações foram em medicina, psicologia, agropecuária, entre outras áreas.

Entre as aplicações de PLSR, destaca-se o trabalho de [1] que apresenta uma análise de deformação tridimensional para um único ponto sobre a barragem. A análise consiste em um modelo PLSR de ajuste aos dados, previsão de deformação e análise da contribuição de fatores individuais. A metodologia foi empregada em uma barragem de terra localizada na região central da China. A conclusão no artigo foi que o modelo PLSR é mais confiável e tem melhor integridade do que o modelo de regressão múltipla, que segundo o autor foi amplamente empregado no monitoramento de barragens.

Neste contexto, desenvolver um modelo estatístico PLSR é a proposta deste artigo, onde as leituras dos pêndulos diretos e bases de alongâmetro instalados no trecho F da Barragem

Principal de ITAIPU (deslocamentos) compõem a matriz de variáveis dependentes e as variáveis independentes (preditoras) são as leituras da variação do nível do reservatório e as leituras dos termômetros de superfície. O aspecto relevante deste modelo é sua característica multivariada, ou seja, será proposto um modelo para prever as leituras futuras de vários pêndulos diretos e bases de alongâmetro simultaneamente, o que pode auxiliar no monitoramento.

2. REGRESSÃO POR MÍNIMOS QUADRADOS PARCIAIS (PLSR)

O método PLSR é uma técnica de estimação do modelo de regressão linear, baseada na decomposição das matrizes de variáveis respostas e de variáveis preditoras. O algoritmo usado PLSR examina ambas as matrizes e extrai componentes que são diretamente relevantes a ambos os conjuntos de variáveis [2].

Segundo [6], o método de regressão PLS apresenta as vantagens de:

- Modelar regressões com múltiplas variáveis respostas;
- Aceitar multicolinearidade e;
- Os fatores produzidos têm alto poder de predição, devido às altas covariâncias com as variáveis resposta.

As desvantagens do método são:

- Dificuldade na interpretação das cargas dos fatores;
- Os testes de significância dos estimadores dos coeficientes de regressão são realizados via métodos de reamostragem, pois suas distribuições não são conhecidas e;
- Falta de estatísticas de teste para o modelo.

O método de regressão por PLS é, preferencialmente, uma técnica de predição, e não interpretação, apesar de existirem trabalhos que fazem interpretação dos fatores extraídos via PLS [6]. Devido ao caráter de predição do método, sua aplicação no monitoramento de barragens se torna interessante, pois a comparação de valores preditos pelo modelo aos valores observados traz informação útil sobre o comportamento da barragem.

3.1. Descrição do modelo PLSR

Segue abaixo a descrição do método de regressão PLS baseada no trabalho de [7].

Como em regressão linear multivariada, o objetivo da regressão por PLS é construir um modelo linear $Y = X\beta + \varepsilon$, Y é uma matriz $n \times m$ de variáveis de resposta, X é uma matriz $n \times r$ de variáveis preditoras, β é uma matriz $r \times m$ dos coeficientes de regressão e ε é a matriz dos resíduos $n \times m$. A regressão PLS é uma abordagem livre de distribuição, dessa forma, os resíduos possuem vetor de médias nulo e matriz de covariâncias igual a $\sigma^2 I$, onde I é a matriz identidade de ordem $n \times n$, mas sem distribuição definida.

O modelo de regressão por PLS encontra poucas “novas” variáveis chamadas de escores de X , ou componentes, ou fatores, que são denotadas por t_a ($a = 1, 2, \dots, A$). Estas componentes são preditoras de Y e também modelam X (equações (4) e (2)), ou seja, ambas variáveis e

são assumidas, ao menos parcialmente, modeladas pelas mesmas variáveis latentes. O objetivo de extrair componentes que consigam capturar as variâncias das variáveis preditoras e também obter correlações com as variáveis de resposta, é alcançado maximizando a covariância entre X , t_a e Y [6].

O número de componentes A é menor do que o número de variáveis preditoras ($A < r$) e estes são ortogonais, obtidos por combinações lineares das variáveis originais $\underline{x}_r$, com os pesos $\underline{w}_a$ ($a=1,2,...,A$), da seguinte forma:

$$T = XW \quad (1)$$

onde $T = (t_1, t_2, \dots, t_A)$ é a matriz $n \times a$ de componentes (escores ou fatores) e $W = (\underline{w}_1, \underline{w}_2, \dots, \underline{w}_A)$ é a matriz $r \times a$ de pesos.

As matrizes X e Y são decompostas na forma:

$$X = TP' + F \quad (2)$$

$$Y = UC' + G \quad (3)$$

onde T e U são matrizes $n \times A$ de componentes (escores ou fatores) de X e Y respectivamente, P' e C' são matrizes $A \times r$ de pesos de X e Y respectivamente e F e G são as matrizes dos resíduos.

As componentes de X são boas preditoras de Y , ou seja,

$$Y = TC' + E \quad (4)$$

onde C' é obtido por mínimos quadrados, dado por

$$\hat{C}' = (T'T)^{-1} T'Y \quad (5)$$

Para conseguir os coeficientes da regressão por PLS referentes as variáveis originais, substitui-se (1) em (4) e obtém-se

$$Y = TC' + E = XWC' + E = XB + E \Rightarrow B = WC'$$

ou seja,

$$\hat{B} = W\hat{C}' \quad (6)$$

A j -ésima coluna da matriz $\hat{B}$ corresponde aos coeficientes estimados para o modelo de regressão por PLS da variável de resposta $\underline{y}_j$, $j = 1, \dots, m$.

Diferentes algoritmos podem ser utilizados na extração das componentes da regressão por PLS. O algoritmo mais popular é o Non-Iterative Partial Least Squares (NIPALS), desenvolvido por Wold em 1966. Outros algoritmos podem ser encontrados na literatura como o algoritmo SIMPLS, publicado por De Jong em 1993 ou o algoritmo Kernel descrito por Lindgren et al. em 1993 e Rannar et al. 1994 [8].

Na construção de qualquer modelo, o erro de estimação aumenta e o erro do modelo diminui à medida que a complexidade do modelo aumenta. O erro de predição é a soma dos dois tipos de

erro, assim, encontrar a complexidade ideal é uma troca entre os dois efeitos e o valor ótimo para fins de predição geralmente situa-se entre os dois extremos [9].

No caso da regressão por PLS, o número de componentes A , determina a complexidade do modelo. Com inúmeras e correlacionadas variáveis preditoras, existe risco de sobreajuste, ou seja, um modelo bem ajustado, porém com pouco, ou nenhum poder de predição. Portanto, a necessidade de verificar o poder de cada componente adicionada no modelo e parar o processo quando a inclusão de componentes for não significativa [7]. A validação cruzada (CV) é normalmente utilizada para determinar este número ótimo de componentes [8].

3.2. Validação Cruzada

Quando possível, é interessante dividir os dados em um conjunto para treinamento (calibragem) e um conjunto para teste (validação). O conjunto de treinamento é então utilizado para a seleção do modelo e com o conjunto para teste, estima-se o desempenho da previsão para novos dados (validação do modelo). O erro obtido a partir deste procedimento é uma medida mais realista, visto que foi validado por observações que não estavam na construção do modelo [4]. Dentre os testes do erro, destacam-se, o erro padrão de previsão (standard error of prediction - SEP), o erro médio quadrático de previsão (mean squared error of prediction - MSEP) e a raiz quadrada do MSEP (square root of the MSEP - RMSEP). Outra maneira de otimizar o desempenho de um modelo é utilizar estratégias de reamostragem como validação cruzada. É possível determinar o número de componentes PLS pelo procedimento a seguir. Divide-se o conjunto total de dados em k subconjuntos mutuamente exclusivos de mesmo tamanho e, a partir disto, um subconjunto é utilizado para teste e os $k-1$ restantes são utilizados para estimação dos parâmetros. Ao final das k iterações calcula-se a acurácia sobre os erros encontrados, através da equação

$$Ac_f = \frac{1}{v} \sum_{i=1}^v \varepsilon_{y_i, \hat{y}_i} = \frac{1}{v} \sum_{i=1}^v (y_i - \hat{y}_i) \quad (7)$$

onde v é o número de dados de validação e $\varepsilon_{y_i, \hat{y}_i}$ é o resíduo dado pela diferença entre o valor real da saída i e o valor predito.

O método leave-one-out de validação cruzada (LOO-CV) é um caso específico do descrito acima, conhecido como k -fold, onde k será igual ao número total de dados n . Nesta abordagem são realizados n cálculos de erro, um para cada dado. O modelo ótimo é o de menor erro de predição possível.

Neste artigo, inicialmente utiliza-se o procedimento (LOO-CV) onde os erros serão encontrados por RMSEP e posteriormente, compara-se com o resultado do RMSEP calculado com o conjunto de teste.

4. MATERIAIS E MÉTODOS

4.1. Usina Hidrelétrica de Itaipu

A Usina Hidrelétrica de Itaipu é uma empresa binacional localizada no Rio Paraná, na

fronteira entre Brasil e Paraguai (Figura 1). A barragem foi construída no período de 1975 a 1982 por ambos os países. Itaipu foi a maior produtora mundial de geração de energia por dois anos consecutivos, 2012 e 2013, porém em 2014, a Barragem das Três Gargantas, na China, produziu 98,8 milhões de megawatts por hora (MWh), contra 98,5 milhões de MWh de Itaipu.

A barragem de Itaipu possui estrutura singular composta por cinco tipos de represas. A de enrocamento, a de terra, a de concreto de gravidade aliviada, a de concreto de gravidade maciça e a de contraforte.

A Barragem Principal (trecho F) é composta de blocos de concreto do tipo gravidade aliviada com 16 tomadas de água para a geração de energia. Este é o trecho com maior influência da variação do nível do reservatório, por apresentar os blocos com maior altura, portanto foi escolhido para análise.

O trecho F possui 18 blocos duplos que vão do F1/2 ao F35/36 sendo quatro deles intensamente instrumentados por serem representativos segundo os engenheiros [10]. Estes blocos mais instrumentados são denominados blocos chave. Os blocos chave no trecho F são os blocos F5/6, F13/14, F19/20 e F35/36. Considera-se esses blocos para análise pela maior representatividade e maior número de instrumentos instalados.



Figura 1: Vista aérea da Usina Hidrelétrica de Itaipu em destaque o Trecho F. Fonte: [11].

4.2. Dados

Itaipu conta com cerca de 2.400 instrumentos (1.358 no concreto, 881 nas fundações e 161

para geodesia), sendo 270 automatizados, e 5.295 drenos (949 no concreto e 4.346 nas fundações) para acompanhar o desempenho das estruturas de concreto e fundações [12].

Os dados obtidos pela instrumentação permitem aos engenheiros analisar o comportamento da estrutura. Assim, observam se está funcionando como o esperado e, também, mantêm uma vigilância contínua da estrutura, detectando quaisquer evoluções atípicas que podem pôr em perigo a segurança da obra. Esses dados se compõem de medidas de deslocamento, tensão, deformação, pressão, cargas em elementos estruturais e de infiltração e drenagem, juntamente com os fatores ambientais que afetam o comportamento da barragem, tais como temperaturas, nível do reservatório e precipitação [2].

O deslocamento é um dos parâmetros mais significativos em monitoramento de segurança de barragens e esses são esperados durante e após a construção, portanto, os instrumentos pêndulo direto e base de alongâmetro foram considerados nesta análise. Esses instrumentos captam os movimentos relativos da estrutura. Os movimentos relativos da estrutura acontecem devido a fatores externos como variação do nível do reservatório e temperatura, logo as leituras do nível do reservatório e as leituras dos termômetros de superfície também são consideradas neste estudo.

As leituras manuais dos três instrumentos (pêndulo direto, base de alongâmetro e termômetros de superfície) e do nível do reservatório estão disponíveis desde 1980 até os dias atuais, porém considera-se neste trabalho o período compreendido entre 2000 e 2013. Pois a partir de 2000, as leituras realizadas pelos técnicos apresentam, em geral, periodicidade mensal e no início do período a frequência de leituras era maior. Com exceção do nível do reservatório que sempre teve leituras diárias. Outra razão é porque foram detectados poucos dados perdidos nesse período, não precisando de um método mais complexo de imputação. Para esses poucos casos, a média da leitura anterior e posterior foi utilizada.

Os dados das leituras são de 33 sensores dos instrumentos nos blocos chave do trecho F e as leituras do nível do reservatório, medida em metros. O Quadro 1 apresenta as 34 variáveis consideradas para o modelo. No caso do pêndulo direto, esse fornece medidas dos deslocamentos na direção X, no sentido do fluxo (direção montante-jusante) e na direção Y, perpendicular ao fluxo (margem direita-esquerda) medido em milímetros. As bases de alongâmetro fornecem medidas dos deslocamentos de abertura e deslizamento/assentamento, onde a unidade de medida é em milímetros. O presente trabalho limita-se a modelar os deslocamentos no sentido do fluxo (direção X) e deslocamentos de abertura de juntas. A unidade de medida para os termômetros de superfície é em graus Celsius.

4.3 Ajuste do modelo de regressão PLS

O número de observações (leituras manuais dos instrumentos com periodicidade mensal) é de 168 para cada variável, onde o período considerado é de 2000 a 2013. Primeiramente, as 168 observações foram divididas em dois conjuntos, um para ajuste do modelo (treinamento) composto da observação no tempo inicial até a 156ª observação e outro conjunto para validação do modelo (teste) composto pela observação 157ª observação até a 168ª observação. Para a

modelagem foi utilizado o software livre R [13].

Variáveis dependentes									
Pêndulo direto (direção X)					Base de alongâmetro (abertura)				
COF10X	COF11X	COF12X	COF13X	COF14X	JSF51AB	JSF6AB	JSF67AB		
COF18X	COF19X	COF2X	COF20X	COF21X	JSF68AB	JSF69AB			
COF22X	COF25X	COF26X	COF27X	COF28X					
COF3X	COF4X	COF5X							
Variáveis independentes									
Termômetro de superfície					Nível do reservatório				
TSF1	TSF11	TSF12	TSF13	TSF14	Nível				
TSF15	TSF16	TSF2	TSF21	TSF901					

Quadro 1: Variáveis consideradas para o modelo.

Para o primeiro ajuste, considera-se até 10 componentes PLS para investigar a variância explicada por cada componente e os valores de validação cruzada leave-one-out (LOO-CV). Apresenta-se abaixo as saídas do software R da validação cruzada para as 4 primeiras variáveis das 23 variáveis dependentes presentes no modelo.

```
Data:  X dimension: 156 11
      Y dimension: 156 23
Fit method: kernelpls
Number of components considered: 10

VALIDATION: RMSEP
Cross-validated using 156 leave-one-out segments.

Response: COF10X
(Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps
CV          0.8048 0.515 0.4765 0.4732 0.4711 0.4659 0.4474
adjCV       0.8048 0.515 0.4765 0.4731 0.4711 0.4656 0.4475
7 comps 8 comps 9 comps 10 comps
CV          0.4458 0.4312 0.4281 0.4354
adjCV       0.4462 0.4313 0.4280 0.4353

Response: COF11X
(Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps
CV          1.138 0.6766 0.6219 0.6031 0.5975 0.5860 0.5605
adjCV       1.138 0.6765 0.6219 0.6031 0.5974 0.5855 0.5606
7 comps 8 comps 9 comps 10 comps
CV          0.5596 0.551 0.5427 0.5506
adjCV       0.5597 0.551 0.5426 0.5504

Response: COF12X
(Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps
CV          1.277 0.6878 0.6182 0.6027 0.5919 0.5819 0.5509
adjCV       1.277 0.6877 0.6182 0.6026 0.5918 0.5814 0.5512
7 comps 8 comps 9 comps 10 comps
CV          0.5475 0.5392 0.5317 0.5420
adjCV       0.5477 0.5392 0.5316 0.5418

Response: COF13X
(Intercept) 1 comps 2 comps 3 comps 4 comps 5 comps 6 comps
CV          1.442 0.7872 0.7256 0.6920 0.6860 0.6800 0.6464
adjCV       1.442 0.7872 0.7255 0.6919 0.6859 0.6795 0.6470
7 comps 8 comps 9 comps 10 comps
CV          0.6410 0.6295 0.6245 0.6368
adjCV       0.6412 0.6295 0.6243 0.6366
```


A Figura 2 apresenta os gráficos do número de componentes considerado no modelo PLS versus RMSEP para as variáveis COF10X e COF11X. O gráfico foi obtido para todas as 23 variáveis de resposta, mas por motivo de espaço não serão apresentados. Geralmente seleciona-se o número de componentes após o qual o erro da validação cruzada não apresenta diminuição significativa, ou seja, se os resultados fossem dispostos graficamente, toma-se o primeiro mínimo local ao invés do mínimo absoluto da curva [8]. Nesta perspectiva, para as variáveis COF10X e COF11X, por exemplo, considera-se apenas 1 componente (Figura 2). No entanto, a análise gráfica para as 23 variáveis pode induzir a uma escolha de número de componente muito baixa, pois o valor do RMSE para a variável COF10X e COF11X diminui até a escolha de 9 componentes quando observa-se as saídas do modelo. Segundo [8], esta escolha do número de componentes é um tanto quanto subjetiva. Cabe então ao pesquisador investigar diferentes aspectos do ajuste de alguns modelos de regressão PLS de forma a auxiliar na decisão sobre o número de componentes do modelo.

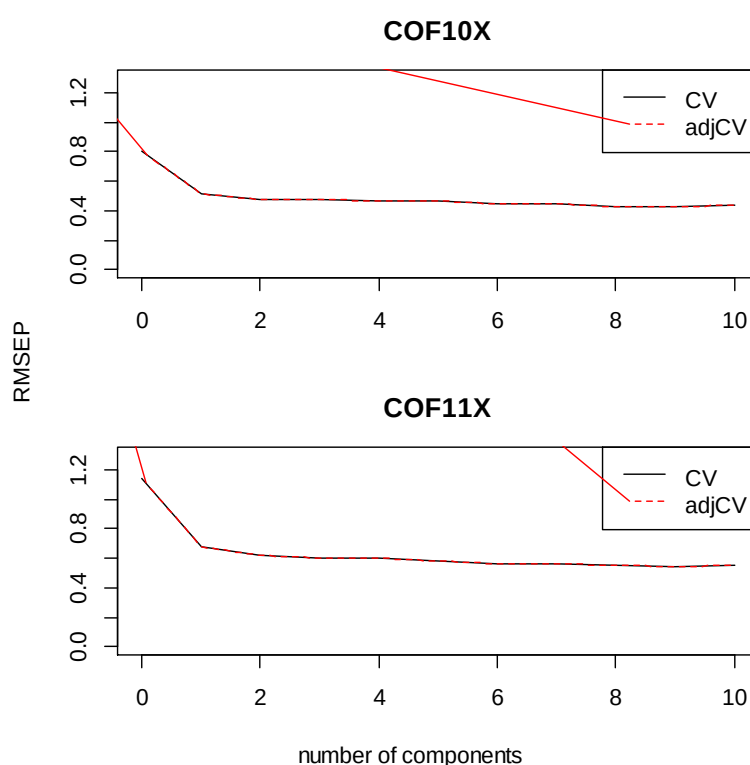


Figura 2: Gráficos número de componentes versus RMSEP para as variáveis dependentes COF10X e COF11X (pêndulos diretos).

Para auxiliar na escolha do número de componentes avalia-se o valor da quantidade de variância relativa de X (variáveis preditoras) explicada por cada componente. A Tabela 1 mostra a porcentagem de variância explicada por cada variável para a escolha de 1 componente até 10 componentes para o ajuste do modelo de regressão PLS.

Pelos valores das porcentagens de variâncias explicadas, um número de 4 componentes seria uma escolha adequada para o ajuste do modelo, pois quando compara-se a coluna referente a 4 componentes e 5 componentes na Tabela 1 nota-se que não houve um ganho significativo na porcentagem de variância explicada por 5 componentes. Isso indica que o ganho com o aumento do número de componentes é pequeno e assim opta-se por um modelo mais parcimonioso.

Treinamento: Porcentagem de variância explicada.										
	1comp	2comp	3comp	4comp	5comp	6comp	7comp	8comp	9comp	10comp
X	54,38	72,35	80,46	90,14	91,61	94,42	97,01	98,98	99,69	99,90
COF10X	59,79	66,46	67,56	68,15	71,99	72,07	73,04	75,07	75,70	75,81
COF11X	65,42	71,51	73,88	74,63	78,42	78,47	78,91	79,55	80,42	80,47
COF12X	71,65	77,79	79,46	80,47	83,42	83,43	84,07	84,61	85,23	85,39
COF13X	70,84	75,90	78,76	79,28	81,97	81,97	82,89	83,62	84,10	84,17
COF14X	62,97	66,35	70,80	71,15	73,19	73,20	74,95	76,77	77,29	77,45
COF18X	66,04	74,37	74,39	75,54	78,47	78,70	79,00	80,14	80,94	81,16
COF19X	74,51	81,73	82,16	83,16	85,64	85,70	85,78	85,97	86,87	87,01
COF2X	65,17	68,05	71,35	71,72	75,83	75,83	76,18	76,87	77,77	78,12
COF20X	74,24	81,09	81,89	82,82	84,98	85,03	85,38	85,78	86,60	86,75
COF21X	75,41	81,53	83,20	84,07	86,00	86,00	86,27	86,58	87,80	87,95
COF22X	67,95	72,66	76,04	76,51	78,40	78,40	79,28	80,21	81,00	81,11
COF25X	65,64	76,79	78,10	79,35	82,44	83,21	83,61	84,03	84,31	84,49
COF26X	78,97	85,13	87,64	88,93	90,15	91,57	91,98	92,04	92,08	92,12
COF27X	81,01	86,92	88,47	89,56	90,94	92,11	92,51	92,63	92,65	92,67
COF28X	81,97	86,82	87,97	88,81	90,09	91,03	91,51	91,57	91,64	91,81
COF3X	72,53	74,16	77,12	78,35	81,13	81,15	81,53	81,88	82,97	83,33
COF4X	74,64	77,00	79,42	80,22	82,39	82,39	83,09	83,34	84,04	84,38
COF5X	67,41	70,15	75,01	75,43	78,28	78,29	79,07	79,97	80,72	81,09
JSF51ab	61,02	64,16	67,98	68,71	68,81	68,83	68,84	71,72	73,06	73,07
JSF6ab	76,97	77,39	81,14	81,22	82,07	83,95	84,63	84,67	85,57	85,70
JSF67ab	70,65	70,68	70,74	70,85	70,85	75,90	76,10	76,17	76,25	76,31
JSF68ab	66,34	71,02	77,75	77,76	77,78	77,78	78,70	78,72	78,72	78,79
JSF69ab	61,62	64,25	71,14	71,15	71,17	71,17	71,85	71,87	72,02	72,47

Tabela 1: Porcentagem de variância explicada para o número de componentes do modelo.

Para o número de 4 componentes todos os valores de RMSE da validação cruzada (usando *leave-on-out*) apresentam valores estabilizados e a variância explicada de cada variável é maior do que 70%, com exceção das variáveis COF10X e JSF51ab em que a porcentagem ficou em aproximadamente 68%. Uma vez definido o número de componentes inspecionam-se diferentes aspectos do ajuste do modelo como gráficos de predições, escores e pesos.

A Figura 3 mostra os gráficos dos valores observados versus valores preditos somente para as

duas primeiras variáveis preditoras (COF10X e COF11X). O ideal seria que os pontos estivessem sobre a reta de meta, ou muito próximos dela. À esquerda os dois gráficos mostram que existe uma dispersão em torno da reta de meta, mas não existe indicativo de curvatura ou anomalias, no entanto, não é possível afirmar que uma variável teve seus valores de resposta melhor ajustados do que outra variável, pois as escalas do eixo y são diferentes. À direita estão os gráficos na mesma escala. Portanto, os gráficos de valores observados versus valores preditos são apenas para identificar a presença de curvaturas ou pontos isolados, para iniciar investigação. Para as outras variáveis de resposta destaca-se apenas o gráfico da variável JSF67ab que apresentou uma leve curvatura (Figura 4). Essa variável é uma base de alongâmetro de piso do bloco F18 que se encontra instalada na cota 214 m. Plotando as leituras dos deslocamentos de abertura das 5 bases de alongâmetro (Figura 5), a variável JSF67ab tem o mesmo comportamento periódico que as demais e destaca-se maior amplitude para as variáveis JSF6ab e JSF67ab localizadas na mesma cota 214 m, indicando que nesta cota as amplitudes de abertura são maiores para estes instrumentos considerados.

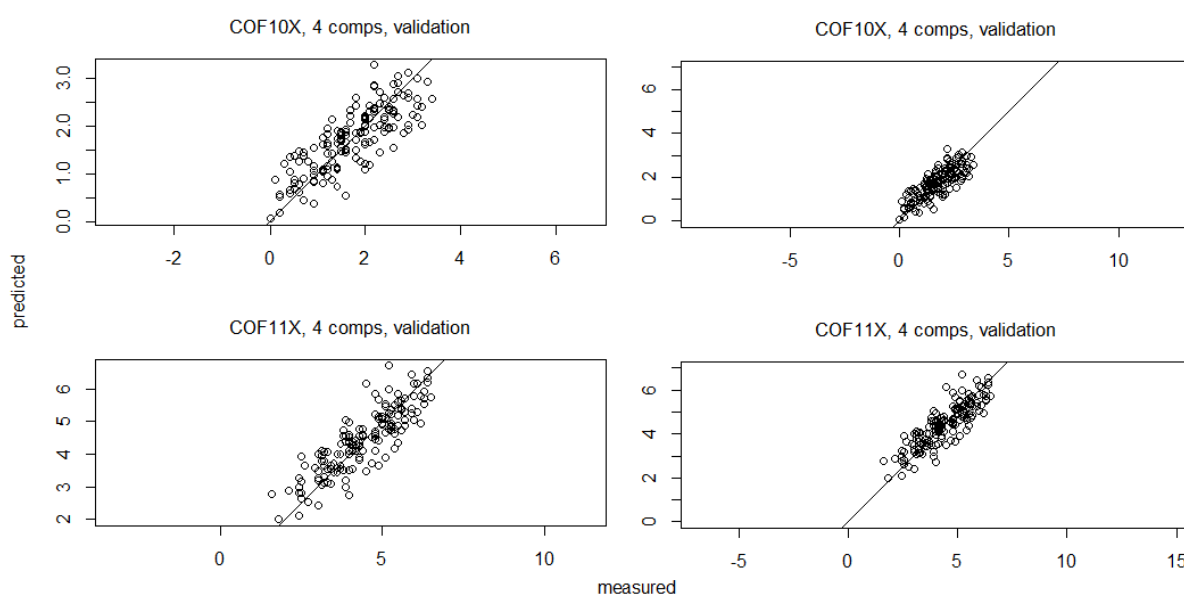


Figura 3: Gráfico dos valores observados versus valores preditos pelo modelo considerando 4 componentes para as variáveis COF10X e COF11X.

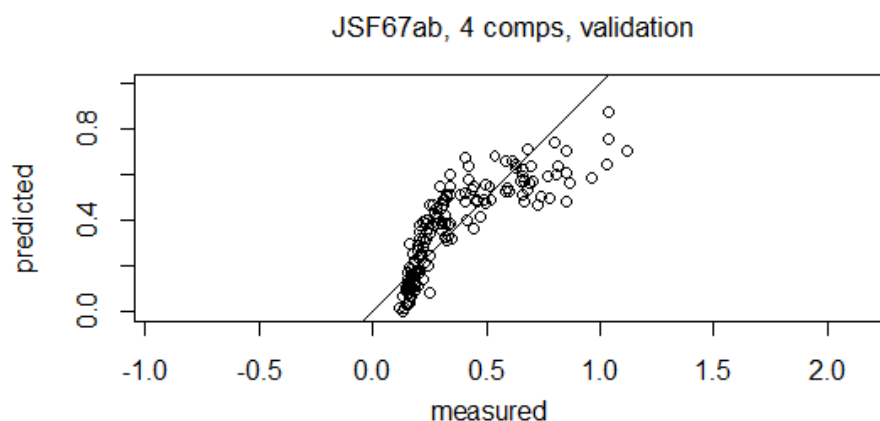


Figura 4: Gráfico dos valores observados versus valores preditos pelo modelo considerando 4 componentes para a variável JSF67ab.

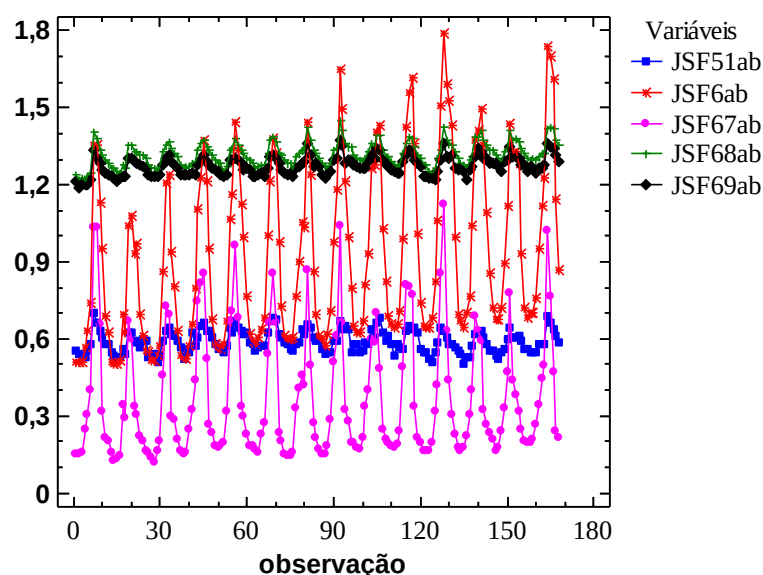


Figura 5: Leituras dos deslocamentos de abertura das bases de alongâmetro.

O gráfico de pares dos valores dos escores para as 4 componentes (Figura 6) é utilizado para procurar padrões, grupos ou outliers nos dados [8]. Na Figura 6 não há indicação clara de agrupamento ou de outliers. Os números entre parênteses são as quantidades relativas de variância de X (variáveis preditoras) explicada por cada componente.

O modelo ajustado é utilizado então para prever os valores de resposta de novas observações. A seguir, na Tabela 2, os valores previstos para as 23 variáveis de resposta (pêndulos diretos e bases de alongâmetro) utilizando o conjunto de teste das variáveis preditoras e considerando 4 componentes para o modelo de regressão PLS.

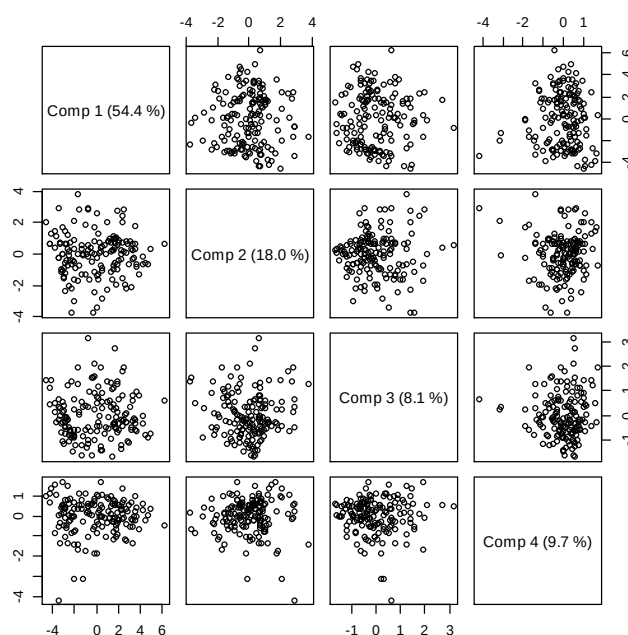


Figura 6: Gráfico dos escores.

n	COF10X	COF11X	COF12X	COF13X	COF14X	COF18X	COF19X	COF2X	COF20X	COF21X	COF22X	COF25X
157	0,273	2,257	4,676	7,617	11,511	1,426	2,956	1,372	5,049	7,378	11,976	1,797
158	0,563	2,686	5,172	8,094	12,026	1,755	3,404	1,616	5,566	7,911	12,536	2,189
159	0,494	2,557	5,044	7,929	11,784	1,722	3,330	1,511	5,463	7,772	12,325	2,215
160	1,511	4,126	6,844	9,846	14,073	2,764	4,887	2,636	7,303	9,758	14,607	3,303
161	1,701	4,266	7,075	10,003	14,033	3,118	5,290	2,671	7,700	10,045	14,721	4,003
162	2,196	5,066	7,970	11,066	15,413	3,551	6,020	3,357	8,592	11,060	16,006	4,350
163	2,493	5,559	8,520	11,740	16,312	3,795	6,455	3,807	9,131	11,687	16,826	4,511
164	2,650	5,817	8,824	12,133	16,844	3,929	6,712	4,094	9,444	12,050	17,297	4,611
165	2,586	5,766	8,765	12,121	16,916	3,822	6,614	4,140	9,350	11,992	17,299	4,406
166	2,541	5,730	8,709	12,021	16,808	3,762	6,517	4,058	9,246	11,898	17,196	4,287
167	1,542	4,247	6,979	10,190	14,695	2,683	4,937	2,989	7,408	9,967	15,044	3,037
168	0,843	3,170	5,735	8,840	13,070	1,969	3,852	2,175	6,126	8,581	13,441	2,291
n	COF26X	COF27X	COF28X	COF3X	COF4X	COF5X	JSF51ab	JSF6ab	JSF67ab	JSF68ab	JSF69ab	
157	2,573	4,809	8,909	3,402	5,003	8,078	0,570	0,487	0,103	1,278	1,240	
158	3,169	5,529	9,549	3,712	5,379	8,502	0,555	0,470	0,088	1,266	1,234	
159	3,303	5,657	9,669	3,594	5,244	8,291	0,553	0,445	0,098	1,262	1,230	
160	5,236	8,124	12,069	5,013	6,976	10,273	0,557	0,690	0,208	1,282	1,249	
161	6,837	9,979	13,883	5,002	7,084	10,197	0,554	0,664	0,308	1,267	1,238	
162	7,570	11,006	14,991	5,836	8,113	11,424	0,585	0,929	0,430	1,307	1,267	
163	7,959	11,579	15,637	6,389	8,790	12,235	0,609	1,127	0,520	1,338	1,291	
164	8,382	12,155	16,304	6,760	9,252	12,734	0,637	1,313	0,633	1,370	1,313	
165	8,141	11,900	16,121	6,865	9,346	12,838	0,656	1,420	0,680	1,393	1,329	
166	7,698	11,346	15,502	6,800	9,215	12,737	0,637	1,332	0,594	1,378	1,319	
167	5,255	8,293	12,493	5,477	7,543	10,913	0,626	1,074	0,423	1,357	1,300	
168	3,798	6,428	10,632	4,445	6,276	9,490	0,611	0,843	0,300	1,331	1,280	

Tabela 2: Valores previstos para as variáveis respostas utilizando o conjunto de teste.

Como os valores observados para estes valores previstos são conhecidos (variáveis de resposta

do conjunto de teste) calcula-se o RMSEP para comparar com o RMSEP do conjunto de treinamento. Considerando 4 componentes para o modelo de regressão PLS, a Tabela 3 mostra que os valores de RMSEP para o conjunto de teste são muito próximos as estimativas validadas anteriormente no RMSEP para o conjunto de treinamento.

$RMSEP_{train}$	$RMSEP_{test}$	$ RMSEP_{train} - RMSEP_{test} $
0,471	0,400	0,071
0,598	0,849	0,251
0,592	0,731	0,139
0,686	0,817	0,131
1,017	1,394	0,377
0,414	0,364	0,050
0,498	0,632	0,134
0,516	0,588	0,072
0,590	0,584	0,006
0,600	0,568	0,032
0,875	0,950	0,075
0,439	0,444	0,005
0,670	0,564	0,106
0,815	0,840	0,025
0,865	1,046	0,181
0,525	0,678	0,153
0,640	0,820	0,181
0,802	1,256	0,454
0,024	0,020	0,004
0,148	0,221	0,073
0,132	0,142	0,010
0,023	0,031	0,008
0,020	0,024	0,004

Tabela 3: Valores da validação para o conjunto de treinamento e para o conjunto de teste.

Segue abaixo os gráficos dos valores observados versus valores preditos para o conjunto de teste.

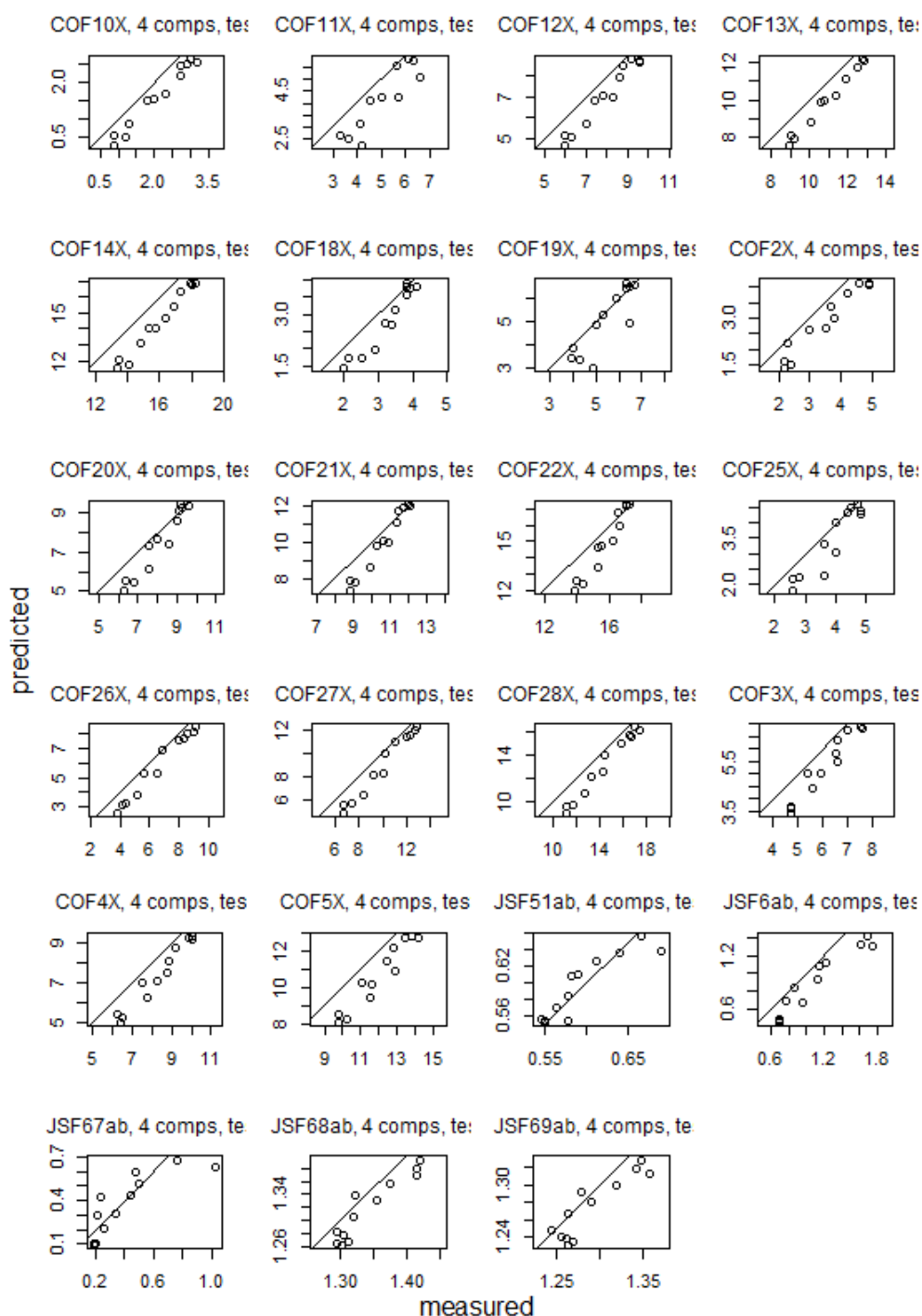


Figura 7: Valores observados versus valores preditos para o conjunto de teste.

5. CONCLUSÃO

O modelo proposto para as respostas de deslocamento da barragem de Itaipu pode ser usado como procedimento auxiliar no monitoramento da barragem.

Com o modelo, é possível prever as leituras dos pêndulos diretos e bases de alongâmetro conhecendo as variações do nível do reservatório e temperatura dos termômetros de superfície. Essas previsões quando comparadas às leituras reais fornecem informação se houve mudança de comportamento com relação ao comportamento anterior, considerado estável.

A regressão por mínimos quadrados parciais extrai do relacionamento entre as 23 variáveis dependentes e 11 variáveis independentes, apenas 4 componentes. Essas 4 componentes explicam aproximadamente 90% da variabilidade das variáveis preditoras e explicam mais que 70% de 21 das 23 variáveis de resposta. As variáveis COF10X e JSF51ab foram as que apresentaram menor porcentagem de variabilidade explicada pelas 4 componentes, nos valores de 68,2% e 68,7%, respectivamente. Isso mostra um bom potencial para o uso da regressão PLS no tratamento de dados de monitoramento de barragem reduzindo o número de variáveis a serem monitoradas.

O modelo é vantajoso no sentido de que na literatura, uma gama de modelos univariados é proposta para dados de monitoramento de barragem, enquanto que ao estimar conjuntamente os parâmetros, no caso de modelos multivariados, obtêm-se um ganho de eficiência dos estimadores e leva-se em conta o relacionamento entre todas as variáveis. Embora, os métodos estatísticos sejam frequentemente utilizados para modelar os dados em monitoramento de barragens, inevitavelmente certas correlações entre as variáveis impossibilitam a utilização da estatística tradicional. Portanto, investigações de técnicas que admitem correlações entre as variáveis são necessárias em aplicações desta área.

Em trabalho futuro, o intervalo de confiança será estabelecido para as previsões de modo a estabelecer limites de controle para as novas observações de deslocamentos.

AGRADECIMENTOS

Agradeço ao CEASB – Itaipu Binacional, pelo apoio financeiro e disponibilização dos dados utilizados neste estudo, em especial ao engenheiro Claudio Neumann Júnior pela disponibilidade e parceria.

Agradeço a Coordenação de Matemática da UTFPR-Toledo pelo apoio.

Agradeço ao PPGMNE – UFPR pela oportunidade desta pesquisa.

REFERÊNCIAS

- [1] N. DENG, J. WANG e A. C. SZOSTAK, *Dam deformation analysis using the partial least squares method*, 13th FIG International Symposium on Deformation Measurements and Analysis e 4th IAG Symp. on Geodesy for Geotechnical and Structural Engineering, Lisbon (2008).
- [2] B. A. NEDUSHAN, *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).
- [3] J. F. HAIR, W. C. BLACK, B. J. BABIN, R. E. ANDERSON e R. L. TATHAM, *Análise multivariada de dados*”, Bookman, (2009).
- [4] H. GARCIA e P. FILZMOSER, “Multivariate Statistical Analysis using the R package chemometrics”, Vienna, Austria, (2011) .
- [5] R. D. Tobias, *An introduction to partial least squares regression*. Proc. Ann. SAS Users Group Int. Conf., 20th, Orlando (1995).
- [6] S. A. Morellato, *Modelos de regressão PLS com erros heteroscedásticos*, Dissertação de Mestrado em Estatística, apresentada à Universidade Federal de São Carlos-UFSCar, São Carlos, (2010).
- [7] S. Wold, M. Sjöström, and L. Eriksson, “PLS-regression: a basic tool of chemometrics.” *Chemometrics and intelligent laboratory systems* Vol. **58**, Issue 2, pp. 109-130, (2001).
- [8] B. H. MEVIK e R. WEHRENS, “The pls package: principal component and partial least squares regression in R”, *Journal of Statistical Software*, Vol. **18**, Issue 2, pp. 1-24, (2007).
- [9] T. Naes, P. B. Brockhoff e O. Tomic, “Multiple Regression, Principal Components Regression and Partial Least Squares Regression”, *Statistics for Sensory and Consumer Science*, John Wiley & Sons, pp. 227-248, (2010).
- [10] I. BINACIONAL, *Itaipu: usina hidrelétrica–projeto: aspectos de engenharia*, TAB Marketing Editorial, Foz do Iguaçu, (2009).
- [11] M. F. BUZZI, A. S. DYMINSKI e A. CHAVES NETO, *Avaliação das correlações de séries temporais de leituras de instrumentos de monitoração geotécnico-estrutural e temperatura ambiente na barragem de ITAIPU – Caso do pêndulo direto*, XXVIII Congress Ibero Latino-Americano de Métodos Computacionais em Engenharia (CILAMCE), Porto, (2007).
- [12] ITAIPU BINACIONAL: *Barragem*. Disponível em <<http://www.itaipu.gov.br/energia/barragem>>. Acesso em 04 fev. 2015.
- [13] R CORE TEAM, “R: A language and environment for statistical computing”, *R Foundation for Statistical Computing*, Vienna, Austria. Disponível em: <http://www.R-project.org/>, (2014).