MULTIVARIATE STATISTICAL ANALYSIS OF MONITORING DATA FOR CONCRETE DAMS

Behrouz Ahmadi Nedushan

December 2002

Department of Civil Engineering and Applied Mechanics

McGill University, Montreal

Thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements of the degree of

Ph.D

(Structural Engineering)

Your file Votre référence
ISBN: 0-612-88408-2
Our file Notre référence
ISBN: 0-612-88408-2

Dedicated To My mother

Abstract

Major dams in the world are often instrumented in order to validate numerical models, to gain insight into the behavior of the dam, to detect anomalies, and to enable a timely response either in the form of repairs, reservoir management, or evacuation. Advances in automated data monitoring system makes it possible to regularly collect data on a large number of instruments for a dam. Managing this data is a major concern since traditional means of monitoring each instrument are time consuming and personnel intensive. Among tasks that need to be performed are: identification of faulty instruments, removal of outliers, data interpretation, model fitting and management of alarms for detecting statistically significant changes in the response of a dam.

Statistical models such as multiple linear regression, and back propagation neural networks have been used to estimate the response of individual instruments. Multiple linear regression models are of two kinds, (1) Hydro-Seasonal-Time (HST) models and (2) models that consider concrete temperatures as predictors.

Univerariate, bivariate, and multivariate methods are proposed for the identification of anomalies in the instrumentation data. The source of these anomalies can be either bad readings, faulty instruments, or changes in dam behavior.

The proposed methodologies are applied to three different dams, Idukki, Daniel Johnson and Chute-à-Caron, which are respectively an arch, multiple arch and a gravity dam. Displacements, strains, flow rates, and crack openings of these three dams are analyzed.

This research also proposes various multivariate statistical analyses and artificial neural networks techniques to analyze dam monitoring data. One of these methods, Principal Component Analysis (PCA) is concerned with explaining the variance-covariance structure of a data set through a few linear combinations of the original variables. The general objectives are (1) data reduction and (2) data interpretation. Other multivariate analysis methods such as canonical correlation analysis, partial least squares and

nonlinear principal component analysis are discussed. The advantages of methodologies for noise reduction, the reduction of number of variables that have to be monitored, the prediction of response parameters, and the identification of faulty readings are discussed. Results indicated that dam responses are generally correlated and that only a few principal components can summarize the behavior of a dam.

Résumé

Les grands barrages sont souvent instrumentés afin de valider les modèles numériques, pour développer une meilleure compréhension du comportement des barrage, détecter des anomalies, et permettre une réponse opportune sous forme de réparations, de gestion de réservoir, ou d'évacuation. Les progrès récents dans les systèmes de surveillance automatisés permettent la collecte simultanée des données sur un grand nombre d'instruments. La gestion de ces données est un souci important puisque les moyens traditionnels d'analyse pour chaque instrument sont laborieux.. Les tâches d'analyse qui doivent être accomplies sont : (1)l'identification des instruments défectueux, (2) l'élimination des valeur errorrées ou aberrantes la sélection et l'ajustement des modéles et la gestion des alarmes pour détecter les changements statistiquement significatif dans la réponse d'un barrage.

Des modèles statistiques tels que la régression linéaire multiple et les réseaux neurologiques ont été employés pour estimer la réponse de différents instruments. Les modèles de régression linéaire multiple sont de deux sortes : (1) les modèles Hydraulique-Saisonnier-Temps (HST) et (2) les modèles qui considèrent les températures du béton parmi les prédicteurs.

Des méthodes à une, deux, ou plusieures variables sont discutées. pour l'identification des anomalies dans les données d'instrumentation. La source de ces anomalies peut être des lectures errorrées, des instruments défectueux, ou un changement de comportement du barrage.

Les méthodologies proposées sont appliquées à trois barrages, Idukki , Daniel Johnson et Chute-à-Caron, qui sont respectivement des barrages voûtes,à voûtes multiples et un barrage poids. Les déplacements, les contraintes, les débits, et les ouvertures de fissures ou de joints de ces trois barrages sont analysés.

Diverses méthodes d'analyse statistique multivariables et d'analyse par réseaux neurologiques sont également discutées pour analyser des données d'instrumentation du barrage. Une de ces méthodes, l'Analyse par Composantes Principales (ACP) a pour objectif de décrire la struture de variance-covariance des données par le biais de quelques combinaisons linéaires des variables initiales. Les avantages principaux de cette méthode sont : (1) le réduction des données et (2) l'intérpretation de données. Le nombre de variables qui doivent être surveillées peut être réduit sans perte significative d'information. D'autres méthodes multivariables telles que l'analyse canonique, l'analyse PLS et l'analyse non-linéaire par de composant principales sont discutées.

Les avantages de ces méthodes pour la réduction du bruit de fond, la réduction du nombre de variables qui doivent être surveillées, la prédiction des observations, et l'identification des lectures erronées sont discutés. Les résultats indiquent que les données d'instrumentation du barrage sont corrélées et que seulement quelques composantes principales peuvent décrire le comportement du barrage.

Acknowledgements

First, I would like to express my sincere gratitude to my supervisor Professor Luc E. Chouinard for his valuable guidance, patience, constructive comments and friendly manner throughout the course of this research. I am also indebted to him for introducing me to field of statistical data analysis and its application in practical engineering problems.

I express my regards to all my Professors in McGill University during my graduate studies, and Sharif University of Technology during my undergraduate studies. My appreciation is also extended to Dr. William Cook who has always been very helpful in solving my computer problems.

I express my gratitude to the McGill University for awarding me the David Stuart Major fellowship in 1995-1998.

The author appreciates the contribution of Hydro-Quebec for providing the data of Daniel Johnson Multiple arch dam, and Snc-Lavalin for providing the data for Idukki arch dam and Chute-à-Caron gravity dam.

On a personal basis I would like to express my deepest gratitude to my parents for their unending support and encouragement at all levels during my many years of study. No words can truly express my deepest and sincere appreciation for all the sacrifices that my mother and my late father have done for all their children.

Last but not least, I would like to thank my wife Mahshid for her understanding, unconditional love and continuous moral support during these past few years. She has also lent me her excellent programming skills at times I needed to quickly write some statistical programs.

TABLE OF CONTENTS

List of Figures

List of Symbols

ε	Residuals
z	Normalized reservoir water level
H	Reservoir water level
H_{min}	Minimum Reservoir water level
H_{max}	Maximum Reservoir water level
t	Time
θ	Temperature
k	Number of principal components
X	Matrix of predictor variables
Y	Matrix of responses
m	Number of responses
n	Number of observations
p	Number of predictor variables
$S_{y,x}$	Standard errors of estimate
b	Vector of regression coefficients
$\hat{b}$	Estimate of regression coefficients
R	Correlation Matrix
S	Covariance matrix
$\underline{\Lambda}$	Diagonal matrix of eigenvalues
λ_i	i^{th} eigenvalue
$\underline{p}_i$	i^{th} eigenvector
$\underline{S}_{xx}$	Sample variance covariance matrix of predictor variables
$\underline{S}_{yy}$	Sample variance covariance matrix of response variables
E	Residuals matrix
$\underline{f}_i$	i^{th} Principal component
$\underline{v}_i$	i^{th} canonical variable for the first set of data
$\underline{w}_i$	i^{th} canonical variable for the second set of data
r_i	Canonical correlation for the i^{th} pair of canonical

CHAPTER 1

1. Introduction

1.1 General

For centuries, dams have provided mankind with such essential benefits as water supply, flood control, recreation, hydropower, and irrigation. They are an integral part of society's infrastructure. Dam failures are rated as one of the major "low probability, high-loss" events. The large number of dams that are 50 or more years old is a matter of great concern, since they are generally characterized by increased risk due to structural deterioration or inadequate spillway capacity (NRC, 1983). Performance monitoring of existing dams is an essential part of a dam safety program.

Performance monitoring of dams is accomplished by conducting visual observations, and reviewing and analyzing data collected from instruments, which measure critical indicators of structural behavior. "Instrumentation of a dam furnishes data to determine if the completed structure is functioning as intended and to provide a continuing surveillance of the structure to warn of any developments which endanger its safety"

(Post, 1985). The means and methods available to monitor phenomena that can lead to dam failure include a wide spectrum of instruments and procedures ranging from very simple to very complex. Any program of dam safety instrumentation must be properly designed and consistent with other project components, must be based on prevailing geological and geotechnical conditions at the dam, and must include consideration of the hydrologic and hydraulic factors present both before and after the project is in operation.

Measurements complement visual observations as a continuing surveillance system of the threat to life, property, and the environment, and assist in investigating unexpected or abnormal performance. A full measurement program covers system design, installation, operation, maintenance, evaluation of instruments and measurement systems for dams, appurtenant structures, and foundations. Instrumented monitoring includes measurements of displacement, strain, stress, pressure, loads on structural members, and seepage and drainage along with environmental factors that affect dam behavior such as temperatures, reservoir level, and precipitation. Data are collected and observations are made, processed, and evaluated by qualified personnel.

1.2 Dam Instrumentation objectives

The principal objectives of dam instrumentation may be generally grouped into three categories, 1) analytical assessment, 2) legal evaluation, 3) development and verification of future designs. A wide variety of instruments may be utilized in a comprehensive monitoring program to ensure that all critical conditions for a given project are covered adequately (USACE, 1995).

1) Analytical assessment: Analysis of data obtained from instruments can be used to verify design parameters, verify design assumptions and construction techniques, analyze adverse events, and verify apparent satisfactory performance as discussed below.

a) Verification of design parameters: Instrumentation may be utilized to verify design parameters with observations of actual performance, thereby enabling engineers to determine the suitability of the design, b) Verification of design assumptions and construction techniques: Experience has shown that most new or modified designs and construction techniques are not readily accepted until proven satisfactory on the basis of actual performance. Data obtained from instrumentation can aid in evaluating the suitability of new or modified designs, c) Analysis of adverse events: When a failure, a partial failure or a severe distress condition has occurred at a dam project, data from instrumentation can be extremely valuable in the determination of the specific cause or causes of the event. Also, instrumentation is often installed prior to, or during, remedial work at a site to determine the effectiveness of the improvements and the effect of the treatment on existing conditions, d) Verification of apparent satisfactory performance: Positive indications of satisfactory performance are very reassuring to evaluating engineers and operators of a dam project. Instrumentation data can prove to be valuable should some future variation in historic data occur, signaling a potential problem.

2) Legal evaluation: Valid instrumentation data can be valuable for potential litigation relative to construction claims. It can also be valuable for evaluation of subsequent claims relative to changed conditions. In many cases, damage claims arising from adverse events can be of such a magnitude that the cost of providing instrumentation is justified on this basis alone.

3) Development and Verification of future designs: Analysis of the performance of existing dams, and instrumentation data generated during operation, can be used to advance the state of the art of design and construction of dams.

Available instrumentation data has to be analyzed thoroughly since it is a main component of dam safety investigation. A wide variety of devices and procedures are used to monitor dams. The following features of dams and dam sites most often monitored by instruments: 1) Movements (horizontal, vertical, rotational and lateral), 2) Pore pressure and uplift pressures, 3) Water level and flow, 4) Seepage flow, 5) Water quality, 6) Temperature, 7) Crack and joint size, 8) Seismic activity, 9) Weather and precipitation, 10) Stress, and 11) Strain.

1.3 Motivation

In most countries throughout the world, interest in dam safety has risen significantly in recent years. Aging dams, new hydrologic information, dam construction and population growth in flood plain areas downstream from dams have resulted in an increased emphasis on dam safety, operation and maintenance. Historical data shows that the most prevalent category of potential failure modes for a concrete dam are those related to loss of foundation support for the dam. For both gravity and arch dams, adequate support from the supporting rock is essential to the structural integrity of the dam. Significant loss of this foundation induces stresses for which the dam is not designed. This leads to cracking of the dam, and potentially its failure. For arch dams, thrust support provided by the abutments is particularly crucial, given the high loadings transmitted to them (Veesart, 1997). The first phase of the dam safety process involves monitoring dams to identify potential deficiencies. Monitoring involves making periodic inspections and collections and evaluating instrumentation data.

In establishing an instrumentation program, it is important to understand the objectives of the program, the need for each type of instrument, the environment in which the instrument will be located, the difficulty in gathering the data, and the time and effort in reducing and understanding data generated. The wrong type of instrument may not measure the desired behavior. Reliability of instruments also has to be regularly checked. It is necessary to determine which of the instruments are reliable and which should be retained (Stateler et al., 1995).

Instrumentation data is often accumulated, but its engineering significance is not fully exploited in dam surveillance. In many instances the amount of effort put into analysis of data is small and out of proportion relative to the effort put in instrumentation of the dam and gathering the data. The output of dam monitoring system, which is a main part of dam surveillance, has to be thoroughly analyzed to alert dam wardens of any possible anomalies. The need for effective analysis tools of dam monitoring data was recently emphasized in the latest International Commission of Large Dams (Dibiagio, 2000). Dam monitoring practice has not been keeping pace with recent advances in statistical analysis methods. There is a need to develop new analysis tools to help dam safety engineers in the evaluation of the dam behavior.

A method, which can extract important features from the data, would be a useful tool in dam safety studies. Since structural responses of a dam are caused by the combination of several factors, the multivariate data analysis methods present several advantages: 1) it is cost effective by reducing the number of individual analyses, 2) it can separate the signal component from noise across a group of instruments given that the noise component is by definition uncorrelated from one instrument to another, 3) it can identify dominant patterns of behavior.

The definition of acceptable ranges for instrumentation readings can be used for immediate data review during data collection, so that anomalies can be quickly identified (Veesaert, 1997). It is recognized that trying to establish the range of expected monitoring data might be difficult in some cases. There is always a trade-off between setting the range of expected dam performance too narrowly where the limits may be exceeded frequently, and setting performance ranges too broadly where the danger is that adverse behavior could occur within the limits of so-called safe performance. When instrumentation data is not within pre-established limits, prompt evaluation of the safety of the dam should be undertaken which may lead to: 1) Assessing, and if needed, resetting the boundaries of satisfactory performance of the dam, as measured by instruments, 2) Heightened awareness of the condition of the dam and intensified monitoring, 3) Reducing the reservoir level, 4) Warning, and potentially evacuating downstream area and, 5) Taking structural corrective actions. Establishing alarm levels is an important part of dam safety programs. In dam safety practice these alarm levels are chosen by statistical analysis for each individual instrument. However, as some of the measurements are noisy or unreliable, this approach increases the chance of randomly finding an instrument out of control. The more variables there are, the more likely it is that one of these instruments may be out of control and indicate an adverse condition when the dam is actually in a safe state. Thus the false alarm rate (or probability of Type 1 error) is increased if each variable is analyzed and controlled separately.

1.4 Objectives

The principal objectives of this thesis are:

(1) Application of Multivariate analysis and neural networks for dam monitoring data exploration/analysis.

(2) The identification of structural anomalies, faulty instruments and readings from data

1.5 Scope and outline of the thesis

Chapter 2 presents a summary of previous research on statistical and numerical analysis methods for dam monitoring data methods used in other fields are also reviewed. Chapter 3 describes in detail the statistical analysis and neural network methods used in this thesis for the analysis of dam monitoring data. Sources of measurement errors are described in Chapter 4, followed by a description of dam instrumentation with an emphasis on concrete arch and gravity dams. In Chapter 5, the methods discussed in Chapter 4 are applied to data from the Idukki dam in India. In chapter 6, the methods are also applied to data from Daniel Johnson dam and to Chute-à-Caron dam. In Chapter 7, a summary of the thesis, and of the major contributions is presented, followed by suggestions for future research.

CHAPTER 2

2. Review of previous research

2.1 Introduction

Dam safety relies on a carefully planned surveillance program, which consists of stability checks, measurements and tests on materials. A key part of such a program is a visual examination of the dam complemented with monitoring data from the dam. A basic requirement of managing dam safety is monitoring of the structure in order to collect data, which are then interpreted to understand the state of the dam (Gresz, 1993).

Automatic instrumentation and data acquisition systems are used to monitor the real time behavior of dams. The output of monitoring systems is presented locally to dam wardens to alert them of possibly dangerous situations. Telemetry systems are used to send the information to a central database, where experts evaluate the status of the structure through the interpretation of data. Monitoring systems produce a large quantity of data, which has to be managed, and in case of automatic dam monitoring, managed in real time (Crépon et al., 1999).

The extent and nature of the instrumentation depends on the complexity of the dam, the size of the impoundment, and the potential for life and economic losses. A number of instruments on existing dams are initially installed to monitor the behavior of the dam during first impoundment and are often irrelevant for monitoring the behavior during its service life; others are incrementally installed as defects become apparent. A comprehensive analysis of data provided by these monitored devices is a valuable tool in dam safety investigation.

The need for effective analysis tools was recently emphasized in the latest International Commission of Large Dams (ICOLD) General Report (Dibiagio, 2000). Two of the biggest problems in monitoring have been data processing and presentation, and the analysis of data. A few decades ago, data gathering and processing was done manually. The time lag between measurements and interpretation was often very long. The introduction of computers and specialised software has more or less eliminated the first problem, but the second problem still exists. In many instances the amount of effort put into analysis and interpretation of data is small and out of proportion relative to the effort put in instrumentation of the dam and gathering the data. Performance data is often accumulated, but its engineering significance is not fully exploited.

When correctly installed and configured, automatic data acquisition systems with properly located geotechnical, structural, and hydrological sensors provide crucial information for operating the reservoir safely over the long term. When the data is scrutinized by human judgment on a regular and periodic basis, a well-maintained instrument automation system can provide a reliable performance database for the structure during its operation over several years. This information helps the owner to recognize an abnormal response under normal conditions, and to monitor the behavior during extreme hydrological or seismic events.

Two types of models have traditionally been used in dam monitoring, (1) statistical and (2) deterministic models. Statistical models are based on correlations between environmental factors (impounded water level, ambient temperatures, ice pressure) and dam responses (displacements, pressures, flows). These correlations are estimated by performing statistical analysis of historical data (Enel, 1980). Statistical models provide the answer to a basic question: Is the dam behaving as it did in the past? Statistical models are used to interpolate the response within the range observed historically and uncertainty and unpredicted responses with statistical models typically increased when extrapolating behind the range of values observed historically. Deterministic models can be used to model long-term non-reversible deformations to predict behavior at extremely low or high reservoir levels, at extreme temperatures, and during seismic events. This method is used to answer the following question: Is the dam behaving properly for given loading conditions? (Lombardi, 1999). A good agreement between statistical forecasts and measurements implies that the dam behaves as it did in the past.

2.2 Statistical Methods of Dam Monitoring

Various statistical procedures have been proposed for the analysis of monitoring data (Silva Gomes et al., 1985). A model of quantitative analysis is a functional relationship between observed effects and corresponding actions. These models rely on some basic assumptions: 1) The analyzed effects correspond to a period which the configuration of the structure remains the same. 2) The response of the dam can be separated in two parts, a) reversible effects due to the variation of hydrostatic level and air temperature, and b) irreversible effects which are function of time and can be induced by creep, alkali aggregation reaction, or other damage. A general statistical model for the response of an instrument can be formulated as follows:

$$D_i(t) = F_i(t) + G_i(H) + H_i(T) + \varepsilon_i \tag{2.1}$$

where F(t) is an irreversible response associated with consolidation, settling, degradation, or creep, G(H) is the response due to the hydrostatic level, H(T) is the response to temperature, and ε is the residual error. In many cases, thermal inertia creates a delayed response between temperature variation and instruments readings. Researchers have proposed various functions for the modeling different components of responses.

The hydrostatic-season-time model (HST), (Crépon et al., 1999). is a regression model, which takes into account the hydrostatic level as a fourth degree polynomial the seasonal effect, as a sum of four Sin functions, and T(t) is irreversible effects. Rain effect can also be included for modeling the piezometer readings. Least square criteria is used to estimate the model coefficients. The HST model is extremely robust and yields satisfactory results. It gives in a simple nonlinear function of water level and a periodic function (Figure 2.1), which is comparable to a delayed response to annual and half-yearly cycles in which the total extreme loads occur. One of the main gaps in this model is the lack of physical information provided by parameters (Bonelli et al., 2001). The polynomial expression for the effects of water level was originally based on a mechanical analysis of the water level on the displacements of an arch dam, based on resistance of materials.

$$D(t) = H(z) + S(\theta) + T(t) \tag{2.2}$$

$$H(z) = a_1 + a_2 z + a_3 z^2 + a_4 z^3 + a_5 z^4 \tag{2.3}$$

$$S(\theta) = a_6 Sin(\theta) + a_7 \cos(\theta) + a_8 Sin(\theta) Cos(\theta) + a_9 Sin^2(\theta) \tag{2.4}$$

$$T(t) = c_1 t' + c_2 t'^2 + c_3 t'^3 \tag{2.5}$$

where

$z = \dfrac{H - H_{min}}{H_{max} - H_{min}}$, $t' = t - t_0$, and $\theta = \dfrac{2\pi t'}{365}$

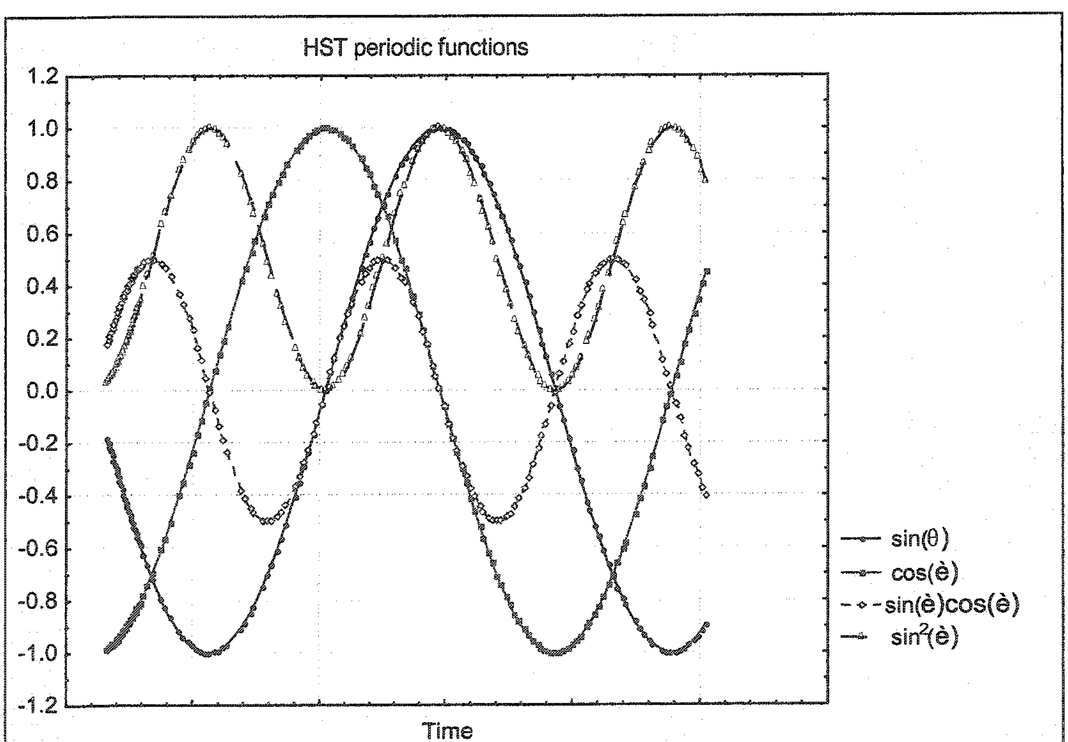


Figure 2.1, Periodic functions used in HST model

H_{min}, H_{max}, and t_0 are respectively the minimum and maximum reservoir water level, and the starting time for data collection considered for statistical analysis.

Enel (1980) use a statistical model to forecast the behavior of dams. The effect of water level, ambient temperature, and creep are considered. A non-linear model is used to describe the effect of reservoir level variation

$$F_k(H) = A_0 + A_1 H + A_2 H^2 ... + A_m H^m \qquad 2.6$$

where F_k (H) is a structural response (e.g. displacement) and H is reservoir water level. The order of the model is usually less than 5.

For estimating the thermal displacements two methods are proposed. The displacements can either be formulated as:

$$E_k(t) = l_1(t)e_{k1} + l_2(t)e_{k2} + ... + l_n(t)e_{kn} \qquad 2.7$$

Where e_{k1}, e_{k2}, ... , e_{kn}, are unknown coefficients and $l_1(t)$,..., $l_n(t)$ are measured concrete temperatures. If no temperature measurements are available, thermal displacements can be formulated as:

$$E_k(t) = E_{11} \sin \omega t + E_{12} \cos \omega t + ... + E_{p1} \sin p\omega t + E_{p2} \cos p\omega t \qquad 2.8$$

where $\omega = \frac{2\pi}{T}$, E_{11}, E_{12}, ... , E_{p2} are some unknown coefficients, and T is the period usually equal to 1 year. Irreversible displacements can be described by a combination of exponential and polynomial functions of time such as:

$$G_k(t) = a_1 e^{k_1 t} + a_2 e^{-k_2 t} + a_3 t \qquad 2.9$$

where a_1, a_2, a_3, k_1, k_2 are some unknown coefficients. The total displacement results from the superposition of the three different types of displacements.

$$\delta_k(t) = E_k(t) + F_k(t) + G_k(t) \qquad 2.10$$

The observed displacements are denoted by Δ_k (t). The residuals $\epsilon(t)$ are

$$\varepsilon(t) = \Delta_k(t) - \delta_k(t) \qquad 2.11$$

Unknown parameters are estimated by least squares. After the estimation of the model parameters the time series of residual $\varepsilon(t)$ is obtained. The Variance of the residuals σ is used to define confidence intervals. It is assumed that residuals follow a normal distribution, therefore, probabilities of $\varepsilon(t) \leq n\sigma$ can be calculated. For example for n=2, it can be assumed that the measurements must be between values of $\delta_k(t)-2\sigma$ and $\delta_k(t)+2\sigma$ with probability of 95%.

Guedes et al. (1985) use linear regression to relate individual dam instrument readings to reservoir, thermal, and time effects. This method is used to separate the effects, to determine whether or not these factors are independent of each other and to determine the best empirical equations.

Kalkani (1989) uses polynomial regression to monitor individual piezometers of the Kremasta embankment dam located in Greece. A portion of the data was used for estimating the relationship between reservoir level and piezometer levels, which was then used to forecast the piezometer levels for another portion of the data set and calculate the mean square of the forecasting errors. Separate models were used for predicting observation for increasing or decreasing reservoir level. Measurements higher than the predicted values plus one standard deviation were used to detect increased seepage through the dam, while measurements lower than the predicted values minus one standard deviation were used as an indication of malfunction of the piezometer. Temperature, time and rain effects were not considered.

Gilg et al. (1982) describe results of data analysis for three Swiss dams. The Mauvesian dam, is an arch dam of 237 m height that shows that the daily oscillation of the air temperature penetrates to a depth of 50-60 cm and 80 cm respectively in the downstream and upstream phases. At a depth of 3 m, the yearly variation of temperature is only 5-6°C, which is equal to about 1/3 of the variation of the monthly average ambient temperature. At a depth of 15 m the variation of temperature is only 0.5-1°C and the average temperature is higher than the average annual ambient temperature. This was attributed to the influence of solar radiation. Time lags of 0.5, 2 and 5 months are observed between concrete temperatures and ambient temperature at depths of 1, 3 and 15 m respectively. A lag of one month is observed between the maximum of reservoir level and the maximum deformation, which is explained by the temperature effect. An irreversible displacement of 18 mm was recorded during the 17 years of observation. Uplift pressures show that there is no water pressure in the upstream part of the central blocks, which shows that the grout curtain is very effective.

Goguel et al. (1992) describe the instrumentation of Kariba dam, a 128-m high arch dam in Kenya. Strain meters show a continuous drift, due to creep and shrinkage effects. The maximum creep rate is 23 micron/meter/year. Scanning of concrete samples with an electron microscope detected the presence of an expansive gel typical of alkali aggregate reaction.

Blas (1989) describes a methodology used in the analysis of an arch dam that was exhibiting moderate irreversible upstream displacements. A statistical model was developed to estimate the irreversible radial components of displacements at the top of three dam blocks. Three different time variables are used to capture trends as a piecewise linear function. The other variables considered in the model are average air temperature over the previous eight weeks, temperature on the day of the observation, and the water level.

The order of importance of the variables is time, averaged air temperature, reservoir level, and daily air temperature. Higher orders of the reservoir level were rejected because reservoir level oscillations were very small during the time interval considered. Irreversible displacements of up to 2mm /year were observed. A finite element analysis of the dam was performed to consider: 1) the non-linear behavior of the material, 2) a time-dependant volumetric expansion due to water penetration, and 3) representation of construction joints. It was concluded that swelling was the main reason for observed anomalies.

Hulea et al. (2000) describe statistical and deterministic models used for monitoring the Tarnita arch dam. Crest displacements were almost 60% larger than predicted displacements (70 mm versus 45 mm). However, the dam structure did not show any significant signs of deterioration. Different functions were used for estimating the temperature effect in statistical models. The analysis performed with statistical models showed that 1) alarm level or threshold for monitored data cannot be established at the design stage. A dynamic process has to be implemented that allows for updating the acceptable limits in terms of the evolution of dam behavior, 2) the best model was based on measured air temperatures as opposed to a HST model.

Chouinard et al. (1995) apply principal component analysis (PCA) to estimate the principal modes of deformation of a dam from a historical record of instruments. The PCA was applied independently to two groups of instruments, one for data from stress meters and the second for data from instrumented cylinders. The correlations between the scores of the principal components and factors such as reservoir water level, ambient temperature, and time were analyzed.

Comité Suisse des barrages (2000) describe a method "measured-calculated" for modeling dam behavior and detecting anomalies. The method consists of the following steps, 1) monitor and model dam behavior through instruments, 2) calculate the same quantities through numerical models, 3) compare the predictions and measure values. There is no restriction for the type of dam to which the method can be applied. In Switzerland it has been applied to both arch and gravity dams. Radial displacements (upstream-downstream) are generally used as response variables but the method could also be applied to tangential displacements, pressure meters, joint movement and other quantities. Reservoir level, concrete temperature and age are the variables considered. The general model can be written as:

$$R(t,env) = P(t,env) + D(t,env) \tag{2.12}$$

in which R is observed value at time t and environmental conditions of env. P is the estimated value and D is the error accounting for modeling and observation errors. Three different methods are used for estimating the observed values, 1) statistical models, 2) numerical models, and 3) hybrid models, which combine the first two methods where some parameters of the numerical model are optimized using measured values. The statistical method is very similar to HST models. Table 2.1 presents the relationship between these different methods.

Bonelli et al. (2001) describe a model for performing delay analysis on pore pressure measurements. The method is based on Darcy's law and Richard's equation of seepage

Table 2.1, Relation between different Methods

		Modeling functions			
		Empirical	By Modeling		
Calibration of parameters	No		Deterministic	No	Statistical Treatment
	Yes	Statistical	Hybrid	Yes	
		No	Yes		
		Modeling of the dam			

and involves the use of a linear dynamic system accounting for the contribution of non-ageing factors. Delayed effects are due to dissipative behaviour (viscoelasticity, seepage, etc.), and are therefore irreversible. The delayed effect of water level was calculated based on convolution of the impulse response of the dam structure and water level.

Bourdarot (2001) presents a simplified method for analysing the magnitude of the observed deformation and the deformation patterns in Arch dams. After an initial phase following the reservoir impounding, during which the irreversible displacements toward the downstream are observed, an inverse evolution of dams towards the upstream is observed after 30 to 50 years of operation. Simple finite element models were developed to illustrate the different pattern of irreversible effects due to shrinkage and foundation settlement. Displacement rates are estimated from actual case histories and applied to an arch dam. Creep, shrinkage, and settlement lead to a beneficial compression on the upstream face countering the effect of the reservoir level. They also cause tensile stresses at downstream base of the dam, which sometimes cannot be balanced at low reservoir levels. The model for shrinkage and creep is very simple, stress dependencies and loading history are not taken into account.

Paxton (2001) describes the structural monitoring system for Milliken dam. A spalled block from the downstream face of the dam in the 1950's led to the installation of Carlson resistance wire joint meters to monitor changes in six of the cracked or separated lift joints. Joint meter readings indicated that crack opening mostly occurs during the last 6 m of reservoir filing. The monitoring system was improved in 1998 by measuring crest deformation, installing crack meters, and replacing the Carlson joint meters. It was concluded that deformations resulting from static loading conditions have stabilized and do not threaten the safe operation of the dam.

Crepon et al. (1999) provide a description of Monitor, a software based on the HST model, developed for dam monitoring data analysis. The database of the program includes: description of the dam, location and description of measuring instruments, special functions such as calculation of physical measurements from raw measurements, and selection of measurements intervals. The software gives a spatial representation of measured quantities, and has built-in functions for estimating the derivatives and integrals of original variables. These new derived variables can be used as explanatory variables and improve the model precision. Considering the accumulated rainfall over a ten-day period and the speed at which the reservoir changes, results in significant improvement for modeling of piezometer levels.

The statistical models discussed analyse relationships between environmental factors (impounded water level, ambient temperatures) and dam performance (i.e. displacements, pressure, flow). Statistical methods are based on the analysis of past behaviour of the dam, which is expected to remain the same during normal loading condition. Another approach is numerical analysis of the dam based on information on loads, properties of materials, and physical laws governing the stress-strain relationship. One advantage of a numerical model is that it can predict the response of a dam to extreme effects such as floods and earthquakes. Some of numerical analysis methods applied in dam monitoring are discussed next.

2.3 Numerical analysis

Arch dams (Figure 2.2) rely significantly on arch action to transfer horizontal loads to the abutements. Arch dams may be divided, according to the geometry of their cross sections, into thin, moderately thin, and thick arch sections. Table 2.2 identifies each of these types with regard to crest thickness (t_c) and base thickness (t_b), each

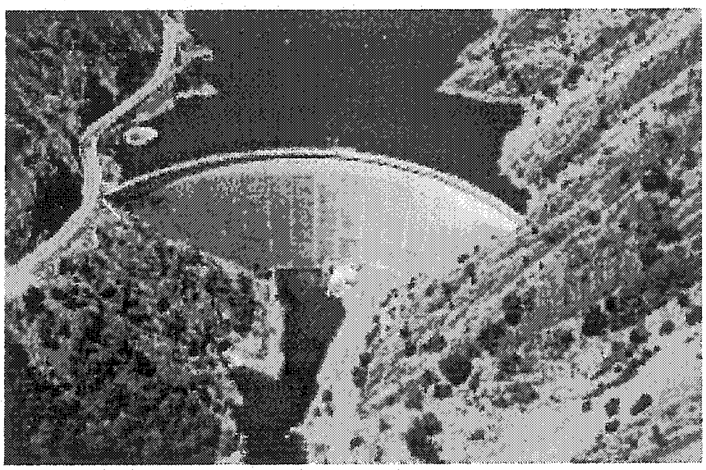


Figure 2.2, A top view of an Arch dam

expressed as a ratio to the height (H), and the ratio of base-to-crest thickness. Under static loads, a well-designed arch dam should develop essentially compressive stresses, which are significantly less than the compressive strength of the concrete. However, the analyses of monolithic arch dams with empty reservoirs, with low water levels, or with severe low temperatures have indicated that zones of horizontal tensile stresses can develop in the dam on the upstream and downstream dam faces (USACE, 1994). Although concrete can resist a limited amount of tensile stress, it is important to keep tension to a minimum so that the arch has sufficient reserve strength if subjected to additional loads.

When the design limits are reached or, as in the case of many existing dams, when the dam is not designed for severe loading conditions, some cracking can occur at the base and near the abutments. These horizontal stresses tend to open the vertical contraction joints, which are expected to have little or no tensile strength. It is apparent that joint opening will relieve any indicated arch tensile stresses, and the corresponding loads can be redistributed to cantilever action provided that tensile arch stresses are limited to only a small portion of the dam.

Table 2.2, Arch dam Types (USACE, 1994)

	t_c/H	t_b/H	t_b/t_c
Thin arch	0.025-0.05	0.09-0.25	2.9-5
Moderately thin	0.025-0.05	0.25-0.40	5-10
Thick gravity-arch	0.05-0.10	0.5-1.0	8-15

The need for an acceptably accurate method of analyzing arch dams led to the development of the trial load method based on structural mechanics concepts in the 1960's. The arch dam is decomposed into a series of horizontal arches and vertical cantilevers. The trial load method is based on the assumption that the hydrostatic load is divided between cantilever and arch elements in a proportion that results in equal arch and cantilever deflections at all points. This method can neither fully represent the solid body of an arch dam nor reflects the effect of foundation. Since the 1970's the linear finite element method (LFEM) has been employed for stress analysis of dams. The finite element method can be applied to complex geometries, and can accommodate variations in material properties whitin the model. Finite element methods have been used for both static and dynamic analysis of dams. Some of the applications for analysis of dams are reviewed next.

Veltrop et al. (1990) develop a finite element model of one of the arches of the Daniel Johnson multiple arch dam. The analysis consists of a 2D transient heat flow analysis and a 3D-stress analysis. The results of the 2D-heat flow analysis were used to define the critical temperature gradient, which then was applied as part of the loading for a 3D-stress analysis. The loading conditions considered were the hydrostatic and dead load and winter thermal load conditions. It was concluded that the first set of cracks that appeared on the arches was due to geometric discontinuities and that the second set of cracks was formed because of winter load conditions.

Léger et al. (1993) describe a finite element modeling procedure for obtaining seasonal temperature and stress distributions in concrete gravity dams. Effects of the reservoir, foundation, and air temperature distributions, and heat supply from solar radiation on the thermal response of a dam are discussed. Two separate analyses are performed, a thermal analysis of a dam, to define the input, followed by the stress analysis of the system. Parametric analyses are performed to evaluate the effects of geometrical, thermal and mechanical properties; reservoir, air and foundation temperature variations, and heat supply from the sun on the thermal and mechanical behavior of gravity dams. Some of main conclusions were: 1) displacements occur when the mean temperature of the top dam section is lowest, 2) the daily air temperatures greatly effects the surface thermal tensile stresses; when actual daily air temperatures are used, the maximum surface stresses increase by a factor of 1.5 to 2, as compared to a model based on average daily temperatures over 22 years period, 3) solar radiation has a significant effect on the depth of frost penetration while its effects on stresses are negligible, 4) the height of the dam has little effect on the depth of frost penetration, 5) maximum crest displacements occur when the mean temperature of the upper section is lowest.

Bouzoubaâ et al. (1997) investigate the effects of external temperature variations on mass concrete through laboratory experiments and finite element analysis. A concrete block, instrumented with thermocouples and vibrating wire extensometers and exposed to temperature variations on one face, was used to simulate the behavior of the downstream face of a concrete gravity dam exposed to thermal cycles. A finite element model for thermal analysis was developed and results were compared with the experiments. The validated numerical model was then used to study the effects of the variation of outside temperatures on the behavior of gravity dam. Some of the main conclusions were that 1) when the effect solar radiation is not included, the depth of frost penetration can be overestimated, 2) the maximum principal tensile stress occurs in coldest month of the year at the toe of the dam near the downstream face.

Zhang et al. (1997) discuss the effect of the initial temperature and convection coefficients in transient thermal analysis of massive concrete structures. Determination of a precise convective coefficient for the heat exchanged between concrete and air is complicated. A finite element model of a generating station structure was developed and results were compared with measurements. Different uniform initial temperatures were considered. The following conclusions were made: 1) the heat transfer behavior of a dam is not sensitive to the variation in the heat convection coefficient when the coefficients are within a certain range, 2) no matter what the assumed initial temperature is, after 9-10 months the results of numerical model coincide with measurement data.

Pedro et al. (1985) evaluate the safety of a Portuguese cracked arch dam in Portugal. The dam is 130 m in height with a thickness varying from 30 m at its base to 4.6 m at its crest, and a span of 60 m. Horizontal cracks developed at the downstream face of the dam near the crest. Triangular flat shell elements are used to model the arch dam. Two different load cases are considered: 1) dead load and hydrostatic load and, 2) the dead load, hydrostatic load, and loads due to temperature variations. It was concluded that the cracking at the downstream face of the dam did not significantly affect its stability. The recommendation was that the reliability of a numerical model should be evaluated by comparison of the results with observations under normal operating conditions (and interpretation of eventual incidents), and by analysis of incidents that occurred in past.

Tahmazian et al. (1989) investigate the stability of the Daniel Johnson multiple arch dam. A three-dimensional non-linear finite element model using the smeared crack technique was developed to reproduce the observed behavior of the dam, to assess its safety, and to help evaluate scenarios for remedial work. First, the winter temperature distribution in the dam was calculated using transient heat flow finite element analysis. The winter temperature was used in structural analysis in addition to water load and dead load. The conclusions were that the finite element model successfully reproduced the observed stresses and deflections, and that observed crack patterns and the thermal cracking did not significantly affect the load carrying capacity of the structure.

Barrie (1995) studies the safety of Gerber arch dam. Gerber dam, a 26-m high variable radius thin arch structure, has experienced seepage and extensive freeze-thaw damage since its construction. On several occasions since 1951, the upstream face of the arch dam has been treated with waterproof membranes to prevent seepage. Inspection of the structure indicated that the treatment is deteriorated and should be considered ineffective. Since the last treatment in 1973, seepage has been reported between lifts. A three dimensional finite element model was used to evaluate the safety of the arch structure. The loading combination of water, dead load and winter temperatures were found to be critical. Winter temperatures contract the arch and displace it in the downstream direction. Tension was reported to be greater on the upstream side. The maximum compressive and tensile stresses were reported to be 2.6 MPa and 2.0 MPa respectively. It was concluded that concrete is not expected to crack under its service load and that the resulting net tensile strength of the joints may be loosened by tensile stress, which increase seepage rates.

Lan et al. (1997) describe a Non-Linear Finite Element model (NLFEM) of an arch dam. The non-linear stress-strain relationship and cracking behaviour are considered. Two kinds of cracking models are usually employed in a non-linear finite element model of concrete structures, which are the discrete cracking model and the smeared cracking model. The discrete cracking model is set up between two adjacent element surfaces. It can model the occurrence and propagation of the tensile cracks in the structure and estimate the crack depth. However, the analysed structure has to be re-meshed when the cracks occur and propagate, which leads to more computation cost. The smeared cracking mode assumes various continuous parallel cracks. This approach can more efficiently address the cracking phenomena. Employing the smeared crack criteria a model of a 250 m high arch dam was developed under the loading condition: normal water pressure, dead load and rising temperature change. The maximum displacement in the arch dam was 157.5 mm for the linear model and 167.7 mm for a (NLFEM), which shows a difference of 6.5%. The maximum difference between two the models was 24.3% at the bottom of the arch dam, due to initial cracking in this region under the working loading conditions. The authors believe that (NLFEM) results are more representative of dam behaviour as the non-linear nature of cracking are properly addressed.

In conclusion, statistical and deterministic methods have been used for modeling dam behavior. In statistical methods, the most common approach in the analysis of dam monitoring data is to proceed at the level of individual instruments. In the presence of noisy data, it can be very difficult to identify significant deviations from normal readings. A procedure that minimizes the effect of noise from individual instruments is to perform simultaneous analysis on several instruments. Successful applications for simultaneous analysis of multiple instruments have been reported in many other fields using multivariate statistical analysis methods and artificial neural networks (ANN). Some of the most relevant studies are reviewed next.

2.4 Multivariate statistical Methods

Multivariate statistical methods are used extensively in chemometrics (MacGregor et al., 1994). Computerized data acquisition systems are routinely utilized to collect real-time data from a multitude of sensors every few seconds in chemical processing plants. Traditionally, operating personnel had been using only a few measurements to monitor the performance of a processing plant.

Kresta et al. (1991) propose a multivariate statistical process control for simultaneously analyzing several process and quality variables. Multivariate statistical procedures (PCA and PLS) are used to reduce the dimensionality of a large and highly correlated data set down to a few factors or components, which contain most of the information about the process under normal operating conditions. The scores of component variables are plotted as a function of time to detect large deviations from normal operating conditions. Plots of the squared errors of prediction are also used to detect major changes in the normal operating condition.

Nomikos et al. (1994) use principal component analysis to extract the information from all the measured process variables, and to project it onto a lower dimensional space defined by the latent variables or principal components. Analysis of process batches is used to classify similar batches by examining the clusters of their projections into an hyperplane. The approach is based on basic statistical process control (SPC) concepts, whereby the performance of a process is assessed by comparing it with past measurements when the process was operating well, and was in control. Control limits for the monitoring charts are derived from statistical properties of the reference data set.

Sun (1996) describes a multivariate regression procedure based on principal component regression. The method corresponds to a simultaneous analysis of several response variables of interest and is referred to as Multivariate Principal Component Regression (MPCR). PCR works on only one response variable. When there is more than one response variable of interest, one way to apply PCR is to analyse each response variable separately using PCR. It is apparent that this approach cannot use the correlation information of the response variables. In MPCR, principal components of response variables and independent variables are calculated, and then regression analysis is used for regressing the principal components of response variables on principal components of independent variables.

2.5 Artificial Neural Networks

Principal component analysis is now widely used for reducing the dimensionality of data set and to obtain a better understanding of processes (Martin et al., 1996). However, the linearity assumption inherent in conventional PCA can lead to misleading conclusions in the analysis of data from highly non-linear processes. Conventional PCA is not effective when the variables are nonlinearly related and in such situations nonlinear principal component analysis (NLPCA) is more appropriate. Nonlinear principal components analysis can be used in a similar way to PCA, that is data summarization, data visualization and data exploration. Neural networks have been applied for extracting both linear and nonlinear principal components from data. The concept of extracting features from highly nonlinear data has been discussed in a number of studies, most of the techniques are based on artificial neural networks (Dong et al., 1996). Diamantras et al. (1996) provide a good review of PCA neural networks.

Fan et al. (1993) present an approach to fault diagnosis of chemical processes during steady-state operations by using artificial neural networks (ANN). The authors indicate that back-propagation networks can learn and perform nonlinear mapping only to a certain extent. For the case of fault detection in chemical processes, nonlinearities can be very complex, especially in the case of multiple faults. They modify the conventional back-propagation ANN by the addition of a number of functional units to the input layer.

Kramer (1991) uses auto-associative neural networks for nonlinear principal component analysis (NLPCA). The Auto-associative neural networks with a bottleneck layer of nodes can be used to reduce the number of input variables. The network is called auto-associative neural network, as it must reproduce the input at the output layer. The network has three layers, with p nodes in the input and output layers and m nodes in the bottleneck layer. Since the dimension of the bottleneck layer is smaller than both input and output layers, the network is forced to develop a compact representation of input data. NLPCA was used to identify and remove correlations among variables as an aid to reduce dimensionality, visualize the data, and for exploratory data analysis. While PCA identifies only linear correlations between variables, NLPCA uncovers both linear and nonlinear relations.

CHAPTER 3

3. Methodology

3.1 Introduction

The traditional approach in dam monitoring is to analyse the response of individual instruments and to set thresholds on the observed values to trigger alarms. However, it can be very difficult to estimate statistically significant deviations from normal readings for individual instruments, given that the fluctuations in stresses, strains, or deformations are small and in the order of magnitude of noise in the measurements. In addition, these are subjected to the simultaneous effects of the water level fluctuations and temperature, which are often highly correlated.

Traditionally, engineers have relied mainly on instruments that integrate strains over large volumes of the dam, such as inverted pendulums, or targets. Measurements from these instruments tend to be less variable due to averaging; however, when significant deviations are detected, damage and deterioration is usually already in an advanced stage, and it may be difficult to fully analyse the problem based solely on these instruments.

For that purpose, data from stress and strain meters can provide useful additional information. The desirable features of analysis procedures for this type of data are that: (1) they should make use of all instruments simultaneously, and (2) they should separate signal from noise. When simultaneous readings are available for different instruments across a structure, estimates of correlation between these instruments can be used to identify the major patterns of deformation of the facility under a variety of external actions. Correlation measures the degree of linear dependency between the variables, which is usually a valid assumption for the behavior of dams under normal operating conditions.

The response of any instrument results from a combination of several reversible or irreversible effects. Irreversible effects are usually associated with time-dependent phenomena such as creep, swelling, and settlements. These phenomena are usually most critical from the point of view of dam safety and it is desirable to monitor their rate of progress. Reversible effects are usually not critical from dam safety point of view and are associated with fluctuations of the reservoir water level and temperature. To estimate a realistic model of the dam the data set should cover as much as possible all the anticipated operational conditions of the dam.

3.1.1 Data processing and presentation

The usefulness of any observation depends strongly on the care with which the calibration and subsequent data processing are carried out. Once the data are collected, further processing is required to check for the errors and to remove erroneous values. Two types of errors must be considered in the editing stage: (1) large "accidental" errors or "spikes" that result from equipment failure or other major data flow disruptions; and (2) small random errors or "noise" that arise from changes in sensor configuration, electrical and environmental noise, and unresolved environmental variability. The noise can be treated using statistical methods while elimination of the larger errors generally requires the use of some subjective evaluation procedure. Data summary diagrams or distributions are useful in identifying large errors as sharp deviations from the general population. By not directly examining the data points in conjunction with adjacent values, one can never be sure that reliable values are thrown away.

Missing data or gaps are observed in many engineering data records. Missing data is frequently the consequence of uneven sampling (in time and/or space), or may result from removal of erroneous values during editing and from sporadic recording system failures, major difficulties arise if the length of the holes exceeds 20-30% (Struges, 1983). Most analysis methods require data values that are regularly spaced in time/space. As a consequence, it is sometimes necessary to use interpolation/estimation procedures to create the required regular set of data values as part of data processing. The analysis of data records necessitates some form of " first look" visual display. Even the editing and processing of the data typically requires a display stage. Plotting time series of temperature, reservoir level, and instrumentation data, scatter plots of different observed data are to be considered in visual data representation. With the advent of the computer and electronic data collection methods, the knowledge of statistical methods has become essential to any reliable interpretation of results.

3.2 Multivariate Analysis of the data

Multivariate analysis is concerned with the empirical analysis of data that is a function of several independent variables. Multivariate calibration designates procedures used to describe how measurements on predictor variables $\underline{X}_1, \underline{X}_2, \ldots, \underline{X}_p$ are related to some target variables $\underline{Y}_1, \underline{Y}_2, \ldots, \underline{Y}_m$. The matrix $\underline{Y}$ (m×n) is formed from n observations on m responses (stresses, strains, displacement, flow) at the same time, where the i^{th} column is the observation vector $\underline{Y}_i$ at time t_i. The matrix $\underline{X}$ (p×n) is formed of observations on p predictor variables (air temperature, concrete temperatures, reservoir level, time) measured at the same time. These methods are applied to a set of simultaneous observations to determine the relationship between a set of dependent variables and a set of independent variables, and to make predictions through extrapolation of available data. The multivariate analysis methods considered in this thesis are reviewed next.

3.2.1 Multiple Linear Regression

Multiple linear regression (MLR) is the most widely applied technique for describing relationships between variables. It is used to describe the relationship between a dependent (response) variable Y and one or more independent (predictor) variables X_1, X_2 ,..., X_p. The relationship can be expressed as:

$$\underline{Y} = \underline{X}.\underline{b} + \underline{\varepsilon} \qquad 3.1$$

Where $\underline{b}$ is the vector of regression coefficients and $\underline{\epsilon}$ is the vector of residuals. The least square method is used to minimize the sum of the squared residuals

$$\underline{\varepsilon}'.\underline{\varepsilon} = (\underline{Y} - \underline{X}.\underline{b})'(\underline{Y} - \underline{X}.\underline{b}) \qquad 3.2$$

and the estimate of $\underline{b}$ is:

$$\hat{\underline{b}} = (\underline{X}'.\underline{X})^{-1}(\underline{X}'\underline{Y}) \qquad 3.3$$

Therefore an estimate of $\underline{Y}$ can be expressed as:

$$\hat{\underline{Y}} = \underline{X}.(\underline{X}'.\underline{X})^{-1}(\underline{X}'.\underline{Y}) \qquad 3.4$$

which corresponds to the projection of $\underline{Y}$ on the $\underline{X}$ space. The standard error of estimate is expressed as:

$$S_{y,x} = \sqrt{\frac{\hat{\underline{Y}}'\hat{\underline{Y}} - \underline{b}'\underline{X}'\underline{Y}}{n-p-1}} \qquad 3.5$$

The regression model can be used to predict future observations on the response y_0 corresponding to values of the p predictor variable ($\underline{x}_0$) as $\hat{y}(\underline{x}_0) = \underline{x}'_0\underline{b}$. A 100(1-α)% prediction interval for this future observation is:

$$\hat{y}(\underline{x}_0) - t_{\alpha/2,n-p}(\sqrt{S_{y,x}^2(1+\underline{x}'_0(\underline{X}'\underline{X})^{-1}\underline{x}_0)} \leq y_0 \leq \hat{y}(\underline{x}_0) + t_{\alpha/2,n-p}(\sqrt{S_{y,x}^2(1+\underline{x}'_0(\underline{X}'\underline{X})^{-1}\underline{x}_0)}$$

3.6

where n and p are the number of observations and predictor variables respectively. Prediction intervals can be used to set alarm levels for different response variables. In predicting new observations care must be taken about extrapolating beyond the region covered by the sample. It is possible that a model that fits well in the sample data cannot predict accurately responses outside of that region.

A number of problems can also occur when some of the independent variables are highly correlated (Jackson, 1991). This situation, which is called multicollinearity, is characterized by columns in $\underline{X}$ that are approximately or exactly linearly dependent. Multicollinearity causes several problems:

1) The inverse of $\underline{X}'\underline{X}$ may be difficult to obtain, since the matrix in nearly singular.
2) The regression coefficients are highly correlated, and the interpretation of these coefficients is unreliable. Sequential procedures such as forward selection and backward elimination can be used to mitigate these problems.

The ordinary least squares (OLS) estimation principle, assumes that the variance on the residuals is constant. If variance of residuals is not constant; weighted least squares (WLS) should be used. In this case, the weighted sum of squared residuals is minimized. The estimate of $\underline{b}$ is:

$$\hat{\underline{b}} = \left(\underline{X}'\underline{V}^{-1}\underline{X}\right)^{-1} \underline{X}'\underline{V}^{-1}\underline{Y} \qquad 3.7$$

where $\underline{V}$ is the priori estimates of the uncertainty variances of observations.

In situations when there is a high degree of correlation among the predictor variables, multivariate regression techniques based on latent variables are used as the preferred method (Martens et al., 1989). These procedures select a few latent variables, which are a linear function of the original variables and used to forecast the response variables. Some of these methods, Principal Component Analysis (PCA), Partial Least Square Regression (PLSR) and Canonical Correlation Analysis (CCA) will be discussed in the following.

3.2.2 Principal Component Analysis (PCA)

PCA is a statistical technique falling under the general title of factor analysis. PCA is concerned with explaining the variance-covariance structure of a data set through a few linear combinations of the original variables. The purpose of PCA is to identify the dependence structure of multivariate observations in order to obtain a compact representation. The general objectives are (1) data reduction and (2) data interpretation. The analysis identifies characteristic and uncorrelated modes of variation of the variables.

Attractive features of this operation are that: (1) it eliminates correlation among the variables, (2) it is efficient method of compressing the data.

In PCA, the original variables are transferred into new, uncorrelated variables called the principal components or factors. Each principal component is a linear combination of the original variables. One measure of the amount of the information conveyed by each principal component is its variance. For this reason the principal components are arranged in order of decreasing variance. When the observed variables are correlated, the number of variables can be reduced without losing much of the information. This objective can be achieved by selecting only the first few principal components.

PCA is applied to either the correlation matrix ($\underline{R}$) or the covariance matrix ($\underline{S}$) of the original variables. PC's are obtained from the solution of the eigenvalue problem:

$$(\underline{R} - \lambda \underline{I})\underline{p} = \underline{0} \qquad 3.8$$

where $\underline{I}$ is the identity matrix of order m. A number of different numerical algorithms can be used to compute the eigenvectors, and eigenvalues (Martens et al., 1989). Solving for Eq. 3.8 results in a set of eigenvalues $\lambda_j (j = 1 : m)$, which can be placed as the elements of a diagonal matrix $\underline{\Lambda}$, and a corresponding set of vectors $\underline{p}_j (j = 1 : m)$. The solution is the vector $\underline{p}$ with maximum resemblance to all observations.

Principal Components (PC's) or scores $\underline{f}_i (i = 1 : m)$ are linear combinations of the variables, where the weights on each variable are given by the eigenvectors. The percentage of variance explained by each principal component is equal to its associated eigenvalue. The percentage of variance explained by the fist k principal components can be expressed by:

$$\frac{\sum_{j=1}^{k} \lambda_j}{\sum_{j=1}^{m} \lambda_j} \qquad 3.9$$

If the purpose of the analysis is data reduction, then retaining only the first k components will suffice. The number of components to be retained can be determined by:

1) Percentage of variance: The number of components to retain can be based on the percentage of variance explained for. The number of components is chosen to explain a relatively high percentage, say 70%-90%.

2) Average eigenvalue: Those components whose eigenvalue is greater than the average eigenvalue, which is also the average variance of the variables, are retained. Therefore, for a correlation matrix, only components whose variances are greater than unity are retained. The average eigenvalue method often works well in practice. Previous studies have shown that this method is fairly accurate when the number of original variables is <30 and the variables are rather highly correlated (Rencher, 1998).

3) Scree graph: Eigenvalues are plotted as a function of the number of eigenvalues. The number of components is selected where the scree graph flattens out.

The original data ($\underline{Y}$) can be reconstructed by using the first k principal components

$$\underline{Y} = \underline{f_1}\,\underline{p_1}' + \underline{f_2}\,\underline{p_2}' + \ldots + \underline{f_k}\,\underline{p_k}' + \underline{E} \qquad 3.10$$

where $\underline{E}$ is the residual matrix, $\underline{f_i}$ are scores or principal components, and $\underline{p_i}$ are eigenvectors (loading vectors). The residual matrix contains that part of the data not explained by the PCA model and most likely represents the noise in the data. The method is useful for separating signal from noise since random noise components are usually uncorrelated and are associated with lower principal components.

These first k components $\underline{f}_1$, $\underline{f}_2$, $\underline{f}_3$,...,$\underline{f}_k$ explain a greater percent of the data variance than the first k terms on any other expansion. PCA can be performed on the correlation $\underline{R}$ or covariance matrix $\underline{S}$ of observations. The components extracted from the covariance matrix are not the same as those found by analysing the correlation matrix. If different types of measurements are considered (displacement, flow rates, stresses), then the structure of PC's derived from the covariance matrix will depend essentially on the type of units of the measurements. If there are large differences between the variances of the variables, those variables whose variances are large will tend to dominate the first few principal components. Variables are typically standardized if they are measured on scales with widely differing ranges or if the measurement units are dissimilar.

In dam monitoring, the number of predictor variables is generally limited and is far less than the number of responses (different instruments). Therefore, the main application of PCA in dam monitoring is when it is applied to responses for data reduction.

PCA method does not take into account the relation between predictors and response variables during the decomposition process. PCA can be applied either to predictors or responses. There are occasions where PCA has been used for both predictor and response variables. In this case, the PC's for the predictors are obtained in one operation, and the PC's for the responses in another. The PC's of responses are then regressed against the PC's of the predictors (Jackson, 1991).

Alternatively, both response and predictor variables can be considered during factor calculations. This leads to another calibration method, Partial Least Squares (PLS), which is explained in the following section. Conventional PCA is not effective when the relations between variables are non-linear. In such situations non-linear principal component analysis (NLPCA) is more appropriate. NLPCA can be used in a similar way to PCA that is for data visualisation, data reduction and data exploration. The techniques for extracting non-linear principal components are based on Artificial Neural Networks (ANNs), which will be described in section 3.3.

3.2.3 Partial Least Square (PLS)

Partial least Squares (PLS), also known as Projection to Latent Structures, is a dimensionality reduction technique for maximizing the covariance between the predictor (independent) matrix $\underline{X}$ and the response (dependent) matrix $\underline{Y}$. Partial least square (PLS) differs from PCA by using both the dependent and independent variables actively during the decomposition process (Martens et al., 1989). The algorithm used in PLS examines both $\underline{X}$ and $\underline{Y}$ and extracts components, which are directly relevant to both sets of variables.

If the model $\underline{Y}=f(\underline{X})$ is considered, the objective is to model $\underline{X}$ in such a way that $\underline{Y}$ can be predicted as well as possible. It can be described by the following equations. The matrix $\underline{X}$ is decomposed into a score matrix $\underline{T}$, a loading matrix $\underline{P}$ and an error matrix $\underline{E_1}$ as:

$$\underline{X} = \underline{T}\,\underline{P}' + \underline{E_1} \qquad 3.11$$

In a similar manner matrix $\underline{Y}$ is decomposed into a score matrix $\underline{U}$, a loading matrix $\underline{Q}$ and an error matrix $\underline{E_2}$ as:

$$\underline{Y} = \underline{U}\,\underline{Q}' + \underline{E_2} \qquad 3.12$$

A relationship between the scores of the data sets can be established to extract the latent variables. The first latent variable is extracted from the matrices $\underline{X}$ and $\underline{Y}$ and explains as much as possible the variance of matrix $\underline{Y}$. Different algorithms can be used to extract the factors. The most popular algorithm used in PLSR is known as Non-Iterative Partial Least Squares (NIPALS) (Galedi, 1986 and Wise, 1990). Another algorithm, known as SIMPLS, can also be used (De Jong, 1993). When the optimal number of latent variables has been determined, the remaining variance is considered to be contributed by noise.

3.2.4 Canonical Correlation Analysis (CCA)

Canonical correlation analysis (CCA) is the generalization of the correlation coefficient. While the correlation coefficient measures the association between two sets of n observations; CCA generalizes this principle to the association between two sets of variables. CCA is useful when there is more than one response variable and especially when the predictor variables are moderately correlated. CCA is not a prediction technique but rather an explanatory technique for portraying the relationship between two sets of multivariate data. In the canonical correlation technique, one is looking for linear combinations of the predictors and linear combinations of the responses, which, themselves, have maximum correlation.

The approach has some similarity to PCA, which searches for patterns whitin a single multivariate data set that represent maximum amounts of the variation in the data.

CCA transforms pairs of responses ($\underline{Y}$) and predictors ($\underline{X}$) into sets of new variables called canonical variates:

$$\underline{v}_k = \sum_{i=1}^{p} a_{k,i}\, \underline{x}_i \qquad k=1,....,\min(p,m) \qquad 3.13$$

and

$$\underline{w}_k = \sum_{j=1}^{m} b_{k,j}\, \underline{y}_j \qquad k=1,....,\min(p,m) \qquad 3.14$$

where $a_{k,i}$ and $b_{k,j}$ are called canonical coefficients and the correlation between them, r_i, is called the canonical correlation coefficient, $\underline{v}_1$ and $\underline{w}_1$ is the first pair of canonical variables which have the maximum correlation, $\underline{v}_2$ and $\underline{w}_2$ is the second pair of canonical variables, independent of the first pair, which has maximum correlation, and so on. Assuming that both S_{xx} and S_{yy} are full rank, the number of pairs of canonical variables will be the minimum of p and m. The canonical correlations are ordered in the same manner as characteristic roots: $r_1 \geq r_2 \geq \ldots \geq r_{\min(p,q)}$.

The information drawn upon by CCA is contained in the joint variance-covariance matrix of the variables $\underline{X}$ and $\underline{Y}$. These correlations may be obtained from the solution of following eigenvalue problems:

$$\left|\underline{S}_{xx}^{-1}\underline{S}_{xy}\underline{S}_{yy}^{-1}\underline{S}_{yx} - r^2\underline{I}\right| = 0 \qquad 3.15$$

and

$$\left|\underline{S}_{yy}^{-1}\underline{S}_{yx}\underline{S}_{xx}^{-1}\underline{S}_{xy} - r^2\underline{I}\right| = 0 \qquad 3.16$$

where $\underline{S}_{xx}$ ($p \times p$) is the variance covariance matrix of predictor variables $\underline{X}$, $\underline{S}_{yy}$ ($m \times m$) is the variance covariance matrix of response variables Y, the matrices $\underline{S}_{xy}$ ($p \times m$) and $\underline{S}_{yx}$ ($m \times p$) contain the covariances between each of elements of $\underline{X}$ and each element of $\underline{Y}$.

Equations 3.15 and 3.16 yield the same eigenvalues, since the two matrixes involved are of the form $\underline{A}\,\underline{B}$ and $\underline{B}\,\underline{A}$ where $\underline{A} = \underline{S_{yy}}^{-1}\underline{S_{yx}}$ and $\underline{B} = \underline{S_{xx}}^{-1}\underline{S_{xy}}$. $\underline{A}\,\underline{B}$ and $\underline{B}\,\underline{A}$ have the same eigenvalues but different eigenvectors. Square roots of the eigenvalues are the canonical correlation between the canonical variables. Canonical coefficients of $\underline{X}$ and, coefficient of Y variables are eigenvectors of (3.15) and (3.16) respectively.

A diagram of different approaches to linear multivariate calibration is presented in Figure 3.1. Predictor variables $\underline{X}_1, \ldots, \underline{X}_p$ are used in a linear regression equation to predict the response variable $\underline{Y}$ (Figure 3.1.a). In principal component regression, the factors extracted from highly correlated $\underline{X}$ variables are used to predict $\underline{Y}$ (Figure 3.1.b). In partial least square regression, information on $\underline{Y}$ is also used for extracting the factors (Figure 3.1.c). In CCA, two sets of factors are extracted to explain relation between two sets of multivariate data (Figure 3.1.d).

3.3 Artificial Neural Networks (ANNs)

In recent years there has been a growing interest in a class of computing devices that operate in a manner analogous to that of biological systems. Artificial neural networks (ANN) have been applied in almost all branches of science and engineering including structural engineering. An overview of applications of ANNs to civil engineering is given in Flood et al (1994). Several factors have simulated this interest, the most notable is the ability to learn and generalize from examples, to produce meaningful results even when input data is incomplete or contain error. Before a neural network can be used with any degree of confidence, there is a need to establish the validity of the results. A network could provide almost perfect answers to the set of the problems with which it was trained, but fails to produce almost perfect answers to other examples.

3.3.1 Basic concepts

Artificial neural networks (ANNs) such as three layer back-propagation networks and radial basis function networks have been proven to be performing complex function approximation. This ability to approximate complex functions has been exploited in applying ANNs as models of processes. Neural networks have been trained to perform complex functions in various fields of application including pattern recognition,

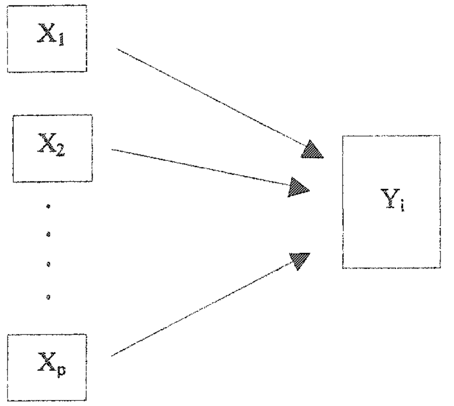


Figure 3.1.a, Multivariate Linear regression

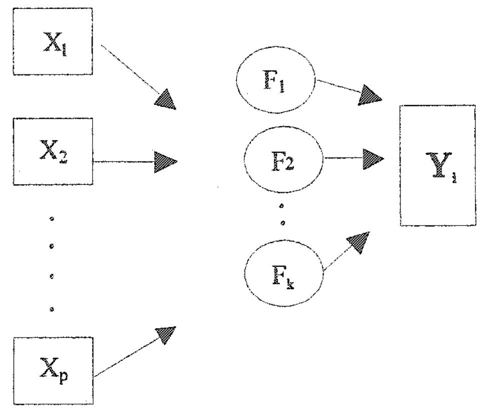


Figure 3.1.b, Principal Component Regression

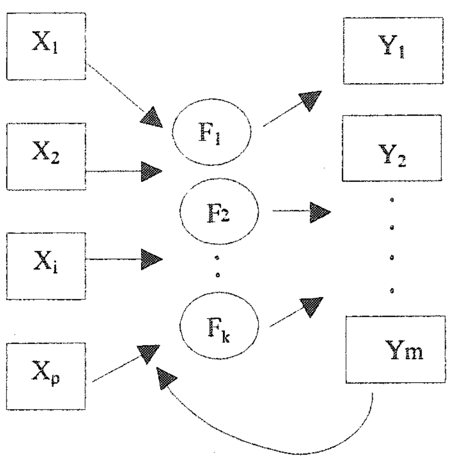


Figure 3.1.c, Partial Least Square Regression

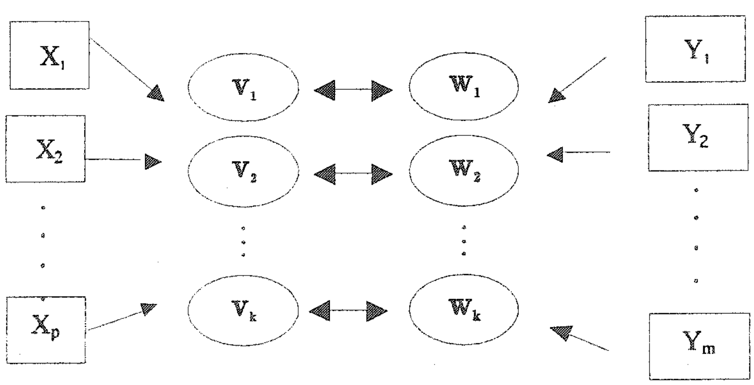


Figure 3.1.d, Canonical Correlation Analysis

Figure 3.1, Comparison of multivariate statistical methods classification, speech, vision, and control systems. The main advantage of ANNs models is that they can be synthesized without detailed knowledge of underlying process. Neural networks are configured from a number of parallel operating processors, termed neurones. Each processor maintains only one piece of dynamic information which is its current level of activation and is capable of doing some simple calculation such as adding inputs, computing a new activation level, or comparing input to a threshold value. But collectively, in the form of a neural network, they are capable of solving complicated problems. The type of activation function adopted, the topology of the connections, and the values of the connection weights determine the task performed by a network. Usually, the activation function and topology of the connections are selected first and so it is left to determine an appropriate set of weights that make the network perform the required task.

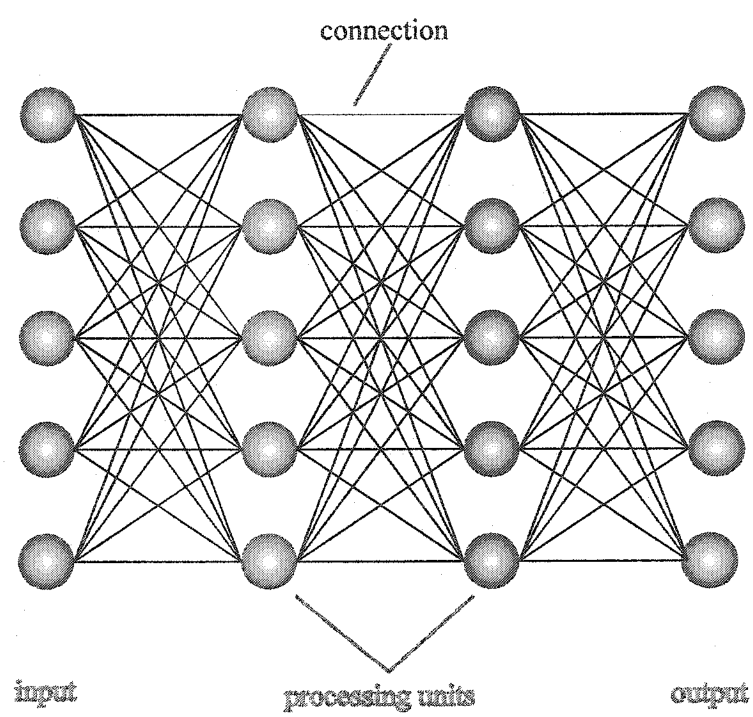


Figure 3.2, Sample Neural Network

3.3.2 The anatomy of a Neural Network

The basic anatomy of neural networks, can be divided into several basic concepts: 1) a set of processing units, 2) the state of activation of a processing unit, 3) the function used to compute output of a processing unit, 4) the pattern of connectivity among the processing units, 5) the rule of Propagation employed, 6) the activation function employed, and 7) the rule of learning employed. An overview of these concepts is presented next.

Set of processing units

All neural networks are composed of a set of processing units. These nodes carry out all processing and calculations. A processing unit receives input from its neighbors, computes an output and sends it to its neighbors. The processing units can be divided to three groups, input units, hidden units, and output units (Figure 3.2).

State of activation

Each unit has an activation function level, which is most often represented as a continuous quantity between values 0 and 1.

Output function employed

Each processing unit transmits its output to its neighbours. This output, which is also a scalar value between 0 and 1, is determined from the level of activation of the processing unit. An output function f is associated with each processing unit, which defines how the output value for the processing unit is determined from its activation. The relationship between the activation level and the output level for any processing unit I can be described as follows:

$$o_i = f_i[a_i] \qquad 3.17$$

The output function can be the unity function or a threshold function.

Pattern of connectivity among the processing units

Processing units are connected to other processing units and communicate with each other via these connections. The pattern of connectivity and the strength of the connections influence how a neural network performs the most. The absolute value of w_{ij}, represents the strength with which the i^{th} unit excites or inhibits the j^{th} unit.

Rule of propagation employed

The rule of propagation describes how the inputs and the strengths of the connections arriving at a node are to be combined to compute the net input. Most often, this rule is simply a weighted summation:

$$N_i = \sum w_{ij} \rho_i \tag{3.18}$$

If outputs and the weights of the connections coming from the other nodes are represented by (o_1, o_2,..., o_n) and (w_1, w_2,..., w_n) respectively, this net input is simply the dot product of these two vectors. The dot product is maximum when row vectors are in the same direction.

Activation function employed

The activation function, F, defines how the net input received by the node and its current level of activation is combined to compute the new level of activation. This is mathematically expressed by:

$$a_{i-new} = F_i[a_{i-new}, N_i] \tag{3.19}$$

The sigmoid function is commonly used in neural network modelling. The sigmoid function keeps the value of activation between 0 and 1.

Rule of learning employed

The learning rule defines how neural network is modified in response to input data and learns from examples. Two general learning rules are used most frequently in neural networks. The first, hebb's rule of learning is stated as "when an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased" (Hebb, 1949). For simple hebbian learning, the learning rule can be simply stated as:

$$\Delta w_{ij} = \eta \, a_j o_i \tag{3.20}$$

Where η is a constant with a value between 0 and 1 representing the degree by which the weights are changing when both units are excited. Another common form of learning rule is the delta rule:

$$\Delta w_{ij} = \eta[t_j - a]_j o_i \qquad 3.21$$

It is called the delta rule because the learning is proportional to the difference between actual and expected activation.

3.3.3 The Back Propagation Neural Network

The Back Propagation Neural Network (BPNN) is currently the most general-purpose and commonly used neural network paradigm (Swingler, 1996). BPNN learns to generate a mapping from the input pattern space by minimizing the error between the output produced by the network and the desired output across a set of input vectors or exemplars. The learning process starts with presentation of an input pattern to BPNN. The training of a multi-layer BPNN, via the generalized delta rule is an iterative process. Input pattern is propagated through the entire network until an output is generated. The error in each layer is calculated with generalized delta rule. Each step involves the determination of error associated with each unit, and then modification of weights on the connections coming out to that unit. The weights in different layers are slightly changed in each step to reduce its error signal and the process is repeated for the next pattern. A set of cycles, made up of one cycle for each row of input data, is called an epoch. The training process for a network requires sometimes thousands of epochs for all the input features to be learnt by the network. The iterations are stopped when the sum of squares of the error for all the input in training set is below a pre-determined value (convergence), or when the maximum number of epochs is performed by the network, in this case, the network does not converge and it has to be re-designed.

3.3.4 PCA Neural Networks

The successful application of ANNs for extracting the principal components has been reported in many studies. Neural networks have been applied for extracting both linear and non-linear principal components from the data. Non-linear principal component analysis is a generalization of PCA.

A good review of linear and non-linear neural network is provided in Diamantaras et al (1996). The Auto associative neural networks with a bottleneck layer of nodes (Figure 3.3) can be used to reduce the dimension of input variables.

The network has three layers, with *p* nodes in the input and output layers and one node in the hidden layer. The activation functions are all linear, so the outputs are given by $(\underline{x}'\underline{w})\underline{w}$, where $\underline{w}$ is a weight vector. Weights are estimated using least-squares method.

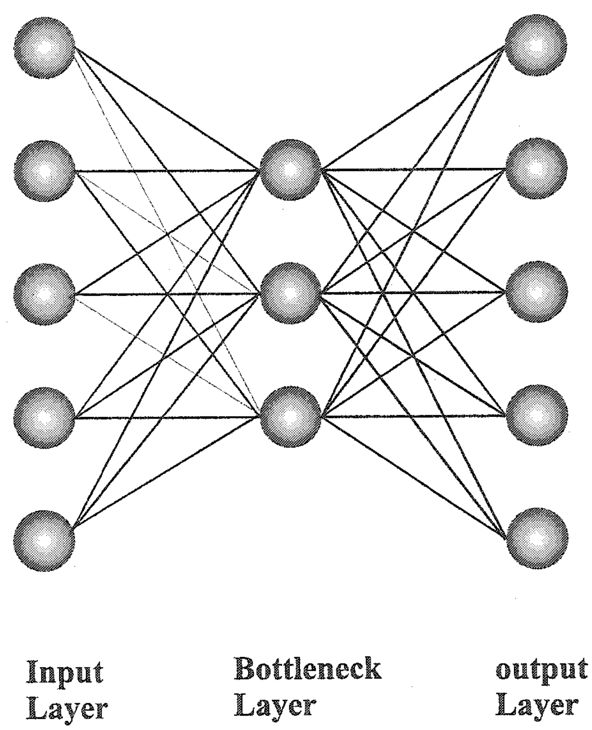


Figure 3.3, Autoassociative Neural Networks

The network is called an auto associative neural network because it is trained to reproduce its inputs. The hidden layer in an auto associative network is also called a bottleneck layer because the *p*-dimensional inputs must pass through the *k*-dimensional bottleneck layer before reproducing the inputs. Data compression therefore occurs in the bottleneck layer. NLPCA is a direct generalization of the neural-network implementation of PCA. NLPCA modifies the PCA networks by adding hidden layers with non-linear activation functions between the input and bottleneck layers and between the bottleneck and output layers, giving a network with a total of five layers. The network models a composition of functions. Figure 3.4 shows an example of a NLPCA network. The five-layer NLPCA network has p nodes in the input layer, $k< p$ nodes in the third (bottleneck) layer, and p nodes in the output layer.

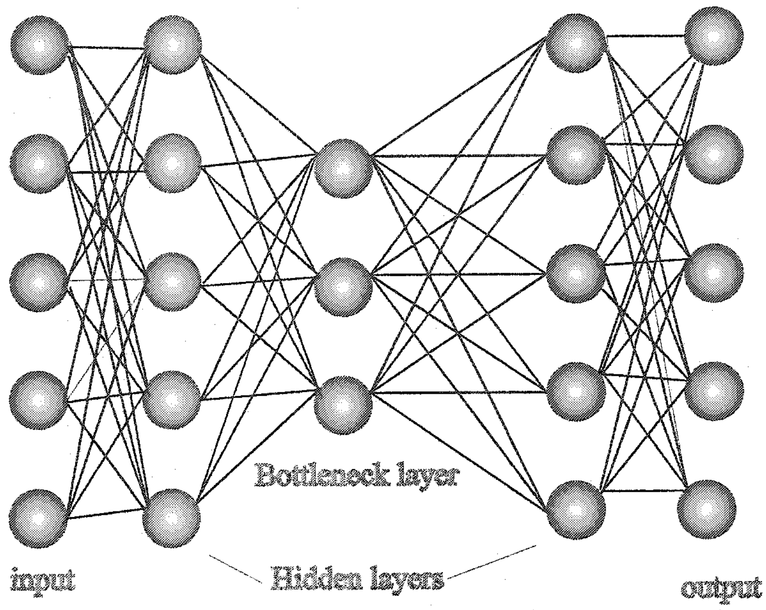


Figure 3.4, Nonlinear PCA Neural Network

The nodes in layers 2 and 4 must have nonlinear activation functions so that layers 1, 2, and 3 and layers 3, 4, and 5 can represent arbitrary smooth functions. The nodes in layers 3 and 5 usually have linear activation functions, although they could be nonlinear. Direct connections are allowed between layer 1 and 3 and between layer 3 and 5, but direct connections are not allowed to cross bottleneck layer 3. As with the linear PCA networks, data compression takes place because the p-dimensional inputs must pass through the k- dimensional bottleneck layer before reproducing the inputs. Once the network has been trained, the bottleneck node activation values give the scores.

3.4 Process control methods

Statistical Process Control (SPC) forms the basis of traditional process performance monitoring and the detection of process malfunctions. The objective of SPC is to monitor the performance of a process over time and to verify that it remains in a state-of-statistical-control. Traditionally, this is achieved by successive plotting and comparison of a chosen sample statistic with appropriate control limits (Efthimiadu et al., 1995). If the plotted variable exceeds the respective control limits, the process is considered to be out of statistical control.

SPC charts such as the Shewhart chart are well-established statistical procedures for monitoring stable univariate processes. The assumption behind them is that a process subjected only to its natural ("common cause") variability will remain in a state of statistical control unless a special event occurs. (Kresta et al., 1991). The control charts represent several statistical hypotheses testing procedures aimed at detecting the occurrence of a special event as quickly as possible.

A Shewhart chart consists of plotting a given statistic sequentially on a graph, which displays a target value, and upper and lower limits (Figure 3.5). The control limits are usually determined by analyzing the variability of the process when the process is under control. The limits are then usually set at plus and minus three standard deviations about the target (Figure 3.6).

When the mean of the statistic is not constant and its trend is predictable, the residuals of a linear regression model are used as control variables. Values of the standardized variables

$$Q_i = \frac{\varepsilon_i}{\sigma} \qquad 3.22$$

can be plotted on Shewhart chart.

The difficulty with plotting several univariate control charts is that response variables are generally not independent. Figure 3.7 illustrates the failure of the univariate control to detect out-of-control state, indicated by a cross. The bivariate control ellipse (Figure 3.7) provides a more detailed and compact representation of the system. The out-of-control point illustrated cannot be detected by univariate charts and can only be detected by the control ellipse.Typically process monitoring applies to systems or processes in which more than one variable is measured and tested. Multivariate Statistical Process Control methods (MSPC) address some of the limitations of univariate monitoring techniques by considering all the data simultaneously, and extracting information on the behavior of one variable relative to another (Martin et al, 1996).

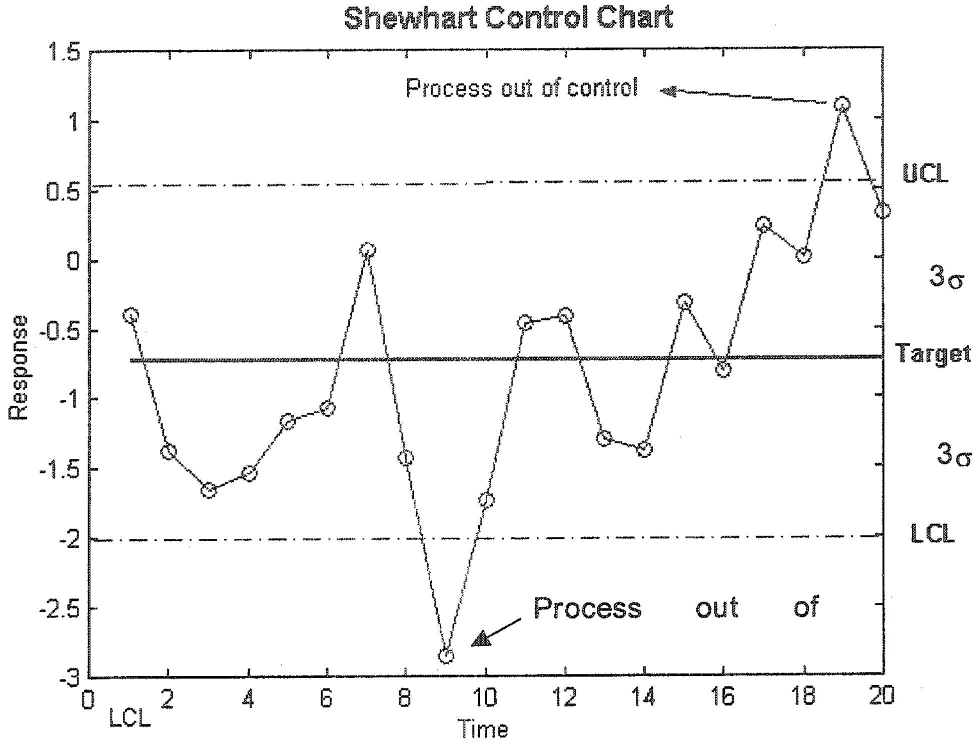


Figure 3.5, A Typical Shewhart Chart

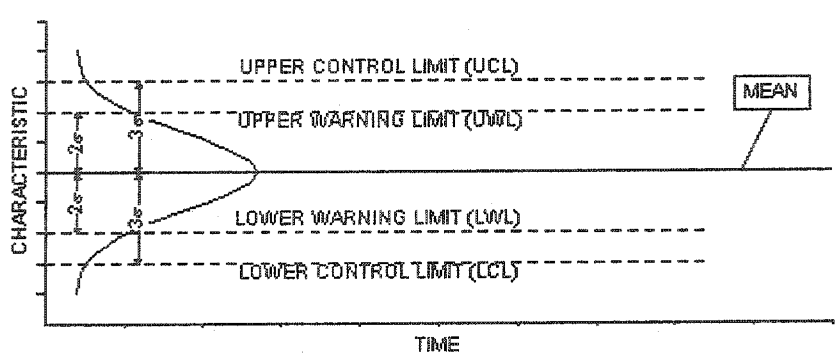


Figure 3.6, Warning and control limits

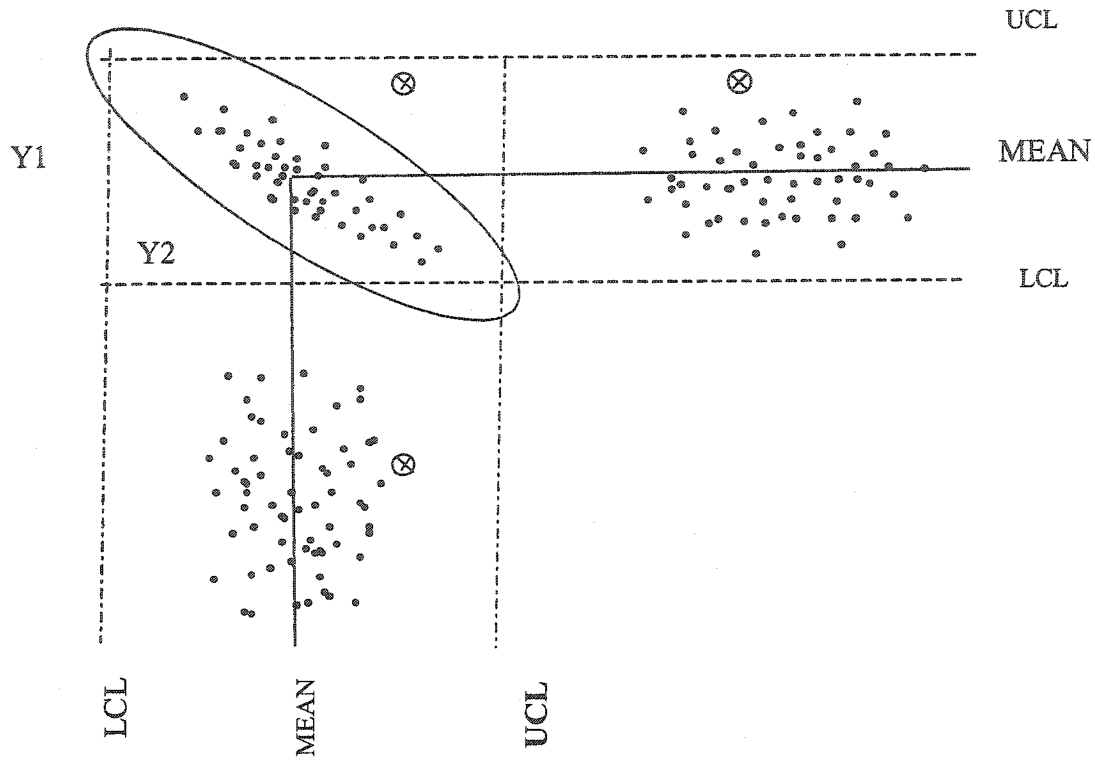


Figure 3.7, Quality Control using two Variables

Multivariate Statistical Process Control (MSPC)

MSPC is increasingly being recognized as a valuable tool for providing early warning of changes in processes and for a better understanding of processes. MSPC is the multivariate extension of univariate statistical process control (SPC). Examples of the MSPC methods are multivariate T^2 statistic, two-dimensional plots of latent variables scores (from PCA or PLS), and the Squared Prediction Error (SPE).

One approach is to extend the univariate analysis by plotting a statistic, which measures the overall deviations of the several statistics from their targets. The most commonly used statistic of this type is the Hotteling T^2.Hotelling statistic was the first to consider the problem of analyzing a correlated set of variables. The procedure is based on the concept of statistical distance.

The T² statistic is defined as:

$$T^2 = (\underline{Y} - \underline{\overline{Y}})' \underline{S}^{-1} (\underline{Y} - \underline{\overline{Y}}) \tag{3.23}$$

where

$$T^2 \frac{n-m}{(n-1)m} \approx F_{m,n-m} \tag{3.24}$$

T^2 has a F distribution with p and n-p degree of freedoms, where n and m are the number of observations and the number of variables respectively. Statistics $\underline{S}$ and $\overline{Y}$ are sample covariance and mean value which are estimated from a sample on available past multivariate observations:

$$\underline{S} = \frac{\sum_{i=1}^{n} (\underline{Y}_i - \underline{\overline{Y}})'(\underline{Y}_i - \underline{\overline{Y}})'}{n-1} \tag{3.25}$$

$$\underline{\overline{Y}} = \frac{\sum_{i=1}^{n} \underline{Y}_i}{n} \tag{3.26}$$

When new multivariate observations are obtained, then the Hotteling T^2 statistics can be plotted as a function of time. Figure 3.8 presents an example chart with α=0.01 for detecting possible anomalies in the system. The T² statistic has emerged as an extremely useful metric for multivariate process control.

Another approach to multivariate quality control is to transform a p-dimensional set of highly correlated data into a lower k dimensional set of data using PCA and PLS models. These models are known to be suitable for handling noisy or highly collinear or highly correlated data (MacGregor et al., 1994).

The most common forms of presenting the information is through one and two-dimensional plots of principal component scores, Hotelling's T^2 (Jackson, 1991) and the Squared Prediction Error (SPE), also known as Q statistics (Jackson, 1991). Once a model has been developed from the nominal data using a reduced set of principal components of latent variables, k, the fitted values can be calculated. These values are then used to evaluate the SPE for each new observation. That is the squared difference between the observed values and the predicted values from a reference model.

$$SPE = \sum_{i=1}^{m} (y_{new,i} - \hat{y}_{new,i})^2 \qquad 3.27$$

where $y_{new,i}$ are observed variable and $\hat{y}_{new,i}$ is computed from PCA reference model. Using the first k principal components, this statistic represents the squared perpendicular distance of a new multivariate observation from the hyper plane.

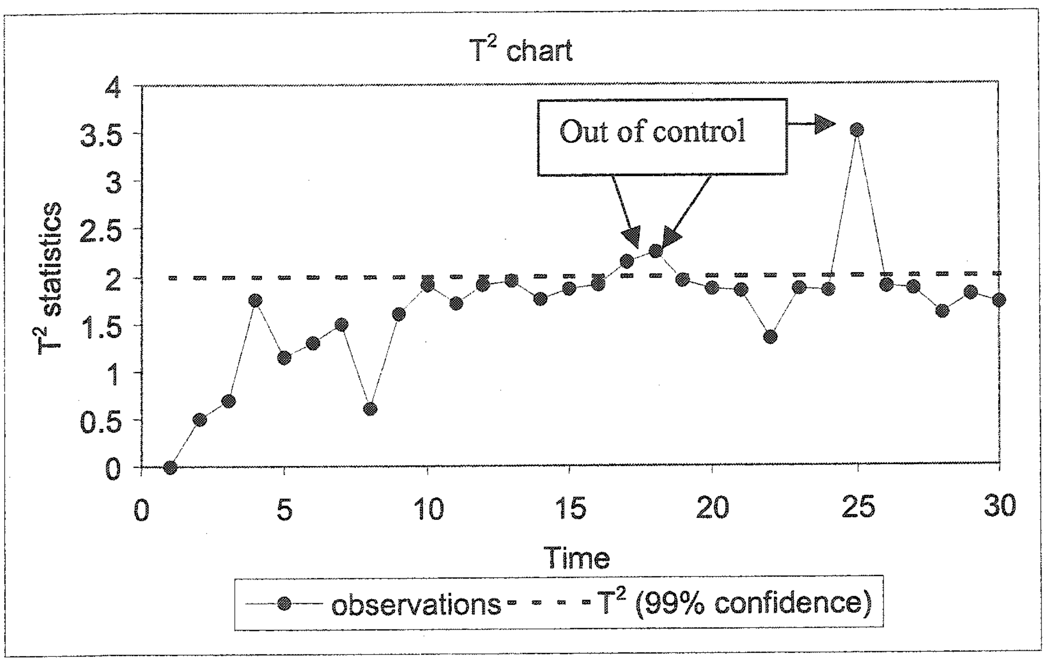


Figure 3.8, T^2 chart for a Multivariate Process

When the system is in control, the value of SPE or Q should be small. When process is in control SPE represents noise that can not be accounted by model. The SPE plot provides the facility to identify a new event not previously captured in data. By adopting an approach similar to that for univariate SPC, action and warning limits can be defined for each latent variable plot based on standard statistical distribution theory. The only requirement for applying these methods is the existence of a good database of past observations when the system was behaving normally.

CHAPTER 4

4. Dam Instrumentation and Monitoring

4.1 Introduction

"When you measure what you are speaking about and express it in numbers, you know something about it, but when you cannot express it in numbers your knowledge about is of a meagre and unsatisfactory kind"' *(Lord Kelvin (1824- 1907)*

"Listening what the dam tells through its monitoring system is an alternative to sophisticated calculation models" *(Dibiagio, 2000)*

The challenge of managing ageing dams is rapidly becoming a principal focus of dam engineering throughout the world. At least a quarter of the dams listed in the U.S. Army Corps of Engineers National Inventory of Dams are more than 50 years old (Bowles et al., 1999). The fact that these dams are the product of old standards and construction practices is generally of greater concern than the ageing process itself. The risks associated with ageing dams are typically of low probability but high consequence.

The safe operation of dams is an extremely important matter of public safety and economics. It is imperative, therefore, to have a means of gathering information that can be used to assess dam performance and safety. Figure 4.1 shows different means of gathering information about a dam state. Field measurements, the art of monitoring and qualifying the behaviour of structures by taking physical measurements has traditionally been used by dam engineers for this purpose. Dam monitoring programs have also contributed significantly to advancements in the state-of-the-art of dam engineering (Dibiagio, 2000).

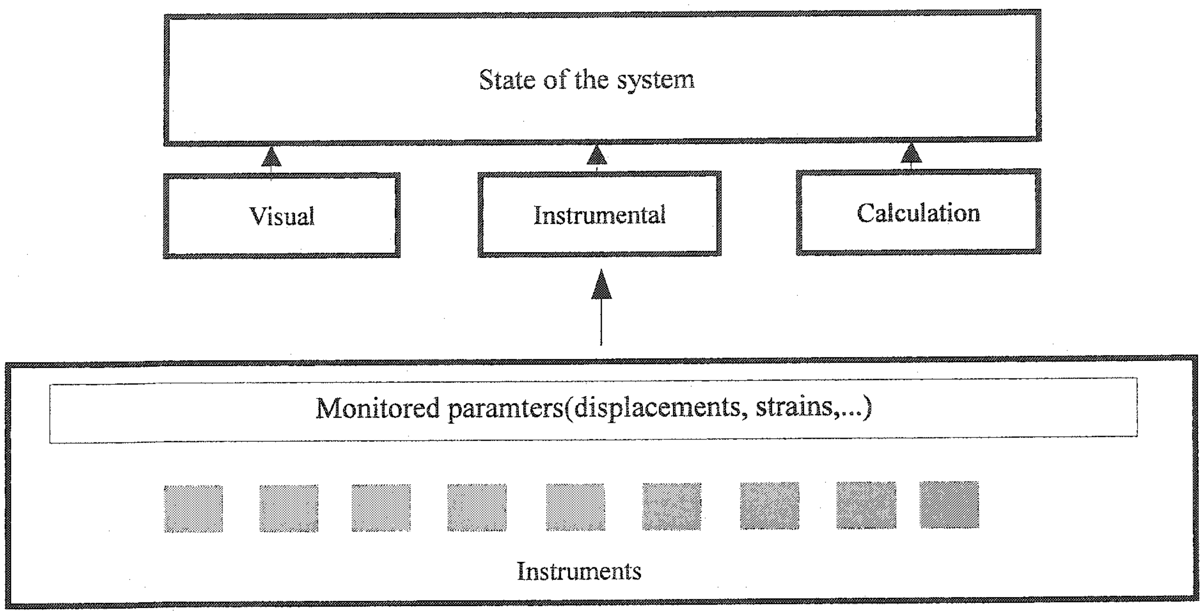


Figure 4.1, Block diagram of dam evaluation methods

Dam monitoring is the most effective defensive line against dam failure by early detection of anomalies (Hulea et al., 2000). Major dams, like other large constructed facilities are equipped with various types of instruments to monitor their behavior. The efficiency of these devices depends on their diagnostic value and the quality of the data. Understanding the role of different types of measurements can improve the quality of a monitoring program.

Measurements are of particular interest to civil engineers because of the uncertainty in predicting the behavior of dams (Dibiagio, 2000). Accurate numerical modeling of field conditions is often impossible or impractical to achieve. Thus, it is frequently necessary to make assumptions regarding the materials properties or important features such as drainage conditions, degree of rock fissuring, in-situ stresses, etc. Consequently, many dam problems cannot be solved strictly on the basis of mathematical analyses and physical experiments, therefore other sources of information are required. Instrumentation and visual inspections are necessary to fill in and bridge the knowledge gap between theory and actual behavior.

The greatest improvements in the instrumentation have been a result of developments in instrumentation technology, material science and information technology. The role of sensors is to convert a measurement, usually an electrical signal, into another quantity, which can be more easily interpreted. The perfect measurement system does not exist nor does the perfect measurement environment. Thus, all measurements are subject to disturbances from many source of error. Measurements should always be validated by theoretical and/or practical verification. Too often data is accepted without questioning its accuracy.

4.2 Measurement errors

Every measurement is always inaccurate to some extent. Measurements are always corrupted with stochastic deviations or noise. Noise is inherent to all physical systems but its level can be reduced by appropriate measures in terms of measurement and system design. Identifying the various errors, which exists in measurement systems, is vital for a good monitoring system. It is necessary to reduce errors in the instrument readings to the minimum possible level, and to quantify the maximum error, which may exist in any output reading. Two main types of measurement errors are generally recognized: systematic errors, or bias in which every measurement is either less than or greater than the correct value by a fixed percentage or amount, and random errors, which are unpredictable variations in the measured signal. This latter type of error is often called noise, by analogy to acoustic noise.

4.2.1 Systematic errors

Systematic errors describe errors in the output readings of a measurement system, which consistently overpredict or underpredict a quantity. Sources of systematic errors are system disturbances during measurement, damaged sensors and use of uncalibrated instruments. Filtering cannot deal with systematic error due to drift or incorrect calibration of the measurement device. Commonly bias can be eliminated by re-calibration of the instruments.

4.2.2 Random errors

Random errors are perturbations of the measurement that can be on either side of the true value. Such perturbations are usually small, but their importance is a function of magnitude of the signals. Random errors are introduced when measurements are taken manually, and when this involves interpolation on a scale. Electrical noise can also be a source of random errors. Other sources of random errors include uncontrolled influential factors, such as air currents, ambient temperature fluctuations, relative humidity, power source disturbances, electromagnetic interference, and connector contact resistance variations. One particular area of concern is the cables which are connected to measurement instruments. Cables can be a major source of problems in dam monitoring. They are extremely vulnerable to damage during the construction and they may become unserviceable with time because of ingress of water. Also cables can function as antennas and become a major source of electromagnetic interference and noise (Dibiagio, 2000).
To a large extent, random errors can be reduced by taking the same measurement several times and averaging the measurements. However, any estimation of the measurement value and its error bounds must be treated statistically.

4.2.3 Signals and noise

The choice of an instrument will in many cases be dependent on the type of output signal produced by the device. The reason is that certain types of signals have definite advantages regarding noise immunity or are more tolerant of changes in electrical characteristics of cables and connections in the monitoring program. Table 4.1 lists the most common types of output signals for typical sensors used in dam monitoring systems (Dibiagio, 2000). The types of signals are listed in order of increasing preference.

Field and experimental measurements are never perfect, even with sophisticated modern instruments. When measurements are corrupted by random variables, they are said to be affected by noise. The standard deviation is a good measure of magnitude of the noise in the signal. One of the fundamental problems in signal measurement is distinguishing the noise from the signal. What really distinguishes signal from the noise is that the noise in not predictable or reproducible, it changes from one measurement to the next. If a signal can be measured more than once, the average of the measurements provides a better representation of the signal.

Table 4.1, Common Types of Output Signals

Types of signals	**Examples of typical sensor**
Low level analogue	Resistance strain gauges
High level analogue	Servo-inclinometers
D.C current (4 – 20 mA)	Many process type instruments
Frequency	Vibrating wire strain gauges
Digital	Encoding and smart sensors

In many measurements in physical science and engineering, the true signal measurements evolves rather smoothly as a function of time or position, whereas noise is characterized by rapid changes in amplitude from one point to the next. Some types of noise can be easily separated from the true signal. For example, it may occur that the signal is disturbed by sharp spikes at a few points. These spikes can be detected by comparing the value of each data point with its neighbors. If the difference with its neighbors is larger than a given threshold value, it can be identified as a spike. The spikes are removed, without affecting the rest of the signal, by replacing the spike by the average value of its neighbors. The threshold might be defined as some multiple of the estimated standard deviation of the noise. It is common practice to reduce the noise by a process called smoothing. As long as the true underlying signal is smooth, then the true signal will not be much distorted by smoothing, but noise will be reduced. The simplest smoothing algorithm is the rectangular or unweighted sliding-average. It simply replaces each point in the signal with the average of m closest points.

$$x_k = \frac{1}{m}\sum_{k-m+1}^{k} x_i \qquad 4.1$$

The triangular (weighted) and exponential smoothing methods can also be used and give more importance to values closest to the current measurement point.

Exponential averaging is obtained by calculating the weighted average of the points in a moving window of m data points. The last point in the window (i.e. the point i to be smoothed) is given the greatest weight and each proceeding point is attributed a lower weight determined by the shape of the exponential function. This filter smoothes the point i by using the points that precede this point. If we consider also the mean with one additional point:

$$\bar{x}_{k+1} = \frac{1}{m+1}\sum_{k-m+1}^{k+1} x_i = \frac{1}{m+1}\left[x_{k+1} + \sum_{i=k-n+1}^{k} x_i\right] = \frac{1}{n+1}x_{k+1} + \frac{n}{n+1}\bar{x}_k \qquad 4.2$$

By shifting the time index back one time-step, the corresponding expression is obtained:

$$\bar{x}_k = \frac{1}{n+1}x_k + \frac{n}{n+1}\bar{x}_{k-1} \qquad 4.3$$

to simplify the notion:

$$\alpha = \frac{n}{n+1} \Rightarrow \bar{x}_k = \alpha\,\bar{x}_{k-1} + (1-\alpha)x_k \qquad 4.4$$

This expression is known as the exponentially weighted moving average filter. The value of the constant α dictates the degree of filtering. The optimum choice depends on the characteristics of the signal and the sampling interval. The exponentially weighted moving average filter is arguably the most commonly used noise reduction algorithm in the process industries, it is also known commonly as the first order low-pass filter. While these methods are effective in reducing the random noise, they introduce some shift into the signal. Values of 0.3-1 are normally used for α. The larger values introducing larger lags in the filtered signals. A compromise has to be made when selecting a value for α for achieving sufficient noise reduction with lags.

4.3 Dam safety

The safety of an existing dam can be improved and its life lengthened by a carefully planned and implemented surveillance program. A key part of such a program is a visual examination of the structure, the reservoir, and the appurtenant works. However, surveillance must be more than visual observations. Settlements may go undetected without proper measurements of the dam. Comparison of seepage quantities from one inspection to another and over the years is difficult by visual observation and estimation. There are also conditions within a dam that cannot be seen but that can be measured by instrumentation. Thus, even for a simple structure, some type of instrumentation may be needed to improve and supplement the visual examination. Dam safety surveillance today is a two-part process based on periodic visual inspection of accessible parts of the dam and its surroundings, and systematic monitoring of the body of the dam and its foundation by means of instrumentation systems designed specifically for this purpose.

4.3.1 Dam Monitoring

The purpose of instrumentation in an existing dam is to furnish data to determine if the completed structure is functioning as intended, and to provide a continuing surveillance of the structure to warn of any developments that endanger its safety (Post, 1985).

Every instrumentation program should include some redundancy; especially with embedded instruments since it is usually not possible to repair damaged embedded instruments. The cost to retrofit replacement instruments will far exceed the cost of providing adequate redundancy. Redundancy, in this case, is more than only the furnishing of additional instruments to account for those that are defective or are damaged during installation. Redundancy includes providing different instruments, which can measure similar behavior with different methods.

The general layout of the instruments is based on the vertical or horizontal sections, which the designers consider to be of utmost importance. The types of field measurements needed to evaluate the behaviour of a designed structure depend on the theoretical concepts that are available to the designer at the time the design is made. In dam monitoring, measurements are taken only on parameters seemed significant and at points judged critical. Stateler et al. (1995) propose that a set of performance parameters be determined for each dam and that these parameters should form the basis of the monitoring system. These parameters can be used in design of an instrumentation system. Dunnicliff (1988) provides examples of possible geotechnical questions associated with the appropriate features and parameters (Table 4.2). The architecture and choice of components for monitoring dams are based on the analysis of the structure's behavior carried out by designers in developing the project. The designers consider certain modes of deformation and their amplitude, as well as certain failure mechanisms.

Concrete and masonry dams are inspected and monitored on a continuous basis following a carefully planned program. To aid in these inspections and in the analysis of the condition of the dam, a number of monitoring methods and devices are used (Figure 4.2). Where these devices are installed, they should be maintained in good condition, and the data obtained should be regularly recorded and evaluated. Because of the higher level of stress, both in the dam and its abutments, the instrumentation of arch dams must be very carefully planned and denser than that of other types of concrete dams (Bordes et al., 1998). A summary of the most important factors to monitor is presented in Table 4.3. Various monitored properties of concrete dams are discussed next.

4.3.1.1 Concrete quality

A great deal can be learned about concrete quality by visual observations. Careful attention should be placed on the appearance of weathered concrete. Pattern cracking might point to drying shrinkage or alkali-aggregate reaction (NRC, 1983). Surfaces subject to rapid flowing water, such as spillways or outlet chutes, must be examined regularly. Where strength is a question, nondestructive tests, such as sonic velocity measurements, and core tests for compressive and tensile strength and modulus of elasticity can be used.

Table 4.2, Steps for Developing an Instrumentation System (Dunnicliff 1988)

Step	Element of Plan
A	Prediction of mechanisms that control behaviour
B	Definition of purpose of instrumentation
C	Definition of geotechnical questions
D	Selection of parameters to monitor
E	Prediction of magnitudes of change
F	Selection of instrument locations
G	Selection of instruments
H	Determination of need for automation
I	Planning for recording of factors which influence measurements
J	Establishment of procedures for ensuring data validity
K	Determination of costs
L	Planning installation
M	Planning long-term protection
N	Planning regular calibration and maintenance
O	Planning data collection and management
P	Coordination of resources
Q	Determination of life cycle costs

4.3.1.2 Abutment or foundation deformations

This is particularly important for an arch dam because excessive deformations can induce high tensile stresses. They may cause cracks in the concrete and high tensile stress measurements may be observed. Also, anomalous decreases in abutment seepage rates may result from closure of openings in the rock mass; and indicate increase of uplift pressures. Abutment deformation can be recognized by fresh cracks in the rock. Sources of deformations are reservoir load, earthquake forces, arch dam thrust, large temperature gradients, freeze and thaw damage, and excessive uplift forces. Remedies are dependent on the cause of the deformation. One or a combination of remedies can be taken, 1) Deep rock anchors or rock bolt to strengthen the abutments, 2) Vertical and horizontal beams across the rock mass and anchored by rock bolts, and 3) Extensive grouting to improve the modulus of elasticity of the rock.

4.3.1.3 Uplift

Uplift pressures are an important factor in stability analysis of a gravity dam. Although uplift pressure is not a critical issue in the stress analysis of arch dams, it is important in the foundation stability analysis and should be included as part of overall instrumentation program. However, measuring water pressures in a rock foundation is always a difficult task since it can change over short distances because of joing and fissuration. Uplift pressures in the foundation and in the dam are measured routinely as indicators of stability. Changes in pressure are looked for; increases may result in instability. Uplift pressures are measured by piezometers inserted in holes drilled into the foundation of the dam. Generally, installation of drains is the most effective and economical solution to reduction of uplift forces. Regular drain flow observations are a must in any surveillance system. When drains become so obstructed as to impair their function, redrilling the old drains or drilling new drains is suggested. In many concrete and masonry dams a foundation drainage system is installed to reduce uplift pressures on the dam. These systems are usually installed during construction but can be installed or supplemented at any time. They consist of holes drilled through the base of the dam into the foundation and may contain pipes. Also, monolith joint drains are commonly installed to intercept seepage along monolith and lift joints. The water maybe checked for chemical and suspended sediment content to aid in evaluation of solution or erosion that may be taking place. The elevation of the reservoir and tail water elevations is recorded at the time of drainage measurements so that relationships between these parameters can be developed.

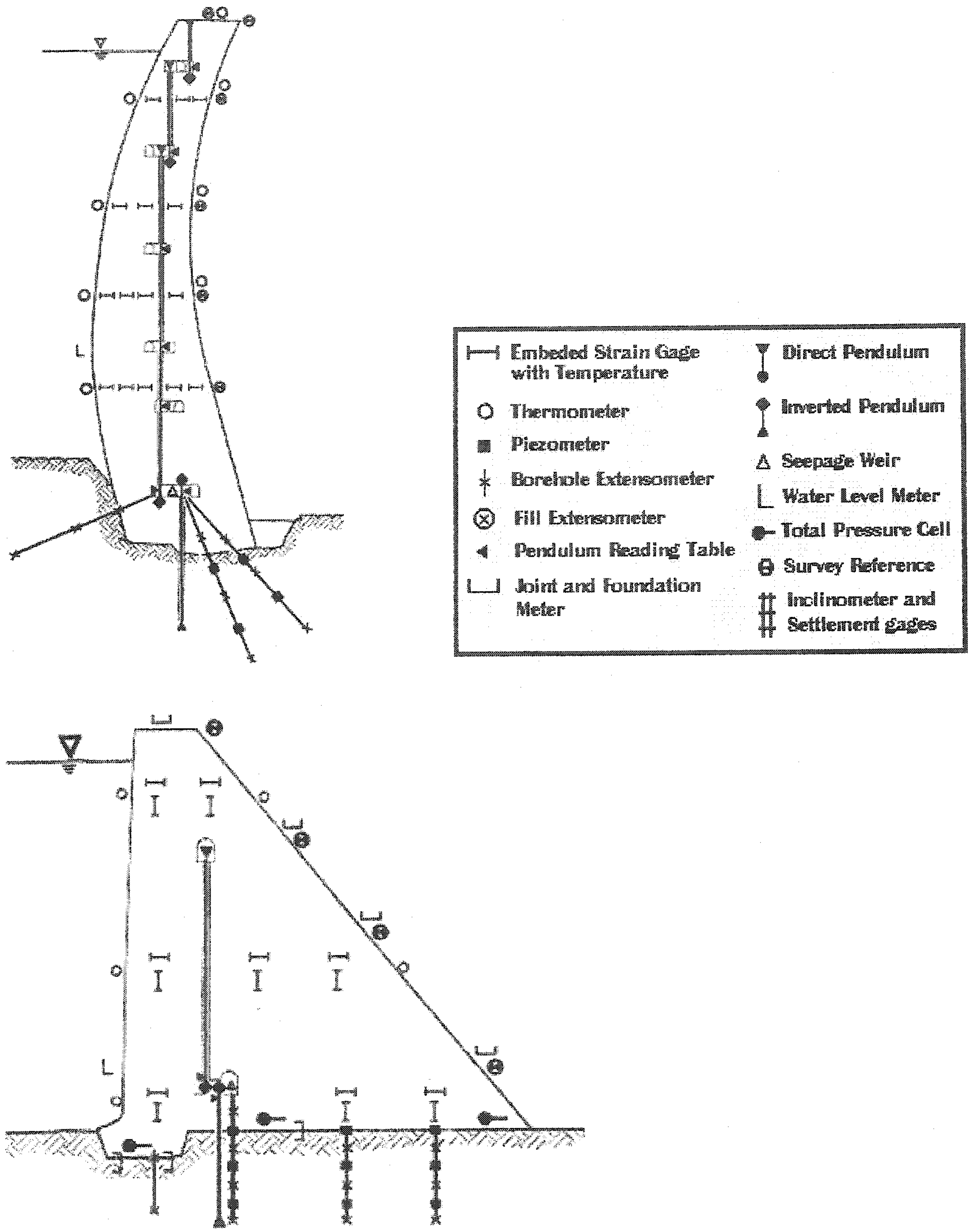


Figure 4.2, Typical instrumentation in arch and gravity dams (Bordes et al., 1998)

Indicator	Possible causes	Possible effects	Potential remedies	Instrumentation
Concrete cracking (general, shallow)	Freeze-thaw cycles Ageing- sulphate attack	Accelerated deterioration Reduction of effective section ,Increased leakage	Conduct quality tests, coring, density, porosity Seal surfaces from exposure	Crack meter Flow meter
Concrete cracking (local)	Stress concentration Freeze-thaw action differential movement	Progressive deterioration Increased leakage Loss of section	Movement monitoring Remove and repair deteriorated sections	Pendulums Joint meters
Deep concrete cracking	Excessive loading Shrinkage (early age) Foundation movement	Increased leakage Accelerated deterioration Increased stresses	Determine extent of cracking Evaluate short-long term effects Seal or grout cracks Increase drainage	Stress/strain meters Flow meters Piezometers Pendulums
Leakage (wet surfaces on concrete)	Cracks Deteriorated concrete Porous concrete	Increased rate of deterioration Loss of strength Increased leakage	Determine path/extent of cracks Seal cracks	Joint meters Pendulums Stress/strain
Leakage (concentrated through concrete)	Cracks, Open joints Differential movements High uplift	Loss of concrete matrix Loss of structural integrity Increased uplift	Map all leak locations Detailed inspection Determine path of water if possible	Water level Flow meter Piezometers Pendulums Stress/strain meters

Table 4.3, Evaluation matrix of a concrete dam (NRC, 1983)

4.3.1.4 Seepage and leakage

Seepage performance is one of the most sensitive early warning indicators (Myers et al., 1997). Seepage through a dam and its foundation is visible evidence that the dam is not a perfect water barrier. Seepage and leakage from the abutments, foundation, and joints or cracks in a dam is collected and measured on a routine basis. It is important to review such flows for changes in magnitude and material, both dissolved and suspended, transported by these flows. Increase in these items is early warning indicators of potential problems. Weirs and venturi flumes with upstream stilling basins are frequently used to measure seepage and leakage.

Conclusions on the performance of seepage control systems can be drawn from several measurements. A common and simple monitoring system is to rely on visual surface inspections at predetermined intervals. Monitoring devices can include piezometers, observation wells, and drainage collection systems to determine a site dependent pattern of behaviour. A regular review of the collected data will generally detect major changes between subsequent readings (UASCE, 1993).

4.3.1.5 Movements

Displacements are probably the most meaningful parameters that can be readily monitored, because of the monolithic behaviour of arch dams. Although displacements occur in all directions, the most significant displacements are usually the ones that take place in a horizontal plane. All concrete arch dams should have provisions for measuring these displacements, including relative movements between points within the dam and movement of the dam relative to a remote fixed point. In new dams, plumblines are still the preferred instrument to monitor the relative horizontal movements within an arch dam. It may be easier to install inclinometers or a series of tiltmeters, in existing dams (UASCE, 1994).

Movement of concrete and masonry dams and their abutments can be expected during and after construction. These movements will occur as the reservoir is first filled, and as it is emptied and filled during succeeding seasons. Small movements are of little concern, but increase in the magnitude of the movement or direction of movement should be immediately evaluated as to their potentially adverse impact on the structure. Movements are measured by surveying the location of the surface monuments located at various points on or adjacent to the dam. The benchmark or starting location for surveys is located outside of the influence of the dam or reservoir if possible. Measurements of the locations of the monuments should be such that changes in vertical, horizontal (both longitudinal and transverse to the dam axis), and angular locations are measured. The number of monuments surveyed depends on the size and type of the structure. The locations are tailored to the structure and, might include locations to measure movement between blocks, displacement at joints and cracks, deflections of various parts of the structure, settlement of the foundation, and movement of the abutments.

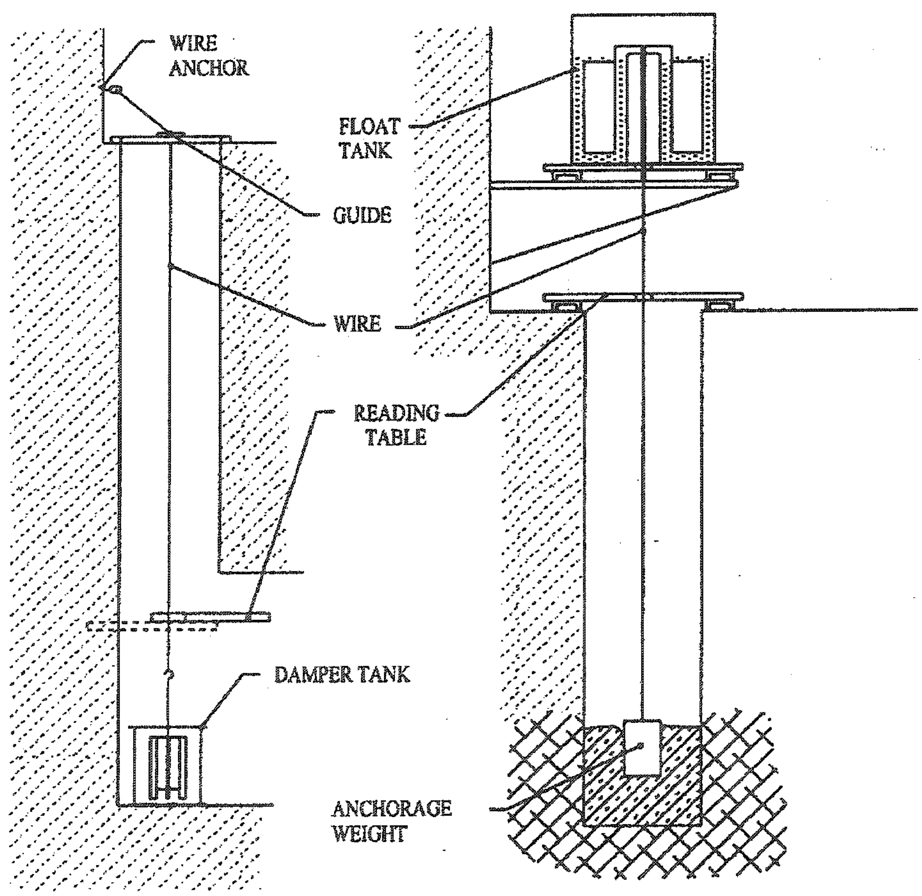


Figure 4.3, Direct and inverted pendulum

Measurements of the monuments should be recorded at relatively short intervals in the initial years of the life of the structure and less frequently as the satisfactory history of the dam lengthens. They should be more frequent if any unsatisfactory performance is indicated. The data collected should be carefully recorded and should include observations on the relative water levels in the reservoir and down stream.

Pendulums are one of the methods of measuring the horizontal and vertical movements of the dam. Direct and inverted pendulums are designed to accurately measure the relative horizontal and vertical displacements of two points along a true vertical line (Figure 4.3). The fixed end of an inverted pendulum is grouted into the lower point of a borehole and a float tensions the wire vertically. The wire position is monitored by a reading table; bolted onto the upper point of the structure. A direct pendulum is comprised of a wire suspended from the upper point and a reading station fixed to the structure at the lower point. The wire is tensioned by a suspended weight that fits into a dashpot to dampen oscillations.

4.3.1.6 Crack and joint measuring devices

Joint meters are used to measure the opening of monolith joints. Depending on the type of device being used, the maximum opening that can be measured may range from 0.08 to 0.4 inches. Joint meters provide information about when the joints have begun to open. They also give an indication of the effectiveness of the grouting and show whether any movement occurs in the joint during and after grouting. In both new and existing structures the development of cracks and movements of joints are indications of the stress on the structure that are sometimes above normal. Measurement of these areas of distress is provided through crack and joint displacement that can be either installed in predetermined locations to monitor expected cracks or to observe joint behaviour, or placed at the site of a known crack or joint as need arises for its monitoring. Relative displacement instruments monitor monolith movement in three dimensions. They measure the relative displacement between two surfaces of the instrument attached to opposite sides of a crack or joint (Figure 4.4)

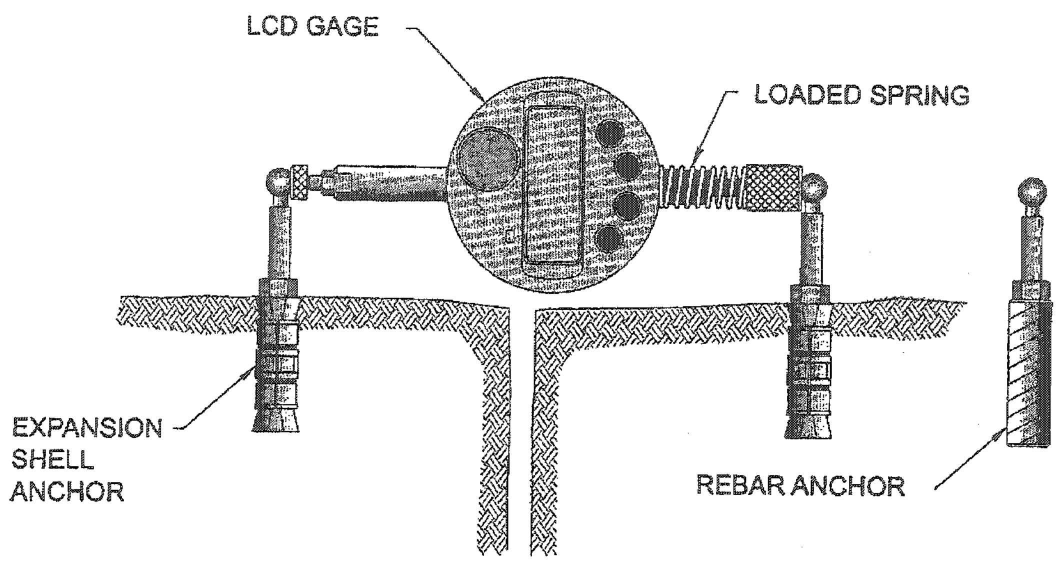


Figure 4.4, A portable fissurometer

4.3.1.7 Stress and strain state

Strain meters measure strain and temperature. Since they measure strain at one location and in one direction only, it is usually necessary to install strain meters in a group of several instruments. Since strain meters do not directly measure stress, it is necessary to convert strains into stresses, which will require knowledge of concrete material properties, which are changing with time, because of creep, shrinkage, and change in modulus of elasticity. These material properties, as well as the coefficient of thermal expansion and Poisson's ratio, are usually determined by laboratory testing. The most common systems of stress and strain measurement are based on the deformations of a hydraulic or pneumatic pressure cell or the deformation of a vibrating wire. In a vibrating wire strain meter, a string or wire is stretched between two points with two electromagnets placed symmetrically with respect to the wire span. At the time of loading the distance between two points will be changed in proportion to the loading, and this will cause a change in the vibration wire frequency. This frequency is a function of the dimensions of the wire, its modulus of elasticity and the strain imposed on the wire.

The strain is converted to stress using the modulus of elasticity E and Poisson's ratio ν of concrete at the measuring point.

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = \frac{E}{(1+\nu)(1-2\nu)} \begin{bmatrix} 1-\nu & \nu & 0 \\ \nu & 1-\nu & 0 \\ 0 & 0 & \frac{1-2\nu}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{Bmatrix} \qquad 4.5$$

$$\begin{Bmatrix} \sigma_x \\ \sigma_y \\ \tau_{xy} \end{Bmatrix} = \frac{E}{1-\nu^2} \begin{bmatrix} 1 & \nu & 0 \\ \nu & 1 & 0 \\ 0 & 0 & \frac{1-\nu}{2} \end{bmatrix} \begin{Bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_z \end{Bmatrix} \qquad 4.6$$

Equations 4.5 and 4.6 can be used for calculation of stresses in plane strain and plane stress conditions respectively. Stress meters measure compressive stresses independently of shrinkage, expansion, creep, or changes in modulus of elasticity. They are used for special applications such as determining vertical stress at the base of a section, and for comparison of results from strain meters. They are also used in the arches for determining horizontal stress normal to the direction of thrust in the thinner arch elements near the top of the dam.

4.3.1.8 Temperature monitoring

The internal temperature of concrete dams is commonly measured both during and after construction. During construction, the heat of hydration of freshly placed concrete can create high stresses, which could result in cracking. After construction is completed and a dam is in operation, it is not uncommon for very significant temperature differentials to exist depending on the season of the year. For example, during the winter, the upstream face of a dam remains relatively warm because of reservoir water temperature, while the temperature of the downstream face of the dam is reduced by cold ambient air temperature. The reverse is true in the summer. Temperature measuring devices are very important in arch dams since volume changes caused by temperature fluctuations have a significant contribution to the loading on an arch dam (USACE, 1994). Thermometers are used to determine the temperature gradients for use in evaluating thermal stresses, which contribute to thermal cracking. They are also used to control the cooling process during the grouting operations and are used to determine the mean concrete temperature. If concrete temperatures are available for enough points in a dam section, the mean concrete temperature for the section can be calculated. The mean concrete temperature can be used for estimating the structural responses of this section (displacements and strains).

4.3.1.9 Seismic instrument program

A seismic instrument program is an essential part of evaluating existing dams in areas of high potential for seismic activity. Devices to measure ground motions and dam responses can facilitate rational design decisions for repairs and strengthening of a structure if damage has occurred as a result of an earthquake. These records are also helpful to compare the performance of the structure with design expectations and to estimate the performance of the structure during larger earthquakes.

4.3.2 Dam inspection

Routine visual inspection of dams is of great value in determining the integrity of the structure. Where signs of deterioration of materials appear, cores and samples are tested in a laboratory to estimate the strength of the material. A routine schedule of nondestructive testing, such as ultra-sonic velocity measurements, can be useful in determining trends of changes in strength.

Careful interpretation of all observations, visual and field instruments must be carried out to assess the situation before methods of repair or upgrading are decided upon. Unexpected observations may not give any reasons for concern if a logical explanation can be found. Fundamental questions have to be answered: (1) what is abnormal behavior? the observed behavior may deviate from what the designer expected, but tested against the professional experience, it may prove not to be so abnormal. (2) Does the deviation from predicted behavior indicate lower safety than aimed for in design? (3) Does the deviation require remedial action, such as repair, upgrading or strengthening?

The frequency of instrument readings or making observations at a dam depends on several factors including: 1) relative hazard to life and property that the dam represents, height or size of the dam, 2) Relative quantity of water impounded by the dam, 3) relative seismic risk at the site, 4) age of the dam, and 5) frequency and amount of water level fluctuation in the reservoir. In general, as each of the above factors increases, the frequency of monitoring should be increased. For example, very frequent (even daily) readings should be taken during the first filling of a reservoir, and more frequent readings should be taken during high water levels and after significant storms and earthquakes. Daily or weekly readings should be made during the first filling, immediate readings should be taken following a storm or earthquake, and significant seepage, movement, and stress-strain readings should probably be made at least monthly.

While instrumentation data are an essential part of a dam surveillance system, the owners of dams also believe, for valid reason, that no automatic data acquisition configuration can replace human judgement when it comes to dam performance and safety monitoring (Bordes et al, 1998). For example, the visual inspection of elements such as cracks in the concrete and the colour of seepage water remains a vital and integral part of any complete monitoring program.

The dam safety philosophy is to promote visual observation as being equally important as instrumentation data. The most dangerous events like, local deformations, cracks, concentrated seepage flows, and wet spots cannot be detected by the instruments. They can eventually be discovered before becoming dangerous by means of visual inspections, which still are the main way of controlling the dam safety (Dibiagio, 2000). Processes affecting safety, but which cannot be measured by instruments installed during construction, can only be detected by visual inspection. Typical examples are cracks being detected and their growth, increasing turbidity in leakage, or the discovery of new leaks downstream of a weir (Post, 1985). An effective inspection program is essential to properly maintain a project in a safe condition. The inspections follow a schedule, which defines the frequency of inspection. This frequency depends on a number of factors. A dam that has not been properly inspected by experts for some years or a new or a reconstructed dam should be inspected rather frequently. The frequency decreases from the construction and impounding periods towards the long-term operation. However, inspection must be performed during the whole lifetime of a structure. Apart from pre-designed schedules the frequency of inspections should be increased at times of exceptional events, such as floods, storm-induced wave action, earthquakes and the like. It is good practice to have inspections under variable operating conditions such as:

- Reservoir level down, so that the upstream face and abutments can be visited.
- Reservoir full. This allows inspection of leakage or piezometer pressure under maximum head conditions. It also helps the inspector to assess hydraulic condition of spillways.

Generally speaking, the opinion of very many individual engineers and panels is that measured data of instrumentation and visual inspections results are both necessary and that a dam cannot be considered safe unless both lead to a favorable conclusion (Post, 1985).

The dam safety engineer will have to combine results of inspection (visual data) and instrumentation data (quantitative) for a dam safety analysis. Monitoring device conditions are assessed in a site inspection and are related to a performance level (ability to monitor). Monitoring device importance factors are determined based on their overall diagnostic value for the safety assessment of a dam. An optimization strategy using a ranking equation based solely on the condition assessment and device importance determination is then used for combining the information. Anderson et al. (1999) propose a ranking procedure for the prioritization of maintenance and rehabilitation tasks on the performance monitoring devices for embankment dams. The priority ranking is the product of the loss of diagnostic function and the importance of that function. It is assumed that the most important devices in the worst condition should be given priority. Device condition are elicited from an expert panel. Condition is defined in terms of ability to function as a monitoring device. The diagnostic value for each of the monitoring devices is determined by the utility function that is based on: (1) subjective conditional probabilities of potential failure modes, (2) subjective conditional probabilities of specific adverse conditions related to those failure modes, (3) diagnostic values of indicators that suggest the presence of the adverse conditions, and (4) the diagnostic value of the monitoring devices for these indicators. This diagnostic value can be updated as conditions on the dam change over time.

Table 4.4 shows different scenarios for probabilities of anomaly detection. As it can be seen there are two types of errors. It is important to take into account the probabilities of both type I and type II errors in instrumentation analysis and design. The probability of type I error is denoted by alpha (α), the level of significance of an error. Significance level is defined as the degree of uncertainty about the statistical statement:

α=Level of significance = P (type I error)=P (reject $H_0 |\ H_0$ is true)

When a statistical test is performed to compare validity of the competing hypothesis statements, the result will cause the null hypothesis H_0 to be either rejected or accepted.

β= P (type II error)=P (accept $H_0 |\ H_0$ is false).

Table 4.4, Probabilities of fault detection

		True Behaviour	
		Fault Present	Fault Not Present
Detection by surveillance	Detected	Good	Type I errors
	Not detected	Type II errors	Good

In statistical testing, it is desirable to keep the probability of a type I error as low as possible. This can be done easily by using small values of α. However, there is a relationship between type I and II errors. As type I errors decrease, type II errors increase.

4.4 Recent trends in Performance monitoring

Some current trends in sensor development and measurement technology that have an impact on dam monitoring methods and equipment are listed below.

- Instruments are more accurate and there is a larger selection to choose from.
- Generally, equipment costs will decrease with time.
- Improved corrosion resistance of sensors due to increased use of noble alloys and materials, for example the use of titanium for wetted parts in pore pressure sensors.
- Intelligent instruments or "smart sensors" are becoming more common and their level of intelligence is constantly being elevated. Some important advantages of smart sensors are: 1) built-in capacity for self-checking and automatic warning of malfunctions, 2) automatic compensation for nonlinearitly and hysterisis errors or systematic errors due to temperature drift, 3) networking capability thereby allowing a number of instruments to be connected to the same instrument cable, thus reducing the amount of cables that have to be installed. The cost of cables and cable installation and terminal work accounts for a large portion of the total cost of monitoring systems.
- Increased use of optical instruments, for example computer controlled surveying instruments for geodetic measurements of deformations.
- Laser position measurement systems for precise static and dynamic displacement measurements.
- Digital photographic techniques combined with image processing for displacement and deformation measurements.
- Improvement in and increased use of optical fibre sensors for monitoring pressure, strain, temperature and displacement, including distributed systems where measurements can be made at many points along a single fibre. The principal advantage is their immunity to electromagnetic interference and elimination of damage caused by induced voltages during electrical storms (Dibiagio, 2000). It is a simple matter to replace conventional electrical cables with fibre optic cables. The equipment needed to do this is now available and easy to use. Likewise there are modems and converters available that can be used to convert and transmit the output

signals from many types of sensors over fibre optic cables. However, with today's technology it is not feasible to replace conventional sensor cables entirely with optical cables. The reason for this is that most sensors require some form of electrical power to operate them. This can not be done as technology available for transmitting power through fibre optic cable that is too expensive, too complicated, or too limited.

- More accurate GPS (global positioning satellite) equipment and methods for monitoring displacements.

Agencies responsible for dam safety have long used conventional surveying methods to measure the displacements of benchmarks as part of dam monitoring programs. Such surveys have provided infrequent though precise estimates of the motion of a dam. With the development of high precision GPS methods to monitor plate tectonic motions and crustal deformation rates, an alternative method for monitoring such structural motions is available (Hudnut et al, 1998). Benefits of GPS lies in a much higher temporal resolution and nearly unattended continuous operation. Figures 4.5 and 4.6 show an example of the use of GPS on the Pacioma dam.

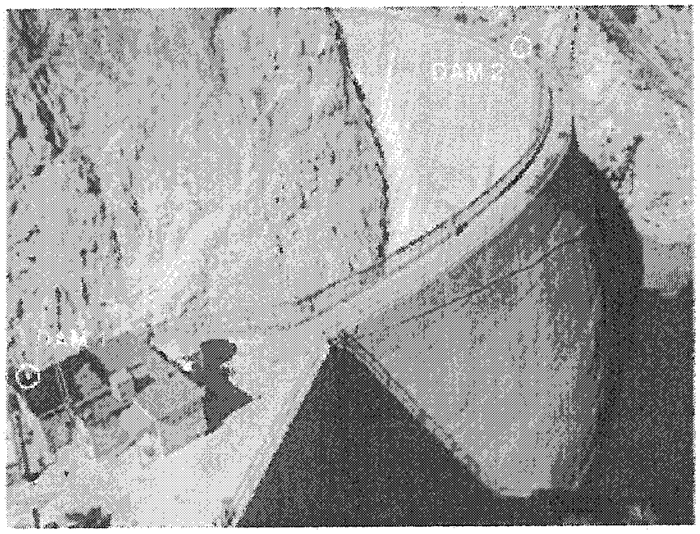


Figure 4.5, Pacoima Dam.

4.4.1 Automatic Data acquisition and processing

The trend towards automated monitoring systems started in the 1970's and eliminated much of tedious work involved in data processing and presentation. Subsequent rapid advances in electronics and computer technology coupled with decreasing costs and extended use of electrical sensors spurred considerable interest in automation of monitoring systems for dams. Automatic data acquisition and processing have now become an integral part of the vast majority of dam monitoring programs for new dams. Recent major advances in electronics, communications and computers have made it possible that readings to be taken automatically on dams and on their periphery (Bordes et al., 1998).

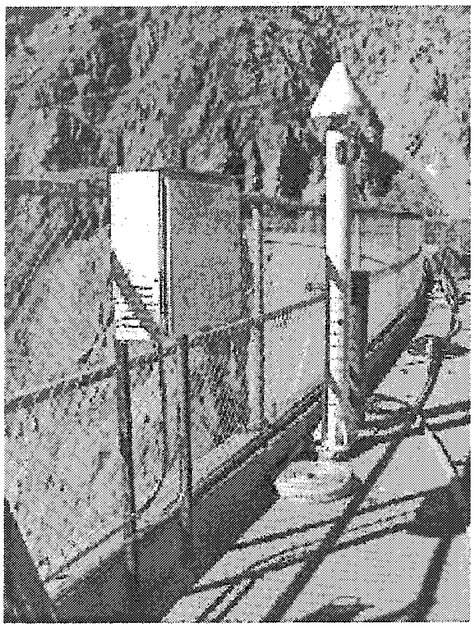


Figure 4.6, The antenna is mounted on a steel pier, and GPS receiver

(Courtesy: Hudnut et al., 1998)

Automation of instrumentation can assist in the assessment of the safety of dams. This is particularly true for monitoring that requires rapid and frequent data collection or for instruments that are inaccessible. In recent years, the technology of devices for measuring seepage, stresses, and movements in dams has improved significantly with respect to accuracy, reliability, and economics. Although the initial installation of an Automated Data Acquisition System (ADAS) may appear to be more expensive than traditional instrumentation systems, the overall long-term cost, in many cases, is now economically competitive. Automation should receive consideration for all systems that are to be installed during new dam construction, major rehabilitations, structural modifications, or any major effort that would support a major instrumentation system. Instrument upgrades

Table 4.5, Advantages and limitations of ADAS

Advantages	Limitations
Increased accuracy, reduced human error.	Produces large volumes of data; overtaxes storage medium
Increased frequency, more data, less system error.	Potentially higher maintenance costs.
Increased data reliability and consistency.	Installation could be expensive.
Timeliness of information, obtain data whenever needed.	Lightning; variable voltage potential Is destructive.
Data and system validity checks enhance data quality.	Excessive downtime
Alarms for exceeding data thresholds and system health.	Requires use of electronic transducers which have least long-term reliability

and replacements could be justified on a case-by-case basis. The instrument automation concept generally includes an instrument or transducer that is linked to a data-logger or computer with communication capability that allows data retrieval locally or from a remote location. The advantages and limitations of an automated system are summarized in Table 4.5. The limitations can be minimized with appropriate attention to planning and use of the system (USACE, 1987).

These readings are taken by Measurement and Control Units (MCUs) that are typically located in close proximity to the instruments they measure. The readings can be a) transmitted to an on-site computer immediately through an on-line network, b) stored temporarily and transmitted periodically to a remote computer using standard telecommunication facilities. MCUs have complete optional control features which can be programmed to run automatically, or respond in a programmed way to remote commands. This allows easy and fast excecution of certain corrective measures. ADAS provides information on a more frequent and uninterrupted basis than manual instrument readings. This becomes especially important during severe weather events, when real-time data is critical for making informed decisions. Adjusting the monitoring rates are easy to implement with an ADAS. As dams age, maintenance needs and safety concerns increase. Dam safety officials are putting greater emphasis on automated instrumentation and use of data acquisition systems to reduce costs of inspection and monitoring. The collection and analysis of large quantities of data, especially over long distances, requires centralized and automated measuring techniques. Data can be processed more rapidly thus enabling efficient alarm systems to be implemented when predetermined thresholds are exceeded.

CHAPTER 5

5. Case studies of dam monitoring for Arch dams

5.1 Introduction

The safety of a dam is determined by its design, construction and supervision during operation. High arch dam failures have dropped dramatically since the early part of century. An essential part of this improvement relates to improved measurement techniques that can make earlier detection of unexpected behavior. The overall safety of a dam can be assessed by the analysis of available measurements. The current practice is to perform statistical analysis of individual instruments using linear regression methods. Confidence intervals are used to set the alarm levels for observed values. However, a large number of instruments must be analyzed, and it can be very difficult to estimate statistically significant deviations from normal readings for individual instruments, given that the fluctuations in stresses, strains, or deformations are small, and in the order of magnitude of noise in some measurements. Data reduction methods can be helpful to overcome these difficulties and provide tools for better management and analysis of dam monitoring data. Some of the multivariate analysis methods discussed in Chapter 3 are applied to the Idukki dam data set.

5.2 Idukki Dam

The Idukki arch dam is situated in south India in the Periyar valley in the state of Kerala. It was the first concrete arch dam to be built in India. It is a double curvature, parabolic, thin asymmetrical arch dam and was constructed between 1969 and 1974.

The 169-m high dam is made up of 24 individual blocks and is seated on massive tightly

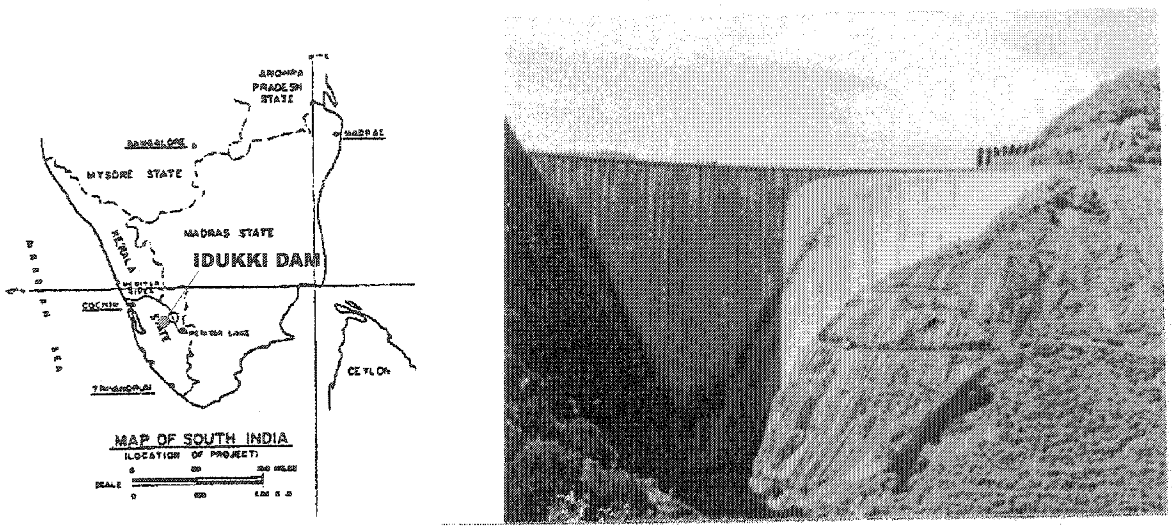


Figure 5.1, Location and view of Idukki arch dam

jointed hypersteine granite of excellent quality. The crest of the dam is 381 m long by 7.6m wide, and the thickness at the base is 24.4m. The concrete for Idukki dam was designed to withstand a maximum compression of 35 MPa and a tension of 1.1 MPa. Arch dams are designed to carry external loads by compression. The major effort in arch dam design is to adjust the geometry to minimize the extent and magnitude of tensile stresses. However, tensile stresses can not be completely avoided and they do exist in some arch dams, and often result in cracks (Veltrop et. al, 1990). The monitoring system of the Idukki arch dam is described next.

5.2.1 Monitoring

Different types of instruments are installed to monitor the deformations of the dam and of the foundations, and the stresses in the arch dam. (Table 5.1). Records of these data have been kept since 1974. The accuracy of a portion of the data may be questionable, particularly regarding the stress, strain measurements, since these instruments are sealed in the concrete and cannot be inspected, or their readings validated by independent means.

Reservoir level

The fluctuation levels of the Idukki reservoir have been monitored on a daily basis since April 1974. Reservoir variations follow a regular cyclic pattern of almost 12 months. Maximum and minimum yearly reservoir levels are in November and June of each year (Figure 5.2). The maximum reservoir level of 731m was reached on Sept 1981. The minimum reservoir level after the filling period was 695m recorded in June 1983.

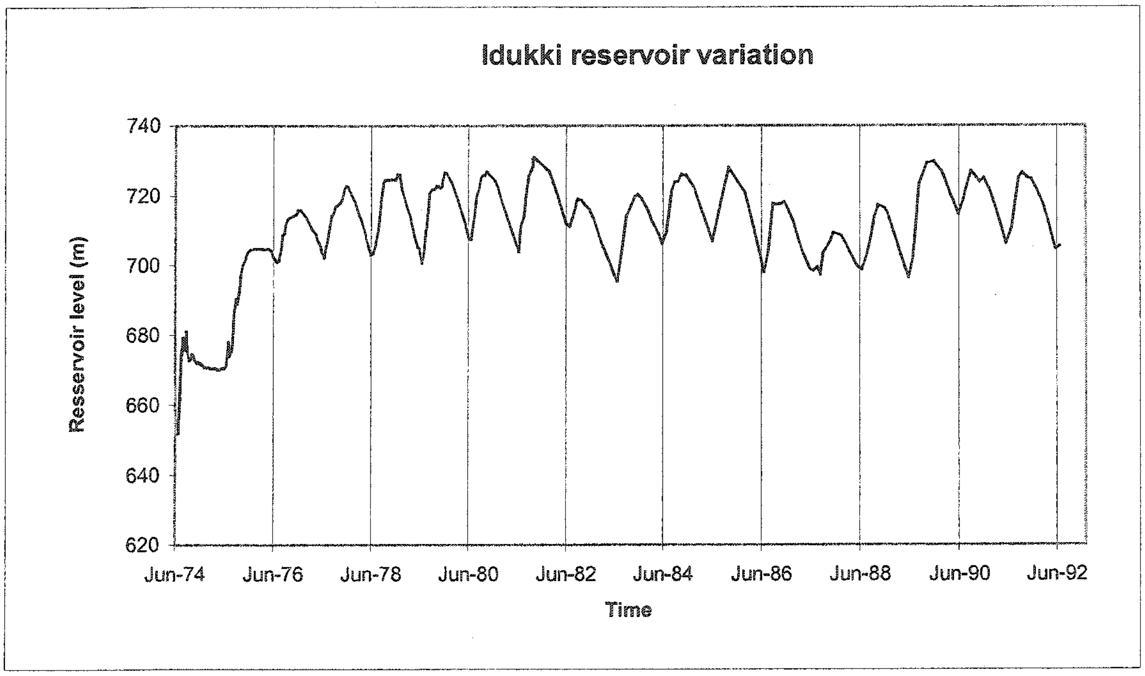


Figure 5.2, Reservoir level fluctuation, Idukki dam

Temperature

Daily air temperatures were recorded at the Idukki site. Minimum and maximum temperatures occur in the months of October and June of each year respectively. The maximum recorded temperature was 32.2 °C in May 1990.

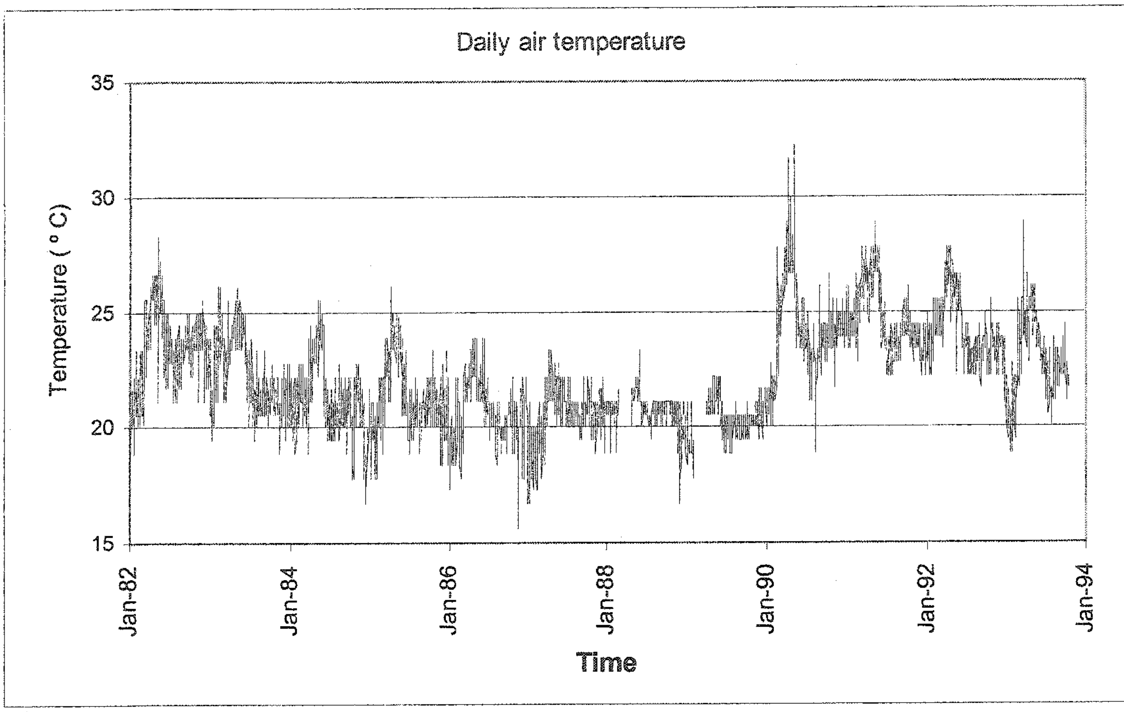


Figure 5.3, Daily air temperature Variation, Idukki site

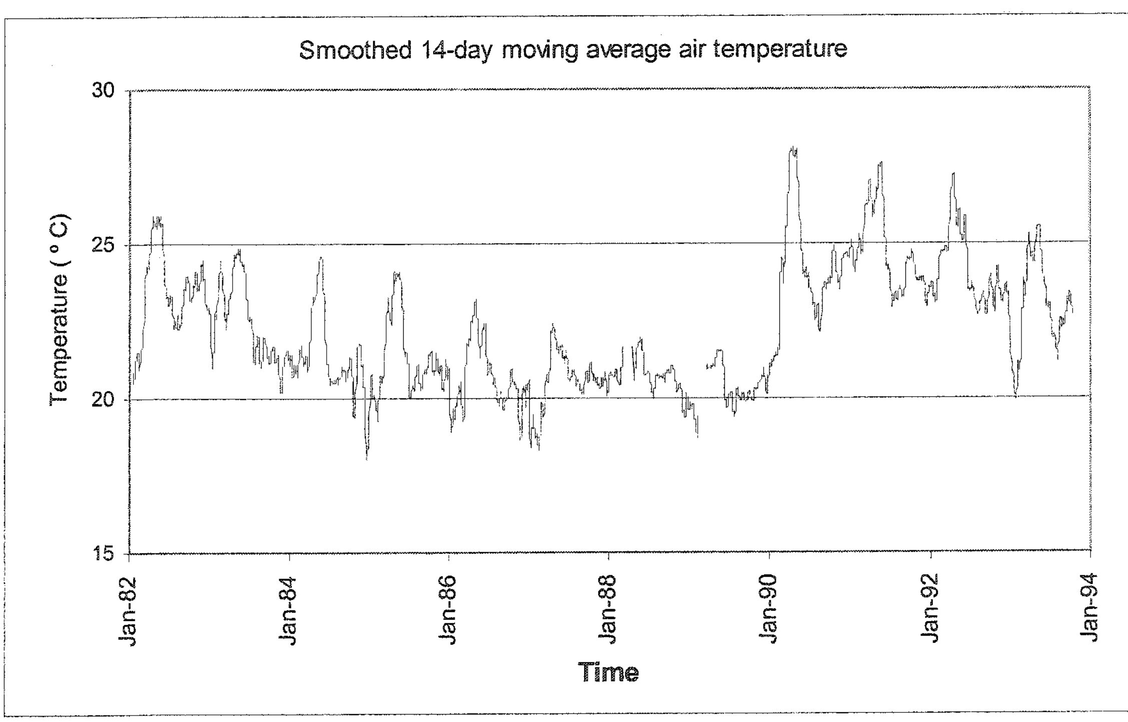


Figure 5.4, Smoothed 14-day air temperature, Idukki site.

Temperature variations are different for 1990 onward and show a trend of warmer temperature. Smoothed temperatures are presented in Figure 5.4. There were also 8 resistance-type water thermometers embedded in the upstream face of the dam between the 579 m and 731 m levels. The data from the five operative instruments shows that the water temperature at the bottom of the reservoir is stable and colder than in the upper parts of the reservoir.

Table 5.1 shows the list of instruments that are installed in Idukki dam. These instruments are either classified as global or local instruments. A local instrument such as an instrumented cylinder measures the strain at a specific location of the dam. A global instrument mainly integrates response parameters (strains) over large volumes of the dam, such as inverted pendulums. Measurements from these instruments tend to be less variable due to averaging. The last column shows number of readings in a year. Instrumented cylinders are among the instruments, which are monitored very frequently; while the stress meters and strain meters are monitored only twice per year, at maximum and minimum reservoir level.

Table 5.1, Type and frequency of reading for different instruments

Measures	Instrument	Global	Local	Number of readings (No/year)
Displacement	Pendulum	*		12
	Crest Collimation	*		4
	Base Meters	*		4
	Rock targets	*		4
	Clinometer	*		12
	Electronic Joint meter		*	
Stress/Strain	Strain meters		*	2
	Stress meters		*	2
	Cylinders		*	104

5.2.1.1 Deformations

There are two types of instruments that are installed in the dam to measure and record the horizontal deformations of the dam: (1) Pendulums, and (2) Crest collimation targets. Pendulums are among the most reliable instruments for measuring the horizontal movements of the dam (KSEB, 1989). A total of 6 pendulums were installed in blocks 1,7 and 8 (Figure 5.5). Simple Pendulums P2, P3, P4 and an inverted pendulum P5 are installed at elevations 2400 ft (732 m), 2300 ft (701 m), 2100 ft (640 m) and 1900 ft (579 m) of the central block (block 1) respectively (Figure 5.6). Pendulums P1 and P6 are installed in blocks 7 and 8 to measure movements at elevations 2100 ft (640 m) and 2300 ft (701 m) respectively. These instruments measure both radial and tangential displacements. Pendulum P5 measures absolute displacements of the dam at elevation 1900 ft (579 m). This displacement is added successively to displacements from other pendulums to obtain their absolute displacements. Displacements on March 1977 were selected as a reference to compare the pendulums and crest collimation results. Therefore displacements represent the relative displacement from this date. The recorded reservoir level was 711.44 m at the reference level.

Average radial displacements are calculated for different reservoir level intervals (Figure 5.7). The dam moves upstream when reservoir levels are less than the reference reservoir level and moves downstream for reservoir levels greater than this reference reservoir level.

Figure 5.8 and Figure 5.9 show the tangential and radial displacements in the central block (block 1). The radial displacements are highly correlated with reservoir level except at the base (elevation 579 m), where the displacements are in order of 1-2 mm. As can be expected radial displacements show higher correlation with reservoir level than tangential displacements.

Displacements of pendulum P4 at elevation 2100 ft (640 m) are downstream for reservoir levels greater than the reference reservoir level of 711 m, and upstream for reservoir

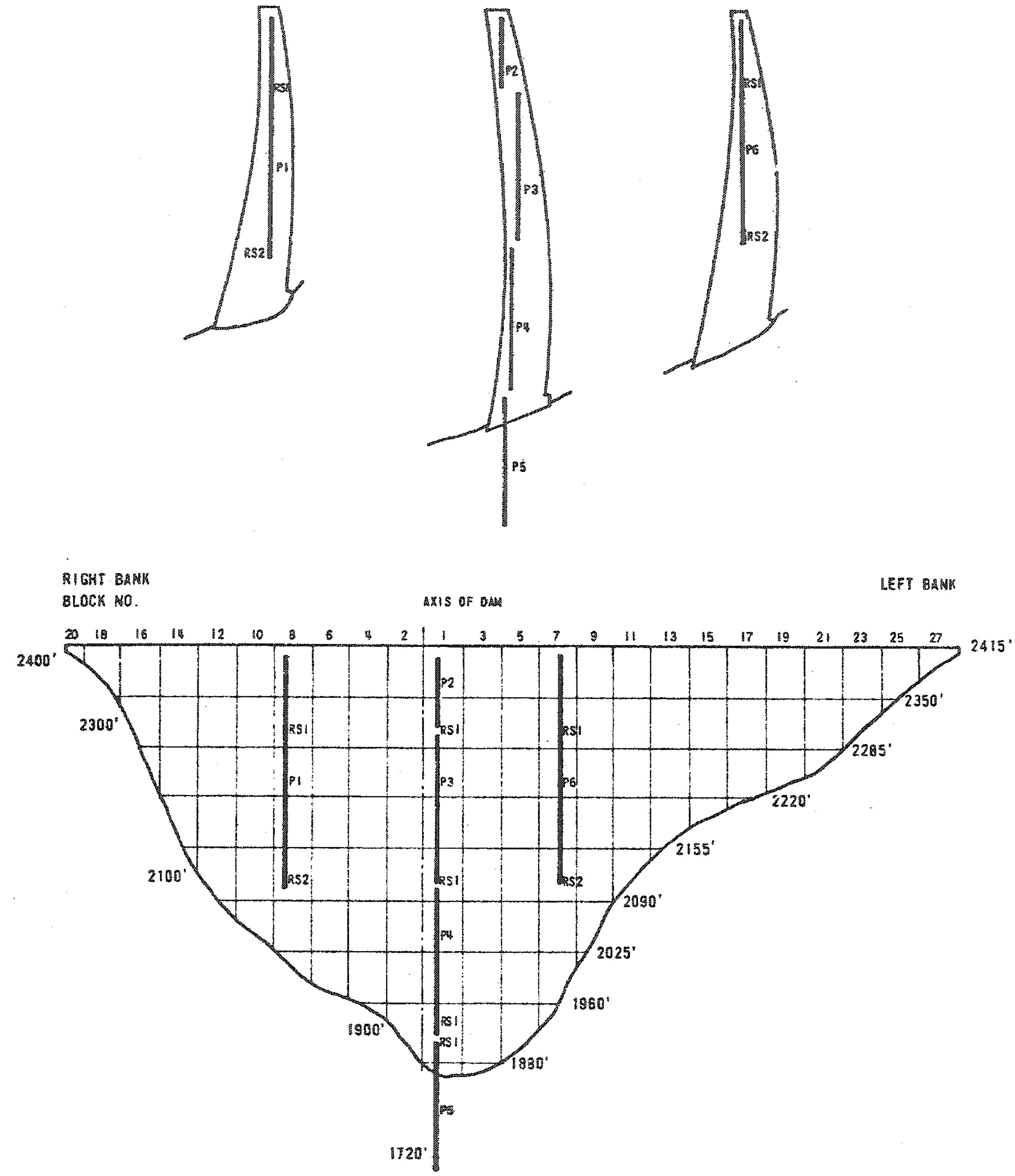


Figure 5.5, Location of pendulums, Idukki dam

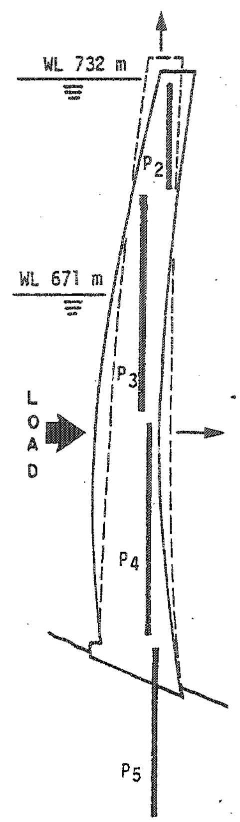


Figure 5.6, Pendulums installed in block 1

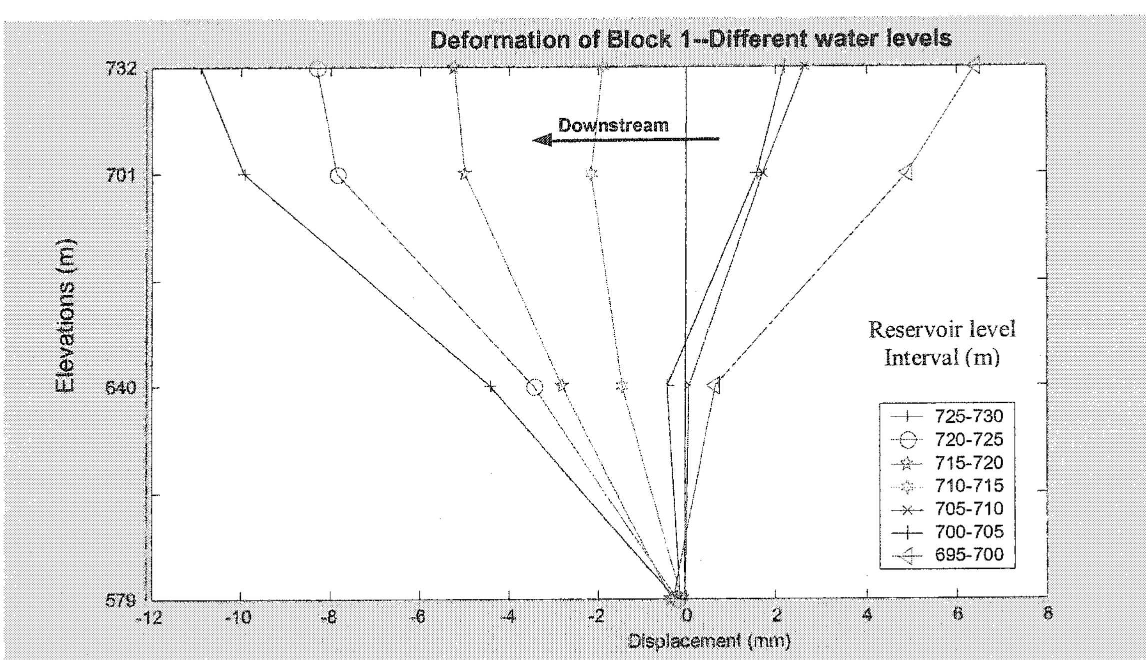
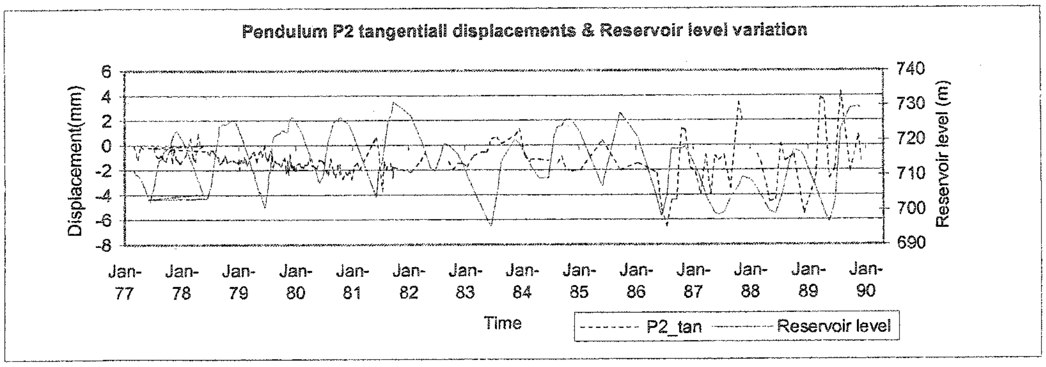
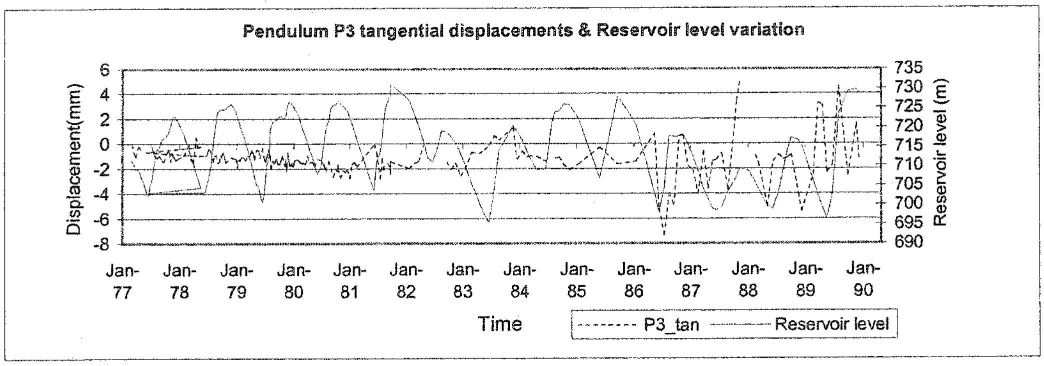
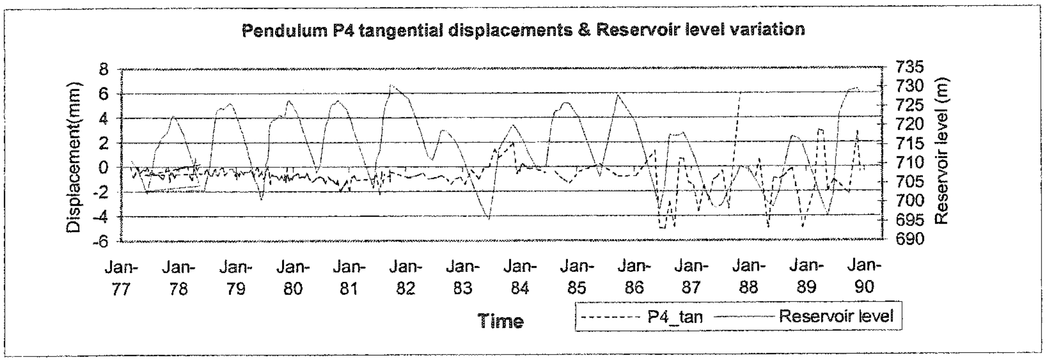


Figure 5.7, Displacements of Block 1 at different reservoir water levels elevations







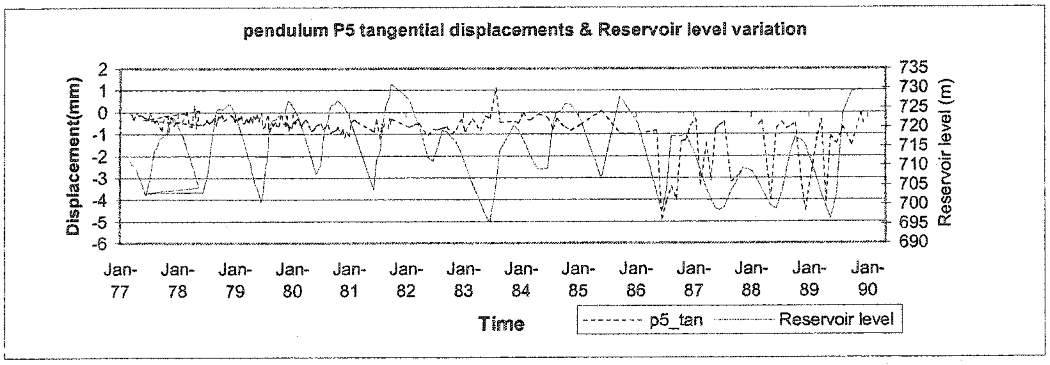
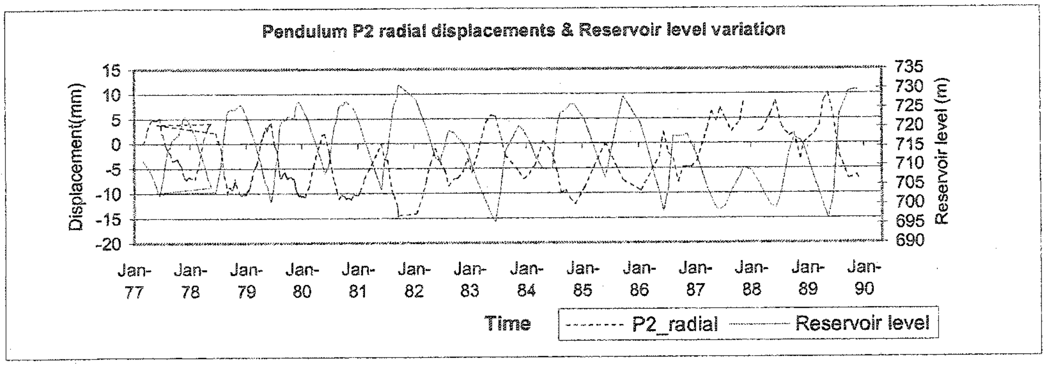
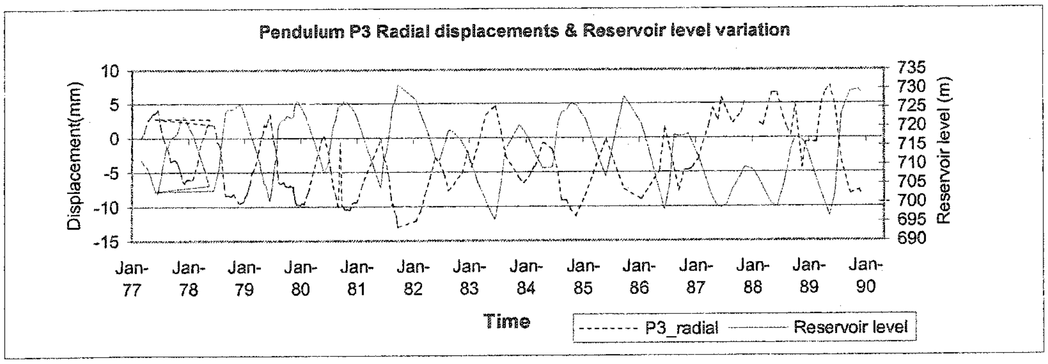
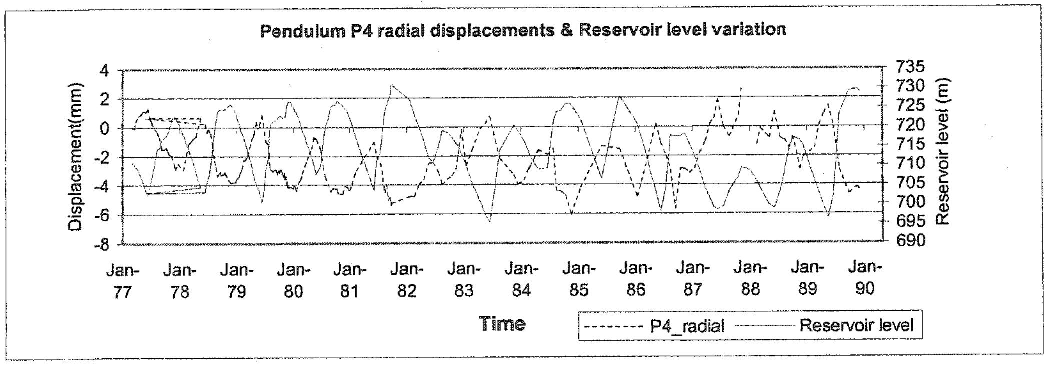


Figure 5.8, Tangential displacements of Block 1 and reservoir level variations







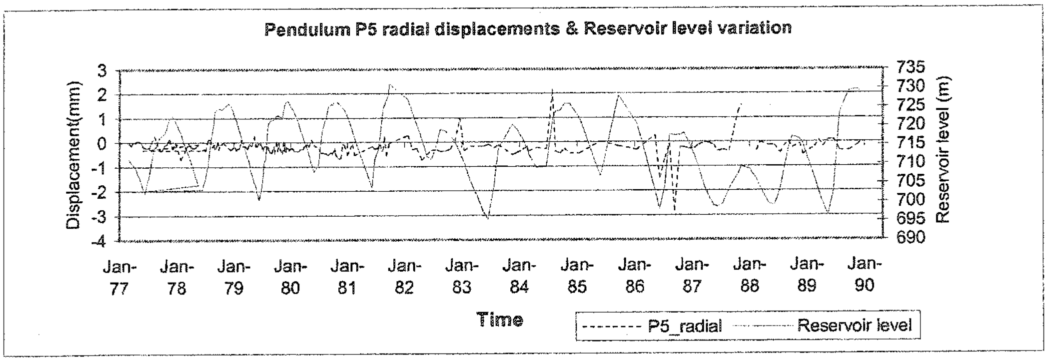


Figure 5.9, Radial displacements Block 1 and reservoir level variations

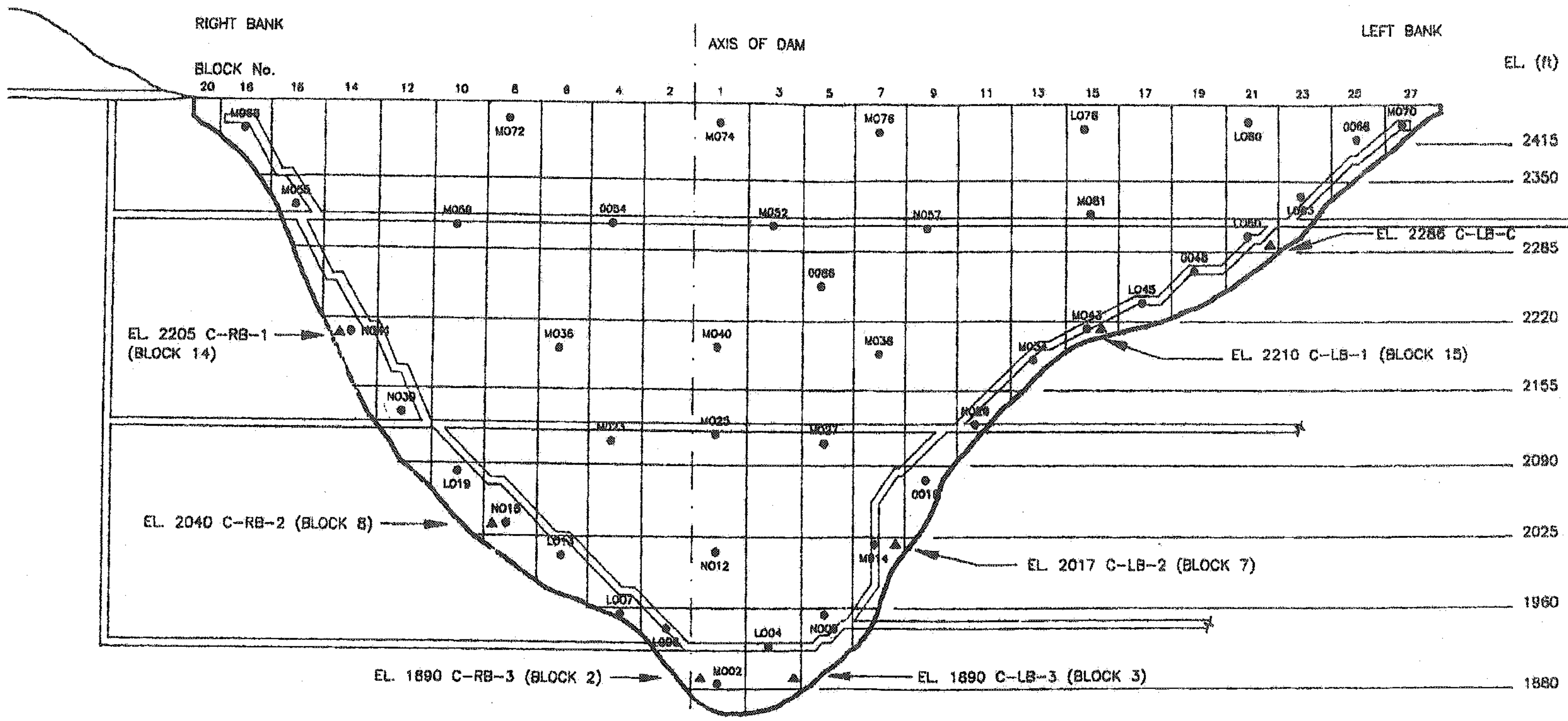


Figure 5.10, Instrumented cylinders and strain meters, Idukki dam

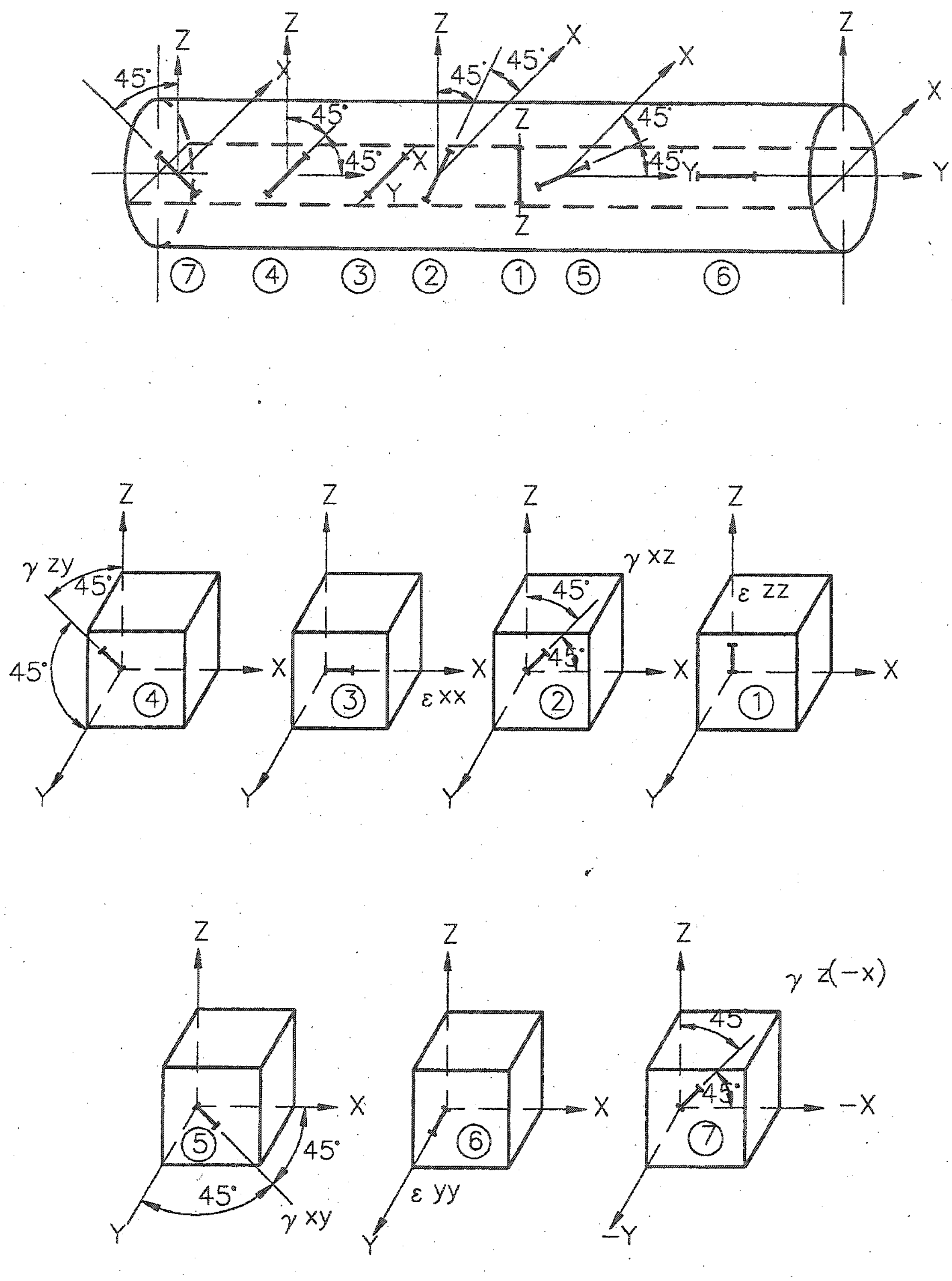


Figure 5.11, Instrumented cylinders

levels less than 711 m. The maximum radial displacement is 6.05 mm downstream on Nov. 1984 for a reservoir level of 725.8 m. The maximum upstream movement is 1.86 m on June 1987 for a reservoir level of 698.6 m. The pendulum at elevation 2300 ft (701 m) shows maximum downstream and upstream displacements of 13.1 and 7.7 mm respectively.

5.2.1.2 Crest Collimation

Displacement of the crest is measured by shifting a movable target on the crest in a direction perpendicular to a fixed line of sight. Movable targets are installed over blocks 1,7 and 8 where, the readings were taken on a monthly schedule. The maximum downstream movement for the period of 1977-1989 was 15 mm at reservoir level of 730m in September 1981. The maximum upstream movement for this period was 10.1 mm in block 1 during May 1989 at reservoir level of 696 m.

5.2.1.3 Stress/strain measurements

Initially 82 rosettes of vibrating wire strain meters were embedded into the body of the dam to measure the strain and stress in different points of the dam (Figure 5.10). One half were installed 3 m deep in the downstream wall and the other half placed 3 m deep in the upstream wall. Some of these instruments were not functional and some were giving unreasonably high readings. In general, tension has been recorded for most of the instruments. The readings from the strain meters are unreliable on the whole and are no longer dependable (KSEB, 1989). High tensile strain/ stress measurements obtained from the instruments are not consistent with results from the visual inspection of the dam. Some of these instruments had to be replaced in order to continue monitoring stresses whitin the dam. Since the strain-stress state of arch dam is very important in monitoring of the dam behavior, new rosettes of vibrating wire strain meters, placed in cylindres instrumenté de l'Universite de Sherbrooke (CIUS), were used to replace some of the more strategically placed instruments. Each concrete cylinder contains seven vibrating wire extensometers. The instrumented cylinder is placed in a 6-inch diameter borehole. The instrumented cylinders are fully described in Simard et al (1993), and Ballivy et al

(1993). Each concrete cylinder comprises vibrating strings from which the three dimensional state of strain at a point can be measured (Figure 5.11). The cylinders were initially installed in October 1991 to validate measurements from the network of strain gauges installed at the time of initial construction. There are a total of six instrumented cylinders installed in the dam.

5.2.2 Statistical analysis of stresses in Idukki dam

A very important part of an arch dam monitoring system is the analysis of stresses to validate and control its safety. The stress meters installed in Idukki measure only the stresses perpendicular to their axis and are located in the plane of the arch of the dam in a vertical or horizontal position.

These instruments are not read frequently and consist only of one measurement during the summer season and one measurement during the winter season corresponding to minimum and maximum reservoir levels. In order .to properly identify stress patterns inside the dam, readings should have been taken much frequently. Preliminary analysis of these instruments indicated that reliability of many of the instruments is questionable.

Due to the poor quality and small number of readings of stress meters and strain meters, it was difficult to obtain good results from the statistical analysis. Instead statistical analysis were performed on instrumented cylinders, which are read twice weekly. Out of 6 cylinders originally installed in the dam, two are defective and their readings are not acceptable. The four remaining cylinders were considered for statistical analysis. There are a total of 24 measurements, which are recorded twice weekly for a period of 17 months.

To eliminate spikes from the data set, the time series of first differences (rate of strain change) was calculated for each component of the 24 readings. Mean and standard deviations were calculated for each component. Considering the sample size, all the data points greater than three-standard deviations from the mean were removed (Myers, 1995).

The strain matrix can be represented by:

$$\begin{bmatrix} \varepsilon_{xx} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{zy} & \varepsilon_{zz} \end{bmatrix} \tag{5.1}$$

The strain matrix has three eigenvalues with three corresponding orthogonal eigenvectors. These three eigenvectors correspond to normal strains, referred to as principal strains, which exist on mutually orthogonal surfaces that contain no shear strain. The principal strain tensor can be expressed by:

$$\begin{bmatrix} \varepsilon_{p_1} & 0 & 0 \\ 0 & \varepsilon_{p_2} & 0 \\ 0 & 0 & \varepsilon_{p_3} \end{bmatrix} \tag{5.2}$$

in which principal strains ϵ_{p1}, ϵ_{p2}, and ϵ_{p3} are the solution of :

$$\begin{vmatrix} \varepsilon_x - \varepsilon_{p_1} & \varepsilon_{xy} & \varepsilon_{xz} \\ \varepsilon_{yx} & \varepsilon_{yy} - \varepsilon_{p_2} & \varepsilon_{yz} \\ \varepsilon_{zx} & \varepsilon_{yz} & \varepsilon_{zz} - \varepsilon_{p_3} \end{vmatrix} = 0 \tag{5.3}$$

When principal strains are obtained, they can be used to calculate principal stresses:

$$\begin{bmatrix} \sigma_{p_1} \\ \sigma_{p_2} \\ \sigma_{p_3} \end{bmatrix} = \frac{E\nu}{(1+\nu)(1-2\nu)} \begin{bmatrix} \varepsilon_{p_1} + \varepsilon_{p_2} + \varepsilon_{p_3} \\ \varepsilon_{p_1} + \varepsilon_{p_2} + \varepsilon_{p_3} \\ \varepsilon_{p_1} + \varepsilon_{p_2} + \varepsilon_{p_3} \end{bmatrix} + \frac{E\nu}{1+\nu} \begin{bmatrix} \varepsilon_{p1} \\ \varepsilon_{p2} \\ \varepsilon_{p_3} \end{bmatrix} \tag{5.4}$$

Where E and ν are the elastic constants of concrete. Principal strains are calculated from Eq. 5.3. The results indicate that instrumented cylinder RB2 measures the maximum principal tensile (Figure 5.12) and compressive stain (Figure 5.13), and is more critical than the other three instruments. High tensile strains are observed in three of the instruments. However, no cracks have been observed to validate the true state of stress. Nonetheless, the data from cylinders can be used to monitor changes in the state of stress in relative terms.

5.2.2.1 Multivariate linear regression

The goal of multiple linear regression analysis is to determine the dependency of the maximum tensile principal strains (corresponding to cylinder RB2) with the reservoir

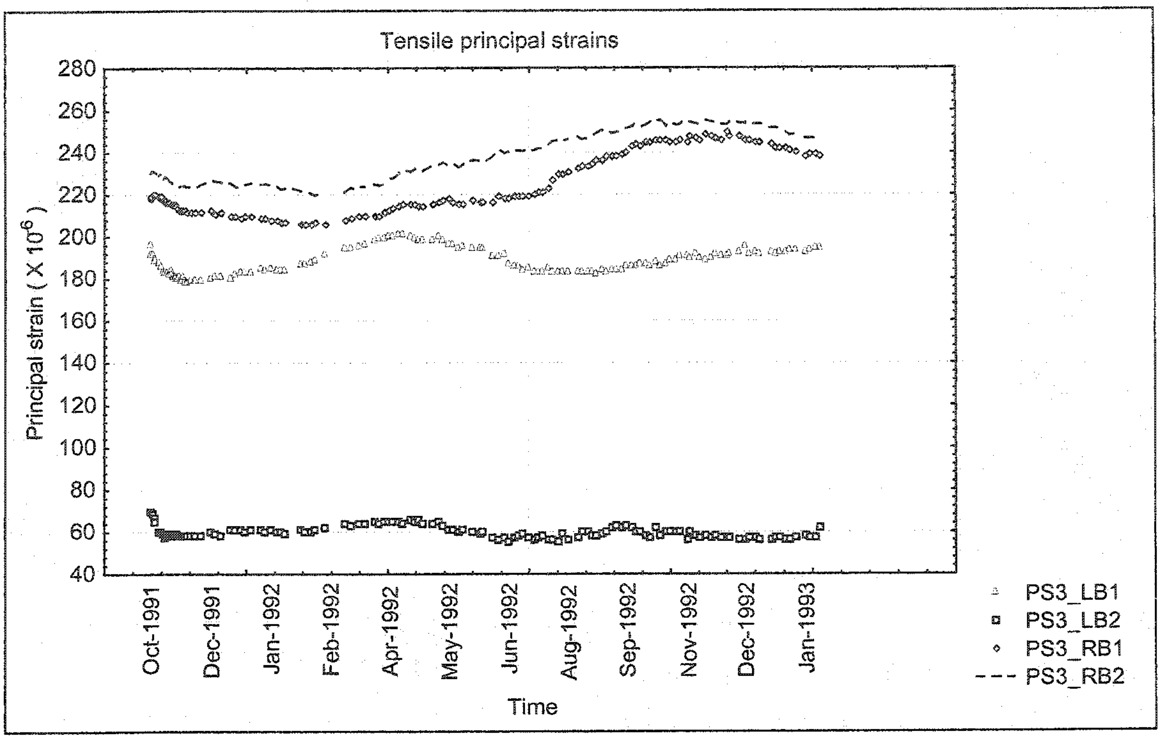


Figure 5.12, Tensile principal strains in four instrumented cylinders

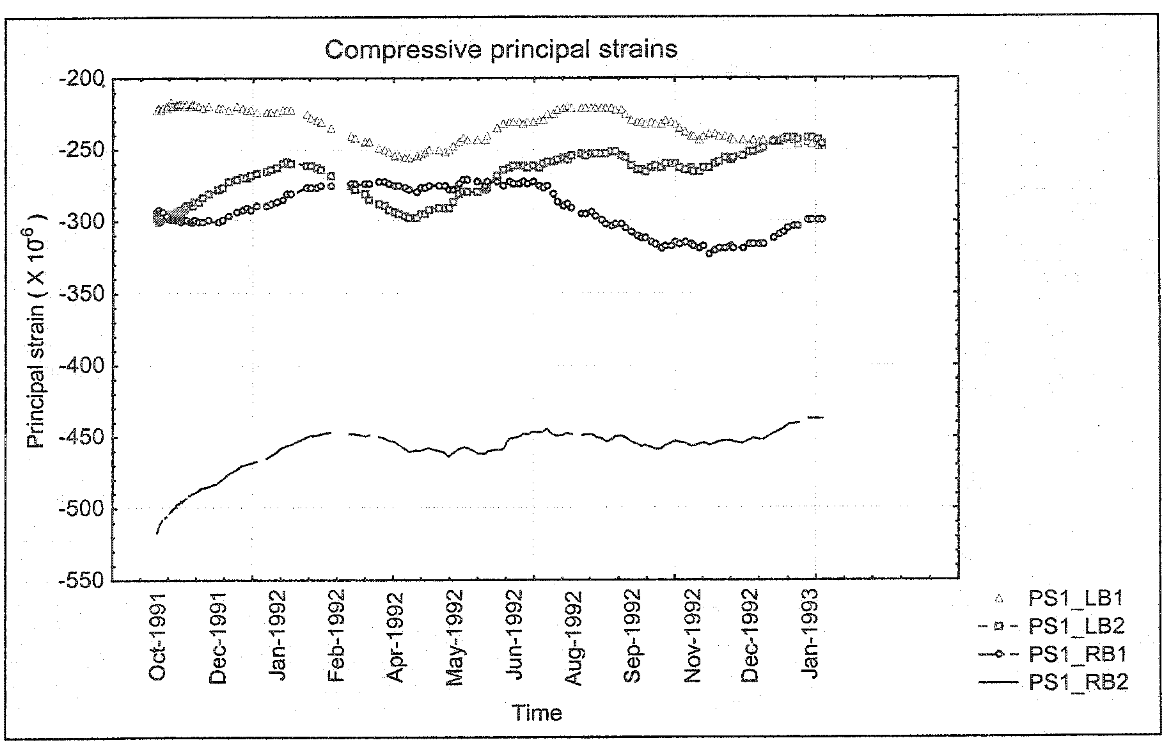


Figure 5.13, Compressive principal strains in four instrumented cylinders

level, time, ambient temperature and internal temperature of concrete. The prediction model can be used to check future readings, establish confidence intervals, and set alarm levels.

In dam monitoring practice two types of statistical models are used. The first category of models is based on ambient temperature, reservoir level, and time effects. The second category of models is the HST model, in which seasonal effects are used instead of temperature data. The HST method provides good results in many cases but is inefficient in predicting of the responses for abnormal temperature cycles. The HST model is used to model the tensile principal strain of cylinder RB2 (denoted PS_RB2 hereafter). Periodic functions (Eq. 5.7), and polynomial of reservoir level (Eq. 5.6) and time functions (Eq. 5.8) are considered.

$$D(t) = H(z) + S(\theta) + T(t) \qquad 5.5$$

$$H(z) = a_1 + a_2 z + a_3 z^2 + a_4 z^3 + a_5 z^4 \qquad 5.6$$

$$S(\theta) = a_6 Sin(\theta) + a_7 \cos(\theta) + a_8 Sin(\theta)Cos(\theta) + a_9 Sin^2(\theta) \qquad 5.7$$

$$T(t) = c_1 t' + c_2 t'^2 + c_2 t'^3 \qquad 5.8$$

where $z = \dfrac{H - H_{min}}{H_{max} - H_{min}}$, $t' = t - t_0$, and $\theta = \dfrac{2\pi t'}{365}$

where H_{min}, H_{max}, are respectively minimum and maximum reservoir water level.

Every regression model must be validated. One practice is to use cross validation by splitting the data into estimation and validation sets. A model is obtained on the estimation set and verified with the validation set. However, this is not easily done with the standard statistical software packages. Press (Prediction Sum of Squares) statistics is a criterion that can be used as a form of validation. Press statistics or deleted residuals are readily available in commercial statistical packages (Myers, 1990).

Consider a set of data in which the first observation is withheld from the data, and the remaining (n-1) observations are used. The first observation is then replaced and the second observation is withheld to estimate the coefficients for a model. The procedure is repeated so that the model is fitted n times. The deleted response is estimated for every data point, resulting in n prediction errors or press residuals. An observation is not used in estimating its estimated press residuals consequently this is a form of validation. Press residuals are calculated from:

$$e_{i,-i} = \frac{y_i - \hat{y}_i}{1 - \underline{x}'_i(\underline{X}'\underline{X})^{-1}\underline{x}_i} = \frac{e_i}{1 - \underline{x}'_i(\underline{X}'\underline{X})^{-1}\underline{x}_i} \qquad 5.9$$

where $\underline{x}_i$ is the vector of the predictor variables at a particular time The PRESS statistics is defined as:

$$PRESS = \sum (e_{i,-i})^2 \qquad 5.10$$

checks are also made to avoid overfitting and underfitting the data set

The best subset of predictor was selected on the basis of C_p statistics. If p predictor variables are selected the C_p is defined as:

$$C_p = \frac{\sum_{i=1}^{n} (\underline{y} - \underline{\hat{y}})^2}{s^2} + 2p - n \qquad 5.11$$

where

$\underline{\hat{y}}$ is the predicted value of y from p predictors

s^2 is the residual mean square after regression on the completed set of k variables

n is the sample size

Stepwise regression summary is presented in Table 5.2. The periodic functions, reservoir level variation and time effect explain about 98% of variation of principal strain of cylinder RB2. The variables in order of importance are time and periodic functions. Parameter estimates and associated t-test values are presented in Table 5.3.

Residuals are normally distributed, have no bias, and have a constant variance (Figure 5.14), and model performance is judged to be satisfactory (Figure 5.15). Predicted values and prediction intervals are shown in Figure 5.16. Note that the prediction intervals can be used to set alarms for detecting significant changes in the principal strain. Hydrostatic, seasonal and irreversible component of the HST model is presented in Figure 5.17.

Table 5.2, Stepwise regression summary of principal strain of cylinder RB2

Variable	Summary of Stepwise Regression Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variabls included
t'	1	0.908	0.825	0.825	622.340	0.000000	1
Cos(Θ)	2	0.977	0.954	0.129	363.813	0.000000	2
Sin(Θ)	3	0.987	0.973	0.020	96.212	0.000000	3
z^4	4	0.991	0.981	0.008	54.216	0.000000	4
Sin(Θ)Cos(Θ	5	0.992	0.983	0.002	13.845	0.000296	5

Table 5.3, Regression summary of principal strain of cylinder RB2

Regression Summary
R= .991 R²= .983
Std.Error of estimate: 1.5819

N=134	Beta	Std.Err. of Beta	B	Std.Err. of B	t(128)	p-level
Intercept			187.402	1.659	112.986	0.000000
t'	0.749	0.034	0.060	0.003	22.133	0.000000
Cos(Θ)	-0.196	0.033	-3.418	0.570	-5.995	0.000000
Sin(Θ)	0.256	0.022	4.537	0.395	11.475	0.000000
z^4	0.267	0.049	10.090	1.856	5.437	0.000000
Sin(Θ)Cos(Θ	0.050	0.013	1.791	0.481	3.721	0.000296

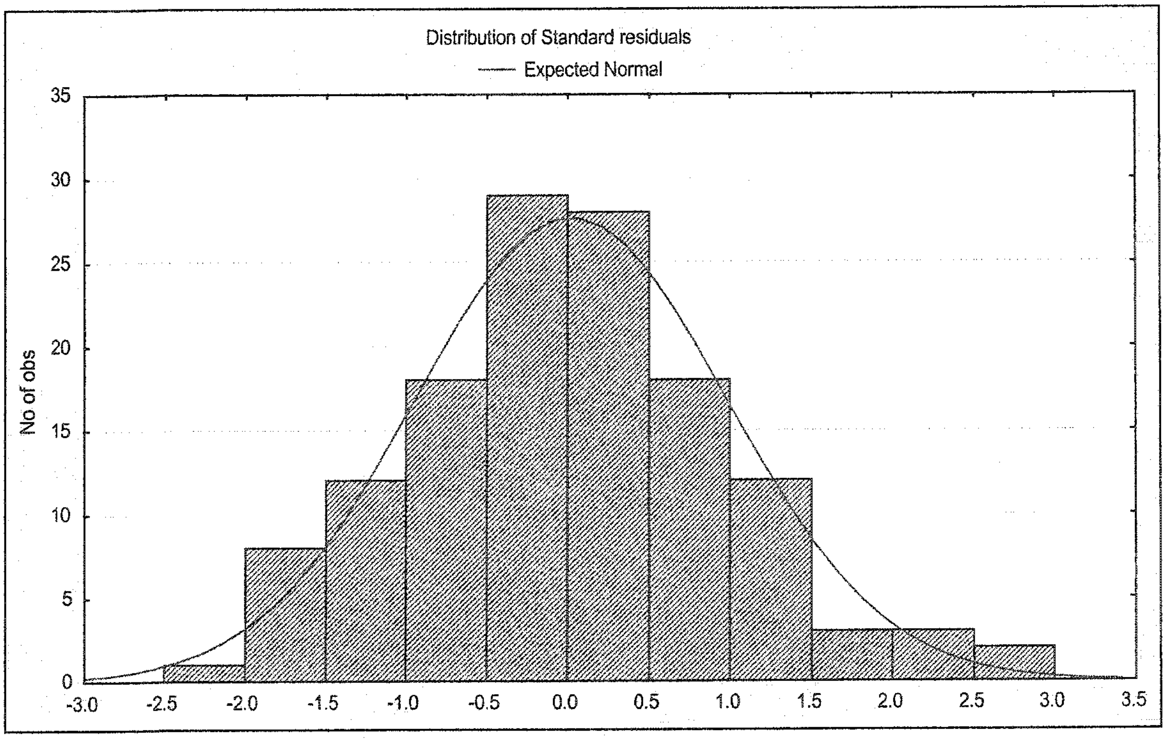


Figure 5.14, Standardized residuals of principal tensile strain cylinder RB2, HST model

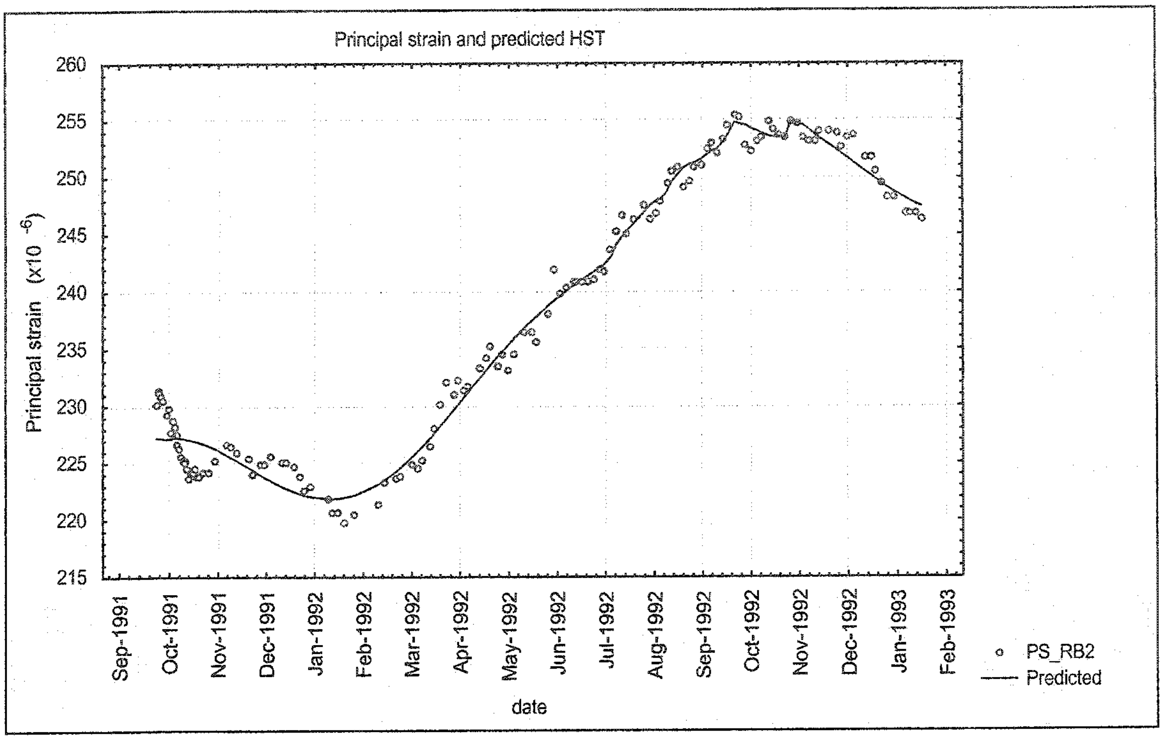


Figure 5.15, Principal strain cylinder RB2 and predicted HST model

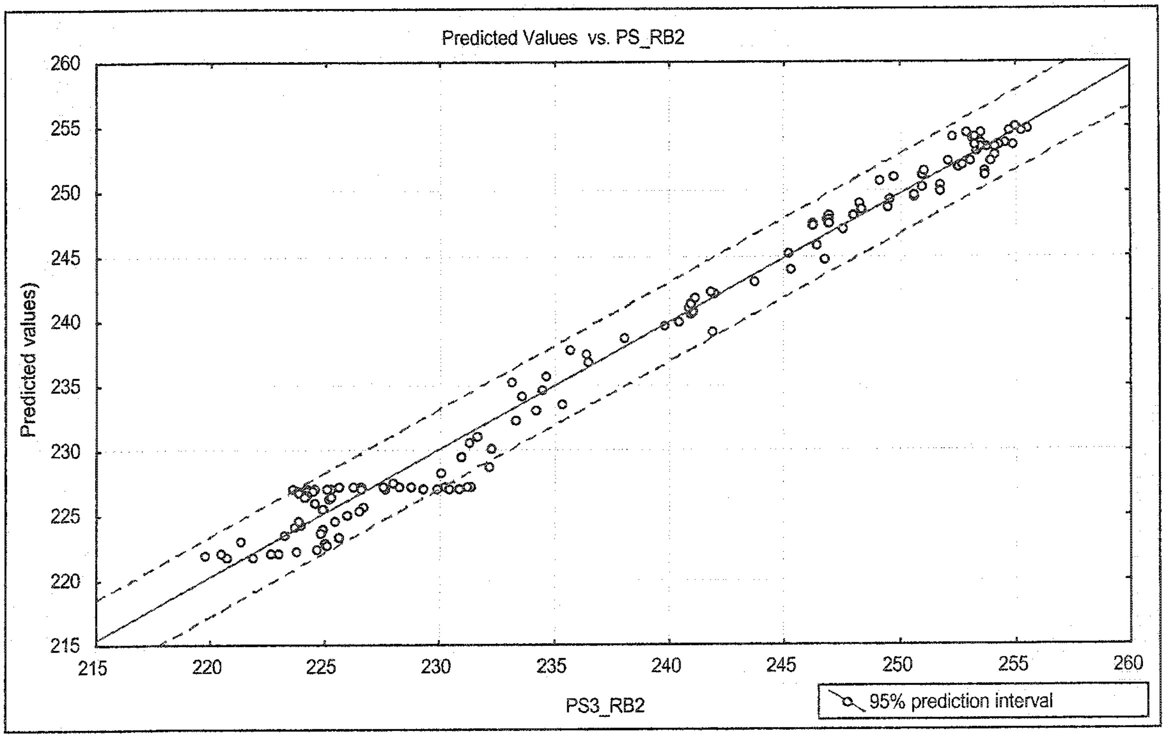


Figure 5.16, Principal strains, predicted values and prediction interval

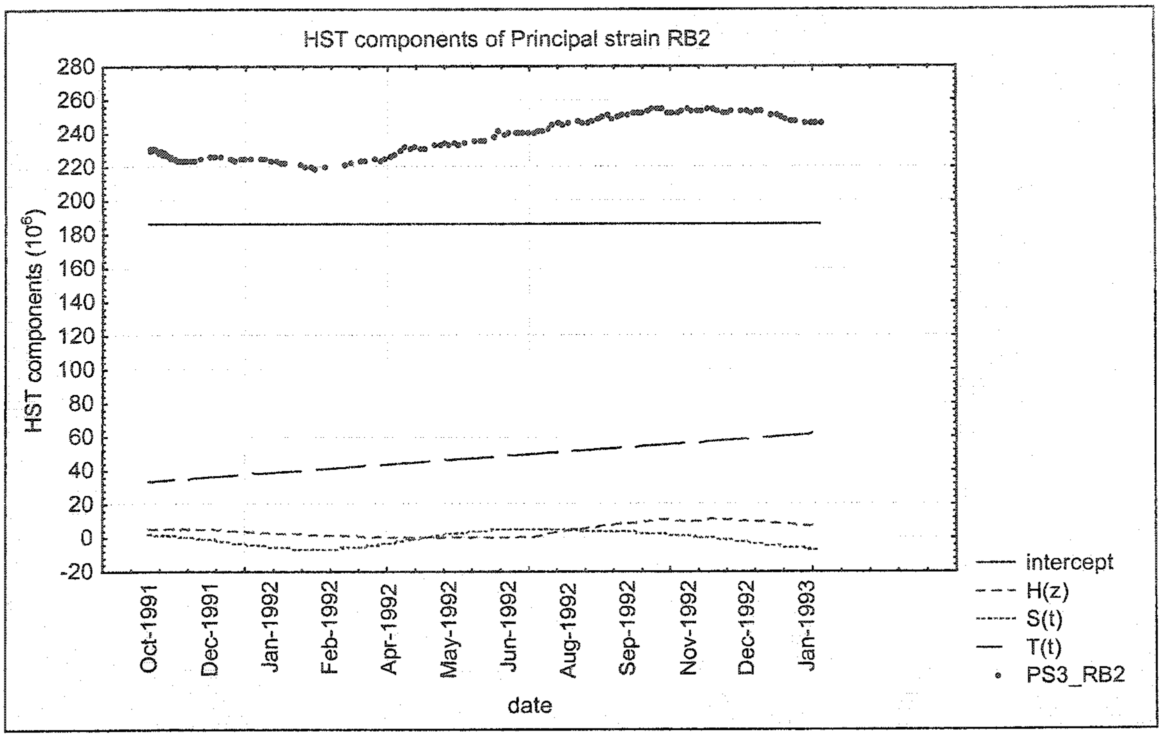


Figure 5.17, HST components of principal strain of cylinder RB2, Idukki dam

HST model indicates that irreversible effects account for more than 82% of the variation of principal strain of cylinder that is rather surprising.

Dimensional analysis indicates that strains induced by reservoir level variations can be expressed as $\varepsilon_H = b_1 z + b_2 z^2 + b_3 z^3$. In consequence, another model was developed which considered reservoir level variations and seasonal effects as predictors. Reservoir level and seasonal effects accounts for 43% and 46% of the variation of principal strain respectively (Table 5.4). Regression summary results are presented in Table 5.5. Predicted values and prediction intervals are shown in Figure 5.20. Comparing the two models, the HST model would explain a larger proportion of total variances, and is more accurate in predicting maximum principal strains.

It must be noted that the data set analyzed covers only a period of 17 months, from Oct. 1992 to Feb 1993. Ideally data sets should cover at least a period of 3 year, in order to obtain a good estimate of seasonal factors. Analysis of data over a longer time period can separate irreversible effects more conclusively.

Table 5.4, Stepwise regression summary of second model for principal strain of cylinder RB2

Variable	Summary of Stepwise Regression Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variabls included
z^3	1	0.66	0.44	0.44	102.56	0.000000	1
sin(θ)	2	0.87	0.75	0.31	163.88	0.000000	2
cos(θ)	3	0.96	0.92	0.17	284.87	0.000000	3
$\sin^2$(θ)	4	0.99	0.97	0.05	225.22	0.000000	4

Table 5.5, Regression summary of second model for principal strain of cylinder RB2

Regression Summary
R= .9856 R²= .9714 Std.Error of estimate: 2.0464

N=134	Beta	Std.Err. of Beta	B	Std.Err. of B	t(129)	p-level
Intercept			214.91	0.45	478.24	0.000000
z3	1.48	0.03	55.40	0.94	58.83	0.000000
Sin(t)	0.81	0.02	14.40	0.31	46.09	0.000000
Cos(t)	0.59	0.02	10.25	0.40	25.66	0.000000
Sin2(t)	0.23	0.02	7.64	0.51	15.01	0.000000

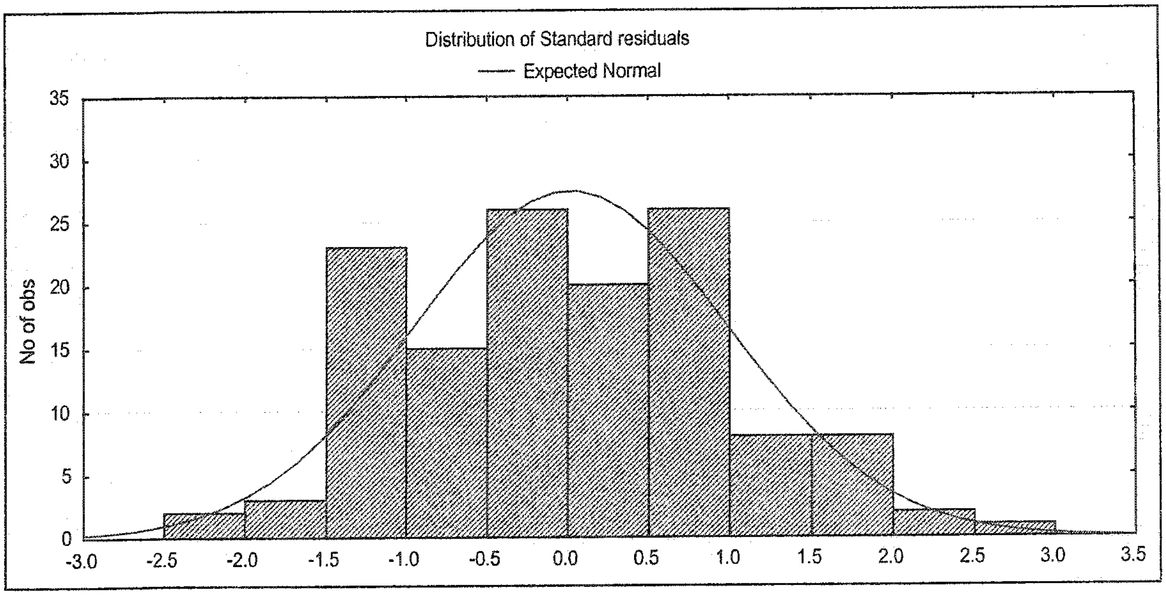


Figure 5.18, Residuals of principal tensile strain cylinder RB2, second model

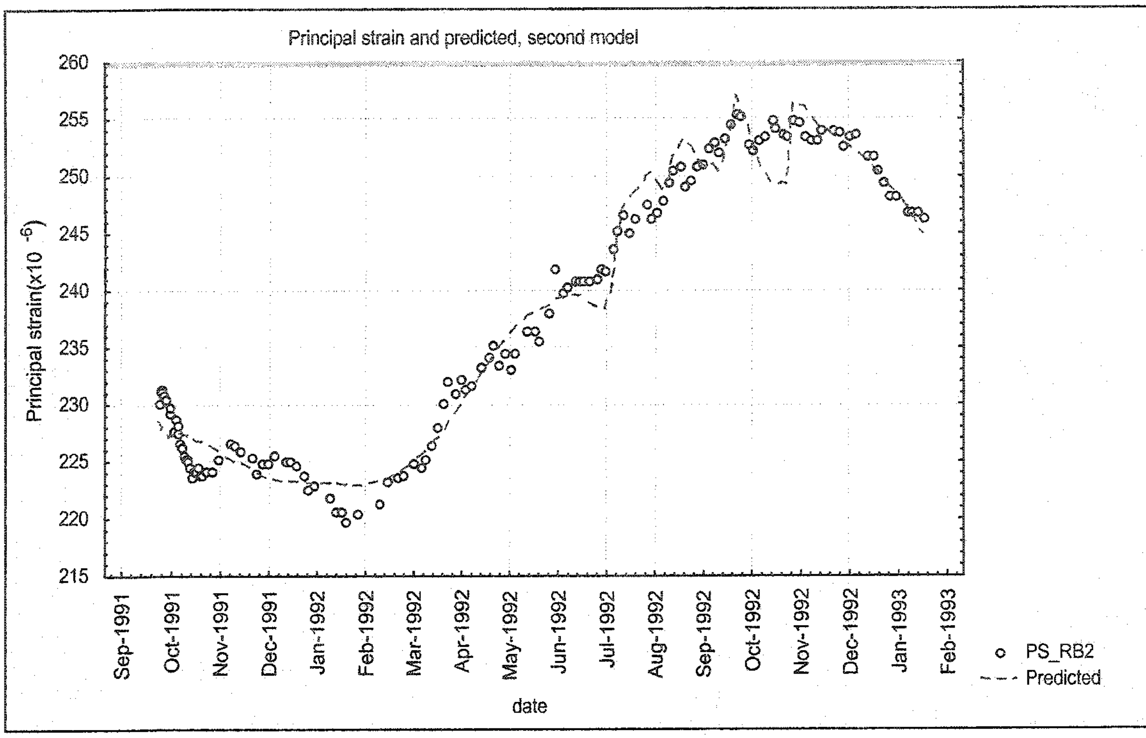


Figure 5.19, Principal strain cylinder RB2 and predicted second model

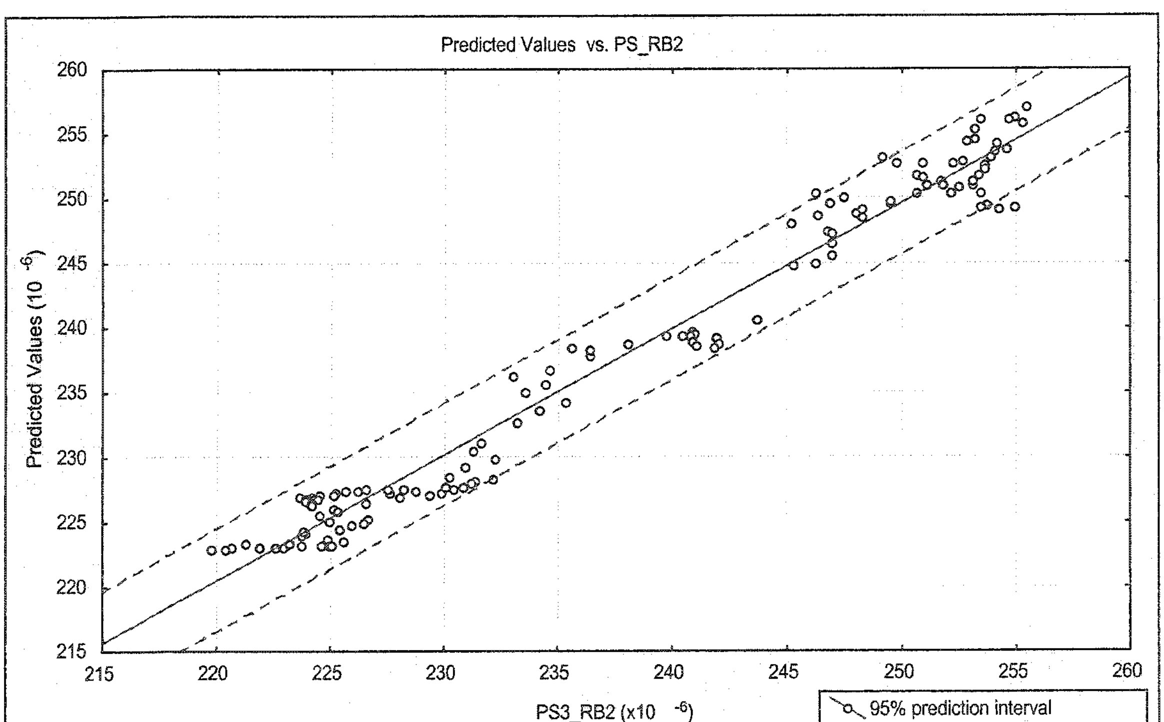


Figure 5.20, Principal strains, predicted values and prediction interval

5.2.2.2 Principal component analysis

The data set analyzed covers 17 months of observations with a frequency of two readings per week. The reliability of two of six cylinders is doubtful based on a preliminary exploratory analysis. Therefore only four cylinders are considered for the analysis. The cylinders are installed at the periphery of the dam next to the rock surface and can only model the localized behavior of the related dam section The data set consists of 24 variables, six for each cylinder.

PCA were performed on the correlation matrix. The number of principal components to retain may be determined by several methods. Figure 5.21 shows the scree plot, which is plot of the eigenvalues as a function of the number of eigenvalues. The number of components is selected by identifying the point at which the scree plot flattens out. Another option is to use only eigenvalues of greater than one. In this application the scree plot flattens out at the fifth eigenvalue that is also less than one. Consequently, only the first four principal components are retained.

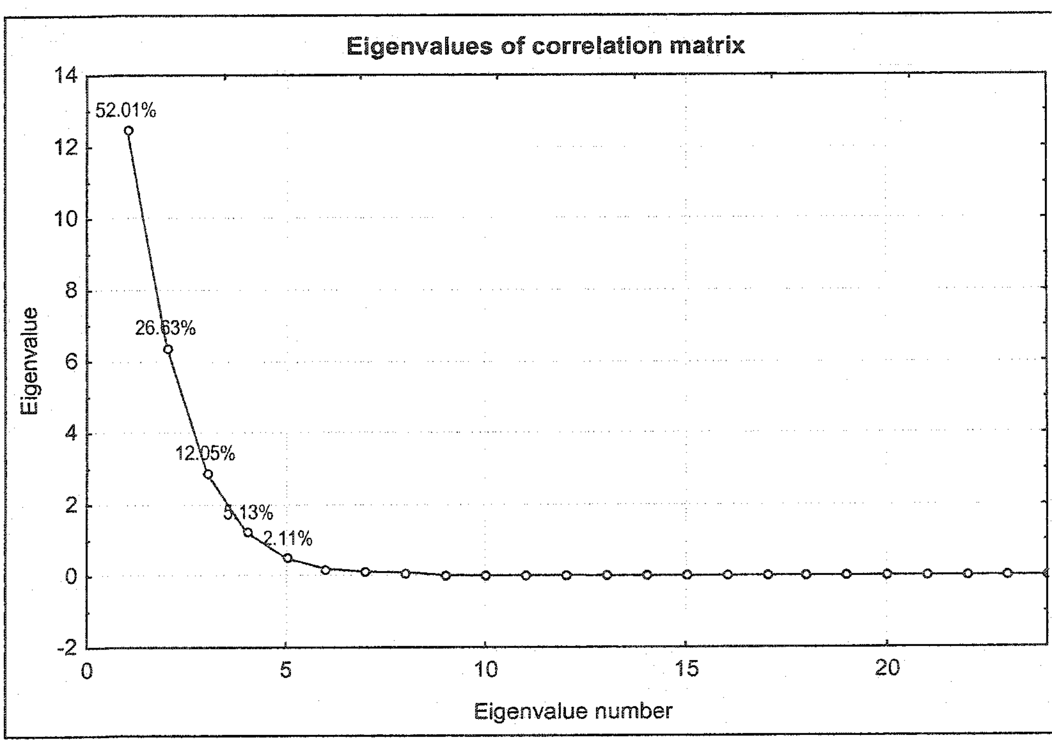


Figure 5.21, Scree plot, PCA correlation matrix of instrumented cylinders

The first four components explain 96% of the total variance (Table 5.6) of the 24 original variables, which is a significant reduction with respect to number of original variables. It can be concluded that PCA can efficiently compress the original data and reduce number of monitored variables.

Table 5.6, PCA results

Principal Component	Eigenvalues	Variance Explained by each PC	Commutative Variance Explained
1	12.48	52%	52%
2	6.39	27%	79%
3	2.89	12%	91%
4	1.23	5%	96%
5	0.5	2.1%	98%

The first principal component explains 52%; the second component 27%; the third 12%; and the fourth 5% of the total variance of the data set (Table 5.6). As can be noted, the remaining components explain only 4% of the variance and, most likely, they correlate with the noise in the data..

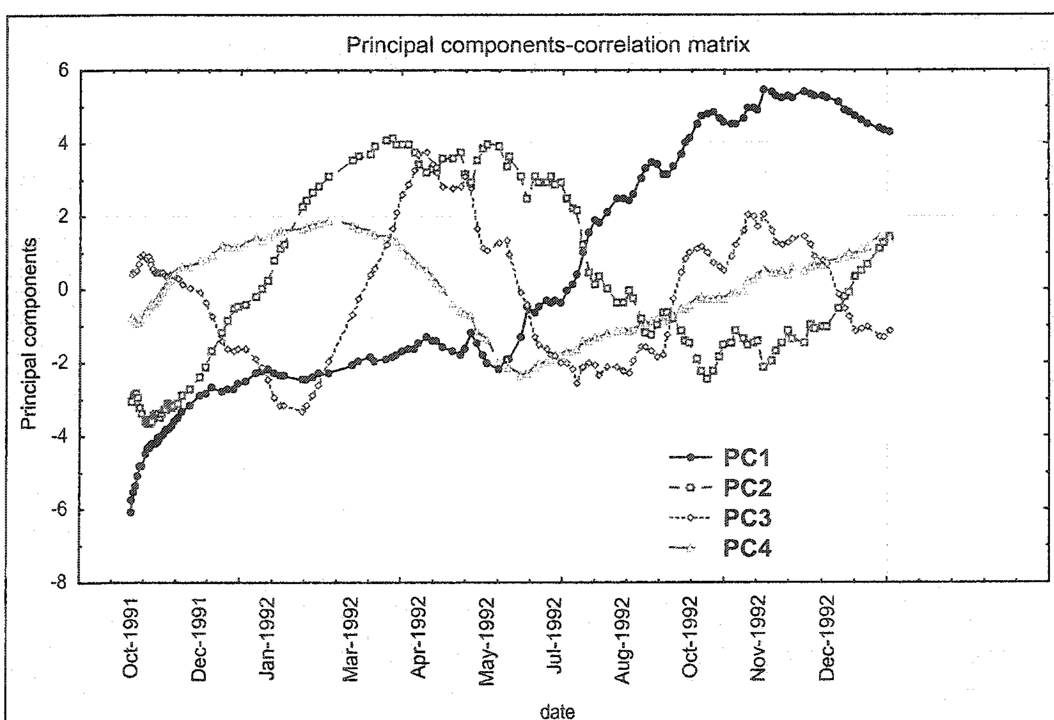


Figure 5.22, Principal components of cylinders, Idukki dam

Table 5.7, Stepwise regression, summary for PC1 of cylinders

Variable	Summary of Stepwise Regression; PC1 Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variabls included
t	1.00	0.96	0.93	0.93	1629.53	0.00	1
z	2.00	0.99	0.98	0.05	281.21	0.00	2

Table 5.8, Stepwise regression, summary for PC2 of cylinders

Variable	Summary of Stepwise Regression PC2 Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variabls included
z	1	0.81	0.65	0.65	241.60	0.0000	1
t	2	0.92	0.84	0.19	154.80	0.0000	2
T_RB1	3	0.97	0.95	0.11	285.54	0.0000	3
T_LB2	4	0.99	0.97	0.02	101.84	0.0000	4
z2	5	0.99	0.98	0.01	54.35	0.0000	5

Table 5.9, Stepwise regression summary for PC3 of cylinders

Variable	Summary of Stepwise Regression PC3 Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variabls included
T_LB2	1	0.70	0.49	0.49	126.27	0.0000	1
z3	2	0.97	0.94	0.45	1001.45	0.0000	2
T_LB1	3	0.98	0.95	0.01	30.90	0.0000	3

Table 5.10, Stepwise regression summary for PC4 of cylinders

Variable	Summary of Stepwise Regression PC4 Step +in/-out	Multiple R	Multiple R-square	R-square change	F - to entr/rem	p-level	Variabls included
T_RB1	1	0.62	0.38	0.38	81.04	0.0000	1
T_LB1	2	0.95	0.91	0.53	764.79	0.0000	2
t	3	0.97	0.93	0.02	48.94	0.0000	3
z3	4	0.98	0.95	0.02	54.96	0.0000	4
T_RB2	5	0.98	0.96	0.01	22.86	0.0000	5

Multiple linear regression models were developed for the first four PC's. Cylinder temperatures, time and polynomial of reservoir level were considered as predictor variables. Temperature of cylinders LB1, LB2, RB1, and RB2 are denoted by T_LB1, T_LB2, T_RB1 and T_RB2 respectively. Results of stepwise regression indicate that PC1 is highly correlated with time, and PC2 is correlated with water level. The first component explains most of the variance and the scores are monotonically decreasing as a function of time (Figure 5.22). The scores could be associated with creep, which shows similar behavior as a function of time. However it is not clear if creep is occurring at the cylinders or in the dam. The second component explains 27% of the variance. Plotting the scores as a function of time together with reservoir level indicated that the scores are correlated to reservoir level variation. Third component (PC3) is correlated with both water level and temperature reservoir level (Table 5.9). Plotting the scores as a function of time together with average cylinder temperature indicated that the scores are correlated to average cylinder temperature. The fourth component explains 6% of the total variance, and is highly correlated with cylinder temperatures (Table 5.10). The scores for the fourth component show a very strong correlation with the variation of the temperature measured by cylinders RB1 (Figure 5.25). The correlation is strongest for a time delay of 45 days due to thermal inertia of the dam. The remaining principal components represent only 4% of the data variation and they represent noise in the data as they correspond to variation type not present in all the measured instruments.

It must be noted that the data set analyzed covers only a period of 17 months, from Oct. 1992 to Feb 1993. Analysis of data over a longer time period can separate irreversible and reversible effects more conclusively.

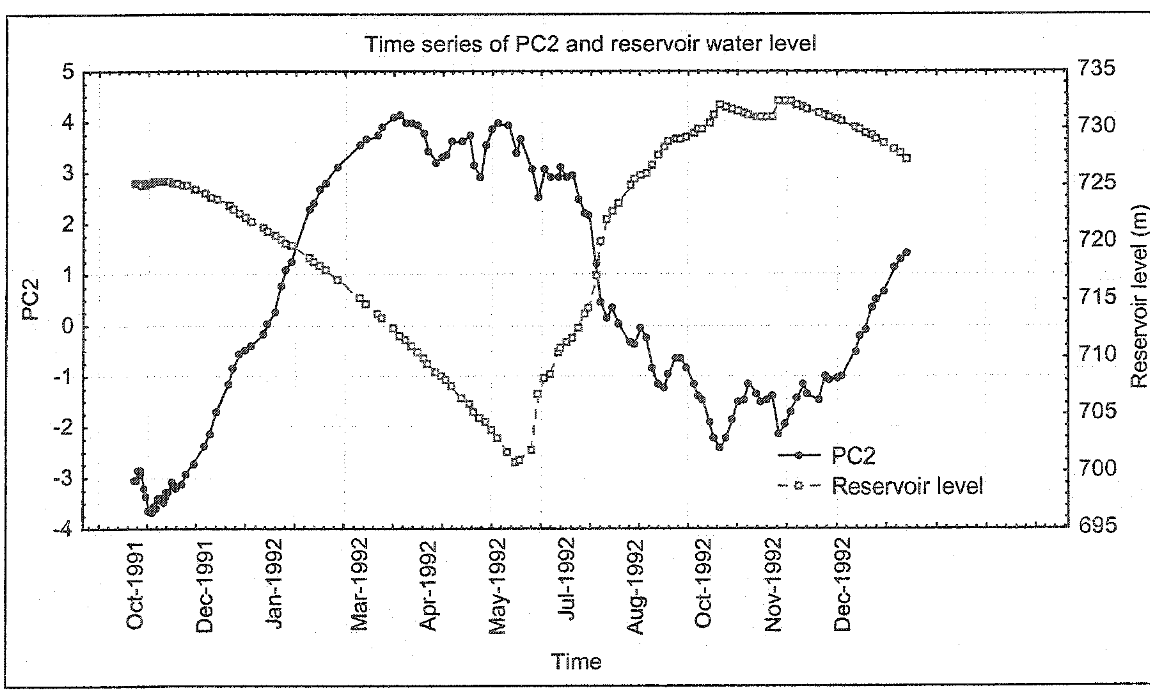


Figure 5.23, Scores of second PC and reservoir level variations, Idukki dam

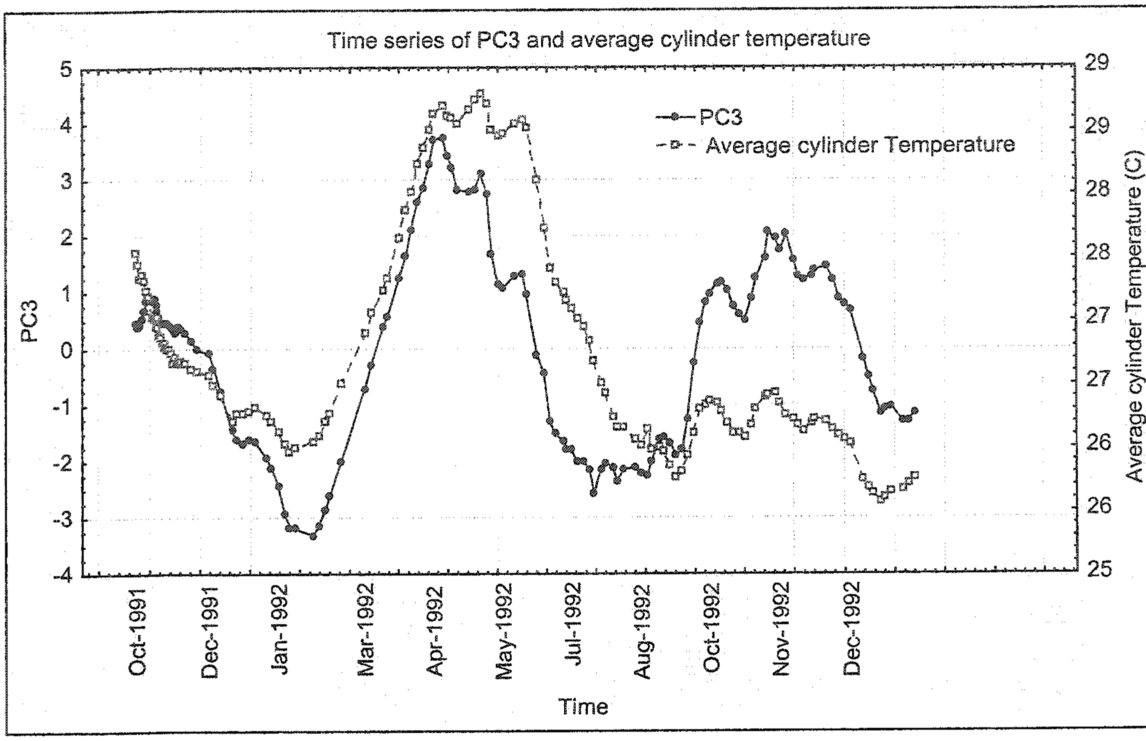


Figure 5.24, Scores of third PC and average cylinder temperatures, Idukki dam

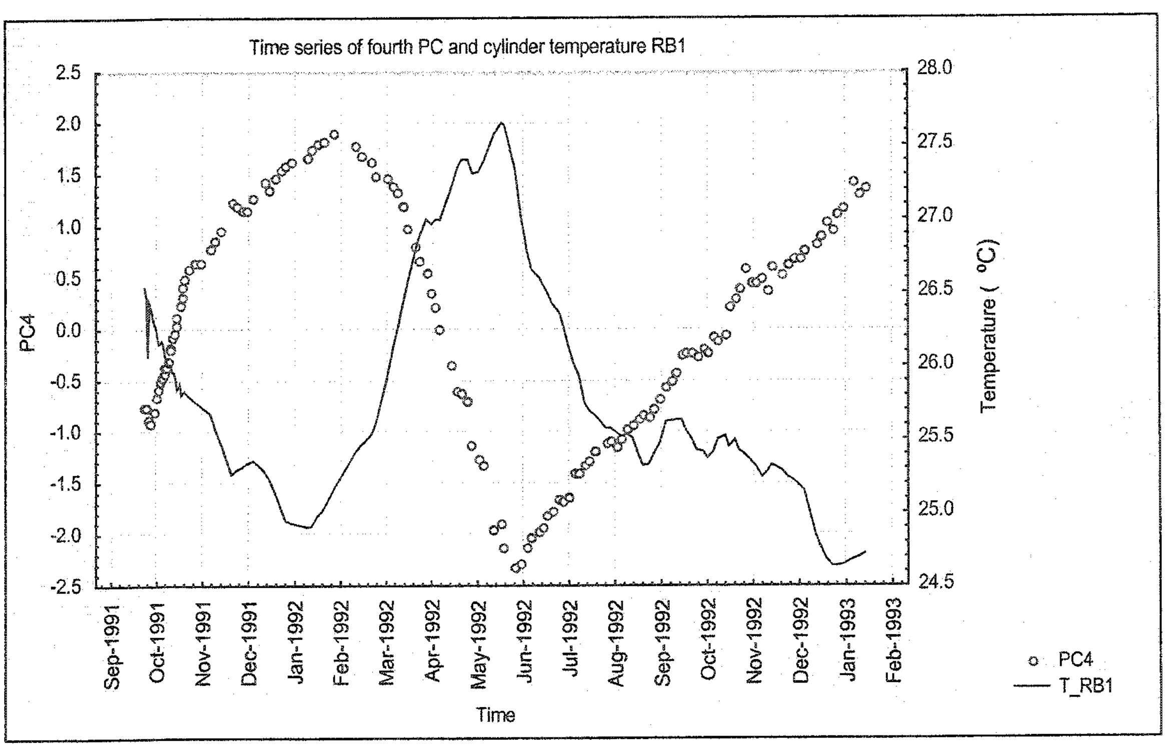


Figure 5.25, Scores of fourth PC and cylinder temperature RB1, Idukki dam

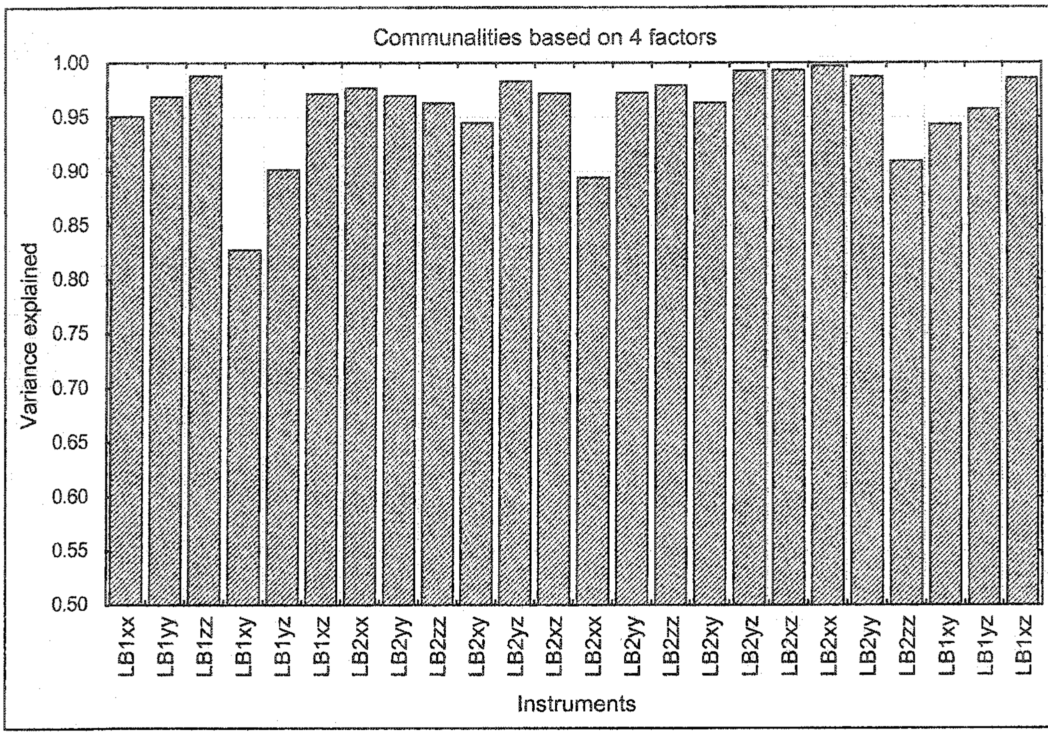


Figure 5.26, Explained variance of instruments by four principal components of instrumented cylinders, Idukki dam

Four principal components can be used to estimate the response variables (24 measured strains). The variance explained for each instrument is a measure of goodness of fit (Figure 5.26). For most of the instruments the variance explained is well over 90%. The four factors only explain about 83% of the variation of the ε_{xy} for cylinder LB1

5.2.3 Statistical analysis of displacements

Simple Pendulums P2, P3, P4 and inverted pendulum P5 are installed at elevations 2400 ft (732 m), 2300 ft (701 m), 2100 ft (640 m) and 1900 ft (579 m) of the central block (block 1) respectively (Figure 5.6). Pendulums P1 and P6 are installed in blocks 7 and 8 to measure movements at elevations 2100 ft (640 m) and 2300 ft (701 m) respectively. These pendulums are considered for statistical analysis.

Pendulum readings are available from 1977 to 1989. Measurements had been recorded more frequently before 1981, Air temperature readings were only available from 1982 and therefore were not considered in the analysis. The name of the variables is defined to indicate the block number, elevation, and type of displacement, for example 1R2400 and 1T2400 are respectively of radial and tangential displacements at elevation 2400 ft (732 m) of block 1.

Table 5.11, Basic statistics for pendulums, Idukki dam

	Descriptive Statistics				
Variable	Valid N	Mean	Minimum	Maximum	Std.Dev.
1R1900	220	-0.20	-2.82	2.17	0.36
1R2100	220	-2.28	-6.05	8.94	1.94
1R2300	220	-4.22	-13.11	17.66	5.11
1R2400	220	-4.26	-15.01	16.92	5.77
1T1900	220	-0.86	-35.59	1.13	2.50
1T2100	220	-0.93	-35.30	5.97	2.81
1T2300	212	-1.41	-36.24	18.32	3.24
1T2400	212	-1.42	-36.07	18.89	3.24
7R2300	192	-0.74	-5.32	6.31	2.58
7R2400	197	-1.26	-7.95	7.34	3.24
7T2100	192	-1.68	-10.91	11.54	4.72
7T2300	197	-2.18	-13.42	12.86	5.42
8R2100	193	-1.24	-5.87	7.41	2.50
8R2300	193	-0.58	-11.60	7.17	3.15
8T2100	192	0.47	-3.93	5.07	1.87
8T2300	193	0.21	-3.20	7.80	1.65

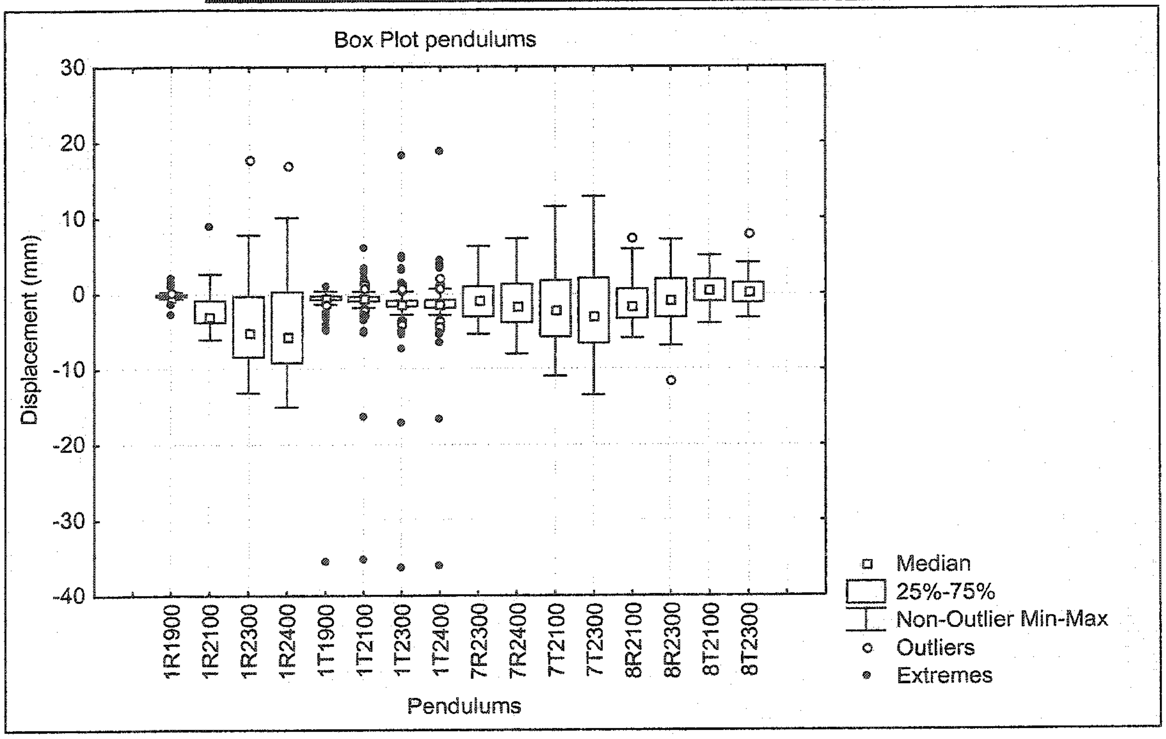


Figure 5.27, Box plot for pendulum readings, Idukki dam

Complete simultaneous readings for all pendulums are available for 182 data points. There are a few extreme points for all instruments, especially for tangential displacements in Block 1 (Figure 5.27). Maximum displacements are observed at elevation 2400 ft (732 m) of Block 1. Figure 5.28 shows the correlation among pendulum readings, which are grouped according to their correlation coefficients.

The following observations can be made regarding pendulum displacements:

1) all radial displacements except 1R1900, which is located at the base, are highly correlated, 2) tangential movements of block 1 are highly correlated, 3) radial displacements and reservoir level variations are highly correlated, 4) there is no correlation between tangential displacements of central block 1 and reservoir level, and 5) radial and tangential displacements of block 7 and 8 are strongly correlated with reservoir level.

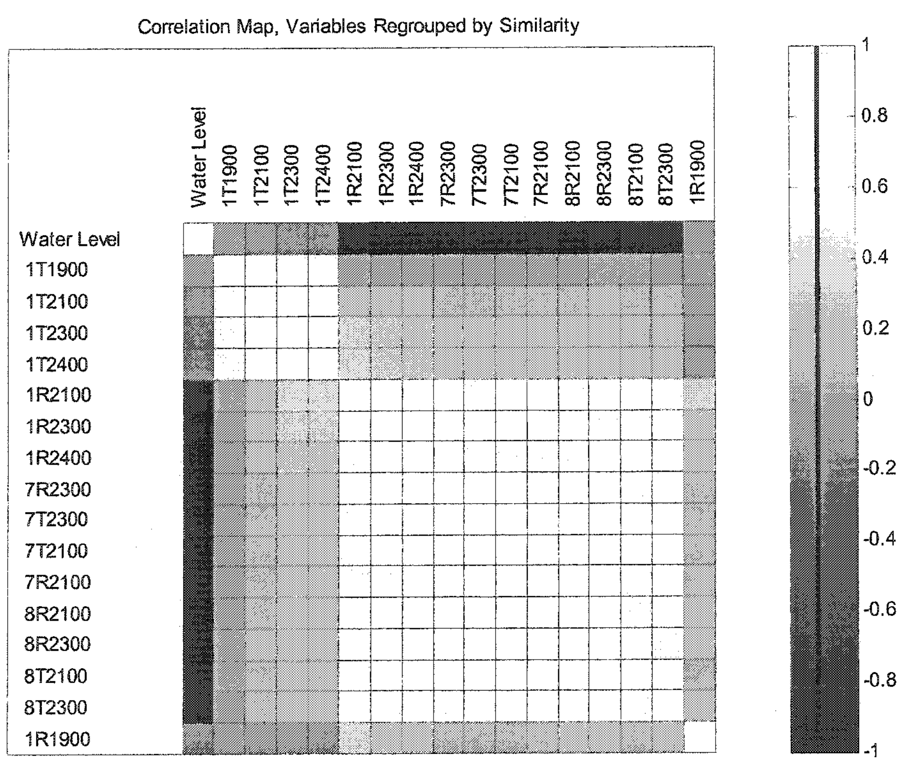


Figure 5.28, Correlation map for pendulum readings, Idukki dam

Table 5.12, Correlation between pendulum displacements blocks 1, Idukki dam

Variable	Correlations between pendulum displacement								
	WL	1R1900	1R2100	1R2300	1R2400	1T1900	1T2100	1T2300	1T2400
WL	1.00	-0.10	**-0.83**	-0.91	-0.92	0.01	-0.08	-0.15	-0.16
1R1900	-0.10	1.00	0.31	0.18	0.19	-0.02	-0.05	-0.08	-0.10
1R2100	**-0.83**	0.31	1.00	**0.94**	**0.93**	0.00	0.09	0.32	0.32
1R2300	**-0.91**	0.18	**0.94**	1.00	**0.98**	-0.02	0.08	0.27	0.28
1R2400	**-0.92**	0.19	**0.93**	**0.98**	1.00	-0.02	0.09	0.26	0.27
1T1900	0.01	-0.02	0.00	-0.02	-0.02	1.00	**0.85**	0.73	0.72
1T2100	-0.08	-0.05	0.09	0.08	0.09	0.85	1.00	0.86	0.84
1T2300	-0.15	-0.08	0.32	0.27	0.26	0.73	**0.86**	1.00	**0.99**
1T2400	-0.16	-0.10	0.32	0.28	0.27	0.72	**0.84**	**0.99**	1.00

Table 5.13, Correlation between pendulum displacements block 1,7 and 8, Idukki dam

Variable	Correlations							
	7R2100	7R2300	7T2100	7T2300	8R2100	8R2300	8T2100	8T2300
WL	-0.86	-0.90	-0.89	-0.90	-0.91	-0.90	-0.82	-0.81
1R1900	0.07	0.06	0.07	0.06	0.16	0.10	0.06	0.10
1R2100	0.73	0.76	0.76	0.77	0.82	0.81	0.69	0.70
1R2300	0.89	0.88	0.89	0.90	0.90	0.88	0.80	0.80
1R2400	0.89	0.92	0.92	0.93	0.92	0.90	0.82	0.84
1T1900	-0.03	-0.02	-0.03	-0.01	-0.02	0.02	0.01	-0.03
1T2100	0.06	0.08	0.07	0.09	0.09	0.10	0.08	0.07
1T2300	0.16	0.18	0.17	0.18	0.19	0.17	0.10	0.13
1T2400	0.17	0.19	0.18	0.19	0.19	0.18	0.11	0.13
7R2100	1.00	0.92	0.96	0.93	0.88	0.82	0.78	0.82
7R2300	0.92	1.00	0.98	1.00	0.88	0.87	0.82	0.83
7T2100	0.96	0.98	1.00	0.99	0.89	0.86	0.81	0.83
7T2300	0.93	1.00	0.99	1.00	0.89	0.88	0.83	0.85
8R2100	0.88	0.88	0.89	0.89	1.00	0.85	0.76	0.84
8R2300	0.82	0.87	0.86	0.88	0.85	1.00	0.86	0.76
8T2100	0.78	0.82	0.81	0.83	0.76	0.86	1.00	0.91
8T2300	0.82	0.83	0.83	0.85	0.84	0.76	0.91	1.00

5.2.3.1 Anomaly detection

One of the main applications of the instrumentation data analysis is the identification of anomalies in dam monitoring systems Two kinds of checks are generally applied on instrumentation data: 1) A comparison of the observation and its rate of change with preset threshold values. The thresholds are obtained from historical data or numerical models; and. 2) A comparison of the observation with predicted values obtained from a reference model.

In current dam safety practice these alarm levels are chosen based on the analysis of each individual instrument. However, as some of the measurements are noisy or unreliable, this approach increases the chance of finding an instrument out of control. Thus the false alarm rate (or probability of Type 1 error) is increased if each variable is analyzed and checked separately since the number of false alarms is directly proportional to the number of instruments. In many cases, the dam will actually be in a safe state but each alarm could require verification. A combination of univariate, bivariate and multivariate statistical methods is proposed to overcome some of these difficulties and reduce the probabilities of false alarms generated by a dam monitoring system.

A model is developed for the reference period and is used for the prediction of instrument readings in real time. Abnormal data, which is not consistent with past readings must be investigated and labeled as erroneous data or possible anomalies.

The data used in generating statistical reference model:

1) must be for the intact structure, where only minor structural modifications have occurred, 2) must be under operating conditions and cover all range of normal operational loadings (i.e reservoir level, temperatures), 3) must have sufficient length and appropriate frequency of readings to obtain a statistically significant model

In developing the statistical reference model it is important to treat outliers properly. The reference model is then used for the prediction of future observations and to establish the confidence limits for range of expected behavior. Any reading, which is not consistent with the reference model, generates a warning message or alarm. When the reading is not within pre-established limits, prompt evaluation of the safety of the dam is normally taken which may lead to:

1) Assessing, and if needed, resetting the boundaries of satisfactory performance of the dam, as measured by instruments, 2) intensified monitoring, 3) lowering the reservoir

level, 4) warning the population, and evacuating downstream areas and, 5) taking structural corrective actions.

Erroneous data can be a source of error when developing reference models, and must be removed before building regression or PCA models (Wise, 1991). This is important because outliers have a great deal of leverage on the regression models and can change them significantly. An outlier is an observation (or subset of observations) that appears to be inconsistent with the remainder of the data set (Barnett, 1994). Any extreme data point can be due to: 1) erroneous data, 2) faulty sensors or, 3) changes in the dam behavior. Outlier identification (and subsequent removal or accommodation) is a part of the data screening process, which should be done routinely before statistical analyses.

Outliers can be classified into one of four categories. The first category contains those outliers arising form a procedural error, such as data entry error or a mistake in coding. These types of outliers can be easily identified and should be eliminated. The second class of outliers are observations that occur as a result of an extraordinary event. In this case an explanation exists for the uniqueness of the observation (Hair et al, 1995). Higher than normal reservoir level and daily temperatures are such examples. A decision must be made whether or not the outlier represents a valid observation in the population. If so, it should be retained; if not, it should be deleted from the analysis. The third class of outlier comprises extraordinary observations for which there is no explanation. Although these are the outliers most likely to be omitted, they may be retained if the analyst feels they represent a valid segment of the population. Other information such as results of inspections can be used to facilitate a decision. The fourth and final class of outliers are,observations that fall within the ordinary range of values on each of the variables, but are unique in their combination of values across the variables (multivariate outliers). In this case, the observations must be retained unless specific evidence is available discounting the outlier as a valid member of the population. Methods used in detection of outliers are a significant part of a statistical study and can be divided into, univariate, bivariate and multivariate. Several of these methods should be utilized, looking for consistent results in identifying outliers.

The following steps are taken in the identification of anomalies in the reference model and for the subsequent comparison of future values to this model. Since results of visual dam inspection and engineering reports are available at the time of developing the reference model, any data that does not conform to the model is generally discarded. However, anomalies detected in the real-time monitoring must be carefully dealt with.

Univariate detection

Outliers are identified by examining the distribution of observations, deleting as outliers those cases that fall in the outer ranges of the distribution. The primary issue concerns the establishment of the threshold for designation of an outlier; the observations (i.e. displacements, pressures) or their rate of change are converted to standard scores, which have a mean of zero and a standard deviation of one. Once the values are expressed is a standardized format, comparisons across variables can be easily made. For small samples, (80 or fewer observations) the guidelines suggest identifying those cases with standard scores of 2.5 or greater as outliers (Myers et al, 1995). When the sample size is larger, the guidelines suggest that the threshold value of standard scores range from 3 to 4. In either case, It should be recognized that a certain number of observations may occur normally in these outer ranges of the distribution. Univariate control charts for tangential displacements of block 1 at elevations 1900 ft (579 m) and 2400 ft (732 m) are shown in Figure 5.29 and Figure 5.30 respectively. Figure 5.29 shows anomalies in tangential displacements of block 1 at the base of dam. The tangential displacements of pendulum P5 at the base are less than 1-2 mm during the period of 1977-1986, while, these displacements increase during the period from 1986 to 1989 and the maximum displacement in this period is 4.91 mm for a reservoir level of 698 m.

Bivariate detection

Pairs of variables can be examined jointly through a scatter plot. Cases that fall markedly outside the range of the other observations can be noted as isolated points in the scatter plot. To assist in determining the expected range of observations, an ellipse representing a specified confidence interval (normally 95% of the distribution) for a bivariate normal

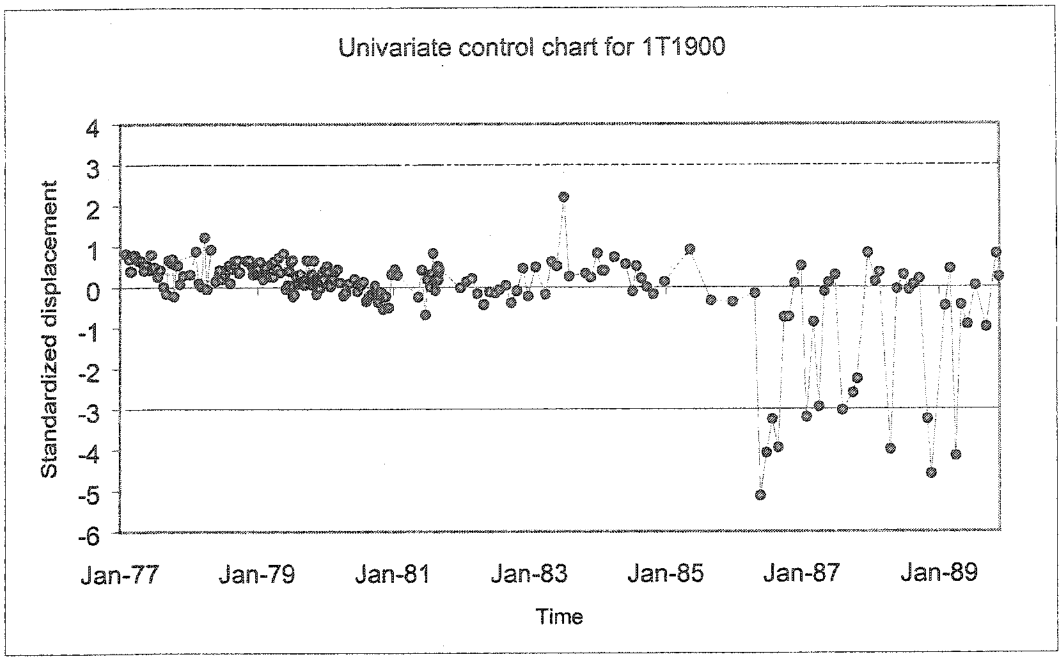


Figure 5.29, Univerariate control chart for tangential displacements of block 1, elevation 1900 ft (571 m)

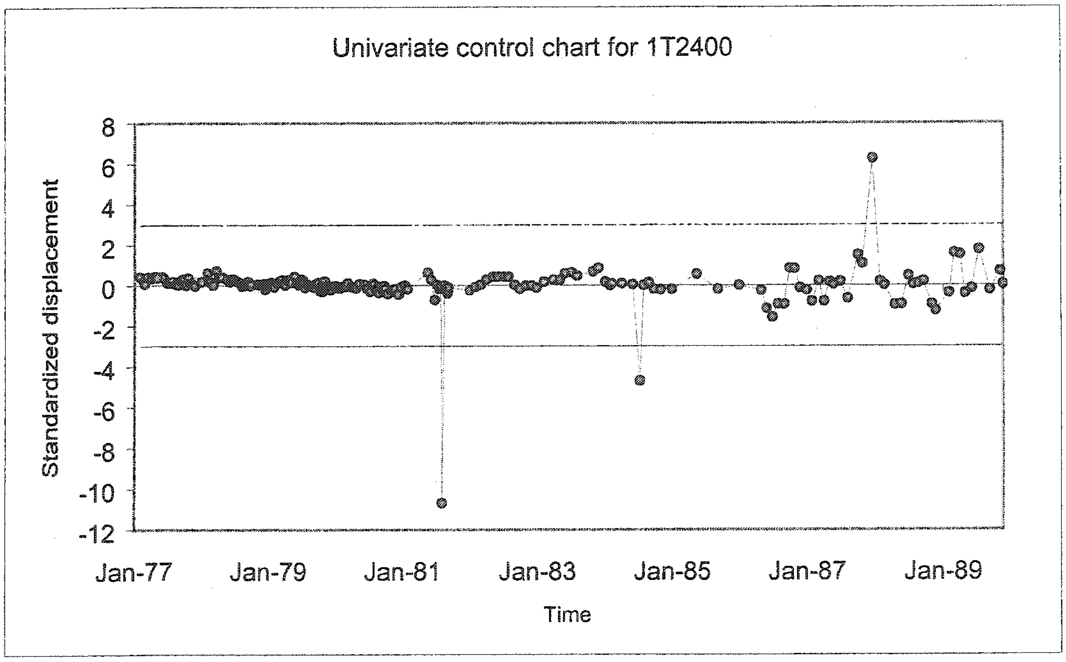


Figure 5.30, Univerariate control chart for tangential displacements of block 1, elevation 2400 ft (732 m)

distribution can be superimposed over the scatter plot. This provides a graphical portrayal of the confidence limits and facilitates identification of the outliers.

Figure 5.31 shows that that radial displacements are highly correlated. Observed displacements of January 1988 are 17.66 mm and 16.92 mm for elevations 2300 ft (701 m) and 2400 ft (731 m) respectively, and is regarded as an outlier. Similarly readings of September 1980 are regarded as an outlier. It must be noted that this point can not be detected from univariate method. Radial displacements are highly correlated with reservoir level variations. Alternatively scatter plots of these displacements and reservoir level can be considered. Similarly, tangential readings of block 8 at elevations 2100 ft (640 m) and 2300 ft (701 m) are highly correlated and outliers can be identified outside of confidence ellipse (Figure 5.32).

Multivariate detection,

Multivariate assessment of each observation across a set of variables can also be used. Dam responses are often correlated, Consequently, the dimensionality can be substantially reduced to a few principal components, much less than the original number of variables. In many cases, only three of four principal components can describe as much as 90% of the variance. Using standard distribution theory, confidence ellipses can be superimposed on the joint principal component scores charts.

PCA was used in a preliminary analysis to detect outliers. Principal components are extracted and their scores calculated. Next the joint principal component scores charts are plotted and 95% Confidence ellipses are then used to identify outliers (Figure **5.33** to Figure 5.35) and several anomalies are identified..

Outlier designation

When observations that are candidates for designation as outliers are identified by the several univariate, bivariate, or multivariate methods observations, which were similar, are deleted prior to fitting the reference model.

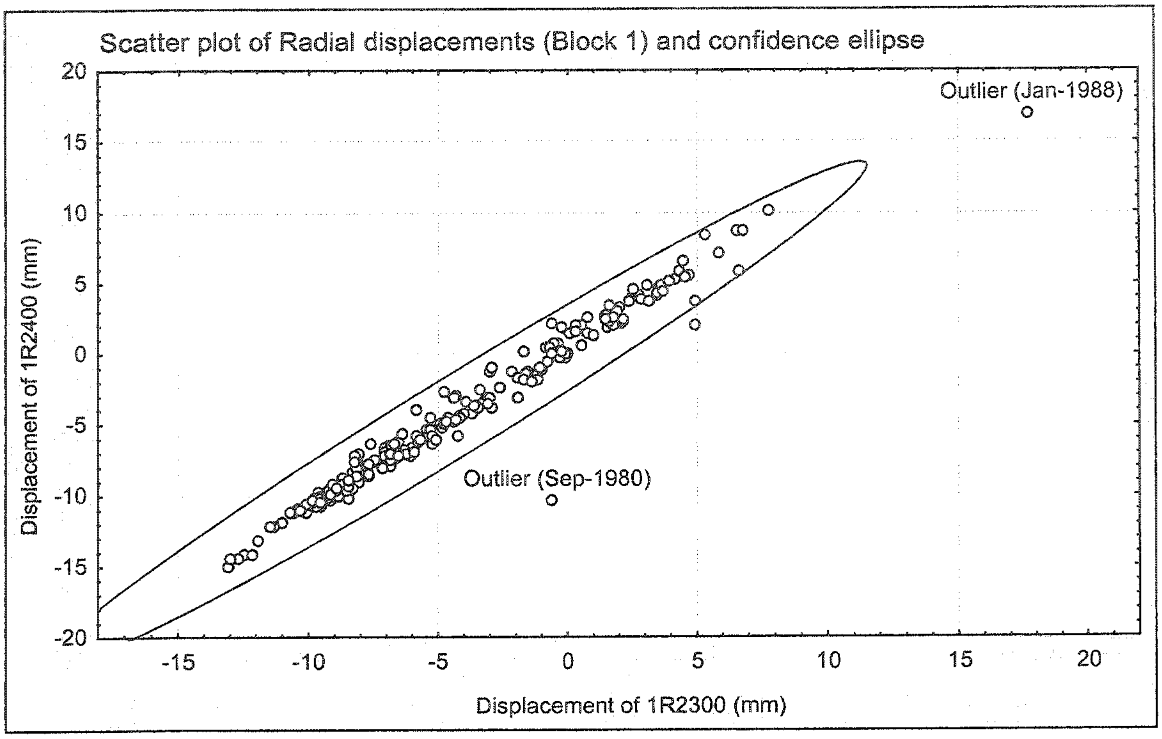


Figure 5.31, Bivariate control chart for radial displacements of Block 1, Idukki dam

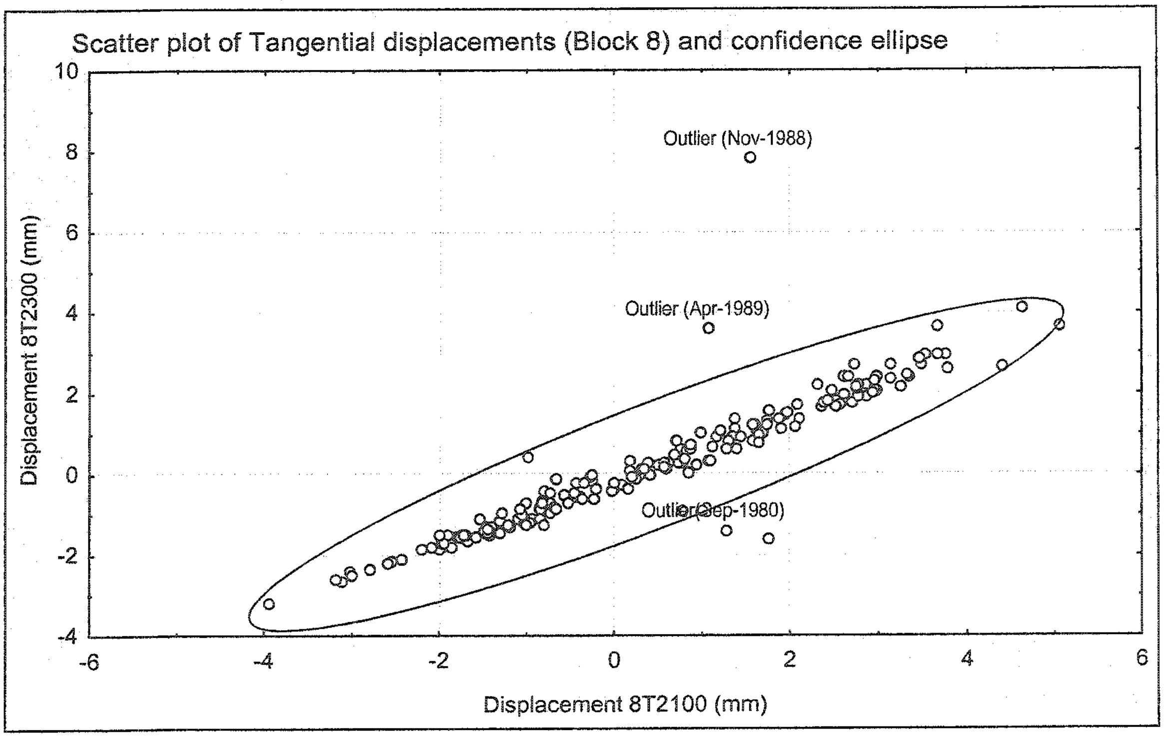


Figure 5.32, Bivariate control chart for tangential displacements of Block 8, Idukki dam

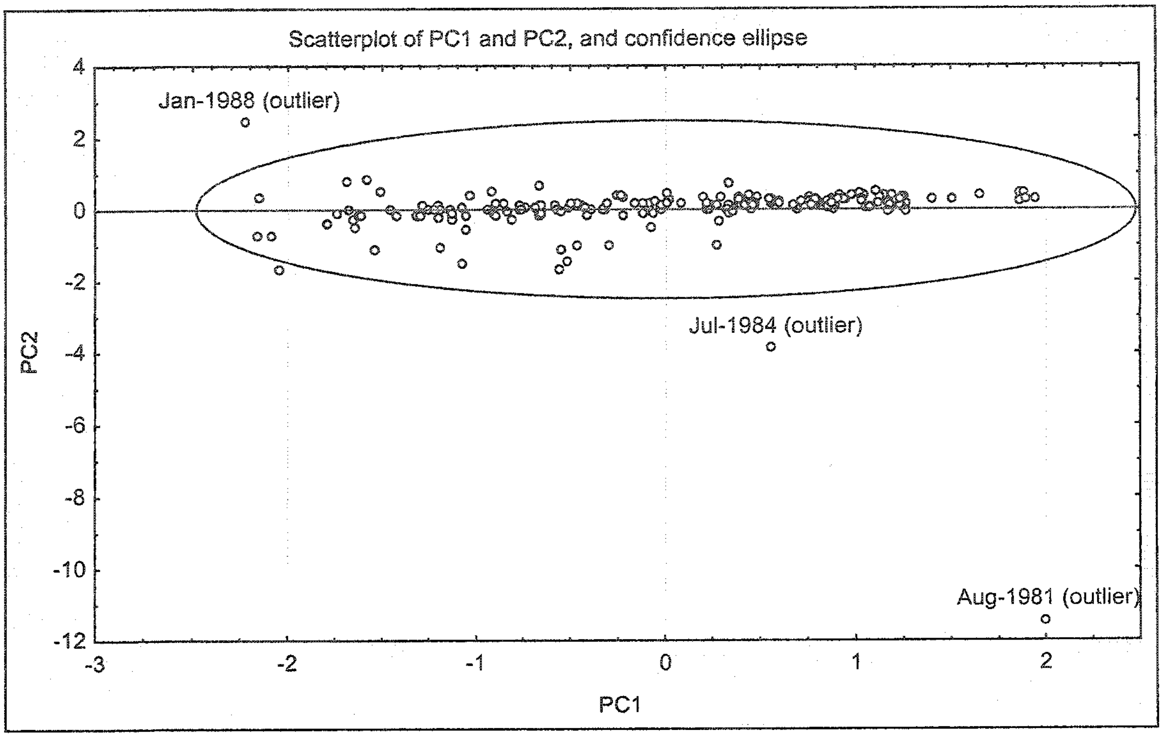


Figure 5.33, Monitoring chart based on first and second PC

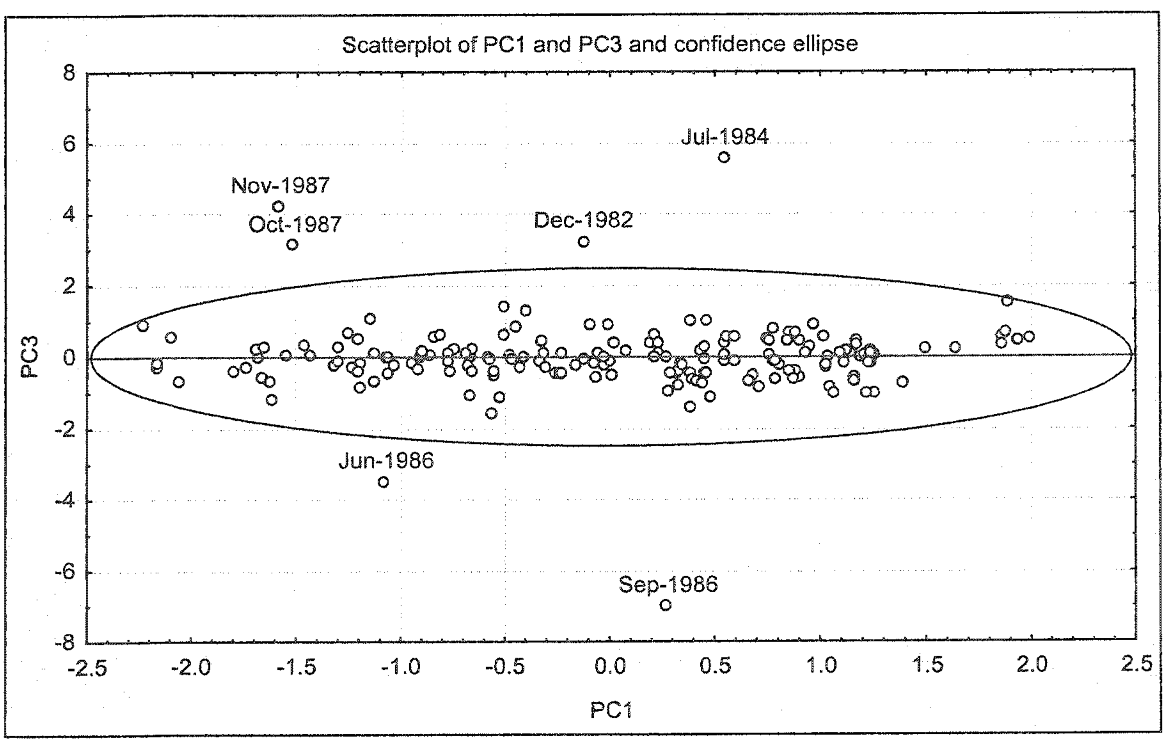


Figure 5.34, Monitoring chart based on first and third PC

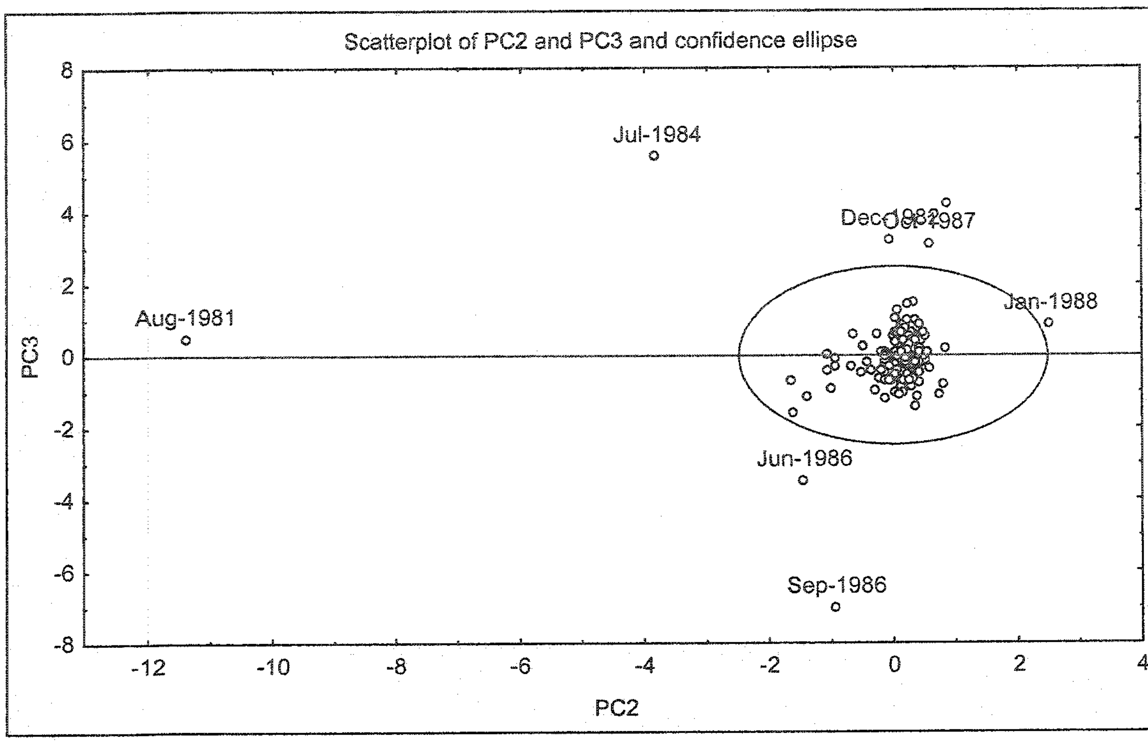


Figure 5.35, Monitoring chart based on second and third PC

5.2.3.2 Principal component analysis

PCA was performed on the correlation matrix of pendulums. The first three components explain 88% of the total variance (Figure 5.36) of the 16 original variables. The first principal component explains 63%; the second component 19%; the third 6%; of the total variance of the data set (Figure 5.36). The first component explains most of the variance and the scores are highly correlated with reservoir level variations (Figure 5 40).

Analyses of loadings for the first and second principal component indicates that all radial displacements of block 1 except at elevation 1900 ft (579 m) and tangential and radial displacements of blocks 7 and 8 are associated with PC1.

PC2 accounts for tangential displacements of block 1. PC3 is only correlated with radial displacements of block 1 at the base (1R1900). Displacements of 1R1900 (located at the base) are not correlated with rest of the pendulums. The principal components are uncorrelated and orthogonal; therefore when one of the variables is not correlated with the other variables, then variance of that variable is one of the eigenvalues and the

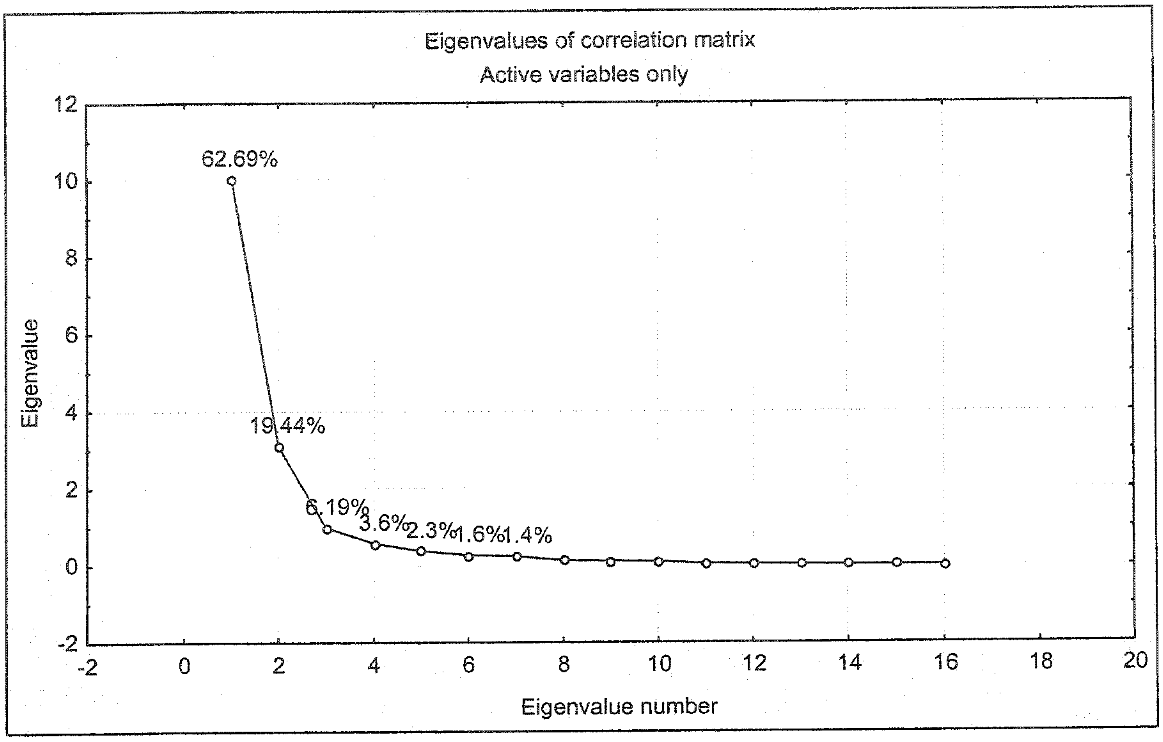


Figure 5.36, Scree plot, pendulum readings, Idukki dam

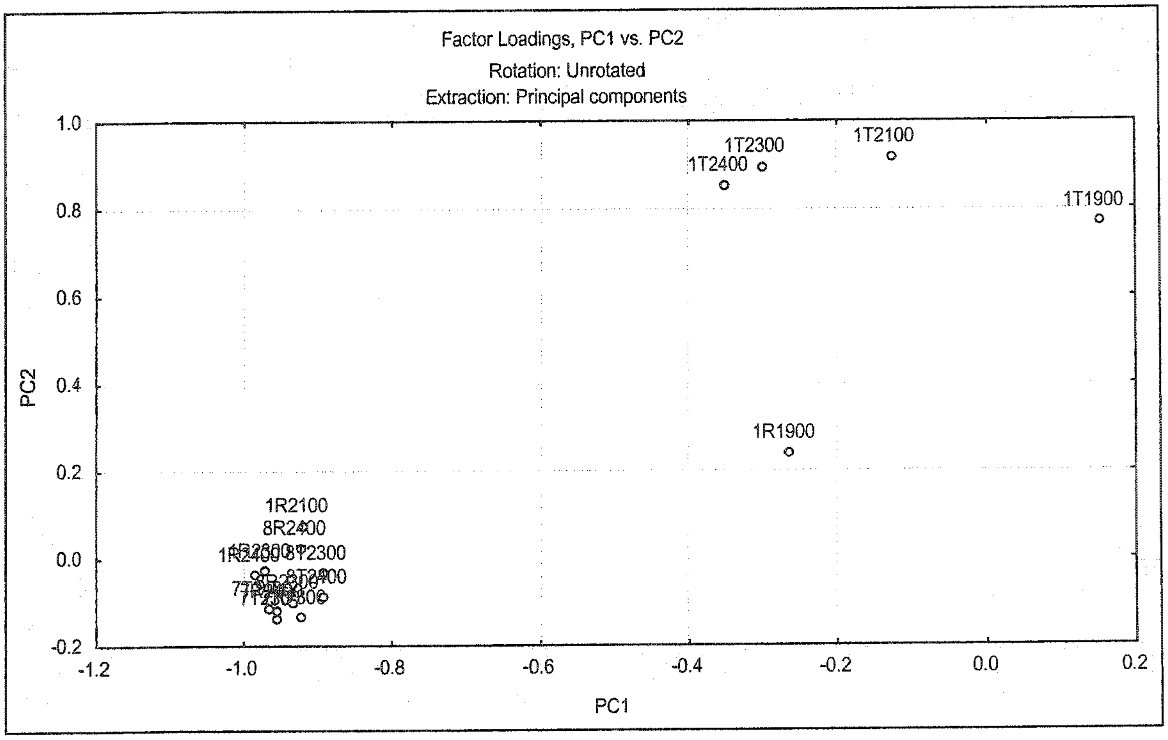


Figure 5.37, Loadings for PC1 and PC2

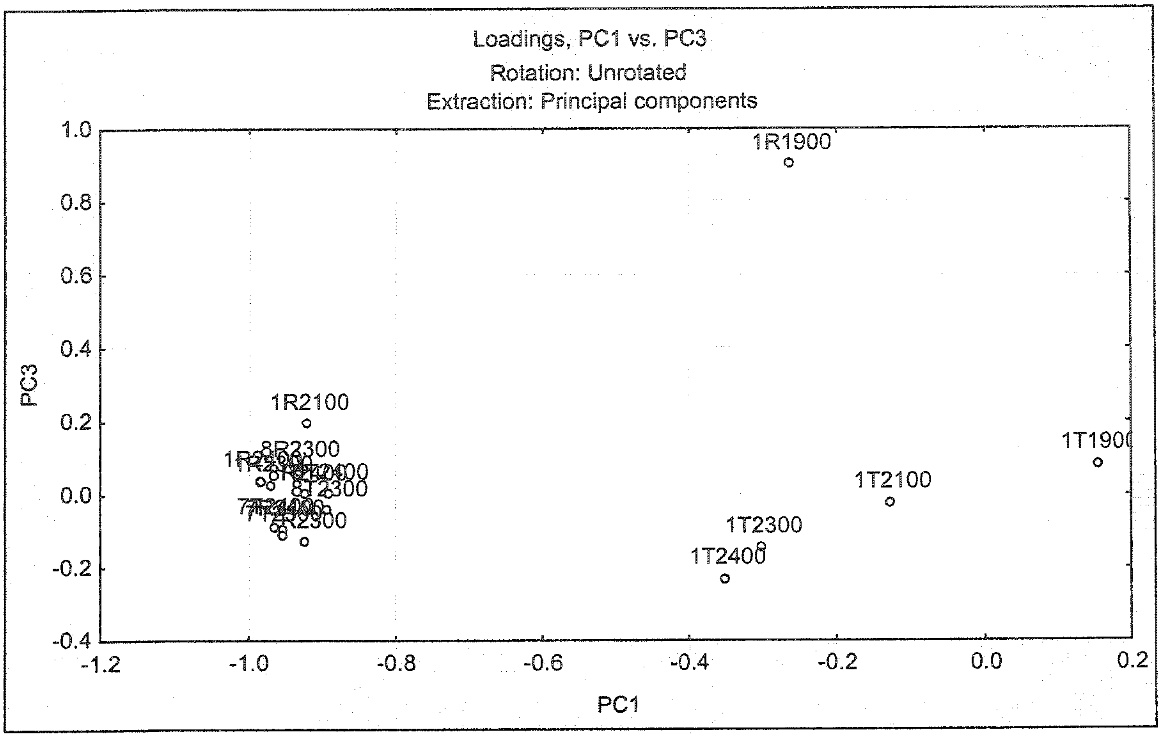


Figure 5.38, Loadings for PC's 1 and 3

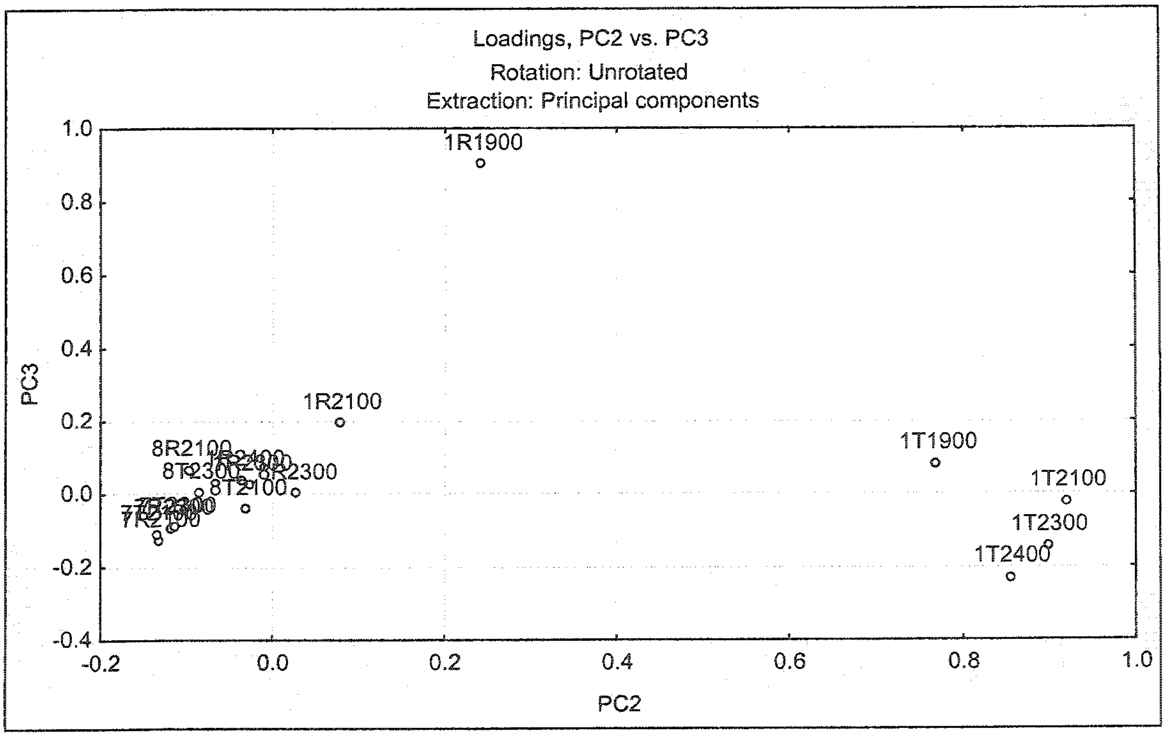


Figure 5.39, Loadings for PC2 and PC3

Table 5.14, Loading of principal components, pendulums

Variable	Loadings PC1	PC2	PC3
1R1900	-0.26	0.24	**0.90**
1R2100	**-0.92**	0.08	0.20
1R2300	**-0.97**	-0.03	0.03
1R2400	**-0.99**	-0.04	0.04
1T1900	0.15	**0.77**	0.08
1T2100	-0.13	**0.92**	-0.02
1T2300	-0.30	**0.90**	-0.14
1T2400	-0.35	**0.85**	-0.23
7R2100	**-0.92**	-0.13	-0.13
7R2300	**-0.96**	-0.12	-0.09
7T2100	**-0.96**	-0.13	-0.11
7T2300	**-0.97**	-0.11	-0.09
8R2100	**-0.93**	-0.10	0.07
8R2300	**-0.92**	0.03	0.00
8T2100	**-0.89**	-0.03	-0.04
8T2300	**-0.89**	-0.09	0.01
Expl.Var	10.03	3.11	0.99
Prp.Totl	0.63	0.19	0.06

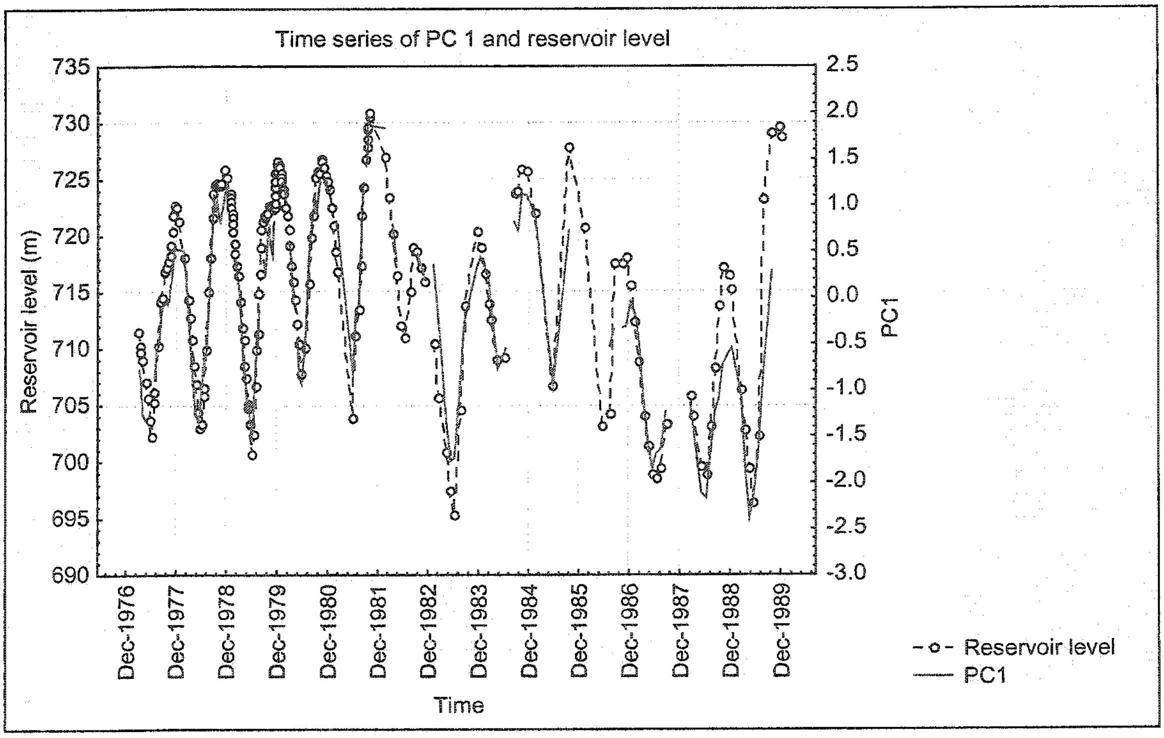


Figure 5 40, Time series of PC1 and Reservoir variations

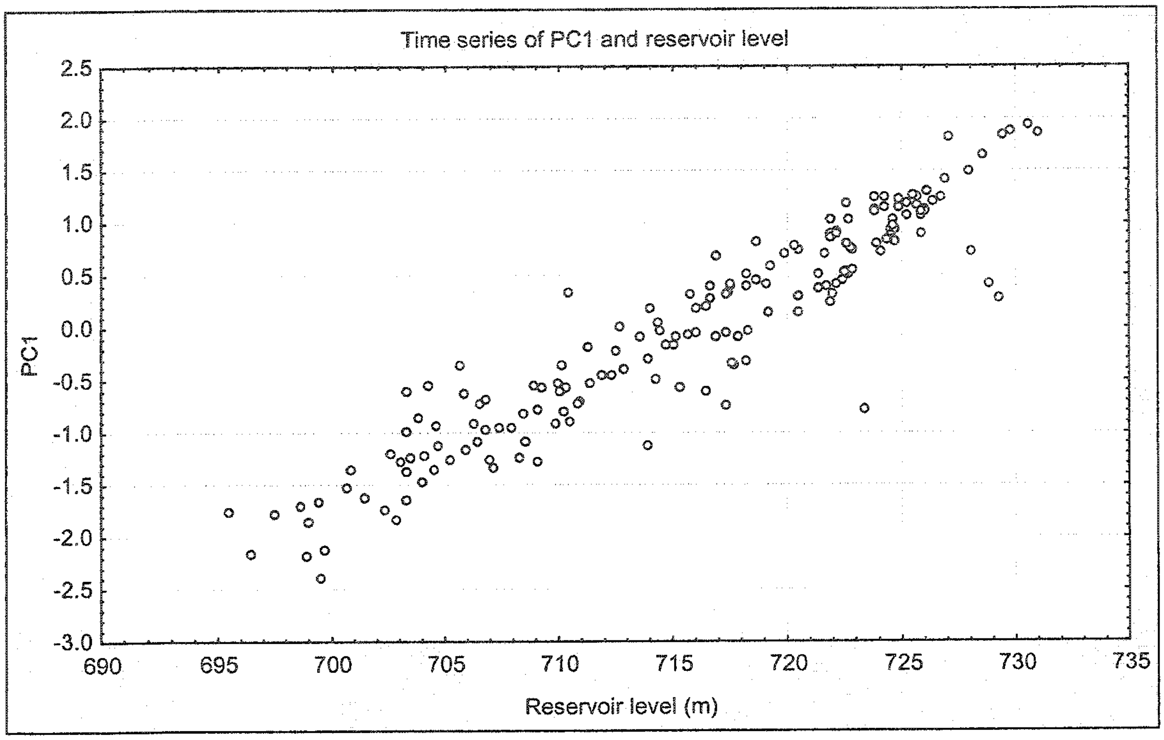


Figure 5.41, Predicted scores of PC1

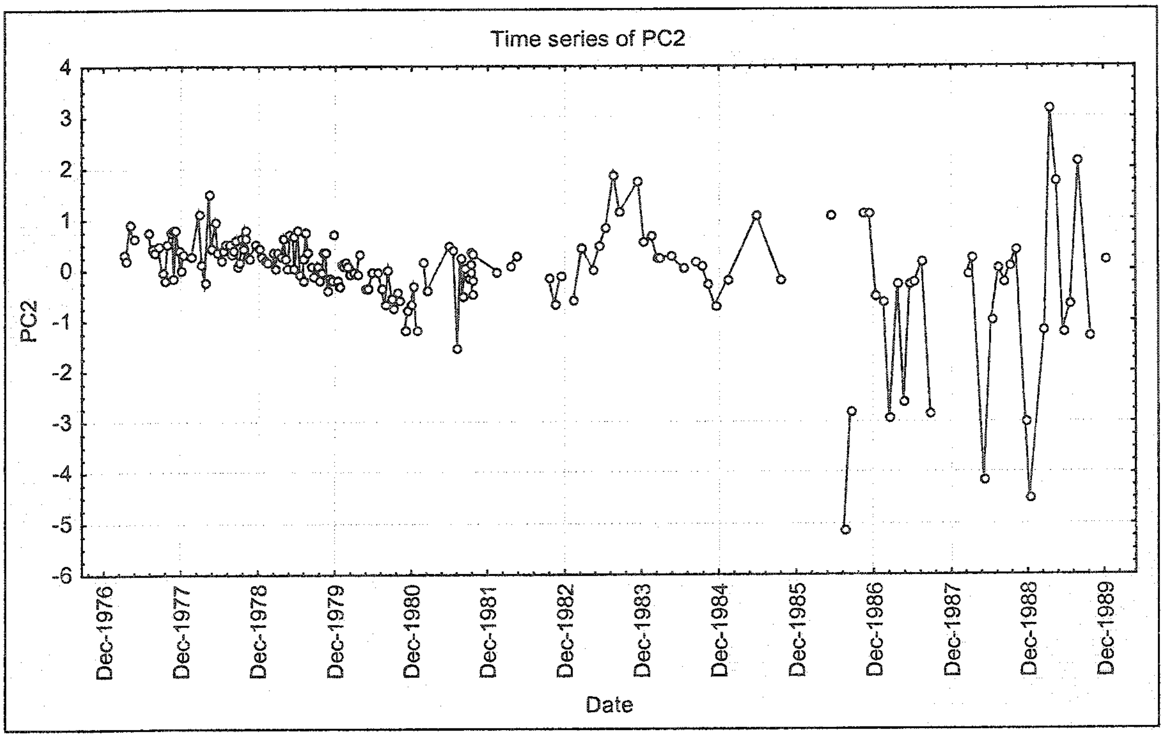


Figure 5.42, Scores of PC2, pendulum readings, Idukki dam

variable itself becomes one of the principal components. In extreme case when there is no correlation between variables, PCA reproduce the variables, thus if correlations are small there is little to be gained with a principal components analysis. PC2 shows movement of the dam towards the right bank after 1986.

It must be noted that Pendulum P5 measures absolute displacements of the dam at elevation 1900 ft.(571 m) This displacement is added successively to displacements from other pendulums to obtain their absolute displacements. Therefore most of anomalies are recorded at the base of the dam. The tangential displacements of pendulum P5, located at the base, are less than 1 mm during the period of 1977-1986. The tangential displacements increase during the period from 1986 to 1989 and the maximum displacement in this period is 4.91 mm for a reservoir level of 698 m.

The PCA model can be used for monitoring the future observations. Number of monitored variables is decreased from 16 (number of original instruments) to the first three PC's. Scores of PC1 is used for monitoring radial displacements of block 1 at higher elevations, and radial and tangential displacements of blocks 7 and 8. Scores of PC2 are used for monitoring tangential displacements of block 1. The signal component of the instruments reading is separated from noise given that the noise component is by definition uncorrelated from one instrument to another. This will result in reducing number of false alarms as compared to the traditional practice of monitoring the 16 instruments independently.

For every new observation scores are calculated and compared with the expected range specified by the reference PCA model. Univariate and bivariate control charts are useful for the detection of anomalies. When an anomaly is detected the instruments associated with that PC can be reviewed.

CHAPTER 6

6. Application of Methodology to Daniel Johnson and Chute-à-Caron dams

6.1 Introduction

In the northern regions of Canada, the ambient air temperature differentials between summer and winter can be as high as 45° C. This variation in air temperature induce thermal stresses that exceed the tensile strength of the concrete. The Daniel-Johnson dam (Fig. 6.1) located in Quebec, Canada, is an example of a large multiple arch dam that has undergone thermal damage. Chute-à-Caron is a gravity dam comprised of three sections. Two of the sections form an angle of 147°. These two segments interact due to thermal expansion and a construction joint, located in the drainage gallery has opened under this action. Some of the multivariate statistical analysis methods discussed in Chapter 3 are applied to these two dams.

6.2 Daniel Johnson Dam

The Daniel Johnson dam (Figure 6.1) is the largest multiple arch dams in the world. It is located 800 km northeast of the city of Montreal on the Manicouagan river. Its maximum height and length are about 165 m and 1315 m respectively (Veltrop et. al, 1990). It consists of 13 cylindrical arches supported by 14 buttresses (Figure 6.2). The dam includes a large main arch, two adjacent symmetrical transition arches, and eleven normal arches of identical geometry. The main arch spans over 150 m, has a thickness of 24 m at its base, and a thickness of 4.8 m at its crest.

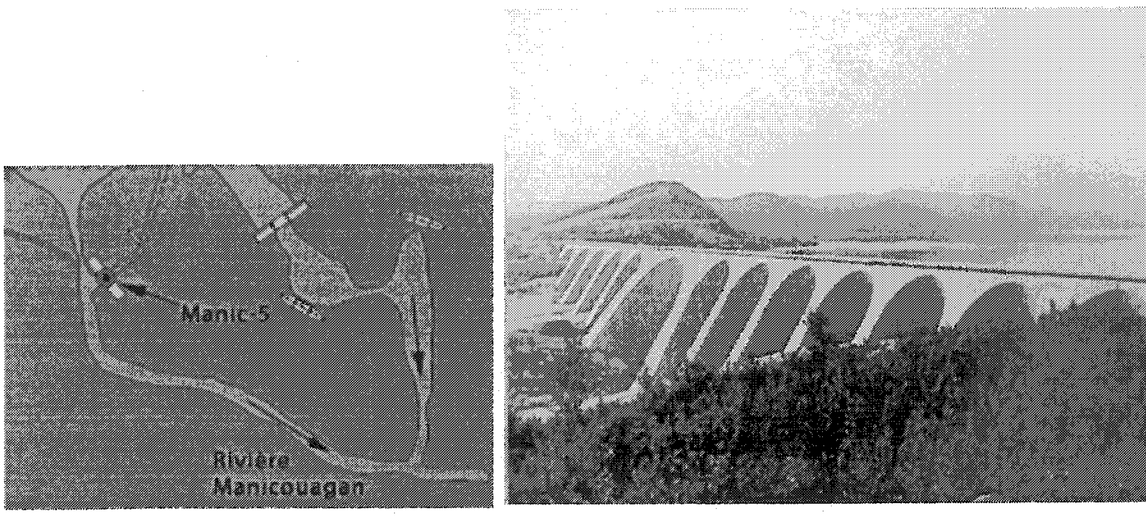


Figure 6.1, Location and view of Daniel Johnson multiple arch dam

Impounding of the lake started in 1964 and was completed in 1977. Shortly after construction, different types of cracks started to appear on the downstream and upstream faces of the dam. Eleven of the thirteen arches are known to have cracks.

Two kinds of cracks are present: The first type of cracks is known as plunging cracks. They cut the bottom of the arch in a plane perpendicular to the upstream face (at elevation 830 ft) and plunge towards the downstream face (Figure 6.3). These cracks were revealed by water seepage in the drainage system (Mamet, 1989). These plunging cracks were caused by stress concentrations due to geometric discontinuities and have been treated by a series of on-going grouting programs. In December 1992, the water infiltration rate of Arch 5-6 suddenly increased by 5 l/s in a matter of six weeks to reach an infiltration rate of 15 l/s (Figure 6.4). Progressive grout erosion was identified as the most probable reason for the increase in the water infiltration. Numerous cracks and joints were grouted in order to reduce water infiltration during and after the construction of the dam. In some cases, as a result of high injection pressures and inaccurate methods, the injection provoked the propagation of existing cracks or the initiation of the new ones (Larivière et al, 1999). As a result of the research work carried out at IREQ (Institut de Recherche d'Hydro-Quebec), last grouting program was successfully conducted in January 1999. The total water infiltration from the plunging cracks dropped from 15 l/s before the campaign to less than 0.1 l/s by the end of the grouting program.

2) The second type of cracks are visible "oblique" cracks on the downstream face of most of the arches, especially at the lower part. These cracks first appeared in 1968 in the lower downstream faces of arches. Field observations indicated that the cumulative length of downstream face oblique cracks continues to increase every year at a nearly constant rate (Tahmazian et al, 1989).

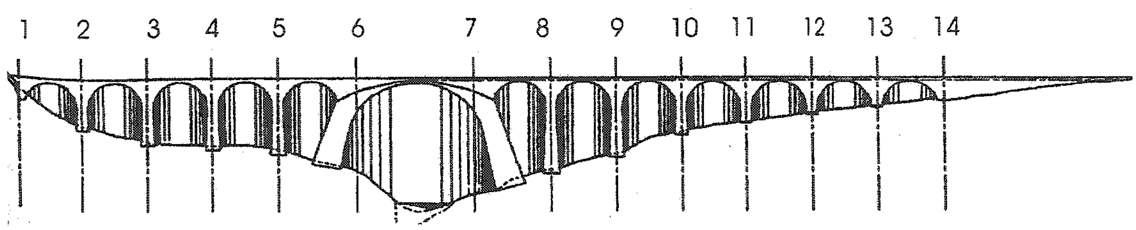


Figure 6.2, Elevation view, Daniel Johnson dam

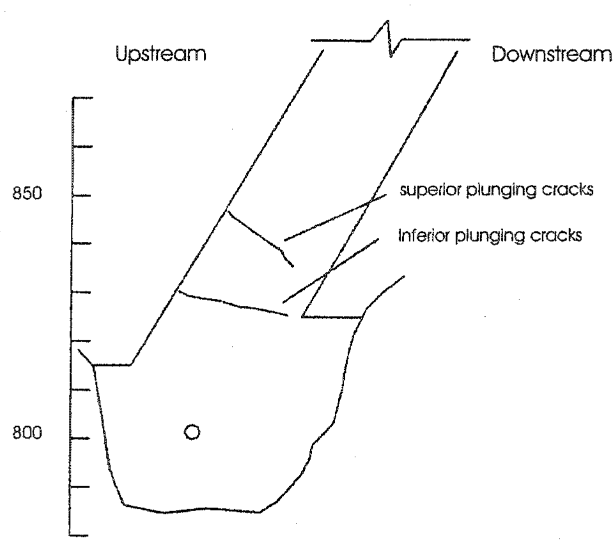


Figure 6.3, Plunging cracks, Arch 5-6, Daniel Johnson dam

Although the dam was stable, there was concern that continued cracking and exposure of the downstream face could lead to concrete deterioration Winter temperatures were identified as the primary cause of oblique cracks (Tahmazian et al, 1989). In 1986, shelters were installed at the downstream face of nine arches to insulate the bottom of the arches from extreme winter temperatures, which reduced the crack propagation. Field observations showed that the annual amount of new cracking in the nine sheltered arches decreased from 488 ft prior to installation of the shelters to 24 ft after installation. (Tahmazian et al, 1989).

6.2.1 Monitoring

A wide variety of devices and procedures are used to monitor the behavior of the Daniel Johnson dam. The following features are monitored: Movements, pore pressures, water level, seepage flows, temperatures, cracks and joints openings, stresses and strains. A comprehensive study of these measurements is necessary for checking the safety of the dam. Reservoir water level, air temperature, and concrete temperature measurements are useful for the prediction of structural responses. Study of dam displacements are useful for estimating the normal and abnormal behavior of a dam through comparison with the past performance data. Arch dams are very sensitive to base cracking (Fanelli, 1998). Analysis of cracks characteristics through visual inspection, extensometers readings, and water infiltration is useful for making decisions regarding safety and related actions regarding rehabilitation of the dam. Some of these measurements are discussed next.

Temperature and reservoir water level

Air temperatures and reservoir water level are recorded almost daily at the site. Maximum reservoir level of 359.6 m was reached in October 1979 (Figure 6.5). Air Temperatures vary between -32° C in winter to 32° C in the summer (Figure 6.6).

Pendulums

Simple and inverted pendulums are installed to measure the displacements. The frequency of available readings varies but the average frequency of readings is about one month.

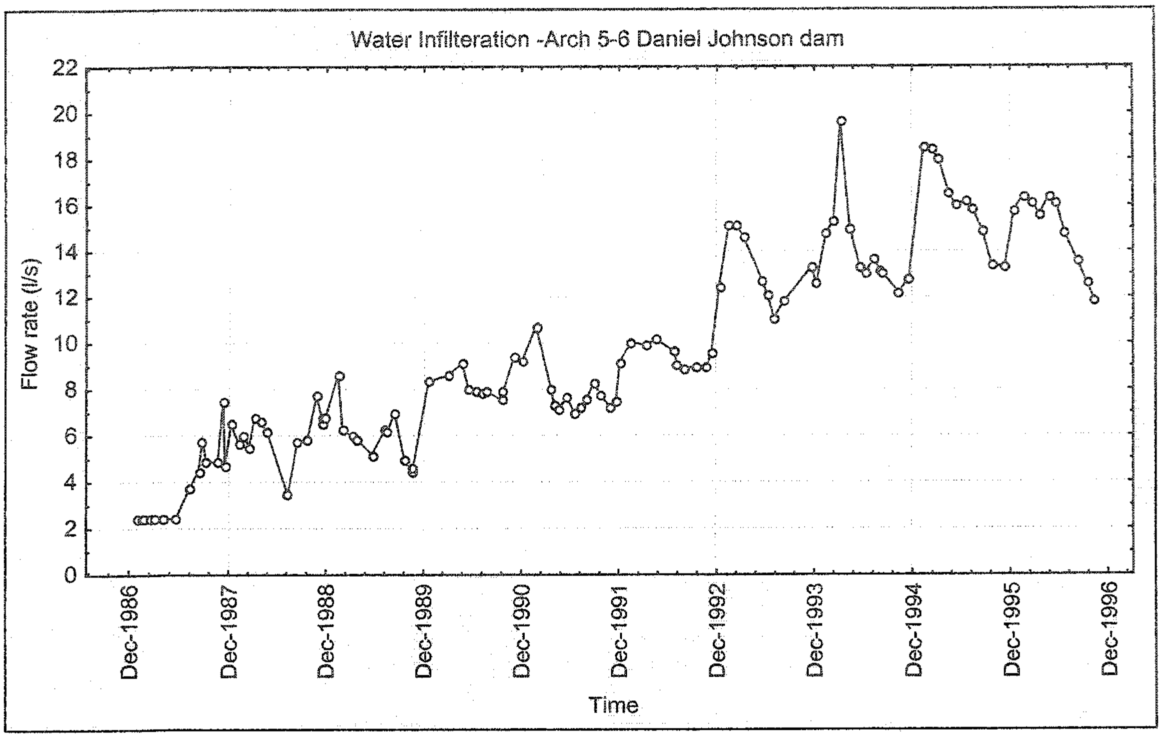


Figure 6.4, Water infiltration, Arch 5-6, Daniel Johnson dam

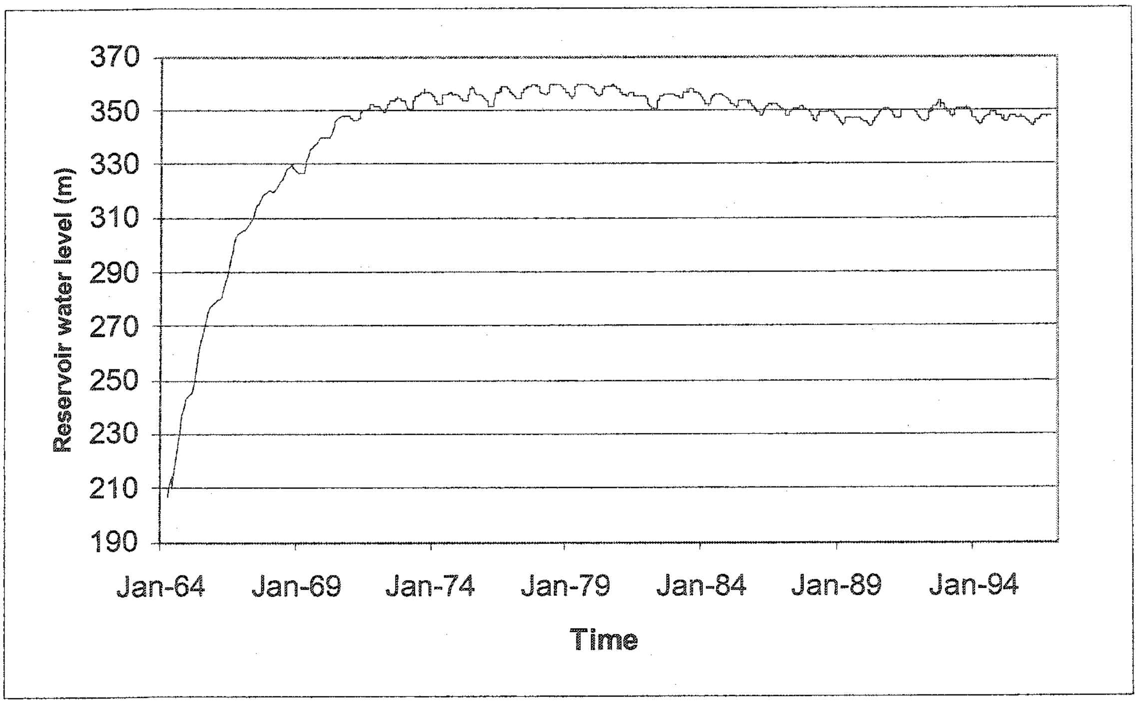


Figure 6.5, Reservoir water level variation, Daniel Johnson dam

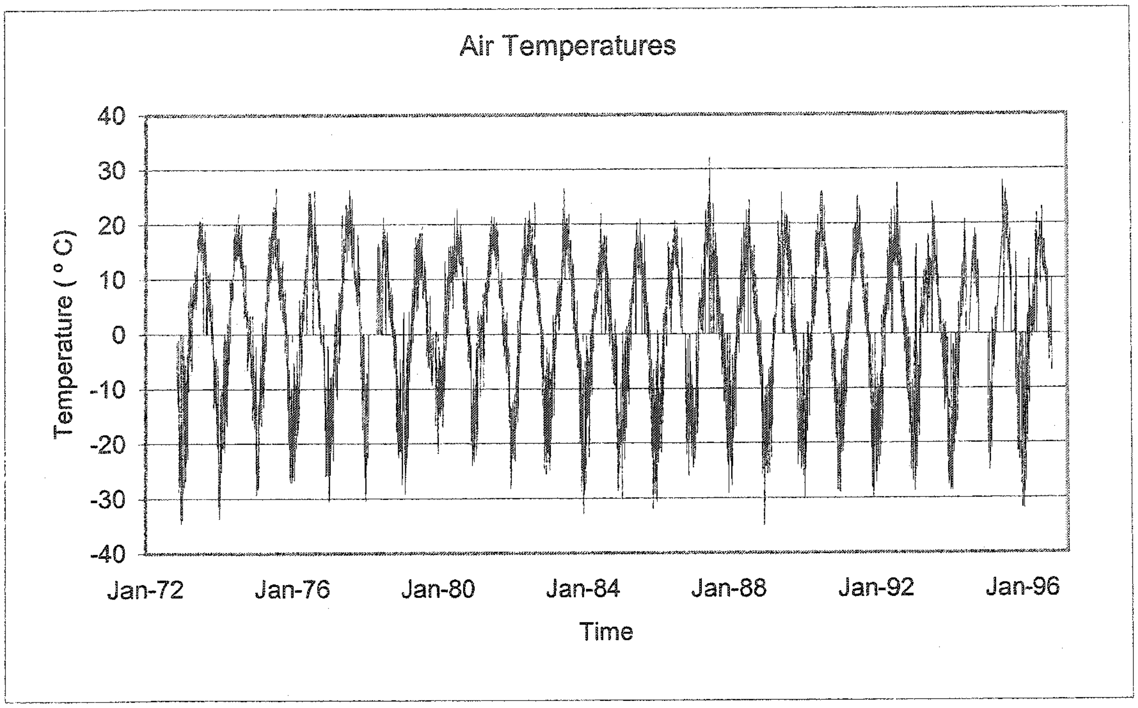


Figure 6.6, Daily air temperature variation, Daniel Johnson dam

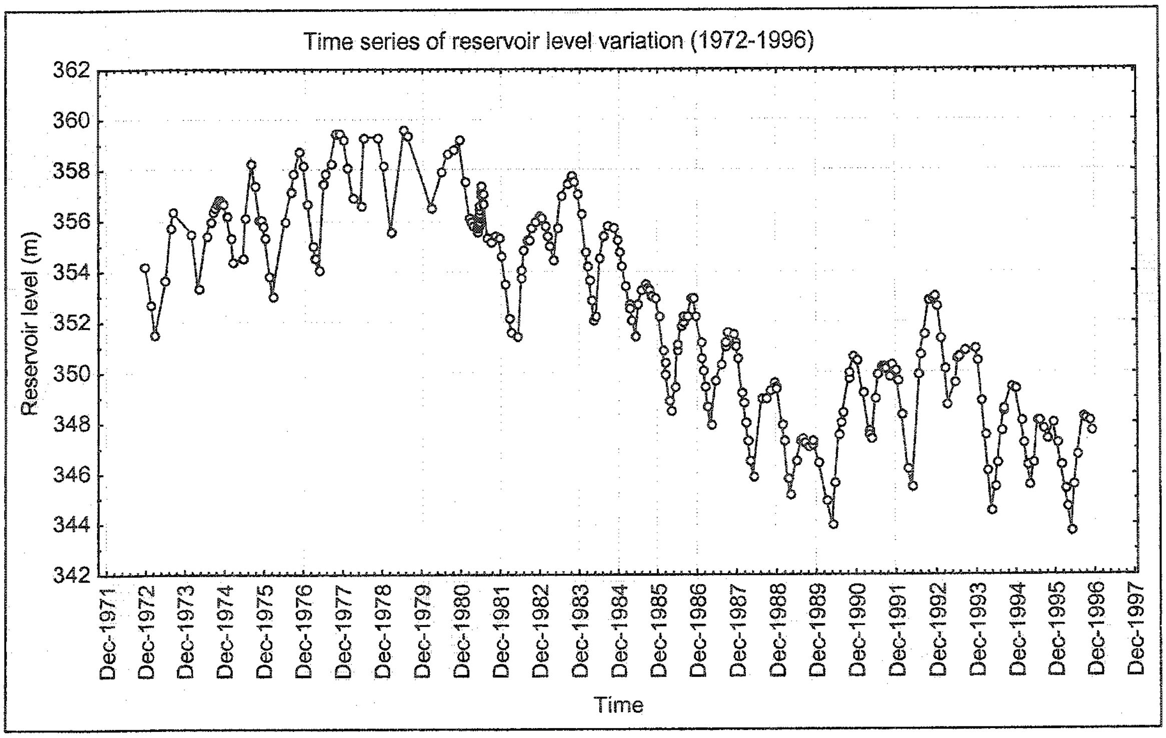


Figure 6.7, Reservoir level variation for the period 1972-1996, Daniel Johnson dam

Flow meters

These instruments measure the water infiltration through the dam, which is as an indicator of severity of cracking in the dam. The analysis of water infiltration data in conjunction with displacements or strains can be used to study the effect of cracks on the structural stability of the dam.

6.2.2 Statistical Analysis of Daniel Johnson Monitoring data

For the purpose of this analysis, data was provided by Hydro-Quebec for pendulums, flow meters and extensometers, located in arch 5-6 from 1966 to 1996. The readings from different instruments are not recorded simultaneously. Pendulum and extensometers measurements are analyzed in the following sections.

6.2.2.1 Analysis of simple Pendulums

Eight simple pendulums (S1 to S8) located in arch 5-6 are considered for statistical analysis. In the analysis of a dam over an extended period of time, care must be taken to insure that a data set is homogeneous and that changes in the structure or response of the dam are clearly identified and modeled. In the extreme cases, the data may be treated as several independent samples. For example during the first impoundment, irreversible displacements associated with settlements are observed and decrease during subsequent cycles of the water level in the reservoir. The analysis was performed on the data starting in 1972 when the normal reservoir level was reached (Figure 6.5). Reservoir level reached its maximum level of 359.6 m in October 1979 (Figure 6.7). The reservoir level has been lowered below this level since that time due to presence of cracks. Annual cycles are also observed with maximum and minimum reservoir levels in October and April.

Characteristics of the pendulums are summarized in Table 6.1. Each pendulum measures the deformations in three principal directions: 1) displacements in the Y direction are downstream-upstream displacements, 2) displacements in the X direction are in-plane displacements, and 3) displacements in the Z direction are vertical displacements.

Vertical displacements are typically small, and are missing for most of the instruments. The maximum displacements are in the Y direction (upstream-downstream), and therefore these displacements were considered for statistical analysis. Measurements of pendulum S6 is different from the trend of other pendulums. Some suspected outliers (spikes) are observed for some of the pendulums (Figure 6.9 to Figure 6.16). As expected the displacements also increase with the elevation of the pendulum (Table 6.1 and Figure 6.8). The movements of pendulums at buttress 6 are higher than those of buttress 5 since buttress 6 is closer to the center line of the dam (Figure 6.2).

Table 6.1, Location of the Pendulums, Daniel Johnson dam

Pendulum	**Buttress**	**Block**	**Elevation(m)**
S1	6	AB	244
S2	6	14	274
S3	6	20	309
S4	6	20	336
S5	6	20	356
S6	5	GH	275
S7	5	V	308
S8	5	W	335

Readings of pendulum S1 located near the base of Buttress 6, vary between 6.5 mm and 9 mm. The maximum displacements for both pendulums S3 and S4 occur in the winter, and are of 22 mm and 28 mm respectively. The absolute maximum displacement is 32.3 mm for pendulum S5 during the winter of 1984. Seasonal variations of displacements are observed in pendulums S3, S4 and S5, which are located at higher elevations. Measurements of S3, S4 and S5 are correlated with daily air temperature variations, which is illustrated for S5 in Figure 6.17 and Figure 6.18. The elliptic shape of the scatter plot indicates a lag between air temperature and displacements. Accurate calculation of lags cannot be done due to the low frequency of readings (once a month) for the pendulums. The maximum correlation (R=0.97) among pendulums is between pendulums S7 and S8 (Table 6.2 and Figure 6.19). This is expected since they are located in neighboring blocks of buttress 6. Pendulums S4 and S3 also show a high correlation, as they are located in the same block (block 20 of buttress 5). Pendulum S6 shows the lowest correlation with the rest of instruments.

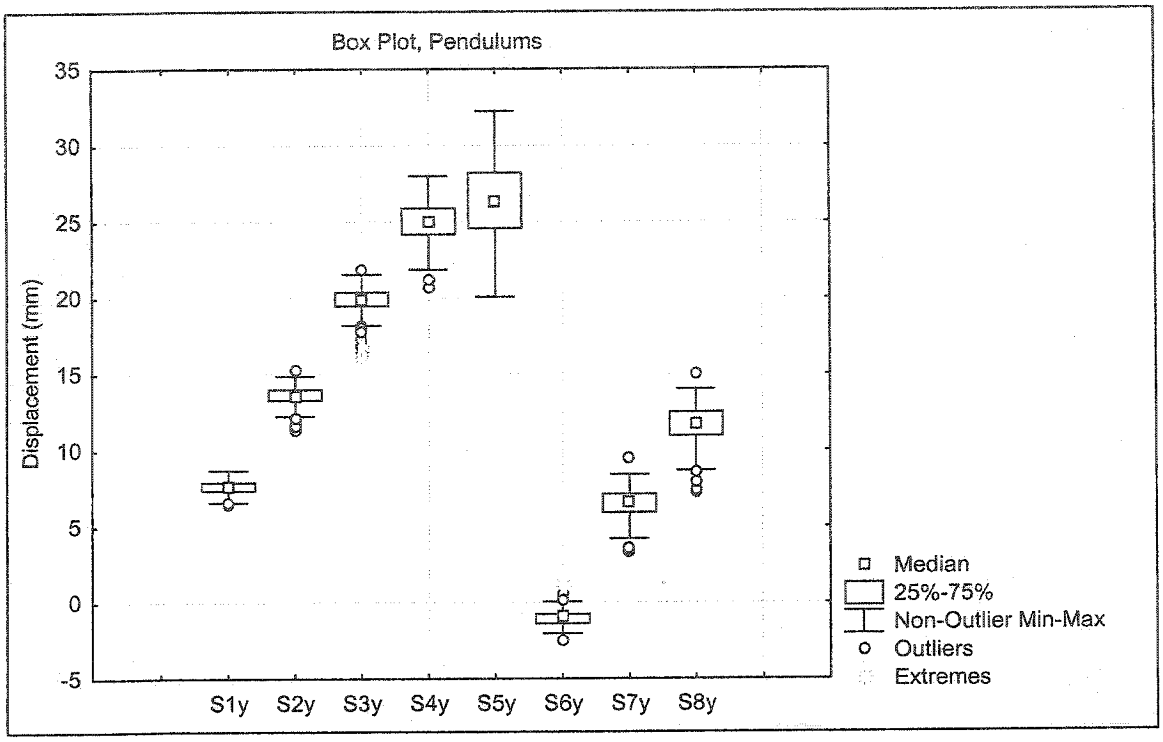


Figure 6.8, Box Plot for Pendulums, Daniel Johnson dam

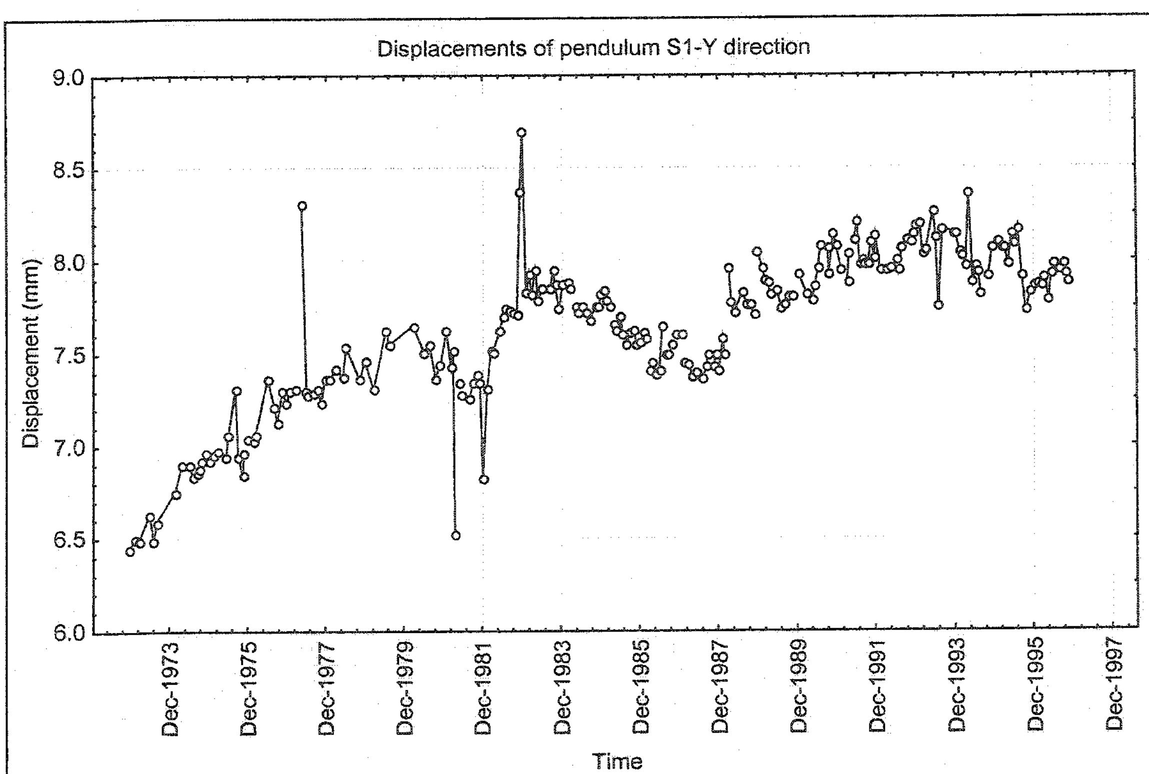


Figure 6.9, Upstream-downstream displacements of pendulum S1, Daniel Johnson dam

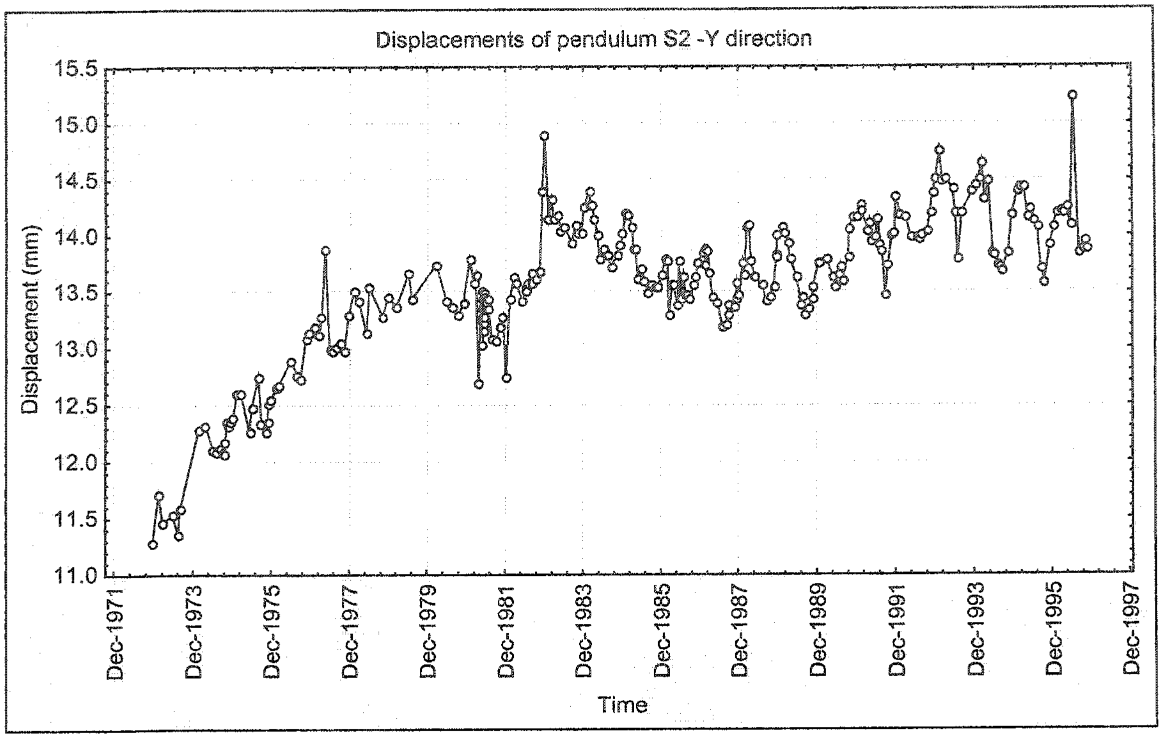


Figure 6.10, Upstream-downstream displacements of pendulum S2, Daniel Johnson dam

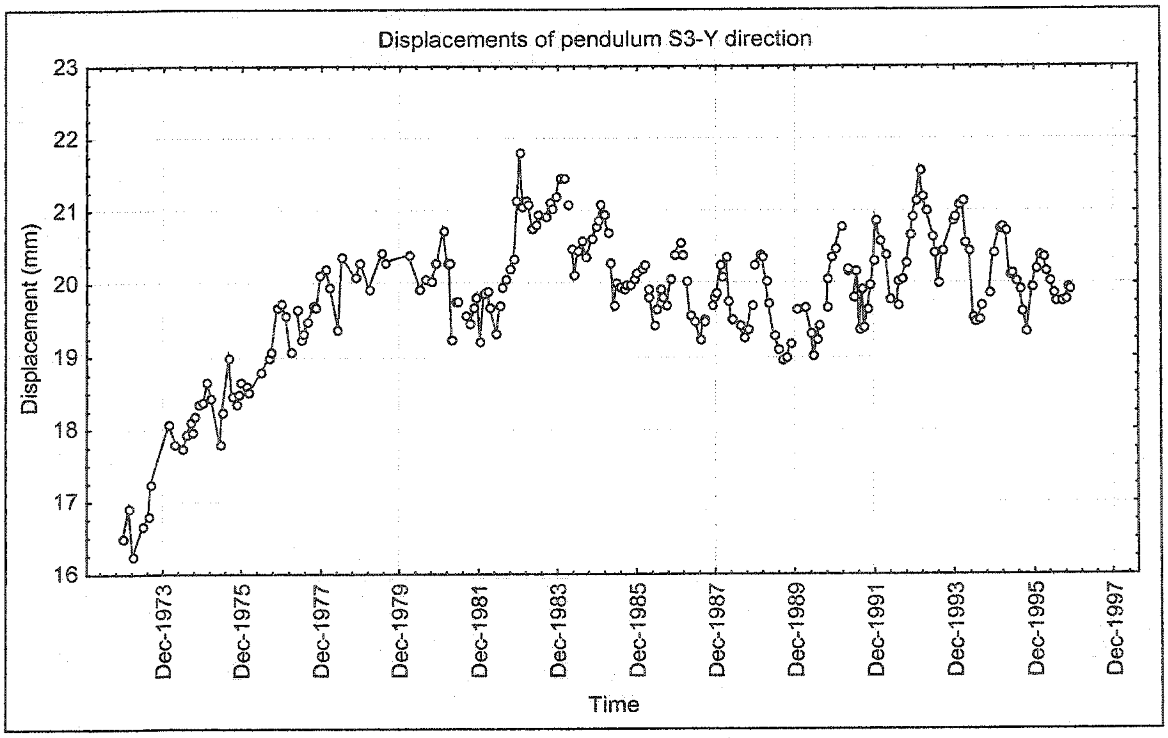


Figure 6.11, Upstream-downstream displacements of pendulum S3, Daniel Johnson dam

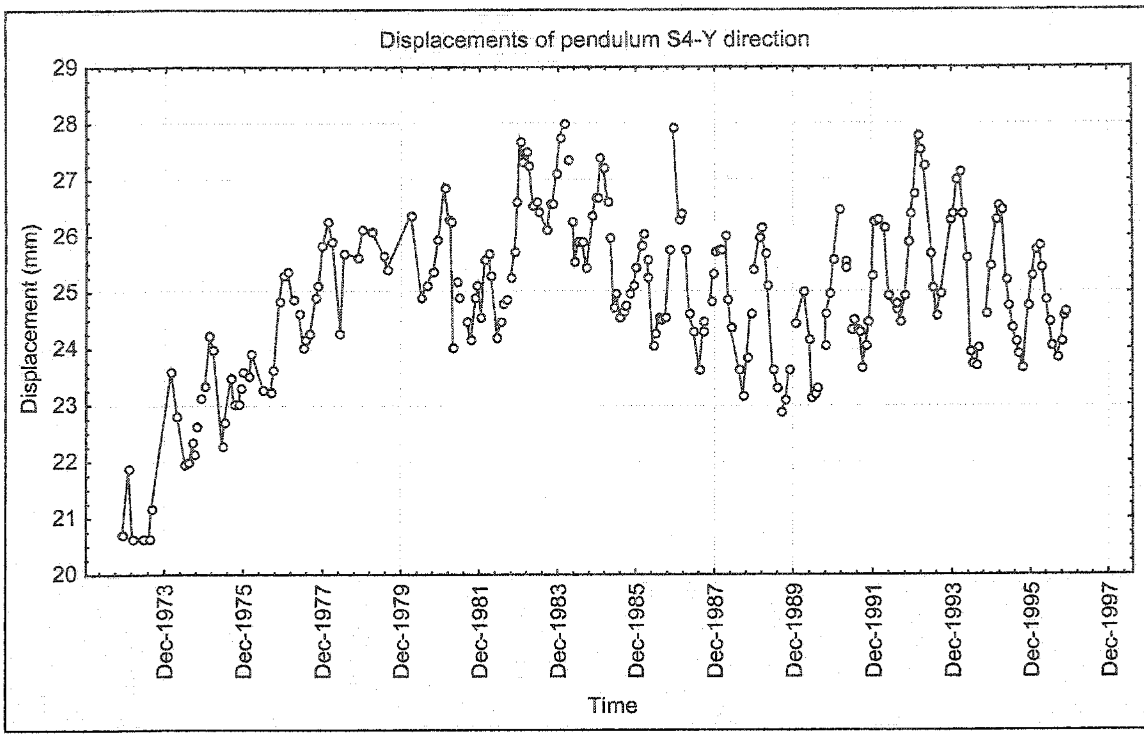


Figure 6.12, Upstream-downstream displacements of pendulum S4, Daniel Johnson dam

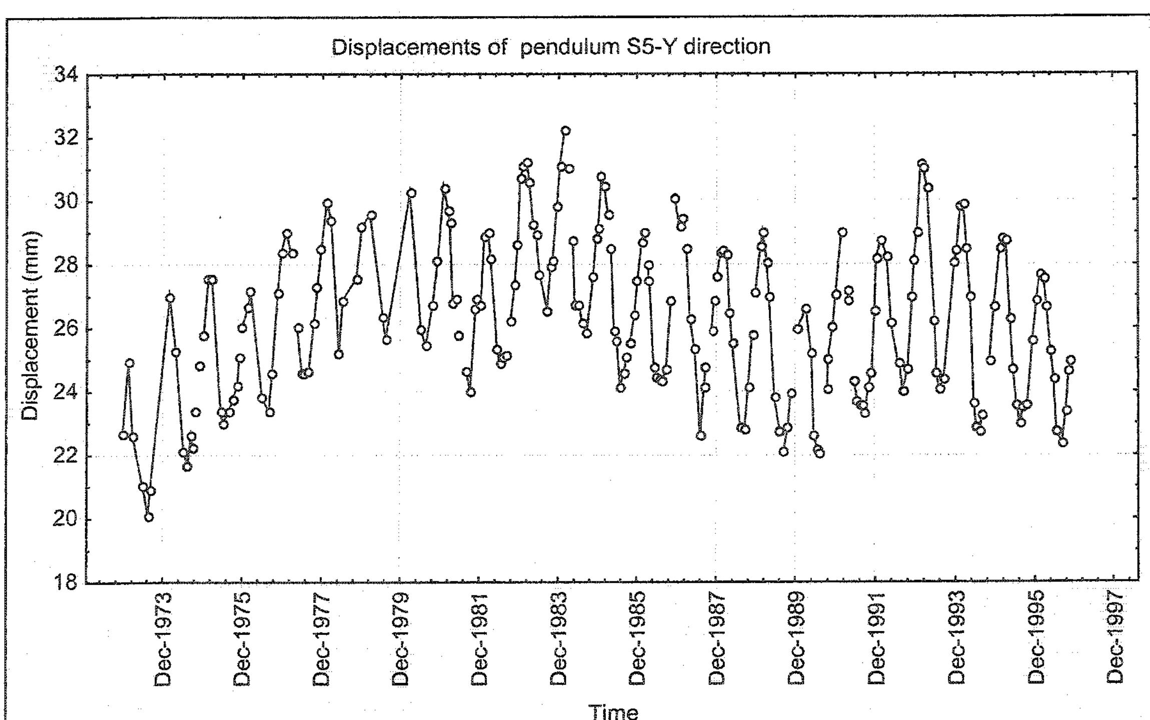


Figure 6.13, Upstream-downstream displacements of pendulum S5, Daniel Johnson dam

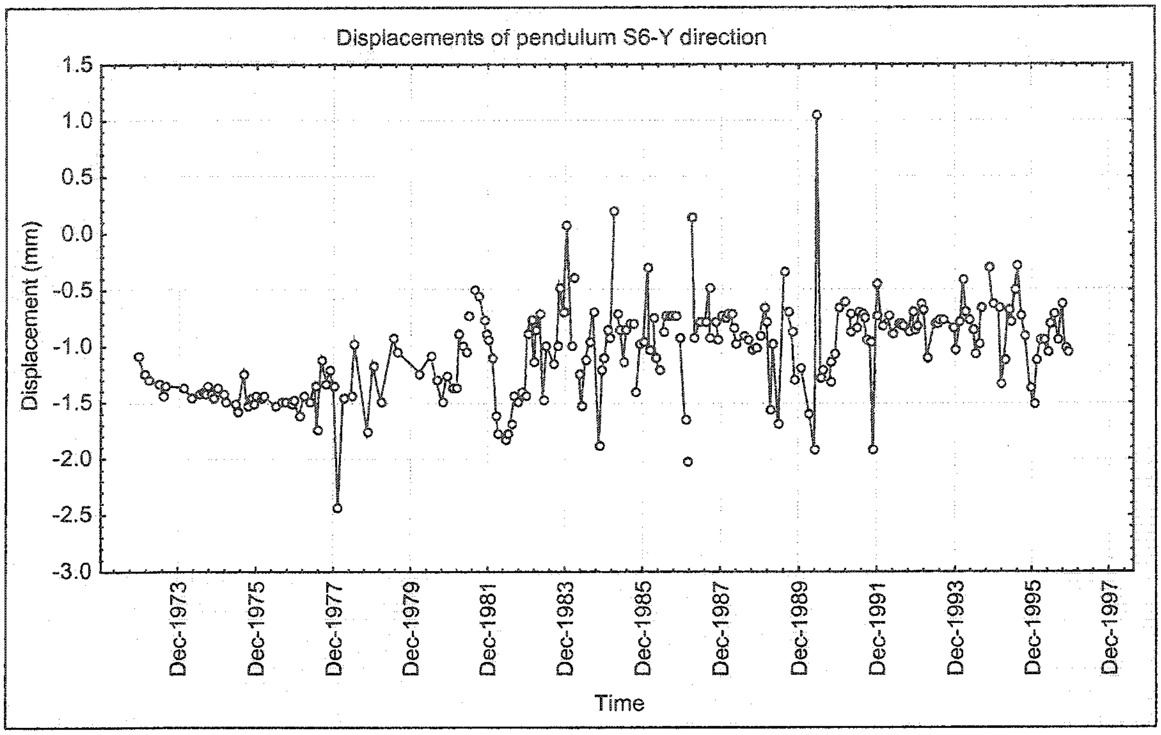


Figure 6.14, Upstream-downstream displacements of pendulum S6, Daniel Johnson dam

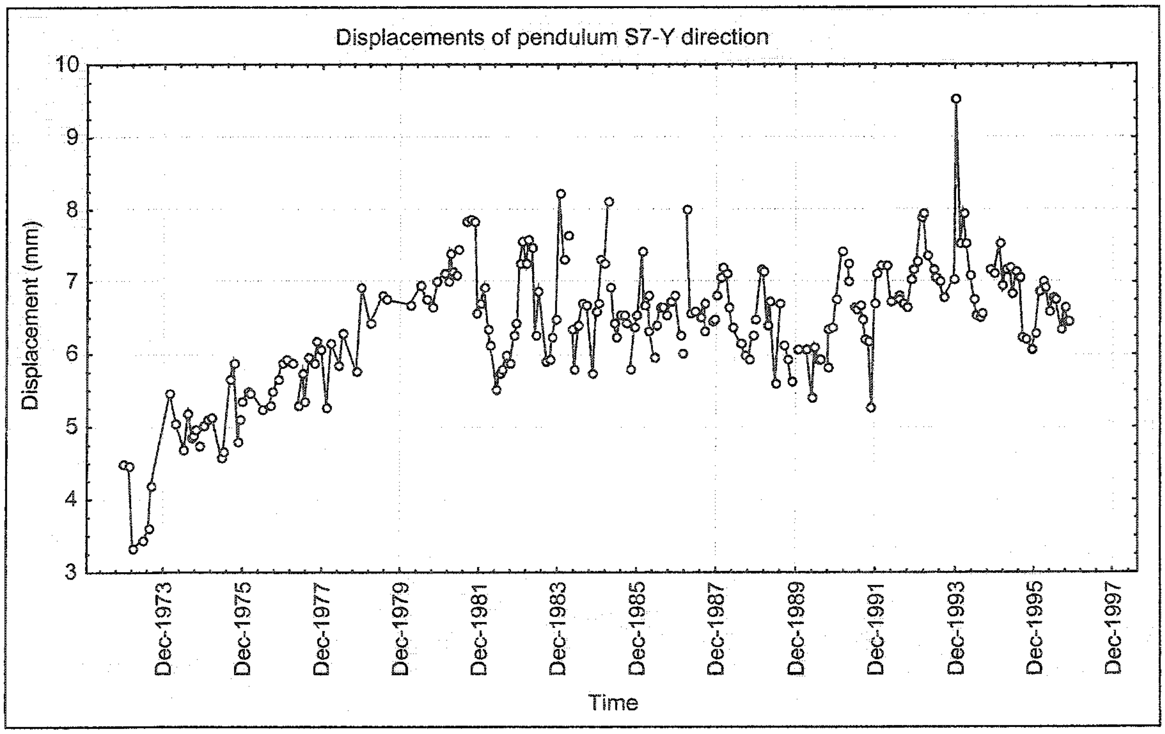


Figure 6.15, Upstream-downstream displacements of pendulum S7, Daniel Johnson dam

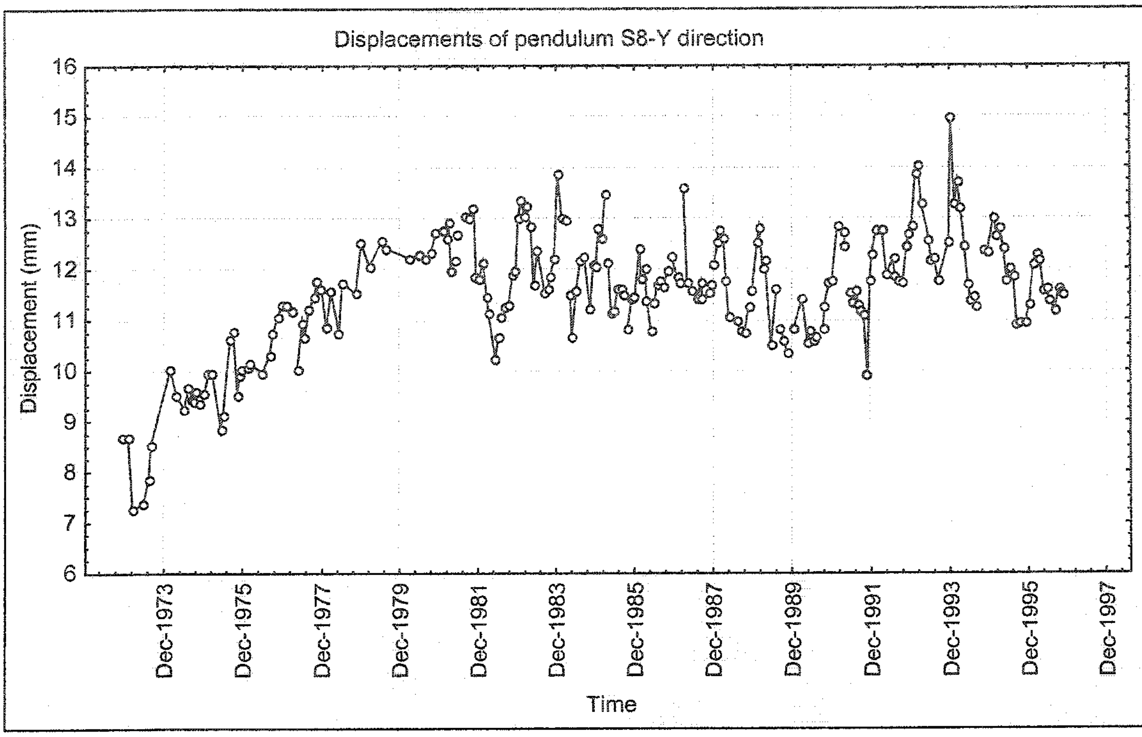


Figure 6.16, Upstream-downstream displacements of pendulum S8, Daniel Johnson dam

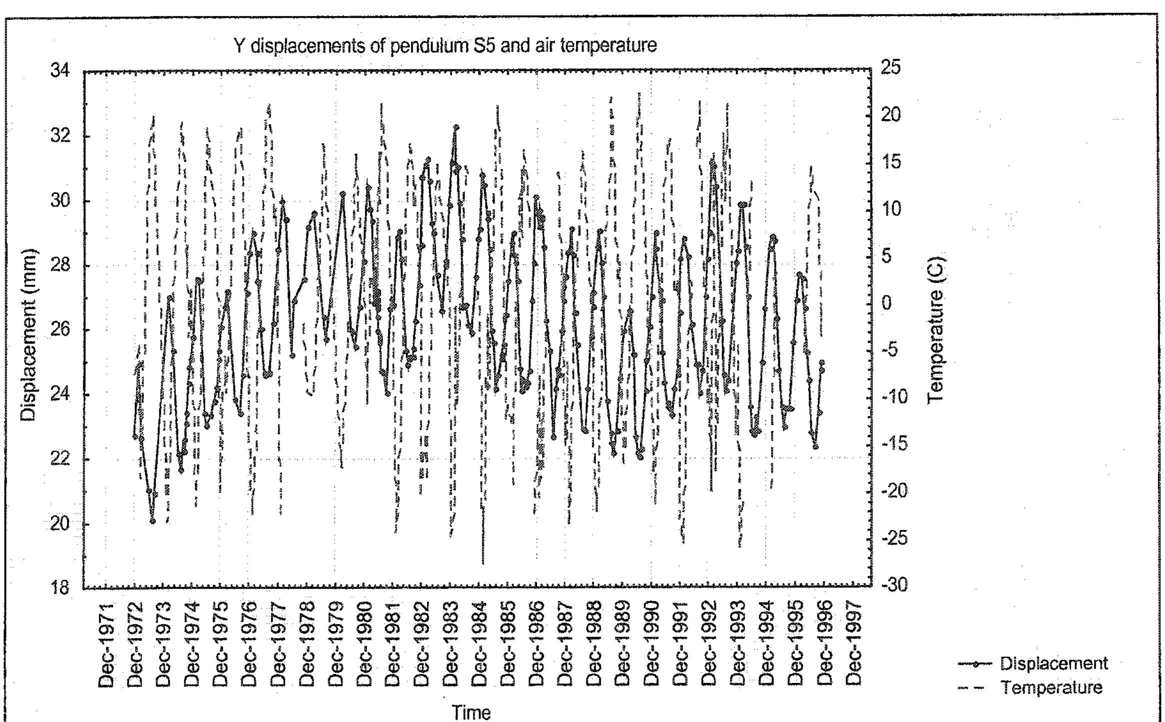


Figure 6.17, Upstream-downstream displacement of pendulum S5, Daniel Johnson dam

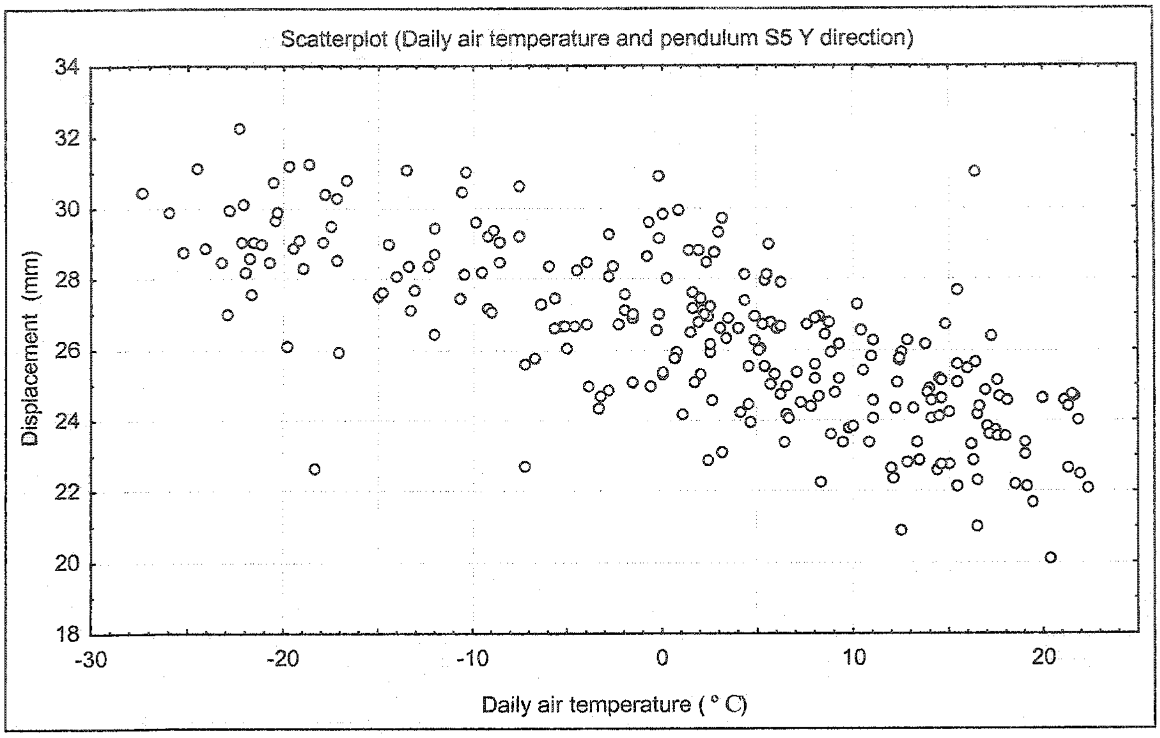


Figure 6.18, Scatter plot of displacements of S5 and daily air temperature, Daniel Johnson dam

Table 6.2, Correlation matrix for simple pendulums, Daniel Johnson dam

Variable	Correlations S1y	S2y	S3y	S4y	S5y	S6y	S7y	S8y
S1y	1.00	**0.92**	0.75	0.54	0.23	0.31	0.52	0.52
S2y	**0.92**	1.00	**0.90**	0.77	0.49	0.36	0.70	0.73
S3y	0.75	**0.90**	1.00	**0.93**	0.70	0.31	0.76	0.84
S4y	0.54	0.77	**0.93**	1.00	**0.90**	0.23	0.72	0.82
S5y	0.23	0.49	0.70	**0.90**	1.00	0.07	0.53	0.65
S6y	0.31	0.36	0.31	0.23	0.07	1.00	0.66	0.55
S7y	0.52	0.70	0.76	0.72	0.53	0.66	1.00	**0.97**
S8y	0.52	0.73	0.84	0.82	0.65	0.55	**0.97**	1.00

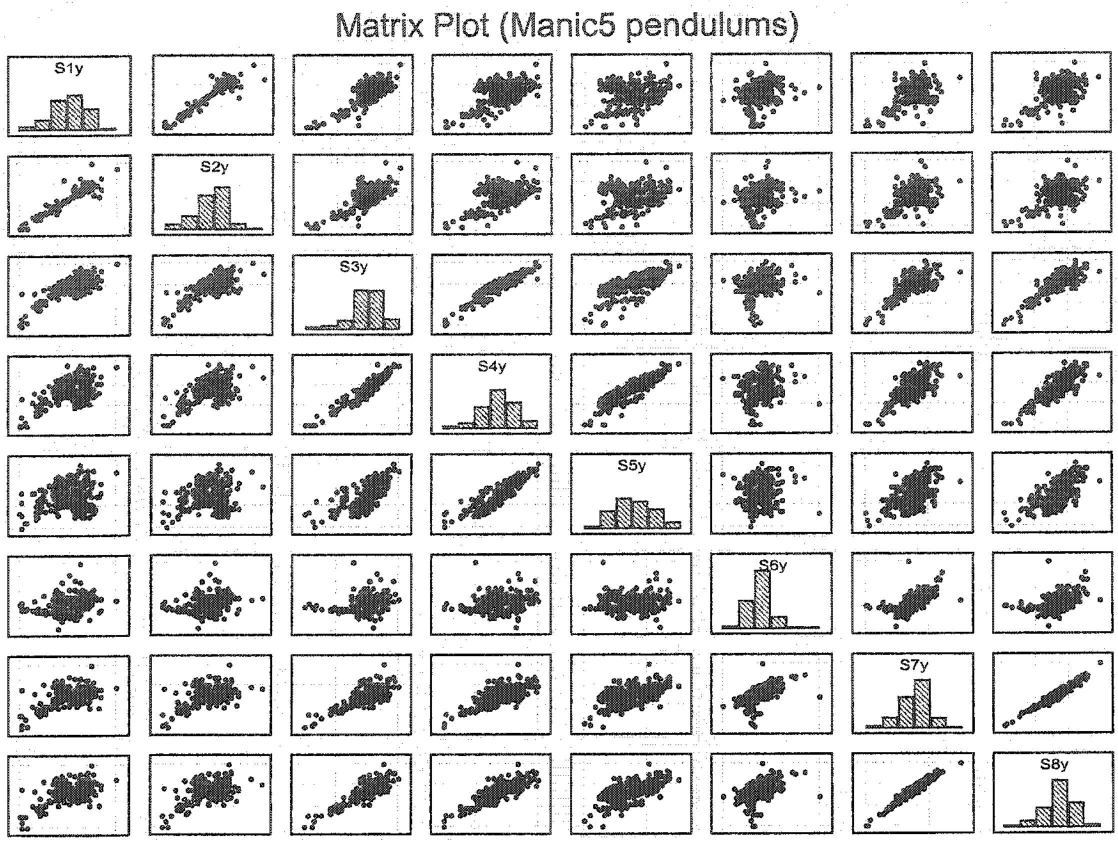


Figure 6.19, Matrix plot of pendulum displacements, Daniel Johnson dam

6.2.2.2 Principal component analysis of simple pendulums

PCA was applied to the correlation matrix of upstream-downstream displacements, which are higher and show more variation. The percentage of the variance explained by each principal component is shown in descending order in Table 6.3. The first three components explain 95.5% of the total variance. The remaining components explain only 4.5% of the variance and, most likely, they represent the noise in the data. The first principal component explains 70%; the second component 14%; and the third 11% of the total variance of the data. Analyses of loadings indicate that principal component 1 (PC1) is highly correlated with all instruments except S6y (Table 6.4 and Figure 6.23). PC2 is rather highly correlated with S6y. This is expected, as S6y is not strongly correlated with other instruments.

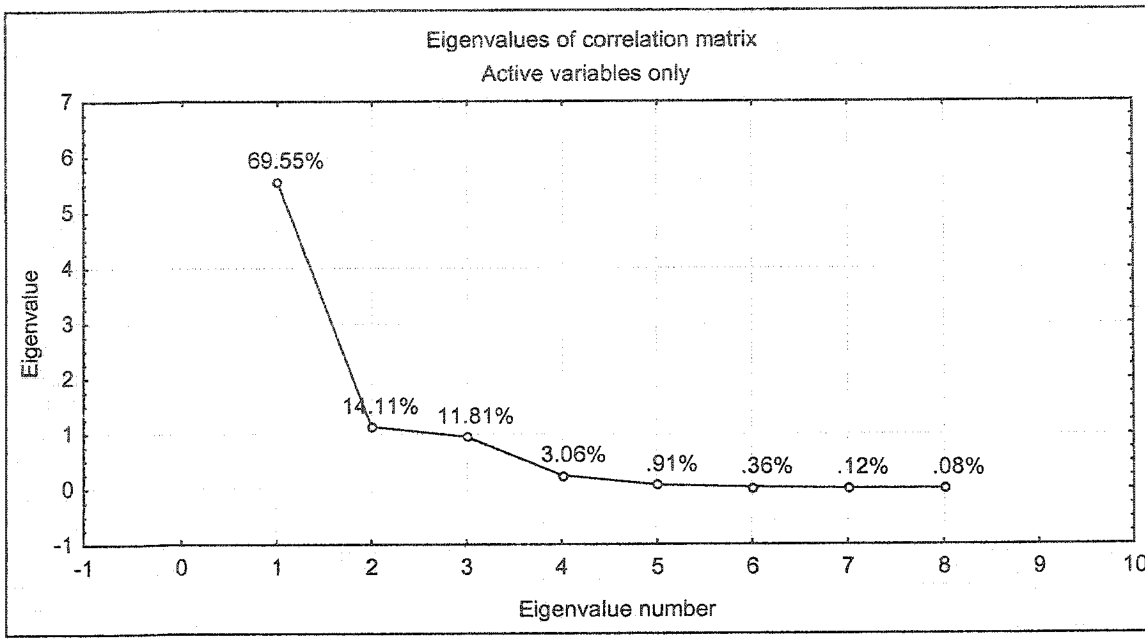


Figure 6.20, Scree plot of pendulums, Daniel Johnson dam

Table 6.3, PCA analysis results for Simple pendulums, Daniel Johnson dam

Principal Component	Eigenvalues	Variance Explained by each PC	Commutative Variance Explained
1	5.56	69.5%	69.5%
2	1.13	14.1%	83.6%
3	0.94	11.8%	95.5%

Table 6.4, Loadings of principal components

	Loadings		
Variable	PC1	PC2	PC3
S1y	-0.73	-0.18	-0.64
S2y	-0.90	-0.05	-0.40
S3y	-0.96	0.16	-0.14
S4y	-0.92	0.36	0.10
S5y	-0.71	0.57	0.33
S6y	-0.49	-0.76	0.31
S7y	-0.89	-0.27	0.28
S8y	-0.93	-0.10	0.27
Expl.Var	5.51	1.18	0.95
Prp.Totl	0.69	0.15	0.12

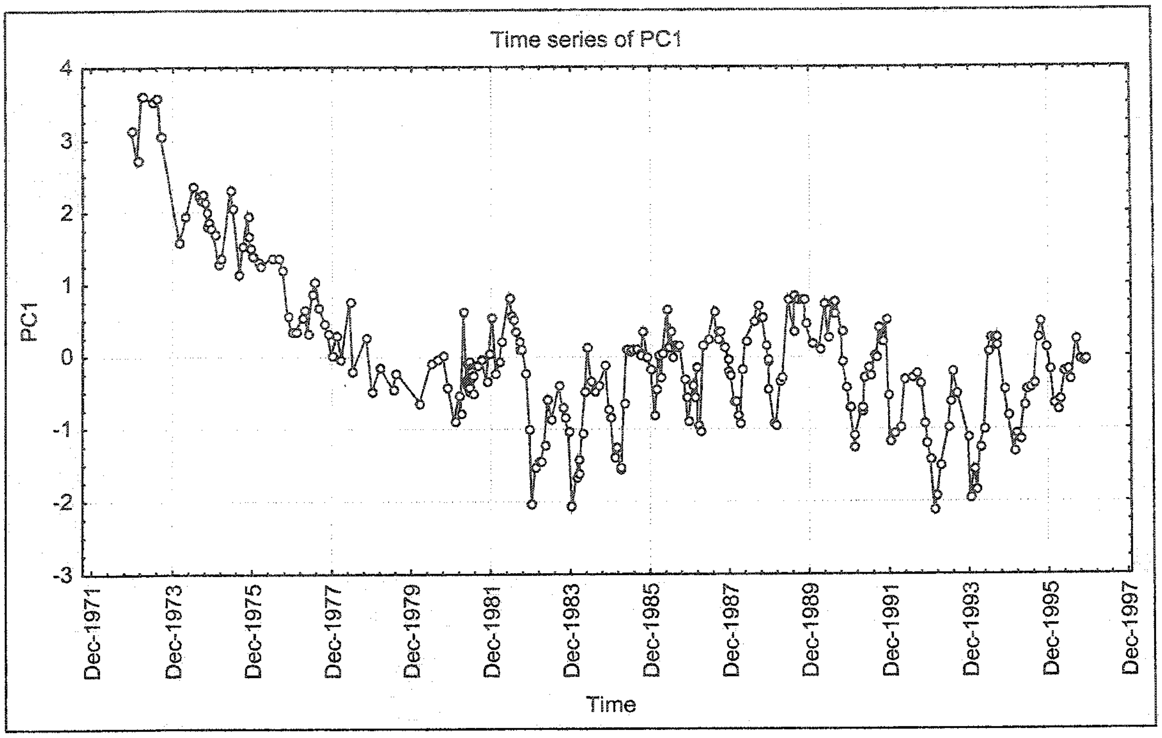


Figure 6.21, Scores of PC1, simple pendulums, Daniel Johnson dam

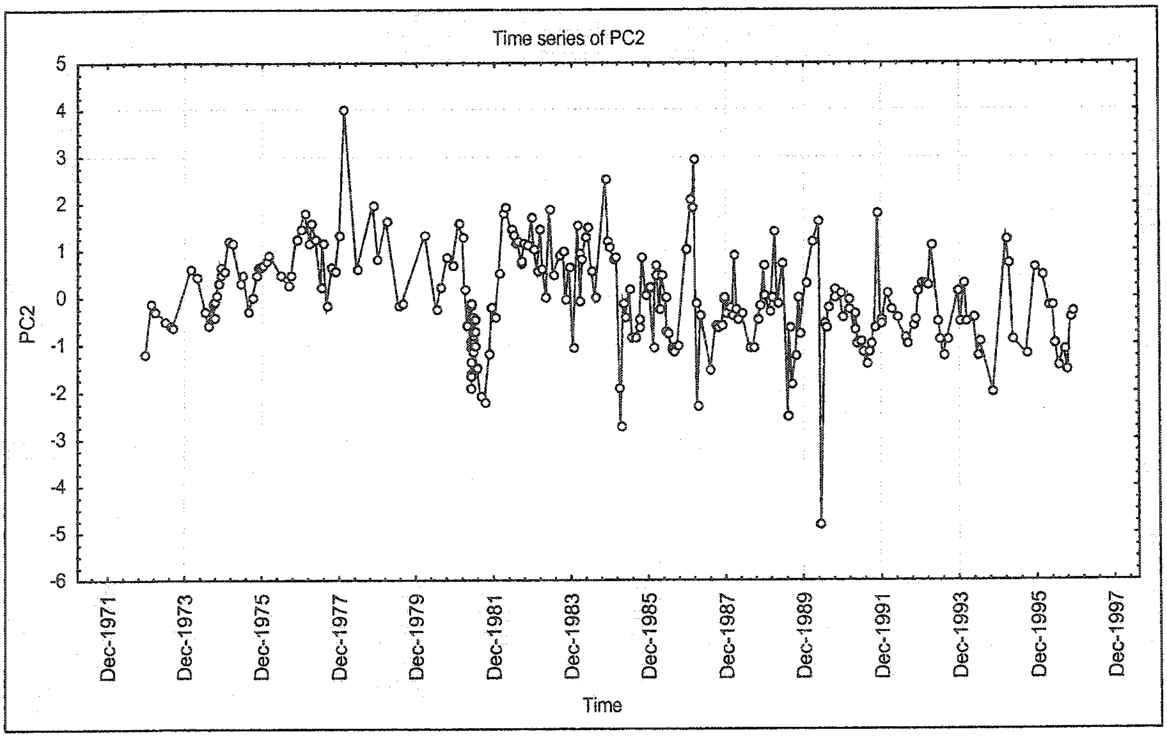


Figure 6.22, Scores of PC2, simple pendulums, Daniel Johnson dam

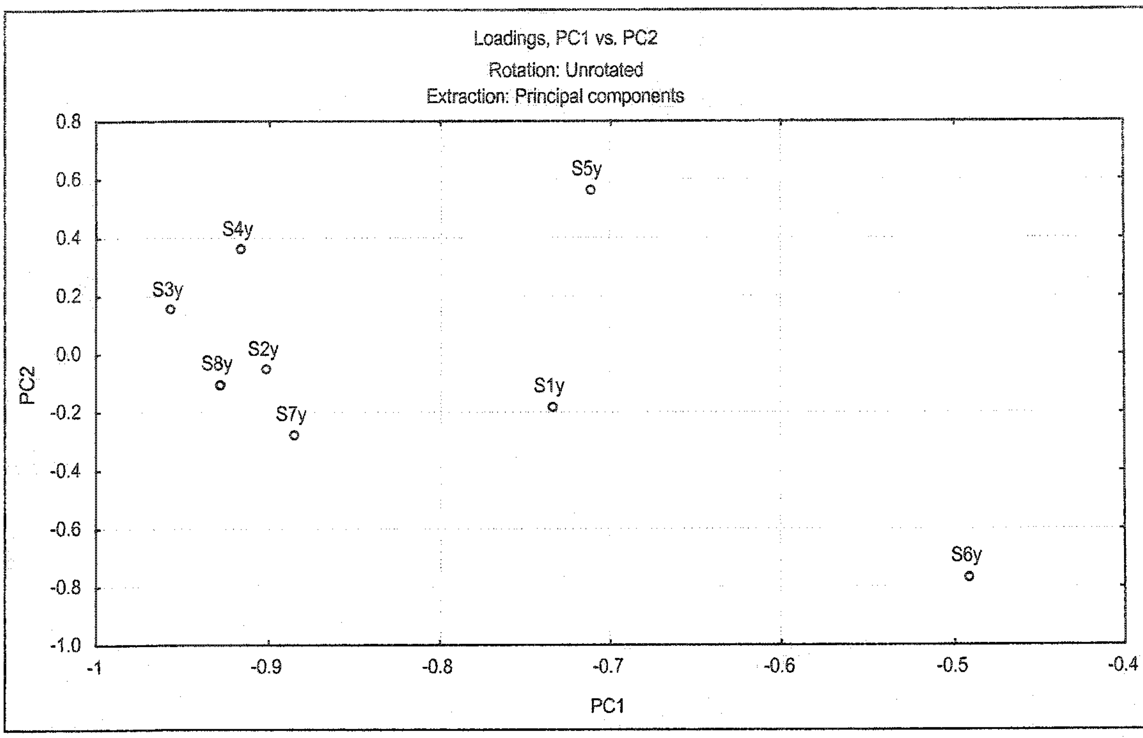


Figure 6.23, Loadings for PC1 and PC2, pendulums, Daniel Johnson dam

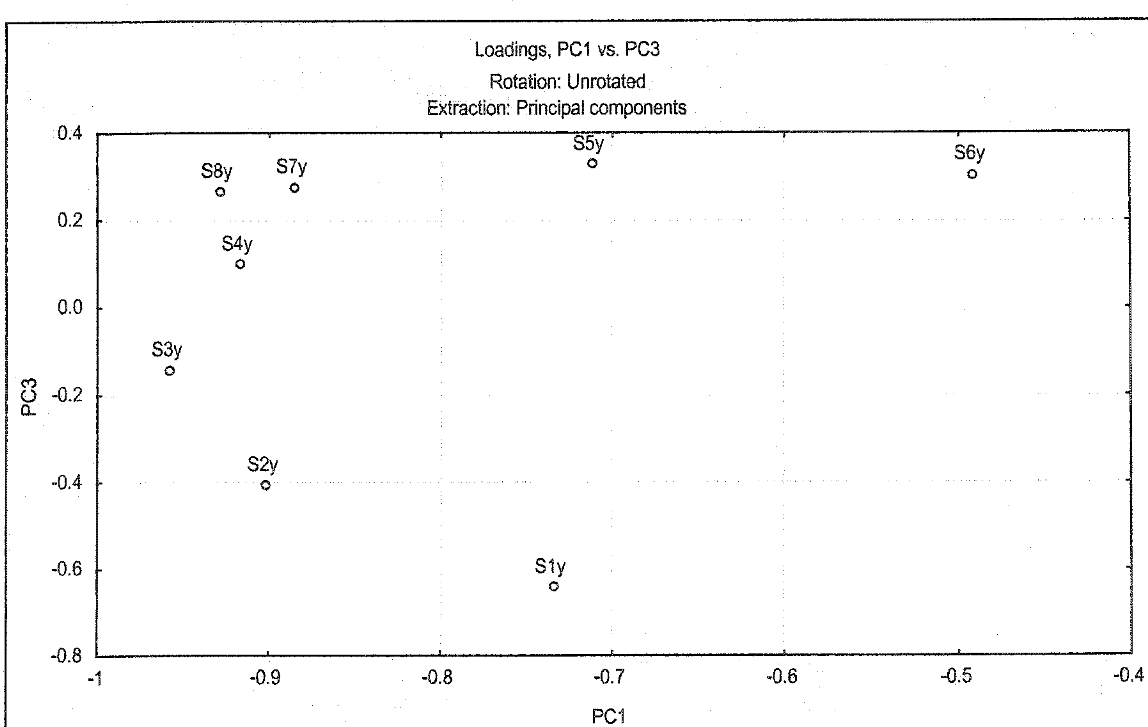


Figure 6.24, Loadings for PC1 and PC3, Daniel Johnson dam

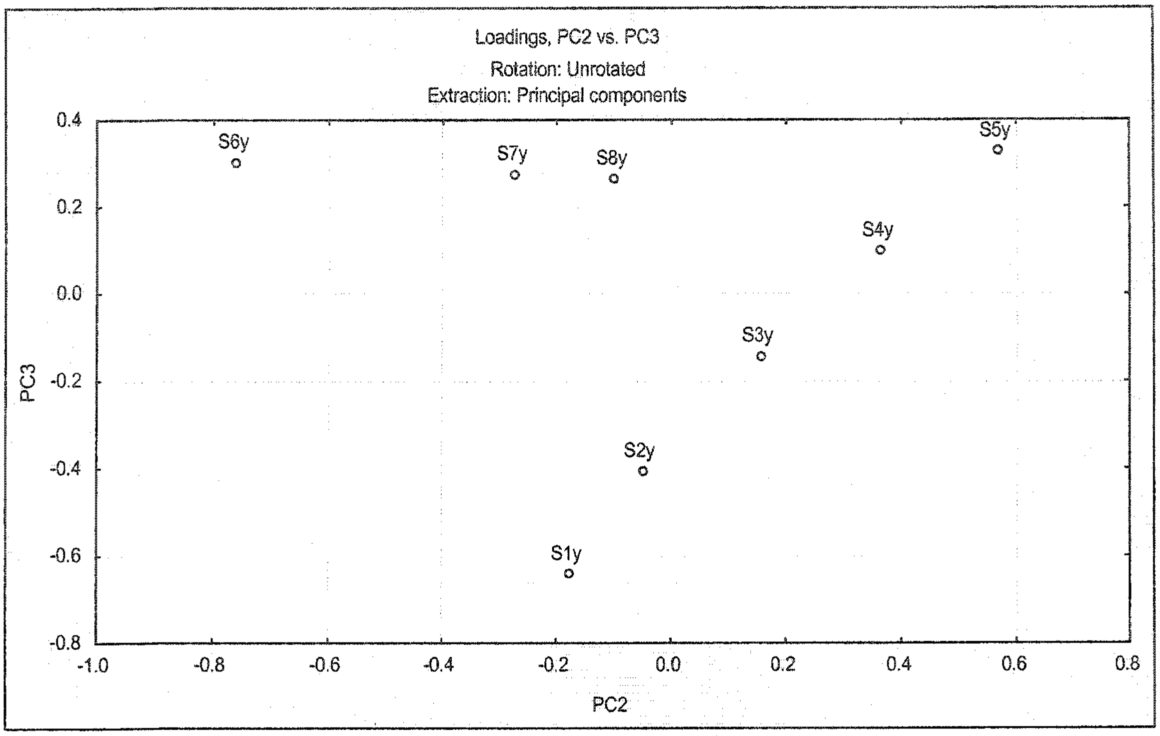


Figure 6.25, Loadings for PCs 2 and 3, Daniel Johnson dam

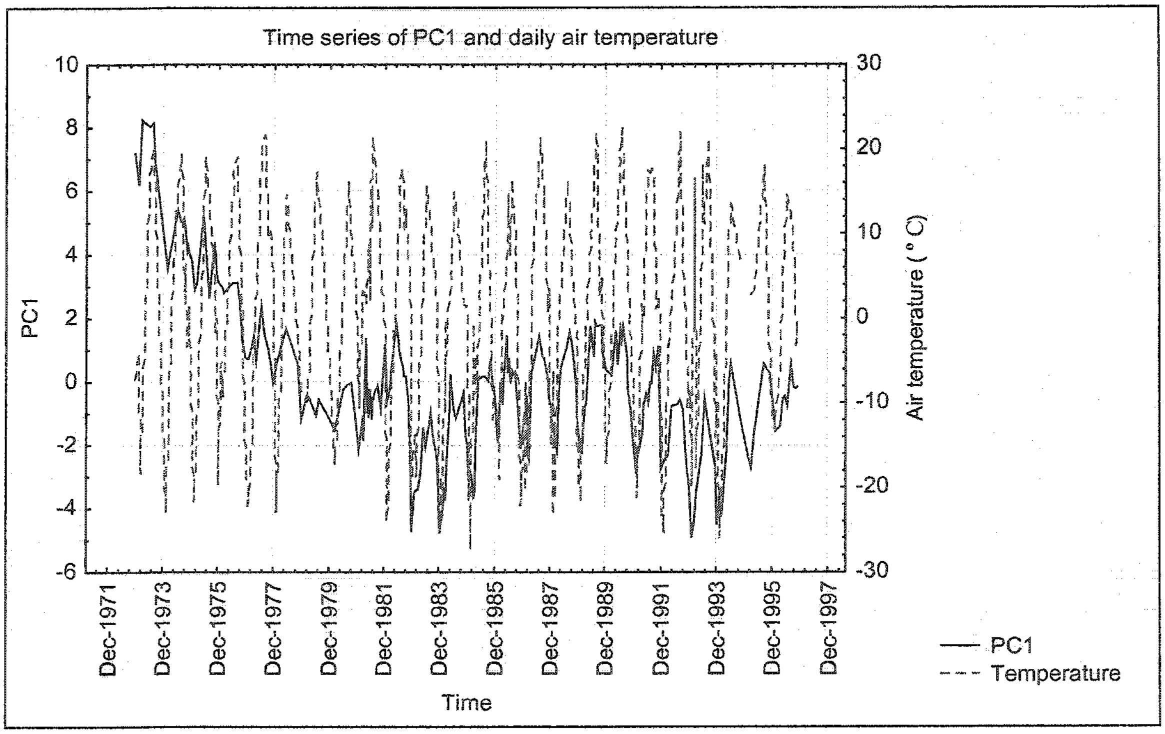


Figure 6.26, Scores of PC1 and daily air temperature, Daniel Johnson dam

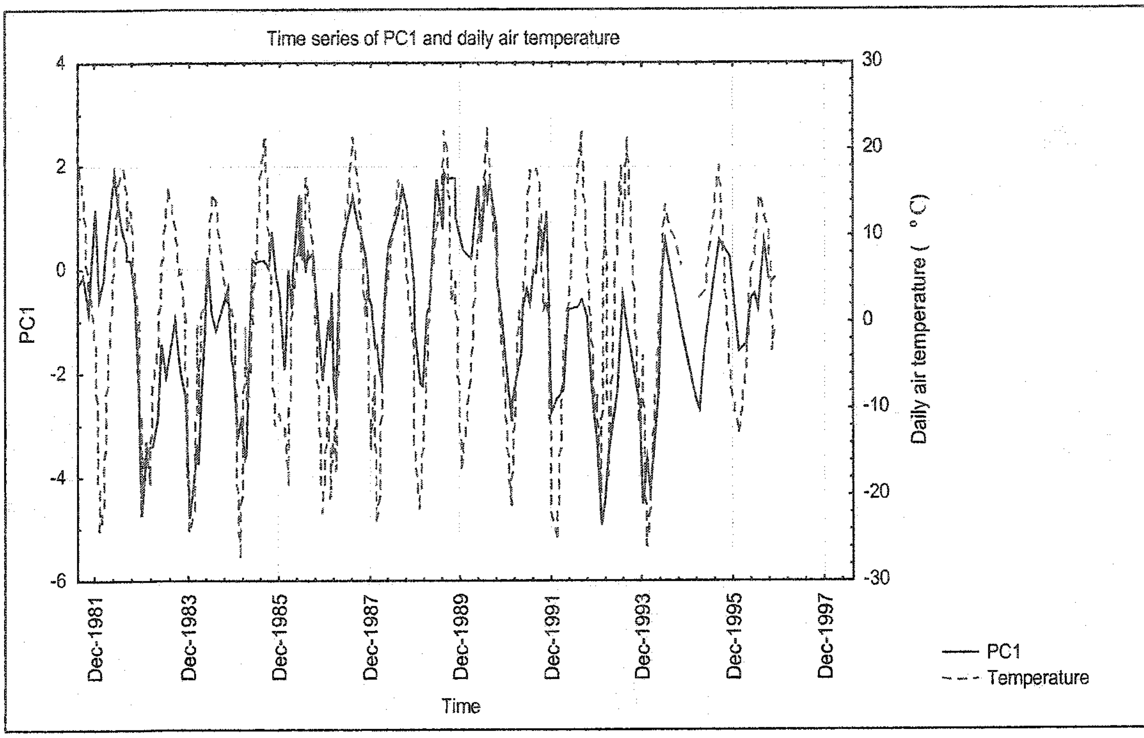


Figure 6.27, Scores of PC 1 and air temperature (1981-1997), Daniel Johnson dam

Following observations can be made regarding PC1: In the period from 1972 to 1979, scores of PC1 are monotonically decreasing. As the loadings are negative, a decrease of PC1 represents increasing displacements during this period. These irreversible displacements can be due to creep. From 1981 to 1995, cyclic trends are observed. Analyses of correlations indicate that these displacements are correlated with daily air temperature (Figure 6.27). PC2 explains the variance of pendulum S6y. There is no strong correlation between PC2 with reservoir level and air temperature. However, displacements of S6y are very small and in the range of 1-2 mm

Grouping instruments

In a large dam like Daniel Johnson, thousands of instruments are used to monitor the behavior of the dam. Traditionally instruments are reviewed and plotted individually. This is time-consuming and in case of faulty instruments can be confusing. In case of a faulty instrument, results will be the inconsistent with readings from similar instruments.

PCA can be used to group instruments that exhibit similar patterns of behavior. Principal components can be monitored instead of the original variables, which reduces the number of variables to monitor. If PC's behave normally there is no need to monitor the original variables. If a statistically significant change is detected in one of the PCs is detected, individual instruments associated with that PC can be reviewed. This can reduce significantly the time and expense of dam monitoring.

6.2.2.3 Analysis of extensometers

Readings of eight extensometers covering the period from 1982 to 1995 was considered for multivariate statistical analysis. These instruments monitor the oblique cracks of buttress 6 (Table 6.5).

Table 6.5, Specification of extensometers

Name(instrument)	**Block**	**Elevation (m)**
EXM1	16	259
EXM2	16	256
EXM3	16	259
EXM4	18	245
EXM5	18	257
EXM6	18	250
EXM7	18	263
EXM8	18	257

A box plot of extensometers data is presented in Figure 6.28. Data from extensometers is more erratic than data from other instruments (Figure 6.29 to Figure 6.37). The maximum crack opening for this time period is 0.9 mm. EXM1 readings are negative for the whole period.

Analyses of correlations between the instruments shows that instruments EXM1, EXM2, EXM5, EXM7, and EXM8 are highly correlated (Figure 6.37 and Table 6.6). These instruments all recorded strong fluctuations in 1986, and a step jump in crack opening in December 1993. EXM3 and EXM4 are also highly correlated (R=0.89).

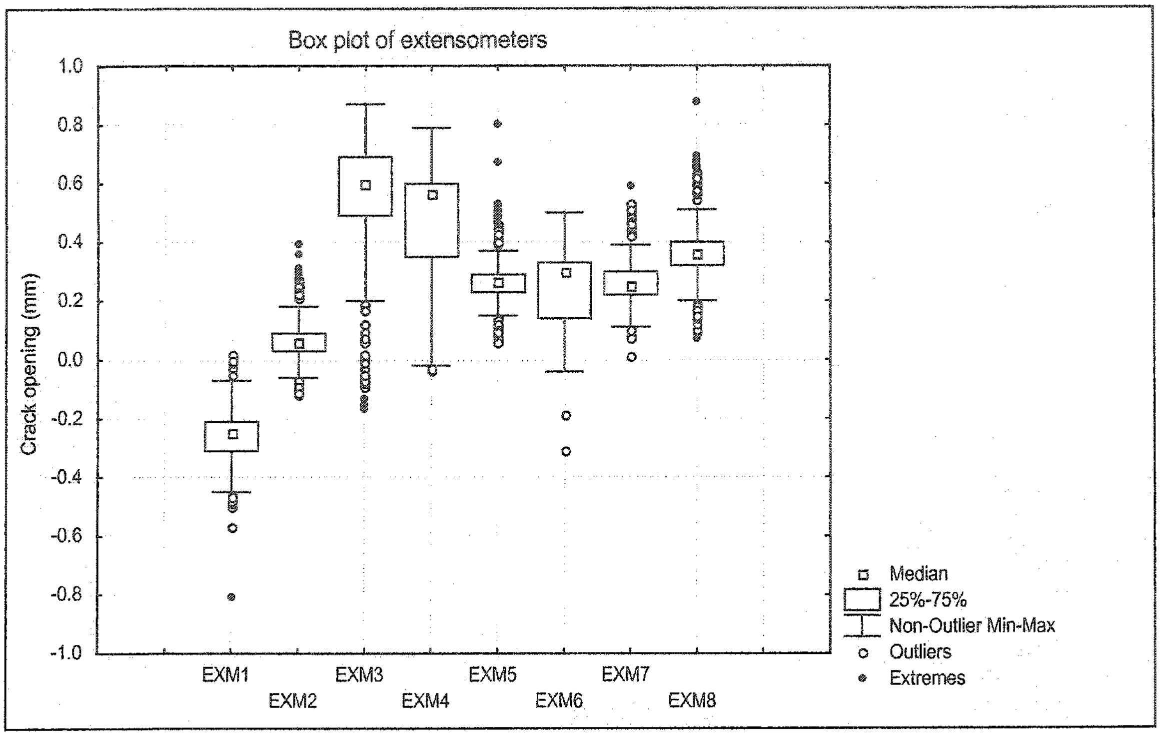


Figure 6.28, Box plot of extensometers, Daniel Johnson dam

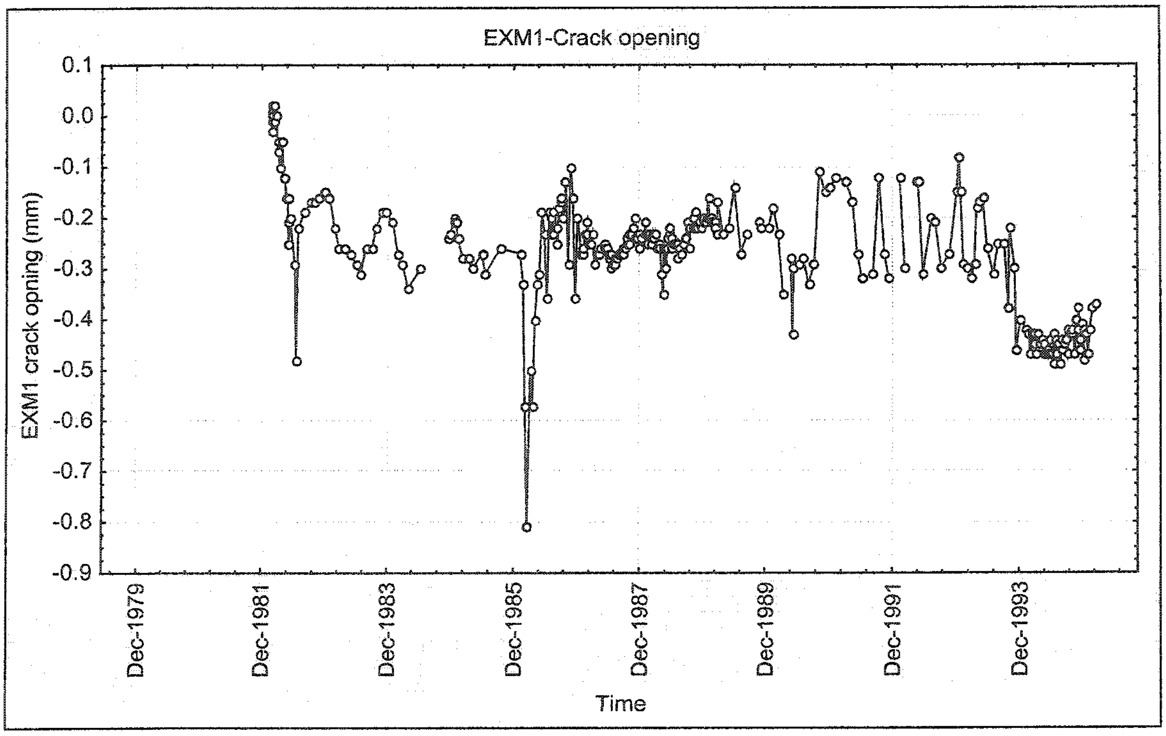


Figure 6.29, Crack opening, EXM1, Daniel Johnson dam

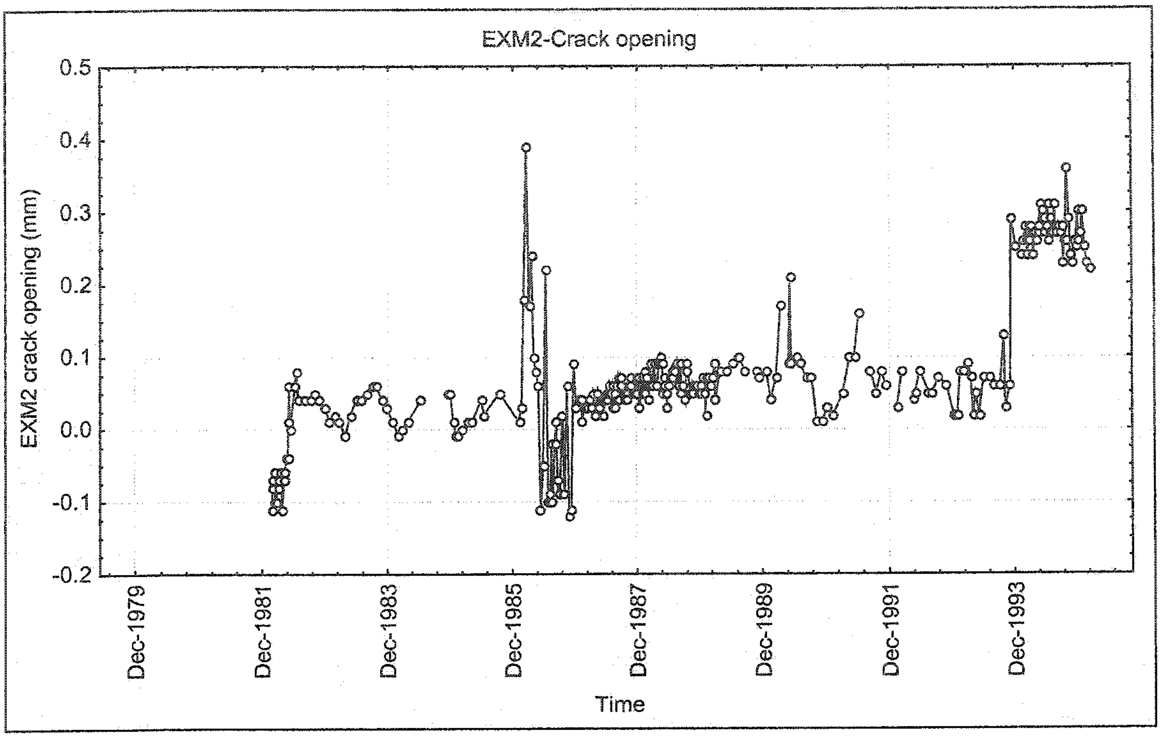


Figure 6.30, Crack opening, EXM2, Daniel Johnson dam

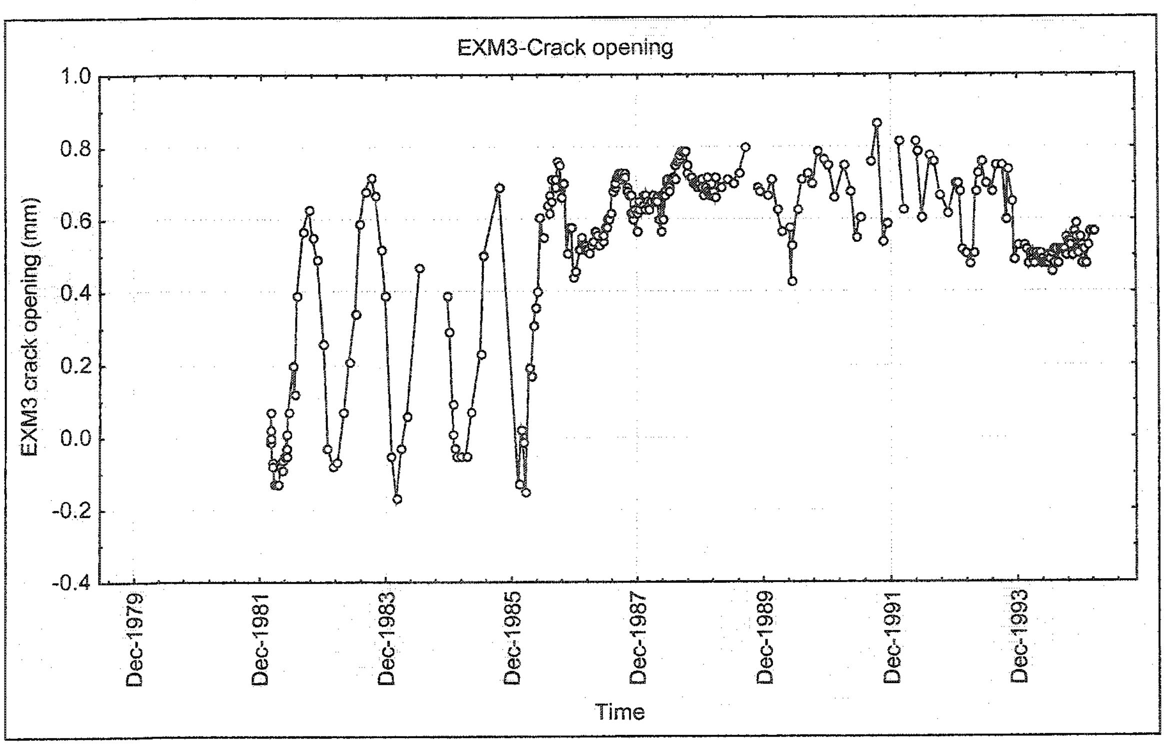


Figure 6.31, Crack opening, EXM3, Daniel Johnson dam

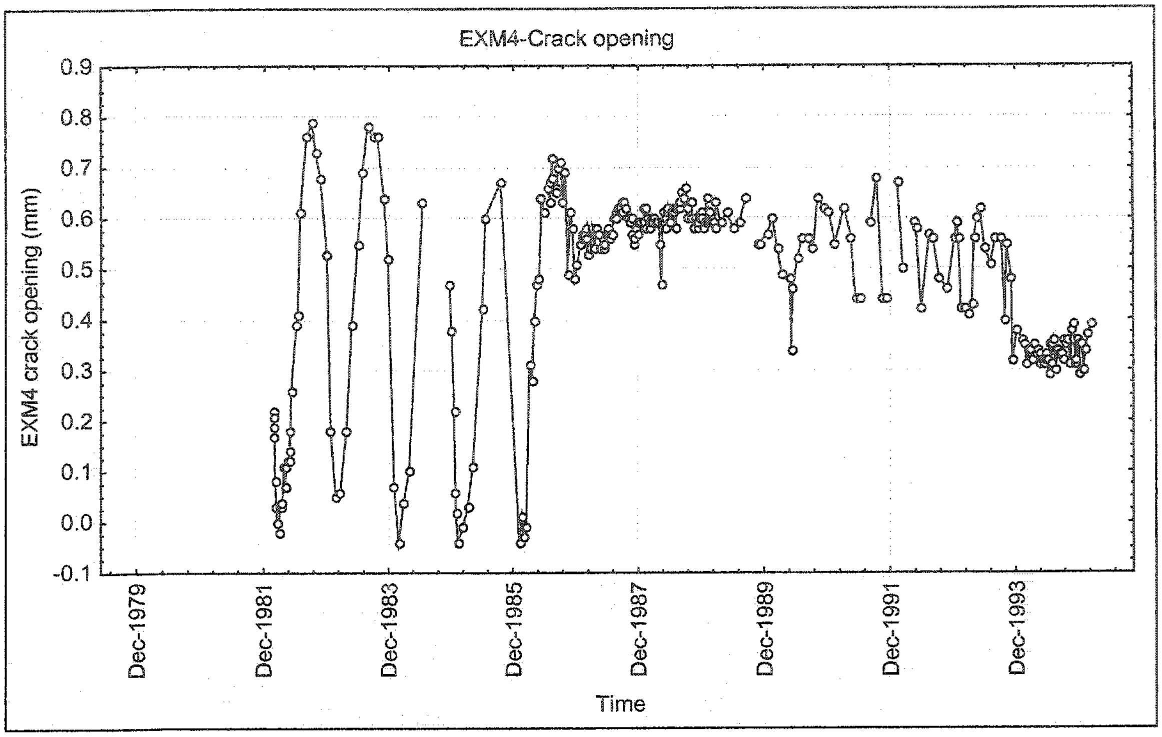


Figure 6.32, Crack opening, EXM4, Daniel Johnson dam

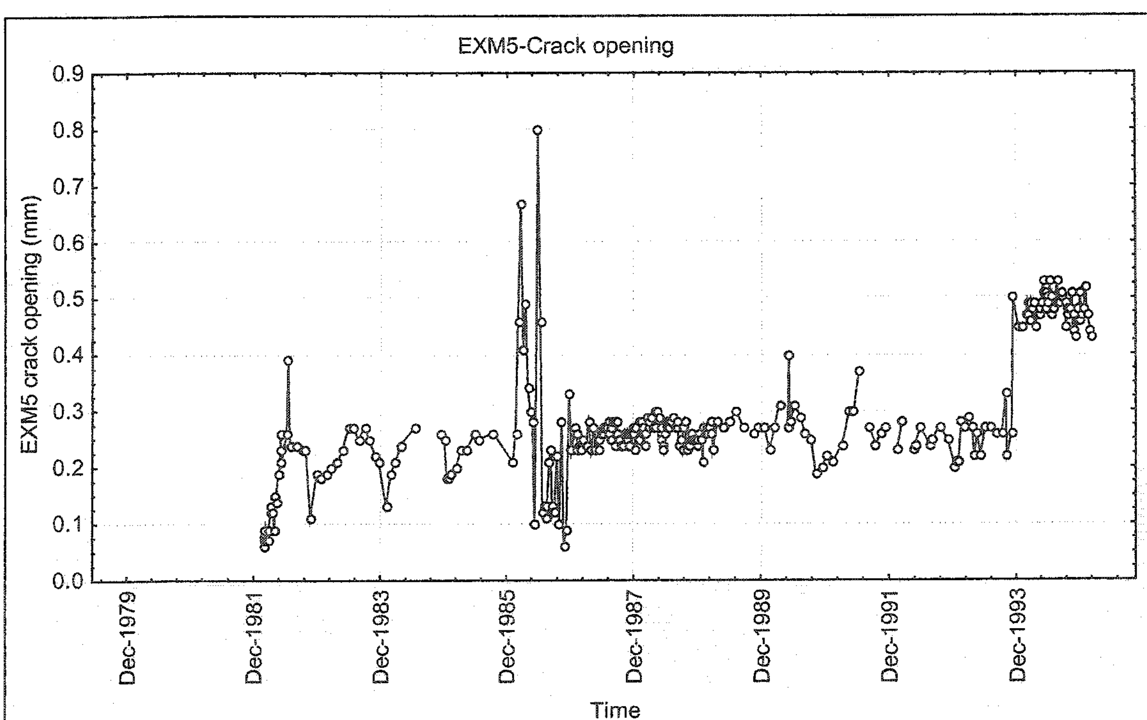


Figure 6.33, Crack opening, EXM5, Daniel Johnson dam

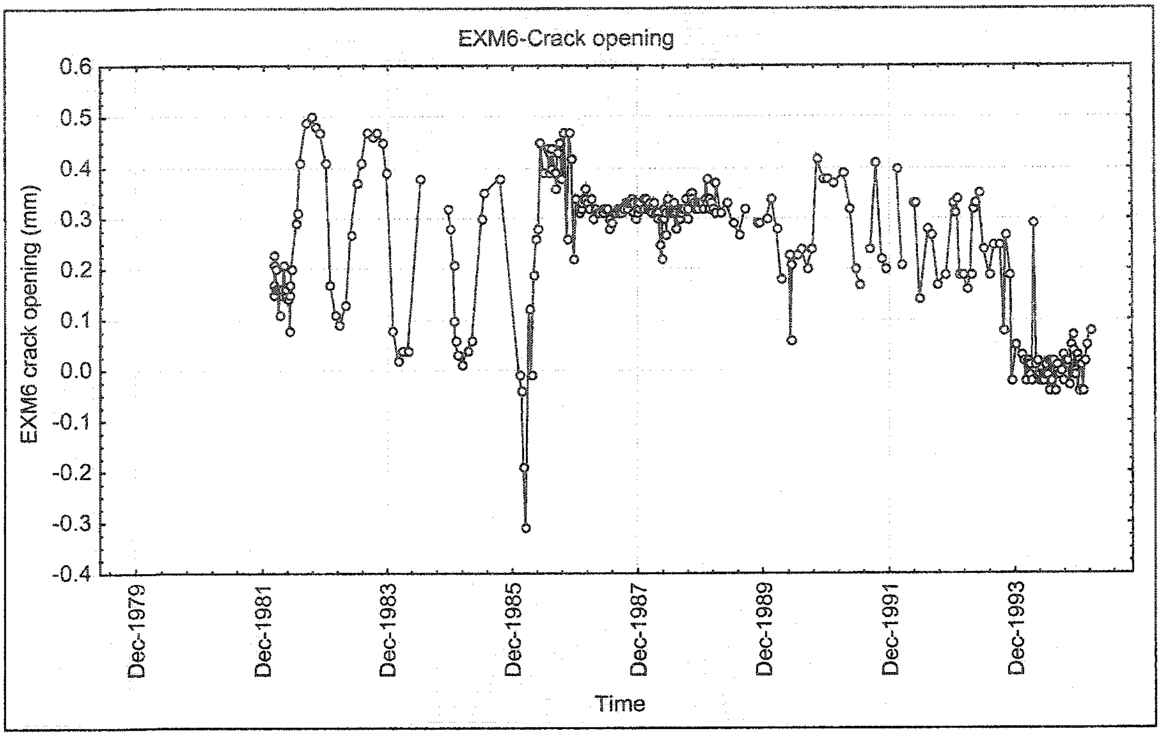


Figure 6.34, Crack opening, EXM6, Daniel Johnson dam

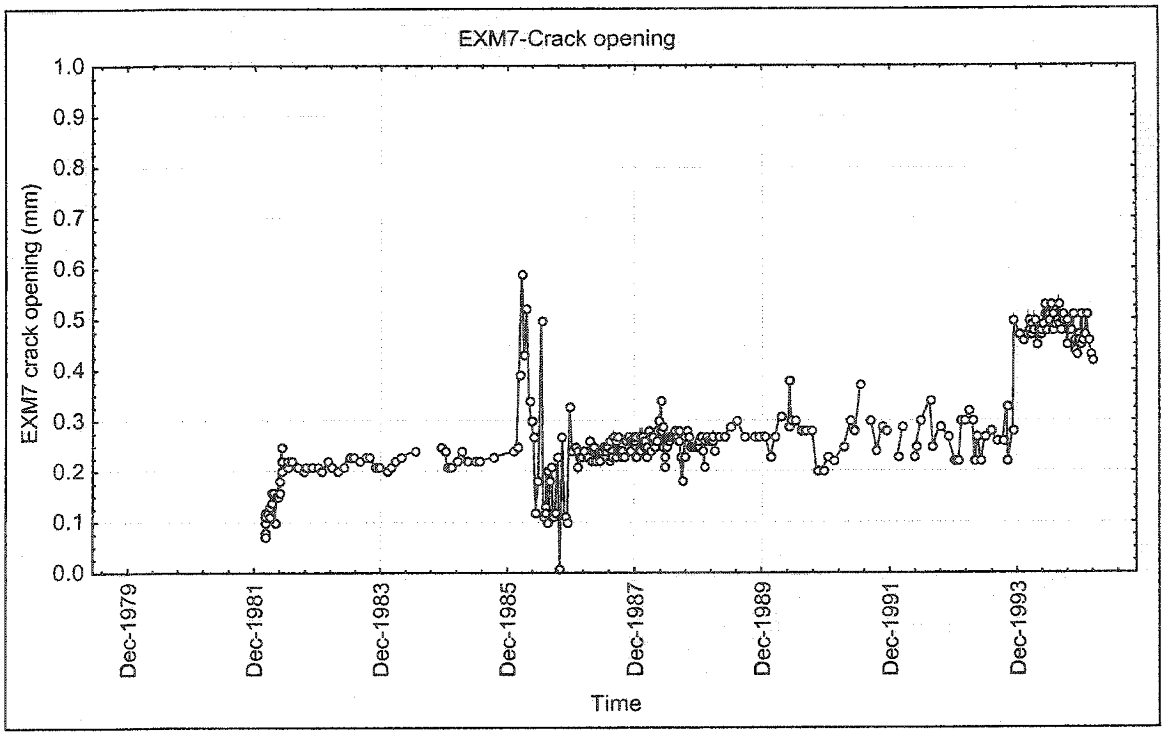


Figure 6.35, Crack opening, EXM7, Daniel Johnson dam

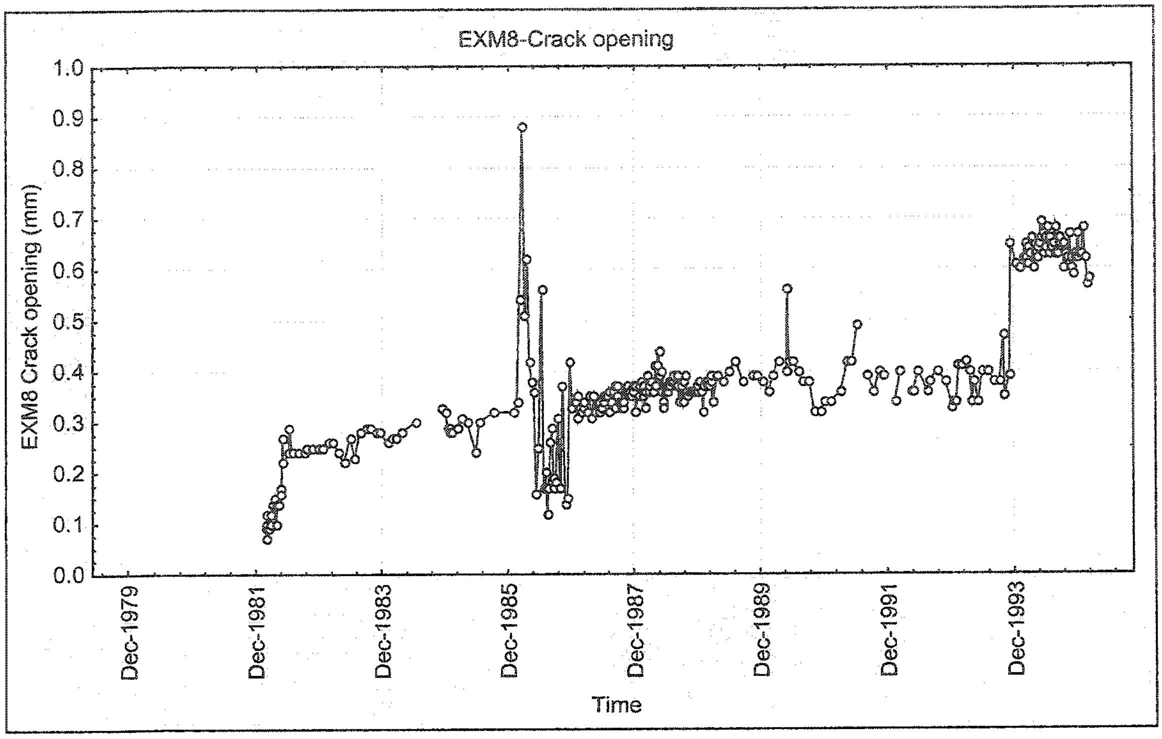


Figure 6.36, Crack opening, EXM8, Daniel Johnson dam

Table 6.6, Correlation matrix of extensometers, Daniel Johnson dam

	Correlations							
Variable	EXM1	EXM2	EXM3	EXM4	EXM5	EXM6	EXM7	EXM8
EXM1	1.00	**-0.87**	-0.02	0.22	**-0.88**	0.67	**-0.88**	**-0.87**
EXM2		1.00	0.14	-0.17	**0.93**	-0.67	**0.98**	**0.98**
EXM3			1.00	**0.89**	0.13	0.50	0.10	0.23
EXM4				**1.00**	-0.15	**0.80**	-0.22	-0.12
EXM5					**1.00**	-0.62	**0.93**	**0.92**
EXM6						1.00	-0.70	-0.65
EXM7							1.00	**0.97**
EXM8								1.00

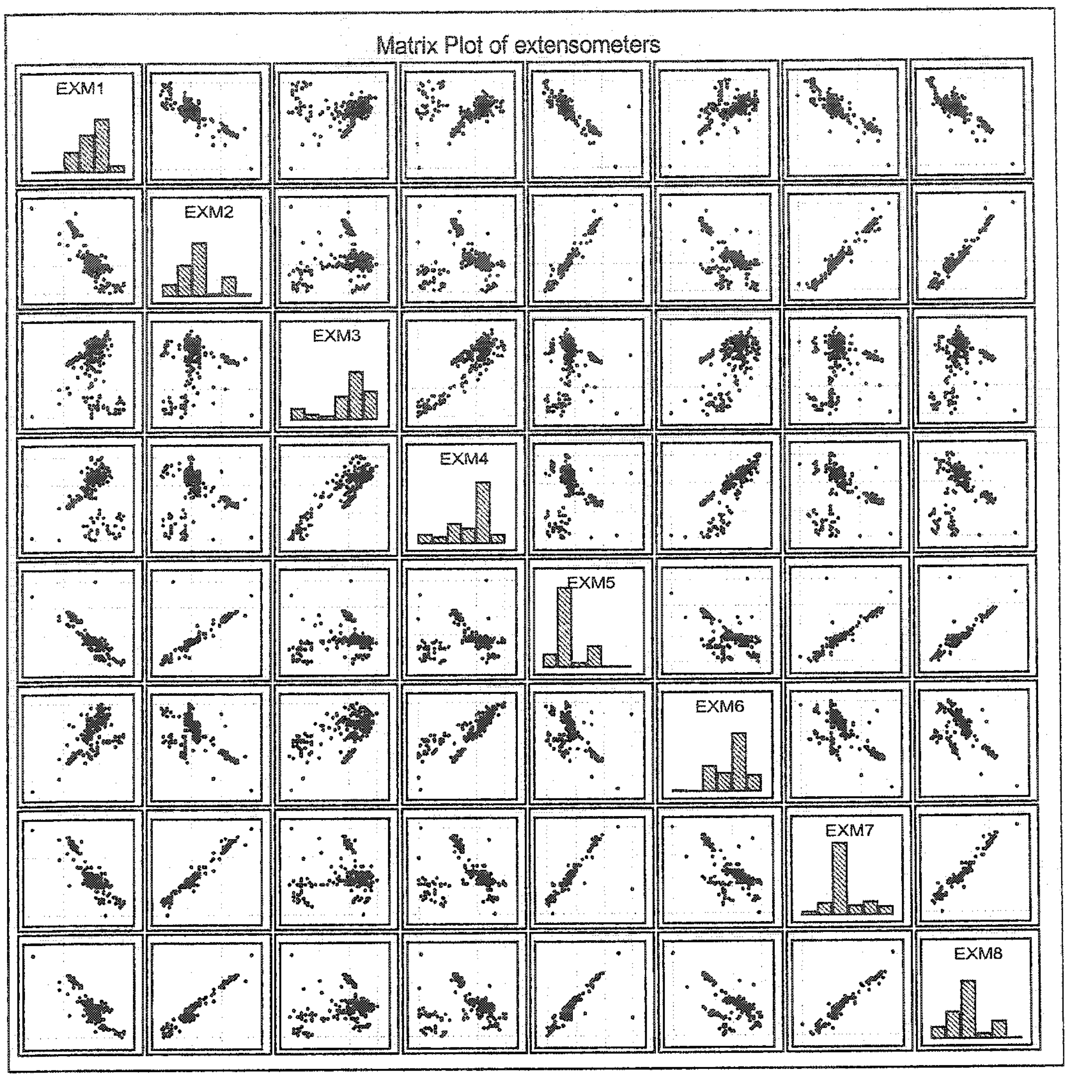


Figure 6.37, Matrix plot of extensometers, Daniel Johnson dam

Principal component analysis

PCA was performed on the correlation matrix of these eight instruments The first two components explain 94% of the total variance (Figure 6.38) of the eight original variables; which is a significant reduction with respect to number of original variables. It can be concluded that PCA can efficiently compress the original data and reduce number of monitored variables. The first principal component and the second component explain 64% and 30% of the total variance of the data set respectively (Figure 6.38). The remaining components explain only 6% of the variance and, most likely, they represent the noise in the data.

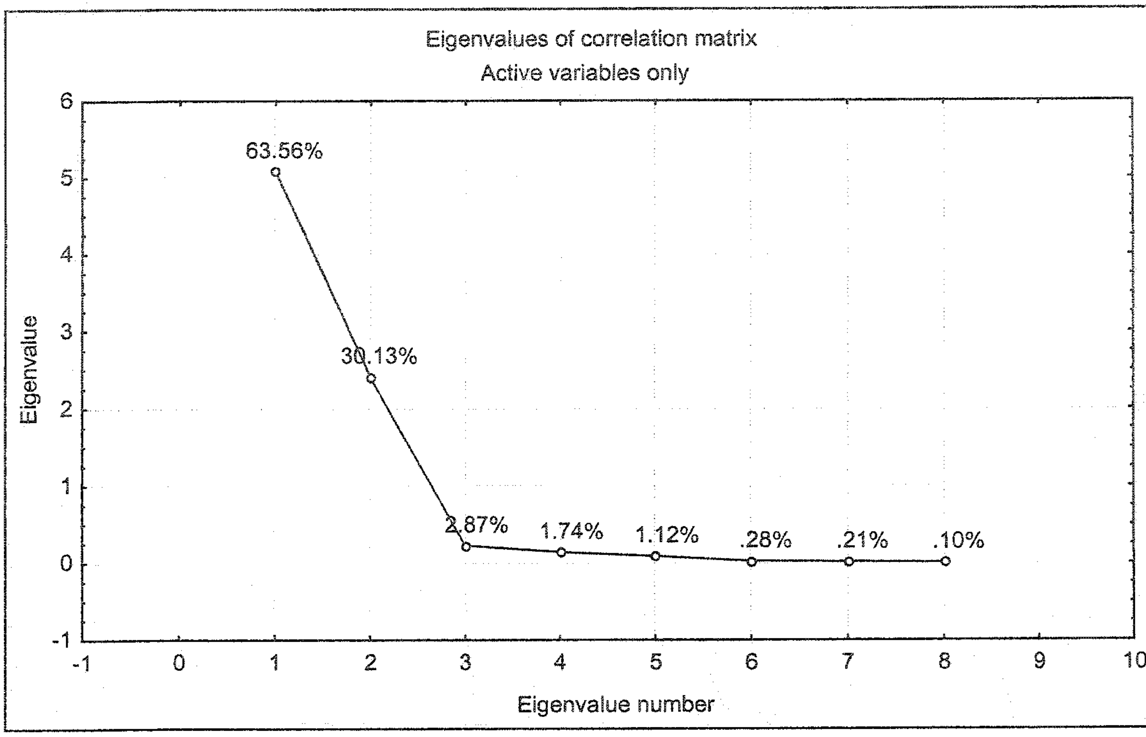


Figure 6.38, Scree plot, correlation matrix of extensometers, Daniel Johnson dam

Rotation of Principal Components

In order to improve the interpretation of the principal components, it is often desirable to rotate a subset of the initial eigenvectors to a second set of new vectors. As a consequence of rotations of the PC's, a second set of scores is computed that are called rotated principal components.

Various rotational strategies have been proposed in the literature. The goal of all of these strategies is to obtain a clear pattern of loadings. The most widely used method is the varimax approach, which rotates the loadings so that the variance of the squared loadings in each component is maximized. The squared loadings in each component are nudged toward 0 and 1 (Rencher, 1998), which aids in the assignments of instruments to PC's, and subsequent grouping of the instruments. This is illustrated using the extensometers data of Daniel Johnson dam.

The varimax criterion was used to rotate the PCs and group the extensometers data. Two groups are observed corresponding to PC1 and PC2. Group 1 comprises instruments EXM1, EXM2, EXM5, EXM7 and EXM8, which are highly correlated with PC1. It must be noted that EXM1 loading for PC1 in negative which indicates high negative correlation between EXM1 and the rest of the group (Figure 6.39). Group 2 comprises instruments EXM3 and EXM4, which are highly correlated with PC2.

As previously mentioned, Daniel Johnson has undergone major rehabilitations (thermal shelters in 1986) and several grouting campaigns. Some of these campaigns have improved the dam section and some have deteriorated the dam section due to excessive high pressures and inaccurate methods. Therefore analyses of correlations of crack openings are hard to perform and interpret. Since the dam has undergone several changes due to cracking and grouting.

Different behaviors are observed for PC1 and PC2. PC1, which represents instruments EXM2, EXM5, EXM7 and EXM8, captures strong fluctuations in crack opening in 1986. There is also a step jump in crack opening after December 1992, which is consistent with reports of increase in seepage in arch 5-6 (Larivière et al, 1999). The amplitude of fluctuations of PC2 is decreased after 1986, which indicates that the thermal shelters, have been effective in reducing crack openings. PC2 shows seasonal behavior before 1986, with a lag of about one month relative to daily air temperature (Figure 6.42 and Figure 6.43). Propagation of cracks is observed in PC2 after 1993. Analyses of correlations between PC1 and reservoir level and ambient temperature are hard to interpret due to continuous changes in crack characteristics due to various interventions in the dam during that period.

6.2.2.4 Analysis of water infiltration measurements

A preliminary analysis was done on simultaneous readings from pendulums and a flow meter for arch 5-6. The number of simultaneous readings is much smaller than the number of readings for either the flow meter or pendulums. Correlations between water infiltration in arch 5-6 and displacements of the pendulums, is maximum with pendulums S1 and S2, which are closest to the location of cracks. The cracks are located at elevations of 253 and 259 m, and propagate from the upstream face to the downstream face.

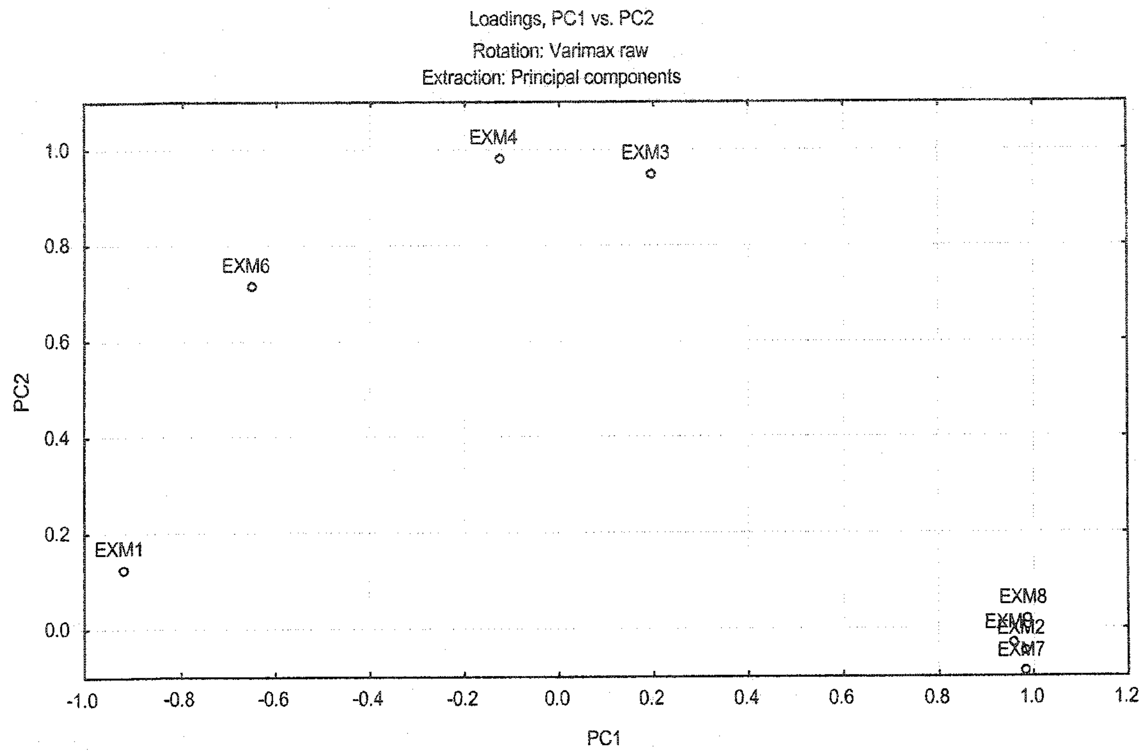


Figure 6.39, Loadings of PC1 and PC2, Daniel Johnson dam

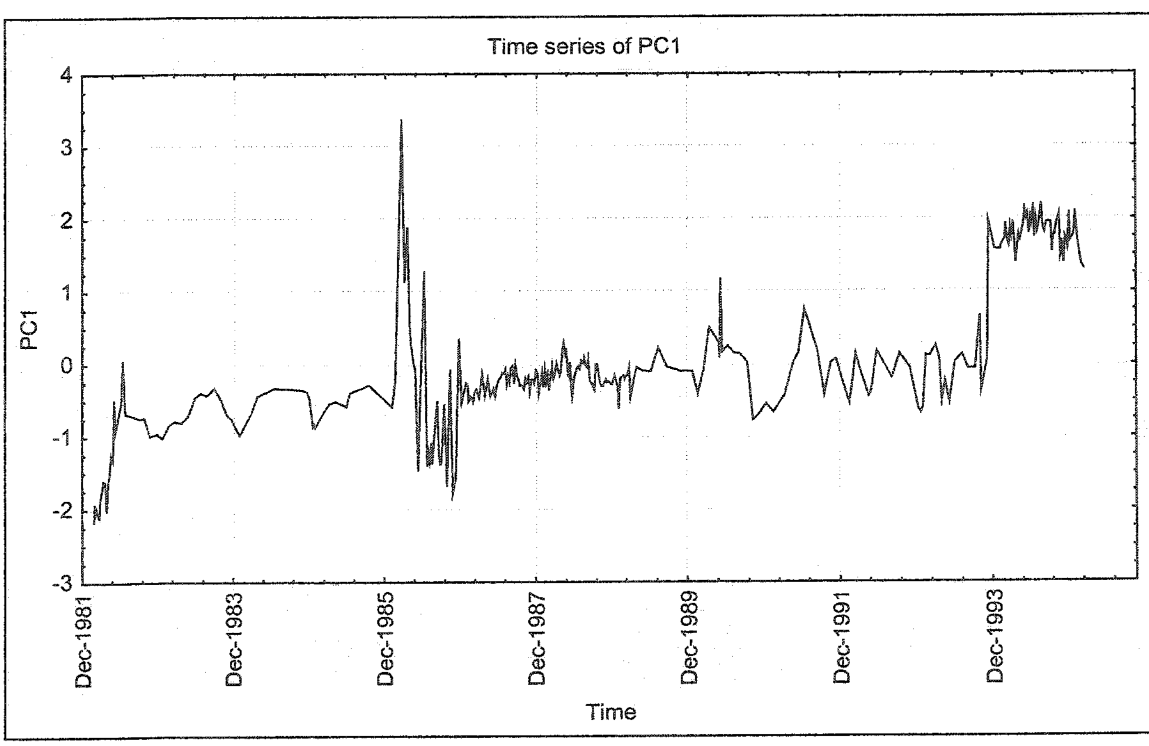


Figure 6.40, Scores of PC 1, extensometers, Daniel Johnson dam

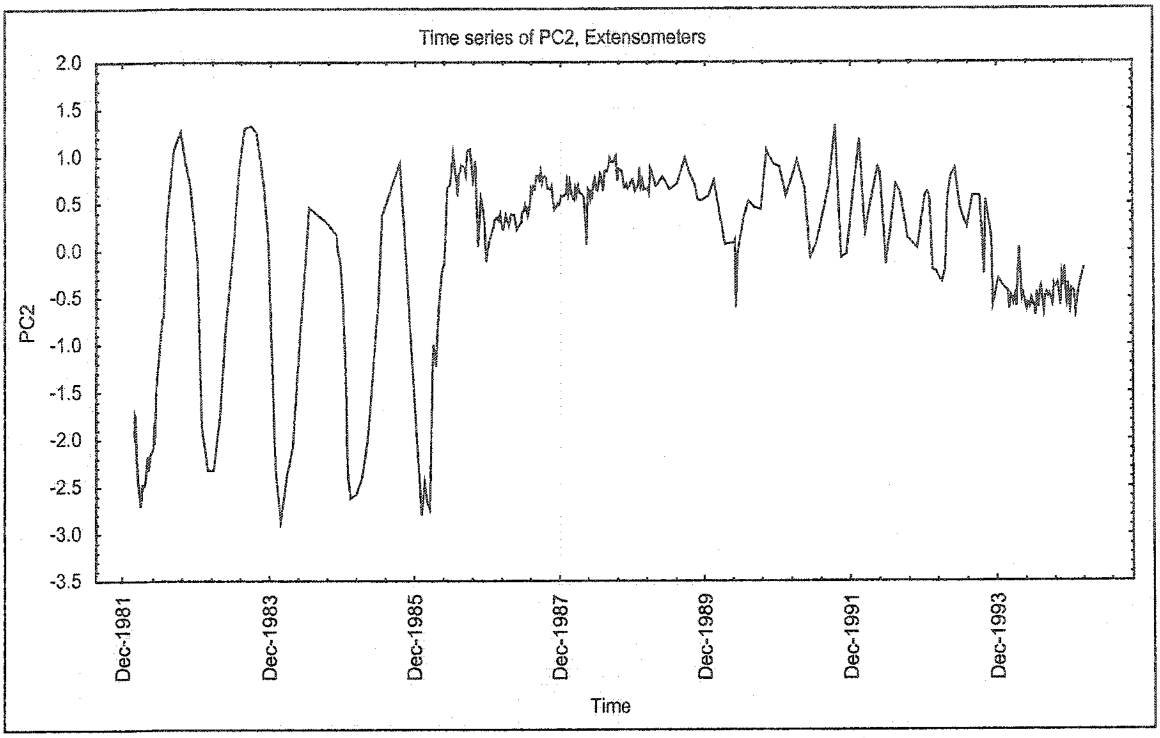


Figure 6.41, Scores of PC2, extensometers, Daniel Johnson dam

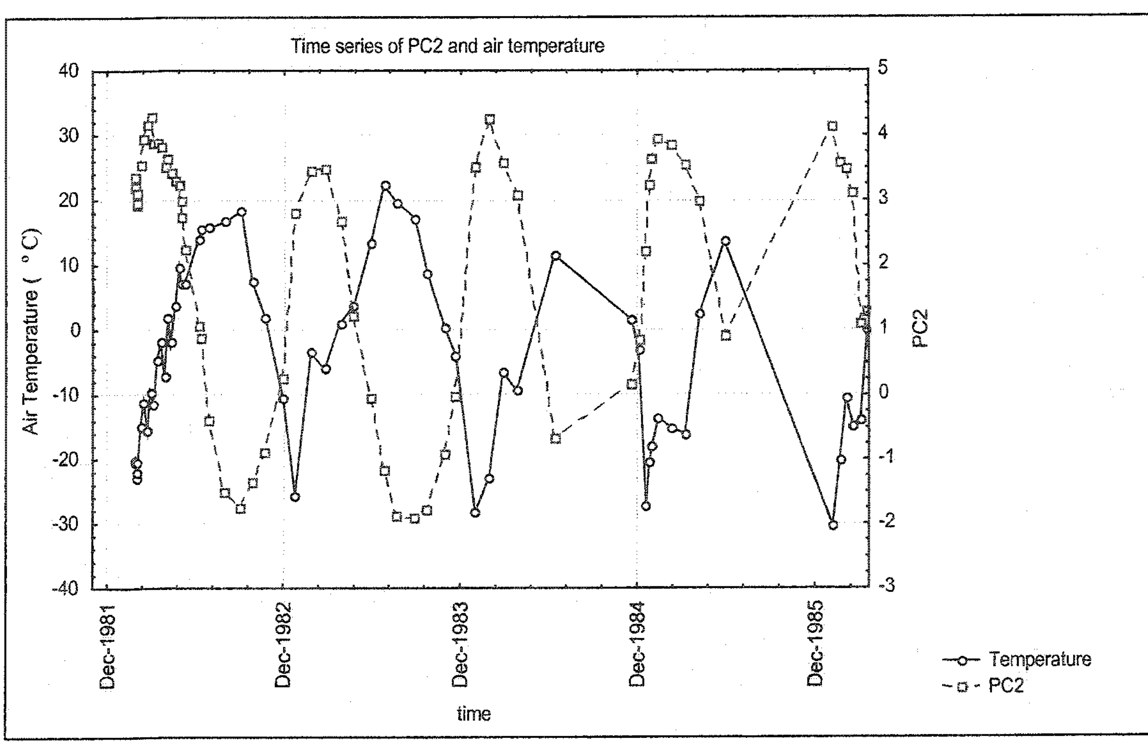


Figure 6.42, Partial plot of PC 2 and air temperature (1982-1986), Daniel Johnson dam

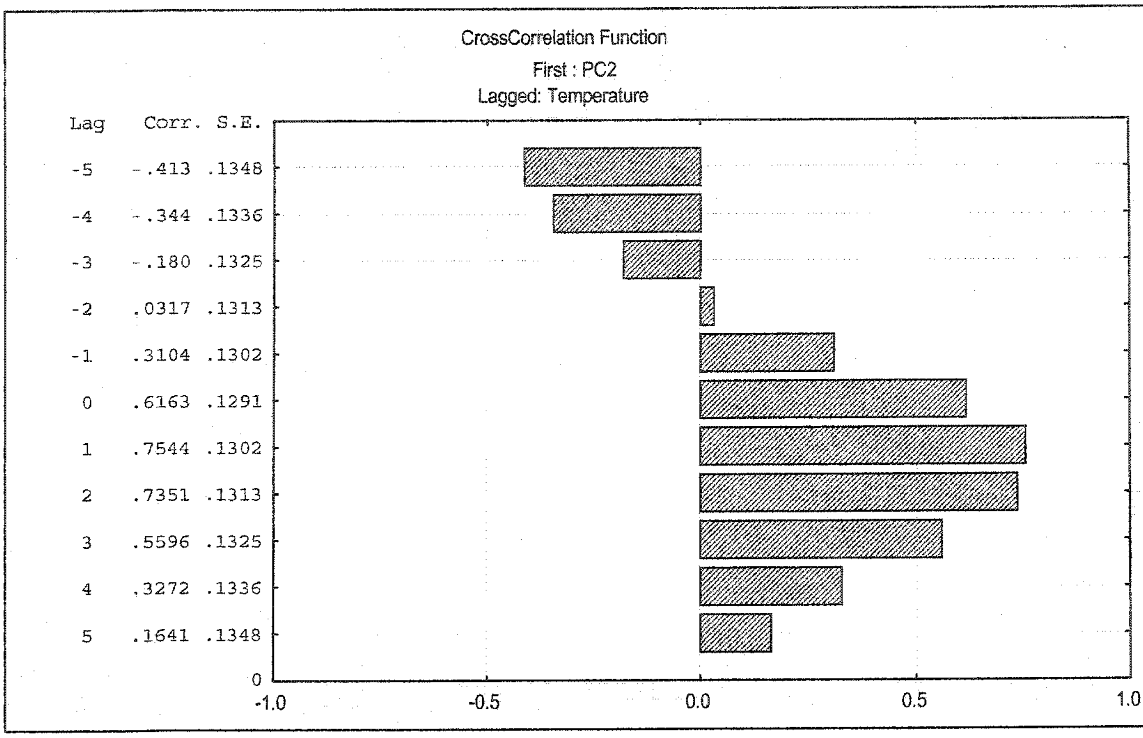


Figure 6.43, Cross correlation of PC2 and air temperature (before 1986), Daniel Johnson dam

Pendulum S1 is located at elevation 244 m. Pendulum S1 and water infiltration were considered for a more detailed analysis. Readings of both instruments that were recorded in the same week were considered for analysis. Figure 6.44 shows respectively the displacement of pendulum S1 in Y direction along with the water infiltration in Arch 5-6. There is a high correlation between measured flows and pendulum displacement. After January 1993, the trend of displacements is different from that of water flows. While water flow is increasing, pendulum displacements seem to stabilize and do not show much variation. Figure 6.45 presents scatter plot of water flow and pendulum S1 displacements. Obviously the trend has changed after 1993. It can be concluded that propagation of cracking after 1993 was due to grout erosion, which increased water flow in the Arch 5-6, does not affect the stiffness of the dam.

Table 6.7, Correlation, pendulum readings and water infiltration (Arch 5-6)

Correlation Matrix								
	S1	S2	S3	S4	S5	S7	S8	Flow
S1	1.00							
S2	0.82	1.00						
S3	0.44	0.40	1.00					
S4	0.18	0.20	0.93	1.00				
S5	0.02	0.06	0.83	0.96	1.00			
S7	0.26	0.30	0.72	0.70	0.60	1.00		
S8	0.20	0.25	0.81	0.84	0.77	0.96	1.00	
Water infiltration	**0.59**	**0.58**	0.41	0.25	0.10	0.41	0.34	1.00

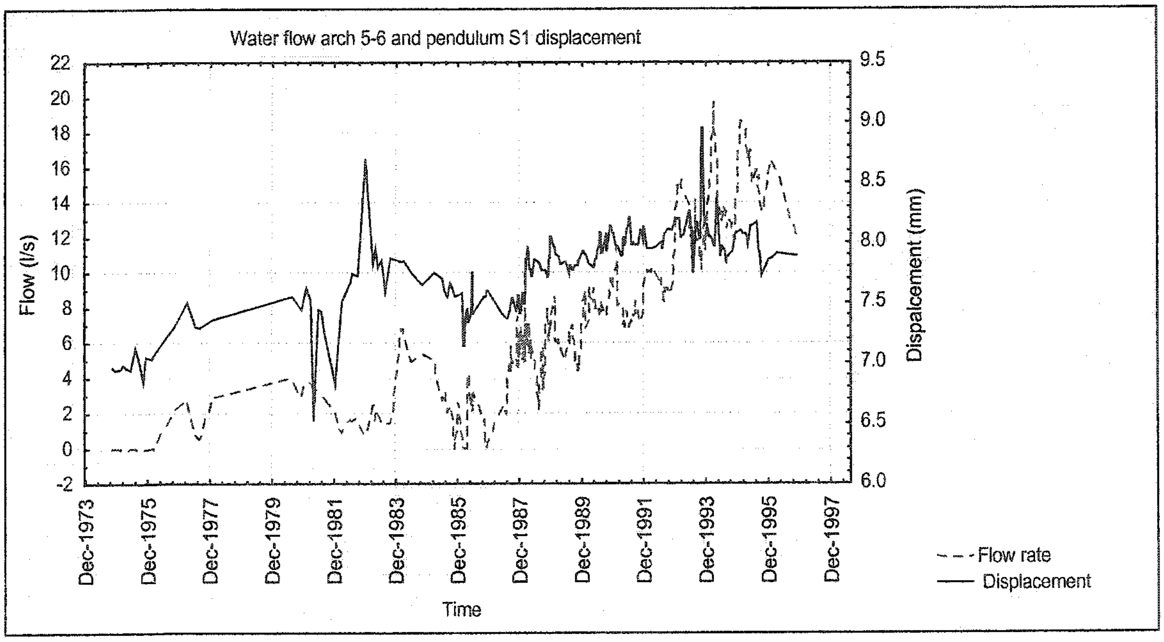


Figure 6.44, Time series of water infiltration and displacement pendulum S1, Daniel Johnson dam

6.3 Chute-à-Caron gravity dam

Chute-à-Caron is a gravity dam that comprises three sections. Two of the segments form an angle of 147 ° (

Figure 6.46). These two segments interact due to thermal expansion and have opened a construction joint, which is located in the drainage gallery.

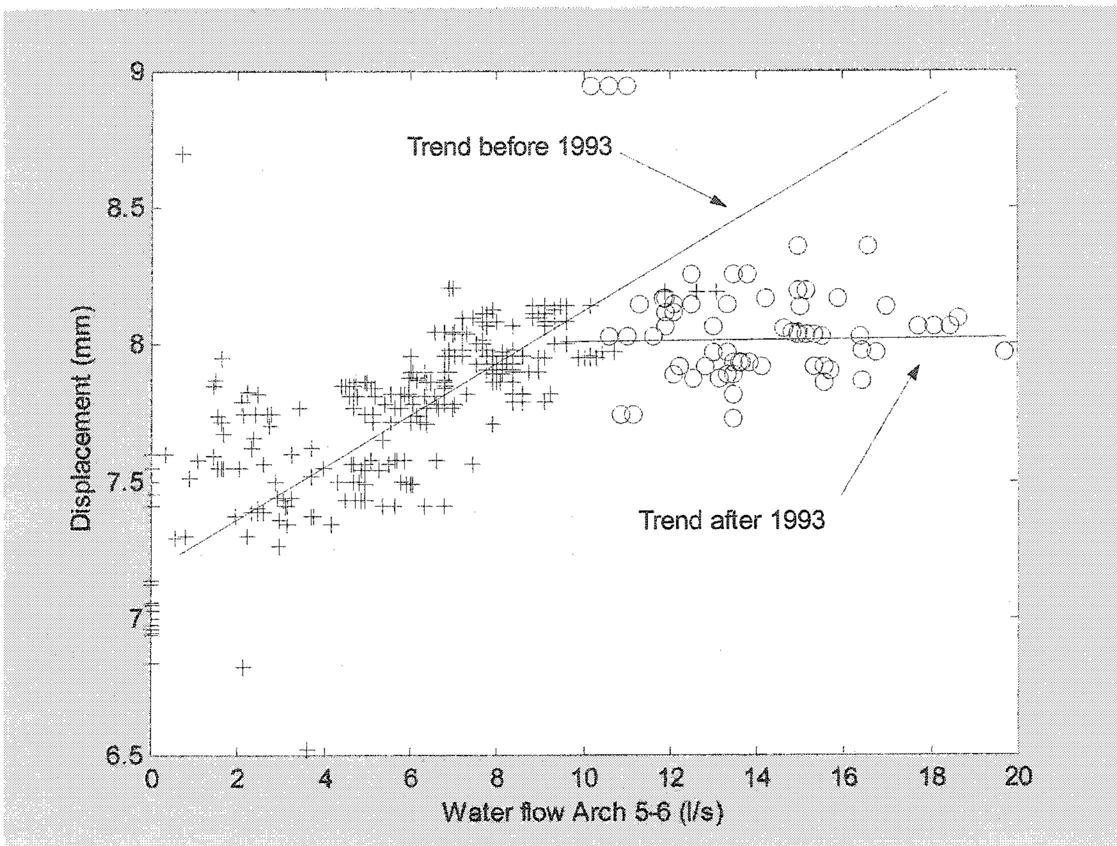


Figure 6.45, Scatter plot of Pendulum S1 displacement and water infiltration Arch 5-6, Daniel Johnson dam

This joint was grouted several times starting in 1991 when water infiltration exceeded 300 l/min. Water flows reduced to 50 l/min but increased to 750 l/min in the summer of 1996. A 15.5 mm wide vertical expansion joint, over 80% of the height of the dam, was cut near the intersection of the two segments in June 1997 using a diamond cable (Figure 6.47). The behavior of the dam is being monitored with inverted pendulums, flow meters, and the joint meters, which measure movement of the construction joint (Table 6.8).

6.3.1 Statistical Analysis of Chute-à-Caron monitoring data

Some instruments, which are located in blocks 12S, 13S and 14S, are considered for statistical analysis. The description of the instruments is summarized in Table 6.8. Thermometers TE14S1, TE14S2, and TE14S3 measure the water temperatures (Figure

6.49), while thermometers TE14S4, TE14S5, TE12S1, TE12S2, and TE12S3 measure the concrete temperatures in the dam. Water temperatures vary between 0 ° C and 22 ° C, while air temperatures vary between -27 ° C and 27 ° C. During winter months (i.e. December through March) water temperature remains at 0 ° C.

The downstream face of the dam is colder than the upstream face in the winter and warmer in the summer. Therefore it is expected that temperature effects cause the dam to move towards downstream in the winter and upstream in the summer.

Table 6.9 presents descriptive statistics on measured temperatures from 1998 to 2000. Concrete temperatures are less variable than air temperatures. The maximum concrete temperature was recorded at TE12S1, which is located near the dam surface, and is higher than maximum air temperature. This can be explained by the effect of solar radiation. Thermometers located deeper inside the dam show less variability. Mean concrete temperatures are higher than the mean air temperature due to solar radiation and exposure of the upstream side to the reservoir, which never goes below 0° C.

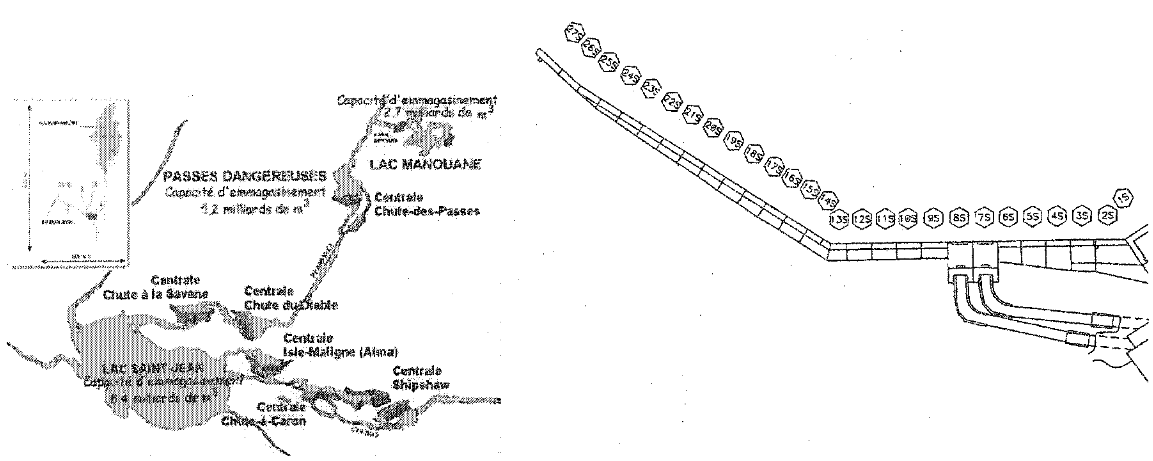


Figure 6.46, Location and plan, Chute -à- Caron gravity dam

Table 6.8, List of instruments, Chute -à- Caron dam

Instrument	Plot 13S, Chute-à-Caron
Inverted pendulums	PD13S-1, PD13S-2
Flow meters	DV13S-1, DV14S-1
Joint meters	FI13S-1, FI13S-2, FI13S-3
Thermometers (water)	TE14S-1, TE14S-2, TE14S-3,
Thermometers (concrete)	TE14S-4, TE14S-5, TE12S-1, TE12S-2, TE12S-3

Inverted pendulums measure the displacements of block 13 in three directions: (1) right-left (X), (2) upstream-downstream (Y), and (3) vertical (Z) (Figure 6.52 and Figure 6.53). It must be noted that because of the position of block 13S in the intersection of two angled segments, upstream-downstream movements of block 13 are related to both upstream-downstream and right-left movements of block 14. Right-left displacements can be used to monitor opening and closing of vertical expansion joint. Joint meters FI13S1, FI13S2, and FI13S3 are located at the ends of block 13 and joint meter FI13S2 is located at the center of the block 13 (Figure 6.48).

Correlations between instruments are presented in Table 6.10. Vertical displacements of the two pendulums are highly correlated (R=0.99). Right-left displacements of pendulums are also highly correlated (R=-0.98). Right left displacements of both pendulums are correlated with displacements of joint FI13S1 and concrete temperatures, TE14S4 and TE12S2. Upstream-downstream displacements of PD13S2 are correlated with concrete temperatures (TE12S1), near the dam surface. Vertical displacements of both pendulums show strong correlation with joint meters FI13S1 and thermometers TE14S5 and TE12S3. Measurements from the two flow meters are correlated (R=0.71), but these two instruments are not strongly correlated with the other instruments. One explanation is that flows usually have a very non-linear response (relative to other instruments) since they are highly dependent on closure and opening of joints.

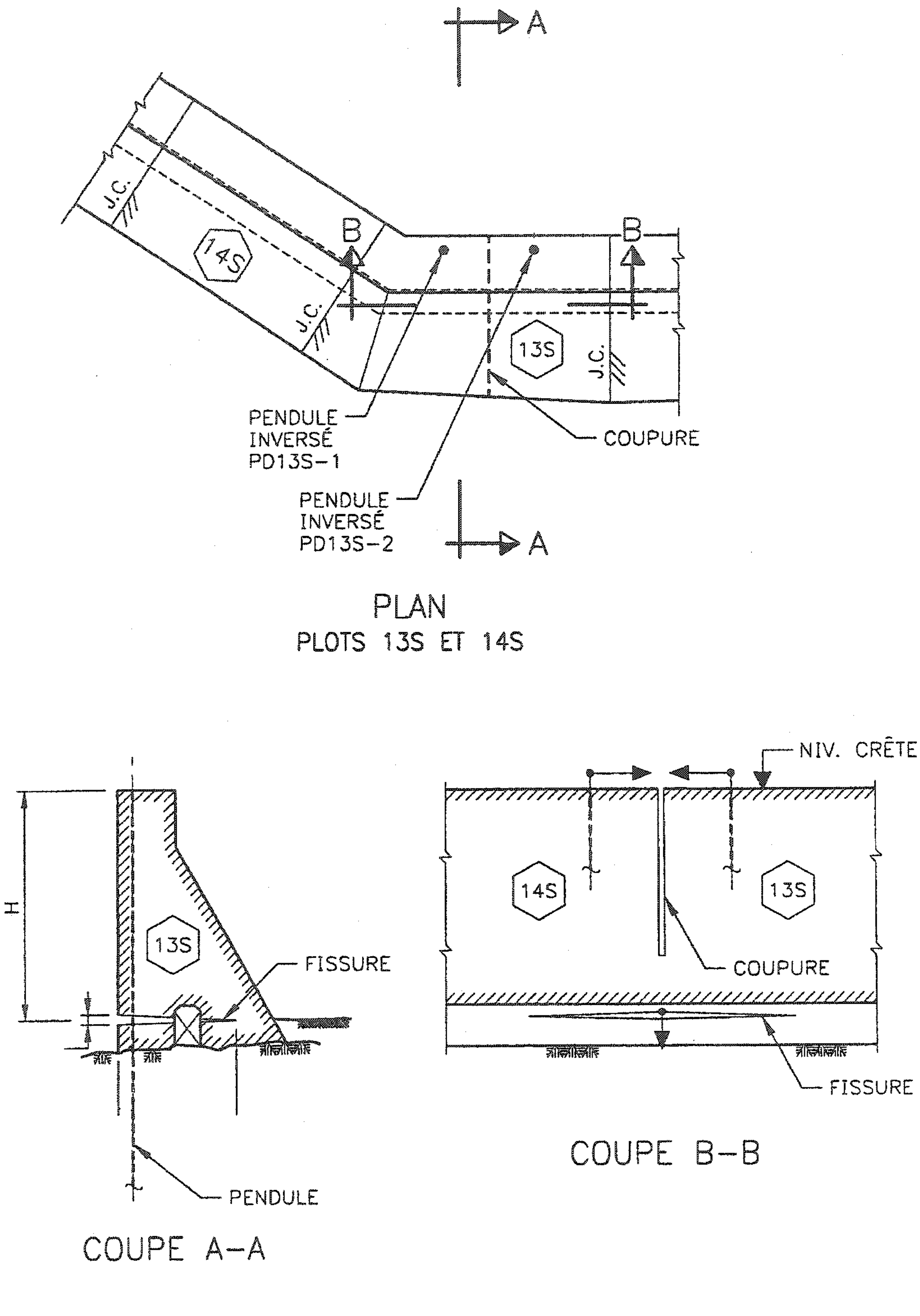


Figure 6.47, Details of Block 13, Chute-à-Caron dam

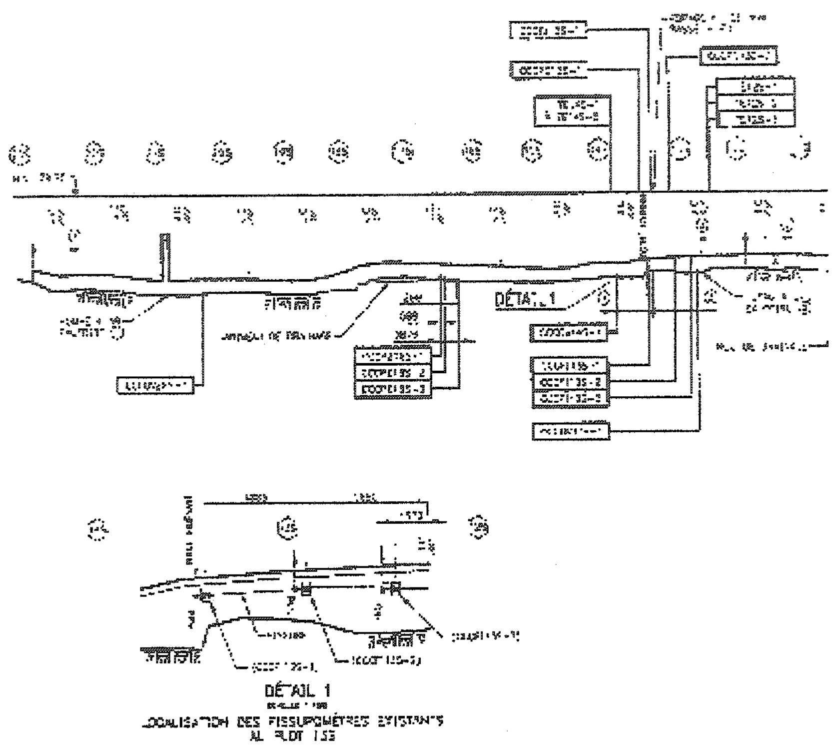


Figure 6.48, Block 13 instrumentation layout, Chute-à-Caron dam

Table 6.9, Statistics for air, water, and concrete temperatures

Variable	Descriptive Statistics Valid N	Mean	Minimum	Maximum	Std.Dev.
TEMPERATURE	1460.0	4.6	-27.3	26.7	12.0
TE14S1	1284.0	7.8	0.2	21.7	7.8
TE14S2	1281.0	7.8	0.2	21.7	7.8
TE14S3	1282.0	7.8	0.2	22.0	7.8
TE14S4	1280.0	6.9	-12.8	23.4	11.0
TE14S5	1283.0	8.4	-5.5	23.6	9.2
TE12S1	1276.0	10.0	-15.7	31.9	11.4
TE12S2	1279.0	10.4	-5.7	26.1	9.2
TE12S3	1283.0	10.6	0.7	22.4	7.3

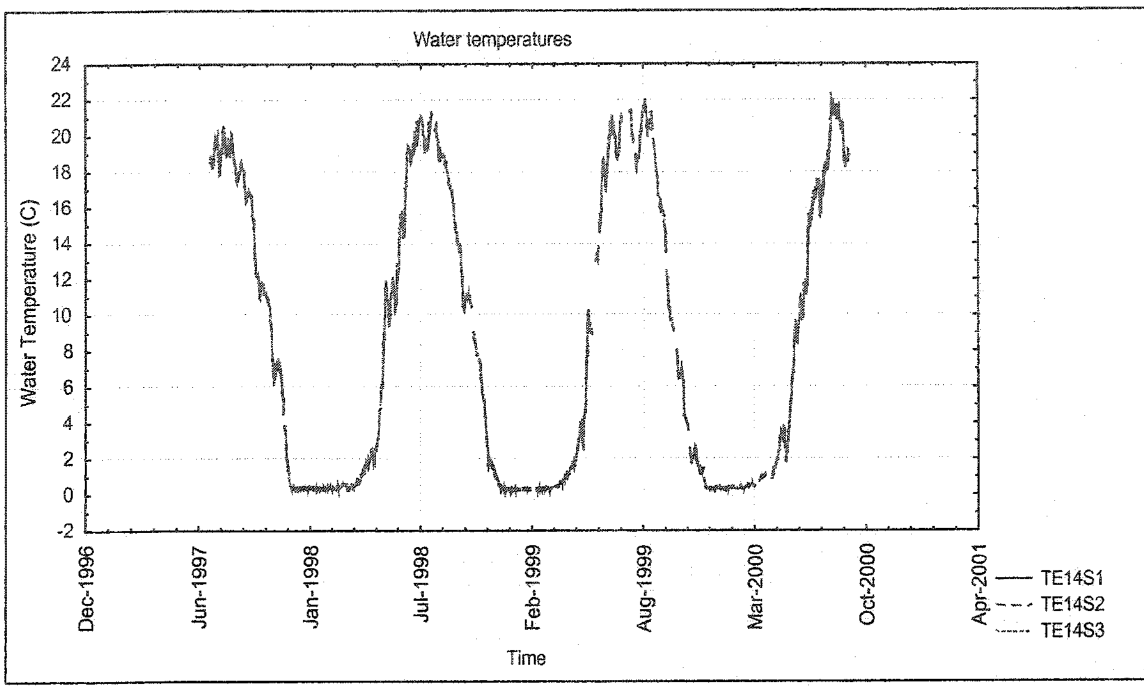


Figure 6.49, Water temperatures, Chute-à-Caron dam

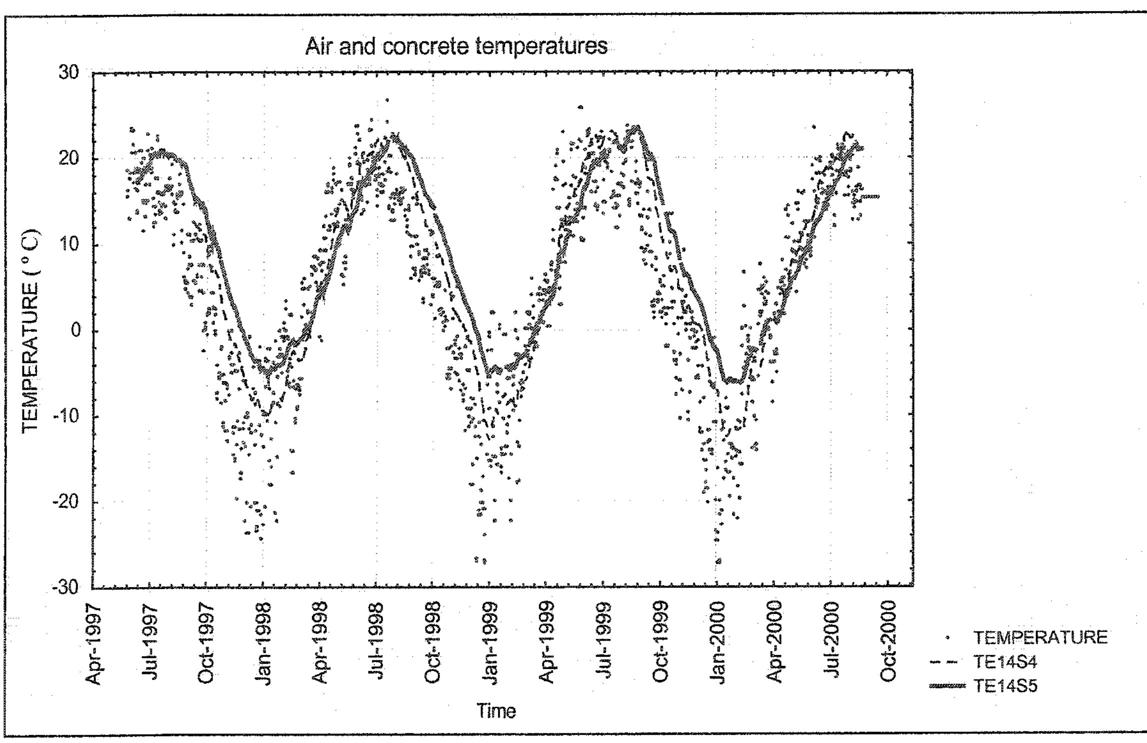


Figure 6.50, Air and concrete temperatures, Chute-à-Caron dam

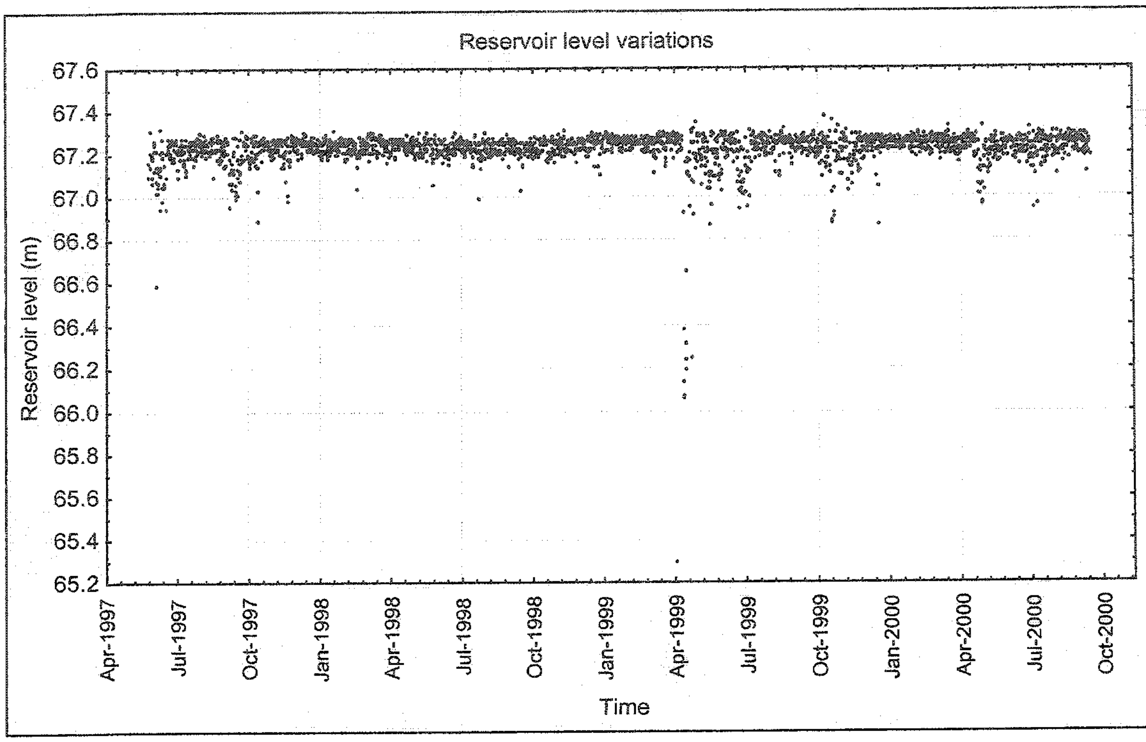


Figure 6.51, Reservoir level variation, Chute-à-Caron dam

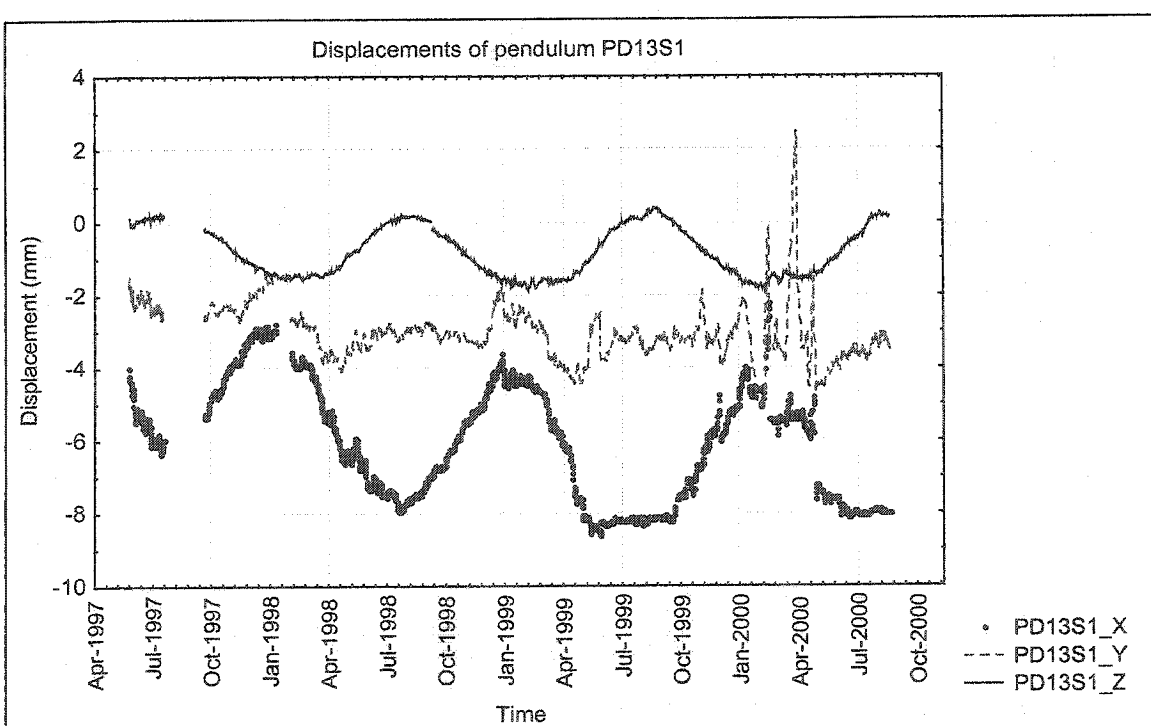


Figure 6.52, Displacements of Pendulum PD13S1, Chute-à-Caron dam

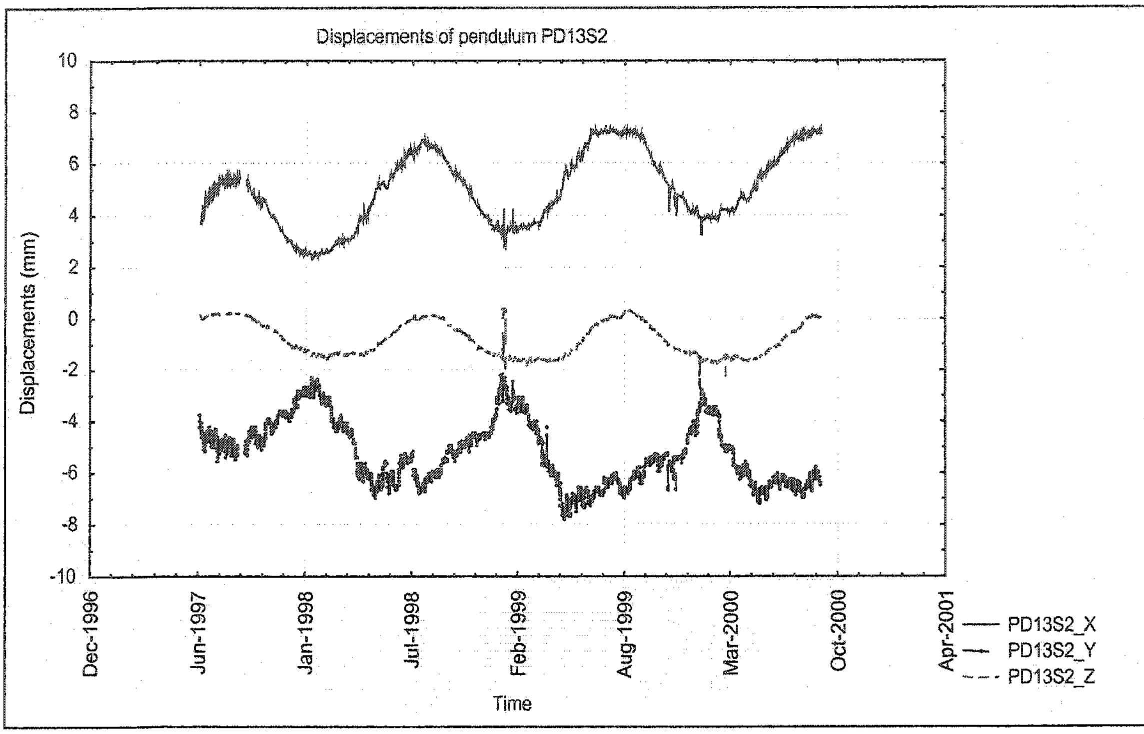


Figure 6.53, Displacement of Pendulum PD13S2, Chute-à-Caron dam

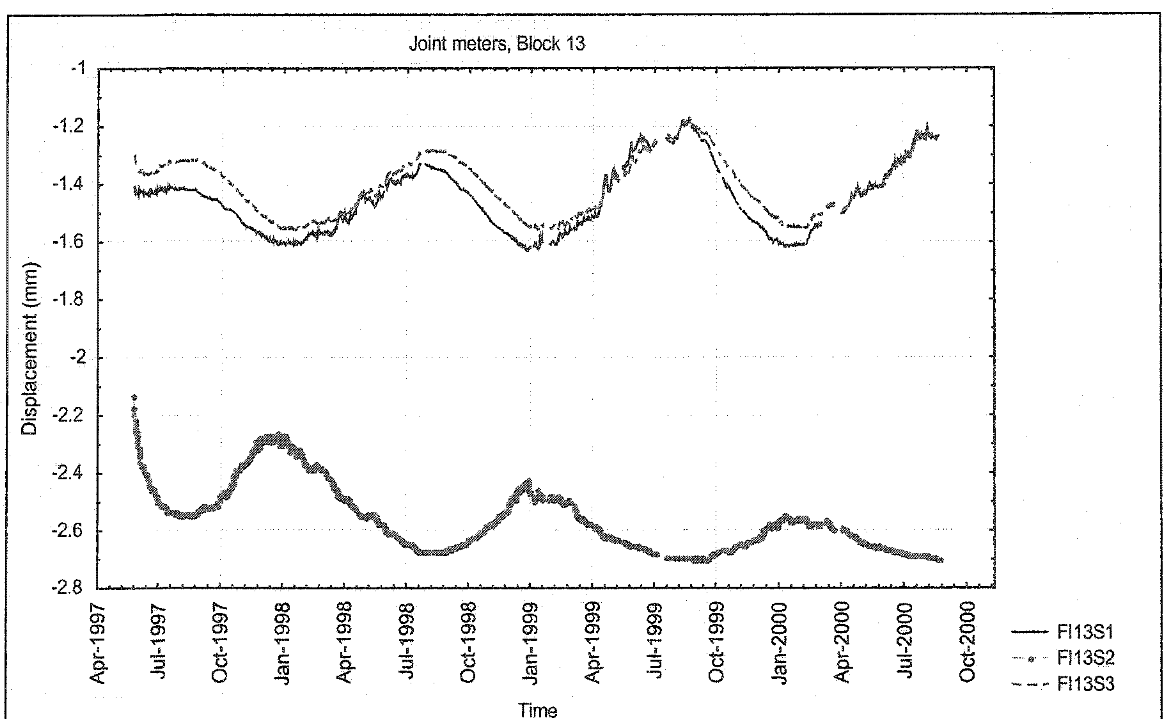


Figure 6.54, Joint meters, Chute-à-Caron dam

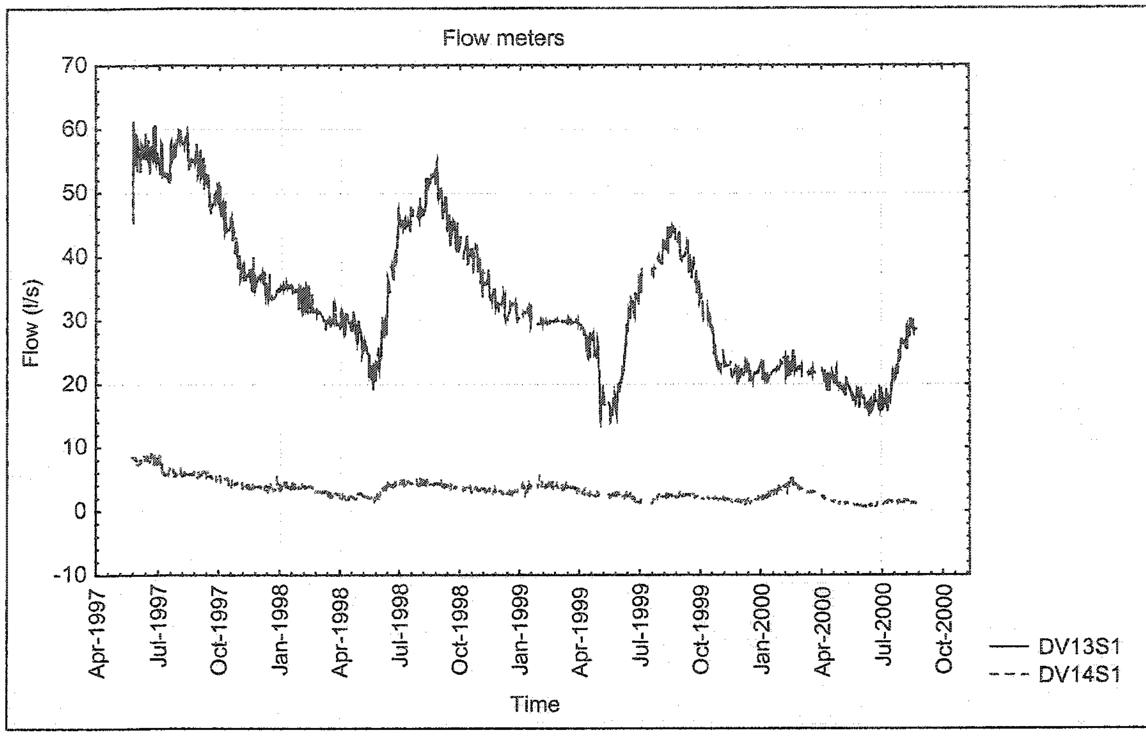


Figure 6.55, Water flow, blocks 13 and 14, Chute-à-Caron dam

Table 6.10, Correlation matrix of instruments Chute-à-Caron

	PD13S1_X	PD13S1_Y	PD13S1_Z	PD13S2_X	PD13S2_Y	PD13S2_Z	DV13S1	DV14S1	FI13S1	FI13S2	FI13S3
TEMPERATUR	-0.81	-0.52	0.61	0.80	-0.84	0.63	0.11	-0.05	0.82	-0.62	0.70
PD13S1_X	1.00	0.60	-0.64	-0.98	0.88	-0.64	0.02	0.27	-0.92	0.92	-0.84
PD13S1_Y	0.60	1.00	0.02	-0.47	0.79	0.03	0.32	0.41	-0.38	0.59	-0.24
PD13S1_Z	-0.64	0.02	1.00	0.74	-0.42	0.99	0.58	0.18	0.80	-0.55	0.93
PD13S2_X	-0.98	-0.47	0.74	1.00	-0.81	0.75	0.10	-0.20	0.94	-0.90	0.90
PD13S2_Y	0.88	0.79	-0.42	-0.81	1.00	-0.42	0.16	0.33	-0.78	0.74	-0.64
PD13S2_Z	-0.64	0.03	0.99	0.75	-0.42	1.00	0.59	0.20	0.80	-0.54	0.92
DV13S1	0.02	0.32	0.58	0.10	0.16	0.59	1.00	0.71	0.19	0.01	0.40
DV14S1	0.27	0.41	0.18	-0.20	0.33	0.20	0.71	1.00	-0.11	0.30	0.00
FI13S1	-0.92	-0.38	0.80	0.94	-0.78	0.80	0.19	-0.11	1.00	-0.74	0.94
FI13S2	0.92	0.59	-0.55	-0.90	0.74	-0.54	0.01	0.30	-0.74	1.00	-0.75
FI13S3	-0.84	-0.24	0.93	0.90	-0.64	0.92	0.40	0.00	0.94	-0.75	1.00
TE14S1	-0.82	-0.17	0.87	0.89	-0.63	0.89	0.33	0.06	0.91	-0.66	0.89
TE14S2	-0.82	-0.18	0.87	0.89	-0.63	0.89	0.33	0.06	0.91	-0.66	0.88
TE14S3	-0.82	-0.18	0.87	0.89	-0.63	0.89	0.33	0.06	0.91	-0.66	0.88
TE14S4	-0.89	-0.40	0.84	0.92	-0.79	0.85	0.27	-0.03	0.94	-0.73	0.91
TE14S5	-0.85	-0.28	0.93	0.90	-0.69	0.93	0.41	0.04	0.92	-0.73	0.97
TE12S1	-0.88	-0.54	0.67	0.87	-0.88	0.69	0.16	-0.03	0.89	-0.68	0.78
TE12S2	-0.89	-0.40	0.86	0.92	-0.79	0.87	0.32	0.00	0.94	-0.75	0.94
TE12S3	-0.70	-0.09	0.96	0.78	-0.44	0.94	0.53	0.08	0.78	-0.70	0.94

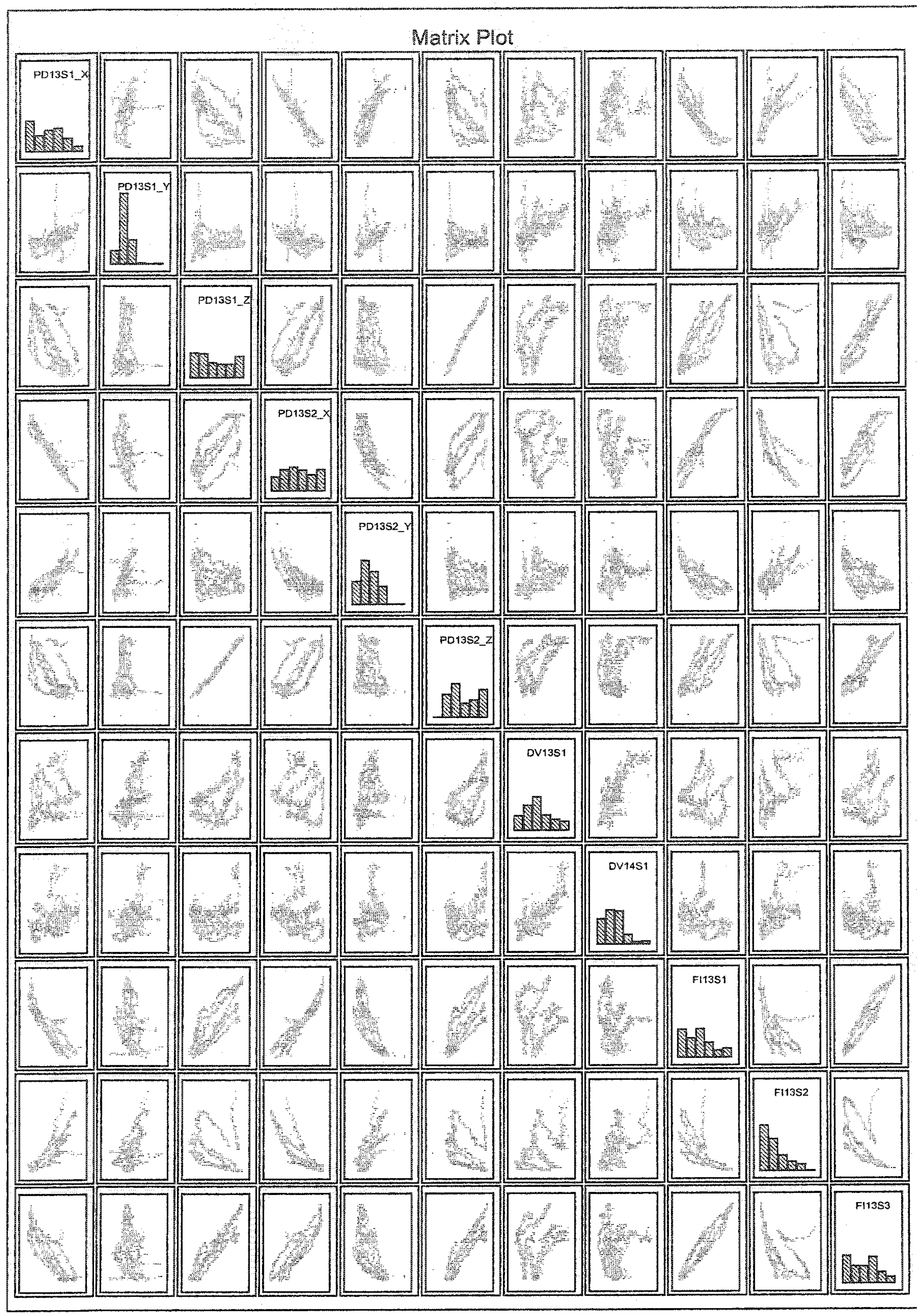


Figure 6. 56, Matrix plot of instruments, Chute-à-Caron dam

6.3.1.1 Principal component analysis

PCA was performed on the correlation matrix of pendulums, flow meters, and joint meters of block 13 of Chute- à -Caron. The first two principal components, which have eigenvalues greater than one, are retained. These two principal components explain 87% of the total variance of the original eleven instruments. The first and second principal component explains 58% and 29% of the total variance of the instruments.

Analyses of loadings (Table 6.11 and Figure 6.57) indicate that components X and Z of pendulums, and joint meters are contributing to PC1. As previously discussed these variables are highly correlated with each other, Flow meters and Y displacements of PD13S1 are correlated with PC2. These variables do not have strong correlations with the rest of instruments (Table 6.10).

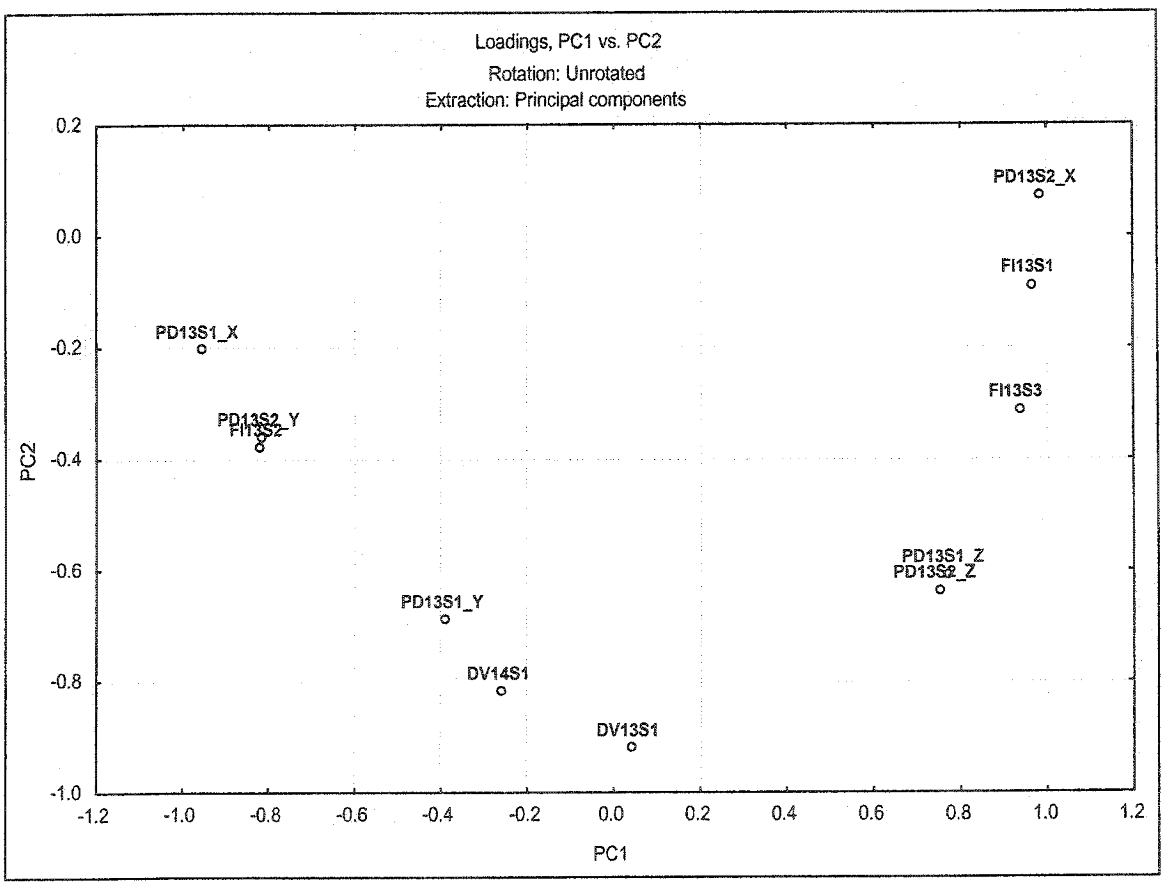


Figure 6.57, Loading of PCs 1 and 2, Chute-à-Caron dam

Table 6.11, Loading for principal components

Variable	Loadings	
	PC1	PC2
PD13S1_X	-0.96	-0.20
PD13S1_Y	-0.39	-0.69
PD13S1_Z	0.77	-0.61
PD13S2_X	0.98	0.07
PD13S2_Y	-0.82	-0.36
PD13S2_Z	0.75	-0.63
DV13S1	0.04	-0.92
DV14S1	-0.26	-0.82
FI13S1	0.96	-0.09
FI13S2	-0.82	-0.38
FI13S3	0.93	-0.31
Expl.Var	6.41	3.17
Prp.Totl	0.58	0.29

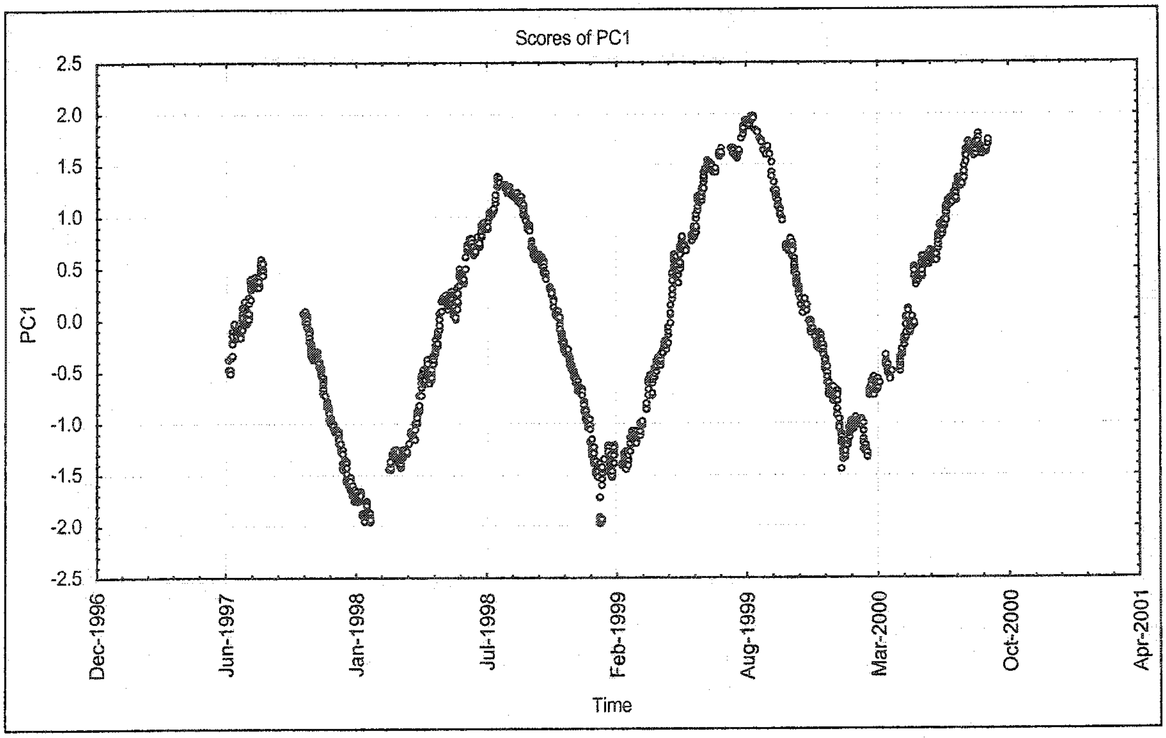


Figure 6.58, Scores of PC 1, Chute-à-Caron dam

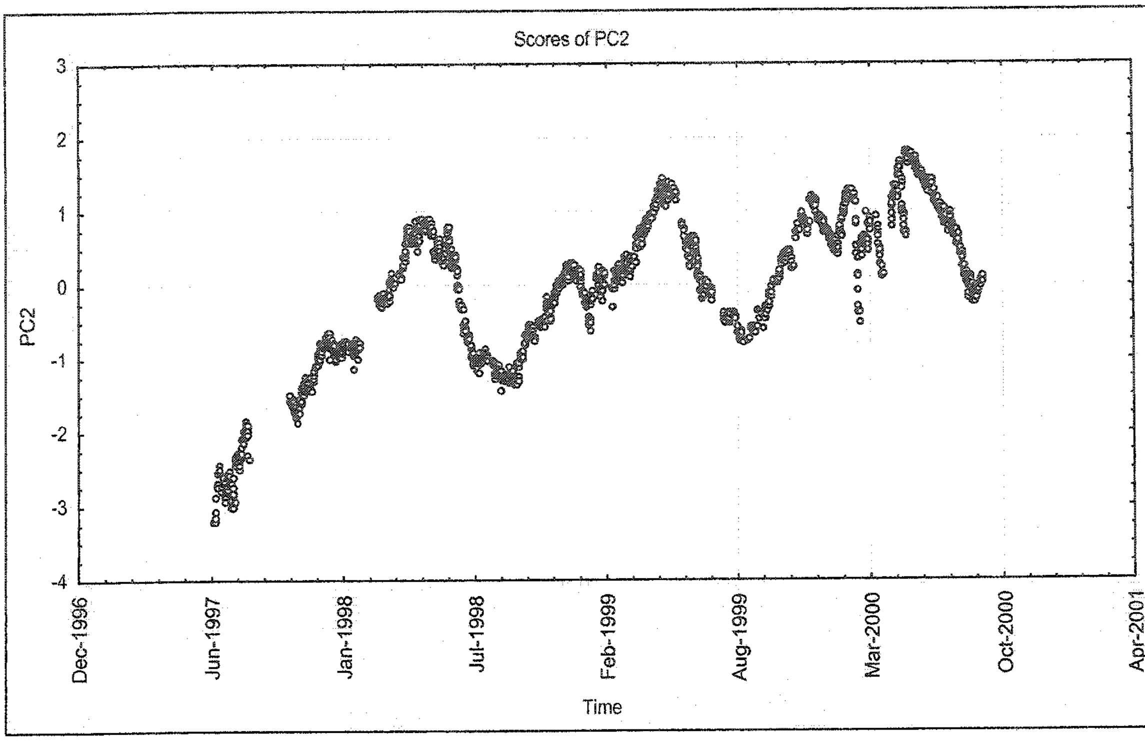


Figure 6.59, Scores of PC 2, Chute-à-Caron dam

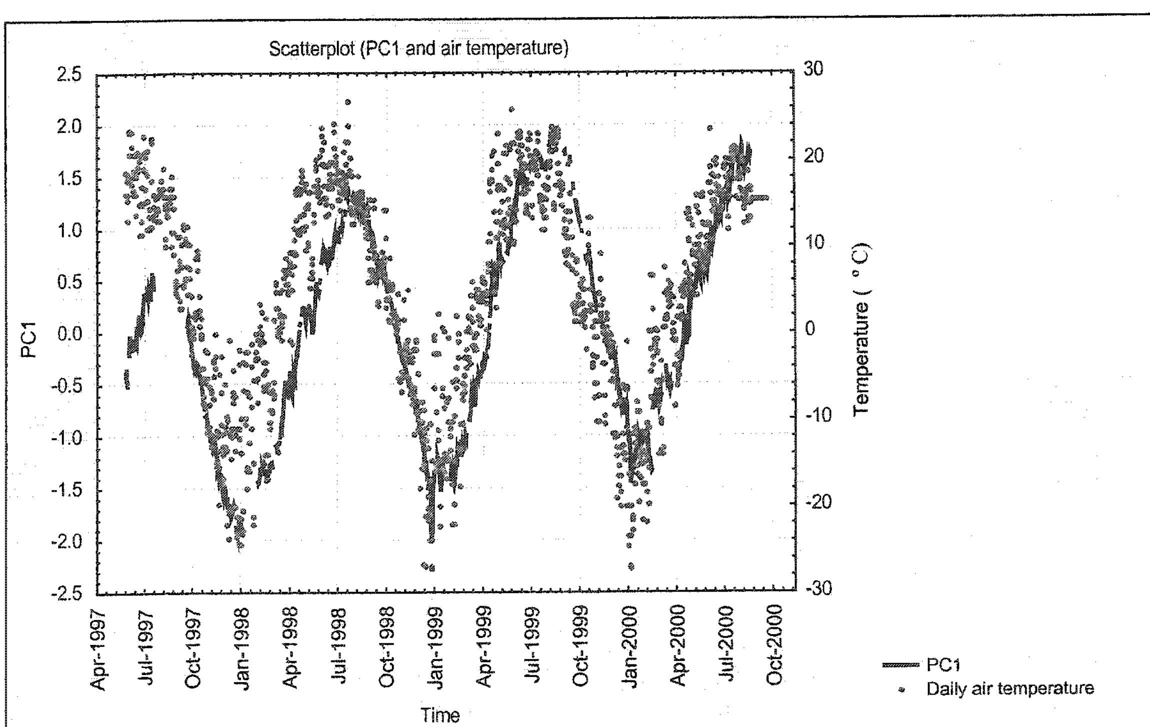


Figure 6.60, Scores of PC 1 and daily air temperature, Chute-à-Caron dam

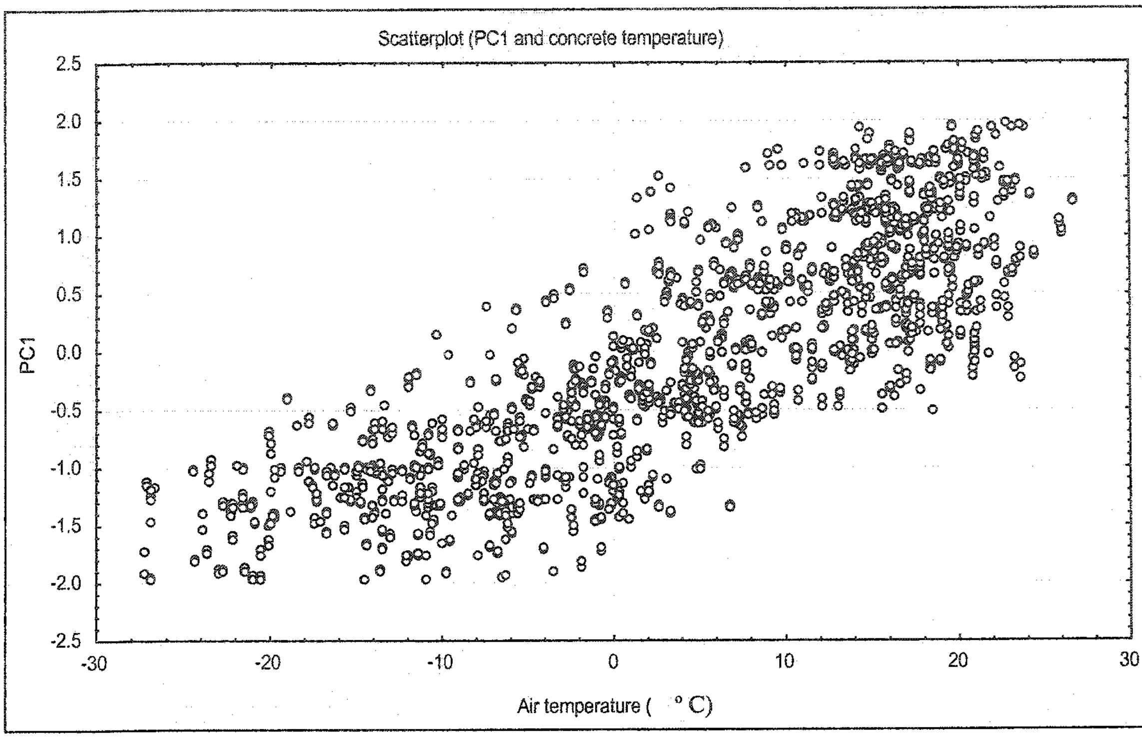


Figure 6.61, Scatter plot of PC 1 and daily air temperature, Chute-à-Caron dam

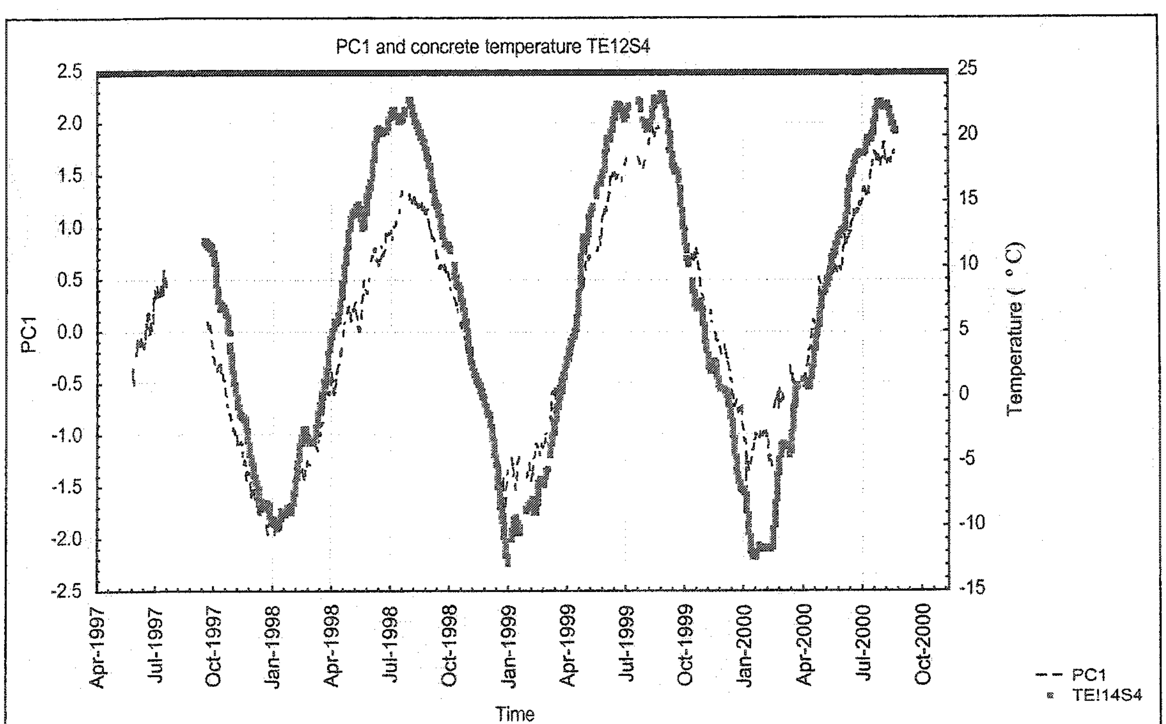


Figure 6.62, PC 1 and concrete temperature, TE14S4, Chute-à-Caron dam

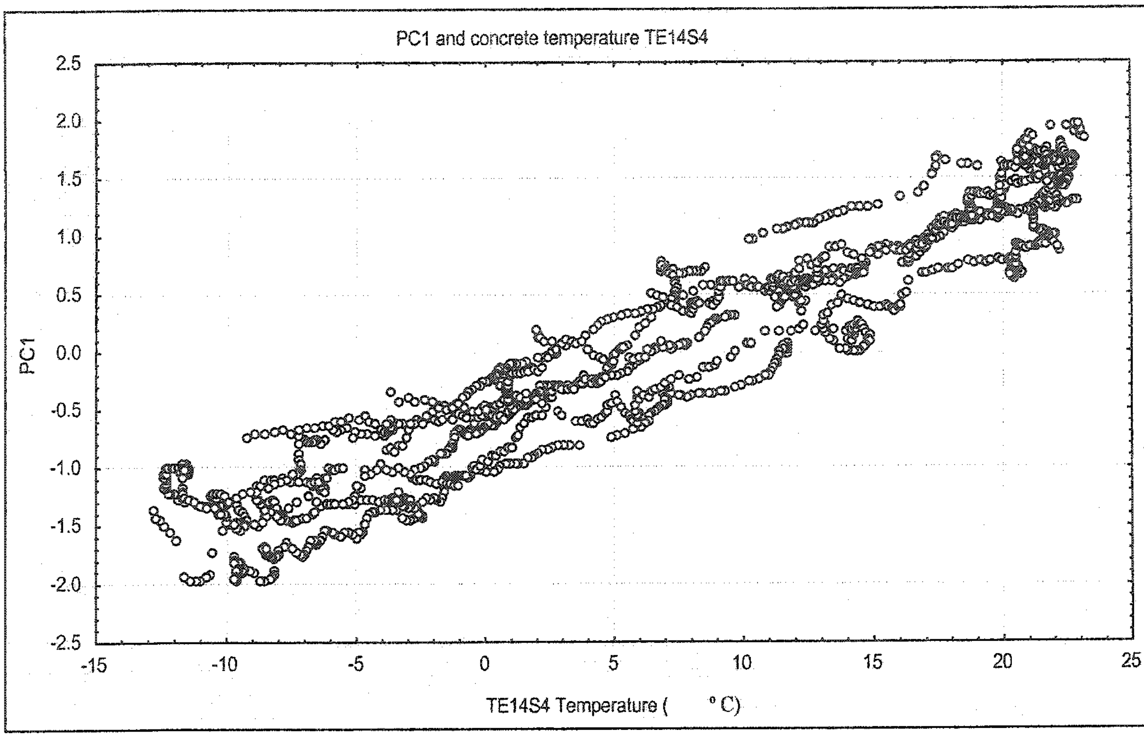


Figure 6.63, Scatter plot of PC 1 and concrete temperature, TE14S4, Chute-à-Caron dam

Scores of PC1 are highly correlated with all temperature measurements. Correlation is highest when it is compared with concrete temperatures (Figure 6.62 and Figure 6.63). Concrete temperatures are less variable than daily air temperatures and are a better predictor of dam instrument responses. These data must be collected whenever it is possible, and used in analysis of dam monitoring data. If concrete temperatures are not available, average values of daily air temperature can be used (with the optimum time lag). This is required in order to reduce the high frequency in the air temperature variations, since dam only responds to lower frequencies of temperature fluctuations

The two principal components can be used for monitoring of the dam instead of the original eleven instruments. HST regression models were developed for both PCs. Explained variances are 0.977 and 0.89 for HST models of PCs 1 and 2 respectively (Table 6.12 and 6.13). Lower R^2 on PC 2 is due to the nonlinear behavior of flow meters. Seasonal and irreversible components of models is presented in Figures 6.65 and 6.67. Irreversible component of PC1 indicate gradual closure of vertical expansion joint.

Table 6.12, Regression summary of HST model for PC 1

N=1646	Regression Summary for Dependent Variable: PC1 R= .9885 R²= .977 Std.Error of estimate: .151					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(1640)	p-level
Intercept			-1.39	0.026	-54.0	0.0000
t	0.93	0.02	2.93E-3	0.000	41.9	0.0000
t^2	-0.70	0.02	-1.33E-6	0.000	-31.4	0.0000
$\sin(\theta)$	-0.66	0.00	-0.98	0.006	-173.6	0.0000
$\cos(\theta)$	-0.73	0.00	-0.98	0.005	-188.0	0.0000
$\sin^2(\theta)$	0.06	0.00	0.18	0.011	17.0	0.0000

Table 6.13, Regression summary of HST model for PC2

N=1646	Regression Summary for Dependent Variable: PC2 R= .944 R²= .891 Std.Error of estimate: .32960					
	Beta	Std.Err. of Beta	B	Std.Err. of B	t(1639)	p-level
Intercept			-2.16E+00	0.056	-38.32	0.000
t	1.181	0.048	3.73E-03	0.000	24.45	0.000
t^2	-0.569	0.048	-1.09E-06	0.000	-11.73	0.000
$\sin(\theta)$	0.552	0.008	8.22E-01	0.012	67.00	0.000
$\cos(\theta)$	-0.124	0.008	-1.68E-01	0.011	-14.74	0.000
$\sin(\theta)\cos(\theta)$	-0.373	0.008	-1.07E+00	0.024	-45.25	0.000
$\sin^2(\theta)$	-0.089	0.008	-2.51E-01	0.023	-10.72	0.000

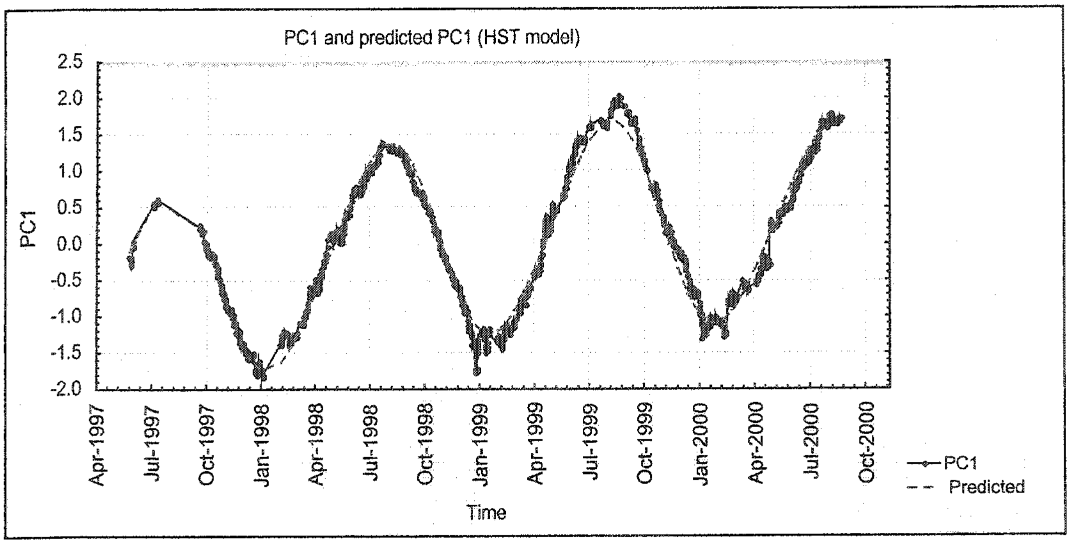


Figure 6.64, Scores of PC 1 and predicted scores using HST model, Chute-à-Caron dam

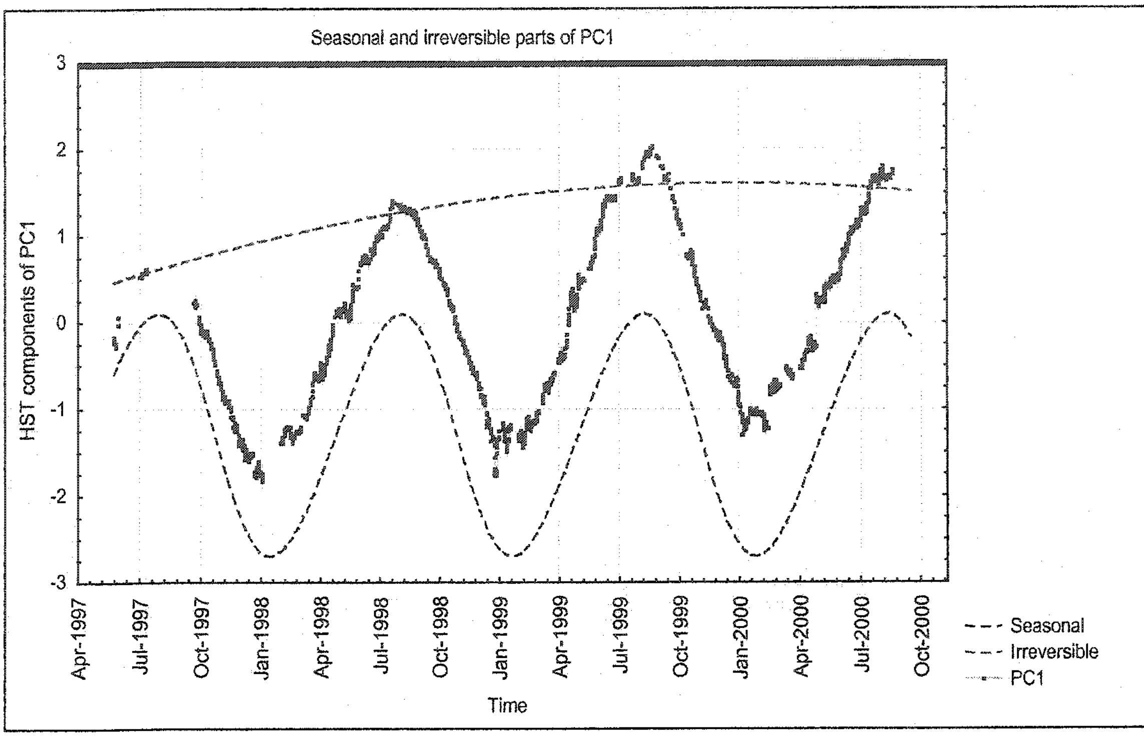


Figure 6.65, Seasonal and irreversible components of PC 1, Chute-à-Caron dam

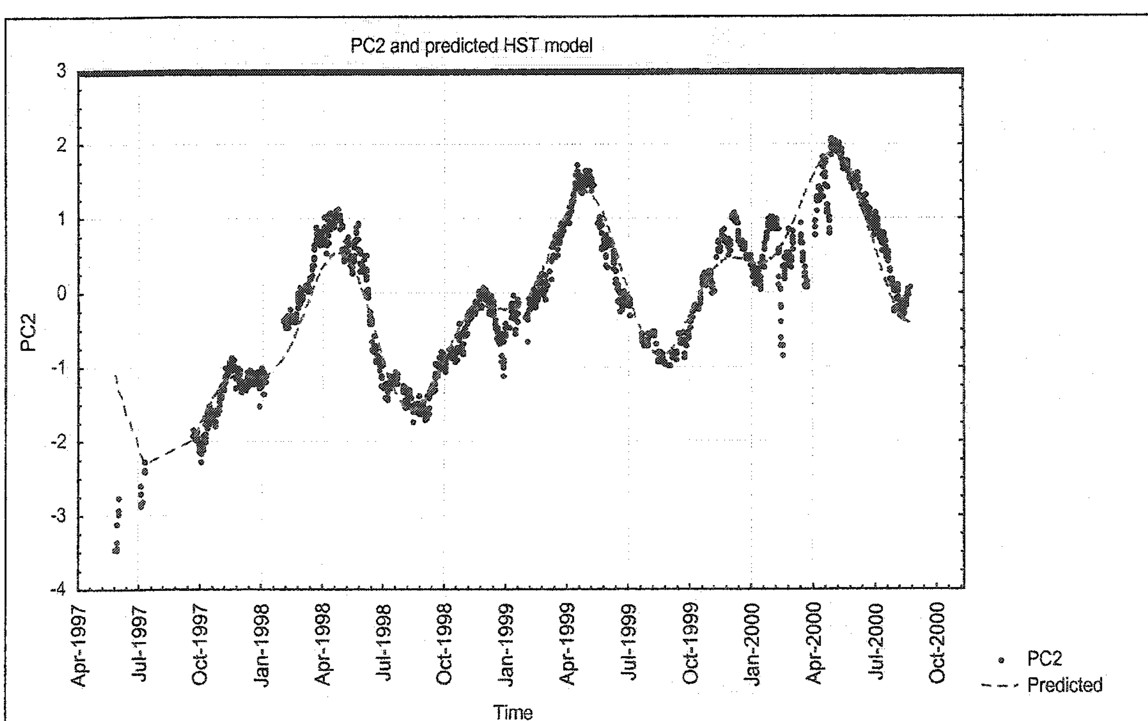


Figure 6.66, Scores of PC2 and predicted scores using HST model, Chute-à-Caron dam

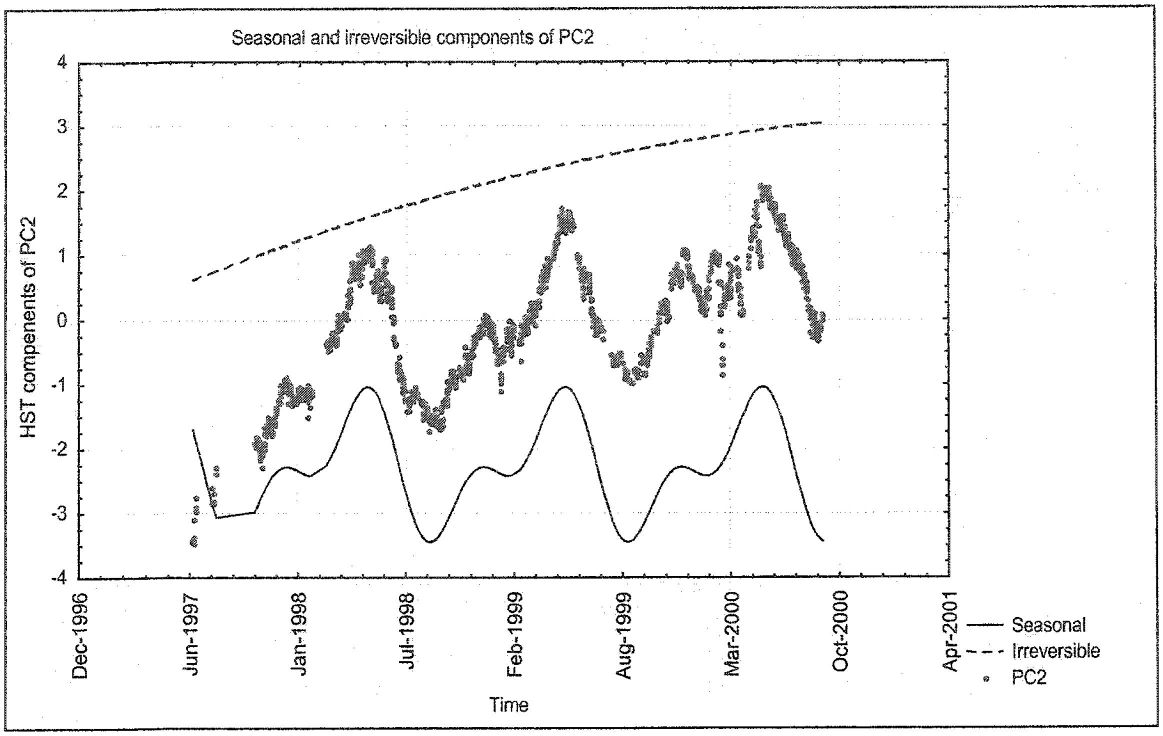


Figure 6.67, Seasonal and irreversible components of PC2, Chute-à-Caron dam

6.3.1.2 Canonical correlation analysis

Canonical correlation analysis is another multivariate method, which can be used for exploratory data analysis. Relationships between two sets of instruments are estimated and latent factors, for these two sets of instruments, are obtained.

CCA was used to explore the relationship between a set of measurements for pendulums PD13S1 and PD13S2 in three different directions (total of 6 measurements), and a second set consisting of flow meters and joint meters (total of 5 instruments).

Table 6.14, Canonical weights for displacements

	Canonical Weights, first set				
Variable	Root 1	Root 2	Root 3	Root 4	Root 5
PD13S1_X	-0.02	0.21	0.25	-0.94	0.96
PD13S1_Y	0.00	0.01	0.35	0.85	-1.28
PD13S1_Z	0.54	-1.89	-8.14	4.31	1.02
PD13S2_X	0.97	1.07	-0.58	-1.90	-0.90
PD13S2_Y	-0.05	-0.08	-1.23	-1.46	-0.67
PD13S2_Z	-0.59	0.48	8.31	-4.27	0.06

Table 6.15, Canonical weights for joint and flow meters

Variable	Canonical Weights, second set Root 1	Root 2	Root 3	Root 4	Root 5
DV13S1	-0.05	-0.02	0.69	-0.55	2.37
DV14S1	0.04	-0.04	0.54	-0.56	-1.76
FI13S1	0.68	1.37	3.43	0.65	1.67
FI13S2	-0.47	-0.71	-0.03	1.48	-0.33
FI13S3	-0.06	-2.23	-3.43	0.52	-2.57

The canonical weights of canonical factor 1 for the first set are high for displacement X of pendulum PD13S1 and Z displacements of both Table 6.14). The canonical coefficients of canonical factor 1 for the second set are high for F13S1 and F13S2 (Table 6.15). Closing of joint FI13S1 and opening of FI13S2 are associated with vertical displacements of both pendulums. Flow meters loadings are very small which indicated that there is no strong relationship between flow meters and displacements. First canonical factor is strongly correlated with measurements of TE14S4 (Figure 6.68).

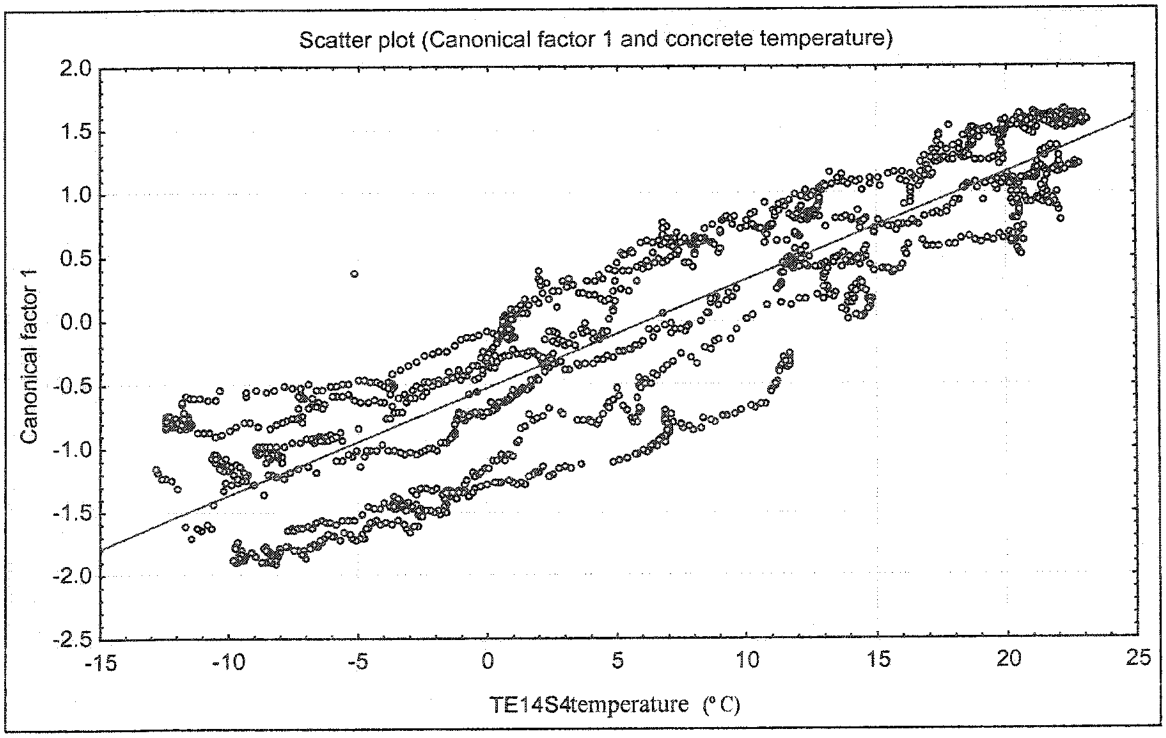


Figure 6.68, Canonical factor 1 and concrete temperature

Canonical factor1 is highly correlated with PC1. CCA results confirm the results obtained from PCA. Temperature effects are the main factor affecting the dam behavior.

6.3.1.3 Multiple linear regression results

HST models were developed for all eleven instrument of block 13. Reservoir level is almost constant during the period and therefore it is not contributing to the variations of instruments readings. Regression models were developed that consider concrete temperatures as an alternative to the seasonal component of HST models. A comparison of explained variance (R^2) by these two models is presented in Table 6.16.

Table 6.16, Comparison of results for two regression models

	Adjusted R squares	
	Model 1 HST	Model 2 concrete temperatures
PD13S1-X	0.95	0.96
PD13S1-Y	0.3	0.45
PD13S1-Z	0.98	0.98
PD13S2-X	0.99	0.99
PD13S2-Y	0.87	0.9
PD13S2-Z	0.97	0.96
DV13S1	0.88	0.82
DV14S1	0.7	0.62
FI13S1	0.91	0.96
FI13S2	0.92	0.95
FI13S3	0.96	0.98

Both models are performing well as R^2 is higher than 0.9 for most of the instruments, except flow meters and PD13S1-Y. However this can be expected, as flow meters did not have strong correlation with temperatures. Both models performed poorly on PD13S1-Y, which exhibits unusually high values in Feb to March 2000. These readings do not conform to the previous readings. Other instruments do not indicate any change in dam behavior at that particular time. These readings are most likely outliers due to interference with the wire of the pendulum.

6.3.1.4 Artificial Neural networks Applications

A back propagation network was used to build a model for forecasting displacements of pendulum PD13S2 in the X, Y and Z direction. Several models were tested using a back propagation neural network with sigmoid transfer functions. Preliminary analysis showed that time and concrete temperatures are the best predictors for the displacement of the pendulums. Concrete temperatures at five locations and time were used as input variables to forecast the displacements of the pendulums.

Unlike statistical forecasting methods, the neural network does not follow well-defined development guidelines. Therefore the model building process is generally conducted by trial and error. Two major elements of the back propagation neural network models are the number of hidden layers, and the number of hidden nodes.

Number of hidden layers

As the number of hidden layers is increased, the training data can be predicted very accurately. However, the prediction skill of the validation set can decrease, as the model gets more complex. The lack of fit of the validation set occurs due to overfitting. The optimal number of layers must strike a balance between overfitting and accuracy (Neuralware, 1996).

Using more than one hidden later increases the complexity of the network, computational time and possibility of model overfitting. Most problems require only one, and sometimes two layers. Generally it is better to start with one hidden layer and increase the number of hidden nodes until satisfactory performance is reached. If the results of one hidden layer are not satisfactory, more hidden layers must be added to the network configuration. In this application, a single hidden layer was tried first and results were found to be satisfactory.

Number of hidden nodes

It is suggested that the number of hidden nodes must be smaller than the number of input nodes (Neuralware, 1996). Since six inputs were used as dependant variables, the number of hidden nodes was increased incrementally between one and six. All the networks were trained until the best performance was achieved.

Performance of the network was measured using R^2 statistics, i.e. percentage of the variance explained by the model. The data was divided into two equal sets. The first set was used to train the network, and the second set was used to validate the model. R^2 statistics were calculated for both the training and validation sets. Every network was trained several thousand times until the model converged and best results were obtained.

Prediction model for PD13S2 in x direction

Results of R^2 for different network configurations are presented in Table 6.17. All the networks, which have more than one hidden unit, produce very satisfactory results. Since the model with two hidden nodes is simpler than the other three models and results are comparable, it was selected as the best model (Figure 6.69).

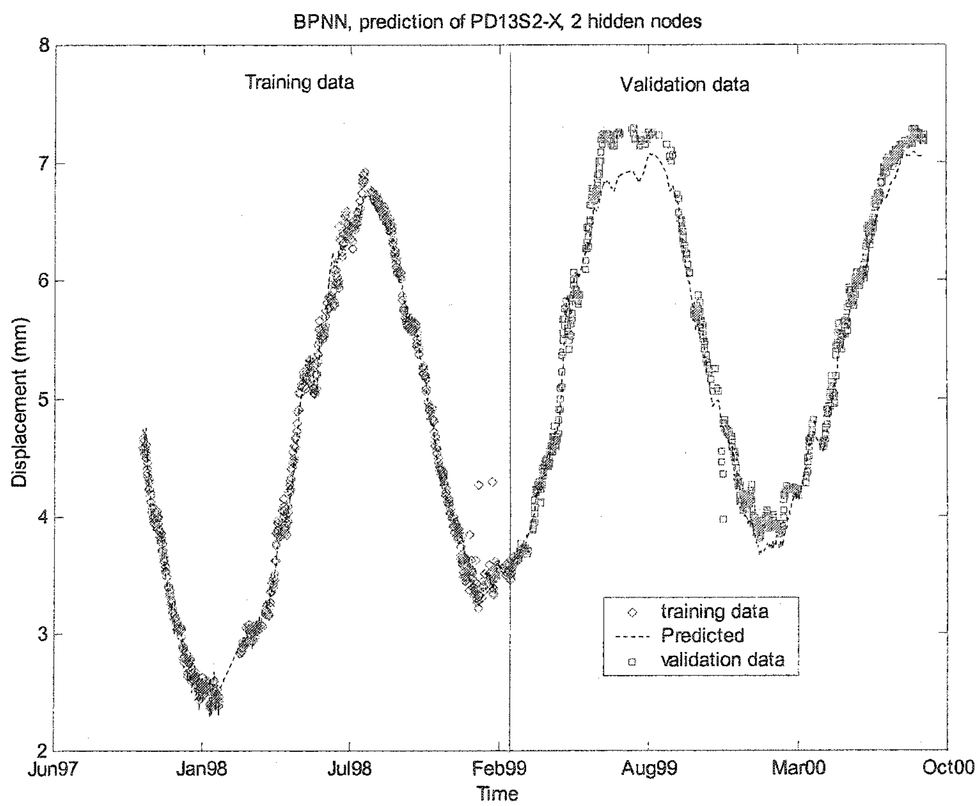


Figure 6.69, Predicted and observed values of pendulum m PD13S2 (X Direction)

Prediction model for PD13S2 in Y direction

Results of R^2 for different network configurations are presented in Table 6.18. All the networks, which have more than one hidden unit produce very satisfactory results. As the case for PD132_X, the model with two hidden nodes is simpler than the other three models and results are comparable, it was selected as the best model (Figure 6.70).

Table 6.17, Variance explained (R^2) for PD13S2_X, different Model configurations

Number of hidden nodes	Training data	Validation data	All data
1	0.982	0.928	0.957
2	0.994	0.971	0.983
3	0.993	0.980	0.978
4	0.996	0.967	0.981
5	0.992	0.972	0.983

Table 6.18, Variance explained (R^2) for PD13S2_Y different model configurations

Number of hidden nodes	Training data	Validation data	All data
1	0.925	0.854	0.892
2	0.962	0.896	0.931
3	0.979	0.897	0.933
4	0.986	0.890	0.941
5	0.974	0.914	0.946

Prediction model for PD13S2 in Z direction

The data was divided into two equal sets. The first set was used to train the network, and the second part was used to validate the model. Results of R^2 for different network configurations are presented in Table 6.19. Like the previous two models, all the networks, which have more than one hidden unit, produce very satisfactory results. The model with 3-hidden nodes was selected as the best model as it produced better predictions at the peaks (Figure 6.71).

Table 6.19, Variance explained (R^2) for PD13S2_Z, and different model configurations

Number of hidden nodes	Training data	Validation data	All data
1	0.945	0.920	0.931
2	0.982	0.946	0.963
3	0.994	0.963	0.971
4	0.996	0.965	0.978
5	0.996	0.961	0.980

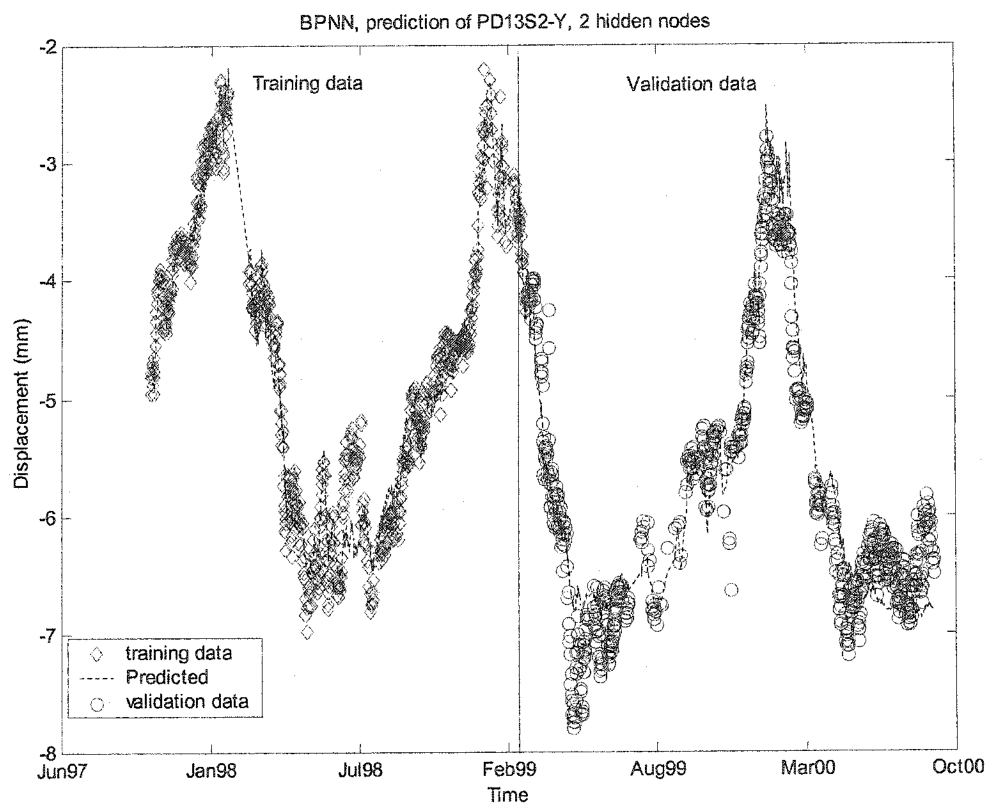


Figure 6.70, Predicted and observed values of pendulum m PD13S2 (Y direction)

Generally one-layer networks with 2 or 3 nodes provided good predictions of pendulum displacements. Explained variances were higher for X and Y direction as vertical displacements are not as highly correlated with temperatures as X and Y displacements.

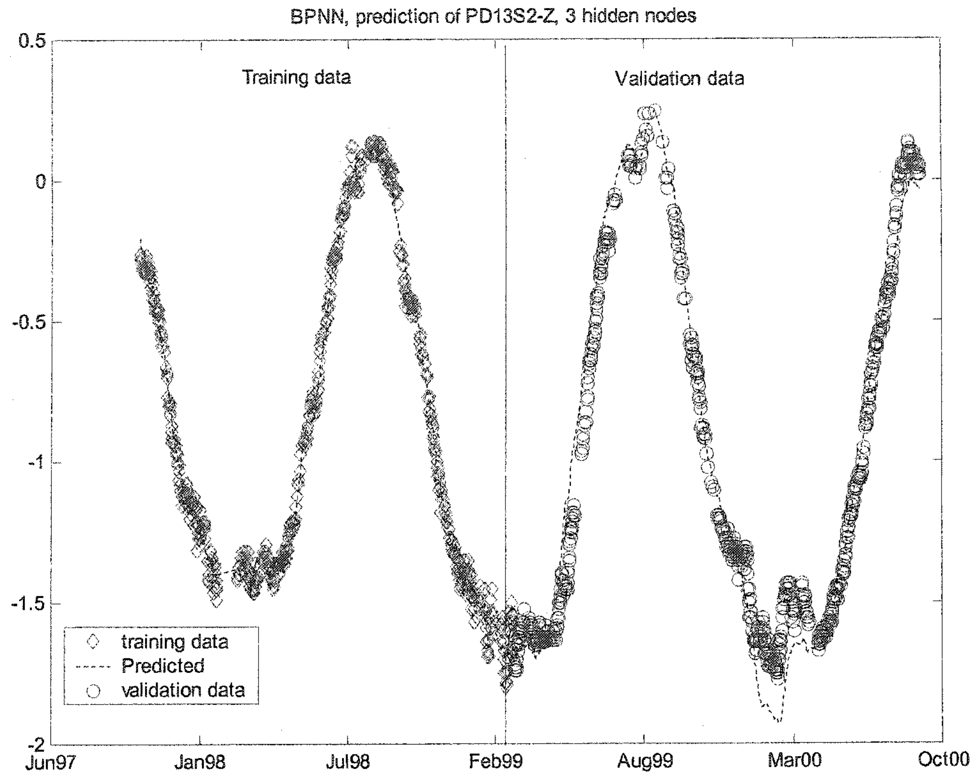


Figure 6.71, Predicted and observed values of pendulum m PD13S2 (Z direction)

Once a model is built by neural networks it can be used to monitor future observations. Discrepancies with observations can be used as an indication of possible anomalies. The neural networks are generally computationally demanding, and training of the networks `takes more time than statistical models. Neural networks are difficult to interpret physically and are not as useful as statistical models to estimate contributions for reservoir fluctuation, temperature effects, and irreversible effects on the total response.

CHAPTER 7

7. Summary and Recommendations

7.1 Summary

Univariate and multivariate statistical methods are used to analyze the behavior of concrete dam. Statistical models such as multiple linear regression, and back propagation neural networks have been used to estimate the response of individual instruments. Multiple linear regression models are of two kinds, (1) Hydro-Seasonal-Time (HST) models and (2) models that consider concrete temperatures as predictors. Univariate, bivariate, and multivariate methods are proposed for the identification of anomalies in instrumentation data. The source of these anomalies can be either from bad readings, faulty instruments, or changes in dam behavior.

Multivariate statistical analysis methods are applied to three different dams, Idukki, Daniel Johnson, and Chute-à-Caron, which are respectively an arch, multiple arch and a gravity dam. Displacements, strains, flow rates, and crack openings of these three dams are analyzed.

Data reduction

Multiple instruments are often used for monitoring the behavior of a dam. The responses of these instruments are often correlated as they are affected by common factors. The response of a dam to external factors can be grouped into reversible and irreversible effects. Reversible effects are usually correlated with reservoir level variations, and air and water temperature variations. Irreversible effects are time dependent, and are due to creep, shrinkage, settlement, and chemical reactions such as alkali aggregate reactions. The simultaneous analysis of instrumentation data was performed using principal component analysis on instrumentation data from three different dams (Table 7.1). Generally less than four factors were needed to explain as much as 90% of the total variance. The unexplained variance is due to noise levels in the instruments and possibly localized behavior.

Table 7.1, comparison of PCA results, three different dams

Type of dam	**Instruments**	**Number of original Instruments**	**Number of PC's**	**Explained Variance**
Idukki	Instrumented cylinders Pendulums	24 16	4 4	91% 92%
Daniel Johnson	Pendulums Extensometers	8 8	3 2	95% 93%
Chute à Caron	Flow meters Displacements Join meters	2 6 3	2	87%

These three analyses on three types of dams indicate that principal component analysis is effective in data reduction for dam monitoring data. These principal components can be effectively used to monitor a dam instead of monitoring all the individual instruments. Confidence intervals can be used to predict the expected minimum and maximum bounds for the scores of each principal component. If principal components behave normally there is no need to review each individual instruments. If a statistically significant change is detected in one of the principal components, individual instruments highly correlated with the principal component can be reviewed. This can significantly reduce the time and expense of dam surveillance.

Individual analysis of Instruments

Models such as multiple linear regression, and back propagation neural networks are used to estimate the response of individual instruments. Multivariate linear regression models are of two kinds, HST models and models that consider concrete temperatures as an alternative to the seasonal model of HST. Several methods must be taken into consideration for choosing a prediction model prediction accuracy, time, cost, and simplicity of prediction method.

The HST method is the easiest to develop and provides satisfactory results in many cases but it is inefficient for predicting responses for abnormal temperature cycles. In contrast, back propagation neural networks are better suited for modeling non-linear relationship, more computationally demanding and training of the networks takes more time than statistical models. Neural networks are difficult to interpret physically and are not as useful as statistical model to estimate contributions for reservoir fluctuation, temperature effects, and irreversible effects to the total response.

Fault detection and faulty instruments

Univariate, bivariate, and multivariate methods are proposed to identify anomalies in instrumentation data. These anomalies can be due to either bad readings or an indication of change in behavior of the dam. Current practice methods are based on univariate methods. The use of bivariate and multivariate methods is an important contribution to dam monitoring methodology.

Collection of the data, and frequency of reading

Concrete temperatures are less variable than daily air temperatures and were shown to be a better predictor of dam instrument response. These data must be collected whenever it is possible, and used in analysis of dam monitoring data. If concrete temperatures are not available, daily air temperature must be recorded, and average values are used (with the optimum time lag). This is required in order to reduce the high frequency in the air temperature variations, since dam only responds to lower frequencies of temperature fluctuations.

Available data for stress and strain meters of Idukki dam were only recorded twice per year, which makes the estimation of parameters very difficult. Whenever possible all instruments should be recorded at least weekly. Simultaneous analysis of such a data set can provide insight about the behavior of the dam and the correlation of readings between different instruments. Available data for instrumented cylinders of Idukki dam were available for a period of 17 months, which makes the estimation of seasonal effects very difficult. The minimum length of the data needed for a meaningful analysis is at least three years of data to separate reversible and irreversible effects.

7.2 *Recommendation for future research*

- Multivariate statistical methods were proved to be useful in data reduction for instrumentation in concrete dams. The simultaneous analysis of different types of instrumentation could not be considered in many cases due to limitations in data availability and quality. More studies are needed and more data should be collected with an appropriate frequency of readings to develop these methods further and integrate them into regular dam behavior studies that are monitored by government control.
- More research is needed to improve the near real time surveillance of the dam using monitoring data and for setting a hierarchy of alarm levels that allows an early detection of potential dam safety related problems while minimizing costly false alarms.
- Measurements complement visual observations as a continuing surveillance system of the threat to life, property, and the environment, and assist in investigating unexpected or abnormal performance. Research is also needed to integrate qualitative visual inspection observations and dam instrumentation data in an efficient dam surveillance methodology.

- Application of multivariate methods must be further developed to model the changes in dam responses.
- Setting a methodology based on modes of failure, A methodology is needed to be developed to assign1) instrumentation which can detect different failure modes, 2)devise analysis and diagnosis tools based bivariate/ multivariate statistical methods

7.3 Contribution to the Knowledge

In most countries throughout the world, interest in dam safety has risen significantly in recent years. Aging dams, new hydrologic information, dam construction and population growth in flood plain areas downstream from dams have resulted in an increased emphasis on dam safety, operation and maintenance.

Instrumentation data is often accumulated, but its engineering significance is not fully exploited in dam surveillance. The output of dam monitoring system, which is a main part of dam surveillance, has to be thoroughly analyzed to alert dam wardens of any possible anomalies. The need for effective analysis tools of dam monitoring data was recently emphasized in the latest International Commission of Large Dams (*Dibiagio, 2000)*. Dam monitoring practice has not been keeping pace with recent advances in statistical analysis methods. 'There is a need to develop new analysis tools to help dam safety engineers in the evaluation of dam behaviors.'

Original contribution of this research can be divided into three main areas:

(1) Data reduction using multivariate statistical methods and in particular principal component analysis

(2) Anomaly detection

(3)-Application of Artificial Neutral Networks to dam monitoring data

Principal component analysis can be used as a powerful tool in analysis of dam monitoring data. Principal components can be effectively used to monitor a dam instead of monitoring all the individual instruments. If a statistically significant change is detected in one of the principal components, individual instruments highly correlated with the principal component can be reviewed. This can significantly reduce the time and cost of dam surveillance.

The second major contribution is with respect to detection of anomalies in instruments readings. In dam safety practice these alarm levels are chosen based on analysis of each individual instrument. However, as some of the measurements are noisy or unreliable, this approach increases the chance of randomly finding an instrument out of control. Thus the false alarm rate (or probability of Type 1 error) is increased if each variable is analyzed and controlled separately because the more variables there are, the more likely it is that one of these instruments may be out of control and indicate an adverse condition when the dam is actually in a safe state. Multivariate statistical methods overcome some of these difficulties and reduce the probabilities of false alarms generated by a dam monitoring system.

Finally, the third contribution is application of Artificial Neural Networks to dam monitoring data. Back propagation neural networks are used as an alternative estimation method in analysis of dam monitoring data. Back propagation neural networks can be effectively used to model linear and non-linear relationships. However, they are more computationally demanding and training of the networks takes more time than statistical models.

References

Andersen, G. R., Chouinard, L. E., Bouvier, C., Edward Back, W., (1999), "Ranking Procedure on Maintenance Tasks for Monitoring of Embankment Dams", Journal of Geotechnical and Geoenvironmental Engineering, April 1999, p. 247-259.

Ballivy, G., Benmokrane, B., Poulin, R., Saleh, K., (1990), "Une Technique Nouvelle d'Inclusion pour la Measure à Long Terme des Déformations dans des Barrages en Béton", Canadian Journal of Civil Engineering, V. 17, No. 6, p. 919-930.

Barnett, V., Lewis, T., (1994), "Outliers in Statistical Data", Chichester; New York: Wiley.

Barrie R. E., (1995), "Structural Safety Evaluation of Gerber Arch Dam", WaterPower 95: Proceeding of the International Conference on HydroPower, V. 3, p. 2147-2156.

Blas, J. R., (1989), "Analysis of Anomalies in the Behavior of an Arch Dam under Concrete Expansion", International Commission on Large Dams, Symposium on Analytical Evaluation of Dam Related Safety Problems, Copenhagen.

Bonelli, S., Royet, P., (2001), "Delayed Response Analysis of Dam Monitoring Data", Dams in a European Context, Swets and Zeitlinger, Lisse, p. 91-99.

Bordes, J. L., Capelle, J. F., Choquet, P., (1998), "Dam Monitoring and Instrumentation", Roctest Ltd.

Bourdarot, E., (2001), "Analysis of the Irreversible Deformations of Arch Dams", Dams in a European Context, Swets and Zeitlinger, Lisse, p. 101-106.

Bouzoubaâ, N., Lachemi, M., Aitcin, P. C., (1997), "Thermal Damage of Mass Concrete: Experimental and Numerical Studies on the Effect of External Temperature Variations", Canadian Journal of Civil Engineering, V. 24, p. 649-657.

Bowles, D. S., Anderson, L. R., Glover, T. F., Chauhan, S. S., (1999), "Understanding and Managing the Risks of Aging dams: Principles and Case Studies", 19th US Commission on Large Dams, Annual Meeting and Lecture, Atlanta, Georgia.

Chouinard, L. E., Bennet, D., Feknous, N., (1995), "Statistical Analysis of Monitoring Data for Concrete Arch Dams", Journal of Performance of Constructed Facilities, Nov. 1995, p. 287-301.

Chouinard, L. E, Ahmadi Nedushan, B., Feknous, N., Sesha, B., (1996), "Statistical Analysis in Real Time of Monitoring Data for Idukki Arch Dam", 2nd International Conference on Dam Safety Evaluation, Nov. 1996, India.

Comité Suisse des Barrages (2000), "Méthodes de Prédiction du Comportement des Barrages: Comparaison Mesuré-Calculé", Zurich.

Crépon, O., Lino, M., (1999), "An Analytical Approach to Monitoring, International Water Power and Dam Construction, June 1999, p. 52-54.

Dunnicliff, J., (1988), "Geotechnical Instrumentation for Monitoring Field Performance", John Wiley, New York.

De Jong, S., (1993), "An Alternative Approach to Partial Least Squares Regression", Chemometrics and Intelligent Laboratory Systems, V. 18, p. 251-263

Diamantaras, K. I., Kung, S. Y., (1996), "Principal Component Neural Networks, Theory and Application", John Wiley & sons.

Dibiagio, E., (2000), "Monitoring of Dams and Their Foundations", 20th International Commission on Large Dams, Beijing, Q. 78, General Report.

Dong, D., McAvoy, T. J., (1996), " Nonlinear Principal Component Analysis: Based on Principal Curves and Neural Networks", Computer and Chemical Engineering, V. 20, p. 65-78.

Efthimiadu, I., Tham, M. T., Adgar, A., Cox, C. S., (1995), "Integrating Statistical Engineering Control Techniques", Measurement and Control, V. 28, April 1995, p. 78-82.

Emery, W. J., Thomson, R. E., (2001), "Data Analysis Methods in Physical Oceanography", Elsevier Science.

Enel, (1980), "Behavior of Enel's Large Dams", Enel's Report, Italian National Power Board, Roma, Italy.

Fan, J. Y., Nikolaou, M., White, R. E., (1993), "An Approach to Fault Diagnosis of Chemical Processes via Neural Networks", AICHE Journal, V. 39, No. 1, p. 82-88.

Fanelli, M., (1998), "Sensivity of High Arch Dams to Base Cracking: Circular versus Parabolic Arches", Hydropower and Dams, Issue 5, p.101-104.

Fanelli, M. A., Giuseppetti, G., Riccioni, R., (1979), 13th International Commission on Large Dams, New Delhi, Q. 49, R. 44, p. 663-677.

Flood, I., Kartam, N., (1994), "Neural Networks in Civil Engineering. II: Systems and Application", Journal of Computing in Civil Engineering, V. 8, No. 2, April 1994, p. 149-162.

Galedi, P., Kowalski, B. R., (1986), "Partial Least-Squares Regression: a Tutorial", Analytica Chimica Acta, V. 185, p. 1-17.

Gilg, B., Sinniger, R., Gavard, M., Torrione, J., Stucky, J. P., (1982), "Long Terms Measurements on Three Swiss Dams, Mauvoisin, Grande Dixence and Mattmark", International Commission on Large Dams, Rio de Janeiro.

Goguel, B., Mpala, A. S., (1992), "Monitoring and Maintenance at Kariba Dam", Water Power & Dam Construction, June 1992, p. 50-54.

Gresz, I., (1993), "Analysis of Structure Surveillance Data by the GARIN Method", International Workshop on Dam Safety Evaluation, V. 3, p. 63-69.

Guedes, Q. M., Coelho, P. S. M., (1985), "Statistical Behavior Model of Dams", 15th International Commission on Large Dams, Q. 56, R. 16, Lausanne.

Hair, J. F., Anderson, R. E., Tatham, R. L., Black, W. C., (1995), " Multivariate Data Analysis with Readings", Prentice-Hall.

Hanna, T. H., (1985),"Field Instrumentation in Geotechnical Engineering", Trans Tech Publication, Clausthal-Zellerfeld, Federal Republic of Germany.

Hebb, D. O., (1949), "The Organization of Behavior", Wiley, New York.

Hudnut, K. W., Behr, J. A., (1998),"Continuous GPS Monitoring of Structural Deformation at Pacoima Dam", California, Seismological Research Letters, V. 69, No. 4, p. 299-308.

Hulea, D., Finkelstein, I., Mihaesteanu, E., Stematiu, D., (2000), "Behavior Models for Tarnita Arch Dam", 20th International Commission on Large Dams, Beijing, p. 467-477.

Jackson, J. E., (1991), " A User's Guide to Principal Components", John Wiley & sons.

Kalkani, E. C., (1989), "Polynomial Regression to Forecast Earth Dam Piezometer

Levels", Journal of Irrigation & Drainage Engineering, V. 115, Aug. 1989, p. 545-55.

Kerala State Electricity Board, (1989), "Idukki Dam Monitoring Report", No. 8.

Kramer, M. A., (1991), "Nonlinear Principal Component Analysis Using Autoregressive Neural Networks", AICHE Journal, V. 37, No. 2, p. 233-243.

Kresta, J. V., Macgregor, J. F., Marlin, T. M., (1991), "Multivariate Statistical Monitoring of Process Operating Performance", The Canadian Journal of Chemical Engineering, V. 69, Feb. 1991, p. 35-47.

Kutzner, C., (1997),"Earth and Rockfill Dams Principles Design and Construction", A.A. Balkema publishers, Netherland.

Lan, S., Yang, J., (1997), "Nonlinear Finite Element Analysis of Arch Dam-II. Nonlinear Analysis", Advanced in Engineering Software, V. 28, p. 409-415.

Larivière, R., Routhier, L., Roy, V., Saleh, K., Tremblay, S., (1999),"Grouting of the Cracks in the Arch 5-6-Daniel-Johnson Dam", 2nd Canadian Dam Association Conference, Oct. 1999, Sudbury, Ontario, Canada.

Ledesma, A., Gens, A., Alonso, E., (1996), "Parameter and Variance Estimation in Geotechnical Backanalysis Using Prior Information", International Journal of Numerical and Analytical Methods in Geomechanics, V. 20, p. 119-141.

Léger, P., Venturelli, J., Bhattacharjee, S. S., (1993), "Seasonal Temperature and Stress Distributions in Concrete Gravity Dams. Part 1: Modeling", Canadian Journal of Civil Engineering, V. 20, p. 999-1017.

Lombardi, G., (1999), "On the Limits of Structural Analysis of Dams", Hydropower and Dams, Issue 5, p. 50-56.

MacGregor, J. F., Jaeckle, C., Kiparissides, C., Koutoudi, M., (1994), "Process Monitoring and Diagnosis by Multiblock PLS Methods", AICHE Journal, V. 40, No. 5, p. 826-838.

MacGregor, J. F., Kourti, T., (1995), "Statistical Process Control of Multivariate Processes", Control Engineering Practice, V. 3, p. 403-414.

Mamet, J. C., (1989), "Nonlinear analysis of the Daniel Johnson Dam", Proceedings ICOLD International Symposium on Analytical Evaluation of Dam Related Safety Problems, Copenhagen, p. 205-213.

Martens, H., Naes, T., (1989), "Multivariate Calibration", John Wiley & sons.

Martin, E. B., Morris, A. J., Zhang, J., (1996), "Process Performance Monitoring Using Multivariate Statistical Process Control", IEE Process & Control Theory Applications, V. 143, No. 2.

Mason, R. L., Champ, C. W., Tracy, N. D., Wierda, S. J., Young, J. C., (1997), "Assessment of Multivariate Process Control Techniques", Journal of Quality Technology, V. 29, No. 2, April 1997.

Myers, R. H., (1990), "Classical and Modern Regression with Applications", PWS-KENT Publishing Company.

Myers, B. K., Marilley, J. M., (1997), "Automated Monitoring at Tolt Dam", Journal Civil Engineering, March 1997, p. 44-46.

Nomikos, P., MacGregore, J. F., (1994), "Monitoring Batch Processes Using Multiway Principal Component Analysis", AICHE Journal, V. 40, No. 8, p. 1361-1375.

National Research Council, (1983), "Safety of Existing Dams Evaluation and Improvement", National Academy Press, Washington D.C.

NeuralWare, (1996), "NeuralWorks Advanced Reference Guide", NeuralWare, Inc.

Paxton, J. F., (2001), "Milliken Dam: Structural Monitoring of a Concrete Arch Dam", 10th FIG International Symposium on Deformation Measurements, p. 193-202.

Pedro, J. O., Pina, C. A. B., Oliveira, S. B. M., (1985), "Structural Safety of a Cracked Arch Dam", 15th International Commission on Large Dams, Q. 56, R. 14, Lausanne.

Post, G., (1985), "Dams And Foundation Monitoring", 15th International Commission on Large Dams, Q. 56, General Report, Lausanne.

Poulin, R. M., Cupitt, P., Lateef, A., (1985), "Instrument Reliability Monitoring the Data from Idukki Arch Dam", 15th International Commission on Large Dams, Q. 56, R. 5, p. 109-124, Lausanne.

Rencher, A. C., (1998), "Multivariate Statistical Inference and Applications", John Wiley & sons.

Roehm, L. H., (1968), "Deformation Measurements of Flaming Gorge Dam", Journal of Surveying and Mapping Division, Jan. 1968, p. 37-48.

Rummelhart, D. E., McClelland, J. L., (1988), "Parallel Distributed Processing", Explorations in the Microstructure of Cognition, V. 1, MIT Press.

Silva Gomes, A. F., Silva Matos, D., (1985), "Quantitative Analysis of Dam Monitoring Results, State of the Art, Applications and Prospects", 15th International Commission on Large Dams, Q. 56, R. 16, Lausanne.

Silveira, A. F., Pedro, J., (1964), "Quantitative Interpretation of Results Obtained in the Observation of Concrete Dams", 8th International Commission on Large Dams, Q. 29, R. 43, Edinburg.

Simard, R., Feknous, N., Poulin, R.,Ballivy, G., Saleh, K., (1993) ,"Réinstrumentation du Barrage de Idukki, Inde", 3ème Colloque sur la Réflection des Infrastructures en Béton, Centre de Recherche Interuniversitaire sur le Béton, Sherbrook-Laval.

Stateler, J., Von Thum, L., Scott, G., Boernge, J., (1995), "Development of Performance Parameters for Dam Safety Monitoring", Geotechnical Instrumentation News, Dec. 1995, p. 44-48.

Struges, W., (1983), "on Interpolating Gappy Records for Time-Series Analysis", Journal of Geophysics. Research, V. 88, p. 9736-9740.

Sun, J., (1996), "A Multivariate Principal Component Regression Analysis of NIR Data", Journal of Chemometrics, V. 10, p.1-9.

Swingler, K., (1996), "Applying Neural Networks a Practical Guide", Academic Press.

Tahmazian, B., Yeh, C. H., Paul, W. J., (1989)," Thermal Cracking and Arch Action in Daniel Johnson Dam", International Symposium on Analytical Evaluation of Dam related Safety problems, Copenhagen, Vol. 1, p. 235-244.

U.S. Army Corps of Engineers, (1987), "Engineering and Design-Instrumentation for Concrete Structures", EM 1110-2-4300, Nov. 1987.

U.S. Army Corps of Engineers, (1993), "Engineering and Design- Seepage Analysis and Control for Dams with CH 1", EM 1110-2-1901, April 1993.

U.S. Army Corps of Engineers, (1994), "Engineering and Design-Deformation Monitoring and Control Surveying", EM 1110-1-1004, Oct. 1994.

U.S. Army Corps of Engineers, (1994), "Engineering and Design-Arch Dam Design", EM 1110-2-2201, May 1994.

U.S. Army Corps of Engineers, (1995), "Engineering and Design- Instrumentation of Embankment Dams and Levees", EM 1110-2-1908, June 1995.

Veesaert, C. J., (1997), "The Use of Performance Parameters in Monitoring the Safety of Dams", Dams and Safety Management at Downstream Valleys, Balkema, Rotterdam, p. 45-50.

Veltrop, J. A., Yeh, C. H., Paul, W. J., (1990), "Evaluation of Cracks in a Multiple Arch Dam", Dam Engineering, V. 1, Issue 5, p. 5-13

Widmann, R., (1997), "The Mathematics of Dam Safety", International Water Power and Dam Construction, May 1997.

Widrow, G., Hoff, M. E., (1960),"Adaptive Switching Circuits", IRE Western Electronic Show and Convention: Convention Record, Part 4, p. 96-104.

Wilks, D. S., (1995), "Statistical Methods in the Atmospheric Science", Academic Press.

Willm, G., Beaujoint, N.,(1967), "Les Methods de Surveillance des Barrages au Service de la Production Hydraulige d'Electricite de France, Problemes Anciens et Solution Nouvelles", 9th International Commission on Large Dams, Q. 34, R. 30, Istanbul.

Wise, B. M., (1991),"Adapting Multivariate Analysis for Monitoring and Modeling of Dynamic Systems", Ph.D. Dissertation, University of Washington, Seattle.

Wise, B. M., Gallagher, N. B., (1996), "The Process Chemometrics Approach to the Detection and Estimation of Jumps in Linear Systems", Journal of Process Control, V. 6, p. 329-348.

Zhang, P. X., Ayari, M. L., Robinson, L. C., (1997), "Role of Fundamental Heat Transfer Frequency in the Computation of Transient Thermal Fields in Massive Structures", Canadian Journal of Civil Engineering, V. 24, p.1059-1065.