INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

MULTIVARIATE ANALYSIS OF THE DISPLACEMENTS OF A CONCRETE DAM WITH RESPECT TO THE ACTION OF ENVIRONMENTAL CONDITIONS

Sheila Regina Oro
Federal University of Technological - Paraná, Brazil
E-mail: sheilaro@utfpr.edu.br

Anselmo Chaves Neto
Federal University of Paraná, Brazil
E-mail: anselmo@ufpr.br

Tereza Rachel Mafioleti
Federal University of Technological - Paraná, Brazil
E-mail: mafioleti@utfpr.br

Suellen Ribeiro Pardo Garcia
Federal University of Technological - Paraná, Brazil
E-mail: suellenrp@utfpr.edu.br

Cláudio Neumann Júnior
Itaipu, Brazil
E-mail: neumann@itaipu.gov.br

Submission: 07/12/2015
Revision: 31/12/2015
Accept: 27/01/2016

ABSTRACT

A review of the concrete dam's structural performance is a complex issue comprised of many dimensions. This article proposes a method to assist in monitoring the displacements of structures and foundations of dams, considering the action of environmental conditions. Multivariate techniques are used to analyze the data pendulums, extensometer bases and multiple rods extensometer, along with environmental variables of the concrete surface temperature, ambient temperature and the reservoir water level. Specifically applies to Canonical Correlation Analysis to evaluate the influence of environmental variables in the displacement of structures and foundations. Factor Analysis identifies the factors inherent to the variability of the data.

This technique makes it possible to order the variables considering the action of factors. This applies also to Cluster Analysis on the data of dates of measurements, according to the similarities present in the observations. Then, Discriminant Analysis evaluates the formed groups for uniformity. The results demonstrate that the method can distinguish the dam responses and identify the effects of variations in environmental conditions over the displacements of structures and foundations.

Keywords: *structural monitoring; concrete dam; multivariate analysis*

1. INTRODUCTION

The concrete dam structures are subject to changes caused by the incidence of phenomena, such as displacements, strains, stresses, pressures, etc. (CARVALHO; ROMANEL, 2007). This occurs because these structures have strong interaction with environmental, hydraulic and geomechanical factors, as the temperature of the concrete, the hydrostatic pressure and the effect of time (LI; WANG; LIU, 2013). Therefore, these factors should be taken into account during structural evaluation.

Structural Monitoring is accomplished by visual inspection, geodetic measurement using vertical and/or horizontal displacements, bathymetric surveys and monitoring instrumentation (CRUZ, 2006). The instruments used in this monitoring include pendulums, extensometer bases, triple orthogonal meters, flow meters, piezometers, multiple rod extensometers (MATOS, 2002).

The instrumentation data set is useful for assessing the safety of dam's performance, especially if the current measures are compared with the entire series data through statistical and structural identification tools (DE SORTIS; PAOLIANI, 2007). Reports on instrumentation and visual inspections are useful in this case because they cover all aspects of the dam since its construction up to the operational phase (KUPERMAN et al., 2005).

Detailed analysis of instrumentation data requires a combination of knowledge, especially of Engineering, Mathematics and Statistics, and should be done by an experienced technical team, with the assistance of computational resources (VILLWOCK et al., 2013).

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

This work presents a helper method in the structural monitoring of concrete dams, combining multivariate statistical techniques to: (1) quantify the influence of environmental conditions on displacements of structures and foundations; (2) identification of the most relevant sensors with respect to the variability of the data; (3) grouping of the dates of measurements, according to the similarities.

The text structure is composed of six sections, with Introduction being the first one. The second section presents the review of the literature on the use of statistical techniques in the structural monitoring of concrete dams. Section 3 discusses the theory related to multivariate techniques of Canonical Correlation Analysis, Cluster Analysis, Discriminant Analysis and Factorial Analysis. The proposed method is described in the fourth section. Then, the results obtained in applying the method are conferred and discussed. The last section grants the conclusion of the article and suggestions for future research.

2. STRUCTURAL SECURITY

The interest in perfecting techniques for structural monitoring of dams has grown in recent decades mainly due to the need for greater security in order to avoid the consequences of disasters caused by structural problems (MEDEIROS; LOPES, 2011).

The methods used to monitor the structural safety of dams usually consist of comparing loads and safety factors used in the projects of dams with the behavior of all structures over the years (KUPERMAN et al., 2005).

The selection of data, during the process of monitoring the behavior of structures, involves the choice and type of method, number and location of the sensor, and hardware acquisition/storage/data transmission. This process is specific to each application. Economic issues play an important role in making these decisions. The time interval in which data should be collected is another point that should be considered (FARRAR; WORDEN, 2007).

In situations that have uncertainties inherent in the system adopted, the statistical analysis is indicated Farrar and Worden (2007) to sort an amendment of the parameters as from the structural condition change (failure) or modification of environmental and/or operating conditions.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

As the data may represent measures of different natures and scales, it is important to standardize the data to enable the damage identification process. Accordingly, Figueiredo et al. (2011) points out that standardization of data is an inherent procedure for data acquisition, feature extraction and statistical modeling in structural monitoring process, such as to remove the effects of operational and environmental variables with the extracted resources.

In addition, Farrar and Worden (2007) it is stressed the need to identify and minimize sources of variability in the data acquisition process and in the monitored system. However, not all sources of variability can be eliminated, for example, the variation caused by various environmental conditions such as temperature, humidity, loading and boundary conditions. Therefore, it is necessary to make appropriate measures so that these resources can be quantified statistically.

According to the literature, the statistical modeling of the structures monitoring data has been applied, mainly, to achieve classification, association, forecast values and outliers detection. Among multivariate statistical techniques, the most used for this purpose are discriminant analysis, canonical correlation, multiple linear regression and principal component analysis. Reports and discussions of these applications can be found at (BUZZI, 2007; CHENG; ZHENG, 2013; DENG; WANG; SZOSTAK-CHRZANOWSKI, 2008; DE SORTIS; PAOLIANI, 2007; FIGUEIREDO et al., 2011; GUEDES; DE FARIA, 2007; JIN-PING; YU-QUN, 2011; LI; WANG; LIU, 2013; MATA, 2011; MATA; TAVARES DE CASTRO; SÁ DA COSTA, 2013; MUJICA et al., 2014; XU; YUE; DENG, 2012).

3. THEORETICAL

A collection of *n* observations of *p* distinct random variables, taken from the same item, compose a multivariate sample, which can be represented in matrix form (Eq. 1), where each X_j vector containing the *n* observations of the variable *j*, for all $j = 1, 2, ..., p$.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

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

or

$$X = \left[\underline{X}_1 \quad \underline{X}_2 \cdots \quad \underline{X}_p\right]$$

Multivariate analysis provides methods and techniques for the theoretical interpretation of jointly sample. The main purposes that justify the use of multivariate analysis methods and techniques are: data reduction and structural simplification; sorting and grouping; investigation of dependence between variables; forecast; construction and hypothesis testing.

The following are the theoretical aspects of multivariate techniques called canonical correlation analysis, factor analysis, cluster analysis and discriminant analysis used in this study.

3.1. Canonical Correlation Analysis

The Canonical Correlation Analysis is an interdependence analysis technique that allows researchers to identify and quantify the associations between two groups of variables (*X* and *Y*). The basic idea is to find the linear combination of variables *X* and linear combination of variables *Y* that produce the highest correlation between the two groups (JOHNSON; WICHERN, 2007).

The first group (*X*) is composed of p decision variables, also called, explanatory variables (independents). While the second (*Y*) is formed by q response variables (dependents on explanatory).

The vectors *X* and *Y* have covariance matrices and, respectively, and the relationship is summarized in the cross-covariance matrix between these vectors, that is.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

Given that *U* and *V* are linear combinations of the vectors in *X* and *Y* (Eq. 2), respectively, the canonical problem is to obtain the vectors of coefficients a and b that maximize the correlation between U and V (Eq. 3). The vectors *a* and *b*, in this case, are solutions of a system of equations (Eq. 4). The linear combinations *U* and *V*, in this case, are called canonical variables.

$$U = a'X \qquad V = b'Y \tag{2}$$

$$Corr(U,V) = \sqrt{\lambda} = \frac{a'\Sigma_{XY}b}{\sqrt{a'\Sigma_X a}\sqrt{b'\Sigma_Y b}} \tag{3}$$

$$\begin{aligned} \left(\Sigma_{XY}\Sigma_Y^{-1}\Sigma_{YX} - \lambda\Sigma_X\right)a = 0 \\ \left(\Sigma_{YX}\Sigma_X^{-1}\Sigma_{XY} - \lambda\Sigma_X\right)b = 0 \end{aligned} \tag{4}$$

Where *λ* is the largest eigenvalue of the matrix or, equivalently, of the matrix $\left(\Sigma_{YY}^{-1}\Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY}\right)$.

Thus, each pair of canonical variables have unit variance, maximum correlation and is not correlated with others pairs of canonical variables. The number of pairs of canonical variables that can be obtained is equal to the lowest value of *p* and *q*. In general, we try to get a few pairs of canonical variables that explain much of the interdependence between the two sets of observable variables.

3.2. Factorial Analysis

The application of multivariate technique of Factorial Analysis allows the explanation of the correlations between many variables of a set of data through a limited number of unobservable random variables, called factors. (JOHNSON; WICHERN, 2007).

Verification of the viability of the factor model used is made by applying the Bartlett Test (Eq. 5), and the quality of fit of the model to the data set is estimated by the Criterion Kaiser-Meyer-Olkin (KMO). The KMO coefficient (Eq. 6) varies between 0 and 1. The closer to 1, the better the adjustment factor model to the data.

$$H_0 : \rho = I_{p\times p} \times H_A : \rho \neq I_{p\times p} \tag{5}$$

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

$$T = -\left[n - \frac{1}{6}(2p+11)\right]\sum_{j=1}^{p} \ln\lambda_j , comT \quad \chi_v^2, v = \frac{p}{2}(p-1)$$

$$KMO = \frac{\sum_{i\neq j} r_{ij}^2}{\sum_{i\neq j} r_{ij}^2 + \sum_{i\neq j} q_{ij}^2} \tag{6}$$

Where q_{ij} is the element belonging to the *i-th* row and *j-th* column of the matrix, $Q = DR^{-1}D$ with $D = \left(\sqrt{diagonal\left(R^{-1}\right)}\right)^{-1}$.

The factor model (Eq. 7) considers that each variable can be written as a linear combination of the common factors (F_k) and specific factors. During the process of obtaining the factors, are estimated the factor loadings (l_{ji}), the commonalities (h_i), the specific variances (ε_j) and the factorial scores (f_{jk}) that are measurements with explanatory properties of great interest to the researcher.

The load factor is a measure of the variable correlation with the factor. The commonality is the portion of the variance of each original variable from the extracted factors. The remainder of the variability, which can be owed to other factors, is measured by specific variance.

$$\begin{aligned}
Z_1 &= l_{11}F_1 + l_{12}F_2 + \ldots + l_{1m}F_m + \varepsilon_1 \\
Z_2 &= l_{21}F_1 + l_{22}F_2 + \ldots + l_{2m}F_m + \varepsilon_2 \\
&\vdots \\
Z_j &= l_{j1}F_1 + l_{j2}F_2 + \ldots + l_{jm}F_m + \varepsilon_j \\
&\vdots \\
Z_p &= l_{p1}F_1 + l_{p2}F_2 + \ldots + l_{pm}F_m + \varepsilon_p
\end{aligned} \tag{7}$$

The factor scores (Eq. 8) are estimates for the values of the factors for each sample element. They may be used to sort sample components or as input variables for further statistical analysis.

$$\underline{f}_i = \left(L'_z\, L_z\right)^{-1} L'_z\left(\underline{x}_i - \underline{x}\right), i = 1,2,\ldots,n \tag{8}$$

3.3. Cluster Analysis

The use of Cluster Analysis seeks to find, within a heterogeneous set of data, a small number of homogeneous groups, whose variation within the group is substantially smaller than the total variability of the data set.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

Initially, in the hierarchical agglomerative method, each observation forms a separate group. At each step of the process, the groups join according to the similarities, forming new groups, until remains only one group with the total number of observations included.

Similarity is a measure of proximity between two groups. One way of calculating this measure is the Mahalanobis distance (Eq. 9).

$$D_{ij}^2 = \left(x_i - x_j\right)' \Sigma^{-1} \left(x_i - x_j\right) \tag{9}$$

Where *Σ* is the complete data set of covariance matrix *X*.

3.4. Discriminant Analysis

Discriminant Analysis is a technique that enables, starting from independent variable, to study the profile, performing classification and differentiation of two or more group elements. The number of groups should be known in advance. The discrimination is made based on a mathematical rule, which minimizes the likelihood of incorrect classification errors.

In the perspective of Mahalanobis (Eq. 10), is calculated the distance (D_g^2) of each observation to the centroid of each group ($\bar{x}_g$). Then, the observation is allocated to the nearest centroid group.

$$D_g^2 = \left(x - \bar{x}_g\right)' \Sigma_W^{-1} \left(x - \bar{x}_g\right) \tag{10}$$

Where Σ_W is the covariance matrix within the group between the independent variables.

4. METHODOLOGY

The evaluation of the responses of a dam structure, considering the instrumentation data and taking into account interaction with the environment, is a problem composed of various dimensions. Therefore, it is necessary to use techniques that allow the joint analysis of monitoring data, reduce the magnitude of the problem and assist decision-making.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

The multivariate statistical techniques, called Canonical Correlation Analysis, Factorial Analysis, Cluster Analysis and Discriminant Analysis meet these requirements and, therefore, constitute the method used in this work.

As illustrated in Figure 1, the method consists of five steps. The first is the selection of instruments, the definition of the time period and data collection. If there is a difference in frequency of the measurement instruments, it is necessary to equate the periods, using for example, the monthly average of observations. Furthermore, in this step are identified and filled in the gaps coming from the absence of readings in the period. The filling is made with modeling and forecasting time series.

Then, the Box-Plot and Scatterplot graphics are used to identify the occurrence of outliers. For each detected outlier, it is need to evaluate its maintenance or exclusion from the data set. If you choose to exclude it, you must make a new value forecast, using the same procedure of filling the gaps.

Thus, it is composed a sample data matrix (X), with n lines (monthly average) and p columns (sensors of the instrument). If there are differences of magnitude and in the scale of the observations, owed to the use of different kinds of instruments, the data must be standardized. The standardized data matrix (Z) is, then, used as input to the procedures listed in the following steps.

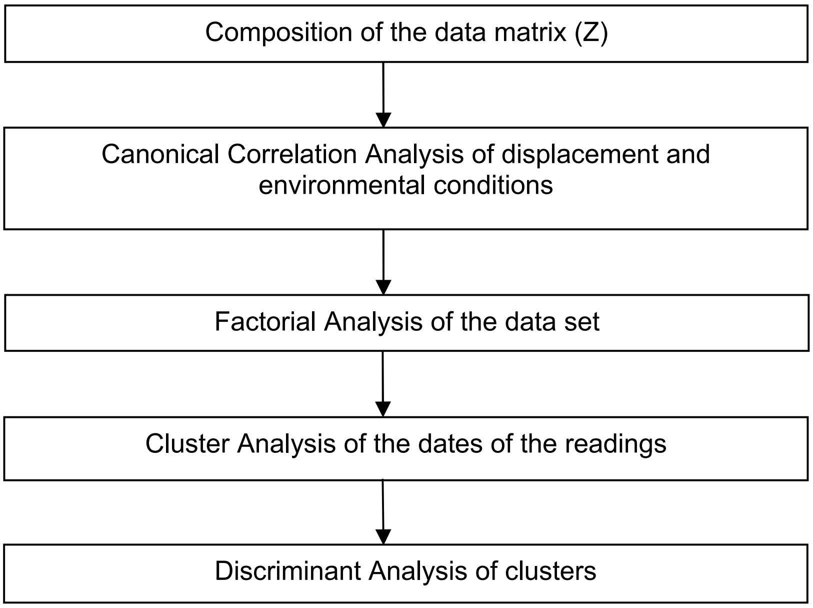

Figure 1: Flow chart of model

INDEPENDENT JOURNAL OF MANAGEMENT & PRODUCTION (IJM&P)
http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

The second step is the application of Canonical Correlation Analysis to study the relationship between the group of sensors, which measure the displacements of the dam structures and their foundations, and the group of indicators of environmental conditions.

Then, in step 3, the Factorial Analysis is used to estimate the influence of environmental conditions in shifts, perform the ranking of the instruments according to their importance in the factor model and identify factors that can be used as criteria for the overall assessment of displacements.

In the next stage, the scores of factors are subjected to Cluster Analysis for the formation of homogeneous groups of measurements dates.

In the last step, Discriminant Analysis tests the formed groups, with reference to the sensors with higher ranking in the Factorial Analysis.

4.1. Application in the Context of Structural Monitoring of Dams

The proposed method was applied to a structural monitoring process of a concrete dam. The data set used consists of the observations recorded in the period between January 1990 and December 2013, which were obtained through manual measurements of the installed instrumentation in D7 and D8 key blocks (Fig. 2), the D portion (Dam Right Side, built in blocks buttresses) of the Itaipu dam, and the hydrometeorological data from the same period.

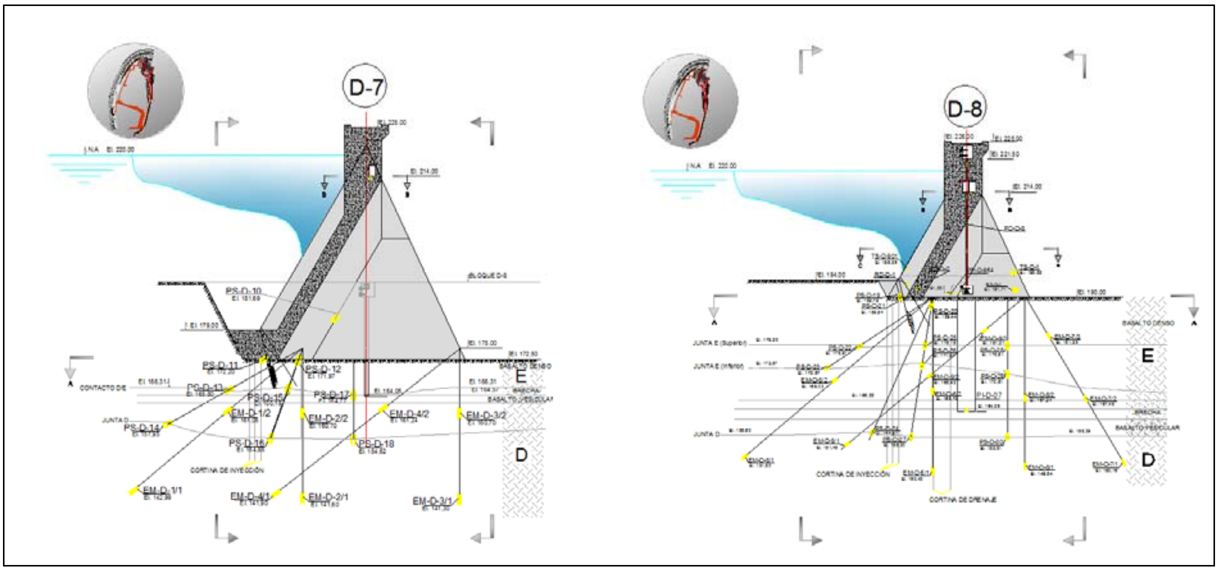

Figure 2: Layout of instrumentation installed in key blocks D7 and D8 of the Itaipu dam

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

Table 1 shows the sensors 42 considered in this study, corresponding to the direct pendulums, inverted pendulums, extensometer bases, multiple rods extensometers, thermometers and Limnimetric ruler.

Table 1: Phenomena monitored by instruments and respective sensors

Phenomena	Instrument	Sensor ID
Radial displacement	Direct pendulum	Z1, Z5
Radial displacement	Inverted pendulum	Z3
Tangential displacement	Direct pendulum	Z2, Z6
Tangential displacement	Inverted pendulum	Z4
Opening and closing of joints between blocks	Extensometer base	Z7, Z9, Z11, Z13, Z15, Z17
Horizontal sliding between blocks	Extensometer base	Z8, Z12, Z14, Z18
Differential settlement between blocks	Extensometer base	Z10, Z16
Surface temperature of block	Concrete thermometer – downstream	Z19
Surface temperature of block and water reservoir	Concrete thermometer – upstream	Z20
Deformations of rocky massive	Multiple rod extensometer	Z21, ..., Z40
Ambient temperature	Thermometer	Z41
Water level of the reservoir	Limnimetric Ruler	Z42

4.1.1. Composition of the Sample Data Matrix

Given that the frequency of measurements gauged with the different instruments (sensors) was not the same, it was decided to use the monthly average of the observations. The implementation of a computational procedure, performed with the help of Matlab software (Matlab R2013, 2013), allowed to create a monthly average observations of each of the 42 sensors and to identify the existence and location of gaps, corresponding to periods (months) that measurements were not performed.

It was found that there are eight incomplete series, totaling 15 deficiencies resulting from the lack of measurements at some point. There were created two (sub) series for each sensor that detects the occurrence of gaps: one with previous observations to the missing data and the other with subsequent information to the same issue. The forecast of this data was performed using the forecasting/backforecasting procedure, which consists in modeling each (sub) series, making value forecast and fill the gap with the average value of the two forecasts.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

The ARIMA models were used for the realization of the forecasts, with the help of Statgraphics software (Statgraphics Centurion XVI, 2010).

After completing the sensor data series, we proceeded the analysis of the Box-plot and Scatterplot graphics of each series, in search of the occurrence of outliers. Outliers were identified in the data reservoir water level (Fig. 2). The cause of this occurrence was the incidence of drought (low rainfall) in the months of the summer season of the years 1999/2000 and 2012/2013. It was decided to keep the values as observed, to check the influence of this occurrence in the dam's answers.

Owing to differences in quantities and measuring scale on the variables, because of the nature of the sensors, it was necessary to carry out the standardization of data.

Thus, it was made the sampling data matrix (*Z*), order 288×42, whose lines correspond to the dates (month/year), the columns to the variables (sensors) and the elements to standardized data.

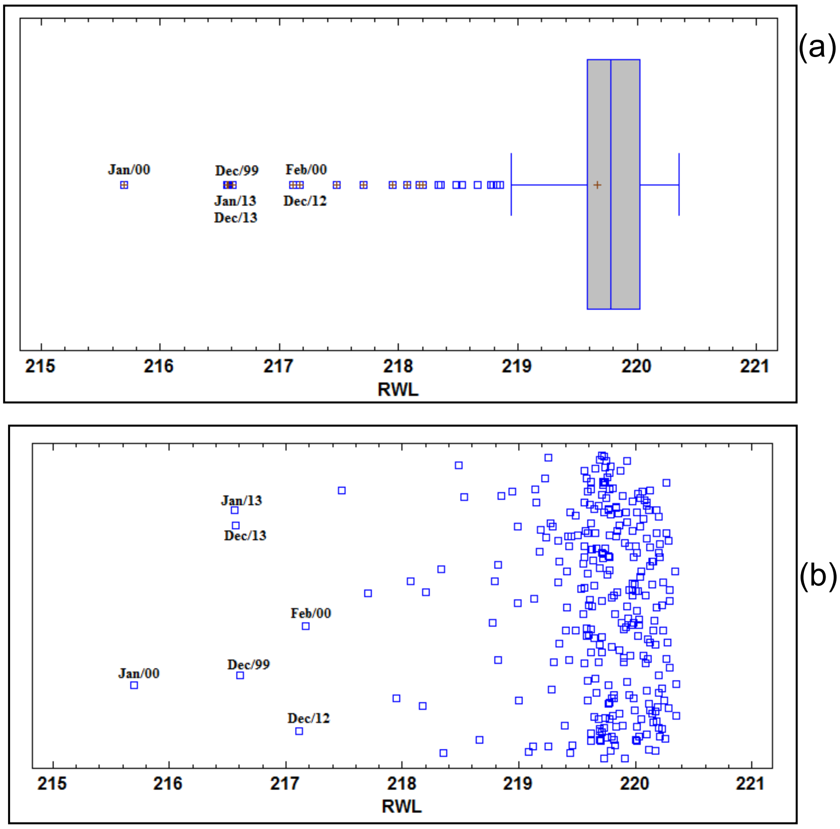

Figure 3: (a) Box-Plot and (b) Scatterplot graphic of the reservoir water level (RWL) from Itaipu dam, from Jan /90 to Dec /13

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

4.1.2. Canonical Correlation Analysis of displacements and Environmental Conditions

The Canonical Correlation Analysis was used to study the relationship between the sensors clusters that measure displacement (Z1 to Z18 and Z21 to Z40) and indicators of environmental conditions (Z18, Z19, Z41 and Z42). As shown in Table 2, all the eigenvalues were considered significant at a confidence level of 95% (p-value <0.05). It was decided to discuss the results of the canonical correlation of higher value.

Table 2: Canonical correlation of displacement sensors vs. environmental conditions

Eigenvalue	Canonical Correlation	Wilks'lambda	χ^2	Degree of Freedom	p-value
λ_1 = 0,973	0.986	0.003	1537.080	152	0
λ_2 = 0,717	0.847	0.114	577.122	111	0
λ_3 = 0,437	0.661	0.402	242.062	72	0
λ_4 = 0,286	0.535	0.714	89.492	35	0

The canonical correlation between the two groups was 0.986. This measure indicates the strong influence that the environmental conditions (V1) have on the set of sensors that measure the displacements of the dam's structures (U1).

The correlations between the 38 variables of the first group and four of the second were estimated. Table 3 lists the variables with the highest correlations ($|\rho| > |0.8|$ and p-value < 0.05), with predominance of multiple rod extensometers (Z23, Z24, Z35, Z38, Z40) related to temperatures (Z19, Z20 and Z41). Appear, also, three sensors pendulums (Z2, Z3 and Z5) associated to the block surface temperatures in the upstream (Z20) and environment (Z41).

Table 3: Strongly correlated variables

Sensors	Correlations
Z35 – Z20	-0.913
Z24 – Z20	-0.910
Z24 – Z41	-0.889
Z23 – Z20	-0.888
Z23 – Z41	-0.872
Z5 – Z41	-0.855
Z40 – Z19	-0.855
Z40 – Z41	-0.843
Z3 – Z41	-0.829
Z2 – Z20	-0.820
Z35 – Z41	-0.815

Sensors	Correlations
Z23 – Z19	-0.814
Z24 – Z19	-0.812
Z38 – Z19	-0.811

In the region where is located the Itaipu dam, the range of monthly average ambient temperatures, observed in the same year, may reach 20° C. The strong correlation of the sensors with temperature variables confirmed this. The negative correlation, with dominant presence in the first rows of the table, indicated an inverse relationship between the variables. That is, in periods of low temperature, the displacements were larger than those registered in periods of high temperature were.

Moreover, the reservoir water level alone showed a small positive correlation with just a few sensors. Possibly because the low variability in this variable led the forces acting on the dam almost constant. Another reason may be the need for the reservoir water level interaction with the temperature to influence the displacements.

Confronting the canonical variables U1 and V1, through the Scatterplot (Fig. 3), it was confirmed the existing linear relationship between these variables, showing the possibility to predict the dam's structural performance in a given time, depending on the measuring sensors offsets. It was also noted the distance of a point compared to the others. This point was referring to measurements made in July 2000, when it was recorded one of the monthly average lower to room temperature.

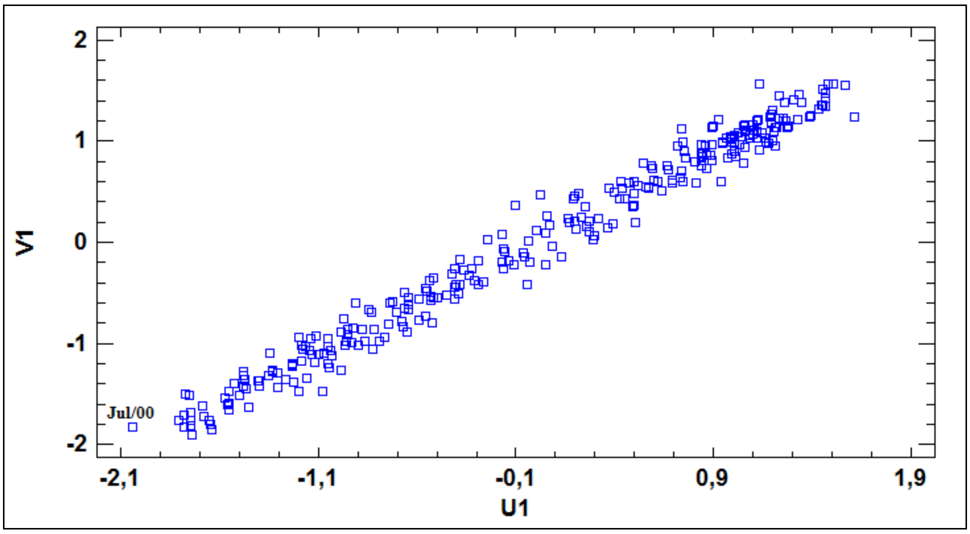

Figure 4: Scatterplot of the canonical variables U1 and V1

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

Table 4 presents the most correlated sensors (| ρ |> | 0.9 | e p-value < 0.05) with the first pair of canonical variables. It is observed that the rod of multiple rod extensometers Z23, Z24, and Z35 were the more influenced sensors by temperatures Z20, Z41.

Table 4: Key correlations between canonical variables and sensors of each group

Sensor	U1	Sensor	V1
Z23	-0.924	Z20	0.989
Z24	-0.943	Z41	0.936
Z35	-0.926		

Quality assessment of the potential of canonical variables was based on the proportion of variance explained by the canonical variables for each group. The canonical variable U1 explained 38.7% of the variance observed in shifts, while the proportion of the variance explained by V1 to the "Environmental Conditions" group was 65.6%. Thus, the groups "Displacement" and "Environmental Conditions" were well represented by the first pair of canonical variables, since the canonical correlation between these groups was 0.986, while the other pairs have lower values.

Therefore, if the "Environmental Conditions" group was the cause of the variability observed in the group "Displacement", then U1 can be used as the best predictor and V1 the most likely criterion for the realization of the dam's structural performance prediction, in what concerns to offsets.

4.1.3. Factorial Analysis of displacements and Environmental Conditions

The application of Factorial Analysis by principal components resulted in a model composed of five factors, identified based on the greatest factor loadings of sensors, able to explain 91.12% of the variance of the set of comments.

The factors were named according to the sensors more correlated with them, that is, as the greatest factor loadings. The first factor, due to its positive correlation with most of the stems of multiple rod extensometers, especially with Z21, Z22, Z25, Z26, Z29, Z30, corresponds to the "Foundation’s Movement". This was the most important among factors identified because accounts for 45.88% of the observed variability in the data set.

The second factor, named "Horizontal Movement Structure in Normal Direction", explained 30.83% of the variance and is associated with the openings and joints between the dam block and the horizontal displacement in the normal direction

ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

(perpendicular to the direction of water flow). The "Horizontal Movement Structure in the Flow Direction", third factor identified, accounted for 9.66% of the variability. The dominant variables in this factor were the temperatures (Z19, Z20, Z41), with the multiple rod extensometers (Z23, Z24, Z38, Z40) and the horizontal displacements in the direction of water flow of D8 block (Z3 and Z5).

The fourth and fifth factors, called "Block D7 Structure Geometry" and "Hydrostatic Pressure" were related to horizontal displacements in the direction of water flow of D7 block (Z1) and the reservoir water level (Z42), respectively. Furthermore, the "Block D7 Structure Geometry" accounted for 2.41% of the total variability, while the influence of "Hydrostatic Pressure" was 2.34%.

The variability in the readings of each sensor, arising from identified factors, were estimated by commonality. Thus, a commonality low (less than 0.60) would indicate that the variable would not be sufficiently explained by the model and could be discarded. The results pointed to the preservation of all sensors considered in this study.

Using as a measure of commonality as a measure of importance of each variable to the factorial model, there was obtained the ranking of the sensors. Therefore, the most important instruments for the D7 and D8 blocks were respectively rods extensometer multiple Z33, Z34, Z35 (even borehole) and extensometer bases (opening) Z7, Z9. Furthermore, the reservoir water level (Z42) was the variable related to the environmental conditions highest classified.

The factor, being a latent variable, cannot be measured directly. However, the values of factors, called factor scores, are estimated based on factor loadings and the sensor values that dominate this factor. Obtained the factor scores for each of the 288 dates of the readings, with respect to each factor, took place the Cluster and the Discriminant analysis.

4.1.4. Grouping the dates of the measurements

Using the method of the group average and the Mahalanobis distance, were identified three homogeneous groups of dates (Fig. 4), adding 190 elements in the first (G1), 83 in the second (G2) and 15 in the third (G3). The first group comprised essentially of the months from November to May, period in which were recorded the highest temperatures, while most months, with lower temperatures, were gathered in

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

G2. The third group brought together the dates in which were recorded the lowest water levels of the reservoir.

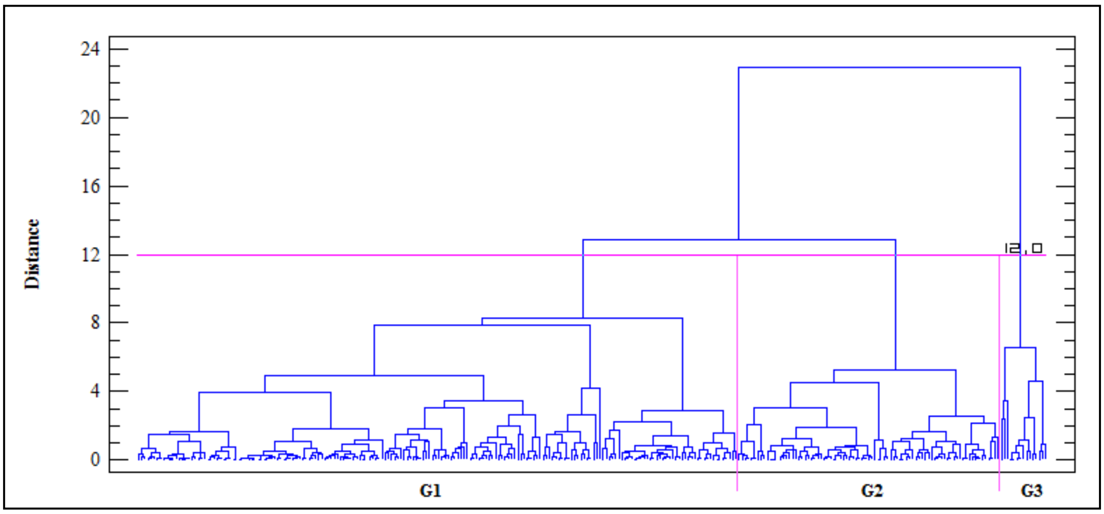

Figure 5: Group by the middle connection method, using the Mahalanobis distance

Discriminant Analysis, considering the sensors Z7, Z9, Z33, Z34, Z35 and Z42, chosen because they have the greatest commonalities, tested the classification of 288 observations. Due to the large difference in size of the groups, we considered the proportionality of the number of observations per group. The results are in Table 5. The high percentage of correct classification, 94.1%, has confirmed the discriminating power of the sensors regarded in the analysis.

Table 5. Classification of the dates of the measurements into three groups

Current Group	Size of the Group	Proportion	Provided Group			Cumulative percentage
			1	2	3	
1	190	65.97%	178	11	1	93.68%
2	83	28.82%	5	78	0	93.98%
3	15	5.21%	0	0	15	100%
TOTAL	288	100%	183	89	16	94.10%

Two functions were considered statistically significant, at the confidence level of 99%, to distinguish the observations belonging to each group (Fig. 5). The first function discriminates High Temperature and Low Temperature groups, while the second group discriminates the Low Water Level of Reservoir (LWLR) group from the others.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

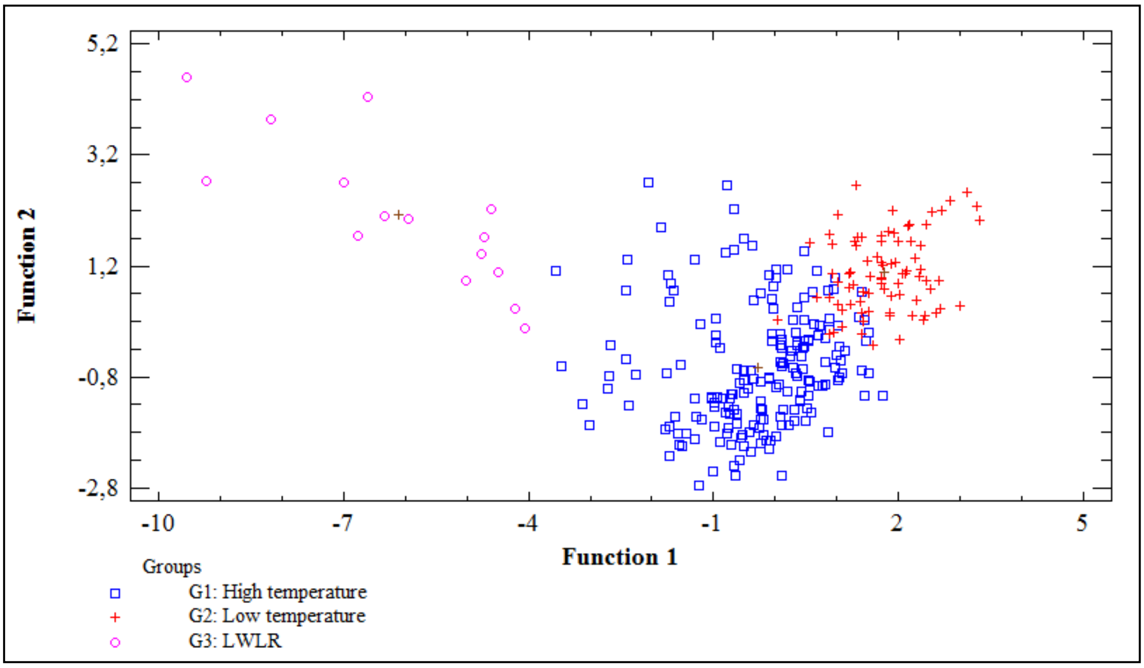

Figure 6: Dispersion of the elements according to discriminant functions between groups.

Ranking functions of observations in groups (Eq. 10) are linear combinations of the sensors. These functions can be used to classify new dates readings. To do so, simply calculate the scores of each new element in each group and, then, allocate it in that with highest score.

$$
\begin{aligned}
G1 &= -0{,}664 + 0{,}664{*}Z53 - 1{,}332{*}Z52 - 0{,}075{*}Z54 + 0{,}323{*}Z9 + 0{,}134{*}Z61 - 0{,}881{*}Z7 \\
G2 &= -3{,}398 - 0{,}446{*}Z53 + 2{,}866{*}Z52 + 0{,}073{*}Z54 - 1{,}284{*}Z9 + 1{,}656{*}Z61 + 2{,}674{*}Z7 \\
G3 &= -23{,}805 + 5{,}944{*}Z53 + 1{,}009{*}Z52 + 0{,}544{*}Z54 + 3{,}013{*}Z9 - 10{,}866{*}Z61 - 3{,}636{*}Z7
\end{aligned}
\quad (1)
$$

5. CONCLUSION

The method proposed in this paper consists in the multivariate analysis of the displacements of the structures and foundations of a concrete dam, taking into account the interaction with the environment.

The results of Canonical Correlation Analysis allow us to infer that these shifts are strongly influenced by environmental conditions. In general, the instrumentation registers larger dislocations during periods of low temperatures. The set of instruments that comprises pendulums, extensometer bases and multiple rods extensometer can be used to predict the structural performance of a dam, with respect to displacement, according to the variability criteria of environmental conditions.

The dates of the observations recorded by instrumentation, when subjected to Cluster and Discriminant analyses, can be grouped into "High Temperature", "Low Temperature" and “Lower Reservoir Water Level”.

Most of the measurement data variability is due to factors: Foundation’s Movement; Horizontal Movement Structure in Normal Direction; Horizontal Movement Structure in the Flow Direction; Block D7 Structure Geometry and Hydrostatic Pressure.

ACKNOWLEDGMENTS

The authors thank for the contributions of the following institutions:

- Ceasb - Itaipu Binacional: technical support and provision of data;

- PROPPG - UTFPR: feasibility of this study by Ordinance no. 0398/2014;

- PPGMNE - UFPR: theoretical support and encourage the development of this work.

REFERENCES

BUZZI, M. F. (2007) **Avaliação das correlações de séries temporais de leituras de instrumentos de monitoração geotécnico-estrutural e variáveis ambientais em barragens - estudo de caso de Itaipu.** Dissertação—Métodos Numéricos em Engenharia: UFPR.

CARVALHO, J. V.; ROMANEL, C. (2007) Redes neurais temporais aplicadas ao monitoramento de barragens. **Revista Eletrônica de Sistemas de Informação**, n. 10.

CHENG, L.; ZHENG, D. (2013) Two online dam safety monitoring models based on the process of extracting environmental effect. **Advances in Engineering Software**, v. 57, p. 48–56, mar.

CRUZ, P. T. (2006) **100 barragens brasileiras: casos históricos, materiais de construção, projeto.** São Paulo: Oficina dos Textos.

DENG, N.; WANG, J.-G.; SZOSTAK-CHRZANOWSKI, A. (2008) **Dam Deformation Analysis Using the Partial Least Squares Method.** Proceedings in 13th FIG Int. Symp. on Deformation Measurements and Analysis & 4th IAG Symp. on Geodesy for Geotechnical and Structural Engineering. Lisbon.

DE SORTIS, A.; PAOLIANI, P. (2007) Statistical analysis and structural identification in concrete dam monitoring. **Engineering Structures**, v. 29, p. 110–120.

FARRAR, C. R.; WORDEN, K. (2007) An introduction to structural health monitoring. **Philosophical Transactions of the Royal Society A**, v. 365, p. 303–315.

FIGUEIREDO, E.; PARK, G.; FARRAR, C. R.; WORDEN, K.; FIGUEIRAS, J. (2011) Machine learning algorithms for damage detection under operational and environmental variability. **Structural Health Monitoring**, v. 10, n. 6, p. 559 – 572, nov.

http://www.ijmp.jor.br v. 7, n. 2, April - June 2016
ISSN: 2236-269X
DOI: 10.14807/ijmp.v7i2.413

GUEDES, Q. M.; DE FARIA, É. F. (2007) **Modelo estatístico de controle do deslocamento monitorado na barragem casca da UHE Funil.** Proceedings in XXVII SEMINÁRIO NACIONAL DE GRANDES BARRAGENS. Belém, PA, jun.

JIN-PING, H.; YU-QUN, S. (2011) Study on TMTD Statistical Model of Arch Dam Deformation Monitoring. **Procedia Engineering**, v. 15, p. 2139 – 2144.

JOHNSON, R. A.; WICHERN, D. W. (2007) **Applied Multivariate Statistical Analysis**. 6. ed. Pearson.

KUPERMAN, S. C.; MORETTI, M. R.; CIFU, S; CELESTINO, T. B.; RE, G.; ZOELLNER, K. (2205) **Criteria to establish limit values of instrumentation readings for old embankment and concrete dams**.

LI, F.; WANG, Z. Z.; LIU, G. (2013) Towards an error correction model for dam monitoring data analysis based on cointegration theory. **Structural Safety**, v. 43, p. 12–20.

MATA, J. (2011) Interpretation of concrete dam behavior with artificial neural network and multiple linear regression models. **Engineering Structures**, v. 33, p. 903–910.

MATA, J.; TAVARES DE CASTRO, A.; SÁ DA COSTA, J. (2013) Time–frequency analysis for concrete dam safety control: Correlation between the daily variation of structural response and air temperature. **Engineering Structures**, v. 48, p. 658 – 665.

Matlab R2013 (2013). Math Works.

MATOS, S. F. (2002) **Avaliação de instrumentos para auscultação de barragem de concreto. Estudo de caso: deformímetros e tensômetros para concreto na Barragem de Itaipu.** Dissertação—Curitiba, PR: UFPR.

MEDEIROS, C. H.; LOPES, M. G. M. (2011) **O Risco da Classificação de Barragens por Categoria de risco, com Base em Método de Ponderação de Fatores.** Sessão Técnica apresentado em XXVIII Seminário Nacional de Grandes Barragens. Rio de Janeiro.

MUJICA, L. E.; RUIZ, M.; POZO, F.; RODELLAR, J. (2014) A structural damage detection indicator based on principal component analysis and statistical hypothesis testing. **Smart Materials and Structures**, v. 23, n. 2, p. 25014 – 25025, Fev.

Statgraphics Centurion XVI (2010). StatPoint Technologies.

VILLWOCK, R.; STEINER, M. T. A.; DYMINSKI, A. S.; CHAVES NETO, A. (2013) **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. [s.l.] InTech, 2013. p. 81–102.

XU, C.; YUE, D.; DENG, C. (2012) Hybrid GA/SIMPLS as alternative regression model in dam deformation analysis. **Engineering Applications of Artificial Intelligence**, v. 25, p. 468–475, abr.