

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

6,900

Open access books available

185,000

International authors and editors

200M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



---

# **Itaipu Hydroelectric Power Plant Structural Geotechnical Instrumentation Temporal Data Under the Application of Multivariate Analysis – Grouping and Ranking Techniques**

---

Rosangela Villwock, Maria Teresinha Arns Steiner,  
Andrea Sell Dyminski and Anselmo Chaves Neto

Additional information is available at the end of the chapter

<http://dx.doi.org/10.5772/53838>

---

## **1. Introduction**

The monitoring of a dam structure can generate an enormous mass of data of which the analysis and interpretation are not always trivial. It is important to select the information that better “explain” the behavior of the dam, making possible the prediction and resolution of eventual problems that may occur.

The world largest hydroelectricity generator, Itaipu hydroelectric power plant, has more than 2.200 instruments that monitor its geotechnical and structural behavior, and these instruments have readings stored on a database for over 30 years. The high dimensionality and the large quantity of records stored on the databases are nontrivial problems that are kept so that one can pursue "knowledge" through these data.

The detailed analysis of the auscultation instrumental data requires a combination of knowledge of Engineering, Mathematics and Statistics, as well as the previous experience of the engineer or the technician responsible for the analysis of these data. That consumes a lot of time, and often makes it impossible to accomplish this task in an efficient way. This is the reason why the use of techniques and computational instrumentation to help the decisions maker is extremely important.

There are no records of the existence of methods that perform the classification of monitoring instruments in dams. In case of reading intensification this hierarchy could be useful to define which instrument to chose.

The aim of this paper is to identify the tools that are the most significant for the analysis of a dam behavior, which maximizes the effectiveness and efficiency of the analysis of the readings. It shows a methodology based on the field of Multivariate Analysis, applied to the Hierarchical Cluster Analysis in order to identify the groups of instruments similar to Ward's linkage method. The factor analysis of the strain gauge of each instrument group was also applied, performing the hierarchical cluster of monitoring instruments in dams, detecting the main instruments.

This chapter is organized as follows: Section 2 features the problem statement which addresses the importance of the safety on dams and the risks faced when dam rupture accident occurs. Section 3 describes the application area focusing on the safety of dams, on the conditions of load and on the conditions of the monitoring instrument. Section 4 approaches a "research course". Section 5 describes the used data and the Multivariate Statistical Analysis techniques. Section 6 shows the status. Section 7 shows the results. Section 8 approaches the future researches. Section 9 shows the results.

## 2. Problem statement

Once the potential risks and losses as a result of rupture accidents on a dam can reach large scales a safe project and adequate construction as well as a correct operation on dams are concerns of Brazilian and worldwide engineers. Additionally, an effectively done monitoring on large dams is essential for the safety of its structure. By aiming the safety of the dams, International Guidelines and many helpful discussions about this subject have been proposed and conducted, such as the one from the [1].

In Brazil, guidelines that aim the safety of the dams were published by the *Comitê Brasileiro de Grandes Barragens* (Brazilian Committee on Large Dams), see [2]. The *Comissão de Constituição e Justiça e Cidadania do Congresso Nacional* (The Constitutional, Justice and Citizenship Committee of the National Congress) approved, on 06/23/2009, the proposal that requires the Executive Power to establish a National Policy on Dams Safety. Its aim was to endow the Public Power with a permanent instrument for the inspection of over 300 thousand dams in the country. The text that has been questioned is the surrogate for the Law Project 1181/03 [3]. The original proposal, the Law Project 1181/03, (*Projeto de Lei – PL 1181/03*), written by Leonardo Monteiro, defines safety guidelines for the construction of dams and landfills of industrial liquid wastes.

The concerns about the Brazilian Constitutional Public Powers is due to the recent rupture of the Dam of Câmara, (in the State of Pará - PA), in 2004; the rupture of the diversion structure of the Dam of Campos Novos, (in the State of Santa Catarina - SC), in 2006; the rupture of the Dam of Algodões I, (in the State of Piauí - PI), in 2009; and other accidents of smaller magnitude.

According to [4], the catastrophes have been opportune signs for the examination of the criteria of the existing projects and for the selection of more efficient methods and monitoring safety of dams.

In [5] show a table which contains the estimative for the most common causes of ruptures on dams. Among them, the following are highlighted: problems of the foundation; inappropriate spillway; structural problem; different declinations; extreme low-pressure; rupture of landfills; defective materials; incorrect operation; actions of war, and earthquakes. All these problems can be diagnosed with the monitoring of the dam instrumentation, with exception of the last two ones, which percentage of frequencies sum just 4%.

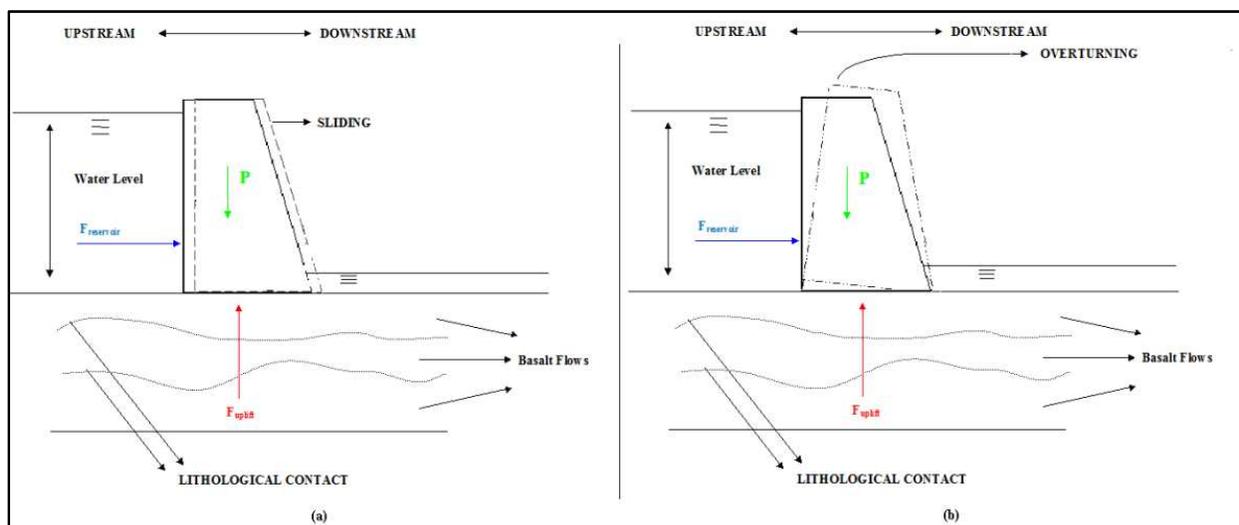
According to [6], the global experience shows that the expenses in order to guarantee the safety of a dam are little when compared to the costs of its rupture. The author quotes the importance of the use of a database of instrumentation for supporting the preliminary analysis of the readings in order to detect problems.

### 3. Application area

#### 3.1. Safety of dams

The principles established on NBR 8681 – *Ações e Segurança das Estruturas* (Actions and safety of structures) [7] conceptualizes the safety of the concrete constructions of a dam. For concrete gravity-dam projects, some verification corresponding to the stability analysis are necessary in order to evaluate the safety of the movements of: sliding, overturning, floating, tension at base and on structure, deformations, consolidation and vibrations.

The stability of dams must be primary analyzed during the phase of the project. The geometry of the structures and the property of the materials involved must be well considered, as well as the load condition. Some of the basic load conditions are shown on Figure 1.



**Figure 1.** Illustration of the basic conditions of load and lack of stability of gravity concrete dams.

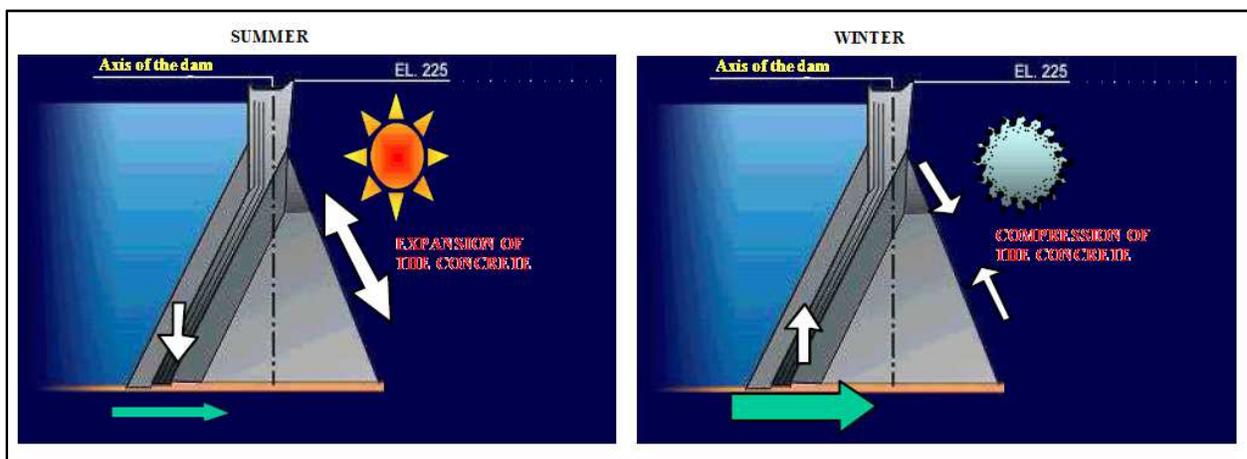
Through Physics, it is possible to explain that the difference of the water level (downstream-upstream) generates a hydraulic gradient between the dam downstream and upstream

making the water of the reservoir to try passing through upstream in order to archive a hydraulic equilibrium. To do so, the water percolates through the foundation mass of the dam. During this process, the infiltrated water generates vertical forces acting upward over the dam, these forces are called uplift pressure in dam. The resultant of these forces is represented by  $F_{\text{uplift}}$ . Furthermore, the water from the reservoir generates horizontal forces that act downstream-upstream over the dam. These forces are called hydrostatic pressures against the dam wall. The resultants of these forces are represented by  $F_{\text{reservoir}}$ . These two resultant forces are called destabilizing forces. As for the force  $P$  (dam weight) it is a stabilizing structure force. The combination of  $F_{\text{uplift}}$  and  $F_{\text{reservoir}}$  can cause the overturning and the slipping of the dam, not just because of the efforts and moment when it is directly applied, but also for the relief of the weight of the structure itself (in case of uplift pressure).

The above described effects of loads on dams can be observed on figure 1, where the slipping (a) and the overturning (b) are emphasized.

The loading conditions and the properties of materials can change over the lifecycle of a dam, and instrumentation can identify some of these changes.

Figure 2 shows the differences in the behavior of the dam in relation to summer and winter climate conditions, as well as its consequences. In summer, an expansion of the concrete occurs, and that causes the block to tumble downstream. This overturning causes the block to compress the foundation. In winter, the concrete is compressed causing the block to tumble upstream, returning to initial position. As a consequence the pressure that occurs in summer over the foundation to be relieved. In this way, it is possible to identify a cyclical behavior of the structure, intrinsically conditioned by the environmental conditions which involve the construction.

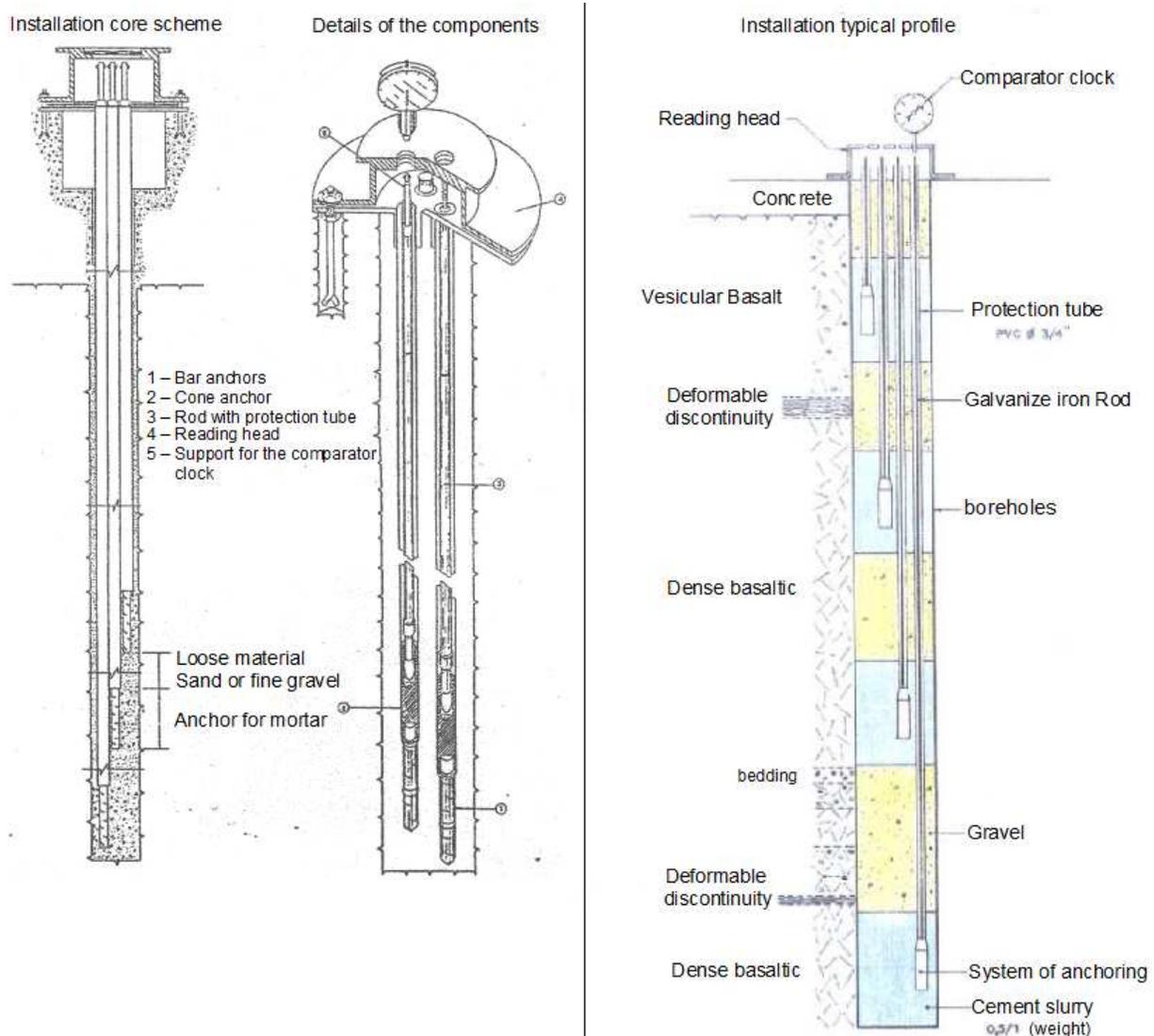


**Figure 2.** Behavior of the dam in relation to summer and winter climate conditions (Adapted from [8]).

According to [9], the instrumentation must be used as supplement to visual inspection when executing the evaluation of the performance and safety of dams. The careful inspection of the instrumentation data can reveal a critical condition.

In [10] shows correlations between the types of instruments that are usually used for the auscultation on concrete dams, and the primary types of deterioration of concrete dams. According to the author, the multiple extensometer for example, is related to the monitoring of deteriorations caused by sliding, different declinations, land subsidence of the upstream base, and the Alkali-Aggregate Reactivity.

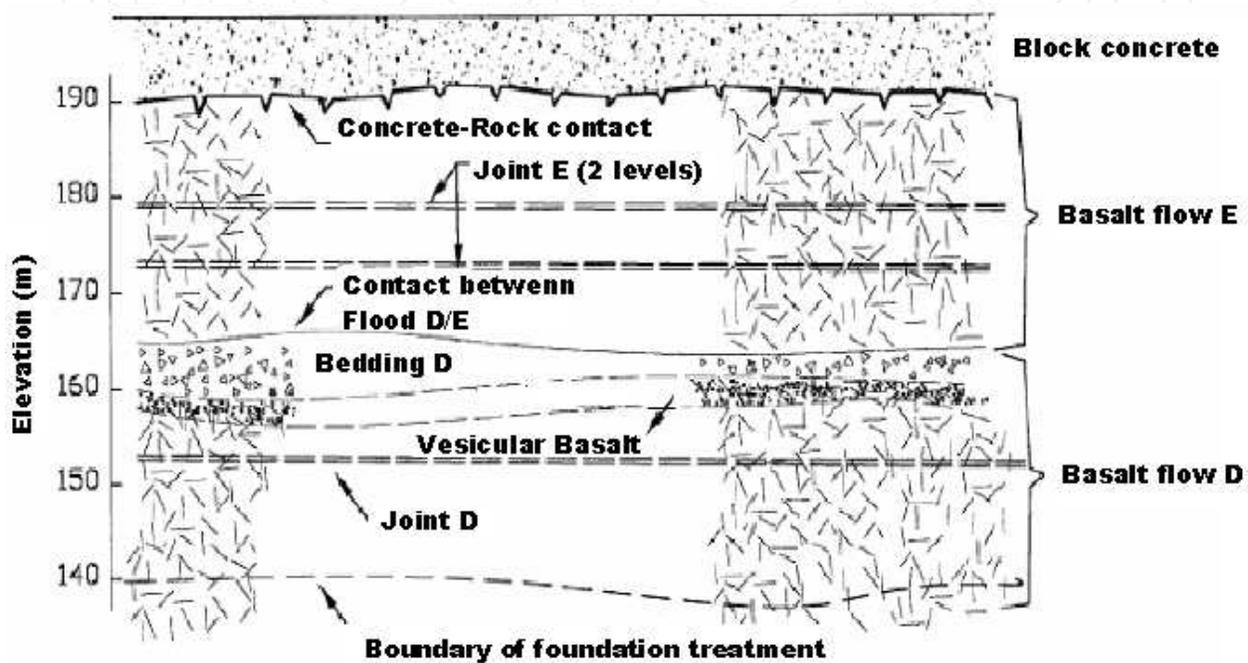
The measurement of the declinations is one of the most important observations for monitoring a dam behavior during the period of construction, of dams filling and operation. The measurement of the declination can be performed through a multiple point rod extensometer installed on boreholes [10]. Figure 3 shows the multiple point rod extensometer and an example of a typical profile of a multiple point rod extensometer at *Itaipu*.



**Figure 3.** Multiple point rod extensometer and a example of a typical profile of a multiple point rod extensometer at *Itaipu* (Adapted from [11]).

The measurements of displacements and deformation can be performed in several parts of the foundation with the usage of various rods. Among these displacement and deformations are

the contact of concrete and rock, joints and faults and other sub-horizontal discontinuities in the foundation. This approach was used at the Itaipu Dam, where different points of foundation mass were instrumented, specially the geological discontinuities. Figure 4 shows a typical geological profile of the foundation mass of the Itaipu Dam part, which has no tunnel in its right-side, where primary geological discontinuities can be found (contacts, joints, and gaps) of that specific site. In blocks where there is a transversal gallery access to the shaft, the installation of downstream-upstream extensometers can help in the measurement of the angular displacement of the dam with the foundation [10].



**Figure 4.** Schematic geological profile of the foundation of Itaipu (ITAIPU BINACIONAL, 1995, *apud* [8]).

The measurement of the horizontal displacement of the ridge is a relevant parameter which is affected by deflections of the concrete structure, by the rotation of the base of the structure (due to the deformability of the foundation), and by thermal and environmental influences. These displacements are affected by the characteristics of the concrete or by the properties of the foundation rock mass, resulting in important information for the auscultation of the behavior of the dam and of its foundation. The horizontal displacements of the ridge can be measured by a direct pendulum, usually installed at the end of the construction process. The measurements are done on the stages of reservoir spillway and of dam operation [10].

The stability of the structure in terms of sliding, overturning or floating is directly affected by the level of the piezometric pressures in the concrete-rock interface and in the sub horizontal discontinuities of low resistance that exists in the foundation. The measurements of low pressures on the concrete dam foundation are important for the monitoring of its safety conditions. The drainage is one of the most efficient ways to ensure adequate safety coefficients. The measurements of low pressure are performed by the piezometer [10].

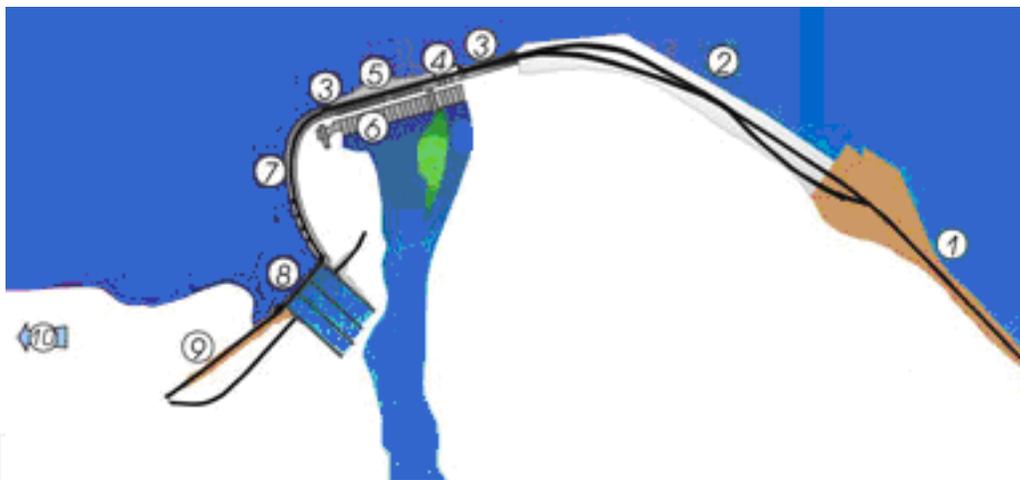
## 4. Research course

### 4.1. Itaipu Binacional

The Itaipu Binacional, the largest energy producer of the world, had its construction started in 1973 at a river stretch of Rio Paraná known as Itaipu, which in Tupi language means “the singing boulder”, located in the heart of Latin America, on the border of Brazil and Paraguay [12]. The construction of the dam ended in 1982 and the last generator unit was completed in 2008.

Nowadays, the Itaipu Dam has 20 generator units of 700 MW (megawatts) each, generating a total potential of 14.000 MW. Itaipu Binacional (Bi-national Itaipu) reached its record in producing energy in 2000, generating over 93,4 billions kilowatts-hour (KWh). It is responsible for supplying 95% of the energy consumed in Paraguay and 24% of all the Brazilian consumption.

The Itaipu Dam has 7.919m of extension and a maximum high of 196m; these dimensions made of the Itaipu construction a reference in concrete, and dam safety studies. Itaipu dam is made of two stretches of earth dam, one stretch of rock-fill dams and concrete stretches, and these forms the higher structures of it. Figure 5 illustrates the whole structure of the Itaipu dam, and table 1 shows the main characteristics of the stretches pointed on figure 5.



**Figure 5.** Whole structure of the Itaipu Complex [12].

It is possible to find in all the Itaipu extension an amount of 2.218 instruments (1.362 in the concrete, and 865 in the foundations and earthen embankments) and from this amount 270 of them are automated, to monitor the performance of the concrete structures and foundations. Furthermore, there are 5.239 drains (949 in the concrete and 4.290 in the foundations). The readings of these instruments occur in different frequencies, they can be, for example, daily, weekly, fortnightly, and monthly, depending on the type of instrument. These readings have been stored for over 30 years.

Even though, every stretch of the dam is instrumented and monitored, one of the stretches, called *Barragem principal* (main dam) (Denominated stretch F and identified as number “5”

on Figure 5), should be highlighted in a deeper study. The turbines for generating energy can be found in stretch F. In addition, this stretch is the most high water column and the most instrumented one. This stretch is made of many blocks, and each of them has instruments in the concrete structures and in the foundation that provides data about its physical behavior. This study was developed based on the data collected in this stretch of the dam.

Stretch	Structure	Lenth (m)	Maximun high (m)	
1	Auxiliary Dam (Saddle Dam)	Earth	2294	30
2	Auxiliary Dam (Saddle Dam)	Rock-filling	1984	70
3 and 7	Lateral Dams	Couterfort	1438	81
4	Deviation Structure	Concrete mass	170	162
5	Main Dam (TStretch F)	Reliefed Gravity	612	196
9	Auxiliary Dam	Terra	872	25
Other strwtches		Características		
6	Powerhouse	20 Generator units		
8	Spillway	350 m length		

**Table 1.** Characteristics of the stretches of Itaipu.

In the stretch F it is possible to find extensometers, piezometers, triothogonal meter, water level gauge and foundation instrumentation (seepage flow meter). Among these instruments, the multiple point rod extensometers, that are installed in boreholes, were selected for the analysis. This type of instrument is considered one of the most important because they are responsible for measuring the vertical displacement. That is one of the most important observations while monitoring the behavior of the dam structure. There are 30 extensometers located in stretch F.

The procedure for the methodology used for the analysis of the problem of *Itaipu* is the following:

In the first phase, the data were selected and it was decided that the methodology would be applied only to the extensometers located in stretch F.

In the second phase, the data given by *Itaipu* were converted into spreadsheets, from which the necessary data used for developing this study were extracted.

In the third phase, the data were standardized in order to receive the subsequent application of the clustering methods.

In the fourth phase, the Factor Analysis and the Clustering Analysis were applied at the same time. The Factor Analysis was also applied within each cluster formed through Clustering Analysis.

## 5. Method used

The methodology used for the analysis was applied to the data of 30 extensometers located in different blocks of stretch F of the dam, which having one or two point rods, totalizes 72 displacement measures. These measurements are identified as follow: equip4\_1, meaning rod 1 of the extensometer 4, and so on.

The data used in this study are monthly stored and they correspond to the period of January/1995 to December/2004, totalizing 120 readings. This period was chosen as a suggestion of the engineer team of Itaipu because it is subsequent the construction of the dam and prior to the system of automatic acquisition of data. During the period of system implementation, some instruments ended up having no manual readings, in addition, a total of 11 automated instruments (totalizing 24 rods) went through modifications that might have influenced the subsequent readings; there was an exchange on the instrument head for a 70 cm longer one. In this way, the referred 120 readings were immune to these irregularities.

During the period of pre-processing the data, it was identified that most of the instruments readings are monthly, but some of them showed more than one reading per month, so for this cases, the monthly average was considered. Moreover, some instruments had missing readings, in these cases; interpolations were performed through temporal series, meaning that, an adequate model was established from the Box & Jenkins methodology, using the Statgraphics [13]. In this way, it was possible to assure that all the 120 instruments had 120 readings (10 years). See [14] for more information about the interpolation techniques with temporal series.

In this way, the Matrix of entrance of structural geotechnical instrumentation data (Matrix  $Q$ ) is of order  $a \times b$ , where  $a$  is the number of patterns and  $b$  is the number of attributes. For the structural geotechnical instrumentation data of Itaipu,  $a = 72$  (number of patterns) and  $b = 120$  (number of attributes).

During the period of the Multivariate Analysis was applied and the patterns were grouped through the Ward's hierarchical clustering method. The grouping was performed in order to find out similar groups of instruments, and the aim of doing it was to establish the technical justifications for its formation. In addition, the Factor Analysis was applied to the referred data. The Factor Analysis was used to rank the rods of the extensometers through a balanced average of factor scores. Next, the Factor Analysis was applied within each group formed by the clustering analysis. Once having groups that have the instrumentations with a similar behavior, a raking of these instruments was performed within each group, in order to indicate the most relevant instruments, which would be chosen, for example, in cases of intensifying the reading.

## 5.1. Statistical multivariate analysis

### 5.1.1. Factor analysis

Factor Analysis is a multivariate statistical method, which objective is to explain the correlations between one large set of variables in terms of a set of unobserved low random variables called factors. Hence, suppose the random vector  $\underline{X}$  resulted from  $p$  random variables;  $\underline{X}' = [x_1 \ x_2 \ x_3 \ \dots \ x_p]$ , and in order to study the covariance structure of this vector, in other words, if  $\underline{X}$  is observed  $n$  times, it happens that its parameters  $E(\underline{X}) = \underline{\mu}$  e  $V(\underline{X}) = \Sigma$  can be estimated and the relation between the evaluated variables represented by matrix of covariance  $\Sigma$  or of correlation  $\rho$ . The factor analysis makes a grouping of variables to explain the influence of latent variables (unobserved) or factors. Within a same group, the variables are highly correlated with each other, and from one group to another, the correlations are low. Each group represents a factor, which is responsible for the observed correlation.

The covariance matrix of the vector  $\underline{X}$  can be placed in an exact form:  $V(\underline{X}) = \Sigma = LL' + \Psi$ , where matrix  $LL'$  has on the main diagonal the called communality defined for each variable considering  $m$  factors by:  $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{ip}^2$ . However, considering the  $m$  main factors, it is given that  $h_i^2 = l_{i1}^2 + l_{i2}^2 + \dots + l_{im}^2$ ,  $i = 1, 2, \dots, p$  variables. In this way, the communality  $h_i^2$  is the part of the variance of the random variable  $x_i$  that comes from  $m$  factors. And, the part of the variance of the random variable  $x_i$  due to the factors  $p - m$  that are not important are called specific variance. Hence,  $V(x_i) = h_i^2 + \psi_i$ ,  $i = 1, 2, \dots, p$ .

There are many criteria to define  $m$  number of factors. The most used one is the Kaiser criterion [15], which suggests that the number of extracted factors, must be equal to the number of eigenvalues higher than one, of  $\Sigma$  or  $\rho$ .

If  $\underline{X}$  is a random vector, with  $p$  components, and the parameters  $E(\underline{X}) = \underline{\mu}$  e  $V(\underline{X}) = \Sigma$ , in factor model orthogonal,  $\underline{X}$  is linearly dependent upon several random unobserved variables,  $F_1, F_2, \dots, F_m$  called common factors and  $p$  sources of joining variables:  $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_p$ , called errors or specific factors.

The model of Factor Analysis is represented below, where  $\mu_i$  is the average of the  $i$ -th variable,  $\varepsilon_i$  is the  $i$ -th error, or specific factor,  $F_j$  is the  $j$ -th common factor and  $l_{ij}$  is the weight of the  $j$ -th  $F_j$  factor on  $i$ -th  $X_i$  variable. Equation 1 shows the model represented in matrix terms.

$$\begin{cases} X_1 - \mu_1 = l_{11} F_1 + l_{12} F_2 + \dots + l_{1m} F_m + \varepsilon_1 & i = 1, 2, \dots, p \\ X_2 - \mu_2 = l_{21} F_1 + l_{22} F_2 + \dots + l_{2m} F_m + \varepsilon_2 & j = 1, 2, \dots, m \\ \dots \\ X_p - \mu_p = l_{p1} F_1 + l_{p2} F_2 + \dots + l_{pm} F_m + \varepsilon_p & m \leq p. \end{cases}$$

$$\underline{X} = \underline{\mu} + LL' + \Psi \quad (1)$$

In order to estimate the loading  $l_{ij}$  and the specific variables  $\psi_i$ , the method of principal components can be used, which is briefly described below [15].

If the pair of eigenvalues and eigenvectors are  $(\lambda_i, \underline{e}_i)$  of the matrix of sample covariance  $S$ , with  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  and  $m < p$  is the number of common factors the matrix of the estimated loadings is given by  $L = CD^{1/2}$ , where  $C$  is the matrix of the eigenvectors and  $D$  is a diagonal matrix of which the diagonal elements are the eigenvalues.

In the application of this method, the observations are primarily centralized or standardized, in other words, the matrix of correlation  $R$  (estimator of  $\rho$ ) is used in order to avoid problems of scale. The specific variances estimated  $\widehat{\Psi}_i$  are given by the diagonal elements of the matrix  $\widehat{\Psi}_i = S - LL'$ .

In multiple actions, it is necessary to estimate the value of the scores of each factor (unobserved) for an individual  $\underline{X}$  observation. These factor values are called factor scores. The estimated factor scores to the original variables are  $\underline{F} = (L'L)^{-1} L'(\underline{X} - \bar{\underline{X}})$  and for the standardized variables are  $\underline{F} = (L'L)'L\underline{z}$ , that is, if the principal components are used in order to estimate the loadings.

According to [15], with the rotation of factors, a structure is obtained for the low or moderated loadings on the other factors. This leads to a more simplified structure to be interpreted. Kaiser suggested an analytical measure known as *Varimax* criteria [15] in order to make the rotation.

The rotation coefficient scaled by the square root of the communalities is defined by  $\tilde{l}_{ij} = \frac{l_{ij}}{h_{ij}}$ . The *Varimax* selects the orthogonal transformation  $T$  that turns  $V$  (given by equation 2) the largest possible, in other words, the procedure starts from  $\sum LT'L'$  and gives the loadings  $\underline{L}^*$  which comes from  $LT$ . Hence, the criterion is to maximize  $V$ .

$$V = \frac{1}{p} \sum_{i=1}^p \left[ \frac{\sum_{i=1}^p l_{-ij}^{*4} - \sum_{i=1}^p (l_{-ij}^{*2})^2}{p} \right] \quad (2)$$

In Factor Analysis, communality  $h_i^2$  is the portion of the variance of the variable that is attributed to the factors and represents the percentage of variation of the variable which is not random but from the factors. Thus, the criterion used to classify the patterns is sort the variables (instruments) according to their factor scores. The factor scores were evaluated by a factor that distinguishes the behavior of the instrument, using it as a practical and simple quality control of the measurement of the instrument.

To perform the ranking of the variables (instruments), a final factor score was used, which is given by equation (3), where  $m$  is the number of factors extracted,  $\lambda_i$  are the eigenvalues and  $f_i$  are the factor scores.

$$final\_factor\_score = \frac{\sum_{i=1}^m \lambda_i f_i}{\sum_{i=1}^m \lambda_i} \quad (3)$$

The Factor Analysis was done with the aid of the computational *Statgraphics* [13].

### 5.1.2. Cluster analysis

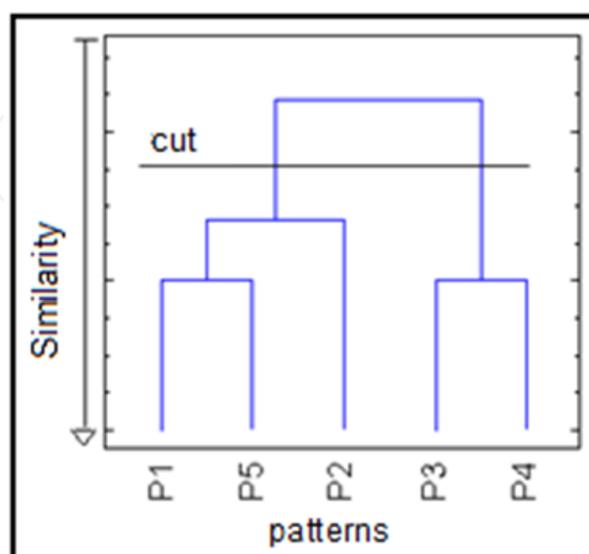
The clustering is a manner of grouping in a way that those patterns inside the same group are very similar to each other, and different from patterns of the other groups. According to [16], cluster analysis is an analytical technique used to develop meaningful subgroups of objects. Its objective is to classify the objects in a small number of groups that are mutually exclusive. According, to [17], it is important to favor a small number of groups in cluster analysis.

The clustering algorithms can be divided into categories in many ways, according to its characteristics. The two main classes of clustering are: the hierarchical methods and the nonhierarchical methods.

The hierarchical methods include techniques that connection of the items assuming obtain various levels of clustering. The hierarchical methods can be subdivided into divisive or agglomerative ones. The agglomerative hierarchical method considers at the beginning each pattern as a group and interactively, clusters a pair of groups that are the most similar with a new group until there is only one group containing all patterns. In the other hand, the divisive hierarchical method, starts with a single group and performs a process of successive subdivisions [18].

The most popular hierarchical clustering methods are: Single Linkage, Complete Linkage, Average Linkage and Ward's Method. The most common method of representing a hierarchical cluster is using a dendrogram that represents the clustering of the patterns and the levels of similarity in which the groups are formed. The dendrograms can be divided in different levels, showing different groups [19].

In the dendrogram (figure 6), two groups can be seen by admitting a cut on the level represented by the figure. The first one composed by patterns  $P1$ ,  $P2$  and  $P5$  and the second one composed by patterns  $P3$  and  $P4$ .



**Figure 6.** Example of dendrogram.

Methods that are not hierarchical or partitioning seek for a way of partitioning without the need of hierarchical associations. Optimizing some criteria, a partition of the elements on  $k$  group is selected [18].

The most known method among the nonhierarchical methods is the  $k$ -means cluster method [15]. Normally, the  $k$  clusters that are found are of better quality than the  $k$  clusters produced by the hierarchical methods. The methods of partitioning are advantageous in applications that involve larger series of data.

The methods of the Multivariate Statistics field were used because these are already common methods. The Multivariate Statistic Analysis is an old method that has been made feasibly recently with the advance of present, fast and economic computation.

The clustering of the patterns is based on the measure of similarity and dissimilarity. The similarity measure evaluates the similarities of the objects, in other words, the highest the measures value are the most similar are the objects. The most known mean of similarity is the correlation coefficient. The means of dissimilarity evaluates whether the objects are dissimilar, this is to say, that the highest the measure value are the less similar the objects are. The most known measure of dissimilarity is the Euclidean distance.

According to [15], Ward's method performs the join of two clusters based on the "loss of information". It is considered to be the criteria of "loss of information" the sum of the error square ( $SQE$ ). For each cluster  $I$ , the measure of the cluster (or centroide) of the cluster and the sum of the cluster error square ( $SQE_i$ ) which is the sum of the error square of each pattern of the cluster in relation to the measure. For cluster  $k$  there is  $SQE_1, SQE_2, \dots, SQE_k$ , where  $SQE$  is defined by equation 4.

$$SQE = SQE_1 + SQE_2 + \dots + SQE_k \quad (4)$$

For each pair of cluster  $m$  and  $n$ , first, the measure (or centroide) of the formed cluster is calculated (cluster  $mn$ ). Then, the sum of error for the square of cluster  $mn$  is calculated ( $SQEmn$ ), according to equation 5.

$$SQE = SQE_1 + SQE_2 + \dots + SQE_k - SQE_m - SQE_n + SQE_{mn} \quad (5)$$

The clusters  $m$  and  $n$  that show the lower increase on the sum of error square ( $SQE$ ) (lower loss of information) will be gathered. According to [16], this method tends to obtain clusters of same size due to the decrease of its internal variation.

Cluster Analysis was applied with the aid of the computational software *Statgraphics* [13]. The measure of dissimilarity used was the Euclidean distance. The data were standardized.

## 6. Status

This research was performed during the first author's (Rosangela Villwock) doctorate process, from 2005 to 2009, in the Post-graduation Program on Numerical Methods in Engineering, of the Federal University of Paraná, guided by the second author of this text

(Maria Teresinha Arns Steiner). This study was part of a project guide by the third author (Andrea Sell Dyminski), called "*Análise de Incertezas e Estimção de Valores de Controle para o Sistema de Monitoração Geotécnico-estrutural na Barragem de Itaipu*" (Estimation of Control Values for the System of Geotechnic-structural Monitoring in the Itaipu Dam). All the research process counted with the collaboration of the fourth author (Anselmo Chaves Neto) and it was also supervised by him.

As mentioned before, the aim of this paper is to identify the instruments that are the most significant to the analysis of the behavior of dams. There are no records of the existence of methods that perform the ranking of the instruments of monitoring dams. In order to achieve this aim, it is necessary to select, cluster and rank geotechnical-structural instruments of an electric power plant looking forward to maximizing the effectiveness and efficiency of the readings analysis, in our case the Itaipu Hydroelectric Power Plant. In case of needing to intensify the reading this hierarchy could be useful to define which instruments to choose.

The choice of instrumentation is performed with no previous knowledge about the location, features, or characteristics of the instruments. In this way, it is possible to think of applying the methodology when making decisions about the automation of the additional instruments. Approaches that are similar to this can be used in many other cases because there are hundreds of large Civil Engineering construction works that rely on systems of instrumentation in Brazil which the data must have an appropriate treatment.

## 7. Results

In the cluster Analysis, the patterns are the rods of the extensometers, and its readings along the months which are compared in order to determine the clusters. The dendrogram on figure 7 shows the formation of the clusters for these data.

Considering the first cut, there are two clusters left. The first cluster, here denominated "cluster 1", is formed by the rods of the extensometers that are considered extremely important to the monitoring of the dam. They are rods of extensometers installed in the axis of the block upstream the dam and inclined  $60^\circ$  towards upstream.

Notice that there is a formation of two additional clusters in the second cut. The first one denominates "cluster 2" which most of its rods of the extensometer installed in the balsatic rocks B, C and D (A and B are called the deepest rocks; C and D are called the superficial rocks), and on the lithological contacts B/C and C/D. The second cluster, denominated "cluster 3" has most of the rods of the extensometers installed in the joints (between the rock layers) A and B and on the lithological contact A/B.

This was the quantity of clusters which are been considered (3 clusters), since it was possible to obtain technical justification for its formation. In a larger subdivision, such justification was not observed.

Notice that at this point it was possible to cluster the instruments according to the relevant geological characteristics of the foundation mass, even though they weren't explicitly



of cluster 1, it means that there is an external phenomenon influencing them. As mentioned before, these rods reflect the effects of the summer and winter. In the same way, each factor is dominated by a set of rods and there is an external phenomenon that explains each set of rods or factor, even though it is not easy to interpret them.

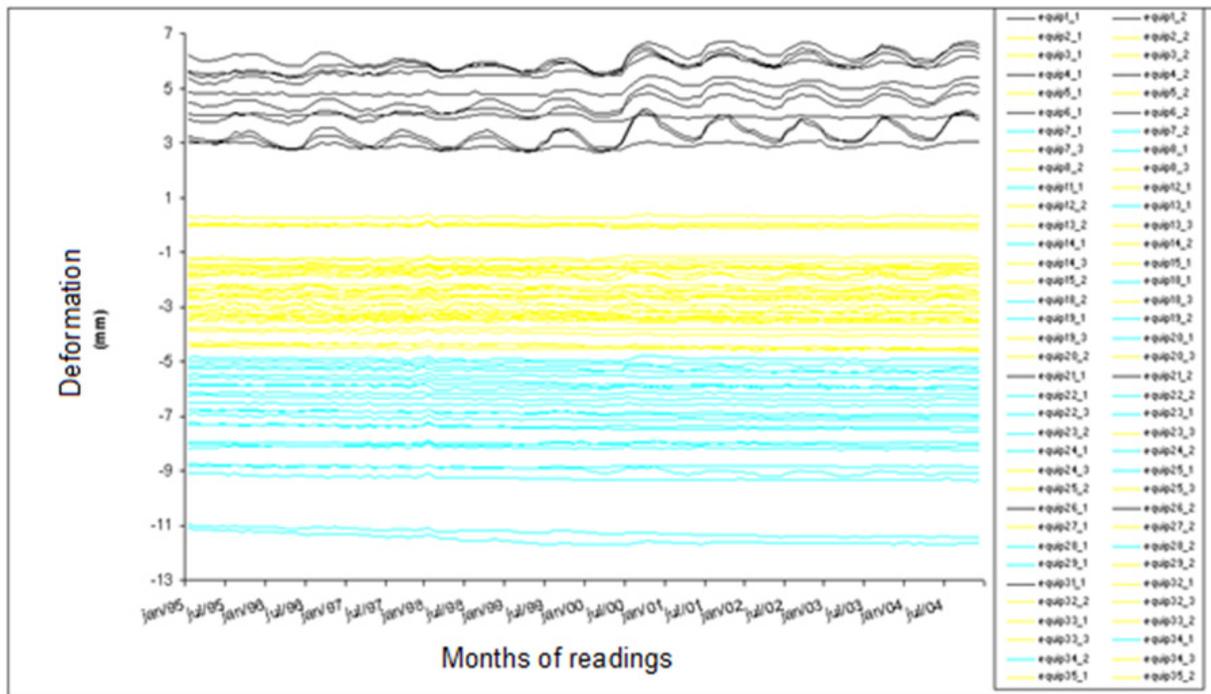


Figure 8. Graphic of all rods of the extensometers from the period of study.

factor1	equip7_1	equip7_2	equip7_3	equip11_1	equip12_1	equip12_2	equip13_1
	equip13_2	equip14_1	equip14_2	equip14_3	equip19_2	equip20_2	equip20_3
	equip22_1	equip22_2	equip22_3	equip23_1	equip23_2	equip23_3	equip24_1
	equip24_2	equip24_3	equip25_1	equip25_2	equip27_2	equip28_1	equip28_2
	equip29_1	equip29_2	equip32_3	equip33_1	equip33_2	equip33_3	equip34_1
	equip34_2	equip34_3	equip35_1	equip35_2			
factor2	equip1_1	equip1_2	equip4_1	equip4_2	equip6_1	equip6_2	equip8_1
	equip8_3	equip21_1	equip21_2	equip25_3	equip26_2	equip31_1	
factor3	equip2_1	equip2_2	equip3_1	equip3_2	equip5_1	equip5_2	
factor4	equip13_3	equip18_1	equip18_2	equip18_3	equip19_1	equip19_3	
factor5	equip15_1	equip15_2					
factor6	equip8_2	equip20_1	equip32_2				
factor7	equip26_1						
factor8	equip32_1						

Table 2. Rods of extensometers that are important to each factor, according to its weights in the Factor Analysis.

A community is the portion of the variation of the extensometer rods which is explained by its factors. A low community within a rod indicates that the same is not greatly affected by the factor because a community is the sum of contributions of each rod in each square factor. Thereon, in this case the influence mainly comes from a random factor. Notice that none of the extensometer rods showed lower community than 0.71, it means that none of the random variations are over 29%. A community that is equal to 0.71 indicates that the 71% of the rods extensometer variations is ascribed to the factors and that only 29% of those variations is random, it means that these correlated rods are working properly. A low community would indicate a need of investigating the rods.

Table 3 shows 25 rods of extensometers with the highest communalities. In case of reading intensification, these rods are the recommended ones. The highlighted rods are part of the system of automatic data acquisition of Itaipu. 24 of the 74 rods that were analyzed were automated by the engineers' team of Itaipu. The method of ranking that was proposed (without the previous clustering of the rods) indentified 14 of the 24 automated rods.

Communality	0,988861	0,981763	0,976523	0,975655	0,972231	0,971971
Rod	equip29_1	equip21_2	equip23_1	equip22_1	equip3_1	equip1_1
Communality	0,971798	0,970804	0,970397	0,968213	0,968083	0,967029
Rod	equip22_3	equip11_1	equip1_2	equip23_2	equip4_1	equip21_1
Communality	0,966632	0,965999	0,965522	0,964925	0,963139	0,960121
Rod	equip4_2	equip29_2	equip34_3	equip6_1	equip6_2	equip22_2
Communality	0,957609	0,953036	0,950395	0,949394	0,949108	0,948646
Rod	equip14_3	equip25_1	equip33_2	equip24_2	equip24_1	equip5_1
Communality	0,943644					
Rod	equip28_1					

**Table 3.** Shows the 25 rods of extensometers with the highest communalities.

After forming three clusters, the ranking of the rods was performed within each cluster with the help of the Factor Analysis. The hierachization within each group can also be used to identify rods used on readings intensification. The advantage of application of ranking within each group is that a separation of the rods with similar behavior is firstly obtained then the indicated rods will well represent the variability of the cluster. Note that the rods of the automated extensometers are, mostly, among the first in the ranking of each cluster.

As mentioned above, a low communality of a rod indicates that this rod is not strongly influenced by the factors and, in this case, the influence comes from random factors. In the application of Factor analysis within each cluster, there are rods of extensometers with communities between 0,6 and 0,7, in other words, random variation between 30% and 40%. It is indicated that the investigation on the rods is performed in this case.

Furthermore, in order to identify the 24 rods that are the most relevant, we opted, in first place, to identify the 8 best ranked rods from each cluster. In this case, there would be 15 out of the 24 automated rods. This number of rods coinciding with the automated ones in Itaipu

would increase with the aid of a specialist for a better interpretation of the results. This specialist would detect that cluster 1, for example, is formed by rods that are extremely important for the monitoring of dams, and that all rods from this cluster should be automated.

This type of analysis was not found in literatures, for this reason the contribution of this study is relevant. It is recommended that this Analysis (process of hierarquization) is repeated periodically (according to the needs indicated by the specialists in this field – in this case, by the engineers' team of Itaipu) what could be done, for example, every two years. This can show the appearance of new rods that are indicated by the performing of readings intensification (that should be investigated), the same could occur with rods that would no longer be indicated.

When there are rods within the clusters with low communalities, it is recommended that they are investigated. Low communality indicates a high percentage of randomness in the data and that can be an indicator of problems with the rods.

These identifications of similar rods can also be used in projecting the control values. In this case, the values of control for each rod can be associated to the readings of the rods that belong to a same cluster.

The final factorial score performs the hierachization of the attributes. In this case the patterns are vectors of which the components (attributes) are the readings of the rods of the extensometers in a certain month. Therefore, the final factorial score performs the hierachization of the months showing whether there is any month that is rather relevant and that deserves greater attention.

Table 4 shows the first 15 months with a higher final factorial score and the last 15 months with lower final factorial scores, considering the 72 rods of extensometers. The values of the 15 first months with a higher final factorial score reveal that all the months are important; there is no month that is rather relevant. Only the month of December does not appear in the first 15 months. Notice that 1995 was the most relevant year and in analyzing the ambient temperature during the period of study it was possible to verify that this occurred due to the high temperature variation. The values of the last 15 months with least final factorial scores revealed that the months of April, May, and June are the most important one, identifying the effects of summer.

As mentioned, cluster 1 shows the effect winter/summer in its readings. For this reason the final factorial score was calculated in order to perform a ranking of the months for cluster 1, to show whether there is any month or some months with greater relevance.

The first 15 months with a higher final factorial score and the last 15 months with least final factorial scores were observed considering only the 11 rods of the extensometer of cluster 1. The values of the 15 first months with higher final factorial scores reveal that the months of September, October, and November are the most relevant ones, identifying the effects of winter. The values of the last 15 months with least final factorial score reveal that the months of March, April, May and June are the most important ones, identifying the effects of summer.

The identification of the months with more significant readings for an external effect (in this case, the effect of summer and of winter on the readings of the rods of the extensometers), can be useful, for example, in the projection of the values of control. Admitting that there are differences in the readings of the rods for the months related above, only the readings performed in these months would be used to define specific values of control for these months.

15 first		15 last	
Final factorial score	Month	Final factorial score	Month
1,755	January/98	-0,761	April/02
1,217	August/95	-0,816	July/03
1,153	January/95	-0,821	February/03
1,091	February/95	-0,821	June/00
0,992	June/95	-0,856	June/02
0,914	March/95	-0,877	May/03
0,902	April/95	-0,904	February/00
0,877	November/96	-0,924	May/00
0,794	May/95	-0,934	May/02
0,781	July/95	-0,965	April/00
0,776	April/97	-0,971	April/04
0,749	October/95	-1,050	April/03
0,741	November/95	-1,061	June/03
0,710	September/95	-1,135	May/00
0,709	February/96	-1,152	March/03

**Table 4.** Shows final factorial scores of the months in which the readings of the 72 rods of the extensometers were performed.

## 8. Further research

The application of this methodology is suggested for other instruments and other periods, and the implementation of it in order to define values of control and for anomaly detection. Once the process of ranking is repeated in several periods (every 2 years, for example.) it can show the appearance of new rods which are indicated for performing readings intensification or the appearance of rods that could no longer be indicated (these should be investigated).

## 9. Conclusions

This manuscript shows a methodology that uses some techniques of the field of Multivariate Analysis, which aim is to select, cluster and rank geotechnical-structural instruments of a Hydroelectric power plant, in our case, the Itaipu hydroelectric power plant, in order to maximize the efficiency and effectiveness of the analysis of the readings.

The methodology showed was applied to the instruments called extensometers, located in different points of block F of the dam, a total of 30 extensometers that with one, two or three point rods totalized 72 measures of monthly displacement. These measures were stored over a period of 10 years, totalizing 120 readings (January/1995 to December/2004). It is important to remember that 24 measures out of the 72 were automated by the company. The ranking of the instruments would be a way to choose the instruments without any previous knowledge about its location, features, or other characteristics. In this way, it is possible to think in applying this methodology in further decision-making when it relates to the automation of additional new instruments.

The methodology used to analyze the problem of the research was composed by the following form: Ward's method was applied in order to cluster 72 rods of extensometers; at the same time, the Factor Analysis was applied in order to rank the rods; latter, the Factor Analysis was applied within each cluster formed by Clustering Analysis.

In the Factor Analysis applied to the 72 rods, there was not need of investigation for any of the rods, once the communality was high for each of them. Observing the 25 rods of extensometers with the highest communality, 14 rods were identified among the ones that were automated by the team of engineers of Itaipu (the automated rods are the ones considered the most important), in other words, the proposed hierachization method (without previous clustering of the rods) identified 14 of the 24 automated rods.

The Clustering Analysis shows that it is possible to find technical justification for the formation of three clusters. The instruments were clustered according the relevant geological characteristics of the foundation mass, although they were not explicitly shown to the technicians.

By observing the clusters 1, 2, and 3, the factor analysis was applied within each cluster in order to perform the ranking of the rods of the extensometers. It was possible to notice that the rods of the automated extensometers are, most of the time, among the first ones of the ranking of each cluster.

In order to identify the 24 rods that are the most relevant, we decided to identify the 8 best ranked rods from each cluster. In this case, there would be 15 of the 24 automated rods. This number of rods coinciding with the automated ones in Itaipu would increase with the aid of a specialist for a better interpretation of the results. For instance, this specialist would detect that cluster 1 is formed by rods that are extremely important for the monitoring of dams and that all rods from this cluster should be automated.

Approaches that are similar to this can be used in many other cases, since there are thousands of large construction works of Civil Engineering that use the system of instrumentation, of which the data can and must receive an appropriate treatment.

The approach of an important problem of engineer, the analysis of the instrumentation data of large construction works, clustering techniques and other techniques were applied, in the context of the Multivariate Statistical Analysis, aiming the identification of the instruments that are the most significant ones to the analysis of the behavior of dams.

## Author details

Rosangela Villwock \*

*Mathematics division, Western Paraná State University, Cascavel, Brasil*

Maria Teresinha Arns Steiner, Andrea Sell Dyminski and Anselmo Chaves Neto

*PPGMNE, Federal University of Paraná, Curitiba, Brasil*

## Acknowledgement

The authors would like to thank Itaipu's Civil Engineering team for instrumentation data and technical contributions.

## 10. References

- [1] Icold - International Commission on Large Dams. <http://www.icold-cigb.org>, 2008.
- [2] CBGB - Comitê Brasileiro de Grandes Barragens. Diretrizes para a inspeção e avaliação de segurança de barragens em operação. Rio de Janeiro, 1983. 26 p.
- [3] Brasil. Projeto de Lei Nº 1.181/2003. Estabelece diretrizes para verificação da segurança de barragens de cursos de água para quaisquer fins e para aterros de contenção de resíduos líquidos industriais. Disponível em [http://www.emtermos.com.br/ABMS/PL\\_1181.pdf](http://www.emtermos.com.br/ABMS/PL_1181.pdf). Acesso em 19/06/2009.
- [4] Kalustyan, E. S. Assessment and role of risk in dam building. *Hydrotechnical Construction*, v. 33, n. 12, 1999.
- [5] Yenigun, K.; Erkek, C. Reliability in dams and the effects of spillway dimensions on risk levels. *Water Resources Management*, v. 21, p. 747-760, 2007.
- [6] MenescaL, R. de A. Gestão da Segurança de Barragens no Brasil - Proposta de um Sistema Integrado, Descentralizado, Transparente e Participativo. 769 f. Tese (Doutorado em Engenharia Civil) - Departamento de Engenharia Hidráulica e Ambiental, Universidade Federal do Ceará, Fortaleza, 2009.
- [7] ABNT – Associação Brasileira De Normas Técnicas. Ações e Segurança nas Estruturas (NBR 8681), 2003.
- [8] Osako, C. I. A Manutenção dos Drenos nas Fundações de Barragens - O Caso da Usina Hidrelétrica de Itaipu. 126 f. Dissertação (Mestrado em Construção Civil). – Setor de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2002.
- [9] FEMA – Federal Emergency Management Agency, Federal Guidelines For Dam Safety, U. S. Department Of Homeland Security, USA, 2004.
- [10] Silveira, J. F. A. Instrumentação e Comportamento de Fundações de Barragens de Concreto. São Paulo: Oficina de Textos, 2003.
- [11] Matos, S. F. Avaliação de Instrumentos para Auscultação de Barragem de Concreto. Estudo de caso: Deformímetros e Tensômetros para Concreto na Barragem de Itaipu.

---

\* Corresponding Author

- 106 f. Dissertação (Mestrado em Construção Civil). – Setor de Ciências Exatas, Universidade Federal do Paraná, Curitiba, 2002.
- [12] ITAIPU – Itaipu Binacional. Disponível em <http://www.itaipu.gov.br/>. Acesso em 28/08/2008.
- [13] Statgraphics Plus 5.1 – Statgraphics Plus 5.1, Statistical Graphics Corp., Rockville, 2001.
- [14] Box, G.E.P.; Jenkins, G.M. Time Series Analysis, forecasting and control. Ed. Holden Day, 1976.
- [15] Johnson, R.A.; Wichern, D.W. Applied Multivariate Statistical Analysis. Sixth Edition. New Jersey: Pearson Prentice Hall, 2007.
- [16] Hair Jr, J.F.; Anderson, R.E.; Tatham, R.L.; Black, W.C. Análise Multivariada de Dados. Tradução de: Santanna, A. S.; Chaves Neto, A. Porto Alegre: Bookman, 2005.
- [17] Freitas, A. A. Data Mining and Knowledge Discovery with Evolutionary Algorithms. New York: Springer, 2002.
- [18] Diniz, C. A. R.; Louzada Neto, F. Data mining: uma introdução. São Paulo: ABE, 2000.
- [19] Jain A. K., Murty M. N., Flynn P. J. Data clustering: a review. ACM Computing Surveys. v. 31, n. 3, 1999.